



Co-expression Networks From Gene Expression Variability Between Genetically Identical Seedlings Can Reveal Novel Regulatory Relationships.

Sandra S. Cortijo, Marcel Bhattarai, James C. W. Locke, Sebastian E. Ahnert

► To cite this version:

Sandra S. Cortijo, Marcel Bhattarai, James C. W. Locke, Sebastian E. Ahnert. Co-expression Networks From Gene Expression Variability Between Genetically Identical Seedlings Can Reveal Novel Regulatory Relationships.. *Frontiers in Plant Science*, 2021, 11, pp.599464. 10.3389/fpls.2020.599464 . hal-03095468

HAL Id: hal-03095468

<https://hal.science/hal-03095468>

Submitted on 4 Jan 2021

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution - NonCommercial - NoDerivatives 4.0 International License



Co-expression Networks From Gene Expression Variability Between Genetically Identical Seedlings Can Reveal Novel Regulatory Relationships

Sandra Cortijo^{1,2}, Marcel Bhattarai¹, James C. W. Locke^{1*} and Sebastian E. Ahnert^{1,3,4,5*}

OPEN ACCESS

Edited by:

Atsushi Fukushima,
RIKEN, Japan

Reviewed by:

Norihito Nakamichi,
Nagoya University, Japan
Torgeir R. Hvidsten,
Norwegian University of Life Sciences,
Norway

*Correspondence:

James C. W. Locke
james.locke@slcu.cam.ac.uk
Sebastian E. Ahnert
sea31@cam.ac.uk

Specialty section:

This article was submitted to
Plant Systems and Synthetic Biology,
a section of the journal
Frontiers in Plant Science

Received: 27 August 2020

Accepted: 05 November 2020

Published: 15 December 2020

Citation:

Cortijo S, Bhattarai M, Locke JCW
and Ahnert SE (2020) Co-expression
Networks From Gene Expression
Variability Between Genetically
Identical Seedlings Can Reveal Novel
Regulatory Relationships.
Front. Plant Sci. 11:599464.
doi: 10.3389/fpls.2020.599464

¹ The Sainsbury Laboratory, University of Cambridge, Cambridge, United Kingdom, ² UMR5004 Biochimie et Physiologie Moléculaire des Plantes, Univ Montpellier, CNRS, INRAE, Institut Agro, Montpellier, France, ³ Theory of Condensed Matter, Cavendish Laboratory, University of Cambridge, Cambridge, United Kingdom, ⁴ Department of Chemical Engineering and Biotechnology, Philippa Fawcett Drive, University of Cambridge, Cambridge, United Kingdom, ⁵ The Alan Turing Institute, British Library, London, United Kingdom

Co-expression networks are a powerful tool to understand gene regulation. They have been used to identify new regulation and function of genes involved in plant development and their response to the environment. Up to now, co-expression networks have been inferred using transcriptomes generated on plants experiencing genetic or environmental perturbation, or from expression time series. We propose a new approach by showing that co-expression networks can be constructed in the absence of genetic and environmental perturbation, for plants at the same developmental stage. For this, we used transcriptomes that were generated from genetically identical individual plants that were grown under the same conditions and for the same amount of time. Twelve time points were used to cover the 24-h light/dark cycle. We used variability in gene expression between individual plants of the same time point to infer a co-expression network. We show that this network is biologically relevant and use it to suggest new gene functions and to identify new targets for the transcriptional regulators GI, PIF4, and PRR5. Moreover, we find different co-regulation in this network based on changes in expression between individual plants, compared to the usual approach requiring environmental perturbation. Our work shows that gene co-expression networks can be identified using variability in gene expression between individual plants, without the need for genetic or environmental perturbations. It will allow further exploration of gene regulation in contexts with subtle differences between plants, which could be closer to what individual plants in a population might face in the wild.

Keywords: networks, gene expression, co-expression analysis, variability, modules, seedlings, *Arabidopsis*

INTRODUCTION

Understanding how transcriptomes are regulated is key to shed light on how plants develop and also respond to environmental fluctuations. A powerful tool often used to reveal transcriptional regulation at a genome-wide level is gene co-expression networks (Stuart et al., 2003; Serin et al., 2016). In gene co-expression networks, genes that co-vary in expression in different conditions are detected and paired together (Usadel et al., 2009; Ruan et al., 2010; Contreras-López et al., 2018). By doing this for the entire transcriptome, a multitude of genes can be linked, indicating a similar gene regulation. Communities of genes, called modules, that are more closely linked can then be identified (Mao et al., 2009). The presence of genes in a given module indicates a close co-regulation and is usually the starting point to look for their implication in the same pathway, or their regulation by the same transcription factor(s) [TF(s)] (Aoki et al., 2007; Zheng et al., 2011; de Luis Balaguer et al., 2017; Liu et al., 2019). Most studies using co-expression networks can be separated into two categories: targeted analyses that use only a subset of genes (selected based on their function or transcriptional regulation) or specific genetic/environmental perturbations and global analyses that make use of hundreds or thousands of transcriptomes performed in various conditions, often publicly available ones, and do not select genes based on their function prior to the co-expression analysis. Co-expression networks are now commonly used in a variety of work in plant research and have allowed the identification or prediction of new genes and TFs involved in development (Xie et al., 2015; Silva et al., 2016; de Luis Balaguer et al., 2017), in metabolic pathways (Wisecaver et al., 2017), and in response to biotic and abiotic stresses (Prasch and Sonnewald, 2013; Shaik and Ramakrishna, 2013; Amrine et al., 2015; Sharma et al., 2018; Liu et al., 2019). It has also been proposed that the topology of the co-expression network and position of genes in the network can be of interest in itself to identify genes involved in natural diversity in development and in the response to environment (Ichihashi et al., 2014; Des Marais et al., 2017).

One limit of gene co-expression networks is that they only provide information about correlation in expression but do not indicate the direction and type of relationship between genes that are co-expressed. In order to define which genes are TFs that regulate the expression of other genes in the network, additional types of data should be used or integrated (Rao and Dixon, 2019). These additional data can be, for example, ChIP-seq (Chen et al., 2018; Kulkarni and Vandepoele, 2019) that provides the list of targets of a given TF, protein–protein interaction (He et al., 2010; Kulkarni and Vandepoele, 2019), as well as the presence of TF binding motifs in the promoter of genes (Vandepoele et al., 2009; Ma et al., 2013). Another limit is that genes should exhibit changes in expression between the different samples used for the analysis in order to detect co-expressed pairs of genes. This is usually achieved by using genetic and/or environmental perturbations in order to cause changes in the transcriptome regulation. However, these perturbations often have large effects, and it can be time-consuming and challenging to produce the large number of samples required. In order to analyze gene regulation in a biologically context that is more relevant, more

subtle changes in expression might be preferred. This could be achieved by using milder genetic or environmental perturbations. Another option would be to analyze changes in expression that occur in the absence of any genetic or environmental perturbation (Bhosale et al., 2013). This can be possible in theory as widespread differences in gene expression levels have been observed between genetically identical plants, in the absence of any environmental perturbation (Hall et al., 2007; Jimenez-Gomez et al., 2011; Shen et al., 2012; Mönchgesang et al., 2016; Cortijo et al., 2019). The idea is to use this variability in gene expression to find potential co-regulation. In mammals, variability in gene expression between single cells of the same cell type has been used to identify co-expression patterns for genes that show a high level of gene expression variability between cells (Mantsoki et al., 2016). Moreover, gene co-expression networks have been inferred using transcriptomes of individual plant leaves, after removing *in silico* the genotype, environment, and genotype $\times$ environment effects on gene expression (Bhosale et al., 2013). The modules identified in this network were functionally relevant, and this study allowed the identification of a new regulator of the jasmonate pathway (Bhosale et al., 2013). It thus shows that the analysis of gene expression regulation can be as powerful in the absence of genetic and environmental fluctuation. However, the first step of the study of Bhosale et al. was to remove *in silico* the genotype, environment, and genotype $\times$ environment effects on gene expression, as the transcriptomes were performed on plants from different genotypes, as well as plants that were grown in different research laboratories. It is thus not clear if co-expression networks can be identified in plants from transcriptomes performed in the absence of genetic and environmental perturbation.

We thus decided to test if it is possible to infer gene co-expression networks using transcriptomes generated on single plants in the absence of any genetic and environmental perturbation. In particular, we wanted to define if such a network would provide different information compared to a network using environmental perturbation. Finally, we wished to determine if modules that would be detected in such a network would have functional relevance. In order to answer these questions, we took advantage of the existence of a set of published transcriptomes carried out on single seedlings of the same genotype that were grown in the same environmental conditions (Cortijo et al., 2019). In this dataset, multiple genetically identical seedlings had been harvested at several time points during a day/night cycle. Differences in expression between seedlings in each time point were previously observed for many genes in this dataset. In particular, 8.7% of the genes in this dataset have been identified as highly variable genes (HVGs), as their expression was statistically more variable between seedlings than the rest of the transcriptome. Using this dataset, we were able to infer co-expression networks in the absence of genetic and environmental perturbations. Based on enrichment in a module for genes involved in flavonoid metabolism, we speculated that AT4G22870, a 2-oxoglutarate (2OG) and Fe(II)-dependent oxygenase, could also have a role in flavonoid metabolism. Finally, we identified new targets for the transcriptional regulators PHYTOCHROME INTERACTING

FACTOR 4 (PIF4), GIGANTEA (GI), and PSEUDO-RESPONSE REGULATOR 5 (PRR5).

RESULTS

Co-expression Networks Can Be Inferred Using Expression Variation Between Individual Seedlings

Co-expression networks in plants are normally inferred using transcriptomes obtained from pools of plants, using genetic or environmental perturbations in order to identify genes that co-vary in expression between these conditions. In order to define if co-expression networks can be inferred from expression

measurements obtained from single seedlings in the absence of genetic and environmental perturbations, we used the previously published dataset of transcriptomes generated on single seedlings grown in the same environment. This dataset contained a total of 14 seedlings per time point, for 12 time points spanning a 24-h day/night cycle (Cortijo et al., 2019). Widespread differences in expression levels have been observed between seedlings in this dataset, which is a prerequisite to be able to infer a co-expression network (Usadel et al., 2009; Ruan et al., 2010; Contreras-López et al., 2018) (**Figure 1A** and **Supplementary Figure S1**). We first detected co-expressed genes in each time point, by measuring Spearman correlation for each pair of genes in profiles of expression in the 14 seedlings of this time point. In order to keep robust correlations in the final network, we used a Benjamini-Hochberg correction with a false-discovery rate of 10% to keep

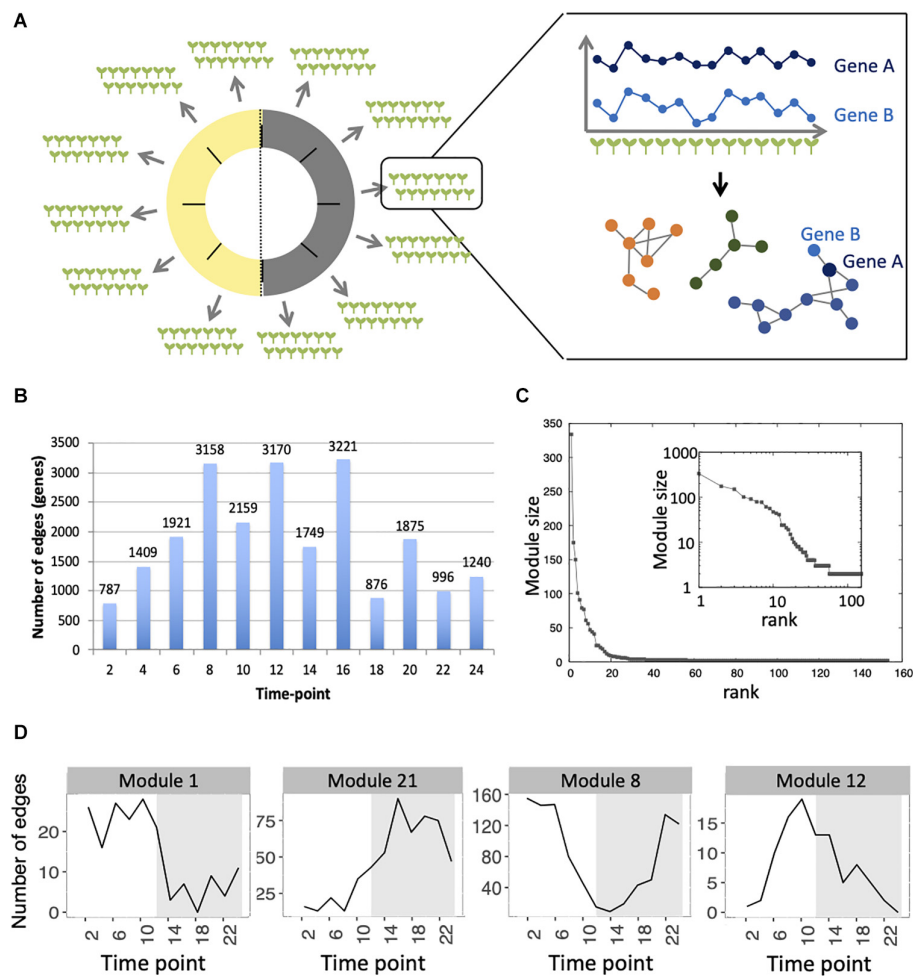


FIGURE 1 | Inference of gene co-expression networks in absence of genetic and environmental perturbations. **(A)** Description of co-expression network inference using transcriptomes performed on single seedlings. Transcriptomes were generated for a total of 14 seedlings per time point, with 12 time points spanning a day/night cycle over 24 h. In each time point, genes with correlated expression profiles in the 14 seedlings were identified. The co-expression network was inferred based on pairs of genes significantly correlated in at least four consecutive time points. Finally, modules in the network, which consist of groups of genes that are densely connected, were detected. **(B)** Total number of edges in the final network that are detected in each time point. **(C)** Distribution of the number of genes present in each of the 153 modules. Inset shows the same data plotted with a logarithmic scale. **(D)** Number of edges that are detected in each time point for four modules: module 1 in which most edges are detected during day time, module 21 in which most edges are detected during the night time, module 8 in which most edges are detected at the transition between night and day, and module 12 in which most edges are detected at the transition between day and night.

the most significant correlations and then selected edges of the network that are detected in at least four consecutive time points, with one gap allowed (section “Materials and Methods”). Using this approach, we find a total of 4715 edges, connecting 1729 genes in this network, from now on referred to as the variability network. The number of edges detected for each time point varies from 787 to 3221, with a higher number of edges being detected at the end of the day and the beginning of the night (**Figure 1B**). We then used the Louvain community detection algorithm in order to identify modules of genes that are densely connected in the network (Blondel et al., 2008). In total, we identified 153 modules (**Supplementary Table S1**), containing between two and 334 genes, with most of the modules only composed of two genes (**Figure 1C**). To test the robustness of the variability network, we also selected significantly correlated edges that are in at least three consecutive time points and compared the detected modules in both networks (**Supplementary Figure S2**). Similar modules with a similar overall connectivity between them are found in the two networks, which confirms the robustness of the modules in our original network. Modules in the network based on four consecutive time points are smaller. In some cases, several modules of the network based on four consecutive time points correspond to a single module in the network based on three consecutive time points and these smaller modules have differences in several features (**Supplementary Figure S2**). That is why we decided to focus our analysis on the network obtained when selecting edges present in four consecutive time points, and in particular for the 28 modules of this network containing five genes or more.

First, we analyzed the number of edges at each time point throughout the time course for each module (**Figure 1D** and **Supplementary Figure S3**). In most modules, the edges are distributed non-uniformly across the 12 time points. Some exhibit a larger number of edges during the day or the night, while in other modules, a larger number of edges are observed at the transitions from night to day, or from day to night. It indicates that genes in these modules are co-regulated at some moments of the day/night cycle but not at others. While for some modules, this is linked to the genes being more expressed at these same times of the day (module 1 for example), this is not the case for other modules in which genes are expressed throughout the time course (module 12 for example; **Figure 2**). Most of the edges in module 1 are observed during the day (**Figure 1D**), and we were able to confirm co-expression of genes in this module by doing an RT-qPCR in a replicate experiment for a few genes in this module. In this replicate experiment, we find a very high correlation during the day (ZT6) and a lower correlation during the night (ZT14) (**Supplementary Figure S4**). On the other hand, most edges in module 21 are observed during the night (**Figure 1D**). We also find in a replicate RT-qPCR experiment that genes of module 21 were more correlated during the night than during the day (**Supplementary Figure S4**). These results confirm that the co-expression of genes in modules of the variability network, and also the differences in co-expression between the day and night, can be reproduced in a replicate experiment. Moreover, we find that modules with a high percentage of edges during the night are more connected to one another than with modules for

which most edges are observed during the day, and vice versa (**Figure 3A**). We can measure this assortativity of the network (i.e., the tendency of similar nodes to be connected to each other) through the Pearson correlation of the daytime-specific edge percentages of connected modules (Pearson correlation = 0.4573, p -value = 0.043). This result shows that modules that are more connected to one another are more similar, at least for this feature, indicating that the community detection in the network worked well and provides modules that are relevant.

Since high gene expression variability between genetically identical plants was previously observed in the transcriptome dataset we used to infer the variability network (Cortijo et al., 2019), we tested if the network is enriched in HVGs. We find a total of 477 HVGs in the network, that is, 27.6% of all genes in the variability network. This is higher than the 8.7% of HVGs that were detected in the full transcriptome dataset (Cortijo et al., 2019). This result suggests that most of the genes in the variability network do not have to display a high level of gene expression variability to be able to detect co-expression between individual seedlings. We find that most modules are either strongly enriched in HVGs, or strongly depleted in HVGs, with only a few modules containing around 27% of HVGs (**Table 1** and **Supplementary Figure S5a**). Modules 37, 43, and 66 for example are only composed of HVGs, while a total of eight modules do not have a single HVG. This result suggests that HVGs can co-vary in expression and are potentially co-regulated. It also suggests that HVGs are not likely to co-vary in expression with non-variable genes. To test if this result could indicate a bias in the method used to construct the variability network and detect modules, we analyzed expression levels in single seedlings for genes in modules with high or low percentage of HVGs (**Supplementary Figure S5b**). We find that modules with high or low percentage of HVGs have different expression profiles in the seedlings, indicating an absence of bias. Moreover, modules with a high percentage of HVGs tend to be more connected to one another than with modules containing a low percentage of HVGs, and vice versa (Pearson correlation = 0.6896, p -value = 0.0007683, **Figure 3B**).

Our results show that gene co-expression networks can be inferred in the absence of genetic or environmental perturbation. Moreover, genes do not need to show a high level of gene expression variability between seedlings to be integrated in the network.

Additional Gene Co-expression Is Identified in the Variability Network Compared to a Network Inferred From Pools of Plants

Next, we decided to test whether the co-expression network based on the variability of expression between genetically identical plants grown in the same environment is different from what is the standard practice in the field. Usually, co-expression networks are inferred by comparing transcriptomes obtained from pools of plants experiencing an environmental or genetic perturbation. Given that our dataset contains data for several time points throughout a day/night cycle, this

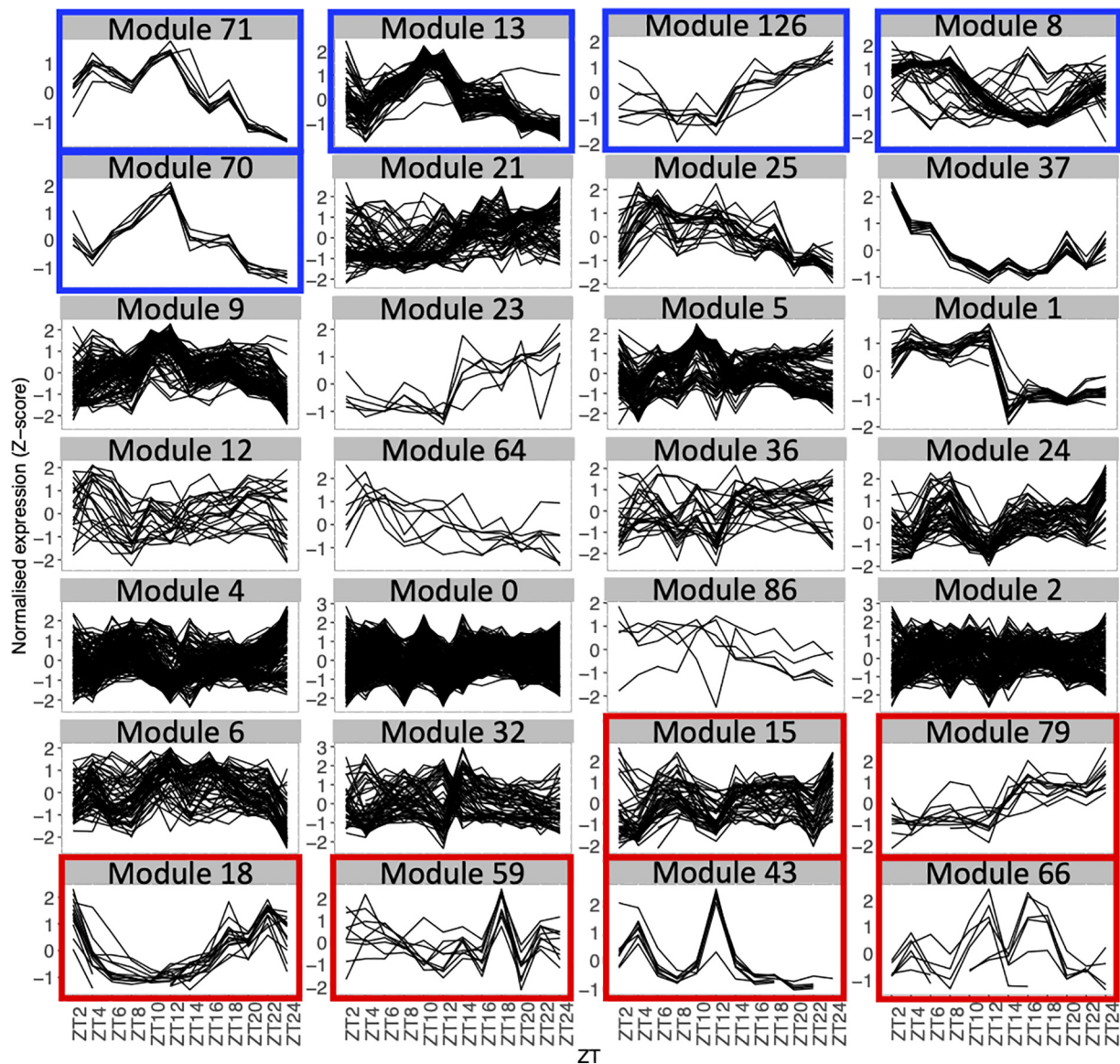


FIGURE 2 | Expression profiles throughout the time course for genes in each module with 5 genes or more, using the average expression of the fourteen seedlings for each time point. Each line represents the normalized expression (z-score) for one gene. Modules are ordered by the percentage of genes in the averaged time course network (high to low). Modules highlighted in blue contain 50% or more of genes that are also in the averaged time course network. Modules highlighted in red contain 15% or less of genes that are also in the averaged time course network.

would correspond to comparing the average expression of the 14 seedlings for each time point and exploiting changes in expression happening during the time course. We used this strategy to infer a co-expression network, referred to as the averaged time course network, that allows the identification of co-expression throughout the time course and is the closest to standard practices using our dataset. Using this approach, we find a total of 9332 edges, connecting 3861 genes in the averaged time course network. A total of 524 genes of this averaged time course network are also present in the variability network, that is, 30% of the genes in the variability network (Table 2). Only 35 edges are shared between the two networks. This result shows that the

majority of the genes and edges present in the variability network are not detected in this dataset using a classical approach with pools of plants.

We find that between 0 and 87.5% of genes in modules of the variability network are also in the averaged time course network, with most of the modules having between 20 and 50% of genes also present in the averaged time course network (Table 2). The modules with the highest percentage of genes also in the averaged time course network are modules 71 (87%: seven out of eight genes) and 13 (76%: 59 out of 77 genes). We find that genes in these modules have very similar and clear expression profiles throughout the time course (Figure 2).

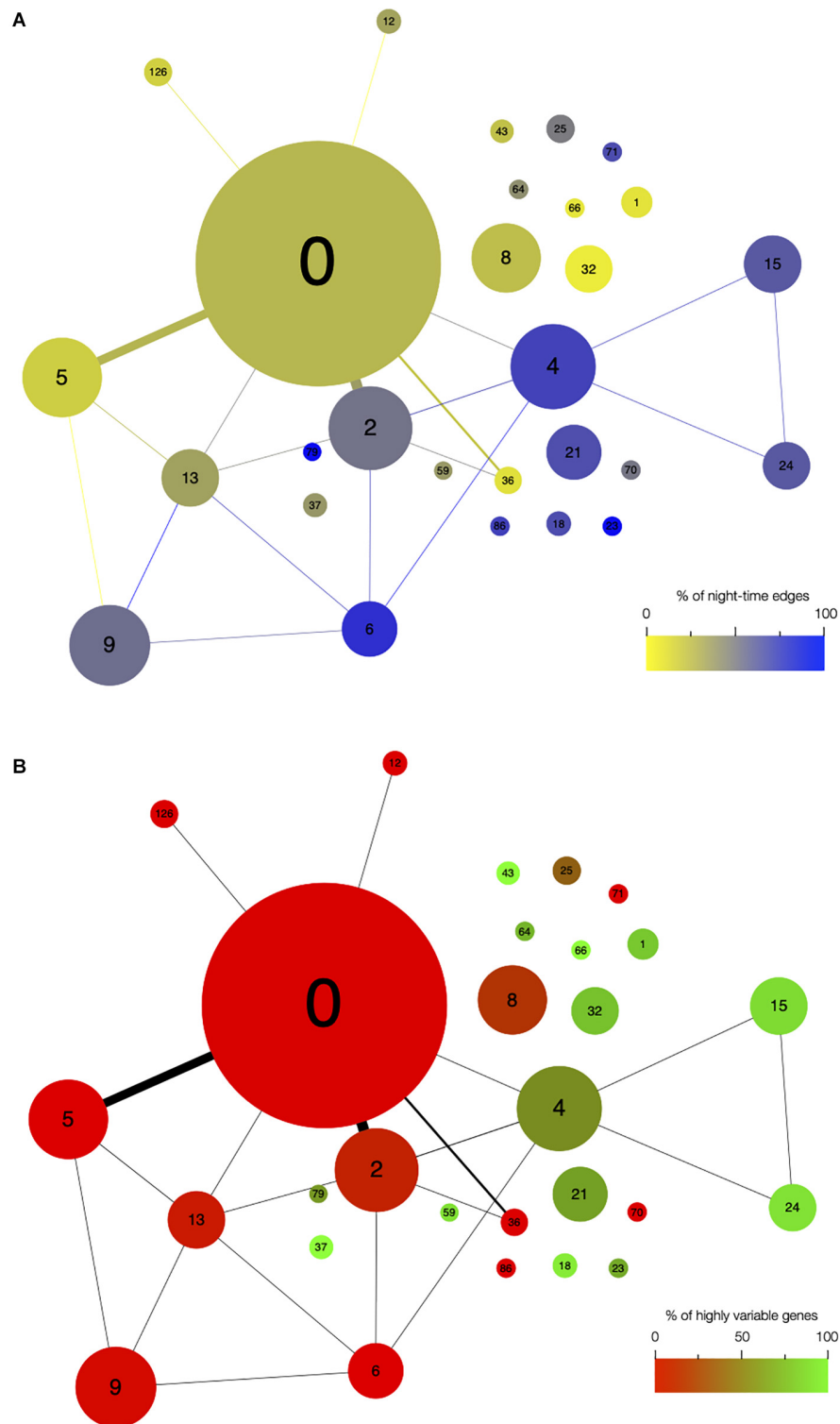


FIGURE 3 | Network architecture is mainly influenced by the time of day when edges are detected and by the presence of highly variable genes in modules. Organization of modules in the network, with the size of circles representing the module size (i.e., number of edges). Number of edges connecting the modules is represented by the thickness of the lines between modules. The number in each module corresponds to the module number. **(A)** Modules are color coded based on the percentage of edges that are detected during the night in each module. Blue modules are composed of a majority of nighttime edges, while yellow modules are mainly composed of daytime edges. **(B)** Modules are color coded based on the percentage of highly variable genes (HVGs) in the modules. Green modules are composed of a majority of HVGs while red modules have a low percentage of HVGs.

TABLE 1 | Number and percentage of HVGs in each module, for modules with at least five genes.

Module	size_modules	nb_HVG	percentage_HVG	Fisher's <i>p</i> -value
37	12	12	100.0	0.0022
43	8	8	100.0	0.0118
66	6	6	100.0	0.0287
18	15	14	93.3	0.0022
24	47	42	89.4	0.0001
59	9	8	88.9	0.0181
15	44	38	86.4	0.0001
1	19	15	78.9	0.0052
32	56	43	76.8	0.0001
64	7	5	71.4	0.1496
23	6	4	66.7	0.2384
21	79	50	63.3	0.0001
79	10	6	60.0	0.1322
4	175	97	55.4	0.0001
25	24	9	37.5	0.4014
8	41	8	19.5	0.482
2	150	17	11.3	0.0002
13	77	6	7.8	0.0009
9	101	3	3.0	0.0001
0	334	3	0.9	0.0001
5	91	0	0.0	0.0001
6	61	0	0.0	0.0001
36	24	0	0.0	0.0047
12	21	0	0.0	0.0127
71	8	0	0.0	0.2142
126	7	0	0.0	0.3578
70	6	0	0.0	0.3512
86	5	0	0.0	0.5915

This is also the case for all modules with at least 50% of genes in the averaged time course network (modules highlighted in blue in **Figure 2**). This result could suggest that the reason why these modules contain many genes also present in the averaged time course network is because their genes have very similar expression patterns throughout the day/night cycle. On the other hand, several modules that only have 15% or less of genes present in the averaged time course network are composed of genes without clear expression patterns during the time course. These results show that additional gene co-expression is identified in the variability network compared to the averaged time course network. Most importantly, using gene expression in single seedlings, co-expression between genes can be detected even in absence of expression patterns throughout the day/night cycle.

Modules Identified in the Variability Network Are Functionally Relevant

In order to define if the modules identified in the variability network are functionally relevant, we performed a Gene Ontology (GO) enrichment analysis. We find that some of the modules have strongly enriched GO (**Supplementary Table S2**).

TABLE 2 | Number and percentage in each module of genes also detected in the averaged time course network, for modules with at least five genes.

Module	Module size	Number genes averaged time course network	Percentage genes averaged time course network	Fisher's <i>p</i> -value
71	8	7	87.5	0.0585
13	77	59	76.6	0.0001
126	7	4	57.1	0.2947
8	41	21	51.2	0.0671
70	6	3	50.0	0.4431
21	79	37	46.8	0.0431
25	24	10	41.7	0.4146
37	12	5	41.7	0.566
9	101	41	40.6	0.1273
23	6	2	33.3	1
5	91	29	31.9	0.8248
1	19	6	31.6	1
12	21	6	28.6	1
64	7	2	28.6	1
36	24	6	25.0	0.8289
24	47	11	23.4	0.5296
4	175	37	21.1	0.0592
0	334	69	20.7	0.0063
86	5	1	20.0	1
2	150	26	17.3	0.0087
6	61	10	16.4	0.0845
32	56	9	16.1	0.0985
15	44	6	13.6	0.0627
79	10	1	10.0	0.4744
18	15	1	6.7	0.1401
59	9	0	0.0	0.1286
43	8	0	0.0	0.2106
66	6	0	0.0	0.3465

For example, module 8 is enriched in multiple GO related to photosynthesis. In particular, 33 genes out of the 41 in this module are members of photosystem I or II, or of the light harvesting complex (**Figure 4A** and **Supplementary Table S3**). Other genes in this module also have functions related to photosynthesis: CURT1A is required for a proper thylakoid morphology (Pribil et al., 2018), while RBCS1A, RBCS3B, and RCA are members of the Rubisco or necessary for Rubisco light activation (Izumi et al., 2012; Carmo-Silva and Salvucci, 2013). Most edges of module 8 are observed at the transition between night and day. This module contains 51% of genes that are also in the averaged time course network, which could be expected as most of the genes have very similar expression patterns throughout the day/night cycle. In particular, all genes of this module present in photosystem I, II, or the light harvesting system have the same expression profile with a peak of expression at dawn and the beginning of the day, while other genes have different expression profiles with a peak of expression at the beginning of the day and another one during the night

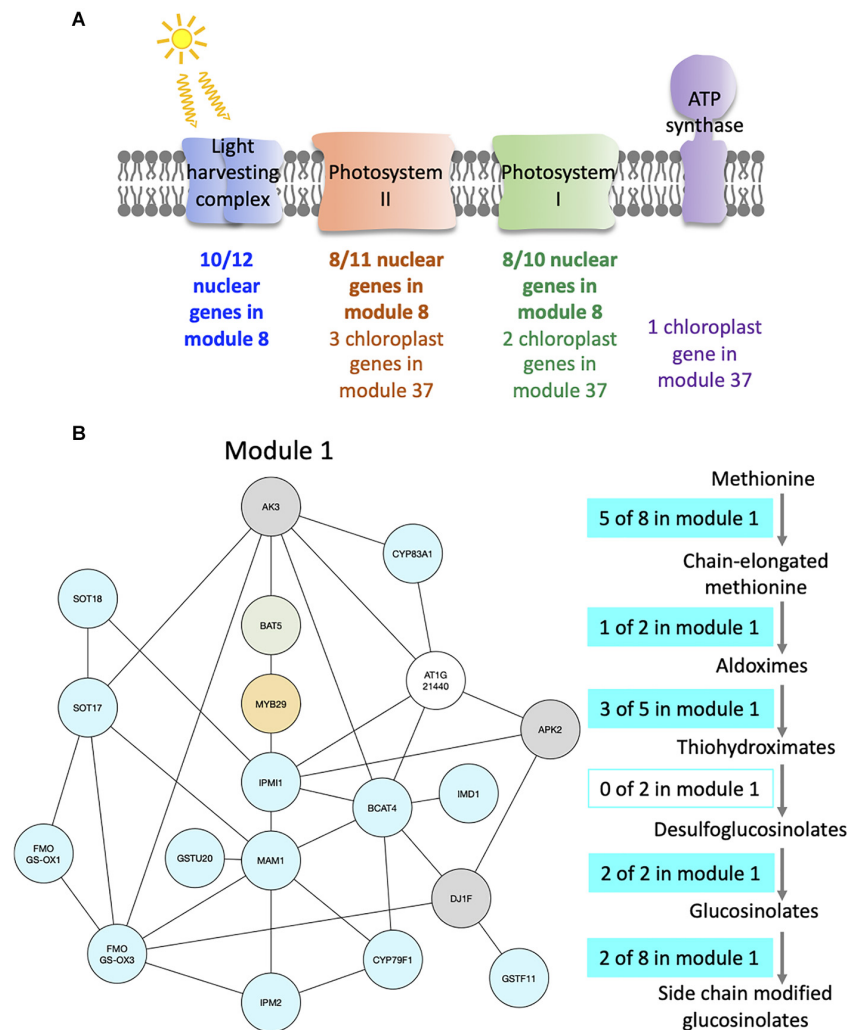


FIGURE 4 | Modules enriched in genes involved in the photosynthesis and the glucosinolate pathway. **(A)** Functional analysis of modules 8 and 37. For each module, the number of genes that are part of photosystem I (green), photosystem II (orange), the light harvesting complex (blue), or the ATP synthase (purple) is indicated. **(B)** Functional analysis of module 1. Genes of the module are color coded depending on their role in the glucosinolate pathway: biosynthesis (turquoise), transport (green), or regulation (orange). Genes previously identified as co-expressed with glucosinolate biosynthesis genes are also indicated (gray). On the right side, the glucosinolate biosynthesis pathway is shown with an indication of the number of genes present in module 1 at each step of the pathway.

(Supplementary Figure S6a). Also, we find that these other genes are at the periphery of module 8 (Supplementary Figure S6b), which highlights that these genes are less well correlated with the dense core of highly correlated photosystem genes in the center of the network. Another module enriched in GO related with photosynthesis is module 37 (Supplementary Table S2), in which nine out of the 12 genes are chloroplast genes, some being present in photosystem I or II, in the Cytochrome b6/f complex, or in the ATP synthase (Figure 4A and Supplementary Table S3). Genes in module 37 are mainly expressed at the beginning of the day. These results suggest that the expression of genes involved in photosynthesis is co-regulated, not only over time but also between plants at a given time.

Module 71 is enriched in GO related to DNA packaging (Supplementary Table S2) and is in fact only composed of

histones, including two variants of H2A, two variants of H2B, and H3.1 (Supplementary Table S3). None of the genes in this module are HVGs, and seven out of eight genes are also present in the averaged time course network.

Module 1 is enriched in GO related to glucosinolate (Supplementary Table S2). We find that 16 out of 19 genes of the module are in the glucosinolate biosynthesis pathway, transporters of glucosinolate, or TFs regulating the pathway (Figure 4B and Supplementary Table S3). All genes in module 1, except one, were previously identified as co-expressed, in a previous study of the glucosinolate pathway (Wisecaver et al., 2017). Among the genes that are not known to be involved in glucosinolate biosynthesis, but are co-expressed with it, AKN2 is regulated by the MYB TF also regulating the glucosinolate pathway (Yatusevich et al., 2010). AKN2 is involved in sulfate

assimilation, which is linked to glucosinolate metabolism (Yatusevich et al., 2010). It thus makes sense that AKN2 is co-expressed with genes of the glucosinolate biosynthesis pathway. Most edges of module 1 are observed during the day, which is when the genes in the module are more expressed. Also, 15 out of the 19 genes of the modules are HVGs.

Module 43 is enriched in GO related to flavonoid metabolism (**Supplementary Table S2**), with six out of eight genes shown to be involved in flavonoid biosynthesis (Saito et al., 2013). Among the other genes, AT4G22870 has not been shown to be involved in the flavonoid pathway, and our result suggests that it might have a role in this pathway. It is a protein of the 2OG and Fe(II)-dependent oxygenase superfamily. We also find that all the genes in module 43 are HVGs. Most edges of the module are observed during the day, and the genes in the module have very similar expression patterns throughout the time course with a peak of expression at the beginning of the day and another one at the end of the day. However, none of the genes in module 43 are also present in the averaged time course network.

Overall, we find that several modules in the variability network are functionally relevant, with modules showing enrichment for functions such as photosynthesis, DNA packaging, and glucosinolate or flavonoid metabolism, even in the absence of genetic and environmental perturbations. Moreover, we could identify a potential role in the enriched pathways for some genes, based on their co-expression with other genes in the same module.

Identification of New Targets for GI, PIF4, and PRR5

To go further in the functional analysis of the modules, we looked for enrichment of targets of transcriptional regulators in the modules. We focused on transcriptional regulators for which ChIP-seq was performed under similar conditions (seedlings grown in day/night cycles) and for which a list of target genes have been previously published (Pfeiffer et al., 2014; Zhang et al., 2014; Liu et al., 2016; Nohales et al., 2019). We define, as targets, genes that are in proximity to regions where transcriptional regulators are binding to the DNA (ChIP-seq peaks), without considering if they are misregulated in the mutant for the transcriptional regulator. This way, we identified an enrichment in modules for targets of the TFs SPL7, PIF4, and PRR5, and of the transcriptional regulator GI (**Supplementary Table S4**). For example, all 41 genes in module 8 are SPL7 targets (**Supplementary Table S4**) (Zhang et al., 2014). This is significantly more compared to the entire network in which 244 genes are SPL7 targets (14%). SPL7 targets have been previously shown to be enriched in multiple GO terms, including photosynthesis (Zhang et al., 2014), in agreement with the predominant role in photosynthesis of genes in this module 8.

Targets for GI have been identified genome-wide, even if it does not bind directly to the DNA. We find that six out of seven genes in module 64 are targets of GI (**Supplementary Table S4** and **Figure 5A**). This is more compared to the entire network in which 394 genes are GI targets (22%). To explore more in detail GI binding at the genes in module 64, we downloaded the ChIP-seq data, mapped it on the *Arabidopsis thaliana* genome, and looked at the ChIP-seq signal for GI at all the seven genes

of module 64 (Nohales et al., 2019). We find a strong signal for the GI ChIP-seq at the promoter of the six genes that were already identified as GI targets (**Figure 5A**). Interestingly, the signal for GI at the seventh gene, AT1G03630, not previously described as a GI target, is equally strong at the promoter of the gene (**Figure 5A**). This result indicates that AT1G03630 is also a target of GI, even if it has not been previously identified as such. AT1G03630, or PORC, encodes for a protein with protochlorophyllide oxidoreductase activity that is NADPH- and light-dependent (Gabruk et al., 2015).

Another TF with enriched targets in some modules is PIF4. We find that three out of the five genes (60%) in module 86 are PIF4 targets (**Supplementary Table S3** and **Figure 5B**), while only 305 genes in the full network are PIF4 targets (17.6%). To explore more in detail PIF4 binding at the genes in module 86, we downloaded the ChIP-seq data, mapped it on the *A. thaliana* genome, and looked at the ChIP-seq signal for PIF4 at all the five genes of module 86 (Pfeiffer et al., 2014). We observe a strong signal for the PIF4 ChIP at the promoters of the three known targets in module 86 (**Figure 5B**). We also see a clear signal for the PIF4 ChIP for the two other genes in module 86, AT4G26542 and AT5G55730, suggesting that they are also targets of PIF4 (**Figure 5B**). AT4G26542 is an anti-sense transcript for AT4G26540. AT5G55730 (FLA1) encodes a fasciclin-like arabinogalactan-protein 1 (Johnson et al., 2011).

Finally, we find an enrichment for PRR5 targets in modules 8 and 21 with, respectively, 78 and 70% of genes in the module that are PRR5 targets (Liu et al., 2016). For comparison, 27% of all genes in the network are PRR5 targets. To explore more in detail PRR5 binding at the genes in module 8, we downloaded the ChIP-seq data, mapped it on the *A. thaliana* genome, and looked at the ChIP-seq signal for PRR5 at all the 41 genes of module 8 (Nakamichi et al., 2012). We find a strong ChIP-seq PRR5 signal at the 32 target genes in module 8, and a similarly strong signal for most of the other nine genes in the module that were not listed as a PRR5 target (**Supplementary Figure S7a**). In order to look for PRR5 targets, and to expand the analysis to other modules, we re-identified peaks for the PRR5 ChIP-seq and looked for PRR5 targets in the modules with a high proportion of already described PRR5 targets (**Supplementary Table S4**). This way, we identified five additional PRR5 targets in module 8, and nine additional PRR5 targets in module 21. When combining the PRR5 targets from both analyses, the total percentage of PRR5 targets is 90% in module 8, and 83% in module 21 (**Supplementary Figure S7b**). These results suggest that most, if not all, genes in modules 8 and 21 are in fact PRR5 targets. PRR5 is a core component of the circadian clock. To test if genes in modules 8 and 21 might be regulated by the circadian clock, we analyzed other TFs involved in the circadian clock: PRR7 (Liu et al., 2016), PRR9 (Liu et al., 2016), TOC1 (Liu et al., 2016), LUX (Ezer et al., 2017), and CCA1 (Kamioka et al., 2016). We found the strongest enrichment in modules 8 and 21 for PRR7 targets (53 and 32%, when 3.9% of all genes are PRR7 targets). On the other hand, none of the modules were strongly enriched in targets for CCA1 or TOC1 (**Supplementary Table S4**). We did not analyze enrichments for targets of PRR9 and LUX as only nine targets of PRR9 and two targets of LUX are present in the network. This result suggests that modules 8 and 21 could

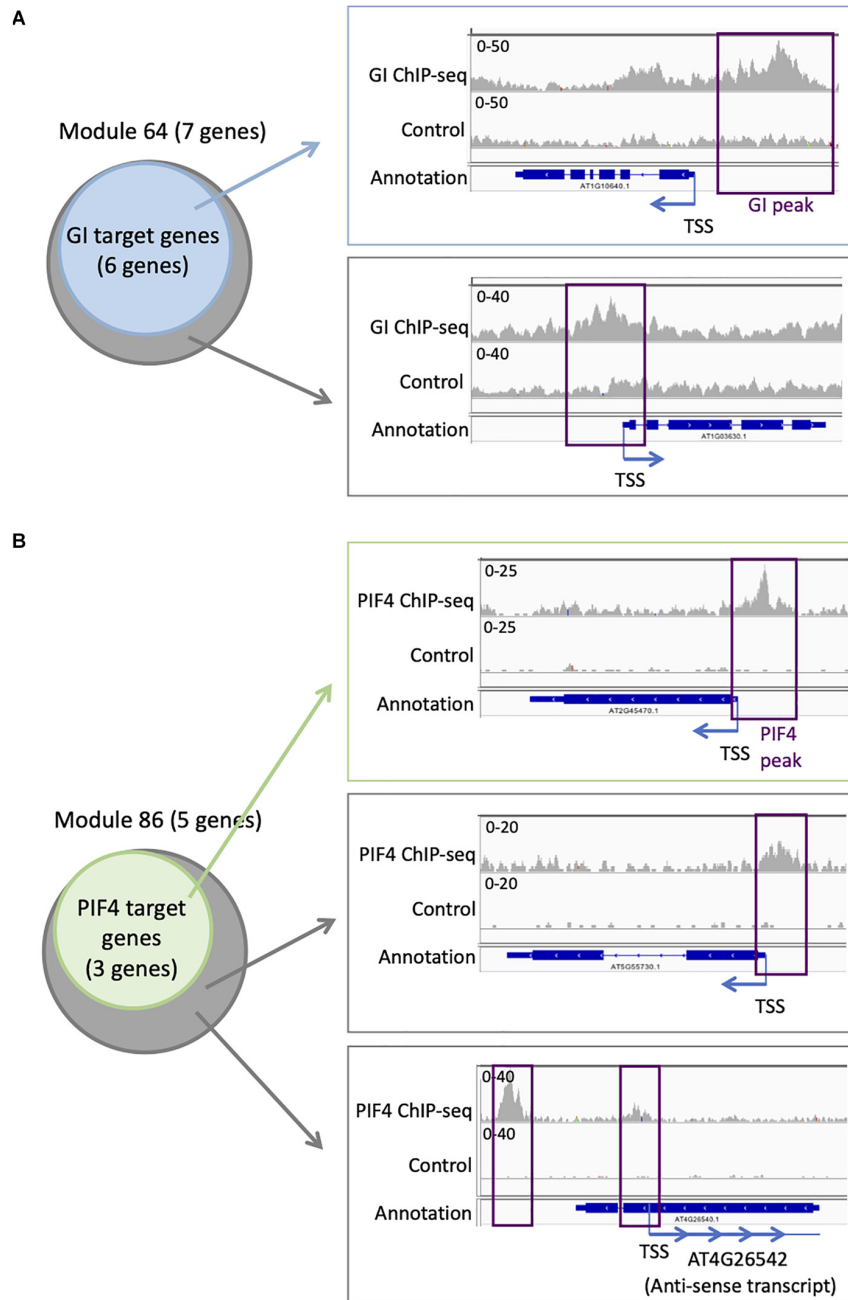


FIGURE 5 | Additional TF targets can be identified using TF target enrichment in modules. **(A)** Analysis of GI transcriptional regulator targets on module 64: six of the seven genes in module 64 are known targets of GI (left). IGVScreenshot showing the signal for the GI ChIP-seq (right) at a known GI target (top) and for the seventh gene in module 64 that is not known as a GI target (bottom). **(B)** Analysis of PIF4 TF targets on module 86: three of the five genes in module 86 are known targets of PIF4 (left). IGVScreenshot showing the signal for the PIF4 ChIP-seq (right) at a known PIF4 target (top) and for the two other genes in module 86 that are not known as a PIF4 target (bottom).

be regulated by several PRR TFs, but probably not by other circadian clock TFs.

Overall, we find that some modules are enriched for targets of several transcriptional regulators and that this enrichment can be used to identify additional targets for the transcriptional regulator in the modules showing enrichment for its targets.

DISCUSSION

In this work, we have analyzed gene co-expression networks inferred using expression data generated in the absence of genetic and environmental perturbations. To do this, we made use of an already published dataset of transcriptomes performed on

single seedlings that were grown in the same environment. We showed that genes do not need a high level of gene expression variability between seedlings to be able to integrate them in the network (**Table 1**). Moreover, we find that modules identified in this network are biologically relevant, as several are strongly enriched in GOs (**Figure 4**) and in targets of transcriptional regulators (**Figure 5**). Based on these enrichments, we speculated that AT4G22870 could also have a role in flavonoid metabolism and identified new targets for the transcriptional regulators GI, PIF4, and PRR5.

We find that it is possible to infer gene co-expression networks using transcriptomes of genetically identical plants of the same age grown in the exact same environment. This is in agreement with previous work, where co-expression networks have been inferred on transcriptomes generated on individual plants and for which genetic and environmental effects have been removed *in silico* (Bhosale et al., 2013). We also find an interesting topology of the network with some modules more connected with one another and that connected modules share similar characteristics in terms of percentage of edges detected during the day or night, and percentage of HVGs (**Figure 3** and **Supplementary Figure S8**). We observe that modules have either a high or low percentage of HVGs, but rarely a mix of HVG with non-HVGs. This suggests that some pathways are more variable than others. We find that, in general, modules with genes involved in the response to the environment are also composed of a high percentage of HVGs. This is the case, for example, for module 37, enriched in photosynthesis (100% of HVGs); module 43, enriched in flavonoid metabolism (100% of HVGs); or module 1, enriched in glucosinolate metabolism (78% of HVGs). Flavonoids are secondary metabolites and have been shown to be involved in many biotic and abiotic responses in plants (Tohge et al., 2017). Glucosinolates are involved in the response to pathogens (Burow et al., 2010). In agreement with our observation, previous work showed that HVGs are usually involved in the response to the environment (Newman et al., 2006; Yin et al., 2009; Hirao et al., 2015; Gasch et al., 2017; Cortijo et al., 2019). In particular, plant-to-plant variability has already been observed for glucosinolates (Mönchgesang et al., 2016), showing that the variability in expression we observe for genes involved in this pathway can lead to differences in glucosinolate content between plants.

Like for Bhosale et al. (2013), we find that the modules of the network identified in the absence of genetic and environmental perturbation are biologically relevant and can be used to speculate new gene function or regulation. We only explored the function for the most obvious GO enrichment in modules as GO can be sparse for some functions and many genes do not have a GO. For example, module 43 is enriched in genes involved in the flavonoid pathway. We speculate that AT4G22870, a member of this module, is also involved in the flavonoid pathway. To support our suggestion, AT4G22870 codes for a protein of the 2OG and Fe(II)-dependent oxygenase superfamily, and three 2OG- and ferrous iron-dependent oxygenases have been previously shown to be involved in flavonoid biosynthesis (Saito et al., 2013). Most importantly, this new potential candidate gene could not have been detected by analyzing the network inferred using day/night

environmental fluctuations as none of the genes in module 43 are also present in the averaged time course network.

We find several modules with enrichment for genes involved in photosynthesis, particularly modules 8 and 37. The main distinction between these two modules is that module 8 is composed of genes from the nuclear genome, while module 37 is mainly composed of genes from the chloroplast genome. Our approach was not designed to specifically identify and separate genes from different organelles, suggesting that genes from the nuclear and chloroplast genomes involved in photosynthesis vary differently in expression between seedlings. Our result is in agreement with the fact that organelle functional modules can be detected in *A. thaliana* (Penga et al., 2016). However, genes that are not from the nuclear genome are usually ignored in network analysis, and it would be of interest to integrate them in the future.

Finally, we identified enrichment for targets of the transcriptional regulators GI, PIF4, and PRR5 in different modules and used this enrichment to highlight new targets. We find that in most cases, when a module is enriched in targets for a transcriptional regulator, the remaining genes of that module are also targets of this regulator. By reanalyzing the ChIP-seq data for PRR5, we could increase the percentage of targets in modules already showing a strong enrichment. This result shows the double interest of combining co-expression networks with ChIP-seq data (Chen et al., 2018; Kulkarni and Vandepoele, 2019). On the one hand, ChIP-seq data add information about the regulation of genes in the co-expression network. On the other hand, the co-expression network is a good way to focus on some of the targets of the TF to better understand their regulation and also to detect extra targets. In the case of PRR5, we find that 90% of the genes in module 8 are targets of this TF. Genes in module 8 are involved in photosynthesis. This is in agreement with the fact that the circadian clock, of which PRR5 is a core member, has been shown to regulate the photosynthesis (Harmer et al., 2000; Schaffer et al., 2001; Dodd et al., 2014).

The functional characterization of the network has been restricted to some modules with obvious GO enrichments and to transcriptional regulators for which ChIP-seq data and lists of targets were available and performed in similar conditions. However, this network, being the first to be performed in the absence of genetic and environmental fluctuation, could bring further information on other pathways we have not explored in this paper. Moreover, our approach could reveal co-regulations that might not be detected using environmental perturbations, as shown by the fact that the variability network provided additional co-expression relationships that were not detected in a network inferred on the same dataset using expression fluctuations caused by the day/night cycle. That is why we encourage readers to look at the modules for their genes or pathway of interest and have developed an interactive website where readers can do so¹.

We show that most genes in the network are not HVGs (**Table 1**), showing that high gene expression variability between seedlings is not needed to be able to detect co-expression. These results indicate that we are not in the presence of random

¹<https://jlggroup.shinyapps.io/VariabilityNetwork/>

fluctuation in expression, or noise, but that pathways are slightly differently regulated in individual seedlings even if the plants are in the same environment. Our approach uses these small differences between seedlings that might be caused by micro-environmental fluctuation or a different state of the plant caused by internal factors. It indicates that plants are very sensitive to minor changes in their environment and that we could harness this sensitivity to better understand gene expression regulation. Phenotypic differences have been observed between genetically identical plants grown in the same environment (Paxman, 1956; Sakai and Shimamoto, 1965; Hall et al., 2007; Forde, 2009; Jimenez-Gomez et al., 2011; Mönchgesang et al., 2016), indicating that the changes in expression of pathways we highlight here might be physiologically relevant (Cortijo and Locke, 2020). It shows that it is not necessary to perform experiments in very different environmental conditions to identify co-expression networks that could be relevant to the studied pathway. This is particularly true for crops growing in outdoor fields, like very recently shown in maize (Felipe Cruz et al., 2020). Strong fluctuations (mutants, over-expressors, and environmental fluctuations) could potentially affect a big part of the transcriptome that could mask some co-expressions of interest showing the usefulness of our approach in some contexts. Our work shows the interest in harnessing gene expression variability between genetically identical individual plants in order to better understand gene regulation in a context where differences between plants are not known and probably very subtle.

MATERIALS AND METHODS

Transcriptome Data

The transcriptomes we used were already published [GSE115583 (Cortijo et al., 2019)] and performed on single seedlings, for a total of 14 seedlings per time point every 2 h over a 24-h cycle. Expression levels and corrected variability levels for all genes were downloaded from <https://jlggroup.shinyapps.io/AraNoisy/>, as these data had already been corrected as previously described (Cortijo et al., 2019).

Network Construction

Variability Network

For each of the 12 time points (0, 2, 4, ... 22 h), we calculated the Spearman correlation between every pair of genes, using their expression profiles across the 14 seedlings (Figure 1A). Using a Benjamini–Hochberg correction with a false discovery rate of 10%, the most significant correlations were selected and further filtered by only considering those for which a significant correlation appeared in four consecutive time points (with one gap allowed, e.g., 8, 10, 14, and 16 h). These correlations form the edges of the variability network. We also calculated a version of the network using a filter that only required three consecutive time points and calculated network modules using the same community detection algorithm. As can be seen in **Supplementary Figure S1**, similar modules with a similar overall connectivity between them are found, which confirms the

robustness of the modules in our original network. All network analysis was carried out using the Python NetworkX and python-louvain libraries.

Averaged Time Course Network

For the averaged time course network, we calculated the mean expression across all seedlings for every time point, generating a time series of average expression for every gene. We again calculated the Spearman correlations for every pair of genes and generated a network by applying the Bonferroni correction as a (highly conservative) significance cutoff. This yielded a network that was similar in size to the variability network. All network analysis was carried out using the Python NetworkX and python-louvain libraries.

Community Detection

The Louvain community detection algorithm (Blondel et al., 2008) was used to identify modules in the networks. This algorithm attempts to maximize the modularity of the network by searching the space of network partitions. Due to the size of the search space, it is unable to find the global maximum. The composition of modules may therefore (as with most community detection algorithms) vary somewhat between runs of the algorithm.

RT-qPCR

Col-0 WT *A. thaliana* seeds were sterilized, stratified for 3 days at 4°C in the dark, and transferred for germination on solid 1/2X Murashige and Skoog (MS) media at 22°C in long days for 24 h. To reduce the level of phenotypic variation between plants, we selected the seeds that were at the same stage of germination with a binocular microscope and transferred them into a new plate containing solid 1/2X MS media. Seedlings were grown in a convion reach-in cabinet at 22°C and 65% humidity, with 12 h of light (170 μ moles) and 12 h of dark. After 7 days of growth, 16 individual seedlings were harvested at ZT6 and at ZT14 into a 96-well plate and flash-frozen in dry ice. All seedlings harvested in a given time point were grown in the same plate. Total RNA was isolated from one seedling. We assessed RNA concentration using Qubit RNA HS assay kit. cDNA synthesis was performed on 700 ng of DNase-treated RNA using the Transcriptor First-Strand cDNA Synthesis Kit. RT-qPCR analysis was performed in the LightCycler 480 instrument using LC480 SYBR green I master, on 0.4 μ l of cDNA in a 10- μ l reaction. Gene expression relative to two control genes (SandF and PP2A) was measured (see **Supplementary Table S5** for the list of primers used for RT-qPCR).

GO Term Enrichment

We used the Ontologizer (Bauer et al., 2008) command line tool with Bonferroni multiple-hypothesis correction to perform GO term enrichment analysis of the network modules. Only the significantly enriched non-redundant GO are shown.

ChIP-seq Data and Analysis

ChIP-seq data were downloaded from GSE45213 for SPL7 (Zhang et al., 2014), from GSE129865 for GI

(Nohales et al., 2019), from GSE43286 for PIF4 (Pfeiffer et al., 2014), and from GSE36361 for PRR5 (Liu et al., 2016).

Reads were aligned to the TAIR10 genome using Bowtie2 (Langmead et al., 2009) and Picard tools were used to remove potential optical duplicates². Peak calling was performed using MASC2 (Zhang et al., 2008), with the corresponding INPUT used as a reference. The Integrative Genomics Viewer (IGV; Robinson et al., 2011) was used to show snapshots of ChIP-seq signal around targets.

DATA AVAILABILITY STATEMENT

Publicly available datasets were analyzed in this study. This data can be found on Gene Expression Omnibus GSE115583: <https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE115583>.

AUTHOR CONTRIBUTIONS

SC conceived the project and interpreted the data. SC, JL, and SA designed the project and wrote the article, with SC writing the first draft. SA inferred the networks. SC and SA performed downstream analyses of the network. MB performed the RT-qPCR validation. All authors contributed to the article and approved the submitted version.

FUNDING

JL was funded by Gatsby Foundation (<https://www.gatsby.org.uk/>, Grant No: GAT3272/GLC) and ERC (<https://erc.europa.eu/>, Grant No. 338060). SA was funded by Gatsby Foundation (<https://www.gatsby.org.uk/>, Grant No. PTAG/021).

ACKNOWLEDGMENTS

This manuscript has been released as a preprint at <https://www.biorxiv.org/content/10.1101/2020.06.15.152314v1> (Cortijo et al., 2020).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fpls.2020.599464/full#supplementary-material>

Supplementary Figure 1 | Expression in seedlings of genes in module 1, from the RNA-seq data, with one line per gene. Expression is mean normalized for each gene.

²<https://github.com/broadinstitute/picard>

Supplementary Figure 2 | Comparison of edges in modules detected in the networks containing edges present in three or four consecutive time points. Modules detected in the network based on edges present in at least three consecutive time points are shown in blue. Modules detected in the network based on edges present in at least three consecutive time points are shown in red. For the later, the percentage of edges of the modules that are also detected in the blue modules is indicated.

Supplementary Figure 3 | Number of edges in the final network that are detected in each time point, for every module containing at least five genes.

Supplementary Figure 4 | (A) Correlation in expression between seedlings for genes of the module 1 and module 21, for the RNA-seq experiment. AT5G07690 at ZT14 was removed as it is not expressed. (B) Correlation in expression between seedlings for genes of the module 1 and module 21, based on an RT-qPCR replicate of the RNA-seq experiment. Sixteen seedlings were harvested at ZT6 and at ZT14. (C) Normalized expression level in the fourteen seedlings for the genes of the module 21, from the RNA-seq data. Expression level for AT4G13250 is shown as the x axis while expression for the other genes of the module is shown on the y axis. Expression is mean normalized for each gene.

Supplementary Figure 5 | (A) Inter-individual gene expression variability profiles throughout the time course for genes in each module with five genes or more. Each line represents the corrected variability level for one gene: $\text{corrected } CV^2 = [\log_2(CV^2/\text{trend})]$, with $CV^2 = \text{variance}/(\text{average}^2)$ (see Cortijo et al., 2019). Modules are ordered by the percentage of HVG (high to low). Modules highlighted in blue contain 75% or more of HVGs. Modules highlighted in red contain 10% or less of HVGs. (B) Heatmap of normalized gene expression for genes in modules 2, 6 (less than 15% HVG), module 4 (55% of HVG), and modules 15 and 24 (more than 85% of HVG). Expression is shown in single seedlings from the time point ZT20. Expression is mean normalized: expression in a seedling/averaged expression in all seedlings. The left color coded bar indicates the module of each gene.

Supplementary Figure 6 | (A) Expression profiles throughout the time course for genes in module 8. Each line represents the normalized expression (z-score) for one gene. Genes of the photosystem I, II, or the light harvesting system are in blue. (B) All edges and nodes of module 8. Genes of the photosystem I, II, or the light harvesting system are in blue.

Supplementary Figure 7 | (A) Analysis of PRR5 TF targets in the module 8. 32 of the 41 genes in the module 8 are known targets of PRR5 (left). IGV screenshot showing the signal for the PRR5 ChIP-seq (right) at a known PRR5 target (top) and for a gene in the module 8 that is not known as a PRR5 target (bottom). (B) Comparison of published (blue) and realized (gray) PRR5 targets in modules 8 (left) and 21 (right).

Supplementary Figure 8 | Organization of modules in the network, with the size of circles representing the module size (i.e., number of edges). Number of edges connecting the modules are represented by the thickness of the lines between modules. (A) Most enriched GOs are written in each module. (B) Modules are color coded based on the percentage of edges in the averaged time course network. Dark blue modules have a high percentage of genes in the averaged time course network while light blue modules have a low percentage of genes in the averaged time course network.

Supplementary Table 1 | List of genes in each module.

Supplementary Table 2 | GO enriched (corrected p -value < 0.5) in each module.

Supplementary Table 3 | Function of genes in modules 1, 43, and 71.

Supplementary Table 4 | Number and percentage in the modules of targets for the TFs SPL7, Gl, PIF4, and PRR5.

Supplementary Table 5 | List of primers used for RT-qPCR.

REFERENCES

- Amrine, K. C. H., Blanco-Ulate, B., and Cantu, D. (2015). Discovery of core biotic stress responsive genes in *Arabidopsis* by weighted gene co-expression network analysis. *PLoS One* 10:e0118731. doi: 10.1371/journal.pone.0118731
- Aoki, K., Ogata, Y., and Shibata, D. (2007). Approaches for extracting practical information from gene co-expression networks in plant biology. *Plant Cell Physiol.* 48, 381–390. doi: 10.1093/pcp/pcm013
- Bauer, S., Grossmann, S., Vingron, M., and Robinson, P. N. (2008). Ontologizer 2.0—a multifunctional tool for GO term enrichment analysis and data exploration. *Bioinformatics* 24, 1650–1651. doi: 10.1093/bioinformatics/btn250
- Bhosale, R., Jewell, J. B., Hollunder, J., Koo, A. J. K., Vuylsteke, M., Michoel, T., et al. (2013). Predicting gene function from uncontrolled expression variation among individual wild-type *Arabidopsis* plants. *Plant Cell* 25, 2865–2877. doi: 10.1105/tpc.113.112268
- Blondel, V. D., Guillaume, J.-L., Lambiotte, R., and Lefebvre, E. (2008). Fast unfolding of communities in large networks. *J. Stat. Mech.* 2008:10008. doi: 10.1088/1742-5468/2008/10/P10008
- Burrow, M., Halkier, B. A., and Kliebenstein, D. J. (2010). Regulatory networks of glucosinolates shape *Arabidopsis thaliana* fitness. *Curr. Opin. Plant Biol.* 13, 348–353. doi: 10.1016/j.pbi.2010.02.002
- Carmo-Silva, A. E., and Salvucci, M. E. (2013). The regulatory properties of Rubisco activase differ among species and affect photosynthetic induction during light transitions. *Plant Physiol.* 161, 1645–1655. doi: 10.1104/pp.112.213348
- Chen, D., Yan, W., Fu, L.-Y., and Kaufmann, K. (2018). Architecture of gene regulatory networks controlling flower development in *Arabidopsis thaliana*. *Nat. Commun.* 9:4534. doi: 10.1038/s41467-018-06772-3
- Contreras-López, O., Moyano, T. C., Soto, D. C., and Gutiérrez, R. A. (2018). Step-by-step construction of gene co-expression networks from high-throughput *Arabidopsis* RNA sequencing data. *Methods Mol. Biol.* 1761, 275–301. doi: 10.1007/978-1-4939-7747-5_21
- Cortijo, S., Aydin, Z., Ahnert, S., and Locke, J. C. (2019). Widespread inter-individual gene expression variability in *Arabidopsis thaliana*. *Mol. Syst. Biol.* 15:e8591. doi: 10.15252/msb.20188591
- Cortijo, S., Bhattarai, M., Locke, J. C. W., and Ahnert, S. E. (2020). Co-expression networks from gene expression variability between genetically identical seedlings can reveal novel regulatory relationships. *BioRxiv* [Preprint]. doi: 10.1101/2020.06.15.152314
- Cortijo, S., and Locke, J. C. W. (2020). Does gene expression noise play a functional role in plants? *Trends Plant Sci.* 25, 1041–1051. doi: 10.1016/j.tplants.2020.04.017
- de Luis Balaguer, M. A., Fisher, A. P., Clark, N. M., Fernandez-Espinoza, M. G., Möller, B. K., Weijers, D., et al. (2017). Predicting gene regulatory networks by combining spatial and temporal gene expression data in *Arabidopsis* root stem cells. *Proc. Natl. Acad. Sci. U.S.A.* 114, E7632–E7640. doi: 10.1073/pnas.1707566114
- Des Marais, D. L., Guerrero, R. F., Lasky, J. R., and Scarpino, S. V. (2017). Topological features of a gene co-expression network predict patterns of natural diversity in environmental response. *Proc. Biol. Sci.* 284:914. doi: 10.1098/rspb.2017.0914
- Dodd, A. N., Kusakina, J., Hall, A., Gould, P. D., and Hanaoka, M. (2014). The circadian regulation of photosynthesis. *Photosyn. Res.* 119, 181–190. doi: 10.1007/s1120-013-9811-8
- Ezer, D., Jung, J.-H., Lan, H., Biswas, S., Gregoire, L., Box, M. S., et al. (2017). The evening complex coordinates environmental and endogenous signals in *Arabidopsis*. *Nat. Plants* 3:17087. doi: 10.1038/nplants.2017.87
- Felipe Cruz, D., De Meyer, S., Ampe, J., Sprenger, H., Herman, D., Van Hautegeem, T., et al. (2020). Using single-plant-omics in the field to link maize genes to functions and phenotypes. *BioRxiv* [Preprint]. doi: 10.1101/2020.04.06.027300
- Forde, B. G. (2009). Is it good noise? The role of developmental instability in the shaping of a root system. *J. Exp. Bot.* 60, 3989–4002. doi: 10.1093/jxb/erp265
- Gabruk, M., Stecka, A., Strzalka, W., Kruk, J., Strzalka, K., and Mysliwa-Kurdziel, B. (2015). Photoactive protochlorophyllide-enzyme complexes reconstituted with PORA, PORB and PORC proteins of *A. thaliana*: fluorescence and catalytic properties. *PLoS One* 10:e0116990. doi: 10.1371/journal.pone.0116990
- Gasch, A. P., Yu, F. B., Hose, J., Escalante, L. E., Place, M., Bacher, R., et al. (2017). Single-cell RNA sequencing reveals intrinsic and extrinsic regulatory heterogeneity in yeast responding to stress. *PLoS Biol.* 15:e2004050. doi: 10.1371/journal.pbio.2004050
- Hall, M. C., Dworkin, I., Ungerer, M. C., and Purugganan, M. (2007). Genetics of microenvironmental canalization in *Arabidopsis thaliana*. *Proc. Natl. Acad. Sci. U.S.A.* 104, 13717–13722. doi: 10.1073/pnas.0701936104
- Harmer, S. L., Hogenesch, J. B., Straume, M., Chang, H. S., Han, B., Zhu, T., et al. (2000). Orchestrated transcription of key pathways in *Arabidopsis* by the circadian clock. *Science* 290, 2110–2113. doi: 10.1126/science.290.5499.2110
- He, F., Zhou, Y., and Zhang, Z. (2010). Deciphering the *Arabidopsis* floral transition process by integrating a protein-protein interaction network and gene expression data. *Plant Physiol.* 153, 1492–1505. doi: 10.1104/pp.110.153650
- Hirao, K., Nagano, A. J., and Awazu, A. (2015). Noise-plasticity correlations of gene expression in the multicellular organism *Arabidopsis thaliana*. *J. Theor. Biol.* 387, 13–22. doi: 10.1016/j.jtbi.2015.09.017
- Ichihashi, Y., Aguilar-Martínez, J. A., Farhi, M., Chitwood, D. H., Kumar, R., Millon, L. V., et al. (2014). Evolutionary developmental transcriptomics reveals a gene network module regulating interspecific diversity in plant leaf shape. *Proc. Natl. Acad. Sci. U.S.A.* 111, E2616–E2621. doi: 10.1073/pnas.1402835111
- Izumi, M., Tsunoda, H., Suzuki, Y., Makino, A., and Ishida, H. (2012). RBCS1A and RBCS3B, two major members within the *Arabidopsis* RBCS multigene family, function to yield sufficient Rubisco content for leaf photosynthetic capacity. *J. Exp. Bot.* 63, 2159–2170. doi: 10.1093/jxb/err434
- Jimenez-Gomez, J. M., Corwin, J. A., Joseph, B., Maloof, J. N., and Kliebenstein, D. J. (2011). Genomic analysis of QTLs and genes altering natural variation in stochastic noise. *PLoS Genet.* 7:e1002295. doi: 10.1371/journal.pgen.1002295
- Johnson, K. L., Kibble, N. A. J., Bacic, A., and Schultz, C. J. (2011). A fasciclin-like arabinogalactan-protein (FLA) mutant of *Arabidopsis thaliana*, fla1, shows defects in shoot regeneration. *PLoS One* 6:e25154. doi: 10.1371/journal.pone.0025154
- Kamioka, M., Takao, S., Suzuki, T., Taki, K., Higashiyama, T., Kinoshita, T., et al. (2016). Direct repression of evening genes by CIRCADIEN CLOCK-ASSOCIATED1 in the *Arabidopsis* circadian clock. *Plant Cell* 28, 696–711. doi: 10.1105/tpc.15.00737
- Kulkarni, S. R., and Vandepoele, K. (2019). Inference of plant gene regulatory networks using data-driven methods: a practical overview. *Biochim. Biophys. Acta Gene Regul. Mech.* 2019:194447. doi: 10.1016/j.bbagr.2019.194447
- Langmead, B., Trapnell, C., Pop, M., and Salzberg, S. L. (2009). Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol.* 10:R25. doi: 10.1186/gb-2009-10-3-r25
- Liu, T. L., Newton, L., Liu, M.-J., Shiu, S.-H., and Farré, E. M. (2016). A G-box-like motif is necessary for transcriptional regulation by circadian pseudo-response regulators in *Arabidopsis*. *Plant Physiol.* 170, 528–539. doi: 10.1104/pp.15.01562
- Liu, W., Lin, L., Zhang, Z., Liu, S., Gao, K., Lv, Y., et al. (2019). Gene co-expression network analysis identifies trait-related modules in *Arabidopsis thaliana*. *Planta* 249, 1487–1501. doi: 10.1007/s00425-019-03102-9
- Ma, S., Shah, S., Bohnert, H. J., Snyder, M., and Dinesh-Kumar, S. P. (2013). Incorporating motif analysis into gene co-expression networks reveals novel modular expression pattern and new signaling pathways. *PLoS Genet.* 9:e1003840. doi: 10.1371/journal.pgen.1003840
- Mantsoki, A., Devailly, G., and Joshi, A. (2016). Gene expression variability in mammalian embryonic stem cells using single cell RNA-seq data. *Comput. Biol. Chem.* 63, 52–61. doi: 10.1016/j.compbiolchem.2016.02.004
- Mao, L., Van Hemert, J. L., Dash, S., and Dickerson, J. A. (2009). *Arabidopsis* gene co-expression network and its functional modules. *BMC Bioinformatics* 10:346. doi: 10.1186/1471-2105-10-346
- Mönchgesang, S., Strehmel, N., Trutschel, D., Westphal, L., Neumann, S., and Scheel, D. (2016). Plant-to-plant variability in root metabolite profiles of 19 *Arabidopsis thaliana* accessions is substance-class-dependent. *Int. J. Mol. Sci.* 17:1565. doi: 10.3390/ijms17091565
- Nakamichi, N., Kiba, T., Kamioka, M., Suzuki, T., Yamashino, T., Higashiyama, T., et al. (2012). Transcriptional repressor PRR5 directly regulates clock-output pathways. *Proc. Natl. Acad. Sci. U.S.A.* 109, 17123–17128. doi: 10.1073/pnas.1205156109
- Newman, J. R. S., Ghaemmaghami, S., Ihmels, J., Breslow, D. K., Noble, M., DeRisi, J. L., et al. (2006). Single-cell proteomic analysis of *S. cerevisiae* reveals the architecture of biological noise. *Nature* 441, 840–846. doi: 10.1038/nature04785

- Nohales, M. A., Liu, W., Duffy, T., Nozue, K., Sawa, M., Pruneda-Paz, J. L., et al. (2019). Multi-level modulation of light signaling by GIGANTEA regulates both the output and pace of the circadian clock. *Dev. Cell* 49, 840–851.e8. doi: 10.1016/j.devcel.2019.04.030
- Paxman, G. J. (1956). Differentiation and Stability in the development of nicotiana rustica. *Ann. Bot.* 20, 331–347. doi: 10.1093/oxfordjournals.aob.a083526
- Penga, J., Wang, T., Huc, J., Wang, Y., and Chen, J. (2016). Constructing networks of organelle functional modules in *Arabidopsis*. *Curr. Genomics* 17, 427–438. doi: 10.2174/1389202917666160726151048
- Pfeiffer, A., Shi, H., Tepperman, J. M., Zhang, Y., and Quail, P. H. (2014). Combinatorial complexity in a transcriptionally centered signaling hub in *Arabidopsis*. *Mol. Plant* 7, 1598–1618. doi: 10.1093/mp/ssu087
- Prasch, C. M., and Sonnewald, U. (2013). Simultaneous application of heat, drought, and virus to *Arabidopsis* plants reveals significant shifts in signaling networks. *Plant Physiol.* 162, 1849–1866. doi: 10.1104/pp.113.221044
- Pribil, M., Sandoval-Ibáñez, O., Xu, W., Sharma, A., Labs, M., Liu, Q., et al. (2018). Fine-tuning of photosynthesis requires CURVATURE THYLAKOID1-mediated thylakoid plasticity. *Plant Physiol.* 176, 2351–2364. doi: 10.1104/pp.17.00863
- Rao, X., and Dixon, R. A. (2019). Co-expression networks for plant biology: why and how. *Acta Biochim. Biophys. Sin. (Shanghai)* 51, 981–988. doi: 10.1093/abbs/gmz080
- Robinson, J. T., Thorvaldsdóttir, H., Winckler, W., Guttman, M., Lander, E. S., Getz, G., et al. (2011). Integrative genomics viewer. *Nat. Biotechnol.* 29, 24–26. doi: 10.1038/nbt.1754
- Ruan, J., Dean, A. K., and Zhang, W. (2010). A general co-expression network-based approach to gene expression analysis: comparison and applications. *BMC Syst. Biol.* 4.8. doi: 10.1186/1752-0509-4-8
- Saito, K., Yonekura-Sakakibara, K., Nakabayashi, R., Higashi, Y., Yamazaki, M., Tohge, T., et al. (2013). The flavonoid biosynthetic pathway in *Arabidopsis*: structural and genetic diversity. *Plant Physiol. Biochem.* 72, 21–34. doi: 10.1016/j.plaphy.2013.02.001
- Sakai, K., and Shimamoto, Y. (1965). Developmental instability in leaves and flowers of NICOTIANA TABACUM. *Genetics* 51, 801–813.
- Schaffer, R., Landgraf, J., Accerbi, M., Simon, V., Larson, M., and Wisman, E. (2001). Microarray analysis of diurnal and circadian-regulated genes in *Arabidopsis*. *Plant Cell* 13, 113–123. doi: 10.1105/tpc.13.1.113
- Serin, E. A. R., Nijveen, H., Hilhorst, H. W. M., and Ligterink, W. (2016). Learning from co-expression networks: possibilities and challenges. *Front. Plant Sci.* 7:444. doi: 10.3389/fpls.2016.00444
- Shaik, R., and Ramakrishna, W. (2013). Genes and co-expression modules common to drought and bacterial stress responses in *Arabidopsis* and rice. *PLoS One* 8:e77261. doi: 10.1371/journal.pone.0077261
- Sharma, R., Singh, G., Bhattacharya, S., and Singh, A. (2018). Comparative transcriptome meta-analysis of *Arabidopsis thaliana* under drought and cold stress. *PLoS One* 13:e0203266. doi: 10.1371/journal.pone.0203266
- Shen, X., Pettersson, M., Rönnegård, L., and Carlborg, Ö. (2012). Inheritance beyond plain heritability: variance-controlling genes in *Arabidopsis thaliana*. *PLoS Genet.* 8:e1002839. doi: 10.1371/journal.pgen.1002839
- Silva, A. T., Ribone, P. A., Chan, R. L., Ligterink, W., and Hilhorst, H. W. M. (2016). A predictive coexpression network identifies novel genes controlling the seed-to-seedling phase transition in *Arabidopsis thaliana*. *Plant Physiol.* 170, 2218–2231. doi: 10.1104/pp.15.01704
- Stuart, J. M., Segal, E., Koller, D., and Kim, S. K. (2003). A gene-coexpression network for global discovery of conserved genetic modules. *Science* 302, 249–255. doi: 10.1126/science.1087447
- Tohge, T., de Souza, L. P., and Fernie, A. R. (2017). Current understanding of the pathways of flavonoid biosynthesis in model and crop plants. *J. Exp. Bot.* 68, 4013–4028. doi: 10.1093/jxb/erx177
- Usadel, B., Obayashi, T., Mutwil, M., Giorgi, F. M., Bassel, G. W., Tanimoto, M., et al. (2009). Co-expression tools for plant biology: opportunities for hypothesis generation and caveats. *Plant Cell Environ.* 32, 1633–1651. doi: 10.1111/j.1365-3040.2009.02040.x
- Vandepoele, K., Quimbaya, M., Casneuf, T., De Veylder, L., and Van de Peer, Y. (2009). Unraveling transcriptional control in *Arabidopsis* using cis-regulatory elements and coexpression networks. *Plant Physiol.* 150, 535–546. doi: 10.1104/pp.109.136028
- Wisecaver, J. H., Borowsky, A. T., Tzin, V., Jander, G., Kliebenstein, D. J., and Rokas, A. (2017). A global coexpression network approach for connecting genes to specialized metabolic pathways in plants. *Plant Cell* 29, 944–959. doi: 10.1105/tpc.17.00009
- Xie, W., Huang, J., Liu, Y., Rao, J., Luo, D., and He, M. (2015). Exploring potential new floral organ morphogenesis genes of *Arabidopsis thaliana* using systems biology approach. *Front. Plant Sci.* 6:829. doi: 10.3389/fpls.2015.00829
- Yatusevich, R., Mugford, S. G., Matthewman, C., Gigolashvili, T., Frerigmann, H., Delaney, S., et al. (2010). Genes of primary sulfate assimilation are part of the glucosinolate biosynthetic network in *Arabidopsis thaliana*. *Plant J.* 62, 1–11. doi: 10.1111/j.1365-313X.2009.04118.x
- Yin, S., Wang, P., Deng, W., Zheng, H., Hu, L., Hurst, L. D., et al. (2009). Dosage compensation on the active X chromosome minimizes transcriptional noise of X-linked genes in mammals. *Genome Biol.* 10:R74. doi: 10.1186/gb-2009-10-7-r74
- Zhang, H., Zhao, X., Li, J., Cai, H., Deng, X. W., and Li, L. (2014). MicroRNA408 is critical for the *HY5-SPL7* gene network that mediates the coordinated response to light and copper. *Plant Cell* 26, 4933–4953. doi: 10.1105/tpc.114.127340
- Zhang, Y., Liu, T., Meyer, C. A., Eeckhoutte, J., Johnson, D. S., Bernstein, B. E., et al. (2008). Model-based analysis of ChIP-Seq (MACS). *Genome Biol.* 9:R137. doi: 10.1186/gb-2008-9-9-r137
- Zheng, X., Liu, T., Yang, Z., and Wang, J. (2011). Large cliques in *Arabidopsis* gene coexpression network and motif discovery. *J. Plant Physiol.* 168, 611–618. doi: 10.1016/j.jplph.2010.09.010

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

Copyright © 2020 Cortijo, Bhattarai, Locke and Ahnert. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.