Identifying the functional downstream consequences of genetic variants that are associated with complex disease

Lude Franke
(University Medical Centre Groningen) September 11, 2009

For various complex diseases dozens of genetic variants have now been identified. An important question is what the functional consequences of these variants are. Through systematic interrogation of genetic variation and gene expression levels, for over 10,000 genetic variants effects on gene expression have now been identified (expression quantitative traits, eQTLs).

However, the effect sizes of these variants are usually small (likely comparable to variants that affect complex traits such as lipid levels, adult height and many diseases). Thus, in order to be adequately powered to detect these eQTLs, considerable amounts of samples are required and expression levels should be quantitated as accurately as possible.

To achieve sufficient statistical power, we conducted a meta-analysis on ¿10 'genetical genomics' datasets (comprising over 2,000 samples). To quantitate expression levels as accurately as possible we used a two step approach: We first applied principal component analysis to unrelated expression data (over 20,000 samples) from the Gene Expression Omnibus, resulting in the identification of 50 principal components that reflect environmental factors, some of which exert strong influences on global gene expression levels.

Subsequently, by superimposing these principal components on the various genetical genomics datasets, we could correct per individual sample or these factors, enabling considerably improved eQTL mapping. Over 15,000 unique variants (both SNPs and CNVs, imputed using TriTyper, Franke et al, AJHG 2008) were identified that affect gene expression levels (FDR 0.05).

For various genetic variants also effects on genes, mapping to different chromosome were detected (trans-eQTLs). In order to gain insight in these, subsequent biological interpretation was conducted by using a gene network (Franke et al, AJHG 2006, constituting over 80,000 known interactions derived from BIND, HPRD, IntAct, Reactome and KEGG). Results from this analysis indicate that the genes within the eQTL locus can usually be connected to the trans-gene within a fewer number of steps through the network than within permuted data.