Understanding Exome Sequencing: Tips for the Pediatrician

DHANYA LAKSHMI NARAYANAN AND KATTA MOHAN GIRISHA

From Department of Medical Genetics, Kasturba Medical College (Manipal Academy of Higher Education), Manipal, India. Correspondence to: Dr Girisha KM, Department of Medical Genetics, Kasturba Medical College, Manipal Academy of Higher Education, Manipal 576 104, India. girish.katta@manipal.edu

Exome sequencing is gaining popularity as a genomic test for the diagnosis of Mendelian disorders in children. It is essential for pediatricians to familiarize themselves with this technique and its interpretation. This brief review discusses some of the key components of a clinical or research report on exome sequencing for a practicing pediatrician, so as to enable them to utilize this test well and provide timely referrals to a clinical geneticist.

Keywords: *Next-generation sequencing, Sanger sequencing, Targeted gene panel, Monogenic*.

***PII*: S097475591600295**

Exome sequencing (ES) has emerged as one of the most powerful and widely available tools in the clinic for diagnosis of monogenic disorders. This has resulted in pediatricians encountering a large number of exome sequencing reports (clinical or research) in their practice. We intend to provide a brief guide to pediatricians, to interpret these reports to enable them determine the clinical relevance of the reported variants and thus help in post-test counseling. Readers may refer to earlier articles for a review of clinical indications, methodology and laboratory workflow of exome sequencing [1-4]. The relevant questions related to exome sequencing are listed here with explanations.

Was Exome Sequencing Ordered for an Appropriate Indication?

Currently, ES is indicated when a monogenic disorder (disease caused by alterations in a single gene) is suspected. It is not a test for diagnosis of aneuploidy (Trisomies 13, 18, 21), copy number variations (Di George syndrome due to 22q11.2 microdeletion), triplet repeat disorders (Fragile *X* syndrome) or multifactorial disorders (neural tube defects). There are easier, accurate and inexpensive tests for the diagnosis of beta thalassemia, spinal muscular atrophy, Duchenne muscular dystrophy, hemophilia A, achondroplasia, Apert syndrome, sickle cell disease and fragile *X* syndrome. In individuals with Duchenne muscular dystrophy, 65-80% of disease causing variants are large deletions or duplications of one or more exons in *DMD* gene [5]. Hence a test like multiplex ligation-dependent probe amplification (MLPA), which detects deletions and duplications of exons, is the investigation of choice. Variations in *HBB* gene (a small gene with only three exons), which cause sickle cell disease and beta thalassemia can be detected by Sanger sequencing of the entire gene. However, rarely ES may end the diagnostic odyssey by diagnosing or ruling out (with less confidence) a monogenic disorder.

Differences Between Trio and Singleton Exome Sequencing?

Trio exome sequencing (ES of proband and parents) is the ideal approach for diagnosis of a monogenic disorder. Additionally, other affected or unaffected family members may be included for a comprehensive analysis. Trio (or more) ES can confirm the segregation of the variant, prioritize the rare disease-causing variants and confirm *de novo* origin of variants. However, to reduce the cost, often a singleton ES (only proband) is considered. In some autosomal dominant conditions with incomplete penetrance (like Treacher Collins syndrome, due to heterozygous variant in *TCOF1* gene), a heterozygous variant identified in the proband may be identified in one of the parents who could be clinically asymptomatic or have unrecognizable mild clinical features. In such instances, it is essential to evaluate the parents in detail before disregarding the variant as benign.

What Is the Relevance of Systematic Phenotyping and Providing Clinical History and Differential Diagnoses to the Laboratory

It is crucial to know that symptoms, signs, results of investigations (haematological, biochemical or radiological) and a three-generation pedigree are used extensively to interpret the exome data by the laboratory. The laboratory would also benefit from a list of differential diagnoses provided by the clinician. Clinical validity of a disease-causing variant is assessed by the information that

the clinician provides to assess whether an identified variant is truly responsible for the phenotype of a patient. A laboratory attempts to achieve this by searching the databases of known disease-causing variants and by comparing the phenotype of previously reported patients. A pedigree helps to infer the possible mode/s of inheritance as well.

Were the Genes in Question (Your Differential Diagnoses) 'Covered' Well by Exome Sequencing?

Coverage refers to the fraction or breadth of the target region that is actually sequenced [6]. Coverage may differ from one service provider to another and most laboratories provide the coverage of genes of interest in the ES report. If a clinical diagnosis of osteogenesis imperfecta is made, then the clinician checks whether all the 20 genes known to cause this condition are covered by ES. Coverage may depend on biases in DNA sample preparation, GC content of the target region and differences in the efficiency of capture kits [6]. The term depth, which is often used interchangeably, denotes the number of times (50x, 100x, etc) a region is sequenced. A sufficient read depth of at least 20x is required to ensure sensitivity and specificity of a variant call and assessing allelic balance (especially to determine heterozygous or mosaic nature of the variant) [3]. It is the responsibility of the laboratory to be confident of calling a variant accurately. If ES does not cover certain genes or regions causing a particular phenotype, then alternate methods like Sanger sequencing may be used to sequence those regions.

What Are the Different Levels of Evidence to Imply Disease Causation of a Variant?

At the variant level, segregation of the variant in diseased individuals in a family, inheritance consistent with the proposed mode of inheritance and mechanism of disease causation (haploinsufficiency, loss of function or gain of function) and a low frequency of the variant in the population may provide evidence for disease causation. The variant may be deleterious if the variant is present in an evolutionarily conserved site or alters the protein domain essential for the function of a protein. Experimental evidence for altered quantity or function of a protein by a variant can also be considered as an evidence for causation. The variant may be considered significant if the variant is present in a gene or a pathway already known to cause a disease phenotype [2].

Is the Interpretation of the Variant Correct?

The laboratory provides a classification of the variant identified based on the recommendations of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology [7]. The variants are classified as pathogenic, likely pathogenic, benign, likely benign and variants of uncertain significance (VUS). While the first two categories confirm the diagnosis, VUS poses a huge challenge. The following points may be considered to check whether the interpretation is correct:

Is the variant already reported to cause the phenotype? Databases like Human Gene Mutation Database *(http://www.hgmd. cf.ac.uk/ac/all.php)*, Leiden Open Variation Database (LOVD) (*https://www.lovd.nl/*) or ClinVar (*https://www.ncbi.nlm.nih.gov/clinvar/*) may be used for this purpose. There are some disease-specific databases like Infevers [variant information on genes related to autoinflammatory diseases (*https://infevers.umai-montpellier.fr/web/index.phpm*)] or Osteogenesis Imperfecta Variant Database (*https://oi.gene.le.ac.uk/home.php?select_db=COL1A1*). LOVD provides access to each of these disease-specific databases.

Does the phenotype of the patient match with the gene and disease identified? It is essential for a clinician to re-examine the patient to check for certain phenotypic features, relevant to the disease in question, which might have skipped attention earlier. Information regarding specific phenotypic features of a particular condition can be obtained from databases like Online Mendelian Inheritance in Man (OMIM; *https://www.omim.org/*).

Does the inheritance pattern of the reported variant and the mechanism of disease causation match with the diagnosis? Mechanism of disease is an essential aspect to consider while evaluating the clinical relevance of a variant. For example, FGFR2 related craniosynostosis syndromes result from heterozygous gain of function variants in *FGFR2* gene. Hence identifying a loss of function variant in *FGFR2* gene may not be clinically relevant for craniosynostosis. Similarly inheritance pattern of a variant needs to be checked. For example, heterozygous variations in *FLNB* can result in Larsen syndrome (MIM#150250) which is inherited in an autosomal dominant manner. Homozygous or compound heterozygous variations in *FLNB* can cause Spondylocarpotarsal synostosis syndrome (MIM#272460), inherited in an autosomal recessive manner.

How do you proceed if a VUS is reported? In case of VUS, the clinician should ensure that adequate effort has been made by the laboratory in reviewing existing literature, disease, variant databases and segregation of variants in the family. It is recommended that VUS should be used with caution in clinical decision-making [7]. Clinical, laboratory (enzyme assay) or radiological re-evaluation of the patient may provide further evidence for reclassification of a VUS as pathogenic or benign. It is appropriate to refer a family to a clinical geneticist for

assessment before providing prenatal diagnosis based on VUS. Reassessment of patient and the variant at a later date might also help in solving the dilemma.

What Is Meant by 'Segregation of Disease-Causing Variants'?

Segregation of variants helps in assessing whether a specific variant segregates with the disease status or a particular phenotype in the family. Segregation of variants can be done by trio (or more) ES or by Sanger sequencing of the candidate variants identified in the proband in parents and other affected or unaffected family members. Hence, it is essential to provide samples of affected or unaffected family members for testing along with that of the patient.

Segregation of variants is less important in autosomal dominant conditions with recurrent or known disease causing variants (Achondroplasia and Apert syndrome). It is also less critical for conditions with well-established disease mechanisms (glycine substitutions in *COL1A1* and *COL1A2* in progressively deforming or perinatal lethal osteogenesis imperfecta). However, to prove or disprove whether a novel variant or a novel gene causes a disease phenotype, segregation is essential. If a disease-causing variant does not segregate in a family, consider reduced or incomplete penetrance, age dependent penetrance, mosaicism, subtle or mild clinical features, adoption, gamete donation or disputed paternity.

Is Sanger Sequencing Always Required to Validate the Variant Detected by Exome Sequencing?

Sanger sequencing is used by many laboratories to establish the analytical validity or accuracy of the ES. ES may result in false positive or false negative results on a few occasions. A laboratory considers parameters like depth and coverage, presence of repetitive sequences and pseudogenes before deciding to undertake analytical validation by Sanger sequencing. The laboratory should determine if the variant is called with sufficient confidence. Sanger sequencing may be done to check for segregation of variants in family members as explained above.

How to Proceed if Exome Sequencing Does Not Yield a Result?

The possible outcomes of ES are provided in **Box I**. It is necessary to check if the clinical indication was appropriate, sufficient clinical information was provided and the laboratory has considered this information while reporting the variants. A negative ES result may indicate the presence of a variant that is not usually detected like large deletions, deep intronic variants, epigenetic changes (methylation abnormalities), triplet repeat disorders, or variants in repetitive regions**.** Even though copy number variants involving exons may be identified in ES (requires a different analytical step), alternate methods like chromosomal microarray (CMA) or MLPA may be required to confirm them. It is pertinent to consider whether the disease in question is really a monogenic disorder. A negative ES report could also indicate a phenotype due to an environmental or teratogenic agent (Aicardi Goutières syndrome, a genetic disease, may resemble congenital cytomegalovirus infection) or multifactorial disease. A chromosomal disorder or microdeletion/duplication syndrome should be considered when there is global developmental delay, facial dysmorphism and major or minor malformations, and a chromosomal microarray or MLPA may be ordered. In some instances, whole genome sequencing may be required to detect variants in non-coding regions if a trio ES is non-diagnostic. Reassessing the phenotype of the patient later may be helpful to identify symptoms and signs, which may evolve with age. It is important to ensure timely referral to a clinical geneticist for expert opinion. Illustrative examples are provided in **Supplementary Table I.**

Box I Possible Outcomes of Exome Sequencing

- Identification of a single disease causing variant.
- No disease-causing variants are identified.
- Identification of multiple plausible variants causing a phenotype (multiple genetic disorders causing a blended phenotype).
- Findings that are unrelated to the phenotype being evaluated. These are known as 'secondary findings' [8].
- Variants of uncertain significance.

CONCLUSION

Pediatricians have an ever-increasing responsibility to be updated with evolving technologies in clinical practice. It is essential for a pediatrician to understand the merits and limitations of widely used ES and make timely referral to a clinical geneticist to provide the best possible care to a family with a genetic disease.

Contributors: DLN: designed and wrote the first draft of the manuscript; KMG: conceived the idea and guided drafting and revising it critically for important intellectual content. Both approve the final version to be published.
Funding: None; *Competing interest*: None stated.

REFERENCES

1. Narayanan DL, Girisha KM. Genomic Testing for Diagnosis of Genetic Disorders in Children: Chromosomal Microarray and Next-Generation Sequencing. Indian Pediatr. 2020;57:549-54.
2. MacArthur DG, Manolio TA, Dimmock DP, et al.

Guidelines for Investigating Causality of Sequence Variants in Human Disease. Nature. 2014; 508:469-76.
3. Seaby EG, Pengelly RJ, Ennis S. Exome sequencing explained: A practical guide to its clinical application. Brief Funct Genomics. 2016;15:374-84.
4. Biesecker LG, Green RC. Diagnostic clinical genome and exome sequencing. N Engl J Med. 2014; 370:2418-25.
5. Darras BT, Urion DK, Ghosh PS. Dystrophinopathies. 2000 Sep 5 [Updated 2018 Apr 26]. *In*: Adam MP, Ardinger HH, Pagon RA, et al., editors. GeneReviews® [Internet]. University of Washington; 1993-2020.
6. Sims D, Sudbery I, Ilott NE, Heger A, Ponting CP. Sequencing depth and coverage: Key considerations in genomic analyses. Nat Rev Genet. 2014;15:121-32.
7. 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.
8. Kalia SS, Adelman K, Bale SJ, et al. Recommendations for Reporting of Secondary Findings in Clinical Exome and Genome Sequencing, 2016 Update (ACMG SF v2.0): A Policy Statement of the American College of Medical Genetics and Genomics. Genet Med. 2017;19:249-55.