

Scientists create world's largest catalog of human genomic variation (Update)

September 30 2015



The 1000 Genomes Project tracked genetic variation across the globe. Credit: EMBL-EBI/Spencer Philipps

An international team of scientists from the 1000 Genomes Project Consortium has created the world's largest catalog of genomic



differences among humans, providing researchers with powerful clues to help them establish why some people are susceptible to various diseases. While most differences in peoples' genomes - called variants - are harmless, some are beneficial, while others contribute to diseases and conditions, ranging from cognitive disabilities to susceptibilities to cancer, obesity, diabetes, heart disease and other disorders. Understanding how genomic variants contribute to disease may help clinicians develop improved diagnostics and treatments, in addition to new methods of prevention.

The National Human Genome Research Institute (NHGRI), part of the National Institutes of Health, helped fund and direct this international public-private consortium of researchers in the United States, the United Kingdom, China, Germany and Canada.

In two studies published online on Sept. 30, 2015, in *Nature*, investigators examined the genomes of 2,504 people from 26 populations across Africa, East and South Asia, Europe and the Americas.

In the main *Nature* study, investigators identified about 88 million sites in the human genome that vary among people, establishing a database available to researchers as a standard reference for how the genomic make-up of people varies in populations and around the world. The catalog more than doubles the number of known variant sites in the human genome, and can now be used in a wide range of studies of human biology and medicine, providing the basis for a new understanding of how inherited differences in DNA can contribute to disease risk and drug response.

Of the more than 88 million variable sites identified, about 12 million had common variants that were likely shared by many of the populations. The study showed that the greatest genomic diversity is in



African populations, consistent with evidence that humans originated in Africa and that migrations from Africa established other populations around the world.

The 26 populations studied included groups such as the Esan in Nigeria; Colombians in Medellin, Colombia; Iberian populations in Spain; Han Chinese in Beijing; and Sri Lankan Tamil in the United Kingdom. All of the individuals studied for the project consented to broad release of their data, and the data can be used by researchers around the world.

"The 1000 Genomes Project was an ambitious, historically significant effort that has produced a valuable resource about human genomic variation," said Eric Green, M.D., Ph.D., director of NHGRI. "The latest data and insights add to a growing understanding of the patterns of variation in individuals' genomes, and provide a foundation for gaining greater insights into the genomics of human disease."





Scientists of the 1000 Genomes Project Consortium has created the world's largest catalog of genomic differences among humans. Credit: Ernesto Del Aguila III

These reports mark the culmination of the 1000 Genomes Project, which found more than 99 percent of variants in the human genome that occur at a frequency of at least 1 percent in the populations studied.

"Some 88 million sites in the genome differ among people. About onequarter of these variants are common and occur in many or all populations, while about three-quarters occur in only 1 percent of people or are even more rare," said Lisa Brooks, Ph.D., program director in the NHGRI Genomic Variation Program. "The 1000 Genomes Project data are a resource for any study in which scientists are looking for genomic



contributions to disease, including the study of both common and rare variants."

One of the more immediate uses of 1000 Genomes Project data is for genome-wide association studies (GWAS), which compare the genomes of people with and without a disease to search for regions of the genome that contain genomic variants associated with that disease. Such studies generally find several genomic regions associated with a disease and many variants in each of those regions. Scientists can now combine GWAS data with the more detailed 1000 Genomes Project data to home in on regions affecting disease more precisely. Instead of sequencing the genomes of all the people in a study, which remains expensive, researchers can use the 1000 Genomes Project data to find most of the variants in those regions that are associated with the disease.

"When the 1000 Genomes Project was first launched in 2008, there wasn't much understanding of how rare genomic variants were distributed among populations around the world and their relationship to other variants," said Adam Auton, Ph.D., the main study senior author and principal investigator who until recently was assistant professor of genetics at the Albert Einstein College of Medicine in New York City.

"The project has been an effort to build a reference dataset of genomic variation. It really tells us about the structure of human genomic variation and diversity," said Dr. Auton, who is now with the company 23andMe in Mountain View, California.

In a companion paper in *Nature*, scientists examined differences in the structure of the genome in the 2,504 samples. They found nearly 69,000 differences, known as structural variants. These genomic differences, many of which affected genes, include deletions (loss of DNA), insertions (added DNA), and duplications (extra DNA copies). The researchers created a map of eight classes of structural variants that



potentially contribute to disease.

"Structural variation is responsible for a large percentage of differences in the DNA among human genomes," said Jan Korbel, Ph.D., group leader and European Research Council Investigator in the Genome Biology Unit of the European Molecular Biology Laboratory in Heidelberg, Germany. Dr. Korbel is senior author and a co-principal investigator for the structural variation study. "No study has ever looked at genomic structural variation with this kind of broad representation of populations around the world."

Dr. Korbel, co-principal investigator Evan Eichler, Ph.D., professor of genome sciences and a Howard Hughes Medical Institute investigator at the University of Washington in Seattle, and their colleagues discovered that structural variants were often more complicated than they originally thought. For example, the majority of inversions, which involve DNA sequences changing their orientation in the genome, frequently occur along with other structural changes.

To Gonçalo Abecasis, Ph.D., chair of biostatistics at the University of Michigan in Ann Arbor and co-principal investigator for the main *Nature* study, the value of the 1000 Genomes Project extends far beyond the data. Advances in DNA sequencing and bioinformatics were vital to completing the project.

"We've learned a great deal about how to do genomics on a large scale," said Dr. Abecasis. "Over the course of the 1000 Genomes Project, we developed new, improved methods for large-scale DNA sequencing, analysis and interpretation of genomic information, in addition to how to store this much data. We learned how to do quality genomic studies in different contexts and parts of the world."

"The 1000 Genomes Project has laid the foundation for others to answer



really interesting questions," said Dr. Auton. "Everyone now wants to know what these variants tell us about human disease."

More information: *Nature*, DOI: 10.1038/nature15394 *Nature*, DOI: 10.1038/nature15393

Provided by NIH/National Human Genome Research Institute

Citation: Scientists create world's largest catalog of human genomic variation (Update) (2015, September 30) retrieved 5 May 2024 from <u>https://medicalxpress.com/news/2015-09-links-genes-large-scale-genetic-world.html</u>

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.