Minireview *Highlight article*

Whole-genome sequencing as a first-tier diagnostic framework for rare genetic diseases

Haseeb Nisar[1,2,*], Bilal Wajid[3,4,5,*], Samiah Shahid[6,], Faria Anwar[7,**], Imran Wajid[4,8,**], Asia Khatoon[2,**], Mian Usman Sattar[9,***] and Saima Sadaf[2,***]**

[1]Office of Research, Innovation and Commercialization, University of Management and Technology, Lahore 54000, Pakistan; [2]School of Biochemistry & Biotechnology, University of the Punjab, Lahore 54000, Pakistan; [3]Department of Electrical Engineering, University of Engineering and Technology, Lahore 54000, Pakistan; [4]Ibn Sina Research & Development Division, Sabz-Qalam, Lahore 54000, Pakistan; [5]Department of Computer Sciences, University of Management and Technology, Lahore 54000, Pakistan; [6]Institute of Molecular Biology and Biotechnology, The University of Lahore, Lahore 54000, Pakistan; [7]Out Patient Department, Mayo Hospital, Lahore 54000, Pakistan; [8]Institute of Social Sciences, Istanbul Commerce University, Istanbul, Turkey; [9]Department of Management Sciences, Beaconhouse National University, Beaconhouse National University, Lahore 54000, Pakistan

Corresponding authors: Bilal Wajid. Email: bilalwajidabbas@hotmail.com; Haseeb Nisar. Email: Haseeb.nisar@umt.edu.pk

*These authors should be regarded as joint first authors.

**These authors should be regarded as joint second authors.

***These authors should be regarded as joint third authors.

Impact statement

Rare diseases affect nearly 300 million people globally with most patients aged five or less. Traditional diagnostic approaches have provided much of the diagnosis; however, there are limitations. For instance, simply inadequate and untimely diagnosis adversely affects both the patient and their families. This review is very important in the current time because the sequencing technologies are rapidly changing and the use of WGS as a diagnostic test is becoming more practical and feasible to solve the increasing number of undiagnosed rare diseases. This review differentiates current sequencing schemes concerning their cost per sample, types of variants detected, sequencing depth along with its pros and cons. Additionally, it advocates the use of WGS in clinical settings for the diagnosis of rare genetic diseases by showcasing five case studies where utilizing the technique has helped in providing relief to patients via correct diagnosis followed by the use of precision medicine.

Abstract

Rare diseases affect nearly 300 million people globally with most patients aged five or less. Traditional diagnostic approaches have provided much of the diagnosis; however, there are limitations. For instance, simply inadequate and untimely diagnosis adversely affects both the patient and their families. This review advocates the use of whole genome sequencing in clinical settings for diagnosis of rare genetic diseases by showcasing five case studies. These examples specifically describe the utilization of whole genome sequencing, which helped in providing relief to patients via correct diagnosis followed by use of precision medicine.

Keywords: Rare genetic disease, whole genome sequencing, whole-exome sequencing, precision medicine, next generation sequencing, national health systems

Experimental Biology and Medicine **2021; 246: 2610–2617. DOI: 10.1177/15353702211040046**

Introduction

A genetic disorder is defined as an abnormality caused by either single (monogenic), multiple (polygenic) gene mutations or chromosomal abnormalities. Monogenic disorders are mostly Mendelian in nature. They usually arise during the development of the fetus, making them visible at birth and are diagnosed based on family history. Regretfully, most monogenic disorders last specific treatment. In contrast, polygenic mutations are multifactorial in nature, showing their signs and symptoms due to a combined effect of multiple polymorphic genes in combination with external environmental factors. Diseases that occur due to polygenic mutations are called "complex" diseases. They are non-Mendelian in nature, show reduced penetrance, yet

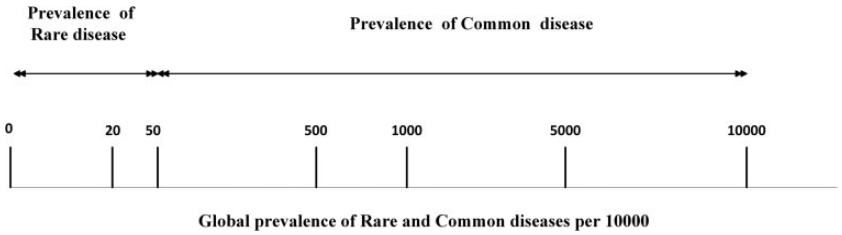

Figure 1. Worldwide prevalence of rare and common disease per 100,000 people. Most rare diseases have low prevalence varying from 10 to 50 per 100,000. On the other hand, common diseases occur more frequently, ranging from 50 to 10,000 per 100,000. (Data adapted from January 2020 Orphanet and World Health Organization (WHO) reports.)

occur more frequently than single-gene disorders. Here, the effect is more gradual, with disease symptoms appearing at a later stage of life. The most significant difference between single and complex gene disorders pertains to the degree to which genetic mutations alter the phenotype.

Rare and common diseases are defined as "rare" or "common" based on their relative prevalence (Figure 1). Rare diseases affect nearly 300 million people worldwide.[1] They vary significantly across different parts of the world each with different mutations, phenotype, and diagnostic methods (Table S1). As per Orphanet, as of 2021, there are about 7000 rare diseases with genetic causes, leading to nearly 80% of all cases. Regretfully, symptoms are often misrepresented leading to incorrect diagnosis and delay in therapy. Moreover, rare diseases are often severe with most of them incurable. As patients affected by rare diseases are few, research in disease diagnostics and therapeutics has not reached its true potential, rendering immense suffering to the patient and their families. Nevertheless, accurate and timely diagnosis is necessary because it helps physicians manage their patients as well as counsel their families.[2,3] Hence, this work focuses on "rare genetic diseases."

In 2011, the International Rare Diseases Research Consortium (IRDRC) started with the aim to provide accurate diagnosis and suitable therapy to rare diseases in the shortest possible time.[4,5] According to IRDRC, since 2010 more than 800 novel rare diseases have been reported with close to 4000 associated genes. As rare genetic diseases are not easily identified on phenotypes, determining the exact mutation causing the genetic disease is necessary. Traditional genetic diagnosis involves both conventional screenings such as chromosomal microarray (CMA) as well as screening entire exomes and genomes to determine the exact cause of the disease.[6] Hence, this review advocates the use of whole-genome sequencing (WGS) for diagnosing rare genetic diseases; enumerates both traditional and WGS based screening frameworks; specifies six case studies where WGS successfully identified rare genetic diseases which were previously undetected via conventional sequencing; and finally highlights the importance of WGS as a first-tier test with a caution on potential hurdles that need to be resolved before bringing it fully into a clinic setting.

Traditional genetic screening

First introduced in the 19th century, conventional genetic screening starts with G-banded karyotyping which helped in identifying chromosomal abnormalities in number, translocations, inversions, or amplifications of chromosomal segments. The process starts by treating metaphase chromosomes with trypsin enzyme causing the chromatin structure to relax. Thereafter, mitotic cells arrested in the metaphase stage of the cell cycle are stained with Giemsa dye producing between 400 and 800 different bands (G-banding) distributed across 23 pairs of human chromosomes. The banding pattern that is numbered on each arm of the chromosome from centromere to telomere is easily identified and any structural chromosomal changes are described accordingly. Examples of its diagnostic capability include showing trisomy 21 leading to Down syndrome and an extra X-chromosome causing Klinefelter syndrome.[7,8] However, karyotyping is limited in its scope because it is unable to detect chromosomal changes smaller than three million base pairs (Mb).[9]

First introduced in 1935, a technique called fluorescence in situ hybridization (FISH) was introduced showing better sensitivity as compared with its predecessor karyotyping.[10,11] Its applications included prenatal screening to detect aneuploidy, suspected malignancies, gene rearrangements, and deletions close to telomeres (as in the case of leukemia). However, it too had limitations in its diagnostic capacity, as FISH exhibited low resolution (300 kb) noticing only those chromosomal locations that were specifically targeted by FISH probes.[12]

Introduced in 1993, chromosome microarray analysis (CMA) replaced both FISH and karyotyping, as it enabled the detection of submicroscopic variations not detected by conventional techniques. The principal behind CMA involves the isolation of genomic DNA of both healthy control and a diseased individual. The two genomes are enzymatically broken down, differentially labeled with different fluorochromes, and co-hybridized on a microscopic glass slide to which cloned DNA segments from a representative genome are immobilized.[9] Copy number variations (CNVs) along the length of chromosomes are detected by measuring the differences in fluorescence signals and normalized to compare data between patient and control samples.[13]

CMA facilitates the diagnosis of novel rare diseases, as it detects CNVs particularly in neonates suffering from congenital birth defects.[14] Nevertheless, CMA is unable to detect small chromosomal rearrangements and somatic mosaicism. Therefore, with limitations still unaddressed, conventional gene discovery necessitated the movement towards next-generation sequencing for diagnosing rare genetic diseases as highlighted in Table 1.

Table 1. Genetic screening tests while diagnosing rare disease.

Diagnostic type	Methodology	Resolution	No. of loci screened	Variants detected	Diagnostic yield
Traditional genetic screening	G-band karyotyping	10Mb	500	Larger than 5Mb	Low
	FISH	>300kb	More than 300	Gene rearrangements, aneuploidy and malignancies	Low
	CMA	100kb	2 million	CNVs	Medium
Next generation sequencing	Targeted sequencing	20kb	40 million	SNVs in coding region	High
	WES	1bp	50 million	Variants in exonic regions	High
	WGS	1bp	3 billion	Variants throughout the genome	Highest

Note: Genetic screening tests ranges from analyzing chromosome via light microscope to detecting copy number variation to detecting specific coding regions to the full genome. With increase in resolution, the number of variants detected also increases. WGS detects the highest number of variants by covering the entire genome showcasing the largest diagnostic yield making it an ideal technique for detecting variants left undiscovered from traditional techniques.

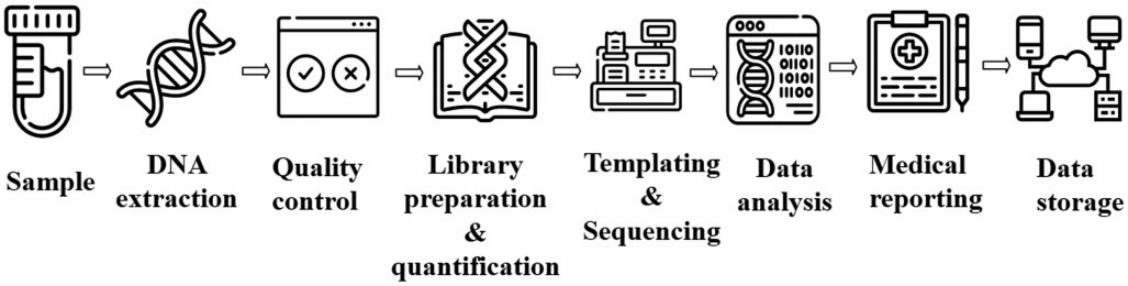

Figure 2. General NGS Workflow starting from sample collection till data storage. The process starts from DNA extraction from samples followed by quality control, library preparation, and subsequent sequencing by a sequencing machine. If the sequencing process completes successfully, bioinformaticians conduct appropriate analysis as per need.

NGS-based screening

Next-generation sequencing (NGS) uses high throughput sequencing technologies to sequence (i) coding regions of targeted genes to (ii) entire exomes and (iii) genomes. The general steps for NGS analysis are depicted in Figure 2. With rapidly decreasing sequencing cost and the advent of long-read sequencing, genomic medicine has allowed clinicians to devise new strategies for prevention, diagnosis, and therapy of rare genetic diseases.[15] The first report of using NGS in diagnosing a rare disorder called Freeman-Sheldon syndrome came in 2009 by identifying the MYH3 gene as a causative agent,[16] followed closely by identifying disease-causing genes for both Miller syndrome[17] and Kabuki syndrome,[18] both of which were not possible via conventional screening methods.

NGS has played a pivotal role in identifying more than 180 pathogenic mutations,[19] including heterozygous mutations where only a single copy of mutant is present in homologous chromosome.[20] NGS-based screening methods include (i) targeted, (ii) whole-exome, and (iii) whole-genome sequencing.[21]

Targeted gene panels

Targeted sequencing is performed by either hybridization-based targeted enrichment or PCR-based amplicon sequencing. It is favored because (i) it provides quick results because it screens only known pathogenic variants to known disease gene, (ii) can detect rare genetic variants at high sequencing depth (500 → 1000×), and (iii) costs less, missing only 10% of the mutations identified by whole-exome sequencing (WES).[22] Nevertheless, targeted gene panel sequencing is unable to (i) detect genetic heterogeneity and novel (ii) genetic causes, and (iii) mechanisms of the disease.

WES

WES covers the entire coding region (exome), which makes up to 2% of the genome. The process involves enrichment of coding regions of the genome, regulatory regions, and other functionally annotated regions of interest such as miRNA. It has been successfully employed to identify genetic causes for neurological disorders,[23] intellectual disability (ID),[24] and autism spectrum disorders[25] to name a few. It is popular because of its (i) low cost, (ii) relative abundance of pathogenic mutations in protein-coding regions, (iii) easy data storage, and (iv) processing. However, as WES only targets the exome (2% of the genome), it is unable to capture pathogenic variations that occur in the remaining 98% leaving us with WGS to look forward to.

WGS

By sequencing the entire genome, WGS can potentially detect all pathogenic variations. Gradually, WGS is becoming an effective first-tier test in cases where physicians face diagnostic ambiguity. General steps for WGS remain the same as that of WES as shown in Figure 3. When compared, WGS outperforms WES with similar coverage, as WGS (i) is less sensitive to GC content; (ii) provides more uniform coverage; (iii) capable of identifying both exome and non-coding pathogenic variants; (iv) suited to detect SNPs, CNVs, inversions, indels (in case of small read WGS),

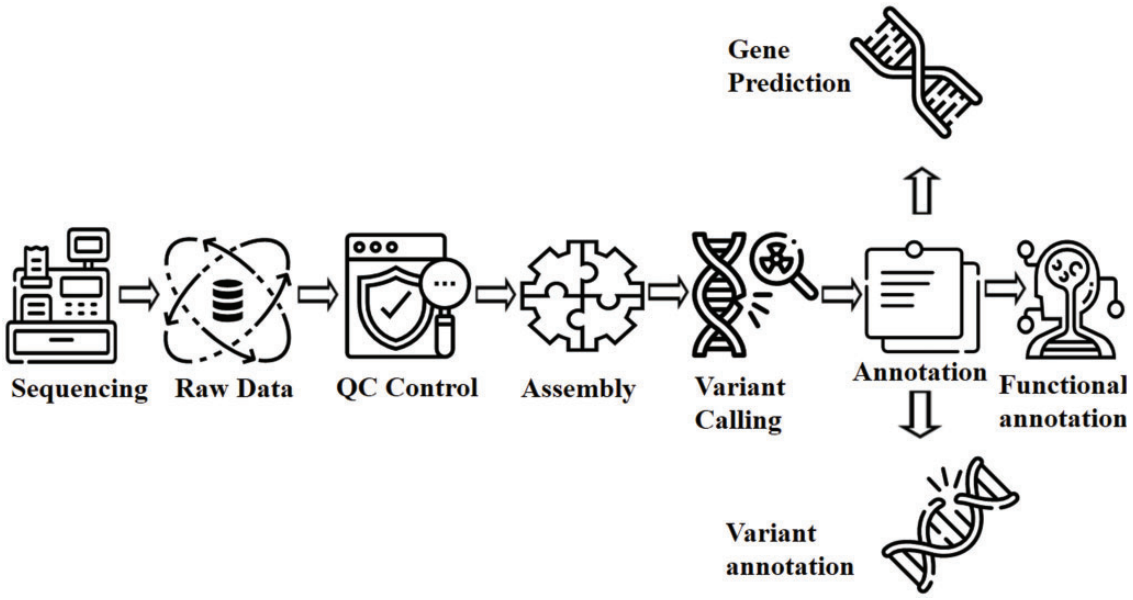

Figure 3. Flow chart for WGS analysis pipeline. Raw data generated from the sequencing process undergoes extensive cleaning and quality control. Thereafter, the filtered reads are joined together via *de-novo* or comparative assembly to form contiguous sequences (contigs). Contigs are connected via scaffolding to obtain draft assemblies. Thereafter, the assembled genome is searched for variants and annotated for identifying gene locations, determining the function of those genes and quantifying the impact of variation on proteins. Readers may employ either Genobuntu (Abbas WA, Genobuntu Package for Next Generation Sequencing. http://sourceforge.net/projects/genobuntu/) or Baari, both providing sufficient tools and software for the entire analysis pipeline.[26–30]

whereas long read WGS can recognize chromosomal rearrangements like tandem repeats[31–34]; (v (effective in trio-based screening[35,36]; (vi) proficient in detecting long repetitive regions (as in the case of Oxford Nanopore and PacBio) helping to diagnose tandem-repeat diseases, (vii) determine structural changes and transposable elements (TE) insertions (in case of Oxford Nanopore and PacBio)[37–42]; and (viii) senses SNVs and large-scale deletions in mitochondrial genome-causing disorders[43,44] like Kearns-Sayre syndrome, Pearson's syndrome,[45] and Addison disease.[46]

Figure 3 summarizes the series of interconnected analyses referred to as "pipeline" in the WGS process. Whereas Table 2 outlines the merits of WGS compared with WES, and Table 3 delineates some case studies where WGS proved more effective than WES and conventional genetic screening.

Diagnosis of some rare diseases based on WGS

Batten's disease

Batten's disease, also called Juvenile Neuronal Ceroid Lipofuscinosis, is primarily caused by a mutation in the CLN3 gene.[47] Batten's disease has an autosomal recessive mode of inheritance with initial symptoms that include sudden onset of blindness, ataxia, dysarthria, dysphagia, and seizures.

In a case study, magnetic resonance imaging (MRI) images of a six-year-old patient's head revealed cerebral and cerebellar atrophy, whereas skin biopsy showed an irregular pattern of lysosomal inclusions. Targeted gene panel revealed heterozygous single known pathogenic mutation in *MFSD8* gene with no other mutation. Thereafter, medical experts conducted a trio-based WGS screening to reveal a group of 2 kb SVA (SINE-VNTR-*Alu*) insertion in *MFSD8* intronic region located in both the patient and her mother, thereby changing *MSD8* splicing and translation effect.[48] This successful diagnosis via WGS helped develop Milasen (a 22-nucleotide antisense oligonucleotide) as personalized antisense oligonucleotide therapy[49,50] with its repeated injections for one year resulted in improvement of patient's condition by reducing the number of seizures by more than 50%.[51]

Pulmonary arterial hypertension

Hereditary pulmonary arterial hypertension (PAH) is a rare disorder characterized by blockage of arterioles in the lung, leading to pulmonary vascular resistance.[52] Earlier, PAH was assumably caused by an injury to smooth blood vessels of the lung. However, this alone could not account for 15–20% of inherited cases of PAH. Later, a study in 2010 found a mutation in BMPR2 gene.[53] Still, some blanks needed filling. It was only when researchers applied WGS that four additional causative genetic variations for PAH were found in ATP13A3, AQP1, SOX17, and GDF2 genes.[54]

Atypical hemolytic uremic syndrome

Atypical hemolytic uremic syndrome (aHUS) is a rare disorder characterized by features of thrombocytopenia, non-immune microangiopathic hemolytic anemia, and acute renal failure.[55] Diagnosing aHUS, without a family history, is difficult, as the exact cause of genetic alteration remains unidentified. WES analysis discovered mutations in at least seven genes with *CFH* gene being termed the most dominant factor linked to aHUS.

Table 2. Comparison of targeted gene sequencing, WES, and WGS.

Region sequenced	Targeted sequencing Selected genes/gene sections	WES Entire exome	WGS Entire genome
Cost per sample (USD) as of 2020	$21	$50	$1000–1600
Variants detected	Depends upon panel size	~20,000	~4,000,000
Sequencing depth	300 → 1000×	100 → 200×	30 → 60×
Methodology	Targeted enrichment using hybridization-based protocol; PCR-based amplicon sequencing	Exome enrichment	PCR-free library preparation
Pros	Low cost, short duration, high coverage for rare variants, customizable	Low cost, identifies majority of mutations in protein coding regions	Identifies novel mutations in both coding and non-coding regions; detects structural and copy number variants, uniform sequencing depth
Cons	Limited to selected genes, requires database to be regularly updated as new genes are discovered, unable to detect CNVs and SNPs	Unable to detect variants in intronic regions and SVs	High cost, large data storage and its processing required, complex data analysis

Note: Comparison of NGS techniques (i) targeted gene sequencing, (ii) WES, and (iii) WGS.
CNV: copy number variant; SNP: single nucleotide polymorphism; SV: structural variants.

Table 3. List of case studies where WGS was used as a diagnostic test.

S. no.	Rare disease	Sex	Gene	Mode of inheritance	Genomic variant
1	Battens disease	F	MFSD8	AR	• c.1102G → C NM_152778.3 • 2 kb SVA insertion
2	Pulmonary arterial hypertension	M, F	ATP13A3, AQP1, SOX17, GDF2	X linked, AR, Ht	• c.583 C > T (p.R195W) • c.527T > A(p.Val176Glu) • c.411C>G(p.Y137*)
3	Atypical hemolytic uremic syndrome	M, F	CFH, MCP, CFI, CFB, C3, THBD, DGKE	AR	c.888 + 40A>G (intronic)
4	Niemann-Pick type C disease	M	NPC1	AR	c.2713 C > T (p.Gln905Ter)
5	Dopa (3,4 dihydroxyphenylalanine) – responsive dystonia	M, F	SPR	Ht	• c.448A>G NM_003124 • c.751A>T NM_003124

Note: Examples of rare disease where WGS was employed as the genetic diagnostic test.
F: female; M: male; AR: autosomal recessive; Ht: heterozygous.

A study inducted two unrelated families and conducted both WES and WGS screening. While WES was unable to link any significant mutation in previously known genes, WGS was able to determine a non-coding mutation (c.888 + 40A>G) in DGKE gene resulting in a disrupted form of DGKE mRNA, thereby adversely affecting protein catalytic sites. These WGS-driven results had direct implications on clinical management of the disease as physicians stopped administering both plasma therapy and eculizumab (a drug commonly used to treat aHUS), as both seemed to have no link with the causative agent DGKE gene.

Niemann-Pick type C disease

Niemann-Pick type C disease (NPC) is a rare autosomal recessive disorder,[56] characterized by intracellular cholesterol trafficking, neurological disorder, reduction in bile flow and liver abnormalities due to lipid accumulation within liver cells (specifically, hepatocytes). Previously, mutations in NPC1 gene addressed up to 95% of the affected families.[57] This was also confirmed in a case study where an infant (male) showed features indicative of NPC, even though the child's parents were not cousins. After liver biopsy and electron microscopy, it was only WGS that confirmed the mutation in NPC1 gene as the causative agent of NPC disease.[58] This enabled the physicians to employ appropriate therapies to (i) delay neurodegeneration and (ii) prevent irreversible damage to the patient's neurons.[59]

Dopa (3,4-dihydroxyphenylalanine) responsive dystonia

Dopa (3,4-dihydroxyphenylalanine) responsive dystonia (DRD), also known as Segawa syndrome, is a heterogeneous rare inherited movement disorder,[60] where the patient's lower limb muscles contract uncontrollably. Patients with DRD lack enzymes involved in dopamine syntheses like GTP cyclohydrolase 1 (GTP-CH-I) or sepiapterin reductase. Previous studies on DRD showed an autosomal dominant mutation in *GCH1* gene as the leading cause of DRD.[61]

However, a study that investigated the entire genome of fraternal twins diagnosed with DRD, revealed heterozygous mutations in the SPR gene. This mutation reduces the synthesis of tetrahydrobiopterin, an essential cofactor required for the synthesis of dopamine and serotonin. This WGS-driven finding had immediate implications on clinical therapy, as physicians administered L-dopa therapy to both fraternal twins. The therapy helped improve movement coordination, enhance sleep and focus, boost exercise capability, and reduce the frequency of laryngeal spasms.[62]

Application of WGS on national healthcare systems

As an important milestone, WGS has shown significant potential when applied to large cohort studies involving both Swedish and UK's healthcare systems. For instance, the Karolinska Institutet in Sweden conducted an extensive study involving 4437 patients under the "clinical academic" collaborative model using WGS. This clinical–academic model proved promising, as their collaboration resulted in determining the cause of rare genetic diseases in nearly 1200 patients of which 54% (~650 patients) were previously undiagnosed using previous diagnostic frameworks.[63]

In another study, the UK's healthcare system applied WGS to 13,037 cohort participants, of whom 9802 were diagnosed to have rare diseases. Out of the 9802 patients, WGS was able to determine the genetic causes of 1138 patients showing the effectiveness of the framework concerning rare genetic diseases.[64]

Conclusions

Rare diseases are chronic and often life-threatening, hence requiring accurate and timely diagnosis both for disease management and personalized therapy. The review advocates employing WGS as a first-tier genetic screening test for rare genetic diseases. With the increase in unsolved cases following WES, more disease-associated genes and variants remain to be explored. This is because most of the knowledge about disease-causing variants revolves around the coding region but less is known about the role of non-coding and structural variants. This calls for alternate approaches such as sequencing the entire genome, third-generation long-read sequencing, and transcriptome sequencing. Clinical WGS promises to deliver its potential in disease management, accurate diagnosis, and solving unknown cases which remains a burden for both patients and healthcare workers. WGS could hugely impact pediatric genomics as a study diagnosed rare genetic diseases in two critically ill newly born children within 50 h of WGS screening as a first-tier test.[65]

Clinical WGS could pave the way for designing personalized therapy for the patient and providing enough information for genetic counselors to guide affected families regarding the risks of genetic mutations running through generations. Due to rapidly decreasing sequencing costs, WGS is becoming more accessible and an important genetic screening test for rare diseases.

Despite its potential, some hurdles need to be solved. This includes (i) the availability of powerful computing systems with (ii) appropriate bioinformatics programs coupled with (iii) technical personnel that can read and interpret data from a clinical standpoint. Moreover, as the data is huge, (iv) storage and transfer of raw data files is challenging, and although regulatory bodies like the American College of Medical Genetics and Genomics have published guidelines for employing WGS in clinical settings, (vi) individual interpretation is still quite varied.

To address some of these challenges, the Medical Genome Initiative was formed with the (initial) goal to publish recommended clinical and laboratory practices for applying WGS into medicine.[66] It is important to develop, constantly update, and maintain a detailed rare mutation database to facilitate diagnostic/prognostic studies across different parts of the world. Moreover, it is imperative to study epigenetics, transcriptome, proteome, and functional analysis of the genome for an improved understanding of the disease mechanism for us to devise targeted therapies. Nevertheless, with rapidly decreasing sequencing costs and an intensely collaborative approach, WGS is expected to become a standard first-tier approach for diagnosing rare genetic diseases.

AUTHORS' CONTRIBUTIONS

HN participated in the design, interpretation and writing of the manuscript; BW supervised the entire study, including writing, proof-reading and editing the article; SS assisted in developing figures; FA highlighted clinical application; IW, AK, US and SS helped edit the manuscript and provided valuable suggestions. All authors have read and agreed to the submitted version of the article.

DECLARATION OF CONFLICTING INTERESTS

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

FUNDING

The paper has been partly supported by Sabz-Qalam, Grant # SQ-2019-Bioinfo-1.

ORCID iD

Haseeb Nisar https://orcid.org/0000-0002-1123-430X

REFERENCES

1. Simone Baldovino MD, Domenica Taruscio MD, Dario Roccatello MD. Rare diseases in Europe: from a wide to a local perspective. *Isr Med Assoc J* 2016;**18**:359–63
2. Bick D, Jones M, Taylor SL, Taft RJ, Belmont J. Case for genome sequencing in infants and children with rare, undiagnosed or genetic diseases. *J Med Genet* 2019;**56**:783–91
3. Wakap SN, Lambert DM, Olry A, Rodwell C, Gueydan C, Lanneau V, Murphy D, Le Cam Y, Rath A. Estimating cumulative point prevalence of rare diseases: analysis of the orphanet database. *Eur J Hum Genet* 2020;**28**:165–73
4. Austin CP, Cutillo CM, Lau LPL, Jonker AH, Rath A, Julkowska D, Thomson D, Terry SF, de Montleau B, Ardigò D. Future of rare diseases research 2017–2027: an IRDiRC perspective. *Clin Transl Sci* 2018;**11**:21–7

5. Dawkins HJS, Draghia-Akli R, Lasko P, Lau LPL, Jonker AH, Cutillo CM, Rath A, Boycott KM, Baynam G, Lochmüller H. Progress in rare diseases research 2010–2016: an IRDiRC perspective. *Clin Transl Sci* 2018;**11**:11–20
6. Hartley T, Balcı TB, Rojas SK, Eaton A, Canada CR, Dyment DA, Boycott KM. The unsolved rare genetic disease atlas? An analysis of the unexplained phenotypic descriptions in OMIM®. *Am J Med Genet C Semin Med Genet* 2018;**178**:458–63
7. Jacobs PA, Strong JA. A case of human intersexuality having a possible XXY sex-determining mechanism. *Nature* 1959;**183**:302–3
8. Lejeune J. Study of somatic chromosomes from 9 mongoloid children. *CR Hebd Seances Acad Sci* 1959;**248**:1721–2
9. Shaffer LG, Bejjani BA. Medical applications of array CGH and the transformation of clinical cytogenetics. *Cytogenet Genome Res* 2006;**115**:303–9
10. Wiegant J, Ried T, Nederlof PM, Ploeg MVd, Tanke HJ, Raap AK. In situ hybridisation with fluoresceinated DNA. *Nucl Acids Res* 1991;**19**:3237–41
11. Langer PR, Waldrop AA, Ward DC. Enzymatic synthesis of biotin-labeled polynucleotides: novel nucleic acid affinity probes. *Proc Natl Acad Sci U S A* 1981;**78**:6633–7
12. Beliveau BJ, Joyce EF, Apostolopoulos N, Yilmaz F, Fonseka CY, McCole RB, Chang Y, Li JB, Senaratne TN, Williams BR. Versatile design and synthesis platform for visualizing genomes with oligopaint FISH probes. *Proc Natl Acad Sci USA* 2012;**109**:21301–6
13. Maciejewski JP, Mufti GJ. Whole-genome scanning as a cytogenetic tool in hematologic malignancies. *Blood* 2008;**112**:965–74
14. Lu X, Shaw CA, Patel A, Li J, Cooper ML, Wells WR, Sullivan CM, Sahoo T, Yatsenko SA, Bacino CA. Clinical implementation of chromosomal microarray analysis: summary of 2513 postnatal cases. *PLoS One* 2007;**2**:e327
15. Boycott KM, Vanstone MR, Bulman DE, MacKenzie AE. Rare-disease genetics in the era of next-generation sequencing: discovery to translation. *Nat Rev Genet* 2013;**14**:681–91
16. Ng SB, Turner EH, Robertson PD, Flygare SD, Bigham AW, Lee C, Shaffer T, Wong M, Bhattacharjee A, Eichler EE. Targeted capture and massively parallel sequencing of 12 human exomes. *Nature* 2009;**461**:272–6
17. Ng SB, Buckingham KJ, Lee C, Bigham AW, Tabor HK, Dent KM, Huff CD, Shannon PT, Jabs EW, Nickerson DA. Exome sequencing identifies the cause of a mendelian disorder. *Nat Genet* 2010;**42**:30–5
18. Ng SB, Bigham AW, Buckingham KJ, Hannibal MC, McMillin MJ, Gildersleeve HI, Beck AE, Tabor HK, Cooper GM, Mefford HC. Exome sequencing identifies MLL2 mutations as a cause of kabuki syndrome. *Nat Genet* 2010;**42**:790–3
19. Cheng AY, Teo Y-Y, Ong RT-H. Assessing single nucleotide variant detection and genotype calling on whole-genome sequenced individuals. *Bioinformatics* 2014;**30**:1707–13
20. Beck TF, Mullikin JC, Biesecker LG. Systematic evaluation of sanger validation of next-generation sequencing variants. *Clin Chem* 2016;**62**:647–54
21. Bhattacharjee A, Sokolsky T, Wyman SK, Reese MG, Puffenberger E, Strauss K, Morton H, Parad RB, Naylor EW. Development of DNA confirmatory and high-risk diagnostic testing for newborns using targeted next-generation DNA sequencing. *Genet Med* 2015;**17**:337–47
22. Saudi Mendeliome Group. Comprehensive gene panels provide advantages over clinical exome sequencing for mendelian diseases. *Genome Biol* 2015;**16**:134
23. Soden SE, Saunders CJ, Willig LK, Farrow EG, Smith LD, Petrikin JE, LePichon J-B, Miller NA, Thiffault I, Dinwiddie DL. Effectiveness of exome and genome sequencing guided by acuity of illness for diagnosis of neurodevelopmental disorders. *Sci Transl Med* 2014;**6**:265ra168
24. Rump P, Jazayeri O, van Dijk-Bos KK, Johansson LF, van Essen AJ, Verheij JBGM, Veenstra-Knol HE, Redeker EJW, Mannens MMAM, Swertz MA. Whole-exome sequencing is a powerful approach for establishing the etiological diagnosis in patients with intellectual disability and microcephaly. *BMC Med Genomics* 2015;**9**:1–9
25. Tammimies K, Marshall CR, Walker S, Kaur G, Thiruvahindrapuram B, Lionel AC, Yuen RKC, Uddin M, Roberts W, Weksberg R. Molecular diagnostic yield of chromosomal microarray analysis and whole-exome sequencing in children with autism spectrum disorder. *Jama* 2015;**314**:895–903
26. Wajid B, Sohail MU, Ekti AR, Serpedin E. The A, C, G, and T of genome assembly. *Biomed Res Int* 2016;**2016**:6329217
27. Wajid B, Serpedin E, Nounou M, Nounou H. MARAGAP: a modular approach to reference assisted genome assembly pipeline. *Ijcbdd* 2015;**8**:226–50
28. Wajid B, Serpedin E. Do it yourself guide to genome assembly. *Brief Funct Genom* 2016;**15**:1–9
29. Wajid B, Serpedin E, Nounou M, Nounou H. Optimal reference sequence selection for genome assembly using minimum description length principle. *EURASIP J Bioinform Syst Biol* 2012;**2012**:1–11
30. Wajid B, Serpedin E. Review of general algorithmic features for genome assemblers for next generation sequencers. *Genomics Proteomics Bioinformatics* 2012;**10**:58–73
31. Belkadi A, Bolze A, Itan Y, Cobat A, Vincent QB, Antipenko A, Shang L, Boisson B, Casanova J-L, Abel L. Whole-genome sequencing is more powerful than whole-exome sequencing for detecting exome variants. *Proc Natl Acad Sci USA* 2015;**112**:5473–78
32. Meynert AM, Ansari M, FitzPatrick DR, Taylor MS. Variant detection sensitivity and biases in whole genome and exome sequencing. *BMC Bioinformatics* 2014;**15**:1–11
33. Meienberg J, Bruggmann R, Oexle K, Matyas G. Clinical sequencing: is WGS the better WES? *Hum Genet* 2016;**135**:359–62
34. Boycott KM, Hartley T, Biesecker LG, Gibbs RA, Innes AM, Riess O, Belmont J, Dunwoodie SL, Jojic N, Lassmann T. A diagnosis for all rare genetic diseases: the horizon and the next frontiers. *Cell* 2019;**177**:32–37
35. Schwarze K, Buchanan J, Fermont JM, Dreau H, Tilley MW, Taylor JM, Antoniou P, Knight SJL, Camps C, Pentony MM. The complete costs of genome sequencing: a microcosting study in cancer and rare diseases from a single center in the United Kingdom. *Genet Med* 2020;**22**:85–94
36. Lionel AC, Costain G, Monfared N, Walker S, Reuter MS, Hosseini SM, Thiruvahindrapuram B, Merico D, Jobling R, Nalpathamkalam T. Improved diagnostic yield compared with targeted gene sequencing panels suggests a role for whole-genome sequencing as a first-tier genetic test. *Genet Med* 2018;**20**:435–43
37. Wright CF, Fitzgerald TW, Jones WD, Clayton S, McRae JF, Van Kogelenberg M, King DA, Ambridge K, Barrett DM, Bayzetinova T. Genetic diagnosis of developmental disorders in the DDD study: a scalable analysis of genome-wide research data. *Lancet* 2015;**385**:1305–14
38. Cumming SA, Hamilton MJ, Robb Y, Gregory H, McWilliam C, Cooper A, Adam B, McGhie J, Hamilton G, Herzyk P. De novo repeat interruptions are associated with reduced somatic instability and mild or absent clinical features in myotonic dystrophy type 1. *Eur J Hum Genet* 2018;**26**:1635–47
39. Mitsuhashi S, Nakagawa S, Ueda MT, Imanishi T, Frith MC, Mitsuhashi H. Nanopore-based single molecule sequencing of the D4Z4 array responsible for facioscapulohumeral muscular dystrophy. *Sci Rep* 2017;**7**:1–8
40. Mizuguchi T, Suzuki T, Abe C, Umemura A, Tokunaga K, Kawai Y, Nakamura M, Nagasaki M, Kinoshita K, Okamura Y. A 12-kb structural variation in progressive myoclonic epilepsy was newly identified by long-read whole-genome sequencing. *J Hum Genet* 2019;**64**:359–68
41. Merker JD, Wenger AM, Sneddon T, Grove M, Zappala Z, Fresard L, Waggott D, Utiramerur S, Hou Y, Smith KS. Long-read genome sequencing identifies causal structural variation in a mendelian disease. *Genet Med* 2018;**20**:159–63
42. Gonçalves A, Oliveira J, Coelho T, Taipa R, Melo-Pires M, Sousa M, Santos R. Exonization of an intronic LINE-1 element causing Becker muscular dystrophy as a novel mutational mechanism in dystrophin gene. *Genes (Basel)* 2017;**8**:253
43. Wallace DC. Mitochondrial DNA variation in human radiation and disease. *Cell* 2015;**163**:33–38
44. Krishnan KJ, Reeve AK, Samuels DC, Chinnery PF, Blackwood JK, Taylor RW, Wanrooij S, Spelbrink JN, Lightowlers RN, Turnbull DM. What causes mitochondrial DNA deletions in human cells? *Nat Genet* 2008;**40**:275–9

45. Jacobs LJ, Jongbloed RJ, Wijburg FA, de Klerk JB, Geraedts JP, Nijland JG, Scholte HR, de Coo IF, Smeets HJ. Pearson syndrome and the role of deletion dimers and duplications in the mtDNA. *J Inherit Metab Dis* 2004;**27**:47–55
46. Duran GP, Martinez-Aguayo A, Poggi H, Lagos M, Gutierrez D, Harris PR. *Large mitochondrial DNA deletion in an infant with Addison disease*. JIMD Reports-Case and Research Reports, 2011/3. Berlin: Springer, 2011, pp.5–9.
47. Wisniewski KE, Zhong N, Philippart M. Pheno/genotypic correlations of neuronal ceroid lipofuscinoses. *Neurology* 2001;**57**:576–81
48. Ray DA, Batzer MA. Reading TE leaves: new approaches to the identification of transposable element insertions. *Genome Res* 2011;**21**:813–20
49. Finkel RS, Chiriboga CA, Vajsar J, Day JW, Montes J, De Vivo DC, Yamashita M, Rigo F, Hung G, Schneider E. Treatment of infantile-onset spinal muscular atrophy with nusinersen: a phase 2, open-label, dose-escalation study. *Lancet* 2016;**388**:3017–26
50. Finkel RS, Mercuri E, Darras BT, Connolly AM, Kuntz NL, Kirschner J, Chiriboga CA, Saito K, Servais L, Tizzano E. Nusinersen versus sham control in infantile-onset spinal muscular atrophy. *N Engl J Med* 2017;**377**:1723–32
51. Kim J, Hu C, Moufawad El Achkar C, Black LE, Douville J, Larson A, Pendergast MK, Goldkind SF, Lee EA, Kuniholm A. Patient-customized oligonucleotide therapy for a rare genetic disease. *N Engl J Med* 2019;**381**:1644–52
52. Thenappan T, Ormiston ML, Ryan JJ, Archer SL. Pulmonary arterial hypertension: pathogenesis and clinical management. *Bmj* 2018;**360**: j5492
53. Portillo K, Santos S, Madrigal I, Blanco I, Paré C, Borderías L, Peinado VI, Roca J, Milà M, Barberà JA. Study of the BMPR2 gene in patients with pulmonary arterial hypertension. *Arch Bronconeumol* 2010;**46**:129–34.
54. Gräf S, Haimel M, Bleda M, Hadinnapola C, Southgate L, Li W, Hodgson J, Liu B, Salmon RM, Southwood M, Machado RD, Martin JM, Treacy CM, Yates K, Daugherty LC, Shamardina O, Whitehorn D, Holden S, Aldred M, Bogaard HJ, Church C, Coghlan G, Condliffe R, Corris PA, Danesino C, Eyries M, Gall H, Ghio S, Ghofrani H-A, Gibbs JSR, Girerd B, Houweling AC, Howard L, Humbert M, Kiely DG, Kovacs G, MacKenzie Ross RV, Moledina S, Montani D, Newnham M, Olschewski A, Olschewski H, Peacock AJ, Pepke-Zaba J, Prokopenko I, Rhodes CJ, Scelsi L, Seeger W, Soubrier F, Stein DF, Suntharalingam J, Swietlik EM, Toshner MR, van Heel DA, Vonk Noordegraaf A, Waisfisz Q, Wharton J, Wort SJ, Ouwehand WH, Soranzo N, Lawrie A, Upton PD, Wilkins MR, Trembath RC, Morrell NW. Identification of rare sequence variation underlying heritable pulmonary arterial hypertension. *Nat Commun* 2018;**9**:1416
55. Noris M, Remuzzi G. Atypical hemolytic–uremic syndrome. *N Engl J Med* 2009;**361**:1676–87
56. Vanier MT, Duthel S, Rodriguez-Lafrasse C, Pentchev P, Carstea ED. Genetic heterogeneity in Niemann-Pick C disease: a study using somatic cell hybridization and linkage analysis. *Am J Hum Genet* 1996;**58**:118–25
57. Vanier MT. Niemann-Pick disease type C. *Orphanet J Rare Dis* 2010;**5**:1–18
58. Hildreth A, Wigby K, Chowdhury S, Nahas S, Barea J, Ordonez P, Batalov S, Dimmock D, Kingsmore S; RCIGM Investigators. Rapid whole-genome sequencing identifies a novel homozygous NPC1 variant associated with Niemann–Pick type C1 disease in a 7-week-old male with cholestasis. *Cold Spring Harb Mol Case Stud* 2017;**3**:a001966
59. Liu B, Turley SD, Burns DK, Miller AM, Repa JJ, Dietschy JM. Reversal of defective lysosomal transport in NPC disease ameliorates liver dysfunction and neurodegeneration in the npc1−/− mouse. *Proc Natl Acad Sci USA* 2009;**106**:2377–82
60. Fink JK, Ravin PD, Filling-Katz M, Argoff CE, Hallett M. Clinical and genetic analysis of progressive dystonia with diurnal variation. *Arch Neurol* 1991;**48**:908–11
61. Furukawa Y, Guttman M, Sparagana SP, Trugman JM, Hyland K, Wyatt P, Lang AE, Rouleau GA, Shimadzu M, Kish SJ. Dopa-responsive dystonia due to a large deletion in the GTP cyclohydrolase I gene. *Ann Neurol* 2000;**47**:517–20
62. Bainbridge MN, Wiszniewski W, Murdock DR, Friedman J, Gonzaga-Jauregui C, Newsham I, Reid JG, Fink JK, Morgan MB, Gingras M-C. Whole-genome sequencing for optimized patient management. *Sci Transl Med* 2011;**3**:87re3–87re3
63. Stranneheim H, Lagerstedt-Robinson K, Magnusson M, Kvarnung M, Nilsson D, Lesko N, Engvall M, Anderlid BM, Arnell H, Johansson CB, Barbaro M, Björck E, Bruhn H, Eisfeldt J, Freyer C, Grigelioniene G, Gustavsson P, Hammarsjö A, Hellström-Pigg M, Iwarsson E, Jemt A, Laaksonen M, Enoksson SL, Malmgren H, Naess K, Nordenskjöld M, Oscarson M, Pettersson M, Rasi C, Rosenbaum A, Sahlin E, Sardh E, Stödberg T, Tesi B, Tham E, Thonberg H, Töhönen V, von Döbeln U, Vassiliou D, Vonlanthen S, Wikström AC, Wincent J, Winqvist O, Wredenberg A, Ygberg S, Zetterström RH, Marits P, Soller MJ, Nordgren A, Wirta V, Lindstrand A, Wedell A. Integration of whole genome sequencing into a healthcare setting: high diagnostic rates across multiple clinical entities in 3219 rare disease patients. *Genome Med* 2021;**13**:40
64. Turro E, Astle WJ, Megy K, Gräf S, Greene D, Shamardina O, Allen HL, Sanchis-Juan A, Frontini M, Thys C, Stephens J, Mapeta R, Burren OS, Downes K, Haimel M, Tuna S, Deevi SVV, Aitman TJ, Bennett DL, Calleja P, Carss K, Caulfield MJ, Chinnery PF, Dixon PH, Gale DP, James R, Koziell A, Laffan MA, Levine AP, Maher ER, Markus HS, Morales J, Morrell NW, Mumford AD, Ormondroyd E, Rankin S, Rendon A, Richardson S, Roberts I, Roy NBA, Saleem MA, Smith KGC, Stark H, Tan RYY, Themistocleous AC, Thrasher AJ, Watkins H, Webster AR, Wilkins MR, Williamson C, Whitworth J, Humphray S, Bentley DR, Kingston N, Walker N, Bradley JR, Ashford S, Penkett CJ, Freson K, Stirrups KE, Raymond FL, Ouwehand WH; NIHR BioResource for the 100,000 Genomes Project. Whole-genome sequencing of patients with rare diseases in a national health system. *Nature* 2020;**583**:96–102
65. van Diemen CC, Kerstjens-Frederikse WS, Bergman KA, de Koning TJ, Sikkema-Raddatz B, van der Velde JK, Abbott KM, Herkert JC, Löhner K, Rump P, Meems-Veldhuis MT, Neerincx PBT, Jongbloed JDH, van Ravenswaaij-Arts CM, Swertz MA, Sinke RJ, van Langen IM, Wijmenga C. Rapid targeted genomics in critically ill newborns. *Pediatrics* 2017;**140**:e20162854
66. Marshall CR, Bick D, Belmont JW, Taylor SL, Ashley E, Dimmock D, Jobanputra V, Kearney HM, Kulkarni S, Rehm H. The medical genome initiative: moving whole-genome sequencing for rare disease diagnosis to the clinic. *Genome Med* 2020;**12**:1–4