Prioritizing Disease Genes by Analysis of Common Elements (PDG-ACE)

Richard C. McEachin, PhD\textsuperscript{1,4}, Benjamin J. Keller, PhD\textsuperscript{2,4}, Peter P. Zandi, PhD\textsuperscript{3}, Mohsen Almani\textsuperscript{1,4}, Melvin G. McInnis, MD\textsuperscript{1,4}

\textsuperscript{1}Univ of Mich., Ann Arbor, MI; \textsuperscript{2}Eastern Mich. Univ., Ypsilanti, MI; \textsuperscript{3}The Johns Hopkins Univ., Baltimore, MD, \textsuperscript{4}National Center for Integrative Biomedical Informatics

Abstract

Complex diseases are characterized by multiple genetic and environmental influences on disease susceptibility. Since the multiple genetic influences converge on a single phenotype, some commonality may be evident among genes that influence the disease. We exploit this potential commonality among candidate disease genes to prioritize genes for further analysis and to pose novel, statistically significant, biologically plausible hypotheses on disease etiology.

Introduction

The relative risk of Bipolar Disorder for first-degree relatives of probands is approximately 3 to 7 times that of the general population, and for monozygotic twins the relative risk is approximately 15. These observations are consistent with genetic influences on bipolar susceptibility, and based on a number of genetic studies, the observed familial incidence is likely due to multiple genetic influences. We expect that if multiple genes predispose a single phenotype, they will show a commonality in annotation because they code for proteins that may function in a common pathway; have a common function, perhaps with common targets; are gene family members with related functions; show common tissue, cellular compartment, and/or temporal expression; have a regulator/regulatee relationship; or are binding partners. Equally, these genes may influence a common phenotype, not necessarily known to be related to disease, or have a common relationship with environmental factors. Combinations of these influences may result in multi-gene interactions.

We consider the problem of identifying relationships between genes that are likely to correspond to genetic interactions in complex disease. We start with a pair of groups of genes that may be identified through genetic linkage analysis, expert researchers, or SNP-SNP associations. In each case, candidate interactions must be probed for potential biological significance.

Method

Our approach PDG-ACE assumes genes influencing a common phenotype will exhibit common elements in annotation. The POCUS algorithm\textsuperscript{1} uses a similar approach based on Gene Ontology and InterPRO domain annotation. In the hope of finding more novel relationships, PDG-ACE instead uses keywords or phrases extracted from Entrez gene records for genes in the interacting groups. We score each keyword by the product of the number of genes from each group, and repeat this scoring for a large number of random group selections. We place the score for the hypothesized interaction into the distribution of random interaction scores, and calculate the \( p \)-value for the keyword as the proportion of scores at least as large as the hypothesized score. After correction for multiple hypothesis tests, the algorithm delivers a list of genes that show significant statistical evidence for interaction, and the biomedical keywords that may explain the interactions. Based on these genes and keywords, the researcher may identify novel hypotheses for follow-up.

Results

The expected interactions are found in documented gene/gene interactions used as positive controls. No significant interactions are seen in randomly selected gene pairs used as negative controls. Based on comparisons with MiMI\textsuperscript{2} and SAGA\textsuperscript{2} results, we find significant molecular interactions in Bipolar Disease etiology.

References


This work done with the support of the National Institutes of Health under #U54 DA021519