Exome Sequencing in early disease diagnosis: Are we on the right track?

Musambil M*

Department of Genetics, Strategic Center for Diabetes Research, King Saud University, Riyadh, Saudi Arabia

Identification of genetic variants associated with monogenic syndromes, complex disorders and related traits opened up an avenue that had not been explored before, which is to translate the genetic screening information into disease predicting tools which could provide more efficient management of the disease by improving risk and its prediction capabilities. The methods to uncover the genetics of these complex disorders have evolved over time. The International HapMap Project, carried out as a part of the Human Genome Project, was successful in providing information about more than one million SNPs (Single Nucleotide Polymorphisms) across the human genome. A revolution in SNP genotyping technology had occurred, making it possible to genotype hundreds of thousands of SNPs, opening new horizons for genetic association studies. The HapMap project also demonstrated that genotyping of about 500,000 SNPs is enough to cover around 75% of the common variants (minor allele frequency >5%) [1]. A multitude of genome-wide association studies (GWAS) for different disorders has been published so far [2]. Although GWAS have resulted in large amounts of data demonstrating the association of hundreds of genetic loci to complex traits in humans, they do not necessarily identify the gene or set of genes in a given locus associated with the disease, and they do not usually give a bigger picture of how the disease gene function [2]. The most associated SNPs in many cases exist only as a marker for the functional variant and the majority of associated regions contain several genes. So far GWAS have been based on the common disease-common variant hypothesis, thus it has failed to detect rare variants with strong effects. In addition, GWAS findings have been inconsistent across populations and what applies to one population does not necessarily apply to another. This is due to different mutational mechanisms caused by differences in demographics, cultural histories, food habits, environment etc. which necessitates identifying population-specific variants [3-4]. The next trend that took rise was the next generation platforms (NGS) also known as high-throughput sequencing which had greater yield and advantages over the conventionally used Sanger's sequencing.

The last decade had witnessed a huge drop in the price of genome sequencing, and NGS have become fully automatic and highly improved. Presently complete genomes can be sequenced quickly and at an affordable price compared to the earlier time. Human genome sequencing enjoys much lower cost per gigabase than other species with the availability of latest sequencers such as Illumina's HiSeq X platform [5]. NGS could identify genetic variants that affect heritable phenotypes, including important disease-causing mutations and natural variation that not only could be used in the medical field but also in agricultural and other social sectors to improve crops and livestock. When considering NGS as a whole, the whole genome sequencing (WGS) is the one that could provide in-depth knowledge of all the variants that affect phenotypic expression when compared to whole exome sequencing (WES) and targeted gene sequencing. But WGS still remains very much expensive compared with WES and targeted sequencing [6] (Tables 1 and 2).

WES could be briefly explained as the process of sequencing exons or the protein-encoding parts of the genes which represent the functional part of the genome. WES gives a clear picture of high penetrance allelic variation and its relationship to disease phenotype [7]. As WES targets exons and with the knowledge that Mendelian or partly Mendelian variations are mediated by non-synonymous, splice site and frameshift variations, exomes remain the most ideal regions to be screened in order to link genetic variation to health and disease. There are clear studies suggesting WES leading to the identification of the causative variant for Mendelian diseases. Most of the mutations (80-90%) that we know to cause Mendelian diseases are reported to be in the exonic region. In 2011, the scientific community was surprised to see a 4-year-old being saved from a life-threatening condition whose intestinal disorder was finally diagnosed after sequencing his exome. Once investigators identified the gene that caused the critical symptoms, a bone marrow transplant was done to save his life [8]. In such cases; we cannot employ conventional diagnostic techniques as we need to find the right cause at the right time. After this success story, we have seen WES being very widely and extensively used in diagnosing novel syndromes and finding novel mutations for known disease phenotypes. Further, WES was also employed in the diagnosis of young patients who may not show the full spectrum of the disease. WES always had an upper hand over the conventional diagnostic techniques [9-12] with a great influence on what we call a personalized medicine, where even treatment plans are altered based on the genetic profile of an individual. WES also had an advantage of being a totally non-invasive technique that could give confirmed diagnosis compared to other traditional techniques [9,13,14]. The data generated from WES throughout the period from different research groups have also added a numerous number of novel variation/mutations to existing large databases of known SNPs, known pathogenic variants, and control genomes. The Exome Aggregation Consortium (ExAC) has come up with a database exclusively for exome sequences of more than 60,000 unrelated individuals, which is freely available (Exome Aggregation Consortium (ExAC)). Such a database could provide researchers with a large set of reference exomes that they could compare with their novel findings.

*Correspondence to: Mohthash Musambil, Department of Genetics, Strategic Center for Diabetes Research, King Saud University, Riyadh, Saudi Arabia, E-mail: mmusambil@dsrcenter.org, mohthashmusambil@gmail.com

Table1. Milestone research studies that employed WES in finding the causative variant

Milestone research studies that employed WES in finding the causative variant.	
Disease name	**Reference**
Amyotrophic Lateral Sclerosis	Johnson, J. O., J. Mandrioli, M. et. al(**2010**) *Neuron 68: 857–864.*
Kaposi Sarcoma	Byun, M., A. Abhyankar, V.et.al (**2010**) *J. Exp. Med. 207: 2307–2312.*
Leber congenital amaurosis	Wang, H., X. Chen, L. et.al (**2011**) *Mol. Vis. 17:3529–3540.*
High Myopia	Shi, Y., Y. Li, et.al (**2011**) *PLoS Genet. 7: e1002084.*
Maturity-Onset Diabetes Of The Young(MODY)	Johansson, S., H. Irgens, K. K. et.al (**2012**) *PLoS One 7: e38050.*
Alzheimer Disease	Sassi, C., R. Guerreiro, R et.al (**2014**) *Neurobiol Aging 35:2422.e13-6.*
Autosomal Recessive Polycystic Kidney Disease	Xu, Y., B. Xiao, W. T. et.al (**2014**) *Gene 551: 33–38.*
Acromelic Frontonasal Dystois	Smith, J. D., A. V. Hing, C. M. et.al (**2014**) *Am. J. Hum.Genet. 95: 235–240.*
Cancer Predisposition Mutations	Yan et .al. (**2011**) *Nat. Genet. 43: 309–315*; Greif et .al. (2012) *Blood 120: 395–403*; Snape et .al. (**2012**) *Breast Cancer Res. Treat. 134: 429–433*; Kiiski et .al. (**2014**) *Proc. Natl. Acad. Sci. USA 111: 15172– 15177*; Cai et .al. (**2015**) *Endocr. Relat. Cancer 22: 23–33.*
Unexplained Intellectual Disability And/Or Developmental Delay	Braun, D. A., Schueler, M., et.al (**2016**) *Kidney international*, 89(2), 468-475.

Table 2. Major bench top platforms used for WES

Major bench top platforms used for WES		
Company	**Platforms/machine name**	**Read length**
Roche	454 GS junior	400-500 bp/35Mb
Illumina	MiSeq	2 × 300 bp/15Gb
	MiSeq Dx	2 × 300 bp/15Gb
	MiSeq FGx	2 × 300 bp/15Gb
	NextSeq 500	2 × 500 bp/120Gb
	NextSeq 550	2 × 500 bp/120Gb
Life Technologies	Ion Proton	200 bp/10Gb

Now being talked about the advantages of WES till here, it is very much necessary to talk about its limitations too. WES cannot be applied for all the genetic diseases as random because the technology in some cases fails to detect all the genes present even inside a single exon. Exons that are located within repeats sequences and out towards the chromosome tips are again undetected by WES. Structural variations (SVs), which are also an important causative factor for Mendelian disease, are sadly not identified easily by WES. Mitochondrial gene mutations are yet another causative factor that cannot be detected by WES. Triplet repeat disorders, such as Friedreich's ataxia and Huntington's are undetected by WES radar. Very importantly genes in introns are not detected by WES as they target only exons. The other areas where WES fails are in detecting epigenetic factors, mosaic mutations, uniparental disomy, mutations in repetitive or high GC rich region and mutations in genes with corresponding pseudogenes or other highly homologous sequences. Moreover, WES does not target 100% of the genes in the human genome; approximately 97% of exons are targeted [7].

Even though we have all these disadvantages for WES, until a new methodology is being discovered to detect genetic disorders (simple/complex) better and faster than WES, we don't have any other choice or alternative than using the best available technology (WES) that could save lives and administer changes in the treatment plan. The great value of WES has been discussed with two important examples, one that had led to the discovery of the causative mutation in a known gene behind an unusual disease such as in "Nicholas Volker", the saved Pulitzer boy; [8] and the other in identifying mutations in novel genes, as 2 of the 7 children being saved in the Duke University clinic by fast and rapid diagnosis using WES [15].

References

1. International hapmap consortium (2005) A haplotype map of the human genome. *Nature* 437: 1299-1320. [Crossref]

2. Cortes A, Brown MA (2011) Promise and pitfalls of the Immunochip. *Arthritis Res Ther* 13: 101. [Crossref]

3. Kivisild T, Bamshad MJ, Kaldma K, Metspalu M, Metspalu E, et al. (1999) Deep common ancestry of indian and western-eurasian mitochondrial DNA lineages. *Current Biology* 9: 1331-1334. [Crossref]

4. Roychoudhury AK (1979) The genetics of Indian populations. *Science and Culture Calcutta* 45: 87-91.

5. Watson M (2014) Illuminating the future of DNA sequencing. *Genome Biol* 15: 108. [Crossref]

6. Warr A, Robert C, Hume D, Archibald A, Deeb N, et al. (2015) Exome sequencing: Current and future perspectives. *G3 (Bethesda)* 5: 1543-1550. [Crossref]

7. Biesecker LG, Shianna KV, Mullikin JC (2011) Exome sequencing: the expert view. *Genome Biol* 12: 128. [Crossref]

8. Worthey EA, Mayer AN, Syverson GD, Helbling D, Bonacci BB, et al. (2011) Making a definitive diagnosis: successful clinical application of whole exome sequencing in a child with intractable inflammatory bowel disease. *Genetics in Medicine* 13: 255. [Crossref]

9. Iglesias A, Anyane-Yeboa K, Wynn J, Wilson A, Truitt Cho M, et al. (2014) The usefulness of whole-exome sequencing in routine clinical practice. *Genetics in Medicine* 16: 922-931. [Crossref]

10. Xu Y, Xiao B, Jiang WT, Wang L, Gen HQ, et al. (2014) A novel mutation identified in PKHD1 by targeted exome sequencing: guiding prenatal diagnosis for an ARPKD family. *Gene* 551: 33-38. [Crossref]

11. Sassi C, Guerreiro R, Gibbs R, Ding J, Lupton MK, et.al. (2014) Exome sequencing identifies 2 novel presenilin 1 mutations (p. l166v and p. s230r) in British early-onset alzheimer's disease. *Neurobiology oaging* 35: e13-e16. [Crossref]

12. Bras JM, Singleton AB (2011) Exome sequencing in parkinson's disease. *Clin Genet* 80: 104-109. [Crossref]

13. Grossmann V, Tiacci E, Holmes AB, Kohlmann A, Martelli MP, et al. (2011) Whole-exome sequencing identifies mutations of BCOR in acute myeloid leukemia with normal karyotype. *Blood* 118: 6153-6163. [Crossref]

14. Rabbani B, Mahdieh N, Hosomichi K, Nakaoka H, Inoue I (2012) Next-generation sequencing: impact of exome sequencing in characterizing Mendelian disorders. *Journal of human genetics* 57: 621. [Crossref]

15. Braun DA, Schueler M, Halbritter J, Gee HY, Porath JD, et al. (2016) Whole exome sequencing identifies causative mutations in the majority of consanguineous or familial cases with childhood-onset increased renal echogenicity. *Kidney Int* 89: 468-475. [Crossref]