Journal of Chemical, Biological and Physical Sciences

An International Peer Review E-3 Journal of Sciences

*Available online at*www.jcbsc.org

Section B: Biological Sciences

CODEN (USA): JCBPAT Research Article

Network Inference for Gene Regulation during Flowering Intitiation in *Arabidopsis Thaliana* using Time-Series Data

Pui Shan Wong and Sachiyo Aburatani*

Biotechnology Research Institute for Drug Discovery, National Institute of AIST, Tokyo, Japan

Received: 26 August 2015; **Revised:** 17 September 2015; **Accepted:** 25 September 2015

Abstract: Time-series gene expression is very useful for tracking changes in expression during environmental changes or experimental conditions. As the majority of gene interactions has a many-to-many relationship with each other, a network can be used to model and analyse such data. We create and use a method that maps, gene interactions onto a network using their time-series expression data. The model relies on coding the differences in gene expression through the time into numerical patterns. The patterns form the vertices of the network and are linked together by edges that are decided by comparing two patterns. The edge weights are calculated using the expression of two connected patterns and represents the degree of expression difference between the two connected patterns. We apply our method to transcription factors in apical meristem time-series data from *Arabidopsis thaliana* and show that the inferred network identifies and clusters transcription factors involved in floral transition. The network consists of 17 subnetworks of varying sizes with the largest containing 36 vertices. The subnetworks are distinguished by gene ontology enrichment terms. We identify the pivotal time period of transition from vegetative growth to flowering to 12 and 13 hours after germination by examining the expression patterns.

Keywords: Arabidopsis thaliana, transcription factors, network, time-series

INTRODUCTION

Genome sequencing and expression experiments are rapidly advancing the field of biology as they become more efficient and accessible. The data produced by these technologies give us great insight into the genomics and metabolomics of living systems. It shows that genes function within multiple systems that include many complex interactions. The intricate processes that take place between genotype and phenotype plays a large role in organism traits.

In order to keep up with the advance in data gathering methods, data analysis techniques also need to be developed. One of the key steps during data analysis lies within developing or choosing the correct data analysis techniques to derive information and meaning from raw data. Next-generation sequencing is acknowledged for generating a wealth of data that tends to be deep and complex, requiring different types of analyses, models and tools such as DESeq2[1] and EBSeq[2], as well as higher interpretation tools such as gene set enrichment analysis (GSEA)[3].

A network is a type of model that is useful for analysing interactions within a cell. As networks are analysis as well as visualization tools, they are helpful for examining systems where one action affects many components. These types of interactions happen frequently in a cell during a response to stimuli, such as metabolites, or environmental queues, such as light. The main elements involved in gene expression regulation are called transcription factors. They are proteins that activate or suppress the gene transcription activity, typically by binding to a transcription factor binding site close to a gene. When a transcription factor is activated, it can affect the expression of other genes in a many–to-many relationship. This type of relationship is important in situations such as the introduction of drug compounds.

We propose an analysis method using gene expression patterns to construct a network that models the interaction of transcription factors within a cellular system. By converting expression values into discrete units, called expression patterns, our method is able to infer a transcription factor network on temporal RNA-Seq data and extract information about interactions occurring within the organism. It is an extension of the phylogenetic profile clustering method used for assigning protein functions[4], altered so that it can work on expression data to produce a visual summary of the relationships between expression patterns. We apply our method to apical meristem growth expression data from *Arabidopsis thaliana* to model the expression of transcription factors associated with floral transition[5].

The transition from vegetative development to flowering is controlled by transcription factors responding to genetic and environmental stimuli such as hormones and light[6]. Several major transcription factors were outlined with the *A. thaliana* data[5]. Some of them positively regulate transitioning and flowering like SUPPRESSOR OF OVEREXPRESSION OF CO 1 (SOC1)[7,8] and FLOWERING LOCUS D (FD)[9,10]. Some of them negatively regulate flowering like FLOWERING LOCUS C (FC)[11] and SHORT VEGETATIVE PHASE (SVP)[12]. Other transcription factors initiate different stages during the transitioning period such as initiating flowering or controlling floral organ identity. These include LEAFY (LFY)[13,14], CAULIFLOWER (CAL)[15,16], PISTILLATA (PI)[17], AGAMOUS and various SQUAMOSA PROMOTER BINDING PROTEIN-LIKE family (SPL)[18-21] that go on to regulate APETALA1 (AP1)[22,23].

We use the constructed network to show that the floral transition transcription factors are clustered together and use them to confirm the identification of additional transcription factors that affect or are affected by known regulators.

Additionally, we use the direction and weight of the network edges to quantify the relationships between groups of transcription factors to highlight important regulation events at each time. Lastly, we broadly describe the processes being regulated in our network by using an enrichment analysis on the transcription factors incorporated in our network.

METHOD

Expression Pattern Creation: The purpose of the expression patterns is to convey the change in expression in time for each gene. As the *A. thaliana* data, we used[5] was already normalized at each time, the difference in expression between adjacent time points could be calculated directly. The transition from vegetative phase to flowering initiation started from M4 so the differences in expression were calculated between M4 to M10. The time intervals were M4-M5, M5-M6, M6-M7, M7-M8, M8-M9, M9-M10. If an experiment used a control and treatment group, it is also possible to use fold change data instead of difference in expression.

The expression values in each time interval were binned into -1, 0 and 1 values, where -1 indicates a decreasing in expression, 0 indicates no change in expression and 1 indicates an increase in expression. To compensate for the variation in expression values, a threshold was used to determine whether a non-zero expression was large enough to be considered -1 or 1. As most genes do not exhibit varied gene expression when compared to genes of interest, we can use their gene expression as a guide for no change in expression. For each gene, we calculated the standard deviation in gene expression across all time intervals. We then took the median of all the standard deviations and used the positive and negative value of that as the threshold boundary for the 0 bin. Our threshold was 90.29 meaning gene expressions between -90.29 and 90.29 were put into the 0 bin and gene expressions were binned into 1 or -1 if they were above or below 90.29 and -90.29 respectively.

A vector was assembled for each gene to create expression patterns using the binned values. The elements of each vector are ordered in chronological order to the time intervals starting with M4-M5 in the first position.

Identifying Transcription Factors: The transcription factors were found using the Plant Transcription Factor Database[24] for *A. thaliana*.

Network Construction

Graph Structure: A network was constructed by using the expression patterns as nodes. Edges in the network were added to signify a relationship between the two patterns while the edge weights quantified the relationship.

A pair of expression patterns, nodes u and v, are denoted as $\sim u$ and $\sim v$ and are made up of values u_i and v_i at each position i within each pattern. All u_i and v_i have possible values of -1, 0 and 1. When $\sim u$ contains at least two different values, the first position where u_i is different from u_1 is named a. As $u_1 = 6\ u_a$, it stands to reason that a 6= 1. For example, in (0, 0, 1, 0, 0, 0), $u_{[1]} = 0$ and $u_{[a]} = 1$ so $a = 3$. An edge connecting $\sim u$ to $\sim v$ is made if there is only one u_i and v_i that is not equal to each other. The position where this difference is located is named b. For example, in $\sim u$ = (1,0,0,1,0,0) and $\sim v$ = (1,0,0,0,0,0) the difference is at u_4 and v_4, so therefore $b = 4$.

Edge Properties: The edge direction was decided by considering the expression of transcription factors at early time points affecting the expression of transcription factors at later time points. We developed the following algorithm to decide whether an edge started from *u* and ended at *v*.

Step 1: Remove the edge from *u* to *v* if either of the vectors are made up of only one value. These are the patterns consisting of only 0s or only 1s or only -1s and so is unrelated to time.

Step 2: Compare $\sim u_{[1:a]}$ and $\sim v_{[1:a]}$. If they are the same, then the edge direction from *u* to *v* is false because it means that the difference in expression pattern is after *a*. If they are different, continue to step 3.

Step 3: For *u* and *v* where *a* = *b*, check if $v_1 = 0$ and $v_a = 0$. If $v_1 = 0$ and $v_a = 0$, then the edge direction from *u* to *v* is true as the pattern in $u_{[1:a]}$ contains 1 or -1 while $v_{[1:a]}$ is 0. If they are both not equal to 0, then the edge direction is false.

Step 4: For patterns where *a* is different from *b*, check if u_b 6= 0 and $v_b = 0$. If u_b 6= 0 and $v_b = 0$, then the edge direction from *u* to *v* is true as the expression pattern 1 or -1 happens in *u* before *v*. If $u_b = 0$ and v_b 6= 0 , then the edge direction is false.

After establishing edge directions, the edge weights were calculated using Equation 1. The weights are a measure of the expression between genes at nodes *u* and *v*, particularly at position u_a and v_a. As there are usually many genes per expression patterns, the median was chosen to represent the average expression value.

$$W(u,v) = \frac{\mathbb{M}_u(F_d(x))}{\mathbb{M}_v(F_d(x))} \frac{\mathbb{M}_v(F_{\forall t}(x))}{\mathbb{M}_u(F_{\forall t}(x))} \quad (1)$$

where $W(u,v)$ is the weight of the edge from node *u* to *v*, *d* is the time at which the expression pattern differs between *u* and *v*, and $\mathrm{M}u(F_d(x))$ is the median expression of genes with pattern *u* at time *d*. Similarly, $\mathrm{M}v(F_d(x))$ is the median expression of genes with pattern *v* at time *d*. The following $\mathrm{M}u(F_{\forall t}(x))$ and $\mathrm{M}v(F_{\forall t}(x))$ is the median of expression at all-time points of genes with pattern *u* and *v* respectively. In the case that a median is 0, a small number is added to all medians such as $1\text{x}10^{-10}$.

Undirected edges and unconnected nodes were removed from the full network, leaving a subnetwork or several subnetworks of connected nodes and directed edges.

Network Visualization: The graphs were created and drawn in R using the igraph package[25,26].

Enrichment Analysis: The gene list of each subnetwork was used in an enrichment analysis at The Plant GSEA[27]. The gene sets used were the three Gene Ontology (GO) components (biological process, cellular component and molecular function) with the species set to *A. thaliana*.

The list of gene ontologies were then summarised using REViGO[28] to collapse redundant terms. This was done using the small setting with the FDR corrected p-values from the enrichment analysis. We used the *A. thaliana* GO term database with the SimRel semantic similarity measure.

Expression Data Source: The *A. thaliana* expression data is publicly available in the NCBI Sequence Read Archive [project ID PRJNA268115][5].

RESULTS

Gene Expression Patterns: Gene expression patterns were created for 1,402 *A. thaliana* transcription factors. There were 242 unique patterns of length six where each element in the vector represented the

difference in expression between two adjacent time points. The time intervals were M4-M5, M5-M6, M6-M7, M7-M8, M8-M9 and M9-M10. Each time point was one day apart starting from 10 days after germination at M4 to 16 days after germination at M10. The values of each vector element indicated if gene expression was up-regulated (1), down-regulated (-1) or unchanged (0). With six sample points and three possible values, there are 729 possible expression patterns. However, we only observed 242 patterns in our data. The top three most common patterns were (0, 0, 0, 0, 0, 0) seen in 40% of the transcription factors, (-1, 1, 0, 0, 0, 0) seen in 2.3% of transcription factors and (-1, 0, 0, 0, 0, 0) seen in 2.1% of the transcription factors. In contrast, there were 111 patterns seen in one gene each and 51 patterns represented by two genes each. The patterns with ten or more genes had on average 3.6 0s, 2 1s and 2.1 -1s while the patterns with one or two genes had on average 2.1 0 s, 1.9 1s and 1.8 -1 s.

The floral transition transcription factors were represented by 15 unique patterns. The prevalence of each value in the patterns reveals that the most important time interval for these transcription factors is M6-M7. This time interval had eight patterns with a 1 value at this time, followed by six patterns with a 0 value and only one pattern with -1. When compared to the counts for the other time periods, it has the highest frequency of a value and the lowest (1 and -1 respectively). The other time intervals had a mixture of 7, 6 and 2 or 6, 5 and 4 frequencies for each value.

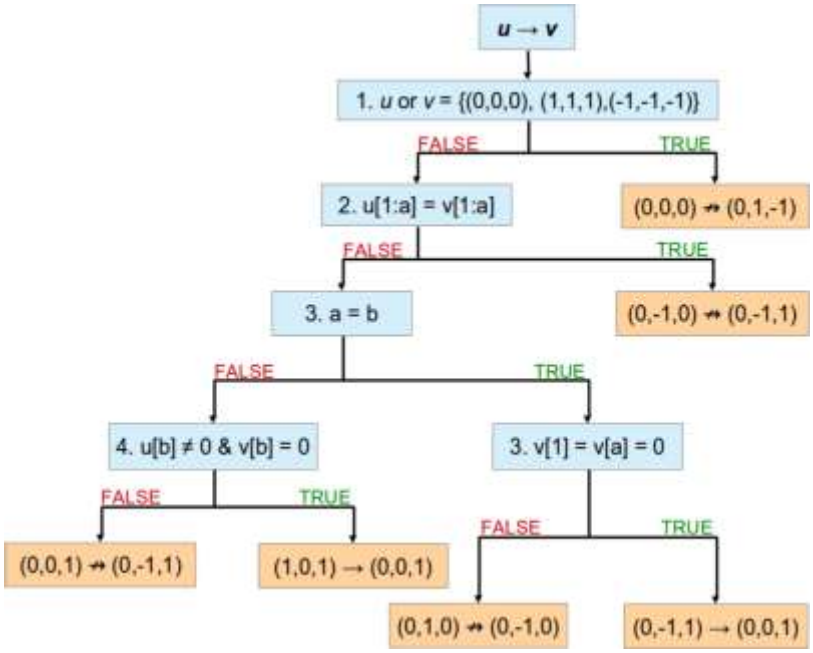

Fig. 1: A step-by-step flow chart showing the application of the edge direction algorithm with examples. The blue boxes show a step in the algorithm and the orange boxes show an example of a node pair following the algorithm.

Transcription Factor Network: The transcription factor expression patterns were used to create a network model of the relationships between expression patterns. As discrete units, the expression patterns were used as vertices to construct the initial undirected network (Fig. 2). The network was connected by joining expression patterns that differed by only one value, e.g. (1, 1, 0, 0, 0, 0) and (0, 1, 0, 0, 0, 0). This created a network made up of 242 vertices connected by 680 edges. The most connected vertex was (1, -1, 0, 1, -1,

0) and had a degree of 12. There were eight vertices with a degree of 1 and one unconnected vertex, (-1, -1, 1, 1, 0, 1).

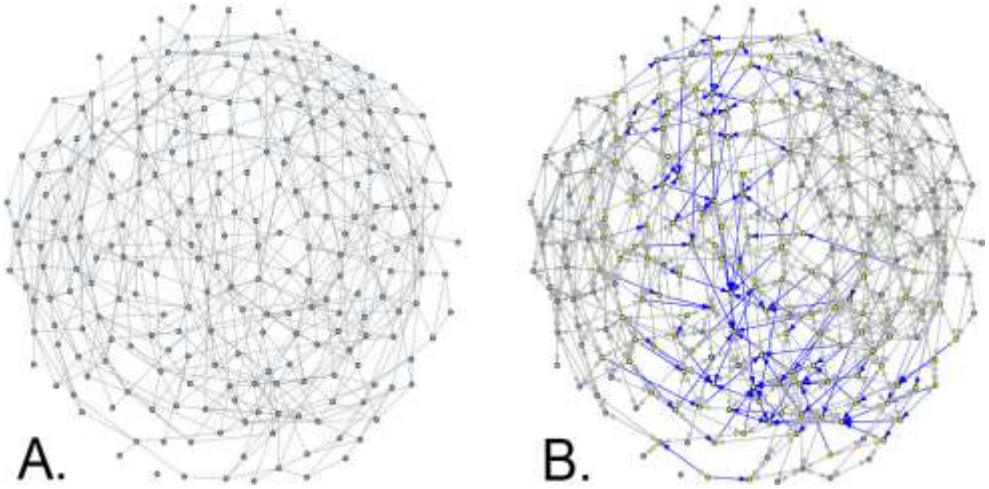

Fig. 2: The network during construction in the first two stages. A. The vertices are joined together to create an undirected network. The vertices are in light blue and are connected by gray edges. B. The edge directions are established and highlighted in dark blue. Vertices with at least one directed edge are in yellow. The final graph includes the yellow vertices and dark blue edges only.

Once the edges were established, the directions and weights were added to the undirected network (Fig. 2). The direction was determined by the temporal effect of gene regulation, while the weights were ratios signifying the impact of gene expression at the time where the expression patterns differed. Connected patterns whose edges could not be directed were removed for clarity. This produced 17 separate subnetworks that contained 136 vertices and 123 edges within them (Fig. 3). There were 6 subnetworks that were made up of only two vertices and an edge, while the largest subnetwork was made up of 36 vertices and 36 edges.

Due to the way the edge directions were determined, most subnetworks contained a central sink vertex where all the connecting edges were incoming edges. The sink vertex patterns usually contained many 0 values such as (0, 0, 0, 0, 0, 1) and were represented by many transcription factors compared to the leaf vertices. While the smaller subnetworks with five or less vertices were mainly linear, the large subnetworks were mainly star-like with fairy separate branches joined to a central spine. A rectangular structure was visible in three of the large subnetworks consisting of four vertices and four edges.

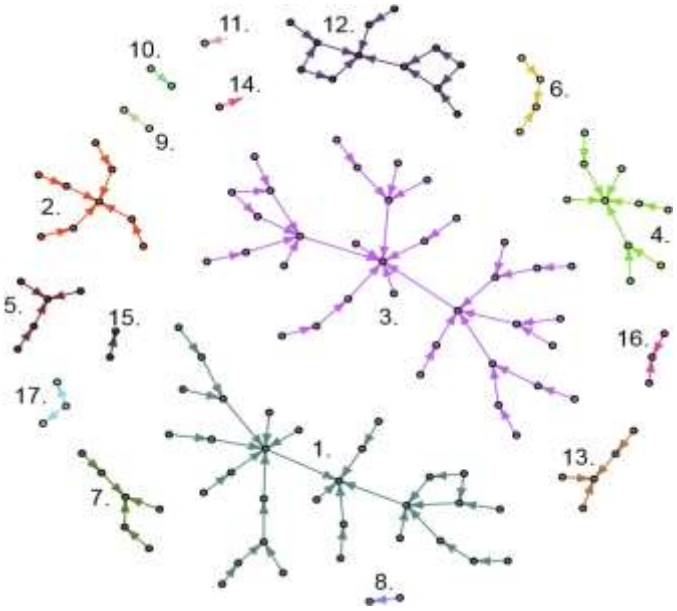

Fig 3: The final network with undirected edges removed from Fig. 2 B. There are 17 unconnected subnetworks labelled 1-17 of varying sizes. In descending order, they are 3 (36 vertices), 1 (31 vertices), 12 (12 vertices), 4 (10 vertices), 2 (9 vertices), 7 (6 vertices), 5 (5 vertices), 13 (5 vertices), 6 (4 vertices), 16 (3 vertices), 17 (3 vertices), 8 (2 vertices), 9 (2 vertices), 10 (2 vertices), 11 (2 vertices), 14 (2 vertices) and 15 (2 vertices). The floral transition transcription factors are in subnetworks 1, 4, and 7.

Most of the edge weights were quite small with a median of 1.76 and a mean of -1.40. There were eight large weights with values over 100 and these were found in the two largest subnetworks. There were two paths of large edge weights with three vertices where the largest weights were between the last two vertex of each path. These were (1, -1, 0, 0, 1, 1) → (0, -1, 0, 0, 1, 1) → (0, 0, 0, 0, 1, 1) and (-1, 1, 0, 0, 1, 1) → (0, 1, 0, 0, 1, 1) → (0, 0, 0, 0, 1, 1). The second path indicates that a regulation step at M5-M6 is related to transcription factor activity at later time points and that the relation is stronger at M4-M5 than at M5-M6.

There were 17 floral transition transcription factors and nine of them were found in three of the subnetworks, 1, 4 and 7. Subgraph 1 was the largest among them and contained seven transcription factors (Fig. 4). The pathway begins with SPL4 and LFY in patterns (1, -1, 1, 1, 1, 1) and (-1, 0, 1, 1, 1, 1), and then flows on to AP1, AP3, PI and SPL3, with patterns (0, 0, 0, 1, 1, 1), (0, 0, 0, 0, 1, 1) and (0, 0, 0, 0, 0, 1). Of the two initial transcription factors, the edge from LFY had a larger edge weight than SPL4. The other transcription factors identified in the subnetworks were CAL in subnetwork 4 and FLC in subnetwork 7. The remaining unconnected transcription factors were SVP, CO1, FD, SPL9, SPL5, SPL15, MYB3R1 and MYB3R4.

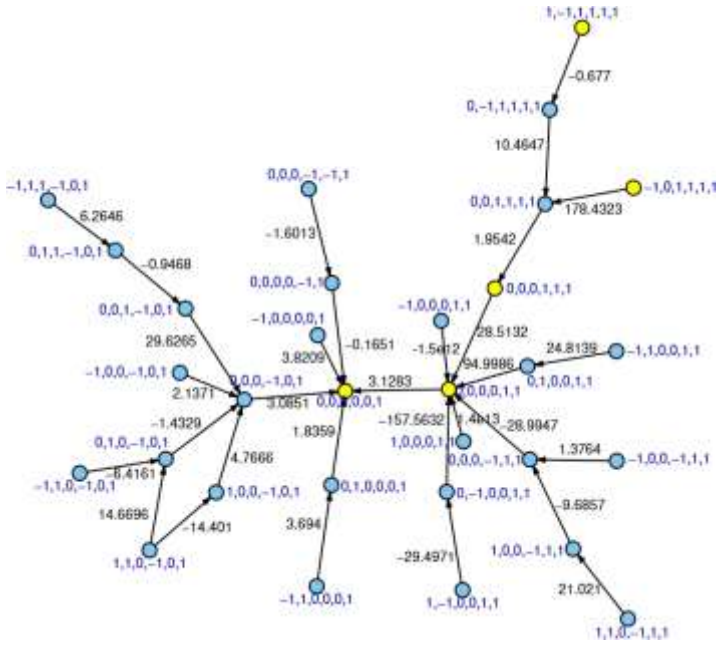

Fig. 4: The weights and vertex patterns of subnetwork 1. The weights are written in black next to the applicable edge and the patterns are written in dark blue next to the applicable vertex. This major subnetwork contains 9 of the 17 floral transition transcription factors. Their expression patterns are highlighted by yellow vertices.

Enrichment Analysis: The genes of the subnetworks were put through a GO enrichment analysis. The four largest subnetworks had 199, 170, 119 and 74 enriched terms respectively. They shared 54 common terms which made up a 73% majority for the fourth largest subnetwork. The enrichment analysis was successful for all subnetworks except for subnetwork 14 which did not contain enough annotated genes.

The resulting lists of enriched gene ontologies were summarised so that they could be compared. Many of the summaries included general terms such as biological regulation or metabolic processes, however, there were several unique and indispensable ontologies that were observed. Subgraph 1 was defined by transcription factors annotated with rhythmic process, circadian rhythm, response to light stimulus, protein acetylation, and regulation of multicellular organismal process, reproduction and reproductive structure development. Subgraph 2 was a lot smaller and was summarised the two terms, long-chain fatty acid metabolic process and defence response to insects. Subgraph 3 was the largest graph and was distinguished by the terms, immune system process, response to endogenous stimulus, glucuronoxylan metabolic process, bract development, reproductive process, positive regulation of biological process and multi-organism process. Subgraph 4 contained a moderate number of genes, but was only defined by DNA-templated regulation of transcription and response to gibberellin.

Subgraph 5 was smaller and was represented by nitrogen compound metabolic process and positive regulation of biological process. Subgraph 6 was roughly the same size as subnetwork 5 and was described by the response to salicyclic acid, respiratory burst and positive regulation of biological process. Subgraph 7 was summarised by the more specific terms, nitrogen compound metabolic process, chloroplast relocation and negative regulation of flower development. Subgraph 8 had only two vertices and its summarising

terms were DNA-templated regulation of transcription, response to gibberellin and embryo development ending in seed dormancy. Similarly subnetwork 9 was the same size and had several unique terms in its summary of response to auxin, multicellular organismal process, developmental process, DNA-templated positive regulation of transcription and gynoecium development.

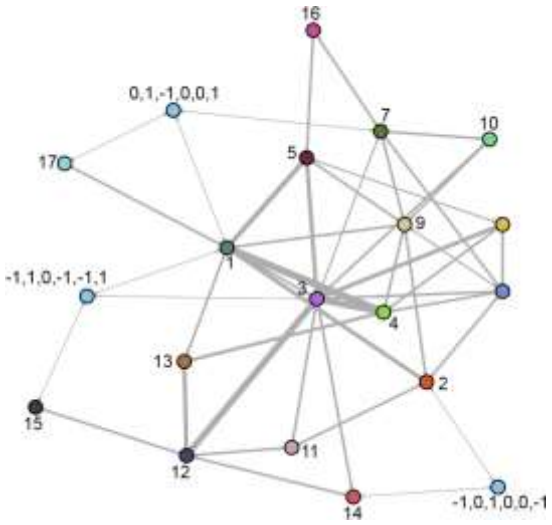

Fig. 5: A simplified view of the subnetworks when connected by the minimum number of undirected edges. Each subnetwork vertex and edge connection is merged into one vertex, keeping the edges that lie along the shortest path between subnetworks. The subnetworks are colored as they were in Fig. 3. Joining vertices that were not part of a subnetwork are marked in light blue and labelled with the expression pattern.

Subgraph 10 and 11 were also two vertex graphs where subnetwork 10 was described by reproduction, response to abscisic acid, seed germination, peptidyl-histidine modification and reproductive process while subnetwork 11 was described by general transcription factor terms like nucleic acid binding, sequence specific DNA binding transcription factor activity and transcription regulator activity. Subgraph 12 was a more moderately sized graph distinguished by the term immune effector process, post embryonic morphogenesis, RNA metabolic process and multicellular organismal process. Subgraph 13 is only defined by two terms, regulation of metabolic process and cotyledon morphogenesis. Subgraphs 15, 16 and 17 are very small structures where subnetwork 15 is involved in demethylation, subnetwork 16 in transcription regulator activity and subnetwork 17 in cellular response to glucose stimulus.

DISCUSSION

Gene expression is a chronologically dependent process where a change in expression in gene A at one time point is related to the change in expression in gene B at a later time point. We developed our method based upon that approach, creating the expression patterns and edge algorithm. When applied to floral transition data, the inferred network successfully identified established connections, such as the positive regulation of AP1 by LFY[14]. The general arrangement of each subnetwork where outlying leaf patterns were

represented by fewer genes than the central patterns verified the cascading effect initiated by a few genes that affect many downstream genes.

The expression patterns were made to group similarly expressed genes together so that examining important expression patterns will identify related genes. Expression pattern importance can be determined through the experiment design or previous research, such as the function of LFY in the floral transition data. The expression pattern for LFY was (-1, 0, 1, 1, 1, 1) which was shared by two other transcription factors, ATREM1 and BETA HLH PROTEIN 93 (BHLH093). ATREM1 has been observed from the vegetative apical meristem to the inflorescence meristem and binds to AP1, AP3, PI and SVP[29,30]. This is confirmed in our network where AP1, AP3 and PI are downstream from ATREM1. BHLH093 has been observed in several different development stages such as the final stage of leaf development, expanded cotyledon stage, and flowering[31,32]. It is possible that BHLH093 is expressed as a regulator of a concurrent process with floral transitioning but is not directly part of the process.

Similarly, connections to identified expression patterns also helped to find related genes. In our network, LFY is connected to AP1 by an additional pattern, (0, 0, 1, 1, 1, 1), which represents the two transcription factors, SQUAMOSA PROMOTER BINDING PROTEIN-LIKE 8 (SPL8) and FOREVER YOUNG

FLOWER (FYF). SPL8 has been seen to act as a developmental regulator with gibberellins and flowering time[33,34] while FYF is a repressor of floral organ senescence[35,36].

The final network summarizes the data into 17 subnetworks, each distinguishable by enriched GO terms. To illustrate the connectivity between the subnetworks, they were collapsed into vertices 5. The reduced graph clearly shows a connection between the two largest subnetworks, 1 and 3. This connection is formed by two connecting patterns, indicating a relatively weak link. In contrast, there are 8 and 7 links between subnetwork 1 and 4, and 3 and 4 respectively. This suggests that gene regulation of rhythmic processes, response to light and other processes found in subnetwork 1 is related to immune responses, response to endogenous stimulus and other processes found in subnetwork 3 primarily through transcription factors related to the plant hormone, gibberellin.

Although the inferred network included many floral transition transcription factors, there were several patterns of important transcription factors that remained unconnected such as SOC1[7]. The edge determination of the network is heavily influenced by the number of time points so that the increase in the number of time points will result in a less connected graph due to the higher number of patterns needed to make connections. This effect can be decreased by focusing on a narrower range of time points, as we did by excluding earlier time periods prior to floral transition. Although it will not always be sufficient, it remains that the remaining transcription factors were able to be connected into the network with undirected edges.

CONCLUSION

By using our method that converts expression values into numeric patterns, we were able to infer a network that summarizes temporal changes in gene expression. The network depicts the associations between patterns, enabling us to see which patterns initiate changes and which associations had a greater effect. This approach improves visualization of time-series data and introduces new ways of investigating biological processes.

We applied our method to floral transition data from *A. thaliana* and identified the relationships of established transcription factors, LFY, AP1, AP3, AGAMOUS, PI and SPL. By investigating shared and connected patterns, we identified other transcription factors that seem to be associated with floral transition. Finally, we detected an association between rhythmic process regulation and immune response regulation via regulation of gibberellins by using GO enrichment on the inferred network.

REFERENCES

1. M. I. Love, W. Huber and S. Anders, Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2, *Genome Biology*, 2014, 15(12), 550. doi:10.1186/s13059-014-0550-8.
2. N. Leng, J. A. Dawson, J. A. Thomson, V. Ruotti, A. I. Rissman, B. M. Smits, J. D. Haag, M. N. Gould, R. M. Stewart and C. Kendziorski, EBSeq: an empirical Bayes hierarchical model for inference in RNA-seq experiments, *Bioinformatics*, 2013, 29(8), 1035–43. doi:10.1093/bioinformatics.
3. A. Subramanian, P. Tamayo, V. K. Mootha, S. Mukherjee, B. L. Ebert, M. A. Gillette, A. Paulovich, S. L. Pomeroy, T. R. Golub, E. S. Lander and J. P. Mesirov, Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles, Proceedings of National Academy of Science USA, 2005, 102(43), 45545–50. doi:10.1073/pnas. 0506580102.
4. M. Pellegrini, E. M. Marcotte, M. J. Thompson, D. Eisenberg and T. O. Yeates, Assigning protein functions by comparative genome analysis: protein phylogenetic profiles, Proceedings of the National Academy of Sciences of the United States of America, 1999, 96(8), 4285–4288. doi: 10.1073/pnas.96.8.4285.
5. A. V. Klepikova, M. D. Logacheva, S. E. Dmitriev and A. A. Penin, RNAseq analysis of an apical meristem time series reveals a critical point in Arabidopsis thaliana flower initiation, *BMC Genomics*, 2000, 16 (466). doi:10. 1186/s12864-015-1688-9.
6. Y. Kobayashi and D. Weigel, ove on up, it's time for change–mobile signals controlling photoperiod-dependent flowering, *Genes and Development*, 2007, 21(19), 2371–84. doi:10.1101/gad.1589007.
7. R. Borner, G. Kampmann, J. Chandler, R. Gleissner, E. Wisman, K. Apel and S. Melzer, A MADS domain gene involved in the transition to flowering in Arabidopsis, *The Plant Journal*, 2000, 24(5), 591–9. doi:10.1046/j.1365-313x.2000.00906.x.
8. H. Lee, S. S. Suh, E. Park, E. Cho, J. H. Ahn, S. G. Kim, J. S. Lee, Y. M. Kwon and I. Lee, The AGAMOUS-LIKE 20 MADS domain protein integrates floral inductive pathways in Arabidopsis, *Genes and Development*, 2000, 14(18), 2366–76. doi:10.1101/gad.813600.
9. M. Abe, Y. Kobayashi, S. Yamamoto, Y. Daimon, A. Yamaguchi, Y. Ikeda, H. Ichinoki, M. Notaguchi, K. Goto and T. Araki, FD, a bZIP protein mediating signals from the floral pathway integrator FT at the shoot apex, *Science*, *2005*, 309(5737), 1052–6. doi:10.1126/science.1115983.
10. P. A. Wigge, M. C. Kim, K. E. Jaeger, W. Busch, M. Schmid, J. U. Lohmann and D. Weigel, Integration of spatial and temporal information during floral induction in Arabidopsis, *Science*, 2005, 309(5737), 1056–9. doi:10.1126/science.1114358.
11. S. D. Michaels and R. M. Amasino, FLOWERING LOCUS C encodes a novel MADS domain protein that acts as a repressor of flowering, *Plant Cell*, 1999, 11(5), 949–56.

12. U. Hartmann, S. H¨ohmann, K. Nettesheim, E. Wisman, H. Saedler and P. Huijser, Molecular cloning of SVP: a negative regulator of the floral transition in Arabidopsis, *The Plant Journal*, 2000, 21(4), 351–60. doi:10.1046/j.1365-313x.2000.00682.x.
13. M. A. Bl´azquez, L. N. Soowal, I. Lee and D. Weigel, LEAFY expression and flower initiation in Arabidopsis, Development, 1997, 124(19), 3835–44.
14. F. D. Hempel, D. Weigel, M. A. Mandel, G. Ditta, P. C. Zambryski, L. J. Feldman and M. F. Yanofsky, Floral determination and expression of floral regulatory genes in Arabidopsis, Development, 1997, 124(19), 3845–53.
15. M. A. Busch and B. K. D. Weigel, Activation of a floral homeotic gene in Arabidopsis, *Science*, 1999, 285(5427), 585–7. doi:10.1126/science.285. 5427.585.
16. S. A. Kempin, B. Savidge and M. F. Yanofsky, Molecular basis of the cauliflower phenotype in Arabidopsis, *Science*, 1995, 267(5197), 522–5.
17. C. Smaczniak, R. G. Immink, J. M. Muin˜o, R. Blanvillain, M. Busscher, J. Busscher-Lange, Q. D. Dinh, S. Liu, A. H. Westphal, S. Boeren, F. Parcy, L. Xu, C. C. Carles, G. C. Angenent, K. Kaufmann, Characterization of MADS-domain transcription factor complexes in Arabidopsis flower development, Proceedings of the National Academy of Sciences of the United States of America, 2012, 109(5), 560–65. doi:10.1073/pnas. 1112871109.
18. G. H. Cardon, S. Hohmann, K. Nettesheim, H. Saedler and P. Huijser, Functional analysis of the Arabidopsis thaliana SBP-box gene SPL3: a novel gene involved in the floral transition, *The Plant Journal*, 1997, 12(4), 367–77. doi:10.1046/j.1365-313X.1997.12020367.x.
19. S. Schwarz, A. V. Grande, N. Bujdoso, H. Saedler and P. Huijser, The microRNA regulated SBP-box genes SPL9 and SPL15 control shoot maturation in Arabidopsis, *Plant Molecular Biology*, 2008, 67(1-2), 183–95. doi:10.1007/s11103-008-9310-z.
20. J. W. Wang, B. Czech and D. Weigel, miR156-regulated SPL transcription factors define an endogenous flowering pathway in Arabidopsis thaliana, Cell, 2009, 138(4), 738–49. doi:10.1016/j.cell.2009.06.014.
21. A. Yamaguchi, M. F. Wu, L. Yang, G. Wu, R. S. Poethig and D. Wagner, The microRNA-regulated SBP-Box transcription factor SPL3 is a direct upstream activator of LEAFY, FRUITFULL, and APETALA1, *Developmental Cell*, 2009, 17(2), 268–78. doi:10.1016/j.devcel.2009.06.007.
22. M. A. Mandel, C. Gustafson-Brown, B. Savidge and M. F. Yanofsky, Molecular characterization of the Arabidopsis floral homeotic gene APETALA1, *Nature*, 1992, 360(6401), 273–7. doi:10.1038/360273a0.
23. M. Ng, M. F. Yanofsky, Activation of the Arabidopsis B class homeotic genes by APETALA1, *Plant Cell*, 2001, 13(4), 739–53.
24. P. P´erez-Rodr´ıguez, D. M. Rian˜o-Pach´on, L. G. G. Corrˆea, S. A. Rensing, B. Kersten and B. Mueller-Roeber, PlnTFDB: updated content and new features of the plant transcription factor database, *Nucleic Acids Research*, 2010, 38, D822–D827. doi:10.1093/nar.
25. R Core Team, R: A Language and Environment for Statistical Computing, R Foundation for Statistical Computing, Vienna, Austria, 2015. URL http://www.r-project.org
26. G. Csardi and T. Nepusz, The igraph software package for complex network research, *Inter. Journal Complex Systems*, 2006, 1695. URL http://igraph.org
27. X. Yi, Z. Du and Z. Su, PlantGSEA: a Gene Set Enrichment Analysis toolkit for plant community, *Nucleic Acids Research*, 2013, 41(W1), 98–103. doi: 10.1093/nar.

28. F. Supek, M. Bosnjak, N. Skunca and T. Smuc, REVIGO summarizes and visualizes long lists of Gene Ontology terms, *Plos One*, 2011, 6(7), e21800. doi:10.1371/journal.pone.0021800.
29. J. M. Franco-Zorrilla, P. Cubas, J. A. Jarillo, B. Fern´andez-Calv´n, J. Salinas and J. M. Mart´nez-Zapater, AtREM1, a Member of a New Family of B3 Domain-Containing Genes, Is Preferentially Expressed in Reproductive Meristems, Plant Physiology 128 (2) (2002) 418–427.
30. O. Mantegazza, V. Gregis, M. A. Mendes, P. Morandini, M. Alves-Ferreira, C. M. Patreze, S. M. Nardeli, M. M. Kater and L. Colombo, Analysis of the arabidopsis REM gene family predicts functions during flower development, *Annals of Botany*, 2014, 114(7), 1507–15. doi:10.1093/aob.
31. Y. Oono, M. Seki, M. Satou, K. Iida, K. Akiyama, T. Sakurai, M. Fujita, K. Yamaguchi-Shinozaki and K. Shinozaku, Monitoring expression profiles of Arabidopsis genes during cold acclimation and deacclimation using DNA microarrays, *Functional & Integrative Genomics,* 6 (14). doi:10.1007/ s10142-005-0014-z.
32. K. Ohashi-Ito and D. C. Bergmann, Arabidopsis FAMA Controls the Final Proliferation/Differentiation Switch during Stomatal Development, *American Society of Plant Biologists*, 2006, 18(10), 2493–2505.
33. Y. Zhang, S. Schwarz, H. Saedler and P. Huijser, SPL8, a local regulator in a subset of gibberellin-mediated developmental processes in Arabidopsis, *Plant Molecular Biology*, 2007, 63(3), 429–439. doi:10.1007/ s11103-006-9099-6.
34. S. A. Jorgensen and J. C. Preston, Differential SPL gene expression patterns reveal candidate genes underlying flowering time and architectural differences in Mimulus and Arabidopsis, *Molecular Phylogenetics and Evolution*, 2014, 73, 129–39. doi:10.1016/j.ympev.2014.01.029.
35. M. K. Chen, W. H. Hsu, P. F. Lee, M. Thiruvengadam, H. I. Chen and C. H. Yang, The MADS box ge ne, FOREVER YOUNG FLOWER, acts as a repressor controlling floral organ senescence and abscission in Arabidopsis, *The Plant Journal*, 2011, 68(1), 168–85. doi:10.1111/j.1365-313X. 2011.04677.x.
36. W. H. Chen, P. F. Li, M. K. Chen, Y. I. Lee and C. H. Yang, FOREVER YOUNG FLOWER Negatively Regulates Ethylene Response DNA-binding Factors (EDFs), by Activating An Ethylene-Responsive Factor (ERF), to Control Arabidopsis Floral Organ Senescence and Abscission, Plant Physiology 10.

Corresponding author: Sachiyo Aburatani;

Biotechnology Research Institute for Drug Discovery, National Institute of AIST, Tokyo, Japan