Planta (2019) 249:1487–1501
https://doi.org/10.1007/s00425-019-03102-9

ORIGINAL ARTICLE

Gene co-expression network analysis identifies trait-related modules in *Arabidopsis thaliana*

Wei Liu[1] · Liping Lin[1] · Zhiyuan Zhang[1] · Siqi Liu[1] · Kuan Gao[1] · Yanbin Lv[1] · Huan Tao[1] · Huaqin He[1]

Received: 11 November 2018 / Accepted: 28 January 2019 / Published online: 30 January 2019

Abstract

Main conclusion **A comprehensive network of the *Arabidopsis* transcriptome was analyzed and may serve as a valuable resource for candidate gene function investigations. A web tool to explore module information was also provided.**

Arabidopsis thaliana is a widely studied model plant whose transcriptome has been substantially profiled in various tissues, development stages and other conditions. These data can be reused for research on gene function through a systematic analysis of gene co-expression relationships. We collected microarray data from National Center for Biotechnology Information Gene Expression Omnibus, identified modules of co-expressed genes and annotated module functions. These modules were associated with experiments/traits, which provided potential signature modules for phenotypes. Novel heat shock proteins were implicated according to guilt by association. A higher-order module networks analysis suggested that the *Arabidopsis* network can be further organized into 15 meta-modules and that a chloroplast meta-module has a distinct gene expression pattern from the other 14 meta-modules. A comparison with the rice transcriptome revealed preserved modules and KEGG pathways. All the module gene information was available from an online tool at http://bioinformatics.fafu.edu.cn/arabi/. Our findings provide a new source for future gene discovery in *Arabidopsis*.

Keywords Rice · Conservation · Hub gene · Transcriptome

Abbreviations

GCN	Gene co-expression network
WGCNA	Weighted gene co-expression network analysis
NCBI	National Centre for Biotechnology Information
GEO	Gene Expression Omnibus
RSD	Relative standard deviation
GO	Gene ontology
KEGG	Kyoto Encyclopedia of Genes and Genomes

Introduction

Microarray-based high-throughput technology has been extensively applied to profile genome-wide gene expression under diverse conditions (Aoki et al. 2007). Using advanced bioinformatics tools, researchers can find candidate genes for a specific phenotype, infer gene functions and regulations (Usadel et al. 2009; Li et al. 2015; Ficklin and Feltus 2011), and perform comparative co-expression analysis (Movahedi et al. 2012; Ruprecht et al. 2017). As a robust system, living beings could response to biotic and abiotic stresses (Amrine et al. 2015; Nishiyama et al. 2018). Complex life activity relies not only on individual genes but also on a dynamic and complex gene network. To identify the gene network, a large-scale analysis of the transcriptome data can be performed.

The gene co-expression network (GCN) method has been used to explore global, temporal and spatial expression of *Arabidopsis* (Schmid et al. 2005). For example, Mao constructed *Arabidopsis* GCN using 1094 arrays from AtGenExpress and functionally annotated 46 modules of the 382 identified modules (Mao et al. 2009). Furthermore, Mutwil

Electronic supplementary material The online version of this article (https://doi.org/10.1007/s00425-019-03102-9) contains supplementary material, which is available to authorized users.

✉ Wei Liu
weilau@fafu.edu.cn

✉ Huaqin He
hehq3@fafu.edu.cn

[1] School of Life Sciences, Fujian Agriculture and Forestry University, Fuzhou 350002, People's Republic of China

used 351 microarray data points and identified 181 gene clusters as well as 27 of the 34 significant clusters; then validated 6 genes predicted to be essential genes (Mutwil et al. 2010). Zheng used 1388 microarrays from ATTED-II to construct GCN and predicted motifs in the promoter regions of co-expressed genes (Zheng et al. 2011). Giorgi used a modification of RMA to normalize 3707 *Arabidopsis* microarrays for correlation analysis (Giorgi et al. 2010). Feltus maximized gene co-expression relationships through pre-clustering of 7105 *Arabidopsis* expression samples (Feltus et al. 2013). There are also several targeted or condition-dependent network analyses (Ficklin et al. 2017). Targeted network analyses focus on GCN of a subset of genes, while condition-dependent analyses emphasize GCN under a limited number of biotic and abiotic conditions. For example, Peng constructed GCN for organelles in *Arabidopsis* based on Gene Ontology Cellular Component information (Penga et al. 2016). Wang identified 2438 cell wall-related genes in *Arabidopsis* under 351 conditions based on GCN (Wang et al. 2012b). Boruc constructed a dynamic interaction network on core cell cycle genes through a combination of GCN and protein–protein interaction information (Boruc et al. 2010). Amrine analyzed 272 microarrays that involved microbial infections of *Arabidopsis* with a wide array of fungal and bacterial pathogens with biotrophic, hemibiotrophic, and necrotrophic lifestyles as well as constructed GCN of core biotic stress-responsive genes (Amrine et al. 2015). Prasch applied triple-stress conditions with heat, drought and virus exposure to *Arabidopsis*, and they revealed significant shifts in signaling networks by GCN (Prasch and Sonnewald 2013). Rasmussen constructed GCN in 10 *Arabidopsis* ecotypes using cold, heat, light, salt and flagellin treatment as single-stress factors as well as their combinations (Rasmussen et al. 2013). Veen used GCN to compare 8 *Arabidopsis* accessions under compound stress imposed by submergence, and they revealed a core of conserved, genotype- and organ-specific responses to flooding stress (van Veen et al. 2016). Furthermore, a major task for scientists is to transfer acquired knowledge from the model organism *Arabidopsis* to crop species. The emerging comparative GCN has become a powerful tool for cross-species analysis. The PlaNet combines sequence and comparative GCN to help in identifying homologs in valuable crop species (Mutwil et al. 2011). Ficklin used GCN to find conserved gene modules between maize and rice (Ficklin and Feltus 2011), while Shaik used GCN to identify common modules for drought and bacterial stress responses between *Arabidopsis* and rice (Shaik and Ramakrishna 2013). Finally, researchers have developed web tools for gene co-expression exploration. These tools have different features, for example, AraNet focuses on functional annotation by combing multiple data sources (Lee et al. 2015); ATTED-II and CressExpress emphasize gene–gene co-expression query (Aoki et al. 2016); and PLANEX also provides Cohen's Kappa statistics for cross-species co-expression gene comparison (Yim et al. 2013).

One of the widely used GCN methods is weighted gene co-expression network analysis (WGCNA) (Zhang and Horvath 2005). It groups genes with similar expression patterns across biological samples, which may be members of the same pathway or biological process. The whole transcriptome can be simplified to several modules, which allows us to look into bio-system components easily. The relationships between genes within modules can be delineated. The higher-order module network can also be described. These network properties can further be correlated with other biological traits to find functional genes or modules. However, the inherent gene–gene connections that exist within a transcriptome can only be detected when enough perturbations viz. biological replications are pooled.

In this study, we applied WGCNA to publicly available microarray data covering several conditions for *Arabidopsis*. Genome-scale modules of co-expressed genes with clear functional annotations were identified. The module association with traits was inferred. Five potential heat shock-responsive genes were found. A higher-order module network analysis indicated the distinct expression pattern of chloroplast genes. Module preservation analysis suggested that there was a similarity between *Arabidopsis* and rice.

Materials and methods

Microarray data acquisition and processing

Microarray datasets were obtained from the National Centre for Biotechnology Information (NCBI) Gene Expression Omnibus (GEO) database under the platform numbers GPL198 for *Arabidopsis* and GPL2025 for rice. The two platforms consist of experimental samples from assays using the Affymetrix *Arabidopsis* ATH1 Genome Array and Affymetrix Rice Genome Array (http://www.affymetrix.com). The *Arabidopsis* array contains 22,810 probesets, and the rice array contains 57,381 probesets. Briefly, 931 *Arabidopsis* datasets with 12,112 samples and 191 rice datasets with 2043 samples were analyzed. Raw gene chip data were analyzed with Expression Console (v1.4.1.46) using the MAS5 method (Pepper et al. 2007). Probe-level gene expression data were retrieved. The duplicated samples were detected by the R (v3.3.1) duplicated function. After removing duplicated and disrupted samples, 11,896 *Arabidopsis* and 2025 rice samples were found. Control probes were removed before quantile normalization in R using the normalize.quantiles function (Bolstad et al. 2003). The probesets were mapped to Entrez gene IDs according to the array annotation table provided by NCBI GEO. The genes labeled with more

than one probeset were filtered by their relative standard deviation (RSD). The probeset with the highest RSD was retained, which guaranteed the useful information. For convenience, we referred to the probeset as the corresponding gene throughout the manuscript. Finally, 21,275 genes from *Arabidopsis* and 19,449 genes from rice were included for downstream analysis. Detailed information for these datasets is provided in Supplementary Tables S1–4.

Weighted gene co-expression network analysis (WGCNA)

Network analysis was performed using the Bioconductor WGCNA package (v1.63) on a Dell PowerEdge R930 Server with the following parameters: networkType = 'signed', softPower = 10 or 14, minModuleSize = 30, deepSplit = 4 (Huber et al. 2015; Langfelder and Horvath 2008). Briefly, signed co-expression networks were constructed for *Arabidopsis* and rice separately. For each gene in the gene expression matrix, a pairwise Pearson correlation coefficient was computed, and an adjacency matrix was calculated by raising the correlation matrix to a power (Zhang and Horvath 2005). A power of 10 and 14 was chosen for *Arabidopsis* and rice, respectively, using the scale-free topology criterion. Then, the adjacency matrix was transformed into a network of topological overlap (TO), which measures not only the correlation of two genes but also the extent of their shared correlations across the weighted network (Zhang and Horvath 2005). The TO matrix was then hierarchically clustered to identify highly co-expressed genes. Finally, co-expression gene modules were identified by the Dynamic Tree Cut algorithm (Oldham et al. 2008). Each module was summarized by a module eigengene (ME) through singular value decomposition, so that each module expression profile was represented by its first principal component (Zhang and Horvath 2005). Thus, ME explains the maximum amount of variation of the module expression levels and is considered the most representative gene expression in a module. To construct the network of modules and identify modules, the same process was applied to the results discussed above. The parameters were power = 6, minModuleSize = 2. The clustering was conducted with the hclust function in the WGCNA package.

Module stability was tested as the average correlation between the original connectivity and the connectivity from half samples that were randomly sampled 1000 times. The process was run for every module. The module preservation of rice compared to *Arabidopsis* was analyzed with the WGCNA modulePreservation function using the following parameters: referenceNetworks = *Arabidopsis* and networkType = "signed", nPermutations = 100. The analysis provides quantitative statistics of module preservation, which provides a rigorous argument that a module is not preserved (Langfelder et al. 2011). Through permutations, the analysis provides a Zsummary value, which summarizes the evidence that a module is preserved and indicative of module robustness and reproducibility. The Zsummary threshold for strongly preserved modules is 10. Zsummary scores between 2 and 10 are for weak to moderately preserved modules, and Zsummary scores < 2 are for modules that are not preserved.

Functional annotation of the modules

Gene ontology (GO) enrichment for network modules was performed using the Database for Annotation, Visualization and Integrated Discovery (DAVID 6.8) (Huang et al. 2009) with the background list from the *Arabidopsis* ATH1-121501 Genome Array genome. DAVID provides not only enrichment results for GO but also information for the Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway and Pfam motif and chromosome enrichment. The overrepresentation of a term is defined as a modified Fisher's exact *P* value with an adjustment for multiple tests using the Benjamini method. For simplicity, the top significant term was recorded. Modular genes enriched within chromosome regions were analyzed with the Positional Gene Enrichment analysis tool (De Preter et al. 2008). Statistical significance was set at a P value of 3E−7. The overrepresented chromosomal regions were visualized using the Ensembl Genome browser.

For gene expression variation analysis, the gene expression relative standard deviation for each gene in a module was calculated, and the average values for each module were provided.

Comparison with rice transcriptome data

Overall, 1094 rice microarray data points from the NCBI GEO database under the platform GPL2025 were collected and processed as mentioned in "Microarray data acquisition and processing". The orthologs between *Arabidopsis* and rice were downloaded through the EnsemblPlants BioMart tool. Orthologs were subjected to module preservation analysis using the R WGCNA package (R Development Core Team 2013). KEGG pathway-based analysis and visualization were also performed in R according to the package tutorial.

Results

A gene co-expression network of *Arabidopsis* was successfully constructed

A total of 11,896 *Arabidopsis* samples were used to construct a scale-free gene co-expression network, which is a property of natural biological networks, by choosing a power

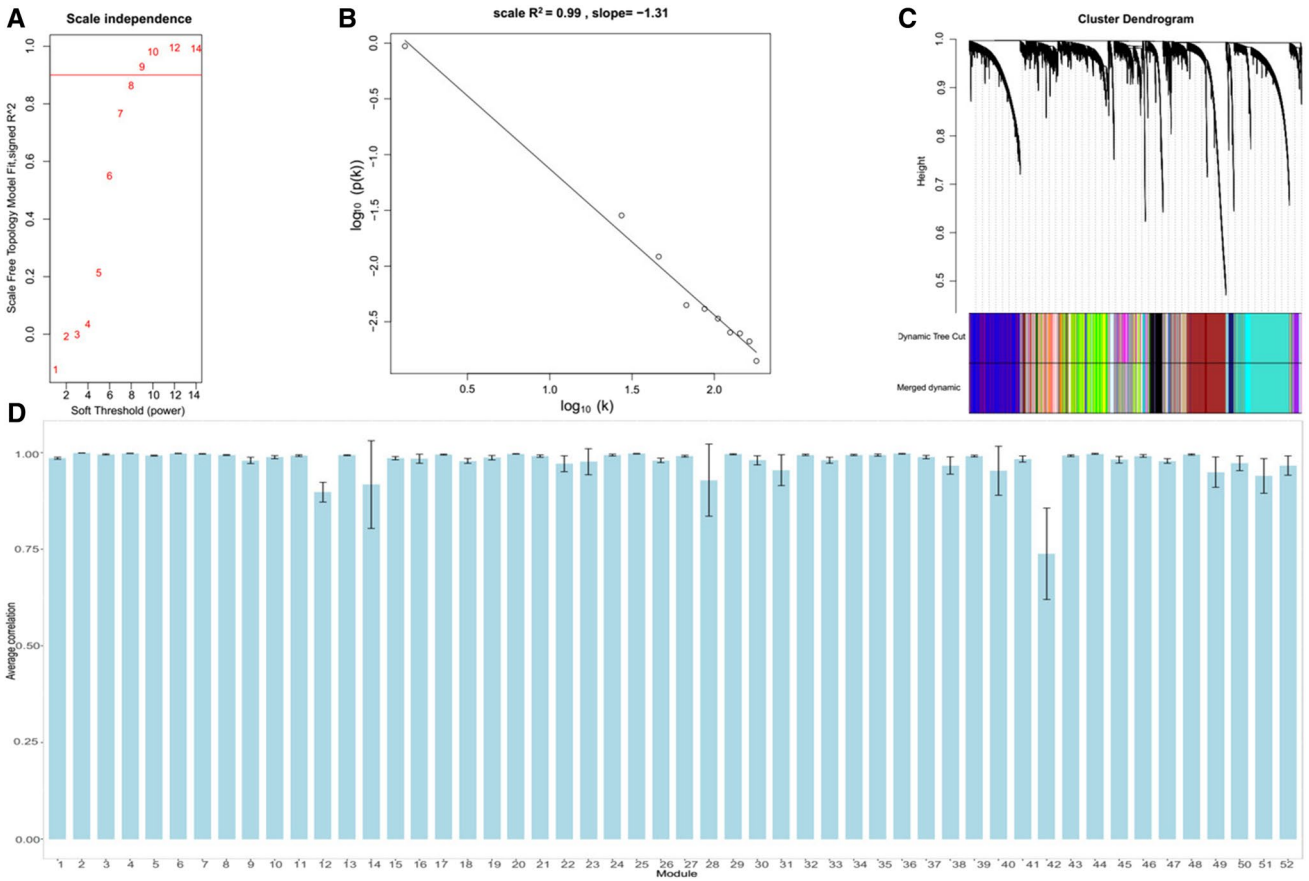

Fig. 1 A scale-free gene co-expression network for *Arabidopsis* was successfully constructed. **a** A signed R^2 against power plot shows the threshold (red line) for a scale-free network. **b** The *Arabidopsis* network obeys the power law when a power of 10 is chosen. The regression line (diagonal) shows a good model fits index $R^2 = 0.99$. **c** Dendrograms produced by average linkage of hierarchical clustering of *Arabidopsis* genes, which is based on a topological overlap matrix (TOM). The modules were assigned colors as indicated in the horizontal bar beneath the dendrogram. **d** Bar plots present the correlation of intramodule connectivity for each module by half-sampling 1000 times with the original one (mean ± SD)

of 10 (Fig. 1a, b). As described in the methods, the hierarchically clustered genes were detected iteratively by dynamic tree cut method to find stable gene clusters (Fig. 1c). Similar clusters were merged to form 52 co-expressed gene modules (Table 1). The module stability was tested by examining the correlation between the original connectivity and the values calculated from the 1000 sampled connectivity values for each module (Fig. 1d). All the modules had an average connectivity correlation larger than 0.9, except for M42 (for simplicity, the modules are presented as M plus the module number, such as M42). M42 has the lowest module stability, while M2 has the highest module stability.

Functional enrichment shows that these modules, except for M50 and M51, are associated with various biological processes (only the top one is shown) (Table 1). Most of the modules are enriched with a specific biological term, which suggests the network analysis is valid. The most significant module is M2 enriched with chloroplast genes, while M50 and M51 had no significant annotation. Some of the modules reflect the co-expression of protein complexes in biological processes, such as the M4 ribosome genes in translation and M6 chloroplast genes in photosynthesis. Modules M17, M24, M28, M34 and M37 are biotic stress responsive, that are involved in response to microorganism; while, M25 and M32 are involved in abiotic stresses, such as light and heat. Modules M15, M16, M23, M35 and M41 are associated with protective tissues, such as cell wall, phloem, Casparian strip and seed coat. Modules M3, M18 and M30 are associated with reproduction, such as pollen tube growth and cell division. These modules are also enriched with specific Pfam terms (Supplementary Table S5). For example, M34 is enriched with a Leucine-Rich Repeat (1.1E−6), and M1 is enriched with a PPR repeat family (8.8E−56).

Connectivity-based analysis

In a network, hubs usually include genes with higher connectivity, so hubs are more important (Batada et al. 2006).

Table 1 Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) enrichment of the 52 gene co-expression modules identified in *Arabidopsis*

Mod	Size	Biological process	Cellular component	Molecular function	KEGG pathway
1	2998	Regulation of transcription, DNA-templated (4E−19)	Nucleus (4E−18)	DNA binding (7E−22)	
2	2746	Photosynthesis (3E−49)	Chloroplast (5E−194)	Chlorophyll binding (9E−15)	Photosynthesis (6E−21)
3	2642	Pollen tube growth (1E−13)	Pollen tube (1E−9)	ATP binding (3E−5)	SNARE interactions in vesicular transport (5E−5)
4	1320	Translation (6E−124)	Ribosome (1E−123)	Structural constituent of ribosomes (2E−145)	Ribosome (2E−114)
5	1147	rRNA processing (5E−29)	Nucleolus (5E−67)	RNA binding (4E−53)	Ribosome biogenesis in eukaryotes (2E−27)
6	854	Chloroplast organization (4E−24)	Chloroplast (3E−151)	rRNA binding (5E−17)	Aminoacyl-tRNA biosynthesis (5E−9) Porphyrin and chlorophyll metabolism (5E−7)
7	725	Lipid storage (6E−7)	Glyoxysome (1E−2)	Nutrient reservoir activity (1E−3)	
8	684	Oxidation reduction (4E−5)	Vacuole (2E−11)	Heme binding (2E−6)	Phenylpropanoid biosynthesis (4E−7)
9	75	Oxalate metabolism (1E−2)	Extracellular region (2E−5)	Nutrient reservoir activity (5E−3)	
10	525	mRNA splicing, via spliceosome (6E−6)	Nucleus (1E−37)	RNA binding (1E−38)	Spliceosome (1E−10)
11	504	Regulation of transcription, DNA-templated (2E−13)	Nucleus (6E−12)	Sequence-specific DNA binding (5E−19)	Plant hormone signal transduction (7E−4)
12	478	SCF-dependent proteasomal ubiquitin-dependent protein catabolism			
13	401	Isocitrate metabolic process (3E−2)	Anchored component of membrane (1E−4)		Biosynthesis of antibiotics (8E−7) and secondary metabolites (6E−4)
14	376	SCF-dependent proteasomal ubiquitin-dependent protein catabolism (6E−6)	Nucleus (2E−3)	RNA-directed DNA polymerase activity (6E−6)	
15	106	Plant-type cell wall organization (3E−20)	Extracellular region (3E−6)	Structural constituent of cell wall (2E−8)	Phenylpropanoid biosynthesis (4E−7)
16	242	Plant-type secondary cell wall biogenesis (2E−24)	Golgi membrane (1E−9)	Hydroquinone: oxygen oxidoreductase activity (1E−4)	Amino sugar and nucleotide sugar metabolism (4E−2)
17	816	Response to chitin (4E−32)	Plasma membrane (4E−22)	Kinase activity (2E−14)	Plant–pathogen interaction (5E−11)
18	170	Cell division (1E−21)	Nucleus (6E−9)	Cyclin-dependent protein serine/threonine kinase regulator activity (4E−10)	DNA replication (4E−13)
19	162	Response to abscisic acid	Nucleus (2E−5)	Protein binding (2E−2)	Endocytosis (2E−2)
20	154	Glycolytic process (1E−17)	Golgi apparatus (3E−35)	Copper ion binding (5E−8)	Biosynthesis of antibiotics (3E−7)
21	144		Integral component of the plasma membrane (1E−6)	Carbohydrate transmembrane transporter activity (5E−2)	Fatty acid degradation (7E−3)
22	133	Oxidation–reduction (3E−5)	Extracellular region (3E−8)	Heme binding (7E−3)	Pentose and glucuronate interconversions (4E−2)
23	126	Phloem development (2E−4)	Anchored component of plasma membrane (1E−3)	Polysaccharide binding (6E−4)	
24	124	Defence response to fungus(2E−5)	Integral component of membrane (2E−5)	Heme binding (2E−3)	Phenylalanine, tyrosine and tryptophan biosynthesis (2E−4)

Table 1 (continued)

Mod	Size	Biological process	Cellular component	Molecular function	KEGG pathway
25	76	Leucine catabolic process (2E−4) Response to absence of light (1E−4)	Mitochondrial matrix (2E−3)		Valine, leucine and isoleucine degradation (1E−8)
26	106		Nucleus (8E−3)	DNA binding (5E−2)	Spliceosome (1E−2)
27	105	Transmembrane receptor protein tyrosine kinase signaling pathway (1E−2)	Extracellular region (3E−4)	(1- > 3)-beta-D-glucan binding (3E−2)	Plant hormone signal transduction (2E−2)
28	104	Killing of cells from other organisms (8E−5)	Extracellular region (7E−15)	Enzyme inhibitor activity (2E−2)	
29	103		Vacuole (1E−2)	SNARE binding (1E−2)	SNARE interactions in vesicular transport (1E−2)
30	100	Pollen exine formation (3E−24)	Extracellular region (7E−11)	Lipid binding (4E−8)	Phenylpropanoid biosynthesis (2E−3)
31	32				Iron/zinc purple acid phosphatase-like protein C (8E−3)
32	98	Response to heat (1E−53)	Cytoplasm (3E−6)	Unfolded protein binding (2E−3)	Protein processing in endoplasmic reticulum (6E−19)
33	92	Cell division (2E−2)		Protein binding (5E−3)	
34	84	Defence response (1E−9)	Plasma membrane (4E−8)	Kinase activity (2E−14)	Plant–pathogen interaction (5E−4)
35	83	Cell–cell junction assembly (1E−7)	Casparian strip (7E−13)	Guiding stereospecific synthesis activity (2E−4)	Phenylpropanoid biosynthesis (8E−8)
36	82	Glutathione metabolism (1E−11)	Cytosol (8E−10)	Glutathione transferase activity (7E−10)	Glutathione metabolism (2E−9)
37	78	Defence response to bacterium (3E−10)	Plasma membrane (2E−3)	Kinase activity (5E−5)	
38	76		Integral component of membrane (3E−2)	Alternative oxidase activity (2E−2)	
39	99		Nucleolus (1E−4)	RNA binding (4E−6)	Ribosome biogenesis in eukaryotes (1E−3)
40	328	SCF-dependent proteasomal ubiquitin-dependent protein catabolism (8E−6)	SCF ubiquitin ligase complex (6E−5)	Ubiquitin-protein transferase activity (2E−4)	Ubiquitin-mediated proteolysis (7E−3)
41	67	Seed coat development (6E−7)	Integral component of membrane (4E−2)		Flavonoid biosynthesis (8E-
42	59	Cellular protein catabolic process (3.5E−2)			
43	56	Suberin biosynthesis (2E−10)	Anchored component of membrane (1E−2)	Hydrolase activity, acting on ester bonds (1E−2)	Cutin, suberin and wax biosynthesis (1E−5)
44	55	Response to chitin (1E−12)		Transcription factor activity, sequence-specific DNA binding (3E−4)	Plant–pathogen interaction (5E−3)
45	54	Protein phosphorylation (4E−2)	Plasma membrane (2E−2)	ATP binding (4E−2)	
46	45	Cutin biosynthetic process (8E−6)		RNA polymerase II regulatory region sequence-specific DNA binding (9E−3)	Cutin, suberin and wax biosynthesis (1E−5)
47	45		Extracellular region (1E−2)	Oxalate decarboxylase activity (4E−2)	
48	45	Photorespiration (1E−8)	Mitochondrial respiratory chain complex I (5E−16)	Ubiquinol–cytochrome-c reductase activity (8E−9)	Oxidative phosphorylation (1E−11)
49	44	Cytidine deamination (4E−14)		Cytidine deaminase activity (7E−14)	Ubiquitin-mediated proteolysis (9E−3)
50	37	Nectar secretion			

Table 1 (continued)

Mod	Size	Biological process	Cellular component	Molecular function	KEGG pathway
51	36	Mucilage pectin metabolism			
52	36	lipid catabolic process (4E−3)	Extracellular region (1E−2)		

First, the most highly connected genes with the highest intramodule connectivity for each module are summarized in Supplementary Table S5. Second, to find highly connected genes in the whole network, all genes were sorted in descending order according to their global connectivity, which represents how densely a gene is connected with others. The top 100 highly connected genes are from either M1 or M3, and they are enriched with kinases (4E−7), which indicates the important role of kinase in the network. To check if these genes are date or party hubs, the proportion of connections within or outside of their own module, which is the intramodule connectivity divided by global connectivity, was calculated (Chang et al. 2013). The average intermodule connectivity proportion for the 100 genes is 8%, which is significantly lower compared to 100 random genes with 1000 times sampling ($P < 0.01$), which suggests that M1- or M3-associated kinases serve as party hubs. However, party hubs interact with most of their partners simultaneously, whereas date hubs bind different partners at different locations and times. A yeast data analysis revealed that kinases fall largely into the date hub category (Agarwal et al. 2010). Recently, it was proposed that a party hub might undergo a rapid transition to a date hub (or vice versa) as expression levels/post-translational modifications changed (Dietz et al. 2010). In our network, a total of 9 modules are annotated with kinase-related terms. To examine if kinases in other modules are party hubs, kinases were extracted and intermodule connectivity proportions were calculated. Compared to whole network genes, kinases in M1, M2 and M3 are party hubs ($P < 1E{-}8$, $P < 0.01$ and $P < 1E{-}9$), but not kinases in M11 and M17. The conclusion also stands when compared to all the kinases.

When incorporating current lethal gene information (Lloyd et al. 2015), a permutation test shows that the lethal genes have a higher connectivity at the $P < 0.05$ level but are not significantly higher at the $P < 0.01$ level (60 tests with $P > 0.01$ in 1000 permutations). A hypergeometric test suggests there are 4 modules enriched with lethal genes, including M6 ($P = 0$), M5 ($P = 1E{-}10$), M18 ($P = 1E{-}5$), and M4 ($P = 1E{-}4$). Another trait is that seed pigment-associated genes are also enriched in M2 ($P = 1E{-}9$) and M6 ($P = 0$). Interestingly, those modules are all preserved in the rice transcriptome, as discussed in the following section.

Gene expression variation in modules

We have reduced the transcriptome data complexity by gene co-expression modules. We suppose that these modules can be treated as function modules. The modular gene expression variation may infer whether the function is more basal or conditional. The relative gene variation (relative standard deviation of gene expression) was calculated for each gene and then averaged across modules. The top 3 highly stable

modules include M48 (photorespiration), M20 (glycolysis) and M4 (translation). The top 3 highly variable modules include M28 (proteolysis), M49 (cytidine deamination), and M30 (pollen exine formation) (Supplementary Table S6).

The correlation between gene expression and connectivity was calculated to observe the relationship for each module (Supplementary Table S6). Two types of correlation were found. One was positive correlation, such as M2, M4, M5, M6 and M15, and most of those modules are involved in synthesis biological processes. Another type was negative correlation, such as M23, M40, M49 and M50, and most of those modules are involved in biological degradation processes.

Correlating modules with experimental conditions/traits

After co-expressed gene module identification, we checked the expression status of a specific module from experiments. Linking the modular gene expression with experimental conditions may help to discover modules functioning under a specific condition (Supplementary Table S6). If a module has a high ME, then the module could be a signature for that trait or experiment. For example, combined heat and anoxia treatment leads to the highest modular expression in M32, which is enriched with gene responses to heat. A significant overlap between the anoxic and the heat responses was reported. The transcription factor heat shock factor A2 (*HsfA2*) is induced by both heat and anoxia, and it was strongly induced by anoxia (Banti et al. 2010). On the other side, the lowest M32 expression is presented in arrested development 3 (add3) mutant *Arabidopsis*. It has been shown that add3 mutation prevents the expansion of leaf blades at high temperature, which suggests that add3 affects genes involved in inherently temperature-sensitive developmental processes (Pickett et al. 1996). The hub genes include *AT5G37670*, *AT1G30070* and *HSP23.6-MITO* (Fig. 2a). Our results confirm these studies from a module-based perspective. The module genes can serve as signatures or candidates for thermo-tolerance.

Although M51 has no significant functional annotation, it is highly expressed in the seed coat at the bending cotyledon stage, which was inferred from the transcriptome atlas of the *Arabidopsis* maternal seed subregions (Khan et al. 2015) and indicates the potential role of high expression of M51 in germination. On the other hand, the lowest M51 expression is induced when seedlings were treated with MG132 to block proteasome function and to increase TOC1 protein levels, which is an essential component of the *Arabidopsis* circadian system (Gendron et al. 2012). It has been demonstrated that MG132 treatment completely inhibited seedling growth from dissected embryos, which suggests that proteasome activity is required for germination (Chiu et al. 2016). Eight out of 36 M51 genes have been reported to be seed coat epidermal-specific genes (Esfandiari et al. 2013). The hubs include AtMES esterase family genes *MES6* and *MES4* (Fig. 2b), while the MES family has been implicated in hormone homeostasis and germination (Vlot et al. 2008;

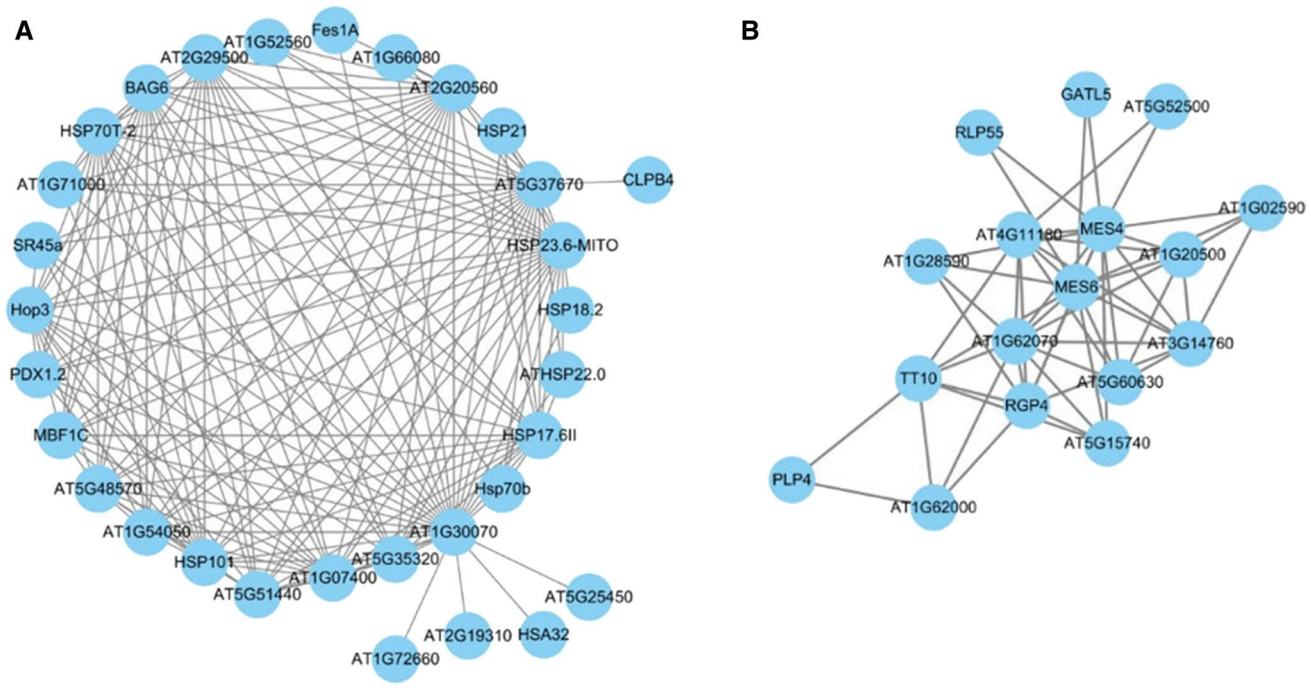

Fig. 2 Representative module network visualization for M32 and M51. **a** Network visualization for M32. **b** Network visualization for M51

Rajjou et al. 2006; Yang et al. 2008). Therefore, M51 may play roles in germination. Similarly, two other modules that have functional annotations, but that were not significant after P value adjustment, include M50, which may perform nectar secretion, and M42, which may perform manganese ion transmembrane transport, as indicated by their expression in lateral nectaries and embryo roots.

Genomic positional gene enrichment analysis

To check whether the 52 modules were associated with specific chromosome regions, modular genes were tested with the Positional Gene Enrichment analysis tool. At a stringent P value (7E−7) and using "more than 3 gene hits" criteria, 5 modules, including M9, M24, M30, M36, and M49, were identified as enriched within a specific chromosome region (Supplementary Table S7).

Genes function prediction

M32, which contains 98 genes, was selected to demonstrate the application of a gene co-expression module in gene function prediction. Only 3 genes were annotated as hypothetical protein coding genes, and the other 95 genes have unambiguous functional annotations. To confirm their role in heat shock, these 98 genes were submitted to NCBI PubMed, NCBI Gene and GoogleScholar to check their association with heat shock. We found 25 HSPs, 6 HSFs, 14

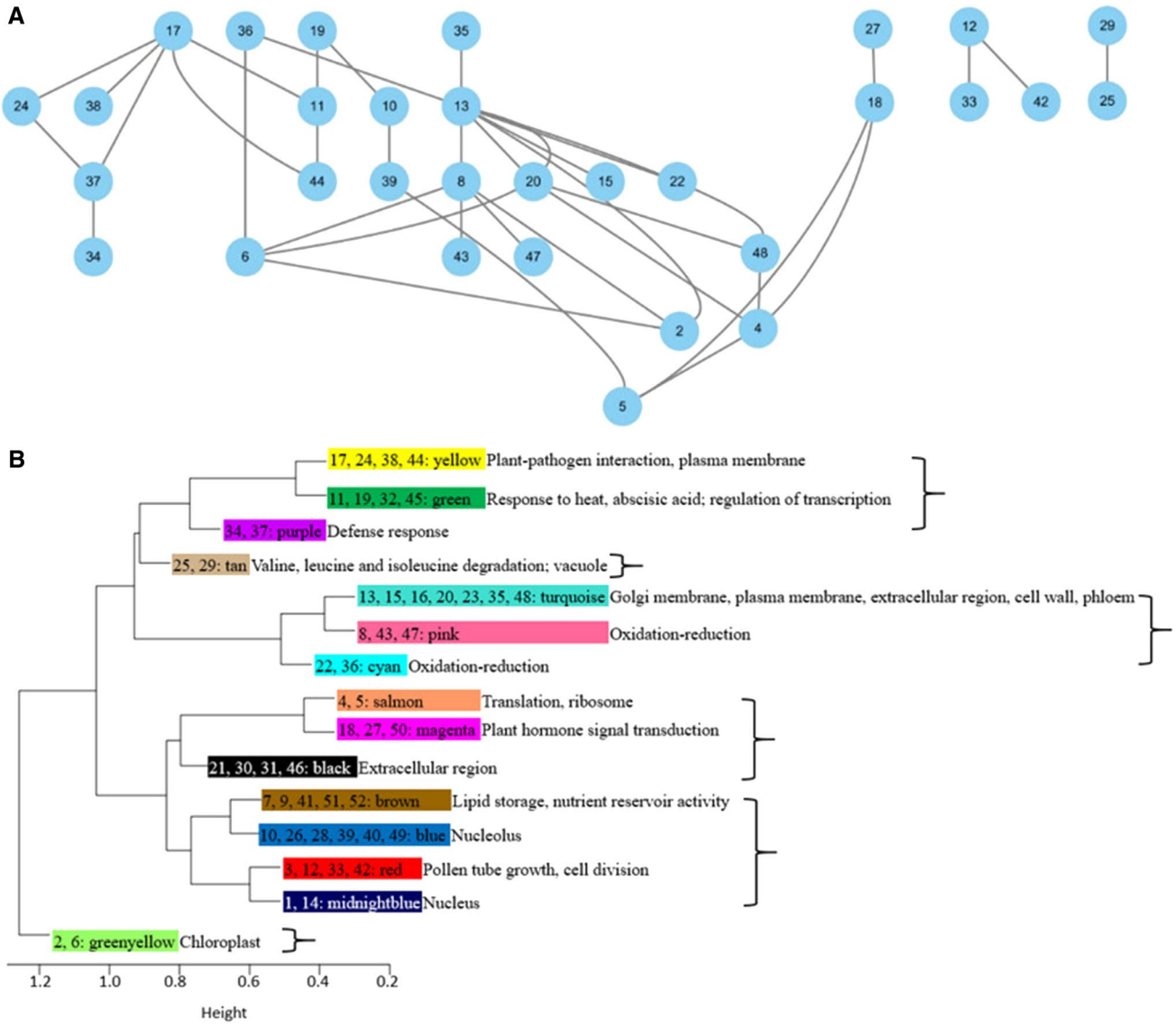

Fig. 3 Higher-order network and its relationships. **a** A network shows the connections between modules. **b** Hierarchical clustering shows the 52 modules organize into 15 meta-modules, which are denoted with colors and module numbers at corresponding leaves

chaperones, 48 heat shock response genes (Supplementary Table S8). However, some of the 48 heat shock response genes are from microarray studies and have no functional experiment support. The remaining 5 genes (*AT2G41170*, *AT4G02550*, *AT1G44414*, *AT1G55530*, and *AT3G56250*) have not been reported to be associated with heat shock in any previous studies. So, these 5 genes could be potential heat shock-responsive genes based on guilt by association, which merits functional verification.

Higher-order module organization

To observe the organization between these modules, the networks of the 52 identified modules were also analyzed. These 52 modules are organized into 15 interconnecting meta-modules. A global connectivity analysis shows that the top 3 highly connected modules were M13, M8, and M4, which have annotations that include biosynthesis of secondary metabolites, oxidation reduction, and translation, respectively (Fig. 3a, Supporting Table S9). The results may indicate the importance and complexity of secondary metabolites (Kliebenstein 2004).

To check the relationships between these meta-modules, a clustering diagram was plotted, which showed that these 15 meta-modules can be divided into 6 major branches. The two orphan branches are meta-modules in green–yellow and tan, which correspond to chloroplast and valine, and leucine and isoleucine degradation (Fig. 3b). From these results, it can be inferred that chloroplast genes have a distinct transcriptional pattern compared to the other 14 meta-modules.

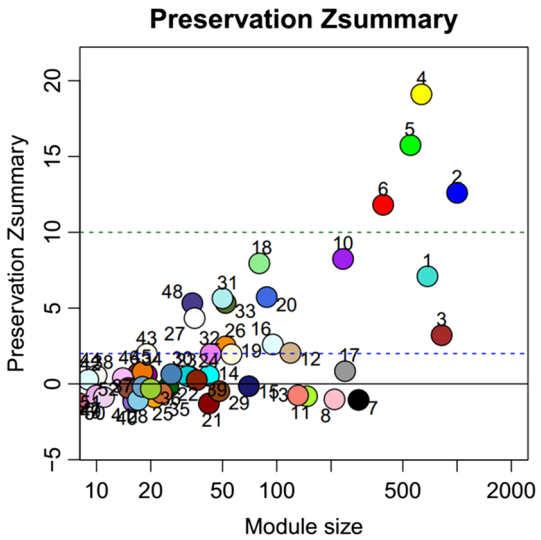

Fig. 4 Rice transcriptome preservation in *Arabidopsis*. Dashed green and blue lines represent the Zsummary threshold for strong ($Z > 10$) and weak–moderate ($2 < Z < 10$) module preservation. Numbers along with coloured dots represent the identified modules. Module size is the number of rice orthologs for *Arabidopsis*

Comparison with previous *Arabidopsis* networks

To confirm our results, multiple publications results were compared. Mao and colleagues constructed a gene expression map from 1094 *Arabidopsis* microarrays and identified 382 sets of highly correlated genes (Mao et al. 2009). They identified 46 modules with significantly enriched GO terms, and 38 of those terms share common GO terms with our modules. However, their modules are small, and the analysis was limited to annotated genes. For example, they identified just 6 genes that respond to heat. Mutwil used 351 microarray data and identified 181 gene clusters, and 27 of the 34 significant clusters share common functional annotations in our results (Mutwil et al. 2010).

Comparison with rice transcriptome

To test if the identified modules were also present in rice, we further collected 2043 rice microarray data points and projected the transcriptomes to *Arabidopsis* modules using the R function modulePreservation in the WGCNA package. The analysis showed that only 4 modules (M2, M4, M5 and M6) were highly preserved, 11 modules (M1, M3, M10, M16, M18, M20, M26, M27, M33, M39 and M48) were weak to moderately preserved, and the other 37 modules had lower preservation Zsummary than the former 15 modules (Fig. 4). M4 showed the strongest preservation, while M21 showed the lowest (Supplementary Table S10). The well-preserved modules include modules associated with translation, rRNA processing, photosynthesis and chloroplast organization. An example of not preserved module is M7 (rank = 47), which involves lipid storage. Evidence suggested that the lipid change patterns in rice are different from those in *Arabidopsis* (Wang et al. 2012a). To provide more details about preservation between *Arabidopsis* and rice, 8 KEGG pathways, including Photosynthesis, Ribosome, Ubiquitin-mediated proteolysis, Endocytosis, Plant–pathogen interaction, Porphyrin and chlorophyll metabolism, Plant hormone signal transduction, and Phenylpropanoid biosynthesis were analyzed. These pathways were identified in both the *Arabidopsis* and rice datasets. Module preservation statistics and module membership correlation analysis show that the Photosynthesis, Ribosome, Endocytosis, Porphyrin and chlorophyll metabolism, and Phenylpropanoid biosynthesis pathways were preserved, while preservation Zsummary was low for Ubiquitin-mediated proteolysis, Plant–pathogen interaction and Plant hormone signal transduction (Fig. 5). Networks demonstrated the similarity between the Photosynthesis and Ribosome pathways as well as differences in hormone signaling between *Arabidopsis* and rice (Fig. 6). The unique complexities of hormone-mediated defence networking have been summarized in recent publications (De Vleesschauwer et al. 2014; Ma et al. 2010). For a better exploration of the KEGG pathways between the two

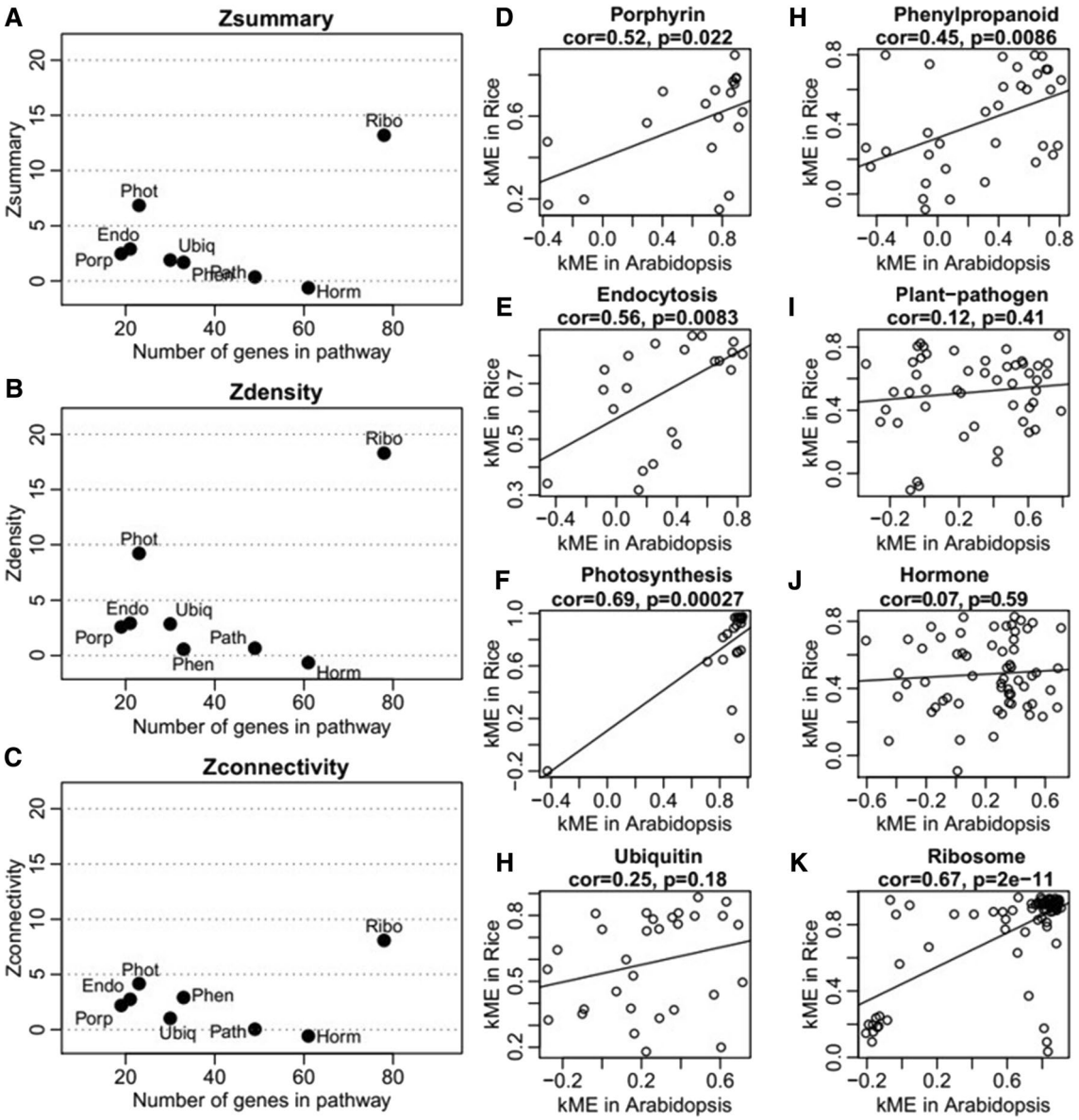

Fig. 5 KEGG pathway-based analysis of the preservation between *Arabidopsis* and rice. **a**, **b** and **c** show the preservation statistics, and $Z > 10$ indicates strong preservation, whereas $2 < Z < 10$ indicates weak–moderate preservation. The dot labels are extracted from the first four letters for each pathway name. **d**, **e**, **f**, **g**, **h**, **i**, **j** and **k** plot the kME correlation for orthologs in the 8 pathways, including the Porphyrin and chlorophyll metabolism, Endocytosis, Photosynthesis, Ubiquitin-mediated proteolysis, Phenylpropanoid biosynthesis, Plant–pathogen interaction, Plant hormone signal transduction, and Ribosome pathways

species, a shiny-based web viewer was developed (Chang et al. 2015). The tool is available here: http://bioinformatics.fafu.edu.cn/arabi/.

Discussion

With the development of high-throughput technologies, large-scale data-based network analysis has become a robust method to discover the gene function, cellular machinery, modularity, conservation or tissue differences (Boruc et al. 2010; He and Maslov 2016; Ruprecht et al. 2017; He et al. 2016). WGCNA has been widely used in biomedical research; however, its application in plants has lagged due to small sample sizes in individual experiments, which is an important factor in GCN analysis. Although the state-of-the-art RNA-Seq data of *Arabidopsis* have accumulated over the years, the diversity in library type, sequencing depth, and instrument type makes it hard to conduct integrative analysis. Studies have

showed that microarray data seem more suited for gene network analysis (Giorgi et al. 2013), and a larger sample size may help to more robustly detect the gene co-expression modules (Oldham et al. 2008). In this study, we collected a compendium of *Arabidopsis* transcriptome data and identified 52 co-expressed gene modules. All the sample data were pooled together to construct a pan network that was non-targeted, which is different from many previous works that were targeted at one or more specific biological conditions, as mentioned in the Introduction. Thus, we can identify pan modules that are a compilation of nearly all possible co-expressed modules under various conditions. We believe that gene–gene connections exist underlying any transcriptome even if those connections cannot be detected when all the samples have the same status (e.g., a static transcriptome). The inherent gene–gene connections can be detected when perturbation occurs within the network, such as the biomedical project Connectivity Map in which cell line transcriptomes were perturbed by drug treatments to find gene signatures (Lamb et al. 2006). Therefore, the tissue origin or experimental treatment can both be considered as a perturbation, which is the reason for our cross-studies data integration. Fortunately, the standardized microarray platform data provide a possibility for integration.

Although early studies have constructed non-targeted GCNs for *Arabidopsis* (Mao et al. 2009; Mutwil et al. 2010; He and Maslov 2016), our analysis improves power and has a different perspective that is more concentrated on biological meanings. First, after identifying the gene modules, module stability was tested by half-sampling. Second, connectivity-based analysis shows the important genes in the global network, and module-based gene variation analysis may infer potential basal and responsive modules. Third, we further analyzed the associations between modules and experimental conditions (phenotype annotation), which may help to identify important modules under specific conditions/traits. These modules could be potential signatures for a trait. Finally, as an important crop, rice gene function research often starts from *Arabidopsis* orthologs. A comparative network analysis of relevant KEGG pathways shows network-based transcriptional conservation of some basal cellular machinery and divergence of some responsive modules between *Arabidopsis* and rice. The divergent components may reflect species diversity, but conserved components define preserved gene models across species that may facilitate standardization of experimental models (Mueller et al. 2017). Although the rank of preservation statistics can reflect the pathway preservation between *Arabidopsis* and rice, the statistics should be interpreted with caution due

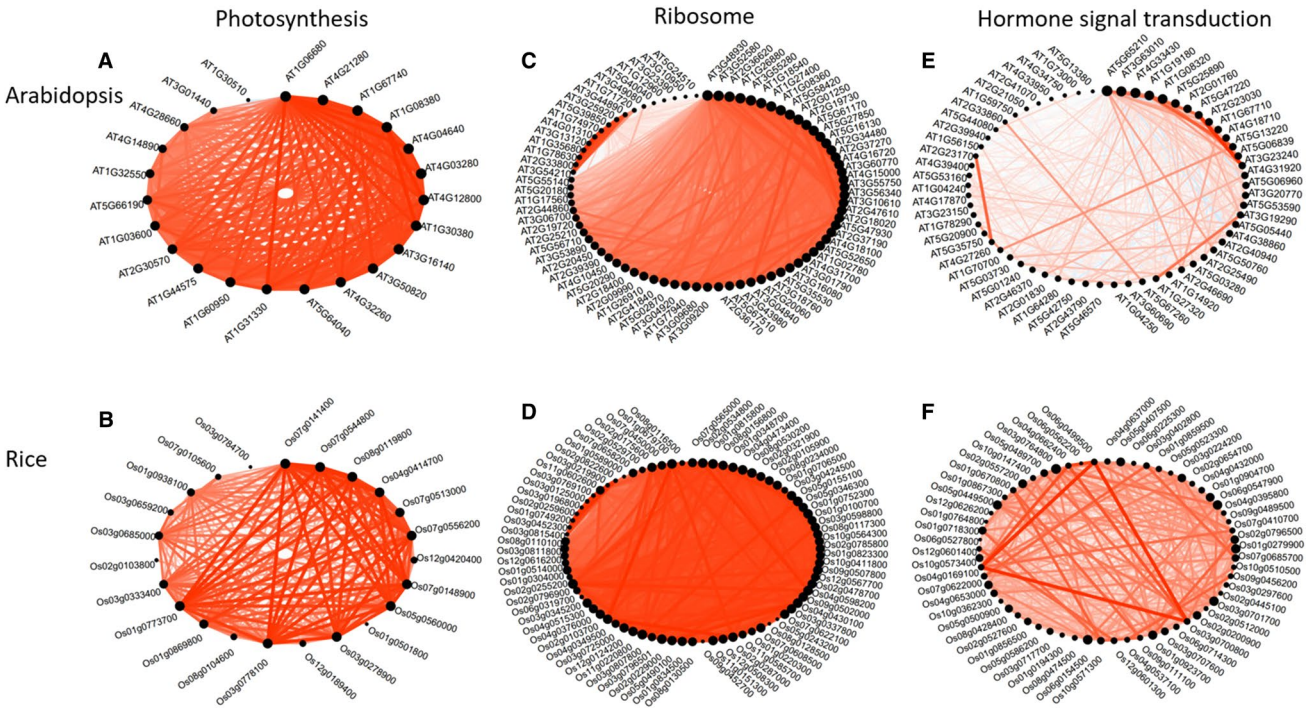

Fig. 6 KEGG pathways-based network visualization for *Arabidopsis* and rice. KEGG pathway-based network for **a** Photosynthesis; **b** Ribosome; **c** Hormone signal transduction. The size of the black dot denotes gene connectivity in *Arabidopsis* and its orthologs in rice. The red line connecting two dots represents the connection strength between them

to the imbalanced tissues/conditions representation in the datasets analyzed.

Overall, our results provide a comprehensive network view for the *Arabidopsis* transcriptome and may serve as a valuable resource for candidate gene function investigation.

Author contribution statement WL, LL, and HH conceived and designed this study. ZZ and YL collected data. SL, KG and HT analyzed the data. WL drafted the manuscript and developed the web tool. All authors read and approved the manuscript.

Acknowledgements There are so many insightful literatures about gene co-expression analysis. The authors apologize that not all related studies were cited due to lack of space.

Funding This work was supported in part by the National Natural Science Foundation of China (Grant numbers 31270454 and 81502091) and Open Project of Key laboratory of Loquat Germplasm Innovation and Utilization, Putian University, Fujian Province (Grant number 2017003).

Compliance with ethical standards

Conflicts of interest The authors have no conflicts of interest to declare.

References

Agarwal S, Deane CM, Porter MA, Jones NS (2010) Revisiting date and party hubs: novel approaches to role assignment in protein interaction networks. PLoS Comput Biol 6(6):e1000817. https://doi.org/10.1371/journal.pcbi.1000817

Amrine KC, Blanco-Ulate B, Cantu D (2015) Discovery of core biotic stress responsive genes in *Arabidopsis* by weighted gene co-expression network analysis. PLoS ONE 10(3):e0118731. https://doi.org/10.1371/journal.pone.0118731

Aoki K, Ogata Y, Shibata D (2007) Approaches for extracting practical information from gene co-expression networks in plant biology. Plant Cell Physiol 48(3):381–390. https://doi.org/10.1093/pcp/pcm013

Aoki Y, Okamura Y, Tadaka S, Kinoshita K, Obayashi T (2016) ATTED-II in 2016: a plant coexpression database towards lineage-specific coexpression. Plant Cell Physiol 57(1):e5. https://doi.org/10.1093/pcp/pcv165

Banti V, Mafessoni F, Loreti E, Alpi A, Perata P (2010) The heat-inducible transcription factor HsfA2 enhances anoxia tolerance in *Arabidopsis*. Plant Physiol 152(3):1471–1483. https://doi.org/10.1104/pp.109.149815

Batada NN, Hurst LD, Tyers M (2006) Evolutionary and physiological importance of hub proteins. PLoS Comput Biol 2(7):e88. https://doi.org/10.1371/journal.pcbi.0020088

Bolstad BM, Irizarry RA, Astrand M, Speed TP (2003) A comparison of normalization methods for high density oligonucleotide array data based on variance and bias. Bioinformatics 19(2):185–193

Boruc J, Van den Daele H, Hollunder J, Rombauts S, Mylle E, Hilson P, Inze D, De Veylder L, Russinova E (2010) Functional modules in the *Arabidopsis* core cell cycle binary protein–protein interaction network. Plant Cell 22(4):1264–1280. https://doi.org/10.1105/tpc.109.073635

Chang X, Xu T, Li Y, Wang K (2013) Dynamic modular architecture of protein-protein interaction networks beyond the dichotomy of 'date' and 'party' hubs. Sci Rep 3:1691. https://doi.org/10.1038/srep01691

Chang W, Cheng J, Allaire JJ, Xie Y, McPherson J (2015) Shiny: web application framework for R. R package version 011 1(4):106

Chiu RS, Pan S, Zhao R, Gazzarrini S (2016) ABA-dependent inhibition of the ubiquitin proteasome system during germination at high temperature in *Arabidopsis*. Plant J 88(5):749–761. https://doi.org/10.1111/tpj.13293

De Preter K, Barriot R, Speleman F, Vandesompele J, Moreau Y (2008) Positional gene enrichment analysis of gene sets for high-resolution identification of overrepresented chromosomal regions. Nucleic Acids Res 36(7):e43. https://doi.org/10.1093/nar/gkn114

De Vleesschauwer D, Xu J, Hofte M (2014) Making sense of hormone-mediated defense networking: from rice to *Arabidopsis*. Front Plant Sci 5:611. https://doi.org/10.3389/fpls.2014.00611

Dietz KJ, Jacquot JP, Harris G (2010) Hubs and bottlenecks in plant molecular signalling networks. New Phytol 188(4):919–938. https://doi.org/10.1111/j.1469-8137.2010.03502.x

Esfandiari E, Jin Z, Abdeen A, Griffiths JS, Western TL, Haughn GW (2013) Identification and analysis of an outer-seed-coat-specific promoter from *Arabidopsis thaliana*. Plant Mol Biol 81(1–2):93–104. https://doi.org/10.1007/s11103-012-9984-0

Feltus FA, Ficklin SP, Gibson SM, Smith MC (2013) Maximizing capture of gene co-expression relationships through pre-clustering of input expression samples: an *Arabidopsis* case study. BMC Syst Biol 7:44. https://doi.org/10.1186/1752-0509-7-44

Ficklin SP, Feltus FA (2011) Gene coexpression network alignment and conservation of gene modules between two grass species: maize and rice. Plant Physiol 156(3):1244–1256. https://doi.org/10.1104/pp.111.173047

Ficklin SP, Dunwoodie LJ, Poehlman WL, Watson C, Roche KE, Feltus FA (2017) Discovering condition-specific gene co-expression patterns using gaussian mixture models: a cancer case study. Sci Rep 7(1):8617. https://doi.org/10.1038/s41598-017-09094-4

Gendron JM, Pruneda-Paz JL, Doherty CJ, Gross AM, Kang SE, Kay SA (2012) *Arabidopsis* circadian clock protein, TOC1, is a DNA-binding transcription factor. Proc Natl Acad Sci USA 109(8):3167–3172. https://doi.org/10.1073/pnas.1200355109

Giorgi FM, Bolger AM, Lohse M, Usadel B (2010) Algorithm-driven artifacts in median polish summarization of microarray data. BMC Bioinform 11:553. https://doi.org/10.1186/1471-2105-11-553

Giorgi FM, Del Fabbro C, Licausi F (2013) Comparative study of RNA-seq- and microarray-derived coexpression networks in *Arabidopsis thaliana*. Bioinformatics 29(6):717–724. https://doi.org/10.1093/bioinformatics/btt053

He F, Maslov S (2016) Pan- and core- network analysis of co-expression genes in a model plant. Sci Rep 6:38956. https://doi.org/10.1038/srep38956

He F, Yoo S, Wang D, Kumari S, Gerstein M, Ware D, Maslov S (2016) Large-scale atlas of microarray data reveals the distinct expression landscape of different tissues in *Arabidopsis*. Plant J 86(6):472–480. https://doi.org/10.1111/tpj.13175

Huang DW, Sherman BT, Lempicki RA (2009) Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. Nucleic Acids Res 37(1):1–13. https://doi.org/10.1093/nar/gkn923

Huber W, Carey VJ, Gentleman R, Anders S, Carlson M, Carvalho BS, Bravo HC, Davis S, Gatto L, Girke T, Gottardo R, Hahne F, Hansen KD, Irizarry RA, Lawrence M, Love MI, MacDonald

J, Obenchain V, Oles AK, Pages H, Reyes A, Shannon P, Smyth GK, Tenenbaum D, Waldron L, Morgan M (2015) Orchestrating high-throughput genomic analysis with Bioconductor. Nat Methods 12(2):115–121. https://doi.org/10.1038/nmeth.3252

Khan D, Millar JL, Girard IJ, Chan A, Kirkbride RC, Pelletier JM, Kost S, Becker MG, Yeung EC, Stasolla C, Goldberg RB, Harada JJ, Belmonte MF (2015) Transcriptome atlas of the *Arabidopsis* funiculus—a study of maternal seed subregions. Plant J 82(1):41–53. https://doi.org/10.1111/tpj.12790

Kliebenstein D (2004) Secondary metabolites and plant/environment interactions: a view through *Arabidopsis thaliana* tinged glasses. Plant Cell Environ 27(6):675–684

Lamb J, Crawford ED, Peck D, Modell JW, Blat IC, Wrobel MJ, Lerner J, Brunet JP, Subramanian A, Ross KN, Reich M, Hieronymus H, Wei G, Armstrong SA, Haggarty SJ, Clemons PA, Wei R, Carr SA, Lander ES, Golub TR (2006) The Connectivity Map: using gene-expression signatures to connect small molecules, genes, and disease. Science 313(5795):1929–1935. https://doi.org/10.1126/science.1132939

Langfelder P, Horvath S (2008) WGCNA: an R package for weighted correlation network analysis. BMC Bioinform 9:559. https://doi.org/10.1186/1471-2105-9-559

Langfelder P, Luo R, Oldham MC, Horvath S (2011) Is my network module preserved and reproducible? PLoS Comput Biol 7(1):e1001057. https://doi.org/10.1371/journal.pcbi.1001057

Lee T, Yang S, Kim E, Ko Y, Hwang S, Shin J, Shim JE, Shim H, Kim H, Kim C, Lee I (2015) AraNet v2: an improved database of co-functional gene networks for the study of *Arabidopsis thaliana* and 27 other nonmodel plant species. Nucleic Acids Res 43:996–1002. https://doi.org/10.1093/nar/gku1053 **(database issue)**

Li Y, Pearl SA, Jackson SA (2015) Gene networks in plant biology: approaches in reconstruction and analysis. Trends Plant Sci 20(10):664–675. https://doi.org/10.1016/j.tplants.2015.06.013

Lloyd JP, Seddon AE, Moghe GD, Simenc MC, Shiu SH (2015) Characteristics of plant essential genes allow for within- and between-species prediction of lethal mutant phenotypes. Plant Cell 27(8):2133–2147. https://doi.org/10.1105/tpc.15.00051

Ma B, Chen S, Zhang J (2010) Ethylene signaling in rice. Chin Sci Bull 55(21):2204–2210

Mao L, Van Hemert JL, Dash S, Dickerson JA (2009) *Arabidopsis* gene co-expression network and its functional modules. BMC Bioinform 10:346. https://doi.org/10.1186/1471-2105-10-346

Movahedi S, Van Bel M, Heyndrickx KS, Vandepoele K (2012) Comparative co-expression analysis in plant biology. Plant Cell Environ 35(10):1787–1798. https://doi.org/10.1111/j.1365-3040.2012.02517.x

Mueller AJ, Canty-Laird EG, Clegg PD, Tew SR (2017) Cross-species gene modules emerge from a systems biology approach to osteoarthritis. NPJ Syst Biol Appl 3:13. https://doi.org/10.1038/s41540-017-0014-3

Mutwil M, Usadel B, Schutte M, Loraine A, Ebenhoh O, Persson S (2010) Assembly of an interactive correlation network for the *Arabidopsis* genome using a novel heuristic clustering algorithm. Plant Physiol 152(1):29–43. https://doi.org/10.1104/pp.109.145318

Mutwil M, Klie S, Tohge T, Giorgi FM, Wilkins O, Campbell MM, Fernie AR, Usadel B, Nikoloski Z, Persson S (2011) PlaNet: combined sequence and expression comparisons across plant networks derived from seven species. Plant Cell 23(3):895–910. https://doi.org/10.1105/tpc.111.083667

Nishiyama S, Onoue N, Kono A, Sato A, Yonemori K, Tao R (2018) Characterization of a gene regulatory network underlying astringency loss in persimmon fruit. Planta 247(3):733–743. https://doi.org/10.1007/s00425-017-2819-0

Oldham MC, Konopka G, Iwamoto K, Langfelder P, Kato T, Horvath S, Geschwind DH (2008) Functional organization of the transcriptome in human brain. Nat Neurosci 11(11):1271–1282

Penga J, Wang T, Huc J, Wang Y, Chen J (2016) Constructing networks of organelle functional modules in *Arabidopsis*. Curr Genom 17(5):427–438. https://doi.org/10.2174/1389202917666160726151048

Pepper SD, Saunders EK, Edwards LE, Wilson CL, Miller CJ (2007) The utility of MAS5 expression summary and detection call algorithms. BMC Bioinform 8:273. https://doi.org/10.1186/1471-2105-8-273

Pickett FB, Champagne MM, Meeks-Wagner DR (1996) Temperature-sensitive mutations that arrest *Arabidopsis* shoot development. Development 122(12):3799–3807

Prasch CM, Sonnewald U (2013) Simultaneous application of heat, drought, and virus to *Arabidopsis* plants reveals significant shifts in signaling networks. Plant Physiol 162(4):1849–1866. https://doi.org/10.1104/pp.113.221044

Rajjou L, Belghazi M, Huguet R, Robin C, Moreau A, Job C, Job D (2006) Proteomic investigation of the effect of salicylic acid on *Arabidopsis* seed germination and establishment of early defense mechanisms. Plant Physiol 141(3):910–923. https://doi.org/10.1104/pp.106.082057

Rasmussen S, Barah P, Suarez-Rodriguez MC, Bressendorff S, Friis P, Costantino P, Bones AM, Nielsen HB, Mundy J (2013) Transcriptome responses to combinations of stresses in *Arabidopsis*. Plant Physiol 161(4):1783–1794. https://doi.org/10.1104/pp.112.210773

R Development Core Team (2013) R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. http://www.R-project.org/. Accessed 1 May 2018

Ruprecht C, Proost S, Hernandez-Coronado M, Ortiz-Ramirez C, Lang D, Rensing SA, Becker JD, Vandepoele K, Mutwil M (2017) Phylogenomic analysis of gene co-expression networks reveals the evolution of functional modules. Plant J 90(3):447–465. https://doi.org/10.1111/tpj.13502

Schmid M, Davison TS, Henz SR, Pape UJ, Demar M, Vingron M, Scholkopf B, Weigel D, Lohmann JU (2005) A gene expression map of *Arabidopsis thaliana* development. Nat Genet 37(5):501–506. https://doi.org/10.1038/ng1543

Shaik R, Ramakrishna W (2013) Genes and co-expression modules common to drought and bacterial stress responses in *Arabidopsis* and rice. PLoS ONE 8(10):e77261. https://doi.org/10.1371/journal.pone.0077261

Usadel B, Obayashi T, Mutwil M, Giorgi FM, Bassel GW, Tanimoto M, Chow A, Steinhauser D, Persson S, Provart NJ (2009) Co-expression tools for plant biology: opportunities for hypothesis generation and caveats. Plant Cell Environ 32(12):1633–1651. https://doi.org/10.1111/j.1365-3040.2009.02040.x

van Veen H, Vashisht D, Akman M, Girke T, Mustroph A, Reinen E, Hartman S, Kooiker M, van Tienderen P, Schranz ME, Bailey-Serres J, Voesenek LA, Sasidharan R (2016) Transcriptomes of eight *Arabidopsis thaliana* accessions reveal core conserved, genotype- and organ-specific responses to flooding stress. Plant Physiol 172(2):668–689. https://doi.org/10.1104/pp.16.00472

Vlot AC, Liu PP, Cameron RK, Park SW, Yang Y, Kumar D, Zhou F, Padukkavidana T, Gustafsson C, Pichersky E, Klessig DF (2008) Identification of likely orthologs of tobacco salicylic acid-binding protein 2 and their role in systemic acquired resistance in *Arabidopsis thaliana*. Plant J 56(3):445–456. https://doi.org/10.1111/j.1365-313X.2008.03618.x

Wang F, Rong W, Wen J, Zhang W (2012a) Quantitative dissection of lipid degradation in rice seeds during accelerated aging. Plant Growth Regul 66(1):49–58

Wang S, Yin Y, Ma Q, Tang X, Hao D, Xu Y (2012b) Genome-scale identification of cell-wall related genes in *Arabidopsis* based on

co-expression network analysis. BMC Plant Biol 12:138. https://doi.org/10.1186/1471-2229-12-138

Yang Y, Xu R, Ma CJ, Vlot AC, Klessig DF, Pichersky E (2008) Inactive methyl indole-3-acetic acid ester can be hydrolyzed and activated by several esterases belonging to the AtMES esterase family of *Arabidopsis*. Plant Physiol 147(3):1034–1045. https://doi.org/10.1104/pp.108.118224

Yim WC, Yu Y, Song K, Jang CS, Lee BM (2013) PLANEX: the plant co-expression database. BMC Plant Biol 13:83. https://doi.org/10.1186/1471-2229-13-83

Zhang B, Horvath S (2005) A general framework for weighted gene co-expression network analysis. Stat Appl Genet Mol Biol 4:Article17. https://doi.org/10.2202/1544-6115.1128

Zheng X, Liu T, Yang Z, Wang J (2011) Large cliques in *Arabidopsis* gene coexpression network and motif discovery. J Plant Physiol 168(6):611–618. https://doi.org/10.1016/j.jplph.2010.09.010