

Evolutionary information improves discovery of mutations associated with diseases

July 18 2012, By Richard Harth



Sudhir Kumar directs the Center for Evolutionary Medicine and Informatics at the Biodesign Institute. He is also Professor in the College of Liberal Arts and Sciences, School of Life Sciences. Credit: The Biodesign Institute at Arizona State University

A book containing misprints may cause annoyance for the reader, but typos in an individual's genetic blueprint (DNA) can mean serious disease or even death. The search for genetic correlates for the wide range of diseases plaguing humankind has inspired a wealth of research falling under the heading of genome-wide association studies (GWAS).

According to Sudhir Kumar, director of the Center for Evolutionary Medicine and Informatics at Arizona State University's Biodesign Institute, however, results from many such studies become less useful when gene variants or alleles implicated in disease in a given population fail to be discovered in subsequent independent studies. "Often, we do not discover the same set of mutations for the same disease in different populations," he says. "This is a huge problem in genomic medicine."

Kumar and colleagues Joel T. Dudley, Rong Chen, Maxwell Sanderford, and Atul J. Butte, have developed a statistical method to remedy this problem by using evolutionary information. It is capable of significantly enhancing the likelihood of identifying disease-associated alleles that show better consistency across populations, improving the reliability of GWAS studies. The method makes use of phylogenetics – the comparative study of species genomes through long-term evolutionary history.

The group's research appeared in the advanced online issue of the journal *Molecular Biology and Evolution*. The new method is now available to use via the web, so that researchers worldwide can apply it as an aid to discovering disease-associated mutations that are more consistently reproducible and therefore useable as diagnostic markers. Kumar refers to this new approach, combining standard comparative genomic studies with phylogenetic data as *phylomedicine*, a rapidly developing field that promises to streamline genomic information and improve its diagnostic power.

"We can take this method and apply it to all the data that has been published," Kumar says. "It will lead to new discoveries that were sitting right there, but nobody knew about."

The new method boosts the discovery of reproducible mutations by integrating evolutionary history of humans with contemporary genomic

information. Applying the new rankings to a large GWAS study improved the discovery of reliable mutation correlates of complex diseases, which will advance personalized medicine based on each patient's genomic code.

The basic idea behind GWAS is simple: compare the genomes of two populations of subjects, one with disease trait and a control group without the disease. Next, identify the disparities at each position of the genome in the two populations. Find the alleles occurring in the diseased population that are less frequent in the healthy population and you have just pinpointed the gene mutations associated with disease.

Or have you?

As Kumar explains, matters are not so simple. The mutations studied in such studies are known as SNPs (for single-nucleotide polymorphisms). This simply means that for a given gene sequence, one of the four nucleotides (A,T,C and G) found commonly in the population is replaced by something else. For example, the majority of healthy subjects may carry the 'A' at a particular position in the genome, but disease individuals may be more likely to carry a 'C' at the same position. If the difference between the groups is striking, the SNP may be associated with the disease trait.

Human genomes are vast structures – consisting of some 3 billion base pairs of nucleotides. Most are littered with SNPs and teasing out which ones sit there without apparent effect and which may translate to disease is often a vexing affair. For some diseases, a one-to-one correspondence between gene mutation and disease exists. Generally, these afflictions, known as monogenic diseases, have particular characteristics. They result from a mutation in just a single gene, rather than multiple genes. They are early-onset diseases, taking their toll when the patient is still young.

Monogenic diseases, which include cystic fibrosis, Tay Sachs disease, sickle cell anemia and Huntington's disease are usually not the targets for genome-wide association studies, because the relationship between gene mutation and occurrence of the disease is straightforward and reliable.

By contrast, so-called complex diseases tend to occur later in life, are triggered by mutations occurring at multiple sites along the genome and often have a significant environmental— that is, non-genetic — component. Finding the alleles responsible for such diseases, which include hypertension, rheumatoid arthritis, Alzheimer's disease, type II diabetes and countless others through GWAS studies has often been a bewildering endeavor, as alleles identified in one study population frequently fail to turn up in different studies with different populations.

GWAS studies compute the odds of an allele along the genome being disease-related and translate this into a statistic known as the P value. Alleles with the lowest P value are least likely to have occurred by random chance. In the current research, a meta-analysis is conducted using results from thousands of previous GWAS studies and phylogenetics are applied to unearth evolutionary trends in the data.

“Every position in the human genome among the billions of base pairs has evolved over time,” Kumar says. “As the genome evolves, some positions permit change frequently while others do not.” The positions least likely to change with time and across mammalian species are known as evolutionarily conserved positions. The group conducted a multispecies genomic analysis of 5,831 putative human risk variants for more than 230 disease phenotypes reported in 2,021 studies. “Even if a GWAS variant does not have a functional role in a disease, evolutionary information is still very relevant, because every position in the human genome has an evolutionary signature that gives us prior information on how alleles at that position are likely to vary in modern human populations,” says Dudley, the study's lead author.

An analysis of existing data found that most of the presumptively disease-related alleles uncovered in the GWAS studies occurred at relatively slow-evolving, highly conserved sites. According to Kumar, this fact accounts for the poor reproducibility of many putative [disease](#) alleles across different populations, as alleles occurring at conserved sites tend to be rare. As Kumar explains “You can keep finding rare alleles like this all day, but they would have limited clinical utility in a broader population.”

The new ranking system, known as E-ranking, incorporates phylogenetic information from multi-species studies of mammals, and applies it to human GWAS data. The effect is to remove the inherent sampling bias for rare alleles, allowing the more common alleles occurring at fast-evolving sites in the genome to be more readily discovered. “Our method removes this bias, which gives a boost to high-frequency common variants that are more likely to reproduce across populations due to the evolutionary history of the genomic position where they are found,” says Dudley.

Provided by Arizona State University

Citation: Evolutionary information improves discovery of mutations associated with diseases (2012, July 18) retrieved 25 April 2024 from <https://medicalxpress.com/news/2012-07-evolutionary-discovery-mutations-diseases.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.