­­Network Walking: Genome-wide validation of a TF-target network that mediates early-to-late Nitrogen responses

**Abstract**

An accounting of transcription factor (TF)-target interactions is necessary to understand how transcriptional networks control the response to environmental and developmental cues. Current approaches to identify these interactions rely on *in vitro* binding assays which lack information on the effect of binding. We have used a cell-based assay to identify direct regulated targets of 20 early nitrogen responsive TFs from Arabidopsis. Targets of each TF are enriched in nitrogen responsive genes and indicate that binding motifs are often associated with a particular direction of regulation. Using a Network Walking approach, we integrate our results with *in planta* perturbation of TGA1, a known regulator of nitrogen, to chart a path from direct targets in cells to indirect targets responding in whole roots. Our approach enables the identification of TFs that can be manipulated individually and in combination to alter nitrogen use efficiency and can be adapted to any transcriptional network in eukaryotic systems.

**Introduction**

Precise control of transcriptional networks enables cells to adapt to a changing environment and carry out complex developmental programs. One of the primary goals of systems biology is to reconstruct the set of transcription factor (TF)-target interactions for the underlying gene regulatory networks (GRNs). Studies have demonstrated that *de novo* network inference is a valuable approach to building GRNs in many organisms, from microbes to plants [1-4]. However, a major challenge especially in higher eukaryotes, is genome-wide validation of the accuracy and predictive power of the resulting GRNs. This is largely due to the lack of methods for high-throughput validation of the inferred TF-target interactions [1, 5].

Several techniques have been developed to provide genome-wide experimental evidence to train and test network inference algorithms. The most commonly used method for identifying direct TF target interactions is chromatin immunoprecipitation (ChIP). However, there are several limitations to this approach. First, TF-target binding performed *in vitro* is a poor predictor of gene regulation [6-9]. Second, at best, ChIP studies provide only a snapshot of the stable TF-binding events under the conditions and at the time-point assayed [10-14]. Other high-throughput approaches, such as Protein Binding Microarrays (PBM) [15] and DNA Affinity Purification Sequencing (DAP-Seq) [16], have been used to identify the genome-wide binding sites and cis-elements for hundreds of TFs. However, these *in vitro* approaches lack cellular context.

Validating TF-target predictions based on a functional assay, e.g. TF-mediated changes in gene expression, takes into account the *in vivo* context. Perturbation of TFs using knockout or overexpressing transgenic lines to identify regulated targets is standard across microbes and higher eukaryotes [17-20]. However, this approach is time-consuming and impossible to scale for most eukaryotes. Additionally, it is not possible to distinguish direct TF effects from indirect targets without additional TF-target binding information, such as ChIP.

Here, we have exploited a cell-based TF perturbation system called *TARGET* [21] to functionally validate predicted TF-target interactions in GRNs. This system overcomes many of the limitations described above, as experiments can be performed and analyzed within weeks, direct TF targets can be identified by TF-mediated gene regulation, and the assay is done on cells from the tissue of interest. The *TARGET* assay identifies TF targets though controlled nuclear localization of a transiently expressed TF [21]. As transient expression systems have been developed for many multicellular organisms as a quicker alternative to the generation of stable transgenics and mutants [22-24], the *TARGET* TF-perturbation approach is broadly applicable.

In this study, we have scaled-up the throughput of the cell-based *TARGET* system for TF-perturbation. The modifications we employ have allowed us to identify the direct genome-wide targets of 20 TFs involved in early nitrogen (N) signaling. Our aim is to understand how this nutrient signal, which is first perceived in root cells [25], triggers longer-term growth responses (e.g. lateral root nutrient foraging) [26]. At present, the connection between the early and late effects of N-signaling is not well understood. In our proof-of-principle “Network Walking” example, we were able to determine the network path for a well-known TF (TGA1) in the N-response [27]. Using this approach, we connect 72% of the indirect targets which respond to TGA1 only *in planta,* back to TGA1 through intermediate TF2s. Elucidating the hierarchy of TF-target interactions underlying the N-signaling response in plants has implications for plant N-use efficiency, a process that impacts agriculture, the environment and human health.

**Results**

**Rapid genome-wide validation of direct regulated targets of 20 “N-early response” TFs.**

We sought to identify which of the ~2,000 Arabidopsis transcription factors facilitate the rapid response to N-signaling in plants. To do this, we targeted TFs for functional testing that themselves have a rapid transcriptional response to N treatment in a fine-scale time course [25]. In that study, the identification of genes that respond to N by time (NxTime) was combined with a “*Just-in-Time*” [I would suggest “first response”; it’s more accurate] classification, in which genes were binned based on the *first* time-point there mRNA reached a fold-change of 1.5 in response to N treatment relative to the control [25]. These NxTime responsive genes identified include 145 TFs in roots and 162 in shoots, with an overlap of 49 TFs (Fig. 1a). We selected a set of 20 TFs that respond to NxTime in both shoots and roots to validate their role in the N-early response network. These 20 N-early response TFs include 11 “first responders” (5-10 min), 6 “early responders” (15-20 min), and 3 “later responders” (45-90 min) (Fig. 1a). The TFs include 6 that have previously been validated in the N-response (e.g. CRF4 & CDF1 [25], NAC4 [28], TGA1/4 [27], LBD38 [29], and an additional 14 TFs with an unknown role in N-signaling.

To determine the genome-wide targets regulated by these 20 “N-early response” TFs, we used the cell-based *TARGET* system for inducible TF perturbation [21]. This system has several advantages over *in planta* TF-target validation; i) Rapid: TF-targets can be identified in 1 week, ii) Direct targets: it can identify *direct* TF-targets based on TF-induced gene regulation [21], and iii) Throughput: with our modifications we can functionally test 20 TFs within 3 weeks. Briefly, the cell-based *TARGET* assay [21] identifies direct TF-targets based on TF-induced gene regulation. TF-nuclear entry is controlled using a subdomain of the glucocorticoid receptor (GR) [30] fused to the TF of interest. The GR-TF fusion protein is held in the cytoplasm by HSP90-GR binding [31]. Dexamethasone (DEX) treatment disrupts this binding, allowing temporal control over TF entry into the nucleus. Additionally, when transfected root cells are pre-treated with Cycloheximide (CHX) to block downstream regulation of secondary TF-targets, direct TF-targets can be identified as those that respond transcriptionally to DEX-induced TF nuclear import in the presence of CHX.

We made two important changes to the cell-based *TARGET* system to increase the throughput of TF perturbations (Fig. 1b). First, by introducing an Empty Vector (EV) control – a construct containing only the GR and no TF – we can identify differential expression resulting from DEX-induced nuclear translocation of all TFs tested on that same day. Second, when two populations of cells are transfected separately with two different vectors – TF1-RFP vector, and TF2-GFP vector – cells expressing the two flurophores can be pooled and efficiently separated by FACS, saving time in the cell-sort process.

As proof-of-principle, we used this higher throughput *TARGET* approach to identify direct targets of our set of 20 “N-early responsive” TFs. The transcriptome responses affected by each of these 20 TFs were examined at 3 hours after DEX-induced nuclear entry of the GR-TF fusion. To obtain a list of differentially expressed (DE) genes, each set of TF samples was compared to the EV using the analysis pipeline shown in Fig. S1. The TF-targets identified as DE between each of the 20 TFs and EV (FDR<0.025) are reported in Tables S1-20.

**The 20 “N-early responsive” TFs directly regulate genome-wide targets enriched in NxTime genes**

The number of direct genome-wide targets for each of the 20 “early N-response” TFs was between 596 (CRF4) and 5,799 (HSFB2A) (Fig. 1c). Each TF is capable of acting as either as an inducer or repressor of target genes (Tables S1-S20). As the 20 TFs were selected based on their N-response [25], we examined whether the targets of each TF overlapped with NxTime responsive genes in the shoot or root. To do so, we calculated the specificity of each TF to the NxTime gene sets (i.e. the percentage of direct TF targets that respond in NxTime) and the influence of each TF on the NxTime genes (i.e. the percent of the NxTime genes regulated by a TF) for both shoots and roots (Fig. 1c, see Methods). The order and shading of the TFs in Fig. 1c is based on the N-specificity Index (p-value), a measure of the significance of the influence a TF has on the NxTime responsive genes (see Methods). Overall, the targets of each of the 20 N-early response TFs significantly overlapped with the NxTime genes in shoots and/or roots, however the ordering of the TFs was organ-dependent. For instance, CRF4, a known TF in the N-response in shoots [25] (Fig. 1c, green arrows), and HYH (Fig. 1c, orange arrows), are examples of TFs whose targets show specificity for the shoot or root, respectively. By contrast, bZIP3 and C2H2, TFs that our study now implicate in the N response control genes that respond to N treatment in both shoots and roots (Fig. 1c black arrows),.

**TF-target regulation data can identify cis-motifs associated with gene repression or activation.**

We used our TF-target regulation data to interrogate TF-DNA binding data and classify cis-elements based on their association with the direction of gene regulation (e.g. induction or repression). Specifically, ten of the 20 “N-early response” TFs in our study also have cis-binding motif data from DAP-seq [16] (16 cis-motifs), Cis-BP [32] (8 cis-motifs) or protein binding microarray (PBM) [33] (1 cis-motif). We looked for enrichment of these cis-motifs in the 500bp promoter of the regulated targets. When all targets of each of the 10 TFs were analyzed, we found a significant enrichment (p-value < 0.05) of at least one cis-motif for 7/10 TFs in our study (15/25 unique cis-motifs, Table S21). We further separated the TF-regulated genes for each TF into induced and repressed targets, and repeated the analysis (Fig. 2). When separated by direction of regulation, an additional 5 motifs (19/25 motifs) and two additional TFs (9/10 TFs) showed enrichment of their motif in either the induced or repressed set of targets. The cis-binding motif was found to be enriched *only* in the promoter of *either* the induced targets or repressed targets for 8/10 TFs. The lone example of the cis-binding motif being enriched in *both* induced and repressed TF-regulated targets was for BEE2. One of the TFs (HHO3), showed no enrichment of the cis-binding motifs in any set of *in vivo* regulated targets. Similar results are observed when the overlap between the target genes identified by DAP-seq are compared to the direct and regulated targets identified by *TARGET* (Table S22).

**N-responsive TFs regulate a connected network of genes enriched for N-related processes**

The above results ranked the 20 N-early response TFs according to their *individual* roles in regulating genes in the N-signaling pathway. We also asked how this set of N-early response TFs *work together* in mediating the N-response. When analyzed collectively, the regulated targets of 20 N-early response TFs are significantly enriched in the root NxTime responsive genes (p-value 1.01E-96) (Table S23). Collectively these 20 TFs regulate 73% of the NxTime genes in roots (1215/1658 genes; p-value 3.08E-96) and 73% of the NxTime genes in shoots (1582/2173; p-value 2.37E-123).

To gain further insight into their collective behavior, we explored the network topology for the 20 N early response TFs and their targets. We compared the distribution of TF-target edges within the validated *TARGET* network (Fig. 3a, orange bars) to a randomized network that contains the same number of TFs, targets and edges (Fig. 3a, grey bars). The distribution of edges in the validated *TARGET* network (orange bars) differs significantly from a randomized network. Specifically, the validated network contains significantly (p-value <0.001) more targets that are “unique” (i.e. targeted by only one or two of the TFs) or shared (i.e. targets of 8 or more TFs), compared to the random network (Fig. 3a).

We next asked whether the shared targets of the 20 N-early response TFs are enriched for N-related processes. To test this, we calculated the enrichment for Gene Ontology (GO) terms, i.e. the frequency of a GO term in the targets of a TF, over the background frequency of the GO term (see Methods). This analysis uncovered enrichment in N-metabolism terms, root development and hormone signaling (Table S24). We found that the enrichment of these GO terms increases as the number of TFs with edges to a set of common targets increases (Fig. 3b). To test whether the enrichment of N-related GO processes in the shared TF targets occurred by chance, we devised a Figure of Merit, which we call Focus (see Methods). The Focus for a TF-target network is greater with respect to set of genes (e.g. GO term) when TFs have more edges to that set of genes. Our significance test determines the probability that the Focus calculated from the edges in the validated network is higher than we would expect to see by chance. For the network of all validated targets of the N-early response TFs, the Foucs for each of the GO terms was significantly greater than for the randomized networks (p-value <0.005) (Fig. 3b).

**The shared targets of the 20 N-early response TFs are enriched in cis-regulatory elements for putative TF partners.**

In order to identify TFs that work together in regulating genes in the dynamic N-response, we looked at the occurrence of cis-motifs in the regulated targets of the 20 early NxTime responsive TFs. Given the large number of TF-binding motifs from high-throughput methods such as DAP-seq [16], PBM [32, 33] and SELEX [34], and the fact that TFs from the same family often have very similar cis-motifs, searching for the >1,200 cis-motifs available is impractical, and the results are hard to interpret. Therefore, we used the RSAT matrix-clustering tool [35] on available Arabidopsis TF cis-motifs (1,282), which identified 80 cis-motifs groups based on hierarchical clustering (Fig. S2). Cis-motifs from TFs belonging to the same family generally fell into the same cluster, which reflects the previous analysis of the 529 cis-motifs identified by DAP-seq [16]. For each cis-element cluster, we obtained a consensus motif and corresponding position weight matrix (PWM) [35] (Tables S25 and S26). Using the PWM for each of the 80 motif clusters, we calculated enrichment of each in the promoters (500bp) of the regulated targets of the 20 N-early response TFs. This analysis shows that the induced vs. repressed targets of the 20 N-early response TFs form two clusters based on the cis-motifs shared in the promoters (Fig. 4). Furthermore, we are able to identify sets of cis-motifs that are primarily enriched in the induced (Fig. 4, green box) or repressed (Fig. 4, red box) targets. For example, cis-motifs for some TF families (e.g. cis-motif cluster 29 and 30 (MYB)) are enriched only in repressed TF-targets. By contrast, other cis-motifs (e.g. cis-motif cluster 8 and 23 AP2EREBP) are enriched only in the induced TF targets.

**Using time-series network inference to predict TF-target edges for all N-responsive TFs in roots.**

We next sought to expand our gene regulatory network of the N-response in roots beyond the initial 20 TFs for which we validated the genome-wide targets using the *TARGET* system (Fig. 1). To do this, we used fine scale time-series transcriptome response to N-treatment [25] in a network inference approach called Dynamic Factor Graphs (DFG) [36] to predict TF-target edges for all 145 TFs in the N-response. DFG is a time-based machine-learning method that uses the time series to estimate the quantitative influence of TFs at time *t* on target genes at time *t+1.* This approach has been used previously to build network models that can predict gene expression states at future time-points, even when few time points are tested [25, 37]*.* The resulting GRN provides an edge score, or measure of influence, for each TF on every target gene in the network (i.e. 243,726 TF-target edges).

To validate our GRN predictions, we used the validated TF-target edges for 18 of the N-early response TFs identified in *TARGET* (Fig. S3a) to calculate a “precision” threshold for TF-target predictions, and to “prune” the DFG network for high-confidence TF-target edges involved in the root N-response [1]. We note that two of the TFs – ZFP4 and HSFB2A – were excluded from this analysis, as they did not meet the stringent FDR threshold used for DFG. Area Under Precision Recall (AUPR) for the TF-target predictions in the DFG predicted GRN (0.209) was significantly greater than for 1000 random Precison-Recall (PR) curves (mean = 0.175) (Fig S3b). From the PR curve (Fig. S3c) generated using the validated edges of the 18 TFs, we chose a Precision threshold cut-off of 0.30. This point, at which the curve begins to flatten out, was used to filter our predicted network for high confidence TF-target edges. We note that our Precision cutoff of 0.30 is of comparable scale to the 0.50 Precision achieved using an ensemble approach of multiple network inference methods in simpler microbial systems [1]. At this cut-off, which corresponds to an edge threshold score of 1.0794, the resulting “pruned” GRN is comprised of 5,577 total edges between 143 TFs and 252 targets (Fig. S4, Tables S27 and S28).

To evaluate their individual contributions of the 18 TFs to the Precision and Recall, we calculated the Precision, Recall and the F-score (harmonic mean of Precision and Recall [38]) for each of the 18 TFs within the pruned DFG network (Table 1). While there is variation in all three metrics for each of the 18 TFs, the mean Precision, Recall, and F-score among all 18 TFs was 0.38, 0.17 and 0.18 respectively, and are close to weighted values for the Precision, Recall and F-score calculated from the *TARGET* edges for all 18 TFs, indicating that they are not biased towards a few TFs with many edges. Importantly, the 90% confidence intervals for Precision (0.31 - 0.46), Recall (0.10 – 0.25) and F-score (0.14 - 0.21), indicate that the TF-target edges predicted by DFG for the remaining 125 NxTime TFs in the network are likely to be reliable.

 **‘Network Walking’ – charting a TF network path from direct to indirect targets.**

We next sought to combine our network inference data (for 125 N-response TFs) with TF-target validation data (for 20 N-early response TFs) to chart a path for an N-early response TF from its direct targets (in root cells, where the N-signal is first detected) to its indirect targets (e.g. in whole roots). In an approach we call “Network Walking” (Fig. 5a), we connect an N-early response TF1 to its direct targets identified in root cells, which include TF2s. Next, we connect TF2s to indirect targets of TF1 (identified *in planta)* using edges for intermediate TF2s validated by the cell-based *TARGET* system (e.g. 20 TFs), and high confidence TF-target predictions for the remaining 125 TFs in the DFG network. As proof-of-principle, we demonstrate how the Network Walking approach can shed light on the mode-of-action and network path of TGA1, an important regulatory component of the N response [27] (Fig. 5b). The dataset includes genes that respond to TGA1 perturbation in cells (Table S20), or *in planta* (Table S29), as well as TF-target edges validated for 20 TFs (Tables S1-19) and for the 125 TF targets as predicted by DFG (Tables S27 and S28).

In Network Walking, ***the first step*** is to use the TF1 targets identified by the cell-based *TARGET* system to identify which DE genes from *in planta* perturbation are direct targets. In our example, *TARGET* assays shows that TGA1 directly regulates 574 root N-responsive genes (Fig. 5b, yellow box), which includes 125 genes that also respond to TGA1 *in planta.* ***The second step***, is to connect a path from TF1 – via a TF2 – to its indirect targets which only respond to *in planta* perturbation of TF1. To connect TGA1 to its indirect targets, we use TF-target validated direct edges for the other 19 TFs from *TARGET* assays (Fig. 1), or the predicted TF2-target edges from the pruned DFG network (Fig. S4). For the TGA1 example, 72% (75/104) of the indirect TGA1 targets identified *in planta* can now be linked back to TGA1 through 43/57 TF2s. The set of 57 TF2s validated to act downstream of TGA1, includes 5 TFs for which we have validated targets in cells using the *TARGET* assay (HHO3, BEE2, GATA17, TGA4, HHO2), and 52 TFs with DFG predicted edges to TGA1 indirect targets. To further determine which of these intermediate TF2s are most important in relaying the signal from the primary TF (TGA1), enrichment of the cis-motifs in the targets for each of the 80 cis-motif clusters (Fig. S2) can be used. To apply this approach to the TGA1 Network Walk, we looked for enrichment of cis-motifs in the promoters (500bp) of TGA1 indirect targets. The most enriched cluster in the TGA1 indirect targets was for the WRKY family (cluster 13) (FDR = 0.109). Interestingly, 6 WRKY TFs respond to TGA1 perturbation in cells or *in planta*, with WRKY54 being a direct TGA1 target and good candidate for future validation studies.

**Discussion**

The goal of system biology is to learn networks that can accurately predict TF-target edges in a Gene Regulatory Network (GRN). A key step in this is to provide experimental validation of predicted edges between TF regulators and their target genes in a GRN. However, ascertaining genome-wide functional targets of a TF in a high throughput manner is a challenge across biological systems. For example, even in the most studied single cell system – *E. coli K12* – less than half of the genes have a known TF regulator, one third of the predicted TFs lack experimental targets, and two thirds lack a high confidence binding site [39]. The sparsity of TF-target regulation data is a limitation to elucidating these networks, and is more striking in multicellular eukaryotes such as Arabidopsis, where there are approximately six times as many genes and TFs.

Despite significant advances in the identification of physical interactions of TF and targets, there is still relatively little known about which genes *are transcriptionally regulated* by a majority of the ~2,000 Arabidopsis TFs. A proliferation of TF-target binding data has recently emerged from high-throughput *in vitro* approaches such as DAP-Seq [16] and PBMs [33]. Collectively, these datasets have identified more than 1,200 cis-binding motifs for nearly 700 TFs (Fig. S2, Table S26). However, one consideration is that these *in vitro* identified motifs rarely account for features present *in vivo,* such as protein-protein interactions and chromatin structure. More importantly, TF-target binding data – obtained *in vitro* or *in vivo* – cannot inform about the effect of this interaction on gene expression, e.g. induction vs. repression.

It is thus necessary to determine how a TF affects gene regulation in a live plant cell. One of the primary limitations to identifying TF regulated targets in whole plants, and other multicellular eukaryotes, is the time it takes to generate mutants and transgenics. Cell-based transient assays facilitate the rapid identification of regulated TF targets based on TF-induced regulation [40-44]. This approach is potentially better than TF-DNA binding because it is a functional assay that identifies targets in plant cells. The direct TF-targets identified by *TARGET* have been demonstrated to have *in planta* relevance, as shown for ABI3 [21], bZIP1 [42] and HRS1 [43]. Importantly, *TARGET* can also capture rapid and transient TF target interactions in isolated cells that are missed in whole tissues *in planta* [42, 45]. However, until now the *TARGET* TF-perturbation system has been used only to identify targets of a single TF [21, 42-44] or few TFs [25]. In this study, we have increased the throughput of the *TARGET* system by making modifications to the protocol. This has enabled the identification of direct targets for 20 N-responsive TFs in less than a month (Fig. 1 and Tables S1-S20).

As proof-of-principle, we selected 20 “N-early responders” (5 min –2 hrs) TFs from a time-course study [25] for medium throughput identification of the TF-target edges in root cells using *TARGET*. As a positive control, we chose six TFs that have previously been validated in the N response *in planta* – TGA1/TGA4 [27], NAC4 [28], LBD38 [29] and CRF4/CDF1 [25]. We also selected 14 TFs for which we have now demonstrated a role in N-signaling (ASR3, BEE2, bZIP3, C2H2, COL5, ERF056/060, GATA17/17-L, HHO2/3, HSFB2A, HYH, and ZFP4). Our *TARGET* experiments revealed that the direct regulated targets of each of the 20 TFs overlap significantly with NxTime responsive genes in root and/or shoot (Fig. 1c). Impressively, these 20 N-early response TFs, when combined, regulate 73% of all NxTime responsive genes from both organs, supporting their collective role in mediating the dynamic N response in plants. The enrichment of N-related processes in the shared targets of these 20 TFs (Fig. 3b), indicates that the influence of these 20 TFs converge on the N response. We also observed that the distribution of edges in the validated TF-target network resembles a scale-free network [46] (Fig. 3a), a common feature of biological networks that provides robustness [47]. Signal integration, an emergent property of biological systems [48], could explain the large number (258) of N-responsive TFs in shoots and roots, roughly 10% of all the predicted Arabidopsis TFs [25]. This type of complex combinatorial interaction between TFs, which can fine-tune the response to signal inputs [49], has been observed previously in plants [50, 51] and other systems [52-54].

The *TARGET* system which reports information on the direct regulated TF targets, provides much needed *in vivo* context to TF-binding data. We found that the enrichment of the TF binding motif in the promoters of TF-regulated targets is specific to the direction of regulation (e.g. induction vs. repression) for a majority of the TFs in our study (Fig. 2). This result suggests that direct TF-binding in the promoter of a target is typically associated with one regulatory effect (induction vs. repression). We also observed that *in vitro* TF-target binding is a poor indicator of TF-regulation, e.g. the vast majority of TF-bound genes are not regulated by the TF [Not so clear. Is this one TF? If so which one? Or should this be TFs?] in plant cells (Table S22). Because each TF both induces and represses direct targets (Tables S1-S20), regulation that cannot be explained by binding to the primary cis-element may involve weaker secondary cis-elements, binding elsewhere (e.g. gene body), or partner TFs which are not present in *in vitro* binding assays.

To facilitate the incorporation of the large amount of available TF binding data into our GRN, we compiled data on TF-target cis-motifs and used hierarchical clustering to compile 1282 cis-motifs into 80 representative cis-motif clusters (Fig. S2). The resulting cis-motif clusters separate into families, as seen for clustering of 529 DAP motifs into 57 clusters [16] and clustering of the 489 plant TF cis-motifs in the JASPAR database into 35 clusters [55]. We found that sets of genes that are directly induced or repressed by the 20 TFs tested in the *TARGET* system, share common cis-elements in their promoters (Fig. 4). This approach facilitates the identification of TFs that work together to regulate shared targets.

To integrate our results with other types of data, we derived an approach called Network Walking. The key feature of the Network Walking strategy is to connect the regulated direct TF1-edges in plant cells (identified using *TARGET*) to the indirect TF1 targets validated *in planta*. This is done using; i) experimental data on a TF1 for its direct regulated TF2 targets, ii) experimental data for the direct targets of TF2s, iii) predicted edges for TF2s to their targets from network inference, and iv) enrichment of the cis-motifs in the indirect targets of TF1. The resulting “Network Walk” which begins with a single TF1 is used to identify TF2s which mediate the signal between the initial TF1 and downstream *in planta* targets. The results can guide combinatorial experiments (e.g. TF stacking [should this be explained?]) and experiments on important TF2s in a systems biology cycle.

As proof-of-principle, we demonstrate the “Network Walking” approach (Fig. 5), starting with TGA1, a known regulator of the N-response in Arabidopsis [27]. TGA1 was previously identified as the highest ranked TF when all Arabidopsis TFs were sorted on several criteria, including the number of connections having a consistent response to nitrogen [27]. Due to the functional redundancy of TGA1 and TGA4, Alvarez et al. validated the role of TGA1/TGA4 in the N response using a double *tga1/tga4* mutant. In our study, we overcame the redundancy issue by identifying TGA1 targets based on overexpression in cells and *in planta* (Tables S20 and S29). Our *TARGET* results show that TGA1 is able to directly regulate 35% of all N-early responsive genes in root cells, and has the third highest N-specificity index in roots of the 20 N-early response TFs studied (Fig.1c), confirming the important role of TGA1 in N-signaling. Combining our *TARGET* and *in planta* data allowed us to identify the *in planta* targets that respond directly to TGA1 (Fig. 5b, yellow box) and separate them from indirect TGA1 targets (Fig. 5b, orange box). These results indicate that there are a large number of direct TGA1 targets detected in root cells that are missed by *in planta* perturbation assays. These may be transient TGA1 targets that can be detected only when the *TARGET* assay is performed in the presence of a protein synthesis inhibitor [21, 42, 43].

The large number of targets (574 genes) directly regulated by TGA1 includes 57 N-early response TF2s. By regulating roughly 40% of the N-responsive TF2s, the effect TGA1 has on the N response is amplified. In the “Network Walk” for TGA1, we were able to use the 57 downstream TF2s to connect TGA1 to a majority of its indirect targets identified *in planta*. These connections come from the validated direct targets of TF2s identified using *TARGET*, and high confidence TF2-target predictions from the pruned inferred GRN. To add to our predictions of which TFs are important in the GRN, we also used enrichment of cis-binding elements in the indirect targets. For example, enrichment of cis-elements points to a member of the WRKY family as being important in propagating the TGA1 signal. This implicates three WRKYs, which are validated direct targets of TGA1for future studies.

The development of tools for high throughput validation of the edges within inferred networks is crucial to obtaining high quality predictive GRNs. We demonstrate that the *TARGET* system for cell-based TF perturbation complements existing approaches by providing *direct* TF-target edges based on TF-mediated gene *regulation* in cells isolated from the tissue of interest. Importantly this system does not require the creation of stable transgenics and scales easily. We have also introduced an approach, Network Walking, which connects the early and direct TF targets in cells to downstream responses observed *in planta*. While our proof-of-principle studies focused on rapid N signaling in Arabidopsis, both of these approaches are generally applicable to study of GRNs involved in transducing signals in any eukaryotic system.

**Methods**

*Genome-wide targets of 20 TFs identified with the modified plant cell-based TARGET system*

The 20 TFs were TOPO cloned into pENTR (Invitrogen) from cDNA or isolated from the Arabidopsis TF collection [56]. TFs were then transferred to the pBeaconRFP\_GR plasmid [21] or a GFP version of the same plasmid (pBeaconGFP\_GR) by Gateway (Invitrogen) cloning. Arabidopsis Col-0 plants were grown in 1% w/v sucrose, 0.5 g/L MES, 1X MS basal salts (-CN), 1 mM KNO3, 2% agar, pH 5.7 for 10 days prior to the *TARGET* experiment. Light conditions were 120 μmol m-2s-1 at constant temperature at 22⁰C, 16 h light, 8 h dark (long day). The preparation of root protoplasts and conditions for the transient transformation were as described previously [21, 42]. For each TF and the EV construct, 4-6 million cells were transformed separately and after washing, a single TF in the RFP vector and a single TF in a GFP vector were combined in 3 replicate wells of a 24 well plate. After overnight incubation, each pool of transformed root protoplasts were treated sequentially with N (20 mM KNO3 + 20 mM NH4 NO3 ) for 2 hours and 35 µM CHX for 20 min before a 10 µM DEX treatment to induce TF nuclear entry. Transformed cells were sorted by FACS into GFP and RFP expressing populations 3 hours after DEX-induced TF nuclear import. Cells over-expressing the candidate TF or empty vector were collected in triplicate and RNA-Seq libraries were prepared from their mRNA using the NEBNext® Ultra™ RNA Library Prep Kit for Illumina®. The libraries were pooled (24 libraires per run) and sequenced on the Illumina NextSeq 500 platform for 75 cycles. The RNA-Seq reads were aligned to the TAIR10 genome assembly using TopHat [57] and gene expression estimated the GenomicFeatures/GenomicAlignments packages [58]. The gene counts for every sample were combined and differential expression between the TF overexpression libraries and the EV libraries were identified by using DESeq2 [59] package with a TF+Batch model at a significance level of FDR < 0.025.

*Nitrogen Specificity, Influence and N-Specificity Index*

Specificity of each TF to the NxTime responsive data set of Varala et al. [25] was calculated by dividing the NxTime responsive targets of a TF by the total number of targets regulated by that TF. The percent of NxTime genes regulated by a TF was calculated as the number of NxTime responsive targets of that TF divided by the total number of NxTime responsive targets. The N-specificity index was calculated as in Varala et al. [25]. [Perhaps repeat that calculation]

*GO Enrichment*

The web application agriGO v2.0 [60] was used to identify GO terms enriched cumulative direct regulated targets of the 20 TFs, and for the intersection of these direct targets with the root NxTime genes.

For Fig. 3b, Let *G* be a list of genes associated with a GO-term and *Tk* be a list of Targets that are targeted by at least *k* TFs. We can calculate the enrichment of the *G* in *Tk* by simply comparing the frequency of *G* in *Tk* (i.e. the fraction of genes in *Tk* that intersect *G*), represented as *Fk*, to the frequency of G in the background (all 27,416 annotated Arabidopsis genes), represented as *Fb*.
$$Eg=\frac{Fk}{Fb}$$

For each gene list (*G*) we also calculated a figure of merit we call Focus (*Fg*) by simply adding the *Eg* for all *k* up to *n* TFs:

$$Fg=\sum\_{k=1}^{n}\frac{Fk}{Fb}$$

To test if the Focus of the validated network is significant for a given GO term, a Monte Carlo test was used to determine an empirical *P* value by comparing the Focus of the validated network to the Focus of 1000 iterations of a randomized network, generated by shuffling the edges within the experimentally validated TF-target network.

*Cis-binding site enrichment and clustering*

Enrichment of the cis-binding for motifs was calculated using the AME tool within the MEME package [61], setting the background to the 500bp promoter regions (upstream of the transcription start site) from all 27,416 TAIR10 annotated genes, and the frequency of bases to the occurrence within that set.

For motif clustering, cis-binding motifs for Arabidopsis transcription factors were collected from DAP-seq [16], Cis-BP [32] and the PBM of Zorilla et al. [33]. Position weight matrices (PWMs) were converted to the MEME motif format [62] and the RSAT matrix-clustering tool [35] was used with the following parameters: hclust\_method = average, calc = sum, metric\_build\_tree = Ncor, lth w 5 lth cor = 0.6, lth Ncor = 0.45, quick = true. To search for the enrichment of each motif in the TF targets, the consensus PWM for each of the 80 clusters was converted to the MEME format and the FIMO tool [63] within the MEME package was used to identify every occurrence of each of the 80 consensus cis-motifs in the 500bp promoters of all 27,416 Arabidopsis genes at a p-value of 0.0001. Overlapping motifs were removed, retaining only the motif with the lowest p-value. For each set of TF targets, enrichment of a motif in the target set relative to their occurrence in all annotated genes was calculated using a Fisher’s exact test. The resulting p-values were FDR corrected using the Benjamini-Hochberg procedure [64]. Heatmaps and hierarchical clustering were generated with euclidean distance and the “ward.D” agglomeration method using the gplots [65] heatmap.2 function in R.

*Identification of direct and indirect TGA1 in planta targets*

The TGA1 overexpression construct was made by Gibson assembly [66] with a three‐part construct. The CaMV‐35S promoter was fused to the TGA1 CDS using in the pGreen vector backbone. Primers used in the assembly are in Table S30.

Arabidopsis Col-0 plants with the 35S:TGA1 transgene were generated using Agrobacterium-mediated floral-dip method [67]. Approximately 100 seeds were sown in Phytatrays (Sigma-Aldrich) in liquid media that was identical to what was used in the *TARGET*: 1% w/v sucrose, 0.5 g/L MES, 1X MS basal salts (-CN), 1 mM KNO3, pH 5.7. Light conditions were 120 μmol m-2s-1 at constant temperature at 22⁰C, 16 h light, 8 h dark (long day).

When seedlings were 13 days old, they were transferred to N-starvation media (1% w/v sucrose, 0.5 g/L MES, 1X MS basal salts (-CN), pH 5.7). After 24 hours, 2 hours after subjective dawn, seedlings were transferred to Phytatrays containing identical media with the addition of 20mM KCl or 20mM KNO3 + 20mM NH4NO3. Plants were incubated within treatment media for 2 hours. After which, root tissue were immediately harvested and flash frozen in liquid nitrogen.

RNA was extracted from root tissue using the QIAGEN RNeasy kit (Qiagen). mRNA was purified with oligo-T beads (Invitrogen), and RNA-seq libraries made using the NEBNext Ultra Library Prep Kit (NEB). Libraries were sequenced the Illumina HiSeq 2500 v4 platform using 1x50 single end chemistry.

Resulting RNA-seq reads were aligned to the TAIR10 Arabidopsis genome using Tophat [57] and gene counts estimated using HTSeq [68]. Reads were then filtered and quantile normalized using the EDASeq [69] R package. Two-way ANOVA was used to identify genes whose expression was significantly different between the treatments.

*Network Inference and Pruning*

The DFG [36] predicted Gene Regulatory Network (GRN) was generated as described previously [25]. Briefly, DFG infers interactions between 145 TFs and 1658 genes that responded to N in the root time course [25]. The full inferred GRN, which contains an edge between every TF and every target gene (240,410 edges) was pruned using the validated TF-target edges from *TARGET.* The validated edges were used to perform an AUPR analysis and identify a precision threshold of 0.30 (Fig. S3b-c). This edge cut-off was chosen to minimize false positives, while recovering as many true positives as possible. The resulting pruned network was visualized (Fig. S4) using Cytoscape [70]. Precision (True Positives/ (True + False Positives)), Recall (True Positives/ (True Positives + False Negatives)) and F-score ((2 \* Precision \* Recall) / (Precision + Recall)) were calculated for the edges in the pruned network to generate Table 1.

**Figure Legends**

**Figure 1 –** **A modified *TARGET* system enables increased throughput for screening early N-response TFs.** A) TFs chosen for early nitrogen response to nitrogen in roots and shoots [25]. TFs in bold have been described in the nitrogen response. TFs with an asterisk do not meet the threshold for DFG.

B) Modifications enable screening of 8 TFs/day. GFP and RFP reporters allow transfection of protoplasts which are divided into replicates. Transfection of samples with an Empty Vector (EV) are compared to TFs to identify regulated genes. FACS - Fluorescent Activated Cell Sorting, N - nitrogen, CHX - cycloheximide, DEX - dexamethasone.

C) The intersection of TF targets with NxTime genes allows identification of new TFs important in both roots and shoots (bZIP3/C2H2, black arrows), tissue specific TFs (shoots CRF4, green arrows, roots HYH, orange arrow).

**Figure 2 – *in vivo* TF-Target regulation data helps interpret TF cis-binding motifs determined *in vitro.***

One or more cis-binding motifs are available for 10 of the 20 TFs from DAP-seq [16], Cis-BP [32] or the PBM of Franco-Zorrilla et al. [33]. Enrichment of these motifs in the 500bp promter of either the induced or repressed targets identified by the cell-based *TARGET* assay reveals that *in vitro* identified binding motifs often are related to only induction or repression of targets by the TF. Only BEE2 shows enrichment of the motif in both the induced and repressed targets.

**Figure 3 – Nitrogen-related processes are enriched in targets shared by multiple early N response TFs.**

A) TF-target edges in the validated network (orange bars) resembles a scale-free distribution, with significantly more unique targets (edges to 1 or 2 TFs) or shared targets (edges to 8 or more TFs), compared to a randomized network which contains the same number of TFs, targets and edges (grey bars). B) Enrichment of gene ontology terms for nitrogen related processes increases as the number of TFs regulating the set of target genes increases. Significance was tested by calculating a figure of merit called Focus (see methods), for the validated network and comparing it to the Focus value for 1000 randomized networks in which the edges within the validated network were shuffled. (\*\* *P* < 0.005, \*\*\* *P* < 0.001)

**Figure 4 – Cis-motifs representative of TF families are enriched in induced or repressed regulated TF targets**

Heatmap and clustering of enriched consensus cis-motifs for the 80 clusters (columns) in the 500bp promoters of the regulated targets of the 20 TFs identified using the *TARGET* system. TF targets were separated based on whether they were induced (green) or repressed (red). The motif logo, cluster number and family representation for the each consensus motif is shown above. The red box indicates the set of consensus motifs that are enriched primarily in the repressed targets and the green box those motifs that are primarily enriched in the induced targets.

**Figure 5 – Network Walking: Connecting early and late TF responses**

A) A schematic representation of the Network Walking approach used to connect early and direct TF targets identified in cells to the late and indirect targets identified only *in planta*. B) Example of a Network Walk from direct TGA1 targets identified in cells (yellow shaded region) to indirect TGA1 targets identified only by in planta TGA1 overexpression. Edges connecting the indirect TGA1 targets back to TGA1 come from validated *TARGET* edges for 5 TFs (HHO3, BEE2, TGA4, GATA17, HHO2) and DFG predicted edges for 52 additional TF2s. To further narrow candidate TF2s for further study, enrichment of cis-motifs for the 80 TF clusters (Fig. S2 and Tables S26 and S26) in the 500

bp promtoers of the indirect TGA1 targets was assessed. The consensus motif for cluster 13, which is predominantly WRKY family TFs is the most enriched. Indirect targets with a WRKY motif present are outlined in pink, and WRKY TFs are filled with pink. The network shown is limited to TFs and targets that respond to NxTime in Varala et al. [25]

**Table 1 – Precision and recall of the 18 TFs within the pruned DFG inferred network**

A comparison of TARGET validated versus DFG predicted edges in the pruned DFG network was used to calculate Precision (TP/(TP+FP)), Recall (TP/(TP+FN)) and F-score (harmonic mean of Precision and Recall). The mean and 90% confidence intervals of these values for the 18 TFs are similar to the values obtained with combining the edges for all 18 TFs. TP – True Positive, FP – False Positive, FN – False Negative.

**Figure S1 – Workflow for analyzing the *TARGET* RNA-seq results to identify differentially expressed genes**

General workflow for processing the RNA-seq results for the *TARGET* experiments with the input files at each step on top. After trimming adapter and low quality reads/bases, Tophat [57] was used to align to the Arabidopsis TAIR10 genome. Gene counts for the aligned reads estimated the GenomicFeatures/GenomicAlignments packages [58]. Expression counts from all 69 samples (20 TFs + 3 empty vectors times 3 replicates) were combined in a single matrix and analyzed using DESeq2 [59] with a TF + Batch model.

**Figure S2 – Heatmap of cis-binding motif clustering for 1282 Arabidopsis TF motifs into 80 groups**

Heatmap generated by the RSAT matrix-clustering tool [35] of the similarity (based on motif length normalized correlation (Ncor)) between 1282 position weight matrices of Arabidopsis TFs. Cis-binding motifs were collected from DAP-seq [16], Cis-BP [32] and the PBM of Zorilla et al. [33]. Description of each motif and TF membership can be found in Tables S25 and S26)

**Figure S3 – Evaluation of the inferred DFG network from roots with validated genome-wide TF target data.**

A) Genome-wide regulated targets of 18 TFs (Fig. 1c) were used to calculate the Precision and Recall of the DFG inferred GRN and “prune” the TF-target edges in the network to a precision threshold of 0.30.

B) Area Under Precision-Recall (AUPR) analysis demonstrates that the ranking of edges in the DFG inferred network is significantly better (*P* < 0.001) than randomizing the order of edge rankings (n=1000). The pruned network has a precision cut-off of 0.30 which corresponds to an Edge Score of 1.0794. The resulting pruned network has 5,577 edges between 143 TFs and 252 target genes (Tables S7 and S28).

C) Plot of the Precision-Recall curve. The 0.30 cut-off was chosen as the highest precision before the curve flattens. The pruned network contains 653 out of 29,844 predicted edges in the full DFG network for the 18 TFs in this study. TP – True Positive, FP – False Positive, FN – False Negative.

**Figure S4 – A pruned network inferred using dynamic factor graphs predicts targets for 143 NxTime TFs**

The root NxTime data from Varala et al [25] was used to infer TF-target influence with a time-based machine learning approach, DFG [36, 37]. Using genome-wide validated targets for 18 of the 20 TFs (Fig. 1) identified by the *TARGET* system, the inferred edges were pruned to a precision threshold of 0.3, chosen based on AUPR analysis (Fig S4). The resulting network is displayed in the context of the ‘Just-in-Time’ bins for each TF (left) and target gene (right) [25]. The size of the nodes is representative of the number of edges for each TF and target. The shading of the edges indicates the edge score. TFs highlighted in red have been previously shown to be involved in nitrogen signaling.

**References**

1. Marbach, D., J.C. Costello, R. Küffner, N.M. Vega, R.J. Prill, D.M. Camacho, K.R. Allison, A. Aderhold, R. Bonneau, and Y. Chen, *Wisdom of crowds for robust gene network inference.* Nature methods, 2012. **9**(8): p. 796.

2. Wilkins, O., C. Hafemeister, A. Plessis, M.M. Holloway-Phillips, G.M. Pham, A.B. Nicotra, G.B. Gregorio, S.V. Jagadish, E.M. Septiningsih, R. Bonneau, and M. Purugganan, *EGRINs (Environmental Gene Regulatory Influence Networks) in Rice That Function in the Response to Water Deficit, High Temperature, and Agricultural Environments.* Plant Cell, 2016. **28**(10): p. 2365-2384.

3. Li, Y., S.A. Pearl, and S.A. Jackson, *Gene Networks in Plant Biology: Approaches in Reconstruction and Analysis.* Trends Plant Sci, 2015. **20**(10): p. 664-675.

4. Gaudinier, A. and S.M. Brady, *Mapping Transcriptional Networks in Plants: Data-Driven Discovery of Novel Biological Mechanisms.* Annu Rev Plant Biol, 2016. **67**: p. 575-94.

5. Arrieta‐Ortiz, M.L., C. Hafemeister, A.R. Bate, T. Chu, A. Greenfield, B. Shuster, S.N. Barry, M. Gallitto, B. Liu, and T. Kacmarczyk, *An experimentally supported model of the Bacillus subtilis global transcriptional regulatory network.* Molecular systems biology, 2015. **11**(11): p. 839.

6. Lickwar, C.R., F. Mueller, S.E. Hanlon, J.G. McNally, and J.D. Lieb, *Genome-wide protein–DNA binding dynamics suggest a molecular clutch for transcription factor function.* Nature, 2012. **484**(7393): p. 251.

7. Marchive, C., F. Roudier, L. Castaings, V. Bréhaut, E. Blondet, V. Colot, C. Meyer, and A. Krapp, *Nuclear retention of the transcription factor NLP7 orchestrates the early response to nitrate in plants.* Nature Communications, 2013. **4**: p. 1713.

8. Swift, J. and G.M. Coruzzi, *A matter of time - How transient transcription factor interactions create dynamic gene regulatory networks.* Biochim Biophys Acta, 2017. **1860**(1): p. 75-83.

9. Gorski, J.J., K.I. Savage, J.M. Mulligan, S.S. McDade, J.K. Blayney, Z. Ge, and D.P. Harkin, *Profiling of the BRCA1 transcriptome through microarray and ChIP-chip analysis.* Nucleic acids research, 2011. **39**(22): p. 9536-9548.

10. Kuo, M.-H. and C.D. Allis, *In vivo cross-linking and immunoprecipitation for studying dynamic protein: DNA associations in a chromatin environment.* Methods, 1999. **19**(3): p. 425-433.

11. Cuvier, O. and B. Fierz, *Dynamic chromatin technologies: from individual molecules to epigenomic regulation in cells.* Nature Reviews Genetics, 2017. **18**(8): p. 457.

12. Todeschini, A.-L., A. Georges, and R.A. Veitia, *Transcription factors: specific DNA binding and specific gene regulation.* Trends in genetics, 2014. **30**(6): p. 211-219.

13. Li, Y., K. Varala, and G.M. Coruzzi, *From milliseconds to lifetimes: tracking the dynamic behavior of transcription factors in gene networks.* Trends Genet, 2015. **31**(9): p. 509-15.

14. Varala, K., Y. Li, A. Marshall-Colon, A. Para, and G.M. Coruzzi, *"Hit-and-Run" leaves its mark: catalyst transcription factors and chromatin modification.* Bioessays, 2015. **37**(8): p. 851-6.

15. Bulyk, M.L., *Protein binding microarrays for the characterization of DNA-protein interactions.* Adv Biochem Eng Biotechnol, 2007. **104**: p. 65-85.

16. O'Malley, R.C., S.C. Huang, L. Song, M.G. Lewsey, A. Bartlett, J.R. Nery, M. Galli, A. Gallavotti, and J.R. Ecker, *Cistrome and Epicistrome Features Shape the Regulatory DNA Landscape.* Cell, 2016. **166**(6): p. 1598.

17. Gardner, T.S., D. Di Bernardo, D. Lorenz, and J.J. Collins, *Inferring genetic networks and identifying compound mode of action via expression profiling.* Science, 2003. **301**(5629): p. 102-105.

18. Cantone, I., L. Marucci, F. Iorio, M.A. Ricci, V. Belcastro, M. Bansal, S. Santini, M. di Bernardo, D. di Bernardo, and M.P. Cosma, *A yeast synthetic network for in vivo assessment of reverse-engineering and modeling approaches.* Cell, 2009. **137**(1): p. 172-81.

19. Vermeirssen, V., I. De Clercq, T. Van Parys, F. Van Breusegem, and Y. Van de Peer, *Arabidopsis ensemble reverse-engineered gene regulatory network discloses interconnected transcription factors in oxidative stress.* The Plant Cell, 2014. **26**(12): p. 4656-4679.

20. Faith, J.J., B. Hayete, J.T. Thaden, I. Mogno, J. Wierzbowski, G. Cottarel, S. Kasif, J.J. Collins, and T.S. Gardner, *Large-scale mapping and validation of Escherichia coli transcriptional regulation from a compendium of expression profiles.* PLoS biology, 2007. **5**(1): p. e8.

21. Bargmann, B.O., A. Marshall-Colon, I. Efroni, S. Ruffel, K.D. Birnbaum, G.M. Coruzzi, and G. Krouk, *TARGET: a transient transformation system for genome-wide transcription factor target discovery.* Mol Plant, 2013. **6**(3): p. 978-80.

22. Codamo, J., T.P. Munro, B.S. Hughes, M. Song, and P.P. Gray, *Enhanced CHO cell-based transient gene expression with the epi-CHO expression system.* Mol Biotechnol, 2011. **48**(2): p. 109-15.

23. Huynh, C.Q. and H. Zieler, *Construction of modular and versatile plasmid vectors for the high-level expression of single or multiple genes in insects and insect cell lines.* J Mol Biol, 1999. **288**(1): p. 13-20.

24. Yin, J., G. Li, X. Ren, and G. Herrler, *Select what you need: a comparative evaluation of the advantages and limitations of frequently used expression systems for foreign genes.* J Biotechnol, 2007. **127**(3): p. 335-47.

25. Varala, K., A. Marshall-Colón, J. Cirrone, M.D. Brooks, A.V. Pasquino, S. Léran, S. Mittal, T.M. Rock, M.B. Edwards, G.J. Kim, S. Ruffel, W.R. McCombie, D. Shasha, and G.M. Coruzzi, *Temporal transcriptional logic of dynamic regulatory networks underlying nitrogen signaling and use in plants.* Proceedings of the National Academy of Sciences, 2018. **115**(25): p. 6494-6499.

26. Ruffel, S., G. Krouk, D. Ristova, D. Shasha, K.D. Birnbaum, and G.M. Coruzzi, *Nitrogen economics of root foraging: transitive closure of the nitrate-cytokinin relay and distinct systemic signaling for N supply vs. demand.* Proc Natl Acad Sci U S A, 2011. **108**(45): p. 18524-9.

27. Alvarez, J.M., E. Riveras, E.A. Vidal, D.E. Gras, O. Contreras-Lopez, K.P. Tamayo, F. Aceituno, I. Gomez, S. Ruffel, L. Lejay, X. Jordana, and R.A. Gutierrez, *Systems approach identifies TGA1 and TGA4 transcription factors as important regulatory components of the nitrate response of Arabidopsis thaliana roots.* Plant J, 2014. **80**(1): p. 1-13.

28. Vidal, E.A., J.M. Alvarez, and R.A. Gutierrez, *Nitrate regulation of AFB3 and NAC4 gene expression in Arabidopsis roots depends on NRT1.1 nitrate transport function.* Plant Signal Behav, 2014. **9**(3): p. e28501.

29. Rubin, G., T. Tohge, F. Matsuda, K. Saito, and W.R. Scheible, *Members of the LBD family of transcription factors repress anthocyanin synthesis and affect additional nitrogen responses in Arabidopsis.* Plant Cell, 2009. **21**(11): p. 3567-84.

30. Aoyama, T. and N.H. Chua, *A glucocorticoid-mediated transcriptional induction system in transgenic plants.* Plant J, 1997. **11**(3): p. 605-12.

31. Yamaguchi, N., C.M. Winter, F. Wellmer, and D. Wagner, *Identification of direct targets of plant transcription factors using the GR fusion technique.* Methods Mol Biol, 2015. **1284**: p. 123-38.

32. Weirauch, M.T., A. Yang, M. Albu, A.G. Cote, A. Montenegro-Montero, P. Drewe, H.S. Najafabadi, S.A. Lambert, I. Mann, K. Cook, H. Zheng, A. Goity, H. van Bakel, J.C. Lozano, M. Galli, M.G. Lewsey, E. Huang, T. Mukherjee, X. Chen, J.S. Reece-Hoyes, S. Govindarajan, G. Shaulsky, A.J.M. Walhout, F.Y. Bouget, G. Ratsch, L.F. Larrondo, J.R. Ecker, and T.R. Hughes, *Determination and inference of eukaryotic transcription factor sequence specificity.* Cell, 2014. **158**(6): p. 1431-1443.

33. Franco-Zorrilla, J.M., I. Lopez-Vidriero, J.L. Carrasco, M. Godoy, P. Vera, and R. Solano, *DNA-binding specificities of plant transcription factors and their potential to define target genes.* Proc Natl Acad Sci U S A, 2014. **111**(6): p. 2367-72.

34. Mathelier, A., X. Zhao, A.W. Zhang, F. Parcy, R. Worsley-Hunt, D.J. Arenillas, S. Buchman, C.Y. Chen, A. Chou, H. Ienasescu, J. Lim, C. Shyr, G. Tan, M. Zhou, B. Lenhard, A. Sandelin, and W.W. Wasserman, *JASPAR 2014: an extensively expanded and updated open-access database of transcription factor binding profiles.* Nucleic Acids Res, 2014. **42**(Database issue): p. D142-7.

35. Castro-Mondragon, J.A., S. Jaeger, D. Thieffry, M. Thomas-Chollier, and J. van Helden, *RSAT matrix-clustering: dynamic exploration and redundancy reduction of transcription factor binding motif collections.* Nucleic Acids Research, 2017. **45**(13).

36. Mirowski, P. and Y. LeCun, *Dynamic Factor Graphs for Time Series Modeling.* Machine Learning and Knowledge Discovery in Databases, Pt Ii, 2009. **5782**: p. 128-143.

37. Krouk, G., P. Mirowski, Y. LeCun, D.E. Shasha, and G.M. Coruzzi, *Predictive network modeling of the high-resolution dynamic plant transcriptome in response to nitrate.* Genome Biol, 2010. **11**(12): p. R123.

38. Hase, T., S. Ghosh, R. Yamanaka, and H. Kitano, *Harnessing diversity towards the reconstructing of large scale gene regulatory networks.* PLoS Comput Biol, 2013. **9**(11): p. e1003361.

39. Gama-Castro, S., H. Salgado, A. Santos-Zavaleta, D. Ledezma-Tejeida, L. Muniz-Rascado, J.S. Garcia-Sotelo, K. Alquicira-Hernandez, I. Martinez-Flores, L. Pannier, J.A. Castro-Mondragon, A. Medina-Rivera, H. Solano-Lira, C. Bonavides-Martinez, E. Perez-Rueda, S. Alquicira-Hernandez, L. Porron-Sotelo, A. Lopez-Fuentes, A. Hernandez-Koutoucheva, V. Del Moral-Chavez, F. Rinaldi, and J. Collado-Vides, *RegulonDB version 9.0: high-level integration of gene regulation, coexpression, motif clustering and beyond.* Nucleic Acids Res, 2016. **44**(D1): p. D133-43.

40. Yoo, S.-D., Y.-H. Cho, and J. Sheen, *Arabidopsis mesophyll protoplasts: a versatile cell system for transient gene expression analysis.* Nature protocols, 2007. **2**(7): p. 1565.

41. Wehner, N., L. Hartmann, A. Ehlert, S. Böttner, L. Oñate‐Sánchez, and W. Dröge‐Laser, *High‐throughput protoplast transactivation (PTA) system for the analysis of Arabidopsis transcription factor function.* The Plant Journal, 2011. **68**(3): p. 560-569.

42. Para, A., Y. Li, A. Marshall-Colon, K. Varala, N.J. Francoeur, T.M. Moran, M.B. Edwards, C. Hackley, B.O. Bargmann, K.D. Birnbaum, W.R. McCombie, G. Krouk, and G.M. Coruzzi, *Hit-and-run transcriptional control by bZIP1 mediates rapid nutrient signaling in Arabidopsis.* Proc Natl Acad Sci U S A, 2014. **111**(28): p. 10371-6.

43. Medici, A., A. Marshall-Colon, E. Ronzier, W. Szponarski, R. Wang, A. Gojon, N.M. Crawford, S. Ruffel, G.M. Coruzzi, and G. Krouk, *AtNIGT1/HRS1 integrates nitrate and phosphate signals at the Arabidopsis root tip.* Nat Commun, 2015. **6**: p. 6274.

44. Sparks, E.E., C. Drapek, A. Gaudinier, S. Li, M. Ansariola, N. Shen, J.H. Hennacy, J. Zhang, G. Turco, J.J. Petricka, J. Foret, A.J. Hartemink, R. Gordan, M. Megraw, S.M. Brady, and P.N. Benfey, *Establishment of Expression in the SHORTROOT-SCARECROW Transcriptional Cascade through Opposing Activities of Both Activators and Repressors.* Dev Cell, 2016. **39**(5): p. 585-596.

45. Doidy, J., Y. Li, B. Neymotin, M.B. Edwards, K. Varala, D. Gresham, and G.M. Coruzzi, *"Hit-and-Run" transcription: de novo transcription initiated by a transient bZIP1 "hit" persists after the "run".* BMC Genomics, 2016. **17**: p. 92.

46. Barabási, A.-L. and R. Albert, *Emergence of scaling in random networks.* science, 1999. **286**(5439): p. 509-512.

47. Albert, R., *Scale-free networks in cell biology.* J Cell Sci, 2005. **118**(Pt 21): p. 4947-57.

48. Bhalla, U.S. and R. Iyengar, *Emergent properties of networks of biological signaling pathways.* Science, 1999. **283**(5400): p. 381-7.

49. Buchler, N.E., U. Gerland, and T. Hwa, *On schemes of combinatorial transcription logic.* Proc Natl Acad Sci U S A, 2003. **100**(9): p. 5136-41.

50. Knight, H. and M.R. Knight, *Abiotic stress signalling pathways: specificity and cross-talk.* Trends Plant Sci, 2001. **6**(6): p. 262-7.

51. Vogel, M.O., D. Gomez-Perez, N. Probst, and K.J. Dietz, *Combinatorial signal integration by APETALA2/Ethylene Response Factor (ERF)-transcription factors and the involvement of AP2-2 in starvation response.* Int J Mol Sci, 2012. **13**(5): p. 5933-51.

52. Ireton, K., D.Z. Rudner, K.J. Siranosian, and A.D. Grossman, *Integration of multiple developmental signals in Bacillus subtilis through the Spo0A transcription factor.* Genes & Development, 1993. **7**(2): p. 283-294.

53. Dolmetsch, R.E., R.S. Lewis, C.C. Goodnow, and J.I. Healy, *Differential activation of transcription factors induced by Ca2+ response amplitude and duration.* Nature, 1997. **386**(6627): p. 855.

54. Chen, X., H. Xu, P. Yuan, F. Fang, M. Huss, V.B. Vega, E. Wong, Y.L. Orlov, W. Zhang, and J. Jiang, *Integration of external signaling pathways with the core transcriptional network in embryonic stem cells.* Cell, 2008. **133**(6): p. 1106-1117.

55. Khan, A., O. Fornes, A. Stigliani, M. Gheorghe, J.A. Castro-Mondragon, R. van der Lee, A. Bessy, J. Cheneby, S.R. Kulkarni, G. Tan, D. Baranasic, D.J. Arenillas, A. Sandelin, K. Vandepoele, B. Lenhard, B. Ballester, W.W. Wasserman, F. Parcy, and A. Mathelier, *JASPAR 2018: update of the open-access database of transcription factor binding profiles and its web framework.* Nucleic Acids Res, 2018. **46**(D1): p. D260-D266.

56. Pruneda-Paz, J.L., G. Breton, D.H. Nagel, S.E. Kang, K. Bonaldi, C.J. Doherty, S. Ravelo, M. Galli, J.R. Ecker, and S.A. Kay, *A genome-scale resource for the functional characterization of Arabidopsis transcription factors.* Cell Rep, 2014. **8**(2): p. 622-32.

57. Kim, D., G. Pertea, C. Trapnell, H. Pimentel, R. Kelley, and S.L. Salzberg, *TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions.* Genome Biol, 2013. **14**(4): p. R36.

58. Lawrence, M., W. Huber, H. Pages, P. Aboyoun, M. Carlson, R. Gentleman, M.T. Morgan, and V.J. Carey, *Software for computing and annotating genomic ranges.* PLoS computational biology, 2013. **9**(8): p. e1003118.

59. Love, M.I., W. Huber, and S. Anders, *Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2.* Genome Biol, 2014. **15**(12): p. 550.

60. Tian, T., Y. Liu, H. Yan, Q. You, X. Yi, Z. Du, W. Xu, and Z. Su, *agriGO v2.0: a GO analysis toolkit for the agricultural community, 2017 update.* Nucleic Acids Res, 2017. **45**(W1): p. W122-W129.

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

62. Bailey, T.L., M. Boden, F.A. Buske, M. Frith, C.E. Grant, L. Clementi, J. Ren, W.W. Li, and W.S. Noble, *MEME SUITE: tools for motif discovery and searching.* Nucleic Acids Res, 2009. **37**(Web Server issue): p. W202-8.

63. Grant, C.E., T.L. Bailey, and W.S. Noble, *FIMO: scanning for occurrences of a given motif.* Bioinformatics, 2011. **27**(7): p. 1017-8.

64. Benjamini, Y. and Y. Hochberg, *Controlling the false discovery rate: a practical and powerful approach to multiple testing.* Journal of the royal statistical society. Series B (Methodological), 1995: p. 289-300.

65. Warnes, G.R., B. Bolker, L. Bonebakker, R. Gentleman, W. Huber, A. Liaw, T. Lumley, M. Maechler, A. Magnusson, and S. Moeller, *gplots: Various R programming tools for plotting data.* R package version, 2009. **2**(4): p. 1.

66. Gibson, D.G., L. Young, R.-Y. Chuang, J.C. Venter, C.A. Hutchison III, and H.O. Smith, *Enzymatic assembly of DNA molecules up to several hundred kilobases.* Nature methods, 2009. **6**(5): p. 343.

67. Clough, S.J. and A.F. Bent, *Floral dip: a simplified method forAgrobacterium‐mediated transformation ofArabidopsis thaliana.* The plant journal, 1998. **16**(6): p. 735-743.

68. Anders, S., P.T. Pyl, and W. Huber, *HTSeq—a Python framework to work with high-throughput sequencing data.* Bioinformatics, 2015. **31**(2): p. 166-169.

69. Risso, D., K. Schwartz, G. Sherlock, and S. Dudoit, *GC-content normalization for RNA-Seq data.* BMC bioinformatics, 2011. **12**(1): p. 480.

70. Shannon, P., A. Markiel, O. Ozier, N.S. Baliga, J.T. Wang, D. Ramage, N. Amin, B. Schwikowski, and T. Ideker, *Cytoscape: a software environment for integrated models of biomolecular interaction networks.* Genome research, 2003. **13**(11): p. 2498-2504.