BMC Bioinformatics

POSTER PRESENTATION   **Open Access**

Toward PWAS: discovering pathways associated with human disorders

Emre Guney, Baldo Oliva[*]

Introduction

The past decade has witnessed dramatic advances in genome sequencing and a substantial shift in the number of genome wide association studies (GWAS). These efforts have expanded considerably our knowledge on the sequential variations in Human DNA and their consequences on the human biology. Nevertheless, complex genetic disorders often involve products of multiple genes acting cooperatively and pinpointing the decisive elements of such disease pathways remains a challenge.

Network biology recently proved its use in identifying candidate disease genes based on the simple observation that proteins translated by phenotypically related genes tend to interact, the so called guilt-by-association principle.

Methods

Here, we present GUILD (Genes Underlying Inheritance Linked Disorders), a network-based candidate disease-gene prioritization framework which reveals the path-
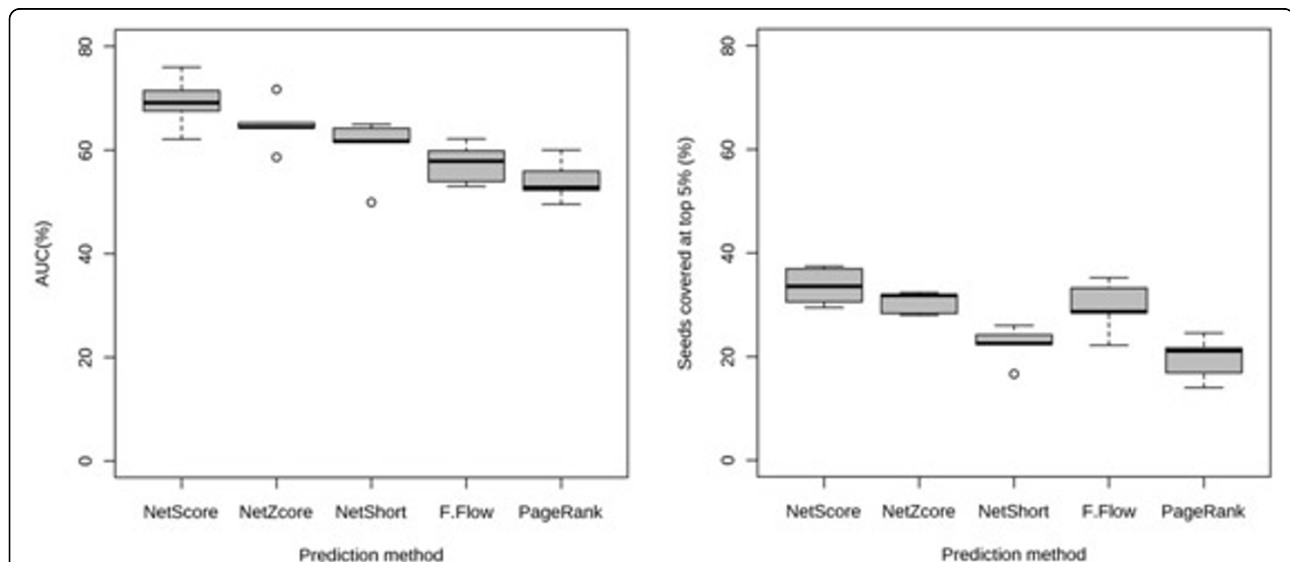

Figure 1 Comparison of prediction performance of proposed candidate disease-gene prioritization algorithms with existing methods. a) Area under ROC curve (AUC) yielded using 5 fold cross validation scheme over all disease-gene associations and different interaction sources used b) seed coverage among high scoring top 5% proteins.

Structural Bioinformatics Group (GRIB-IMIM), Universitat Pompeu Fabra,
Barcelona Research Park of Biomedicine (PRBB), 08003, Barcelona, Catalonia,
Spain

ways associated with the disorder (pathway wide association study, PWAS). We exploit several distinct plausible communication mechanisms of known genes associated with the phenotype emerging from the topology of the interaction network. We used three sources of gene-phenotypic association to specify nodes involved in a disorder (seeds for the methods proposed): Online Mendelian Inheritance in Man (OMIM) database [1] and two published data sets (by Goh et al. [2], and Chen et al. [3]).

Results and discussion

Analysis on multiple human disease phenotypes demonstrate that the methods proposed in GUILD surpass state-of-the-art prioritization methods such as PageRank with priors [3] and Functional-Flow [4] (Fig. 1). We tested the robustness of the approaches proving the effect of the network properties and the independence with the number of original genes/proteins associated with the function or phenotype. We applied the prioritization methods to study the implication of pathways in various diseases and highlight the relationship between AD and aging. Our findings confirm that most prioritization methods introduced in this study are able to distinguish between groups of connected genes with functions identical to those of the known disease-associated genes (disease pathways). In addition, using prioritization methods, we increased the coverage of genes known to play important roles in the interplay of AD and aging, most of which would not be otherwise identified by just inspecting the direct neighborhood in the network.

References

1. Hamosh A, Scott AF, Amberger JS, Bocchini CA, McKusick VA: **Online Mendelian Inheritance in Man (OMIM), a knowledgebase of human genes and genetic disorders.** *Nucleic Acids Res* 2005, **33**(Database): D514-517.
2. Goh KI, Cusick ME, Valle D, Childs B, Vidal M, Barabasi AL: **The human disease network.** *Proc Natl Acad Sci USA* 2007, **104**(21):8685-8690.
3. Chen J, Xu H, Aronow BJ, Jegga AG: **Improved human disease candidate gene prioritization using mouse phenotype.** *BMC Bioinformatics* 2007, **8**:392.
4. Nabieva E, Jim K, Agarwal A, Chazelle B, Singh M: **Whole-proteome prediction of protein function via graph-theoretic analysis of interaction maps.** *Bioinformatics* 2005, **21**(Suppl 1):i302-310.