Co-expression networks from gene expression variability between genetically identical seedlings can reveal novel regulatory relationships

Short title: Co-expression networks from noisy gene expression

Cortijo S[1,2], Bhattarai M[1], Locke J C W[1]*, Ahnert S E[1,3]*

[1] The Sainsbury Laboratory, University of Cambridge, Bateman Street, Cambridge CB2 1LR, UK

[2] BPMP, Univ Montpellier, CNRS, INRAE, Institut Agro, Montpellier, France

[3] Theory of Condensed Matter, Cavendish Laboratory, University of Cambridge, JJ Thomson Avenue, Cambridge CB3 0HE, UK

* Corresponding authors:

James Locke: Tel: +44 1223 761110; E-mail: james.locke@slcu.cam.ac.uk

Sebastian Ahnert: Tel: +44 1223 761148 ; E-mail: sebastian.ahnert@slcu.cam.ac.uk

Abstract

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 in the same conditions and for the same amount of time. Twelve time points were used to cover the 24h 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 transcription factors 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.

Author summary

Plant development and response to changes in the environment are strongly regulated at the level of gene expression. That is why understanding how gene expression is regulated is key, and transcriptome approaches have allowed the analysis of transcription for all genes of the genome. Extracting useful information from the high amount of data generated by transcriptomes is a challenge, and gene co-expression networks are a powerful tool to do this. The principle is to find genes that co-vary in expression in different conditions and to pair them together. Communities of genes that are more closely linked are then identified and this is the starting point to look for their implication in the same pathway. Co-expression networks have been used to identify new regulation and function of genes involved in plant development and their response to the environment. They were constructed using transcriptomes generated on plants experiencing genetic or environmental perturbation. We show that co-expression networks can in fact be constructed in the absence of genetic and environmental perturbation. Our work will allow further exploration of gene co-regulation in contexts with subtle differences between plants, which could be closer to what individual plants in a population might face in the wild.

Introduction

Understanding how transcriptomes are regulated is key to shedding 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 [1,2]. In gene co-expression networks, genes that co-vary in expression in different conditions are detected and paired together [3–5]. 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 [6]. 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) [7–10]. 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 transcription factors involved in development [9,11,12], in metabolic pathways [13] and in response to biotic and abiotic stresses [8,14–17]. 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 [18,19].

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 transcription factors (TF) that regulate the expression of other genes in the network, additional types of data should be used or integrated [20]. This additional data can be for example ChIP-seq [21,22] that provides the list of targets of a given transcription factor, protein-protein interaction [21,23], as well as the presence of TF binding motifs in the promoter of genes [24,25]. 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 analyse gene regulation in a biologically context that is more relevant, more subtle changes in expression might be prefered. This could be achieved by using milder genetic or environmental perturbations. Another option would be to analyse changes in expression that occur in the absence of any genetic or environmental perturbation [26]. 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 [27–31]. 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 [32]. Moreover, gene co-expression networks have been inferred using transcriptomes of individual plant leaves, after removing in silico the genotype, environment and genotype x environment effects on gene expression [26]. The modules identified in this network were functionally relevant and this study allowed the identification of a new regulator of the jasmonate pathway [26]. 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 and colleagues was to remove *in silico* the genotype, environment and genotype x 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 [28]. In this dataset, multiple genetically identical seedlings had been harvested at several time points during a day/night cycle. Differences in expression between seedlings were previously observed for many genes in each time point of this dataset. In particular, 8.7% of the genes in this dataset have been identified as Highly Variable Genes (HVG), 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 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 TFs 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 fourteen seedlings per time point, for twelve time points spanning a 24 hours day/night cycle [28]. 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 ([3–5] Fig 1a and 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 then selected edges of the network that are detected in at least four consecutive time points, with one gap allowed (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 (Fig 1b). We then used the Louvain community detection algorithm in order to identify modules of genes that are densely connected in the network [33]. In total, we identified 153 modules (Table S1), containing between 2 and

334 genes, with most of the modules only composed of two genes (Fig 1c). To test the robustness of the variability network, we also selected edges that are in at least 3 consecutive time points and compared the detected modules in both networks (Fig 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 4 consecutive time points are smaller. In somes cases, several modules of the network based on 4 consecutive time points correspond to a single module in the network based on 3 consecutive time points and these smaller modules have differences in several features (Fig S2). That is why we decided to focus our analysis on the network obtained when selecting edges present in 4 consecutive time points, and in particular for the 28 modules of this network containing 5 genes or more.

Fig 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 fourteen seedlings per time point, with twelve time points spanning a day/night cycle over 24 hours. In each time point, genes with correlated expression profiles in the fourteen seedlings were identified. The co-expression network was inferred based on pairs of genes correlated for 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: the module 1 in which most edges are detected during day time, the module 21 in which most edges are detected during the night time, the module 8 in which most edges are detected at the transition between night and day and the module 12 in which most edges are detected at the transition between day and night.

First we analysed the number of edges at each time point throughout the time course for each module (Fig 1d and 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 is 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 with 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, Fig 2). Most of the edges in module 1 are observed during the day (Fig 1d) and we were able to confirm co-expression of genes in this module by doing a 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) (Fig S4). On the other hand, most edges in module 21 are observed during the night (Fig 1d). We also find in a replicate RT-qPCR experiment that genes of module 21 were more correlated during the night than the day (Fig 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 (Fig 5a). 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.

Fig 2. Comparison of the variability network and the averaged time course network

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 normalised 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.

Fig 5. Network architecture is mainly influenced by the time of day when edges are detected, and by the presence of highly variable genes

Organisation 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. The number in each

250 module corresponds to the module number. (A) Modules are color coded based on
251 the percentage of edges that are detected during the night in each module. Blue
252 modules are composed of a majority of night-time edges, while yellow modules are
253 mainly composed of day-time edges. (B) Modules are color coded based on the
254 percentage of highly variable genes (HVG) in the modules. Green modules are
255 composed of a majority of HVG while red modules have a low percentage of HVG.
256
257 Since high gene expression variability between genetically identical plants was
258 previously observed in the transcriptome dataset we used to infer the variability
259 network [28], we tested if the network is enriched in highly variable genes (HVGs). We
260 find a total of 477 HVGs in the network, that is 27.6% of all genes in the variability
261 network. This is higher than the 8.7% of HVGs that were detected in the full
262 transcriptome dataset [28]. This result suggests that most of the genes in the variability
263 network do not have to display a high level of gene expression variability to be able to
264 detect co-expression between individual seedlings. We find that most modules are
265 either strongly enriched in HVGs, or strongly depleted in HVGs, with only a few
266 modules containing around 27% of HVGs (Table 1, Fig S5a). Modules 37, 43 and 66
267 for example are only composed of HVGs, while a total of 8 modules do not have a
268 single HVG. This result suggests that HVGs can co-vary in expression and are
269 potentially co-regulated. It also suggests that HVGs are not likely to co-vary in
270 expression with non variable genes. To test if this result could indicate a bias in the
271 method used to construct the variability network and detect modules, we analysed
272 expression levels in single seedlings for genes in modules with high or low percentage
273 of HVG (Fig S5b). We find that modules with high or low percentage of HVG have
274 different expression profiles in the seedlings, indicating an absence of bias. Moreover,

275 modules with a high percentage of HVG tend to be more connected to one another
276 than with modules containing a low percentage of HVG, and vice versa (Pearson
277 correlation=0.6896, p-value= 0.0007683, Fig 5b).
278

module name	number of genes in module	number of HVGs in module	percentage of HVGs in module	Fisher's p-value
37	12	12	100,00	0,0022
43	8	8	100,00	0,0118
66	6	6	100,00	0,0287
18	15	14	93,33	0,0022
24	47	42	89,36	0,0001
59	9	8	88,89	0,0181
15	44	38	86,36	0,0001
1	19	15	78,95	0,0052
32	56	43	76,79	0,0001
64	7	5	71,43	0,1496
23	6	4	66,67	0,2384
21	79	50	63,29	0,0001
79	10	6	60,00	0,1322
4	175	97	55,43	0,0001
25	24	9	37,50	0,4014
8	41	8	19,51	0,4820
2	150	17	11,33	0,0002
13	77	6	7,79	0,0009
9	101	3	2,97	0,0001
0	334	3	0,90	0,0001
5	91	0	0,00	0,0001
6	61	0	0,00	0,0001
36	24	0	0,00	0,0047
12	21	0	0,00	0,0127
71	8	0	0,00	0,2142
126	7	0	0,00	0,3578
70	6	0	0,00	0,3512
86	5	0	0,00	0,5915

279

280 **Table 1.**

281 Number and percentage of HVGs in each module, for modules with at least 5 genes.
282

283 Our results show that gene co-expression networks can be inferred in the absence of
284 genetic or environmental perturbation. Moreover, genes don't need to show a high
285 level of gene expression variability between seedlings to be integrated in the network.
286

Additional gene co-expression is identified in the variability network compared to a network inferred from a pool 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 a co-expression network inferred from a pool of plants. Since the transcriptome dataset contains data for several time points throughout a day/night cycle, we decided to infer a co-expression network using the average expression of the fourteen seedlings for each time point and thus exploit changes in expression happening during the time course. This network, referred to as the averaged time course network, allows the identification of co-expression throughout the time course. 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.

module name	number of genes in module	number of genes in module also in averaged time course network	percentage of genes in module also in averaged time course network	Fisher's p-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

Table 2.

Number and percentage in each module of genes also detected in the averaged time course network.

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%: 7 out of 8 genes) and 13 (76%: 59 out of 77 genes).

315 We find that genes in these modules have very similar and clear expression profiles
316 throughout the time course (Fig 2). This is also the case for all modules with at least
317 50% of genes in the averaged time course network (modules highlighted in blue in Fig
318 2). This result could suggest that the reason why these modules contain many genes
319 also present in the averaged time course network, is because their genes have very
320 similar expression patterns throughout the day/night cycle. On the other hand, several
321 modules that only have 15% or less of genes present in the averaged time course
322 network are composed of genes without clear expression patterns during the time
323 course. These results show that additional gene co-expression is identified in the
324 variability network compared to the averaged time course network. Most importantly,
325 using gene expression in single seedlings, co-expression between genes can be
326 detected even in absence of expression patterns throughout the day/night cycle.
327
328 **Modules identified in the variability network are functionally relevant**
329
330 In order to define if the modules identified in the variability network are functionally
331 relevant, we performed a gene ontology (GO) enrichment analysis. We find that some
332 of the modules have strongly enriched GO (Table S2).
333 For example, the module 8 is enriched in multiple GO related to photosynthesis. In
334 particular, 33 genes out of the 41 in this module are members of the photosystem I or
335 II, or of the light harvesting complex (Fig 3a, Table S3). Other genes in this module
336 also have functions related to photosynthesis: CURT1A is required for a proper
337 thylakoid morphology [34], while RBCS1A, RBCS3B and RCA are members of the
338 Rubisco or necessary to Rubisco light activation [35,36]. Most edges of module 8 are
339 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 the 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 (Fig S6a). Also, we find that these other genes are at the periphery of the module 8 (Fig S6b), which highlights that these genes are less well correlated with the dense core of highly correlated photosystem genes in the centre of the network. Another module enriched in GO related with photosynthesis is module 37 (Table S2), in which 9 out of the 12 genes are chloroplast genes, some being present in the Photosystems I or II, in the Cytochrome b6/f complex or in the ATP synthase (Fig 3a, 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 are co-regulated, not only over time, but also between plants at a given time.

Fig 3. Modules enriched in genes involved in the photosynthesis and the glucosynolate pathway

(A) Functional analysis of modules 8 and 37. For each module, the number of genes that are part of the Photosystem I (green), Photosystem II (orange), the Light harvesting complex (blue) or the ATP synthase (purple) are indicated. (B) Functional analysis of module 1. Genes of the module are color coded depending on their role in the glucosynolate pathway: biosynthesis (turquoise), transport (green) or regulation (orange). Genes previously identified as co-expressed with glucosynolate biosynthesis genes are also indicated (grey). On the right side, the glucosinolate

biosynthesis pathway is shown with an indication of the number of genes present in the module 1 at each step of the pathway.

Module 71 is enriched in GO related to DNA packaging (Table S2), and is in fact only composed of histones, including 2 variants of H2A, 2 variants of H2B, and H3.1 (Table S3). None of the genes in this module are HVGs, and 7 genes out of 8 are also present in the averaged time course network.

Module 1 is enriched in GO related to glucosinolate (Table S2). We find that 16 out of 19 genes of the module are in the glucosinolate biosynthesis pathway, transporters of glucosinolate, or transcription factors (TF) regulating the pathway (Fig 3b, Table S3). All genes in module 1, except one, were previously identified as co-expressed, in a previous study of the glucosinolate pathway [13]. 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 [37]. AKN2 is involved in sulfate assimilation which is linked to glucosinolate metabolism [37]. 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 (Table S2), with 6 out of 8 genes shown to be involved in flavonoid biosynthesis [38]. 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 2-

oxoglutarate (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 the 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 TFs in the modules. We focussed on TFs for which ChIP-seq were performed in similar conditions (seedlings grown in day/night cycles), and for which a list of target genes have been previously published [39–42]. This way, we identified an enrichment in modules for targets of SPL7, GI, PIF4 and PRR5 (Table S4). For example, all 41 genes in the module 8 are SPL7 targets (Table S4) [39]. 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,

414 including photosynthesis [39], in agreement with the predominant role in
415 photosynthesis of genes in this module 8.
416 We also find that 6 out of 7 genes in the module 64 are targets of GI (Table S4, Fig
417 4a). This is more compared to the entire network in which 394 genes are GI targets
418 (22%). To explore more in detail GI binding at the genes in the module 64, we
419 downloaded the ChIP-seq data, mapped it on the *Arabidopsis thaliana* genome and
420 looked at the ChIP-seq signal for GI at all the 7 genes of the module 64 [40]. We find
421 a strong signal for the GI ChIP-seq at the promoter of the 6 genes that were already
422 identified as GI targets (Fig 4a). Interestingly, the signal for GI at the 7th gene,
423 AT1G03630, not previously described as a GI target, is equally strong at the promoter
424 of the gene (Fig 4a). This result indicates that AT1G03630 is also a target of GI, even
425 if it has not been previously identified as such. AT1G03630, or PORC, encodes for a
426 protein with protochlorophyllide oxidoreductase activity that is NADPH- and light-
427 dependent [43].
428
429 **Fig 4. Additional TF targets can be identified using TF targets enrichment in**
430 **modules**
431 (A) Analysis of GI TF targets on the module 64: 6 of the 7 genes in the module 64 are
432 known targets of GI (left). IGV screenshot showing the signal for the GI ChIP-seq
433 (right) at a known GI target (top) and for the seventh gene in the module 64 that is not
434 known as a GI target (bottom). (B) Analysis of PIF4 TF targets on the module 86: 3 of
435 the 5 genes in the module 86 are known targets of PIF4 (left). IGV screenshot showing
436 the signal for the PIF4 ChIP-seq (right) at a known PIF4 target (top) and for the two
437 other genes in the module 86 that are not known as a PIF4 target (bottom).
438

Another TF with enriched targets in some modules is PIF4. We find that 3 out of the 5 genes (60%) in the module 86 are PIF4 targets (Table S3, Fig 4b), while only 305 genes in the full network are PIF4 targets (17.6%). To explore more in detail PIF4 binding at the genes in the module 86, 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 5 genes of the module 86 [41]. We observe a strong signal for the PIF4 ChIP at the promoters of the 3 known targets in the module 86 (Fig 4b). We also see a clear signal for the PIF4 ChIP for the two other genes in the module 86, AT4G26542 and AT5G55730, suggesting that they are also targets of PIF4 (Fig 4b). AT4G26542 is an anti-sens transcript for AT4G26540. AT5G55730 (FLA1) encodes a fasciclin-like arabinogalactan-protein 1 [44].

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 [42]. For comparison, 27% of all genes in the network are PRR5 targets. To explore more in detail PRR5 binding at the genes in the module 8, we downloaded the ChIP-seq data, mapped it on the *Arabidopsis thaliana* genome and looked at the ChIP-seq signal for PRR5 at all the 41 genes of the module 8 [45]. We find a strong ChIP-seq PRR5 signal at the 32 target genes in the module 8, and a similarly strong signal for most of the other 9 genes in the module that were not listed as a PRR5 target (Fig 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 (Table S4). This way, we identified 5 additional PRR5 targets in the module 8, and 9 additional PRR5 targets in the module 21. When combining the PRR5 targets from both analyses, the total percentage of PRR5 targets

464 is 90% in the module 8, and 83% in the module 21 (Fig S7b). These results suggest
465 that most, if not all, genes in the module 8 and 21 are in fact PRR5 targets.
466
467 Overall, we find that some modules are enriched for TFs targets, and that this
468 enrichment can be used to identify additional targets for the TF in the modules showing
469 enrichment for its targets.
470
471 **Discussion**
472
473 In this work, we have analysed gene co-expression networks inferred using expression
474 data generated in the absence of genetic and environmental perturbations. To do this,
475 we made use of an already published dataset of transcriptomes performed on single
476 seedlings that were grown in the same environment. We showed that genes do not
477 need a high level of gene expression variability between seedlings to be able to
478 integrate them in the network (Table 1). Moreover we find that modules identified in
479 this network are biologically relevant, as several are strongly enriched in GOs (Fig 3)
480 and in TF targets (Fig 4). Based on these enrichments, we speculated that AT4G22870
481 could also have a role in flavonoid metabolism and identified new targets for the TFs
482 GI, PIF4 and PRR5.
483
484 We find that it is possible to infer gene co-expression networks using transcriptomes
485 of genetically identical plants grown in the exact same environment. This is in
486 agreement with previous work, where co-expression networks have been inferred on
487 transcriptomes generated on individual plants and for which genetic and
488 environmental effects have been removed *in silico* [26]. 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 HVG (Fig 5 and S8). We observe that modules have either a high or low percentage of HVG, but rarely a mix of HVG with non HVG. 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 HVG. This is the case for example for the module 37, enriched in photosynthesis (100% of HVG), the module 43, enriched in flavonoid metabolism (100% of HVG), or the module 1, enriched in glucosinolate metabolism (78% of HVG). Flavonoids are secondary metabolites and have been shown to be involved in many biotic and abiotic responses in plants [46]. And glucosinolates are involved in the response to pathogens [47]. In agreement with our observation, previous work showed that HVG are usually involved in the response to the environment [28,48–51]. In particular, plant-to-plant variability has already been observed for glucosinolates [30], 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 and colleagues, we find that the modules of the network identified in 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 the 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,

514 AT4G22870 codes for a protein of the 2-oxoglutarate (2OG) and Fe(II)-dependent
515 oxygenase superfamily and three 2-oxoglutarate- and ferrous iron-dependent
516 oxygenases have been previously shown to be involved in flavonoid biosynthesis [38].
517 Most importantly, this new potential candidate gene could not have been detected by
518 analysing the network inferred using day/night environmental fluctuations as none of
519 the genes in the module 43 are also present in the averaged time course network.
520 We find several modules with enrichment for genes involved in photosynthesis. In
521 particular, the modules 8 and 37. The main distinction between these two modules is
522 that module 8 is composed of genes from the nuclear genome, while module 37 is
523 mainly composed of genes from the chloroplast genome. Our approach was not
524 designed to specifically identify and separate genes from different organelles,
525 suggesting that genes from the nuclear and chloroplast genomes involved in
526 photosynthesis vary differently in expression between seedlings. Our result is in
527 agreement with the fact that organelle functional modules can be detected in
528 *Arabidopsis thaliana* [52]. However, genes that are not from the nuclear genome are
529 usually ignored in network analysis, and it would be of interest to integrate them in the
530 future.
531
532 Finally, we identified enrichment for targets of the TFs GI, PIF4 and PRR5 in different
533 modules, and used this enrichment to highlight new targets. We find that in most
534 cases, when a module is enriched in targets for a TF, the remaining genes of that
535 module are also targets of this TF. By reanalysing the ChIP-seq data for PRR5, we
536 could increase the percentage of targets in modules already showing a strong
537 enrichment. This result shows the double interest of combining co-expression
538 networks with ChIP-seq data [21,22]. On the one hand, ChIP-seq data adds

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 the module 8 are targets of this TF. Genes in the 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 [53–55].

The functional characterisation of the network has been restricted to some modules with obvious GO enrichments, and to TFs 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 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 so (https://jlgroup.shinyapps.io/VariabilityNetwork/).

We show that most genes in the network are not HVG (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 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

564 uses these small differences between seedlings that might be caused by micro-
565 environmental fluctuation, or a different state of the plant caused by internal factors. It
566 indicates that plants are very sensitive to minor changes in their environment, and that
567 we could harness this sensitivity to better understand gene expression regulation.
568 Phenotypic differences have been observed between genetically identical plants
569 grown in the same environment [27,29,30,56–58], indicating that the changes in
570 expression of pathways we highlight here might be physiologically relevant. It shows
571 that it is not necessary to perform experiments in very different environmental
572 conditions to identify co-expression networks that could be relevant to the studied
573 pathway. Strong fluctuations (mutants, over-expressors, environmental fluctuations)
574 could potentially affect a big part of the transcriptome that could mask some co-
575 expressions of interest showing the usefulness of our approach in some contexts. Our
576 work shows the interest in harnessing gene expression variability between genetically
577 identical individual plants in order to better understand gene regulation in a context
578 where differences between plants are not known and probably very subtle.
579

580 **Materials and Methods**

581 **Transcriptome data**

582 The transcriptomes we used were already published (GSE115583; [28]), and
583 performed on single seedlings, for a total of fourteen seedlings per time point every
584 two hours over a 24 hours cycle. Expression levels and corrected variability levels for
585 all genes were downloaded from https://jlgroup.shinyapps.io/AraNoisy/, as this data
586 had been already corrected as previously described [28].
587

588 **Network construction**

589 *Variability network*: For each of the 12 time points (0h, 2h, 4h, … 22h) we calculated
590 the Spearman correlation between every pair of genes, using their expression profiles
591 across the 14 seedlings (Fig 1a). Using a Benjamini-Hochberg correction with a false-
592 discovery rate of 10% the most significant correlations were selected, and further
593 filtered by only considering those for which a significant correlation appeared in four
594 consecutive time points (with one gap allowed, e.g. 8h, 10h, 14h, 16h). These
595 correlations form the edges of the variability network. We also calculated a version of
596 the network using a filter that only required three consecutive time points, and
597 calculated network modules using the same community detection algorithm. As can
598 be seen in Fig S1, similar modules with a similar overall connectivity between them
599 are found, which confirms the robustness of the modules in our original network. All
600 network analysis was carried out using the Python NetworkX and python-louvain
601 libraries.
602
603 *Averaged time course network:* For the averaged time course network we calculated
604 the mean expression across all seedlings for every time point, generating a time series
605 of average expression for every gene. We again calculated the Spearman correlations
606 for every pair of genes and generated a network by applying the Bonferroni correction
607 as a (highly conservative) significance cutoff. This yielded a network that was similar
608 in size to the variability network. All network analysis was carried out using the Python
609 NetworkX and python-louvain libraries.
610
611 **Community detection**
612 The Louvain algorithm [33] community detection algorithm was used to identify
613 modules in the networks. This algorithm attempts to maximise the modularity of the

614 network by searching the space of network partitions. Due to the size of the search
615 space it is unable to find the global maximum. The composition of modules may
616 therefore (as with most community detection algorithms) vary somewhat between runs
617 of the algorithm.
618
619 **RT-qPCR**
620 Col-0 WT *Arabidopsis thaliana* seeds were sterilised, stratified for 3 days at 4°C in
621 dark and transferred for germination on solid 1/2X Murashige and Skoog (MS) media
622 at 22°C in long days for 24 hours. Using a binocular microscope, seeds that were at
623 the same stage of germination were transferred into a new plate containing solid 1/2X
624 MS media. Seedlings were grown at 22°C, 65% humidity, with 12 hours of light (170
625 µmoles) and 12 hours of dark in a conviron reach-in cabinet. After 7 days of growth,
626 seedlings were harvested individually into a 96-well plate and flash-frozen in dry ice.
627 Sixteen seven-day old Col-0 WT *Arabidopsis thaliana* seedlings were harvested
628 individually and flash-frozen in dry ice at ZT6 and at ZT14. All seedlings harvested in
629 a given time point were grown in the same plate. Total RNA was isolated from 1 ground
630 seedling. RNA concentration was assessed using Qubit RNA HS assay kit. cDNA
631 synthesis was performed on 700ng of DNAse treated RNA using the Transcriptor First
632 Strand cDNA Synthesis Kit. For RT-qPCR analysis, 0.4 µl of cDNA was used as
633 template in a 10 µl reaction performed in the LightCycler 480 instrument using LC480
634 SYBR green I master. Gene expression relative to two control genes (SandF and
635 PP2A) was measured (See Table S5 for the list of primers used for RT-qPCR).
636
637 **Gene Ontology term enrichment**

638 We used the Ontologizer [59] command line tool with Bonferroni multiple-hypothesis
639 correction to perform Gene Ontology (GO) term enrichment analysis of the network
640 modules. Only the significantly enriched non redundant GO are shown.
641
642 **ChIP-seq data and analysis**
643 ChIP-seq data were downloaded from GSE45213 for SPL7 [39], from GSE129865 for
644 GI [40], from GSE43286 for PIF4 [41] and from GSE36361 for PRR5 [42].
645 ChIP-seq data were analysed in house, using a combination of publicly available
646 software and in-house scripts. Reads were aligned to the TAIR10 genome using
647 Bowtie2 [60]. Potential optical duplicates were removed using Picard tools
648 (https://github.com/broadinstitute/picard). Peak calling was performed using MASC2
649 [61], with the corresponding INPUT used as a reference. Snapshots of ChIP-seq signal
650 around targets were shown using the Integrative Genomics Viewer (IGV [62]).
651
652 **Authors' contribution**
653 SC conceived the project. SC, JCWL and SA designed the project. SA inferred the
654 networks. SC and SA performed downstream analyses of the network and SC
655 interpreted the data. MB performed the RT-qPCR validation. SC, JCWL and SA
656 wrote the article, with SC writing the first draft.
657
658 **Bibliography**

659 1. Stuart JM, Segal E, Koller D, Kim SK. A gene-coexpression network for global
660 discovery of conserved genetic modules. Science. 2003;302: 249–255.
661 doi:10.1126/science.1087447

662 2. Serin EAR, Nijveen H, Hilhorst HWM, Ligterink W. Learning from Co-

663 expression Networks: Possibilities and Challenges. Front Plant Sci. 2016;7:
664 444. doi:10.3389/fpls.2016.00444

665 3. Contreras-López O, Moyano TC, Soto DC, Gutiérrez RA. Step-by-Step
666 Construction of Gene Co-expression Networks from High-Throughput
667 Arabidopsis RNA Sequencing Data. Methods Mol Biol. 2018;1761: 275–301.
668 doi:10.1007/978-1-4939-7747-5_21

669 4. Ruan J, Dean AK, Zhang W. A general co-expression network-based
670 approach to gene expression analysis: comparison and applications. BMC
671 Syst Biol. 2010;4: 8. doi:10.1186/1752-0509-4-8

672 5. Usadel B, Obayashi T, Mutwil M, Giorgi FM, Bassel GW, Tanimoto M, et al.
673 Co-expression tools for plant biology: opportunities for hypothesis generation
674 and caveats. Plant Cell Environ. 2009;32: 1633–1651. doi:10.1111/j.1365-
675 3040.2009.02040.x

676 6. Mao L, Van Hemert JL, Dash S, Dickerson JA. Arabidopsis gene co-
677 expression network and its functional modules. BMC Bioinformatics. 2009;10:
678 346. doi:10.1186/1471-2105-10-346

679 7. Aoki K, Ogata Y, Shibata D. Approaches for extracting practical information
680 from gene co-expression networks in plant biology. Plant Cell Physiol.
681 2007;48: 381–390. doi:10.1093/pcp/pcm013

682 8. Liu W, Lin L, Zhang Z, Liu S, Gao K, Lv Y, et al. Gene co-expression network
683 analysis identifies trait-related modules in Arabidopsis thaliana. Planta.
684 2019;249: 1487–1501. doi:10.1007/s00425-019-03102-9

685 9. de Luis Balaguer MA, Fisher AP, Clark NM, Fernandez-Espinosa MG, Möller
686 BK, Weijers D, et al. Predicting gene regulatory networks by combining spatial

and temporal gene expression data in Arabidopsis root stem cells. Proc Natl Acad Sci USA. 2017;114: E7632–E7640. doi:10.1073/pnas.1707566114

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

11. Silva AT, Ribone PA, Chan RL, Ligterink W, Hilhorst HWM. A Predictive Coexpression Network Identifies Novel Genes Controlling the Seed-to-Seedling Phase Transition in Arabidopsis thaliana. Plant Physiol. 2016;170: 2218–2231. doi:10.1104/pp.15.01704

12. Xie W, Huang J, Liu Y, Rao J, Luo D, He M. Exploring potential new floral organ morphogenesis genes of Arabidopsis thaliana using systems biology approach. Front Plant Sci. 2015;6: 829. doi:10.3389/fpls.2015.00829

13. Wisecaver JH, Borowsky AT, Tzin V, Jander G, Kliebenstein DJ, Rokas A. A global coexpression network approach for connecting genes to specialized metabolic pathways in plants. Plant Cell. 2017;29: 944–959. doi:10.1105/tpc.17.00009

14. Sharma R, Singh G, Bhattacharya S, Singh A. Comparative transcriptome meta-analysis of Arabidopsis thaliana under drought and cold stress. PLoS ONE. 2018;13: e0203266. doi:10.1371/journal.pone.0203266

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

16. Amrine KCH, Blanco-Ulate B, Cantu D. Discovery of core biotic stress responsive genes in Arabidopsis by weighted gene co-expression network

analysis. PLoS ONE. 2015;10: e0118731. doi:10.1371/journal.pone.0118731

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

18. Ichihashi Y, Aguilar-Martínez JA, Farhi M, Chitwood DH, Kumar R, Millon LV, et al. Evolutionary developmental transcriptomics reveals a gene network module regulating interspecific diversity in plant leaf shape. Proc Natl Acad Sci USA. 2014;111: E2616-21. doi:10.1073/pnas.1402835111

19. Des Marais DL, Guerrero RF, Lasky JR, Scarpino SV. Topological features of a gene co-expression network predict patterns of natural diversity in environmental response. Proc Biol Sci. 2017;284. doi:10.1098/rspb.2017.0914

20. Rao X, Dixon RA. Co-expression networks for plant biology: why and how. Acta Biochim Biophys Sin (Shanghai). 2019;51: 981–988. doi:10.1093/abbs/gmz080

21. Kulkarni SR, Vandepoele K. 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.bbagrm.2019.194447

22. Chen D, Yan W, Fu L-Y, Kaufmann K. Architecture of gene regulatory networks controlling flower development in Arabidopsis thaliana. Nat Commun. 2018;9: 4534. doi:10.1038/s41467-018-06772-3

23. He F, Zhou Y, Zhang Z. Deciphering the Arabidopsis floral transition process by integrating a protein-protein interaction network and gene expression data. Plant Physiol. 2010;153: 1492–1505. doi:10.1104/pp.110.153650

24. Vandepoele K, Quimbaya M, Casneuf T, De Veylder L, Van de Peer Y.

735 Unraveling transcriptional control in Arabidopsis using cis-regulatory elements
736 and coexpression networks. Plant Physiol. 2009;150: 535–546.
737 doi:10.1104/pp.109.136028

738 25. Ma S, Shah S, Bohnert HJ, Snyder M, Dinesh-Kumar SP. Incorporating motif
739 analysis into gene co-expression networks reveals novel modular expression
740 pattern and new signaling pathways. PLoS Genet. 2013;9: e1003840.
741 doi:10.1371/journal.pgen.1003840

742 26. Bhosale R, Jewell JB, Hollunder J, Koo AJK, Vuylsteke M, Michoel T, et al.
743 Predicting gene function from uncontrolled expression variation among
744 individual wild-type Arabidopsis plants. Plant Cell. 2013;25: 2865–2877.
745 doi:10.1105/tpc.113.112268

746 27. Jimenez-Gomez JM, Corwin JA, Joseph B, Maloof JN, Kliebenstein DJ.
747 Genomic analysis of QTLs and genes altering natural variation in stochastic
748 noise. PLoS Genet. 2011;7: e1002295. doi:10.1371/journal.pgen.1002295

749 28. Cortijo S, Aydin Z, Ahnert S, Locke JC. Widespread inter-individual gene
750 expression variability in Arabidopsis thaliana. Mol Syst Biol. 2019;15: e8591.
751 doi:10.15252/msb.20188591

752 29. Hall MC, Dworkin I, Ungerer MC, Purugganan M. Genetics of
753 microenvironmental canalization in Arabidopsis thaliana. Proc Natl Acad Sci
754 USA. 2007;104: 13717–13722. doi:10.1073/pnas.0701936104

755 30. Mönchgesang S, Strehmel N, Trutschel D, Westphal L, Neumann S, Scheel D.
756 Plant-to-Plant Variability in Root Metabolite Profiles of 19 Arabidopsis thaliana
757 Accessions Is Substance-Class-Dependent. Int J Mol Sci. 2016;17.
758 doi:10.3390/ijms17091565

759 31. Shen X, Pettersson M, Rönnegård L, Carlborg Ö. Inheritance beyond plain
760 heritability: variance-controlling genes in Arabidopsis thaliana. PLoS Genet.
761 2012;8: e1002839. doi:10.1371/journal.pgen.1002839

762 32. Mantsoki A, Devailly G, Joshi A. Gene expression variability in mammalian
763 embryonic stem cells using single cell RNA-seq data. Comput Biol Chem.
764 2016;63: 52–61. doi:10.1016/j.compbiolchem.2016.02.004

765 33. Blondel VD, Guillaume J-L, Lambiotte R, Lefebvre E. Fast unfolding of
766 communities in large networks. J Stat Mech. 2008;2008: P10008.
767 doi:10.1088/1742-5468/2008/10/P10008

768 34. Pribil M, Sandoval-Ibáñez O, Xu W, Sharma A, Labs M, Liu Q, et al. Fine-
769 Tuning of Photosynthesis Requires CURVATURE THYLAKOID1-Mediated
770 Thylakoid Plasticity. Plant Physiol. 2018;176: 2351–2364.
771 doi:10.1104/pp.17.00863

772 35. Izumi M, Tsunoda H, Suzuki Y, Makino A, Ishida H. RBCS1A and RBCS3B,
773 two major members within the Arabidopsis RBCS multigene family, function to
774 yield sufficient Rubisco content for leaf photosynthetic capacity. J Exp Bot.
775 2012;63: 2159–2170. doi:10.1093/jxb/err434

776 36. Carmo-Silva AE, Salvucci ME. The regulatory properties of Rubisco activase
777 differ among species and affect photosynthetic induction during light
778 transitions. Plant Physiol. 2013;161: 1645–1655. doi:10.1104/pp.112.213348

779 37. Yatusevich R, Mugford SG, Matthewman C, Gigolashvili T, Frerigmann H,
780 Delaney S, et al. Genes of primary sulfate assimilation are part of the
781 glucosinolate biosynthetic network in Arabidopsis thaliana. Plant J. 2010;62:
782 1–11. doi:10.1111/j.1365-313X.2009.04118.x

783 38. Saito K, Yonekura-Sakakibara K, Nakabayashi R, Higashi Y, Yamazaki M,
784 Tohge T, et al. The flavonoid biosynthetic pathway in Arabidopsis: structural
785 and genetic diversity. Plant Physiol Biochem. 2013;72: 21–34.
786 doi:10.1016/j.plaphy.2013.02.001

787 39. Zhang H, Zhao X, Li J, Cai H, Deng XW, Li L. MicroRNA408 is critical for the
788 *HY5-SPL7* gene network that mediates the coordinated response to light and
789 copper. Plant Cell. 2014;26: 4933–4953. doi:10.1105/tpc.114.127340

790 40. Nohales MA, Liu W, Duffy T, Nozue K, Sawa M, Pruneda-Paz JL, et al. Multi-
791 level Modulation of Light Signaling by GIGANTEA Regulates Both the Output
792 and Pace of the Circadian Clock. Dev Cell. 2019;49: 840-851.e8.
793 doi:10.1016/j.devcel.2019.04.030

794 41. Pfeiffer A, Shi H, Tepperman JM, Zhang Y, Quail PH. Combinatorial
795 complexity in a transcriptionally centered signaling hub in Arabidopsis. Mol
796 Plant. 2014;7: 1598–1618. doi:10.1093/mp/ssu087

797 42. Liu TL, Newton L, Liu M-J, Shiu S-H, Farré EM. A G-Box-Like Motif Is
798 Necessary for Transcriptional Regulation by Circadian Pseudo-Response
799 Regulators in Arabidopsis. Plant Physiol. 2016;170: 528–539.
800 doi:10.1104/pp.15.01562

801 43. Gabruk M, Stecka A, Strzałka W, Kruk J, Strzałka K, Mysliwa-Kurdziel B.
802 Photoactive protochlorophyllide-enzyme complexes reconstituted with PORA,
803 PORB and PORC proteins of A. thaliana: fluorescence and catalytic
804 properties. PLoS ONE. 2015;10: e0116990. doi:10.1371/journal.pone.0116990

805 44. Johnson KL, Kibble NAJ, Bacic A, Schultz CJ. A fasciclin-like arabinogalactan-
806 protein (FLA) mutant of Arabidopsis thaliana, fla1, shows defects in shoot

807 regeneration. PLoS ONE. 2011;6: e25154. doi:10.1371/journal.pone.0025154

808 45. Nakamichi N, Kiba T, Kamioka M, Suzuki T, Yamashino T, Higashiyama T, et
809 al. Transcriptional repressor PRR5 directly regulates clock-output pathways.
810 Proc Natl Acad Sci USA. 2012;109: 17123–17128.
811 doi:10.1073/pnas.1205156109

812 46. Tohge T, de Souza LP, Fernie AR. Current understanding of the pathways of
813 flavonoid biosynthesis in model and crop plants. J Exp Bot. 2017;68: 4013–
814 4028. doi:10.1093/jxb/erx177

815 47. Burow M, Halkier BA, Kliebenstein DJ. Regulatory networks of glucosinolates
816 shape Arabidopsis thaliana fitness. Curr Opin Plant Biol. 2010;13: 348–353.
817 doi:10.1016/j.pbi.2010.02.002

818 48. Gasch AP, Yu FB, Hose J, Escalante LE, Place M, Bacher R, et al. Single-cell
819 RNA sequencing reveals intrinsic and extrinsic regulatory heterogeneity in
820 yeast responding to stress. PLoS Biol. 2017;15: e2004050.
821 doi:10.1371/journal.pbio.2004050

822 49. Newman JRS, Ghaemmaghami S, Ihmels J, Breslow DK, Noble M, DeRisi JL,
823 et al. Single-cell proteomic analysis of S. cerevisiae reveals the architecture of
824 biological noise. Nature. 2006;441: 840–846. doi:10.1038/nature04785

825 50. Yin S, Wang P, Deng W, Zheng H, Hu L, Hurst LD, et al. Dosage
826 compensation on the active X chromosome minimizes transcriptional noise of
827 X-linked genes in mammals. Genome Biol. 2009;10: R74. doi:10.1186/gb-
828 2009-10-7-r74

829 51. Hirao K, Nagano AJ, Awazu A. Noise-plasticity correlations of gene expression
830 in the multicellular organism Arabidopsis thaliana. J Theor Biol. 2015;387: 13–

22. doi:10.1016/j.jtbi.2015.09.017

52. Penga J, Wang T, Huc J, Wang Y, Chen J. Constructing networks of organelle functional modules in arabidopsis. Curr Genomics. 2016;17: 427–438. doi:10.2174/1389202917666160726151048

53. Schaffer R, Landgraf J, Accerbi M, Simon V, Larson M, Wisman E. Microarray analysis of diurnal and circadian-regulated genes in Arabidopsis. Plant Cell. 2001;13: 113–123. doi:10.1105/tpc.13.1.113

54. Harmer SL, Hogenesch JB, Straume M, Chang HS, Han B, Zhu T, et al. Orchestrated transcription of key pathways in Arabidopsis by the circadian clock. Science. 2000;290: 2110–2113. doi:10.1126/science.290.5499.2110

55. Dodd AN, Kusakina J, Hall A, Gould PD, Hanaoka M. The circadian regulation of photosynthesis. Photosyn Res. 2014;119: 181–190. doi:10.1007/s11120-013-9811-8

56. Forde BG. Is it good noise? The role of developmental instability in the shaping of a root system. J Exp Bot. 2009;60: 3989–4002. doi:10.1093/jxb/erp265

57. Paxman GJ. Differentiation and Stability in the Development of Nicotiana rustica. Ann Bot. 1956;20: 331–347. doi:10.1093/oxfordjournals.aob.a083526

58. Sakai K, Shimamoto Y. Developmental instability in leaves and flowers of NICOTIANA TABACUM. Genetics. 1965;51: 801–813.

59. Bauer S, Grossmann S, Vingron M, Robinson PN. Ontologizer 2.0--a multifunctional tool for GO term enrichment analysis and data exploration. Bioinformatics. 2008;24: 1650–1651. doi:10.1093/bioinformatics/btn250

60. Langmead B, Trapnell C, Pop M, Salzberg SL. Ultrafast and memory-efficient

alignment of short DNA sequences to the human genome. Genome Biol. 2009;10: R25. doi:10.1186/gb-2009-10-3-r25

61. Zhang Y, Liu T, Meyer CA, Eeckhoute J, Johnson DS, Bernstein BE, et al. Model-based analysis of ChIP-Seq (MACS). Genome Biol. 2008;9: R137. doi:10.1186/gb-2008-9-9-r137

62. Robinson JT, Thorvaldsdóttir H, Winckler W, Guttman M, Lander ES, Getz G, et al. Integrative genomics viewer. Nat Biotechnol. 2011;29: 24–26. doi:10.1038/nbt.1754

Supplementary Figures and Tables legends

Supplementary figure 1

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

Supplementary figure 2

Comparison of edges in modules detected in the networks containing edges present in 3 or 4 consecutive time points. Modules detected in the network based on edges present in at least 3 consecutive time points are shown in blue. Modules detected in the network based on edges present in at least 3 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 are indicated.

Supplementary figure 3

877 Number of edges in the final network that are detected in each time point, for every
878 module containing at least 5 genes.

879

880 **Supplementary figure 4**

881 a. Correlation in expression between seedlings for genes of the module 1 and module
882 21, for the RNA-seq experiment. AT5G07690 at ZT14 was removed as it is not
883 expressed.

884 b. Correlation in expression between seedlings for genes of the module 1 and module
885 21, based on a RT-qPCR replicate of the RNA-seq experiment. Sixteen seedlings
886 where harvested at ZT6 and at ZT14.

887 c. Normalised expression level in the fourteen seedlings for the genes of the module
888 21, from the RNA-seq data. Expression level for AT4G13250 is shown as the x axis
889 while expression for the other genes of the module are shown on the y axis.
890 Expression is mean normalised for each gene.

891

892 **Supplementary figure 5**

893 a. Inter-individual gene expression variability profiles throughout the time course for
894 genes in each module with 5 genes or more. Each line represents the corrected
895 variability level for one gene: : corrected $CV^2=[\log_2(CV^2/\text{trend})]$, with
896 $CV^2=\text{variance}/(\text{average}^2)]$ (see Cortijo et al., 2019).

897 Modules are ordered by the percentage of HVG (high to low). Modules highlighted in
898 blue contain 75% or more of HVGs. Modules highlighted in red contain 10% or less of
899 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 normalised: 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 normalised 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 on 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 64 that is not known as a PRR5 target (bottom).

b. Comparison of published (blue) and realised (grey) PRR5 targets in modules 8 (left) and 21 (right).

Supplementary figure 8

Organisation 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.

Table S1

List of genes in each module

Table S2

GO enriched (corrected p-value<0.5) in each module

Table S3

Function of genes in modules 1, 43 and 71

Table S4

Number and percentage in the modules of targets for the TFs SPL7, GI, PIF4 and PRR5

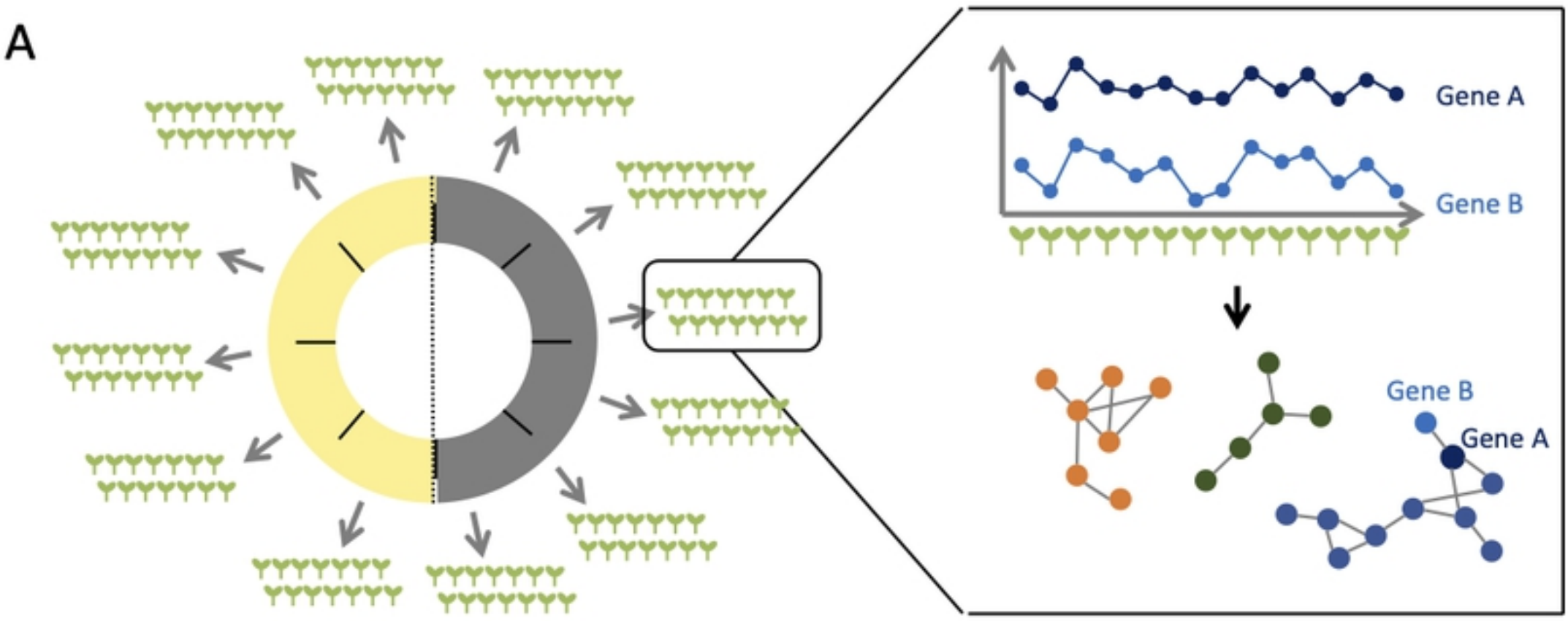


Fig1

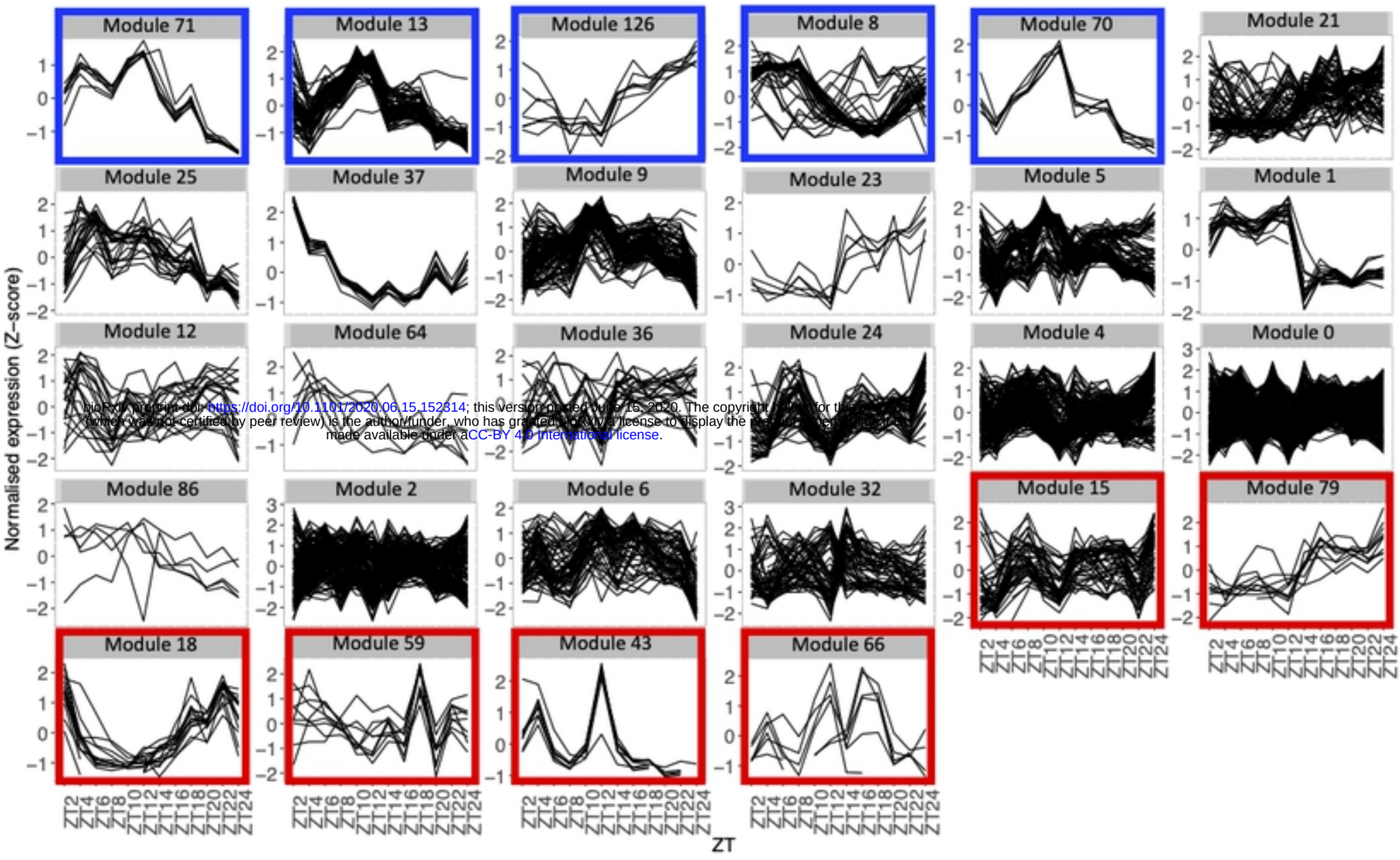
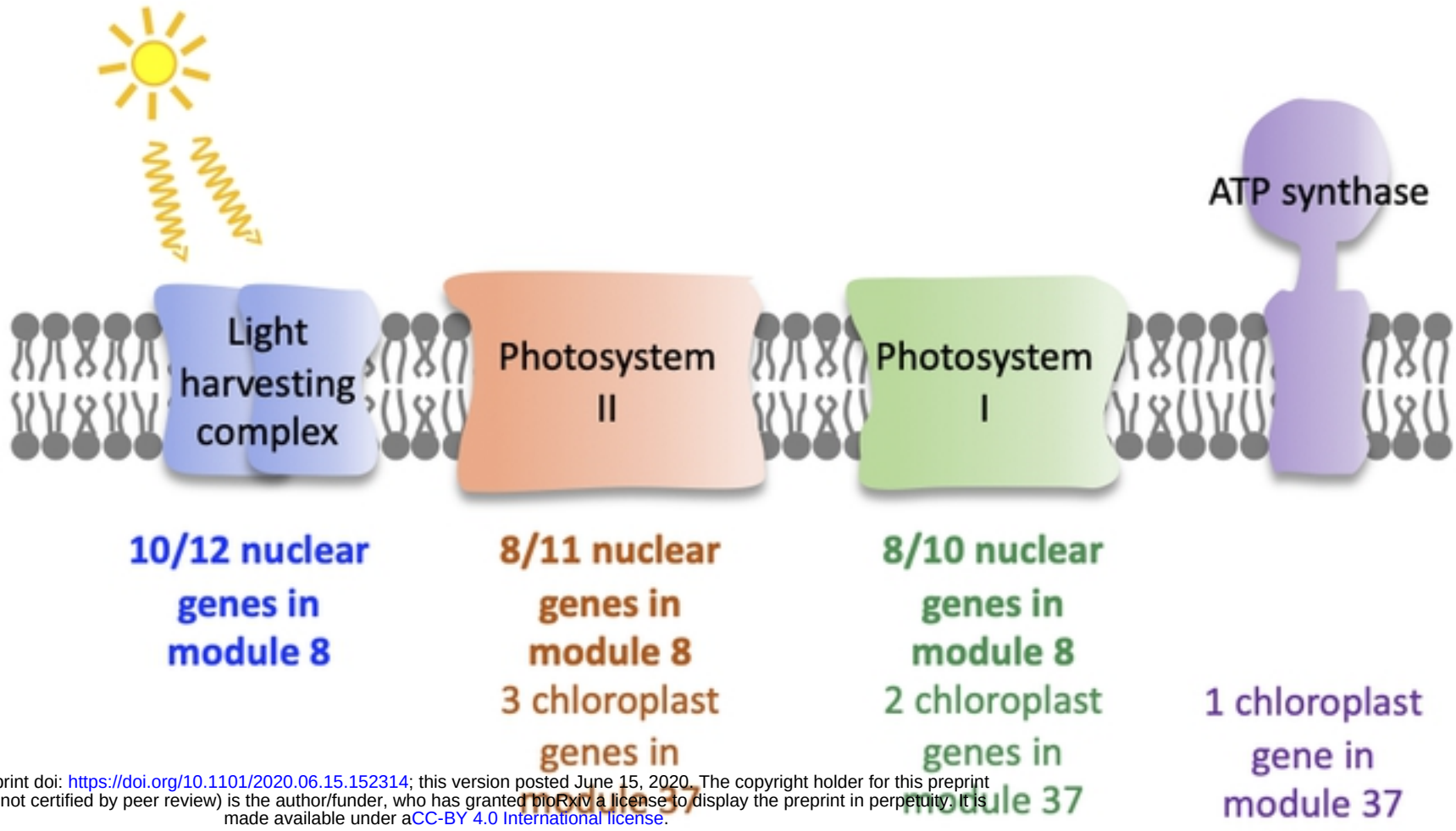
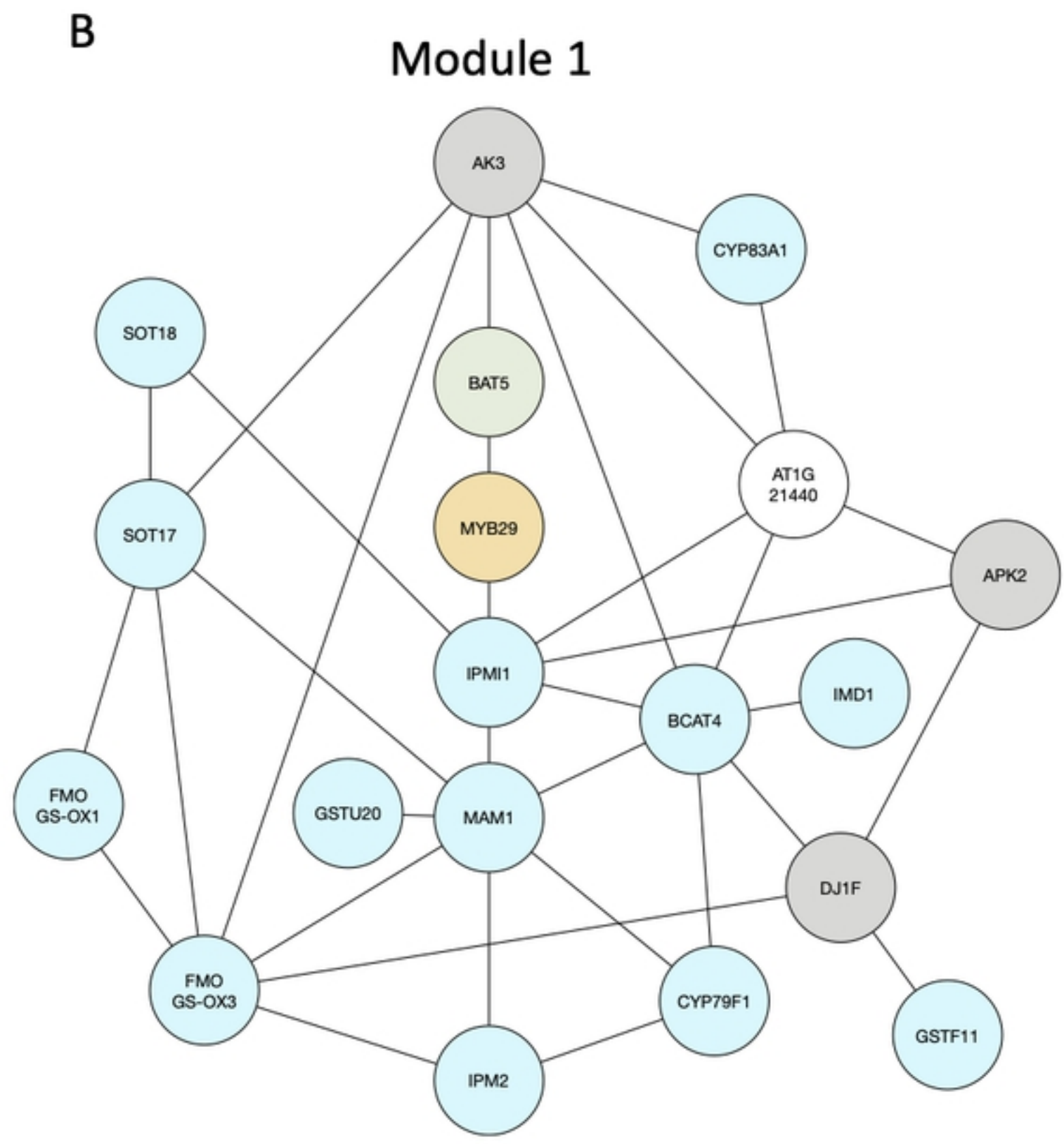


Fig2

A



B



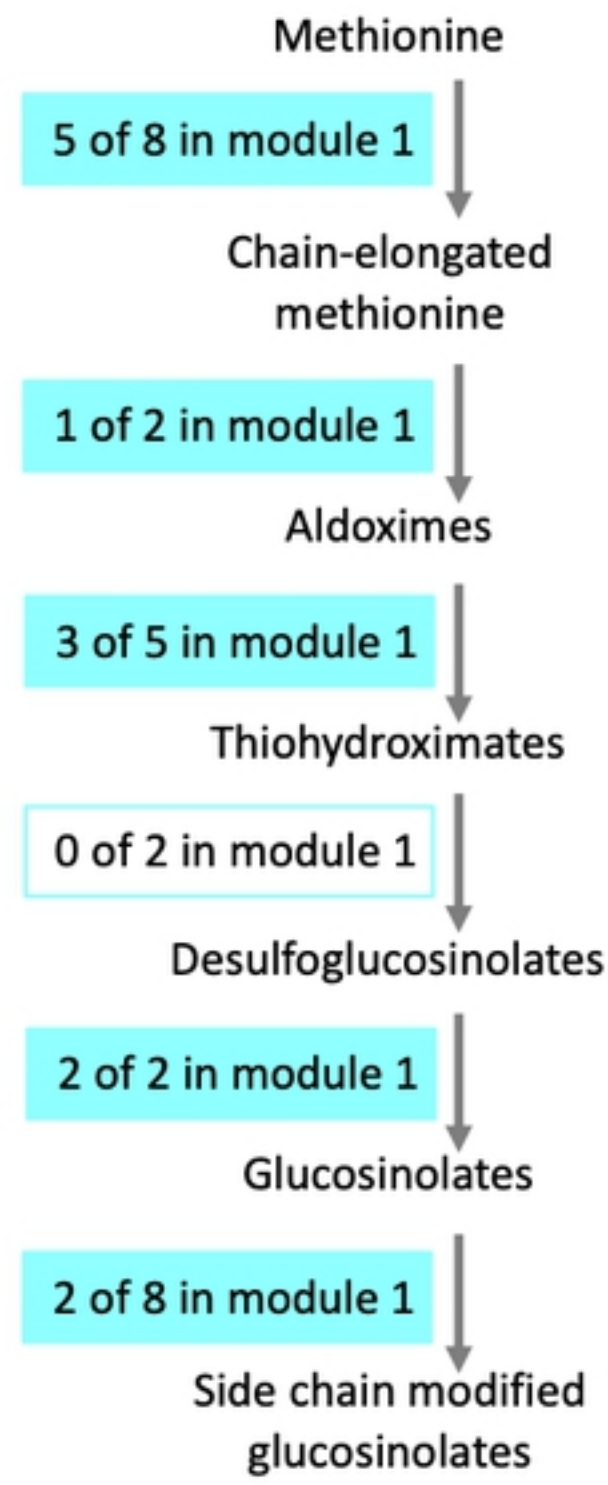
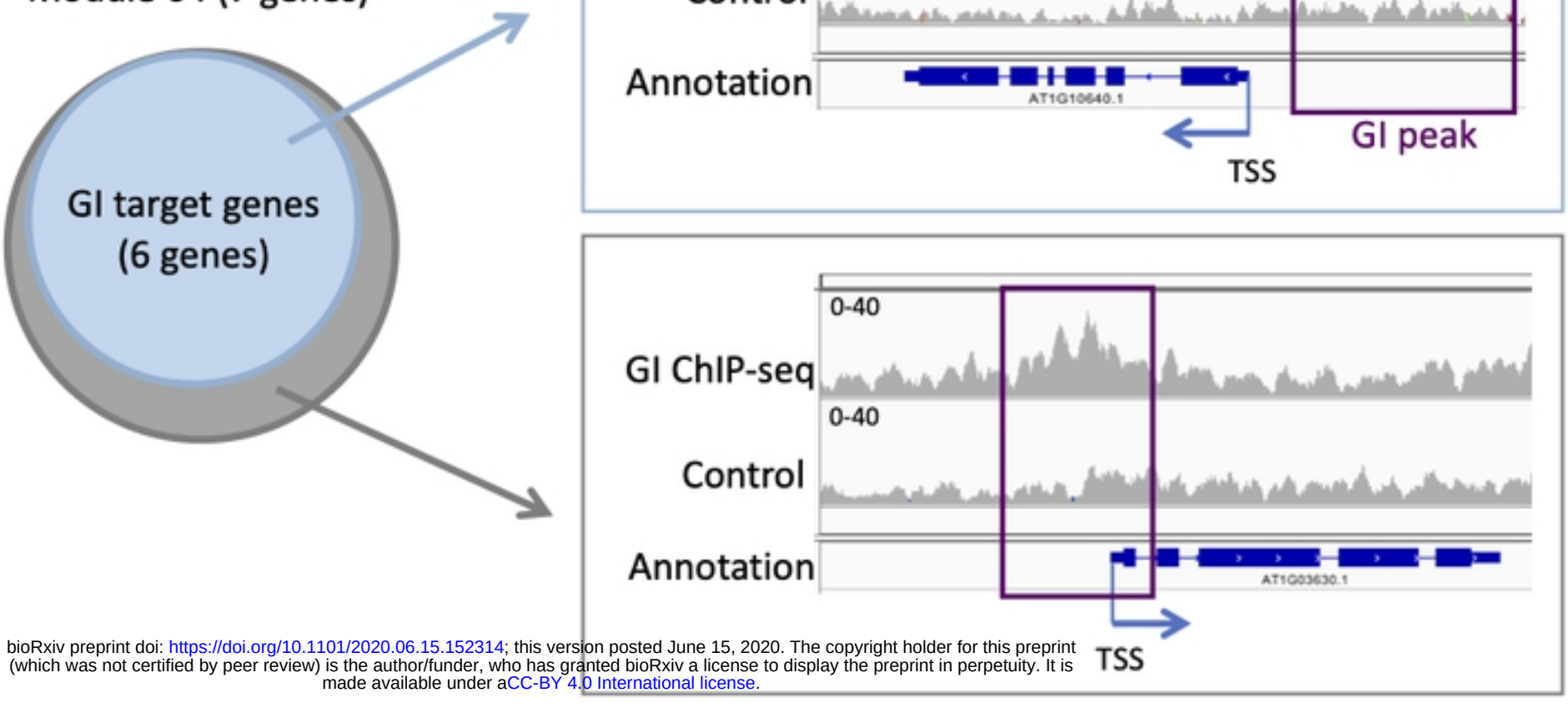


Fig3

A



B

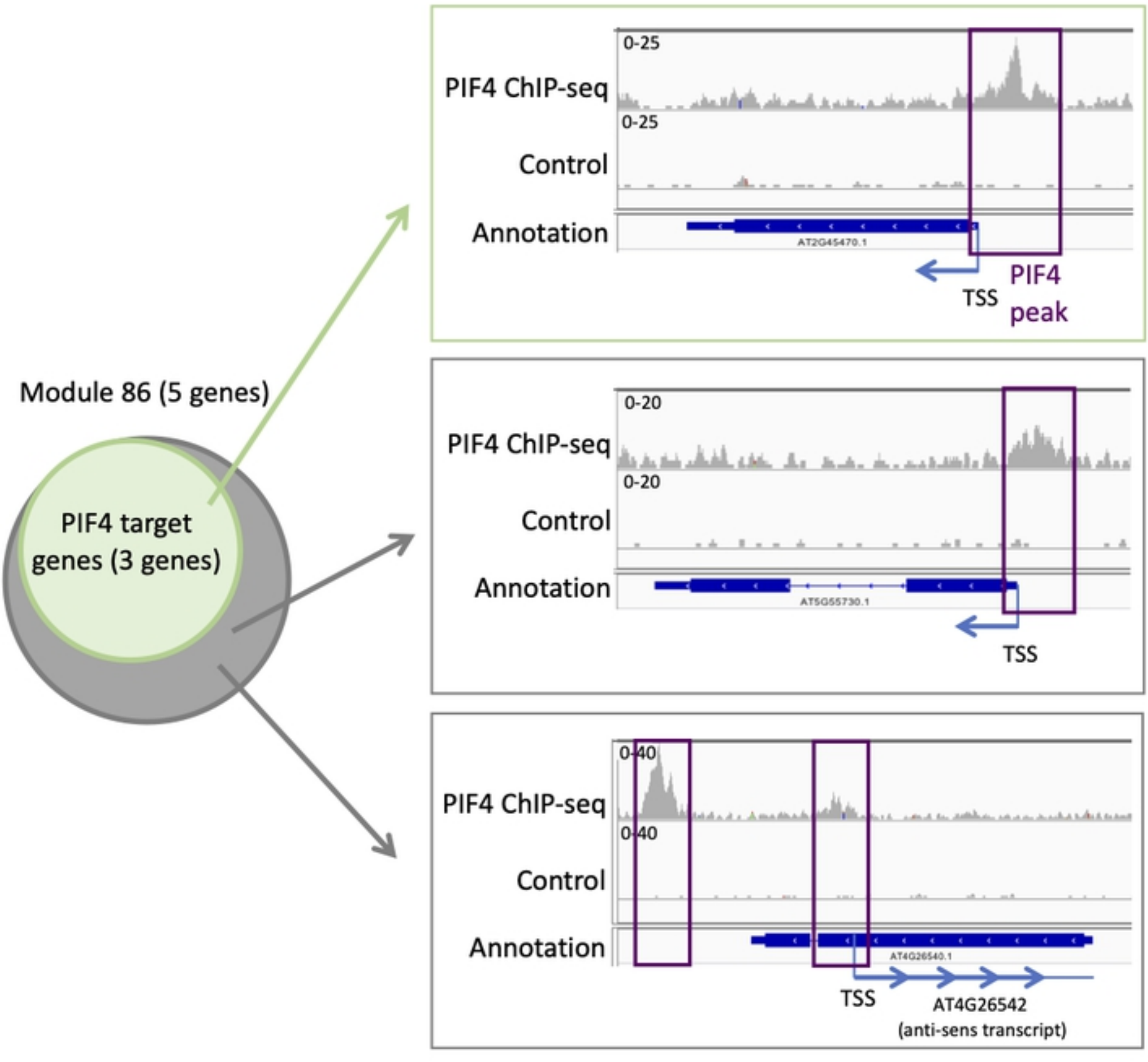
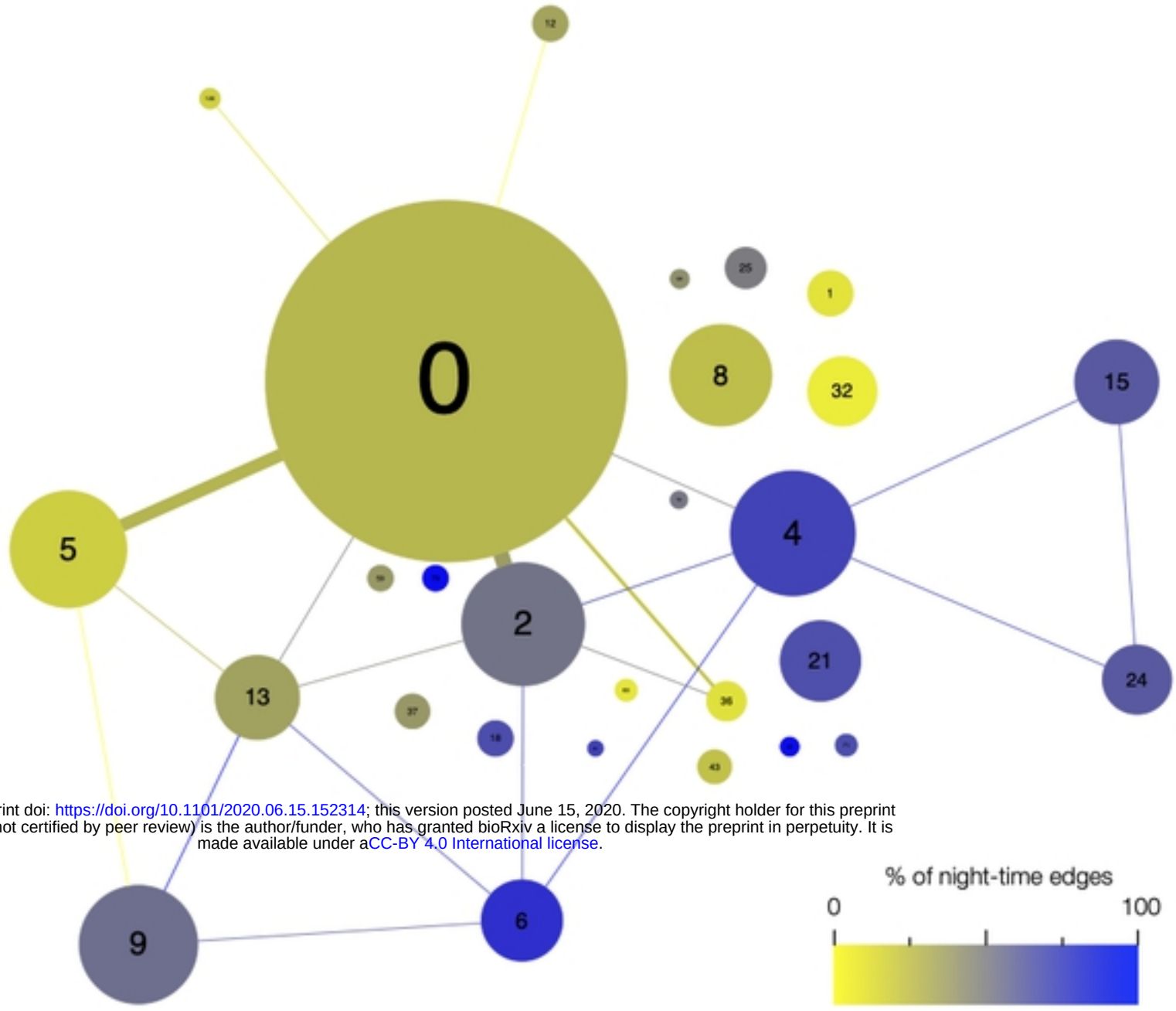


Fig4

A



B

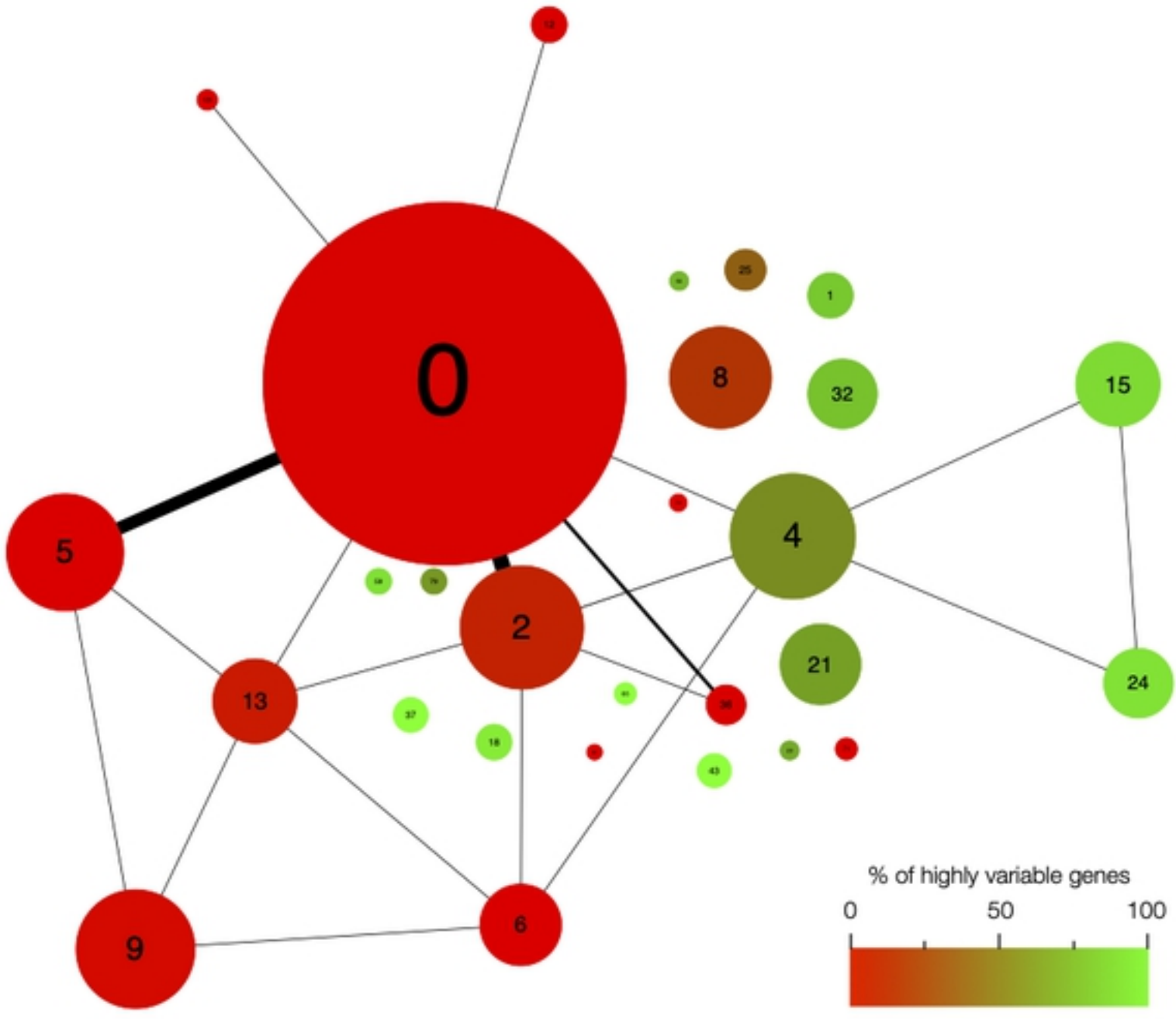


Fig5