

# Largest human exomes data reveals an excess of low frequency non-synonymous coding variants

October 5 2010

---

In a paper appearing in *Nature Genetics* today, an international research group reported the resequencing and analysis of 200 human exomes, established the largest data set for human exomes published so far and reveal an excess of low frequency deleterious non-synonymous genetic mutations. The collaborative team includes investigators from BGI-Shenzhen, UC Berkeley, University of Copenhagen and some other European institutions.

The team used NimbleGen 2.1M exon capture array to targeted capture 18,654 coding genes of human [genome](#) and sequenced 200 individuals from Denmark. The average sequencing depth for each exome is 12X coverage and about 95% of targeted regions were covered by at least 1 read. In total, 121,870 SNPs were identified in the population, about 44% were novel SNPs. 53,081 coding SNPs (cSNPs), 25,275 synonymous and 27,806 non-synonymous, were identified, of which 42.6% were novel.

Based on the large population data, statistical analysis was performed for SNP calling and calculate distribution of allele frequencies. The allele frequency spectrum of cSNPs with a minor allele frequency  $> 2\%$  was developed to exclude false positive SNPs. By comparing the distribution of allele frequencies among non-synonymous and synonymous cSNPs, a 1.8 fold excess of deleterious, non-synonymous over synonymous cSNPs was identified in the low allele frequency range between 2-5%.

Moreover, this excess was higher for [X chromosome](#) SNPs, suggesting that deleterious mutations on the X chromosome are primarily recessive. The team further analyzed the potential effects of methylation over allele frequencies by comparing the frequency distribution for sites potentially affected by CpG methylation or [gene conversion](#) with unaffected sites, where no strong effect was detected at a genome-wide scale.

The study provides an valuable data set for studying the allele frequency specturm and population genetic patterns, said Dr Yingrui Li, the project investigator from BGI-Shenzhen. We found more low frequency deleterious mutations in coding regions than previously expected, and most of them are recessive, thus we support the idea that much of the heritable variation affecting fitness is caused by low frequency mutations.

Association studies have only detect limited heritable variation associated with common polygenic traits and genotyping analysis generally overlooks the effects of low frequency mutations. The results obtained in this study further demonstrate that exome sequencing is an effective and promising approach to identify genetic variants associated with human traits and study population genetics. The team expects that Future analyses of non-coding regions and ethnically diverse samples will help build a complete picture of human genomic variation and an understanding of the interaction between genetic drift, mutation, recombination, and selection in the human genome.

Previouly, a paper in *Science* (*Science*. 2010 July; 329(5987): 75-78) reported sequencing the exomes of 50 Tibetan individuals and found evidence for high altitude adapdation of Tibetan populations. It shows that next generaton sequencing is getting more applications and will have great potential in genomics research, drug discovery and personalized medical treatment.

Source: BGI Shenzhen

Citation: Largest human exomes data reveals an excess of low frequency non-synonymous coding variants (2010, October 5) retrieved 27 April 2024 from <https://medicalxpress.com/news/2010-10-largest-human-exomes-reveals-excess.html>

This document is subject to copyright. Apart from any fair dealing for the purpose of private study or research, no part may be reproduced without the written permission. The content is provided for information purposes only.