










## Article

# Gene Functional Networks from Time Expression Profiles: A Constructive Approach Demonstrated in Chili Pepper (*Capsicum annuum* L.)

Alan Flores-Díaz <sup>1,†</sup>, Christian Escoto-Sandoval <sup>1,†</sup>, Felipe Cervantes-Hernández <sup>1,†</sup>, José J. Ordaz-Ortiz <sup>1</sup>, Corina Hayano-Kanashiro <sup>2</sup>, Humberto Reyes-Valdés <sup>3</sup>, Ana Garcés-Claver <sup>4</sup>, Neftalí Ochoa-Alejo <sup>5</sup> and Octavio Martínez <sup>1,\*</sup>

- <sup>1</sup> Unidad de Genómica Avanzada (Langebio), Centro de Investigación y de Estudios Avanzados del Instituto Politécnico Nacional (Cinvestav), Irapuato 36824, Mexico
  - <sup>2</sup> Departamento de Investigaciones Científicas y Tecnológicas de la Universidad de Sonora, Hermosillo 83000, Mexico
  - <sup>3</sup> Department of Plant Breeding, Universidad Autónoma Agraria Antonio Narro, Saltillo 25315, Mexico
  - <sup>4</sup> Unidad de Hortofruticultura, Centro de Investigación y Tecnología Agroalimentaria de Aragón, Instituto Agroalimentario de Aragón-IA2 (CITA-Universidad de Zaragoza), 50059 Zaragoza, Spain
  - <sup>5</sup> Departamento de Ingeniería Genética, Centro de Investigación y de Estudios Avanzados del Instituto Politécnico Nacional (Cinvestav), Irapuato 36824, Mexico
- \* Correspondence: octavio.martinez@cinvestav.mx  
† These authors contributed equally to this work.



**Citation:** Flores-Díaz, A.; Escoto-Sandoval, C.; Cervantes-Hernández, F.; Ordaz-Ortiz, J.J.; Hayano-Kanashiro, C.; Reyes-Valdés, H.; Garcés-Claver, A.; Ochoa-Alejo, N.; Martínez, O. Gene Functional Networks from Time Expression Profiles: A Constructive Approach Demonstrated in Chili Pepper (*Capsicum annuum* L.). *Plants* **2023**, *12*, 1148. <https://doi.org/10.3390/plants12051148>

Academic Editor: Alexey A. Dmitriev

Received: 31 January 2023

Revised: 20 February 2023

Accepted: 27 February 2023

Published: 3 March 2023



**Copyright:** © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

**Abstract:** Gene co-expression networks are powerful tools to understand functional interactions between genes. However, large co-expression networks are difficult to interpret and do not guarantee that the relations found will be true for different genotypes. Statistically verified time expression profiles give information about significant changes in expressions through time, and genes with highly correlated time expression profiles, which are annotated in the same biological process, are likely to be functionally connected. A method to obtain robust networks of functionally related genes will be useful to understand the complexity of the transcriptome, leading to biologically relevant insights. We present an algorithm to construct gene functional networks for genes annotated in a given biological process or other aspects of interest. We assume that there are genome-wide time expression profiles for a set of representative genotypes of the species of interest. The method is based on the correlation of time expression profiles, bound by a set of thresholds that assure both, a given false discovery rate, and the discard of correlation outliers. The novelty of the method consists in that a gene expression relation must be repeatedly found in a given set of independent genotypes to be considered valid. This automatically discards relations particular to specific genotypes, assuring a network robustness, which can be set a priori. Additionally, we present an algorithm to find transcription factors candidates for regulating hub genes within a network. The algorithms are demonstrated with data from a large experiment studying gene expression during the development of the fruit in a diverse set of chili pepper genotypes. The algorithm is implemented and demonstrated in a new version of the publicly available R package “Salsa” (version 1.0).

**Keywords:** gene expression; RNA-Seq; time expression profile; fruit development; *Capsicum*

## 1. Introduction

Gene co-expression networks (GCN) [1] are graphs that connect genes (vertices) by lines (edges), which represent the fact that genes were co-expressed in space (cell, tissue, organ, etc.), time (for example, during development), or under particular environmental conditions. In general, graph theory has been found to be useful to analyze biological networks [2], and, in particular, GCN have been used to quantify relationships between

co-expression, co-regulation, and gene function [3], and some functional modules have been found to be conserved in very distant species, as in yeast and humans [4].

GCN can be estimated using different statistical approaches, such as, for example, Pearson's correlation, graphical Gaussian models, Bayesian approaches, or methods based on mutual information [5]. Aside the statistical significance of the relations estimated, the main point of GCN is to summarize information about gene modules associated with particular biological processes, providing novel insights into the system-level understanding of cellular processes [6]. This process could lead to unraveling gene function in crops [7], which, in turn, will allow the efficient application of molecular-aided methods of plant breeding [8].

Nonetheless, when the number of genes and relations in a GCN are large, its plot resembles a "hairy ball" [9], and thus could be almost useless for direct interpretation, given the human limits for the perception of relations [10]. In Supplementary SI.1, we present an example of a large GCN.

Beyond a GCN—which will simply show significant co-expression of genes in a single genotype—we must look for gene functional networks (GFN), which will display robust functional relations between genes. An ideal GFN must show causal relationships between genes, which follow an alike time expression profile and are robust in the sense of being independently repeated in various genotypes.

Causality, our first requirement for a GFN, demands a deep knowledge of the process involved, which is rarely found out of model organisms, as *Arabidopsis* for plants or mice for animals. However, gene ontology (GO) [11] groups genes into biological process (BP), and gene orthology permits us to infer genes with the same BP in non-model species, allowing the prediction of GO BP from temporal gene expression patterns, even for non-annotated genes [12]. It appears reasonable to assign causality to well annotated genes that share identical BP in the GO, and we will assume that as a fact. However, we must take into account that the gene to BP relation is not one to one, i.e., a gene is usually annotated in different BPs, forming directed acyclic graphs, and this, as well as the quality of annotations, must be taken into account to avoid mistakes [13].

Asking for time co-expression of genes, our second requirement for a GFN, limits our definition to cases where gene expression was estimated at a grid of increasing times, presenting significant changes in at least one pair of neighbor times. This requirement excludes genes that, even when causally participating in a BP, do not show significant expression changes during the whole interval observed, and, thus, these could be labeled as "housekeeping" genes [14] for the BP of interest. On the other hand, it is axiomatic that genes with highly similar expression profiles are likely to be regulated via the same mechanisms [3,15], and thus this fact is integrated in our GFN definition.

Accuracy, scalability, robustness and reproducibility are important qualities of GFN [16]. GFN can vary depending on the genotype, for example, GFN including WRKY transcription factors in rice show variation between the Indica and Japonica genotypes [17]. In our GFN definition, robustness is evaluated by the number of times that a relation between two genes is independently detected in different genotypes at a fixed false discovery rate (FDR) [18] selected by the researcher. If a gene relation is found only in a single accession or genotype, but absent in all the others, that relation cannot be part of a general GFN for the species of interest, and it is more likely to occur by chance or by particularities of that genotype. In contrast, gene relations that are found in all independent genotypes studied are likely to be causal and must form part of the GFN for the species of interest. We designed here a measure of robustness for gene relations, which allows the researcher to set a level of confidence for the relations to be included in the network (see Section 2).

The main goal of our work was to obtain a method to estimate robust and small GFN modules for specific biological processes, using independent time expression profiles from different genotypes. Such modules can then be linked together into larger networks by the genes shared between them. The novelty of this approach consists in the design and implementation of an algorithm that uses standardized expression profiles (SEPs) [19,20]

from independent experiments. This allows us to judge and tune the robustness of the GFN modules, assuring that they show only relations, which are confirmed in a set of representative and independent genotypes. Later, by orderly linking together GFN modules, we avoid the “hairy ball” effect, obtaining a network that is easier to interpret than the ones obtained by methods that estimate a whole network from all genes at once, as is the case with ARACNE [21] and other approaches [22].

A secondary goal of our approach was to formalize an algorithm to estimate transcription factor (TF) candidates to be regulating genes present in a GFN. Briefly, if no TFs are included in a GFN, the algorithm looks for TF genes, which have highly correlated time expression profiles with the genes in the GFN, demanding also that such correlation will be present in various independent genotypes. If found, those TFs are presented as candidates to regulate the gene or genes of interest within the GFN. That method has successfully applied to recover TFs that regulate the *AT3* gene, involved in capsaicin’s biosynthesis in chili pepper [23].

The methods are demonstrated in a collection of transcriptomes obtained from 12 accessions during the development of the chili pepper fruits [19], and they are implemented in a new version of the publicly available R package “*Salsa*” [24].

## 2. Methods

### 2.1. Prerequisites

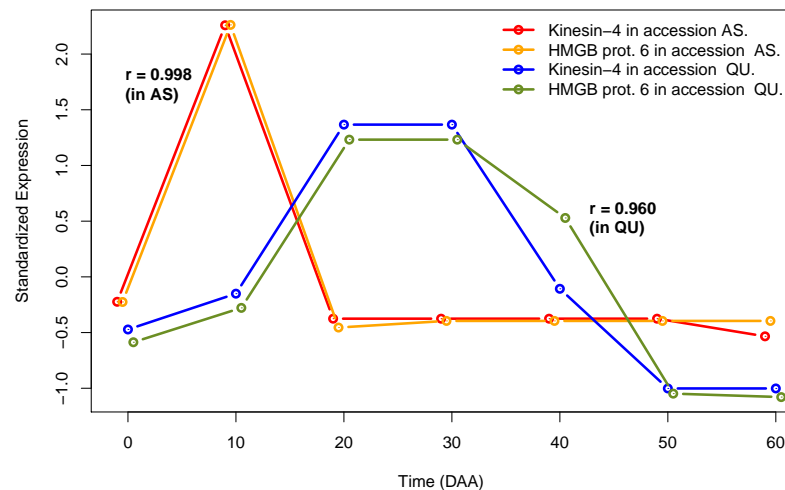
To apply our method to estimate GFN, the researcher must have genome-wide gene expression profiles, resulting from a time-course experiment in different genotypes of the species of interest. Even when such gene expression profiles could be the result of microarray experiments—as exemplified in [25]—here we assume that the data were obtained with the RNA-Seq technology [26,27], which in general appears to have a better performance for GCN estimation than microarrays [22].

Given that both the genotype [28], as well as the environment [29] and their interaction [30], can influence gene expression, the generality of GFN can be assured only for the diverse genotypes and environments where it was tested. For this reason, our method demands that the expression profiles must be obtained from a set of different genotypes in independent experiments performed under the same environmental conditions. If those requirements are fulfilled, then our method will give robust GFNs with the stringency that can be fixed by the researcher.

### 2.2. Congruent Gene Correlations

Given that our method is based on congruent gene expression profiles, which are repeated in independent genotypes, Figure 1 presents an illustrative example of that case.

Figure 1 presents standardized expression profiles (SEPs) [19,20] for two genes (Kinesin-4 and HMGB prot. 6) in two different chili pepper accessions, “Ancho San Luis”, a domesticated accession, and the wild accession, “Piquín Quéretaro”, with keys “AS” and “QU” in the figure legend. We can see how these two genes present a highly congruent expression behavior within each one of the two independent genotypes (AS and QU), a fact that is reflected in very high values of Pearson’s correlations, 0.998 for AS and 0.960 for QU. Additionally, these correlation coefficients are highly significant, having  $p$ -values  $< 6 \times 10^{-4}$ . In contrast, the two highly correlated gene pairs show distinct and uncorrelated expression profiles in the two different accessions (see Supplementary SI.2.1). We will see that, if correlation between the same pair of genes is found in various independent accessions, this strongly suggests that such a co-expressing gene pair could be indeed functionally related—even when the genes could have different and uncorrelated time expression profiles, as shown in Figure 1.



**Figure 1.** Example of SEPs for two highly correlated genes in two independent genotypes.

### 2.3. Robustness of a Relation between Genes

Any method to estimate GFN must give an associated measure of robustness to judge how strong the gene relations displayed in the network are. A now customary method to protect against false positives when many statistical tests are performed is the transformation of  $p$ -values into  $q$ -values, which allows the setting of a False Discovery Rate (FDR) threshold [18].

Our method to predict the minimum robustness of a GFN couples the setting of a FDR threshold with the fact that the relation must be found in various independent genotypes. Assume, for example, that we have a total of  $a = 12$  independent genotypes, representing a fair sample of the genetic diversity existent in the species of interest, and that we set a maximum FDR of 5%. Then, the probability of finding a fortuitous relation between two genes in a particular genotype will be 5%. However, given that the genotypes are evaluated in independent experiments, we can ask that the relation must be corroborated with the same FDR in various genotypes to be taken into account in the GFN; we can ask, for example, that the relation must be repeatedly found in  $x = 2, 3, \dots, a$ . In Supplementary SI.2.2, we show that the probability of finding the same gene relation independently repeated in  $x$  genotypes depends on the binomial distribution, and it rapidly decreases when  $x$  increases, being minimal at  $x = a$ . For example, by fixing a FDR of 5%, if we ask that a relation must be found in a minimum of  $x = 5$  genotypes, then the probability of that relation being by pure chance is small, approximately 0.0002 or one in 5436. If we want to be more strict, we could ask that the relation must be repeatedly found in the whole set of genotypes, i.e., in the  $x = 12$  accessions. In that last case, the probability of that relation being random is very small,  $\approx 2.4 \times 10^{-16}$ ; this is less than one in many billions and, for all practical purposes, we could be “almost sure” that the relation is real and not given by chance.

In summary, when a reasonable number of independent genotypes are available, the researcher could obtain robust GFN values, which very likely represent true functional relations between genes.

### 2.4. The “Gene2Gene” Algorithm

The objective of the *Gene2Gene* algorithm is to obtain a GFN. As mentioned before, the algorithm assumes the existence of time expression profiles in a group of genotypes, and it needs as input:

1. The genes of interest, say a set of “ $g$ ” gene identifiers.
2. The genotypes (or accessions) where the estimation will be performed, i.e., a set of “ $a$ ” genotypes.
3. A threshold for the FDR, “ $f$ ”.
4. A threshold for the minimum value of Pearson’s determination coefficient, “ $mr^2$ ”.

5. A threshold to eliminate putative regression outliers, “ $q$ ”.
6. The minimum number of genotypes where the gene relation must be found to be reported in the output, “ $x$ ” ( $x \leq a$ ).

Given the input, in a first step, the algorithm estimates and tests all  $g(g-1)/2$  pairs of correlations between the  $g$  genes of interest in each one of the  $a$  genotypes. In a second step, the algorithm selects a set of genes that fulfill all criteria given in the input, and then it outputs all the relevant statistics to construct a GFN. The output could be null—when there are no genes that fulfill the input thresholds or a set of structures that define a GFN.

Even when the *Gene2Gene* algorithm has been implemented in R [31] within the package *Salsa* [24], specifically for data from a *Capsicum* experiment [19], it can be adapted to process any set of gene expression profiles. Details of the *Gene2Gene* algorithm are given in Supplementary SI.2.3.

### 2.5. Constructing GFN: A Bottom-Up Approach

One important property of biological networks is their modular structure. In many relevant cases, as in *Arabidopsis* [32], humans [33], or yeast [33], genes within functional modules are densely connected, while, in contrast, connections between modules can be sparse or inexistent.

It is certainly possible to construct a general gene co-expression network with data for all genes in the genome, such as, for example, the one constructed for *Arabidopsis* [32], and then one can detect and order the modularity, as reviewed in [34]. Modularity can be based on different criteria, such as, for example, biological process, cell component, molecular function or metabolic pathway, and modules from different criteria will in general present highly complex and difficult to interpret relations. Additionally, co-expression networks estimated for all genes in the genome will include not identified or not annotated genes, which will complicate interpretation even further.

We suggest that a bottom-up approach, beginning with a careful selection of a relatively small set of genes annotated in a process of interest, and progressing to networks of different modules, i.e., sets of GFN or “Meta Networks” (MN), will be easier to interpret and could lead to relevant and immediate discoveries. Additionally, with our method, the generality (robustness) of each GFN can be statistically evaluated, and the analysis could be performed with different stringency levels.

The same bottom-up approach could be employed with different criteria, but not necessarily with genes that share a known annotation in a given biological process. For example, assume that there is a set of genes that are of special interest for a research group, without being annotated in any particular aspect. In that case, the *Gene2Gene* algorithm can be run with that set of genes to determine if they form a significant gene expression network. See supplementary SI.2.4 and SI.2.5.

## 3. Results

### 3.1. Data Analyses

As an example of the application of the *Gene2Gene* algorithm, we present the estimation of GFN for three biological processes (BP). The original data were duplicated RNA-Seq libraries from fruits of 12 genotypes (accessions) of chili pepper (*Capsicum annuum* L.) sampled in each case at seven times of development. These genome-wide data were annotated with GO categories [11], and they were processed to give standardized time expression profiles (SEPs) [19,20]. All data and functions employed here to perform the analyses are implemented in a new version of the publicly available R package “*Salsa*” [24], and details of the computations and results are presented in supplementary SI.3.

#### 3.1.1. GFN for Three Biological Processes (BPs)

We selected the BP “*cell cycle*”, “*reproduction*” and “*response to virus*”, which are abbreviated here as “*celcy*”, “*rep*” and “*vir*”, respectively. The parameters for the *Gene2Gene* algorithm were set to a FDR of 10%, a minimum of  $\hat{r}^2$  of 0.7, and a threshold to eliminate

5% of putative regression outliers. We also asked that the output will give gene relations that were repeated in all the 12 genotypes in the collection. Table 1 summarizes the results obtained for the three BPs.

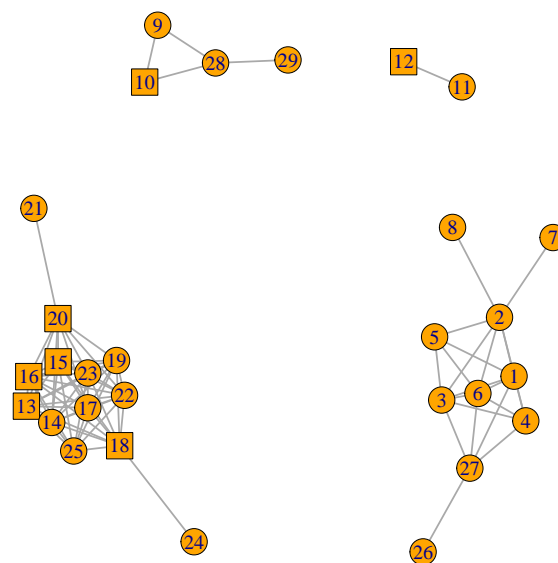
**Table 1.** Summary of results for GFN estimation.

BP	In	Out	% Out/In	Con.	$\bar{r}$	$\bar{q}$
<i>celcy</i>	352	29	8	81	0.9859	0.00729
<i>rep</i>	228	10	4	29	0.9934	0.00257
<i>vir</i>	33	4	12	6	0.9936	0.00126

BP—Abbreviation for the Biological Process, In—Number of genes in the input, Out—Number of genes in the output, Con.—Number of connections in the GFN,  $\bar{r}$ —Average value of  $\hat{r}$  in all connections.  $\bar{q}$ —Average value of  $\hat{q}$  in all connections.

In Table 1, we can see that the algorithm began with 352 genes that were annotated in the *celcy* BP, but, from all the  $(352 \times (351 - 1))/2 = 61,600$  pair correlations tested per genotype, only 81 of them, involving 29 genes, passed the thresholds imposed and were repeated consistently in the full set formed by the 12 genotypes; thus, only approximately 8% of the total set of genes is corroborated to have co-expressions. Even when the algorithm was run with parameters of moderate stringency, i.e., a minimum  $\hat{r}^2 = 0.7$ —which implies a positive value of  $\hat{r} \geq 0.83$ , and a FDR of 10% ( $\hat{q} \leq 0.1$ ), the actual values of the averages of  $\hat{r}$  and  $\hat{q}$ , presented as  $\bar{r}$  and  $\bar{q}$  respectively in Table 1, are highly superior to the threshold values set;  $\bar{r} = 0.9859$  is much larger than the threshold of  $\hat{r} \geq 0.83$ , while the values of  $\bar{q} = 0.00729$  is much smaller than the threshold of  $\hat{q} \leq 0.1$  asked to filter by FDR. The same behavior on the comparisons between realized values of  $\hat{r}$  and  $\hat{q}$  can be observed in the other two GFNs, *rep* and *vir* in Table 1, i.e., the average values are far away from the thresholds set. This is a direct result of the fact that we are asking that the correlations must be present in the full set of 12 different genotypes, and this highly stringent criterion is likely to give only highly congruent and truthful relations that happen in domesticated, wild, as well as  $F_1$  crosses of *Capsicum* genotypes [15,19].

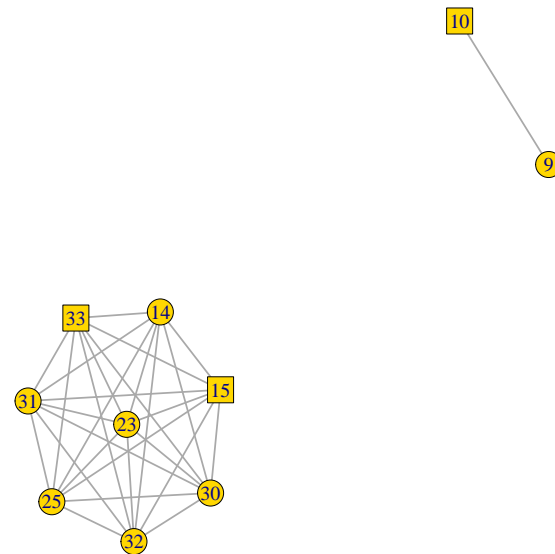
Figures 2–4 present the plots for the *celcy*, *rep*, and *vir* GFN, respectively. In these three figures, transcription factors (TF) are represented by squares, while other genes are represented by circles, and the identity of each one of these genes is presented in Supplementary SI.5.



**Figure 2.** Cell Cycle (*celcy*) GFN. 29 genes and 81 pair connections in four disconnected sub-graphs. TF represented by squares, while other genes are represented by circles (Supplementary SI.5 for gene descriptions).

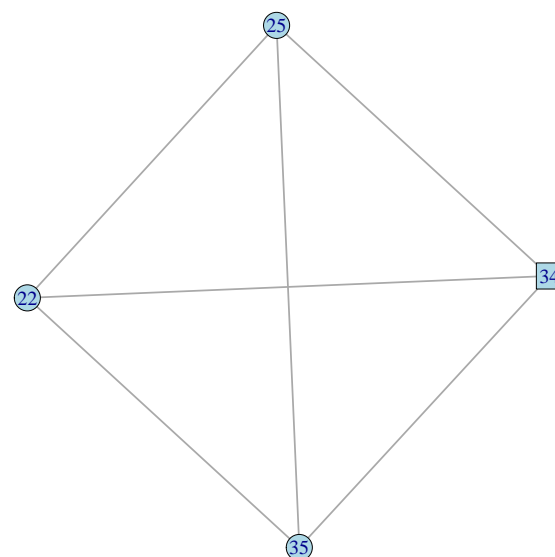
In Figure 2 we can see the 29 genes form four disconnected sub-graphs. This indicates that there are four contrasting time expression profiles in the 29 genes annotated in *celcy*; the first groups four genes in the upper left hand side corner of the figure and follows that corner in clock-wise direction, a second group includes two genes, a third one includes 10 genes, and a fourth one includes 13 genes in the down left hand side corner.

In Figure 3, we see that the 10 genes of the *rep* GFN are organized into two disconnected sub-graphs. The largest, at the lower left hand side corner, contains eight interconnected genes linked by 28 connections, and the second, at the upper right side corner, is formed by only two genes with one connection.



**Figure 3.** Reproduction (*rep*) GFN. 10 genes and 29 pair connections in two disconnected sub-graphs. TF represented by squares, while other genes are represented by circles (Supplementary SI.5 for gene descriptions).

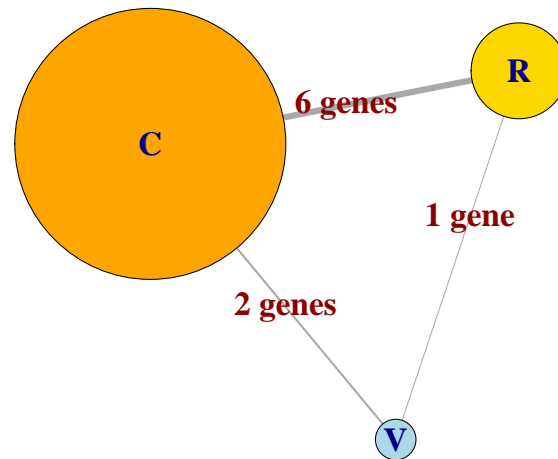
In Figure 4 we see that the four genes of the *vir* GFN are organized into one fully connected graph, with all the six possible connections between those four genes.



**Figure 4.** Response to virus (*vir*) GFN. Four genes and six connections in a fully connected graph. TF is represented by squares, while other genes are represented by circles (Supplementary SI.5 for gene descriptions).

### 3.1.2. A Meta-Network

Having estimated GFN for three processes, *celcy*, *rep*, and *vir*, we can now note that those three networks share some common genes, and thus they can be linked to form a meta-network (MN). Figure 5 shows a schematic representation of the MN, where each GFN is represented by a single circle, and the number of genes shared by pairs of GFN is given at the links between GFN.



**Figure 5.** Schematic representation of the MN formed by the GFN *celcy*, *rep* and *vir*, represented as the circles labeled as C, R and V, respectively. Sizes of circles are proportional to the number of genes at each GFN and links are annotated with the number of genes shared between pairs of GFN.

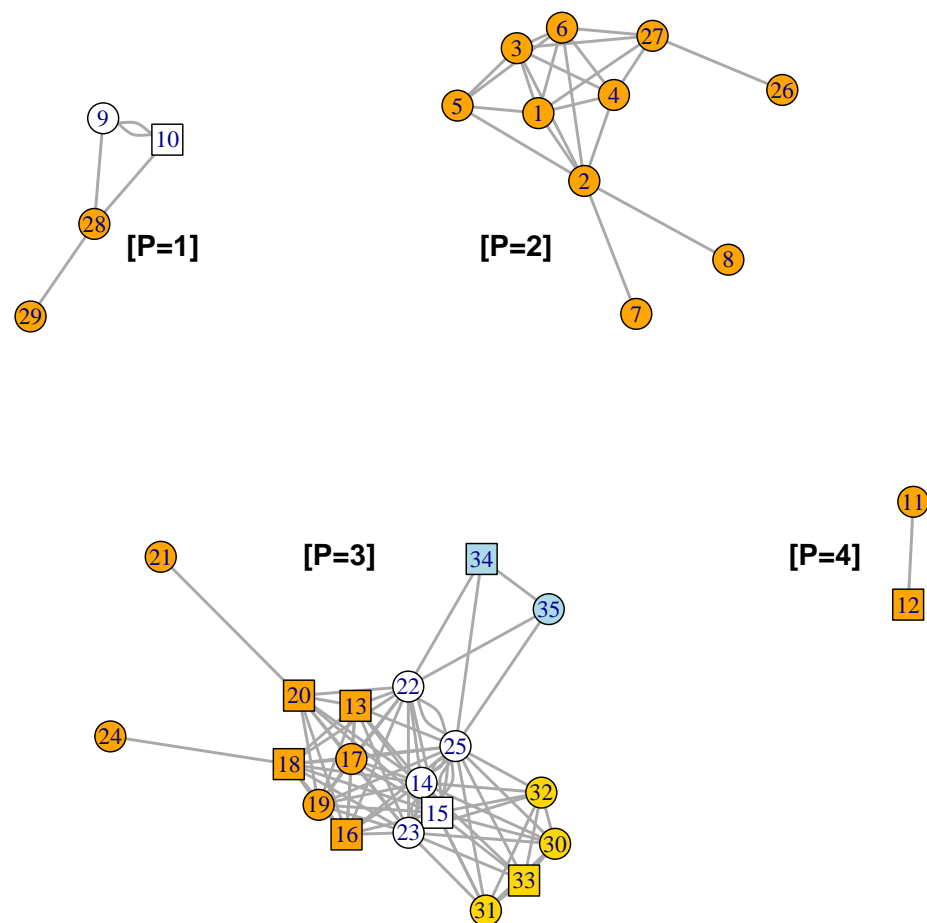
In Figure 5, we can appreciate the differences in size of the three GFN that form the MN, which include 29, 10, and four genes for the *celcy* (C), *rep* (R), and *vir* (V) GFN, respectively. We can also see that the three GFNs are inter-connected by sharing some genes between the GFN; six between *celcy* and *rep*, two between *celcy* and *vir*, and one between *rep* and *vir*. This last fact indicates that cell cycle (*celcy*), reproduction (*rep*), and response to virus (*vir*) in chili pepper are inter-dependent biological processes (BP).

Figures 2–4 presented the graphs of the three GFNs, coloring genes in orange for the *celcy*, in golden yellow for the *rep*, and in blue for the *vir* GFN. Figure 6 presents the MN formed by the three GFNs, following that color scheme, but changing to white the color of genes shared by two or more of the GFN.

Before, in Figures 2 and 3, we have seen that the GFN for *celcy* and *rep* included disconnected sub-networks; four for *celcy* and two for *rep*, while the GFN for *vir* presented a single fully connected graph (Figure 4). Disconnected sub-graph within a single GFN indicates that there are heterogeneous gene expression profiles, each one of them including a subset of the genes in the corresponding GFN.

In Figure 6, we can appreciate that the MN is formed by four disconnected sub-graphs, annotated with labels [P = 1] to [P = 4]. These four disconnected sub-graphs include highly correlated genes, which present characteristic gene expression profiles. Figure 7 presents the plots of the average standardized expression profiles (SEPs) on time for gene patterns [P = 1] to [P = 4], in panels (a) to (d), respectively. In these four panels, grey lines represent the average in the 12 accessions—which include two  $F_1$  crosses between a D and a W genotypes [19], while red and blue lines represent the average in the six domesticated (D) and four wild (W) accessions, respectively. Thin vertical lines at each time point give the 95% confidence interval (CI) for the mean of standardized expression at each time and for each one of the groups.





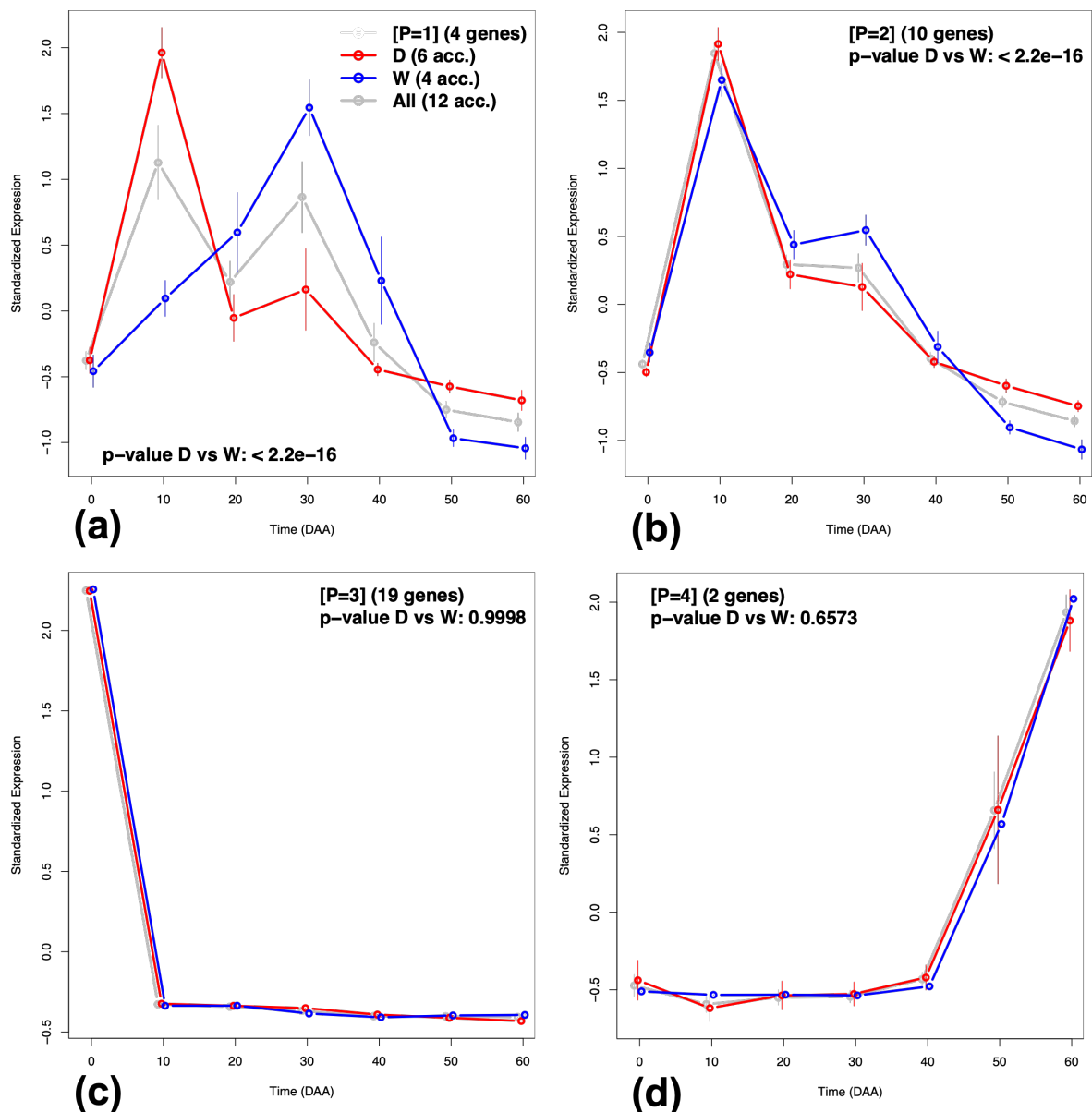
**Figure 6.** Graph of the MN formed by the 3 GFN. Genes found exclusively in BP *celcy* are colored in orange, the ones exclusively in *rep* are colored in golden yellow, and the ones exclusively in *vir* are colored in blue, while genes shared by two or more BP are in white. Disconnected sub-networks are annotated with labels [P = 1] to [P = 4] (Supplementary SI.5 for gene descriptions).

Figure 7a,b show that there are differences between SEP from the domesticated (D, in red) and wild (W, in blue) accessions—the 95% CI are not overlapped in some fruit developing times, while Figure 7c,d do not show such differences between the D and W groups.

In summary, we found that the three GFN are inter-connected forming a MN (Figure 6), which consist of four disconnected sub-graphs, each one with a typical time expression profile (Figure 7). This implies that neither gene annotation per BP, nor expression profile in isolation, can fully explain the complexity of the phenomenon observed; for a deep understanding, we need to consider, in concurrence, both the BP annotation, as well as the particularities of typical time expression profiles. Supplementary SI.3 presents the interpretation and biological implications for the development of chili pepper fruit of the four disconnected sub-networks that form the MN.

### 3.2. Finding Transcription Factors for Hub Genes

Transcription Factors (TF), in particular pioneer TF, are key regulators that initiate network changes [35]. In our GFN algorithm, only TF that are already among the genes initially input could be selected to form part of the output network. Nevertheless, using a similar approach to the one employed by the GFN algorithm, it is possible to obtain strong TF candidates to be regulating one or more of the genes that are already in a GFN. This is specially important for highly interconnected genes, i.e., “hub genes”, given that such TF could be playing a central role in the GFN regulation.



**Figure 7.** Average Standardized Expression Profiles (SEPs) patterns for genes included in the Meta-Network in Figure 6. (a) Pattern [P = 1]. (b) Pattern [P = 2]. (c) Pattern [P = 3]. (d) Pattern [P = 4].

The algorithm “Gene2TF” to perform the selection of TF candidates is an adaptation of the *Gene2Gene* function, which calculates pairwise correlations between a target gene and all TFs in all genotypes available, using thresholds as the ones previously described to assure both, a significant co-expression, as well as robustness. Robustness is guaranteed by asking that the co-expression must be present in different genotypes. Details of the *Gene2TF* algorithm are presented and exemplified in Supplementary SI.3.5.

In the example presented with our *Capsicum* data, the only sub-network within the MN that does not include TF is the one with expression pattern [P = 2] within the *celcy* GFN (pattern [P = 2] in Figure 6). That sub-network includes 10 genes, which have a maximum expression at 10 DAA (Figure 7b); however, none of those genes is a TF. To find TF candidates to be controlling this sub-network, we ran the *Gene2TF* for every gene with expression pattern [P = 2], with the same thresholds than the ones used to obtain the GFN, and then we asked that the TF candidates must be present in the 12 genotypes available. This procedure selected 10 TF candidates for the genes with pattern [P = 2]. Furthermore, a

new run of the GFN algorithm, including both the 10 genes with  $[P = 2]$ , as well as the 10 TF candidates found, produced a highly connected and robust network, which is presented in Supplementary SI.3.5. However, experimental evidence will be needed to corroborate or discard the control role of those TF candidates on the genes with pattern  $[P = 2]$  within the *celcy* GFN; see, for example, [36–38].

### 3.3. Sensitivity Analyses

To evaluate the acuity of the *Gene2Gene* algorithm to input parameters, we designed a fully balanced grid on the combinations of the algorithm's input thresholds, excepting the number of genotypes where the relations must be found, which was always set to the maximum level of stringency: 12 genotypes.

We ran the algorithm with the set of the *celcy* genes using the balanced grid of parameters, and we measured as output variables the number of genes and relations reported in the output. We found that the number of genes and relations reported by the algorithm have a strong linear correlation ( $\hat{r} \approx 0.99$ ;  $p < 2 \times 10^{-16}$ ), and the more important parameter to determine the number of genes and relations reported was the FDR threshold, followed by the minimum  $\hat{r}^2$ . The less important parameter that determines the numbers of genes and relations reported by the algorithm was the threshold to avoid regression outliers.

In summary, we conclude that the *Gene2Gene* algorithm allows the recovery of GFN, with consistency and flexibility for an ample set of reasonable combinations of stringency thresholds defined by the user. Details of the results of the sensitivity analyses are presented in Supplementary SI.3.2.

## 4. Discussion

Even when GCNs are powerful tools to disentangle the high complexity of the transcriptome, genome wide networks always result in graphs that are too large and complex for direct interpretation, as also happens with genome-scale metabolic networks [39]. In such large networks, it is possible to zoom in particular regions of interest, perhaps centered in a given gene, but, in that case, much of the information obtained will be wasted or not adequately noticed.

In contrast with the whole genome GCN, we advocate here for a bottom-up constructive approach, which begins with a small set of genes of interest and produces small and easy to interpret GFN. Furthermore, given that we ask that the relations must be repeatedly found in many genotypes, which ideally represent the diversity of the specie, the resulting GFN will have an evaluable level of robustness. We also argue that highly robust GFN are likely to present causal relations between genes because the majority of the not-causal relations, which could be either random or particular to a given genotype, are filtered out by the algorithm. Even when there are other statistical means to approach causality of GFN [40], we reason that repeatedly finding the same gene relation on independent experiments and genotypes gives the strongest rationale for causality. This common sense argument is analogous to the decision that could be taken about the fairness of a coin that repeatedly shows the same result when flipped; even when the coin could be fair, the likelihood of fairness compared with the likelihood of un-fairness rapidly decreases when we observe only one of the two possible results in many assays. In Supplementary SI.2.2 and SI.3, we further discuss the implication of our rule to infer causality for a gene relation.

To the best of our knowledge, our approach is the only one that explicitly takes into account repeatability in different genotypes to assure robustness of the inferred networks. For example, computational tools, such as WGCNA [41], DICER [42], CoXpress [43], or DiffCoEx [44], estimate networks, but without asking for repeatability in different genotypes so as to consider as valid each one of the relations found.

In the examples presented, we have seen that a single GFN could contain disconnected sub-graphs. The presence of such independent modules directly indicates that there are two or more dissimilar time expression profiles that share the same annotation, and that fact improves the understanding of the different processes within a single GFN.

On the other hand, we have also seen in the examples that two or more GFNs could be joined when they share genes, forming what, here, we call a “Meta-Network” (MN). Such joining process could proceed with as many GFNs as the researcher considers reasonable for a hierarchical ordering of BP. In the examples, we showed how the GFN obtained from the cell cycle (*celcy*), reproduction (*rep*), and response to virus (*vir*) BP could be joined in a single MN, whose biological relevance is discussed in Supplementary SI.3.3 and SI.3.4.

Furthermore, the *Gene2Gene* algorithm could be applied not only to a specific BP to obtain a GFN, but to any set of genes of particular interest of the researcher, for example, to genes annotated in an specific metabolic pathway, or to a set of lncRNA as done in [45], etc.

Besides, the *Gene2TF* algorithm allows the selection of TF candidates to be regulating any gene of interest (see Supplementary SI.3.5). Even when the correlation in time of expression of a target gene with a TF candidate in various genotypes could suggest the possibility of co-regulation [46,47], other bioinformatic or experimental means are needed to corroborate such a role [36–38].

One point worth discussing is the information given by genes that are excluded from a GFN when asking congruent correlation in a set of genotypes. For example, the first row in Table 1 shows that only approximately 8% (29/352) of the genes annotated in the *celcy* (cell cycle) BP are finally included in the GFN. This implies the existence of heterogeneity in the time expression patterns of the  $352 - 29 = 323$  excluded genes, which even when functionally annotated as related to cell cycle, present a diverse time expression profile among genotypes. Mining such specific time expression profiles could lead to a better understanding of the differences that exist among the set of tested genotypes—in this case with reference to cell cycle, but in general to any estimated GFN.

While in three out of four of the estimated sub-networks there is at least one TF that could be regulating expression (patterns [P = 1], [P = 3] and [P = 4] in Figure 6), in the fourth one (pattern [P = 2] in Figure 6), there is none. The inclusion of regulatory elements (TF) is of paramount importance to understand gene network changes [35], and thus all GFNs must include such elements—even when their functional role could be only putative. In this regard, the *Gene2TF* algorithm showed high efficiency by providing a set of 10 TF candidates that form a highly connected network with the genes previously found in the cell reproduction (*celcy*) GFN (Figure S12 in SI.3.5). These results give a solid working hypothesis to investigate the regulation of cell reproduction, a biological process of foremost practical, as well as theoretical importance, given its role in fruit size [48]—one of the key traits in fruit production.

As the final result of our method—applying the *Gene2Gene* and *Gene2TF* algorithms to the *Capsicum* data—we obtained a robust gene functional meta network (MN), linking three important biological processes and including 45 genes (Tables S6 and S7 in SI.5), of which 29 (64%) are TFs. This MN is easier to interpret than genome-wide networks because the researcher can focus on relations between specific processes of interest, with the additional advantage that sub-networks show particular time expression profiles (Figure 7). This approach can be directly applied using the *Salsa* software to other sets of *Capsicum* genes during fruit development, grouped by different criteria, or can also be implemented for other organisms or development processes by following the guidelines given in [20].

About the design of future experiments using our methodology, the *Gene2Gene* and *Gene2TF* algorithms are only applicable to development processes for which time expression profiles exist in a set of different genotypes. The first question is how many time points need to be fixed to estimate time expression profiles—which can always be transformed to SEPs [20]. In this regard, any case with less than five time points will be too flimsy to obtain useful correlation coefficients, and of course, more time points will give a more precise discrete description of the continuous changes of gene expression through time. It is important to take into account that each time point within each genotype will demand the construction and sequencing of a minimum of two RNA-Seq libraries.

The second issue in this regard concerns the nature and number of the genotypes to be used. In any real case, it will be impossible to include all the relevant genetic

diversity present in a crop or in a model organism. However, usefulness of the method will be maximized by selecting the largest number of almost independent and contrasting genotypes. In our case, we selected six D chili pepper accessions with contrasting fruit characteristics, four W accessions presenting also diversity in fruit as well as in geographic origin, and two crosses between a D and a W accessions to include  $F_1$  hybrids.

The function “`pred.error.g2g()`” in the *Salsa* package calculates the predicted or realized error in finding a correlation between two genes significant at a fixed false discovery rate (FDR). By exploring a grid of values for the total number of accessions available, the number of genotypes where the relation must be found to be considered in the GFN and the maximum FDR that could be tolerated, the researcher can reach a reasonable decision on the number of genotypes to be employed (see Supplementary SI.2.2).

## 5. Conclusions

The *Gene2Gene* and *Gene2TF* algorithms, now implemented and demonstrated in the R package *Salsa* (version 1.0), provide powerful tools to study gene co-expression of gene sets in a hierarchical way. When there are gene expression profiles from a representative set of genotypes, these tools allow the estimation of robust and easy to interpret GFNs (see also Supplementary SI.4).

We expect that the research community interested in fruit development (or other development processes) will adopt our methodology. We also anticipate that the results and functions already present in the *Salsa* R package can lead to further biologically relevant findings.

**Supplementary Materials:** Supporting information can be downloaded at: <https://www.mdpi.com/article/10.3390/plants12051148/s1>. Supporting information includes details of the algorithms presented as well as of the analyses of the *Capsicum* data [49–71]. Specific sections are referred as “supplementary SI.#”, where “#” is the section number in the supporting information file.

**Author Contributions:** Conceptualization, O.M. and N.O.-A.; methodology, J.J.O.-O., C.H.-K., H.R.-V. and A.G.-C.; software, A.F.-D., C.E.-S. and O.M.; validation, J.J.O.-O., C.H.-K., H.R.-V. and A.G.-C.; formal analysis, A.F.-D., C.E.-S., H.R.-V. and O.M.; investigation, N.O.-A., J.J.O.-O. and O.M.; resources, N.O.-A., C.H.-K. and H.R.-V.; data curation, A.F.-D., C.E.-S. and F.C.-H.; writing—original draft preparation, O.M.; writing—review and editing, J.J.O.-O., C.H.-K., H.R.-V., A.G.-C. and N.O.-A.; visualization, A.F.-D., C.E.-S. and F.C.-H.; supervision, J.J.O.-O., O.M. and N.O.-A.; project administration, N.O.-A.; funding acquisition, N.O.-A. and O.M. All authors have read and agreed to the published version of the manuscript.

**Funding:** This work was funded by the Consejo Nacional de Ciencia y Tecnología (Conacyt, México), “Fronteras de la Ciencia” project 1570. AF-D, CE-S and FC-H acknowledge Conacyt scholarship for PhD studies; grant numbers 700545, 630487, and 707294, respectively.

**Data Availability Statement:** Original RNA-Seq data have been deposited into NCBI’s Gene Expression Omnibus and are accessible through the GEO Series accession number GSE165448: <https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE165448> (accessed on 27 February 2023). All curated data and functions for the analyses and algorithms are publicly available in the R package “*Salsa*” (version 1.0): <https://zenodo.org/record/4767445#.YmAHvZPMLaY> (accessed on 27 February 2023).

**Acknowledgments:** We thank Magda L. Arce-Rodríguez, Fernando Hernández-Godínez and Fernando G. Razo-Mendivil for help in the acquisition and curation of the *Capsicum* plant materials.

**Conflicts of Interest:** The authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript; or in the decision to publish the results.

## Abbreviations

The following abbreviations are used in this manuscript:

BP	Biological Process
<i>celcy</i>	Cell cycle (BP)
D	Domesticated accession (genotype)
DAA	Days After Anthesis
FDR	False Discovery Rate
GCN	Gene Coexpression Network
GFN	Gene Functional Network
MN	Meta Network
<i>rep</i>	Reproduction (BP)
SEP	Standardized Expression Profile
TF	Transcription Factor
W	Wild accession (genotype)
<i>vir</i>	Response to virus (BP)

## References

1. Reverter, A.; Chan, E.K. Combining partial correlation and an information theory approach to the reversed engineering of gene co-expression networks. *Bioinformatics* **2008**, *24*, 2491–2497. [[CrossRef](#)]
2. Pavlopoulos, G.A.; Secrier, M.; Moschopoulos, C.N.; Soldatos, T.G.; Kossida, S.; Aerts, J.; Schneider, R.; Bagos, P.G. Using graph theory to analyze biological networks. *BioData Min.* **2011**, *4*, 10. [[CrossRef](#)] [[PubMed](#)]
3. Allocco, D.J.; Kohane, I.S.; Butte, A.J. Quantifying the relationship between co-expression, co-regulation and gene function. *BMC Bioinform.* **2004**, *5*, 18. [[CrossRef](#)] [[PubMed](#)]
4. Liu, W.; Li, L.; Ye, H.; Chen, H.; Shen, W.; Zhong, Y.; Tian, T.; He, H. From *Saccharomyces cerevisiae* to human: The important gene co-expression modules. *Biomed. Rep.* **2017**, *7*, 153–158. [[CrossRef](#)] [[PubMed](#)]
5. Lopez-Kleine, L.; Leal, L.; Lopez, C. Biostatistical approaches for the reconstruction of gene co-expression networks based on transcriptomic data. *Briefings Funct. Genom.* **2013**, *12*, 457–467. [[CrossRef](#)] [[PubMed](#)]
6. Aoki, K.; Ogata, Y.; Shibata, D. Approaches for extracting practical information from gene co-expression networks in plant biology. *Plant Cell Physiol.* **2007**, *48*, 381–390. [[CrossRef](#)]
7. Schaefer, R.J.; Michno, J.M.; Myers, C.L. Unraveling gene function in agricultural species using gene co-expression networks. *Biochim. Biophys. Acta (BBA)-Gene Regul. Mech.* **2017**, *1860*, 53–63. [[CrossRef](#)]
8. Xu, Y. *Molecular Plant Breeding*; CAB International: Wallingford, UK, 2010.
9. Serrano, M.a.; De Los Rios, P. Interfaces and the edge percolation map of random directed networks. *Phys. Rev. E* **2007**, *76*, 056121. [[CrossRef](#)]
10. Hafri, A.; Firestone, C. The perception of relations. *Trends Cogn. Sci.* **2021**, *25*, 475–492. [[CrossRef](#)]
11. Ashburner, M.; Ball, C.A.; Blake, J.A.; Botstein, D.; Butler, H.; Cherry, J.M.; Davis, A.P.; Dolinski, K.; Dwight, S.S.; Eppig, J.T.; et al. Gene ontology: Tool for the unification of biology. *Nat. Genet.* **2000**, *25*, 25–29. [[CrossRef](#)]
12. Lægreid, A.; Hvidsten, T.R.; Midelfart, H.; Komorowski, J.; Sandvik, A.K. Predicting gene ontology biological process from temporal gene expression patterns. *Genome Res.* **2003**, *13*, 965–979. [[CrossRef](#)]
13. Rhee, S.Y.; Wood, V.; Dolinski, K.; Draghici, S. Use and misuse of the gene ontology annotations. *Nat. Rev. Genet.* **2008**, *9*, 509. [[CrossRef](#)] [[PubMed](#)]
14. Eisenberg, E.; Levanon, E.Y. Human housekeeping genes, revisited. *Trends Genet.* **2013**, *29*, 569–574. [[CrossRef](#)] [[PubMed](#)]
15. Martínez, O.; Arce-Rodríguez, M.L.; Hernandez-Godínez, F.; Escoto-Sandoval, C.; Cervantes-Hernandez, F.; Hayano-Kanashiro, C.; Ordaz-Ortiz, J.J.; Reyes-Valdes, M.H.; Razo-Mendivil, F.G.; Garces-Claver, A.; et al. Transcriptome analyses throughout chili pepper fruit development reveal novel insights into the domestication process. *Plants* **2021**, *10*, 585. [[CrossRef](#)] [[PubMed](#)]
16. Wang, W.; Tan, H.; Sun, M.; Han, Y.; Chen, W.; Qiu, S.; Zheng, K.; Wei, G.; Ni, T. Independent component analysis based gene co-expression network inference (ICAnet) to decipher functional modules for better single-cell clustering and batch integration. *Nucleic Acids Res.* **2021**, *49*, e54. [[CrossRef](#)]
17. Jimmy, J.L.; Babu, S. Variations in the structure and evolution of Rice WRKY genes in Indica and japonica genotypes and their co-expression network in mediating disease resistance. *Evol. Bioinform.* **2019**, *15*, 1176934319857720. [[CrossRef](#)]
18. Benjamini, Y.; Hochberg, Y. Controlling the false discovery rate: A practical and powerful approach to multiple testing. *J. R. Stat. Soc. Ser. B (Methodol.)* **1995**, *57*, 289–300. [[CrossRef](#)]
19. Escoto-Sandoval, C.; Flores-Díaz, A.; Reyes-Valdes, M.H.; Ochoa-Alejo, N.; Martínez, O. A method to analyze time expression profiles demonstrated in a database of chili pepper fruit development. *Sci. Rep.* **2021**, *11*, 13181. [[CrossRef](#)]
20. Martínez, O. Time course gene expression experiments. In *Transcriptome Profiling: Progress and Prospects*; Ali, M.A., Lee, J., Eds.; Elsevier: Amsterdam, The Netherlands, 2022; Chapter 4, pp. 85–110.
21. Margolin, A.A.; Nemenman, I.; Basso, K.; Wiggins, C.; Stolovitzky, G.; Dalla Favera, R.; Califano, A. ARACNE: An algorithm for the reconstruction of gene regulatory networks in a mammalian cellular context. *BMC Bioinform.* **2006**, *7*, S7. [[CrossRef](#)]
22. Izadi, F.; Zarrini, H.N.; Kiani, G.; Jelodar, N.B. A comparative analytical assay of gene regulatory networks inferred using microarray and RNA-seq datasets. *Bioinformation* **2016**, *12*, 340. [[CrossRef](#)]

23. Arce-Rodríguez, M.L.; Ochoa-Alejo, N. An R2R3-MYB transcription factor regulates capsaicinoid biosynthesis. *Plant Physiol.* **2017**, *174*, 1359–1370. [[CrossRef](#)] [[PubMed](#)]
24. Martínez, O.; Escoto-Sandoval, C. Salsa: An R Package of Data Mining Facilities for *Capsicum* Gene Expression Profiles. *Zenodo* **2021**. [[CrossRef](#)]
25. Ma, P.; Castillo-Davis, C.I.; Zhong, W.; Liu, J.S. A data-driven clustering method for time course gene expression data. *Nucleic Acids Res.* **2006**, *34*, 1261–1269. [[CrossRef](#)] [[PubMed](#)]
26. Cloonan, N.; Forrest, A.R.; Kolle, G.; Gardiner, B.; Faulkner, G.J.; Brown, M.K.; Taylor, D.F.; Steptoe, A.L.; Wani, S.; Bethel, G.; et al. Stem cell transcriptome profiling via massive-scale mRNA sequencing. *Nat. Methods* **2008**, *5*, 613–619. [[CrossRef](#)]
27. Wang, Z.; Gerstein, M.; Snyder, M. RNA-Seq: A revolutionary tool for transcriptomics. *Nat. Rev. Genet.* **2009**, *10*, 57–63. [[CrossRef](#)]
28. Casati, P.; Walbot, V. Gene expression profiling in response to ultraviolet radiation in maize genotypes with varying flavonoid content. *Plant Physiol.* **2003**, *132*, 1739–1754. [[CrossRef](#)]
29. Gibson, G. The environmental contribution to gene expression profiles. *Nat. Rev. Genet.* **2008**, *9*, 575–581. [[CrossRef](#)]
30. Grishkevich, V.; Yanai, I. The genomic determinants of genotype × environment interactions in gene expression. *Trends Genet.* **2013**, *29*, 479–487. [[CrossRef](#)]
31. R Core Team. *R: A Language and Environment for Statistical Computing*; R Foundation for Statistical Computing: Vienna, Austria, 2013.
32. Mao, L.; Van Hemert, J.L.; Dash, S.; Dickerson, J.A. Arabidopsis gene co-expression network and its functional modules. *BMC Bioinform.* **2009**, *10*, 346. [[CrossRef](#)]
33. Langfelder, P.; Horvath, S. Eigengene networks for studying the relationships between co-expression modules. *BMC Syst. Biol.* **2007**, *1*, 54. [[CrossRef](#)]
34. Mitra, K.; Carvunis, A.R.; Ramesh, S.K.; Ideker, T. Integrative approaches for finding modular structure in biological networks. *Nat. Rev. Genet.* **2013**, *14*, 719–732. [[CrossRef](#)] [[PubMed](#)]
35. Zaret, K.S. Pioneer transcription factors initiating gene network changes. *Annu. Rev. Genet.* **2020**, *54*, 367–385. [[CrossRef](#)] [[PubMed](#)]
36. Arce-Rodríguez, M.L.; Ochoa-Alejo, N. Silencing *AT3* gene reduces the expression of *pAmt*, *BCAT*, *Kas*, and *Acl* genes involved in capsaicinoid biosynthesis in chili pepper fruits. *Biol. Plant.* **2015**, *59*, 477–484. [[CrossRef](#)]
37. Arce-Rodríguez, M.L.; Ochoa-Alejo, N. Biochemistry and molecular biology of capsaicinoid biosynthesis: Recent advances and perspectives. *Plant Cell Rep.* **2019**, *38*, 1017–1030. [[CrossRef](#)] [[PubMed](#)]
38. Kim, G.B.; Gao, Y.; Palsson, B.O.; Lee, S.Y. DeepTFactor: A deep learning-based tool for the prediction of transcription factors. *Proc. Natl. Acad. Sci. USA* **2021**, *118*, e2021171118. [[CrossRef](#)]
39. Lugar, D.J.; Mack, S.G.; Sriram, G. NetRed, an algorithm to reduce genome-scale metabolic networks and facilitate the analysis of flux predictions. *Metab. Eng.* **2021**, *65*, 207–222. [[CrossRef](#)]
40. Opgen-Rhein, R.; Strimmer, K. From correlation to causation networks: A simple approximate learning algorithm and its application to high-dimensional plant gene expression data. *BMC Syst. Biol.* **2007**, *1*, 37. [[CrossRef](#)]
41. Langfelder, P.; Horvath, S. WGCNA: An R package for weighted correlation network analysis. *BMC Bioinform.* **2008**, *9*, 559. [[CrossRef](#)]
42. Amar, D.; Safer, H.; Shamir, R. Dissection of regulatory networks that are altered in disease via differential co-expression. *PLoS Comput. Biol.* **2013**, *9*, e1002955. [[CrossRef](#)]
43. Watson, M. CoXpress: Differential co-expression in gene expression data. *BMC Bioinform.* **2006**, *7*, 509. [[CrossRef](#)]
44. Tesson, B.M.; Breitling, R.; Jansen, R.C. DiffCoEx: A simple and sensitive method to find differentially coexpressed gene modules. *BMC Bioinform.* **2010**, *11*, 497. [[CrossRef](#)] [[PubMed](#)]
45. Mukherjee, S.; Banerjee, B.; Karasik, D.; Frenkel-Morgenstern, M. mRNA-lncRNA Co-Expression Network Analysis Reveals the Role of lncRNAs in Immune Dysfunction during Severe SARS-CoV-2 Infection. *Viruses* **2021**, *13*, 402. [[CrossRef](#)] [[PubMed](#)]
46. Oliveira de Biagi, C.A., Jr.; Nociti, R.P.; Brotto, D.B.; Funicheli, B.O.; Cassia Ruy, P.d.; Bianchi Ximenez, J.P.; Alves Figueiredo, D.L.; Araujo Silva, W., Jr.; et al. CeTF: An R/Bioconductor package for transcription factor co-expression networks using regulatory impact factors (RIF) and partial correlation and information (PCIT) analysis. *BMC Genom.* **2021**, *22*, 624. [[CrossRef](#)] [[PubMed](#)]
47. Yin, W.; Mendoza, L.; Monzon-Sandoval, J.; Urrutia, A.O.; Gutierrez, H. Emergence of co-expression in gene regulatory networks. *PLoS ONE* **2021**, *16*, e0247671. [[CrossRef](#)]
48. Cong, B.; Liu, J.; Tanksley, S.D. Natural alleles at a tomato fruit size quantitative trait locus differ by heterochronic regulatory mutations. *Proc. Natl. Acad. Sci. USA* **2002**, *99*, 13606–13611. [[CrossRef](#)] [[PubMed](#)]
49. Chen, M.; Thelen, J.J. Plastid uridine salvage activity is required for photoassimilate allocation and partitioning in arabidopsis. *Plant Cell* **2011**, *23*, 2991–3006. [[CrossRef](#)]
50. Csardi, G.; Nepusz, T. The igraph software package for complex network research. *Interj. Complex Syst.* **2006**, *1695*, 1–9.
51. Escoto-Sandoval, C.; Ochoa-Alejo, N.; Martínez, O. Inheritance of gene expression throughout fruit development in chili pepper. *Sci. Rep.* **2021**, *11*, 16. [[CrossRef](#)]
52. Gómez-García, M.d.R.; Ochoa-Alejo, N. Biochemistry and molecular biology of carotenoid biosynthesis in chili peppers (*Capsicum* spp.). *Int. J. Mol. Sci.* **2013**, *14*, 19025–19053. [[CrossRef](#)]
53. Góngora-Castillo, E.; Ibarra-Laclette, E.; Trejo-Saavedra, D.L.; Rivera-Bustamante, R.F. Transcriptome analysis of symptomatic and recovered leaves of geminivirus-infected pepper (*Capsicum annuum*). *Viol. J.* **2012**, *9*, 295. [[CrossRef](#)]

54. Hödl, M.; Basler, K. Transcription in the absence of histone h3. 2 and h3k4 methylation. *Curr. Biol.* **2012**, *22*, 2253–2257. [[CrossRef](#)] [[PubMed](#)]
55. Jeevalatha, A.; Siddappa, S.; Kumar, A.; Kaundal, P.; Guleria, A.; Sharma, S.; Nagesh, M.; Singh, B.P. An insight into differentially regulated genes in resistant and susceptible genotypes of potato in response to tomato leaf curl new delhi virus-[potato] infection. *Virus Res.* **2017**, *232*, 22–33. [[CrossRef](#)] [[PubMed](#)]
56. Kwak, K.J.; Kim, J.Y.; Kim, Y.O.; Kang, H. Characterization of transgenic arabidopsis plants overexpressing high mobility group b proteins under high salinity, drought or cold stress. *Plant Cell Physiol.* **2007**, *48*, 221–231. [[CrossRef](#)] [[PubMed](#)]
57. Lachmann, A.; Giorgi, F.M.; Lopez, G.; Califano, A. Aracne-ap: Gene network reverse engineering through adaptive partitioning inference of mutual information. *Bioinformatics* **2016**, *32*, 2233–2235. [[CrossRef](#)] [[PubMed](#)]
58. Laurencelle, L. The exact binomial test between two independent proportions: A companion. *Quant. Methods Psychol.* **2021**, *17*, 76–79. [[CrossRef](#)]
59. Lee, Y.R.J.; Liu, B. Cytoskeletal motors in arabidopsis. sixty-one kinesins and seventeen myosins. *Plant Physiol.* **2004**, *136*, 3877–3883. [[CrossRef](#)]
60. Ohler, L.; Niopek-Witz, S.; Mainguet, S.E.; Möhlmann, T. Pyrimidine salvage: Physiological functions and interaction with chloroplast biogenesis. *Plant Physiol.* **2019**, *180*, 1816–1828. [[CrossRef](#)]
61. Petricka, J.J.; Clay, N.K.; Nelson, T.M. Vein patterning screens and the defectively organized tributaries mutants in arabidopsis thaliana. *Plant J.* **2008**, *56*, 251–263. [[CrossRef](#)]
62. Redon, C.; Pilch, D.; Rogakou, E.; Sedelnikova, O.; Newrock, K.; Bonner, W. Histone h2a variants h2ax and h2az. *Curr. Opin. Genet. Dev.* **2002**, *12*, 162–169. [[CrossRef](#)]
63. Song, Q.; Chen, S.; Wu, Y.; He, Y.; Feng, J.; Yang, Z.; Lin, W.; Zheng, G.; Li, Y.; Chen, H. Comparative transcriptome analyses of gene response to different light conditions of camellia oleifera leaf using illumina and single-molecule real-time-based rna-sequencing. *Forests* **2020**, *11*, 91. [[CrossRef](#)]
64. Spies, D.; Ciaudo, C. Dynamics in transcriptomics: Advancements in rna-seq time course and downstream analysis. *Comput. Struct. Biotechnol. J.* **2015**, *13*, 469–477. [[CrossRef](#)] [[PubMed](#)]
65. Spies, D.; Renz, P.F.; Beyer, T.A.; Ciaudo, C. Comparative analysis of differential gene expression tools for rna sequencing time course data. *Briefings Bioinform.* **2019**, *20*, 288–298. [[CrossRef](#)]
66. Strober, B.; Elorbany, R.; Rhodes, K.; Krishnan, N.; Tayeb, K.; Battle, A.; Gilad, Y. Dynamic genetic regulation of gene expression during cellular differentiation. *Science* **2019**, *364*, 1287–1290. [[CrossRef](#)] [[PubMed](#)]
67. Wang, N.; Yang, Y.; Moore, M.J.; Brockington, S.F.; Walker, J.F.; Brown, J.W.; Liang, B.; Feng, T.; Edwards, C.; Mikenas, J.; et al. Evolution of portulacineae marked by gene tree conflict and gene family expansion associated with adaptation to harsh environments. *Mol. Biol. Evol.* **2019**, *36*, 112–126. [[CrossRef](#)] [[PubMed](#)]
68. Zhang, Y.; Li, Z.; Ma, B.; Hou, Q.; Wan, X. Phylogeny and functions of lob domain proteins in plants. *Int. J. Mol. Sci.* **2020**, *21*, 2278. [[CrossRef](#)] [[PubMed](#)]
69. Zhang, Z.; Liu, J.; Huber, D.J.; Qu, H.; Yun, Z.; Li, T.; Jiang, Y. Transcriptome, degradome and physiological analysis provide new insights into the mechanism of inhibition of litchi fruit senescence by melatonin. *Plant Sci.* **2021**, *308*, 110926. [[CrossRef](#)] [[PubMed](#)]
70. Zhao, X.; Bai, X.; Jiang, C.; Li, Z. Phosphoproteomic analysis of two contrasting maize inbred lines provides insights into the mechanism of salt-stress tolerance. *Int. J. Mol. Sci.* **2019**, *20*, 1886. [[CrossRef](#)] [[PubMed](#)]
71. Zoppoli, P.; Morganello, S.; Ceccarelli, M. Timedelay-aracne: Reverse engineering of gene networks from time-course data by an information theoretic approach. *BMC Bioinform.* **2010**, *11*, 54. [[CrossRef](#)]

**Disclaimer/Publisher’s Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.