

RESEARCH

Open Access



Cellular clarity: a logistic regression approach to identify root epidermal regulators of iron deficiency response

Selene R. Schmittling¹, DurreShahwar Muhammad², Samiul Haque³, Terri A. Long⁴ and Cranos M. Williams^{1*}

Abstract

Background Plants respond to stress through highly tuned regulatory networks. While prior works identified master regulators of iron deficiency responses in *A. thaliana* from whole-root data, identifying regulators that act at the cellular level is critical to a more comprehensive understanding of iron homeostasis. Within the root epidermis complex molecular mechanisms that facilitate iron reduction and uptake from the rhizosphere are known to be regulated by bHLH transcriptional regulators. However, many questions remain about the regulatory mechanisms that control these responses, and how they may integrate with developmental processes within the epidermis. Here, we use transcriptional profiling to gain insight into root epidermis-specific regulatory processes.

Results Set comparisons of differentially expressed genes (DEGs) between whole root and epidermis transcript measurements identified differences in magnitude and timing of organ-level vs. epidermis-specific responses. Utilizing a unique sampling method combined with a mutual information metric across time-lagged and non-time-lagged windows, we identified relationships between clusters of functionally relevant differentially expressed genes suggesting that developmental regulatory processes may act upstream of well-known Fe-specific responses. By integrating static data (DNA motif information) with time-series transcriptomic data and employing machine learning approaches, specifically logistic regression models with LASSO, we also identified putative motifs that served as crucial features for predicting differentially expressed genes. Twenty-eight transcription factors (TFs) known to bind to these motifs were not differentially expressed, indicating that these TFs may be regulated post-transcriptionally or post-translationally. Notably, many of these TFs also play a role in root development and general stress response.

Conclusions This work uncovered key differences in -Fe response identified using whole root data vs. cell-specific root epidermal data. Machine learning approaches combined with additional static data identified putative regulators of -Fe response that would not have been identified solely through transcriptomic profiles and reveal how developmental and general stress responses within the epidermis may act upstream of more specialized -Fe responses for Fe uptake.

Keywords Iron deficiency, Epidermis, Arabidopsis, Logistic regression, Gene regulatory network, RNA-seq

*Correspondence:

Cranos M. Williams
cmwilli5@ncsu.edu

Full list of author information is available at the end of the article



© The Author(s) 2023. **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

Background

Iron (Fe) plays a critical role in essential physiological and biochemical pathways in plants [1–3]. Consequently, iron deficiency (-Fe) results in reduced plant growth and nutritional quality [4, 5], while excess iron can be lethal due to the accumulation of reactive oxygen species [6]. Plants are the primary conduit by which Fe and other essential minerals are mined from soil for animal uptake [7]. However, while Fe is the fourth most abundant element in soil, it is not easily utilized in its common form [8]. Thus, the challenge of engineering plants capable of maximizing iron uptake, while efficiently storing iron for consumption, plays an important role in addressing human and plant health issues.

Recent studies using the model plant *Arabidopsis thaliana* have led to the development of new and predictive mathematical models that describe regulatory processes involved in -Fe response in roots [9–14]. One of the most well-studied regulatory processes involved in -Fe response is Strategy I iron acquisition. During this response, the master regulator, basic helix-loop-helix (bHLH) transcription factor FER-LIKE FE DEFICIENCY INDUCED TRANSCRIPTION FACTOR (FIT) binds with bHLH heterodimers (bHLH100, bHLH101, bHLH38, and bHLH39), ETHYLENE INSENSITIVE3 (EIN3), and ETHYLENE INSENSITIVE3-LIKE1 (EIL1) [9, 14, 15] to transcriptionally activate the H⁺-ATPase (AHA2), which acidifies the rhizosphere and increases the solubility of Fe(III) near the root epidermis. Following acidification, FERRIC REDUCTASE OXIDASE2 (FRO2), reduces Fe(III) to Fe(II), and IRON REGULATED TRANSPORTER1 (IRT1), a membrane-localized metal ion transporter, transports iron across the plasma membrane [9, 16–20]. Upon iron uptake, genes such as *YSL-Like* and *FRD3*, which encode transporters of Fe chelates nicotianamine and citrate, respectively, are expressed to facilitate long-distance transport of Fe throughout the vasculature [21, 22].

High-throughput transcriptomic temporal analysis of *Arabidopsis* responses to iron deprivation has been critical in moving the field of iron homeostasis forward [9–11, 23, 24]. However, well-studied Strategy I iron acquisition transcriptional regulators (FIT, bHLH100/101, bHLH38/39) and their subsequent targets (AHA2, FRO2, IRT1) [9, 14–19, 25] are not completely described from these analyses, as molecular processes controlling Strategy I iron acquisition primarily occur in the outer cell layer of the root, the epidermis [9]. While every cell contains the same genetic information, distinct cell processes arise largely due to differences in gene regulation, thus stress response varies between cells and cell types [11, 24, 26–30]. For example, 48% of salt-responsive

genes are regulated in the cortex under high salinity, while 28% and 31% are regulated in the stele and epidermis, respectively [11, 24, 26, 27]. POPEYE (PYE), another iron-responsive transcription factor, exhibits opposing cell-type specific regulation to mediate iron bioavailability [29]. Current iron-response networks typically represent gene regulatory relationships inferred from *Arabidopsis* whole root data and do not consider the cell-specific nature of regulation. In light of the critical role of the epidermis in Fe uptake, here, we focused on this cell type as a model for identifying cell-specific regulators of -Fe response.

While clustering transcriptional data generates groups of genes with common transcriptional profiles [31–33], inferring relationships between genes from different clusters remains challenging. Inter-cluster networks capture regulatory relationships where network nodes represent clusters of genes and edges represent relationships between clusters. Connections are typically inferred using cluster centroids or mean transcription profiles [10, 34]. Using cluster centroids may not be ideal as genes within the cluster may deviate from the mean cluster profile, potentially resulting in missed relationships or connections across clusters. A method that builds inter-cluster relationships based on individual transcriptional profiles might reveal crucial transcriptional responses, particularly in cases with significant intra-cluster variations of transcriptional profiles.

TFs in inter-cluster networks derived from stress response transcriptional profiles represent potentially important regulators that are differentially expressed. When working with time-course data with a control at each time point, differential expression is defined by the induction or repression of transcript abundance under the experimental conditions (e.g., iron deprivation) with respect to the control at that time point. Potential regulators that are not differentially regulated with respect to the control but have non-negligible transcriptional abundance would not be identified by this definition of differential expression, and thus would not be included in the set of genes used for downstream gene regulatory network inference. Transcription factors that fall into this category, however, may be differentially regulated outside of transcriptional mechanisms (e.g., post-transcriptional regulation), and still have the potential to differentially regulate genes that show up as differentially expressed. Prior work has shown that post-transcriptional regulation plays a key role in stress response [35–37]. Methods that integrate transcriptional profiles with other static data sets (e.g., DNA motif information [38–40]) have the potential to identify stress regulators that would otherwise be missed by traditional approaches.

In this work, we analyzed root epidermal transcriptomic data of plants exposed to -Fe for 36 h, taking samples at 6-h increments, and compared these profiles to existing whole root time course data. Set comparisons of differentially expressed genes allowed us to identify organ-level versus cell-specific responses and to parse out developmental genes from those induced or repressed by iron deficiency. We identified several TFs with no previously known role in -Fe response that were induced or repressed by -Fe. Utilizing a unique sampling method combined with a mutual information metric across time-lagged and non-time-lagged windows, we inferred a cluster-based GRN that identified relationships between clusters of functionally relevant differentially expressed genes. This network suggested that developmental regulatory processes may act upstream of well-known Fe-specific responses. By integrating time-series transcriptome profiles with static DNA motif information and interpretable machine learning models [41–43] (e.g., logistic regression with LASSO), we identified motifs that served as crucial features for predicting differentially expressed genes. Twenty-eight transcription factors (TFs) known to bind to these motifs were themselves not differentially expressed. This results in a putative list of TFs that may contribute to the differential expression of genes under -Fe but which are themselves potentially regulated outside of transcription and would otherwise be missed by traditional approaches. Overall, this work contributes to our overall understanding of -Fe response in epidermal cells and reveals how developmental and general stress responses within the epidermis may act upstream of more specialized -Fe responses for Fe uptake.

Results

Cell-specific experimental data uncovers genes associated with -Fe response in the epidermis

Differences between genes expressed at the organ level and genes expressed specifically in the epidermis were first identified. Microarray data capturing expression activity at 7 time points (0, 3, 6, 12, 24, 48, and 72 h) and previously analyzed relative to the control sample at the 0-h time point were used to capture root (organ)-level gene expression activity [24]. Epidermis transcriptome data were obtained by sequencing marker-assisted fluorescence-activated cell sorted samples targeting root epidermal cells under -Fe and +Fe (control) at 7 time points (0, 3, 6, 12, 18, 24, and 36 h) (see Methods). These two time-course datasets were compared based on three sets of differentially expressed genes: Set R[ed], which was composed of genes differentially expressed in the microarray data at any time point with respect to the 0-h +Fe (control) time point (Fig. 1A Set R[ed]); Set B[lue], which

was composed of genes differentially expressed in the epidermis RNA-Seq data at any time point with respect to the 0-h +Fe (control) time point (Fig. 1A Set B[lue]); and Set G[reen], which was composed of genes differentially expressed in epidermis RNA-Seq data at any time point with respect to the +Fe control at that time point (Fig. 1A Set G[reen]).

Set R (organ-level) contained 2,893 DEGs (Supplementary Table S01), while Set B (cell-specific) which also used a control at the 0-h time point only, uncovered 1,950 DEGs (Fig. 1A Set B). A comparison of these two sets, which both used a 0-h control only, identified differences that were not associated with having a control at all time points. We identified 2,739 DEGs (Set G) in the epidermis using +Fe controls at all time points (Supplementary Table S02). Figure 1A compares the numbers of genes in each set. Comparing whole root and epidermal DEGs identified using only the 0-h +Fe control revealed 467 common genes (intersection of red and blue sets in Fig. 1A), suggesting we missed approximately 50% of the epidermis-specific iron response DEGs when solely using whole root data.

Control samples at every time point identify iron response genes in the epidermis

Differential expression with respect to only the 0-h time point (control) identifies genes that change significantly at the time stress is applied, resulting in a set of genes that change as a result of stress and genes that change as a result of development. Differential expression analysis with respect to a control at every time point ensures that DEGs are associated with stress response, rather than development. We isolated the iron response from developmental effects by comparing the epidermis-specific DEGs identified using a 0-h +Fe control only (Set B) to those identified using an +Fe control at every time point (Set G; Fig. 1A).

We identified 1,709 iron-responsive genes that would have been missed had we identified DEGs solely with respect to the 0-h time point control using the whole root or epidermis data (Fig. 1A). While microarray and RNA-Seq both measure transcript expression, microarray results are limited to genes with representative probes on the array, and produce relative measurements based on fluorescent activity [33, 44, 45]. RNA-Seq, on the other hand, identifies all transcripts in the sample within a detectable range using a reference genome and produces absolute transcript counts [45]. RNA-Seq's ability to identify a broader group of genes with increased sensitivity means it can identify and quantify more genes than microarray data. This might account for the differences we observed with iron-responsive genes between the

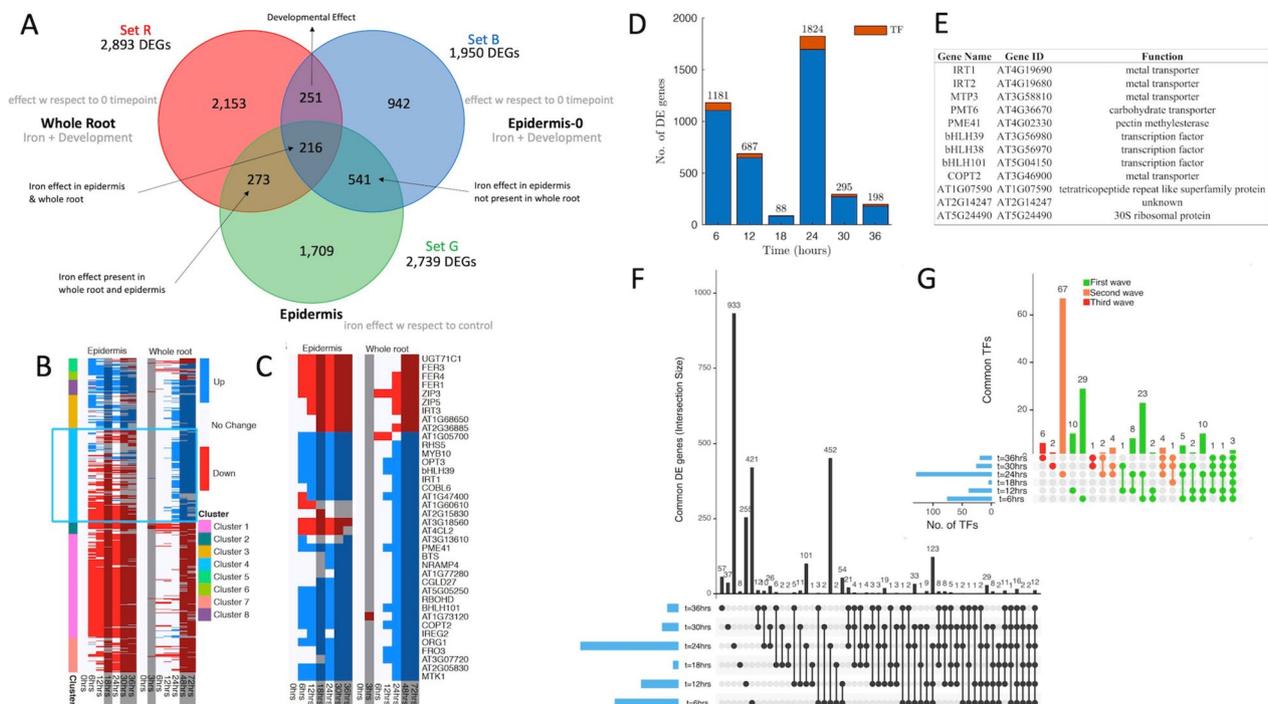


Fig. 1 -Fe whole root and epidermis specific transcriptional comparative analysis. **A** DEG counts found in 1) whole root data using control at 0 time point only (Set Red), 2) cell-specific epidermis data using control at 0 time point only (Set Blue), and 3) cell-specific epidermis data using controls at all 7 time points (Set Green). The first two show differentially expressed genes involved in stress response & development. The first captures average behavior over several cell types and the second shows epidermis only. The third shows differentially expressed genes associated with iron response in the epidermis. **B** Heatmap comparing DEG expression in whole root and epidermis using control at 0 time point only. Blue box shows genes with significantly different expression patterns. Blue indicates up-regulation, Red indicates down-regulation. **C** Heatmap of expression for 38 known iron homeostasis DEGs in whole root and epidermis data using 0-time point control showing earlier activation in the epidermis vs. whole root. **D** DEG Gene counts at each time point identified using epidermis data using a control at each time point. The largest number of DEGs is seen at 24 h. The second largest is seen at 6 h. **E** List of genes that are differentially expressed at all 7 time points including 3 metal transporters and 3 known regulators of -Fe. **F** Upset diagram comparing sets of differentially expressed (under -Fe in the epidermis) genes at each time point. **G** Upset diagram showing the subsets of differentially expressed (under -Fe in the epidermis) transcription factors at each time point

microarray and RNA-Seq datasets. However, given the comparable number of genes identified using microarray or RNA-Seq, it seems likely that this effect is negligible.

Epidermal transcriptomic expression patterns reveal expression that differs from whole root patterns

A comparison of the profiles of the 467 genes that were differentially expressed in both the whole root and the epidermis transcript data (Fig. 1A) uncovered several differences (Figs. 1B and C). Figure 1B reveals differences across all 467 genes as a heatmap of clustered expression patterns for the epidermis vs. the whole root. Eight clusters were identified using hierarchical clustering of expression patterns from whole root and epidermis (Fig. 1B). Clusters 1, 3, 5, 6, 7, and 8 showed delayed expression in the whole root (48 h) compared to the epidermis (6 h) and in some cases, reversed directionality (i.e., activated vs. repressed and vice versa). Cluster 4 showed a delayed activation pattern (12 and 24 h vs.

6 and 12 h in the epidermis) and an almost complete change in directionality: up-regulation in the whole root vs. down-regulation in the epidermis.

Thirty-eight known iron homeostasis genes exhibited earlier activation or repression in the epidermis, compared to their expression in the whole roots (Fig. 1C, 6 hours vs. 12 and 24 h in the whole root). IRT1, regulated by FIT and considered the main iron transporter involved in iron uptake from the rhizosphere, [9, 16, 25] and MYB10, an -Fe-induced transcription factor [46], were upregulated in the epidermis at 6 h, whereas they did not appear upregulated in the whole root until 24 h. Two FIT interactors, bHLH39 and bHLH100, as well as OPT3 and COPT2, iron and copper transporters, respectively, were upregulated after 12 h of exposure to -Fe in the epidermis, but at 24 h in the whole root dataset. FER1 and FER4, genes that bind iron to prevent oxidative stress [47], were repressed in the epidermis as early as 6 h but repressed at 24 h of -Fe in the whole root

dataset. Earlier induction or repression of known iron homeostasis genes in the epidermis compared to that in the whole root suggests that early epidermal activity is masked by collective activity in the whole root. These findings are supported by previous studies that characterized cell-type specific transcriptional changes under salt and -Fe stress [11, 24, 26].

Epidermal expression patterns reveal response waves

We analyzed the 2,739 DEGs identified using the epidermis data and a control at every time point (Set G) to assess overall epidermal expression dynamics. We found the most DEGs and differentially expressed TFs at 6 (1,181 DEGs) and 24 h (1,824 DEGs) (Fig. 1D), while the 18-h time point resulted in the lowest number of transcriptional profile changes (88 DEGs) and the fewest TFs (4) (Fig. 1D). This suggests that these time points represent the initiation of transcriptional changes that induce a cascade of downstream responses. Twelve genes, listed in Fig. 1E, were differentially expressed at all sample time points, including four metal transporters and three known -Fe regulators. STOP2, previously identified as regulating several genes for aluminum stress and low pH tolerance, was differentially expressed in our data at 5 out of 6 time points [48]. Tsai and Schmidt (2020), investigating transcriptomic changes under optimal and high pH under -Fe, found that STOP2 was moderately induced under optimal pH, but highly repressed under high pH in iron-deficient conditions [49]. Since -Fe triggers the activity of epidermal proton pumps to acidify the rhizosphere and increase iron solubility [19], STOP2 likely plays an indirect role in iron response by regulating prevailing pH conditions. While STOP2 was previously identified as a minor isoform of STOP1, which controls pH and Al response [48], we did not find STOP1 to be substantially regulated in our datasets. Thus, STOP2 may play a more prominent role in controlling -Fe-induced pH changes at the rhizosphere-epidermal interface.

To characterize gene activity, we described each gene's activity by the time points at which they were differentially expressed (Fig. 1F). Focusing on activity patterns of TFs (Fig. 1G), we identified three waves of transcriptional responses: wave 1 (activity before 12 h) contained the largest number of TFs (94), wave 2 (activity between 12 and 18 h) included 78 TFs and wave 3 (activity only after 30 h), 9 TFs. This activity suggests an organized cascade of transcriptional stress response which is discrete from wave to wave.

Expression analysis and inter-cluster GRN reveal epidermis-specific regulators and response genes

Using the Dirichlet process Gaussian process mixture model (DPGP) clustering algorithm [50], we generated

50 co-expression clusters using the max normalized gene expression of our 2,739 DEGs (Supplementary Table S02). Two clusters, 25 and 26, were enriched for iron-related Gene Ontology (GO) terms (adjusted p-value of 0.05). Most genes and TFs in these two clusters were previously known iron response genes. We did not find enriched GO terms in other clusters, probably due to smaller cluster sizes. However, we identified multiple cluster-specific TFs within the smaller clusters with known regulation adjacent or related to -Fe roles in development and stress response. Clusters 36 and 37 contained GL2, TRY, RHD6, and LRL1, TFs related to root hair development, while Cluster 1 contained several TFs involved in ABA response. Expression patterns for the 50 clusters are provided in Supplemental Figure S1.

To identify new functional relationships, we inferred and examined inter-cluster relationships. Previous studies and our results indicate that genes with similar biological functions show similar expression patterns and are more likely to belong to the same cluster (intra-cluster relationships). Likewise, it is logical to assume clusters containing regulator genes should have some causal relationship to clusters containing target genes (inter-cluster relationships). Previous works that developed cluster networks used cluster means or centroids to quantify cluster correlation [34, 51]. However, the range of transcriptional profiles within a cluster (intra-cluster variation) may result in the cluster centroid being a poor representation of all genes within the cluster (Fig. 2A) and result in the inability to infer informative causal inter-cluster relationships. To address this, we implemented a sampling-based scheme that measured the strength of a cluster relationship based on the distribution of Mutual Information (MI) [52] between individual genes belonging to clusters. This allowed us to incorporate gene-level expression patterns while maintaining computational efficiency. MI is a well-known metric for quantifying causality between pairs of genes [53].

Using our sampling technique for each inter-cluster connection, we generated probability distribution functions of MI scores (Fig. 2B). High MI scores indicate strong pairwise relationships between genes, which we expected to find in the tails of these distributions (higher MI values). We ranked all cluster connections using no time lag and a 6-h time lag to catch connections that would be missed using only a 0-time lag. We defined α_{ij} as the 90th percentile value of the MI score distribution calculated for genes in the i th and j th clusters. Larger values of α indicated MI distributions with stronger connections in the top 10% of their MI values while smaller values of α indicated MI distributions with weaker connections in the top 10% of their MI values. We generated the distribution of α_{ij} for all

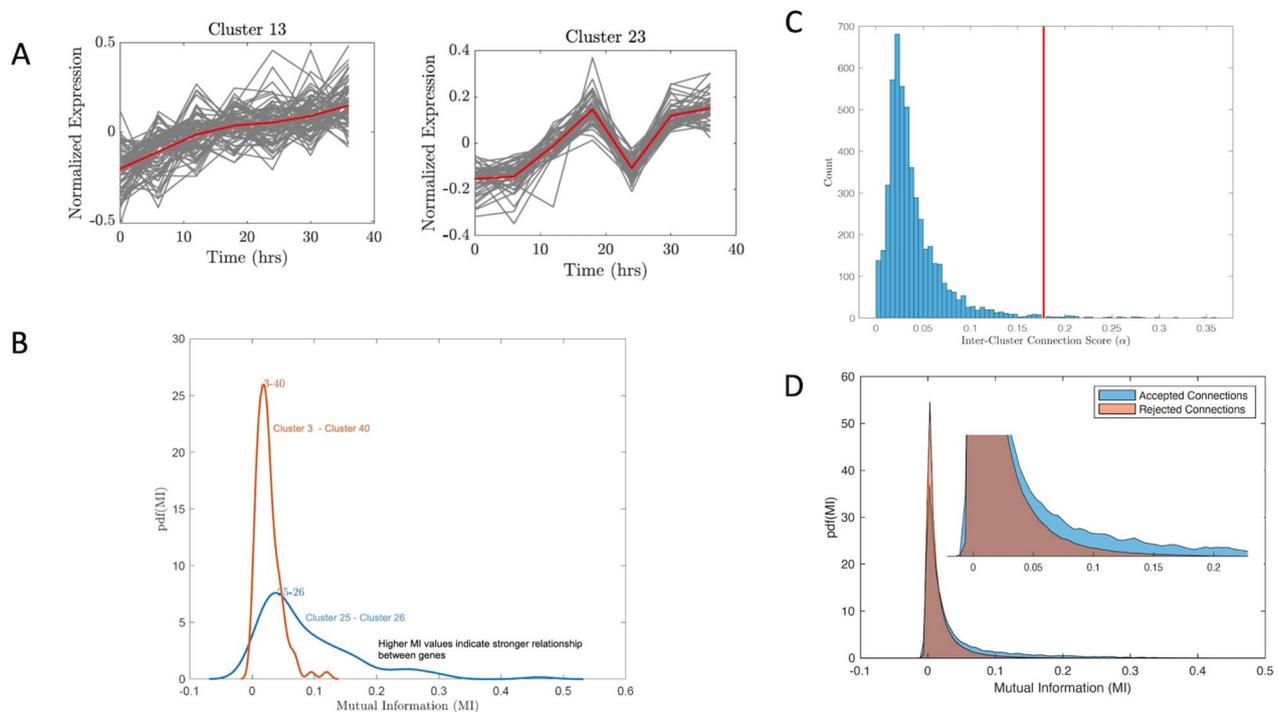


Fig. 2 Computationally efficient mutual information sampling technique to infer gene regulatory network. **A** These examples show cases in which the mean is not a good representation of individual gene expression, **B** Example distribution of inter-cluster MI scores for 2 clusters, **C** Distribution of alpha values for all clusters **D** MI scores for accepted vs. rejected connections for all clusters

ij cluster pairs. We then identified a natural separation in this distribution (Fig. 2C) from which we could distinguish MI distributions with strong connections from distributions with weak connections. We identified a threshold of $\alpha = 0.1775$ as shown in Fig. 2C, which was the center value of an empty bin that created a natural separation between lower and higher values of α . Examination of the distribution of Mutual Information values for connections that were accepted vs. rejected using this threshold of α (Fig. 2D) shows that while the mode of their distributions is similar, the accepted cluster connections have a higher density in the tails.

We identified 33 potential interactions among 14 clusters (Fig. 3) from the accepted connections. Connections uncovered between FIT and binding partners, bHLH100, bHLH101, and bHLH39 have been confirmed experimentally in prior works [14, 54]. We used known TFs in the clusters to categorize biological functionality using existing literature and the TAIR database [55]. Our network included connections between Cluster 1, which contains FER1; Cluster 28, which contains FER3 and 4; and Cluster 39, which contains NAC042 (JUB1) and AT3G20340. Sudre et al. (2013) found 54 genes, including NAC042 and AT3G20340, were significantly affected under 3 conditions: *atfer1-3-4* mutant vs. WT both under iron sufficiency, *atfer1-3-4* under excess iron vs. WT, and

atfer1-3-4 vs. WT both under excess iron [56]. These findings support our network's biological relevance and provide future avenues of investigation on -Fe response in epidermal cells. Potential novel -Fe responsive genes identified also have other biological functions, e.g., root hair development (GL2, TRT, RHD6), ABA response (WRKY31, CTH/AtTZF1), and nitrate/phosphate response (WRKY42, HRS1, NAC42). Two clusters containing TFs associated with root hair development, 1 and 37, are connected to Cluster 39 which contains nitrate/phosphate response TFs. Liu et al. (2020) reviewed nitrate regulation associated with lateral root and root hair development and reported that both are gradually inhibited in a homogeneous nitrate environment [57]. These relationships suggest that iron deprivation triggers developmental and more general abiotic stress response within the epidermis that then induces downstream specialized processes for iron uptake activities.

Logistic regression models identify TFs associated with motifs important for predicting differential expression in epidermal data

In the previous section regulators of iron epidermal deficiency response were identified from a pool of differentially expressed genes. By integrating static data (DNA motif information) and our time-series transcriptomic

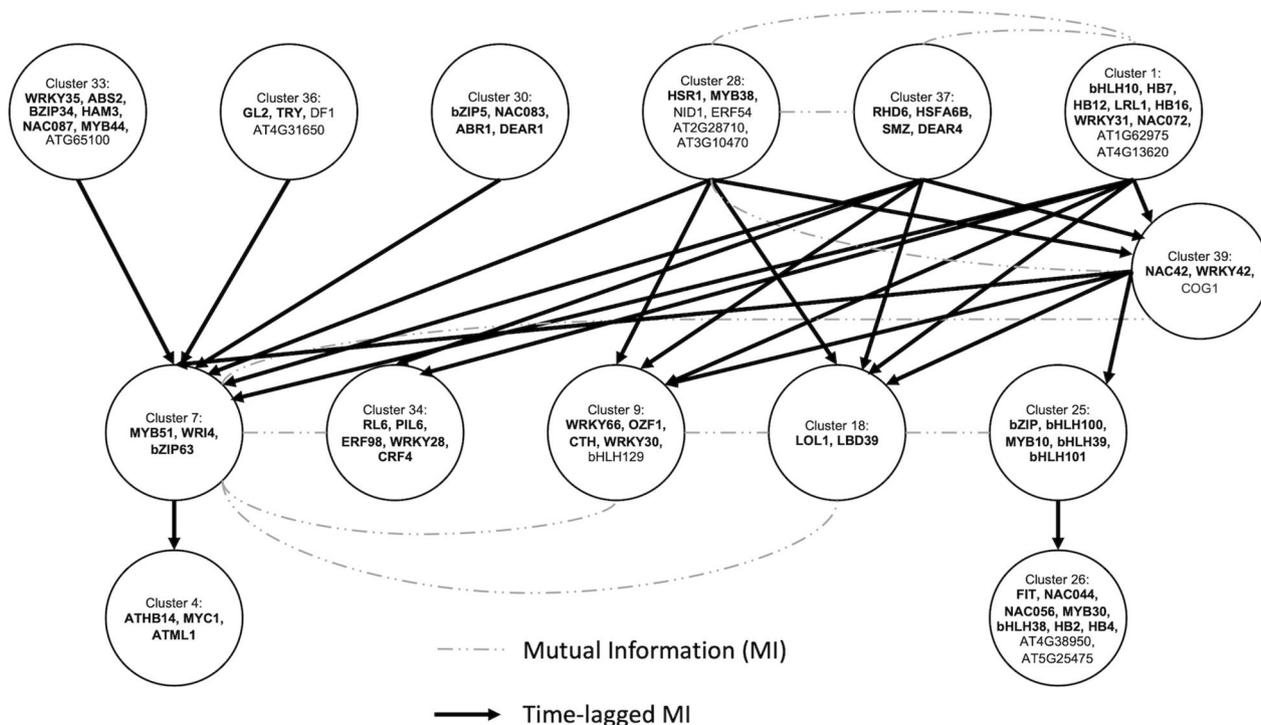


Fig. 3 Inter-cluster GRN. Inter-cluster GRN generated using DPGP clustering coupled with sampling-based scheme to measure the strength of a cluster relationship. Important TFs for each cluster are labeled. Dash lines indicate MI between gene pairs and solid arrows represent time-lagged MI

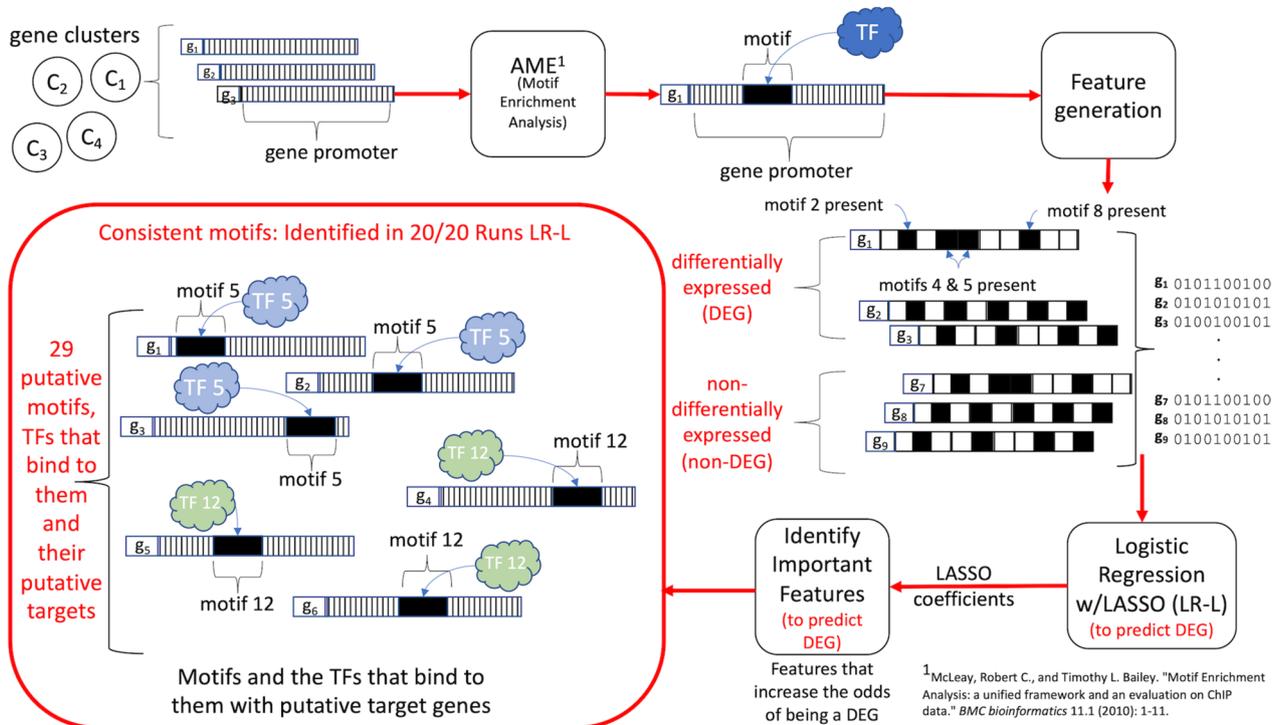


Fig. 4 Machine learning pipeline to identify motifs/TFs important for predicting differential expression

data into interpretable machine-learning models, we uncovered regulators that influence genes that were differentially expressed under -Fe, but which were not differentially expressed themselves. In our pipeline (Fig. 4), we first identified known cis-elements enriched in the promoter region 1000 bp upstream of the transcription start site as in Schwarz et al. (2020) using genes in the clusters previously generated using the DPGP algorithm [50, 58]. We removed cluster 50 from further analysis since no motifs were identified for that cluster.

We found 337 motifs enriched in our gene clusters (Supplementary Table S03) and generated features using the presence (0) or absence (1) of a motif in the promoter of each gene (see Methods). These features were input into a logistic regression (classification) model to identify motifs important for predicting differential expression. Given known class labels (e.g., differentially expressed, 1, and non-differentially expressed, 0), classification models can identify predictors that are associated with a response [58–62]. While previous works have successfully used Random Forest and Support Vector Machines, logistic regression models with LASSO regularization provide highly interpretable results in addition to feature selection for this task [41–43, 63].

We hypothesized that motifs that increased the likelihood of a gene being differentially expressed would be associated with stress response. We used the model coefficients to identify important motifs. In logistic regression, coefficients represent how a predictor increases or decreases the log odds of the response relative to the positive (i.e., response=1 or “differentially expressed”) class. Since coefficients are associated with changes in the log odds, we chose motifs for which $e^x > 1$, where x is the coefficient [64]. Due to stochasticity introduced by LASSO regularization, we ran the algorithm 20 times and identified motifs that were consistent predictors across all runs. This uncovered 29 consistent motifs of interest (MOI; Supplementary Table S04). By generating a model that included all differentially expressed genes, as opposed to one model per cluster, we were able to identify MOIs that were influential over multiple clusters. We extracted TFs known to bind to the 29 consistent MOIs (putative regulators, provided in Supplementary Table S04) and the set of genes with the MOI in their promoter (putative targets, provided in Supplementary Table S05) using the motif-finding tool, AME [65], which identifies biologically verified TFs using the DAP-seq database [66].

Predictive models identify TFs that may be post-transcriptionally regulated

To better characterize the function of the identified TFs, we explored their expression patterns and found that

only one TF, ANAC070 (AT4G10350), was differentially expressed with respect to the control (+Fe) at any time point. We performed GO analysis on these putative regulators to determine their biological relevance and found that 6 putative regulators were associated with terms specific to root development, one with the epidermis, and two with hormone pathways (Supplementary Table S04). All of these activities have been associated with iron stress response and root hair development, which is affected by -Fe [67, 68]. Next, we determined the extent to which putative targets of these TFs included known iron homeostasis genes. Thirteen of our 29 putative regulators had at least one known iron homeostasis gene in their putative target sets. Putative targets included known iron stress-responsive genes: BRUTUS(BTS), IRT1, and FRO2, as well as regulators (TFs) of the iron stress response including FIT, bHLH100, and bHLH101 (Supplementary Table S05).

To assess the gene function of the putative targets for each of our TFs, we performed GO enrichment analysis and identified 876 unique, statistically significant (adjusted p -value < 0.05) terms across all putative regulators. Supplementary Figure S02 shows a subset of terms identified. A complete list of TFs, their GO terms, and FDR are included in supplemental data (Supplementary Table S06). The putative targets of six of our 29 consistent TFs produced no statistically significant terms.

Since changes in root hair development are a hallmark phenotype of iron deprivation and these changes occur in the epidermis [11, 24], we focused on statistically significant terms from the GO analysis that included “iron”, “epidermis” and “root”. Targets for eleven of our TFs had terms associated with iron response and homeostasis, while 9 had terms associated with root morphogenesis, development, or regeneration, and 5 had terms associated with epidermis development and cell fate (Supplementary Tables S07, S08, and S09). This suggests the MOIs/TFs we identified could be involved in iron response in the epidermis. Two of the MOIs/TFs with “root” in their GO terms and one with “epidermis” had no known iron homeostasis genes in their targets and could represent previously unidentified iron response genes.

Non-differentially expressed regulators are associated with iron-related pathways and activity

Many of the putative regulators we identified were expressed at some level (Supplementary Figure S03) and several have -Fe RPKM measurements with similar magnitude to bHLH101, a FIT binding partner that is differentially expressed under -Fe compared to +Fe control at every time point (Supplementary Figure S03). Thus, we suspect that these non-differentially expressed transcription factors represent high-level regulators that

A

Transcription Factor		Literature Review					Additional Information
AGI	Alias	DELLA/GA Pathway	Associated with Abiotic Stress Response	Root Development	Epidermis	Evidence of Post-Transcriptional Regulation	Some Putative Targets
AT2G35530	bZIP16	Hsieh et al., 2012					BTS, bHLH100, ZIF1
AT5G61270	PIF7	Liu et al., 2021	Paik et al., 2017			Huang et al., 2018; Zhang and Ding, 2021	BTS, bHLH100, ZIF1, bHLH38
AT3G45610	DOF6		Noguero et al., 2013	Derbyshire et al., 2008		Noquero et al., 2013	bHLH100/101, IRT1, FRO2
AT4G30980	bHLH69/LRL2			Han et al., 2020; Lin et al., 2015	Bruex et al., 2012		
AT3G47500	CDF3		Barah et al., 2016; Kreps et al., 2002				FRO2, bHLH29
AT5G66940	ATDOF5.8		He et al., 2015; Waqas et al., 2008				bHLH100, BTS
AT3G50700	IDD2	Jin, Y., 2015					FIT
AT3G15170	CUC1					Mitsuda et al., 2013	
AT1G01060	LHY1	Maurer et al., 2014	Roeber et al., 2011			Hernando et al., 2017; Kojima et al., 2011; Mateos et al., 2018; Staiger and Köster, 2011	
AT5G18090	AT5G18090		Wei et al., 2008	Wei et al., 2008			

B

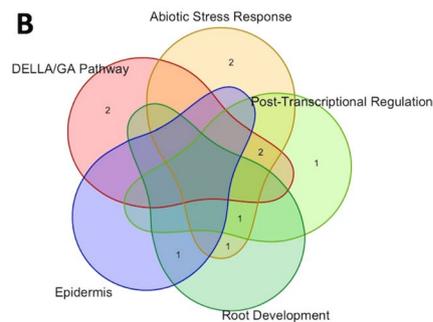


Fig. 5 Subset of putative regulators identified using logistic regression with LASSO. **A** Putative TFs investigated in literature review. Subset included 13 TFs with known iron response genes in their putative targets and three additional TFs. References were identified for the topics listed in the table for 10 of the 16 TFs investigated. **B** Venn diagram showing overlap between topics covered in the literature review for TFs listed in **A**

themselves are regulated post-transcriptionally, not unlike several other -Fe-responsive TFs [69, 70]. A review of the literature indicates that post-transcriptional regulation is associated with four of our non-differentially expressed putative regulators, including the DOF family of transcription factors [71]. CUC1 was found to be targeted by miR164, which provides evidence of post-transcriptional regulation [72], and Huang et al. (2018) determined that PIF7 activity is regulated by phosphorylation in shade-induced nuclear localization [73], while LHY1 has also been found to be post-transcriptionally regulated [74–77].

Post-transcriptional regulation is critical in abiotic stress response [35, 37, 78]. To explore whether the regulators we identified might be involved in iron deficiency response in the epidermis, we performed an extensive literature review. We focused on the subset of 13 putative regulators with known iron response genes in their targets and included three additional TFs with no known iron genes in their targets (CUC1, LHY1, and AT5G18090) to explore how these putative regulators and their targets might be associated with -Fe. Ten of the 16 TFs we reviewed were associated with one or more activities implicated in -Fe response (Fig. 5A and B) [71–77, 79–94]. These TFs represent a small testable set of previously unidentified iron stress response regulators.

Six of the TFs are implicated with abiotic stress response including, PIF7 [90], DOF6 [71], CDF3 [79, 86], ATDOF5.8 [83, 91], LHY1 [88] and AT5G18090 [92]. Putative targets associated with some of these TFs include BTS, bHLH100/101, bHLH38, bHLH29, IRT1, FRO2, and

ZIF1 (Fig. 5A). While a universal stress response pathway could explain this, research by Iyer-Pascuzzi et al. (2011) did not find evidence of one [27]. However, they and others have found that regulators may respond to multiple stresses [24, 27, 86].

Four TFs are associated with the DELLA/gibberellin pathway, which has been implicated in FIT regulation [68]: bZIP16 [84], PIF7 [89], IDD2 [85], and LHY1 [88]. Putative targets of these four TFs include BTS, FIT, bHLH100, bHLH38, and ZIF1. LHY1 has not previously been identified as an iron stress response regulator. However, Maurer et al. (2014) found that LHY1 is more highly expressed under -Fe in a triple mutant background (*bhlh39-1 bhlh100-1 bhlh101-1*) [88]. Given that changes in LHY1 are associated with mutants of its putative targets a feedback mechanism may be involved.

Three TFs are associated with root expression or development: DOF6 [81], bHLH69/LRL2 [80], and AT5G18090 [92]. Root development and changes in root hair length are known phenotypes of -Fe [11, 95, 96]. Putative DOF6 targets include bHLH100/101, IRT1 and FRO2. DOF6 is associated with root radial pattern formation according to TAIR and is expressed in all three developmental zones (meristematic, elongation, and differentiation), but is pericycle-specific [81]. bHLH69 is involved in root hair formation and root epidermis development, and mutants show root hair defects [80]. AT5G18090 is induced in response to manganese deficiency, which is associated with stimulated root hair elongation [92]. Notably, there is evidence showing iron

and manganese crosstalk via IRT1, an iron transporter that nonspecifically transports manganese under -Fe [25], increasing its levels in roots [11, 29].

Only one of our TFs, IDD2, is associated with epidermal root hair differentiation [85]. However, the relationship between this TF and the DELLA/gibberellin pathway makes it a prime TF for further investigation to determine its role in the -Fe response. While biological confirmation is required, this work uncovered connections between non-differentially expressed regulators and mechanisms associated with iron stress response and provides several avenues of future inquiry.

Discussion

Gene set comparisons between organ- and cell-level data using a 0-h +Fe control sample identified 1,483 genes associated with epidermal activity that would not have been found without cell-specific measurements (Fig. 1A). By including a +Fe control sample at every time point, an additional 1,709 genes were identified that would have been missed if only a 0-h control were used (Fig. 1A). These results highlight the need for cell-level data with controls at every time point, as well as the benefit of gene set comparisons across datasets. While a complete comparison between whole root and epidermis data was not feasible since microarray data does not contain the whole transcriptome, our comparative transcriptional analysis shows evidence of early activation in the epidermis in response to -Fe along with complete transcriptional changes between the whole root (up-regulated) and the epidermis (down-regulated; Fig. 1B and C).

The cluster GRN (Fig. 3), developed using gene-level comparisons in a computationally efficient manner, identified relationships between transcription factors that mirror those found using whole root data, while also

revealing novel putative regulators that act upstream of known -Fe response genes. In many cases, these early response regulators play a role in development and stress responses. A logistic regression with LASSO modeling approach uncovered regulators that may act upstream of known -Fe response processes, yet are not transcriptionally regulated by -Fe. Plants rely on a highly tuned regulatory system for stress response due to their sessile nature. Regulators revealed by our approach, such as CUC1, LHY1, and IDD2 may be post-transcriptionally controlled to fine-tune responses and enable quick transitions in response to Fe deprivation within the epidermis [36]. Notably, many of these putative upstream regulators also play a role in development and general stress response [85, 88, 97].

We propose a potential connection between the high-level putative regulators we identified, which are associated with development and general stress response, and -Fe response in the epidermis. The root epidermis is highly sensitive to stress and readily exhibits developmental, physiological, and biochemical alterations in response to nutrient availability [93, 94]. One of the most well-studied responses to -Fe stress is the induction of the Strategy I response within the root epidermis. This response relies upon membrane localization and induction of the critical FRO2, IRT1, and AHA2 protein complex that induces rhizosphere acidification, ferric reduction, and Fe uptake into epidermal cells [98].

Our analysis revealed that several known developmental regulators, such as DOF6, bHLH69/LRL2, WRKY42, and RHD6, might directly or indirectly target FRO2 and IRT1, as well as their bHLH regulator, FIT, and FIT's binding partners, bHLH100 and bHLH101, as well as other Fe homeostasis genes. Taken together our spatiotemporal RNA-seq and modeling of epidermal-specific

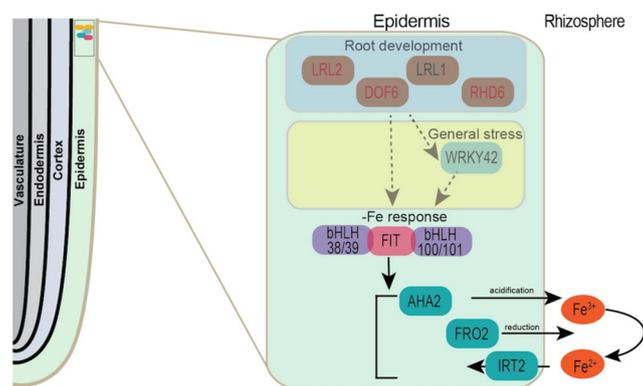


Fig. 6 Hypothesized root epidermal -Fe mechanism. Model depicting epidermal root development (blue box) and general stress responses (yellow box) epistatic to the known -Fe response. Dash grey arrows indicate direct or indirect connections identified by the DPGP clustering and Logistic regression with LASSO modeling algorithms on the -Fe spatiotemporal RNA-seq data. Solid black arrows indicate confirmed direct -Fe interactions. Genes in red are potentially post-transcriptionally regulated

transcriptional responses reveal a mechanism whereby developmental and general stress responses are epistatic or parallel to the early -Fe stress response within the root epidermis, which then later induce more specialized -Fe responses for Fe uptake (Fig. 6). Figure 6 hypothesizes connections between transcription factors associated with development, general stress response, and -Fe response.

Conclusions

This work uncovered key differences in -Fe response identified using whole root data vs. cell-specific root epidermal data, and using a 0-h control vs. controls at every time point. Machine learning approaches, which employed interpretable models combined with additional static data, identified potential high-level regulators of -Fe response that would not have been identified solely through transcriptomic profiles. These regulators reveal how developmental and general stress responses within the epidermis may act upstream of more specialized -Fe responses for Fe uptake.

Methods

The *Arabidopsis* pWER::GFP epidermis marker line (WEREWOLF; AT5G14750: Lee & Schiefelbein, 1999) in Columbia (Col-0) accession was used to perform all experiments [99]. Seeds were surface sterilized using 70% ethanol for 2 min followed by 30% bleach and 0.02% Triton X-100 solution for 15 min. Seeds were rinsed three times with sterile water and stratified at 4 °C for 2 to 3 d before being plated on nylon mesh (Genesee Scientific Cat 57–103) on top of solidified media to facilitate transfer. Seeds were grown in a vertical position in a Percival incubator with 16 h of daily illumination and 8 h of dark at 22 °C. Iron-sufficient (+Fe) and deficient media is standard Murashige and Skoog media (Caisson Labs Cat MSP33-50LT) with 0.05% MES, 1% sucrose, 1% agar, and 0.1 mM FeEDTA. Ferrozine (300 μM), an iron chelator, was added to make deficient media.

Seedlings were shifted from iron-sufficient media to sufficient and deficient media, on the 7th day of growth, equating to 0-h. After 7d transfer, an additional 6 h, 12 h, 18 h, 24 h, 30 h, and 36 h of growth occurred. After which, roots were harvested for protoplasting followed by fluorescence-activated cell sorting [100]. Three replicates of the epidermis samples were collected and analyzed.

RNA was extracted using RNeasy Plant RNA Purification Kit (Qiagen Cat. 74,904). cDNA synthesis and amplification were performed using the SMARTer Low Input RNA Kit (TaKaRa Cat. 634,940) for sequencing. North Carolina State University Genomic Science Library facilities sequenced the library using an Illumina HiSeq 2500 sequencing machine, with 125 bp single-end reads.

Adapter contamination and low-quality reads were assessed using fastQC [101]. Adapter contents were trimmed, and low-quality reads discarded using fastq-mcf [102]. Clean reads were mapped to The Arabidopsis Information Resource (TAIR)10 reference genome [55] using tophat2 [103–105]. More than 92% of reads were mapped to the reference genome for all the samples. From the mapped reads we obtained the read count (RPKM) for each gene using the RSubread package [106].

Differentially expressed genes (DEGs) were identified using edgeR [107] package, using glmQLFtest for DE analysis. DE genes were identified with respect to the 0-time point or with respect to the control using adjusted p-values [108]. A maximum false discovery rate (FDR) of 0.05 and minimum log fold change threshold of 0.75 was applied for identifying differentially expressed genes.

To cluster gene expression trajectories, we used a Dirichlet Process Gaussian Process (DPGP)-based algorithm [50]. DPGP can capture time course dependencies and is suitable for modeling time series data. Data was max normalized prior to clustering using the following equation

$$N_t^i = \frac{R_t^i - \frac{\sum_{k=1}^T R_k^i}{T}}{R_k^i}$$

Where N_t^i is normalized expression of gene i at sampling time point t . T is the total number of time samples, and R_k^i is the RPKM gene expression value of gene i at time t .

Latin Hypercube Sampling (LHS) [109] was used to randomly sample TFs from each DPGP-generated cluster (with repetition). We calculated 0- and 6-h time lagged mutual information (MI) scores between pairs of randomly sampled TFs from different clusters using [53]. Mutual information between the expression patterns of two genes (random variables) (G_1, G_2) is defined as

$$I(G_1; G_2) = \iint f(G_1, G_2) \log \log \left(\frac{f(G_1)f(G_2)}{f(G_1, G_2)} \right) dG_1 dG_2$$

Where $f(G_1, G_2)$ is the joint distribution, and $f(G_1)$, $f(G_2)$ are marginal distributions of the random variables representing the gene expression patterns. We generated a distribution from the MI scores calculated for each pair of clusters. We ranked all possible cluster connections based on the 90th percentile (near the tail end) of the MI value and called this score α . We calculated α for all possible cluster connections, generated a distribution of α and identified an appropriate threshold which was applied to identify accepted connections between clusters.

Promoter regions 1000 bp upstream [58] of the transcription start site for genes in each of the 50 DPGP

clusters were identified. We then used the Analysis of Motif Enrichment (AME) algorithm [65] using default settings and the DAP-seq database [66] to identify enriched motifs. AME provided a list of motifs, TFs that bind to them, and the genes that contained the motif in their promoter (putative targets). We generated a binary feature for each gene by encoding presence (1) or absence (0) of the motif for each gene. We encoded an equal number of non-differentially expressed genes (absolute log₂FC for all time points < 0.251). We chose a more restrictive threshold than Schwarz et al. (2020) (0.4), to provide us with enough negative samples while ensuring that expression was negligible [58]. We set class labels for differentially expressed to 1, and non-differentially expressed to 0. These features and labels were used as input to Logistic Regression models with LASSO.

We performed logistic regression with LASSO using the R package glmnet. Important features were identified by using model coefficients for which (exp(x)) was > 1 where x was the model coefficient [64]. Since the use of LASSO can result in different coefficients from run to run, we generated 20 models and focused on motifs identified in all 20 runs. This allowed us to capture the variability of the algorithm and maintain the exploratory nature of the task.

GO terms associated with TFs that are known to bind to our motifs were identified using the TAIR database [55]. GO enrichment of putative targets was performed using the R library topGO [110].

Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s12864-023-09714-6>.

Additional file 1: Figure. S01. Expression profiles of 2,739 DEGs. 50 clusters were identified using DPGP clustering algorithm. Cluster 25 and 26 were enriched in iron related GO terms. **Figure. S02** GO Enrichment of putative targets for a subset of GO terms. **Figure. S03.** Expression values of 29 TFs identified using Logistic Regression.

Additional file 2: Table 01. List of 2,893 differentially expressed genes identified in whole root data. **Table 02.** List of 2,739 differentially expressed genes identified using epidermis specific RNA-seq data with controls at all time points. Fields included are AGI, Alias(es), Cluster number, and a flag indicating whether the gene was differentially expressed at 6, 12, 18, 24, 30 or 36 hours. **Table 03.** List of 337 motifs identified using AME (65) with the DAP-seq (66) database. **Table 04.** List of 29 Consistent transcription factors identified in 20 runs of logistic regression with LASSO. Name, AGI and GO terms identified on arabidopsis.org (TAIR) are provided. **Table 05.** List of putative targets associated with each of the 29 consistent transcription factors. List includes the name of transcription factor, AGI of target, flag indicating whether target is 1) a known iron gene and/or 2) a transcription factor. **Table 06.** GO terms enriched in the group of targets associated with each of 29 consistent transcription factors. **Table 07.** Subset of GO terms from Supplementary Table 03 that include reference to "iron". GO ID, GO term, adjusted p-value and transcription factor name or AGI provided. **Table 08.** Subset of GO terms from Supplementary Table 03 that include reference to "root". GO ID, GO term, adjusted p-value and transcription factor name or AGI provided. **Table 09.**

Subset of GO terms from Supplementary Table 03 that include reference to "epidermis". GO ID, GO term, adjusted p-value and transcription factor name or AGI provided.

Acknowledgements

We would like to thank Dr. Anna Matthiadis for her assistance with cell sorting, as well as Logan Kallam and Dr. Hsuan "Tina" Chou for technical assistance.

Authors' contribution

D.M. experimental design, experimental data preparation and analysis, co-wrote article. S.H. RNA-seq, differential expression, and gene set analysis, GRN generation, and co-wrote article. S.S. cis-element identification, Logistic regression model and interpretation, literature review for TFs identified, co-wrote article. C.M.W. and T.A.L. design methodology and drafting of manuscript.

Funding

This material is based upon work supported by the National Science Foundation under Grant No. 1247427, by the USDA National Institute of Food and Agriculture, Hatch project 1010901, and by the National Institute of Health under Grant No. K99GM146026.

Availability of data and materials

Raw sequence data (FASTQ files) and RPKM values have been submitted to NCBI GEO and are available for download (Accession # GSE228011). *Arabidopsis thaliana* pWER::GFP epidermis marker line (WEREWOLF; AT5G14750: Lee & Schiefelbein, 1999) in Columbia (Col-0) seeds were generated as described in Lee & Schiefelbein, 1999, and donated to, maintained by and purchased from Arabidopsis Biological Resource Center; stock number CS66493 and this material was used with permission.

Declarations

Ethics approval and consent to participate

Arabidopsis pWER::GFP epidermis marker line (WEREWOLF; AT5G14750: Lee & Schiefelbein, 1999) in Columbia (Col-0) seeds were generated as described in Lee & Schiefelbein, 1999, and donated to, maintained by and purchased from Arabidopsis Biological Resource Center; stock number CS66493 and this material was used with permission. The experimental research works on the plants described comply with institutional, national and international guidelines.

Consent for publication

Not applicable.

Competing interests

The authors declare no competing interests.

Author details

¹Department of Electrical & Computer Engineering, North Carolina State University, Raleigh, USA. ²Department of Biosciences, Rice University, Houston, USA. ³Life Sciences Customer Advisory, SAS Institute Inc, Cary, USA. ⁴Department of Plant & Microbial Biology, North Carolina State University, Raleigh, USA.

Received: 14 March 2023 Accepted: 3 October 2023

Published online: 18 October 2023

References

- Connorton JM, Balk J, Rodríguez-Celma J. Iron homeostasis in plants—a brief overview. *Metallomics*. 2017;9:813–23.
- Kim SA, Gueriot ML. Mining iron: Iron uptake and transport in plants. *FEBS Lett*. 2007;581:2273–80.
- Vose PB. Iron nutrition in plants: A world overview. *J Plant Nutr*. 1982;5:233–49.
- Walker EL, Connolly EL. Time to pump iron: iron-deficiency-signaling mechanisms of higher plants. *Curr Opin Plant Biol*. 2008;11:530–5.

5. Zuo Y, Zhang F. Soil and crop management strategies to prevent iron deficiency in crops. *Plant Soil*. 2011;339:83–95.
6. Zaid A, Ahmad B, Jaleel H, Wani SH, Hasanuzzaman M. A critical review on iron toxicity and tolerance in plants: role of exogenous phytoprotectants. In: *Plant Micronutrients*. 2020. 83–99.
7. Saini RK, Nile SH, Keum Y-S. Food science and technology for management of iron deficiency in humans: A review. *Trends Food Sci Technol*. 2016;53:13–22.
8. Guerinot ML, Yi Y. Iron: Nutritious, Noxious, and Not Readily Available. *Plant Physiol*. 1994;104:815–20.
9. Colangelo EP, Guerinot ML. The essential basic helix-loop-helix protein FIT1 is required for the iron deficiency response. *Plant Cell*. 2004;16:3400–12.
10. Koryachko A, Matthiadis A, Ducoste JJ, Tuck J, Long TA, Williams C. Computational approaches to identify regulators of plant stress response using high-throughput gene expression data. *Curr Plant Biol*. 2015;3–4:20–9.
11. Long TA, Tsukagoshi H, Busch W, Lahner B, Salt DE, Benfey PN. The bHLH Transcription Factor POPEYE Regulates Response to Iron Deficiency in *Arabidopsis* Roots. *Plant Cell*. 2010;22:2219–36.
12. Matthiadis A, Long TA. Further insight into BRUTUS domain composition and functionality. *Plant Signal Behav*. 2016;11:e1204508.
13. Selote D, Samira R, Matthiadis A, Gillikin JW, Long TA. Iron-Binding E3 Ligase Mediates Iron Response in Plants by Targeting Basic Helix-Loop-Helix Transcription Factors. *Plant Physiol*. 2015;167:273–86.
14. Yuan Y, Wu H, Wang N, Li J, Zhao W, Du J, et al. FIT interacts with AtbHLH38 and AtbHLH39 in regulating iron uptake gene expression for iron homeostasis in *Arabidopsis*. *Cell Res*. 2008;18:385–97.
15. Lingam S, Mohrbacher J, Brumbarova T, Potuschak T, Fink-Straube C, Blondet E, et al. Interaction between the bHLH Transcription Factor FIT and ETHYLENE INSENSITIVE3/ETHYLENE INSENSITIVE3-LIKE1 Reveals Molecular Linkage between the Regulation of Iron Acquisition and Ethylene Signaling in *Arabidopsis*. *Plant Cell*. 2011;23:1815–29.
16. Eide D, Broderius M, Fett J, Guerinot ML. A novel iron-regulated metal transporter from plants identified by functional expression in yeast. *Proc Natl Acad Sci*. 1996;93:5624–8.
17. Robinson NJ, Procter CM, Connolly EL, Guerinot ML. A ferric-chelate reductase for iron uptake from soils. *Nature*. 1999;397:694–7.
18. Römheld V, Müller C, Marschner H. Localization and Capacity of Proton Pumps in Roots of Intact Sunflower Plants 1. *Plant Physiol*. 1984;76:603–6.
19. Santi S, Schmidt W. Dissecting iron deficiency-induced proton extrusion in *Arabidopsis* roots. *New Phytol*. 2009;183:1072–84.
20. Vert G, Barberon M, Zelazny E, Séguéla M, Briat J-F, Curie C. *Arabidopsis* IRT2 cooperates with the high-affinity iron uptake system to maintain iron homeostasis in root epidermal cells. *Planta*. 2009;229:1171–9.
21. Curie C, Cassin G, Couch D, Divol F, Higuchi K, Le Jean M, et al. Metal movement within the plant: contribution of nicotianamine and yellow stripe 1-like transporters. *Ann Bot*. 2009;103:1–11.
22. Durrett TP, Gassmann W, Rogers EE. The FRD3-Mediated Efflux of Citrate into the Root Vasculature Is Necessary for Efficient Iron Translocation. *Plant Physiol*. 2007;144:197–205.
23. Buckhout TJ, Yang TJ, Schmidt W. Early iron-deficiency-induced transcriptional changes in *Arabidopsis* roots as revealed by microarray analyses. *BMC Genomics*. 2009;10:147.
24. Dinneny JR, Long TA, Wang JY, Jung JW, Mace D, Pointer S, et al. Cell Identity Mediates the Response of *Arabidopsis* Roots to Abiotic Stress. *Science*. 2008;320:942–5.
25. Vert G, Grotz N, Dédaléchamp F, Gaymard F, Guerinot ML, Briat J-F, et al. IRT1, an *Arabidopsis* Transporter Essential for Iron Uptake from the Soil and for Plant Growth. *Plant Cell*. 2002;14:1223–33.
26. Geng Y, Wu R, Wee CW, Xie F, Wei X, Chan PMY, et al. A Spatio-Temporal Understanding of Growth Regulation during the Salt Stress Response in *Arabidopsis*. *Plant Cell*. 2013;25:2132–54.
27. Iyer-Pascuzzi AS, Jackson T, Cui H, Petricka JJ, Busch W, Tsukagoshi H, et al. Cell Identity Regulators Link Development and Stress Responses in the *Arabidopsis* Root. *Dev Cell*. 2011;21:770–82.
28. Jakoby M, Schnittger A. Cell cycle and differentiation. *Curr Opin Plant Biol*. 2004;7:661–9.
29. Muhammad D, Clark NM, Haque S, Williams CM, Sozzani R, Long TA. POPEYE intercellular localization mediates cell-specific iron deficiency responses. *Plant Physiol*. 2022;kiac357.
30. Trofimov K, Ivanov R, Eutebach M, Acaroglu B, Mohr I, Bauer P, et al. Mobility and localization of the iron deficiency-induced transcription factor bHLH039 change in the presence of FIT. *Plant Direct*. 2019;3:e00190.
31. D'haeseleer P. How does gene expression clustering work? *Nat Biotechnol*. 2005;23:1499–501.
32. Oyelade J, Isewon I, Oladipupo F, Uwoghien E, Ameh F, et al. Clustering Algorithms: Their Application to Gene Expression Data. *Bioinforma Biol Insights*. 2016;10:BBI.S38316.
33. Quackenbush J. Computational Analysis of Microarray Data. *Nat Rev Genet*. 2001;2:418–27.
34. de Luis Balaguer MA, Fisher AP, Clark NM, Fernandez-Espinosa MG, Möller BK, Weijers D, et al. Predicting gene regulatory networks by combining spatial and temporal gene expression data in *Arabidopsis* root stem cells. *Proc Natl Acad Sci*. 2017;114:E7632–40.
35. Floris M, Mahgoub H, Lanet E, Robaglia C, Menand B. Post-transcriptional Regulation of Gene Expression in Plants during Abiotic Stress. *Int J Mol Sci*. 2009;10:3168–85.
36. Hernández-Elvira M, Sunnerhagen P. Post-transcriptional regulation during stress. *FEMS Yeast Res*. 2022;22:foac025.
37. Guerra D, Crosatti C, Khoshroo HH, Mastrangelo AM, Mica E, Mazzucotelli E. Post-transcriptional and post-translational regulations of drought and heat response in plants: a spider's web of mechanisms. *Front Plant Sci*. 2015;6(Article 57):1–14.
38. D'haeseleer P. What are DNA sequence motifs. *Nat Biotechnol*. 2006;24:423–5.
39. Aerts S. Computational Strategies for the Genome-Wide Identification of cis-Regulatory Elements and Transcriptional Targets. *Curr Top Dev Biol*. 2012;98:121–45.
40. Hashim FA, Mabrouk MS, Al-Atabany W. Review of Different Sequence Motif Finding Algorithms. *Avicenna J Med Biotechnol*. 2019;11:19.
41. Kocbek P, Fijacko N, Soguero-Ruiz C, Mikalsen KØ, Maver U, Povalej Brzan P, et al. Maximizing Interpretability and Cost-Effectiveness of Surgical Site Infection (SSI) Predictive Models Using Feature-Specific Regularized Logistic Regression on Preoperative Temporal Data. *Comput Math Methods Med*. 2019;2019:e2059851.
42. Kost S, Rheinbach O, Schaeben H. Using logistic regression model selection towards interpretable machine learning in mineral prospectivity modeling. *Geochemistry*. 2021;81:125826.
43. Langone R, Cuzzocrea A, Skantzos N. Interpretable Anomaly Prediction: Predicting anomalous behavior in industry 4.0 settings via regularized logistic regression tools. *Data Knowl Eng*. 2020;130:101850.
44. Quackenbush J. Microarray data normalization and transformation. *Nat Genet*. 2002;32:496–501.
45. Mantione KJ, Kream RM, Kuzelova H, Ptacek R, Raboch J, Samuel JM, et al. Comparing Bioinformatic Gene Expression Profiling Methods: Microarray and RNA-Seq. *Med Sci Monit Basic Res*. 2014;20:138–41.
46. Palmer CM, Hindt MN, Schmidt H, Clemens S, Guerinot ML. MYB10 and MYB72 Are Required for Growth under Iron-Limiting Conditions. *PLOS Genet*. 2013;9:e1003953.
47. Briat J-F, Ravet K, Arnaud N, Duc C, Boucherez J, Touraine B, et al. New insights into ferritin synthesis and function highlight a link between iron homeostasis and oxidative stress in plants. *Ann Bot*. 2010;105:811–22.
48. Kobayashi Y, Ohyama Y, Kobayashi Y, Ito H, Iuchi S, Fujita M, et al. STOP2 Activates Transcription of Several Genes for Al- and Low pH-Tolerance that Are Regulated by STOP1 in *Arabidopsis*. *Mol Plant*. 2014;7:311–22.
49. Tsai H-H, Schmidt W. pH-dependent transcriptional profile changes in iron-deficient *Arabidopsis* roots. *BMC Genomics*. 2020;21:694.
50. McDowell IC, Manandhar D, Vockley CM, Schmid AK, Reddy TE, Engelhardt BE. Clustering gene expression time series data using an infinite Gaussian process mixture model. *PLOS Comput Biol*. 2018;14:e1005896.
51. Koryachko A, Matthiadis A, Muhammad D, Foret J, Brady SM, Ducoste JJ, et al. Clustering and Differential Alignment Algorithm: Identification

- of Early Stage Regulators in the Arabidopsis thaliana Iron Deficiency Response. *PLoS ONE*. 2015;10:e0136591.
52. Kraskov A, Stögbauer H, Grassberger P. Estimating mutual information. *Phys Rev E*. 2004;69:066138.
 53. Qiu P, Gentles AJ, Plevritis SK. Fast calculation of pairwise mutual information for gene regulatory network reconstruction. *Comput Methods Programs Biomed*. 2009;94:177–80.
 54. Sivitz AB, Hermand V, Curie C, Vert G. Arabidopsis bHLH100 and bHLH101 Control Iron Homeostasis via a FIT-Independent Pathway. *PLoS ONE*. 2012;7:e44843.
 55. Lamesch P, Berardini TZ, Li D, Swarbreck D, Wilks C, Sasidharan R, et al. The Arabidopsis Information Resource (TAIR): improved gene annotation and new tools. *Nucleic Acids Res*. 2012;40:D1202–10.
 56. Sudre D, Gutierrez-Carbonell E, Lattanzio G, Rellán-Álvarez R, Gaymard F, Wohlgenuth G, et al. Iron-dependent modifications of the flower transcriptome, proteome, metabolome, and hormonal content in an Arabidopsis ferritin mutant. *J Exp Bot*. 2013;64:2665–88.
 57. Liu B, Wu J, Yang S, Schiefelbein J, Gan Y. Nitrate regulation of lateral root and root hair development in plants. *J Exp Bot*. 2020;71:4405–14.
 58. Schwarz B, Azodi CB, Shiu S-H, Bauer P. Putative cis-Regulatory Elements Predict Iron Deficiency Responses in Arabidopsis Roots. *Plant Physiol*. 2020;182:1420–39.
 59. Azodi CB, Lloyd JP, Shiu S-H. The cis-regulatory codes of response to combined heat and drought stress in Arabidopsis thaliana. *NAR Genomics Bioinforma*. 2020;2:iqaa049.
 60. Uygun S, Seddon AE, Azodi CB, Shiu S-H. Predictive Models of Spatial Transcriptional Response to High Salinity. *Plant Physiol*. 2017;174:450–64.
 61. Uygun S, Azodi CB, Shiu S-H. Cis-Regulatory Code for Predicting Plant Cell-Type Transcriptional Response to High Salinity. *Plant Physiol*. 2019;181:1739–51.
 62. Zou C, Sun K, Mackaluso JD, Seddon AE, Jin R, Thomashow MF, et al. Cis-regulatory code of stress-responsive transcription in Arabidopsis thaliana. *Proc Natl Acad Sci*. 2011;108:14992–7.
 63. Tibshirani R. Regression Shrinkage and Selection Via the Lasso. *J R Stat Soc Ser B Methodol*. 1996;58:267–88.
 64. Pampel FC. Interpreting logistic regression coefficients. *Logistic Regression*. 2000:19–40.
 65. McLeay RC, Bailey TL. Motif Enrichment Analysis: a unified framework and an evaluation on ChIP data. *BMC bioinformatics*. 2010;11(1):1–1.
 66. O'Malley RC, Huang SC, Song L, Lewsey MG, Bartlett A, Nery JR, et al. Cistrome and Epistrome Features Shape the Regulatory DNA Landscape. *Cell*. 2016;165:1280–92.
 67. Feng Y, Xu P, Li B, Li P, Wen X, An F, et al. Ethylene promotes root hair growth through coordinated EIN3/EIL1 and RHD6/RSL1 activity in Arabidopsis. *Proc Natl Acad Sci U S A*. 2017;114:13834–9.
 68. Wild M, Davière J-M, Regnault T, Sakvarelidze-Achard L, Carrera E, Lopez Diaz I, et al. Tissue-Specific Regulation of Gibberellin Signaling Fine-Tunes Arabidopsis Iron-Deficiency Responses. *Dev Cell*. 2016;37:190–200.
 69. Schwarz B, Bauer P. FIT, a regulatory hub for iron deficiency and stress signaling in roots, and FIT-dependent and -independent gene signatures. *J Exp Bot*. 2020;71:1694–705.
 70. Meiser J, Lingam S, Bauer P. Posttranslational Regulation of the Iron Deficiency Basic Helix-Loop-Helix Transcription Factor FIT Is Affected by Iron and Nitric Oxide. *Plant Physiol*. 2011;157:2154–66.
 71. Noguero M, Atif RM, Ochatt S, Thompson RD. The role of the DNA-binding One Zinc Finger (DOF) transcription factor family in plants. *Plant Sci*. 2013;209:32–45.
 72. Mitsuda N, Ohme-Takagi M. Functional Analysis of Transcription Factors in Arabidopsis. *Plant Cell Physiol*. 2009;50:1232–48.
 73. Huang X, Zhang Q, Jiang Y, Yang C, Wang Q, Li L. Shade-induced nuclear localization of PIF7 is regulated by phosphorylation and 14–3–3 proteins in Arabidopsis. *eLife*. 2018;7:e31636.
 74. Hernando CE, Romanowski A, Yanovsky MJ. Transcriptional and post-transcriptional control of the plant circadian gene regulatory network. *Biochim Biophys Acta BBA - Gene Regul Mech*. 2017;1860:84–94.
 75. Kojima S, Shingle DL, Green CB. Post-transcriptional control of circadian rhythms. *J Cell Sci*. 2011;124:311–20.
 76. Mateos JL, de Leone MJ, Torchio J, Reichel M, Staiger D. Beyond Transcription: Fine-Tuning of Circadian Timekeeping by Post-Transcriptional Regulation. *Genes*. 2018;9:E616.
 77. Staiger D, Köster T. Spotlight on post-transcriptional control in the circadian system. *Cell Mol Life Sci*. 2011;68:71–83.
 78. Robles P, Quesada V. Transcriptional and Post-transcriptional Regulation of Organellar Gene Expression (OGE) and Its Roles in Plant Salt Tolerance. *Int J Mol Sci*. 2019;20:1056.
 79. Barah P, B N MN, Jayavelu ND, Sowdhamini R, Shameer K, Bones AM. Transcriptional regulatory networks in Arabidopsis thaliana during single and combined stresses. *Nucleic Acids Res*. 2016;44:3147–64.
 80. Bruex A, Kainkaryam RM, Wieckowski Y, Kang YH, Bernhardt C, Xia Y, et al. A Gene Regulatory Network for Root Epidermis Cell Differentiation in Arabidopsis. *PLOS Genet*. 2012;8:e1002446.
 81. Derbyshire P, Drea S, Shaw PJ, Doonan JH, Dolan L. Proximal–distal patterns of transcription factor gene expression during Arabidopsis root development. *J Exp Bot*. 2008;59:235–45.
 82. Han G, Wei X, Dong X, Wang C, Sui N, Guo J, et al. Arabidopsis ZINC FINGER PROTEIN1 Acts Downstream of GL2 to Repress Root Hair Initiation and Elongation by Directly Suppressing bHLH Genes. *Plant Cell*. 2020;32:206–25.
 83. He L, Wang Y, Wei Z. ATDOF5.8 protein is the upstream regulator of ANAC069 and is responsive to abiotic stress. *Biochimie*. 2015;110:17–24.
 84. Hsieh W-P, Hsieh H-L, Wu S-H. Arabidopsis bZIP16 Transcription Factor Integrates Light and Hormone Signaling Pathways to Regulate Early Seedling Development. *Plant Cell*. 2012;24:3997–4011.
 85. Jin Y. Functional Analysis of INDETERMINATE DOMAIN 1 and 2 in Gibberellin Signaling in Arabidopsis thaliana. Doctor of Philosophy: Duke University; 2015.
 86. Kreps JA, Wu Y, Chang H-S, Zhu T, Wang X, Harper JF. Transcriptome Changes for Arabidopsis in Response to Salt, Osmotic, and Cold Stress. *Plant Physiol*. 2002;130:2129–41.
 87. Lin Q, Ohashi Y, Kato M, Tsuge T, Gu H, Qu LJ, Aoyama T. GLABRA2 directly suppresses basic helix-loop-helix transcription factor genes with diverse functions in root hair development. *Plant Cell*. 2015;27(10):2894–906.
 88. Maurer F, Naranjo Arcos MA, Bauer P. Responses of a Triple Mutant Defective in Three Iron Deficiency-Induced BASIC HELIX-LOOP-HELIX Genes of the Subgroup Ib(2) to Iron Deficiency and Salicylic Acid. *PLoS ONE*. 2014;9:e99234.
 89. Paik I, Kathare PK, Kim J-I, Huq E. Expanding Roles of PIFs in Signal Integration from Multiple Processes. *Mol Plant*. 2017;10:1035–46.
 90. Roeber VM, Bajaj I, Rohde M, Schmölling T, Cortleven A. Light acts as a stressor and influences abiotic and biotic stress responses in plants. *Plant Cell Environ*. 2021;44:645–64.
 91. Waqas M, Shahid L, Shoukat K, Aslam U, Azeem F, Atif RM. Role of DNA-binding with one finger (Dof) transcription factors for abiotic stress tolerance in plants. In: *Transcription factors for abiotic stress tolerance in plants 2020 Jan 1* (pp. 1–14). Academic Press.
 92. Wei Yang TJ, Perry PJ, Ciani S, Pandian S, Schmidt W. Manganese deficiency alters the patterning and development of root hairs in Arabidopsis. *J Exp Bot*. 2008;59:3453–64.
 93. Liu Y, Jafari F, Wang H. Integration of light and hormone signaling pathways in the regulation of plant shade avoidance syndrome. *ABIOTECH*. 2021;2:131–45.
 94. Zhang H, Ding Y. Novel insights into the pervasive role of RNA structure in post-transcriptional regulation of gene expression in plants. *Biochem Soc Trans*. 2021;49:1829–39.
 95. Schmidt W, Tittel J, Schikora A. Role of Hormones in the Induction of Iron Deficiency Responses in Arabidopsis Roots. *Plant Physiol*. 2000;122:1109–18.
 96. Vissenberg K, Claeijs N, Balcerowicz D, Schoenaers S. Hormonal regulation of root hair growth and responses to the environment in Arabidopsis. *J Exp Bot*. 2020;71:2412–27.
 97. Mallory AC, Dugas DV, Bartel DP, Bartel B. MicroRNA Regulation of NAC-Domain Targets Is Required for Proper Formation and Separation of Adjacent Embryonic, Vegetative, and Floral Organs. *Curr Biol*. 2004;14:1035–46.
 98. Martín-Barranco A, Spielmann J, Dubeaux G, Vert G, Zelazny E. Dynamic Control of the High-Affinity Iron Uptake Complex in Root Epidermal Cells. *Plant Physiol*. 2020;184:1236–50.

99. Lee MM, Schiefelbein J. WEREWOLF, a MYB-Related Protein in *Arabidopsis*, Is a Position-Dependent Regulator of Epidermal Cell Patterning. *Cell*. 1999;99:473–83.
100. Birnbaum K, Shasha DE, Wang JY, Jung JW, Lambert GM, Galbraith DW, et al. A Gene Expression Map of the *Arabidopsis* Root. *Science*. 2003;302:1956–60.
101. Andrews S. FASTQC. A quality control tool for high throughput sequence data. <http://www.bioinformatics.babraham.ac.uk/projects/fastqc/>.
102. Aronesty E. ea-utils: Command-line tools for processing biological sequencing data. ea-utils: FASTQ processing utilities. 2011. <http://code.google.com/p/ea-utils>.
103. Kim D, Pertea G, Trapnell C, Pimentel H, Kelley R, Salzberg SL. TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. *Genome Biol*. 2013;14:R36.
104. Trapnell C, Pachter L, Salzberg SL. TopHat: discovering splice junctions with RNA-Seq. *Bioinformatics*. 2009;25:1105–11.
105. Trapnell C, Roberts A, Goff L, Pertea G, Kim D, Kelley DR, et al. Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. *Nat Protoc*. 2012;7:562–78.
106. Liao Y, Smyth GK, Shi W. The R package Rsubread is easier, faster, cheaper and better for alignment and quantification of RNA sequencing reads. *Nucleic Acids Res*. 2019;47:e47.
107. Robinson MD, McCarthy DJ, Smyth GK. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics*. 2010;26:139–40.
108. Benjamini Y, Hochberg Y. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *J R Stat Soc Ser B Methodol*. 1995;57:289–300.
109. Stein M. Large Sample Properties of Simulations Using Latin Hypercube Sampling. *Technometrics*. 1987;29:143–51.
110. Alexa A, Rahnenführer J. Gene set enrichment analysis with topGO. *Bioconductor Improv*. 2009;27:1–26.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Ready to submit your research? Choose BMC and benefit from:

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

At BMC, research is always in progress.

Learn more biomedcentral.com/submissions

