Title: How bad is the human genome or what can we learn from sequencing many humans?

Abstract:

Systematic human resequencing datasets including a new dataset of 54 protein coding genes sequenced in over 1,500 individual human chromosomes shed light on the fate of mutations in the human population. We combined analysis of the deep re-sequencing datasets with human-chimpanzee divergence and mutations causing human Mendelian diseases and found that about 20% of new missense mutations in humans result in a loss of function, while about 25% are effectively neutral. Thus, more than half of new missense mutations have mildly deleterious effects. These mutations give rise to many low frequency deleterious allelic variants in the human population. Surprisingly, up to 73% of low frequency missense alleles are mildly deleterious and associated with a heterozygous fitness loss in the range $10^{-2.5}$– 10^{-3}. Thus, the low allele frequency of an amino acid variant can by itself serve as a predictor of its functional significance. Several recent studies have reported a significant excess of rare missense variants in disease populations compared to controls in candidate genes or pathways. These studies would be unlikely to work if most rare variants were neutral or if rare variants were not a significant contributor to the genetic component of phenotypic inheritance. Our results provide a justification for these types of candidate gene (pathway) association studies and imply that mutation-selection balance may be a feasible mechanism for evolution of some common diseases. An example of a recent study of human obesity will be discussed.

Important genetic variation is not limited to protein coding regions of the genome. It is widely assumed that human non-coding sequences comprise a substantial reservoir for functional variants impacting gene regulation and other chromosomal processes. Evolutionarily conserved non-coding sequences (CNSs) in the human genome have attracted considerable attention for their potential to simplify the search for functional elements and phenotypically important human alleles. A major outstanding question is whether functionally significant human non-coding variation is concentrated in CNSs or distributed more broadly across the genome. We combined whole-genome sequence data from four non-human species (chimp, dog, mouse, and rat) with comprehensive human resequencing data to analyze selection at single nucleotide resolution. We showed that a substantial fraction of active selection in non-coding sequences occurs outside of CNSs and is diffusely distributed across the genome. This suggests the existence of a large complement of human non-coding variants that may impact gene expression and phenotypic traits, the majority of which will escape detection using current approaches to genome analysis.