CrY2H-seq: a massively-multiplexed assay for deep coverage interactome mapping

Shelly A. Trigg[1,2], **Renee M. Garza**[1], **Andrew MacWilliams**[1], **Joseph R. Nery**[1], **Anna Bartlett**[1], **Rosa Castanon**[1], **Adeline Goubil**[1,4], **Joseph Feeney**[1,4], **Ronan O'Malley**[1,4], **Shao-shan Carol Huang**[1], **Zhuzhu Z. Zhang**[1], **Mary Galli**[1,4], and **Joseph R. Ecker**[1,2,3,*]

[1]Genomic Analysis and Plant Biology Laboratory, The Salk Institute for Biological Studies, La Jolla, California, USA

[2]Division of Biological Sciences, University of California San Diego, La Jolla, California, USA

[3]Howard Hughes Medical Institute, The Salk Institute for Biological Studies, La Jolla, California, USA

Abstract

Broad-scale protein-protein interaction mapping is a major challenge given the cost, time, and sensitivity constraints of existing technologies. Here, we present a massively-multiplexed yeast two-hybrid method, CrY2H-seq, that uses a Cre recombinase interaction reporter to intracellularly fuse the coding sequences of two interacting proteins, and next-generation DNA sequencing to identify these interactions *en masse*. We applied CrY2H-seq to investigate sparsely annotated combinatorial interactions among plant transcription factors. By performing ten independent CrY2H-seq screens each testing 3.6 million interaction combinations, and reporting a deep coverage network of 8,577 interactions among 1,453 transcription factors, we demonstrate CrY2H-seq's improved capacity, efficiency, and sensitivity over existing technologies. In addition to recapitulating one third of previously reported interactions derived from diverse methods, we expand the number of reported plant transcription factor interactions by three-fold, revealing

Data availability

Protein interaction data from this study are included in this published article (and its supplementary information files), and can be found at http://signal.salk.edu/interactome/AtTFIN-1.html. Raw read data files and alignment indexes can be found at http://neomorph.salk.edu/download/CrY2H-seq_TFxTF.tar.

Code availability

Code generated for analysis during this study is included in this published article (and its supplementary information files).

previously unknown family-specific interaction module associations with plant reproductive development, root architecture, and circadian coordination.

INTRODUCTION

The yeast two-hybrid (Y2H) assay is one of the most widely adopted methods for high-throughput mapping of binary protein-protein interactions. Y2H datasets[1–3] have largely contributed to widely used protein interaction repositories[4] and probabilistic interactome databases[5,6]. Y2H data have revealed complexes regulating disease[7] and improved the interpretation of disease phenotypes arising from genomic or transcriptomic variation[8,9]. However, broad-scale Y2H data acquisition remains constrained by the cost and labor requirements of tracking interactions and the iterative screening necessary to generate complete interactome maps[10].

Advancements that leverage next-generation sequencing to identify interactions have made large-scale Y2H screening more feasible[1,11,12]. To circumvent the isolated screening of bait proteins for tracking interactions, multiplexed screening strategies that enable pools of baits to be screened against pools of preys were recently developed[12,13]. Barcode Fusion Genetics (BFG-Y2H) uses intracellular DNA recombination of barcoded open reading frame (ORFs) clones to identify interacting proteins, allowing Y2H positive colonies to be pooled and sequenced simultaneously. However, this technology still requires isolating and sequencing each barcoded bait and prey clone prior to screening in order to make barcode-ORF associations, which quickly becomes costly in large-scale screening. To more efficiently enable iterative screening, we developed CrY2H-seq (Cre reporter-mediated yeast two-hybrid coupled with next-generation sequencing). CrY2H-seq uses Cre recombinase as a Y2H protein-protein interaction reporter that functions intracellularly to covalently and unidirectionally link interacting bait and prey plasmids via specialized loxP sites that flank the protein-coding sequences. The linked protein-coding sequences serve as interaction-identifying DNA molecules that enable massively-multiplexed screening coupled with next-generation DNA sequencing to detect protein-protein interactions.

We applied CrY2H-seq to comprehensively screen a collection of 1,956 *Arabidopsis* transcription factors and regulators (hereafter collectively called TFs)[14] against itself in ten independent "all-by-all" screens. We report a deep coverage Arabidopsis transcription factor interactome composed of 8,577 binary interactions, 7,994 of which are novel. After experimentally and computationally validating interactions, we identified several network modules associated with plant reproductive development, root growth, environmental regulation of circadian rhythms, and stress- and hormone-response pathway crosstalk.

RESULTS

CrY2H-seq assay development

To establish CrY2H-seq, we first generated a yeast strain, CRY8930, that carries both a Gal4-inducible *GAL7::CRE* expression cassette in addition to two well-characterized *GAL1::HIS3* and *GAL2::ADE2* auxotrophic expression cassettes[1] (Fig. 1a). We then

modified a widely used ARS/CEN Gateway-compatible plasmid set[1,3] to contain unidirectional lox sequences[15] flanking the 3' end of ORF inserts, such that upon Cre recombination both ORF inserts would be on the same DNA molecule in a fixed orientation (Fig. 1b). By screening yeast transformants harboring known positive and negative interaction pairs in these modified plasmids (**Online Methods**), we confirmed that positive pairs induced Cre expression in addition to enabling growth selection (Supplementary Fig. 1). Yeast colony PCR with Gal4-AD and Gal4-DB primers (Fig. 1b and Supplementary Table 1) produced amplicons only for positive pairs, indicating plasmids underwent Cre-recombination (Supplementary Fig. 1b–c). Sanger sequencing of Cre-recombination PCR products verified that a newly formed double mutant lox site became sandwiched between the two ORF sequences, and recombination occurred in a fixed 3'-end to 3'-end fashion (Fig. 1c). Moreover, interactions gave the same either positive or negative result regardless of whether CRY8930 or the unmodified Y8930 was used (Supplementary Fig. 2).

There are two main distinctions between CrY2H-seq and existing multiplexed Y2H technologies[12,13]. First, interactions detected by CrY2H-seq require the parallel activation of two reporter genes driven by distinct promoters for detection of interactions; an auxotrophic rescue reporter and *CRE*. We used *HIS3* in conjunction with *CRE* because *GAL1:HIS3* is known to be more sensitive than *ADE2* for detecting interacting proteins[16], and the use of the independent *GAL7* promoter to drive *CRE* expression reduces promoter-specific false positives[17]. Furthermore, including *CRE* as a secondary reporter gene minimizes the time and reagents required of a steroid-inducible Cre expression system[12,13]. The second distinction is that CrY2H-seq uses interacting protein coding sequences themselves to form an intracellular DNA identifier (Fig. 1c) rather than barcode identifiers[12] that could become a bottleneck in large-scale screens. These key features allowed us to circumvent current Y2H limitations and establish a general CrY2H-seq pipeline for all-by-all massively-multiplexed screening (Fig. 2).

Deep interaction screening of an Arabidopsis TF ORFeome

We loaded a set of 1,956 Arabidopsis TFs[14] into the CrY2H-seq pipeline and performed ten all-by-all screens with final bait and prey libraries containing 1,877 and 1,933 unique yeast clones respectively (Supplementary Table 2a and **Online Methods**). These starting library populations showed an ORF size distribution consistent with the expected size distribution (Supplementary Fig. 3a), and the data showed minimal ORF size bias (Supplementary Fig. 3b–c). While bait proteins are typically screened for self-activation prior to Y2H screening, we chose to eliminate this step in order to rigorously challenge whether the assay would be able detect real interaction signal above the "noise" from self-activator interactions. Instead, to internally control for self-activating bait proteins[18], we spiked into each screen an excess amount of a Y8800 strain harboring an empty pADlox plasmid. Libraries were mated and underwent *HIS3* reporter selection ten independent times. This deep screening tested 3.6 million potential protein combinations approximately 300 times, for an estimated total of one billion interactions surveyed (**Online Methods**). After carrying out multi-template PCR amplification on plasmid pools isolated from each screen, we randomly sheared the PCR products to ~300 bp and generated standard Illumina-based DNA sequencing libraries (Fig. 2). We then performed 100 bp paired-end Illumina sequencing, aiming for a previously

established optimized coverage of 40 million reads per screen (Supplementary Fig. 4 and **Online Methods**). Paired-end reads were mapped and quality filtered, and fragments corresponding to Cre-recombined ORF junctions were extracted (Supplementary Fig. 5a–e and **Online Methods**). We applied a pre-determined basal fragment cutoff to eliminate any putative interactors that were represented by less than three junction fragments (Supplementary Fig. 5f and **Online Methods**). The remaining interaction-identifying fragments (**Online Methods**) were normalized across the ten independent screens to control for variation between sequencing runs (Supplementary Fig. 5g and **Online Methods**), and were classified as 'normalized protein interaction fragments' (NPIFs; Fig. 2). Very minimal amplicon size bias was observed in our dataset (Supplementary Fig. 3d–e), although fragments mapping to homodimers were notably absent from the data likely due to difficulty in amplification of the hairpin structure formed by fused identical ORFs as was previously observed in small scale experiments (Supplementary Fig. 6). In total, 10.9 million NPIFs were identified from the ten CrY2H-seq screens, mapping to 173,000 unique Cre-recombined ORF junctions (Fig. 3a). Among these were 299 different pDBlox ORFs fused to an empty pADlox vector, indicating that 16% of baits exhibited self-activation (Supplementary Table 3a). All 164,293 unique ORF combinations containing these TFs (Supplementary Table 3b) were excluded from the data. The remaining 1.4 million (13%) NPIFs mapped to 8,577 protein interactions, with a median of 7 NPIFs per interaction (Fig. 3b). The 8,577 interactions form the deep coverage interactome we refer to as "*Arabidopsis thaliana* transcription factor interaction network, version 1" (AtTFIN-1) (http://signal.salk.edu/interactome/AtTFIN-1.html, Supplementary Table 2b–c, **Online Methods**).

Validation of AtTFIN-1 Interactions

To estimate sampling sensitivity, the fraction of all identifiable interactions found in one screen[10], we simulated results for all possible orderings of replicate screens and found that one screen alone on average yielded 2012 ± 354 interactions (mean $\pm$ standard deviation). Calculating the average number of new interactions gained after each of the ten screens (Fig. 3c) revealed that even after ten screens, saturation was not reached. We fit this data to a Michaelis-Menton modeled curve to estimate the degree of saturation and determined that of the $15,610 \pm 2,661$ interactions that could have been maximally detected (Supplementary Fig. 7, **Online Methods**), we detected more than half (54.6%).

To estimate reproducibility, we retested 771 (9%) AtTFIN-1 interactions (678 of which were novel) that showed a range of NPIFs and screen occurrences (Supplementary Table 4) using a standard pairwise 1×1 array style Y2H screen[18] (Supplementary Fig. 8a). Excluding *de novo* self-activating baits identified by parallel plating on cycloheximide selection media[18], we observed an overall retest rate of 73% (422/580 novel interactions and 57/76 'known' interactions, defined below). Additionally, we observed an increased retest rate for interactions appearing in multiple screens (Fig. 4a), but a relatively similar retest rate among interactions showing different ranges of NPIFs (Supplementary Fig. 8b). We also tested 94 AtTFIN-1 interactions (59 of which were novel) (Supplementary Table 5a) using the wNAPPA assay[19] and observed that 50% of all AtTFIN-1 interactions and 25.4% of novel AtTFIN-1 interactions tested positive (Fig. 4b, Supplementary Fig. 9). These rates

contrasted significantly with the 2.8% positive rate observed for 36 random TF interactions tested in wNAPPA.

To estimate assay sensitivity, the fraction of all detectable biophysical interactions[10], we mined both literature[3] and databases[4–6] for TF interactions that were screened in CrY2H-seq (Supplementary Table 2b). We refer to these mined interactions collectively as 'known' interactions. Interactions involving self-activating TFs and homodimers were excluded from this analysis. AtTFIN-1 showed the greatest overlap (52.2%) with Arabidopsis Interactome-1 interactions[3] and the least overlap with AraNet[6] interactions (13.4%) (Fig. 4c). We estimated a false positive rate of 0.69% ± 0.12% (mean ± standard deviation), by calculating the overlap of AtTFIN-1 interactions with ten different datasets, each composed of 8,577 randomly generated TF interactions (**Online Methods**). Overall, AtTFIN-1 interactions showed significantly greater recapitulation of known interactions, including those derived from a variety of assays (Supplementary Fig. 10a), relative to random interactions (Fig. 4c). A precision-recall curve of these detection rates plotted as a function of the number of screen occurrences, showed a large drop in precision with little gain in recall between one and two screens, leading us to classify high-confidence interactions as those identified in two or more screens (Fig. 4d).

To measure performance improvements over array-based high throughput Y2H (HT-Y2H), we compared TF interaction detection rates between CrY2H-seq and HT-Y2H used to generate the Arabidopsis-Interactome-1[3]. CrY2H-seq showed a five-fold increase in general TF interaction detection relative to HT-Y2H (Supplementary Fig. 11a). Of the commonly screened TF interactions, CrY2H-seq showed a seven-fold increase in detection, recovering 1,609 TF interactions whereas HT-Y2H detected only 229 (Supplementary Fig. 11b). Of the commonly tested literature curated interaction (LCI) pairs[3], CrY2H-seq recalled 33.3% while HT-Y2H recalled only 12.3% (Supplementary Fig. 11c). While CrY2H-seq showed a clear overall improvement to HT-Y2H, it should be noted that the Arabidopsis Interactome-1 was based on the union of two primary screens and was filtered by pairwise retesting, where AtTFIN-1 was based on ten primary screens that were not filtered by pairwise retesting.

To evaluate the biological relevance of AtTFIN-1 interactions, we compared expression correlations between AtTFIN-1 interactions and a random interaction dataset using 6,057 different expression datasets[20]. We observed significantly higher expression correlation for transcripts encoding AtTFIN-1 interactions than for transcripts encoding random interactions (Supplementary Fig. 12), supporting their potential to interact *in vivo*.

AtTFIN-1 defines expanded transcription factor modules

We further investigated the biological significance of the 3,086 high-confidence AtTFIN-1 interactions (2,578 novel) by looking for 'preferential' intra- and interfamily interactions that occurred more frequently than would be expected by chance. AtTFIN-1 interactions classified by previously assigned families[14] were compared to those in 10,000 randomly rewired degree-conserved networks (Fig. 5a, Supplementary Fig. 13 and **Online Methods**). We observed highly significant preferential intrafamily interactions among family members known to dimerize including the bHLH, MADS, bZIP, NAC, WRKY, AUX-IAAs, and ARF families. We also observed highly significant preferential interfamily interactions between

plant-specific families known to dimerize including Growth Regulating-Factors (GRFs) and Growth Regulating-Factor Interacting Factors (GIFs)[21], LUGs and YABBYs[22], and AUX-IAAs and ARFs[23]. The TCP family (Teosinte-branched/Cycloidea/Proliferating Cell Factor) showed significant preference for 18 TF families (Supplementary Fig. 13) consistent with previously observations of TCPs as 'hub' proteins[3,24].

We further examined highly significant, unknown preferential interfamily interactions, and found the preference of the ABI3-VP1/B3 family for GeBP and TRIHELIX proteins was driven by one ABI3-VP1/B3 member, AT5G60142, that showed many interactions with various TRIHELIX and GeBP members (Fig. 5b). While the GeBP and TRIHELIX members have sparse GO annotations, AT5G60142 has recently been found up-regulated in isolated early stage gynoecium medial domain cells[25]. Interestingly, not only were AT5G60142 and 93% (13/14) of its TRIHELIX and GeBP interacting partners found co-expressed in this study, but five of AT5G60142's partners (ASIL2, AT3G58630, AT1G76870, AT3G04930, and STKL1) were significantly up-regulated in cells from the same distinct domain. These interactions may form part of a previously unrecognized module underlying early stage reproductive development. We also found the preference of G2-like proteins for the GRAS family was driven by multiple phosphate response-like factors and the scarecrow-like factors (Fig. 5c). This network reveals a logical link between phosphate sensing and root development, consistent with the notion that phosphate deprivation drives altered root architecture and increased root hair density[26,27]. C2C2-CO-like TFs showed significant preferential interaction with the "orphans" category of unassigned TFs (Fig. 5d). Closer examination of these interactions revealed that all proteins contained BBX domains, including the C2C2-CO-like proteins themselves. These interactions could be mediated by BBX domains as these have been shown to be crucial in mediating protein-protein interactions and transcriptional regulation[28]. Many BBX domain-containing proteins are known to have specific and sometimes opposing functions in regulating flowering, circadian clock, biotic or abiotic stress response[28]. Moreover, it was recently reported that overexpressing AtBBX32 in soybean plants increased grain yield by altering light input and expression patterns of clock genes necessary for initiation of different stages of reproductive development[29]. This AtTFIN-1 module suggests that combinatorial complexity among BBX proteins may play a role in integrating environmental signals and flowering time potentially through feedback or feed-forward loops.

Beyond the well-characterized interfamily interaction between ARFs and AUX/IAAs[23], for which we observed a significant preferential family interaction between eight ARF members and 23 AUX/IAA members, individual AUX-IAA members very interestingly showed distinct interactions with other families (Fig. 6). For instance, IAA17 heavily interacted with TCPs compared to other IAAs, suggesting IAA17 could be the main player mediating crosstalk between auxin and TCP transcriptional regulation. IAA2, 10, 17, and 18 commonly interacted with MBD (Methyl-CpG binding domain) proteins indicating their potential involvement in regulating DNA methylation. Particular IAAs and ARFs showed interactions with specific hormone and stress associated TFs: IAA11 with hormone/abiotic stress response factors ERF70 and DRIP2, IAA10 with defense response factors LOL2 and GEBP, and ARF18 with abscisic acid response factors VAL1 and VAL2, indicating their potential roles in integrating auxin response with different hormone and stress signals. This

expanded ARF-AUX-IAA interactome reveals how particular TFs may play specific roles in mediating cross-talk between auxin response and other plant pathways.

DISCUSSION

CrY2H-seq offers an untargeted, highly scalable screening approach to directly assay binary protein-protein interactions in yeast. We demonstrated that nearly four million interactions could be assayed to >50% saturation with ten cost-effective and time efficient CrY2H-seq replicate screens (Supplementary Fig. 14), a scale which has not been achievable in the past. The increased interaction detection rates and significantly greater overlap with previously reported interactions (Fig. 4c, Supplementary Fig. 10a, Supplementary Fig. 11) suggest CrY2H-seq could increase overlap between inter-laboratory Y2H screens[30]. We attribute these increases to using next-generation sequencing for interaction detection and the ease of iterative screening. Moreover, the reported CrY2H-seq sensitivity may even be an underestimate, and removal of self-activating proteins prior to screening could lead to the detection of missed interactions. Nonetheless, our CrY2H-seq screening was not exhaustive nor did it completely capture all known interactions, alluding to inherent yeast two-hybrid limitations including sub-optimal protein expression levels or strain copy number in pools. CrY2H-seq could be further optimized to reduce sequencing costs by applying strategies for targeted capture of fused lox-containing DNA fragments and depletion of over-abundant DNA from sequencing libraries. Additionally, the incorporation of a unique DNA sequence into the lox region on one of the CrY2H-seq plasmids could disrupt the hairpin structure to allow the potential detection of homodimers and optimized tracking of bait/prey orientations.

The widely applicable resource, AtTFIN-1, generated from CrY2H-seq screening substantially expands the available interaction data among Arabidopsis TFs, tripling the 3,170 interactions documented in BioGRID[4]. The novel interactions we identified reveal potential involvement of poorly annotated TFs in various biological processes including root and reproductive development, and the integration of environmental stimulus with circadian rhythms. These data can be used for future genomic analyses and data integration pipelines to further define these network modules and help identify candidate genes that could be used for crop improvement. This expanded TF network can be used to generate hypotheses regarding the specific roles of individual TFs or TF families throughout development and in response to a multitude of biotic and abiotic stressors. For instance, the activity of AtTFIN-1 interactions could be tested on different promoters to examine how interactions affect target gene expression[31]. Further understanding the roles of TF interaction partners in combinatorial gene regulation is particularly valuable for improving crop optimization strategies that currently target individual TFs[32].

Lastly, CrY2H-seq technology could be applied to Y2H assay variations. For instance, CrY2H-seq could be adapted to the split-ubiquitin system[33] for screening hydrophobic proteins, or to yeast one-hybrid for screening genome-wide protein-DNA interactions[34]. The ease of setting up CrY2H-seq replicate experiments permits screening on multiple media types for selection of different reporter genes, or selection on media supplemented with various hormones that may influence interactions[35]. Furthermore, while we used an array

cloning strategy[18] here for mobilizing ORFs into CrY2H-seq plasmids, *en masse* cloning strategies[36,37] can be used to reduce cost and importantly extend the application of CrY2H-seq to cDNA library-against-cDNA library screening. This would enable comparisons of unprecedentedly large-scale interactomes derived from different ecotypes, growth conditions, or tissue types, and identification of network differences underlying different phenotypes. Interaction maps generated by CrY2H-seq could be integrated with other 'omics' data to provide deeper insight into the functional relationships between genotype and phenotype, the network effects of variants, and interactome modules that certain transcriptional programs give rise to.

ONLINE METHODS

Strain and plasmid construction

Primers used to modify plasmids and the CRY8930 strain are listed in Supplementary Table 1. The genotype of CRY8930 is MATα *leu2-3,112 trp1-901 his3-200 ura3-52 gal4 gal80 P_{GAL2}-ADE2 LYS2::P_{GAL1}-HIS3 MET2::P_{GAL7}-CRE-HPHMX6 cyh2[R]*. The genotype of Y8800 is MATa *leu2-3,112 trp1-901 his3-200 ura3-52 gal4 gal80 P_{GAL2}-ADE2 LYS2::P_{GAL1}-HIS3 MET2::P_{GAL7}-lacZ cyh2[R]*. Y8800 and CRY8930 strain stocks, and pADlox and pDBlox plasmid stocks, have been deposited with the *Arabidopsis* Biological Resource Center (https://abrc.osu.edu/).

Cre reporter strain construction—The bacteriophage P1 Cre recombinase gene[38] was PCR amplified from pQTL123 GST-Cre with flanking SalI and PacI sites and ligated into SalI/PacI digested pFA6α-HPHMX6. The Cre-hygromycin resistance marker cassette was PCR amplified from the resulting plasmid and used in a homologous recombination reaction to replace the *LacZ* reporter gene within the *GAL7::LacZ* cassette integrated at the *MET2* locus of yeast strain Y8930[1]. Correct integration of *CRE* in the modified strain, referred to as CRY8930, was confirmed by sequencing of the *MET2* locus. To test *CRE* reporter gene expression, RNA was extracted from a histidine positive diploid culture containing the known interaction pair AD-bZIP53 and DB-bZIP63[39] using the Qiagen RNeasy kit. Reverse transcription was carried out on DNAse treated RNA extract using SuperScript II (Life Technologies) followed by PCR to detect the presence of Cre cDNA (Supplementary Fig. 1a, primers listed in Supplementary Table 1).

Construction of lox site-containing bait and prey destination vectors—Lox71 and lox66 sequences[40] were inserted into the XmaI and AatII sites located downstream of the attB2 site in pDEST-AD[1] and pDEST-DB[1] respectively, using standard cloning methods. The resulting destination vectors, pADlox and pDBlox, were Sanger sequenced confirmed and transformed into One Shot ccdB Survival 2 T1R competent cells (Life Technologies). Lox71 and lox66 sites are modified versions of the standard loxP sites that display favorable forward recombination reaction equilibrium[13,15].

Pilot tests for Cre-lox recombination functionality in yeast two-hybrid—Small-scale tests were conducted to confirm the functionality of the CrY2H-seq system in yeast (Supplementary Fig. 1b–c, Supplementary Fig. 2, and Supplementary Fig. 6). In brief, ORFs

were Gateway™ cloned into pADlox and pDBlox using LR clonase and transformed into DH5α chemically competent cells. pAD-ORF-lox and pDB-ORF-lox plasmids were purified using a QIAprep Spin Miniprep kit (Qiagen) and transformed into yeast strains Y8800 and CRY8930 respectively, using a standard lithium acetate method. ORFs were also transformed into the Y8930 parental strain to serve as negative controls. Strains were mated according to published protocols[18], and grown for 3 days on interaction selection media (-Leu/-Trp/-His + 1mM 3-Amino-1,2,4-Triazole (3-AT)). For Supplementary Fig. 1, The known positive interaction pair bZIP53/bZIP63[39] and non-interacting pairs bZIP53/ZTL, and ZTL/bZIP63 were tested. Mated strains were also grown in parallel on diploid selection media (-Leu/-Trp). Colonies were then picked from all plates, and in the case of the non-interacting pair on interaction selection media all background cells were scraped. Lysates were prepared as described previously[18], and PCR using AD and DB primers (Supplementary Table 1) was performed to detect Cre recombination products. For Supplementary Fig. 2, prior to plating diploids on selection media, culture concentrations (OD_{600}) were measured on a Tecan Safire2 plate reader (Supplementary Fig. 2b). CRY8930/Y8800 diploids were plated adjacent to Y8930/Y8800 diploids to assess strain differences (Supplementary Fig. 2c). For Supplementary Fig. 6, *HIS3* positive colonies were picked, lysates prepared as described previously[18], and PCR using AD and DB primers (Supplementary Table 1) was performed to detect Cre recombination products. All PCR reactions were prepared with 1 μL of template, 0.1 μL Phusion Polymerase (NEB), 2 μL 5xGC buffer (NEB), 2 μL 5 M betaine, 200 μM each dNTP, and 0.25 uM of AD and DB primers (Supplemental Table 1). Reactions were run at 98°C for 2 minutes, 30 cycles of 98°C for 10 seconds, 60°C for 30 seconds, and 72°C for 90 seconds, then a final extension at 72°C for 7 minutes. 5 μL of each PCR reaction was run on a 1% agarose 1x TAE gel.

Transcription factor library construction for CrY2H-seq screening

All cloning and transformations were carried out according to published protocols[18]. Briefly, 1,956 entry clones from an Arabidopsis transcription factor ORF collection[14] were individually Gateway™ cloned in 96-well format into both pADlox and pDBlox vectors using LR clonase and transformed into chemically competent DH5α-T1[R] cells. Transformants were selected in Terrific Broth containing carbenicillin, and plasmid DNA was extracted and purified using QiaPrep 96 turbo kits (Qiagen). Next, pADlox TF plasmids and pDBlox TF plasmids were individually transformed into the yeast strains Y8800 and CRY8930 respectively using a 96-well lithium acetate transformation protocol[18] as follows: Plasmid DNA and yeast competent cells were combined, 96-well plates were incubated at 42°C for 1 hour, cells were centrifuged, washed, spotted on SC –Trp (pADlox clones) or SC –Leu (pDBlox clones), and grown at 30°C for three days. Colonies were then picked and inoculated into liquid SC –Trp or -Leu, and cultures were grown for three days at 30°C at 200 rpm to reach saturation. Equal volumes of cells from individual TF clones were pooled to make the CrY2H-seq libraries for mating. Aliquots of 1 mL containing ~3 OD_{600} were mixed with 500 μL of 50% glycerol and stored at −80°C. Additionally, 96-well glycerol stocks of individual TF clones were also made for archival storage purposes.

Characterizing starting bait and prey libraries—Plasmid DNA was purified from a 1 mL aliquot of each library, from which ORF DNA was PCR amplified with either AD or DB

primer and a primer that anneals to a common sequence downstream ORF inserts (Supplementary Table 1). An Illumina sequencing library was then prepared from each starting library by fragmenting ORF amplicons to 300 bp with a Covaris S2 sonicator, end-repairing fragments with the End-It DNA End-Repair Kit (Epicentre-Illumina), A-tailing repaired fragments with Klenow 3'-5'exo-(NEB), and ligating Illumina Truseq adapters to fragments using T4 ligase (NEB) overnight at 16°C. The adapter ligated libraries were then run on a 2% agarose gel and a 400–600 bp region was excised and purified using a QIAquick gel extraction kit (Qiagen). Purified DNA was then amplified with Phusion Polymerase supplemented with 1 M betaine and Illumina Truseq primers for three cycles using Illumina recommended conditions. A final purification with SeraMag Speedbeads (GE; 2% v./v. SeraMag Speedbeads, 18% w./v. PEG-8000, 1M NaCl, 10mM Tris HCl, 1mM EDTA) at a 1:1 bead to DNA ratio was performed to remove unincorporated Truseq primers, and libraries were sequenced on an Illumina paired-end 200 cycle Rapid Run on an Illumina HiSeq 2500 platform. Each library was sequenced to ~$1000\times$ coverage (bait library, 3.7M reads; prey library, 2.3M reads; equivalent to 1.7% of a Rapid Run flowcell). Reads were analyzed following the next generation sequencing analysis pipeline detailed below with the following difference: paired reads for which each of the mates aligned to the same ORF and showed different strand orientation underwent a size filter that required that the difference of the start position of one read and the end position of the read pair fall within the expected library size of 400–600bp. After this filtering, ORF-mapped fragments were totaled and libraries were further characterized by plotting the size distribution and representation of detected ORFs (Supplementary Fig. 3a–c). A total of 1,933 and 1,877 unique AD and DB clones respectively were identified, giving rise to ~3.6 million possible combinations.

CrY2H-seq screening of transcription factor libraries

Each replicate screen consisted of mating ~20 OD_{600} of each TF clone library (pADlox in Y8800 and pDBlox in CRY8930). Based on cell titers of 2×10^7 cells/OD that we observed for each library, we estimated that each replicate screen would test the ~3.6 million possible protein combinations at 10-fold excess, assuming a 10% mating efficiency.

Frozen aliquots of the 1,933 TF pADlox library and the 1,877 TF pDBlox library were thawed, separately inoculated into 200 mL of YEPD media, and grown for 1 hour at 30°C and 150 rpm prior to mating. Cell concentrations were measured and libraries were combined such that each replicate screen contained ~20 OD_{600} of each CrY2H-seq library. To internally test for self-activating proteins, a pADlox empty plasmid in the Y8800 strain was spiked into each replicate mating batch in at least three-fold excess of the average individual clone population (~2×10^5 cells/clone). For each replicate, mating in liquid YEPD was carried out at 30°C for 4.5 hours with shaking at 50 rpm. Subsequently, a 10 μL aliquot of the mated culture was diluted and plated on -Leu, -Trp, and -Leu/-Trp media to determine mating efficiency, which was on average 6% with ~1.25×10^8 diploids formed per screen. Assuming all combinations of proteins were equally represented among the diploid population, we estimate that each possible combination was sampled ~$34\times$ in each screen (1.25×10^8 diploids/3.63×10^6 total protein combinations).

The remainder of the mated cultures were washed with 1× SC and individually resuspended in 100 mL 1× SC –Leu/-Trp supplemented with 125 μg/mL hygromycin to enrich for diploids and reduce background growth. These cultures were grown at 30°C overnight shaking at 150 rpm. Diploid cells for each screen were then collected, washed with 1× SC, and resuspended in water at 1 OD_{600} per mL. Cells were plated at roughly 0.5 OD per plate on SC–Leu –Trp –His +1mM 3-AT plates (~48 plates per screen) and grown for three days at 30°C to select for interactors. 48 plates, each containing more than 10,000 colony forming units, were individually scraped into 48 wells of a 96-well deepwell plate. Cells were heated at 75°C for 20 minutes to inactivate Cre recombinase. Cells were next treated with 300 μL zymolyase buffer (0.1 M sodium phosphate buffer pH 7.4, 1% betamercaptethanol, and 2.5 mg/mL Zymolyase 20T (US Biological), and 100 μg/mL RNase A (Qiagen) and incubated at 37°C for 1 hour at 50 rpm. Zymolyase-treated cell suspensions were split into two wells of a 96-well deepwell plate, and plasmid DNA was prepared following the QiaPrep 96 turbo miniprep kit protocol and recommendations for purifying low-copy plasmids. DNA concentrations were measured using the dsDNA Quantifluor System (Promega) and ~5–10 nanograms from each well was used to PCR amplify Cre recombined ORF pairs using Phusion Polymerase (NEB), 1xGC buffer (NEB), 1 M betaine, 200 μM each dNTP, and 0.25 uM of AD and DB primers (Supplemental Table 1). Reactions were run at 98°C for 2 minutes, 21 cycles of 98°C for 10 seconds, 65°C for 30 seconds, and 72°C for 90 seconds, then a final extension at 72°C for 7 minutes. 5 μL of each PCR reaction was run on a 1% agarose gel and showed a DNA smear corresponding to the size range expected for Cre recombined products (~1 kb to > 4 kb). Amplicons from each PCR reaction were pooled, isopropanol precipitated, and purified with SeraMag Speedbeads (GE; 2% v./v. SeraMag Speedbeads, 18% w./v. PEG-8000, 1M NaCl, 10mM Tris HCl, 1mM EDTA) at a 1:1 bead to DNA ratio to remove primers, typically yielding ~2 μg of DNA. Illumina sequencing libraries were then prepared following the exact same steps as previously mentioned for the starting bait and prey libraries.

Pilot sequencing test to determine optimal sequencing depth—The same sequencing library from one CrY2H-seq screen was sequenced to a read depth of 20 million (20M) and 80 million (80M) reads. We observed that interactions with at least three distinct identifying fragments in 20M showed an expected increase in coverage of about 4× at 80M, while those with less than 3 fragments in 20M were not consistently reproducible (Supplementary Fig. 4). We therefore established a cutoff requiring at least 3 fragments for a PPI to be included in a screen dataset. Moreover, since deeper sequencing predominantly revealed PPIs represented by less than 3 fragments (i.e. below our cutoff), we concluded that 20 million reads was sufficient and aimed for 40 million reads per screen library.

Sequencing of CrY2H-seq screen libraries—Libraries were sequenced with an Illumina paired-end 200 cycle Rapid Run on an Illumina HiSeq 2500 platform. The total paired reads obtained from sequencing was 583M equivalent to 1.65 Rapid Run flowcells.

Next-generation sequence analysis of CrY2H-seq screen libraries

Reads were mapped using Bowtie2-2.0.2[41] local alignment with default settings to a custom genome composed of *Arabidopsis* TF coding sequences from TAIR10, the *Saccharomyces*

cerevisiae genome, Gal4 AD and Gal4 DB domain sequences, and the empty CrY2H-seq plasmid sequences (Supplementary Fig. 5a). A quality filter was applied requiring reads to map with at least 30 matching bases, allowing a maximum of 2 mismatches, 2 insertions or deletions, and 2 bases of trimming from the beginning of the read (Supplementary Fig. 5b). Reads were then joined with their corresponding read pairs and included in the next analysis step only if both reads passed the first filter and mapped to Arabidopsis TF ORF sequences. Clonal fragments were removed from read pairs if both reads in a fragment contained the same start positions. Paired reads for which each of the mates aligned to a different ORF and showed the same strand orientation (Cre recombination occurs such that ORFs on pADlox and pDBlox plasmids become inverted in a 3'-to-3' orientation, Supplementary Fig. 5c) were included in further analysis. Fragments were further subjected to a size filter that required that the sum of the lengths of each read (start position of each read to the end of each ORF) and the lox region conformed to the expected library size of 400–600bp (Supplementary Fig. 5c). Remaining fragments that mapped to Cre-recombined ORF junctions were totaled (Supplementary Fig. 5e). Each screen had on average ~1.4 million fragments corresponding to ORF junction sites and ~16 million fragments mapping to gene bodies. Remaining data mapped to priming site region ORF junctions or did not align. Analysis scripts can be found in Supplementary Software. After applying the basal fragment cutoff mentioned above to all data sets (Supplementary Fig. 5f), fragments were normalized by the median filtered fragments as follows: A scale factor for each replicate dataset was determined by dividing the filtered protein interaction fragments by the median filtered protein interaction fragments. The number of fragments per protein pair was multiplied by this scale factor and rounded down to the nearest integer to normalize protein interaction fragments (Supplementary Fig. 5g).

Identification and removal of self-activating bait proteins—Any TF found to be linked with an empty pADlox plasmid by the mapping pipeline was labeled self-activating and not included in AtTFIN-1. A list of proteins identified as self-activating can be found in Supplementary Table 3.

Bait and prey orientation analysis of AtTFIN-1 interaction fragments—As the double mutant lox sequence from Cre-recombined plasmids is not a full palindrome, the middle region can be used to determine bait and prey orientations of interacting proteins (Supplementary Fig. 15a). An analysis script was written to assess the bases at this middle region for fragments where at least one read mapped to one ORF and 15 base pairs into the lox77 sequence (Supplementary Software). It should be noted that the region of the read being mapped to lox77 was within the last 10bp of the read where sequencing quality is known to be low due to the nature of sequencing by synthesis. Of fragments mapping to non-self-activating PPIs, 5.5% (9662266/14588892) could identify bait and prey orientations of 49.71% of (4264/8577) AtTFIN-1 pairs (Supplementary Fig. 15a–b, Supplementary Table 2c). We acknowledge that this is a partial analysis and more data would be needed to confirm the bait and prey orientations for all pairs in AtTFIN-1.

Estimating CrY2H-seq screen saturation

To estimate CrY2H-seq screening saturation (the number of interactions detected out of the number of interactions CrY2H-seq could detect for this ORF collection), we simulated results for all possible orderings (10!) for the 10 replicate screens. We calculated the average number and standard deviation of interactions detected at each step, considering all possible orderings (Fig. 3c). We built a model based on the average new interaction detection rate after each replicate, and fit it to a Michaelis-Menten curve to predict the number of interactions detectable by CrY2H-seq after any number of screens (Supplementary Software, Supplementary Fig. 7).

Yeast two-hybrid retest

A set of 950 interaction pairs that showed a range of screen occurrences and NPIFs was selected for use in a retest assay carried out using standard 1×1 array-style HT-Y2H methods. Clones corresponding to interaction pairs were cherry picked from pAD-lox and pDB-lox plasmid stock plates and freshly transformed into yeast strains Y8800 and CRY8930 as described above. 771 yeast transformant pairs were recovered that could be screened in both bait and prey orientations (Supplementary Table 4). This ensured that both orientations in which the interaction could have been initially detected were accounted for. A Y2H screening pipeline was followed as described previously[18], including inoculation of individual AD and DB yeast cultures, 1×1 mating onto YEPD medium, replica-plating onto selective SC –Leu, -Trp for diploid selection, and replica-plating onto selective SC –Leu, -Trp, -His +1mM 3-AT plates and SC -Leu, -His +1mM 3-AT plates containing 1mg/L cycloheximide. Cycloheximide containing plates select for cells that do not have the AD plasmid due to plasmid shuffling and can identify spontaneous self-activators[18]. After replica-plating onto SC –Leu, -Trp, -His +1mM 3-AT, plates were incubated at 30°C overnight, then replica-cleaned by placing each plate on a piece of velvet stretched over a replica-plating block and pressing evenly to remove excess yeast cells. Plates were incubated an additional three days at 30°C and phenotypes were independently scored by two researchers (for representative colonies and scoring, refer to Supplementary Fig. 8a). Only pairs scored as positive for *HIS3* reporter gene activation and negative for growth on cycloheximide by both researchers were considered positive interactions in the retest assay. 115 pairs (~15%) activated the *HIS3* reporter gene and showed growth on cycloheximide. These interactions were scored as self-activating and not included in subsequent analysis of the retest dataset.

wNAPPA assay

TFs corresponding to 59 novel interactions that showed a range of screen occurrences and NPIFs were selected for validation in the wNAPPA assay. Additionally, 35 previously reported protein interactions that were present in At-TFIN-1 and 36 random interactions not present in AtTFIN-1 were also processed in parallel. Clones were cherry picked from TF entry clone stock plates and recombined into pIX-GST and pIX-HA destination vectors[3] using LR clonase. Reactions were transformed into DH5α-T1^R and plasmid DNA was purified using QiaPrep 96 Turbo kits. Plasmid DNA was measured using the Quantifluor dsDNA System and a Tecan SafireII plate reader. DNA was concentrated to roughly 250

ng/μL and 1 μg of each plasmid was combined for use *in vitro* transcription/translation reactions as follows. Bait and prey proteins were co-expressed using the TNT SP6 Coupled Wheat Germ Extract System (Promega) following manufacturer recommendations. Protein expression reactions were then added to anti-GST antibody-coated detection plates (GE Healthcare) and incubated at 15°C for 2 hours. Wells were washed and blocked with 1× PBS with 0.1% Tween and 5% non-fat dry milk (PBS/T/NFM) for 1 hour at room temperature, then incubated with mouse anti-HA monoclonal antibody (Covance) diluted 1:5000 in PBS/T/NFM for 1 hour at room temperature. Antibody was washed from wells with PBS/T/NFM with three quick washes followed by three longer washes each with a five-minute room temperature incubation period with gentle rotation. Wells were then incubated with anti-mouse HRP-coupled secondary antibody (GE Healthcare) diluted 1:2000 in PBS/T/NFM for 1 hour at room temperature. Secondary antibody was washed from the wells with PBS/T with three quick washes followed by three 5-minute washes. Wells were rinsed twice with 1× PBS before adding Supersignal ELISA Femto substrate (Pierce), and then incubated for 1.5 minutes at room temperature with gentle shaking. Luminescence (RLU) was measured using a Tecan SafireII plate reader. Interactions were tested in both vector combinations and observed z-scores are listed in Supplementary Table 5a.

To control for plate-to-plate variation, a set of 16 pairs previously used for normalization[3] (Supplementary Table 5b) was included on each plate. Plate normalization and scoring were done according to previously described methods[3]. Briefly, for each plate the normalization pair average and standard deviation was calculated after subtracting the average blank (empty pIX GST and empty HA plasmid mix) and taking the $\log_2$ RLU value. A z-score for each well was then calculated by first subtracting the normalization pair average from the RLU value and then dividing by the normalization pair standard deviation. To determine the recall rates, the maximum z-score of the two orientations tested for each pair was considered and a scoring threshold was determined by maximizing for the number of positively scoring known interactions and minimizing for the number of positively scoring random interactions (Supplementary Fig. 9). A scoring threshold of 1.6 was selected based on these criteria.

Literature, database, and randomly generated data comparison with AtTFIN-1

Literature and database interaction data files were downloaded from links listed in Supplementary Table 6, and all interactions between TFs screened in CrY2H-seq were compiled. Interactions from different sources showed some overlap, but also many unique interactions (Supplementary Fig. 10b). For this reason, comparisons were made between AtTFIN-1 and individual datasets (Fig. 4c). Only high confidence STRING and AraNet interactions with scores above 900 and 4.5 were used. To generate random TF interactions, a list of all possible combinations was first generated. From this list, 8,577 interactions were selected randomly using the script in Supplementary Software. This step was done a total of 10 times to produce 10 random interaction datasets. From each of these datasets, we excluded homodimers and interactions with TFs detected as self-activating in the CrY2H-seq screens. Comparisons between AtTFIN-1 and each list were performed and the average overlap was reported (Fig. 4c). Supplementary Fig. 10b was generated using the web interface provided by VIB/University of Ghent Bioinformatics and Evolutionary Genomics

Division, Belgium (http://bioinformatics.psb.ugent.be/webtools/Venn/). The precision-recall curve (Fig. 4d) was generated using the R package PRROC[42].

Preferential family-specific interaction analysis

The R package igraph[43] was used to generate randomly rewired interactions from a list of high confidence AtTFIN-1 interactions using the rewire function with degree conservation. The gene IDs in the subsequent list of random interactions were converted into family names, sorted and family interactions were counted. This was done 10,000 times. The high confidence AtTFIN-1 interactions were similarly converted to family names and family interactions were counted. The AtTFIN-1 family interaction observations were then compared to the 10,000 random observations and P values were calculated based on where the AtTFIN-1 family interaction observation occurred in the empirical distribution of all observations for each family interaction. Heatmaps (Fig. 5 and Supplementary Fig. 13) were generated using the R package, Heatmap3[44]. Interaction networks (Fig. 5 and 6) were generated using Cytoscape[45].

Cost and time comparisons to existing HT-Y2H methods

Traditional Y2H and BFG-Y2H cost approximations (Supplementary Fig. 14) are based on appendix figure S4 in Yachie N. *et al.* (2016) *Mol. Syst. Biol*[12]. Costs for traditional Y2H were calculated on a per plate basis assuming minipools of 50 preys, and assuming the recovery of 500, and 10,000 positive interactions from 1,000,000, and 900,000,000 PPIs screened, respectively. CrY2H-seq sequencing costs are estimated from 1 Illumina HiSeq Rapid PE Sequencing Run (cluster kit and 200 cycle kit) costing $3126, and yielding on average 350,000,000 reads.

Statistics

Exact n values are reported in main text and legends for Fig. 4a–c, and Supplementary Fig. 8b, 11, and 12. For Fig. 4b–c and Supplementary Fig. 12, a one-sided Fisher's exact test was done to compare the detection rates of known and novel interactions to random interactions. For Fig. 5a and Supplementary Fig. 13, empirical P values were calculated by ranking the observed family interaction frequency among frequencies generated from 10,000 different degree conserved network re-wirings.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

# References

1. Yu H, et al. Next-generation sequencing to generate interactome datasets. Nat. Methods. 2011; 8:478–480. [PubMed: 21516116]

2. Rolland T, et al. A proteome-scale map of the human interactome network. Cell. 2014; 159:1212–1226. [PubMed: 25416956]

3. Arabidopsis Interactome Mapping Consortium. Evidence for Network Evolution in an Arabidopsis Interactome Map. Science (80-.). 2011; 333:601–607.

4. Stark C, et al. BioGRID: a general repository for interaction datasets. Nucleic Acids Res. 2006; 34:D535–9. [PubMed: 16381927]

5. Szklarczyk D, et al. STRING v10: Protein-protein interaction networks, integrated over the tree of life. Nucleic Acids Res. 2015; 43:D447–D452. [PubMed: 25352553]

6. Lee T, et al. AraNet v2: An improved database of co-functional gene networks for the study of Arabidopsis thaliana and 27 other nonmodel plant species. Nucleic Acids Res. 2015; 43:D996–D1002. [PubMed: 25355510]

7. Wang X, et al. Three-dimensional reconstruction of protein networks provides insight into human genetic disease. Nat. Biotechnol. 2012; 30:159–64. [PubMed: 22252508]

8. Hofree M, Shen JP, Carter H, Gross A, Ideker T. Network-based stratification of tumor mutations. Nat. Methods. 2013; 10:1108–1115. [PubMed: 24037242]

9. Jiang Z, Dong X, Zhang Z. Network-Based Comparative Analysis of Arabidopsis Immune Responses to Golovinomyces orontii and Botrytis cinerea Infections. Scientific reports. 2016; 6:19149. [PubMed: 26750561]

10. Venkatesan K, et al. An empirical framework for binary interactome mapping. Nat. Methods. 2009; 6:83–90. [PubMed: 19060904]

11. Weimann M, et al. A Y2H-seq approach defines the human protein methyltransferase interactome. Nat. Methods. 2013; 10:339–42. [PubMed: 23455924]

12. Yachie N, et al. Pooled-matrix protein interaction screens using Barcode Fusion Genetics. Mol. Syst. Biol. 2016; 12:863. [PubMed: 27107012]

13. Hastie AR, Pruitt SC. Yeast two-hybrid interaction partner screening through in vivo Cre-mediated Binary Interaction Tag generation. Nucleic Acids Res. 2007; 35

14. Pruneda-Paz JL, et al. A Genome-Scale Resource for the Functional Characterization of Arabidopsis Transcription Factors. Cell Rep. 2014; 8:622–632. [PubMed: 25043187]

15. Oberdoerffer P, Otipoby KL, Maruyama M, Rajewsky K. Unidirectional Cre-mediated genetic inversion in mice using the mutant loxP pair lox66/lox71. Nucleic Acids Res. 2003; 31:e140. [PubMed: 14602933]

16. Stynen B, Tournu H, Tavernier J, Van Dijck P. Diversity in genetic in vivo methods for protein-protein interaction studies: from the yeast two-hybrid system to the mammalian split-luciferase system. Microbiol. Mol. Biol. Rev. 2012; 76:331–82. [PubMed: 22688816]

17. Jacquier A. Two-hybrid systems — methods and protocols. Edited by Paul N. MacDonald, published by Humana Press, 2001, 336 p. Biochimie. 2002; 84

18. Dreze M, et al. High-quality binary interactome mapping. Methods Enzymol. 2010; 470:281–315. [PubMed: 20946815]

19. Braun P, et al. An experimentally derived confidence score for binary protein-protein interactions. Nat. Methods. 2009; 6:91–97. [PubMed: 19060903]

20. He F, et al. Large-scale atlas of microarray data reveals the distinct expression landscape of different tissues in Arabidopsis. Plant J. 2016; 86:472–80. [PubMed: 27015116]

21. Debernardi JM, et al. Post-transcriptional control of GRF transcription factors by microRNA miR396 and GIF co-activator affects leaf size and longevity. Plant J. 2014; 79:413–426. [PubMed: 24888433]

22. Cho WK, et al. Time-course RNA-Seq analysis reveals transcriptional changes in rice plants triggered by rice stripe virus infection. PLoS One. 2015; 10

23. Guilfoyle TJ. The PB1 domain in auxin response factor and Aux/IAA proteins: a versatile protein interaction module in the auxin response. Plant Cell. 2015; 27:33–43. [PubMed: 25604444]

24. Mukhtar MS, et al. Independently Evolved Virulence Effectors Converge onto Hubs in a Plant Immune System Network. Science (80-.). 2011; 333:596–601.

25. Villarino GH, et al. Temporal and spatial domain-specific transcriptomic analysis of a vital reproductive meristem in Arabidopsis thaliana. Plant Physiol. 2016; 171:42–61. [PubMed: 26983993]

26. Madmon O, et al. Expression of MAX2 under SCARECROW promoter enhances the strigolactone/MAX2 dependent response of Arabidopsis roots to low-phosphate conditions. Planta. 2016; :1–9. DOI: 10.1007/s00425-016-2477-7

27. Sun L, Song L, Zhang Y, Zheng Z, Liu D. Arabidopsis PHL2 and PHR1 Act Redundantly as the Key Components of the Central Regulatory System Controlling Transcriptional Responses to Phosphate Starvation. Plant Physiol. 2016; 170:499–514. [PubMed: 26586833]

28. Gangappa SN, Botto JF. The BBX family of plant transcription factors. Trends in Plant Science. 2014; 19:460–471. [PubMed: 24582145]

29. Preuss SB, et al. Expression of the Arabidopsis thaliana BBX32 gene in soybean increases grain yield. PLoS One. 2012; 7

30. Huang H, Bader JS. Precision and recall estimates for two-hybrid screens. Bioinformatics. 2009; 25:372–378. [PubMed: 19091773]

31. Llorca CM, et al. The elucidation of the interactome of 16 arabidopsis bZIP factors reveals three independent functional networks. PLoS One. 2015; 10

32. Century K, Reuber TL, Ratcliffe OJ. Regulating the regulators: the future prospects for transcription-factor-based agricultural biotechnology products. Plant Physiol. 2008; 147:20–9. [PubMed: 18443103]

33. Snider J, Kittanakom S, Curak J, Stagljar I. Split-ubiquitin based membrane yeast two-hybrid (MYTH) system: a powerful tool for identifying protein-protein interactions. J. Vis. Exp. 2010; :e1698.doi: 10.3791/1698

34. Wilson TE, Fahrner TJ, Johnston M, Milbrandt J. Identification of the DNA binding site for NGFI-B by genetic selection in yeast. Science. 1991; 252:1296–1300. [PubMed: 1925541]

35. Lumba S, et al. A mesoscale abscisic acid hormone interactome reveals a dynamic signaling landscape in arabidopsis. Dev. Cell. 2014; 29:360–372. [PubMed: 24823379]

36. Cao S, Siriwardana CL, Kumimoto RW, Holt BF. Construction of high quality Gateway™ entry libraries and their application to yeast two-hybrid for the monocot model plant *Brachypodium distachyon*. BMC Biotechnol. 2011; 11:53. [PubMed: 21595971]

37. Benatuil L, Perez JM, Belk J, Hsieh CM. An improved yeast transformation method for the generation of very large human antibody libraries. Protein Eng. Des. Sel. 2010; 23:155–159. [PubMed: 20130105]

38. Abremski K, Hoess R. Bacteriophage P1 site-specific recombination. Purification and properties of the Cre recombinase protein. J. Biol. Chem. 1984; 259:1509–1514. [PubMed: 6319400]

39. Ehlert A, et al. Two-hybrid protein-protein interaction analysis in Arabidopsis protoplasts: Establishment of a heterodimerization map of group C and group S bZIP transcription factors. Plant J. 2006; 46:890–900. [PubMed: 16709202]

40. Albert H, Dale EC, Lee E, Ow DW. Site-specific integration of DNA into wild-type and mutant lox sites placed in the plant genome. Plant J. 1995; 7:649–659. [PubMed: 7742860]

41. Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. Nat Methods. 2012; 9:357–359. [PubMed: 22388286]

42. Keilwagen J, et al. Area under Precision-Recall Curves for Weighted and Unweighted Data. PLoS One. 2014; 9:e92209. [PubMed: 24651729]

43. Csardi G, Nepusz T. The igraph software package for complex network research. Inter Journal Complex Syst. 2006; :1695.doi: 10.1109/ICCSN.2010.34

44. Zhao S, Guo Y, Sheng Q, Shyr Y. Heatmap3: an improved heatmap package with more powerful and convenient features. BMC Bioinformatics. 2014; 15:P16.

45. Shannon P, et al. Cytoscape: a software environment for integrated models of biomolecular interaction networks. Genome Res. 2003; 13:2498–504. [PubMed: 14597658]
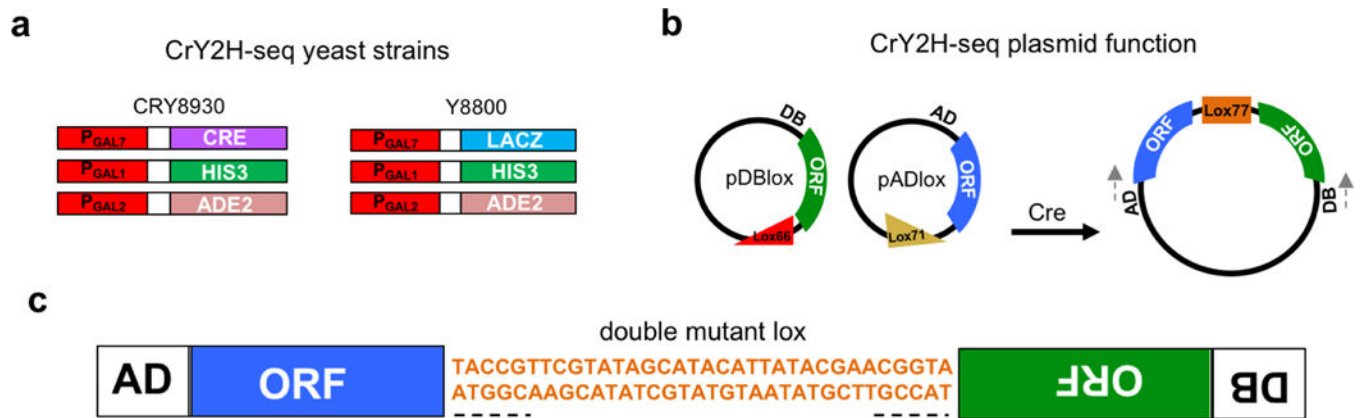
Figure 1.
CrY2H-seq strain and plasmid design. (**a**) CrY2H-seq uses yeast strains CRY8930 and
Y8800. (**b**) CrY2H-seq bait and prey plasmids pDBlox and pADlox contain mutant lox sites
(lox66 and lox71, respectively) flanking the 3' end of ORF inserts. Upon Cre/lox-
recombination of plasmids, a fused ORF product can be recovered by PCR amplification
using activation (AD) and DNA binding (DB) domain specific primers, indicated by the grey
arrows. (**c**) Representative PCR amplicon from AD and DB primers showing fused ORFs.
Mutant lox sites are underlined.
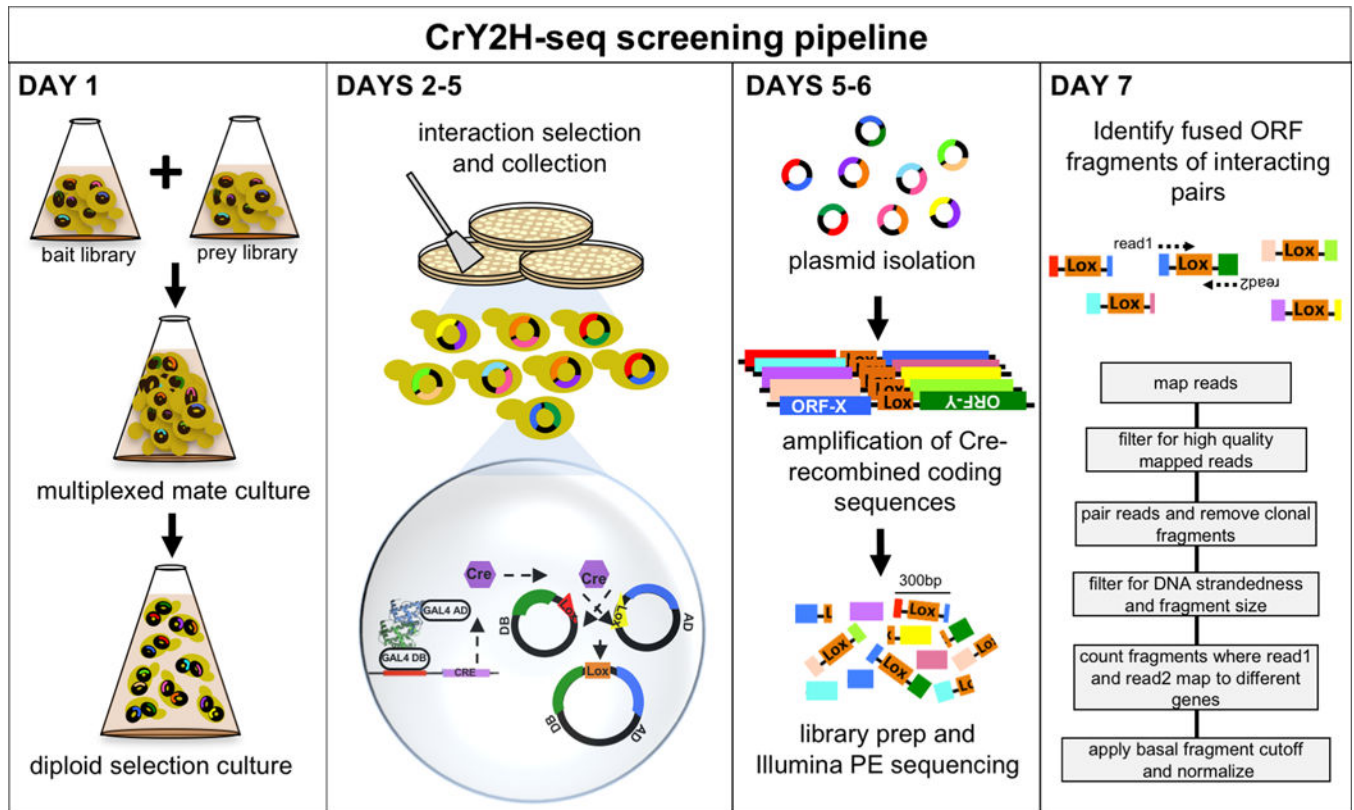
Figure 2.
The CrY2H-seq screening pipeline. On day 1, archival stocks of bait and prey libraries are combined in one massively-multiplexed mate culture that undergoes diploid selection overnight. On day 2, the diploid culture is plated on media to select for cells with protein interaction-mediated Gal4 reconstitution and subsequent transcriptional activation of the *HIS3* and *CRE* reporter genes. *HIS3* expression allows cells to survive on selection media and *CRE* expression permits unidirectional plasmid linkage, where ORF combinations corresponding to protein-protein interactions become fixed together inside cells. After 3 days of selection, surviving cells are harvested *en masse*, plasmids are purified in a single prep, and Cre-recombined ORF junctions are amplified in multi-template PCR reactions. From these amplicons, an Illumina sequencing library is prepared and sequenced. A bioinformatics pipeline is used to identify fragments derived from Cre recombination PCR products (see Supplementary Fig. 5 and **Online Methods** for more details, including media composition).
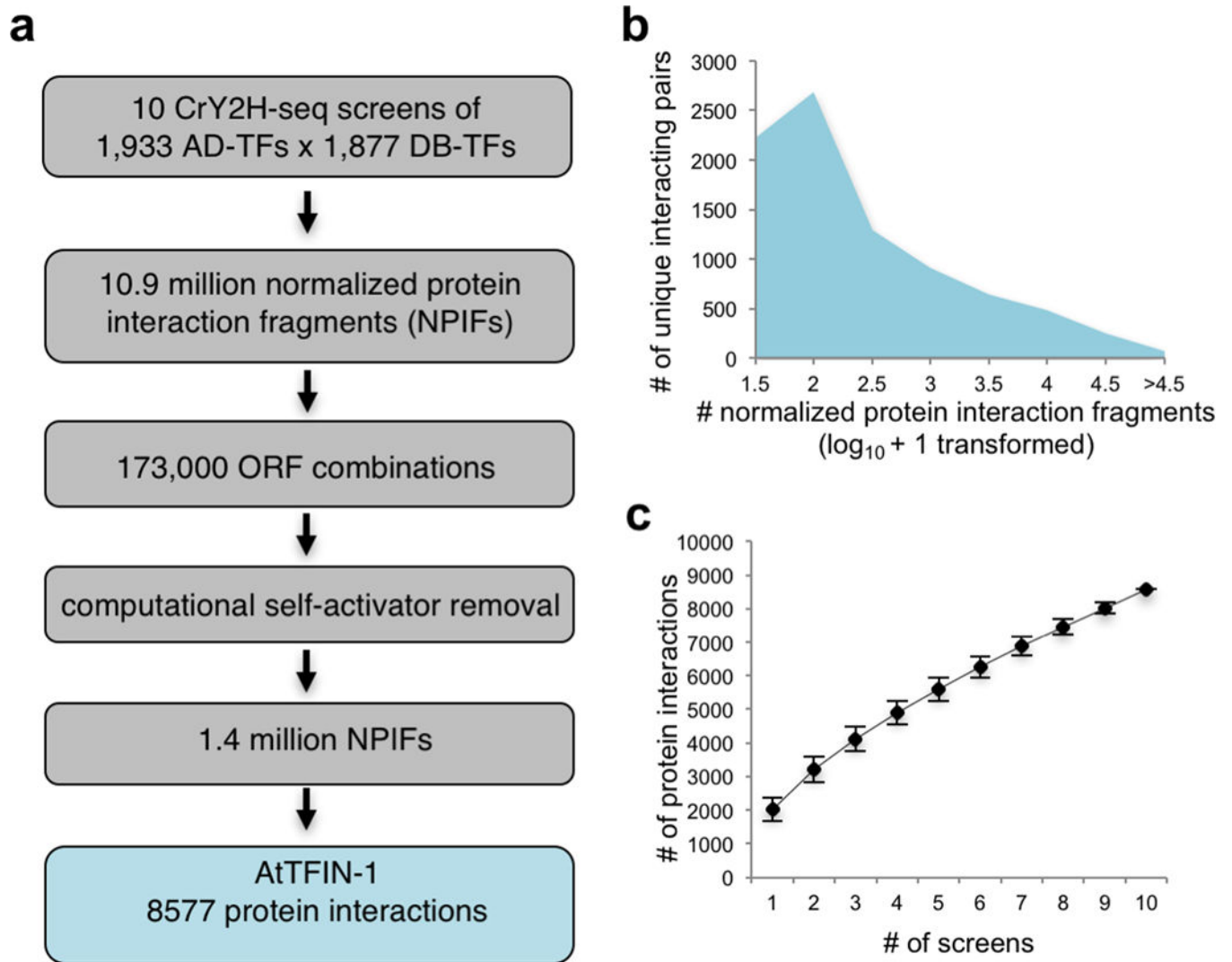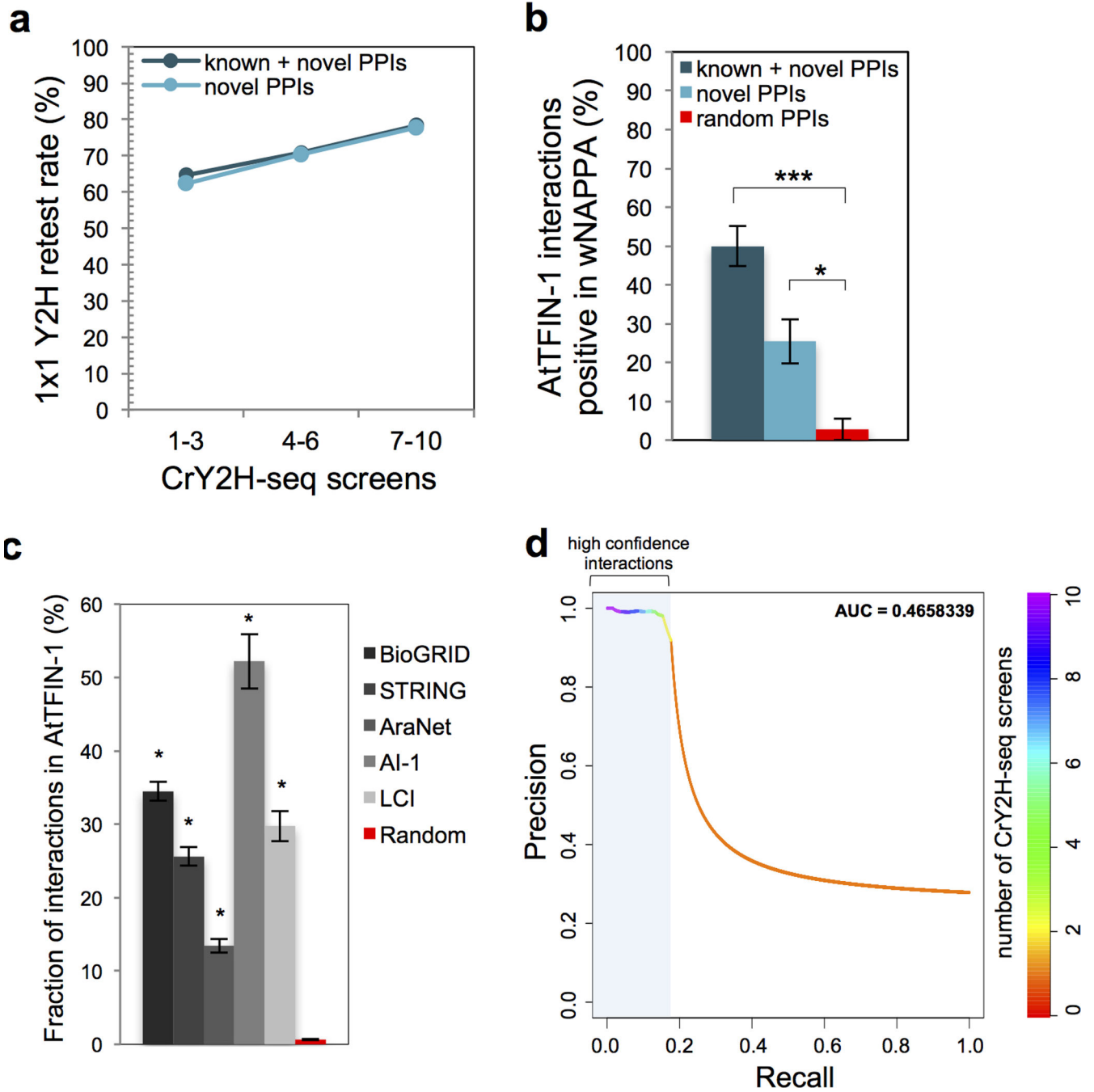
Figure 3.
Coverage of AtTFIN-1. (**a**) Summary of TF ORFeome screening. (**b**) Cumulative coverage of unique interacting pairs detected in paired-end sequencing of all ten CrY2H-seq screens after self-activator removal. (**c**) Sampling sensitivity shown by the average number of new interactions detected after each CrY2H-seq screen considering all possible (10!) orderings of screens. Error bars, standard deviation.

Figure 4.

Quality of AtTFIN-1. (**a**) Fraction of AtTFIN-1 protein-protein interactions (PPIs) that were positive in 1×1 matrix style Y2H retest screen (retest rate) as a function of the number of CrY2H-seq screens that interactions were observed in. Bin sizes, 1–3: 65, 4–6: 342, and 7–10: 249. (**b**) Fraction of AtTFIN-1 PPIs that were positive in wNAPPA. Error bars, standard error of proportion. *P* values, one-sided Fisher's exact test (*** = 3.57e-08, and * = 0.002395). (**c**) Fraction of 1,368 BioGRID, 1,198 STRING, 1,355 AraNet, 182 Arabidopsis Interactome-1 (AI-1), 501 Arabidopsis Interactome literature-curated interactions (LCI), and

8,577 random interactions in AtTFIN-1. Error bars, standard error of proportion. Literature and database interactions are detected significantly more often than random interactions (P values, one-sided Fisher's exact test, * = 2.2e-16). (**d**) Precision-recall curve calculated using the union of known interactions as true positives and a random interaction dataset as false positives plotted as a function of the number of CrY2H-seq screens that interactions were observed in. Interactions observed in two or more replicate experiments are classified as high-confidence interactions as indicated by the pale blue box.
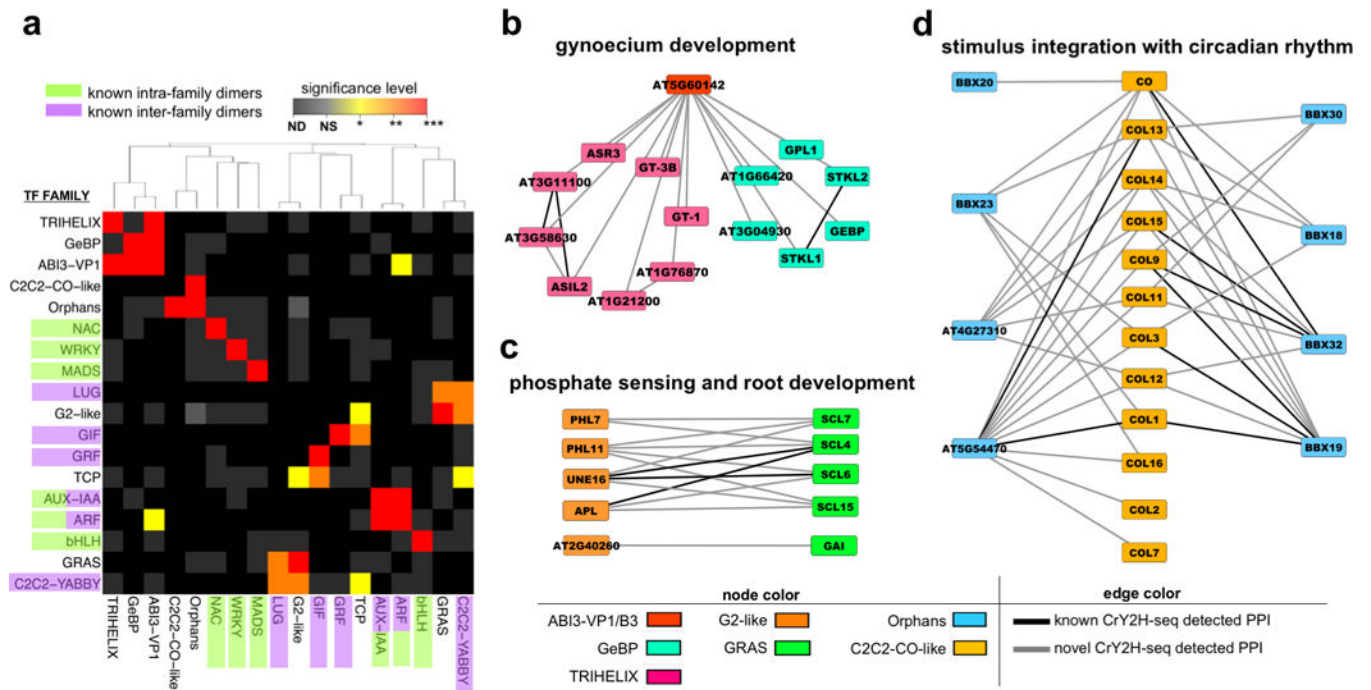
Figure 5.
Biological functions underlying TF family interactions in AtTFIN-1. (**a**) Discrete empirical
P values of family interactions observed more frequently in AtTFIN-1 than expected by
random chance. Families are hierarchically clustered by common family interactions. Color
key: ND = not detected, NS = not significant, * p<0.05, ** p< 0.01, *** p<0.001. Examples
of known intra-family and inter-family dimers are highlighted in green and purple,
respectively. See Supplementary Fig. 13 for a matrix showing all TF family interactions
observed. (**b**) An ABI3-VP1/B3 transcription factor preferentially interacts with many
members of TRIHELIX and GeBP families, a module potentially involved in gynoecium
development. (**c**) GRAS family members preferentially interact with G2-like family
members providing a potential molecular link between phosphate sensing and the regulation
of root development. (**d**) Preferential interaction between BBX domain-containing
"Orphans" proteins and C2C2-CO-like family members suggest a potential means by which
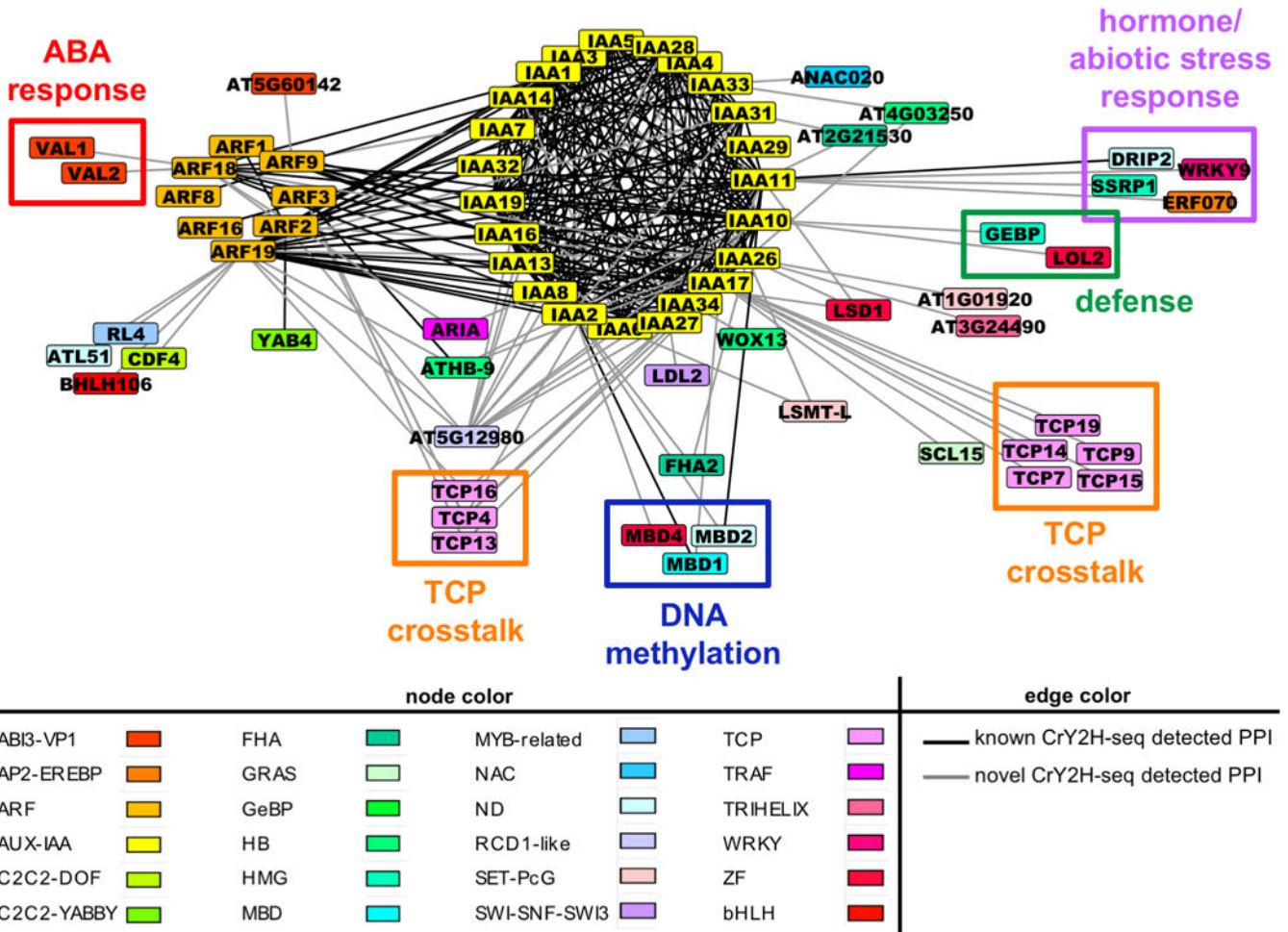stimulus signals are integrated with circadian rhythms.

Figure 6.
An expanded ARF-AUX-IAA transcription factor network. Distinct interactions among AUX-IAA and ARF proteins suggest certain family members have specific functions. IAA17 shows preferential enrichment for TCP family members. IAA2, 10, 17, and 18 commonly interact with MBD proteins. IAA11 shows distinct interactions with hormone and water stress related factors, ERF70 and DRIP2. ARF18 specifically interacts with VAL1 and VAL2 abscisic acid response factors. IAA10 interacts with LOL2 and GEBP defense response-related factors.