



For rare pediatric diseases, genome sequencing can increase diagnostic yield

By Nabihah Sachedina, MBBS, MPP, MBA

It is one of the most traumatic things new parents can experience: the discovery that something is wrong with their young baby. Perhaps there is some obvious illness, or the infant is not developing as expected. Whatever the cause, the event sets off a cascade of medical visits, tests, and evaluations that may span years in the aptly named diagnostic odyssey.

When it comes to rare pediatric diseases, shortening the time to diagnosis must be a key goal for clinical teams, especially those in laboratory medicine. For many diseases and disorders, early intervention with therapies or modifications

to diet or lifestyle can make a significant difference in future quality of life. Those interventions are enabled by a clear clinical diagnosis delivered as quickly as possible.

While incremental progress has been possible in recent decades through technical advances such as microarrays and syndromic panel testing, the real shift in improving diagnostic yield has come from the application of modern DNA sequencing platforms. Beginning with whole exome sequencing and moving to the more comprehensive whole genome sequencing approach, the ability to scan broadly for potentially pathogenic variants has led to faster, more



Photo 56463226 © Avrahuismannphotography | Dreamstime.com

definitive results for pediatric diseases. This is particularly true for rare diseases, which are often tied to a single highly pathogenic variant — in some cases a variant that has not been previously reported. These variants would be virtually impossible to detect with conventional technologies, but with exome or genome coverage they become much easier to spot.

At this point, there have been enough reported studies deploying a variety of sequencing tools for rare disease diagnosis that it is now possible to review the advantages and disadvantages of different approaches. For clinical laboratory teams, it can be instructive to consider the diagnostic outcomes of exome versus genome sequencing, proband versus trio sequencing, short-read versus long-read sequencing, the value of epigenetics, and more.

Pediatric rare diseases

What qualifies as a rare disease differs across countries, leading to significant variation in overall numbers reported. A landmark paper written in 2019 by scientists from Orphanet in Europe puts the number of clinically defined rare diseases at more than 6,000, while the American nonprofit National Organization for Rare Disorders cites the tally as greater than 10,000.^{1,2} Some better-known examples include sickle cell

Earning CEUs

See test online at <https://ce.mlo-online.com/courses/for-rare-pediatric-diseases-genome-sequencing-can-increase-diagnostic-yield/>. Passing scores of 70 percent or higher are eligible for 1 contact hour of P.A.C.E. credit.



Scan code to go directly to the CE test.

LEARNING OBJECTIVES

Upon completion of this article, the reader will be able to:

1. List past testing paradigms in the diagnosis of rare pediatric diseases.
2. Describe the factors and benefits for the advancement of molecular diagnostics in rare pediatric diseases.
3. Describe the type of mutations that different molecular sequencing tests can identify.
4. Discuss the prevalence and disease rates and cause of rare pediatric diseases.

disease, cystic fibrosis, Prader-Willi syndrome, hemophilia, and Duchenne muscular dystrophy. While rare diseases have low individual prevalence, taken collectively, they affect 300 million people globally.³

The total disease count may vary by region, but key characteristics of rare disease do not. Around 80% percent of rare diseases have a genetic cause and 70% begin in childhood.³ Newborn screening can detect some of these conditions, but it is used inconsistently across geographic regions and typically for a small number of diseases and disorders. For more comprehensive investigations, extensive genetic analysis is indispensable.

But genome sequencing has only been available at a cost-effective price point in the last several years. Typically, serial conventional tests are used to evaluate one hypothesis after another, subjecting babies and children to a number of potentially invasive and costly procedures — and placing a significant burden on their families.

The diagnostic odyssey that ensues is unique to each family, but almost universally complex. One study found that 38 percent of families had to consult at least six physicians to receive a diagnosis.⁴ Another, focused on adults with rare diseases, found that nearly a third of respondents had waited more than five years for an accurate diagnosis, while half had been diagnosed incorrectly along the way.⁵ The specific disease can be a significant contributor to the odyssey: for example, patients with cystic fibrosis are often diagnosed more quickly (25 percent of patients achieved a correct diagnosis in about 15 months) while patients with Ehlers-Danlos syndrome can wait for decades (25 percent of patients waited 28 years for an accurate diagnosis).⁶

Devastatingly, 30% of children with a rare disease die before their fifth birthday, in part due to delayed diagnosis and limited treatment options.³

Sequencing approaches

Clearly, a different model is needed, and comprehensive genomic sequencing may answer that call. As the costs of next-generation DNA sequencing technologies fell and instruments became more widely available, researchers began to

assess whole exome sequencing for identifying the genetic cause of rare diseases in children. In the past decade, whole exome sequencing has been used successfully to increase the diagnostic yield for pediatric rare diseases compared to conventional testing. A meta-analysis of 37 studies found that the molecular diagnostic rate for whole exome sequencing ranged from 24 percent to 68 percent, while diagnostic utility of the commonly used chromosomal microarray ranged from 0 percent to 17 percent.⁷

As sequencing costs continued to fall, many researchers started to explore the impact of sequencing the genome rather than the exome. While results vary by study, the meta-analysis of studies performed between 2013 and 2017 deemed exome and genome analysis to be roughly equivalent in achieving a molecular diagnosis.⁷ In a further effort to increase diagnostic yield, researchers began identifying instances where trio sequencing — that is, sequencing whole genomes of the child and both parents — could resolve additional cases and improve the detection of *de novo* variants.

Virtually all of the studies included in this meta-analysis involved short-read sequencing, the most common type of next-generation sequencing technology. These platforms produce sequencing reads of a few hundred bases, requiring sophisticated assembly and alignment procedures to stitch reads together into a representation of the full genome.

More recently, researchers have been exploring the use of long-read sequencing to uncover answers in pediatric rare disease. Sequencers that are able to produce long reads — often hundreds of kilobases in length — require far less assembly and are therefore less prone to errors in alignment and orientation. Because of this, long-read whole genome sequencing is able to visualize previously ‘dark’ regions of the genome and in some cases can visualize disease-causing variants even without the additional step of trio sequencing. With extremely long reads, genetic regions can be phased into maternal and paternal haplotypes based solely on the child’s sequencing data. This has the potential to be particularly advantageous in cases when it is not possible to analyze both parents’ DNA.

Variants to target

Because rare diseases may be caused by many different types of variants — and *de novo* mutations in particular can arise in various forms — the ideal technical approach would allow for the detection of all types of genetic variation.

Short-read sequencers, unfortunately, are not up to this ambitious task. While excellent at detecting single-nucleotide variants and small insertions or deletions, the short reads generated by these platforms are not long enough to span larger variants. Structural variants such as copy number alterations, repeat expansions, inversions, and translocations, among many others, cannot be consistently and reliably identified with short-read data because these reads can collapse during assembly, misrepresenting the original DNA sequence.

To view a broader range of variants, many researchers have turned to long-read sequencing platforms. These technologies may produce reads kilobases to even megabases in length, long enough to span even large genomic variants in a single read to eliminate the ambiguities associated with assembly. For example, scientists have recently used a long-read sequencing technology to target the *SMN1* and *SMN2* genes associated with spinal muscular atrophy, creating a single-platform workflow that generates amplicons as long as 11 kilobases and calls single-nucleotide variants, insertions, deletions, and copy number variants with at least 98 percent accuracy.⁸

For a truly comprehensive approach to uncovering answers about rare disease, clinical laboratory teams may want to consider sequencing technologies that go beyond DNA variants. Transcriptome (RNA-based) studies can provide valuable insights into gene activity, or expression. Long-read sequencers again demonstrate more complete results because longer reads can span full isoforms. Short reads, on the other hand, must be assembled together to infer isoforms, and this process is particularly error-prone for representing splice variants.

Methylation is another area that has been informative for rare diseases, especially for imprinting disorders such as Angelman syndrome. To track methylation patterns, most sequencing technologies require a bisulfite conversion step. This adds a layer of complexity, time, and cost on top of the exome or genome sequencing process. Some sequencing platforms can now directly detect methylation profiles as they sequence DNA, enabling valuable insights without additional cost or processing.

Rare disease studies

The use of long-read whole genome sequencing to resolve pediatric rare disease cases is expanding. Studies have already started to demonstrate the potential for this approach.

For example, researchers in France evaluated the performance of short-read and long-read whole genome sequencing for rare disease cases in which whole exome sequencing had not produced definitive results. In a small pilot study including five probands, they were able to make molecular diagnoses for three cases based on data not found in the coding regions of the genome.⁹ In one case, the use of long-read sequencing allowed the team to identify a balanced inversion known to be a rare cause of Sotos syndrome.

Tracking cryptic variants was also the goal for a study from researchers in Spain and France looking to diagnose the genetic disorder congenital aniridia. From a prior cohort of 110 patients, two cases remained unsolved after short-read sequencing efforts. The team deployed long-read nanopore sequencing to characterize the *PAX6* locus related to the disorder and found balanced chromosomal rearrangements, including a *de novo* inversion and a translocation.¹⁰ The specific breakpoint was mapped to a highly repetitive region in the

Long-read sequencers again demonstrate more complete results because longer reads can span full isoforms.

centromere of chromosome 6. According to the team, “Our study underscores the limitations of traditional short-read sequencing in uncovering pathogenic [structural variants] affecting low-complexity regions of the genome and the value of [long-read sequencing] in providing insight into hidden sources of variation in rare genetic diseases.” In a recent preprint study, researchers detected a broad range of pathogenic changes and identified diagnoses with long-read sequencing for 13% of cases that short-read sequencing had failed to solve.¹¹

Studies have also tackled the all-important component of time. A team in California deployed nanopore sequencing for an ultra rapid genome sequencing workflow that reported results as quickly as 7 hours and 18 minutes after the lab received a blood sample.¹² The study included 12 critically ill patients; five received an initial genetic diagnosis through this process. One patient was a three-month-old baby experiencing seizures and other symptoms. Previous tests, including interictal electroencephalography and magnetic resonance imaging, did not provide information leading to a clear diagnosis. Genome sequencing and variant analysis performed in about eight and a half hours identified a heterozygous variant in *CSNK2B* considered likely pathogenic. This enabled a definitive diagnosis of Poirier–Bienvenu neurodevelopmental syndrome, a *CSNK2B*-related disorder. “This result halted further planned diagnostic testing, facilitated disease-specific counseling and prognostication, and aided in management of epilepsy by providing insight about reported seizure types and treatment response to common antiseizure medications,” the team noted. Interestingly, an epilepsy gene panel ordered before the patient was enrolled in the study returned results two weeks later but offered no clear diagnostic information.

Researchers are starting to explore methylation patterns associated with rare diseases. In Japan, researchers used a sequencing technology that directly detects methylation while it sequences DNA as the basis for an assay to identify individuals with Prader-Willi syndrome or Angelman syndrome.¹³ Their focus on the key region of chromosome 15 associated with these imprinting disorders allowed them to spot aberrant methylation and to diagnose four patients with Prader-Willi syndrome and three others with Angelman syndrome. Because they also had DNA sequence data for the same region, the researchers were able to analyze copy number, homozygosity, and structural variation to characterize pathogenic mechanisms responsible for the two conditions.

Another team used a similar approach to explore methylation associated with developmental and epileptic encephalopathies. They targeted rare, differentially methylated regions of the genome in 10 individuals and used long-read sequencing to identify associated genetic variants, such as repeat expansions with high GC content, copy number variants, and a balanced translocation.¹⁴ Some pathogenic sequence variants linked to the methylated regions had been missed in previous exome analyses, with overall increases in the diagnostic yield.

Looking ahead

The true benefits of full genome characterization, transcriptomics, and epigenetics are likely to be significant. As more data is generated and matched with clinical phenotypes, it is expected that the diagnostic yield of sequencing approaches will substantially increase.

To date, many of the published studies are small-scale or clinical research efforts. However, the ability to truly identify the genomic basis of disease is likely to result in growing implementation of long-read whole genome sequencing to support rare disease diagnosis in babies and children. Clinical laboratories can prepare for this by evaluating the needs of their patient populations and determining how those needs can be answered by the various sequencing technologies available today. Ultimately, the most comprehensive approach possible is likely to provide the most answers for families facing the effects of an undiagnosed rare disease. 📌



Scan code to go directly to the CE test.

REFERENCES

1. 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.* 2020;28(2):165-173. doi:10.1038/s41431-019-0508-0.
2. List of rare diseases. National Organization for Rare Disorders. Accessed May 7, 2024. <https://rarediseases.org/rare-diseases/>.
3. The Lancet Global Health. The landscape for rare diseases in 2024. *Lancet Glob Health.* 2024;12(3):e341. doi: 10.1016/S2214-109X(24)00056-1.
4. Zurynski Y, Deverell M, Dalkeith T, et al. Australian children living with rare diseases: experiences of diagnosis and perceived consequences of diagnostic delays. *Orphanet J Rare Dis.* 2017;11;12(1):68. doi:10.1186/s13023-017-0622-4.
5. Molster C, Urwin D, Di Pietro L, et al. Survey of healthcare experiences of Australian adults living with rare diseases. *Orphanet J Rare Dis.* 2016;24;11:30. doi:10.1186/s13023-016-0409-z.
6. Kole A, Faurisson F. Rare diseases social epidemiology: analysis of inequalities. *Adv Exp Med Biol.* 2010;686:223-50. doi:10.1007/978-90-481-9485-8_14.
7. Clark MM, Stark Z, Farnaes L, et al. Meta-analysis of the diagnostic and clinical utility of genome and exome sequencing and chromosomal microarray in children with suspected genetic diseases. *NPJ Genom Med.* 2018;9;3:16. doi:10.1038/s41525-018-0053-8.
8. Hall B, Alyafei S, Ramaswamy S, Sinha S, et al. A dual-mode targeted Nanopore sequencing assay for comprehensive *SMN1* and *SMN2* variant analysis. PREPRINT. medRxiv 2024.02.22.24303180; doi: <https://doi.org/10.1101/2024.02.22.24303180>.
9. Lecoquierre F, Quenez O, Fournieux S, et al. High diagnostic potential of short and long read genome sequencing with transcriptome analysis in exome-negative developmental disorders. *Hum Genet.* 2023;142(6):773-783. doi:10.1007/s00439-023-02553-1.
10. Damián A, Núñez-Moreno G, Jubin C, Tamayo A, et al. Long-read genome sequencing identifies cryptic structural variants in congenital aniridia cases. *Hum Genomics.* 2023;2;17(1):45. doi:10.1186/s40246-023-00490-8.
11. Tayoun AA, Sinha S, Rabea F, et al. Long read sequencing enhances pathogenic and novel variation discovery in patients with rare diseases. *Research Square.* Published online 2024. doi:10.21203/rs.3.rs-4235049/v1.
12. Gorzynski JE, Goenka SD, Shafin K, et al. Ultrarapid Nanopore Genome Sequencing in a Critical Care Setting. *N Engl J Med.* 2022;17;386(7):700-702. doi:10.1056/NEJMc2112090.
13. Yamada M, Okuno H, Okamoto N, et al. Diagnosis of Prader-Willi syndrome and Angelman syndrome by targeted nanopore long-read sequencing. *Eur J Med Genet.* 2023;66(2):104690. doi:10.1016/j.ejmg.2022.104690.
14. LaFlamme CW, Rastin C, Sengupta S, et al. Diagnostic utility of genome-wide DNA methylation analysis in genetically unsolved developmental and epileptic encephalopathies and refinement of a CHD2 epismature. *bioRxiv.* Published online 2023. doi:10.1101/2023.10.11.23296741.



Nabihah Sachedina, MBBS, MPP, MBA currently serves as Vice President, Health Programs, at Oxford Nanopore Technologies. She is a UK-trained pediatrician who has spent more than a decade developing and implementing city-wide, regional, and national public health and healthcare policy.