

GENOMICS ENGLAND AND THE 100,000 GENOMES PROJECT

Genomics England, with the consent of participants and the support of the public, is creating a lasting legacy for patients, the NHS and the UK economy, through the sequencing of 100,000 genomes.

In April 2003 one of the most significant scientific breakthroughs of modern times was announced. After years of painstaking research carried out by thousands of dedicated scientists across the world, the complete genetic code of a human being – their genome – could now be published.

The Human Genome Project, as this work was known, was the largest international collaboration ever undertaken in biology with British scientists leading the global race to read the human genome, which is made of DNA, letter by letter, a technique called sequencing. The UK has often led the world in scientific breakthroughs and DNA was no exception. Crick and Watson won the Nobel Prize for discovering the double helix structure of DNA and it was a British double Nobel Prize winning scientist, Fred Sanger, who discovered how to sequence it.

Now there is a real opportunity to turn the very important scientific discoveries about DNA and the way it works into a potentially life-saving reality for NHS patients across the country.

What is genomics?

Most of us have heard of genetics, the study of the way particular features or diseases are inherited through genes passed down from one generation to the next. But the more we learn about genes, the more we understand that the old idea of having a single gene for this, or a single gene for that, which determines your fate is not – except in the case of unusual inherited diseases - a good way of describing the complexity of genes. In fact, groups of genes work together and their activity is influenced by a huge variety of environmental and other factors. And we now know that the DNA between your genes is also very important.

You have a complete set of genes in almost every healthy cell in your body. One set of all these genes, (plus the DNA between the genes), is called a genome. Genomics is the study of the whole genome and how it works but has also come to have a broader meaning to include the way that the genome is interpreted and the technologies that have been developed to help do this.

Why this focus on genomics now?

When the first draft of the whole human genome was announced it was claimed that it would revolutionise medical treatment. It had taken 13 years and over £2 billion to laboriously read every letter of the human genetic code. It took such a long time because the DNA sequence of humans is very long – 3 billion letters – and because the sequencing machines available at the time were so slow and laborious. Now a human genome can be sequenced in a few days for less than £1000. It's the leap in the speed and cost of technology that has opened up the potential of genomics and brought it within reach of mainstream healthcare.

But haven't we already got a good understanding of genetics? One of the great surprises from the Human Genome Project was that there were only about 20,000 genes– about the same number as a

starfish. The role of the remainder of a human's genome – in fact a staggering 95 percent of it - was a mystery. Now we know that the remaining DNA is not irrelevant as was once thought but that much of it has a critically important role, influencing, regulating and controlling the rest. That's why it's necessary to sequence the whole human genome (rather than just looking at the 20,000 genes currently used for diagnosis in medicine) if we are to really understand the role of genes in health and disease.

But people are very different, so studying only a small number of genomes would not be enough to give doctors and scientists a true picture of our genes and their relationship to disease. Another key point is that by itself, a genome can't tell you very much. To make sense of it, it is essential to know much more about the person who donated it; details like their symptoms and when they first started, along with physiological measurements, such as heart rate or blood pressure (this sort of information is provided by clinicians and called phenotypic data). Another set of information which may be important in interpreting genomic data comes from their past medical records and would include such things as previous illnesses, medications and birth weight.

And this is where the NHS comes in. The way in which the NHS is able to link a whole lifetime of medical records with a person's genome data and the fact it can do this on a large scale is unique. The richness of this data can help to understand disease and to tease apart the complex relationship between our genes, what happens to us in our lives and illness.

So what can genomics do? You can use it to predict how well a person will respond to a treatment or find one that will work best for them – so called personalised medicine. A good example in use already is whether or not a woman's breast cancer is HER2 positive. If it is, Herceptin will be very effective for her but not for someone who doesn't have HER2. You can also use genomics to test how well a cancer might respond to radiotherapy. For some that can mean far fewer radiotherapy sessions. Or use it to find the 30,000 people who currently use insulin for their Type 1 diabetes but would do better on simple tablets. Genomics can be used to track infectious disease, precisely pinpointing the source and nature of the outbreak through looking at the whole genomes of bugs. The potential of genomics is huge, leading to more precise diagnostics for earlier diagnosis, new medical devices, faster clinical trials, new drugs and treatments and potentially, in time, new cures.

The supersonic age of genomics has begun. And just as the NHS has been at the forefront of scientific breakthroughs before, we want the NHS to be at the forefront again, with its patients benefiting from all that genomics offers, becoming the first mainstream health service in the world to offer genomic medicine as part of routine care for NHS patients.

100,000 Genomes Project

To bring the predicted benefits of genomics to NHS patients is why the Prime Minister launched the 100,000 Genomes Project in late 2012.

Genomics England, a company wholly owned and funded by the Department of Health, was set up to deliver this flagship project which will sequence 100,000 whole genomes from NHS patients by 2017. Its four main aims are; to create an ethical and transparent programme based on consent; to bring benefit to patients and set up a genomic medicine service for the NHS; to enable new scientific discovery and medical insights; and to kick start the development of a UK genomics industry.

The project will focus on patients with a rare disease and their families and patients with cancer. The first samples for sequencing are being taken from patients living in England with discussions taking place with Scotland, Wales and Northern Ireland about potential future involvement.

Why cancer and rare diseases?

In the UK, just fewer than 160,000 people died from cancer in 2011 with over 330,000 new cases reported every year. Because cancer is more likely to occur as people age, we expect the number of cancer cases to rise as people live longer. And although rare diseases are individually very uncommon, because there are between 5000 and 8000 of them, a surprisingly large number of people are affected in total – 3 million – or, put another way one in 17 (or between 6 and 7 percent) of the UK population. Genomics has great potential for both because both rare disease and cancer are strongly linked to changes in the genome. Cancer begins because of changes in genes within what was a normal cell. Although a cancer starts with the same DNA as the patient, it develops mutations or changes which enable the tumour to grow and spread. By taking DNA from the tumour and DNA from the patient's normal cells and comparing them, the precise changes are detected. Knowing and understanding them strongly indicates which treatments will be the most effective. Genomics has already started to guide and inform doctors about the best treatment for individual patients. We've already mentioned Herceptin for HER2 positive breast cancer but we are only at the beginning. Many more cancer types, including those for whom there is hardly any successful current treatments such as lung cancer could be helped if only we knew which gene changes were important.

At least 80 percent of rare diseases are genomic with half of new cases found in children. Knowledge of the whole genome sequence may identify the cause of some rare diseases and help point the way to new treatments for these devastating conditions – vital progress given that some rare diseases take two or more years just to identify. As most rare diseases are inherited, the genomes of the affected individual (usually a child) plus two of their closest blood relatives will be included to pinpoint the cause of the condition.

In all, it is anticipated that about 75,000 people will be involved of which 40,000 will be patients with serious illness. The numbers add up like this: 50,000 genomes from cancer – two per patient, therefore 25,000 patients. 50,000 from rare disease – three per patient (affected person plus two blood relatives) – therefore roughly 17,000 patients, 33,000 others. In all, just over 40,000 patients, and about 75,000 people involved in total. There has already been an extraordinary response by patients and their families wanting to take part in the Genomics England pilot.

Some patients involved in the 100,000 Genomes Project have already benefitted, because a better treatment is identified for them or their condition is diagnosed for the first time. See our news section for more details. However, for most, the benefit will be in knowing that they will be helping people like them in the future through research on the genome data they generously allow to be studied but all will

know that because of their involvement, an infrastructure will be developed which, in the future will enable the NHS to offer genomic services much more widely, to any patient who might benefit.

THE CHALLENGES

To make genomics a reality for the NHS it has to be of high quality, fast and affordable with results that are readily understood. How can this be done?

The sequencing challenge

Genomics England has invested in the latest, state of the art sequencing machines to sequence the 100,000 genomes in the project. This is the first time sequencing has been done on such a scale in the UK. At first, all results will be double checked using existing clinical testing. This is to ensure that the information it delivers is of such high quality that doctors are confident to use it in making major decisions about care.

The data challenge

The first step after sequencing is to compare the possibly millions of differences between the patient's genome and a reference genome, a process called variant calling. The next hurdle - annotation - is to interpret the meaning and importance of those differences which are important. Some of the differences will just be natural harmless variations between individuals, but some will be damaging and almost certainly involved in the development of disease. In truth, much of the genome is still a mystery, requiring an immense amount of work to understand.

The raw data from one genome is about 200GB which would occupy most of the average laptop's hard drive. Just the annotations would easily fill a DVD by themselves. This mountain of data needs to be sifted, analysed and presented in a way that is helpful to doctors, most of whom will not have specialist knowledge of gene changes. Genomics England is investing in the people, expertise and technology to undertake this work.

The security challenge

The genome data is large in size and also precious and will be stored securely and respectfully with rigorous conditions for access which the public can have confidence in. Each one of these challenges involves science at the cutting edge in a field that is moving very rapidly, and at a scale never seen before. Genomics England is having to be very flexible, changing its plans frequently to reflect new advances but also being humble enough to learn from things that don't go right, especially in the pilot stage. There is much to learn from patients, from clinicians and Genomics England will be working with them to develop protocols that are robust, practical and efficient.

Delivering benefit to patients

An additional and critically important spin off is the importance of this huge amount of data to researchers. This includes those wanting to understand more about the genome but also to those wanting to develop new treatments, diagnostics, devices and medicines including academics and those in life science industries which involve not just well known pharmaceutical and biotechnology companies but also a great number of innovative small and medium enterprises (SMEs).

WHAT ARE THE CONCERNS?

Some people feel companies should not benefit commercially from patients who have donated their genome data without receiving any payment. Or that participant's data might not be secure and that they could be identified if they take part, or their data used by researchers in a way that is not fair.

Ethical issues

The 100,000 Genomes Project has a high ethical importance as without it, the NHS won't be able to get the genomics service that patients deserve. High standards of ethical practice must underpin a future service and ethics has to have a major contribution in its development. Genomics England has its own independent Ethics Advisory Committee which advises the Genomics England board on the ethical aspect of everything Genomics England does. Issues already scrutinised include what information patients should receive about their results as well as policies on consent. A series of engagement and involvement activities with patients, clinicians and other groups about these issues has been undertaken. The outputs of these discussions is available [here](#).

Commercialisation and who benefits

Patients donate their samples and information using models of informed consent which have been approved by an independent NHS ethics committee ([see the approved protocol here](#)). Patients are explicitly asked if they are willing for commercial companies to be able to conduct approved research on their data. Those people that have already generously consented to take part understand the challenges about sharing data in their own case but they are keen to see their data used to help progress research into the condition that affects them. If innovative treatments are to be found to extend or save lives then commercial companies will need to invest in the research, development and manufacture of new drugs and diagnostic tests. It has always been the case that this work is carried out in the commercial sector and not by government or within the NHS itself.

Genomics England is developing ways of charging for its data services to ensure that the costs of maintaining the data are shared with companies and that the UK tax payer will benefit should companies successfully develop drugs, devices, treatments, diagnostic tests or other services through its use. If successful products are developed, it means that patients are benefiting.

Privacy and confidentiality issues

Any relevant information about a patient will be returned to their doctor. For other medical researchers and companies to gain access to Genomics England's data services they will have to first pass a rigorous ethical review and have their research proposal approved under policies being devised by Genomics England's Ethics Advisory Committee. Insurers and marketing companies will not be allowed access. Oversight by the Genomics England Data Advisory Committee will ensure that any researchers wanting access to data will go through rigorous identity checks and their use of the data will be closely supervised. No raw genome data can be taken away. The data will be kept within Genomics England's data structures and will be constantly under its control. Genomics England commits itself to constant testing and re-testing of its security systems to ensure data safety.

While Genomics England has the data, patient identifiers (such as NHS number or postcode are removed) to reduce the risk of re-identification of clinical and genomic information with a particular individual. Only when data is used for a patient's own care will identifiable data be made available to the patient's doctor and medical team.

Patients are told that participant anonymity cannot be absolutely guaranteed as in theory, any non-trivial piece of health records data can be re-identified by someone who already has access to sufficiently detailed information about an individual. In practice, this is very hard to do and harder still to achieve undetected. Genomics England can't promise that no researcher would be able to do this but what it can promise is that it will be made so difficult that there would be far easier ways to achieve the same goal.

Genomics England is talking constantly to patients about their concerns to make sure that any issues they may have are addressed at an early stage. Patients have been involved from the outset and are at the very heart of this project. In particular, the commitment to consent is of paramount importance.

KICK STARTING A UK GENOMICS INDUSTRY

It is not just patients and the NHS that stand to benefit from the 100,000 Genomes Project. There will be numerous knock-on advantages for the country. An example from the past might be the introduction of the railways in the Victorian era. . Individuals and families benefited from cheap travel but the infrastructure created by the new railways also triggered an economic boom. Whilst the growth of some companies say, those making railway tracks, was predicted other economic benefits were not. For instance, there was a boom in holidays resorts, the sale of postcards and travel guides.

The 100,000 Genomes Project has some parallels. Whilst primarily for the benefit of people who are sick, there are potentially many economic benefits for the nation. We can be certain of benefits such as new medicines and diagnostic tests but just as with railways, some of the companies that may develop will be unexpected, built on new, as yet undiscovered technologies that will emerge over the next five years.

The 100,000 Genomes Project cannot be guaranteed to succeed, in the same way that there was no guarantee for the railways. So only the government is willing to take the risk and make the necessary investment in it. And just as Victorian England with its great engineers was the perfect place for the birth of the railways, the UK, which not only leads the world in life sciences but has the unique benefit of the NHS, is the best place in the world to initiate the practical use of genome sequencing and interpretation for patient benefit. Our vision was one where England is the leader in a new industry where genomics is used to help patients get better, more personalised care and treatment.

Delivering the project

When the project ends, the NHS will need to be ready to use genomics as part of its routine care, so it is vital that more scientists, geneticists and doctors are trained to interpret the data and understand what it means for a patient's medical condition. In parallel with Genomics England's work, a skills and training programme for workers in the NHS is currently being set up by the organisation responsible for doing this – Health Education England.

The 100,000 Genomes Project will use the generosity of patients and the outstanding skills and talent found in the medical and the life sciences' sectors in the UK to help deliver this project. Genomic England's legacy will be a genomics service ready for adoption by the NHS, high ethical standards and public support for genomics, new medicines, treatments and diagnostics and a country which hosts the world's leading genomic companies. It is a bold ambition with benefits for all.