Putative cis-Regulatory Elements Predict Iron Deficiency Responses in Arabidopsis Roots

Birole Schwarz, a Christina B. Azodi, b,c Shin-Han Shiu, b,c,d, and Petra Bauer a,e,2,3

a Institute of Botany, Heinrich Heine University, Düsseldorf 40225 Germany
b Department of Plant Biology, Michigan State University, East Lansing, Michigan 48824
c DOE-Great Lake Bioenergy Research Center, Michigan State University, East Lansing, Michigan 48824
d Department of Computational, Mathematics, Science, and Engineering, Michigan State University, East Lansing, Michigan 48824

Cluster of Excellence on Plant Science (CEPLAS), Heinrich Heine University, Düsseldorf 40225 Germany

Putative cis-Regulatory Elements Predict Iron Deficiency Responses in Arabidopsis Roots

Plant iron deficiency (−Fe) activates a complex regulatory network that coordinates root Fe uptake and distribution to sink tissues. In Arabidopsis (Arabidopsis thaliana), FER-LIKE FE DEFICIENCY–INDUCED TRANSCRIPTION FACTOR (FIT), a basic helix-loop-helix (bHLH) transcription factor (TF), regulates root Fe acquisition genes. Many other −Fe-induced genes are FIT independent, and instead regulated by other bHLH TFs and by yet unknown TFs. The cis-regulatory code, that is, the cis-regulatory elements (CREs) and their combinations that regulate plant −Fe-responses, remains largely elusive. Using Arabidopsis root transcriptome data and coexpression clustering, we identified over 100 putative CREs (pCREs) that predicted −Fe-induced gene expression in computational models. To assess pCRE properties and possible functions, we used large-scale in vitro TF binding data, positional bias, and evolutionary conservation. As one example, our approach uncovered pCREs resembling IDE1 (iron deficiency-responsive element 1), a known grass −Fe response CRE. Arabidopsis IDE1-likes were associated with FIT-dependent gene expression, more specifically with biosynthesis of Fe-chelating compounds. Thus, IDE1 seems to be conserved in grass and nongrass species. Our pCREs matched among others in vitro binding sites of B3, NAC, bZIP, and TCP TFs, which might be regulators of −Fe responses. Altogether, our findings provide a comprehensive source of cis-regulatory information for −Fe-responsive genes that advance our mechanistic understanding and inform future efforts in engineering plants with more efficient Fe uptake or transport systems.

The micronutrient iron (Fe) is crucial for survival of all organisms. Plants encounter Fe deficiency (−Fe) on calcareous and alkaline soils or during developmental phases with increased sink demands. As a central component of heme and Fe-sulfur (FeS) clusters, Fe acts in redox processes in plants in basically all important metabolic processes, such as the respiratory and photosynthetic electron transport chains, chlorophyll biosynthesis, DNA replication and repair, and nitrogen and sulfur assimilation. Consequently, plants react to −Fe with a range of molecular, physiological, and morphological adjustments, which is reflected in transcriptional alterations of more than 1000 genes in Arabidopsis (Arabidopsis thaliana; Dinneny et al., 2008; Rodriguez-Celma et al., 2013; Mai et al., 2016). In the shoots, the photosynthetic machinery is remodeled, leading to visible leaf chlorosis symptoms, and essential Fe-requiring processes are prioritized, which can be achieved through break-down of dispensable Fe-bound proteins and Fe redistribution between organelles (Blaby-Haas and Merchant, 2013; Balk and Schaedler, 2014; Hantzis et al., 2018). In the roots, genes controlling soil Fe uptake and detoxification of other transition metal ions acquired along with Fe are up-regulated. Additionally, Fe is mobilized from internal storages and distributed to Fe sinks. −Fe also leads to changes in root architecture and root hair morphology (Brumbarova et al., 2015; Curie and Mari, 2017; Jeong et al., 2017; Li and Lan, 2017).

To acquire soil Fe, grasses secrete mugineic acid (MA) family phytosiderophores and import Fe^{3+}-MA...
complexes into the root ("Strategy II"). In contrast, non-grass monocots and dicots, such as Arabidopsis, acquire Fe via a reduction-based mechanism, in which soil Fe\(^{3+}\) is solubilized by lowering the local pH through proton extrusion, followed by reduction to plant-accessible Fe\(^{2+}\) at the root epidermis and Fe\(^{2+}\) uptake ("Strategy I"; Marschner and Römheld, 1994). In Strategy I, secreted chelators (mainly phenylpropanoid-derived coumarins or riboflavin derivatives) aid efficient Fe\(^{3+}\) solubilization and reduction (Fourcroy et al., 2014; Schmid et al., 2014). Thus, Fe chelation is important during acquisition in both strategies.

Transcriptional control plays an important role in –Fe responses. A regulatory cascade ultimately controls a set of –Fe response genes. In both Strategy I and II, the current cascade model involves related subgroups of basic helix-loop-helix (bHLH) transcription factors (TFs). When Arabidopsis plants experience –Fe, subgroup IVc bHLH proteins, IAA-LEU RESISTANT3 (ILR3), bHLH34, bHLH104, and bHLH115, together with the subgroup IVb bHLH121, activate subgroup Ib bHLH genes and subgroup IVb BHLH POPEYE (PYE; Zhang et al., 2015; Li et al., 2016; Liang et al., 2017; Kim et al., 2019; Tissot et al., 2019; Gao et al., 2020). In addition, IVc bHLH and possibly bHLH121 protein levels are controlled by Fe-responsive genes. A similar function is exerted by subgroup IVc bHLH proteins POSITIVE REGULATOR OF IRON HOMEOSTASIS1 (PR11), PR12, PR13, and PR14 in rice (Oryza sativa; Zhang et al., 2017; Kobayashi et al., 2019; Zhang et al., 2020). Subgroup Ib bHLH proteins, bHLH38, bHLH39, bHLH100, and bHLH101, and PYE in Arabidopsis, as well as orthologs IRON-RELATED TRANSCRIPTION FACTOR2 (IRO2) and IRO3 in rice, regulate downstream responses (Ogo et al., 2007; Yuan et al., 2008; Long et al., 2010; Zheng et al., 2010). In addition, IVc bHLH and possibly bHLH121 protein levels are controlled by Fe-regulated E3 ligases (Selote et al., 2015; Zhang et al., 2017; Kim et al., 2019; Kobayashi et al., 2019).

Strategy I Fe acquisition requires the bHLH TF FER-LIKE IRON DEFICIENCY-INDUCED TRANSCRIPTION FACTOR (FIT; Colangelo and Guerinot, 2004; Jakoby et al., 2004), and is activated upon –Fe mainly through interaction of FIT with subgroup Ib TFs (Yuan et al., 2008; Sivitz et al., 2012; Wang et al., 2013). FIT is essential for up-regulation of Fe\(^{3+}\) reduction, Fe\(^{2+}\) uptake, and coumarin biosynthesis and export genes (Colangelo and Guerinot, 2004; Jakoby et al., 2004; Sivitz et al., 2012; Schmid et al., 2014; Mai et al., 2016). Similarly, FIT ortholog bHLH156 from rice activates Strategy II in conjunction with IRO2 (Wang et al., 2019).

Despite these conserved regulatory and functional interactions of bHLH proteins between Arabidopsis and rice, it remains unclear if other regulatory components between Strategy I and II are conserved. For example, in grasses, IRON DEFICIENCY-RESPONSIVE ELEMENT BINDING FACTOR1 (IDE1), ABI3VP1 subfamily of B3 TF and IDE2 (NAC TF) coordinate –Fe responses through binding to IDE1 (iron deficiency-responsive element 1) and IDE2 (Kobayashi et al., 2007; Ogo et al., 2008). IDE1 has been connected to induction of genes involved in Strategy II MA biosynthesis and Fe-MA uptake (Kobayashi et al., 2005; Ogo et al., 2007). However, whereas barley (Hordeum vulgare) IDE1 can drive reporter gene expression in tobacco (Nicotiana tabacum) in a –Fe-dependent manner and IDE1-like motifs are present in several Arabidopsis –Fe response genes, a function for IDE1 has not been shown in Strategy I plants (Kobayashi et al., 2003; Kobayashi et al., 2005; Kobayashi et al., 2007; Murgia et al., 2011).

A coexpression network built with –Fe-responsive genes has provided insight into the –Fe regulatory system in Arabidopsis (Ivanov et al., 2012; Schwarz and Bauer, 2020). Among coexpression clusters, one contains root-specific and FIT-dependent genes involved in Fe acquisition, whereas another one is composed of root- and shoot-expressed FIT-independent genes, regulated by IVc and IVb bHLH proteins. In this work, we refer to robust (i.e. consistently identified in different studies) FIT-dependent and FIT-independent genes as the gold standard (GS; i.e. benchmark) –Fe-induced genes. The concept to discriminate FIT-dependent and FIT-independent coexpression clusters is informative for interpreting mutant phenotypes and for placing regulators into the –Fe response cascade (e.g. Zhang et al., 2015; Liang et al., 2017; Gratz et al., 2019; Kim et al., 2019; Schwarz and Bauer, 2020). FIT-independent network genes mostly act in subcellular and long-distance transport and distribution of Fe and in Fe signaling, and they include subgroup Ib BHLH genes and PYE (Ivanov et al., 2012; Schwarz and Bauer, 2020). The latter, PYE, controls NICOTIANAMINE SYNTHASE4 (NAS4), ZINC-INDUCED FACILITATOR1 (ZIF1), and FERRIC REDUCTION OXIDASE3 (FRO3) of the same coexpression regulon (Long et al., 2010).

For most –Fe-responsive genes, including reliable marker genes (Ivanov et al., 2012; Mai et al., 2016), the specific cis-regulatory elements (CREs) that coordinate their expression are unknown. Computational approaches uncover regulatory connections on a genome-wide scale, such as through elucidating the cis-regulatory code, i.e. the collection of CREs and the genes they regulate in a given regulatory context (Yáñez-Cuna et al., 2013). Putative CREs (pCREs) could be identified computationally by the overrepresentation of sequences in the promoter regions of coregulated genes. Combining with data for TF binding motifs (TFBMs) in Arabidopsis (Weirauch et al., 2014; O’Malley et al., 2016), regulatory connections can be made between TFs, binding sequences, and target genes. To further improve the confidence of computationally derived cis-regulatory code, machine learning algorithms (for review, see Ma et al., 2014) can be applied to build models with pCREs to predict gene expression levels or transcriptional responses. These models can be evaluated by how well they can predict expression levels or responses of genes that are not used for model training. Most importantly, a good model indicates that the pCREs used are most likely important for regulating the expression/response of interest. Previous work has demonstrated the suitability of machine
learning for elucidating the cis-regulatory code of environmental stress responses in Arabidopsis (Zou et al., 2011; Uygun et al., 2019).

To get a deeper understanding of Fe response regulation, we elucidate its underlying cis-regulatory code using a combination of coexpression, motif finding, and machine learning approaches. Because some TFs have well-established roles in Fe response, we use these TFs to validate our findings. In addition, we combined genome, transcriptome, and in vitro protein-DNA interaction data to uncover links between pCREs controlling Fe responses and their upstream TFs.

With pCREs overrepresented in promoters of coexpressed genes we modeled Fe-induced up-regulation and identified over 100 informative pCREs of Fe-responsive processes.

RESULTS AND DISCUSSION

Overview of Approach and Functions of Fe-Responsive Genes

To identify root Fe-associated CREs at a genomewide scale, we defined root Fe-response coexpression clusters; then we identified k-mers (sequences with k number of nucleotides) enriched in the promoter regions of those genes, and finally we modeled Fe response on the basis of the enriched promoter k-mers. An overview of our complete workflow including functional analysis of the identified pCREs is shown in Figure 1A. Because many factors impact the discovery of functional connections between genes, such as the choice of data set or the measure used to define expression similarity (Uygun et al., 2016), we used multiple expression data combinations and algorithms with varying parameters (see “Materials and Methods”).

Fe-responsive genes (log2 fold-change [log2FC] ≥ 1 or ≤ −1, q < 0.05) were identified using transcriptomic data available for six time points after a Fe treatment in Arabidopsis seedling roots (Dinneny et al., 2008). Enrichment analysis (Fisher’s exact test, or FET; q < 0.05) of biological process gene ontologies (GOs) showed that, next to Fe-related GOs (e.g., Fe transport, homeostasis, and Fe-Sulfur [FeS] cluster assembly), responses to several hormones were overrepresented, including auxin, ethylene, abscisic acid (ABA), and jasmonic acid (Fig. 1B). This is consistent with the roles of hormones in Fe response (Brumbarova et al., 2015) and in root and root hair morphology and development (Schmidt et al., 2000), which were also enriched GOs. Fe affects the photosynthetic machinery and often correlates with oxidative stress responses (Rodríguez-Celma et al., 2013), which is reflected in enrichment of GOs regarding oxidative stress, photosynthesis, and primary metabolism even in roots (Fig. 1B; Supplemental Fig. S1).
Using Multiple Expression Data Sets to Define –Fe Coexpression Clusters

We next grouped differentially regulated –Fe response genes into coexpression clusters using two approaches: k-means clustering and correlation to robust GS –Fe response genes. K-means clustering was based on the transcriptional responses to –Fe alone and on –Fe responses combined with different responses to other stress and developmental conditions (Fig. 1C; Schmid et al., 2005; Kilian et al., 2007; Dinneny et al., 2008; Goda et al., 2008). Correlation-based clusters were generated for each gene in our curated list of 28 GS –Fe response genes (see “Materials and Methods”; Supplemental Table S1). These clusters contain the query GS gene, and all –Fe response genes (GS and non-GS) with a significantly similar expression pattern to the GS gene (Pearson’s Correlation Coefficient, or PCC; see “Materials and Methods”). Again, the different combinations of transcriptional data were used (Fig. 1C).

The main reason for using both k-means and correlation-based approaches was to maximize chances of detecting clusters of coregulated genes. Our earlier studies demonstrated that the clustering approach used impacts how well we can recover genes with similar functions from expression data and that different clustering approaches complement each other (Uygun et al., 2016). The k-means algorithm is commonly used and generally provides good results; however, k-means clusters are impacted by the arbitrarily chosen value of k. We took this into account by testing different k values (see “Materials and Methods”); yet coexpression associations, most importantly with GS genes, might have been missed. Thus, in addition to using k-means, we applied the widely used coexpression measure PCC. For this correlation-based approach, GS genes were used as queries to establish GS-centric clusters that regulate central –Fe response mechanisms. Genes that did not correlate with GS genes were not considered in this PCC-based approach but would be captured by the k-means approach. Therefore, these two approaches are complementary and enabled us later to compare regulatory elements of cluster types with and without GS genes.

In total, 1,520 k-means and 439 PCC-based clusters were defined. To characterize them, we determined enriched GOs. We grouped clusters with similar biological functions according to their enriched GOs (FET, q < 0.05) into “superclusters” (Fig. 2A; Supplemental Fig. S2A; see “Materials and Methods”). Superclusters were defined as groups of at least 20 clusters with significantly higher similarity to each other than the average of all clusters (all Mann-Whitney U, P < 2.2e-16; Supplemental Fig. S2, B and C). We found that not all k-means clusters have directly Fe-related functions. For example, while k-means supercluster C was enriched in the Fe-related GO “cellular response to Fe” (GO shared by ≥75% of coexpression clusters within the supercluster), k-means superclusters A and B shared GOs related to different stress responses.

Because the current The Arabidopsis Information Resource (TAIR; http://www.arabidopsis.org) GO annotation for –Fe response-related processes does not contain all –Fe-responsive genes of interest (missing, for example, the GS genes MYB10, UDP-GLUCOSYL TRANSFERASE 72E1 [UGT72E1], AT3G07720, FE UPTAKE-INDUCING PEPTIDE3 [FEP3] or NAS4; Ivanov et al., 2012), we also determined if k-means coexpression clusters were enriched with GS genes (“GS-enriched”; FET, q < 0.05; Fig. 2A, right). While many of these clusters were part of the Fe-related GO supercluster A, the GS approach allowed us to identify an additional 23 Fe-related coexpression clusters that would have been overlooked by conventional GO enrichment analysis. In total, 7% of the k-means clusters were GS enriched (Fig. 2B, left). Applying this same analysis to the PCC-based clusters (Fig. 2A; Supplemental Fig. S2A), we found higher levels of similarity between PCC-based clusters compared with k-means clusters (Mann-Whitney U, P < 2.2e-16; Supplemental Fig. S2D), because we precondition their identification on GS genes, some of which are tightly coregulated (Ivanov et al., 2012). Accordingly, we found that 93% were GS enriched (Fig. 2B, right).

GS genes are either regulated by FIT (“FIT-dependent”, typically Fe acquisition genes) or by other factors (“FIT-independent” genes, typically involved in Fe distribution). We found this reflected in our GS-enriched clusters: 71% of the GS-enriched clusters were specifically enriched with either FIT-dependent (39%) or FIT-independent genes (32%; FET, q < 0.05; Fig. 2B, bottom). The remaining GS-enriched clusters were enriched with both (“mixed”). Interestingly, clusters based on combined expression data (i.e. data combinations [DC] 2, 3, 5a/b, 6) were more often enriched with FIT-dependent or FIT-independent genes, whereas –Fe time course data alone (DC 1) produced mainly mixed category clusters (Fig. 2C). The utility of including spatial or developmental data (DC 2, 6) to define coexpression clusters reflects that –Fe-responsive genes act at different time points and in different tissues and organs (Dinneny et al., 2008; Ivanov et al., 2012; Jeong et al., 2017). Finally, genes in clusters not enriched with GS genes (“nonenriched”) tended to respond to particular abiotic stresses, for example cold (Supplemental Fig. S3 e.g. clusters 937, 973) or salt (e.g. clusters 915, 936, 987), whereas expression of GS-enriched cluster genes tended to randomly oscillate under different abiotic stresses (e.g. clusters 818, 835, 858, 889). This might indicate different regulatory networks and highlight the usefulness of incorporating additional abiotic stress data (as in DC 3, 5, 6) when defining coexpression clusters that are likely coregulated.

In summary, by using different expression data sets and clustering methods, we defined 1,959 –Fe coexpression clusters, many of which were enriched with FIT-dependent and/or FIT-independent GS genes. These represent possibly coregulated functional units in
Fe acquisition and Fe homeostasis processes, well-suited to identify pCREs, which can explain Fe-induced up-regulation. Genes in coexpression clusters that were enriched with Fe-responsive genes but not GS genes (nonenriched clusters) are presumably regulated by mechanisms different from GS-enriched clusters.

A Machine Learning Approach to Model Regulation of Fe-Responsive Coexpression Clusters

The machine learning algorithm Random Forest (RF) has been successfully used to model stress transcriptional responses using cis-regulatory sequences in plants (Deng et al., 2017; Uygun et al., 2017; Zou et al., 2011). Here, for each coexpression cluster, we used pCREs (enriched k-mers in putative promoter sequences) to build a RF model that classifies genes as belonging to the cluster in question or as a nonresponsive gene (see “Materials and Methods”). Here the model performance is measured using the F1 score defined as the harmonic mean of precision and recall, where a perfect model would have a F1 of 1 and, because we used balanced data sets (50% responsive and 50% nonresponsive) in model training, the F1 for a model performing no better than random guessing would have an F1 of 0.5. The pCREs from models performing above a defined threshold (moderate model performance, F1 ≥ 0.7) were considered further. Out of 1,959 coexpression clusters, 28% of the models passed the performance threshold, 60% performed poorly, and for 12% no model could be built due to small sample size (median size = 2 genes; Supplemental Fig. S4, A and B). Poorly performing models (median F1 = 0.62) were mostly for small clusters (median size = 12; Supplemental Fig. S4B; Supplemental Table S2), likely due to the lack of training data. Nonetheless, 66 large clusters (>100 genes, median size = 135) also performed poorly (median F1 = 0.64). This is likely because these large clusters are too heterogeneous containing genes with multiple regulatory codes (Uygun et al., 2016; Uygun et al., 2017), and/or are controllable but not at the transcriptional level (e.g. posttranslationally controlled). Interestingly, of the 28% of clusters with models above the threshold, only 36% were GS-enriched clusters (Supplemental Fig. S4A). Nonetheless, models built for GS-enriched clusters (median F1 = 0.68) tended to perform better than models built for nonenriched clusters (median F1 = 0.65; Mann-Whitney U, P < 2.358e-09; Fig. 3, A–C). The 159 GS-enriched (27 k-mers, 132 PCC based) and 358 nonenriched (349 k-mers, 9 PCC based) cluster models (total n = 517) performed above the threshold and were considered further.

Good model performance indicates that genes in a cluster are more likely coregulated, and, because pCREs were used to build the model, these pCREs are likely the regulatory sequences contributing to the coregulation. Each pCRE had an importance score,
reflecting how useful a pCRE was for predicting −Fe response genes in a cluster (see "Materials and Methods"). Because the models were trained on presence/absence of pCREs, the rate of a gene being correctly predicted (true positive; TP) is positively correlated with the total number of pCREs present in its promoter (Spearman’s rank correlation coefficient, calculated for 159 GS-enriched clusters; median = 0.8; Supplemental Fig. S4D). However, not all pCREs were necessary for overall good model performance, mostly due to pCRE redundancy. The minimum set of pCREs (termed min-pCREs), which are sufficient to predict a cluster model, will be addressed in a later section. Figure 3, D–F, demonstrates the principle of RF models for coexpression clusters with good (Fig. 3, D and E) and poor (Fig. 3F) performance. Roughly speaking, the more genes are correctly predicted as −Fe responsive, the better the overall model performance.

Across 517 well-performing cluster models, we identified 5,639 pCREs enriched in promoters of −Fe-responsive genes that may be predictive of −Fe-induced up- or down-regulation. The pCREs of each cluster model were specific for the coexpressed genes in these clusters. As control, we applied the models to all −Fe response genes in general. As expected, the specific models could not generally predict any −Fe response genes (Supplemental Fig. 54E; Supplemental Table S3). Instead, pCREs were even specific for models of −Fe-responsive expression at certain time points or in different root zones. For example, expression of cluster 44 (expression starts 24 h upon −Fe treatment) is predicted by a different set of pCREs than cluster 120 (expression starts at 12 h). Similarly, cluster 532 (genes expressed in root zone 4) and cluster 552 (genes expressed in root zones 3 and 4) are predicted by different sets of pCREs (Supplemental Fig. S5).

To further evaluate the biological relevance of pCREs, in the following sections, we assess pCREs based on their association with GS-enriched or non-enriched clusters, importance for model performance, and similarity to known TF binding sites.
grass, for example E-/G-boxes (bHLH TF binding sites) and IDE1, will serve as positive controls.

Identifying Common pCREs Across Coexpression Clusters

We expect true −Fe response CREs to be (1) important for building models with good performance in predicting −Fe response and (2) reliably identified in coexpression clusters with similar gene content. For an overview of the 5,639 pCREs, we calculated for each the proportion of clusters enriched with the pCRE and its average importance rank across those clusters (Supplemental Table S4). The importance rank of a pCRE was derived from its importance score in the trained RF models.

pCREs were identified in between 1 (0.6%) and 56 (35%) GS-enriched clusters and in between 1 (0.3%) and 54 (15%) nonenriched clusters. Also, 173 pCREs were considered “frequent pCREs” (fq-pCREs), because they were identified in >5% of GS-enriched or nonenriched clusters (Supplemental Table S4). Across GS-enriched clusters, pCREs tended to occur at higher proportions and with higher importance ranks than across nonenriched clusters (Mann-Whitney U, P < 2.2e-16; Fig. 4, A and B; Supplemental Fig. S6, A and B). The higher proportion of GS-enriched cluster pCREs can be explained partly by the fact that those clusters are more homogenous in terms of gene contents than nonenriched clusters (Mann-Whitney U, P < 2.2e-16; Supplemental Fig. S4C).

We next determined whether GS-enriched clusters are regulated by a different set of pCREs than nonenriched clusters. Out of the 5,639 pCREs, 15% (n = 860) were unique to GS-enriched clusters, whereas 73% (n = 4,109) were unique to nonenriched clusters. Twelve percent (n = 670) were found in both GS-enriched and nonenriched clusters (Fig. 4A, inset). This indicates that GS-enriched clusters and nonenriched clusters are regulated partly by different pCREs, but also by a fraction of shared pCREs. However, 43% (n = 286) of the 670 shared pCREs were predominant to GS-enriched clusters (i.e. having only low proportion and low importance rank in nonenriched clusters; Supplemental Fig. S6C, top). This indicates that pCREs that were categorized as shared might not be equally important for regulating both GS-enriched and nonenriched clusters. Interestingly, unique GS-enriched freq-pCREs represented 59% (n = 102) of the 173 freq-pCREs, whereas 35.5% (n = 58) were unique nonenriched, and only 7.5% (n = 13) freq-pCREs were shared between GS-enriched and nonenriched clusters, indicating that pCREs with high proportion are also the ones that seem to regulate almost exclusively the GS-enriched or nonenriched cluster functions, but not both (Fig. 4A, inset; Supplemental Fig. S6C, bottom). Furthermore, freq-pCREs tended to have higher importance ranks than nonfrequent pCREs (Mann-Whitney U, P < 1.924e-14; Supplemental Fig. S6D). Together, this suggests that freq-pCREs could be particularly relevant for regulation of −Fe response mechanisms.

To characterize the freq-pCREs, we grouped them according to sequence similarity using pairwise PCC distances of pCRE position weight matrices (PWM; see “Materials and Methods”). Sixty-two percent (n = 107) of all freq-pCREs could be placed into one of eight pCRE groups (Fig. 4, C and D; Supplemental Fig. S6E). Freq-pCREs of the same group tended to be predictors of the same cluster category (GS-enriched/nonenriched).

In summary, we identified more than 100 −Fe pCREs that were reliably associated either exclusively with GS-enriched or nonenriched coexpression clusters or with high preference for one of the categories. Those pCREs were also ranked as important for machine learning models and might therefore be candidates for functionally relevant motifs to different responses to −Fe.

Similarity of −Fe pCREs to Known TFBMs

CREs are recognized by TFs to modulate gene expression. To identify what types of TFs may bind to the identified pCREs, we compared them with known TFBMs from two sources of Arabidopsis TF large-scale in vitro binding data: DNA affinity purification sequencing (DAP-seq; O’Malley et al., 2016) and Catalog of Inferred Sequence Binding Preferences (CIS-BP; Weiраuch et al., 2014; see “Materials and Methods”).

Based on threshold similarities that differentiate between TFBMs and random sequences (Uygun et al., 2017), we were able to match a specific TF and/or a specific TF family to each of the 173 freq-pCREs (see “Materials and Methods”; Fig. 5A; Supplemental Fig. S7). This indicates that these freq-pCREs are not randomly occurring promoter sequences. To gain an overview that TF families could be associated with GS-enriched clusters, we calculated the proportions of TF family TFBMs that match freq-pCREs from GS-enriched and nonenriched clusters (Fig. 5B). We found most TF families with higher proportion in either GS-enriched clusters (14 TF families) or nonenriched clusters (12), whereas only four were equally distributed between both categories.

Most known −Fe regulators in Arabidopsis are bHLH TFs (e.g. FIT, subgroup Ib, Ivb, and Ivc bHLH proteins; Jakoby et al., 2004; Wang et al., 2007; Long et al., 2010; Zhang et al., 2015; Gao et al., 2020). Two more important TFs are MYB10 and MYB72 (Palmer et al., 2013). Consistently, we identified bHLH and MYB TF families associated with our clusters, and even with higher proportion in GS-enriched clusters than in nonenriched clusters. Other matching TF families overrepresented in GS-enriched clusters were bZIP (FET, q < 0.05), B3, TCP, and NAC. Although a B3 TF (IDF1; ABI3VP1 subfamily of B3) and a NAC TF (IDF2) are important regulators of Strategy II Fe acquisition in grasses (Kobayashi et al., 2007; Ogo et al., 2008), the role of these TF families in Strategy I nongrass
plant species has not yet been described. In contrast, ARID, WRKY (both FET, q < 0.05), Homebox, and CAMTA TF families were matched more in non-enriched clusters. That these TF families were not primarily associated with GS genes might point toward them regulating Fe responses, which are not directly Fe uptake or distribution. Alternatively, because several nonenriched clusters contain down-regulated Fe-responsive genes, some of these TFs might be repressors.

Next, freq-pCRE-TFBM matches from GS-enriched clusters served to infer specific upstream regulators of Fe-responsive modules. More than 50% of freq-pCREs (60 out of 115) matched TFBMs of a specific TF (Fig. 5A). Of those, 28 freq-pCREs shared perfect sequence similarity (PCC = 1) to the TFBM. This shows that, in principle, TFs can bind to these freq-pCREs in vitro. From these perfect matches, 22 pCREs were unique for GS-enriched clusters. TF candidates for these 22 cases were for example FUS3 (an ABI3VP1/B3 TF), bHLH104, bZIP3, 16 and 42, PLASTID TRANSCRIPTION FACTOR1 (PTF1, a TCP TF), and FAR-RED IMPAIRED RESPONSE1 (FAR1; Supplemental Table S5). These results are explored in the next section.

While the DAP-seq and CIS-BP TFBM databases contain binding information for many TFs, they are far from exhaustive. For example, out of 162 known Arabidopsis bHLH proteins (Bailey et al., 2003), only 46 were available to be included in the analysis. Therefore, some TF families were likely underrepresented in our analysis and some top match TFBMs may not accurately reflect the binding partner for certain pCREs. Consequently, some important pCRE-TFBM matches might not be detectable at this time. However, as new experimental TF binding data are collected, we might gain more biological insight into our Fe pCREs.
Inferring Upstream Regulators of the −Fe Response

Because our genome-wide approach for identifying regulatory elements may shed light on areas of −Fe responses that are less well understood, we next interpreted our findings in context with open questions in the field. For example, the ABI3VP1/B3-type TF IDE1, a key regulatory factor of −Fe responses in rice and barley roots, binds in vivo and in vitro to the CATGC core of IDE1 (Kobayashi et al., 2007; Kobayashi et al., 2009; Kobayashi et al., 2010). With ten of our freq-pCREs having a CATGC (or GCATG) core and matching ABI3VP1/B3 family TFBMs, IDE1-likes were fairly dominant among the freq-pCREs and unique to GS-enriched clusters (Supplemental Table S4). Besides validating our approach, this strongly suggests an important function for IDE1-like motifs in Arabidopsis. IDE1-containing grass promoters function in a heterologous manner in nongrass plants, as shown in vivo in transgenic tobacco (Kobayashi et al., 2003). Corresponding tobacco TFs are not known yet. The closest homologs of the rice IDE1 in Arabidopsis are AFLs (B3 family TFs ABA-INSENSITIVE3 [ABI3]/FUSCA3 [FUS3]/LEAFY COTYLEDON2 [LEC2]). In fact, ABI3 and FUS3 bind in vitro to Ry-like elements (CATGCA), regulating FeS cluster subunit formation during seed maturation (Roschitztardt et al., 2009). However, ABI3 or FUS3 functions during later developmental stages, particularly in the root during −Fe response, remain to be elucidated. Because the FUS3 TFBM matched our top most abundant IDE1-like (CATGCC; Supplemental Table S5), and because FUS3 is expressed in the root epidermis and in lateral root primordia during later developmental stages (Boulard et al., 2017; Tang et al., 2017), FUS3 might be an IDE1 homolog in Strategy I plants.

Another −Fe response-related TF in rice and barley, IDE2, belongs to the NAC family and binds to the CA(A/C)G(T/C)(T/C/A)(T/C/A) core in IDE2 (Ogo et al., 2008). Although we did not have a perfect (PCC = 1) pCRE-NAC TFBM match, we found matched NAC TFBMs slightly overrepresented in GS-enriched clusters, but have no known direct roles in −Fe responses. However, bZIP TFs are known regulators of the zinc (Zn) deficiency response, which, together with the fact that one GS gene, ZIP9, is also responsive to Zn deficiency, could indicate an interdependency of Zn and Fe homeostasis (Assunção et al., 2010; Sinclair et al., 2018). Furthermore, two matched

**Figure 5.** Similarity of freq-pCREs to in vitro identified TFBMs. A, Sequence similarity of freq-pCREs from GS-enriched clusters to the best matching TFBM from DAP-seq or CIS-BP data. If the freq-pCRE matched TFBMs from both data sets, DAP-seq match was preferred over CIS-BP match (Supplemental Table S5). Bars = 95th percentile (PCC) significance thresholds for within TF family (red, pCRE sequence is more similar to a specific TFBM than to other TFBMs from the same family), between TF families (light blue, pCRE more similar to a TFBM in a family than to TFBMs from other families), or random (dark blue, pCRE more similar to a TFBM from a family than to random 6-mers). Similarity of freq-pCREs from nonenriched clusters to TFBMs: Supplemental Figure S7. B, Proportion of TF family TFBMs (representing freq-pCRE matches meeting at least “between” threshold) in GS-enriched clusters (x axis) and nonenriched clusters (y axis). TFBM matches significantly overrepresented (FET, q < 0.05) in the GS-enriched or nonenriched cluster category are depicted in red and marked with an “X.” Dashed line marks theoretical position for TF family TFBMs with the same proportion in both categories.

**Inferring Upstream Regulators of the −Fe Response**

Because our genome-wide approach for identifying regulatory elements may shed light on areas of −Fe responses that are less well understood, we next interpreted our findings in context with open questions in the field. For example, the ABI3VP1/B3-type TF IDE1, a key regulatory factor of −Fe responses in rice and barley roots, binds in vivo and in vitro to the CATGC core of IDE1 (Kobayashi et al., 2007; Kobayashi et al., 2009; Kobayashi et al., 2010). With ten of our freq-pCREs having a CATGC (or GCATG) core and matching ABI3VP1/B3 family TFBMs, IDE1-likes were fairly dominant among the freq-pCREs and unique to GS-enriched clusters (Supplemental Table S4). Besides validating our approach, this strongly suggests an important function for IDE1-like motifs in Arabidopsis. IDE1-containing grass promoters function in a heterologous manner in nongrass plants, as shown in vivo in transgenic tobacco (Kobayashi et al., 2003). Corresponding tobacco TFs are not known yet. The closest homologs of the rice IDE1 in Arabidopsis are AFLs (B3 family TFs ABA-INSENSITIVE3 [ABI3]/FUSCA3 [FUS3]/LEAFY COTYLEDON2 [LEC2]). In fact, ABI3 and FUS3 bind in vitro to Ry-like elements (CATGCA), regulating FeS cluster subunit formation during seed maturation (Roschitztardt et al., 2009). However, ABI3 or FUS3 functions during later developmental stages, particularly in the root during −Fe response, remain to be elucidated. Because the FUS3 TFBM matched our top most abundant IDE1-like (CATGCC; Supplemental Table S5), and because FUS3 is expressed in the root epidermis and in lateral root primordia during later developmental stages (Boulard et al., 2017; Tang et al., 2017), FUS3 might be an IDE1 homolog in Strategy I plants.

Another −Fe response-related TF in rice and barley, IDE2, belongs to the NAC family and binds to the CA(A/C)G(T/C)(T/C/A)(T/C/A) core in IDE2 (Ogo et al., 2008). Although we did not have a perfect (PCC = 1) pCRE-NAC TFBM match, we found matched NAC TFBMs slightly overrepresented in GS-enriched clusters (Fig. 5B). Furthermore, two of the top ten most abundant freq-pCREs unique to GS-enriched clusters matched NAC TFBMs (PCC > 0.9), with one freq-pCRE being highly similar to the IDE2 core (CACGCC). This indicates that IDE2-like motifs might also play a role during Arabidopsis −Fe responses.

One freq-pCRE (CGTGCC) perfectly matched to a bHLH104 TFBM (Supplemental Table S5). We identified CGTGCC among others in the promoter of BHLH101 (AT5G04150; Supplemental Table S2), which is an experimentally validated target of bHLH104 during −Fe responses (Zhang et al., 2015; Li et al., 2016). Other freq-pCREs matched to known TFBMs from TFs with unknown roles in −Fe responses. For example, bZIP TFBMs were significantly overrepresented in GS-enriched clusters, but have no known direct roles in −Fe responses. However, bZIP TFs are known regulators of the zinc (Zn) deficiency response, which, together with the fact that one GS gene, ZIP9, is also responsive to Zn deficiency, could indicate an interdependency of Zn and Fe homeostasis (Assunção et al., 2010; Sinclair et al., 2018). Furthermore, two matched
TFs, bZIP3 and bZIP16, are involved in ABA signaling, which is connected to –Fe responses by modulating root growth (Séguelé et al., 2008; Miatelli et al., 2011; Hsieh et al., 2012). Possible functions of bZIP TFs in response to –Fe stress should be explored in the future.

PTF1 and FAR1 TF BMs are two more examples for perfect freq-pCRE matches with yet unknown specific roles of the TFs during –Fe. Their specificity to GS-enriched clusters points toward important roles in regulating GS genes. TCPs are involved in plant development, but also in signaling of hormones that influence –Fe responses (Davière et al., 2014; Brumbarova et al., 2015; Resentini et al., 2015; Nicolas and Cubas, 2016). For example, TCP20 was reported to bind to the BHLH39 promoter (Andriankaja et al., 2014), indicating a possible connection of TCPs and –Fe responses during plant development. PTF1 is involved in regulating responses to light shade signals through PHYTOCHROME INTERACTING FACTORS (PIFs; Zhou et al., 2018). Interestingly, FAR1 and its homolog FAR-RED ELONGATED HYOCOTYL3 (FHY3) also act in phytochrome-PIF signaling (Wang and Wang, 2015). Together, this suggests a connection of light perception and –Fe responses mediated through these TFs, which is consistent with the known diurnal influence on Fe uptake (Vert et al., 2003; Santi and Schmidt, 2009; Hong et al., 2013; Salomé et al., 2013). In addition, FAR1/FHY3 act in the regulation of the phosphate starvation response, together with ethylene regulator ETHYLENE-INSENSITIVE3 (EIN3; Liu et al., 2017), which also binds FIT to promote Fe uptake (Lingam et al., 2011). Therefore, FAR1 might also regulate Fe acquisition via the ethylene pathway.

Finally, a perfect freq-pCRE-WRKY11 match indicates that WRKY TFs, although significantly overrepresented in nonenriched clusters, are also important for regulating GS-enriched clusters. WRKY11 is involved in abiotic stress tolerance in Arabidopsis (Ali et al., 2018), with no specific role known during –Fe response yet. However, WRKY TFs in general have already been connected to –Fe, for example as putative regulators of the coumarin transporter gene PLEIOTROPIC DRUG RESISTANCE9 (PDR9; Ito and Gray, 2006) and of PYE (Koryachko et al., 2015). Furthermore, WRKY46 negatively regulates the vacuolar Fe importer gene VACUOLAR IRON TRANSPORTER-LIKE1 (VTL1; Gollhofer et al., 2011; Gollhofer et al., 2014; Yan et al., 2016). We found a WRKY TF BM (GTCAAC) in several nonenriched clusters containing down-regulated Fe-responsive genes, including the VTL1 homolog VTL5 (AT3G25190; Supplemental Table S2), indicating that some of the TFs matching nonenriched cluster pCREs might act as repressors of Fe excess genes.

In summary, many pCREs commonly found among GS-enriched clusters share significant sequence similarity with known –Fe CREs, such as IDE1, or with binding sites of known –Fe-associated TF families, such as AB3VPI/B3, NAC, MYB, and bHLH. Notably, we found evidence for IDE1-like motifs being relevant not only in Strategy II plants, but also in the Strategy I plant Arabidopsis. Our results also suggest a role of bZIPs or TCPs in –Fe responses. We assessed in the next paragraph specific –Fe response processes in which pCREs of particular interest, such as IDE1-likes, might be involved.

Identifying the Most Important pCREs (min-pCREs) and Associating Them with FIT-Dependent or FIT-Independent Functions

After identifying potential regulators in the –Fe response, we pinpointed some of those that could best explain models of –Fe-responsive up-regulation and explored their potential functions.

More than 1,500 pCREs were identified in total in GS-enriched clusters, raising the question of a core set of important pCREs needed to robustly predict –Fe responses in each cluster. Using pCRE abundance (freq-pCREs) across GS-enriched clusters as the only selection criterion could result in missing motifs, simply due to the fact that some coexpression clusters were more unique than others. This is supported by the fact that rare pCREs (i.e. pCREs identified in few clusters) still can have a high importance rank (Fig. 4A). Those pCREs were not included in the set of freq-pCREs, although they seem to be important for regulating individual GS-enriched clusters. We identified the most important pCREs for a cluster model (defined as the minimum set of pCREs; min-pCREs) by building RF models iteratively with successively deleting the least important pCREs in each round (see “Materials and Methods”; Supplemental Fig. S8A). We applied this approach to each of the 159 GS-enriched clusters. In most cases (117 of the clusters), 50% to 90% of the pCREs were not important for model performance, mostly due to sequence redundancy with a more important pCRE (Supplemental Table S2). pCREs were regarded as redundant if they were reverse complements or if one was the subset of the other. We calculated that, on average, 27.8 pCREs were identified in a GS-enriched cluster, and 10.3 of them were min-pCREs. Supplemental Figure S8B shows the pCRE presence/absence patterns of Figure 3, D and E, example clusters (IDs: 493, 823), with the min-pCREs highlighted. Genes in cluster 493 had on average 9.2 min-pCREs and genes in cluster 823 had on average 4.09 min-pCREs. We also counted the min-pCREs per gene in the other 157 GS-enriched clusters (Supplemental Table S2) and calculated that, on average, 7.01 different min-pCREs were present in a gene’s promoter. This indicates that several pCREs are needed for a gene to be –Fe-responsive.

The collective set was 615 different min-pCREs across GS-enriched clusters. Each min-pCRE was identified in at least 1 (0.6%) and in up to 48 (30%) GS-enriched clusters. The IDE1-like CATGCC was the top most abundant min-pCRE. Together with CATGCC, two more IDE1-like motifs, TCATGC and CCATGC, were among the top ten most abundant min-pCREs.
(Supplemental Table S4). This supports a previous computational analysis of rice promoters, in which IDE1-like was among the top scoring motifs (Kakei et al., 2013). Together with our previous finding that IDE1-like ABI3VP1/B3 TFBMs are associated with GS-enriched clusters, it suggests an important, yet unknown, function of IDE1-like motifs in Arabidopsis – Fe response regulation. Min-pCREs matching a bHLH (CGTGAC), a MYB (TAACTA), and the IDE2-like NAC TFBM (CAGGCC; all Supplemental Table S5) were also among the top ten most abundant min-pCREs (Supplemental Table S4), further demonstrating the utility of our approach.

To determine in which processes min-pCREs might function during −Fe, we tested if min-pCREs were more likely to be found in FIT-dependent or FIT-independent coexpression clusters. More than 60% of the 159 GS-enriched clusters were classified as either FIT dependent (35% out of 159) or FIT independent (28% out of 159; Fig. 6A). We then calculated the proportion of min-pCREs (present in ≥5 GS-enriched clusters) in each cluster category (Supplemental Table S6). Interestingly, the two IDE1-like motifs, CATGCC, CCAATGC, and the related ABI3VP1/B3 TFBM match ATGCAT, were predominantly identified in FIT-dependent clusters, but IDE2-like CACGCC had no preference for either FIT-dependent or FIT-independent clusters (Fig. 6B). This suggests that the IDE1-like pCREs tend to be more important for FIT-dependent root Fe acquisition rather than FIT-independent Fe sensing, signaling, and distribution. This is also consistent with the role of grass IDE1 in Fe uptake (Kobayashi et al., 2003; Kobayashi et al., 2005). Two ARF TFBM-matching min-pCREs (AACGTA/ARF16, GTCGGA/ARF2) were also preferentially found in FIT-dependent clusters. ARFs are involved in auxin signaling, thereby controlling, among other functions, root hair elongation (Pitts et al., 1998; Mangano et al., 2017; Choi et al., 2018). Different studies reported that −Fe responses can be accompanied by an increase of root hair number, by elongation of root hairs, and by deformed or short root hairs (Schmidt et al., 2000; Müller and Schmidt, 2004; Dinneny et al., 2008). ARF2 and ARF16 TFs are root hair growth repressors (Choi et al., 2018), which would be consistent with a short root hair phenotype and downregulation of respective GO terms under −Fe (Dinneny et al., 2008; Supplemental Fig. S1). During −Fe, several root hair-acting genes are coexpressed in a regulon, which also contains IRON-REGULATED TRANSPORTER2 (IRT2; Ivanov et al., 2012; Schwarz and Bauer, 2020), indicating a possible connection of ARF TFBM-matching min-pCREs with these root hair processes.

In contrast, three bZIP TFBM-matching min-pCREs, GTGCGA, CACCTC, and CACTAC, were predominately identified in FIT-independent clusters. As described in the previous section, bZIPs are involved in ABA signaling. ABA negatively regulates FIT-dependent Fe³⁺ reductase gene FERRIC REDUCTION OXIDASE2 (FRO2) and Fe²⁺ importer gene IRON-REGULATED TRANSPORTER1 (IRT1; Séguéla et al., 2008). However, ABA signaling also leads to enhanced apoplastic and vacuolar Fe utilization and root to shoot transport under −Fe (Séguéla et al., 2008; Lei et al., 2014). Lei et al., 2014 propose that ABA-responsive gene regulation and Fe remobilization and transport are connected through bZIPS, which is consistent with our results that bZIP TFBMs are preferentially found in FIT-independent coexpression clusters of genes involved in Fe mobilization and translocation. Similarly, two TCP TFBMs, GAC-CAC and ACCCCAC, were identified almost exclusively in FIT-independent clusters, which is in agreement with TCP20 regulating BHLH39 in a FIT-independent manner (Andriankaja et al., 2014). Finally, bHLH TFBMs were identified in all cluster categories with a preference for mixed clusters. This matches the ubiquitous nature of bHLH target motifs (E-/G-boxes), which act at many levels in the −Fe bHLH regulatory cascade. In summary, we can propose plausible roles for pCREs as TFBMs in FIT-dependent and FIT-independent processes.

Positional Bias, Conservation, and Inferred Functions of min-pCREs

Next, to explore if min-pCREs displayed significant positional bias in the promoter regions of coexpressed genes, we compared the observed min-pCRE frequencies in 100-bp bins of −1000 to +500 bp and of −500 to +1000 bp flanking regions adjacent to the transcription start site (TSS) and transcription termination site (TTS), respectively, with the expected frequencies from shuffled pCRE sequences (according to Uygun et al., 2017; Supplemental Fig. S9). Furthermore, we examined the noncoding as well as the coding sequences of the transcribed regions. Overall, the distributions of min-pCREs revealed a slight positional bias in the promoter regions (Fig. 6C, top). We investigated the distribution plots separately for ten selected min-pCREs: FIT-dependent ABI3VP1/B3 TFBM-matching min-pCREs (containing the two IDE1-likes), the IDE2-like (NAC TFBM match), FIT-independent bZIP TFBM matches, and bHLH TFBM matches (Fig. 6C). These min-pCREs had significant location bias in the putative promoters up to 1000 bp upstream of the TSS. Because known CREs often exhibit positional bias (Zou et al., 2011; Heyndrickx et al., 2014; Yu et al., 2016), this provides additional support for these pCREs having regulatory functions in Fe uptake and homeostasis. Interestingly, some of the pCREs common to mixed clusters did not show position bias in any of the genomic regions tested (e.g. ABI3VP1/CTTATA and MYB/TAACTA; Fig. 6C, bottom), indicating that genes of such clusters are less likely to be truly transcriptionally coregulated.

Next, we sought to pinpoint the specific processes of Fe homeostasis, which these ten min-pCREs might regulate. To assess this, we determined the genes containing the respective min-pCRE and counted the number of incidents in which these genes were likely
Figure 6. Characteristics of the most informative pCREs (min-pCREs). A, Proportion of GS-enriched clusters enriched with FIT-dependent genes (FIT, blue), FIT-independent genes (non-FIT, red), or both (mixed, gray). B, Ternary plot of min-pCREs identified in >3% (n = 5) GS-enriched clusters. Position of the min-pCREs corresponds to the normalized proportions of FIT, non-FIT, and mixed clusters in which the min-pCRE was identified. Bubble size corresponds to the overall proportion of GS-enriched clusters with min-pCRE. Labeled min-pCREs are shown in (C) and (D) or mentioned in the main text. C, Positional bias of all (mean with SD; top) and selected min-pCREs (below) in the putative promoter region (1st column), all introns and all exons (Ex; 2nd column; mean with SD), and in the putative noncoding region (3rd column). 1st and 3rd column: position distributions in all coexpression clusters with min-pCRE (gray areas). Red line: mean distribution. TFBM matches (PCC) for each min-pCRE are shown (4th column), and min-pCREs are sorted by TF family. Log2(obs/exp), log2 of the number of observed (obs) min-pCRE occurrences divided by the number of min-pCRE occurrences in randomized sequences (expected, exp). D, Genes which might be regulated by the selected min-pCREs. Count, number of GS-enriched clusters in which the min-pCRE was identified and which included the respective gene having the min-pCRE in its promoter. Dashed line, total number of GS-enriched clusters with the min-pCRE. Genes in which the min-pCRE overlaps with a CNS are highlighted black. A high-resolution image is available as Supplemental Figure S10.
regulated by the min-pCRE (Fig. 6D; high-resolution image Supplemental Fig. S10; Supplemental Table S7). For example, CATGCC was found in 48 GS-enriched clusters, and 36 of those clusters (75%) included the min-pCRE-containing gene MYB10, whereas only four of those clusters (8%) included MEK KINASE16 (MAPKKK16; Fig. 6D, second row at right). We inferred that CATGCC might be regulating predominantly processes in which MYB10 is required. As an additional line of evidence for pCRE functionality, we determined if min-pCREs overlapped with conserved noncoding sequences (CNS) of the Brassicaceae family (Supplemental Table S8; Haudry et al., 2013). As expected, bHLH TFBM-matching min-pCREs were identified in many gene promoters, including BHLH39/BHLH101 (direct targets of bHLH IVc TFs; Zhang et al., 2015), NAS4 (direct PYE target; Long et al., 2010), and IRT1 and GENERAL REGULATORY FACTOR11 (GRF11; FIT targets; Fig. 6D; Sivitz et al., 2012; Yang et al., 2013). In BHLH39/BHLH101, IRT1, and GRF11, the respective min-pCREs overlapped with CNS, further supporting the importance of these motifs. Interestingly, bHLH matched min-pCREs were also located in CNS of BRUTUS (BTS) and BRUTUS-LIKE1 (BTSL1), two genes that negatively regulate Fe uptake by marking positive regulators for degradation (Selote et al., 2015). If BTS were to be regulated by bHLH proteins from the same regulatory cascade, this may indicate a negative feedback loop.

FIT-dependent IDE1-likes CATGCC and CCATGC were located in the IRT1 promoter, overlapping with CNS (Fig. 6D, second row). Interestingly, both IDE1-likes were found in several genes encoding enzymes and TFs involved in coumarin biosynthesis (cytochrome P450 gene CYP82C4, SCOPOLETIN 8-HYDROXYLASE [S8H], FERULOYL-COA 6'-HYDROXYLASE [F6'H1], MYB72/MYB10; Kai et al., 2008; Murgia et al., 2011; Fourcroy et al., 2014; Schmid et al., 2014; Zamiodus et al., 2014; Rajniak et al., 2018; Siwinska et al., 2018). In grasses, IDE1 is common to genes of MA synthesis (Kobayashi et al., 2003; Kobayashi et al., 2005). We propose that IDE1 is important for synthesis of Fe chelators in both monocots and dicots. The IDE2-like min-pCRE was located among others in BTLSL1 (within a CNS), BHLH39, OBP3-RESPONSIVE GENE1 (ORGI), and a number of uncharacterized genes. While the role of IDE2 in the Strategy II Fe response has not been comprehensively explored, it regulates expression of rice YELLOW STRIPE LIKE2 (YSL2), a phloem Fe2+-nicotianamine transporter (Kobayashi et al., 2003; Ogo et al., 2008). This putative involvement in Fe translocation suggests that the IDE2-like might preferentially associate with FIT-independent gene functions. However, we could not assess the relationship between IDE2 and YSLs in this analysis because YSL1/2/3 were not expressed above the log2FC ≥ 1 threshold.

Next, we explored the bZIP TFBM-matching min-pCREs, since they had the strongest FIT-independent preference. These min-pCREs were located in a number of genes involved in translocation of Fe/Fe chelates or the synthesis of Fe chelators (e.g. NATURAL RESISTANCE ASSOCIATED MACROPHAGE PROTEIN4 [NRAMP4], ZIF1, OLIGOPEPTIDE TRANSPORTER3 [OPT3]; Lanquar et al., 2005; Haydon et al., 2012; Mendoza-Cózatl et al., 2014; Zhai et al., 2014) or in Fe sensing and signaling (e.g. OPT3, BTS; Mendoza-Cózatl et al., 2014; Zhai et al., 2014; Selote et al., 2015; Khan et al., 2018). Furthermore, GTGGCA (matched to bZIP TFBM) overlapped with a CNS of CONSERVED IN THE GREEN LINEAGE AND DIATOMS27 (CGLD27; Fig. 6D, third row), which has been associated with photoprotection in leaves during Fe (Ruiz-Sola and Rodríguez-Concepción, 2012; Rodríguez-Celma et al., 2013). However, its function in roots remains elusive. Taken together, our findings suggest diverse roles for bZIP TFBMs, including Fe transport and adjustment of the plastid proteome.

In summary, we identified robust Fe pCREs, which, in addition to sharing significant sequence similarity to known TFBMs, were also part of the core sets of pCREs needed for prediction of Fe responses of GS-enriched clusters (min-pCREs). Furthermore, they were preferentially located in promoter regions upstream of the TSS, and even in CNS’ of some genes. Together, these findings indicate that these pCREs might be authentic Fe CREs. From the biological context of the genes, which are likely regulated by some of the pCREs, we were able to generate hypotheses relevant for understanding Fe response regulation in Arabidopsis. For example, our work highlighted that in addition to the bHLH TFBMs, IDE1-like motifs and bZIP TFBMs are likely involved in different responses to Fe and should be considered of high interest for future work.

**CONCLUSION**

We identified 5,639 pCREs that were enriched in promoters of coexpressed Fe-responsive genes and were used as features to predict Fe-responsive regulation of root-expressed genes on a genome-wide scale. Of those, 173 pCREs reliably predicted Fe response genes of >5% of our defined coexpression clusters (freg-pCREs). Because most of those pCREs were either unique to coexpression clusters enriched with our GS Fe acquisition and homeostasis genes, or unique to coexpression clusters lacking those genes, we conclude that our approach had captured motifs specifically regulating different responses during Fe. Several pCREs contained the previously experimentally validated core motif of IDE1. To take advantage of the publicly available in vitro TF binding information, we compared the freg-pCREs to TFBMs from two studies (Weirauch et al., 2014; O’Malley et al., 2016) and found that our approach had captured known Strategy I Fe recognition motifs for bHLH and MYB proteins. Our approach also led to regulatory connections of bZIP, B3, NAC, and TCP families to Strategy I Fe response regulation. While bZIP and bHLH TF families are also
associated with high salinity stress response (Uygun et al., 2017), other high salinity stress response-associated TF families (e.g., WRKY and AP2) were not common among our −Fe pCREs regulating GS-enriched coexpression clusters, highlighting that our approach pinpoints regulators specific to a stress condition.

We inferred possible functions of pCREs, which were most important for modeling −Fe responses (min-pCREs) from their enrichment in FIT-dependent or FIT-independent coexpression clusters and their location bias in promoters of particular −Fe-responsive genes (Fig. 6, B and D). Our results provide evidence that IDE1-like pCREs are linked to coumarin synthesis, indicating that the function of IDE1 to ensure supply of Fe-chelating compounds for Fe acquisition could be an evolutionarily conserved function, at least among flowering plants. While our results highlight the importance of IDE1-like motifs for Fe acquisition, it was not the only prominent −Fe pCRE. This is in contrast to Zn deficiency, where ZDRE seems to be singularly associated with multiple Zn deficiency responses (Assunção et al., 2010), and indicates that despite of overlaps of Zn and Fe homeostasis control (Briat et al., 2015), their transcriptional regulation must follow different mechanisms.

Our results support a concept in which −Fe is not regulated by only one or few regulatory elements. In fact, 7.01 min-pCREs were present on average in the promoter of a −Fe-responsive gene. Of the many important pCREs for −Fe response, many share significant similarity with TFBSMs of TF families known to undergo hetero-dimerization and protein interaction across families, such as bHLH, MYB, bZIP, TCP, and ABI3VP1 (Bemer et al., 2017). A combinatorial mechanism would dramatically increase the flexibility of transcriptional responses driven by a set of few TFs. It might be that some pCREs not as important in our prediction models, would become informative in combination, as suggested for high salinity stress response (Zou et al., 2011; Uygun et al., 2017). A next step would therefore be to build −Fe prediction models that explicitly account for interactions between pCREs. A limitation of our approach is that our coexpression clusters were based on ATH1 chip microarray data, the only comprehensive −Fe time course transcriptome set available to date. Some important −Fe marker genes (e.g., FRO2) are not represented on the chip and others might have not passed our significance threshold because of sensitivity issues with the microarray technology. Additionally, we restricted our analysis to the promoter regions 1,000 bp upstream of the TSS. While this is expected to cover most important cis-acting elements and reduce the occurrence of promoters overlapping with adjacent genes, introns as well as more distal promoter regions are known to harbor cis-acting elements (Rose et al., 2008; Rose et al., 2016).

The large number of TFs known to be involved in −Fe-induced up-regulation points toward the importance of transcriptional regulation. However, −Fe responses are also heavily controlled at the posttranscriptional and posttranslational level (Lingam et al., 2011; Meiser et al., 2011; Sivitz et al., 2011; Selote et al., 2015; Zhang et al., 2015; Gratza et al., 2019). Naturally, our approach cannot cover such regulatory aspects. However, it allows us to predict TF families, that may act upstream of the known −Fe-responsive genes. Besides bHLH and MYB TFs, we suggest TFs of the bZIP, ABI3VP1/B3, NAC, and TCP families as upstream regulators of −Fe responses in the root. Because major Fe sinks are located in the shoot, a systemic shoot-to-root signal must exist for proper Fe supply (Vert et al., 2003; García et al., 2013). Integrating shoot transcriptomic data would expand our knowledge on how responses to −Fe stress are coordinated at the whole-plant level.

In conclusion, we demonstrate that our machine-learning-based approach can identify pCREs for −Fe-induced gene up-regulation. This strategy can be applied to various stresses and developmental conditions to elucidate regulatory mechanisms, especially when cis- and/or transacting elements were previously elusive (Zou et al., 2011; Uygun et al., 2017). We provide a comprehensive source of potential −Fe response cis-regulators for a wide range of −Fe-responsive genes. Because the identified pCREs are potentially involved in enhancing Fe uptake and translocation, they generate potential for future applications in engineering plants with improved plant performance traits, for example, higher nutritional value because of better Fe allocation and coping with unfavorable soil conditions.

**MATERIALS AND METHODS**

**Expression Data Processing and Generation of Multiple Expression Data Combinations**

Arabidopsis (Arabidopsis thaliana) expression data (Affymetrix ATH1) from a −Fe treatment time course experiment with six time points and of four −Fe-treated root zones (both Dinneny et al., 2008) were downloaded from Gene Expression Omnibus (GEO; https://www.ncbi.nlm.nih.gov/geo/; GSE10502, GSE10497) in the Affymetrix CEL data file format (stores probe intensities from a single ATH1 chip) preprocessed, normalized, and contrasted as described below. AtGenExpress expression data (Affymetrix ATH1) of abiotic stresses (Kilian et al., 2007; GSE5620-5628 or TAIR-ME00325-330, only data of root samples were used), hormone treatment (Goda et al., 2008; GSE39384 or TAIR- ME00333-340, ME00343-344, ME00350-352, ME00356), and plant development (Schmid et al., 2005; GSE6529-5634 or TAIR-ME00319) were downloaded from The Arabidopsis Information Resource (TAIR; https://www.arabidopsis.org/portals/expression/microarray/AtGenExpress.jsp), preprocessed, normalized, and contrasted by S. Uygun (Uygun et al., 2016) as described below. Background correction and quantile normalization of CEL files were performed with Robust Multi-Array Average expression measure using the Bioconductor affy package (Gautier et al., 2004). The log2 fold-change (log2FC) in expression was calculated for all data sets except developmental data by pairwise comparison of treatment and control experiments for each treatment and time point. Contrast matrices and linear model fits were created using R and the Bioconductor LIMMA package (Ritchie et al., 2015; Phipson et al., 2016). Because developmental stages have no control treatment, absolute normalized fluorescence intensity values were used. The p-values for log2FC or fluorescence intensities were corrected for multiple testing (adjusted p-values = q) using the BH (Benjamini-Hochberg) method (Benjamini and Hochberg, 1995). Genes were regarded as −Fe-responsive if abs(log2FC) ≥ 1, and q < 0.05 at least at one −Fe treatment time point or in at least one −Fe treated root zone. −Fe deficiency time course data were combined with −Fe root zone expression data or AtGenExpress data sets in different combinations (Fig. 1C) either including only genes up-regulated (“up”) or all genes up- or down-regulated (“up and...
Coexpression Clustering using k-Means

To cluster genes with similar expression pattern, k-means clustering was implemented in R using the Hartigan-Wong algorithm (Hartigan and Wong, 1979). Because k-means returns a local optimum solution depending on the number of clusters (k) created and the random selection of genes as initial “means”, the outcome varies with run (i.e. nondeterministic). Therefore, different k (25, 30, 35, 40, 50, 70, 80, 100) were tested, and the clustering was repeated up to four times. Machine learning models were built (see below) with all clusters generated from expression data combinations (DC) 1, 2, 3, and 5 (Fig. 1C). To prevent confusion, it should be noted that the total number of k-means clusters used to build models represents several repeated clustering events of always the same two sets of – Fe-responsive genes (up; up and down, see above). The clustering events differ in the DC, which was used and in the k. Two DC were excluded from the analysis: DC 4 produced identical clusters as DC 1, which were therefore not considered. DC 6 contained different measuring units (log2FC and absolute normalized fluorescence intensity) and could not be handled by the k-means algorithm.

Coexpression Clustering by Correlation with GS Genes

To generate coexpression clusters based on gold standard (GS) genes (Supplemental Table S1), each GS gene was used as a query to identify genes with similar expression patterns. Briefly, for each expression DC (Fig. 1C), the PCC was calculated between the query gene and each gene in DC using SciPy (http://www.scipy.org; Jones et al., 2001). Similar to Uygun et al., 2016, a random background PCC was calculated representing the null distribution of expression correlation by calculating the PCC of 10,000 randomly selected gene pairs in DC and the 95th percentile of PCCs was tried as the threshold for classifying a pair of genes as significantly correlated. For some DC (mostly those containing only up-regulated – Fe-responsive genes), a significance threshold was lowered below 90% down to 45%, because the PCC between random – Fe-responsive genes was already very high. On the other hand, when >50 genes were considered significantly correlated, the threshold was raised above 95% to 99% to home in on genes most likely to be coregulated. In addition, a second version of clusters was generated with >50 genes, containing only the 10 genes with highest PCC. Machine learning models were built (see below) with both versions, and the results were further used from the better performing version only. Percentiles used for each PCC-based cluster are given in Supplemental Table S2. Two DC were excluded from the analysis: DC 4 (up), because the resulting clusters were identical to those generated from DC 1 (up) and DC 6 (up and down), because developmental data seemed to have a disproportional influence on the PCC with the result that even – Fe up- and down-regulated gene pairs were identified as strongly correlating. As in the k-means clustering, the total number of PCC-based clusters used for modeling represented clustering events of the same two sets of – Fe-responsive genes (as described above).

The Coexpression Clusters: GO and GS/FIT-Dependent/ FIT-Independent Gene Enrichment and GO/Gene Content Similarity

GO associations for Arabidopsis were downloaded from TAIR (ftp://ftp.arabidopsis.org/home/tair/Genes_Ontology; Berardini et al., 2004). Biological process (BP) GO annotations were downloaded from GO (http://purl.obolibrary.org/obo/go.obo) and parsed for BP information. Enrichment of GO terms in genes that were significantly differentially regulated (q < 0.05, abs(log2FC) > 3) in the – Fe time course data set was determined with a FET implemented in Python using SciPy (Jones et al., 2001), and q-values were corrected for multiple testing (q α) using the “qvalue” function in R (Storey, 2002; Supplemental Table S9).

All coexpression clusters were tested for enrichment of GO terms as described above. The similarity of enriched GOs between coexpression clusters was assessed using the Jaccard index (J), or the intersection of GOs divided by the union of the GOs, where J = 1 if the exact same GOs were enriched in both coexpression clusters. Coexpression clusters were grouped by hierarchical clustering using the JI with the UPGMA method in the R cluster package (Maechler et al., 2017). Groups containing ≥20 coexpression clusters and having a within-mean JI that was significantly higher than the mean JI of all clusters were defined as “superclusters.” Biological functions of superclusters were defined through GOs shared by ≥75% (k-means clustering) or ≥90% (GS gene correlation; PCC) of the clusters. Similarly, FET with p-value correction for multiple testing was used to identify coexpression clusters enriched with (A – Fe GS genes, (B) FIT-dependent genes, and/or (C) FIT-independent genes.

K-mer Enrichment and Identification of pCREs Predictive of – Fe Response using Random Forest (pCRE Identification Pipeline)

Promoter sequences 1 kb upstream from the transcription start site (TSS) were downloaded from TAIR (ftp://ftp.arabidopsis.org/home/tair/Sequences/blast_databases/TAIR10_blastsets/upstream_sequences/TAIR10_upstream_1000_20101104). A list of all possible 6-mers of A, T, C, G was generated with the Python itertools function and the Biopython Bio.Seq module (https://biopython.org/docs/dev/api/Bio.Seq.html; Cock et al., 2009). Only one 6-mer for each reverse complement pair was kept (resulting in 2,080 6-mers).

Potentially meaningful cis-regulatory elements for – Fe response were identified in two steps, where we first looked for enriched k-mers in the promoters of – Fe-responsive genes and then determined how well the enriched k-mers predicted – Fe response using machine learning. The code for this analysis is available on GitHub (https://github.com/ShiuLab/MotiDiscovery). For the first step, the promoter sequences of the genes in coexpression clusters (positive set) were searched for enriched 6-mers in comparison with promoter sequences of nonresponsive genes (negative set, as defined below; 10,068 genes). These enriched 6-mers were elongated by one base and tested again for enrichment. This process was repeated until no longer k-mer was more enriched than the shorter k-mer. Enrichment was calculated using a one-sided FET (P < 0.01).

For the second step, to determine which sets of enriched k-mers were predictive of – Fe response, we generated features based on presence or absence of each enriched k-mer and used these features to build machine learning models using the RF algorithm. RF was implemented in Python using Scikit-Learn (Pedregosa et al., 2011). To avoid building biased models, 50 models were generated for each coexpression cluster. Each time, the same number of positive set genes were randomly drawn from the negative set to generate balanced (i.e. size positive set equals size of negative set) input data sets. A 10-fold cross-validation approach was used to train and test the models. Briefly, the balanced data sets were divided randomly into ten even groups with a 1:1 ratio of positive to negative class genes. The model was trained on the 1~9 folds and applied to the 10th (and successively trained on 1~8~10 and applied to the 9th, etc.). This cross-validation scheme was repeated ten times. Each RF model was made up of 500 decision trees, each trained on a random subset of enriched k-mers and of training set genes. The final model performance is represented by the mean F1 score (i.e. F-measure) across all 50 balanced models. The F1 score is the harmonic mean of precision, P = TP/(TP + FP) and recall, R = TP/(TP + FN), where TP = true positive, FP = false positive, and FN = false negative. Only coexpression clusters for which the enriched k-mers were deemed as good predictors (F1 ≥ 0.7) were used in the downstream analysis. Predictive k-mers (then referred to as putative cis-regulatory elements, pCREs) were ranked by importance. The importance score is based on the Gini index (Breiman et al., 1994), built into the Scikit-Learn implementation of RF. The Gini index is defined as the total decrease in node impurity (i.e. how accurately positive [– Fe-responsive] and negative [nonresponsive] class genes are split by a feature [pCRE] at a particular node) averaged over all decision trees. The importance score of a pCRE increases with the frequency it was selected for a node split and with its overall ability to correctly classify –Fe-responsive and nonresponsive genes. To determine how well the models predicted individual – Fe-responsive genes, the RF models were rerun with an updated code, which outputs the prediction scores for each gene during each of the 50 balanced replicates (https://github.com/ShiuLab/ML-Pipeline). Using the scores the ratio of correct predictions (TP) was calculated by counting the number of times that a gene was correctly predicted out of the 50 replicates.

Genes were considered robustly – Fe nonresponsive if they were not significantly differentially expressed (abs(log2FC) < 0.4) at any time point during the – Fe time course experiment or in any – Fe-treated root zone or in four additional – Fe treatment experiments (Li and Schmidt, 2010; GSE16964; Long et al., 2010; GSE21443; Schuler et al., 2011; GSE24348; Shivitz et al., 2012; GSE40076). The four additional data sets were downloaded in CEL format from GEO and processed as described in the first section of “Materials and Methods.”
The log2FC threshold for nonresponsive genes was chosen far below abs(log2FC) < 1 to create a safe log2FC distance between positive and negative class genes. This was regarded as helpful for the RF classifier, since genes with abs(log2FC) ≥ 0.858 (corresponds to FC ≥ 1.5) can be considered weakly differentially expressed. However, to maintain a large pool of negative set genes for the random draws during RF, the threshold was not pushed below abs(log2FC) < 0.4.

**pCRE Sequence Similarity**

To assess sequence similarities between the 173 pCREs that were frequently identified (in >5% freq-pCREs) in GS-enriched or nonenriched coexpression clusters with good model performance (F1 ≥ 0.7), sequence dissimilarity of pCRE position weight matrices (PWMs) was calculated by pairwise PCC distance and a distance matrix was generated using the TAMO package in Python (Gordon et al., 2005). Freq-pCREs were grouped by hierarchical clustering of the PCC distance matrix using the UPGMA method in the R cluster package (Maechler et al., 2017), and visualized in a dendrogram (Supplemental Fig. S6E). Due to groupwise averaging of PCC distances during hierarchical clustering, the algorithm produced skewed PCC distances of some similar pCRE pairs. Therefore, freq-pCRE clusters were additionally visualized as a network, in which freq-pCREs with PCC distance = 0 (identical freq-pCREs or subsets of each other) were connected with black bold edges and freq-pCREs with PCC distance ≤ 0.22 were connected with light gray edges (Fig. 4C). Highly interconnected nodes were arranged in groups. The network was created using the Cytoscape software (Shannon et al., 2003). To show a consensus of freq-pCREs within a network group, freq-pCRE sequences were aligned using ClustalX (Larkin et al., 2007) with default parameters and a sequence logo was created with weblogo (https://weblogo.berkeley.edu/logo.cgi).

**Identification of Most Informative pCREs (min-pCREs) by Nonlinear Regression**

The most informative pCREs of a coexpression cluster were defined as the minimum set of pCREs (min-pCREs) needed for RF models without sacrificing performance. To identify min-pCREs, for each GS-enriched coexpression cluster, the pCREs used as features were stepwise reduced, with the least important pCREs deleted at each step. First, for pCREs that were subsets of each other (PCC distance = 0), the lower ranked one was removed. Then, from this list of pCREs and for successively shorter lists of pCREs (n = 40, 30, 25, 20, 15, 12, 10, 8, 6, 5, 4, 3, 2, 1), 10 replicates of RF models were trained on balanced data sets. F1 scores were plotted against the number of pCREs (n) and a nonlinear regression curve was fitted to the data points. An exponential recovery function:

\[
F_1(x) = a \left(1 - e^{-c \cdot n} \right)
\]

was found to best describe the data behavior. Starting values for variables a and c were approximated by fitting a linear model to the logarithmic transformation of the function. The set of pCREs with the highest F1 closest to the inflection point of the regression curve was defined as min-pCRE set (example in Supplemental Fig. S8A).

**pCRE Similarity to TFBMs**

Two different sources of in vitro TF binding data were used. In the DNA Affinity Purification Sequencing database (DAP-seq; O’Malley et al., 2016), TFBMs were determined by in vitro binding of Arabidopsis TFs to genomic DNA and sequencing of the bound DNA fraction. In the Catalog of Inferred Sequence Binding Preferences database (CIS-BP; Weirauch et al., 2014), TFBMs are based on in vitro experiments (protein binding microarrays) or inferred TRAFAC models. pCREs were compared with 510 different DAP-seq TFBMs from 49 different TF families and 297 different CIS-BP TFBMs from 32 different families (Supplemental Table S10). DAP-seq TFBMs were preferred over CIS-BP TFBMs when a pCRE-TFBM match was present in both databases.

PWMs of pCREs were compared with PWMs of TFBMs using PCC, and the pCREs were classified as similar to (A) a specific TF, (B) a TF family, or (C) to TFs generally, based on the degree of similarity to their best matching TFBM (Uygur et al., 2017). A pCRE is similar to a specific TFBM (A) if the pCRE between the pCRE and the TFBM was ≥ 95th percentile of PCCs between that TFBM and TFBMs from the same TF family. Alternatively, a pCRE is similar to TFBMs from a TF family (B) if the pCRE between the pCRE and a TFBM from that family was ≥ 95th percentile of PCCs between TFBMs from that family and TFBMs from other TF families. Finally, a pCRE was similar to TFBMs (C) if the pCRE between the pCRE and any known TFBM was ≥ 95th percentile of PCCs between TFBMs and randomly generated 6-mers. For 95th percentile PCC thresholds see Supplemental Table S11. To determine if pCREs similar to specific TF families were enriched in GS-enriched versus nonenriched coexpression clusters, the percentage of pCREs similar to TFBMs (significance level A or B) from each TF family was calculated for each coexpression cluster category. Then, FET with multiple testing correction (q ≤ 0.05) was used to determine if GS-enriched coexpression clusters were enriched with TF families compared with non-GS-enriched coexpression clusters and vice versa.

**Positional Distribution of pCREs**

To determine the positional distribution of the min-pCREs for each GS-enriched coexpression cluster, min-pCREs were converted to PWMs adjusted to the Arabidopsis background AT (0.33) and CG (0.17) content using the TAMO package (Gordon et al., 2005) and mapped to the promoter sequences ranging from 1,000 bp upstream to 500 bp downstream of the transcription start site (1000-TSS-500), using Motility (https://github.com/cbt/motility). For comparison, min-pCRE PWMs were also mapped to exons and introns, respectively, and to the region 500 bp upstream and 1,000 bp downstream of the transcription termination site (500-TTS-1000). Arabidopsis sequences were downloaded from TAIR (ftp://ftp.arabidopsis.org/Sequences/blast_databases/TAIR10/blastsets/). Positional distributions were calculated as described in Uygur et al., 2017. In brief, min-pCREs were mapped to 100 bp bins of 1000-TSS-500 and 500-TTS-1000 and to whole exons and introns. For comparison, min-pCREs were mapped to randomized versions of the sequences. Randomization was performed within each 100-bp bin and in each exon or intron, respectively, to maintain nucleotide composition and therefore GC content. Positional distribution was calculated as log2FC of number of observed mappings divided by number of randomly expected mappings (log2FC[observed/expected]).

**pCRE Coordinate Overlap with CNS Coordinates**

All 615 min-pCRE PWMs were mapped to the putative promoter region (1 kb upstream of TSS) of 11 Fe response genes (described above). The min-pCRE coordinates were then compared with the coordinates reported as CNS across nine species in the Brassicaceae family (Haudry et al., 2013) downloaded from the UCSC Genomics Bioinformatics Web site (http://mustang.biolo.montg.rr.ca:8885/download/A.thaliana/gff/AT_CNS.gff; Supplemental Table S8).

**Accession Numbers**

Sequence data from genes mentioned in this article can be found in The Arabidopsis Information Resource (TAIR) and in The Rice Annotation Project (RAP) under the accession numbers AB3 (AT3G24650), bHLH34 (AT3G23210), bHLH65 (AT5G56970), bHLH104 (AT4G14410), bHLH113 (AT3G15100), bHLH121 (AT3G19860), bZIP3 (AT5G15830), bZIP6 (AT5G23370), EIN3 (AT5G20770), F6/H1 (AT1G31600), FAR1 (AT4G15900), FYH3 (AT2G22170), FIT (AT2G28160), FRO2 (AT1G01580), FUS3 (AT5G26790), IDE1 (AT5G24600), IDE2 (AT5G0426200), ILR3 (AT5G34680), IRO2 (AT5G0932800), IRO3 (AT5G0973000), IRT2 (AT4G19680), LEC2 (AT2G8300), MAPK16 (AT2G6690), PDR9 (AT5G3480), PR11 (AT5G013890), PR2 (AT5G055400), PR5 (AT5G0116600), PR4 (AT5G053800), PT1 (AT3G22150), TCP20 (AT3G22701), TLT1 (AT1G22140), TLT5 (AT3G25190), WRKY11 (AT3G23550), YSL2 (AT5G0649000). For accession numbers of Arabidopsis genes not listed here, please see Supplemental Table S1.

**Supplemental Data**

The following supplemental materials are available.

**Supplemental Figure S1.** Complete GO enrichment analysis of -Fe responsive genes.

**Supplemental Figure S2.** GO terms and -Fe GS gene enrichments of the defined coexpression clusters containing up-/down-regulated genes, and mean similarity within designated superclusters of up-regulated and up-/down-regulated genes.
Supplemental Figure S3. Expression plots of all well-performing coexpression clusters.

Supplemental Figure S4. Coexpression cluster RF model performance.

Supplemental Figure S5. Different sets of pCREs predict gene expression at different time points of -Fe treatment or in different -Fe treated root zones.

Supplemental Figure S6. Comparison of the pCRE abundance and importance in GS-enriched clusters vs. nonenriched clusters and hierarchical clustering of freq-pCRE sequences.

Supplemental Figure S7. Sequence similarity for freq-pCRE from nonenriched clusters and the best matching known TFBM (DAP-seq or CIS-BP).

Supplemental Figure S8. Example of a nonlinear regression curve to determine the minimum set of pCREs (min-pCREs) for a coexpression cluster and min-pCREs highlighted in two more example clusters.

Supplemental Figure S9. Positional distribution plots of all 615 min-pCREs.

Supplemental Figure S10. High-resolution image of Figure 6.

Supplemental Table S1. Robust -Fe-responsive gold standard (GS) genes.

Supplemental Table S2. All generated coexpression clusters. Table shows the input expression data combinations, the algorithm and parameters used for clustering, significant enrichment in GS genes and more detailed enrichment in FIT-dependent genes (= FIT), FIT-independent genes (= non-FIT) or both (= mixed).

Supplemental Table S3. Well-performing models (F1 ≥ 0.7) built from coexpressed -Fe-responsive genes ("specific") applied to all -Fe-responsive genes as bulk ("general").

Supplemental Table S4. All pCREs (n = 5,639) identified in well-performing clusters (GS-enriched and nonenriched).

Supplemental Table S5. Most relevant pCREs (freq-pCREs and min-pCREs) and their similarity to DAP-seq and CIS-BP TFBMs.

Supplemental Table S6. Proportional association of min-pCRE to FIT-dependent, FIT-independent or both cluster categories.

Supplemental Table S7. Total counts of incidences in which each min-pCREs (n = 615) located in the putative promoter of a gene was also identified as predictive for the cluster containing the gene.

Supplemental Table S8. Overlap of min-pCRE coordinates with conserved noncoding sequences (CNS) in Brassicaceae.

Supplemental Table S9. P- and q-values of complete GO enrichment analysis of -Fe-responsive genes (Supplemental Fig S1).

Supplemental Table S10. DAP-seq and CIS-BP TFBMs used in this study.

Supplemental Table S11. TF family 95th percentiles of within, between and random PCC thresholds determining the pCRE-TFBM similarity.

ACKNOWLEDGMENTS

We thank Sahra Uygun, Bethany Moore, Nicholas Panchy, and Peipei Wang for providing scripts and help with programming. B.S. is a member of the international graduate school iGRAD-Plant, Düsseldorf.

Received August 19, 2019; accepted December 22, 2019; published January 14, 2020.

LITERATURE CITED

Ali MA, Azeem F, Nawaz MA, Aacet T, Abbas A, Imran MQ, Shah KH, Rehman HM, Chung G, Yang SH, Bohlmann H (2018) Transcription factors WRKY11 and WRKY17 are involved in abiotic stress responses in Arabidopsis. J Plant Physiol 226: 12–21

Andriankaja ME, Danisman S, Mignolet-Spruyt LF, Claey’s H, Kochanek I, Vermeersch M, De Milde L, De Bodt S, Storme V, Skirycz A, et al (2014) Transcriptional coordination between leaf cell differentiation and chloroplast development established by TCP20 and the subgroup Ib bHLH transcription factors. Plant Mol Biol 85: 233–245

Assunção AG, Herrero E, Lin YF, Huettel B, Talukdar S, Smacznik C, Immink RG, van Eldik M, Fiers M, Schat H, Aarts MG (2010) Arabidopsis thaliana transcription factors bZIP19 and bZIP23 regulate the adaptation to zinc deficiency. Proc Natl Acad Sci USA 107: 10296–10301

Bailey PC, Martin C, Toledo-Ortiz G, Quail PH, Huq E, Heim MA, Jakoby M, Werber M, Weisshaar S (2003) Update on the basic helix-loop-helix transcription factor gene family in Arabidopsis thaliana. Plant Cell 15: 2497–2502

Balk J, Schaadler TA (2014) Iron cofactor assembly in plants. Annu Rev Plant Biol 65: 125–153

Bemer M, van Dijk ADJ, Immink RGH, Angenent GC (2017) Cross-family transcription factor interactions: an additional layer of gene regulation. Trends Plant Sci 22: 66–80

Benjamini Y, Hochberg Y (1995) Controlling the false discovery rate: A practical and powerful approach to multiple testing. J R Stat Soc B 57: 12

Berardini TZ, Mundodi S, Reiser L, Huala E, Garcia-Hernandez M, Zhang P, Mueller LA, Yoon J, Doyle A, Lander G, et al (2004) Functional annotation of the Arabidopsis genome using controlled vocabularies. Plant Physiol 138: 745–755

Blaby-Haas CE, Merchant SS (2013) Iron sparing and recycling in a compartmentalized cell. Curr Opin Microbiol 16: 67–73

Bourdages C, Fatouhi A, Legrand B, Dubreucq B (2017) Regulation and evolution of the interaction of the seed B3 transcription factors with NF-Y subunits. Biochim Biophys Acta Gen Rech Mech 1860: 1069–1078

Breiman L, Friedman JM, Olshen RA, Stone CJ (1984) Classification and regression trees. Wadsworth International Group, Belmont, CA

Briat JF, Rouached H, Tissot N, Gaymard F, Dubos C (2015) Integration of P, Fe, and Zn nutrition signals in Arabidopsis thaliana: Potential involvement of PHOSPHATE STARVATION RESPONSE 1 (PHR1). Front Plant Sci 6: 290

Brumberova T, Bauer P, Ivanov R (2015) Molecular mechanisms governing Arabidopsis iron uptake. Trends Plant Sci 20: 124–133

Choi HS, Seo M, Cho HT (2018) Two TPL-binding motifs of ARF2 are involved in repression of auxin responses. Front Plant Sci 9: 372

Cock PJ A, Antao T, Chang JT, Chapman BA, Cox CJ, Dalke A, Friedberg I, Hamelryck T, Kauff F, Wilczynski B, de Hoon MJL (2004) The Gene ontology database and information resource. Nucleic Acids Res 32: D429–D433

Colangelo EP, Guérinot ML (2004) The essential basic helix-loop-helix protein FTII is required for the iron deficiency response. Plant Cell 16: 3400–3412

Curie C, Mari S (2017) New routes for plant iron mining. New Phytol 214: 521–525

Davie JM, Wild M, Regnault T, Baumberger N, Eisler H, Genschik P, Achard P (2014) Class I TCP-DELLA interactions in inflorescence shoot apex determine plant height. Curr Biol 24: 1923–1928

Deng W, Zhang K, Busov V, Wei H (2017) Recursve random forest algorithm for constructing multilayered hierarchical gene regulatory networks that govern biological pathways. PLoS One 12: e0171522

Denneny JR, Long TA, Wang YJ, Jung JW, Mace D, Pointer S, Barron C, Brady SM, Schiefelbein J, Benfey PN (2008) Cell identity mediates the response of Arabidopsis roots to abiotic stress. Science 320: 942–945

Fourcroy P, Sísó-Terraza P, Sudre D, Saviron M, Rey G, Gaymard F, Abadia A, Abadia J, Alvarez-Fernández A, Briat JF (2014) Involvement of the ABCG34 transporter in secretion of scopoletin and derivatives by Arabidopsis roots in response to iron deficiency. New Phytol 201: 155–167

Gao F, Robe K, Bettembourg M, Navarro N, Rofidal V, Santoni V, Gaymard F, Vignols F, Roschtzardtz H, Izquierdo E, et al (2020) The transcription factor bHLH112 interacts with bHLH1105 (ILR3) and its closest homologs to regulate iron homeostasis in Arabidopsis. Plant Cell 32: 508–524

Garcia MJ, Romera FJ, Stacey MG, Stacey G, Villar E, Alcáñara E, Pérez-Vicente I (2013) Shoot to root communication is necessary to control the expression of iron-acquisition genes in Strategy I plants. Planta 237: 65–75

Gautier L, Cope L, Bolstad BM, Irizarry RA (2004) Affy—analysis of Affymetrix GeneChip data at the probe level. Bioinformatics 20: 307–315

Godha H, Sasaki E, Akiyama K, Maruyama-Nakashita A, Nakabayashi K, Li W, Ogawa M, Yamauchi Y, Preston J, Aoki K, et al (2008) The AtGenExpress hormone and chemical treatment data set: Experimental design, data evaluation, model data analysis and data access. Plant J 55: 526–542
Abiradopsis Fe Deficiency Response pCNs

Gollhofer J, Schläwicke C, Jungnick N, Schmidt W, Buckhout TJ (2011) Members of a small family of nodulin-like genes are regulated under iron deficiency in roots of Arabidopsis thaliana. Plant Physiol Biochem 49: 557–564

Gollhofer J, Timofeev R, Lan P, Schmidt W, Buckhout TJ (2014) Vacuolar-Iron-Transporter-Like proteins mediate iron homeostasis in Arabidopsis. PLoS One 9: e104488

Gordon DB, Nekludova L, McCallum S, Fraenkel E (2012) Fitting into the harsh reality: Members of a small family of nodulin-like genes are regulated under iron deficiency in roots of Arabidopsis thaliana. Plant Physiol 159: 276–740.e10

Hantizis LJ, Kroh GE, Jahn CE, Campbell M, Peers G, Pilon M, Ravet K (2018) A program for iron economy during deficiency targets specific Fe proteins. Plant Physiol 176: 596–610

Hartigan JA, Wong MA (1979) A k-means clustering algorithm. J R Stat Soc Ser C Appl Stat 28: 9

Haudry A, Platts AE, Vello E, Hoen DR, Leclercq M, Williamson RJ, Forczek E, Joly-Lopez Z, Steffen JC, Hassouri KM, et al (2013) An atlas of Arabidopsis thaliana noncoding sequences provides insights into cruciferous regulatory regions. Nat Genet 45: 891–898

Haydon MJ, Kawachi M, Wirtz M, Hillmer S, Hell R, Krämer U (2012) Vacular nicotianamine has critical and distinct roles under iron deficiency and for zinc sequestration in Arabidopsis. Plant Cell 24: 724–737

Heydrickx KS, Van de Velde J, Wang C, Weigel D, Vandepoele K (2014) A functional and evolutionary perspective on transcription factor binding in Arabidopsis thaliana. Plant Cell 26: 3894–3910

Huang S, Kim SA, Clough CR (2013) Reciprocal interaction of the circadian clock with the iron homeostasis network in Arabidopsis. Plant Physiol 161: 893–903

Hsiep WP, Hsieh HL, Wu SH (2012) Arabidopsis bZIP16 transcription factor integrates light and hormone signaling pathways to regulate early seedling development. Plant Cell 24: 3997–4011

Ito H, Gray WM (2006) A gain-of-function mutation in the Arabidopsis plant disease resistance transporter PDR9 confers resistance to auxin herbicides. Plant Physiol 142: 63–74

Ivanov R, Brumberova T, Bauer P (2012) Fitting into the harsh reality: Regulation of iron-deficiency responses in dicotyledonous plants. Mol Plant 5: 27–42

Jakoby M, Wang HY, Reidt W, Weisshaar B, Bauer P (2004) FRU (bHLH) requires for induction of iron mobilization genes in Arabidopsis thaliana. FEMS Lett 577: 99–105

Jeong J, Merkovich A, Clyne M, Connolly EL (2010) Scopoletin is biosynthesized via ortho-kinase in Arabidopsis thaliana. Plant J 62: 451–461

Jones J, Merkovich A, Clyne M, Connolly EL (2017) Directing iron transport in dicots: Regulation of iron acquisition and translocation. Curr Opin Plant Biol 39: 106–113

Jones E, Oliphant T, Peterson P, et al (2001) SciPy: Open source scientific tools for Python. In

Kai K, Mizutani M, Kawamura N, Yamamoto R, Tamai M, Yamaguchi H, Sakata K, Shimizu B (2008) Scopoletin is biosynthesized via ortho-hydroxylation of feraloyl CoA by a 2-oxoglutarate-dependent dioxygenase in Arabidopsis thaliana. Plant J 55: 989–999

Kakai Y, Ogo Y, Itai RN, Kobayashi T, Yamakawa T, Nakanishi H, Nishizawa NK (2013) Development of a novel prediction method of cis-elements to hypothesize collaborative functions of cis-element pairs in iron-deficient rice. (N rice) Y 6: 22

Khan MA, Castro-Guerrero NA, McInturf SA, Nguyen NT, Dame AN, Wang J, Bindereif RK, Joshi T, Jiruziak SS, Nusinowa D, Mendoza-Cazall D (2018) Changes in iron availability in Arabidopsis are rapidly sensed in the leaf vasculature and impaired sensing leads to opposite transcriptional programs in leaves and roots. Plant Cell Environ 41: 2263–2276

Kilian J, Whitehead D, Horak J, Wanke D, Weinl S, Batistic O, D’Angelo C, Börnberg-Eber E, Kudla J, Harker K (2007) The AtGenExpress global stress expression data set: Protocols, evaluation and model data analysis of UV-B light, drought and cold stress responses. Plant J 50: 347–363

Kim SA, LaCroix IS, Gerber SA, Guerino ML (2019) The iron deficiency response in Arabidopsis thaliana requires the phosphorylated transcription factor URI. Proc Natl Acad Sci USA 116: 24933–24942

Kobayashi T, Itai RN, Ogo Y, Kakei Y, Nakanishi H, Takahashi M, Nishizawa NK (2009) The rice transcription factor cDEF1 is essential for the early response to iron deficiency, and induces vegetative expression of late embryogenesis abundant genes. Plant J 60: 948–961

Kobayashi T, Nakayama Y, Itai RN, Nakanishi H, Yoshitara T, Morii S, Nishizawa NK (2003) Identification of novel cis-acting elements, IDE1 and IDE2, of the barley IDS2 gene promoter conferring iron-deficiency, root-specific expression in heterogeneous tobacco plants. Plant J 36: 780–793

Kobayashi T, Ogo Y, Aung MS, Nozoye T, Itai RN, Nakanishi H, Yamakawa T, Nishizawa NK (2010) The spatial expression and regulation of transcription factors IDE1 and IDE2. Ann Bot 105: 1109–1117

Kobayashi T, Ogo Y, Itai RN, Nakanishi H, Takahashi M, Morii S, Nishizawa NK (2007) The transcription factor IDE1 regulates the response to and tolerance of iron deficiency in plants. Proc Natl Acad Sci USA 104: 19150–19155

Kobayashi T, Ozu A, Kobayashi S, An G, Jeon JS, Nishizawa NK (2019) OshBHLH058 and OshBHLH059 transcription factors positively regulate iron deficiency responses in rice. Plant Mol Biol 101: 471–486

Kobayashi T, Suzuki M, Inoue H, Itai RN, Takahashi M, Nakanishi H, Morii S, Nishizawa NK (2005) Expression of iron-acquisition-related genes under iron-deficiency is co-ordinately induced by partially balanced iron-served deficiency responsive elements. J Exp Bot 56: 1305–1316

Koryachko A, Matthiadis A, Muhammad D, Forel J, Brady SM, Ducoste JJ, Tuck J, Long TA, Williams C (2015) Clustering and differential alignment algorithm: Identification of early stage regulators in the Arabidopsis thaliana iron deficiency response. PLoS One 10: e0136591

Liang G, Zhang H, Ai Q, Liang G, Yu D, Li X, Zhang H, Ai Q, Liang G, Yu D, Li X, Zhang H, Ai Q, Liang G, Yu D (2014) Abscisic acid alleviates iron deficiency by promoting root iron utilization and transport from root to shoot in Arabidopsis. Plant Cell Environ 37: 852–863

Li W, Lan P (2017) The understanding of the plant iron deficiency responses in Strategy I plants and the role of ethylene in this process by omic approaches. Front Plant Sci 8: 40

Li W, Schmidt W (2010) A lysine-63-linked ubiquitin chain-forming conjugase, UBC13, promotes the developmental responses to iron deficiency in Arabidopsis roots. Plant J 62: 330–343

Li X, Zhang H, Ai Q, Liang G, Yu D (2016) Two bHLH transcription factors, bHLH34 and bHLH104, regulate iron homeostasis in Arabidopsis thaliana. Plant Physiol 170: 2479–2493

Liang G, Zhang H, Li X, Ai Q, Yu D (2017) bHLH transcription factor bHLH115 regulates iron homeostasis in Arabidopsis thaliana. J Exp Bot 68: 1743–1755

Lingam S, Mohracher J, Brumberova T, Potuschak T, Fink-Strobl C, Blondet E, Genschik P, Bauer P (2011) Interaction between the bHLH transcription factor FIT and ETHYLENE INSENSITIVE/Ethylene INSENSITIVES-LIKE1 reveals molecular linkage between the regulation of iron acquisition and ethylene signaling in Arabidopsis. Plant Cell 23: 1815–1829

Liu J, Xie Y, Wang H, Ma X, Yao W, Wang H (2017) Light and ethylene coordinately regulate the phosphate starvation response through transcriptional regulation of PHOSPHATE STARVATION RESPONSE1. Plant Cell 29: 2269–2284

Long TA, Tsukagoshi H, Busch W, Lahner B, Salt DE, Benfey PN (2010) The bHLH transcription factor POFEYE regulates response to iron deficiency in Arabidopsis roots. Plant Cell 22: 2219–2236

Ma C, Zhang HH, Wang X (2014) Machine learning for Big Data analytics in plants. Trends Plant Sci 19: 798–808

Maechler M, Rousseeuw P, Struyf A, Hubert M, Hornik K (2017) cluster: Cluster Analysis Basics and Extensions. In Ed R package version 2.06

Maui HJ, Pateyron S, Bauer P (2016) Iron homeostasis in Arabidopsis thaliana: Transcriptomic analyses reveal novel FIT-regulated genes, iron deficiency marker genes and functional gene networks. BMC Plant Biol 16: 211

Mangano S, Denita-Juarez SP, Choi HS, Marzol E, Hwang Y, Ranocha P, Velasquez SM, Borassi C, Barbeneri ML, Aptekmann AA, et al (2017) Molecular link between auxin and ROS-mediated polar growth. Proc Natl Acad Sci USA 114: 5289–5294
Salomé PA, Oliva M, Weigel D, Krämer U (2013) Circadian clock adjustment to plant iron status depends on chloroplast and phytocrome function. EMBO J 32: 511–523

Santi S, Schmidt W (2009) Dissecting iron deficiency-induced proton extrusion in Arabidopsis roots. New Phytol 183: 1072–1084

Schmid M, Davison TS, Henz SR, Pape UJ, Demar M, Vinrong M, Schönkopf B, Weigel D, Lohmann JU (2005) A gene expression map of Arabidopsis thaliana development. Nat Genet 37: 501–506

Schmid NR, Giehl RF, Döll S, Mock HP, Strehmel N, Scheel D, Kong X, Hider RC, von Wirén N (2014) Feruloyl-CoA 6′-Hydroxylase1-dependent coumarins mediate iron acquisition from alkaline substrates in Arabidopsis. Plant Physiol 164: 160–172

Schmidt W, Tittel J, Schikora A (2000) Role of hormones in the induction of iron deficiency responses in Arabidopsis roots. Plant Physiol 122: 1199–1118

Schuler M, Keller A, Backes C, Philippark K, Lenhof HP, Bauer P (2011) Transcriptome analysis by GeneTrail revealed regulation of functional categories in response to alterations of iron homeostasis in Arabidopsis thaliana. BMC Plant Biol 11: 87

Schwarz B, Bauer P (2020) FIT, a regulatory hub for iron deficiency and stress signaling in roots, and FIT-dependent and -independent gene signatures. J Exp Bot online 02/12

Selote D, Samira R, Matthiads A, Gillikin JW, Long TA (2015) Iron-binding E3 ligase mediates iron response in plants by targeting basic helix-loop-helix transcription factors. Plant Physiol 167: 273–286

Séguelà M, Briat JF, Vert G, Curie C (2008) Cytokinins negatively regulate the root iron uptake machinery in Arabidopsis through a growth-dependent pathway. Plant J 55: 289–300

Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, Amin N, Schiwkowski B, Iderker T (2003) Cytoscape: A software environment for integrated models of biomolecular interaction networks. Genome Res 13: 2498–2504

Sinclair SA, Senger T, Talke IN, Cobbett CS, Haydon MJ, Krämer U (2018) Systemic upregulation of MTP2- and HMA2-mediated Zn partitioning to the shoot supplements local Zn deficiency responses. Plant Cell 30: 2463–2479

Sivitz A, Grinvalds C, Barberon M, Curie C, Vert G (2011) Proteasome-mediated turnover of the transcriptional activator FIT is required for plant iron-deficiency responses. Plant J 66: 1044–1052

Sivitz AB, Hermand V, Curie C, Vert G (2012) Arabidopsis bHLH100 and bHLH101 control iron homeostasis via a FIT-independent pathway. PLoSOne 7: e4843

Siwinska J, Siatkowska K, Ory A, Grosjean J, Henn A, Bourgaud F, Mhearg AA, Carey M, Lojowska E, Ihnatonivici A (2018) Scopeatin 8-hydroxylase: A novel enzyme involved in coumarin biosynthesis and iron-deficiency responses in Arabidopsis. J Exp Bot 69: 1735–1748

Storey JD (2002) A direct approach to false discovery rates. J R Stat Soc Series B Methodol 64: 479–498

Tang JF, Zhou C, Wang SS, Yuan J, Zhang XS, Su YH (2017) FUSCA3 interacting with LEAFY COTYLEDON2 controls lateral root formation through regulating YUCCA4 gene expression in Arabidopsis thaliana. New Phytol 213: 1740–1754

Tissot N, Robe K, Gao F, Grant-Sant S, Bouchez J, Bellegarde F, Maghiaoui A, Marcelin R, Izquierdo E, Benhamed M, et al (2019) Transcriptomic integration of the responses to iron availability in Arabidopsis by the bHLH factor ILK3. New Phytol 225: 1433–1446

Uygur S, Azodi CB, Shiu SH (2019) Co-regulatory code for predicting plant cell-type transcriptional response to high salinity. Plant Physiol 181: 1739–1751

Uygur S, Peng C, Lehti-Shiu MD, Last RL, Shiu SH (2016) Utility and limitations of using gene expression data to identify functional associations. PLOS Comput Biol 12: e1005244

Uygur S, Seddon AE, Azodi CB, Shiu SH (2017) Predictive models of spatial transcriptional response to high salinity. Plant Physiol 174: 450–464

Vert GA, Briat JF, Curie C (2003) Dual regulation of the Arabidopsis high-affinity root iron uptake system by local and long-distance signals. Plant Physiol 131: 794–801

Wang H, Wang H (2015) Multifaceted roles of FHY3 and FAR1 in light signaling and beyond. Trends Plant Sci 20: 453–461

Wang HY, Klatte M, Jakoby M, Bäumllein H, Weisshaar B, Bauer P (2007) Iron deficiency-mediated stress regulation of four subgroup Ib BHLH genes in Arabidopsis thaliana. Planta 226: 897–908
Wang N, Cui Y, Liu Y, Fan H, Du J, Huang Z, Yuan Y, Wu H, Ling HQ (2013) Requirement and functional redundancy of Ib subgroup bHLH proteins for iron deficiency responses and uptake in Arabidopsis thaliana. Mol Plant 6: 503–513

Wang S, Li L, Ying Y, Wang J, Shao JF, Yamaji N, Whelan J, Ma JF, Shou H (2019) A transcription factor OsbHLH156 regulates Strategy II iron acquisition through localising IRO2 to the nucleus in rice. New Phytol 225: 1247–1260

Weirauch MT, Yang A, Albu M, Cote AG, Montenegro-Montero A, Drewe P, Najafabadi HS, Lambert SA, Mann I, Cook K, et al (2014) Determination and inference of eukaryotic transcription factor sequence specificity. Cell 158: 1431–1443

Yan JY, Li CX, Sun L, Ren JY, Li GX, Ding ZJ, Zheng SJ (2016) A WRKY transcription factor regulates Fe translocation under Fe deficiency. Plant Physiol 171: 2017–2027

Yáñez-Cuna JO, Kvon EZ, Stark A (2013) Deciphering the transcriptional cis-regulatory code. Trends Genet 29: 11–22

Yang JL, Chen WW, Chen LQ, Qin C, Jin CW, Shi YZ, Zheng SJ (2013) The 14-3-3 protein GENERAL REGULATORY FACTOR11 (GRF11) acts downstream of nitric oxide to regulate iron acquisition in Arabidopsis thaliana. New Phytol 197: 815–824

Yu CP, Lin JJ, Li WH (2016) Positional distribution of transcription factor binding sites in Arabidopsis thaliana. Sci Rep 6: 25164

Yuan Y, Wu H, Wang N, Li J, Zhao W, Du J, Wang D, Ling HQ (2008) FIT interacts with AtbHLH38 and AtbHLH39 in regulating iron uptake gene expression for iron homeostasis in Arabidopsis. Cell Res 18: 385–397

Zamioudis C, Hanson J, Pieterve MM (2014) β-Glucosidase BGLU42 is a MYB72-dependent key regulator of rhizobacteria-induced systemic resistance and modulates iron deficiency responses in Arabidopsis roots. New Phytol 204: 368–379

Zhai Z, Gayomba SR, Jung HI, Vimalakumari NK, Piñeros M, Craft E, Rutzke MA, Danku J, Lahner B, Punshon T, et al (2014) OPT3 is a phloem-specific iron transporter that is essential for systemic iron signaling and redistribution of iron and cadmium in Arabidopsis. Plant Cell 26: 2249–2264

Zhang H, Li Y, Po M, Xu P, Liang G, Yu D (2020) Oryza sativa POSITIVE REGULATOR OF IRON DEFICIENCY RESPONSE 2 (OsPRI2) and OsPRI3 are involved in the maintenance of Fe homeostasis. Plant Cell Environ 43: 261–274

Zhang H, Li Y, Yao X, Liang G, Yu D (2017) POSITIVE REGULATOR OF IRON HOMEOSTASIS1, OsPRI1, facilitates iron homeostasis. Plant Physiol 175: 543–554

Zhang J, Liu R, Li M, Feng D, Jin H, Wang P, Liu J, Xiong F, Wang J, Wang HB (2015) The bHLH transcription factor bHLH104 interacts with IAA-LEUCINE RESISTANT3 and modulates iron homeostasis in Arabidopsis. Plant Cell 27: 787–805

Zheng L, Ying Y, Wang L, Wang F, Whelan J, Shou H (2010) Identification of a novel iron regulated basic helix-loop-helix protein involved in Fe homeostasis in Oryza sativa. BMC Plant Biol 10: 166

Zhou Y, Zhang D, An J, Yin H, Fang S, Chu J, Zhao Y, Li J (2018) TCP transcription factors regulate shade avoidance via directly mediating the expression of both PHOTOTRICHous INTERACTING FACTORS and auxin biosynthetic genes. Plant Physiol 176: 1850–1861

Zou C, Sun K, Mackaluso JD, Seddon AE, Jin R, Thomashow MF, Shiu SH (2011) Cis-regulatory code of stress-responsive transcription in Arabidopsis thaliana. Proc Natl Acad Sci USA 108: 14992–14997