

1000 Genomes Project reveals human variation

An international effort to map variability in the genome hits its first landmark.

Alla Katsnelson

The long-awaited results from the pilot phase of the first large-scale initiative to sequence individual genomes have identified 95% of the variation found across the human genome and revealed some 15 million gene variants, more than half of which had never been observed before. The data represent the most thorough effort so far to understand the depth of genetic differences between individuals and populations, but the results also highlight the fact that there is still an enormous amount left to learn.

The **1000 Genomes Project**, a consortium of researchers from more than 75 universities and companies around the world, two years ago embarked on a mission to catalogue genetic variants — small inter-individual differences in specific regions of the genome — that are found in all human populations. Such differences are quite common, the results of the survey revealed, with each person's genome carrying some 250 or 300 so-called 'loss-of-function' mutations that incapacitate the gene in which they occur.



A project to sequence hundreds of human genomes has found millions of gene variants.

L. WILLATT, EAST ANGLIAN
REGIONAL GENETICS SERVICE /
SCIENCE PHOTO LIBRARY

"That's quite a lot — it's on the order of 1% of all genes," says Richard Durbin, a genomicist at the Wellcome Trust Sanger Institute in Hinxton, UK, and one of the chief architects of the project.

Data goldmine

This first leg of the project consisted of three components: sequencing the complete genomes of 179 individuals from West Africa, Europe, China and Japan at a fairly low level of accuracy; sequencing the complete genomes of two sets of trios (a child and both its parents) at a high level of accuracy; and sequencing the 'exome' regions — the segments of the genome that contain protein-coding genes — for an additional 697 individuals.

The paper describing the analysis of the sequencing data, published in *Nature* today¹, may not be a great surprise to geneticists as much of it has already been disseminated in the community, says David Goldstein, a geneticist at Duke University in Durham, North Carolina, who is not involved with the consortium.

But its effects will be long-lasting and instructive, he says. In addition to producing an enormous quantity of raw data, analyzing them has compelled researchers to develop tools to work with this type of data. Goldstein calls it a "treasure-trove of goodies for how to do whole-genome sequencing".

In its next phase, the project will expand its sequencing efforts further — to 2,500 individuals.

That catalogue will give researchers a baseline with which to compare mutations they identify in patients. Also, notes Durbin, comparing the variation in the genomes of different populations will provide a better

picture of how evolutionary forces act on specific genes to create different traits.

Rarer variants

A companion paper, published today in *Science*², describes a new method for probing variation in regions of the genome that contain multiple copies of genes. This type of 'structural' variation isn't picked up by traditional techniques and has been extremely difficult to track, says Evan Eichler of the University of Washington in Seattle, the lead author of the paper and chair of the 1000 Genomes Project's working group on structural variation.

On comparing stretches of DNA containing multiple copies of genes in humans and great apes, the group identified several copy-number differences between species in genes related to brain development. "Individually they are all a bunch of just-so stories," says Eichler, "but to me, in the aggregate, there is a signal that suggests these particular genes and gene families have been enriched" in humans compared with great apes – findings that highlight the bounty of genomic information still left to mine.

Technologically, too, the *Science* study is "hugely encouraging," says Goldstein. "It's a massive leap forward in our ability to accurately represent this kind of variation."

Genome sequencing is not technologically straightforward, and current techniques are prone to errors that must be corrected or accounted for at the analysis stage. One of the primary benefits of the 1000 Genomes Project is that it has served as a collaborative research laboratory, in which researchers could hammer out and agree on a set of best practices, geneticist David Altshuler, director of medical and population genetics at the Broad Institute in Cambridge, Massachusetts, and a member of the consortium, said in a press briefing.

For the past several years, the most comprehensive way of studying the genetics of disease was the genome-wide association study, which looked for mutations in distinct regions of the genome that were known to vary in about 5% of the population and tried to link them to specific diseases. But those studies have come up short, and many researchers have proposed that rarer variants might play a more important role.

"What's exciting is that we are now going to have the tools and methods to just answer that question with empirical data rather than speculation," Altshuler said.

References

1. The 1000 Genomes Project Consortium *Nature* 467, 1061-1073 (2010). | [Article](#) | [OpenURL](#)
2. Sudmant, P. H. et al. *Science* 330, 641-646 (2010). | [ChemPort](#) |