Plant Biotechnology 26, 3–7 (2009)

Minireview

Practical network approaches and biologic interpretations of co-expression analyses in plants

Yoshiyuki Ogata, Daisuke Shibata*

Kazusa DNA Research Institute, Kisarazu, Chiba 292-0818, Japan
* E-mail: shibata@kazusa.or.jp Tel: +81-438-52-3900 Fax: +81-438-52-3948

Received December 26, 2008; accepted January 27, 2009 (Edited by T. Obayashi)

Abstract The completion of plant genome sequences and advances in DNA microarray technology have contributed to the accumulation of a vast number of plant gene expression datasets. Co-expression analyses using such datasets can be used to predict the functions of many plant genes. A network approach has been incorporated into co-expression analysis to visualize gene-to-gene functional relatedness (co-expression network analysis). Applying of this analysis to plant gene expression datasets has led to the accumulation of large quantities of information on plant gene function. Plant gene expression datasets and genome-level information obtained using co-expression analyses in plants can be retrieved from various types of databases. Here, we summarize practical approaches for detecting co-expressed genes in plants and review recent progress in plant co-expression analyses.

Key words: Co-expression analysis, gene expression, microarray, network approaches.

Since the completion of the *Arabidopsis thaliana* genome sequence (Arabidopsis Genome Initiative 2000), genome sequences of other plants have also been extensively or completely decoded. Recent advances in DNA microarray technologies allow for the application of microarray analyses to the genes of such plants, leading to the accumulation of datasets in public databases such as Gene Expression Omnibus (GEO; Barrett et al. 2006) and ArrayExpress (Parkinson et al. 2007). As of November 2008, the GEO and ArrayExpress databases contained datasets based on Affymetrix GeneChip microarray slides (or assays) of the following plants: *Arabidopsis thaliana* (4186 and 7308 slides for GEO and ArrayExpress, respectively), *Glycine max* (soy bean; 2999 and 3083 slides), *Hordeum vulgare* (barley; 151 and 904 slides), *Triticum aestivum* (wheat; 309 and 724 slides), *Zea mays* (maize; 249 and 342 slides), *Vitis vinifera* (grape; 104 and 131 slides), *Medicago truncatula* (44 and 127 slides), *Populus trichocarpa* (black cottonwood; 14 and 4 slides), *Oryza sativa* (rice; 230 slides in GEO), *Solanum lycopersicum* (tomato; 50 slides in GEO), *Saccharum officinarum* (sugarcane; 20 slides in GEO), and *Citrus sinensis* (orange; 12 slides in GEO).

The large number of microarray datasets can be used for high-throughput prediction of co-expressed genes, which show similar expression profiles throughout the datasets (Shen-Orr et al. 2002; Wu et al. 2002). Co-expression analysis is based on the hypothesis that co-expressed genes are functionally related to each other; e.g., in metabolic pathways (Li 2002), transcriptional regulation (Lee et al. 2002), protein-protein interactions (Eisen et al. 1998; Ge et al. 2001; Fraser et al. 2004) and biologic process (Eisen et al. 1998; Wen et al. 1998; Tavazoie et al. 1999). Tools and databases that facilitate co-expression analyses for plant genes have been comprehensively reviewed by Coulibaly and Page (2008), Page and Coulibaly (2008), and Suwabe and Yano (2008).

A network approach is a powerful tool for co-expression analysis (Marcotte et al. 1999). A co-expression network comprises vertices (or nodes), representing genes, and vertex-to-vertex links (or edges), representing co-expression relationships for gene pairs. In the network, a group of tightly interconnected genes is called a ‘co-expression module’ (also known as a motif, cluster, subgraph, or subnetwork), in which the genes represent functional relatedness based on co-expression analysis. Thus, a challenging goal in co-expression network analysis is to extract co-expression modules from the whole network. The extraction of co-expression modules requires that an adequate threshold of gene-to-gene association is determined. Luo et al. (2006) and Gupta et al. (2006) reported the relationship between threshold setting and changes in membership of co-expression modules. Various indices used in network analysis have been introduced

Abbreviations: GEO, Gene Expression Omnibus.
This article can be found at http://www.jspcmb.jp/

(D'haeseleer 2005; Dong and Horvath 2007). Co-expression network analysis in plants are recently reviewed by Aoki et al. (2007).

Co-expression analysis, in general, is accompanied by false-positive (co-expressed, but with no effective causality) and false-negative (not co-expressed, but with effective causality) gene-to-gene association (Ge et al. 2003). Yanai et al. (2006) reported on the false-positive association of genes that are expressed specifically in different tissues. False-negative associations are thought to arise from the non-transcriptional regulation, such as post-transcriptional regulation (Saito et al. 2008). Biases in gene expression datasets, e.g., due to the unequal assembly of various datasets, may also create false gene-to-gene associations.

To avoid such biases, a network approach is more useful than methods that use an individual association index for a pair of genes. An approach that eliminates this bias is likely to improve the prediction of gene co-expression (e.g., the Confeito algorithm; submitted by Y.O. et al.).

Practical co-expression network approaches in plants

In this section, we discuss practical approaches for predicting co-expressed genes in plants are categorized (Table 1). The URLs and publications of databases and tools for these approaches are listed in Table 2.

Basic approaches using web-browsing databases

To detect co-expressed genes in plants, a user can browse such genes in public databases for co-expression network analysis. When a user selects one or several gene(s) for this purpose, the genes that are co-expressed with the selected gene(s) can be extracted from databases such as ACT, ATTED-II, and CoP. In the databases, the user

Table 1. Approaches for co-expression network analyses in plants.

Items of interest	Repository	Data processing	Analysis
Basic approaches using web-browsing databases			
One or several genes	Inquiry using gene name(s) or identifier(s) in ACT, ATTED-II, CoP		
A specific biologic process	Inquiry using a name or Gene Ontology identifier in CoP		
Advanced approaches using stand-alone tools			
Specific experiment(s)	Selection of an expression dataset	Calculation of gene-to-gene correlation coefficients	Depiction of a network using Cytoscape and Pajek
General (unselected) items	Acquisition of datasets from GEO and ArrayExpress	Calculation of gene-to-gene correlation coefficients	Extraction of co-expression modules using DP-Clus
General (unselected) items	Acquisition of correlation data from ATTED-II	Not applicable	Extraction of co-expression modules using DP-Clus
General (unselected) items	Acquisition of datasets from GEO and ArrayExpress	Not applicable	Extraction of co-expression modules using ARACNE
Metabolic pathway(s)	Overlay of co-expression links onto metabolic pathways using KaPPA-View 3		

Each row in the Items of interest column represents an individual approach.

Table 2. The URLs and publications of public databases and tools.

Types	Database names and URLs	Organisms	Slides
Repository	GEO (Gene Expression Omnibus; Barrett et al. 2008) http://ncbi.nlm.nih.gov/geo/	12 plants	8368
Repository	ArrayExpress (Parkinson et al. 2008) http://ebi.ac.uk/microarray-as/ae/	10 plants	12759
Viewer	KaPPA-View 3 (Tokimatsu et al. 2005) http://kpv.kazusa.or.jp/kappa-view3/	3 plants	1388
Viewer	Cytoscape (Shannon et al. 2003) http://cytoscape.org/	–	–
Viewer	Pajek (Batagelj and Mrval 1998) http://vlabo.fmf.uni-lj.si/pub/networks/pajek/	–	–
Retrieval	ACT (Manfield et al. 2006) http://www.arabidopsis.leeds.ac.uk/act/	*Arabidopsis thaliana*	322
Retrieval	ATTED-II (Obayashi et al. 2009) http://atted.jp/	*Arabidopsis thaliana*	1388
Retrieval	CoP (Co-expressed biological processes) http://pmnedo.kazusa.or.jp/kagiana/coexprocess	*Arabidopsis thaliana*	3654
Computation	ARACNE (Margolin et al. 2006) http://amdec-bioinfo.cu-genome.org/html/ARACNE.htm	Any	–
Computation	DP-Clus (Altaf-Ul-Amin et al. 2006) http://kanaya.naist.jp/DPClus/	Any	–

The slides column represents the number of datasets based on Affymetrix GeneChip slides in each database.

queries gene identifier(s) or name(s) to obtain a list of co-expressed genes. In the ACT database, the user has a choice of gene expression datasets. ATTED-II provides a co-expression network view of a query gene and its co-expressed genes.

When a user selects a specific biologic process for co-expression analysis, the CoP database allows the user to retrieve co-expressed genes involved in the process on the basis of co-expression network analysis. In this database, the user queries a name or Gene Ontology identifier (The Gene Ontology Consortium 2000) of a biologic process to obtain a list of genes involved in the queried process. Then, by clicking a gene identifier (e.g., an AGI code for an Arabidopsis gene) in the list, a list of co-expressed genes is displayed.

Advanced approaches using stand-alone tools

While databases for co-expression network analysis provide a basic approach to detect co-expressed genes, a detailed co-expression analysis can also be executable using stand-alone tools. When a user searches genes co-expressed in specific experiments for a practical purpose, a practical approach is as follows. The user 1) extracts genes that are specifically expressed in the experiments using a scatter plot to compare the experimental data with the control data; 2) calculates pairwise Pearson correlation coefficients between the extracted genes; and 3) depicts the co-expression network, in which a gene is connected to other genes on the basis of a selected threshold value of the coefficient (e.g., 0.6). Freeware tools such as Cytoscape and Pajek are available for visualizing a network. Datasets for this approach can also be obtained from public databases of expression datasets such as GEO and ArrayExpress.

When a user extracts co-expression modules from a genome-wide network using a large number of microarray datasets, the user 1) obtains the datasets from databases such as GEO and ArrayExpress; 2) calculates Pearson correlation coefficients between all pairs of genes; and 3) extracts co-expression modules using co-expression network analysis tools such as ARACNE, for which the second step is skipped, and DP-Clus. To skip the first and second steps for DP-Clus, the gene-to-gene correlation dataset based on 1388 microarray datasets is available at ATTED-II. To evaluate the relationships between co-expressed genes and metabolic pathways, the KaPPA-View 3 tool allows users to overlay co-expression links onto metabolic pathways of interest.

Co-expression analysis using a large expression dataset

Using a large number of microarray datasets, co-expression analysis enables the genome-wide prediction of gene function (Boutros and Okey 2005; Hasegawa 2006; Michalak 2008). Such co-expression analyses have been merged with 'omics' datasets such as genome (genes as a whole), proteome (proteins), and metabolome (metabolites) to reveal relationships between genes, proteins, and metabolites. Using large-scale gene expression data for plants, co-expression is associated with similarity in the genome sequence in cis-elements (Haberer et al. 2006), in a gene family (Nijhawan et al. 2008), or interspecies (Ma et al. 2005; Ren et al. 2007; Krom and Ramakrishna 2008). Williams and Bowles (2004) reported high co-expression of neighboring genes of Arabidopsis. Schmid et al. (2005) and Hirai et al. (2007) reported similar expression levels between transcription factor genes and metabolic genes in Arabidopsis. Mentzen and Wurtele (2008) predicted 998 sets of co-regulated genes using 963 microarray datasets. Information on gene expression and protein-protein interactions is frequently combined to predict the relationship between co-expression and such interactions (Geisler-Lee et al. 2007). Wienkoop et al. (2008) proposed an approach to integrate complex molecular data including transcripts, proteins, and metabolites. Co-expressed genes may be mapped onto a specific metabolic pathway (Li 2002). To evaluate the effects of genes that encode transcription factors and enzymes on the biosynthesis of metabolites, microarray analysis for plant genes has been integrated with mass spectrometry (Rischer et al. 2006, Yonekura-Sakakibara et al. 2008).

Combining various experimental datasets may lead to biased expression profiles. To alleviate such a bias, Schreiber et al. (2008) and Fukushima et al. (2008) recommended performing the 'singular value decomposition' for expression datasets. To minimize the bias in microarray analysis, it is important to use an expression dataset that was collected with the proper control for a particular purpose. Co-expression relationships that are predicted on the basis of a large-scale microarray analysis that includes a variety of experimental datasets are likely to represent a coexistence or co-absence of genes, not necessarily their functional relatedness. For example, Booth et al. (2005) reported an unexpected correlation between the expression and function of genes involved in various aspects of chemotaxis. To associate co-expressed genes by their functional relatedness, knowledge about their functions and localization may be required.

Perspectives for high-throughput co-expression analysis

Co-expression analysis using large DNA microarray datasets that are available in public databases have been used to predict gene function, as reviewed here.

Moreover, recent advances in whole-genome tiling array technology, which covers the whole genome sequence of an organism with a single chip, allows for the analysis of not only mRNA but also small interfering RNAs and microRNAs (Jones-Rhoades et al. 2007; Gregory et al. 2008). Therefore, co-expression analyses with the new technology will provide new insights into the relationships between mRNA and such previously undetected RNA species. High-throughput sequencing technology, such as 454 sequencing (454 Life Sciences, Branford, CT), Illumina Sequencing (Illumina Inc., San Diego, CA), and SOLiD System Sequencing (Applied Biosystems Inc., Foster City, CA) have recently emerged. The sequencers can be used for large-scale quantification of transcripts as cDNA species (Jones-Rhoades et al. 2007; Hicks et al. 2008; Sittka et al. 2008; Heisel et al. 2008). Therefore, if cost of sequencing becomes as low as that of microarrays, co-expression analysis using such high-throughput sequencing will begin to replace co-expression analysis based on DNA microarray analysis, even for plants for which little genomic information is available.

References

Altaf-Ul-Amin M, Shinbo Y, Mihara K, Kurosawa K, Kanaya S (2006) Development and implementation of an algorithm for detection of protein complexes in large interaction networks. *BMC Bioinformatics* 7: 207

Aoki K, Ogata Y, Shibata D (2007) Approaches for extracting practical information from gene co-expression networks in plant biology. *Plant Cell Physiol* 43: 381–390

Batagelj V, Mrvar A (1998) Pajek—Program for large network analysis. *Connections* 21: 47–57

Barrett T, Troup DB, Wilhite SE, Ledoux P, Rudnev D, Evangelista C, Kim IF, Soboleva A, Tomashevsky M, Marshall KA, Phillippy KH, Sherman PM, Muertter RN, Edgar R (2007) NCBI GEO: archive for high-throughput functional genomic data. *Nucl Acids Res* 37: D885–D890

Booth EO, Van Driessche N, Zhuchenko O, Kuspa A, Shaulsky G (2005) Microarray phenotyping in *Dictyostelium* reveals a regulon of chemotaxis genes. *Bioinformatics* 21: 4371–4377

Boutros PC, Okey AB (2005) Unsupervised pattern recognition: an introduction to the whys and wherefores of clustering microarray data. *Brief Bioinformatics* 6: 331–343

Coulibaly I, Page GP (2008) Bioinformatic tools for inferring functional information from plant microarray data II: analysis beyond single gene. *Int J Plant Genomics* 2008: 893941

D'haeseleer P (2005) How does gene expression clustering work? *Nature Biotechnol* 23: 1499–1501

Dong J, Horvath S (2007) Understanding network concepts in modules. *BMC Systems Biol* 1: 24

Eisen MB, Spellman PT, Brown PO, Botstein D (1998) Cluster analysis and display of genome-wide expression patterns. *Proc Natl Acad Sci USA* 95: 14863–14868

Fraser HB, Hirsh AE, Wall DP, Eisen MB (2004) Coevolution of gene expression among interacting proteins. *Proc Natl Acad Sci USA* 101: 9033–9038

Fukushima A, Wada M, Kanaya S, Arita M (2008) SVD-based anatomy of gene expressions for correlation analysis in Arabidopsis thaliana. *DNA Res* 15: 367–374

Ge H, Liu Z, Church GM, Vidal M (2001) Correlation between transcriptome and interactome mapping data from *Saccharomyces cerevisiae.* *Nature Genet* 29: 482–486

Ge H, Walhout AJM, Vidal M (2003) Integrating 'omics' information: a bridge between genemics and systems biology. *Trends Genet* 19: 551–560

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

Gregory BD, Yazaki J, Ecker JR (2008) Utilizing tiling microarrays for whole-genome analysis in plants. *Plant J* 53: 636–644

Gupta A, Maranas CD, Albert R (2006) Elucidation of directionality for co-expressed genes: predicting intra-operon termination sites. *Bioinformatics* 22: 209–214

Haberer G, Mader MT, Kosarev P, Spannagl M, Yang L, Mayer KFX (2006) Large-scale cis-element detection by analysis of correlated expression and sequence conservation between Arebidopsis and *Brassica oleracea.* *Plant Physiol* 142: 1589–1602

Hasegawa Y, Seki M, Mochizuki Y, Heida N, Hirosawa K, Okamoto N, Sakurai T, Satou M, Akiyama K, Iida K, Lee K, Kanaya S, Demura T, Shinozaki K, Konagaya A, Toyoda T (2006) A flexible representation of omic knowledge for thorough analysis of microarray data. *Plant Methods* 2: 5

Heisel SE, Zhang Y, Allen E, Guo L, Reynolds TL, Yang X, Kovalic D, Roberts JK (2008) Characterization of unique small RNA populations from rice grain. *PLoS ONE* 3: e2871

Hicks JA, Tembhurne P, Liu H (2008) MicroRNA expression in chicken embryos. *Poult Sci* 87: 2335–2343

Hirai MY, Sugiyama K, Sawada Y, Tohge T, Obayashi T, Suzuki A, Araki R, Sakurai N, Suzuki H, Aoki K, Goda H, Nishizawa OI, Shibata D, Saito K (2007) Omics-based identification of *Arabidopsis* Myb transcription factors regulating aliphatic glucosinolate biosynthesis. *Proc Natl Acad Sci USA* 104: 6478–6483

Jones-Rhoades MW, Borevitz JO, Preuss D (2007) Genome-wide expression profiling of the *Arabidopsis* female gametophyte identifies families of small, secreted proteins. *PLoS Genet* 3: e171

Krom N, Ramakrishna W (2008) Comparative analysis of divergent and convergent gene pairs and their expression patterns in rice, Arabidopsis, and *Populus.* *Plant Physiol* 147: 1763–1773

Lee TI, Rinaldi NJ, Robert F, Odom DT, Bar-Joseph Z, Gerber GK, Hannett NM, Harbison CT, Thompson CM, Simon I, Zeitlinger J, Jennings EG, Murray HL, Gordon DB, Ren B, Wyrick JJ, Tagne J, Volkert TL, Fraenkel E, Gifford DK, Young RA (2002) Transcriptional regulatory networks in *Saccharomyces cerevisiae.* *Science* 298: 799–804

Li K (2002) Genome-wide coexpression dynamics: Theory and application. *Proc Natl Acad Sci USA* 99: 16875–16880

Luo F, Zhong J, Yang Y, Zhou J (2006) Application of random matrix theory to microarray data for discovering functional gene modules. *Physical Rev E* 73: 031924

Ma L, Chen C, Liu X, Jiao Y, Su N, Li L, Wang X, Cao M, Sun N, Zhang X, Bao J, Li J, Pedersen S, Bolund L, Zhao H, Yuan L, Wong GK, Wang J, Deng XW, Wang J (2005) A microarray analysis of the rice transcriptome and its comparison to *Arabidopsis*. *Genome Res* 15: 1274–1283

Manfield IW, Jen C, Pinney JW, Michalopoulos I, Bradford JR,

Gilmartin PM, Westhead DR (2006) Arabidopsis Co-expression Tools (ACT): web server tools for microarray-based gene expression analysis. *Nucl Acids Res* 34: W504–W509

Marcotte EM, Pellegrini M, Thompson MJ, Yeates TO, Eisenberg D (1999) A combined algorithm for genome-wide prediction of protein function. *Nature* 402: 83–86

Margolin AA, Nemenman I, Basso K, Wiggins C, Stolovitzky G, Favera RD, Califano A (2006) ARACNE: an algorithm for the reconstruction of gene regulatory networks in a mammalian cellular context. *BMC Bioinformatics* 7: S7

Mentzen WI, Wurtele ES (2008) Regulon organization of Arabidopsis. *BMC Plant Biol* 8: 99

Michalak P (2008) Coexpression, coregulation, and cofunctionality of neighboring genes in eukaryotic genomes. *Genomics* 91: 243–248

Nijhawan A, Jain M, Tyagi AK, Khurana JP (2008) Genomic survey and gene expression analysis of the basic leucine zipper transcription factor family in rice. *Plant Physiol* 146: 333–350

Obayashi T, Hayashi S, Saeki M, Ohta H, Kinoshita K (2009) ATTED-II provides coexpressed gene networks for Arabidopsis. *Nucl Acids Res* 37: D987–D991

Page GP, Coulibaly I (2008) Bioinformatic tools for inferring functional information from plant microarray data: tools for the first steps. *Int J Plant Genomics* 2008: 147563

Parkinson H, Kapushesky M, Shojatalab M, Abeygunawardena N, Coulson R, Farne A, Holloway E, Kolesnykov N, Lilja P, Lukk M, Mani R, Rayner T, Sharma A, William E, Sarkans U, Brazma A (2007) ArrayExpress—a public database of microarray experiments and gene expression profiles. *Nucl Acids Res* 35: D747–D750

Ren X, Stiekema WJ, Nap J (2007) Local coexpression domains in the genome of rice show no microsynteny with Arabidopsis domains. *Plant Mol Biol* 65: 205–217

Rischer H, Orešič M, Seppänen-Laakso T, Katajamaa M, Lammertyn F, Ardiles-Diaz W, Van Montagu MCE, Inzé D, Oksman-Caldentey K, Goossens A (2006) Gene-to-metabolite networks for terpenoid indole alkaloid biosynthesis in Catharanthus roseus cells. *Proc Natl Acad Sci USA* 103: 5614–5619

Saito K, Hirai MY, Yonekura-Sakakibara K (2008) Decoding genes with coexpression networks and metabolomics— 'majority report by precogs'. *Trends Plant Sci* 13: 36–43

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

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. *Nature Genet* 37: 501–506

Schreiber AW, Shirley NJ, Burton RA, Fincher GB (2008) Combining transcriptional datasets using the generalized singular value decomposition. *BMC Bioinformatics* 9: 335

Shen-Orr SS, Milo R, Mangan S, Alon U (2002) Network motifs in the transcriptional regulation network of *Escherichia coli*. *Nature Genet* 31: 64–68

Sittka A, Lucchini S, Papenfort K, Sharma CM, Rolle K, Binnewies TT, Hinton JCD, Vogel J (2008) Deep sequencing analysis of small noncoding RNA and mRNA targets of the global post-transcriptional regulator, Hfq. *PLoS Genet* 4: e1000163

Suwabe K, Yano K (2008) Omics databases in plant science: key to systems biology. *Plant Biotechnol* 25: 413–422

Tavazoie S, Hughes JD, Campbell MJ, Cho RJ, Church GM (1999) Systematic determination of genetic network archtechture. *Nature Genet* 22: 281–285

The Arabidopsis Genome Initiative (2000) Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. *Nature* 408: 796–815

The Gene Ontology Consortium (2000) Gene Ontology: tool for the unification of biology. *Nature Genet* 25: 25–29

Tokimatsu T, Sakurai N, Suzuki H, Ohta H, Nishitani K, Koyama T, Umezawa T, Misawa N, Saito K, Shibata D (2005) KaPPA-View. A web-based analysis tool for integration of transcript and metabolite data on plant metabolic pathway maps. *Plant Physiol* 138: 1289–1300

Wen X, Fuhrman S, Michaels GS, Carr DB, Smith S, Barker JL, Somogyi R (1998) Large-scale temporal gene expression mapping of central nervous system development. *Proc Natl Acad Sci USA* 95: 334–339

Wienkoop S, Morgenthal K, Wolschin F, Scholz M, Selbig J, Weckwerth W (2008) Integration of metabolomic and proteomic phenotypes. *Mol Cell Proteomics* 7: 1725–1736

Williams EJB, Bowles DJ (2004) Coexpression of neighboring genes in the genome of *Arabidopsis thaliana*. *Genome Res* 14: 1060–1067

Wu LF, Hughes TR, Davierwala AP, Robinson MD, Stoughton R, Altschuler SJ (2002) Large-scale prediction of *Saccharomyces cerevisiae* gene function using overlapping transcriptional clusters. *Nature Genet* 31: 255–265

Yanai I, Korbel JO, Boue S, McWeeney SK, Bork P, Lercher MJ (2006) Similar gene expression profiles do not imply similar tissue functions. *Trends Genet* 22: 132–138

Yonekura-Sakakibara K, Tohge T, Matsuda F, Nakabayashi R, Takayama H, Niida R, Watanabe-Takahashi A, Inoue E, Saito K (2008) Comprehensive flavonol profiling and transcriptome coexpression analysis leading to decoding gene-metabolite correlations in *Arabidopsis*. *Plant Cell* 20: 2160–2176