biobank uk

Improving the health of future generations

WHOLE GENOME SEQUENCING: Transforming health research

A ground-breaking initiative to sequence the whole genomes of 450,000 UK Biobank participants has been announced. The venture is funded by government, charity and industry and is in addition to the 50,000 volunteer samples already being sequenced by UK Biobank with funding from the MRC. It is the biggest endeavour of its kind ever undertaken, and will transform the way in which scientists study health.

What is the purpose of whole genome sequencing?
Whole genome sequencing turns blood samples into usable data. Sequencing all of the participants in UK Biobank means that these data can be readily shared among all approved researchers, anywhere in the world, whenever they are needed. Each individual has a unique genetic code of 3 billion building blocks. These make up the 24,000 active genes (and much more) inside a human cell that control biochemical processes that underpin life. The vast majority of genes are the same in everyone, but small changes account for our differences, and can make us more susceptible to illness, or affect the way we respond to treatments. In UK Biobank's case, the genome from a single person is of no real interest. However, by making it possible for researchers to compare thousands of genomes – or hundreds of thousands – patterns will emerge, differences will become clearer and the impact on a wide range of diseases of individual genes or combination of genes, along with lifestyle and envrionment, are likely to be found. This might suggest new ways to tackle disease, and help identify people at higher risk, so that preventive strategies such as screening programmes can be better targeted.

Why is this important?
UK Biobank's goal, with your support and the support of our other 500,000 volunteer participants, is to create a resource of detailed information that will help scientists find out why some people get illnesses like cancer, heart disease, diabetes and depression and others do not. With knowledge, comes the power to produce change. Detailed information on genetics, linked to lots of other information about your lives, the illnesses you get and how you are treated will boost understanding, and speed up advances in health care.

How will participants' privacy be protected?
UK Biobank has a number of systems in place to ensure that you are not identified when data you have generously provided is used for health research. For instance, data are separated at the point of collection from personal information (such as name, date of birth) and retained separately. Data are also extensively encrypted and access to the decryption keys is very limited (to those that need them). Further, when data are released to researchers they are fully de-identified. In addition, all research is undertaken under a legal agreement with UK Biobank which prohibits any researchers from identifying participants.

Will UK Biobank participants be identifiable through the data? No, identifiable data are not provided to researchers.

Who is funding the work?
A pilot project (the Vanguard) to sequence the whole genomes of 50,000 UK Biobank participants is already underway. (You may remember reading about its launch in last year's Newsletter.) The £30 million funding was provided to UK Biobank by the UK's Medical Research Council (MRC) and the sequencing is being done at the Wellcome Sanger Institute in Cambridge.

This new project, to sequence the other 450,000 participants, is funded by government, charity and industry:

- £50 million from the government's UK Research & Innovation (UKRI) Industrial Strategy Challenge Fund;
- £50 million from the charity The Wellcome Trust;
- £100 million (£25 million each) from four of the world's leading pharmaceutical companies, GlaxoSmithKline (GSK), AstraZeneca (AZ), Johnson & Johnson (J&J) and Amgen.

Where will the sequencing take place?
Having gained UK Biobank approval for the sequencing, the applicants appointed the Wellcome Sanger Institute and deCODE in Iceland to conduct the next phase of work in roughly equal measures. UK Biobank will provide them with a small portion of blood, currently stored at minus 80 degrees in our Coordinating Centre in Stockport, from each participant. They will then provide the sequence data to UK Biobank.

Why now?
In recent years, advances in technology have revolutionised the time it takes to sequence the human genome, and the cost involved. The first human genome was sequenced in 2003 after several years and at a cost of nearly US $3Bn. It is now feasible to sequence large numbers very much more rapidly and at far lower costs. Sequencing all 500,000 UK Biobank participants now will deliver the project's genetic promise very much quicker than would have been expected when recruitment began in 2006.

When will the genetic data be used for research?
Scientists are already using participants' genetic data in the form of genotype and exome sequence data. However, these data make up only a proportion of the human genome. It is intended to complete the whole genome sequencing of all 500,000 participants during the next 2 years. An outline timetable is as follows:

- At the end of May 2020, the four pharmaceutical companies will be provided with access for analysis to the first tranche of sequence data (anticipated to be for about 125,000 participants) linked to all of the other data in the UK Biobank resource.
- After an exclusive access period of 9 months to the benefit of the applicants, these data will be made available to all other approved researchers around the world in March 2021.
- Similarly, after completing the sequencing of all 500,000 participants in early 2022, the consortium will again have exclusive access to these linked data for 9 months before all of the data are made widely available by the end of 2022.

Why do the industry partners get an exclusive period to use the data?
With government, industry and charity working together, a vast amount of genetic data will be made available to the health research community. Without this investment from industry, these data would not be available. Providing an exclusive period has already been used successfully to secure the investment for exome sequencing led by Regeneron and partners. Those exome data are now available to health researchers, and more than 100 health studies using them are now under way.

What other genetic analyses have taken place?

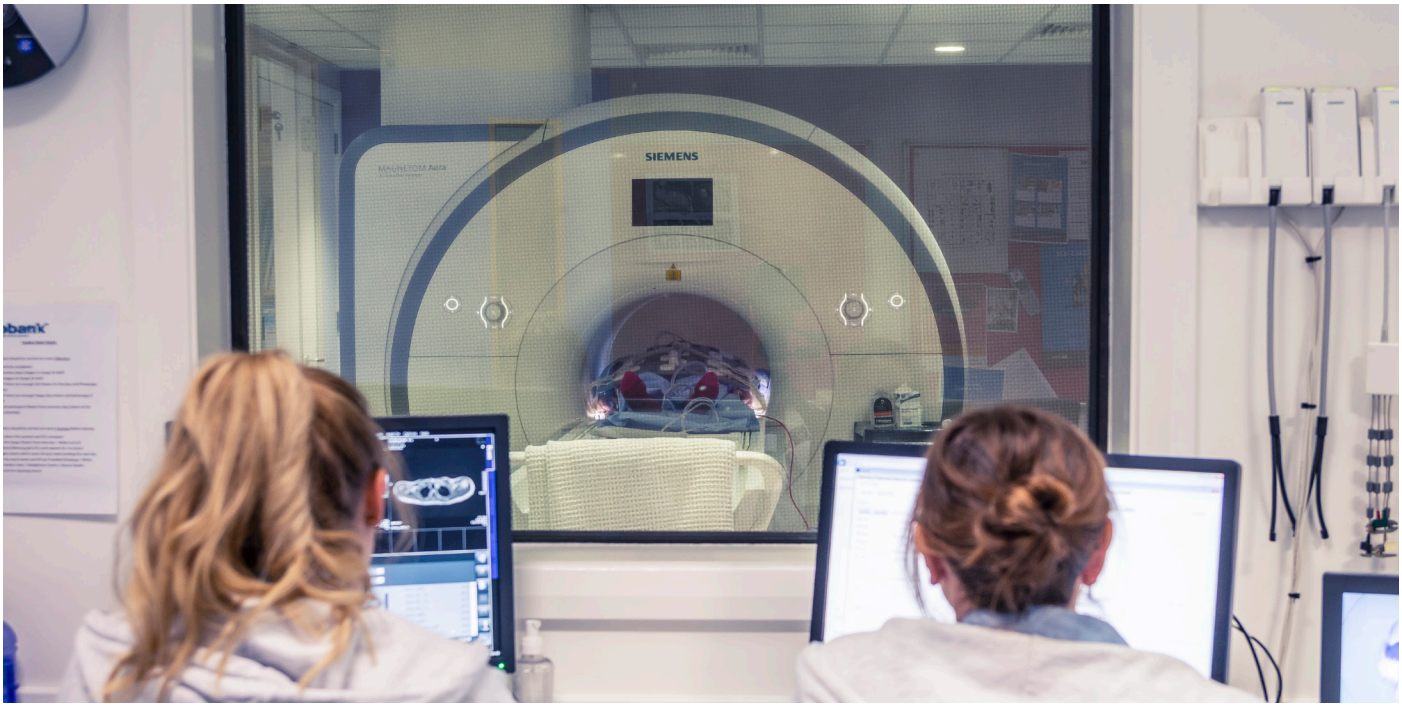
This new sequencing work supports other genetic analyses which are already contributing to research:

- Genotyping (completed) – analysis of around 800,000 key points of the 3 billion building blocks that make up the human genome. This enables statisticians to impute (i.e. estimate, based on past analysis) a further 90 million points;

- Exome sequencing (under way) – this involves the study of about 3% of the human genome, and targets the active genes that produce proteins that build cells and control cell activity. 50,000 exomes were sequenced in a joint venture by the pharmaceutical companies GSK in the UK and Regeneron in the US. The first exome sequenced data were released to the global research community in March 2019. Regeneron and a consortium of five other companies are now exome sequencing the other 450,000 samples. Again, these data will be made available to other researchers, with an additional 100,000 exomes to be released in April 2020.

How will UK Biobank and health research benefit from this collaboration?
Researchers are already using UK Biobank genetic information, linked to a wide range of health outcomes, to study health. Many studies have published in expert scientific journals, and last year UK Biobank featured in one of the most prestigious, Nature. New release of sequence data will provide even more impetus for research to bring about advances in preventing and treating a wide range of disorders more quickly.

Is there a requirement that results must be published within a particular time? If so, when and where?
All researchers using the UK Biobank resource are required to report annually on the progress of their work. They are expected to make their findings public and tell us what, if anything, they have patented or otherwise found out that is significant for health. Researchers are also obliged to return their results to UK Biobank so that other researchers can benefit from them.



How big – and how important – is this development?
UK Biobank is of national and international strategic importance in the global effort to improve health. The addition of whole genome sequencing is tremendously exciting. It builds on the baseline data you provided when you joined the study, and subsequent enhancements (genotyping all 500,000 participants, exome sequencing currently under way, detailed imaging of 100,000 participants, links to health records, biochemistry on 500,000, week-long activity data on 100,000, eye measures on 80,000 and a number of well-supported online questionnaires). Research based on all of this information will contribute to better treatments and improved prevention and diagnosis in the coming years.

What is the benefit for UK Biobank participants and the wider public?
On joining UK Biobank, you were told that your participation may be of no direct benefit to you, but that it was an investment for future generations. UK Biobank does not provide feedback to participants, but the potential benefits for the wider public (including UK Biobank participants) of having these whole genome sequence data available for research are enormous in understanding and treating diseases. Your data and samples now comprise one of (if not the) most extensive and widely used research resources worldwide.

Relevant links:

Whole genome sequencing – first 50,000 announced: https://www.ukbiobank.ac.uk/2018/04/whole-genome-sequencing-will-transform-the-research-landscape-for-a-wide-range-of-diseases/

UK Biobank genetics in the news: https://www.ukbiobank.ac.uk/genetics-in-the-news/

Tantalising and exciting: UK Biobank genetics opens the door to a new era of health research: http://www.ukbiobank.ac.uk/2018/10/tantalising-and-exciting-uk-biobank-genetics-opens-the-door-to-a-new-era-of-health-research-2/