

1,092 genomes and counting

November 16 2012, by Veronica Meade-Kelly



The 1000 Genomes Project Consortium set out to create a complete catalog of all common genetic variation across human populations. The consortium has sequenced 1,092 human genomes from 14 different populations thus far. Credit: Lauren Solomon, Broad Communications

Focusing on fine features in order to see "the big picture" seems almost counterintuitive, but that is exactly what is happening in the field of genomics. Researchers are sequencing human genomes, cataloging the variation in people's genetic code – the As, Ts, Gs, and Cs of human DNA that serve as each individual's biological blueprint – to get a

broader view of human health, a deeper knowledge of human genetic history, and a clearer understanding of why some people develop certain diseases while others do not.

The 1000 Genomes Project is contributing significantly to that "big genetic picture." Launched in 2008, the consortium, which includes Broad Institute scientists and crosses several scientific disciplines, institutions, and continents, set out to create a complete catalog of all [common genetic variation](#) across human populations. The research team also set out to make that catalog open and accessible to researchers worldwide. This month, the consortium published its second paper in *Nature*.

The story of the consortium's progress could be told in numbers: the original goal of the project was to sequence 1000 human genomes, and the team has exceeded that mark. This week's paper reports on the 1,092 genomes from 14 different [human populations](#) that the team has sequenced so far, and they are now poised to complete a total of 2,500 genomes from 26 populations by the end of the project.

From those 1,092 samples, the 1000 Genomes Project Consortium has increased manifold the catalog of [human genetic variation](#) that researchers rely upon to test the genetic differences that exist among people and populations. Their findings include 38,000,000 [single nucleotide polymorphisms](#), or SNPs (pronounced "snips"), which are single-letter variations in the A, T, G, and C bases that form DNA. SNPs are the most common type of genetic variation among people, and they suggest to researchers where in the [genome](#) the [genetic code](#) differs between people. The paper's authors estimate that they have now identified about 98% of all [SNPs](#) that occur with a frequency of 1% or higher (that is, those variants that occur in at least 1 in 100 people).

The team, led in part by Broad researchers Steve McCarroll, Bob

Handsaker, and Charles Lee, also identified structural variants in the sequenced DNA. These included 1.4 million indels (small insertions or deletions of anywhere from 1 to 50 bases in the DNA code), and over 14,000 larger deletions. These structural variants are relatively new to the catalog, having been difficult to detect in previous genome-sequencing projects.

The numbers do not tell the whole story of the project's impact, however. Broad core member and 1000 Genomes Project co-chair David Altshuler, and fellow author Mark DePristo, who serves as co-director of the Broad's Program in Medical and Population Genetics, say that the project's biggest contributions stem from its roots as a "community resource project." They say that the methods and technology the project has fostered and developed have been incorporated with accessibility in mind and they expect those innovations to facilitate genetic research in the coming years.

From the start, the consortium made a concerted effort to ensure that all of the 2,500 samples collected for the project were obtained along with the proper informed consent, allowing the sequenced data to be included in an open, online database. As a result, all of the genomic data produced by the 1000 Genomes Project is publicly available for use by any researcher, anywhere, for use in any research field.

The project has also ensured that the new sequencing technologies and computational methods used by the project and developed under its auspices remained reasonably standard, so that data generated in one time and place could be used by other researchers. The scope of the 1000 Genomes Project collaboration helped in this regard, as the many public and private entities involved, including the Broad's Genome Sequencing and Analysis Program, had to make sure that the data produced was accessible not only to their project partners, but also to the end user – the research community at large.

"The 1000 Genomes Project has been a free trade zone where these different technologies have been successfully brought together," Altshuler said. "The consortium has developed methods and paradigms for sharing and integrating data that ensure that the data generated around the world is more homogeneous, and can be more easily combined, shared, and analyzed."

The comprehensive catalog of human genetic variants that this collaboration has yielded is already impacting research.

"The 1000 Genomes Project data is a core resource for an enormous number of research projects," DePristo said.

Those projects range from population studies, in which researchers use the genomic data to trace migration patterns and the transmission of genetic information throughout human evolution, to disease studies.

In the study of human disease, the 1000 Genomes Project is expected to be a particular boon. Just as the slight variations in human genes can explain why people differ in appearance, genetic variation can also account for differences in disease risk. Medical researchers are already using the project's catalog of genomic data to locate [genetic differences](#) that may be contributing to rare and common diseases.

This research is being conducted in two fundamental ways: one route is to test those genetic variants that have been newly added to the public catalog for their role in disease. Another is to use the data collected by the 1000 Genomes Project as a filter, to rule out many candidate genes that are too common to be the "guilty" gene in the case of rare disease. In other words, if researchers are looking for a gene that contributes to a disease that appears in very few people, they can essentially rule out the millions of variants identified by the 1000 Genomes Project (which all occur at a frequency greater than 1%) because they're just too common.

"The 1000 Genomes Project is used by almost every rare disease genetic study that I'm aware of for this filtering step," Altshuler said.

In addition to the catalog of genetic variants that the 1000 Genomes Project has produced, it has also yielded some interesting scientific findings, the most significant of which could influence how genomic sequencing studies are conducted in the future.

One of the first genomic sequencing studies to analyze human genomes on a large scale across populations, the 1000 Genomes Project has been able to show that the rarer a mutation is, the more localized it is within its geographical region. That is, rare mutations tend to be passed on within local populations, and tend not to be present in populations that are farther away.

DePristo notes that this finding is not necessarily surprising, given that mutations tend to be passed on within families, but the scope of the 1000 Genomes Project, which has extended the reach of genome sequencing farther into the Americas and will later include previously neglected populations in Africa and the Indian subcontinent, is allowing researchers to compare genomes across populations and to understand the extent of this localization.

"We are one of the rare projects that has systematic surveys of populations across continental groups, which has allowed us to make meaningful comparisons across populations," he said.

This finding has implications for the design of future medical studies because it shows that, as researchers start to look at rarer mutations, these genetic alterations get more and more specific to populations.

"What this research says is that, as you get rarer and rarer mutations, if you're only working in Europeans then you're going to completely miss

the mutations that are rare in Asians. You are going to need to do these pan-ethnic studies," DePristo said. "You're going to need a lot of people from many different ethnic groups to really see the effect of rare mutations on disease."

More findings are expected to come out of the project, along with an even more expansive catalog, as the 1000 Genomes Project enters its next phase. A final report will be released after all 2,500 genomes have been sequenced.

More information: The 1000 Genomes Project Consortium, An integrated map of genetic variation from 1,092 human genomes, *Nature*. DOI: [10.1038/nature11632](https://doi.org/10.1038/nature11632)

Provided by Broad Institute of MIT and Harvard

Citation: 1,092 genomes and counting (2012, November 16) retrieved 16 April 2024 from <https://medicalxpress.com/news/2012-11-genomes.html>

<p>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.</p>
--