

## TOOLBOX

# Long-awaited databases reveal breadth of genetic variation

BY KATE YANDELL

4 NOVEMBER 2015

Two massive efforts to sequence the DNA of more than 11,000 people are finally complete. Together, they provide the most detailed picture yet of genetic variation across the general population. They also give researchers a starting point for finding genetic variants tied to a variety of conditions, including autism.

The **1,000 Genomes Project**, described in two papers published 1 October in *Nature*, includes the genomes of more than 2,500 people across five continents<sup>1,2</sup>. A second database, described in the same issue of *Nature* and called the **UK10K project**, provides sequence data for nearly 9,000 people from the U.K., including some people with health conditions<sup>3</sup>.

For the **1,000 Genomes Project**, researchers sequenced whole genomes and then performed more in-depth sequencing of protein-coding regions. They uncovered 88 million genetic variants — or 4.1 to 5 million per person.

These variants include spots at which single DNA letters have been swapped with others, as well as deletions and duplications of small and large swaths of DNA. Structural variants, which span more than 50 nucleotides, are difficult to detect and can sometimes obliterate the function of genes. The researchers also identified 240 genes whose deletion does no obvious harm.

Scientists have cited data from the project in several hundred papers that mention autism. Many of these studies compare sequences from people who have autism with sequences from this database to determine whether **certain rare variants** are unique to people with autism.

## Efficient decoding:

Other researchers have used these early data to make it easier to detect autism-linked variants

that are **common in the general population** in genome-wide association studies. These studies typically require sequencing data from thousands of individuals with the disorder — which can be expensive. To survey genomes at a lower cost, researchers can identify common variants in people with autism at a limited number of genomic sites and then make educated guesses about the rest of their variants. The 1,000 Genomes Project provides an atlas of which variants tend to be inherited together.

For the **UK10K project**, researchers sequenced the whole genomes of 3,781 individuals without known disorders and only the protein-coding regions in 5,182 people with autism, schizophrenia, obesity or rare diseases, or their family members.

The project uncovered 46 million genetic variants, including 24 million that were previously unknown. These include single-nucleotide swaps as well as duplicated or deleted DNA segments.

The researchers found no variants clearly associated with autism. But when they added data from the **Autism Sequencing Consortium** to boost their numbers, they identified 13 genes that are mutated unusually often in people with autism. Many of these genes are linked to developmental disorders and intellectual impairment.

Researchers can access data from both projects freely online. They can also purchase cell lines from 1,000 Genomes Project participants who carry specific variants through the **Coriell Institute for Medical Research**, a nonprofit research organization in Camden, New Jersey, focused on understanding the human genome.

### REFERENCES:

1. Auton A. *et al. Nature* **526**, 68-74 (2015) [PubMed](#)
2. Sudmant P.H. *et al. Nature* **526**, 75-81 (2015) [PubMed](#)
3. Walter K. *et al. Nature* **526**, 82-90 (2015) [PubMed](#)