RESOURCE

Rational association of genes with traits using a genome-scale gene network for *Arabidopsis thaliana*

Insuk Lee[1,2,5], Bindu Ambaru[4,5], Pranjali Thakkar[4], Edward M Marcotte[2,3] & Seung Y Rhee[4]

We introduce a rational approach for associating genes with plant traits by combined use of a genome-scale functional network and targeted reverse genetic screening. We present a probabilistic network (AraNet) of functional associations among 19,647 (73%) genes of the reference flowering plant *Arabidopsis thaliana*. AraNet associations are predictive for diverse biological pathways, and outperform predictions derived only from literature-based protein interactions, achieving 21% precision for 55% of genes. AraNet prioritizes genes for limited-scale functional screening, resulting in a hit-rate tenfold greater than screens of random insertional mutants, when applied to early seedling development as a test case. By interrogating network neighborhoods, we identify AT1G80710 (now *DROUGHT SENSITIVE 1*; *DRS1*) and AT3G05090 (now *LATERAL ROOT STIMULATOR 1*; *LRS1*) as regulators of drought sensitivity and lateral root development, respectively. AraNet (http://www.functionalnet.org/aranet/) provides a resource for plant gene function identification and genetic dissection of plant traits.

Manipulating plant traits that affect the production of food, fiber and renewable energy has important agricultural and economic consequences. What is the best approach for identifying genes for important plant traits? Forward genetics is limited as mutations in many genes may generate only moderate or weak phenotypes. Similarly, although reverse genetics allows for directed assay of gene perturbations[1], saturated phenotyping for many plant traits is impractical. A pragmatic near-term solution is the computational identification of likely candidate genes for desired traits, allowing for focused, efficient use of reverse genetics. This solution is not unlike rational drug design in which computer-assisted and expert knowledge are combined with targeted screening for the desired drug, or in this case, trait.

One emerging approach for prioritizing candidate genes is network-guided guilt by association. In this approach, functional associations are first determined between genes in a genome on the basis of extensive experimental data sets, encompassing millions of individual observations. Such a map of functional associations is often represented as a graph model and referred to as a functional gene network[2].

Probabilistic functional gene networks integrate heterogeneous biological data into a single model, enhancing both model accuracy and coverage. Once a suitable network is generated, new candidate genes are proposed for traits based upon network associations with genes previously linked to these traits. Such network-guided screening has been successfully applied to unicellular organisms[3,4] and *Caenorhabditis elegans*[5,6], and is a proposed strategy for identifying human disease genes[4,5,7–10].

We demonstrate here that this approach successfully identifies genes affecting specified traits for a reference flowering plant, *A. thaliana*, and we introduce a genome-wide, functional gene network for *Arabidopsis* suitable for prioritizing candidate genes for a wide variety of traits of economic and agricultural importance.

RESULTS

Reconstruction of an *Arabidopsis* gene network

We integrated diverse functional genomics, proteomics and comparative genomics data sets into a genome-wide functional gene network, using data integration and benchmarking methods customized for *Arabidopsis* genes (**Supplementary Methods**). The data sets included mRNA co-expression patterns measured from DNA microarray data sets (**Supplementary Table 1**), known *Arabidopsis* protein-protein interactions[11–14], protein sequence features including sharing of protein domains, similarity of phylogenetic profiles[15–17] or genomic context of bacterial or archaebacterial homologs[18–20], and diverse gene-gene associations (mRNA co-expression, physical protein interactions, multiprotein complexes, genetic interactions, literature mining) transferred from yeast[21], fly[11,22–24], worm[5] and human genes based on orthology[25] (**Supplementary Table 2**). In total, 24 distinct types of gene-gene associations, encompassing >50 million individual experimental or computational observations, were scored for their ability to correctly reconstruct shared membership in *Arabidopsis* biological processes. Then, these were incorporated into a single integrated network model, dubbed AraNet. AraNet contains 1,062,222 functional linkages among 19,647 genes (~73% of the total *Arabidopsis* genes), with each linkage weighted by the log likelihood of the linked genes to participate in the same biological processes.

Integrating data improves network coverage and accuracy, as tested by recovery of known functional associations (**Fig. 1a** and **Supplementary Fig. 1**). AraNet extends substantially beyond

[1]Department of Biotechnology, College of Life Science and Biotechnology, Yonsei University, Seodaemun-gu, Seoul, Korea. [2]Center for Systems and Synthetic Biology, and [3]Department of Chemistry and Biochemistry, Institute for Cellular and Molecular Biology, University of Texas at Austin, Austin, Texas, USA. [4]Department of Plant Biology, Carnegie Institution for Science, Stanford, California, USA. [5]These authors contributed equally to this work. Correspondence should be addressed to I.L. (insuklee@yonsei.ac.kr) or E.M.M. (marcotte@icmb.utexas.edu) or S.Y.R. (rhee@acoma.stanford.edu).

Received 10 June 2009; accepted 23 December 2009; published online 31 January 2010; doi:10.1038/nbt.1603

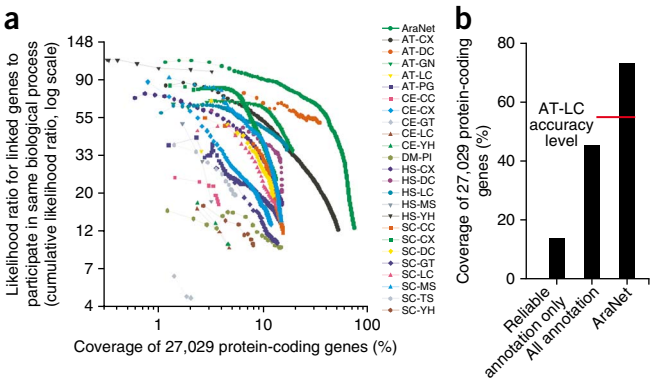

Figure 1 Construction, accuracy and coverage of AraNet, a functional gene network for *Arabidopsis*. (**a**) Pairwise gene linkages derived from 24 diverse functional genomics and proteomics data sets, representing >50 million experimental or computational observations, were integrated into a composite network with higher accuracy and genome coverage than any individual data set. The integrated network (AraNet) contains 1,062,222 functional linkages among 19,647 (73%) of the 27,029 protein-coding *Arabidopsis* genes. The plot *x* axis indicates the log-scale percentage of the 27,029 protein-coding genes[13] covered by functional linkages derived from the indicated data sets (curves); the *y* axis indicates predictive quality of the data sets, measured as the cumulative likelihood ratio of linked genes to share GO biological process term annotations, tested using 0.632 bootstrapping and plotted for successive bins of 1,000 linkages each (symbols). Data sets are named as XX-YY, where XX indicates species of data origin (AT, *A. thaliana*; CE, *C. elegans*; DM, *D. melanogaster*; HS, *H. sapiens*; SC, *S. cerevisiae*) and YY indicates data type (CC, co-citation; CX, mRNA co-expression; DC, domain co-occurrence; GN, gene neighbor; GT, genetic interaction; LC, literature-curated protein interactions; MS, affinity purification/mass spectrometry; PG, phylogenetic profiles; PI, fly protein interactions; TS, tertiary structure; YH, yeast two-hybrid. Relative contribution of each data type and combining different evidences for inferring function is discussed in **Supplementary Discussion** (**Supplementary Figs. 13–15**). (**b**) AraNet spans ~73% of the protein-coding genes, far in excess of current GO biological process annotations for *Arabidopsis*, for which ~12.2% of genes are annotated by reliable experimental evidence (GO evidence codes IDA, IMP, IGI, IPI, IEP) or traceable author statements (GO evidence code TAS), or ~45% annotated by any evidence including computational inferences or sequence homology. The subset of AraNet linkages stronger than the likelihood ratio for literature-curated protein interactions (AT-LC, corresponding to a likelihood ratio of 35:1) covers 55% of the genes.

well-characterized *Arabidopsis* genes (**Fig. 1b**): 23,720 *Arabidopsis* genes are unannotated with Gene Ontology (GO) biological process annotations by reliable experimental evidence[13]. AraNet includes more than half (7,465) of genes lacking even sequence homology–based annotations (14,847 genes). These genes' functions can now be hypothesized based upon their network neighborhoods. AraNet implicates specific processes for 4,479 uncharacterized genes using guilt by association. There are 2,986 uncharacterized genes associated only with other uncharacterized genes in AraNet, suggesting many still-uncharacterized cellular processes in plants.

Evaluating the accuracy of AraNet

To verify the reliability of functional associations in AraNet, we tested their consistency with known *Arabidopsis* gene annotations by applying guilt by association in AraNet to identify genes associated with specific biological processes. Each gene in the genome was scored for association with a particular process by summing network edge weights connecting that gene to known genes in that process. A gene's resulting score corresponds to the naive Bayes estimate for the gene to belong to that process given network evidence (**Fig. 2a**). Performing cross-validation of this test allows us to assess predictive power with a receiver-operator characteristic (ROC) curve, measuring the true-positive prediction rate versus false-positive prediction rate as a function of prediction score. We use the area under the ROC curve (AUC) to summarize performance. AUC values of ~0.5 and 1 indicate random and perfect performance, respectively.

Using cross-validation, we tested AraNet's ability to correctly associate genes with each GO biological process, observing significantly better than random predictability for the majority of biological processes ($P < 10^{-53}$; Wilcoxon signed rank test unless noted otherwise) (**Fig. 2b**). AraNet incorporates data from other organisms; we correspondingly observed higher predictability for evolutionarily conserved processes than for GO processes annotated only with plant genes ($P < 10^{-24}$, Wilcoxon rank sum test) (**Figs. 2c,d**). Nonetheless, genes were correctly associated with plant-specific processes at significantly higher rates than expected by chance ($P < 10^{-28}$) (**Fig. 2d**). For example, many important plant traits showed high predictability, including abiotic stress responses (**Fig. 2e**), organ development (**Fig. 2f**), biotic stress responses (**Supplementary Fig. 2a**) and hormonal signaling (**Supplementary Fig. 2b**). Tests on two additional independent data sets—the set of reliable GO cellular component annotations describing 86 subcellular locations or protein complexes of *Arabidopsis* proteins[13] and the KEGG definitions of 82 *Arabidopsis* biochemical pathways[26]—show similarly high predictive power ($P \leq 10^{-16}$ and $P \leq 10^{-14}$, respectively, compared to chance) (**Fig. 3**). We find that AraNet shows far stronger predictability than previous smaller-scale networks of *Arabidopsis* genes (**Supplementary Table 3** and **Supplementary Fig. 3**)[27–30], stemming at least in part from greater coverage, which allows for stronger guilt by association due to higher query gene coverage. Thus, AraNet is strongly predictive for many *Arabidopsis* processes, including those specific to plants.

Contributions of plant versus non-plant data to AraNet

Many AraNet linkages derive from orthology to animals and yeast, organisms evolutionarily distant from *Arabidopsis*. Therefore, we asked the extent to which non-plant–derived linkages contribute to AraNet's accuracy. A version of AraNet composed only of links from yeast and animal data sharply reduces predictive power ($P < 10^{-4}$) (**Fig. 2b**). A version of only plant-derived links performs substantially better ($P \leq 10^{-11}$), implying that plant-derived data underlies much of AraNet's predictive power (**Fig. 2b**). This plant-derived data dependence is stronger when predicting plant-specific than conserved processes ($P \leq 10^{-19}$, Wilcoxon rank sum test) (**Fig. 2c,d**).

Any method for linking genes to plant traits must perform well for plant-specific processes, so we examined evidence supporting these cases. Notably, even for well-predicted plant-specific pathways (AUC scores ranging from ~0.7 to 1, **Fig. 4**), supporting evidence did not derive entirely from plants. For example, photorespiration genes were identified by combining evidence from *Arabidopsis*, human and *C. elegans*. Similarly, trichome differentiation genes were recovered using predominantly human and yeast-derived evidence. Genes of abscisic acid–mediated signaling drew support from all organisms, including fly protein interactions. Thus, although plant-derived links provide most of AraNet's predictive power, non-plant–derived linkages help substantially in associating genes with plant-specific processes, as processes unique to plants nonetheless often involve conserved genes with conserved interactions.

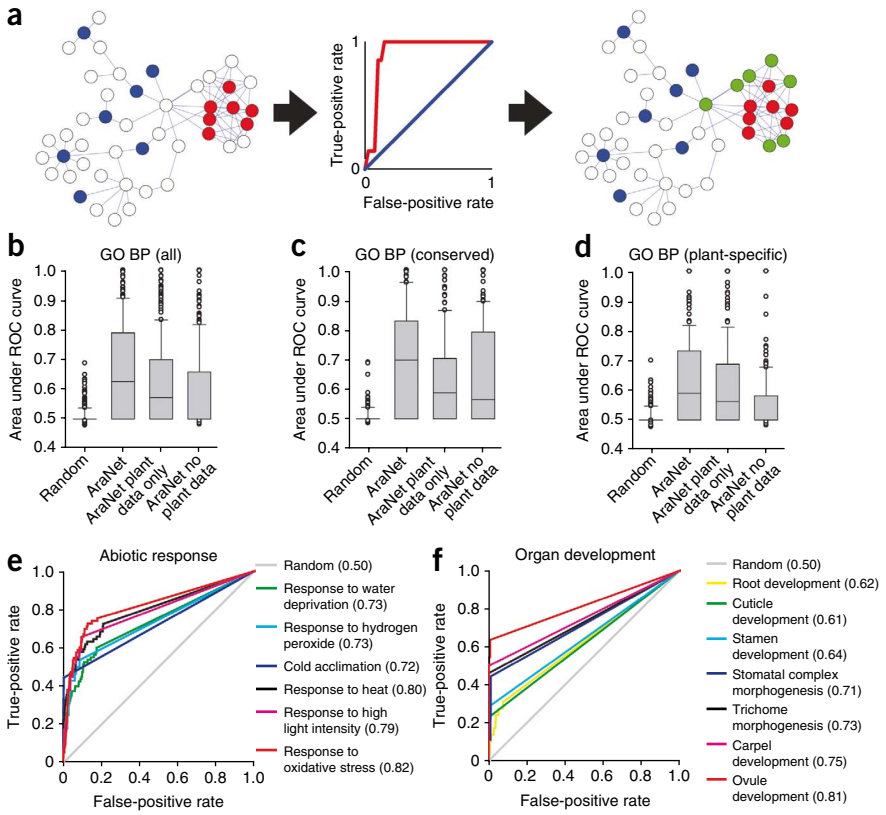

Figure 2 Predictive power of AraNet for conserved and plant-specific biological processes. AraNet's predictive capacity was measured using cross-validated receiver operator characteristic (ROC) curve analysis, as illustrated in (**a**). For a given process, each gene in the *Arabidopsis* genome is rank-ordered by the sum of its network linkage scores to the set of 'bait' genes already associated with that process (omitting each bait gene from the bait set for purposes of evaluation). High-scoring genes are most tightly connected to the bait set and are the most likely new candidates to participate in that process. This trend is evident in a ROC plot measuring recovery of bait genes as a function of rank, calculating the true-positive prediction rate (sensitivity; TP/(TP+FN)) versus the false-positive prediction rate (1–specificity; FP/(FP+TN)). If bait genes are highly interconnected (red circles), unlike random genes (blue circles), additional genes connected to the bait genes (green circles) are more likely to be involved in the same process. The area under the cross-validated ROC curve (AUC) provides a measure of predictability, ranging from ~0.5 for random expectation (blue curve) to 1 for perfect predictions (red curve). (**b**) Distributions of AUC values are plotted for network-based identification of genes for each of the 318 GO biological process terms with annotations, (**c**) for each of the 151 biological process terms with annotations shared between plant and animal or between plant and yeast and (**d**) for each of the 167 biological process terms with annotations found in plants but absent from animals and fungi. In bar-and-whiskers plots, the central horizontal line in the box indicates the median AUC and the boundaries of the box indicate the first and third quartiles of the AUC distribution. Whiskers indicate the 10th and 90th percentiles, and circles indicate individual outliers. AraNet specifically identified genes associated with (**e**) plant abiotic stress response genes and (**f**) organ developmental processes, as annotated by GO. AUC values are indicated in parentheses.

Linked genes share cell type–specific expression patterns

Many traits in multicellular organisms pertain to specific tissues or cell types. The predictive strength shown by AraNet for such processes raises the question of how a global gene network, incorporating diverse samples and data from orthologs, can correctly identify genes for cell type– and tissue-specific processes. Using measurements of transcript observations in 20 root cell types[31] that were not used in building AraNet, we measured the extent to which genes linked in AraNet were spatiotemporally co-expressed in these cells. We find that linked genes show strong cell-specific co-expression in *Arabidopsis* (**Fig. 3c**)—indeed, far stronger than in previous networks of *Arabidopsis* genes (**Supplementary Table 3**)[27–30]—with linked genes four times more likely to be expressed in the same cell types than expected by chance. Thus, although different individual networks were not constructed for each cell type, such cell and tissue specificity is nonetheless at least in part implicitly encoded in AraNet linkages. This correlation between functional association and spatiotemporal co-expression of genes likely enhances prediction strength for many traits, and is evident even for linkages between characterized and uncharacterized genes (**Fig. 3c**), supporting applicability of AraNet to uncharacterized genes.

Associating genes with specific mutant phenotypes

Because linked genes in AraNet tend to operate in the same processes (**Figs. 1–4**), we might expect that they often affect the same phenotypic traits[3,5]. This allows association of new candidate genes with traits of interest based on network connections. To test this, we used results from large-scale mutant seed phenotyping[32] and analyzed genes whose disruption induced embryonic lethality or changes in seed (embryo) pigmentation. Genes involved in each trait were interlinked significantly more often compared to chance ($p < 10^{-31}$ for embryonic lethality and $P < 10^{-10}$ for seed pigmentation, normal distribution) (**Fig. 3d**). Unlike AraNet, previous *Arabidopsis* gene networks[27–30] do not significantly predict either phenotype (**Supplementary Fig. 4**). Thus, AraNet offers a feasible approach for selecting genes likely to be associated with specific plant traits.

Tenfold enrichment for seed pigmentation genes

To experimentally test the association of new genes with a trait, we used 23 known seed pigmentation genes (**Supplementary Table 4**) to search AraNet for new pigmentation genes. Genes in this phenotypic class generally affect chloroplast development or photomorphogenesis, and mutant seedlings show early developmental defects, with albino, pale green, purple or variegated leaves[33].

From AraNet's top 200 candidate genes, we screened all genes with available homozygous T-DNA insertional mutant lines (**Supplementary Table 5**). We screened 90 candidate genes (represented by 118 mutant lines), of which 14 genes (represented by 17 lines) exhibited color and morphology defects in young seedlings, reminiscent of seed pigmentation mutants (**Supplementary Tables 6** and **7**). This represents a tenfold enrichment in the discovery rate of the mutant phenotype ($P \leq 10^{-12}$, binomial distribution) over that observed during screens of T-DNA insertional lines[33] (see Online Methods). This discovery rate compares well to animal networks, for example, in *C. elegans* 16 tumor suppressor effectors were identified from 170 candidates[5].

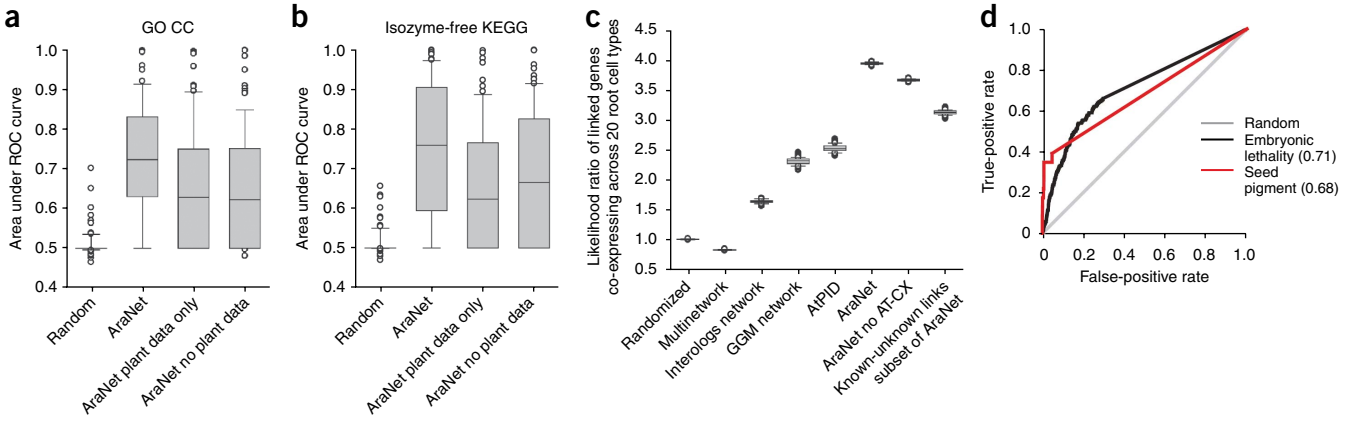

Figure 3 Validation of AraNet by independent data sets. (**a**) The predictive power of AraNet for 86 GO cellular component terms. (**b**) AraNet predictive power for 82 KEGG metabolic pathways (excluding isozymes). Bar-and-whiskers plots are as in **Figure 2b–d**. (**c**) Co-occurrence of mRNA transcripts across 20 root cell-types[31] for network-linked genes. Genes linked in AraNet (AraNet, lane 6) were more co-expressed in each root cell-type than gene pairs from random networks (repeating the calculation for 100 randomized networks and plotting the distribution of the 100 resulting odds ratios (randomized, lane 1; see **Supplementary Methods** for details). This trend cannot be explained simply by the incorporation of *Arabidopsis* mRNA expression data into AraNet, as a version of the network lacking this data shows similarly high cell-type specificity (AraNet no AT-CX, lane 7). Compared to other previous *Arabidopsis* networks (lanes 2–5, **Supplementary Table 3**), AraNet shows 1.6- to 4-fold higher cell-type specificity. Subset of AraNet composed of only known-unknown genes shows high cell type–specificity as well, supporting predictability of AraNet using cell-specific expression data (lane 8). (**d**) AraNet shows predictability for genes affecting embryonic lethality or changes in seed pigmentation, as identified in the SeedGenes database[32]. AUC values are indicated in parentheses.

Of the 14 genes with mutant phenotypes, 3 genes (AT5G45620, AT4G26430 (also known as *CSN6B*) and AT5G50110) exhibited the phenotypes in two alleles, 6 genes in only one of the two alleles, and 5 genes were tested in only one allele (**Fig. 5a** and **Supplementary Table 7**). The 6 genes in which only one of the two alleles showed phenotype are likely to be untagged and were not characterized further. Expressivity of the phenotypes of the 11 lines representing 8 genes (6 lines for 3 genes and 5 lines for 5 genes) varied among individual plants within the homozygous population, ranging from delayed or failed germination, arrested or delayed development, anthocyanin accumulation, clear or white patches on the shoot to pale green shoot. As expected from known seed pigmentation mutants, survival rate in soil was <100% in most lines (**Supplementary Table 8**).

To determine how these genes are associated with seed pigmentation, we examined linkages among the known and newly identified (3 supported by two alleles and 5 by one allele) genes (**Fig. 5b**). These genes form five network components, two belonging to photomorphogenesis and three to chloroplast development (**Supplementary Table 9**).

The largest component includes members of the COP9 signalosome complex (CSN), an evolutionarily conserved post-translational regulatory complex involved in cell proliferation, response to DNA damage and gene expression[34]. In plants, the CSN complex is essential for photomorphogenesis[35]. Two of the three genes supported by two independent alleles belong to this component, AT4G26430 and AT5G45620. AT4G26430 (*CSN6B*) encodes a subunit of CSN, CSN6. Using a single allele, *CSN6B* has been shown to be genetically redundant to another gene (*CSN6A*) under white light, though only partially redundant in dark and blue light[36]. AT5G45620 encodes a protein with sequence similarity to a subunit of the lid subcomplex of 26S proteasome, but its biological role is unknown[13]. Supporting evidence for its prediction comes from protein domain co-occurrence with FUS5, FUS6 and COP8, and among their human orthologs (**Supplementary**

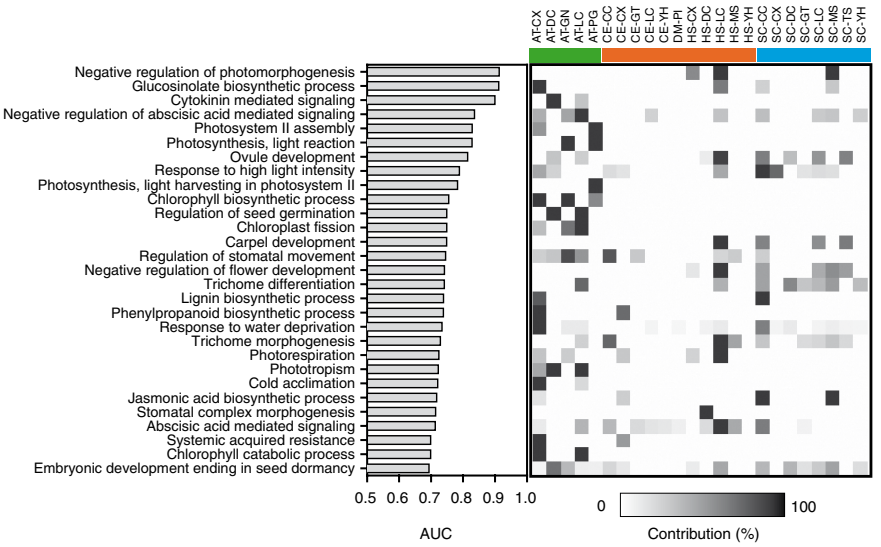

Figure 4 AraNet correctly associates genes with many processes unique to plants, nonetheless relying at least in part on data from animals and yeast, which contribute evidence for linkages among genes that are broadly conserved but whose roles in *Arabidopsis* are in plant-specific processes. The performance at associating genes with each of 29 biological processes specific to plants (that is, annotated only with plant genes in GO and known to occur only in plants) is summarized as the area under a cross-validated ROC curve (AUC). Even though these processes are absent in animals or fungi, the associated genes often have orthologs in these taxa, and AraNet draws upon data from these orthologs in making the associations. Each gray square demarks the support provided by a data set, measured as percentage of a sum of log likelihood scores contributing to that process, with darker gray indicating higher support. Data sets are labeled as in **Figure 1**.

Table 5). The CSN and the lid subcomplex of 26S proteasome complexes share structural and functional similarities[34], suggesting involvement of other protein degradation machineries in photomorphogenesis and early seedling development. The self-pollinated progeny of mutants that survived to make seeds did not show the seedling defects under standard growth conditions as their progenitors (data not shown). However, when grown in dark or under blue light, the mutants showed slight (5–25%) but significant ($P < 0.01$, paired *t*-test, see Online Methods) differences in hypocotyl length (**Fig. 5c**,**d**). CSN6B mutants showed reduced hypocotyl length in dark but slightly increased hypocotyl length when grown under blue light compared to wild type. The other two mutants had longer hypocotyls than wild type under both dark and blue light conditions. All three genes have paralogs in the genome. As double mutants of *CSN6A* and *CSN6B* show a constitutive photomorphogenic phenotype, severe seedling dwarfism and high levels of anthocyanin accumulation[36], it is possible that AT4G26430 and AT5G45620 may also exhibit genetic redundancy with their respective paralogs.

The remaining hits from our screen link to each other and known seed pigmentation genes in three components relevant to thylakoid biogenesis and chlorophyll biosynthesis, processes affecting chloroplast development and function (**Fig. 5b**). **Supplementary Table 9** details possible roles for the newly discovered genes. These results confirm that AraNet can efficiently associate new genes with a specific phenotypic trait.

Discovering functions for uncharacterized *Arabidopsis* genes

Given that AraNet can successfully associate genes with traits of interest, we wished to test hypothesized roles for uncharacterized *Arabidopsis* genes *in planta*. AraNet predicts biological roles for 4,479 previously uncharacterized genes. We selected three uncharacterized genes (AT1G80710, AT2G17900 and AT3G05090) based on several criteria: (i) no known biological process assigned; (ii) predicted by AraNet to be involved in developmental regulatory processes; and (iii) exist as single-copy genes. These represent extremely stringent tests of the network-based association method, and are all cases in which prediction based on sequence homology has failed.

AraNet predicts GO biological process annotations, ordering predictions by the sum of the log likelihood scores linking a gene to genes already annotated by each term (**Supplementary Table 10**). For the three genes selected, we tested for morphological and physiological phenotypes in the top ten predicted processes. Two control genes, AT1G15772 and AT2G34170, were chosen randomly from genes lacking AraNet functional predictions. Mutant plants were confirmed for homozygosity (**Supplementary Table 11**) and lack of detectable transcripts (data not shown). Self-pollinated progeny of homozygous plants were subjected to a bank of phenotypic assays based on the top ten predictions (see Online and **Supplementary Methods**). Of the three mutants, two exhibited phenotypes in the predicted processes.

AT1G80710 is a regulator of drought sensitivity

AraNet implicated the gene AT1G80710 in the response to water deprivation, among other processes, drawing support from affinity purifications of yeast orthologs (SC-MS)[37,38] (**Supplementary Table 10**). This gene is expressed in all tissues examined, with highest abundance in flowers (**Supplementary Fig. 5**). We asked whether the ability to retain water differed in the mutants. Under drought, mutant plants retained ~80% of the water of wild type ($P \leq 0.001$, unpaired *t*-test, **Fig. 6a**). Reduced water retention was not observed in control mutants (**Supplementary Fig. 6**).

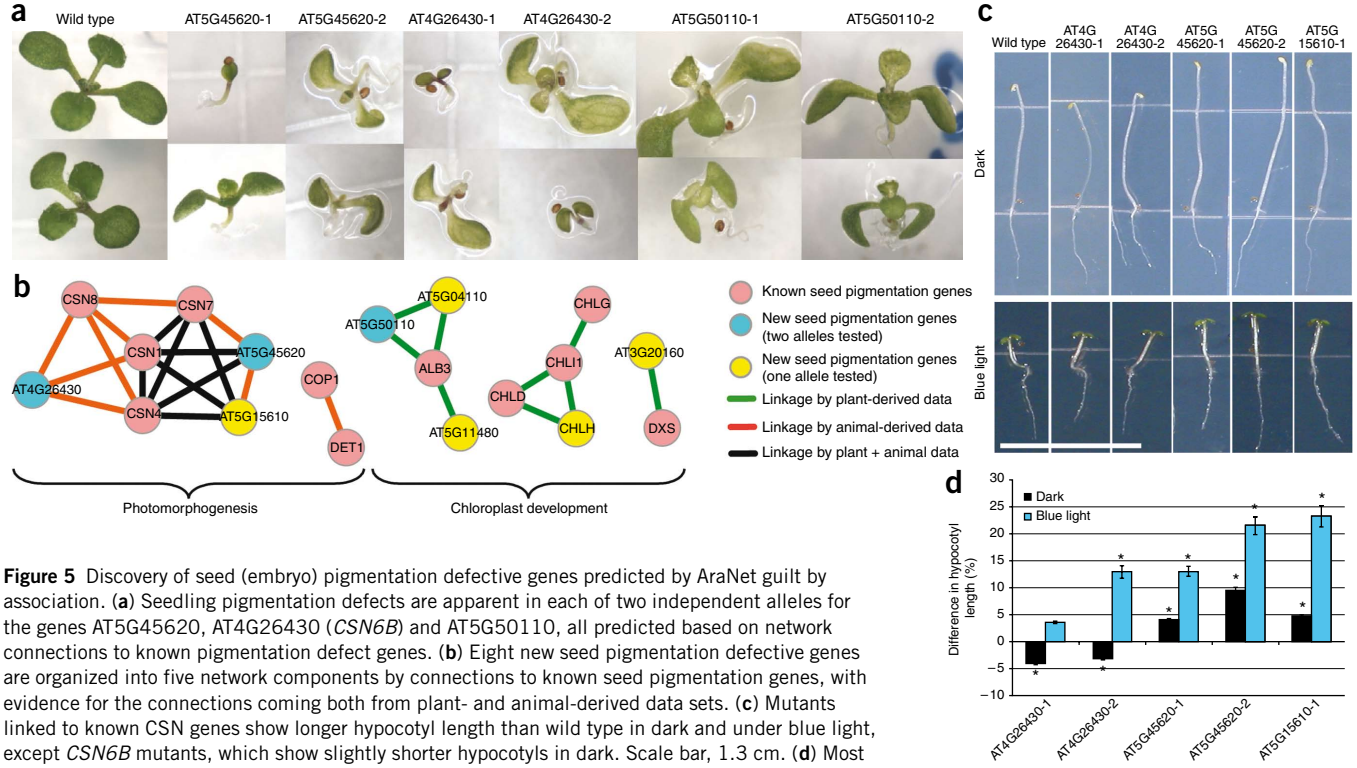

Figure 5 Discovery of seed (embryo) pigmentation defective genes predicted by AraNet guilt by association. (**a**) Seedling pigmentation defects are apparent in each of two independent alleles for the genes AT5G45620, AT4G26430 (*CSN6B*) and AT5G50110, all predicted based on network connections to known pigmentation defect genes. (**b**) Eight new seed pigmentation defective genes are organized into five network components by connections to known seed pigmentation genes, with evidence for the connections coming both from plant- and animal-derived data sets. (**c**) Mutants linked to known CSN genes show longer hypocotyl length than wild type in dark and under blue light, except *CSN6B* mutants, which show slightly shorter hypocotyls in dark. Scale bar, 1.3 cm. (**d**) Most of the differences in hypocotyl length of these mutants are slight (5–25%) but significant. Significant differences from wild type are indicated by asterisks ($P < 0.01$, paired *t*-test, $n = 26$ (dark), 32 (blue light)). *n* indicates the number of plates, each plate containing 7–8 plants of wild type and mutant genotype. Results are from seven (dark) or eight (blue light) independent experiments, plotting mean ± s.e.m.

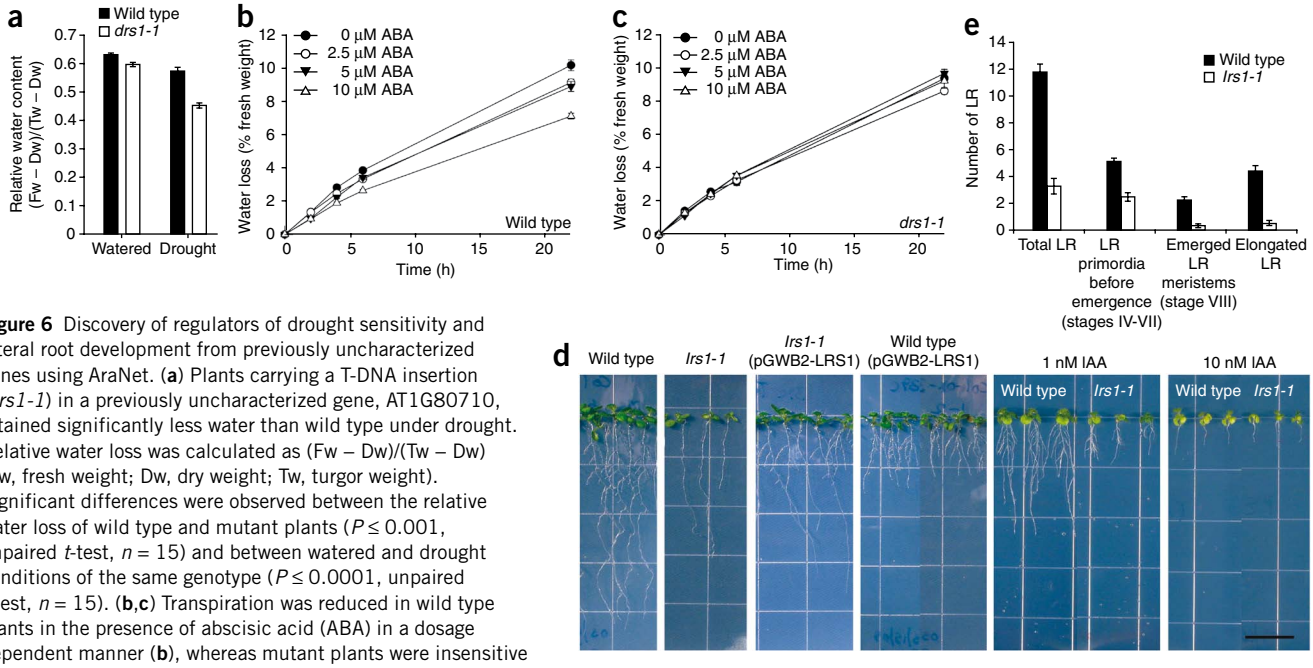

Figure 6 Discovery of regulators of drought sensitivity and lateral root development from previously uncharacterized genes using AraNet. (**a**) Plants carrying a T-DNA insertion (*drs1-1*) in a previously uncharacterized gene, AT1G80710, retained significantly less water than wild type under drought. Relative water loss was calculated as (Fw – Dw)/(Tw – Dw) (Fw, fresh weight; Dw, dry weight; Tw, turgor weight). Significant differences were observed between the relative water loss of wild type and mutant plants ($P \leq 0.001$, unpaired *t*-test, $n = 15$) and between watered and drought conditions of the same genotype ($P \leq 0.0001$, unpaired *t*-test, $n = 15$). (**b**,**c**) Transpiration was reduced in wild type plants in the presence of abscisic acid (ABA) in a dosage dependent manner (**b**), whereas mutant plants were insensitive to ABA (**c**). (**d**) The number of lateral roots is strongly reduced in lines carrying a T-DNA insertion (*lrs1-1*) in another previously uncharacterized gene AT3G05090. This phenotype can be complemented by reintroduction of the functional gene. When additional copies of the gene are expressed in a wild type strain, lateral roots increase, whereas the primary root decreases in length. Six other independent transformants in each background gave similar phenotypes (data not shown). 1 nM Auxin (IAA) increases the number and length of lateral roots in both the wild-type and mutant seedlings. Contrarily, 10 nM IAA severely reduces the primary root length in both genotypes. Scale bar, 1.4 cm. (**e**) Different stages of the lateral root formation are affected in the *lrs1-1* mutant. Wild-type lateral roots are distributed fairly evenly among lateral root primordia, emerged lateral root and elongated lateral root. The mutant has reduced numbers of the lateral roots in all of these stages, though the reduction is more severe in the emerged and elongated lateral root than that in the lateral root primordia. LR, lateral root. Error bars indicate s.e.m.

Drought response is mediated by several signaling pathways in *Arabidopsis*, including the hormone abscisic acid (ABA), transcription factor DREB2A, ERD1 and E3 SUMO ligase SIZ1[39,40]. To determine whether the reduced water retention upon drought stress is mediated by ABA, we examined the effect of ABA on transpiration of detached leaves. The mutant was insensitive to ABA on water loss, whereas the wild type lost significantly less water in the presence of 10 μM ABA ($P \leq 0.0004$, unpaired *t*-test, **Fig. 6b**,**c**). At this ABA concentration, mutant leaves lost 30% more water than wild type ($P \leq 0.04$, unpaired *t*-test, **Fig. 6b**,**c**). ABA showed no effect on germination rate in the mutant (data not shown), indicating that not all ABA-mediated processes are affected in the mutant.

Both the water retention and ABA-insensitive water transpiration response segregated as a single recessive Mendelian locus and linked to the T-DNA insertion (**Supplementary Table 12** and **Supplementary Fig. 7**). We designate AT1G80710 as *DRS1* (*DROUGHT SENSITIVE 1*) and the T-DNA insertion allele (Salk_001238C) as *drs1-1*. An independent T-DNA allele (Salk_149366C) that we designate as *drs1-2* exhibited the same phenotypes in relative water content after drought and ABA-insensitive water transpiration (**Supplementary Fig. 8**), confirming that the phenotypes are linked to mutations in *DRS1*. *DRS1* is a WD-40 repeat family protein containing a DWD (DDB1 binding WD-40) motif[41]. Some DWD-containing proteins are substrate receptors for DDB1-Cul4 ubiquitin ligase machinery in humans, yeast and *Arabidopsis*[41,42]. Combination of AraNet prediction and experimental testing thus demonstrates that *DRS1* promotes tolerance to drought stress, possibly mediated by ABA, and suggests involvement of DDB1-Cul4–mediated protein degradation in drought response. Given that the a priori odds of selecting a gene affecting the response to water deprivation are ~1 in 318 (currently only 85 of 27,029 *Arabidopsis* genes are annotated for response to water deprivation), these tests strongly support the network-based approach to rationally associate even entirely uncharacterized genes with plant traits.

AT3G05090 is a regulator of lateral root development

The second candidate gene, AT3G05090, was implicated in cell proliferation and meristem organization, drawing support from phylogenetic profiling of bacterial homologs of *Arabidopsis* proteins and domain co-occurrence patterns of yeast orthologs (**Supplementary Table 10**). We examined both shoot and root development in at3g05090-1 seedlings. We did not observe shoot phenotypes, but the number of lateral roots was significantly reduced ($P \leq 10^{-37}$, unpaired *t*-test, **Fig. 6d**,**e** and **Supplementary Fig. 9**). This phenotype segregated as a single recessive Mendelian locus linked to the T-DNA insertion (**Supplementary Table 12** and **Supplementary Fig. 10a**). The length of the primary root was shorter than in wild type (**Fig. 6d**) but this phenotype was unlinked to the T-DNA insertion (**Supplementary Fig. 10b**), showing that the lateral root phenotype is separable and independent from the primary root phenotype. We designate AT3G05090 as *LRS1* (*LATERAL ROOT STIMULATOR 1*) and the at3g05090-1 allele as *lrs1-1*. Homozygous lines transformed with a wild-type coding sequence driven under a 35S CaM virus promoter complemented the lateral root phenotype (**Fig. 6d** and **Supplementary Fig. 9**). To determine if the lateral root formation is blocked before the lateral root meristem emergence, we examined the number of lateral root primordia and meristems (lateral root stages IV–VIII[43]). Wild-type lateral roots are distributed fairly evenly among lateral root primordia, emerged lateral root and elongated lateral root (**Fig. 6e**). The mutant has reduced numbers of lateral roots at all of these stages, though the reduction is more severe in the

emerged and elongated lateral root than in the lateral root primordia (**Fig. 6e**). Transforming wild-type lines with the 35S::LRS1 construct did not increase the number of lateral roots, but we observed a dramatic increase in the length of the lateral roots and a decrease in the primary root length (**Fig. 6d** and **Supplementary Fig. 9**).

Regulation of root architecture and function, modulated by both intrinsic and extrinsic signals, is critical for efficient nutrient and water use for plants. Auxin, a plant hormone, is a key regulator for lateral root development, including lateral root initiation, primordium development and emergence[44]. The reduction in number of lateral roots in the mutant and the increase in lateral root length concomitant with the decrease in the primary root length in the overexpressed lines evoke defects in auxin accumulation or perception[44]. We thus asked whether exogenous auxin could alleviate the phenotype by growing plants in the presence of native auxin, indole acetic acid (IAA). IAA increased the lateral root number in the mutant (**Fig. 6d** and **Supplementary Fig. 11a**,**b**), demonstrating that auxin perception was not altered in the mutant and suggesting that auxin accumulation is compromised in the mutant. Auxin accumulation can be altered by changing synthesis, degradation, sequestration or transport[45]. To test for auxin transport defects, we examined effects of an auxin transport inhibitor, N-(1-naphthyl)phthalamic acid (NPA) on root growth. NPA decreases both the number and length of lateral roots in both genotypes (**Supplementary Fig. 11a**,**b**). *LRS1* encodes another DCAF protein[41], suggesting involvement of DDB1-Cul4–mediated protein degradation in lateral root development. These results demonstrate that the *lrs1-1* mutant is defective in lateral root development and suggest roles for DDB1-Cul4–mediated protein degradation in regulating auxin accumulation during lateral root primordium development and lateral root meristem emergence, consistent with its hypothesized roles in cell proliferation and meristem organization.

DISCUSSION

We demonstrate here that genes can be rationally associated with plant traits through guilt by association in a gene network. For this purpose, we created AraNet, a genome-wide gene network for *A. thaliana*, a reference organism for flowering plants, including many crops. AraNet is the most extensive gene network for any plant thus far; gene annotations derived by network guilt by association extend substantially beyond current gene annotations. We validated the network's predictive power by cross-validation tests, independent pathway and phenotype data sets, cell-specific expression data sets, and by experiments on computationally selected candidate genes.

AraNet generates at least two main types of testable hypotheses. The first type uses a set of genes known to be involved in a specific process as bait to find new genes involved in that process. This test is useful if the bait genes are well connected (that is, high AUC). We used the set of genes conferring seed pigmentation defects (AUC = 0.68) as bait and found a tenfold enrichment in identifying mutants with comparable phenotypes. Of the 318 GO biological processes with more than five genes, ~43% have AUCs of at least 0.68 (**Supplementary Table 13**), suggesting that AraNet will be useful in identifying new genes in nearly half of these biological processes. Similar distributions of AUCs are found for GO cellular component terms and KEGG pathways (**Supplementary Tables 14** and **15**). In practice, this translates into identifying a small set of new genes from a relatively limited-scale screen of the top network-predicted candidates (e.g., computer simulations suggest finding an average of 4–7 novel genes from tests of the top 200 candidates for biological processes with AUC >0.6; **Supplementary Fig. 12**). The second type of hypothesis involves predicting functions for uncharacterized genes. We assayed predicted phenotypes for three uncharacterized genes, two of which showed phenotypes in the predicted processes, response to drought and meristem development. There are 4,479 uncharacterized genes in AraNet (16% of protein-coding genes) with links to characterized genes, suggesting broad utility for AraNet in identifying candidate functions. Both of these modes of operation can be easily performed on the AraNet website.

Although AraNet currently shows high accuracy for many processes (**Figs. 2–4**), there are nonetheless specific processes that are poorly represented, with this trend stronger among plant-specific processes (**Fig. 2d**). This trend manifested in our experimental validation of only two of three tested candidate genes, although these intentionally represented challenging cases lacking any current functional annotation and for which sequence homology approaches had failed. Although we observed that non-plant–derived data sets helped identify genes for plant-specific processes, it is clear that more plant data sets will enhance the utility of gene networks for finding trait-relevant genes.

Three major causes underlie such cases of poor predictive performance. First, our current knowledge of genetic factors for a process may be so sparse that AraNet cannot link them efficiently. Second, AraNet may lack linkages or data relevant to the poorly predicted processes. These two trends likely explain the lower performance among plant-specific processes relative to more broadly studied, evolutionarily conserved processes. Additional plant-specific data sets, such as protein interactions, should help here, as should considering both indirect and direct network linkages for ranking candidates. Third, strongly implicated candidate genes that nonetheless test negative for a trait, resulting in apparent false positives, might be masked by epistatic effects, thus actually representing true predictions and false-negative assay results. This trend may be reasonably common and has been previously observed in yeast[46].

AraNet represents a step toward the goal of computationally identifying gene-trait associations in plants. This work suggests that gene networks for food and energy crops will facilitate manipulation of traits of economic importance and crop genetic engineering.

METHODS

Methods and any associated references are available in the online version of the paper at http://www.nature.com/naturebiotechnology/.

Note: Supplementary information is available on the Nature Biotechnology website.

ACKNOWLEDGMENTS

We are grateful to M. Ahn, A. Noorani and V. Bakshi for technical assistance, J. Shin for assistance on AraNet web design, T. Nakagawa (Shimane University, Japan) for providing pGWB2, K. Barton for providing lab space and D. Meinke, M. Running, W. Briggs, Z. Wang and K. Dreher for helpful discussions. This work was supported by Carnegie Institution for Science (B.A., S.Y.R.), a grant from the National Science Foundation (MCB-0520140) to S.Y.R. and by the National Research Foundation of Korea (NRF) grant funded by the Korean government (no. 2009-0063342, 2009-0070968) and Yonsei University (no. 2008-7-0284, 2008-1-0018) to I.L. and from the National Science Foundation, National Institutes of Health, and Welch (F1515) and Packard Foundations to E.M.M.

AUTHOR CONRIBUTIONS

I.L. and S.Y.R. conceived the project, I.L. created AraNet using approaches developed with E.M.M., B.A. performed the experimental tests, P.T. assisted in seed pigmentation mutant analysis, S.Y.R. supervised the experimental tests, I.L., E.M.M. and S.Y.R. analyzed AraNet and wrote the manuscript.

COMPETING INTERESTS STATEMENT

The authors declare no competing financial interests.

Published online at http://www.nature.com/naturebiotechnology/.
Reprints and permissions information is available online at http://npg.nature.com/reprintsandpermissions/.

1. Alonso, J.M. *et al.* Genome-wide insertional mutagenesis of *Arabidopsis thaliana*. *Science* **301**, 653–657 (2003).

2. Marcotte, E.M., Pellegrini, M., Thompson, M.J., Yeates, T.O. & Eisenberg, D. A combined algorithm for genome-wide prediction of protein function. *Nature* **402**, 83–86 (1999).
3. McGary, K.L., Lee, I. & Marcotte, E.M. Broad network-based predictability of *Saccharomyces cerevisiae* gene loss-of-function phenotypes. *Genome Biol.* **8**, R258 (2007).
4. Fraser, H.B. & Plotkin, J.B. Using protein complexes to predict phenotypic effects of gene mutation. *Genome Biol.* **8**, R252 (2007).
5. Lee, I. *et al.* A single gene network accurately predicts phenotypic effects of gene perturbation in *Caenorhabditis elegans*. *Nat. Genet.* **40**, 181–188 (2008).
6. Zhong, W. & Sternberg, P.W. Genome-wide prediction of *C. elegans* genetic interactions. *Science* **311**, 1481–1484 (2006).
7. Lage, K. *et al.* A human phenome-interactome network of protein complexes implicated in genetic disorders. *Nat. Biotechnol.* **25**, 309–316 (2007).
8. Franke, L. *et al.* Reconstruction of a functional human gene network, with an application for prioritizing positional candidate genes. *Am. J. Hum. Genet.* **78**, 1011–1025 (2006).
9. Linghu, B., Snitkin, E.S., Hu, Z., Xia, Y. & Delisi, C. Genome-wide prioritization of disease genes and identification of disease-disease associations from an integrated human functional linkage network. *Genome Biol.* **10**, R91 (2009).
10. Huttenhower, C. *et al.* Exploring the human genome with functional maps. *Genome Res.* **19**, 1093–1106 (2009).
11. Hermjakob, H. *et al.* IntAct: an open source molecular interaction database. *Nucleic Acids Res.* **32**, D452–D455 (2004).
12. Alfarano, C. *et al.* The Biomolecular Interaction Network Database and related tools 2005 update. *Nucleic Acids Res.* **33**, D418–D424 (2005).
13. Swarbreck, D. *et al.* The *Arabidopsis* Information Resource (TAIR): gene structure and function annotation. *Nucleic Acids Res.* **36**, D1009–D1014 (2008).
14. de Folter, S. *et al.* Comprehensive interaction map of the *Arabidopsis* MADS Box transcription factors. *Plant Cell* **17**, 1424–1433 (2005).
15. Huynen, M., Snel, B., Lathe, W. III & Bork, P. Predicting protein function by genomic context: quantitative evaluation and qualitative inferences. *Genome Res.* **10**, 1204–1210 (2000).
16. Pellegrini, M., Marcotte, E.M., Thompson, M.J., Eisenberg, D. & Yeates, T.O. Assigning protein functions by comparative genome analysis: protein phylogenetic profiles. *Proc. Natl. Acad. Sci. USA* **96**, 4285–4288 (1999).
17. Wolf, Y.I., Rogozin, I.B., Kondrashov, A.S. & Koonin, E.V. Genome alignment, evolution of prokaryotic genome organization, and prediction of gene function using genomic context. *Genome Res.* **11**, 356–372 (2001).
18. Bowers, P.M. *et al.* Prolinks: a database of protein functional linkages derived from coevolution. *Genome Biol.* **5**, R35 (2004).
19. Dandekar, T., Snel, B., Huynen, M. & Bork, P. Conservation of gene order: a fingerprint of proteins that physically interact. *Trends Biochem. Sci.* **23**, 324–328 (1998).
20. Overbeek, R., Fonstein, M., D'Souza, M., Pusch, G.D. & Maltsev, N. The use of gene clusters to infer functional coupling. *Proc. Natl. Acad. Sci. USA* **96**, 2896–2901 (1999).
21. Lee, I., Li, Z. & Marcotte, E.M. An improved, bias-reduced probabilistic functional gene network of baker's yeast, *Saccharomyces cerevisiae*. *PLoS ONE* **2**, e988 (2007).
22. Breitkreutz, B.J. *et al.* The BioGRID Interaction Database: 2008 update. *Nucleic Acids Res.* **36**, D637–D640 (2007).
23. Chatr-aryamontri, A. *et al.* MINT: the Molecular INTeraction database. *Nucleic Acids Res.* **35**, D572–D574 (2007).
24. Giot, L. *et al.* A protein interaction map of *Drosophila melanogaster*. *Science* **302**, 1727–1736 (2003).
25. Remm, M., Storm, C.E. & Sonnhammer, E.L. Automatic clustering of orthologs and in-paralogs from pairwise species comparisons. *J. Mol. Biol.* **314**, 1041–1052 (2001).
26. Ogata, H. *et al.* KEGG: Kyoto Encyclopedia of Genes and Genomes. *Nucleic Acids Res.* **27**, 29–34 (1999).
27. Cui, J. *et al.* AtPID: *Arabidopsis thaliana* protein interactome database an integrative platform for plant systems biology. *Nucleic Acids Res* **36**, D999–D1008 (2007).
28. Geisler-Lee, J. *et al.* A predicted interactome for *Arabidopsis*. *Plant Physiol.* **145**, 317–329 (2007).
29. Gutierrez, R.A. *et al.* Qualitative network models and genome-wide expression data define carbon/nitrogen-responsive molecular machines in *Arabidopsis*. *Genome Biol.* **8**, R7 (2007).
30. Ma, S., Gong, Q. & Bohnert, H.J. An *Arabidopsis* gene network based on the graphical Gaussian model. *Genome Res.* **17**, 1614–1625 (2007).
31. Brady, S.M. *et al.* A high-resolution root spatiotemporal map reveals dominant expression patterns. *Science* **318**, 801–806 (2007).
32. Meinke, D., Muralla, R., Sweeney, C. & Dickerman, A. Identifying essential genes in *Arabidopsis thaliana*. *Trends Plant Sci.* **13**, 483–491 (2008).
33. McElver, J. *et al.* Insertional mutagenesis of genes required for seed development in *Arabidopsis thaliana*. *Genetics* **159**, 1751–1763 (2001).
34. Wei, N., Serino, G. & Deng, X.W. The COP9 signalosome: more than a protease. *Trends Biochem. Sci.* **33**, 592–600 (2008).
35. Peng, Z., Serino, G. & Deng, X.W. Molecular characterization of subunit 6 of the COP9 signalosome and its role in multifaceted developmental processes in *Arabidopsis*. *Plant Cell* **13**, 2393–2407 (2001).
36. Gusmaroli, G., Figueroa, P., Serino, G. & Deng, X.W. Role of the MPN subunits in COP9 signalosome assembly and activity, and their regulatory interaction with *Arabidopsis* Cullin3-based E3 ligases. *Plant Cell* **19**, 564–581 (2007).
37. Gavin, A.C. *et al.* Proteome survey reveals modularity of the yeast cell machinery. *Nature* **440**, 631–636 (2006).
38. Krogan, N.J. *et al.* Global landscape of protein complexes in the yeast *Saccharomyces cerevisiae*. *Nature* **440**, 637–643 (2006).
39. Catala, R. *et al.* The *Arabidopsis* E3 SUMO ligase SIZ1 regulates plant growth and drought responses. *Plant Cell* **19**, 2952–2966 (2007).
40. Shinozaki, K. & Yamaguchi-Shinozaki, K. Gene networks involved in drought stress response and tolerance. *J. Exp. Bot.* **58**, 221–227 (2007).
41. Lee, J.H. *et al.* Characterization of *Arabidopsis* and rice DWD proteins and their roles as substrate receptors for CUL4-RING E3 ubiquitin ligases. *Plant Cell* **20**, 152–167 (2008).
42. Jin, J., Arias, E.E., Chen, J., Harper, J.W. & Walter, J.C. A family of diverse Cul4-Ddb1-interacting proteins includes Cdt2, which is required for S phase destruction of the replication factor Cdt1. *Mol. Cell* **23**, 709–721 (2006).
43. Casimiro, I. *et al.* Dissecting *Arabidopsis* lateral root development. *Trends Plant Sci.* **8**, 165–171 (2003).
44. Fukaki, H., Okushima, Y. & Tasaka, M. Auxin-mediated lateral root formation in higher plants. *Int. Rev. Cytol.* **256**, 111–137 (2007).
45. Vanneste, S. & Friml, J. Auxin: a trigger for change in plant development. *Cell* **136**, 1005–1016 (2009).
46. Li, Z. *et al.* Rational extension of the ribosome biogenesis pathway using network-guided genetics. *PLoS Biol.* **7**, e1000213 (2009).

ONLINE METHODS

All analyses are based on the set of 27,029 predicted protein coding loci of *Arabidopsis* (genome release version TAIR7)[13]. Reference and benchmark sets, raw data sets, and the construction and computational validation of AraNet are described in **Supplementary Methods** (**Supplementary Figs. 16–18**).

Targeted reverse genetics screening for seed pigmentation mutants. We searched AraNet with 23 confirmed seed pigmentation mutants (**Supplementary Table 4**) from the SeedGenes database[32] as bait. We retrieved the homozygous T-DNA insertional lines for the top 200 candidates from the SIGnAL database[1] and obtained the stocks from the *Arabidopsis* Biological Resource Center (**Supplementary Tables 5** and **6**). Seven to 9 seeds for each line were sterilized as described below (Plant material). Seeds were stratified at 4 °C for 2 d in the dark and grown in Murashige and Skoog (MS) medium with 1% sucrose under continuous illumination of 50–80 μmol/m^2 s at 22 °C. Seedlings were observed under a dissecting microscope (Leica MZ125) 6 d after germination and followed up 10–12 d after germination. For each of the lines where sufficient seeds were available, the assay was conducted at least twice. T-DNA insertions and genotypes of the seeds for the 11 lines with the mutant phenotypes described in this study (6 lines representing two alleles of three genes and 5 lines representing single alleles of five genes) were confirmed using PCR (**Supplementary Methods** and **Supplementary Table 16**).

To determine the significance of the discovery rate, we used the results of a large-scale screening of T-DNA insertional mutants for embryo defective or seed pigmentation mutants[32] as the background rate. This study found ~1,260 seed pigmentation mutants from screening 120,000 T-DNA lines. Because this was a forward-genetics screening whereas our screen was reverse-genetic (that is, preselected for intragenic insertions), we adjusted the total number of lines in the background to 84,000 based on the genome-wide distribution of T-DNA insertion sites of 70% insertion events in intragenic regions[1].

Candidate selection for uncharacterized genes and experimental validation of mutant phenotypes. To test the predictive power of AraNet *in planta*, we analyzed mutant phenotypes of genes of unknown function, whose biological roles were inferred by the annotations of the neighbors of these genes in AraNet. Of the 27,029 protein-encoding genes in *Arabidopsis*, 14,847 have no information about the biological processes in which they are involved. More than half of these uncharacterized genes (7,465 genes) are included in AraNet. Of these, 4,479 genes are inferred to be associated with specific biological processes based upon annotated AraNet neighbors (using only IDA, IMP, IGI, IPI, IEP and TAS evidence). To test the accuracy of such inferences made by AraNet, we chose three genes to characterize experimentally. These were chosen on the basis of available homozygous knockout lines, absence of paralogs and AraNet inferences of involvement in specific biological processes. The genes chosen were AT3G05090, AT1G80710 and AT2G17900, whose top ten AraNet predictions are shown in **Supplementary Table 10**. From the predictions, we assayed all of the phenotypes that could be measured with available resources at Carnegie. For AT1G80710, the following were tested: response to water deprivation (rank 3); trichome differentiation (rank 8); leaf development (rank 9). For AT3G05090, the following phenotypes were tested: trichome differentiation (rank 1); leaf development (rank 3); cell proliferation (rank 5); meristem organization (rank 8); and regulation of flower development (rank 10). For AT2G17900, the following were tested: meristem organization (rank 2); leaf morphogenesis (rank 3); hyperosmotic salinity response (rank 4); brassinosteroid mediated signaling (rank 5); multidimensional cell growth (rank 6); response to auxin stimulus (rank 7); detection of brassinosteroid stimulus (rank 8). In addition, we selected two genes randomly, AT1G15772 and AT2G34170, which were included in AraNet but were neighbors of other uncharacterized genes, to test specificity of all observed phenotypes.

Plant material. Seeds of homozygous T-DNA knockout mutants were obtained from the *Arabidopsis* Biological Resource Center. The stock numbers for the 118 seed pigmentation candidate genes are listed in **Supplementary Table 6**. Seeds for the five uncharacterized genes were SALK_059570C (AT3G05090), SALK_001238C (AT1G80710), SALK_127952C (AT2G17900), Salk_118634C (AT1G15772) and Salk_099804C (AT2G34170). For experiments conducted in soil, seeds were sown in soil (Premier Pro-mix) supplemented with fertilizer (Osmocote Classic, Hummert International). Seeds were stratified at 4 °C in the dark for 2 d and grown under 16/8 h of light/dark (90–100 μmol/m^2 s) and 30% humidity at 22 °C. For experiments conducted in agar plates, seeds were surface sterilized with 15% commercial bleach (6.25% sodium hypochlorite), containing a few drops of Tween-20 detergent and rinsed with sterile water five times. Seeds were sown on agar plates containing 0.43% MS salts, 0.5% 2-(N-Morpholino)ethanesulfonic acid (MES), 0.5% sucrose, 0.8% agar, pH 5.7. Plants on agar plates were grown under constant illumination of 50–80 μmol/m^2 s at 22 °C. For root assays, 50 ml of the MS medium was prepared poured into 100 × 100 × 15 mm square plates (Fisher Scientific) 1 d before planting to minimize plate-to-plate variability.

T-DNA insertions were confirmed, and genetic linkage, complementation, and overexpression tests were performed as described in **Supplementary Methods** (**Supplementary Tables 16** and **17**).

Visible phenotype assays. The following traits were observed by naked eye and using dissecting (Leica MZ125) and compound (Nikon Eclipse E600 with Nomarski optics) microscopes throughout the life cycle of the mutant plants: trichome differentiation (observations made on rosette and cauline leaves and sepals), leaf development and morphogenesis, cell proliferation, meristem organization and multidimensional cell growth (observations made on leaf, floral, inflorescence and root organs). To detect phenotypes in the regulation of flower development, we observed floral organs and flowering time under long days (16/8 of light/dark) and short days (8/16 of light/dark).

Hypocotyl length measurements. Seeds were germinated and grown vertically for 4 d in dark or under 4 μM/m^2 s continuous blue light. Seven to eight seeds of mutant and wild-type Col-0 (a reference strain that is the genetic background for the T-DNA mutants) were planted per plate and two plates per genotype were tested in each experiment. Each condition was tested in 7–8 independent experiments. Hypocotyl length was measured using ImageJ on photographs of the plates after 4 d of growth. The average hypocotyl length of each genotype was determined from each plate and the difference in hypocotyl length between wild type and mutant was determined using one-tailed, paired *t*-test.

Root length and number measurements. Seeds were germinated and grown vertically. Root measurements were taken 10–11 d after germination. Lateral roots were counted using a dissecting microscope or from digital images of plants using ImageJ. Different stages of the lateral roots were determined using a compound microscope. The root length was measured by tracing the length of the root using ImageJ on digital images of the seedlings.

Auxin response. Auxin and auxin transport inhibitor treatments were carried out as described[47]. Seeds were sown on MS agar medium and grown under continuous light. After 4 d, seedlings were transferred to either MS agar medium (control) or MS agar medium containing 1 nM, 5 nM, 10 nM or 30 nM indole acetic acid (IAA, Sigma) or 1 nM, 10 nM, 100 nM or 1 μM of naphthylphthalamic acid (NPA, Chem Service). Both wild type (Col-0) and mutant seedlings were transferred to the same plates. On the tenth or eleventh day of growth, the primary root length, number of lateral roots and length of lateral roots were measured as described above. Significant differences were determined by unpaired *t*-test. Each experiment was conducted with 2–3 plates of 7–10 plants each of wild type and mutant per plate. At least three independent experiments were carried out for each hormone assay.

Abscisic acid (ABA) response. The effect of ABA on detached leaf transpiration was determined as described[48] with some modifications: plants were grown in soil under long-day conditions (16/8 h light/dark) under white light of 90–100 μM m^{-2} s^{-1} at 22 °C. The largest, fully-open rosette leaves of 4–week-old plants were excised at the bottom of the petioles and were placed into a Parafilm-sealed 1.5 ml centrifuge tubes containing 1.4 ml of 0, 2.5 μM, 5 μM and 10 μM of ABA in an artificial xylem sap solution (15 mM KNO_3, 1 mM $CaCl_2$, 0.7 mM $MgSO_4$, and 1 mM $(NH_4)_2HPO_4$, with pH adjusted to 5.0 with 1 M phosphoric acid[49]). Transpiration was measured by weighing total weight of the tubes at times 0, 2, 4, 6 and 22 h. All of the excisions took place between 10 am and noon (4–6 h after the onset of illumination).

For the F2 linkage test of *drs1-1*, four leaves were excised from each plant and two were treated with the sap solution only and the other two were treated with 10 μM of ABA in the sap solution. Details on the linkage test are found in **Supplementary Methods**. For wild-type and mutant comparisons, each experiment used two to four leaves from three to four plants per genotype at each time point and was conducted in triplicate. Four independent experiments were conducted.

Drought response assay. Response to water deprivation was determined by measuring relative water content as described[50]. Plants were grown in soil under long-day conditions (16/8 h light/dark) under white light of 90–100 μM m^{-2} s^{-1} at 22 °C for 4–5 weeks. Watering was stopped for the drought treatment and relative water content was measured on day 0, 4, 7 and 10 of droughting. Control plants were watered every 2–3 d. To measure relative water content, we excised plants at the shoot/root junction, removed any bolts and weighed rosettes to determine the fresh weight (Fw). The rosettes were then completely submerged in water for 4 h and weighed to determine the turgid weight (Tw). Rosettes were then dried overnight at 80 °C and weighed to obtain the dry weight (Dw). Three plants from each genotype for each condition were measured. Relative water content was calculated as (Fw − Dw)/(Tw − Dw) and the significance of differences was determined by unpaired *t*-test. Three plants of each genotype were used for each time point per experiment. Four independent experiments were conducted.

An interactive web tool for AraNet-based candidate gene selection is available at http://www.functionalnet.org/aranet/.

47. Cho, H.T. & Cosgrove, D.J. Regulation of root hair initiation and expansin gene expression in *Arabidopsis*. *Plant Cell* **14**, 3237–3253 (2002).
48. Munns, R. & King, R.W. Abscisic acid is not the only stomatal inhibitor in the transpiration stream of wheat plants. *Plant Physiol.* **88**, 703–708 (1988).
49. Goodger, J.Q., Sharp, R.E., Marsh, E.L. & Schachtman, D.P. Relationships between xylem sap constituents and leaf conductance of well-watered and water-stressed maize across three xylem sap sampling techniques. *J. Exp. Bot.* **56**, 2389–2400 (2005).
50. Giraud, E. *et al.* The absence of ALTERNATIVE OXIDASE1a in *Arabidopsis* results in acute sensitivity to combined light and drought stress. *Plant Physiol.* **147**, 595–610 (2008).

doi:10.1038/nbt.1603