# Generation and analysis of a human-mouse conserved co-expression network

ID - 198

Ala Ugo[1], Provero Paolo[1], Piro Rosario[1], Damasco Christian[1], Grassi Elena[1], Di Cunto Ferdinando[1]
[1]Molecular Biotechnology Center, University of Turin, Torino

**Motivation**

Genes and proteins of living organisms deploy their functions through a complex series of interactions with other genes and proteins. These relationships can be more or less direct, and can be inferred from different types of experimental evidences.

The most obvious relation is a direct molecular interaction, which can be shown both with biochemical methods and with molecular biological techniques such as the yeast two hybrid system. Nevertheless, very close functional relationships are even possible in the absence of direct molecular binding. Considering that genes involved in the same functions tend to show very similar expression pattern and given the availability of massive gene expression data repository, co-expression analysis represents one of the most powerful tools for exploring the complexity of functional relationships among genes. In particular, phylogenetic conservation of co-expression has been proposed as a very strong criterion to identify functionally relevant links among genes. We have previously described CLOE (Pellegrino et al., MBC Bioinformatics 2004), a data-mining approach based on such kinds of meta-analysis which enables to make high confidence predictions about proteins function and interaction. As a logical evolution, we have applied this method at a global-scale and here we present a network of the co-expression relationships conserved between human and mouse, based on the analysis of cDNA microarray databases.

**Methods**

DNA microarray data have been obtained from published studies performed with cDNA microarray technology. In our work, we collected human and mouse data from the Stanford Microarray Database (http://smd.stanford.edu/cgi-bin/search/QuerySetup.pl) which is the most relevant repository of microarray experiments. All the probes were identified by GeneBank id and we have completely re-mapped them to the most significant biological databases, and in particular, Unigene, Entrez Gene and Ensembl. Relationships between orthologous genes were assigned on the base of HomoloGene tables. The computational analysis starts with the generation, for every probe in the dataset, of all the other datasets probes, ordered by a co-expression index, computed with the Pearsons correlation coefficient. The top 1% of every list is selected and we collapse probes lists referred to the same gene (identified by Entrez Gene id). We introduce phylogenetic conservation to compare lists of orthologous genes. We construct the network using meta-genes (where a meta-gene represents an orthologous genes pair) as nodes and one-to-one phylogenetic conserved co-expression links as edges. We have performed a statistical analysis to evaluate the enrichment in Gene Ontology (GO) terms for all the genes and in order to explore the possible predictive value in terms of human phenotypes, we focused our attention on the OMIM terms used in MimMiner (http://www.cmbi.ru.nl/MimMiner/cgi-bin/main.pl).

Finally, we have evaluated the overlapping of this network with literature and two hybrid-based human interactomes.

**Results**

Our preliminary results have revealed that nearly 30% of the lists, composed by the first neighbours of every gene, is enriched for at least one GO term and that there is a seven time enrichment in number of links between two genes characterized by related disease phenotype.

These results strongly suggest that co-expression relationships conserved between human and mouse are very relevant for exploring the function of mammalian genes and for studying human genetic disease.

**Email:** ugo.ala@unito.it