



Long-read next-generation sequencing for molecular diagnosis of pediatric endocrine disorders

Yoko Kuroki^{1,2},
Atsushi Hattori^{1,3},
Keiko Matsubara^{1,3},
Maki Fukami^{1,3}

¹Division of Diversity Research, National Research Institute for Child Health and Development, Tokyo, Japan
²Department of Genome Medicine, National Research Institute for Child Health and Development, Tokyo, Japan
³Department of Molecular Endocrinology, National Research Institute for Child Health and Development, Tokyo, Japan

Recent advances in long-read next-generation sequencing (NGS) have enabled researchers to identify several pathogenic variants overlooked by short-read NGS, array-based comparative genomic hybridization, and other conventional methods. Long-read NGS is particularly useful in the detection of structural variants and repeat expansions. Furthermore, it can be used for mutation screening in difficult-to-sequence regions, as well as for DNA-methylation analyses and haplotype phasing. This mini-review introduces the usefulness of long-read NGS in the molecular diagnosis of pediatric endocrine disorders.

Keywords: Gene, Molecular analysis, Mutation, Nanopore, PacBio, Sequencer

Highlights

- Long-read next-generation sequencing (NGS) can provide sequence reads of several kilobases or megabases and therefore identify several pathogenic variants that have been overlooked by other methods.
- Despite its relatively high error rate and high cost, long-read NGS is becoming an important tool for molecular diagnosis of congenital diseases, including endocrine disorders.

Introduction

Nucleotide substitutions and structural variants in the genome play a major role in the development of congenital diseases, including endocrine disorders.¹⁻³ The identification of pathogenic variants in patients with congenital disorders helps to optimize management procedures and increase the accuracy of genetic counseling.⁴ Currently, short-read next-generation sequencing (NGS) and array-based comparative genomic hybridization (CGH), together with conventional methods such as karyotyping, Sanger sequencing, and fluorescence *in situ* hybridization, are used to identify pathogenic mutations and copy number variants (CNVs).^{1,2,5} Still, these methods have several technical limitations. For example, short-read NGS typically yields sequence reads measuring 100 or 150 bp in length and is optimized to detect simple nucleotide substitutions and indels.¹ Likewise, array-based CGH primarily focuses on CNVs in single-copy genomic regions (<https://www.agilent.com/en/product/cgh-cgh-snp-microarray-platform>). As a result, these methods often overlook chromosomal rearrangements, particularly retrotransposon insertions, repeat expansions, and copy number neutral inversions. These missing genomic variants may account for a certain percentage of cases with congenital disorders. Moreover, short-read NGS and array-based CGH cannot be used to evaluate epigenetic abnormalities.

Recent advances in long-read NGS have enabled researchers to identify several previously unrecognized genetic variants.⁶⁻¹⁰ Furthermore, long-read NGS can be used for mutation screening in difficult-to-sequence regions, as well as for DNA-methylation analyses and haplotype phasing.^{7,9} Currently, the relatively high error rate and the high cost of long-read

See the commentary on "Long-read next-generation sequencing for molecular diagnosis of pediatric endocrine disorders" via <https://doi.org/10.6065/apem.24224014edi03>.

Received: 9 February, 2024
Accepted: 23 April, 2024

Address for correspondence:

Maki Fukami
Department of Molecular Endocrinology, National Research Institute for Child Health and Development, 2-10-1 Okura, Setagaya, Tokyo 157-8535, Japan
Email: fukami-m@ncchd.go.jp
<https://orcid.org/0000-0001-9971-4035>

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

ISSN: 2287-1012(Print)
ISSN: 2287-1292(Online)

NGS are being improved.¹⁰ In this mini-review, we introduce the usefulness of long-read NGS in the molecular diagnosis of congenital disorders, particularly of pediatric endocrine disorders.

Long-read NGS

Nature Methods described long-read NGS as the "Method of the Year 2022."⁷ Long-read NGS is becoming a popular tool for molecular analysis of clinical samples. Two technologies, i.e., single-molecule real-time sequencing by Pacific Biosciences (Menlo Park, CA, USA) and Oxford Nanopore Technologies (ONT) (Oxford, UK), are predominantly used in research and clinical settings.⁷ These 2 technologies can provide sequence reads of several kilobases⁹ and therefore are capable of characterizing complex structural variants, chromosomal inversions, retrotransposon insertions, and repeat expansions. Of course, long-read NGS can be exploited for the detection of simple nucleotide substitutions and indels, although it leads to a higher error rate in sequence data than short-read NGS. Notably, long-read NGS can be used not only for whole-genome sequencing but also for target sampling for specific regions of interest.^{8,11} Target sequencing significantly reduces the cost of sequence analyses. In particular, ONT has provided a system for software-based target assignment designated as adaptive sampling.¹¹ Furthermore, long-read NGS has advantages in DNA-methylation analysis and rapid workflow.⁷

Long-read NGS is particularly beneficial for the detection of specific types of genomic abnormalities (Fig. 1). In the following sections, we provide some examples of molecular diagnoses of endocrine disorders achieved by long-read NGS.

Identification of pathogenic structural variants and repeat expansions

Recently, long-read NGS succeeded in identifying several missing variants associated with congenital diseases.⁶⁻¹⁰ Identified variants included CNVs involving transposable elements.^{6,7,9} It is known that such elements account for a substantial percentage of the human genome and that insertions of these elements can cause disease phenotypes by disrupting exons or the cis-regulatory machinery of gene expression.^{11,12}

Nevertheless, CNVs of transposable elements are barely discernible by short-read NGS and array-based CGH. In 2022, Miller et al.¹³ performed targeted long-read NGS for a family with autosomal dominant pseudohypoparathyroidism and successfully identified a ~2.8-kb insertion in the *GNAS* region on chromosome 20. This insertion was a retrotransposon element composing an SVA-VNTR-*Alu* sequence and is assumed to have caused the disease phenotype through epigenetic dysregulation of nearby genes. Similarly, an *Alu* insertion was identified in an *ALMS1* intron of patients with Alstrom syndrome,¹⁴ and an insertion of the SINE-VNTR-*Alu* retroelement was detected in an *NR5A1* intron of patients with disorders of sex development.¹⁵ Many other examples support the usefulness of long-read NGS in the detection of missing structural variants.^{6,7,9}

Furthermore, long-read NGS is capable of analyzing copy number alterations of repeats and repetitive sequences that are often overlooked by other methods.^{6,7,9} Expansions of short tandem repeats, particularly triplet repeat expansions, are known to be one of the major causes of neurological disorders and have also been linked to other types of disorders.^{16,17} For example, a trinucleotide repeat expansion in *DMD* intron was identified in patients with myopathy.¹⁸ Furthermore, Miyatake et al.¹⁹ succeeded in identifying pathogenic repeat expansions in several patients with neurological and neuromuscular diseases. The authors concluded that ONT-based adaptive sampling is superior to conventional diagnostic methods in terms of speed, accuracy, and comprehensiveness.

In addition, long-read NGS is beneficial in determining the structure of complex rearrangements.^{6,7,9} In this context, recent studies have discovered unique cellular events designated as chromothripsis or chromoanagenesis, which result in highly complex rearrangements involving one or a few chromosomes (Fig. 2).²⁰⁻²² To date, chromothripsis-/chromoanagenesis-compatible rearrangements have been identified in multiple patients with growth failure, congenital malformations, and endocrine abnormalities.²⁰⁻²² However, whole-genome sequencing using short-read NGS and array-based CGH show difficulties in determining the alignments of chromosomal fragments of these cases. Lei et al.²³ employed long-read NGS and determined the structure of a massively rearranged chromosome in a patient with atypical Langer-

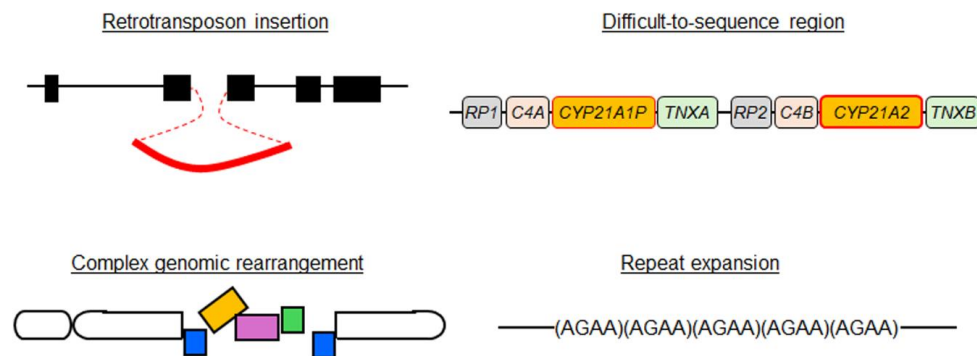


Fig. 1. Genomic features for which long-read next-generation sequencing is advantageous over conventional methods.

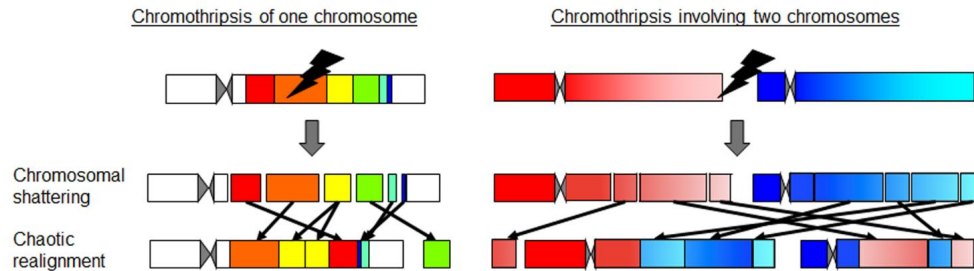


Fig. 2. Schemas of chromothripsis/chromoanagenesis. This phenomenon is characterized by the shattering of one or a few chromosomes and their subsequent random realignment. This realignment can include deletion and duplication of chromosomal fragments.

Giedion syndrome and Cornelia de Lange syndrome type IV. Furthermore, long-read NGS can detect copy number neutral inversions and translocations that are usually undetectable by array-based CGH.

Sequencing for difficult-to-sequence regions

The human genome contains several difficult-to-sequence regions.²³ These regions often harbor homologous sequences due to segmental amplification.²³ An important example of difficult-to-sequence regions from the viewpoint of pediatric endocrinology is 6p21.33, which contains *CYP21A2*.²⁴ *CYP21A2* is the causative gene for congenital adrenal hyperplasia due to 21-hydroxylase deficiency.²⁵ This gene is located within a segmentally duplicated region of approximately 40 kb (Fig. 1).²⁵ The presence of a pseudogene (*CYP21A1P*) renders *CYP21A2* difficult for short-read NGS and Sanger sequencing to analyze.²⁵ Aberrant recombination between *CYP21A2* and *CYP21A1P* is often observed in patients with 21-hydroxylase deficiency.²⁵ Thus, efforts were made to sequence this region using long-read NGS.^{26,27} Adachi et al. enabled mutation screening of *CYP21A2* through long-read NGS for PCR-amplified DNA fragments.²⁷ The authors proposed that the cost of this analysis can be reduced by using a barcode system.

DNA-methylation analysis

Long-read NGS can be used to analyze DNA modification.^{6,7,9} In particular, ONT-based NGS can analyze DNA methylation without bisulfite treatment of DNA samples. Information on DNA methylation is critical for molecular diagnosis of imprinting disorders, which explain a certain percentage of neonatal diabetes, pseudohypoparathyroidism, pubertal disorders, and short stature in individuals considered small for their gestational age.²⁸ Yamada et al.²⁹ performed DNA-methylation analyses using targeted ONT-based NGS and successfully diagnosed patients with Prader-Willi syndrome or Angelman syndrome. Furthermore, ONT-based NGS is useful in the genome-wide screening for DNA-methylation changes associated with disorders (epi-variants).

Haplotype phasing

Another advantage of long-read NGS over short-read NGS and array-based CGH is its capability for haplotype phasing.⁹ Haplotype phasing is a method to differentiate 2 alleles inherited from the mother and the father.⁹ Haplotype phasing provides critical information on compound heterozygosity of mutations associated with autosomal recessive diseases; this method can determine whether 2 variants in a gene are located on the same or different alleles. Furthermore, it is useful to determine whether a *de novo* mutation is located on the maternally or paternally derived allele. This information helps to predict the consequences of variants of imprinted genes. The combination of haplotype phasing with DNA-methylation analysis enables researchers to determine epigenetic abnormalities on each allele. In addition, haplotype phasing can be used to clarify the parental origin of each DNA fragment of complex rearrangements and to examine the clonality of multiple mosaic variants in tumor tissues.⁹

Limitations and further applications of long-read NGS

Long-read NGS has some limitations, which include a relatively high error rate and cost.^{6,7,9} However, the accuracy of sequence results is continuously improving, and technical advances are helping to reduce the cost per base.^{9,10} It is expected that long-read NGS will become a major platform for molecular diagnosis of congenital disorders. In this context, ONT-based NGS is characterized by rapid workflow.⁷ A pilot study at Stanford University showed that ultra-rapid ONT-based NGS was able to diagnose rare genetic diseases of critically ill patients within an average of 8 hours.⁷ In addition, long-read NGS has several further applications in research and clinical settings, such as cancer genetics and infection surveillance,^{7,9} although this is beyond the scope of the present article.

Conclusions

Accumulating evidence suggests that long-read NGS can identify several pathogenic variants in the genome that have been overlooked by short-read NGS and array-based CGH.

Long-read NGS is becoming an increasingly important tool for molecular diagnosis of endocrine disorders.

Notes

Conflicts of interest: No potential conflict of interest relevant to this article was reported.

Funding: This study was supported by grants from the Takeda Science Foundation and the National Center for Child Health and Development.

Author contribution: Conceptualization: MF; Funding acquisition: MF; Visualization: YK, AH; Writing - original draft: MF; Writing - review & editing: YK, AH, KM

ORCID

Yoko Kuroki: 0000-0003-4697-8007

Atsushi Hattori: 0000-0003-1511-4820

Keiko Matsubara: 0000-0001-6254-9216

Maki Fukami: 0000-0001-9971-4035

References

- Eichler EE. Genetic variation, comparative genomics, and the diagnosis of disease. *N Engl J Med* 2019;381:64-74.
- Fukami M, Miyado M. Next generation sequencing and array-based comparative genomic hybridization for molecular diagnosis of pediatric endocrine disorders. *Ann Pediatr Endocrinol Metab* 2017;22:90-4.
- Lee C, Iafrate AJ, Brothman AR. Copy number variations and clinical cytogenetic diagnosis of constitutional disorders. *Nat Genet* 2007;39(7 Suppl):S48-54.
- Izumi Y, Suzuki E, Kanzaki S, Yatsuga S, Kinjo S, Igarashi M, et al. Genome-wide copy number analysis and systematic mutation screening in 58 patients with hypogonadotropic hypogonadism. *Fertil Steril* 2014;102:1130-6.
- Suga K, Imoto I, Ito H, Naruto T, Goji A, Osumi K, et al. Next-generation sequencing for the diagnosis of patients with congenital multiple anomalies and/or intellectual disabilities. *J Med Invest* 2020;67:246-9.
- Logsdon GA, Vollger MR, Eichler EE. Long-read human genome sequencing and its applications. *Nat Rev Genet* 2020;21:597-614.
- Oehler JB, Wright H, Stark Z, Mallett AJ, Schmitz U. The application of long-read sequencing in clinical settings. *Hum Genomics* 2023;17:73.
- Miller DE, Sulovari A, Wang T, Loucks H, Hoekzema K, Munson KM, et al. Targeted long-read sequencing identifies missing disease-causing variation. *Am J Hum Genet* 2021;108:1436-49.
- Conlin LK, Aref-Eshghi E, McEldrew DA, Luo M, Rajagopalan R. Long-read sequencing for molecular diagnostics in constitutional genetic disorders. *Hum Mutat* 2022;43:1531-44.
- Mastrososa FK, Miller DE, Eichler EE. Applications of long-read sequencing to Mendelian genetics. *Genome Med* 2023;15:42.
- Sharp AJ, Cheng Z, Eichler EE. Structural variation of the human genome. *Annu Rev Genomics Hum Genet* 2006;7:407-42.
- Stankiewicz P, Lupski JR. The genomic basis of disease, mechanisms and assays for genomic disorders. *Genome Dyn* 2006;1:1-16.
- Miller DE, Hanna P, Galey M, Reyes M, Linglart A, Eichler EE, et al. Targeted long-read sequencing identifies a retrotransposon insertion as a cause of altered GNAS exon A/B methylation in a family with autosomal dominant pseudohypoparathyroidism type 1b (PHP1B). *J Bone Miner Res* 2022;37:1711-9.
- Taşkesen M, Collin GB, Evsikov AV, Güzel A, Özgül RK, Marshall JD, et al. Novel Alu retrotransposon insertion leading to Alström syndrome. *Hum Genet* 2012;131:407-13.
- Del Gobbo GF, Wang X, Couse M, Mackay L, Goldsmith C, Marshall AE, et al. Long-read genome sequencing reveals a novel intronic retroelement insertion in NR5A1 associated with 46,XY differences of sexual development. *Am J Med Genet A* 2024;194:e63522.
- Pfaff AL, Singleton LM, Köks S. Mechanisms of disease-associated SINE-VNTR-Alus. *Exp Biol Med* (Maywood) 2022;247:756-64.
- Stevanovski I, Chintalaphani SR, Gamaarachchi H, Ferguson JM, Pineda SS, Scriba CK, et al. Comprehensive genetic diagnosis of tandem repeat expansion disorders with programmable targeted nanopore sequencing. *Sci Adv* 2022;8:eabm5386.
- Kekou K, Sofocleous C, Papadimas G, Petichakis D, Svingou M, Pons RM, et al. A dynamic trinucleotide repeat (TNR) expansion in the DMD gene. *Mol Cell Probes* 2016;30:254-60.
- Miyatake S, Koshimizu E, Fujita A, Doi H, Okubo M, Wada T, et al. Rapid and comprehensive diagnostic method for repeat expansion diseases using nanopore sequencing. *NPJ Genom Med* 2022;7:62.
- Liu P, Erez A, Nagamani SC, Dhar SU, Kołodziejaska KE, Dharmadhikari AV, et al. Chromosome catastrophes involve replication mechanisms generating complex genomic rearrangements. *Cell* 2011;146:889-903.
- Kloosterman WP, Guryev V, van Roosmalen M, Duran KJ, de Bruijn E, Bakker SC, et al. Chromothripsis as a mechanism driving complex de novo structural rearrangements in the germline. *Hum Mol Genet* 2011;20:1916-24.
- Hattori A, Fukami M. Established and novel mechanisms leading to de novo genomic rearrangements in the human germline. *Cytogenet Genome Res* 2020;160:167-76.
- Lei M, Liang D, Yang Y, Mitsuhashi S, Katoh K, Miyake N, et al. Long-read DNA sequencing fully characterized chromothripsis in a patient with Langer-Giedion syndrome and Cornelia de Lange syndrome-4. *J Hum Genet* 2020;65:667-74.
- Mantere T, Kersten S, Hoischen A. Long-read sequencing

- emerging in medical genetics. *Front Genet* 2019;10:426.
25. Stephens Z, Milosevic D, Kipp B, Grebe S, Iyer RK, Kocher JA. PB-Motif-A method for identifying gene/pseudogene rearrangements with long reads: an application to CYP21A2 genotyping. *Front Genet* 2021;12:716586.
 26. Zhang R, Cui D, Song C, Ma X, Cai N, Zhang Y, et al. Evaluating the efficacy of a long-read sequencing-based approach in the clinical diagnosis of neonatal congenital adrenocortical hyperplasia. *Clin Chim Acta* 2024;555:117820.
 27. Adachi E, Nakagawa R, Tsuji-Hosokawa A, Gau M, Kirino S, Yogi A, et al. A MinION-based long-read sequencing application with one-step PCR for the genetic diagnosis of 21-hydroxylase deficiency. *J Clin Endocrinol Metab* 2024;109:750-60.
 28. Soellner L, Begemann M, Mackay DJ, Grønskov K, Tümer Z, Maher ER, et al. Recent advances in imprinting disorders. *Clin Genet* 2017;91:3-13.
 29. Yamada M, Okuno H, Okamoto N, Suzuki H, Miya F, Takenouchi T, et al. Diagnosis of Prader-Willi syndrome and Angelman syndrome by targeted nanopore long-read sequencing. *Eur J Med Genet* 2023;66:104690.