

REPORT

A stele-enriched gene regulatory network in the Arabidopsis root

Siobhan M Brady^{1,2}, Lifang Zhang³, Molly Megraw⁴, Natalia J Martinez⁵, Eric Jiang¹, Charles S Yi¹, Weilin Liu¹, Anna Zeng², Mallorie Taylor-Teeple², Dahae Kim², Sebastian Ahnert⁶, Uwe Ohler⁴, Doreen Ware^{3,7,*}, Albertha JM Walhout^{5,*} and Philip N Benfey^{1,*}

¹ Department of Biology and IGSP Center for Systems Biology, Duke University, Durham, NC, USA, ² Department of Plant Biology and Genome Center, UC Davis, Davis, CA, USA, ³ Cold Spring Harbor Laboratory, Cold Spring Harbor, NY, USA, ⁴ Institute for Genome Sciences and Policy, Duke University, Durham, NC, USA, ⁵ Program in Gene Function and Expression and Program in Molecular Medicine, University of Massachusetts Medical School, Worcester, MA, USA, ⁶ Theory of Condensed Matter Group, Cavendish Laboratory, University of Cambridge, Cambridge, UK and ⁷ USDA-ARS, Robert W. Holley Center for Agriculture and Health, Ithaca, NY, USA

* Corresponding authors. D Ware, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY 11724, USA. Tel.: +1 516 367 6979; Fax: +1 516 367 8389; E-mail: ware@cshl.edu or AJM Walhout, Program in Gene Function and Expression and Program in Molecular Medicine, University of Massachusetts Medical School, Worcester, MA 01605, USA. Tel.: +1 508 856 4364; Fax: +1 508 856 5460; E-mail: marian.walhout@umassmed.edu or PN Benfey, Department of Biology and IGSP Center for Systems Biology, Duke University, French Family Science Center, Durham, NC 27708, USA. Tel.: +1 919 613 8202; Fax: +1 919 660 7293; E-mail: benfey@duke.edu

Received 1.8.10; accepted 17.11.10

Tightly controlled gene expression is a hallmark of multicellular development and is accomplished by transcription factors (TFs) and microRNAs (miRNAs). Although many studies have focused on identifying downstream targets of these molecules, less is known about the factors that regulate their differential expression. We used data from high spatial resolution gene expression experiments and yeast one-hybrid (Y1H) and two-hybrid (Y2H) assays to delineate a subset of interactions occurring within a gene regulatory network (GRN) that determines tissue-specific TF and miRNA expression in plants. We find that upstream TFs are expressed in more diverse cell types than their targets and that promoters that are bound by a relatively large number of TFs correspond to key developmental regulators. The regulatory consequence of many TFs for their target was experimentally determined using genetic analysis. Remarkably, molecular phenotypes were identified for 65% of the TFs, but morphological phenotypes were associated with only 16%. This indicates that the GRN is robust, and that gene expression changes may be canalized or buffered.

Molecular Systems Biology 7: 459; published online 18 January 2011; doi:10.1038/msb.2010.114

Subject Categories: metabolic and regulatory networks; plant biology

Keywords: Arabidopsis; gene regulatory network; miRNA; transcription factor

This is an open-access article distributed under the terms of the Creative Commons Attribution Noncommercial Share Alike 3.0 Unported License, which allows readers to alter, transform, or build upon the article and then distribute the resulting work under the same or similar license to this one. The work must be attributed back to the original author and commercial use is not permitted without specific permission.

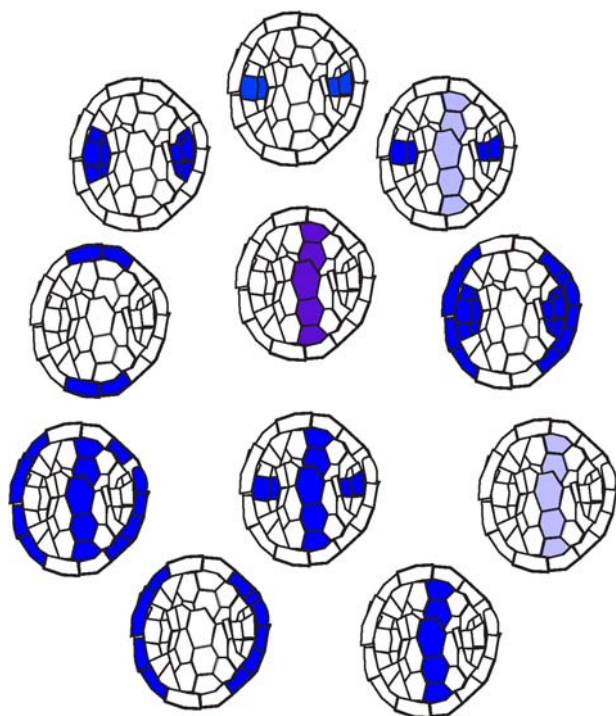
Introduction

Arabidopsis root development provides a remarkably tractable system to delineate tissue-specific, developmental gene regulatory networks (GRNs) and study their functionality in a complex multicellular model system. The simplifying aspects of root development include a small number of cell types arranged primarily in concentric cylinders and a set of stem cells at the root tip that give rise to all the other cells in the root. Previously, cell type-specific gene expression profiling revealed a large number of cell-type specific transcription factors (TFs), as well as TFs that are broadly expressed across cell types within a tissue and across tissues (Brady *et al.*, 2007) (Box 1). Spatially restricted 'master' transcriptional regulators

have been identified that specify root cell type identity and patterning (Levesque *et al.*, 2006). The expression of microRNAs (miRNAs) is also tightly spatiotemporally controlled throughout plant development in a similar manner to protein-coding genes (Válóczi *et al.*, 2006). This regulation contributes to tissue-specific activity of miRNAs and can result in a reduction of activity of their target mRNAs in specific tissues. However, little is known about the upstream regulators of these TFs and miRNAs.

The objective of this study was to initiate the experimental mapping of an Arabidopsis GRN using TF and miRNA-encoding genes that act within the root stele. The stele forms the central part of the root and contains xylem, phloem, procambium and pericycle cell types. We focused on TFs that

Box 1 Transcriptional patterns observed in the stele



Within stele tissue, 11 distinct dominant transcriptional patterns exist (Brady *et al*, 2007). These patterns occur in individual cell types of the tissue, and across multiple cell types of the tissue. Blue indicates relative enrichment in a cell type as marked by a GFP reporter line. Light blue indicates the meristematic stage of development, dark blue indicates the maturation stage of development, and purple indicates expression throughout all stages of development.

are primarily enriched within one cell type or multiple cell types of stele tissue, as well as miRNAs that target these TFs (Box 1, Supplementary Table S1). TFs and their targets have been shown to display similar tissue-specific expression enrichment (Deplancke *et al*, 2006a; Vermeirssen *et al*, 2007a). Therefore, we reasoned that this would provide an excellent starting point as it could increase our ability to identify spatially relevant regulatory interactions. In order to map the first Arabidopsis tissue-specific GRN, we employed gene-centered yeast one-hybrid (Y1H) assays to identify TFs that can bind TF and miRNA gene promoters (Deplancke *et al*, 2006a), as well as yeast two-hybrid (Y2H) assays to identify protein-protein interactions (PPIs) between the selected TFs. We then used the GRN to determine whether TFs that demonstrate spatially restricted expression patterns are regulated by TFs that are coexpressed. Finally, our initial GRN allowed us to identify emerging design principles of Arabidopsis gene regulation at a systems level.

Results and discussion

We selected a set of 167 primarily stele-enriched (in individual cell types or multiple cell types) TFs as protein prey for use in Y1H matrix assays (Box 1) (Vermeirssen *et al*, 2007b). These

TFs were selected based on their enriched expression in at least one cell type of the stele relative to all other tissues (> 1.2 fold, $q < 0.0001$) (Brady *et al*, 2007) (Supplementary Table S1). We next selected and cloned a subset (60) of the promoters of the genes that encode these TFs for the first mapping of a stele GRN (Supplementary Table S2). The promoters of five additional TFs that were identified as prey interactors were also cloned and included as DNA baits in a second round of Y1H screening. In total, the promoters from 65 TFs were used (Supplementary Table S2).

We identified 46 protein-DNA interactions (PDIs) between 16 TF-encoding gene promoters and 21 TFs (Supplementary Figure S1A, Supplementary Table S3). At least one interacting TF was identified for $\sim 25\%$ of the promoters. The majority of these promoters bound more than one TF. The number of interactions is relatively low compared with other gene-centered GRNs, where interactors were identified for 65–97% of promoters tested (Deplancke *et al*, 2006a; Vermeirssen *et al*, 2007a; Martinez *et al*, 2008a; Arda *et al*, 2010). This is likely due to TF coverage, as we only screened against 24% of all TFs expressed in Arabidopsis stele tissue, which represents 11% of all TFs encoded by the genome, while in the *Caenorhabditis elegans* studies, promoters were screened against cDNA and TF libraries that together cover more than 70% of all TFs (Deplancke *et al*, 2006a; Vermeirssen *et al*, 2007a; Martinez *et al*, 2008a; Arda *et al*, 2010).

Many experimentally validated miRNA targets exhibit highly localized expression within the stele (Lee *et al*, 2006). We selected the promoters of 28 miRNAs with representation from 16 families (Supplementary Table S3) giving priority to those directly targeting TFs with expression in the stele, and used these promoters in Y1H assays versus the same set of TFs. In all, 20 PDIs were found between eight miRNA promoters and 15 TFs (Supplementary Figure S1B, Supplementary Table S3). Six of these TFs interact with multiple miRNA promoters. Promoters of miRNAs within the same family (i.e., miR166) were bound by different TFs, suggesting that they are differentially regulated. This is in agreement with our previous observations in a *C. elegans* miRNA GRN (Martinez *et al*, 2008a).

To identify PPIs between the 167 stele-enriched TFs, we used Y2H matrix assays (Walhout and Vidal, 2001). In all, 25 interactions were identified between 26 TFs (Supplementary Figure S1C, Supplementary Table S3). Five of these involve homodimers, eight are heterodimers between TFs of the same family, and 12 are interactions between TFs of different families (Supplementary Table S4). Overall, 4 out of 6 previously reported interactions were identified, indicating that our false negative rate is relatively low (Supplementary Table S5) (Ehlert *et al*, 2006; Tan and Irish, 2006). When we visualized all PPIs in a network diagram, six discrete interaction groups could be identified (Supplementary Figure S1C). Novel interactions between different TF families were identified in most groups. These groups increase the spectrum of putative gene regulatory expression interactions requiring PPIs.

We integrated the PDI and PPI data into a GRN graph (Figure 1). We also incorporated putative stele-enriched downstream TF targets of miRNA genes (miRNA-miRNA interactions). TFs that bound TF promoters also bound miRNA

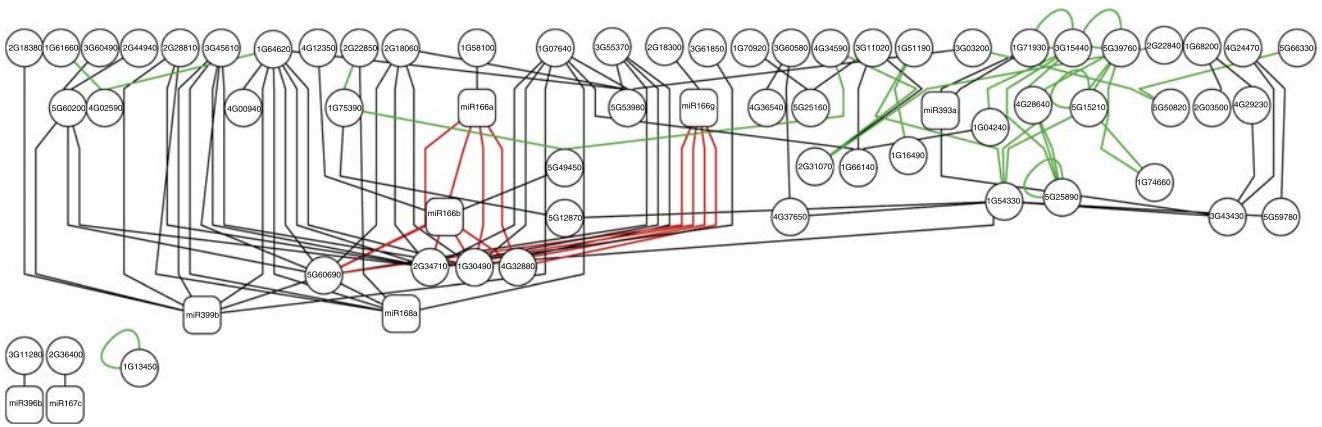


Figure 1 A stele-enriched Arabidopsis GRN. Black edge, PDI; red edge, miRNA–mRNA interaction; green edge, PPI; circle, TF; square, miRNA. Interactions are represented from the top down, with targets located below their interacting TF.

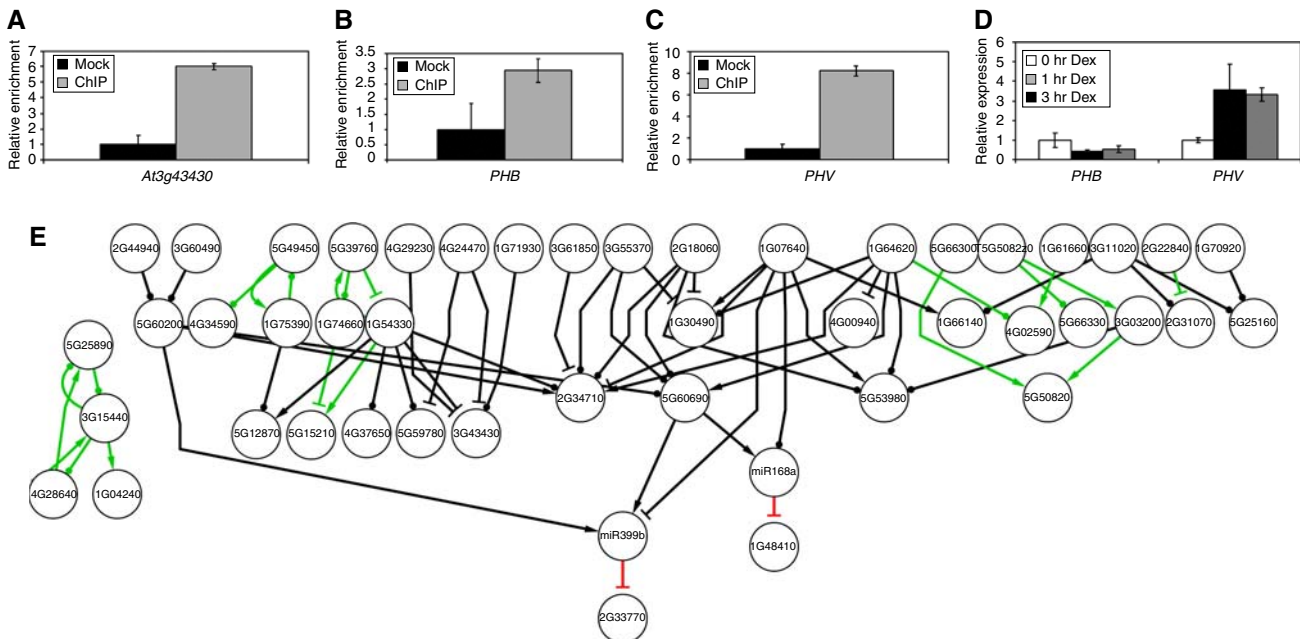


Figure 2 *In planta* validation of protein–DNA interactions (PDIs) using chromatin immunoprecipitation (ChIP)-quantitative PCR (qPCR) (A–C), conditional induction (D), or genetic analyses (E). Results are representative of three (A–C) two (D, E (protein-coding genes)), and four (E, miRNAs) biological replicates. (A–C) Black bar, mock precipitation; gray bar, anti-GFP immunoprecipitated fraction. (A) The –1952 to –1851 base pair region of *At3g43430* containing a NAC domain consensus sequence is enriched in the *VND7pro::VND7::GFP* immunoprecipitated fraction. (B) The –422 to –253 base pair region of *PHB* is enriched in the *OBP2pro::OBP2::GFP* immunoprecipitated fraction. (C) The –2437 to –2246 base pair region of *PHV* is enriched in the *OBP2pro::OBP2::GFP* anti-GFP immunoprecipitated fraction. (D) Transcriptional activity of the 35S:*OBP2::glucocorticoid receptor* line was induced with 10 μ M dexamethasone (Dex) and gene expression monitored at 1 and 3 h after Dex exposure. At 1 h after Dex exposure, *PHB* expression was downregulated, while *PHV* expression was upregulated. (E) Interactions that are repressive, activating, or exerting no effect (circle) were determined using qPCR of the TF and its target in a mutant TF background. Edges are as described in Figure 1.

promoters, indicating a lack of bias toward one or the other type of gene. The GRN is highly connected with only a few edges not linked to the main network component. Only 8 out of the 26 TFs that engage in a PPI with another TF also interact with a promoter, suggesting that dimerization may be a prerequisite for the other 18 to bind to DNA.

To determine whether the PDIs that we identified occur *in planta*, we first used chromatin immunoprecipitation (ChIP) coupled with quantitative PCR (qPCR) to examine interactions of two TFs with three promoters. We confirmed

binding of VND7 to the putative NAC domain-binding site in the *AT3G43430* promoter (Figure 2A). We also confirmed *in planta* binding of OBP2 to the proximal and distal Dof-consensus regions of the *PHB* and *PHV* promoters, respectively (Figure 2B and C). Bound regions showed little correlation with the abundance of Dof-consensus sequences (Supplementary Figure S2). This observation is in agreement with those in animal systems where the abundance of consensus TF-binding sites is a poor predictor of TF binding (Farnham, 2009).

As OBP2 bound both the *PHB* and *PHV* promoters in ChIP assays, we explored the ability of OBP2 to regulate *PHB* and *PHV* gene expression *in vivo* using a conditional OBP2/glucocorticoid receptor overexpression line (Skirycz *et al*, 2006). OBP2 activity was induced using dexamethasone (Dex) and gene expression monitored at 1 and 3 h after Dex exposure (Figure 2D). At both time points, OBP2 repressed *PHB* expression and activated *PHV* expression. Together, these data suggest that OBP2 directly interacts with the *PHB* and *PHV* promoters, and is able to act as both a repressor and activator. In summary, Y1H assays are able to identify physical PDIs that likely function to regulate gene expression *in planta*.

To further assess the *in planta* relevance of the interactions retrieved by Y1H and Y2H assays, we used qPCR to analyze the effects of mutations in TFs on the expression of their putative target genes. This allowed us to infer the consequence of the regulatory relationship (e.g., activator or repressor), as well as to quantify the degree of activation or repression. We also examined the effect of mutating TFs that physically interact with each other as some have been shown to regulate the expression of their partner, either indirectly or directly (Cui *et al*, 2007). We examined 54 putative regulatory interactions in 27 TF hypomorphic or hypermorphic allelic backgrounds using qPCR, or TaqMan miRNA expression assays within the whole root. For PPIs, if homozygous alleles were available for both interacting TFs, then expression of the interacting TF was measured in each mutant background. In all, 8 out of the 11 physically interacting TF pairs demonstrated transcriptional regulatory relationships in at least one direction (Figure 2E).

As was observed for OBP2, some TFs demonstrate both activating and repressing activity (AT1G54330, AT1G64620). Including interactions validated by ChIP-qPCR, 32 (59%) of interactions tested were validated *in vivo*, which is a much higher verification rate than those that have been reported from networks elucidated using ChIP methods in plants and animals (Li *et al*, 2008; Farnham, 2009; Moreno-Risueno *et al*, 2010). Of the TF genes mutated and analyzed for expression phenotypes, 16% had a root length or vascular mutant phenotype, several of which have been published (Supplementary Table S5). Moreover, 65% demonstrated a reproducible expression phenotype (Figure 2E). This demonstrates the striking developmental robustness of the stele GRN to tolerate and perhaps canalize expression changes without affecting overall morphology (Waddington, 1942).

We next employed a modeling approach to predict the regulatory potential of each of the relationships. We first plotted qPCR expression values of the TF and its target TF or miRNA in a wild-type and TF mutant background. We then used weighted least squares regression (WLSR) to fit a line that incorporates variance measurements across technical replicates for each biological replicate (Figure 3, Supplementary Figures S3 and S4) where the slope represents the degree of activation or repression and the *P*-value represents the confidence for the slope. As shown in Figure 3, the TF AT5G60200 is predicted to repress the expression of *mir399b* much more strongly than it activates the expression of *AT5G60690*. This predictive framework enables the identification of the most influential upstream activators or repressors to be manipulated in order to regulate expression of a target.

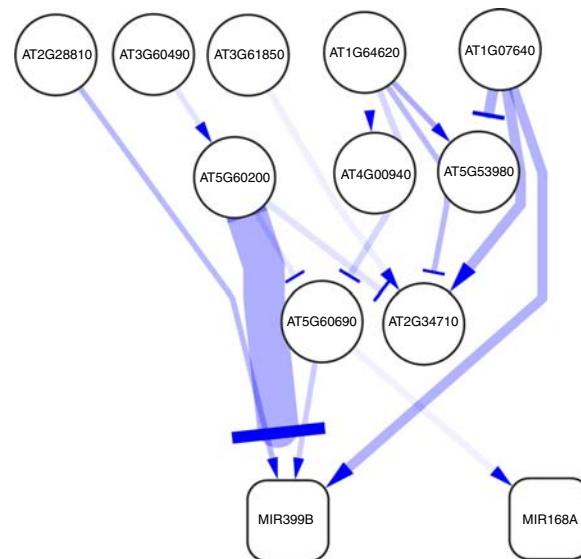


Figure 3 A predictive framework identifying regulatory relationships likely to have important roles in regulating expression of *At5g60690* (*REV*), *At2g34710* (*PHB*), *mir399b* and *mir168a*. Regulatory interaction strength is represented by edge and arrow width with thicker lines or arrowheads representing steeper slopes. *P*-value strength is represented by edge opacity with darker edges representing more significant interactions. Slopes were determined by plotting quantitative PCR (qPCR) expression values of the transcription factor (TF) and its target and fitting a line using weighted least squares regression (WLSR).

To determine whether the relative connectivity of TFs correlates with their function in development, we searched the literature for characterized gene function and mutant phenotypes (Supplementary Table S6). Genes that are bound by relatively large number of TFs are significantly positively correlated with conferring a morphological (root) phenotype ($P=0.05$; Supplementary Tables S7 and S8, Supplementary Figure S5). Conversely, no significant correlation was observed for TFs that bind a relatively large number of promoters ($P=0.19$). Although our sample size is relatively small, it is interesting to note that the relative morphological phenotypic importance of genes whose promoters are bound by a large number of TFs has also been observed in yeast pseudohyphal growth and in a *C. elegans* neuronal GRN (Borneman *et al*, 2006; Vermeirssen *et al*, 2007a). This may indicate a general property of GRNs in both unicellular and multicellular organisms in different kingdoms. The increased number of interactions with these promoters may reflect the need to tightly control the expression of the target and thus ensure robust expression. Furthermore, as suggested in yeast, these mapping methods can be used as a method to identify previously uncharacterized key developmental regulators (Borneman *et al*, 2006).

We next asked whether TFs and their targets are co-expressed within the same cell types of the stele. First, for each stele cell type, we calculated the expression of the TF and the target relative to their mean across the stele, and determined whether expression was enriched in one or more cell types. Surprisingly, very little overlap in expression enrichment between TFs and their targets was observed (Supplementary Figure S6). This suggests that future GRN mapping efforts with larger sets of TFs will likely uncover many more TF-target

interactions. We next examined the level of expression overlap between a TF and its target promoter by determining whether either the TF or the target is present (as opposed to enriched) within a cell type, and whether these expression domains overlap using the tissue overlap coefficient (TsOC) metric (the number of cell types shared between the TF and its target, divided by the smallest of the total number of tissues where either TF or target is expressed) (Martinez *et al*, 2008b). Overall, 83% of interacting nodes demonstrate overlap of 1.0 (Supplementary Figure S7a) (P -value=0.0157, Materials and methods). In 59% of these cases the upstream TF is more broadly expressed than its target, while in 23% of cases the target is more broadly expressed than its interacting TF, and in 18% of cases the TF and the target display identical expression domains (Supplementary Figure S7b) (P -value=0.0043, Materials and methods). Therefore, although TFs and targets show expression domain overlap, little coexpression was observed with respect to relative expression levels. These results should be taken into account when computationally reconstructing GRNs given gene expression data and expression similarity to define regulatory modules. As only a small proportion of TFs and targets are coexpressed, such computationally reconstructed GRNs will likely be incomplete.

The stele GRN that we delineated integrates 103 protein–DNA, protein–protein, and miRNA–mRNA interactions between 64 TFs and 8 miRNAs. Overall, 59% of the interactions tested were validated, confirming that our experimental approach is well suited to delineate Arabidopsis GRNs. Although our validation rate is much higher than those reported in plants and animals using ChIP-based systems, some interactions identified in yeast may not be relevant *in vivo*. However, and importantly, lack of validation by a biochemical or genetic assay does not necessarily mean that an interaction does not occur *in planta*. For instance, ChIP-chip and ChIP-seq experiments have revealed that many TFs bind to thousands of regulatory regions without observable regulation of gene expression (Li *et al*, 2008; Farnham, 2009; Moreno-Risueno *et al*, 2010). Furthermore, changes in gene expression may not be detectable by profiling expression in the whole root (Brady *et al*, 2007) or these may represent cases of genetic redundancy.

Genetic redundancy of TFs belonging to large families in Arabidopsis has been suggested because single loss-of-function mutants often fail to result in a morphological phenotype (Nawy *et al*, 2005). In this study, we found that morphological phenotypes were associated with only 16% of the TFs tested. However, molecular or expression phenotypes were identified for 65% of TFs. This suggests that the GRN is highly robust and points to canalization or buffering (Waddington, 1942). Functional redundancy between TFs, however, is still a possibility, for interactors without a gene regulatory or morphological phenotype. Our genetic validation approach, in addition to defining activators and repressors, will help refine our understanding of redundancy within GRNs, as well as provide a framework to unravel the combinatorial control of gene expression. For instance, nine TFs bind to the *PHB* promoter, seven of which belong to the Dof family, and two of which belong to the NAC family. These seven Dof TFs show expression domain overlap and sequence similarity, suggesting that they could act redundantly. Our genetic validation

approach can help refine this hypothesis. In all, four out of the seven Dof TFs were identified as regulators of *PHB* expression—two as activators and two as repressors. Using this information, one can now determine whether the two activators and repressors, respectively, act synergistically or additively to control expression. If more than two activators or repressors were identified, our modeling approach would allow us to predict the most likely mutant combination to maximally alter expression of the target. The remaining three Dof TFs that bind to the *PHB* promoter, but did not show an expression or morphological phenotype, can be tested for redundancy by making double and higher order mutants and observing both molecular and morphological phenotypes. An analogous situation involving genetic redundancy was observed in *C. elegans* where a TF–miRNA target interaction identified in a Y1H assay was not observed to have an effect on gene regulation using a genetic assay (Ow *et al*, 2008). This TF was then shown to redundantly act with two other TFs from the same family that also bound the same target in Y1H assays.

Expansion of the Arabidopsis root stele GRN to include modules that underlie differentiation of distinct cell types within this tissue as well as their temporal dynamics, in a similar manner to the sea urchin endomesodermal GRN will be a natural extension of this work (Smith and Davidson, 2008). Our gene-centered approach that uses genetic validation and modeling, coupled with high spatiotemporal resolution gene expression data (Brady *et al*, 2007), and increased coverage of all TFs within this tissue will allow us to comprehensively and precisely determine these spatiotemporal modules. Further data to be integrated in the GRN include how these gene regulatory interactions influence differential abundance of target expression in distinct cell types, tissues and time points. Similar approaches to those employed in sea urchin, where the effects of TF knockdowns are determined by visualizing target gene expression using reporter genes or *in situ* hybridization, will provide this added dimension to understanding the spatial and temporal dynamics of the stele GRN (Smith and Davidson, 2008).

Materials and methods

Plant growth

Arabidopsis thaliana lines in the Columbia (Col-0) ecotypes were used for all analyses. All plants were grown vertically on 1 × Murashige and Skoog salt mixture, 1% sucrose, and 2.3 mM 2-(*N*-morpholino)ethanesulfonic acid (pH 5.8) in 1% agar. All tissues sampled were 5 days old. Plants to be used for RNA extraction were plated on nylon mesh as in (Brady *et al*, 2007).

TF open reading frame cloning

Root RNA was isolated using the Qiagen RNeasy kit, and cDNA synthesized using the Superscript III First Strand Synthesis kit. Primers with GATEWAY-compatible attB1 and attB2 tails were designed to amplify the full-length coding sequence. PCR products were recombined with the pDONR221 vector using BP Clonase II and verified by sequencing. A number of TF clones were obtained from (Lee *et al*, 2006) and from the Arabidopsis Biological Resource Center (Supplementary Table 2). The resulting pENTR clones were recombined with pDEST-AD or pDB-DEST with LR Clonase II and verified by sequencing.

TF promoter cloning

Genomic DNA was extracted from seedlings using the Qiagen DNeasy Plant DNA Extraction kit. Primers were designed according to (Lee *et al*, 2006) and nested PCR performed to amplify up to 3000 bp of sequence upstream from the translational start according to (Lee *et al*, 2006), except for AT3G45610 which contains 2000 bp of upstream sequence. Amplicons were recombined with the pDONRP4P1-R vector using BP Clonase II and sequence verified. Resulting pENTR clones were recombined with pMW2 (Y1H *HIS3* reporter vector) and pMW3 (Y1H *LacZ* reporter vector) (Deplancke *et al*, 2006a) using LR clonase and sequence verified. Promoter sequences or sources can be found in Supplementary Table S3.

miRNA promoter cloning

Genomic DNA was extracted from seedlings. Promoter-specific primers with recombination sites were designed to amplify up to 2 kb region of upstream of the transcription start site of pre-miRNA (Xie *et al*, 2005). The purified PCR segments were cloned into pDONRp4-P1R using BP clonase and verified by sequencing. The correct clones were then recombined with pMW2 and pMW3 as described above. Promoter sequences can be found in Supplementary Table S3.

Y1H assays

Promoter::*HIS3* and promoter::*LacZ* reporter constructs were integrated in the yeast genome, selected on media lacking histidine and uracil and tested for auto-activation according to (Deplancke *et al*, 2006b). Genomic DNA was isolated from strains with minimal auto-activation, and promoter::reporter integration was verified by PCR and sequencing. AD-TF-encoding plasmids were directly transformed into bait yeast strains, and transformants were selected on media lacking histidine, uracil and tryptophan. Positive interactions were identified by activation of both the *HIS3* (by including 3-aminotriazole in the media) and the *LacZ* reporters (for up to 24 h after addition of Xgal) according to (Deplancke *et al*, 2006b), and compared with a pDEST-AD control and published Y1H controls (Deplancke *et al*, 2004). To minimize the inclusion of false positives, only double positive interactions in which both reporters were activated were considered. All interactions were validated by retesting using the same procedure.

Y2H assays

DB-TF and AD-TF constructs were transformed into MAV103 and MAV203 yeast strains and selected on media lacking leucine and tryptophan, respectively (Walhout and Vidal, 2001). DB-TF-containing yeast strains were selected for minimal auto-activation (<100 mM 3-AT with the majority showing activation between 20 and 80 mM 3-AT, <30 min exposure to X-gal; Supplementary Table 8). Mating was performed according to (Walhout and Vidal, 2001), and positive interactors were selected by activation of both the *HIS3* (on a competitive 3-AT series for up to 10 days after transformation) and *LacZ* reporter (for up to 24 h after addition of X-gal) relative to the minimal auto-activation of the DB-TF strains and published Y2H controls (Walhout and Vidal, 2001). DB-TF strains demonstrating auto-activation are found in Supplementary Table S2. All interactions were validated by retesting using the same procedure.

Mutant analysis: genotyping

In the stele GRN, 44 TFs had at least one interactor (PPI or PDI), other than itself. We obtained homozygous alleles of 31 TF-encoding genes of which 27 displayed altered expression of the gene disrupted as determined by qPCR. Up to 3 FLAG, SALK or SAIL insertion lines were identified per TF using the SIGNAL T-DNA Express Gene Mapping Tool (<http://signal.salk.edu/cgi-bin/tdnaexpress>) and obtained from the Arabidopsis Biological Resource Center. Primers were designed to amplify the insertion using the iSect Primer Design tool (<http://signal.salk.edu/tdnaprimers.2.html>). Genomic DNA was isolated

using the Sigma REExtract-N-Amp Plant PCR Kit (XNAP). Segregating insertion lines were tested for a gene-specific amplicon or a T-DNA-gene amplicon, with Columbia wild type as a control. Homozygous individuals contained only a T-DNA-gene amplicon, and no gene-specific amplicon. Genotyping primers are described in Supplementary Table S9.

qPCR—protein-coding genes

Primers were designed to amplify a 100 bp region containing 3'UTR sequence, if available. RNA was isolated from 5-day-old wild-type and homozygous mutant roots (30–50 each) respectively, grown on the same plate using the Qiagen RNeasy kit, and cDNA synthesized using the Superscript III First Strand Synthesis kit. Each plate is considered a biological replicate. cDNA was diluted to 500 ng/μl. Primers were tested on wild-type DNA by qPCR using the Applied Biosystems StepOne PCR system (Applied Biosystems) for artifacts as determined by the melting curve and amplification efficiency. Gene expression was measured for two biological replicates across technical replicates each of a Columbia wild-type and mutant pair using the $\Delta\Delta C^T$ method. If expression was reproducibly different across both biological replicates for the mutant line, then gene expression was measured for the target of the TF whose expression was measured, in both wild-type and mutant as described. For measurement of *PHABULOSA*, *PHAVOLUTA* and *REVOLUTA* expression, due to extensive sequence similarity, primers amplifying 400 bp of their cDNA sequence were designed. A primer set amplifying 400 bp of β -tubulin was used as a control in these cases. Primer sequences can be found in Supplementary Table S10. Expression was expressed relative to that of wild type.

qPCR-miRNAs

MiRNA was isolated from 5-day-old wild-type or homozygous roots grown on the same plate using the Ambion miRvana kit. Applied Biosystems TaqMan MicroRNA (Applied Biosystems) assays selected were miR168a (AB 4373403), miR399b (AB 4373425) and miR393a (AB 4373414). Two control small RNA primers for Arabidopsis (snoR85 and snoR41; AB 4395197 and 4395772) were tested for their expression variation in wild-type and three mutant root samples (At4g24470/SALK_008200, At1g64620/SALK_130584, and At1g75390/SALK_084241). SnoR85 showed higher expression in all samples with less variance than SnoR41 (SnoR85 C^T mean=22.96, variance=0.3 [2 biological replicates of wild-type, 2 biological replicates of At4g24470/SALK_008200 \times 4 technical replicates each], SnoR85 C^T mean=22.57, variance=0.13 [2 biological replicates of wild-type, 2 biological replicates of At1g75390/SALK_084241 \times 4 technical replicates each], SnoR41 C^T mean=25.99, variance=0.54 [2 biological replicates of wild-type, 2 biological replicates of At4g24470/SALK_008200 \times 4 technical replicates each]) and was selected for further use (Supplementary Figure S3). Gene expression for each wild-type and mutant pair was measured as described above for protein coding genes.

Chromatin immunoprecipitation—QPCR

Translational GFP fusions with high GFP fluorescence for At1g71930 and At1g07640 (Lee *et al*, 2006) were selected for immunoprecipitation. Root tissue of 5-day-old plants was fixed in 0.3% formaldehyde in MC buffer (10 mM Na₂PO₄ (pH7), 50 mM NaCl and 0.1 M sucrose) and washed three times. Tissue was prepared and immunoprecipitation was performed according to (Busch *et al*, 2010). Immunoprecipitation was performed with anti-GFP antibody (Abcam ab290). Primers were designed to amplify regions containing Dof-binding sites (TTTC/AAAG) or NAC-binding sites CATGTG/GTACTC (Supplementary Table S11). One region amplified did not contain a Dof-binding site for At1g07640 as an additional negative control. Primers were tested as described above for the qPCR experiments with Col-0 genomic DNA. Immunoprecipitation was performed for three biological replicates with two technical replicates each for each primer set per translational GFP line, and with Col-0 plants as a mock control with the same number of biological and technical replicates.

Root-length assay

A minimum of three Columbia and three homozygous mutant roots that displayed reproducible changes in gene expression were tested for changes in root length. If a significant change in root length was observed (unpaired Student's *t*-test, $P < 0.001$), root length was measured in one to two subsequent biological replicates for consistent statistical significant deviations from wild-type. Root length was determined using ImageJ at 7 days after germination. Root-length graphs of all mutants are displayed in Supplementary Table 8. An * indicates significance at $P < 0.001$, ** for $P < 0.0001$.

Expression overlap: TsOC

For all TFs and targets, their expression enrichment in the stele was determined by identifying expression in the JO121, J2661, S17, S18, S4, S32, SUC2, and WOL marker lines as determined by (Brady *et al*, 2007). If expression was > 1.0 , expression was listed in that cell type as follows: JO121=xylem pole pericycle (XPP), S17=phloem pole pericycle (PPP), S18=maturing xylem (MAX), S4=meristematic xylem (MEX), S32=phloem, SUC2=phloem companion cells. If expression was not identified in one of those marker lines, but was expressed in the overlapping WOL line, then expression enrichment in the procambium was determined using (Cartwright *et al*, 2009). The TsOC was measured according to (Martinez *et al*, 2008b). The data are summarized in Supplementary Table S13.

Receiver operating characteristic (ROC) curve

We used the ROC to measure the correlation between the presence of a phenotype and the number of targets to which a TF binds. We also used the ROC to measure the correlation between the presence of a phenotype and the number of TFs that interact with a single promoter. Nodes were ranked according to their in- or out-degree: nodes that display a phenotype were recorded as true positives and those that do not display a phenotype, or for which a phenotype has not been reported, as false positives. As we moved down the ranking, we recorded the relative proportions of true and false positives and plotted into an ROC curve. The deviation of this curve from the diagonal (that indicates a random distribution of phenotypes) is measured by considering the area between the curve and the diagonal. The statistical significance of this deviation was assessed by randomizing the phenotypes by 10^5 times and observing the distribution of areas. The area value of 0.1944 obtained for the data lies in the tail of this distribution, corresponding to a P -value of 0.051.

P-value calculation for TsOC

A P -value of 0.0157 for the overlap percentage of the TF and target expression domain (83%) was determined analytically and confirmed computationally. The analytical expression for the P -value is given by the probability of finding 38 or more TsOC=1 interactions among 46 randomly selected interactions from a total of $T=167^2=27\,889$ possible directed interactions (including self-interactions) between the 167 distinct TFs:

$$p = \sum_{r=38}^{46} \binom{N}{r} \binom{T-N}{46-r} / \binom{T}{46}$$

where $N=18\,743$ is the total number of TsOC=1 cases among all possible $T=27\,889$ interactions. The above expression yields a P -value of 0.0157, which we confirmed (to within ± 0.0005) by running a computational test, employing 10^6 sets of 46 randomly drawn interactions.

For the breakdown of the TsOC=1 cases into the three categories in which: (a) the TF is more broadly expressed than its target (59%), (b) less broadly expressed (23%), and (c) identically expressed (18%), we established a P -value of 0.0043 analytically by considering, for each possible number of TsOC=1 cases, all possible breakdowns into the three categories that represented a more extreme distribution, i.e., a proportion in category A of 59% or greater, in category B of 23% or less, and in category C of 18% or less. To improve accuracy, we used

the fraction of interaction counts, such as 22/38 instead of the rounded figures such as 59%. The P -value is, therefore, given by:

$$p = \sum_{r=0}^{46} \binom{T-N}{46-r} / \binom{T}{46} \sum_{k=\text{int}(f_A r)}^r \binom{N_A}{k} \sum_{m=F(r,k)}^{G(r,k)} \binom{N_B}{m} \binom{N_C}{r-k-m}$$

where $F(r,k)=\max[0, \text{int}(f_B r)] - k + \text{int}(f_A r)$, $G(r,k)=\min[\text{int}(f_B r), r-k]$, $f_A=22/38$, and $f_B=9/38$, and where $N_A=N_B=8041$, and $N_C=2661$ are the total numbers of TsOC=1 interactions in categories A, B, and C. Similar to the above P -value, this P -value of 0.0043 was confirmed by a similar computational test to be accurate within ± 0.0001 .

Modeling

For each putative interaction between any pair of TFs of interest, noted here as TF1 and TF2, we examined that any trend observed in our data is consistent with TF1 as an activator or repressor of TF2 under the conditions of the experiment. In the case that the data may suggest an activating or repressive effect, we estimated the degree of the effect relative to other TF pairs under examination. For each TF pair, the observed data consist of normalized qPCR measurement values computed as $\text{EXP}=2^{-(\text{Ct}_{\text{NOR}})}$. As described above, measurements were performed on each of two to three different wild-type and mutant (TF1 TDNA insertion) plants, and for each plant three technical replicates are performed to measure TF1 and TF2 gene expression. A biological replicate consisted of one wild-type and one mutant plant on a single plate. Thus, each experiment had two to three biological replicates per plant type, each with three technical replicates, for a total of 12–18 measurements of both TFs.

We visualized the relationship between TF1 and TF2 measurements in wild-type and mutant plants by determining the best-fit WLSR line between all TF1–TF2 biological replicates. In this procedure, the TF1 coordinate of each biological replicate is computed by taking the average of the TF1 technical replicate measurements, and likewise for the TF2 coordinate. Thus, each biological replicate was represented by a (TF1, TF2) coordinate pair. As the coordinate values are computed as averages, we weighted biological replicates that result from tightly clustered technical replicate measurements more heavily than those resulting from widely varying measurements. That is, we intuitively place more trust in biological replicates resulting from low-variance technical replicate measurements. We incorporate this judgment into the regression procedure by weighting each biological replicate point as $1/[\text{VAR}(\text{TF1 measurements}) + \text{VAR}(\text{TF2 measurements})]$. Resulting plots are displayed in Supplementary Figure S3.

Next, we computed the slope and P -value for the WLSR line. The slope of the regression line can help to suggest whether TF1 is an activator, repressor or has no effect on TF2. The steepness of the slope can also provide a rough estimate of the severity of the repressive or enhancing effect. A P -value is calculated using the *t*-test and the null hypothesis that the slope is equal to zero. Conceptually, the P -value answers the question 'If there were no linear relationship between TF1 and TF2 overall, what is the probability that randomly selected points would result in a regression line as far from horizontal (or further) as the observed line?'

As resources limit the number of biological replicates in the study to four to six replicates per pair of TFs, it is impractical to verify that all least squares regression assumptions are met (small number of values in residual plots). Nonetheless, we find the WLSR line to provide a useful visual suggestion about the data trend, and that the P -values provide a ranking among the plots that agrees well with what a human analyst would conclude about the data. That is, that TF1–TF2 relationships with steep slopes as well as replicates that fall very near to the best-fit line are the most convincing of a trend, and other relationships may still be suggestive but become less convincing with increasingly shallow slope and/or replicates scattered further from the best-fit line.

Overall, we observed 30 TF1–TF2 plots that suggested a repressive or activating trend. Among these TF pairs, 15 have a strongly convincing slope associated with a low P -value, and 6 have a moderately convincing slope associated with P -values in the range 0.14–0.2. In general, we used the P -values to rank the plots from most

to least convincing trends, and displayed this data in Supplementary Figure S3 and summarized in Supplementary Table S12.

Data access

Protein–DNA and protein–protein interaction data can be visualized in the VirtualPlant multinet network (<http://www.virtualplant.org>) and downloaded in .sif format at (http://virtualplant.bio.nyu.edu/public/download/multinet/Brady_data.sif), AGRIS (<http://arabidopsis.med.ohio-state.edu/submissions.html?id=agris10315>) (Palaniswamy *et al*, 2006), and AREX (http://www.arexdb.org/download_archive.jsp). The protein interactions from this publication have been submitted to the IMEx (<http://www.imexconsortium.org>) consortium through IntAct (Aranda *et al*, 2009) and assigned the identifier IM-15165.

Supplementary information

Supplementary information is available at the *Molecular Systems Biology* website (www.nature.com/msb).

Acknowledgements

This work was funded in part by an NSF grant to PNB and UO (Arabidopsis 2010: 0618304) and an NIH P50 grant for the Duke Center for Systems Biology, and UC Davis startup grants (SMB). AJMW is supported by NIH grants DK068429 and GM082971. LZ and DW are supported by USDA ARS 1907-21000-030-00D. SMB was supported by an NSERC PDF. MM is supported by an NSF PDF in Biological Informatics (0805648). SEA was supported by The Leverhulme Trust and The Royal Society (UK). We would like to thank Wolfgang Busch, Denis Dupuy, John Harada, Terri Long, Rossangela Sozzani, and Jaimie VanNorman for critical review of the manuscript, and Alexandra Skiryecz for the gift of the OBP2 mutant and induction lines.

Author contributions: SMB, LZ, DW, AJMW and PNB designed the research and SMB performed data analysis. SMB, LZ, SA, DW, AJMW and PNB wrote the manuscript. SMB, EJ and NJM cloned TFs, and SMB, LZ and EJ cloned promoters. SMB, LZ and NJM performed Y1H experiments. SMB and EJ performed Y2H experiments. SMB, CSY, WL and MTT genotyped mutants and performed qPCR and TaqMan miRNA assay experiments. SMB, AZ, MTT and DK performed mutant root phenotyping experiments. MM and UO designed the WLSR analysis and MM performed the WLSR analysis and contributed text. SA designed and implemented the statistical analysis of TF connectivity with function and expression domain overlap enrichment and contributed text.

Conflict of interest

The authors declare that they have no conflict of interest.

References

Aranda B, Achuthan P, Alam-Faruque Y, Armean I, Bridge A, Derow C, Feuermann M, Ghanbarian AT, Kerrien S, Khadake J, Kerssemakers J, Leroy C, Menden M, Michaut M, Montecchi-Palazzi L, Neuhauser SN, Orchard S, Perreau V, Roehert B, van Eijk K *et al* (2009) The IntAct molecular interaction database in 2010. *Nucleic Acids Res* **38** (Database issue): D525–D531

Arda HE, Taubert S, MacNeil LT, Conine C, Tsuda B, van Gilst M, Sequerra R, Doucette-Stamm L, Yamamoto KR, Walhout AJM (2010) Functional modularity of nuclear hormone receptors in a *C. elegans* metabolic gene regulatory network. *Mol Syst Biol* **6**: 367

Borneman AR, Leigh-Bell JA, Yu H, Bertone P, Gerstein M, Snyder M (2006) Target hub proteins serve as master regulators of development in yeast. *Genes Dev* **20**: 435–448

Brady SM, Orlando DA, Lee J-Y, Wang JY, Koch J, Dinneny JR, Mace D, Ohler U, Benfey PN (2007) A high-resolution root spatiotemporal map reveals dominant expression patterns. *Science* **318**: 801–806

Busch W, Miotk A, Ariel FD, Zhao Z, Forner J, Daum G, Suzaki T, Schuster C, Schultheiss SJ, Leibfried A, Haubeiff S, Ha N, Chan RL, Lohmann JU (2010) Transcriptional control of a plant stem cell niche. *Dev Cell* **18**: 849–861

Cartwright DA, Brady SM, Orlando DA, Sturmfels B, Benfey PN (2009) Reconstructing spatiotemporal gene expression data from partial observations. *Bioinformatics* **25**: 2581–2587

Cui H, Levesque MP, Vernoux T, Jung JW, Paquette AJ, Gallagher KL, Wang JY, Bllou I, Scheres B, Benfey PN (2007) An evolutionarily conserved mechanism delimiting SHR movement defines a single layer of endodermis in plants. *Science* **316**: 421–425

Deplancke B, Dupuy D, Vidal M, Walhout AJM (2004) A gateway-compatible yeast one-hybrid system. *Genome Res* **14**: 2093–2101

Deplancke B, Mukhopadhyay A, Ao W, Elewa AM, Grove CA, Martinez NJ, Sequerra R, Doucette-Stamm L, Reece-Hoyes JS, Hope IA, Tissenbaum HA, Mango SE, Walhout AJM (2006a) A gene-centered *C. elegans* protein–DNA interaction network. *Cell* **125**: 1193–1205

Deplancke B, Vermeirssen V, Arda HE, Martinez NJ, Walhout AJM (2006b) Gateway-compatible yeast one-hybrid screens. *Cold Spring Harb Protoc* doi:10.1101/pdb.prot4590

Ehler A, Weltmeier F, Wang X, Mayer CS, Smeekens S, Vicente-Carbajosa J, Dröge-Laser W (2006) Two-hybrid protein–protein interaction analysis in Arabidopsis protoplasts: establishment of a heterodimerization map of group C and group S bZIP transcription factors. *Plant J* **46**: 890–900

Farnham PJ (2009) Insights from genomic profiling of transcription factors. *Nat Rev Genet* **10**: 605–616

Lee J-Y, Colinas J, Wang JY, Mace D, Ohler U, Benfey PN (2006) Transcriptional and posttranscriptional regulation of transcription factor expression in Arabidopsis roots. *PNAS* **103**: 6055–6060

Levesque MP, Vernoux T, Busch W, Cui H, Wang JY, Bllou I, Hassan H, Nakajima K, Matsumoto N, Lohmann JU, Scheres B, Benfey PN (2006) Whole-genome analysis of the SHORT-ROOT developmental pathway in Arabidopsis. *PLoS Biology* **4**: e143

Li XY, MacArthur S, Bourgon R, Nix D, Pollard DA, Iyer VN, Hechmer A, Simirenko L, Stapleton M, Luengo Hendriks CL, Chu HC, Ogawa N, Inwood W, Sementchenko V, Beaton A, Weiszmann R, Celniker SE, Knowles DW, Gingeras T, Speed TP *et al* (2008) Transcription factors bind thousands of active and inactive regions in the *Drosophila* blastoderm. *PLoS Biol* **6**: e27

Martinez NJ, Ow MC, Barassa MI, Hammell M, Sequerra R, Doucette-Stamm L, Roth FP, Ambros VR, Walhout AJM (2008a) A *C. elegans* genome-scale microRNA network contains composite feedback motifs with high flux capacity. *Genes Dev* **22**: 2535–2549

Martinez NJ, Ow MC, Reece-Hoyes JS, Barrasa MI, Ambros VR, Walhout AJM (2008b) Genome-scale spatiotemporal analysis of *Caenorhabditis elegans* microRNA promoter activity. *Genome Res* **18**: 2005–2015

Moreno-Risueno MA, Busch W, Benfey PN (2010) Omics meet networks—using systems approaches to infer regulatory networks in plants. *Curr Opin Plant Biol* **13**: 126–131

Nawy T, Lee J-Y, Colinas J, Wang JY, Thongrod SC, Malamy JE, Birnbaum K, Benfey PN (2005) Transcriptional profile of the Arabidopsis root quiescent center. *Plant Cell* **17**: 1908–1925

Ow MC, Martinez NJ, Olsen PH, Silverman HS, Barrasa MI, Conradt B, Walhout AJ, Ambros V (2008) The FLYWCH transcription factors FLH-1, FLH-2, and FLH-3 repress embryonic expression of microRNA genes in *C. elegans*. *Genes Dev* **22**: 2520–2534

Palaniswamy SK, James S, Sun H, Lamb RS, Davuluri RV, Grotewold E (2006) AGRIS and AtRegNet: A platform to link cis-regulatory elements and transcription factors into regulatory networks. *Plant Physiol* **140**: 818–829

Skiryecz A, Reichelt M, Burow M, Birkemeyer C, Rolcik J, Kopka J, Zanol MI, Gershenzon J, Strnad M, Szopa J, Mueller-Roeber B, Witt I (2006) DOF transcription factor AtDof1.1 (OBP2) is part of a

- regulatory network controlling glucosinolate biosynthesis in Arabidopsis. *Plant J* **47**: 10–24
- Smith J, Davidson EH (2008) Gene regulatory network subcircuit controlling a dynamic spatial pattern of signaling in the sea urchin embryo. *PNAS* **105**: 20090–20094
- Tan QK-G, Irish VF (2006) The Arabidopsis zinc finger-homeodomain genes encode proteins with unique biochemical properties that are coordinately expressed during floral development. *Plant Physiol* **140**: 1095–1108
- Válóczi A, Várallyay E, Kauppinen S, Burgyán J, Havelda Z (2006) Spatio-temporal accumulation of microRNAs is highly coordinated in developing plant tissues. *Plant J* **47**: 140–151
- Vermeirssen V, Barrasa MI, Hidalgo CA, Babon JAB, Sequerra R, Doucette-Stamm L, Barabasi A-L, Walhout AJM (2007a) Transcription factor modularity in a gene-centered *C. elegans* core neuronal protein-DNA interaction network. *Genome Res* **17**: 1061–1071
- Vermeirssen V, Deplancke B, Barrasa MI, Reece-Hoyes JS, Arda HE, Grove CA, Martinez NJ, Sequerra R, Doucette-Stamm L, Brent MR, Walhout AJ (2007b) Matrix and Steiner-triple-system smart pooling assays for high-performance transcription regulatory network mapping. *Nat Methods* **4**: 659–664
- Waddington CH (1942) canalization of Development and the Inheritance of Acquired Characters. *Nature* **3811**: 563–565
- Walhout AJM, Vidal M (2001) High-throughput yeast two-hybrid assays for large-scale protein interaction mapping. *Methods* **24**: 297–306
- Xie Z, Allen E, Fahlgren N, Calamar A, Givan SA, Carrington JC (2005) Expression of Arabidopsis MIRNA genes. *Plant Physiol* **138**: 2145–2154



Molecular Systems Biology is an open-access journal published by *European Molecular Biology Organization* and *Nature Publishing Group*. This work is licensed under a Creative Commons Attribution-Noncommercial-Share Alike 3.0 Unported License.