

Analysis of tissue-specific co-expression networks

Somaye Hashemifar, Daniela Boernigen, Natalia Maltsev and Jinbo Xu

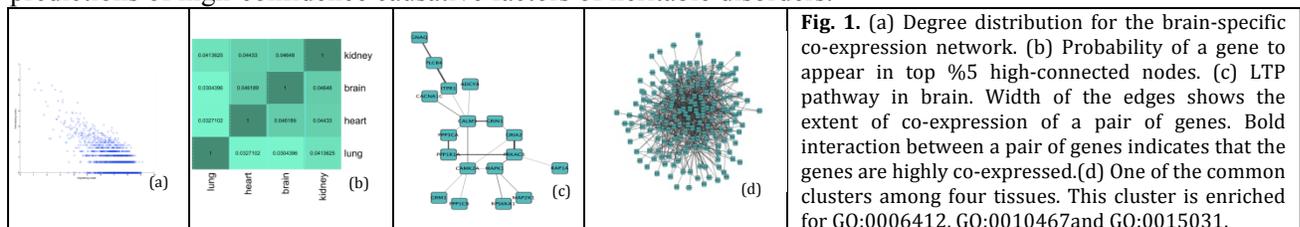
Understanding context-specific behavior of biological systems plays a fundamental role toward analyzing molecular mechanisms contributing to genetic diseases and the development of efficient therapeutic strategies [1]. The majority of genes show significant differences in expression levels in different tissues that reflect variations in biological pathways characteristic for a particular tissue or cell types. Thus, the reconstruction and analysis of the gene co-expression networks provides an approach to understanding tissue-specific mechanisms in health and disease. In this study, we reconstructed and analytically assessed tissue-specific co-expression networks in order to identify and characterize molecular pathways specific for a particular tissue (e.g. heart, brain, kidneys) or shared between a number of tissues. The reconstructed tissue-specific networks will be further used for the development of network-based disease models and subsequent predictions of causative genetic factors. To accomplish this goal we have reconstructed gene networks applying a well-established method for modeling the gene expression correlations as a multivariate Gaussian distribution with an L2 norm penalty.

$$\frac{n}{2} (\log(\det(S) - tr(S\Sigma)) - \lambda \|S\|^2)$$

Where n is the number of samples and S and Σ are the covariance matrix and precision matrix respectively. This method relies on gene expression data to infer tissue-specific networks. The global map of human gene expression (Lukk et al.[2]) for various tissues under normal (non-disease) condition was used for the reconstruction tissue-specific networks.

Tissue-specific networks were computed for four different tissues (heart, brain, lung, and kidney). The accuracy of the method for generating such networks was estimated as follows: (1) We observed a scale-free degree distribution for these tissue-specific networks (Fig. 1(a)) which is in line with previous findings [1]. (2) We determined the highest connecting hub genes in each network by calculating the probability of a gene to appear in the top 5% of each pair of networks' degree distributions. Here, different hubs in the tissue-specific networks suggest a discrepancy in the functions of the tissues (Fig. 1(b)). To find out what tissues are involved in a particular biological process, we analyzed the networks to identify tissue-specific pathways. For example, a long-term potentiation (LTP) of the signal transmission between two neurons, was found to have the most number of common interactions in brain rather than other tissues. The subgraph of LTP in brain is shown in Fig. 1(c).

The identification of sub-networks shared between different tissues-specific networks was performed by using MCODE algorithm [4]. The enrichment for characterization of the common clusters between the networks was performed using in-house Lynx system [3]. As expected, the subnetwork shared by all four tissues was enriched for the genes involved in central metabolism (e.g. carbohydrate, lipid and amino acid biosynthesis), information pathways and other housekeeping cellular functions (e.g. genes involved in the cell cycle, exo- and endocytosis, stress response, transmembrane transport). The described approach allows to identify and explore the molecular mechanisms unique or shared between the particular tissues. In future, we plan to use the reconstructed tissue-specific networks and sub-networks for the development of the high-resolution models for phenotypes of interest (e.g. diseases, developmental stages) and predictions of high-confidence causative factors of heritable disorders.



1. Watts, D.J. and S.H. Strogatz, Collective dynamics of "small-world" networks. nature, 1998. 393(6684): p. 440-442.
2. Lukk, M., et al., A global map of human gene expression. Nature biotechnology, 2010. 28(4): p. 322-324.
3. Sulakhe, D., et al., Lynx: a database and knowledge extraction engine for integrative medicine. Nucleic acids research, 2014. 42(D1): p. D1007-D1012.
4. Bader, Gary D., and Christopher WV Hogue. "An automated method for finding molecular complexes in large protein interaction networks." BMC bioinformatics 4.1 (2003): 2.s