sinc(*i*) Research Center for Signals, Systems and Computational Intelligence (fich.unl.edu.ar/sinc)
L. Kamenetzky, M. López, G. Stegmayer, D. H. Milone, J. Giovannoni, A. Fernie & F. Carrari; "Metabolic pathway reconstruction by a neural clustering"
1er Congreso Argentino de Bioinformática y Biología Computacional, may, 2010.

Metabolic pathway reconstruction by a neural clustering

Laura Kamenetzky[1], Mariana López[1], Georgina Stegmayer[2], Diego Milone[3], James Giovanonni[4], Alisdair Fernie[5] and Fernando Carrari[1]. [1]IB-INTA, Argentina. [2]CIDISI, UTN-FRSF, CONICET, Argentina. [3]Sinc(*i*), FICH-UNL, CONICET, Argentina. [4]BTI, Cornell, USA. [5]MPIMP-Golm, Germany.

Generation of large scale datasets from 'omics' technologies requires the integration of heterogeneous information in order to describe and reconstruct complex biological networks. Several analytical tools have been developed to identify patterns of gene expression that are responsible for potent biological effects by integrating large-scale transcriptomic data with diverse biological information such as pathways and associated metabolites [1]. The aim of this work is to assess the powerful of neuronal clustering for the discovery of new relationships between gene expression and metabolite content in ripen tomato fruits. Data input consisted in expression µarray and metabolic profiles obtained from fruits of tomato cultivars differing in their genomic structure (introgressed lines, ILs). After pre-processing, filtering, selection and normalization steps, 1159 transcripts and 70 metabolites were selected. Variation patterns of these molecular entities were integrated and analyzed with a self-organizing neural network model named IL-SOM [2](Figure 1).

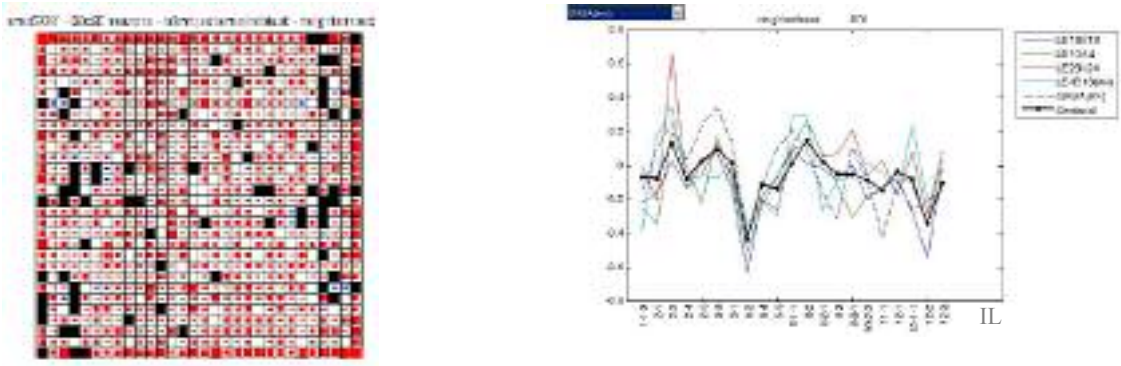

Figure 1. Snapshot of IL-SOM visualizations. Left: IL-SOM map painted according to type of patterns grouped in neurons: black (metabolites and transcripts), blue (only metabolites) and red (only transcripts). Right: detail of data clustered in neuron 876.

According to the total data points, a range of candidate IL-SOMs were proposed having 225 to 1600 neurons. The maps in this range have been tested on several training configurations, including: number of epochs, neighborhood functions, learning rate reductions, influence neighborhood of the winning neuron, batch versus sequential training and types of topology grids. Map quality was evaluated considering only nodes grouping both metabolites and transcripts data. Thus, a 30x30 IL-SOM trained over 100 epochs with the standard batch-training learning algorithm was selected as the reference map. All obtained clusters were linked to currently available databases (SGN http://solgenomics.net/ and KEGG, http://www.genome.ad.jp/kegg/). Preliminary results allowed reconstructing an exemplary network: the known GABA biosynthetic pathway. Moreover, elements from an unrelated metabolic pathways (i.e anthocyanin biosynthesis) grouped with various components of the GABA metabolism suggests novel links between these two pathways (Table 1).

Table 1. Neighborhood neurons clustering transcripts and metabolites from known metabolic pathways.

Neuron #	Transcript/**Metabolite**	Metabolic pathway (KEGG)	Cohesion
876 (Vn=0)	**GABA (γ-aminobutyric acid)**	GABA shunt (ko00250)	0.674
	LE23K24 (Transmembrane amino acid transporter protein)		
	LE4B18 (anthocyanin acyltransferase)	Anthocyanin biosynthesis (ko00942)	
875 (Vn=1)	**Arginine**	Arginine and proline biosynthesis (ko00330)	0.705
	Ornithine		
874 (Vn=2)	LE10E20 (glycosyl hydrolase)	Citrate cycle (ko00020)	0.694
	LE13L02 (UDP-glucosyltransferase)	Anthocyanin biosynthesis (ko00942)	
	Proline	Arginine and proline biosynthesis (ko00330)	

Vn= visualization neighborhood. Cohesion: average measure of the distance between each data point in a cluster and the cluster centroid.

Conclusion: The IL-SOM model presented in this work provides easy identification of data clusters that groups both metabolites and transcripts having similar measured values in each tomato cultivar facilitating pathway reconstruction of metabolic networks.

References:

1. Szymanski J, Bielecka M, Carrari F, Fernie AR, Hoefgen R, Nikiforova VJ. On the processing of metabolic information through metabolite-gene communication networks: an approach for modelling causality. Phytochemistry. 2007, 68:2163-75.

2. Stegmayer G., Milone D., Kamenetzky L., Lopez M., Carrari F., Neural network model for integration and visualization of introgressed genome and metabolite data,, pp. 2983-2989, International Joint Conference on Neural Networks (IJCNN), 2009.