

An international consortium of scientists known as the 1000 Genomes Project has published a long-awaited map of variation in the human genome, cataloging the subtle differences that shape our bodies and influence our risk of disease. The results, published today (31 October) in *Nature*, were derived from the genome sequences of 1,092 volunteers hailing from 14 populations in Europe, East Asia, Africa, and the Americas. They should help scientists more efficiently hunt for the genetic causes of disease, by comparing mutations in a patient's genome against those seen in his own country or ethnic group.

"The 1000 genomes project is the backbone of our understanding for human variation," said [Ewan Birney](#) from the European Bioinformatics Institute, who is not part of the consortium. "Both the data and the methods will be reused many times in the forthcoming decade."

The 1000 Genomes Project takes advantage of the dramatically falling cost of genome sequencing—and even contributed to the continued reduction of those costs by increasing demand. The pilot phase, [published in 2010](#), had already documented most of the common variants found in more than 5 percent of people. In the new study, the researchers focused on the world's rarer variants, by sequencing every nucleotide in the volunteers' exome about 50 to 100 times over. (The rest of participants' DNA was also sequenced, but at a much lower coverage rate of just 2 to 6 times.)

The results revealed 38 million single nucleotide polymorphisms (SNPs), 1.4 million short insertions and deletions, and more than 14,000 larger deletions. The team believes that this represents at least 98 percent of the variants found in at least 1 percent of people, and around half of those found in 0.5 percent.

Consortium leader [Gil McVean](#) from Oxford University said that the "single most important result" is that common mutations are shared by people across the globe, while rarer ones are confined to certain ethnic groups or nations. Variants that could strongly increase one population's risk of a disease might be non-existent in another group of people.

Rare variants can also hint at historical connections. For example, if one of the project's Spaniards shares a rare mutation with just one of the 1,091 other volunteers, there's a 48 percent chance that person comes from Central or Southern America. "It reflects the history of colonization," said McVean. "These rare variants might help us to pick up unexpected connections between populations that we didn't know about."

Importantly, the data generated by the is completely open. "Everything is fully accessible," McVean said. "The world can see the data at the same time as I can."

This will provide a useful anchor for future research, said [Eric Topol](#), a geneticist from the Scripps Research Institute who was not part of the study. "But a serious limitation, and one that I have stressed for years, is that none of the 1,092 individuals in this project had their phenotype characterized," he said. Without that information, he explained, researchers cannot link their genetic differences to the state of their health.

"That's deliberate," countered McVean. Because the consortium always planned to make their data publicly available, they could not also release physical traits of their subjects. Plus, he added, even a study of a thousand people is too small to identify genetic changes that will affect disease risk. Instead, the project's data is more suited to refining the results of larger studies.

For example, [in a study published in August](#), McVean analyzed a region of the genome that had been linked to risk of multiple sclerosis. Using the 1000 Genomes catalog, his team pinpointed one variant—in a gene called TNFRSF1A, which encodes a receptor for an inflammatory molecule called TNF—that drives the increased risk of the disease.

"The 1000 Genomes Project has already benefited countless studies," said [Chris Gunter](#), a geneticist from the HudsonAlpha Institute for Biotechnology who was not involved in the study. "[This paper] formalizes a dataset which people have been using already to make discoveries. We absolutely have to have this resource to have the proper power to study both human history and medical genomics going forward."

In the final phase of the project, which is expected to finish next year, the teams will sequence people from more populations still, including those in South Asia and parts of Africa. McVean anticipates that other scientists will then plug the remaining gaps, such as gathering and sequencing genomes from people in Australasia. "My hope is that [the 1000 Genomes Project] will be superseded," he said.

**The 1000 Genomes Project Consortium., "An integrated map of genetic variation from 1,092 human genomes," *Nature*, 491: 56-65, 2012.**