




OPEN

## Optical genome mapping unveils hidden structural variants in neurodevelopmental disorders

Isabelle Schrauwen<sup>1</sup> , Yasmin Rajendran<sup>1</sup>, Anushree Acharya<sup>1</sup>, Susanna Öhman<sup>2</sup>, Maria Arvio<sup>3</sup>, Ritva Paetau<sup>4</sup>, Auli Siren<sup>5</sup>, Kristiina Avela<sup>6</sup>, Johanna Granvik<sup>7</sup>, Suzanne M. Leal<sup>1,8</sup>, Tuomo Määttä<sup>9</sup>, Hannaleena Kokkonen<sup>10</sup> & Irma Järvelä<sup>11</sup>

While short-read sequencing currently dominates genetic research and diagnostics, it frequently falls short of capturing certain structural variants (SVs), which are often implicated in the etiology of neurodevelopmental disorders (NDDs). Optical genome mapping (OGM) is an innovative technique capable of capturing SVs that are undetectable or challenging-to-detect via short-read methods. This study aimed to investigate NDDs using OGM, specifically focusing on cases that remained unsolved after standard exome sequencing. OGM was performed in 47 families using ultra-high molecular weight DNA. Single-molecule maps were assembled de novo, followed by SV and copy number variant calling. We identified 7 variants of interest, of which 5 (10.6%) were classified as likely pathogenic or pathogenic, located in *BCL11A*, *OPHN1*, *PHF8*, *SON*, and *NFIA*. We also identified an inversion disrupting *NAALADL2*, a gene which previously was found to harbor complex rearrangements in two NDD cases. Variants in known NDD genes or candidate variants of interest missed by exome sequencing mainly consisted of larger insertions (> 1kbp), inversions, and deletions/duplications of a low number of exons (1–4 exons). In conclusion, in addition to improving molecular diagnosis in NDDs, this technique may also reveal novel NDD genes which may harbor complex SVs often missed by standard sequencing techniques.

**Keywords** Optical genome mapping, Neurodevelopmental disorders, Structural variants, Copy number variants, Unsolved cases

Neurodevelopmental disorders (NDDs) encompass conditions linked to abnormal brain development. Those with intellectual disability (ID), marked by notable limitations in intellectual functioning and adaptive behavior, impact 1% of the global population<sup>1</sup>. Rare genetic variants, particularly in severe NDDs, have been identified as crucial contributors to their development.

Since the completion of the human genome project, our understanding of the genetic factors implicated in NDDs has greatly improved. However, current gene panel based genetic diagnostic tests based on modern short-read sequencing techniques for NDDs have a diagnostic rate of ~ 50%. This can be improved with short-read exome/genome sequencing with 16–25%<sup>2,3</sup>. This means a large gap remains in modern short-read genetic testing as it fails to identify the cause of NDDs in 25–50% of affected individuals, depending on the approach (gene panel, exome, or genome) used. Affected individuals and/or parents who still fail to receive a molecular diagnosis after such testing often embark on a diagnostic odyssey associated with years of misdirected diagnoses and treatments in addition to a large cost to our healthcare. Part of this failure to molecularly diagnose these conditions can be due to current technological limitations to assess variation in the human genome.

<sup>1</sup>Department of Neurology, Center for Statistical Genetics, Gertrude H. Sergievsky Center, Columbia University Medical Center, Columbia University, 630 W 168th St, New York, NY 10032, USA. <sup>2</sup>Kärkulla Samkommun, Kirjala, Finland. <sup>3</sup>Päijät-Häme Wellbeing Services, Neurology, Lahti, Finland. <sup>4</sup>Department of Child Neurology, University of Helsinki and Helsinki University Hospital, Helsinki, Finland. <sup>5</sup>Kanta-Häme Central Hospital, Hämeenlinna, Finland. <sup>6</sup>Institute of Biomedicine, University of Turku, Turku, Finland. <sup>7</sup>The Wellbeing Services County of Ostrobothnia, Kokkola, Finland. <sup>8</sup>Taub Institute for Alzheimer's Disease and the Aging Brain, Columbia University Medical Center, New York, NY, USA. <sup>9</sup>The Wellbeing Services County of Kainuu, Kajaani, Finland. <sup>10</sup>Northern Finland Laboratory Centre NordLab and Medical Research Centre, Oulu University Hospital and University of Oulu, Oulu, Finland. <sup>11</sup>Department of Medical Genetics, University of Helsinki, Helsinki, Finland. ✉email: is2632@cumc.columbia.edu

Current state-of-the-art techniques utilizing short-read next-generation sequencing (NGS) technologies, do not fully capture the breadth of variation present in the human genome<sup>4</sup>. Recently, several new genomic mapping and sequencing technologies have emerged that can provide a better picture of the full variant spectrum and better characterize complex structural variation<sup>5,6</sup>. These technologies have the promise to detect a larger variety of variants, complex rearrangements, assess homologous regions (“dead zones”) and highly repetitive events<sup>7</sup>. Various studies using newer technologies have reported approximately > 20 K structural variants (SVs) per human genome exist, most of which could not be detected using short-read sequencing<sup>6,8</sup>.

Optical genome mapping (OGM) is a technique with the capability to identify SVs that are undetectable or difficult to detect via short-read sequencing techniques. It fluorescently tags long, linearized DNA molecules at specific sites to create a detailed map of genomic variation including repetitive regions. This study aims to investigate the use of OGM in the molecular diagnosis of NDDs. We performed OGM in families which remained genetically unsolved after standard genomic approaches to identify pathogenic variants that had remained hidden to these techniques.

## Materials and methods

### Participant recruitment and clinical assessment

A total of 47 families including 51 affected participants with mild to profound ID, and both non-syndromic and syndromic forms, were enrolled in the study (Suppl. Table 1). This included 10 trios, 3 larger families, 4 families with one parent and/or one sibling available for the OGM analysis, and 30 singleton affected individuals. Affected individuals were clinically evaluated by a child neurologist and clinical geneticist for the study. Photographs display syndromic features from affected individuals and, in relevant cases, magnetic resonance imaging (MRI) was also obtained. Research was performed in accordance with the Declaration of Helsinki. The participants, parents or legal guardians of all participants in this study provided written informed consent to participate and additional consent was obtained for the publication of photographs.

### Sample inclusion and prior genetic testing

We selected families previously investigated via conventional exome sequencing techniques with no conclusive causative variant(s) identified. All affected individuals have a syndromic or non-syndromic neurodevelopmental disorder with intellectual disability. Unaffected individuals, including parents and unaffected siblings, were included as well in the analysis if a blood sample was available or could be obtained (Suppl. Table 1).

Exome sequencing was performed prior to optical genome mapping on all 51 affected individuals and their family members with the same exome enrichment kit, with exception of family FIN44, where one of the two affected individuals was X-exome sequenced. All exome sequencing in these 47 families resulted in inconclusive results (no likely pathogenic nor pathogenic variants), and these families also did not have a promising candidate gene found in their exome data. In short, exomic libraries were prepared using the SureSelect Human All Exon V6 kit (Agilent Technologies, Santa Clara, CA, USA) following by paired-end sequencing was performed on a HiSeq or NovaSeq instrument (Illumina Inc, San Diego, CA, USA). Data analysis was performed as described previously<sup>8</sup>. In addition to single nucleotide variants (SNVs) and small insertion/deletions (InDels), copy number variants (CNVs) were also analyzed using Copy Number Inference from Exome Reads (CoNIFERv.0.2.2)<sup>9</sup>. CNV analysis revealed an average of 8 exonic deletions and 18 duplications per sample, but did not uncover any likely pathogenic nor pathogenic variants.

The majority of families also had karyotyping and/or chromosomal microarray analysis done prior to exome sequencing with negative results (68%; Suppl. Table 2). In short, a chromosomal microarray analysis was performed from DNA extracted from peripheral blood using 50mer oligochips, the HumanCytoSNP-12 (v2.1) or Infinium CytoSNP-850K (v1.2) BeadChip (Illumina), which contain ~ 300,000 (300 K) or over 840,000 (850 K) oligonucleotide probes spaced 6–18 kbp or 1–5 kbp apart genome-wide, respectively.

### Optical Genome Mapping (OGM)

Ultra-high molecular weight (UHMW) DNA was extracted using the Prep SP Blood and Cell Culture DNA Isolation Kit (Bionano Genomics, San Diego, CA) from frozen blood for 75 samples (51 affected individuals and 24 unaffected relatives). OGM was performed on the Saphyr platform (Bionano Genomics, San Diego, USA). In short, HMW DNA was fluorescently labeled using DLE-1 followed by automated electrophoresis into the nanochannel array of a Saphyr Chip<sup>®</sup> and automated imaging of the linearized DNA by the Bionano Saphyr<sup>®</sup> instrument. The average filtered molecule N50 was 279 kbp (Suppl. Table 3).

Single-molecule maps were assembled de novo into consensus maps using the Bionano Solve™ data analysis software (v3.7), followed by SV, CNV and aneuploidy calling against the hg38/GRCh38 reference. The average coverage of the reference genome was 189x (Suppl. Table 3) and 4812 SVs were called on average for every individual (Suppl. Table 4). The following databases of SVs/CNVs were used for filtering and/or identify variants in disease-associated genomic regions: The Database of Genomic Variants (DGV)<sup>10</sup>, dbVar (NCBI) and gnomAD<sup>10</sup>. In addition, a dataset of 279 controls analyzed with OGM was been made available by Bionano Genomics, with representation from diverse ethnic backgrounds (24.6% African, 8.9% Admixed American, 9.5% East Asian, 24.6% European, 8.4% South Asian, 24% unknown) (Bionano Solve Theory of Operation; CG-30190). In addition, data was annotated with AnnotSV<sup>11</sup>, biomaRt<sup>12</sup>, which allows us to annotate genes and disease phenotypes. Samples were analyzed together with parental DNA, when available. We considered different inheritance models depending on the family structure, such as autosomal dominant (de novo or inherited), autosomal recessive, X-linked dominant and recessive. In general, we used allele frequency thresholds of  $< 5.0 \times 10^{-3}$  and  $< 5.0 \times 10^{-4}$  for recessive and dominant models. Unique variants absent from databases and controls, which occurred de novo,

were prioritized. Rare variants of interest (1–3 per family) were further manually reconstructed and curated using Bionano Access via visualization. A schematic of the workflow is presented in Supplementary Fig. 1.

Variants were followed up with various techniques, depending on the variant type. We performed Sanger sequencing and high resolution microarray to further fine map the breakpoints in variants. Variants were classified following the American College of Medical Genetics and Genomics (ACMG) and the Clinical Genome Resource (ClinGen) guidelines for sequence variants and copy number variants<sup>13,14</sup>. For intragenic CNVs, we followed adaptive recommendations for PVS1<sup>15</sup> and single gene copy number variants<sup>16</sup>. Manual investigation of the gene structure and sequence was done to review variants for loss-of-function and Franklin Software (Genoox, Tel Aviv-Yafo, Israel) was used to assist in this analysis and variant classification.

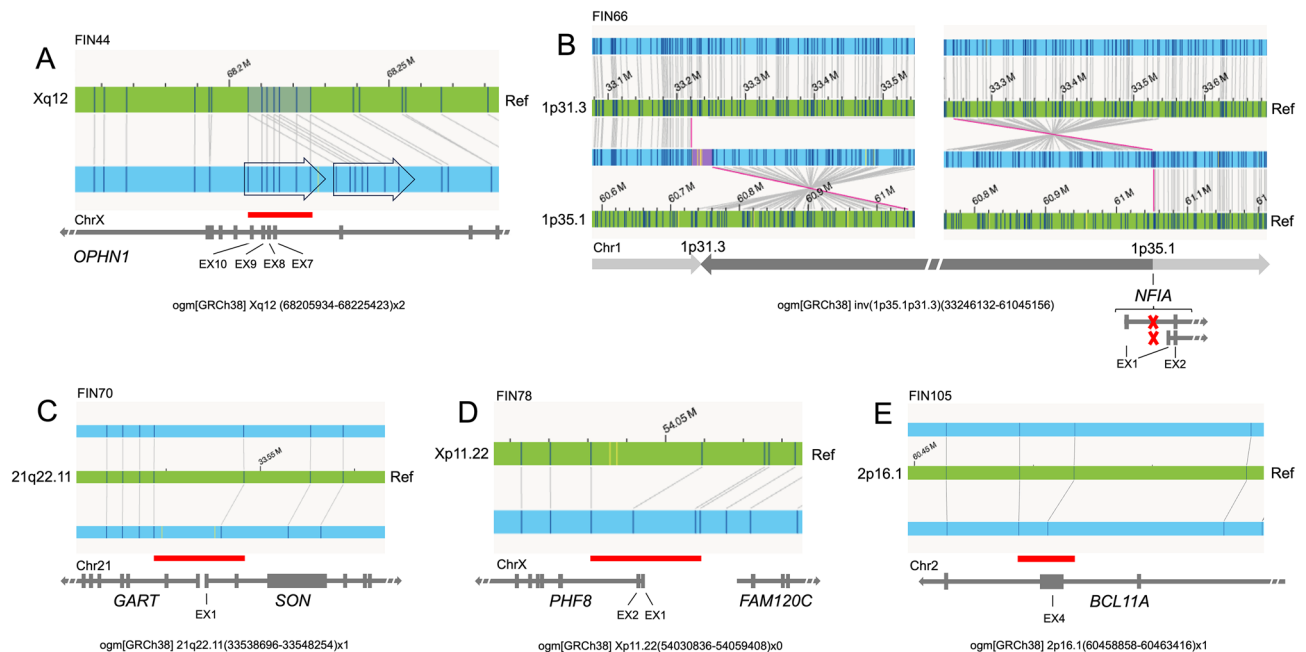
### Ethics declaration

The study was approved by the Hospital District of Helsinki (HUS 2532/2017) and the Institutional Review Board of Columbia University (IRB-AAAS433). Written consent was obtained from all participants and/or legal guardians/parents.

### Results

The majority of variants identified via OGM consisted of insertions (67.6%) and deletions (29.2%) (Suppl. Table 5). In the 47 unsolved families, OGM identified 7 rare variants of interest, including 2 variants of unknown significance and 5 likely pathogenic or pathogenic variants (10.6%). The likely pathogenic or pathogenic variants SVs were found in known NDD genes (Fig. 1; Suppl. Table 5), including *BCL11A*, *OPHN1*, *PHF8*, *SON*, and *NFIA*. A detailed clinical description of these cases is available in the Supplementary Material (Suppl. Document 1).

In family FIN44, which included two affected males and two unaffected females of a larger family described previously<sup>17</sup>, we identified a 19.5 kb intragenic hemizygous duplication in *OPHN1* ogm[GRCh38] Xq12 (68205934\_68225423) × 2 in the affected individuals (Fig. 1A; Suppl. Fig. 2A). The unaffected mother of both cases was identified as a carrier and the unaffected female sibling was not. We were able to verify this duplication using a high resolution microarray (Infinium CytoSNP-850 K v1.2 BeadChip Illumina) as arr[GRCh38] Xq12(68198545 × 168203479\_68224263 × 268231507 × 1) (Suppl. Fig. 3). The duplication covers exons 7–10 (NM\_002547.3), retains a reading frame and is predicted to lead to an 149 amino acid tandem duplication (AA163–311) in the 802-amino acid sized protein, which disturbs the BAR and PH domains, where the large majority of pathogenic variants have been identified<sup>18</sup>. The variant was classified as likely pathogenic and the clinical features are compatible with the phenotype reported in pathogenic *OPHN1* variants (#MIM 300486).



**Figure 1.** Variant identified via OGM in this study. Reference map: Green; Case map: Blue. **(A)** In family FIN44, a 19.5 kb intragenic hemizygous duplication in *OPHN1* affecting 4 exons was identified as ogm[GRCh38] Xq12 (68205934\_68225423) × 2; **(B)** In family FIN66, a mosaic 27.8 Mb balanced paracentric inversion was found as ogm[GRCh38] inv(1p35.1p31.3)(33246132\_61045156). One of the two breakpoints affects the *NFIA* gene. **(C)** In family FIN70, we identified a heterozygous 2.4 kb deletion verified to affect the first exon of *SON*, ogm[GRCh38] 21q22.11(33538696\_33548254) × 1. **(D)** In family FIN78, we identified a hemizygous 17.7 kb deletion covering the first two coding exons of *PHF8*, ogm[GRCh38] Xp11.22(54030836\_54059408) × 0. **(E)** In family FIN105, we identified a heterozygous 2.1 kb deletion ogm[GRCh38] 2p16.1(60458858\_60463416) × 1 partially deleting an exon of *BCL11A*. Molecule support for each variant is presented in Suppl. Fig. 2.

In family FIN66, a 27.8 Mb likely pathogenic mosaic (26%) balanced paracentric inversion was found in the affected individual *ogm*[GRCh38] *inv*(1p35.1p31.3) (33246132\_61045156) (Fig. 1B; Suppl. Fig. 2B). The variant was copy neutral, but one of the two breakpoints (between Chr1:61045156–61052358) disrupts the first intron of the *NFIA* gene, or the promoter based on the transcript. Several balanced and unbalanced events with breakpoints disrupting *NFIA* have been reported previously<sup>19</sup>, suggesting that this gene may be prone to rearrangement. Similar to the previous cases with variants in *NFIA* (#MIM 613735), which causes brain malformations with or without urinary tract defect<sup>19–21</sup>, FIN66 had polymicrogyria (Suppl. Fig. 4), seizures, and long-lasting feeding problems.

In family FIN70, which consisted of an affected individual and his unaffected brother, we identified a likely pathogenic heterozygous 2.4 kb deletion including *SON*, *ogm*[GRCh38] 21q22.11(33538696\_33548254) × 1 (Fig. 1C; Suppl. Fig. 2C), that was absent from the unaffected sibling. The exact breakpoints were verified with Sanger sequencing as NC\_000021.9:g.33542069\_33544417del (Suppl. Fig. 5). This deletion affects the first coding exon of *SON* and is a predicted loss of function. The clinical features of FIN70 are compatible with Zhu-Tokita-Takenouchi-Kim syndrome, caused by variants in *SON* (#MIM 617140)<sup>22</sup>.

For the affected proband of family FIN78, we identified a hemizygous 17.7 kb deletion covering *PHF8*, *ogm*[GRCh38] Xp11.22(54030836\_54059408) × 0 (Fig. 1D; Suppl. Fig. 2D). This deletion was verified via a posteriori visualization of exome data in Integrative Genomics Viewer (IGV) and with the CytoSure Medical Research Exome Array as *arr*[GRCh38] Xp11.22 (54038034\_54049437) × 0 (Suppl. Fig. 6), and was inherited from an unaffected heterozygous mother. His affected brother was also found to be hemizygous for this deletion via microarray analysis (data not shown). The deletion was classified as pathogenic, includes the first two coding exons of *PHF8* and is a predicted loss-of-function. Recently, the clinical phenotype of *PHF8* X-linked neurodevelopmental disorders was reported as variable (#MIM 300263)<sup>23</sup>. In fact, this is demonstrated in FIN78 family where the phenotype of the brothers varies quite remarkably concerning facial features (Supplementary Document 1; Suppl. Fig. 7C,D). The older brother has mild ID and cleft lip/palate. The younger brother had a diagnosis of ADHD and the variant was suspected only after the finding in the older brother. He has no cleft lip/palate.

In the affected case from FIN105, we identified a heterozygous 2.1 kb deletion *ogm*[GRCh38] 2p16.1(60458858\_60463416) × 1 covering *BCL11A* (Fig. 1E; Suppl. Fig. 2E). It was verified via Sanger sequencing as 2.3 kb NC\_000002.12:g.60458959\_60461237del (Suppl. Fig. 8). The deletion was not present in the unaffected mother. The variant deletes part of the second to last or last exon, depending on the isoform, and is predicted to lead to a frameshift and/or a truncated and mutant protein for the majority of transcripts. It was classified as likely pathogenic. Her fetal hemoglobin (Hb) was increased (8.5%, normal < 1%), a feature of Dias-Logan syndrome (#MIM 617101), and her clinical features are also compatible with this disorder<sup>24</sup>.

In two additional families, we found rare or unique SVs of interest that remain variants of unknown significance (Suppl. Figs. 9 and 10; Suppl. Table 5). In FIN25, we identified an 897 kb inversion, *ogm*[GRCh38] *inv*(3q26.31)(175056794\_175953843) that disrupts *NAALADL2* by the inversion of several exons, a possible candidate gene prone to rearrangement (Suppl. Figs. 9A and 10A). In FIN83, we identified a 3.5 kb homozygous insertion in *NUP133* *ogm*[GRCh38] *ins*(1q42.13)(229464712\_229467770) of unknown significance (Suppl. Figs. 9B and 10B).

## Discussion

This study aimed to take a more comprehensive view into the genomic landscape of rare variants implicated in NDDs via OGM. We studied 47 families in which exome sequencing did not yield a diagnosis via OGM, and in 5 families (10.6%) we found rare and likely pathogenic or pathogenic SVs in known NDD genes. In addition to improving molecular diagnosis, this technique may also uncover novel candidate genes which harbor complex SVs often missed by standard sequencing techniques. Variants in known NDD genes or candidate variants of interest missed by exome sequencing mainly consisted of inversions, deletions/duplications of a low number of exons (1–4 exons) and larger insertions (> 1 kbp).

Inversions, other balanced rearrangements and larger insertions are often invisible to many genomic techniques, including microarray, short-read exome sequencing, and short-read genome sequencing<sup>25</sup>. Challenges arise in detecting certain SVs due to limitations associated with short-reads, including restricted mappability and limited ability to span SVs. Despite improvements achieved through the integration of multiple methods (e.g., split-read, read-depth, paired-end, and assembly-based) for SV identification in short-read data<sup>26</sup>, detecting copy-neutral events remains challenging. Additionally, SV breakpoints are often flanked by repetitive elements, hindering their detection<sup>27</sup>. Therefore, short-read genome sequencing still fails to detect the majority of insertions (83%) and inversions (59%), even when using 13 different SV detection algorithms<sup>28</sup>. Indeed, estimates based on newer technologies suggest approximately 22–27 K SVs (> 50 bp) per healthy genome exist, most are not detected by short reads<sup>28,29</sup>. Integrating different technologies (long-read, short-read, linked-read, strand-seq, Hi-C, OGM) can lead to a sevenfold increase in SV detection compared to short-read WGS<sup>28</sup>. Via OGM, we identified a likely pathogenic inversion in an unsolved NDD case that was undetectable by exome sequencing and microarray.

Additionally, through OGM, we successfully identified small copy number variants < 20 kb (CNVs). These small CNVs are frequently elusive to microarray detection due to limited resolution. Although this is based on probe spacing on the array used, this is often limited to 5–10 kb, and frequently extending to tens of kilobases<sup>30</sup>. The ~ 2 kb deletions in *SON* and *BCL11A* would therefore evade detection in such cases, and the 17.7 kb and 19.5 kb CNVs would also not be detected on many standard microarrays with limited resolution. In addition, small CNVs also pose a challenge for exomes due to limited resolution. Most exome CNV calling algorithms typically identify CNVs spanning 4–10 exons<sup>31</sup>. In contrast, OGM enabled the identification of the small CNVs approximately ~ 2 kb in size in two instances encompassing either a single exon or a partial single exon. Additionally, the two CNVs of 17.7 kb and 19.5 kb spanning 2–4 exons each, were also detected via OGM.

In one family, we found a rare and unique mosaic inversion in the gene *NAALADL2*. The inversion is surrounded by repeats, encompassing long terminal repeats (LTR) retrotransposon repeats and long interspersed elements (LINE-1), suggesting a potential role in facilitating this genetic rearrangement. Previously, complex rearrangements in *NAALADL2* have been found in two cases with a neurodevelopmental disorder<sup>32,33</sup>. *NAALADL2* is also a frequently altered fragile sites in the cancer genome<sup>34</sup>. This suggests that this gene may be prone particularly to complex rearrangements, yet loss-of-function structural variants in this gene are rare in gnomADv4.0.0<sup>10</sup>. *NAALADL2* is a member of the glutamate carboxypeptidase II family and involved in Glutamate metabolism<sup>35</sup>. The latter is a prevalent and abundant excitatory neurotransmitter within the central nervous system. This could imply that this gene is a potential candidate as the causative factor for a neurodevelopmental disorder.

Previous research on OGM in cases of NDDs, along with other constitutional cases, has predominantly centered on comparing results between OGM and conventional diagnostic tests, instead of focusing on identifying variants in unsolved cases. These studies have shown high concordance (98% with CMA; 98.6% with one or more standard-of-care diagnostic tests) and highlight OGM as a platform that can detect a large number of classes of SVs in the genome<sup>36–38</sup>. In addition, OGM has been shown to be a helpful technique to further reconstruct complex SVs in NDDs<sup>38,39</sup>. However, a systematic investigation of unsolved cases in NDDs, to our knowledge, has been missing. Nonetheless, some studies have found OGM to be useful in the diagnosis of several types of unsolved constitutional diseases. In Duchenne muscular dystrophy (DMD), a severe progressive muscle disease, a study was able to identify an inversion disrupting the *DMD* gene (~1.28 Mb) missed by exome sequencing, MLPA and Sanger sequencing<sup>40</sup>. In a case with a retinal disease with negative genome sequencing data, OGM identified a pericentric inversion with 1 breakpoint disrupting *USH2A*. Retrospectively, the variant could be observed via genome sequencing data but was previously considered a false positive<sup>41</sup>.

The study's limitations involve the need for a more comprehensive database of controls analyzed with OGM, which would aid in data filtering. We anticipate an updated resource will become available soon (personal communication with Bionano Genomics). A larger dataset and routine implementation into diagnostics may reveal a better diagnostic rate of NDDs via the implementation of this technique.

In conclusion, this investigation sought to further study unsolved cases with a neurodevelopmental disorder via OGM, and gain a broader understanding of the genomic spectrum of variants associated with NDDs. This approach unveiled various types of variants implicated in NDDs frequently overlooked by short-read sequencing methods, including small CNVs, insertions, and balanced inversions.

## Data availability

Variants reported in this study were deposited into ClinVar under accession numbers: SCV004183567-SCV004183571 (<https://www.ncbi.nlm.nih.gov/clinvar/>). Optical genome mapping data was deposited into dbGAP under phs002937.v1.p1.

Received: 4 January 2024; Accepted: 13 May 2024

Published online: 16 May 2024

## References

- Maulik, P. K., Mascarenhas, M. N., Mathers, C. D., Dua, T. & Saxena, S. Prevalence of intellectual disability: A meta-analysis of population-based studies. *Res. Dev. Disabil.* **32**, 419–436 (2011).
- Shashi, V. *et al.* The utility of the traditional medical genetics diagnostic evaluation in the context of next-generation sequencing for undiagnosed genetic disorders. *Genet. Med.* **16**, 176–182 (2014).
- Thevenon, J. *et al.* Diagnostic odyssey in severe neurodevelopmental disorders: Toward clinical whole-exome sequencing as a first-line diagnostic test. *Clin. Genet.* **89**, 700–707 (2016).
- Marks, P. *et al.* Resolving the full spectrum of human genome variation using Linked-Reads. *Genome Res.* **29**, 635–645 (2019).
- Sedlazeck, F. J., Lee, H., Darby, C. A. & Schatz, M. C. Piercing the dark matter: Bioinformatics of long-range sequencing and mapping. *Nat. Rev. Genet.* **19**, 329–346 (2018).
- Wang, O. *et al.* Efficient and unique cobarcoding of second-generation sequencing reads from long DNA molecules enabling cost-effective and accurate sequencing, haplotyping, and de novo assembly. *Genome Res.* **29**, 798–808 (2019).
- Barseghyan, H. *et al.* Next-generation mapping: A novel approach for detection of pathogenic structural variants with a potential utility in clinical diagnosis. *Genome Med.* **9**, 90 (2017).
- Järvelä, I. *et al.* Exome sequencing reveals predominantly de novo variants in disorders with intellectual disability (ID) in the founder population of Finland. *Hum. Genet.* **140**, 1011–1029 (2021).
- Gambin, T. *et al.* Homozygous and hemizygous CNV detection from exome sequencing data in a Mendelian disease cohort. *Nucleic Acids Res.* **45**, 1633–1648 (2017).
- Karczewski, K. J. *et al.* The mutational constraint spectrum quantified from variation in 141,456 humans. *Nature* **581**, 434–443 (2020).
- Geoffroy, V. *et al.* AnnotSV: An integrated tool for structural variations annotation. *Bioinformatics* **34**, 3572–3574 (2018).
- Zhang, J. *et al.* BioMart: A data federation framework for large collaborative projects. *Database (Oxford)* **2011** (2011).
- Riggs, E. R. *et al.* Technical standards for the interpretation and reporting of constitutional copy-number variants: A joint consensus recommendation of the American College of Medical Genetics and Genomics (ACMG) and the Clinical Genome Resource (ClinGen). *Genet. Med.* **22**, 245–257 (2020).
- Richards, 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.* **17**, 405–424 (2015).
- Abou Tayoun, A. N. *et al.* Recommendations for interpreting the loss of function PVS1 ACMG/AMP variant criterion. *Hum. Mutat.* **39**, 1517–1524 (2018).
- Brandt, T. *et al.* Adapting ACMG/AMP sequence variant classification guidelines for single-gene copy number variants. *Genet. Med.* **22**, 336–344 (2020).
- Philips, A. K. *et al.* X-exome sequencing in Finnish families with intellectual disability—four novel mutations and two novel syndromic phenotypes. *Orphanet. J. Rare Dis.* **9**, 49 (2014).
- Schwartz, T. S. *et al.* Expanding the phenotypic spectrum associated with OPHN1 variants. *Eur. J. Med. Genet.* **62**, 137–143 (2019).

19. Lu, W. *et al.* NFIA haploinsufficiency is associated with a CNS malformation syndrome and urinary tract defects. *PLoS Genet.* **3**, e80 (2007).
20. Bayat, A., Kirchhoff, M., Madsen, C. G., Roos, L. & Kreiborg, S. Familial craniofacial abnormality and polymicrogyria associated with a microdeletion affecting the NFIA gene. *Clin. Dysmorphol.* **26**, 148–153 (2017).
21. Negishi, Y. *et al.* Truncating mutation in NFIA causes brain malformation and urinary tract defects. *Hum. Genome Var.* **2**, 15007 (2015).
22. Kim, J.-H. *et al.* De Novo mutations in SON disrupt RNA splicing of genes essential for brain development and metabolism, causing an intellectual-disability syndrome. *Am. J. Hum. Genet.* **99**, 711–719 (2016).
23. Sobering, A. K. *et al.* Variants in PHF8 cause a spectrum of X-linked neurodevelopmental disorders and facial dysmorphism. *HGG Adv.* **3**, 100102 (2022).
24. Dias, C. *et al.* BCL11A haploinsufficiency causes an intellectual disability syndrome and dysregulates transcription. *Am. J. Hum. Genet.* **99**, 253–274 (2016).
25. Kadlubowska, M. K. & Schrauwen, I. Methods to improve molecular diagnosis in genomic cold cases in pediatric neurology. *Genes (Basel)* **13**, 333 (2022).
26. English, A. C. *et al.* Assessing structural variation in a personal genome-towards a human reference diploid genome. *BMC Genom.* **16**, 286 (2015).
27. Ho, S. S., Urban, A. E. & Mills, R. E. Structural variation in the sequencing era. *Nat. Rev. Genet.* **21**, 171–189 (2020).
28. Chaisson, M. J. P. *et al.* Multi-platform discovery of haplotype-resolved structural variation in human genomes. *Nat. Commun.* **10**, 1784 (2019).
29. Audano, P. A. *et al.* Characterizing the major structural variant alleles of the human genome. *Cell* **176**, 663–675.e19 (2019).
30. Batzir, N. A., Shohat, M. & Maya, I. Chromosomal microarray analysis (CMA) a clinical diagnostic tool in the prenatal and post-natal settings. *Pediatr. Endocrinol. Rev.* **13**, 448–454 (2015).
31. Gordeeva, V. *et al.* Benchmarking germline CNV calling tools from exome sequencing data. *Sci. Rep.* **11**, 14416 (2021).
32. Tonkin, E. T. *et al.* A giant novel gene undergoing extensive alternative splicing is severed by a Cornelia de Lange-associated translocation breakpoint at 3q26.3. *Hum. Genet.* **115**, 139–148 (2004).
33. Borg, K. *et al.* Molecular analysis of a constitutional complex genome rearrangement with 11 breakpoints involving chromosomes 3, 11, 12, and 21 and a approximately 0.5-Mb submicroscopic deletion in a patient with mild mental retardation. *Hum. Genet.* **118**, 267–275 (2005).
34. Simpson, B. S., Pye, H. & Whitaker, H. C. The oncological relevance of fragile sites in cancer. *Commun. Biol.* **4**, 567 (2021).
35. Whitaker, H. C. *et al.* N-Acetyl-L-aspartyl-L-glutamate peptidase-like 2 is overexpressed in cancer and promotes a pro-migratory and pro-metastatic phenotype. *Oncogene* **33**, 5274–5287 (2014).
36. Mantere, T. *et al.* Optical genome mapping enables constitutional chromosomal aberration detection. *Am. J. Hum. Genet.* **108**, 1409–1422 (2021).
37. Broeckel, U. *et al.* Detection of constitutional structural variants by optical genome mapping: A multisite study of postnatal samples. *J. Mol. Diagn.* **26**, 213–226 (2024).
38. Barseghyan, H. *et al.* Comparative benchmarking of optical genome mapping and chromosomal microarray reveals high technological concordance in CNV identification and additional structural variant refinement. *Genes (Basel)* **14**, 1868 (2023).
39. Nicolle, R. *et al.* 16p13.11p11.2 triplication syndrome: A new recognizable genomic disorder characterized by optical genome mapping and whole genome sequencing. *Eur. J. Hum. Genet.* **30**, 712–720 (2022).
40. Erbe, L. S. *et al.* Exome sequencing and optical genome mapping in molecularly unsolved cases of duchenne muscular dystrophy: Identification of a causative X-chromosomal inversion disrupting the DMD gene. *Int. J. Mol. Sci.* **24**, 14716 (2023).
41. de Bruijn, S. E. *et al.* Optical genome mapping and revisiting short-read genome sequencing data reveal previously overlooked structural variants disrupting retinal disease-associated genes. *Genet. Med.* **25**, 100345 (2023).

## Acknowledgements

We thank the participants for their partaking in this study.

## Author contributions

Conceptualization: I.S., I.J.; Data curation: I.S., Y.R., A.A.; Formal analysis: I.S., Y.R., A.A., H.K.; Investigation: S.O., M.A., R.P., A.S., K.A., J.G., T.M., H.K.; Resources: S.M.L, Visualization: I.S.; Writing-original draft: I.S., I.J.; Writing-review & editing: All authors.

## Funding

This research was supported by a pilot Grant from the Columbia University Sergievsky Center (Gertrude H. Sergievsky Center Award), R21 NS123325 to IS and R01 HD109342 to SML and IS.

## Competing interests

The authors declare no competing interests.

## Additional information

**Supplementary Information** The online version contains supplementary material available at <https://doi.org/10.1038/s41598-024-62009-y>.

**Correspondence** and requests for materials should be addressed to I.S.

**Reprints and permissions information** is available at [www.nature.com/reprints](http://www.nature.com/reprints).

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2024