Full title

Unraveling transcription factor functions through integrative inference of transcriptional networks in *Arabidopsis thaliana* yields novel regulators involved in reactive oxygen species stress signaling

Short title

Integrative inference of transcriptional networks in *Arabidopsis thaliana* yields novel regulators involved in reactive oxygen species stress signaling

Inge De Clercq[1,2+], Jan Van de Velde[1,2,3+], Xiaopeng Luo[1,2], Li Liu[1,2,3], Veronique Storme[1,2], Robin Pottie[1,2], Dries Vaneechoutte[1,2,3], Frank Van Breusegem[1,2], Klaas Vandepoele[1,2,3*]

(1) Ghent University, Department of Plant Biotechnology and Bioinformatics, Technologiepark 71, 9052 Ghent, Belgium

(2) VIB Center for Plant Systems Biology, Technologiepark 71, 9052 Ghent, Belgium

(3) Bioinformatics Institute Ghent, Ghent University, Technologiepark 71, 9052 Ghent, Belgium

+ both authors contributed equally

* corresponding author: klaas.vandepoele@psb.vib-ugent.be

ABSTRACT

Gene regulation is a dynamic process in which transcription factors (TFs) play an important role to control spatiotemporal gene expression. While gene regulatory networks describe the interactions between TFs and their target genes, our global knowledge about the complexity of TF control for different genes and biological processes is incomplete. To enhance our understanding of the global regulatory lexicon in *Arabidopsis thaliana*, different regulatory input networks capturing complementary information about DNA motifs, open chromatin, TF binding and expression-based regulatory interactions, were combined using a supervised learning approach, resulting in an integrated gene regulatory network (iGRN) covering 1,491 TFs and 31,393 target genes (1.7 million interactions). This iGRN outperforms the different input networks to predict known regulatory interactions and has a similar performance to recover functional interactions compared to state-of-the-art experimental methods like yeast one-hybrid and ChIP-seq. The iGRN correctly inferred known functions for 681 TFs and predicted new gene functions for hundreds of unknown TFs. For regulators predicted to be involved in reactive oxygen species stress regulation, we confirmed in total 75% of TFs with a function in ROS and/or physiological stress responses. This includes 13 novel ROS regulators, previously not connected to any ROS or stress function, that were experimentally validated in our ROS-specific phenotypic assays of loss- or gain-of-function lines. In conclusion, the presented iGRN offers a high-quality starting point integrating different experimental data types at the network level to enhance our understanding of gene regulation in plants.

46 **INTRODUCTION**

47 Accurate spatiotemporal gene expression is of key importance for plant development and
48 molecular responses to environmental cues. Key factors controlling this process are
49 transcription factors (TFs) which regulate their target genes by recognizing and binding short *cis*-
50 regulatory DNA sequences called TF binding sites (TFBSs or DNA motifs) [1]. The full set of
51 regulatory interactions between TFs and their target gene(s) forms a gene regulatory network
52 (GRN) and depending on the underlying source of data, GRNs are composed of physical and/or
53 regulatory protein-DNA interactions. Efficiently identifying regulatory interactions is pivotal to
54 understand how a gene's transcriptional activity is controlled through the interplay of one or
55 multiple regulators. Driven by different modes of gene duplication, specific TF families have
56 been expanded within the green plant lineage, with the model species *Arabidopsis thaliana*
57 containing around 1700 TFs [2]. Although specific regulatory networks and the associated TFs have
58 been studied and are often conserved across species [3], still for hundreds of *Arabidopsis* TFs
59 experimental information is missing [4].

60 Several studies have used high-throughput assays to systematically identify GRNs under
61 specific conditions or in specific developmental and/or anatomical contexts in *Arabidopsis.*
62 Examples include yeast one-hybrid (Y1H) screens to map tissue- or process-specific GRNs as well
63 as TF chromatin immunoprecipitation (ChIP) experiments describing regulatory interactions
64 between TFs and target genes involved in flower development, light and clock signaling,
65 hormone response, and (a)biotic stress response [5]. The latter revealed the high complexity of
66 plant gene regulation and suggested that a single gene can potentially be regulated by up to 75
67 different TFs [6,7]. Open chromatin profiling, either using DNase-Seq or Assay for Transposase-
68 Accessible Chromatin using sequencing (ATAC-Seq), allows studying the organization and
69 dynamics of accessible chromatin regions, and provides a template to infer GRNs in different
70 organs, in response to specific treatments or in different cell types [8-12]. Whereas all these
71 methods aid the construction of GRNs by reporting physical protein-DNA interactions, a major
72 challenge remains the identification of functional interactions, referring to the binding of a
73 specific TF to a cis-regulatory sequence leading to a regulatory effect on a target genes'
74 expression.

75 In this paper we present a network-based approach based on supervised learning for large-
76 scale functional data integration to enhance our understanding of TF gene regulation in
77 *Arabidopsis*, hence referred to as an integrative GRN (iGRN). Supervised learning approaches

78 make efficient use of the available prior knowledge to train a machine learning algorithm, called
79 a classifier, for handling new data. Compared to unsupervised learning approaches, supervised
80 learning is considered more powerful when a large amount of existing knowledge is available to
81 guide the classification task, which here is the identification of functional TF-target interactions
82 [13]. Construction and validation of the iGRN, leveraging a diverse set of complementary
83 experimental datasets, revealed a strong enrichment for functional interactions and a high
84 predictive power to correctly infer functions for both functionally characterized and
85 uncharacterized TFs involved in a variety of biological processes. The latter is demonstrated in a
86 case study by the experimental validation of novel iGRN-based function predictions for TFs
87 involved in plant defense responses to oxidative stresses.
88

89 **RESULTS**

90 **An integrated gene regulatory network for *Arabidopsis thaliana***

91 To generate an integrated gene regulatory network (*Arabidopsis* iGRN), different types of
92 complementary regulatory datasets were first processed, resulting in different input networks
93 (Table 1). In total seven input networks, encompassing information about DNA motifs, open
94 chromatin, TF binding and expression-based regulatory interactions, were obtained by applying
95 different network inference and data analysis approaches (see Methods for underlying datasets;
96 Figure 1A). Subsequently, through the application of a supervised learning approach, a classifier
97 was built to identify those interactions from the input networks that most likely represent
98 functional interactions.
99 The first network is based on the presence of a TF motif, modelled using Position Weight
100 Matrices (PWMs), in the promoter of a target gene. A total of 1,260 high quality PWMs, covering
101 874 TFs, were mapped onto the *Arabidopsis* genome resulting in a TF - PWM target gene
102 network. Using the genomic locations of publicly available DNaseI datasets the PWM network
103 was filtered resulting in a second open chromatin regulatory input network (called "PWM in
104 DHS"). DNaseI hypersensitive sites (DHS) are considered to be a proxy for open chromatin and
105 potential protein binding to the DNA [14]. As a third network, the set of PWMs was also evaluated
106 for evolutionary conservation across 13 dicotyledonous genomes using a comparative motif
107 mapping algorithm [15], which resulted in an evolutionarily conserved regulatory network
108 ("Conserved PWM"). To integrate experimental data, a fourth "TF ChIP" network was created by

integrating experimentally determined TF binding occupancy profiles from 27 TFs [7]. Next to a physical TF binding network, two co-expression-based regulatory networks were also generated. Co-expression approaches measure similarity at the output level of transcriptional regulation and as such can provide a complementary view compared to promoter-centered regulatory networks. First, the Pearson correlation between a TF and a target gene was used to build co-expression clusters around each TF, resulting in a fifth "CoE TF-target" input network. Secondly, GENIE3, a reverse engineering tool for reconstructing regulatory networks based on expression data, was used to build a sixth input network [16]. GENIE3 was among the best performing tools in reconstructing regulatory interactions in a large-scale benchmark study [17]. The last network was generated using a hybrid approach of co-expression and PWM based methods ("COE+PWM"). Initially a co-expression cluster was constructed for each gene and PWM enrichment was subsequently performed. Only genes with significantly enriched PWMs in their co-expression cluster were retained, resulting in a seventh network where TF binding site information is combined with co-regulatory information (see Methods).

Figure 1. Construction and evaluation of the integrated gene regulatory network (iGRN).

A. Overview of the different input networks used to construct the supervised iGRN. Whereas colored circles indicate the type of data that is used to construct the input network, the information below each

127 network depicts how edges in that network are weighted. B. Overview of the sizes of the different
128 networks together with the recall and enrichment for edges in the evaluation test dataset (green and blue
129 bars, respectively). The yellow bars report the fraction of each input network that is present in the iGRN
130 while the red bars report which fraction of the iGRN edges is supported by this input network.
131
132 As the different input networks capture different aspects of gene regulation, we next applied
133 supervised learning to construct, starting from these input networks, an integrative regulatory
134 network reporting functional interactions. A score for each possible TF-target interaction was
135 devised as a function of the feature-specific weights provided by the seven input networks. For
136 building a classifier, we compiled a gold standard dataset comprising interactions based on all
137 validated interactions in AtRegNet [18], an exhaustive literature-based network reporting
138 interactions in secondary cell wall development [19], and a large-scale data mining effort on 974
139 peer reviewed articles called ATRM [20]. This gold standard entails interactions for 47 TF families
140 involved in different biological processes, with on average 11 interactions per TF (Table 1 and
141 Supplemental Table S1). Eighty percent of the interactions present in this gold standard were
142 retained as training dataset for the machine learning classification algorithms while the
143 remaining twenty percent was used as an evaluation set. Initially different classifiers were
144 learned and their performance to identify functional interactions in the evaluation set was
145 assessed (Supplemental Table S2). The classifier based on gradient boosting model using
146 decision trees showed the best performance, with an area under the curve (AUC) value of 0.77,
147 and Precision and Recall of 71% and 48% on the evaluation set, respectively (Supplemental
148 Table S2). Applying this classifier to the complete set of interactions present in the seven input
149 networks resulted in an integrative network of 1,709,871 interactions for 1,491 TFs, containing
150 31,393 target genes (29,137 protein-coding genes; see Data file 1).
151 In order to obtain an unbiased performance estimate of the input networks compared to the
152 iGRN, all input networks were compared to the evaluation dataset of known interactions and
153 benchmarked measuring the recovery of true edges (referred to as the recall, Figure 1B).
154 Furthermore, we also determined the enrichment of a network towards the evaluation dataset,
155 which functions as a proxy for specificity [21]. Focusing on the different input networks, Figure 1B
156 shows that the network based on evolutionary conserved TF binding sites ("Conserved PWM")
157 has the highest enrichment (6.4 fold, Supplemental Table 3). Evolutionary conservation is known
158 to be a strong predictor of functional regulatory elements explaining this high enrichment [22,23].

159 The TF ChIP-seq network also performs well with both a high recovery (29%) and enrichment
160 (6.0 fold). This overlap can partially be explained by the fact that a subset of the validated
161 AtRegNet interactions in the gold standard are based on ChIP studies [7]. Other TF motif mapping
162 strategies recover larger fractions of true edges but suffer from low specificity, resulting in low
163 enrichment values. The simple mapping of TF motifs ("PWM") is known to result in many false
164 positives because TF binding sites are often short and typically contain some level of degeneracy
165 [11,24]. Both co-expression based networks do not consider TF binding site information and are
166 therefore able to also identify regulatory interactions for TFs lacking DNA binding motif
167 information. Both these networks perform relatively well: correlation-based co-expression
168 between TF and target genes, implemented in the "CoE TF-target" network, has the third
169 highest enrichment towards the evaluation dataset (10% recovery and 3.0 fold enrichment),
170 while GENIE3 displays a good balance between recovery of true edges and specificity
171 (Supplemental Table 3). Focusing on the integrative network (iGRN) shows it has the strongest
172 enrichment (8.8 fold) as well as a high recovery for true regulatory relationships (46% recovery
173 of interactions in the evaluation set), clearly outperforming the different computational and
174 experimental approaches used to construct the input networks.
175 Apart from determining the power of the iGRN network to recover known functional
176 interactions, we also measured the contribution of each of the seven input networks to the
177 iGRN. Ideally, a good integration approach should favor interactions from the input networks
178 with high enrichment values towards the evaluation set, as these would boost the recovery of
179 correct interactions. As reported in Figure 1B (yellow and red bar charts), the evolutionary
180 conserved motifs ("Conserved PWM") and the "TF ChIP" network contributed the largest part of
181 their edges to the integrative network (91% and 80%, respectively), while networks showing low
182 specificity, such as "PWM", only contributed a small fraction of interactions (7%). This result
183 corroborates that the applied supervised approach successfully identified functional interactions
184 coming from high-quality input networks. Next, we measured which iGRN interactions are
185 supported by the different input networks (Figure 1B, red bar plot). Overall, the integrative
186 network has 70% of the edges supported by a PWM mapping, meaning that a majority of
187 regulatory interactions is supported by the presence of TF binding site(s). Furthermore, 36-48%
188 of the interactions are supported by at least one of the two co-expression networks, indicating
189 for nearly half of all interactions a regulatory effect at the transcript level. Overall, 92% of the

190 iGRN interactions are supported by more than one input network (Figure S1), confirming both
191 the commonality and complementarity of the input networks.
192

193 **iGRN interactions are complementary to TF regulatory events identified from**
194 **experimental methods**

195 To assess whether the iGRN identifies functional TF-target gene interactions that are
196 supported by experimental data, detailed comparisons of the iGRN interactions were made
197 against three different datasets, capturing a variety of TFs and biological processes, which were
198 not used during the iGRN construction (see Methods). These datasets cover i) a set of 2,759 Y1H
199 interactions reflecting direct *in vitro* interactions ii) 28 TF ChIP-seq experiments collected from
200 different studies (148,949 interactions, see also Methods) reflecting primarily direct *in vivo*
201 binding and iii) a set of 40 differentially expressed (DE) gene sets upon genetic TF perturbation
202 (51,178 interactions, see Methods). The significance of the observed overlaps was examined
203 through comparison to a background distribution of expected overlaps generated by
204 randomizing the iGRN (see Methods). First, we detected a weak but significant overlap between
205 edges in the integrative network and regulatory edges determined by Y1H (Figure 2A). An
206 overlap of 336 edges (12%) was observed reflecting a 1.6 higher overlap than expected by
207 chance (Figure 2B). TF ChIP-seq interactions displayed an overlap of 15,487 interactions (10%)
208 with a fold enrichment of 2.4. Only considering TFs that occur in both the iGRN and this ChIP-
209 Seq dataset, for 25/28 regulators there was a significant overlap between the iGRN and ChIP
210 defined target genes (Supplemental Table S4). Overall we found that TF ChIP-seq data confirmed
211 39% of predicted interactions in the iGRN, demonstrating that a large fraction of iGRN
212 interactions are supported by *in vivo* TF binding. This fraction is comparable with the
213 concordance between TF ChIP-based regulatory interactions and altered target gene transcript
214 levels after genetic TF perturbation, an approach that is frequently used to identify bound target
215 genes that are also regulated [25]. Next to TF-DNA binding assays, we also investigated the
216 recovery of directly and indirectly regulated target genes based on differential target gene
217 expression. The iGRN recovered 9,672 differentially expressed interactions (19%) resulting in a
218 3.6 fold enrichment (Figure 2A and 2B). Taken together, both the comparison with binding and
219 regulatory datasets demonstrate that the iGRN represents high-quality interactions supported
220 by experimental data.

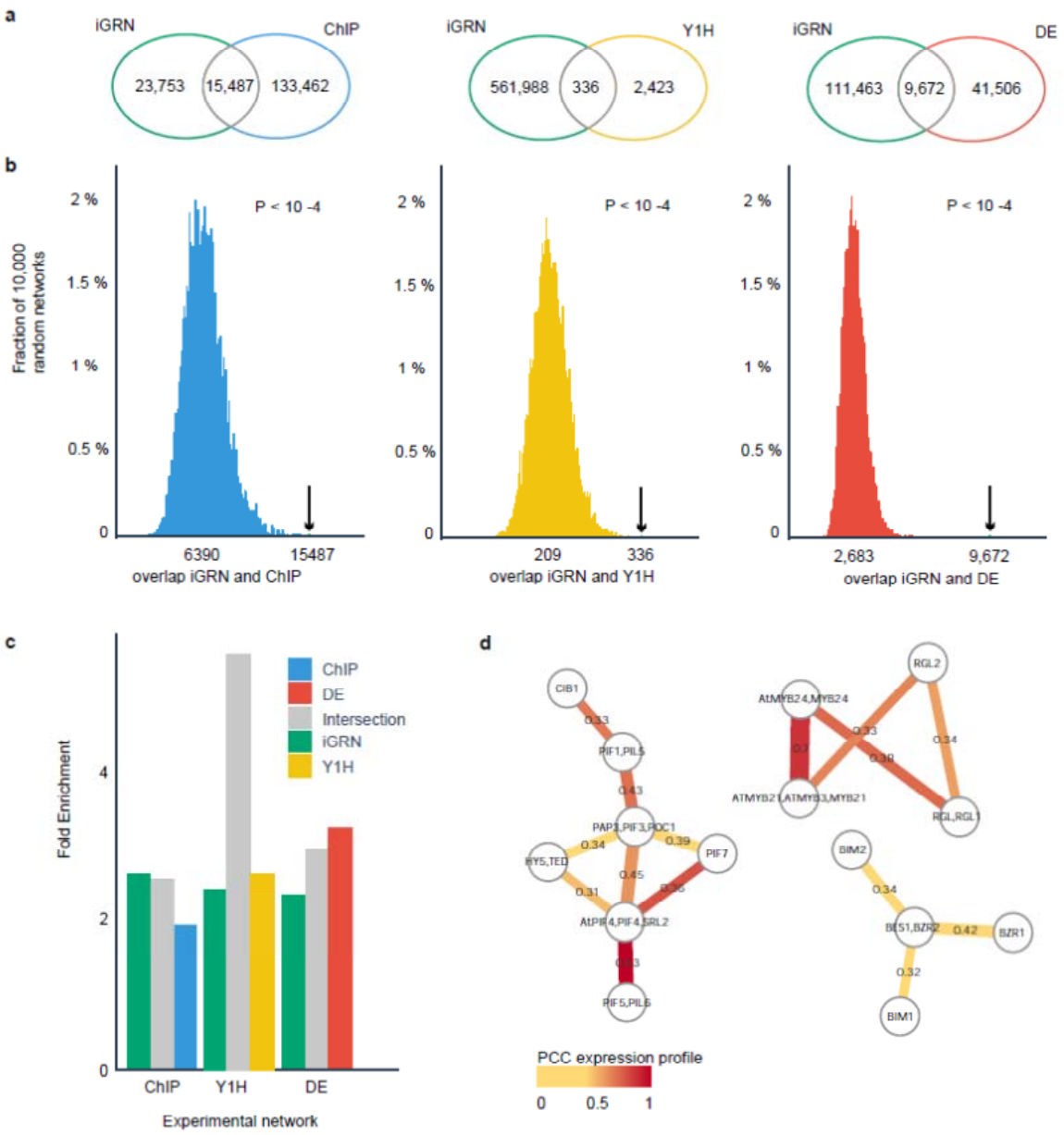


Figure 2. Overlap between iGRN interactions and different experimental data sources describing gene regulatory and functional information.

A. Venn diagrams showing the number of shared interactions between the iGRN and TF ChIP, Y1H and differentially expressed (DE) genes after TF perturbation (depicted in blue, yellow and red, respectively). B. Histograms showing the expected overlap based on 10,000 randomized networks, together with the observed overlap indicated by the black arrow. The p-values for observing the measured overlap by chance are also reported. C. Fold enrichment for a functional metric describing how well target genes of a specific TF are involved in similar biological processes (average values for all TFs present in the different data series are reported). Whereas the grey bars indicate the functional enrichment for shared

interactions, blue, red, yellow and green depict fold enrichment for interactions unique to the ChIP, DE, Y1H and iGRN, respectively. D. Sub-networks showing experimental protein-protein interactions for TFs sharing a large number of target genes (Jaccard index >0.3). Whereas the edge thickness and values reports the fraction of shared target genes for a pair of TFs quantified using the Jaccard index, the color represents the absolute value of the Pearson Correlation Coefficient between the expression profiles of a pair of TFs.

To evaluate the relevance of the regulatory interactions uniquely predicted by the iGRN and hence not overlapping with the three experimental datasets (Y1H, TF ChIP-Seq, and DE), a functional metric was used to measure whether a TF's target genes participate in similar biological processes or have similar functions. This metric starts from the assumption that a set of bona fide target genes will have a higher enrichment for functional annotations than randomized networks [21]. For each TF, the enriched Gene Ontology (GO) Biological Process terms based on its target genes were determined, summarized in a distribution and compared to the background distribution based on randomized networks. In Figure 2C, the iGRN-predicted interactions were first divided in different sets based on their overlap with the TF ChIP, Y1H and DE experimental datasets (as in Figure 2A), respectively. Next, for each subset the fold enrichments for functional GO Biological Process terms were assessed. Whereas the grey series in Figure 2C show the regulatory interactions shared between the iGRN and the experimental dataset, the green series shows the interactions unique to the iGRN. Likewise, the blue, yellow and red series show the regulatory interactions present in the ChIP, Y1H and DE dataset not overlapping with the iGRN.

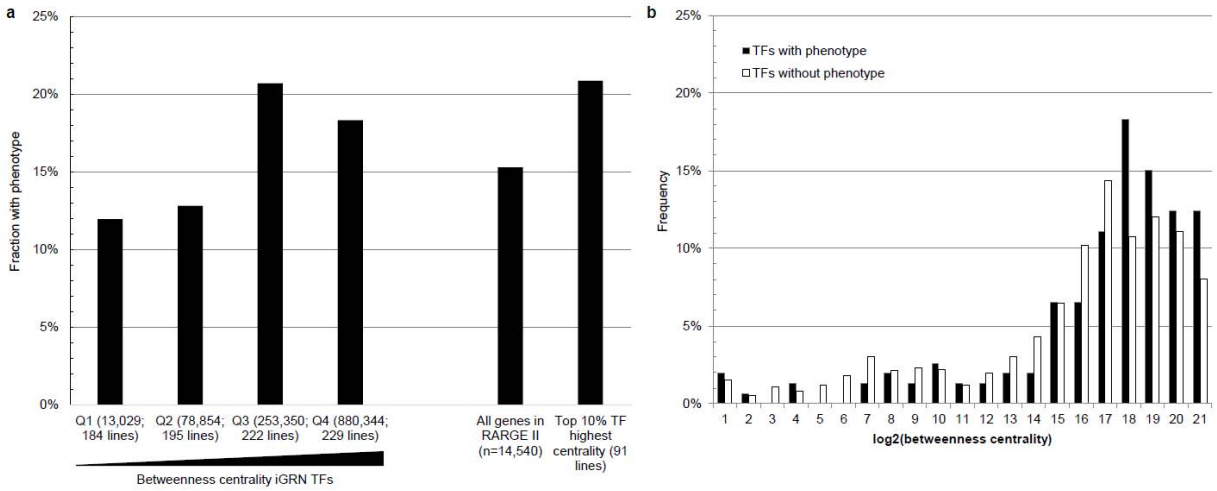
Focusing on the overlap with TF ChIP dataset revealed that the functional enrichment was similar for the interactions unique to the iGRN compared to those shared between the iGRN and TF ChIP dataset (2.63 and 2.55, respectively). However, the regulatory interactions unique to the TF ChIP datasets showed lower fold enrichments (1.94), indicating that these show less functional coherence. For the Y1H comparison, the interactions uniquely predicted by the iGRN had a similar fold enrichment as the interactions only reported by the Y1H assays (2.41 and 2.61). The interactions that are reported by both approaches had a considerable higher enrichment compared to the other networks (5.55; though note the small set size, n=336). The comparison with the DE dataset showed a similar result with an increase in enrichment fold from uniquely predicted edges in the iGRN to the intersection and uniquely predicted edges by the DE gene sets (2.33, 2.96 and 3.23, respectively). Taken together, these results indicate that

264 the predicted interactions, including those unique to the iGRN, convey functional gene
265 regulatory events.
266

267 **Network architecture reveals different levels of gene regulatory complexity affecting**
268 **plant phenotypes**

269 Starting from the complete iGRN comprising 1.7 million interactions, we investigated the
270 structure and complexity of the network and determined if network architecture can be linked
271 to specific biological functions. On average, each gene in the network has 54 incoming TFs
272 (median 32), while each TF is linked to 1,147 target genes (median 802; Supplemental Figure S2).
273 There was a large range in connectivity with half of the genes being regulated by 8-79 TFs
274 (interquartile range), and a small number of TFs (107/1491, or 7%) regulating more than 3,000
275 target genes. Focusing on the top 1% genes with the largest number of incoming TFs revealed a
276 significant bias towards genes with transcription factor activity, corroborating previous findings
277 reporting that regulators themselves are complexly regulated by other regulators [7,26]. In
278 contrast, the 2172 genes having only one incoming TF showed enrichment for RNA binding and
279 translation, suggesting that these functions are less complexly regulated by TFs.
280 Next, we examined if network structure can offer new insights in the role of individual
281 regulators. More precisely, we investigated whether the degree (number of target genes) and
282 betweenness centrality (a measure for the importance based on number of shortest paths in the
283 network graph) of TFs in the iGRN can be predictive for potential master regulators showing
284 prominent phenotypes upon genetic perturbation. Whereas degree refers to the number of
285 interactions a node, or gene, has in the network, betweenness centrality measures how
286 important a node is to the shortest paths through the network. Nodes with a high centrality are
287 assumed to be more important in the network as more information is passing through these
288 nodes. After processing large-scale phenotyping data from RARGE II, an integrated phenotype
289 database of Arabidopsis mutants scoring traits using a controlled vocabulary for 14,540 genes,
290 phenotypes were recorded for 2,224 genes [27]. When splitting all TFs present in the iGRN and
291 part of the RARGE II dataset in four quartiles (Q1-Q4) based on their degree values, we observed
292 that 17.4% of the high-degree (Q3+Q4) and 15.1% of the low-degree (Q1+Q2) TFs had a
293 reported phenotype (p-value 0.43; two-tailed Mann-Whitney U test). When comparing the
294 quartiles of TFs based on betweenness centrality, a significant difference was found when

295 scoring regulators with a reported phenotype (19.6% for Q3+Q4 and 12.4% for Q1+Q2,
296 respectively; p-value 0.0088; two-tailed Mann-Whitney U test; Figure 3A). This pattern is
297 confirmed when comparing the betweenness centrality for TFs with and without phenotypes, as
298 regulators showing RARGE II phenotypes have more of the shortest network paths going
299 through their nodes (log2 median value of 17.50 and 16.46 for TFs with and without reported
300 phenotypes, respectively; p-value 1e-05; two-tailed Mann-Whitney U test; Figure 3B). iGRN TFs
301 within the top 10% genes with highest centrality for which mutants have reported phenotypes
302 (19/91) are significantly biased towards post-embryonic development and include ATGTL1, PIF4,
303 SEP3, IDD4 and IDD8, ANAC056, DOF5.6, LFY3 and TCP4. Interestingly, both for TFs with and
304 without phenotypes, a similar number of TFs in their target genes was found (9.45% and 9.44%,
305 respectively), indicating that TFs showing a phenotype do not regulate more TFs and thus are
306 not higher in the TF hierarchy.
307



308

309 **Figure 3. Phenotypes reported after TF perturbation for TFs with different network centrality**
310 **properties.**
311 A. RARGE-II phenotypes reported for iGRN TFs with different betweenness centrality. Q1-Q4 refer to the
312 four quartiles of all iGRN TFs (sorted based on increasing betweenness centrality) and values in
313 parenthesis refer to the median centrality value together with the number of lines available in the RARGE
314 II dataset for that quartile. B. Betweenness centrality distribution for TFs with and without phenotypes
315 reported in RARGE-II (n= 153 and 1589, respectively).
316

317 As transcriptional regulation is mediated through interactions between TFs, co-activators or
318 co-repressors, and the general transcriptional machinery, we determined whether the iGRN has

319 predictive power to identify cooperative TF regulation. Therefore, we measured if regulators
320 sharing a large number of target genes are more likely to interact with each other, based on
321 experimentally determined protein-protein interactions (PPIs). We found that TF pairs sharing
322 many target genes (Jaccard index >0.3) have a 15-fold higher occurrence of PPIs compared to TF
323 pairs having less shared target genes (Jaccard index <0.3), which is highly significant (2.71%
324 versus 0.31%; p-value <2.2e-16, hypergeometric distribution). Whereas most of these PPIs cover
325 TFs from the same gene family, probably due to the high similarity of their corresponding
326 binding sites yielding similar sets of target genes, their potential to act as heterodimers and/or
327 potential functional redundancy, also cases of interacting TF pairs from different gene families
328 were found. Examples of the latter case include HY5-PIF3/PIF4 involved in far-red signaling,
329 MYB21-RGL1/RGL2 controlling hormone-mediated cell proliferation and differentiation, and
330 BES1-BIM1/BIM2 acting synergistically to control brassinosteroid signaling (Figure 2D). Highly
331 correlated expression profiles of these TFs sharing a large number of target genes further
332 suggest they exert their regulatory roles in a cooperative manner.
333

334 **Network-based identification of gene functions for Arabidopsis TFs**

335 Based on the validation results indicating the functional relevance of regulatory edges in the
336 iGRN, we also investigated if individual TF functions can be inferred. Therefore, we performed
337 GO enrichment analysis for all target genes of each TF, only considering experimental Biological
338 Process annotations and excluding very general GO terms (see Methods). Next, we determined
339 if the enriched Biological Process terms were already known for that specific TF or whether a
340 novel functionality was predicted (Data file 2. TF_function_all). In total we could infer functions
341 for 1,301 TFs and for 54% (681) of the 1,259 TFs with experimental Biological Process
342 annotations we could correctly infer one or more of the known functions. For 268 TFs lacking
343 experimental annotations, novel iGRN-based functions were predicted. Reciprocally, per
344 Biological Process term, on average 21% of the known regulators were recovered, while for
345 several distinct processes the recovery was 50% or higher. Examples of processes with high
346 recovery rates include response to wounding (14/20, 70%), plant-type secondary cell wall
347 biogenesis (18/28, 64%), response to hormone (174/343, 51%), chloroplast organization (6/12,
348 50%) and DNA replication (3/6, 50%).

349 The global recovery of known TF functions based on the iGRN was lower compared to the
350 probabilistic functional gene network AraNet v2 [28], which recovered known functions for 739 TFs
351 (28% average recovery of known regulators per GO BP term). However, both approaches are
352 highly complementary, as 207 and 265 TFs had known functions only recovered using iGRN and
353 AraNet, respectively (Supplemental Figure S3 iGRN_TF_GOBPpredictions.pdf). Examples of
354 known TF functions uniquely recovered through the iGRN cover EIN3 (ethylene-activated
355 signaling pathway), APETALA2 (flower development), LFY (flower development, gibberellic acid-
356 mediated signaling pathway), FUS3 (somatic embryogenesis), TINY (multidimensional cell
357 growth), and AMS, ATWRKY34 and CRF12 (pollen development). Focusing on specific GO terms,
358 biological processes with a higher recovery of known regulators in the iGRN compared to AraNet
359 cover response to biotic stimulus (47% versus 35%), chloroplast organization (50% versus 0%),
360 response to abiotic stimulus (47% versus 26%), pollination (33% versus 0%), and root hair cell
361 differentiation (35% versus 6%). For the latter process we correctly identified the regulators
362 ZPF5, RSL2 and RSL4, RHD6, LRL2 and LRL3, as well as MYB36. MYB36 regulates the
363 developmental transition in the root endodermis, where the differentiated endodermis forms a
364 protective waxy barrier called the Casparian strip. In our network, MYB36 is predicted to
365 regulate genes involved in cell wall organization such as CASP1-5, PER64 and ESB1, which are
366 known to be involved in Casparian strip formation and have been shown to be differentially
367 expressed in the *myb36* mutant [29]. Furthermore, our iGRN network also predicts MYB36 to be
368 involved in root hair elongation, which is in agreement with the longer root hairs observed in
369 *myb36* mutants.
370

371 **Identification of novel TFs in reactive oxygen species (ROS) stress regulation**

372 Apart from recovering known functions of TFs involved in development and stress responses,
373 the predictive power of the iGRN was used to identify novel TF functions. As a test case, we
374 extracted TFs from the iGRN with predicted oxidative stress signaling and/or responsiveness
375 functions. As an increase in ROS levels is a common theme during abiotic and biotic stresses,
376 oxidative stress responsiveness was used to validate novel regulators of plant stress responses.
377 One hundred twenty-four ROS-related TFs were extracted from the iGRN by identifying those
378 TFs that entailed an enrichment for core ROS responsive transcripts in their target genes [30].
379 Among these TFs, 17 had an already reported function in plant responses to ROS or cellular ROS-

380 inducing agents (hydrogen peroxide [H_2O_2], methyl viologen [MV] and 3-amino triazole [3-AT]),
381 here referred to as 'KNOWN_ROS' TFs (Table 2, Supplemental Table S5). Fifty-eight ROS TFs
382 were previously reported to function in plant responses to environmental stresses, such as
383 salinity, drought, heavy metals, cold, hypoxia, high light and infection by bacterial and fungal
384 pathogens (and that we refer to as 'KNOWN_STRESS' TFs, Supplemental Table S5). The
385 remaining 49 extracted ROS TFs had previously not been linked to any role in
386 ROS/environmental stress (referred to as 'NOVEL' TFs in Table 2), at the time the iGRN was
387 constructed (see Methods). In the meantime, five of the 49 novel TFs were validated for a role in
388 environmental stress responses by recent literature evidence (Table 2): NAC102 (high light; [31]), ,
389 BEH2 (drought; [32]), NAC103 (genotoxic; [33]), ZAT18 (drought; [34]) and ERF019 (drought, bacterial
390 and fungal pathogens; [35,36]).
391 For experimental validation of a function in ROS responses, we selected the top 32 novel TFs,
392 ranked based on enrichment for ROS marker target genes, for which at least one *Arabidopsis*
393 loss- or gain-of-function transgenic line was available (Supplemental Table S6). Plant
394 performance based on rosette area growth of the mutant lines compared to the wild type was
395 monitored when grown at low and high MV and 3-AT stress, respectively, using automated
396 image analysis (see Methods).
397
398

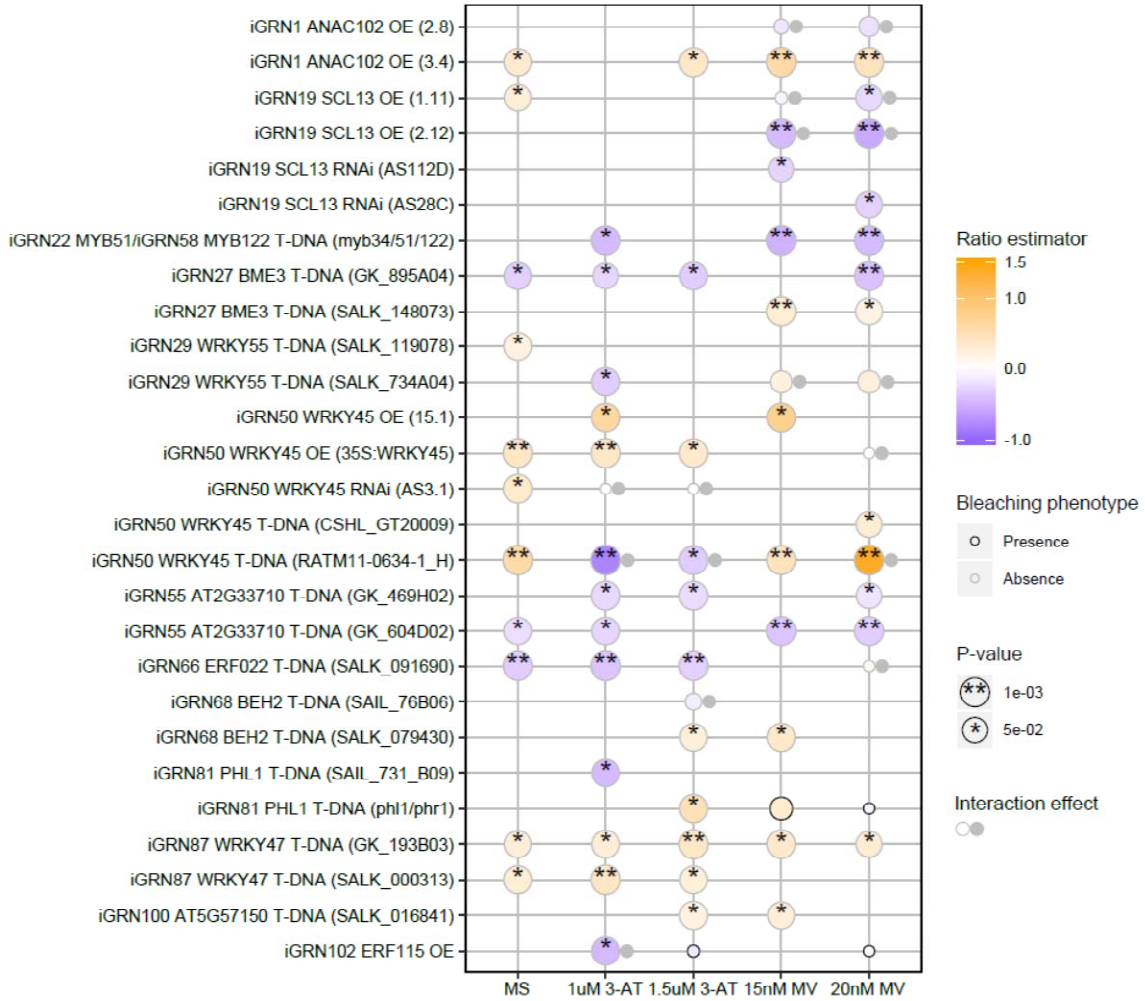
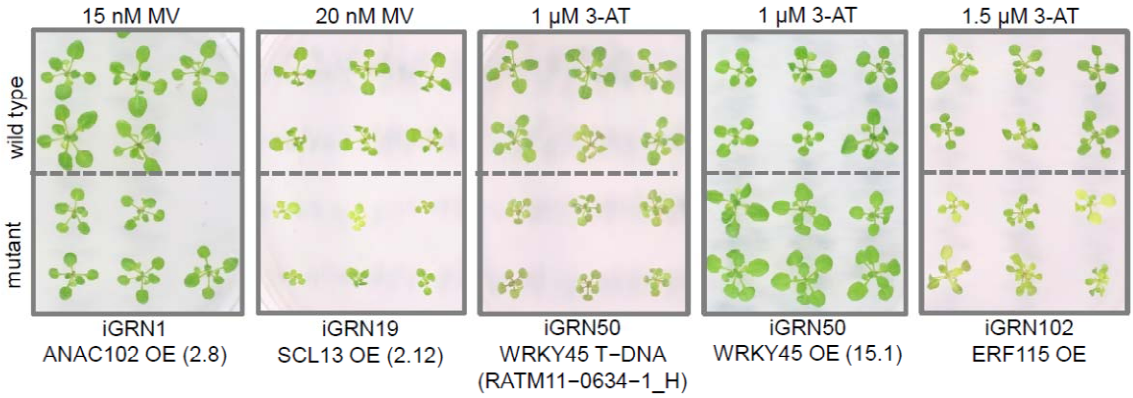


Figure 4. Oxidative stress phenotyping results of novel ROS TFs.

Heatmap presenting phenotypes associated with mutant alleles of iGRN-extracted novel ROS TFs under different oxidative stress conditions. Mutant lines with a statistically significant phenotype under oxidative stress (1 µM and 1.5 µM 3-AT and 15 and 20 nM MV) and/or normal growth conditions (MS) are displayed. Results of other mutant lines that were tested in this study can be found in Supplemental Table S5 and S6. Statistically significant differences relative to wild type (Col-0) are indicated by asterisk; *, $P < 0.05$ and **, $P < 0.001$ as determined by the hypothesis tests on the fixed effects of the mixed model (see Methods). Color scale indicates the ratio of the rosette area of 14-day-old mutant alleles relative to wild type (Col-0). Black and grey borders indicate presence or absence of a bleaching phenotype of the mutant plants, respectively. Small grey circle indicates a statistically significant genotype-by-stress treatment interaction effect.

For 13 of the 32 phenotyped novel ROS TFs we found at least one mutant allele with significant oxidative stress-induced changes in rosette area relative to the wild type (i.e.

414 genotype effect dependent on the oxidative stress treatment and/or genotype-by-treatment
415 effect under at least one of the tested stress treatments; Figure 4; Table 2). For the other novel
416 TFs, no significant changes were detected by the genotype under all tested conditions, either
417 there was an overall genotype effect under all tested conditions (similar changes were
418 observed) or that was independent of the stress treatment (e.g. WRKY47; Figure 4). For SCL13,
419 WRKY45, BEH2 and PHL1, experimental evidence is provided by at least two mutant alleles.
420 *WRKY45* loss-of-function (T-DNA mutant RATM11−0634−1_H and RNAi line AS3.1) affected the
421 rosette area growth during low and high 3-AT stress compared to the wild type in a genotype-
422 by-treatment dependent matter, whereas, *WRKY45* OE (line 15.1) caused an increased rosette
423 yield that was dependent on stress treatments low 3-AT and low MV (Figure 5). Two *SCL13*
424 overexpression alleles (1.11 and 2.12) displayed decreased rosette growth dependent on both
425 MV conditions in a genotype-by-treatment-dependent manner (Figure 5). Similarly, knockdown
426 of *SCL13* (AS112D and AS28C) negatively affected rosette growth under MV stress, but no
427 genotype-by-treatment interaction effect was observed. *PHL1* T-DNA and *PHL1/PHR1* double T-
428 DNA lines displayed altered growth on low and high 3-AT, respectively, the latter dependent on
429 the genotype-by-treatment. *BEH2* mutation altered the rosette area that was dependent on the
430 stress treatments high 3-AT and low MV for line SALK_079430 and dependent on the genotype-
431 by-treatment for line SAIL_76B06.
432



433
434 **Figure 5. Representative images of oxidative stress phenotypes for a selection of mutants.**
435 Representative images of novel ROS TFs with a phenotype during oxidative stress. 17-days-old seedlings
436 of selected mutant lines compared to wild type (Col-0).
437

438 We further demonstrated that ANAC102, BME3, MYB51, WRKY55, AT2G33710, ERF022,
439 AT5G57150 and ERF115 have a function in oxidative stress responses based on one mutant
440 allele (for *ERF022* and *ERF115* only one line was available) (Figure 4 and Table 2). For example,
441 *ANAC102* OE line 3.4 (strong OE; ˜ 100-fold) displayed a larger rosette area that was
442 independent of the stress treatment, while rosette area changes by *ANAC102* OE2.8 (weak OE;
443 ˜4.5-fold) were dependent on the genotype-treatment interaction (Figure 5). *ERF115* OE
444 affected plant performance in a manner that was dependent on the stress treatment, with a
445 decreased rosette area under low 3-AT (dependent on the genotype-by-stress treatment) and
446 severe bleaching under high MV and 3-AT concentrations (Figure 5).
447 Next, the expression profiles of KNOWN_ROS and validated novel ROS TFs during different
448 oxidative stress conditions were analyzed and compared [30]. Fourteen out of the 17 known ROS
449 TFs were identified as responsive to a wide range of oxidative stress conditions ($|\log_2$ fold
450 change$| > 1$ and FDR < 0.01; Figure 6; Supplemental Table S7). Among the 13 newly validated
451 novel ROS TFs, only half were ROS responsive at the transcript level. *WRKY55*, *MYB122* (for
452 which the triple mutant *myb34/51/122* shows a phenotype, but not the *myb122* single mutant),
453 *ERF022*, *BEH2*, *PHL1*, AT5G57150 and *ERF115* expression levels are not or only under very
454 specific ROS-inducing conditions/mutants regulated, and thus, would not have been predicted
455 from studies solely relying on differential expression at the whole plant or organ level. For
456 example, *ERF115* has tissue-specific expression patterns, confined to the stem cell niche within
457 the root meristem [37], indicating that transcript analysis on the whole plant level might not reveal
458 stress-induced alterations in this specific tissue [38].
459

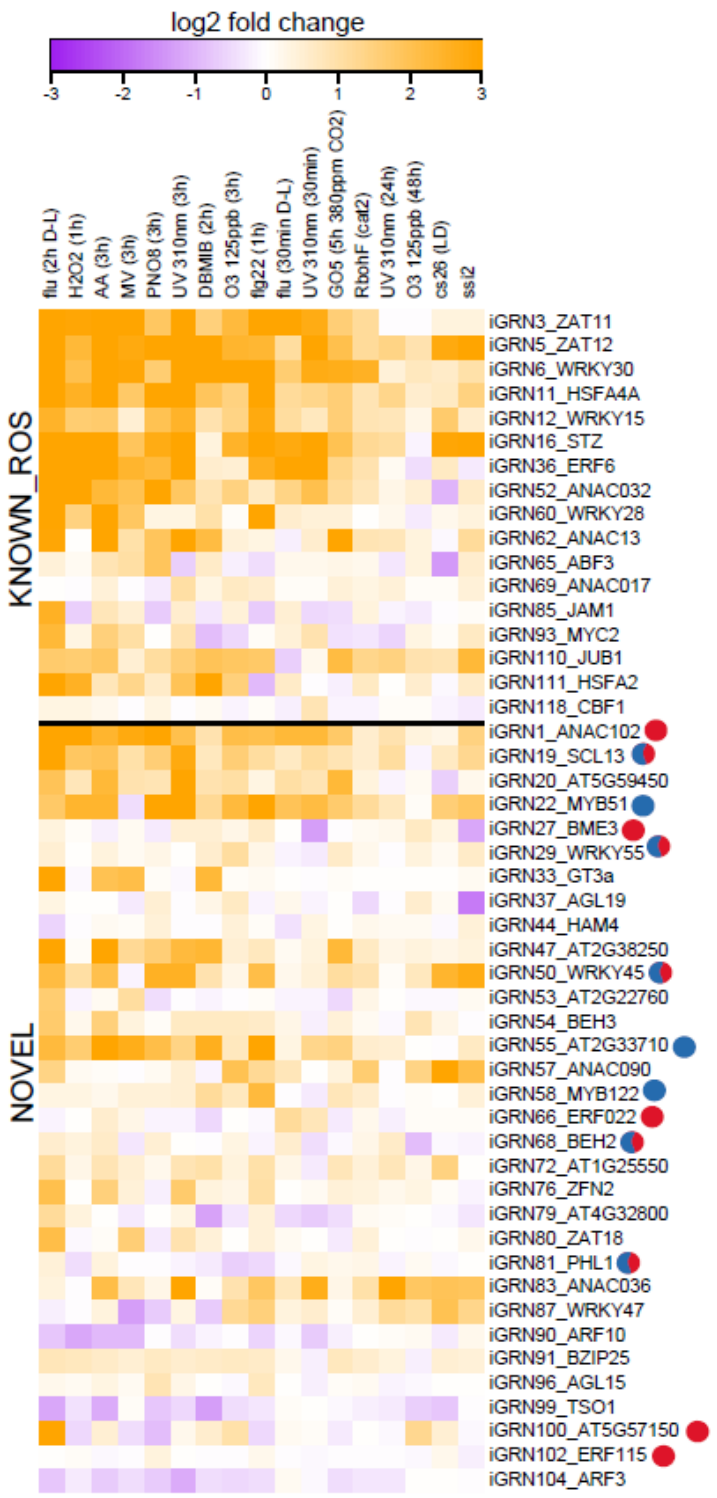


Figure 6. Expression of KNOWN and NOVEL ROS TFs during various conditions provoking cellular ROS/redox imbalances.

Heatmap showing the expression response of ROS TFs extracted from the iGRN during a selection of conditions provoking cellular ROS production or redox imbalances (a more elaborate set of oxidative stress conditions is displayed in Supplemental Table S7). TFs with previously known function in ROS responses and TFs with no functional link with any oxidative stress and/or (a)biotic stress responses (NOVEL) that were tested in our phenotyping assays are displayed. NOVEL TFs validated in our study are shown using blue and red circles, indicating a positive and negative effect by the TF on oxidative stress tolerance, respectively.

490

491 Overall, among the 124 novel ROS TFs extracted from the iGRN, 80 (65%) were validated by
492 previously published studies, including recent publications after the iGRN was constructed, to be
493 involved in the plant's responses to ROS perturbation and/or (a)biotic stresses. By analyzing

494 transgenic lines perturbed in one of the 32 novel ROS TFs, 13 additional regulators were
495 validated for a function in ROS responses by our phenotyping assays. Among the novel ROS TFs
496 newly identified by our phenotyping assays, WRKY55, WRKY47, AT2G33710 [ERF] and
497 AT5G57150 [bHLH] had previously not been associated with any function. Other validated novel
498 ROS TFs are known to function in age-triggered senescence (WRKY45, ERF022 [AP2/EREBP];
499 [39,40]), response to phosphate starvation (WRKY45, PHL1 [G2 family]; [41,42]), submergence stress
500 (WRKY45; [43]), glucosinolate (antimicrobial compounds involved in plant immunity) biosynthesis
501 (MYB51, MYB122; [44]), brassinosteroid signaling (BEH2 [BES1 TF family]; [45]), phytochrome-
502 dependent light signaling (SCL13 [GRAS family]; [46]), hypocotyl growth (BEH2 [BES1 family]; [47]),
503 seed germination (BME3 [GATA family]; [48]) and somatic embryogenesis (ERF022; [49]). Overall,
504 among the novel ROS TFs validated for an oxidative stress function by literature evidence or in
505 our phenotyping assays, the AP2-EREBP (5), NAC (5) and WRKY (5) families are highly
506 represented (Supplemental Table S5). These families are typically associated with abiotic, biotic
507 and/or oxidative stress responses [50] and the iGRN identified additional members with an
508 oxidative stress phenotype. In addition, the iGRN correctly predicted an oxidative stress
509 response function for one or more members of the bHLH, ZAT (or C2H2), HSF, MYB and bZIP
510 families that are also often associated with functions in plant stress responses. The iGRN also
511 identified novel ROS TFs from families that were known to be mainly involved in plant
512 growth/developmental processes, such as GRAS (SCL13), BES1 (BEH2) and GATA (BME3) ([45,51,52]) .
513 So far there is very little evidence for the involvement of GATA TFs in plant stress responses:
514 during low nitrogen stress [53,54] and by the identification of differentially expressed alternative
515 splice products of rice GATA TFs during salt and drought stress [55]. BME3 is therefore one of the
516 first GATA transcription factors with a proven function in the regulation of stress responses.
517 Exploring the GATA family might provide novel leads for regulators and novel functional links
518 with plant stress responses.
519

520 **DISCUSSION**

521 The comprehensive experimental mapping of plant GRNs is a major challenge and different
522 technologies, either *in vitro* or *in vivo*, provide partial snapshots of the regulatory logic in
523 *Arabidopsis*. While different integrative approaches have been applied to identify regulatory
524 networks controlling specific biological processes in plants [12,56-59], in this study we leveraged a

525 wide variety of regulatory datasets capturing different properties of TF-mediated transcriptional
526 control. Through the application of an advanced machine learning approach for large-scale
527 functional data integration, we performed for the first time a systematic regulatory annotation
528 for nearly all *Arabidopsis* genes. Although the presented iGRN is generic and not tailored
529 towards specific plant organs or growth stages, extensive validation using different types of
530 experimental datasets, measuring TF binding and/or regulation, revealed a strong enrichment
531 for functional interactions. In addition, the iGRN outperformed all input networks when
532 assessing the prediction power using a gold standard evaluation set, demonstrating the success
533 of the supervised learning approach. An assessment of the functional coherence of a TF's target
534 genes suggests that the iGRN regulatory interactions are of equally high quality compared to
535 those obtained through experimental methods. Taken together, our validation results indicate
536 that the iGRN is strongly enriched for functional interactions and that the applied integrative
537 approach clearly complements GRNs inferred through a single experimental assay.
538 Investigating the relationship between network architecture and the complexity of TF control
539 confirmed that TFs themselves are complexly regulated by other TFs. In addition, through the
540 integration of large-scale phenotyping data, a significant association was found between a TF's
541 network centrality (but not degree) and observing a phenotype when knocking out that TF. This
542 finding indicates that these TFs occupy key positions in the GRN and loss of their gene products
543 has major consequences for plant growth and development. Furthermore, TFs showing a
544 phenotype when screening mutant lines were not having more TFs among their target genes
545 compared to TFs lacking phenotypes, suggesting that master regulators high in the TF network
546 hierarchy are not more frequently showing phenotypes compared to lower-tier TFs. This finding
547 complements previous results reporting that higher-tier TFs do not show more differential
548 expression during ABA response [26]. Finally, TF pairs sharing many target genes are indicative of
549 combinatorial control, where TFs physically interacting with each other form higher-order
550 protein complexes mediating increased DNA-binding specificity [60]. Consequently, TF pairs with
551 many shared target genes in our integrative network which also show correlated expression
552 profiles are good candidates for future research to shed light on the role of cooperative TF
553 control in specific biological processes in plants.
554 Predicted TF functions based on the iGRN regulatory interactions indicated that for various
555 biological processes many known regulators could be recovered. Starting from our systematic
556 regulatory annotation, known functional annotations were confirmed for 681 TFs. For several

557 biological processes our predicted TF functions are highly complementary to the AraNet
558 probabilistic functional gene network and for 268 TFs lacking experimental annotations, new
559 functions were inferred. To validate the power of our integrative network, novel regulators
560 predicted to play a role in oxidative stress signaling were experimentally validated using ROS-
561 specific phenotyping of loss- or gain-of-function lines. By analyzing transgenic lines perturbed in
562 one of the 32 novel ROS TFs, an additional 13 (41%) regulators were validated for a function in
563 ROS responses by our phenotyping studies. In addition to TF families that have often been
564 associated with oxidative or other environmental stress responses, the iGRN identified novel
565 ROS TFs from the GRAS, BES1 and GATA families that are not, or to a lesser extent, known to be
566 involved in stress responses. Notable, compared to previous studies that predicted gene
567 functions or prioritized candidate regulators based on (co-)expression analysis [61], our method
568 enables the identification of novel regulators that would have been missed when solely relying
569 on (differential) expression information. To conclude, our integrative network offers a high-
570 quality resource to study TF control for individual target genes and to identify, through the
571 interrogation of specific or condition-specific gene sets, novel regulators controlling different
572 plant traits.
573

574 **METHODS**

575 **Input datasets**

576 We assembled a comprehensive compendium of genome-wide datasets that were used as
577 primary data sources to construct the seven input networks and to build the integrative gene
578 regulatory network. Starting from the TAIR10 Arabidopsis gene annotation, we added a set of
579 5711 non-coding RNAs described in [62] resulting in a dataset covering 38,966 genes. A first
580 dataset consisted of 1,260 PWM binding site profiles for 874 transcription factors. This dataset
581 contained 108 and 623 positional frequency matrices that were obtained from protein binding
582 microarray studies performed by [63] and [64], respectively, and 529 positional frequency matrices
583 obtained from DAP-seq experiments [65]. A second dataset contained all available DNaseI
584 hypersensitive data obtained from the plant regulome data resource
585 (http://www.plantregulome.org/, [14]). All separate conditions were merged into a single open
586 chromatin dataset. A third dataset, composed of evolutionary conserved binding sites for 874

587 TFs, was obtained using the comparative motif mapping algorithm as described in [15], starting
588 from the input motif dataset as described above.
589 A fourth dataset comprised gene expression profiles generated by RNA-seq. The RNA-Seq
590 expression compendium was built with public datasets from NCBI's Sequence Read Archive
591 (SRA) using Curse [66]. The compendium contains gene-level expression values for 40 manually
592 selected samples (Supplemental Table S8) of different treatment and tissue combinations (in
593 total >2,7 billion reads). SRA files for each sequencing run were downloaded from the SRA and
594 converted to the FASTQ format using fastq-dump (v2.4.4) from the SRA toolkit. FASTQ files from
595 runs of the same sample were concatenated. Paired-end reads were unpaired by randomly
596 selecting either the forward or reverse read and processing it as single-end. FastQC (v0.9.1) was
597 used to detect overrepresented adapter sequences, which were subsequently clipped with
598 fastx_clipper from the FASTX toolkit (v0.0.13). Nucleotides with Phred quality scores lower than
599 20 were trimmed with fastq_quality_trimmer from the FASTX toolkit. Reads shorter than 20
600 nucleotides after quality trimming were discarded. To obtain raw read counts for each transcript
601 in the TAIR10 annotation, Sailfish (v0.6.3) [67] was run with a k-mer length of 20. For genes with
602 multiple transcripts, the raw read counts of its transcripts were summed to get a gene-level read
603 count. Counts were then normalized for the entire compendium with the Variance Stabilizing
604 Transformation (VST) from the DESeq R package (v1.14.0) [68]. VST was chosen since it results in
605 correlation coefficients between genes that are most comparable to those obtained with
606 microarray data [69]. A fifth dataset consisted of a compendium of 27 genome-wide TF binding
607 datasets obtained through ChIP-seq [7].

608 **Network inference for the different input datasets**

609 We derived input networks from each dataset and a weight was assigned to each TF-target
610 gene interaction reflecting the certainty of the regulatory interaction. The motif target gene
611 network was created by mapping all binding site data on an extended locus for each gene in the
612 *Arabidopsis thaliana* genome spanning 2kb upstream, intron and 1kb downstream regions, with
613 the exons masked. All binding site specificity matrices were mapped genome-wide using Cluster-
614 Buster with C parameter set to zero and M parameter set to 5 [70]. Both the C and M scores were
615 used as weights. Due to the short length of motifs, many instances of motifs in a single genome
616 are expected to be coincidental matches that may not be bound *in vivo* or have no functional
617 regulatory effect. Therefore, the resulting set of binding sites was filtered using DNaseI

618 hypersensitive sites [14] in order to obtain an open chromatin filtered motif network. The
619 signalValue reported by the authors was retained as weight for motif matches in open
620 chromatin, this is a measurement of overall enrichment for the region to DNaseI
621 hypersensitivity. A second filtering approach was based on the evolutionarily conservation of
622 these binding sites resulting in a conserved motif network. Weights in the conserved motif
623 network are defined by the species conservation score according to the CMM method as
624 described in [23]. Co-expression clusters were built around all TFs obtained from PlantTFDB [71]
625 using the Pearson correlation coefficient and only the top 1000 most co-expressed genes were
626 retained as putative target genes. GENIE3 was run using standard parameters with the same list
627 of regulators, only predicted interactions with a variance in expression explained larger than
628 0.005 were retained [16]. A k-nearest neighbor (k-NN) co-expression cluster of size 100 was built
629 around each gene in the *Arabidopsis thaliana* genome using the RNA-seq compendium as
630 defined above. PWM motif enrichment was performed on this co-expression cluster using the
631 hypergeometric distribution and significantly enriched motifs (p-value < 0.001 after Benjamini–
632 Hochberg correction for multiple hypotheses testing) were assigned to the members of the co-
633 expression cluster in order to obtain the k-NN motif enrichment network. The enrichment fold
634 of the motif in the co-expression cluster is used as a weight. The peak scores for each TF bound
635 region assigned to the closest gene were used as weight in the ChIP-seq network [7].

636 **Integrative network inference**

637 The integrative network inference method is based on a binary classifier, where the class label
638 represents the presence or absence of an interaction. The set of true interactions (positives)
639 used in the training set was obtained from all validated interactions in AtRegNet [18], a review
640 article on known interactions in secondary cell wall development [19] and large scale data mining
641 effort on 974 peer reviewed articles called ATRM [20], and has 5,732 edges, connecting 522 TFs
642 with 3,738 targets. The remaining interactions between these TFs and potential target genes are
643 considered to be negatives. In other words, while no true "negatives" are present in the dataset,
644 we assign negative labels to TF–target pairs where the TF appears in one of the positive pairs.
645 The negative labels reflect the fact that the TF has been studied and since no interaction is
646 reported for it, it is more likely to represent a negative interaction. This dataset was split into a
647 dataset for learning (80%) and a dataset for independent evaluation (20%). As such, the training
648 set consists of 1,951,236 TF–target gene pairs, out of which 4,586 are positives (true

649 interactions) and 1,946,650 negatives. The remaining gene pairs for TFs absent from the
650 learning/evaluation dataset are considered to be negatives. Using the weights derived from the
651 input networks we tested a number of different classification methods. We extracted the 4,586
652 validated interacting pairs from the true dataset and obtained the negative pairs for training by
653 selecting a random subset of 13,758 (three times the number of positives; 3x 4,586) negatively
654 labeled interacting pairs. Using this labeled training set different classifiers were built using the
655 weights of the input networks as features. All features were normalized to have zero mean and
656 unit variance. Classifiers were evaluated using the 20% evaluation dataset. In this analysis the
657 gradient boosting using decision trees, random forests, and logistic regression were compared.
658 The gradient boosting model made use of 1,000 trees, shrinkage of 0.01 and an interaction
659 depth of 3. The random forest model used 500 trees selecting a random set of features of size
660 square root (features) was sampled in each iteration. Logistic regression was used with a
661 binomial distribution suitable for binary classification. All classifiers were compared using the
662 area under the curve (Supplemental Table S2).
663 The final network was generated using the gradient boosting classifier, as this was the best
664 performing throughout the training and validation steps. A total set of 60,714,507 unlabeled
665 interactions was classified. We included all interactions that are predicted to interact with a
666 probability of 0.35 or higher in the integrative network as this cutoff maximized the F1 measure
667 on the test dataset (Supplemental Table S2).

668 **Functional evaluation using experimental datasets**

669 Three experimentally obtained gene regulatory networks were used for further validation of
670 the predicted network. An accumulation of publicly available Y1H data from literature was used
671 to generate a Y1H network containing 2,759 interactions [72-75]. A set of differentially expressed
672 genes upon TF perturbation was obtained from the CORNET database, containing 51,178
673 interactions [76]. An independent set of interactions determined by ChIP-seq was obtained from
674 literature, reporting 148,949 interactions [26,77-80]. From Song et al. 2016 only non-ABA conditions
675 were considered.
676 The overlap between the iGRN and three experimentally determined gene regulatory
677 networks was evaluated by counting how many interactions from the experimental networks
678 were present in the integrative network. The enrichment between two networks was defined as
679 the number of interactions that are present in both networks divided by the number of

680 interactions expected by chance. The number of common interactions expected by chance is
681 given by the mean overlap of 10,000 randomized networks [21]. Network randomization was done
682 by permuting the labels of all TFs and the labels of all genes, which preserves the network
683 structure. This assures that the observed enrichment is not due to potential biases arising from
684 structural properties of the network (Supplemental Table S3).
685 Functional enrichment was determined for each sub network (intersection, FN, FP) using
686 biological metrics as described in [23] using Gene Ontology information. Note that no functional
687 GO information was used as input for the iGRN construction.

688 **Analysis of network architecture and overlap with plant phenotypes**

689 Based on the target genes for all TFs present in the iGRN, degree and betweenness centrality
690 were computed using the igraph R package. Based on bulk downloads of the RARGE II database
691 ([27]; download November 2018), which reports ontology-based phenotype information for
692 48,077 lines (17,671 RIKEN Arabidopsis Ds transposon mutant lines, 16,337 CSHL Arabidopsis
693 genetrap mutant lines, 14,069 RIKEN Arabidopsis full-length cDNA overexpressed Arabidopsis
694 lines). The presence of phenotypes in the RARGE II dataset was scored for different sets of
695 regulators, defined based on degree and betweenness centrality.

696 **TF function prediction**

697 For each TF, iGRN target genes were subjected to GO enrichment analysis using the
698 hypergeometric distribution combined with Benjamini–Hochberg (B&H) correction for multiple
699 hypotheses testing. A q-value cutoff 1e-03 was applied since this resulted in the best F1 score, a
700 metric that combines precision and recall to recover known GO annotations. Only experimental
701 and curated GO BP annotations were considered (version August 2018), GO annotations were
702 propagated to include parental GO terms, and a set of 18 general GO terms, corresponding with
703 terms located at the root of the GO hierarchy, were excluded (GO:0008150, GO:0009987,
704 GO:0008152, GO:0044237, GO:0071704, GO:0050896, GO:0065007, GO:0032502, GO:0050789,
705 GO:0032501, GO:0007275, GO:0050794, GO:0006355, GO:0045893, GO:0045892).
706 To predict ROS TFs, core genes part of the ROS wheel clusters IV-VIII, covering genes related to
707 chemical and genetic perturbation of cellular ROS/redox, were used to define a custom category
708 of ROS marker genes [30]. Subsequently, the over-representation of these ROS markers in the set
709 of iGRN target genes per TF, was used to predict novel ROS TFs and rank them based on the q-

710 value for enrichment (hypergeometric distribution, q-value cutoff 1e-05). The assignment of ROS
711 TFs to the categories KNOWN_ROS, KNOWN_STRESS and NOVEL_TF was based on the
712 information available in the literature at the moment the initial iGRN was constructed
713 (December 2016).

714 **Phenotyping candidate ROS TFs**

715 T-DNA insertion mutants were obtained from the European Arabidopsis Stock Centre (NASC;
716 http://arabidopsis.info/), GABI-Kat (http://www.gabi-kat.de/db/), Versailles Arabidopsis Stock
717 Center (http://publiclines.versailles.inra.fr/tdna/index) (Supplemental Table S6). Homozygous T-
718 DNA insertion lines were selected by PCR analysis (Supplemental Table S9). Mutant lines
719 previously published were provided by the authors (Supplemental Table S6 (overview lines)).
720 Overexpression lines were generated by Gateway cloning in the binary destination vector
721 pK7WG2D (18056864) (Supplemental Table S9 (primers)). Plants were grown on half-strength
722 Murashige and Skoog (1/2MS) medium (Duchefa) supplemented with 1% (w/v) Suc, 0.75% (w/v)
723 agar, and B5 vitamins, pH 5.7, at 21°C and 60 µE m−2 s−1 light intensity in a 16-h/8-h light/dark
724 photoperiod. To induce oxidative stress, the 1/2MS medium was supplemented with 15 or 20
725 nM MV; or 1 or 1.5 µM 3-AT, causing on average 15% (15 nM MV), 30% (20 nM MV), 20% (1 µM
726 3-AT) and 55% (1.5 µM 3-AT) growth reduction compared to the nontreated seedlings. Rosette
727 area was determined by automated area measurements of individual plant rosettes using
728 ImageJ. Raw phenotyping results are available in Data file 3. Presence or absence of
729 chlorosis/bleaching was monitored by visual assessment. For each mutant line, the rosette area
730 of mutant seedlings was compared with wild-type plants grown in the same Petri plate. Data
731 were obtained using at least two biological replicates per experiment and from independent
732 experiments.

733 **Statistical analysis leaf phenotyping data ROS TFs**

734 A mixed model was fitted to the rosette area with genotype and treatment as fixed effects as
735 well as the interaction term with the mixed procedure from SAS (SAS Institute Inc. 2018.
736 SAS/STAT® 15.1, Cary, NC). The factor plate was added as random effect in the model to take
737 into account the correlation between observations made on the same plate. The Kenward-
738 Rogers method was used to calculate the denominator degrees of freedom of the fixed effects.
739 The null hypotheses of interest were an equal genotype effect upon treatment in the mutant

line and the wild-type line and an equal treatment versus MS effect between the mutant line and the wild-type line (i.e. no interaction effect). A significance level of 0.05 was taken. To correct for multiple testing, the maxT procedure was used as implemented in the plm procedure.

ACKNOWLEDGEMENTS

The authors wish to thank Cordelia Bolle (SCL13), Xuemei Chen (ARF3), Yi-Fang Chen (WRKY45), Donna Fernandez (AGL15), Henning Frerigmann (MYB51, MYB122), Małgorzata Gaj (ERF022), Tamara gigolashvili (MYB51), Dwayne Hegedus (SCL15), Lars Hennig (AGL19), Ildoo Hwang (BME3), Zhongchi Liu (TSO1), Lars Østergaard (ARF3), Javier Paz-Ares (PHL1) and Olivier Van Aken (WRKY45) for kindly providing Arabidopsis mutant lines. We thank Zoé Joly-Lopez and Olivia Wilkins for providing in-depth comments on an earlier version of the manuscript, as well as Camilla Ferrari and Indeewari Dissanayake for proofreading. I.D.C. is supported by the Research Foundation Flanders (postdoctoral fellowship 12N2415N and grant for a long stay abroad V400620N). J.V.d.V. is indebted to the Agency for Innovation by Science and Technology in Flanders for a predoctoral fellowship. X.L. and L.L. are supported by the China Scholarship Council for a PhD fellowship (201706910099 and 201808530499). F.V.B. is supported by the Fonds Wetenschappelijk Onderzoek – Vlaanderen (FWO) and the Fonds de la Recherche Scientifique – FNRS under EOS Project No. 30829584 and the Research Foundation Flanders (project nos. G0D7914N and G055416N).

CONTRIBUTIONS

J.V.d.V., I.D.C. and K.V. designed the research. I.D.C., X.L., L.L. and R.P. performed the phenotype screening. V.S. performed statistical data analysis. D.V. generated the expression compendium. J.V.d.V. and K.V. performed network analysis and function prediction. J.V.d.V., I.D.C., F.V.V. and K.V. wrote the manuscript.

765 **TABLES**

766 **Table 1. Overview of different networks used to construct the iGRN.**

Type	Network name	#interactions	#TFs	Median #interactions (average)
Input network	PWM	17,053,785	874	21,797 (19,512)
	Conserved PWM	340,893	874	247 (390)
	PWM in DHS	5,231,258	874	6,195 (5,985)
	COE + PWM	324,864	703	290 (462)
	TF ChIP	46,619	27	1,026 (1,727)
	COE TF-target	1,547,002	1547	1,000 (1,000)
	GENIE3	2,588,312	1261	1,985 (2,053)
Gold standard	Gold standard	5,732	522	2 (11)
Evaluation	Test dataset	1,146	257	1 (4)
Output network	iGRN	1,709,871	1491	802 (1,147)

767

Table 2. Overview phenotype information for known and validated novel ROS regulators.

RANK iGRN	Symbol	q-val	Status (1)	Known oxidative stress phenotype from literature	Reference	Known other (a/biotic) stress phenotype from literature	Reference	Validated oxidative stress phenotype from this study
iGRN1	ANAC102	1.84E-35	**NOVEL**			high light	[31]	**MV**
iGRN3	ZAT11	2.81E-25	KNOWN_ROS	MV	[81]	heavy metal	[82]	
iGRN5	ZAT12	8.70E-25	KNOWN_ROS	MV	[83]	osmotic/salt/high light/heat	[84]	
iGRN6	WRKY30	1.42E-24	KNOWN_ROS	MV	[85]	salt	[85]	
iGRN11	HSFA4A	5.09E-23	KNOWN_ROS	MV	[86]	salt, anoxia	[86]	
iGRN12	WRKY15	7.46E-23	KNOWN_ROS	H2O2	[87]	salt/osmotic	[87]	
iGRN16	STZ	4.90E-22	KNOWN_ROS	MV	[88]	osmotic/salt/heat	[88]	
iGRN19	SCL13	3.76E-21	**NOVEL**					**MV**
iGRN22	MYB51	1.84E-20	**NOVEL**					**3-AT, MV**
iGRN27	BME3	2.01E-18	**NOVEL**					**MV**
iGRN29	WRKY55	6.69E-18	**NOVEL**					**3-AT, MV**
iGRN36	ERF6	3.13E-15	KNOWN_ROS	H2O2	[89]	high light, B. cinerea, osmotic	[89 90 91]	

iGRN50	WRKY45	5.93E-13	**NOVEL(**)**			submergence stress	[43]	**3-AT, MV**
iGRN52	ANAC032	1.26E-12	KNOWN_ROS	3-AT	[92]	P. syringae, salt/high light	[92 93]	
iGRN55	AT2G33710	9.71E-12	**NOVEL**					**3-AT, MV**
iGRN58	MYB122	1.42E-10	**NOVEL**					**3-AT, MV**
iGRN60	WRKY28	1.70E-10	KNOWN_ROS	MV, H2O2	[94]	salt/osmotic/ drought, S. sclerotiorum, A. brassicicola	[94 95 96]	
iGRN62	ANAC013	1.93E-10	KNOWN_ROS	MV	[97]			
iGRN65	ABF3	8.06E-10	KNOWN_ROS	MV	[98]	cold/freezing/heat /drought	[98]	
iGRN66	ERF022	9.55E-10	**NOVEL**					**MV**
iGRN68	BEH2	2.75E-09	**NOVEL(**)**			drought	[32]	**3-AT, MV**
iGRN69	ANAC017	3.04E-09	KNOWN_ROS	MV	[99]	drought/high light	[100]	
iGRN81	PHL1	2.76E-08	**NOVEL**					**3-AT, MV**
iGRN85	JAM1	4.07E-08	KNOWN_ROS	MV, H2O2	[94]	salt/osmotic/ drought	[94]	
iGRN93	MYC2	4.19E-07	KNOWN_ROS	MV	[101]	salt, B. Cinerea	[102 103]	
iGRN100	AT5G57150	1.50E-06	**NOVEL**					**3-AT, MV**
iGRN102	ERF115	1.73E-06	**NOVEL**					**3-AT, MV**
iGRN110	JUB1	3.20E-06	KNOWN_ROS	H2O2	[104]	salt/heat, P. syringae, drought	[104 105 106]	
iGRN111	HSFA2	3.21E-06	KNOWN_ROS	MV	[107]	high light, heat, cold, hypoxia	[107 108 109 110 111]	
iGRN118	CBF1	5.32E-06	KNOWN_ROS	MV	[112]	salt, drought, freezing	[112 113 114]	

(1) Asterisk indicates NOVEL TFs that were validated for a function in ROS (*) or (a)biotic stress (**) responses by literature evidence after 2016.

FIGURE LEGENDS

Figure 1. Construction and evaluation of the integrated gene regulatory network (iGRN).

A. Overview of the different input networks used to learn the supervised iGRN. Whereas colored circles indicate the type of data that is used to construct the input network, the information below each network depicts how edges in that network are weighted. B. Overview of the sizes of the different networks together with the recall and enrichment for edges in the evaluation test dataset (green and blue bars, respectively). The yellow bars report the fraction of each input netwo rk that is present in the iGRN (yellow) while the red bars report which fraction of the iGRN edges is supported by this input network.

Figure 2. Overlap between iGRN interactions and different experimental data sources describing gene regulatory and functional information.

A. Venn diagrams showing the number of shared interactions between the iGRN and TF ChIP, Y1H and differentially expressed (DE) genes after TF perturbation (depicted in blue, yellow and red, respectively). B. Histograms showing the expected overlap based on 10,000 randomized networks, together with the observed overlap indicated by the black arrow. The p-values for observing the measured overlap by chance are also reported. C. Fold enrichment for a functional metric describing how well target genes of a specific TF are involved in similar biological processes. Whereas the grey bars indicate the functional enrichment for shared interactions, blue, red, yellow and green depict fold enrichment for interactions unique to the ChIP, DE, Y1H and iGRN, respectively. D. Sub-networks showing experimental protein-protein interactions for TFs sharing a large number of target genes (Jaccard index >0.3). Whereas the edge thickness and values reports the fraction of shared target genes for a pair of TFs quantified using the Jaccard index, the color represents the absolute value of the Pearson Correlation Coefficient between the expression profiles of a pair of TFs.

Figure 3. Phenotypes reported after TF perturbation for TFs with different network centrality properties.

A. RARGE-II phenotypes reported for iGRN TFs with different betweenness centrality. Q1-Q4 refer to the four quartiles of all iGRN TFs (sorted based on increasing betweenness centrality)

and values in parenthesis refer to the median centrality value together with the number of lines available in the RARGE II dataset for that quartile. B. Betweenness centrality distribution for TFs with and without phenotypes reported in RARGE-II (n= 153 and 1589, respectively).

Figure 4. Oxidative stress phenotyping results of novel ROS TFs.

Heatmap presenting phenotypes associated with mutant alleles of iGRN extracted novel ROS TFs under different oxidative stress conditions. Mutant lines with a statistically significant phenotype under oxidative stress (1 µM and 1.5 µM 3-AT and 15 and 20 nM MV) and/or normal growth conditions (MS) are displayed. Overview (of the results) of other mutant lines that were tested in this study can be found in Supplemental Table S5 and S6. Statistically significant differences relative to wild type (Col-0) are indicated by asterisk; *, $P < 0.05$ and **, $P < 0.001$ as determined by the hypothesis tests on the fixed effects of the mixed model (see Methods). Color scale indicates the ratio of the rosette area of 14-day-old mutant alleles relative to wild type (Col-0). Black and grey borders indicate presence or absence of a bleaching phenotype of the mutant plants, respectively. Small grey circle indicates a statistically significant genotype-by-stress treatment interaction effect.

Figure 5. Representative images of oxidative stress phenotypes for a selection of mutants.

Representative images of novel ROS TFs with a phenotype during oxidative stress. 17-days-old seedlings of selected mutant lines compared to wild type (Col-0).

Figure 6. Expression of KNOWN and NOVEL ROS TFs during various conditions provoking cellular ROS/redox imbalances.

Heatmap showing the expression response of ROS TFs extracted from the iGRN during a selection of conditions provoking cellular ROS production or redox imbalances (a more elaborate set of oxidative stress conditions is displayed in Supplemental Table S7). TFs with known function in ROS responses (KNOWN_ROS) and TFs with no functional link with any oxidative stress and/or (a)biotic stress responses (NOVEL) that were tested in our phenotyping assays are displayed. NOVEL TFs validated in our study are indicated: blue and red circles indicate a positive and negative effect by the TF on oxidative stress tolerance, respectively. * indicates validated by recently published studies after the time the iGRN was constructed.

OVERVIEW OF SUPPLEMENTAL FIGURES AND TABLES

Figure S1. Overview of the iGRN interactions supported by different input networks.

Figure S2. Overview of the in- and out-degree for the complete iGRN.

Figure S3. Recovery of known GO annotations for TF belonging to different biological processes.

Supplemental Table S1. Overview of TF families in gold standard.

Supplemental Table S2. Overview of data types, training methods and performance results for different machine learning classification algorithms.

Supplemental Table S3. Overlap and enrichment between different networks.

Supplemental Table S4. Overlap of target genes between iGRN and ChIP-Seq data.

Supplemental Table S5. Overview of all TFs predicted by the iGRN with a function in ROS responses.

Supplemental Table S6. Overview of mutant lines used for phenotypic analysis.

Supplemental Table S7. Expression of KNOWN_ROS and NOVEL ROS TFs during various conditions provoking cellular ROS/redox imbalances.

Supplemental Table S8. Overview of samples included in the RNA-Seq expression compendium.

Supplemental Table S9. Primers used in this study.

DATA AVAILABILITY

Source data for Figs. 1,4 and S3 are provided with the paper.

Data file 1. Data_file1_iGRN_supervised_network_support.txt All TF-target interactions in the iGRN.

Columns report: TF, target genes, score, rank, support input networks.

Overview of generated, analysed and used iGRN regulatory interaction data for Fig. 1.

Data file 2. TF_function_all.txt reporting predicted functions for all TFs present in the iGRN.

The sheet README explains the different columns and offers 3 examples how this file can be explored for specific TFs, functions or a specific target gene.

Overview of generated, analysed and used iGRN TF function data for Fig. S3.

Data file 3. Raw phenotype results that were used as input for the statistical analysis.

Overview of raw phenotyping measurements for Fig. 4.

REFERENCES

1 Franco-Zorrilla, J. M. & Solano, R. Identification of plant transcription factor target sequences. *Biochim Biophys Acta* **1860**, 21-30, doi:10.1016/j.bbagrm.2016.05.001 (2017).

2 Lehti-Shiu, M. D., Panchy, N., Wang, P., Uygun, S. & Shiu, S. H. Diversity, expansion, and evolutionary novelty of plant DNA-binding transcription factor families. *Biochim Biophys Acta Gene Regul Mech* **1860**, 3-20, doi:10.1016/j.bbagrm.2016.08.005 (2017).

3 Jones, D. M. & Vandepoele, K. Identification and evolution of gene regulatory networks: insights from comparative studies in plants. *Curr Opin Plant Biol* **54**, 42-48, doi:10.1016/j.pbi.2019.12.008 (2020).

4 Rhee, S. Y. & Mutwil, M. Towards revealing the functions of all genes in plants. *Trends Plant Sci* **19**, 212-221, doi:10.1016/j.tplants.2013.10.006 (2014).

5 Gaudinier, A. & Brady, S. M. Mapping Transcriptional Networks in Plants: Data-Driven Discovery of Novel Biological Mechanisms. *Annu Rev Plant Biol* **67**, 575-594, doi:10.1146/annurev-arplant-043015-112205 (2016).

6 Ferrier, T., Matus, J. T., Jin, J. & Riechmann, J. L. Arabidopsis paves the way: genomic and network analyses in crops. *Curr Opin Biotechnol* **22**, 260-270, doi:10.1016/j.copbio.2010.11.010 (2011).

7 Heyndrickx, K. S., Van de Velde, J., Wang, C., Weigel, D. & Vandepoele, K. A functional and evolutionary perspective on transcription factor binding in Arabidopsis thaliana. *Plant Cell* **26**, 3894-3910, doi:10.1105/tpc.114.130591 (2014).

8 Lu, Z., Hofmeister, B. T., Vollmers, C., DuBois, R. M. & Schmitz, R. J. Combining ATAC-seq with nuclei sorting for discovery of cis-regulatory regions in plant genomes. *Nucleic Acids Res* **45**, e41, doi:10.1093/nar/gkw1179 (2017).

9 Maher, K. A. *et al.* Profiling of Accessible Chromatin Regions across Multiple Plant Species and Cell Types Reveals Common Gene Regulatory Principles and New Control Modules. *Plant Cell* **30**, 15-36, doi:10.1105/tpc.17.00581 (2018).

10 Sijacic, P., Bajic, M., McKinney, E. C., Meagher, R. B. & Deal, R. B. Changes in chromatin accessibility between Arabidopsis stem cells and mesophyll cells illuminate cell type-specific transcription factor networks. *Plant J* **94**, 215-231, doi:10.1111/tpj.13882 (2018).

11 Kulkarni, S. R., Jones, D. M. & Vandepoele, K. Enhanced Maps of Transcription Factor Binding Sites Improve Regulatory Networks Learned from Accessible Chromatin Data. *Plant Physiol* **181**, 412-425, doi:10.1104/pp.19.00605 (2019).

12 Pajoro, A. *et al.* Dynamics of chromatin accessibility and gene regulation by MADS-domain transcription factors in flower development. *Genome biology* **15**, R41 (2014).

13 Lee, I., Ambaru, B., Thakkar, P., Marcotte, E. M. & Rhee, S. Y. Rational association of genes with traits using a genome-scale gene network for Arabidopsis thaliana. *Nat Biotechnol* **28**, 149-156, doi:10.1038/nbt.1603 (2010).

14 Sullivan, A. M. *et al.* Mapping and dynamics of regulatory DNA and transcription factor networks in A. thaliana. *Cell reports* **8**, 2015-2030 (2014).

15 Van de Velde, J., Van Bel, M., Vaneechoutte, D. & Vandepoele, K. A Collection of Conserved Noncoding Sequences to Study Gene Regulation in Flowering Plants. *Plant Physiol* **171**, 2586-2598, doi:10.1104/pp.16.00821 (2016).

16 Huynh-Thu, V. A., Irrthum, A., Wehenkel, L. & Geurts, P. Inferring regulatory networks from expression data using tree-based methods. *PLoS One* **5**, doi:10.1371/journal.pone.0012776 (2010).

17 Marbach, D. *et al.* Wisdom of crowds for robust gene network inference. *Nat Methods* **9**, 796-804, doi:10.1038/nmeth.2016 (2012).
18 Yilmaz, A. *et al.* AGRIS: the Arabidopsis Gene Regulatory Information Server, an update. *Nucleic Acids Res* **39**, D1118-1122, doi:10.1093/nar/gkq1120 (2011).
19 Hussey, S. G., Mizrachi, E., Creux, N. M. & Myburg, A. A. Navigating the transcriptional roadmap regulating plant secondary cell wall deposition. *Front Plant Sci* **4**, 325, doi:10.3389/fpls.2013.00325 (2013).
20 Jin, J. *et al.* An Arabidopsis Transcriptional Regulatory Map Reveals Distinct Functional and Evolutionary Features of Novel Transcription Factors. *Mol Biol Evol* **32**, 1767-1773, doi:10.1093/molbev/msv058 (2015).
21 Marbach, D. *et al.* Predictive regulatory models in Drosophila melanogaster by integrative inference of transcriptional networks. *Genome Res* **22**, 1334-1349, doi:10.1101/gr.127191.111 (2012).
22 Kheradpour, P., Stark, A., Roy, S. & Kellis, M. Reliable prediction of regulator targets using 12 Drosophila genomes. *Genome Res* **17**, 1919-1931, doi:10.1101/gr.7090407 (2007).
23 Van de Velde, J., Heyndrickx, K. S. & Vandepoele, K. Inference of transcriptional networks in Arabidopsis through conserved noncoding sequence analysis. *The Plant Cell* **26**, 2729-2745 (2014).
24 Tompa, M. *et al.* Assessing computational tools for the discovery of transcription factor binding sites. *Nat Biotechnol* **23**, 137-144, doi:10.1038/nbt1053 (2005).
25 O'Maoileidigh, D. S., Graciet, E. & Wellmer, F. Gene networks controlling Arabidopsis thaliana flower development. *New Phytol* **201**, 16-30, doi:10.1111/nph.12444 (2014).
26 Song, L. *et al.* A transcription factor hierarchy defines an environmental stress response network. *Science* **354**, aag1550 (2016).
27 Akiyama, K. *et al.* RARGE II: an integrated phenotype database of Arabidopsis mutant traits using a controlled vocabulary. *Plant Cell Physiol* **55**, e4, doi:10.1093/pcp/pct165 (2014).
28 Lee, T. *et al.* AraNet v2: an improved database of co-functional gene networks for the study of Arabidopsis thaliana and 27 other nonmodel plant species. *Nucleic Acids Res* **43**, D996-1002, doi:10.1093/nar/gku1053 (2015).
29 Kamiya, T. *et al.* The MYB36 transcription factor orchestrates Casparian strip formation. *Proc Natl Acad Sci U S A* **112**, 10533-10538, doi:10.1073/pnas.1507691112 (2015).
30 Willems, P. *et al.* The ROS Wheel: Refining ROS Transcriptional Footprints. *Plant Physiol* **171**, 1720-1733, doi:10.1104/pp.16.00420 (2016).
31 D'Alessandro, S., Ksas, B. & Havaux, M. Decoding beta-Cyclocitral-Mediated Retrograde Signaling Reveals the Role of a Detoxification Response in Plant Tolerance to Photooxidative Stress. *Plant Cell* **30**, 2495-2511, doi:10.1105/tpc.18.00578 (2018).
32 Kim, H. *et al.* Transcriptional network regulation of the brassinosteroid signaling pathway by the BES1-TPL-HDA19 co-repressor complex. *Planta* **250**, 1371-1377, doi:10.1007/s00425-019-03233-z (2019).
33 Ryu, T. H. *et al.* SOG1-dependent NAC103 modulates the DNA damage response as a transcriptional regulator in Arabidopsis. *Plant J* **98**, 83-96, doi:10.1111/tpj.14201 (2019).
34 Yin, M. *et al.* The Arabidopsis Cys2/His2 zinc finger transcription factor ZAT18 is a positive regulator of plant tolerance to drought stress. *J Exp Bot* **68**, 2991-3005, doi:10.1093/jxb/erx157 (2017).
35 Huang, P. Y. *et al.* NINJA-associated ERF19 negatively regulates Arabidopsis pattern-triggered immunity. *J Exp Bot* **70**, 1033-1047, doi:10.1093/jxb/ery414 (2019).

36 Scarpeci, T. E., Frea, V. S., Zanor, M. I. & Valle, E. M. Overexpression of AtERF019 delays plant growth and senescence, and improves drought tolerance in Arabidopsis. *J Exp Bot* **68**, 673-685, doi:10.1093/jxb/erw429 (2017).
37 Heyman, J. *et al.* ERF115 controls root quiescent center cell division and stem cell replenishment. *Science* **342**, 860-863, doi:10.1126/science.1240667 (2013).
38 Kong, X. *et al.* PHB3 Maintains Root Stem Cell Niche Identity through ROS-Responsive AP2/ERF Transcription Factors in Arabidopsis. *Cell Rep* **22**, 1350-1363, doi:10.1016/j.celrep.2017.12.105 (2018).
39 Chen, L., Xiang, S., Chen, Y., Li, D. & Yu, D. Arabidopsis WRKY45 Interacts with the DELLA Protein RGL1 to Positively Regulate Age-Triggered Leaf Senescence. *Mol Plant* **10**, 1174-1189, doi:10.1016/j.molp.2017.07.008 (2017).
40 Hinckley, W. E., Keymanesh, K., Cordova, J. A. & Brusslan, J. A. The HAC1 histone acetyltransferase promotes leaf senescence and regulates the expression of ERF022. *Plant Direct* **3**, e00159, doi:10.1002/pld3.159 (2019).
41 Wang, H. *et al.* Arabidopsis WRKY45 transcription factor activates PHOSPHATE TRANSPORTER1;1 expression in response to phosphate starvation. *Plant Physiol* **164**, 2020-2029, doi:10.1104/pp.113.235077 (2014).
42 Bustos, R. *et al.* A central regulatory system largely controls transcriptional activation and repression responses to phosphate starvation in Arabidopsis. *PLoS Genet* **6**, e1001102, doi:10.1371/journal.pgen.1001102 (2010).
43 Meng, X. *et al.* Mitochondrial signalling is critical for acclimation and adaptation to flooding in Arabidopsis thaliana. *Plant J* **103**, 227-247, doi:10.1111/tpj.14724 (2020).
44 Frerigmann, H. & Gigolashvili, T. MYB34, MYB51, and MYB122 distinctly regulate indolic glucosinolate biosynthesis in Arabidopsis thaliana. *Mol Plant* **7**, 814-828, doi:10.1093/mp/ssu004 (2014).
45 Yin, Y. *et al.* A new class of transcription factors mediates brassinosteroid-regulated gene expression in Arabidopsis. *Cell* **120**, 249-259, doi:10.1016/j.cell.2004.11.044 (2005).
46 Torres-Galea, P., Huang, L. F., Chua, N. H. & Bolle, C. The GRAS protein SCL13 is a positive regulator of phytochrome-dependent red light signaling, but can also modulate phytochrome A responses. *Mol Genet Genomics* **276**, 13-30, doi:10.1007/s00438-006-0123-y (2006).
47 Lachowiec, J., Mason, G. A., Schultz, K. & Queitsch, C. Redundancy, Feedback, and Robustness in the Arabidopsis thaliana BZR/BEH Gene Family. *Front Genet* **9**, 523, doi:10.3389/fgene.2018.00523 (2018).
48 Liu, P. P., Koizuka, N., Martin, R. C. & Nonogaki, H. The BME3 (Blue Micropylar End 3) GATA zinc finger transcription factor is a positive regulator of Arabidopsis seed germination. *Plant J* **44**, 960-971, doi:10.1111/j.1365-313X.2005.02588.x (2005).
49 Nowak, K., Wojcikowska, B. & Gaj, M. D. ERF022 impacts the induction of somatic embryogenesis in Arabidopsis through the ethylene-related pathway. *Planta* **241**, 967-985, doi:10.1007/s00425-014-2225-9 (2015).
50 Hoang, X. L. T., Nhi, D. N. H., Thu, N. B. A., Thao, N. P. & Tran, L. P. Transcription Factors and Their Roles in Signal Transduction in Plants under Abiotic Stresses. *Curr Genomics* **18**, 483-497, doi:10.2174/1389202918666170227150057 (2017).
51 Hirsch, S. & Oldroyd, G. E. GRAS-domain transcription factors that regulate plant development. *Plant Signal Behav* **4**, 698-700, doi:10.4161/psb.4.8.9176 (2009).

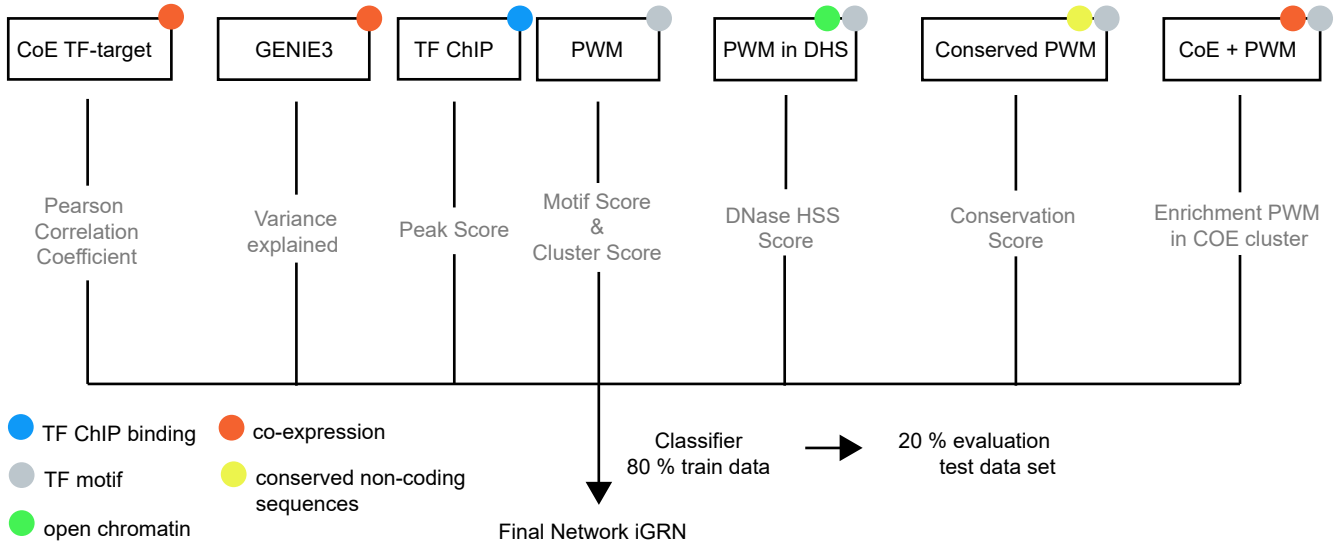
52 Behringer, C. & Schwechheimer, C. B-GATA transcription factors - insights into their structure, regulation, and role in plant development. *Front Plant Sci* **6**, 90, doi:10.3389/fpls.2015.00090 (2015).
53 Hudson, D. *et al.* Rice cytokinin GATA transcription Factor1 regulates chloroplast development and plant architecture. *Plant Physiol* **162**, 132-144, doi:10.1104/pp.113.217265 (2013).
54 Zhang, C. *et al.* Genome-wide survey of the soybean GATA transcription factor gene family and expression analysis under low nitrogen stress. *PLoS One* **10**, e0125174, doi:10.1371/journal.pone.0125174 (2015).
55 Gupta, P., Nutan, K. K., Singla-Pareek, S. L. & Pareek, A. Abiotic Stresses Cause Differential Regulation of Alternative Splice Forms of GATA Transcription Factor in Rice. *Front Plant Sci* **8**, 1944, doi:10.3389/fpls.2017.01944 (2017).
56 Brooks, M. D. *et al.* Network Walking charts transcriptional dynamics of nitrogen signaling by integrating validated and predicted genome-wide interactions. *Nat Commun* **10**, 1569, doi:10.1038/s41467-019-09522-1 (2019).
57 Gaudinier, A. *et al.* Transcriptional regulation of nitrogen-associated metabolism and growth. *Nature* **563**, 259-264, doi:10.1038/s41586-018-0656-3 (2018).
58 Zander, M. *et al.* Integrated multi-omics framework of the plant response to jasmonic acid. *Nat Plants* **6**, 290-302, doi:10.1038/s41477-020-0605-7 (2020).
59 Reynoso, M. A. *et al.* Evolutionary flexibility in flooding response circuitry in angiosperms. *Science* **365**, 1291-1295, doi:10.1126/science.aax8862 (2019).
60 Smaczniak, C., Muino, J. M., Chen, D., Angenent, G. C. & Kaufmann, K. Differences in DNA Binding Specificity of Floral Homeotic Protein Complexes Predict Organ-Specific Target Genes. *Plant Cell* **29**, 1822-1835, doi:10.1105/tpc.17.00145 (2017).
61 Ransbotyn, V. *et al.* A combination of gene expression ranking and co-expression network analysis increases discovery rate in large-scale mutant screens for novel Arabidopsis thaliana abiotic stress genes. *Plant Biotechnol J* **13**, 501-513, doi:10.1111/pbi.12274 (2015).
62 Liu, J. *et al.* Genome-wide analysis uncovers regulation of long intergenic noncoding RNAs in Arabidopsis. *Plant Cell* **24**, 4333-4345, doi:10.1105/tpc.112.102855 (2012).
63 Franco-Zorrilla, J. M. *et al.* DNA-binding specificities of plant transcription factors and their potential to define target genes. *Proceedings of the National Academy of Sciences* **111**, 2367-2372 (2014).
64 Weirauch, M. T. *et al.* Determination and inference of eukaryotic transcription factor sequence specificity. *Cell* **158**, 1431-1443, doi:10.1016/j.cell.2014.08.009 (2014).
65 O'Malley, R. C. *et al.* Cistrome and Epicistrome Features Shape the Regulatory DNA Landscape. *Cell* **165**, 1280-1292 (2016).
66 Vaneechoutte, D. & Vandepoele, K. Curse: building expression atlases and co-expression networks from public RNA-Seq data. *Bioinformatics* **35**, 2880-2881, doi:10.1093/bioinformatics/bty1052 (2019).
67 Patro, R., Mount, S. M. & Kingsford, C. Sailfish enables alignment-free isoform quantification from RNA-seq reads using lightweight algorithms. *Nat Biotechnol* **32**, 462-464, doi:10.1038/nbt.2862 (2014).
68 Anders, S. & Huber, W. Differential expression analysis for sequence count data. *Genome Biol* **11**, R106, doi:10.1186/gb-2010-11-10-r106 (2010).
69 Giorgi, F. M., Del Fabbro, C. & Licausi, F. Comparative study of RNA-seq- and microarray-derived coexpression networks in Arabidopsis thaliana. *Bioinformatics* **29**, 717-724, doi:10.1093/bioinformatics/btt053 (2013).

70 Frith, M. C., Li, M. C. & Weng, Z. Cluster-Buster: Finding dense clusters of motifs in DNA sequences. *Nucleic Acids Res* **31**, 3666-3668 (2003).
71 Jin, J., Zhang, H., Kong, L., Gao, G. & Luo, J. PlantTFDB 3.0: a portal for the functional and evolutionary study of plant transcription factors. *Nucleic Acids Res* **42**, D1182-1187, doi:10.1093/nar/gkt1016 (2014).
72 Brady, S. M. *et al.* A stele-enriched gene regulatory network in the Arabidopsis root. *Mol Syst Biol* **7**, 459, doi:10.1038/msb.2010.114 (2011).
73 Li, B. *et al.* Promoter-based integration in plant defense regulation. *Plant Physiol* **166**, 1803-1820, doi:10.1104/pp.114.248716 (2014).
74 Taylor-Teeples, M. *et al.* An Arabidopsis gene regulatory network for secondary cell wall synthesis. *Nature* **517**, 571-575 (2015).
75 Tian, C. *et al.* An organ boundary-enriched gene regulatory network uncovers regulatory hierarchies underlying axillary meristem initiation. *Mol Syst Biol* **10**, 755, doi:10.15252/msb.20145470 (2014).
76 De Bodt, S., Hollunder, J., Nelissen, H., Meulemeester, N. & Inze, D. CORNET 2.0: integrating plant coexpression, protein-protein interactions, regulatory interactions, gene associations and functional annotations. *New Phytol* **195**, 707-720, doi:10.1111/j.1469-8137.2012.04184.x (2012).
77 Nagel, D. H. *et al.* Genome-wide identification of CCA1 targets uncovers an expanded clock network in Arabidopsis. *Proc Natl Acad Sci U S A* **112**, E4802-4810, doi:10.1073/pnas.1513609112 (2015).
78 Verkest, A. *et al.* A generic tool for transcription factor target gene discovery in Arabidopsis cell suspension cultures based on tandem chromatin affinity purification. *Plant Physiol* **164**, 1122-1133, doi:10.1104/pp.113.229617 (2014).
79 Birkenbihl, R. P., Diezel, C. & Somssich, I. E. Arabidopsis WRKY33 is a key transcriptional regulator of hormonal and metabolic responses toward Botrytis cinerea infection. *Plant Physiol* **159**, 266-285, doi:10.1104/pp.111.192641 (2012).
80 Merelo, P. *et al.* Genome-wide identification of KANADI1 target genes. *PLoS One* **8**, e77341, doi:10.1371/journal.pone.0077341 (2013).
81 Qureshi, M. K., Sujeeth, N., Gechev, T. S. & Hille, J. The zinc finger protein ZAT11 modulates paraquat-induced programmed cell death in Arabidopsis thaliana. *Acta Physiologiae Plantarum* **35**, 1863–1871 (2013).
82 Liu, X. M. *et al.* ZAT11, a zinc finger transcription factor, is a negative regulator of nickel ion tolerance in Arabidopsis. *Plant Cell Rep* **33**, 2015-2021, doi:10.1007/s00299-014-1675-7 (2014).
83 Rizhsky, L., Davletova, S., Liang, H. & Mittler, R. The zinc finger protein Zat12 is required for cytosolic ascorbate peroxidase 1 expression during oxidative stress in Arabidopsis. *J Biol Chem* **279**, 11736-11743, doi:10.1074/jbc.M313350200 (2004).
84 Davletova, S., Schlauch, K., Coutu, J. & Mittler, R. The zinc-finger protein Zat12 plays a central role in reactive oxygen and abiotic stress signaling in Arabidopsis. *Plant Physiol* **139**, 847-856, doi:10.1104/pp.105.068254 (2005).
85 Scarpeci, T. E., Zanor, M. I., Mueller-Roeber, B. & Valle, E. M. Overexpression of AtWRKY30 enhances abiotic stress tolerance during early growth stages in Arabidopsis thaliana. *Plant Mol Biol* **83**, 265-277, doi:10.1007/s11103-013-0090-8 (2013).
86 Perez-Salamo, I. *et al.* The heat shock factor A4A confers salt tolerance and is regulated by oxidative stress and the mitogen-activated protein kinases MPK3 and MPK6. *Plant Physiol* **165**, 319-334, doi:10.1104/pp.114.237891 (2014).

87 Vanderauwera, S. *et al.* AtWRKY15 perturbation abolishes the mitochondrial stress response that steers osmotic stress tolerance in Arabidopsis. *Proc Natl Acad Sci U S A* **109**, 20113-20118, doi:10.1073/pnas.1217516109 (2012).
88 Mittler, R. *et al.* Gain- and loss-of-function mutations in Zat10 enhance the tolerance of plants to abiotic stress. *FEBS Lett* **580**, 6537-6542, doi:10.1016/j.febslet.2006.11.002 (2006).
89 Sewelam, N. *et al.* Ethylene response factor 6 is a regulator of reactive oxygen species signaling in Arabidopsis. *PLoS One* **8**, e70289, doi:10.1371/journal.pone.0070289 (2013).
90 Moffat, C. S. *et al.* ERF5 and ERF6 play redundant roles as positive regulators of JA/Et-mediated defense against Botrytis cinerea in Arabidopsis. *PLoS One* **7**, e35995, doi:10.1371/journal.pone.0035995 (2012).
91 Dubois, M. *et al.* Ethylene Response Factor6 acts as a central regulator of leaf growth under water-limiting conditions in Arabidopsis. *Plant Physiol* **162**, 319-332, doi:10.1104/pp.113.216341 (2013).
92 Mahmood, K., Xu, Z., El-Kereamy, A., Casaretto, J. A. & Rothstein, S. J. The Arabidopsis Transcription Factor ANAC032 Represses Anthocyanin Biosynthesis in Response to High Sucrose and Oxidative and Abiotic Stresses. *Front Plant Sci* **7**, 1548, doi:10.3389/fpls.2016.01548 (2016).
93 Allu, A. D., Brotman, Y., Xue, G. P. & Balazadeh, S. Transcription factor ANAC032 modulates JA/SA signalling in response to Pseudomonas syringae infection. *EMBO Rep* **17**, 1578-1589, doi:10.15252/embr.201642197 (2016).
94 Babitha, K. C. *et al.* Co-expression of AtbHLH17 and AtWRKY28 confers resistance to abiotic stress in Arabidopsis. *Transgenic Res* **22**, 327-341, doi:10.1007/s11248-012-9645-8 (2013).
95 Chen, X. *et al.* Overexpression of AtWRKY28 and AtWRKY75 in Arabidopsis enhances resistance to oxalic acid and Sclerotinia sclerotiorum. *Plant Cell Rep* **32**, 1589-1599, doi:10.1007/s00299-013-1469-3 (2013).
96 Wu, L., Zhong, G., Wang, J., Li, X. & Yang, Y. Functional analysis of an arabidopsis transcription factor WRKY28 in resistance to alternaria brassicicola. *Journal of Agricultural Science and Technology* **14**, 65-71 (2012).
97 De Clercq, I. *et al.* The membrane-bound NAC transcription factor ANAC013 functions in mitochondrial retrograde regulation of the oxidative stress response in Arabidopsis. *Plant Cell* **25**, 3472-3490, doi:10.1105/tpc.113.117168 (2013).
98 Kim, J. B., Kang, J. Y. & Kim, S. Y. Over-expression of a transcription factor regulating ABA-responsive gene expression confers multiple stress tolerance. *Plant Biotechnol J* **2**, 459-466, doi:10.1111/j.1467-7652.2004.00090.x (2004).
99 Van Aken, O. *et al.* Mitochondrial and Chloroplast Stress Responses Are Modulated in Distinct Touch and Chemical Inhibition Phases. *Plant Physiol* **171**, 2150-2165, doi:10.1104/pp.16.00273 (2016).
100 Ng, S. *et al.* A membrane-bound NAC transcription factor, ANAC017, mediates mitochondrial retrograde signaling in Arabidopsis. *Plant Cell* **25**, 3450-3471, doi:10.1105/tpc.113.113985 (2013).
101 Dombrecht, B. *et al.* MYC2 differentially modulates diverse jasmonate-dependent functions in Arabidopsis. *Plant Cell* **19**, 2225-2245, doi:10.1105/tpc.106.048017 (2007).
102 Verma, D., Jalmi, S. K., Bhagat, P. K., Verma, N. & Sinha, A. K. A bHLH transcription factor, MYC2, imparts salt intolerance by regulating proline biosynthesis in Arabidopsis. *FEBS J* **287**, 2560-2576, doi:10.1111/febs.15157 (2020).

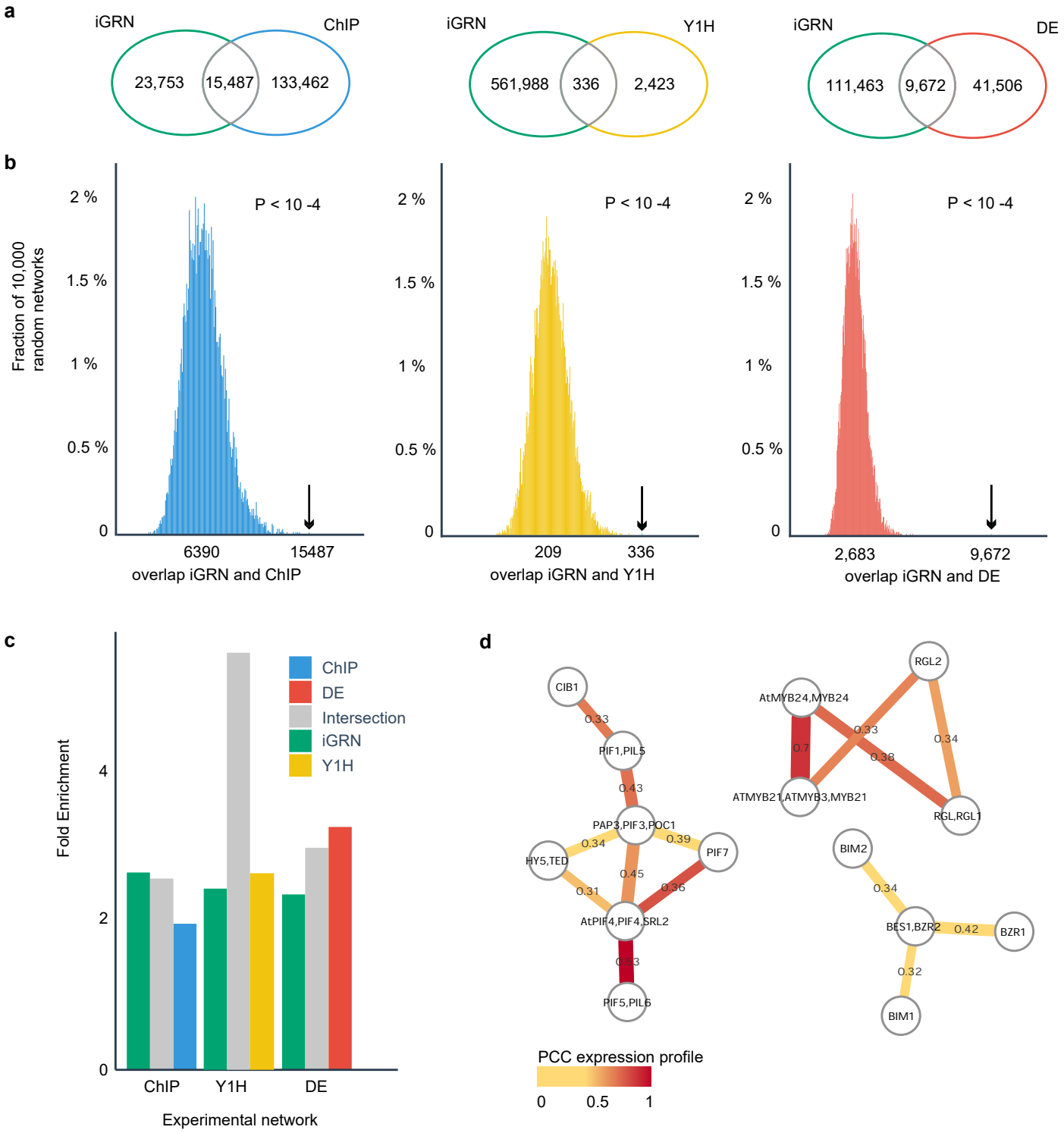
103 Song, S. *et al.* Interaction between MYC2 and ETHYLENE INSENSITIVE3 modulates antagonism between jasmonate and ethylene signaling in Arabidopsis. *Plant Cell* **26**, 263-279, doi:10.1105/tpc.113.120394 (2014).
104 Wu, A. *et al.* JUNGBRUNNEN1, a reactive oxygen species-responsive NAC transcription factor, regulates longevity in Arabidopsis. *Plant Cell* **24**, 482-506, doi:10.1105/tpc.111.090894 (2012).
105 Shahnejat-Bushehri, S., Nobmann, B., Devi Allu, A. & Balazadeh, S. JUB1 suppresses Pseudomonas syringae-induced defense responses through accumulation of DELLA proteins. *Plant Signal Behav* **11**, e1181245, doi:10.1080/15592324.2016.1181245 (2016).
106 Ebrahimian-Motlagh, S. *et al.* JUNGBRUNNEN1 Confers Drought Tolerance Downstream of the HD-Zip I Transcription Factor AtHB13. *Front Plant Sci* **8**, 2118, doi:10.3389/fpls.2017.02118 (2017).
107 Nishizawa, A. *et al.* Arabidopsis heat shock transcription factor A2 as a key regulator in response to several types of environmental stress. *Plant J* **48**, 535-547, doi:10.1111/j.1365-313X.2006.02889.x (2006).
108 Nishizawa-Yokoi, A., Yabuta, Y. & Shigeoka, S. The contribution of carbohydrates including raffinose family oligosaccharides and sugar alcohols to protection of plant cells from oxidative damage. *Plant Signal Behav* **3**, 1016-1018, doi:10.4161/psb.6738 (2008).
109 Zhang, L., Li, Y., Xing, D. & Gao, C. Characterization of mitochondrial dynamics and subcellular localization of ROS reveal that HsfA2 alleviates oxidative damage caused by heat stress in Arabidopsis. *J Exp Bot* **60**, 2073-2091, doi:10.1093/jxb/erp078 (2009).
110 Olate, E., Jimenez-Gomez, J. M., Holuigue, L. & Salinas, J. NPR1 mediates a novel regulatory pathway in cold acclimation by interacting with HSFA1 factors. *Nat Plants* **4**, 811-823, doi:10.1038/s41477-018-0254-2 (2018).
111 Enomoto, T. *et al.* STOP1 regulates the expression of HsfA2 and GDHs that are critical for low-oxygen tolerance in Arabidopsis. *J Exp Bot* **70**, 3297-3311, doi:10.1093/jxb/erz124 (2019).
112 Hsieh, T. H. *et al.* Heterology expression of the Arabidopsis C-repeat/dehydration response element binding factor 1 gene confers elevated tolerance to chilling and oxidative stresses in transgenic tomato. *Plant Physiol* **129**, 1086-1094, doi:10.1104/pp.003442 (2002).
113 Jaglo-Ottosen, K. R., Gilmour, S. J., Zarka, D. G., Schabenberger, O. & Thomashow, M. F. Arabidopsis CBF1 overexpression induces COR genes and enhances freezing tolerance. *Science* **280**, 104-106, doi:10.1126/science.280.5360.104 (1998).
114 Hsieh, T. H., Lee, J. T., Charng, Y. Y. & Chan, M. T. Tomato plants ectopically expressing Arabidopsis CBF1 show enhanced resistance to water deficit stress. *Plant Physiol* **130**, 618-626, doi:10.1104/pp.006783 (2002).

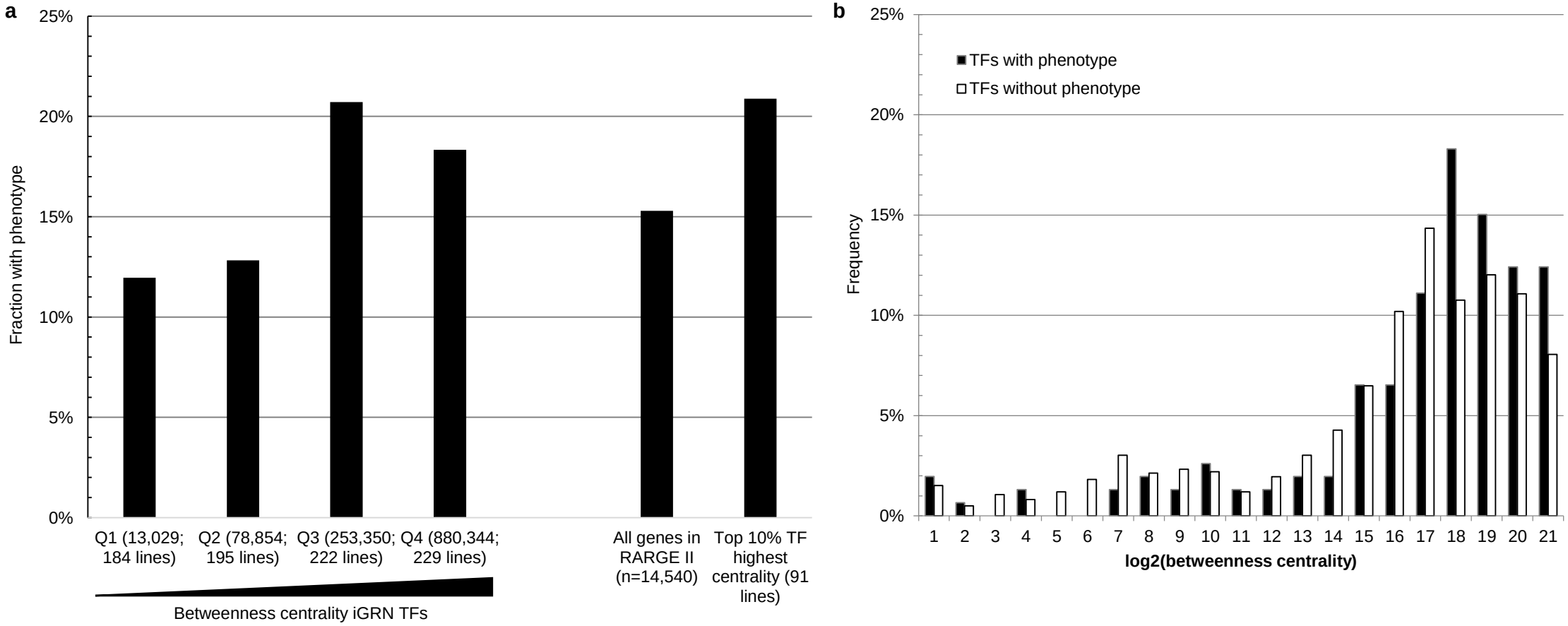
a

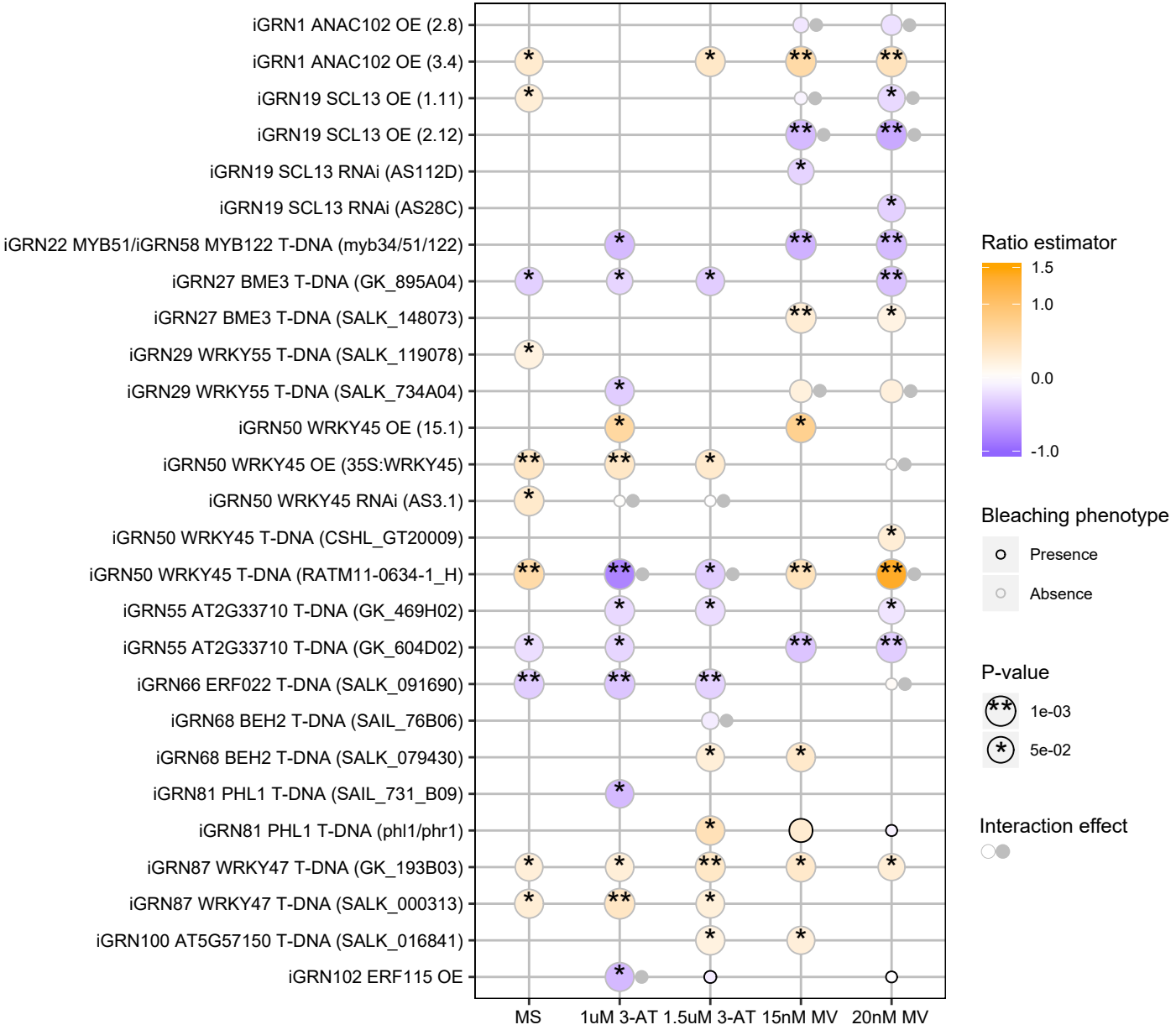


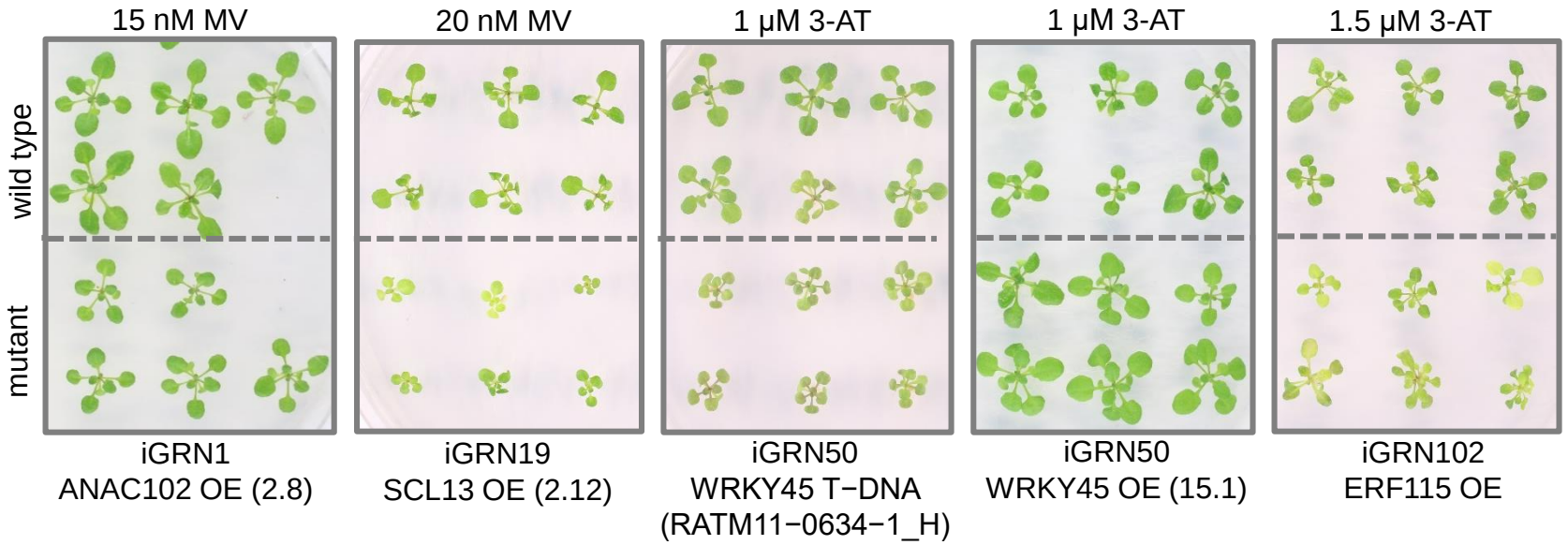
b

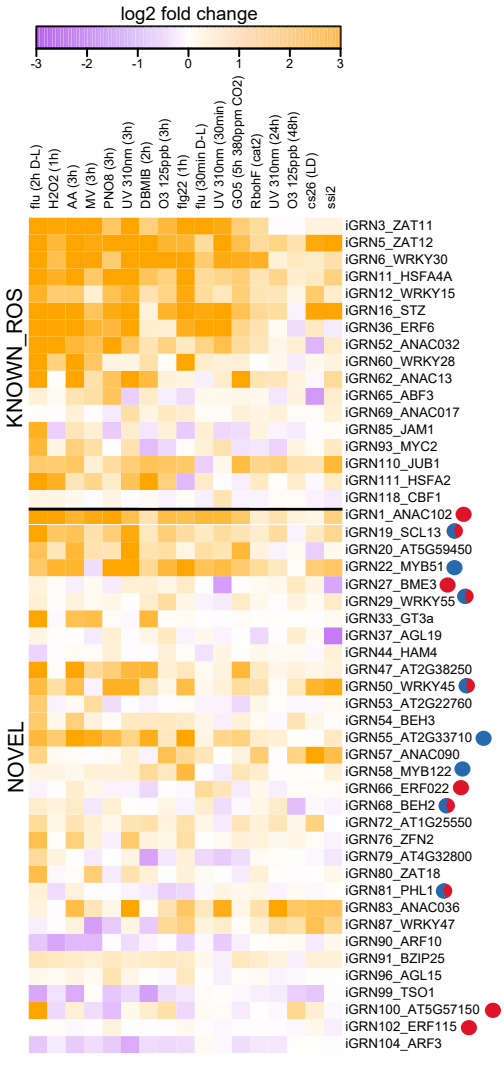
Approach	Number of interactions	Recall of true positives	Enrichment evalation set	Network contribution	Percentage of supervised
TF ChIP	46,619	29 %	6.0	80 %	2 %
PWM	17,053,785	60 %	1.0	7 %	70 %
Conserved PWM	340,893	11 %	6.4	91 %	17 %
PWM in DHS	5,231,258	34 %	1.5	17 %	49 %
CoE + PWM	324,864	3 %	2.2	18 %	3 %
CoE TF- target	1,547,002	10 %	3.0	40 %	36 %
GENIE3	2,588,312	17 %	1.7	32 %	48 %
Supervised Learning iGRN	1,709,871	46 %	8.8		


a
iGRN
ChIP
23,753
15,487
133,462
iGRN
Y1H
561,988
336
2,423
iGRN
DE
111,463
9,672
41,506
b
Fraction of 10,000 random networks
2 %
1.5 %
1 %
0.5 %
0
P < 10 -4
6390
15487
overlap iGRN and ChIP
209
336
overlap iGRN and Y1H
2,683
9,672
overlap iGRN and DE
c
Fold Enrichment
4
2
0
ChIP
DE
Intersection
iGRN
Y1H
ChIP
Y1H
DE
Experimental network
d
CIB1
PIF1,PIL5
PAP3,PIF3,POC1
HY5,TED
PIF7
AtPIF4,PIF4,SRL2
PIF5,PIL6
0.33
0.43
0.34
0.39
0.45
0.31
0.36
0.53
RGL2
AtMYB24,MYB24
ATMYB21,ATMYB3,MYB21
RGL,RGL1
0.33
0.34
0.7
0.38
BIM2
BES1,BZR2
BZR1
BIM1
0.34
0.42
0.32
PCC expression profile
0
0.5
1


a
25%
20%
15%
10%
5%
0%
Fraction with phenotype
Q1 (13,029; 184 lines)
Q2 (78,854; 195 lines)
Q3 (253,350; 222 lines)
Q4 (880,344; 229 lines)
All genes in RARGE II (n=14,540)
Top 10% TF highest centrality (91 lines)
Betweenness centrality iGRN TFs
b
TFs with phenotype
TFs without phenotype
Frequency
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21
log2(betweenness centrality)


iGRN1 ANAC102 OE (2.8)
iGRN1 ANAC102 OE (3.4)
iGRN19 SCL13 OE (1.11)
iGRN19 SCL13 OE (2.12)
iGRN19 SCL13 RNAi (AS112D)
iGRN19 SCL13 RNAi (AS28C)
iGRN22 MYB51/iGRN58 MYB122 T-DNA (myb34/51/122)
iGRN27 BME3 T-DNA (GK_895A04)
iGRN27 BME3 T-DNA (SALK_148073)
iGRN29 WRKY55 T-DNA (SALK_119078)
iGRN29 WRKY55 T-DNA (SALK_734A04)
iGRN50 WRKY45 OE (15.1)
iGRN50 WRKY45 OE (35S:WRKY45)
iGRN50 WRKY45 RNAi (AS3.1)
iGRN50 WRKY45 T-DNA (CSHL_GT20009)
iGRN50 WRKY45 T-DNA (RATM11-0634-1_H)
iGRN55 AT2G33710 T-DNA (GK_469H02)
iGRN55 AT2G33710 T-DNA (GK_604D02)
iGRN66 ERF022 T-DNA (SALK_091690)
iGRN68 BEH2 T-DNA (SAIL_76B06)
iGRN68 BEH2 T-DNA (SALK_079430)
iGRN81 PHL1 T-DNA (SAIL_731_B09)
iGRN81 PHL1 T-DNA (phl1/phr1)
iGRN87 WRKY47 T-DNA (GK_193B03)
iGRN87 WRKY47 T-DNA (SALK_000313)
iGRN100 AT5G57150 T-DNA (SALK_016841)
iGRN102 ERF115 OE
MS
1uM 3-AT
1.5uM 3-AT
15nM MV
20nM MV
Ratio estimator
1.5
1.0
0.0
-1.0
Bleaching phenotype
Presence
Absence
P-value
1e-03
5e-02
Interaction effect


15 nM MV
20 nM MV
1 µM 3-AT
1 µM 3-AT
1.5 µM 3-AT
wild type
mutant
iGRN1
ANAC102 OE (2.8)
iGRN19
SCL13 OE (2.12)
iGRN50
WRKY45 T−DNA
(RATM11−0634−1_H)
iGRN50
WRKY45 OE (15.1)
iGRN102
ERF115 OE


log2 fold change
-3
-2
-1
0
1
2
3
flu (2h D-L)
H2O2 (1h)
AA (3h)
MV (3h)
PNO8 (3h)
UV 310nm (3h)
DBMIB (2h)
O3 125ppb (3h)
flg22 (1h)
flu (30min D-L)
UV 310nm (30min)
GO5 (5h 380ppm CO2)
RbohF (cat2)
UV 310nm (24h)
O3 125ppb (48h)
cs26 (LD)
ssi2
KNOWN_ROS
NOVEL
iGRN3_ZAT11
iGRN5_ZAT12
iGRN6_WRKY30
iGRN11_HSFA4A
iGRN12_WRKY15
iGRN16_STZ
iGRN36_ERF6
iGRN52_ANAC032
iGRN60_WRKY28
iGRN62_ANAC13
iGRN65_ABF3
iGRN69_ANAC017
iGRN85_JAM1
iGRN93_MYC2
iGRN110_JUB1
iGRN111_HSFA2
iGRN118_CBF1
iGRN1_ANAC102
iGRN19_SCL13
iGRN20_AT5G59450
iGRN22_MYB51
iGRN27_BME3
iGRN29_WRKY55
iGRN33_GT3a
iGRN37_AGL19
iGRN44_HAM4
iGRN47_AT2G38250
iGRN50_WRKY45
iGRN53_AT2G22760
iGRN54_BEH3
iGRN55_AT2G33710
iGRN57_ANAC090
iGRN58_MYB122
iGRN66_ERF022
iGRN68_BEH2
iGRN72_AT1G25550
iGRN76_ZFN2
iGRN79_AT4G32800
iGRN80_ZAT18
iGRN81_PHL1
iGRN83_ANAC036
iGRN87_WRKY47
iGRN90_ARF10
iGRN91_BZIP25
iGRN96_AGL15
iGRN99_TSO1
iGRN100_AT5G57150
iGRN102_ERF115
iGRN104_ARF3