Congenica

100k Genomes Pilot

on Rare Disease Diagnosis in Healthcare

New England Journal of Medicine:

100,000 Genomes Pilot on Rare Disease Diagnosis in Health Care

Preliminary Report

Author:
Andrea Haworth, Lead Clinical Scientist for Rare Disease, Congenica and co-author of *'100,000 Genomes Pilot on Rare Disease Diagnosis in Health Care - Preliminary Report'*

Introduction

The UK 100,000 Genomes Project was a globally unique translational research program established to collect clinical data and undertake whole genome sequencing (WGS) on 85,000 patients with undiagnosed rare disease or cancer from the UK National Health Service (NHS).

The aims of the project were:

- To bring benefit to patients

- To create an ethical and transparent program based on consent

- To enable new scientific discovery and medical insights

- To kickstart the development of a UK genomics industry

The project has sparked huge global interest and has provided a template for other largescale / national genome projects.

Congenica is a key partner of Genomics England from supporting the initial pilot project, the findings of which are described in this paper, to being chosen as the sole clinical decision support platform provider for rare disease for the new national genomic medicine service.

With the publication of a recent paper in New England Journal of Medicine, **100,000 Genomes Pilot on rare disease diagnosis in healthcare – Preliminary Report** [1], the 100,000 Genomes Project Pilot investigators, including Congenica, outline in detail the findings of the recruitment, analysis and reporting of the initial 4660 participants from 2183 families with 161 rare disorders. All participants had previously undergone normal NHS standard of care prior to recruitment and had remained undiagnosed.

The Project

Participants and their families

The systems processes and frameworks developed to deliver the 100,000 Genomes pilot project were done in collaboration with many stakeholders, with patients, families and their advocates playing the most important role.

Consequently, a detailed consenting process was developed for the project which ensured that participants' health care activities can be followed over their life course using electronic health records (EHR), registries, and data repositories. These data were critical to assessing impact and health outcomes and providing an overview of the rare disease participant's journey to diagnosis. In addition, they allow assessment of the impact on health care resources and provide information that allows physicians/scientists to look for potential patterns within disease categories.

The analysis of genomic data must be undertaken in the context of the patient's clinical presentation. To enable this, the pilot project developed systems to allow referring physicians to collect the patient's signs and symptom (phenotype data) in a structured format using Human Phenotype Ontology (HPO). These data have also been used to improve diagnostic capabilities via use of algorithms such as Exomiser and other tools embedded within clinical decision support platforms. Additional information such as previous genomic testing (e.g., chromosome microarray, gene panel testing etc.) was also collected where possible.

Adults with rare disease were more commonly enrolled than children (74% vs. 26%), which is in line with the general population of England and Wales. A lower percentage of females were recruited than males, which was especially true for children and may reflect the increased susceptibility of males to recessive X-linked conditions. The inferred ancestry of affected individuals was consistent with that expected from the population, indicating that the diversity of the participants was broadly representative of the population at large (2011 census of England and Wales).

Samples and clinical data from were collated and processed by regional genomic laboratories/centers and exported to the sequencing provider and Genomics England Ltd.

Genome sequencing (GS)

GS was undertaken in partnership with Illumina Corp. PCR-free library preparation was used and short read sequencing at a mean depth of 32× (range from 27× to 54×) was undertaken on an Illumina platform. At least 95% of the reference human genome was sequenced at 15x. For the Pilot Project WGS reads were aligned to the Genome Reference Consortium human genome build 37 (GRCh37). After completion of the pilot reads were aligned to GRCh38.

Finding the needle in the haystack: data analysis and variant identification strategy

A multipronged approach collectively contributed to an overall diagnostic yield of 25% for the pilot project.

The 100,000 Genomes Project is the largest genome sequencing project completed to date, and construction of high quality scalable of data analytics for variant identification during the pilot project was key to success. An automated analytical pipeline was constructed to filter the 3 billion letters of the human genome. The goal was to filter the genome to remove variants unlikely to be causal and highlight the most likely ones in the remainder (variant prioritization).

A multi-step approach was used in order to efficiently and rapidly identify high quality causal variants and achieve maximum diagnostic yield. This process is described at length in the paper and is referred to as "variant tiering" and "variant prioritization". In addition, research analysis across the cohort was undertaken to identify new genes and gene/disorder associations.

Variant tiering pipelines categorize and filter variants based on the predicted impact on a gene/protein, the presence or absence in phenotypically relevant genes and whether the variant inheritance pattern matched the disorder (segregation) and gene.

Variant prioritization was undertaken using Exomiser, a sophisticated algorithm that utilizes phenotype terms (HPO) to drive variant analysis, taking into account knowledge of the phenotypic presentation of the gene in humans and animal models, plus a number of other parameters. In the pilot project, Exomiser was used to prioritize variants both within gene panels and across all genes in the genome. In addition, clinical decision support platforms, tools and associated data analytics provided by partner organizations (such as Congenica) were utilized to prioritize and identify additional candidate variants. Candidate variants identified by any of the aforementioned methods were reviewed, and classifications agreed by genomic experts at the recruiting center before being returned to participants.

Assessing impact by collecting participant outcomes

A range of final clinical outcomes were collected including whether a genetic diagnosis was confirmed and whether the variant(s) found explained all, or some of the participants' phenotypes. In addition, resulting healthcare benefits in respect of any change in medication, additional surveillance for the rare disease individual or relatives, clinical trial eligibility or whether the results informed future reproductive choices were collected.

Findings

Providing diagnoses for patients

A genetic diagnosis was achieved for 25% of probands in total and ended long diagnostic odysseys for many patients and their families.

Diagnostic yields were highest when samples and clinical information was available from multiple family members, for example trios where genetic data is available from the mother and father as well as the affected individual. This is because variants can be automatically filtered out if they do not segregate within the family as expected.

The chance of being diagnosed was not evenly distributed across the cohort. Unsurprisingly, cases where the referring clinician suspected a specific diagnosis or in which genes were known to underpin the disorder had much higher diagnostic yields (35%), with intellectual disability, hearing and vision disorders achieving yields between 40% and 55%. Those with more complex etiologies had an overall 11% yield. Tumor syndromes had a low diagnostic yield at 6%, which may be due to the fact that most of the established tumor predispositions genes had already been tested for prior to entry into the 100,000 Genomes Project.

Variant tiering initially identified 1041 candidate variants to the referring centers for review. 291 (28%) were reported to be diagnostic and a total of 60% of the confirmed

diagnoses involved variants identified by the variant tiering process. The analysis was re-run in December 2019 on data from the pilot and a subsection of the main program using updated versions of the gene panels, pipelines and platforms, increasing the number of detected genetic diagnoses from 322 to 377. In addition, a median of 1 candidate variant per case was returned to the referring centers.

A review of variant prioritization (SNVs and indels) by Exomiser demonstrated that in diagnosed cases it ranked the causal variant as the top variant in 77% of cases. 86% and 88% of diagnoses were in the top 3 and top 5 rankings, respectively.

Combined use of Exomiser and updated and improved virtual panels ranked the causal variant in the top 5 in 92% of these diagnoses. Diagnostic discoveries derived by a combination of research, clinical decision support platform discoveries, clinical validation, and assessment yielded a total of 72 additional diagnoses.

10% of probands were classified with variants of uncertain clinical significance in genes consistent with the patient's clinical presentation. When possible, these variants are subjected to additional investigation, e.g., assessment of functional impact.

Improving variant detection and increasing diagnoses

13% of the diagnoses obtained via WGS were caused by variants that are difficult to detect using other methodologies, such as causal variants in non-coding region of genes, complex structural variations, variants impacting the mitochondrial genome, and tandem repeat disorders. Additionally, 2% of the diagnoses made in the study were discovered in coding regions that had low coverage using exome sequencing tests.

The results of the study provide new evidence of the value of genome sequencing compared to other tests, on a large scale across a broad range of 161 disorders. They mirror the results of other smaller studies in which more than 50% of participants who received a new diagnosis by genome sequencing had previously been tested by exome sequencing [2].

Enabling new scientific discoveries and medical insights

The dataset generated by the 100,000 Genomes Project provides a unique resource to allow scientists and physicians to expand knowledge of known rare disorders and discover new disease genes. For example, cohort-wide burden testing across 57,000 genomes (including the cases analyzed as part of the pilot project) has to date enabled the discovery of 3 new disease genes which were independently confirmed – namely UBAP1 in hereditary spastic paraplegia, FOXJ1 in non-CF bronchiectasis, and SORD in Charcot-Marie-Tooth disease. In addition, 22 candidate genes have been identified which likely represent new mendelian disease genes.

Healthcare benefits of genome sequencing

Of the genetic diagnoses made:

25%
had immediate impact on clinical decision making for the patients or their relatives

and only

0.2%
were described as having no benefit.

The clinical utility of a genomic diagnosis was described in detail in the paper using participant stories and included:

13 cases
allowing eligibility for clinical trial, e.g., a 36-year-old participant whose diagnosis meant he was eligible for a gene replacement trial for vision loss

4 cases
where a diagnosis led to a suggested change in medication, e.g., a diagnosis in a child which allowed testing and therapeutic intervention for a younger sibling within weeks of birth or a case where a diagnosis in a 10-year-old admitted to intensive care enabled a curative bone marrow transplant

26 cases
where the diagnosis impacted suggesting additional surveillance for the proband or relatives

59 cases
where diagnosis informed future reproductive choices

32 cases
with other benefits

Does WGS improve patient care above and beyond existing genetic testing strategies?

Genome sequencing resulted in a substantial increase in diagnoses across a broad spectrum of rare diseases, whether or not participants had undergone previous genetic testing, and for 25% of those who received a genetic diagnosis via WGS, there was immediate clinical utility.

Data was available from 1177 rare disease participants on the presence or absence of any previous genetic tests. The median number of genetic tests undertaken prior to entry into the 100,000 Genomes Project Pilot was 1, with a range of 0-16. Among these 1177 patients, genome sequencing resulted in a substantial increase in diagnoses across a broad spectrum of rare diseases. This was observed regardless of whether participants had been previous genetically tested (31%) or not (33%), and for 25% of those who received a genetic diagnosis via WGS, there was immediate clinical utility.

Legacy: creation of the genomic medicine service

The patients and families that were part of the pilot study were recruited and sequenced throughout 2014-2016. The infrastructure to collect, process, analyze and return the data was developed over the same period. Genomic results were returned to the referring centers from May 2016 to April 2019.

The pilot project was intended to inform the 100,000 Genomes Project Main Programme, and in November 2020, it was reported that 122,945 whole genomes, including 86,073 from rare disease participants, have been sequenced in the project and the NHS, and post-pilot phase results reported to have been returned to the referring centers within 6 weeks.

The pilot project findings described in this publication have informed the development of the NHS National Genomic Test Directory, which in turn catalogues the rare diseases which will receive genome sequencing as a first or second line test in the new NHS Genomic Medicine Service. With an ambitious target of sequencing 500,000 whole genomes from rare disease and cancer cases within the first 5 years the NHS will be the first national health care system to offer whole genome sequencing as part of routine care.

Conclusion

Based on performance in the pilot, GEL and NHSE subsequently selected Congenica as the exclusive clinical decision support platform and sole provider of genomic data analysis for rare disease cases in the UK Genomic Medicine Service – the world's first health system to provide whole- genome sequencing to patients at a national level. Ambitious commitments in the NHS Long Term Plan include sequencing 500,000 whole genomes by 2023/24. [3]

The results of this study results demonstrate the value of genome sequencing for unmet diagnostic needs in rare diseases, and we hope that the findings will help other health systems in their integration of genome sequencing and analysis in the care of patients with rare diseases.

Do you want to learn more?
Read the full paper
www.nejm.org/doi/full/10.1056/NEJMoa2035790

Congenica
Enabling genomic medicine

References

[1]   The 100,000 Genomes Project: early impact on rare disease diagnosis and 1 management in a national healthcare system. New England Journal of Medicine
https://www.nejm.org/doi/full/10.1056/NEJMoa2035790

[2]   Splinter K, Adams DR, Bacino CA, et al. Undiagnosed Diseases Network. Effect of Genetic Diagnosis on Patients with Previously Undiagnosed Disease.
N Engl J Med 2018; 29;379(22):2131-2139.

[3]   NHS Genomic Medicine Service
https://www.england.nhs.uk/genomics/nhs-genomic-med-service/

[4]   Genomics England case study
https://genomics.congenica.com/genomics_england_case_study

Partner with Congenica

Discover how easy it is to perform NGS testing when you partner with Congenica. Talk to a clinical expert today.

⊕   www.congenica.com/cis
✉   hello@congenica.com
🐦   @congenica

Request a consultaton

⊕   www.congenica.com/consult