



# Co-expression and co-responses: within and beyond transcription

Takayuki Tohge\* and Alisdair R. Fernie

Max-Planck-Institute of Molecular Plant Physiology, Potsdam-Golm, Germany

## Edited by:

Bjoern Usadel, RWTH Aachen University, Germany

## Reviewed by:

Atsushi Fukushima, RIKEN, Japan  
Kentaro Yano, Meiji University, Japan

## \*Correspondence:

Takayuki Tohge, Max-Planck-Institute of Molecular Plant Physiology, Am Mühlenberg 1, 14476 Potsdam-Golm, Germany.  
e-mail: tohge@mpimp-golm.mpg.de

Whole genome sequencing, the relative ease of transcript profiling by the use of microarrays and latterly RNA sequencing approaches have facilitated the capture of vast amounts of transcript data. However, despite the enormous progress made in gene annotation a substantial proportion of genes remain to be annotated at the functional level. Considerable progress has, however, been made by searching for transcriptional coordination between genes of known function and non-annotated genes on the premise that such co-expressed genes tend to be functionally related. Here we review progress made following this approach as well as its expansion to include phenotypic information from other levels of cellular organization such as proteomic and metabolomic data as well as physiological and developmental phenotypes.

**Keywords:** gene annotation, network analysis, gene expression, plant metabolism, correlation analysis

## INTRODUCTION

Despite the laudable aim of the Arabidopsis 2010 project we remain a long way from knowing the function of every gene of this plant, notwithstanding unprecedented research effort with recent estimates suggesting in the region of 50% of genes are functionally annotated by gene homology and between 10 and 15% have an experimentally verified biological function (Saito et al., 2008; Tohge and Fernie, 2010; Mutwil et al., 2011). The simplicity of homology searches means that at least for dicots the number of genes annotated by homology in rice and soybean or in the more recently published maize (Schnable et al., 2009), poplar (Tuskan et al., 2006), or tomato (Tomato Genome Consortium, 2012) genomes remains reasonable. However, the proportion of genes for which function has been verified experimentally is, at least in most of these species, negligible rendering predictive gene annotation and subsequent validation thereof a vital task for genomics both in model and crop species.

The development and widespread adoption of unbiased RNA sequencing (RNAseq) approaches by plant researchers (Bao et al., 2011; Matas et al., 2011; Hamilton and Buell, 2012; Lohse et al., 2012) effectively increases the scale of this task since it circumvents the need for in depth *a priori* knowledge that was a pre-requisite for microarray hybridizations. Despite the fact that we have as yet not reached satisfactory levels of gene annotation several approaches – all of which are based on a common principal – have recently greatly facilitated gene annotation. This is particularly in the case of pathways under strict transcriptional regulation such as cell wall associated genes and those involved in the various pathways of secondary metabolism as well as leading to the classification of process-associated gene including those linked to cold stress and jasmonate signaling, operon-like genes and seed germination (Hannah et al., 2005; McGrath et al., 2005; Tohge et al., 2005; Saito et al., 2008; Srinivasasainagendra et al., 2008; Mutwil et al., 2009; Obayashi et al., 2009; Usadel et al., 2009; Ogata et al., 2010; Tohge and Fernie, 2010; Bassel et al., 2011; Wada et al., 2012).

These approaches are based on the guilt-by-association approach which assumes that if transcript levels of a gene of unknown function co-respond tightly with those of a gene of known function then it is highly likely that the gene of unknown function plays a role in the same biological process as the known gene. Whilst by no means foolproof, providing a number of considerations and caveats are taken into account, as pointed out in an excellent review by many of the leading investigators in the field (Usadel et al., 2009), then this strategy can prove very powerful. In this mini-review we detail (i) how such approaches have been utilized in a “stand-alone” fashion to successfully predict gene function in Arabidopsis, (ii) how such approaches can be translated for gene functional prediction in crop species for which suitable transcriptomic datasets are publically available, and finally (iii) how other phenotypic data can be incorporated into such studies to support successful gene annotation.

## PREDICTION OF THE FUNCTION OF ARABIDOPSIS GENES

In spite of the clear advantage of biological co-expression network approaches based on gene expression, protein interaction, and genetic interactions for microorganisms such as yeast (see an example, Zhang et al., 2005), co-expression network approaches in plant research have largely been developed solely on the basis of microarray data. This has revealed clear correlations between genes in multiple biosynthetic pathways (Tohge et al., 2007; Movahedi et al., 2011; Mutwil et al., 2011). In addition, *Arabidopsis thaliana* is currently the most useful model plant for integrative analysis due to the availability of several resources such as knockout mutants, cDNA library, tag counts of ESTs, microarray, data and metabolite profiling data. Furthermore, several co-expression gene network analyses and integrative analysis with metabolite profiles have been used to understand the transcriptional correlation networks and discover novel gene functions in this species (Noji et al., 2006; Saito et al., 2008; Mao et al., 2009; Tohge and Fernie, 2010; Mutwil et al., 2011). For this purpose, several web-based

co-expression applications, for example ATTED-II (Obayashi et al., 2009, 2011), AraNet (Hwang et al., 2011), Expression Angler of the Bio-Array Resource (BAR; Toufighi et al., 2005), CressExpress (Srinivasasainagendra et al., 2008), CSB.DB (Steinhauser et al., 2004), KappaViewer (Sakurai et al., 2011), GeneCAT (Mutwil et al., 2008), Genevestigator (Zimmermann et al., 2004), OryzaExpress (Hamada et al., 2011), and VirtualPlant (Katari et al., 2010) have been developed (Table 1).

One of the best examples of co-expression analysis in *Arabidopsis* is cellulose synthase (CESA) genes in the secondary cell wall metabolism (Brown et al., 2005; Persson et al., 2005b), and the primary wall hemicellulose xyloglucan (Cocuron et al., 2007). These studies used the three major secondary wall CESA genes as the baits to construct networks and find novel functional genes displaying similar expression patterns. The CESA gene network has been in several publications (Persson et al., 2005a; Mutwil et al., 2008, 2009, 2011; Ruprecht et al., 2011). A second successful example of the co-expression approach is that of plant secondary metabolism, since this is of the directly regulated at the transcriptional level by a range of different transcription factors including the MYB transcription factors. Since a framework of flavonoid co-expression network was constructed for identify the flavonol-3'-*O*-methyltransferase (AtOMT1; Tohge et al., 2007), such co-expression network approaches have been expanded to find other flavonoid biosynthetic genes such as flavonol-7-*O*-rhamnosyltransferase (At1g06000) and flavonol-3-*O*-arabinosyltransferase (At5g17030; Yonekura-Sakakibara et al., 2007, 2008; Tohge and Fernie, 2010). In addition, this approach was also utilized in the identification of glucosinolate MYB regulators (AtMYB28 and AtMYB29; Hirai et al., 2007), monolignol transporter (AtABCG29) involved in lignin biosynthesis (Alejandro et al., 2012) and novel signaling related candidate genes and transporters following the exposure of *Arabidopsis* to UV-B (Tohge et al., 2011a).

In addition to its utility in understanding the regulation of individual metabolic pathways or even metabolic networks co-expression analysis has also been applied at a much broader level to look at tissue-specific transcriptional networks (Song et al., 2010) and at diverse biological processes including seed germination and dark-induced senescence (Araújo et al., 2011; Bassel et al., 2011). Studying gene sharing networks of *Arabidopsis* and rice Song et al. (2010) discovered that tissues or cell types from the same organ system tend to group together to form network modules. The operon-like clusters in *Arabidopsis* using genome-based co-expression network analysis has been found (Wada et al., 2012). Furthermore, plant tissues in consecutive developmental stages or sharing physiological functions are highly connected. Extending their comparisons to mouse and human gene expression data they were able to observe common principles of gene-sharing across the species and hypothesize that gene sharing evolved as a fundamental organizing feature of gene expression in eukaryotes. The co-expression approach was also successfully applied to microarray data across the entire seed germination process (Bassel et al., 2011). The output, which the authors termed SeedNet (<http://vseed.nottingham.ac.uk>), facilitated the definition of two state-dependent interactions associated with either dormancy or germination with an intermediate transition region between

**Table 1 | Co-expression databases presented in this article.**

Species	Database	URL
<b>Cross species</b>		
	PlaNet	<a href="http://aranet.mpimp-golm.mpg.de/">http://aranet.mpimp-golm.mpg.de/</a>
<b><i>Arabidopsis thaliana</i></b>		
	ATTED-II	<a href="http://atted.jp/">http://atted.jp/</a>
	AraNet	<a href="http://www.functionalnet.org/aranet/">http://www.functionalnet.org/aranet/</a>
	BAR	<a href="http://bar.utoronto.ca/welcome.htm">http://bar.utoronto.ca/welcome.htm</a>
	COP	<a href="http://webs2.kazusa.or.jp/kagiana/cop">http://webs2.kazusa.or.jp/kagiana/cop</a>
	CORNet	<a href="http://bioinformatics.psb.ugent.be/cornet">http://bioinformatics.psb.ugent.be/cornet</a>
	CSB.DB	<a href="http://csbdb.mpimp-golm.mpg.de/csbdb/dbcor/ath.html">http://csbdb.mpimp-golm.mpg.de/csbdb/dbcor/ath.html</a>
	CressExpress	<a href="http://www.cressexpress.org/index">http://www.cressexpress.org/index</a>
	GeneCAT	<a href="http://genecat.mpg.de/cgi-bin/Ainitiator.py">http://genecat.mpg.de/cgi-bin/Ainitiator.py</a>
	GeneVestigator	<a href="https://www.genevestigator.com/gv/plant.jsp">https://www.genevestigator.com/gv/plant.jsp</a>
	KappaViewer4	<a href="http://kpv.kazusa.or.jp/en/">http://kpv.kazusa.or.jp/en/</a>
	SeedNet	<a href="http://bree.cs.nott.ac.uk/arabidopsis/">http://bree.cs.nott.ac.uk/arabidopsis/</a>
<b><i>Oryza sativa</i></b>		
	OryzaExpress	<a href="http://bioinf.mind.meiji.ac.jp/Rice_network_public/script/index.html">http://bioinf.mind.meiji.ac.jp/Rice_network_public/script/index.html</a>
	ATTED-II	<a href="http://atted.jp/">http://atted.jp/</a>
<b><i>Populus trichocarpa</i></b>		
	COP	<a href="http://webs2.kazusa.or.jp/kagiana/cop/">http://webs2.kazusa.or.jp/kagiana/cop/</a>

the two being characterized by an enrichment of genes involved in cellular phase transitions. Moreover, the dormancy region of the co-expression network was strongly associated to abiotic stress response genes. The combined findings were thus taken to suggest that seed dormancy is an adaptive trait that arose evolutionarily late and evolved by coopting existing biosynthetic pathways regulating cellular phase transitions and abiotic stress response genes. During dark-induced senescence there is a dramatic switch from respiration of sugars to respiration of protein which is underpinned by dramatic transcriptional reprogramming of metabolism (Araújo et al., 2010; Araujo et al., 2011), including the degradation of lysine and branched chain amino acids by as yet undefined pathways. In this case the co-expression response was able to provide a high number of candidate genes involved in this process (Araújo et al., 2011), however these remain to be functionally verified.

## PREDICTION OF THE FUNCTION OF CROP GENES: WITHIN SPECIES COMPARISONS

Although considerably fewer microarray experiments have been reported for crop species, with the possible exceptions of rice, several examples exist of the power of the approach in stand-alone network analyses for rice and tomato (Ficklin et al., 2010; Ozaki et al., 2010; Rohrmann et al., 2011; Sakurai et al., 2011; Fukushima et al., 2012). We will here shortly review these studies and highlight the important knowledge inference for studies in tomato and the grasses. In tomato the most comprehensive

study was that performed by Fukushima et al. (2012) who constructed co-response networks from 327 tomato Affymetrix arrays. Although this dataset was substantially smaller than that regularly used for Arabidopsis a number of important conclusion could be drawn including biologically relevant co-expression networks including DNA endoreduplication, response to cold, jasmonate-associated metabolic processes, and the ubiquitous photosynthetic gene cluster. The study also revealed that duplicated genes often displayed differential co-expression when tissue-type was studied a fact highlighted by genes of lycopene and flavonoid biosynthesis (Fukushima et al., 2012). In two more targeted analyses co-expression analysis was also linked to metabolite levels in tomato fruit (Rohrmann et al., 2011; Lee et al., 2012), however, we will return to these studies later when discussing *layering in other phenotypes to aid annotation strategies*.

In addition to these recent studies in tomato there have also been studies in barley (*Hordeum vulgare*; Faccioli et al., 2005; Mochida et al., 2011; Tohge et al., 2011b), wheat (Manickavelu et al., 2012), rice (Fukushima et al., 2009; Lee et al., 2009; Ficklin et al., 2010; Childs et al., 2011; Hamada et al., 2011), maize (Ficklin and Feltus, 2011), poplar (*Populus* spp.; Ogata et al., 2010), and tobacco (*Nicotiana tabacum*; Edwards et al., 2010). Studies in rice revealed that gene co-expression analysis facilitated elucidation of gene function. With the study of Ficklin et al. (2010) returning 45 co-expressed gene modules and 76 cofunctional gene clusters some of which were enriched for previously characterized mutant phenotypes thus providing strong hints toward molecular functions of unknown genes within the clusters with similar outcomes being achieved for the other species mentioned above.

### PREDICTION OF THE FUNCTION OF CROP GENES: BETWEEN SPECIES COMPARISONS

Whilst the above described studies show that there is considerable benefit from co-expression analysis in species such as tomato for which genome scale microarray platforms do not yet exist another approach that has been demonstrated to be highly powerful is combining comparisons of gene cluster networks and sequence homology as a method of assigning gene function and was recently published under the acronym PlaNet (Mutwil et al., 2011). PlaNet builds on the concept first published in 2008 by the same group which already described the search for barley gene orthologs of annotated Arabidopsis genes (Mutwil et al., 2008). PlaNet extended this to include the crop species barley, medicago, poplar, rice, soybean, and wheat, and used a comparative network algorithm to estimate similarities between network structures. The algorithm was exemplified using the canonical the photosystem I reaction center (*PSA-D*) family gene-related networks as well as those related to chalcone synthase suggesting that the rapid transfer of knowledge between species will be possible. That this is so was recently also demonstrated by the same group in a study of secondary wall cellulose biosynthesis (Ruprecht et al., 2011). In this study, the authors compared co-expressed gene vicinity networks of primary and secondary wall CESAs in all species housed in PlaNet to identify those genes consistently co-regulated with cellulose biosynthesis. In addition to the expected polysaccharide acting enzymes, they also found many gene families associated with cytoskeleton, signaling, transcriptional regulation, oxidation,

and protein degradation. Based on these analyses, they selected and biochemically analyzed T-DNA insertion lines corresponding to approximately 20 genes from gene families that re-occur in the co-expressed gene vicinity networks of secondary wall CESAs across the seven species. One of the mutants, corresponding to a pinorensin reductase gene, was subsequently characterized as displaying disturbed xylem morphology and containing lower levels of lignin than the wild-type.

The very same seven species used for the PlaNet study were used in an independent study to generate a pipeline within the BAR software suite (Toufighi et al., 2005) to rank ortholog predictions based on sequence and expression profile similarity with the best fitting on this criteria being defined as the *expressolog* (Patel et al., 2012). Interestingly, global analyses revealed that orthologs with the highest sequence similarity do not necessarily exhibit the highest expression pattern similarity. Moreover, other putative orthologs show highly distinct expression patterns suggesting they may need re-annotating or at best to be given a more specific annotation. A similar comprehensive comparison between maize and rice was additionally recently carried out using the IsoRank tool (Ficklin and Feltus, 2011). It thus appears likely that both these tools as well as PlaNet will likely greatly aid translational efforts to translate the huge knowledge we have gained from Arabidopsis studies into crop species.

### LAYERING IN OTHER PHENOTYPES TO AID ANNOTATION STRATEGIES

The above examples have by and large only relied on data from transcript profiling and have neither harnessed information derived from other molecular approaches, such as proteomics and metabolomics, nor indeed of end-phenotypes such as total yield and harvest indexes. Several recent studies have however incorporated such data collected in order to complement transcriptomic efforts of gene functional annotation (Hirai et al., 2007; Horan et al., 2008; Yonekura-Sakakibara et al., 2008; Sulpice et al., 2009; Allen et al., 2010; Tohge and Fernie, 2010; Araujo et al., 2011; Rohrmann et al., 2011; Tohge et al., 2011a,b). Returning to the tomato examples mentioned above, in order to exploit the impact of tomato genetic diversity on carotenoids, Lee et al. (2012) used *Solanum pennellii* introgression lines as a source of defined natural variation and as a resource for the identification of candidate regulatory genes. For this purpose ripe fruits were analyzed for numerous fruit metabolites and transcriptome profiles generated using a 12,000 unigene oligoarray. Correlation analysis between carotenoid content and gene expression profiles revealed 953 carotenoid-correlated genes. A subnetwork analysis of carotenoid-correlated transcription narrowed this down to 38 candidates. One of which, *Solanum lycopersicum* ethylene response factor 6 (*SIERF6*), was subsequently functionally characterized revealing that it indeed influences carotenoid biosynthesis and additional ripening phenotypes. In a similar approach Rohrmann et al. (2011) developed a quantitative real-time PCR platform allowing accurate quantification of the expression level of approximately 1000 tomato transcription factors. In addition to utilizing this novel approach, they performed cDNA microarray analysis and metabolite profiling of primary and secondary metabolites using gas chromatography–mass spectrometry (GC–MS) and

liquid chromatography–mass spectrometry (LC–MS), respectively. Applying these platforms to pericarp material harvested throughout fruit development and studying both wild-type *Solanum lycopersicum* cv. Ailsa Craig and the *hp1* (*high pigment*) mutant which is functionally deficient in the tomato homolog of the negative regulator of the light signal transduction gene UV-DAMAGED DNA BINDING PROTEIN 1 (*DDBI*) from Arabidopsis. They chose this particular mutant since it had previously been shown to harbor dramatic alterations in the content of several important fruit metabolites but relatively little impact on other ripening phenotypes. The combined dataset was extensively mined searching for co-responsive metabolites and transcription factors, and, where possible, the respective transcriptional expression network underlying this control. Two further studies in tomato merit discussion here. Mounet et al. (2009), used a combination of metabolite profiling and transcript profiling to identify candidate for the key factor of fruit composition and development. More recently Osorio et al. (2011), used a combination of transcriptomics, proteomics, and metabolomics alongside network computation to assess ripening across a range of classical ripening mutants and recently extended this analysis to compare ripening in tomato with that in pepper (Osorio et al., 2012).

Staying with the integration of transcriptomic, proteomic, and metabolomic data we recently combined data from all three platforms to infer function within the tonoplast proteome (Tohge et al., 2011b). In order to do so we performed metabolic profiling of both primary and secondary metabolites in highly purified vacuoles of barley or the protoplast preparations from which they were isolated. This gave us quantitative data on 59 primary metabolites for which we knew the exact chemical structure and some 200 secondary metabolites for which we had strong predicted chemical formulae. This data was then compared to the 88 tonoplast proteins reported for barley (Endler et al., 2006) and evaluating their co-expression using PlaNet. This strategy allowed us to putatively assign transport function for phenylpropanoids, flavonoids, storage proteins, and mugonic acid, as well as a potential transport system for phytosiderophores.

Proteomic data are also an important component of the interaction networks that form part of the CORNET tools (De Bodt et al., 2010, 2012) which combine co-expression analysis with protein–protein interaction searches. The latter is similar to other tools such as those in CressExpress, BAR, and VirtualPlant (Toufighi

et al., 2005; Srinivasasainagendra et al., 2008; Katari et al., 2010), however, it presents microarray data with the corresponding meta-data including sample information, protein–protein interaction data, localization data, and functional information within a single central database. Developed CORNET 2.0 includes the majority of interaction databases, six different protein–protein interaction dataset, and three sets of regulatory interaction data, thereby providing with consistently updated data sets for versatile searches (De Bodt et al., 2012). The efficacy of computational classification to enrich potential protein–protein interactions to predict putative interactions of Arabidopsis membrane protein has been applied (Chen et al., 2012). This method is also an important to fill gaps to biological networks and suggest hypothetical process and genes involving signal transduction and transport.

## CONCLUSION

It is hopefully clearly apparent from this mini-review that co-expression analyses are a very powerful tool in gene annotation not only in model systems such as Arabidopsis and rice but also in less well characterized plant species. To date, it has found great utility in improving our understanding of pathways which are known to be regulated at the transcriptional level such as cell wall biosynthesis and various pathways of secondary metabolism, however, recent examples also demonstrate its utility in elucidating novel players in various developmental processes. The guilt-by-association response is clearly powerful even in stand-alone single species approaches. However, the increasing availability of data from multiple species and at multiple different levels of the cellular hierarchy will likely facilitate the adoption of integrative genomics approaches by many more laboratories in the near future. Even some 6–9 years ago the power of combining transcript and metabolite profiling for (candidate) gene discovery was demonstrated for non-sequenced species (Urbanczyk-Wochniak et al., 2003; Rischer et al., 2006). Recent developments in RNA sequencing (Schneeberger and Weigel, 2011), will likely render this considerably easier in the near future.

## ACKNOWLEDGMENTS

Research activity of Takayuki Tohge is supported by the Alexander von Humboldt Foundation. Funding from the Max-Planck-Society (to Takayuki Tohge and Alisdair R. Fernie) is gratefully acknowledged.

## REFERENCES

- Alejandro, S., Lee, Y., Tohge, T., Sudre, D., Osorio, S., Park, J., et al. (2012). AtABCG29 is a monolignol transporter involved in lignin biosynthesis. *Curr. Biol.* 10, 1207–1212.
- Allen, E., Moing, A., Ebbels, T. M., Maucourt, M., Tomos, A. D., Rolin, D., et al. (2010). Correlation network analysis reveals a sequential reorganization of metabolic and transcriptional states during germination and gene-metabolite relationships in developing seedlings of Arabidopsis. *BMC Syst. Biol.* 4, 62. doi: 10.1186/1752-0509-4-62
- Araújo, W. L., Ishizaki, K., Nunes-Nesi, A., Larson, T. R., Tohge, T., Krahnert, I., et al. (2010). Identification of the 2-hydroxyglutarate and isovaleryl-CoA dehydrogenases as alternative electron donors linking lysine catabolism to the electron transport chain of Arabidopsis mitochondria. *Plant Cell* 22, 1549–1563.
- Araújo, W. L., Ishizaki, K., Nunes-Nesi, A., Tohge, T., Larson, T. R., Krahnert, I., et al. (2011). Analysis of a range of catabolic mutants provides evidence that phytanoyl-coenzyme A does not act as a substrate of the electron-transfer flavoprotein/electron-transfer flavoprotein:ubiquinone oxidoreductase complex in Arabidopsis during dark-induced senescence. *Plant Physiol.* 157, 55–69.
- Araújo, W. L., Tohge, T., Ishizaki, K., Leaver, C. J., and Fernie, A. R. (2011). Protein degradation – an alternative respiratory substrate for stressed plants. *Trends Plant Sci.* 16, 489–498.
- Bao, E., Jiang, T., Kaloshian, I., and Girke, T. (2011). SEED: efficient clustering of next-generation sequences. *Bioinformatics* 27, 2502–2509.
- 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.
- Brown, D. M., Zeef, L. A. H., Ellis, J., Goodacre, R., and Turner, S. R. (2005). Identification of novel genes in Arabidopsis involved in secondary cell wall formation using expression profiling and reverse genetics. *Plant Cell* 17, 2281–2295.

- Chen, J., Lalonde, S., Obrdlik, P., Noorani Vatani, A., Parsa, S., Vilarino, C., et al. (2012). Uncovering Arabidopsis membrane protein interactome enriched in transporters using mating-based split ubiquitin assays and classification models. *Front. Plant Sci.* 3:124. doi: 10.3389/fpls.2012.00124
- Childs, K. L., Davidson, R. M., and Buell, C. R. (2011). Gene coexpression network analysis as a source of functional annotation for rice genes. *PLoS ONE* 6, e22196. doi: 10.1371/journal.pone.0022196
- Cocuron, J.-C., Lerouxel, O., Drakakaki, G., Alonso, A. P., Liepman, A. H., Keegstra, K., et al. (2007). A gene from the cellulose synthase-like C family encodes a beta-1,4 glucan synthase. *Proc. Natl. Acad. Sci. U.S.A.* 104, 8550–8555.
- De Bodt, S., Carvajal, D., Hollunder, J., Van den Cruyce, J., Movahedi, S., and Inze, D. (2010). CORNET: a user-friendly tool for data mining and integration. *Plant Physiol.* 152, 1167–1179.
- De Bodt, S., Hollunder, J., Nelissen, H., Meulemeester, N., and Inzé, D. (2012). CORNET 2.0: integrating plant coexpression, protein-protein interactions, regulatory interactions, gene associations and functional annotations. *New Phytol.* 195, 707–720.
- Edwards, K. D., Bombarely, A., Story, G. W., Allen, F., Mueller, L. A., Coates, S. A., et al. (2010). TobEA: an atlas of tobacco gene expression from seed to senescence. *BMC Genomics* 11, 142. doi: 10.1186/1471-2164-11-142
- Endler, A., Meyer, S., Schelbert, S., Schneider, T., Weschke, W., Peters, S. W., et al. (2006). Identification of a vacuolar sucrose transporter in barley and Arabidopsis mesophyll cells by a tonoplast proteomic approach. *Plant Physiol.* 141, 196–207.
- Faccioli, P., Provero, P., Herrmann, C., Stanca, A. M., Morcia, C., and Terzi, V. (2005). From single genes to co-expression networks: extracting knowledge from barley functional genomics. *Plant Mol. Biol.* 58, 739–750.
- Ficklin, S. P., and Feltus, F. A. (2011). Gene coexpression network alignment and conservation of gene modules between two grass species: maize and rice. *Plant Physiol.* 156, 1244–1256.
- Ficklin, S. P., Luo, F., and Feltus, F. A. (2010). The association of multiple interacting genes with specific phenotypes in rice using gene coexpression networks. *Plant Physiol.* 154, 13–24.
- Fukushima, A., Kanaya, S., and Arita, M. (2009). Characterizing gene coexpression modules in *Oryza sativa* based on a graph-clustering approach. *Plant Biotechnol.* 26, 485–493.
- Fukushima, A., Nishizawa, T., Hayakumo, M., Hikosaka, S., Saito, K., Goto, E., et al. (2012). Exploring tomato gene functions based on coexpression modules using graph clustering and differential coexpression approaches. *Plant Physiol.* 158, 1487–1502.
- Hamada, K., Hongo, K., Suwabe, K., Shimizu, A., Nagayama, T., Abe, R., et al. (2011). OryzaExpress: an integrated database of gene expression networks and omics annotations in rice. *Plant Cell Physiol.* 52, 220–229.
- Hamilton, J. P., and Buell, C. R. (2012). Advances in plant genome sequencing. *Plant J.* 70, 177–190.
- Hannah, M. A., Heyer, A. G., and Hincha, D. K. (2005). A global survey of gene regulation during cold acclimation in *Arabidopsis thaliana*. *PLoS Genet.* 1, 179–196. doi: 10.1371/journal.pgen.0010026
- Hirai, M. Y., Sugiyama, K., Sawada, Y., Tohge, T., Obayashi, T., Suzuki, A., et al. (2007). Omics-based identification of Arabidopsis Myb transcription factors regulating aliphatic glucosinolate biosynthesis. *Proc. Natl. Acad. Sci. U.S.A.* 104, 6478–6483.
- Horan, K., Jang, C., Bailey-Serres, J., Mittler, R., Shelton, C., Harper, J. E., et al. (2008). Annotating genes of known and unknown function by large-scale coexpression analysis. *Plant Physiol.* 147, 41–57.
- Hwang, S., Rhee, S. Y., Marcotte, E. M., and Lee, I. (2011). Systematic prediction of gene function in *Arabidopsis thaliana* using a probabilistic functional gene network. *Nat. Protoc.* 6, 1429–1442.
- Katari, M. S., Nowicki, S. D., Aceituno, F. F., Nero, D., Kelfer, J., Thompson, L. P., et al. (2010). VirtualPlant: a software platform to support systems biology research. *Plant Physiol.* 152, 500–515.
- Lee, J. M., Joungh, J.-G., McQuinn, R., Chung, M.-Y., Fei, Z., Tieman, D., et al. (2012). Combined transcriptome, genetic diversity and metabolite profiling in tomato fruit reveals that the ethylene response factor SIERF6 plays an important role in ripening and carotenoid accumulation. *Plant J.* 70, 191–204.
- Lee, T.-H., Kim, Y.-K., Pham, T. T. M., Song, S. I., Kim, J.-K., Kang, K. Y., et al. (2009). RiceArrayNet: a database for correlating gene expression from transcriptome profiling, and its application to the analysis of coexpressed genes in rice. *Plant Physiol.* 151, 16–33.
- Lohse, M., Bolger, A., Nagel, A., Fernie, A., Lunn, J., Stitt, M., et al. (2012). RobiNA: a user-friendly, integrated software solution for RNA-seq-based transcriptomics. *Nucleic Acids Res.* 40, W622–W627.
- Manickavelu, A., Kawaura, K., Oishi, K., Shin-I, T., Kohara, Y., Yahiaoui, N., et al. (2012). Comprehensive functional analyses of expressed sequence tags in common wheat (*Triticum aestivum*). *DNA Res.* 19, 165–177.
- Mao, L., Van Hemert, J. L., Dash, S., and Dickerson, J. A. (2009). Arabidopsis gene co-expression network and its functional modules. *BMC Bioinformatics* 10, 346. doi: 10.1186/1471-2105-10-346
- Matas, A. J., Yeats, T. H., Buda, G. J., Zheng, Y., Chatterjee, S., Tohge, T., et al. (2011). Tissue- and cell-type specific transcriptome profiling of expanding tomato fruit provides insights into metabolic and regulatory specialization and cuticle formation. *Plant Cell* 23, 3893–3910.
- McGrath, K. C., Dombrecht, B., Manners, J. M., Schenk, P. M., Edgar, C. I., Maclean, D. J., et al. (2005). Repressor- and activator-type ethylene response factors functioning in jasmonate signaling and disease resistance identified via a genome-wide screen of Arabidopsis transcription factor gene expression. *Plant Physiol.* 139, 949–959.
- Mochida, K., Uehara-Yamaguchi, Y., Yoshida, T., Sakurai, T., and Shinozaki, K. (2011). Global landscape of a co-expressed gene network in barley and its application to gene discovery in Triticeae crops. *Plant Cell Physiol.* 52, 785–803.
- Mounet, F., Moing, A., Garcia, V., Petit, J., Maucourt, M., Deborde, C., et al. (2009). Gene and metabolite regulatory network analysis of early developing fruit tissues highlights new candidate genes for the control of tomato fruit composition and development. *Plant Physiol.* 149, 1505–1528.
- 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.
- Mutwil, M., Obro, J., Willats, W. G. T., and Persson, S. (2008). GeneCAT – novel webtools that combine BLAST and co-expression analyses. *Nucleic Acids Res.* 36, W320–W326.
- Mutwil, M., Ruprecht, C., Giorgi, F. M., Bringmann, M., Usadel, B., and Persson, S. (2009). Transcriptional wiring of cell wall-related genes in Arabidopsis. *Mol. Plant* 2, 1015–1024.
- Noji, M., Kawashima, C. G., Obayashi, T., and Saito, K. (2006). *In silico* assessment of gene function involved in cysteine biosynthesis in Arabidopsis: expression analysis of multiple isoforms of serine acetyltransferase. *Amino Acids* 30, 163–171.
- 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.
- Obayashi, T., Nishida, K., Kasahara, K., and Kinoshita, K. (2011). ATTED-II updates: condition-specific gene coexpression to extend coexpression analyses and applications to a broad range of flowering plants. *Plant Cell Physiol.* 52, 213–219.
- 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.
- Osorio, S., Alba, R., Damasceno, C. M. B., Lopez-Casado, G., Lohse, M., Zanor, M. I., et al. (2011). Systems biology of tomato fruit development: combined transcript, protein, and metabolite analysis of tomato transcription factor (nor, rin) and ethylene receptor (Nr) mutants reveals novel regulatory interactions. *Plant Physiol.* 157, 405–425.
- Osorio, S., Alba, R., Nikoloski, Z., Kochevenko, A., Fernie, A. R., and Giovannoni, J. J. (2012). Integrative comparative analyses of transcript and metabolite profiles from pepper and tomato ripening and development stages uncovers species-specific patterns of network regulatory behaviour. *Plant Physiol.* 159, 1713–1729.
- Ozaki, S., Ogata, Y., Suda, K., Kurabayashi, A., Suzuki, T., Yamamoto, N., et al. (2010). Coexpression analysis of tomato genes and experimental verification of coordinated expression of genes found in a functionally enriched coexpression module. *DNA Res.* 17, 105–116.

- Patel, R. V., Nahal, H. K., Breit, R., and Provart, N. J. (2012). BAR expression profile similarity ranking of homologous genes in plant species. *Plant J.* 71, 1038–1050.
- Persson, S., Wei, H., Milne, J., Page, G., and Somerville, C. (2005a). Large-scale coexpression analysis reveals novel genes involved in cellulose biosynthesis. *Plant Biol. (Rockville)* 2005, 176.
- Persson, S., Wei, H. R., Milne, J., Page, G. P., and Somerville, C. R. (2005b). Identification of genes required for cellulose synthesis by regression analysis of public microarray data sets. *Proc. Natl. Acad. Sci. U.S.A.* 102, 8633–8638.
- Rischer, H., Oresic, M., Seppanen-Laakso, T., Katajamaa, M., Lammerly, F., Ardiles-Diaz, W., et al. (2006). Gene-to-metabolite networks for terpenoid indole alkaloid biosynthesis in *Catharanthus roseus* cells. *Proc. Natl. Acad. Sci. U.S.A.* 103, 5614–5619.
- Rohrmann, J., Tohge, T., Alba, R., Osorio, S., Caldana, C., McQuinn, R., et al. (2011). Combined transcription factor profiling, microarray analysis and metabolite profiling reveals the transcriptional control of metabolic shifts occurring during tomato fruit development. *Plant J.* 68, 999–1013.
- Ruprecht, C., Mutwil, M., Saxe, F., Eder, M., Nikoloski, Z., and Persson, S. (2011). Large-scale co-expression approach to dissect secondary cell wall formation across plant species. *Front. Plant Sci.* 2:23. doi: 10.3389/fpls.2011.00023
- Saito, K., Hirai, M. Y., and Yonekura-Sakakibara, K. (2008). Decoding genes with coexpression networks and metabolomics – “majority report by precogs.” *Trends Plant Sci.* 13, 36–43.
- Sakurai, N., Ara, T., Ogata, Y., Sano, R., Ohno, T., Sugiyama, K., et al. (2011). KaPPA-View4: a metabolic pathway database for representation and analysis of correlation networks of gene co-expression and metabolite co-accumulation and omics data. *Nucleic Acids Res.* 39, D677–D684.
- Schnable, P. S., Ware, D., Fulton, R. S., Stein, J. C., Wei, F., Pasternak, S., et al. (2009). The B73 maize genome: complexity, diversity, and dynamics. *Science* 326, 1112–1115.
- Schneeberger, K., and Weigel, D. (2011). Fast-forward genetics enabled by new sequencing technologies. *Trends Plant Sci.* 16, 282–288.
- Song, W. Y., Park, J., Mendoza-Cozatl, D. G., Suter-Grotemeyer, M., Shim, D., Hortensteiner, S., et al. (2010). Arsenic tolerance in *Arabidopsis* is mediated by two ABC-type phytochelatin transporters. *Proc. Natl. Acad. Sci. U.S.A.* 107, 21187–21192.
- Srinivasainagendra, V., Page, G. P., Mehta, T., Coulibaly, 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.
- Sulpice, R., Pyl, E. T., Ishihara, H., Trenkamp, S., Steinfath, M., Witucka-Wall, H., et al. (2009). Starch as a major integrator in the regulation of plant growth. *Proc. Natl. Acad. Sci. U.S.A.* 106, 10348–10353.
- Tohge, T., and Fernie, A. R. (2010). Combining genetic diversity, informatics and metabolomics to facilitate annotation of plant gene function. *Nat. Protoc.* 5, 1210–1227.
- Tohge, T., Kusano, M., Fukushima, A., Saito, K., and Fernie, A. (2011a). Transcriptional and metabolic programs following exposure of plants to UV-B irradiation. *Plant Signal Behav.* 6, 1987–1992.
- Tohge, T., Nishiyama, Y., Hirai, M. Y., Yano, M., Nakajima, J., Awazuhara, M., et al. (2005). Functional genomics by integrated analysis of metabolome and transcriptome of *Arabidopsis* plants over-expressing an MYB transcription factor. *Plant J.* 42, 218–235.
- Tohge, T., Ramos, M. S., Nunes-Nesi, A., Mutwil, M., Giavalisco, P., Steinhäuser, D., et al. (2011b). Toward the storage metabolome: profiling the barley vacuole. *Plant Physiol.* 157, 1469–1482.
- Tohge, T., Yonekura-Sakakibara, K., Niida, R., Watanabe-Takahashi, A., and Saito, K. (2007). Phytochemical genomics in *Arabidopsis thaliana*: a case study for functional identification of flavonoid biosynthesis genes. *Pure Appl. Chem.* 79, 811–823.
- Tomato Genome Consortium. (2012). The tomato genome sequence provides insights into fleshy fruit evolution. *Nature* 485, 635–641.
- Toufighi, K., Brady, S. M., Austin, R., Ly, E., and Provart, N. J. (2005). The Botany Array Resource: e-Northern, expression angling, and promoter analyses. *Plant J.* 43, 153–163.
- Tuskan, G. A., DiFazio, S., Jansson, S., Bohlmann, J., Grigoriev, I., Hellsten, U., et al. (2006). The genome of black cottonwood, *Populus trichocarpa* (Torr. & Gray). *Science* 313, 1596–1604.
- Urbanczyk-Wochniak, E., Luedemann, A., Kopka, J., Selbig, J., Roessner-Tunali, U., Willmitzer, L., et al. (2003). Parallel analysis of transcript and metabolic profiles: a new approach in systems biology. *EMBO Rep.* 4, 989–993.
- Usadel, B., Obayashi, T., Mutwil, M., Giorgi, F. M., Bassel, G. W., Tanimoto, M., et al. (2009). Co-expression tools for plant biology: opportunities for hypothesis generation and caveats. *Plant Cell Environ.* 32, 1633–1651.
- Wada, M., Takahashi, H., Altaf-Ul-Amin, M., Nakamura, K., Hirai, M. Y., Ohta, D., et al. (2012). Prediction of operon-like gene clusters in the *Arabidopsis thaliana* genome based on co-expression analysis of neighboring genes. *Gene* 503, 56–64.
- Yonekura-Sakakibara, K., Tohge, T., Matsuda, F., Nakabayashi, R., Takayama, H., Niida, R., et al. (2008). Comprehensive flavonol profiling and transcriptome coexpression analysis leading to decoding gene-metabolite correlations in *Arabidopsis*. *Plant Cell* 20, 2160–2176.
- Yonekura-Sakakibara, K., Tohge, T., Niida, R., and Saito, K. (2007). Identification of a flavonol 7-O-rhamnosyltransferase gene determining flavonoid pattern in *Arabidopsis* by transcriptome coexpression analysis and reverse genetics. *J. Biol. Chem.* 282, 14932–14941.
- Zhang, L., King, O., Wong, S., Goldberg, D., Tong, A., Lesage, G., et al. (2005). Motifs, themes and thematic maps of an integrated *Saccharomyces cerevisiae* interaction network. *J. Biol.* 4, 6.
- Zimmermann, P., Hirsch-Hoffmann, M., Hennig, L., and Gruissem, W. (2004). GENEVESTIGATOR. *Arabidopsis* microarray database and analysis toolbox. *Plant Physiol.* 136, 2621–2632.

**Conflict of Interest Statement:** The authors declare 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: 25 July 2012; paper pending published: 17 August 2012; accepted: 20 October 2012; published online: 08 November 2012.

Citation: Tohge T and Fernie AR (2012) Co-expression and co-responses: within and beyond transcription. *Front. Plant Sci.* 3:248. doi: 10.3389/fpls.2012.00248 This article was submitted to *Frontiers in Plant Systems Biology, a specialty of Frontiers in Plant Science*.

Copyright © 2012 Tohge and Fernie. 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.