

New genomics study shows ancestry could help solve disease riddles

October 25 2012

October 25, 2012 – Explosive advancement in human genome sequencing opens new possibilities for identifying the genetic roots of certain diseases and finding cures. However, so many variations among individual genomes exist that identifying mutations responsible for a specific disease has in many cases proven an insurmountable challenge. But now a new study by scientists at The Scripps Research Institute (TSRI), Scripps Health, and Scripps Translational Science Institute (STSI) reveals that by comparing the genomes of diseased patients with the genomes of people with sufficiently similar ancestries could dramatically simplify searches for harmful mutations, opening new treatment possibilities.

The work, reported recently in the journal *Frontiers in Genetics: Applied Genetic Epidemiology*, should speed the search for the causes of many diseases and provide critical guidance to the genomics field for maximizing the potential benefits of growing genome databases.

Much work is already under way to sequence the DNA of people suffering from diseases with unknown causes, called idiopathic conditions, to find the roots of their problems. Unlike more complex conditions such as diabetes, in some cases a limited number of genetic defects, or even a single mutation, can cause an idiopathic disease. Identifying those critical mutations can lead to effective treatments for previously mysterious problems.

Complicated Searches

While there have been some successes, in many other instances the genetic basis of an idiopathic disease remains elusive. Among other groups, The National Human Genome Research Institute runs searches for idiopathic disease sufferers and is able to find offending [gene sequences](#) only about 30 percent of the time. "One explanation for that other 70 percent might be that the diseases are enormously complex," said the new study's senior author Nicholas Schork, a professor at TSRI, director of research for Scripps Health's [genomic medicine](#) program, and director of biostatistics and bioinformatics at STSI, "but it could be that they're still searching in the noise."

The new work offers a likely filter for much of that noise. The results show that comparing a person's DNA sequence against existing genomes for those whose ancestry is not sufficiently similar, as is typically the case, can cause serious problems. Countless differences that seem unique to a patient might instead be DNA variants carried by everyone with the same ancestry. A researcher might, for instance, identify hundreds of variants and not be able to zero in on the one responsible for a disease.

But the new results show that comparing closer ancestry matches will dramatically reduce the number of variants identified as potentially responsible for a disease, reducing a search to a workable number.

For the work, the team developed a tool called the Scripps Genome Adviser. This processing framework uses a supercomputer to incorporate a variety of databases and algorithms to identify DNA variants in a particular genome relative to reference genomes. It then uses algorithms to analyze these variants and predict whether they have any physiological effects, and if so what those might be.

The team began with nearly 60 whole human genome databases and ran

three key types of computing experiments. First the researchers identified the number of variants in the reference human genomes and found that on average each has millions of variants, about 12,000 of which have functional effects. Then the scientists looked at the rates at which variants appeared in various ancestry lines.

Honoring Ancestry

Importantly, the scientists didn't stop there. They deliberately inserted a mutation known to cause disease into a genome, then ran this genome through the Adviser to see how effectively it could identify that known variant as unique.

When the team ran the searches comparing that altered genome against a reference panel of genomes that included different ancestries, the known variant remained effectively lost in a sea of other variants. But comparison against genomes of similar ancestry dramatically reduced the number of variants identified, allowing identification of the inserted disease-causing gene.

A study published simultaneously with the Scripps team's paper by Professor Carlos Bustamante and colleagues from Stanford University also pointed to ancestry's importance, but this is the first time a team has been able to look at the problem on the whole-genome scale. "Others have indeed recognized ancestry as important," said Schork, "but no one had shown how much it could haunt a particular study, especially on a whole genome basis."

As importantly, prior to this study it wasn't clear how to address the ancestry issue. But the new study provides clear direction. The team calculated that identification of the vast majority of ancestral variants can be performed successfully with a reference panel of less than 20 genomes—though it could well take more to identify a particular

ancestry group's rarest deviations. Of course, most people have more than one ancestry line, meaning that in practice a patient's reference panel would need to include multiple reference groups.

This result should act as a guide for continuing genomics work. Many ancestries are already well represented, meaning that assembling an effective reference panel is possible in some cases. But the number of whole genomes from a particular ancestral group isn't the only consideration. Ideally, reference genomes need to be from relatively disease-free people, meaning subjects who lived to an old age without major complications from genetic conditions.

Recognizing the importance of ancestral comparisons, researchers and companies can now deliberately work to fill any holes. "Building those sorts of resources could only benefit the community," said Schork. In fact, Schork, Ali Torkamani and others at Scripps are collaborating with Complete Genomics, Inc., a whole genome sequencing company in Mountain View, CA, to develop appropriate reference panels for clinicians and researchers.

Deciphering Diseases

Schork and his colleagues are already working toward broader application of their results using an increasingly advanced version of the tool. While processing a single person's genome to identify and analyze variants took about four days when the project began, today the Adviser can accomplish the task in about 30 minutes.

Along with the paper's lead author Torkamani, Schork is a founder of a company called Cypher Genomics that has licensed the Scripps Genome Adviser for disease-focused research. The teams in both industry and academia hope not only to continue idiopathic disease research, but also to apply similar principles to search for the causes of more complex

congenital conditions. "The broader message of our work is that you have to take ancestry into account no matter what disease you're studying," said Schork.

More information: *Frontiers in Genetics: Applied Genetic Epidemiology*
[doi: 10.3389/fgene.2012.00211](https://doi.org/10.3389/fgene.2012.00211)

Provided by Scripps Research Institute

Citation: New genomics study shows ancestry could help solve disease riddles (2012, October 25)
retrieved 25 April 2024 from
<https://medicalxpress.com/news/2012-10-genomics-ancestry-disease-riddles.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.