First report from the world's most ambitious sequencing project
20 July 2022

Scientists at deCODE genetics—a subsidiary of Amgen—together with collaborators from Denmark report on the whole genome sequences of 150 thousand participants in the UK biobank in a paper published in the journal *Nature* today.

This is the first report from the largest whole genome sequencing effort to date where scientists from deCODE genetics and from the Wellcome Trust Sanger Institute are set to sequence 500 thousand whole genomes in three years.

The scientists at deCODE genetics found 600 millions SNPs and indels in these 150 thousand genomes corresponding to 7% of those that can theoretically occur in the human genome. It is however likely that some of the theoretically possible variants are incompatible with life.

This large dataset allowed the scientists to separate regions that are tolerant to large diversity in sequence from those that are not. The assumption is that regions that are intolerant to sequence diversity are important to human survival and procreation. It has long been held that coding exons are the regions most important to human survival. However, when the 1% of the genome with sequences that are best conserved are examined only 13% of them are coding exons.

"Data of this type and quantity are going to revolutionize our ability to identify and characterize intergenic sequences of importance to human diversity, be it to risk of disease and response to treatment or some other attributes," said Kari Stefansson the founder of deCODE and one of the authors of the paper.

Furthermore, scientists at deCODE also report on the association of variants that were not identified through whole exome sequencing with diseases and other phenotypes.

Participants in the UK biobank are of diverse genetic ancestry and have forefathers from most of the countries of the world. The scientists determined that 85% of the participants could trace most of their ancestry to the British Isles. The scientists also found a large group of participants who can trace their ancestry mostly to Africa and South Asia. This study is likely to represent the largest set of whole genome sequenced individuals of African and South-Asian origin. However, the imbalance in the ethnic mix of those contributing sequences to this study as well as to other studies already published is unfortunate from both societal and scientific point of views. Scientists at deCODE genetics are determined to work towards more ethnically balanced sequencing cohorts in the future.


Provided by deCODE genetics