Finding new genes for non-syndromic hearing loss through an in silico prioritization study

Accetturo M[1], Creanza TM[1], Giordano A[1] , Leo P[1], Santoro C[1], Scioscia G[1] , Tria G[1], Vaccina A[1]

Motivation

Prioritizing genes is a major concern for all those complex disorders whose genetic causes have not been yet completely understood. Due to its extremely heterogeneous genetics, non-syndromic Hereditary Hearing Loss (HHL) is one of the best candidates for such an approach: there are indeed 51 genes already known to be responsible, if mutated, of this phenotype, and 111 chromosomal regions linked to this disease over the years where one or more genes causing HHL are located. These regions are often large, containing hundreds of genes, making the systematic screening of all the genes they contain (candidate genes) in search of causative mutations not feasible. In this scenario a computational help to select the candidate genes according to their probability to cause, if mutated, the disease is strongly needed. To address this issue we built a gene scoring system based on Gene Ontology (GO), which scores the candidate genes for HHL by comparing them with the 51 HHL disease genes, relying on the rationale that genes whose dysfunction cause a disease, tend to be functionally related.

Methods

We defined a semantic similarity measure exploiting the information contained in GO to quantify the functional similarity between genes. Starting from Lin's metric (Lin 1998) which measures the semantic similarity between two GO terms through their Information Content (IC), i.e. a measure of how specific and informative a term is, we estimated the semantic similarity between two gene products looking at their GO annotations, measuring the information they share normalized by the information contained in their total descriptions. As GO allows multiple parents for each concept, two terms can share parents by multiple paths, therefore we chose for each GO term pair the more specific parent term shared by both of them. We defined a set of candidate genes for HHL as all the genes contained in the susceptibility loci known so far, and we prioritized them for the association with the

[1] IBM GBS BAO Advanced Analytics Services and MBLab, Via P. Leonida Laforgia, 14 - 70125 Bari. Italy

disease by measuring their similarity respect to the disease gene set. All candidate genes were ranked by computing the semantic similarity measure for each candidate-disease gene pair; then the final score used to prioritize each candidate was obtained as the mean of the scores estimated for that candidate against all the disease genes.

Results

The twenty top-scored genes were then examined to evaluate their possible involvement in HHL. We found that half of them are reported in literature to be expressed in human inner ear and/or cochlea, while six are reported to be expressed in other organisms' inner ear and/or cochlea, mainly mouse or chicken, and four have no gene expression data for these tissues. Due to the limited availability of gene expression data for human inner ear, these findings support the goodness of the ranking we produced in respect to the HHL. Moreover, looking at their functions, we found that most of the top-ranked genes play roles compatible with a possible involvement in HHL phenotype, such as a) processes of remodeling and organization of actin, an essential component of the hair-cell bundle; b) formation and maintenance of cilia, the sensory organelles devoted to receive the mechanical stimulus; c) K+ cycling and pH homeostasis in cochlear fluids, essentials for the generation and maintenance of the endocochlear potential; d) signal transduction, that again support our ranking. We also validated our results adding 15% of disease genes randomly drawn 1000 times from the disease gene set to the candidates, and testing if the number of disease genes ranked in the top 100, 75, 50 and 25 genes was significantly greater than expected when a random extraction of 100, 75, 50 and 25 genes was performed from the total set. In this case we always observed a p-value smaller than 0.05. We therefore are extremely confident that our metric was able to suggest excellent candidate genes for HHL to be screened in patients and controls for causative mutations.

Contact e-mail

pietro_leo@it.ibm.com