



UNIVERSITY OF HELSINKI

# Gene expression profiling using clustering and functional annotations

P124

Knuutila J. E. A.<sup>1</sup>, Törönen P.<sup>2</sup> & Castrén E.<sup>1</sup>  
(juha.knuutila@helsinki.fi)

1. Neuroscience Center, University of Helsinki,  
FINLAND

2. A. I. Virtanen Institute  
for Molecular Sciences, University of Kuopio,  
FINLAND

## INTRODUCTION & AIMS

Microarrays can be used to identify candidate genes that might be involved in a disease or treatment of interest.

However when validated by methods such as qPCR or *in situ* hybridisation, microarray experiments are shown to produce substantial amount false positives.

To facilitate the process, we suggest the use of bioinformatic validation methods prior to the wet lab validation.

## MATERIALS & METHODS

We measured gene expressions in rat *prefrontal cortex* at different time points during chronic antidepressant drug (Imipramine) treatment. Expression profiles of 1090 genes were clustered using the Self-Organizing Map (SOM).

The SOM algorithm has an ability to plot the gene expression profiles to an 2-dimensional plane so that:

- similar expression patterns are placed in the same proximity and
- dissimilar patterns are placed far off from each other

So there is dependency also between adjacent nodes and not only inside the node!

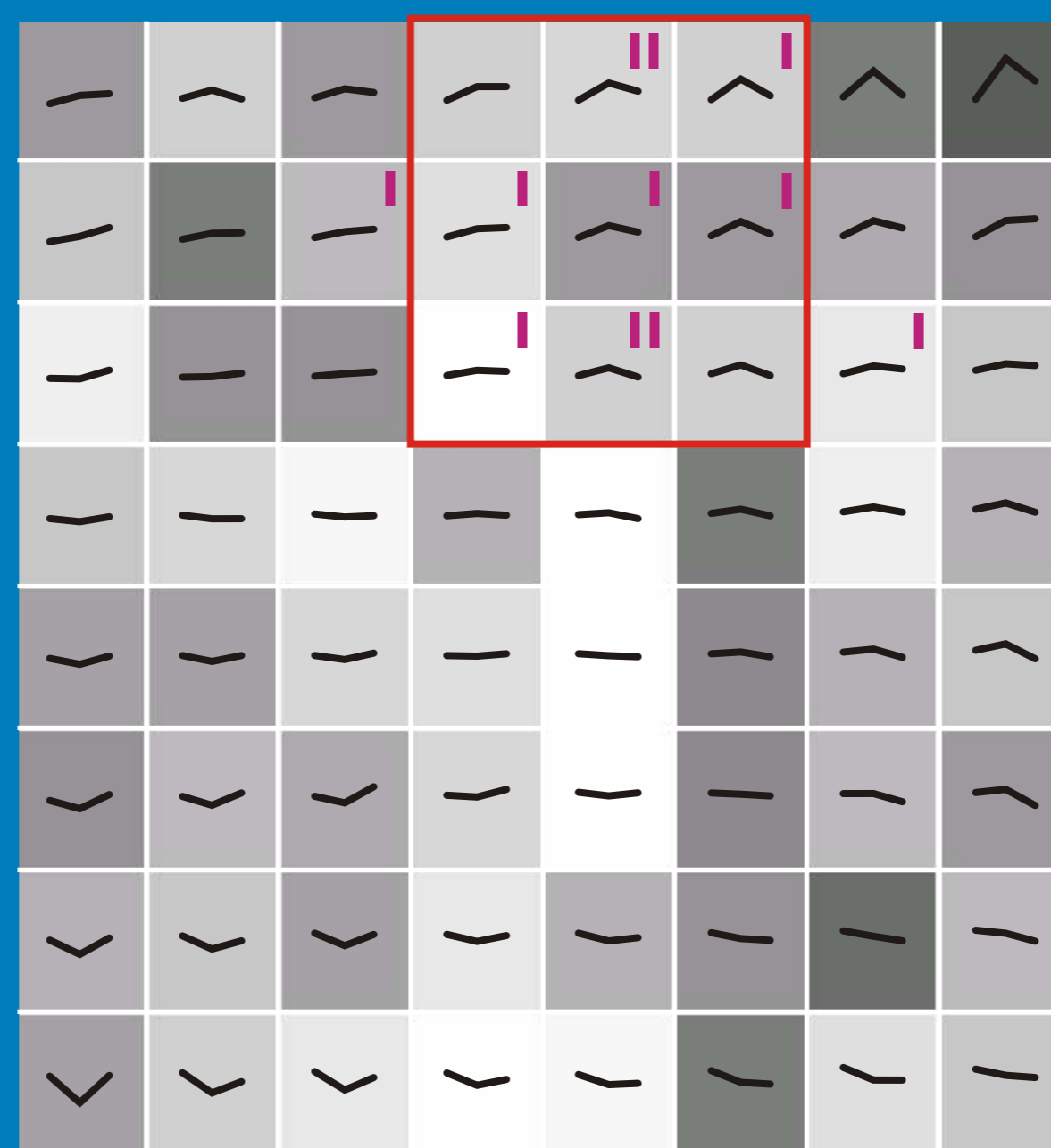
We then analysed whether genes belonging to the same functional group (which were formed using keywords from Swiss-Prot) were overrepresented in a defined cluster area.

## RESULTS

We found that several keywords showed correlation to a specific sub-area of the SOM.

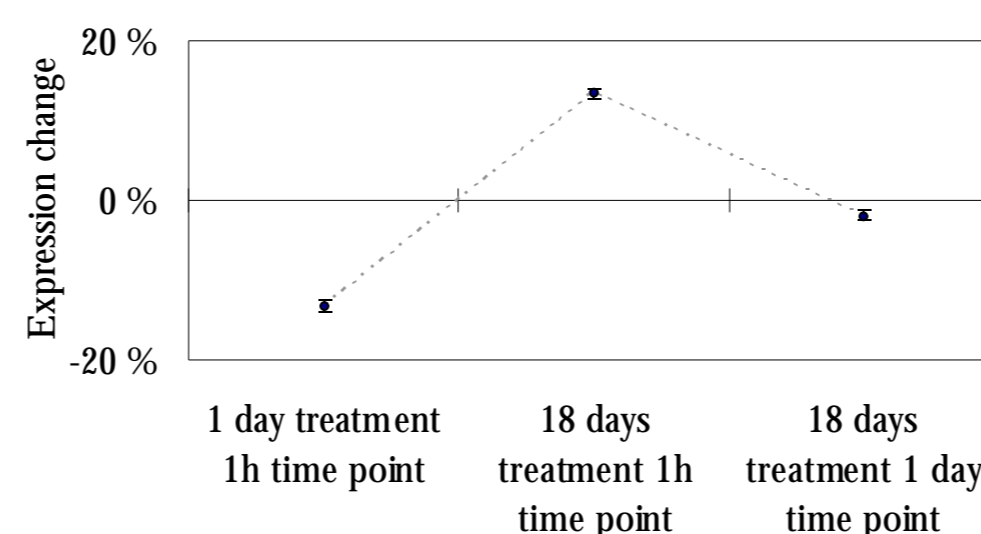
For example, genes annotated by the *cAMP* and *cAMP synthesis* among other keywords had similar expression patterns. (Figure 1)

Figure 1. Gene expression profiles placed on a 2-dimensional plane by Self-Organizing Map algorithm



- Individual black curve is the average expression pattern of those genes located in the area of one node (64 little squares)
- Purple bars indicate the map location of genes that are annotated by *cAMP* and *cAMP synthesis* keywords by Swiss-Prot data bank. Note that the expression patterns of these genes are located in the same proximity
- Red square defines an area of interest that was further analyzed
- Brighter color in the map indicate that those nodes contain more genes than the darker ones

Average expression pattern of genes in the cluster



Functional Classes Overrepresented in the Cluster

Key word class	P value from Fisher's exact test	Size of the class	Observed number of genes	Expected number of genes
cAMP & cAMP synthesis	0,00000178	11	9	1,71
Lyase	0,00055919	25	11	3,88
Serine protease	0,00120425	16	8	2,48

Number of genes in this cluster area: 169 genes out of the total number of 1090 genes

cAMP & cAMP synthesis genes in this cluster area	Genbank #
Adenylyl cyclase 2	M80550
Adenylyl cyclase 3	M55075
Adenylyl cyclase 4	M80633
Adenylyl cyclase 5	M96159
Adenylyl cyclase 6	L01115
Adenylyl cyclase 8	L26986
cAMP-specific phosphodiesterase 4B	J04563
cAMP-specific phosphodiesterase 4C	M25347
cAMP-specific phosphodiesterase 4D	U09457

Lower panel: Genes belonging to the key word class "cAMP" and "cAMP synthesis" within the cluster area

For the following validation steps we have then selected not only those genes that are the most differentially expressed but also those gene groups which showed synexpression but were not necessarily evidently differentially expressed.

## CONCLUSION

Validation of microarray data can be quite laborious task and careful considerations should be made when selecting which genes are meaningful in the in context of experiment set-up.

Selection should not be made based on the previous knowledge of genes and experiment in hand because this might prevent researcher to find new insights about the biological mechanisms laying behind the test set up.

In this study we observed that several gene classes are synexpressed throughout the time course. Some of these results where something that we were not expecting and without our bioinformatical analysis we would not have thought these genes to be interesting for the future studies.

## REFERENCES

- Kohonen T. 2001 in T.S. Huang and M.R. Schoeder. (eds.), *Self-Organizing Maps*, ed. 3., Springer, Berlin.
- Törönen, P., Kolehmainen, M., Wong, G., and Castrén, E. 1999. *Analysis of gene expression data using self-organizing maps. FEBS Lett.* 451:142-6.
- Nikkilä, J., Törönen, P., Kaski, S., Venna, J., Castrén, E., and Wong, G. 2002. *Analysis and visualization of gene expression data using Self-Organizing Maps. Neural Netw.* 15:953-66.
- Boeckmann, B., Bairoch, A., Apweiler, R., Blatter, M.C., Estreicher, A., Gasteiger, E., Martin, M.J., Michoud, K., O'Donovan, C., Phan, I., Pilbout, S., and Schneider, M. 2003. *The SWISS-PROT protein knowledgebase and its supplement TrEMBL in 2003. Nucleic Acids Res.* 31:365-70.