



Correlation networks visualization

Nicholas Provart*

Department of Cell and Systems Biology/Centre for the Analysis of Genome Evolution and function, University of Toronto, Toronto, ON, Canada

Edited by:

Bjoern Usadel, RWTH Aachen University, Germany

Reviewed by:

Takayuki Tohge, Max Planck Institute for Molecular Plant Physiology, Germany

Ming Chen, Zhejiang University, China

*Correspondence:

Nicholas Provart, Department of Cell and Systems Biology/Centre for the Analysis of Genome Evolution and function, University of Toronto, 25 Willcocks Street, Room 3051, Toronto, ON, Canada M5S 3B2.
e-mail: nicholas.provart@utoronto.ca

New, *in silico* ways of generating hypotheses based on large data sets have emerged in the past decade. These data sets have been used to investigate different aspects of plant biology, especially at the level of transcriptome, from tissue-specific expression patterns to patterns in as little as a few cells. Such publicly available data are a boon to researchers for hypothesis generation by providing a guide for experimental work such as phenotyping or genetic analysis. More advanced computational methods can leverage these data via gene coexpression analysis, the results of which can be visualized and refined using network analysis. Other kinds of networks of, e.g., protein–protein interactions, can also be used to inform biology. These networks can be visualized and analyzed with additional information on gene expression levels, subcellular localization, etc., or with other emerging kinds of information. Finally, cross-level correlation is an area that will become increasingly important. Visualizing these cross-level correlations will require new data visualization tools.

Keywords: coexpression analysis, network visualization, hypothesis generation, transcriptomics, protein–protein interactions

INTRODUCTION

The past decade of plant research has seen an unprecedented increase in the amount of data being generated from various levels of an organism. Advances in genome sequencing, gene expression profiling, determining all-by-all interactomes, and other areas, have allowed new, *in silico* ways of generating hypotheses, in addition to traditional and powerful forward genetic screens and techniques. These methods have been used to investigate different aspects of plant biology, especially at the level of transcriptome, from tissue-specific expression patterns at the centimeter scale to those in as little as a few cells. While such publicly available data can be used on their own by researchers to generate hypothesis, one of the most powerful methods for hypothesis generation to date is gene coexpression analysis. Here, the correlation in expression pattern between pairs of genes is measured, and those exhibiting strong correlations are “joined” in a graphical representation to create a network, which can be visualized with graph network viewers. Other kinds of networks of, e.g., protein–protein interactions, a kind of biological, *in vivo* correlation, can also be used to inform biology. These networks can be visualized and analyzed with additional information on gene expression levels, subcellular localization, etc., or with other kinds of information just becoming available, such as patterns of positive selection along a gene, genes with similar patterns of expression in equivalent tissues in other species, etc. Finally, cross-level correlation is an area that will become more important as genome wide association studies (GWAS) could be used to link genotype to environmental factors or perturbations through changes in the transcriptome, epigenome, or other ‘omes of a plant. Visualizing these cross-level correlations will require new data visualization tools, the most promising of which will allow the plant at all levels to be an “app” on a smart phone, laptop computer, or scientific conference video wall. Such efforts will serve to make the proposed open access/open source Arabidopsis Information Portal, in which

zoomable user interfaces and powerful data analysis and visualization algorithms will be an important part, an immediate resource not only for *Arabidopsis* researchers, but for all plant biologists worldwide.

GENE EXPRESSION VISUALIZATION

Large amount of gene expression data for *Arabidopsis thaliana* and other plants have been generated in the past decade, and these can be examined to provide further understanding about one’s gene(s) or biological system of interest. One can use such data to ask questions such as: within my gene’s family, are family members uniquely expressed in specific organs/tissues or is one uniquely induced in expression by a specific stress? Such data can be explored with several easy-to-use web-based tools, such as Expression Browser (Toufighi et al., 2005) or electronic Fluorescent Pictograph (eFP) Browser (Winter et al., 2007) at the Bio-Array Resource (BAR), Genevestigator (Zimmermann et al., 2004), the TileViz tool developed for the At-TAX data set generated by Laubinger et al. (2008), and others. The use of these powerful tools is described in a review by Brady and Provart (2009). While such gene expression visualization tools can be very useful for some biological questions, in order to identify patterns of correlation between different genes using these tools one would have to manually examine outputs looking for genes with similar patterns of expression, a rather onerous and subjective task. Thus it is advantage to use computational methods to identify such coexpressed genes. Coexpressed genes can be used as a “primary screen” to identify novel genes involved in one’s biological process of interest, and examples of such new insights are described in Usadel et al. (2009).

GENE COEXPRESSION NETWORKS

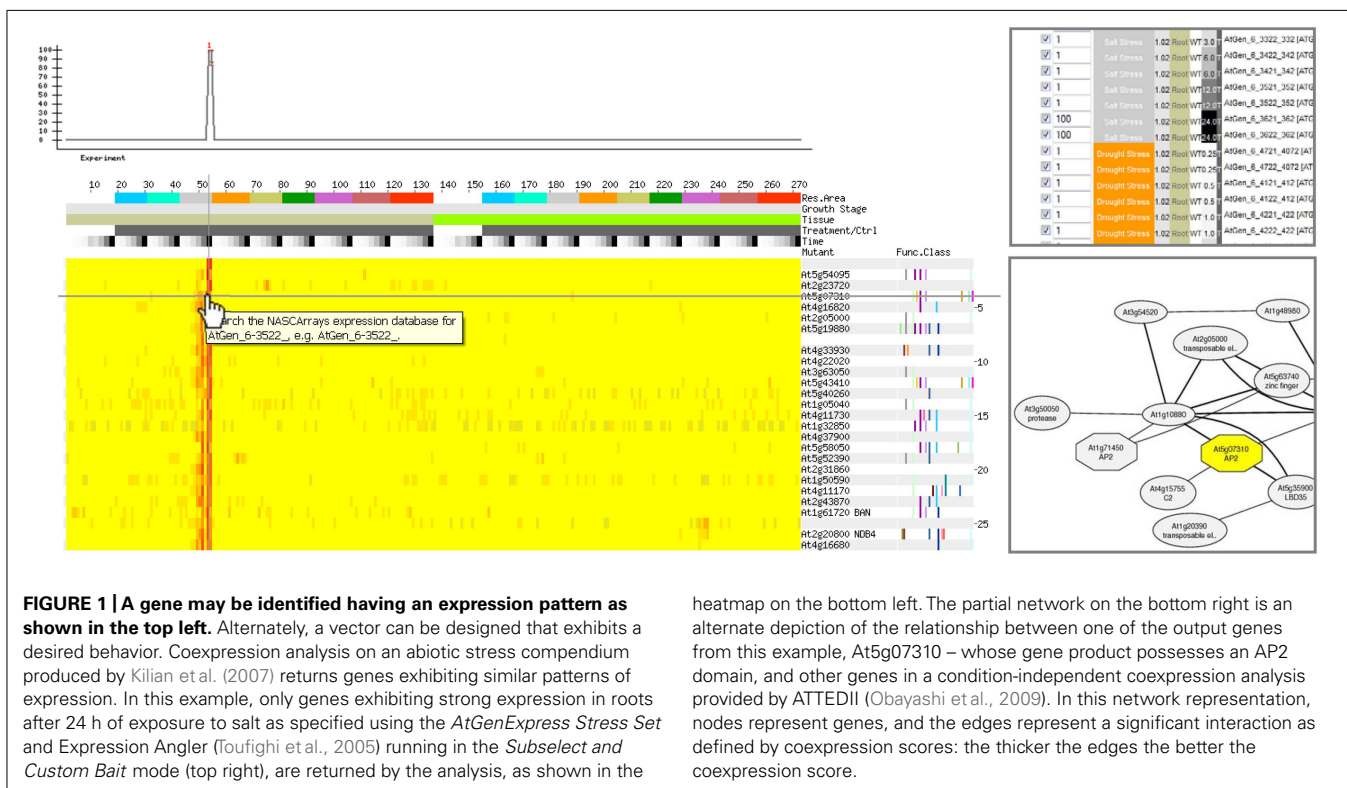
Four components are necessary for conducting a gene coexpression analysis. These are: (1) a collection of gene expression profiles

across as many genes as possible from different samples and/or different perturbations; (2) a method for computing expression pattern similarity; (3) a way for assessing the degree of significance of expression pattern similarity; and (4) a tool to visualize and analyze statistically significant coexpression patterns. A recent review by Usadel et al. (2009) describes how the selection of gene expression profiles to create a collection (or “compendium”) of profiles for coexpression analysis can influence the results that one gets from such an analysis. It is possible to select samples in a “condition-independent” or “condition-dependent” manner. In some cases, such as SeedNet (Bassel et al., 2011), it clearly makes sense to use data sets from specific samples (seeds in that case), while in other, more exploratory cases, condition-independent data sets can be used as a starting point for hypothesis generation. The Usadel et al. (2009) article also describes commonly used metrics for assessing expression pattern similarity. The most commonly used metrics are the Pearson’s correlation coefficient (PCC), along with Spearman’s (Rank) correlation coefficient, and mutual information. Further, the Usadel et al. (2009) review also describes ways of assessing significance of coexpression scores. It should be pointed out that if large numbers of data sets are used when computing expression pattern similarity, PCC values of as low as 0.2 will have good *P* values. Thus it is often useful to use a higher cutoff, and to consider the r^2 value, achieved by squaring the PCC score, for the purpose of selecting coexpressed genes. The r^2 value reflects the amount of variance in common. Thus a pair of genes with a coexpression score of 0.2 as measured by the PCC will have an r^2 value of $(0.2)^2 = 0.04$, or only 4% of variance in common. This fact also often explains why a PCC value of 0.75

is commonly used as a cutoff, because then the r^2 value would be 0.56, meaning that genes exhibiting such a score would have 56% of their variance in common. This seems biologically “meaningful” in the sense that this variance might be directed by *cis*-regulatory elements in common in the promoters of coexpressed genes. In terms of visualizing expression patterns, often it is instructive to examine heatmap outputs for all of the highly correlated genes for a given query gene, or for those matching a desired gene expression pattern, as shown in **Figure 1**. When the values for coexpression are significant between several of the genes in the single gene query example just mentioned, it is also useful to depict these as a graph network as shown in **Figure 1**.

The network depicted in **Figure 1** can be generated using the ATTEDII coexpression tool (Obayashi et al., 2009). Nodes that are not related to the original query gene may be added to the network based on precomputed coexpression data in the ATTEDII database. Another nice feature of the ATTEDII database is the ability to easily explore coexpression scores in a condition-independent compendium or in several focused condition-dependent compendia, such as stress, hormone treatments, etc.

Other useful tools for generating and visualizing gene coexpression networks in *Arabidopsis* are CressExpress (Srinivasasainagendra et al., 2008), PlaNet (Mutwil et al., 2011), and CSB.DB (Steinhauser et al., 2004). With each of these it is possible to easily retrieve data for use in generating a coexpression network, which may be visualized online or with network viewers as described in the next paragraph. A couple of other of tools, which are not strictly coexpression-based, are also useful for generating



correlation networks. GeneMANIA (Mostafavi et al., 2008) and AraNet (Lee et al., 2010) permit coexpression data to be used to connect and extend genes in a network. Further information can be used to deduce new members of a network. Such information includes subcellular localization, protein–protein interaction data, and shared protein domains. Another tool was recently developed for exploring “gene-sharing networks” for tissues that are connected based on the degree of shared, leptokurtically expressed genes, i.e., genes that are preferentially expressed in certain tissues (Li et al., 2012).

Once a coexpression network has been generated, for example, using the ATTEDII database or by a custom “all-by-all” analysis, as was the case for SeedNet, it is often useful to import the network into a powerful network analysis tool, Cytoscape (Shannon et al., 2003; Kohl et al., 2011). This is easily done from an Excel table containing the interactors (i.e., genes exhibiting a strong correlation in their expression patterns) in two columns along with their interaction strength (e.g., coexpression score). Using tools

within Cytoscape, it is possible to analyze the network to identify nodes (i.e., genes) that are “hubs”, meaning these genes have large numbers of coexpression partners. These hub nodes may be analyzed in conjunction with other data, such as whether the nodes genes exhibit increased or decreased expression levels under particular conditions, to identify significant hub genes (Horvath and Dong, 2008). These significant hub genes may play important roles biologically. In the case of SeedNet, 50% of the genes identified as hub genes were accurately predicted to be regulators of germination based on follow up biological experiments, in contrast to a 22% accuracy rate based on candidates identified by differential gene expression analysis alone. “Modules”, that is, genes coexpressed as a group, which would have many coexpression connections between all of the genes in the module, can be identified using the MCODE algorithm (Bader and Hogue, 2003). Further analysis of the genes in a module can be done using several Gene Ontology (Ashburner et al., 2000) term enrichment analysis plugins available for Cytoscape, to see whether the genes in a

Table 1 | Selected correlation network databases and tools.

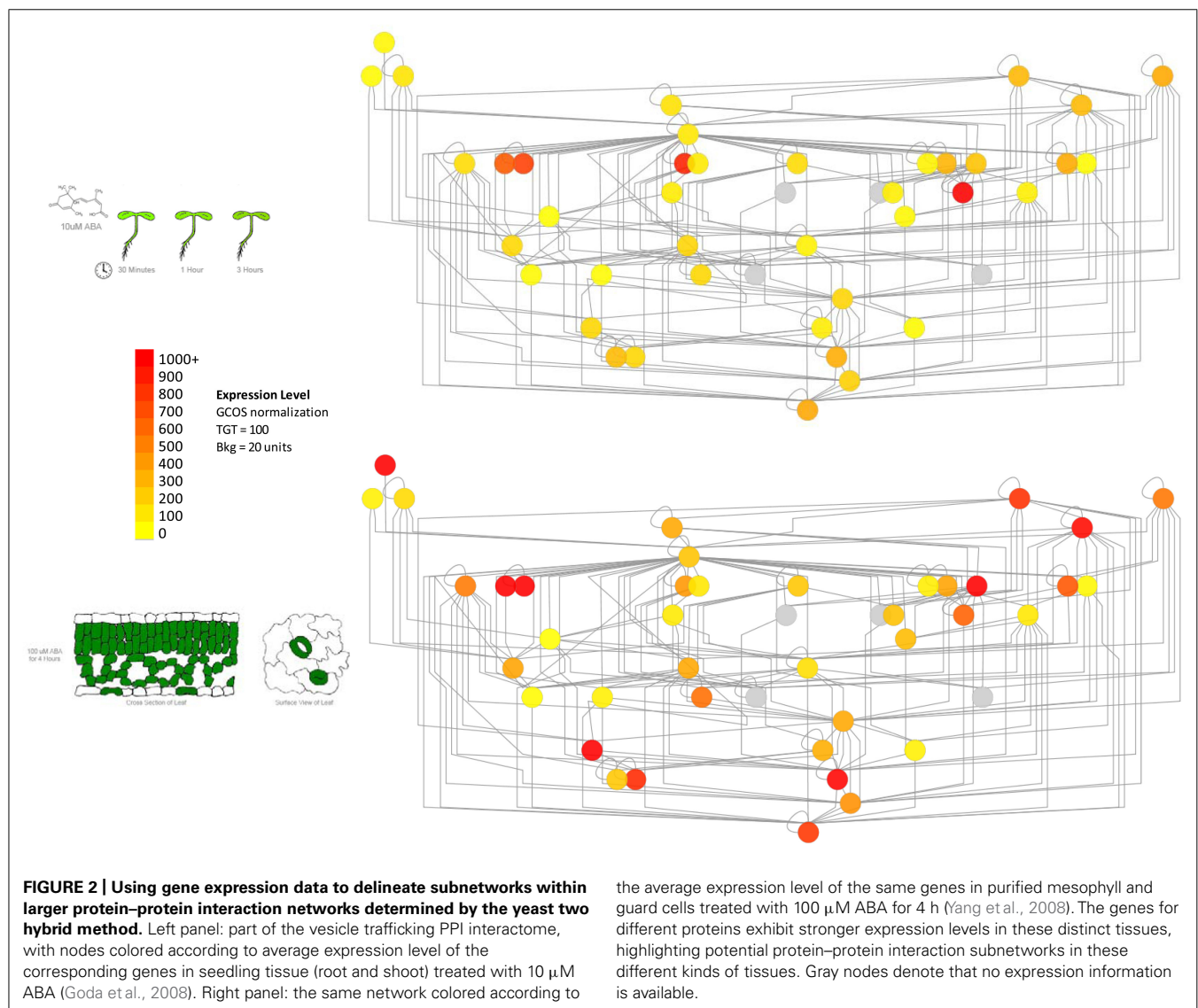
URL and comments	
Correlation network DB	
ATTEDII (Obayashi et al., 2009)	http://atted.jp/ ; explore condition-independent coexpression networks for up to 100 genes in <i>Arabidopsis</i> using the NetworkDrawer tool . Coexpression analyses may also be performed for rice.
CressExpress (Srinivasasainagendra et al., 2008)	http://cressexpress.org/ ; generate condition-independent coexpression analyses or custom condition-dependent coexpression analyses for <i>Arabidopsis</i> with up to 30 genes. Results are easily imported into Cytoscape for visualization.
PlaNet (Mutwil et al., 2011)	http://aranet.mpimp-golm.mpg.de/ ; use this tool to explore condition-independent coexpression networks in seven plant species. Networks are displayed as static SVG images, but networks may also be downloaded for easy viewing and further manipulation in Cytoscape or Pajek.
CSB.DB (Steinhauser et al., 2004)	http://csbdb.mpimp-golm.mpg.de/ ; use this tool to explore both condition-independent or condition-dependent coexpression networks in <i>Arabidopsis</i> for up to 60 genes. Networks may be immediately viewed as images, or downloaded for further manipulation into the visualization tools below.
GeneMANIA (Mostafavi et al., 2008)	http://genemania.org/ ; this tool allows functional network generation in <i>Arabidopsis</i> based on user-selected or default expression data sets, protein–protein interactions, subcellular localization, shared protein domains, etc. Results are easily visualized via an embedded Cytoscape Web (Lopes et al., 2010) application.
SeedNet (Bassel et al., 2011)	http://bree.cs.nott.ac.uk/arabidopsis/ ; explore condition specific (i.e., seed-expressed) gene networks from <i>Arabidopsis</i> in a custom network explorer.
AraNet (Lee et al., 2010)	http://www.functionalnet.org/aranet/ ; like GeneMANIA this tool allows functional network generation in <i>Arabidopsis</i> . Results may be visualized via activation of Cytoscape Web.
Network visualization tools	
Cytoscape (Shannon et al., 2003; Kohl et al., 2011)	http://www.cytoscape.org/ ; use this powerful open-source desktop tool to visualize coexpression and other networks, such as those generated by protein–protein interaction studies. Nodes and edges may be appended with additional, user-defined information.
Biolayout (Theocharidis et al., 2009)	http://biolayout.org/ ; the current iteration of this desktop tool, Biolayout Express ^{3D} , permits visualization of coexpression and other networks in three-dimensional space. Cytoscape “.sif” files may be imported into the tool.
Pajek (Batagelj and Mrvar, 1998)	http://pajek.imfm.si/doku.php

given module are involved with a particular biological process or molecular function. Other graph network visualization tools such as Pajek (Batagelj and Mrvar, 1998) or Biolayout (Theocharidis et al., 2009) can also be used to visualize networks. The aforementioned databases and visualization tools are summarized with their availability and features in **Table 1**.

TRANSCRIPTOME DATA IN OTHER KINDS OF CORRELATION NETWORKS

Elucidating protein–protein interaction networks has proven to be useful for the identification of novel members of a biological process. In the case of *Arabidopsis thaliana*, many groups have used biochemical or yeast-based methods, such as the yeast two hybrid system (Fields and Song, 1989), to determine whether or not small subsets of proteins interact with one another. The largest application of the yeast two hybrid method to date for plants was published last year by the Arabidopsis Interactome Mapping Consortium (2011), and covers an estimated 2% of the potential

protein–protein interactions in this species. Computational methods can also be used to predict protein–protein interactions in a species based on the presence of interacting orthologs in other species (Geisler-Lee et al., 2007). An important caveat to protein–protein interaction analysis methods, especially those based on protein expression in heterologous systems such as yeast, is that even though proteins can interact *in vitro* or *in vivo*, doesn't mean that they necessarily interact *in vivo* in every tissue of the plant or under every perturbation. Other factors, such as expression levels, protein degradation, and subcellular localization, are important for an interaction to occur. For instance, if two proteins can physically interact based on yeast two hybrid data, but one of these proteins is not expressed in a given tissue or organ *in planta*, then that interaction cannot occur in that tissue or organ. Likewise, proteins that are not localized in the same subcellular compartment would seem to have a lower likelihood of being able to interact *in planta*, although it is clear that proteins, especially signaling proteins, can move between compartments.



In order to address the above caveat, it is possible to overlay subcellular localization data or gene expression data onto protein–protein interaction networks. Such data are available in the former case from SUBA, the *Arabidopsis* subcellular database (Heazlewood et al., 2006), and in the latter from extensive expression repositories at NCBI's GEO (Edgar et al., 2002), the BAR (Toufighi et al., 2005), Genevestigator (Zimmermann et al., 2004) and other sources. An example of how such expression data can be used to delineate subnetworks within PPI networks can be seen in **Figure 2**. The network shows some proteins involved in vesicle trafficking, based both on literature-documented and predicted protein–protein interactions (Geisler-Lee et al., 2007).

As shown in **Figure 2**, only certain members of the “supernetwork” of all possible protein–protein interactions are expressed in either seedlings or guard cells. While further experiments are necessary to show whether the nodes with higher expression levels, shown in red or orange in the figure, are actually key to vesicle trafficking in these tissues, such information can clearly be useful in terms of ordering T-DNA knockout lines of specific genes (Alonso et al., 2003) and identifying under which conditions or in which tissues to look for a phenotype.

CORRELATION NETWORKS ACROSS DIFFERENT SPECIES

The above two sections have highlighted two ways in which correlation networks can be used to guide biology in one very well studied species, *Arabidopsis thaliana*. Recent work by Mutwil et al. (2011) has computed coexpression networks in different plant species, and has used similar coexpressed network vicinities to help identify functional homologs across species. These data are accessible through the PlaNet tool. CoP by Ogata et al. (2010) and STARNET2 by Jupiter et al. (2009) offer a similar functionality. In a slightly different approach, Patel et al. (2012) have used gene expression atlases for seven plant species – *Arabidopsis*, soybean, *Medicago truncatula*, poplar, barley, maize, and rice – to identify equivalent tissues between these species. Based on these tissue equivalencies, the authors then computed the expression pattern similarity scores for a set of homologs from one species to a homolog from another species. The homolog exhibiting the best expression pattern similarity score is termed the “expressolog”. The authors showed that the number of instances in which the expressologs are not the best sequence similarity matches ranges from a low of 15.4% between poplar and *M. truncatula*, to a high of 50.7%

between barley and soybean. The expression pattern and sequence similarity scores may be viewed with the Expressolog Tree Viewer tool available at <http://bar.utoronto.ca>. Given the complexity of plant genomes in terms of whole genome and segmental duplication events (Arabidopsis Genome Initiative, 2000; Jiao et al., 2011), it is perhaps not surprising that expression patterns and sequences can follow different evolutionary trajectories. Interestingly, Movahedi et al. (2011) postulate that coexpression networks have undergone concerted rewiring in *Arabidopsis* as compared to rice.

CORRELATION NETWORKS ACROSS DIFFERENT SCALES AND A LOOK TO THE FUTURE

It is becoming easier to produce transcriptome data for specific cell types, following the protocol initially developed in Philip Benfey's lab (Birnbaum et al., 2003). A high resolution spatiotemporal map of root development covering more than 120 different cell types and maturity stages provides unprecedented insight into root biology (Brady et al., 2007). Recent efforts to generate epigenomic information using bisulfite sequencing or “BS-seq”, first established by Lister et al. (2008) can be extended down to the cell and even allele level (Peng and Ecker, 2012). Further information on the genomes of ecotypes of *Arabidopsis* is being generated by the 1001 Arabidopsis Genomes effort (Weigel and Mott, 2009). Correlating genomic and epigenomic variation with gene expression levels will provide new insight into plant biology. For instance, it was recently shown by Downen et al. (2012) that plants differentially methylate their genomes in response to biotic stress, and that these patterns are closely associated with differential gene expression responses. Incorporating transcriptome data into GWAS efforts will also allow greater insight. Such an approach was recently published by Gan et al. (2011). The challenge for researchers will be to easily tap into such data sets to ask whether their gene or system of interest might be under some kind of regulation described in the literature, and to visualize these data sets integratively. It is hoped that the development of a new Arabidopsis Information Portal (International Arabidopsis Informatics Consortium, 2012) will easily enable such analyses, and that the further development of visualization tools, such as a “multi-omics” plugin for Cytoscape (Enjalbert et al., 2011), will allow us to visualize the results in a comprehensible manner.

REFERENCES

- Alonso, J. M., Stepanova, A. N., Leisse, T. J., Kim, C. J., Chen, H., Shinn, P., et al. (2003). Genome-wide insertional mutagenesis of *Arabidopsis thaliana*. *Sci. STKE* 301, 653.
- Arabidopsis Genome Initiative. (2000). Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. *Nature* 408, 796–815.
- Arabidopsis Interactome Mapping Consortium. (2011). Evidence for network evolution in an *Arabidopsis* interactome map. *Science* 333, 601.
- Ashburner, M., Ball, C. A., Blake, J. A., Botstein, D., Butler, H., Cherry, J. M., et al. (2000). Gene Ontology: tool for the unification of biology. *Nat. Genet.* 25, 25.
- Bader, G., and Hogue, C. (2003). An automated method for finding molecular complexes in large protein interaction networks. *BMC Bioinformatics* 4, 2. doi: 10.1186/1471-2105-4-2
- Bassel, G. W., Lan, H., Glaab, E., Gibbs, D. J., Gerjets, T., Krasnogor, N., et al. (2011). Genome-wide network model capturing seed germination reveals coordinated regulation of plant cellular phase transitions. *Proc. Natl. Acad. Sci. U.S.A.* 108, 9709–9714.
- Batagelj, V., and Mrvar, A. (1998). Pajek-program for large network analysis. *Connections* 21, 47–57.
- Birnbaum, K., Shasha, D., Wang, J., Jung, J., Lambert, G., Galbraith, D., et al. (2003). A gene expression map of the *Arabidopsis* root. *Sci. STKE* 302, 1956.
- Brady, S., and Provart, N. (2009). Web-queryable large-scale data sets for hypothesis generation in plant biology. *Plant Cell* 21, 1034–1051.
- Brady, S. M., Orlando, D. A., Lee, J.-Y., Wang, J. Y., Koch, J., Dinneny, J. R., et al. (2007). A high-resolution root spatiotemporal map reveals dominant expression patterns. *Science* 318, 801–806.
- Downen, R. H., Pelizzola, M., Schmitz, R. J., Lister, R., Downen, J. M., Nery, J. R., et al. (2012). Widespread dynamic DNA methylation in response to biotic stress. *Proc. Natl. Acad. Sci. U.S.A.* 109, E2183–E2191.
- Edgar, R., Domrachev, M., and Lash, A. E. (2002). Gene Expression Omnibus: NCBI gene expression and hybridization array data

- repository. *Nucleic Acids Res.* 30, 207–210.
- Enjalbert, B., Jourdan, F., and Portais, J.-C. (2011). Intuitive visualization and analysis of multi-omics data and application to *Escherichia coli* carbon metabolism. *PLoS ONE* 6, e21318. doi: 10.1371/journal.pone.0021318
- Fields, S., and Song, O. (1989). A novel genetic system to detect protein-protein interactions. 340, 245–246.
- Gan, X., Stegle, O., Behr, J., Steffen, J. G., Drewe, P., Hildebrand, K. L., et al. (2011). Multiple reference genomes and transcriptomes for *Arabidopsis thaliana*. *Nature* 477, 419–423.
- Geisler-Lee, J., O'Toole, N., Ammar, R., Provart, N. J., Millar, A. H., and Geisler, M. (2007). A predicted interactome for *Arabidopsis*. *Plant Physiol.* 145, 317–329.
- Goda, H., Sasaki, E., Akiyama, K., Maruyama-Nakashita, A., Nakabayashi, K., Li, W., 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.
- Heazlewood, J., Verboom, R., Tonti-Filippini, J., Small, I., and Millar, A. (2006). SUBA: the *Arabidopsis* sub-cellular database. *Nucleic Acids Res.* 35, D213–D218.
- Horvath, S., and Dong, J. (2008). Geometric interpretation of gene coexpression network analysis. *PLoS Comput. Biol.* 4, e1000117. doi:10.1371/journal.pcbi.1000117
- International Arabidopsis Informatics Consortium. (2012). Taking the next step: building an Arabidopsis information portal. *Plant Cell* 24, 2248–2256.
- Jiao, Y., Wickett, N. J., Ayyampalayam, S., Chandrabali, A. S., Landherr, L., Ralph, P. E., et al. (2011). Ancestral polyploidy in seed plants and angiosperms. *Nature* 473, 97–100.
- Jupiter, D., Chen, H., and Vanburen, V. (2009). STARNET 2: a web-based tool for accelerating discovery of gene regulatory networks using microarray co-expression data. *BMC Bioinformatics* 10, 332. doi: 10.1186/1471-2105-10-332
- Kilian, J., Whitehead, D., Horak, J., Wanke, D., Weigl, S., Batistic, O., et al. (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.
- Kohl, M., Wiese, S., and Warscheid, B. (2011). Cytoscape: software for visualization and analysis of biological networks. *Methods Mol. Biol.* 696, 291–303.
- Laubinger, S., Zeller, G., Henz, S., Sachsenberg, T., Widmer, C., Naouar, N., et al. (2008). At-TAX: a whole genome tiling array resource for developmental expression analysis and transcript identification in *Arabidopsis thaliana*. *Genome Biol.* 9, R112.
- Lee, I., Ambaru, B., Thakkar, P., Marcotte, E. M., and Rhee, S. Y. (2010). Rational association of genes with traits using a genome-scale gene network for *Arabidopsis thaliana*. *Nat. Biotechnol.* 28, 149–156.
- Li, S., Pandey, S., Gookin, T. E., Zhao, Z., Wilson, L., and Assmann, S. M. (2012). Gene-Sharing networks reveal organizing principles of transcriptomes in *Arabidopsis* and other multicellular organisms. *Plant Cell* 24, 1362–1378.
- Lister, R., O'Malley, R. C., Tonti-Filippini, J., Gregory, B. D., Berry, C. C., Millar, A. H., et al. (2008). Highly integrated single-base resolution maps of the epigenome in *Arabidopsis*. *Cell* 133, 523–536.
- Lopes, C. T., Franz, M., Kazi, F., Donaldson, S. L., Morris, Q., and Bader, G. D. (2010). Cytoscape Web: an interactive web-based network browser. *Bioinformatics* 26, 2347–2348.
- Mostafavi, S., Ray, D., Warde-Farley, D., Grouios, C., and Morris, Q. (2008). GeneMANIA: a real-time multiple association network integration algorithm for predicting gene function. *Genome Biol.* 9, S4.
- Movahedi, S., Van De Peer, Y., and Vandepoele, K. (2011). Comparative network analysis reveals that tissue specificity and gene function are important factors influencing the mode of expression evolution in *Arabidopsis* and rice. *Plant Physiol.* 156, 1316–1330.
- Mutwil, M., Klie, S., Tohge, T., Giorgi, F. M., Wilkins, O., Campbell, M. M., et al. (2011). PlaNet: combined sequence and expression comparisons across plant networks derived from seven species. *Plant Cell* 23, 895–910.
- Obayashi, T., Hayashi, S., Saeki, M., Ohta, H., and Kinoshita, K. (2009). ATTED-II provides coexpressed gene networks for *Arabidopsis*. *Nucleic Acids Res.* 37, D987–D991.
- Ogata, Y., Suzuki, H., Sakurai, N., and Shibata, D. (2010). CoP: a database for characterizing co-expressed gene modules with biological information in plants. *Bioinformatics* 26, 1267–1268.
- Patel, R. V., Nahal, H. K., Breit, R., and Provart, N. J. (2012). BAR expresso: expression profile similarity ranking of homologous genes in plant species. *Plant J.* 71, 1038–1050.
- Peng, Q., and Ecker, J. R. (2012). Detection of allele-specific methylation through a generalized heterogeneous epigenome model. *Bioinformatics* 28, i163–i171.
- Shannon, P., Markiel, A., Ozier, O., Baliga, N. S., Wang, J. T., Ramage, D., et al. (2003). Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res.* 13, 2498–2504.
- Srinivasasainagendra, V., Page, G. P., Mehta, T., Coulbaly, I., and Loraine, A. E. (2008). CressExpress: a tool for large-scale mining of expression data from *Arabidopsis*. *Plant Physiol.* 147, 1004–1016.
- Steinhauser, D., Usadel, B., Luedemann, A., Thimm, O., and Kopka, J. (2004). CSB.DB: a comprehensive systems-biology database. *Bioinformatics* 20, 3647–3651.
- Theocharidis, A., Van Dongen, S., Enright, A. J., and Freeman, T. C. (2009). Network visualization and analysis of gene expression data using BioLayout Express^{3D}. *Nat. Protoc.* 4, 1535–1550.
- Toufighi, K., Brady, M., Austin, R., Ly, E., and Provart, N. (2005). The Botany Array Resource: e-Northern, Expression Angling, and promoter analyses. *Plant J.* 43, 153–163.
- Usadel, B., Obayashi, T., Mutwil, M., Giorgi, F., Bassel, G., Tanimoto, M., et al. (2009). Co-expression tools for plant biology: opportunities for hypothesis generation and caveats. *Plant Cell Environ.* 32, 1633–1651.
- Weigel, D., and Mott, R. (2009). The 1001 Genomes Project for *Arabidopsis thaliana*. *Genome Biol.* 10, 107.
- Winter, D., Vinegar, B., Nahal, H., Ammar, R., Wilson, G. V., and Provart, N. J. (2007). An 'Electronic Fluorescent Pictograph' browser for exploring and analyzing large-scale biological data sets. *PLoS ONE* 2, e718. doi: 10.1371/journal.pone.0000718
- Yang, Y., Costa, A., Leonhardt, N., Siegel, R., and Schroeder, J. (2008). Isolation of a strong *Arabidopsis* guard cell promoter and its potential as a research tool. *Plant Methods* 4, 6.
- Zimmermann, P., Hirsch-Hoffman, M., Hennig, L., and Gruissem, W. (2004). GENEVESTIGATOR. *Arabidopsis* microarray database and analysis toolbox. *Plant Physiol.* 136, 2621–2632.

Conflict of Interest Statement: The author declares that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Received: 07 August 2012; accepted: 09 October 2012; published online: 29 October 2012.

Citation: Provart N (2012) Correlation networks visualization. *Front. Plant Sci.* 3:240. doi: 10.3389/fpls.2012.00240

This article was submitted to *Frontiers in Plant Systems Biology*, a specialty of *Frontiers in Plant Science*.

Copyright © 2012 Provart. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in other forums, provided the original authors and source are credited and subject to any copyright notices concerning any third-party graphics etc.