Network analysis of differential expression for the identification of disease-causing genes

Daniela Nitsch

Genetic studies that aim to identify the causes of genetic disorders, such as linkage or association studies, identify chromosomal loci involved in a disease or a phenotype of interest. A frequent limitation is that those regions often contain tens or hundreds of genes, offering a serious limitation to further studies. Recently, methods for prioritizing candidate genes have been proposed, but they usually rely heavily on prior knowledge, which prevents them from tackling truly innovative discoveries.

A standard practice in genetics consists of checking variations in the expression level of a candidate gene in patient-derived EBV transformed lymphoblastoid cells or fibroblast cells against wild-type cells. Using a gene network, our approach extends this strategy by assessing the level of differential expression in the neighborhood of a candidate gene under the hypothesis that a strong candidate has many differentially expressed neighbors, meaning that it belongs to a disrupted expression module.

A network based on the STRING database (von Mering et al. 2007) was built, for which the distances between genes could be computed. Candidate genes located in highly differentially expressed neighborhoods are strong candidates. Different kernel matrices were used as distance measures, e.g. the Exponential Diffusion Kernel (Konder et al. 2002). To improve the enormous computing time, the kernel matrices were approximated by the Cholesky decomposition and the Reduced Eigenvalue decomposition.

We score candidates by weighting and aggregating the differential expression of neighbors as a function of distance. Through a randomization procedure, we rank candidates by their p-values.

We illustrate our approach on four diseases and successfully prioritize the known disease causing genes, with the key advantage of using unbiased expression data.

Reference: Nitsch D, Tranchevent LC, Thienpont B, Thorrez L, Van Esch H, Devriendt K, Moreau Y (2008). Network analysis of differential expression for the identification of disease-causing genes. In Review.