

CORNET: A User-Friendly Tool for Data Mining and Integration^{1[W]}

Stefanie De Bodt*, Diana Carvajal, Jens Hollunder, Joost Van den Cruyce, Sara Movahedi, and Dirk Inzé

Department of Plant Systems Biology, Flanders Institute for Biotechnology, and Department of Plant Biotechnology and Genetics, Ghent University, 9052 Ghent, Belgium

As an overwhelming amount of functional genomics data have been generated, the retrieval, integration, and interpretation of these data need to be facilitated to enable the advance of (systems) biological research. For example, gathering and processing microarray data that are related to a particular biological process is not straightforward, nor is the compilation of protein-protein interactions from numerous partially overlapping databases identified through diverse approaches. However, these tasks are inevitable to address the following questions. Does a group of differentially expressed genes show similar expression in diverse microarray experiments? Was an identified protein-protein interaction previously detected by other approaches? Are the interacting proteins encoded by genes with similar expression profiles and localization? We developed CORNET (for CORrelation NETWORKs) as an access point to transcriptome, protein interactome, and localization data and functional information on *Arabidopsis thaliana*. It consists of two flexible and versatile tools, namely the coexpression tool and the protein-protein interaction tool. The ability to browse and search microarray experiments using ontology terms and the incorporation of personal microarray data are distinctive features of the microarray repository. The coexpression tool enables either the alternate or simultaneous use of diverse expression compendia, whereas the protein-protein interaction tool searches experimentally and computationally identified protein-protein interactions. Different search options are implemented to enable the construction of coexpression and/or protein-protein interaction networks centered around multiple input genes or proteins. Moreover, networks and associated evidence are visualized in Cytoscape. Localization is visualized in pie charts, thereby allowing multiple localizations per protein. CORNET is available at <http://bioinformatics.psb.ugent.be/cornet>.

A high number of experiments have been performed to unravel molecular mechanisms underlying diverse biological processes active in the model plant *Arabidopsis thaliana*. For instance, in microarray experiments, diverse tissues from wild-type plants as well as mutant or transgenic plants are sampled at different developmental stages and treated with numerous compounds. Although enormous amounts of data have been generated, it remains a hurdle to sift through the heterogeneous information to find data relevant to a particular biological question. Data resulting from profiling studies are stored in different formats in various databases. Microarray data and, in particular, corresponding meta-data (e.g.

sampled tissue, time point, treatment) are stored in an unstructured manner, which complicates data retrieval and interpretation. Also, proteomics data, such as protein-protein interaction (PPI) data, are dispersed over several databases in somewhat different formats. Although some efforts have been made, such as setting up MIAME (for Minimum Information About a Microarray Experiment) and MIAPE (for Minimum Information About a Proteomics Experiment), these systems are not (yet) generally employed (Taylor et al., 2007; Brazma, 2009). Using transcript profiling data, one can investigate how genes are expressed, when genes are active and/or differentially expressed, and which other genes show similar expression profiles. Integration of microarray data with PPI data can, for instance, lead to the identification of protein complexes and/or coregulated genes, a better understanding of a group of differentially expressed genes, and the prediction of putative functions for unknown genes (Brown et al., 2005; Gachon et al., 2005; Lisso et al., 2005; Rautengarten et al., 2005; Usadel et al., 2009).

Databases and tools such as ACT (Manfield et al., 2006), AtCOECiS (Vandepoele et al., 2009), ATTED-II (Obayashi et al., 2007, 2009), Bio-Array Resource (BAR Toufighi et al., 2005), CressExpress (Srinivasasainagendra et al., 2008), CSB.DB (Steinhauser et al., 2004), GeneCAT (Mutwil et al., 2008), Genevestigator (Zimmermann et al., 2004), Plant Gene Expression Database (Horan et al., 2008), and PRIME (Akiyama et al., 2008) have been developed with the aim to easily find similarity

¹ This work was supported by Ghent University (Bijzonder Onderzoeksfonds Methusalem project no. BOF08/01M00408), the Interuniversity Attraction Poles Programme (grant nos. IUAP VI/25 [BioMaGNet] and VI/33) initiated by the Belgian State Science Policy Office, the European Union 6th Framework Programme (AGRONOMICS grant no. LSHG-CT-2006-037704), the Institute for the Promotion of Innovation by Science and Technology in Flanders-funded industrial collaboration project with BASF Cropscience, and the Research Foundation-Flanders (postdoctoral fellowship to S.D.B.).

* Corresponding author; e-mail stefanie.debodt@psb.vib-ugent.be.

The author responsible for distribution of materials integral to the findings presented in this article in accordance with the policy described in the Instructions for Authors (www.plantphysiol.org) is: Stefanie De Bodt (stefanie.debodt@psb.vib-ugent.be).

^[W] The online version of this article contains Web-only data.

www.plantphysiol.org/cgi/doi/10.1104/pp.109.147215

between genes based on their expression, and databases such as IntAct (Hermjakob et al., 2004), BioGRID (Stark et al., 2006), DIP (Salwinski et al., 2004), MINT (Chatr-aryamontri et al., 2007), BIND (Bader et al., 2003), BAR Arabidopsis Interactions Viewer (Geisler-Lee et al., 2007), and AtPID (Cui et al., 2008) have been built to easily retrieve PPIs. In the next paragraph, we describe the tools that have been developed to enable small-scale coexpression studies rather than the numerous large-scale coexpression studies often involving clustering. For a more detailed discussion on the coexpression tools as well as other databases and tools developed for hypothesis generation in plant biology, we point to two recent reviews (Brady and Provart, 2009; Usadel et al., 2009).

Current coexpression tools allow the visualization of gene expression profiles and/or the search for genes that are coexpressed with one or more genes of interest. To identify coexpression, these tools employ a measure, such as the Pearson correlation coefficient, a correlation rank (Obayashi et al., 2009), or linear regression (Srinivasasainagendra et al., 2008), followed by either applying an absolute cutoff or selecting the top x most correlated genes. The implementation of coexpression analysis is well advanced, with some tools that provide a flexible choice of input expression data sets. For instance, BAR Expression Angler allows the use of different types of expression data sets, among them the AtGenExpress compendia (Schmid et al., 2005; Toufighi et al., 2005; Kilian et al., 2007; Goda et al., 2008), whereas CressExpress allows the selection of microarray experiments based on tissue types (Srinivasasainagendra et al., 2008). Intuitively, the use of different expression data sets can yield different degrees of expression correlation between genes because some genes might behave similarly under certain conditions and differently under others. In other words, condition-dependent and condition-independent coexpression analyses have to be distinguished (Usadel et al., 2009). Therefore, a flexible and efficient compilation of the expression data sets used to calculate expression correlation needs to be enabled.

In contrast to the coexpression analysis, only a few tools provide additional functionalities, such as retrieval of PPIs, functions, pathways, and cis-regulatory elements, and the network visualization. The following tools have incorporated PPI data from one or more of the above-mentioned PPI databases. The output of the BAR Expression Angler displays Gene Ontology (GO) functional categories and PPI data from the BAR Arabidopsis Interactions Viewer (Toufighi et al., 2005; Geisler-Lee et al., 2007). ATTED-II provides PPIs, Kyoto Encyclopedia of Genes and Genomes pathway information, and cis-regulatory elements in addition to coexpression links (Obayashi et al., 2009). Virtual Plant provides a network analysis tool that compiles PPI data (BIND, interolog detection, and AtPID), microRNA:RNA associations, enzymatic reactions (both primary and secondary), and regulatory links based on binding site occurrence (Gutierrez et al., 2007).

To a large extent, the representation of the output determines the accessibility and interpretability of the results. The aforementioned tools came up with different solutions to represent coexpression and interaction data. In most tools, the output is in tabular format (such as in CressExpress [Srinivasasainagendra et al., 2008]). Although this format has many advantages for the advanced user who can import the results in other software tools, it does not allow immediate inspection of the results by less experienced users. With the BAR Expression Angler, the viewing and downloading of results are possible in both text and matrix formats (Toufighi et al., 2005), and with the DataMetaFormatter, functional classification of the coexpressed genes and PPIs are displayed on a clickable map of the matrix of coexpression data linking to other BAR tools. The BAR Arabidopsis Interactions Viewer allows the export of PPI networks to Cytoscape sif format. ATTED-II generates a network representation of the results (Obayashi et al., 2009). Although intuitively very comprehensive, the network views are static, ruling out visualization and exploration of large networks (Obayashi et al., 2007). In addition, network visualization is only possible in within-query gene searches. Only these small networks can be downloaded in tab-delimited, Pajek or Cytoscape sif formats (Shannon et al., 2003; de Nooy et al., 2005). In the latest version of ATTED-II, precalculated networks of particular genes can be viewed using the Google Maps API (Obayashi et al., 2009). PRIME allows coexpression analysis of multiple genes, provides the results in network files that can be viewed in dedicated software, such as Pajek (de Nooy et al., 2005) or Biolayout (Goldovsky et al., 2005), and thus allows the exploration of larger networks (Akiyama et al., 2008). The network analysis tool of Virtual Plant visualizes the resulting networks in Cytoscape Web Start (Gutierrez et al., 2007).

Taking into account all these features, we developed a new user-friendly tool for data mining and integration, with the acronym CORNET (for CORrelation NETworks), that is accessible through <http://bioinformatics.psb.ugent.be/cornet>. We collected the majority of the currently available microarray expression data; corresponding meta-data describing sampled tissues, treatments, and time points of sampling; PPI data; localization data; and functional information in a central database. A user-friendly interface allows one to query the database, enabling coexpression analysis through a multitude of search options addressing diverse biological questions. Several predefined expression data sets, such as global compendia representing diverse experimental conditions as well as tissue- or treatment-specific expression data sets, are provided. In addition, the user can compile expression data sets from public as well as private microarray data or can upload personal processed expression data sets. Directed selection of microarray experiments is possible, as all meta-data are described with standardized ontology terms and stored in the database. Not only is it possible

to calculate expression correlation based on one particular data set, but coexpression also can be assessed simultaneously among several expression data sets. PPI networks can be reconstructed with both experimentally identified and computationally predicted data. Moreover, coexpression and PPI networks can be integrated. CORNET generates a comprehensive visualization that provides a bird's eye view of the results and the different degrees of reliability of the extracted information. The tool makes use of Cytoscape Web Start, which has the advantage that all functionalities of Cytoscape itself and numerous plug-ins can be exploited to further explore the constructed networks (Shannon et al., 2003).

RESULTS AND DISCUSSION

Primarily, CORNET is composed of two tools, namely the coexpression tool and the PPI tool, constructing coexpression and PPI networks, respectively (Fig. 1). Both tools can be used autonomously but can also be used consecutively to build a network of coexpression links as well as PPIs. Additionally, localization and functional information (GO terms and protein domain information) can be displayed on the constructed networks.

Annotation of Microarray Experiments

All expression data available at Gene Expression Omnibus (Barrett and Edgar, 2006) and resulting from experiments carried out on Affymetrix ATH1 arrays were incorporated into the CORNET database. As the meta-data of the microarray experiments available in public databases are very unstructured and hard to process automatically, and as information on growth conditions, treatments, sampled tissues, and genotypes is difficult to retrieve, we described the meta-data of the microarray experiments by manually assigning ontology terms. Existing ontologies were exploited to avoid confusion and redundancy. Plant Ontology (Plant Structures and Plant Growth and Developmental Stages; Bruskiewich et al., 2002; Pujar et al., 2006; Ilic et al., 2007; Avraham et al., 2008), Plant Environmental Conditions (www.gramene.org), and MGED Ontology (Whetzel et al., 2006) were used (Supplemental Fig. S1). The final aim of this ontology-based annotation was the automatic and comprehensive retrieval of microarray experiments, similar to the selection of microarray experiments in Genevestigator (Zimmermann et al., 2004). With these microarray data, we compiled different, so-called predefined expression compendia (see "Materials and Methods"). Compendium 1 is a set of microarray experiments covering diverse conditions but somewhat biased toward growth and development. Compendium 2 also contains diverse conditions, but biases toward particular design types (see "Materials and Methods") are reduced as much as possible. In addition, highly re-

dundant experiments are removed, resulting in a set of approximately 100 experiments. Finally, we compiled several specific expression compendia (abiotic stress, biotic stress, development, flower, genetic modification, hormone, leaf, root, seed, and abiotic plus biotic stress data sets) using the design types and ontology terms (see "Materials and Methods"). The user can temporarily upload personal raw expression data and annotate and incorporate the data into user-defined expression data sets (see "Materials and Methods").

Coexpression Tool

Using the coexpression tool, genes with similar expression profiles in a number of experimental conditions can be identified. When performing a coexpression analysis, first, one needs to decide on the input expression data that will be used. The user can either select the predefined expression compendia or compile user-defined data sets. In the latter option, the user is directed to the "Browse experiments" page, where a set of microarray experiments can be assembled using ontology terms that describe the meta-data (Supplemental Fig. S1). The ontology terms allow an easily reproducible and intuitive selection of the microarray experiments without going through each individual experiment. Users should keep in mind that user-defined expression data sets should be large enough to enable reliable calculation of the correlation coefficients (Usadel et al., 2009). In the next step, the coexpression tool page is displayed, where one or more genes can be introduced for coexpression analysis (step 1; Supplemental Fig. S1). Subsequently, one or more of the predefined, previously generated user-defined, or personal, preprocessed expression data sets needs to be selected (step 2). Below, more details are given on the importance of different expression compendia. Next, one can choose to calculate either Pearson or Spearman correlation coefficients (Spearman can only be chosen when one expression compendium is selected; see "Materials and Methods"). Then, thresholds to limit the number of results and search options, referring to different biological questions, can be chosen (Fig. 2). Either an absolute or a relative threshold or both can be chosen, namely a correlation coefficient threshold and/or a number of the most highly coexpressed genes (top x), respectively. In the case of multiple compendia, each compendium is treated separately in a first step, and subsequently, these results are combined, depending on the "all/at least" parameter. When coexpression in at least one compendium is chosen (at least = 1), all coexpression links found in the different compendia are reported (union). When coexpression needs to hold true for multiple expression compendia (at least >1), the intersection of the individual results is reported. In addition, average, minimum, and maximum correlation coefficients over all expression compendia that meet the chosen thresholds are reported. The

search options entail the calculation of expression correlation in a pairwise manner between given genes or a list of gene pairs uploaded as a tab-delimited file

(“Pairwise correlations”), between one or more given genes and all genes in the genome [“Correlation of query gene(s) with neighbors”], and between genes

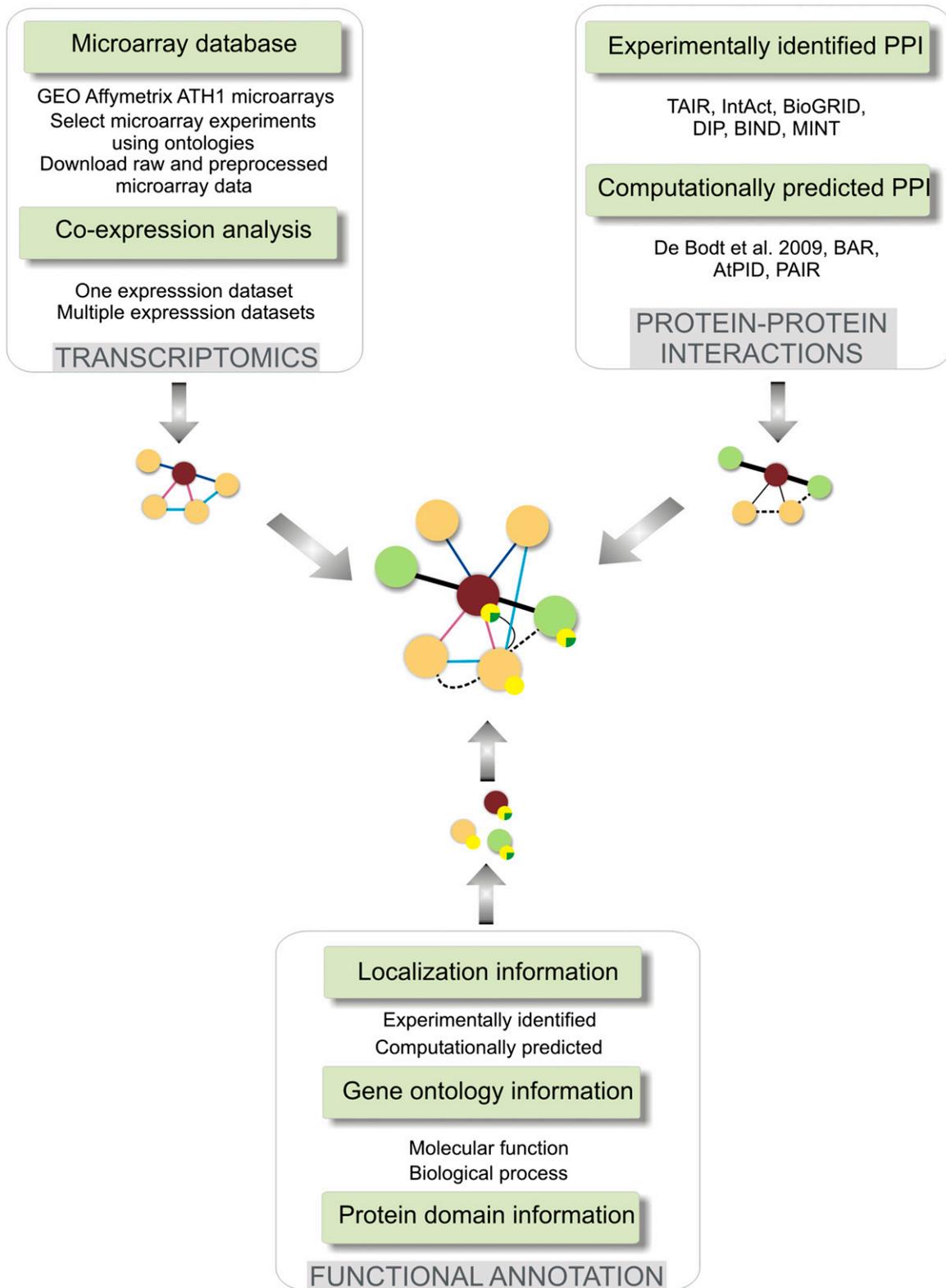


Figure 1. Functionalities of CORNET.

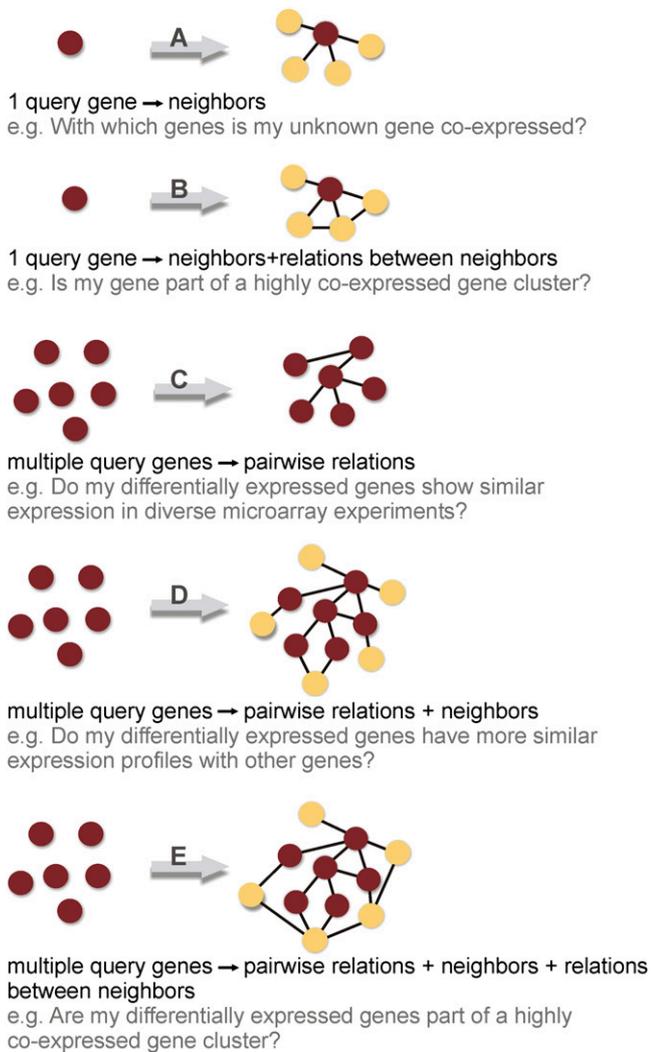


Figure 2. Search options of coexpression and PPI tools. A, Correlations or interactions of query gene/protein with neighbors. B, Correlations or interactions of query gene/protein with neighbors plus correlations or interactions between neighbors. C, Pairwise correlations or interactions of query genes/proteins with neighbors. E, Pairwise correlations or interactions plus correlations or interactions of query genes/proteins with neighbors plus correlations between neighbors.

that show coexpression with the given gene(s) (“Correlations between neighbors”; Fig. 2). These search options are more flexible and intuitive than the functionalities of most existing coexpression tools, which allow the search for coexpressed genes of one or a few query genes. Evidently, thresholds need to be chosen carefully when the two latter search options are selected because huge amounts of results can be generated. We advise starting off by choosing a relatively high correlation coefficient threshold (such as 0.9) or the top 10 genes and gradually lowering the stringency. Finally, one can integrate localization information and/or PPIs using the PPI tool as discussed below.

Selection of Multiple Microarray Expression Compendia

When comparing the expression profiles of two genes, the input microarray data are expected to influence the observations (Usadel et al., 2009). For instance, due to pleiotropic functions or the combinatorial nature of cis-regulation, some genes may coexpress under certain conditions but differ in expression under others and/or coexpress with other genes under these conditions. Using a global measure of coexpression, such as the Pearson correlation coefficient, all conditions in the input expression compendium are used. Thus, depending on the nature of the studied genes and the interest of the user, different input expression compendia can be imagined, some being subsets of each other. For instance, when looking for genes that are similar to a drought stress-responsive gene, an expression compendium representing abiotic stress conditions can be used to identify specific and relevant relations.

To investigate the variability of expression correlation over different expression compendia, several subsets of expression data, namely the abiotic stress, biotic stress, development, flower, genetic modification, hormone, leaf, root, seed, and abiotic plus biotic stress sets, as well as three global compendia, namely the AtGenExpress compendium, Compendium 1, and Compendium 2, were considered (Table I; see “Materials and Methods”) and are provided as predefined compendia in CORNET. We observe that expression correlation can vary with the input expression compendia independently from the chosen threshold (Fig. 3). Overall, few coexpression links held true when taking into account multiple compendia. Twenty-five percent to 30% of all gene pairs are correlated based on one expression compendium, while about 10% of the correlated gene pairs also show correlation in three other expression compendia, and this number gradually drops when more expression compendia are taken into account. Only 0.01% to 0.05% of the gene pairs show coexpression in all 14 compendia simultaneously (Fig. 3). The majority of the genes that show coexpression over all compendia encode ribosomal proteins or are involved in photosynthesis (Pearson correlation coefficient threshold of 0.9). The coexpressed genes have a high average expression level and are highly variable across conditions. Based on this study, we can conclude that, when performing coexpression analyses, expression correlation can vary significantly when using different expression conditions. Consequently, coexpression tools need to enable the estimation of coexpression over diverse expression compendia or in specific expression compendia. The identification of coexpression under specific conditions can be employed to seek genes with characteristics common to genes of interest (such as similar expression upon abiotic stress treatment). In contrast, coexpression analysis can be carried out with a collection of expression compendia, representing diverse conditions, and lead to the identification of those con-

Table 1. Predefined expression compendia

Expression Compendium	Description	No. of Experiments
AtGenExpress	All experiments performed by AtGenExpress	425
Compendium 1	Collection of microarray experiments oriented toward growth, development, and cell cycle studies	454
Compendium 2	Collection of microarray experiments; very similar experiments were removed; similar numbers of experiments of each design type	111
Abiotic stress	Abiotic stress series (cold, drought, genotoxic, heat, osmotic, oxidative, salt, UV-B, wounding)	256
Biotic stress	Biotic stress series (<i>Botrytis</i> , <i>Pseudomonas</i> , <i>Phytophthora</i> , etc.)	69
Abiotic plus biotic stress	Combination of the abiotic and biotic stress data sets	336
Hormone treatment	Hormone treatment series (abscisic acid, brassinosteroids, GA, cytokinin, etc., and inhibitors)	140
Development	Developmental series (different tissues, developmental stages, developmental mutants)	135
Flower	Microarray experiments in which floral tissues are sampled	72
Leaf	Microarray experiments in which leaf tissues are sampled	212
Root	Microarray experiments in which root tissues are sampled	258
Seed	Microarray experiments in which seed tissues are sampled	83
Whole plant	Microarray experiments in which whole plants are sampled	85
Genetic modification	Microarray experiments in which transgenic lines are profiled (gene overexpression [knockin], gene knockout, transient transgene expression)	313

ditions in which the genes of interest show similar expression patterns. For genes with limited functional information, expression compendia delivering high correlations hint at possible functional activities of the unknown genes. Using CORNET, this can either be done using the predefined expression compendia or user-defined compendia individually or by considering several predefined data sets. In the latter, coexpression links found in all selected compendia or in at least x compendia are reported (see above). As such, coexpression in, for instance, root only, leaf only, or both root and leaf can be studied in one analysis.

PPI Tool

As for the coexpression tool, the PPI tool needs one or more proteins as input (step 1; Supplemental Fig. S1). Next, different PPI databases can be chosen (step 2; see “Materials and Methods”) to extract only experimentally identified PPIs, only computationally predicted interactions, or both. As in the coexpression tool, different search options can be selected, namely search for PPI in a pairwise manner (Pairwise interactions), search for proteins that interact with the given protein(s) [“Interactions of query protein(s) with neighbors”], and/or search if the proteins that interact with the given protein(s) also interact (“Interactions between neighbors”; step 3; Fig. 2). Finally, also here, integration of coexpression and localization information is possible.

Integration of Coexpression and PPI Networks

CORNET allows the integration of coexpression and PPIs networks, which can be approached in two ways. One can start with the coexpression tool, identify

ing genes that coexpress with each other, and subsequently test if the corresponding proteins interact and if these proteins interact with other proteins or vice versa. One needs to keep in mind that the order of the analysis will yield different results when choosing the option “Correlations/interactions with neighbors.” For instance, when a coexpression analysis is followed by a PPI search, all coexpressed genes of the first analysis are used as input for the PPI search and genes that do not show coexpression with other genes are not included as input for the PPI tool. Conversely, when first performing a PPI search, only proteins for which interactions have been found will be used as input for the subsequent coexpression analysis. Depending on the question to be addressed, one or both approaches can be opted for.

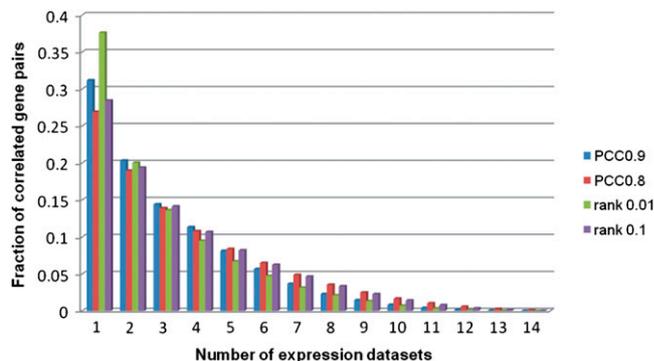


Figure 3. Variability of expression correlation. Only a few genes are coexpressed based on multiple expression compendia. Expression correlation is defined by a Pearson correlation threshold of 0.9 (blue), 0.8 (red), top 1% most correlated gene pairs (green), and top 10% most correlated gene pairs (purple).

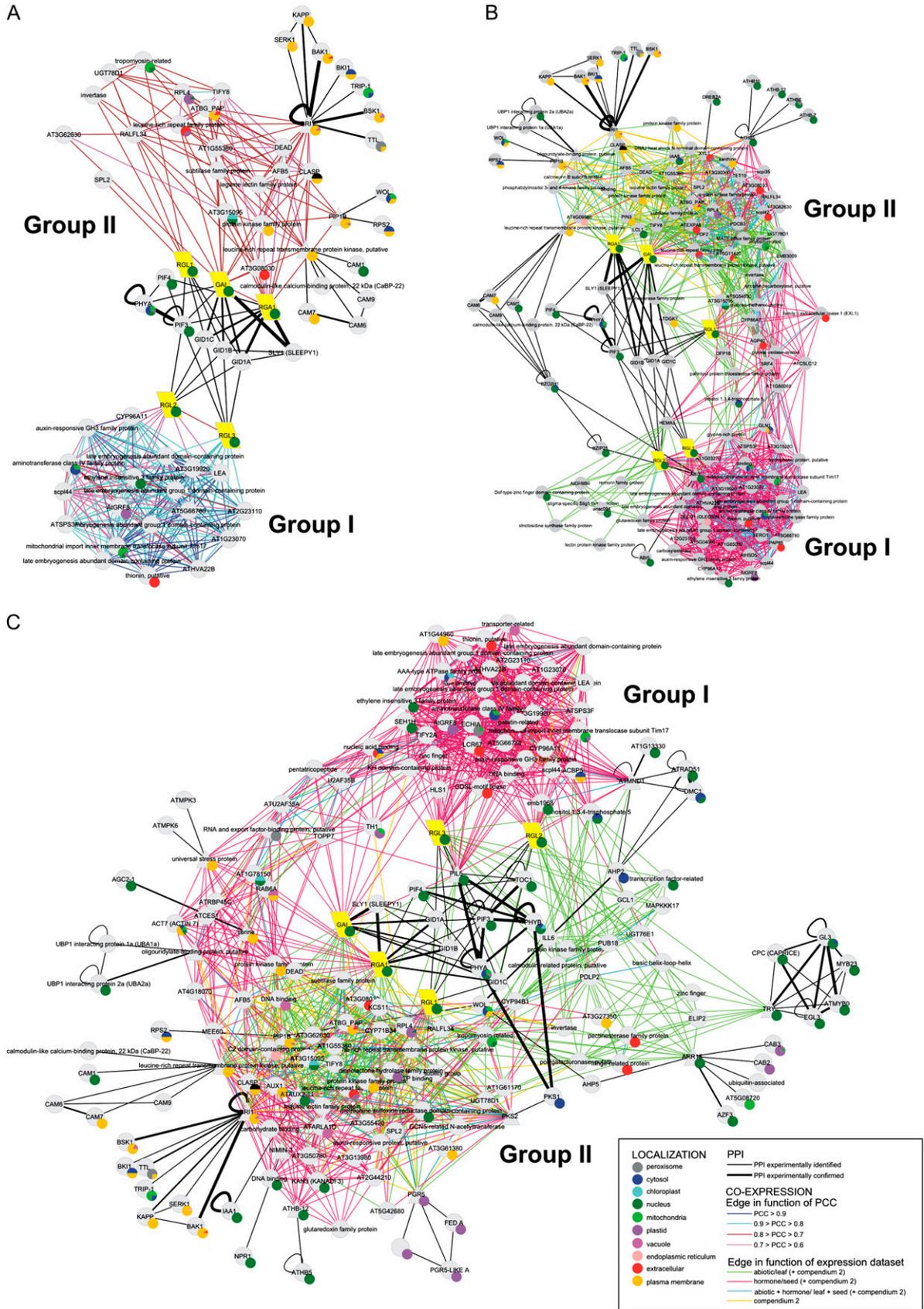


Figure 4. (Legend appears on following page.)

When integrating coexpression and PPI data, the degree of coexpression of genes encoding for interacting proteins can be studied using CORNET. A global study of all experimentally identified PPIs showed the relatively low concordance between coexpression and interactions (see Fig. 2 in De Bodt et al., 2009). This observation can be confirmed using the different expression compendia available in CORNET. The mean correlation coefficient for experimentally identified PPIs ranges between 0.13 and 0.21. As previously mentioned in the literature, this low degree of expression similarity is probably due to the transient nature of PPIs.

Visualization of Coexpression and PPI Networks

For network visualization, the existing software Cytoscape was designated (Shannon et al., 2003; Fig. 1) because its functionalities allow browsing and zooming into the constructed networks, a visual as well as textual representation of diverse attributes (e.g. correlation coefficient, localization databases; see Frequently Asked Questions [FAQ] page and VizMapper in Cytoscape) and further exploration and analysis of the networks. The degree of expression correlation is represented by the color of the edges (or correlations) in the coexpression network, where, for instance, blue edges correspond to highly correlated expression profiles. In case multiple expression compendia were selected in the coexpression tool, the minimum, maximum, or average Pearson correlation coefficient (as requested by the user) is displayed. However, all retrieved Pearson correlation coefficients are reported in the text output and as composite attributes (data set coefficients) in Cytoscape. PPIs are depicted by black edges, whereas their reliability can be assessed through the width (number of data sources) and the style (detection method) of the edges (or interactions) in the PPI network. Experimentally identified interactions are represented by solid lines, while computationally predicted interactions are represented by dashed lines. The shape of the nodes (or genes/proteins) depicts the nature of the gene/protein as query or as neighbor in either the coexpression or PPI tool. All attributes can be displayed in the lower data panel of Cytoscape by clicking on the Select Attributes button (see FAQ) and copied and/or exported in tabular format at any time.

Case Study: DELLA Network

To demonstrate the functionalities of CORNET and, particularly, the use of multiple expression compendia in the coexpression tool, we have investigated the molecular context of DELLA proteins. The Arabidopsis genome encodes five DELLA proteins (REPRESSOR OF GA [RGA], GIBBERELLIC ACID INSENSITIVE [GAI], RGA-LIKE1 [RGL1], RGL2, and RGL3) that act as negative regulators of the GA signaling pathway. DELLA proteins bind to the GIBBERELLIN INSENSITIVE DWARF1 (GID1) receptor in the presence of GA. The GA-GID1-DELLA complex is then targeted by an SCF-E3 ligase (SLEEPY), resulting in ubiquitinylation and degradation of DELLA by the 26S proteasome, thereby relieving DELLA-mediated repression of GA responses. DELLAs are involved in diverse processes such as flower development, seed germination, leaf growth, and abiotic stress response (Tyler et al., 2004; Achard et al., 2006; Schwechheimer, 2008; Achard and Genschik, 2009). Although the functional differences between the DELLAs are not completely clear, RGA and GAI have been shown to repress stem elongation (Dill and Sun, 2001; King et al., 2001), RGL2 inhibits seed germination (Lee et al., 2002), and RGA, RGL1, and RGL2 together regulate floral development (Cheng et al., 2004; Tyler et al., 2004; Yu et al., 2004). Whereas RGA and GAI are highly expressed in most tissues, *RGL1*, *RGL2*, and *RGL3* are mainly expressed in germinating seeds, young seedlings, and flowers, indicating that these signaling molecules might be transcriptionally regulated (Tyler et al., 2004). Thus, DELLA proteins are highly pleiotropic and serve as an ideal test case.

Different functions of the DELLA proteins surface when employing particular expression compendia (Fig. 4). First, we have generated a network using Compendium 2, which represents diverse conditions and gives a global estimate of coexpression. Subsequently, experimentally identified PPIs were added between the resulting proteins and all other proteins (neighbors) as well as between those other interacting proteins (neighbors). Two main groups of highly connected genes could be delineated in this network (Fig. 4A). The degree of correlation between the genes of both groups clearly differed (blue edges [group I] versus red edges [group II]). Group I consists of RGL2 and RGL3 and several LEA (for late embryogenesis abundant) genes with roles in seed development,

Figure 4. DELLA networks generated by CORNET using multiple compendia. The DELLA network based on Compendium 2 (A) shows two groups of coexpressed genes where group I is highly coexpressed (blue edges) and group II shows a lower degree of coexpression (red edges). The networks based on Compendium 2 plus abiotic stress plus hormone treatment compendia (B) and Compendium 2 plus leaf plus seed compendia (C) reveal new coexpression links and new coexpressed genes. Red edges represent coexpression links found in the hormone (B) and seed (C) compendia, and green edges represent expression links found in the abiotic stress (B) and leaf (C) compendia. Edges starting and ending at the same node represent homodimerization of proteins.

Table II. Statistics on the CORNET database

Data	No.
Microarray	
Arrays	3,055
Experiments	1,209
Series	200
Experiments with two replicates	634
Experiments with more than two replicates	575
PPIs	
Experimentally identified interactions	4,302
Computationally predicted interactions	89,181
Total interactions	93,109
Localization	
Experimentally identified localizations	14,135
Computationally predicted localizations	11,310
Total localizations	18,824

while group II contains GAI, RGA, and RGL1, some hormone-related genes such as BRASSINOSTEROID INSENSITIVE1 and AFB5 (auxin F-box), and CLIP-associating protein-related genes involved in mitosis, growth, and protein stability, as well as many other genes. Both groups are connected to each other through PPI links between the DELLAs and the GID receptors, SLEEPY1, PHYTOCHROME B, PHYTOCHROME INTERACTING FACTOR3, and PIF4, interactions that have been identified by various experimental approaches (Supplemental Files S1, “Features” attribute in Cytoscape; Dill et al., 2004; Fu et al., 2004; Griffiths et al., 2006; Nakajima et al., 2006; Achard et al., 2007; Ariizumi et al., 2008; de Lucas et al., 2008; Feng et al., 2008). Next, we investigated how this network varied with different, more specific expression compendia, namely the abiotic stress and hormone treatment compendia on the one hand (Fig. 4B) and the leaf and seed compendia on the other hand (Fig. 4C; see “Materials and Methods”), again followed by a search for experimentally identified PPIs. Strikingly, a high number of new coexpression links and coexpressed genes were identified that had not been found with Compendium 2 (Fig. 4, B and C; Supplemental Tables S1 and S2). Nevertheless, the two groups of genes and especially the degree of expression correlation between the respective genes were still demarcated, although some coexpression links between the two groups appeared. When examining the expression compendia that supported the different coexpression links, an additional distinction between group I and group II could be noticed. Group I links were mainly retrieved using the hormone treatment and the seed compendia, while some group II links were found using the abiotic stress and leaf compendia (Fig. 4, B and C). In other words, genes in each group had similar expression patterns in particular tissues and conditions and probably had associated functions primarily in those tissues and conditions. Through CORNET, the expression profiles of these genes for the different expression compendia could

be explored to further pinpoint the possible similarities and functions of the two groups. Group I genes showed similar and high expression in the imbibition stage (the first stage of seed germination) as well as in later stages of embryo development. In addition, these genes were highly expressed in gibberellic acid and abscisic acid experiments, in which seeds or embryos were sampled, as well as in methyl jasmonate and cytokinin experiments, with sampling of stamens and shoots, respectively.

Accordingly, Cao et al. (2006) observed that the set of genes that are presumably regulated by DELLAs for seed germination overlaps little with and is largely distinct from the set of DELLA-regulated genes involved in floral development when investigating transcript profiling results from *gal-3* and *gal-3 gai-t6 rga-t2 rgl1-1 rgl2-1* seeds and flowers. This observation suggests that GA-mediated seed germination and floral development are under the control of distinct DELLA-dependent transcriptomes. Our analysis suggests a similar distinction between the sets of genes involved in processes ongoing in seed and leaf and between sets of genes involved in hormone and abiotic stress-related processes.

CONCLUSION

We have developed CORNET, a tool for the construction of coexpression and PPI networks and their functional annotation in Arabidopsis (bioinformatics.psb.ugent.be/cornet). With this tool, we aim at providing biologists with the means to investigate the associations between genes and between encoded proteins. Thereby, we provide the ability to better understand the functional context of a gene, leading to function prediction of unknown genes or prediction of (indirect) regulatory interactions between known genes and proteins.

In the case study of DELLA proteins, coexpression and PPI networks demonstrate the importance of careful coexpression analysis. Both the correct selec-

Table III. Number of microarray experiments in each design type

Design Type	No. of Experiments
Stimulus_or_stress_design	13
Abiotic_stress_design	255
Biotic_stress_design	322
Compound_treatment_design	23
Hormone_treatment_design	140
Genetic_modification_design	314
Growth_condition_design	69
Time_series_design	357
Circadian_rhythm_design	15
Development_or_differentiation_design	134
Individual_genetic_characteristics_design	3
Organism_part_comparison_design	121
Strain_or_line_design	22
Translational_bias_design	2

tion of expression compendia and of adequate search options and thresholds can greatly enhance the power of coexpression tools to uncover new associations between genes and between genes and processes/microarray conditions. Moreover, the comprehensive visualization, as illustrated by Figure 4, allows a bird's eye view of the constructed networks, instantly discovering those associations without any requirement for bioinformatics skills, such as scripting or complicated database queries.

In the future, we plan to add different data types as they become available. First, we foresee the incorporation of cis-regulatory elements, which can be tightly linked with the coexpression results, as is worked out in AtCOECiS (Vandepoele et al., 2009). Second, through comparative genomics approaches, constructed networks can be transferred to other plant species, such as poplar (*Populus* species), tomato (*Solanum lycopersicum*), rice (*Oryza sativa*), and crops of interest. Last, the inclusion of protein-DNA (or transcription factor-target) interactions will be very profitable, as data are generated through chromatin immunoprecipitation (ChIP)-chip and ChIP-seq or indirectly through the analysis of transcriptome data.

MATERIALS AND METHODS

CORNET Troubleshooting

CORNET can be accessed through the following URL: <http://bioinformatics.psb.ugent.be/cornet>. The tool is fully functional in Firefox and Safari browsers. First users might need to accept a security certificate before accessing the Web site. The site is ideally viewed at $1,280 \times 1,024$ resolution. You need to allow pop-ups in your browser before clicking the "go" button. After calculations and database queries, Cytoscape will start automatically from the Web. In other words, Cytoscape does not have to be installed on your computer. However, to enable the Cytoscape WebStart, an up-to-date version of Java is required. More details can be found on our FAQ page.

Construction of the Microarray Database

All expression data available at Gene Expression Omnibus (Barrett and Edgar, 2006) and resulting from experiments with Affymetrix ATH1 arrays were uploaded in the CORNET database (Table II). Only experiments where two or more replicates were performed are included. Meta-data were described using ontology terms (for more details, see "Results and Discussion"). However, in some instances, these ontologies did not suffice to describe the microarray experiments. Existing ontologies were extended as necessary to allow a more detailed description of tissues, transgenic lines, and experimental designs (indicated by "EXT"). The experimental designs proved very valuable in the compilation of specific expression compendia (e.g. development_or_differentiation_design, genetic_modification_design, compound_treatment_design, abiotic_stress_design, biotic_stress_design, time_series_design, hormone_treatment_design; see below; Table III).

Microarray Experiment Browsing and Retrieval

Microarray experiments and their meta-data can be browsed through CORNET (Browse experiments). Using a tree-based representation, the ontology terms can be browsed and selected in order to compile specific sets of microarray experiments. For each microarray experiment, the different replicates and the link between control and treatment arrays are shown. On the one hand, raw data together with their annotation can be downloaded. On the other hand, processed data can be downloaded or used as input for the coexpression tool. The microarray data are processed with the Robust Multiarray Average procedure implemented in BioConductor (Irazzary et al., 2003a,

2003b; Gautier et al., 2004; Gentleman et al., 2004) using an alternative Chip Description File that takes into account possible cross-hybridization (tinsath1cdf; Casneuf et al., 2007).

Uploading Personal Expression Data

CORNET allows the incorporation of personal expression data ("Upload"). Personal data can be uploaded temporarily, processed with Robust Multiarray Average, and downloaded for later use. In addition, the uploaded data can be combined with public expression data in user-defined data sets and subsequently downloaded for later use and/or used as input for the coexpression tool. The submitter can access these personal data for 24 h after submission. The personal data are linked to the session identifier of the browser.

Predefined Expression Compendia

The coexpression tool of CORNET makes use of the following predefined expression compendia. First, an expression data set ("AtGenExpress") including the publicly available AtGenExpress compendia (such as abiotic stress, pathogen, development, and hormones) is provided (Schmid et al., 2005; Kilian et al., 2007; Goda et al., 2008). Second, we have compiled two different microarray compendia, which cover diverse conditions and consequently can be used to estimate an overall degree of coexpression taking into account all conditions. The first compendium (Compendium 1) covers diverse conditions that mainly deal with plant development and growth. The second compendium (Compendium 2) was built to reduce the bias toward particular conditions as much as possible. Moreover, the redundant information that microarray experiments can generate is assessed by calculating the correlation between experiments rather than between genes. A cutoff of 0.99 is used to identify redundant experiments. From each group of redundant experiments, one or more experiments are selected, taking the experiments with the highest number of replicates and the noncontrol experiments. Subsequently, an equal number of experiments of each type of condition (according to the assigned "design" terms) is selected. In addition to these "global" compendia, we have compiled specific expression data sets using the design and ontology terms mentioned above. As such, expression compendia were generated that are specific to certain tissues (leaf, root, flower, seed, and whole plant) or certain treatments (abiotic stress, biotic stress, and hormone treatment), compendia of experiments in which development is perturbed, or in which genes are modified (overexpression, knockout, or silencing lines; Tables I and III). Expression profiles for particular genes can be viewed in automatically generated line graphs.

Coexpression Tool

To quantify the similarity in expression profiles, the commonly used Pearson and Spearman correlation coefficients can be calculated. Pearson is a parametric method based on actual expression values, while Spearman is a nonparametric method based on ranks. Both measures range from -1 (anticorrelation) over 0 (no correlation) to 1 (correlation). When only one expression compendium is chosen for coexpression analysis, correlation coefficients are calculated in real time. However, when more than one expression compendium is chosen, it is no longer feasible to do the calculations in real time. Alternatively, the correlation coefficients are extracted from the database containing precalculated Pearson correlation coefficients higher than 0.6 and lower than -0.6 (this limitation is due to space and time constraints).

Arabidopsis Genome Initiative codes are used to describe the genes. Only genes represented on the Affymetrix ATH1 array are taken into account in the coexpression analysis (in contrast to the PPI tool, where all proteins are taken into account). In total, 20,777 genes are accounted for using an alternative Chip Description File that takes into account possible cross-hybridization (tinsath1cdf; Casneuf et al., 2007). There is no limit to the number of genes that can be given as query.

PPI Tool

We have assembled currently available experimentally identified PPIs for *Arabidopsis thaliana* from BIND (Bader et al., 2003), IntAct (Hermjakob et al., 2004), BioGRID (Stark et al., 2006), DIP (Salwinski et al., 2004), MINT (Chatr-aryamontri et al., 2007), and The Arabidopsis Information Resource (Rhee et al., 2003), the predicted PPIs from BAR (Geisler-Lee et al., 2007) and AtPID (Cui et al., 2008), and the filtered (high-stringency) and

predicted (low-stringency) interactions identified in our own study (De Bodt et al., 2009; Table II). To be able to grasp the reliability of the PPIs, we distinguish between experimental and predicted PPIs and indicate the different data sources (database, experiment type, evidence code, and PubMed identifier) as edge attributes in Cytoscape (see “Results and Discussion”; Shannon et al., 2003). Where possible, gene names are mapped on the nodes of the network. Alternatively, gene descriptions are shown. Gene names and full gene descriptions downloaded from The Arabidopsis Information Resource are stored as “description” and “descriptionLong” attributes in Cytoscape.

Gene Information

To allow easy interpretation of the constructed networks, we add localization data and functional information to the genes. The localization data are a collection of both experimental and predicted localizations retrieved from SUBA (Heazlewood et al., 2007), IPSort (Bannai et al., 2002), LocTree (Nair and Rost, 2005), MITOPRED (Guda et al., 2004), MitoProt (Claros, 1995), MultiLoc (Høglund et al., 2006), PeroxP (Emanuelsson et al., 2003), Predotar (Small et al., 2004), SubLoc (Chen et al., 2006), TargetP (Emanuelsson et al., 2007), and WoLF_PSORT (Horton et al., 2007; Table II). These localization data are depicted in pie charts, allowing multiple localizations for one gene. The fractions of the pie chart are based on the fraction of databases in which a particular localization was found. As for PPIs, sources for localization data are reported in the Cytoscape attributes. In addition to the localization data, we integrate InterPro protein domain information (Hunter et al., 2009) and GO Biological Process and GO Molecular Function data (Harris et al., 2004), which can be viewed as node attributes in Cytoscape (see “Results and Discussion”; Shannon et al., 2003).

Supplemental Data

The following materials are available in the online version of this article.

Supplemental Figure S1. Screen shots of CORNET.

Supplemental Table S1. Group I genes in the DELLA network (Fig. 4A).

Supplemental Table S2. Group II genes in the DELLA network (Fig. 4B).

Supplemental Files S1. Cytoscape session files of the DELLA networks (http://bioinformatics.psb.ugent.be/supplementary_data/stbod/cornet/).

ACKNOWLEDGMENTS

We thank Klaas Vandepoele and Fabio Fiorani for fruitful discussions, our colleagues for helpful comments on the tool, Stijn Vermeir, Thomas Van Parys, and Lieven Baeyens for technical assistance, and Martine De Cock for help in preparing the manuscript. We acknowledge three anonymous reviewers for helpful comments on the manuscript.

Received September 8, 2009; accepted January 4, 2010; published January 6, 2010.

LITERATURE CITED

- Achard P, Cheng H, De Grauwe L, Decat J, Schoutteten H, Moritz T, Van Der Straeten D, Peng J, Harberd NP (2006) Integration of plant responses to environmentally activated phytohormonal signals. *Science* **311**: 91–94
- Achard P, Genschik P (2009) Releasing the brakes of plant growth: how GAs shutdown DELLA proteins. *J Exp Bot* **60**: 1085–1092
- Achard P, Liao L, Jiang C, Desnos T, Bartlett J, Fu X, Harberd NP (2007) DELLAs contribute to plant photomorphogenesis. *Plant Physiol* **143**: 1163–1172
- Akiyama K, Chikayama E, Yuasa H, Shimada Y, Tohge T, Shinozaki K, Hirai MY, Sakurai T, Kikuchi J, Saito K (2008) PRIME: a Web site that assembles tools for metabolomics and transcriptomics. *In Silico Biol* **8**: 339–345
- Ariizumi T, Murase K, Sun TP, Steber CM (2008) Proteolysis-independent downregulation of DELLA repression in *Arabidopsis* by the gibberellin receptor GIBBERELLIN INSENSITIVE DWARF1. *Plant Cell* **20**: 2447–2459
- Avraham S, Tung CW, Ilic K, Jaiswal P, Kellogg EA, McCouch S, Pujar A, Reiser L, Rhee SY, Sachs MM, et al (2008) The Plant Ontology Database: a community resource for plant structure and developmental stages controlled vocabulary and annotations. *Nucleic Acids Res* **36**: D449–D454
- Bader GD, Betel D, Hogue CWV (2003) BIND: the Biomolecular Interaction Network Database. *Nucleic Acids Res* **31**: 248–250
- Bannai H, Tamada Y, Maruyama O, Nakai K, Miyano S (2002) Extensive feature detection of N-terminal protein sorting signals. *Bioinformatics* **18**: 298–305
- Barrett T, Edgar R (2006) Gene expression omnibus: microarray data storage, submission, retrieval, and analysis. *Methods Enzymol* **411**: 352–369
- Brady SM, Provart NJ (2009) Web-queryable large-scale data sets for hypothesis generation in plant biology. *Plant Cell* **21**: 1034–1051
- Brazma A (2009) Minimum Information About a Microarray Experiment (MIAME): successes, failures, challenges. *ScientificWorldJournal* **9**: 420–423
- Brown DM, Zeef LAH, Ellis J, Goodacre R, Turner SR (2005) Identification of novel genes in *Arabidopsis* involved in secondary cell wall formation using expression profiling and reverse genetics. *Plant Cell* **17**: 2281–2295
- Bruskiewich R, Coe EH, Jaiswal P, McCouch S, Polacco M, Stein L, Vincent L, Ware D (2002) The plant ontology consortium and plant ontologies. *Comp Funct Genomics* **3**: 137–142
- Cao D, Cheng H, Wu W, Soo HM, Peng J (2006) Gibberellin mobilizes distinct DELLA-dependent transcriptomes to regulate seed germination and floral development in *Arabidopsis*. *Plant Physiol* **142**: 509–525
- Casneuf T, Van de Peer Y, Huber W (2007) In situ analysis of cross-hybridisation on microarrays and the inference of expression correlation. *BMC Bioinformatics* **8**: 461
- Chatr-aryamontri A, Ceol A, Montecchi Palazzi L, Nardelli G, Schneider MV, Castagnoli L, Cesareni G (2007) MINT: the Molecular INteraction database. *Nucleic Acids Res* **35**: D572–D574
- Chen H, Huang N, Sun Z (2006) SubLoc: a server/client suite for protein subcellular location based on SOAP. *Bioinformatics* **22**: 376–377
- Cheng H, Qin L, Lee S, Fu X, Richards DE, Cao D, Luo D, Harberd NP, Peng J (2004) Gibberellin regulates *Arabidopsis* floral development via suppression of DELLA protein function. *Development* **131**: 1055–1064
- Claros MG (1995) MitoProt, a Macintosh application for studying mitochondrial proteins. *Comput Appl Biosci* **11**: 441–447
- Cui J, Li P, Li G, Xu F, Zhao C, Li Y, Yang Z, Wang G, Yu Q, Li Y, et al (2008) ATPID: *Arabidopsis thaliana* protein interactome database. An integrative platform for plant systems biology. *Nucleic Acids Res* **36**: D999–D1008
- De Bodt S, Proost S, Vandepoele K, Rouze P, Van de Peer Y (2009) Predicting protein-protein interactions in *Arabidopsis thaliana* through integration of orthology, gene ontology and co-expression. *BMC Genomics* **10**: 288
- de Lucas M, Davière JM, Rodríguez-Falcón M, Pontin M, Iglesias-Pedraz JM, Lorrain S, Fankhauser C, Blázquez MA, Titarenko E, Prat S (2008) A molecular framework for light and gibberellin control of cell elongation. *Nature* **451**: 480–484
- de Nooy W, Mrvar A, Batagelj V (2005) *Exploratory Social Network Analysis with Pajek*. Cambridge University Press, Cambridge, UK
- Dill A, Sun TP (2001) Synergistic derepression of gibberellin signaling by removing RGA and GAI function in *Arabidopsis thaliana*. *Genetics* **159**: 777–785
- Dill A, Thomas SG, Hu J, Steber CM, Sun TP (2004) The *Arabidopsis* F-box protein SLEEPY1 targets gibberellin signaling repressors for gibberellin-induced degradation. *Plant Cell* **16**: 1392–1405
- Emanuelsson O, Brunak S, von Heijne G, Nielsen H (2007) Locating proteins in the cell using TargetP, SignalP and related tools. *Nat Protoc* **2**: 953–971
- Emanuelsson O, Elofsson A, von Heijne G, Cristóbal S (2003) In silico prediction of the peroxisomal proteome in fungi, plants and animals. *J Mol Biol* **330**: 443–456
- Feng S, Martinez C, Gusmaroli G, Wang Y, Zhou J, Wang F, Chen L, Yu L, Iglesias-Pedraz JM, Kircher S, et al (2008) Coordinated regulation of *Arabidopsis thaliana* development by light and gibberellins. *Nature* **451**: 475–479
- Fu X, Richards DE, Fleck B, Xie D, Burton N, Harberd NP (2004) The

- Arabidopsis* mutant sleepy1^{gar2-1} protein promotes plant growth by increasing the affinity of the SCF^{SLY1} E3 ubiquitin ligase for DELLA protein substrates. *Plant Cell* **16**: 1406–1418
- Gachon CMM, Langlois-Meurinne M, Henry Y, Saindrenan P (2005) Transcriptional co-regulation of secondary metabolism enzymes in *Arabidopsis*: functional and evolutionary implications. *Plant Mol Biol* **58**: 229–245
- Gautier L, Cope L, Bolstad BM, Irizarry RA (2004) affy: analysis of Affymetrix GeneChip data at the probe level. *Bioinformatics* **20**: 307–315
- Geisler-Lee J, O'Toole N, Ammar R, Provart NJ, Millar AH, Geisler M (2007) A predicted interactome for *Arabidopsis*. *Plant Physiol* **145**: 317–329
- Gentleman RC, Carey VJ, Bates DM, Bolstad B, Dettling M, Dudoit S, Ellis B, Gautier L, Ge Y, Gentry J, et al (2004) Bioconductor: open software development for computational biology and bioinformatics. *Genome Biol* **5**: R80
- Goda H, Sasaki E, Akiyama K, Maruyama-Nakashita A, Nakabayashi K, Li W, Ogawa M, Yamauchi Y, Preston J, Aoki K, et al (2008) The AtGenExpress hormone and chemical treatment data set: experimental design, data evaluation, model data analysis and data access. *Plant J* **55**: 526–542
- Goldovsky L, Cases I, Enright AJ, Ouzounis CA (2005) BioLayout(Java): versatile network visualisation of structural and functional relationships. *Appl Bioinformatics* **4**: 71–74
- Griffiths J, Murase K, Rieu I, Zentella R, Zhang ZL, Powers SJ, Gong F, Phillips AL, Hedden P, Sun TP, et al (2006) Genetic characterization and functional analysis of the GID1 gibberellin receptors in *Arabidopsis*. *Plant Cell* **18**: 3399–3414
- Guda C, Fahy E, Subramaniam S (2004) MITOPRED: a genome-scale method for prediction of nucleus-encoded mitochondrial proteins. *Bioinformatics* **20**: 1785–1794
- Gutierrez RA, Lejay LV, Dean A, Chiaromonte F, Shasha DE, Coruzzi GM (2007) Qualitative network models and genome-wide expression data define carbon/nitrogen-responsive molecular machines in *Arabidopsis*. *Genome Biol* **8**: R7
- Harris MA, Clark J, Ireland A, Lomax J, Ashburner M, Foulger R, Eilbeck K, Lewis S, Marshall B, Mungall C, et al (2004) The Gene Ontology (GO) database and informatics resource. *Nucleic Acids Res* **32**: D258–D261
- Heazlewood JL, Verboom RE, Tonti-Filippini J, Small I, Millar AH (2007) SUBA: the *Arabidopsis* Subcellular Database. *Nucleic Acids Res* **35**: D213–D218
- Hermjakob H, Montecchi-Palazzi L, Lewington C, Mudali S, Kerrien S, Orchard S, Vingron M, Roechert B, Roepstorff P, Valencia A, et al (2004) IntAct: an open source molecular interaction database. *Nucleic Acids Res* **32**: D452–D455
- Hoglund A, Dönnies P, Blum T, Adolph HW, Kohlbacher O (2006) MultiLoc: prediction of protein subcellular localization using N-terminal targeting sequences, sequence motifs and amino acid composition. *Bioinformatics* **22**: 1158–1165
- Horan K, Jang C, Bailey-Serres J, Mittler R, Shelton C, Harper JE, Zhu JK, Cushman JC, Gollery M, Girke T (2008) Annotating genes of known and unknown function by large-scale coexpression analysis. *Plant Physiol* **147**: 41–57
- Horton P, Park KJ, Obayashi T, Fujita N, Harada H, Adams-Collier CJ, Nakai K (2007) WoLF SORT: protein localization predictor. *Nucleic Acids Res* **35**: W585–W587
- Hunter S, Apweiler R, Attwood TK, Bairoch A, Bateman A, Binns D, Bork P, Das U, Daugherty L, Duquenne L, et al (2009) InterPro: the integrative protein signature database. *Nucleic Acids Res* **37**: D211–D215
- Ilic K, Kellogg EA, Jaiswal P, Zapata F, Stevens PF, Vincent LP, Avraham S, Reiser L, Pujar A, Sachs MM, et al (2007) The plant structure ontology, a unified vocabulary of anatomy and morphology of a flowering plant. *Plant Physiol* **143**: 587–599
- Irizarry RA, Bolstad BM, Collin F, Cope LM, Hobbs B, Speed TP (2003a) Summaries of Affymetrix GeneChip probe level data. *Nucleic Acids Res* **31**: e15
- Irizarry RA, Hobbs B, Collin F, Beazer-Barclay YD, Antonellis KJ, Scherf U, Speed TP (2003b) Exploration, normalization, and summaries of high density oligonucleotide array probe level data. *Biostatistics* **4**: 249–264
- Kilian J, Whitehead D, Horak J, Wanke D, Weinl S, Batistic O, D'Angelo C, Bornberg-Bauer E, Kudla J, Harter K (2007) The AtGenExpress global stress expression data set: protocols, evaluation and model data analysis of UV-B light, drought and cold stress responses. *Plant J* **50**: 347–363
- King KE, Moritz T, Harberd NP (2001) Gibberellins are not required for normal stem growth in *Arabidopsis thaliana* in the absence of GAI and RGA. *Genetics* **159**: 767–776
- Lee S, Cheng H, King KE, Wang W, He Y, Hussain A, Lo J, Harberd NP, Peng J (2002) Gibberellin regulates *Arabidopsis* seed germination via RGL2, a GAI/RGA-like gene whose expression is up-regulated following imbibition. *Genes Dev* **16**: 646–658
- Lisso J, Steinhäuser D, Altmann T, Kopka J, Müssig C (2005) Identification of brassinosteroid-related genes by means of transcript co-response analyses. *Nucleic Acids Res* **33**: 2685–2696
- Manfield IW, Jen CH, Pinney JW, Michalopoulos I, Bradford JR, Gilmartin PM, Westhead DR (2006) *Arabidopsis* Co-expression Tool (ACT): Web server tools for microarray-based gene expression analysis. *Nucleic Acids Res* **34**: W504–W509
- Mutwil M, Øbro J, Willats WGT, Persson S (2008) GeneCAT: novel Webtools that combine BLAST and co-expression analyses. *Nucleic Acids Res* **36**: W320–W326
- Nair R, Rost B (2005) Mimicking cellular sorting improves prediction of subcellular localization. *J Mol Biol* **348**: 85–100
- Nakajima M, Shimada A, Takashi Y, Kim YC, Park SH, Ueguchi-Tanaka M, Suzuki H, Katoh E, Iuchi S, Kobayashi M, et al (2006) Identification and characterization of *Arabidopsis* gibberellin receptors. *Plant J* **46**: 880–889
- Obayashi T, Hayashi S, Saeki M, Ohta H, Kinoshita K (2009) ATTED-II provides coexpressed gene networks for *Arabidopsis*. *Nucleic Acids Res* **37**: D987–D991
- Obayashi T, Kinoshita K, Nakai K, Shibaoka M, Hayashi S, Saeki M, Shibata D, Saito K, Ohta H (2007) ATTED-II: a database of co-expressed genes and cis elements for identifying co-regulated gene groups in *Arabidopsis*. *Nucleic Acids Res* **35**: D863–D869
- Pujar A, Jaiswal P, Kellogg EA, Ilic K, Vincent L, Avraham S, Stevens P, Zapata F, Reiser L, Rhee SY, et al (2006) Whole-plant growth stage ontology for angiosperms and its application in plant biology. *Plant Physiol* **142**: 414–428
- Rautengarten C, Steinhäuser D, Büssis D, Stintzi A, Schaller A, Kopka J, Altmann T (2005) Inferring hypotheses on functional relationships of genes: analysis of the *Arabidopsis thaliana* subtilase gene family. *PLoS Comput Biol* **1**: e40
- Rhee SY, Beavis W, Berardini TZ, Chen G, Dixon D, Doyle A, Garcia-Hernandez M, Huala E, Lander G, Montoya M, et al (2003) The *Arabidopsis* Information Resource (TAIR): a model organism database providing a centralized, curated gateway to *Arabidopsis* biology, research materials and community. *Nucleic Acids Res* **31**: 224–228
- Salwinski L, Miller CS, Smith AJ, Pettit FK, Bowie JU, Eisenberg D (2004) The Database of Interacting Proteins: 2004 update. *Nucleic Acids Res* **32**: D449–D451
- Schmid M, Davison TS, Henz SR, Pape UJ, Demar M, Vingron M, Schölkopf B, Weigel D, Lohmann JU (2005) A gene expression map of *Arabidopsis thaliana* development. *Nat Genet* **37**: 501–506
- Schwechheimer C (2008) Understanding gibberellin acid signaling: are we there yet? *Curr Opin Plant Biol* **11**: 9–15
- Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, Amin N, Schwikowski B, Ideker T (2003) Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res* **13**: 2498–2504
- Small I, Peeters N, Legeai F, Lurin C (2004) Predotar: a tool for rapidly screening proteomes for N-terminal targeting sequences. *Proteomics* **4**: 1581–1590
- Srinivasainagendra V, Page GP, Mehta T, Coulibaly I, Loraine AE (2008) CressExpress: a tool for large-scale mining of expression data from *Arabidopsis*. *Plant Physiol* **147**: 1004–1016
- Stark C, Breitkreutz BJ, Reguly T, Boucher L, Breitkreutz A, Tyers M (2006) BioGRID: a general repository for interaction datasets. *Nucleic Acids Res* **34**: D535–D539
- Steinhäuser D, Usadel B, Luedemann A, Thimm O, Kopka J (2004) CSB-DB: a comprehensive systems-biology database. *Bioinformatics* **20**: 3647–3651
- Taylor CF, Paton NW, Lilley KS, Binz PA, Julian RK Jr, Jones AR, Zhu W, Apweiler R, Aebersold R, Deutsch EW, et al (2007) The minimum information about a proteomics experiment (MIAPE). *Nat Biotechnol* **25**: 887–893

- Toufighi K, Brady SM, Austin R, Ly E, Provart NJ** (2005) The Botany Array Resource: e-northern, expression angling, and promoter analyses. *Plant J* **43**: 153–163
- Tyler L, Thomas SG, Hu J, Dill A, Alonso JM, Ecker JR, Sun TP** (2004) DELLA proteins and gibberellin-regulated seed germination and floral development in Arabidopsis. *Plant Physiol* **135**: 1008–1019
- Usadel B, Obayashi T, Mutwil M, Giorgi FM, Bassel GW, Tanimoto M, Chow A, Steinhauser D, Persson S, Provart NJ** (2009) Co-expression tools for plant biology: opportunities for hypothesis generation and caveats. *Plant Cell Environ* **32**: 1633–1651
- Vandepoele K, Quimbaya M, Casneuf T, De Veylder L, Van de Peer Y** (2009) Unraveling transcriptional control in Arabidopsis using cis-regulatory elements and coexpression networks. *Plant Physiol* **150**: 535–546
- Whetzel PL, Parkinson H, Causton HC, Fan L, Fostel J, Fragoso G, Game L, Heiskanen M, Morrison N, Rocca-Serra P, et al** (2006) The MGED Ontology: a resource for semantics-based description of microarray experiments. *Bioinformatics* **22**: 866–873
- Yu H, Ito T, Zhao Y, Peng J, Kumar P, Meyerowitz EM** (2004) Floral homeotic genes are targets of gibberellin signaling in flower development. *Proc Natl Acad Sci USA* **101**: 7827–7832
- Zimmermann P, Hirsch-Hoffmann M, Hennig L, Gruissem W** (2004) GENEVESTIGATOR: Arabidopsis microarray database and analysis toolbox. *Plant Physiol* **136**: 2621–2632