The strengths and weaknesses of whole-genome sequencing

Emily Ferguson
Year 5, Medicine, Cardiff University
Email: FergusonEC@cardiff.ac.uk

Introduction

Recent advancements in genetics have increased our DNA sequencing capabilities. The sequencing of the whole genome has become more commonplace, primarily due to the development of next-generation sequencing (NGS).[1] NGS platforms sequence millions of fragments of DNA in parallel, allowing the entire genome to be sequenced at once.[1] Bioinformatics piece these fragments together using the human reference genome.[1] In this way, the nucleotides in the human genome are sequenced multiple times, providing accurate data much quicker than previous methods.[1] This technology can also be used to sequence the protein-coding region of the genome (the exome), which is called whole-exome sequencing (WES).[1] Both of these methods have led to breakthroughs in genetic research, allowing diagnoses and extending our knowledge of disease mechanisms and potential therapeutic targets.[2] This article seeks to discuss the potential benefits and challenges associated with the use of whole-genome sequencing (WGS).

Advantages

The most obvious advantage of WGS is that the entire genome is scrutinised. This means that, in one test, every single variant in the genome is identified, whether small (a single nucleotide variant) or large (a copy number variant or translocation).[2] This information is incredibly valuable and could identify a monogenic disease, for example, thalassaemia, or could identify a predisposition to developing a polygenic disease, such as type 2 diabetes mellitus.[3]

WGS can be used to diagnose genetic mutations that would affect a particular individual, but it could also act as a screening tool that could allow identification of genetic carriers of recessive diseases, such as cystic fibrosis.[3] This could aid family planning by encouraging partners to be fully tested, and would allow future parents to receive genetic counselling or even pre-implantation genetic screening.[4]

The first ever genome cost 2.7 billion US dollars to sequence, but with the advent of NGS, the cost continues to plummet and it seems possible that, in the future, it will cost less than $1,000 per genome.[5] Despite the obvious cost associated with this, many would argue that the identification of disease and disease risk could save money in the future.[6]

Although not commonly used in clinical practice, WGS in cancer research could allow the identification of genetic drivers of tumours and new biological therapies.[7] However, NGS means that WGS can be provided in a clinical environment due to reduced cost and time frame. If the results can be provided within a clinically relevant time frame, it could have an impact on a patient's cancer management.[7] This is called 'precision oncology'; if WGS can be used to genotype cancer cells, we can ascertain which genes have become mis-regulated and provide a more personalised treatment plan.[7] Non-small cell lung cancer is an example where molecular profiling of the tumour has been proven to be key in order to provide the optimal treatment plan for each patient.[8]

Disadvantages

The major difficulty associated with WGS is the sheer mass of information provided, which must be analysed and assessed to determine what is important or what is not.[9] Although our knowledge in genomics is growing, the roles of many genes are still undetermined and huge numbers of variants across the genome have not yet been distinguished as being benign or pathogenic. This means that, although WGS can produce a large volume of data, most of this may be misleading or useless.[9]

The large amount of data produced by WGS will not only need to be analysed, but it will also need to be stored. This in itself raises some challenges, not limited to the large capacity and cost required for this, but also to the privacy of data, which could raise ethical dilemmas with insurance companies and family members.[10]

WGS may uncover unsought secondary findings; it may reveal a diagnosis of a genetic condition that is untreatable and may not present itself for many years, such as Huntington's disease.[3] This could have a negative psychological impact on the individual and also on family members.[3] It could affect family relationships, as other family members may not want to be aware of such information whilst others would rather know. In addition, it raises the question of which, if any, results should be disclosed to a patient.[3]

Is there already a solution?

Many researchers have turned to WES to overcome some of the challenges of WGS. WES is a more cost-effective method because only 1% of the entire genome needs to be sequenced and over 85% of the mutations are located here.[11] It also means there are fewer variants of unknown significance detected and requiring analysis. WES has already been used successfully to locate genes in which variants can increase the risk of breast and colorectal cancer.[12,13]

However, it can be difficult to detect structural variants using WES, and researchers have found that DNA variants outside of exons, which would only be picked by WGS, can be pathogenic. Not only this, but it is difficult to sequence areas of DNA that are rich in GC nucleotides using WES, which can cause inaccurate sequencing leading to both false negatives and false positives.[14] The large number of false negatives produced by WES means all the variants need to be confirmed by Sanger sequencing (a method of DNA sequencing), which is time consuming, wasting valuable research time. [1,14]

Conclusion

Both WGS and WES have their own set of strengths and weaknesses (each outlined in Table 1), not limited to those mentioned in this article. Nonetheless, both are valuable research methods and will undoubtedly be used to translate the variations between individual human genomes into medically useful information.[14,15]

Table 1. WGS vs WES.

Advantages	Disadvantages
WGS	
Detects both coding and non-coding variants	High cost (currently around $1500 per genome)
Detects structural variants	Huge volume of data to process and store
	Genetic variants need validation using Sanger sequencing
	Variants of unknown significance: limited knowledge to fully understand the implication of each variant and huge numbers of variants (~ 3.5 million) can be found in non-coding regions, which could be relevant or not
WES	
Only around 20,000 variants to analyse	Difficult to detect structural variants
Reduced cost	Sequences only coding DNA
Less data produced so less to be filtered, researched and stored	Difficult to capture sections of DNA with a high GC nucleotide percentage, leading to false positives and negatives
	Genetic variants need validation using Sanger sequencing

Table based on data from[14,15,16]

References

1. Behjati S, Tarpey PS. What is next generation sequencing? *Arch Dis Child Educ Pract Ed*, 2013; 98(6):236–238.
2. Wrzeszczynski KO, Felice V, Shah M, et al (2018) Whole genome sequencing-based discovery of structural variants in glioblastoma. In: Placantonakis D. (eds) Glioblastoma. Methods in molecular biology, vol 1741. Humana Press, New York, NY, pp 1–29.
3. Katsanis SH, Katsanis N. Molecular genetic testing and the future of clinical genomics. *Nat Rev Genet*, 2013; 14(6):415-26.
4. Berberich AJ, Ho R, Hegele RA. Whole genome sequencing in the clinic: empowerment or too much information? *CMAJ*, 2018; 190(5):E124–E125.
5. Radford C, Prince A, Lewis K, et al. Factors which impact the delivery of genetic risk assessment services focused on inherited cancer genomics: expanding the role and reach of certified genetics professionals. *Journal of Genetic Counseling*, 2013; 23(4):522-530.
6 Christensen KD, Dukhovny D, Siebert U, et al. Assessing the costs and cost-effectiveness of genomic Sequencing. *J Pers Med*, 2015; 5(4):470–486.
7. Zhao EY, Jones M, Jones SJM. Whole-genome sequencing in cancer. *Cold Spring Harb Perspect Med*, 2019; 9(3):a034579.
8. Brown NA, Aisner DL, Oxnard GR. Precision medicine in non-small cell lung cancer: current standards in pathology and biomarker interpretation. *Am Soc Clin Oncol Educ book*, 2018; 38:708-715.
9. Botkin JR, Rothwell E. Whole genome sequencing and newborn screening. *Curr Genet Med Rep*, 2016; 4(1):1–6.
10. Howard HC, Knoppers BM, et al. Whole-genome sequencing in newborn screening? A statement on the continued importance of targeted approaches in newborn screening programmes. *Eur J Hum Genet*, 2015; 23(12):1593–1600.
11. Choi M, Scholl UI, Ji W, et al. Genetic diagnosis by whole exome capture and massively parallel DNA sequencing. *Proc Natl Acad Sci U S A*, 2009; 106(45):19096-19101.
12. Seguí N, Mina LB, Lázaro C, et al. Germline mutations in FAN1 cause hereditary colorectal cancer by impairing DNA repair. *Gastroenterology*, 2015; 149(3):563–566.
13. Rahman N, Seal S, Thompson D, et al. PALB2, which encodes a BRCA2-interacting protein, is a breast cancer susceptibility gene. *Nat Genet*, 2007; 39(2):165–167.
14. Meienberg J, Bruggmann R, Oexle K, et al. Clinical sequencing: is WGS the better WES? *Hum Genet*, 2016; 135(3):359-36213.
15. Gong J, Pan K, Fakih M, Pal S, Salgia R. Value-based genomics. *Oncotarget*, 2018; 9(21):15792-15815.
16. Gonzaga-Jauregui C, Lupski JR, Gibbs RA. Human genome sequencing in health and disease. *Annu Rev Med*, 2012; 63:35-61.