

## Why the UK wants to share everything it knows about the genetics of 500,000 Britons



ly into Britain's Manchester Airport these days and you might spot a new landmark amid the urban sprawl on the ground below. Two huge white cylinders stand sentinel: the only outward sign of a massive biomedical project that promises a revolution in science and health care. And like all revolutions, this one is born in blood.

The cylinders pump liquid nitrogen into a facility called the UK Biobank. Inside the walls of this anonymous-looking industrial unit, scientists hold the bodily fluids of half a million Britons in state-of-the-art, robot-managed freezers. Research does not come more open-access than this. Blood biochemistry, genetic analysis, images of brains, hearts and other organs — all the internal secrets of volunteers — are combined with intimate personal confessions about lifestyle, such as how many sexual partners someone's had, how much alcohol they drink and if they routinely drive faster than the motorway speed limit.

The data are packaged with confidential medical histories such as hospital visits and surgeries and then handed to all qualified to use them, wherever they may be.

The results of that largesse are flowing. In a given month, dozens of scientific studies can appear based on UK Biobank data. They range from the curious — how many cups of coffee can safely be consumed in a single day — to the fundamental, such as the discovery that specific gene variants are associated with disease or healthy life expectancy. And in an area of research where size is crucial, such studies count their volunteers not by the hundred or the thousand, but by the hundred thousand. More than a century after Ernest Rutherford's Manchester lab showed the world how to [unlock the secrets inside the atom](#), the city is showcasing how Big Data can answer fundamental questions about human health.

knowable magazine logo

"The UK Biobank is the gold standard right now," says Josh Denny, a researcher in biomedical informatics at Vanderbilt University Medical Center in Nashville, Tennessee. "Worldwide it's the benchmark of an open-access large database with rich information and genetics." Denny published an article on this subject — using clinical data to get the most out of genomic research — for the [Annual Review of Biomedical Data Science](#) in 2018. "What we do when we bring health care and genetics data together is to get at the outcomes that are important to us," he says.

Even as results emerge touching on everything from aging to susceptibility to asthma, the biobank effort isn't without its detractors or bumps in the road. Some worry that the broad nature of the research done with the samples makes it impossible for volunteers to give proper consent. And in October, a high-profile paper was withdrawn because of technical problems in the way biobank data were analyzed.

But to scientists like Denny, the promise is clear. "This is a resource for the world," he says.

### Making the genetic connection

The principle behind the UK Biobank is ambitious: to link health outcomes to the genetic data that pour from DNA sequencing machines across the world. Medicine traditionally is guided by a patient's physical symptoms and measurable changes to physiology — what biologists call the phenotype. Integrating genetic data — a patient's genotype — into these deliberations could help tailor treatments to boost their effectiveness, or even identify people at higher risk of developing a given disease, who could be offered help earlier. But to make that work, scientists need to connect the dots: match genotype to phenotype, find patterns and connections in the way people's DNA varies and the way their health does, too.

Those connections are becoming clearer. In February this year, for example, scientists found genetic markers in the biobank data that [linked high cholesterol to the development of motor neuron disease](#). Cholesterol-lowering drugs like statins, the results suggest, might prevent this deadly and incurable condition. Last month, a different team combed through the genetics of 334,000 of the half-million people signed up to the biobank project to identify genes associated with problematic metabolism of uric acid, which causes health problems including the painful condition gout. From head to toe, month by month scientists are using the biobank information to reveal everything from the benefits of being left-handed to the genetic basis for hearing loss and the damage that diabetes can do to the heart.

The UK project isn't the first to recruit volunteers to identify links between genes and disease. National efforts are also under way in Estonia, Sweden, Iceland, China and Mexico. And back in the 1990s, the Icelandic company [deCODE](#) set out to build a database of the genes they found in the country's population. Analysis of the Icelandic data, now owned by the US biopharmaceutical giant Amgen, continues — for example, the company is now working on medicines to mimic the heart-protecting effects of a gene variant carried by one in 120 Icelanders.

Even with its high volunteer numbers, the UK Biobank isn't the largest project of its type either. The British effort can call on data from some 500,000 recruits — but, as it often does, the US military has gone further. In April, the US [Million Veterans Program](#) signed up its 750,000th participant since it began in 2011, and still wants more to reach its eponymous goal. The MVP screens the health and genomes of veterans to probe the genetics of post-traumatic stress disorder, diabetes, heart disease, suicide prevention and other topics of particular relevance to that community.

all of us biobank to store million samples for medical research

Image not found or type unknown

Image: Mayo Clinic

So, what's so great about Great Britain's project? Access. Other biobanks set up around the world are useful projects that can help answer some specific questions, Denny says. But it's often difficult for outside scientists to get access to the data. Some national projects guard their secrets from foreign eyes as a way to give their own researchers a head start. Others fret about privacy and losing the trust of participants if they were to start sharing their information more widely.

The UK Biobank is unique because open and free data access for everyone was the plan from day one, says [Rory Collins](#), an epidemiologist at the University of Oxford and chief executive of the UK Biobank project. "We wanted to build something, a resource, in the same way as they built CERN," the European particle physics lab near Geneva, he says. "This wasn't a grant application which has to have a specific hypothesis." It's a point that other people attached to the Biobank project make repeatedly: This is a basic science project. If they built it, they thought that scientists would come and want to use it.

They have come, and continue to do so. At last count, 13,000 scientists in 77 countries, from Australia and Malaysia to Russia and Jordan, have been given access to data on topics from cognition and sleep to mental health.

## Marshaling money and volunteers

Prompted by a call from British scientists to invest in the promise of DNA, the biobank started life as a funding pledge from Tony Blair's new Labour government in 1998. Backed by the Medical Research Council, a state funder, and the Wellcome Trust, a biomedical charity, the project was based on the principles of the famous [Framingham Heart Study](#), an influential population cohort study that followed 5,200 residents of Framingham, Massachusetts, as a way to find factors that influence cardiac illness.

The UK started to recruit volunteers to its study in 2006 and reached its half-million goal four years later. It focused on individuals ages 40 to 69 because organizers figured it would be most useful to study older people, who tend to more quickly show the signs of ill health that researchers are interested in. (Indeed, the carefully preserved samples at the Manchester HQ now represent the earthly remains of at least 20,000 volunteers who have since passed away.)

Participants weren't paid and had to spend hours at one of several regional centers, where they surrendered blood and urine, had their health examined and filled in surveys on their habits and lifestyle. As a result, the biobank population is not as diverse as geneticists might like, Collins admits, especially if the results are supposed to be useful around the world. Some 94 percent of people the biobank signed up are white, and certain socioeconomic groups, including young, low-income white men, are underrepresented.

Initially, the blood samples were analyzed for simple variations in genetic sequence, such as single nucleotide polymorphisms. These single base-pair changes in DNA occur at specific places in the genome and can explain traits such as eye color and inherited diseases such as cystic fibrosis and sickle-cell anemia. They also act as markers to indicate risk of complex diseases, including diabetes and [Alzheimer's](#).

The first of these genotyping data were made available for a group of 150,000 biobank participants in May 2015. Results from the other 350,000 people were added two years later. That fulfilled the original plan, but as genetic sequencing has become faster and cheaper, other researchers wanted to go further — some of them in private companies. In 2017, the drug firms GSK and Regeneron offered to sequence the "exome" of 50,000 UK Biobank participants. This gives a readout of the sections of DNA that actually code for proteins, and is seen as a more powerful way to locate useful information that could be used to develop medicines.

The companies agreed to pay the bill, but wanted something in return: exclusive access to the data. They were given 6 to 12 months, then earlier this year the information was released to the wider scientific community. A larger group of pharma companies is now working on exome sequences for the remaining 450,000 volunteers under the same arrangement, with the data due to be fully released next year. The total cost of the exome sequencing — paid for by the commercial firms — will be \$150 million.

That's a sensible and pragmatic way for publicly funded science to work with commercial companies, says cognitive epidemiologist [Simon Cox](#), who uses UK Biobank data to study cognitive aging at the University of Edinburgh. "It's not necessarily ideal," he says. "But if this is the model we have to adopt to get this stuff

done then I don't think it's spectacularly untoward."

Cox recently published a study that exploits another extension of the biobank's original goals. In April 2016, organizers started to invite thousands of the participants back to take detailed MRI scans of their bodies, brains and other organs and add them to the database. Some 40,000 of the original volunteers have been scanned this way so far, and the project wants to make that 100,000 by 2022. Some of these volunteers also agreed to take a battery of psychological tests when they returned — to probe reasoning and intelligence with questions not asked in the original assessment. Cox and his colleagues took advantage of this to revisit one of the most controversial historic questions in psychology: [Does a bigger brain make someone smarter?](#) (Spoiler: It does, but not much.) Particularly important, the research showed, is the structure of white matter and the relative size of individual regions such as parts of the frontal cortex.

To do this analysis, the scientists compared the test scores and brain volumes of more than 18,000 people — way more than had ever been gathered for such a study before. At a stroke, Cox says, the biobank's massive numbers address one of the strongest criticisms of findings based on MRI scans — that they include too few people to give reliable results.

[de biobank x](#)

“It’s the biggest dataset there is,” says Lisa Nobis, a neuroscientist at the University of Oxford. And it shows the way that science is going. “When MRI first came out, you could get a paper in *Nature* or *Science* with just 15 people. But in the last few years, studies have shown that less than 50 people is not reliable at all” as a general rule. She is using the biobank scans to build up a picture of how the brain’s [hippocampus shrinks with age](#) and with dementia. The plan is to produce a reference tool of normal hippocampus size for different age groups that future clinicians can use to spot problems earlier.

It’s the kind of study, Nobis says, that can only be done with a very large sample size. Scientists’ attempts to boost numbers of brains by combining results from sets of smaller MRI studies run into their own set of problems: Technical differences between the way scans are done in different centers often make it difficult to meaningfully compare them.

The investment in MRI scanners and other shiny, new pieces of equipment gives the Manchester headquarters of the biobank the air of a wealthy private hospital. Elderly volunteers are helped along its spotless wide corridors and shuttled among different scanners, depending on the level of tissue detail required. In one room, staff are preparing to take ultrasound scans of the thickness of the wall of a patient’s carotid arteries. Again, it’s a measurement that is being taken because it can be, rather than because there’s a solid scientific or medical need for the data at the moment.

The scientists hope this type of basic anatomical information will be useful one day in the future — especially as the biobank database links to details about hospital admissions, treatments and surgery held in the electronic records of Britain’s National Health Service and aims to expand the links to even more granular clinical information about complaints, symptoms and prescriptions held in primary care health records of thousands of doctors’ offices. According to biobank staff, most volunteers realize that the data gathered from them and others is unlikely to benefit them directly, but they say they hope it will help future generations, either of their own family or someone else’s.

One such volunteer is Irene Soulsby, a 61-year-old from Newcastle upon Tyne, in northeast England, who is happy to disclose and talk about her involvement. (The biobank promises confidentiality, if preferred.) Soulsby enrolled in 2008 and did so, she says, because she had previously benefited from successful treatment for cancer, which she knew was based on prior medical research. “I’m here because somebody’s done this for me,” she says. There is another motive too: Given her history, she welcomes the extra tests and ongoing scrutiny of her health. “It’s an amazing resource and I’m really proud to be part of it,” she says.

## **From the start, some skepticism**

The broad scope of the project was not always viewed as an asset. Before it even got off the ground, a 2003 editorial in the leading medical journal the *Lancet* voiced skepticism. It called the UK Biobank “[a project in search of a protocol](#)” that many researchers in Britain saw “as an ill-conceived, politically motivated project, in which consultations have only been done to give an appearance of legitimacy and in which the scientific case has not been made for its design.”

And despite the apparent success and cranking up of the scientific output — more than 270 research papers appeared in 2018 alone based on UK Biobank data — such criticisms have not gone away entirely. [Kenneth Weiss](#) is an emeritus professor of anthropology and genetics at Pennsylvania State University and a long-standing critic of mega-genomics projects based on the promise of big data. “There is a need for skepticism. There are better ways to find and document the complex causation of genetics than this open-ended genomic approach,” he says. Having swallowed more than \$250 million so far, the UK project is certainly an expensive option, and much of the burden is carried by UK taxpayers.

As with all medical studies, the biobank results are only as strong as the data and assumptions they rest on. And in October, scientists discovered a flaw in a biobank analysis that proved fatal for the high-profile study that was based on it. A technical glitch led to the biobank undercounting how many older people carried a specific mutation in the *CCR5* gene that confers resistance to HIV. Researchers used this as evidence in a paper that concluded that carrying this mutation could shorten people’s lives. The finding was significant because this is the genetic change the Chinese scientist [He Jiankui](#) claimed to have deliberately introduced to twin baby girls in 2018.

The paper was retracted, but genetic experts who reanalyzed the data said the problem was a one-off that doesn’t undermine other studies using biobank data.

There are also concerns about ethics and consent. When participants sign up for the project, they do so without knowing exactly what will be done with the information they provide. Soulsby says she trusts the project officials to use it properly, but not everybody is happy with the arrangement. Traditionally, scientific research projects ask for explicit consent from volunteers after outlining exactly how the samples or data they provide will be used. The broad spectrum of research that blossoms from a biobank project make this impractical, so a more general type of “open consent” is used instead, under which volunteers essentially give blanket and open-ended approval for their samples to be used however project organizers deem fit, even after their own deaths.

Follow the latest news and policy debates on sustainable agriculture, biomedicine, and other ‘disruptive’ innovations. Subscribe to our newsletter.

[SIGN UP](#)

In a critical essay published in 2017 in the journal *PLOS Biology*, law and ethics experts Timothy Caulfield and Blake Murdoch, of the University of Alberta, Canada, [argued that this isn’t good enough](#). “There remains a deep lack of clarity around basic legal and ethical principles,” they wrote of UK Biobank and projects like it. “The international research community has built a massive and diverse research infrastructure on a foundation that has the potential, however slight, to collapse, in bits or altogether.”

One issue that Caulfield and Murdock highlight is what lawyers call the “rights of control.” It’s possible for example, that participants, or their families, could seek payment for profitable outcomes, or even challenge the use of their data for specific projects they disagree with. And there could be more to disagree with than one might expect. The project was sold as a biomedicine effort — and the terms of the UK Biobank open consent process ask volunteers to agree to vague-sounding “health-related research



purposes” — but some of the research does stretch the definition of what many people might think of as medicine.

Earlier this year, for example, an international team of researchers used the biobank data to show how genetic variants associated with lower educational attainment and lower socioeconomic status bunched together in former coal-mining districts of the UK, a report that raised concerns among some in the field. Such analyses are controversial because they attempt to extend the links between genotype and phenotype from medical traits to complex — and contested — social behaviors and values.

[Peter Visscher](#), a geneticist at the University of Queensland in Brisbane, Australia, who worked on the study, says it does have a strong link to health — and that’s why the UK Biobank officials approved it. The genetic patterns on educational attainment that the team analyzed closely associate with risk of conditions like coronary heart disease and schizophrenia, he says.

Genes, lifestyle, environment, health, social values, behavior: Scientists across the world are using the biobank data to help piece together and sort out the complexity of what makes people — of Britain, of Iceland, of the US military, and of the world — tick.

It’s a collaborative effort, and the UK project — and UK science — is playing its part. At a time when political divisions over Brexit have triggered endless debates about Britain’s place in the world, UK scientists are spending UK taxpayers money to send the health and genetic data of UK citizens to all who can use it, whoever and wherever they are, for the common good. It’s science sans frontiers: a principle perhaps out of step with the prevailing political winds, but one that should be cherished, Denny says.

“You accomplish a lot more with a free trade of scientific ideas than if you keep them to yourself,” he says. “I guess the world is very fortunate that the scientific community has always been international and collaborative.”

**David Adam is a freelance journalist based near London. Follow him on Twitter [@davidneiladam](#)**

**A version of this article was originally published at [Knowable Magazine](#) and has been republished here with permission. Newsletter [signup](#).**