HOUSES OF PARLIAMENT
PARLIAMENTARY OFFICE OF SCIENCE & TECHNOLOGY

POSTNOTE

Number 504 September 2015

The 100,000 Genomes Project

In late 2012, the Prime Minister announced plans to sequence 100,000 genomes.[1] The project aims to establish a genomic medicine service within the NHS and support the Government Strategy for UK Life Sciences.[2,3] This POSTnote provides an introduction to genomic medicine and an overview of the aims, structure and major challenges of the 100,000 Genomes Project.

Overview

- A genome is the complete set of DNA found within a cell. Genomic medicine uses genome data to support clinical treatment.
- The 100,000 Genomes Project aims to sequence 100,000 genomes through NHS England by 2017, with a focus on rare diseases, cancers and infectious diseases.
- The project aims to incorporate genomic medicine into the NHS to benefit patients. NHS clinicians will collaborate with academia and industry to interpret clinical and genome data and aid the long-term development of new medicines.
- Ethical considerations include governance of data access for research purposes and achieving the valid consent of participants.
- It is not yet clear how best to deliver whole genome sequencing within the NHS when the project ends in 2017.

The Human Genome

Biological information within short sections of DNA (genes) hold the instructions for producing proteins, which are the building blocks of all organisms. Within a human cell there are roughly 20,000 genes, but these only account for around 1% of the DNA present in each cell.[4] The remaining 99% of the human DNA sequence does not have a known function but is increasingly the subject of wider study.[5] The entirety of this DNA is known as the 'human genome'.

Until recently genetic studies focused on disorders such as cystic fibrosis, which are caused by a single gene. Tests for such disorders are routinely used in the NHS.[6] Genomics is an extension of genetics and focuses on all of the DNA present in a cell, both in genes and the non-protein coding regions. Determining the exact composition of the DNA within a gene or genome relies on DNA sequencing. The Human Genome Project, completed in 2003, took 13 years to sequence fully the human genome for the first time and cost nearly £2bn ($2.7bn).[4,7] Improvements in technology and computing have vastly reduced human genome costs (~$1,000 or ~£650) and time (days) and made sequencing large numbers of genomes achievable.[8,9]

What is Genomic Medicine?

Completion of the Human Genome Project brought increased interest in the potential use of genome information to inform healthcare (Box 1). Genomic medicine aims to use individuals' DNA sequence data to support their clinical diagnosis and treatment. In the longer-term, sequence data may help to identify new therapeutic targets and inform disease risk prediction.[10,11,12] Having a patient's genome information raises the possibility of personalised medicine, which tailors treatment to an individual.[10] The UK is not the only country interested in this area; in early 2015 the US Precision Medicine Initiative was launched with an aim to realise personalised medicine through the use of clinical research volunteers, genome data and electronic health records.[13] This POSTnote outlines the aims of the 100,000 Genomes Project (the project), its targeted disease areas, implementation and the ethical issues it raises.

Aims of the 100,000 Genomes Project

The project will sequence whole genomes from consenting NHS patients and their family members in England, who will be recruited through NHS centres (see Project Implementation).[2] It aims to develop a framework that will integrate genomic research and clinical treatment, making the NHS a global pioneer in the field of genomic medicine.[2]

Box 1. Converting genetic knowledge into healthcare

- Clinical diagnosis – genomic technologies have proved useful in determining a diagnosis in cases where this has not been possible using clinical symptoms or other techniques.[14] In a study of over 1,000 children with previously investigated yet undiagnosed developmental conditions, the UK-based Deciphering Developmental Disorders programme used gene sequencing and reported a diagnostic rate of 27%.[15]
- Pharmacogenomics – genetic tests can be used to detect a patient's specific genetic mutations to inform the choice of drug treatment. For example, around 20% of breast cancer patients produce high levels of a protein, that increases tumour aggression.[16] Directly targeting this protein using an antibody treatment[17] in combination with chemotherapy can reduce mortality rates by one-third.[18] Genetic tests can also be used to identify those patients that are likely to have an adverse reaction to a drug treatment.[12] For instance, they can identify patients for whom the drug Abacavir, a drug used to treat HIV, is highly toxic.[19]
- Developing new diagnostics and medicines – identification of genes and their role in the clinical presentation of a disease can give new potential diagnostic and drug discovery targets. However, functional research of these pathways and the development of new drugs is a lengthy process and can take years to reach the clinical trial stage.[20] An alternative to this is the repurposing of already licensed drugs to treat a different condition, which can be supported by genomic approaches.[21,22,23,24]

To deliver the project by the end of 2017, the Department of Health (DH) has set up a fully government-owned private company, Genomics England, to oversee the project and manage the sequencing and data centres.[2] Over £300 million in funding has been committed by various partners.[25] In addition to sequencing 100,000 genomes and producing a dataset that pairs genomes with medical data by 2017, other major goals are to:

- **Benefit NHS patients.** Among possible benefits to patients are improved diagnosis of their condition (examples of this have already been announced)[26] and better matching of patients to treatments (pharmacogenomics, see Box 1).
- **Encourage scientific discovery and medical insights.** Researchers will be able to access the datasets under monitored conditions (Box 2). Collaboration with these groups is hoped to spark new research opportunities and maximise the potential future benefits for NHS patients.[2]
- **Stimulate development of a UK genomics industry.** Allowing industrial access to the collated datasets could make the UK attractive to the life sciences industry. Ten commercial partners have formed a pre-competitive consortium to access the project's dataset (Box 2) and the sequencing company Illumina will invest a net sum of £84 million over the course of the project.[25]
- **Increase public support for genomic medicine.** The project aims to ensure trust through the creation of an ethical programme and engage the public through strategies being developed by Genomics England alongside a range of partners including NHS England.

Disease Areas

The project will focus on three main branches of disease: rare diseases, patients with particular cancers and patients with certain infectious diseases.[2] The relationship between genetics and disease is complicated (Box 3) and varies between diseases. Participants are told during recruitment that only a small number are likely to clinically benefit directly during the project.[28] In the long-term, however, it is hoped that the project will stimulate research that supports the development of new treatments for patients.

Box 2. Collaborations with Academic and Commercial Research

Genomics England Clinical Interpretation Partnership (GeCIP)
To enhance the interpretation of patient data, the 100,000 Genomes Project will bring together clinicians involved in NHS Genomic Medicine Centres with academic researchers and trainees. This partnership will work mainly in disease specific groups to foster an environment that aims to:

- support innovative clinical interpretation and validation
- develop new computational methods for analysis
- deliver research on implicated disease pathways
- assess the repurposing of licensed drugs to treat other diseases
- train clinicians and researchers in genomics.[2]

Genomics Expert Network for Enterprises (GENE) Consortium
Engagement with pharmaceutical companies is needed to develop any new medicines. In March 2015, Genomics England launched a consortium consisting of ten industrial partners, including AstraZeneca and GlaxoSmithKline. These partners will collaborate with each other to find the most effective way to translate data from the project into diagnostics and treatments. Partners have invested either £250,000 or £25,000 depending on the size of the company.[27]

Rare Diseases

A rare disease is one that affects less than 1 in 2,000 people. However, the large number of rare genetic diseases – an estimated 7,000 different conditions – means that collectively they affect ~3.5 million people in the UK.[29] Around half of these rare genetic conditions, have a known gene implicated in their diagnosis.[21] The rare disease branch will sequence up to 50,000 participants[2] across more than 120 diseases, selected on the basis of the need for better clinical treatments.[30] For each patient included in the study, two family members will be sequenced to help identify the disease-causing mutation.[2]

Cancers

The uncontrolled cell growth that is characteristic of cancers arises from novel genetic mutations during a person's lifetime (POSTnote 406).[31,32] Within the cancer branch of the project, up to 50,000 genomes will be sequenced from NHS patients. However, this does not equate to 50,000 people, as for each individual the genome from a non-cancerous cell is sequenced for comparison with the cancer genome. Some people may be sequenced multiple times to follow the genetic evolution of their cancer.[2] The aim is to identify mutations driving the emergence of cancer and increase understanding of individual drug treatment responses, which could support personalised medicine approaches. Moreover, the identification of new driver mutations could offer new treatment targets.[2]

Infectious Diseases

An as yet unspecified number of people infected by HIV, hepatitis C or tuberculosis will be studied. Participants within

Box 3. Genetics and disease

Certain genetic mutations can influence the likelihood of a person developing a specific disease. Where a mutation is highly likely to lead to development of a disease, the mutation is described as highly penetrant (POSTnote 445). Such mutations are the causes of many rare conditions such as cystic fibrosis.[33] However, only a small number of cancer-associated mutations are highly penetrant. One example are the mutations in the *BRCA1* and *BRCA2* genes that are estimated to have a cumulative risk of developing breast cancer by the age of 70 of 57% and 49% respectively.[34]

Many common diseases such as cancers, heart disease and diabetes are described as multi-factorial, as they are influenced by a variety of genetic and environmental factors (POSTnote 451). Numerous studies, known as genome-wide association studies, have performed genome analysis on cancer patients to identify associated predisposing mutations with their cancer type.[12] Many of the individual mutations identified have been estimated to increase only slightly the risk of susceptibility. These can be minor risk factors when compared against other environmental traits such as smoker status.[35,36]

this branch of the project, which is being led by Public Health England, will have the pathogen they are infected with sequenced. This will provide data on local transmission networks and potentially influence treatment through the identification of drug resistant infections.[2,37] Patients who exhibit a severe response to infection could also have their own genome sequenced for analysis by a clinical interpretation partnership (Box 2).

Project Implementation

The project will recruit consenting participants through specialist NHS Genomic Medicine Centres and perform the DNA sequencing and analysis within secure data centres. Controlled access to de-identified data will be available to permitted academic researchers and commercial groups. The eventual aim is to sequence a participant within 2 weeks of enrolment, so an efficient system will be needed.[2] This includes designating new NHS Genomic Medicine Centres, creating sequencing and data storage capacity and supporting interpretation.

NHS Genomic Medicine Centres

To recruit participants for the project, eleven NHS Genomic Medicine Centres (GMCs) have so far been selected by NHS England through a competitive process. These centres are locally administered and have committed to deliver a quota of samples. The GMCs will be the interface for participant interaction through the course of the project.[2] They will obtain the consent of potential participants, collect samples and clinical data, conduct clinical validation of diagnoses and feedback findings to participants. Additional training courses have been set up (Box 4) to train staff in clinical genetics. A challenge that has already arisen is DNA extraction from tumour biopsy samples, as the current preservation method is not designed for DNA sequencing. Within the projects cancer pilot, high quality DNA was obtained from only 50% of such samples.[38] Genomics England is currently optimising new methods of preserving samples to increase the success rate, which could improve the quality of NHS samples collected in the future.

Box 4. Training the NHS workforce

Surveys of NHS cancer genetics centres have identified workforce education as a key challenge in offering more genetic testing.[39] To support this, Health Education England has set up online learning resources and a fully-funded Master's course in Genomic Medicine for 550 NHS clinical staff, to begin in September 2015.[40] This course aims to give GMC staff training in genomics and counselling to provide support for patients. However, attendance for staff based in GMCs is not mandatory.

Sequencing and Genomics England Data Centre

Genomics England selected Illumina through a competitive tender to perform the DNA sequencing. Alongside Illumina and the Wellcome Trust Sanger Institute, Genomics England is developing an NHS Genomic Medicine Sequencing Centre where the genome sequencing will take place. This has been supported by £27 million from the Wellcome Trust.[2,25] Data from the Sequencing Centre will be transferred to the Genomics England Data Centre, which has received £24 million in Medical Research Council funding.[2,25] Each human genome sequence is around 100 gigabytes of data, with a total of roughly 15 petabytes (15 million gigabytes) of overall memory required.[41] Within the Data Centre, all sequence and clinical data available to researchers will be de-identified (Box 5). External data access will be managed through an approval process, with any breaches to conditions potentially leading to exclusion of access and legal proceedings.[2]

Clinical Interpretation and Data Analysis

Data interpretation in the project will be performed iteratively. The statistical power to detect, for instance, any new mutations that may cause a rare disease or drive cancerous growth, will increase as more participants are recruited. Genomics England has estimated the genome numbers required to identify such mutations. For rare diseases caused by a single mutation, it is possible that only small numbers (~10) will be required.[2,42] For genetically complex cancers where many genes are involved, detecting newly implicated genes could require over 5,000 patients.[2,43] The detailed clinical information provided by the GMCs will be essential in getting the most of these studies. Defined clinical characteristic terms will be used to provide standardisation across the different centres.[2]

Ethical Considerations

Genomics England has built an independent Ethics Advisory Committee into the project and its working protocol has been approved by an NRES Research Ethics Committee.[2,44] Key ethical issues raised by the project include participant consent, data access and the feedback of findings.

Participant Consent

The project is using a broad consent process with patient information sheets and consent forms developed through feedback by patient group representatives and their families under the guidance of the Ethics Advisory Committee.[2] To provide informed consent, participants need to understand the project and the nature of potential uses of their data. During recruitment, potential participants undergo a

Box 5. De-identification of data
The project will collect personal identifiers such as name and address, but will not make these available to researchers within the Genomics England Data Centre. Instead, participant data will be replaced by temporary codes (POSTnote 474) with no external meaning. However, this can still carry a risk of identification, for example if a patient has a particularly rare disease and can be linked to a particular treatment centre.[45] Research activity on the de-identified data is monitored and participant data cannot be taken out of the secure research environment: only aggregate results or data assessed as anonymous can be exported.

discussion with a clinician where consent is sought. This process is designed to highlight the key issues within the project that affect participants including the re-identification risk (Box 5), uncertainty over any personal clinical benefit and the long-term access to de-identified data by approved researchers.[28]

If proposed amendments to Articles 81 and 83 of the EU Data Protection Regulation (POSTnote 473) to protect research participants are implemented, the project's consent model could be affected, as the Regulation would require explicit consent for each new use of patient data.[46] This would also affect many other health research studies that use personal datasets. Repeatedly contacting participants to obtain consent for each study is a concern as it may discourage participation and impede research through increased administration.[47]

Data Access and Usage

Consultations suggest that patients are supportive of research on their own genome data and health records, but want to be reassured that data access will be subject to good governance in the public interest.[48] There are restrictions on the purposes for which data can be accessed by external researchers and an independent advisory panel assesses all applications for access. Researchers are not allowed to remove the raw sequence or personal data form the secure data environment. A common public concern is whether data use might lead to stigmatisation or discrimination, for example, by affecting insurance premiums.[49] Any genetic finding from the project is covered by a pre-existing agreement between the Association of British Insurers and the Government that insurers do not require results from clinical research.[50] This agreement is in place until 2019. However, it does not cover patients who subsequently receive treatment on the basis of a genetic diagnosis; such patients are required to inform insurers if the information is requested.[2]

Feedback of Findings

Three main types of genetic finding could potentially be uncovered by participation within the project.[2,51] Genomics England operates a specific policy for each.

- **Main findings** are genetic findings that are relevant to the disease for which the participant was recruited into the project. All main findings are provided to the clinician to discuss with the participant. Participants must agree to receive main findings to be accepted into the project.
- **Additional findings** refer to 10 additional conditions such as the *BRCA1* and *BRCA2* mutations for cancer, for which there is a known mutation that can inform healthcare and for which NHS treatment is available. To receive such findings from their clinician, patients must opt in during the consent process. Conditions may be added to this list throughout the project as new evidence emerges. Participants who are pregnant or planning a pregnancy may also be offered the option of receiving information from tests to see whether they carry a mutation for a small number of rare inherited diseases such as cystic fibrosis, which could affect their baby.
- **Incidental findings** are findings that do not relate to the main condition and are not included within the additional findings. There is debate over whether such findings should be fed back to participants, particularly where they relate to conditions like Huntington's disease that are not treatable. A survey carried out for the project noted that more than half of respondents desired the return of incidental findings.[49] However, participants in the project are told during the consent process that such findings will generally not be fed back to them unless there is evidence that healthcare could be influenced. If, say, a mutation for Huntington's disease was uncovered a group within the Genomics England Clinical Interpretation Partnership (Box 2) would review whether it should be fed back. No mutations that only slightly increase propensity to a disease will be fed back, but the de-identified data will be available to external researchers.

Genomic Medicine in the NHS Beyond 2017

Genomics England is working with a range of stakeholders including NHS Genomic Medicine Centres, funding bodies, research institutes, patient support groups and industry to aid engagement.[52] The controversy over the care.data project (POSTnote 474) illustrates that public engagement is required to raise awareness about genomics, build confidence in the project and support a societal discussion on the ethical future use of genomic medicine in the NHS.

The sequencing of participants will finish by the end of 2017, but the dataset will continue to be available to researchers. Further genome sequencing beyond 2017 would need to be commissioned via the NHS or require new public or private investment where the NHS could not justify sequencing as cost effective. While in the short-term, participants in the project are unlikely to receive information that will directly inform their healthcare, the project could deliver:

- cases of direct patient benefit
- new structures for conducting clinical research in the NHS
- increased public confidence in genomic medicine
- high quality academic publications
- new targets for drug discovery
- commercial interest in funding NHS clinical research.

Endnotes

1 DNA tests to revolutionise fight against cancer and help 100,000 NHS patients www.gov.uk/government/news/dna-tests-to-revolutionise-fight-against-cancer-and-help-100000-nhs-patients
2 Caulfield M et al, *The 100,000 Genomes Project Protocol*, 2015 www.genomicsengland.co.uk/library-and-resources/

POST is an office of both Houses of Parliament, charged with providing independent and balanced analysis of policy issues that have a basis in science and technology. POST is grateful to Lee Murray for researching this briefing, to BBSRC for funding his parliamentary fellowship, and to all contributors and reviewers. For further information on this subject, please contact the co-author, Pete Border.

3 *Strategy for UK Life Sciences*, BIS 2011
4 IHGSC, *Nature*, 431, 931-45 2004
5 The ENCODE Project Consortium, *Nature*, 489, 57-75, 2012
6 Promoting Gene Testing, Fourth Report of the UKGTN, 2014, http://ukgtn.nhs.uk/fileadmin/uploads/ukgtn/Documents/Resources/Library/Reports_Guidelines/UKGTN%204th%20Biennial%20Report%202014.pdf
7 National Human Genome Research Institute, www.genome.gov/11006943
8 National Human Genome Research Institute, www.genome.gov/sequencingcosts/
9 Gudbjartsson DF et al, *Nature Genetics*, 47, 435-44, 2015
10 House of Lords Science and Technology Committee, *Genomic Medicine* Volume I: Report, 2009
11 Bell J, *Nature*, 429, 453-456, 2004
12 McCarthy JJ et al, *Science Translational Medicine*, 5, 189sr4, 2013
13 Collins FS and Varmus H, *NEJM*, 372, 793-795, 2015
14 Martin HC et al, *Human Molecular Genetics*, 23, 2014
15 Wright CF et al, *The Lancet*, 385, 1305-1314, 2015
16 Owens MA et al, *Clinical Breast Cancer*, 5, 63-69, 2004
17 Slamon DJ et al, *NEJM*, 344, 783-792, 2001
18 Moja L et al, *Cochrane Database of Systematic Reviews*, CD006243, 2012
19 Mallal S et al, *NEJM*, 358, 568-579, 2008
20 Chin et al, *Nature Medicine*, 17, 297-303, 2011
21 Boycott KM et al, *Nature Genetics*, 14, 681-691, 2013
22 Sanseau P et al, *Nature Biotechnology*, 317-320, 2012.
23 Rilstone JJ et al, *NEJM*, 368, 543-550, 2013
24 Dudley JT et al, *Science Translational Medicine,* 3, 96ra76, 2011
25 www.genomicsengland.co.uk/uk-to-become-world-number-one-in-dna-testing-with-plan-to-revolutionise-fight-against-cancer-and-rare-diseases/
26 www.genomicsengland.co.uk/first-patients-diagnosed-through-the-100000-genomes-project/
27 *A framework for industry engagement*, Genomics England, 2015
28 *Patient Information Sheet: For adults with cancer,* Genomics England, 2015
29 Rare Disease UK www.raredisease.org.uk/about-rare-diseases.htm
30 Genomics England Rare Disease List, see http://www.genomicsengland.co.uk/library-and-resources/
31 Vogelstein B et al, *Science*, 339, 1546-1558, 2013
32 Stratton MR et al, *Nature*, 458, 719-724, 2009
33 WHO Factsheet see www.who.int/genomics/public/geneticdiseases/en/index2.html
34 Chen S & Parmigiani G, *Journal of Clinical Oncology*, 25, 1329-1333, 2007
35 Burger M et al. *European Urology*, 63, 234-241, 2013.
36 Yang IA et al, *Journal of Thoracic Diseases*, 5, S454-S462, 2013.
37 Peacock S et al, *Strategic Priorities in Infectious Diseases*. See www.genomicsengland.co.uk/library-and-resources/
38 100,000 Genomes Project Update see www.genomicsengland.co.uk/100000-genomes-project-update/
39 Slade et al, *Genome Medicine*, 7:18, 2015.
40 Genomics Education Programme, NHS Health Education England, 2015
41 Genomics England Data Working Group Report see www.genomicsengland.co.uk/library-and-resources/
42 Ng SB et al, *Nature Genetics*, 42, 790-793, 2010
43 Lawrence MS et al, *Nature*, 505, 495-501. 2014
44 Parker M, Ethics Working Group, Genomics England see www.genomicsengland.co.uk/library-and-resources/
45 *Anonymisation: managing data protection risk code of practice*, ICO, 2012 https://ico.org.uk/media/1061/anonymisation-code.pdf
46 *Position of non-commercial research organisations and academics, Protecting health and scientific research in the Data Protection Regulations*, (2012/0011 (COD)), May 2015.
47 McGuire AL & Beskow LM, *Annu Rev Genom Hum Genet*, 22, 361-381, 2010.
48 Simpson A & Hunter A. Genetic Alliance UK, *Research Findings for the 100,000 Genomes Project*, 2014
49 Hazleton A & Petchey L, Genetic Alliance UK, *Genome sequencing: What do patients think?* Patient Charter, 2015
50 HM Government, Association of British Insurers. Concordat and Moratorium on Genetic and Insurance. 2011 www.gov.uk/government/uploads/system/uploads/attachment_data/file/216821/Concordat-and-Moratorium-on-Genetics-and-Insurance-20111.pdf
51 MRC/Wellcome *Trust Framework on the feedback of health-related findings in research*, March 2014
52 Genomics England Engagement Strategy see http://www.genomicsengland.co.uk/library-and-resources/