Review

Diagnostic strategies in patients with undiagnosed and rare diseases

Dídac Casas-Alba[1,2], Janet Hoenicka[1,3,4], Alba Vilanova-Adell[3], Lourdes Vega-Hanna[1], Jordi Pijuan[3,4], Francesc Palau[1,2,3,4,5,6] (ID)

[1]Department of Genetic Medicine, Pediatric Institute of Rare Diseases (IPER), Hospital Sant Joan de Déu, Esplugues de Llobregat 08950, Spain.
[2]European Reference Network, ERN ITHACA, Esplugues de Llobregat 08950, Spain.
[3]Laboratory of Neurogenetics and Genomic Medicine, Institut de Recerca Sant Joan de Déu, Esplugues de Llobregat 08950, Spain.
[4]Centro de Investigación Biomédica en Red de Enfermedades Raras (CIBERER), ISCIII, Madrid 28029, Spain.
[5]Clinic Institute of Medicine and Dermatology (ICMiD), Hospital Clínic, Barcelona 08036, Spain.
[6]Division of Pediatrics, University of Barcelona School of Medicine and Health Sciences, Esplugues de Llobregat 08950, Spain.

Correspondence to: Prof./Dr. Francesc Palau, Department of Genetic Medicine, Hospital Sant Joan de Déu, Passeig Sant Joan de Déu 2, Esplugues de Llobregat 08950, Barcelona, Spain. E-mail: francesc.palau@sjd.es

Abstract

Rare diseases are life-threatening or chronically debilitating conditions affecting millions of people worldwide. In many instances, the patients experience a delay in their diagnosis or remain undiagnosed despite extensive investigations carried out by specialists. There are several explanations to account for this phenomenon including the socioeconomic context and the lack of an established consensus for diagnostic testing. Nonetheless, the widespread use of genetic and genomic tests in the past decades has had a major impact on clinical reasoning paradigms, and new troves of data are constantly being generated and analyzed. This requires constantly updating tools to match the discovery rate and allow reanalysis. In this review, we summarize the latest international recommendations and guidelines to address the problem of diagnostic deficit as well as present the current diagnostic workflows. Increasing access to exome and genome sequencing technologies and biological validation, gaining insight into the interpretation of multi-omics datasets, and fostering data sharing would reduce the long diagnostic odyssey and diagnostic gap.

INTRODUCTION

In this review, we summarize the latest international recommendations and guidelines to classify undiagnosed patients as well as present the current diagnostic workflows. We depict the advanced sequencing techniques revolutionizing genetic diagnostic practices, the future of multiple omics technologies (such as epigenomics, transcriptomics, proteomics, and metabolomics), and the use of *in silico* prediction of variant pathogenicity and functional genomics.

Rare diseases (RDs) are very numerous (more than 6000), with many of them being ultra-rare. In the European Union (EU), the definition of RDs was established in the EU Regulation on orphan medicinal products (1999) as life-threatening or chronically debilitating conditions affecting no more than 5 in 10,000 individuals[1]. The American Orphan Drug Act (1983) defined RDs as disorders affecting fewer than 200,000 individuals in the United States (US)[2]. Nguengang *et al.* calculated that 71.9% of RDs are genetic (including all diseases known or suspected to be familial), and 69.9% are exclusively pediatric onset[3]. Currently, the estimated prevalence of RDs is at least 3.5%-5.9%, which equates to 263-446 million individuals worldwide[3]. Moreover, RDs pose an economic burden as direct medical costs per patient are estimated to be around 3-5 times higher than controls of the same age without RDs[4].

The high frequency with which RDs remain undiagnosed[5] is a major challenge, as reflected in one of the International Rare Diseases Research Consortium (IRDiRC) goals for 2017-2027, which encourages to achieve a diagnosis of patients within one year if their disorder is known in the medical literature[6]. The term diagnostic odyssey refers to the time between when a potential RD is noted until the final diagnosis is made, while diagnostic impasse refers to the difficulty in achieving a diagnosis after performing all currently available procedures. The diagnostic deficit is usually associated with patients with complex phenotypes, the lack of genotype/phenotype correlation, or the lack of certainty of the clinical impact of a given genetic variant. Diagnosis in RD patients not only provides answers for patients and families but also has a potential clinical impact, which includes gaining knowledge on the natural history of disease and prognosis, providing genetic counseling, guiding personalized treatments, offering patient support networks, enabling participation in research studies, informing reproductive choices, and impacting the health of relatives[7].

Many but not all patients with an undiagnosed disease have an RD[8]. Undiagnosed and rare diseases (URDs) are conditions that elude diagnosis by a referring specialist or center of expertise for a long time and despite extensive state-of-the-art investigations[9]. Graessner *et al.* estimated that around 50% of patients with RDs remain undiagnosed even in advanced expert clinical settings where genome sequencing techniques are applied routinely[5]. Available investigations vary in each socioeconomic context, and there is not a consensus list of laboratory and ancillary tests to be performed before concluding that a patient is undiagnosed. Our ability to diagnose URDs is limited by our incomplete knowledge of the natural history and clinical expression of the disease, the genotype/phenotype correlations, the full mutational spectrum associated with all RDs, and the number of unique RDs that have yet to be discovered[10]. The IRDiRC Solving the Unsolved Task Force proposes to classify URD patients into specific subsets with significant utility for optimizing diagnostic strategies [Table 1][10]. However, in many cases, it is more important to take into account the potential explanations for the diagnostic deficit before undergoing follow-up tests. In these cases, we can use an alternative classification system [Table 2].

Table 1. Clinical groupings and diagnostic strategies for patients with undiagnosed and rare diseases (adapted from the IRDiRC solving the Unsolved Task Force[10])

Clinical groupings	Recommended diagnostic strategies
Patients with clinically recognizable disorders	
No causative variant after an appropriate highly sensitive test (e.g., single-gene disorders such as neurofibromatosis type 1)	Explore new tests and computational tools (e.g., RNAseq might be useful in patients with suspected neurofibromatosis type 1 and negative conventional tests)
No identified causative variant in the context of genetic heterogeneity (e.g., retinitis pigmentosa)	Get large datasets of patients, detailed phenotypic and genomic information, and share data (data sharing of patients with a similar phenotype can help in the identification of disease-causing variants in yet-to-be-discovered disease genes)
Unsolved but recognizable disorder (e.g., VACTERL association)	Explore new tests and computational tools
Patients without clinically recognizable disorders	
Patients with syndromes without a name (SWAN), which are not recognizable as a previously described disorder	Exome or genome sequencing as first-line tests

The widespread use of genetic tests in the clinical setting has had a great impact on the paradigms of clinical reasoning in the field of clinical genetics. Classical linear reasoning, in which phenotype assessment leads to a clinical suspicion that is further confirmed or refuted by genetic testing, has increasingly been replaced by circular reasoning, where both phenotype and genotype are assessed in parallel[14,15]. Moreover, some diagnoses are only considered after the identification of a specific genetic variant. However, considering phenotype assessment and clinical suspicion is still an essential piece of the diagnostic process.

THE GENOMIC SCENARIO: EXOME AND GENOME SEQUENCING

Exome and genome sequencing (ES and GS, respectively) are now recommended by the American College of Medical Genetics and Genomics (ACMG) to be considered as first- or second-tier tests (after chromosomal microarray (CMA) or focused gene testing) for patients with congenital anomalies (CA), developmental delay (DD), or intellectual disability (ID)[16]. Only patients with clinical presentations suggestive of a specific diagnosis should undergo targeted testing first[16]. This may include patients with suspicion of a chromosomal disorder, a diagnosis in which sequencing may not be diagnostic (e.g., fragile X syndrome), or known family history of a disorder[16]. This recommendation supersedes the previous advice to perform CMA as first-line tests in patients with CA and DD/ID (setting aside patients with autism spectrum disorder without other delays)[17].

The recent change in recommendations is due to the high diagnostic yield of ES and GS. The Ontario Health Technology Assessment reported a diagnostic yield of 43% for GS and 34% for ES (not differentiating between the trio and singleton testing), compared with the diagnostic yield of 21% for standard genetic testing (which typically included CMA, candidate single-gene testing, or large gene panel testing)[18]. Although GS, by definition, captures more comprehensively all types of variants in testing, ES is expected to be used for a long time in clinical genetics given the lower cost, focused approach, and reduced burden on downstream analysis compared to GS[19]. When performing ES and GS, best practice includes trio testing if available to help contextualize rare variants, but they also can be effectively performed as singleton testing, with diagnostic yield being slightly reduced[16].

Although first-line ES or GS has still not been established as standard clinical practice in all settings, the increased use of ES in the past decade has had a significant impact in the field of URDs. If a patient remains undiagnosed after performing ES, it is essential to know the limitations of ES to continue with the diagnostic process [Table 3]. In addition to recommending ES/GS as first-tier tests (or second-tier tests in certain cases), the ACMG suggests a diagnostic algorithm for the evaluation of patients with GDD/ID if ES

Table 2. Potential explanations of the diagnostic deficit and diagnostic strategies

Potential explanations	Recommended diagnostic strategies
No variant has been identified	
• The disorder has not been recognized by the referring specialist (because of recent discovery, atypical presentation, concomitant comorbidities, and/or lack of expertise), and thus the appropriate analysis of existing data (e.g., focused on candidate genes) or the appropriate test (e.g., Angelman syndrome methylation analysis) has not been performed	• Precision phenotype assessment; includes deep phenotyping using Human Phenotype Ontology (HPO) terminology[11] • Literature review
• Changes in the classification of variants over time	• Periodic reanalysis
• Limitations of current testing methodologies [Table 3]	• Consider alternative tests/new computational tools and tests
• Difficulties in the interpretation of complex inheritance patterns (e.g., genetic modifiers and polygenic inheritance)	• No general recommendation can be made. Genome-wide association studies can be used to identify single nucleotide polymorphisms significantly associated with a complex trait (e.g., human height)[12]
A variant has been identified (adapted from Pijuan et al. 2021)[13]	
A variant has been identified in one of the known disease genes: • The genotype matches the phenotype (at least partially), but there is insufficient evidence for candidate variant (variant of unknown significance) • The genotype does not match the phenotype: - Difficulties in the interpretation of reduced penetrance and variable expressivity of known disorders - Yet-to-be discovered allelic disease*	• Precision phenotype assessment -phenotype • Literature review • *In silico* prediction of pathogenicity • Experimental functional validation studies - functional genomics
• A variant has been identified in a yet-to-be-discovered disease gene, but there is insufficient evidence for candidate variant or gene causality*	

*In any group of patients with a clinically recognizable disorder or suspected new disease, consider a large dataset of patients, detailed phenotypic and genomic information, and data sharing.

is nondiagnostic[20]. The next step would be CMA, and, in the case it is nondiagnostic, then the clinician could consider further evaluation/testing including periodic ES reanalysis every 1-3 years, fragile X syndrome testing, metabolic testing and/or mitochondrial DNA sequencing depending on clinical presentation (although mtDNA analysis is sometimes performed with ES), and karyotyping to assess for abnormal segregation of balanced chromosomal rearrangements[20].

CMA is still the standard of care for detecting copy number variants (CNVs) in clinical laboratories[17]. However, many efforts have been made to develop ES-based CNV detection tools, such as the modified ExomeDepth workflow[19]. It is expected that GS will eventually replace CMA as the gold standard for CNV detection, considering that it covers both coding and noncoding regions. The next frontier in next-generation sequencing (NGS) technologies is long-read GS, which might outperform the limitations of short-read sequencing concerning the identification of structural variants, sequencing repetitive regions, phasing of alleles, and distinguishing highly homologous genomic regions[23]. The main limitations of current long-read NGS technologies are difficulties in library preparation, higher error rate compared to short-read NGS, higher costs, and difficulties in data analysis and storage[23].

Table 3. Types of variants missed by ES and recommended diagnostic strategies[10,20-22]

Types of variants missed	Recommended diagnostic strategies
• Large CNVs (except when specifically included in analysis pipeline)	CMA GS
• Small CNVs (except when specifically included in analysis pipeline)	MLPA (targeted approach) High-resolution CMA GS
• Balanced chromosomal rearrangements	Karyotype GS
• Low coverage regions	GS
• Low-level mosaicism	Deep sequencing of multiple tissues
• Repeat expansions	Repeat expansion testing (targeted approach) Long-read GS
• Splicing mutations (synonymous, splice site, or intronic mutations)	Sanger sequencing of small fragments of genome (targeted approach) GS RNA-seq
• Regulatory DNA mutations (promoter, enhancer, and others)	GS RNA-seq
• Imprinting	Methylation arrays
• Transposable elements (retrotransposons)	New computational tools

CMA: Chromosomal microarray; CNVs: copy number variants; GS: genome sequencing; MLPA: multiplex ligation-dependent probe amplification; RNA-seq: RNA sequencing.

Genomic data reanalysis

The adoption of ES has accelerated the rate of novel gene discovery for Mendelian conditions. The annual number of discoveries of genes underlying RDs peaked between 2012 and 2015 (approximately 285 per year), and it has declined slightly thereafter[24]. This is one of the arguments in favor of periodic reanalysis of ES/GS data, but it is only the tip of the iceberg. There are several other reasons causative mutations might be unrecognized. Data are usually analyzed according to the reported patient phenotype, and key elements might not be available to the clinical laboratory or might not have emerged at the time of the first analysis[25]. The constantly growing knowledge of gene networks is a useful resource for hypothesis generation and prioritizing candidate genes. The processes of both calling variants from short sequence reads and to annotating the impact of variants are performed with imperfect and constantly evolving bioinformatics tools[25,26]. Similarly, current phenotype-driven genomic diagnostics software (which usually uses HPO terminology) and databases that collect published mutations (e.g., Human Gene Mutation Database®) are incomplete[25]. Besides, a significant amount of time is required for an expert geneticist to analyze ES data, and there are various sources of variability and bias that make it difficult to exactly replicate the analysis[25]. According to various reports, reanalysis of ES data may enhance the diagnostic yield by 10%-18.9%[25,27-29].

Deep genomic sequencing by next-generation sequencing

It is worth noting that sensitivity for detection of mosaicism can be lower on ES/GS compared with panel testing, due to greater read depth and sequence coverage[20]. When somatic mosaicism is suspected, the best diagnostic strategy is to combine deep NGS (generally using a customized panel of candidate genes associated with the phenotype) and variant search in multiple tissues[30]. The utility of this approach may be limited by difficulties in obtaining tissues other than blood, saliva or buccal mucosa, and skin fibroblasts for genetic analysis. Somatic mutations have been described in several noncancerous disorders, such as *PI3K/AKT/mTOR* pathway mutations in patients with hemimegalencephaly, *GNAQ* mutations in patients with Sturge-Weber syndrome, *GNAS* mutations in patients with McCune Albright syndrome, and *NIPBL* mutations in patients with Cornelia de Lange syndrome (approximately 30% of clinically diagnosed patients have somatic *NIPBL* mutations)[30,31]. Deep NGS has also been found to be useful for detecting somatic

mutations in patients with other brain malformations, such as double-cortex syndrome or polymicrogyria[32,33].

Classical karyotype studies

Despite new technological advances, it should be noted that chromosome analysis still has a role in the study of patients with URDs. Chromosome analysis remains particularly valuable in the detection of mosaic abnormalities; the clarification of unbalanced translocations, rings, and complex rearrangements; and the detection of balanced rearrangements[34]. Patients with *de novo* apparently balanced rearrangements might harbor cryptic deletions or potential gene disruptions, which would require further testing[35].

MULTI-OMICS APPROACHES

Massively parallel technologies appear as powerful approaches to achieve diagnosis integrated with both DNA sequencing and phenotypic data. Different biological molecules can be approached with omics techniques beyond DNA, such as RNA, metabolites, and proteins or epigenetic modifications. The analysis of these molecules offers complementary data that could help to understand diseases mechanism and provide new diagnostic tools. A general drawback of multi-omics approaches is the difficulty in obtaining the most informative tissue for the analysis when it is not blood or skin fibroblasts.

Epigenomics

Epigenomics is the analysis of the entire set of epigenetic modifications on the genetic material of a cell, generally intending to identify alterations in DNA methylation or histone modification in diagnosing specific disorders. DNA methylation is one of the most commonly studied epigenetic modifications, and aberrant DNA methylation has long been associated with several RDs, such as Angelman and fragile X syndromes. Several methods have been developed to assess genome-wide DNA methylation in peripheral blood[36-38]. These methods have proven particularly useful to diagnose patients for whom there is a clinical suspicion of an imprinting disorder or Mendelian disorders caused by mutations in genes that code for proteins involved in epigenetics (e.g., Coffin-Siris, Kabuki, and Sotos syndromes)[36,37]. The potential utility of diagnostic clinical genomic DNA methylation testing in patients with RDs has been recently reported[38]. Using a clinical genome-wide methylation test for patients with RDs, 35% of patients in a targeted cohort were positive for a diagnostic episignature versus 11% of patients in a screening cohort, proving that this approach can be applied for the untargeted study of patients with URDs[38].

Transcriptomics

The utility of analyzing the patient's transcriptome for RDs has already been proven. Single-gene RNA sequencing has been carried out to elucidate the role of candidate genetic variants of uncertain significance (VUS) by combining reverse transcriptase to create cDNA and RT-PCR to measure RNA[39]. Performing RNA sequencing (RNA-seq) of the entire transcriptome in a single run can facilitate the detection of aberrant expression, aberrant splicing, and mono-allelic expression[40]. RNA-seq has proven particularly useful in specific clinical settings, such as the analysis of muscle biopsies from patients with rare muscle disorders and cultured fibroblasts from patients with mitochondrial disorders, or recessive conditions for which only one mutation has been identified[41-43]. Some studies have assessed the utility of RNA-seq, in combination with ES/GS, as a diagnostic tool for URDs of diverse disease categories, with an additional yield of 17%-18%[44,45]. The main challenges of RNA-seq are the accessibility of appropriate tissues for mRNA extraction, the large amount of sequencing required to detect changes, and the necessity of large control cohorts[46].

Proteomics and metabolomics

The analysis of proteins and the proteome can provide information about the activity, interaction, location, and composition of protein complexes of clinical interest. On the other hand, metabolomic profiles have been used as biomarkers for disease progression and response to treatment, and they are now applied in diagnostics to determine the functional consequences of a given VUS. Mass spectrometry is the most widely used platform in proteomics and metabolomics. Proteomics is the quantitative and qualitative analysis of the entire set of proteins in a given specimen, generally to identify proteins that are consistently modified or present at abnormal concentrations in specific disorders[47]. Proteomic strategies have been applied to investigate the pathophysiology of metabolic disorders, such as methylmalonic acidemia[48,49]. Metabolomics is the quantitative and qualitative analysis of all metabolites derived from sugars, lipids, proteins, and nucleic acids in a given specimen[50]. The most successful application of targeted metabolomics analysis is the newborn screening of inborn errors of metabolism[51]. Overall, significant work needs to be done to implement proteomics or metabolomics in the untargeted study of patients with URDs in routine clinical practice.

IN SILICO BIOLOGY AND EXPERIMENTAL FUNCTIONAL STUDIES

One of the main challenges for genetic diagnosis using NGS is the interpretation of the pathogenicity of variants, particularly when the phenotype is a source of uncertainty (e.g., reduced penetrance and variable expressivity) or when a variant is classified as VUS or localizes in a yet-to-be-discovered disease gene. Indeed, during the analysis and interpretation of NGS data in a particular patient, the probability of detecting a VUS is higher than the probability of detecting a pathogenic variant[52]. The ACMG has elaborated standards and guidelines to classify genetic variants into five criteria-based categories using different types of variant evidence (e.g., population data, computational data, functional data, and segregation data)[53]. In addition, certain bioinformatic tools facilitate the classification process (e.g., Available from: Varsome, https://varsome.com/; Franklin, Available from: https://franklin.genoox.com/clinical-db; CADD, Available from: https://cadd.gs.washington.edu/), but each geneticist must provide specific knowledge to be able to classify a certain variant[13]. Sometimes, segregation studies can clear up doubts, but those variants that are not reclassified remain as VUS until further studies allow their pathogenicity to be determined.

In silico prediction of the pathogenicity of a VUS can be improved with several tools such as literature review and data mining, pathogenicity predictors, and 3D protein modeling[13]. After that, functional genomics can be used for the validation of the genetic variant through molecular and cellular experiments (e.g., subcellular localization studies, expression levels, and specific studies related to protein function)[13,54]. At Sant Joan de Déu (SJD) Hospital and Research Institute, we developed the in-house Translational Diagnostics Program (TDP) to functionally validate both candidate VUS and variants found in patients with phenotype-genotype incongruity[13,55]. The TDP uses different tools of experimental and computational biology to analyze VUS and determine the function and possible pathophysiological alteration of the encoded protein. The objective is to delineate the impact of VUS by combining four stages including the in-depth and precision phenotyping with functional genomics, for the validation of the genetic variant through molecular and cellular experiments. Using this pipeline in RDs patients, we can assist the process of reclassifying variants concerning a patient's phenotype and improve the diagnostic deficit.

OTHER STRATEGIES FOR DIAGNOSIS AND DISEASE-RELATED GENE DISCOVERY

In recent years, the greater integration of genomic science and medicine has made it possible to explore other strategies to achieve the diagnosis of patients or to define new genes as responsible for the disease. The director's board of the ACMG released a position statement on how responsible sharing of genomic variant

and phenotype data is crucial to improving genetic healthcare[56]. Data sharing is necessary to describe the key features of the phenotype of those with RDs, establish the association between genetic RDs and the causative genes, classify genomic variants, and improve standards used in variant classification[56]. Data sharing is compatible with the imperative of protecting privacy in healthcare[56]. Currently, there are online data-sharing resources such as Matchmaker Exchange, open collaboration between different platforms (including (GeneMatcher, PhenomeCentral, DECIPHER, and others) to facilitate the matching of cases with similar phenotypic and genotypic profiles (matchmaking) through standardized application programming interfaces and procedural conventions[57]. With the advent of multi-omics approaches, there is an increasing need to share data on tools and methods for data integration and interpretation. Regarding the discovery of new genes associated with disease, the analysis of genomic population data allows the evaluation of the strength of natural selection to identify genes and genomic regions that are constrained for variation compared to the expected mutation rates. This information reveals which genes are most intolerant to loss-of-function or missense variants[58]. The predicted loss-of-function intolerance (pLI) score and the lower observed/expected upper-bound fraction (LOEUF) score can be used to identify candidate haploinsufficient disease genes[7,58,59]. Another alternative strategy for the discovery of new genes is the analysis of the phenotypic effects of gene disruption using model organisms when the gene of interest is evolutionarily conserved[7]. Four related projects that use animal models are the Monarch Initiative, the Mouse Genome Database, the Knockout Mouse Project, and the International Mouse Phenotyping Consortium[7]. However, it should be noted that animal models often fail to recapitulate human disease phenotypes. Other options are the possibility of deriving human-induced pluripotent stem cells (hiPSCs) from patient cells (e.g., fibroblasts) or generating pathogenic variants in wild-type hiPSCs using editing approaches (e.g., CRISPR/Cas9 technologies), which could be coupled with the generation of 3D organoids.

CONCLUSION

We are currently witnessing how genetics and genomics of URDs are one of the fields in which precision medicine and translational research are opening new paths and opportunities in etiological diagnosis. Advances in genetic and genomic testing have markedly improved the rate and time to diagnosis of patients with URDs. However, it is important to note that all techniques have limitations. Novel multi-omics techniques are rapidly advancing toward clinical practice, and *in silico* studies and functional analyses allow us to validate the significance of the findings.

Increasing access to exome and genome sequencing technologies and biological validation, gaining insight into the interpretation of multi-omics datasets, and fostering data sharing would reduce the long diagnostic odyssey and diagnostic gap.

DECLARATIONS
Authors' contributions
Conception or design of the work, drafting the article, critical revision of the article, final approval of the version to be published: Casas-Alba D, Pijuan J
Conception or design of the work, critical revision of the article, final approval of the version to be published: Hoenicka J, Vilanova-Adell A, Vega-Hanna L, Palau F

Availability of data and materials
Not applicable.

Conflicts of interest
All authors declared that there are no conflicts of interest.

Ethical approval and consent to participate
Not applicable.

Consent for publication
Not applicable.

Copyright
© The Author(s) 2022.

REFERENCES

1. Regulation (EC) No 141/2000 of the European Parliament and the Council of 16 December 1999 on orphan medicinal products. European Parliament. Available from: https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=celex%3A32000R0141 [Last accessed on 9 June 2022].
2. National Institute of Health. Public law 97-414 97th congress. Jan 4, 1983. Available from: https://history.nih.gov/research/downloads/PL97-414.pdf [Last accessed on 9 June 2022].
3. Nguengang Wakap S, Lambert DM, Olry A, et al. Estimating cumulative point prevalence of rare diseases: analysis of the Orphanet database. *Eur J Hum Genet* 2019;28:165-73. DOI PubMed PMC
4. Tisdale A, Cutillo CM, Nathan R, et al. The IDeaS initiative: pilot study to assess the impact of rare diseases on patients and healthcare systems. *Orphanet J Rare Dis* 2021;16:429. DOI PubMed PMC
5. Graessner H, Zurek B, Hoischen A, Beltran S. Solving the unsolved rare diseases in Europe. *Eur J Hum Genet* 2021;29:1319-20. DOI PubMed PMC
6. Austin CP, Cutillo CM, Lau LPL, et al; Future of rare diseases research 2017-2027: an IRDiRC perspective. *Clin Transl Sci* 2018;11:21-7. DOI PubMed PMC
7. Seaby EG, Rehm HL, O'Donnell-Luria A. Strategies to uplift novel mendelian gene discovery for improved clinical outcomes. *Front Genet* 2021;12:674295. DOI PubMed PMC
8. Kliegman RM, Bordini BJ, Basel D, Nocton JJ. How doctors think: common diagnostic errors in clinical judgment-lessons from an undiagnosed and rare disease program. *Pediatr Clin N Am* 2017;64:1-15. DOI PubMed
9. About undiagnosed diseases. Available from: https://www.udninternational.org/schede-38-about_undiagnosed_diseases [Last accessed on 9 June 2022].
10. Boycott KM, Hartley T, Biesecker LG, et al. A diagnosis for all rare genetic diseases: the horizon and the next frontiers. *Cell* 2019;177:32-7. DOI PubMed
11. Köhler S, Carmody L, Vasilevsky N, et al. Expansion of the human phenotype ontology (HPO) knowledge base and resources. *Nucleic Acids Res* 2019;47:D1018-27. DOI PubMed PMC
12. Yang J, Benyamin B, McEvoy BP, et al. Common SNPs explain a large proportion of the heritability for human height. *Nat Genet* 2010;42:565-9. DOI PubMed PMC
13. Pijuan J, Rodríguez-Sanz M, Natera-de Benito D, et al. Translational diagnostics: an in-house pipeline to validate genetic variants in children with undiagnosed and rare diseases. *J Mol Diagn* 2021;23:71-90. DOI PubMed
14. Palau F. Medicina personalizada o de precisión: la homeostasis de la individualidad. *Rev Soc Esp Bioq Biol Mol* 2020;8-13. Available from: https://revista.sebbm.es/pdf.php?id=643 [Last accessed on 9 June 2022].
15. Palau F. Pediatric genomics and precision medicine in childhood. In: Faintuch J, Faintuch S, editors. Precision medicine for investigators, practitioners and providers. Cambridge: Academic Press, 2020; pp. 143-152. Available from: https://www.elsevier.com/books/precision-medicine-for-investigators-practitioners-and-providers/faintuch/978-0-12-819178-1 [Last

accessed on 9 June 2022].

16.   Manickam K, McClain MR, Demmer LA, et al; 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. DOI PubMed

17.   Manning M, Hudgins L; Array-based technology and recommendations for utilization in medical genetics practice for detection of chromosomal abnormalities. *Genet Med* 2010;12:742-5. DOI PubMed

18.   Health (Quality). Genome-Wide Sequencing for Unexplained Developmental Disabilities or Multiple Congenital Anomalies: A Health Technology Assessment. *Ont Health Technol Assess Ser* 2020;20:1. PubMed

19.   Rajagopalan R, Murrell JR, Luo M, Conlin LK. A highly sensitive and specific workflow for detecting rare copy-number variants from exome sequencing data. *Genome Med* 2020;12:14. DOI PubMed PMC

20.   Srivastava S, Love-Nichols JA, Dies KA, et al; Meta-analysis and multidisciplinary consensus statement: exome sequencing is a first-tier clinical diagnostic test for individuals with neurodevelopmental disorders. *Genet Med* 2019;21:2413-21. DOI PubMed

21.   Gardner EJ, Prigmore E, Gallone G, et al. Contribution of retrotransposition to developmental disorders. *Nat Commun* 2019;10:4630. DOI PubMed PMC

22.   Wang Q, Shashikant CS, Jensen M, Altman NS, Girirajan S. Novel metrics to measure coverage in whole exome sequencing datasets reveal local and global non-uniformity. *Sci Rep* 2017;7:885. DOI PubMed PMC

23.   Mantere T, Kersten S, Hoischen A. Long-read sequencing emerging in medical genetics. *Front Genet* 2019;10:426. DOI PubMed PMC

24.   Bamshad MJ, Nickerson DA, Chong JX. Mendelian gene discovery: fast and furious with no end in sight. *Am J Hum Genet* 2019;105:448-55. DOI PubMed PMC

25.   Wenger AM, Guturu H, Bernstein JA, Bejerano G. Systematic reanalysis of clinical exome data yields additional diagnoses: implications for providers. *Genet Med* 2017;19:209-14. DOI PubMed

26.   Gardner EJ, Sifrim A, Lindsay SJ, et al. Detecting cryptic clinically relevant structural variation in exome-sequencing data increases diagnostic yield for developmental disorders. *Am J Hum Genet* 2021;108:2186-94. DOI PubMed PMC

27.   Basel-Salmon L, Orenstein N, Markus-Bustani K, et al. Improved diagnostics by exome sequencing following raw data reevaluation by clinical geneticists involved in the medical care of the individuals tested. *Genet Med* 2019;21:1443-51. DOI PubMed

28.   Salinas V, Martínez N, Maturo JP, et al. Clinical next generation sequencing in developmental and epileptic encephalopathies: Diagnostic relevance of data re-analysis and variants re-interpretation. *Eur J Med Genet* 2021;64:104363. DOI PubMed

29.   Bullich G, Matalonga L, Pujadas M, et al; Systematic collaborative reanalysis of genomic data improves diagnostic yield in neurologic rare diseases. *J Mol Diagn* 2022;24:529-42. DOI PubMed

30.   McNulty SN, Evenson MJ, Corliss MM, et al. Diagnostic utility of next-generation sequencing for disorders of somatic mosaicism: a five-year cumulative cohort. *Am J Hum Genet* 2019;105:734-46. DOI PubMed PMC

31.   Krawczynska N, Wierzba J, Wasag B. Genetic mosaicism in a group of patients with cornelia de lange syndrome. *Front Pediatr* 2019;7:203. DOI PubMed PMC

32.   Jamuar SS, Lam AT, Kircher M, et al. Somatic mutations in cerebral cortical malformations. *N Engl J Med* 2014;371:733-43. DOI PubMed PMC

33.   Stutterd CA, Brock S, Stouffs K, et al. Genetic heterogeneity of polymicrogyria: study of 123 patients using deep sequencing. *Brain Commun* 2021;3:fcaa221. DOI PubMed PMC

34.   Bi W, Borgan C, Pursley AN, et al. Comparison of chromosome analysis and chromosomal microarray analysis: what is the value of chromosome analysis in today's genomic array era? *Genet Med* 2013;15:450-7. DOI PubMed

35.   Astbury C, Christ LA, Aughton DJ, et al. Detection of deletions in de novo "balanced" chromosome rearrangements: further evidence for their role in phenotypic abnormalities. *Genet Med* 2004;6:81-9. DOI PubMed

36.   Aref-Eshghi E, Schenkel LC, Lin H, et al. Clinical validation of a genome-wide DNA methylation assay for molecular diagnosis of imprinting disorders. *J Mol Diagn* 2017;19:848-56. DOI PubMed

37.   Aref-Eshghi E, Bend EG, Colaiacovo S, et al. Diagnostic utility of genome-wide DNA methylation testing in genetically unsolved individuals with suspected hereditary conditions. *Am J Hum Genet* 2019;104:685-700. DOI PubMed PMC

38.   Sadikovic B, Levy MA, Kerkhof J, et al. Clinical epigenomics: genome-wide DNA methylation analysis for the diagnosis of Mendelian disorders. *Genet Med* 2021;23:1065-74. DOI PubMed

39.   Eriksson M, Brown WT, Gordon LB, et al. Recurrent de novo point mutations in lamin A cause Hutchinson-Gilford progeria syndrome. *Nature* 2003;423:293-8. DOI PubMed

40.   Luo HM, Zhang Y, Wang XQ, et al. Identification and control of a poliomyelitis outbreak in Xinjiang, China. *N Engl J Med* 2013;369:1981-90. DOI PubMed

41.   Cummings BB, Marshall JL, Tukiainen T, et al; Improving genetic diagnosis in Mendelian disease with transcriptome sequencing. *Sci Transl Med* 2017;9:eaal5209. DOI PubMed

42.   Kremer LS, Bader DM, Mertes C, et al. Genetic diagnosis of Mendelian disorders via RNA sequencing. *Nat Commun* 2017;8:15824. DOI PubMed PMC

43. Kernohan KD, Frésard L, Zappala Z, et al. Whole-transcriptome sequencing in blood provides a diagnosis of spinal muscular atrophy with progressive myoclonic epilepsy. *Hum Mutat* 2017;38:611-4. DOI PubMed PMC

44. Lee H, Huang AY, Wang LK, et al; Diagnostic utility of transcriptome sequencing for rare Mendelian diseases. *Genet Med* 2020;22:490-9. DOI PubMed PMC

45. Frésard L, Smail C, Ferraro NM, et al; Identification of rare-disease genes using blood transcriptome sequencing and large control cohorts. *Nat Med* 2019;25:911-9. DOI PubMed

46. Hartley T, Lemire G, Kernohan KD, Howley HE, Adams DR, Boycott KM. New diagnostic approaches for undiagnosed rare genetic diseases. *Annu Rev Genomics Hum Genet* 2020;21:351-72. DOI PubMed

47. Mischak H, Allmaier G, Apweiler R, et al. Recommendations for biomarker identification and qualification in clinical proteomics. *Sci Transl Med* 2010;2:46ps42. DOI PubMed

48. Caterino M, Chandler RJ, Sloan JL, et al. The proteome of methylmalonic acidemia (MMA): the elucidation of altered pathways in patient livers. *Mol Biosyst* 2016;12:566-74. DOI PubMed PMC

49. Caterino M, Pastore A, Strozziero MG, et al. The proteome of cblC defect: in vivo elucidation of altered cellular pathways in humans. *J Inherit Metab Dis* 2015;38:969-79. DOI PubMed

50. Zhan X, Long Y, Lu M. Exploration of variations in proteome and metabolome for predictive diagnostics and personalized treatment algorithms: Innovative approach and examples for potential clinical application. *J Proteomics* 2018;188:30-40. DOI PubMed

51. Costanzo M, Zacchia M, Bruno G, Crisci D, Caterino M, Ruoppolo M. Integration of proteomics and metabolomics in exploring genetic and rare metabolic diseases. *Kidney Dis (Basel)* 2017;3:66-77. DOI PubMed PMC

52. Marcke C, Collard A, Vikkula M, Duhoux FP. Prevalence of pathogenic variants and variants of unknown significance in patients at high risk of breast cancer: A systematic review and meta-analysis of gene-panel data. *Crit Rev Oncol Hematol* 2018;132:138-44. DOI PubMed

53. Richards S, Aziz N, Bale S, et al; 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. DOI PubMed PMC

54. Klee EW, Cousin MA, Pinto E Vairo F, et al. Impact of integrated translational research on clinical exome sequencing. *Genet Med* 2021;23:498-507. DOI PubMed

55. Yubero D, Natera-de Benito D, Pijuan J, et al. The Increasing impact of translational research in the molecular diagnostics of neuromuscular diseases. *Int J Mol Sci* 2021;22:4274. DOI PubMed PMC

56. Board of Directors. Laboratory and clinical genomic data sharing is crucial to improving genetic health care: a position statement of the American College of Medical Genetics and Genomics. *Genet Med* 2017;19:721-2. DOI PubMed

57. Philippakis AA, Azzariti DR, Beltran S, et al. The matchmaker exchange: a platform for rare disease gene discovery. *Hum Mutat* 2015;36:915-21. DOI PubMed PMC

58. Karczewski KJ, Francioli LC, Tiao G, et al; The mutational constraint spectrum quantified from variation in 141; 456 humans. *Nature* 2020;581:434-43. DOI PubMed

59. Lek M, Karczewski KJ, Minikel EV, et al; Analysis of protein-coding genetic variation in 60; 706 humans. *Nature* 2016;536:285-91. DOI PubMed