Putative *cis*-regulatory elements and protein interaction networks control iron deficiency responses in *Arabidopsis thaliana*

Inaugural-Dissertation

zur Erlangung des Doktorgrades
der Mathematisch-Naturwissenschaftlichen Fakultät
der Heinrich-Heine-Universität Düsseldorf

vorgelegt von

Birte Schwarz
aus Henstedt-Ulzburg

Düsseldorf, Juni 2019

aus dem Institut für Botanik
der Heinrich-Heine-Universität Düsseldorf

Gedruckt mit der Genehmigung der
Mathematisch-Naturwissenschaftlichen Fakultät der
Heinrich-Heine-Universität Düsseldorf

Berichterstatter:

1. Prof. Dr. Petra Bauer

2. Prof. Dr. Georg Groth

Tag der mündlichen Prüfung:

18.09.2019

Eidesstattliche Erklärung

Ich versichere an Eides Statt, dass die Dissertation von mir selbständig und ohne unzulässige fremde Hilfe unter Beachtung der „Grundsätze zur Sicherung guter wissenschaftlicher Praxis an der Heinrich-Heine-Universität Düsseldorf“ erstellt worden ist.

Ich habe die Dissertation in dieser oder ähnlicher Form bisher keiner anderen Fakultät vorgelegt.

Ich habe bisher keine erfolglosen Promotionsversuche unternommen.

Birte Schwarz
Düsseldorf, 03.06.2019

Acknowledgements

First of all, I want to thank Prof. Petra Bauer for giving me the opportunity to work on an exciting project. Thank you Petra for your support and for your enthusiasm and also thank you for giving me the freedom of doing things my way.

I want to thank Prof. Georg Groth for supporting me as my mentor during these years and for valuable input in my committee meetings. Furthermore thanks to my committee members Dr. Marion Eisenhut and Prof. Shin-Han Shiu for helpful and encouraging discussions.

Shin-Han, thank you for a wonderful time at MSU, for fun group activities and for introducing me to computational biology. Thanks to the whole Shiu lab for making my stay so great! Special thanks go to Christina for constant help and advice with the manuscript and beyond, and to Beth, Sahra, Peipei and Nick for helping me through programming struggles.

A huge thanks to the Bauer lab! Dani - it has been a pleasure collaborating with you. Thank you and Anna for having my back, bearing with me and making lab life good. Regina and Maria, thank you for making me feel so welcome in this group from day one. You were missed during my last years here. Thank you, Regina, for all these evenings of scientific and non-scientific discussions and your moral support. Maria, I will be forever grateful for you letting me inherit your apartment. I want to express my great gratitude to Rumen and Tzvetina for advice whenever it was needed. Thanks to “my” students Chris, Christin, Sarah, Theresa, and Shayna, and of course to Elke and Ginte for technical support; this project could not have been done without you! Hansi, Inga, Ksenia, Jannik, and everyone else, you have all helped me in many ways (including respecting my need for peace and quiet during this last year), so thank you very much!

Thank you, Sieglinde, for looking after me, it was great having you down the corridor. Thanks Alex, Jenia, Woogie, Flo and everyone from iGRAD Plant.

Zum Schluss möchte ich all denen danken, die mir nahe stehen. Eure Liebe, Freundschaft und Unterstützung in den ganzen Jahren haben mir den nötigen Schwung gegeben, die Sache durchzuziehen!

Table of Contents

1 Preface III

2 Summary IV

3 Zusammenfassung V

4 Abbreviation List VII

5 Introduction 1

5.1 Iron in humans and plants 1

5.2 Plant Fe acquisition strategies 2

5.2.1 Strategy I 2

5.2.2 Strategy II 2

5.3 Fe homeostasis regulation is dominated by basic helix-loop-helix transcription factors 3

5.4 Co-expression networks and significance of transcriptional control during Fe deficiency responses 5

5.4.1 Known *cis*-regulatory elements of Fe homeostasis 7

5.5 Post-translational control of Fe deficiency responses 8

5.6 Fe sensing and signaling 8

5.6.1 Potential plant Fe sensors 9

5.6.2 Signals for Arabidopsis Fe acquisition 9

5.7 Small proteins and peptides in nutrient signaling 11

6 Thesis Objectives 13

7 Manuscript 1 22

8 Manuscript 2 71

9 Concluding Remarks 136

1 Preface

This thesis is structured in three parts, preceded by a short summary of the thesis in English and German language. The first part contains an introduction into plant iron acquisition, including transcriptional and post-transcriptional regulation as well as iron status sensing and signaling. The part closes with a paragraph presenting the thesis objectives. The second and main part contains two manuscripts:

1. **Putative *cis*-regulatory elements predict iron deficiency responses in Arabidopsis**
   Birte Schwarz, Christina B. Azodi, Shin-Han Shiu, Petra Bauer
   - *in preparation* - online available as preprint (*bioRxiv* DOI: 10.1101/603290)

In this study, we identify more than 100 putative *cis*-regulatory elements (pCREs) which robustly predict different transcriptional iron deficiency responses in Arabidopsis roots. Several of these pCREs resemble Iron Deficiency-responsive Element 1 (IDE1), which had only been characterized in grasses. We provide first evidence for a regulatory function of IDE1-likes in non-grasses: regulating synthesis and extrusion of coumarins, which are essential compounds for iron uptake. Furthermore, we show that many of the remaining pCREs resemble experimentally identified binding sites for transcription factors, and we pinpoint B3, NAC, bZIP and TCP transcription factor families as novel regulators of non-grass iron deficiency responses.

2. **Novel insights into the protein interaction network regulating iron deficiency responses in Arabidopsis roots**
   Birte Schwarz[q], Daniela M. Lichtblau[q], Christopher Endres, Christin Sieberg, Petra Bauer
   [q]Authors contributed equally
   - *in preparation* -

With this study, we introduce the recently discovered small peptide FE-UPTAKE-INDUCING PEPTIDE3 (FEP3) as inhibitor of two E3 ligases and negative regulators of Fe uptake, BRUTUS-LIKE1 (BTSL1) and BTSL2. As motivation for this hypothesis, we show that FEP3 physically interacts with BTSL1 and BTSL2 and that stable Arabidopsis lines over-expressing FEP3 phenocopy a *btsl1-btsl2* double loss-of-function line. We identified the interaction in a comprehensive yeast two-hybrid targeted screen, which also revealed several more novel interactions between proteins acting in iron deficiency responses. The presented protein interactome expands our current understanding of Arabidopsis iron uptake regulation and provides a valuable source for follow-up studies.

The third part contains concluding remarks that put the most important findings of both manuscripts into context.

2 Summary

Iron (Fe) is essential for all plants and animals. Because of its redox properties, Fe acts in all important metabolic processes. In plants, the photosynthetic machinery is a major Fe sink. Lack of Fe results in visible leaf chlorosis, developmental retardation and yield loss. Because soil Fe is often oxidized and precipitated, plants have developed different acquisition strategies. Grasses usually make Fe accessible by releasing mugineic acid-based Fe^{3+} chelators into the soil ("Strategy II"). Non-grasses like Arabidopsis (*Arabidopsis thaliana*) acquire Fe via a multi-step mechanism which involves local soil acidification and Fe reduction, accompanied by release of Fe-chelating coumarins ("Strategy I"). More than 1000 genes respond to Fe deprivation in Arabidopsis, coordinating Fe acquisition, allocation, and related functions. The basic helix-loop-helix (bHLH) transcription factor (TF) FIT (FER-LIKE IRON DEFICIENCY INDUCED TRANSCRIPTION FACTOR) regulates Strategy I Fe acquisition genes. In a co-expression network, FIT target genes form a sub-cluster. FIT-independent genes form another sub-cluster, containing *e.g.* Fe distribution genes and group Ib and IVa *BHLHs*. Group Ib bHLHs are needed in heterodimers with FIT to initiate Fe uptake, demonstrating that both sub-clusters are interconnected.

Relevant *cis*-regulatory elements (CREs) coordinating the co-regulation are largely unknown. Using a machine learning algorithm we modeled Fe deficiency-responses in Arabidopsis roots and identified more than 100 putative CREs (pCREs) which could robustly predict Fe deficiency-induced up-regulation. Among them was Strategy II CRE IDE1 (Iron Deficiency-responsive Element1), which we put forward as regulator of coumarin synthesis during Strategy I. Large-scale *in vitro* TF-DNA binding data revealed expected bHLH binding sites among the pCREs, and also binding sites for B3, NAC, bZIP and TCP TFs, which we thus propose as novel upstream regulators.

Another puzzle is how the co-regulated Fe deficiency-responsive genes work together at the protein level, especially with regard to the yet unknown mechanism of Fe deficiency sensing and signaling. With targeted yeast two-hybrid assays we screened for novel protein-protein interactions. An emerging interaction network connects amongst others three homologous E3 ligases and negative regulators of Fe uptake, BRUTUS (BTS), BTS-LIKE1 (BTSL1) and BTSL2 with individual sets of bHLHs and with a small peptide, FEP3 (FE-UPTAKE-INDUCING PEPTIDE3). FEP3 is a potential phloem-mobile signal, while BTS/BTSL1/BTSL2 are potential Fe sensors. Because FEP3 physically interacted with BTSL1/2 and FEP3 over-expression lines phenocopied a *btsl1-btsl2* double loss-of-function line, we introduce FEP3 as inhibitor of BTSL1/2 to de-repress Fe uptake.

With the two studies, this thesis brings to light novel elements of transcriptional regulation and uncovers novel links between proteins which are being discussed in the field as promising candidates of Fe deficiency sensing and signaling.

3 Zusammenfassung

Eisen ist essentiell für Pflanzen und Tiere. Als Katalysator von Redox-Reaktionen ist Eisen an allen wichtigen Stoffwechselprozessen beteiligt. Ein Großteil des Eisens in Pflanzen wird vom Photosyntheseapparat benötigt. Eisenmangel äußert sich unter anderem in Blattchlorose und führt langfristig zu vermindertem Wachstum und Ertragsausfällen. Weil Eisen im Boden oft oxidiert und als Präzipitat vorliegt, haben Pflanzen spezielle Eisenaufnahmestrategien. Gräser geben Mugineinsäure-basierte Chelatoren in den Boden ab, welche Eisen(III) komplexieren und verfügbar machen („Strategie II“). Alle anderen Bedecktsamer, unter ihnen Arabidopsis, machen Eisen über einen mehrschrittigen Mechanismus verfügbar, einschließlich Senken des Rhizosphären-pH Werts und Reduktion von Eisen an der Epidermis („Strategie I“). Unterstützend werden Cumarine (Eisenchelatoren) in die Rhizosphäre abgegeben. Mehr als 1000 Gene sind in *Arabidopsis thaliana* eisenmangelreguliert. Sie koordinieren unter anderem Eisenaufnahme, Weitertransport und viele weitere damit zusammenhängende Prozesse. Der basic helix-loop-helix (bHLH) Transkriptionsfaktor (TF) FIT (FER-LIKE IRON DEFICIENCY INDUCED TRANSCRIPTION FACTOR) reguliert die Gene, die an der Strategie I Eisenaufnahme beteiligt sind. Im Koexpressions-Netzwerk sind die FIT-abhängigen Gene als Gruppe klar von der Gruppe FIT-unabhängiger Gene abgegrenzt. Letztere sind unter anderem am Weitertransport von Eisen beteiligt und mehrere von ihnen sind ebenfalls *BHLHs*. Untergruppe Ib bHLHs sind als Dimere mit FIT für die Eisenaufnahme notwendig. Dies zeigt, dass die Gruppen FIT-abhängiger und FIT-unabhängiger Gene teilweise voneinander abhängen.

Cis-regulatorische Elemente (CREs) welche diese Koregulation koordinieren sind kaum bekannt. Unter Anwendung von maschinellem Lernen wurde die transkriptionelle Eisenmangelantwort in Arabidopsiswurzeln modelliert. Dadurch wurden mehr als 100 putative CREs (pCREs) identifiziert, die zuverlässig eisenmangelinduzierte Genexpression vorhersagen konnten. Unter anderem wurde das Strategie II-CRE IDE1 (Iron Deficiency-responsive Element1) identifiziert, welches hier mit der Regulation von der Cumarinsynthese während Strategie I in Verbindung gebracht werden konnte. Vergleiche mit *large-scale in vitro* TF-DNA-Bindungsdaten zeigten wie erwartet, dass einige der pCREs bHLH-Bindestellen entsprachen. Weitere pCREs entsprachen Bindestellen für B3, NAC, bZIP und TCP TFs, welche wir daher als vielversprechende Kandidaten vorschlagen, die oberhalb der bekannten transkriptional induzierten Gene agieren.

Es ist immer noch relativ wenig darüber bekannt, wie die unter Eisenmangel koregulierten Gene auf Proteinebene in Beziehung zueinander stehen. Dies besser zu verstehen ist besonders interessant im Hinblick darauf, dass man noch immer nicht weiß, wie Eisenmangel von Pflanzen wahrgenommen und wie diese Information weitergeleitet wird. Mit einer Reihe von Hefe Zwei-Hybrid Experimenten haben wir nach neuen Protein-Protein-

Interaktionen gesucht. Aus den Ergebnissen formiert sich ein Interaktionsnetzwerk, welches unter anderem drei homologe E3 Ligasen und negative Regulatoren der Eisenaufnahme, BRUTUS (BTS), BTS-LIKE1 (BTSL1) und BTSL2 mit je eigenen Kombinationen an bHLH TFs in Verbindung bringt. Dazu gehört außerdem ein kleines Peptid, FEP3 (FE-UPTAKE-INDUCING PEPTIDE3), welches ein potentielles Phloem-mobiles Signal sein könnte. BTS/BTSL1/BTSL2 sind potentielle Eisensensoren. Weil FEP3 mit BTSL1/2 interagieren konnte und weil FEP3 Überexpressionslinien bezüglich des molekularen Phänotyps stark einer *btsl1-btsl2* doppel-*knock-out* Linie ähnelte, wurde ein hypothetisches Modell aufgestellt, in dem FEP3 die Funktion von BTSL1/2 inhibiert und damit die negative Regulation der Eisenaufnahme aufhebt.

Diese Arbeit, mit den beiden enthaltenen Studien, bringt neue Elemente der transkriptionalen Regulation ans Licht und deckt neue Verbindungen zwischen Proteinen auf, welche im Moment als vielversprechende Kandidaten für eine Beteiligung am Eisenmangel-Signalweg diskutiert werden.

4 Abbreviation List

bHLH	Basic helix-loop-helix
BTS / BTSL	BRUTUS / BRUTUS-LIKE
CRE / pCRE	*cis*-regulatory element / putative *cis*-regulatory element
Fe / FeS	Iron / Iron-sulfur
FEP	FE-UPTAKE-INDUCING PEPTIDE
FIT	FER-LIKE IRON DEFICIENCY-INDUCED TRANSCRIPTION FACTOR
HHE	Hemerythrin/HHE cation binding motif
IDE1	Iron Deficiency-responsive Element 1
IDEF	IRON DEFICIENCY-RESPONSIVE ELEMENT BINDING FACTOR
IRT1	IRON REGULATED TRANSPORTER1
MA	Mugineic acid
TF	Transcription factor

5 Introduction

5.1 Iron in humans and plants

Iron (Fe) as essential micronutrient is crucial for all living organisms. As important cofactor of many enzymes, Fe plays a role in numerous central functions of cell metabolism, such as respiration and DNA replication and repair (Hentze *et al.*, 2004; Netz *et al.*, 2012; Blaby-Haas and Merchant, 2013; Zhang, 2014). In humans, most Fe (~80%) is bound in hemoglobin, which is needed for the oxygen transport function of red blood cells (Hentze *et al.*, 2004). Fe deficiency is the most common micronutrient deficiency and the primary cause for anemia, which is a major health risk especially for children and pregnant women (Hentze *et al.*, 2004; Petry *et al.*, 2016). An estimated 25% of the world's population suffer from anemia (McLean *et al.*, 2009), with highest prevalence in regions in which diets mainly consist of starchy staple crops low in trace minerals or vitamins, such as wheat, rice, maize or cassava. Biofortification, *i.e.* increasing micronutrient content of crops by breeding or biotechnological approaches, is one major motivation which drives plant Fe metabolism research.

In plants, the photosynthetic machinery is an additional major Fe sink. As central component of heme groups or Fe-sulfur (FeS) clusters in all major complexes and cytochromes, Fe is an important redox catalysator of the electron transport chain (Blaby-Haas and Merchant, 2013; Briat *et al.*, 2015). FeS clusters probably account for the main share of Fe usage (Balk and Pilon, 2011). Alone Arabidopsis (*Arabidopsis thaliana*) is estimated to have over 100 holoproteins binding FeS clusters, which are amongst others involved in chlorophyll biosynthesis, DNA replication/repair, and sulfur and nitrogen assimilation (Eggink *et al.*, 2004; Balk and Pilon, 2011; Fuss *et al.*, 2015).

Plants respond to Fe limitation with several adjustments, including modifications of root architecture and root hair morphology to increase Fe uptake (Giehl *et al.*, 2014), and metabolic modifications. For example, to economize Fe usage, the photosynthetic machinery and mitochondrial functions are adapted, while essential processes, such as DNA replication/repair, are maintained (Vigani *et al.*, 2009; Urzica *et al.*, 2012; Blaby-Haas and Merchant, 2013; Rodríguez-Celma *et al.*, 2013; Balk and Schaedler, 2014; Saito *et al.*, 2014; Hantzis *et al.*, 2018). Fe limitation manifests itself in visible interveinal leaf chlorosis, impaired growth and development, and yield loss (Briat *et al.*, 2015). Although Fe is the fourth most abundant element in the earth's crust, its tendency to oxidize and form insoluble Fe^{3+} oxide or hydroxide precipitates in aerated soils of neutral or basic pH (Lindsay, 1988) makes it difficult for plants to access. An estimated third of the world's soils are neutral or mildly alkaline (Atlas of the Biosphere, Center for Sustainability and the Global Environment, University of Wisconsin, Madison). In light of the growing global population, agriculture will be facing major challenges to keep up with the required amount and quality of food. Therefore, generating

crops that can be cultivated despite unfavorable soil conditions is another motivation for understanding the mechanisms underlying plant Fe uptake, allocation and storage.

5.2 Plant Fe acquisition strategies

Higher plants have developed different strategies to make precipitated soil Fe bioavailable (**Figure 1**). The current model distinguishes reduction-based Strategy I, performed by non-grass species, and chelation-based Strategy II, performed by grass species (Marschner and Römheld, 1994; Kobayashi and Nishizawa, 2012). In brief, during Strategy I, soil Fe is solubilized by lowering the rhizosphere pH, and subsequently taken up in its reduced form. During Strategy II, chelators are secreted to acquire the oxidized form of Fe as complexes. Many key molecular components are different between the two strategies (Kobayashi and Nishizawa, 2012). However, some aspects are common to both, which implies that the mechanisms underlying Strategy I and Strategy II are not mutually exclusive.

5.2.1 Strategy I

Arabidopsis as Strategy I plant acquires Fe in a three-step mechanism, executed by proteins located at the root epidermis plasma membrane. (1) Protons are extruded into the rhizosphere via the P-type H^+-ATPASE2 (AHA2), which solubilizes Fe^{3+} (Santi and Schmidt, 2009). (2) FERRIC REDUCTION OXIDASE2 (FRO2) reduces Fe^{3+} to Fe^{2+} (Robinson *et al.*, 1999) and (3) IRON REGULATED TRANSPORTER1 (IRT1) subsequently imports Fe^{2+} (Eide *et al.*, 1996; Vert *et al.*, 2002). All components are transcriptionally induced under Fe deficient conditions (Santi and Schmidt, 2009). Efficient Fe^{3+} solubilization and reduction is also dependent on phenolic compounds (phenylpropanoid-derived coumarins or riboflavin derivatives) (Schmid *et al.*, 2014; Rajniak *et al.*, 2018; Tsai *et al.*, 2018), which are another type of Fe chelators. They are secreted into the rhizosphere potentially via the ATP-binding-cassette transporter PLEIOTROPIC DRUG RESISTANCE 9 (PDR9) (Fourcroy *et al.*, 2014). Chelation during Fe acquisition is thus common to both Strategy I and II.

5.2.2 Strategy II

Strategy II plants, including many important crops, such as rice (*Oryza sativa*), barley (*Hordeum vulgare*) and maize (*Zea mays*), synthesize different types of mugineic acid (MA) family phytosiderophores in response to Fe deficiency and release them into the rhizosphere through TRANSPORTER OF MUGINEIC ACID1 (TOM1) where they complex and solubilize Fe^{3+} (Takagi *et al.*, 1984; Römheld and Marschner, 1986; Nozoye *et al.*, 2011). Instead of being reduced, Fe^{3+} is imported as Fe^{3+}-MA chelate into the root via YELLOW STRIPE1 (YS1) and YELLOW STRIPE1-LIKE (YSL) transporters (Curie *et al.*, 2001; Inoue *et al.*, 2009). Interestingly, rice and barley possess IRT1 orthologues which are transcriptionally induced by Fe deficiency and capable of transporting Fe^{2+}, suggesting that some grasses also take up Fe in its reduced form (Ishimaru *et al.*, 2006; Pedas *et al.*, 2008). Since Fe exists mainly in its

reduced form when dissolved in water, rice using IRT1 was suggested as an adaptation to the submerged cultivation conditions (Ishimaru *et al.*, 2006). In accordance with that, the barley IRT1 orthologue was recently shown to transport manganese, but not Fe, *in planta* (Long *et al.*, 2018).

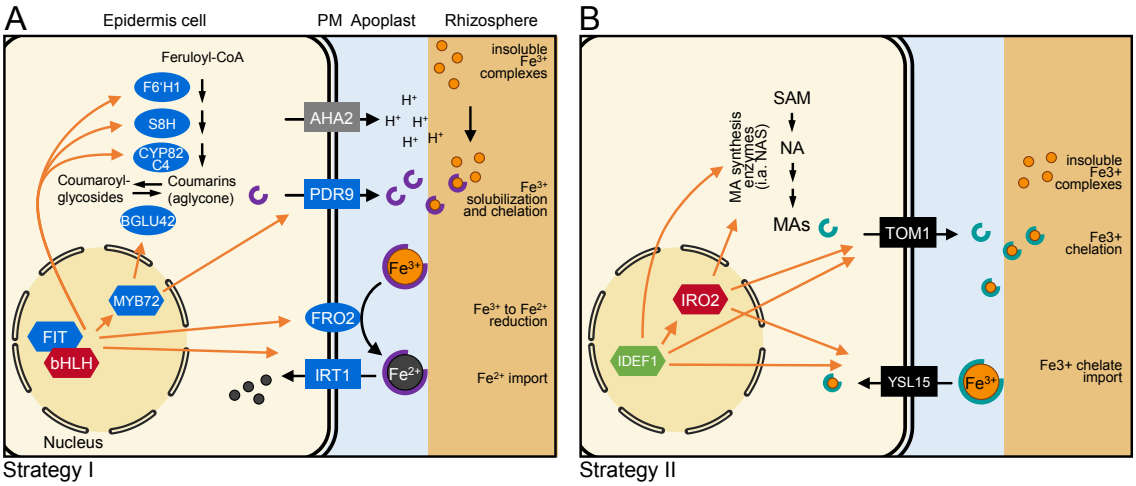


Figure 1. Models of Fe acquisition strategies and the most important transcriptional regulators in non-grass and grass angiosperms. A: Strategy I Fe acquisition in Arabidopsis. Protons are released by AHA2 to acidify the rhizosphere, which solubilizes Fe^{3+}. Fe^{3+} is chelated and kept soluble by phenylpropanoid-derived coumarins, which are secreted via ABC transporter PDR9. FRO2 reduces chelated Fe^{3+} to Fe^{2+}, which is imported by IRT1 into the root. Genes of the Fe uptake machinery and of the coumarin pathway enzymes are activated by FIT, upon forming heterodimers with other TFs from bHLH group Ib. **B:** Strategy II Fe acquisition in rice. MA family phytosiderophores are released into the rhizosphere by TOM1, where they chelate Fe^{3+}. Fe^{3+}-MA complexes are imported into the root via YSL15. IDEF1 directly induces IRO2 (an Arabidopsis Ib bHLH homolog) and both directly and indirectly induce Fe acquisition genes, including MA synthesis enzymes. F6'H1: FERULOYL-COA 6'-HYDROXYLASE; S8H: SCOPOLETIN 8-HYDROXYLASE; CYP82C4: CYTOCHROME P450 family protein; BGLU42: BETA GLUCOSIDASE42; IDEF1: IRON DEFICIENCY-RESPONSIVE ELEMENT BINDING FACTOR1; IRO2: IRON-RELATED TRANSCRIPTION FACTOR2; PM: plasma membrane; SAM: S-adenosyl-L-methionine; NA: nicotianamine; NAS: nicotianamine synthase; MA: mugineic acid; orange arrows: transcriptional induction; black arrows: Fe movement or reduction; blue proteins: FIT-dependent; red proteins: bHLH homologs. Literature: Kobayashi and Nishizawa, 2012; Zamioudis *et al.*, 2014; Brumbarova *et al.*, 2015; Sisó-Terraza *et al.*, 2016; Rajniak *et al.*, 2018; Stringlis *et al.*, 2018.

5.3 Fe homeostasis regulation is dominated by basic helix-loop-helix transcription factors

Most transcriptional regulators of Fe homeostasis that have been identified to date in both grasses and non-grasses belong to the basic helix-loop-helix (bHLH) family of transcription factors (TFs) (Gao *et al.*, 2019). As part of a regulatory cascade in roots (**Figure 2**), group IVc bHLHs (AtbHLH34/104/105/115; OsPRI1 (POSITIVE REGULATOR OF IRON HOMEOSTASIS1)) transcriptionally induce group Ib bHLHs (AtbHLH38/39/100/101; OsIRO2 (IRON-RELATED TRANSCRIPTION FACTOR2)) and group IVb bHLH AtPYE (OsIRO3) (Zhang *et al.*, 2015; Li *et al.*, 2016; Liang *et al.*, 2017; Zhang *et al.*, 2017), which then control Fe uptake and distribution genes (Ogo *et al.*, 2007; Yuan *et al.*, 2008; Long *et al.*, 2010; Zheng *et al.*, 2010; Sivitz *et al.*, 2012; Wang *et al.*, 2013). That this model may be too simplistic is indicated by recent studies which provide evidence for regulatory feedback loops going on between bHLH105 (IAA-LEUCINE RESISTANT3 (ILR3)) and PYE (Samira *et al.*,

2018; Tissot *et al.*, 2019). IVc bHLHs can dimerize among each other and with PYE (Long *et al.*, 2010; Selote *et al.*, 2015; Zhang *et al.*, 2015; Li *et al.*, 2016). Therefore, the system holds a high potential of flexibility, which was recently demonstrated with ILR3 acting as activator when dimerizing with other IVc bHLHs, and as repressor when dimerizing with PYE (Tissot *et al.*, 2019). Despite of what is already known, the regulation and interplay of the above genes is not fully understood, not least because cell type specificity has not been taken into account in the regulatory cascade model. Additionally, AtFEP1 (FE-UPTAKE-INDUCING PEPTIDE1, a.k.a. IRON MAN3 (IMA3)) and its homolog AtFEP3/IMA1, putative small peptides, act as positive regulators of Fe uptake upstream of Ib bHLHs (Grillet *et al.*, 2018; Hirayama *et al.*, 2018). FEPs/IMAs will be introduced in more detail in a later section of the introduction.

Ib bHLHs need to heterodimerize with the group IIIa bHLH FIT (FER-LIKE IRON DEFICIENCY-INDUCED TRANSCRIPTION FACTOR) to activate Fe uptake genes (Yuan *et al.*, 2008; Wang *et al.*, 2013) (**Figure 1A**, **Figure 2**). Because FIT is required for Fe uptake, it is considered the "master regulator" of Strategy I Fe acquisition (Colangelo and Guerinot, 2004; Jakoby *et al.*, 2004; Mai *et al.*, 2016). A FIT orthologue in grasses has not been identified. Instead, two TFs named IRON DEFICIENCY-RESPONSIVE ELEMENT BINDING FACTOR1 (IDEF1; **Figure 1B**) and IDEF2, belonging to the ABI3VP1 and NAC TF families, respectively, control Fe deficiency response genes directly and also via IRO2 and IRO3 (Kobayashi *et al.*, 2007; Ogo *et al.*, 2008; Kobayashi *et al.*, 2010; Kobayashi *et al.*, 2014). The upstream response cascade is controlled by E3 ligases (AtBTS (BRUTUS); OsHRZ1/2 (HEMERYTHRIN MOTIF-CONTAINING REALLY INTERESTING NEW GENE- AND ZINC-FINGER PROTEINS)), which target IVc bHLH (AtILR3/bHLH115, OsPRI) proteins for proteasomal degradation (Selote *et al.*, 2015; Zhang *et al.*, 2017).

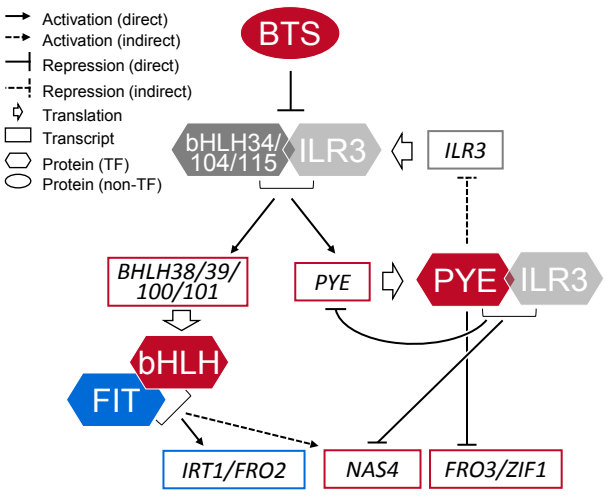


Figure 2. Current model of Fe homeostasis regulation in Arabidopsis. The transcriptional regulatory cascade contains several bHLH TFs which regulate Fe uptake and distribution. Group Ib bHLHs (bHLH38/39/100/101) together with FIT induce Fe uptake machinery genes and are synergistically induced by IVc bHLHs (ILR3/bHLH34/104/115) under Fe deficiency. IVc bHLHs also induce PYE, which is a direct repressor of *NAS4*, *FRO3* and *ZIF1*. *NAS4* is also positively regulated by FIT-dependent MYB TFs (MYB10/72). Regarding *NAS4* repression, PYE acts synergistically with ILR3. In negative feedback loops, PYE represses *ILR3* transcription and PYE-ILR3 dimers repress *PYE* transcription. ILR3 and bHLH115 proteins are targeted for proteasomal degradation by the RING-type E3 ligase BTS. Blue: FIT and FIT-targets; red: Fe deficiency-induction FIT-independent. Literature: Long *et al.*, 2010; Selote *et al.*, 2015; Zhang *et al.*, 2015; Li *et al.*, 2016; Liang *et al.*, 2017; Samira *et al.*, 2018; Tissot *et al.*, 2019.

Despite its importance for plant metabolism, uptake of surplus Fe is harmful to cells due to its tendency to trigger formation of reactive oxygen species (ROS) via the Fenton reaction (Hell and Stephan, 2003). Strictly controlled Fe homeostasis therefore is crucial for the health of the plant. As negative regulators of Fe uptake, BTS/HRZ are thought to balance the amount of Fe intake during Fe acquisition or to rapidly switch off Fe acquisition once Fe levels are restored (Kobayashi *et al.*, 2013; Selote *et al.*, 2015).

Developmental phases and fluctuations of environmental conditions, such as the amount of light available for photosynthesis, are some of the factors which influence Fe requirement. Therefore, Fe acquisition is amongst others tightly interconnected with many plant hormone signaling pathways (reviewed in Brumbarova *et al.*, 2015) and with the circadian clock (Hong *et al.*, 2013; Salomé *et al.*, 2013; Nozoye *et al.*, 2014). In Arabidopsis, several of these factors act on Fe acquisition by modulating FIT protein activity (Brumbarova *et al.*, 2015), showing that Fe deficiency responses are not only transcriptionally, but also heavily post-transcriptionally regulated.

5.4 Co-expression networks and significance of transcriptional control during Fe deficiency responses

Alone in Arabidopsis roots, more than 500 genes respond to Fe deficiency (Dinneny *et al.*, 2008; Mai *et al.*, 2016). These genes encode several transporters, chelators, enzymes and regulatory proteins needed for Fe uptake and distribution, and are also involved in structural modifications in the root, hormone responses and generalized stress responses (Dinneny *et al.*, 2008). Depending on the study, 70 to over 400 Fe deficiency-responsive genes were mis-regulated in *fit-1* loss-of-function plants, suggesting that they are directly or indirectly dependent on FIT (Colangelo and Guerinot, 2004; Mai *et al.*, 2016). FIT-regulated genes typically act in Fe acquisition (**Figure 1A**; *FRO2*, *IRT1*, coumarin synthesis/extrusion genes; Colangelo and Guerinot, 2004; Sivitz *et al.*, 2012; Mai *et al.*, 2016) or in related processes. For example, several metal transporters are up-regulated by FIT, which sequester non-Fe heavy metals that are taken up during Fe acquisition (Vert *et al.*, 2002; Colangelo and Guerinot, 2004; Yang *et al.*, 2010).

In a co-expression network of Fe deficiency-responsive genes, which was generated based on expression data from the ATTED-II data base (Version 8.0; Aoki *et al.*, 2016), FIT-dependent genes form a root-specific sub-network (“FIT target network”; **Figure 3**). Another sub-network, which is not root-specific (Ivanov *et al.*, 2012), contains FIT-independent genes. Because several of these genes were associated with Fe distribution, the sub-network was termed “Fe homeostasis network”. The Fe homeostasis network contains e.g. *NAS4* (*NICOTIANAMINE SYNTHASE4*), catalyzing synthesis of Fe chelator nicotianamine (NA), or *OPT3 (OLIGOPEPTIDE TRANSPORTER3)*, needed for phloem Fe loading, and also Ib *BHLHs* and *PYE*. *NAS4* is one of the genes negatively regulated by PYE (Long *et al.*, 2010)

(**Figure 2**), showing that there are different hierarchical levels within the sub-network. Furthermore, regulatory crosstalk exists between both sub-networks: Ib bHLHs induce Fe acquisition genes as dimers with FIT, and *NAS4* is also positively regulated by partially FIT-dependent R2R3 MYB family TFs (MYB10/72; Palmer *et al.*, 2013). MYB72 (and presumably MYB10) also control synthesis and secretion of phenolic compounds through *BGLU42* (*BETA GLUCOSIDASE42*) and *PDR9* (Palmer *et al.*, 2013; Zamioudis *et al.*, 2014). Thus, MYB10/72 are important regulators of enzymes that act in biosynthesis pathways of different Fe chelators.

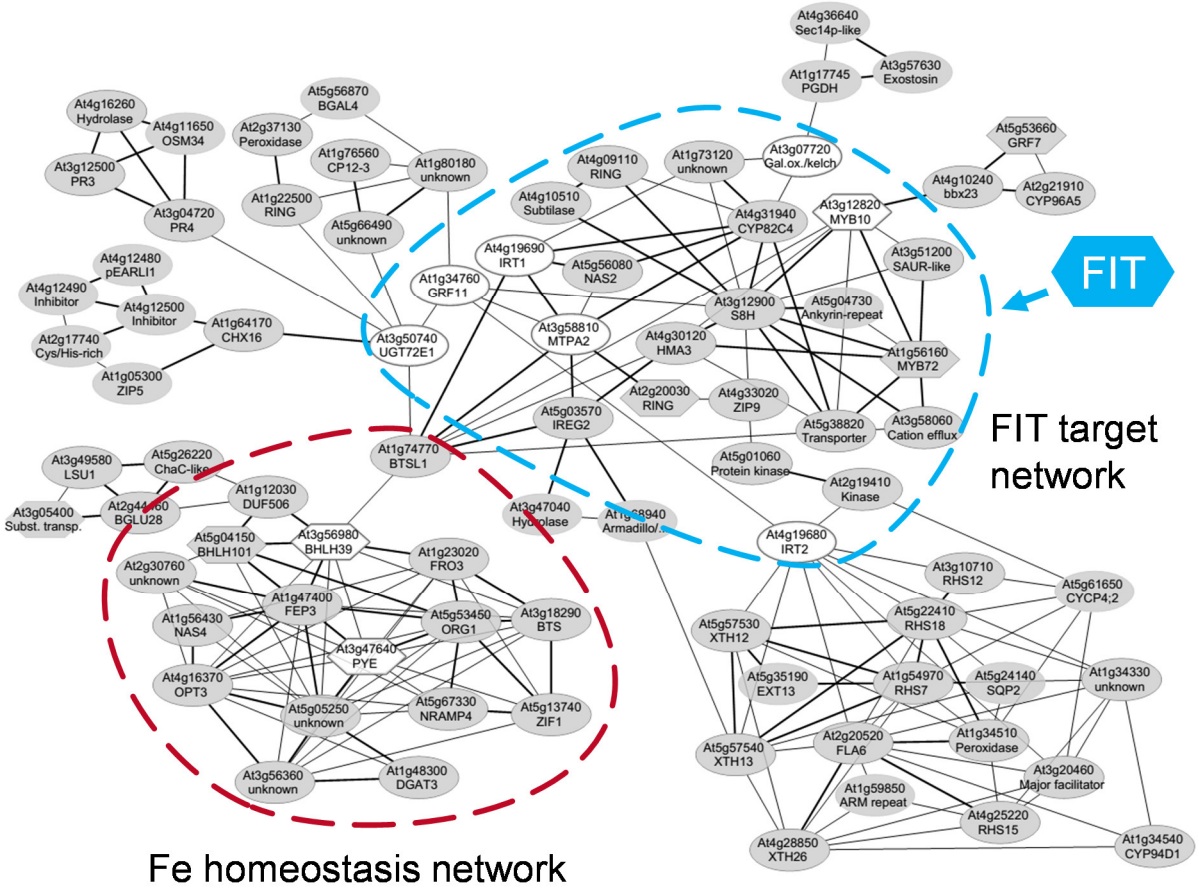


Figure 3. Co-expression network of root-expressed Fe deficiency-responsive genes in Arabidopsis (based on Ivanov *et al.*, 2012). **Blue:** FIT target network containing genes which are directly or indirectly dependent on FIT. **Red:** Fe homeostasis network containing FIT-independent genes, which are root- and shoot-expressed. The root-expressed IRT2 regulon contains several root hair genes. White: Fe deficiency marker genes used as guide genes for construction of the network. Hexagons: Transcription factors. The network was constructed using the ATTED-II tool (ver. 8.0; http://atted.jp/) and processed with Cytoscape. Thickness of the edges corresponds to expression correlation based on the ATTED-II Mutaual Rank scoring system (Obayashi and Kinoshita, 2009).

Besides bHLH family TFs and MYB10/72, only few other TFs have a known function during Fe deficiency-responsive transcriptional regulation (Gao *et al.*, 2019). One example is WRKY46, which is an important regulator of Fe root-to-shoot translocation as repressor of *VITL1* (*VACUOLAR IRON TRANSPORTER1-LIKE1*; Yan *et al.*, 2016), an Fe excess marker which sequesters Fe into the vacuole (Gollhofer *et al.*, 2011). As the most prominent Fe excess markers, *FER* (*FERRITIN*) gene expression is likewise repressed during Fe starvation (Briat

et al., 2010). Ferritins are almost universal multimeric Fe storage proteins (Arosio *et al.*, 2017) and plant ferritins function as Fe excess buffer in plastids and mitochondria (Briat *et al.*, 2010; Vigani *et al.*, 2013). TFs mediating *FER* repression in an Fe status-dependent manner had not been identified, until it was shown recently that ILR3 together with PYE acts as repressor of *FER1*, *FER2* and *FER3* (Tissot *et al.*, 2019).

5.4.1 Known *cis*-regulatory elements of Fe homeostasis

TFs specifically bind to *cis*-regulatory elements (CREs), usually located in promoter regions, in order to modulate gene expression. Considering the number of genes which transcriptionally respond to Fe deficiency (Dinneny *et al.*, 2008), surprisingly few CREs are known to mediate these responses. W-boxes (TTTGAC(C/T)) in the promoter of *VITL1* were identified as binding sites of WRKY46 (Yan *et al.*, 2016). Furthermore, *FER1* repression is controlled by IDRS (Iron Dependent Regulatory Sequence) through yet unknown TFs (Petit *et al.*, 2001) and by a G-box element, which is bound by ILR3 (Tissot *et al.*, 2019). E-boxes (CANNTG) or G-boxes (canonical E-boxes; CACGTG) are typical bHLH binding promoter motifs (Pires and Dolan, 2010). In accordance with the dominance of bHLH TFs during Fe deficiency response regulation, E-/G-boxes have been validated as direct binding sites in several bHLH-regulated genes (FIT target: Yang *et al.*, 2013; FIT/bHLH38/bHLH39 targets: Wild *et al.*, 2016; IVc bHLH targets: Zhang *et al.*, 2015; Liang *et al.*, 2017; ILR3 targets: Tissot *et al.*, 2019).

In grasses, Iron Deficiency-responsive Element1 (IDE1; core motif CATGC) and IDE2 (core motif CA(A/C)G(T/C)(T/C/A)(T/C/A)) have been identified as binding motifs for IDEF1 and IDEF2, respectively (Kobayashi *et al.*, 2003; Kobayashi *et al.*, 2007; Ogo *et al.*, 2008). Many Strategy II Fe acquisition genes, including *OsIRO2* and MA synthesis genes, contain IDE1-like motifs (Kobayashi *et al.*, 2005; Kobayashi *et al.*, 2009). IDE1 is therefore considered a leading CRE of Strategy II. IDE1-like sequences have been identified in several Arabidopsis Fe deficiency response genes (Kobayashi *et al.*, 2005; Murgia *et al.*, 2011), and barley IDE1-driven reporter genes were effectively expressed in tobacco, suggesting that IDE1-recognizing TFs exist in non-grass species (Kobayashi *et al.*, 2003). However, it has not yet been confirmed that IDE1 regulates Strategy I Fe acquisition. Taken together, although individual promoter motifs were identified which likely act as CREs during Fe deficiency, for most of the >500 Fe deficiency-responsive genes the regulators are not known. Thus, a Fe response *cis*-regulatory code, *i.e.* CREs and their combinations which control gene expression (Zou *et al.*, 2011; Yáñez-Cuna *et al.*, 2013), remains largely elusive. Deciphering the code would substantially improve our understanding of how transcriptional Fe deficiency responses are coordinated.

5.5 Post-translational control of Fe deficiency responses

Protein activity and stability can be modulated by numerous different post-translational modifications, such as phosphorylation or ubiquitination. Several Fe deficiency response proteins are controlled at the post-translational level. For example, it was recently shown that FIT protein activity is post-translationally modulated by CIPK11-mediated phosphorylation, which enables it to interact with bHLH39, and to induce downstream transcription (Gratz *et al.*, 2019). The current model describes a system of larger inactive and smaller active pools of FIT, which can be rapidly adjusted to respond to Fe deficiency (Lingam *et al.*, 2011; Meiser *et al.*, 2011; Sivitz *et al.*, 2011; Gratz *et al.*, 2019). Besides FIT activity, proton pumping of AHA2 is also regulated by phosphorylation (Fuglsang *et al.*, 1999; Fuglsang *et al.*, 2007; Fuglsang *et al.*, 2014). It has been suggested recently that IRT1 senses excess non-Fe metals in the local soil, which triggers IRT1 phosphorylation and subsequent degradation as a protective layer of non-Fe metal over-accumulation during Fe acquisition (Dubeaux *et al.*, 2018). IRT1 protein turnover is mediated by ubiquitination via the E3 ligase IRT1 DEGRADATION FACTOR1 (IDF1) (Barberon *et al.*, 2011; Shin *et al.*, 2013; Ivanov *et al.*, 2014; Dubeaux *et al.*, 2018).

E3 ligases are the substrate-recruiting component of the ubiquitin-proteasome system, a universal pathway to rapidly adjust the cellular proteome. Ubiquitination usually leads to protein degradation by the 26S proteasome, but can also lead to change of protein localization or altered function (Mukhopadhyay and Riezman, 2007; Vierstra, 2009; Komander and Rape, 2012; Stone, 2014; Romero-Barrios and Vert, 2018). E3 ligases are key regulators amongst others in plant abiotic stress responses and hormone signaling (Santner and Estelle, 2010; Shu and Yang, 2017). The Arabidopsis genome encodes for more than 1,400 predicted E3 ligases (Vierstra, 2009). This number is substantially larger than predicted in the human (>600) or yeast (80) genome (Li *et al.*, 2008), which reflects the plant's need to constantly adapt to ever-changing environmental cues.

As mentioned earlier, the E3 ligase BTS controls protein levels of at least two IVc bHLHs (ILR3, bHLH115; Selote *et al.*, 2015). BTS has two paralogs in Arabidopsis, which are also up-regulated during Fe deficiency: BTS-LIKE1 (BTSL1) and BTSL2 (Hindt *et al.*, 2017). Their ubiquitination targets are not known, however their existence further increases the share of post-translationally acting regulators during Fe deficiency. BTS (and presumably BTSL1/2) are themselves likely controlled post-translationally. Although the precise mechanism is not yet clear, Fe levels seem to directly modulate their protein stability and function (Selote *et al.*, 2015). This makes BTS (and BTSL1/2) potential Fe sensing proteins.

5.6 Fe sensing and signaling

All living systems studied so far sense the Fe status through a mechanism in which Fe^{2+}, FeS clusters or heme directly bind to regulatory proteins. This binding triggers functional alterations, which links Fe levels mechanistically to downstream regulation (reviewed in

Kobayashi and Nishizawa, 2014). In bacterial systems, the transcriptional repressor Fur (ferric uptake regulation) loses its DNA binding ability when not bound to Fe^{2+}, allowing Fe uptake genes to be active under Fe deficiency (Andrews *et al.*, 2003). In mammals, the positive Fe uptake regulator IRP1 (Iron Regulatory Protein1) assembles an FeS cluster under Fe repletion, which inhibits IRP1 function (Thompson and Bruick, 2012). IRP2 is targeted for proteasomal degradation by a multimeric E3 ligase complex. The Fe sensing subunit of this complex, FBXL5 (F-BoX and Leucine-rich repeat5), binds Fe^{2+} via a hemerythrin/HHE cation binding motif (HHE) domain. This stabilizes the protein and allows the complex to act as repressor of IRP2 (Salahudeen *et al.*, 2009).

5.6.1 Potential plant Fe sensors

Neither the bacterial Fur, nor the mammalian FBXL5-IRP2 system are found in plants. However, the domain structure of BTS(L) (OsHRZ1/2) is similar to the FBXL5-E3 ligase complex (Kobayashi *et al.*, 2013). In addition to their E3 ligase domain, BTS(L)/HRZ have two to three N-terminal HHE domains (Kobayashi *et al.*, 2013; Rodríguez-Celma *et al.*, 2019); a combination which is unique to this protein family and to the plant kingdom (Matthiadis and Long, 2016). Encouraged by the structural similarities and the fact that BTS/HRZ protein stability is influenced by Fe presence or absence, it is thought that BTS/HRZ (and presumably BTSL1/2) act in a similar manner as the FBXL5-E3 ligase complex (Kobayashi *et al.*, 2013; Selote *et al.*, 2015; Rodríguez-Celma *et al.*, 2019). However, whether BTS/HRZ protein is stabilized or de-stabilized by Fe is still a matter of debate (Kobayashi *et al.*, 2013; Selote *et al.*, 2015; Hindt *et al.*, 2017). Taken together, the concrete roles of BTSL1/2 during Fe deficiency, the mechanism of how BTS(L) act together and how their putative Fe sensing function can be incorporated into the bigger picture are still elusive.

BTS(L)/HRZ as negative regulators can only prevent excess Fe uptake (Aung *et al.*, 2018). The mechanism of how Fe acquisition is initiated still remains unclear. In grasses, IDEF1 was suggested as another sensor, acting as upstream activator of Strategy II Fe acquisition. The protein seems to be capable of binding divalent metal ions (Kobayashi *et al.*, 2012). However, evidence for a direct mechanistic link of Fe binding and downstream regulation of Fe homeostasis is lacking.

5.6.2 Signals for Arabidopsis Fe acquisition

Many Fe sinks are located in shoots, such as the photosynthetic machinery and reproductive tissue (*e.g.* seeds). Consequently, plants must sense the Fe status in shoots and communicate Fe demand to roots. It has been known for more than one decade that the Arabidopsis Fe uptake machinery responds to local, as well as to systemic signals (Vert *et al.*, 2003; Kumar *et al.*, 2017). However, the chemical nature of the Fe status signal(s), its transduction and perception and how local and systemic signals are integrated, are still largely obscure.

Small signaling molecules, such as ethylene, auxin and NO are known to influence Fe acquisition, but they fail to activate the Fe uptake machinery when phloem Fe levels are high, indicating that they cannot be the only signals determining Fe acquisition (Garcia *et al.*, 2013, and references therein). Furthermore, cytosolic Ca^{2+} as second messenger might influence Fe uptake via modulation of FIT protein activity. However, the signal also seems to act downstream of low phloem Fe levels (Gratz *et al.*, 2019).

The current model suggests that phloem Fe levels are important for systemic Fe status sensing, with sufficient phloem Fe levels suppressing root Fe uptake (Vert *et al.*, 2003; García *et al.*, 2018; Khan *et al.*, 2018). This model is motivated by the observations (i) that sufficient phloem Fe overrules the enhancing effect of hormone treatment (Lucena *et al.*, 2006; Romera *et al.*, 2011), and (ii) that roots of Arabidopsis mutants with impaired Fe phloem levels fail to adjust Fe uptake to the plant Fe content (Stacey *et al.*, 2008; Zhai *et al.*, 2014; Kumar *et al.*, 2017). For example, leaves of an Arabidopsis mutant line lacking YSL1 (YELLOW STRIPE-LIKE1) and YSL3, which putatively export Fe-NA from the xylem into the surrounding vegetative tissue, develop interveinal leaf chlorosis, because Fe is retained in the vasculature (Waters *et al.*, 2006; Kumar *et al.*, 2017). However, the root Fe acquisition machinery in *ysl1ysl3* loss-of-function plants does not respond to the shoot Fe deficiency. It was hypothesized that the wild-type-level Fe in the phloem sends a false Fe sufficiency signal (Kumar *et al.*, 2017). Conversely, Arabidopsis mutant lines with decreased levels of the putative phloem Fe importer OPT3 have decreased phloem Fe levels (Zhai *et al.*, 2014). In accordance with the model, the *opt3* partial loss-of-function mutants displayed constant root Fe deficiency responses although Fe was overall accumulating in the plant body (Stacey *et al.*, 2008). *opt3* shoots - in contrast to roots - properly sense Fe overload but are unable to communicate this information to roots, indicating that OPT3 is needed for shoot-to-root transduction of Fe demand via a yet unknown phloem-located signal (Khan *et al.*, 2018)

In signaling systems of other nutrients, such as phosphate (Pi) or nitrogen (N), miRNAs and peptides have been identified as long-distance signals (Ruffel, 2018). The phloem-mobile miR399 migrates from shoots to roots to induce root Pi uptake (Pant *et al.*, 2008; Huang *et al.*, 2013). During N starvation responses, different small peptides mediate root N uptake in a root-to-shoot-to-root signaling system (Tabata *et al.*, 2014; Ohkubo *et al.*, 2017). Although Fe deficiency-induced miRNAs have been identified (Kong and Yang, 2010), an accumulation upon Fe deficiency in the phloem sap could not be shown (Buhtz *et al.*, 2010). Recently, because they were shown to promote Fe uptake and to be located in the phloem, the members of the FEP/IMA family of small peptides (hereafter only called FEP) were proposed as phloem-mobile signaling molecules (Grillet *et al.*, 2018; Hirayama *et al.*, 2018).

5.7 Small proteins and peptides in nutrient signaling

Small proteins and peptides (generally defined as polypeptides with less than 100 amino acid (aa) residues) are found in all organisms as regulators of essential processes, acting amongst others in cell-to-cell and long-distance signaling (Scheidler *et al.*, 2019). Their biogenesis, structure and functions are as diverse as the current nomenclature (see reviews Delcourt *et al.*, 2018; Hsu and Benfey, 2018; Olsson *et al.*, 2019).

In Arabidopsis, CLV3 (CLAVATA3) is one of the best studied small peptides, acting in stem cell differentiation. CLV3 is cell-to-cell mobile and regulates shoot apical meristem maintenance through binding as ligand to membrane-localized receptor-like kinases (Shinohara and Matsubayashi, 2015). Very few peptides have been linked to long-distance nutrient signaling yet (Ruffel, 2018). The best studied are peptides of the Arabidopsis CEP family (C-TERMINALLY ENCODED PEPTIDE) which were identified as N starvation root-to-shoot mobile signals, where they bind as ligands to phloem-localized receptor kinases (CEP receptors (CEPR)) (Tabata *et al.*, 2014). CLV3 and CEPs are examples of so-called peptide hormones. They are usually very small (<20 aa) and derived from larger precursor proteins by several post-translational processing steps (Matsubayashi, 2014). These precursors typically have a highly variable N-terminal region, often harboring a secretion signal, and a single conserved C-terminal domain which will be cleaved off as mature bioactive peptide (Stührwohldt and Schaller, 2019). Hormone peptides are one of several categories of small signaling proteins (Scheidler *et al.*, 2019). The N starvation signal perceived in shoots by CEPRs triggers synthesis of another type phloem-mobile peptides, called CEPD (CEP-DOWNSTREAM), which transmit the signal back to roots, and induce transcription of root N importer *NITRATE TRANSPORTER2.1* (*NRT2.1*) (Ohkubo *et al.*, 2017). *NRT2.1* is only induced in root parts located in a N deprived soil patch, thus N uptake is regulated by an interplay of systemic and local signals (Ruffel and Gojon, 2017), similar to the mechanism which is assumed for Fe uptake regulation. CEPDs are glutaredoxin-like proteins of ~100 aa and are not classified as secreted hormone peptides (Ohkubo *et al.*, 2017). Similar to CEPDs, FEP1/FEP3 are phloem-localized and transcriptionally respond to deprivation of the nutrient (Grillet *et al.*, 2018; Hirayama *et al.*, 2018). In addition, although the FEP protein domain organization resembles hormone peptides, there is no evidence for a secretion signal and post-translational cleavage reported (Grillet *et al.*, 2018; Hirayama *et al.*, 2018). Reciprocal grafting of FEP3 over-expression lines with wild-type plants as well as CRISPR-Cas9 loss-of-function lines of all eight identified *FEP* genes with wild-type plants showed that the Fe uptake machinery responds mostly locally, but also systemically, to FEP3, while both responses together result in the most effective Fe uptake (Grillet *et al.*, 2018). It is however not clear from these experiments, if FEP3 itself is mobile or if it facilitates transmission of another type of signaling molecule.

With the recent advances regarding transcriptional regulation and Fe sensing and signaling, we are only beginning to understand the complex mechanisms by which Fe homeostasis is maintained at the whole-plant level. Especially our knowledge regarding a *cis*-regulatory code underlying coordination at the transcriptional level, as well as the interplay of proteins leading to activation of the Fe acquisition machinery - for example as part of a signaling cascade - is still very limited. Knowing these processes will be of great importance in order to develop efficient approaches for generating crops with improved Fe bioavailability or with overall better performance in soils with unfavorable properties.

6 Thesis Objectives

Transcriptomic and proteomic studies, for example in Arabidopsis and rice, have demonstrated the large number of protein-coding genes which act in plant Fe deficiency responses (Colangelo and Guerinot, 2004; Dinneny *et al.*, 2008; Lan *et al.*, 2011; Rodríguez-Celma *et al.*, 2013; Bashir *et al.*, 2014; Chen *et al.*, 2015; Mai *et al.*, 2016). Many of them are strongly co-regulated, implying a high degree of coordination between these players (Ivanov *et al.*, 2012). However, for most of them it is not known how they are functionally connected. Therefore, the questions were raised, (1.) which are the *cis*- and *trans*-regulators coordinating Fe deficiency responses of co-regulated genes, and (2.) how are these genes linked at the protein level. To tackle these questions, the following thesis objectives were formulated:

1. Identify *cis*-regulatory elements and novel potential upstream transcriptional regulators that control Arabidopsis root Fe deficiency responses.

On the basis of Fe deficiency time course transcriptomic data, co-expressed Fe deficiency-responsive genes are clustered and these clusters are used to identify over-represented *k*-mers in the putative promoter sequences of the genes. With a machine learning approach, we model transcriptional Fe deficiency response, meaning that *k*-mers are tested for their ability to predict Fe deficiency-induced expression. Those which robustly predict Fe deficiency responses, termed putative *cis*-regulatory elements (pCREs), are further evaluated. Their properties as plausible CREs are assessed and large-scale *in vitro* protein-DNA interaction data is used to deduce possible TFs that can bind to them. CREs and TFs with established roles during Fe deficiency are used to validate the findings.

2. Uncover connections between Fe deficiency response-related proteins to build an interaction network of Arabidopsis root proteins and to pinpoint key players.

Candidate proteins are carefully selected and tested in a targeted yeast two-hybrid screen in pair-wise manner for interaction. A protein interaction network is generated from all interactions and selected protein pairs are explored in more detail. We study promoter activity, subcellular protein localization and co-localization as well as their ability to undergo protein-protein interactions *in planta*. With stable Arabidopsis gain-of-function and loss-of-function lines, the roles of these proteins and their interplay during Fe deficiency are assessed. A final model is generated, putting the findings into the context of Fe homeostasis.

References

Andrews SC, Robinson AK, Rodríguez-Quiñones F (2003) Bacterial iron homeostasis. FEMS Microbiol Rev 27: 215-237

Aoki Y, Okamura Y, Tadaka S, Kinoshita K, Obayashi T (2016) ATTED-II in 2016: a plant coexpression database towards lineage-specific coexpression. Plant Cell Physiol 57: e5

Arosio P, Elia L, Poli M (2017) Ferritin, cellular iron storage and regulation. Iubmb Life 69: 414-422

Aung MS, Kobayashi T, Masuda H, Nishizawa NK (2018) Rice HRZ ubiquitin ligases are crucial for response to excess iron. Physiol Plant

Balk J, Pilon M (2011) Ancient and essential: the assembly of iron-sulfur clusters in plants. Trends in Plant Science 16: 218-226

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

Barberon M, Zelazny E, Robert S, Conéjéro G, Curie C, Friml J, Vert G (2011) Monoubiquitin-dependent endocytosis of the IRON-REGULATED TRANSPORTER 1 (IRT1) transporter controls iron uptake in plants. Proceedings of the National Academy of Sciences of the United States of America 108: E450-E458

Bashir K, Hanada K, Shimizu M, Seki M, Nakanishi H, Nishizawa NK (2014) Transcriptomic analysis of rice in response to iron deficiency and excess. Rice (N Y) 7: 18

Blaby-Haas CE, Merchant SS (2013) Iron sparing and recycling in a compartmentalized cell. Current Opinion in Microbiology 16: 677-685

Briat JF, Dubos C, Gaymard F (2015) Iron nutrition, biomass production, and plant product quality. Trends Plant Sci 20: 33-40

Briat JF, Duc C, Ravet K, Gaymard F (2010) Ferritins and iron storage in plants. Biochimica Et Biophysica Acta-General Subjects 1800: 806-814

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

Buhtz A, Pieritz J, Springer F, Kehr J (2010) Phloem small RNAs, nutrient stress responses, and systemic mobility. BMC Plant Biol 10: 64

Chen L, Ding C, Zhao X, Xu J, Mohammad AA, Wang S, Ding Y (2015) Differential regulation of proteins in rice (Oryza sativa L.) under iron deficiency. Plant Cell Rep 34: 83-96

Colangelo EP, Guerinot ML (2004) The essential basic helix-loop-helix protein FIT1 is required for the iron deficiency response. Plant Cell 16: 3400-3412

Curie C, Panaviene Z, Loulergue C, Dellaporta SL, Briat JF, Walker EL (2001) Maize yellow stripe1 encodes a membrane protein directly involved in Fe(III) uptake. Nature 409: 346-349

Delcourt V, Staskevicius A, Salzet M, Fournier I, Roucou X (2018) Small proteins encoded by unannotated ORFs are rising stars of the proteome, confirming shortcomings in genome annotations and current vision of an mRNA. Proteomics 18

Dinneny JR, Long TA, Wang JY, 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

Dubeaux G, Neveu J, Zelazny E, Vert G (2018) Metal sensing by the IRT1 transporter-receptor orchestrates its own degradation and plant metal nutrition. Molecular Cell 69: 953-+

Eggink LL, LoBrutto R, Brune DC, Brusslan J, Yamasato A, Tanaka A, Hoober JK (2004) Synthesis of chlorophyll b: localization of chlorophyllide a oxygenase and discovery of a stable radical in the catalytic subunit. BMC Plant Biol 4: 5

Eide D, Broderius M, Fett J, Guerinot ML (1996) A novel iron-regulated metal transporter from plants identified by functional expression in yeast. Proc Natl Acad Sci U S A 93: 5624-5628

Fourcroy P, Sisó-Terraza P, Sudre D, Savirón M, Reyt G, Gaymard F, Abadía A, Abadia J, Alvarez-Fernández A, Briat JF (2014) Involvement of the ABCG37 transporter in secretion of scopoletin and derivatives by Arabidopsis roots in response to iron deficiency. New Phytol 201: 155-167

Fuglsang AT, Guo Y, Cuin TA, Qiu QS, Song CP, Kristiansen KA, Bych K, Schulz A, Shabala S, Schumaker KS, Palmgren MG, Zhu JK (2007) Arabidopsis protein kinase PKS5 inhibits the plasma membrane H^+-ATPase by preventing interaction with 14-3-3 protein. Plant Cell 19: 1617-1634

Fuglsang AT, Kristensen A, Cuin TA, Schulze WX, Persson J, Thuesen KH, Ytting CK, Oehlenschlæger CB, Mahmood K, Sondergaard TE, Shabala S, Palmgren MG (2014) Receptor kinase-mediated control of primary active proton pumping at the plasma membrane. Plant J 80: 951-964

Fuglsang AT, Visconti S, Drumm K, Jahn T, Stensballe A, Mattei B, Jensen ON, Aducci P, Palmgren MG (1999) Binding of 14-3-3 protein to the plasma membrane H^+-ATPase AHA2 involves the three C-terminal residues Tyr^{946}-Thr-Val and requires phosphorylation of Thr^{947}. Journal of Biological Chemistry 274: 36774-36780

Fuss JO, Tsai CL, Ishida JP, Tainer JA (2015) Emerging critical roles of Fe-S clusters in DNA replication and repair. Biochimica Et Biophysica Acta-Molecular Cell Research 1853: 1253-1271

Gao F, Robe K, Gaymard F, Izquierdo E, Dubos C (2019) The transcriptional control of iron homeostasis in plants: a tale of bHLH transcription factors? Frontiers in Plant Science 10

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

García MJ, Corpas FJ, Lucena C, Alcántara E, Pérez-Vicente R, Zamarreño Á, Bacaicoa E, García-Mina JM, Bauer P, Romera FJ (2018) A shoot Fe signaling pathway requiring the OPT3 transporter controls GSNO reductase and ethylene in *Arabidopsis thaliana* roots. Front Plant Sci 9: 1325

Giehl RFH, Gruber BD, von Wirén N (2014) Its time to make changes: modulation of root system architecture by nutrient signals. Journal of Experimental Botany 65: 769-778

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

Gratz R, Manishankar P, Ivanov R, Köster P, Mohr I, Trofimov K, Steinhorst L, Meiser J, Mai HJ, Drerup M, Arendt S, Holtkamp M, Karst U, Kudla J, Bauer P, Brumbarova T (2019) CIPK11-dependent phosphorylation modulates FIT activity to promote Arabidopsis iron acquisition in response to calcium signaling. Dev Cell

Grillet L, Lan P, Li WF, Mokkapati G, Schmidt W (2018) IRON MAN is a ubiquitous family of peptides that control iron transport in plants. Nature Plants 4: 953-+

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

Hell R, Stephan UW (2003) Iron uptake, trafficking and homeostasis in plants. Planta 216: 541-551

Hentze MW, Muckenthaler MU, Andrews NC (2004) Balancing acts: Molecular control of mammalian iron metabolism. Cell 117: 285-297

Hindt MN, Akmakjian GZ, Pivarski KL, Punshon T, Baxter I, Salt DE, Guerinot ML (2017) BRUTUS and its paralogs, BTS LIKE1 and BTS LIKE2, encode important negative regulators of the iron deficiency response in *Arabidopsis thaliana*. Metallomics 9: 876-890

Hirayama T, Lei GJ, Yamaji N, Nakagawa N, Ma JF (2018) The putative peptide gene FEP1 regulates iron deficiency response in Arabidopsis. Plant and Cell Physiology 59: 1739-1752

Hong S, Kim SA, Guerinot ML, McClung CR (2013) Reciprocal interaction of the circadian clock with the iron homeostasis network in Arabidopsis. Plant Physiology 161: 893-903

Hsu PY, Benfey PN (2018) Small but mighty: functional peptides encoded by small ORFs in plants. Proteomics 18: e1700038

Huang TK, Han CL, Lin SI, Chen YJ, Tsai YC, Chen YR, Chen JW, Lin WY, Chen PM, Liu TY, Chen YS, Sun CM, Chiou TJ (2013) Identification of downstream components of

ubiquitin-conjugating enzyme PHOSPHATE2 by quantitative membrane proteomics in Arabidopsis roots. Plant Cell 25: 4044-4060

Inoue H, Kobayashi T, Nozoye T, Takahashi M, Kakei Y, Suzuki K, Nakazono M, Nakanishi H, Mori S, Nishizawa NK (2009) Rice OsYSL15 is an iron-regulated iron(III)-deoxymugineic acid transporter expressed in the roots and is essential for iron uptake in early growth of the seedlings. Journal of Biological Chemistry 284: 3470-3479

Ishimaru Y, Suzuki M, Tsukamoto T, Suzuki K, Nakazono M, Kobayashi T, Wada Y, Watanabe S, Matsuhashi S, Takahashi M, Nakanishi H, Mori S, Nishizawa NK (2006) Rice plants take up iron as an Fe^{3+}-phytosiderophore and as Fe^{2+}. Plant J 45: 335-346

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

Ivanov R, Brumbarova T, Blum A, Jantke AM, Fink-Straube C, Bauer P (2014) SORTING NEXIN1 is required for modulating the trafficking and stability of the Arabidopsis IRON-REGULATED TRANSPORTER1. Plant Cell 26: 1294-1307

Jakoby M, Wang HY, Reidt W, Weisshaar B, Bauer P (2004) FRU (029) is required for induction of iron mobilization genes in *Arabidopsis thaliana*. FEBS Lett 577: 528-534

Khan MA, Castro-Guerrero NA, McInturf SA, Nguyen NT, Dame AN, Wang J, Bindbeutel RK, Joshi T, Jurisson SS, Nusinow DA, Mendoza-Cozatl DG (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

Kobayashi T, Itai RN, Aung MS, Senoura T, Nakanishi H, Nishizawa NK (2012) The rice transcription factor IDEF1 directly binds to iron and other divalent metals for sensing cellular iron status. Plant J 69: 81-91

Kobayashi T, Itai RN, Nishizawa NK (2014) Iron deficiency responses in rice roots. Rice 7

Kobayashi T, Itai RN, Ogo Y, Kakei Y, Nakanishi H, Takahashi M, Nishizawa NK (2009) The rice transcription factor IDEF1 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, Nagasaka S, Senoura T, Itai RN, Nakanishi H, Nishizawa NK (2013) Iron-binding haemerythrin RING ubiquitin ligases regulate plant iron responses and accumulation. Nat Commun 4: 2792

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

Kobayashi T, Nishizawa NK (2012) Iron uptake, translocation, and regulation in higher plants. Annu Rev Plant Biol 63: 131-152

Kobayashi T, Nishizawa NK (2014) Iron sensors and signals in response to iron deficiency. Plant Sci 224: 36-43

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 IDEF1 and IDEF2. Ann Bot 105: 1109-1117

Kobayashi T, Ogo Y, Itai RN, Nakanishi H, Takahashi M, Mori S, Nishizawa NK (2007) The transcription factor IDEF1 regulates the response to and tolerance of iron deficiency in plants. Proc Natl Acad Sci U S A 104: 19150-19155

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

Komander D, Rape M (2012) The ubiquitin code. Annu Rev Biochem 81: 203-229

Kong WW, Yang ZM (2010) Identification of iron-deficiency responsive microRNA genes and *cis*-elements in Arabidopsis. Plant Physiol Biochem 48: 153-159

Kumar RK, Chu HH, Abundis C, Vasques K, Rodriguez DC, Chia JC, Huang R, Vatamaniuk OK, Walker EL (2017) Iron-nicotianamine transporters are required for proper long distance iron signaling. Plant Physiol 175: 1254-1268

Lan P, Li W, Wen TN, Shiau JY, Wu YC, Lin W, Schmidt W (2011) iTRAQ protein profile analysis of Arabidopsis roots reveals new aspects critical for iron homeostasis. Plant Physiol 155: 821-834

Li W, Bengtson MH, Ulbrich A, Matsuda A, Reddy VA, Orth A, Chanda SK, Batalov S, Joazeiro CAP (2008) Genome-wide and functional annotation of human E3 ubiquitin ligases identifies MULAN, a mitochondrial E3 that regulates the organelle's dynamics and signaling. Plos One 3

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: 2478-2493

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

Lindsay WL (1988) Solubility and redox equilibria of iron compounds in soils. *In* S J.W., G B.A., S U., eds, Iron in Soils and Clay Minerals, Vol 217. Springer, Dordrecht, p 894

Lingam S, Mohrbacher J, Brumbarova T, Potuschak T, Fink-Straube C, Blondet E, Genschik P, Bauer P (2011) Interaction between the bHLH transcription factor FIT and ETHYLENE INSENSITIVE3/ETHYLENE INSENSITIVE3-LIKE1 reveals molecular linkage between the regulation of iron acquisition and ethylene signaling in Arabidopsis. Plant Cell 23: 1815-1829

Long LZ, Persson DP, Duan FY, J ø rgensen K, Yuan LX, Schjoerring JK, Pedas PR (2018) The iron-regulated transporter 1 plays an essential role in uptake, translocation and grain-loading of manganese, but not iron, in barley. New Phytologist 217: 1640-1653

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

Lucena C, Waters BM, Romera FJ, García MJ, Morales M, Alcántara E, Pérez-Vicente R (2006) Ethylene could influence ferric reductase, iron transporter, and H^+-ATPase gene expression by affecting *FER* (or *FER*-like) gene activity. J Exp Bot 57: 4145-4154

Mai 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

Marschner H, Römheld V (1994) Strategies of plants for acquisition of iron. Plant and Soil 165: 261-274

Matsubayashi Y (2014) Posttranslationally modified small-peptide signals in plants. Annu Rev Plant Biol 65: 385-413

Matthiadis A, Long TA (2016) Further insight into BRUTUS domain composition and functionality. Plant Signal Behav 11: e1204508

McLean E, Cogswell M, Egli I, Wojdyla D, de Benoist B (2009) Worldwide prevalence of anaemia, WHO Vitamin and Mineral Nutrition Information System, 1993-2005. Public Health Nutr 12: 444-454

Meiser J, Lingam S, Bauer P (2011) Posttranslational regulation of the iron deficiency basic helix-loop-helix transcription factor FIT is affected by iron and nitric oxide. Plant Physiol 157: 2154-2166

Mukhopadhyay D, Riezman H (2007) Proteasome-independent functions of ubiquitin in endocytosis and signaling. Science 315: 201-205

Murgia I, Tarantino D, Soave C, Morandini P (2011) Arabidopsis CYP82C4 expression is dependent on Fe availability and circadian rhythm, and correlates with genes involved in the early Fe deficiency response. J Plant Physiol 168: 894-902

Netz DJA, Stith CM, Stumpfig M, Köpf G, Vogel D, Genau HM, Stodola JL, Lill R, Burgers PMJ, Pierik AJ (2012) Eukaryotic DNA polymerases require an iron-sulfur cluster for the formation of active complexes. Nature Chemical Biology 8: 125-132

Nozoye T, Nagasaka S, Bashir K, Takahashi M, Kobayashi T, Nakanishi H, Nishizawa NK (2014) Nicotianamine synthase 2 localizes to the vesicles of iron-deficient rice roots, and its mutation in the YXXφ or LL motif causes the disruption of vesicle formation or movement in rice. Plant Journal 77: 246-260

Nozoye T, Nagasaka S, Kobayashi T, Takahashi M, Sato Y, Uozumi N, Nakanishi H, Nishizawa NK (2011) Phytosiderophore efflux transporters are crucial for iron acquisition in graminaceous plants. J Biol Chem 286: 5446-5454

Obayashi T, Kinoshita K (2009) Rank of correlation coefficient as a comparable measure for biological significance of gene coexpression. DNA Res 16: 249-260

Ogo Y, Itai RN, Nakanishi H, Kobayashi T, Takahashi M, Mori S, Nishizawa NK (2007) The rice bHLH protein OsIRO2 is an essential regulator of the genes involved in Fe uptake under Fe-deficient conditions. Plant J 51: 366-377

Ogo Y, Kobayashi T, Nakanishi Itai R, Nakanishi H, Kakei Y, Takahashi M, Toki S, Mori S, Nishizawa NK (2008) A novel NAC transcription factor, IDEF2, that recognizes the iron deficiency-responsive element 2 regulates the genes involved in iron homeostasis in plants. J Biol Chem 283: 13407-13417

Ohkubo Y, Tanaka M, Tabata R, Ogawa-Ohnishi M, Matsubayashi Y (2017) Shoot-to-root mobile polypeptides involved in systemic regulation of nitrogen acquisition. Nat Plants 3: 17029

Olsson V, Joos L, Zhu S, Gevaert K, Butenko MA, De Smet I (2019) Look closely, the beautiful may be small: precursor-derived peptides in plants. Annu Rev Plant Biol 70: 153-186

Palmer CM, Hindt MN, Schmidt H, Clemens S, Guerinot ML (2013) MYB10 and MYB72 are required for growth under iron-limiting conditions. PLoS Genet 9: e1003953

Pant BD, Buhtz A, Kehr J, Scheible WR (2008) MicroRNA399 is a long-distance signal for the regulation of plant phosphate homeostasis. Plant J 53: 731-738

Pedas P, Ytting CK, Fuglsang AT, Jahn TP, Schjoerring JK, Husted S (2008) Manganese efficiency in barley: Identification and characterization of the metal ion transporter HvIRT1. Plant Physiology 148: 455-466

Petit JM, van Wuytswinkel O, Briat JF, Lobréaux S (2001) Characterization of an iron-dependent regulatory sequence involved in the transcriptional control of *AtFer1* and *ZmFer1* plant ferritin genes by iron. Journal of Biological Chemistry 276: 5584-5590

Petry N, Olofin I, Hurrell RE, Boy E, Wirth JP, Moursi M, Angel MD, Rohner F (2016) The proportion of anemia associated with iron deficiency in low, medium, and high human development index countries: a systematic analysis of national surveys. Nutrients 8

Pires N, Dolan L (2010) Origin and diversification of basic-helix-loop-helix proteins in plants. Mol Biol Evol 27: 862-874

Rajniak J, Giehl RFH, Chang E, Murgia I, von Wirén N, Sattely ES (2018) Biosynthesis of redox-active metabolites in response to iron deficiency in plants. Nat Chem Biol 14: 442-450

Robinson NJ, Procter CM, Connolly EL, Guerinot ML (1999) A ferric-chelate reductase for iron uptake from soils. Nature 397: 694-697

Rodríguez-Celma J, Chou H, Kobayashi T, Long TA, Balk J (2019) Hemerythrin E3 Ubiquitin Ligases as Negative Regulators of Iron Homeostasis in Plants. Front Plant Sci 10: 98

Rodríguez-Celma J, Pan IC, Li W, Lan P, Buckhout TJ, Schmidt W (2013) The transcriptional response of Arabidopsis leaves to Fe deficiency. Front Plant Sci 4: 276

Romera FJ, García MJ, Alcántara E, Pérez-Vicente R (2011) Latest findings about the interplay of auxin, ethylene and nitric oxide in the regulation of Fe deficiency responses by Strategy I plants. Plant Signal Behav 6: 167-170

Romero-Barrios N, Vert G (2018) Proteasome-independent functions of lysine-63 polyubiquitination in plants. New Phytologist 217: 995-1011

Ruffel S (2018) Nutrient-related long-distance signals: common players and possible crosstalk. Plant & Cell Physiology

Ruffel S, Gojon A (2017) Systemic nutrient signalling: On the road for nitrate. Nat Plants 3: 17040

Römheld V, Marschner H (1986) Evidence for a specific uptake system for iron phytosiderophores in roots of grasses. Plant Physiol 80: 175-180

Saito A, Shimizu M, Nakamura H, Maeno S, Katase R, Miwa E, Higuchi K, Sonoike K (2014) Fe deficiency induces phosphorylation and translocation of Lhcb1 in barley thylakoid membranes. FEBS Lett 588: 2042-2048

Salahudeen AA, Thompson JW, Ruiz JC, Ma HW, Kinch LN, Li Q, Grishin NV, Bruick RK (2009) An E3 ligase possessing an iron-responsive hemerythrin domain is a regulator of iron homeostasis. Science 326: 722-726

Salomé PA, Oliva M, Weigel D, Krämer U (2013) Circadian clock adjustment to plant iron status depends on chloroplast and phytochrome function. Embo Journal 32: 511-523

Samira R, Li BH, Kliebenstein D, Li CY, Davis E, Gillikin JW, Long TA (2018) The bHLH transcription factor ILR3 modulates multiple stress responses in Arabidopsis. Plant Molecular Biology 97: 297-309

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

Santner A, Estelle M (2010) The ubiquitin-proteasome system regulates plant hormone signaling. Plant J 61: 1029-1040

Scheidler CM, Kick LM, Schneider S (2019) Ribosomal peptides and small proteins on the rise. Chembiochem

Schmid NB, 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

Selote D, Samira R, Matthiadis 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

Shin LJ, Lo JC, Chen GH, Callis J, Fu HY, Yeh KC (2013) IRT1 DEGRADATION FACTOR1, a RING E3 Ubiquitin Ligase, Regulates the Degradation of IRON-REGULATED TRANSPORTER1 in Arabidopsis. Plant Cell 25: 3039-3051

Shinohara H, Matsubayashi Y (2015) Reevaluation of the CLV3-receptor interaction in the shoot apical meristem: dissection of the CLV3 signaling pathway from a direct ligand-binding point of view. Plant J 82: 328-336

Shu K, Yang W (2017) E3 ubiquitin ligases: ubiquitous actors in plant development and abiotic stress responses. Plant Cell Physiol 58: 1461-1476

Sisó-Terraza P, Luis-Villarroya A, Fourcroy P, Briat JF, Abadía A, Gaymard F, Abadía J, Álvarez-Fernández A (2016) Accumulation and secretion of coumarinolignans and other coumarins in *Arabidopsis thaliana* roots in response to iron deficiency at high pH. Front Plant Sci 7: 1711

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 Journal 66: 1044-1052

Sivitz AB, Hermand V, Curie C, Vert G (2012) Arabidopsis bHLH100 and bHLH101 control iron homeostasis via a FIT-independent pathway. PLoS One 7: e44843

Stacey MG, Patel A, McClain WE, Mathieu M, Remley M, Rogers EE, Gassmann W, Blevins DG, Stacey G (2008) The Arabidopsis AtOPT3 protein functions in metal homeostasis and movement of iron to developing seeds. Plant Physiol 146: 589-601

Stone SL (2014) The role of ubiquitin and the 26S proteasome in plant abiotic stress signaling. Front Plant Sci 5: 135

Stringlis IA, Yu K, Feussner K, de Jonge R, Van Bentum S, Van Verk MC, Berendsen RL, Bakker PAHM, Feussner I, Pieterse CMJ (2018) MYB72-dependent coumarin exudation shapes root microbiome assembly to promote plant health. Proc Natl Acad Sci U S A 115: E5213-E5222

Stührwohldt N, Schaller A (2019) Regulation of plant peptide hormones and growth factors by post-translational modification. Plant Biol (Stuttg) 21 Suppl 1: 49-63

Tabata R, Sumida K, Yoshii T, Ohyama K, Shinohara H, Matsubayashi Y (2014) Perception of root-derived peptides by shoot LRR-RKs mediates systemic N-demand signaling. Science 346: 343-346

Takagi S, Nomoto K, Takemoto T (1984) Physiological aspect of mugineic acid, a possible phytosiderophore of graminaceous plants. Journal of Plant Nutrition 7: 469-477

Thompson JW, Bruick RK (2012) Protein degradation and iron homeostasis. Biochim Biophys Acta 1823: 1484-1490

Tissot N, Robe K, Gao F, Grant-Grant S, Boucherez J, Bellegarde F, Maghiaoui A, Marcelin R, Izquierdo E, Benhamed M, Martin A, Vignols F, Roschzttardtz H, Gaymard F, Briat JF, Dubos C (2019) Transcriptional integration of the responses to iron availability in Arabidopsis by the bHLH factor ILR3. New Phytol

Tsai HH, Rodríguez-Celma J, Lan P, Wu YC, Vélez-Bermúdez IC, Schmidt W (2018) Scopoletin 8-hydroxylase-mediated fraxetin production is crucial for iron mobilization. Plant Physiol 177: 194-207

Urzica EI, Casero D, Yamasaki H, Hsieh SI, Adler LN, Karpowicz SJ, Blaby-Haas CE, Clarke SG, Loo JA, Pellegrini M, Merchant SS (2012) Systems and trans-system level analysis identifies conserved iron deficiency responses in the plant lineage. Plant Cell 24: 3921-3948

Vert G, Grotz N, Dédaldéchamp F, Gaymard F, Guerinot ML, Briat JF, Curie C (2002) IRT1, an Arabidopsis transporter essential for iron uptake from the soil and for plant growth. Plant Cell 14: 1223-1233

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 Physiology 132: 796-804

Vierstra RD (2009) The ubiquitin-26S proteasome system at the nexus of plant biology. Nature Reviews Molecular Cell Biology 10: 385-397

Vigani G, Maffi D, Zocchi G (2009) Iron availability affects the function of mitochondria in cucumber roots. New Phytol 182: 127-136

Vigani G, Tarantino D, Murgia I (2013) Mitochondrial ferritin is a functional iron-storage protein in cucumber (*Cucumis sativus*) roots. Frontiers in Plant Science 4

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

Waters BM, Chu HH, Didonato RJ, Roberts LA, Eisley RB, Lahner B, Salt DE, Walker EL (2006) Mutations in Arabidopsis *yellow stripe-like1* and *yellow stripe-like3* reveal their roles in metal ion homeostasis and loading of metal ions in seeds. Plant Physiol 141: 1446-1458

Wild M, Davière JM, Regnault T, Sakvarelidze-Achard L, Carrera E, Diaz IL, Cayrel A, Dubeaux G, Vert G, Achard P (2016) Tissue-specific regulation of gibberellin signaling fine-tunes Arabidopsis iron-deficiency responses. Developmental Cell 37: 190-200

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

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

Yang TJ, Lin WD, Schmidt W (2010) Transcriptional profiling of the Arabidopsis iron deficiency response reveals conserved transition metal homeostasis networks. Plant Physiol 152: 2130-2141

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

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

Zamioudis C, Hanson J, Pieterse CM (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, Guerinot ML, Salt DE, Kochian LV, Vatamaniuk OK (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 C (2014) Essential functions of iron-requiring proteins in DNA replication, repair and cell cycle control. Protein Cell 5: 750-760

Zhang HM, Li Y, Yao XN, Liang G, Yu DQ (2017) POSITIVE REGULATOR OF IRON HOMEOSTASIS1, OsPRI1, facilitates iron homeostasis. Plant Physiology 175: 543-554

Zhang J, Liu B, 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

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 U S A 108: 14992-14997

7 Manuscript 1

Putative *cis*-regulatory elements predict iron deficiency responses in Arabidopsis roots

Putative *cis*-regulatory elements predict iron deficiency responses in Arabidopsis roots

Birte Schwarz[1], Christina B. Azodi[2], Shin-Han Shiu[2,3], Petra Bauer[1,4]

[1]Institute of Botany, Heinrich Heine University, Universitätsstr. 1, Düsseldorf, Germany
[2]Department of Plant Biology, Michigan State University, East Lansing, MI, USA
[3]Department of Computational, Mathematics, Science, and Engineering, Michigan State University, East Lansing, MI, USA
[4]Cluster of Excellence on Plant Science (CEPLAS), Heinrich Heine University, Düsseldorf, Germany

One sentence summary

>100 putative *cis*-regulatory elements robustly predict Arabidopsis root Fe deficiency-responses in computational models, and shed light on the mechanisms of transcriptional regulation.

Funding: This work was supported by the German Research Foundation grant through the DFG International Research Training group 1525 to P.B., a NSF Graduate Research Fellowship to C.B.A, and by grants to S.-H.S. from the US National Science Foundation IOS-1546617, DEB-1655386, and DGE-1828149, and the US Department of Energy (DOE) Great Lakes Bioenergy Research Center (DOE Office of Science BER DE-SC0018409). This work received funding from Germany´s Excellence Strategy, EXC 2048/1, Project ID: 390686111.

Author contributions: B.S., P.B., and S.-H.S. conceived the project; B.S., and S.-H.S. designed the research plan; B.S., and C.B.A. analyzed the data; B.S. wrote the original draft; B.S., C.B.A., S.-H.S., and P.B. reviewed and edited the article; P.B., C.B.A, and S.-H.S. acquired funding. P.B. and S.-H.S. agree to serve as the authors responsible for contact and ensure communication.

Corresponding authors: Petra Bauer, petra.bauer@hhu.de; Shin-Han Shiu, shius@msu.edu

Keywords: Arabidopsis, Fe deficiency, machine learning, Random Forest, co-expression clustering, *cis*-regulatory element, transcription factor binding motif, FIT, IDE1, coumarin

Abbreviations

CIS-BP	Catalog of inferred sequence binding preferences
CNS	Conserved non-coding sequence
DAP-seq	DNA affinity purification sequencing
Fe / -Fe	Iron / Iron deficiency
FeS	Iron-Sulfur
FET	Fisher's exact test
FIT	FE DEFICIENCY-INDUCED TRANSCRIPTION FACTOR
freq-pCRE	Frequent pCRE
GO	(Biological process) Gene ontology
GS	Gold standard
IDE	Iron Deficiency-responsive Element
log2FC	log_2 fold-change
MA	Mugineic acid
min-pCRE	Minimum set pCRE
PCC	Pearson's correlation coefficient
pCRE	Putative *cis*-regulatory element
PWM	Position weight matrix
RF	Random Forest
TF	Transcription factor
TFBM	Transcription factor binding motif
TSS	Transcription start site
TTS	Transcription termination site
Zn	Zinc

Abstract

Iron (Fe) is a key cofactor in many cellular redox processes, including respiration and photosynthesis. Plant Fe deficiency (-Fe) activates a complex regulatory network which coordinates root Fe uptake and distribution to sink tissues, while avoiding over-accumulation of Fe and other metals to toxic levels. In Arabidopsis (*Arabidopsis thaliana*), FIT (FER-LIKE FE DEFICIENCY-INDUCED TRANSCRIPTION FACTOR), a bHLH transcription factor (TF), is required for up-regulation of root Fe acquisition genes. However, other root and shoot -Fe-induced genes involved in Fe allocation and signaling are FIT-independent. The *cis*-regulatory code, i.e. the *cis*-regulatory elements (CREs) and their combinations that regulate plant -Fe-responses, remains largely elusive. Using Arabidopsis genome and transcriptome data, we identified over 100 putative CREs (pCREs) that were predictive of -Fe-induced up-regulation of genes in root tissue. We used large-scale *in vitro* TF binding data, association with FIT-dependent or FIT-independent co-expression clusters, positional bias, and evolutionary conservation to assess pCRE properties and possible functions. In addition to bHLH and MYB TFs, also B3, NAC, bZIP, and TCP TFs might be important regulators for -Fe responses. Our approach uncovered IDE1 (Iron Deficiency-responsive Element 1), a -Fe response CRE in grass species, to be conserved in regulating genes for biosynthesis of Fe-chelating compounds also in Arabidopsis. Our findings provide a comprehensive source of *cis*-regulatory information for -Fe-responsive genes, that advances our mechanistic understanding and informs future efforts in engineering plants with more efficient Fe uptake or transport systems.

Introduction

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; Rodríguez-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 rice and Arabidopsis plants experience -Fe, subgroup IVc bHLH proteins activate subgroup Ib and IVb *BHLH* genes (Zhang et al., 2015; Liang et al., 2017). Downstream from IVc bHLH TFs (ILR3/bHLH34/bHLH104/bHLH115 in Arabidopsis, PRI1 in rice), subgroup Ib bHLH TFs (bHLH38/39/100/101 in Arabidopsis, IRO2 in rice) and subgroup IVb bHLH TFs (PYE in Arabidopsis, IRO3 in rice) regulate responses further downstream (Ogo et al., 2007; Yuan et al., 2008; Long et al., 2010; Zheng et al., 2010). In addition, IVc bHLH protein levels are controlled by Fe-regulated E3 ligases (Selote et al., 2015; Zhang et al., 2017).

Despite these conserved regulatory and functional interactions of subgroup IVc, Ib, and IVb bHLH TFs between grass and non-grass species, it remains unclear if other regulatory components between Strategy I and II are conserved. For example, in grasses, IDEF1 (IRON DEFICIENCY-RESPONSIVE ELEMENT BINDING FACTOR1, ABI3VP1 subfamily of B3 TF) and IDEF2 (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, while barley IDE1 can drive reporter gene expression in tobacco 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). Strategy I Fe acquisition requires the bHLH TF FIT (FER-LIKE IRON DEFICIENCY-INDUCED TRANSCRIPTION FACTOR) that is absent in rice (Colangelo and Guerinot, 2004; Jakoby et al., 2004), and is activated upon -Fe mainly through interaction 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 chelator biosynthesis and export (Colangelo and Guerinot, 2004; Jakoby et al., 2004; Sivitz et al., 2012; Schmid et al., 2014; Mai et al., 2016).

A co-expression network built with -Fe-responsive genes gives insight into the complex -Fe regulatory system in Arabidopsis (Ivanov et al., 2012). Among co-expression clusters, one contains root-specific and FIT-dependent genes involved in Fe acquisition, while another one is composed of root- and shoot-expressed FIT-independent genes. 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) -Fe-induced genes. The concept to discriminate FIT-dependent and FIT-independent co-expression clusters has proven very informative for interpreting mutant phenotypes and to place novel regulators into the -Fe response cascade (e.g. Zhang et al., 2015; Liang et al., 2017; Gratz et al., 2019). FIT-independent network genes mostly act in sub-cellular 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). Only few upstream regulators for FIT-independent gene expression have been identified yet, namely bHLH IVc TFs, controlling Ib *BHLH* and *PYE*, and PYE controlling *NAS4*, *ZIF1* and *FRO3* of the same co-expression 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) which 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 over-representation of sequences in the promoter regions of co-regulated 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 (reviewed in Ma et al., 2014) can be applied to build models with pCREs to predict gene expression or transcriptional responses. These models can be evaluated by making predictions on expression of genes that are not part of the 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., 2017).

To get a deeper understanding of -Fe response regulation, we elucidate the underlying *cis*-regulatory code. Because some TFs have well established roles in -Fe response, we can use these to validate our findings. 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 over-represented in promoters of co-expressed 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 genome-wide scale, we defined root -Fe response co-expression clusters, then we identified *k*-mers 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, such as the choice of data set or the measure used to define expression similarity, impact the discovery of functional connections between genes (Uygun et al., 2016), we used multiple expression data combinations and algorithms with varying parameters (see **Methods**).

-Fe-responsive genes ($\log_2$ fold-change (log2FC) >1 or <-1, q<0.05) were identified using transcriptomic data available for six time points after an -Fe treatment in Arabidopsis seedling roots (Dinneny et al., 2008). Enrichment analysis (Fisher's exact test (FET), q<0.05) of biological process gene ontologies (GOs) showed that, next to Fe-related GOs (e.g. Fe transport, homeostasis and FeS cluster assembly), responses to several hormones, including auxin, ethylene, abscisic acid and jasmonic acid, were over-represented (**Figure 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 (**Figure 1B**; **Supplemental Figure S1**).

Using multiple expression data sets to define -Fe co-expression clusters

We next grouped differentially regulated -Fe response genes into co-expression clusters using two approaches: *k*-means clustering and correlation to gold standard. *K*-means clustering was based on the transcriptional responses to -Fe alone and combined with different responses to other stress and developmental conditions (**Figure 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 gold standard (GS) -Fe response genes (see **Methods**; **Supplemental Table S1**), by selecting the differentially regulated -Fe response genes with a significantly similar (Pearson's Correlation Coefficient (PCC); see **Methods**) expression pattern to the GS gene, also using the different combinations of transcriptional data

(**Figure 1C**). To identify co-expression clusters with similar biological functions, we grouped them according to their enriched GOs (FET, q<0.05) into "superclusters" (**Figure 2A**; **Supplemental Figure S2A**; see **Methods**), which were defined as groups of at least 20 clusters with significantly higher similarity to each other than the average similarity of all clusters (all Mann-Whitney U, p<2.2e-16; **Supplemental Figure S2B, C**). While *k*-means supercluster C was enriched in an Fe-related GO (cellular response to Fe, GO shared by ≥75% of co-expression clusters within each supercluster), *k*-means superclusters A and B shared GOs related to different stress responses.

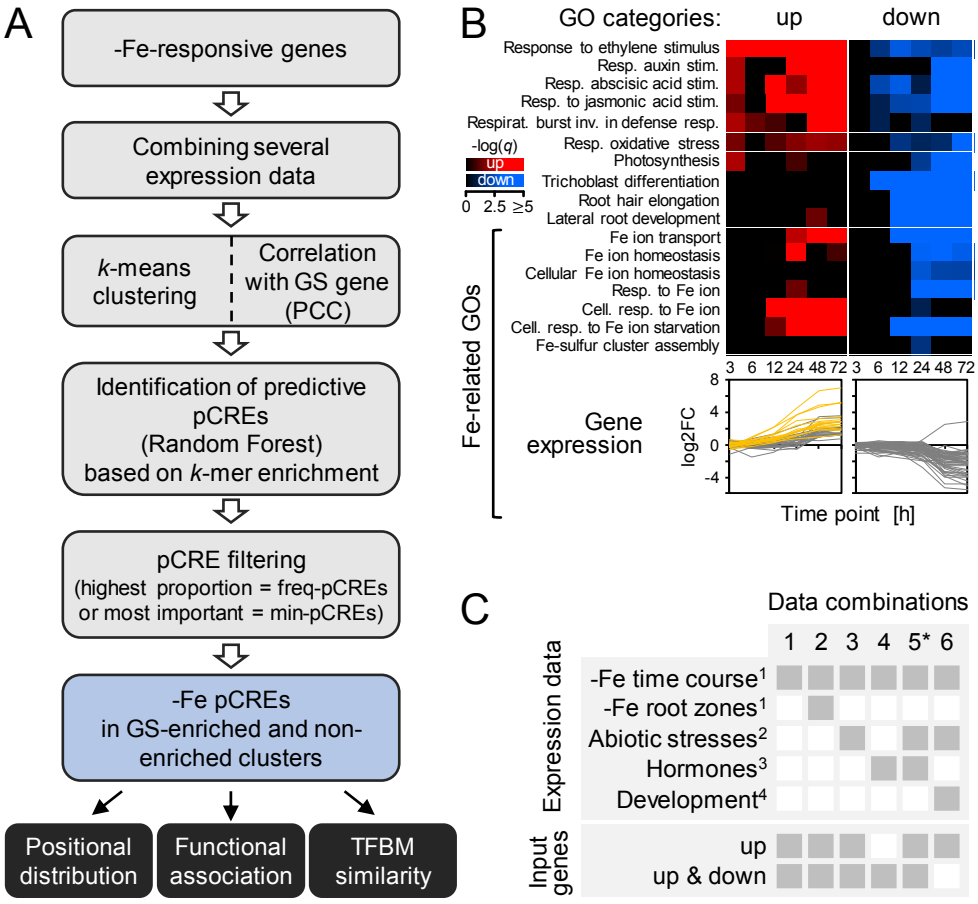


Figure 1. -Fe pCRE identification workflow and transcriptomic data.
A: pCRE identification workflow. **B:** Heatmap of enrichment (FET, q<0.05) of selected GO terms in genes that were significantly up- (red) or down-regulated (blue) (q<0.05) at ≥1 of 6 time points in -Fe-treated roots of 6 d-old seedlings (Dinneny et al., 2008). Differential regulation was defined as $\log_2$ fold-change (log2FC) >1 or <-1 (treatment vs. control). GOs are sorted by category, and expression patterns of genes corresponding to Fe-related GOs are shown below the heatmap. Yellow genes indicate -Fe GS genes. **C:** Transcriptomic data combinations which were used for clustering of co-expressed genes. Gray filled boxes in columns depict (top) expression data used in the combination and (bottom) if up (up-regulated only) or up & down (up- and down-regulated) genes were included. [1](Dinneny et al., 2008), [2](Kilian et al., 2007), [3](Goda et al., 2008), [4](Schmid et al., 2005), *Tested with (5a) and without (5b) genotoxic stress data, (5b) input only up-regulated genes.

Because the current TAIR GO annotation for -Fe response-related processes does not contain all -Fe-responsive genes of interest (e.g. *MYB10*, *UGT72E1*, *AT3G07720*, *FEP3* or *NAS4*, (Ivanov et al., 2012)), we also determined if *k*-means-generated co-expression clusters

were enriched for GS genes ("GS-enriched"; FET, q<0.05; **right**, **Figure 2A**). 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 co-expression clusters that would have been overlooked by conventional GO enrichment analysis. In total, 7% of the *k*-means-generated clusters were GS-enriched (**Figure 2B**). Applying this same analysis to the correlation-based clusters (**Figure 2A**; **Supplemental Figure S2A**), we found higher levels of similarity between correlation-based clusters compared to *k*-means clusters (Mann-Whitney U, p<2.2e-16; **Supplemental Figure S2D**), because we pre-condition their identification on GS genes, some of which are tightly co-regulated (Ivanov et al., 2012). Accordingly, we found that 93% were GS-enriched (**right**, **Figure 2B**).

GS genes are either FIT target ("FIT-dependent") or FIT-independent Fe homeostasis ("FIT-independent") genes, which we found reflected in our GS-enriched clusters: 71% of the GS-enriched clusters were more specifically enriched for FIT-dependent (39%) and/or FIT-independent genes (32%) (FET, q<0.05; **bottom**, **Figure 2B**). The remaining GS-enriched clusters were enriched for both ("mixed"). Interestingly, clusters based on combined expression data (i.e. data combinations (dc) 2, 3, 5a/b, 6) were more often enriched for FIT-dependent or FIT-independent genes, while -Fe time course data alone (dc1) produced mainly mixed category clusters (**Figure 2C**). The utility of including spatial or developmental data (dc2, 6) to define co-expression 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 for GS-genes ("non-enriched") tended to respond to particular abiotic stresses, for example cold (**Supplemental Figure S8**; e.g. clusters 937, 973) or salt (e.g. clusters 900, 915, 936, 987), whereas gene expression for GS-enriched cluster genes tended to randomly oscillate under different abiotic stresses (e.g. clusters 818, 835, 858, 889), which might indicate different regulatory networks and highlight the usefulness of incorporating additional abiotic stress data (as in dc3, 5, 6) when defining co-expression clusters that are likely co-regulated.

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

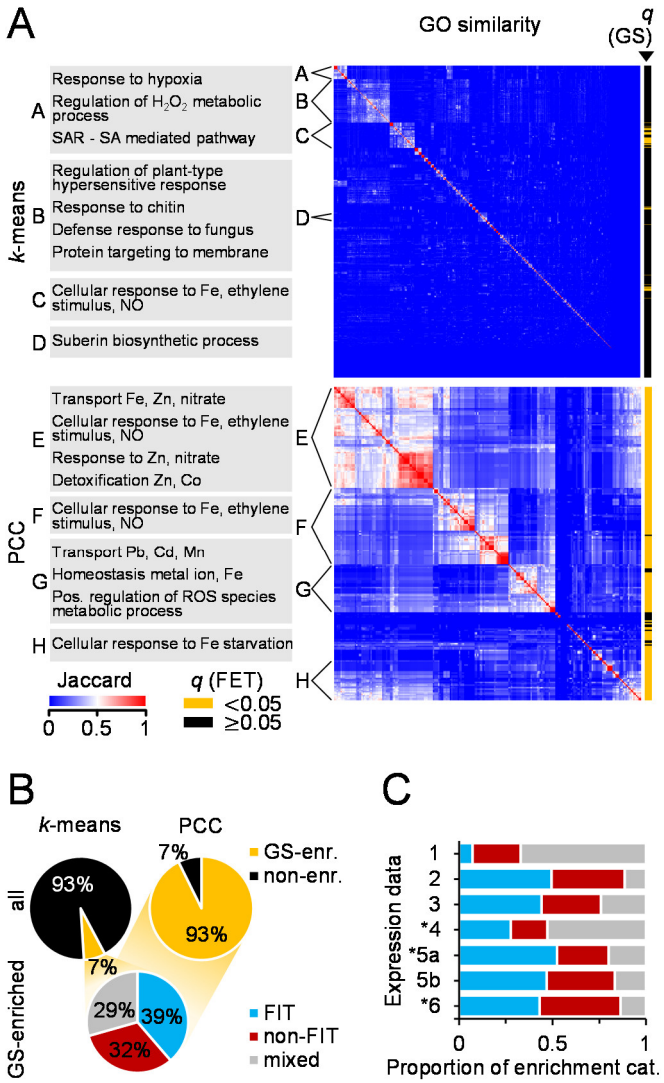


Figure 2. Characterization of the defined co-expression clusters by GO terms, -Fe GS gene content and FIT-dependent/FIT-independent gene content.
A: Heatmap of GO similarity between co-expression clusters from *k*-means clustering (top, n=985 clusters) and GS gene correlation (PCC; bottom, n=238), containing up-regulated genes (Figure 1C; up- and down-regulated genes: Supplemental Figure S2A). Clusters were grouped by hierarchical clustering and superclusters (A-F) were defined as groups of >20 clusters that have a within-mean Jaccard Index significantly higher than the mean Jaccard Index of all clusters. Enriched GO terms shared by ≥75% (*k*-means) and ≥90% (PCC) of the clusters in each supercluster are shown (left). Co-expression clusters enriched for -Fe GS genes are designated (yellow, right). **B:** Proportions of all *k*-means (top left) and PCC (top right) co-expression clusters in which -Fe GS genes are significantly over-represented (yellow). Of those (bottom), the proportion enriched for FIT-dependent genes (FIT, blue), FIT-independent genes (non-FIT, red) or for both (mixed, gray) was calculated. **C:** Proportion of enrichment categories FIT, non-FIT, and mixed clusters found using each expression data combination (as in Figure 1C). 1: -Fe time course, 2: time course + root zones, 3: time course + abiotic stresses, 4: time course + hormone treatments, 5a: time course + abiotic stresses + hormones, 5b: as 5a, genotoxic stress deleted, 6: time course + abiotic stresses + developmental data. *PCC clusters only. All enrichment analyses: FET, q<0.05.

A machine learning approach to model regulation of -Fe responsive co-expression clusters

The machine learning algorithm Random Forest (RF) has been successfully used to model stress transcriptional response using *cis*-regulatory sequences in plants (Zou et al., 2011; Deng et al., 2017; Uygun et al., 2017). Here, for each co-expression cluster, we used pCREs (enriched *k*-mers in putative promoter sequences; see **Methods**) to build a RF model that classifies genes as belonging to the cluster in question or as a non-responsive gene (see **Methods**). The pCREs from models performing above a defined threshold (F1≥0.7; see **Methods**) were then considered further. Out of 1,959 co-expression clusters, 28% of the models passed the performance threshold, 60% performed poorly, and for 12% no model could be built due to small size (median size=2 genes; **Supplemental Figure S3A, B**). Poor performing models (median F1=0.62) were mostly for small clusters (median size=12) (**Supplemental Figure S3B**; **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 co-regulated but not at the transcriptional level (e.g. post-translationally controlled). Interestingly, of the 28% of clusters with models above the threshold, only 36% were GS-enriched clusters (**Supplemental Figure S3A**). Nonetheless, models built for GS-enriched clusters (median F1=0.68) tended to perform better than models built for non-enriched clusters (median F1=0.65; Mann-Whitney U, *p*<2.358e-09; **Figure 3A-C**).

Good model performance indicates that genes in a cluster are more likely co-regulated, and, because pCREs were used to build the model, these pCREs are likely the regulatory sequences contributing to the co-regulation. Taken together, we identified 5,639 pCREs enriched in promoters of -Fe-responsive genes that may be predictive of -Fe-induced up- or down-regulation. 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. Known -Fe CREs from Arabidopsis and also from grasses, for example E-/G-boxes (bHLH TF binding sites) and IDE1, will serve as positive controls.

Identifying common pCREs across co-expression clusters

We expect true -Fe response CREs to be: (i) important for building models with good performance in predicting -Fe response, and (ii) reliably identified in co-expression clusters with similar gene content. Therefore, for each pCRE we calculated the proportion of clusters enriched for the pCRE and its average importance rank across those clusters (**Supplemental Table S3**). The importance rank of a pCRE was derived from an importance score for the pCRE in question from the trained RF models that reflects how useful a pCRE was for predicting -Fe response genes in a cluster. This allowed us to get a snapshot of pattern of presence and absence of important pCREs for genes correctly predicted (true positives (TP)) in co-expression clusters with good (**Figure 3D, E**) and poor (**Figure 3F**) performance. The pCREs were enriched in between 1 (0.6%) and 56 (35%) GS-enriched clusters and in between 1 (0.3%) and 54 (15%) non-enriched clusters, with 173 pCREs considered frequent pCREs (freq-pCREs, enriched in >5% of GS-enriched or non-enriched clusters) (**Supplemental Table S3**). Across GS-enriched clusters, pCREs tended to have higher proportions with higher importance ranks than across clusters that were not GS-enriched (Mann-Whitney U, *p*<2.2e-16; **Figure 4A, B**; **Supplemental Figure S4A, 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 non-enriched clusters (Mann-Whitney U, *p*<2.2e-16; **Supplemental Figure S3C**).

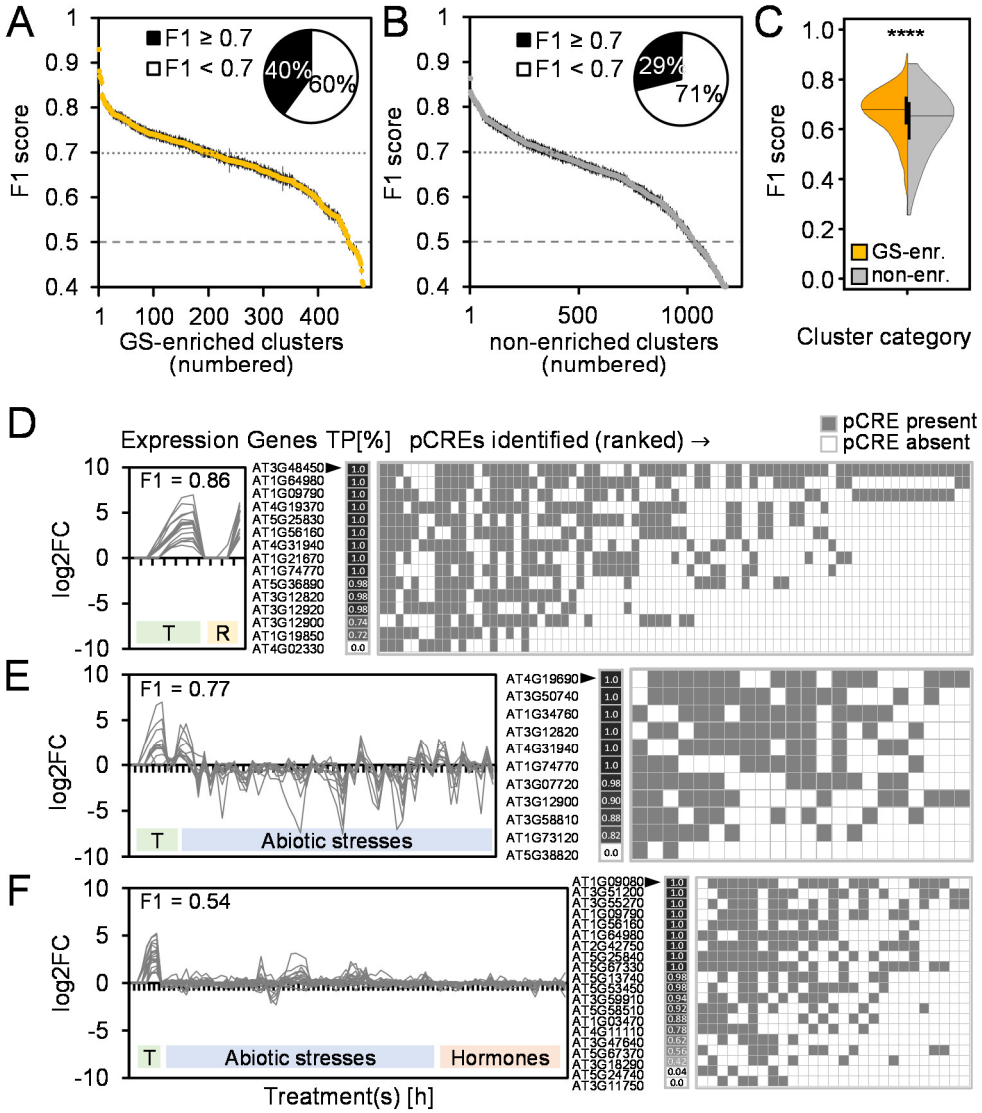


Figure 3. Performance of -Fe response RF prediction models.
A: F1 scores of all GS-enriched clusters (n=495). Inset: Proportions of well-performing (F1≥0.7) and poorly performing clusters among the GS-enriched clusters. **B:** F1 scores of all non-enriched clusters (n=1,240). Inset: Proportions of well-performing and poorly performing clusters among the non-enriched clusters. **C:** Mean F1 score distributions of all GS-enriched clusters (yellow) and non-enriched clusters (gray). Statistical analysis: Mann-Whitney U (**** p<2.358e-09). **D-F:** Example GS-enriched co-expression clusters with good (**D**, **E**; cluster IDs: 493, 823) and bad (**F**; cluster ID: 1297) model performance. (Left) Expression (log_2 fold-change: log2FC) profile of all genes in the co-expression cluster. (Center) Percent of times across RF replicates each gene was correctly predicted as -Fe-responsive (true positive (TP); black=100%, white=0%). (Right) pCREs sorted by importance rank (top ranked pCRE on the left) with heatmap designating when pCRE was present (gray) or absent (white) in a gene's promoter. T: -Fe treatment time course. R: -Fe-treated root zones 1-4. F1 score: harmonic mean of precision and recall, with 1=perfect prediction and 0.5=random guessing. Cluster IDs and details: Supplemental Table S2.

We next determined whether GS-enriched clusters are regulated by a different set of pCREs than non-enriched clusters. Out of the 5,639 pCREs, 15% (n=860) were unique to GS-enriched clusters, while 73% (n=4109) were unique to non-enriched clusters and 12% (n=670) were found in both GS-enriched and non-enriched clusters (**inset**, **Figure 4A**). This indicates that GS-enriched clusters and non-enriched 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 non-enriched clusters; **top**, **Supplemental Figure S4C**). This indicates that pCREs that were categorized as shared might not be equally important for regulating both GS-enriched and non-enriched clusters. Interestingly, unique GS-enriched freq-pCREs represented 59% (n=102) of the 173 freq-pCREs, while 35.5% (n=58) were unique non-enriched, and only 7.5% (n=13) freq-pCREs were shared between GS-enriched and non-enriched clusters, indicating that pCREs with high proportion are also the ones which seem to regulate almost exclusively either GS-enriched or non-enriched cluster functions, but not both (**inset**; **Figure 4A**; **bottom**, **Supplemental Figure S4C**). Furthermore, freq-pCREs tended to have higher importance ranks than non-frequent pCREs (Mann-Whitney U, p<1.924e-14; **Supplemental Figure S5D**). Together, this suggests that freq-pCREs could be particularly relevant for regulation of -Fe response mechanisms.

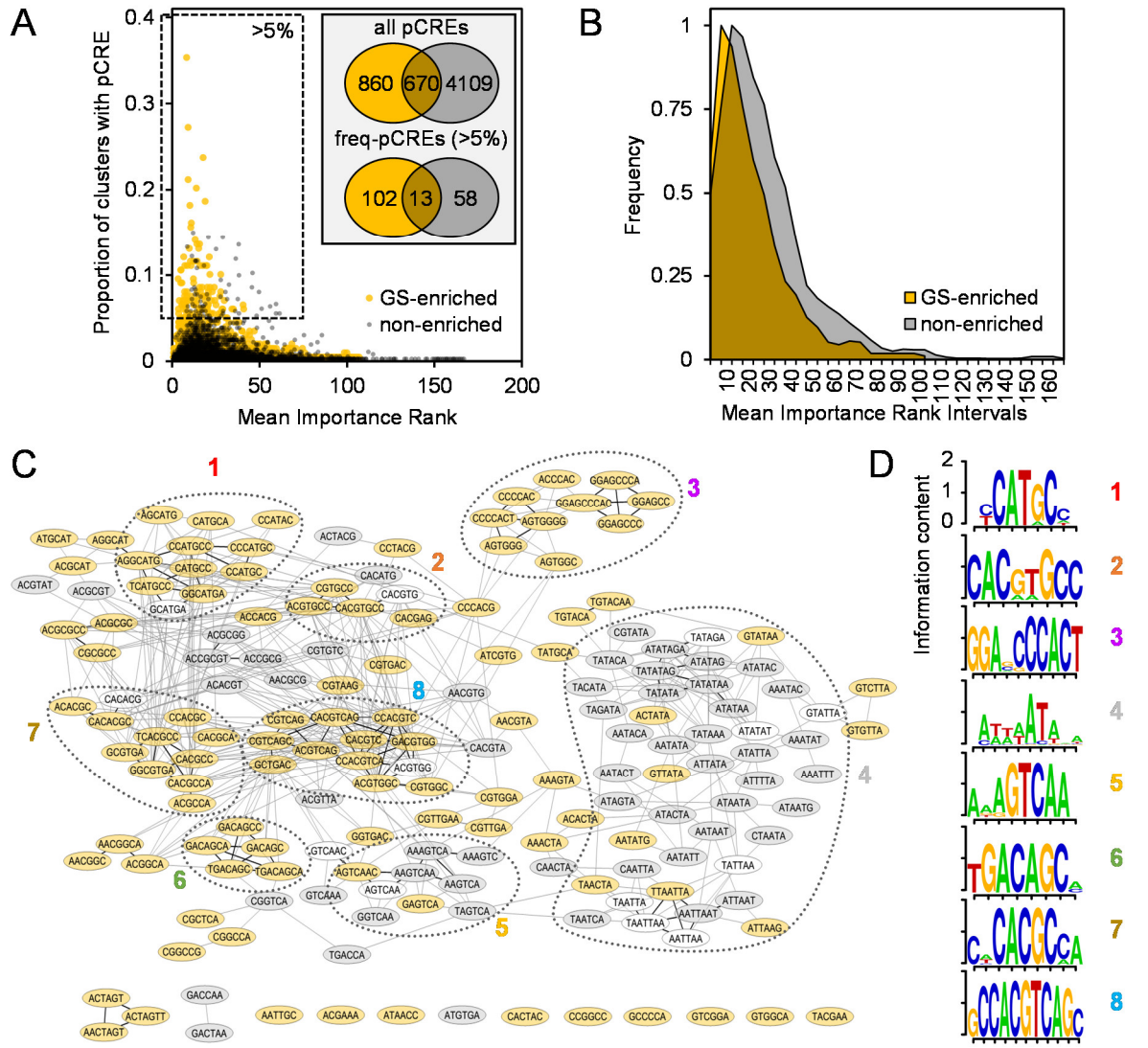


Figure 4. Analysis of pCREs predictive of -Fe co-expression clusters.
A: Proportion of GS-enriched (yellow) and non-enriched (gray) clusters in which each pCRE (total n=5,639) was identified (y-axis) and mean importance rank (1=most important) of that pCRE in those clusters (x-axis). Inset: Numbers of unique and shared pCREs of GS-enriched and non-enriched cluster categories. Upper: all 5,639 pCREs. Lower: pCREs identified in >5% of GS-enriched or non-enriched clusters (n=173; freq-pCREs). *(legend continued on next page)*

B: Frequency of normalized mean importance ranks across all pCREs in GS-enriched (yellow) and non-enriched (gray) clusters. **C:** Cytoscape network of the 173 freq-pCREs based on sequence similarity, where similar pCREs (nodes) are connected by edges representing pair-wise correlation (PCC) distance of freq-pCRE PWMs. Bold black edges: distance=0. Light gray edges: distance ≤0.22. Highly interconnected freq-pCREs were arranged in groups and numbered. Hierarchical clustering representation of PCC distances: Supplemental Figure S5E. Yellow filled: freq-pCRE unique for GS-enriched clusters, gray filled: freq-pCRE unique for non-enriched clusters, not filled: shared freq-pCRE. **D:** PWMs of merged freq-pCREs from the same group (as in 4C).

To characterize the freq-pCREs, we grouped them according to sequence similarity using pair-wise PCC distances of pCRE position weight matrices (PWM; see **Methods**). 62% (n=107) of all freq-pCREs could be placed into one of eight pCRE groups (**Figure 4C, D**; **Supplemental Figure S4E**). Freq-pCREs of the same group tended to be predictors of the same cluster category (GS-enriched/non-enriched).

In summary, we identified more than 100 -Fe pCREs that were reliably associated either exclusively to GS-enriched or non-enriched co-expression 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 examined the similarities between the -Fe pCREs and known TF binding motifs (TFBMs) from two sources (see **Methods**). Based on threshold similarities, we were able to match a specific TF and/or a specific TF family to each of the 173 freq-pCREs (see **Methods**; **Figure 5A**; **Supplemental Figure S5**). To gain an overview which TF families might be associated with GS-enriched clusters and how specific these TF families are, we asked which families contained over-represented numbers of TFs that likely bound pCREs from GS-enriched and non-enriched cluster categories. We found that most TF families were found with higher proportion in either GS-enriched clusters (14 TF families) or non-enriched clusters (12), while only four were similarly distributed between both categories (**Figure 5B**).

Most known -Fe regulators in Arabidopsis are bHLH TFs (FIT, subgroup Ib and IVc bHLH proteins, PYE, e.g. (Jakoby et al., 2004; Wang et al., 2007; Long et al., 2010; Palmer et al., 2013; Zhang et al., 2015)). BHLH and MYB TF families were identified, and even with higher proportion in GS-enriched clusters than in non-enriched clusters, which is indeed consistent with their role in -Fe response regulation. Other matching TF families over-represented in GS-enriched clusters were bZIP (FET, q<0.05), B3, TCP and NAC. Although a B3 TF (ABI3VP1 subfamily; IDEF1) and a NAC TF (IDEF2) are important regulators of Strategy II Fe acquisition in grasses (Kobayashi et al., 2007; Ogo et al., 2008), the role for these TF families in Strategy I non-grass plant species has not yet been described. In contrast, ARID, WRKY (both FET, q<0.05), Homeobox, and CAMTA TF families were matched more in non-enriched than GS-

enriched clusters, pointing towards roles during -Fe stress other than Fe uptake or homeostasis, in which GS genes are mostly involved.

Next, freq-pCRE-TFBM matches from GS-enriched clusters served to infer specific upstream regulators of -Fe-responsive modules. More than 50% freq-pCREs (60 out of 115) matched TFBMs of a specific TF (**Figure 5A**). Of those, 29 freq-pCREs shared perfect sequence similarity (PCC=1) to the TFBM, which were then of particular interest. From these perfect matches, 23 pCREs were unique for GS-enriched clusters. Example TF candidates for these 23 cases were FUS3 (an ABI3VP1/B3 TF), bHLH104, bZIP3, 16 and 42, TCP13 (PTF1), and FAR1 (**Supplemental Table S4A**). 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 bHLHs (Bailey et al., 2003), only 46 were available to be included in the analysis. Therefore, some TF families were likely under-represented 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 is collected, we might gain more biological insight into our -Fe pCREs.

Inferring upstream regulators of the -Fe response

Because we believe our genome-wide approach for identifying regulatory elements may shed light on areas of -Fe response that are less well understood, we next put our findings in context with open questions in the field. For example, the ABI3VP1/B3-type TF IDEF1, a key regulatory factor of -Fe responses in rice and barley roots, recognizes the CATGC core of IDE1 (Kobayashi et al., 2003; Kobayashi et al., 2005; Kobayashi et al., 2007; Kobayashi et al., 2009; Kobayashi et al., 2010). With ten of our freq-pCREs having an IDE1 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 S3**). This strongly suggests an important function for IDE1-like motifs in Arabidopsis. Arabidopsis AFLs (B3 family TFs ABI3/FUS3/LEC2), are the closest homologs of the rice IDEF1, and may bind to the IDE1-likes. In fact, ABI3 and FUS3 bind to RY-like elements (CATGCA), regulating FeS cluster subunit formation during seed maturation (Roschzttardtz et al., 2009). However, ABI3 or FUS3 functions during later developmental stages, particularly in the root during -Fe response, remain to be elucidated. Since the FUS3 TFBM matched our top most abundant IDE1-like (CATGCC; **Supplemental Table S4A**), 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 IDEF1 homolog in Strategy I plants.

Another -Fe response-related TF in rice and barley, IDEF2, 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 over-represented in GS-enriched clusters (**Figure 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.

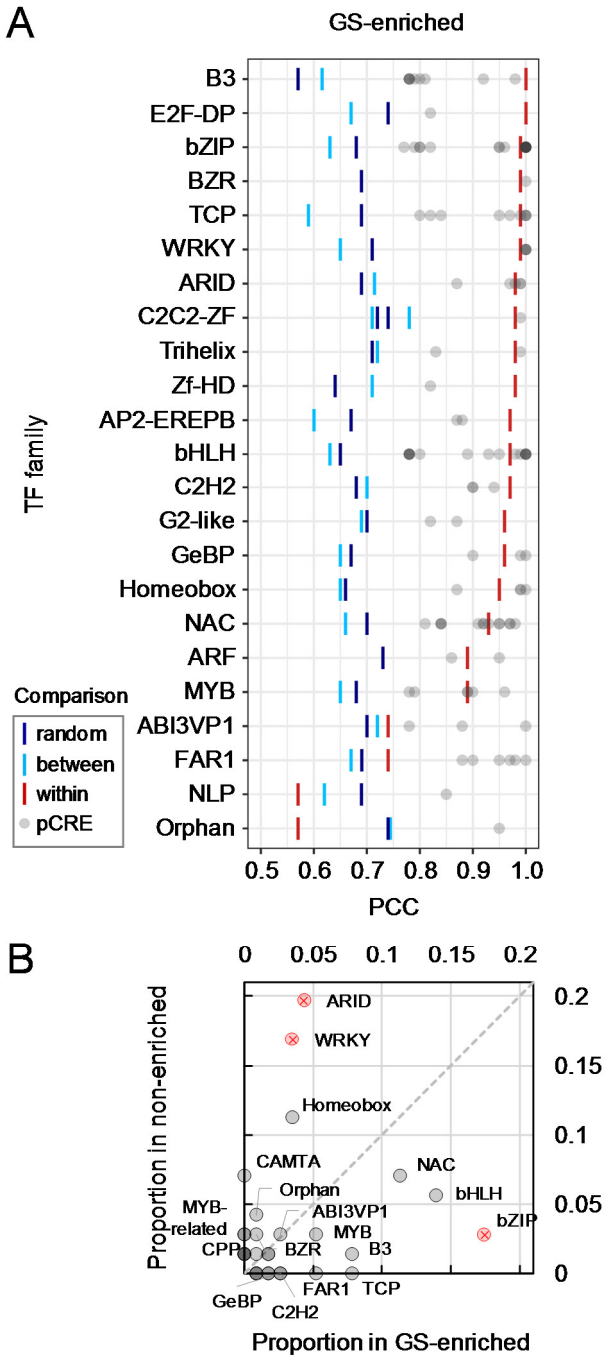


Figure 5. Similarity of freq-pCREs to *in vitro* TFBMs.
A: Significance of sequence similarity for freq-pCREs from GS-enriched clusters and the best matching known TFBM. Bars represent 95th percentile (PCC) significance thresholds for within TF family (red, pCRE sequence is more similar to a specific TFBM than other TFBMs from the same family), between TF families (light blue, pCRE sequence is more similar to a TFBM in a TF family than TFBMs from other TF families), or random (dark blue, pCRE sequence is more similar to a TFBM from a family than random 6-mers). Similarity of freq-pCREs from non-enriched clusters to TFBMs: Supplemental Figure S5. **B:** Proportion of TF family TFBMs (representing freq-pCRE matches meeting at least "between" threshold) in GS-enriched clusters (x-axis) and non-enriched clusters (y-axis). TFBM matches significantly over-represented (FET, q<0.05) in the GS-enriched or non-enriched cluster category are depicted in red and marked with "X". Dashed line marks theoretical position for TF family TFBMs with the same proportion in both categories.

One freq-pCRE (CGTGCC) perfectly matched to a bHLH104 TFBM (**Supplemental Table S4A**). bHLH104 binds to the promoters of subgroup Ib *BHLH* genes *BHLH38*/*39*/*100*/*101* (Zhang et al., 2015; Li et al., 2016), positively regulating Fe uptake. Consistently, we found CGTGCC in clusters containing *BHLH101* (*AT5G04150*; **Supplemental Table S2**).

Other freq-pCREs matched to known TFBMs from TFs with unknown roles in -Fe response. For example, bZIP TFBMs were significantly over-represented in GS-enriched clusters, but have no known direct roles in -Fe response. However, bZIP TFs are known regulators of the 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 response amongst

others by modulating root growth (Séguéla et al., 2008; Matiolli et al., 2011; Hsieh et al., 2012). Possible functions of bZIP TFs in response to -Fe stress should be explored in the future.

TCP13 (PTF1) and FAR1 TFBMs are two more examples for perfect freq-pCRE matches with yet unknown specific roles of the TFs during -Fe, although their specificity to GS-enriched clusters points towards important roles in regulating GS genes. TCPs are involved in plant development, but also act 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. TCP13 is involved in regulating responses to light shade signals through *PHYTOCHROME INTERACTING FACTORS* (*PIF*s) (Zhou et al., 2018). Interestingly, FAR1 and its homolog 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 phosphate starvation response, together with ethylene regulator 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 over-represented in non-enriched 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, WRKYs in general have already been connected to -Fe, for example as putative regulators of the coumarin transporter gene *PDR9* (Ito and Gray, 2006) and of *PYE* (Koryachko et al., 2015). Furthermore, WRKY46 negatively regulates the vacuolar Fe importer gene *VTL1*/*VITL1* (Gollhofer et al., 2011; Gollhofer et al., 2014; Yan et al., 2016). We found a WRKY TFBM (GTCAAC) in several non-enriched clusters containing down-regulated Fe-responsive genes, including the *VTL1* homolog *VTL5* (*AT3G25190*; **Supplemental Table S2**), indicating that some of the TFs matching non-enriched cluster pCREs might act as repressors of Fe excess genes.

In summary, many pCREs commonly found among GS-enriched clusters shared significant sequence similarity with known -Fe CREs, such as IDE1, or with binding sites of known -Fe-associated TF families, such as ABI3VP1/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 novel associations, such as the role of bZIPs or TCPs in -Fe responses. We assessed in the next paragraph in which specific -Fe response processes pCREs of particular interest, such as IDE1-likes, might be involved in.

Associating important pCREs with FIT-dependent or FIT-independent functions

After identifying novel potential regulators in the -Fe response, we pinpointed some of those which 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 response in each cluster. Using pCRE abundance (freq-pCREs) among GS-enriched clusters as the only criteria for selecting informative motifs for those clusters could result in missing motifs simply due to the fact that some co-expression clusters were more unique than others. This is supported by the fact that rare pCREs still can have a high importance rank (**Figure 4A**), meaning that 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 (defined as the minimum set of pCREs; min-pCREs) by building RF models iteratively with successively deleting the least important pCREs in each round (**Supplemental Figure S6**). Applying this approach to the 159 GS-enriched clusters resulted in a collective set of 615 min-pCREs. They were part of the minimum sets of between 1 (0.6%) and 48 (30%) GS-enriched clusters, with the IDE1-like CATGCC being 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 S3**). 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 typing IDE1-like ABI3VP1/B3 TFBMs to 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 (CACGCC; all **Supplemental Table S4A**) were also among the top ten most abundant min-pCREs (**Supplemental Table S3**), 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 co-expression clusters. More than 60% of the 159 GS-enriched clusters were classified as either FIT-dependent with a likely function in root iron acquisition (35% out of 159) or FIT-independent with either a function in internal Fe homeostasis in shoots and roots or in -Fe response regulation (28% out of 159; **Figure 6A**). We then calculated the proportion of min-pCREs (present in ≥5 GS-enriched clusters) in each cluster category (**Supplemental Table S5**). Interestingly, the two IDE1-like motifs, CATGCC, CCATGC and the related ABI3VP1/B3 TFBM matched ATGCAT, were predominantly identified in FIT-dependent clusters, but IDE2-like CACGCC had no preference for either FIT-dependent or FIT-independent clusters (**Figure 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 matched 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, elongation of root hairs, 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 down-regulation of respective GO terms under -Fe (Dinneny et al., 2008) (**Supplemental Figure S1**). During -Fe, several root hair-acting genes are co-expressed in a regulon which also contains *IRT2* (Ivanov et al., 2012), indicating a possible connection of ARF TFBM matched min-pCREs with these root hair processes.

In contrast, three bZIP TFBM matched min-pCREs, GTGGCA, CACGTC and CACTAC, were predominantly identified in FIT-independent clusters. As described in the previous section, bZIPs are involved in ABA signaling. ABA negatively regulates FIT-dependent Fe^{3+} reductase gene *FRO2* and Fe^{2+} importer gene *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 co-expression clusters of genes involved in Fe mobilization and translocation. Similarly, two TCP TFBMs, GACCAC and ACCCAC, 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 cascade. In summary, we can propose plausible roles for pCREs as TFBMs in FIT-dependent and FIT-independent processes.

Distribution and conservation of min-pCREs in co-expression cluster gene promoters

Next, to explore if min-pCREs displayed significant positional bias in the promoter regions of co-expressed 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) for all 615 min-pCREs (**Supplemental Figure S9**). Furthermore, we examined the non-coding 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 (**top**, **Figure 6C**). We investigated the distribution plots separately for ten selected min-pCREs: FIT-dependent ABI3VP1/B3 TFBM-matched min-pCREs (containing the two IDE1-likes), the IDE2-like (NAC TFBM match), FIT-independent bZIP TFBM matches, and bHLH TFBM matches (**Figure 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; **bottom**, **Figure 6C**), indicating that genes of such clusters are less likely to be transcriptionally co-regulated.

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 regulated by the min-pCRE (**Figure 6D**; **Supplemental Table S6A**). For example, CATGCC was found in 48 GS-enriched clusters, and 36 of those clusters (75%) included the min-pCRE-containing gene *MYB10*, while only four of those clusters (8%) included *MAPKKK16* (**second row right**, **Figure 6D**). 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 (Haudry et al., 2013) (**Supplemental Table S6B**). As expected, bHLH TFBM matched 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*, *AT3G07720* and *GRF11* (FIT targets; **Figure 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 *BTS* and *BTSL1*, two genes that negatively regulate Fe uptake by marking positive regulators (e.g. bHLH IVc TFs) 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 (**second row**, **Figure 6D**). Interestingly, both IDE1-likes were found in several genes encoding enzymes and TFs involved in coumarin biosynthesis (*CYP82C4*, *S8H*, *F6'H1*, *MYB72*/*MYB10*; (Kai et al., 2008; Murgia et al., 2011; Fourcroy et al., 2014; Schmid et al., 2014; Zamioudis et al., 2014; Rajniak et al., 2018; Siwinska et al., 2018)).

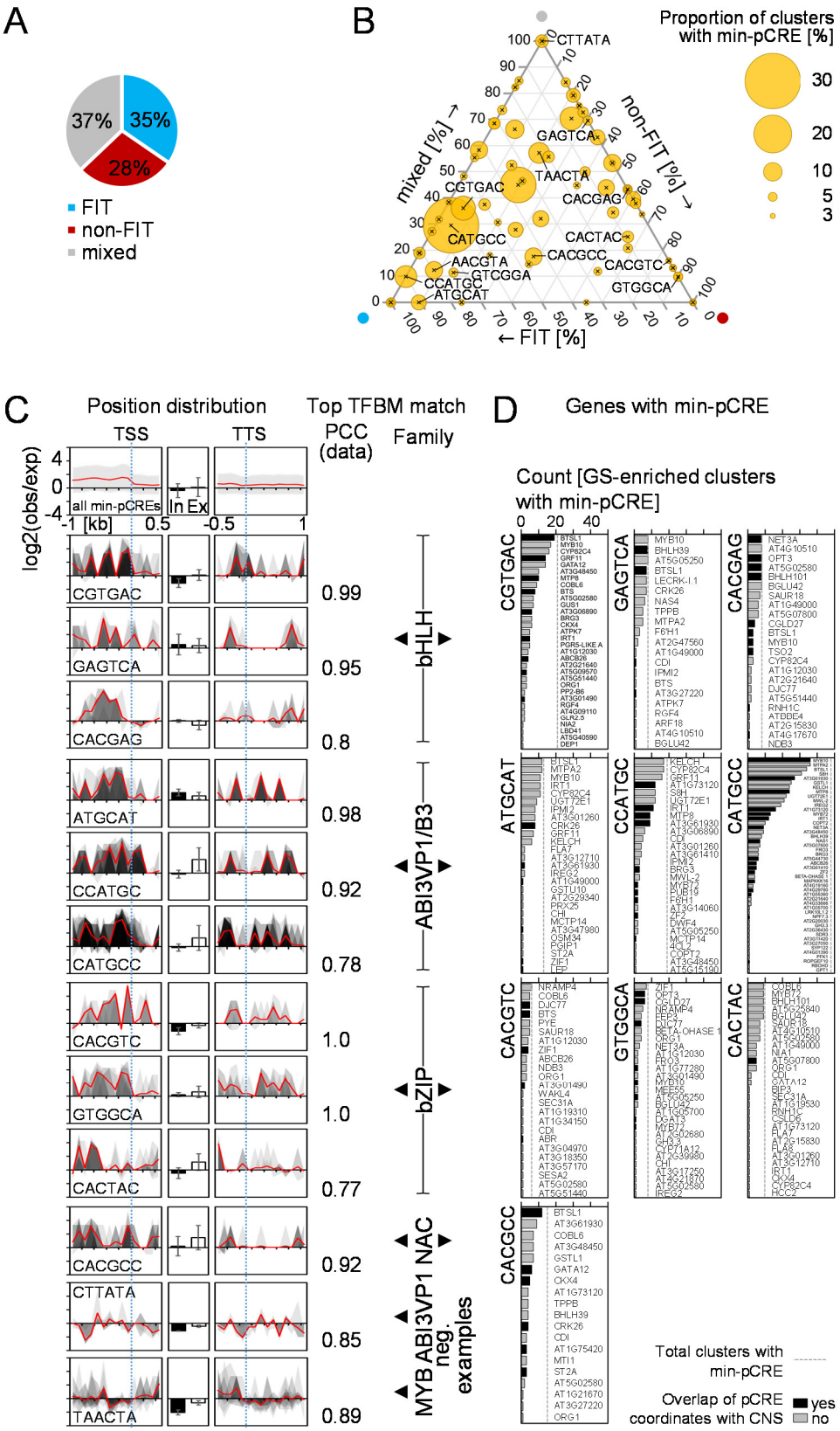


Figure 6. Characteristics of the most informative pCREs (min-pCREs).
A: Proportion of GS-enriched co-expression clusters enriched for FIT-dependent genes (FIT, blue), FIT-independent genes (non-FIT, red) or both (mixed, gray). **B:** Ternary plot including min-pCREs identified in >3% (n=5) GS-enriched 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 6C, D or mentioned in the main text. *(legend continued on next page)*

C: Positional bias of all (mean with standard deviation; top) and selected min-pCREs (below) in the putative promoter region (1st column), all introns (In) and all exons (Ex) (2nd column; mean with standard deviation), and in the putative non-coding region (3rd column). 1st and 3rd column: position distributions in all co-expression clusters with min-pCRE (gray areas) with mean distribution (red line). TFBM matches (PCC) for each min-pCRE are shown (4th column) and min-pCREs are sorted by TF family. log2(obs/exp): log_2 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 designated with black bars. A high-resolution image is available as Supplemental Figure S7.

We propose that IDE1 is important for synthesis of Fe chelators in response to low Fe conditions in both monocots and dicots (see Kobayashi et al., 2003; Kobayashi et al., 2005). The IDE2-like min-pCRE was located amongst others in *BTSL1* (within a CNS), *BHLH39*, *ORG1*, and a number of uncharacterized genes. While the role of IDE2 in the Strategy II Fe response has not been comprehensively explored, it is known to regulate expression of *OsYSL2*, a phloem Fe^{2+}-nicotianamine transporter (Kobayashi et al., 2003; Ogo et al., 2008). This putative involvement in phloem translocation of Fe suggests that the IDE2-like might preferably 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-matched 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. *NRAMP4*, *ZIF1*, *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 *CGLD27* (**third row**, **Figure 6D**), 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 more than 100 -Fe pCREs which, in addition to sharing significant sequence similarity to known TFBMs, were also part of the core sets of pCREs needed for robust 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 greatly improve our understanding of -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 enriched in promoters of co-expressed -Fe-responsive genes that were used as features to predict -Fe-responsive regulation of root-expressed genes on a genome-wide scale. Of those, 173 reliably predicted -Fe response genes of >5% of our defined co-expression clusters (freq-pCREs). Because most of those pCREs were either unique to co-expression clusters enriched for our gold standard Fe acquisition and homeostasis genes, or unique to co-expression clusters lacking those genes, we conclude that our approach had captured motifs specifically regulating different responses during -Fe. To take advantage of the publicly available *in vitro* TF binding information, we compared the freq-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 novel 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 co-expression clusters, highlighting the usefulness of this approach to pinpoint 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 co-expression clusters and their location bias in promoters of particular -Fe-responsive genes (**Figure 6B, D**). Our results provide evidence that B3 TFBM pCREs containing the IDE1 core motif CATGC are linked to coumarin synthesis, indicating that the function of IDE1-like motifs 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. Of the many important pCREs for -Fe response, many share significant similarity with TFBMs 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 co-expression 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 not have passed our significance threshold because of sensitivity issues with the microarray technology. Additionally, we restricted our analysis to the promoter region 1000 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 towards the importance of transcriptional regulation. However -Fe responses are also heavily controlled at the post-transcriptional and post-translational level (Lingam et al., 2011; Meiser et al., 2011; Sivitz et al., 2011; Selote et al., 2015; Zhang et al., 2015; Gratz 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. We suggest TFs of the bZIP, ABI3VP1/B3, NAC, and TCP families as upstream regulators of -Fe response 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; Garcia 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 *trans*-acting 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, e.g. higher nutritional value because of better Fe allocation and coping with unfavorable soil conditions.

Methods

Expression data processing and generation of multiple expression data combinations

Expression data (Affymetrix ATH1) from an -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 CEL format, 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), GSE5629-5634 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 (RMA) using the Bioconductor affy package (Gautier et al., 2004). The $\log_2$ 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 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 was combined with -Fe root zone expression data or ATGenExpress datasets in different combinations (**Figure 1C**) either including only genes up-regulated ("up") or all genes up- or down- regulated ("up & down") in ≥1 -Fe time point or root zone. This resulted in 12 different expression data combinations.

Co-expression clustering using *k*-means

To cluster genes with similar expression pattern, *k*-means clustering (Hartigan and Wong, 1979) was applied using the Euclidean distance as the similarity measure. 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. non-deterministic). Therefore, different *k* (25, 30, 35, 40, 50, 70, 80, 100) were tested and the clustering was repeated up to four times. We build machine learning models (see below) with all clusters generated from expression data combinations (DC) 1, 2, 3 and 5 (**Figure 1C**). To prevent confusion, we point out that the total number of *k*-means-generated clusters used to build models represents several repeated clustering events of always the same two sets of -Fe-responsive genes (up; up & 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.

Co-expression clustering by correlation with GS genes

To generate co-expression 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 data combination (DC; **Figure 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), we allowed a significance threshold 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 hone in on genes most likely to be co-regulated. In addition, we generated a second version of clusters with >50 genes, containing only the 10 genes with highest PCC. We build machine learning models (see below) with both versions and further used the results from the better performing version only. Percentiles used for each PCC-generated 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 & 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-generated clusters used for modeling represents repeated clustering events of the same two sets of -Fe-responsive genes (as described above).

The co-expression clusters: GO and GS/FIT-dependent/FIT-independent gene enrichment and GO/gene content similarity

Gene ontology (GO) associations for *A. thaliana* were downloaded from TAIR (ftp://ftp.arabidopsis.org/home/tair/Ontologies/Gene_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)≥1) in the -Fe time course data set was determined with a Fisher's exact test (FET, http://www.scipy.org, (Jones et al., 2001)), and *p*-values were corrected for multiple testing (=q) using the "qvalue" function in R (Storey, 2002) (**Supplemental Table S7**).

All co-expression clusters were tested for enrichment of GO terms as described above. The similarity of enriched GOs between co-expression clusters was assessed using the Jaccard Index (JI), or the intersection of GOs divided by the union of the GOs, where JI=1 if the exact same GOs were enriched in both co-expression clusters. Co-expression 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 co-expression 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 co-expression clusters enriched for (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_datasets/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 (http://biopython.org/ wiki/Biopython, (Cock et al., 2009)). Only one 6-mer for each reverse complement pair was kept (resulting in 2,080 6-mers). Genes were considered -Fe non-responsive if they were not significantly differentially expressed (abs(log2 FC) <0.4) during 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, (Sivitz 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 the **Methods** part.

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/MotifDiscovery, https://github.com/ShiuLab/ML_Pipeline). For the first step, the promoter sequences of the genes in co-expression clusters (positive set) were searched for enriched 6-mers in comparison to promoter sequences of non-responsive genes (negative set). 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 Random Forest (RF) algorithm (Pedregosa et al., 2011). To avoid building biased models, 50 models were generated for each co-expression cluster by randomly drawing from the negative set to generate balanced (i.e.

size positive set equals size of negative set) input datasets. A 10-fold cross-validation approach was used to train and test the models. Briefly, the balanced datasets 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 10^{th} (and successively trained on 1-8+10 and applied to the 9^{th}, 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/(FP+TP)) and recall (R=TP/(FN+TP)), where TP=true positive, FP=false positive, and FN=false negative. Only co-expression 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, which is a measure of node purity, where important pCREs separate positive from negative class genes well and low ranked pCREs are less informative. To determine how well the models predicted specific -Fe responsive genes, we calculated the percent of times each gene was correctly predicted (TP) out of the 50 balanced replicates.

pCRE sequence similarity

To assess sequence similarities between the 173 pCREs that were frequently identified (in >5%; freq-pCREs) in GS-enriched or non-enriched co-expression clusters with good model performance (F1≥0.7), sequence dissimilarity of pCRE position weight matrices (PWMs) was calculated by pair-wise PCC distance and a distance matrix was generated using the TAMO package (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 Figure S4E**). Due to group-wise 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 (**Figure 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 non-linear regression

The most informative pCREs of a co-expression 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 co-expression cluster, the pCREs used as features were step-wise 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 datasets. F1 scores were plotted against the number of pCREs (x) and a non-linear regression curve was fitted to the data points. An exponential recovery function

$$F1(x) = a\left(1 - e^{(-nx)}\right)$$

was found to best describe the data behavior. Starting values for variables a and n 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 Figure S6**).

pCRE similarity to TFBMs

In vitro binding data of Arabidopsis TFs to genomic DNA (DNA Affinity Purification Sequencing, DAP-seq, (O'Malley et al., 2016)) and TF binding data based on protein-binding microarray data or the TRANSFAC data base (Catalog of Inferred Sequence Binding Preferences, CIS-BP, (Weirauch et al., 2014)) were used (**Supplemental Table S4B**), with DAP-seq TFBMs used over CIS-BP TFBMs when the TF was present in both databases. PWMs of pCREs were compared to 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 (Uygun et al., 2017). A pCRE was similar to a specific TFBM (A) if the PCC between the pCRE and the TFBM was ≥95th percentile of PCCs between that TFBM and TFBMs from the same TF family. Alternatively a pCRE was similar to TFBMs from a TF family (B) if the PCC 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 PCC 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 S4C**. To determine if pCREs similar to specific TF families were enriched in GS-enriched versus non-enriched co-expression clusters, the percentage of pCREs similar to TFBMs (significance level A or B) from each TF family was calculated for each co-expression cluster category. Then, FET with multiple testing correction ($q \leq 0.05$) was used to determine if GS-enriched co-expression

clusters were enriched for TF families compared to non-GS-enriched co-expression clusters and vice versa.

Positional distribution of pCREs

To determine the positional distribution of the min-pCREs for each GS-enriched co-expression 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 1000 bp upstream to 500 bp downstream of the transcription start site (1000-TSS-500), using Motility (http://cartwheel. caltech.edu). For comparison, min-pCRE PWMs were also mapped to exons and introns, respectively, and to the region 500 bp upstream and 1000 bp downstream of the transcription termination site (500-TTS-1000). Arabidopsis sequences were downloaded from TAIR (ftp://ftp.arabidopsis.org/Sequences/blast_datasets/TAIR10_blastsets/). Positional distributions were calculated as described in (Uygun 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, in order 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 -Fe response genes (described above). The min-pCRE coordinates were then compared to the coordinates reported as conserved non-coding sequences (CNS) across nine species in the Brassicaceae family (Haudry et al., 2013) downloaded from the UCSC Genomics Bioinformatics website (http://mustang.biol.mcgill.ca:8885/download/A.thaliana/gff/AT_CNS.gff; **Supplemental Table S6B**).

Supplemental Material

Supplemental Figures S1-S7 and Supplemental Table S1 are included at the end of this manuscript. Supplemental Figures S8/S9 and Supplemental Tables S2-S7 are available as "Supplementary Material" online (https://doi.org/10.1101/603290).

Supplemental Figure S1. Complete GO enrichment analysis of -Fe-responsive genes.
Supplemental Figure S2. GO terms and -Fe GS gene enrichments of the defined co-expression clusters containing up-/down-regulated genes, and mean similarity within designated superclusters of up-regulated and up-/down-regulated genes.
Supplemental Figure S3. Co-expression cluster RF model performance of cluster category (GS-enriched and non-enriched) and cluster size.
Supplemental Figure S4. Comparison of the pCRE abundance and importance in GS-enriched clusters vs. non-enriched clusters and hierarchical clustering of freq-pCRE sequences.
Supplemental Figure S5. Significance of sequence similarity for freq-pCRE from non-enriched clusters and the best matching known TFBM.
Supplemental Figure S6. Example of a non-linear regression curve to determine the minimum set of pCREs for a co-expression cluster.
Supplemental Figure S7. High-resolution image of Figure 6.
Supplemental Figure S8. Expression plots of all well-performing co-expression clusters.
Supplemental Figure S9. Positional distribution plots of all 615 min-pCREs.
Supplemental Table S1. Robust -Fe-responsive GS genes (FIT-dependent/FIT-independent).
Supplemental Table S2. Detailed information of all generated co-expression clusters: input expression data combinations, algorithm and parameters used for clustering, enrichment of GS genes, FIT-dependent/FIT-independent/both genes, F1 score, gene content, all identified pCREs, min-pCREs.
Supplemental Table S3. List of all pCREs (n=5,639) identified in well-performing clusters (GS-enriched and non-enriched).
Supplemental Table S4. A: List of most relevant pCREs (freq-pCREs and min-pCREs) and their similarity to DAP-seq and CIS-BP TFBMs. **B:** DAP-seq and CIS-BP TFBMs used in this study. **C:** TF family 95th percentiles of within, between and random PCC thresholds determining the pCRE-TFBM similarity.
Supplemental Table S5. Association of min-pCREs to FIT-dependent, FIT-independent or both cluster categories.

Supplemental Table S6. A: List of all 615 min-pCREs with total counts of GS-enriched clusters and genes having the min-pCRE. **B:** Overlap of min-pCRE coordinates with Brassicaceae conserved non-coding sequences (CNS).

Supplemental Table S7. *p*- and *q*-values of complete GO enrichment analysis of -Fe-responsive genes (**Supplemental Figure S1**).

Acknowledgements

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. Funding from the German Research Foundation through the DFG International Research Training group 1525 to P.B. is greatly acknowledged. C.B.A. was supported in part by an NSF Graduate Research Fellowship. This work was partly supported by grants to S.-H.S. from the US National Science Foundation IOS-1546617, DEB-1655386, and DGE-1828149, and the US Department of Energy (DOE) Great Lakes Bioenergy Research Center (DOE Office of Science BER DE-SC0018409). This work received funding from Germany´s Excellence Strategy, EXC 2048/1, Project ID: 390686111.

Literature Cited

Ali MA, Azeem F, Nawaz MA, Acet T, Abbas A, Imran QM, 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, Claeys H, Kochanke I, Vermeersch M, De Milde L, De Bodt S, Storme V, Skirycz A, Maurer F, Bauer P, Mühlenbock P, Van Breusegem F, Angenent GC, Immink RG, Inzé D (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, Smaczniak 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 U S A 107: 10296-10301

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

Balk J, Schaedler 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. Journal of the Royal Statistical Society. Series B (Methodological) 57: 12

Berardini TZ, Mundodi S, Reiser L, Huala E, Garcia-Hernandez M, Zhang P, Mueller LA, Yoon J, Doyle A, Lander G, Moseyko N, Yoo D, Xu I, Zoeckler B, Montoya M, Miller N, Weems

D, Rhee SY (2004) Functional annotation of the Arabidopsis genome using controlled vocabularies. Plant Physiol 135: 745-755
Blaby-Haas C, Merchant S (2013) Iron sparing and recycling in a compartmentalized cell. Current Opinion in Microbiology 16: 677-685
Boulard C, Fatihi A, Lepiniec L, Dubreucq B (2017) Regulation and evolution of the interaction of the seed B3 transcription factors with NF-Y subunits. Biochim Biophys Acta Gene Regul Mech 1860: 1069-1078
Briat JF, Rouached H, Tissot N, Gaymard F, Dubos C (2015) Integration of P, S, Fe, and Zn nutrition signals in Arabidopsis thaliana: potential involvement of PHOSPHATE STARVATION RESPONSE 1 (PHR1). Frontiers in Plant Science 6
Brumbarova 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 PJA, Antao T, Chang JT, Chapman BA, Cox CJ, Dalke A, Friedberg I, Hamelryck T, Kauff F, Wilczynski B, de Hoon MJL (2009) Biopython: freely available Python tools for computational molecular biology and bioinformatics. Bioinformatics 25: 1422-1423
Colangelo EP, Guerinot ML (2004) The essential basic helix-loop-helix protein FIT1 is required for the iron deficiency response. Plant Cell 16: 3400-3412
Curie C, Mari S (2017) New routes for plant iron mining. New Phytologist 214: 521-525
Davière 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) Recursive random forest algorithm for constructing multilayered hierarchical gene regulatory networks that govern biological pathways. PLoS One 12: e0171532
Dinneny JR, Long TA, Wang JY, 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, Sisó-Terraza P, Sudre D, Savirón M, Reyt G, Gaymard F, Abadía A, Abadia J, Alvarez-Fernández A, Briat JF (2014) Involvement of the ABCG37 transporter in secretion of scopoletin and derivatives by Arabidopsis roots in response to iron deficiency. New Phytol 201: 155-167
Garcia MJ, Romera FJ, Stacey MG, Stacey G, Villar E, Alcantara E, Perez-Vicente R (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
Goda H, Sasaki E, Akiyama K, Maruyama-Nakashita A, Nakabayashi K, Li W, Ogawa M, Yamauchi Y, Preston J, Aoki K, Kiba T, Takatsuto S, Fujioka S, Asami T, Nakano T, Kato H, Mizuno T, Sakakibara H, Yamaguchi S, Nambara E, Kamiya Y, Takahashi H, Hirai MY, Sakurai T, Shinozaki K, Saito K, Yoshida S, Shimada Y (2008) The AtGenExpress hormone and chemical treatment data set: experimental design, data evaluation, model data analysis and data access. Plant J 55: 526-542
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-Transporter1-Like proteins mediate iron homeostasis in Arabidopsis. PLoS One 9: e110468
Gordon DB, Nekludova L, McCallum S, Fraenkel E (2005) TAMO: a flexible, object-oriented framework for analyzing transcriptional regulation using DNA-sequence motifs. Bioinformatics 21: 3164-3165
Gratz R, Manishankar P, Ivanov R, Köster P, Mohr I, Trofimov K, Steinhorst L, Meiser J, Mai HJ, Drerup M, Arendt S, Holtkamp M, Karst U, Kudla J, Bauer P, Brumbarova T (2019) CIPK11-dependent phosphorylation modulates FIT transcription factor activity,

promoting Arabidopsis iron acquisition in response to calcium signaling. Developmental Cell 48
Hantzis LJ, Kroh GE, Jahn CE, Cantrell 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. Journal of the Royal Statistical Society. Series C (Applied Statistics) 28: 9
Haudry A, Platts AE, Vello E, Hoen DR, Leclercq M, Williamson RJ, Forczek E, Joly-Lopez Z, Steffen JG, Hazzouri KM, Dewar K, Stinchcombe JR, Schoen DJ, Wang XW, Schmutz J, Town CD, Edger PP, Pires JC, Schumaker KS, Jarvis DE, Mandáková T, Lysak MA, van den Bergh E, Schranz ME, Harrison PM, Moses AM, Bureau TE, Wright SI, Blanchette M (2013) An atlas of over 90,000 conserved noncoding sequences provides insight into crucifer regulatory regions. Nature Genetics 45: 891-U228
Haydon MJ, Kawachi M, Wirtz M, Hillmer S, Hell R, Krämer U (2012) Vacuolar nicotianamine has critical and distinct roles under iron deficiency and for zinc sequestration in Arabidopsis. Plant Cell 24: 724-737
Heyndrickx KS, Van de Velde J, Wang CM, Weigei D, Vandepoele K (2014) A functional and evolutionary perspective on transcription factor binding in Arabidopsis thaliana. Plant Cell 26: 3894-3910
Hong S, Kim SA, Guerinot ML, McClung CR (2013) Reciprocal interaction of the circadian clock with the iron homeostasis network in Arabidopsis. Plant Physiology 161: 893-903
Hsieh 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 pleiotropic drug resistance transporter PDR9 confers resistance to auxinic herbicides. Plant Physiol 142: 63-74
Ivanov R, Brumbarova 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 (BHLH029) is required for induction of iron mobilization genes in Arabidopsis thaliana. FEBS Lett 577: 528-534
Jeong 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, al. e (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 feruloyl CoA by a 2-oxoglutarate-dependent dioxygenase in Arabidopsis thaliana. Plant J 55: 989-999
Kakei 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. Rice (N Y) 6: 22
Khan MA, Castro-Guerrero NA, McInturf SA, Nguyen NT, Dame AN, Wang J, Bindbeutel RK, Joshi T, Jurisson SS, Nusinow DA, Mendoza-Cozatl DG (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
Kilian J, Whitehead D, Horak J, Wanke D, Weinl S, Batistic O, D'Angelo C, Bornberg-Bauer E, Kudla J, Harter 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
Kobayashi T, Itai RN, Ogo Y, Kakei Y, Nakanishi H, Takahashi M, Nishizawa NK (2009) The rice transcription factor IDEF1 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, Yoshihara T, Mori S, Nishizawa NK (2003) Identification of novel cis-acting elements, IDE1 and IDE2, of the barley IDS2 gene promoter conferring iron-deficiency-inducible, 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 IDEF1 and IDEF2. Ann Bot 105: 1109-1117

Kobayashi T, Ogo Y, Itai RN, Nakanishi H, Takahashi M, Mori S, Nishizawa NK (2007) The transcription factor IDEF1 regulates the response to and tolerance of iron deficiency in plants. Proc Natl Acad Sci U S A 104: 19150-19155

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

Koryachko A, Matthiadis A, Muhammad D, Foret 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

Lanquar V, Lelièvre F, Bolte S, Hamès C, Alcon C, Neumann D, Vansuyt G, Curie C, Schröder A, Krämer U, Barbier-Brygoo H, Thomine S (2005) Mobilization of vacuolar iron by AtNRAMP3 and AtNRAMP4 is essential for seed germination on low iron. EMBO J 24: 4041-4051

Larkin MA, Blackshields G, Brown NP, Chenna R, McGettigan PA, McWilliam H, Valentin F, Wallace IM, Wilm A, Lopez R, Thompson JD, Gibson TJ, Higgins DG (2007) Clustal W and clustal X version 2.0. Bioinformatics 23: 2947-2948

Lei GJ, Zhu XF, Wang ZW, Dong F, Dong NY, Zheng SJ (2014) Abscisic acid alleviates iron deficiency by promoting root iron reutilization and transport from root to shoot in Arabidopsis. Plant Cell Environ 37: 852-863

Li WF, 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. Frontiers in Plant Science 8

Li WF, Schmidt W (2010) A lysine-63-linked ubiquitin chain-forming conjugase, UBC13, promotes the developmental responses to iron deficiency in Arabidopsis roots. Plant Journal 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: 2478-2493

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, Mohrbacher J, Brumbarova T, Potuschak T, Fink-Straube C, Blondet E, Genschik P, Bauer P (2011) Interaction between the bHLH transcription factor FIT and ETHYLENE INSENSITIVE3/ETHYLENE INSENSITIVE3-LIKE1 reveals molecular linkage between the regulation of iron acquisition and ethylene signaling in Arabidopsis. Plant Cell 23: 1815-1829

Liu Y, Xie Y, Wang H, Ma X, Yao W (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 POPEYE 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 2.0.6

Mai 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, Barberini ML, Aptekmann AA, Muschietti JP, Nadra AD, Dunand C, Cho HT,

Estevez JM (2017) Molecular link between auxin and ROS-mediated polar growth. Proc Natl Acad Sci U S A 114: 5289-5294
Marschner H, Römheld V (1994) Strategies of plants for acquisition of iron. Plant and Soil 165: 261-274
Matiolli CC, Tomaz JP, Duarte GT, Prado FM, Del Bem LE, Silveira AB, Gauer L, Corrêa LG, Drumond RD, Viana AJ, Di Mascio P, Meyer C, Vincentz M (2011) The Arabidopsis bZIP gene AtbZIP63 is a sensitive integrator of transient abscisic acid and glucose signals. Plant Physiol 157: 692-705
Meiser J, Lingam S, Bauer P (2011) Posttranslational regulation of the iron deficiency basic helix-loop-helix transcription factor FIT is affected by iron and nitric oxide. Plant Physiol 157: 2154-2166
Mendoza-Cózatl DG, Xie Q, Akmakjian GZ, Jobe TO, Patel A, Stacey MG, Song L, Demoin DW, Jurisson SS, Stacey G, Schroeder JI (2014) OPT3 is a component of the iron-signaling network between leaves and roots and misregulation of OPT3 leads to an over-accumulation of cadmium in seeds. Mol Plant 7: 1455-1469
Murgia I, Tarantino D, Soave C, Morandini P (2011) Arabidopsis CYP82C4 expression is dependent on Fe availability and circadian rhythm, and correlates with genes involved in the early Fe deficiency response. J Plant Physiol 168: 894-902
Müller M, Schmidt W (2004) Environmentally induced plasticity of root hair development in Arabidopsis. Plant Physiol 134: 409-419
Nicolas M, Cubas P (2016) TCP factors: new kids on the signaling block. Curr Opin Plant Biol 33: 33-41
O'Malley RC, Huang SSC, Song L, Lewsey MG, Bartlett A, Nery JR, Galli M, Gallavotti A, Ecker JR (2016) Cistrome and epicistrome features shape the regulatory DNA landscape. Cell 165: 1280-1292
Ogo Y, Itai RN, Nakanishi H, Kobayashi T, Takahashi M, Mori S, Nishizawa NK (2007) The rice bHLH protein OsIRO2 is an essential regulator of the genes involved in Fe uptake under Fe-deficient conditions. Plant J 51: 366-377
Ogo Y, Kobayashi T, Nakanishi Itai R, Nakanishi H, Kakei Y, Takahashi M, Toki S, Mori S, Nishizawa NK (2008) A novel NAC transcription factor, IDEF2, that recognizes the iron deficiency-responsive element 2 regulates the genes involved in iron homeostasis in plants. J Biol Chem 283: 13407-13417
Palmer CM, Hindt MN, Schmidt H, Clemens S, Guerinot ML (2013) MYB10 and MYB72 are required for growth under iron-limiting conditions. PLoS Genet 9: e1003953
Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, Blondel M, Prettenhofer P, Weiss R, Dubourg V, Vanderplas J, Passos A, Cournapeau D, Brucher M, Perrot M, Duchesnay E (2011) Scikit-learn: Machine Learning in Python. Journal of Machine Learning Research 12: 2825-2830
Phipson B, Lee S, Majewski IJ, Alexander WS, Smyth GK (2016) Robust hyperparameter estimation protects against hypervariable genes and improves power to detect differential expression. Ann Appl Stat 10: 946-963
Pitts RJ, Cernac A, Estelle M (1998) Auxin and ethylene promote root hair elongation in Arabidopsis. Plant J 16: 553-560
Rajniak J, Giehl RFH, Chang E, Murgia I, von Wirén N, Sattely ES (2018) Biosynthesis of redox-active metabolites in response to iron deficiency in plants. Nat Chem Biol 14: 442-450
Resentini F, Felipo-Benavent A, Colombo L, Blázquez MA, Alabadí D, Masiero S (2015) TCP14 and TCP15 mediate the promotion of seed germination by gibberellins in Arabidopsis thaliana. Mol Plant 8: 482-485
Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W, Smyth GK (2015) limma powers differential expression analyses for RNA-sequencing and microarray studies. Nucleic Acids Res 43: e47
Rodríguez-Celma J, Pan IC, Li W, Lan P, Buckhout TJ, Schmidt W (2013) The transcriptional response of Arabidopsis leaves to Fe deficiency. Front Plant Sci 4: 276
Roschzttardtz H, Fuentes I, Vásquez M, Corvalán C, León G, Gómez I, Araya A, Holuigue L, Vicente-Carbajosa J, Jordana X (2009) A nuclear gene encoding the iron-sulfur subunit

of mitochondrial complex II is regulated by B3 domain transcription factors during seed development in Arabidopsis. Plant Physiol 150: 84-95
Rose AB, Carter A, Korf I, Kojima N (2016) Intron sequences that stimulate gene expression in Arabidopsis. Plant Mol Biol 92: 337-346
Rose AB, Elfersi T, Parra G, Korf I (2008) Promoter-proximal introns in Arabidopsis thaliana are enriched in dispersed signals that elevate gene expression. Plant Cell 20: 543-551
Ruiz-Sola M, Rodríguez-Concepción M (2012) Carotenoid biosynthesis in Arabidopsis: a colorful pathway. Arabidopsis Book 10: e0158
Salomé PA, Oliva M, Weigel D, Krämer U (2013) Circadian clock adjustment to plant iron status depends on chloroplast and phytochrome function. Embo Journal 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, Vingron M, Schölkopf B, Weigel D, Lohmann JU (2005) A gene expression map of Arabidopsis thaliana development. Nat Genet 37: 501-506
Schmid NB, 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: 1109-1118
Schuler M, Keller A, Backes C, Philippar 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 Biology 11
Selote D, Samira R, Matthiadis 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
Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, Amin N, Schwikowski B, Ideker T (2003) Cytoscape: A software environment for integrated models of biomolecular interaction networks. Genome Research 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 Journal 66: 1044-1052
Sivitz AB, Hermand V, Curie C, Vert G (2012) Arabidopsis bHLH100 and bHLH101 control iron homeostasis via a FIT-independent pathway. PLoS One 7: e44843
Siwinska J, Siatkowska K, Olry A, Grosjean J, Hehn A, Bourgaud F, Meharg AA, Carey M, Lojkowska E, Ihnatowicz A (2018) Scopoletin 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. Journal of the Royal Statistical Society Series B-Statistical Methodology 64: 479-498
Séguéla 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
Tang LP, 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
Uygun 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
Uygun 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 Physiology 132: 796-804
Wang H, Wang HY (2015) Multifaceted roles of FHY3 and FAR1 in light signaling and beyond. Trends in Plant Science 20: 453-461

Wang HY, Klatte M, Jakoby M, Bäumlein 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

Weirauch MT, Yang A, Albu M, Cote AG, Montenegro-Montero A, Drewe P, Najafabadi HS, Lambert SA, Mann I, Cook K, Zheng H, Goity A, van Bakel H, Lozano JC, Galli M, Lewsey MG, Huang EY, Mukherjee T, Chen XT, Reece-Hoyes JS, Govindarajan S, Shaulsky G, Walhout AJM, Bouget FY, Ratsch G, Larrondo LF, Ecker JR, Hughes TR (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

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. Scientific Reports 6

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

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

Zamioudis C, Hanson J, Pieterse CM (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, Guerinot ML, Salt DE, Kochian LV, Vatamaniuk OK (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 HM, Li Y, Yao XN, Liang G, Yu DQ (2017) POSITIVE REGULATOR OF IRON HOMEOSTASIS1, OsPRI1, facilitates iron homeostasis. Plant Physiology 175: 543-554

Zhang J, Liu B, 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 DZ, An JX, Yin HJ, Fang S, Chu JF, Zhao YD, Li J (2018) TCP transcription factors regulate shade avoidance via directly mediating the expression of both PHYTOCHROME INTERACTING FACTORs and auxin biosynthetic genes. Plant Physiology 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 U S A 108: 14992-14997

Supplemental Figures and Tables

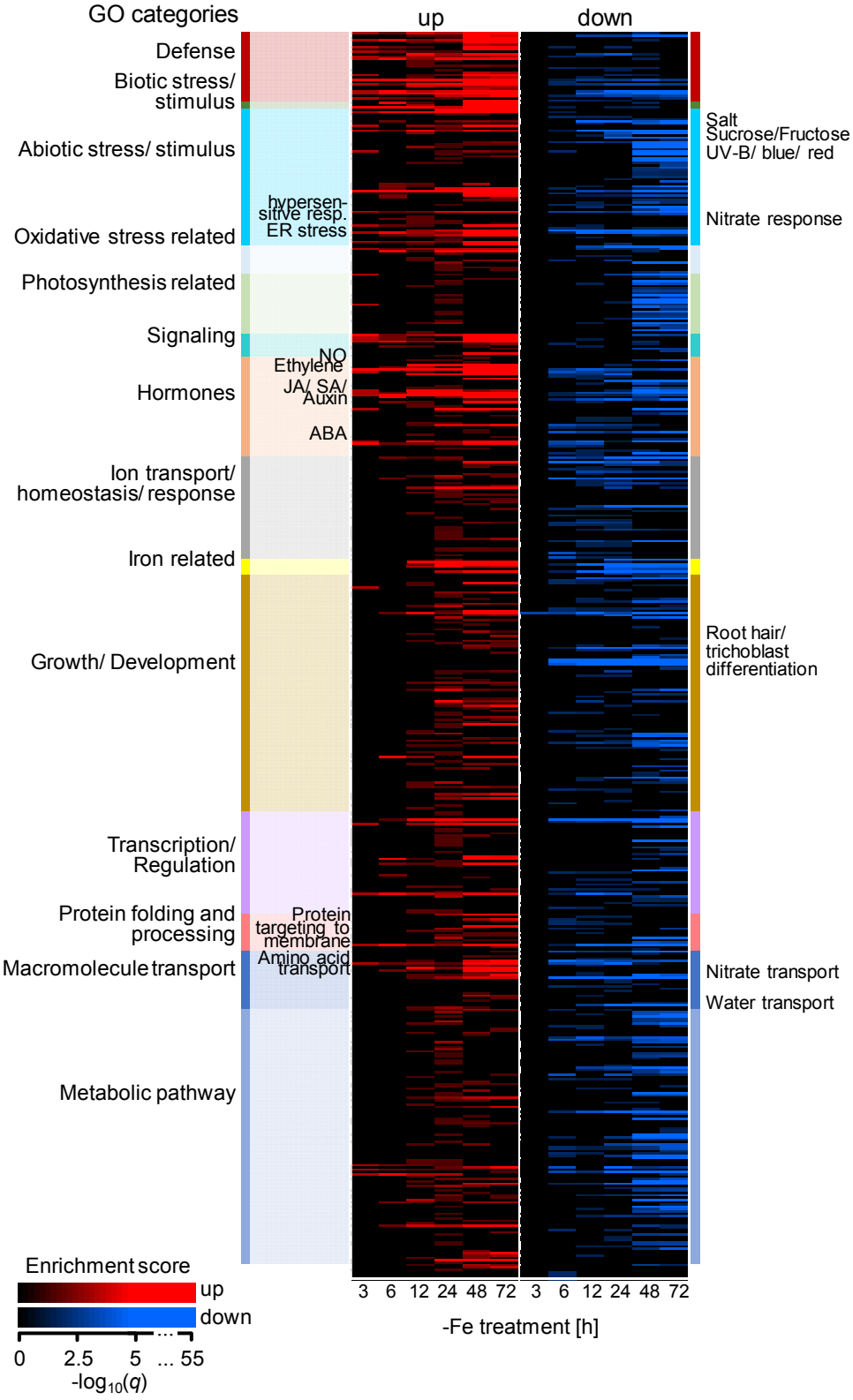


Figure S1. Complete GO enrichment analysis of -Fe-responsive genes.
Heatmap of enrichment (FET, q<0.05) of GO terms in genes that were significantly up- (red) or down-regulated (blue) (q<0.05) at ≥1 of 6 time points in -Fe-treated roots of 6 d-old seedlings (Dinneny et al., 2008). Differential regulation was defined as $\log_2$ fold-change (log2FC) >1 or <-1 (treatment vs. control). Selected highly enriched GO terms are indicated. Maximum color intensity threshold was set to $-\log_{10}(q)$=5 for visualization purposes. (Original p- and q-values: Supplemental Table S7).

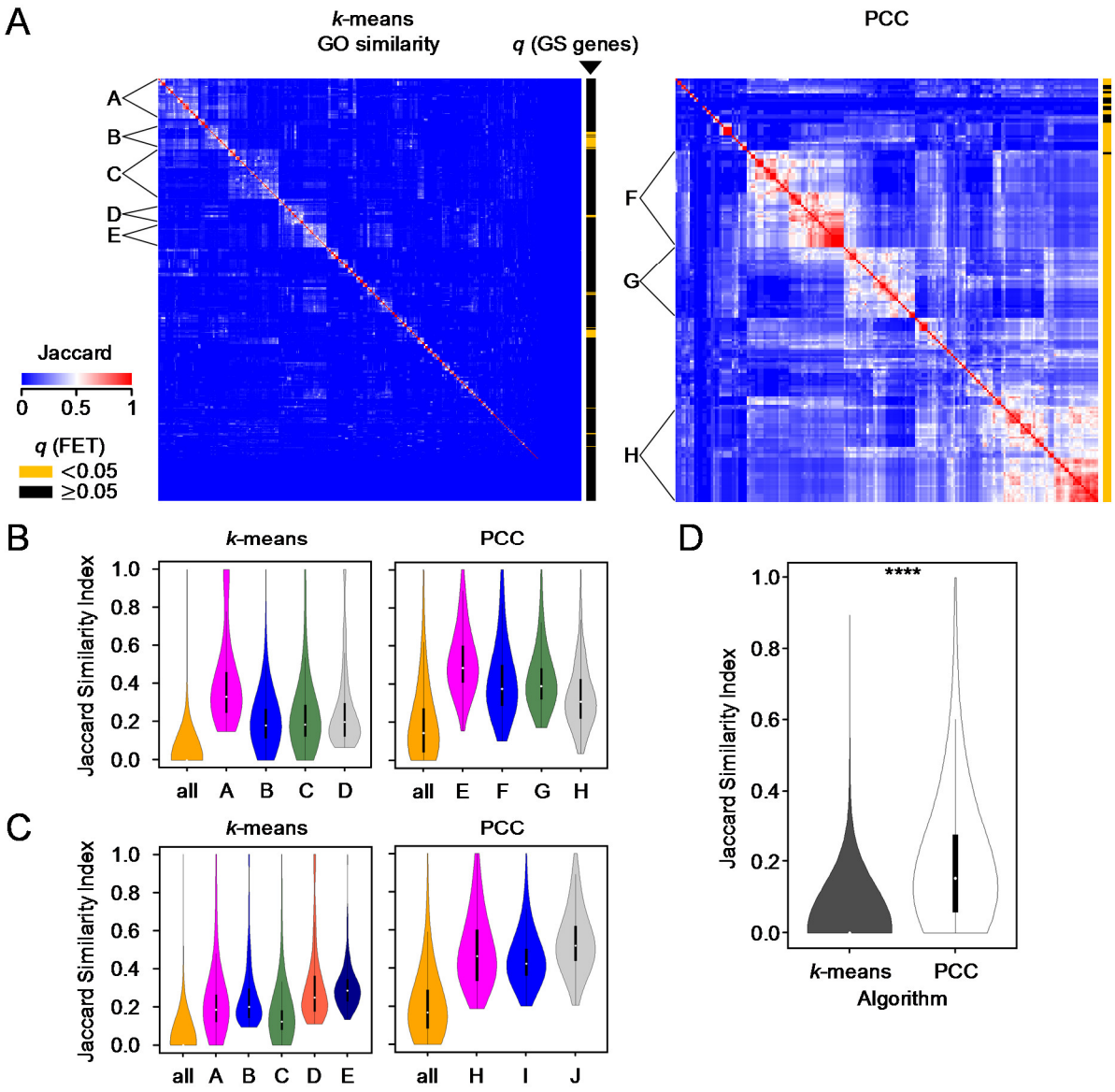


Figure S2. GO terms and -Fe GS gene enrichments of the defined co-expression clusters containing up-/down-regulated genes, and mean similarity within designated superclusters of up-regulated and up-/down-regulated genes.

A: Heatmap of GO similarity, measured using the Jaccard Index, between co-expression clusters from *k*-means clustering (left, n=535) and GS gene correlation (PCC; right, n=201), generated from expression data containing up- and down-regulated genes. Clusters were grouped by hierarchical clustering and superclusters (A-H) were defined as groups of >20 clusters that have a within-mean Jaccard Index significantly higher than the mean Jaccard Index of all clusters. Enriched GO terms shared by ≥75% (*k*-means) and ≥90% (PCC) of the clusters in each supercluster are shown (left). Co-expression clusters enriched for -Fe GS genes are designated (yellow, right). *k*-means superclusters: A, Transport nitrate, Response to nitrate. B, Cellular response to Fe/ethylene stimulus/NO. C, Response to red light. D, Regulation of H_2O_2 metabolic process, Response to hypoxia, SAR-SA mediated pathway. E, Defense response to fungus, Response to chitin. PCC superclusters: F, Cellular response to Fe/ethylene stimulus/NO. G, Transport Pb/Cd/Mn, Homeostasis metal ion/Fe, Pos. regulation of ROS species metabolic process. H, Transport Fe/Zn/nitrate, Cellular response to Fe/ethylene stimulus/NO, Response to Zn/nitrate, Detoxification Zn/Co. All enrichment analyses: FET, q<0.05.

B: Violin plots showing Jaccard Index distributions of all *k*-means and all PCC clusters ("all"), and of superclusters A-H, generated from expression data containing up-regulated genes (Figure 2A). Statistical analysis of each *k*-means or PCC supercluster compared to the respective "all": Mann-Whitney U (p<2.2e-16).

C: As in S2B, for superclusters shown in S2A. Mann-Whitney U of each supercluster compared to "all" (p<2.2e-16).

D: Gene content redundancy of -Fe co-expression clusters, corresponding to Figure 2A and S2A. Gene content similarity within all *k*-means and all PCC clusters. Clusters were compared in pair-wise manner with the Jaccard Index (1=identical gene content). Statistical analysis comparing *k*-means and PCC mean Jaccard Index: Mann-Whitney U (**** p<2.2e-16).

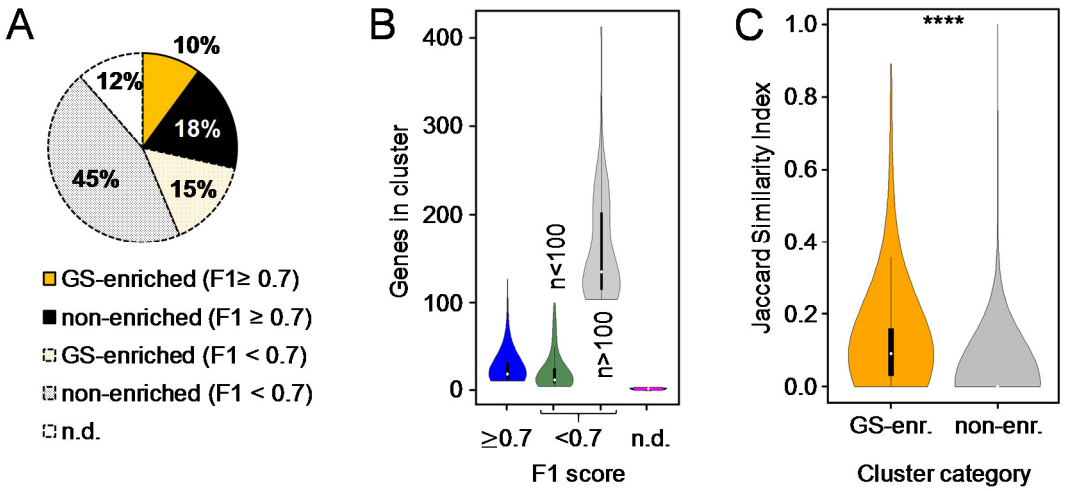


Figure S3. Co-expression cluster RF model performance of cluster category (GS-enriched and non-enriched) and cluster size.
A: Proportions of co-expression clusters with good (F1≥0.7; filled) and poor (dotted) RF performance organized by cluster category (GS-enriched; yellow, and non-enriched; black), and clusters which did not pass RF (n.d.; white).
B: Violin plots of cluster size distributions of all well-performing clusters (n=558), poor-performing clusters with <100 genes (n=1111), poor-performing clusters with >100 genes (n=66) and clusters which did not pass RF (n.d.; n=224).
C: Gene content similarity of well-performing GS-enriched (yellow; n=159) and non-enriched (gray; n=358) co-expression clusters (redundant correlation-based clusters not included; see Methods). Clusters within each category were compared in pair-wise manner with the Jaccard Index (1=identical gene content). Statistical analysis comparing mean Jaccard Index of both categories: Mann-Whitney U (**** p<2.2e-16). Cluster details: Supplemental Table S2.

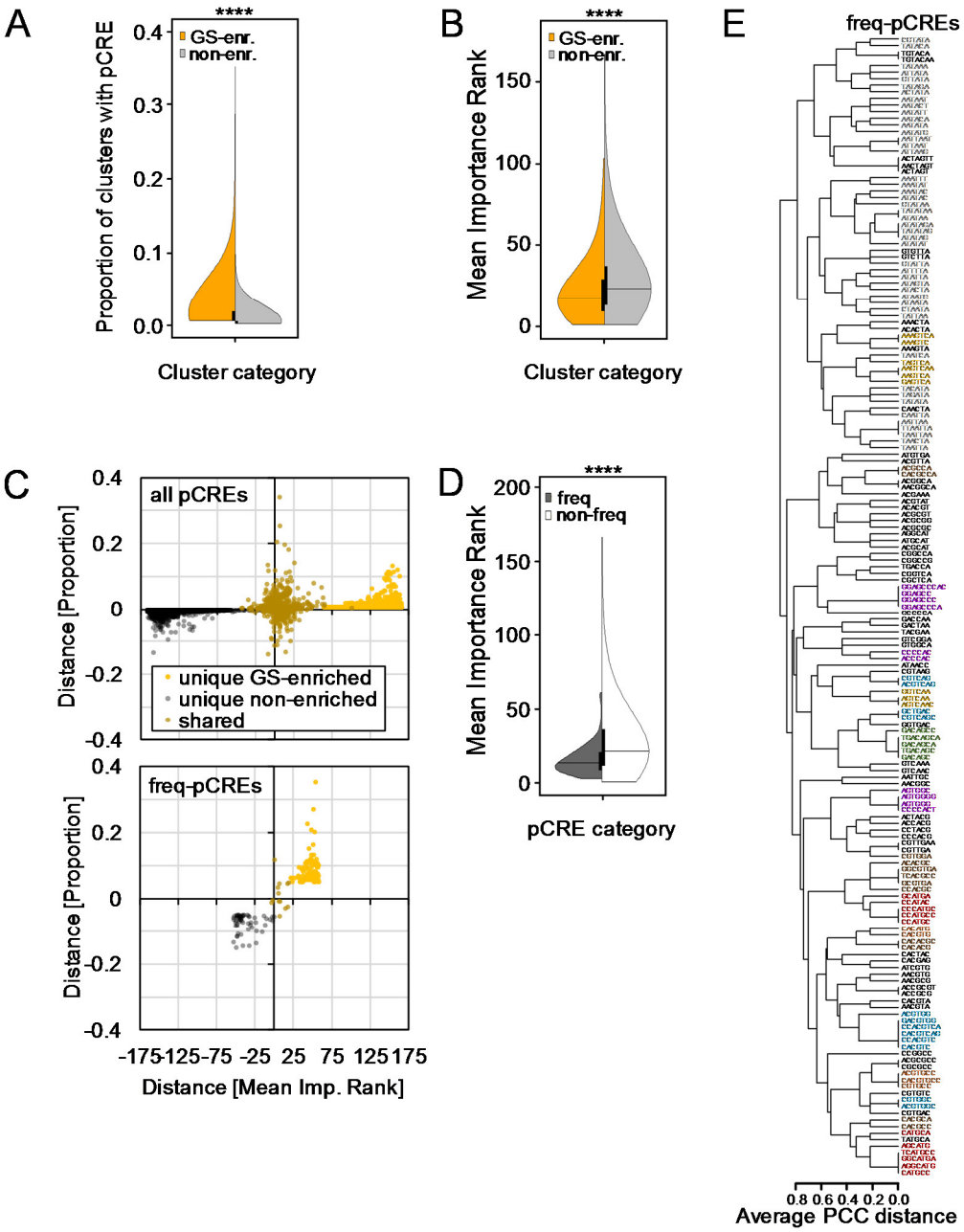


Figure S4. Comparison of the pCRE abundance and importance in GS-enriched clusters vs. non-enriched clusters and hierarchical clustering of freq-pCRE sequences.
A: Distribution of proportions of GS-enriched (yellow) and non-enriched (gray) clusters in which each pCRE (n=5,639) was identified. Statistical analysis: Mann-Whitney U (**** p<2.2e-16).
B: Distribution of the mean importance rank (1=most important) of each pCRE across GS-enriched and non-enriched clusters. Statistical analysis: Mann-Whitney U (**** p<2.2e-16).
C: Distance plots showing mean importance rank and proportion in GS-enriched vs. non-enriched clusters of all 5,639 pCREs (top) and 173 freq-pCREs (bottom). Distances were calculated as follows: prop[GS]−prop[non] and rank[non]−rank[GS] (positive value: pCRE has a higher proportion and/or is more important in GS-enriched clusters). Note that most shared pCREs in (top) become unique freq-pCREs in (bottom). In contrast to S4A, B, and Figure 4A, B, which depict shared pCREs twice, each pCRE/freq-pCRE is represented once. Therefore, a preference for GS-enriched or non-enriched clusters can be derived for shared pCREs/freq-pCREs.
D: Mean importance rank distributions of freq-pCREs (n=173) and non-freq-pCREs (n=5,466). Statistical analysis: Mann-Whitney U (**** p<1.924e-14).
E: PCC distances of freq-pCRE PWMs (hierarchical clustering). Due to group-wise averaging of PCC distances during hierarchical clustering, the algorithm produced skewed PCC distances of some pCRE pairs. Therefore, all pair-wise PCC distances were evaluated individually and corrected, if necessary, by hand in the freq-pCRE network (Figure 4C). freq-pCRE colors correspond to freq-pCRE network groups in Figure 4C, red: group1, orange: group2, pink: group3, gray: group4, yellow: group5, green: group6, brown: group7, blue: group8.

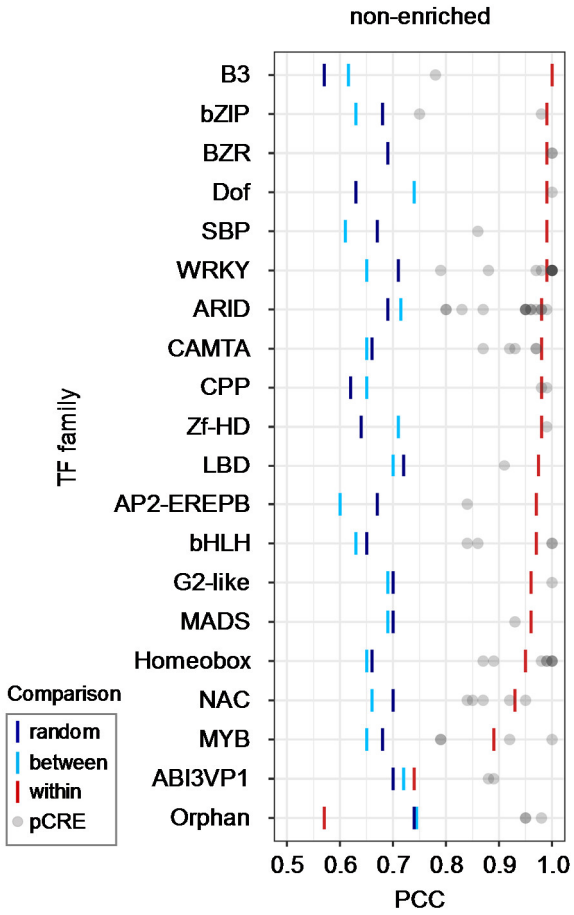


Figure S5. Significance of sequence similarity for freq-pCRE from non-enriched clusters and the best matching known TFBM.
Bars represent 95th percentile (PCC) significance thresholds for within TF family (red, pCRE sequence is more similar to a specific TFBM than other TFBMs from the same family), between TF families (light blue, pCRE sequence is more similar to a TFBM in a TF family than TFBMs from other TF families), or random (dark blue, pCRE sequence is more similar to a TFBM from a family than random 6-mers).

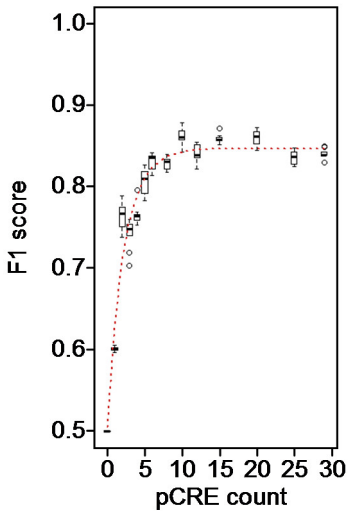


Figure S6. Example of a non-linear regression curve to determine the minimum set of pCREs for a co-expression cluster.
The minimum set was defined as the minimum number of pCREs (min-pCREs) needed for the model to perform at its maximum (F1 remains unchanged). Low-ranked pCREs or pCREs that were redundant (one is subset of the other) to a higher ranked pCRE were removed in step-wise manner and the predictive power of the remaining pCREs was tested in 10 RF runs. In this example, the top 10 pCREs were kept. Cluster ID: 513 (Supplemental Table S2).

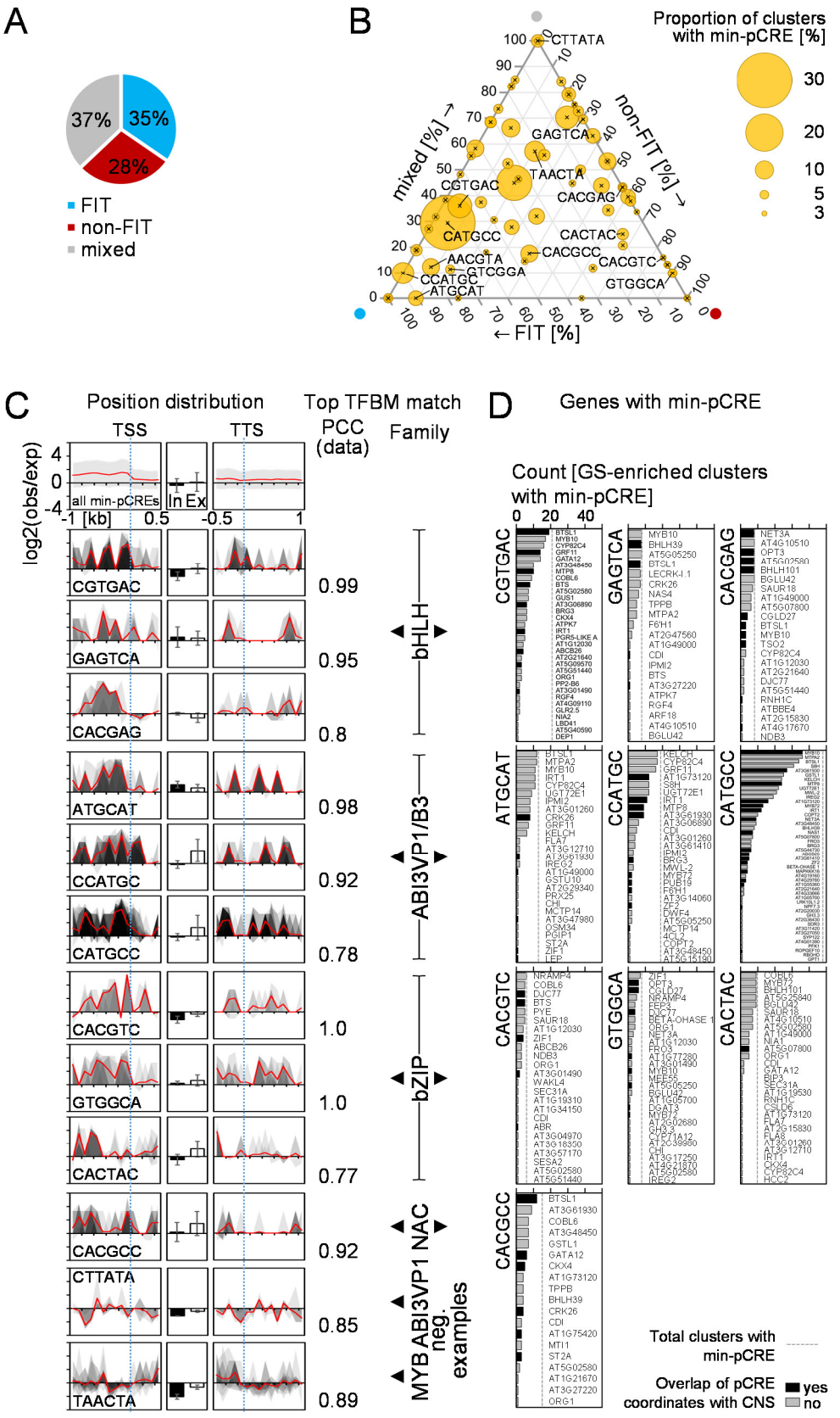


Figure S7. High-resolution image of Figure 6.

Table S1. Robust -Fe-responsive gold standard (GS) genes. -Fe GS genes are reliably up-regulated under -Fe in Arabidopsis roots or seedlings across several independent studies, some of which are transcriptomic and others are targeted. A subset of -Fe GS genes is FIT-dependent whereas other GS genes are FIT-independent.

AGI	Description	References
FIT-dependent		
AT1G34760	ROOT HAIR SPECIFIC 5 (RHS5), GENERAL REGULATORY FACTOR 11 (GRF11), 14-3-3 protein	1-6, 8, 9, 10
AT1G56160	MYB72, transcription factor	1-9, 11, 12
AT3G12820	MYB10, transcription factor	1-6, 8, 9, 11
AT1G73120	F-box/RNI superfamily protein	1, 2, 4, 5, 7-9
AT2G20030	RING/U-box superfamily protein	1, 3, 4, 6, 7, 9
AT3G07720	Galactose oxidase/kelch repeat superfamily protein	1-9
AT3G12900	SCOPOLETIN 8-HYDROXYLASE (S8H)	1-9, 13-15
AT3G50740	UDP-GLUCOSYL TRANSFERASE 72E1 (UGT72E1)	1, 2, 5-9
AT4G31940	CYTOCHROME P450, FAMILY 82, SUBFAMILY C, POLYPEPTIDE 4 (CYP82C4)	1, 2, 5-8, 16
AT4G30120	HEAVY METAL ATPASE (HMA3)	1, 2, 5, 6, 8
AT4G19690	IRON-REGULATED TRANSPORTER 1 (IRT1)	1-9, 17, 18
AT5G03570	IRON-REGULATED PROTEIN 2 (IREG2)	1-3, 5-9, 19
AT3G58810	METAL TOLERANCE PROTEIN A2 (MTPA2), putative zinc transporter	1-9, 20
AT4G33020	ZIP9, metal ion transporter	1, 2, 5, 6, 8
AT5G38820	Putative amino acid transporter	1-5, 8, 9
FIT-independent		
AT3G56980	OBP3-RESPONSIVE GENE 3 (ORG3), bHLH039, transcription factor	2, 3, 5-8, 21-23
AT5G04150	bHLH101, transcription factor	2, 3, 5, 6, 8, 21
AT3G47640	POPEYE (PYE), bHLH transcription factor	2, 3, 5-8
AT3G18290	EMBRYO DEFECTIVE 2454 (EMB2454), BRUTUS (BTS), putative E3 ligase	2, 3, 5-7, 24
AT1G74770	BRUTUS LIKE 1 (BTSL1)	2, 3, 6-8, 25
AT5G53450	OBP3-RESPONSIVE GENE 1 (ORG1)	2, 3, 5-8
AT1G56430	NICOTIANAMINE SYNTHASE 4 (NAS4)	2, 5-7, 26
AT1G23020	FERRIC REDUCTION OXIDASE 3 (FRO3)	2, 3, 5-8, 27
AT4G16370	OLIGOPEPTIDE TRANSPORTER 3 (OPT3)	2, 3, 5-8, 28, 29
AT5G67330	NATURAL RESISTANCE ASSOCIATED MACROPHAGE PROTEIN 4 (NRAMP4), metal ion transporter	2, 3, 5-7, 30
AT5G13740	ZINC INDUCED FACILITATOR 1 (ZIF1), small organic molecule transporter	2, 3, 5-8, 31
AT1G47400	FE-UPTAKE-INDUCING PEPTIDE3 (FEP3), IRONMAN 1 (IMA1)	2, 3, 5-8, 32, 33
AT5G05250	Hypothetical protein	2, 3, 5-8

[1]Colangelo and Guerinot, 2004, [2]Dinneny et al., 2008 (time course, up in at least 1 of 6 time points), [3]Buckhout et al., 2009, [4]García et al., 2010, [5]Long et al., 2010, [6]Yang et al., 2010, [7]Ivanov et al., 2012, [8]Sivitz et al., 2012, [9]Mai et al., 2016, [10]Yang et al., 2013, [11]Palmer et al., 2013, [12]Zamioudis et al., 2015, [13]Siwinska et al., 2018, [14]Tsai et al., 2018, [15]Rajniak et al., 2018, [16]Murgia et al., 2011, [17]Vert et al., 2002, [18]Jakoby et al., 2004, [19]Schaaf et al., 2006, [20]Arrivault et al., 2006, [21]Wang et al., 2007, [22]Yuan et al., 2008, [23]Naranjo-Arcos et al., 2017, [24]Selote et al., 2015, [25]Hindt et al., 2017, [26]Klatte et al., 2009, [27]Mukherjee et al., 2006, [28]Mendoza-Cózatl et al., 2014, [29]Zhai et al., 2014, [30]Lanquar et al., 2005, [31]Haydon et al., 2012, [32]Hirayama et al., 2018, [33]Grillet et al., 2018. Gene annotation based on the The Arabidopsis Information Resource (TAIR) 10.0 genome release, www.arabidopsis.org (accessed 03/2019).

Supplemental Literature Cited

Arrivault S, Senger T, Krämer U (2006) The Arabidopsis metal tolerance protein AtMTP3 maintains metal homeostasis by mediating Zn exclusion from the shoot under Fe deficiency and Zn oversupply. Plant Journal 46: 861-879

Buckhout TJ, Yang TJW, Schmidt W (2009) Early iron-deficiency-induced transcriptional changes in Arabidopsis roots as revealed by microarray analyses. Bmc Genomics 10

Colangelo EP, Guerinot ML (2004) The essential basic helix-loop-helix protein FIT1 is required for the iron deficiency response. Plant Cell 16: 3400-3412

Dinneny JR, Long TA, Wang JY, 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

García MJ, Lucena C, Romera FJ, Alcántara E, Pérez-Vicente R (2010) Ethylene and nitric oxide involvement in the up-regulation of key genes related to iron acquisition and homeostasis in Arabidopsis. Journal of Experimental Botany 61: 3885-3899

Goda H, Sasaki E, Akiyama K, Maruyama-Nakashita A, Nakabayashi K, Li W, Ogawa M, Yamauchi Y, Preston J, Aoki K, Kiba T, Takatsuto S, Fujioka S, Asami T, Nakano T, Kato H, Mizuno T, Sakakibara H, Yamaguchi S, Nambara E, Kamiya Y, Takahashi H, Hirai MY, Sakurai T, Shinozaki K, Saito K, Yoshida S, Shimada Y (2008) The AtGenExpress hormone and chemical treatment data set: experimental design, data evaluation, model data analysis and data access. Plant J 55: 526-542

Grillet L, Lan P, Li WF, Mokkapati G, Schmidt W (2018) IRON MAN is a ubiquitous family of peptides that control iron transport in plants. Nature Plants 4: 953-+

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

Hindt MN, Akmakjian GZ, Pivarski KL, Punshon T, Baxter I, Salt DE, Guerinot ML (2017) BRUTUS and its paralogs, BTS LIKE1 and BTS LIKE2, encode important negative regulators of the iron deficiency response in Arabidopsis thaliana. Metallomics 9: 876-890

Hirayama T, Lei GJ, Yamaji N, Nakagawa N, Ma JF (2018) The putative peptide gene FEP1 regulates iron deficiency response in Arabidopsis. Plant and Cell Physiology 59: 1739-1752

Ivanov R, Brumbarova 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 (BHLH029) is required for induction of iron mobilization genes in Arabidopsis thaliana. FEBS Lett 577: 528-534

Kilian J, Whitehead D, Horak J, Wanke D, Weinl S, Batistic O, D'Angelo C, Bornberg-Bauer E, Kudla J, Harter 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

Klatte M, Schuler M, Wirtz M, Fink-Straube C, Hell R, Bauer P (2009) The analysis of Arabidopsis nicotianamine synthase mutants reveals functions for nicotianamine in seed iron loading and iron deficiency responses. Plant Physiol 150: 257-271

Lanquar V, Lelièvre F, Bolte S, Hamès C, Alcon C, Neumann D, Vansuyt G, Curie C, Schröder A, Krämer U, Barbier-Brygoo H, Thomine S (2005) Mobilization of vacuolar iron by AtNRAMP3 and AtNRAMP4 is essential for seed germination on low iron. EMBO J 24: 4041-4051

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

Mai 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

Mendoza-Cózatl DG, Xie Q, Akmakjian GZ, Jobe TO, Patel A, Stacey MG, Song L, Demoin DW, Jurisson SS, Stacey G, Schroeder JI (2014) OPT3 is a component of the iron-signaling network between leaves and roots and misregulation of OPT3 leads to an over-accumulation of cadmium in seeds. Mol Plant 7: 1455-1469

Mukherjee I, Campbell NH, Ash JS, Connolly EL (2006) Expression profiling of the Arabidopsis ferric chelate reductase (FRO) gene family reveals differential regulation by iron and copper. Planta 223: 1178-1190

Murgia I, Tarantino D, Soave C, Morandini P (2011) Arabidopsis CYP82C4 expression is dependent on Fe availability and circadian rhythm, and correlates with genes involved in the early Fe deficiency response. J Plant Physiol 168: 894-902

Naranjo-Arcos MA, Maurer F, Meiser J, Pateyron S, Fink-Straube C, Bauer P (2017) Dissection of iron signaling and iron accumulation by overexpression of subgroup Ib bHLH039 protein. Scientific Reports 7

Palmer CM, Hindt MN, Schmidt H, Clemens S, Guerinot ML (2013) MYB10 and MYB72 are required for growth under iron-limiting conditions. PLoS Genet 9: e1003953

Rajniak J, Giehl RFH, Chang E, Murgia I, von Wirén N, Sattely ES (2018) Biosynthesis of redox-active metabolites in response to iron deficiency in plants. Nat Chem Biol 14: 442-450

Schaaf G, Honsbein A, Meda AR, Kirchner S, Wipf D, von Wirén N (2006) AtIREG2 encodes a tonoplast transport protein involved in iron-dependent nickel detoxification in Arabidopsis thaliana roots. Journal of Biological Chemistry 281: 25532-25540

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

Selote D, Samira R, Matthiadis 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

Sivitz AB, Hermand V, Curie C, Vert G (2012) Arabidopsis bHLH100 and bHLH101 control iron homeostasis via a FIT-independent pathway. PLoS One 7: e44843

Siwinska J, Siatkowska K, Olry A, Grosjean J, Hehn A, Bourgaud F, Meharg AA, Carey M, Lojkowska E, Ihnatowicz A (2018) Scopoletin 8-hydroxylase: a novel enzyme involved in coumarin biosynthesis and iron-deficiency responses in Arabidopsis. J Exp Bot 69: 1735-1748

Tsai HH, Rodríguez-Celma J, Lan P, Wu YC, Vélez-Bermúdez IC, Schmidt W (2018) Scopoletin 8-hydroxylase-mediated fraxetin production is crucial for iron mobilization. Plant Physiol 177: 194-207

Vert G, Grotz N, Dédaldéchamp F, Gaymard F, Guerinot ML, Briat JF, Curie C (2002) IRT1, an Arabidopsis transporter essential for iron uptake from the soil and for plant growth. Plant Cell 14: 1223-1233

Wang HY, Klatte M, Jakoby M, Bäumlein H, Weisshaar B, Bauer P (2007) Iron deficiency-mediated stress regulation of four subgroup Ib BHLH genes in Arabidopsis thaliana. Planta 226: 897-908

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

Yang TJ, Lin WD, Schmidt W (2010) Transcriptional profiling of the Arabidopsis iron deficiency response reveals conserved transition metal homeostasis networks. Plant Physiol 152: 2130-2141

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, Korteland J, Van Pelt JA, van Hamersveld M, Dombrowski N, Bai Y, Hanson J, Van Verk MC, Ling HQ, Schulze-Lefert P, Pieterse CM (2015) Rhizobacterial volatiles and photosynthesis-related signals coordinate MYB72 expression in Arabidopsis roots during onset of induced systemic resistance and iron-deficiency responses. Plant J 84: 309-322

Zhai Z, Gayomba SR, Jung HI, Vimalakumari NK, Piñeros M, Craft E, Rutzke MA, Danku J, Lahner B, Punshon T, Guerinot ML, Salt DE, Kochian LV, Vatamaniuk OK (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

Author contributions to Manuscript 1

Birte Schwarz

Designed the research and analyzed the data (except pCRE-TFBM similarity analysis). Prepared all figures and tables, with contribution to Figure 5A and Supplemental Figure S5. Wrote the manuscript, reviewed/edited the manuscript.

Christina Azodi

Wrote the machine learning pipeline and performed pCRE-TFBM similarity analysis. Prepared Figure 5A and Supplemental Figure S5. Helped with programming and data analysis. Acquired funding, contributed to writing of the manuscript, reviewed/edited the manuscript.

Shin-Han Shiu, Petra Bauer

Conceived and supervised the study, helped with the research plan (S.-H.S.), acquired funding, reviewed/edited the manuscript.

8 Manuscript 2

Novel insights into the protein interaction network regulating iron deficiency responses in Arabidopsis roots

Novel insights into the protein interaction network regulating iron deficiency responses in Arabidopsis roots

Birte Schwarz[1,q], Daniela M. Lichtblau[1,q], Christopher Endres[1], Christin Sieberg[1], Petra Bauer[1,2,*]

[1]Institute of Botany, Heinrich Heine University, Düsseldorf 40225, Germany
[2]Cluster of Excellence on Plant Science (CEPLAS), Heinrich Heine University, Düsseldorf 40225, Germany
[q]These authors contributed equally
***Corresponding author:** Petra Bauer, petra.bauer@hhu.de

Funding: This work received funding from the German Research Foundation grant through the DFG International Research Training group 1525 to P.B. and from Germany's Excellence Strategy, EXC 2048/1, Project ID: 390686111.

Author contributions: B.S., D.M.L., C.E., and C.S. performed the experiments and analyzed the data; B.S., D.M.L., and P.B. designed the experiments and supervised the research; B.S., D.M.L., and P.B. conceived the project; B.S. wrote the original draft with contributions of D.M.L. and C.E.; B.S., D.M.L., and P.B. reviewed/edited the article; P.B. acquired funding. P.B. agrees to serve as the author responsible for contact and ensures communication.

One sentence summary: E3 ligases BTSL1 and BTSL2 interact with bHLH proteins and with a small peptide that acts as their antagonist during Arabidopsis iron deficiency responses.

Highlights:

- A targeted yeast two-hybrid screen reveals several novel protein-protein interactions potentially functioning in Arabidopsis Fe deficiency responses
- A regulatory protein interactome emerges containing E3 ligases BTS/BTSL1/BTSL2, bHLH TFs and the small peptide FEP3
- FEP3 is introduced as inhibitor of BTSL1/BTSL2, acting upstream of ILR3, bHLH104 and PYE

Keywords: Fe deficiency, protein interaction network, regulation, signaling, BTSL1/2, E3 ligase, peptide, bHLH, yeast two-hybrid, Arabidopsis.

Abbreviations:

3AT	3-amino-1,2,4-triazole
aa	Amino acid
AD	Activation domain
BD	Binding domain
bHLH	Basic helix-loop-helix
BiFC	Bimolecular fluorescence complementation
GFP	Green fluorescent protein
GUS	β-Glucuronidase
HHE	Hemerythrin/HHE cation-binding motif
MA	Mugineic acid
mCherry	Second generation mRFP derivative
mRFP	Monomeric red fluorescent protein
OX	Over-expression
RT-qPCR	Reverse transcription quantitative PCR
SD	Standard deviation (Statistics) / Synthetic defined medium (Y2H)
TF	Transcription factor
WT	Wild type
Y2H	Yeast two-hybrid
YFP	Yellow fluorescent protein

Abstract

Iron (Fe) is an essential micronutrient for plants, acting as co-factor in all important metabolic processes, including photosynthesis. Because Fe is also cytotoxic when present in excess, plants must carefully balance internal Fe levels. Complex regulatory mechanisms coordinate Fe homeostasis transcriptionally and post-transcriptionally. More than 500 genes are induced upon Fe deficiency stress in Arabidopsis (*Arabidopsis thaliana*) roots, which are amongst others involved in Fe acquisition and different adaptations to Fe deficiency. Many of these genes are strongly co-regulated, but their interplay at the protein level is not well understood. Furthermore, the mechanism by which Fe demand is sensed and signaled is still obscure. To uncover novel functions and connections between Fe deficiency response proteins, we conducted a comprehensive yeast two-hybrid interaction screen. We tested 24 proteins with either unknown function, putative Fe sensing or signaling function, or important regulatory functions during Fe deficiency responses. The screen resulted in 14 novel and four known interactions. Together they form a root-specific putative protein interaction network in which we identified BRUTUS-LIKE 1 (BTSL1) as one network hub. BTSL1 is a homolog of BTS, a putative Fe sensor which negatively controls Fe uptake by targeting specific basic-helix-loop-helix (bHLH) transcription factors (TFs) for degradation. BTSL1 interacted amongst others with Fe uptake-regulating bHLH TFs (bHLH104, IAA-LEUCINE RESISTANT3 (ILR3) and POPEYE (PYE)), and with small proteins. BTS and BTSL2 also interacted with some of these proteins. By testing a series of protein deletion mutants, we show that one of the small proteins, FE UPTAKE-INDUCING PEPTIDE3 (FEP3), interacts with BTSL1 at a similar C-terminal region as ILR3 and bHLH104. FEP3 positively regulates Fe uptake. With transgenic Arabidopsis plants over-expressing *FEP3* we could copy the phenotype of *btsl1-btsl2* loss-of-function plants. Together, our results suggest that FEP3 inhibits BTSL protein function, possibly by preventing BTSL-bHLH interaction.

Introduction

Iron (Fe) is essential for all organisms. In plants, Fe is a crucial component of chlorophyll synthesis (Tottey et al., 2003) as well as in photosynthetic and respiratory electron transport chains (Nouet et al., 2011). Fe deficiency stress is reflected in visible interveinal leaf chlorosis, and can result in compromised development and yield loss. Although very abundant in the soil, Fe is often not readily accessible for plants, because it precipitates as insoluble Fe^{3+} oxides at neutral or basic pH (Lindsay, 1988; Wedepohl, 1995). Graminaceaous plants like rice extrude mugineic acids (MAs) as chelating compounds into the rhizosphere to solubilize Fe^{3+} and import Fe^{3+}-MA complexes into the root (Fe acquisition "Strategy II") (Marschner and Römheld, 1994). Non-graminaceaous monocots and dicots, like Arabidopsis, solubilize soil Fe by lowering the local pH through proton extrusion, accompanied by release of coumarins as chelating compounds. Fe^{3+} is then reduced and imported into the root as Fe^{2+} ("Strategy I") (Marschner and Römheld, 1994; Schmid et al., 2014). Because of its reactivity, free internal Fe^{2+} has to be avoided. Therefore, once it has entered the symplast, Fe is rapidly chelated and transported to local and distant sinks, or sequestered (von Wirén et al., 1999; Hell and Stephan, 2003; Briat et al., 2007; Schuler et al., 2012; Curie and Mari, 2017). To ensure sufficient Fe supply while avoiding cellular Fe content rising to toxic levels, plants need to carefully balance Fe homeostasis. This includes Fe status sensing and accordingly signaling Fe demand, as well as regulation of internal Fe redistribution and uptake of external Fe. The need to orchestrate different processes is reflected in a complex transcriptomic network of co-regulated genes encoding several transporters, enzymes and transcription factors (TFs) (Ivanov et al., 2012).

In Strategy II as well as Strategy I, a number of basic helix-loop-helix (bHLH) TFs regulate Fe uptake and distribution genes in roots (Gao et al., 2019). In Arabidopsis (*Arabidopsis thaliana*), FER-LIKE IRON DEFICIENCY INDUCED TRANSCRIPTION FACTOR (FIT) controls a set of co-expressed root genes acting in Fe acquisition (Colangelo and Guerinot, 2004; Jakoby et al., 2004; Bauer et al., 2007; Sivitz et al., 2012; Mai et al., 2016). For example, *FERRIC REDUCTION OXIDASE2* (*FRO2*), which reduces Fe^{3+} to Fe^{2+} (Robinson et al., 1999), and *IRON-REGULATED TRANSPORTER1* (*IRT1*), which imports Fe^{2+} (Eide et al., 1996; Vert et al., 2002), are directly regulated by FIT. FIT protein exerts its function upon heterodimerization with bHLH TFs from group Ib (bHLH38/39/100/101) (Heim et al., 2003; Yuan et al., 2008; Wang et al., 2013). Ib *BHLHs* are transcriptionally induced by bHLH TFs from group IVc (bHLH34/104/105/115) (Zhang et al., 2015; Liang et al., 2017). IVc bHLH TFs have redundant functions and act in a synergistic manner (Li et al., 2016; Liang et al., 2017). bHLH34/104/115 also induce transcription of *POPEYE* (*PYE*) (Zhang et al., 2015; Liang et al., 2017), a group IVb bHLH TF which is considered a direct negative regulator of Fe

distribution genes *NICOTIANAMINE SYNTHASE4* (*NAS4*), *FRO3* and *ZINC-INDUCED FACILITATOR1* (*ZIF1*) (Long et al., 2010).

IVc bHLH protein levels are likely controlled through proteasomal degradation. bHLH105 (ILR3 (IAA-LEUCINE RESISTANT3)) and bHLH115 are targeted by BRUTUS (BTS), an Fe deficiency-induced E3 ligase and negative regulator of Fe uptake (Selote et al., 2015; Matthiadis and Long, 2016). The somewhat paradox situation of *BTS* up-regulation under Fe deficiency, while the protein negatively regulates Fe uptake, was explained by the need to have a rapid shut-down mechanism of Fe uptake (Hindt et al., 2017). Another hypothesis is that BTS ensures constant turnover of “fresh” IVc bHLH proteins (Selote et al., 2015), a mechanism that is known from FIT protein activity (Sivitz et al., 2011). BTS, IVc bHLHs, Ib bHLHs and PYE have orthologous counterparts in rice (Kobayashi et al., 2018), which indicates that the general mechanism is conserved in Fe acquisition Strategy I and Strategy II.

Interestingly, BTS might have Fe sensing functions, as indicated by its unique domain composition. BTS has a C-terminal REALLY INTERESTING NEW GENE (RING) domain with E3 ligase activity and N-terminal hemerythrin/HHE cation-binding motif (HHE) domains which can bind Fe^{2+} (Kobayashi et al., 2013; Selote et al., 2015; Matthiadis and Long, 2016). This structure resembles the mammalian Fe sensing E3 ligase complex (Kobayashi et al., 2013), which has the subunit FBXL5, that is stabilized when Fe^{2+} is bound to its HHE domain (Salahudeen et al., 2009). Indeed, BTS protein stability and function was also shown to be directly linked to Fe presence or absence (Selote et al., 2015). Two BTS homologs, BTS-LIKE1 (BTSL1) and BTSL2, have been recently identified and also regulate Fe uptake in a negative manner (Hindt et al., 2017). BTS, BTSL1 and BTSL2 have partly redundant functions, but only *BTS* is expressed in roots and shoots, while *BTSL1/2* are expressed only in roots (Hindt et al., 2017). It is assumed that BTSL1/2 work in a similar manner as BTS, however, the proteins are not functionally characterized yet.

The root Fe uptake machinery responds to local and systemic Fe deficiency signaling (Vert et al., 2003; Kumar et al., 2017), however, the chemical nature of the signals is still unknown. Hormones and small molecules modulate Fe acquisition (Brumbarova et al., 2015), but their activating effect can be overruled by phloem Fe content (Garcia et al., 2013). Sufficient phloem Fe levels in shoots seems to be a key element to repress root Fe uptake (Garcia et al., 2013, and references therein; Zhai et al., 2014; Khan et al., 2018). That the shoot vasculature is important for root Fe deficiency responses, implies phloem-localized shoot-to-root signaling (García et al., 2018 and references therein). Recently, phloem-located small peptides, conserved in angiosperms, have been suggested as mobile signals that activate root Fe deficiency responses (termed FE UPTAKE-INDUCING PEPTIDE (FEP) or IRON MAN (IMA)) (Grillet et al., 2018; Hirayama et al., 2018). Small peptides in nutrient-related signaling

are best described in nitrogen deprivation responses. Two polypeptides are phloem-mobile shoot-to-root signals and induce expression of a root nitrate transporter (Tabata et al., 2014; Ohkubo et al., 2017). Regarding Fe signaling, it is currently unclear if FEPs themselves are a shoot-to-root signal, or if they trigger another kind of mobile signal which ultimately leads to up-regulation of Fe acquisition genes.

In view of the number of Fe deficiency-co-regulated genes - even in roots alone - surprisingly little is known about their interplay at the protein level. This work aims to uncover yet unknown functions and connections between proteins, to improve our mechanistic understanding of Fe deficiency response regulation in Arabidopsis roots.

We conducted a comprehensive targeted yeast two-hybrid screen with a set of 24 selected proteins. Most candidates were co-expressed under Fe deficiency, and had either unknown functions or a likely function in Fe sensing or signaling, such as BTS, BTSL1 and FEP3. We also included important transcriptional regulators, such as FIT, bHLH39, PYE and IVc bHLHs. We discovered a protein interaction network, with several novel interactions that deserve thorough follow-up experiments. Here, we concentrate on the “BTS(L) interactome”, a part of the protein network which connects BTS, BTSL1 and BTSL2 with individual sets of bHLH TFs, and with FEP3. We shed light on the multilayered Fe deficiency response regulation and add new players. We discuss an emerging mechanism in which FEP3 inhibits BTSL1/2 E3 ligase activity, thereby de-repressing downstream Fe deficiency responses.

Results

Identifying novel protein-protein interactions in a targeted yeast two-hybrid screen

Because transcriptional co-regulation has turned out in many cases as a good indicator for protein-protein interaction, we updated the root Fe deficiency co-expression network (based on (Ivanov et al., 2012) and the ATTED-II database (Aoki et al., 2016)) and used it to select candidates for a comprehensive targeted protein-protein interaction screen (**Supplemental Figure S1A**). In the selection process, we focused on proteins with possible connection to regulation or signaling. Since it is not yet known how exactly Fe deficiency is signaled, criteria for candidates were (i) unknown functions during Fe deficiency responses (at the time the study was initiated; UP1-4, KELCH, S8H, DUF506, FEP3, DGAT3, BTSL1, ORG1, SDI1) and (ii) regulatory functions of Fe homeostasis (TFs: FIT, bHLH39, MYB72, PYE, ILR3, bHLH104; E3 ligase: BTS; 14-3-3 protein: GRF11; enzymes: NAS2, NAS4 (Jakoby et al., 2004; Yuan et al., 2008; Klatte et al., 2009; Long et al., 2010; Palmer et al., 2013; Yang et al., 2013; Selote et al., 2015; Zhang et al., 2015)). In addition, we included PRS2 and JAL12, which were suspected to be FIT targets based on previous experiments (data not shown). Known or predicted integral membrane proteins (e.g. transporters) were excluded due to difficulties to

investigate such proteins with standard yeast two-hybrid (Y2H) assays. All 24 candidates are listed in **Table 1**.

Table 1. List of candidates tested in the Y2H screen.

AGI	Short name - Description
FIT-dependent[1, 2]	
AT1G34760	GRF11 - GENERAL REGULATORY FACTOR 11, 14-3-3 protein
AT1G56160	MYB72 - TF
AT1G73120	UP1 - unknown protein (109 aa)
AT3G06890	UP2 - unknown protein (128 aa)
AT3G07720	KELCH - galactose oxidase/kelch repeat superfamily protein
AT3G12900	S8H - SCOPOLETIN 8- HYDROXYLASE
AT5G56080	NAS2 - NICOTIANAMINE SYNTHASE2
Potentially FIT-dependent[3]	
AT1G32380	PRS2 - PHOSPHORIBOSYL PYROPHOSPHATE SYNTHASE2
AT1G52120	JAL12 - mannose-binding lectin superfamily protein
FIT-independent	
AT1G12030	DUF506 - DOMAIN OF UNKNOWN FUNCTION506, putative phosphoenolpyruvate carboxylase
AT1G47400	FEP3 - FE-UPTAKE-INDUCING PEPTIDE3, IRONMAN1 (IMA1) (50 aa)
AT1G48300	DGAT3 - DIACYLGLYCEROL ACYLTRANSFERASE3, 2Fe-2S cluster protein
AT1G56430	NAS4 - NICOTIANAMINE SYNTHASE4
AT1G74770	BTSL1 - BRUTUS-LIKE1, putative RING E3 ligase
AT3G18290	BTS - BRUTUS, putative RING E3 ligase
AT3G47640	PYE - POPEYE, bHLH TF
AT3G56360	UP3 - unknown protein (233 aa)
AT3G56980	bHLH39 - TF
AT5G05250	UP4 - unknown protein (239 aa)
AT5G53450	ORG1 - OBP3-RESPONSIVE GENE1, predicted protein kinase activity
AT5G48850	SDI1 - SULPHUR DEFICIENCY-INDUCED1
TFs regulating Fe deficiency responses*	
AT2G28160	FIT - FER-LIKE IRON DEFICIENCY INDUCED TRANSCRIPTION FACTOR, bHLH29
AT4G14410	bHLH104
AT5G54680	ILR3 - IAA-LEUCINE RESISTANT3, bHLH105

[1](Colangelo and Guerinot, 2004); [2](Mai et al., 2016); [3](Mai, 2013). *Not co-expressed with the other Fe deficiency response genes (**Supplemental Figure S1**). Short name and description according to The Arabidopsis Information Resource (TAIR), 10.0 genome release. If no short name available, we provided a shortened description version.

Using the Y2H system, all 24 candidates were systematically tested against each other, including themselves (referred to as the “Y2H screen” in this work). Protein pairs were tested reciprocally to capture a majority of interactions (total: 576 tests). To efficiently assay all protein pairs, we transformed bait and prey constructs separately and combined them by yeast mating. Around 6% of all tested interactions were positive (**Supplemental Figure S1B**). Among them, expected interactions were recovered (FIT+bHLH39, BTS+ILR3, BTS+bHLH104, PYE+ILR3, ILR3+ILR3) (Yuan et al., 2008; Long et al., 2010; Selote et al., 2015). In addition, we found 14

novel between-protein interactions and five homodimers. 11 of the 14 interactions involved BTSL1 or PYE: BTSL1 interacted with PYE, ILR3, MYB72, DUF506, SDI1, PRS2, FEP3 and UP3. Besides with BTSL1, PYE interacted with UP1, UP3 and SDI1. Thus, BTSL1 and PYE emerged as interaction hubs in our Y2H screen (**Figure 1**).

BTS, BTSL1, and BTSL2 interact with individual sets of bHLH TFs and a small peptide

BTS and BTSL1 both interacted with ILR3, but only BTSL1 interacted with PYE. Furthermore, BTSL1 (but not BTS) interacted with the putative mobile peptide FEP3. To validate the results of the Y2H screen, and to test to what extent BTS and BTSL1 have similar interaction partners, we conducted targeted Y2H assays using co-transformation. We included IVc bHLHs (ILR3, bHLH104), PYE and FEP3. We also included BTSL2, which was not considered in the initial Y2H screen because it did not appear in the Fe deficiency response network by the time we initialized the experiments. Because full-length bHLH104 was auto-activating the Y2H system (**Supplemental Figure S3D**), we cloned the interacting C-terminal part and used bHLH104-C for all further experiments (**Supplemental Figure S4A**). Our results confirmed that BTS and BTSL1 both interact with ILR3 and also with bHLH104-C, but BTSL2 did not interact with ILR3. Both BTSL1 and BTSL2, but not BTS, interacted with PYE and FEP3 (**Supplemental Figure S4B**). BTS, BTSL1 and BTSL2 did not interact with each other (**Supplemental Figure S4D**), but could interact with themselves in the Y2H screen (BTSL2 not tested). Taken together, BTS, BTSL1 and BTSL2 have partly the same and partly different interaction partners. Together with ILR3, bHLH104, PYE and FEP3, they form a highly interconnected part of the protein interaction network, referred to as the “BTS(L) interactome” in this work (**Figure 1**).

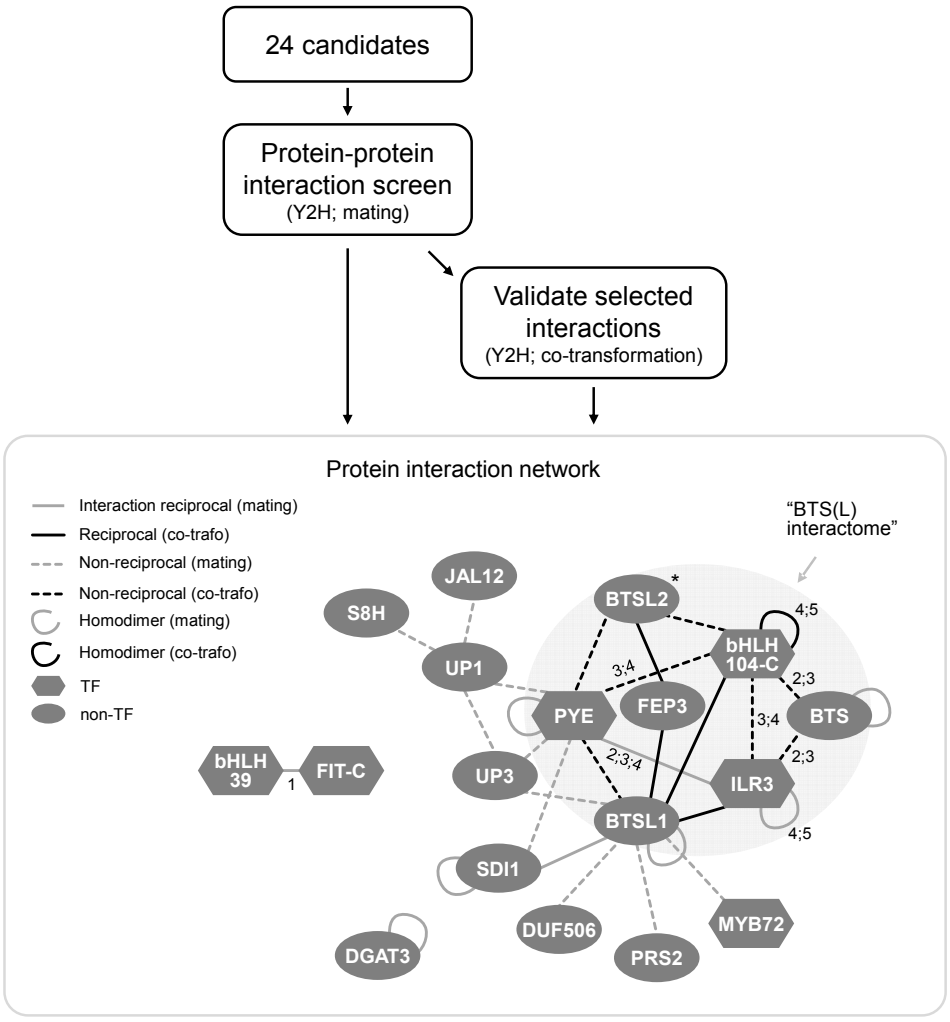


Figure 1. Workflow and protein interaction network based on Y2H data.
Each candidate was tested against all 24 (including itself; total 576 tests). The protein interaction network summarizes all new and known interactions identified with the initial Y2H screen. Seven candidates did not interact with any of the other proteins (GRF11, UP2, UP4, KELCH, NAS2, NAS4, ORG1). Interactions within the "BTS(L) interactome" (BTSL1, BTSL2, BTS, ILR3, bHLH104, PYE, FEP3) were validated by another round of Y2H experiments, in which yeast was co-transformed with bait and prey ("co-trafo"). *BTSL2 was not part of the initial Y2H screen, and tested only against the proteins of the BTS(L) interactome, FIT and bHLH39. BTSL2 was not tested against any of the other proteins or against itself. The 24 candidates are listed in **Table 1**. Overview matrix of Y2H screen interactions and non-interactions: **Supplemental Figure S1B**. Original data of the Y2H screen: **Supplemental Figures 2, 3**. Original data of validated selected interactions and non-interactions: **Supplemental Figures 4, 5**. Literature of known interactions: [1](Yuan et al., 2008); [2](Long et al., 2010); [3](Selote et al., 2015); [4](Zhang et al., 2015); [5](Li et al., 2016).

Next, we wanted to show that the proteins interact in the plant system. Using bimolecular fluorescence complementation (BiFC) we were only able to detect the interaction between BTSL1 and PYE (**Figure 2C(1)**), but not between BTSL1 and ILR3, bHLH104 or FEP3, or any of the BTSL2 interactions. One possible explanation is that full-length BTS and BTSL proteins are unstable (Selote et al., 2015; Rodríguez-Celma et al., 2019). To overcome this, we used BTSL1-C (C-terminal part of BTSL1; **Figure 2A**). Via Y2H experiments, we showed that BTSL1-C contained the domain needed for protein-protein interactions (**Figure 2B**), similar to what is known from BTS (Selote et al., 2015). By switching to BTSL1-C in BiFC experiments we still detected the interaction with PYE, and in addition an interaction with ILR3 (**Figure 2C(2, 3)**), but still none of the other interactions which involve either BTSL1-C (or

BTSL2-C, negative BiFC data not shown). Interestingly, full-length BTSL1 interacted with PYE outside the nucleus, whereas BTSL1-C interacted with PYE and with ILR3 inside the nucleus.

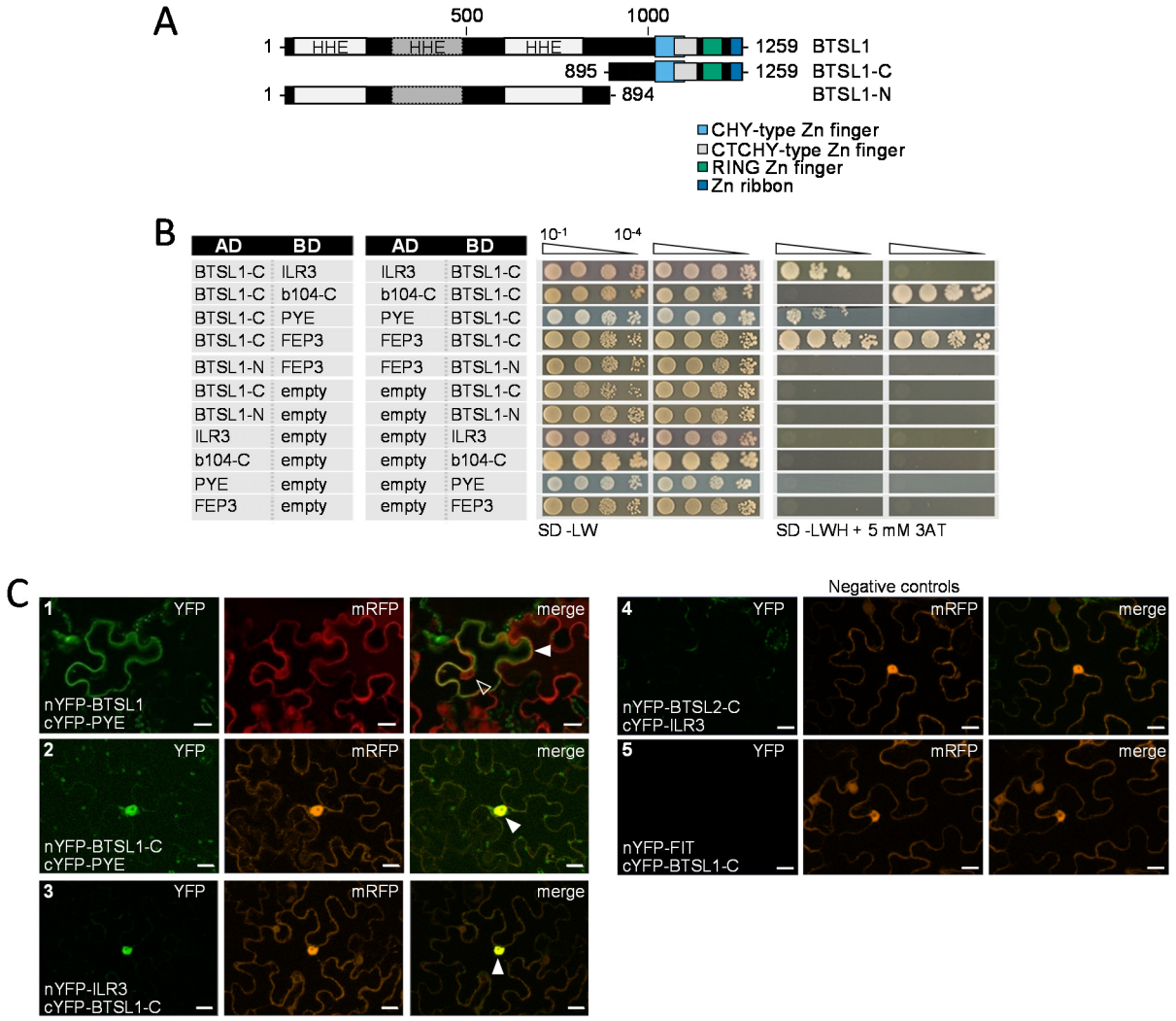


Figure 2. The BTSL1 C-terminus is sufficient for interaction with ILR3, bHLH104-C, PYE and FEP3, and can be used to validate interactions *in planta*.
A: Schematic representation of full-length BTSL1 and the truncated versions BTSL1-C (C-terminal part) and BTSL1-N (N-terminal part). Domain predictions according to (Rodríguez-Celma et al., 2019) (Hemerythrin/HHE cation binding motifs), and InterPro (www.ebi.ac.uk/interpro; C-terminal domains). **B:** Reciprocal targeted Y2H protein interaction assay between BTSL1-C and ILR3, bHLH104-C (C-terminal part of bHLH104; b104-C), PYE, FEP3, and between BTSL1-N and FEP3. Protein pairs were fused to the GAL4 DNA-binding domain (BD) and the GAL4 activation domain (AD), and vice-versa. Yeast co-transformed with the AD and BD combinations were spotted in 10-fold dilution series (A_{600}=10^{-1}-10^{-4}) on SD-LW (transformation control) and on SD-LWH supplemented with 5 mM 3AT (selection for protein interaction). Negative controls: empty AD with BD-proteins and empty BD with AD-proteins. **C:** BiFC experiments showing interactions between full-length nYFP-BTSL1 and cYFP-PYE (1), nYFP-BTSL1-C and cYFP-PYE (2), cYFP-BTSL1-C and nYFP-ILR3 (3) in transiently transformed tobacco leaf epidermis cells. Transformation control: mRFP. Non-interacting related proteins were chosen as negative controls: nYFP-BTSL2-C + cYFP-ILR3 (4) for BTSL1-C + ILR3; n-YFP-FIT + cYFP-BTSL1-C (5) for ILR3/PYE + BTSL1-C. The localization shift in (2) compared to (1) additionally validates specificity of BTSL1 + PYE and BTSL1-C + PYE signals. The YFP and mRFP signals were imaged with a fluorescence microscope and an ApoTome for enhanced resolution. Scale bars: 20 µm.

In summary, using Y2H and BiFC, we detected novel protein-protein interactions between several Fe deficiency response proteins. The interactions involving different small proteins (UP1, UP3, FEP3) are of particular interest, especially with regard to Fe deficiency signaling. With FEP3 being able to specifically interact with BTSL1/2, we uncovered a novel

connection between a small peptide and putative E3 ligases. In addition, BTSL1/2 interacted with different IVc bHLHs and with PYE, which could be partly validated *in planta*. In the following, we focus on BTSL1 and BTSL2, and their interactions with FEP3, IVc bHLHs and PYE.

ILR3, bHLH104-C and FEP3 interact with a similar C-terminal region of BTSL1

As shown above, BTSL-C is sufficient for interaction with FEP3, ILR3 and bHLH104-C. BTSL-C lacks the Fe^{2+}-binding HHE domains, but has predicted CHY- and CTCHY-type zinc (Zn) finger domains, a Zn ribbon domain and a RING domain (Kobayashi et al., 2013; Rodríguez-Celma et al., 2019), with putative E3 ligase function (Freemont, 2000; Selote et al., 2015). It was suggested that BTS interacts with its degradation targets ILR3 and bHLH115 via the RING domain (Selote et al., 2015). We used Y2H to investigate which of the C-terminal domains of BTSL1 interacts with ILR3, bHLH104-C and FEP3. We cloned four BTSL1-C sub-fragments in a way that allowed us to assess each of the domains individually (**Figure 3A, B**).

BTSL1-C.2, which has CHY and CTCHY domains, but lacks the predicted RING domain and the Zn ribbon domain, interacted with ILR3, bHLH104-C and FEP3. This indicates that BTSL1 RING and Zn ribbon are not needed for interaction. However, the interacting domain must be located close to RING, since a slightly shorter sub-fragment (BTSL1-C.1), lacking the RING-adjacent CTCHY, was no longer able to interact with any of the proteins. Interestingly, also CTCHY was not needed for FEP3 interaction, since BTSL1-C.4 interacted with FEP3. All three deletion constructs interacting with FEP3 (BTSL1-C.2, C.3 and C.4) have one common region of 14 amino acids (aa), located between CTCHY and RING (**Figure 3C**, yellow box). In case of ILR3 and bHLH104-C, results were less clear. For interaction with ILR3, CTCHY might also be important (**Figure 3A, B**).

Because BTSL2 and BTS did not interact with ILR3 and FEP3, respectively, we looked for differences in the putative interacting region of BTSL1 and found that “RTLVEH” with the evolutionary conserved R and L residues was different to both BTSL2 (“RLIVEH”) and BTS (“MKLVNH”). We focused on this part of the 14-aa region because we initially had reason to believe that FEP3 only interacted with BTSL1 and neither BTSL2 nor BTS. Consistent with our later findings that FEP3 interacted with both BTSL1 and BTSL2, deleting RTLVEH in BTSL1 (BTSL1-dRH) only slightly reduced interaction with FEP3 (**Figure 3D, E**). Substituting RTLVEH with a sextuple G residue spacer (BTSL1-6G) fully restored the interaction with FEP3. Possibly, the evolutionary conserved aa adjacent to RTLVEH are more important for interaction with FEP3 (**Figure 3C**). However, to our surprise, ILR3 and bHLH104-C could no longer interact with BTSL1-dRH. ILR3 did also not interact with BTSL1-6G, and interaction of bHLH104-C with BTSL1-6G was severely reduced (**Figure 3D, E**).

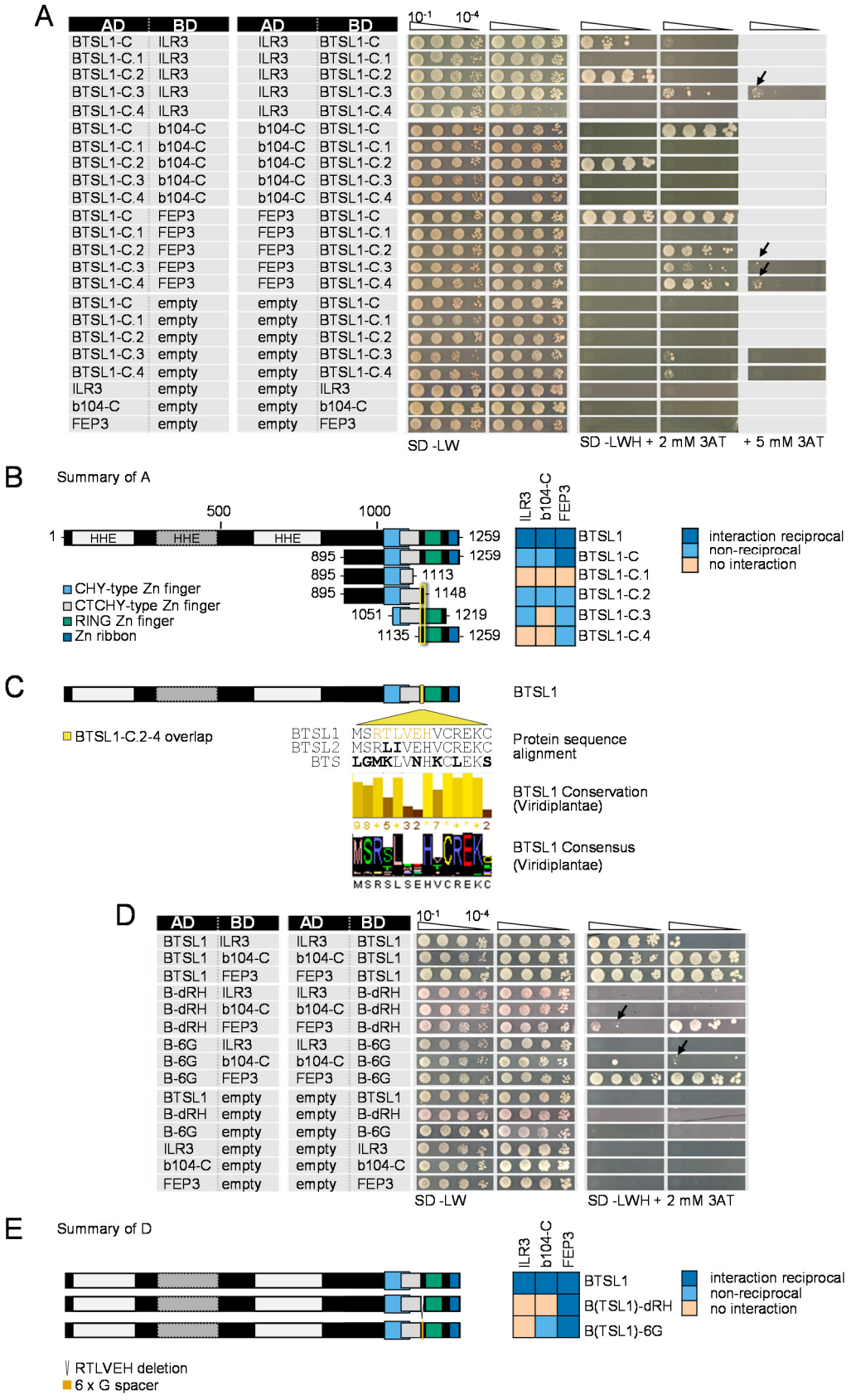


Figure 3. Determining the region in BTSL1 which mediates protein-protein interactions.
A: Reciprocal targeted Y2H protein interaction assay of BTSL1-C and its four sub-fractions (BTSL1-C.1, BTSL1-C.2, BTSL1-C.3, BTSL1-C.4) with ILR3, bHLH104-C (b104-C), PYE, FEP3. Yeast co-transformed with AD and BD combinations were spotted as described in Figure 2. Arrows indicate interaction. *(legend continued on next page)*

B: Schematic representation of full-length BTSL1 protein, BTSL1-C and the four sub-fragments of BTSL1-C. Heatmap summarizes the results of A (except interaction data of full-length BTSL1, which corresponds to D). Sub-fragments interacting with FEP3 (BTSL1-C.2, C.3 and C.4) have one common region (yellow box). **C:** Magnification of the yellow box and alignment of its 14 aa with BTSL2 and BTS. Bold: aa in BTSL2 and BTS that are different to BTSL1. "RTLVEH" (orange): contains aa that are different between BTSL1 and BTSL2 and between all three proteins (bold letters). "RTLVEH" was chosen for deletion/substitution (see below). Conservation scores of the aa and consensus correspond to a BTSL1 BLAST against the Viridiplantae data base. **D:** Interactions of full-length and deletion forms of BTSL1 with ILR3, bHLH104-C, FEP3. Yeast were co-transformed with full-length BTSL1, a BTSL1 deletion construct lacking RTLVEH (B-dRH), or a BTSL1 deletion-substitution construct in which RTLVEH was substituted with a sextuple glycine spacer (B-6G), in combination with either ILR3, bHLH104-C or FEP3 and spotted as described in Figure 2. Arrows indicate interaction. The single AD-B-6G + BD-b104-C yeast colony growing at $A=10^{-2}$ was not interpreted as reliable interaction. **E:** Schematic representation of full-length BTSL1, B-dRH and B-6G proteins. Heatmap summarizes the results of D. Full alignment of BTSL1, BTSL2, BTS: **Supplemental Figure S6**.

In summary, we identified a 14-aa region close to the BTSL1 E3 RING domain which is needed for interaction with FEP3, and possibly also with ILR3 and bHLH104-C. Within this region, RTLVEH is potentially important for interaction with ILR3 and bHLH104-C, but not with FEP3. This indicates that FEP3 and the IVc bHLHs do not bind to the same residues in BTSL1, but to a similar region. However, it should be noted that we cannot rule out that deleting or substituting RTLVEH affects BTSL1 protein folding in a way that only the small FEP3, but not ILR3 or bHLH104-C are able to bind.

FEP3 C-terminal residues are conserved and needed for interaction

Next, we were interested which domains of FEP3 are important for protein-protein interactions. A BLAST search of the FEP3 protein sequence revealed a single domain consisting of the final 17 aa at the C-terminus (**Supplemental Figure S7**), consistent with findings of (Grillet et al., 2018). This domain is conserved in dicots as well as monocots with Strategy II Fe acquisition, such as rice, wheat and maize. We cloned an N-terminal and a C-terminal half of FEP3 (termed FEP3-N and FEP3-C, respectively) and determined, not surprisingly, FEP3-C as the interacting part (**Figure 4A**). Next, we tested two truncated FEP3 versions lacking the whole conserved domain (FEP3-d17) and lacking the last seven aa (FEP3-d7), respectively (**Figure 4B**). Neither of the two interacted with BTSL1, showing that the last seven aa in FEP3 ("YDYAPAA") are crucial. We noted a certain aa sequence similarity between YDYAPAA and the C-terminus of IVc bHLHs (**Supplemental Figure S8**). As a matter of fact, bHLH104 end FEP3 peptide sequences both end with "PAA". In comparison, Ib bHLH protein C-termini did not align with FEP3 YDYAPAA (data not shown). Therefore we hypothesized that FEP3 might mimic IVc bHLHs during interaction with BTSL1/2.

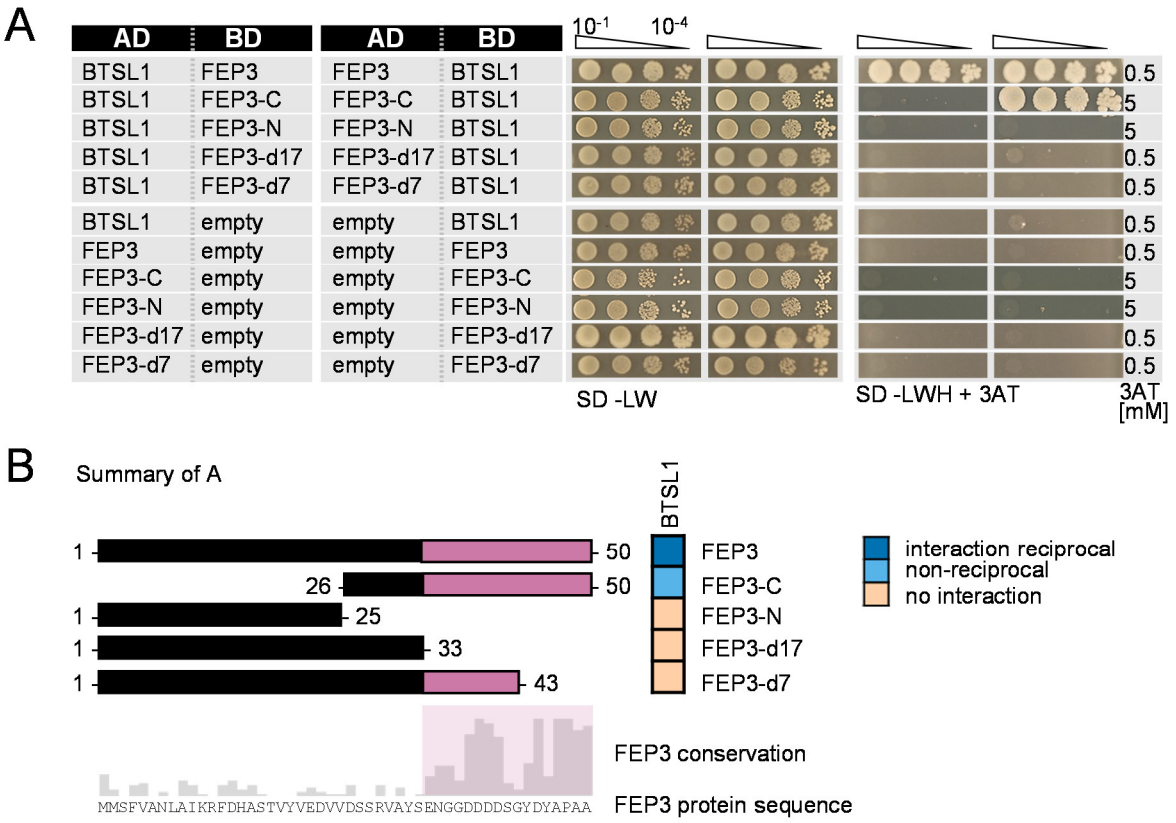


Figure 4. Determining the region in FEP3 needed for protein-protein interaction.
A: Reciprocal targeted Y2H protein interaction assay of full-length FEP3 or four truncated FEP3 versions with BTSL1. Yeast were co-transformed with AD-BTSL1 or BD-BTSL1 combined with BD or AD of FEP3, the C-terminal half of FEP3 (FEP3-C), the N-terminal counterpart (FEP3-N), a version lacking the last 17 aa (FEP3-d17), a version lacking the last 7 aa (FEP3-d7). Yeast were spotted as described in Figure 2. **B:** Schematic representation of FEP3 full-length and truncated versions. The protein sequence and aa conservation in angiosperms is indicated. Heatmap summarizes results of A. Full alignment of FEP3 angiosperm BLAST search results and consensus plot: **Supplemental Figure S7**.

Because we used bHLH104-C in all Y2H experiments, we already knew that the C-terminal part of bHLH104 is sufficient for interaction with BTSL1, BTSL2 and BTS. To test, if the interacting sites lie in the last 25 aa (aligning with FEP3 YDYAPAA), we cloned ILR3-d25 and bHLH104-C-d25 (lacking the last 25 aa) and the counterparts ILR3-CC and bHLH104-CC (only the last 25 aa; **Supplemental Figure S9**). The results were somewhat inconclusive and we might have pushed the Y2H method to the limit. However, ILR3-d25 and bHLH104-C-d25 still interacted with BTSL1/2 (**Supplemental Figure S9**). Therefore we can conclude that the C-terminal ends of ILR3 and bHLH104 are not required for interaction, despite the somewhat resemblance to FEP3. These results do not support the idea that FEP3 mimics bHLH IVc proteins.

FEP3 and BTSL1/2 are both upstream regulators of Fe deficiency responses and act in an opposite manner

FEP3 is a positive regulator of Fe uptake and a putative mobile signal (Grillet et al., 2018). However, the mechanism by which FEP3 acts is unknown. BTSL1 and BTSL2 are

negative regulators of Fe uptake and putative Fe sensors (Hindt et al., 2017). Because FEP3 physically interacted with BTSL1 and BTSL2, we hypothesized that FEP3 inhibits their function. To investigate this, we compared transgenic Arabidopsis plants over-expressing *FEP3* (FEP3-OX) to *btsl1-btsl2* loss-of-function mutants.

We generated two independent FEP3-OX lines (FEP3-OX1, FEP3-OX3). By reverse transcription quantitative PCR (RT-qPCR) we confirmed that both strongly over-expressed *FEP3* in seedlings and roots under Fe sufficient (+Fe) and Fe deficient (-Fe) conditions (**Figure 5A**; **Supplemental Figure S10A**). Transcript accumulation was strongest in FEP3-OX1 (up to 12,000-fold increase compared to the Col-0 wild type (WT) control, in roots of plants grown 14 d at +Fe, and shifted to +Fe or -Fe for an additional 3 d; **Figure 5A**). *FEP3* was not further up-regulated under -Fe, showing that its expression in FEP3-OX lines was independent from Fe supply. We detected HA_3-FEP3 in whole protein extracts of 6 d-old FEP3-OX1 seedlings under +Fe and under -Fe (**Supplemental Figure S10B, C**). HA_3-FEP3 protein has a calculated molecular weight of 12.5 kDa, however, we always detected a single band of the fusion protein at approximately 20 kDa in SDS protein gels. For very acidic proteins like FEP3 (predicted isoelectric point: 3.83; www.arabidopsis.org) slower migration than expected is a common phenomenon (Graceffa et al., 1992; Peterman et al., 2004). In addition, this behavior could be an indication for post-translational modifications of FEP3 protein. HA_3-FEP3 bands from FEP3-OX1 protein extracts were very weak, other than expected by the amount of transcript over-accumulation. We did not detect HA_3-FEP3 in FEP3-OX3 protein extracts, consistent with the lower transcript over-accumulation compared to FEP3-OX1. Difficulties to detect HA_3-FEP3 might point towards post-translational control of FEP3 protein levels.

Next, we assessed at which hierarchical level FEP3 acts during root Fe deficiency response. We determined expression levels of genes that are part of the Fe deficiency regulatory cascade (*FRO2*, *IRT1*, *FIT*, *BHLH38*/*39*, *PYE*, *FRO3*, *NAS4*, *ILR3*, *BHLH104*) in FEP3-OX roots grown in the 14+3 d system. We also tested and compared their expression levels in the *btsl1-btsl2* mutant line (**Figure 5A-N;** verification of *btsl1-btsl2* loss-of-function line see **Supplemental Figure S10D-F**). Fe uptake genes *FRO2* and *IRT1* were up-regulated under +Fe in FEP3-OX1 compared to the WT control, indicating that the Fe uptake machinery is constantly active irrespective of Fe supply (**Figure 5H, I**). To test if FEP3-OX lines accumulate more Fe, we measured the seed Fe content. Indeed, FEP3-OX1/3 seeds contained significantly more Fe than the WT control (**Figure 5O**). However, *FIT* expression levels were not changed in FEP3-OX1, or even decreased under +Fe in FEP3-OX3 (**Figure 5G**). Therefore, enhanced Fe uptake did not seem to be caused by increased *FIT* expression. However group Ib *BHLH38* and *BHLH39*, which activate *FRO2* and *IRT1* expression as heterodimers with FIT (Yuan et al., 2008), were up-regulated under +Fe and -Fe compared to the WT control (**Figure 5E, F**). This explains the activity of the Fe uptake machinery under

+Fe. Up-regulation of *FRO2*, *IRT1*, *BHLH38* and *BHLH39* under +Fe in FEP3-OX lines was also confirmed by others (IMA1c Ox; (Grillet et al., 2018)). It was reported that Ib *BHLH* expression is controlled by IVc bHLHs, which also regulate *PYE* (Zhang et al., 2015). Indeed, we also found *PYE* up-regulated under +Fe and -Fe in FEP3-OX1 (**Figure 5J**). Consequently, we expected *FRO3* and *NAS4*, which are negatively regulated by PYE (Long et al., 2010), to be down-regulated. However, *NAS4* regulation was not changed, and *FRO3* was even up-regulated under +Fe in FEP3-OX1 (**Figure 5K, L**), suggesting that other mechanisms in addition to PYE control *NAS4* and *FRO3* expression. Together, this data confirms that FEP3 acts as a positive regulator upstream in the Fe deficiency response cascade.

Fe deficiency response genes were regulated similarly in the *btsl1-btsl2* mutant: *FRO2*, *IRT1*, *BHLH38*/*39*, *PYE* and *FRO3* were up-regulated (even stronger than in FEP3-OX), while *FIT* and *NAS4* were not (**Figure 5E-L**). However, *BTS* was up-regulated in *btsl1-btsl2*, but not in FEP3-OX (**Figure 5D**). This might reflect the need to compensate for the missing BTSLs in *btsl1-btsl2*. Unexpectedly, *ILR3* and *BHLH104* were down-regulated in FEP3-OX1/3 and *btsl1-btsl2* under +Fe (**Figure 5M, N**). Thus, *ILR3*/*BHLH104* were transcriptionally regulated in an opposite manner as their putative downstream targets *BHLH38*/*39*/*PYE* and *FRO2*/*IRT1*. This observation was confirmed in an independent experiment using 6 d-old seedlings (**Supplemental Figure S11**).

In line with the RT-qPCR results, *btsl1-btsl2* seeds also accumulated Fe (**Figure 5O**). Otherwise, both FEP3-OX and *btsl1-btsl2* displayed only weak phenotypic traits. Chlorophyll a and b content in rosette leaves of 10 d-old plants were not changed (**Figure 5P**). We also measured the length of primary roots. With our growth conditions, we typically observe longer primary roots in 6 d-old WT plants grown under -Fe compared to +Fe (**Figure 5Q**). This effect was not changed in FEP3-OX and *btsl1-btsl2*, however, FEP3-OX3 and *btsl1-btsl2* roots were overall significantly longer. Therefore, FEP3-OX and *btsl1-btsl2* also have similar - however weak - morphological phenotypes.

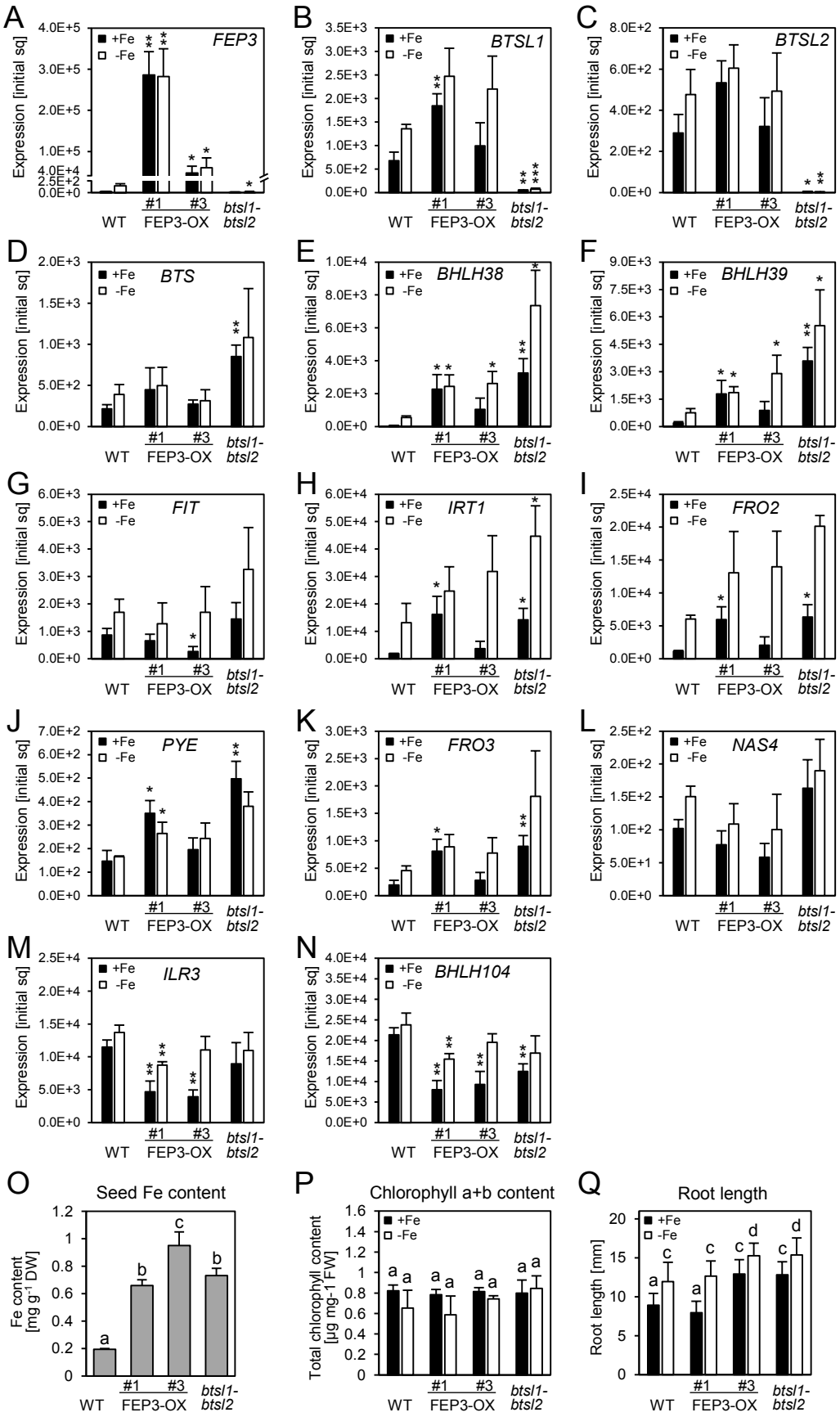


Figure 5. FEP3 and BTSL1/2 oppositely regulate Fe deficiency responses.
A-N: Expression analysis (RT-qPCR) of Fe deficiency response genes in roots of two FEP3-OX lines (#1, #3, WT background) and a *btsl1-btsl2* mutant line (see also **Supplemental Figure S10**) in comparison to WT. Expression [initial sq]=Absolute normalized expression [initial starting quantity]. *(legend continued on next page)*

Plants were grown in the 14+3 d system with sufficient (+Fe, black bars) or deficient (-Fe, white bars) Fe supply. Asterisks indicate statistically significant differences to the WT of the same growth condition (n=3) (Student's T-Test, *p<0.05, **p<0.01, ***p<0.001). **A:** *FEP3*, **B:** *BTSL1*, **C:** *BTSL2*, **D:** *BTS*, **E:** *BHLH38*, **F:** *BHLH39*, **G:** *FIT*, **H:** *IRT1*, **I:** *FRO2*, **J:** *PYE*, **K:** *FRO3*, **L:** *NAS4*, **M:** *ILR3*, **N:** *BHLH104*. **O:** Seed Fe content per dry weight (DW) (n=3). **P:** Chlorophyll a and b content per fresh weight (FW) leaves of WT, FEP3-OX (#1, #3) and *btsl1-btsl2* grown in the 10 d system with +Fe or -Fe supply (n=4). **Q:** Root lengths of WT, FEP3-OX (#1, #3) and *btsl1-btsl2* grown in the 6 d system with +Fe or -Fe supply (n>12). **A-Q:** Data are represented as mean and SD. **O-Q:** Different letters indicate statistically significant differences (one-way ANOVA and Tukey's post-hoc test, p<0.05).

Our results show that FEP3 and BTSL1/2 regulate a similar set of genes. To test whether they act consecutively by one being a transcriptional target of the other, or whether they act independently, we measured their transcript levels in the respective other mutant. *FEP3* was down-regulated in the *btsl1-btsl2* mutant under -Fe compared to the WT control (**Figure 5A**). However, since *FEP3* is Fe deficiency-induced, this relative transcript decrease might be a secondary effect of increased internal Fe in *btsl1-btsl2* plant tissue (Hindt et al., 2017; Rodríguez-Celma et al., 2017). *BTSL2* transcript levels were slightly but not significantly increased in FEP3-OX1 and unchanged in FEP3-OX3, which indicates that FEP3 and BTSL2 proteins do not influence each other transcriptionally. However, *BTSL1* levels were increased in FEP3-OX1 under +Fe compared to the WT control (**Figure 5B, C**).

Taken together, FEP3-OX plants resemble the phenotype of *btsl1-btsl2* loss-of-function plants, indicating that FEP3 and BTSL1/2 regulate downstream genes in a complementary manner. BTSL1/2 proteins interacted with FEP3 and with IVc bHLH/PYE in yeast. We hypothesize that FEP3 may bind and inhibit BTSL function to prevent degradation of their TF targets under -Fe. One prerequisite of protein-protein interactions being relevant *in vivo* is the existing possibility of the putative interaction partners to meet *in planta*. Therefore, we next determined and compared the promoter activities and sub-cellular protein localizations of FEP3, BTSL1, BTS, ILR3, bHLH104 and PYE.

Promoter activities of "BTS(L) interactome" genes

We assessed whether the "BTS(L) interactome" genes are expressed in similar plant tissues by studying their promoter activities. We cultivated stable Arabidopsis lines expressing the *GUS* (β-glucuronidase) reporter gene under the control of *ILR3*, *BHLH104*, *FEP3*, *PYE*, *BTSL1* and *BTS* promoters under +Fe and -Fe. As expected from previous reports (Li et al., 2016), we detected GUS staining, indicative of *proILR3* and *proBHLH104* activity, in the vasculature of roots, hypocotyl and cotyledons under +Fe and -Fe (**Figure 6A, B, G, H**). *ProFEP3* activity was also detected in root and cotyledon vasculature under +Fe and -Fe, but not in the hypocotyl (**Figure 6C, I**). In addition, true leaves were heavily stained with GUS under -Fe, but not under +Fe. We detected *proBTS* and *proPYE* activity mainly outside the root vasculature (**Figure 6D, F, J, L**; **Supplemental Figure S12A**), in contrast to other studies reporting expression mostly or exclusively inside the vasculature (Long et al., 2010; Selote et al., 2015). Only one of three independent *proBTS:GUS* lines had GUS staining in the

vasculature and this was only observed under +Fe (**Figure 6F**). Other than reported, we also did not detect GUS in *proBTS:GUS* cotyledon, however, true leaves had small stained areas under +Fe and -Fe (**Figure 6F, Supplemental Figure 12A**). Others have measured *BTS* expression in shoots of 14 d-old plants (Hindt et al., 2017), indicating that our 6 d-old seedlings might be too young for proper shoot *GUS* expression. We detected *proPYE* activity throughout the length of the whole root under +Fe and -Fe, with highest activity in the root meristem (**Figure 6D, J**). Because we used identical *BTS* and *PYE* promoter regions as in (Long et al., 2010; Selote et al., 2015), the different promoter activity patterns in the other studies might be due to different cultivation protocols. For example, we used custom-made Hoagland medium, instead of standard MS, to be able to carefully adjust all nutrients. We also avoided adding the Fe chelator Ferrozine to -Fe medium, which can cause off-target effects because it also complexes copper (Duffy and Gaudin, 1977). We detected weak *proBTSL1* activity under +Fe and -Fe only in root tissue mainly outside the vasculature (**Figure 6E, K**). Thus, it overlapped with *proPYE* activity. To overlap with *proILR3*, *proBHLH104* and *proFEP3*, *proBTSL1* should be active in the central cylinder. However, GUS was either absent from the vasculature, which would be in agreement with reports of others (Rodríguez-Celma et al., 2017), or by far not as pronounced as in *proILR3:GUS*, *proBHLH104:GUS*, and *proFEP3:GUS* lines. Interestingly, we observed *proILR3*, *proPYE*, *proBTSL1*, and *proBTS* activity under +Fe and -Fe in the root tip, possibly around the quiescent center and in the columella (**Figure 6A, D, E, G, J, L**; **Supplemental Figure 12A**). Noteworthy, others have measured Fe accumulation in the quiescent center root area, a pattern that was disturbed in *bts* mutant roots (Hindt et al., 2017).

In conclusion, *BTSL1* seems to be expressed mainly in the outer root layers, which is in contrast to its putative interaction partners *ILR3*, *BHLH104*, and *FEP3*, which are all expressed predominantly in the vascular tissue. Interestingly, *BTSL1* expression pattern overlaps with *PYE* expression pattern in the root differentiation zone. However, location of gene expression does not necessarily restrict the protein to the same location. To test if FEP3 protein can move to outer root layers, we generated a *pFEP3:FEP3-GUS* line, but could not observe GUS accumulation outside the root vascular tissue (**Supplemental Figure S12B**). However, since GUS (68 kDa) is much larger than FEP3 (5.5 kDa), tagged FEP3 might behave differently.

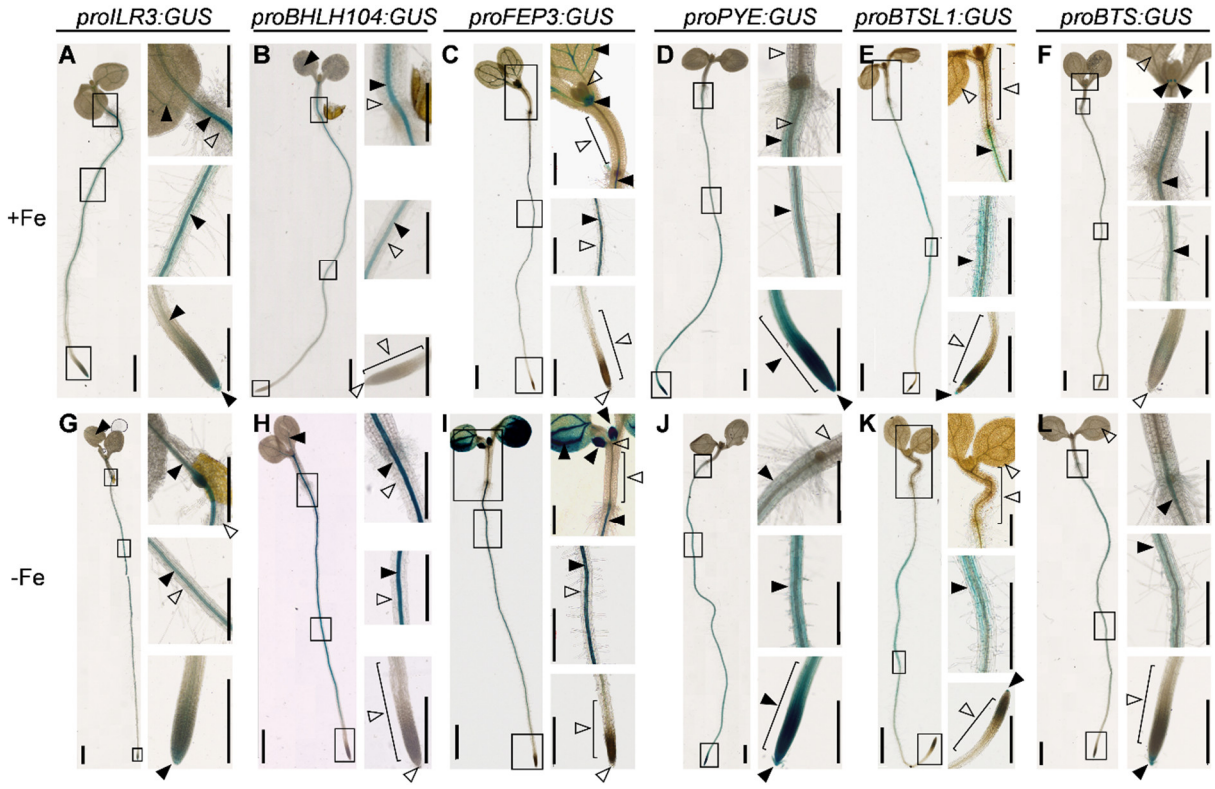


Figure 6. "BTS(L) interactome" genes are partly expressed in the same tissue.
GUS reporter activity, driven by *ILR3, BHLH104, FEP3, PYE, BTSL1*, *BTS* putative promoters. Transgenic Arabidopsis plants with *proILR3:GUS* (**A, G**), *proBHLH104:GUS* (**B, H**), *proFEP3:GUS* (**C, I**), *proPYE:GUS* (**D, J**), *proBTSL1:GUS* (**E, K**), *proBTS:GUS* (**F, L**) in WT background were grown in the 6 d system under sufficient (+Fe, **A-F**) and deficient (-Fe, **G-L**) Fe supply. Reporter gene expression was visualized by a chemical reaction (GUS) resulting in blue staining of the respective tissues. Plants were imaged with brightfield microscopy. Rectangles in whole-seedling images indicate positions of the enlarged images. Filled arrowheads indicate selected areas of GUS staining, non-filled arrowheads indicate selected non-stained areas. Vascular tissue is indicated by the arrowhead pointing directly at it, non-vascular tissue is indicated by the arrowhead pointing more to an outside area. Scale bars of whole seedling images: 1 mm; of magnifications: 0.5 mm. See also **Supplemental Figure S12A**.

Next, we were interested in the subcellular localizations of the proteins. We have stated earlier that BTS/BTSL proteins are difficult to detect in plant systems due to instability. This might be one reason why to date we were unable to image BTSL1-GFP protein in stable transgenic Arabidopsis lines (data not shown). We therefore switched to a transient expression system using tobacco leaf epidermis cells.

Subcellular localization and co-localization of "BTS(L) interactome" proteins

BiFC experiments of BTS with ILR3 or bHLH104 have shown that the proteins interact in the nucleus (Selote et al., 2015). Our data of fluorophore-tagged BTS, ILR3 and bHLH104 confirmed nuclear localization (**Figure 7A (4-6)**). Because BTSL1/2 interacted with ILR3 and/or bHLH104 in our Y2H experiments, we expected nuclear localizations similar to BTS. To our surprise, fluorophore-tagged BTSL1 mainly localized to the plasma membrane and only weakly to the nucleus (**Figure 5A (1)**). The fluorophore position (N-/C-terminal) did not affect BTSL1 localization (**Supplemental Figure S13**). Yellow fluorescent protein (YFP)-tagged BTSL2 localized stronger to the nucleus and to the cytoplasm (**Figure 7A (2)**). Noteworthy, BTSL1-C-

GFP (lacking the HHE domains) localized more to the cytoplasm and to the nucleus than the full-length version (**Figure 7A (8)**). This raises interesting questions regarding functionality of BTSL1 and the role of the HHE domains.

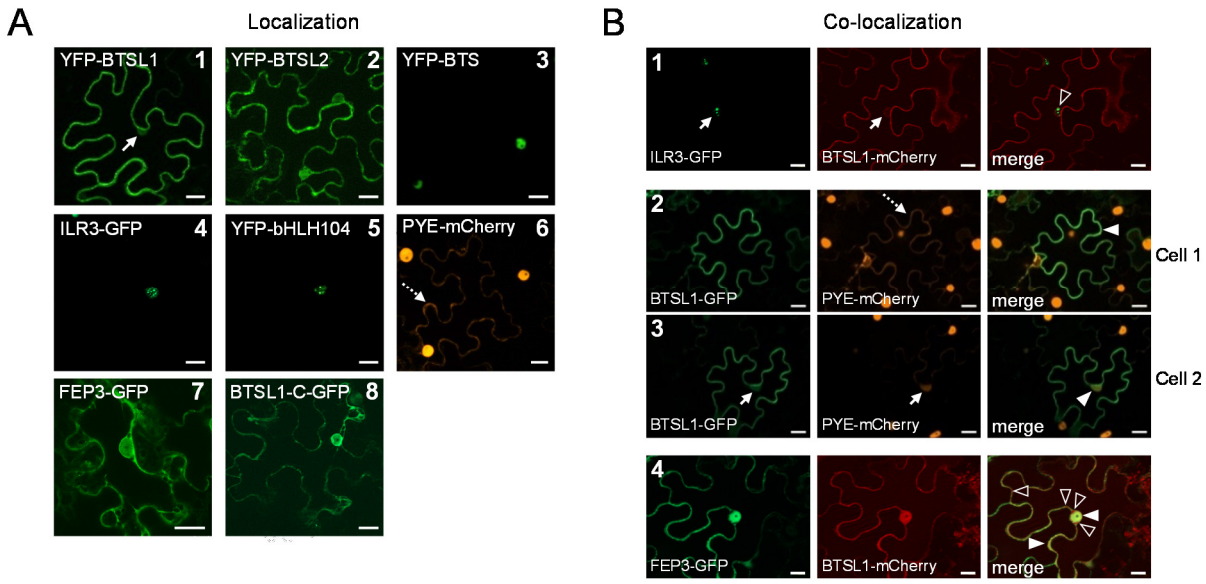


Figure 7. Subcellular localization and co-localization of "BTS(L) interactome" proteins.
A: Subcellular localization of YFP-BTSL1 (1), YFP-BTSL2 (2), YFP-BTS (3), ILR3-GFP (4), YFP-bHLH104 (5), PYE-mCherry (6), FEP3-GFP (7), BTSL1-C-GFP (8) in tobacco leaf epidermis cells (see also **Supplemental Figure S13**). **B:** Co-localization of BTSL1 with ILR3, PYE, FEP3. (1) ILR3-GFP and BTSL1-mCherry, (2, 3) BTSL1-GFP and PYE-mCherry, (4) FEP3-GFP and BTSL1-mCherry. Arrowheads show examples of co-localized GFP and mCherry signals (filled arrowheads) and areas of no co-localization (non-filled arrowheads). **A, B:** Non-dashed arrows indicate nucleus, dashed arrow indicates cytoplasm. The YFP, GFP and mCherry signals were imaged with a fluorescence microscope and an ApoTome for enhanced resolution. Scale bars: 20 µm.

FEP3-GFP and PYE-mCherry localized to nucleus and cytoplasm, with a preference of FEP3-GFP for the cytoplasm, and of PYE-mCherry for the nucleus (**Figure 7A (6, 7)**). BTSL1-GFP and PYE-mCherry co-localized inside the nucleus and possibly also outside the nucleus, although the PYE-mCherry signal at the cell periphery was weak (**Figure 7B (2, 3)**). In support of the peripheral co-localization, our BiFC experiments indicated BTSL1 and PYE interaction only outside the nucleus (**Figure 2C (1)**). Full-length BTSL1-mCherry and ILR3-GFP did not co-localize. In fact, they localized to mutually exclusive areas within the nucleus (**Figure 7B (1)**). Interestingly, when BTSL1-mCherry was co-expressed with FEP3-GFP, mCherry signal was detected much stronger inside the nucleus, as when expressed alone (**Supplemental Figure S13 (3)**). GFP and mCherry signals co-localized inside and outside the nucleus, except for a small area around the nucleus and small patches at the cell periphery (**Figure 7B (4), non-filled arrowheads**). It remains to be tested whether the mCherry signal corresponds to free fluorophore, to full-length BTSL1-mCherry or to a truncated version of BTSL1-mCherry. In any case, our observations might indicate FEP3-triggered post-translational modification of BTSL1.

Taken together, our results suggest that some, but not all, of the “BTS(L) interactome” proteins are able to meet *in planta*, when heterologously expressed. The putative interacting proteins BTSL1/PYE and BTSL1/FEP3 co-localized in different sub-cellular compartments, whereas BTSL1/ILR3 did not co-localize. The same might be true for BTSL1/bHLH104, since bHLH104 and ILR3 single localizations were identical. Single localization experiments further suggest that BTSL2 might co-localize with ILR3 and/or bHLH104, since BTSL2 accumulated more in the nucleus than BTSL1. These co-localization experiments are work in progress.

Discussion

With the help of a comprehensive targeted Y2H screen we tested 24 Fe deficiency response-related proteins for interactions, consisting of six TFs and 18 non-TFs. The screen recovered previously reported interactions (BTS+ILR3/bHLH104, FIT+bHLH39), confirming suitability of the approach. We discovered a protein interaction network in which BTSL1 and PYE stand out as interaction hubs. Since the putative Fe sensors BTSL1/2 and the putative mobile signal FEP3 are connected in the network, we might have discovered components of a Fe deficiency sensing system. In this context, we acknowledge the presence of two more small proteins (of yet unknown function), linked to BTSL1 and PYE. Noteworthy, FIT did not interact with most proteins. The protein interaction network thus rather reflects FIT-independent processes. In conclusion, the Y2H screen was well suited to identify robust protein interactions that are potentially relevant for Fe deficiency responses. The results are a reliable data source for detailed studies. In this study, we concentrated on the highly interconnected “BTS(L) interactome”, containing the three homologous RING-type E3 ligases BTS, BTSL1, BTSL2 with PYE, bHLH TFs of subgroup IVc (ILR3, bHLH104), and the small peptide (FEP3).

Unique properties of BTS, BTSL1 and BTSL2

Within the “BTS(L) interactome”, BTSL1 interacted with ILR3, bHLH104, PYE and FEP3. BTSL2 interacted with bHLH104, PYE and FEP3, but not with ILR3. BTS interacted with ILR3 and bHLH104, but not with PYE or FEP3. Therefore, BTS, BTSL1 and BTSL2 can partly interact with the same proteins, but seem to have individual sets of interacting partners. It was assumed from studying single, double and triple loss-of-function lines that BTS, BTSL1 and BTSL2 have partly redundant functions as negative Fe uptake regulators during Fe deficiency response (Hindt et al., 2017). We found *BTS* significantly up-regulated in the *btsl1-btsl2* mutant, possibly to compensate for lack of BTSL1 and BTSL2. However, *btsl1-btsl2* plants still accumulated more Fe than the WT, indicating that BTS can partly, but not completely, take over BTSL functions.

Apart from this work, experimental data on BTSL1 and BTSL2 is scarce. Other studies reported that *BTS*, *BTSL1* and *BTSL2* are expressed partly in different and partly in the same root tissues, and it was suggested that BTSL1/2 and BTS represent different checkpoints to

avoid Fe overload during its way from the rhizosphere into the root vasculature (Rodríguez-Celma et al., 2017). The authors of a bioRxiv preprint present FIT as BTSL1/2 degradation target, because they found increased FIT protein levels in *btsl1-btsl2* mutants and found BTSL1/2 interacting with FIT protein *in vitro* (Rodríguez-Celma et al., 2017). In our *in vivo* experiments, FIT did not interact with BTS, BTSL1 or BTSL2. We also tested BTS and BTSL1 promoter activities and could not confirm a clear expression partitioning in root tissue. Contrary to reports of others, *proBTS* activity in our study was detected mostly outside the central cylinder. At this point it is difficult to interpret these results, however, different media compositions of the other versus our study might be one reason for the effect. Noteworthy, the same study reporting vascular *BTS* expression (Long et al., 2010), also reported that *PYE* was expressed only in the vasculature, and that the PYE protein was able to move to outer root layers. However, we detected *proPYE* activity in all root tissues, indicating that PYE protein does not need to be cell-to-cell mobile in order to act outside the vasculature.

Although our results did not show that *BTSL1* and *BTS* have distinct promoter activities, we found intriguingly distinct subcellular localizations of BTS, BTSL1 and BTSL2. BTSL1 was mostly located at the plasma membrane and only weakly in the nucleus, in contrast to BTS, which was located only in the nucleus. BTSL2 had an intermediate localization pattern (cytoplasm and nucleus). Nuclear BTS localization was reported by others (Selote et al., 2015), but BTSL1 and BTSL2 localizations were not known before. However, consistent with our results, the BTS homolog in rice, HRZ1, also localized to the nucleus while HRZ2 localized to nucleus and cytoplasm (Kobayashi et al., 2013). Because BTS, BTSL1 and BTSL2 appear to cover different subcellular compartments, they might target different parts of the Fe deficiency response cascade. For example, both BTS and its degradation target ILR3 were localized (this work) and interacted (Selote et al., 2015) in the nucleus. We showed that ILR3 also interacted with BTSL1 in yeast. However, although BTSL1-mCherry and ILR3-GFP were both detected in nuclei of tobacco leaf cells, they localized to mutually exclusive areas. ILR3-GFP (and YFP-bHLH104) formed nuclear bodies, a yet puzzling phenomenon that has been best described in plants for phytochromes and which is thought to represent locations of transcriptional activity or of protein degradation (Van Buskirk et al., 2012; Klose et al., 2015). BTSL1-mCherry was localized in the nucleus only outside ILR3-GFP-containing nuclear bodies. This might indicate that BTSL1 does not target ILR3 for proteasomal degradation. However, degradation experiments will be needed in the future to clarify this matter.

BTSL1/2 physically interacted with and might target bHLH104 protein for degradation. bHLH104 is a more likely target of BTSL1/2 than ILR3, since bHLH104 is not targeted by BTS (Selote et al., 2015). We found *BHLH38*/*39* and *PYE* were expressed at high levels in *btsl1-btsl2* plants, suggesting highly active IVc bHLHs. Since ILR3 and bHLH115 should be

degraded by BTS (Selote et al., 2015) in *btsl1-btsl2* plants, those active IVc bHLHs are likely bHLH104 (and/or bHLH34).

BTSL1 and BTSL2, but not BTS, interacted with PYE in yeast. In tobacco leaf cells, BTSL1 and PYE co-localized inside and outside the nucleus, but interacted only at the cell periphery. In addition, *PYE* and *BTSL1* promoter activities overlapped in outer layers of Arabidopsis roots. PYE positively regulates overall Fe deficiency responses, since *pye-1* mutants are chlorotic, and show decreased rhizosphere acidification and reductase activity (both of which are indicators of Fe acquisition when increased). Furthermore, *IRT1* is less expressed under Fe deficient conditions (Long et al., 2010). PYE might therefore be a likely target of BTSL1 (and/or BTSL2).

A localization shift of BTSL1?

BTSL1 localization changed from peripheral to more cytoplasmic and nuclear when the N-terminal part harboring the HHE domains was cleaved off (BTSL1-C). Interestingly, similar experiments with BTS showed the same effect, however in this case the protein shifted from nuclear to cytoplasmic localization (Selote et al., 2015). We conclude that subcellular localizations of full-length BTSL1 (and BTS) depend on the N-terminus and possibly on the HHE domains. By binding to HHE domains, Fe^{2+}possibly influences BTS protein stability (Selote et al., 2015) and presumably BTSL1 protein is regulated in a similar manner. Because HHE domains and BTSL1 localization seem to be linked, it is tempting to speculate that BTSL1 protein can exist in two forms, depending on the intracellular Fe status: as full-length protein mostly at the plasma membrane and as truncated version that moves into the nucleus, perhaps to engage in different regulatory processes. Other E3 ligases are known to move into the nucleus upon stress, such as AtHOS1 (HIGH EXPRESSION OF OSMOTICALLY RESPONSIVE GENES1) moving from the cytoplasm into the nucleus during cold stress (Lee et al., 2001; Dong et al., 2006) or AtRGLG1 (RING domain ligase 1) and AtRGLG2, moving from the plasma membrane into the nucleus upon ABA or salt stress treatment (Cheng et al., 2012, 2016; Belda-Palazon et al., 2019). Because HOS1, RGLG1 and RGLG2 target nuclear proteins for degradation, the nuclear localized E3 ligases could be the active forms during the stress conditions. Our BiFC experiments showed that full-length BTSL1 and BTSL1-C both interacted with PYE. Both forms of BTSL1 can therefore engage in physical protein-protein interactions *in planta*, but act in different compartments. In contrast, only BTSL1-C, but not full-length BTSL1 interacted with ILR3 in the nucleus. This holds potential for a mechanism in which BTSL1 switches interaction partners upon truncation. However, it remains speculation if BTSL1-C is a biologically relevant form of BTSL1. Moreover, since all our tobacco experiments represent an Fe sufficiency situation, it needs to be determined if BTSL1 protein structure or localization is altered upon Fe deficiency stress.

Possible function of FEP3-BTSL interaction during Fe acquisition

The small peptide FEP3 is an interesting new player of the Fe deficiency response. Except from being a positive regulator (Grillet et al., 2018), its specific role during Fe deficiency responses is not known. Non-secreted peptides (<100 aa) are one of several classes of short proteins that are encoded by evolutionary conserved small open reading frames (ORFs; (Hsu and Benfey, 2018)). These small ORF-encoded peptides (often termed SEPs (Delcourt et al., 2018)) can act in many ways, for example as ligands to receptors or by modulating protein-protein interactions (Makarewich and Olson, 2017). As opposed to peptide hormones, another class of short proteins, SEPs do not have to be post-translationally processed in order to become bioactive peptides (Stührwohldt and Schaller, 2019). Although FEPs structurally resemble hormone peptide precursors, we and others could not find evidence for FEP3 or FEP1 cleavage and secretion (Grillet et al., 2018; Hirayama et al., 2018), since N-terminally tagged full-length protein was detectable in Arabidopsis stable transgenic lines and protein extracts. We therefore regard FEP3 a SEP. Our data shows that FEP3 acts upstream in the Fe deficiency response cascade. That FEP3-OX lines phenocopy the *btsl1-btsl2* mutant indicates that BTSL1/2 function was inhibited in FEP3-OX lines. The molecular phenotype of FEP3-OX was not as pronounced as of the *btsl1-btsl2* mutant, which probably means that a small amount of BTSL1/2 protein was still functional. Since *BTSL1*/*2* genes were not down-regulated in FEP3-OX lines, rather the contrary, the inhibition likely takes place on the protein level. We propose that FEP3 acts as direct inhibitor of BTSL1/2, thereby allowing the Fe uptake machinery to be active when Fe levels are low.

To our knowledge, there is no reported case of a SEP interacting with an E3 ligase in plants or inhibiting its function. Interestingly, SEP-E3 ligase interactions are known from animal systems. For example the Drosophila SEP *pri* interacts with the E3 ligase Ubr3, thereby facilitating Ubr3 binding to the TF Svb. This changes Svb function (Zanet et al., 2015). From Drosophila and mammals, more examples are known in which SEPs alter protein localization or bind to enzymes to affect their activity, either by direct competition with the substrate or in an allosteric manner (Cabrera-Quio et al., 2016). Our Y2H experiments showed that FEP3 and ILR3/bHLH104 likely interact with adjacent, but not identical residues of BTSL1. We therefore hypothesize that FEP3 is an allosteric inhibitor, preventing that degradation targets of BTSLs can be bound. However, protein crystal structure analyses are needed to clarify how BTSL, FEP3, and bHLH proteins interact.

Noteworthy, when BTSL1-mCherry was co-expressed with FEP3-GFP in tobacco, we observed overall stronger mCherry signal and a localization shift more to the nucleus (resembling the pattern of BTSL1-C). Therefore, FEP3 might trigger BTSL function change by altering the protein structure, stability or localization. The plasma membrane localization of BTSL1 might also indicate a ligand-receptor-like relationship of FEP3 and BTSL1. In a

hypothetical scenario, FEP3 as mobile signal binds to BTSL1/2 in roots, which leads to de-repression of Fe acquisition. A somewhat comparable principle is known from phosphate (Pi) starvation responses, where a microRNA acts as a shoot-to-root signal, targeting the mRNA of a root E2 conjugase, and thereby de-repressing the function of Pi transporters (Huang et al., 2013).

Several internal layers of control carefully balance Fe deficiency response regulation

Our experimental results add to the complexity of the Fe deficiency response regulation cascade both on the transcriptional and on the post-translational level. Next to BTS, we add BTSL1/2 as negative upstream regulators, representing another layer of controlling surplus Fe uptake. We state above that FEP3 and BTSL1/2 do not regulate each other transcriptionally, although *BTSL1* was significantly up-regulated in FEP3-OX. We believe that bHLH39 induces *BTSL1/2* expression as part of a regulatory feedback loop, because our results match an earlier observation by us that *BTSL1* (and *BTSL2*) were up-regulated in transgenic plants over-expressing *BHLH39* (39Ox; (Naranjo-Arcos et al., 2017)). 39Ox plants accumulate massive amounts of Fe under +Fe, which represses Fe deficiency-induced genes that are not transcriptionally controlled by bHLH39. For example, *FEP3* itself was down-regulated in 39Ox, therefore up-regulation of *BTSL1* cannot be dependent on FEP3. Increased *BHLH39* expression in our FEP3-OX lines can explain *BTSL1* up-regulation. The feedback loop between bHLH39 and *BTSL1/2* likely acts as a protective layer to prevent surplus Fe uptake. In comparison, *BTS* transcript levels were not significantly up-regulated in FEP3-OX and 39Ox plants (Naranjo-Arcos et al., 2017), indicating that *BTS* is not regulated by bHLH39.

We initially struggled to explain why *ILR3* and *BHLH104* transcripts were down-regulated in FEP3-OX and *btsl1-btsl2* mutant lines, although all their putative downstream target genes were highly expressed. IVc *BHLH* transcription was never measured in *bts* or *btsl* loss-of-function lines by others. Our data suggests that ILR3/bHLH104 levels are also controlled transcriptionally, possibly for a case in which the proteins are not sufficiently degraded. This might be an additional layer of control to avoid excessive Fe uptake. This scenario actually could explain why neither *btsl1-btsl2*, nor *btsl1-btsl2 bts-3* triple mutants (Hindt et al., 2017) or - at least in our hands - FEP3-OX plants show signs of severe Fe toxicity under +Fe. It was recently reported that PYE represses *ILR3* expression (Samira et al., 2018), hence *ILR3* down-regulation in FEP3-OX and *btsl1-btsl2* can be explained by the elevated *PYE* levels. In another study, ILR3 dimerized with PYE to repress *PYE* transcription (Tissot et al., 2019). We propose that IVc bHLH proteins together with PYE control their own transcription. Overall, the regulatory cascade is not fully understood yet. For example, we expected that PYE and the genes which are negatively regulated by it (*NAS4/ZIF1/FRO3*; (Long et al., 2010)) have opposite expression patterns. However, although PYE was highly expressed in our mutant lines, *NAS4* and *FRO3* were not down-regulated. This fits other

reports investigating IVc bHLH gain-of-function lines (Zhang et al., 2015; Li et al., 2016), and indicates that PYE function can be bypassed or that IVc bHLHs and PYE act antagonistically to fine-tune downstream Fe acquisition and distribution genes.

A working model of Fe deficiency response regulation in Arabidopsis roots

Based on our results and state-of-the-art knowledge, we put forward a hypothetical model of Fe deficiency response regulation in Arabidopsis (**Figure 8A**): BTS, BTSL1 and BTSL2 negatively regulate Fe uptake genes by targeting individual sets of group IVc bHLHs (represented by ILR3 and bHLH104) and PYE for degradation. Interplay of IVc bHLHs and PYE on the protein as well as the transcriptional level impact Fe acquisition and Fe distribution genes positively and negatively, which ensures carefully balanced Fe uptake and homeostasis. In addition, *BTSL1/2* expression is regulated by bHLH39 in a negative feedback loop. Degradation of group IVc bHLHs and PYE by BTSL1/2 is counteracted by the positive Fe uptake regulator FEP3, which blocks BTSL1/2 function by physically interacting with the proteins (**Figure 8B**).

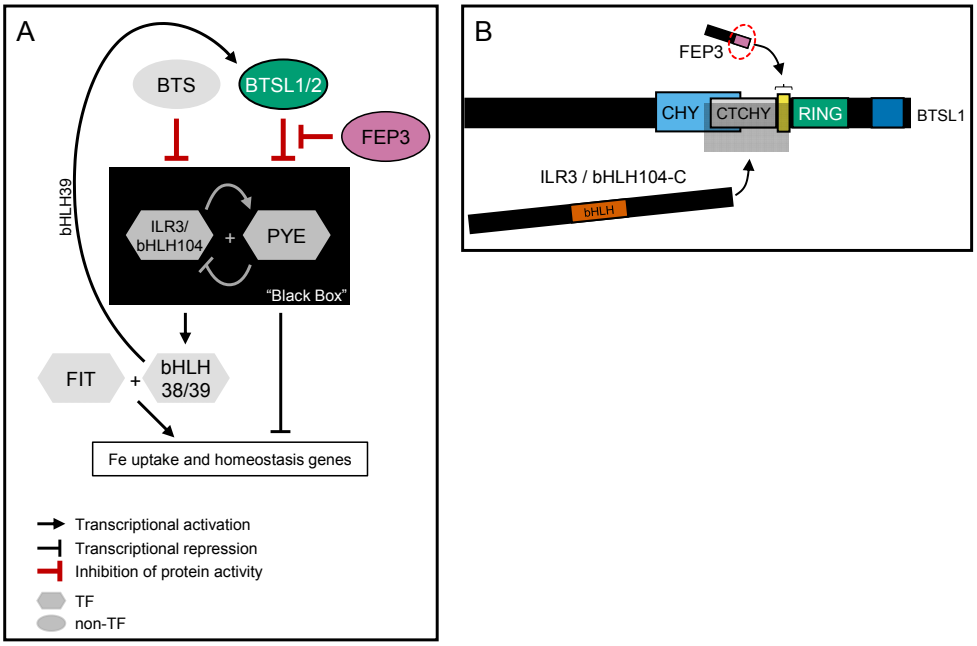


Figure 8. Working hypothesis of BTSL1/2 and FEP3 function, based on expression and protein-protein interaction data.
A: Fe deficiency response data of FEP3-OX and *btsl1-btsl2* lines from this study are incorporated into the current model of the Arabidopsis Fe homeostasis regulatory cascade (Selote et al., 2015; Zhang et al., 2015; Tissot et al., 2019). Going from bottom to top of the model, group Ib bHLHs (represented by bHLH38/39) as dimers with FIT directly induce Fe uptake genes (Yuan et al., 2008; Wang et al., 2013). PYE, possibly as dimer with ILR3, represses Fe homeostasis/distribution genes (*e.g. NAS4*) (Tissot et al., 2019). Group Ib *BHLHs* and *PYE* are transcriptionally up-regulated by group IVc bHLHs (represented by ILR3/bHLH104), which positively regulate Fe acquisition (Zhang et al., 2015; Liang et al., 2017). In turn, PYE represses *ILR3* transcription (Samira et al., 2018). In line with this, *ILR3* and *BHLH104* were down-regulated in our FEP3-OX and *btsl1-btsl2* lines while *PYE* was up-regulated, indicating that PYE might also repress transcription of other IVc *BHLHs* besides *ILR3*. IVc bHLHs and PYE are put in a "Black Box", since data regarding these proteins are not yet fully conclusive. In summary, IVc bHLHs and PYE can homo- and heterodimerize in different combinations, and activate or repress downstream genes, and also each other. BTS/BTSL1/BTSL2 physically interact with different "Black Box" proteins and presumably target some of them for degradation (BTS target: ILR3 (Selote et al., 2015)). *(legend continued on next page)*

We found that BTSL1/2 physically interacts with FEP3, a positive regulator of Fe uptake, and FEP3-OX plants phenocopy *btsl1-btsl2* mutant plants. Our model therefore presents FEP3 as inhibitor of BTSL1/2. *BTSL1/2* are possibly regulated by bHLH39 (but not FIT) in a negative feedback loop that transcriptionally induces its own upstream negative regulator (data of this study and (Mai et al., 2016; Naranjo-Arcos et al., 2017)). It should be noted that the model does not take into account different root tissues or cell types. **B:** Proposed mechanistic model of BTSL1 inhibition by FEP3. The model summarizes Y2H data from **Figures 3** and **4**. FEP3 interacts with BTSL1 in a 14-aa region depicted in yellow. The 7 aa at the FEP3 C-terminus are needed for the interaction (red dashed oval). ILR3 and bHLH104-C interact with a similar region of BTSL1 (gray shaded area, possibly also the yellow box). FEP3 might be an allosteric inhibitor, preventing that BTSL1 can bind and target the bHLHs for degradation. This allows the Fe acquisition cascade downstream of ILR3/bHLH104 (shown in A) to be active, and consequently Fe is taken up.

Concluding remarks

We provide evidence that FEP3 over-accumulation phenocopies *btsl1-btsl2* loss-of-function. FEP3 physically interacts with BTSL1 and BTSL2. Together, our data strongly suggests that FEP3 inhibits BTSL1/2 protein function, possibly by outcompeting BTSL1/2 degradation targets. However, additional experiments are needed to confirm that IVc bHLHs and/or PYE are targeted for degradation by BTSL1/2 and that this function is abolished by FEP3. Also, it needs to be clarified how the FEP3-BTSL interaction is disintegrated, once functional BTSLs are needed. Since FEP3 and BTSL1/2 have been attributed with Fe^{2+} binding abilities, we speculate that Fe itself might trigger disruption of the complex. Since the FEP family contains at least three members in Arabidopsis, we propose that BTS is regulated in a similar manner by one of the FEP3 homologs. A mechanism in which small peptides control the function of E3 ligases through direct binding has not been shown in plants yet. Therefore our study might have uncovered a novel fundamental mechanism of post-translational regulatory control in plants.

Methods

Plant Material

Arabidopsis (*Arabidopsis thaliana*) ecotype Columbia-0 (Col-0) was used as wild type (WT) and as background for transgenic lines. The *btsl1-btsl2* loss-of-function double mutant was described previously (*btsl1-1 btsl2-2*, crossed SALK_015054 and SALK_048470, (Rodríguez-Celma et al., 2017, bioRxiv)). *btsl1-btsl2* seeds were kindly provided by Dr. Janneke Balk, JIC, Norwich, UK. T-DNA insertion sites were verified with primer pairs LBb1.3/btsl1-1_RP (*btsl1*) and LBb1.3/btsl2-2_RP (*btsl2*) and homozygosity was verified with the primer pairs btsl1-1_LP/btsl1-1_RP and btsl2-2_LP/btsl2-2_RP (all primers are listed in **Supplemental Table S1**). For plant lines ectopically over-expressing triple HA-tagged *FEP3* (FEP3-OX) under the control of a double CaMV 35S promoter, the coding sequence was amplified from cDNA of Fe deficiency treated Arabidopsis WT roots with primers carrying B1 and B2 attachment sites, respectively, transferred into the entry vector pDONR207 (Invitrogen)

according to the manufacturer's recommendations (BP reaction, Gateway, Thermo Fisher Scientific) and sequenced. Final constructs were obtained by transferring all candidate genes subsequently into Gateway compatible plant binary destination vector pAlligator2 (N-terminal triple HA fusions=HA_3) (Bensmihen et al., 2004) via LR reactions (Thermo Fisher Scientific), and sequenced. Constructs were transformed into Agrobacteria (*Rhizobium radiobacter*) strain GV3101 (pMP90) (Koncz and Schell, 1986; Young et al., 2001). Stable transgenic Arabidopsis lines were generated via the Agrobacterium-mediated floral dip method (Clough and Bent, 1998). In brief, inflorescences of 6-7 weeks old Arabidopsis WT plants with removed open flowers and siliques were dipped around 20 s into a suspension (OD_{600}=0.8) of Agrobacteria in 5% (w/v) sucrose supplemented with 0.05% (v/v) Silwet Gold (Spiess-Urania). Positive transformants were selected based on seed GFP expression and genotyping PCR on the transgenic cassette, selfed and propagated to T3 generation. Insertion sites of the transgenic cassettes in FEP3-OX1 and FEP3-OX3 were determined by thermal asymmetric interlaced (TAIL) PCR (Liu and Whittier, 1995) with the primers S1_AL2_LB (template: gDNA), S2_AL2_LB (template: S1_AL2_LB amplicon), S3_AL2_LB (template: S2_AL2_LB amplicon), each combined with AD1, AD2, AD3, AD4, AD5, AD6. The insertion sites were verified by genotyping PCR with primer pairs FEP3-OX1_chr5 fw/S3_AL2_LB (FEP3-OX1), FEP3-OX3_chr1 fw/S3_AL2_LB (FEP3-OX3). Homozygosity was determined with primer pairs FEP3-OX1_chr5 fw/ FEP3-OX1_chr5 rev (FEP3-OX1), FEP3-OX3_chr1 fw/FEP3-OX3_chr1 rev (FEP3-OX3). T3 plants were used for physiological analyses. Promoter sequences of *BTS* (2,994 bp upstream of start codon), *BTSL1* (880 bp upstream of start codon), *PYE* (1,120 bp upstream of start codon) and *FEP3* (1,614 bp upstream of start codon) were amplified from Arabidopsis WT gDNA with primer pairs proBTS_-2994_B1 fw/proBTS_-2994_B2 rev (for *proBTS*), proBTSL1_-880_B1 fw/proBTSL1_-880_B2 rev (for *proBTSL1*), proPYE_-1120_B1 fw/proPYE_-1120_B2 rev (for *proPYE*), proFEP3_-1614_B1 fw/proFEP3_-1614_B2 rev (for *proFEP3*) and proFEP3_-1614_B1 fw/FEP3ns_B2 rev (for *proFEP3:FEP3*), respectively, cloned into the vector pDONR207 (Invitrogen) via the Gateway system (BP reaction, Thermo Fisher Scientific) and sequenced. Sequences were transferred into the Gateway-compatible destination vector pGWB3 (Nakagawa et al., 2007) (kindly provided by Dr. Andreas Weber, HHU, Düsseldorf, Germany) via LR reaction (Thermo Fisher Scientific), generating *proBTS:GUS*, *proBTSL1:GUS*, *proPYE:GUS*, *proFEP3:GUS* and *proFEP3:FEP3-GUS* constructs, followed by sequencing. Constructs were transformed into Arabidopsis WT plants as described above. Positive transformants were selected based on hygromycin resistance and genotyping PCR, selfed and propagated to T2 or T3 generation. *ProILR3:GUS*/WT and *proBHLH104:GUS*/WT Arabidopsis lines were kindly provided by Dr. Diqiu Yu, Chinese Academy of Sciences, Kunming, China. Three to four weeks old tobacco (*Nicotiana benthamiana*) plants were used for subcellular localization and interaction studies.

Plant Growth Conditions

For propagation and seed production, Arabidopsis seeds were surface-sterilized and stratified as described in (Lingam et al., 2011), germinated and cultivated on soil in a climate chamber under long day conditions (16 h light, 8 h dark, 21°C). For phenotypic analyses, physiological or histochemical assays, mRNA or protein extraction, surface-sterilized seeds were distributed to sterile plates containing modified half-strength Hoagland medium (1.5 mM $Ca(NO_3)_2$, 1.25 mM KNO_3, 0.75 mM $MgSO_4$, 0.5 mM KH_2PO_4, 50 µM KCl, 50 µM H_3BO_3, 10 µM $MnSO_4$, 2 µM $ZnSO_4$, 1.5 µM $CuSO_4$, 0.075 µM $(NH_4)_6Mo_7O_{24}$, 1% (w/v) sucrose, pH 5.8, and 1.4% w/v plant agar (Duchefa)) with (Fe sufficient, +Fe) or without (Fe deficient, -Fe) 50 µM FeNaEDTA. Seeds on plates were stratified at 4°C in the dark for two days and subsequently grown vertically in plant growth chambers (CLF Plant Climatics) under long day conditions. Seeds were germinated and grown for six or ten days directly on +Fe or -Fe medium (6 day (d) system/10 d system). Alternatively, seeds were germinated and grown for 14 days on +Fe medium and then transferred for three days to either +Fe or -Fe medium (14+3 d system). The 6 d system was used for whole-seedling mRNA extraction for RT-qPCR, protein extraction and root length measurements, the 10 d system was used for chlorophyll measurements and the 14+3 d system was used for root mRNA extraction for RT-qPCR.

Yeast Two-Hybrid (Y2H) Assays

Y2H screen: In a targeted manner, all 24 candidate genes were screened for interactions by pair-wise testing each as N-terminal AD and BD fusion proteins in both reciprocal combinations. pACT2-GW:*X* and pGBKT7-GW:*Y* constructs were used, with *X* and *Y* being the coding sequences of the tested gene pair. Coding sequences were amplified from cDNA of Fe deficiency treated Arabidopsis WT roots with primers carrying B1 and B2 attachment sites, respectively, transferred into the entry vector pDONR207 (Invitrogen) according to the manufacturer's recommendations (BP reaction, Gateway, Thermo Fisher Scientific) and sequenced. Final constructs were obtained by transferring all candidate genes subsequently into destination vectors pACT2-GW and pGBKT7-GW (kindly provided by Dr. Yves Jacob, Institut Pasteur, Paris, France) via LR reactions (Thermo Fisher Scientific), and sequenced. Yeast (*Saccharomyces cerevisiae*) strain Y187 was transformed with pACT2-GW:*X* (AD-X) constructs and yeast strain AH109 was transformed with pGBKT7-GW:*Y* (BD-Y) constructs via the lithium acetate (LiAc) method, based on (Gietz and Schiestl, 2007). In brief, 50 ml YPDA liquid culture (1% (w/v) yeast extract, 2% (w/v) Bacto peptone, 2% (w/v) glucose, 30 mg l^{-1} adenine hemisulfate), OD_{600}=0.5, were made competent with 100 mM LiAc. For transformation, 33.3% (w/v) PEG 4000, 0.1 M LiAc, 50 µg denatured Calf Thymus DNA (Invitrogen) and 0.3-0.7 µg of the construct were mixed in a total volume of 360 µl. Heat shock was carried out 20 min at 42°C. Transformants were selected by cultivation for 2 days on minimal synthetic defined (SD) media (Clontech) lacking Leu (pACT2-GW:*X*) or Trp (pGBKT7-

GW:*Y*). Yeast expressing both *AD-X* and *BD-Y* were obtained by mating. One fresh colony of each transgenic yeast carrying a pACT2-GW*:X* and a pGBKT7-GW*:Y* construct, respectively, were mixed in YPD (without adenine hemisulfate) liquid medium and cultivated 2-3 d on minimal SD media lacking Leu and Trp to select for diploid cells carrying both constructs. To test for protein-protein interaction, a fresh diploid colony was resuspended in sterile H_2O to OD_{600}=1 and 10 µl of the suspensions were dropped onto minimal SD media lacking Leu, Trp and His, containing appropriate concentrations of 3-amino-1,2,4-triazole (3AT, suppression of background growth). Plates were cultivated at 30°C for up to 14 days (growth documentation every 2-5 days). Diploid cells expressing each pACT2-GW:*X* construct in combination with an empty pGBKT7-GW and vice versa were used as negative controls. Combination of pGBT9.BS:*CIPK23* and pGAD.GH:*cAKT1* was used as a positive control of the system (Xu et al., 2006).

Targeted Y2H of selected protein pairs: Selected protein pairs of the Y2H screen and mutagenized/truncated protein versions were assayed as N-terminal AD and BD fusion proteins in both reciprocal combinations. Mutagenized *BTSL1* versions *BTSL1-dRH* and *BTSL1-6G* were created as described in "site-directed mutagenesis". Truncated versions *BTSL1-N*, *BTSL1-C*, *BTSL1-C.1*, *BTSL1-C.2*, *BTSL1-C.3*, *BTSL1-C.4*, *FEP3-N*, *FEP3-C*, *FEP3-d7*, *ILR3-d25*, *ILR3-CC*, *bHLH104-C*, *bHLH104-C-d25*, *bHLH104-CC* were amplified with primers listed in **Supplemental Table S1** and cloned into pACT2-GW and pGBKT7-GW as described in the previous section. Yeast strain AH109 was co-transformed with both pACT2-GW:*X* (AD-X) and pGBKT7-GW:*Y* (BD-Y) (including empty vector controls) as described in the previous section. X and Y represent proteins of a tested protein pair. Haploid double transformants were selected on minimal SD media lacking Leu and Trp. To select for protein-protein interaction, overnight liquid cultures were adjusted to OD_{600}=1 with sterile H_2O and dilution series down to $OD_{600}=10^{-4}$ were prepared. 10 µl of the suspensions were dropped onto SD media lacking Leu, Trp and His and containing the appropriate 3-AT concentration and cultivated as described in the previous section.

BTSL1 mutagenesis

Two deletion forms of BTSL1 lacking the region 1137-1142 (RTLVEH) were created: BTSL1-dRH (deletion) and BTSL1-6G (deletion spaced with GGGGGG). Deletion and 6G substitution were introduced by overlap-extension PCR as described in (Le et al., 2016). Two partially overlapping parts of each sequence were amplified from pDONR207:BTSL1 constructs (see "Plant Material"), using the primer pairs BTSL1 B1 fw/BTSL1_dRH rev, BTSL1_dRH fw/BTSL1 B2 rev (BTSL1-dRH), BTSL1 B1 fw/BTSL1_dRH-G rev, BTSL1_dRH-G fw/BTSL1 B2 rev (BTSL1-6G). Outer primers used to amplify the final construct are underlined. They carry B1 and B2 Gateway attachment sites. Amplicons were transferred into pDONR207, sequenced, subsequently transferred into Y2H destination

vectors pACT2-GW and pGBKT7-GW, transformed into yeast and assayed as described in the previous section.

Histochemical β-glucuronidase (GUS) Assay

Arabidopsis lines expressing *proBTS:GUS*, *proBTSL1:GUS, proILR3:GUS, proBHLH104:GUS, proPYE:GUS*, *proFEP3:GUS* and *proFEP3:FEP3-GUS* in WT background (see "Plant Material") were cultivated under Fe sufficient and Fe deficient conditions in the 6 d system (see "Plant Growth Conditions"). From *proBTS:GUS*, *proILR3:GUS, proBHLH104:GUS* and *proPYE:GUS* lines, four to six seedlings were transferred into 1 ml GUS staining solution based on (Jefferson et al., 1987), containing phosphate buffer (50 mM Na_2HPO_4, 50 mM NaH_2PO_4, pH 7.2) and substrate solution (2 mM $K_4[Fe(CN)_6]Fe^{2+}$, 2 mM $K_3[Fe(CN)_6]Fe^{3+}$, 0.2% Triton-X, 2 mM 5-bromo-4-chloro-3-indoyl-b-D-glucuronic acid (X-Gluc)) and incubated at 37°C in the dark for 15 min up to 12 h until blue tissue staining was visible. From *proBTSL1:GUS, proFEP3:GUS* and *proFEP3:FEP3-GUS* lines, four to six seedlings were fixed in ice cold 90% acetone for 1 h and washed in phosphate buffer prior to incubation in the GUS staining solution, which was vacuum infiltrated to obtain better staining. Incubation was performed as described above and stained tissue was fixed in 75% ethanol and 25% acetic acid for 2 h at RT. Chlorophyll of all stained seedlings was removed by incubation in 70% ethanol for 24 h (replaced by fresh ethanol occasionally). Seedlings were imaged with the Axio Imager M2 (Zeiss, 10x objective magnification). The Stitching function of the ZEN 2 BLUE Edition software (Zeiss) was applied to assemble single images to an image of the entire seedling.

Subcellular (co-) localization

To observe subcellular localization, proteins were tagged C-terminally to GFP and/or mCherry fluorophores and/or N-terminally to YFP fluorophore and expressed transiently in tobacco leaf epidermal cells. For N- and C-terminal fusions, coding sequences were amplified with and without the stop codon, respectively, from cDNA of Fe deficiency treated Arabidopsis WT roots with primers carrying B1 and B2 attachment sites, respectively, transferred into the entry vector pDONR207 (Invitrogen) via BP reactions (Gateway, Thermo Fisher Scientific), and sequenced. Final constructs were obtained by transferring all candidate genes subsequently into Gateway-compatible destination vectors pMDC83 (C-terminal GFP fusions) (Curtis and Grossniklaus, 2003), pH7WGY2 (N-terminal YFP) (Karimi et al., 2005), and β-estradiol-inducible pABind-GFP and pABind-mCherry (C-terminal GFP and mCherry fusions, used in co-localization studies) (Bleckmann et al., 2010) and into via LR reactions, and sequenced. The constructs were transformed into Agrobacteria as described in "Plant Material". A suspension (OD_{600}=0.4) of Agrobacteria carrying the construct of interest in infiltration solution (2 mM NaH_2PO_4, 0.5% (w/v) glucose, 50 mM MES, 100 µM acetosyringone (in DMSO), pH 5.6) was infiltrated into two leaves of the same tobacco plant with the help of a

1 ml syringe pressed to the abaxial leaf side. For co-localization of GFP-tagged and mCherry-tagged fusion proteins, corresponding Agrobacteria suspensions were mixed 1:1 (each to an OD_{600}=0.4) prior to infiltration. For more efficient expression, Agrobacteria carrying the p19 plasmid were co-infiltrated (suppression of RNA interference) (Voinnet et al., 2003, 2015). Images in **Figure 7** and **Supplemental Figure S13 (3)** were taken with added p19. Transformed plants were kept at RT under long day conditions (16 h light, 8 h dark) and punched-out 0.5 cm leaf discs were imaged after 48-72 h with a LSM 510 meta confocal laser scanning microscope (Zeiss; **Supplemental Figure S13 (1, 2)**) or an Axio Imager M2 with ApoTome (Zeiss; **Figure 2C**, **Figure 7**, **Supplemental Figure S13 (3)**). GFP and YFP were imaged at an excitation wavelength of 488 nm and emission wavelength of 500 to 530 nm, mCherry was imaged at an excitation wavelength at 563 nm and emission wavelength of 560 to 615 nm. Expression of pABind constructs was induced by spraying β-estradiol mix (20 µM β-estradiol (in DMSO), 0.1% (v/v) Tween20) to the abaxial leaf side 24-48 h post-infiltration. To capture the best moment of expression, which can vary between proteins, leaves were imaged 24 h and 48 h post-induction. The (co-)localization experiments were performed in at least two independent replicates with two infiltrated leaves each. Plasmolysis of cells expressing *BTSL1-GFP* was achieved through treatment of the leaf sample with 1 M mannitol solution for 30 min.

Bimolecular Fluorescence Complementation (BiFC)

Protein-protein interactions *in planta* were explored with BiFC in transiently transformed tobacco epidermal leaf cells. CDS' of gene pairs to be tested were amplified from cDNA of Fe deficiency treated Arabidopsis WT roots. Amplicons generated with primers carrying B3 and B2 attachment sites were transferred into pDONR221-P3P2 (Thermo Fisher Scientific, for nYFP fusion) and amplicons generated with primers carrying B1 and B4 attachment sites were transferred into pDONR221-P1P4 (Thermo Fisher Scientific, for cYFP fusion), respectively, via BP reactions (Gateway, Thermo Fisher Scientific). Primer sequences are listed in **Supplemental Table S1**. Inserts were sequenced. In a multisite Gateway LR reaction, both genes were transferred simultaneously into destination vector pBiFCt-2in1-NN (N-terminal nYFP and cYFP fusions) (Grefen and Blatt, 2012), performed according to the manufacturer's recommendations (Thermo Fisher Scientific), to create pBiFCt-2in1-NN:PYE-BTSL1, pBiFCt-2in1-NN:PYE-BTSL1-C and pBiFCt-2in1-NN:ILR3-BTSL1-C. The constructs carry a monomeric red fluorescent protein (*mRFP*) as internal transformation control. Construct accuracy was verified by sequencing. Because "empty" split-YFP fusion proteins tend to self-assemble, thereby giving false positive signals (Kudla and Bock, 2016), we used structurally similar proteins known to not interact from previous experiments as negative controls (negative controls: pBiFCt-2in1-NN:ILR3-BTSL2-C, pBiFCt-2in1-NN:FIT-BTSL1-C). Constructs were transformed into Agrobacteria and subsequently infiltrated into tobacco leaves, as described

in the previous section. 48 h to 52 h after infiltration, mRFP and YFP signals were detected in punched-out 0.5 cm leaf discs with an Axio Imager M2 (Zeiss). YFP was imaged at an excitation wavelength of 488 nm and emission wavelength of 500 to 530 nm, mRFP was imaged at an excitation wavelength at 563 nm and emission wavelength of 560 to 615 nm. The BiFC experiments were performed in two independent replicates with two infiltrated leaves each. The vector pBiFCt-2in1-NN was kindly provided by Dr. Christopher Grefen, RUB, Bonn, Germany.

Gene Expression Analysis by RT-qPCR

Gene expression analysis was performed as described previously (Abdallah and Bauer, 2016). In brief, mRNA was extracted from whole seedlings grown in the 6 d system (n>60 per replicate) or from roots grown in the 14+3 d system (n>15 per replicate) (see “Plant Growth Conditions”) and used for cDNA synthesis. RT-qPCR was performed using the iTaq™ Universal SYBR® Green Supermix (Bio-Rad) according to the manufacturer’s recommendations and the SFX96 Touch™ RealTime PCR Detection System (Bio-Rad). Data was processed with the Bio-Rad SFX Manager™ software (version 3.1). Absolute gene expression values were calculated from a gene specific mass standard dilution series and normalized to the Arabidopsis elongation factor *EF1Bα*. Primers for mass standards and RT-qPCR are listed in **Supplemental Table S1**. The analysis was performed with three biological and two technical replicates.

Immunoblot analysis

Total proteins were extracted from ground plant material (tobacco leaves or Arabidopsis whole seedlings grown in the 6 d system, n=30-60 seedlings) with 2x Laemmli buffer (124 mM Tris-HCl, pH 6.8, 5% (w/v) SDS, 4% (w/v) dithithreitol, 20% (v/v) glycerol, with 0.002% (w/v) bromophenol blue) and denatured at 95°C for 10 min. Equal amounts of total protein (max. volume 15 µl) were separated on SDS-polyacrylamide gels, proteins were transferred to a Protran nitrocellulose membrane and stained with PonceauS as described in (Le et al., 2016). To detect HA_3-tagged FEP3 protein, the nitrocellulose membrane was blocked with 5% (w/v) milk powder solution (Roth) in 1xPBST (137 mM NaCl, 2.7 mM KCl, 10.14 mM Na_2HPO_4, 1.76 mM KH_2PO_4, 0.1% (v/v) Tween® 20, pH 7.4) for 30 min to avoid unspecific antibody (AB) binding, and subsequently incubated 1 h with anti-HA-peroxidase high-affinity monoclonal rat antibody (3F10; Roche [catalog no. 12013819001]) diluted 1:1,000 in 2.5% (w/v) milk solution. After three wash steps, each for 15 min in PBST, the nitrocellulose membrane was imaged as described in (Le et al., 2016). Chemiluminescent protein bands were detected with the FluorChem Q system (ProteinSimple) and images were processed with the AlphaView® software (version 3.4.0.0, ProteinSimple).

Root Length Measurement

Arabidopsis WT, FEP3-OX and *btsl1-btsl2* DM lines were cultivated in the 6 d system (see "Plant Growth Conditions") and photographed at day six. Length of primary roots of individual seedlings was measured using the JMicroVision software (version 1.2.7, http://www.jmicrovision.com), as described previously (Ivanov et al., 2014). For calculation of mean root lengths and standard deviations, n=13-29 roots per line and condition were measured.

Colorimetric Seed Fe Content Measurement

To determine seed Fe content, 1-3 plants from each line were grown on soil under long day conditions (16 h light, 8 h dark, 21°C). Seeds were harvested, pooled by plant genotype, and dried for 16 h at 100°C. Fe was extracted from ground seed material by incubation in 500 µl 3% (v/v) HNO3 for 16 h at 100°C. Fe^{2+} content in the supernatant was determined as described in (Tamarit et al., 2006). In short, to 400 µl supernatant, 160 µl (38 g l^{-1}) $C_6H_7NaO_6$ (sodium ascorbate) was added to prevent oxidation, 320 µl (1.7 g l^{-1}) bathophenanthrolinedisulfonic acid disodium salt (BPDS, Fe^{2+} chelator) was added and 126 µl of a 33% $C_2H_7NO_2$ (ammonium acetate) solution (1 part saturated solution was diluted with 2 parts H_2O) was added to buffer the system. After an incubation time for 5 min at RT, absorption of the Fe^{2+}-BPDS complex was measured at a wavelength of 535 nm. Total Fe^{2+} content in the sample was calculated with the help of a standard curve (3.125 µM, 6.25 µM, 12.5 µM, 25 µM, 50 µM FeNaEDTA in 3% (v/v) HNO3) and normalized to seed dry weight (typically 1-5 mg). Per seed pool, n=3 samples were measured.

Chlorophyll Content Measurement

Chlorophyll content was measured in leaves of Arabidopsis lines grown in the 10 d system. Chlorophyll a and chlorophyll b were extracted with 100% acetone added to ground leaf material. The supernatant was collected and the washes were repeated until the collected acetone remained colorless. Absorbances of chlorophyll a and b were measured at a wavelength of 662 nm and 645 nm, respectively. Using the absorption coefficients that apply to 100% acetone, chlorophyll a concentration in the measured sample was calculated with $c_{Chl\text{-}a}$ [mg μl^{-1}]=11.75*A_{662}-2.35*A_{645} and chlorophyll b concentration was calculated with $c_{Chl\text{-}b}$ [mg μl^{-1}]=18.61*A_{645}-3.96*A_{662} (Lichtenthaler and Wellburn, 1983), where A is the absorbance at the indicated wavelength in nm. Values were normalized to fresh weight. Four biological replicates were measured, with each biological replicate containing n=5-7 rosettes.

Multiple Sequence alignments and protein sequence conservation

Multiple sequence alignments of BTS/BTSL1/BTSL2, ILR3/bHLH34/104/115/FEP3 and FEP3 orthologs were performed with ClustalX using default settings (Larkin et al., 2007). To

determine conservation scores of aa in BTSL1, the full BTSL1 aa sequence was uploaded to the Basic Local Alignment Search Tool (BLAST, (Altschul et al., 1990); https://blast.ncbi.nlm.nih.gov/) and run against the Viridiplantae database using the standard blastp (protein-protein BLAST) algorithm. The top 100 hits were downloaded, duplicates were removed and the remaining sequences were used for multiple sequence alignment using the Clustal Omega algorithm (Sievers et al., 2011) and visualized with Jalview ((Waterhouse et al., 2009); http://www.jalview.org/). The full aa sequence of FEP3 run against the Viridiplantae database as described above. Hits were only found within the Brassicaceae family, but alignments showed sequence conservation specifically towards the C-terminus. Subsequent blastp of the C-terminal half of FEP3 (25 aa) resulted in several angiosperm hits. FEP3 ortholog sequence hits from exemplary angiosperm orders were downloaded and aligned.

Statistical Analysis

Null hypothesis between normally distributed groups was tested with a two-tailed Student's t-test. Null hypothesis was rejected, when the *p*-value (*p*) was below 0.05. Statistically significantly different groups are indicated by one asterisk for $p<0.05$, two asterisks for $p<0.01$ and three asterisks for $p<0.001$. When comparing more than two groups, null hypotheses were tested with one-way analysis of variance (ANOVA) and a Tukey's post-hoc test. Null hypotheses were rejected when $p<0.05$. Statistically significantly different groups are indicated by different lower-case letters. Number of technical and biological repetitions of the individual experiments are indicated in the respective Methods sections and in the Figure legends.

Accession Numbers

AKT1 (AT2G26650), BHLH34 (AT3G23210), BHLH38 (AT3G56970), BHLH39 (AT3G56980), BHLH100 (AT2G41240), BHLH101 (AT5G04150), BHLH104 (AT4G14410), BTS (AT3G18290), BTSL1 (AT1G74770), BTSL2 (AT1G18910), CIPK23 (AT1G30270), DGAT3 (AT1G48300), DUF506 (AT1G12030), FEP3 (AT1G47400), FIT (AT2G28160), FRO2 (AT1G01580), FRO3 (AT1G23020), GRF11 (AT1G34760), ILR3 (AT5G54680), IRT1 (AT4G19690), JAL12 (AT1G52120), KELCH (AT3G07720), MYB72 (AT1G56160), NAS2 (AT5G56080), NAS4 (AT1G56430), ORG1 (AT5G53450), PRS2 (AT1G32380), PYE (AT3G47640), SDI1 (AT5G48850), S8H (AT3G12900), TCP20 (AT3G27010), UIP1 (AT1G73120), UIP2 (AT3G06890), UIP3 (AT3G56360), UIP4 (AT5G05250)

Supplemental Material

Supplemental Figure S1. Fe deficiency co-expression network and Y2H screen interaction summary.

Supplemental Figure S2. Original Y2H screen images (part 1).

Supplemental Figure S3. Original Y2H screen images (part 2).

Supplemental Figure S4. Verified "BTS(L) interactome" protein-protein interactions and non-interactions.

Supplemental Figure S5. Verified non-interactions between BTS, BTSL1, BTSL2 and different bHLH TFs.

Supplemental Figure S6. BTSL1/BTSL2/BTS multiple sequence alignment, protein domains and regions relevant in this study.

Supplemental Figure S7. FEP3 conserved domain.

Supplemental Figure S8. Multiple sequence alignment of IVc bHLHs and FEP3 to identify potential sequence similarities.

Supplemental Figure S9. The last 25 aa in ILR3 and bHLH104 are not crucial for interaction with BTSL1 or BTSL2.

Supplemental Figure S10. Validation of FEP3-OX and *btsl1-btsl2* mutant lines.

Supplemental Figure S11. Regulation of selected Fe deficiency response genes in whole seedlings of FEP3-OX and *btsl1-btsl2*.

Supplemental Figure S12. *ProBTS* activity in two additional Arabidopsis lines, and FEP3-GUS protein localization in Arabidopsis seedlings.

Supplemental Figure S13. Peripheral localization of BTSL1 in tobacco leaf epidermis cells (additional images).

Supplemental Table S1. Primers used in this study.

Acknowledgements

We thank Elke Wieneke and Gintaute Matthäi for excellent technical assistance. We acknowledge the contributions of Sarah Plicht, Theresa Priebe, and Kai Blaeser. We thank Rumen Ivanov, Ksenia Trofimov, Inga Mohr, and Tzvetina Brumbarova for help and advice with microscopy and with the implementation of lab protocols. B.S. and D.M.L. are members of the international graduate school iGRAD-Plant, Düsseldorf. Funding from the German Research Foundation through the DFG International Research Training group 1525 to P.B. is greatly acknowledged. This work received funding from Germany's Excellence Strategy, EXC 2048/1, Project ID: 390686111.

Literature Cited

Abdallah HB, Bauer P (2016) Quantitative reverse transcription-qPCRbased gene expression analysis in plants. Methods Mol Biol 1363: 9-24

Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990) Basic local alignment search tool. Journal of Molecular Biology 215: 403-410

Aoki Y, Okamura Y, Tadaka S, Kinoshita K, Obayashi T (2016) ATTED-II in 2016: a plant coexpression database towards lineage-specific coexpression. Plant Cell Physiol 57: e5

Bauer P, Ling HQ, Guerinot ML (2007) FIT, the FER-LIKE IRON DEFICIENCY INDUCED TRANSCRIPTION FACTOR in Arabidopsis. Plant Physiol Biochem 45: 260-261

Belda-Palazon B, Julian J, Coego A, Wu Q, Zhang X, Batistic O, Alquraishi SA, Kudla J, An C, Rodriguez PL (2019) ABA inhibits myristoylation and induces shuttling of the RGLG1 E3 ligase to promote nuclear degradation of PP2CA. Plant J

Bensmihen S, To A, Lambert G, Kroj T, Giraudat J, Parcy F (2004) Analysis of an activated ABI5 allele using a new selection method for transgenic Arabidopsis seeds. Febs Letters 561: 127-131

Bleckmann A, Weidtkamp-Peters S, Seidel CAM, Simon R (2010) Stem cell signaling in Arabidopsis requires CRN to localize CLV2 to the plasma membrane. Plant Physiology 152: 166-176

Briat JF, Curie C, Gaymard F (2007) Iron utilization and metabolism in plants. Current Opinion in Plant Biology 10: 276-282

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

Cabrera-Quio LE, Herberg S, Pauli A (2016) Decoding sORF translation - from small proteins to gene regulation. Rna Biology 13: 1051-1059

Cheng MC, Hsieh EJ, Chen JH, Chen HY, Lin TP (2012) Arabidopsis RGLG2, functioning as a RING E3 ligase, interacts with AtERF53 and negatively regulates the plant drought stress response. Plant Physiology 158: 363-375

Cheng MC, Hsieh EJ, Chen JH, Chen HY, Lin TP (2016) CORRECTION. Arabidopsis RGLG2, Functioning as a RING E3 Ligase, Interacts with AtERF53 and Negatively Regulates the Plant Drought Stress Response. Plant Physiology 170: 1162-1163

Clough SJ, Bent AF (1998) Floral dip: a simplified method for *Agrobacterium*-mediated transformation of *Arabidopsis thaliana*. Plant Journal 16: 735-743

Colangelo EP, Guerinot ML (2004) The essential basic helix-loop-helix protein FIT1 is required for the iron deficiency response. Plant Cell 16: 3400-3412

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

Curtis MD, Grossniklaus U (2003) A gateway cloning vector set for high-throughput functional analysis of genes in planta. Plant Physiology 133: 462-469

Delcourt V, Staskevicius A, Salzet M, Fournier I, Roucou X (2018) Small proteins encoded by unannotated ORFs are rising stars of the proteome, confirming shortcomings in genome annotations and current vision of an mRNA. Proteomics 18

Dong CH, Agarwal M, Zhang Y, Xie Q, Zhu JK (2006) The negative regulator of plant cold responses, HOS1, is a RING E3 ligase that mediates the ubiquitination and degradation of ICE1. Proc Natl Acad Sci U S A 103: 8281-8286

Duffy JR, Gaudin J (1977) Copper interference in the determination of iron in serum using ferrozine. Clin Biochem 10: 122-123

Eide D, Broderius M, Fett J, Guerinot ML (1996) A novel iron-regulated metal transporter from plants identified by functional expression in yeast. Proc Natl Acad Sci U S A 93: 5624-5628

Freemont PS (2000) Ubiquitination: RING for destruction? Curr Biol 10: R84-87

Gao F, Robe K, Gaymard F, Izquierdo E, Dubos C (2019) The transcriptional control of iron homeostasis in plants: a tale of bHLH transcription factors? Frontiers in Plant Science 10

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

García MJ, Corpas FJ, Lucena C, Alcántara E, Pérez-Vicente R, Zamarreño Á, Bacaicoa E, García-Mina JM, Bauer P, Romera FJ (2018) A shoot Fe signaling pathway requiring the OPT3 transporter controls GSNO reductase and ethylene in *Arabidopsis thaliana* roots. Front Plant Sci 9: 1325

Gietz RD, Schiestl RH (2007) High-efficiency yeast transformation using the LiAc/SS carrier DNA/PEG method. Nature Protocols 2: 31-34

Graceffa P, Jancsó A, Mabuchi K (1992) Modification of acidic residues normalizes sodium dodecyl sulfate-polyacrylamide gel electrophoresis of caldesmon and other proteins that migrate anomalously. Archives of Biochemistry and Biophysics 297: 46-51

Grefen C, Blatt MR (2012) A 2in1 cloning system enables ratiometric bimolecular fluorescence complementation (rBiFC). Biotechniques 53: 311-314

Grillet L, Lan P, Li WF, Mokkapati G, Schmidt W (2018) IRON MAN is a ubiquitous family of peptides that control iron transport in plants. Nature Plants 4: 953-+

Heim MA, Jakoby M, Werber M, Martin C, Weisshaar B, Bailey PC (2003) The basic helix-loop-helix transcription factor family in plants: a genome-wide study of protein structure and functional diversity. Mol Biol Evol 20: 735-747

Hell R, Stephan UW (2003) Iron uptake, trafficking and homeostasis in plants. Planta 216: 541-551

Hindt MN, Akmakjian GZ, Pivarski KL, Punshon T, Baxter I, Salt DE, Guerinot ML (2017) BRUTUS and its paralogs, BTS LIKE1 and BTS LIKE2, encode important negative regulators of the iron deficiency response in *Arabidopsis thaliana*. Metallomics 9: 876-890

Hirayama T, Lei GJ, Yamaji N, Nakagawa N, Ma JF (2018) The putative peptide gene FEP1 regulates iron deficiency response in Arabidopsis. Plant and Cell Physiology 59: 1739-1752

Hsu PY, Benfey PN (2018) Small but mighty: functional peptides encoded by small ORFs in plants. Proteomics 18: e1700038

Huang TK, Han CL, Lin SI, Chen YJ, Tsai YC, Chen YR, Chen JW, Lin WY, Chen PM, Liu TY, Chen YS, Sun CM, Chiou TJ (2013) Identification of downstream components of ubiquitin-conjugating enzyme PHOSPHATE2 by quantitative membrane proteomics in Arabidopsis roots. Plant Cell 25: 4044-4060

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

Ivanov R, Brumbarova T, Blum A, Jantke AM, Fink-Straube C, Bauer P (2014) SORTING NEXIN1 is required for modulating the trafficking and stability of the Arabidopsis IRON-REGULATED TRANSPORTER1. Plant Cell 26: 1294-1307

Jakoby M, Wang HY, Reidt W, Weisshaar B, Bauer P (2004) FRU (BHLH029) is required for induction of iron mobilization genes in *Arabidopsis thaliana*. FEBS Lett 577: 528-534

Jefferson RA, Kavanagh TA, Bevan MW (1987) GUS fusions: β-glucuronidase as a sensitive and versatile gene fusion marker in higher plants. Embo Journal 6: 3901-3907

Karimi M, De Meyer B, Hilson P (2005) Modular cloning in plant cells. Trends in Plant Science 10: 103-105

Khan MA, Castro-Guerrero NA, McInturf SA, Nguyen NT, Dame AN, Wang J, Bindbeutel RK, Joshi T, Jurisson SS, Nusinow DA, Mendoza-Cozatl DG (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

Klatte M, Schuler M, Wirtz M, Fink-Straube C, Hell R, Bauer P (2009) The analysis of Arabidopsis nicotianamine synthase mutants reveals functions for nicotianamine in seed iron loading and iron deficiency responses. Plant Physiol 150: 257-271

Klose C, Viczián A, Kircher S, Schäfer E, Nagy F (2015) Molecular mechanisms for mediating light-dependent nucleo/cytoplasmic partitioning of phytochrome photoreceptors. New Phytol 206: 965-971

Kobayashi T, Nagasaka S, Senoura T, Itai RN, Nakanishi H, Nishizawa NK (2013) Iron-binding haemerythrin RING ubiquitin ligases regulate plant iron responses and accumulation. Nat Commun 4: 2792

Kobayashi T, Nozoye T, Nishizawa NK (2018) Iron transport and its regulation in plants. Free Radic Biol Med

Koncz C, Schell J (1986) The promoter of T_L-DNA gene *5* controls the tissue-specific expression of chimeric genes carried by a novel type of Agrobacterium binary vector. Molecular & General Genetics 204: 383-396

Kudla J, Bock R (2016) Lighting the way to protein-protein interactions: recommendations on best practices for Bimolecular Fluorescence Complementation analyses. Plant Cell 28: 1002-1008

Kumar RK, Chu HH, Abundis C, Vasques K, Rodriguez DC, Chia JC, Huang R, Vatamaniuk OK, Walker EL (2017) Iron-nicotianamine transporters are required for proper long distance iron signaling. Plant Physiol 175: 1254-1268

Larkin MA, Blackshields G, Brown NP, Chenna R, McGettigan PA, McWilliam H, Valentin F, Wallace IM, Wilm A, Lopez R, Thompson JD, Gibson TJ, Higgins DG (2007) Clustal W and clustal X version 2.0. Bioinformatics 23: 2947-2948

Le CTT, Brumbarova T, Ivanov R, Stoof C, Weber E, Mohrbacher J, Fink-Straube C, Bauer P (2016) ZINC FINGER OF ARABIDOPSIS THALIANA12 (ZAT12) interacts with FER-LIKE IRON DEFICIENCY-INDUCED TRANSCRIPTION FACTOR (FIT) linking iron deficiency and oxidative stress responses. Plant Physiology 170: 540-557

Lee H, Xiong L, Gong Z, Ishitani M, Stevenson B, Zhu JK (2001) The Arabidopsis HOS1 gene negatively regulates cold signal transduction and encodes a RING finger protein that displays cold-regulated nucleo-cytoplasmic partitioning. Genes & development 15: 912-924

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: 2478-2493

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

Lichtenthaler HK, Wellburn AR (1983) Determinations of total carotenoids and chlorophylls a and b of leaf extracts in different solvents. Biochemical Society Transactions 11: 591-592

Lindsay WL (1988) Solubility and redox equilibria of iron compounds in soils. *In* S J.W., G B.A., S U., eds, Iron in Soils and Clay Minerals, Vol 217. Springer, Dordrecht, p 894

Lingam S, Mohrbacher J, Brumbarova T, Potuschak T, Fink-Straube C, Blondet E, Genschik P, Bauer P (2011) Interaction between the bHLH transcription factor FIT and ETHYLENE INSENSITIVE3/ETHYLENE INSENSITIVE3-LIKE1 reveals molecular linkage between the regulation of iron acquisition and ethylene signaling in Arabidopsis. Plant Cell 23: 1815-1829

Liu YG, Whittier RF (1995) Thermal Asymmetric Interlaced PCR: automatable amplification and sequencing of insert end fragments from P1 and YAC clones for chromosome walking. Genomics 25: 674-681

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

Mai HJ (2013) Analysis of differential protein and gene expression in *Arabidopsis thaliana* depending on iron supply and the abundance of the central iron uptake regulator FIT (FER-LIKE IRON DEFICIENCY-INDUCED TRANSCRIPTION FACTOR) and investigations on possible post-translational modifications of FIT. Saarländische Universitäts- und Landesbibliothek [Dissertation]: doi:10.22028/D22291-22974

Mai 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

Makarewich CA, Olson EN (2017) Mining for Micropeptides. Trends Cell Biol 27: 685-696

Marschner H, Römheld V (1994) Strategies of plants for acquisition of iron. Plant and Soil 165: 261-274

Matthiadis A, Long TA (2016) Further insight into BRUTUS domain composition and functionality. Plant Signal Behav 11: e1204508

Nakagawa T, Kurose T, Hino T, Tanaka K, Kawamukai M, Niwa Y, Toyooka K, Matsuoka K, Jinbo T, Kimura T (2007) Development of series of gateway binary vectors, pGWBs, for realizing efficient construction of fusion genes for plant transformation. Journal of Bioscience and Bioengineering 104: 34-41

Naranjo-Arcos MA, Maurer F, Meiser J, Pateyron S, Fink-Straube C, Bauer P (2017) Dissection of iron signaling and iron accumulation by overexpression of subgroup Ib bHLH039 protein. Scientific Reports 7

Nouet C, Motte P, Hanikenne M (2011) Chloroplastic and mitochondrial metal homeostasis. Trends Plant Sci 16: 395-404

Ohkubo Y, Tanaka M, Tabata R, Ogawa-Ohnishi M, Matsubayashi Y (2017) Shoot-to-root mobile polypeptides involved in systemic regulation of nitrogen acquisition. Nat Plants 3: 17029

Palmer CM, Hindt MN, Schmidt H, Clemens S, Guerinot ML (2013) MYB10 and MYB72 are required for growth under iron-limiting conditions. PLoS Genet 9: e1003953

Peterman TK, Ohol YM, McReynolds LJ, Luna EJ (2004) Patellin1, a novel Sec14-like protein, localizes to the cell plate and binds phosphoinositides. Plant Physiology 136: 3080-3094; discussion 3001-3082

Robinson NJ, Procter CM, Connolly EL, Guerinot ML (1999) A ferric-chelate reductase for iron uptake from soils. Nature 397: 694-697

Rodríguez-Celma J, Chou H, Kobayashi T, Long TA, Balk J (2019) Hemerythrin E3 Ubiquitin Ligases as Negative Regulators of Iron Homeostasis in Plants. Front Plant Sci 10: 98

Rodríguez-Celma J, Green RT, Connorton JM, Kruse I, Cui Y, Ling HQ, Balk J (2017) BRUTUS-LIKE proteins moderate the transcriptional response to iron deficiency in roots. *bioRxiv* [Preprint]: doi.org/10.1101/231365

Salahudeen AA, Thompson JW, Ruiz JC, Ma HW, Kinch LN, Li Q, Grishin NV, Bruick RK (2009) An E3 ligase possessing an iron-responsive hemerythrin domain is a regulator of iron homeostasis. Science 326: 722-726

Samira R, Li BH, Kliebenstein D, Li CY, Davis E, Gillikin JW, Long TA (2018) The bHLH transcription factor ILR3 modulates multiple stress responses in Arabidopsis. Plant Molecular Biology 97: 297-309

Schmid NB, 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

Schuler M, Rellán-Álvarez R, Fink-Straube C, Abadía J, Bauer P (2012) Nicotianamine functions in the phloem-based transport of iron to sink organs, in pollen development and pollen tube growth in Arabidopsis. Plant Cell 24: 2380-2400

Selote D, Samira R, Matthiadis 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

Sievers F, Wilm A, Dineen D, Gibson TJ, Karplus K, Li WZ, Lopez R, McWilliam H, Remmert M, Söding J, Thompson JD, Higgins DG (2011) Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. Molecular Systems Biology 7

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 Journal 66: 1044-1052

Sivitz AB, Hermand V, Curie C, Vert G (2012) Arabidopsis bHLH100 and bHLH101 control iron homeostasis via a FIT-independent pathway. PLoS One 7: e44843

Stührwohldt N, Schaller A (2019) Regulation of plant peptide hormones and growth factors by post-translational modification. Plant Biol (Stuttg) 21 Suppl 1: 49-63

Tabata R, Sumida K, Yoshii T, Ohyama K, Shinohara H, Matsubayashi Y (2014) Perception of root-derived peptides by shoot LRR-RKs mediates systemic N-demand signaling. Science 346: 343-346

Tamarit J, Irazusta V, Moreno-Cermeño A, Ros J (2006) Colorimetric assay for the quantitation of iron in yeast. Analytical Biochemistry 351: 149-151

Tissot N, Robe K, Gao F, Grant-Grant S, Boucherez J, Bellegarde F, Maghiaoui A, Marcelin R, Izquierdo E, Benhamed M, Martin A, Vignols F, Roschzttardtz H, Gaymard F, Briat JF, Dubos C (2019) Transcriptional integration of the responses to iron availability in Arabidopsis by the bHLH factor ILR3. New Phytol

Tottey S, Block MA, Allen M, Westergren T, Albrieux C, Scheller HV, Merchant S, Jensen PE (2003) Arabidopsis CHL27, located in both envelope and thylakoid membranes, is required for the synthesis of protochlorophyllide. Proc Natl Acad Sci U S A 100: 16119-16124

Van Buskirk EK, Decker PV, Chen M (2012) Photobodies in light signaling. Plant Physiol 158: 52-60

Vert G, Grotz N, Dédaldéchamp F, Gaymard F, Guerinot ML, Briat JF, Curie C (2002) IRT1, an Arabidopsis transporter essential for iron uptake from the soil and for plant growth. Plant Cell 14: 1223-1233

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 Physiology 132: 796-804

Voinnet O, Rivas S, Mestre P, Baulcombe D (2003) An enhanced transient expression system in plants based on suppression of gene silencing by the p19 protein of tomato bushy stunt virus (Retracted article. See vol. 84, pg. 846, 2015). Plant Journal 33: 949-956

Voinnet O, Rivas S, Mestre P, Baulcombe D (2015) An enhanced transient expression system in plants based on suppression of gene silencing by the p19 protein of tomato bushy stunt virus (Retraction of Vol 33, Pg 949, 2003). Plant Journal 84: 846-846

von Wirén N, Klair S, Bansal S, Briat JF, Khodr H, Shioiri T, Leigh RA, Hider RC (1999) Nicotianamine chelates both Fe-III and Fe-II. Implications for metal transport in plants. Plant Physiology 119: 1107-1114

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

Waterhouse AM, Procter JB, Martin DMA, Clamp M, Barton GJ (2009) Jalview Version 2-a multiple sequence alignment editor and analysis workbench. Bioinformatics 25: 1189-1191

Wedepohl KH (1995) The composition of the continental crust. Geochimica Et Cosmochimica Acta 59: 1217-1232

Xu J, Li HD, Chen LQ, Wang Y, Liu LL, He L, Wu WH (2006) A protein kinase, interacting with two calcineurin B-like proteins, regulates K^+ transporter AKT1 in Arabidopsis. Cell 125: 1347-1360

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

Young JM, Kuykendall LD, Martinez-Romero E, Kerr A, Sawada H (2001) A revision of Rhizobium Frank 1889, with an emended description of the genus, and the inclusion of all species of *Agrobacterium* Conn 1942 and *Allorhizobium undicola* de Lajudie *et al.* 1998 as new combinations: *Rhizobium radiobacter*, *R. rhizogenes*, *R. rubi*, *R. undicola* and *R. vitis*. International Journal of Systematic and Evolutionary Microbiology 51: 89-103

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

Zanet J, Benrabah E, Li T, Pélissier-Monier A, Chanut-Delalande H, Ronsin B, Bellen HJ, Payre F, Plaza S (2015) Pri sORF peptides induce selective proteasome-mediated protein processing. Science 349: 1356-1358

Zhai Z, Gayomba SR, Jung HI, Vimalakumari NK, Piñeros M, Craft E, Rutzke MA, Danku J, Lahner B, Punshon T, Guerinot ML, Salt DE, Kochian LV, Vatamaniuk OK (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 J, Liu B, 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

Supplemental Figures and Tables

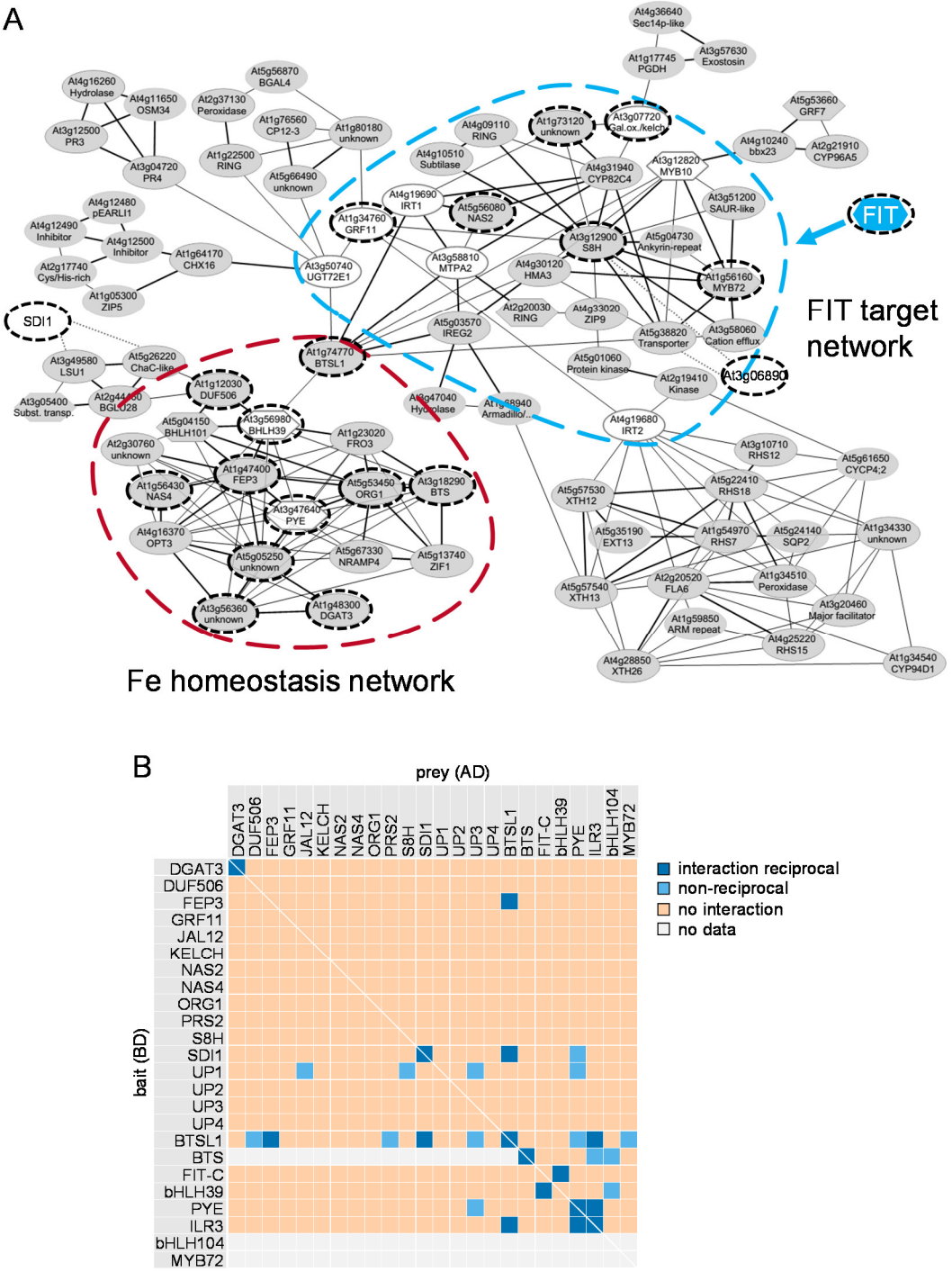


Figure S1. Fe deficiency co-expression network and Y2H screen interaction summary.
A: Genes up-regulated under Fe deficiency stress. Highlighted sub-networks: Root-specific FIT target genes (blue dashed line), root- and shoot-expressed FIT-independent genes (red dashed line). The network was generated with the ATTED-II tool (Ver. 8.0), using the white labeled genes as basis, and visualized with Cytoscape. The network is an updated version of (Ivanov et al., 2012). Candidate genes used in the Y2H screen are highlighted with a dashed border. Candidate genes *ILR3*, *BHLH104*, *PRS2*, *JAL12* do not appear in the network. SDI1 and UP2 appeared in an older ATTED-II (Ver. 7.1) network version, initially used to select candidates.
B: Summary of all protein-protein interactions detected with the Y2H screen. 24 candidates were tested reciprocally in all possible combinations. Bait: protein fused to the GAL4 DNA-binding domain (BD), prey: protein fused to the GAL4 activation domain (AD). The color code distinguishes reciprocal interactions (dark blue), non-reciprocal interactions (light blue), no interactions (light orange). Gray: no data. BD-bHLH104 and BD-MYB72 could not be tested because the proteins auto-activated the Y2H system. Original images and negative controls: **Supplemental Figures S2, S3**.

Figure S2. Original Y2H screen images (part 1).
Targeted Y2H screen between 17 candidates (all non-TFs except BTS) in all pair-wise combinations including homodimerization. Yeast containing both AD- and BD-plasmids were obtained by mating of single transformants and spotted as A_{600}=1 on SD-LW (mating control) and SD-LWH supplemented with different concentrations 3AT (selection for protein interaction). Positive control (+ control): CIPK23/cAKT1 (Xu et al., 2006); negative controls (empty): AD-protein/empty and BD-protein/empty. Rows should be read in pairs: Each upper row tests interaction between an AD-protein and the protein shown in column header (as BD-fusion). Each lower row tests the reciprocal interaction. All protein pairs appear twice in the matrix, except homodimerization (black **X**). Protein names in rows appear in same order as in **Supplemental Figure S1B**. Protein names in column header appear in different order (order in which yeast transformants were spotted during the experiment).

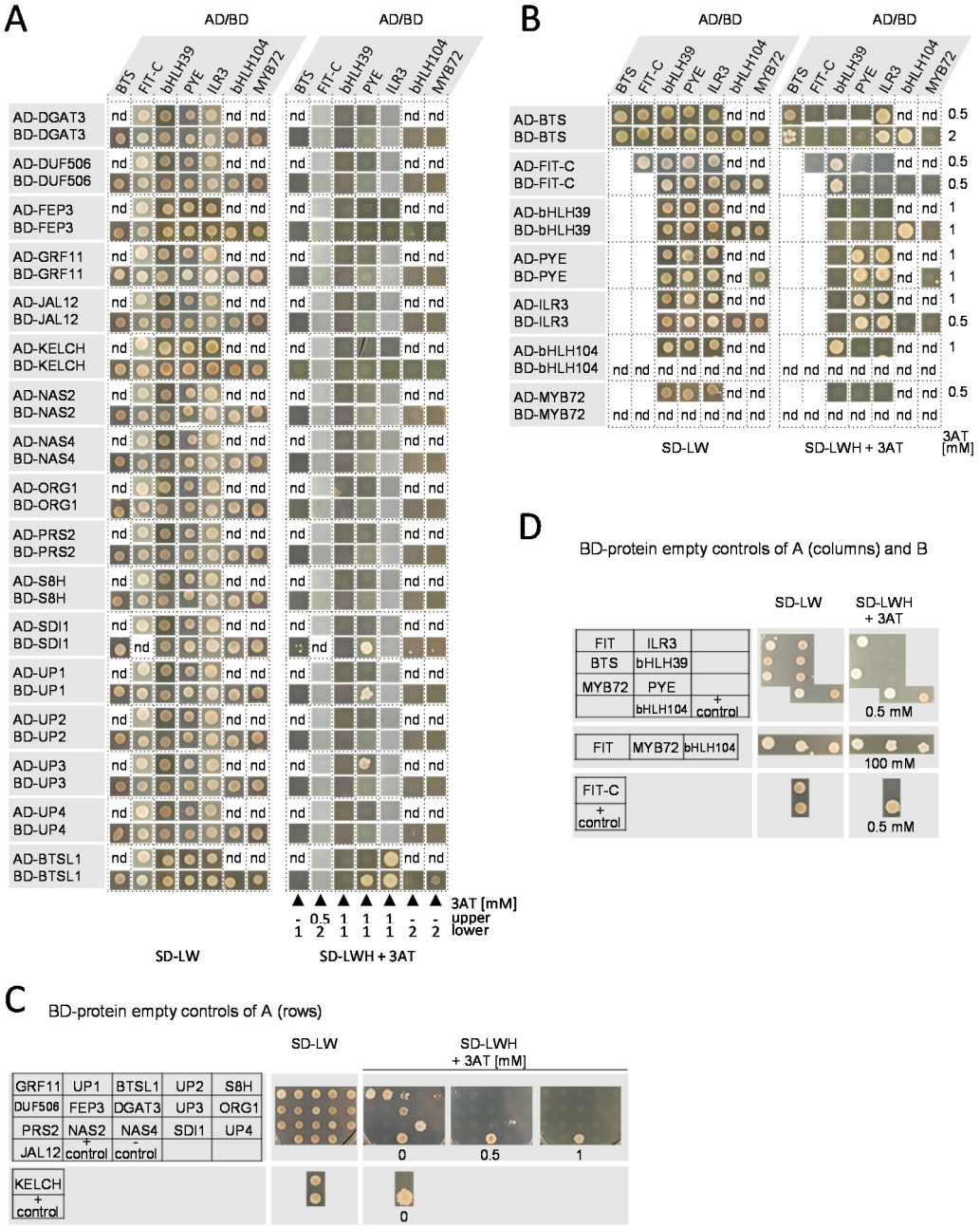


Figure S3. Original Y2H screen images (part 2).
A: Targeted Y2H screen between the 17 candidates shown in part 1 (rows) (**Supplemental Figure S2**) and the 7 remaining candidates (columns) (6 TFs and BTS).
B: Targeted Y2H screen between 6 TFs and BTS in all pair-wise combinations, including homodimerization. Except BTS and FIT-C, all protein pairs appear twice in the matrix. Interaction between AD-bHLH104/BD-bHLH39 could not be verified (see **Supplemental Figure S5B**).
C: Determining the 3AT concentration necessary to suppress background activity induced by the 17 non-TF BD-fusions (= BD-protein/empty negative controls). Positions of protein names in the table correspond to yeast spot positions in the images. Negative controls correspond to the 17 BD-proteins in A (rows). Positive control (+ control): CIPK23/cAKT1 (Xu et al., 2006); negative control of the Y2H system (- control): empty/empty.
D: Determining necessary 3AT concentration for TFs and BTS BD-fusions. Negative controls correspond to the 7 BD-proteins in A (columns) and in B (rows and columns), and of full-length FIT (see also Supplemental Figure S5). Control as described in C.
A, B: Yeast containing both AD- and BD-plasmids were obtained by mating of single transformants and spotted as described in **Supplemental Figure S2**. Rows should be read in pairs: Each upper row tests interaction between an AD-protein and the protein shown in column header (as BD-fusion). Each lower row tests the reciprocal interaction. FIT-C: C-terminal part of FIT. nd: no data. BD-bHLH104 and BD-MYB could not be tested because the proteins auto-activated the Y2H system.

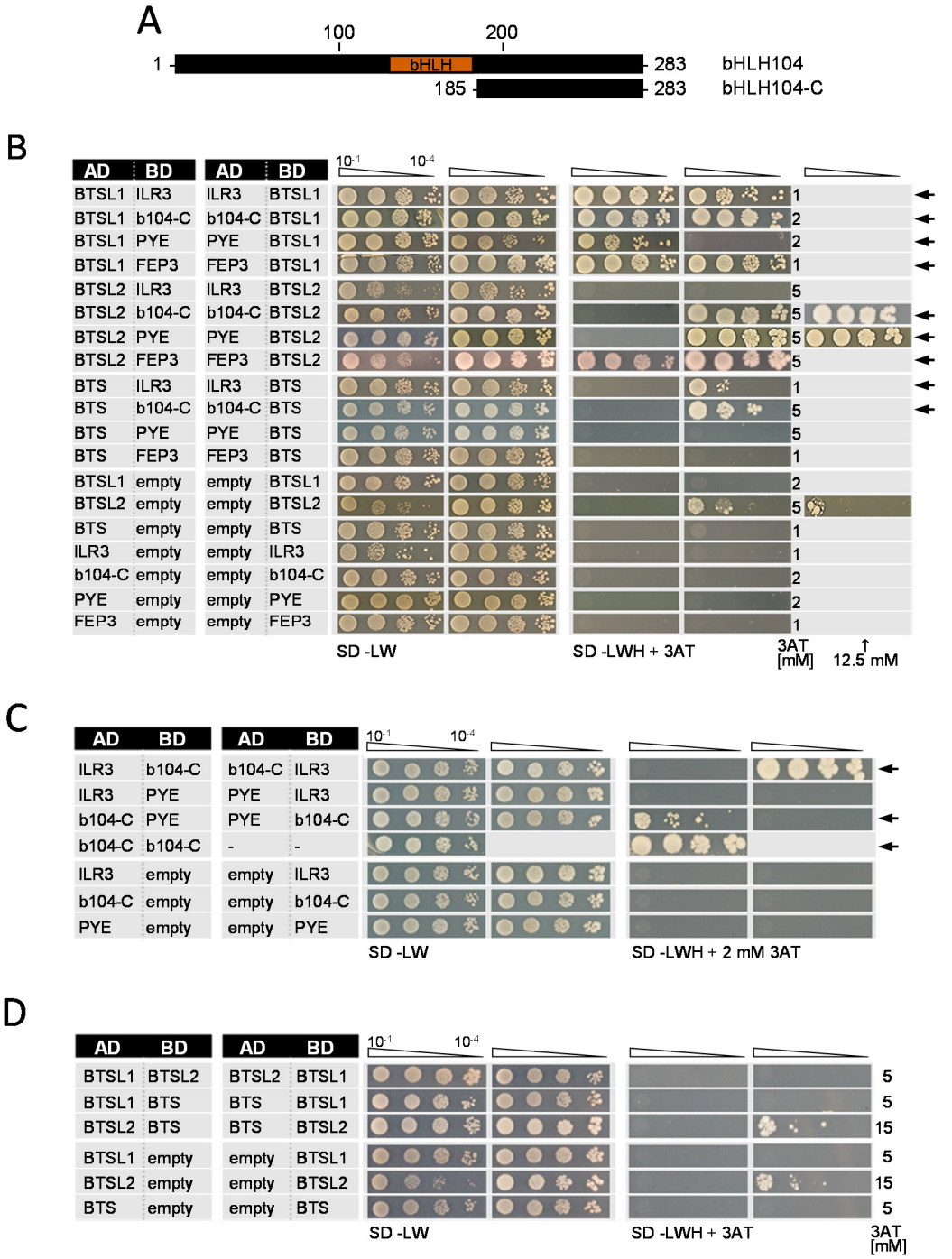


Figure S4. Verified "BTS(L) interactome" protein-protein interactions and non-interactions.
A: Schematic representation of full-length bHLH104 and its C-terminal part lacking the DNA binding domain (bHLH104-C). Because BD-bHLH104 was auto-activating the Y2H system (see **Supplemental Figure S3D**) bHLH104-C was used in all Y2H except the Y2H screen. Protein domain structure based on information deposited on UniProt (www.uniprot.org).
B-D: Reciprocal targeted Y2H protein interaction assays between different combinations of "BTS(L) interactome" proteins. Yeast co-transformed with the AD and BD combinations were spotted in 10-fold dilution series (A_{600}=10^{-1}-10^{-4}) on SD-LW (transformation control) and SD-LWH supplemented with different concentrations of 3AT (selection for protein intraction). Negative controls: empty AD/BD-proteins and empty BD/AD proteins. Arrows indicate interaction.
B: Interactions of BTSL1, BTSL2, BTS with ILR3, bHLH104-C (b104-C), PYE, FEP3. BD-BTSL2 was spotted additionally on higher 3AT concentrations to better distinguish interaction and auto-activation.
C: Interactions of ILR3, bHLH104-C, PYE amongst each other and homodimerization of bHLH104-C.
D: Interactions of BTSL1, BTSL2, BTS amongst each other.

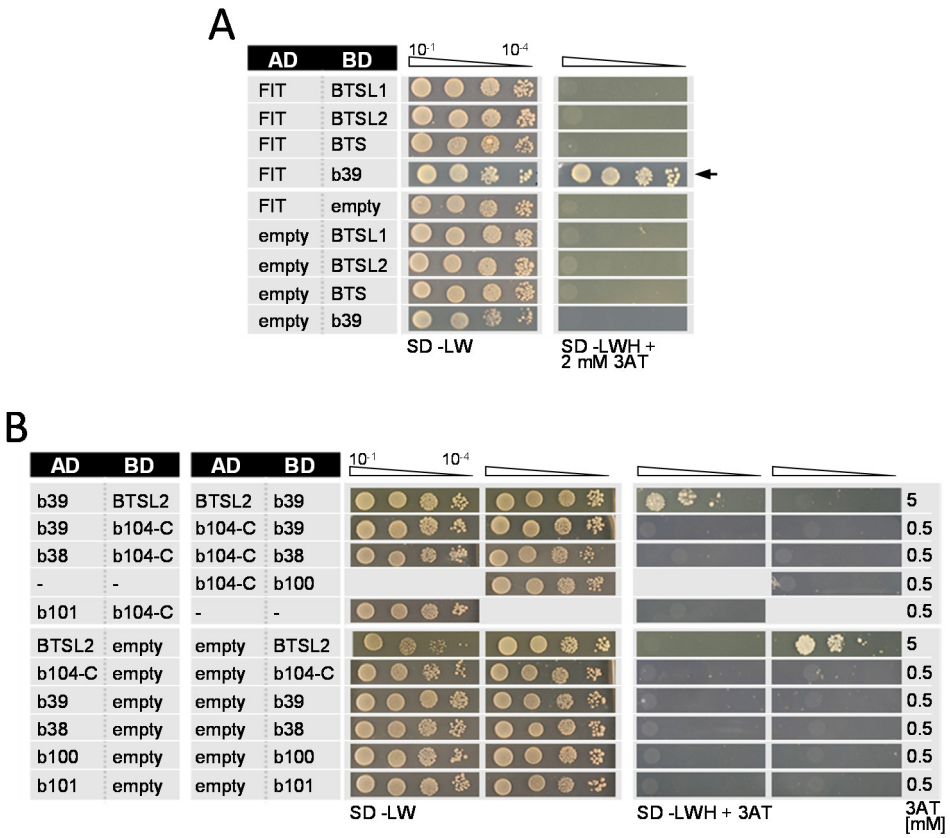


Figure S5. Verified non-interactions between BTS, BTSL1, BTSL2 and different bHLH TFs.
A: Full-length AD-FIT does not interact with BD-BTSL1, BD-BTSL2, BD-BTS in yeast. The known interaction between FIT and b39 (bHLH39) served as positive control (arrow). BD-FIT was not tested because it auto-activates the Y2H system (see **Supplemental Figure S3D**).
B: BTSL2 does not interact with b39. Tested for sake of completeness because BTSL2 was missing in the Y2H screen. b104-C (bHLH104-C) does not interact with b38, b39, b100, b101. Combinations were tested to verify the expected non-interactions (reported as negative in (Zhang et al., 2015)).
A, B: Yeast co-transformed with the AD and BD combinations were spotted as described in **Supplemental Figure S4.**

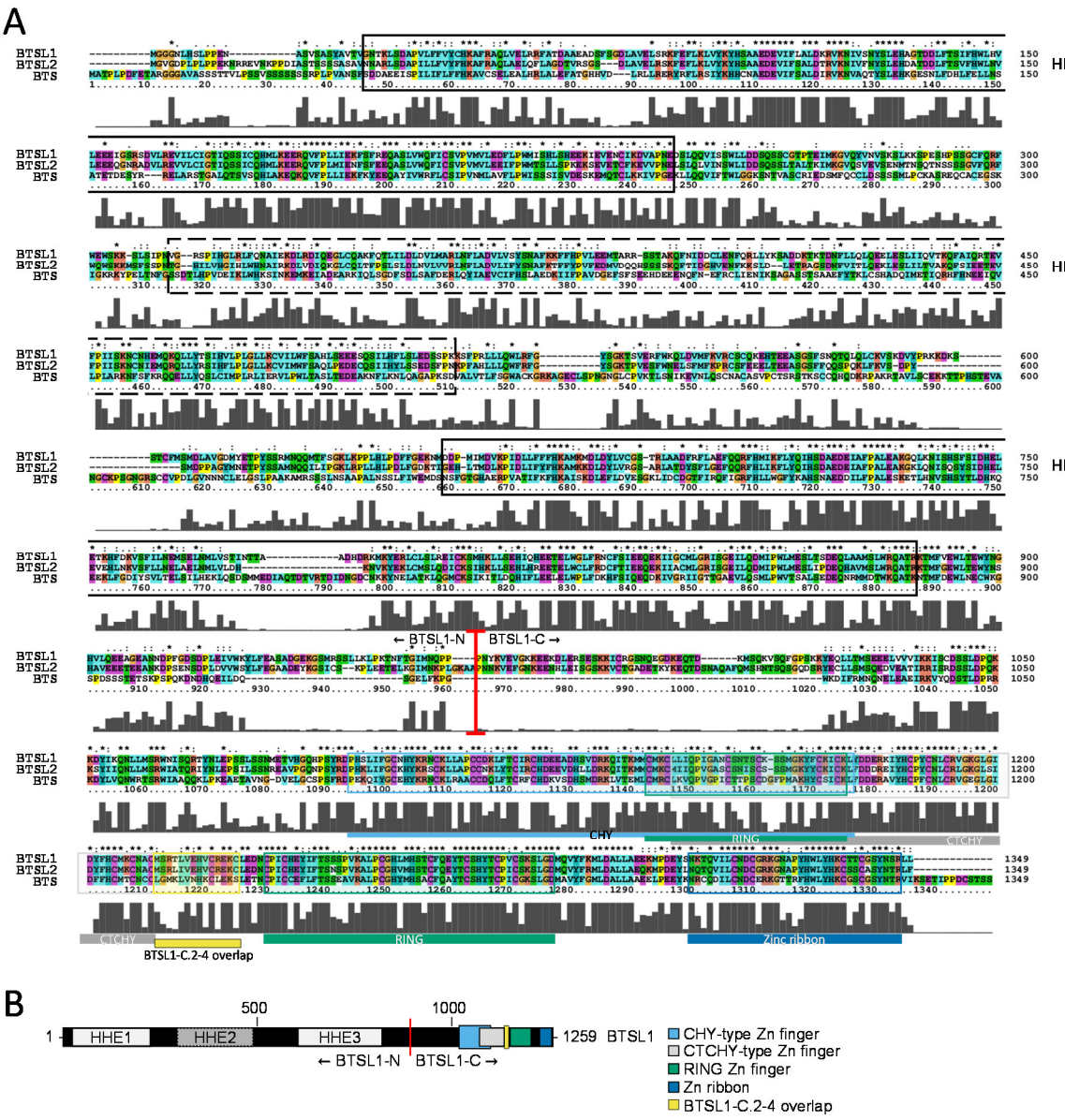


Figure S6. BTSL1/BTSL2/BTS multiple sequence alignment, protein domains and regions relevant in this study.
A: Multiple sequence alignment of BTSL1, BTSL2 and BTS protein sequences. Predicted HHE domains (Rodríguez-Celma et al., 2019) and C-terminal domains (InterPro; (Mitchell et al., 2019)) are indicated. Putative interacting region in BTSL1 indicated in yellow (BTSL1-C.2-4 overlap) (see also **Figure 3**).
B: Schematic representation of BTSL1 protein structure, including all features shown in A.

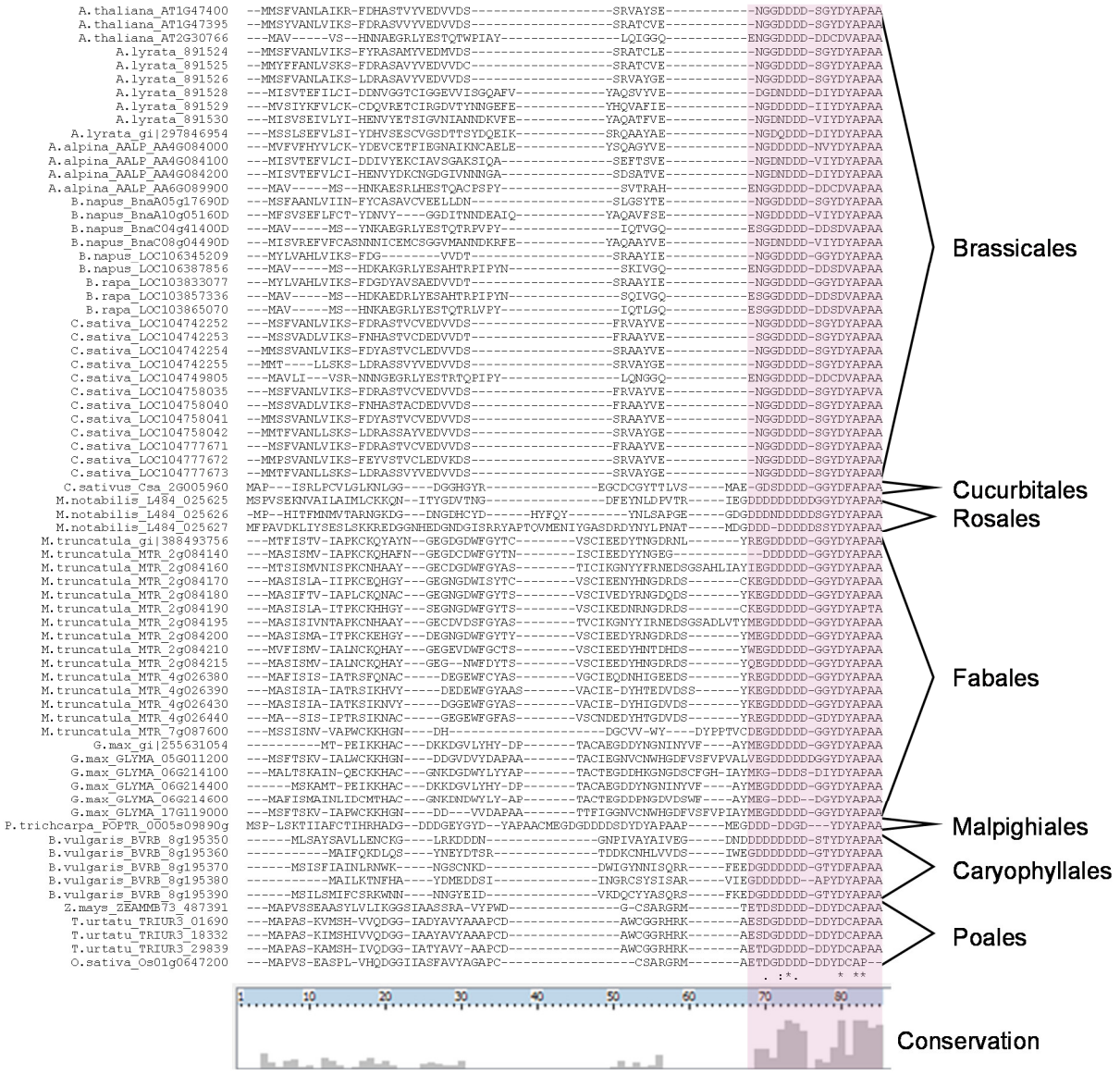


Figure S7. FEP3 conserved domain.
Multiple sequence alignment of FEP3 (AT1G47400) protein sequence and BLAST search hits of selected angiosperm species. Example orders within the angiosperm phylogeny are shown. Corresponds to **Figure 4**.

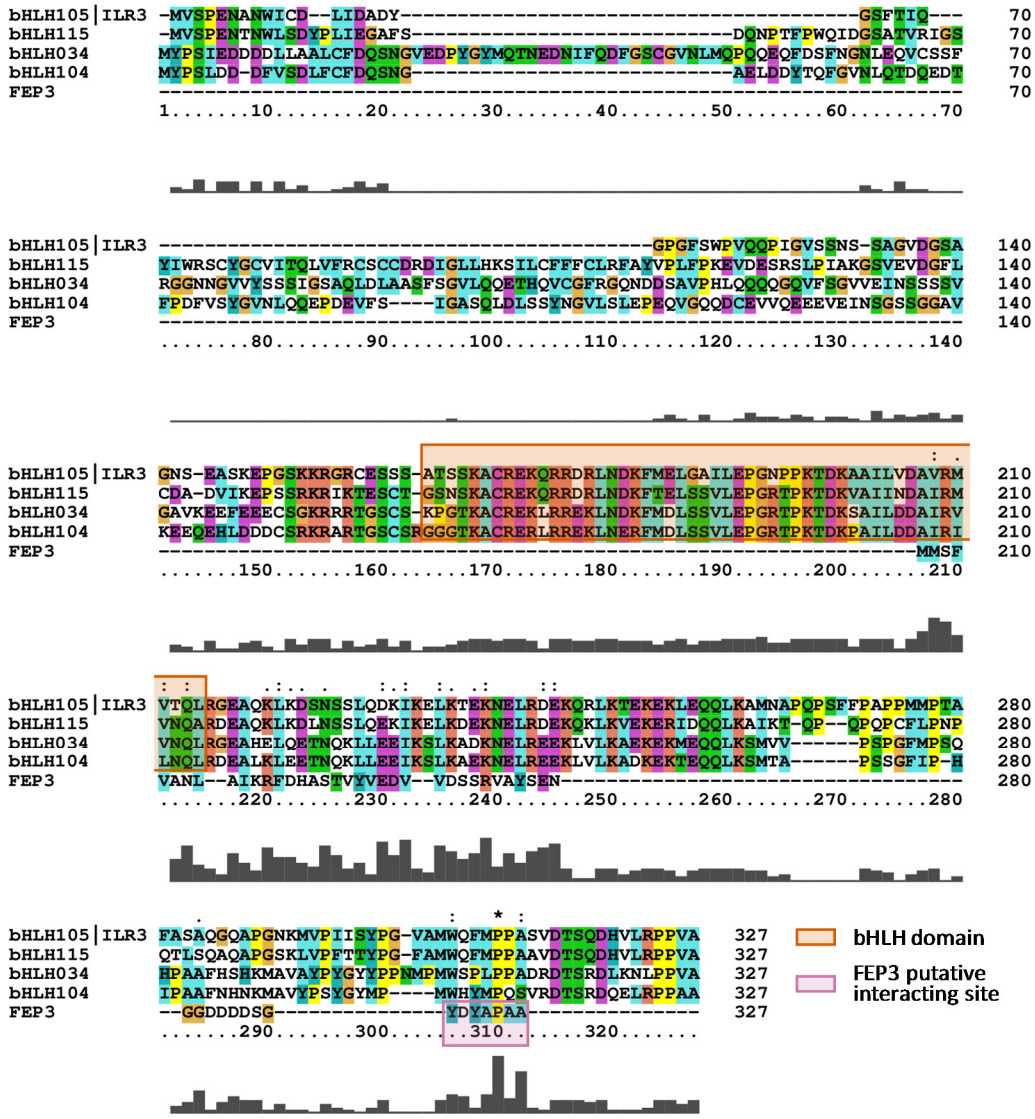


Figure S8. Multiple sequence alignment of IVc bHLHs and FEP3 to identify potential sequence similarities. Multiple sequence alignment of IVc bHLHs (ILR3/bHLH115/34/104) and FEP3 protein sequences. Indicated regions: bHLH domain (orange) and the seven C-terminal amino acids of FEP3 needed for protein-protein interaction (see **Figure 4**) (pink). bHLH domain annotation based on information deposited on UniProt (www.uniprot.org).

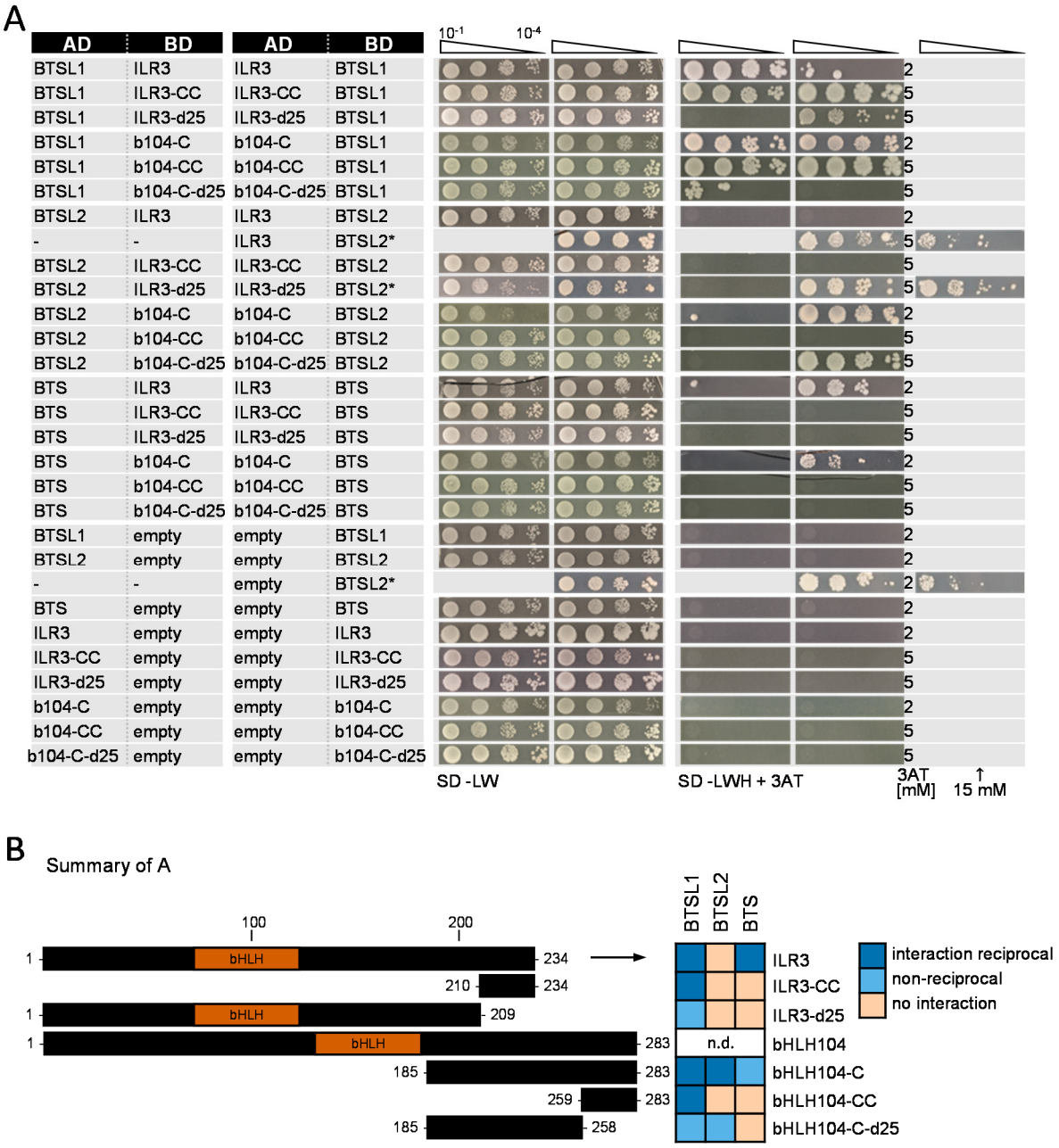


Figure S9. The last 25 aa in ILR3 and bHLH104 are not crucial for interaction with BTSL1 or BTSL2.
A: Reciprocal targeted Y2H protein interaction assay testing different truncated versions of ILR3, bHLH104. Full-length ILR3 or b104-C (bHLH104-C) (=positive controls), truncated versions consisting of the last 25 aa (ILR3-CC, b104-CC) and the respective counterparts, lacking the last 25 aa (ILR3-d25, b104-C-d25) were co-transformed with BTSL1, BTSL2, BTS into yeast. Yeast containing the AD and BD combinations were spotted as described in **Supplemental Figure S4.** *BD-BTSL2 was occasionally auto-activating the system, and spotted additionally on higher 3AT concentrations to better distinguish interaction and auto-activation.
B: Schematic representations of full-length ILR3 and truncated versions, as well as bHLH104-C and truncated versions tested in A. Full-length bHLH104 not tested in A, shown to easily relate the bHLH104 fragments. Heatmap summarizes results of A.

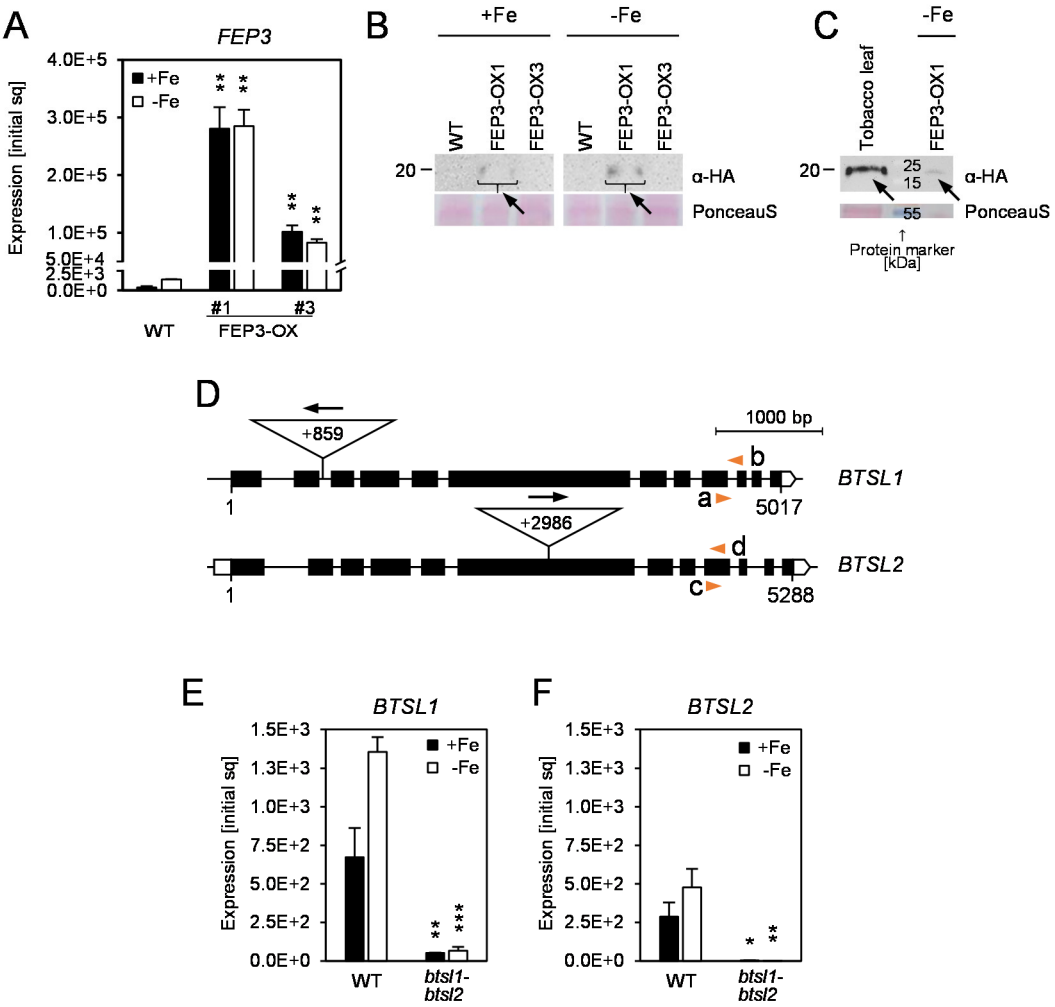


Figure S10. Validation of FEP3-OX and *btsl1-btsl2* mutant lines.
A: *FEP3* gene expression (RT-qPCR) in seedlings of WT and two FEP3-OX lines (#1, #3, WT background) generated in this study. Plants were grown in the 6 d system with sufficient (+Fe, black bars) or deficient (-Fe, white bars) Fe supply.
B, C: Anti-HA immunodetection of HA_3-FEP3 fusion protein (indicated by arrows) in seedlings of FEP3-OX (#1, #3). Seedlings were grown as in A. Tobacco leaf transiently transformed with the HA_3-FEP3 construct served as positive control. WT served as negative control. Loading control: PonceauS staining of the membrane. Molecular weight of the protein (in kDa) is indicated. Because bands in (**B**) are hardly recognizable (protein accumulated in SDS gel pocket corners), a second membrane with positive control is shown (**C**).
D: Schematic representation of *BTSL1* and *BTSL2* genes and mutant alleles of the *btsl1-btsl2* line. Black boxes indicate exons, white boxes indicate 5' and 3' untranslated regions. Positions of T-DNA insertions (triangles) and primers used for RT-qPCR (orange arrowheads; a, b, c, d) are indicated. Black arrows indicate orientation of left border primers used for genotyping (see **Supplemental Table S1**).
E, F: *BTSL1* (**E**) and *BTSL2* (**F**) gene expression (RT-qPCR) in WT and *btsl1-btsl2* mutant roots. Plants were grown in the 14+3 d system with +Fe or -Fe supply.
A, E, F: Expression [initial sq]=Absolute normalized expression [initial starting quantity]. Data are represented as mean and SD. Asterisks indicate statistically significant differences to the WT of the same growth condition (n=3) (Student's T-Test, **p*<0.05, ***p*<0.01, ****p*<0.001).

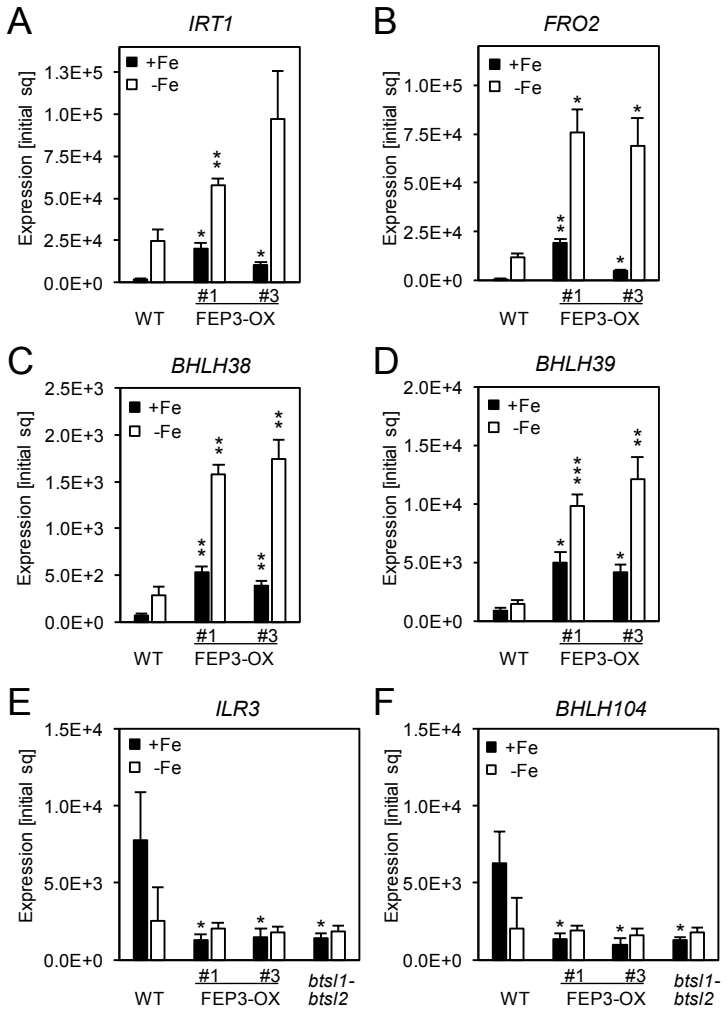


Figure S11. Regulation of selected Fe deficiency response genes in whole seedlings of FEP3-OX and *btsl1-btsl2*.

A-D: Expression analysis (RT-qPCR) of Fe deficiency response genes in seedlings of two FEP3-OX lines (#1, #3) in comparison to WT. A: IRT1, B: FRO2, C: BHLH38, D: BHLH39.

E, F: Expression analysis of ILR3 (E) and BHLH104 (F) in seedlings of FEP-OX (#1, #3) and *btsl1-btsl2* mutant lines.

A-F: Plants were grown in the 6 d system with sufficient (+Fe, black bars) or deficient (-Fe, white bars) Fe supply. Expression [initial sq]=Absolute normalized expression [initial starting quantity]. Data are represented as mean and SD. Asterisks indicate statistically significant differences to the WT of the same growth condition (n=3) (Student's T-Test, *p<0.05, **p<0.01, ***p<0.001). Data corresponds to **Figure 5**.

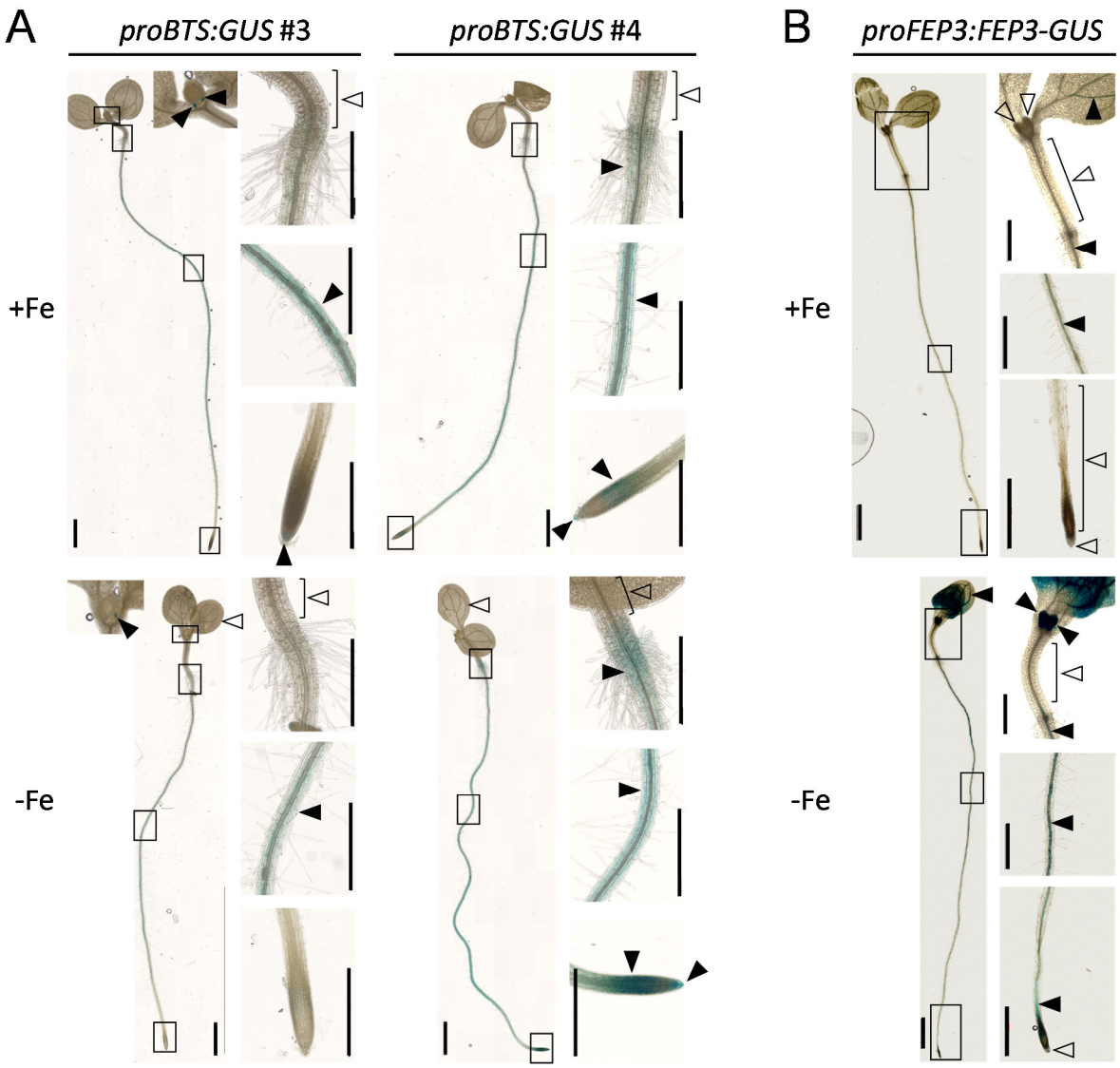


Figure S12. *ProBTS* activity in two additional Arabidopsis lines, and FEP3-GUS protein localization in Arabidopsis seedlings.
A: *BTS*-promoter-driven GUS reporter activity in two additional transgenic Arabidopsis lines expressing *proBTS:GUS* in WT background (correspond to **Figure 6F, L**). Plants were grown in the 6 d system under sufficient (+Fe, **top**) and deficient (-Fe, **bottom**) Fe supply. Reporter gene expression was visualized by a chemical reaction (GUS) resulting in blue staining of the respective tissues. Plants were imaged with brightfield microscopy. Rectangles in whole-seedling images indicate positions of the enlarged images. Filled arrowheads indicate selected areas of GUS staining, non-filled arrowheads indicate selected non-stained areas. Vascular tissue is indicated by the arrowhead pointing directly at it, non-vascular tissue is indicated by the arrowhead pointing more to an outside area. Scale bars of whole seedling images: 1 mm; of magnifications: 0.5 mm.
B: GUS activity indicating FEP3-GUS fusion protein localization. Transgenic Arabidopsis plants expressing *proFEP3:FEP3-GUS* in WT background were grown and imaged as described in A. Images are presented as described in A.

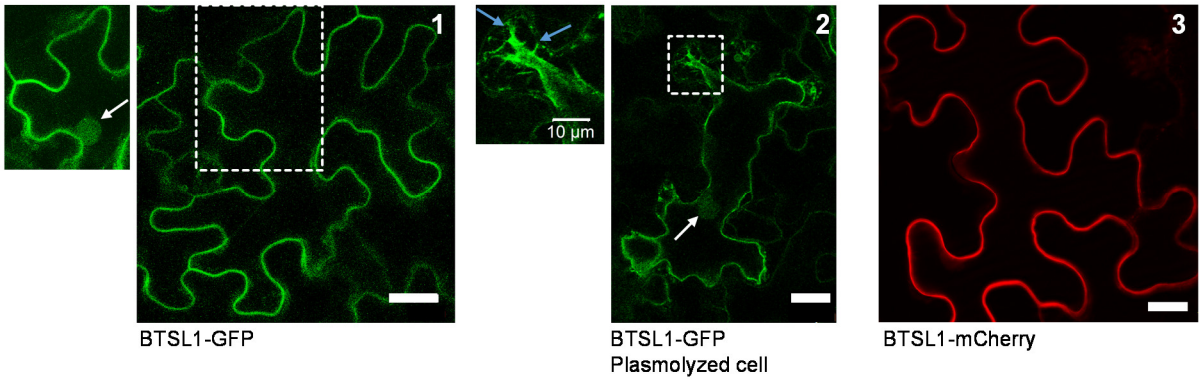


Figure S13. Peripheral localization of BTSL1 in tobacco leaf epidermis cells (additional images).
Images correspond to **Figure 7A (1)** and show localization of BTSL1 with as C-terminal GFP **(1, 2)** and mCherry **(3)** fusions. Magnification of 1 shows a different layer of the Z stack to better show the nucleus. To show that BTSL1-GFP localizes to the plasma membrane **(1)**, cells were plasmolyzed (blue arrows indicate Hechtian strands) **(2)**. Arrows indicate nuclei. (1, 2) were imaged by laser-scanning confocal fluorescence microscopy. (3) was imaged with a fluorescence microscope and an ApoTome for enhanced resolution. Scale bars: 20 µm.

Supplemental Table 1. Primers used in this study.
fw = forward, rev = reverse, s = stop, ns = no stop, d = delta a.k.a. deletion, -C = C-term, -CC = 25 C-terminal amino acids, -N = N-term, (att)B1/B2/B3/B4 = Gateway B attachment sites for BP reaction.

Primer name	Sequence 5‘ → 3‘	Purpose	Origin
Amplification of full-length CDS or deletion constructs for Y2H, BiFC, (co-) localization and OX lines			
BTSL1_B1 fw	GGGGACAAGTTTGTACAAAAAAGCAGGCTT CATGGGAGGCGGAAATCTTC	cloning of BTSL1, BTSL1-N, BTSL1-dRH, BTSL1-6G (attB1)	This study
BTSL1s_B2 rev	GGGGACCACTTTGTACAAGAAAGCTGGGTC CTAAAGGAGCCTTGAGTTGTAG	cloning of BTSL1, BTSL1-C, BTSL1-C.4, BTSL1-dRH, BTSL1-6G (attB2)	This study
BTSL1ns_B2 rev	GGGGACCACTTTGTACAAGAAAGCTGGGTC AAGGAGCCTTGAGTTGTAGGA	cloning of BTSL1, BTSL1-C, no stop codon (attB2)	This study
BTSL1-Ns_B2 rev	GGGGACCACTTTGTACAAGAAAGCTGGGTC CTAAGGAGGCTGATTCATAATACCTG	cloning of BTSL1-N (attB2)	This study
BTSL1-C_B1 fw	GGGGACAAGTTTGTACAAAAAAGCAGGCTT CATGCCTAATTACAAGGTTGAAGTTGGC	cloning of BTSL1-C, BTSL1-C.1, BTSL1-C.2 (attB1)	This study
BTSL1_B3 fw	GGGGACAACTTTGTATAATAAAGTTGTAATG GGAGGCGGAAATCTTC	cloning of BTSL1 (attB3)	This study
BTSL1s_B4 rev	GGGGACAACTTTGTATAGAAAAGTTGGGTG CTAAAGGAGCCTTGAGTTGTAG	cloning of BTSL1-C (attB4)	This study
BTSL1-C.1s_B2 rev	GGGGACCACTTTGTACAAGAAAGCTGGGTC CTAGTTGCAGTAAGGACAGTGATAA	cloning of BTSL1-C.1 (attB2)	This study
BTSL1-C.2s_B2 rev	GGGGACCACTTTGTACAAGAAAGCTGGGTC CTAGCACTTTTCCCTGCAAACAT	cloning of BTSL1-C.2 (attB2)	This study
BTSL-C.3_B1 fw	GGGGACAAGTTTGTACAAAAAAGCAGGCTT CATGGAAGAAGCTGATCACTCGGT	cloning of BTSL1-C.3 (attB1)	This study
BTSL1-C.3s_B2 rev	GGGGACCACTTTGTACAAGAAAGCTGGGTC CTAATCAGGCATCTTCTCTTCTGC	cloning of BTSL1-C.3 (attB2)	This study
BTSL1-C.4_B1 fw	GGGGACAAGTTTGTACAAAAAAGCAGGCTT CATGTCACGTACCTTAGTAGAGC	cloning of BTSL1-C.4 (attB1)	This study
BTSL1_dRH rev	CTTTTCCCTGCAAACTGACATACAAGCATTG CAT	cloning BTSL1-dRH for overlap extension PCR (N-term part)	This study
BTSL1_dRH fw	AATGCTTGTATGTCAGTTTGCAGGGAAAAG TG	cloning BTSL1-dRH for overlap extension PCR (C-term part)	This study
BTSL1_dRH-G rev	GCAAACGCCTCCGCCTCCGCCTCCTGACAT ACAAGCATTGCAT	cloning BTSL1-6G for overlap extension PCR (C-term part)	This study
BTSL1_dRH-G fw	ATGTCAGGAGGCGGAGGCGGAGGCGTTTG CAGGGAAAAGTG	cloning BTSL1-6G for overlap extension PCR (C-term part)	This study
BTSL2_B1 fw	GGGGACAAGTTTGTACAAAAAAGCAGGCTT CATGGGAGTCGGAGATCCT	cloning of BTSL2 (attB1)	This study
BTSL2s_B2 rev	GGGGACCACTTTGTACAAGAAAGCTGGGTC CTAGAAAAGTCTGGTGTTGTAGG	cloning of BTSL2 , BTSL2-C (attB2)	This study
BTSL2-C_B3 fw	GGGGACAACTTTGTATAATAAAGTTGTAATG GGAGTCGGAGATCCT	cloning of BTSL2-C (attB1)	This study
BTS_B1 fw	GGGGACAAGTTTGTACAAAAAAGCAGGCTT CATGGCGACGCCGTTACCA	cloning of BTS (attB1)	This study
BTSs_B2 rev	GGGGACCACTTTGTACAAGAAAGCTGGGTC TCAGGATGAGGTTGAGCAGT	cloning of BTS (attB2)	This study
BTSs_B4 rev	GGGGACAACTTTGTATAGAAAAGTTGGGTG TCAGGATGAGGTTGAGCAGT	cloning of BTS (attB4)	This study
FEP3_B1 fw	GGGGACAAGTTTGTACAAAAAAGCAGGCTT CATGATGTCTTTTGTCGCAAAC	cloning of FEP3, FEP3-N, FEP3-d7 (attB1)	This study
FEP3s_B2 rev	GGGGACCACTTTGTACAAGAAAGCTGGGTC TCACGCAGCAGGAGCATA	cloning of FEP3, FEP3-C (attB2)	This study
FEP3ns_B2 rev	GGGGACCACTTTGTACAAGAAAGCTGGGTC CGCAGCAGGAGCATAATC	cloning of FEP3, proFEP3-FEP3, no stop codon (attB2)	This study
FEP3-Ns_B2 rev	GGGGACCACTTTGTACAAGAAAGCTGGGTC TCATACCACATCTTCAACATACACG	cloning of FEP3-N (attB2)	This study

FEP3-C_B1 fw	GGGGACAAGTTTGTACAAAAAAGCAGGCTT CATGGATAGTTCTCGAGTGGCATATAG	cloning of FEP3-C (attB1)	This study
FEP3-d7s_B2 rev	GGGGACCACTTTGTACAAGAAAGCTGGGTC TCATCAGCCACTGTCATCGTC	cloning of FEP3-d7 (attB2)	This study
ILR3_B1 fw	GGGGACAAGTTTGTACAAAAAAGCAGGCTT CATGGTGTCACCCGAAAACG	cloning of ILR3, ILR3-d25 (attB1)	This study
ILR3s_B2 rev	GGGGACCACTTTGTACAAGAAAGCTGGGTC TTAAGCAACAGGAGGACGAAG	cloning of ILR3 (attB2)	This study
ILR3ns_B2 rev	GGGGACCACTTTGTACAAGAAAGCTGGGTC AGCAACAGGAGGACGAAG	cloning of ILR3, no stop codon (attB2)	This study
ILR3_B3 fw	GGGGACAACTTTGTATAATAAAGTTGTAATG GTGTCACCCGAAAACG	cloning of ILR3 (attB3)	This study
ILR3s_B4 rev	GGGGACAACTTTGTATAGAAAAGTTGGGTG TTAAGCAACAGGAGGACGAAG	cloning of ILR3 (attB4)	This study
ILR3-d25s_B2 rev	GGGGACCACTTTGTACAAGAAAGCTGGGTC TTATCCTGGGTAACTGATGAT	cloning of ILR3-d25 (attB2)	This study
ILR3-CC_B1 fw	GGGGACAAGTTTGTACAAAAAAGCAGGCTT CATGGTTGCCATGTGG	cloning of ILR3-CC (attB1)	This study
bHLH104_B1 fw	GGGGACAAGTTTGTACAAAAAAGCAGGCTT CATGTATCCTTCTCTCGACGA	cloning of bHLH104 (attB1)	This study
bHLH104s_B2 rev	GGGGACCACTTTGTACAAGAAAGCTGGGTC TTAAGCAGCAGGAGGCCTG	cloning of bHLH104, bHLH104-C, bHLH104-CC (attB1)	This study
bHLH104-C_B1 fw	GGGGACAAGTTTGTACAAAAAAGCAGGCTT CATGGCTCT	cloning of bHLH104-C, bHLH104-C-d25 (attB1)	This study
bHLH104-d25s_B2 rev	GGGGACCACTTTGTACAAGAAAGCTGGGTC TTAGTAACCGTAACTTGGATAAAC	cloning of bHLH104-C-d25 (attB2)	This study
bHLH104-CC_B1 fw	GGGGACAAGTTTGTACAAAAAAGCAGGCTT CATGATGCCAATGTGGC	cloning of bHLH104-CC (attB1)	This study
PYE_B1 fw	GGGGACAAGTTTGTACAAAAAAGCAGGCTT CATGGTATCGAAAACTCCTTC	cloning of PYE (attB1)	This study
PYEs_B2 rev	GGGGACCACTTTGTACAAGAAAGCTGGGTC TCATTCACTGGCTTTCAGCC	cloning of PYE (attB2)	This study
PYEns B2 rev	GGGGACCACTTTGTACAAGAAAGCTGGGTC TTCACTGGCTTTCAGCC	cloning of PYE, no stop codon (attB2)	This study
PYE_B3 fw	GGGGACAACTTTGTATAATAAAGTTGTAATG GTATCGAAAACTCCTTC	cloning of PYE (attB3)	This study
PYEs_B4 rev	GGGGACAACTTTGTATAGAAAAGTTGGGTG TCATTCACTGGCTTTCAGCC	cloning of PYE (attB4)	This study
FIT_B1 fw	GGGGACAAGTTTGTACAAAAAAGCAGGCTT CATGGAAGGAAGAGTCAACG	cloning of FIT (attB1)	This study
FITs_B2 rev	GGGGACCACTTTGTACAAGAAAGCTGGGTC TCAAGTAAATGACTTGATG	cloning of FIT, FIT-C (attB2)	This study
FIT-C_B1 rev	GGGGACAAGTTTGTACAAAAAAGCAGGCTT AACTCAACCTTTTCGCGGTATC	cloning of FIT-C (attB1)	(Gratz et al., 2019)
FIT_B3 fw	GGGGACAACTTTGTATAATAAAGTTGTAATG GAAGGAAGAGTCAACG	cloning of FIT (attB3)	This study
bHLH38_B1 fw	GGGGACAAGTTTGTACAAAAAAGCAGGCTT CATGTGTGCATTAGTCCCTTCATTT	cloning of bHLH38 (attB1)	This study
bHLH38_B2 rev	GGGGACCACTTTGTACAAGAAAGCTGGGTC CTAGTTAAACGAGTTTTCACATT	cloning of bHLH38 (attB2)	This study
bHLH39_B1 fw	GGGGACAAGTTTGTACAAAAAAGCAGGCTT CATGTGTGCATTAGTACCTC	cloning of bHLH39 (attB1)	This study
bHLH39_B2 rev	GGGGACCACTTTGTACAAGAAAGCTGGGTC TCATATATATGAGTTTCCAC	cloning of bHLH39 (attB2)	This study
bHLH100_B1 fw	GGGGACAAGTTTGTACAAAAAAGCAGGCTT CATGTGTGCACTTGTCCCTCCATT	cloning of bHLH100 (attB1)	This study

bHLH100_B2 rev	GGGGACCACTTTGTACAAGAAAGCTGGGTC TCATGTAAACGAGTGTCCACATT	cloning of bHLH100 (attB2)	This study
bHLH101_B1 fw	GGGGACAAGTTTGTACAAAAAAGCAGGCTT CATGGAGTATCCATGGCTGCAGTC	cloning of bHLH101 (attB1)	This study
bHLH101_B2 rev	GGGGACCACTTTGTACAAGAAAGCTGGGTC TTATGATTGGCGTAATCCCAAGA	cloning of bHLH101 (attB2)	This study
MYB72_B1 fw	GGGGACAAGTTTGTACAAAAAAGCAGGCTT CATGGGGAAAGGAAGAGCAC	cloning of MYB72 (attB1)	This study
MYB72_B2 rev	GGGGACCACTTTGTACAAGAAAGCTGGGTC TCATAGACATACTTCTCCGA	cloning of MYB72 (attB2)	This study
DGAT3_B1 fw	GGGGACAAGTTTGTACAAAAAAGCAGGCTT CATGGAGAAGGAGAAGAAGGC	cloning of DGAT3 (attB1)	This study
DGAT3s_B2 rev	GGGGACCACTTTGTACAAGAAAGCTGGGTC TCAATATGAGACAGAACCGAGT	cloning of DGAT3 (attB2)	This study
DUF506_B1 fw	GGGGACAAGTTTGTACAAAAAAGCAGGCTT CATGGTAGAGATAGGAGGACGA	cloning of DUF506 (attB1)	This study
DUF506s_B2 rev	GGGGACCACTTTGTACAAGAAAGCTGGGTC CTAAAATATTTGCAACCCCACTT	cloning of DUF506 (attB2)	This study
GRF11_B1 fw	GGGGACAAGTTTGTACAAAAAAGCAGGCTT CATGGAGAACGAGAGAGCG	cloning of GRF11 (attB1)	This study
GRF11s_B2 rev	GGGGACCACTTTGTACAAGAAAGCTGGGTC TTAGATTTTGTTTACCTCATCTTG	cloning of GRF11 (attB2)	This study
JAL12_B1 fw	GGGGACAAGTTTGTACAAAAAAGCAGGCTT CATGTCTCAAGATTCGAATGC	cloning of JAL12 (attB1)	This study
JAL12s_B2 rev	GGGGACCACTTTGTACAAGAAAGCTGGGTC TCACCAGCTTGTGTGTATG	cloning of JAL12 (attB2)	This study
KELCH_B1 fw	GGGGACAAGTTTGTACAAAAAAGCAGGCTT CATGGCAGCGACTCCAATG	cloning of KELCH (attB1)	This study
KELCHs_B2 rev	GGGGACCACTTTGTACAAGAAAGCTGGGTC TCACTTGAGTAGAGTGTTGG	cloning of KELCH (attB2)	This study
NAS2_B1 fw	GGGGACAAGTTTGTACAAAAAAGCAGGCTT CATGGCTTGCGAAAACAACCT	cloning of NAS2 (attB1)	This study
NAS2s_B2 rev	GGGGACCACTTTGTACAAGAAAGCTGGGTC TTACTCGATGGCACTATACTCCT	cloning of NAS2 (attB2)	This study
NAS4_B1 fw	GGGGACAAGTTTGTACAAAAAAGCAGGCTT CATGGGTTATTGCCAAGACGA	cloning of NAS4 (attB1)	This study
NAS4s_B2 rev	GGGGACCACTTTGTACAAGAAAGCTGGGTC CTAGGTAAGTTGTTCTTCATTAGCA	cloning of NAS4 (attB2)	This study
ORG1_B1 fw	GGGGACAAGTTTGTACAAAAAAGCAGGCTT CATGGCACTTTGTGGTGTTTG	cloning of ORG1 (attB1)	This study
ORG1s_B2 rev	GGGGACCACTTTGTACAAGAAAGCTGGGTC CTACATAGACTTATGGTCCAAGC	cloning of ORG1 (attB2)	This study
PRS2_B1 fw	GGGGACAAGTTTGTACAAAAAAGCAGGCTT CATGGCGTCGTTGGCTCT	cloning of PRS2 (attB1)	This study
PRS2s_B2 rev	GGGGACCACTTTGTACAAGAAAGCTGGGTC TCAAAGGAAAATACTACTAACGGAG	cloning of PRS2 (attB2)	This study
S8H_B1 fw	GGGGACAAGTTTGTACAAAAAAGCAGGCTT CATGGGTATCAATTTCGAGGAC	cloning of S8H (attB1)	This study
S8Hs_B2 rev	GGGGACCACTTTGTACAAGAAAGCTGGGTC TCACTCGGCACGTGC	cloning of S8H (attB2)	This study
SDI1_B1 fw	GGGGACAAGTTTGTACAAAAAAGCAGGCTT CATGGAGAGAAGCTTGAAGAA	cloning of SDI1 (attB1)	This study
SDI1s_B2 rev	GGGGACCACTTTGTACAAGAAAGCTGGGTC CTAGCAAACTAATGTATTTCTAAAAG	cloning of SDI1 (attB2)	This study
UIP1_B1 fw	GGGGACAAGTTTGTACAAAAAAGCAGGCTT CATGGCGACTTCTACCTTCT	cloning of UIP1 (attB1)	This study

UIP1s_B2 rev	GGGGACCACTTTGTACAAGAAAGCTGGGTC TCAAGAGAACCAATTAACGAG	cloning of UIP1 (attB2)	This study
UIP2_B1 fw	GGGGACAAGTTTGTACAAAAAAGCAGGCTT CATGTATCAAGATCGTCAAGGC	cloning of UIP2 (attB1)	This study
UIP2s_B2 rev	GGGGACCACTTTGTACAAGAAAGCTGGGTC TCAGTCATCGGAACCATCA	cloning of UIP2 (attB2)	This study
UIP3_B1 fw	GGGGACAAGTTTGTACAAAAAAGCAGGCTT CATGGCGACGTCTGCGA	cloning of UIP3 (attB1)	This study
UIP3s_B2 rev	GGGGACCACTTTGTACAAGAAAGCTGGGTC TCATTGACCAGTCTGCACC	cloning of UIP3 (attB2)	This study
UIP4_B1 fw	GGGGACAAGTTTGTACAAAAAAGCAGGCTT CATGGCGTATGCAAAGATCG	cloning of UIP4 (attB1)	This study
UIP4s_B2 rev	GGGGACCACTTTGTACAAGAAAGCTGGGTC TCAATTTCCCACAAGCCAAC	cloning of UIP4 (attB2)	This study
Amplification of promoter regions for GUS lines			
proBTSL1_-880_B1 fw	GGGGACAAGTTTGTACAAAAAAGCAGGCTT CAGTATTGATTTTGTGAGCCCAATT	cloning of BTSL1 promoter (attB1)	This study
proBTSL1_-880_B2 rev	GGGGACCACTTTGTACAAGAAAGCTGGGTC CCACCGCAACAACAATAACG	cloning of BTSL1 promoter (attB2)	This study
proBTS_-2994_B1 fw	GGGGACAAGTTTGTACAAAAAAGCAGGCTT CATGAGATGAAATGTCTTATCTTTAT	cloning of BTS promoter (attB1)	This study
proBTS_-2994_B2 rev	GGGGACCACTTTGTACAAGAAAGCTGGGTC TTCCCCCAAAGCTTATCTCCGTTTT	cloning of BTS promoter (attB2)	This study
proPYE_-1120_B1 fw	GGGGACAAGTTTGTACAAAAAAGCAGGCTT CACCGCAAAACTATATATAGTATTT	cloning of PYE promoter (attB1)	This study
proPYE_-1120_B2 rev	GGGGACCACTTTGTACAAGAAAGCTGGGTC CTTTGCTTTTATTACAGAACAAGA	cloning of PYE promoter (attB2)	This study
proFEP3_-1614_B1 fw	GGGGACAAGTTTGTACAAAAAAGCAGGCTT CGGCACAAAGAAATCAGACCAAT	cloning of FEP3 promoter (attB1)	This study
proFEP3_-1614_B2rev	GGGGACCACTTTGTACAAGAAAGCTGGGTC TGATATTTTTTGTTGGATATGAATGAAAGT	cloning of FEP3 promoter (attB2)	This study
TAIL PCR and genotyping			
S1_AL2_LB	ATGCTCTTACGTTGTTGTCGGG	TAIL, binds in T-DNA LB of pAlligator2	
S2_AL2_LB	ACCACTCATCATAGCTCCGC	TAIL, binds in T-DNA LB of pAlligator2	
S3_AL2_LB	TTCAGTACATTAAAAACGTCCGC	TAIL, binds in T-DNA LB of pAlligator2, genotyping of FEP3-OX lines	
AD1	NGTCGASWGANAWGAA	TAIL	AD2 in (Liu et al., 1995)
AD2	TGWGNAGSANCASAG	TAIL	AD1 in (Liu and Whittier, 1995)
AD3	AGWGNAGWANCAWAGG	TAIL	AD2 in (Liu and Whittier, 1995)
AD4	STTGNTASTNCTNTGC	TAIL	AD2 in (Tsugeki et al., 1996)
AD5	NTCGASTWTSGWGTT	TAIL	AD1 in (Liu et al., 1995)
AD6	WGTGNAGWANCANAGA	TAIL	AD3 in (Liu et al., 1995)
LBb1.3	ATTTTGCCGATTTCGGAAC	*btsl1-btsl2* DM genotyping	Salk Institute Genomic Analysis Laboratory
FEP3-OX1_chr5 fw	GCAAACCAAGCTTCCATGC	FEP3-OX1 genotyping	This study
FEP3-OX1_chr5 rev	CTCGCCTCCAATACCTCCTT	FEP3-OX1 genotyping	This study
FEP3-OX3_chr1 fw	CGAATGTACTTCGCTGACTTT	FEP3-OX3 genotyping	This study
FEP3-OX3_chr1 rev	TGTTGTTCTAAATTTTGGGATCTT	FEP3-OX3 genotyping	This study

btsl1-1_LP	TGCTTGGCATAATCCTTCTTG	*btsl1* allele genotyping	(Rodríguez-Celma et al., 2017)
btsl1-1_RP	GAACTCTCTTGCTTCCTGAAGC	*btsl1* allele genotyping	(Rodríguez-Celma et al., 2017)
btsl2-2_LP	TCGGTTATTCAGGCAAAACAC	*btsl2* allele genotyping	(Rodríguez-Celma et al., 2017)
btsl2-2_RP	CCCTTTGTACTCATCAGCAGC	*btsl2* allele genotyping	(Rodríguez-Celma et al., 2017)
RT-qPCR			
FEP3_stn fw	GGCCATCAAGAGATTTGACC	*FEP3* RT-qPCR mass standard	This study
FEP3_stn rev	TGGAAACCATGTTTGTTCATCT	*FEP3* RT-qPCR mass standard	This study
FEP3_qPCR fw	TGCTTCCACCGTGTATGTTG	*FEP3* RT-qPCR	This study
FEP3_qPCR rev	CAGGAGCATAATCATAGCCACTG	*FEP3* RT-qPCR	This study
BTSL1_stn fw	TCCCTGGATCCTCAGAAGAA	*BTSL1* RT-qPCR mass standard	This study
BTSL1_stn rev	CGGCTGGTTCTACAATGATG	*BTSL1* RT-qPCR mass standard	This study
BTSL1_qPCR fw	CTCCCCAGTGAAGGCTCTTC	*BTSL1* RT-qPCR	This study
BTSL1_qPCR rev	GACCTGCATATCTCCAAGCGA	*BTSL1* RT-qPCR	This study
BTSL2_stn fw	ATGAGCCGTTGGATTGCTAC	*BTSL2* RT-qPCR mass standard	This study
BTSL2_stn rev	CAAAATCAAATGCTTCCAAAAA	*BTSL2* RT-qPCR mass standard	This study
BTSL2_qPCR fw	GCTTGTATGTCGCGACTCAT	*BTSL2* RT-qPCR	This study
BTSL2_qPCR rev	AGAGCCTTCACCGGAGAATT	*BTSL2* RT-qPCR	This study
BTS_stn fw	AACTTGGATGTTCCCCGTCT	*BTS* RT-qPCR mass standard	This study
BTS_stn rev	ATCAACGGGCTTCTTCACAT	*BTS* RT-qPCR mass standard	This study
BTS_qPCR fw	CGGGGAAGGACTAGGAATCG	*BTS* RT-qPCR	This study
BTS_qPCR rev	CAGCAGATGGGGCAATTTGT	*BTS* RT-qPCR	This study
STD-BHLH038-5'	GGAGATAACCTAAATAACGGC	*BHLH38* RT-qPCR mass standard	(Naranjo Arcos, 2017)
STD-BHLH038-3'	GGTCCAGATCAGTGTTAGATTCA	*BHLH38* RT-qPCR mass standard	(Naranjo Arcos, 2017)
RT 5'bHLH38	AGCAGCAACCAAAGGCG	*BHLH38* RT-qPCR	(Wang et al., 2007)
RT 3'bHLH38	CCACTTGAAGATGCAAAGTGTAG	*BHLH38* RT-qPCR	(Wang et al., 2007)
STD-BHLH039-5'	AACCAAAGCAGCTTCCAAG	BHLH39 RT-qPCR mass standard	(Naranjo Arcos, 2017)
STD-BHLH039-3'	CGAAGAGAAAAAGGACGACA	BHLH39 RT-qPCR mass standard	(Naranjo Arcos, 2017)
bHLH039-RT 5'	GACGGTTTCTCGAAGCTTG	BHLH39 RT-qPCR	(Wang et al., 2007)
RT 3'bHLH39	GGTGGCTGCTTAACGTAACAT	BHLH39 RT-qPCR	(Wang et al., 2007)
STD-FIT-5'(MN)	AAGACATGACCAAAAATGTGTGT	*FIT* RT-qPCR mass standard	(Naranjo Arcos, 2017)
STD-FIT-3'(MN)	TGCATCTCCAACAATGGATGC	*FIT* RT-qPCR mass standard	(Naranjo Arcos, 2017)
FIT F (g166-187)	CCCTGTTTCATAGACGAGAACC	*FIT* RT-qPCR	(Naranjo Arcos, 2017)
RT-FIT-3'(MN)	ATCCTTCATACGCCCTCTCC	*FIT* RT-qPCR	(Bauer, 2016)
AtIRT1-temp-5'(898)	TAGCCATTGACTCCATGGC	*IRT1* RT-qPCR mass standard	(Klatte, 2008)
AtIRT1-temp-3'(1910)	AGAAAACTATGAATCGTGGGG	*IRT1* RT-qPCR mass standard	(Klatte, 2008)
AtIRT1-c-5'new	AAGCTTTGATCACGGTTGG	*IRT1* RT-qPCR	(Wang et al., 2007)
AtIRT1-c-3' (1622)	TTAGGTCCCATGAACTCCG	*IRT1* RT-qPCR	(Wang et al., 2007)
AtFRO2-temp-5'(3110)	CCATGCTCGATCTTGTCTTG	*FRO2* RT-qPCR mass standard	(Bauer, 2016)
AtFRO2-temp-3'(4105)	ATTCCGGAACTTTTGAAAGG	*FRO2* RT-qPCR mass standard	(Bauer, 2016)
FRO2-c5-RT5'	CTTGGTCATCTCCGTGAGC	*FRO2* RT-qPCR	(Wang et al., 2007)
FRO2-c3-RT3'	AAGATGTTGGAGATGGACGG	*FRO2* RT-qPCR	(Wang et al., 2007)
PYE_stn fw	ACCGAAAAGGATCAACAAGG	*PYE* RT-qPCR mass standard	This study
PYE _stn rev	CCATCAAGGCCATAACTTCC	*PYE* RT-qPCR mass standard	This study

PYE _qPCR fw	GTTCCCAGGACTTCCCATTT	*PYE* RT-qPCR	This study
PYE _qPCR rev	GTGTCTGGGGATCAGGTTGT	*PYE* RT-qPCR	This study
FRO3_stn fw	AATCAGATCGACCACCTTGC	*FRO3* RT-qPCR mass standard	This study
FRO3_stn rev	TTCTTTTGGTGAGAAGATTTTGG	*FRO3* RT-qPCR mass standard	This study
FRO3_qPCR fw	ATCGACCACCTTGCTGTTTC	*FRO3* RT-qPCR	This study
FRO3_qPCR rev	TTATCCCACTGCCTCCACTC	*FRO3* RT-qPCR	This study
NAS4_stn fw	CACTCTCTTCAAGCAGCTCGT	*NAS4* RT-qPCR mass standard	This study
NAS4_stn rev	CTGTAGCAAAAACAGCCAACA	*NAS4* RT-qPCR mass standard	This study
AtNAS4-RT810-5'	TGTAATCTCAAGGAAGCTAGGTG	*NAS4* RT-qPCR	(Klatte et al., 2009)
AtNAS4-RT947-3'	GCGAACTCCTCGATAATGC	*NAS4* RT-qPCR	(Schuler, 2011)
ILR3_stn fw	TGATGGCTCGGCTGGAAAC	*ILR3* RT-qPCR mass standard	This study
ILR3_stn rev	CTAAGAAAGCCGAGAAAGAGAGGAG	*ILR3* RT-qPCR mass standard	This study
ILR3_qPCR fw	GCATGTAGAGAGAAGCAGCGAC	*ILR3* RT-qPCR	This study
ILR3_qPCR rev	TGCGGACAGCATCAACCAAG	*ILR3* RT-qPCR	This study
bHLH104_stn fw	GAATTTGCAGCAGGAGCCAG	*BHLH104* RT-qPCR mass standard	This study
bHLH104_stn rev	GCCAAACGGAAGAATCCTAAACC	*BHLH104* RT-qPCR mass standard	This study
bHLH104_qPCR fw	GGTTGAGGAGGGAGAAGCTAAATG	*BHLH104* RT-qPCR	This study
bHLH104_qPCR rev	ACGGATTGCATCATCGAGTATAGC	*BHLH104* RT-qPCR	This study
STD-EF1Balpha2-5'	GCTGCTAAGAAGGACACCAAG	*EF1Balpha* (genomic) RT-qPCR mass standard	(Bauer, 2016)
STD-EF1Balpha2-3'	TGTTCTGTCCCTACTGGATCC	*EF1Balpha* (genomic) RT-qPCR mass standard	(Bauer, 2016)
EFc-5'	TATGGGATCAAGAAACTCACAAT	*EF1Balpha* RT-qPCR	(Bauer, 2016)
EFc-3'	CTGGATGTACTCGTTGTTAGGC	*EF1Balpha* RT-qPCR	(Wang et al., 2007)
AtEF-gen-5' (2522)	TCCGAACAATACCAGAACTAC	*EF1Balpha* (genomic) RT-qPCR	(Mai et al., 2016)
AtEF-gen-3' (2726)	CCGGGACATATGGAGGTAAG	*EF1Balpha* (genomic) RT-qPCR	(Wang et al., 2007)

Supplemental Literature Cited

Bauer P (2016) Regulation of iron acquisition responses in plant roots by a transcription factor. Biochemistry and Molecular Biology Education 44: 438-449

Gratz R, Manishankar P, Ivanov R, Köster P, Mohr I, Trofimov K, Steinhorst L, Meiser J, Mai HJ, Drerup M, Arendt S, Holtkamp M, Karst U, Kudla J, Bauer P, Brumbarova T (2019) CIPK11-dependent phosphorylation modulates FIT activity to promote Arabidopsis iron acquisition in response to calcium signaling. Dev Cell

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

Klatte M (2008) Characterisation of the nicotianamine synthase gene family in *Arabidopsis thaliana* in the context of metal homeostasis. Saarländische Universitäts- und Landesbibliothek [Dissertation]: doi:10.22028/D22291-22550

Klatte M, Schuler M, Wirtz M, Fink-Straube C, Hell R, Bauer P (2009) The analysis of Arabidopsis nicotianamine synthase mutants reveals functions for nicotianamine in seed iron loading and iron deficiency responses. Plant Physiol 150: 257-271

Liu YG, Mitsukawa N, Oosumi T, Whittier RF (1995) Efficient isolation and mapping of *Arabidopsis thaliana* T-DNA insert junctions by thermal asymmetric interlaced PCR. Plant Journal 8: 457-463

Liu YG, Whittier RF (1995) Thermal Asymmetric Interlaced PCR: automatable amplification and sequencing of insert end fragments from P1 and YAC clones for chromosome walking. Genomics 25: 674-681

Mai 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

Naranjo Arcos MA (2017) Investigating the role of the iron dependent bHLH039 transcription factor in coordinating Fe homeostasis in *Arabidopsis thaliana*. Saarländische Universitäts- und Landesbibliothek [Dissertation]: doi:10.22028/D22291-26817

Rodríguez-Celma J, Chou H, Kobayashi T, Long TA, Balk J (2019) Hemerythrin E3 Ubiquitin Ligases as Negative Regulators of Iron Homeostasis in Plants. Front Plant Sci 10: 98

Rodríguez-Celma J, Green RT, Connorton JM, Kruse I, Cui Y, Ling HQ, Balk J (2017) BRUTUS-LIKE proteins moderate the transcriptional response to iron deficiency in roots. *bioRxiv* [Preprint]: doi.org/10.1101/231365

Schuler M (2011) The role of nicotianamine in the metal homeostasis of *Arabidopsis thaliana*. Saarländische Universitäts- und Landesbibliothek [Dissertation]: doi:10.22028/D22291-22738

Tsugeki R, Kochieva EZ, Fedoroff NV (1996) A transposon insertion in the Arabidopsis SSR16 gene causes an embryo-defective lethal mutation. Plant Journal 10: 479-489

Wang HY, Klatte M, Jakoby M, Bäumlein H, Weisshaar B, Bauer P (2007) Iron deficiency-mediated stress regulation of four subgroup Ib BHLH genes in *Arabidopsis thaliana*. Planta 226: 897-908

Zhang J, Liu B, 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

Author contributions to Manuscript 2

Birte Schwarz

Designed, performed and analyzed the following experiments: Y2H screen (Supplemental Figure S2), targeted Y2H assays (Figures 2B; 3A; 4A; Supplemental Figures S4B-D), subcellular protein localization analyses (Supplemental Figure S13 (1, 2)), co-expression network construction (Supplemental Figure S1A), multiple sequence alignments (Supplemental Figures S6; S7; S8). Prepared stable transgenic lines FEP3-OX, *proFEP3:GUS*, *proBTSL1:GUS*. Supervised experiments. Prepared figures and tables, wrote the manuscript, reviewed/edited the manuscript.

Daniela M. Lichtblau

Designed, performed and analyzed the following experiments: Y2H screen (Supplemental Figure S3), targeted Y2H assays (Supplemental Figure S4A; S5), BiFC assays (Figure 2C (1)), subcellular protein localization and co-localization analyses (Figures 7A (1-3, 5, 6); 7B (2, 3), GUS assays (Figure 4A, B, D, F, G, H, J, L); Supplemental Figure S12A). Prepared stable transgenic lines *proBTS:GUS*, *proPYE:GUS*. Supervised experiments. Contributed to writing of the manuscript, reviewed/edited the manuscript.

Christopher Endres

Performed and analyzed the following experiments: targeted Y2H assays (Figure 3D; Supplemental Figures S4B; S9A), RT-qPCRs, immunoblots, seed Fe content measurements, chlorophyll content measurements, root length measurements, remaining subcellular protein localization and co-localization analyses, remaining GUS assays. Helped with writing of the Methods part.

Christin Sieberg

Performed and analyzed the remaining BiFC experiments.

Petra Bauer

Conceived and supervised the study, acquired funding, reviewed/edited the manuscript.

9 Concluding Remarks

The very initiation of the Fe deficiency response is still a great enigma. By pinpointing novel *cis*- and *trans*-acting regulators, and with the discovery of a potential link between putative Fe deficiency signaling and sensing proteins, the two studies in this thesis provide new mechanistic insights into Fe deficiency response regulation.

Finding *cis*-regulatory sequences that regulate specific genes or responses can be extremely tedious. By using machine learning - a powerful tool to uncover hidden patterns in genomes - we identified more than 100 promoter motifs that robustly predicted Fe deficiency-induced transcription (termed pCREs) of Arabidopsis genes. The underlying procedure of our approach can be applied to any other response, given that high-quality transcriptomic data is available. Several of our pCREs were similar to the important Strategy II Fe acquisition CRE IDE1, which regulates MA synthesis and Fe uptake genes. While the presence of IDE1-likes in Arabidopsis Fe deficiency response genes has been acknowledged by others (Kobayashi *et al.*, 2003; Murgia *et al.*, 2011), we link the motif to synthesis of coumarins, which are needed for Strategy I Fe acquisition. Because IDE1-likes were also conserved in the promoter of IRT1, IDE1 might coordinate release of coumarins and Fe import. In addition to IDE1-likes, we identified several more pCREs and connected them to different responses during Fe deficiency. We provide a comprehensive data source of potential *cis*- and corresponding *trans*-regulators of Fe deficiency responses along with the genes which are predicted to be regulated by them. To mention a specific example, our models predicted that Fe deficiency-induced *BTS* and *BTSL1* are regulated by bHLH TFs. Indeed, previous results from us provide experimental evidence that bHLH39 likely regulates *BTSL1* (Naranjo-Arcos *et al.*, 2017). Our findings therefore suggest that testing other bHLHs will lead to identifying *BTS* upstream regulators.

BTS, BTSL1 and BTSL2 have recently become the center of attention for their potential function as Fe sensors (Selote *et al.*, 2015; Matthiadis and Long, 2016; Hindt *et al.*, 2017; Rodríguez-Celma *et al.*, 2019), however, research especially regarding BTSL1/2 is still in the early stages. We were excited to find in our targeted protein interaction screen that BTSL1 and BTSL2 can physically interact with FEP3, another recently emerged player and small peptide with a possible function in long-distance signaling (Grillet *et al.*, 2018). We show that FEP3-OX lines resemble the molecular phenotype of a *btsl1-btsl2* double loss-of-function mutant, and hypothesize that FEP3 acts as inhibitor of BTSL1/2, thereby de-repressing Fe acquisition. In addition to interacting with FEP3, we provide experimental evidence that BTSL1/2 - as suspected from what is known about BTS - can physically interact with IVc bHLHs. From a mechanistic perspective, FEP3 binding to BTSL1/2 could prevent IVc bHLH protein degradation by BTSL1/2, a theory which is encouraged by our finding that transcript levels of all genes which are supposed to be induced by IVc bHLHs are increased in FEP3-OX and *btsl1-btsl2* lines. BTSL1/2 could also physically interact with PYE, which was unexpected and

opens up new perspectives on Fe acquisition regulation. Apart from *in vivo* protein interaction data, we also show an unexpected, yet intriguing localization pattern of BTS, BTSL1 and BTSL2, with BTS being localized only in the nucleus, BTSL2 being localized in nucleus and cytoplasm and BTSL1 being localized primarily at the cell periphery and at low levels in the nucleus. Taken together, our study provides fascinating insights into the molecular mechanisms regulating Fe homeostasis, and raises several new questions. For example, what determines the different subcellular localizations of BTS, BTSL1 and BTSL2 and how are they functionally coordinated? What are their individual protein targets during Fe deficiency responses? Is FEP3 part of the long-sought Fe deficiency signal, and do BTSL1/2 act as its receptors in this scenario? Future efforts will shed light on these aspects.

Both studies in this work are based on Fe deficiency-responsive co-expression information. Co-expression networks have thus proven to be an excellent starting point to uncover regulatory connections and protein interaction networks. In addition to the BTS(L) interactome, our protein interaction screen approach has revealed several more novel protein-protein interactions, some of which strikingly include other small proteins. Following up on those with in-depth studies will further improve our mechanistic understanding of Fe deficiency responses.

References

Grillet L, Lan P, Li WF, Mokkapati G, Schmidt W (2018) IRON MAN is a ubiquitous family of peptides that control iron transport in plants. Nature Plants 4: 953-+

Hindt MN, Akmakjian GZ, Pivarski KL, Punshon T, Baxter I, Salt DE, Guerinot ML (2017) BRUTUS and its paralogs, BTS LIKE1 and BTS LIKE2, encode important negative regulators of the iron deficiency response in *Arabidopsis thaliana*. Metallomics 9: 876-890

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

Matthiadis A, Long TA (2016) Further insight into BRUTUS domain composition and functionality. Plant Signal Behav 11: e1204508

Murgia I, Tarantino D, Soave C, Morandini P (2011) Arabidopsis CYP82C4 expression is dependent on Fe availability and circadian rhythm, and correlates with genes involved in the early Fe deficiency response. J Plant Physiol 168: 894-902

Naranjo-Arcos MA, Maurer F, Meiser J, Pateyron S, Fink-Straube C, Bauer P (2017) Dissection of iron signaling and iron accumulation by overexpression of subgroup Ib bHLH039 protein. Scientific Reports 7

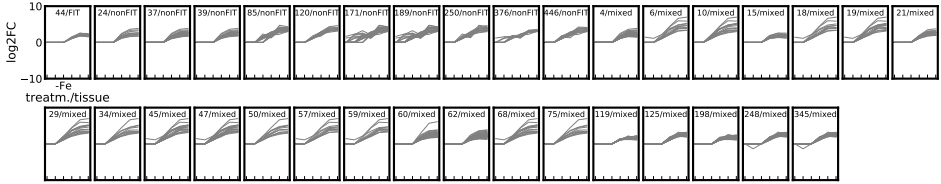
Rodríguez-Celma J, Chou H, Kobayashi T, Long TA, Balk J (2019) Hemerythrin E3 Ubiquitin Ligases as Negative Regulators of Iron Homeostasis in Plants. Front Plant Sci 10: 98

Selote D, Samira R, Matthiadis 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

Figure S8. Expression plots of all well-performing (F1≥0.7; n=517) co-expression clusters.

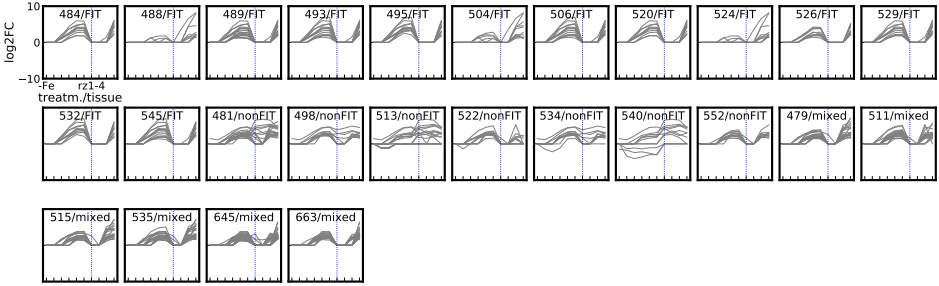
Clusters are organized according to transcriptomic data combinations (**Figure 1C**). GS-enriched (n=159) and non-enriched (n=358) clusters of each data combination are shown on separate pages. Each cluster plot is labeled with its cluster ID, as in **Supplemental Table S2**, and with its enrichment category (FIT-dependent/FIT, FIT-independent/non-FIT, both/mixed, none (if not GS-enriched)). **Pages 1, 2**: -Fe time course[1] data only (Combination 1**). Pages 3, 4**: Time course + -Fe-treated root zones 1-4[1] (Combination 2). **Pages 5, 6**: Time course + hormones[3] (Combination 4). **Pages 7-10**: Time course + abiotic stresses[2] (Combination 3). **Page 11**: Time course + abiotic stresses + hormones (Combination 5a). **Pages 12-15**: Time course + abiotic stresses (genotoxic data deleted) + hormones (Combination 5b). **Pages 16, 17**: Time course + abiotic stresses + development[4] (Combination 6). Total number of plots: 517 (159 GS-enriched, 358 non-enriched). Blue dotted lines separate transcriptomic data within combinations. [1](Dinneny et al., 2008), [2](Kilian et al., 2007), [3](Goda et al., 2008), [4](Schmid et al., 2005). Units: for all data except development: $\log_2$ fold-change (log2FC); development: fluorescence intensity.

Fe deficiency time course, GS-enriched

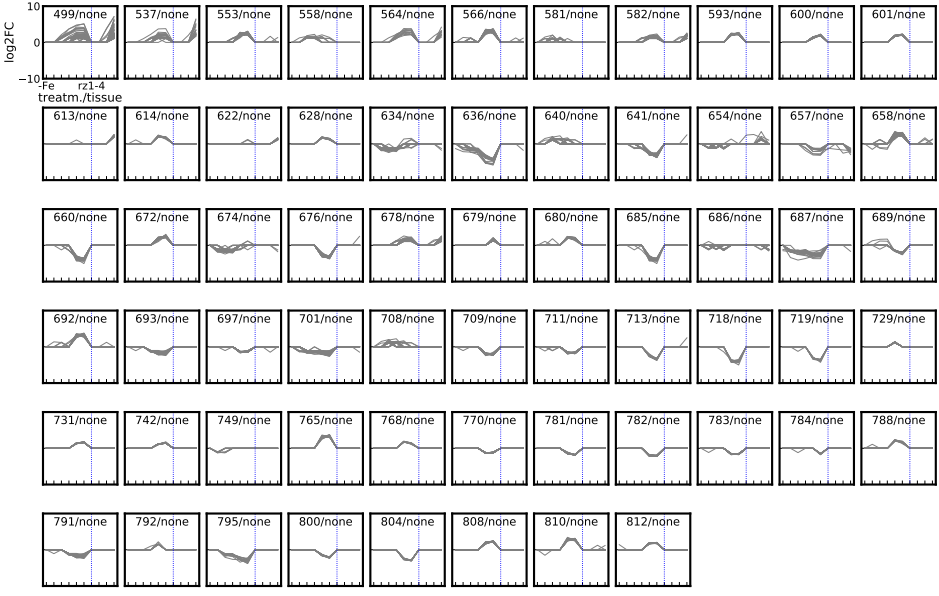


Fe deficiency time course, non-enriched

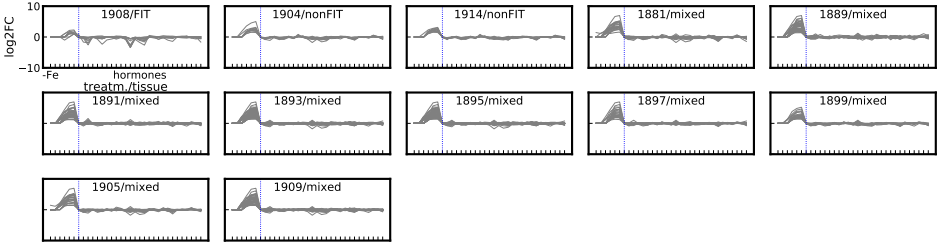
Fe deficiency time course/root zones, GS-enriched



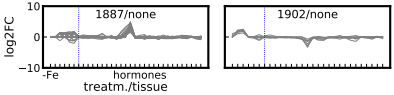
Fe deficiency time course/root zones, non-enriched



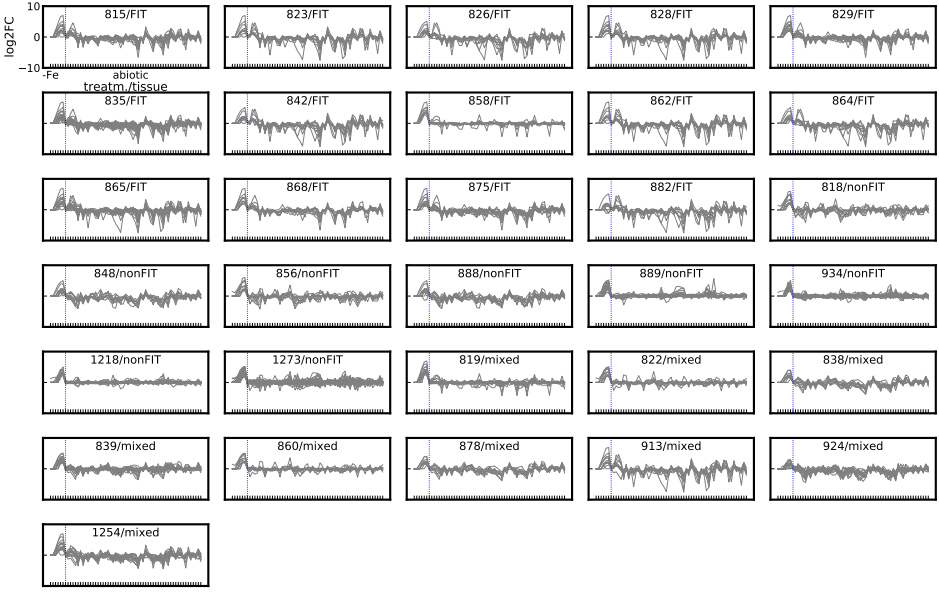
Fe deficiency time course/hormones, GS-enriched



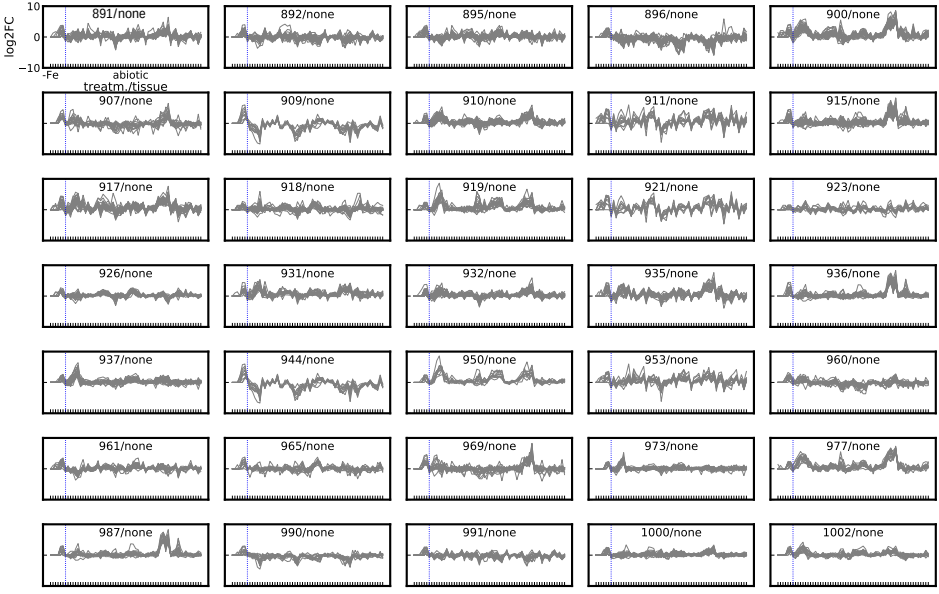
Fe deficiency time course/hormones, non-enriched



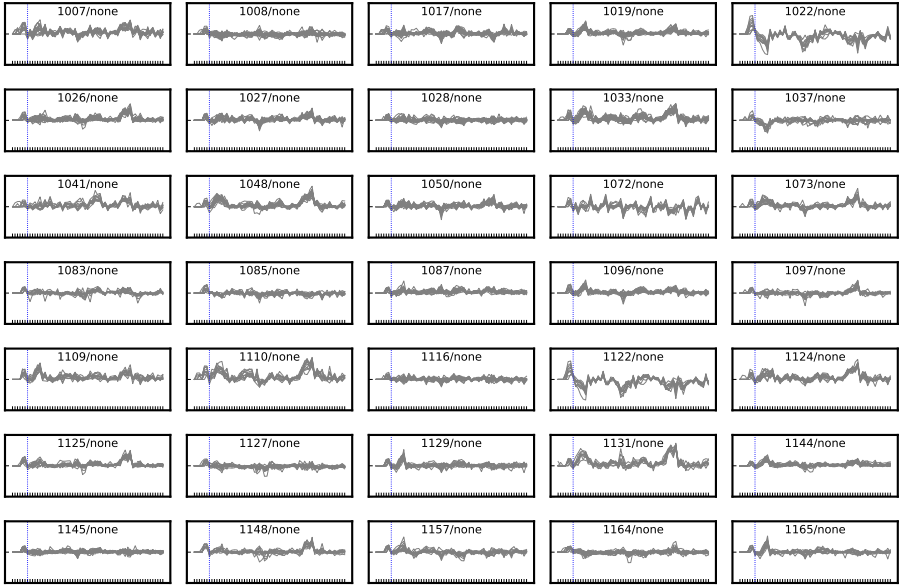
Fe deficiency time course/abiot. stresses, GS-enriched



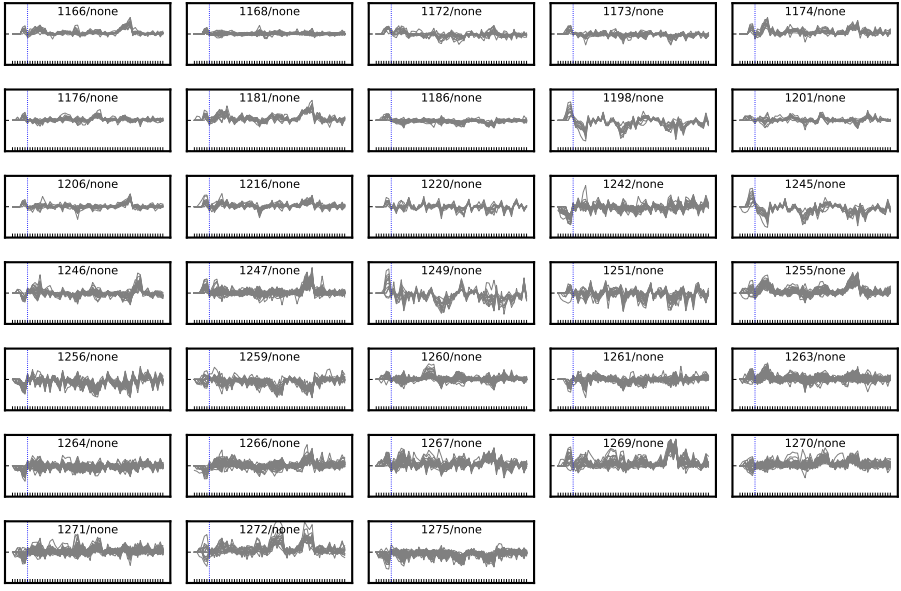
Fe deficiency time course/abiot. stresses, non-enriched (1)



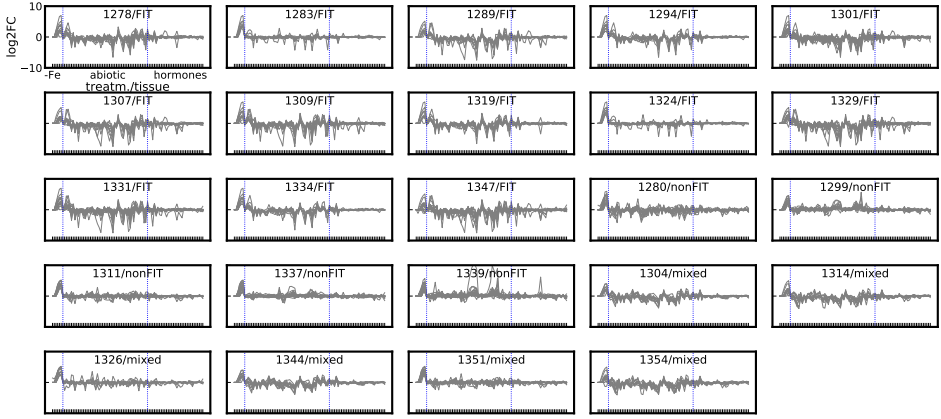
Fe deficiency time course/abiot. stresses, non-enriched (2)



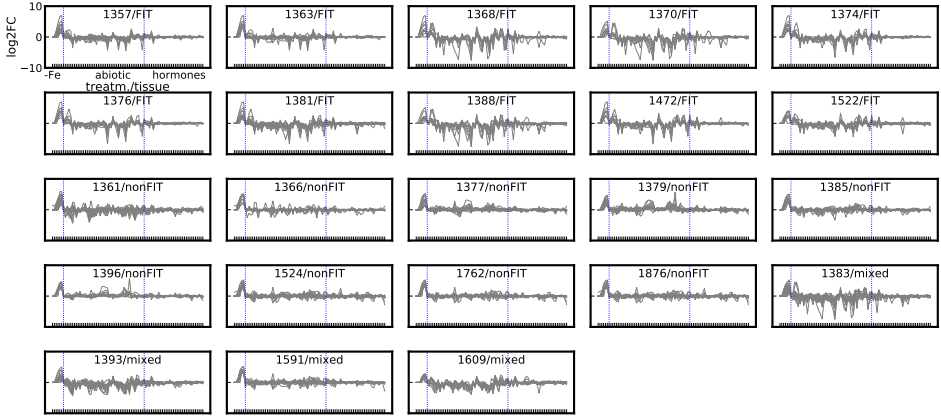
Fe deficiency time course/abiot. stresses, non-enriched (3)



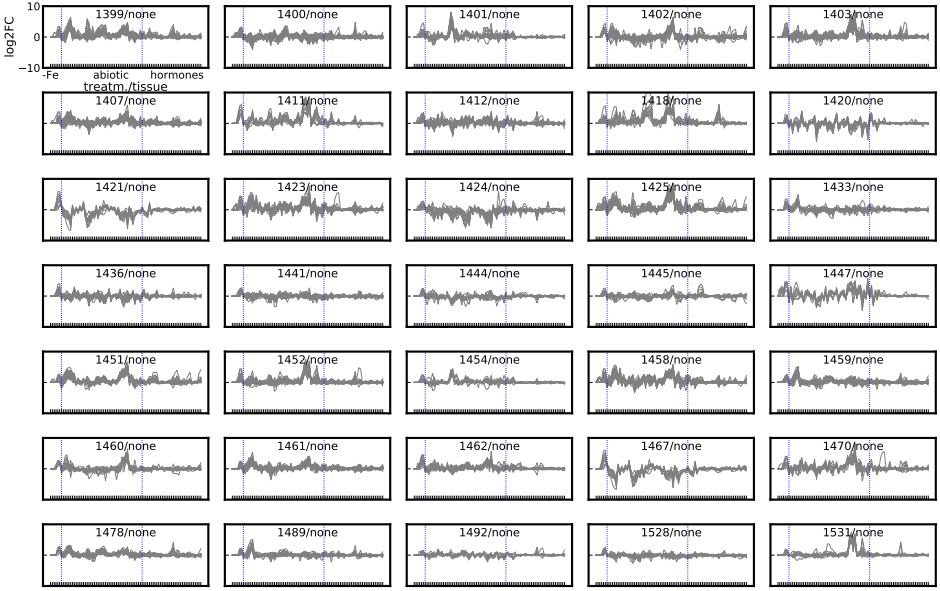
Fe deficiency time course/abiot. stresses/hormones, GS-enriched



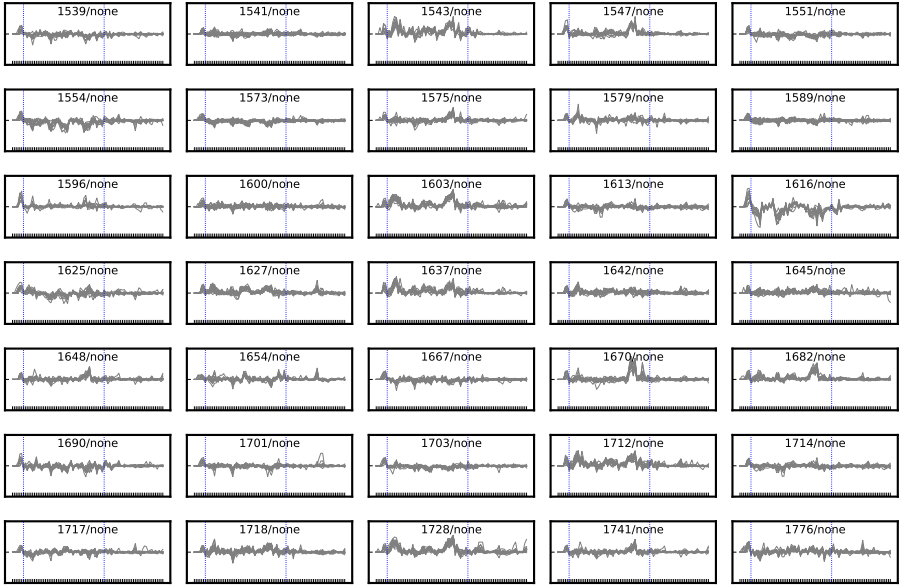
Fe deficiency time course/abiot. stresses (ng)/hormones, GS-enriched



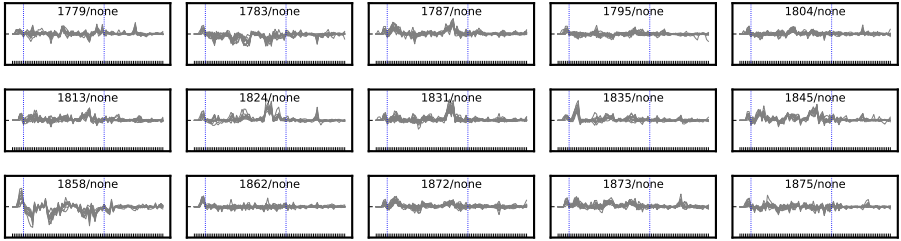
Fe deficiency time course/abiot. stresses (ng)/hormones, non-enriched (1)



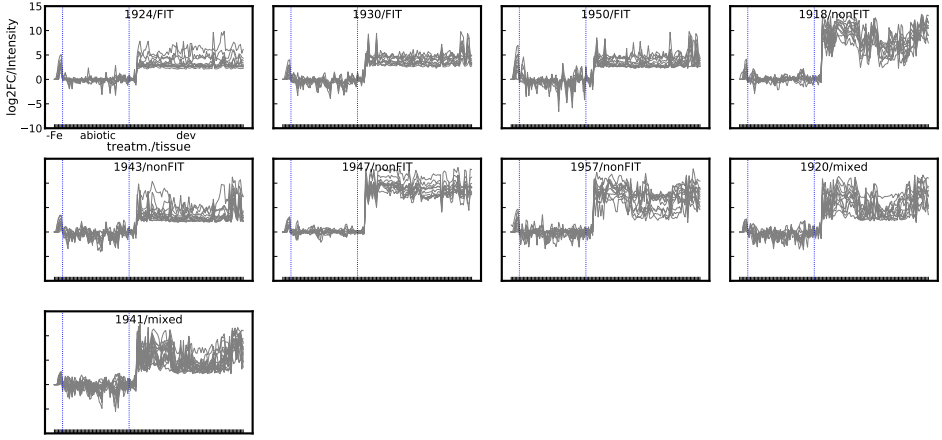
Fe deficiency time course/abiot. stresses (ng)/hormones, non-enriched (2)



Fe deficiency time course/abiot. stresses (ng)/hormones, non-enriched (3)



Fe deficiency time course/abiot. stresses/developm., GS-enriched



Fe deficiency time course/abiot. stresses/developm., non-enriched

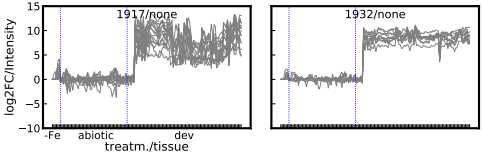


Figure S9. Positional distribution plots of all 615 min-pCREs.

Page 1: Introductory table into min-pCRE ID system (numbered 1-615, sorted by alphabet). **Pages 2-11:** Positional distribution plots. **1st column:** plots of putative promoter region; 1000 bp upstream and 500 bp downstream of TSS (transcription start site). **2nd column:** plots of all introns (In) and all exons (Ex). **3rd column:** plots of putative non-coding region; 500 bp upstream and 1000 bp downstream of TTS (transcription termination site). 1st and 3rd column: position distributions of all co-expression clusters with min-pCRE (gray, filled) plotted as overlay and mean distribution (red, line). 2nd column: bars represent mean with standard deviation. min-pCREs are sorted in columns extended over two pages. **Pages 2+3:** min-pCREs 1-125, **pages 4+5:** min-pCREs 126-250, **pages 6+7:** min-pCREs 251-375, **pages 8+9:** min-pCREs 376-500, **pages 10+11:** min-pCREs 501-615.

Part1

ID (as in S3/S4/S5_Table)	min-pCRE	Count GS-enriched clusters with min-pCRE	ID in this file
8	AAAAACATT	1	1
46	AAAAGT	1	2
50	AAAAGTC	2	3
87	AAACAAC	2	4
91	AAACAGC	1	5
96	AAACATTA	1	6
104	AAACGA	1	7
112	AAACGC	1	8
122	AAACTA	7	9
123	AAACTAC	1	10
128	AAACTG	1	11
150	AAAGGCT	1	12
152	AAAGTA	9	13
170	AAAGTCT	1	14
179	AAATAAC	2	15
189	AAATAG	3	16
191	AAATAGT	1	17
206	AAATATT	2	18
209	AAATATTAT	1	19
211	AAATATTT	2	20
213	AAATCT	1	21
215	AAATCTG	1	22
230	AAATGTA	1	23
235	AAATTA	1	24
282	AACACT	2	25
286	AACAGC	1	26
287	AACAGG	1	27
297	AACATGG	1	28
298	AACATT	1	29
299	AACATTA	2	30
310	AACCCA	1	31
333	AACCTC	1	32
343	AACGAG	1	33
352	AACGCA	2	34
358	AACGCC	2	35
368	AACGGA	1	36
370	AACGGC	14	37
371	AACGGCA	1	38
382	AACGTA	15	39
411	AACTAC	4	40
412	AACTACA	2	41
418	AACTAG	1	42
423	AACTAGT	1	43
426	AACTAT	1	44
427	AACTATA	1	45
435	AACTGA	1	46
469	AAGACGA	2	47
476	AAGAGT	1	48
481	AAGATG	1	49
483	AAGCAA	3	50
484	AAGCAC	1	51
501	AAGCTGC	2	52
511	AAGGAG	1	53
527	AAGGTC	1	54
532	AAGGTTA	1	55
556	AAGTCC	1	56
558	AAGTCG	1	57
570	AAGTGC	1	58
597	AATAAT	3	59
598	AATAATA	1	60
599	AATAATAA	1	61
624	AATACTA	2	62
629	AATAGA	4	63
638	AATAGG	1	64
640	AATAGT	3	65
644	AATAGTT	1	66
645	AATATA	1	67
655	AATATAT	1	68
661	AATATC	2	69
665	AATATG	8	70
674	AATATTAG	1	71
677	AATATTT	1	72
680	AATCAC	1	73
684	AATCAT	3	74
687	AATCCC	1	75
688	AATCCCC	1	76
696	AATCTG	1	77
708	AATGATA	1	78
712	AATGCA	3	79
721	AATGCG	6	80
744	AATGTG	2	81
747	AATTAA	3	82
756	AATTAAT	1	83
760	AATTAC	1	84
793	AATTGC	13	85
801	AATTTAA	1	86
822	ACAACA	1	87
823	ACAACC	2	88

Part2

ID (as in S3/S4/S5_Table)	min-pCRE	Count GS-enriched clusters with min-pCRE	ID in this file
873	ACACGA	1	89
877	ACACGC	2	90
885	ACACGT	3	91
905	ACACTA	9	92
911	ACACTTA	1	93
916	ACAGCA	5	94
935	ACATAAC	3	95
938	ACATAG	1	96
940	ACATAT	1	97
948	ACATCA	1	98
958	ACATGC	3	99
963	ACATGGA	1	100
967	ACATTA	1	101
973	ACATTC	1	102
991	ACCACG	3	103
996	ACCAGC	2	104
999	ACCATA	2	105
1001	ACCATAC	1	106
1011	ACCCAC	14	107
1023	ACCGAA	4	108
1045	ACCGTC	1	109
1054	ACCTAC	3	110
1068	ACCTCG	2	111
1069	ACCTGG	2	112
1072	ACCTTA	3	113
1078	ACGAAA	10	114
1080	ACGAAAAA	1	115
1081	ACGAAAT	2	116
1090	ACGAAT	1	117
1113	ACGAGA	2	118
1114	ACGAGG	2	119
1125	ACGCAA	4	120
1126	ACGCAAA	3	121
1134	ACGCAT	31	122
1141	ACGCCA	4	123
1145	ACGCCG	1	124
1148	ACGCGC	6	125
1151	ACGCGCC	2	126
1158	ACGCGT	3	127
1166	ACGCTC	1	128
1177	ACGGAG	2	129
1178	ACGGCA	5	130
1191	ACGGGGT	1	131
1196	ACGGTG	1	132
1197	ACGTAA	1	133
1208	ACGTAG	1	134
1213	ACGTAGT	1	135
1229	ACGTCA	1	136
1231	ACGTCAG	1	137
1244	ACGTGA	1	138
1246	ACGTGAC	1	139
1261	ACGTGG	4	140
1266	ACGTGGC	2	141
1283	ACGTTA	2	142
1296	ACTAAG	1	143
1306	ACTACG	1	144
1310	ACTACT	3	145
1316	ACTAGG	6	146
1317	ACTAGGT	1	147
1318	ACTAGT	10	148
1323	ACTAGTT	1	149
1324	ACTATA	5	150
1325	ACTATAA	3	151
1328	ACTATAT	1	152
1339	ACTATTC	1	153
1346	ACTCCA	1	154
1348	ACTCCG	2	155
1357	ACTGAT	1	156
1365	ACTGTA	2	157
1386	ACTTGTA	1	158
1400	AGAAATA	1	159
1406	AGAATA	2	160
1413	AGAATG	1	161
1424	AGACTT	1	162
1426	AGACGA	2	163
1431	AGACGC	2	164
1438	AGACTG	2	165
1450	AGAGGC	1	166
1453	AGAGTGA	1	167
1468	AGATTA	3	168
1469	AGATTAA	1	169
1472	AGCAAT	2	170
1474	AGCACA	1	171
1481	AGCATA	1	172
1483	AGCATG	10	173
1485	AGCATGT	1	174
1489	AGCCAAA	3	175
1499	AGCCGG	2	176

Part3

ID (as in S3/S4/S5_Table)	min-pCRE	Count GS-enriched clusters with min-pCRE	ID in this file
1500	AGCCGGA	3	177
1510	AGCGCC	1	178
1555	AGGCAG	1	179
1558	AGGCAT	6	180
1575	AGGGAAA	1	181
1576	AGGGAC	3	182
1594	AGGTCA	1	183
1600	AGGTTA	4	184
1604	AGTAAC	6	185
1619	AGTAGG	4	186
1625	AGTATA	5	187
1641	AGTCAA	8	188
1647	AGTCAAC	5	189
1667	AGTCAT	4	190
1682	AGTCGG	1	191
1686	AGTCTA	1	192
1695	AGTGAAT	1	193
1696	AGTGAC	1	194
1701	AGTGAT	2	195
1704	AGTGGC	7	196
1705	AGTGGCA	1	197
1711	AGTGGG	4	198
1713	AGTGGGG	3	199
1715	AGTGTA	1	200
1724	AGTTAAA	2	201
1728	AGTTAT	1	202
1730	AGTTATAT	1	203
1734	AGTTATT	1	204
1769	ATAAATA	2	205
1770	ATAAATAA	1	206
1775	ATAACA	1	207
1779	ATAACC	6	208
1784	ATAACG	1	209
1794	ATAAGC	4	210
1801	ATAATA	2	211
1807	ATAATAC	2	212
1820	ATAATCA	3	213
1822	ATAATCC	1	214
1824	ATAATG	2	215
1835	ATAATTA	1	216
1841	ATACAAT	1	217
1842	ATACAC	2	218
1850	ATACAT	2	219
1860	ATACCC	3	220
1883	ATACTA	2	221
1893	ATACTGA	1	222
1899	ATAGTAT	1	223
1900	ATAGAC	1	224
1913	ATAGGC	1	225
1939	ATATAC	2	226
1940	ATATACA	2	227
1948	ATATAG	1	228
1956	ATATAT	6	229
1967	ATATATT	5	230
1971	ATATCA	2	231
1974	ATATCG	3	232
1979	ATATGA	1	233
1990	ATATTAA	1	234
1995	ATATTAT	1	235
2021	ATCACG	1	236
2035	ATCCCG	1	237
2037	ATCCTA	2	238
2042	ATCGCA	4	239
2043	ATCGCC	1	240
2055	ATCGTG	10	241
2064	ATCTTC	1	242
2089	ATGAGC	2	243
2091	ATGAGCT	1	244
2100	ATGCAA	2	245
2101	ATGCAC	1	246
2105	ATGCAT	13	247
2110	ATGCATTA	3	248
2111	ATGCCA	2	249
2114	ATGCCC	1	250
2117	ATGGGA	2	251
2128	ATGGTG	1	252
2142	ATGGCG	3	253
2149	ATGGTA	1	254
2152	ATGTAA	5	255
2168	ATGTGG	1	256
2187	ATTAAC	2	257
2192	ATTAAG	15	258
2206	ATTACA	3	259
2213	ATTACG	2	260
2219	ATTAGA	4	261
2221	ATTAGC	3	262
2222	ATTAGCT	1	263
2226	ATTATA	6	264

Part4

ID (as in S3/S4/S5_Table)	min-pCRE	Count GS-enriched clusters with min-pCRE	ID in this file
2236	ATTATC	1	265
2238	ATTATGA	1	266
2239	ATTCAA	1	267
2247	ATTCCA	1	268
2248	ATTCCAC	2	269
2253	ATTCGG	5	270
2266	ATTGAA	1	271
2278	ATTGAG	3	272
2280	ATTGCC	1	273
2284	ATTGGAT	2	274
2285	ATTGTA	1	275
2286	ATTGTAT	1	276
2298	ATTTAC	4	277
2301	ATTTAGA	4	278
2349	CAAATAG	1	279
2365	CAACCAG	2	280
2374	CAACGC	3	281
2387	CAACTA	1	282
2409	CAAGCG	1	283
2413	CAAGGG	1	284
2429	CAATAA	1	285
2433	CAATAC	2	286
2445	CAATTA	2	287
2446	CAATTAA	3	288
2473	CACACG	2	289
2477	CACACGC	2	290
2479	CACACGT	2	291
2506	CACCACG	1	292
2530	CACGAG	8	293
2534	CACGCA	8	294
2541	CACGCC	15	295
2542	CACGCCA	1	296
2547	CACGCG	1	297
2559	CACGTA	1	298
2572	CACGTC	6	299
2573	CACGTCA	1	300
2576	CACGTCAG	1	301
2584	CACGTG	3	302
2609	CACTAC	10	303
2612	CACTACT	1	304
2616	CACTCG	1	305
2620	CACTTA	1	306
2630	CAGACAA	1	307
2637	CAGCAC	1	308
2656	CAGGTC	1	309
2666	CAGTCG	1	310
2676	CAGTTTA	1	311
2679	CATAACA	1	312
2685	CATACA	2	313
2692	CATACG	2	314
2741	CATGCA	10	315
2748	CATGCC	48	316
2751	CATGCG	3	317
2765	CATGTA	1	318
2771	CATTAC	1	319
2785	CATTCC	3	320
2792	CATTGG	1	321
2794	CATTTA	1	322
2797	CATTTAG	1	323
2824	CCAAGG	3	324
2832	CCACAA	1	325
2843	CCACGA	2	326
2846	CCACGC	4	327
2855	CCACGG	1	328
2858	CCACGTC	3	329
2864	CCACTA	3	330
2865	CCACTC	2	331
2868	CCAGCAC	1	332
2869	CCAGCC	1	333
2870	CCAGCG	1	334
2879	CCATAC	6	335
2881	CCATACT	1	336
2887	CCATGC	19	337
2895	CCATTA	1	338
2897	CCATTC	4	339
2907	CCCACG	9	340
2911	CCCACTC	1	341
2920	CCCCAC	4	342
2923	CCCCACT	4	343
2951	CCCTCG	1	344
2960	CCGAGG	2	345
2972	CCGCGG	1	346
2980	CCGGCC	7	347
2981	CCGGCCG	1	348
2992	CCGTCA	1	349
3006	CCTACG	5	350
3038	CCTTGA	1	351
3042	CGAAAA	1	352

Part5

ID (as in S3/S4/S5_Table)	min-pCRE	Count GS-enriched clusters with min-pCRE	ID in this file
3048	CGAAGA	1	353
3049	CGAAGC	1	354
3050	CGAATA	2	355
3055	CGACAC	3	356
3059	CGACCA	1	357
3079	CGAGGC	3	358
3085	CGATGA	2	359
3086	CGATTA	2	360
3104	CGCATA	1	361
3105	CGCCATC	2	362
3108	CGCCAC	1	363
3120	CGCGAC	1	364
3123	CGCGCC	1	365
3140	CGCTCA	12	366
3150	CGCTTA	5	367
3155	CGGACC	1	368
3164	CGGCCA	8	369
3166	CGGCCG	1	370
3177	CGGTAC	1	371
3190	CGTAAA	3	372
3194	CGTAAC	5	373
3197	CGTAAG	6	374
3200	CGTACA	1	375
3212	CGTATAA	1	376
3217	CGTATC	1	377
3223	CGTCAG	3	378
3225	CGTCAGC	1	379
3235	CGTCGA	4	380
3238	CGTCGAC	2	381
3244	CGTGAC	21	382
3247	CGTGAG	1	383
3255	CGTGCC	2	384
3262	CGTGGA	16	385
3263	CGTGGAA	1	386
3267	CGTGGC	2	387
3280	CGTGTC	3	388
3291	CGTTAG	1	389
3299	CGTTGA	3	390
3300	CGTTGAA	4	391
3303	CGTTGGA	1	392
3321	CTAATA	3	393
3329	CTAATTA	1	394
3333	CTACAC	3	395
3354	CTAGGC	1	396
3355	CTAGTA	4	397
3367	CTATAC	1	398
3388	CTATTCG	1	399
3406	CTCAGC	2	400
3419	CTCCTCT	1	401
3426	CTCTAAA	1	402
3429	CTCTACT	1	403
3442	CTGAGA	1	404
3443	CTGATA	2	405
3455	CTGGAA	1	406
3461	CTGTAA	2	407
3477	CTTATA	6	408
3493	CTTGGC	1	409
3494	CTTGTA	1	410
3495	CTTGTAT	2	411
3497	CTTGTC	1	412
3500	CTTTAAT	1	413
3546	GAACGCA	4	414
3568	GAAGGTA	1	415
3574	GAAGTAG	1	416
3576	GAATAA	2	417
3601	GAATTA	3	418
3611	GACAAA	3	419
3623	GACAGC	5	420
3639	GACCAC	5	421
3645	GACCGC	1	422
3656	GACGAA	1	423
3674	GACGTA	1	424
3680	GACGTGG	2	425
3701	GACTCC	4	426
3723	GAGCCCA	1	427
3733	GAGGGA	7	428
3741	GAGTAC	1	429
3742	GAGTCA	8	430
3763	GATAAC	6	431
3773	GATAGA	4	432
3789	GATCTA	1	433
3802	GATGGC	1	434
3807	GATTAA	1	435
3808	GATTAAG	1	436
3810	GATTACA	2	437
3818	GATTTA	1	438
3819	GATTTAA	1	439
3821	GCAAAC	1	440

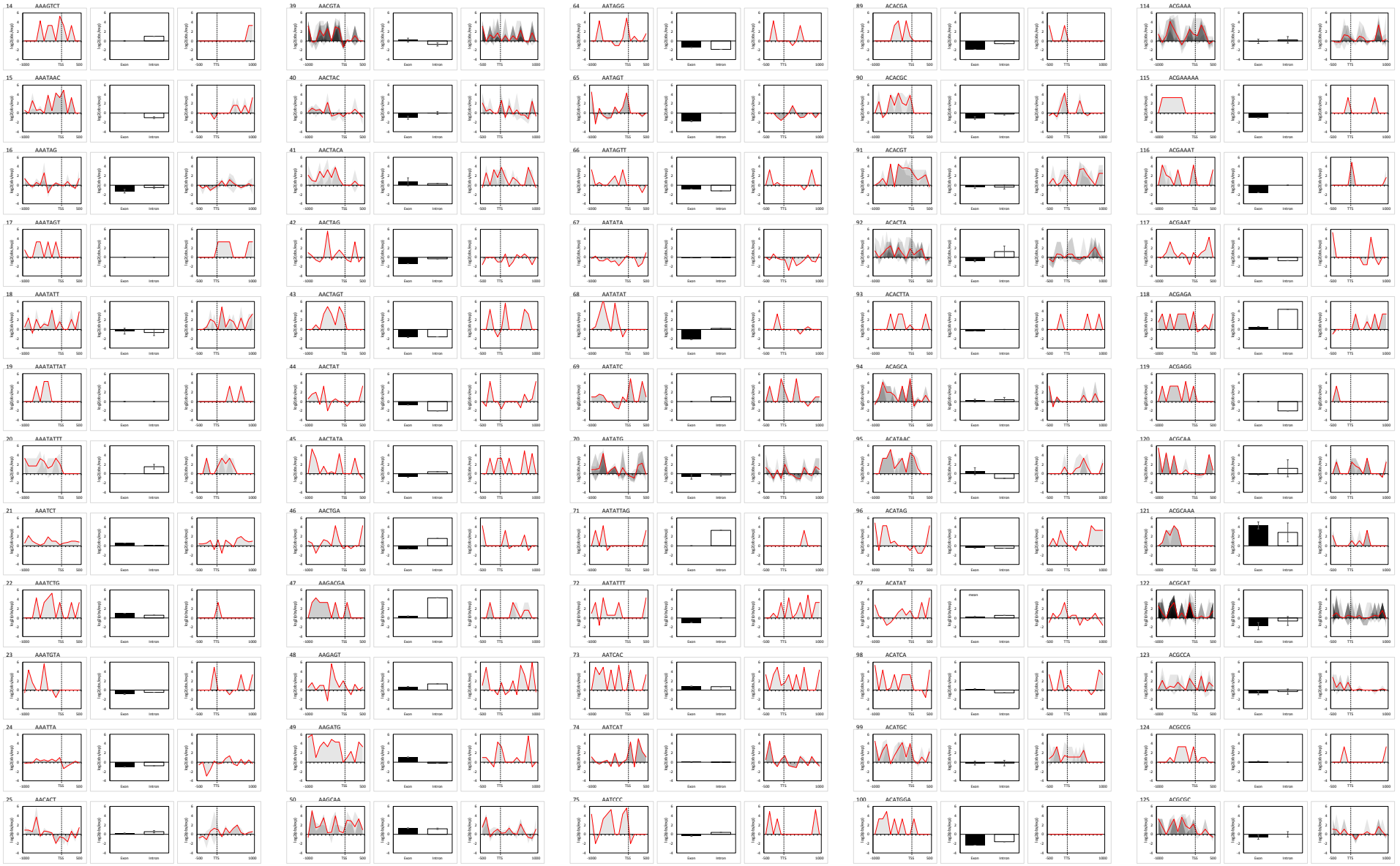
Part6

ID (as in S3/S4/S5_Table)	min-pCRE	Count GS-enriched clusters with min-pCRE	ID in this file
3827	GCACAA	1	441
3838	GCACGC	2	442
3854	GCAGCA	1	443
3875	GCATGA	20	444
3876	GCATGAA	1	445
3885	GCATGC	3	446
3889	GCATTA	4	447
3893	GCATTAT	3	448
3917	GCCCAC	3	449
3918	GCCCATA	1	450
3919	GCCCCA	5	451
3941	GCCTAC	2	452
3955	GCGCAC	1	453
3961	GCGCTGA	1	454
3968	GCGTAA	1	455
3977	GCGTCA	1	456
3980	GCGTCC	1	457
3981	GCGTGA	6	458
4003	GCTATA	1	459
4002	GCTATAA	1	460
4017	GCTGAC	5	461
4050	GGACCA	1	462
4053	GGACCC	1	463
4060	GGAGCC	11	464
4062	GGAGCCCA	3	465
4083	GGATTA	1	466
4087	GGCAAAA	1	467
4096	GGCAGA	1	468
4107	GGCCAC	2	469
4115	GGCCCC	2	470
4130	GGCGTA	1	471
4134	GGCGTGA	1	472
4140	GGGACC	1	473
4143	GGGATA	1	474
4148	GGGCCA	1	475
4158	GGGGCA	2	476
4172	GGGTCA	3	477
4175	GGGTTA	1	478
4179	GGTAAA	2	479
4196	GGTCAC	3	480
4201	GGTCGA	3	481
4207	GGTGAC	5	482
4210	GGTGTAA	1	483
4211	GGTTAA	1	484
4216	GGTTCA	1	485
4226	GTAAAC	1	486
4231	GTATAA	1	487
4235	GTAATCA	1	488
4237	GTACAA	4	489
4242	GTACCA	1	490
4252	GTAGAA	2	491
4253	GTAGAAA	1	492
4262	GTATAA	10	493
4276	GTATTA	13	494
4278	GTATTAAA	3	495
4282	GTATTAT	2	496
4285	GTCAAA	2	497
4292	GTCAAC	15	498
4310	GTCAGA	5	499
4311	GTCATA	1	500
4318	GTCCAC	1	501
4330	GTCGGA	8	502
4337	GTCTTA	4	503
4343	GTCTTAT	1	504
4349	GTGCAA	5	505
4354	GTGCCA	5	506
4364	GTGGAA	2	507
4367	GTGGAAT	1	508
4370	GTGGCA	8	509
4397	GTGTCAA	1	510
4401	GTGTTA	4	511
4403	GTGTTAT	3	512
4404	GTTAAA	2	513
4410	GTTATA	9	514
4420	GTTCTA	3	515
4426	GTTGCA	3	516
4427	GTTGGA	1	517
4457	TAAACTA	1	518
4462	TAAAGA	2	519
4463	TAAAGTA	3	520
4467	TAAATA	3	521
4468	TAAATAA	1	522
4472	TAAATAG	2	523
4474	TAAATTAA	2	524
4499	TAACTA	18	525
4506	TAAGAA	1	526
4508	TAAGCA	4	527
4509	TAAGCAC	1	528

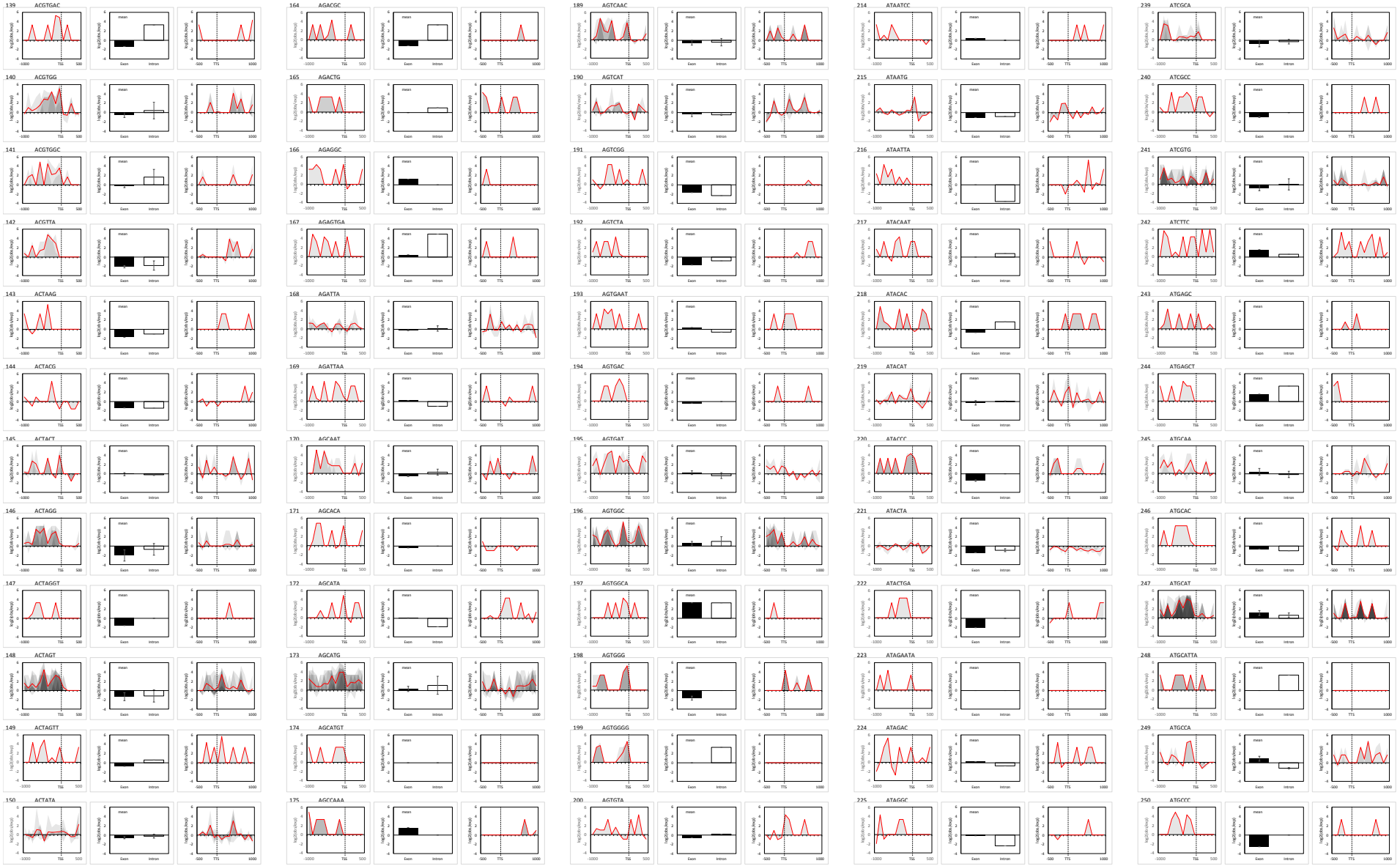
Part7

ID (as in S3/S4/S5_Table)	min-pCRE	Count GS-enriched clusters with min-pCRE	ID in this file
4526	TAAGTATA	1	529
4535	TAATAAT	1	530
4553	TAATATG	2	531
4563	TAATGA	1	532
4580	TAATTA	8	533
4581	TAATTAA	13	534
4583	TAATTAAC	1	535
4588	TAATTAT	2	536
4593	TAATTCC	1	537
4595	TAATTGC	1	538
4606	TACAATA	1	539
4621	TACAGA	5	540
4622	TACAGAA	1	541
4640	TACCATAC	1	542
4648	TACCGA	3	543
4654	TACCTA	1	544
4659	TACGAA	6	545
4660	TACGAAA	1	546
4663	TACGAAT	1	547
4685	TACGTA	1	548
4707	TACTAA	6	549
4715	TACTACT	1	550
4726	TACTCA	1	551
4733	TAGAATA	1	552
4741	TAGATA	1	553
4797	TAGTGGC	1	554
4803	TAGTTAC	1	555
4805	TAGTTTA	1	556
4816	TATAACC	4	557
4821	TATAATA	1	558
4830	TATACAA	2	559
4835	TATACAT	1	560
4844	TATAGA	8	561
4860	TATATA	1	562
4866	TATATACA	2	563
4874	TATATAT	1	564
4904	TATGAA	1	565
4910	TATGATA	1	566
4911	TATGCA	8	567
4912	TATGCAA	6	568
4923	TATGTTA	1	569
4924	TATTAA	3	570
4939	TATTCA	1	571
4947	TATTGA	1	572
4959	TATTTAC	1	573
4979	TCAACTA	2	574
5010	TCACGCC	1	575
5032	TCAGCGC	2	576
5036	TCATAA	1	577
5061	TCATTCC	1	578
5062	TCATTCCA	2	579
5064	TCATTGG	1	580
5065	TCATTTA	1	581
5088	TCCCCA	3	582
5128	TCGGAA	2	583
5142	TCGTAA	1	584
5161	TCGTTCC	1	585
5165	TCTAAAT	1	586
5184	TCTCAA	3	587
5194	TCTTAA	2	588
5210	TGAAGCA	1	589
5218	TGACAA	3	590
5225	TGACAGC	5	591
5259	TGATAA	1	592
5269	TGATTAA	1	593
5270	TGCACA	2	594
5279	TGCATGA	1	595
5281	TGCATGC	2	596
5331	TGGACA	2	597
5353	TGGGAA	1	598
5363	TGGTAA	3	599
5373	TGTACA	3	600
5374	TGTACAA	2	601
5377	TGTATAA	1	602
5380	TGTATCA	2	603
5418	TGTTAA	2	604
5422	TGTTAAT	3	605
5434	TTAACTA	1	606
5441	TTAATAAA	3	607
5444	TTAATTA	3	608
5451	TTAATTTA	1	609
5484	TTAGCTA	1	610
5494	TTATAA	3	611
5502	TTATATA	1	612
5550	TTCGTAA	1	613
5583	TTGTACA	1	614
5619	TTTATTAA	1	615

1 AAAAACATT
2 AAAAGT
3 AAAAGTC
4 AAACAAC
5 AAACAGC
6 AAACATTA
7 AAACGA
8 AAACGC
9 AAACTA
10 AAACTAC
11 AAACTG
12 AAAGGCT
13 AAAGTA
26 AACAGC
27 AACAGG
28 AACATGG
29 AACATT
30 AACATTA
31 AACCCA
32 AACCTC
33 AACGAG
34 AACGCA
35 AACGCC
36 AACGGA
37 AACGGC
38 AACGGCA
51 AAGCAC
52 AAGCTGC
53 AAGGAG
54 AAGGTC
55 AAGGTTA
56 AAGTCC
57 AAGTCG
58 AAGTGC
59 AATAAT
60 AATAATA
61 AATAATAA
62 AATACTA
63 AATAGA
76 AATCCCC
77 AATCTG
78 AATGATA
79 AATGCA
80 AATGCG
81 AATGTG
82 AATTAA
83 AATTAAT
84 AATTAC
85 AATTGC
86 AATTTAA
87 ACAACA
88 ACAACC
101 ACATTA
102 ACATTC
103 ACCACG
104 ACCAGC
105 ACCATA
106 ACCATAC
107 ACCCAC
108 ACCGAA
109 ACCGTC
110 ACCTAC
111 ACCTCG
112 ACCTGG
113 ACCTTA


14 AAAGTCT
15 AAATAAC
16 AAATAG
17 AAATAGT
18 AAATATT
19 AAATATTAT
20 AAATATTT
21 AAATCT
22 AAATCTG
23 AAATGTA
24 AAATTA
25 AACACT
39 AACGTA
40 AACTAC
41 AACTACA
42 AACTAG
43 AACTAGT
44 AACTAT
45 AACTATA
46 AACTGA
47 AAGACGA
48 AAGAGT
49 AAGATG
50 AAGCAA
64 AATAGG
65 AATAGT
66 AATAGTT
67 AATATA
68 AATATAT
69 AATATC
70 AATATG
71 AATATTAG
72 AATATTT
73 AATCAC
74 AATCAT
75 AATCCC
89 ACACGA
90 ACACGC
91 ACACGT
92 ACACTA
93 ACACTTA
94 ACAGCA
95 ACATAAC
96 ACATAG
97 ACATAT
98 ACATCA
99 ACATGC
100 ACATGGA
114 ACGAAA
115 ACGAAAAA
116 ACGAAAT
117 ACGAAT
118 ACGAGA
119 ACGAGG
120 ACGCAA
121 ACGCAAA
122 ACGCAT
123 ACGCCA
124 ACGCCG
125 ACGCGC

126 ACGCGCC
127 ACGCGT
128 ACGCTC
129 ACGGAG
130 ACGGCA
131 ACGGGGT
132 ACGGTG
133 ACGTAA
134 ACGTAG
135 ACGTAGT
136 ACGTCA
137 ACGTCAG
138 ACGTGA
151 ACTATAA
152 ACTATAT
153 ACTATTC
154 ACTCCA
155 ACTCCG
156 ACTGAT
157 ACTGTA
158 ACTTGTA
159 AGAAATA
160 AGAATA
161 AGAATG
162 AGACCTT
163 AGACGA
176 AGCCGG
177 AGCCGGA
178 AGCGCC
179 AGGCAG
180 AGGCAT
181 AGGGAAA
182 AGGGAC
183 AGGTCA
184 AGGTTA
185 AGTAAC
186 AGTAGG
187 AGTATA
188 AGTCAA
201 AGTTAAA
202 AGTTAT
203 AGTTATAT
204 AGTTATT
205 ATAAATA
206 ATAAATAA
207 ATAACA
208 ATAACC
209 ATAACG
210 ATAAGC
211 ATAATA
212 ATAATAC
213 ATAATCA
226 ATATAC
227 ATATACA
228 ATATAG
229 ATATAT
230 ATATATT
231 ATATCA
232 ATATCG
233 ATATGA
234 ATATTAA
235 ATATTAT
236 ATCACG
237 ATCCCG
238 ATCCTA


139 ACGTGAC
140 ACGTGG
141 ACGTGGC
142 ACGTTA
143 ACTAAG
144 ACTACG
145 ACTACT
146 ACTAGG
147 ACTAGGT
148 ACTAGT
149 ACTAGTT
150 ACTATA
164 AGACGC
165 AGACTG
166 AGAGGC
167 AGAGTGA
168 AGATTA
169 AGATTAA
170 AGCAAT
171 AGCACA
172 AGCATA
173 AGCATG
174 AGCATGT
175 AGCCAAA
189 AGTCAAC
190 AGTCAT
191 AGTCGG
192 AGTCTA
193 AGTGAAT
194 AGTGAC
195 AGTGAT
196 AGTGGC
197 AGTGGCA
198 AGTGGG
199 AGTGGGG
200 AGTGTA
214 ATAATCC
215 ATAATG
216 ATAATTA
217 ATACAAT
218 ATACAC
219 ATACAT
220 ATACCC
221 ATACTA
222 ATACTGA
223 ATAGAATA
224 ATAGAC
225 ATAGGC
239 ATCGCA
240 ATCGCC
241 ATCGTG
242 ATCTTC
243 ATGAGC
244 ATGAGCT
245 ATGCAA
246 ATGCAC
247 ATGCAT
248 ATGCATTA
249 ATGCCA
250 ATGCCC

251 ATGCGA
252 ATGCTG
253 ATGGCG
254 ATGGTA
255 ATGTAA
256 ATGTGG
257 ATTAAC
258 ATTAAG
259 ATTACA
260 ATTACG
261 ATTAGA
262 ATTAGC
263 ATTAGCT
276 ATTGTAT
277 ATTTAC
278 ATTTAGA
279 CAAATAG
280 CAACCAG
281 CAACGC
282 CAACTA
283 CAAGCG
284 CAAGGG
285 CAATAA
286 CAATAC
287 CAATTA
288 CAATTAA
301 CACGTCAG
302 CACGTG
303 CACTAC
304 CACTACT
305 CACTCG
306 CACTTA
307 CAGACAA
308 CAGCAC
309 CAGGTC
310 CAGTCG
311 CAGTTTA
312 CATAACA
313 CATACA
326 CCACGA
327 CCACGC
328 CCACGG
329 CCACGTC
330 CCACTA
331 CCACTC
332 CCAGCAC
333 CCAGCC
334 CCAGCG
335 CCATAC
336 CCATACT
337 CCATGC
338 CCATTA
351 CCTTGA
352 CGAAAA
353 CGAAGA
354 CGAAGC
355 CGAATA
356 CGACAC
357 CGACCA
358 CGAGGC
359 CGATGA
360 CGATTA
361 CGCATA
362 CGCATC
363 CGCCAC

264 ATTATA
265 ATTATC
266 ATTATGA
267 ATTCAA
268 ATTCCA
269 ATTCCAC
270 ATTCGG
271 ATTGAA
272 ATTGAG
273 ATTGCC
274 ATTGGAT
275 ATTGTA
289 CACACG
290 CACACGC
291 CACACGT
292 CACCACG
293 CACGAG
294 CACGCA
295 CACGCC
296 CACGCCA
297 CACGCG
298 CACGTA
299 CACGTC
300 CACGTCA
314 CATACG
315 CATGCA
316 CATGCC
317 CATGCG
318 CATGTA
319 CATTAC
320 CATTCC
321 CATTGG
322 CATTTA
323 CATTTAG
324 CCAAGG
325 CCACAA
339 CCATTC
340 CCCACG
341 CCCACTC
342 CCCCAC
343 CCCCACT
344 CCCTCG
345 CCGAGG
346 CCGCGG
347 CCGGCC
348 CCGGCCG
349 CCGTCA
350 CCTACG
364 CGCGAC
365 CGCGCC
366 CGCTCA
367 CGCTTA
368 CGGACC
369 CGGCCA
370 CGGCCG
371 CGGTAC
372 CGTAAA
373 CGTAAC
374 CGTAAG
375 CGTACA

376 CGTATAA
377 CGTATC
378 CGTCAG
379 CGTCAGC
380 CGTCGA
381 CGTCGAC
382 CGTGAC
383 CGTGAG
384 CGTGCC
385 CGTGGA
386 CGTGGAA
387 CGTGGC
388 CGTGTC
401 CTCCTCT
402 CTCTAAA
403 CTCTACT
404 CTGAGA
405 CTGATA
406 CTGGAA
407 CTGTAA
408 CTTATA
409 CTTGGC
410 CTTGTA
411 CTTGTAT
412 CTTGTC
413 CTTTAAT
426 GACTCC
427 GAGCCCA
428 GAGGGA
429 GAGTAC
430 GAGTCA
431 GATAAC
432 GATAGA
433 GATCTA
434 GATGGC
435 GATTAA
436 GATTAAG
437 GATTACA
438 GATTTA
451 GCCCCA
452 GCCTAC
453 GCGCAC
454 GCGCTGA
455 GCGTAA
456 GCGTCA
457 GCGTCC
458 GCGTGA
459 GCTATA
460 GCTATAA
461 GCTGAC
462 GGACCA
463 GGACCC
476 GGGGCA
477 GGGTCA
478 GGGTTA
479 GGTAAA
480 GGTCAC
481 GGTCGA
482 GGTGAC
483 GGTGTAA
484 GGTTAA
485 GGTTCA
486 GTAAAC
487 GTAATA
488 GTAATCA
-1000
TSS
500
Exon
Intron
-500
TTS
1000
mean

389 CGTTAG
390 CGTTGA
391 CGTTGAA
392 CGTTGGA
393 CTAATA
394 CTAATTA
395 CTACAC
396 CTAGGC
397 CTAGTA
398 CTATAC
399 CTATTCG
400 CTCAGC
414 GAACGCA
415 GAAGGTA
416 GAAGTGA
417 GAATAA
418 GAATTA
419 GACAAA
420 GACAGC
421 GACCAC
422 GACCGC
423 GACGAA
424 GACGTA
425 GACGTGG
439 GATTTAA
440 GCAAAC
441 GCACAA
442 GCACGC
443 GCAGCA
444 GCATGA
445 GCATGAA
446 GCATGC
447 GCATTA
448 GCATTAT
449 GCCCAC
450 GCCCATA
464 GGAGCC
465 GGAGCCCA
466 GGATTA
467 GGCAAAA
468 GGCAGA
469 GGCCAC
470 GGCCCC
471 GGCGTA
472 GGCGTGA
473 GGGACC
474 GGGATA
475 GGGCCA
489 GTACAA
490 GTACCA
491 GTAGAA
492 GTAGAAA
493 GTATAA
494 GTATTA
495 GTATTAAA
496 GTATTAT
497 GTCAAA
498 GTCAAC
499 GTCAGA
500 GTCATA

501 GTCCAC
502 GTCGGA
503 GTCTTA
504 GTCTTAT
505 GTGCAA
506 GTGCCA
507 GTGCAA
508 GTGGAAT
509 GTGGCA
510 GTGTCAA
511 GTGTTA
512 GTGTTAT
513 GTTAAA
526 TAAGAA
527 TAAGCA
528 TAAGCAC
529 TAAGTATA
530 TAATAAT
531 TAATATG
532 TAATGA
533 TAATTA
534 TAATTAA
535 TAATTAAC
536 TAATTAT
537 TAATTCC
538 TAATTGC
551 TACTCA
552 TAGAATA
553 TAGATA
554 TAGTGGC
555 TAGTTAC
556 TAGTTTA
557 TATAACC
558 TATAATA
559 TATACAA
560 TATACAT
561 TATAGA
562 TATATA
563 TATATACA
576 TCAGCGC
577 TCATAA
578 TCATTCC
579 TCATTCCA
580 TCATTGG
581 TCATTTA
582 TCCCCA
583 TCGGAA
584 TCGTAA
585 TCGTTCC
586 TCTAAAT
587 TCTCAA
588 TCTTAA
601 TGTACAA
602 TGTATAA
603 TGTATCA
604 TGTTAA
605 TGTTAAT
606 TTAACTA
607 TTAATAAA
608 TTAATTA
609 TTAATTTA
610 TTAGCTA
611 TTATAA
612 TTATATA
613 TTCGTAA

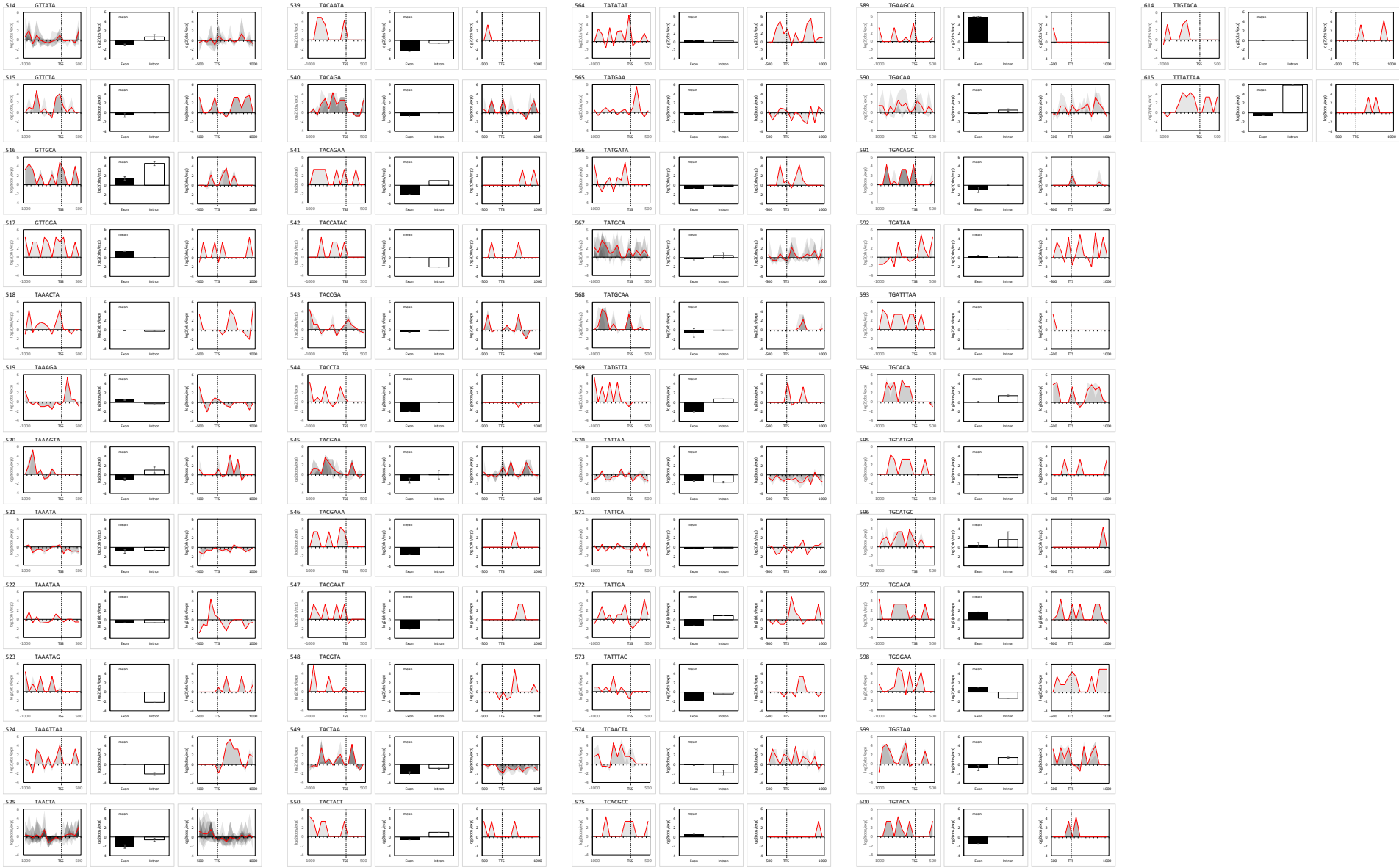

514 GTTATA
515 GTTCTA
516 GTTGCA
517 GTTGGA
518 TAAACTA
519 TAAAGA
520 TAAAGTA
521 TAAATA
522 TAAATAA
523 TAAATAG
524 TAAATTAA
525 TAACTA
539 TACAATA
540 TACAGA
541 TACAGAA
542 TACCATAC
543 TACCGA
544 TACCTA
545 TACGAA
546 TACGAAA
547 TACGAAT
548 TACGTA
549 TACTAA
550 TACTACT
564 TATATAT
565 TATGAA
566 TATGATA
567 TATGCA
568 TATGCAA
569 TATGTTA
570 TATTAA
571 TATTCA
572 TATTGA
573 TATTTAC
574 TCAACTA
575 TCACGCC
589 TGAAGCA
590 TGACAA
591 TGACAGC
592 TGATAA
593 TGATTTAA
594 TGCACA
595 TGCATGA
596 TGCATGC
597 TGGACA
598 TGGGAA
599 TGGTAA
600 TGTACA
614 TTGTACA
615 TTTATTAA

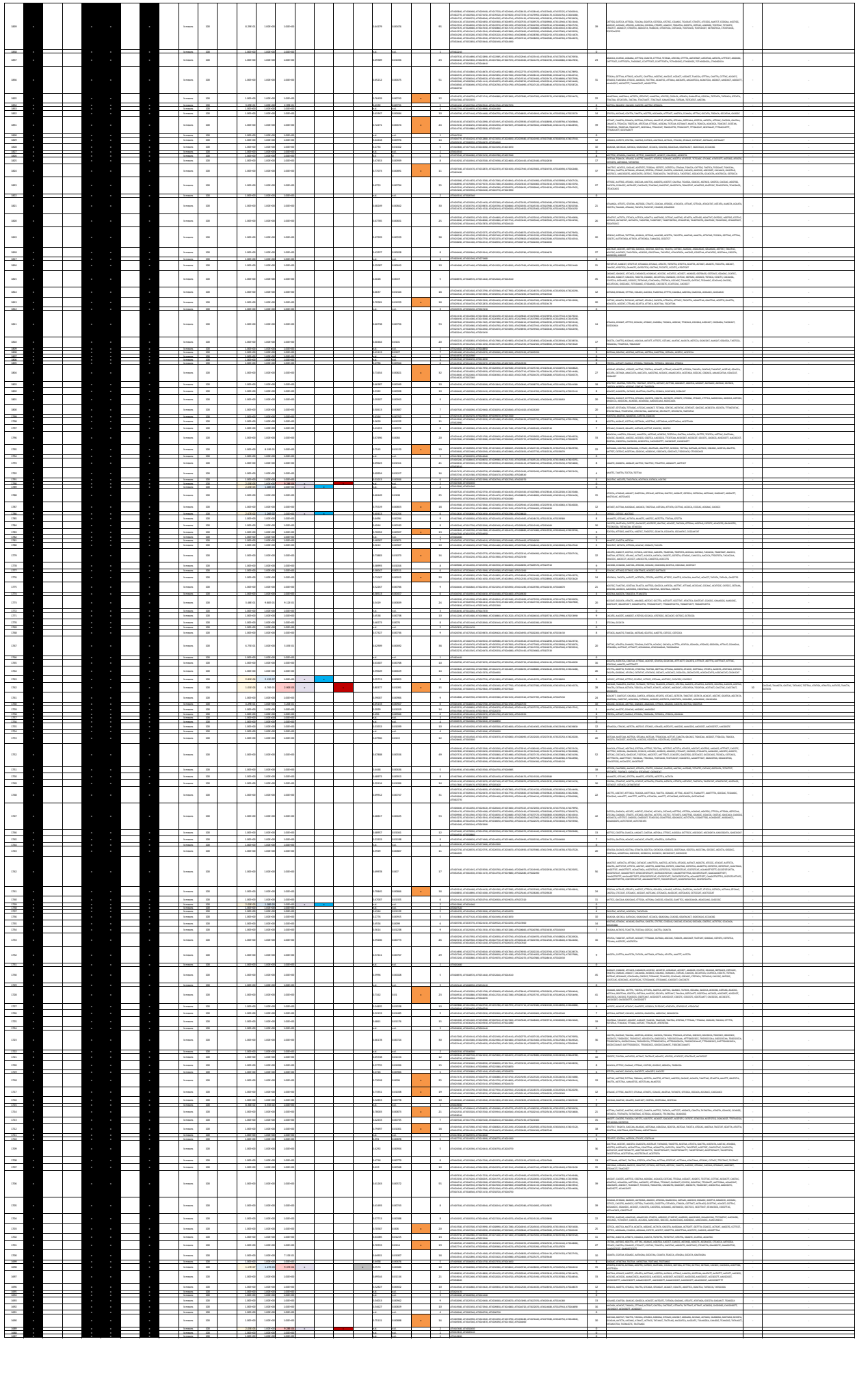
Table S2. All generated co-expression clusters. Table shows the input expression data combinations, the algorithm and parameters used for clustering, significant enrichment in GS genes and more detailed enrichment in PF-dependent genes (>FC, FF-independent genes (<one FF) or both (crossed). Performance of Random Forest model with standard error of the mean was added. The last columns list the genes within each cluster and the pCSCs identified in the cluster, ranked by importance for prediction (Gini index). In GS-enriched clusters, the minimal set of pCSCs (min-pCSCs) was determined and min-pCSCs are listed as well.

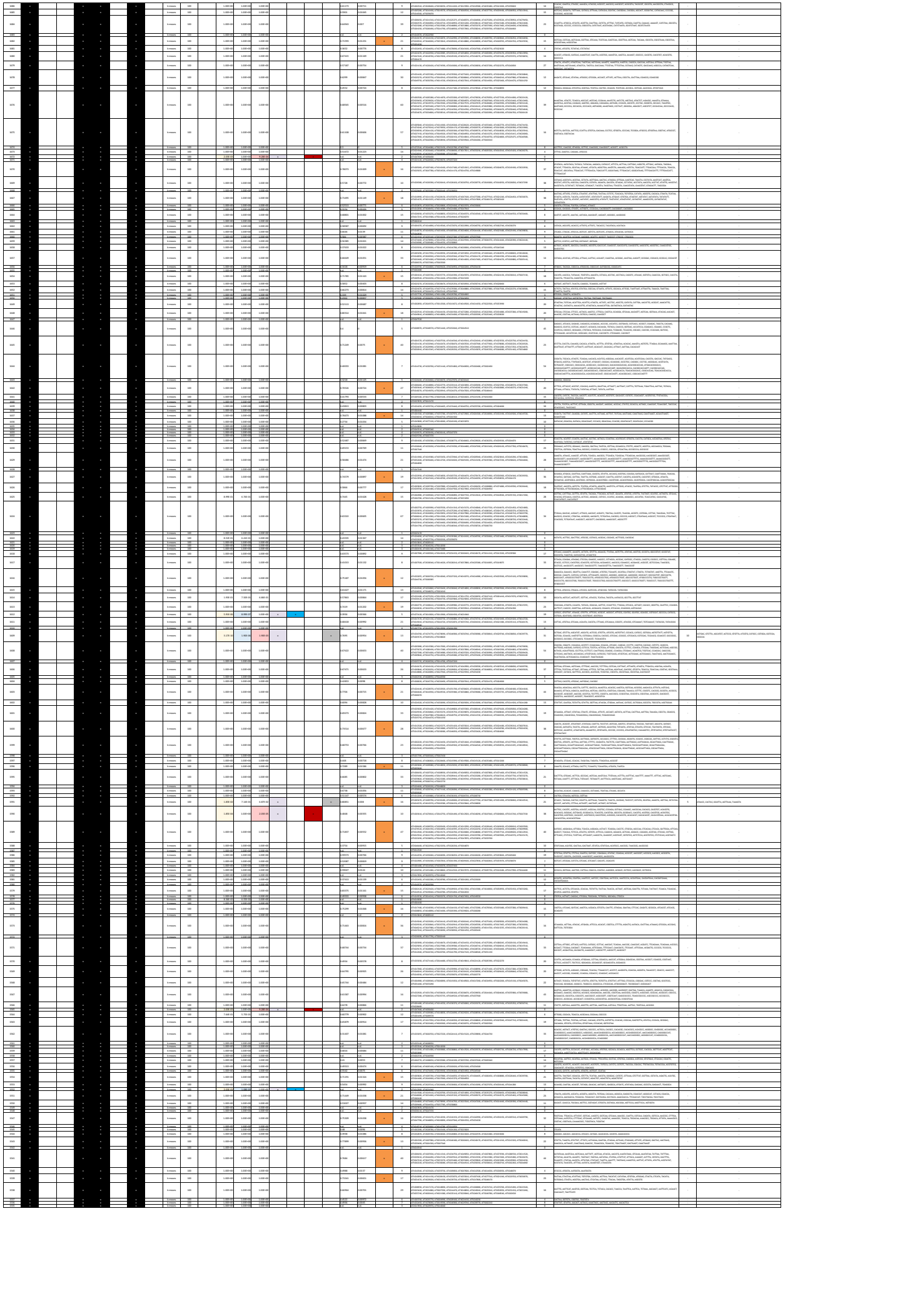
FC = fire deficiency, FCC = Pearson's correlation coefficient, FET = Fisher's exact test, RF = Random Forest, F1 = prediction score, E = Standard error of the mean

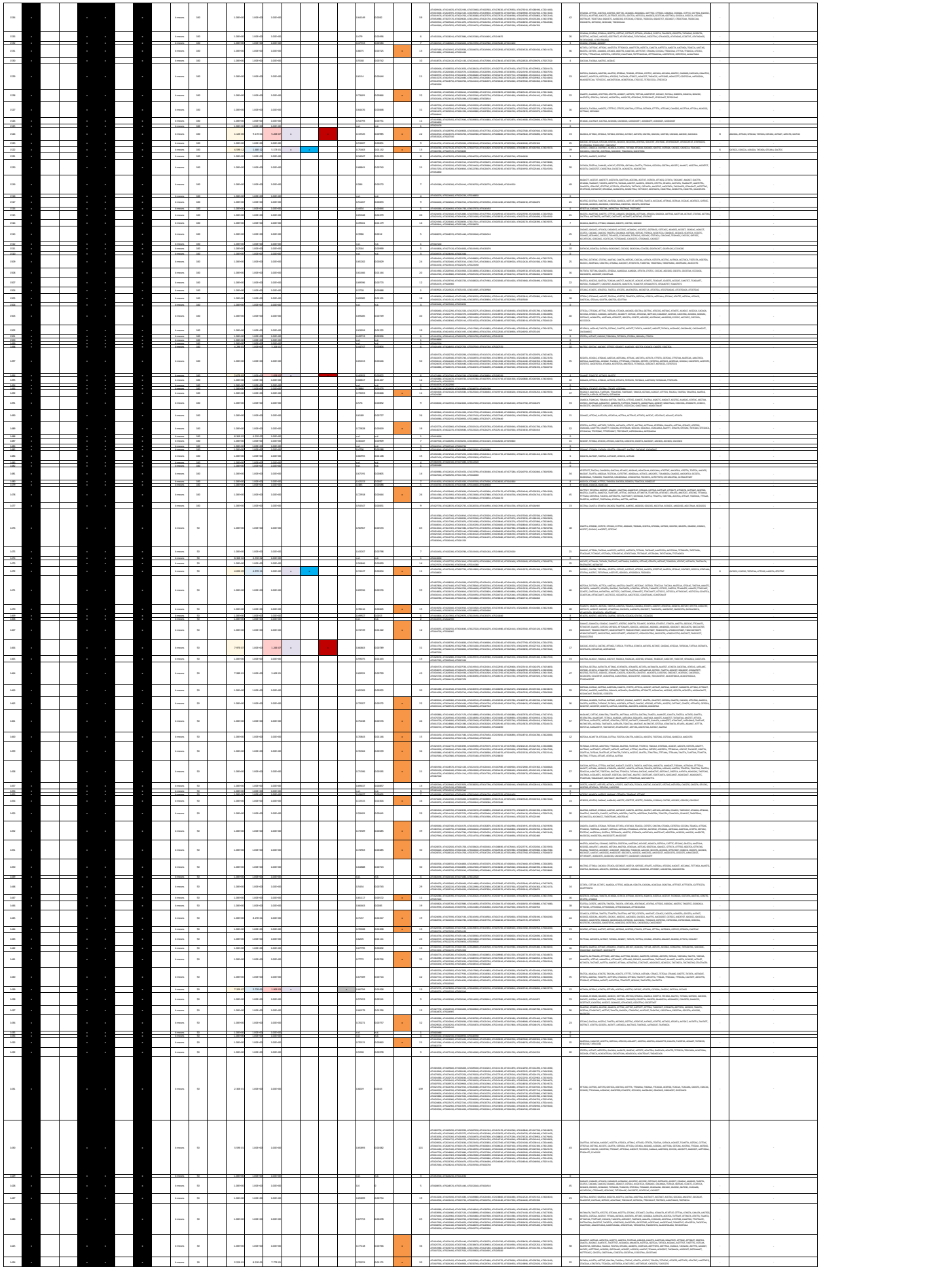
[1] Zimmerly et al. (2010), Science
[2] Kilian et al. 2007, The Plant Journal
[3] Hecker et al. (2007), Genome Biol.
[4] Schmitz et al. (2013), Nature Genetics

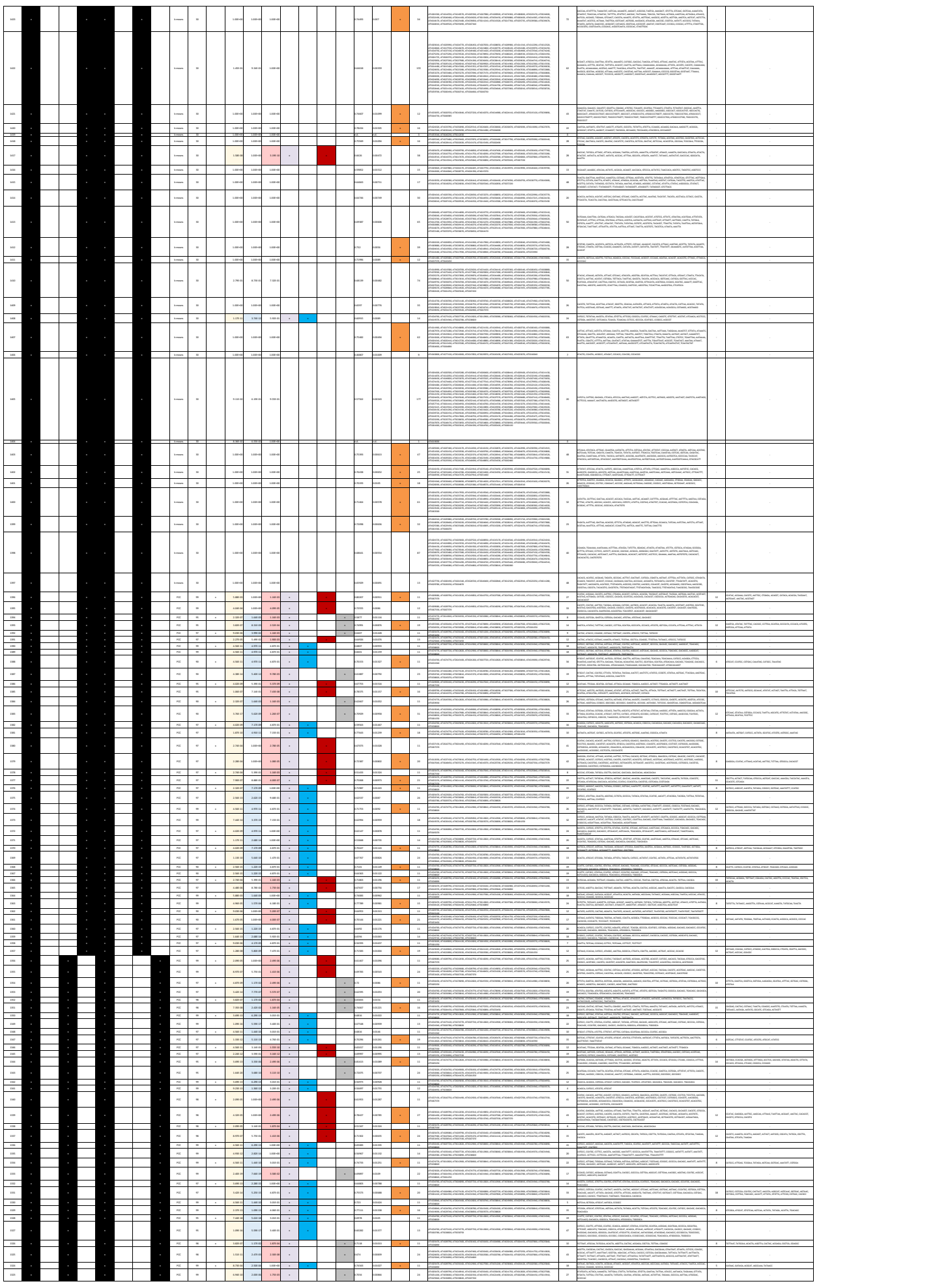
[5] Well performing clusters: pCSCs identified in these clusters are subject of our analyses. If FCC clusters contain <11 genes, a top10-version was generated with the top 10 best correlating genes to the query. iron-top10.0ng/l0.pat has lower consecutive IDs. 2159 GS-enriched clusters with 10, FF, 45 non-FF, 58 mixed, 658 non-enriched clusters).
[6] Clusters with FC >0.7 and only the cluster with higher F1 in a near-top10%pat0.pat.

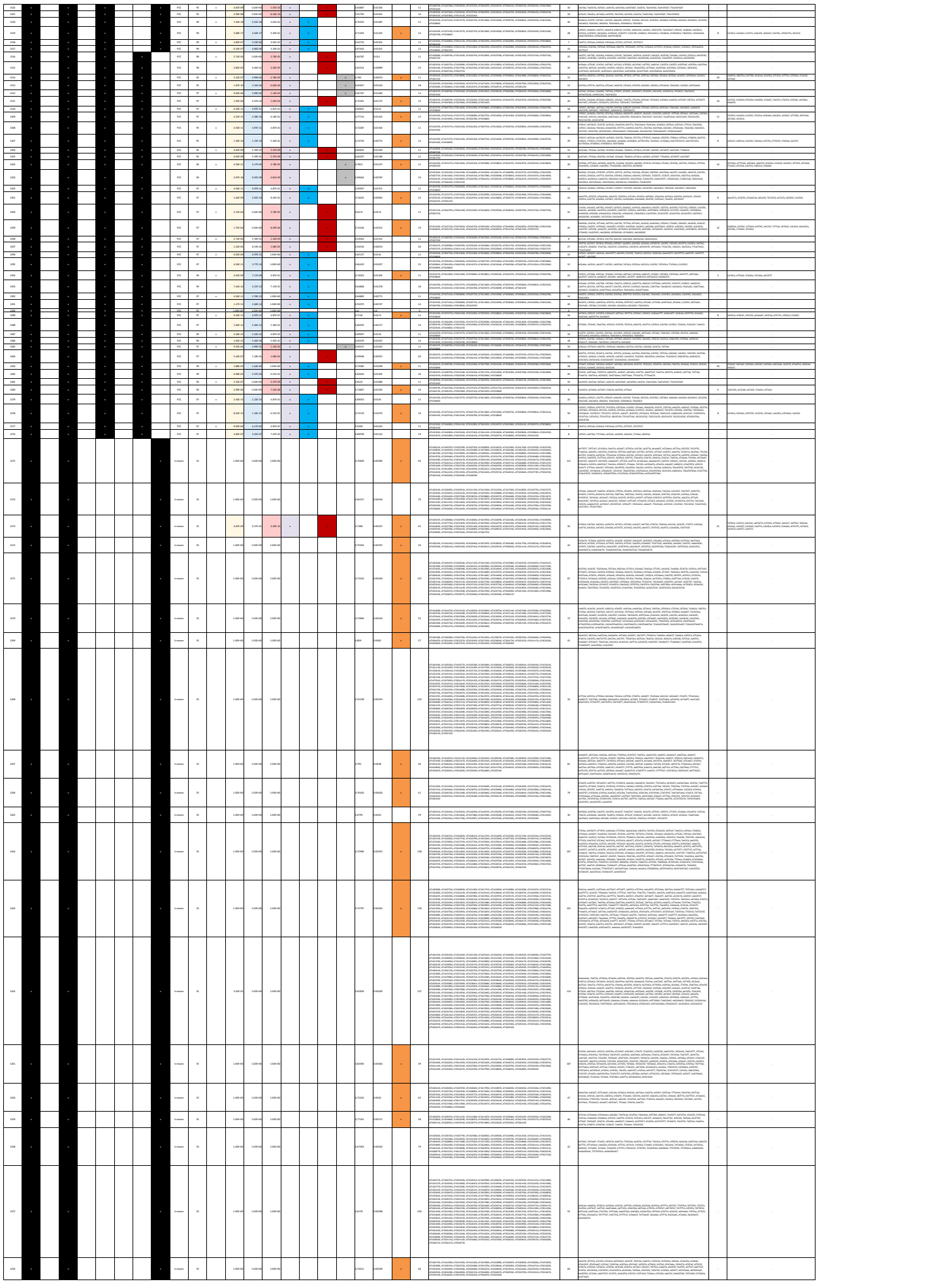
Cluster ID	Expression data					Input genes		Algorithm and parameters			FET		Category (FET, p<0.05)			RF		seed for pCSC creation[5]	# genes in cluster	Genes	pCSCs identified		# min-pCSCs (GS-enriched)	min-pCSCs
	Pol-type mutation	Pol-exit variant[1]	adaptive (ts)	abiotic (ex-genetic)	hormones[2]	developmental[3]	Pol-gp (high FC in GS)	Pol-gp (high FC in non-GS)	algorithm	parameters pre-k	mcl(I) (granularity)	p(GS genes)	p(FF genes)	p(non-FF genes)	GS-enriched	FCC-only	non-FCC only	mixed (FF & non-FF GS)	F1	E			# pCSCs found	pCSCs (ranked by importance)		
1950									PCC	90	x	7.12E-53	1.03E-07	1.47E-05					0.9972	0.01482		17	AT5G14750, AT1G07610, AT5G03840, AT5G07010, AT1G61520, AT1G61250, AT1G61240, AT1G07370, AT5G07090, AT5G07100, AT5G07110	AT5G14750, AT1G07610, AT5G03840, AT5G07010, AT1G61520, AT1G61250, AT1G61240, AT1G07370, AT5G07090, AT5G07100, AT5G07110	
1958	x		x		x		x		PCC	99		1.00E+00	1.00E+00	1.00E+00					0.97756	0.00352		76	AT4G34240, AT1G07370, AT5G03840, AT5G07010, AT1G61520, AT1G61250, AT1G61240, AT1G07370, AT5G07090, AT5G07100, AT5G07110, AT5G07090, AT5G07100, AT5G07110, AT5G07090, AT5G07100, AT5G07110



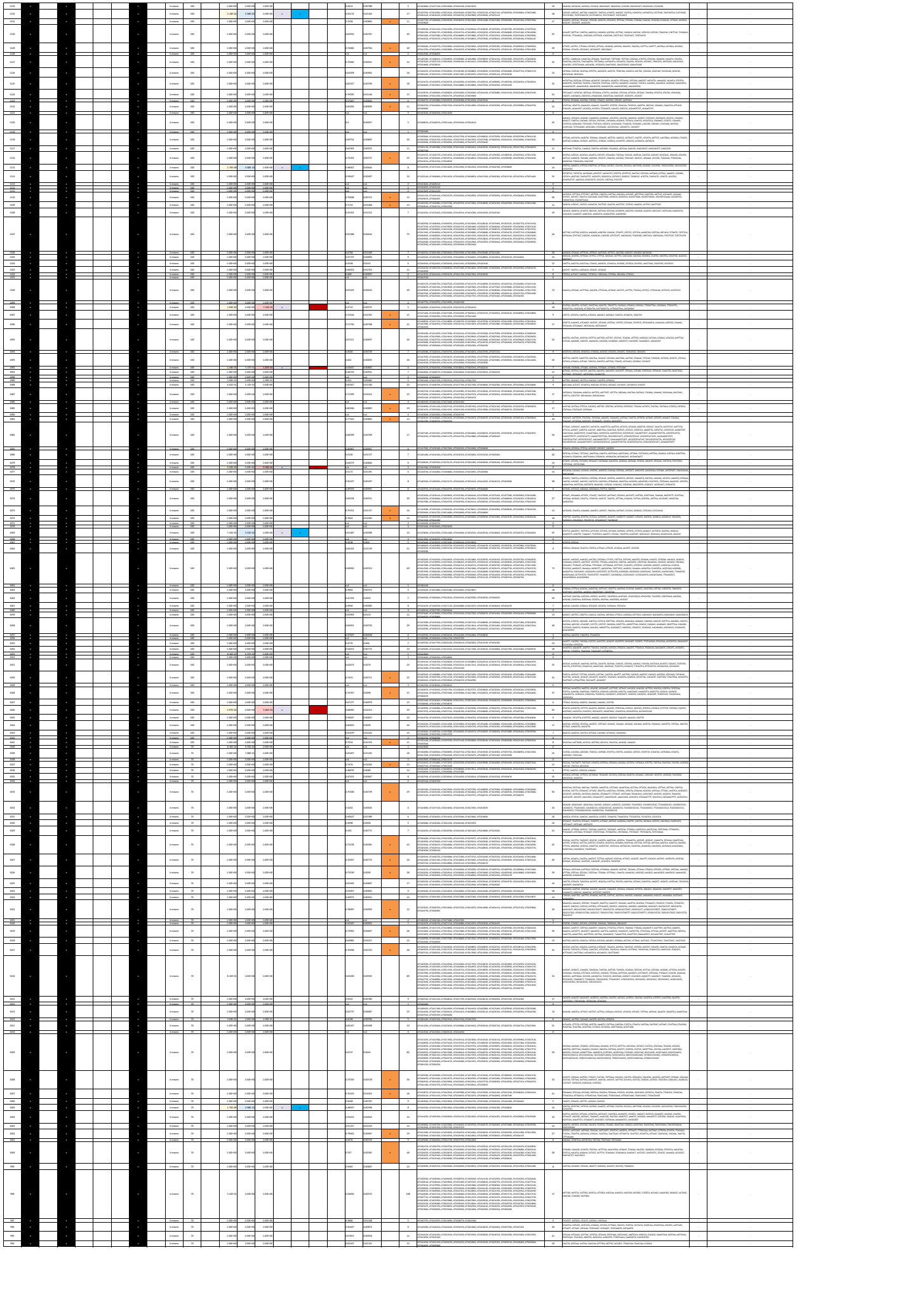


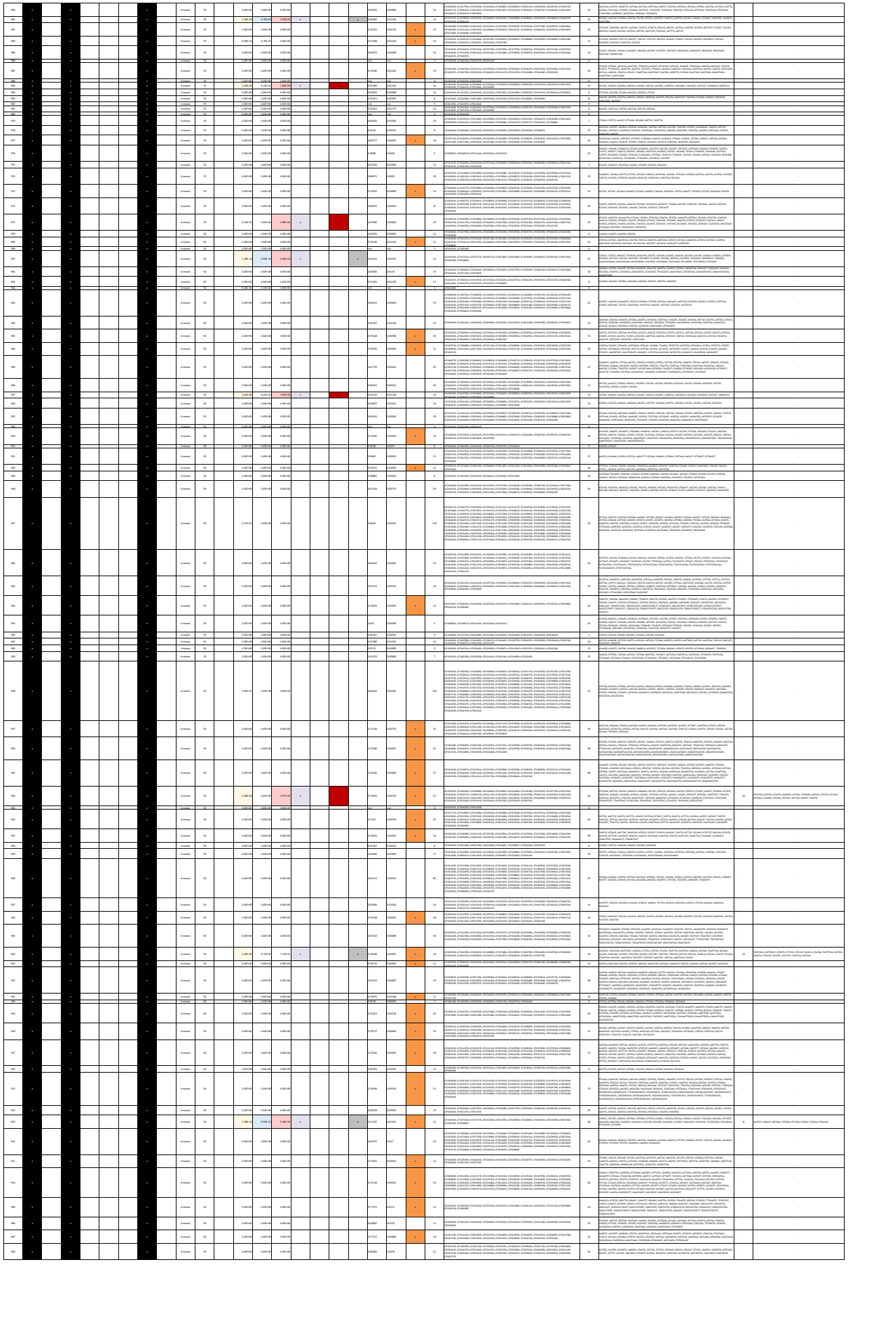


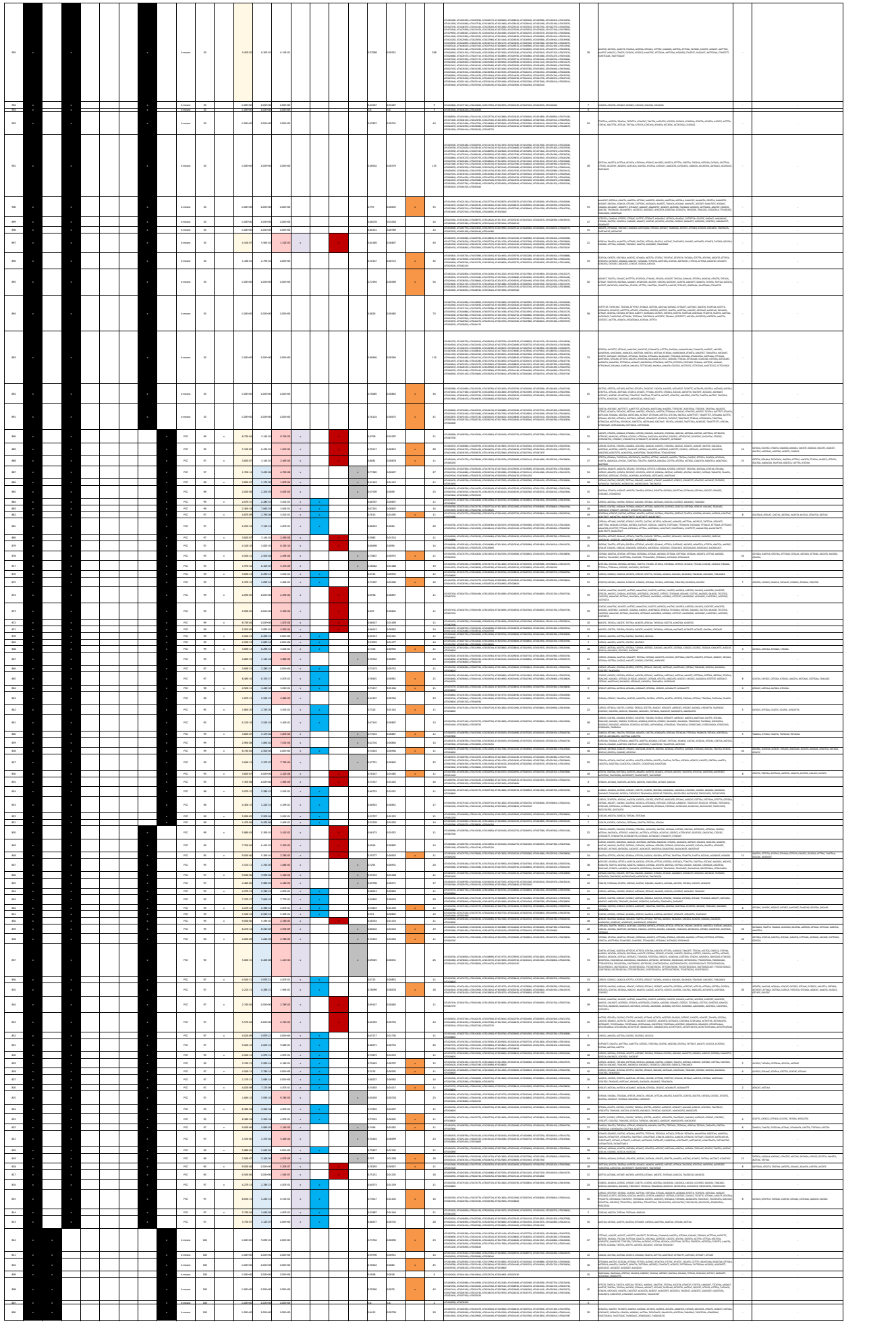


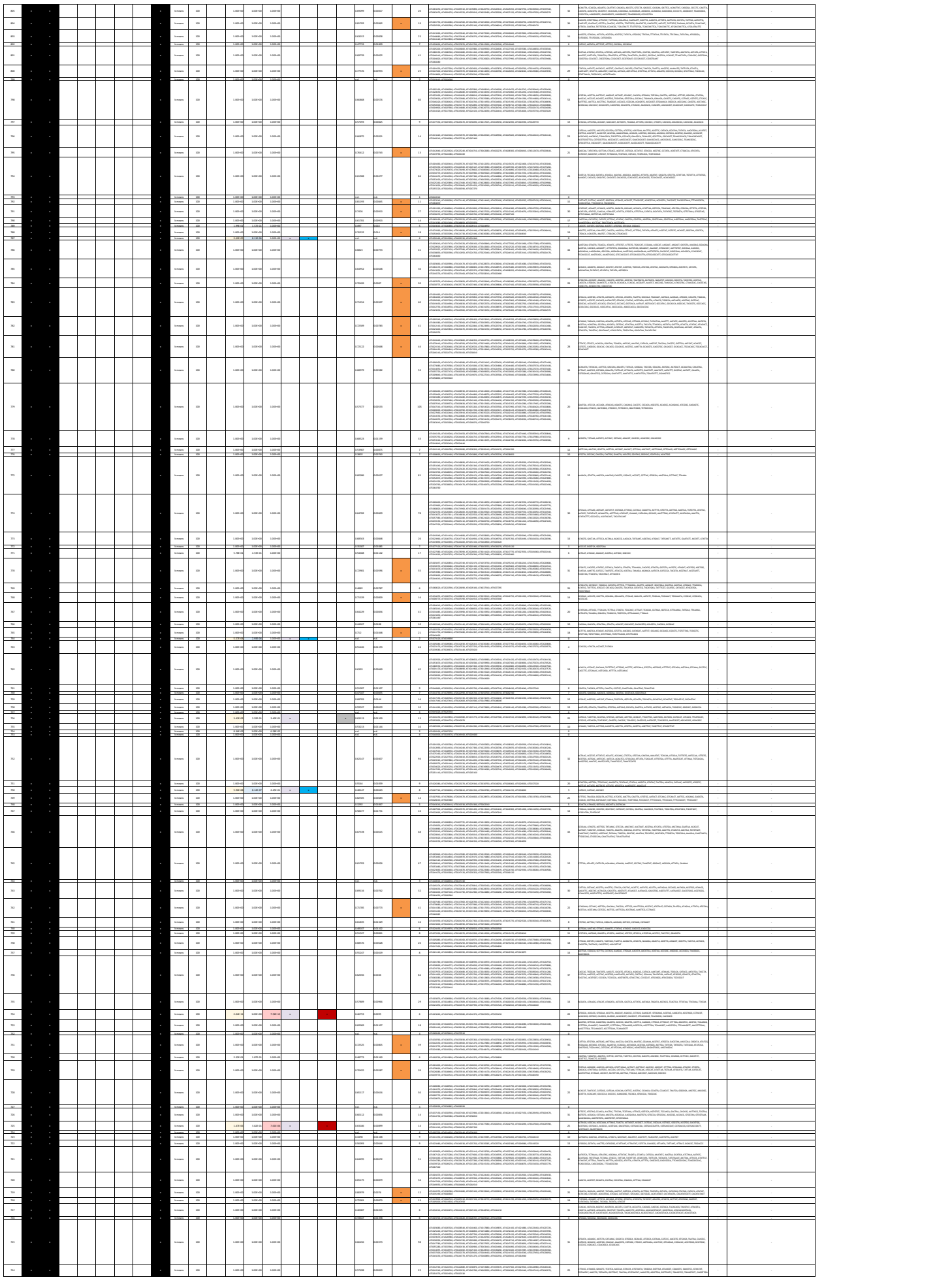


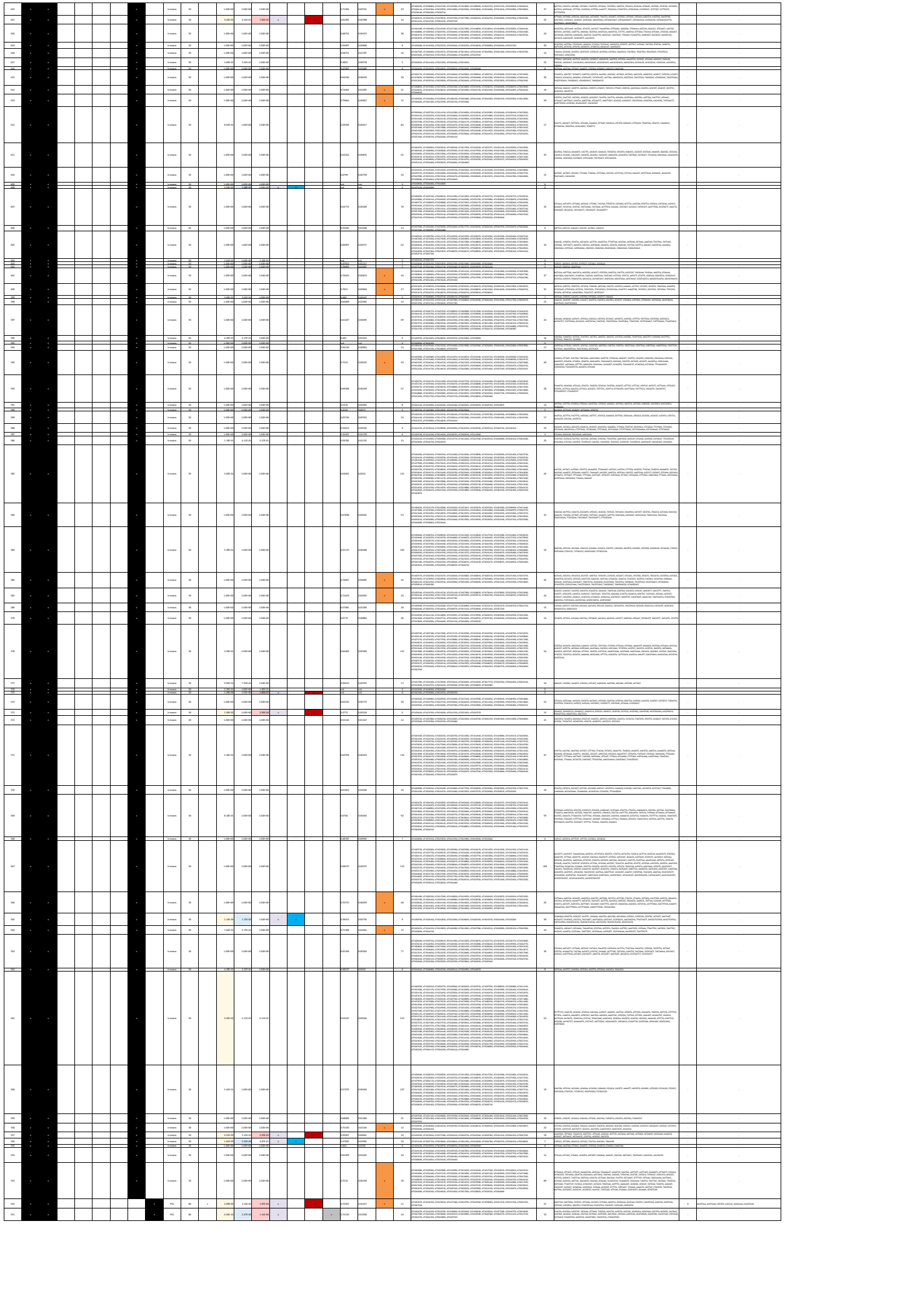
| | | |

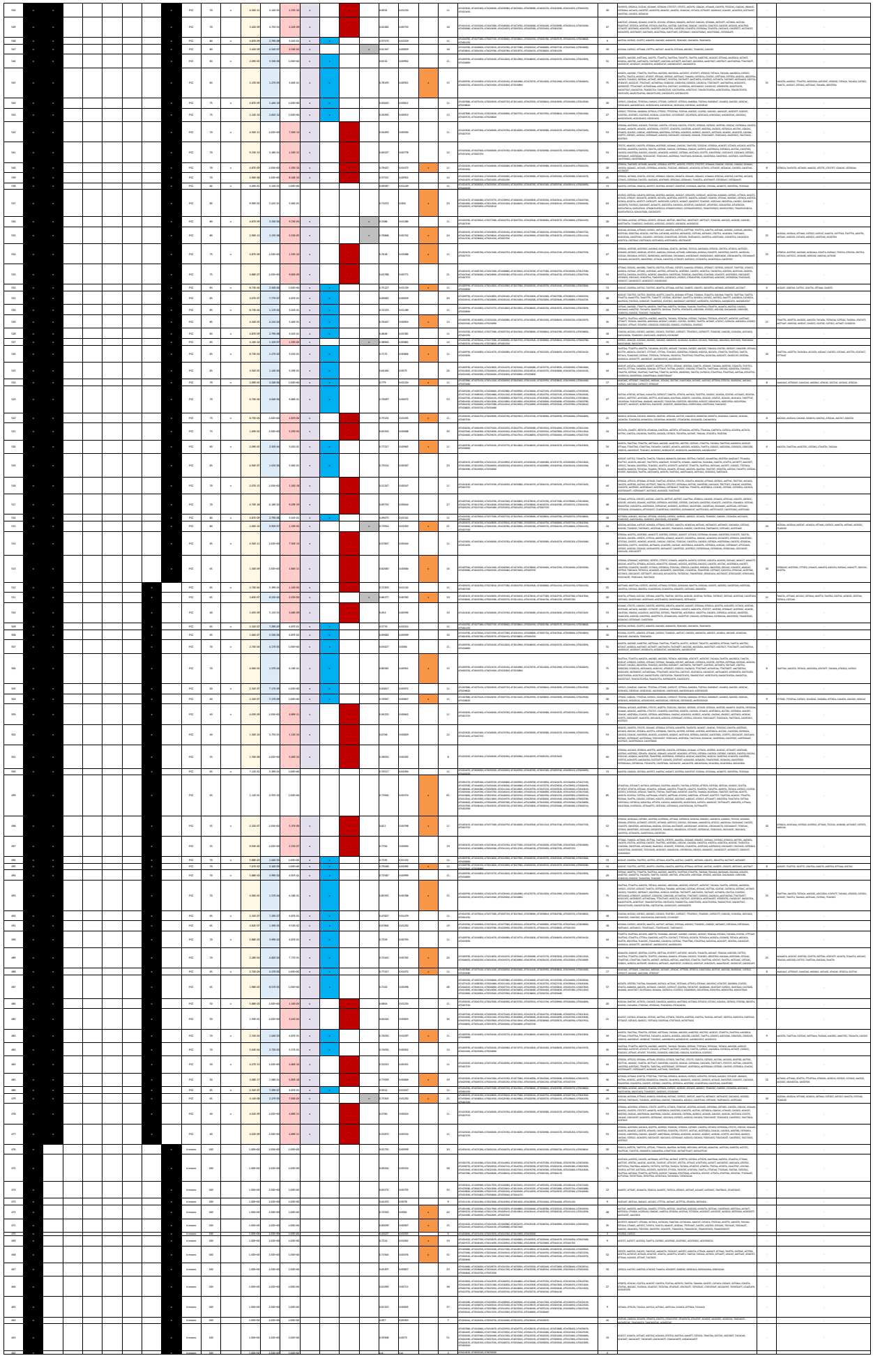


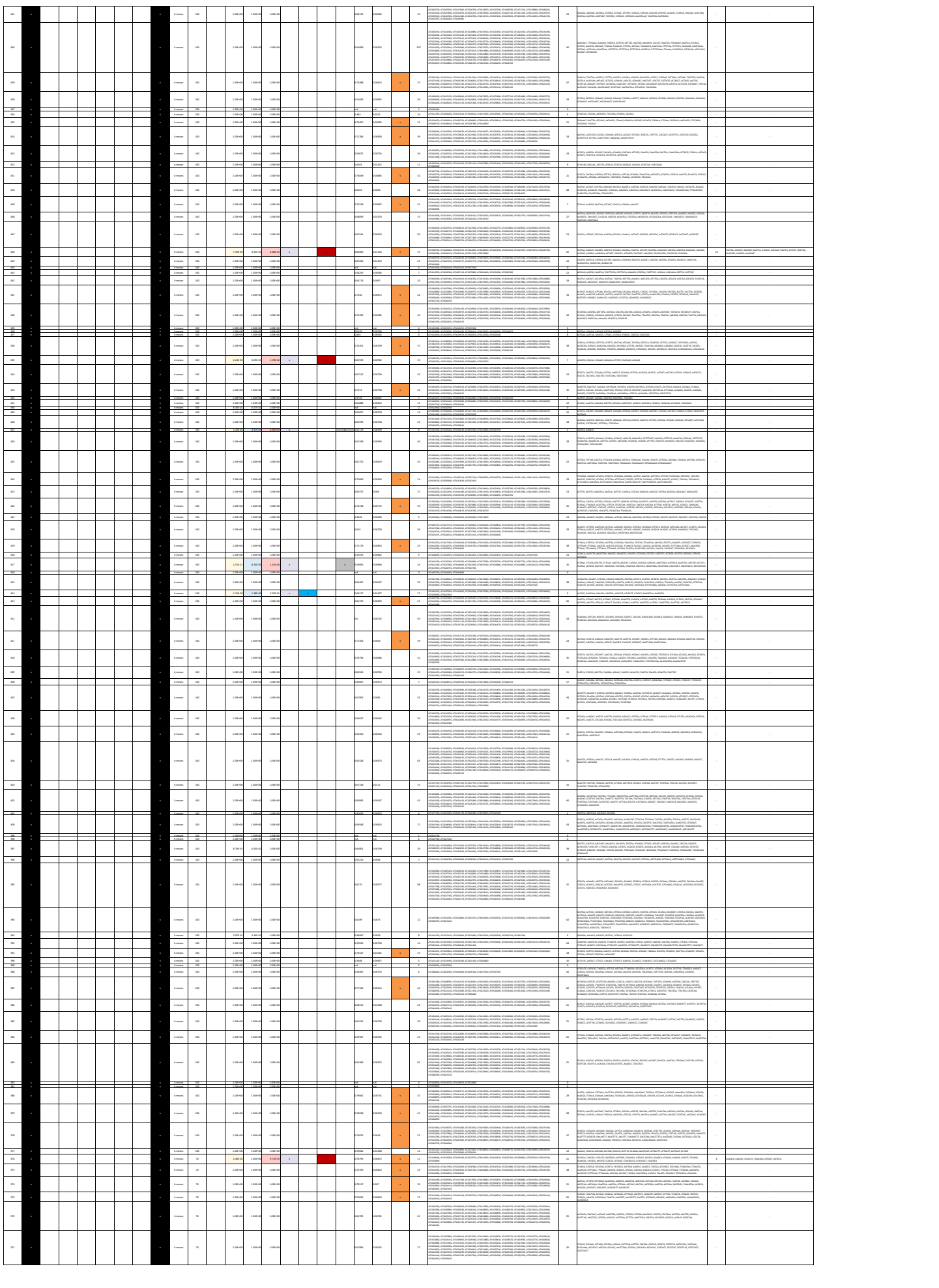


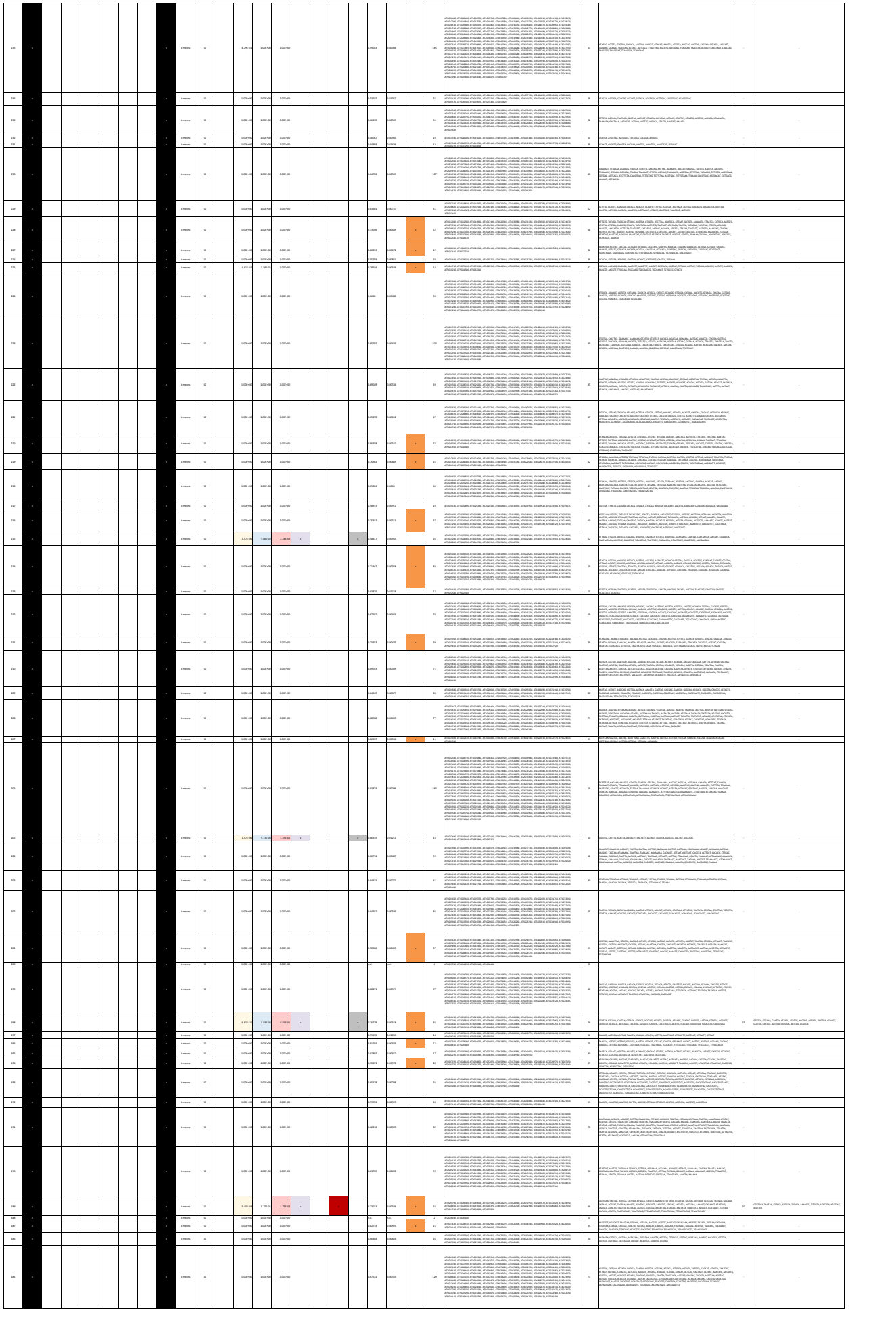


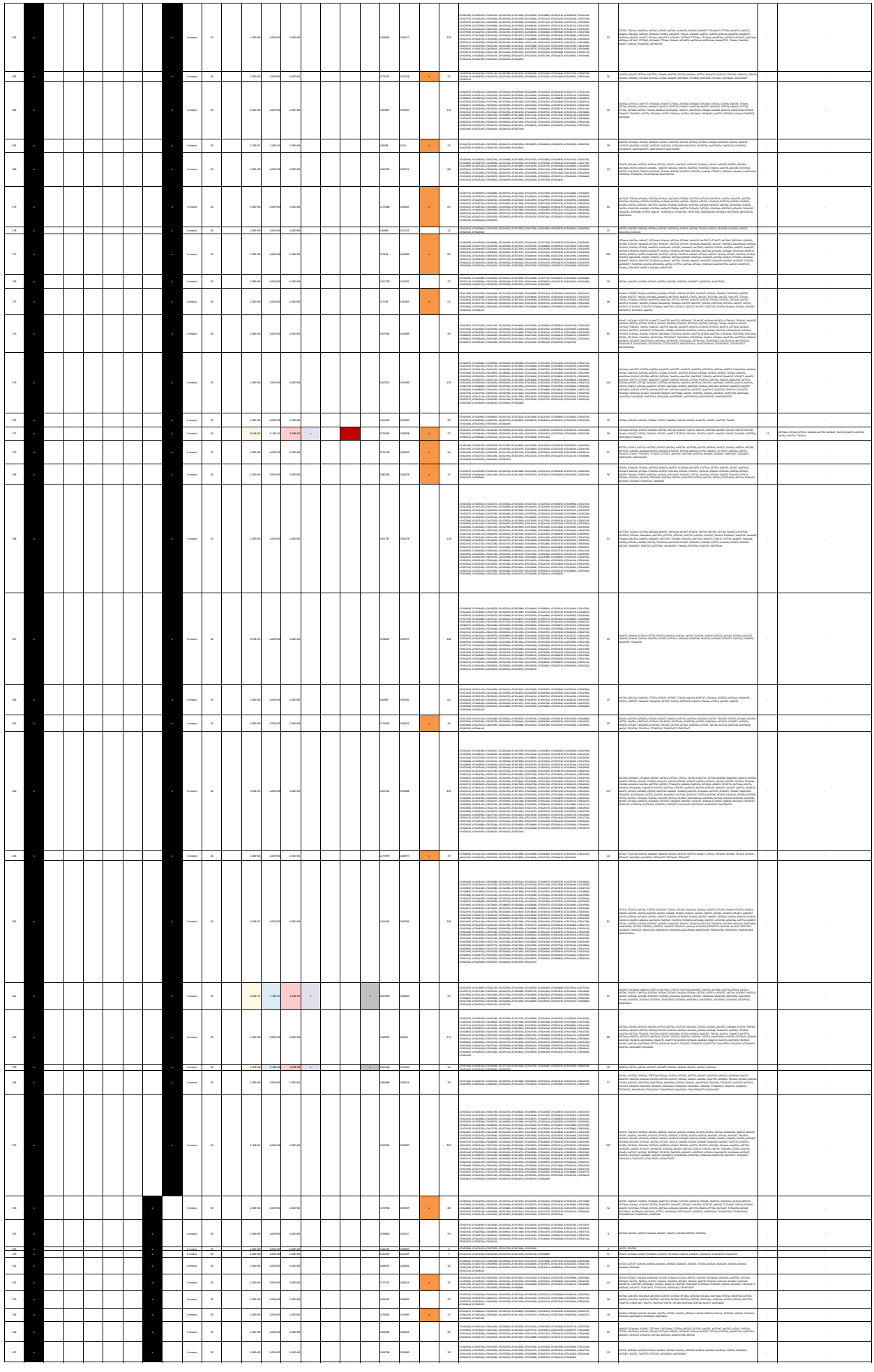


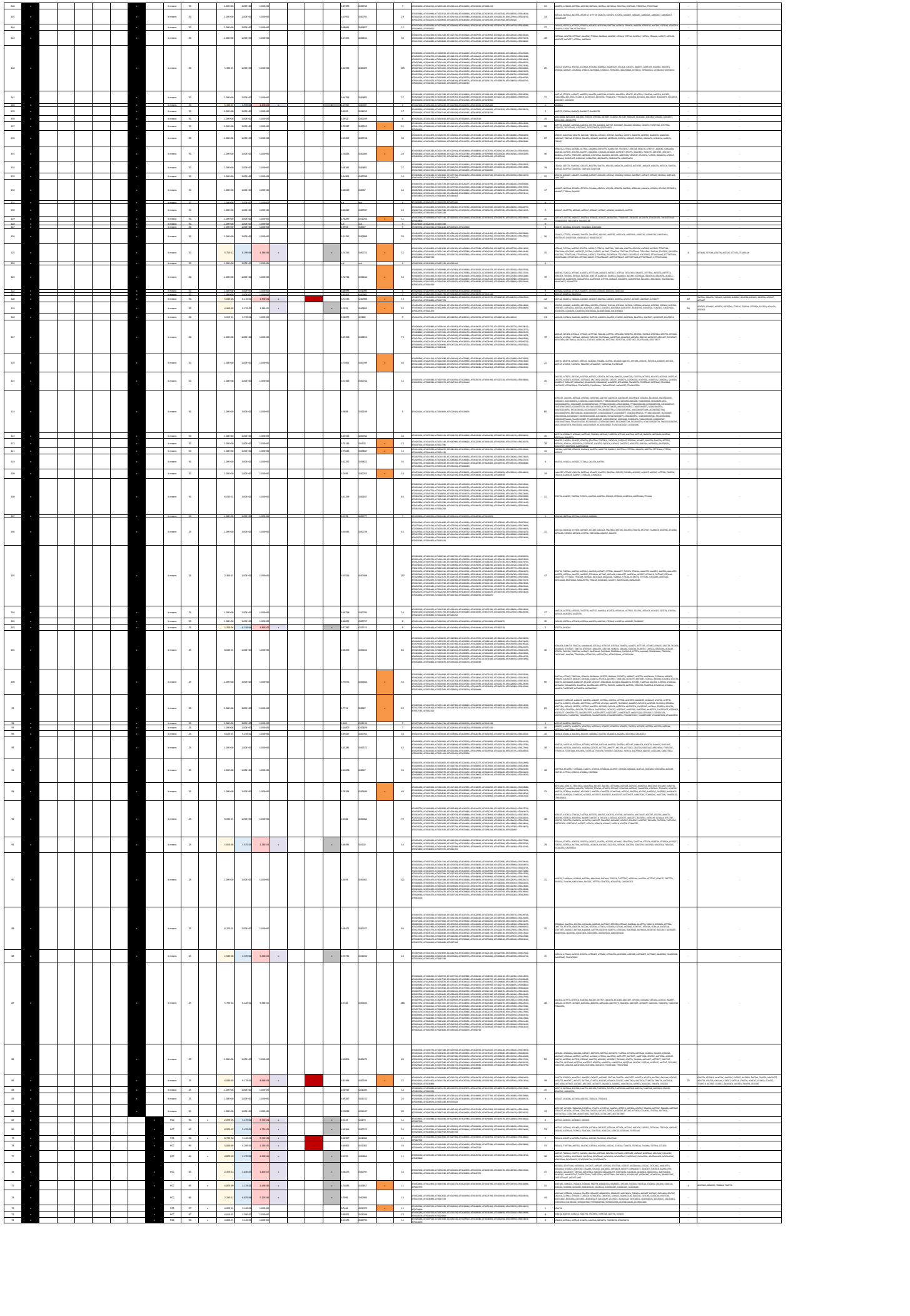








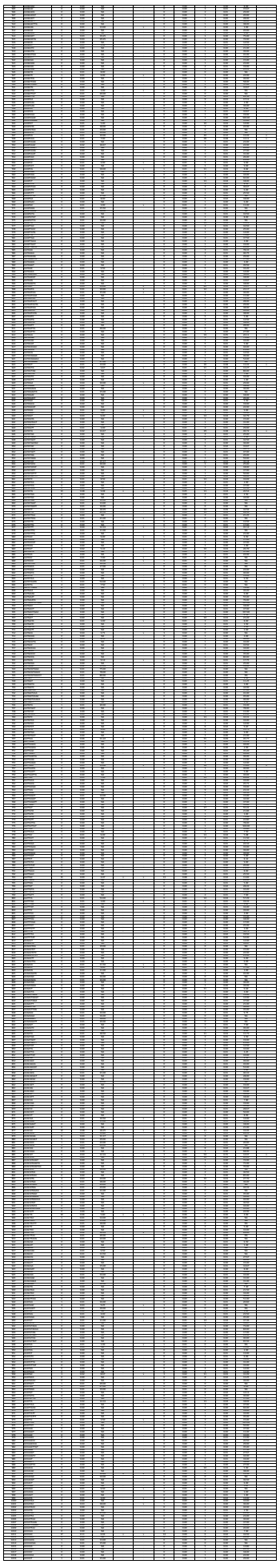


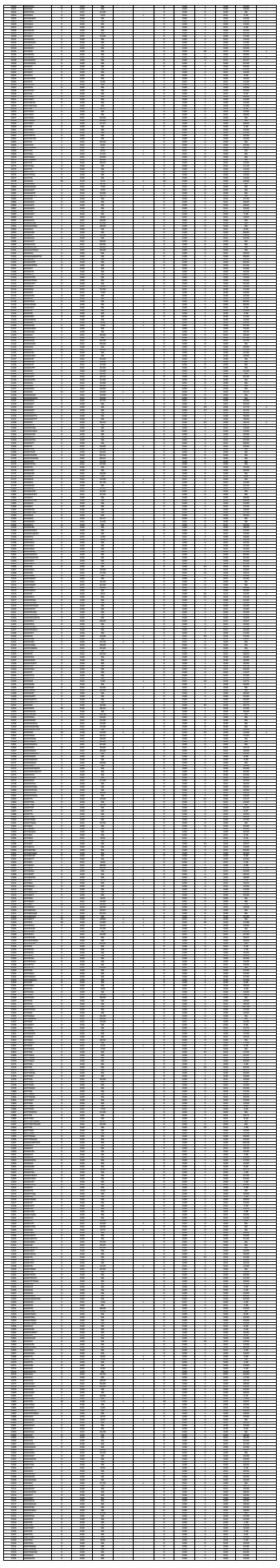


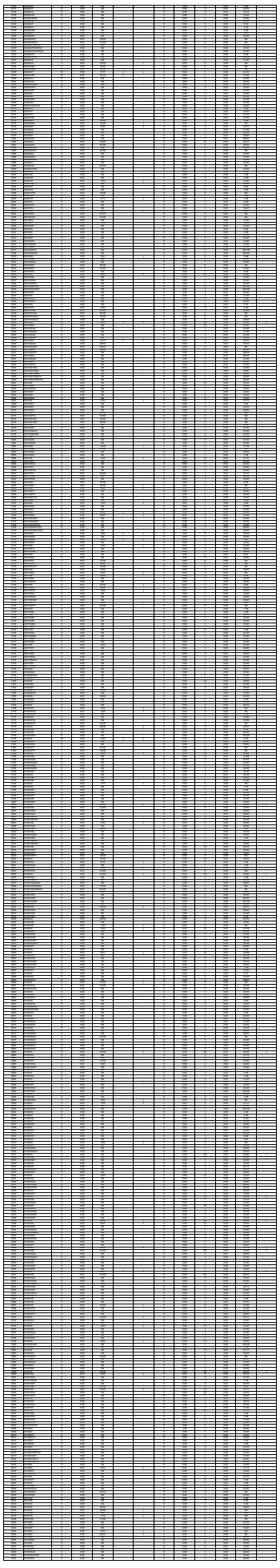
| | | |

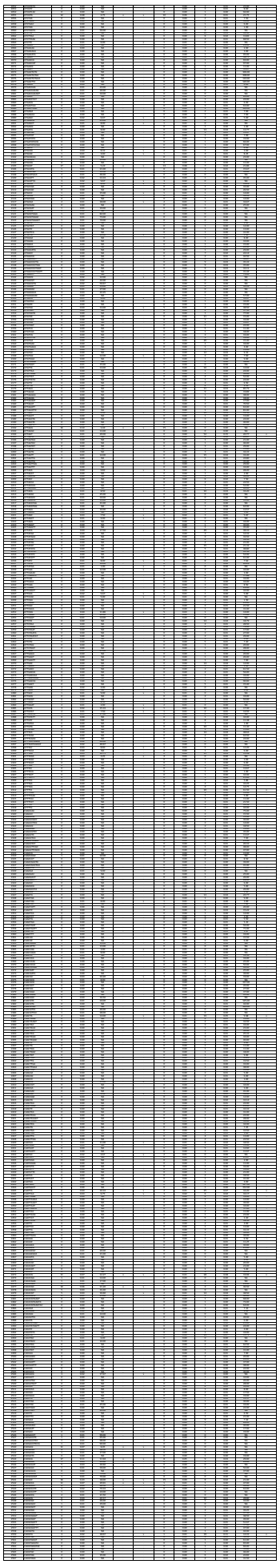
Table S1. All pCR1 identified in well performing models (ES-enriched and non-enriched). Absolute and proportional counts of pCR1 within ES-enriched clusters and non-enriched clusters are provided. Also provided are the number of responders and non-responders within each pCR1 and the number of pCR1 within each cluster. Asterisk (*) marked indicates that the pCR1 was observed more important for prediction of its response outcome. NA= is and because pCR1 can be defined in this cluster category. Group pCR: pCR1 among those which were identified in >5% of the ES_enriched or non-enriched clusters. min. pCR: pCR1 was among the minimum cut-off pCR1 needed for good model performance. Cluster Membership: Clusters in which pCR1 were found in. Table S2: to correlate to values in **Figure 5A**. *To see the most relevant results, sort Note by min.pCR (enriched) Asterisk, Freq.pCR (1 to 4) or min.pCR (1 to 4).*

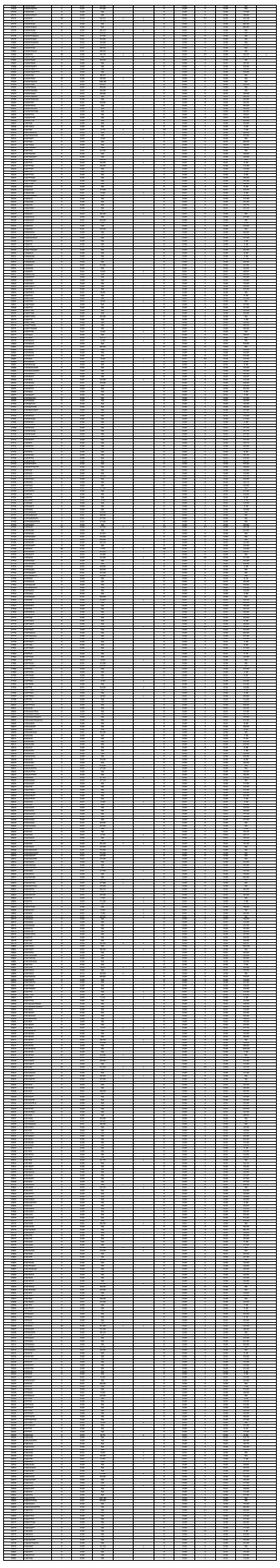
Patient.barcode	pCR1	ES-enriched Clusters (Total: 1000)								Non-Enriched Clusters (Total: 1000)			
		Freq.pCR	Proportion	Group.Size	Proportion	min.pCR	Cluster	Proportion	Group.Size	Freq.pCR	Proportion	Group.Size	Proportion
1	TCGA-A2-A0CM	1	0.001	1000	NA					1	0.001	1000	NA
2	TCGA-A2-A0D0	1	0.001	1000	NA					1	0.001	1000	NA
3	TCGA-A2-A0D2	1	0.001	1000	NA					1	0.001	1000	NA
4	TCGA-A2-A0EM	1	0.001	1000	NA					1	0.001	1000	NA
5	TCGA-A2-A0EN	1	0.001	1000	NA					1	0.001	1000	NA
6	TCGA-A2-A0EQ	1	0.001	1000	NA					1	0.001	1000	NA
7	TCGA-A2-A0ES	1	0.001	1000	NA					1	0.001	1000	NA
8	TCGA-A2-A0EV	1	0.001	1000	NA					1	0.001	1000	NA
9	TCGA-A2-A0EW	1	0.001	1000	NA					1	0.001	1000	NA
10	TCGA-A2-A0EX	1	0.001	1000	NA					1	0.001	1000	NA
11	TCGA-A2-A0EY	1	0.001	1000	NA					1	0.001	1000	NA
12	TCGA-A2-A0SW	1	0.001	1000	NA					1	0.001	1000	NA
13	TCGA-A2-A0SX	1	0.001	1000	NA					1	0.001	1000	NA
14	TCGA-A2-A0T0	1	0.001	1000	NA					1	0.001	1000	NA
15	TCGA-A2-A0T1	1	0.001	1000	NA					1	0.001	1000	NA
16	TCGA-A2-A0T2	1	0.001	1000	NA					1	0.001	1000	NA
17	TCGA-A2-A0T3	1	0.001	1000	NA					1	0.001	1000	NA
18	TCGA-A2-A0T4	1	0.001	1000	NA					1	0.001	1000	NA
19	TCGA-A2-A0T5	1	0.001	1000	NA					1	0.001	1000	NA
20	TCGA-A2-A0T6	1	0.001	1000	NA					1	0.001	1000	NA
21	TCGA-A2-A0T7	1	0.001	1000	NA					1	0.001	1000	NA
22	TCGA-A2-A0YC	1	0.001	1000	NA					1	0.001	1000	NA
23	TCGA-A2-A0YD	1	0.001	1000	NA					1	0.001	1000	NA
24	TCGA-A2-A0YE	1	0.001	1000	NA					1	0.001	1000	NA
25	TCGA-A2-A0YF	1	0.001	1000	NA					1	0.001	1000	NA
26	TCGA-A2-A0YG	1	0.001	1000	NA					1	0.001	1000	NA
27	TCGA-A2-A0YH	1	0.001	1000	NA					1	0.001	1000	NA
28	TCGA-A2-A0YI	1	0.001	1000	NA					1	0.001	1000	NA
29	TCGA-A2-A0YJ	1	0.001	1000	NA					1	0.001	1000	NA
30	TCGA-A2-A0YK	1	0.001	1000	NA					1	0.001	1000	NA
31	TCGA-A2-A0YL	1	0.001	1000	NA					1	0.001	1000	NA
32	TCGA-A2-A0YM	1	0.001	1000	NA					1	0.001	1000	NA
33	TCGA-A2-A0YT	1	0.001	1000	NA					1	0.001	1000	NA
34	TCGA-A2-A1FV	1	0.001	1000	NA					1	0.001	1000	NA
35	TCGA-A2-A1FW	1	0.001	1000	NA					1	0.001	1000	NA
36	TCGA-A2-A1FX	1	0.001	1000	NA					1	0.001	1000	NA
37	TCGA-A2-A1FZ	1	0.001	1000	NA					1	0.001	1000	NA
38	TCGA-A2-A1G0	1	0.001	1000	NA					1	0.001	1000	NA
39	TCGA-A2-A1G1	1	0.001	1000	NA					1	0.001	1000	NA
40	TCGA-A2-A1G4	1	0.001	1000	NA					1	0.001	1000	NA
41	TCGA-A2-A1G5	1	0.001	1000	NA					1	0.001	1000	NA
42	TCGA-A2-A1G6	1	0.001	1000	NA					1	0.001	1000	NA
43	TCGA-A2-A1G7	1	0.001	1000	NA					1	0.001	1000	NA
44	TCGA-A2-A1G9	1	0.001	1000	NA					1	0.001	1000	NA
45	TCGA-A2-A1GB	1	0.001	1000	NA					1	0.001	1000	NA
46	TCGA-A2-A1GC	1	0.001	1000	NA					1	0.001	1000	NA
47	TCGA-A2-A1GD	1	0.001	1000	NA					1	0.001	1000	NA
48	TCGA-A2-A1GE	1	0.001	1000	NA					1	0.001	1000	NA
49	TCGA-A2-A1GF	1	0.001	1000	NA					1	0.001	1000	NA
50	TCGA-A2-A1GH	1	0.001	1000	NA					1	0.001	1000	NA
51	TCGA-A2-A1GI	1	0.001	1000	NA					1	0.001	1000	NA
52	TCGA-A2-A1GJ	1	0.001	1000	NA					1	0.001	1000	NA
53	TCGA-A2-A1GK	1	0.001	1000	NA					1	0.001	1000	NA
54	TCGA-A2-A1GN	1	0.001	1000	NA					1	0.001	1000	NA
55	TCGA-A2-A1GP	1	0.001	1000	NA					1	0.001	1000	NA
56	TCGA-A2-A1GQ	1	0.001	1000	NA					1	0.001	1000	NA
57	TCGA-A2-A1GR	1	0.001	1000	NA					1	0.001	1000	NA
58	TCGA-A2-A1GS	1	0.001	1000	NA					1	0.001	1000	NA
59	TCGA-A2-A1GU	1	0.001	1000	NA					1	0.001	1000	NA
60	TCGA-A2-A1GW	1	0.001	1000	NA					1	0.001	1000	NA
61	TCGA-A2-A1GX	1	0.001	1000	NA					1	0.001	1000	NA
62	TCGA-A2-A1GY	1	0.001	1000	NA					1	0.001	1000	NA
63	TCGA-A2-A1GZ	1	0.001	1000	NA					1	0.001	1000	NA
64	TCGA-A2-A1H0	1	0.001	1000	NA					1	0.001	1000	NA
65	TCGA-A2-A1H1	1	0.001	1000	NA					1	0.001	1000	NA
66	TCGA-A2-A1H2	1	0.001	1000	NA					1	0.001	1000	NA
67	TCGA-A2-A1H3	1	0.001	1000	NA					1	0.001	1000	NA
68	TCGA-A2-A1H4	1	0.001	1000	NA					1	0.001	1000	NA
69	TCGA-A2-A1H5	1	0.001	1000	NA					1	0.001	1000	NA
70	TCGA-A2-A1H6	1	0.001	1000	NA					1	0.001	1000	NA
71	TCGA-A2-A1H7	1	0.001	1000	NA					1	0.001	1000	NA
72	TCGA-A2-A1H8	1	0.001	1000	NA					1	0.001	1000	NA
73	TCGA-A2-A1H9	1	0.001	1000	NA					1	0.001	1000	NA
74	TCGA-A2-A1HA	1	0.001	1000	NA					1	0.001	1000	NA
75	TCGA-A2-A1HB	1	0.001	1000	NA					1	0.001	1000	NA
76	TCGA-A2-A1HC	1	0.001	1000	NA					1	0.001	1000	NA
77	TCGA-A2-A1HD	1	0.001	1000	NA					1	0.001	1000	NA
78	TCGA-A2-A1HE	1	0.001	1000	NA					1	0.001	1000	NA
79	TCGA-A2-A1HF	1	0.001	1000	NA					1	0.001	1000	NA
80	TCGA-A2-A1HG	1	0.001	1000	NA					1	0.001	1000	NA
81	TCGA-A2-A1HH	1	0.001	1000	NA					1	0.001	1000	NA
82	TCGA-A2-A1HI	1	0.001	1000	NA					1	0.001	1000	NA
83	TCGA-A2-A1HJ	1	0.001	1000	NA					1	0.001	1000	NA
84	TCGA-A2-A1HK	1	0.001	1000	NA					1	0.001	1000	NA
85	TCGA-A2-A1HL	1	0.001	1000	NA					1	0.001	1000	NA
86	TCGA-A2-A1HM	1	0.001	1000	NA					1	0.001	1000	NA
87	TCGA-A2-A1HN	1	0.001	1000	NA					1	0.001	1000	NA
88	TCGA-A2-A1HO	1	0.001	1000	NA					1	0.001	1000	NA
89	TCGA-A2-A1HP	1	0.001	1000	NA					1	0.001	1000	NA
90	TCGA-A2-A1HQ	1	0.001	1000	NA					1	0.001	1000	NA
91	TCGA-A2-A1HR	1	0.001	1000	NA					1	0.001	1000	NA
92	TCGA-A2-A1HS	1	0.001	1000	NA					1	0.001	1000	NA
93	TCGA-A2-A1HT	1	0.001	1000	NA					1	0.001	1000	NA
94	TCGA-A2-A1HU	1	0.001	1000	NA					1	0.001	1000	NA
95	TCGA-A2-A1HV	1	0.001	1000	NA					1	0.001	1000	NA
96	TCGA-A2-A1HW	1	0.001	1000	NA					1	0.001	1000	NA
97	TCGA-A2-A1HX	1	0.001	1000	NA					1	0.001	1000	NA
98	TCGA-A2-A1HY	1	0.001	1000	NA					1	0.001	1000	NA
99	TCGA-A2-A1HZ	1	0.001	1000	NA					1	0.001	1000	NA
100	TCGA-A2-A1I0	1	0.001	1000	NA					1	0.001	1000	NA
101	TCGA-A2-A1I1	1	0.001	1000	NA					1	0.001	1000	NA
102	TCGA-A2-A1I2	1	0.001	1000	NA					1	0.001	1000	NA
103	TCGA-A2-A1I3	1	0.001	1000	NA					1	0.001	1000	NA
104	TCGA-A2-A1I4	1	0.001	1000	NA					1	0.001	1000	NA
105	TCGA-A2-A1I5	1	0.001	1000	NA					1	0.001	1000	NA
106	TCGA-A2-A1I6	1	0.001	1000	NA					1	0.001	1000	NA
107	TCGA-A2-A1I7	1	0.001	1000	NA					1	0.001	1000	NA
108	TCGA-A2-A1I8	1	0.001	1000	NA					1	0.001	1000	NA
109	TCGA-A2-A1I9	1	0.001	1000	NA					1	0.001	1000	NA
110	TCGA-A2-A1IA	1	0.001	1000	NA					1	0.001	1000	NA
111	TCGA-A2-A1IB	1	0.001	1000	NA					1	0.001	1000	NA
112	TCGA-A2-A1IC	1	0.001	1000	NA					1	0.001	1000	NA
113	TCGA-A2-A1ID	1	0.001	1000	NA					1	0.001	1000	NA
114	TCGA-A2-A1IE	1	0.001	1000	NA					1	0.001	1000	NA
115	TCGA-A2-A1IF	1	0.001	1000	NA					1	0.001	1000	NA
116	TCGA-A2-A1IG	1	0.001	1000	NA					1	0.001	1000	NA
117	TCGA-A2-A1IH	1	0.001	1000	NA					1	0.001	1000	NA
118	TCGA-A2-A1II	1	0.001	1000	NA					1	0.001	1000	NA
119	TCGA-A2-A1IJ	1	0.001	1000	NA					1	0.001	1000	NA
120	TCGA-A2-A1IK	1	0.001	1000	NA					1	0.001	1000	NA
121	TCGA-A2-A1IL	1	0.001	1000	NA					1	0.001	1000	NA
122	TCGA-A2-A1IM	1	0.001	1000	NA					1	0.001	1000	NA
123	TCGA-A2-A1IN	1	0.001	1000	NA					1	0.001	1000	NA
124	TCGA-A2-A1IO	1	0.001	1000	NA					1	0.001	1000	NA
125	TCGA-A2-A1IP	1	0.001	1000	NA					1	0.001	1000	NA
126	TCGA-A2-A1IQ	1	0.001	1000	NA					1	0.001	1000	NA
127	TCGA-A2-A1IR	1	0.001	1000	NA					1	0.001	1000	NA
128	TCGA-A2-A1IS	1	0.001	1000	NA					1	0.001	1000	NA
129	TCGA-A2-A1IT	1	0.001	1000	NA					1	0.001	1000	NA
130	TCGA-A2-A1IU	1	0.001	1000	NA					1	0.001	1000	NA
131	TCGA-A2-A1IV	1	0.001	1000	NA					1	0.001	1000	NA
132	TCGA-A2-A1IW	1	0.001	1000	NA					1	0.001	1000	NA
133	TCGA-A2-A1IX	1	0.001	1000	NA					1	0.001	1000	NA
134	TCGA-A2-A1IY	1	0.001	1000	NA					1	0.001	1000	NA
135	TCGA-A2-A1IZ	1	0.001	1000	NA					1	0.001	1000	NA
136	TCGA-A2-A1J0	1	0.001	1000	NA					1	0.001	1000	NA
137	TCGA-A2-A1J1	1	0.001	1000	NA					1	0.001	1000	NA
138	TCGA-A2-A1J2	1	0.001	1000	NA					1	0.001	1000	NA
139	TCGA-A2-A1J3	1	0.001	1000	NA					1	0.001	1000	NA
140	TCGA-A2-A1J4	1	0.001	1000	NA					1	0.001	1000	NA
141	TCGA-A2-A1J5	1	0.001	1000	NA					1	0.001	1000	NA
142	TCGA-A2-A1J6	1	0.001	1000	NA					1	0.001	1000	NA
143	TCGA-A2-A1J7	1	0.001	1000	NA					1	0.001	1000	NA
144	TCGA-A2-A1J8	1	0.001	1000	NA					1	0.001	1000	NA
145	TCGA-A2-A1J9	1	0.001	1000	NA					1	0.001	1000	NA
146	TCGA-A2-A1JA	1	0.001	1000	NA					1	0.001	1000	NA
147	TCGA-A2-A1JB	1	0.001	1000	NA					1	0.001	1000	NA
148	TCGA-A2-A1JC	1	0.001	1000	NA					1	0.001	1000	NA
149	TCGA-A2-A1JD	1	0.001	1000	NA					1	0.001	1000	NA
150	TCGA-A2-A1JE	1	0.001	1000	NA					1	0.001	1000	NA
151	TCGA-A2-A1JF	1	0.001	1000	NA					1	0.001	1000	NA
152	TCGA-A2-A1JG	1	0.001	1000	NA					1	0.001	1000	NA
153	TCGA-A2-A1JH	1	0.001	1000	NA					1	0.001	1000	NA
154	TCGA-A2-A1JI	1	0.001	1000	NA					1	0.001	1000	NA
155	TCGA-A2-A1JJ	1	0.001	1000	NA					1	0.001	1000	NA
156	TCGA-A2-A1JK	1	0.001	1000	NA					1	0.001	1000	NA
157	TCGA-A2-A1JL	1	0.001	1000	NA					1	0.001	1000	NA
158	TCGA-A2-A1JM	1	0.001	1000	NA					1	0.001	1000	NA
159	TCGA-A2-A1JN	1	0.001	1000	NA					1	0.001	1000	NA
160	TCGA-A2-A1JO	1	0.001	1000	NA					1	0.001	1000	NA
161	TCGA-A2-A1JP	1	0.001	1000	NA					1	0.001	1000	NA
162	TCGA-A2-A1JQ	1	0.001	1000	NA					1	0.001	1000	NA
163	TCGA-A2-A1JR	1	0.001	1000	NA					1	0.001	1000	NA
164	TCGA-A2-A1JS	1	0.001	1000	NA					1	0.001	1000	NA
165	TCGA-A2-A1JT	1	0.001	1000	NA					1	0.001	1000	NA
166	TCGA-A2-A1JU	1	0.001	1000	NA					1	0.001	1000	NA
167	TCGA-A2-A1JV	1	0.001	1000	NA					1	0.001	1000	NA
168	TCGA-A2-A1JW	1	0.001	1000	NA					1	0.001	1000	NA
169	TCGA-A2-A1JX	1	0.001	1000	NA					1	0.001	1000	NA
170	TCGA-A2-A1JY	1	0.001	1000	NA					1	0.001	1000	NA
171	TCGA-A2-A1JZ	1	0.001	1000	NA					1	0.001	1000	NA
172	TCGA-A2-A1K0	1	0.001	1000	NA					1	0.001	1000	NA
173	TCGA-A2-A1K1	1	0.001	1000	NA					1	0.001	1000	NA
174	TCGA-A2-A1K2	1	0.001	1000	NA					1	0.001	1000	NA
175	TCGA-A2-A1K3	1	0.001	1000	NA					1	0.001	1000	NA
176	TCGA-A2-A1K4	1	0.001	1000	NA					1	0.001	1000	NA
177	TCGA-A2-A1K5	1	0.001	1000	NA					1	0.001	1000	NA
178	TCGA-A2-A1K6	1	0.001	1000	NA					1	0.001	1000	NA
179	TCGA-A2-A1K7	1	0.001	1000	NA					1	0.001	1000	NA
180	TCGA-A2-A1K8	1	0.001	1000	NA					1	0.001	1000	NA
181	TCGA-A2-A1K9	1	0.001	1000	NA					1	0.001	1000	NA
182	TCGA-A2-A1KA	1	0.001	1000	NA					1	0.001	1000	NA
183	TCGA-A2-A1KB	1	0.001	1000	NA					1	0.001	1000	NA
184	TCGA-A2-A1KC	1	0.001	1000	NA					1	0.001	1000	NA
185	TCGA-A2-A1KD	1	0.001	1000	NA					1	0.001	1000	NA
186	TCGA-A2-A1KE	1	0.001	1000	NA					1	0.001	1000	NA
187	TCGA-A2-A1KF	1	0.001	1000	NA					1	0.001	1000	NA
188	TCGA-A2-A1KG	1	0.001	1000	NA					1	0.001	1000	NA
189	TCGA-A2-A1KH	1	0.001	1000	NA					1	0.001	1000	NA
190	TCGA-A2-A1KI	1	0.001	1000	NA					1	0.001	1000	NA
191	TCGA-A2-A1KJ	1	0.001	1000	NA					1	0.001	1000	NA
192	TCGA-A2-A1KK	1	0.001	1000	NA					1	0.001	1000	NA
193	TCGA-A2-A1KL	1	0.001	1000	NA					1	0.001	1000	NA
194	TCGA-A2-A1KM	1	0.001	1000	NA					1	0.001	1000	NA
195	TCGA-A2-A1KN	1	0.001	1000	NA					1	0.001	1000	NA
196	TCGA-A2-A1KO	1	0.001	1000	NA					1	0.001	1000	NA
197	TCGA-A2-A1KP	1	0.001	1000	NA					1	0.001	1000	NA
198	TCGA-A2-A1KQ	1	0.001	1000	NA					1	0.001	1000	NA
199	TCGA-A2-A1KR	1	0.001	1000	NA					1	0.001	1000	NA
200	TCGA-A2-A1KS	1	0.001	1000	NA					1	0.001	1000	NA
201	TCGA-A2-A1KT	1	0.001	1000	NA					1	0.001	1000	NA
202	TCGA-A2-A1KU	1	0.001	1000	NA					1	0.001	1000	NA
203	TCGA-A2-A1KV	1	0.001	1000	NA					1	0.001	1000	NA
204	TCGA-A2-A1KW	1	0.001	1000	NA					1	0.001	1000	NA
205	TCGA-A2-A1KX	1	0.001	1000	NA					1	0.001	1000	NA
206	TCGA-A2-A1KY	1	0.001	1000	NA					1	0.001	1000	NA
207	TCGA-A2-A1KZ	1	0.001	1000	NA					1	0.001	1000	NA
208	TCGA-A2-A1L0	1	0.001	1000	NA					1	0.001	1000	NA
209	TCGA-A2-A1L1	1	0.001	1000	NA					1	0.001	1000	NA
210	TCGA-A2-A1L2	1	0.001	1000	NA					1	0.001	1000	NA
211	TCGA-A2-A1L3	1	0.001	1000	NA					1	0.001	1000	NA
212	TCGA-A2-A1L4	1	0.001	1000	NA					1	0.001	1000	NA
213	TCGA-A2-A1L5	1	0.001	1000	NA					1	0.001	1000	NA
214	TCGA-A2-A1L6	1	0.001	1000	NA					1	0.001	1000	NA
215	TCGA-A2-A1L7	1	0.001	1000	NA					1	0.001	1000	NA
216	TCGA-A2-A1L8	1	0.001	1000	NA					1	0.001	1000	NA
217	TCGA-A2-A1L9	1	0.001	1000	NA					1	0.001	1000	NA
218	TCGA-A2-A1LA	1	0.001	1000	NA					1	0.001	1000	NA
219	TCGA-A2-A1LB	1	0.001	1000	NA					1	0.001	1000	NA
220	TCGA-A2-A1LC	1	0.001	1000	NA					1	0.001	1000	NA
221	TCGA-A2-A1LD	1	0.001	1000	NA					1	0.001	1000	NA
222	TCGA-A2-A1LE	1	0.001	1000	NA					1	0.001	1000	NA
223	TCGA-A2-A1LF	1	0.001	1000	NA					1	0.001	1000	NA
224	TCGA-A2-A1LG	1	0.001	1000	NA					1	0.001	1000	NA
225	TCGA-A2-A1LH	1	0.001	1000	NA					1	0.001	1000	NA
226	TCGA-A2-A1LI	1	0.001	1000	NA					1	0.001	1000	NA
227	TCGA-A2-A1LJ	1	0.001	1000	NA					1	0.001	1000	NA
228	TCGA-A2-A1LK	1	0.001	1000	NA					1	0.001	1000	NA
229	TCGA-A2-A1LL	1	0.001	1000	NA					1	0.001	1000	NA
230	TCGA-A2-A1LM	1	0.001	1000	NA					1	0.001	1000	NA
231	TCGA-A2-A1LN	1	0.001	1000	NA					1	0.001	1000	NA
232	TCGA-A2-A1LO	1	0.001	1000	NA					1	0.001	1000	NA
233	TCGA-A2-A1LP	1	0.001	1000	NA					1	0.001	1000	NA
234	TCGA-A2-A1LQ	1	0.001	1000	NA					1	0.001	1000	NA
235	TCGA-A2-A1LR	1	0.001	1000	NA					1	0.001	1000	NA
236	TCGA-A2-A1LS	1	0.001	1000	NA					1	0.001	1000	NA
237	TCGA-A2-A1LT	1	0.001	1000	NA					1	0.001	1000	NA
238	TCGA-A2-A1LU	1	0.001	1000	NA					1	0.001	1000	NA
239	TCGA-A2-A1LV	1	0.001	1000	NA					1	0.001	1000	NA
240	TCGA-A2-A1LW	1	0.001	1000	NA					1	0.001	1000	NA
241	TCGA-A2-A1LX	1	0.001	1000	NA					1	0.001	1000	NA
242	TCGA-A2-A1LY	1	0.001	1000	NA					1	0.001	1000	NA
243	TCGA-A2-A1LZ	1	0.001	1000	NA					1	0.001	1000	NA
244	TCGA-A2-A1M0	1	0.001	1000	NA					1	0.001	1000	NA
245	TCGA-A2-A1M1	1	0.001	1000	NA					1	0.001	1000	NA
246	TCGA-A2-A1M2	1	0.001	1000	NA					1	0.001	1000	NA
247	TCGA-A2-A1M3	1	0.001	1000	NA					1	0.001	1000	NA
248	TCGA-A2-A1M4	1	0.001	1000	NA					1	0.001	1000	NA
249	TCGA-A2-A1M5	1	0.001	1000	NA					1	0.001	1000	NA
250	TCGA-A2-A1M6	1	0.001	1000	NA					1	0.001	1000	NA
251	TCGA-A2-A1M7	1	0.001	1000	NA					1	0.001	1000	NA
252	TCGA-A2-A1M8	1	0.001	1000	NA					1	0.001	1000	NA
253	TCGA-A2-A1M9	1	0.001	1000	NA					1	0.001	1000	NA
254	TCGA-A2-A1MA	1	0.001	1000	NA					1	0.001	1000	NA
255	TCGA-A2-A1MB	1	0.001	1000	NA					1	0.001	1000	NA
256	TCGA-A2-A1MC	1	0.001	1000	NA					1	0.001	1000	NA
257	TCGA-A2-A1MD	1	0.001	1000	NA					1	0.001	1000	NA
258	TCGA-A2-A1ME	1	0.001	1000	NA					1	0.001	1000	NA
259	TCGA-A2-A1MF	1	0.001	1000	NA					1	0.001	1000	NA
260	TCGA-A2-A1MG	1	0.001	1000	NA					1	0.001	1000	NA
261	TCGA-A2-A1MH	1	0.001	1000	NA	

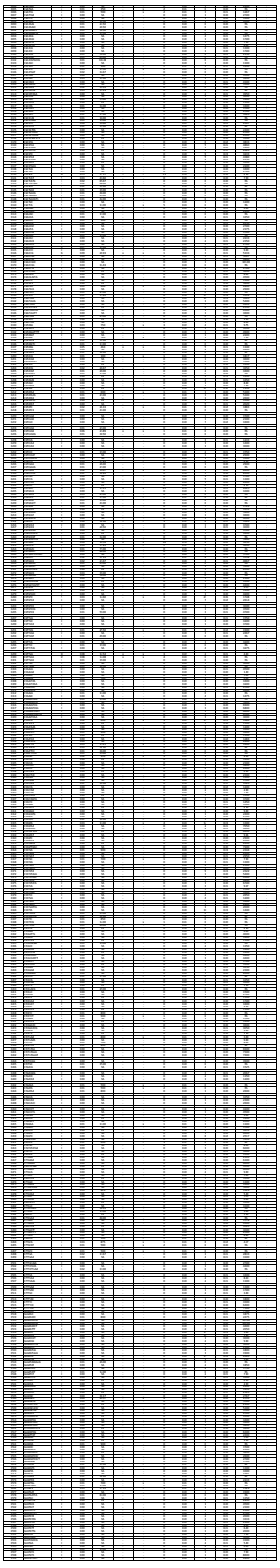


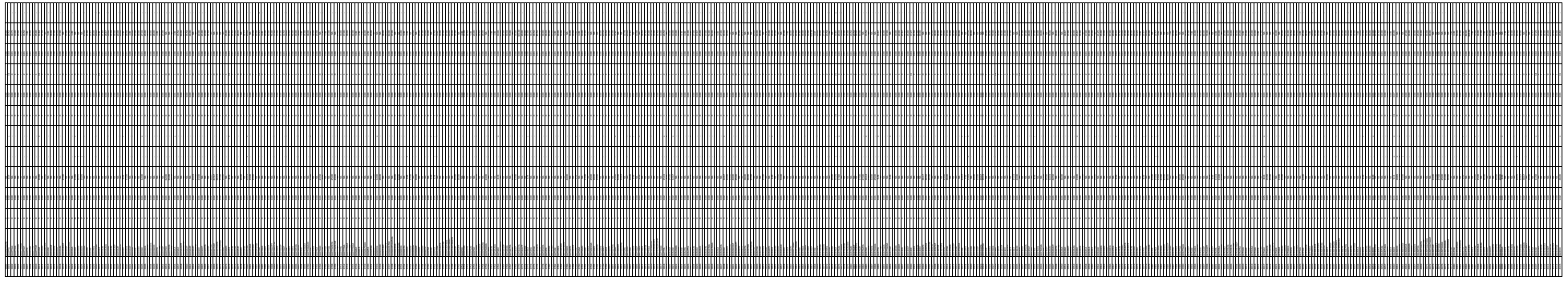


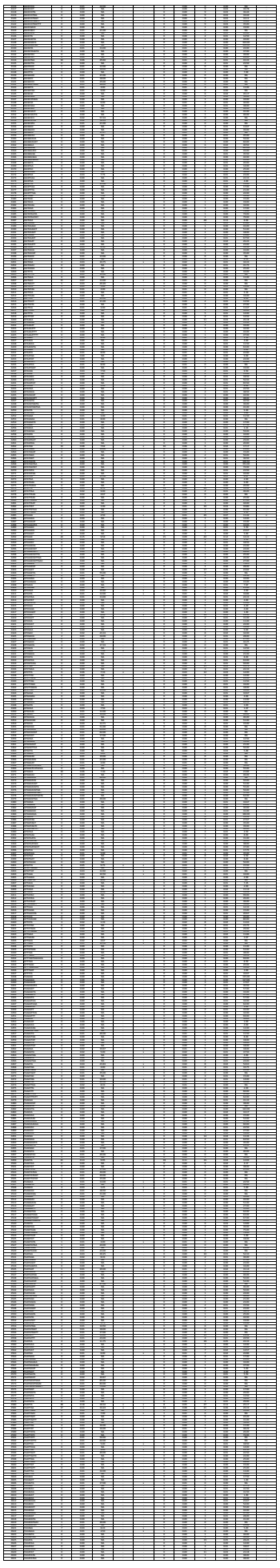


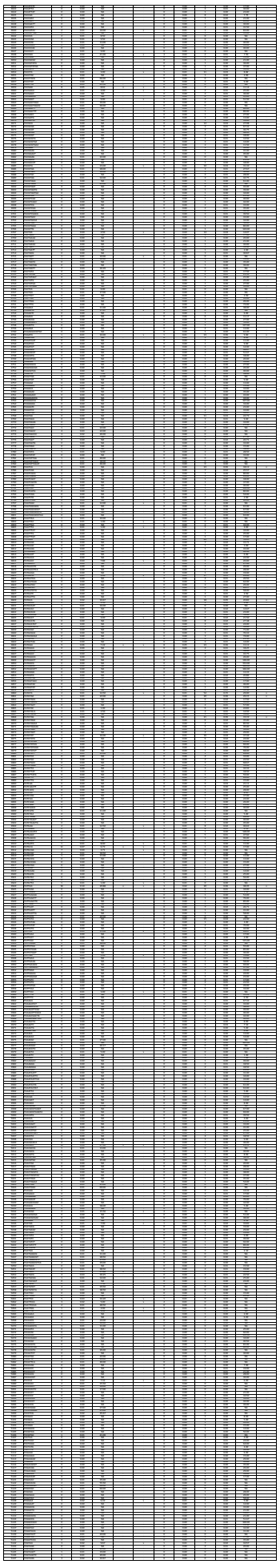












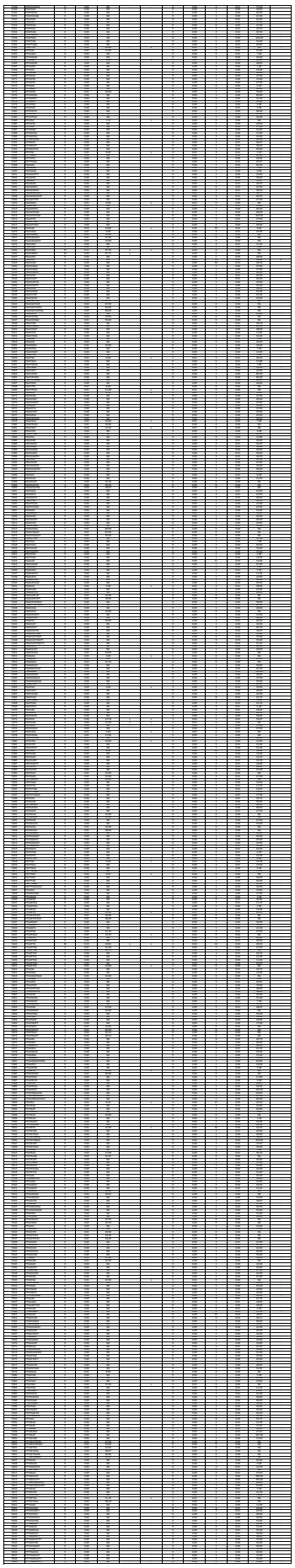


Table S2A. Most relevant pfAMA1 (hsp_pfAMA1) and mtrec pfAMA1 predicted structures (mtrec_pfAMA1) with similarity to AMA1 as a control (SCALOP Pfrtbs). Structures provide by pfAMA1 are given in Supplementary Table S3. Structures provided by mtrec_pfAMA1 are given in Supplementary Table S4. Structures provided by hsp_pfAMA1 are given in Supplementary Table S5. Structures provided by Pfrtbs are given in Supplementary Table S6. Within pfAMA1 sequence is more similar to a specific TMRM that to the rest due to searching TMRM of the same family. Between pfAMA1 sequence is more similar to mtrec searching TMRM of family than to hsp searching TMRM of other TF families. Random pfAMA1 sequence is more similar to mtrec searching TMRM from a family than to a random mtrec TF family. With parameters see **Table S1C**.

Proteinmodel	pfAMA1	pfAMA1 mtrec TF TMRM				pfAMA1 hsp				TMRM mtrec to Pfs				Pfs vs pfAMA1	Pfs Pfrtbs1	Pfs Pfrtbs2
		mtrec pfAMA1	mtrec pfAMA1	mtrec pfAMA1	mtrec pfAMA1	HSP	TF family	TF	HSP	mtrec	TF family	TF	mtrec			
1	1ama1a	1	A			0.99	1ama1a	1ama1a	1ama1a	1ama1a	1ama1a	1ama1a	1ama1a	SHP	1ama1a	1ama1a
2	1ama1b	1	A			0.99	1ama1b	1ama1b	1ama1b	1ama1b	1ama1b	1ama1b	1ama1b	SHP	1ama1b	1ama1b
3	1ama1c	1	A			0.99	1ama1c	1ama1c	1ama1c	1ama1c	1ama1c	1ama1c	1ama1c	SHP	1ama1c	1ama1c
4	1ama1d	1	A			0.99	1ama1d	1ama1d	1ama1d	1ama1d	1ama1d	1ama1d	1ama1d	SHP	1ama1d	1ama1d
5	1ama1e	1	A			0.99	1ama1e	1ama1e	1ama1e	1ama1e	1ama1e	1ama1e	1ama1e	SHP	1ama1e	1ama1e
6	1ama1f	1	A			0.99	1ama1f	1ama1f	1ama1f	1ama1f	1ama1f	1ama1f	1ama1f	SHP	1ama1f	1ama1f
7	1ama1g	1	A			0.99	1ama1g	1ama1g	1ama1g	1ama1g	1ama1g	1ama1g	1ama1g	SHP	1ama1g	1ama1g
8	1ama1h	1	A			0.99	1ama1h	1ama1h	1ama1h	1ama1h	1ama1h	1ama1h	1ama1h	SHP	1ama1h	1ama1h
9	1ama1i	1	A			0.99	1ama1i	1ama1i	1ama1i	1ama1i	1ama1i	1ama1i	1ama1i	SHP	1ama1i	1ama1i
10	1ama1j	1	A			0.99	1ama1j	1ama1j	1ama1j	1ama1j	1ama1j	1ama1j	1ama1j	SHP	1ama1j	1ama1j
11	1ama1k	1	A			0.99	1ama1k	1ama1k	1ama1k	1ama1k	1ama1k	1ama1k	1ama1k	SHP	1ama1k	1ama1k
12	1ama1l	1	A			0.99	1ama1l	1ama1l	1ama1l	1ama1l	1ama1l	1ama1l	1ama1l	SHP	1ama1l	1ama1l
13	1ama1m	1	A			0.99	1ama1m	1ama1m	1ama1m	1ama1m	1ama1m	1ama1m	1ama1m	SHP	1ama1m	1ama1m
14	1ama1n	1	A			0.99	1ama1n	1ama1n	1ama1n	1ama1n	1ama1n	1ama1n	1ama1n	SHP	1ama1n	1ama1n
15	1ama1o	1	A			0.99	1ama1o	1ama1o	1ama1o	1ama1o	1ama1o	1ama1o	1ama1o	SHP	1ama1o	1ama1o
16	1ama1p	1	A			0.99	1ama1p	1ama1p	1ama1p	1ama1p	1ama1p	1ama1p	1ama1p	SHP	1ama1p	1ama1p
17	1ama1q	1	A			0.99	1ama1q	1ama1q	1ama1q	1ama1q	1ama1q	1ama1q	1ama1q	SHP	1ama1q	1ama1q
18	1ama1r	1	A			0.99	1ama1r	1ama1r	1ama1r	1ama1r	1ama1r	1ama1r	1ama1r	SHP	1ama1r	1ama1r
19	1ama1s	1	A			0.99	1ama1s	1ama1s	1ama1s	1ama1s	1ama1s	1ama1s	1ama1s	SHP	1ama1s	1ama1s
20	1ama1t	1	A			0.99	1ama1t	1ama1t	1ama1t	1ama1t	1ama1t	1ama1t	1ama1t	SHP	1ama1t	1ama1t
21	1ama1u	1	A			0.99	1ama1u	1ama1u	1ama1u	1ama1u	1ama1u	1ama1u	1ama1u	SHP	1ama1u	1ama1u
22	1ama1v	1	A			0.99	1ama1v	1ama1v	1ama1v	1ama1v	1ama1v	1ama1v	1ama1v	SHP	1ama1v	1ama1v
23	1ama1w	1	A			0.99	1ama1w	1ama1w	1ama1w	1ama1w	1ama1w	1ama1w	1ama1w	SHP	1ama1w	1ama1w
24	1ama1x	1	A			0.99	1ama1x	1ama1x	1ama1x	1ama1x	1ama1x	1ama1x	1ama1x	SHP	1ama1x	1ama1x
25	1ama1y	1	A			0.99	1ama1y	1ama1y	1ama1y	1ama1y	1ama1y	1ama1y	1ama1y	SHP	1ama1y	1ama1y
26	1ama1z	1	A			0.99	1ama1z	1ama1z	1ama1z	1ama1z	1ama1z	1ama1z	1ama1z	SHP	1ama1z	1ama1z
27	1ama1aa	1	A			0.99	1ama1aa	1ama1aa	1ama1aa	1ama1aa	1ama1aa	1ama1aa	1ama1aa	SHP	1ama1aa	1ama1aa
28	1ama1ab	1	A			0.99	1ama1ab	1ama1ab	1ama1ab	1ama1ab	1ama1ab	1ama1ab	1ama1ab	SHP	1ama1ab	1ama1ab
29	1ama1ac	1	A			0.99	1ama1ac	1ama1ac	1ama1ac	1ama1ac	1ama1ac	1ama1ac	1ama1ac	SHP	1ama1ac	1ama1ac
30	1ama1ad	1	A			0.99	1ama1ad	1ama1ad	1ama1ad	1ama1ad	1ama1ad	1ama1ad	1ama1ad	SHP	1ama1ad	1ama1ad
31	1ama1ae	1	A			0.99	1ama1ae	1ama1ae	1ama1ae	1ama1ae	1ama1ae	1ama1ae	1ama1ae	SHP	1ama1ae	1ama1ae
32	1ama1af	1	A			0.99	1ama1af	1ama1af	1ama1af	1ama1af	1ama1af	1ama1af	1ama1af	SHP	1ama1af	1ama1af
33	1ama1ag	1	A			0.99	1ama1ag	1ama1ag	1ama1ag	1ama1ag	1ama1ag	1ama1ag	1ama1ag	SHP	1ama1ag	1ama1ag
34	1ama1ah	1	A			0.99	1ama1ah	1ama1ah	1ama1ah	1ama1ah	1ama1ah	1ama1ah	1ama1ah	SHP	1ama1ah	1ama1ah
35	1ama1ai	1	A			0.99	1ama1ai	1ama1ai	1ama1ai	1ama1ai	1ama1ai	1ama1ai	1ama1ai	SHP	1ama1ai	1ama1ai
36	1ama1aj	1	A			0.99	1ama1aj	1ama1aj	1ama1aj	1ama1aj	1ama1aj	1ama1aj	1ama1aj	SHP	1ama1aj	1ama1aj
37	1ama1ak	1	A			0.99	1ama1ak	1ama1ak	1ama1ak	1ama1ak	1ama1ak	1ama1ak	1ama1ak	SHP	1ama1ak	1ama1ak
38	1ama1al	1	A			0.99	1ama1al	1ama1al	1ama1al	1ama1al	1ama1al	1ama1al	1ama1al	SHP	1ama1al	1ama1al
39	1ama1am	1	A			0.99	1ama1am	1ama1am	1ama1am	1ama1am	1ama1am	1ama1am	1ama1am	SHP	1ama1am	1ama1am
40	1ama1an	1	A			0.99	1ama1an	1ama1an	1ama1an	1ama1an	1ama1an	1ama1an	1ama1an	SHP	1ama1an	1ama1an
41	1ama1ao	1	A			0.99	1ama1ao	1ama1ao	1ama1ao	1ama1ao	1ama1ao	1ama1ao	1ama1ao	SHP	1ama1ao	1ama1ao
42	1ama1ap	1	A			0.99	1ama1ap	1ama1ap	1ama1ap	1ama1ap	1ama1ap	1ama1ap	1ama1ap	SHP	1ama1ap	1ama1ap
43	1ama1aq	1	A			0.99	1ama1aq	1ama1aq	1ama1aq	1ama1aq	1ama1aq	1ama1aq	1ama1aq	SHP	1ama1aq	1ama1aq
44	1ama1ar	1	A			0.99	1ama1ar	1ama1ar	1ama1ar	1ama1ar	1ama1ar	1ama1ar	1ama1ar	SHP	1ama1ar	1ama1ar
45	1ama1as	1	A			0.99	1ama1as	1ama1as	1ama1as	1ama1as	1ama1as	1ama1as	1ama1as	SHP	1ama1as	1ama1as
46	1ama1at	1	A			0.99	1ama1at	1ama1at	1ama1at	1ama1at	1ama1at	1ama1at	1ama1at	SHP	1ama1at	1ama1at
47	1ama1au	1	A			0.99	1ama1au	1ama1au	1ama1au	1ama1au	1ama1au	1ama1au	1ama1au	SHP	1ama1au	1ama1au
48	1ama1av	1	A			0.99	1ama1av	1ama1av	1ama1av	1ama1av	1ama1av	1ama1av	1ama1av	SHP	1ama1av	1ama1av
49	1ama1aw	1	A			0.99	1ama1aw	1ama1aw	1ama1aw	1ama1aw	1ama1aw	1ama1aw	1ama1aw	SHP	1ama1aw	1ama1aw
50	1ama1ax	1	A			0.99	1ama1ax	1ama1ax	1ama1ax	1ama1ax	1ama1ax	1ama1ax	1ama1ax	SHP	1ama1ax	1ama1ax
51	1ama1ay	1	A			0.99	1ama1ay	1ama1ay	1ama1ay	1ama1ay	1ama1ay	1ama1ay	1ama1ay	SHP	1ama1ay	1ama1ay
52	1ama1az	1	A			0.99	1ama1az	1ama1az	1ama1az	1ama1az	1ama1az	1ama1az	1ama1az	SHP	1ama1az	1ama1az
53	1ama1ba	1	A			0.99	1ama1ba	1ama1ba	1ama1ba	1ama1ba	1ama1ba	1ama1ba	1ama1ba	SHP	1ama1ba	1ama1ba
54	1ama1bb	1	A			0.99	1ama1bb	1ama1bb	1ama1bb	1ama1bb	1ama1bb	1ama1bb	1ama1bb	SHP	1ama1bb	1ama1bb
55	1ama1bc	1	A			0.99	1ama1bc	1ama1bc	1ama1bc	1ama1bc	1ama1bc	1ama1bc	1ama1bc	SHP	1ama1bc	1ama1bc
56	1ama1bd	1	A			0.99	1ama1bd	1ama1bd	1ama1bd	1ama1bd	1ama1bd	1ama1bd	1ama1bd	SHP	1ama1bd	1ama1bd
57	1ama1be	1	A			0.99	1ama1be	1ama1be	1ama1be	1ama1be	1ama1be	1ama1be	1ama1be	SHP	1ama1be	1ama1be
58	1ama1bf	1	A			0.99	1ama1bf	1ama1bf	1ama1bf	1ama1bf	1ama1bf	1ama1bf	1ama1bf	SHP	1ama1bf	1ama1bf
59	1ama1bg	1	A			0.99	1ama1bg	1ama1bg	1ama1bg	1ama1bg	1ama1bg	1ama1bg	1ama1bg	SHP	1ama1bg	1ama1bg
60	1ama1bh	1	A			0.99	1ama1bh	1ama1bh	1ama1bh	1ama1bh	1ama1bh	1ama1bh	1ama1bh	SHP	1ama1bh	1ama1bh
61	1ama1bi	1	A			0.99	1ama1bi	1ama1bi	1ama1bi	1ama1bi	1ama1bi	1ama1bi	1ama1bi	SHP	1ama1bi	1ama1bi
62	1ama1bj	1	A			0.99	1ama1bj	1ama1bj	1ama1bj	1ama1bj	1ama1bj	1ama1bj	1ama1bj	SHP	1ama1bj	1ama1bj
63	1ama1bk	1	A			0.99	1ama1bk	1ama1bk	1ama1bk	1ama1bk	1ama1bk	1ama1bk	1ama1bk	SHP	1ama1bk	1ama1bk
64	1ama1bl	1	A			0.99	1ama1bl	1ama1bl	1ama1bl	1ama1bl	1ama1bl	1ama1bl	1ama1bl	SHP	1ama1bl	1ama1bl
65	1ama1bm	1	A			0.99	1ama1bm	1ama1bm	1ama1bm	1ama1bm	1ama1bm	1ama1bm	1ama1bm	SHP	1ama1bm	1ama1bm
66	1ama1bn	1	A			0.99	1ama1bn	1ama1bn	1ama1bn	1ama1bn	1ama1bn	1ama1bn	1ama1bn	SHP	1ama1bn	1ama1bn
67	1ama1bo	1	A			0.99	1ama1bo	1ama1bo	1ama1bo	1ama1bo	1ama1bo	1ama1bo	1ama1bo	SHP	1ama1bo	1ama1bo
68	1ama1bp	1	A			0.99	1ama1bp	1ama1bp	1ama1bp	1ama1bp	1ama1bp	1ama1bp	1ama1bp	SHP	1ama1bp	1ama1bp
69	1ama1bq	1	A			0.99	1ama1bq	1ama1bq	1ama1bq	1ama1bq	1ama1bq	1ama1bq	1ama1bq	SHP	1ama1bq	1ama1bq
70	1ama1br	1	A			0.99	1ama1br	1ama1br	1ama1br	1ama1br	1ama1br	1ama1br	1ama1br	SHP	1ama1br	1ama1br
71	1ama1bs	1	A			0.99	1ama1bs	1ama1bs	1ama1bs	1ama1bs	1ama1bs	1ama1bs	1ama1bs	SHP	1ama1bs	1ama1bs
72	1ama1bt	1	A			0.99	1ama1bt	1ama1bt	1ama1bt	1ama1bt	1ama1bt	1ama1bt	1ama1bt	SHP	1ama1bt	1ama1bt
73	1ama1bu	1	A			0.99	1ama1bu	1ama1bu	1ama1bu	1ama1bu	1ama1bu	1ama1bu	1ama1bu	SHP	1ama1bu	1ama1bu
74	1ama1bv	1	A			0.99	1ama1bv	1ama1bv	1ama1bv	1ama1bv	1ama1bv	1ama1bv	1ama1bv	SHP	1ama1bv	1ama1bv
75	1ama1bw	1	A			0.99	1ama1bw	1ama1bw	1ama1bw	1ama1bw	1ama1bw	1ama1bw	1ama1bw	SHP	1ama1bw	1ama1bw
76	1ama1bx	1	A			0.99	1ama1bx	1ama1bx	1ama1bx	1ama1bx	1ama1bx	1ama1bx	1ama1bx	SHP	1ama1bx	1ama1bx
77	1ama1by	1	A			0.99	1ama1by	1ama1by	1ama1by	1ama1by	1ama1by	1ama1by	1ama1by	SHP	1ama1by	1ama1by
78	1ama1bz	1	A			0.99	1ama1bz	1ama1bz	1ama1bz	1ama1bz	1ama1bz	1ama1bz	1ama1bz	SHP	1ama1bz	1ama1bz
79	1ama1ca	1	A			0.99	1ama1ca	1ama1ca	1ama1ca	1ama1ca	1ama1ca	1ama1ca	1ama1ca	SHP	1ama1ca	1ama1ca
80	1ama1cb	1	A			0.99	1ama1cb	1ama1cb	1ama1cb	1ama1cb	1ama1cb	1ama1cb	1ama1cb	SHP	1ama1cb	1ama1cb
81	1ama1cc	1	A			0.99	1ama1cc	1ama1cc	1ama1cc	1ama1cc	1ama1cc	1ama1cc	1ama1cc	SHP	1ama1cc	1ama1cc
82	1ama1cd	1	A			0.99	1ama1cd	1ama1cd	1ama1cd	1ama1cd	1ama1cd	1ama1cd	1ama1cd	SHP	1ama1cd	1ama1cd
83	1ama1ce	1	A			0.99	1ama1ce	1ama1ce	1ama1ce	1ama1ce	1ama1ce	1ama1ce	1ama1ce	SHP	1ama1ce	1ama1ce
84	1ama1cf	1	A			0.99	1ama1cf	1ama1cf	1ama1cf	1ama1cf	1ama1cf	1ama1cf	1ama1cf	SHP	1ama1cf	1ama1cf
85	1ama1cg	1	A			0.99	1ama1cg	1ama1cg	1ama1cg	1ama1cg	1ama1cg	1ama1cg	1ama1cg	SHP	1ama1cg	1ama1cg
86	1ama1ch	1	A			0.99	1ama1ch	1ama1ch	1ama1ch	1ama1ch	1ama1ch	1ama1ch	1ama1ch	SHP	1ama1ch	1ama1ch
87	1ama1ci	1	A			0.99	1ama1ci	1ama1ci	1ama1ci	1ama1ci	1ama1ci	1ama1ci	1ama1ci	SHP	1ama1ci	1ama1ci
88	1ama1cj	1	A			0.99	1ama1cj	1ama1cj	1ama1cj	1ama1cj	1ama1cj	1ama1cj	1ama1cj	SHP	1ama1cj	1ama1cj
89	1ama1ck	1	A			0.99	1ama1ck	1ama1ck	1ama1ck	1ama1ck	1ama1ck	1ama1ck	1ama1ck	SHP	1ama1ck	1ama1ck
90	1ama1cl	1	A			0.99	1ama1cl	1ama1cl	1ama1cl	1ama1cl	1ama1cl	1ama1cl	1ama1cl	SHP	1ama1cl	1ama1cl
91	1ama1cm	1	A			0.99	1ama1cm	1ama1cm	1ama1cm	1ama1cm	1ama1cm	1ama1cm	1ama1cm	SHP	1ama1cm	1ama1cm
92	1ama1cn	1	A			0.99	1ama1cn	1ama1cn	1ama1cn	1ama1cn	1ama1cn	1ama1cn	1ama1cn	SHP	1ama1cn	1ama1cn
93	1ama1co	1	A			0.99	1ama1co	1ama1co	1ama1co	1ama1co	1ama1co	1ama1co	1ama1co	SHP	1ama1co	1ama1co
94	1ama1cp	1	A			0.99	1ama1cp	1ama1cp	1ama1cp	1ama1cp	1ama1cp	1ama1cp	1ama1cp	SHP	1ama1cp	1ama1cp
95	1ama1cq	1	A			0.99	1ama1cq	1ama1cq	1ama1cq	1ama1cq	1ama1cq	1ama1cq	1ama1cq	SHP	1ama1cq	1ama1cq
96	1ama1cr	1	A			0.99	1ama1cr	1ama1cr	1ama1cr	1ama1cr	1ama1cr	1ama1cr	1ama1cr	SHP	1ama1cr	1ama1cr
97	1ama1cs	1	A			0.99	1ama1cs	1ama1cs	1ama1cs	1ama1cs	1ama1cs	1ama1cs	1ama1cs	SHP	1ama1cs	1ama1cs
98	1ama1ct	1	A			0.99	1ama1ct	1ama1ct	1ama1ct	1ama1ct	1ama1ct	1ama1ct	1ama1ct	SHP	1ama1ct	1ama1ct
99	1ama1cu	1	A			0.99	1ama1cu	1ama1cu	1ama1cu	1ama1cu	1ama1cu	1ama1cu	1ama1cu	SHP	1ama1cu	1ama1cu
100	1ama1cv	1	A			0.99	1ama1cv	1ama1cv	1ama1cv	1ama1cv	1ama1cv	1ama1cv	1ama1cv	SHP	1ama1cv	1ama1cv
101	1ama1cw	1	A			0.99	1ama1cw	1ama1cw	1ama1cw	1ama1cw	1ama1cw	1ama1cw	1ama1cw	SHP	1ama1cw	1ama1cw
102	1ama1cx	1	A			0.99	1ama1cx	1ama1cx	1ama1cx	1ama1cx	1ama1cx	1ama1cx	1ama1cx	SHP	1ama1cx	1ama1cx
103	1ama1cy	1	A			0.99	1ama1cy	1ama1cy	1ama1cy	1ama1cy	1ama1cy	1ama1cy	1ama1cy	SHP	1ama1cy	1ama1cy
104	1ama1cz	1	A			0.99	1ama1cz	1ama1cz	1ama1cz	1ama1cz	1ama1cz	1ama1cz	1ama1cz	SHP	1ama1cz	1ama1cz
105	1ama1da	1	A			0.99	1ama1da	1ama1da	1ama1da	1ama1da	1ama1da	1ama1da	1ama1da	SHP	1ama1da	1ama1da
106	1ama1db	1	A			0.99	1ama1db	1ama1db	1ama1db	1ama1db	1ama1db	1ama1db	1ama1db	SHP	1ama1db	1ama1db
107	1ama1dc	1	A			0.99	1ama1dc	1ama1dc	1ama1dc	1ama1dc	1ama1dc	1ama1dc	1ama1dc	SHP	1ama1dc	1ama1dc
108	1ama1dd	1	A			0.99	1ama1dd	1ama1dd	1ama1dd	1ama1dd	1ama1dd	1ama1dd	1ama1dd	SHP	1ama1dd	1ama1dd
109	1ama1de	1	A			0.99	1ama1de	1ama1de	1ama1de	1ama1de	1ama1de	1ama1de	1ama1de	SHP	1ama1de	1ama1de
110	1ama1df	1	A			0.99	1ama1df	1ama1df	1ama1df	1ama1df	1ama1df	1ama1df	1ama1df	SHP	1ama1df	1ama1df
111	1ama1dg	1	A			0.99	1ama1dg	1ama1dg	1ama1dg	1ama1dg	1ama1dg	1ama1dg	1ama1dg	SHP	1ama1dg	1ama1dg
112	1ama1dh	1	A			0.99	1ama1dh	1ama1dh	1ama1dh	1ama1dh	1ama1dh	1ama1dh	1ama1dh	SHP	1ama1dh	1ama1dh
113	1ama1di	1	A			0.99	1ama1di	1ama1di	1ama1di	1ama1di	1ama1di	1ama1di	1ama1di	SHP	1ama1di	1ama1di
114	1ama1dj	1	A			0.99	1ama1dj	1ama1dj	1ama1dj	1ama1dj	1ama1dj	1ama1dj	1ama1dj	SHP	1ama1dj	1ama1dj
115	1ama1dk	1	A			0.99	1ama1dk	1ama1dk	1ama1dk	1ama1dk	1ama1dk	1ama1dk	1ama1dk	SHP	1ama1dk	1ama1dk
116	1ama1dl	1	A			0.99	1ama1dl	1ama1dl	1ama1dl	1ama1dl	1ama1dl	1ama1dl	1ama1dl	SHP	1ama1dl	1ama1dl
117	1ama1dm	1	A			0.99	1ama1dm	1ama1dm	1ama1dm	1ama1dm	1ama1dm	1ama1dm	1ama1dm	SHP	1ama1dm	1ama1dm
118	1ama1dn	1	A			0.99	1ama1dn	1ama1dn	1ama1dn	1ama1dn	1ama1dn	1ama1dn	1ama1dn	SHP	1ama1dn	1ama1dn
119	1ama1do	1	A			0.99	1ama1do	1ama1do	1ama1do	1ama1do	1ama1do	1ama1do	1ama1do	SHP	1ama1do	1ama1do
120	1ama1dp	1	A			0.99	1ama1dp	1ama1dp	1ama1dp	1ama1dp	1ama1dp	1ama1dp	1ama1dp	SHP	1ama1dp	1ama1dp
121	1ama1dq	1	A			0.99	1ama1dq	1ama1dq	1ama1dq	1ama1dq	1ama1dq	1ama1dq	1ama1dq	SHP	1ama1dq	1ama1dq
122	1ama1dr	1	A			0.99	1ama1dr	1ama1dr	1ama1dr	1ama1dr	1ama1dr	1ama1dr	1ama1dr	SHP	1ama1dr	1ama1dr
123	1ama1ds	1	A			0.99	1ama1ds	1ama1ds	1ama1ds	1ama1ds	1ama1ds	1ama1ds	1ama1ds	SHP	1ama1ds	1ama1ds
124	1ama1dt	1	A			0.99	1ama1dt	1ama1dt	1ama1dt	1ama1dt	1ama1dt	1ama1dt	1ama1dt	SHP	1ama1dt	1ama1dt
125	1ama1du	1	A			0.99	1ama1du	1ama1du	1ama1du	1ama1du	1ama1du	1ama1du	1ama1du	SHP	1ama1du	1ama1du
126	1ama1dv	1	A			0.99	1ama1dv	1ama1dv	1ama1dv	1ama1dv	1ama1dv	1ama1dv	1ama1dv	SHP	1ama1dv	1ama1dv
127	1ama1dw	1	A			0.99	1ama1dw	1ama1dw	1ama1dw	1ama1dw	1ama1dw	1ama1dw	1ama1dw	SHP	1ama1dw	1ama1dw
128	1ama1dx	1	A			0.99	1ama1dx	1ama1dx	1ama1dx	1ama1dx	1ama1dx	1ama1dx	1ama1dx	SHP	1ama1dx	1ama1dx
129	1ama1dy	1	A			0.99	1ama1dy	1ama1dy	1ama1dy	1ama1dy	1ama1dy	1ama1dy	1ama1dy	SHP	1ama1dy	1ama1dy
130	1ama1dz	1	A			0.99	1ama1dz	1ama1dz	1ama1dz	1ama1dz	1ama1dz	1ama1dz	1ama1dz	SHP	1ama1dz	1ama1dz
131	1ama1ea	1	A			0.99	1ama1ea	1ama1ea	1ama1ea	1ama1ea	1ama1ea	1ama1ea	1ama1ea	SHP	1ama1ea	1ama1ea
132	1ama1eb	1	A			0.99	1ama1eb	1ama1eb	1ama1eb	1ama1eb	1ama1eb	1ama1eb	1ama1eb	SHP	1ama1eb	1ama1eb
133	1ama1ec	1	A			0.99	1ama1ec	1ama1ec	1ama1ec	1ama1ec	1ama1ec	1ama1ec	1ama1ec	SHP	1ama1ec	1ama1ec
134	1ama1ed	1	A			0.99	1ama1ed	1ama1ed	1ama1ed	1ama1ed	1ama1ed	1ama1ed	1ama1ed	SHP	1ama1ed	1ama1ed
135	1ama1ee	1	A			0.99	1ama1ee	1ama1ee	1ama1ee	1ama1ee	1ama1ee	1ama1ee	1ama1ee	SHP	1ama1ee	1ama1ee
136	1ama1ef	1	A			0.99	1ama1ef	1ama1ef	1ama1ef	1ama1ef	1ama1ef	1ama1ef	1ama1ef	SHP	1ama1ef	1ama1ef
137	1ama1eg	1	A			0.99	1ama1eg	1ama1eg	1ama1eg	1ama1eg	1ama1eg	1ama1eg	1ama1eg	SHP	1ama1eg	1ama1eg
138	1ama1eh	1	A			0.99	1ama1eh	1ama1eh	1ama1eh	1ama1eh	1ama1eh	1ama1eh	1ama1eh	SHP	1ama1eh	1ama1eh
139	1ama1ei	1	A			0.99	1ama1ei	1ama1ei	1ama1ei	1ama1ei	1ama1ei	1ama1ei	1ama1ei	SHP	1ama1ei	1ama1ei
140	1ama1ej	1	A			0.99	1ama1ej	1ama1ej	1ama1ej	1ama1ej	1ama1ej	1ama1ej	1ama1ej	SHP	1ama1ej	1ama1ej
141	1ama1ek	1	A			0.99	1ama1ek	1ama1ek	1ama1ek	1ama1ek	1ama1ek	1ama1ek	1ama1ek	SHP	1ama1ek	1ama1ek
142	1ama1el	1	A			0.99	1ama1el	1ama1el	1ama1el	1ama1el	1ama1el	1ama1el	1ama1el	SHP	1ama1el	1ama1el
143	1ama1em	1	A			0.99	1ama1em	1ama1em	1ama1em	1ama1em	1ama1em	1ama1em	1ama1em	SHP	1ama1em	1ama1em
144	1ama1en	1	A			0.99	1ama1en	1ama1en	1ama1en	1ama1en	1ama1en	1ama1en	1ama1en	SHP	1ama1en	1ama1en
145	1ama1eo	1	A			0.99	1ama1eo	1ama1eo	1ama1eo	1ama1eo	1ama1eo	1ama1eo	1ama1eo	SHP	1ama1eo	1ama1eo
146	1ama1ep	1	A			0.99	1ama1ep	1ama1ep	1ama1ep	1ama1ep	1ama1ep	1ama1ep	1ama1ep	SHP	1ama1ep	1ama1ep
147	1ama1eq	1	A			0.99	1ama1eq	1ama1eq	1ama1eq	1ama1eq	1ama1eq	1ama1eq	1ama1eq	SHP	1ama1eq	1ama1eq
148	1ama1er	1	A			0.99	1ama1er	1ama1er	1ama1er	1ama1er	1ama1er	1ama1er	1ama1er	SHP	1ama1er	1ama1er
149	1ama1es	1	A			0.99	1ama1es	1ama1es	1ama1es	1ama1es	1ama1es	1ama1es	1ama1es	SHP	1ama1es	1ama1es
150	1ama1et	1	A			0.99	1ama1et	1ama1et	1ama1et	1ama1et	1ama1et	1ama1et	1ama1et	SHP	1ama1et	1ama1et
151	1ama1eu	1	A			0.99	1ama1eu	1ama1eu	1ama1eu	1ama1eu	1ama1eu	1ama1eu	1ama1eu	SHP	1ama1eu	1ama1eu
152	1ama1ev	1	A			0.99	1ama1ev	1ama1ev	1ama1ev	1ama1ev	1ama1ev	1ama1ev	1ama1ev	SHP	1ama1ev	1ama1ev
153	1ama1ew	1	A			0.99	1ama1ew	1ama1ew	1ama1ew	1ama1ew	1ama1ew	1ama1ew	1ama1ew	SHP	1ama1ew	1ama1ew
154	1ama1ex	1

Table S4B. Self-ranq and CVS-AT TRMs used in this study.

SELF-ranq		CVS-AT	
TF Family	TFMax	TF Family	TFMax
AP2-ERF	AP2-ERF_ERF1	AP2-ERF	AP2-ERF_ERF1
AP2-ERF	AP2-ERF_ERF2	AP2-ERF	AP2-ERF_ERF2
AP2-ERF	AP2-ERF_ERF3	AP2-ERF	AP2-ERF_ERF3
AP2-ERF	AP2-ERF_ERF4	AP2-ERF	AP2-ERF_ERF4
AP2-ERF

Table S4C. TF family 95th percentiles of within, between and random PCC thresholds determining the pCRE-TFBM similarity. PCCs were calculated as in Uygun et al. 2017.

TF Family	TF Family size	Comparison	Mean	Stdev	95perc
ABI3VP1	7	within	0.5529	0.1514	0.74
ABI3VP1	7	between	0.4377	0.1631	0.72
ABI3VP1	7	random	0.3784	0.2026	0.7
AP2-EREBP	115	within	0.731	0.1884	0.97
AP2-EREBP	115	between	0.3542	0.157	0.6
AP2-EREBP	115	random	0.3802	0.1747	0.67
ARF	3	within	0.69	0.164	0.89
ARF	3	between	0.4605	0.1392	0.73
ARF	3	random	0.4321	0.1571	0.73
ARID	4	within	0.8167	0.1526	0.98
ARID	4	between	0.3998	0.2127	0.7145
ARID	4	random	0.3139	0.2139	0.69
AT-hook	2	within	0.54	0	0.54
AT-hook	2	between	0.4083	0.2189	0.72
AT-hook	2	random	0.2592	0.2082	0.58
B3	5	within	0.676	0.2658	1
B3	5	between	0.3652	0.1369	0.6155
B3	5	random	0.3363	0.1585	0.57
BBR/BPC	3	within	1	0	1
BBR/BPC	3	between	0.3389	0.1211	0.55
BBR/BPC	3	random	0.3112	0.1553	0.57
bHLH	46	within	0.6922	0.1995	0.97
bHLH	46	between	0.3635	0.1491	0.63
bHLH	46	random	0.3626	0.1846	0.65
bZIP	50	within	0.6567	0.2072	0.99
bZIP	50	between	0.4059	0.1281	0.63
bZIP	50	random	0.4055	0.1537	0.68
BZR	4	within	0.955	0.0385	0.99
BZR	4	between	0.3851	0.1567	0.69
BZR	4	random	0.4052	0.1532	0.69
C2C2-CO-like	2	within	0.98	0	0.98
C2C2-CO-like	2	between	0.4745	0.1304	0.71
C2C2-CO-like	2	random	0.522	0.1267	0.74
C2C2-Dof	21	within	0.8771	0.1224	0.98
C2C2-Dof	21	between	0.4408	0.2022	0.78
C2C2-Dof	21	random	0.4016	0.1785	0.72
C2C2-YABBY	1	within	NaN	NaN	NaN
C2C2-YABBY	1	between	0.402	0.1778	0.6875
C2C2-YABBY	1	random	0.2913	0.1851	0.55
C2H2	36	within	0.5152	0.2089	0.97
C2H2	36	between	0.4293	0.1535	0.7
C2H2	36	random	0.378	0.1847	0.68
C3H	8	within	0.5429	0.147	0.78
C3H	8	between	0.4525	0.1646	0.74
C3H	8	random	0.3242	0.204	0.64
CAMTA	4	within	0.835	0.1519	0.98
CAMTA	4	between	0.3937	0.1348	0.65
CAMTA	4	random	0.3851	0.1575	0.66
CPP	8	within	0.9132	0.046	0.98
CPP	8	between	0.3729	0.1914	0.65
CPP	8	random	0.2843	0.2075	0.62
CSD	1	within	NaN	NaN	NaN
CSD	1	between	0.3958	0.243	0.8
CSD	1	random	0.2252	0.2248	0.53
DBP	1	within	NaN	NaN	NaN
DBP	1	between	0.4308	0.1297	0.67
DBP	1	random	0.45	0.1324	0.68
Dof	12	within	0.8359	0.1353	0.99
Dof	12	between	0.46	0.1733	0.74
Dof	12	random	0.3186	0.1773	0.63
E2F-DP	5	within	0.796	0.0971	1
E2F-DP	5	between	0.4081	0.1522	0.67
E2F-DP	5	random	0.4082	0.1756	0.74
EIL	5	within	0.727	0.1763	0.94
EIL	5	between	0.4005	0.127	0.62
EIL	5	random	0.396	0.1721	0.68
FAR1	2	within	0.74	0	0.74
FAR1	2	between	0.4037	0.1399	0.67
FAR1	2	random	0.4298	0.1525	0.6905
FHA	1	within	NaN	NaN	NaN
FHA	1	between	0.4769	0.1255	0.7075
FHA	1	random	0.5234	0.1263	0.73
G2-like	17	within	0.7692	0.1334	0.96
G2-like	17	between	0.4036	0.1657	0.69
G2-like	17	random	0.3901	0.1751	0.7
GATA	14	within	0.7464	0.1638	0.9595
GATA	14	between	0.3605	0.1222	0.57
GATA	14	random	0.3204	0.1648	0.6
GeBP	5	within	0.383	0.2931	0.96
GeBP	5	between	0.3667	0.186	0.65
GeBP	5	random	0.3378	0.2051	0.67
GRAS	1	within	NaN	NaN	NaN
GRAS	1	between	0.3931	0.1378	0.5975
GRAS	1	random	0.3588	0.149	0.6
GRF	3	within	0.6733	0.2106	0.9
GRF	3	between	0.4552	0.1346	0.68
GRF	3	random	0.4383	0.1501	0.7305
HMG	1	within	NaN	NaN	NaN
HMG	1	between	0.4018	0.2107	0.89
HMG	1	random	0.4185	0.1418	0.69
Homeobox	44	within	0.6606	0.1833	0.95
Homeobox	44	between	0.4056	0.1544	0.65
Homeobox	44	random	0.3491	0.1872	0.66
HSF	10	within	0.9798	0.0226	1
HSF	10	between	0.3788	0.1318	0.64
HSF	10	random	0.4189	0.147	0.65
LBD	9	within	0.7297	0.1405	0.9745
LBD	9	between	0.4006	0.1603	0.7
LBD	9	random	0.4286	0.1559	0.72
LFY	1	within	NaN	NaN	NaN
LFY	1	between	0.3725	0.1278	0.61
LFY	1	random	0.4315	0.1347	0.68
LIM	1	within	NaN	NaN	NaN
LIM	1	between	0.4035	0.1501	0.62
LIM	1	random	0.2602	0.1793	0.55
MADS	27	within	0.685	0.2062	0.96
MADS	27	between	0.4327	0.1592	0.69
MADS	27	random	0.3857	0.1939	0.7
mTERF	1	within	NaN	NaN	NaN
mTERF	1	between	0.3664	0.2512	0.93
mTERF	1	random	0.3578	0.1646	0.6405
MYB	92	within	0.514	0.1899	0.89
MYB	92	between	0.4199	0.1389	0.65
MYB	92	random	0.4066	0.1579	0.68
MYB-related	26	within	0.6625	0.203	0.99
MYB-related	26	between	0.4214	0.1656	0.68
MYB-related	26	random	0.4076	0.1804	0.73
NAC	60	within	0.6956	0.1577	0.93
NAC	60	between	0.436	0.1279	0.66
NAC	60	random	0.4471	0.1452	0.7
NLP	3	within	0.3967	0.1565	0.57
NLP	3	between	0.4131	0.12	0.62
NLP	3	random	0.443	0.1418	0.69
Orphan	3	within	0.44	0.1008	0.57
Orphan	3	between	0.4268	0.1813	0.7445
Orphan	3	random	0.4257	0.1629	0.74
RAV	1	within	NaN	NaN	NaN
RAV	1	between	0.4611	0.1192	0.67
RAV	1	random	0.4439	0.1529	0.72
REM	2	within	0.7	0	0.7
REM	2	between	0.4238	0.1987	0.76
REM	2	random	0.3242	0.1928	0.6305
S1Fa-like	1	within	NaN	NaN	NaN
S1Fa-like	1	between	0.3849	0.1473	0.6675
S1Fa-like	1	random	0.4231	0.1438	0.64
SBP	18	within	0.7514	0.1468	0.99
SBP	18	between	0.3975	0.1288	0.61
SBP	18	random	0.396	0.1566	0.67
SRS	1	within	NaN	NaN	NaN
SRS	1	between	0.3858	0.1326	0.59
SRS	1	random	0.446	0.1486	0.7705
TCP	21	within	0.809	0.1201	0.99
TCP	21	between	0.3539	0.143	0.59
TCP	21	random	0.3767	0.1847	0.69
Trihelix	13	within	0.6053	0.1998	0.98
Trihelix	13	between	0.4602	0.1525	0.72
Trihelix	13	random	0.4301	0.1507	0.71
Unknown	2	within	0.64	0	0.64
Unknown	2	between	0.3865	0.1549	0.69
Unknown	2	random	0.405	0.1454	0.66
WRKY	60	within	0.8766	0.1159	0.99
WRKY	60	between	0.4312	0.1259	0.65
WRKY	60	random	0.4106	0.1603	0.71
ZF-HD	6	within	0.8033	0.1723	0.98
ZF-HD	6	between	0.4146	0.1861	0.71
ZF-HD	6	random	0.3608	0.1751	0.64

Table S5. Proportional association of min-pCREs to FIT, non-FIT or mixed GS-enriched clusters. All min-pCREs identified in >3% (n = 5) GS-enriched clusters were included in the analysis, to be able to derive a preference of the min-pCRE for one of the categories. The min-pCRE proportions were normalized to the total number of FIT, non-FIT and mixed clusters. FIT: enriched in FIT-dependent genes. non-FIT: enriched in FIT-independent genes. Mixed: GS-enriched, but inconclusive (enriched in both FIT and non-FIT genes). Enrichment analysis: FET, q <0.05. Table shows process of calculation of normalized proportions as they were used for the ternary plot (**Figure 6B**).

Cluster counts: total GS-enriched = 159, FIT = 55, non-FIT = 45, mixed = 59 (see Supplemental Table S2)

ID (alphabetical)	min-pCRE	Total count	Total proportion (used for plot Figure 6B)	Counts			Proportions				Normalization factor	Normalized proportions (used for plot Figure 6B)			
		GS-enriched		FIT	non-FIT	mixed	FIT	non-FIT	mixed	Sum		FIT	non-FIT	mixed	Sum
122	AAACTA	7	4.40	1	0	6	0.02	0.00	0.10	0.12	8.34	0.15	0.00	0.85	1.00
152	AAAGTA	9	5.66	1	3	5	0.02	0.07	0.08	0.17	5.90	0.11	0.39	0.50	1.00
370	AACGGC	14	8.81	5	4	5	0.09	0.09	0.08	0.26	3.78	0.34	0.34	0.32	1.00
382	AACGTA	15	9.43	12	1	2	0.22	0.02	0.03	0.27	3.65	0.80	0.08	0.12	1.00
665	AATATG	8	5.03	0	7	1	0.00	0.16	0.02	0.17	5.80	0.00	0.90	0.10	1.00
721	AATGCG	6	3.77	0	3	3	0.00	0.07	0.05	0.12	8.51	0.00	0.57	0.43	1.00
793	AATTGC	13	8.18	0	7	6	0.00	0.16	0.10	0.26	3.89	0.00	0.60	0.40	1.00
905	ACACTA	9	5.66	3	1	5	0.05	0.02	0.08	0.16	6.19	0.34	0.14	0.52	1.00
916	ACAGCA	5	3.14	3	0	2	0.05	0.00	0.03	0.09	11.31	0.62	0.00	0.38	1.00
1011	ACCCAC	14	8.81	1	6	7	0.02	0.13	0.12	0.27	3.70	0.07	0.49	0.44	1.00
1078	ACGAAA	10	6.29	5	1	4	0.09	0.02	0.07	0.18	5.53	0.50	0.12	0.37	1.00
1134	ACGCAT	31	19.50	11	5	15	0.20	0.11	0.25	0.57	1.77	0.35	0.20	0.45	1.00
1148	ACGCGC	6	3.77	0	1	5	0.00	0.02	0.08	0.11	9.35	0.00	0.21	0.79	1.00
1178	ACGGCA	5	3.14	0	1	4	0.00	0.02	0.07	0.09	11.11	0.00	0.25	0.75	1.00
1316	ACTAGG	6	3.77	0	5	1	0.00	0.11	0.02	0.13	7.81	0.00	0.87	0.13	1.00
1318	ACTAGT	10	6.29	1	5	4	0.02	0.11	0.07	0.20	5.07	0.09	0.56	0.34	1.00
1324	ACTATA	5	3.14	3	1	1	0.05	0.02	0.02	0.09	10.67	0.58	0.24	0.18	1.00
1483	AGCATG	10	6.29	3	0	7	0.05	0.00	0.12	0.17	5.77	0.31	0.00	0.69	1.00
1558	AGGCAT	6	3.77	4	0	2	0.07	0.00	0.03	0.11	9.38	0.68	0.00	0.32	1.00
1604	AGTAAC	6	3.77	1	0	5	0.02	0.00	0.08	0.10	9.72	0.18	0.00	0.82	1.00
1625	AGTATA	5	3.14	0	4	1	0.00	0.09	0.02	0.11	9.45	0.00	0.84	0.16	1.00
1641	AGTCAA	8	5.03	8	0	0	0.15	0.00	0.00	0.15	6.88	1.00	0.00	0.00	1.00
1647	AGTCAAC	5	3.14	5	0	0	0.09	0.00	0.00	0.09	11.00	1.00	0.00	0.00	1.00
1704	AGTGGC	7	4.40	3	0	4	0.05	0.00	0.07	0.12	8.17	0.45	0.00	0.55	1.00
1779	ATAACC	6	3.77	2	1	3	0.04	0.02	0.05	0.11	9.14	0.33	0.20	0.46	1.00
1956	ATATAT	6	3.77	0	5	1	0.00	0.11	0.02	0.13	7.81	0.00	0.87	0.13	1.00
1967	ATATATT	5	3.14	4	1	0	0.07	0.02	0.00	0.09	10.53	0.77	0.23	0.00	1.00
2055	ATCGTG	10	6.29	8	0	2	0.15	0.00	0.03	0.18	5.58	0.81	0.00	0.19	1.00
2105	ATGCAT	13	8.18	12	1	0	0.22	0.02	0.00	0.24	4.16	0.91	0.09	0.00	1.00
2152	ATGTAA	5	3.14	0	1	4	0.00	0.02	0.07	0.09	11.11	0.00	0.25	0.75	1.00
2192	ATTAAG	15	9.43	0	6	9	0.00	0.13	0.15	0.29	3.50	0.00	0.47	0.53	1.00
2226	ATTATA	6	3.77	1	0	5	0.02	0.00	0.08	0.10	9.72	0.18	0.00	0.82	1.00
2253	ATTCGG	5	3.14	3	0	2	0.05	0.00	0.03	0.09	11.31	0.62	0.00	0.38	1.00
2530	CACGAG	8	5.03	0	4	4	0.00	0.09	0.07	0.16	6.38	0.00	0.57	0.43	1.00
2534	CACGCA	8	5.03	0	1	7	0.00	0.02	0.12	0.14	7.10	0.00	0.16	0.84	1.00
2541	CACGCC	15	9.43	7	5	3	0.13	0.11	0.05	0.29	3.46	0.44	0.38	0.18	1.00
2572	CACGTC	6	3.77	0	5	1	0.00	0.11	0.02	0.13	7.81	0.00	0.87	0.13	1.00
2609	CACTAC	10	6.29	1	6	3	0.02	0.13	0.05	0.20	4.94	0.09	0.66	0.25	1.00
2741	CATGCA	10	6.29	2	2	6	0.04	0.04	0.10	0.18	5.48	0.20	0.24	0.56	1.00
2748	CATGCC	48	30.19	31	2	15	0.56	0.04	0.25	0.86	1.16	0.65	0.05	0.29	1.00
2879	CCATAC	6	3.77	0	5	1	0.00	0.11	0.02	0.13	7.81	0.00	0.87	0.13	1.00
2887	CCATGC	19	11.95	17	0	2	0.31	0.00	0.03	0.34	2.92	0.90	0.00	0.10	1.00
2907	CCCACG	9	5.66	0	2	7	0.00	0.04	0.12	0.16	6.13	0.00	0.27	0.73	1.00
2980	CCGGCC	7	4.40	5	0	2	0.09	0.00	0.03	0.12	8.01	0.73	0.00	0.27	1.00
3006	CCTACG	5	3.14	0	4	1	0.00	0.09	0.02	0.11	9.45	0.00	0.84	0.16	1.00
3140	CGCTCA	12	7.55	0	2	10	0.00	0.04	0.17	0.21	4.67	0.00	0.21	0.79	1.00
3150	CGCTTA	5	3.14	0	2	3	0.00	0.04	0.05	0.10	10.49	0.00	0.47	0.53	1.00
3164	CGGCCA	8	5.03	2	0	6	0.04	0.00	0.10	0.14	7.24	0.26	0.00	0.74	1.00
3194	CGTAAC	5	3.14	0	4	1	0.00	0.09	0.02	0.11	9.45	0.00	0.84	0.16	1.00
3197	CGTAAG	6	3.77	3	1	2	0.05	0.02	0.03	0.11	9.04	0.49	0.20	0.31	1.00
3244	CGTGAC	21	13.21	12	1	8	0.22	0.02	0.14	0.38	2.66	0.58	0.06	0.36	1.00
3262	CGTGGA	16	10.06	4	1	11	0.07	0.02	0.19	0.28	3.55	0.26	0.08	0.66	1.00
3477	CTTATA	6	3.77	0	0	6	0.00	0.00	0.10	0.10	9.83	0.00	0.00	1.00	1.00
3623	GACAGC	5	3.14	4	0	1	0.07	0.00	0.02	0.09	11.15	0.81	0.00	0.19	1.00
3639	GACCAC	5	3.14	0	5	0	0.00	0.11	0.00	0.11	9.00	0.00	1.00	0.00	1.00
3733	GAGGGA	7	4.40	2	4	1	0.04	0.09	0.02	0.14	7.03	0.26	0.63	0.12	1.00
3742	GAGTCA	8	5.03	0	2	6	0.00	0.04	0.10	0.15	6.84	0.00	0.30	0.70	1.00
3763	GATAAC	6	3.77	3	2	1	0.05	0.04	0.02	0.12	8.63	0.47	0.38	0.15	1.00
3875	GCATGA	20	12.58	1	4	15	0.02	0.09	0.25	0.36	2.77	0.05	0.25	0.70	1.00
3919	GCCCCA	5	3.14	4	0	1	0.07	0.00	0.02	0.09	11.15	0.81	0.00	0.19	1.00
3981	GCGTGA	6	3.77	2	1	3	0.04	0.02	0.05	0.11	9.14	0.33	0.20	0.46	1.00
4017	GCTGAC	5	3.14	0	1	4	0.00	0.02	0.07	0.09	11.11	0.00	0.25	0.75	1.00
4060	GGAGCC	11	6.92	0	0	11	0.00	0.00	0.19	0.19	5.36	0.00	0.00	1.00	1.00
4207	GGTGAC	5	3.14	4	0	1	0.07	0.00	0.02	0.09	11.15	0.81	0.00	0.19	1.00
4262	GTATAA	10	6.29	3	0	7	0.05	0.00	0.12	0.17	5.77	0.31	0.00	0.69	1.00
4276	GTATTA	13	8.18	0	4	9	0.00	0.09	0.15	0.24	4.14	0.00	0.37	0.63	1.00
4292	GTCAAC	15	9.43	6	0	9	0.11	0.00	0.15	0.26	3.82	0.42	0.00	0.58	1.00
4310	GTCAGA	5	3.14	0	2	3	0.00	0.04	0.05	0.10	10.49	0.00	0.47	0.53	1.00
4330	GTCGGA	8	5.03	6	1	1	0.11	0.02	0.02	0.15	6.74	0.74	0.15	0.11	1.00
4349	GTGCAA	5	3.14	0	3	2	0.00	0.07	0.03	0.10	9.94	0.00	0.66	0.34	1.00
4354	GTGCCA	5	3.14	2	3	0	0.04	0.07	0.00	0.10	9.71	0.35	0.65	0.00	1.00
4370	GTGGCA	8	5.03	0	7	1	0.00	0.16	0.02	0.17	5.80	0.00	0.90	0.10	1.00
4410	GTTATA	9	5.66	0	5	4	0.00	0.11	0.07	0.18	5.59	0.00	0.62	0.38	1.00
4499	TAACTA	18	11.32	4	3	11	0.07	0.07	0.19	0.33	3.07	0.22	0.20	0.57	1.00
4580	TAATTA	8	5.03	0	4	4	0.00	0.09	0.07	0.16	6.38	0.00	0.57	0.43	1.00
4581	TAATTAA	13	8.18	6	3	4	0.11	0.07	0.07	0.24	4.11	0.45	0.27	0.28	1.00
4621	TACAGA	5	3.14	4	1	0	0.07	0.02	0.00	0.09	10.53	0.77	0.23	0.00	1.00
4659	TACGAA	6	3.77	3	0	3	0.05	0.00	0.05	0.11	9.49	0.52	0.00	0.48	1.00
4707	TACTAA	6	3.77	0	6	0	0.00	0.13	0.00	0.13	7.50	0.00	1.00	0.00	1.00
4844	TATAGA	8	5.03	8	0	0	0.15	0.00	0.00	0.15	6.88	1.00	0.00	0.00	1.00
4911	TATGCA	8	5.03	1	5	2	0.02	0.11	0.03	0.16	6.13	0.11	0.68	0.21	1.00
4912	TATGCAA	6	3.77	1	2	3	0.02	0.04	0.05	0.11	8.81	0.16	0.39	0.45	1.00
5225	TGACAGC	5	3.14	4	0	1	0.07	0.00	0.02	0.09	11.15	0.81	0.00	0.19	1.00

Table S6A. Total counts of each gene that have the mir-279 3' UTR element in the paired-as promoter. Example: if mir-61, this means that the gene was part of the coexpression clusters in which when the mir-279 3' was identified. These members are basis for **Figure S5**.

mir-279 3' UTR element	gene_name	Common Name	Coexpressed with mir-279 3' UTR element
mir-279 3' UTR element	WBGene00000001	aap-1	1
	WBGene00000002	aat-1	1
	WBGene00000003	aat-2	1
	WBGene00000004	abf-1	1
	WBGene00000005	abf-2	1
	WBGene00000006	abi-1	1
	WBGene00000007	abi-2	1
	WBGene00000008	abl-1	1
	WBGene00000009	abnormal	1
	WBGene00000010	abnormal	1
	WBGene00000011	abnormal	1
	WBGene00000012	abnormal	1
	WBGene00000013	abnormal	1

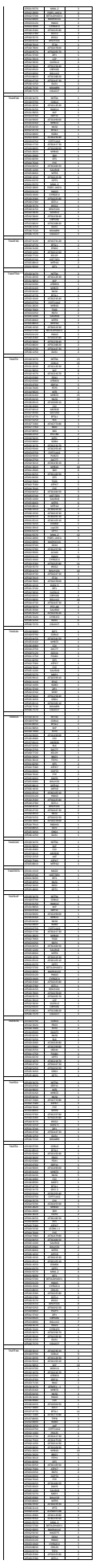


Table S10. Overlap of the piRNA, with Bos taurus/ovis CNE-Genes in all 50 centi-tech clusters in which the mir-piRNA was identified were checked for presence of the mir-piRNA in the putative promoter (500 bp upstream of TSS). The table shows for each mir-piRNA the set alignment amount in at most one bit and most base , the chromosome, strand/start, and orientation of the mir-piRNA in each promoter. Core count of each gene with mir-piRNA shown in **Table S10**. TSS-centred data were downloaded from the genome browser page (UCSC Gene Sorter) (www.genome.ucsc.edu) (August 2012, Bos taurus/bosTau4, and Ovis aries/oviAri1, Ensembl Genes track) using the Ensembl gene predictions. Primary gene symbols according to The Vertebrate Genome Nomenclature Committee (HUGO Gene Nomenclature Committee).

mir-piRNA ID	Genes with mir-piRNA in promoter	Unique Genes with mir-piRNA in promoter	mir-piRNA chromosome	mir-piRNA start position	mir-piRNA end position	mir-piRNA strand	TSS +/- 500 bp	mir-piRNA orientation	
								Sense	Anti-sense
piR_000000277	ADAM23	ADAM23	chr1	127,868,619	127,868,661	+	50	Yes	No

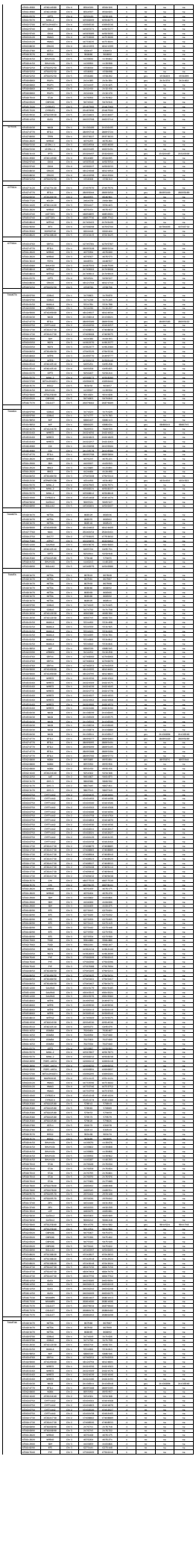
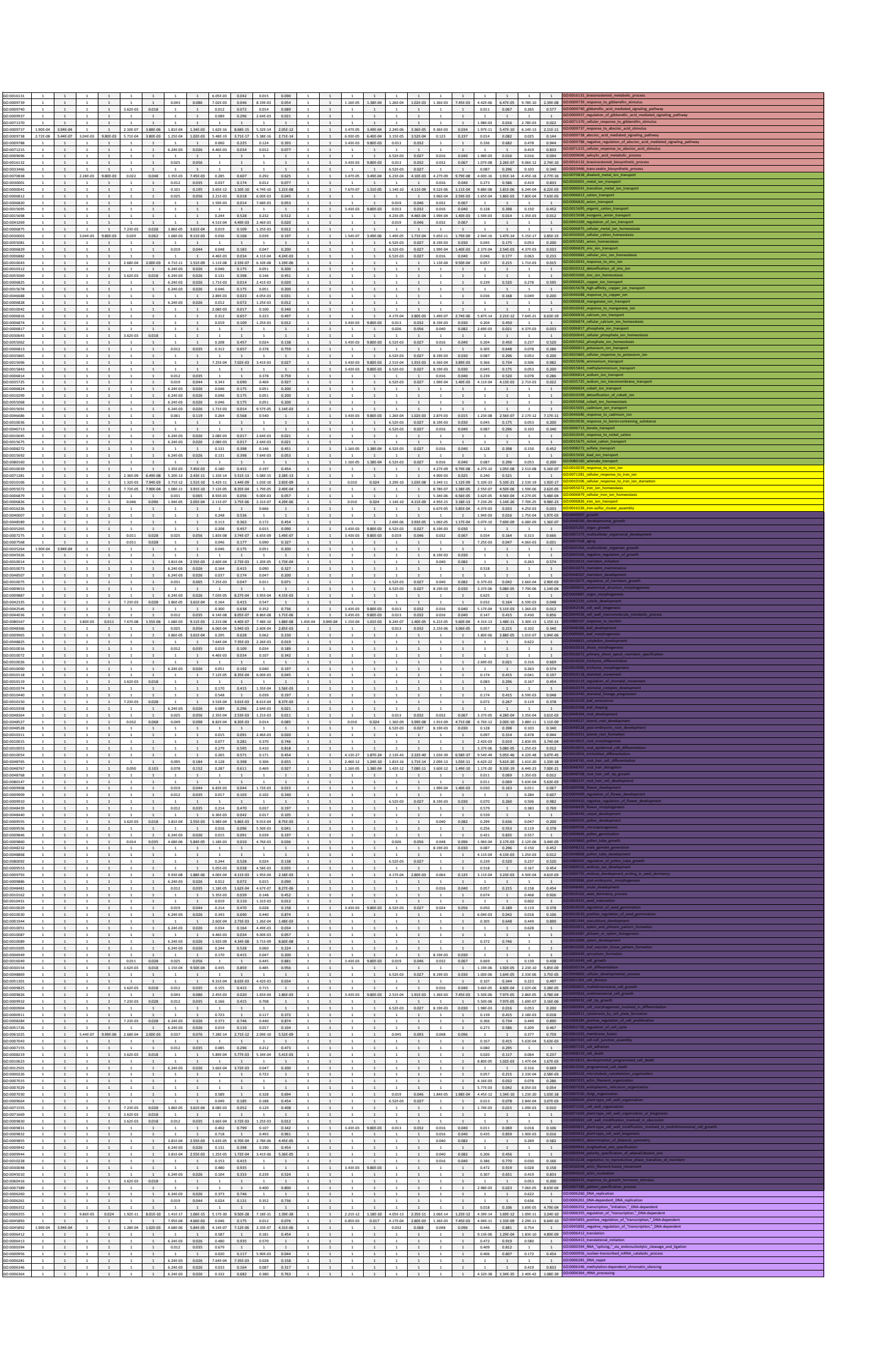


Table S7. *p* - and *q* -values (FET) of GO term enrichment test of -Fe-regulated genes. *q*- values: *p* -values corrected for multiple testing. -log10(q) if *q* <0.05 was calculated for heatmap representation in **Figure 1B** and **Supplemental Figure S1**.

	up										down											
-Fe treatment	3h		6h		12h		24h		48h		72h		3h		6h		12h		24h		48h		72h		
GO	*p*	*q*	*p*	*q*	*p*	*q*	*p*	*q*	*p*	*q*	*p*	*q*	*p*	*q*	*p*	*q*	*p*	*q*	*p*	*q*	*p*	*q*	*p*	*q*	GO description (colors as in Supplemental Figure S1)
GO:0006952	1.90E-04	3.94E-04	1.52E-03	9.80E-03	2.52E-06	3.76E-05	2.76E-05	2.72E-04	2.74E-16	1.38E-14	4.10E-15	1.85E-13	1	1	1	1	1	1	1	1	1.22E-03	0.011	7.68E-03	0.053	GO:0006952_defense_response
GO:0009814	1	1	1	1	0.011	0.028	0.019	0.044	5.41E-07	9.08E-06	3.13E-06	4.93E-05	1	1	1	1	1.26E-04	1.02E-03	1.99E-04	1.40E-03	4.72E-10	1.15E-08	9.51E-09	2.09E-07	GO:0009814_defense_response,_incompatible_interaction
GO:0031347	1	1	1	1	1	1	0.019	0.044	1.13E-07	2.03E-06	9.07E-12	2.74E-10	1	1	1	1	1	1	1	1	0.466	0.910	0.762	1	GO:0031347_regulation_of_defense_response
GO:0031348	1.90E-04	3.94E-04	3.63E-10	1.17E-08	3.57E-04	2.45E-03	1.94E-05	2.05E-04	5.04E-29	8.10E-27	1.03E-25	1.43E-23	1	1	1	1	0.051	0.106	1	1	3.32E-03	0.026	0.033	0.183	GO:0031348_negative_regulation_of_defense_response
GO:0045087	1	1	1	1	1.62E-03	0.018	0.012	0.035	1.18E-09	2.72E-08	1.03E-07	1.96E-06	1	1	1	1	1	1	1	1	1	1	1	0.520	GO:0045087_innate_immune_response
GO:0050776	1	1	1	1	1	1	1	1	2.11E-05	2.79E-04	7.20E-05	8.79E-04	1	1	1	1	1	1	1	1	1	1	1	1	GO:0050776_regulation_of_immune_response
GO:0002679	1.90E-04	3.94E-04	2.28E-03	9.80E-03	0.018	0.044	0.061	0.119	6.18E-34	1.39E-31	5.41E-31	1.09E-28	1	1	0.010	0.024	0.032	0.068	2.87E-03	0.015	0.756	1	0.584	1	GO:0002679_respiratory_burst_involved_in_defense_response
GO:0052542	1	1	1	1	1	1	1	1	1.07E-11	3.07E-10	1.44E-08	3.09E-07	1	1	1	1	1	1	1	1	0.098	0.316	0.143	0.450	GO:0052542_defense_response_by_callose_deposition
GO:0052544	1.90E-04	3.94E-04	7.62E-04	5.48E-03	7.23E-03	0.028	0.012	0.035	0.087	0.236	0.012	0.077	1	1	1	1	1	1	1	1	0.011	0.059	0.013	0.082	GO:0052544_defense_response_by_callose_deposition_in_cell_wall
GO:0009627	1	1	1	1	2.10E-07	3.88E-06	8.21E-05	7.15E-04	8.42E-25	9.84E-23	3.39E-27	5.35E-25	1	1	1	1	1.17E-04	9.87E-04	4.08E-06	5.19E-05	5.56E-03	0.041	0.044	0.200	GO:0009627_systemic_acquired_resistance
GO:0009682	1.90E-04	3.94E-04	1.52E-03	9.80E-03	3.87E-05	4.23E-04	2.30E-04	1.60E-03	1.35E-04	1.53E-03	1.74E-05	2.44E-04	1	1	6.85E-03	0.017	0.019	0.046	0.032	0.067	0.273	0.586	0.352	0.736	GO:0009682_induced_systemic_resistance
GO:0009862	1.90E-04	3.94E-04	3.63E-10	1.17E-08	2.68E-04	2.00E-03	1.51E-10	4.52E-09	1.44E-37	4.33E-35	2.45E-39	9.27E-37	1	1	1	1	0.045	0.093	1	1	2.88E-03	0.023	0.030	0.165	GO:0009862_systemic_acquired_resistance,_salicylic_acid_mediated_signaling_pathway
GO:0009866	1	1	1	1	3.62E-03	0.018	6.24E-03	0.026	0.089	0.296	0.100	0.340	1	1	1	1	1	1	1	1	1	1	1	1	GO:0009866_induced_systemic_resistance,_ethylene_mediated_signaling_pathway
GO:0009871	1	1	1	1	3.62E-03	0.018	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	GO:0009871_jasmonic_acid_and_ethylene-dependent_systemic_resistance,_ethylene_mediated_signaling_pathway
GO:0010112	1	1	1	1	3.62E-03	0.018	6.24E-03	0.026	8.83E-04	8.20E-03	2.41E-03	0.020	1	1	1	1	1	1	1	1	0.204	0.450	0.278	0.595	GO:0010112_regulation_of_systemic_acquired_resistance
GO:0002229	1	1	1	1	1	1	1	1	1	1	1	1	1	1	3.43E-03	9.80E-03	6.52E-03	0.027	8.19E-03	0.030	1.98E-03	0.016	0.025	0.144	GO:0002229_defense_response_to_oomycetes
GO:0002238	1	1	1	1	1	1	0.012	0.035	0.019	0.110	0.046	0.200	1	1	1	1	1	1	0.016	0.040	0.018	0.106	0.049	0.200	GO:0002238_response_to_molecule_of_fungal_origin
GO:0009610	1	1	1	1	1	1	1	1	1.85E-05	2.48E-04	8.00E-05	9.74E-04	1	1	1	1	1	1	1	1	1	1	1	1	GO:0009610_response_to_symbiotic_fungus
GO:0009620	1.90E-04	3.94E-04	1	1	0.014	0.035	1.35E-03	7.45E-03	9.20E-09	1.96E-07	1.11E-09	2.68E-08	1	1	1	1	0.026	0.056	2.31E-03	0.012	7.11E-03	0.046	0.022	0.130	GO:0009630_response_to_fungus
GO:0009817	1	1	1	1	0.011	0.028	0.025	0.056	4.79E-05	5.78E-04	7.05E-06	1.04E-04	1	1	1	1	0.019	0.046	0.032	0.067	0.158	0.415	0.049	0.200	GO:0009817_defense_response_to_"fungus,"_incompatible_interaction
GO:0050832	2.72E-08	5.44E-07	3.37E-12	1.39E-10	3.30E-11	1.32E-09	5.10E-08	1.02E-06	5.98E-37	1.73E-34	1.42E-31	2.98E-29	1	1	0.020	0.045	2.06E-04	1.60E-03	3.37E-04	2.32E-03	8.52E-06	1.20E-04	2.13E-04	2.37E-03	GO:0050832_defense_response_to_fungus
GO:0002237	1	1	1	1	0.014	0.035	0.019	0.044	4.89E-08	9.43E-07	2.22E-08	4.65E-07	1	1	3.43E-03	9.80E-03	0.026	0.056	0.024	0.056	1	1	0.729	1	GO:0002237_response_to_molecule_of_bacterial_origin
GO:0009617	1.90E-04	3.94E-04	1.52E-03	9.80E-03	1.58E-06	2.39E-05	9.31E-08	1.88E-06	1.57E-11	1.25E-11	2.84E-14	1.15E-12	1	1	6.85E-03	0.017	8.69E-04	5.38E-03	2.93E-05	2.88E-04	2.19E-05	2.38E-04	1.58E-04	1.79E-03	GO:0009617_response_to_bacterium
GO:0009816	1	1	1	1	1.29E-05	1.51E-04	1.15E-04	9.50E-04	4.79E-05	5.78E-04	2.28E-05	3.06E-04	1	1	1	1	1	1	1	1	0.158	0.415	0.506	0.982	GO:0009816_defense_response_to_"bacterium,"_incompatible_interaction
GO:0016045	1	1	1	1	1	1	1	1	2.11E-05	2.79E-04	1.26E-04	1.48E-03	1	1	1	1	1	1	1	1	1	1	1	1	GO:0016045_detection_of_bacterium
GO:0042742	1.90E-04	3.94E-04	3.63E-10	1.17E-08	1.61E-05	1.85E-04	3.23E-06	4.21E-05	1.94E-23	2.03E-21	2.41E-22	2.38E-20	1	1	1	1	9.38E-05	9.51E-04	1.99E-07	3.47E-06	2.76E-10	6.96E-09	8.11E-11	2.23E-09	GO:0042742_defense_response_to_bacterium
GO:0009611	1.90E-04	3.94E-04	1.27E-08	3.59E-07	2.30E-09	6.46E-08	1.71E-11	5.99E-10	1.20E-29	6.71E-27	9.90E-31	1.87E-28	1	1	0.024	0.052	2.97E-06	4.25E-05	5.86E-06	7.13E-05	2.75E-11	7.48E-10	4.97E-14	1.93E-12	GO:0009611_response_to_wounding
GO:0009624	1	1	1	1	0.025	0.056	2.76E-05	2.72E-04	5.49E-06	7.97E-05	7.87E-09	1.73E-07	1	1	0.010	0.024	8.69E-04	5.38E-03	6.21E-05	5.60E-04	2.23E-04	2.43E-03	4.02E-05	5.18E-04	GO:0009624_response_to_nematode
GO:0002213	1.90E-04	3.94E-04	7.62E-04	5.48E-03	3.62E-03	0.018	0.012	0.035	0.029	0.154	0.032	0.178	1	1	1	1	1	1	1	1	3.60E-03	0.028	5.44E-04	5.49E-03	GO:0002213_defense_response_to_insect
GO:0009625	1	1	1	1	1	1	1	1	1.17E-09	2.70E-08	1.28E-10	3.48E-09	1	1	3.43E-03	9.80E-03	6.52E-03	0.027	0.016	0.040	0.560	1	1	1	GO:0009625_response_to_insect
GO:0009595	1	1	5.44E-07	9.99E-06	0.014	0.035	0.025	0.056	9.24E-14	3.39E-12	5.70E-15	2.53E-13	1	1	1	1	1	1	1	1	2.53E-03	0.020	0.023	0.134	GO:0009595_detection_of_biotic_stimulus
GO:0051707	1	1	1	1	7.72E-05	7.80E-04	0.031	0.065	1.63E-12	5.22E-11	5.85E-12	1.81E-10	1	1	1	1	0.026	0.056	0.040	0.082	0.227	0.497	0.120	0.381	GO:0051707_response_to_other_organism
GO:0010200	2.72E-08	5.44E-07	3.37E-12	1.39E-10	6.83E-14	4.32E-12	1.40E-05	1.52E-04	9.18E-60	8.26E-57	6.25E-59	6.57E-56	1	1	0.020	0.045	0.099	0.204	0.016	0.040	2.85E-03	0.023	0.089	0.323	GO:0010200_response_to_chitin
GO:0001666	1	1	1	1	1	1	2.15E-08	4.56E-07	3.40E-16	1.65E-14	1.58E-19	1.22E-17	1	1	1	1	1	1	1	1	0.180	0.415	0.444	0.881	GO:0001666_response_to_hypoxia
GO:0071456	1.90E-04	3.94E-04	3.63E-09	1.06E-07	4.89E-17	4.86E-15	2.15E-08	4.56E-07	6.36E-15	2.73E-13	4.45E-13	1.54E-11	1	1	0.017	0.038	0.051	0.106	0.048	0.096	1	1	1	1	GO:0071456_cellular_response_to_hypoxia
GO:0070481	1	1	1	1	1	1	6.24E-03	0.026	8.91E-03	0.042	7.64E-03	0.053	1	1	1	1	1	1	1	1	0.128	0.398	0.150	0.452	GO:0070481_detection_of_hypoxia
GO:0034059	1	1	1	1	1	1	1	1	0.046	0.175	0.051	0.200	1	1	3.43E-03	9.80E-03	6.52E-03	0.027	8.19E-03	0.030	1	1	1	1	GO:0034059_response_to_anoxia
GO:0071454	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1.98E-03	0.016	2.78E-03	0.022	GO:0071454_cellular_response_to_anoxia
GO:0009670	1	1	5.44E-07	9.80E-03	1	1	1.68E-03	9.11E-03	1.06E-04	3.19E-03	6.76E-04	6.69E-03	1	1	1	1	5.98E-08	1.29E-06	8.78E-07	1.38E-05	9.15E-06	1.29E-04	3.18E-08	8.50E-07	GO:0009670_response_to_osmotic_stress
GO:0006972	1	1	1	1	1	1	1	1	0.482	0.939	0.740	1	1	1	1	1	6.52E-03	0.027	0.032	0.067	6.39E-06	9.23E-05	1.42E-07	2.66E-06	GO:0006972_hyperosmotic_response
GO:0042538	1.90E-04	3.94E-04	1	1	1.32E-04	1.50E-03	1.35E-03	7.45E-03	9.90E-15	2.15E-13	9.71E-12	2.93E-10	1	1	1	1	1	1	0.071	0.139	0.014	0.086	0.076	0.277	GO:0042538_hyperosmotic_salinity_response
GO:0071475	1	1	1	1	1	1	6.24E-03	0.026	0.046	0.175	0.051	0.200	1	1	1	1	1	1	1	1	1	1	1	1	GO:0071475_cellular_hyperosmotic_salinity_response
GO:0009651	1.90E-04	3.94E-04	3.80E-03	0.011	1.06E-09	3.26E-08	3.35E-04	2.31E-03	2.72E-06	4.16E-05	2.15E-05	2.89E-04	1	1	0.017	0.038	1.99E-10	6.38E-09	3.36E-10	9.49E-09	1.68E-22	1.63E-20	8.43E-35	2.04E-32	GO:0009651_response_to_salt_stress
GO:0009744	1	1	1	1	3.57E-04	2.45E-03	0.043	0.086	808	1	835	1	1	1	1	1	0.051	0.106	1.84E-05	1.98E-04	3.27E-34	7.75E-32	9.40E-37	3.17E-34	GO:0009744_response_to_sucrose_stimulus
GO:0009749	1	1	1	1	7.72E-05	7.80E-04	0.031	0.065	1.63E-12	5.22E-11	5.85E-12	1.81E-10	1	1	1	1	0.026	0.056	0.040	0.082	0.227	0.497	0.120	0.381	GO:0009749_response_to_glucose_stimulus
GO:0009750	1	1	1	1	1	1	1	1	0.338	0.682	1	1	1	1	1	1	1	1	0.026	0.056	6.44E-08	1.24E-06	3.25E-10	8.09E-09	1.35E-14	5.70E-13	GO:0009750_response_to_fructose_stimulus
GO:0043617	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	2.78E-03	0.022	GO:0043617_cellular_response_to_sucrose_starvation
GO:0009416	1	1	1	1	1	1	0.037	0.076	1.90E-03	0.016	5.06E-04	5.15E-03	1	1	3.43E-03	9.80E-03	1.26E-04	1.02E-03	9.79E-04	5.68E-03	1.17E-10	3.02E-09	3.02E-12	9.79E-11	GO:0009416_response_to_light_stimulus
GO:0009411	1	1	1	1	1	1	0.012	0.035	2.16E-03	0.017	2.88E-03	0.023	1	1	1	1	1	1	1	1	6.58E-09	1.42E-07	9.10E-10	2.25E-08	GO:0009411_response_to_UV
GO:0010224	1	1	1	1	1.92E-04	1.50E-03	0.019	0.044	0.137	0.415	0.099	0.340	1	1	1	1	1	1	0.024	0.056	1.42E-22	3.14E-20	7.32E-24	8.66E-22	GO:0010224_response_to_UV-B
GO:0009642	1	1	1	1	1	1	1	1	0.279	0.595	0.309	0.657	1	1	1	1	1	1	1	1	2.69E-03	0.021	4.37E-03	0.033	GO:0009642_response_to_light_intensity
GO:0009644	1.90E-04	3.94E-04	1	1	1	1	0.019	0.044	2.31E-04	2.50E-03	7.57E-04	7.45E-03	1	1	3.43E-03	9.80E-03	0.032	0.068	0.024	0.056	2.14E-11	5.91E-10	9.14E-11	2.51E-09	GO:0009644_response_to_high_light_intensity
GO:0009637	1	1	1	1	1	1	0.012	0.035	1	1	1	1	1	1	1	1	1	1	1	1	1.61E-25	1.99E-23	4.31E-24	5.29E-22	GO:0009637_response_to_blue_light
GO:0009639	1	1	1	1	1	1	1	1	5.05E-03	0.018	6.65E-04	6.59E-03	1	1	1	1	1	1	1	1	0.518	1	0.325	0.888	GO:0009639_response_to_red_or_far_red_light
GO:0010218	1	1	1	1	1	1	0.019	0.044	0.087	0.296	0.206	0.460	1	1	1	1	1	1	1	1	1.94E-27	2.64E-25	2.77E-24	3.49E-22	GO:0010218_response_to_far_red_light
GO:0010114	1	1	1	1	1	1	0.031	0.065	0.312	0.657	0.542	1	1	1	1	1	1	1	1	1	1.04E-29	1.80E-27	8.42E-29	1.42E-26	GO:0010114_response_to_red_light
GO:0043481	1	1	1	1	7.23E-03	0.028	0.019	0.044	0.108	0.347	0.305	0.652	1	1	1	1	1	1	1	1	1.75E-03	0.015	1.39E-04	1.61E-03	GO:0043481_anthocyanin_accumulation_in_tissues_in_response_to_UV_light
GO:0071482	1	1	1	1	1	1	1	1	0.028	0.149	0.034	0.189	1	1	1	1	1	1	1	1	0.026	0.144	1.06E-04	1.25E-03	GO:0071482_cellular_response_to_light_stimulus
GO:0071483	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	3.40E-04	3.49E-03	7.71E-06	1.13E-04	GO:0071483_cellular_response_to_blue_light
GO:0071486	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	8.22E-04	7.75E-03	1.35E-03	0.012	GO:0071486_cellular_response_to_high_light_intensity
GO:0071490	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1.98E-03	0.016	2.78E-03	0.022	GO:0071490_cellular_response_to_far_red_light
GO:0071491	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1.98E-03	0.016	2.78E-03	0.022	GO:0071491_cellular_response_to_red_light
GO:0071492	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1.40E-04	1.49E-03	5.61E-04	5.63E-03	GO:0071492_cellular_response_to_UV-A
GO:2000030	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	8.19E-03	0.030	0.087	0.296	0.103	0.340	GO:2000030_regulation_of_response_to_red_or_far_red_light
GO:0009638	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0.018	0.106	7.71E-06	1.13E-04	GO:0009638_phototropism
GO:0009646	1	1	7.62E-04	5.48E-03	3.62E-03	0.018	1	1	3.11E-05	3.99E-04	0.014	0.089	1	1	1	1	1	1	1	1	0.495	0.964	0.016	0.096	GO:0009646_response_to_absence_of_light
GO:0009266	1	1	7.62E-04	5.48E-03	0.011	0.028	1	1	0.077	0.282	0.196	0.454	1	1	1	1	0.019	0.046	0.032	0.067	9.74E-06	1.36E-04	4.15E-07	7.41E-06	GO:0009266_response_to_temperature_stimulus
GO:0009408	1	1	1.52E-03	9.80E-03	0.056	2.35E-03	3.38E-05	2.32E-04	5.23E-08	2.58E-07	6.03E-08	1.68E-06	1	1	6.85E-03	0.017	0.045	0.093	0.040	0.082	0.014	0.023	1.11E-03	0.024	GO:0009408_response_to_heat
GO:0009409	2.72E-08	5.44E-07	5.44E-07	9.99E-06	5.13E-08	1.13E-06	3.62E-08	7.37E-07	2.14E-13	7.63E-12	1.06E-14	4.62E-13	1	1	1	1	4.31E-05	4.50E-04	2.80E-04	1.93E-03	2.38E-19	1.70E-17	4.71E-23	5.24E-21	GO:0009409_response_to_cold
GO:0010286	1	1	1	1	1	1	1	1	6.40E-09	1.39E-07	1.82E-08	3.84E-07	1	1	3.43E-03	9.80E-03	1	1	1	1	0.087	0.296	0.044	0.200	GO:0010286_heat_acclimation
GO:0034605	1	1	7.62E-04	5.48E-03	1	1	1	1	1.04E-06	1.70E-05	1.21E-06	2.03E-05	1	1	1	1	0.019	0.046	8.19E-03	0.030	1.69E-05	2.28E-04	2.65E-05	3.53E-04	GO:0034605_cellular_response_to_heat
GO:0070370	1	1	7.62E-04	5.48E-03	1	1	1	1	0.012	0.072	5.20E-04	5.28E-03	1	1	1	1	1	1	1	1	0.167	0.415	0.195	0.454	GO:0070370_cellular_heat_acclimation
GO:0070417	1	1	1	1	1	1	1	1	7.25E-04	7.02E-03	7.09E-04	6.99E-03	1	1	1	1	1	1	1	1	6.61E-04	6.43E-03	7.84E-04	7.63E-03	GO:0070417_cellular_response_to_cold
GO:0006995	1	1	1	1	1	1	1	1	0.029	0.154	0.178	0.454	1	1	1	1	0.013	0.032	0.024	0.056	2.71E-05	3.50E-04	7.80E-04	7.63E-03	GO:0006995_cellular_response_to_nitrogen_starvation
GO:0010438	1	1	1	1	1	1	0.012	0.035	0.089	0.296	0.100	0.340	1	1	3.43E-03	9.80E-03	6.52E-03	0.027	0.016	0.040	1	1	0.103	0.340	GO:0010438_cellular_response_to_sulfur_starvation
GO:0010167	1	1	1	1	0.029	0.062	0.072	0.141	0.016	0.094	0.018	0.108	1	1	1	1	1.15E-03	6.98E-03	0.094	0.182	4.18E-07	7.17E-06	4.15E-07	7.41E-06	GO:0010167_response_to_phosphate_starvation
GO:0009269	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	5.01E-04	4.97E-03	1.59E-06	2.62E-05	GO:0009269_response_to_desiccation
GO:0009414	1.90E-04	3.94E-04	3.04E-03	9.80E-03	5.71E-04	3.80E-03	2.48E-06	3.31E-05	2.22E-18	1.40E-16	2.37E-15	1.11E-13	1	1	6.93E-05	6.40E-04	3.15E-05	3.52E-04	2.80E-04	1.93E-03	1.04E-09	2.44E-08	1.35E-10	3.66E-09	GO:0009414_response_to_water_deprivation
GO:0009415	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	8.22E-04	7.75E-03	3.69E-05	4.79E-04	GO:0009415_response_to_water
GO:0080148	1	1	1	1	3.62E-03	0.018	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0.053	0.200	GO:0080148_negative_regulation_of_response_to_water_deprivation
GO:0009652	1	1	7.62E-04	5.48E-03	3.62E-03	0.018	1	1	0.046	0.175	0.100	0.340	1	1	1	1	1	1	1	1	1	1	0.103	0.340	GO:0009652_thigmotropism
GO:0009958	1	1	1	1	3.62E-03	0.018	6.24E-03	0.026	0.019	0.109	0.047	0.200	1	1	1	1	1	1	1	1	0.018	0.106	0.049	0.200	GO:0009958_positive_gravitropism
GO:0009410	1	1	1	1	3.62E-03	0.018	6.24E-03	0.026	0.085	0.296	6.10E-03	0.045	1	1	1	1	1	1	1	1	0.080	0.295	0.028	0.158	GO:0009410_response_to_xenobiotic_stimulus
GO:0009612	1.90E-04	3.94E-04	1.52E-03	9.80E-03	0.011	0.028	1	1	5.65E-13	1.92E-11	5.78E-12	1.79E-10	1	1	1	1	0.019	0.046	0.024	0.056	0.389	0.776	0.678	1	GO:0009612_response_to_mechanical_stimulus
GO:0009970	1	1	1	1	1	1	1	1	0.244	0.528	0.232	0.512	1	1	3.43E-03	9.80E-03	6.52E-03	0.027	1	1	1.59E-03	0.014	1.35E-03	0.012	GO:0009970_cellular_response_to_sulfate_starvation
GO:0010167	1	1	1	1	1.92E-04	1.50E-03	1.29E-06	2.32E-05	5.35E-07	5.88E-06	1.48E-07	3.30E-06	1	1	1.35E-15	1.53E-13	1.22E-20	2.81E-18	1.30E-27	5.65E-25	1.69E-29	3.31E-27	4.57E-34	6.55E-32	GO:0010167_response_to_nitrate
GO:0010363	1.90E-04	3.94E-04	3.37E-12	1.39E-10	5.78E-05	5.98E-04	5.55E-04	3.37E-03	1.70E-35	4.36E-33	3.11E-30	5.55E-28	1	1	0.020	0.045	2.06E-07	3.87E-06	1.99E-06	2.92E-05	4.65E-10	1.14E-08	3.42E-08	6.95E-07	GO:0010363_regulation_of_plant-type_hypersensitive_response
GO:0010583	1	1	1.52E-03	9.80E-03	3.57E-04	2.45E-03	1.94E-05	2.05E-04	4.07E-18	2.49E-16	3.79E-17	2.14E-15	1	1	1	1	1	1	1	1	3.34E-04	3.45E-03	1.17E-03	0.011	GO:0010583_response_to_cyclopentenone
GO:0033591	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	6.52E-03	0.027	1	1	0.167	0.415	0.016	0.094	GO:0033591_response_to_L-ascorbic_acid
GO:0046685	1	1	1	1	1	1	6.24E-03	0.026	0.089	0.175	0.039	0.197	1	1	1	1	1	1	1	1	0.174	0.415	0.186	0.454	GO:0046685_response_to_arsenic-containing_substance
GO:0030968	1.90E-04	3.94E-04	7.62E-04	5.48E-03	1.28E-



GO:0016246	1	1	1	1	1	1	6.24E-03	0.026	0.061	0.225	0.123	0.391	1	1	1	1	1	1	1	1	0.598	1	0.616	1	GO:0016246_RNA_interference
GO:0031048	1	1	1	1	1	1	6.24E-03	0.026	0.077	0.281	0.147	0.452	1	1	1	1	1	1	1	1	1	1	0.400	0.799	GO:0031048_chromatin_silencing_by_small_RNA
GO:0042254	1	1	1	1	1	1	1	1	0.647	1	0.277	0.595	1	1	1	1	1	1	1	1	2.81E-04	2.93E-03	2.82E-04	3.07E-03	GO:0042254_ribosome_biogenesis
GO:0051225	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0.127	0.398	1.95E-03	0.017	GO:0051225_spindle_assembly
GO:0051567	1	1	1	1	1	1	6.24E-03	0.026	0.166	0.415	1	1	1	1	1	1	1	1	1	1	0.648	1	1	1	GO:0051567_histone_H3-K9_methylation
GO:2001141	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	5.77E-03	0.042	1.47E-04	1.67E-03	GO:2001141_regulation_of_RNA_biosynthetic_process
GO:0042991	1	1	1	1	1	1	1	1	5.42E-07	9.09E-06	1.31E-06	2.17E-05	1	1	1	1	1	1	1	1	1	1	0.622	1	GO:0042991_transcription_factor_import_into_nucleus
GO:0043900	1	1	5.44E-07	9.99E-06	0.014	0.035	0.025	0.056	8.50E-11	2.21E-09	6.00E-12	1.84E-10	1	1	1	1	1	1	1	1	8.68E-04	8.31E-03	0.011	0.069	GO:0043900_regulation_of_multi-organism_process
GO:0048519	1	1	1	1	1	1	6.24E-03	0.026	0.101	0.326	9.64E-03	0.061	1	1	1	1	1	1	1	1	1	1	1	1	GO:0048519_negative_regulation_of_biological_process
GO:0045088	1	1	7.62E-04	5.48E-03	3.62E-03	0.018	6.24E-03	0.026	3.07E-07	5.37E-06	1.42E-08	3.06E-07	1	1	1	1	1	1	1	1	1	1	0.623	1	GO:0045088_regulation_of_innate_immune_response
GO:0006792	1	1	1	1	1	1	6.24E-03	0.026	0.046	0.175	0.051	0.200	1	1	1	1	1	1	1	1	1	1	1	1	GO:0006792_regulation_of_sulfur_utilization
GO:0008361	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0.015	0.089	1.75E-04	1.97E-03	GO:0008361_regulation_of_cell_size
GO:0010155	1	1	1	1	1	1	1	0.264	0.568	0.727	1	1	1	1	1	1	1	1	1	1	4.56E-22	4.06E-20	4.73E-21	4.22E-19	GO:0010155_regulation_of_proton_transport
GO:0010439	1	1	1	1	1	1	1	1	1	1	1	1	1	3.43E-03	9.80E-03	6.52E-03	0.027	8.19E-03	0.030	1	1	1	1	1	GO:0010439_regulation_of_glucosinolate_biosynthetic_process
GO:0010508	1	1	7.62E-04	5.48E-03	1	1	1	0.089	0.296	0.100	0.340	1	1	1	1	1	1	1	1	1	0.087	0.296	0.103	0.340	GO:0010508_positive_regulation_of_autophagy
GO:0010564	1	1	1	1	1	1	1	4.46E-03	0.014	0.060	0.324	1	1	1	1	1	1	1	1	1	1	1	1	1	GO:0010564_regulation_of_cell_cycle_process
GO:0010951	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	6.52E-03	0.027	8.19E-03	0.030	0.128	0.398	0.103	0.340	GO:0010951_negative_regulation_of_endopeptidase_activity
GO:0019216	1	1	1	1	1	1	1	0.312	0.657	0.271	0.584	1	1	1	1	1	1	1	1	1	2.37E-04	2.54E-03	1.06E-04	1.25E-03	GO:0019216_regulation_of_lipid_metabolic_process
GO:0030162	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	6.52E-03	0.027	8.19E-03	0.030	0.128	0.398	0.103	0.340	GO:0030162_regulation_of_proteolysis
GO:0032880	1	1	1	1	3.62E-03	0.018	6.24E-03	0.026	0.373	0.746	0.410	0.818	1	1	1	1	1	1	1	1	1	1	1	1	GO:0032880_regulation_of_protein_localization
GO:0042752	1	1	1	1	1	1	1	1	0.034	0.164	2.26E-03	0.019	1	3.43E-03	9.80E-03	6.52E-03	0.027	8.19E-03	0.030	4.63E-03	0.035	0.016	0.096	GO:0042752_regulation_of_circadian_rhythm
GO:0043067	1	1	1	1	1	1	6.24E-03	0.026	0.089	0.296	0.146	0.451	1	1	1	1	1	1	1	1	1	1	1	1	GO:0043067_regulation_of_programmed_cell_death
GO:0043069	1.90E-04	3.94E-04	1.45E-09	4.38E-08	1.27E-05	1.49E-04	1.02E-04	8.69E-04	4.27E-28	6.20E-26	6.24E-25	8.32E-23	1	0.014	0.031	1.36E-08	3.14E-07	5.53E-06	6.76E-05	0.287	0.611	0.345	0.727	GO:0043069_negative_regulation_of_programmed_cell_death
GO:0043085	1	1	1	1	1	1	3.86E-05	3.61E-04	0.400	0.798	1	1	1	1	1	1	1	1	1	1	2.81E-04	2.93E-03	1.01E-05	1.48E-04	GO:0043085_positive_regulation_of_catalytic_activity
GO:0043433	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1.98E-03	0.016	2.78E-03	0.022	GO:0043433_negative_regulation_of_sequence-specific_DNA_binding_transcription_factor_activity
GO:0043496	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1.98E-03	0.016	2.78E-03	0.022	GO:0043496_regulation_of_protein_homodimerization_activity
GO:0051453	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	8.19E-03	0.030	0.045	0.175	0.053	0.200	GO:0051453_regulation_of_intracellular_pH
GO:0080153	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1.98E-03	0.016	2.78E-03	0.022	GO:0080153_regulation_of_reductive_pentose-phosphate_cycle
GO:0010581	1	1	1	1	1	1	1	0.089	0.296	0.100	0.340	1	1	3.43E-03	9.80E-03	6.52E-03	0.027	1	1	0.087	0.296	0.103	0.340	GO:0010581_regulation_of_starch_biosynthetic_process
GO:0080090	1	1	1	1	3.62E-03	0.018	6.24E-03	0.026	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	GO:0080090_regulation_of_primary_metabolic_process
GO:0080091	1	1	1	1	3.62E-03	0.018	6.24E-03	0.026	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	GO:0080091_regulation_of_raffinose_metabolic_process
GO:0006457	1	1	1	1	1	1	1.15E-04	9.50E-04	9.39E-08	1.73E-06	4.68E-07	8.27E-06	1	1	1	0.026	0.056	0.024	0.056	0.091	0.296	0.101	0.340	GO:0006457_protein_folding
GO:0006464	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	6.52E-03	0.027	8.19E-03	0.030	0.239	0.520	0.316	0.669	GO:0006464_protein_modification_process
GO:0006468	1	1	1	1	1	1	0.043	0.086	7.00E-05	8.34E-04	1.58E-06	5.59E-05	1	6.85E-03	0.017	1	1	1	1	0.012	0.075	0.011	0.069	GO:0006468_protein_phosphorylation
GO:0006486	1	1	1	1	1	1	1	1	0.020	0.117	2.46E-03	0.020	1	3.43E-03	9.80E-03	6.52E-03	0.027	0.024	0.056	1	1	0.722	1	GO:0006486_protein_glycosylation
GO:0006487	1	1	1	1	1	1	1	1	1	1	0.172	0.454	1	3.43E-03	9.80E-03	6.52E-03	0.027	8.19E-03	0.030	1	1	1	1	GO:0006487_protein_N-linked_glycosylation
GO:0006499	1	1	1	1	1	1	1	1	1.16E-06	1.88E-05	7.55E-07	1.31E-05	1	1	1	1	1	1	1	1	0.725	1	0.520	1	GO:0006499_N-terminal_protein_myristoylation
GO:0006508	1	1	1	1	1	1	2.30E-04	1.60E-03	6.97E-03	0.045	0.017	0.102	1	6.85E-03	0.017	0.019	0.046	0.032	0.067	0.018	0.106	1.81E-04	2.03E-03	GO:0006508_proteolysis
GO:0010731	1	1	1	1	3.62E-03	0.018	6.24E-03	0.026	0.046	0.175	0.100	0.340	1	1	1	1	1	1	1	1	1	1	1	1	GO:0010731_protein_glutathionylation
GO:0016567	1	1	1	1	1	1	1	1	1.78E-03	0.015	1.21E-03	0.011	1	1	1	1	6.52E-03	0.027	1	1	0.727	1	0.754	1	GO:0016567_protein_ubiquitination
GO:0018171	1	1	1	1	1	1	6.24E-03	0.026	2.08E-03	0.017	2.64E-03	0.021	1	1	1	1	1	1	1	1	1	1	1	1	GO:0018171_peptidyl-cysteine_oxidation
GO:0035304	1	1	1	1	1	1	3.86E-05	3.61E-04	8.40E-03	0.053	7.79E-03	0.054	1	1	1	1	1	1	1	1	1.64E-15	7.44E-14	2.75E-14	1.11E-12	GO:0035304_regulation_of_protein_dephosphorylation
GO:0043246	1	1	1	1	1	1	1	1	1.31E-03	0.026	1.50E-03	0.041	1	1	1	1	1	1	1	1	0.624	1	0.632	1	GO:0043246_proteasome_assembly
GO:0046777	1	1	1	1	1	1	0.012	0.035	0.077	0.282	0.071	0.263	1	1	1	1	1	1	1	1	9.74E-06	1.36E-04	3.42E-05	4.48E-04	GO:0046777_protein_autophosphorylation
GO:0006612	1.90E-04	3.94E-04	3.17E-12	1.39E-10	5.78E-05	5.98E-04	5.55E-04	3.37E-03	1.70E-35	4.36E-33	3.11E-30	5.55E-28	1	0.020	0.045	2.06E-07	3.87E-06	1.99E-06	2.92E-05	4.65E-10	1.14E-08	3.42E-08	6.95E-07	GO:0006612_protein_targeting_to_membrane
GO:0006626	1	1	1	1	1	1	1	0.018	0.106	5.83E-04	5.83E-03	1	1	1	1	1	1	1	1	0.404	0.803	0.266	0.579	GO:0006626_protein_targeting_to_mitochondrion
GO:0045036	1	1	1	1	1	1	1	1	1	1	0.633	1	1	1	1	1	1	1	1	5.84E-03	0.042	0.035	0.191	GO:0045036_protein_targeting_to_chloroplast
GO:0006810	1	1	1	1	1	1	1	1	0.033	0.164	4.02E-03	0.031	1	3.47E-05	3.49E-04	4.17E-04	2.80E-03	9.79E-04	5.68E-03	6.20E-13	2.09E-11	1.68E-13	6.29E-12	GO:0006810_transport
GO:0016192	1	1	1	1	1	1	1	0.262	0.565	0.521	1	1	1	1	1	1	1	0.032	0.067	7.11E-03	0.046	0.051	0.200	GO:0016192_vesicle-mediated_transport
GO:0055085	1	1	1	1	1	1	1	2.89E-03	0.023	6.24E-03	0.046	1	1	1	1	1	1	1	1	1	1	1	1	GO:0055085_transmembrane_transport
GO:0006897	1	1	1	1	1	1	1	0.170	0.415	0.232	0.512	1	1	1	1	6.52E-03	0.027	8.19E-03	0.030	0.167	0.415	0.237	0.520	GO:0006897_endocytosis
GO:0006857	1	1	1	1	3.62E-03	0.018	6.24E-03	0.026	0.046	0.175	0.100	0.340	1	1	1	1	1	1	1	1	2.57E-05	3.45E-04	8.79E-06	1.28E-04	GO:0006857_oligopeptide_transport
GO:0006865	1.90E-04	3.94E-04	3.04E-03	9.80E-03	4.58E-04	3.06E-03	4.38E-03	0.023	1.84E-15	8.29E-14	4.31E-16	2.19E-14	1	6.93E-05	6.40E-04	2.22E-05	2.53E-04	2.07E-09	5.06E-08	6.64E-03	0.043	0.012	0.075	GO:0006865_amino_acid_transport
GO:0015802	1	1	1	1	1	1	1	7.25E-04	7.02E-03	3.41E-03	0.027	1	1	1	1	0.019	0.046	0.032	0.067	1	1	1	1	GO:0015802_basic_amino_acid_transport
GO:0015824	1	1	1	1	4.53E-07	7.82E-06	5.70E-04	3.44E-03	1.77E-08	3.62E-07	6.01E-07	1.05E-05	1	1	1	1	1	0.048	0.096	0.127	0.398	0.424	0.844	GO:0015824_proline_transport
GO:0043090	1	1	7.62E-04	5.48E-03	0.014	0.035	0.043	0.086	3.44E-08	6.73E-07	1.65E-09	3.96E-08	1	1	1	0.026	0.056	0.056	0.110	0.409	0.812	0.263	0.574	GO:0043090_amino_acid_import
GO:0006862	1	1	1	1	1	1	1	4.51E-04	4.49E-03	2.46E-03	0.020	1	1	1	1	0.019	0.046	0.032	0.067	1	1	1	1	GO:0006862_nucleotide_transport
GO:0006869	1	1	1	1	0.011	0.028	0.019	0.044	2.35E-04	2.53E-03	7.55E-07	1.31E-05	1	1	1	1	0.019	0.046	0.024	0.056	6.99E-07	1.36E-05	6.59E-06	9.83E-05	GO:0006869_lipid_transport
GO:0015706	1	1	1	1	4.28E-03	0.021	1.22E-04	1.00E-03	2.36E-06	3.62E-05	4.66E-06	7.17E-05	1	1.35E-15	1.53E-13	2.01E-31	5.24E-29	1.90E-27	3.49E-25	5.60E-32	1.15E-29	3.89E-27	6.04E-25	GO:0015706_nitrate_transport
GO:0015743	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	8.19E-03	0.030	0.128	0.398	0.150	0.452	GO:0015743_malate_transport
GO:0015770	1	1	1	1	1	1	1	0.170	0.415	0.232	0.512	1	1	1	1	1	1	8.19E-03	0.030	0.011	0.069	0.025	0.144	GO:0015770_sucrose_transport
GO:0015840	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1.98E-03	0.016	8.05E-03	0.054	GO:0015840_urea_transport
GO:0034635	1	1	1	1	1	1	1	1	1	1	1	1	1	3.43E-03	9.80E-03	6.52E-03	0.027	8.19E-03	0.030	0.045	0.175	0.103	0.340	GO:0034635_glutathione_transport
GO:0080055	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0.087	0.296	2.78E-03	0.022	GO:0080055_low_affinity_nitrate_transport
GO:0080135	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	8.19E-03	0.030	0.087	0.296	0.103	0.340	GO:0080135_cruciferophyll_acid_transport
GO:0080606	1	1	1	1	1	1	1	4.85E-03	0.036	0.059	0.223	1	1	1	1	1	1	1	1	0.627	1	0.403	0.805	GO:0080606_protein_import_into_nucleus
GO:0006888	1	1	1	1	1	1	1	4.06E-05	5.07E-04	5.83E-04	5.83E-03	1	1	1	1	0.019	0.046	0.048	0.096	0.416	0.825	0.266	0.578	GO:0006888_ER_to_Golgi_vesicle-mediated_transport
GO:0016558	1	1	1	1	1	1	1	0.144	0.415	2.09E-03	0.018	1	1	1	1	1	1	1	1	0.639	1	0.262	0.571	GO:0016558_protein_import_into_peroxisome_matrix
GO:0045037	1	1	1	1	1	1	1	6.05E-03	0.042	0.146	0.451	1	1	1	1	1	1	1	1	1	1	1	1	GO:0045037_protein_import_into_chloroplast_stroma
GO:0015986	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1.98E-03	0.016	2.78E-03	0.022	GO:0015986_ATP_synthesis_coupled_proton_transport
GO:0006833	1	1	1	1	1	1	0.055	0.108	0.576	1	1	1	1	1	1	1	2.51E-04	1.91E-03	5.30E-07	8.60E-06	1.76E-19	1.27E-17	2.90E-22	2.80E-20	GO:0006833_water_transport
GO:0010233	1	1	1	1	3.62E-03	0.018	1.15E-04	9.50E-04	6.05E-03	0.042	0.024	0.138	1	1	1	1	1	0.024	0.056	1	1	1	1	GO:0010233_phloem_transport
GO:0008152	1	1	1	1	0.014	0.035	0.012	0.035	6.30E-05	7.44E-04	5.06E-04	5.15E-03	1	6.85E-03	0.017	0.026	0.056	0.016	0.040	0.189	0.422	0.027	0.157	GO:0008152_metabolic_process
GO:0019748	1	1	1	1	1	1	2.30E-04	1.60E-03	0.077	0.281	0.062	0.250	1	1	1	1	1	1	1	1	3.21E-05	4.11E-04	2.13E-05	2.88E-04	GO:0019748_secondary_metabolic_process
GO:0006096	1	1	1	1	1	1	1	0.420	0.844	0.805	1	1	1	1	1	1	6.52E-03	0.027	0.048	0.096	1.43E-05	1.95E-04	1.04E-07	1.39E-06	GO:0006096_glycolysis
GO:0006098	1	1	1	1	1	1	0.012	0.035	1	1	0.799	1	1	1	1	1	1	1	1	3.57E-15	1.59E-13	1.25E-19	9.70E-18	GO:0006098_pentose-phosphate_shunt
GO:0019253	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	3.40E-04	3.49E-03	1.47E-04	1.67E-03	GO:0019253_reductive_pentose-phosphate_cycle
GO:0000023	1	1	1	1	1	1	3.86E-05	3.61E-04	0.189	0.422	0.485	0.956	1	1	1	1	1	1	1	5.24E-04	5.17E-03	3.41E-04	3.64E-03	GO:0000023_maltose_metabolic_process
GO:0000271	1	1	1	1	3.62E-03	0.018	6.24E-03	0.026	0.074	0.271	0.460	0.909	1	1	1	1	1	1	1	1.64E-05	2.22E-04	6.20E-07	1.08E-05	GO:0000271_polysaccharide_biosynthetic_process
GO:0005975	1	1	1	1	1	1	1	0.155	0.415	0.020	0.118	1	1	6.85E-03	0.017	6.52E-03	0.027	0.032	0.067	0.147	0.415	9.62E-04	9.29E-03	GO:0005975_carbohydrate_metabolic_process
GO:0005982	1	1	1	1	1	1	1	0.012	0.072	1.72E-03	0.015	1	1	1	1	6.52E-03	0.027	1	1	0.057	0.214	0.046	0.200	GO:0005982_starch_metabolic_process
GO:0005986	1	1	1	1	1	1	1	0.028	0.149	1.25E-03	0.012	1	1	1	1	1	1	1	1	1	1	0.237	0.520	GO:0005986_sucrose_biosynthetic_process
GO:0005992	1	1	1	1	3.62E-03	0.018	1	4.51E-04	4.49E-03	0.124	0.393	1	1	1	1	1	1	1	1	0.336	0.682	0.478	0.944	GO:0005992_trehalose_biosynthetic_process
GO:0006730	1	1	1	1	1	1	1	6.05E-03	0.042	0.024	0.138	1	1	1	1	1	1	1	1	1	1	1	1	GO:0006730_one-carbon_metabolic_process
GO:0010264	1	1	1	1	1	1	1	0.051	0.192	0.679	1	1	1	3.43E-03	9.80E-03	2.69E-06	3.93E-05	3.96E-04	2.59E-03	5.50E-06	7.97E-05	3.57E-05	5.58E-05	GO:0010264_myo-inositol_hexakisphosphate_biosynthetic_process
GO:0010411	1	1	1	1	1	1	1	1	1	1	1	1	1	1.16E-05	1.38E-04	4.23E-05	4.46E-04	8.19E-03	0.030	5.45E-05	6.49E-04	7.79E-06	1.14E-04	GO:0010411_xyloglucan_metabolic_process
GO:0019252	1	1	1	1	1	1	1.15E-04	9.50E-04	0.053	0.200	0.058	0.218	1	1	1	1	1	1	1	2.25E-05	2.96E-04	1.92E-05	2.64E-04	GO:0019252_starch_biosynthetic_process
GO:0000272	1	1	1	1	1	1	1	0.132	0.402	0.236	0.520	1	1	1	1	6.52E-03	0.027	1	1	0.127	0.398	0.056	0.244	GO:0000272_polysaccharide_catabolic_process
GO:0000096	1	1	1	1	1	1	0.019	0.044	0.283	0.603	0.255	0.558	1	1	1	1	1	1	1	4.03E-04	4.09E-03	3.19E-04	3.45E-03	GO:0000096_sulfur_amino_acid_metabolic_process
GO:0006520	1	1	1	1	1	1	3.81E-04	2.55E-03	0.048	0.183	0.047	0.200	1	1	1	1	1	1	1	1	1	1	1	1	GO:0006520_cellular_amino_acid_metabolic_process
GO:0006561	1	1	1	1	3.62E-03	0.018	1	1	0.131	0.398	0.146	0.451	1	1	1	1	1	1	1	1	1	1	1	1	GO:0006561_proline_biosynthetic_process
GO:0008652	1	1	1	1	1	1	0.019	0.044	0.300	0.638	0.275	0.591	1	1	1	1	1	1	1	5.17E-04	5.11E-03	4.35E-04	4.46E-03	GO:0008652_cellular_amino_acid_biosynthetic_process
GO:0009072	1	1	1	1	1	1	0.019	0.044	0.607	1	0.197	0.454	1	1	1	1	1	1	1	1.74E-04	1.95E-03	1.18E-04	1.27E-03	GO:0009072_aromatic_amino_acid_family_metabolic_process
GO:0009073	1	1	1	1	1	1	1	1	0.628	1	0.633	1	1	1	1	1	1	1	1	4.37E-03	0.033	7.44E-03	0.053	GO:0009073_aromatic_amino_acid_family_biosynthetic_process
GO:0009086	1	1	1	1	1	1	0.012	0.035	0.351	0.705	0.673	1	1	1	1	1	1	1	1	0.632	1	0.678	1	GO:0009086_methionine_biosynthetic_process
GO:0009095	1	1	1	1	1	1	6.24E-03	0.026	0.046	0.175	0.051	0.200	1	1	1	1	1	1	1	1	1	1	1	1	GO:0009095_aromatic_amino_acid_family_biosynthetic_"process,"_prephenate_pathway
GO:0009098	1	1	1	1	1	1	1	1	1	1	0.100	0.340	1	1	1	1	1	8.19E-03	0.030	1	1	0.103	0.340	GO:0009098_leucine_biosynthetic_process
GO:0019344	1	1	1	1	1	1	0.055	0.108	0.108	0.346	0.056	0.213	1	1	1	1	1.26E-04	1.02E-03	4.37E-05	4.04E-04	9.64E-35	2.34E-32	1.99E-38	7.22E-36	GO:0019344_cysteine_biosynthetic_process
GO:0042398	1	1	1	1	1	1	1	1	0.022	0.123	7.98E-03	0.054	1	1	1	1	0.013	0.032	6.56E-04	3.89E-03	5.17E-04	5.11E-03	2.83E-05	3.74E-04	GO:0042398_cellular_modified_amino_acid_biosynthetic_process
GO:0006546	1	1	1	1	1	1	0.019	0.044	1	1	1	1	1	1	1	1	1	1	1	2.34E-04	2.53E-03	4.18E-06	6.44E-05	GO:0006546_glycine_catabolic_process
GO:0006569	1	1	1	1	1	1	0.043	0.086	0.025	0.140	3.90E-03	0.030	1	1.16E-05	1.38E-04	0.026	0.056	1.84E-05	1.98E-04	5.46E-03	0.040	9.95E-04	9.58E-03	GO:0006569_tryptophan_catabolic_process
GO:0006749	1	1	1	1	7.23E-03	0.028	1	1	0.131	0.398	0.190	0.454	1	1	1	1	1	1	1	1	1	0.195	0.454	GO:0006749_glutathione_metabolic_process
GO:0006511	1	1	1	1	1	1	1	0.246	0.531	0.163	0.454	1	1	6.85E-03	0.017	0.019	0.046	0.016	0.040	1	1	1	1	GO:0006511_ubiquitin-dependent_protein_catabolic_process
GO:0043161

GO:0042128	1	1	1	1	1	1	1	1	4.51E-04	4.49E-03	9.57E-05	1.14E-03	1	1	1	1	6.52E-03	0.027	0.032	0.067	1	1	1	1	GO:0042128_nitrate_assimilation
GO:0006914	1	1	1	1	3.62E-03	0.018	1	1	7.45E-04	7.20E-03	0.015	0.092	1	1	1	1	1	1	1	1	1	1	1	1	GO:0006914_autophagy
GO:0009063	1	1	1	1	1	1	2.37E-07	4.08E-06	1.13E-06	1.84E-05	9.28E-10	2.28E-08	1	1	1	1	1	1	1	1	1	1	1	1	GO:0009063_anaerobic_respiration
GO:0045333	1	1	1	1	1	1	6.24E-03	0.026	0.012	0.072	0.024	0.138	1	1	1	1	1	1	1	1	0.167	0.415	0.237	0.520	GO:0045333_cellular_respiration
GO:0007623	1	1	1	1	0.014	0.035	0.025	0.056	2.65E-05	3.43E-04	3.30E-04	3.56E-03	1	1	1	1	1	1	1	1	1.17E-04	1.35E-03	1.60E-03	0.015	GO:0007623_circadian_rhythm
GO:0009581	1	1	1	1	1	1	1	1	1.85E-04	2.07E-03	6.71E-06	1.00E-04	1	1	1	1	1	1	1	1	1	1	1	1	GO:0009581_detection_of_external_stimulus
GO:0033500	1	1	1	1	1	1	6.24E-03	0.026	0.089	0.236	0.190	0.454	1	1	1	1	1	1	1	1	1	1	1	1	GO:0033500_carbohydrate_homeostasis
GO:0051645	1	1	1	1	1	1	1	1	0.429	0.849	1	1	1	1	3.43E-03	9.80E-03	1	1	1	1	0.421	0.835	0.023	0.133	GO:0051645_Golgi_localization
GO:0051646	1	1	1	1	1	1	1	1	0.429	0.849	1	1	1	1	3.43E-03	9.80E-03	1	1	1	1	0.421	0.835	0.023	0.133	GO:0051646_mitochondrion_localization
GO:0060151	1	1	1	1	1	1	1	1	0.429	0.849	1	1	1	1	3.43E-03	9.80E-03	1	1	1	1	0.421	0.835	0.023	0.133	GO:0060151_peroxisome_localization