Integrative inference of transcriptional networks in Arabidopsis yields novel ROS signalling regulators

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], Michiel Van Bel[1,2], Robin Pottie[1,2], Dries Vaneechoutte[1,2,3], Frank Van Breusegem[1,2], Klaas Vandepoele[1,2,3*]

(1) Department of Plant Biotechnology and Bioinformatics, Ghent University, 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 authors: Klaas.Vandepoele@psb.vib-ugent.be and Inge.DeClercq@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 global understanding of regulatory interactions 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 to enhance our understanding of gene regulation in plants by integrating different experimental data types at the network level.

38 **INTRODUCTION**

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

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

66 In this paper we present a network-based approach based on supervised learning for large-
67 scale functional data integration to enhance our understanding of TF gene regulation in
68 *Arabidopsis*, hence referred to as an integrative GRN (iGRN). Supervised learning approaches
69 make efficient use of the available prior knowledge to train a machine learning algorithm, called

70 a classifier, for handling new data. Compared to unsupervised learning approaches, supervised
71 learning is considered more powerful when a large amount of existing knowledge is available to
72 guide the classification task, which here is the identification of functional TF-target
73 interactions[13]. Construction and validation of the iGRN, leveraging a diverse set of
74 complementary experimental datasets, revealed a strong enrichment for functional interactions
75 and a high predictive power to correctly infer functions for both functionally characterized and
76 uncharacterized TFs involved in a variety of biological processes. The latter is demonstrated in a
77 case study by the experimental validation of novel iGRN-based function predictions for TFs
78 involved in plant responses to oxidative stresses.
79

80 **RESULTS**

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

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

101 from 27 TFs [7]. Next to a physical TF binding network, two co-expression-based regulatory
102 networks were also generated. Co-expression approaches measure similarity at the output level
103 of transcriptional regulation and as such can provide a complementary view compared to
104 promoter-centered regulatory networks. First, the Pearson correlation between a TF and a
105 target gene was used to build co-expression clusters around each TF, resulting in a fifth "CoE TF-
106 target" input network. Secondly, GENIE3, a reverse engineering tool for reconstructing
107 regulatory networks based on expression data, was used to build a sixth input network [16].
108 GENIE3 was among the best performing tools in reconstructing regulatory interactions in a
109 large-scale benchmark study [17]. The last network was generated using a hybrid approach of co-
110 expression and PWM based methods ("COE+PWM"). Initially a co-expression cluster was
111 constructed for each gene and PWM enrichment was subsequently performed. Only genes with
112 significantly enriched PWMs in their co-expression cluster were retained, resulting in a seventh
113 network where TF binding site information is combined with co-regulatory information (see
114 Methods).
115
116 As the different input networks capture different aspects of gene regulation, we next applied
117 supervised learning to construct, starting from these input networks, an integrative regulatory
118 network reporting functional interactions. A score for each possible TF-target interaction was
119 devised as a function of the feature-specific weights provided by the seven input networks. For
120 building a classifier, we compiled a gold standard dataset comprising interactions based on all
121 validated interactions in AtRegNet [18], an exhaustive literature-based network reporting
122 interactions in secondary cell wall development [19], and a large-scale data mining effort on 974
123 peer reviewed articles [20]. This gold standard entails interactions for 47 TF families involved in
124 different biological processes, with on average 11 interactions per TF (Table 1 and Supplemental
125 Table S1). Eighty percent of the interactions present in this gold standard were retained as
126 training dataset for the machine learning classification algorithms while the remaining twenty
127 percent was used as an evaluation set. Initially different classifiers were learned and their
128 performance to identify functional interactions in the evaluation set was assessed
129 (Supplemental Table S2). The classifier based on gradient boosting model using decision trees
130 showed the best performance, with an area under the curve (AUC) value of 0.77, and Precision
131 and Recall of 71% and 46% on the evaluation set, respectively (Supplemental Table S2). Applying
132 this classifier to the complete set of interactions present in the seven input networks resulted in

133 an integrative network of 1,709,871 interactions for 1,491 TFs, containing 31,393 target genes
134 (29,137 protein-coding and 2,256 non-coding genes; see Data file 1).
135 In order to obtain an unbiased performance estimate of the input networks compared to the
136 iGRN, all input networks were compared to the evaluation dataset of known interactions (20%
137 of the gold standard) and benchmarked measuring the recovery of true edges (referred to as the
138 recall, Figure 1b). Furthermore, we also determined the enrichment of a network towards the
139 evaluation dataset, which functions as a proxy for specificity [21]. Focusing on the different input
140 networks, Figure 1b shows that the network based on evolutionary conserved TF binding sites
141 ("Conserved PWM") has the highest enrichment (6.4 fold, Supplemental Table 3). Evolutionary
142 conservation is known to be a strong predictor of functional regulatory elements explaining this
143 high enrichment [22,23]. The TF ChIP-seq network also performs well with both a high recovery
144 (29%) and enrichment (6.0 fold). This overlap can partially be explained by the fact that a subset
145 of the validated AtRegNet interactions in the gold standard are based on ChIP studies [7]. Other TF
146 motif mapping strategies recover larger fractions of true edges but suffer from low specificity,
147 resulting in low enrichment values. The simple mapping of TF motifs ("PWM") is known to result
148 in many false positives because TF binding sites are often short and typically contain some level
149 of degeneracy [11,24]. Both co-expression based networks do not consider TF binding site
150 information and are therefore able to also identify regulatory interactions for TFs lacking DNA
151 binding motif information. Both these networks perform relatively well: correlation-based co-
152 expression between TF and target genes, implemented in the "CoE TF-target" network, has the
153 third highest enrichment towards the evaluation dataset (10% recovery and 3.0 fold
154 enrichment), while GENIE3 displays a good balance between recovery of true edges and
155 specificity (Supplemental Table S3). Focusing on the integrative network (iGRN) shows it has the
156 strongest enrichment (8.8 fold) as well as a high recovery for true regulatory relationships (46%
157 recovery of interactions in the evaluation set), clearly outperforming the different
158 computational and experimental approaches used to construct the input networks.
159 Apart from determining the power of the iGRN network to recover known functional
160 interactions, we also measured the contribution of each of the seven input networks to the
161 iGRN. Ideally, a good integration approach should favor interactions from the input networks
162 with high enrichment values towards the evaluation set, as these would boost the recovery of
163 correct interactions. As reported in Figure 1b (brown and red bar charts), the evolutionary
164 conserved motifs ("Conserved PWM") and the "TF ChIP" network contributed the largest part of

165 their edges to the integrative network (91% and 80%, respectively), while networks showing low
166 specificity, such as "PWM", only contributed a small fraction of interactions (7%). This result
167 corroborates that the applied supervised approach successfully identified functional interactions
168 coming from high-quality input networks. Next, we measured which iGRN interactions are
169 supported by the different input networks (Figure 1b, red bar plot). Overall, the integrative
170 network has 70% of the edges supported by a PWM mapping, meaning that most regulatory
171 interactions are supported by the presence of TF binding site(s). Furthermore, 36-48% of the
172 interactions are supported by at least one of the two co-expression networks, indicating for
173 nearly half of all interactions a regulatory effect at the transcript level. Overall, 92% of the iGRN
174 interactions are supported by more than one input network (Figure S1), confirming both the
175 commonality and complementarity of the input networks.
176

177 **iGRN interactions are complementary to TF regulatory events identified from**
178 **experimental methods**

179 To assess whether the iGRN identifies functional TF-target gene interactions that are
180 supported by experimental data, detailed comparisons of the iGRN interactions were made
181 against three different datasets, capturing a variety of TFs and biological processes, which were
182 not used during the iGRN construction (see Methods). These datasets cover i) a set of 2,759 Y1H
183 interactions reflecting direct *in vitro* interactions ii) 28 TF ChIP-seq experiments collected from
184 different studies (148,949 interactions, see also Methods) reflecting primarily direct *in vivo*
185 binding and iii) a set of 40 differentially expressed (DE) gene sets upon genetic TF perturbation
186 (51,178 interactions, see Methods). The significance of the observed overlaps was examined
187 through comparison to a background distribution of expected overlaps generated by
188 randomizing the iGRN (see Methods). First, we detected a weak but significant overlap between
189 edges in the integrative network and regulatory edges determined by Y1H (Figure 2a). An
190 overlap of 336 edges (12%) was observed reflecting a 1.6 higher overlap than expected by
191 chance (Figure 2b). TF ChIP-seq interactions displayed an overlap of 15,487 interactions (10%)
192 with a fold enrichment of 2.4. Only considering TFs that occur in both the iGRN and this ChIP-
193 Seq dataset, for 25/28 regulators there was a significant overlap between the iGRN and ChIP
194 defined target genes (Supplemental Table S4). Overall, we found that TF ChIP-seq data
195 confirmed 39% of predicted interactions in the iGRN, demonstrating that a large fraction of iGRN

196 interactions are supported by *in vivo* TF binding. This fraction is comparable with the
197 concordance between TF ChIP-based regulatory interactions and altered target gene transcript
198 levels after genetic TF perturbation, an approach that is frequently used to identify bound target
199 genes that are also regulated [25]. Next to TF-DNA binding assays, we also investigated the
200 recovery of directly and indirectly regulated target genes based on differential target gene
201 expression. The iGRN recovered 9,672 differentially expressed interactions (19%) resulting in a
202 3.6 fold enrichment (Figure 2a and 2b). Taken together, both the comparison with binding and
203 regulatory datasets demonstrate that the iGRN represents high-quality interactions supported
204 by experimental data.
205
206 To evaluate the relevance of the regulatory interactions uniquely predicted by the iGRN and
207 hence not overlapping with the three experimental datasets (Y1H, TF ChIP-Seq, and DE), a
208 functional metric was used to measure whether a TF's target genes participate in similar
209 biological processes. This metric starts from the assumption that a set of bona fide target genes
210 will have a higher enrichment for functional annotations than randomized networks [21]. For each
211 TF, the enriched Gene Ontology (GO) Biological Process terms based on its target genes were
212 determined, summarized in a distribution and compared to the background distribution based
213 on randomized networks. In Figure 2c, the iGRN-predicted interactions were first divided in
214 different sets based on their overlap with the TF ChIP, Y1H and DE experimental datasets (as in
215 Figure 2a), respectively. Next, for each subset the fold enrichments for functional GO Biological
216 Process terms were assessed. Whereas the grey series in Figure 2c show the regulatory
217 interactions shared between the iGRN and the experimental dataset, the green series shows the
218 interactions unique to the iGRN. Likewise, the blue, yellow and red series show the regulatory
219 interactions present in the ChIP, Y1H and DE dataset not overlapping with the iGRN.
220 Focusing on the overlap with TF ChIP dataset revealed that the functional enrichment was
221 similar for the interactions unique to the iGRN compared to those shared between the iGRN and
222 TF ChIP dataset (2.63 and 2.55, respectively). However, the regulatory interactions unique to the
223 TF ChIP datasets showed lower fold enrichments (1.94), indicating that these show less
224 functional coherence. For the Y1H comparison, the interactions uniquely predicted by the iGRN
225 had a similar fold enrichment as the interactions only reported by the Y1H assays (2.41 and
226 2.61). The interactions that are reported by both approaches had a considerable higher
227 enrichment compared to the other networks (5.55; though note the small set size, n=336). The

228 comparison with the DE dataset showed a similar result with an increase in enrichment fold
229 from uniquely predicted edges in the iGRN to the intersection and uniquely predicted edges by
230 the DE gene sets (2.33, 2.96 and 3.23, respectively). Taken together, these results indicate that
231 the predicted interactions, including those unique to the iGRN, convey functional gene
232 regulatory events.
233

234 **Network architecture reveals different levels of gene regulatory complexity affecting**
235 **plant phenotypes**

236 Starting from the complete iGRN comprising 1.7 million interactions, we investigated the
237 structure and complexity of the network and determined if network architecture can be linked
238 to specific biological functions[26]. On average, each gene in the network has 54 incoming TFs
239 (median 32), while each TF is linked to 1,147 target genes (median 802; Supplemental Figure S2).
240 There was a large range in connectivity with half of the genes being regulated by 8-79 TFs
241 (interquartile range), and a small number of TFs (107/1491, or 7%) regulating more than 3,000
242 target genes. Focusing on the top 1% genes with the largest number of incoming TFs revealed a
243 significant bias towards genes with transcription factor activity, corroborating previous findings
244 reporting that regulators themselves are complexly regulated by other regulators [7,27]. In
245 contrast, 2172 genes having only one incoming TF showed enrichment for RNA binding and
246 translation, suggesting that these functions are less complexly regulated by TFs.
247 Next, we examined if network structure can offer new insights in the role of individual
248 regulators. Concretely, we investigated whether the degree (number of target genes) and
249 betweenness centrality (a measure for the importance based on number of shortest paths in the
250 network graph) of TFs in the iGRN can be predictive for potential master regulators showing
251 prominent phenotypes upon genetic perturbation. Whereas degree refers to the number of
252 interactions a node, or gene, has in the network, betweenness centrality measures how
253 important a node is to the shortest paths through the network. Nodes with a high centrality are
254 assumed to be more important in the network as more information is passing through these
255 nodes. After processing large-scale phenotyping data from RARGE II, an integrated phenotype
256 database of Arabidopsis mutants scoring traits using a controlled vocabulary for 14,540 genes,
257 phenotypes were recorded for 2,224 genes [28]. When splitting all TFs present in the iGRN and
258 represented in the RARGE II dataset in four quartiles (Q1-Q4) based on their degree values, we

259 observed that 17.4% of the high-degree (Q3+Q4) and 15.1% of the low-degree (Q1+Q2) TFs had
260 a reported phenotype (p-value 0.43; two-tailed Mann-Whitney U test). When comparing the
261 quartiles of TFs based on betweenness centrality, a significant difference was found when
262 scoring regulators with a reported phenotype (19.6% for Q3+Q4 and 12.4% for Q1+Q2,
263 respectively; p-value 0.0088; two-tailed Mann-Whitney U test; Figure 3a). This pattern is
264 confirmed when comparing the betweenness centrality for TFs with and without phenotypes, as
265 regulators showing RARGE II phenotypes have more of the shortest network paths going
266 through their nodes (log2 median value of 17.50 and 16.46 for TFs with and without reported
267 phenotypes, respectively; p-value 1e-05; two-tailed Mann-Whitney U test; Figure 3b). iGRN TFs
268 within the top 10% genes with highest centrality for which mutants have reported phenotypes
269 (19/91) are significantly biased towards post-embryonic development and include ATGTL1, PIF4,
270 SEP3, IDD4 and IDD8, ANAC056, DOF5.6, LFY3 and TCP4. After assigning regulators to three
271 levels based on their network hierarchy [29], we observed that top-level TFs high in the hierarchy
272 (n=313) have higher degree and centrality values compared to bottom-level TFs (n=704, median
273 degree for top and bottom TFs is 1923 and 529, respectively, while log2 median betweenness
274 centrality for top and bottom TFs is 18.87 and 16.10, respectively). Top-level TFs were not
275 significantly enriched for specific TF families. Interestingly, both for TFs with and without
276 phenotypes, a similar number of TFs in their target genes was found (9.45% and 9.44%,
277 respectively), indicating that TFs showing a phenotype do not regulate more TFs and thus are
278 not higher in the TF hierarchy.
279
280 As transcriptional regulation is mediated through interactions between TFs, co-activators or
281 co-repressors, and the general transcriptional machinery, we determined whether the iGRN has
282 predictive power to identify cooperative TF regulation. Therefore, we measured if regulators
283 sharing a large number of target genes are more likely to interact with each other, based on
284 experimentally determined protein-protein interactions (PPIs). We found that TF pairs sharing
285 many target genes (Jaccard index >0.3) have a 15-fold higher occurrence of PPIs compared to TF
286 pairs having less shared target genes (Jaccard index <0.3), which is highly significant (2.71%
287 versus 0.31%; p-value <2.2e-16, hypergeometric distribution). Whereas most of these PPIs cover
288 TFs from the same gene family, probably due to the high similarity of their corresponding
289 binding sites yielding similar sets of target genes, their potential to act as heterodimers and/or
290 potential functional redundancy, also cases of interacting TF pairs from different gene families

291 were found. Examples of the latter case include HY5-PIF3/PIF4 involved in far-red signaling,
292 MYB21-RGL1/RGL2 controlling hormone-mediated cell proliferation and differentiation, and
293 BES1-BIM1/BIM2 acting synergistically to control brassinosteroid signaling (Supplemental Figure
294 S3). The highly correlated expression profiles of these TFs sharing many target genes further
295 suggest they exert their regulatory roles in a cooperative manner.
296

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

298 Based on the validation results indicating the functional relevance of regulatory edges in the
299 iGRN, we also investigated if individual TF functions can be inferred. Therefore, we performed
300 GO enrichment analysis for all target genes of each TF, only considering experimental Biological
301 Process annotations and excluding very general GO terms (see Methods). Next, we determined
302 if the enriched Biological Process terms were already known for that specific TF or whether a
303 novel functionality was predicted (Data file 2. TF_function_all). In total we could infer functions
304 for 1,301 TFs and for 54% (681) of the 1,259 TFs with experimental Biological Process
305 annotations we could correctly infer one or more of the known functions. For 268 TFs lacking
306 experimental annotations, novel iGRN-based functions were predicted. Reciprocally, per
307 Biological Process term, on average 21% of the known regulators were recovered, while for
308 several distinct processes the recovery was 50% or higher. Examples of processes with high
309 recovery rates include response to wounding (14/20, 70%), plant-type secondary cell wall
310 biogenesis (18/28, 64%), response to hormone (174/343, 51%), chloroplast organization (6/12,
311 50%) and DNA replication (3/6, 50%).
312 The overall recovery of known TF functions based on the iGRN was lower compared to the
313 probabilistic functional gene network AraNet v2 [30], which recovered known functions for 739 TFs
314 (28% average recovery of known regulators per GO BP term). However, both approaches are
315 highly complementary, as 207 and 265 TFs had known functions only recovered using iGRN and
316 AraNet, respectively (Supplemental Figure S4). Examples of known TF functions uniquely
317 recovered through the iGRN cover EIN3 (ethylene-activated signaling pathway), APETALA2
318 (flower development), LFY (flower development, gibberellic acid-mediated signaling pathway),
319 FUS3 (somatic embryogenesis), TINY (multidimensional cell growth), and AMS, ATWRKY34 and
320 CRF12 (pollen development). Focusing on specific GO terms, biological processes with a higher
321 recovery of known regulators in the iGRN compared to AraNet cover response to biotic stimulus

322 (47% versus 35%), chloroplast organization (50% versus 0%), response to abiotic stimulus (47%
323 versus 26%), pollination (33% versus 0%), and root hair cell differentiation (35% versus 6%). For
324 the latter process we correctly identified the regulators ZPF5, RSL2 and RSL4, RHD6, LRL2 and
325 LRL3, as well as MYB36. MYB36 regulates the developmental transition in the root endodermis,
326 where the differentiated endodermis forms a protective waxy barrier called the Casparian strip.
327 In our network, MYB36 is predicted to regulate genes involved in cell wall organization such as
328 CASP1-5, PER64 and ESB1, which are known to be involved in Casparian strip formation and
329 have been shown to be differentially expressed in the *myb36* mutant [31]. Furthermore, the iGRN
330 also predicts MYB36 to be involved in root hair elongation, which agrees with the longer root
331 hairs observed in *myb36* mutants [31].
332

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

334 Apart from recovering known functions of TFs involved in development and stress responses,
335 the predictive power of the iGRN was used to identify novel TF functions. As a test case, we
336 extracted TFs from the iGRN with predicted oxidative stress signaling and/or responsiveness
337 functions. As an increase in ROS levels is a common theme during abiotic and biotic stresses,
338 oxidative stress responsiveness was used to validate novel regulators of plant stress responses.
339 One hundred twenty-four ROS-related TFs were extracted from the iGRN by identifying those
340 TFs that entailed an enrichment for core ROS responsive transcripts (ROS marker genes) in their
341 target genes [32]. Among these TFs, 17 had an already reported function in plant responses to ROS
342 or ROS-causing agents (hydrogen peroxide [H_2O_2], methyl viologen [MV] and 3-amino triazole [3-
343 AT]), here referred to as 'KNOWN_ROS' TFs (Table 2, Supplemental Table S5). Fifty-eight ROS TFs
344 were previously reported to function in plant responses to environmental stresses, such as
345 salinity, drought, heavy metals, cold, hypoxia, high light and infection by bacterial and fungal
346 pathogens (and that we refer to as 'KNOWN_STRESS' TFs, Supplemental Table S5). The
347 remaining 49 identified ROS TFs had previously not been linked to any role in
348 ROS/environmental stress (referred to as 'NOVEL' TFs in Table 2), at the time the iGRN was
349 constructed (see Methods). However, five of the 49 novel TFs were validated for a role in
350 environmental stress responses by studies that were published after the iGRN construction
351 (Table 2 and Supplemental Table S5): ANAC102 and BME3 (high light [33-35]), NAC103 (genotoxic
352 [36]), ZAT18 (drought [37]) and ERF019 (drought, bacterial and fungal pathogens [38,39]).

353 For experimental validation, we selected the top 32 novel TFs, ranked based on enrichment
354 for ROS marker target genes[32], for which at least one *Arabidopsis* loss- or gain-of-function
355 transgenic line was available (Supplemental Table S6). Plant performance under oxidative stress
356 was assessed based on rosette growth of the mutant lines compared to the wild type when
357 grown at low and high MV and 3-AT concentrations, respectively, using automated image
358 analysis (see Methods), and, in addition, chlorosis symptoms were visually scored.
359
360 For 13 of the 32 phenotyped novel ROS TFs we found at least one mutant allele with
361 significant oxidative stress-induced changes in rosette area relative to the wild type (i.e.
362 genotype effect dependent on the oxidative stress treatment and/or genotype-by-treatment
363 effect under at least one of the tested stress treatments; Figure 4; Table 2). For the other novel
364 TFs, no significant changes were detected by the genotype under all tested conditions, either
365 there was an overall genotype effect under all tested conditions (similar changes were
366 observed) or the genotype effect was independent of the stress treatments (e.g. WRKY47;
367 Figure 4). For SCL13, WRKY45, BEH2 and PHL1, experimental evidence is provided by at least
368 two mutant alleles. *WRKY45* loss-of-function (T-DNA mutant RATM11−0634−1_H and RNAi line
369 AS3.1) affected the rosette growth during low and high 3-AT stress compared to the wild type in
370 a genotype-by-treatment dependent matter, whereas, *WRKY45* OE (line 15.1) caused an
371 increased rosette yield that was dependent on stress treatments low 3-AT and low MV (Figure
372 5). Two *SCL13* overexpression alleles (1.11 and 2.12) displayed decreased rosette growth
373 dependent on both MV conditions in a genotype-by-treatment-dependent manner (Figure 5).
374 Similarly, knockdown of *SCL13* (RNAi lines AS112D and AS28C) negatively affected rosette
375 growth under MV stress, but no genotype-by-treatment interaction effect was observed. *PHL1*
376 T-DNA and *PHL1/PHR1* double T-DNA lines displayed altered growth on low and high 3-AT,
377 respectively, the latter dependent on the genotype-by-treatment. *BEH2* mutation altered the
378 rosette area that was dependent on the stress treatments high 3-AT and low MV for line
379 SALK_079430 and dependent on the genotype-by-treatment for line SAIL_76B06.
380
381 We further demonstrated that ANAC102, BME3, MYB51, WRKY55, AT2G33710, ERF022,
382 AT5G57150 and ERF115 have a function in oxidative stress responses based on one mutant
383 allele (for *ERF022* and *ERF115* only one line was available) (Figure 4 and Table 2). For example,
384 *ANAC102* OE line 3.4 (strong OE; ˜ 100-fold) displayed a larger rosette area that was

385 independent of the stress treatment, while rosette area changes by *ANAC102* OE2.8 (weak OE;
386 ~4.5-fold) were dependent on the genotype-treatment interaction (Figure 5). *ERF115* OE
387 affected plant performance in a manner that was dependent on the stress treatment, with a
388 decreased rosette area under low 3-AT (dependent on the genotype-by-stress treatment) and
389 severe bleaching under high MV and 3-AT concentrations (Figure 5).
390 Apart from validating the identified novel ROS TFs based on an oxidative stress-related growth
391 phenotype, we evaluated the correct prediction of ROS responsive target genes. Therefore, we
392 compared the predicted target genes of the highest ranked novel ROS TF ANAC102 with
393 genome-wide target genes identified by ChIP-Seq[27]. Among the 543 ANAC102 target genes
394 predicted by the iGRN, 185 (34%) were confirmed by ChIP-Seq, indicating a significant overlap
395 (2.28 fold enrichment, p-value<2e-26). Meta-analysis of transcript profiles during various
396 oxidative stress conditions shows that 103 of the 185 confirmed target genes are responsive to a
397 wide range of ROS-inducing conditions[32] (Supplemental Table S8, Supplemental Figure S5) and
398 include 20 out of the 35 ROS marker genes predicted to be ANAC102 targets. Among the 543
399 predicted ANAC102 target genes are also 62 TFs of which 28 are confirmed by ChIP-Seq. Among
400 these ChIP-confirmed TFs are the ROS marker genes *WRKY40, ZAT6, ZAT10, ZAT12* and *ERF5*,
401 indicating extensive TF-TF regulation of ROS responses downstream of ANAC102. Furthermore,
402 to be a true regulatory interaction, the spatial-temporal expression characteristics of the TF and
403 its target genes must be similar. To assess this, we explored transcriptome profiles of different
404 rosette and root tissues from Genevestivator[40] to compare tissue-specific expression profiles of
405 *ANAC102* with its ChIP-Seq-confirmed target genes. We observed common expression
406 signatures, namely *ANAC102* and its target genes being expressed in juvenile, adult and
407 senescent leaves as well as in different primary root zones and in the lateral root (Supplemental
408 Figure S5). Taken together, using ANAC102 as a test case, we confirmed the prediction of true
409 target genes, and consistency in oxidative stress responsiveness and spatial-temporal (basal)
410 expression profiles between *ANAC102* and its targets.
411 Finally, the expression profiles of KNOWN_ROS and validated novel ROS TFs during different
412 oxidative stress conditions were analyzed and compared [32]. Fourteen out of the 17 known ROS
413 TFs were identified as responsive to a wide range of oxidative stress conditions ($|\log_2$ fold
414 change$| > 1$ and FDR < 0.01; Figure 6; Supplemental Table S7). Among the 13 newly validated
415 novel ROS TFs, only half were ROS responsive at the transcript level. *WRKY55*, *MYB122* (for
416 which the triple mutant *myb34/51/122* shows a phenotype, but not the *myb122* single mutant),

417 *ERF022*, *BEH2*, *PHL1*, AT5G57150 and *ERF115* expression levels are not regulated by oxidative
418 stress, or only regulated in a small number or specific set of oxidative stress-related conditions,
419 and thus, would not have been predicted from studies solely relying on differential expression at
420 the whole plant or organ level. For example, *ERF115* has tissue-specific expression patterns,
421 confined to the stem cell niche within the root meristem [41], indicating that transcript analysis on
422 the whole plant level might not reveal stress-induced alterations in this specific tissue [42].
423 Interestingly, we observed that novel ROS TFs showing a phenotype under oxidative stress have
424 a 2-fold higher degree and 5-fold higher betweenness centrality compared to novel ROS TFs
425 lacking such phenotype (median degree of 1670 and 749, median betweenness centrality of
426 699,318 and 142,777, for novel ROS TFs with and without oxidative stress phenotype,
427 respectively).
428
429 Overall, among the 124 novel ROS TFs extracted from the iGRN, 75 (60%) were validated by
430 previously published studies (before the iGRN was constructed) to be involved in the plant's
431 responses to ROS perturbation and/or (a)biotic stresses. By analyzing mutant lines perturbed in
432 one of the analyzed 32 novel ROS TFs, 13 additional regulators were validated for a function in
433 ROS responses in our phenotyping assays. Among the novel ROS TFs newly identified by our
434 phenotyping assays, WRKY55, WRKY47, AT2G33710 [ERF] and AT5G57150 [bHLH] had previously
435 not been associated with any function. Other validated novel ROS TFs are known to function in
436 high-light stress tolerance or acclimation responses (NAC102, BME3 [GATA family] [33-35]), age-
437 triggered senescence (WRKY45, ERF022 [AP2/EREBP]; [43,44]), response to phosphate starvation
438 (WRKY45, PHL1 [G2 family]; [45,46]), submergence stress (WRKY45; [47]), glucosinolate (antimicrobial
439 compounds involved in plant immunity) biosynthesis (MYB51, MYB122; [48]), brassinosteroid
440 signaling (BEH2 [BES1 TF family]; [49]), phytochrome-dependent light signaling (SCL13 [GRAS
441 family]; [50]), hypocotyl growth (BEH2 [BES1 family]; [51]), seed germination (BME3; [52]) and somatic
442 embryogenesis (ERF022; [53]). Overall, among the novel ROS TFs validated for an oxidative stress
443 function by literature evidence or in our phenotyping assays, the AP2-EREBP (5), NAC (5) and
444 WRKY (5) families are highly represented (Supplemental Table S5). These families are typically
445 associated with abiotic, biotic and/or oxidative stress responses [54] and the iGRN identified
446 additional members with an oxidative stress phenotype. In addition, the iGRN correctly
447 predicted an oxidative stress response function for one or more members of the bHLH, ZAT (or
448 C2H2), HSF, MYB and bZIP families that are also often associated with functions in plant stress

449 responses. The iGRN also identified novel ROS TFs from families that were known to be mainly
450 involved in plant growth/developmental processes, such as GRAS (SCL13), BES1 (BEH2) and
451 GATA (BME3) ([49,55,56]) . So far there is very little evidence for the involvement of GATA TFs in
452 plant stress responses except for reported functions for rice and soybean GATA TFs in low
453 nitrogen stress tolerance [57,58], the identification of salt and drought stress-induced alternative
454 splice products for rice GATA TFs [59], and the recent characterization of BME3 in systemic
455 acquired acclimation to high light stress, downstream of a systemic signaling cascade mediated
456 by ROS signaling [34,35]. Consistent with recently published studies [34,35], we show that BME3 is one
457 of the first GATA transcription factors with a proven function in the regulation of stress
458 responses.
459

460 **DISCUSSION**

461 The comprehensive experimental mapping of plant GRNs is a major challenge and different
462 technologies, either *in vitro* or *in vivo*, each provide partial snapshots of the regulatory logic in
463 *Arabidopsis*. While different integrative approaches have been applied to identify regulatory
464 networks controlling specific biological processes in plants [12,60-63], in this study we leveraged a
465 wide variety of regulatory datasets capturing different properties of TF-mediated transcriptional
466 control. Through the application of an advanced machine learning approach for large-scale
467 functional data integration, we performed for the first time a systematic regulatory annotation
468 for nearly all *Arabidopsis* genes. Although the presented iGRN is generic and not tailored
469 towards specific plant organs or growth stages, extensive validation using different types of
470 experimental datasets, measuring TF binding and/or regulation, revealed a strong enrichment
471 for functional interactions. In addition, the iGRN outperformed all input networks when
472 assessing the predictive power using a gold standard evaluation set, demonstrating the success
473 of the supervised learning approach. An assessment of the functional coherence of a TF's target
474 genes suggests that the iGRN regulatory interactions are of equally high quality compared to
475 those obtained through experimental methods. Taken together, our validation results indicate
476 that the iGRN is strongly enriched for functional interactions and that the applied integrative
477 approach clearly complements GRNs inferred through a single experimental assay.
478 Investigating the relationship between network architecture and the complexity of TF control
479 confirmed that TFs themselves are complexly regulated by other TFs. In addition, through the

480 integration of large-scale phenotyping data, a significant association was found between a TF's
481 network centrality (but not degree) and observing a phenotype when knocking out that TF. A
482 similar pattern was observed for the novel ROS TFs showing an oxidative stress phenotype. This
483 finding indicates that these TFs occupy key positions in the GRN and loss of their gene products
484 has major consequences for plant growth and development. Furthermore, TFs showing a
485 phenotype when screening mutant lines were not having more TFs among their target genes
486 compared to TFs lacking phenotypes, suggesting that master regulators high in the TF network
487 hierarchy are not more frequently showing phenotypes compared to lower-tier TFs. Obviously,
488 these findings depend on the global quality of the iGRN but complement previous results
489 reporting that higher-tier TFs in ABA networks do not show more differential expression during
490 ABA response [27]. Finally, TF pairs sharing many target genes are indicative of combinatorial
491 control, where TFs physically interacting with each other form higher-order protein complexes
492 mediating increased DNA-binding specificity [64]. Consequently, TF pairs with many shared target
493 genes in our integrative network which also show correlated expression profiles are good
494 candidates for future research to shed light on the role of cooperative TF control in specific
495 biological processes in plants.
496 Predicted TF functions based on the iGRN regulatory interactions indicated that for various
497 biological processes many known regulators could be recovered. Starting from our systematic
498 regulatory annotation, known functional annotations were confirmed for 681 TFs. For several
499 biological processes our predicted TF functions are highly complementary to the AraNet
500 probabilistic functional gene network and for 268 TFs lacking experimental annotations, new
501 functions were inferred. To validate the power of our integrative network, novel regulators
502 predicted to play a role in oxidative stress signaling were experimentally validated using ROS-
503 specific phenotyping of loss- or gain-of-function lines. By analyzing transgenic lines perturbed in
504 one of the 32 novel ROS TFs, an additional 13 (41%) regulators were validated for a function in
505 ROS responses by our phenotyping studies. In addition to TF families that have often been
506 associated with oxidative or other environmental stress responses, the iGRN identified novel
507 ROS TFs from the GRAS, BES1 and GATA families that are not, or to a lesser extent, known to be
508 involved in stress responses. Notable, compared to previous studies that predicted gene
509 functions or prioritized candidate regulators based on (co-)expression analysis [65], our method
510 enables the identification of novel regulators that would have been missed when solely relying
511 on (differential) expression information. To conclude, our integrative network offers a high-

512 quality resource to study TF control for individual target genes and to identify, through the
513 interrogation of specific or condition-specific gene sets, novel regulators controlling different
514 plant traits.
515

516 **METHODS**

517 **Input datasets**

518 We assembled a comprehensive compendium of genome-wide datasets that were used as
519 primary data sources to construct the seven input networks and to build the integrative gene
520 regulatory network. Starting from the TAIR10 Arabidopsis gene annotation, we added a set of
521 5711 non-coding RNAs described in [66] resulting in a dataset covering 38,966 genes. A first
522 dataset consisted of 1,260 PWM binding site profiles for 874 transcription factors. This dataset
523 contained 108 and 623 positional frequency matrices that were obtained from protein binding
524 microarray studies performed by [67] and [68], respectively, and 529 positional frequency matrices
525 obtained from DAP-seq experiments [69]. A second dataset contained all available DNaseI
526 hypersensitive data obtained from the plant regulome data resource
527 (http://www.plantregulome.org/, [14]). All separate conditions were merged into a single open
528 chromatin dataset. A third dataset, composed of evolutionary conserved binding sites for 874
529 TFs, was obtained using the comparative motif mapping algorithm as described [15], starting from
530 the input motif dataset as described above.
531 A fourth dataset comprised gene expression profiles generated by RNA-seq. The RNA-Seq
532 expression compendium was built with public datasets from NCBI's Sequence Read Archive
533 (SRA) using Curse [70]. The compendium contains gene-level expression values for 40 manually
534 selected samples (Supplemental Table S9) of different treatment and tissue combinations (in
535 total >2,7 billion reads). SRA files for each sequencing run were downloaded from the SRA and
536 converted to the FASTQ format using fastq-dump (v2.4.4) from the SRA toolkit. FASTQ files from
537 runs of the same sample were concatenated. Paired-end reads were unpaired by randomly
538 selecting either the forward or reverse read and processing it as single-end. FastQC (v0.9.1) was
539 used to detect overrepresented adapter sequences, which were subsequently clipped with
540 fastx_clipper from the FASTX toolkit (v0.0.13). Nucleotides with Phred quality scores lower than
541 20 were trimmed with fastq_quality_trimmer from the FASTX toolkit. Reads shorter than 20
542 nucleotides after quality trimming were discarded. To obtain raw read counts for each transcript

543 in the TAIR10 annotation, Sailfish (v0.6.3) [71] was run with a k-mer length of 20. For genes with
544 multiple transcripts, the raw read counts of its transcripts were summed to get a gene-level read
545 count. Counts were then normalized for the entire compendium with the Variance Stabilizing
546 Transformation (VST) from the DESeq R package (v1.14.0) [72]. VST was chosen since it results in
547 correlation coefficients between genes that are most comparable to those obtained with
548 microarray data [73]. A fifth dataset consisted of a compendium of 27 genome-wide TF binding
549 datasets obtained through ChIP-seq [7].

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

551 We derived input networks from each dataset and a weight was assigned to each TF-target
552 gene interaction reflecting the certainty of the regulatory interaction. The motif target gene
553 network (PWM) was created by mapping all binding site data on an extended locus for each
554 gene in the *Arabidopsis thaliana* genome spanning 2kb upstream, intron and 1kb downstream
555 regions, with the exons masked. All binding site specificity matrices were mapped genome-wide
556 using Cluster-Buster with C parameter set to zero and M parameter set to 5 [74]. Both the C and M
557 scores were used as weights, which refer to the cluster and motif score of Cluster-Buster
558 (Supplemental Table S3). Due to the short length of motifs, many instances of motifs in a single
559 genome are expected to be coincidental matches that may not be bound *in vivo* or have no
560 functional regulatory effect. Therefore, the resulting set of binding sites was filtered using
561 DNaseI hypersensitive sites [14] in order to obtain an open chromatin filtered motif network
562 (PWM in DHS). The signalValue reported by the authors was retained as weight for motif
563 matches in open chromatin, this is a measurement of overall enrichment for the region to
564 DNaseI hypersensitivity. A second filtering approach was based on the evolutionarily
565 conservation of these binding sites resulting in a conserved motif network (Conserved PWM).
566 Weights in the conserved motif network are defined by the species conservation score according
567 to the Comparative Motif Mapping (CMM) method as described in [23]. Co-expression clusters
568 were built around all TFs obtained from PlantTFDB [75] using the Pearson correlation coefficient
569 and only the top 1000 most co-expressed genes were retained as putative target genes (CoE TF-
570 target). GENIE3 was run using standard parameters with the same list of regulators, only
571 predicted interactions with a variance in expression explained larger than 0.005 were retained
572 (GENIE3) [16]. A k-nearest neighbor (kNN) co-expression cluster of size 100 was built around each
573 gene in the *Arabidopsis thaliana* genome using the RNA-seq compendium as defined above.

574 PWM motif enrichment was performed on this co-expression cluster using the hypergeometric
575 distribution and significantly enriched motifs (p-value < 0.001 after Benjamini–Hochberg
576 correction for multiple hypotheses testing) were assigned to the members of the co-expression
577 cluster in order to obtain the kNN motif enrichment network (CoE + PWM). The enrichment fold
578 of the motif in the co-expression cluster is used as a weight. The peak scores for each TF bound
579 region assigned to the closest gene were used as weight in the ChIP-seq network (TF ChIP) [7].

580 **Integrative network inference**

581 The integrative network inference method is based on a binary classifier, where the class label
582 represents the presence or absence of an interaction. The set of true interactions (positives)
583 used in the training set was obtained from all validated interactions in AtRegNet [18], a review
584 article on known interactions in secondary cell wall development [19] and large scale data mining
585 effort on 974 peer reviewed articles called ATRM [20], and has 5,732 edges, connecting 522 TFs
586 with 3,738 targets. The remaining interactions between these TFs and potential target genes are
587 considered to be negatives. In other words, while no true "negatives" are present in the dataset,
588 we assign negative labels to TF–target pairs where the TF appears in the TFs of one of the
589 positive pairs and this TF-target pair is not in the positive pairs. The negative labels reflect the
590 fact that the TF has been studied and since no interaction is reported for it, it is more likely to
591 represent a negative interaction. This dataset was split into a dataset for learning (80%) and a
592 dataset for independent evaluation (20%). As such, the training set consists of 1,951,236 TF–
593 target gene pairs, out of which 4,586 are positives (true interactions) and 1,946,650 negatives.
594 The remaining gene pairs for TFs absent from the learning/evaluation dataset are considered to
595 be negatives. Using the weights derived from the input networks we tested a number of
596 different classification methods (gene pairs missing in specific input networks received a weight
597 of zero in the matrix used for learning). We extracted the 4,586 validated interacting pairs from
598 the true dataset and obtained the negative pairs for training by selecting a random subset of
599 13,758 (three times the number of positives; 3x 4,586) negatively labeled interacting pairs. Using
600 this labeled training set different classifiers were built using the weights of the input networks as
601 features. All features were normalized to have zero mean and unit variance. Classifiers were
602 evaluated using the 20% evaluation dataset. In this analysis the gradient boosting using decision
603 trees, random forests, and logistic regression were compared. The gradient boosting model
604 made use of 1,000 trees, shrinkage of 0.01 and an interaction depth of 3. The random forest

605 model used 500 trees selecting a random set of features of size square root (features) was
606 sampled in each iteration. Logistic regression was used with a binomial distribution suitable for
607 binary classification. All classifiers were compared using the area under the curve (Supplemental
608 Table S2).
609 The final network was generated using the gradient boosting classifier, as this was the best
610 performing throughout the training and validation steps. A total set of 60,714,507 unlabeled
611 interactions was classified. We included all interactions that are predicted to interact with a
612 probability of 0.35 or higher in the integrative network as this cutoff maximized the F1 measure
613 on the test dataset (Supplemental Table S2). The full iGRN can be mined using overlap and TF
614 enrichment analysis at http://bioinformatics.psb.ugent.be/webtools/iGRN/. Code, datasets and
615 documentation describing the supervised learning approach are available at
616 https://github.com/VIB-PSB/iGRN.
617 The overlap between the iGRN and the evaluation dataset was evaluated by counting how
618 many interactions from these experimental interactions were present in the integrative
619 network. The enrichment between two networks was defined as the number of interactions that
620 are present in both networks divided by the number of interactions expected by chance. The
621 number of common interactions expected by chance is given by the mean overlap of 10,000
622 randomized networks [21]. Network randomization was done by permuting the labels of all TFs
623 and the labels of all genes, which preserves the network structure.

624 **Functional evaluation using experimental datasets**

625 Three experimentally obtained gene regulatory networks were used for further validation of
626 the predicted network. An accumulation of publicly available Y1H data from literature was used
627 to generate a Y1H network containing 2,759 interactions [76-79]. A set of differentially expressed
628 genes upon TF perturbation was obtained from the CORNET database, containing 51,178
629 interactions [80]. An independent set of interactions determined by ChIP-seq was obtained from
630 literature, reporting 148,949 interactions [27,81-84]. From Song et al. 2016 only non-ABA conditions
631 were considered.
632 The overlap between the iGRN and three experimentally determined gene regulatory
633 networks was evaluated by counting how many interactions from the experimental networks
634 were present in the integrative network. The enrichment between two networks was defined as
635 the number of interactions that are present in both networks divided by the number of

636 interactions expected by chance, as explained above. Network randomization was done by
637 permuting the labels of all TFs and the labels of all genes, which preserves the network structure
638 and assures that the observed enrichment is not due to potential biases arising from structural
639 properties of the network (Supplemental Table S3).
640 Functional enrichment was determined for each sub network (intersection, FN, FP) using
641 biological metrics as described in [23] using Gene Ontology information. Note that no functional
642 GO information was used as input for the iGRN construction.

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

644 Based on the target genes for all TFs present in the iGRN, degree and betweenness centrality
645 were computed using the igraph R package, considering directed edge information. Three levels
646 of network hierarchy were calculated using the R code available from
647 http://archive.gersteinlab.org/proj/hinet/ [29]. The Jaccard index, defined as the intersection over
648 union, was used to compute the overlap of target genes between two TFs. Based on bulk
649 downloads of the RARGE II database ([28]; download November 2018), which reports ontology-
650 based phenotype information for 48,077 lines (17,671 RIKEN Arabidopsis Ds transposon mutant
651 lines, 16,337 CSHL Arabidopsis genetrap mutant lines, 14,069 RIKEN Arabidopsis full-length
652 cDNA overexpressed Arabidopsis lines). The presence of phenotypes in the RARGE II dataset was
653 scored for different sets of regulators, defined based on degree and betweenness centrality.

654 **TF function prediction**

655 For each TF, iGRN target genes were subjected to GO enrichment analysis using the
656 hypergeometric distribution combined with Benjamini–Hochberg (B&H) correction for multiple
657 hypotheses testing. A q-value cutoff 1e-03 was applied since this resulted in the best F1 score, a
658 metric that combines precision and recall to recover known GO annotations. Only experimental
659 and curated GO BP annotations were considered (version August 2018), GO annotations were
660 propagated to include parental GO terms, and a set of 18 general GO terms, corresponding with
661 terms located at the root of the GO hierarchy, were excluded (GO:0008150, GO:0009987,
662 GO:0008152, GO:0044237, GO:0071704, GO:0050896, GO:0065007, GO:0032502, GO:0050789,
663 GO:0032501, GO:0007275, GO:0050794, GO:0006355, GO:0045893, GO:0045892).
664 To predict ROS TFs, core genes that are part of the ROS wheel clusters IV-VIII, covering genes
665 related to chemical and genetic perturbation of cellular ROS/redox, were used to define a

666 custom category of ROS marker genes [32]. Subsequently, the over-representation of these ROS
667 markers in the set of iGRN target genes per TF, was used to predict novel ROS TFs and rank them
668 based on the q-value for enrichment (hypergeometric distribution, q-value cutoff 1e-05). The
669 assignment of ROS TFs to the categories KNOWN_ROS, KNOWN_STRESS and NOVEL_TF was
670 based on the information available in the literature at the moment the initial iGRN was
671 constructed (December 2016).

672 **Phenotyping candidate ROS TFs**

673 T-DNA insertion mutants were obtained from the European Arabidopsis Stock Centre (NASC;
674 http://arabidopsis.info/), GABI-Kat (http://www.gabi-kat.de/db/), Versailles Arabidopsis Stock
675 Center (http://publiclines.versailles.inra.fr/tdna/index) (Supplemental Table S6). Homozygous T-
676 DNA insertion lines were selected by PCR analysis (Supplemental Table S10). Mutant lines
677 previously published were provided by the authors (Supplemental Table S6 (overview lines)).
678 Overexpression lines were generated by Gateway cloning in the binary destination vector
679 pK7WG2D (18056864) (Supplemental Table S10). Plants were grown on half-strength Murashige
680 and Skoog (1/2MS) medium (Duchefa) supplemented with 1% (w/v) sucrose, 0.75% (w/v) agar,
681 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
682 photoperiod. To induce oxidative stress, the 1/2MS medium was supplemented with 15 or 20
683 nM MV; or 1 or 1.5 µM 3-AT, causing on average 15% (15 nM MV), 30% (20 nM MV), 20% (1 µM
684 3-AT) and 55% (1.5 µM 3-AT) growth reduction compared to the nontreated seedlings. Rosette
685 area was determined by automated area measurements of individual plant rosettes using
686 ImageJ. Raw phenotyping results are available in Data file 3. Presence or absence of
687 chlorosis/bleaching was monitored by visual assessment. For each mutant line, the rosette area
688 of mutant seedlings was compared with wild-type plants grown in the same Petri plate. Data
689 were obtained from at least two biological replicates per experiment and from independent
690 experiments.

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

692 A mixed model was fitted to the rosette area with genotype and treatment as fixed effects as
693 well as the interaction term with the mixed procedure from SAS (SAS Institute Inc. 2018.
694 SAS/STAT® 15.1, Cary, NC). The factor plate was added as random effect in the model to take
695 into account the correlation between observations made on the same plate. The Kenward-

696 Rogers method was used to calculate the denominator degrees of freedom of the fixed effects.
697 The null hypotheses of interest were an equal genotype effect upon treatment in the mutant
698 line and the wild-type line and an equal treatment versus MS effect between the mutant line
699 and the wild-type line (i.e. no interaction effect). A significance level of 0.05 was taken. To
700 correct for multiple testing, the maxT procedure was used as implemented in the plm
701 procedure.
702

703 **REFERENCES**

704 1 Franco-Zorrilla, J. M. & Solano, R. Identification of plant transcription factor target
705 sequences. *Biochim Biophys Acta* **1860**, 21-30, doi:10.1016/j.bbagrm.2016.05.001
706 (2017).
707 2 Lehti-Shiu, M. D., Panchy, N., Wang, P., Uygun, S. & Shiu, S. H. Diversity, expansion, and
708 evolutionary novelty of plant DNA-binding transcription factor families. *Biochim Biophys*
709 *Acta Gene Regul Mech* **1860**, 3-20, doi:10.1016/j.bbagrm.2016.08.005 (2017).
710 3 Jones, D. M. & Vandepoele, K. Identification and evolution of gene regulatory networks:
711 insights from comparative studies in plants. *Curr Opin Plant Biol* **54**, 42-48,
712 doi:10.1016/j.pbi.2019.12.008 (2020).
713 4 Rhee, S. Y. & Mutwil, M. Towards revealing the functions of all genes in plants. *Trends*
714 *Plant Sci* **19**, 212-221, doi:10.1016/j.tplants.2013.10.006 (2014).
715 5 Gaudinier, A. & Brady, S. M. Mapping Transcriptional Networks in Plants: Data-Driven
716 Discovery of Novel Biological Mechanisms. *Annu Rev Plant Biol* **67**, 575-594,
717 doi:10.1146/annurev-arplant-043015-112205 (2016).
718 6 Ferrier, T., Matus, J. T., Jin, J. & Riechmann, J. L. Arabidopsis paves the way: genomic and
719 network analyses in crops. *Curr Opin Biotechnol* **22**, 260-270,
720 doi:10.1016/j.copbio.2010.11.010 (2011).
721 7 Heyndrickx, K. S., Van de Velde, J., Wang, C., Weigel, D. & Vandepoele, K. A functional
722 and evolutionary perspective on transcription factor binding in Arabidopsis thaliana.
723 *Plant Cell* **26**, 3894-3910, doi:10.1105/tpc.114.130591 (2014).
724 8 Lu, Z., Hofmeister, B. T., Vollmers, C., DuBois, R. M. & Schmitz, R. J. Combining ATAC-seq
725 with nuclei sorting for discovery of cis-regulatory regions in plant genomes. *Nucleic*
726 *Acids Res* **45**, e41, doi:10.1093/nar/gkw1179 (2017).
727 9 Maher, K. A. *et al.* Profiling of Accessible Chromatin Regions across Multiple Plant
728 Species and Cell Types Reveals Common Gene Regulatory Principles and New Control
729 Modules. *Plant Cell* **30**, 15-36, doi:10.1105/tpc.17.00581 (2018).
730 10 Sijacic, P., Bajic, M., McKinney, E. C., Meagher, R. B. & Deal, R. B. Changes in chromatin
731 accessibility between Arabidopsis stem cells and mesophyll cells illuminate cell type-
732 specific transcription factor networks. *Plant J* **94**, 215-231, doi:10.1111/tpj.13882
733 (2018).
734 11 Kulkarni, S. R., Jones, D. M. & Vandepoele, K. Enhanced Maps of Transcription Factor
735 Binding Sites Improve Regulatory Networks Learned from Accessible Chromatin Data.
736 *Plant Physiol* **181**, 412-425, doi:10.1104/pp.19.00605 (2019).
737 12 Pajoro, A. *et al.* Dynamics of chromatin accessibility and gene regulation by MADS-
738 domain transcription factors in flower development. *Genome biology* **15**, R41 (2014).

739 13 Lee, I., Ambaru, B., Thakkar, P., Marcotte, E. M. & Rhee, S. Y. Rational association of
740 genes with traits using a genome-scale gene network for Arabidopsis thaliana. *Nat*
741 *Biotechnol* **28**, 149-156, doi:10.1038/nbt.1603 (2010).
742 14 Sullivan, A. M. *et al.* Mapping and dynamics of regulatory DNA and transcription factor
743 networks in A. thaliana. *Cell reports* **8**, 2015-2030 (2014).
744 15 Van de Velde, J., Van Bel, M., Vaneechoutte, D. & Vandepoele, K. A Collection of
745 Conserved Noncoding Sequences to Study Gene Regulation in Flowering Plants. *Plant*
746 *Physiol* **171**, 2586-2598, doi:10.1104/pp.16.00821 (2016).
747 16 Huynh-Thu, V. A., Irrthum, A., Wehenkel, L. & Geurts, P. Inferring regulatory networks
748 from expression data using tree-based methods. *PLoS One* **5**,
749 doi:10.1371/journal.pone.0012776 (2010).
750 17 Marbach, D. *et al.* Wisdom of crowds for robust gene network inference. *Nat Methods*
751 **9**, 796-804, doi:10.1038/nmeth.2016 (2012).
752 18 Yilmaz, A. *et al.* AGRIS: the Arabidopsis Gene Regulatory Information Server, an update.
753 *Nucleic Acids Res* **39**, D1118-1122, doi:10.1093/nar/gkq1120 (2011).
754 19 Hussey, S. G., Mizrachi, E., Creux, N. M. & Myburg, A. A. Navigating the transcriptional
755 roadmap regulating plant secondary cell wall deposition. *Front Plant Sci* **4**, 325,
756 doi:10.3389/fpls.2013.00325 (2013).
757 20 Jin, J. *et al.* An Arabidopsis Transcriptional Regulatory Map Reveals Distinct Functional
758 and Evolutionary Features of Novel Transcription Factors. *Mol Biol Evol* **32**, 1767-1773,
759 doi:10.1093/molbev/msv058 (2015).
760 21 Marbach, D. *et al.* Predictive regulatory models in Drosophila melanogaster by
761 integrative inference of transcriptional networks. *Genome Res* **22**, 1334-1349,
762 doi:10.1101/gr.127191.111 (2012).
763 22 Kheradpour, P., Stark, A., Roy, S. & Kellis, M. Reliable prediction of regulator targets
764 using 12 Drosophila genomes. *Genome Res* **17**, 1919-1931, doi:10.1101/gr.7090407
765 (2007).
766 23 Van de Velde, J., Heyndrickx, K. S. & Vandepoele, K. Inference of transcriptional
767 networks in Arabidopsis through conserved noncoding sequence analysis. *The Plant Cell*
768 **26**, 2729-2745 (2014).
769 24 Tompa, M. *et al.* Assessing computational tools for the discovery of transcription factor
770 binding sites. *Nat Biotechnol* **23**, 137-144, doi:10.1038/nbt1053 (2005).
771 25 O'Maoileidigh, D. S., Graciet, E. & Wellmer, F. Gene networks controlling Arabidopsis
772 thaliana flower development. *New Phytol* **201**, 16-30, doi:10.1111/nph.12444 (2014).
773 26 Luscombe, N. M. *et al.* Genomic analysis of regulatory network dynamics reveals large
774 topological changes. *Nature* **431**, 308-312, doi:10.1038/nature02782 (2004).
775 27 Song, L. *et al.* A transcription factor hierarchy defines an environmental stress response
776 network. *Science* **354**, aag1550 (2016).
777 28 Akiyama, K. *et al.* RARGE II: an integrated phenotype database of Arabidopsis mutant
778 traits using a controlled vocabulary. *Plant Cell Physiol* **55**, e4, doi:10.1093/pcp/pct165
779 (2014).
780 29 Cheng, C. *et al.* An approach for determining and measuring network hierarchy applied
781 to comparing the phosphorylome and the regulome. *Genome Biol* **16**, 63,
782 doi:10.1186/s13059-015-0624-2 (2015).
783 30 Lee, T. *et al.* AraNet v2: an improved database of co-functional gene networks for the
784 study of Arabidopsis thaliana and 27 other nonmodel plant species. *Nucleic Acids Res* **43**,
785 D996-1002, doi:10.1093/nar/gku1053 (2015).

31 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).
32 Willems, P. *et al.* The ROS Wheel: Refining ROS Transcriptional Footprints. *Plant Physiol* **171**, 1720-1733, doi:10.1104/pp.16.00420 (2016).
33 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).
34 Fichman, Y. *et al.* MYB30 Orchestrates Systemic Reactive Oxygen Signaling and Plant Acclimation. *Plant Physiol* **184**, 666-675, doi:10.1104/pp.20.00859 (2020).
35 Zandalinas, S. I., Sengupta, S., Burks, D., Azad, R. K. & Mittler, R. Identification and characterization of a core set of ROS wave-associated transcripts involved in the systemic acquired acclimation response of Arabidopsis to excess light. *Plant J* **98**, 126-141, doi:10.1111/tpj.14205 (2019).
36 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).
37 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).
38 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).
39 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).
40 Hruz, T. *et al.* Genevestigator v3: a reference expression database for the meta-analysis of transcriptomes. *Adv Bioinformatics* **2008**, 420747, doi:10.1155/2008/420747 (2008).
41 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).
42 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).
43 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).
44 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).
45 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).
46 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).
47 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).
48 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).

833 49 Yin, Y. *et al.* A new class of transcription factors mediates brassinosteroid-regulated
834 gene expression in Arabidopsis. *Cell* **120**, 249-259, doi:10.1016/j.cell.2004.11.044
835 (2005).
836 50 Torres-Galea, P., Huang, L. F., Chua, N. H. & Bolle, C. The GRAS protein SCL13 is a
837 positive regulator of phytochrome-dependent red light signaling, but can also modulate
838 phytochrome A responses. *Mol Genet Genomics* **276**, 13-30, doi:10.1007/s00438-006-
839 0123-y (2006).
840 51 Lachowiec, J., Mason, G. A., Schultz, K. & Queitsch, C. Redundancy, Feedback, and
841 Robustness in the Arabidopsis thaliana BZR/BEH Gene Family. *Front Genet* **9**, 523,
842 doi:10.3389/fgene.2018.00523 (2018).
843 52 Liu, P. P., Koizuka, N., Martin, R. C. & Nonogaki, H. The BME3 (Blue Micropylar End 3)
844 GATA zinc finger transcription factor is a positive regulator of Arabidopsis seed
845 germination. *Plant J* **44**, 960-971, doi:10.1111/j.1365-313X.2005.02588.x (2005).
846 53 Nowak, K., Wojcikowska, B. & Gaj, M. D. ERF022 impacts the induction of somatic
847 embryogenesis in Arabidopsis through the ethylene-related pathway. *Planta* **241**, 967-
848 985, doi:10.1007/s00425-014-2225-9 (2015).
849 54 Hoang, X. L. T., Nhi, D. N. H., Thu, N. B. A., Thao, N. P. & Tran, L. P. Transcription Factors
850 and Their Roles in Signal Transduction in Plants under Abiotic Stresses. *Curr Genomics*
851 **18**, 483-497, doi:10.2174/1389202918666170227150057 (2017).
852 55 Hirsch, S. & Oldroyd, G. E. GRAS-domain transcription factors that regulate plant
853 development. *Plant Signal Behav* **4**, 698-700, doi:10.4161/psb.4.8.9176 (2009).
854 56 Behringer, C. & Schwechheimer, C. B-GATA transcription factors - insights into their
855 structure, regulation, and role in plant development. *Front Plant Sci* **6**, 90,
856 doi:10.3389/fpls.2015.00090 (2015).
857 57 Hudson, D. *et al.* Rice cytokinin GATA transcription Factor1 regulates chloroplast
858 development and plant architecture. *Plant Physiol* **162**, 132-144,
859 doi:10.1104/pp.113.217265 (2013).
860 58 Zhang, C. *et al.* Genome-wide survey of the soybean GATA transcription factor gene
861 family and expression analysis under low nitrogen stress. *PLoS One* **10**, e0125174,
862 doi:10.1371/journal.pone.0125174 (2015).
863 59 Gupta, P., Nutan, K. K., Singla-Pareek, S. L. & Pareek, A. Abiotic Stresses Cause
864 Differential Regulation of Alternative Splice Forms of GATA Transcription Factor in Rice.
865 *Front Plant Sci* **8**, 1944, doi:10.3389/fpls.2017.01944 (2017).
866 60 Brooks, M. D. *et al.* Network Walking charts transcriptional dynamics of nitrogen
867 signaling by integrating validated and predicted genome-wide interactions. *Nat*
868 *Commun* **10**, 1569, doi:10.1038/s41467-019-09522-1 (2019).
869 61 Gaudinier, A. *et al.* Transcriptional regulation of nitrogen-associated metabolism and
870 growth. *Nature* **563**, 259-264, doi:10.1038/s41586-018-0656-3 (2018).
871 62 Zander, M. *et al.* Integrated multi-omics framework of the plant response to jasmonic
872 acid. *Nat Plants* **6**, 290-302, doi:10.1038/s41477-020-0605-7 (2020).
873 63 Reynoso, M. A. *et al.* Evolutionary flexibility in flooding response circuitry in
874 angiosperms. *Science* **365**, 1291-1295, doi:10.1126/science.aax8862 (2019).
875 64 Smaczniak, C., Muino, J. M., Chen, D., Angenent, G. C. & Kaufmann, K. Differences in
876 DNA Binding Specificity of Floral Homeotic Protein Complexes Predict Organ-Specific
877 Target Genes. *Plant Cell* **29**, 1822-1835, doi:10.1105/tpc.17.00145 (2017).
878 65 Ransbotyn, V. *et al.* A combination of gene expression ranking and co-expression
879 network analysis increases discovery rate in large-scale mutant screens for novel

880 Arabidopsis thaliana abiotic stress genes. *Plant Biotechnol J* **13**, 501-513,
881 doi:10.1111/pbi.12274 (2015).
882 66 Liu, J. *et al.* Genome-wide analysis uncovers regulation of long intergenic noncoding
883 RNAs in Arabidopsis. *Plant Cell* **24**, 4333-4345, doi:10.1105/tpc.112.102855 (2012).
884 67 Franco-Zorrilla, J. M. *et al.* DNA-binding specificities of plant transcription factors and
885 their potential to define target genes. *Proceedings of the National Academy of Sciences*
886 **111**, 2367-2372 (2014).
887 68 Weirauch, M. T. *et al.* Determination and inference of eukaryotic transcription factor
888 sequence specificity. *Cell* **158**, 1431-1443, doi:10.1016/j.cell.2014.08.009 (2014).
889 69 O'Malley, R. C. *et al.* Cistrome and Epicistrome Features Shape the Regulatory DNA
890 Landscape. *Cell* **165**, 1280-1292 (2016).
891 70 Vaneechoutte, D. & Vandepoele, K. Curse: building expression atlases and co-expression
892 networks from public RNA-Seq data. *Bioinformatics* **35**, 2880-2881,
893 doi:10.1093/bioinformatics/bty1052 (2019).
894 71 Patro, R., Mount, S. M. & Kingsford, C. Sailfish enables alignment-free isoform
895 quantification from RNA-seq reads using lightweight algorithms. *Nat Biotechnol* **32**, 462-
896 464, doi:10.1038/nbt.2862 (2014).
897 72 Anders, S. & Huber, W. Differential expression analysis for sequence count data.
898 *Genome Biol* **11**, R106, doi:10.1186/gb-2010-11-10-r106 (2010).
899 73 Giorgi, F. M., Del Fabbro, C. & Licausi, F. Comparative study of RNA-seq- and microarray-
900 derived coexpression networks in Arabidopsis thaliana. *Bioinformatics* **29**, 717-724,
901 doi:10.1093/bioinformatics/btt053 (2013).
902 74 Frith, M. C., Li, M. C. & Weng, Z. Cluster-Buster: Finding dense clusters of motifs in DNA
903 sequences. *Nucleic Acids Res* **31**, 3666-3668 (2003).
904 75 Jin, J., Zhang, H., Kong, L., Gao, G. & Luo, J. PlantTFDB 3.0: a portal for the functional and
905 evolutionary study of plant transcription factors. *Nucleic Acids Res* **42**, D1182-1187,
906 doi:10.1093/nar/gkt1016 (2014).
907 76 Brady, S. M. *et al.* A stele-enriched gene regulatory network in the Arabidopsis root. *Mol*
908 *Syst Biol* **7**, 459, doi:10.1038/msb.2010.114 (2011).
909 77 Li, B. *et al.* Promoter-based integration in plant defense regulation. *Plant Physiol* **166**,
910 1803-1820, doi:10.1104/pp.114.248716 (2014).
911 78 Taylor-Teeples, M. *et al.* An Arabidopsis gene regulatory network for secondary cell wall
912 synthesis. *Nature* **517**, 571-575 (2015).
913 79 Tian, C. *et al.* An organ boundary-enriched gene regulatory network uncovers regulatory
914 hierarchies underlying axillary meristem initiation. *Mol Syst Biol* **10**, 755,
915 doi:10.15252/msb.20145470 (2014).
916 80 De Bodt, S., Hollunder, J., Nelissen, H., Meulemeester, N. & Inze, D. CORNET 2.0:
917 integrating plant coexpression, protein-protein interactions, regulatory interactions,
918 gene associations and functional annotations. *New Phytol* **195**, 707-720,
919 doi:10.1111/j.1469-8137.2012.04184.x (2012).
920 81 Nagel, D. H. *et al.* Genome-wide identification of CCA1 targets uncovers an expanded
921 clock network in Arabidopsis. *Proc Natl Acad Sci U S A* **112**, E4802-4810,
922 doi:10.1073/pnas.1513609112 (2015).
923 82 Verkest, A. *et al.* A generic tool for transcription factor target gene discovery in
924 Arabidopsis cell suspension cultures based on tandem chromatin affinity purification.
925 *Plant Physiol* **164**, 1122-1133, doi:10.1104/pp.113.229617 (2014).

926 83 Birkenbihl, R. P., Diezel, C. & Somssich, I. E. Arabidopsis WRKY33 is a key transcriptional
927 regulator of hormonal and metabolic responses toward Botrytis cinerea infection. *Plant*
928 *Physiol* **159**, 266-285, doi:10.1104/pp.111.192641 (2012).
929 84 Merelo, P. *et al.* Genome-wide identification of KANADI1 target genes. *PLoS One* **8**,
930 e77341, doi:10.1371/journal.pone.0077341 (2013).
931 85 Qureshi, M. K., Sujeeth, N., Gechev, T. S. & Hille, J. The zinc finger protein ZAT11
932 modulates paraquat-induced programmed cell death in Arabidopsis thaliana. *Acta*
933 *Physiologiae Plantarum* **35**, 1863–1871 (2013).
934 86 Liu, X. M. *et al.* ZAT11, a zinc finger transcription factor, is a negative regulator of nickel
935 ion tolerance in Arabidopsis. *Plant Cell Rep* **33**, 2015-2021, doi:10.1007/s00299-014-
936 1675-7 (2014).
937 87 Rizhsky, L., Davletova, S., Liang, H. & Mittler, R. The zinc finger protein Zat12 is required
938 for cytosolic ascorbate peroxidase 1 expression during oxidative stress in Arabidopsis. *J*
939 *Biol Chem* **279**, 11736-11743, doi:10.1074/jbc.M313350200 (2004).
940 88 Davletova, S., Schlauch, K., Coutu, J. & Mittler, R. The zinc-finger protein Zat12 plays a
941 central role in reactive oxygen and abiotic stress signaling in Arabidopsis. *Plant Physiol*
942 **139**, 847-856, doi:10.1104/pp.105.068254 (2005).
943 89 Scarpeci, T. E., Zanor, M. I., Mueller-Roeber, B. & Valle, E. M. Overexpression of
944 AtWRKY30 enhances abiotic stress tolerance during early growth stages in Arabidopsis
945 thaliana. *Plant Mol Biol* **83**, 265-277, doi:10.1007/s11103-013-0090-8 (2013).
946 90 Perez-Salamo, I. *et al.* The heat shock factor A4A confers salt tolerance and is regulated
947 by oxidative stress and the mitogen-activated protein kinases MPK3 and MPK6. *Plant*
948 *Physiol* **165**, 319-334, doi:10.1104/pp.114.237891 (2014).
949 91 Vanderauwera, S. *et al.* AtWRKY15 perturbation abolishes the mitochondrial stress
950 response that steers osmotic stress tolerance in Arabidopsis. *Proc Natl Acad Sci U S A*
951 **109**, 20113-20118, doi:10.1073/pnas.1217516109 (2012).
952 92 Mittler, R. *et al.* Gain- and loss-of-function mutations in Zat10 enhance the tolerance of
953 plants to abiotic stress. *FEBS Lett* **580**, 6537-6542, doi:10.1016/j.febslet.2006.11.002
954 (2006).
955 93 Sewelam, N. *et al.* Ethylene response factor 6 is a regulator of reactive oxygen species
956 signaling in Arabidopsis. *PLoS One* **8**, e70289, doi:10.1371/journal.pone.0070289 (2013).
957 94 Moffat, C. S. *et al.* ERF5 and ERF6 play redundant roles as positive regulators of JA/Et-
958 mediated defense against Botrytis cinerea in Arabidopsis. *PLoS One* **7**, e35995,
959 doi:10.1371/journal.pone.0035995 (2012).
960 95 Dubois, M. *et al.* Ethylene Response Factor6 acts as a central regulator of leaf growth
961 under water-limiting conditions in Arabidopsis. *Plant Physiol* **162**, 319-332,
962 doi:10.1104/pp.113.216341 (2013).
963 96 Mahmood, K., Xu, Z., El-Kereamy, A., Casaretto, J. A. & Rothstein, S. J. The Arabidopsis
964 Transcription Factor ANAC032 Represses Anthocyanin Biosynthesis in Response to High
965 Sucrose and Oxidative and Abiotic Stresses. *Front Plant Sci* **7**, 1548,
966 doi:10.3389/fpls.2016.01548 (2016).
967 97 Allu, A. D., Brotman, Y., Xue, G. P. & Balazadeh, S. Transcription factor ANAC032
968 modulates JA/SA signalling in response to Pseudomonas syringae infection. *EMBO Rep*
969 **17**, 1578-1589, doi:10.15252/embr.201642197 (2016).
970 98 Babitha, K. C. *et al.* Co-expression of AtbHLH17 and AtWRKY28 confers resistance to
971 abiotic stress in Arabidopsis. *Transgenic Res* **22**, 327-341, doi:10.1007/s11248-012-9645-
972 8 (2013).

973 99 Chen, X. *et al.* Overexpression of AtWRKY28 and AtWRKY75 in Arabidopsis enhances
974 resistance to oxalic acid and Sclerotinia sclerotiorum. *Plant Cell Rep* **32**, 1589-1599,
975 doi:10.1007/s00299-013-1469-3 (2013).
976 100 Wu, L., Zhong, G., Wang, J., Li, X. & Yang, Y. Functional analysis of an arabidopsis
977 transcription factor WRKY28 in resistance to alternaria brassicicola. *Journal of*
978 *Agricultural Science and Technology* **14**, 65-71 (2012).
979 101 De Clercq, I. *et al.* The membrane-bound NAC transcription factor ANAC013 functions in
980 mitochondrial retrograde regulation of the oxidative stress response in Arabidopsis.
981 *Plant Cell* **25**, 3472-3490, doi:10.1105/tpc.113.117168 (2013).
982 102 Kim, J. B., Kang, J. Y. & Kim, S. Y. Over-expression of a transcription factor regulating
983 ABA-responsive gene expression confers multiple stress tolerance. *Plant Biotechnol J* **2**,
984 459-466, doi:10.1111/j.1467-7652.2004.00090.x (2004).
985 103 Van Aken, O. *et al.* Mitochondrial and Chloroplast Stress Responses Are Modulated in
986 Distinct Touch and Chemical Inhibition Phases. *Plant Physiol* **171**, 2150-2165,
987 doi:10.1104/pp.16.00273 (2016).
988 104 Ng, S. *et al.* A membrane-bound NAC transcription factor, ANAC017, mediates
989 mitochondrial retrograde signaling in Arabidopsis. *Plant Cell* **25**, 3450-3471,
990 doi:10.1105/tpc.113.113985 (2013).
991 105 Dombrecht, B. *et al.* MYC2 differentially modulates diverse jasmonate-dependent
992 functions in Arabidopsis. *Plant Cell* **19**, 2225-2245, doi:10.1105/tpc.106.048017 (2007).
993 106 Verma, D., Jalmi, S. K., Bhagat, P. K., Verma, N. & Sinha, A. K. A bHLH transcription
994 factor, MYC2, imparts salt intolerance by regulating proline biosynthesis in Arabidopsis.
995 *FEBS J* **287**, 2560-2576, doi:10.1111/febs.15157 (2020).
996 107 Song, S. *et al.* Interaction between MYC2 and ETHYLENE INSENSITIVE3 modulates
997 antagonism between jasmonate and ethylene signaling in Arabidopsis. *Plant Cell* **26**,
998 263-279, doi:10.1105/tpc.113.120394 (2014).
999 108 Wu, A. *et al.* JUNGBRUNNEN1, a reactive oxygen species-responsive NAC transcription
1000 factor, regulates longevity in Arabidopsis. *Plant Cell* **24**, 482-506,
1001 doi:10.1105/tpc.111.090894 (2012).
1002 109 Shahnejat-Bushehri, S., Nobmann, B., Devi Allu, A. & Balazadeh, S. JUB1 suppresses
1003 Pseudomonas syringae-induced defense responses through accumulation of DELLA
1004 proteins. *Plant Signal Behav* **11**, e1181245, doi:10.1080/15592324.2016.1181245
1005 (2016).
1006 110 Ebrahimian-Motlagh, S. *et al.* JUNGBRUNNEN1 Confers Drought Tolerance Downstream
1007 of the HD-Zip I Transcription Factor AtHB13. *Front Plant Sci* **8**, 2118,
1008 doi:10.3389/fpls.2017.02118 (2017).
1009 111 Nishizawa, A. *et al.* Arabidopsis heat shock transcription factor A2 as a key regulator in
1010 response to several types of environmental stress. *Plant J* **48**, 535-547,
1011 doi:10.1111/j.1365-313X.2006.02889.x (2006).
1012 112 Nishizawa-Yokoi, A., Yabuta, Y. & Shigeoka, S. The contribution of carbohydrates
1013 including raffinose family oligosaccharides and sugar alcohols to protection of plant cells
1014 from oxidative damage. *Plant Signal Behav* **3**, 1016-1018, doi:10.4161/psb.6738 (2008).
1015 113 Zhang, L., Li, Y., Xing, D. & Gao, C. Characterization of mitochondrial dynamics and
1016 subcellular localization of ROS reveal that HsfA2 alleviates oxidative damage caused by
1017 heat stress in Arabidopsis. *J Exp Bot* **60**, 2073-2091, doi:10.1093/jxb/erp078 (2009).
1018 114 Olate, E., Jimenez-Gomez, J. M., Holuigue, L. & Salinas, J. NPR1 mediates a novel
1019 regulatory pathway in cold acclimation by interacting with HSFA1 factors. *Nat Plants* **4**,
1020 811-823, doi:10.1038/s41477-018-0254-2 (2018).

1021 115 Enomoto, T. *et al.* STOP1 regulates the expression of HsfA2 and GDHs that are critical for
1022 low-oxygen tolerance in Arabidopsis. *J Exp Bot* **70**, 3297-3311, doi:10.1093/jxb/erz124
1023 (2019).
1024 116 Hsieh, T. H. *et al.* Heterology expression of the Arabidopsis C-repeat/dehydration
1025 response element binding factor 1 gene confers elevated tolerance to chilling and
1026 oxidative stresses in transgenic tomato. *Plant Physiol* **129**, 1086-1094,
1027 doi:10.1104/pp.003442 (2002).
1028 117 Jaglo-Ottosen, K. R., Gilmour, S. J., Zarka, D. G., Schabenberger, O. & Thomashow, M. F.
1029 Arabidopsis CBF1 overexpression induces COR genes and enhances freezing tolerance.
1030 *Science* **280**, 104-106, doi:10.1126/science.280.5360.104 (1998).
1031 118 Hsieh, T. H., Lee, J. T., Charng, Y. Y. & Chan, M. T. Tomato plants ectopically expressing
1032 Arabidopsis CBF1 show enhanced resistance to water deficit stress. *Plant Physiol* **130**,
1033 618-626, doi:10.1104/pp.006783 (2002).
1034

1035 **ACKNOWLEDGEMENTS**

1036 The authors wish to thank Cordelia Bolle (SCL13), Xuemei Chen (ARF3), Yi-Fang Chen
1037 (WRKY45), Donna Fernandez (AGL15), Henning Frerigmann (MYB51, MYB122), Małgorzata Gaj
1038 (ERF022), Tamara gigolashvili (MYB51), Dwayne Hegedus (SCL15), Lars Hennig (AGL19), Ildoo
1039 Hwang (BME3), Zhongchi Liu (TSO1), Lars Østergaard (ARF3), Javier Paz-Ares (PHL1) and Olivier
1040 Van Aken (WRKY45) for kindly providing Arabidopsis mutant lines. We thank Zoé Joly-Lopez and
1041 Olivia Wilkins for providing in-depth comments on an earlier version of the manuscript, as well
1042 as Camilla Ferrari and Indeewari Dissanayake for proofreading. I.D.C. is supported by the
1043 Research Foundation Flanders (postdoctoral fellowship 12N2415N and grant for a long stay
1044 abroad V400620N). J.V.d.V. is indebted to the Agency for Innovation by Science and Technology
1045 in Flanders for a predoctoral fellowship. X.L. and L.L. are supported by the China Scholarship
1046 Council for a PhD fellowship (201706910099 and 201808530499). F.V.B. is supported by the
1047 Fonds Wetenschappelijk Onderzoek – Vlaanderen (FWO) and the Fonds de la Recherche
1048 Scientifique – FNRS under EOS Project No. 30829584 and the Research Foundation Flanders
1049 (project nos. G0D7914N and G055416N).

1050 **AUTHOR CONTRIBUTIONS STATEMENT**

1051 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
1052 phenotype screening. V.S. performed statistical data analysis. D.V. generated the expression
1053 compendium. J.V.d.V. and K.V. performed network analysis and function prediction. J.V.d.V.,
1054 I.D.C., F.V.B. and K.V. wrote the manuscript.

COMPETING INTERESTS STATEMENTS

The authors declare no competing interests.

1058 **FIGURE LEGENDS**

1059 **Figure 1. Construction and evaluation of the integrated gene regulatory network**
1060 **(iGRN).**

1061 a. Overview of the different input networks used to construct the supervised iGRN. Whereas
1062 colored circles indicate the type of data that is used to construct the input network, the
1063 information below each network depicts how edges in that network are weighted. b. Overview
1064 of the sizes of the different networks together with the recall and enrichment for edges in the
1065 evaluation test dataset (black and white bars, respectively). The brown bars report the fraction
1066 of each input network that is present in the iGRN while the red bars report which fraction of the
1067 supervised iGRN edges is supported by this input network.

1068 **Figure 2. Overlap between iGRN interactions and different experimental data sources**
1069 **describing gene regulatory and functional information.**

1070 a. Venn diagrams showing the number of shared interactions between the iGRN and TF ChIP,
1071 Y1H and differentially expressed (DE) genes after TF perturbation (depicted in blue, yellow and
1072 red, respectively). b. Histograms showing the expected overlap based on 10,000 randomized
1073 networks, together with the observed overlap indicated by the black arrow. The p-values for
1074 observing the measured overlap by chance are also reported (permutation based statistics, one-
1075 sided). c. Fold enrichment for a functional metric describing how well target genes of a specific
1076 TF are involved in similar biological processes (average values for all enriched GO terms and all
1077 TFs present in the different data series are reported). Whereas the grey bars indicate the
1078 functional enrichment for shared interactions, blue, red, yellow and green depict fold
1079 enrichment for interactions unique to the ChIP, DE, Y1H and iGRN, respectively.
1080

1081 **Figure 3. Phenotypes reported after TF perturbation for TFs with different network**
1082 **centrality properties.**

1083 a. RARGE-II phenotypes reported for iGRN TFs with different betweenness centrality. Q1-Q4
1084 refer to the four quartiles of all iGRN TFs (sorted based on increasing betweenness centrality)
1085 and values in parenthesis refer to the median betweenness centrality value together with the
1086 number of lines available in the RARGE II dataset for that quartile. b. Betweenness centrality

1087 distribution for TFs with and without phenotypes reported in RARGE-II (n= 153 and 1589 TFs,
1088 respectively).

1089 **Figure 4. Oxidative stress phenotyping results of novel ROS TFs.**

1090 Heatmap presenting phenotypes associated with mutant alleles of iGRN extracted novel ROS TFs
1091 under different oxidative stress conditions. Mutant lines with a statistically significant
1092 phenotype under oxidative stress (1 µM and 1.5 µM 3-AT and 15 and 20 nM MV) and/or normal
1093 growth conditions (MS) are displayed. Results of other mutant lines that were tested in this
1094 study can be found in Supplemental Table S5 and S6. Statistically significant differences relative
1095 to wild type (Col-0) are indicated by asterisk; *, $P < 0.05$ and **, $P < 0.001$ as determined by the
1096 hypothesis tests on the fixed effects of the mixed model (two-sided, see Methods). Data were
1097 obtained from at least two biological replicates per experiment and from independent
1098 experiments. Color scale indicates the ratio of the rosette area of 14-day-old mutant alleles
1099 relative to wild type (Col-0). Black and grey borders indicate presence or absence of a bleaching
1100 phenotype of the mutant plants, respectively. Small grey circle indicates a statistically significant
1101 genotype-by-stress treatment interaction effect.

1102 **Figure 5. Representative images of oxidative stress phenotypes for a selection of**
1103 **mutants.**

1104 Representative images of novel ROS TFs with a phenotype during oxidative stress. 17-days-old
1105 seedlings of selected mutant lines compared to wild type (Col-0). Scale bars, 1 cm.

1106 **Figure 6. Expression of KNOWN and NOVEL ROS TFs during various conditions**
1107 **provoking cellular ROS/redox imbalances.**

1108 Heatmap showing the expression response of ROS TFs extracted from the iGRN during a
1109 selection of conditions provoking cellular ROS production or redox imbalances (a more elaborate
1110 set of oxidative stress conditions is displayed in Supplemental Table S7). TFs with previously
1111 known function in ROS responses and TFs with no functional link with any oxidative stress
1112 and/or (a)biotic stress responses (NOVEL) that were tested in our phenotyping assays are
1113 displayed. NOVEL TFs validated in our study are shown using blue and red circles, indicating a
1114 positive and negative effect by the TF on oxidative stress tolerance, respectively.

1115 **TABLES**

1116 **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)

1117

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	[33]	**MV**
iGRN3	ZAT11	2.81E-25	KNOWN_ROS	MV	[85]	heavy metal	[86]	
iGRN5	ZAT12	8.70E-25	KNOWN_ROS	MV	[87]	osmotic/salt/high light/heat	[88]	
iGRN6	WRKY30	1.42E-24	KNOWN_ROS	MV	[89]	salt	[89]	
iGRN11	HSFA4A	5.09E-23	KNOWN_ROS	MV	[90]	salt, anoxia	[90]	
iGRN12	WRKY15	7.46E-23	KNOWN_ROS	H2O2	[91]	salt/osmotic	[91]	
iGRN16	STZ	4.90E-22	KNOWN_ROS	MV	[92]	osmotic/salt/heat	[92]	
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	[93]	high light, B. cinerea, osmotic	[93 94 95]	

iGRN50	WRKY45	5.93E-13	**NOVEL(*)**			submergence stress	[47]	**3-AT, MV**
iGRN52	ANAC032	1.26E-12	KNOWN_ROS	3-AT	[96]	P. syringae, salt/high light	[96 97]	
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	[98]	salt/osmotic/ drought, S. sclerotiorum, A. brassicicola	[98 99 100]	
iGRN62	ANAC013	1.93E-10	KNOWN_ROS	MV	[101]			
iGRN65	ABF3	8.06E-10	KNOWN_ROS	MV	[102]	cold/freezing/heat /drought	[102]	
iGRN66	ERF022	9.55E-10	**NOVEL**					**MV**
iGRN68	BEH2	2.75E-09	**NOVEL**					**3-AT, MV**
iGRN69	ANAC017	3.04E-09	KNOWN_ROS	MV	[103]	drought/high light	[104]	
iGRN81	PHL1	2.76E-08	**NOVEL**					**3-AT, MV**
iGRN85	JAM1	4.07E-08	KNOWN_ROS	MV, H2O2	[98]	salt/osmotic/ drought	[98]	
iGRN93	MYC2	4.19E-07	KNOWN_ROS	MV	[105]	salt, B. Cinerea	[106 107]	
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	[108]	salt/heat, P. syringae, drought	[108 109 110]	
iGRN111	HSFA2	3.21E-06	KNOWN_ROS	MV	[111]	high light, heat, cold, hypoxia	[111 112 113 114 115]	
iGRN118	CBF1	5.32E-06	KNOWN_ROS	MV	[116]	salt, drought, freezing	[116 117 118]	

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

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. Examples of predicted cooperative TFs supported by protein-protein interactions.

Figure S4. Recovery of known GO annotations for TFs belonging to different biological processes.

Figure S5. Oxidative stress-responsive and tissue-specific expression profiles of ANAC102 and its ChIP-confirmed target genes.

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. Expression of ANAC102 and its ChIP-confirmed target genes during various conditions provoking cellular ROS/redox imbalances.

Supplemental Table S9. Overview of samples included in the RNA-Seq expression compendium.

Supplemental Table S10. 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.

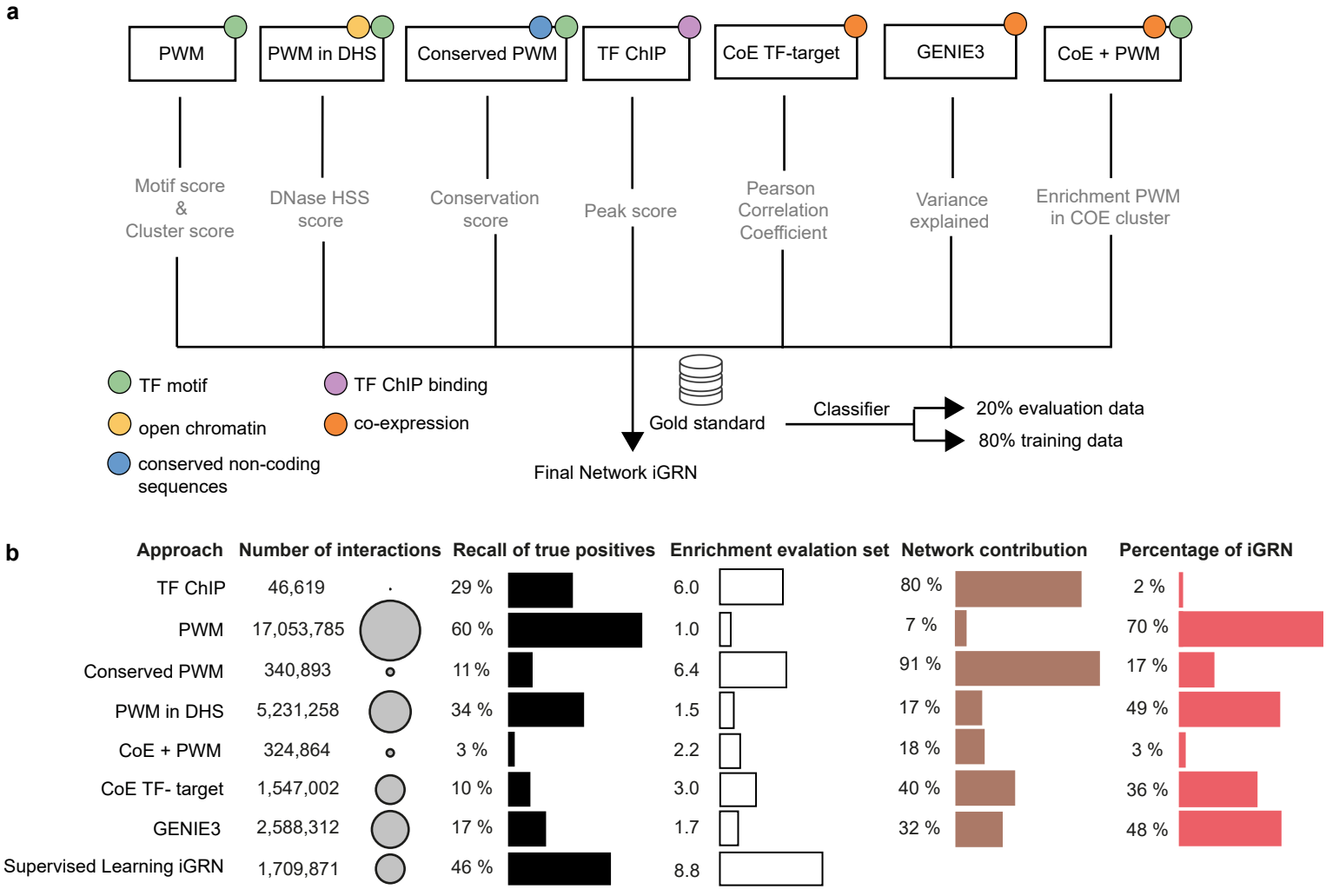
Data file 2. TF_function_all.txt reporting predicted functions for all TFs present in the iGRN.

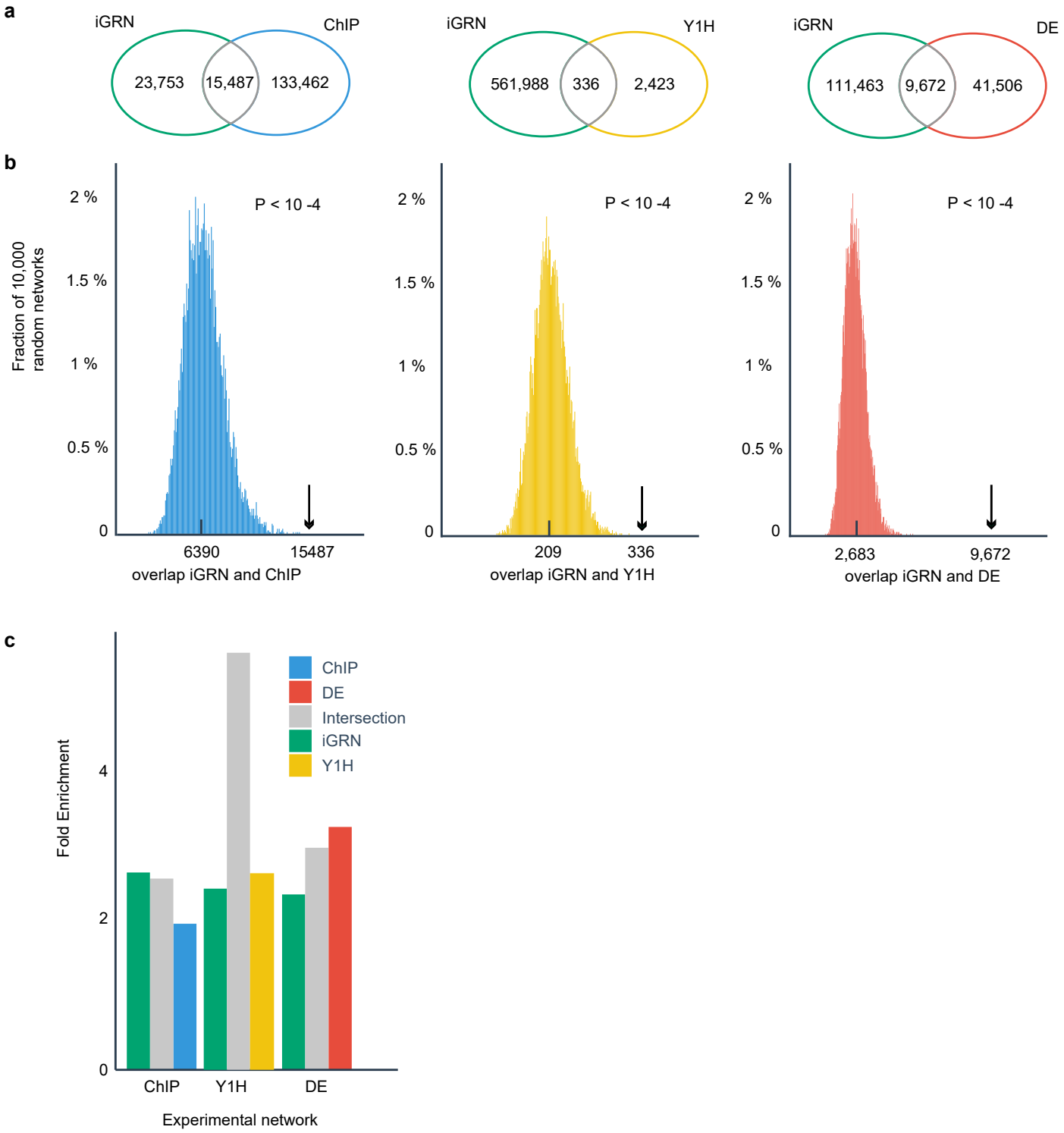
The sheet README explains the different columns and offers 3 examples how this file can be explored for specific TFs, functions or a specific target gene.

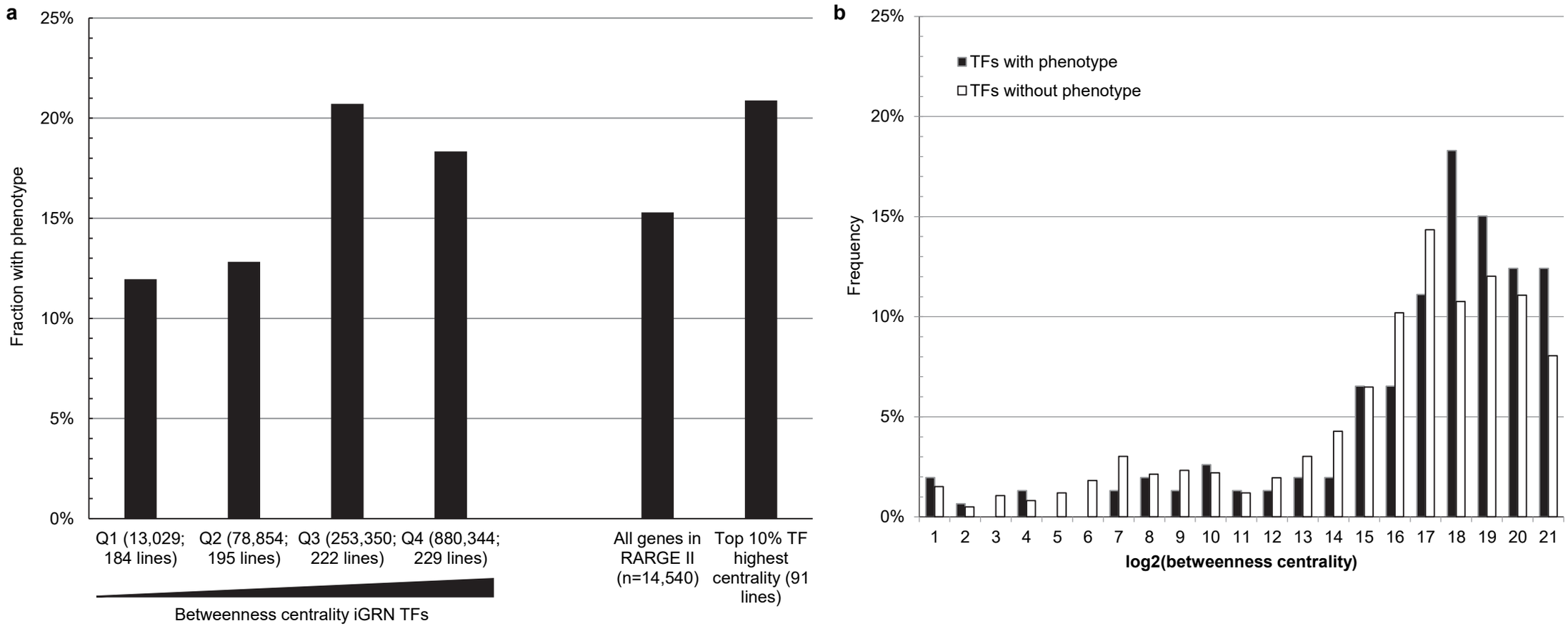
Overview of generated, analysed and used iGRN TF function data for Fig. S3.

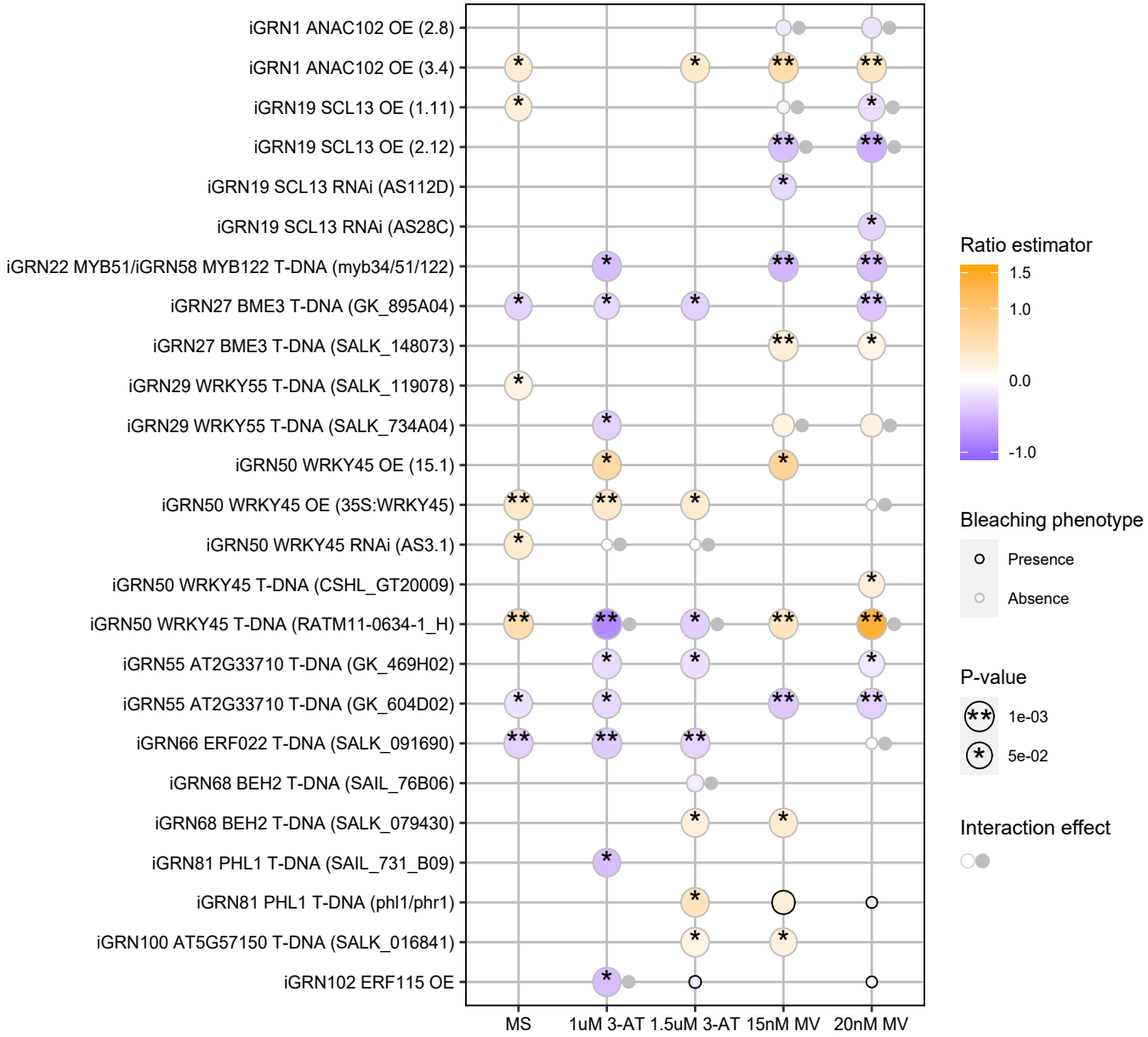
Data file 3. Raw phenotype results that were used as input for the statistical analysis.

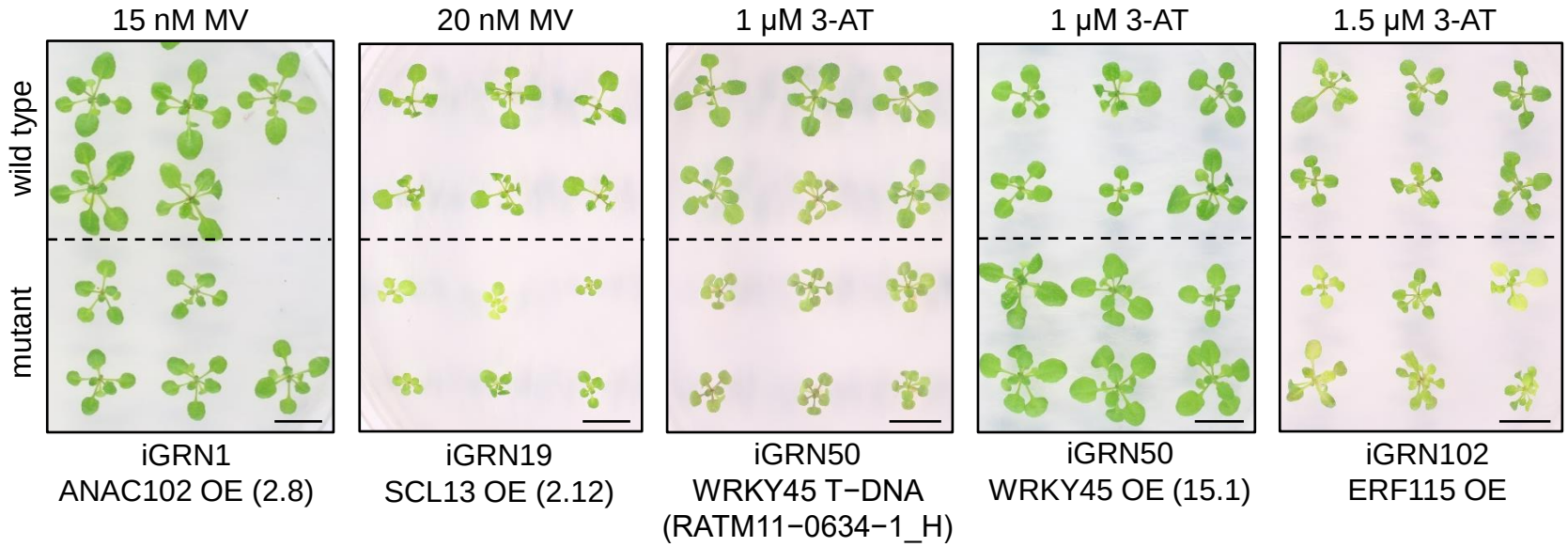
Overview of raw phenotyping measurements for Fig. 4.

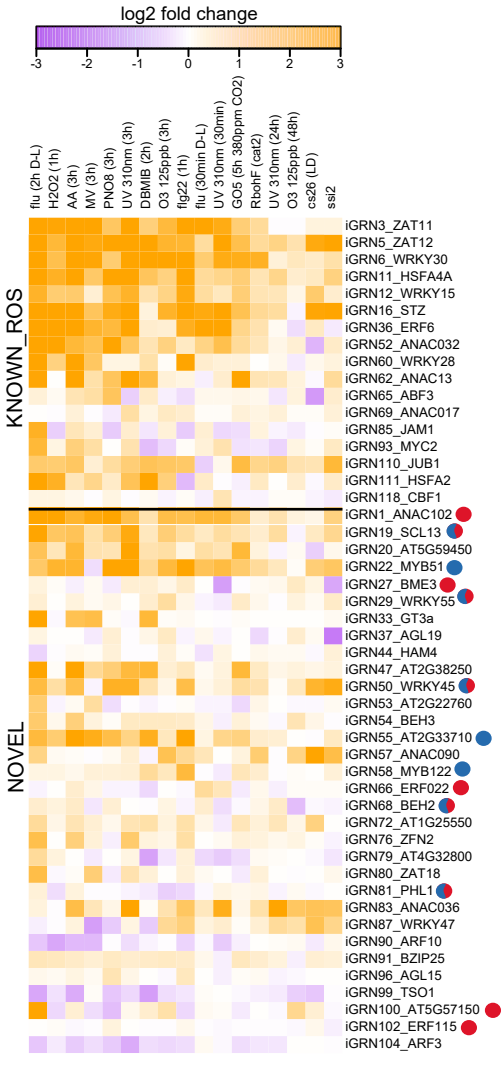

a
PWM
PWM in DHS
Conserved PWM
TF ChIP
CoE TF-target
GENIE3
CoE + PWM
Motif score & Cluster score
DNase HSS score
Conservation score
Peak score
Pearson Correlation Coefficient
Variance explained
Enrichment PWM in COE cluster
TF motif
open chromatin
conserved non-coding sequences
TF ChIP binding
co-expression
Gold standard
Classifier
20% evaluation data
80% training data
Final Network iGRN
b
Approach
Number of interactions
Recall of true positives
Enrichment evalation set
Network contribution
Percentage of iGRN
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


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)
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