Disease gene prediction based on tissue-specific conserved coexpression

Rosario M. Piro, Ugo Ala, Ivan Molineris, Elena Grassi, Christian Damasco, Chiara Bracco, Paolo Provero, Ferdinando Di Cunto
Molecular Biotechnology Center and Department of Genetics, Biology and Biochemistry
University of Turin, Italy – Contact: **rosario.piro@unito.it**

We discuss a method for disease gene prediction based on high-throughput microarray expression data that uses the conservation of coexpression as a powerful filter for biological significance and allows to specifically focus on tissue-specific relationships between disease and candidate genes, providing novel high-confidence candidates for several genetic disorders.

METHOD: We build both multi-tissue and tissue-specific conserved coexpression networks (CCNs) and exploit the gene coexpression clusters they provide to uncover global and tissue-specific relationships between candidate genes and "reference genes" known to be involved in the given phenotype and/or similar disorders.

RESULTS: We show that the novelty of our approach, the tissue-specificity, provides disease-related information (and thus biologically meaningful predictions) that is highly complementary to the information obtained from multi-tissue coexpression. We discuss the obtained results and present a user-friendly web tool for custom analysis.

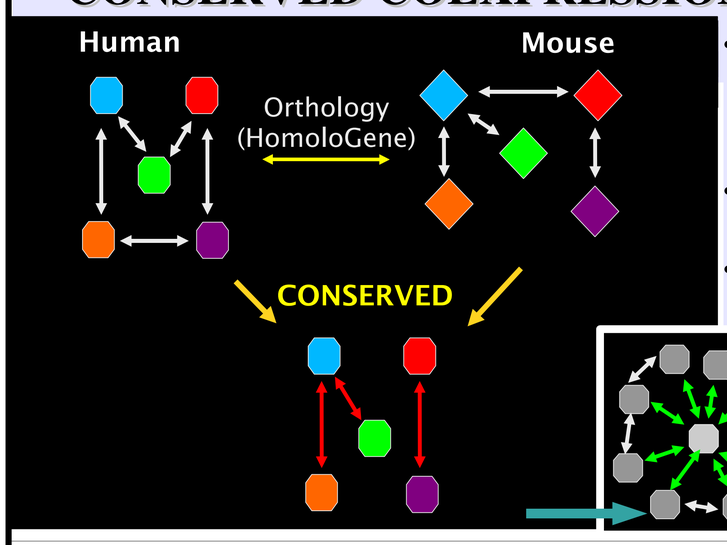
DISCUSSION: The conservation of coexpression between human and mouse allows to focus on biologically relevant relationships that have been preserved during evolution. The analysis of tissue-specific coexpression profiles takes into account the dynamic nature of such relationships that may vary over different tissues, cell types and/or conditions. The integration of both concepts, the tissue-specific conserved coexpression, can provide important clues about molecular mechanisms underlying human hereditary disorders and thus provide a powerful tool for disease gene prediction.

Webtool available at http://www.cbu.mbcunito.it/ts-coexp/

CONSERVED CO-EXPRESSION:

Phylogenetic conservation as a very strong criterion to **identify functionally relevant coexpression** links between genes [1][2]: significant coexpression that is phylo-genetically conserved is likely due to selective advantage, suggesting a functional relation. Conserved co-expression can be used for **disease gene prediction** [3][4].

CONSERVED COEXPRESSION NETWORKS (CCNs):



- Single Species Networks (SSNs): significant coexpression if two genes are among the 1% most coexpressed genes of each other.
- CCNs: genes coexpressed in both species; intersection of human SSN and mouse SSN.
- Tissue-/condition-specific SSNs/CCNs: network construction from a selected subset of the expression data.
- Conserved coexpression clusters:
  – gene plus all its next neighbors
  – one cluster for each gene in the network

Entrez GEO DataSets

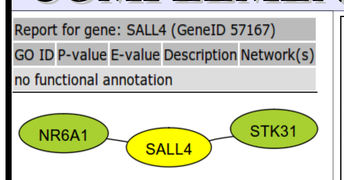
Human: 5,188 experiments on Affymetrix HG-U133 Plus 2.0
Mouse: 2,310 experiments on Affymetrix MG-430 2.0

To allow the selection of tissue- and/or condition-specific subsets, each experiment was classified (manual curation) according to: **anatomical annotation** (hierarchical MeSH ontology), **condition** (normal, tumor, other disease/condition), **stage** (adult, embryo), **sample type** (tissue, cell line)

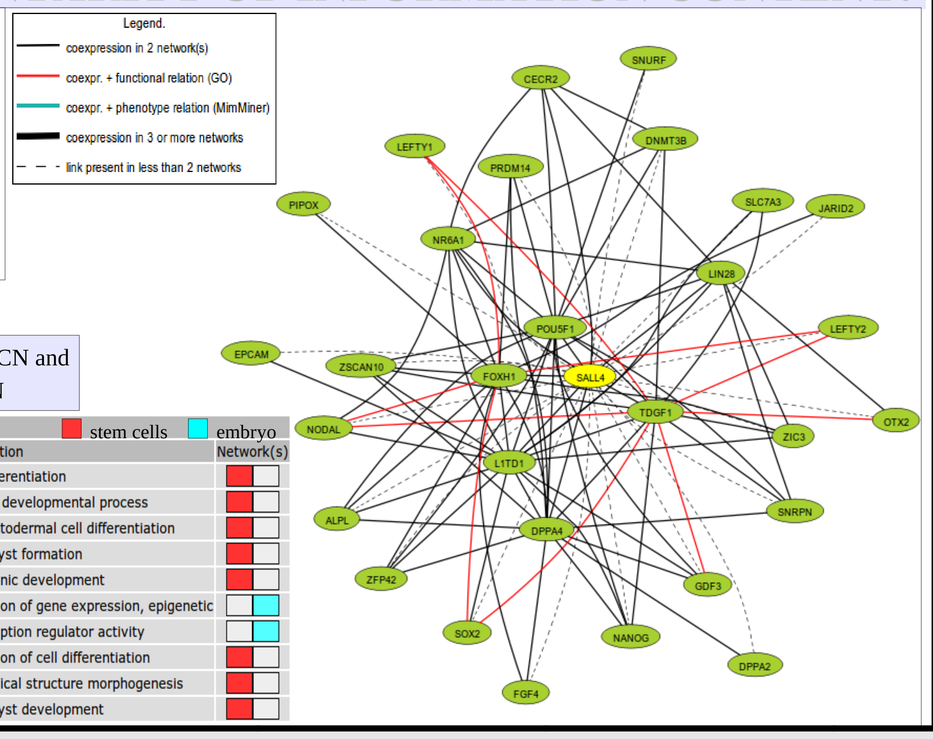
TISSUE-SPECIFIC CO-EXPRESSION:

While many functionally relevant expression correlations can be extracted from hetero-geneous (multi-tissue) datasets, many others are likely to remain unnoticed, because they **specifically occur in one or few tissues, cell types and/or conditions** [4].

COMPLEMENTARITY OF INFORMATION CONTENT:



SALL4 in the CCN derived from " all samples"



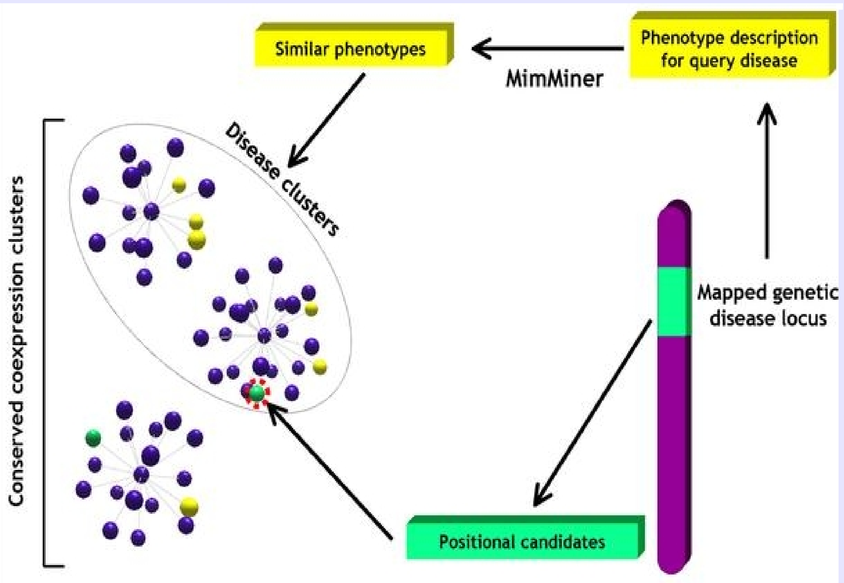
SALL4 in the " stem cells" CCN and the " embryonic tissues" CCN

Report for gene: SALL4 (GeneID 57167) — stem cells — embryo

GO ID	P-value	E-value	Description	Network(s)
GO:0030154	2.55242e-07	0.00263920228	cell differentiation	
GO:0048869	5.09597e-07	0.00526923298	cellular developmental process	
GO:0001829	5.24509e-06	0.0542342306	trophectodermal cell differentiation	
GO:0001825	9.04939e-06	0.0935706926	blastocyst formation	
GO:0009790	1.80276e-05	0.186405384	embryonic development	
GO:0040029	2.20393e-05	0.227886	regulation of gene expression, epigenetic	
GO:0030528	3.76275e-05	0.389068	transcription regulator activity	
GO:0045595	4.56878e-05	0.472411852	regulation of cell differentiation	
GO:0009653	5.6349e-05	0.58264866	anatomical structure morphogenesis	
GO:0001824	7.07775e-05	0.73183935	blastocyst development	

DISEASE GENE CANDIDATE SELECTION:

As **best candidates** among the *positional candidate genes* in a **disease-associated orphan locus** we select those that appear in a gene cluster together with at least two *reference genes* known to be involved in similar phenotypes (i.e. they ***show conserved co-expression with other genes that cause similar phenotypes***)



Importance: 1,074 mapped "orphan loci" for 1,028 OMIM phenotypes with unknown molecular basis (June 2009)

Contemporary use of gene gene clusters from **both multi-tissue and tissue-specific CCNs**

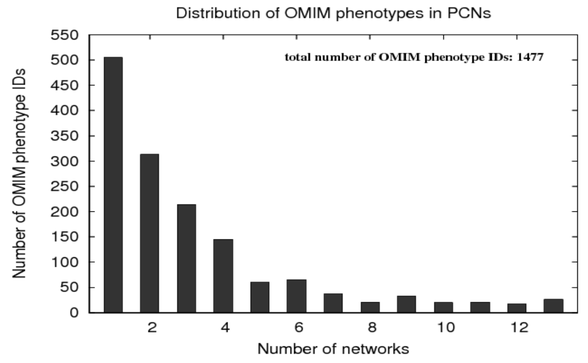
Application: our predictions for 238 loci associated to OMIM phenotypes with un-known molecular basis can be found in [4].

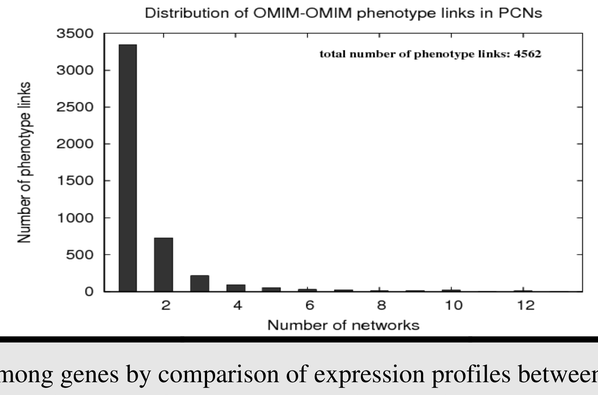
PHENOTYPE SIMILARITY: MIMMINER

We use **MimMiner** [5] to determine the similarity between two OMIM phenotype entries. MimMiner provides normalized scores for phenotype similarity (from 0 to 1); a threshold of 0.4 is used to denote similarity [5]. Important: the **integration of phenomics** allows to predict disease genes also for OMIM phenotypes with so far **unknown molecular basis** (OMIM category: %).

PHENOTYPE COEXPRESSION NETWORKS (PCNs):

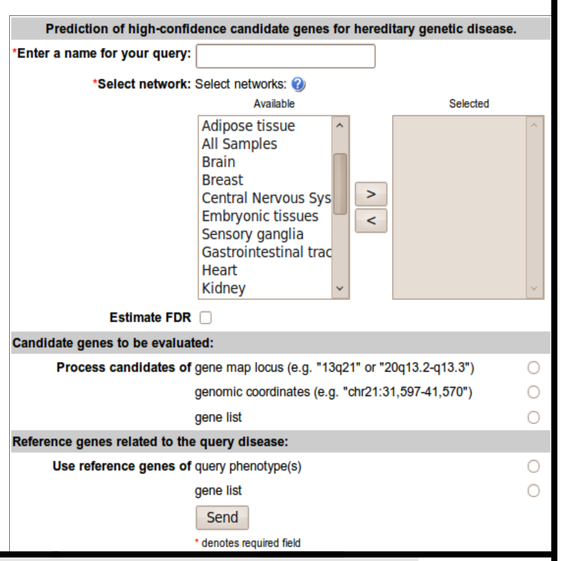
Translating gene-gene links from CCNs into **OMIM-OMIM links** between similar phenotypes shows that CCNs contains both **significant and complementary information** regarding relationships between disorders and their associated genes.





TS-COEXP Webtool:

20 prebuilt CCNs; CCN generation engine; GO enrichment analysis, and **disease gene prediction interface:**



REFERENCES: 1. Pellegrino M, *et al.* (2004) CLOE: identification of putative functional relationships among genes by comparison of expression profiles between two species. *BMC Bioinformatics* 5: 179.
2. Stuart JM, Segal E, Koller D, Kim SK (2003) A Gene-Coexpression Network for Global Discovery of Conserved Genetic Modules. *Science* 302: 249–255.
3. Ala U, Piro RM, Grassi E, Damasco C, Silengo L, *et al.* (2008) Prediction of Human Disease Genes by Human-Mouse Conserved Coexpression Analysis. *PLoS Comput Biol* 4(3): e1000043.
4. Piro RM, Ala U, Molineris I, Grassi E, Bracco C, *et al.* (submitted) An Atlas of Tissue-Specific Conserved Coexpression for Functional Annotation and Disease Gene Prediction.
5. van Driel MA, Bruggeman J, Vriend G, Brunner HG, Leunissen JA (2006) A text-mining analysis of the human phenome. *Eur J Hum Genet* 14: 535–542.