Mining Brain-Related Transcription Factor-Disease Relationships for Novel Linkages

Warren A Cheung[*†‡]
wcheung@dnahelix.org

BF Francis Ouellette[†]
francis@bioinformatics.ca

Wyeth W Wasserman[‡]
wyeth@cmmt.ubc.ca

March 26, 2007

Abstract

Integrated approaches to the computational analysis of diverse data collections offer the possibility to predict links between genes and diseases. We focus on the analysis of biomedical literature for the identification of genes encoding DNA binding transcription factors which play a previously unknown functional role in the pathology of one or more neurological diseases. Existing databases enumerating human transcription factors, online repositories of abstracts from the biomedical literature and organized ontologies and vocabularies for both gene and disease annotation will be integrated. For example, over one thousand human genes in Entrez Gene are labeled as transcription factors via Gene Ontology (GO) terms. Using Medical Subject Heading (MeSH) terms, over half a million articles are identified as relevant to brain diseases in PubMed. To connect these data sources, we use both manually and automatically annotated linkages, such as the reviewed user-submitted Gene Reference into Function (GeneRIF) annotations in Entrez Gene and the computationally generated Related Articles from PubMed.

By distilling this interconnected network of relationships into an integrated database, we will provide a framework to identify known, direct relationships between transcription factors and brain diseases. This will also allow us to experiment with predicting novel relationships by the study of indirect linkages (e.g. transcription factor–characteristic and disease–characteristic intersections). Statistical scoring methodologies, such as over-representation analysis, will be developed to assess putative links between transcription factors and disease. Predictions will be verified using curated sources on gene-related diseases such as the Online Mendelian Inheritance in Man.

[*]Bioinformatics Program, University of British Columbia, Vancouver, BC, Canada

[†]UBC Bioinformatics Centre, Department of Medical Genetics, University of British Columbia

[‡]Centre for Molecular Medicine and Therapeutics, Child and Family Research Institute, Department of Medical Genetics, University of British Columbia