**Next Generation Sequencing’ as a diagnostic tool in Paediatrics**

Corresponding author:

Prof. Diana Baralle

Faculty of Medicine | University of Southampton | Human Development and Health, Duthie Building, Southampton General Hospital, Tremona Road, Southampton, SO16 6YD, UK

+44 (0)23 8120 4264

email: d.baralle@soton.ac.uk

Co-author:

Dr Vardha ISMAIL,

UNIVERSITY HOSPITAL SOUTHAMPTON NHS FOUNDATION TRUST, Southampton, UK

email: vardha.ismail1@nhs.net

Keywords: Genetics, genomics, next generation sequencing, molecular biology

Word count - 1351

**‘Next Generation Sequencing’ as a diagnostic tool in Paediatrics**

Two recent studies in Asia illustrate the potential of Next Generation sequencing (NGS) and the value of large-scale studies in Asian cohorts to represent variation in the reference genome. The United Kingdom itself, has a diverse population and acknowledging the genetic variation that exists within differing ethnic groups is important to deliver a high-quality Genomic service for all. The paper from Heming et al1 demonstrates that an understanding of what each NGS test provides allowed for the use of a large exome gene panel rather than whole exome sequencing. This still increased the diagnostic yield to almost 40% in Mendelian disorders. Bhatia et al2 further showed that utilising whole exome and whole genome sequencing led to a diagnostic yield of 38% and 33% respectively in their Asian cohort. Particularly in children with neuromuscular and skeletal dysplasia phenotypes where performing a ‘trio exome’ also contributed to a higher diagnostic yield. Bhatia et al additionally demonstrate that 61% of the variants found in their multi-ethnic Asian population were novel. This information is crucial to help collate accurate reference data sets, which tend to have a European bias, with Asian ancestry represented by 14% of samples.3

The human genome was first sequenced in 2003 and helped to unravel the complexities behind disease causing alterations in our DNA. Although genetic testing has evolved a great deal since then, the original and ‘first generation’ method used to sequence the genome was ‘Sanger sequencing.’

Named after Fred Sanger who developed this in 1975, involves using DNA as a template to generate a set of fragments that differ in length. The fragments are labelled and reassembled into the original sequence of DNA. Although this is a reliable method, the greatest disadvantage is only one small fragment of DNA can be sequenced at a time, which proves to be costly and time consuming.

Genetic testing was already available in the UK healthcare system however; from March 2017 a National Genomic Medicine Service in the UK was formed. The main strategy being to utilise the power genetic testing can have in “predicting and diagnosing inherited and acquired disease, and to personalise treatments and interventions” 4. This service comprises of genomic laboratory hubs, a national test directory and Whole Genomic Sequencing (WGS) provision. The Royal College of Paediatrics also echoes this vision for WGS provision for paediatric patients and is further reiterated in the new government paper “Genome UK the future of healthcare’5

**What is Next Generation Sequencing (NGS)?**

Next Generation Sequencing allows for hundreds of thousands of DNA fragments to be sequenced at the same time. This is known as “massively parallel sequencing.” The term NGS does not refer to one single technique as NGS utilises a number of different platforms such as sequencing by synthesis, pyrosequencing, ion semi-conductor sequencing or sequencing by ligation. These vary in how the multiple DNA fragments are all separated, but still utilise the principle of massively parallel sequencing. The most commonly used method is ‘sequencing by synthesis’ as summarised: 6

1. **Library Preparation:**

DNA is randomly fragmented by an enzyme into smaller pieces and adding an ‘adaptor’ to both ends of the fragment.

1. **Cluster Generation:**

Each fragment is loaded onto a glass slide called a “flow cell”. This has oligonucleotides attached that anchor the fragment to the flow cell. Each fragment is amplified into multiple copies to form different “clusters” which originate from one fragment. The amplification of each fragment allows for better detection during sequencing.

1. **Sequencing:**

Modified nucleotides (A,C,T,G) are fluorescently labelled with different colours and added to the flow cell. The corresponding nucleotide will bind to the fragment. A laser is used to identify which nucleotide was bound. This cycle continues to be repeated on each cluster and the number of cycles performed to sequence the DNA is known as a “read.”

1. **Data Analysis:**

The sequence “reads” are aligned against a reference genome with computer software and any variation is detected.

**How can NGS be used?**

Depending on the phenotype and index of suspicion a targeted or a broad approach can be taken. A targeted approach would involve sequencing a single gene associated with the suspected genetic condition, or using a ‘gene panel’ where a group of genes associated with the phenotype are sequenced. If a broader approach is needed a ‘clinical exome’ can sequence all of the genes known to be associated with human disease. Broader still, would be ‘whole exome sequencing’ (WES) which sequences all the known coding regions (exons) in the genome which is almost 20,000 genes. Beyond this, WGS can be utilised where all the exons and non-coding (introns) regions are sequenced.

**What are the benefits of NGS?**

*Diagnostic yield*

For the targeted approach where perhaps a diagnosis is already in mind, using large gene panels can improve the diagnostic yield. However, the importance of accurate phenotyping cannot be underestimated. The clinical assessment will provide clarity for which gene panel is most likely to yield a positive diagnosis. For the broader NGS testing such as WES and WGS the benefits lie in being able to cast the net wider which can capture genetic changes such as structural variants or variants of uncertain significance. WGS in particular can deeply interrogate the genome which may yield deep intronic changes that can affect gene regulation. The adaptation of the bioinformatics pipelines to detect these kinds of complex changes requires a significant amount of work.

*Testing efficiency*

Owing to the multiple ’reads’ NGS can provide, it can detect low level mosaicism, previously undetected on Sanger sequencing. It can also detect small base changes (substitutions), insertions and deletions of DNA, large genomic deletions of exons or whole genes. This is helpful in genetically heterogeneous conditions such as early infantile epileptic encephalopathy (EIEE) which can be due to sequence variants, translocations, deletions or duplications.7 This is also useful in terms of maximising the testing utility, as Sanger sequencing often requires other assays or tests to confirm larger insertions, deletions or DNA rearrangements.8

**What are the limitations of NGS?**

*Low read depth*

When reviewing any results from NGS it is important to assess how well the gene or region of interest was sequenced. Most clinical labs used a read depth of 30x and can increase this further in conditions known to be mosaic. If the ‘read’ depth is low owing to poor gene coverage this would be important as further Sanger sequencing of that region may be required.

*Variants of uncertain significance*

NGS yields genetic variants that are interpreted and tiered as ‘pathogenic’, ‘likely pathogenic’, ‘uncertain significance’, ‘likely benign’ or ‘benign’ 9. However, difficulties can arise when novel genetic variants which may not have been seen before are found. These are known as ‘variants of uncertain significance’ (VUS). The interpretation of a VUS is an evolving aspect of Clinical Genetics. This involves reviewing control population databases, understanding the functional impact of the VUS on protein production and clinical assessment in determining how convincing the phenotype is for the genetic condition. This information may change how a variant is tiered, moving it from a VUS to pathogenic. For this reason, communication around the patient phenotype between specialities and Clinical Genetics is incredibly important.

*Ethnic diversity*

The reference genome used to compare sequenced DNA against, has a strong European bias. For ethnic groups out of this European framework, how reliable is the reference genome to make sense of genetic variation? Conducting further studies of large WES or WGS of multiple ethnic groups is important to better reflect the genetic variation of populations in the reference genome.

**NGS and the future**

As NGS can increase diagnostic yield, it is growing to have a vital role in healthcare. Particularly in paediatric patients, where earlier diagnosis can end the ‘diagnostic odyssey’ families endure, impact on the immediate care and greater inform reproductive counselling. It may also be that tests such as WGS become a routine part of Neonatal care in the future. For these reasons, as NGS becomes more accessible both in the UK and globally it is important for non-geneticists to feel comfortable with when to consider NGS and what the potential benefits and limitations are.

**References**

1. *Heming W, Angeline LHM, Ee-Shien T, Mark KJA, Ivy N, Teck WT, et al. Genetic Landscape of Congenital Disorders in Patients from Southeast Asia: Results from Sequencing Using a Gene Panel for Mendelian Phenotypes. Arch Dis Child 2020.*
2. *Bhatia NS, Lim JY, Bonnard C, Kuan JL, Brett M, Wei H, et al. The Singapore Undiagnosed Disease Program: Genomic Analysis Aids Diagnosis and Clinical Management. Arch Dis Child* 2020.
3. P*opejoy AB, Fullerton SM. Genomics is failing on diversity. Nature. 2016; 538 (7624):161-4.*
4. *NHS England. NHS Genomic Medicine Service. Available from:* [*https://www.england.nhs.uk/genomics/nhs-genomic-med-service/*](https://www.england.nhs.uk/genomics/nhs-genomic-med-service/) *[Accessed 1st August 2020].*
5. *Genome UK: The Future of Healthcare. Available from:* [*https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment\_data/file/920378/Genome\_UK\_-\_the\_future\_of\_healthcare.pdf*](https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/920378/Genome_UK_-_the_future_of_healthcare.pdf) *[Accessed 4th October 2020].*
6. *Illumina. An introduction to Next-Generation Sequencing Technology. Available from: https://emea.illumina.com/content/dam/illumina-marketing/documents/products/illumina\_sequencing\_introduction.pdf [Accessed 1st August 2020].*
7. Ostrander BEP, Butterfield RJ, Pedersen BS, Farrell AJ, Layer RM, Ward A, Miller C, DiSera T, Filloux FM, Candee MS, Newcomb T, Bonkowsky JL, Marth GT, Quinlan AR. Whole-genome analysis for effective clinical diagnosis and gene discovery in early infantile epileptic encephalopathy. NPJ Genom Med. 2018; 1(3):22.
8. *Behjati S, Tarpey PS. What is next generation sequencing? Arch. Dis. Child. Educ. Pract. Ed. 2013;98(6):236–8*
9. *Ellard S, Baple EL, Berry I, Forrester N, Turnbull C, Owens M, et al. Association for Clinical Genomic Science. Best Practice Guidelines for Variant Classification 2019. Available from:* [*https://www.acgs.uk.com/quality/best-practice-guidelines/*](https://www.acgs.uk.com/quality/best-practice-guidelines/) *[Accessed 1st August 2020].*