

Exome and genome sequencing for diagnosing patients with suspected rare genetic disease

Go Hun Seo*¹ and Hane Lee¹

Medical Genetics Division, 3billion, Inc., Seoul, Korea

Rare diseases, even though defined as fewer than 20,000 in South Korea, with over 8,000 rare Mendelian disorders having been identified, they collectively impact 6-8% of the global population. Many of the rare diseases pose significant challenges to patients, patients' families, and the healthcare system. The diagnostic journey for rare disease patients is often lengthy and arduous, hampered by the genetic diversity and phenotypic complexity of these conditions. With the advent of next-generation sequencing technology and clinical implementation of exome sequencing (ES) and genome sequencing (GS), the diagnostic rate for rare diseases is 25-50% depending on the disease category. It is also allowing more rapid new gene-disease association discovery and equipping us to practice precision medicine by offering tailored medical management plans, early intervention, family planning options. However, a substantial number of patients remain undiagnosed, and it could be due to several factors. Some may not have genetic disorders. Some may have disease-causing variants that are not detectable or interpretable by ES and GS. It's also possible that some patient might have a disease-causing variant in a gene that hasn't yet been linked to a disease. For patients who remain undiagnosed, reanalysis of existing data has shown promises in providing new molecular diagnoses achieved by new gene-disease associations, new variant discovery, and variant reclassification, leading to a 5-10% increase in the diagnostic rate. More advanced approach such as long-read sequencing, transcriptome sequencing and integration of multi-omics data may provide potential values in uncovering elusive genetic causes.

Key words: Rare diseases, Exome sequencing, Whole genome sequencing.


Introduction

The definition of a rare disease is set by each country. In the United States and South Korea, a rare disease is defined as one that affects fewer than 2,000 and 20,000 people, respectively [1-3]. Over 80% of the rare diseases are reported to have a genetic basis and are commonly referred to as rare Mendelian diseases [4]. To date, more than 8,000 Mendelian disorders have been documented [2,5]. While the prevalence of each of these diseases is rare, collectively, there are millions of individuals and

families worldwide affected by rare diseases which equates to 6-8% of the global population. Many rare diseases do not have a cure and need long-term clinical care and management. This not only places a heavy personal burden on patients in terms of their medical, financial, and psychosocial well-being but also has a substantial public health and the economic impact [6].

Patients who are suspected of having a rare disease often undergo a series of testing, but often doesn't receive a clear answer and end up in a long and arduous diagnostic journey. Although there are rare diseases with such clear phenotype and no or low

Received: 15 November 2023, Accepted: 5 December 2023, Published: 31 December 2023

*Corresponding author: Go Hun Seo, M.D., Ph.D.  <https://orcid.org/0000-0003-1518-1791>

Medical Genetics Division, 3billion, Inc., 416 Teheran-ro, Gangnam-gu, Seoul 06193, Korea.

Tel: +82-70-4855-1010, E-mail: ghseo@3billion.io

Conflict of interest: The authors declare that they do not have any conflicts of interest.

© This is an open-access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/4.0/>) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

© Copyright 2023 by the Korean Society of Medical Genetics and Genomics

www.e-kjgm.org

genetic heterogeneity, a single gene or a small gene panel testing is enough to find the molecular diagnosis, most rare diseases have high phenotypic complexity and genetic diversity, making them challenging to find a molecular diagnosis by conventional single/small panel genetic testing using Sanger sequencing method [7]. Diagnostic approach for rare disease have dramatically changed with the advent of next-generation sequencing (NGS), in particular exome sequencing (ES) and genome sequencing (GS) [8-10]. ES simultaneously sequences almost all protein-coding region of nearly all genes (20,000), while GS sequences all non-coding regions as well. ES and GS achieve a diagnostic rate of approximately 25% to 50%, although this rate varies depending on the disease category [8-13]. The unbiased assessment by ES or GS significantly speeds up the process of making an accurate diagnosis, especially for patients whose symptoms are not specific [14]. It has also led to a dramatic increase in the discovery of new disease-genes association with approximately 250 new disease genes identified each year over the past decade [4]. Finally, like any other diagnostic tests, patients receiving a clear diagnosis by ES and GS will be able to receive tailored medical managements such as initiation of precise treatments, better monitoring for additional symptoms, and receiving customized family planning [15,16]. The patient may also be eligible for clinical trials. There are studies reporting how early application of ES/GS could be overall cost-effective even though the cost for ES/GS is initially high, as cost for multiple diagnostic testing and medical intervention within the undiagnosed period is dropped [17,18]. This is consistent with the American College of Medical Genetics and Genomics (ACMG) recommending ES and GS as a first-line test for patients with congenital anomalies or intellectual disability [19]. Currently, ES is widely implemented in clinical practice and GS is being adopted in national projects and research areas for the diagnosis of rare diseases [9,12,13,19]. In this review, we discuss the overall workflow of ES and GS, as well as their features, limitations, and future directions.

Indication for Exome and Genome Sequencing

For patients with symptoms that strongly point to a specific disease, a single gene test is recommended as a first-tier test. Examples include patients likely to have Down syndrome, Duchenne muscular dystrophy or cystic fibrosis [20-22]. Then there are diseases such as cardiomyopathy or hearing loss with high genetic heterogeneity but still targeted panel sequencing test may suffice although there could be cases where the patient actually has a more complex syndromic disorder that cannot be

detected by panels but is too young to have shown all symptoms or has symptoms mild or overlooked [23,24]. Therefore, an unbiased approach such as ES or GS may be more appropriate to find the molecular diagnosis more rapidly [25] and that's why recently, ES and GS are being recommended as the first-line test for patients not only with pediatric neurodevelopmental delay and/or one or more congenital anomalies, but also with seemingly single system disorders [19,25-27].

Testing Workflow of Exome and Genome Sequencing

The NGS-based genomic sequencing test workflow can be divided into wet-lab part and dry-lab part and the dry-lab part is typically divided into three stages: primary, secondary, and tertiary [28-30]. Wet-lab workflow starts with sample accessioning, genomic DNA extraction and quality check, NGS library preparation including exome capture, and sequencing. The main difference between ES and GS is that ES captures and sequences only the protein-coding regions of almost all genes which make up about 1% to 2% of the genome, whereas GS sequences the entire genome including the non-coding regions. Exome capture step involves hybridizing the sample with capture probes. There are various versions of commercially available capture probe kits [31]. It is possible to enhance the exome performance by adding custom capture probes to augment the coverage of difficult-to-sequence regions, intronic regions with known disease-causing variants and mitochondrial genome. Once the sequencing is complete, the primary analysis starts with the base call files being converted to FASTQ files while demultiplexing the samples based on the index information that was attached to each DNA fragment during the library preparation step [29].

Secondary analysis starts with aligning all sequencing reads to the human reference genome. The current human reference genome version is GRCh38. However, there are still many laboratories using GRCh37 because switching the reference version is a major task. When variants are compared between the data that was mapped to GRCh37 and GRCh38 versions, GRCh38 is in general more accurate although not always [32]. After post-alignment fine tuning including potential PCR duplicate marking and base recalibration, variants are called. Multiple variant calling programs are employed to detect different types of variants, including single nucleotide variant (SNV), small insertion/deletion (INDEL), copy number variant (CNV), structural variant (SV), repeat expansion variant, and mobile element insertion variant [28,29,33]. In general, ES can detect SNVs, small INDELS,

and large CNVs that affect more than three consecutive exons within the protein-coding region, while GS has a broader scope and can detect almost all types of variants due to continuous sequencing data of the entire genome [28,34,35].

Tertiary analysis constitutes of interpreting variants including annotation, filtering, classification, prioritization of variants, and finally identification of variants that are most likely to explain patient's phenotype [28,29]. Given the long list of variants generated by secondary analysis ranging from ~80,000 variants to ~5 million variants for ES and GS, respectively, it is critical to have an efficient yet consistent tertiary analysis algorithm with high sensitivity and specificity. All variants are classified into five groups: pathogenic (P), likely pathogenic (LP) or variants of uncertain significance (VUS), likely benign (LB), and benign (B), according to the ACMG guidelines released in 2015 and 2019 [36,37]. To assess variant pathogenicity, these guidelines take into consideration multiple factors such as population frequency of a given variant, functional impact the variant is reported to have on the protein or may have on the protein, segregation status, and how well the patient's phenotype matches to the disease associated with the gene the variant is found in [36,37]. The first filtering step typically starts by removing variants commonly found in population databases as they cannot cause a rare disease and therefore classified as B/LB. This step removes more than ~90% of all variants and therefore it is the most effective filtering. However, as this filter solely relies on the population information, and variants that are common in sub-population that are underrepresented in public database would not be filtered out as effectively and could be misclassified as P/LP/VUS when it is actually benign. After the common variants are removed, the remaining variants are classified as P/LP/VUS based on various factors mentioned above. Then, the variants are prioritized by how similar the patient's clinical symptoms are to the reported symptoms of the disease associated with the gene the variant occurred in. Since this process is labor-intensive and time-consuming, there are now variant recommendation algorithms being developed using artificial intelligence (AI) technology [38-40]. Top-k, the frequency the diagnostic variant is found in the top-k variants, is a commonly used measure to assess the performance of the algorithms. The higher the frequency is at a small number of k, the better the performance is. As the algorithms improve, variant interpretation efficiency and accuracy will be improved.

Finally, one or two variants that are most likely to have caused the patient's symptoms are selected and reported to clinicians, who then will evaluate the reported variants in the clinical con-

text to make a final diagnosis [28]. Additional phenotyping or further genetic testing may be required in case of a VUS report.

Diagnostic Yield and Technical Limitation of Exome and Genome Sequencing

The diagnostic yield of ES is reported to range from 12% to 63% depending on the patient's symptoms and onset of clinical presentation [8,10,11,13,41], while the diagnostic yield for GS ranges from 21% to 73%, also depending on the phenotypes and ages of the patients being studied [12,41,42]. A modest increase in the diagnostic yield observed with GS compared to ES is attributed to the detection of disease-causing variants that are non-coding, and complex SVs [12,41]. Even some of the coding variants may only be called by GS as there are genomic regions that are difficult to sequence by ES but not GS. These regions include GC-rich regions that suffer with low-coverage due to PCR bias [43]. Also, there could be CNVs affecting the coding regions but too small to be called by chromosomal microarray (CMA) or ES [34,35]. The resolution of CMA and ES with average depth-of-coverage of ~100× are 30-50 kb and 3 consecutive exons, respectively. Therefore, a CNV of a size less than 30kb or affecting fewer than 3 consecutive exons may be missed by both CMA and ES. Lastly, lower-level heteroplasmic mitochondrial genome variants may only be identified GS and not ES as GS typically has significantly higher mitochondrial genome coverage at the level of ~1000X while ES has much lower coverage (<100×) unless it's specifically targeted [44,45].

However, GS still has its limitations. Lower mean depth-of-coverage of ~30 to 50× achieved for GS reduces its sensitivity to identify low-level of mosaic variants in the nuclear genome [46]. A substantial number of non-coding variants identified by GS remain as of uncertain significance as they require functional studies to determine their protein consequence such as altered splicing or abnormal expression pattern [47,48]. That is why the diagnostic rate increases when GS is complemented with transcriptome sequencing (RNAseq) [47-50]. However, RNAseq is mostly performed as research because not all genes are expressed in accessible tissues [50,51].

Next Steps for Exome/Genome-Negative Patients

Despite the advances in comprehensive genomic testing, more than half of patients remain still undiagnosed [8-13,41,52]. There are several reasons for this. First, some patients may not

actually have a genetic disorder. Some patients may exhibit clinical features resembling Mendelian disease, but their conditions may have other underlying causes such as environmental factors, such as fetal alcohol syndrome or complications from preterm delivery. Secondly, due to technical limitations of short-read sequencing, certain variant types cannot be detected. There are a lot of repeat sequences in the human genome but it's difficult to align short sequence reads to these regions [53]. Also, ES and GS cannot detect methylation abnormalities, potentially leading to missed diagnoses [54]. Thirdly, interpretation of VUS with insufficient evidence to be clearly classified as either pathogenic or benign could make up to 50% of reports and patients receiving inconclusive results with VUS will remain undiagnosed [37,55]. They will have to wait until additional evidence is collected through studies such as transcriptome sequencing, functional assays, and segregation analysis in family members. At many times, performing more such analyses is not easily doable, requiring significant amount of time and resources [56-58]. Finally, it's possible that the patient might have a disease-causing variant in a gene that hasn't yet been linked to a disease. The total number of phenotypes and disease genes continues to grow and OMIM database grows by ~250 new gene-disease each year [4]. Novel gene discovery is one major reason how reanalysis of the existing ES or GS data resulting in a new diagnosis [51,58,59]. Studies have shown that reanalyzing the ES data can lead to 3-10% increase of the diagnostic rate [52,59-62]. This increase is attributed to several factors, including reclassification of variant, identification of new variant, and novel gene discoveries [52,59-62]. A VUS may be reclassified as pathogenic or LP if additional evidence such as new functional data or test results that were not available for the initial analysis becomes available [52,59-62]. For instance, a heterozygous VUS could be reclassified as LP if it is found to be *in trans* with a pathogenic variant for a recessive disorder, or if it turns out to be *de novo* for a dominant disorder after parental testing [36,37]. Advances in the bioinformatics pipeline that results in improvement in variant calling could identify a variant that was not identified before and result in a new diagnosis [52,59-62]. Patient's clinical spectrum can evolve overtime and with new phenotypic information, a previously unclear variant-disease association could become diagnostic. [52,59-62]. Finally, as mentioned above, the contribution of novel gene discoveries to diagnostic reanalysis is the largest among all factors resulting in successful reanalysis. The ACMG has published guidelines outlining how reassessment and reanalysis of ES/GS data should be performed [63]. Reference laboratories typically offer a complimentary reanalysis of

the ES or GS data six months to a year after the initial testing is done [59,64] although more frequent free reanalysis at a shorter interval could be offered [52]. Achieving a new molecular diagnosis through reanalysis is a cost-effective approach for patients undiagnosed after ES or GS, and it is expected to become a routine clinical practice [17,64].

Besides periodic reanalysis, there are several next steps that can address some of the limitations GS and ES have. Long-read genome sequencing (LRGS) is one. LRGS generates sequence reads that are significantly longer (ranging from 1 kb to several mega bases) compared to the 150-300 base pair reads generated by short-read NGS [65]. It is easier to uniquely align long reads to genomic regions with repeat sequences, allowing variant detection within these regions that could have been missed when short-read sequencing was performed [65]. LRGS could also be useful for phasing two variants that are far apart from each other [66]. In addition, LRGS can detect CpG methylation to find variants dysregulating the epigenetic machinery [54,66,67]. LRGS comes with challenges on data management, storage, and analysis, and the application of LRGS in diagnosing undiagnosed patients is currently mostly performed as research, as it proves its utility in molecular diagnosis and the cost further decreases, it may become a clinical test in a near future [65]. Even though GS and LRGS can detect almost all types of variants, it is still impossible to interpret the non-coding variants. The AI-based *in silico* prediction tools such as SpliceAI or Pangolin are useful but they are still predictions and until a functional study is performed to show abnormal splicing, the non-coding variant cannot be determined as disease-causing with confidence [68,69]. Confirmation of aberrant splicing or expression levels for non-coding variants requires RNAseq, which could increase the diagnostic rate by 7.5% to 18% depending on phenotype and tissues sampled [47-50]. The biggest challenge of RNAseq is acquiring appropriate tissue to detect the transcript-level changes. [47,50,51]. For instance, approximately 50% of the genes associated with neurodevelopmental disorders are reported to be expressed at sufficient level in whole blood, muscle, or skin-derived fibroblasts, three relatively accessible tissues. For the rest of the genes, a method to express them needs to be developed or a ultra-high coverage RNAseq would need to be performed. Lastly, the combination of multi-omics data, including metabolomics, epigenomics, and proteomics, can complement NGS and contribute to diagnosis and new disease gene discoveries [70,71].

Conclusion

ES and GS have proven to be extremely useful tools, providing comprehensive and unbiased search for a diagnostic variant. They have also led to a large number of novel gene discoveries and understanding the underlying molecular mechanisms of many diseases, which in turn enabled more patients to be diagnosed through reanalysis. However, continuous effort is needed to improve the diagnostic yield of both tests by more effectively prioritizing variants and functionally assessing the impact of each variant. The bioinformatics pipeline could also further improve to identify more variants within the complex and repetitive genomic regions. For patients who are still undiagnosed after ES or GS, periodic reanalysis of existing data in light of growing medical knowledge is essential and approaches using more advanced technologies such as LRGS, RNAseq and/or integration of multi-omics data could also be considered, particularly for patients undiagnosed even after reanalysis.

Acknowledgements

None.

Funding

No fundings to declare.

Authors' Contributions

Conception and design: GHS. Drafting the article: GHS. Critical revision of the article: GHS, HL. Final approval of the version to be published: GHS.

References

- Richter T, Nestler-Parr S, Babela R, Khan ZM, Tesoro T, Molsen E, et al.; International Society for Pharmacoeconomics and Outcomes Research Rare Disease Special Interest Group. Rare disease terminology and definitions—a systematic global review: report of the ISPOR Rare Disease Special Interest Group. *Value Health* 2015;18:906-14.
- Nguengang Wakap S, Lambert DM, Olry A, Rodwell C, Gueydan C, Lanneau V, et al. Estimating cumulative point prevalence of rare diseases: analysis of the Orphanet database. *Eur J Hum Genet* 2020;28:165-73.
- National Health Insurance Service. The statistics of national health insurance: population with health insurance coverage. [https://nhiss.nhis.or.kr/bd/ad/bdada013cv.do]
- Chong JX, Buckingham KJ, Jhangiani SN, Boehm C, Sobreira N, Smith JD, et al. The genetic basis of Mendelian phenotypes: discoveries, challenges, and opportunities. *Am J Hum Genet* 2015;97:199-215.
- Amberger JS, Bocchini CA, Scott AF, Hamosh A. OMIM.org: leveraging knowledge across phenotype-gene relationships. *Nucleic Acids Res* 2019;47:D1038-43.
- Yang G, Cintina I, Pariser A, Oehlein E, Sullivan J, Kennedy A. The national economic burden of rare disease in the United States in 2019. *Orphanet J Rare Dis* 2022;17:163.
- Shashi V, McConkie-Rosell A, Rosell B, Schoch K, Vellore K, McDonald M, et al. The utility of the traditional medical genetics diagnostic evaluation in the context of next-generation sequencing for undiagnosed genetic disorders. *Genet Med* 2014;16:176-82. Erratum in: *Genet Med* 2013;15:849.
- Lee H, Deignan JL, Dorrani N, Strom SP, Kantarci S, Quintero-Rivera F, et al. Clinical exome sequencing for genetic identification of rare Mendelian disorders. *JAMA* 2014;312:1880-7.
- Shickh S, Mighton C, Uleryk E, Pechlivanoglou P, Bombard Y. The clinical utility of exome and genome sequencing across clinical indications: a systematic review. *Hum Genet* 2021;140:1403-16.
- Yang Y, Muzny DM, Reid JG, Bainbridge MN, Willis A, Ward PA, et al. Clinical whole-exome sequencing for the diagnosis of Mendelian disorders. *N Engl J Med* 2013;369:1502-11.
- Seo GH, Kim T, Choi IH, Park JY, Lee J, Kim S, et al. Diagnostic yield and clinical utility of whole exome sequencing using an automated variant prioritization system, EVIDENCE. *Clin Genet* 2020;98:562-70.
- Smedley D, Smith KR, Martin A, Thomas EA, McDonagh EM, Cipriani V, et al. 100,000 Genomes pilot on rare-disease diagnosis in health care – preliminary report. *N Engl J Med* 2021;385:1868-80.
- Wright CF, Campbell P, Eberhardt RY, Aitken S, Perrett D, Brent S, et al.; DDD Study. Genomic diagnosis of rare pediatric disease in the United Kingdom and Ireland. *N Engl J Med* 2023;388:1559-71.
- Shashi V, Petrovski S, Schoch K, Crimian R, Case LE, Khalid R, et al. Sustained therapeutic response to riboflavin in a child with a progressive neurological condition, diagnosed by whole-exome sequencing. *Cold Spring Harb Mol Case Stud* 2015;1:a000265.
- Iglesias A, Anyane-Yeboah K, Wynn J, Wilson A, Truitt Cho M, Guzman E, et al. The usefulness of whole-exome sequencing in routine clinical practice. *Genet Med* 2014;16:922-31.
- Splinter K, Adams DR, Bacino CA, Bellen HJ, Bernstein JA, Cheattle-Jarvela AM, et al.; Undiagnosed Diseases Network. Effect of genetic diagnosis on patients with previously undiagnosed disease. *N Engl J Med* 2018;379:2131-9.
- Ferket BS, Baldwin Z, Murali P, Pai A, Mittendorf KF, Russell HV, et al. Cost-effectiveness frameworks for comparing genome and exome

- sequencing versus conventional diagnostic pathways: a scoping review and recommended methods. *Genet Med* 2022;24:2014-27.
18. Lavelle TA, Feng X, Keisler M, Cohen JT, Neumann PJ, Prichard D, et al. Cost-effectiveness of exome and genome sequencing for children with rare and undiagnosed conditions. *Genet Med* 2022;24:1349-61.
 19. Manickam K, McClain MR, Demmer LA, Biswas S, Kearney HM, Malinowski J, et al.; ACMG Board of Directors. Exome and genome sequencing for pediatric patients with congenital anomalies or intellectual disability: an evidence-based clinical guideline of the American College of Medical Genetics and Genomics (ACMG). *Genet Med* 2021;23:2029-37.
 20. Moskowitz SM, Chmiel JF, Sternen DL, Cheng E, Gibson RL, Marshall SG, et al. Clinical practice and genetic counseling for cystic fibrosis and CFTR-related disorders. *Genet Med* 2008;10:851-68.
 21. Nallamilli BR, Ankala A, Hegde M. Molecular diagnosis of Duchenne muscular dystrophy. *Curr Protoc Hum Genet* 2014;83:9.25.1-29.
 22. Driscoll DA, Gross S. Clinical practice. Prenatal screening for aneuploidy. *N Engl J Med* 2009;360:2556-62.
 23. Vona B, Müller T, Nanda I, Neuner C, Hofrichter MA, Schröder J, et al. Targeted next-generation sequencing of deafness genes in hearing-impaired individuals uncovers informative mutations. *Genet Med* 2014;16:945-53.
 24. Dellefave-Castillo LM, Cirino AL, Callis TE, Esplin ED, Garcia J, Hatchell KE, et al. Assessment of the diagnostic yield of combined cardiomyopathy and arrhythmia genetic testing. *JAMA Cardiol* 2022;7:966-74.
 25. Quaio CRDC, Obando MJR, Perazzo SF, Dutra AP, Chung CH, Moreira CM, et al. Exome sequencing and targeted gene panels: a simulated comparison of diagnostic yield using data from 158 patients with rare diseases. *Genet Mol Biol* 2021;44:20210061.
 26. Hussain HMJ, Wang M, Huang A, Schmidt R, Qian X, Yang P, et al. Novel pathogenic mutations identified from whole-genome sequencing in unsolved cases of patients affected with inherited retinal diseases. *Genes (Basel)* 2023;14:447.
 27. Kim YG, Kwon H, Park JH, Nam SH, Ha C, Shin S, et al. Whole-genome sequencing in clinically diagnosed Charcot-Marie-Tooth disease undiagnosed by whole-exome sequencing. *Brain Commun* 2023;5:fcad139.
 28. Austin-Tse CA, Jobanputra V, Perry DL, Bick D, Taft RJ, Venner E, et al.; Medical Genome Initiative. Best practices for the interpretation and reporting of clinical whole genome sequencing. *NPJ Genom Med* 2022;7:27.
 29. Koboldt DC. Best practices for variant calling in clinical sequencing. *Genome Med* 2020;12:91.
 30. Pereira R, Oliveira J, Sousa M. Bioinformatics and computational tools for next-generation sequencing analysis in clinical genetics. *J Clin Med* 2020;9:132.
 31. Seaby EG, Pengelly RJ, Ennis S. Exome sequencing explained: a practical guide to its clinical application. *Brief Funct Genomics* 2016;15:374-84.
 32. Guo Y, Dai Y, Yu H, Zhao S, Samuels DC, Shyr Y. Improvements and impacts of GRCh38 human reference on high throughput sequencing data analysis. *Genomics* 2017;109:83-90.
 33. Ewing AD. Transposable element detection from whole genome sequence data. *Mob DNA* 2015;6:24.
 34. Royer-Bertrand B, Cisarova K, Niel-Butschi F, Mittaz-Crettol L, Fodstad H, Superti-Furga A. CNV detection from exome sequencing data in routine diagnostics of rare genetic disorders: opportunities and limitations. *Genes (Basel)* 2021;12:1427.
 35. Dharmadhikari AV, Ghosh R, Yuan B, Liu P, Dai H, Al Masri S, et al. Copy number variant and runs of homozygosity detection by microarrays enabled more precise molecular diagnoses in 11,020 clinical exome cases. *Genome Med* 2019;11:30.
 36. Riggs ER, Andersen EF, Cherry AM, Kantarci S, Kearney H, Patel A, et al. Technical standards for the interpretation and reporting of constitutional copy-number variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics (ACMG) and the Clinical Genome Resource (ClinGen). *Genet Med* 2020;22:245-57. Erratum in: *Genet Med* 2021;23:2230.
 37. Richards S, Aziz N, Bale S, Bick D, Das S, Gastier-Foster J, et al.; ACMG Laboratory Quality Assurance Committee. Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. *Genet Med* 2015;17:405-24.
 38. De La Vega FM, Chowdhury S, Moore B, Frise E, McCarthy J, Hernandez EJ, et al. Artificial intelligence enables comprehensive genome interpretation and nomination of candidate diagnoses for rare genetic diseases. *Genome Med* 2021;13:153.
 39. Kim HH, Woo J, Kim DW, Lee J, Seo GH, Lee H, et al. Disease-causing variant recommendation system for clinical genome interpretation with adjusted scores for artefactual variants. *bioRxiv* 2022, in press.
 40. Owen MJ, Lefebvre S, Hansen C, Kunard CM, Dimmock DP, Smith LD, et al. An automated 13.5 hour system for scalable diagnosis and acute management guidance for genetic diseases. *Nat Commun* 2022;13:4057.
 41. Chung CCY, Hue SPY, Ng NYT, Doong PHL, Chu ATW, Chung BHY. Meta-analysis of the diagnostic and clinical utility of exome and genome sequencing in pediatric and adult patients with rare diseases across diverse populations. *Genet Med* 2023;25:100896.
 42. Lionel AC, Costain G, Monfared N, Walker S, Reuter MS, Hosseini SM, et al. 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.
43. Clark MJ, Chen R, Lam HY, Karczewski KJ, Chen R, Euskirchen G, et al. Performance comparison of exome DNA sequencing technologies. *Nat Biotechnol* 2011;29:908-14.
44. Poole OV, Pizzamiglio C, Murphy D, Falabella M, Macken WL, Bugiardini E, et al. Mitochondrial DNA analysis from exome sequencing data improves diagnostic yield in neurological diseases. *Ann Neurol* 2021;89:1240-7.
45. Davis RL, Kumar KR, Puttick C, Liang C, Ahmad KE, Edema-Hildebrand F, et al. Use of whole-genome sequencing for mitochondrial disease diagnosis. *Neurology* 2022;99:e730-42.
46. Acuna-Hidalgo R, Bo T, Kwint MP, van de Vorst M, Pinelli M, Veltman JA, et al. Post-zygotic point mutations are an underrecognized source of de novo genomic variation. *Am J Hum Genet* 2015;97:67-74.
47. Lee H, Huang AY, Wang LK, Yoon AJ, Renteria G, Eskin A, et al. Diagnostic utility of transcriptome sequencing for rare Mendelian diseases. *Genet Med* 2020;22:490-9.
48. Kremer LS, Bader DM, Mertes C, Kopajtich R, Pichler G, Iuso A, et al. Genetic diagnosis of Mendelian disorders via RNA sequencing. *Nat Commun* 2017;8:15824.
49. Frésard L, Smail C, Ferraro NM, Teran NA, Li X, Smith KS, et al. Identification of rare-disease genes using blood transcriptome sequencing and large control cohorts. *Nat Med* 2019;25:911-9.
50. Murdock DR, Dai H, Burrage LC, Rosenfeld JA, Ketkar S, Müller MF, et al. Transcriptome-directed analysis for Mendelian disease diagnosis overcomes limitations of conventional genomic testing. *J Clin Invest* 2021;131:e141500.
51. Aicher JK, Jewell P, Vaquero-Garcia J, Barash Y, Bhoj EJ. Mapping RNA splicing variations in clinically accessible and nonaccessible tissues to facilitate Mendelian disease diagnosis using RNA-seq. *Genet Med* 2020;22:1181-90.
52. Seo GH, Lee H, Lee J, Han H, Cho YK, Kim M, et al. Diagnostic performance of automated, streamlined, daily updated exome analysis in patients with neurodevelopmental delay. *Mol Med* 2022;28:38.
53. Mandelker D, Schmidt RJ, Ankala A, McDonald Gibson K, Bowser M, Sharma H, et al. Navigating highly homologous genes in a molecular diagnostic setting: a resource for clinical next-generation sequencing. *Genet Med* 2016;18:1282-9.
54. Miller DE, Sulovari A, Wang T, Loucks H, Hoekzema K, Munson KM, et al. Targeted long-read sequencing identifies missing disease-causing variation. *Am J Hum Genet* 2021;108:1436-49.
55. Bertier G, Héту M, Joly Y. Unsolved challenges of clinical whole-exome sequencing: a systematic literature review of end-users' views. *BMC Med Genomics* 2016;9:52.
56. Kwong A, Ho CYS, Shin VY, Au CH, Chan TL, Ma ESK. How does reclassification of variants of unknown significance (VUS) impact the management of patients at risk for hereditary breast cancer? *BMC Med Genomics* 2022;15:122.
57. Liu Y, Wang H, Wang X, Liu J, Li J, Wang X, et al. Prevalence and reclassification of BRCA1 and BRCA2 variants in a large, unselected Chinese Han breast cancer cohort. *J Hematol Oncol* 2021;14:18.
58. Horton C, Hoang L, Zimmermann H, Young C, Grzybowski J, Durda K, et al. Diagnostic outcomes of concurrent DNA and RNA sequencing in individuals undergoing hereditary cancer testing. *JAMA Oncol* 2023, in press.
59. Costain G, Jobling R, Walker S, Reuter MS, Snell M, Bowdin S, et al. Periodic reanalysis of whole-genome sequencing data enhances the diagnostic advantage over standard clinical genetic testing. *Eur J Hum Genet* 2018;26:740-4.
60. Liu P, Meng L, Normand EA, Xia F, Song X, Ghazi A, et al. Reanalysis of clinical exome sequencing data. *N Engl J Med* 2019;380:2478-80.
61. Fung JLF, Yu MHC, Huang S, Chung CCY, Chan MCY, Pajusalu S, et al. A three-year follow-up study evaluating clinical utility of exome sequencing and diagnostic potential of reanalysis. *NPJ Genom Med* 2020;5:37.
62. James KN, Clark MM, Camp B, Kint C, Schols P, Batalov S, et al. Partially automated whole-genome sequencing reanalysis of previously undiagnosed pediatric patients can efficiently yield new diagnoses. *NPJ Genom Med* 2020;5:33.
63. Deignan JL, Chung WK, Kearney HM, Monaghan KG, Rehder CW, Chao EC; ACMG Laboratory Quality Assurance Committee. Points to consider in the reevaluation and reanalysis of genomic test results: a statement of the American College of Medical Genetics and Genomics (ACMG). *Genet Med* 2019;21:1267-70.
64. Ewans LJ, Schofield D, Shrestha R, Zhu Y, Gayevskiy V, Ying K, et al. Whole-exome sequencing reanalysis at 12 months boosts diagnosis and is cost-effective when applied early in Mendelian disorders. *Genet Med* 2018;20:1564-74.
65. De Coster W, Weissensteiner MH, Sedlazeck FJ. Towards population-scale long-read sequencing. *Nat Rev Genet* 2021;22:572-87.
66. Mastroiosa FK, Miller DE, Eichler EE. Applications of long-read sequencing to Mendelian genetics. *Genome Med* 2023;15:42.
67. Simpson JT, Workman RE, Zuzarte PC, David M, Dursi LJ, Timp W. Detecting DNA cytosine methylation using nanopore sequencing. *Nat Methods* 2017;14:407-10.
68. Jaganathan K, Kyriazopoulou Panagiotopoulou S, McRae JF, Darbandi SF, Knowles D, Li YI, et al. Predicting splicing from primary sequence with deep learning. *Cell* 2019;176:535-48.e24.
69. Zeng T, Li YI. Predicting RNA splicing from DNA sequence using Pangolin. *Genome Biol* 2022;23:103.
70. Almeida LS, Pereira C, Aanicai R, Schröder S, Bochinski T, Kaune A, et al. An integrated multiomic approach as an excellent tool for the

- diagnosis of metabolic diseases: our first 3720 patients. *Eur J Hum Genet* 2022;30:1029-35.
71. Stenton SL, Kremer LS, Kopajtich R, Ludwig C, Prokisch H. The diagnosis of inborn errors of metabolism by an integrative "multi-omics" approach: a perspective encompassing genomics, transcriptomics, and proteomics. *J Inherit Metab Dis* 2020;43:25-35.