MADRID
22ND CONGRESS
JUNE 22 - 25 | 2017
European Hematology Association

Hereditary hematological disorders - Section 3

The 100,000 Genomes Project

Willem H. Ouwehand on behalf of the NIHR BioResource – Rare Diseases

University of Cambridge and Wellcome Trust Sanger Institute & NHS Blood and Transplant Cambridge, United Kingdom

Take-home messages

- The NHS 100,000 Genomes Project shows that the clinical application of whole genome sequencing for rare inherited hematological diseases is feasible and scalable.
- Clinical and laboratory phenotype information from patients with inherited hematological diseases should be coded with Human Phenotype Ontology (HPO) terms.
- Patients with inherited hematological diseases and their close relatives should be invited to consent on the wide sharing of their phenotype and genotype data via 'safe haven' models.

The 100,000 Genomes Project

The reduction in sequencing costs has enabled initiatives like the 1000 Genomes,[1] UK10K[2] and more recently the 100,000 Genomes Projects (100KGP).[3] The 100KGP aims to achieve analysis by whole genome sequencing (WGS) of the DNA samples of 100,000 NHS patients (Figure 1). The key objective is to establish WGS as standard care in the domains of infection, cancer and rare diseases. The rare diseases element commenced in 2013 and so far samples from just over 36,000 individuals have been analyzed by WGS. Sequencing services are provided by Illumina Cambridge Ltd and are to clinically accredited standard. For the pilot phase for rare diseases, which comprised the first 13,000 DNA samples the sequencing results were transferred to the High Performance Compute Service at the University of Cambridge. For the 100KGP main phase a dedicated data center has been commissioned by Genomics England Ltd (GEL). GEL is a not-for-profit organization entirely owned by the Department of Health and tasked to coordinate the delivery of the 100KGP.

Rare diseases

Twelve rare disease projects were initiated and patients and close relatives were enrolled using a single NIHR BioResource participant information and consent leaflet. The allocated WGS capacity ranged from 1,250 samples/project for five large projects to hundreds of samples for each of the remaining projects. The projects named Bleeding, Thrombotic and Platelet Disorders (BPD), Primary Immune Disorders (PID) and Stem Cell and Myeloid Disorders (SMD) are relevant to the immunology, hematology and hemostasis communities. For these projects clinicians were asked to only enroll patients with molecularly unexplained rare diseases with a high likelihood of being inherited. The sequencing of 13,000 DNA samples for the pilot projects was completed in 2016 and analysis has commenced.

The Human Phenotype Ontology system

Clinical and laboratory phenotype data have been captured using Human Phenotype Ontology (HPO) terms. The HPO is an open source project for phenotypic annotation of genetic disorders. More than 10,000 terms in the HPO are connected via a hierarchy of is-a relationships. Appending of HPO terms allows automated grouping of patients according to phenotypic similarities.[4,5]

Pathogenic and likely pathogenic variants in known genes

Experts in statistical genomics and bioinformatics work with members of the clinical care teams to analyze the phenotyping and genotyping results. There are 264 known genes for BPD, PID and SMD. In the first round of analysis these known genes are reviewed for pathogenic or likely pathogenic variants by a multi-disciplinary team (MDT). So far the analysis has focused on the 65 MB of coding space or exome. As expected for patient with unexplained disorders, causal variants were only identified in <15% of cases and for these samples research reports were issued to

the referring clinician. To address the important question which level of diagnostic sensitivity can be achieved, one of the rare disease pilot projects focused on retinal disease and relaxed eligibility criteria were applied. All patients with visual impairment with a high likelihood of being inherited were enrolled in a cohort of 722 individuals. Likely pathogenic and pathogenic variants were identified for 404/722 (56%) individuals.[6]

Discovery of novel genes

Samples from patients without causal variants are entered in a second round analysis to identify possible new genes. This analysis relies on the development and application of new statistical methods. Such methods can be applied to automatically cluster patients based on phenotypic similarities and also exploits the richness of information from the Mouse Genome Informatics (MGI) and the Online Mendelian Inheritance in Man databases.[7] These approaches, which are critically dependent on the use of HPO terms have shown to be effective in identifying variants in several novel genes.[8-11] For example the identification of gain-of-function (GOF) variants in *DIAPH1* underlying macrothrombocytopenia and deafness illustrates the usefulness of the new methods.[8] Similarly, the discovery that a GOF variant in the kinase *SRC* results in a Grey Platelet Syndrome-like disorder characterized by bleeding, myelofibrosis and osteoporosis was the result of a more integrated approach to data analysis. In addition our approach of bringing genotype and phenoype information from many patients together in a single database also revealed that macrothrombocytopenia, sometimes accompanied by bleeding can be caused by autosomal dominant acting variants in *GP1BB*, one of the well known genes implicated in inherited platelet disorders.[12] Furthermore,

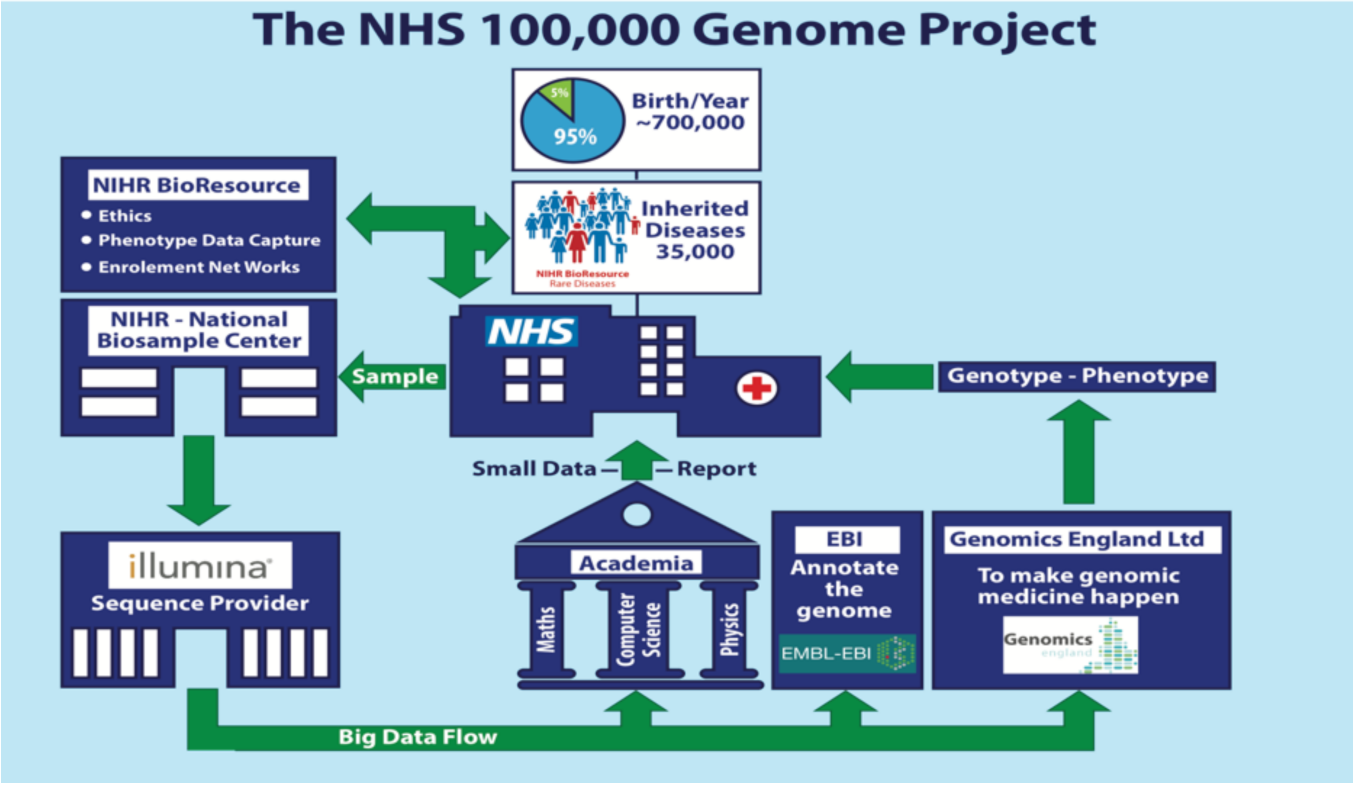

Figure 1. A diagram illustrating the different elements of the 100,000 Genomes Project. There are an estimated 700,000 birth per annum in the United Kingdom and 1 in 20 will experience ill-health during the first decades of life because of an inherited disease. There are an estimated 7,000 inherited diseases of man and the implicated genes have been identified for slightly more than halve of the known diseases

the sequencing of a large number of samples in a single programe has also been helpful in replicating several recent gene discoveries, thereby reducing the risk of erroneous discoveries.[13-15] Another avenue to gene discovery is by combing the results from different genotyping studies. Meyer *et al.* recently illustrated the power of this approach by combining the results from the UK10K Consortium, Deciphering Developmental Disorders study and the NIHR BioResource. They identified 27 unrelated individuals with a childhood-onset dystonia caused by variants in *KMT2B*, the gene encoding the mixed-lineage leukemia protein 4.[16]

Incorporating novel gene findings in diagnostic tests

Historically the cost of Sanger sequencing has prevented the genetic analysis of DNA samples from patients with an assumed inherited disorder. High throughput sequencing (HTS)-based gene panel tests provide an opportunity to analyse a DNA sample at an affordable cost. An international collaboration of BPD experts illustrated that such HTS platforms can achieve excellent sensitivity and specificity.[17-19] Similar HTS panel tests are now available for PID, SMD, pulmonary arterial hypertension and hereditary hemorrhagic telangiectasia. It is expected that by 2022 a WGS test will cost ~€200 rendering HTS gene panel tests obsolete.

Data sharing

The labelling of DNA variants with their level of pathogenicity requires a careful approach because of the risks associated with erroneous genetic diagnosis. So far 1,279 and 635 BPD patient samples have been sequenced by the WGS and HTS tests, respectively. The MDT identified 186 pathogenic and 174 likely pathogenic variants in the 78 known BPD genes, illustrating that nearly halve of the causal variants are novel. To enhance the quality of variant labelling information about genotype and phenotype should be shared through 'safe haven' models. An example of such a safe environment for data sharing is the European Genome-phenome Archive at the European Bioinformatics Institute. Similarly, information of variants should be deposited in freely accessible databases. The ClinVar database, which is being maintained by the National Center for Biotechnology Information provides such an archive on the relationships among DNA variants and phenotypes. The results of the NIHR BioResource rare disease projects will be deposited into ClinVar and EGA.

Conclusions

The 100KGP project commenced in 2013 and so far 36,000 DNA samples have been analyzed. The WGS analysis is now a fully accredited service. Genomic medicine centers have been established at 11 leading academic centers and the sequencing results are reviewed by MDTs with input from clinicians, clinical geneticists and bioinformaticians. The 100KGP has laid the blueprint for the delivery of genomics services in the NHS.

References

1. The 1000 Genomes Project, et al. A global reference for human genetic variation. Natur. 2015;526:68-74.
2. UK10K Consortium, et al. The UK10K project identifies rare variants in health and disease. Nature 2015 1;526:82-90.
3. Peplow M. The 100,000 Genomes Project. BMJ 2016;353:i1757.
4. Kohler S, Vasilevsky NA, Engelstad M, et al. The Human Phenotype Ontology in 2017. Nucleic Acids Resh 2017 ;45:D865-D76.

*5. Westbury SK, Turro E, Greene D, et al. Human phenotype ontology annotation and cluster analysis to unravel genetic defects in 707 cases with unexplained bleeding and platelet disorders. Genome Med 2015;7:36.

The manuscript reports on the importance of coding clinical and laboratory phenotype data from patients with inherited haematological disorders using a systematic library of terms, like the Human Phenotype Ontology system.

6. Carss KJ, Arno G, Erwood M, et al. Comprehensive rare variant analysis via whole-genome sequencing to determine the molecular pathology of inherited retinal disease. Am J Hum Genet 2017 05;100:75-90.
7. Greene D, BioResource N, Richardson S, Turro E. Phenotype similarity regression for identifying the genetic determinants of rare diseases. Am J Hum Genet 2016;98:490-9.

*8. Stritt S, Nurden P, Turro E, et al. A gain-of-function variant in DIAPH1 causes dominant macrothrombocytopenia and hearing loss. Blood 2016;127:2903-144.

The manuscript reports on the discovery of variants in a new gene underlying macrothrombocytopenia and hearing loss. The discovery of this novel gene was made possible by making use of a recently developed statistical method.

9. Stritt S, Nurden P, Favier R, , et al. Defects in TRPM7 channel function deregulate thrombopoiesis through altered cellular Mg(2+) homeostasis and cytoskeletal architecture. Nature Commun 2016;7:11097.

*10. Turro E, Greene D, Wijgaerts A, et al. A dominant gain-of-function mutation in universal tyrosine kinase SRC causes thrombocytopenia, myelofibrosis, bleeding, and bone pathologies. Sci Transl Med 2016;8:328ra30.

The manuscript reports on the use of a new statistical method to analyse whole genome sequencing data by use of Human Phenotype Ontology coding and blood cell progenitor RNA-sequencing data.

11. Pleines I, Woods J, Chappaz S, et al. Mutations in tropomyosin 4 underlie a rare form of human macrothrombocytopenia. J Clin Invest. 2017;127:814-29.
12. Sivapalaratnam S, Westbury SK, Stephens JC, et al. Rare variants in GP1BB are responsible for autosomal dominant macrothrombocytopenia. Blood 2017;129:520-4.

Hereditary hematological disorders - Section 3

13. Chen L, Kostadima M, Martens JH, et al. Transcriptional diversity during lineage commitment of human blood progenitors. Science 2014;345:1251033.
14. Burley K, Whyte CS, Westbury SK, et al. Altered fibrinolysis in autosomal dominant thrombomodulin-associated coagulopathy. Blood 2016 pii: blood-2016-05-716092. [Epub ahead of print].
15. Poggi M, Canault M, Favier M, et al. Germline variants in ETV6 underlie reduced platelet formation, platelet dysfunction and increased levels of circulating CD34+ progenitors. Haematologica 2017;102:282-94.
16. Meyer E, Carss KJ, Rankin J, et al. Mutations in the histone methyltransferase gene KMT2B cause complex early-onset dystonia. Nature Genet 2017;49:223-37.
*17. Simeoni I, Stephens JC, Hu F, et al. A high-throughput sequencing test for diagnosing inherited bleeding, thrombotic, and platelet disorders. Blood 2016;127:2791-803.
The manuscript reports on the validation of a high throughput sequencing (HTS) gene panel test for the comprehensive diagnostic testing of DNA samples from patients assumed to have an inherited bleeding, thrombotic or platelet disorder.
18. Lentaigne C, Freson K, Laffan MA, et al. Inherited platelet disorders: toward DNA-based diagnosis. Blood 2016;127:2814-23.
19. Bariana TK, Ouwehand WH, Guerrero JA, et al. Dawning of the age of genomics for platelet granule disorders: improving insight, diagnosis and management. Br J Haematol 2017;176:705-20.