Exploring the Relationship between Gene Expression and Topological Properties of *Arabidopsis thaliana* Interactome Network.

Silvia M. Gimenez[1], Guillermo Rodrigo[#2], Brian Jiménez[+3], Juan Fernandez-Recio[+4].

[#]*Instituto de Biología Molecular y Celular de Plantas, Consejo Superior de Investigaciones Científicas – Universidad Politécnica de Valencia, 46022, Valencia, Spain*

[+]*Barcelona Supercomputing Center (BSC). Joint BSC-CRG-IRB Research Program in Computational Biology, Barcelona, Spain*

[1]silvia.santam@gmail.com, [2]guirodta@ibmcp.upv.es, [3]brian.jimenez@bsc.es, juanf@bsc.es

Systems Biology, Protein-protein Interaction Network, Statistical correlations.

EXTENDED ABSTRACT

A. Background

Regulatory mechanisms as dynamical responses by a living cell, that is, how it reacts and adapts to a different environmental stimuli, is highly unknown yet at molecular level.

In the field of Systems biology research, models of interaction networks are a powerful tool to approach living complex systems. Presented as a method that facilitates the understanding the structure and dynamics of complex intercellular web of interactions that contribute to the structure and function of a living cell [1].

The protein-protein interaction (PPI) network of *Arabidopsis thaliana* L., previously validated through experimental procedures and published, has the characteristic properties of hierarchical networks [2, 3]. It is shown that PPI networks are dynamically organized into functional modules.

The aim of this study is to integrate and link up transcriptomic data with biological networks approaches. The main objective was to determinate the correlation of transcriptomic profiles with PPI topology, seeking to demonstrate relational or structural patterns within the network internal organization.

B. Results

Correlation coefficients between gene expression profiles derived from microarray data evaluated at steady-state RNA expression levels from several growth conditions, developmental stages biotic and abiotic stresses, and a variety of mutant genotypes [4], and topological properties, such as connectivity degree or clustering coefficient, in PPI context have been evaluated (Fig 1, section b). Genes with high and low expression profiles were located within the global PPI network (Fig 1, section a).

Based on microarray data analyzed, a matrix of coexpression between expressed genes was calculated. Twelve coexpressed nodes were identified and highlighted into the network (Fig 1, section c). Selecting its direct interactions and nodes, a model of functional modules was proposed. Table 1 contains a functional analysis of nodes involved in the proposed modules. Connector nodes between modules were detected and analyzed.

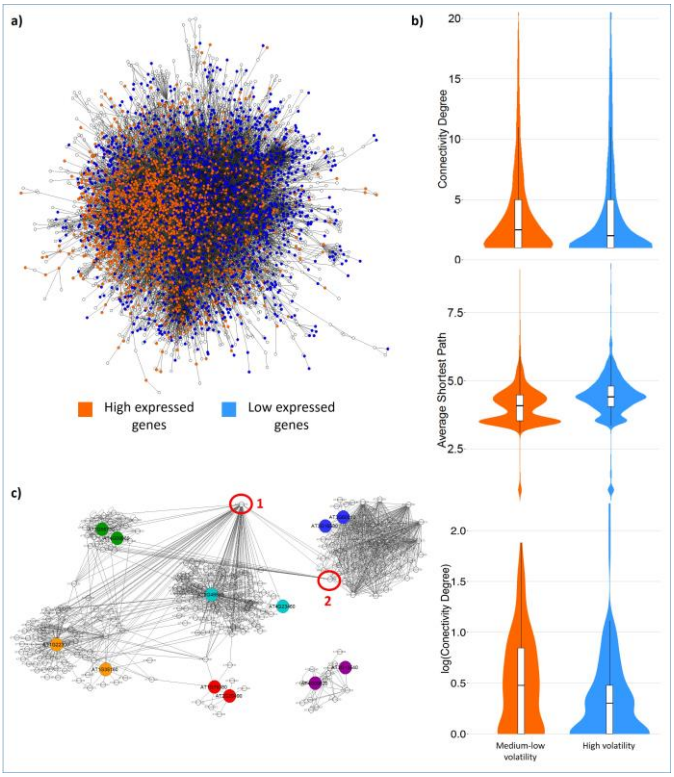

Fig 1 a) Highlight of nodes that represent genes with high and low gene expression level, included in correlation tests, located at its global *A. thaliana* interactome context. b) Violin plots combined with boxplots as representation of significant U-test evaluating correlations. pV = 1.88e-6; pV < 2.2e-16; pV = 0.00053. c) Highlight of coexpressed genes and its functional modules proposed. Circled potential modules regulatory targets, UBQ3 (1) and FBW2 (2) genes.

TABLE I
FUNCTIONAL MODULES NODES DESCRIPTION

Coexpressed nodes	Gene ID	Biological activity	Nº genes
AT1G78380 AT2G29490	GST8 GST19	Glutation transferase activity	9
AT3G16580 AT3G60010	SK13 --	F-Box proteins. Unknown function	77
AT1G22300 AT1G35160	GF14 GF14 PHI	ATP binding	46
AT3G58780 AT4G09960	AGL1/SHP1 AGL11/STK	Transcriptional factors	22
AT3G13540 AT4G09820	MYB5 ATTT8/TT8	Transcriptional factors.	12
AT5G46630 AT4G23460	AP2M --	Protein intercellular transport	25

C. Conclusion

Correlations between gene expression level and topological properties were determined, being positive for connectivity degree (*C*) and negative in relation with shortest path length (*l*). Furthermore, a negative correlation was determined for the relation between volatility coefficient of gene expression levels and connectivity degree (*C*). This conclusions implies that most connected nodes are the genes with higher expression level. At the same time they present the capability to transmit information through the network faster, connecting any other arbitrary node passing by less number of nodes. In addition, related to volatility, the most connected nodes present less fluctuating gene expression levels.

We identified 12 important genes showing relation of coexpression and direct interaction in the network; their functional modules were predicted and characterized. Determinant nodes, understood as connector elements between functional modules were detected. UBQ and FBW2 genes, detected as regulatory network nodes, have important roles in suppression of gene silencing and protein ubiquitination. Those genes are potential targets involved in the transmission procedures of biological information through the network.

D. Future develop/ways

The present study evidences the requirement of a time parameter integration for dynamic network future studies, to well understand how molecular regulatory mechanisms regulate the dynamic responses to environmental stimuli and intracellular signals.

References

[1] H. Kitano. “Systems Biology: A Brief Overview”. Science, Vol. 295, Pp. 1662-1664, 2002.

[2] A.L. Barabási, Z.N. Oltvar. “Network biology: understanding the cell’s functional organization”. Nat. Rev. Genet, Vol. 5, No. 2, Pp. 101-113, 2004.

[3] T. Ideker, N.J. Krogan. “Differential network biology”. Mol. Syst. Biol., Vol. 8, Pp. 565, 2012.

[4] J. Carrera, G. Rodrigo, A. Jaramillo, S. Elena. “Reverse-engineering the *Arabidopsis thaliana* transcriptional network under changing environmental conditions”. Genome biology, Vol. 10, No 9, Pp. R96, 2009.

Author biography

Silvia M. Giménez Santamarina was born in Valencia, Spain, in 1992. He received the B.E. degree in Biotechnology from the Catholic University of Valencia, Spain, in 2015. Nowadays she is finishing the MS. degree in Bioinformatics and Biostatistics from the Universitat Oberta de Catalunya (UOC) Barcelona, Spain, currently in the last semester.

Since 2014, she has been focused in to learn and growth in the field of Bioinformatics. Her first work experience in this field was in Brazil during a Bachelor’s traineeship. After that, her Bachelor Final Thesis was focused on the study of the mode of action of a virus, integrating network biology approaches, and the design and implementation of a software to map proteomic data from non-model organism to a model organism. In 2015 she did a second Bachelor’s traineeship in Uppsala, Sweden, at SLU University. She worked analyzing RNA-Seq data and differential gene expression. The presented study is her MS Final Thesis, which obtained the maximum punctuation. Currently she is working at the BSC, Life Science department, with the Protein Interactions and Docking team, lead by Juan Fernández-Recio.