Institution: Queen Mary University of London		
Unit of Assessment: 1		
Title of case study: The 100,000 Genomes Project: Developing New National Infrastructure for Whole-Genome Sequencing		
Period when the underpinning research was undertaken: 2013 - 2019		
Details of staff conducting the underpinning research from the submitting unit:		
Name(s):	**Role(s) (e.g. job title):**	**Period(s) employed by submitting HEI:**
1) Sir Mark Caulfield	1) Co-Director of Queen Mary's William Harvey Research Institute (WHRI), Chief Scientist for the 100,000 Genomes Project	1) 05/1989 - present
2) Clare Turnbull	2) Professor of Medical Genomics, Clinical Lead for Cancer Genomics for the 100,000 Genomes Project	2) 12/2014 - 03/2020
3) Damian Smedley	3) Reader, Lead for Genomic Interpretation for the 100,000 Genomes Project	3) 01/2016 - present
Period when the claimed impact occurred: 01/2015 - 2020		
Is this case study continued from a case study submitted in 2014? N		

1. Summary of the impact (indicative maximum 100 words)

Prof. Sir Mark Caulfield at Queen Mary has led the delivery of Genomics England's 100,000 Genomes Project as Chief Scientist since July 2013. The project is now internationally recognised as an exemplar national sequencing project at the forefront of genomics innovation, and has transformed the UK's genetic sequencing landscape by:

- Returning genome sequencing results to more than 74,000 NHS rare disease patients, with up to 25% of these receiving a new, and potentially life-changing, diagnosis
- Completing over 900 reports for cancer patients, with approximately 60% of these receiving a finding that could influence their choice of treatment
- Establishing new national infrastructure in the UK for standardised gene sequencing
- Improving NHS services by creating a network of more than 100 hospitals involved in delivering standardised and coordinated whole-genome sequencing
- Positioning the UK as a leader in genomics clinical care and diagnostics. The NHS is now the first national healthcare system in the world to offer whole-genome sequencing in routine care as part of the new UK genomics strategy, which drew on Caulfield's work as part of the 100,000 Genomes Project.

2. Underpinning research (indicative maximum 500 words)

Many diseases are caused by faults in genes. Although research has rapidly advanced in past decades, many people with rare diseases still remain undiagnosed or must undergo long and expensive diagnosis journeys. Additionally, most people with cancer do not receive precision treatments tailored to their specific condition. New gene sequencing technologies could address such issues, and hold the clinical potential to identify the genes associated with some rare diseases and cancers. However, the requirements for making this technology a routine diagnostic approach within a national healthcare system had remained unclear — a key issue being defining the patient sample size needed to successfully detect disease-causing genes.

Due to his work as part of the Wellcome Trust Case Control Consortium and international blood pressure consortia, Queen Mary's Prof. Caulfield is a noted expert in the field of genome-wide association studies for complex traits, having been a corresponding author and primary investigator on several consortia-led *Nature* papers, including key papers in hypertension and blood pressure genetics [3.1–3.3]. Caulfield's leadership in his field includes early recognition [3.1] that detailed patient clinical characterisation and extensive quality assurance of genetic

variant datasets is vital to successful gene discovery in complex traits. Building from the *Nature* 2007 paper where no loci for hypertension were discovered, Caulfield standardised the clinical data using direct blood pressure measurements instead of hypertension, which significantly influenced the discovery and validation of gene loci in both blood pressure and cardiovascular disease risk [3.1, 3.2, 3.3],

Together with Prof. Turnbull, Caulfield has carried out early work on next-generation sequencing (exomes) in breast cancer; used whole-exome sequencing to characterise the tumour landscape of cancer [3.4]; and performed detailed power simulations to identify the sample sizes needed to detect disease-causing genes. This work reinforced the need for high fidelity quality assurance of genomic datasets derived from next-generation sequencing to make discoveries with clinical impact.

Caulfield was seconded to the role of Chief Scientist at Genomics England in 2013 to lead the delivery of the 100,000 Genomes Project and, using his prior work as a basis, immediately standardised clinical data collection using the Human Phenotyping Ontology. This standardisation enabled the project to adopt the software 'Exomiser', developed by Queen Mary's Prof. Smedley, to prioritise genomic variants that identify diagnoses in rare disease patients [3.5, 3.6]. The 100,000 Genomes Project has incorporated Exomiser into its ISO-accredited NHS genome analysis pipeline and as a result, has translated this research into real clinical impact for NHS patients. This work has led to a transformed NHS Genomic Medicine Service, offering 500,000 whole-genome sequences for clinical care over the next 5 years, positioning the UK as a world leader in genomics.

3. References to the research (indicative maximum of six references)

[3.1] Wellcome Trust Case Control Consortium [One of the primary investigators]. (2007). Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. *Nature, 447*, 661-678. https://doi.org/10.1038/nature05911

[3.2] Newton-Cheh, C., Johnson, T., Gateva, V., Tobin, M. D., Bochud, M., Coin, L., (…) Caulfield, M. [A Corresponding Author] & Munroe, P. B. (2009). Genome-wide association study identifies eight loci associated with blood pressure. *Nature Genetics, 41*, 666-676. https://doi.org/10.1038/ng.361

[3.3] The International Consortium for Blood Pressure Genome-Wide Association Studies [Corresponding Author]. (2011). Genetic variants in novel pathways influence blood pressure and cardiovascular disease risk. *Nature, 478*, 103-109. https://doi.org/10.1038/nature10405

[3.4] Ruark, E., Snape, K., Humburg, P., Loveday, C., Bajrami, I., Brough, R., Rodrigues, D. N. (…). (2013). Mosaic *PPM1D* mutations are associated with predisposition to breast and ovarian cancer. *Nature*, *493*, 406-410. https://doi.org/10.1038/nature11725

[3.5] Smedley, D., Jacobsen, J. O., Jäger, M., Köhler, S., Holtgrewe, M., Schubach, M., Siragusa, E., Zemojtel, T., Buske, O. J., Washington, N. L., Bone, W. P., Haendel, M. A., & Robinson, P. N. (2015). Next-generation diagnostics and disease-gene discovery with the Exomiser. *Nature Protocols*, *10* (12), 2004-2015. https://doi.org/10.1038/nprot.2015.124

[3.6] Smedley, D., Schubach, M., Jacobsen, J. O. B., Köhler, S., Zemojtel, T., Spielmann, M., Jäger, M., Hochheiser, H., Washington, N. L., McMurry, J. A., Haendel, M. A., Mungall, C. J., Lewis, S. E., Groza, T., Valentini, G. & Robinson, P.N. (2016). A whole-genome analysis framework for effective identification of pathogenic regulatory variants in mendelian disease. *American Journal of Human Genetics*, *99* (3), 595-606. https://doi.org/10.1016/j.ajhg.2016.07.005

4. Details of the impact (indicative maximum 750 words)

Prof. Caulfield's research on genome-wide association studies for complex traits [3.1–3.4] led him to standardise clinical data for the 100,000 Genomes Project using the Human Phenotyping Ontology [5.1]. This critical work enabled the project team to adopt Exomiser and subsequently improve the diagnostic yield, treatment and care of patients with rare diseases, cancers and infections.

Achievement of uniquely life-changing diagnoses for patients and families in need

By providing a novel precision medicine service, the 100,000 Genomes Project has enabled

more than 74,000 NHS rare disease patients with unmet diagnostic needs to have their genome sequencing results returned to them. 20-25% of these patients have now gained new diagnoses and, therefore, can potentially access new treatments to improve prognosis, mitigating the development of complications, and avoiding the possible negative consequences of their diseases. In cancer patients, the project team has returned over 900 reports, approximately 60% of which achieved a finding that could influence a patient's choice of therapy or offer them a novel clinical trial opportunity.

The project has provided truly life-changing diagnoses for patients and their families, and many children are alive and well in the UK today as a result. For example, the first child to receive a diagnosis from the project had developmental delays and intractable seizures, and had reached her fourth birthday without diagnosis despite extensive testing. This case underscores the diagnostic odyssey experienced by people with rare diseases, as emphasised by the child's mother: "If we had had this done when she was born and found out the results straight away, we would have been on the right track immediately, rather than having years of diagnostic work done" [5.2]. The 100,000 Genomes Project identified a glucose transporter mutation and revealed that the child suffered from low brain glucose – something treatable with a high-fat diet. Her seizures and developmental delay have both improved since diagnosis. The mutation was seen in only her, so her parents, who were planning not to have further children, were reassured of an extremely low likelihood of transmission to a future child, and know they can safely conceive again without high risk of a heritable genetic condition.

One participant, 57, has a life-long history of high blood pressure that has resulted in kidney failure and two transplants. He found out that his kidney failure was caused by a particular genetic variant. As his father, brother and uncle all died of kidney failure and his 34-year-old daughter also has early signs of kidney damage, this discovery has opened up possibilities for early detection, which alleviated many concerns for this family: "I was keen to take part in the project as I felt it was important to try and find out as much as possible about my condition for my daughter and granddaughter. Now that my daughter, Terri, has been given a diagnosis it means that her condition can be monitored every year to see if there are any changes" [5.2]. Not only have father and daughter received a diagnosis but their concern that the 57-year-old's grandchild would also be affected has been assuaged: "The 100,000 Genomes Project is hugely important in gaining as much insight into rare diseases as possible. I wanted to be part of the project for the benefit of my daughter, Katie, and it was such a relief when we found out that she did not have the same rare kidney disease as myself and other family members" [5.2].

Development of new UK infrastructure for genome sequencing

Following its inception in July 2013, the 100,000 Genomes Project has stimulated substantive investment of over GBP210,000,000, including [5.3–5.8]:

GBP162,000,000 from the Medical Research Council, the Wellcome Trust and the NIHR, as well as GBP78,000,000 of investment in the UK by US genomics company Illumina…	…to support the construction of a dedicated NHS Genomic Sequencing Centre in Hinxton
GBP24,000,000 from the Medical Research Council…	…to support a national data centre, architecture, and research environment capable of housing >100,000 genomes, billions of linked data items, and >3,000 researchers from 33 countries working concurrently

GBP20,000,000 from NHS England…	…to support the re-organisation of genetics laboratories and services from >100 hospitals into 13 Genomic Medicine Centres

Improvement of NHS services via standardised sequencing and optimised tumour pathways

The 100,000 Genomes Project has dramatically impacted delivery of NHS clinical care and clinical genetic, molecular genetic and molecular pathology services:

Over 100 hospitals have been brought together to deliver harmonised, standardised, coordinated whole-genome sequencing, with clinical data collected in common formats, and possible rare-disease-associated variants classified and interpreted according to common rules and standards. Molecular tumour boards and multi-disciplinary genomics teams have also been established to facilitate multidisciplinary review and management of genomic test results. Thus, the 100,000 Genomes Project, through Caulfield's leadership, has played a pivotal role in establishing a national infrastructure that now exists for the submission of biological samples and clinical data for whole-genome sequencing.

Transformative pathways in tumour tissue collection have been widely rolled out across the NHS, replacing the centuries-old pathways that were based on formalin-fixed paraffin embedded tissue, which damaged DNA and made it profoundly degrade. These pathways include re-engineering 400 molecular pathology pathways across the NHS; routine refrigeration and vacuum-packing of surgical specimens from theatre; processing of diagnostic biopsies using cryospraying and/or neutral transport media; and the development of new laboratory tissue fixation pathways involving tissue-friendly alcohol-based fixatives (PaxGene). As a result, the NHS molecular pathology service can now routinely process fresh tumour samples across more than 100 laboratories, enabling 17,400 cancer patients to have high fidelity whole-genome sequencing in the 100,000 Genomes Project. These advancements created the basis for an NHS Genomic Medicine Service in cancer.

Creation of a genomics database with linked clinical data

As well as aiding diagnosis and treatment, whole-genome sequencing data from the project are available to researchers in de-identified form, enabling them to derive additional value from these data. This is further enhanced by the integration of these data with clinical datasets. More than 100,000 genome sequences have been assembled in the 100,000 Genomes Project data centre alongside linked clinical data, amounting to more than 3,800,000,000 clinical data points (including routinely collected datasets of hospital events (HES), deaths (Office of National Statistics), cancer diagnostics and pathology (ENCORE), and cancer therapeutics (Systemic Anti Cancer Therapy)). This collaborative framework acts as a national pipeline for rapid, clinical-diagnostic-grade processing of integrated clinical and genomic data. The Medical Research Council awarded GBP24,000,000 to Queen Mary in 2014 to create a multipetabyte data centre which has enabled access for 3,000 researchers from 33 countries to the 100,000 Genomes Project dataset. These researchers have self-organised and nominated 42 research domains for rare disease, cancer, and cross-cutting themes of focus.

Delivery of the UK's Genomic Medicine Service

Due to the benefit the 100,000 Genomes Project has brought to patients, the UK Secretary of State for Health announced a UK genomics strategy in September 2020. As reported in a government press release on 27th February 2019, "The focus on genomics follows the success of the 100,000 Genomes Project, which helped 1 in 4 patients with rare diseases receive a diagnosis for the first time" [5.9]. This strategy will support the delivery of 500,000 whole-genome analyses by 2023/24 to empower health improvement and disease prevention [5.9], making the NHS the first national healthcare system in the world to offer whole-genome sequencing in routine care. NHS England and Genomics England have

partnered to deliver this national 'Genomic Medicine Service', which comprises of [5.10]:

- An NHS National Directory of available tests for genetic screening
- A national test-ordering service to ensure systematic and equitable access to and provision of genomic testing, with central tracking and capture of all tests ordered
- Every patient being offered the same Patient Choice, the new model of consent for genomic testing that allows for longitudinal life course followup, recall for research and data being available to international researchers and industry from a UK Data Centre
- A single central repository of all genomic data generated within the NHS known as the National Genomic Research Library. This will enable comprehensive review across all available clinical data in order to best support diagnosis of rare diseases and clinical prediction of cancer behaviour.

5. Sources to corroborate the impact (indicative maximum of 10 references)

[5.1] Caulfield, M., Davies, J., Dennys, M., Elbahy, L., Fowler, T., Hill, S., Hubbard, T., Jostins, L., Maltby, N., Mahon-Pearson, J., McVean, G., Nevin-Ridley, K., Parker, M., Parry, V., Rendon, A., Riley, L., Turnbull, C., & Woods, K. (2015). The 100,000 Genomes Project Protocol v1. *Genomics England*. https://doi.org/10.6084/m9.figshare.4530893.v1

[5.2] Genomics England. *Participants from NHS Genomic Medicine Centres have shared their stories with us*. https://www.genomicsengland.co.uk/taking-part/participant-stories/. Accessed 8 December 2020.

[5.3] Wellcome Trust (2014, 1 August). *Wellcome Trust invests £27m in world-class sequencing facility for Genomics England and Sanger Institute*. https://wellcome.org/press-release/wellcome-trust-invests-%C2%A327m-world-class-sequencing-facility-genomics-england-and. Accessed 8 December 2020

[5.4] Genomics England (2014, 22 December). *NHS Genomic Medicine Centres announced for 100,000 Genomes Project*. https://www.genomicsengland.co.uk/genomic-medicine-centres/. Accessed 8 December 2020.

[5.5] Genomics England (2016, 21 November). *UK Prime Minister Opens New Sequencing Centre*. https://www.genomicsengland.co.uk/uk-prime-minister-opens-100000-genomes-project-sequencing-centre/. Accessed 8 December 2020.

[5.6] UK Government (2014, 1 August). *Human genome*: *UK to become world number 1 in DNA testing.* https://www.gov.uk/government/news/human-genome-uk-to-become-world-number-1-in-dna-testing. Accessed 8 December 2020.

[5.7] Medical Research Council. *Clinical Research Capabilities and Technologies Initiative*. https://www.mrc.ac.uk/research/initiatives/clinical-research-capabilities-and-technologies-initiative/. Accessed 8 December 2020.

[5.8] NHS. *Genomic Laboratory Service Re-design*. https://www.engage.england.nhs.uk/consultation/genomic-laboratories/. Accessed 8 December 2020.

[5.9] UK Government (2019, 27 February). *Health minister: NHS must lead the world in genomic healthcare*. https://www.gov.uk/government/news/health-minister-nhs-must-lead-the-world-in-genomic-healthcare. Accessed 8 December 2020.

[5.10] UK Government (2020). *Genome UK: The Future of Healthcare*. https://www.gov.uk/government/publications/genome-uk-the-future-of-healthcare