

DNA 'Big Data': Geneticists build massive databases to fight diseases

Early last year, three researchers set out to create one genetic data set to rule them all. The trio wanted to assemble the world's most comprehensive catalogue of human genetic variation, a single reference database that would be useful to researchers hunting rare disease-causing genetic variants.

Unlike past 'big data' projects, which have involved large groups of scientists, this one deliberately kept itself small, deploying just five analysts. Nearly two years in, it has identified about 50 million genetic variants — points at which one person's DNA differs from another's — in whole-genome sequence data collected by 23 other research collaborations. The group, called the Haplotype Reference Consortium, will unveil its database in San Diego, California, on 20 October, at the annual meeting of the American Society of Human Genetics.

Geneticists have not always been so willing to share data. But that seems to be changing. "It's been surprisingly easy to bring all these data sets together," says Jonathan Marchini, a statistical geneticist at the University of Oxford, UK, and one of the consortium's leaders. "There is a lot of goodwill between the people in the field; they all understand the benefits of doing this and have worked hard to make their data available."

Read full, original article: [Giant gene banks take on disease](#)