

THE IDENTIFICATION OF FUNCTIONAL NONCODING HUMAN SNPS

Daniel C. Koboldt and Raymond D. Miller.

Department of Genetics, Washington University School of Medicine, St. Louis, MO

A major objective of human genetics is the identification of functional SNPs, i.e. those whose alleles differentially affect some phenotype including complex disease. The traditional candidates for functional SNPs have been those that cause coding changes in proteins. However, a variety of additional genomic annotations have become available, raising the question: Do SNPs in any of these features have functional importance? To investigate this question, we have cross-referenced a refined set of over 9 million SNPs from public databases to the coordinates of several types of annotated genomic features (results in SNPseek database available at <http://snp.wustl.edu>). Using frequency data from the HapMap Project and extensions of methods developed by Fay *et al.* (Genetics: 158, 1227), we have tested each annotation category by examining the allele frequency distributions and population-specificity of SNPs. The underlying idea is that beyond genetic drift, allele frequencies could only have been changed by natural selection, and natural selection is blind to all but functional SNPs. We found that SNPs in many of these annotated features exhibited an excess of rare variation, as expected due to the action of weak purifying selection. We found that SNPs in annotated regions for transcription factor binding sites, exon splicing enhancers, conserved regions, splice sites, and 3' UTR microRNA binding sites are important classes of functional variation in the human genome. In fact, tests of SNPs in the latter two categories had more dramatic results than those of SNPs that cause coding changes in proteins.