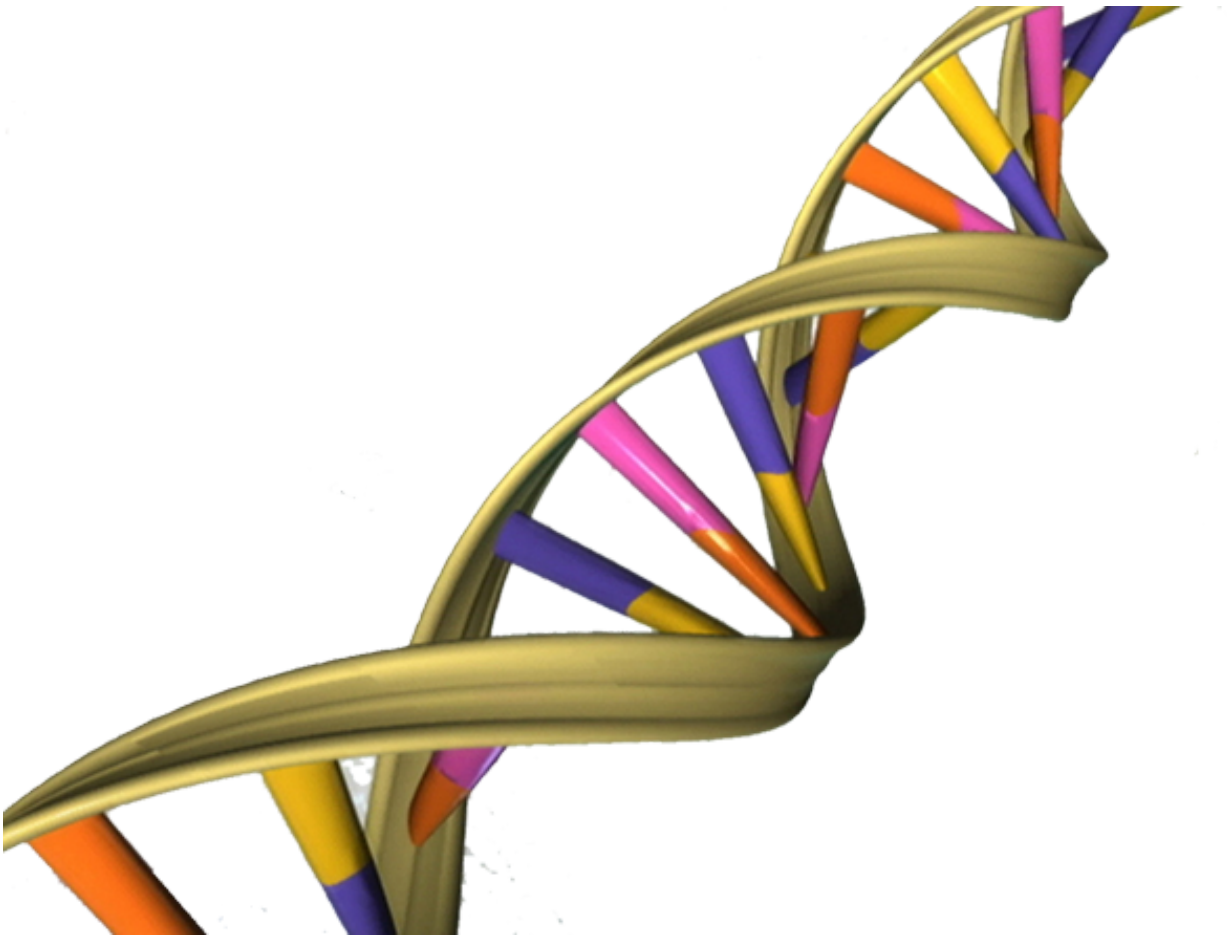
10K genomes project explores contribution of rare variants to human disease and risk factors

September 14 2015



A depiction of the double helical structure of DNA. Its four coding units (A, T, C, G) are color-coded in pink, orange, purple and yellow. Credit: NHGRI

The largest population genome sequencing effort to date is published today in *Nature*. A series of papers describing resources and application of the data is published at the same time in *Nature*, *Nature Genetics*, *Bioinformatics* and *Nature Communications*.

Rare genetic variants are changes in DNA that are carried only by relatively few people in a population. The UK10K study was designed to explore the contribution of these rare genetic variants to human disease and its risk factors.

"The project has made important new contributions towards describing the role of rare genetic variants in a broad range of disease scenarios and human traits." says Dr Nicole Soranzo, corresponding author from the Wellcome Trust Sanger Institute. "It has shown that the value of sequencing a few thousand individuals is high for highly penetrant, rare diseases, but that for complex traits and diseases much larger sample sizes will be required in future studies. The data and results produced by this project will be instrumental for these future efforts."

The project studied nearly 10,000 individuals, both healthy and affected by disease. The conditions included very rare disorders inherited in families, and more common diseases such as autism, schizophrenia and obesity. In healthy people, 64 different biomedical risk factors such as blood pressure or cholesterol levels were studied. By characterising the DNA sequence of these individuals, the project gained insight into the contribution of rare variants to a broad range of disease scenarios, and discovered new genetic variants and genes underpinning disease risk.

"The UK10K project has increased the resolution of genetic discoveries. It has enabled access to a much denser set of variants within the genome in the UK population, which can be used to refine our understanding of genetic effect on phenotypic traits," explains Richard Durbin, senior UK10K researcher at the Sanger Institute. "In earlier studies either very

rare variants with big effects or common variants, which usually only have small effects, could be analysed. Now we have been able to explore an increased part of the spectrum of variation in between the very rare and the common ones."

A series of papers published today in *Nature* and *Nature Genetics* in collaboration with other investigators demonstrates the value of these data for genetic discoveries.

As efforts continue to characterise the genetic underpinnings of complex diseases, the data and results of this study are expected to enable the next wave of discoveries. The UK10K sequence reference panel, described in greater detail in a companion paper published in *Nature Communications*, has been shown to greatly increase the ability to characterise rare variants in large population samples available to the worldwide research community. This resource will enable researchers to 'fill in' missing data from lower resolution genotype studies, allowing them to explore full genotypes more quickly and cheaply.

In addition, the authors have developed a web-based browser of association based on the Dalliance platform, described in a companion paper in *Bioinformatics*. This genome browser allows the easy retrieval of association results for all disease risk traits analysed in the study. Scientists investigating these specific disease risk factors will be able to directly access the consequence of a person's DNA sequence to see how common any genetic variants they have are and what traits these variations are associated with.

"The UK10K project was an enormous undertaking and has laid the ground for future studies, " says Klaudia Walter, a leading author from the Sanger Institute. "For instance, the benefits of the new UK10K haplotype reference panel are already being realised in analyses of international consortia as well as the 0.5M people UK BioBank study."

More information: The UK10K Consortium (2015). The UK10K project identifies rare variants in health and disease. *Nature*. DOI: dx.doi.org/10.1038/nature14962

Zheng H, Forgetta V et al. (2015). Whole-genome sequencing identifies EN1 as a determinant of bone density and fracture. *Nature*. DOI : dx.doi.org/10.1038/nature14878

Huang J, Howie B, et al. (2015). Improved imputation of low-frequency and rare variants using the UK10K haplotype reference panel. *Nature Communications*. DOI: dx.doi.org/10.1038/ncomms9111

Geihs M, Yan Y, et al. (2015). An interactive genome browser of association results from the UK10K cohorts project. *Bioinformatics*. DOI: 10.1093/bioinformatics/btv491