Poster Session
Complex Trait Consortium 2004
Johnson, A.D.

NOVEL METHOD FOR INFERRING GENOTYPE-PHENOTYPE ASSOCIATIONS TO IDENTIFY DISEASE SUSCEPTIBILITY LOCI

Johnson, A.D.[1], Habib, F.A.[2], Janies, D.A.[3]

[1]Integrated Biomedical Sciences Program, Pharmacogenomics Core, College of Medicine
[2]Department of Physics [3]Department of Biomedical Informatics, College of Medicine,
Ohio State University, Columbus, OH, USA

We present a novel method to identify genotype-phenotype associations through the construction and analysis of cladograms based on genetic polymorphisms and phenotypic variation. Phenotypic datasets were chosen for their relation to coronary artery disease (CAD) indicators and risk factors. Initially, trees were constructed using POY searches on a dataset containing 3,473 SNPs (MIT/Roche) and 17 quantitated phenotypes (Jackson Labs) for 15 inbred mouse strains. Tree visualization (Mesquite) allowed for identification of multiple parallel instances of changes in phenotype relative to ancestors and provided a basis to search for correlated genetic changes. Interrogation of intersections in sets of SNP characters that change in step with phenotype was conducted using a novel program (iVENN). This yielded 69 unique associations. These associations were further screened for biological relevance through annotations. A subset of these associations were found in, or near, genes with human homologs that have been studied or proposed for association with CAD. We next examined a 439,942 polymorphism dataset that included single and multiple nucleotide changes. Local database integration tools were developed with PERL to facilitate information extraction, screening and validation against candidate lists combined from multiple Programs for Genomic Applications (PGAs). This approach could be valuable in rapidly identifying new susceptibility loci and choosing high priority loci from lists of known candidates for further experimental investigation. The method is especially promising given 1) the availability of larger, highly validated and annotated polymorphism and phenotypic datasets, 2) the ability to apply these generalizable tools to any phenotype, and 3) the potential to automate much of the process.