1-1-2016

# The Challenge of Analyzing the Results of Next-Generation Sequencing in Children.

Isabelle Thiffault
*Children's Mercy Hospital*

John Lantos
*Children's Mercy Hospital*



## Recommended Citation
Thiffault I, Lantos J. The Challenge of Analyzing the Results of Next-Generation Sequencing in Children. Pediatrics. 2016;137 Suppl 1:S3-S7. doi:10.1542/peds.2015-3731C

# The Challenge of Analyzing the Results of Next-Generation Sequencing in Children

Isabelle Thiffault, MSc, PhD[a,b,c] John Lantos, MD[a,b,d]

abstract

In recent years, next-generation sequencing technologies have revolutionized approaches to genetic studies. Whole-exome or whole-genome sequencing allows diagnoses in many patients who have complex phenotypes and unusual clinical presentations. As genomic and exomic testing expands in both the research and clinical settings, pediatricians will need to understand the technology of next-generation sequencing and the complexity of interpreting genomic variants relevant to patient phenotypic features. This article briefly explains the technology by which genomes are sequenced and discusses some of the complexity related to interpreting genomic variants. We conclude with some thoughts on the clinical applications of such testing.

[a]*Center for Pediatric Genomic Medicine, Children's Mercy Hospital and Clinics, Kansas City, Missouri;* [b]*Department of Pathology and Laboratory Medicine, Children's Mercy Hospital and Clinics, Kansas City, Missouri;* [c]*University of Missouri–Kansas City School of Medicine, Kansas City, Missouri; and* [d]*Department of Pediatrics, Children's Mercy Hospitals, Kansas City, Missouri*

Drs Thiffault and Lantos both contributed to the conceptualization of the manuscript. Dr Lantos conceptualized the article, and Dr Thiffault drafted the initial manuscript. Both authors then reviewed and revised the manuscript, approved the final manuscript as submitted, and agree to be accountable for all aspects of the work.

**DOI:** 10.1542/peds.2015-3731C

Accepted for publication Nov 10, 2015

Address correspondence to John Lantos, MD, Children's Mercy Hospital, Bioethics Center, 2401 Gillham Rd, Kansas City, MO 64108. E-mail: jlantos@cmh.edu

PEDIATRICS (ISSN Numbers: Print, 0031-4005; Online, 1098-4275).



**FINANCIAL DISCLOSURE:** The authors have indicated they have no financial relationships relevant to this article to disclose.

**FUNDING:** Research reported in this publication was supported by the Eunice Kennedy Shriver National Institute of Child Health and Human Development of the National Institutes of Health (U19HD077627). Funded by the National Institutes of Health (NIH).

**POTENTIAL CONFLICT OF INTEREST:** The authors have indicated they have no potential conflicts of interest to disclose.

**To cite:** Thiffault I and Lantos J. The Challenge of Analyzing the Results of Next-Generation Sequencing in Children. *Pediatrics.* 2016;137(s1):e20153731C

Clinical genetics is changing. Next-generation sequencing (NGS) of DNA is slowly replacing traditional technologies for the diagnosis of genetic disorders. The key difference between NGS and older technologies is a matter of precision and scale. NGS can precisely reveal multiple variations in ~19 000 genes simultaneously. The challenge with older technologies was in deciding which variations to search for. The challenge with NGS is in interpreting the meaning of those variations in the context of clinical care.

As the cost of NGS drops, it becomes feasible to use this powerful technology in clinical medical practice. In some cases, this use allows the diagnosis of rare disorders that may not be diagnosable with other methods.[1] In other cases, however, NGS generates confusing results that are difficult to interpret or use in a way that improves clinical care. Because NGS tests for so many things at once, it is unclear how to assess its sensitivity, specificity, accuracy, or clinical utility; or how to compare it with more traditional approaches.[2]

Diagnosis is a complex process. It requires an accurate medical history, a good physical examination, laboratory tests, and imaging studies. Even with these tactics, however, clinical diagnosis often remains elusive. The process of extensive testing to diagnose rare conditions has been referred to as the "diagnostic odyssey." Even when the index of suspicion of a genetic condition is very high, the process may not yield a diagnosis.[3–8]

NGS adds an additional source of data to the process of diagnosis. It is clearly changing the diagnostic paradigm. The success rate of NGS for the identification of a causative variant fluctuates considerably between studies, however.[3,5–18] Thus, many questions remain. Which patients or diseases should be prioritized for NGS analysis? Who should interpret the results and using what criteria? How can we maximize true-positive and -negative results while avoiding false ones? Answers to these questions are essential for determining how best to use NGS in conjunction with other diagnostic approaches.

To help clinicians tackle these challenges, the present article briefly explains the technology of NGS and offers our insights into how to interpret NGS data in the context of pediatric patients.

## TECHNICAL OVERVIEW OF NGS

The human genome has 6.2 billion base pairs. Of the entire genome, only ~1% codes for proteins; this 1% is called the exome. The exome contains ~85% of known or potential disease-causing variants. It is organized into ~19 000 genes, and these genes contain the code for ≥1 protein. The entire DNA content (coding and noncoding) is called the genome.

Sequencing of the entire genetic code of a person is called "whole-genome sequencing"; sequencing parts of the genome that contain genes is called "whole exome sequencing." Whole-genome and whole exome sequencing use the same laboratory processes, which begin with the extraction of DNA from cells (usually white blood cells). After extraction, the DNA is broken into small fragments. These fragments are then put through a process called library preparation, a step that is required for both the exome and genome. For the exome, an enrichment procedure is necessary to "capture" only the information of the exons (the expressed region); that is, the protein-coding regions of the genome. The enrichment step can also be used for targeted gene panels. This process allows sequencing of only preselected genes.

The sequencing instrument "reads" the genetic code of these short sequences and generates millions of short sequence reads. These short sequence reads are then aligned and matched to specific positions in the human genome reference sequence with the use of bioinformatic tools. A computerized annotation of genotype (A, C, G, or T) at each position in the exome or genome is performed. Similarities and differences between the patient's sequence and the reference sequence can then be highlighted.

To assure accuracy, the patient's genomic sequence is read multiple times. This process allows a quantification of the accuracy of the genotype at each base pair position. The next task is to determine which variations in the patient's genome, compared with the reference genome, may be clinically significant or relevant.

Variants can be classified according to frequency, type, and previous reported association with particular clinical conditions. Typically, the file is filtered for rare variants (ie, allele frequency inferior to 1% in the general population) because only rare variants are likely to be pathogenic (variants that are common in the general population seldom cause rare Mendelian disorders). Some variants are known to be benign; others are of a type that generally causes loss of function or altered function of a gene. Many variants have been previously reported to cause disease, but many others remain of unknown clinical significance. Depending on present knowledge, variant analysis is imperfect, and the variant interpretation does not imply 100% certainty. The American College of Medical Genetics and Genomics offers guidelines for variant interpretation.[19]

The yield of sequence reads is inherently uneven across the exome and genome. Typically, NGS results provide adequate coverage of 85% to 98% of the targeted sequence

regions. The biggest challenges in interpreting NGS results are not a product of the inaccuracy of the technology; they instead arise from the difficulty in interpreting the meaning of the numerous variants.

## INTERPRETATION OF VARIANTS

Variants can only be interpreted after a good clinical history, family history, and physical examination have been performed. Data from these preliminary steps allow physicians to assess whether there are similar or related phenotypes in other family members; if so, the inheritance pattern can then be evaluated and assessed.

Physical examination findings allow physicians to begin a search for potentially relevant genes. The patient should be examined for "major features" of genetic disease as well as other potentially relevant "minor features." The Human Phenotype Ontology (HPO) categorization aims to provide a standardized vocabulary of phenotypic abnormalities encountered in human disease. HPO currently contains ~11 000 terms and >115 000 annotations to hereditary diseases.[20]

Mode of inheritance and a comprehensive phenotype can then be used to review the published literature and to search relevant databases. Search engines include tools such as Google and PubMed; more specialized searching can use tools such as Orphanet, DECIPHER, and OMIM (Online Mendelian Inheritance in Man).[20] Bioinformatic tools such as Phenomizer can help develop a differential diagnosis using HPO to identify candidate diseases/disease genes that best explain a patient's set of clinical features.[20,21]

These tools allow some classification of the patient's genomic variants. NGS may lead to the discovery of a known pathogenic variant, a novel pathogenic variant that is likely to be disease-causing, or a variant of unknown clinical significance in a gene known to cause human disease. Novel variants of unknown clinical significance or apparently pathogenic variants in genes not yet known to cause human disease require additional clinical and laboratory research to judge the pathogenicity of the variants. Freely accessible Web sites such as GeneMatcher are designed to enable connections between clinicians and researchers from around the world who share an interest in the same gene or genes. This availability allows matching based on phenotypic features for individuals with novel disorders or novel clinical presentation with or without candidate genes to enable diagnosis for very rare cases.

Clinical validity is a complicated and challenging aspect of NGS. Evidence is required to prove that a specific rare variant in a particular gene, detected by NGS, is indeed pathogenic and responsible for a particular clinical phenotype.

NGS analysis is influenced by the expected inheritance patterns (autosomal dominant, recessive, or X-linked) and whether other family members are available for phenotyping and genetic testing. Biological parental testing is important when a de novo variant is suspected; if neither parent has the variant, and biologic parentage is confirmed by using rare single-nucleotide polymorphisms, the variant is confirmed as de novo. Recent studies suggest that up to 65% of diagnoses are associated with a de novo variant.[7,9,16,22]

In other situations, NGS performed in only 1 affected child, followed by genotyping of just a few variants in affected and unaffected relatives, may show cosegregation of the variant and the disease. These findings support the pathogenicity of the variant.[8,12,16,19,22]

## INTERPRETING AND REPORTING NGS RESULTS

The most challenging aspect of NGS testing is the analytic validity.[19] The highest level of analytic validity occurs when there is a variant in a gene that has been previously associated with the patient's condition and when functional test results of that gene's function exhibit abnormalities. There are, however, few functional studies of the effect of individual variants in their biological context. This limitation hampers effective and comprehensive interpretation.

The next level occurs when a variant has been previously associated with the patient's condition but no functional studies. These findings must be interpreted cautiously, however, because in databases (as well as in literature), there are many false attributions of disease-causing variants. Rare nontruncating variants (synonymous, nonsynonymous, and noncoding variants) that have been described as "pathogenic" and associated with a phenotype should be carefully interpreted for their clinical significance.[19] The major challenge for interpreting and reporting variants is the need for critical and rigorous interpretation of variants associated with clinical indications. Several databases (eg, the Human Gene Mutation Database, ClinVar, the LOVD) document disease-causing variants and attempt to improve variant curation.

Clinicians reviewing NGS clinical reports should apply critical thinking and be aware of the possibility of a false attribution of pathogenicity to a variant. To achieve better diagnostic accuracy, clinicians should extensively review the medical literature and consult with experts in genomic analysis. Clinical geneticists, molecular geneticists, genetic counselors, and pediatric subspecialists may all be helpful. These experts can help a clinician understand the acknowledged

limitations of NGS. For example, NGS is known to miss some particular genetic variations, such as trinucleotide repeat disorders, mitochondrial DNA mutations, large indels, translocation, and disorders of epigenetic regulation.

Given the uncertainties regarding the meaning of many NGS results, NGS cannot be used as a substitute for a careful clinical evaluation. The interpretation of sequence variants could be significantly improved by encouraging data sharing and transparent exchange of curated variants associated with the phenotype.

## CLINICAL USE OF NGS

NGS is likely to be used more in pediatrics than in other clinical settings, mainly because many genetic conditions have a poor prognosis and children who have those conditions do not survive until adulthood. Thus, pediatricians need to be aware of the promise and pitfalls of NGS and be prepared to decide when it will be useful for patients.

Recent studies suggest that less than one-half of patients who have genetic conditions are diagnosed by using standard genetic approaches.[1,5,9,10,12,23] It is possible that NGS will allow a precise diagnosis in a much higher percentage of infants.[1,5,9,10,12,13,16,17,24] At present, many of those infants will have conditions for which no treatment exists. The major benefit of an accurate diagnosis will be to allow precise prognosis and better-informed discussions about the desirability of life-prolonging treatment. To be better prepared for these discussions, pediatricians should be familiar with the technology of testing, the ambiguities in diagnosis, and the possibility for false-positive and false-negative findings that are associated with different strategies for interpreting genomic variants. With such caveats, NGS may prove useful in the care of infants and children with rare conditions that have not been diagnosed with the use of more traditional tests.


## ACKNOWLEDGMENTS

We are grateful to the staff of the Center for Pediatric Genomic Medicine, Children's Mercy Kansas City, and the NSIGHT teams for discussions.


Web sites for Bioinformatics Resources:

1000 Genomes Project, http://www.1000genomes.org/

NHBLI Exome Sequencing Project (ESP) Exome Variant Server, http://evs.gs.washington.edu/EVS/

Online Mendelian Inheritance in Man (OMIM), http://www.omim.org

SNPdb, http://www.ncbi.nlm.nih.gov/projects/SNP

UCSC Genome Browser, http://genome.ucsc.edu/

PolyPhen-2, http://www.genetics.bwh.harvard.edu/pph2/

ClinSeq, http://genome.gov/20519355

RefSeq, http://www-ncbi-nlm-nih-gov.proxy2.library.mcgill.ca/RefSeq

SIFT, sift.jcvi.org/

ExAC, http://exac.broadinstitute.org/

HGMD, http://www.hgmd.cf.ac.uk/ac/index.php

## ABBREVIATIONS

HPO: Human Phenotype Ontology
NGS: next-generation sequencing

## REFERENCES


1. Biesecker LG, Green RC. Diagnostic clinical genome and exome sequencing. *N Engl J Med*. 2014;370(25):2418–2425
2. Shashi V, McConkie-Rosell A, Rosell B, 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(2):176–182
3. Yang Y, Muzny DM, Xia F, et al. Molecular findings among patients referred for clinical whole-exome sequencing. *JAMA*. 2014;312(18):1870–1879
4. Zemojtel T, Köhler S, Mackenroth L, et al. Effective diagnosis of genetic disease by computational phenotype analysis of the disease-associated genome. *Sci Transl Med*. 2014;6(252):252ra123
5. Vrijenhoek T, Kraaijeveld K, Elferink M, et al. Next-generation sequencing-based genome diagnostics across clinical genetics centers: implementation choices and their effects [published correction appears in Eur J Hum Genet. 2015;23(9):1270]. *Eur J Hum Genet*. 2015;23(9):1142–1150
6. Berg JS. Genome-scale sequencing in clinical care: establishing molecular diagnoses and measuring value. *JAMA*. 2014;312(18):1865–1867
7. Deciphering Developmental Disorders Study. Large-scale discovery of novel genetic causes of developmental disorders. *Nature*. 2015;519(7542):223–228
8. Williams HJ, Hurst JR, Ocaka L, et al. The use of whole-exome sequencing to disentangle complex phenotypes [published online ahead of print June 10, 2015]. Eur J Hum Genet. doi: 10.1038/ejhg.2015.121
9. Srivastava S, Cohen JS, Vernon H, et al. Clinical whole exome sequencing in child neurology practice. *Ann Neurol*. 2014;76(4):473–483
10. Need AC, Shashi V, Hitomi Y, et al. Clinical application of exome sequencing in undiagnosed genetic conditions. *J Med Genet*. 2012;49(6):353–361
11. Lupski JR, Gonzaga-Jauregui C, Yang Y, et al. Exome sequencing resolves apparent incidental findings and reveals further complexity of SH3TC2

variant alleles causing Charcot-Marie-Tooth neuropathy. *Genome Med*. 2013;5(6):57

12. Lee H, Deignan JL, Dorrani N, et al. Clinical exome sequencing for genetic identification of rare Mendelian disorders. *JAMA*. 2014;312(18):1880–1887
13. Jacob HJ. Genetic diagnosis through whole-exome sequencing. *N Engl J Med*. 2014;370(11):1069
14. Farwell KD, Shahmirzadi L, El-Khechen D, et al Enhanced utility of family-centered diagnostic exome sequencing with inheritance model-based analysis: results from 500 unselected families with undiagnosed genetic conditions. Genet Med. 2014;17(7):578–586
15. Worthey EA, Mayer AN, Syverson GD, et al Making a definitive diagnosis: successful clinical application of whole exome sequencing in a child with intractable inflammatory bowel disease. Genet Med. 2011;13(3):255–262
16. Soden SE, Saunders CJ, Willig LK, et al. Effectiveness of exome and genome sequencing guided by acuity of illness for diagnosis of neurodevelopmental disorders. *Sci Transl Med*. 2014;6(265):265ra168
17. Saunders CJ, Miller NA, Soden SE, et al. Rapid whole-genome sequencing for genetic disease diagnosis in neonatal intensive care units. *Sci Transl Med*. 2012;4(154):154ra135
18. Dixon-Salazar TJ, Silhavy JL, Udpa N, et al. Exome sequencing can improve diagnosis and alter patient management. *Sci Transl Med*. 2012;4(138):138ra78
19. 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(5):405–424
20. Ullah MZ, Aono M, Seddiqui MH. Estimating a ranked list of human hereditary diseases for clinical phenotypes by using weighted bipartite network. In: Proceedings from the Annual International Conference of the IEEE Engineering in Medicine and Biology Society; 2013;3475–3478; July 3-7, 2013; Osaka, Japan
21. Köhler S, Schulz MH, Krawitz P, et al. Clinical diagnostics in human genetics with semantic similarity searches in ontologies. *Am J Hum Genet*. 2009;85(4):457–464
22. Willig LK, Petrikin JE, Smith LD, et al. Whole-genome sequencing for identification of Mendelian disorders in critically ill infants: a retrospective analysis of diagnostic and clinical findings. *Lancet Respir Med*. 2015;3(5):377–387
23. Kingsmore SF, Saunders CJ. Deep sequencing of patient genomes for disease diagnosis: when will it become routine? *Sci Transl Med*. 2011;3(87):87ps23
24. Yang Y, Muzny DM, Reid JG, et al. Clinical whole-exome sequencing for the diagnosis of mendelian disorders. *N Engl J Med*. 2013;369(16):1502–1511

# The Challenge of Analyzing the Results of Next-Generation Sequencing in Children

Isabelle Thiffault and John Lantos

*Pediatrics* 2016;137;S3
DOI: 10.1542/peds.2015-3731C

PEDIATRICS®
OFFICIAL JOURNAL OF THE AMERICAN ACADEMY OF PEDIATRICS

# The Challenge of Analyzing the Results of Next-Generation Sequencing in Children

Isabelle Thiffault and John Lantos

*Pediatrics* 2016;137;S3
DOI: 10.1542/peds.2015-3731C