Pathway Thinking in Genome-wide Times

David Conti, PhD
Department of Preventive Medicine, Keck School of Medicine, USC

Abstract

The field of epidemiology has long been interested in the role that genes and the environment play in complex disease etiology. However, the analysis of multiple gene polymorphisms and environmental factors often reduces to determining the independent association of each polymorphism via contingency tables or regression models. Some researchers expand the analysis to include statistical interactions - often independent of the underlying biology or with an ad hoc idea of which interactions may be biologically important. These expanded analyses often encounter multiple comparison issues, as well as unstable and biased estimates due to sparse data. Here, we present extensions to hierarchical modeling to analyze multiple polymorphisms, exposures and their interactions. Hierarchical modeling with prior covariates aims to stabilize and inform estimation by incorporating similarities among genes and exposures using categories delaminating biologic relations. To narrow the space of possible regression models, we further structure the prior probability of including any variable as a function of known biology. This is accomplished via Bayes model averaging using stochastic variable selection. A major obstacle in this approach is the specification of prior knowledge and information. As an example, we discuss a computational representation or "ontology" of numerous sub-disciplines relevant to the complex, phenotypes of nicotine dependence and smoking cessation. We demonstrate the integration of this prior information in statistical analysis using data from the volunteer-based Northern California Twin

Registry. Using our ontology we specify prior covariates for all main effects and two-way interactions in relation to each outcome of interest and demonstrate how they can be incorporated into our hierarchical framework. We show that prior information aids in the stochastic model search and posterior inference. Finally, we discuss extensions to GWAS and how this modeling framework can be used to address systems-level hypotheses.

.