Long-Read Sequencing and Optical Genome Mapping Identify Causative Gene Disruptions in Noncoding Sequence in Two Patients with Neurologic Disease and Known Chromosome Abnormalities

Ethan D. Sperry, MD, PhD[a,b], Kristen L. Sund, PhD[a], Jie Liu, PhD[a,b], Joyce Lee, PhD[c], John Garbe, PhD[d], Zakia Abdelhamed, MBChB, PhD[a], Chelsey Maag[a], Barbara Hallinan, MD, PhD[b,e], Steven Wu, MD[b,e], Archana Deshpande, MS[d], Rolf Stottmann, PhD[a,b], Teresa Smolarek, PhD[a,b], Lisa M. Dyer, PhD[a,b], and Matthew S. Hestand, PhD[a,b]

Abstract

Background: Despite advances in next generation sequencing (NGS), genetic diagnoses remain elusive for many patients with neurologic syndromes. Long-read sequencing (LRS) and optical genome mapping (OGM) technologies improve upon existing capabilities in the detection and interpretation of structural variation in repetitive DNA, on a single haplotype, while also providing enhanced breakpoint resolution.

Objective: To demonstrate utility of LRS and OGM for identification of clinically-actionable genetic diagnoses for patients in whom other genetic testing strategies (chromosome microarray, whole exome and genome sequencing) have been non-diagnostic.

Methods: We performed LRS and OGM on two patients with known chromosomal rearrangements and inconclusive Sanger or NGS.

Results: The first patient, who had epilepsy and developmental delay, had a complex translocation between two chromosomes that included insertion and inversion events. The second patient, who had a movement disorder, had an inversion on a single chromosome disrupted by multiple smaller inversions and insertions. Sequence level resolution of the rearrangements identified pathogenic breaks in noncoding sequence in or near known disease-causing genes with relevant neurologic phenotypes (*MBD5*, *NKX2-1*). These specific variants have not been reported previously, but expected molecular consequences are consistent with previously reported cases.

Conclusions: As the use of LRS and OGM technologies for clinical testing increases and data analyses become more standardized, these methods along with multiomic data to validate noncoding variation effects will improve diagnostic yield and increase the proportion of probands with detectable pathogenic variants for known genes implicated in neurogenetic disease.

Introduction

Many clinical genetic investigations, including chromosome microarray (CMA) and whole exome (WES) and genome (WGS) sequencing, are non-diagnostic for a plurality of individuals with syndromic presentations. Studies suggest that genetic diagnoses remain elusive for greater than 50% of individuals with chronic neurodevelopmental symptoms presenting in childhood[3,6]. Up to 5-10% of individuals with neurodevelopmental symptoms (but no better genetic diagnosis) may exhibit apparently balanced structural chromosome rearrangements not detectable using CMA, WES, or WGS[1,7].

Structural chromosome rearrangements are classically considered unbalanced (change in copy number of genetic material) or balanced (no change). Together, these rearrangements account for a significant minority of genetic disease[4,9]. While detection of grossly unbalanced rearrangements is achievable with CMA (to a resolution of 25-50 kilobases for most laboratories), detection of apparently balanced rearrangements can be much more challenging.

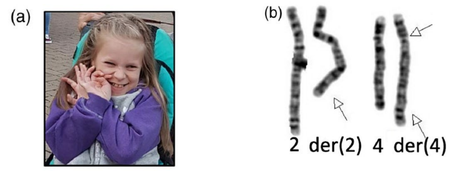
Apparently balanced structural chromosome rearrangements can be clinically silent, whereby they do not disrupt any coding or non-coding regulatory regions with no significant change in quantity of genetic material. However, rearrangements which disrupt canonical gene sequences, nearby regulatory regions, or distant sites important for chromatin packing can result in significant clinical disease, even in the absence of a detectable change to DNA quantity using CMA and standard next-generation sequencing technologies. More recently, long-read sequencing (LRS) and optical genome mapping (OGM) have emerged as clinically-useful technologies which allow for resolution of visible and submicroscopic structural rearrangement within the genome.

In this study, we describe two individuals with neurodevelopmental symptoms and non-diagnostic genetic testing who were found to have complex structural chromosome rearrangements via use of LRS and OGM.

Patient #1: *MBD5*-Associated Neurodevelopmental Disorder

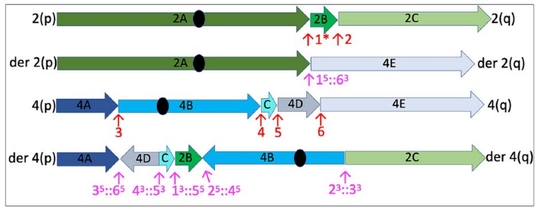
Clinical Course:
- Neonatal onset dystonic posturing of hands and seizures (managed on topiramate)
- Exam significant for prominent forehead and deep-set eyes
- Epilepsy gene panel (in infancy) with a paternally inherited pathogenic variant in *CNTNAP2*, associated with an autosomal recessive genetic epilepsy; this result is non-diagnostic
- Karyotype (age 7 years) significant for 46,XX,ins(4;2)(p15.2;q21.3q23)t(2;4)(q23;q25)

LRS and OGM performed to clarify structural rearrangement:

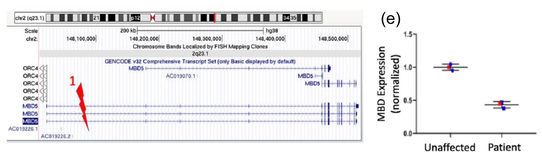


a) Dysmorphic features: prominent forehead, deep-set eyes, dystonic hand posturing

b) Karyotype with complex structural rearrangement involving chromosomes 2 and 4



c) Schematic of patient's complex structural chromosome rearrangement interrogated with LRS and OGM. Breakpoints on normal and derivative chromosomes labeled in red and magenta respectively. Derivative chromosomes with multiple rearrangements.



d) Breakpoint 1 disrupts intron 1 of *MBD5*
e) Quantitative real time PCR using patient's leukocytes demonstrates significant reduction in *MBD5* mRNA expression, consistent with breakpoint 1 disruption of canonical *MBD5* gene

Patient 1

Break-point	Gene	Repetitive elements	Break end	Chr	Long read	Optical mapping	Sanger
1	*MBD5* (ENSG00000204406.13, Intron 1/13, pLI = 1)	LINE	1^5	2	148,089,557	148,088,334	148,089,554
			1^3	2	148090061*	148,100,811	Unconfirmed
2	*LINC01931* (ENSG00000162947.5, Intron 2/4)	LINE	2^5	2	149,836,233	149,835,977	149,836,230
			2^3	2	149836193*	149,839,603	Unconfirmed
3	None	Seg Dup	3^5	4	28823922*	28,821,755	28,823,921,922
			3^3	4	28823989*	28,827,293	Unconfirmed
4	*ALPK1* (ENSG00000073331.18, Intron 1/16, pLI = 0)	LINE	4^5	4	112,302,477	112,297,656	112,302,478
			4^3	4	112,302,502	112,302,707	112,302,501
5	*ANK2* (ENSG00000145362.19, Intron 14/47, pLI = 1)	None	5^5	4	113265028*	113,264,433	Unconfirmed
			5^3	4	113,265,037	113,264,433	113,265,030
6	None	None	6^5	4	121784990*	121,782,613	121,784,989/990
			6^3	4	121,785,003	121,788,694	121,784,998

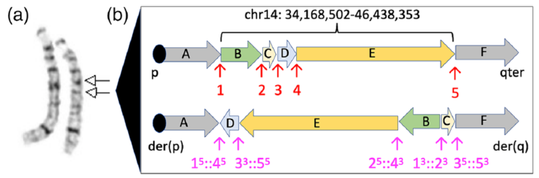
Table 1: Detailed breakpoint characterization for Patient 1

LRS and OGM allowed for resolution of breakpoints in a patient with a chromosome 2;4 translocation with insertion, resulting in a pathogenic disruption of *MBD5* intron 1.

Patient #2: Brain-Lung-Thyroid Syndrome (*NKX2-1*-Related Disorder)

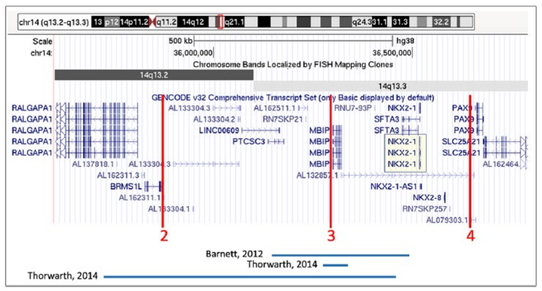
Clinical Course:
- Infant-onset abnormal involuntary choreiform movements and dystonic gait and myoclonus
- Other symptoms: developmental delay, intellectual difference, hypothyroidism, and severe anxiety
- Family history of father with similar clinical presentation, now deceased from lung cancer at age 41 (no relevant exposures)
- Karyotype at age 5 years with paracentric inversion of chromosome 14 (46,XX,inv(14)(q13q21)), inclusive of *NKX2-1*
- Whole genome sequencing at age 21 years was non-diagnostic

LRS and OGM performed to clarify structural rearrangement:



a) Karyotype with paracentric inversion of chromosome 14

b) LRS and OGM enabled detection of multiple breakpoints, labeled in red on normal chromosome and magenta on derivative chromosome. Derivative chromosome with evidence of multiple inversion events.



c) Breakpoint 3 disrupts a region just centromeric to *MBIP*, a gene crucial for proper *NKX2-1* expression[5]. Other published studies support copy number variations affecting this region (blue lines)[2,8].

Patient 2

Break-point	Gene	Repetitive elements	Break end	Chr	Long read	Optical mapping	Sanger
1	None	LINE	1^5	14	34,168,500	34159204*	34,168,504
			1^3	14	34168507*	34171770*	34,168,502
2	*BRMS1L* (ENSG00000100916.14, Intron 9/9, pLI = 0.06)	None	2^5	14	35,869,157	35,868,470	35,869,158
			2^3	14	35869245*	35875520*	35,869,245
3	None	SINE	3^5	14	36,293,562	36,292,963	36,293,457
			3^3	14	36,293,569	36,292,963	36,293,565
4	None	None	4^5	14	36,639,245	36638297*	36,639,246
		None	4^3	14	36,639,246	36,639,440	36,639,244
5	*LINC00871* (ENSG00000258700.6, Intron 3/5)	SINE	5^5	14	46,437,821	46,436,588	46,437,822
		LINE	5^3	14	46,438,353	46,441,930	46,438,349

Table 2: Detailed breakpoint characterization for Patient 2

LRS and OGM allowed for the resolution of breakpoints in a patient with a balanced chromosome 14 inversion resulting in likely pathogenic disruption to a centromeric regulatory region of *MIBP*, itself critical for expression of *NKX2-1*.

Conclusions and Future Directions

Long-read sequencing (LRS) and optical genome mapping (OGM) allow detection of complex structural non-coding genetic changes otherwise overlooked using CMA, WES, and WGS.

- OGM might be more sensitive to breakpoint detection in some situations (see Patient #2).
- Incorporation of LRS and OGM into clinical genetic diagnostic pipelines is expected to improve overall diagnostic yield and shorten the diagnostic odyssey for patients and their families.

Opportunities for future study:

- How can we ensure that LRS and OGM provide interpretable results for patients with complex findings? What orthogonal techniques and testing pipelines can be employed to ensure these results are clinically valid?
- Are there differences in structural genomic variation across clinical presentations and body systems? Can we be more deliberate about when to employ LRS and OGM?
- What is the sensitivity and specificity of OGM relative to LGS for specific clinical situations?
- Do LRS and OGM carry sufficient clinical efficacy to balance their expansive technologic and financial considerations?

References

1. Albuz et al. The high frequency of chromosomal copy number variations and candidate genes in epilepsy patients. Clin Neurol Neurosurg. 2021 Mar;202:106487. doi: 10.1016/j.clineuro.2021.106487. Epub 2021 Jan 11. PMID: 33484953.
2. Barnett et al. Choreoathetosis, congenital hypothyroidism and neonatal respiratory distress syndrome with intact NKX2-1. Am J Med Genet A. 2012 Dec;158A(12):3168-73. doi: 10.1002/ajmg.a.35456. Epub 2012 Nov 20. PMID: 23169673.
3. Clark 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 Jul 9;3:16. doi: 10.1038/s41525-018-0053-8. PMID: 30002876; PMCID: PMC6037748.
4. Halgren et al. Risks and Recommendations in Prenatally Detected De Novo Balanced Chromosomal Rearrangements from Assessment of Long-Term Outcomes. Am J Hum Genet. 2018 Jun 7;102(6):1090-1103. doi: 10.1016/j.ajhg.2018.04.005. Epub 2018 May 24. PMID: 29805044; PMCID: PMC5992120.
5. Invernizzi et al. Benign hereditary chorea and deletions outside NKX2-1: What's the role of MBIP? Eur J Med Genet. 2018 Oct;61(10):581-584. doi: 10.1016/j.ejmg.2018.03.011. Epub 2018 Apr 3. PMID: 29621620.
6. Srivastava et al; NDD Exome Scoping Review Work Group. Meta-analysis and multidisciplinary consensus statement: exome sequencing is a first-tier clinical diagnostic test for individuals with neurodevelopmental disorders. Genet Med. 2019 Nov;21(11):2413-2421. doi: 10.1038/s41436-019-0554-6. Epub 2019 Jun 11. Erratum in: Genet Med. 2020 Oct;22(10):1731-1732. doi: 10.1038/s41436-020-0913-3. PMID: 31182824; PMCID: PMC6831729.
7. Tabet et al. Complex nature of apparently balanced chromosomal rearrangements in patients with autism spectrum disorder. Mol Autism. 2015 Mar 25;6:19. doi: 10.1186/s13229-015-0015-2. PMID: 25844147; PMCID: PMC4384291.
8. Thorwarth et al. Comprehensive genotyping and clinical characterisation reveal 27 novel NKX2-1 mutations and expand the phenotypic spectrum. J Med Genet. 2014 Jun;51(6):375-87. doi: 10.1136/jmedgenet-2013-102248. Epub 2014 Apr 8. PMID: 24714694; PMCID: PMC5240655.
9. Yu et al. Evaluating the Clinical Utility of Genome Sequencing for Cytogenetically Balanced Chromosomal Abnormalities in Prenatal Diagnosis. Front Genet. 2021 Jan 27;11:620162. doi: 10.3389/fgene.2020.620162. PMID: 33584815; PMCID: PMC7873444.

Affiliations and Funding Support

[a]Division of Human Genetics, Cincinnati Children's Hospital Medical Center, Cincinnati, OH, USA
[b]Department of Pediatrics, University of Cincinnati, Cincinnati, OH, USA
[c]Bionano Genomics, San Diego, CA, USA
(with funding through SMRT Structural Variation Grant)
[d]Genomics Center, University of Minnesota, Minneapolis, MN, USA
[e]Division of Neurology, Cincinnati Children's Hospital Medical Center, Cincinnati, OH, USA

This work is published in the *American Journal of Medical Genetics Part A*: Sund et al. Long-read sequencing and optical genome mapping identify causative gene disruptions in noncoding sequence in two patients with neurologic disease and known chromosome abnormalities. Am J Med Genet A. 2024 Dec;194(12):e63818. doi: 10.1002/ajmg.a.63818. PMID: 39041659.