

AtGGM2014, an *Arabidopsis* gene co-expression network for functional studies

MA ShiSong^{1*}, BOHNERT Hans J² & DINESH-KUMAR Savithramma P^{1*}

¹Department of Plant Biology and the Genome Center, College of Biological Sciences, University of California, Davis, CA 95616, USA;

²Departments of Plant Biology and of Crop Sciences, University of Illinois at Urbana-Champaign, Urbana, IL 61801, USA

Received November 10, 2014; accepted December 25, 2014; published online February 9, 2015

Gene co-expression networks provide an important tool for systems biology studies. Using microarray data from the ArrayExpress database, we constructed an *Arabidopsis* gene co-expression network, termed AtGGM2014, based on the graphical Gaussian model, which contains 102,644 co-expression gene pairs among 18,068 genes. The network was grouped into 622 gene co-expression modules. These modules function in diverse house-keeping, cell cycle, development, hormone response, metabolism, and stress response pathways. We developed a tool to facilitate easy visualization of the expression patterns of these modules either in a tissue context or their regulation under different treatment conditions. The results indicate that at least six modules with tissue-specific expression pattern failed to record modular regulation under various stress conditions. This discrepancy could be best explained by the fact that experiments to study plant stress responses focused mainly on leaves and less on roots, and thus failed to recover specific regulation pattern in other tissues. Overall, the modular structures revealed by our network provide extensive information to generate testable hypotheses about diverse plant signaling pathways. AtGGM2014 offers a constructive tool for plant systems biology studies.

***Arabidopsis*, gene co-expression network, graphical Gaussian model, plant development, stress response, hormone response**

Citation: Ma SS, Bohnert HJ, Dinesh-Kumar SP. AtGGM2014, an *Arabidopsis* gene co-expression network for functional studies. *Sci China Life Sci*, 2015, 58: 276–286, doi: 10.1007/s11427-015-4803-x

Gene co-expression networks provide important tools for systems biology studies. Such networks, typically built for a species, connect gene pairs with similar expression patterns, as inferred from a sufficiently large collection of transcriptome data. Following a rule of “guilt by association”, genes with similar expression patterns are thought to typically function in a similar way or function in the same pathway. These networks can then be used to study transcriptome landscapes or the co-expression module organization of an organism that can inform on phenotypic distinctiveness [1–4].

A variety of approaches have been used to construct co-expression networks, employing different ways to meas-

ure the similarity of expression between genes. Most common is the relevance network that employs Pearson correlation coefficients (r) or mutual information for similarity measurement [5,6]. Also used is the Bayesian network model to infer dependency structures among genes [7]. Yet another approach is the graphical Gaussian model (GGM), which uses partial correlation to infer gene pairs with co-expression patterns [8,9]. This partial correlation is defined as the correlation that exists between two genes after removing the effects from other genes. Partial correlation surpasses Pearson correlation in gene network inference [9,10].

Several gene networks have been described for the model plant *Arabidopsis thaliana*. Mao et al. [2] have constructed a network with 6,206 genes, revealing gene co-expression

*Corresponding author (email: spdineshkumar@ucdavis.edu; sma@ucdavis.edu)

module functions, for example, in photosynthesis, vitamin metabolism, and cell cycle regulation. Mentzen and Wurtele [1] constructed a network to infer ‘regulon’ organization in *Arabidopsis*. Both networks were derived using Pearson correlation. Previously, we had constructed a gene co-expression network based on GGM using microarray data from ~1,000 Affymetrix ATH1 microarray slides [8]. This network model revealed a number of functional modules in various stress responses and metabolic pathways.

In this paper, we describe a substantially expanded GGM network that we termed AtGGM2014. This network was generated using ~10,000 *Arabidopsis* Affymetrix ATH1 microarray data and includes 18,068 genes. Clustering resulted in 662 co-expression modules. Gene ontology (GO) analysis revealed that these modules include genes belonging to diverse cellular pathways. We present a tool that facilitates easy visualization of the expression pattern of the genes within modules in various tissues or their regulation by various treatments. The expression visualization tool provides indications about the functions contained in a module. This greatly facilitates the generation of rational hypotheses that can be tested by systems biology approaches.

1 Materials and methods

The co-expression network was constructed using *Arabidopsis* microarray datasets generated with the Affymetrix ATH1 slides that have been deposited in the ArrayExpress database [11]. The raw data files (.cel) from 915 datasets were downloaded and subjected to quality analyses using arrayQualityMetrics [12] in Bioconductor (<http://www.bioconductor.org/>). The quality was assessed with six metrics: maplet, spatial, boxplot, heatmap, nuse, and rle. Low quality arrays failing two or more metrics were removed. Based on this, 10,906 slides from 899 experiment datasets were used for further analysis. The arrays in each dataset were processed into expression values using the gcRMA algorithm [13]. The ATH1 probe sets and the associated gene annotation were obtained from the TAIR database [14]. The expression datasets were combined to a single expression matrix. After quantile normalization [15], the matrix was used for co-expression network construction.

Gene co-expression network based on GGM was constructed by following a previously described method [8]. Gene pairs with partial correlation $pcor \geq 0.05$ and Pearson correlation $r \geq 0.35$, as well as gene pairs with $pcor \leq -0.05$ and $r \leq -0.35$, were selected for network construction. The network containing 102,644 co-expression gene pairs among 18,068 genes were then analyzed and visualized with BioLayoutExpress 3D and Graphviz [16,17]. The network was clustered via the Markov Cluster Algorithm, with parameters “-I 1.5 - scheme 7” [18,19]. The resulted modules were subjected to gene ontology analysis using GOstats [20]. The public available microarray datasets from AtGen-

Express were then used to visualize the expression patterns of the genes within the modules [21–23]. The samples used are listed in Table S1 in Supporting Information. An in-house developed algorithm in R was used for expression visualization. The algorithm is available for download from our lab website (<http://dinesh-kumarlab.genomecenter.ucdavis.edu/downloads.html>).

2 Results

2.1 An expanded *Arabidopsis* GGM co-expression network

An updated *Arabidopsis* gene co-expression network was constructed via GGM following a previously described method [8,9]. The network model used publicly available microarray data from the ArrayExpress database [11]. Microarray data generated from 10,906 Affymetrix ATH1 microarray slides were used for the network construction. These slides originated from 899 experimental datasets covering aspects of *Arabidopsis* development and biotic and abiotic stress responses. The partial correlation coefficients between all gene pairs were calculated, and their distribution is shown in Figure 1A. 99% of the gene pairs have their $pcor$ values in the range of -0.02 and 0.02 . For our network construction, we chose the gene pairs with $|pcor| \geq 0.05$ (P value $\leq 2.2 \times 10^{-16}$), which represents a very stringent criterion. The Pearson’s correlation r of these gene pairs were also calculated. Gene pairs with $|r| < 0.35$ were removed, because r values close to zero indicates weak associations. Also removed were those gene pairs carrying opposite signs for $pcor$ and r . This resulted in 102,343 gene pairs with $pcor \geq 0.05$ and $r \geq 0.35$ as well as 301 gene pairs with $pcor \leq -0.05$ and $r \leq -0.35$. These were chosen for construction of the network.

The network contains 102,644 co-expression gene pairs among 18,068 genes (Table S2). Via a widely used network clustering program—Markov Cluster Algorithm (MCL) [18,19]—the network was assembled into 662 clusters, which contain between six and 275 genes in each cluster (Table S3). These clusters were then treated as co-expression modules. The top 30 modules contain 3,855 genes and each of these modules includes 98–275 genes (Figure 1B). Among all 662 modules, GO analysis [20] identified 180 subdivisions with enriched GOs (P value $\leq 1 \times 10^{-5}$, or 0.0067 after Bonferroni correction) (Table S4). In contrast, if the network were to be permuted by swapping all gene IDs randomly, only four modules would have shown such GO enrichment. Thus, our network and clustering analysis identified many modules enriched with genes sharing the same GO terms, and such enrichment did not arise from randomness. Next, 110 top-ranking modules from our network, with a P value $\leq 1 \times 10^{-6}$, were included in further analyses (Table S4). According to GO enrichment configuration they

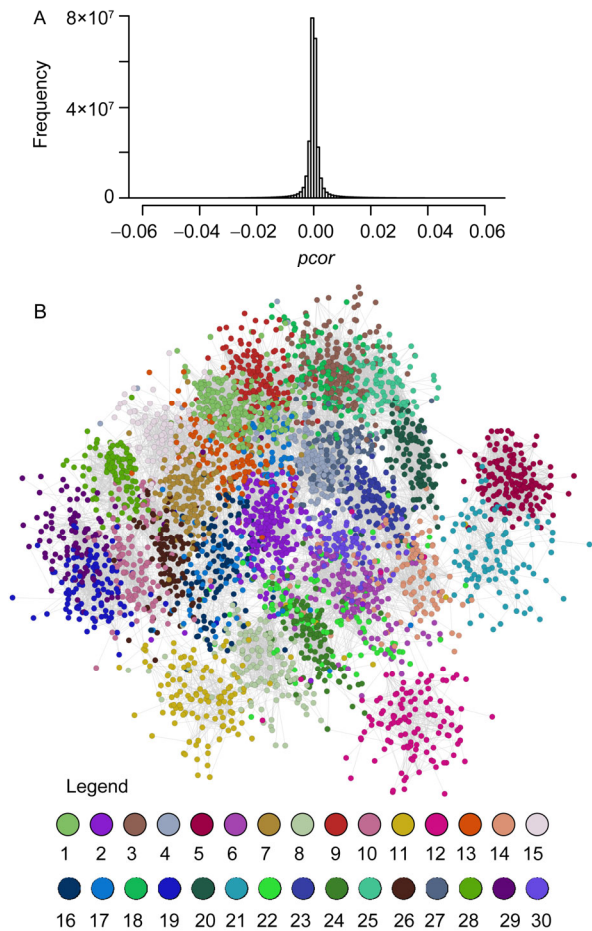


Figure 1 The overall properties of the network. A, A histogram showing the distribution of partial correlation coefficients. 99.1% of the gene pairs have their *pcor* in range between -0.02 and 0.02 . B, A sub-network representing the top 30 modules. Each dot represents a gene, and a connection between gene pairs indicates they have similar expression patterns. Shown are the genes included in the top 30 modules. Their module identities are indicated by colors. See the legend for the corresponding modules for a specific color.

can be divided into diverse categories: I, cell cycle; II, development; III, hormone; IV, house-keeping functions; V, metabolism; VI, stress response; and VII, transporter related. In addition to GO analysis, the gene expression data from the AtGeneExpress project were also used to visualize the expression patterns of genes within these modules (Table S1) [21–23]. See below for discussion on the examples of these modules.

2.2 House-keeping, cell-cycle, and transport related modules

There are five modules related to photosynthesis in the network (Table S4). Module #14 and #23 are enriched with nuclear-encoded genes that functions in photosynthesis. Module #23 includes a large number of genes encoding light harvesting proteins, while #14 comprises many genes

encoding the chloroplast NAD(P)H dehydrogenase complex. (See Table S3 for the list of genes in each module. The same applies to all modules discussed in this paper.) Genes in both modules are induced by light and repressed by stresses that affect leaves. Interestingly, genes encoding light harvesting functions (module #23) are slightly up-regulated by cold, osmotic, and salinity stresses in the roots (Figure 2). Similar up-regulation is observed for genes in module #109 and 120, both enriched for chloroplast encoded photosynthesis genes. Enriched in module #47 are genes related to plastid development. Figure 2 shows the expression patterns of selected genes from these modules in response to stresses or light treatments. The expression patterns of all genes within these module are shown in Figure S1 in Supporting Information. (Note: Heatmaps for the expression patterns for some modules discussed below are also shown as supplemental figures, which are too large to display as regular figures.)

The network also identified modules related to functions in RNA/mRNA metabolism (#46, 416), ribosome biogenesis (#21), translation (#5, 74, 164), protein folding (#69, 136), proteasome components (#56), and microtubule-based processes (#121). Intriguingly, while ribosome biogenesis genes are down-regulated by most abiotic stresses, they are up-regulated by cold in the leaves. Genes in module #69 function in ER-stress responses and they are collectively up-regulated by biotic stresses (Figure S1). These genes, such as *CNX1*, *BIP2*, *CRT1b*, and *UNE5*, are primary targets of NPR1 and required for systemic acquired resistance [24]. Similarly, genes encoding components of proteasomes are also slightly up-regulated by biotic stresses, while the biological significance of such up-regulations remains unknown.

Five modules involved in cell-cycle regulation were identified. As expected, four of these modules (#8, 45, 51, 87) show their genes expressed highly in shoot apex tissues, where cell division is most active (Figure S2). Genes in module #101, involved in chromatin functions, are highly expressed in almost all tissues.

Four identified modules are enriched with transporters or functioning in transport process such as those for iron ions (#84), water (#174), vesicle-mediated transport processes (#197), and nucleocytoplasmic (#236) transport. Also identified are two modules (#350, 360) which include genes encoding different subunits of the vacuolar ATPase complex.

2.3 Development related modules

Many modules related to functions in development were identified (Table S4). The tissue specific expression patterns of these modules are shown in Figure S3. For example, three modules are related to pollen tube growth (#1, 9, 288), module #92 is enriched with lateral root growth related genes, while module #118 regulates stomata development.

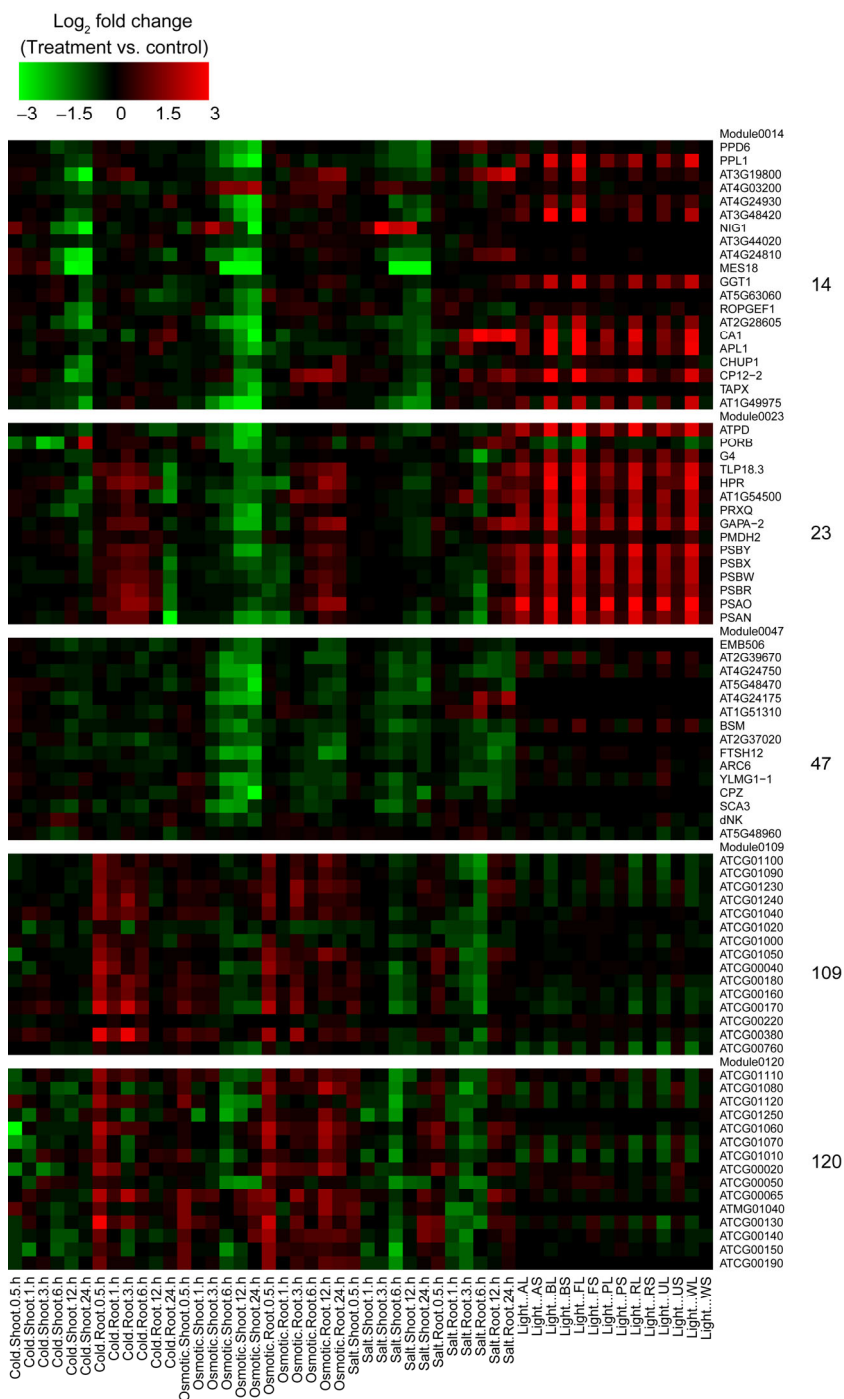


Figure 2 Regulation of selected genes in photosynthesis-related modules by various treatments. In this heatmap, each row represents a gene, and each column represents a treatment (see Table S1 for details on treatment information). The gene regulation values (log₂ treatment vs. control) are indicated by colors. Genes are grouped by modules. Next to the heatmap are module names.

See below for discussion on other modules.

Three modules are related to meristem function and general organ development. Included in module #114 are genes functioning in meristem initiation and the establishment of pattern, such as *MP*, *REV*, *PHB*, *PHV*, and *PINI* (Figure 3A) [25,26]. Module #662 contains several genes of auxin pathways important for pattern formation: *TOPELESS*, *TPR2*

and *TPR4 (WSIP2)* [27]. Module #85 is enriched with genes for multicellular organismal development. Genes in these modules are expressed in multiple tissues (Figure S3).

Three modules are related to different stages of flower development. Module #105 (Figure 3B) contains 43 genes specially expressed in the shoot apex. Their expression is reduced in early flowers (stage 9 & 11), and disappears in

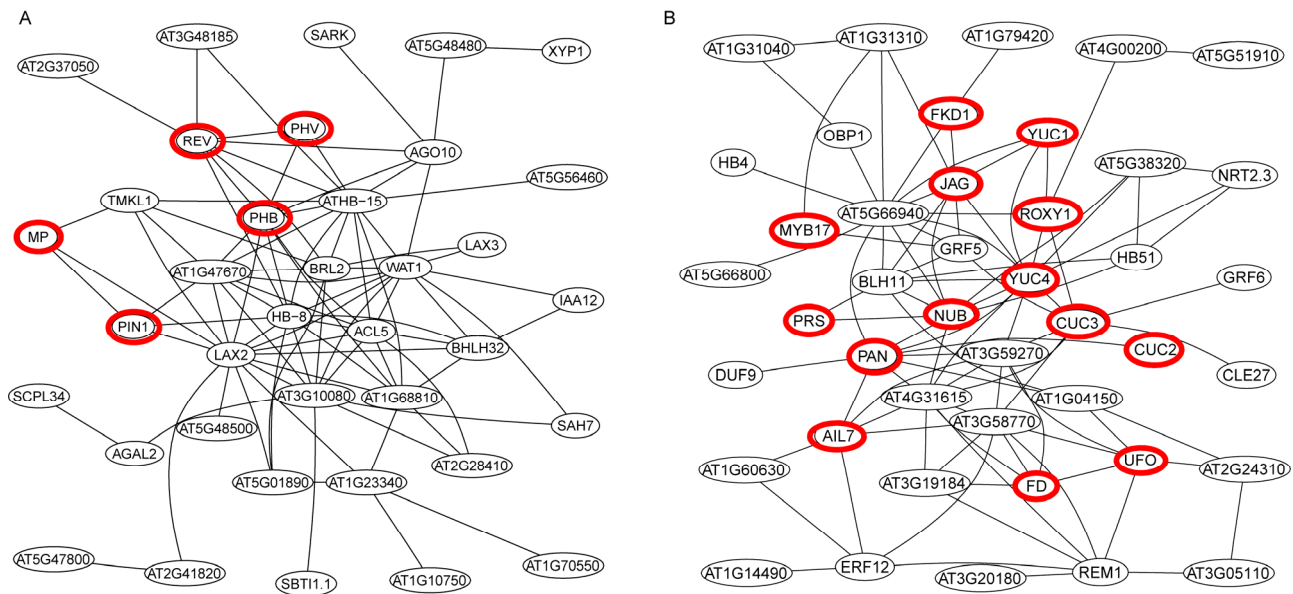


Figure 3 Sub-networks for two developmental modules. Shown are extracted sub-network for two developmental modules, # 114 (A), and #105 (B). In these networks, each node represents a gene, and a connection between genes indicates they have similar expression patterns. Highlighted in red circle are the genes discussed in the text.

mature flowers (stage 12 & 15) (Figure 4). Among these genes developmental regulators are highly enriched. Included are genes required for proper floral meristem formation, such as *PAN*, *LMI2* (*MYB17*), *UFO*, *FD*, *PRS*, genes for proper organ boundary formation, such as *JAG*, *NUB*, *ROXY1*, *CUC2*, *CUC3*, and genes for auxin biosynthesis and distribution, such as *YUC1*, *YUC4*, *AIL7*, and *FKD1* [28–30]. Module #16 comprises 111 genes expressed in both shoot apex and flowers (stage 9 to 15) (Figure 4). It includes genes functioning in vegetative to floral meristem transition, such as *LFY*, *FT*, *TSF*, *SPL3/4/5*, and floral homeotic genes that specify floral organ identities according to the ABC model, such as *API3*, *SEPI2/3*, *AG*, *PI*, and *CAL* [31,32]. In contrast, module #13 includes 116 genes expressed in late flowers (stage 12 to 15) (Figure 4). Among them are genes critical for carpel (including ovule) and stamen development, such as *INO*, *CRC*, *VERDANDI*, *MYB21/24*, *TT2/16*, *BAN*, *SHPI2*, and *NTT* [33–37]. Thus, our network identified many known genes for flower development and revealed uncharacterized genes within these three modules that might play similar or related roles especially those genes encoding transcription factors. More importantly, the genes in these modules provide the basis for future development of mechanistic models regarding how these genes interact and regulate each other to archive the functions associated with the correspondent modules. Similarly, three modules are also identified for leaf development—leaf differentiation (#246), leaf morphogenesis (#290), and leaf growth (#258). These modules include known and functionally uncharacterized genes.

Four cell wall related modules were identified with different tissue specific expression patterns (Figure S3). Mod-

ule #112, enriched with primary cell-wall biogenesis genes, is ubiquitously expressed, while module #7 genes are specifically expressed in roots. Module #33 is enriched for genes related to secondary cell wall biogenesis that are expressed in stems and roots. The genes in module #48, enriched for cell wall loosening and modification genes, are expressed in multiple tissues, which are involved in cell growth.

2.4 Hormone-centered modules

Hormones as signaling molecules govern important roles in plant signal transduction. Our network identified many gene modules involved in hormone metabolism as well as activities related to signaling. Modules enriched for functions in hormone include abscisic acid (ABA) (#2), auxin (#93, 170, 439), cytokinin (#130), ethylene (#139), gibberellin (#473), brassinosteroid (#11, 492), and jasmonic acid (JA) (#72), and salicylic acid (SA) (#19) (Table S4). It is noteworthy that the network identifies multiple modules for auxin and brassinosteroids that might be related to their involvement in spatially regulated function during different developmental stages.

Shown in Figure 5 are modules for ethylene (#139), one of three auxin (#170), and cytokinin (#130) responses. The ethylene response module contains several regulator genes in the ethylene signaling pathway, *EIN2*, *EIN3*, *ERS1*, *ETR2*, and *CTR1* [38,39]. They have similar expression patterns to three genes in the same module: *ARL*, *ARGOS*, *ORS1*, which function in regulating cell expansion and organ size [40,41]. Interestingly, *ETR2*, *ARGOS*, and *ARL* are connected to several genes in module #106 in the co-expression

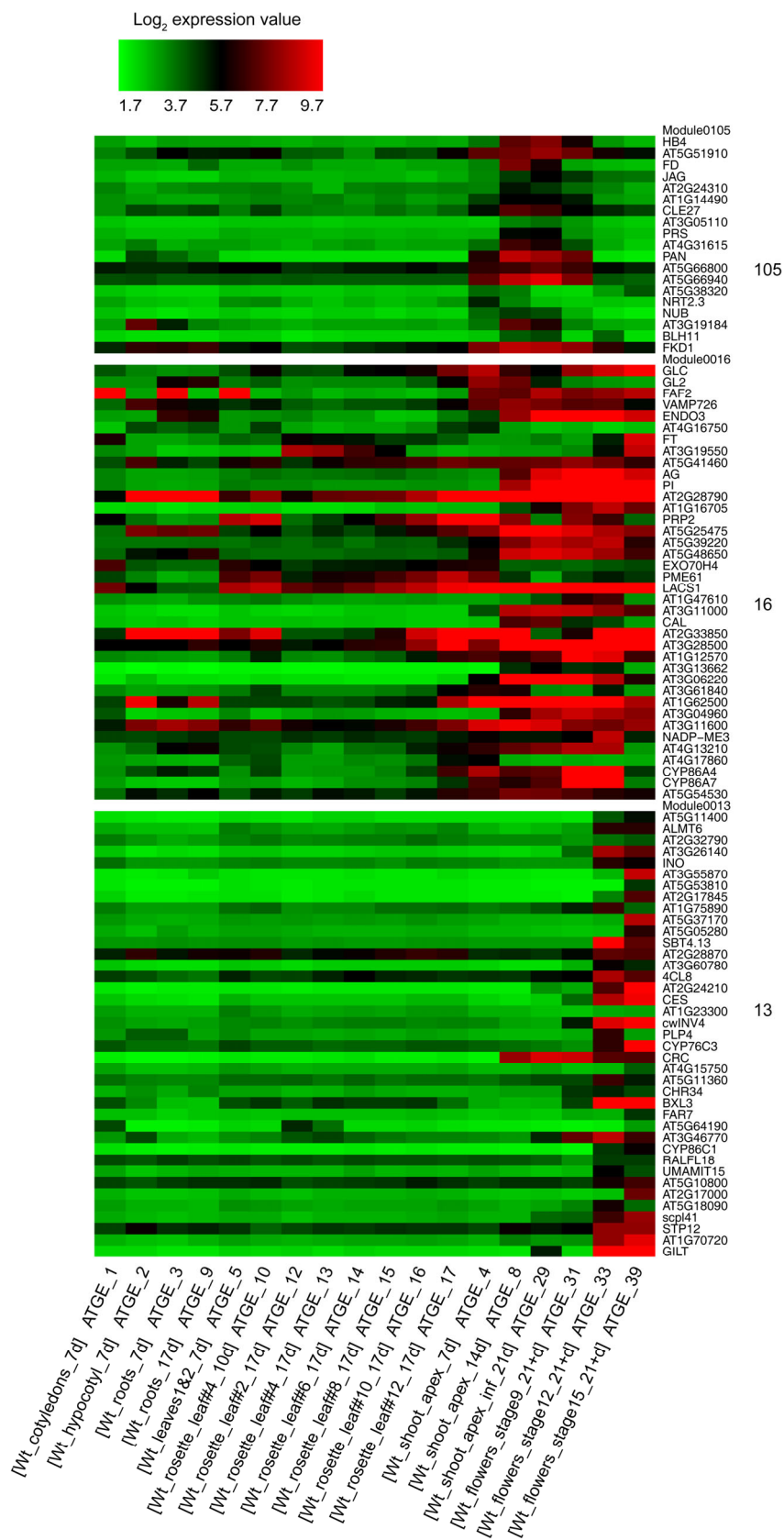


Figure 4 Tissue specific expression of selected genes in flower development modules. In this heatmap, each row represents a gene, and each column represents a tissue (see Table S1 for details on sample information). The gene expression levels are indicated by colors. Genes are grouped by modules. Next to the heatmap are the module names.

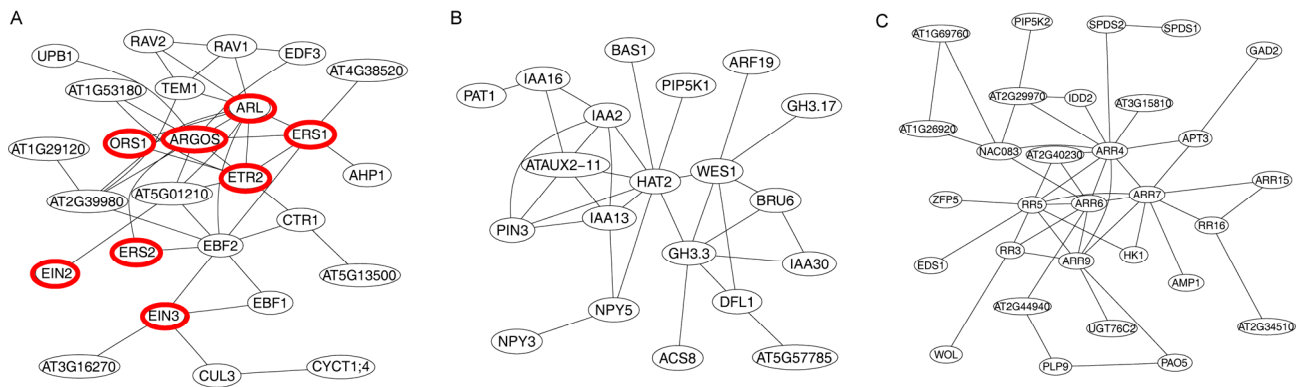


Figure 5 Sub-networks for three hormones response modules. Shown are extracted sub-networks for ethylene (A, #139), auxin (B, #170), and cytokinin (C, #130) response modules. Highlighted in red circle are the genes discussed in the text.

network (Figure S4). Genes in module #106, such as *HBI* and *ADHI*, mainly function in the hypoxia response. In plants, hypoxia stress induces ethylene production, which in turn promotes upward leaf movement (hyponastic growth) that involves cell growth and expansion [42,43]. *ARL* & *ORS1*'s connection to both ethylene signaling and hypoxia response genes indicates that these two genes might play an important role in hyponastic growth.

2.5 Modules enriched for metabolism functions

Many pathways guiding plant metabolism are subject to intense transcriptional regulation. Identified by our network are more than 20 co-expression modules related to metabolism (Table S4). They function in processes such as the generation and metabolism of starch (#6), suberin (#28), lipids (#37), sulfur (#43), hexoses (#49), fatty acids (#52, 90), amino acids (#63, 208), cytidine (#79), wax (#148), sterols (#214, 376), and tetrapyrroles (295). This also includes modules preferential to cellular respiration (#20) and photorespiration (#71).

The genes in many of these modules are repressed by either biotic or abiotic stresses such as those for lipids (#37), sulfur (#43), fatty acids (#90), amino acids (#208), sterols (#214, 376), and tetrapyrrole (#295) metabolism. In general these modules appear to include the genes for functions that are affected by stress conditions (Figure S5). However, modules including genes induced by stresses are found as well. For example, the starch metabolism module (#6) is induced by cold stress at 24 h, while the hexose metabolism module (#49) is induced by salt stress in roots and also by elicitors or bacterial treatment. The module on suberin metabolism (#28) is induced by osmotic stress in roots, and equally induced are amine catabolic processes (#63) after treatment with *Pseudomonas syringae* pv. tomato DC3000 (*Pst* DC3000) (Figure S5), constituting a biotic stress. These inductions indicate the importance of the related metabolites in fighting against the stresses.

2.6 Stress response modules

As discussed previously [8], the GGM network excels in identifying stress related gene modules. From the current network, at least 22 different modules functioning in various aspects of plant stress responses were identified (Table S4). Shown in Figure S6 are the expression patterns of 10 of these modules following treatment with different stresses or hormones. These modules are induced by abiotic and/or biotic stress treatments. Among them, module #77 and 153 are also induced by ABA, #62 by JA, and #19 by SA, indicating the roles of these hormones in regulating gene expression. Interestingly, both module #19 and 50 are induced by biotic stresses, such as bacterial elicitors (*Flagellin* 22, *hrpZ*, etc.), *Pst* DC3000 *hrcC*⁻, and *Phytophthora infestans*, and #19 is also induced by SA, highlighting their importance in the pathogen response. However, genes in #50 are transiently repressed at 6 h post infection (hpi), but this repression is absent in the *Pst* DC3000 *hrcC*⁻ infection that cannot deliver bacterial effector proteins to the host cell [44]. This indicates that effectors delivered from the bacteria might interfere in this module's gene expression characteristics. In contrast, genes in #19 are repressed in both 6 and 24 hpi upon *Pst* DC3000 infection, but such repression is not seen in *Pst* DC3000 *hrrC*⁻, indicating another set of effectors from bacteria down-regulating this module's expression. Indeed, cluster #19 contains many genes critical for immunity against pathogens, such as *MEK1*, *NPRI*, *NDRI*, and *WRKY54* [45–48].

Gene regulatory responses to stresses have gained attention because they often foreshadow phenotypic reactions that can then be related to metabolic perturbations. However, equally instructive are the genes in many modules that show little or no regulation upon stresses or hormone treatments. These modules are grouped in Figure S7, which have tissue specific expression patterns and functions, and may play roles in development (Figure S8). For example, modules #1 and #9 are specifically expressed in pollen and stamen cells and tissues, module #13 and #60 are flower-specific, and

modules #53 and #104 identify silique specific expression. These tissues were typically not sampled, or at least not enriched, in stress or hormone treatment experiments. Therefore, the lack of stress or hormone regulation of these genes might not reflect the real situation, but may be due to the lack of sampling. In other words, we know very little about how the transcriptome might be changed during stresses in these specific tissues.

2.7 Connections between modules

While gene modules function as individual units, they can also have functional connections between each other. To reveal any functional connections between modules, we constructed a higher order network by connecting module pairs that share inter-module co-expressed gene pairs [1]. Shown in Figure 6 are the connections between the ethylene signaling module #139 and other modules. Interestingly, almost all connected modules function in processes regulated by ethylene, for example trehalose (#40) and glucose (#22) metabolism, response to hypoxia (#106), chitin (#36), and biotic stresses (#10), leaf senescence (#250), and cotyledon development (#176) [39,43,49–52]. Likewise, AtGGM2014 also confirmed signaling cross-talk between ethylene and jasmonate (#72) [39,53]. Interestingly, modules #127 and #145 that are connected to module #139 contain genes involved in multiple hormone response pathways such as ABA, auxin, brassinosteroids, and ethylene (see Table S3 for gene list). It is possible that modules #127 and

#145 function as hormone signals integrators involved in deciding the exact gene expression output in specific hormone concentration environments.

3 Discussion

In this paper, we present an updated Arabidopsis gene co-expression network based on the graphical Gaussian model. The network includes 18,068 genes, much more than other Arabidopsis co-expression networks based on Pearson's correlation. The one by Mao et al. includes 6,206 genes, while the one by Mentzen and Wurtele contains 13,456 genes [1,2]. In addition, our network analyses identified top modules with fewer genes compared to those identified in the other networks. The largest module in our network comprises 270 genes, while the two Pearson based network models contain 1,381 and 1,629 genes, respectively, in their largest modules. Thus, the AtGGM2014 network distinguishes refined modular structures and seemingly discriminates conditions more sensitively. For example, our network identified five photosynthesis related modules, each with slightly different expression patterns, while in the other two networks only one photosynthesis-related cluster or regulon could be identified. The current network also shows a vastly broadened coverage when compared with our previous GGM network that included 6,760 genes [8]. This may be due to the higher number of microarray datasets that could be included into the current network analy-

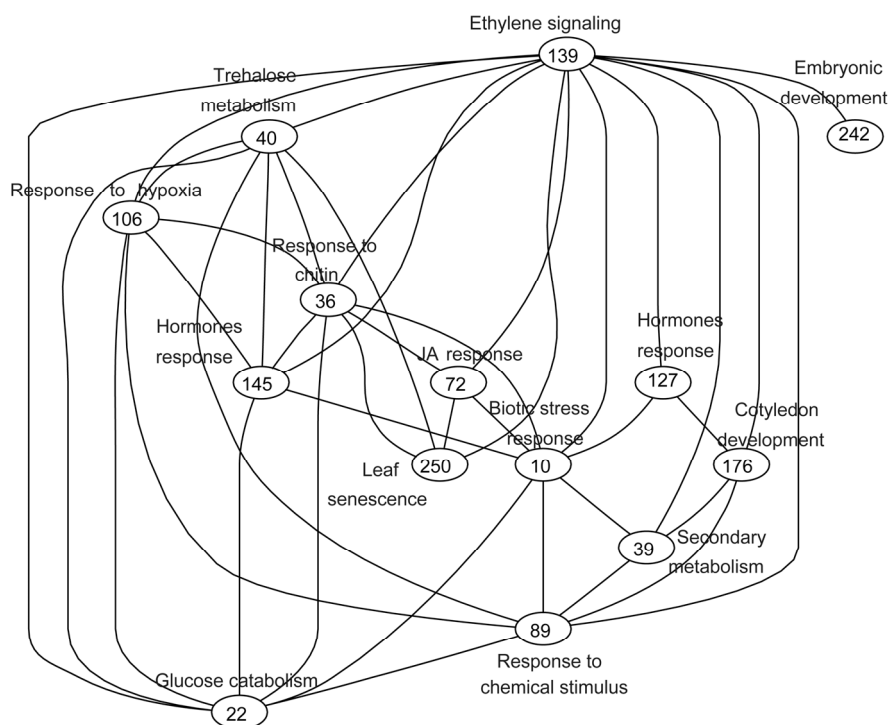


Figure 6 Connections between modules revealing higher-order networks. In this network, co-expression modules are represented by nodes. If two modules have three or more inter-module co-expressed gene pairs, they were connected by a line. Within the nodes are the module numbers, while the labels next to the nodes indicate the module's enriched function. Shown here are modules connected with module #139.

sis. It could be also because a different *pcor* cutoff value was used here (0.1 vs. 0.05). However, even with a $|pcor|$ set at 0.05, the selection is quite stringent (P value $\leq 2.2 \times 10^{-16}$). Owing to the broadened coverage, more relevant genes and co-expressed gene pairs could be recovered for the co-expression modules. This is demonstrated by comparing the ethylene signaling module (#139) of the current network to the equivalent module obtained in the previous network (Figure S9).

Using MCL clustering, we were able to divide the co-expression networks into 662 co-expression modules. As exemplified by their enriched GOs (Table S4), these modules cover many aspects of plant signaling pathways, such as those in relation to house-keeping, development, metabolism, and stress responses. Importantly, the genes within a module have similar expression patterns as they either have specific expression in certain tissues or respond to hormones or stress treatments similarly. The publicly available Arabidopsis gene expression data from AtGenExpress can be used to visualize the gene expression patterns for the genes within these modules. Such visualization can reveal minor but common regulation patterns within a group of genes in the modules. One example is the up-regulation of the proteasome-related genes in module #56 by pathogen treatments (Figure S1), which could bear biological significance. Such regulation will be missed if the analysis were to be carried out on an individual gene level but not on a co-expression module level. Thus, the coupling co-expression module identification, GO analysis, and expression visualization can facilitate thought-provoking and ultimately testable hypothesis generation.

By studying the expression patterns of modules following a variety of treatments, many co-expression modules regulated by different stresses were identified. Using a resulting expression profile in various tissues without stresses, we recognized that genes in these modules were united by that they have expression in leaves and/or roots. In contrast, many modules including genes with specific expression in other tissues did not show any stress regulated patterns, which may possibly be due to lack of sampling during stresses experiments. Therefore, only very limited information is known for the complexity of stress-regulated transcriptome responses in other tissues, while such information could provide critical information for future sustainable agriculture development. For example, how abiotic or biotic stresses affect growing siliques will be important for the development of higher yielding crops under adverse conditions. It remains to be tested if the same stress response mechanisms identified in leaf and root tissues extend to other organs as well.

Overall, we have constructed an Arabidopsis gene co-expression network using the graphical Gaussian model. The network was then clustered to identify gene co-expression modules. These modules group together genes with similar functions in time, tissue or, to some degree,

cells. Often these clusters include genes that belong to the same pathways. However, it must be stated that gene co-expression networks at this time in their development do not represent mechanistic models for regulatory gene expression, because co-expression between two genes does not necessarily translate into regulatory interaction between them. Instead, the genes included in a module are the probable components of a mechanistic or regulatory model. Additionally or alternatively, the gene co-expression network may also be used to identify the cis-regulatory elements and transcription factors that regulate the expression of genes in the modules within the network [54]. It can be expected that in the future it will become possible to incorporate into the network other types of interaction data, such as protein-protein interaction, protein-DNA interaction, or kinase-substrate interactions, to assemble a truly comprehensive regulatory network [55–58].

Our AtGGM2014 network presented here can be used in two aspects. First, it identifies functional co-expression modules in their complexity. As illustrated, such modules group known regulators and uncharacterized genes into a number of precisely circumscribed cellular pathways. This allows for the identification of candidate genes for future functional studies. Second, AtGGM2014 reveals functional connections between specifically characterized modules. As demonstrated by one of the examples, the core ethylene signaling modules can be linked to and used in diverse signaling pathways.

Finally, the AtGGM2014 program will be freely available for download from our lab's website (<http://dineshkumarlab.genomecenter.ucdavis.edu/downloads.html>). We believe it will be helpful for future studies in plant systems biology.

This work was supported by US National Science Foundation grants DBI-0723722 and DBI-1042344 to SPDK, and UC Davis funds to SPDK.

- 1 Mentzen WI, Wurtele ES. Regulon organization of *Arabidopsis*. *BMC Plant Biol*, 2008, 8: 99
- 2 Mao L, Van Hemert JL, Dash S, Dickerson JA. *Arabidopsis* gene co-expression network and its functional modules. *BMC Bioinformatics*, 2009, 10: 346
- 3 Zhang W, Morris QD, Chang R, Shai O, Bakowski MA, Mitsakakis N, Mohammad N, Robinson MD, Zirngibl R, Somogyi E, Laurin N, Eftekharpour E, Sat E, Grigull J, Pan Q, Peng WT, Krogan N, Greenblatt J, Fehlings M, van der Kooy D, Aubin J, Bruneau BG, Rossant J, Blencowe BJ, Frey BJ, Hughes TR. The functional landscape of mouse gene expression. *J Biol*, 2004, 3: 21
- 4 Ghazalpour A, Doss S, Zhang B, Wang S, Plaisier C, Castellanos R, Brozell A, Schadt EE, Drake TA, Lusis AJ, Horvath S. Integrating genetic and network analysis to characterize genes related to mouse weight. *PLoS Genet*, 2006, 2: e130
- 5 Song L, Langfelder P, Horvath S. Comparison of co-expression measures: mutual information, correlation, and model based indices. *BMC Bioinformatics*, 2012, 13: 328
- 6 Zhang B, Horvath S. A general framework for weighted gene co-expression network analysis. *Stat Appl Genet Mol Biol*, 2005, 4: Article17
- 7 Friedman N, Linial M, Nachman I, Pe'er D. Using Bayesian net-

- works to analyze expression data. *J Comput Biol*, 2000, 7: 601–620
- 8 Ma S, Gong Q, Bohnert HJ. An *Arabidopsis* gene network based on the graphical Gaussian model. *Genome Res*, 2007, 17: 1614–1625
 - 9 Schafer J, Strimmer K. A shrinkage approach to large-scale covariance matrix estimation and implications for functional genomics. *Stat Appl Genet Mol Biol*, 2005, 4: Article32
 - 10 Wille A, Zimmermann P, Vranova E, Furholz A, Laule O, Bleuler S, Hennig L, Prelic A, von Rohr P, Thiele L, Zitzler E, Gruissem W, Buhlmann P. Sparse graphical Gaussian modeling of the isoprenoid gene network in *Arabidopsis thaliana*. *Genome Biol*, 2004, 5: R92
 - 11 Rustici G, Kolesnikov N, Brandizi M, Burdett T, Dylag M, Emam I, Farne A, Hastings E, Ison J, Keays M, Kurbatova N, Malone J, Mani R, Mupo A, Pedro Pereira R, Pilicheva E, Rung J, Sharma A, Tang YA, Ternent T, Tikhonov A, Welter D, Williams E, Brazma A, Parkinson H, Sarkans U. ArrayExpress update—trends in database growth and links to data analysis tools. *Nucleic Acids Res*, 2013, 41: D987–990
 - 12 Kauffmann A, Gentleman R, Huber W. arrayQualityMetrics—a bioconductor package for quality assessment of microarray data. *Bioinformatics*, 2009, 25: 415–416
 - 13 Wu ZJ, Irizarry RA, Gentleman R, Martinez-Murillo F, Spencer F. A model-based background adjustment for oligonucleotide expression arrays. *J Am Stat Assoc*, 2004, 99: 909–917
 - 14 Lamesch P, Berardini TZ, Li D, Swarbreck D, Wilks C, Sasidharan R, Muller R, Dreher K, Alexander DL, Garcia-Hernandez M, Karthikeyan AS, Lee CH, Nelson WD, Ploetz L, Singh S, Wensel A, Huala E. The *Arabidopsis* Information Resource (TAIR): improved gene annotation and new tools. *Nucleic Acids Res*, 2012, 40: D1202–1210
 - 15 Bolstad BM, Irizarry RA, Astrand M, Speed TP. A comparison of normalization methods for high density oligonucleotide array data based on variance and bias. *Bioinformatics*, 2003, 19: 185–193
 - 16 Theocharidis A, van Dongen S, Enright AJ, Freeman TC. Network visualization and analysis of gene expression data using BioLayout Express(3D). *Nat Protoc*, 2009, 4: 1535–1550
 - 17 Gansner ER, North SC. An open graph visualization system and its applications to software engineering. *Softw Pract Exp*, 2000, 30: 1203–1233
 - 18 Enright AJ, Van Dongen S, Ouzounis CA. An efficient algorithm for large-scale detection of protein families. *Nucleic Acids Res*, 2002, 30: 1575–1584
 - 19 van Dongen S. Graph clustering by flow simulation. Dissertation for Doctoral Degree. Utrecht: Utrecht University, 2000
 - 20 Falcon S, Gentleman R. Using GOSTats to test gene lists for GO term association. *Bioinformatics*, 2007, 23: 257–258
 - 21 Goda H, Sasaki E, Akiyama K, Maruyama-Nakashita A, Nakabayashi K, Li W, Ogawa M, Yamauchi Y, Preston J, Aoki K, Kiba T, Takatsuto S, Fujioka S, Asami T, Nakano T, Kato H, Mizuno T, Sakakibara H, Yamaguchi S, Nambara E, Kamiya Y, Takahashi H, Hirai MY, Sakurai T, Shinozaki K, Saito K, Yoshida S, Shimada Y. The AtGenExpress hormone and chemical treatment data set: experimental design, data evaluation, model data analysis and data access. *Plant J*, 2008, 55: 526–542
 - 22 Kilian J, Whitehead D, Horak J, Wanke D, Weinl S, Batistic O, D'Angelo C, Bornberg-Bauer E, Kudla J, Harter K. The AtGenExpress global stress expression data set: protocols, evaluation and model data analysis of UV-B light, drought and cold stress responses. *Plant J*, 2007, 50: 347–363
 - 23 Schmid M, Davison TS, Henz SR, Pape UJ, Demar M, Vingron M, Scholkopf B, Weigel D, Lohmann JU. A gene expression map of *Arabidopsis thaliana* development. *Nat Genet*, 2005, 37: 501–506
 - 24 Wang D, Weaver ND, Kesarwani M, Dong X. Induction of protein secretory pathway is required for systemic acquired resistance. *Science*, 2005, 308: 1036–1040
 - 25 Benkova E, Michniewicz M, Sauer M, Teichmann T, Seifertova D, Jurgens G, Friml J. Local, efflux-dependent auxin gradients as a common module for plant organ formation. *Cell*, 2003, 115: 591–602
 - 26 McConnell JR, Emery J, Eshed Y, Bao N, Bowman J, Barton MK. Role of PHABULOSA and PHAVOLUTA in determining radial patterning in shoots. *Nature*, 2001, 411: 709–713
 - 27 Long JA, Ohno C, Smith ZR, Meyerowitz EM. TOPLESS regulates apical embryonic fate in *Arabidopsis*. *Science*, 2006, 312: 1520–1523
 - 28 Cheng Y, Dai X, Zhao Y. Auxin synthesized by the YUCCA flavin monooxygenases is essential for embryogenesis and leaf formation in *Arabidopsis*. *Plant Cell*, 2007, 19: 2430–2439
 - 29 Hou H, Erickson J, Meservy J, Schultz EA. FORKED1 encodes a PH domain protein that is required for PIN1 localization in developing leaf veins. *Plant J*, 2010, 63: 960–973
 - 30 O'Maoileidigh DS, Graciet E, Wellmer F. Gene networks controlling *Arabidopsis thaliana* flower development. *New Phytol*, 2014, 201: 16–30
 - 31 Bowman JL, Smyth DR, Meyerowitz EM. The ABC model of flower development: then and now. *Development*, 2012, 139: 4095–4098
 - 32 Wagner D, Meyerowitz EM. Switching on Flowers: transient LEAFY induction reveals novel aspects of the regulation of reproductive development in *Arabidopsis*. *Front Plant Sci*, 2011, 2: 60
 - 33 Crawford BC, Ditta G, Yanofsky MF. The NTT gene is required for transmitting-tract development in carpels of *Arabidopsis thaliana*. *Curr Biol*, 2007, 17: 1101–1108
 - 34 Baudry A, Heim MA, Dubreucq B, Caboche M, Weissshaar B, Lepiniec L. TT2, TT8, and TTG1 synergistically specify the expression of BANYULS and proanthocyanidin biosynthesis in *Arabidopsis thaliana*. *Plant J*, 2004, 39: 366–380
 - 35 Reeves PH, Ellis CM, Ploense SE, Wu MF, Yadav V, Tholl D, Chetelat A, Haupt I, Kennerley BJ, Hodgens C, Farmer EE, Nagpal P, Reed JW. A regulatory network for coordinated flower maturation. *PLoS Genet*, 2012, 8: e1002506
 - 36 Matias-Hernandez L, Battaglia R, Galbiati F, Rubes M, Eichenberger C, Grossniklaus U, Kater MM, Colombo L. VERDANDI is a direct target of the MADS domain ovule identity complex and affects embryo sac differentiation in *Arabidopsis*. *Plant Cell*, 2010, 22: 1702–1715
 - 37 Cucinotta M, Colombo L, Roig-Villanova I. Ovule development, a new model for lateral organ formation. *Front Plant Sci*, 2014, 5: 117
 - 38 Guo H, Ecker JR. The ethylene signaling pathway: new insights. *Curr Opin Plant Biol*, 2004, 7: 40–49
 - 39 Zhao Q, Guo HW. Paradigms and paradox in the ethylene signaling pathway and interaction network. *Mol Plant*, 2011, 4: 626–634
 - 40 Hu Y, Poh HM, Chua NH. The *Arabidopsis* ARGOS-LIKE gene regulates cell expansion during organ growth. *Plant J*, 2006, 47: 1–9
 - 41 Hu Y, Xie Q, Chua NH. The *Arabidopsis* auxin-inducible gene ARGOS controls lateral organ size. *Plant Cell*, 2003, 15: 1951–1961
 - 42 Millenaar FF, Cox MC, van Berkel YE, Welschen RA, Pierik R, Voeseek LA, Peeters AJ. Ethylene-induced differential growth of petioles in *Arabidopsis*. Analyzing natural variation, response kinetics, and regulation. *Plant Physiol*, 2005, 137: 998–1008
 - 43 Hebelstrup KH, van Zanten M, Mandon J, Voeseek LA, Harren FJ, Cristescu SM, Moller IM, Mur LA. Haemoglobin modulates NO emission and hyponasty under hypoxia-related stress in *Arabidopsis thaliana*. *J Exp Bot*, 2012, 63: 5581–5591
 - 44 Roine E, Wei W, Yuan J, Nurmiho-Lassila EL, Kalkkinen N, Romantschuk M, He SY. Hrp pilus: an hrp-dependent bacterial surface appendage produced by *Pseudomonas syringae* pv. tomato DC3000. *Proc Natl Acad Sci USA*, 1997, 94: 3459–3464
 - 45 Liu Y, Schiff M, Dinesh-Kumar SP. Involvement of MEK1 MAPKK, NTF6 MAPK, WRKY/MYB transcription factors, COI1 and CTR1 in N-mediated resistance to tobacco mosaic virus. *Plant J*, 2004, 38: 800–809
 - 46 Wang D, Amornsiripanitch N, Dong X. A genomic approach to identify regulatory nodes in the transcriptional network of systemic acquired resistance in plants. *PLoS Pathog*, 2006, 2: e123
 - 47 Century KS, Shapiro AD, Repetti PP, Dahlbeck D, Holub E, Staskawicz BJ. NDR1, a pathogen-induced component required for *Arabidopsis* disease resistance. *Science*, 1997, 278: 1963–1965
 - 48 Durrant WE, Dong X. Systemic acquired resistance. *Annu Rev Phytopathol*, 2004, 42: 185–209
 - 49 Boutilier K, Offringa R, Sharma VK, Kieft H, Ouellet T, Zhang L, Hattori J, Liu CM, van Lammeren AA, Miki BL, Custers JB, van Lookeren Campagne MM. Ectopic expression of BABY BOOM triggers a conversion from vegetative to embryonic growth. *Plant*

- Cell, 2002, 14: 1737–1749
- 50 Millet YA, Danna CH, Clay NK, Songnuan W, Simon MD, Werck-Reichhart D, Ausubel FM. Innate immune responses activated in *Arabidopsis* roots by microbe-associated molecular patterns. *Plant Cell*, 2010, 22: 973–990
- 51 Zhou L, Jang JC, Jones TL, Sheen J. Glucose and ethylene signal transduction crosstalk revealed by an *Arabidopsis* glucose-insensitive mutant. *Proc Natl Acad Sci USA*, 1998, 95: 10294–10299
- 52 Gan S, Amasino RM. Making sense of senescence (molecular genetic regulation and manipulation of leaf senescence). *Plant Physiol*, 1997, 113: 313–319
- 53 Bari R, Jones JD. Role of plant hormones in plant defence responses. *Plant Mol Biol*, 2009, 69: 473–488
- 54 Ma S, Shah S, Bohnert HJ, Snyder M, Dinesh-Kumar SP. Incorporating motif analysis into gene co-expression networks reveals novel modular expression pattern and new signaling pathways. *PLoS Genet*, 2013, 9: e1003840
- 55 Popescu SC, Popescu GV, Bachan S, Zhang Z, Gerstein M, Snyder M, Dinesh-Kumar SP. MAPK target networks in *Arabidopsis thaliana* revealed using functional protein microarrays. *Genes Dev*, 2009, 23: 80–92
- 56 Chang KN, Zhong S, Weirauch MT, Hon G, Pelizzola M, Li H, Huang SS, Schmitz RJ, Urich MA, Kuo D, Nery JR, Qiao H, Yang A, Jamali A, Chen H, Ideker T, Ren B, Bar-Joseph Z, Hughes TR, Ecker JR. Temporal transcriptional response to ethylene gas drives growth hormone cross-regulation in *Arabidopsis*. *Elife*, 2013, 2: e00675
- 57 Lee I, Ambaru B, Thakkar P, Marcotte EM, Rhee SY. Rational association of genes with traits using a genome-scale gene network for *Arabidopsis thaliana*. *Nat Biotechnol*, 2010, 28: 149–156
- 58 Arabidopsis Interactome Mapping Consortium. Evidence for network evolution in an *Arabidopsis* interactome map. *Science*, 2011, 333: 601–607

Open Access This article is distributed under the terms of the Creative Commons Attribution License which permits any use, distribution, and reproduction in any medium, provided the original author(s) and source are credited.

Supporting Information

Figure S1 Regulation of genes in selected house-keeping modules by various treatments and conditions. In this heatmap, each row represents a gene, and each column represents a condition. The gene regulation values (\log_2 treatment vs. control) are indicated by colors. Genes are grouped by modules. The module # can be found on top of the list of gene names for each module. This description also applies for Figures S5–S7.

Figure S2 Tissue specific expression of genes in cell-cycle related modules. In this heatmap, each row represents a gene, and each column represents a tissue. The gene expression levels are indicated by colors. Genes are grouped by modules. The module # can be found on top of the list of gene names for each module. This description also applies for Figures S3 and S8.

Figure S3 Tissue specific expression of genes in modules related to development.

Figure S4 The connection between the ethylene signaling module (#139) and the hypoxia response module (#106). Shown are the sub-network for genes from module #139 and #106. Each node represents a gene, while connections between genes indicate shared expression pattern.

Figure S5 Regulation of genes in selected metabolism modules in response to various treatments and conditions.

Figure S6 Regulation of genes in selected stress-related modules by various treatments and conditions.

Figure S7 Regulation of genes in selected modules in response to diverse treatments and conditions.

Figure S8 Tissue specific expression characteristics of genes in selected modules.

Figure S9 Comparison between the current GGM network (A) and the previously published GGM network (B). Shown are the sub-networks for the ethylene signaling modules as they appear in both networks.

Table S1 The treatment conditions and tissue samples information for the data used in the expression visualization

Table S2 The 102,644 co-expression gene pairs contained in the co-expression gene network

Table S3 The genes used for the co-expression network construction, their annotation, and their module identities within the network

Table S4 Gene Ontology analysis results for the modules identified in the network

The supporting information is available online at life.scichina.com and link.springer.com. The supporting materials are published as submitted, without typesetting or editing. The responsibility for scientific accuracy and content remains entirely with the authors.