The Extent of Genetic Variation in Human Genes

Eskender McCoy, Eran Elhaik, and Aravinda Chakravarti
McKusick-Nathans Institute of Genetic Medicine, Johns Hopkins University School of Medicine, Baltimore, MD

Introduction

The publication of the draft Human Genome in 2001 initiated an era of better understanding of the role of genes and their variants in complex diseases. In the following years, sequencing technology for finding genomic variants, such as single nucleotide polymorphisms (SNPs), has greatly advanced. Sequencing technology has enabled us to create a comprehensive catalog of coding SNPs (cSNPs) that may affect the final protein product. Where GWAS (Genome Wide Association Study) studies have genotyped random SNPs in patients and tested for association with a certain disease, limiting the study to cSNPs focuses the search to specific genes.

Method

The cSNPs catalogue, was obtained from the NCBI website (http://www.ncbi.nlm.nih.gov/variation/SNP) and processed by the Galaxy browser (http://main.g2.bx.psu.edu/). We used NCBI tables: Snp130CodingDbSnp, SnpFunctionCode, and SNPAlleleFreq.

Terms used

Genetic Variation: The nucleotide diversity between humans is about 0.1% (Halushka et al, 1999, Nature, 22, 239-247). SNPs along our genome are responsible for most of the variation among two individuals. Although most gene variants are neutral, some may cause functional changes in proteins resulting in phenotypic differences between Individuals.

HapMap: The international HapMap project attempts to create a haplotype map of the human genome. By mapping all SNPs scientists hope to describe the common patterns of human variation.

Hypertension: High blood pressure (HBP) or hypertension means high pressure (tension) in the arteries: arteries being the vessels that carry blood from the heart to all the tissues and organs of the body.

Data Analysis

Table 1: Expected (EXP) number of cSNPs, was calculated using the codon table and assuming all possible single nucleotide changes for each codon.

Identifying cSNPs associated with hypertension

Figure 2a-b: QQ-plots for the p-values of hypertension SNPs. SNPs with the p-values that deviate from a straight are considered to be associated with hypertension. Genes harboring the cSNPs associated with those SNPs are marked.

Conclusions

We demonstrated the usefulness of our computational methods to find coding SNPs and genes related with a complex disorder by analyzing GWAS data from a hypertension study. We identified several cSNPs related to hypertension and candidate genes. The exact role of these genes in hypertension should be further studied.

Acknowledgments

I would like to use this space to thank Vick Schneider and everyone else who runs the Center Scholars Program for giving me the another chance to work in a lab this year.