Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Closing the diagnostic gap in developmental disorders through comprehensive long-read sequencing

Kris Van Den Bogaert

kris.vandenbogaert@uzleuven.be

KU Leuven

Mathilde Geysens

KU Leuven

Benjamin Huremagic

KU Leuven

Erika Souche

University of Leuven - Gasthuisberg - Leuven

Greet Peeters

KU Leuven

Natalia Olszewska

KU Leuven

Senne Meynants

KU Leuven https://orcid.org/0009-0006-7055-3737

Giancarlo Mancuso

KU Leuven

Jeroen Breckpot

Catholic University Leuven, Belgium https://orcid.org/0000-0002-0734-5154

Ellen Denayer

UZ Leuven

Koenraad Devriendt

Catholic University Leuven https://orcid.org/0000-0002-3651-2548

Hilde Peeters

KU Leuven

Björn Menten

Ghent University https://orcid.org/0000-0001-8182-659X

Geert Mortier

Antwerp University Hospital and University of Antwerp https://orcid.org/0000-0001-9871-4578

Annick Vogels

University Hospitals Leuven
Hilde Van Esch
KU Leuven https://orcid.org/0000-0002-9604-5726
Joris Vermeesch
KU Leuven

Article

Keywords: long-read sequencing, nanopore sequencing, developmental disorders, structural variants, episignatures, X-chromosome inactivation

Posted Date: April 27th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-9108777/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: There is **NO** Competing Interest.

Closing the diagnostic gap in developmental disorders through comprehensive long-read sequencing

Mathilde Geysens[1,2*], Benjamin Huremagic[1*], Erika Souche[1], Greet Peeters[1], Natalia Olszewska[1], Senne Meynants[1], Giancarlo Mancuso[1], Jeroen Breckpot[2], Ellen Denayer[2], Koenraad Devriendt[2], Hilde Peeters[2], Björn Menten[3], Geert Mortier[2], Annick Vogels[2], Hilde Van Esch[2], Joris Robert Vermeesch[1,2†], Kris Van Den Bogaert[1,2†]

[1]Laboratory of Cytogenetics and Genome Research, Center for Human Genetics, KU Leuven, Leuven 3000, Belgium.

[2] Department of Human Genetics, Center for Human Genetics, University Hospitals Leuven, Leuven 3000, Belgium.

[3]Department of Biomolecular Medicine, Ghent University, Ghent 9000, Belgium

*Joint first authors

†Co-last authors

CORRESPONDANCE

Prof. Kris Van Den Bogaert
kris.vandenbogaert@uzleuven.be

Prof. Joris Vermeesch
joris.vermeesch@uzleuven.be

Laboratory for Cytogenetics and Genome Research
Department of Human Genetics
ON1 Herestraat 49 - box 606
3000 Leuven - Belgium

Keywords: long-read sequencing, nanopore sequencing, developmental disorders, structural variants, episignatures, X-chromosome inactivation

ABSTRACT

Purpose: A substantial fraction of structural variants (SVs) and complex genomic regions escapes detection by short-read sequencing, while clinical epigenetic analyses remain restricted to a small number of targeted assays. Long-read sequencing (LRS) technologies represent a promising approach to reduce the diagnostic gap in genetic testing for developmental disorders (DD).

Methods: We developed a comprehensive nanopore LRS analytical pipeline enabling detection of genetic and epigenetic variation and applied it to 28 patient-parents trios with unsolved DD following short read exome or genome sequencing.

Results: LRS resolved nine (32.1%) of previously undiagnosed individuals and identified potential diagnoses in an additional six (21.4%). LRS was essential for the detection of three SVs of interest and one single nucleotide variant (SNV) and allowed detection of three (likely) pathogenic SNVs missed by exome sequencing. Systematic methylation analysis revealed skewed X-inactivation, an isolated imprinting defect as well as disease-associated episignatures.

Conclusion: The added diagnostic yield attributable to LRS was 7.1%, rising to 14.3% when potential diagnoses were included, highlighting the value of LRS in reducing the diagnostic gap while streamlining the diagnostic workflow. Concurrent methylation profiling provides functional context that may help prioritize and interpret SVs, addressing one of the major remaining challenges associated with their increased detection.

BACKGROUND

Since the first conventional karyotypes and the initial description of Trisomy 21 as a genetic cause of intellectual disability (ID) in 1959, technological advances have been instrumental for the discovery of molecular mechanisms underlying developmental disorders (DD). Following the application of Sanger sequencing, and later chromosomal microarrays (CMA), the advent of massively parallel sequencing (MPS) ~15 years ago further accelerated the discovery of novel gene–disease associations and enabled the analysis of all genes within a single assay[1]. This represented a major advance in a field characterized by recognizable but also nonspecific phenotypes, which may be caused by variants in hundreds of different genes. Clinical implementation of exome sequencing (ES) has yielded diagnostic rates of approximately 30-40%[2,3], with extension to whole-genome sequencing (GS) providing an additional 0–10%[4–7]; however, more than half of patients with DD remain without molecular diagnosis. Although a subset of DD has non-genetic etiologies, a substantial fraction of unsolved cases is believed to result from genetic variation that escapes detection by current technologies[8].

While MPS enables genome-wide detection of single-nucleotide variants (SNVs) and small insertions and deletions (indels), additional methodologies are still required for the detection of copy number variants (CNVs), short tandem repeats (STRs) such as Fragile-X syndrome, and imprinting disorders (*e.g.* Angelman syndrome). Despite the integration of these complementary assays in clinical care, part of human genetic variation remains undetected. Structural variants (SVs) contribute more to nucleotide-level diversity between human genomes than any other class of genetic variation[9]; however, significant challenges persist in their detection. CNVs in the 50 bp–50 kb size range lie between the effective resolution of CMA and short-read MPS (SRS), which also fail to detect insertions, translocations, and complex chromosomal rearrangements (CCR) below the resolution of conventional karyotyping. In addition, STRs and methylation analyses are confined to targeted assays, limiting their systematic evaluation. Short reads are also difficult to align in low or high complexity genomic regions, leaving part of the genome unassessed.

Over the last decade, long-read sequencing technologies (LRS) emerged, enabling the sequencing of megabase-long DNA strands without clonal amplification, bridging SVs and STRs while maintaining base modifications[10]. LRS was shown to allow the detection of more than 90% of variants requiring complementary analyses in diagnostics laboratories[11]. Importantly, LRS also sheds light on >40% additional SVs, undetectable with other technologies, and was demonstrated to permit the assembly of

highly contiguous haplotypes[12], enabling the characterization of complex SVs at the base pair level[13]. Moreover, LRS-based methylome analyses are suitable for the assessment of all imprinted loci[14] as well as the detection of skewed X chromosome inactivation (XCI)[15] and episignatures[16]. Beyond enhancing the detection of (epi)genetic variation, LRS uncovered 200 Mb of sequence missing in hg38[17], leading to the new T2T reference genome, resolving the structure of 207 protein-coding genes, and reducing false positives in medically relevant genes by up to 12-fold[18].

The added value of LRS for DD was first demonstrated in targeted settings, confirming clinical diagnoses or identifying second hits for recessive diseases[19]. Small proof of concept studies also evaluated the added value of PacBio LRS in 5 to 8 individuals with unsolved DD[20–22], identifying a (possible) diagnosis in 2/18 individuals evaluated across the three studies. More recently, three larger cohorts applying Oxford Nanopore or PacBio LRS for 41 to 114 trios reported yields varying between 10 and 27%[23–25]. However, two of the three studies included multiple individuals with either Duchenne muscular dystrophy or congenital adrenal hyperplasia. The former is caused by variants in the deletion-prone *DMD* gene[26], in which LRS enabled the identification of inversions; the latter results from variants in the *CYP21A2*-gene, a gene for which variant detection has been shown to benefit from LRS given the presence of a pseudogene[27]. Diagnoses also included multiple STR expansions in the context of spinocerebellar ataxia[23] and large CNVs in patients that did not undergo prior CMA[25]. A recent study prospectively comparing LRS and standard of care (SoC) in 100 individuals with neurological disorders observed an improved characterization of detected variants using LRS but no difference in diagnostic yield[28]. Further broad and untargeted studies are therefore still needed to assess the clinical value of LRS for DD, including systematic assessments of the impact of concurrent epigenetic analyses.

In the present study, we evaluated the diagnostic value of LRS in a cohort of 28 individuals with unsolved DD, using a comprehensive pipeline developed to systematically assess not only genetic but also epigenetic variation.

RESULTS

Participants inclusion and sequencing

Thirty patients with DD, including 28 probands and 2 affected siblings, characterized by ID and/or congenital anomalies (CA), were included in the study. All remained unsolved after SoC genetic testing

including CMA and ES (n=22) or short read GS (n=6) (Suppl. Table 1A). High molecular weight DNA was extracted from all probands and their parents and used for Oxford nanopore sequencing, using one Promethion flow cell per sample, generating an average of 86.1 ± 24.7 GB per sample with an average read length N50 of 41.6 ± 13.6 kb (Suppl. Table 1B). Following base- and methylation calling, reads were aligned both to the hg38 and T2T reference genomes. A systematic analysis of SNVs, SVs, CNVs and STRs was performed for all probands. In addition, comprehensive profiling of disease-associated methylation abnormalities – including XCI, imprinting loci and episignatures – was conducted (see *Methods*, Figure 1).

Single nucleotide variants

The analysis of SNVs provided a molecular diagnosis for 8 participants and shed light on potential diagnoses in an additional 4 participants (Table 1). All variants were confirmed by orthogonal methods (Suppl. Table 1C). Two of the diagnoses involved non-coding variants which could not be detected by ES: the recurrent *RNU4-2* n.64_65insT variant[29] as well as an intronic variant in *TBCK* NM_001163435.3:c.782+46A>G, located in *trans* of the pathogenic frameshift variant NM_001163435.3:c.741_744del p.(Asp248fs). The impact on splicing of the c.782+46A>G variant was confirmed on cDNA (insertion of 45bp between exons 9 and 10), resulting in an in-frame insertion of 15 amino acids (Suppl. Figure 1). Two participants were diagnosed with pathogenic variants that remained unidentified during SoC because the gene-disease associations for *NUSAP1* and *CRELD1* have only recently been described[30,31]. Variants of unknown significance (VUS) in *ASHL1, LONP1* and *OSBP* were retained as strong candidate diagnoses following new phenotypic insights from GeneMatcher exchanges[32]. A pathogenic variant in *SLC6A1* and a VUS in *COL11A1* were filtered out during SoC because these variants cause autosomal dominant diseases and were inherited from seemingly unaffected parents. Both mutation carrier parents only presented minor signs of the disease, in contrast to the more severely affected children. Finally, three pathogenic variants were missed using short reads due to technical limitations. The *SETD5* indel had been filtered out of the SR ES *de novo* analysis as a result of insufficient quality of the parental data at the locus and the *SHANK3* variant was located in an exon with insufficient coverage through ES (sequencing gap). The *SYNGAP1* variant was filtered out of the SR ES analysis due to a low quality score resulting from a difficult mapping of short reads in the affected genomic region. The latter would also have been missed applying SR GS since similar mismatches and low quality sequencing is observed in SR GS data (Suppl Figure 2).

Structural variants

Compared to the reference genome hg38 / T2T, an average of 23058 / 21136 SVs were called per sample, including 13154 / 9894 insertions, 9793 / 11189 deletions, 29 / 21 inversions, 9 / 9 duplications and 72 / 23 translocations. Upon inheritance filtering, 87 / 117 *de novo* variants were retained, reduced to 71 / 71 when taking only 'precise specific' calls from Sniffles2 into account. Manual curation using IGV showed most of these to be false calls. A substantial proportion (30-45%) of false calls were located in the pericentromeric regions. Sniffles2 was set up to detect SVs based on minimum two supporting reads to increase sensitivity for potential mosaic SVs. The quality of the calling using Sniffles2 can be further improved by using a minimum of three supporting reads instead of two (Suppl. Figure 3). The amount of SVs to assess and interpret was reduced by filtering both frequent SVs and recurrent false calls using 50 parents as control population. Filtering out all SVs retrieved in any of the 50 parents reduced the number of SVs to an average of 8.9 / 11.1 *de novo* and 2.5 / 3.4 X-linked variants per sample. Manual curation retained 3 / 5 *de novo* deletions and one *de novo* inversion across the whole cohort, as well as 18 / 22 X-linked variants in the 13 male participants (Figure 2). All *de novo* variants retained with the hg38 reference genome were also detected using T2T. Of the X-linked variants, two were detected exclusively using hg38, and six were detected exclusively using T2T. Orthogonal validation of the *de novo* insertion and the 3 *de* novo deletions identified using hg38 was performed through PCR and gel electrophoresis (Suppl. Figure 4). Two *de novo* and one X-linked SV were potentially associated with the patient's phenotype (Figure 3). The remaining three *de novo* deletions could not be linked with the patient's phenotype with current knowledge and functional annotation.

The *de novo* deletion identified in P8 is a heterozygous 11.1 kb deletion (2:47919650-47929780, hg38) located upstream the *FBXO11* gene, a gene for which loss-of-function is known to cause a neurodevelopmental disorder with phenotypic features overlapping the characteristics of the participant[33]. Given the phenotypic overlap and proximity to enhancer sequences, we hypothesized that the deletion might have an impact on the regulation of *FBXO11*'s expression. Local methylation analysis excluded an effect on promoter methylation. Evaluation of the gene expression on the patient's EBV-transformed cells through qPCR did not allow to confirm a reduced expression of the *FBXO11* gene (Suppl. Figure 5). Yet, these results using EBV-transformed cells do not exclude a potential regulatory effect in other tissues (or times) of interest, as the activity of some enhancers has been demonstrated

to be tissue-specific[34]. The GTEx database points to the predominant expression of different isoforms in brain tissues as compared to EBV cell lines (Suppl. Figure 6). Hence this SV remains a VUS.

The *de novo* 444.9 kb inversion that was detected in P22 (21:37022957-37467808, hg38) was found to be pathogenic as one of the breakpoints is disrupting the *DYRK1A* gene, a gene of which haploinsufficiency is leading to a neurodevelopmental disorder characterized by microcephaly, epilepsy, moderate to severe ID and facial features matching the phenotype of P22[35]. The other breakpoint is not affecting a gene or known regulatory sequence. The impact of the inversion was supported by phenotypic evaluation using Face2Gene, which yielded a high gestalt score for *DYRK1A* (Suppl. figure 7), as well as episignature analysis (cfr. infra).

Among X-linked variants, an 114.6 kb inversion involving *CLIC2* was detected in P23 (X:155326442-155441024, hg38). Manual inspection of the coverage and read alignments revealed a CCR of the region, involving both an inversion and duplication(s). Whereas prior CMA data and optical genome mapping confirmed the duplication, they did not detect the inversion, which was confirmed by PCR spanning the breakpoints (Suppl. Figure ^²3). Duplications involving the int22h-1/int22h-2 LCR-flanked region and overlapping the identified Xq28 inversion have been associated with ID, with *CLIC2* and/or *RAB39B* suspected to be dosage-sensitive genes contributing to variant pathogenicity[36]. While the participant shared phenotypic features with previous descriptions of patients with Xq28 duplications, segregation analyses through PCR showed the rearrangement to be inherited from the unaffected maternal grandfather. Two recent publications also reported a reduced penetrance of the Xq28 duplications[37,38], suggesting the effect of variants in this region might be much lower than initially suspected. Additional LRS of the grandfather confirmed hemizygosity for the same complex rearrangement, excluding the possibility that the variant arose *de novo* in the mother mediated by an inversion present in the grandfather. The contribution of this variant to the phenotype of this patient therefore remains uncertain.

To investigate SVs potentially involved in autosomal dominant inherited or autosomal recessive diseases, we also evaluated all inherited SVs absent from our reference set of 50 parents (mean of 28 SVs per individual). This analysis identified carriership for α-thalassemia in one individual, caused by a 3.8kb deletion encompassing *HBA2,* but did not identify (likely) pathogenic SVs associated with the patients' DD.

Evaluation of the DD-associated STR expansions in the *AFF2*, *AFF3*, *DIP2B*, and *FMR1* genes did not detect pathogenic expansions (Figure 3). One of the mothers carried an intermediate *FMR1* allele (51 repeats). As pathogenicity thresholds are less well established for the three other genes, methylation was evaluated for the four outlier repeat lengths and was found to be normal (Suppl. Figure 8).

Table 1 : Overview of identified variants of interest.

Study ID	Variant type	Gene	Variant g.	Transcript	Variant c. p.	Inheritance	Variant class
P0	missense	*ASH1L*	1:155357722 A>G	NM_018489.3	c.6823T>C p.(Cys2275Arg)	*de novo*	VUS
P1	splice	*COL11A1*	1:102984138 C>T	NM_001854.4	c.2556G>A p.(Lys852=)	paternally inherited	VUS
P2 & P2B	missense	*LONP1*	19:5692073 A>C	NM_004793.4	c.2839T>G p.(Phe947Val)	mat. & pat. inherited*	VUS
P6	frameshift	*TBCK*	4:106248282 GATCT>G	NM_001163435.3	c.741_744del p.(Asp248fs)	maternally inherited	likely pathogenic
	splice	*TBCK*	4:106248199 T>C	NM_001163435.3	c.782+46A>G	paternally inherited	likely pathogenic
P7	missense	*OSBP*	11:59615319 G>A	NM_002556.3	c.346C>T p.(Leu116Phe)	*de novo*	VUS
P8	11.13kb deletion	*FBXO11*	2:47919650-47929780			*de novo*	VUS
P16	missense	*SLC6A1*	3:11025842 G>A	NM_003042.4	c.919G>A p.(Gly307Arg)	paternally inherited	pathogenic
P17	nonsense	*SHANK3*	22:50731148 G>T	NM_001372044.2	c.5257G>T p.(Glu1753*)	*de novo*	likely pathogenic
P20	frameshift	*CRELD1*	3:9943425 CA>C	NM_001077415.3	c.959del p.(Gln320fs)	paternally inherited	pathogenic
	missense	*CRELD1*	3:9940964 G>A	NM_001077415.3	c.575G>A p.(Cys192Tyr)	maternally inherited	pathogenic
P21	nonsense	*NUSAP1*	15:41377281 C>A	NM_016359.5	c.1209C>A p.(Tyr403*)	*de novo*	pathogenic
P22	inversion	*DYRK1A,...*	21:37022957-37467808			*de novo*	pathogenic
P23	chromosomal rearrangement	*CLIC2,...*				maternally inherited	risk factor
P24	non-coding RNA	*RNU4-2*	12:120291839 T>TA	NR_002750.2	n.64_65insT	*de novo*	pathogenic
P27	frameshift	*SYNGAP1*	6:33432684 G>GT 6:33432688 GGCTTCCTGA>G	NM_006772.3	c.388-1_388insT, c.394_402del	*de novo*	likely pathogenic
P28	frameshift	*SETD5*	3:9447103 GAA>G	NM_001080517.3	c.1581_1582del p.(Arg528Lysfs*14)	*de novo*	likely pathogenic

Summary of identified variants of interest. The individuals with conclusive molecular diagnoses, corresponding to (likely) pathogenic variants that are considered to explain the phenotype are indicated in bold. Variants of unknown significance (VUS) with a high suspicion of potential association to the phenotype, and (likely) pathogenic variants with an uncertain contribution to the phenotype (e.g. risk factor identified in P23), are considered as potential diagnoses. *The *LONP1* variant identified in P2 and his brother P2B is homozygous. All other variants are heterozygous.

Epigenetic variation

Systematic evaluation of imprinting loci revealed the expected allelic differential methylation across all individuals and loci, with the exception of the *MEG8* locus in P19B, the brother of P19. For this locus, both haplotypes were found to be unmethylated, in contrast to the expected pattern in which only the

paternal allele is unmethylated. Interestingly, the other DMRs of this region, *MEG3* and the 'intergenic' DMR[39], exhibited normal differential methylation patterns (Figure 4). Evaluation of read-level data and coverage excluded an underlying deletion. SNP analysis revealed a localized loss of heterozygosity within the region; however this finding was consistent with the inherited parental haplotypes, identical at this locus. This imprinting disturbance was not detected in P19, the brother of P19B, who presented with a similar specific phenotype (jejunal atresia and hydrocephaly). It is therefore unlikely that this epigenetic variant underlies their peculiar clinical presentation. No variants were identified in *PLAG1,* a gene that has recently been associated with isolated *MEG8* imprinting defects[40].

Systematic evaluation of XCI revealed skewing in 3/28 mothers (P9M, P19M, P21M) and 4/15 female patients (P0, P3, P12, P24) (Suppl. Table 1D). Notably, DMRs involved in XCI consistently showed greater CpG methylation variability on the methylated allele compared with imprinted DMRs (Figure 4, Suppl. Figure 9). Two of the four female patients (P0 and P3) with skewed XCI presented depigmented maculae, supporting an underlying mosaicism or X-linked disorder. One of the mothers with skewed XCI (P19M) has three affected sons with a highly similar phenotype, also consistent with an X-linked recessive inheritance pattern. However, no (likely) pathogenic variants on chromosome X could be identified.

Episignature analysis using our in house developed and public tool[16] identified a positive Tatton-Brown-Rahman syndrome (TBRS) signature in P7. The SVM score (0.378) exceeded the positive threshold but remained substantially lower than that of a positive TBRS control (1.869). Two-dimensional UMAP analysis of methylation levels at TBRS episignature loci[41] confirmed similarities between the methylation profile of P7 and the TBRS control (Suppl. Figure 10). Despite phenotypic similarities between P7 and TBRS, thorough *DNMT3A* analysis, including mosaic variant assessment, did not identify any (likely) pathogenic variants. Analysis of rare *de novo* variants identified a VUS in *OSBP,* a gene functioning within molecular networks including *ARF1* and *CERT1*, two genes previously implicated in intellectual disability[42,43]. Through GeneMatcher, the same variant was found to be shared with another individual exhibiting similar phenotypic features. Further studies are required to determine whether *OSBP* variants are associated with consistent methylation changes, potentially overlapping with the TBRS methylation signature. Two-dimensional UMAP analysis of methylation levels at DYRK1A syndrome episignature loci, reported by Courraud *et al.*[35], demonstrated a close clustering of P22 with three positive controls (Figure 4), confirming the pathogenicity of the identified inversion. Similarly, UMAP analysis of

methylation levels at KBG syndrome episignature loci[44] demonstrated a clustering of P28 with 2 positive controls (Suppl. Figure 10).

DISCUSSION

Comprehensive genome and epigenome analyses using LRS in twenty-eight individuals with unsolved DD enabled the identification of nine conclusive molecular diagnoses as well as six potential diagnoses, corresponding to diagnostic yields of 32.1% and 21.4% respectively. A subset of the (potential) diagnoses (26.6%) could not have been detected using short reads GS, including both SVs (n=3) and SNVs (n=1). Another three diagnoses were missed in earlier genetic testing due to the limitations inherent to ES. Concurrent methylome analysis enabled the detection of one localized imprinting defect, XCI skewing in four patients and three mothers, as well as three (potential) episignatures.

The improved sequencing contiguity of LRS, which has been reported to enable detection of up to 42% more SVs compared with SRS[12], was essential for the identification of a pathogenic inversion disrupting *DYRK1A*, one of the most frequently affected ID genes[35]. The improved contiguity and mapping also enabled characterization of a CCR of the Xq28 int22h-1/int22h-2 locus, a region known to be prone to recurrent duplications because of the presence of homologous low copy repeats[45]. The penetrance of duplications of this region has recently been shown to be much lower than initially described[37,38]. Our data shows the potentially higher structural complexity of variation in this region, where duplications have mostly been described based on CMA data and prompts future studies using LRS to evaluate the prevalence of more complex rearrangements and its potential contribution to incomplete penetrance and variable expressivity.

Importantly, LRS also enables the detection of previously undetectable SNVs[46,47], as illustrated by the clinically relevant *SYNGAP1* variant identified in this study through enhanced mapping and sequencing quality. Although SNV and indel detection has long been considered a limitation of LRS[10], advances in sequencing chemistry, base-calling algorithms and SNV detection models have resulted in performance comparable to SRS technologies, with the exception of indels in homopolymer and STR regions[48]. Moreover, LRS also allows phasing of variants without reference parental data, exemplified by the variants identified in *TBCK* and *CRELD1.* Phasing is essential for the interpretation of variants affecting

genes associated with AR disorders in single cases and enables applications such as haplotype-based preimplantation genetic testing for *de novo* variants[49].

We further evaluated the added value of LRS by using the new T2T reference genome as compared to hg38. Although T2T reduced the total number of SV calls, it was associated with an increased number of false *de novo* calls, potentially driven by the increased representation of complex genomic regions in the analysis. The use of T2T uncovered two true *de novo* variants affecting newly resolved regions in the T2T assembly and thus undetectable using hg38. Both variants are however located in heterochromatin regions (pericentromeric and acrocentric) that lack clinical annotation, limiting the assessment of their potential clinical relevance. Independent of the reference genome used, the clinical interpretation of non-coding variation remains a major limitation of SV analysis[50]. Despite the existence of databases enabling the evaluation of potential disruptions of topologically associating domains and/or transcription factor binding sites, functional validation of variants in these regions remains difficult, as illustrated by the deletion we detected upstream of *FBXO11.* Ongoing large efforts generating population databases of SVs[51,52] and pangenome references[53] are therefore essential to improve the filtering of common variation. Further development of *in silico* and *in vitro* technologies to improve the prediction of the impact of SVs is also necessary.

The other major breakthrough of long-read sequencing technologies such as Oxford nanopore and PacBio sequencing is the concurrent assessment of native base-modifications[16]. Assessment of XCI skewing and episignatures is currently mainly used as a secondary analysis to facilitate the interpretation of identified variants in patients with DD (reclassification of VUS)[54], and in some instances as primary analysis, to target further genetic analyses. By enabling the integration of genetic and epigenetic variation within a single assay, LRS improves the efficiency of first-tier analyses. Simultaneous assessment of local and genome-wide methylation enhances the interpretation of non-coding SVs. For instance, potential effects of the deletion upstream of *FBXO11* and outlier STRs on promoter methylation were readily excluded. Episignature analysis also enabled to confirm the pathogenicity of the inversion disrupting *DYRK1A* and supported the impact of the loss-of-function variant identified in *SETD5*. In addition, methylation levels overlapping the signature of TBRS syndrome were detected in P7. Interestingly, no potential pathogenic variant was identified in *DNMT3A.* However, we did detect a rare variant in *OSBP*, shared with another patient presenting a similar phenotype. This finding warrants

further investigation into a possible *OSBP*–disease association and the presence of an associated episignature, potentially overlapping with the TBRS signature.

In addition to the ~30-fold increase in assessed CpGs compared with current array-based methylome profiling platforms, LRS offers the advantage of enabling haplotype-aware interpretation. Phasing enables direct confirmation of the inactivation of the X chromosome carrying the structural or other potentially pathogenic variant, whereas this is mostly inferred using conventional XCI testing. It may also provide insights into the mechanisms underlying imprinting disturbances, such as the isolated *MEG8* imprinting defect identified in this study. Methylation analysis of XCI revealed substantially higher variability in CpG methylation along inactivated alleles compared with imprinted alleles. This observation raises questions about the reliability of the classic assays for XCI, targeting only one or a few CpGs. Discrepancies between results from the classical methylation-sensitive restriction enzyme–based assays and gene expression analyses have been reported previously[55,56]. Upon validation, nanopore sequencing–based assessment of all CpGs across multiple differentially methylated promoter regions could therefore improve XCI analyses without requiring gene expression measurements.

While LRS was essential for identifying a subset of diagnoses and will therefore be required to close part of the diagnostic gap, the present study also underscores the continued importance of novel gene–disease associations. Most (5) of the (potential) diagnoses in this cohort became possible because of new insights from the literature or through GeneMatcher exchanges[32]. Two other (potential) diagnoses (7.1%) had been filtered out during SoC genetic testing because the variants were inherited from a seemingly unaffected parent, highlighting the challenges of variable expression and reduced penetrance[7].

Although the present study highlights the added value of LRS for DD, both in resolving part of the diagnostic gap and in potentially replacing a multi-step diagnostic odyssey, it also has limitations. The study design does not address whether variants detected by SoC testing would be missed if LRS were implemented as first-tier, single-test approach. Höps *et al.* addressed this question for PacBio sequencing[11]; however, comparable evaluations using nanopore sequencing are still needed. Moreover, advances in data analysis, such as systematic *de novo* assembly, and extension of analytical approaches, including the assessment of *e.g. de novo* epivariation, were not systematically evaluated and have the potential to further increase diagnostic yield. While our cohort consisted of individuals with

a strong suspicion of monogenetic disease, large studies assessing the diagnostic yield of LRS across different indications are needed to define its added value for the full spectrum of clinical applications.

In summary, our comprehensive reanalysis of unsolved DD cases using LRS demonstrates the clinical utility of this technology for systematic evaluation of both genetic and epigenetic variation. Our analyses revealed novel (potential) diagnoses in 53.6% of cases, illustrating multiple origins of the diagnostic gap in standard genetic testing workflows. LRS was required for 2 of 9 conclusive diagnoses and 2 of 6 potential diagnoses. This corresponds to an added diagnostic yield of 7.1% for (likely) pathogenic variants alone, and 14.3% when including potential diagnoses, highlighting LRS' potential to reduce the diagnostic gap while streamlining patient's diagnostic odyssey. Beyond its clinical value, concomitant and direct sequencing of the genome and methylome at base-pair resolution provides new insights into the molecular mechanisms underlying (epi)genetic variation.

METHODS

Participants inclusion and sequencing

Twenty-eight patients with DD were analyzed as parent–offspring trios. For two probands, an affected sibling was also sequenced, forming quads. Inclusion criteria were (1) moderate to severe developmental delay or ID, (2) mild developmental delay or ID with at least one major CA and/or dysmorphism, (3) multiple CA. CA are considered major when they have a significant impact on the health of the child, whereas rare morphological features with a limited functional or esthetic impact are considered as minor anomalies. Dysmorphism was defined as the presence of at least 3 minor anomalies. Exclusion criteria were the suspicion of an acquired cause, *e.g.* congenital infection and toxic exposure. Participants were included without age-related restrictions. All participants previously underwent standard of care genetic testing through CMA and short-read ES or GS, without achieving a molecular diagnosis. The study was approved by the Ethical Commission of UZ Leuven (S65991) and informed consent was obtained from the legal representatives of all participants. One patient (P13) withdrew during the inclusion process.

Fresh blood was used for HMW DNA extraction from the participants and their parents (Promega Wizard®, Monarch®). For one case available fibroblasts were used for HMW DNA extraction. Libraries

were prepared from 3µg of HMW DNA using the Oxford Nanopore Technologies (ONT) Ligation Sequencing Kit (SQK-LSK110 and SQK-LSK114). With the release of the new SQK-LSK114 chemistry, lower N50 were obtained. Needle shearing, in-house modifications of the LSK114 protocol and adding a short read elimination step were tested and all worked to retrieve mean N50 >30kb. After this transition phase, our standard protocol included a short-read elimination step using the PacBio SRE kit, performed prior to SQK-LSK114 library preparation. Each sample was loaded on a R.9.4.1/R.10.4.1 Flow Cell and reloaded once after 24-48h, for a total sequencing period of 72-96h on a PromethION®.

Data processing

Base-calling was performed using the superaccurate model of Dorado (v0.3.0)[57]. Reads were aligned to hg38 and T2T using minimap2 (v2.24)[58]. SVs were called using Sniffles2 (v2.0.2)[59] and QDNAseq (v1.3.8)[60], STRs were called using TRGT (v4.1.0)[61]. SNV calling was performed with Clair3 (v1.0.4)[62]. Clair3 was also used for the generation of haplotagged BAM files. Methylation was visualized with Methylartist[63] and Integrative Genomics Viewer (IGV)[64]. Episignature detection was performed using NSBEPI[16], using bedMethyl files generated from aligned BAM files using modkit[65] (v0.1.13) .

Genetic variation analysis

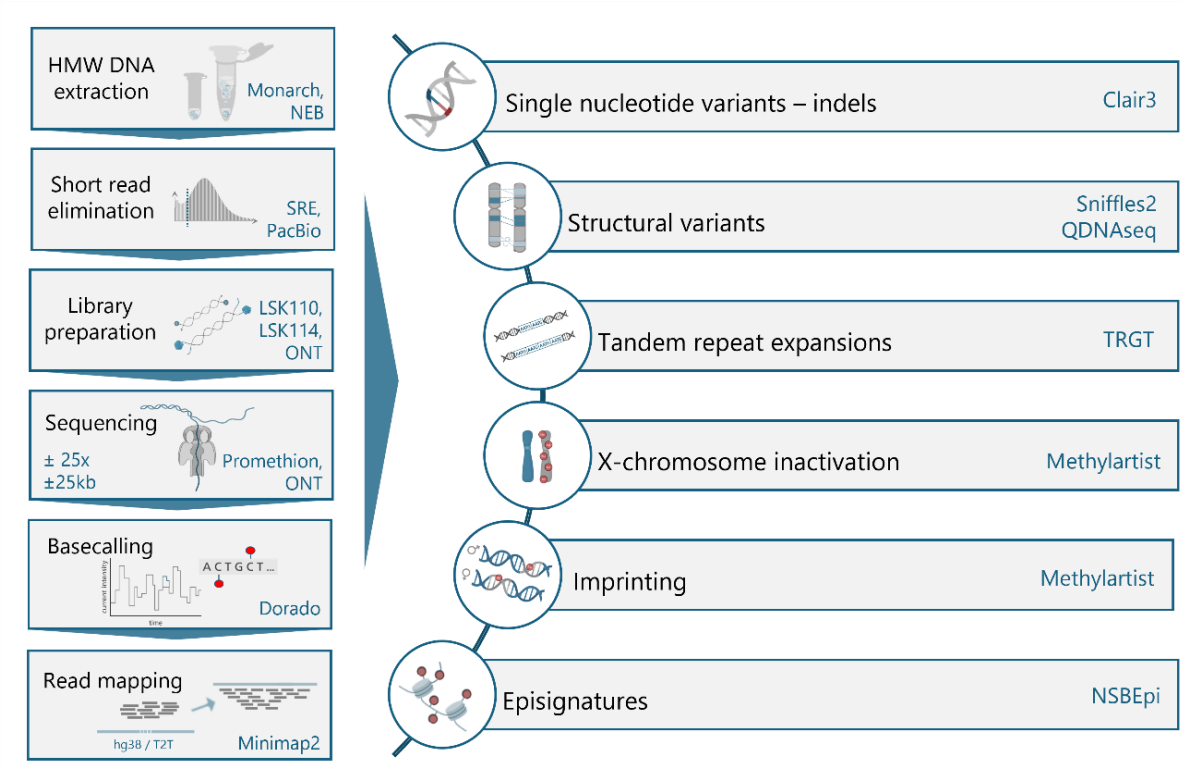
SVs were first assessed according to Mendelian inheritance modes, starting with *de novo* variants, being absent in both parents, and further investigating X-linked variants, inherited from the mother, as well as inherited variants on the autosomes that might lead to autosomal recessive conditions or autosomal dominant conditions with variable penetrance. Additional filtering of SVs was conducted using 50 unaffected parents as reference, excluding all SVs present in any parent for the *de novo* analysis, any father for the X-linked analysis, and any other parent than the parent-of-origin for the analysis on inherited SVs. Both inheritance- and population-based filtering were conducted using Sniffles2. Upon filtering, manual inspection of remaining SVs as well as large CNVs called by QDNAseq was performed using IGV to exclude false calls. Variant classification was performed according to the American College of Medical Genetics and Genomics/Association for Molecular Pathology classification guidelines[66,67]. (Likely) pathogenic variants (classes 4 and 5) that could explain the phenotype of the patient were considered as conclusive diagnoses. VUS with a high suspicion of association to the patient's phenotype were considered as potential diagnoses. The repeat lengths of the STRs that are relevant for the assessed cohort, *i.e.* in *AFF2*, *AFF3*, *DIP2B* and *FMR1* genes, were assessed in all individuals. The

analysis of SNVs was performed using the Geneyx[68] platform, prioritizing variants based on inheritance, HPO-based filtering, population frequency and *in silico* prediction of pathogenicity. The Geneyx platform was also used to perform a HPO-driven reanalysis of SVs.

Epigenetic variation analysis

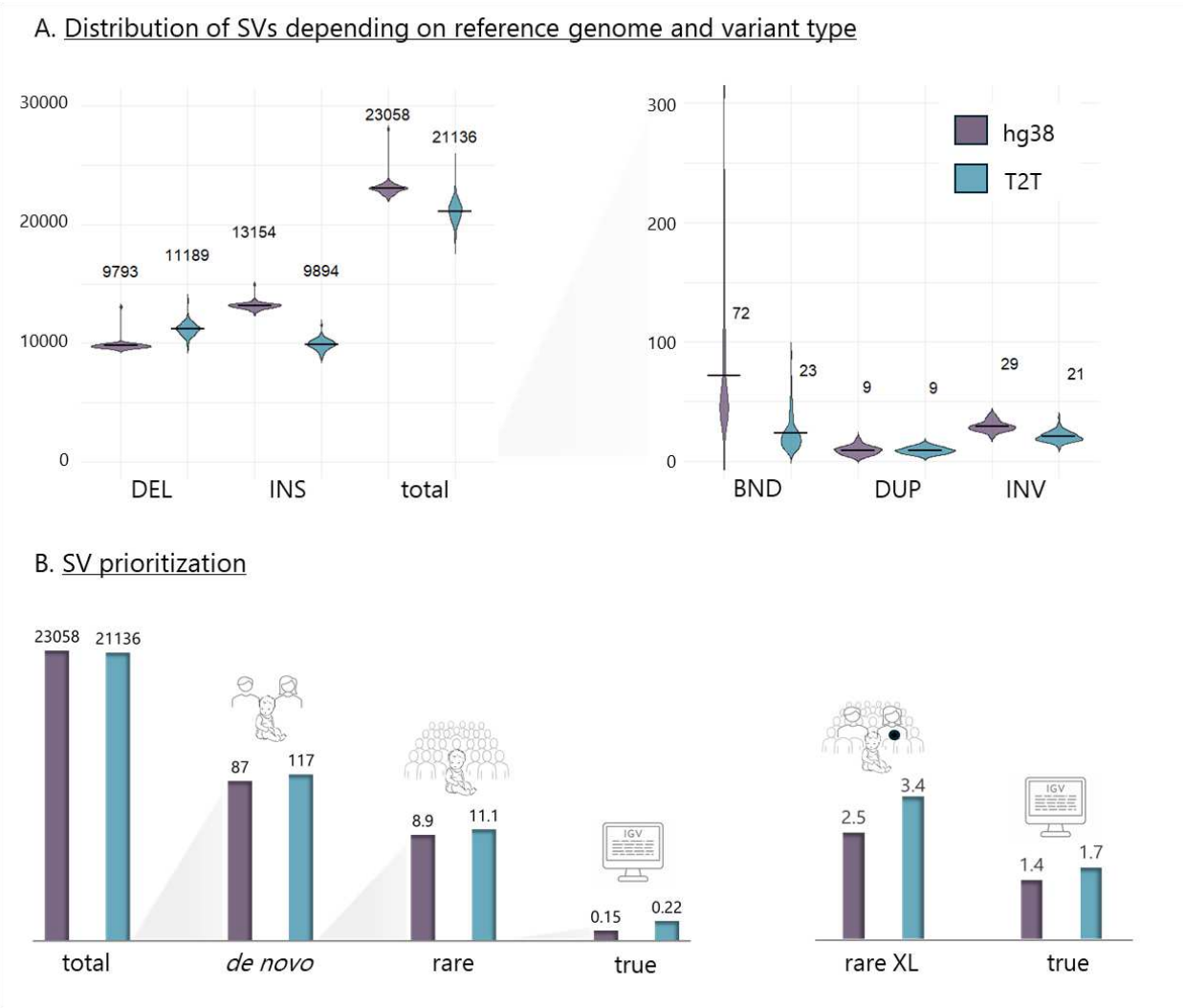
Systematic evaluation of imprinting disorders was performed in all participants as described previously[16]. Skewing of XCI was also systematically evaluated through visualization and quantification of phased methylation levels at the promoter loci of the *AR* and *RP2* genes in all females (probands and mothers). All samples were screened for the presence of an episignature using NSBEpi[16]. Additional evaluation of potential episignatures was performed by UMAP (umap-learn_v0.5.7[69], n_neighbors = 2) for dimensional reduction, and hierarchical clustering (seaborn_v0.12.0[70]).

Figure 1 : Study protocol



The left panel illustrates the wet-lab, (modified)base calling, and read-mapping workflow; the right panel represents the analyses conducted on aligned reads. Applied methods and tools are indicated in blue.

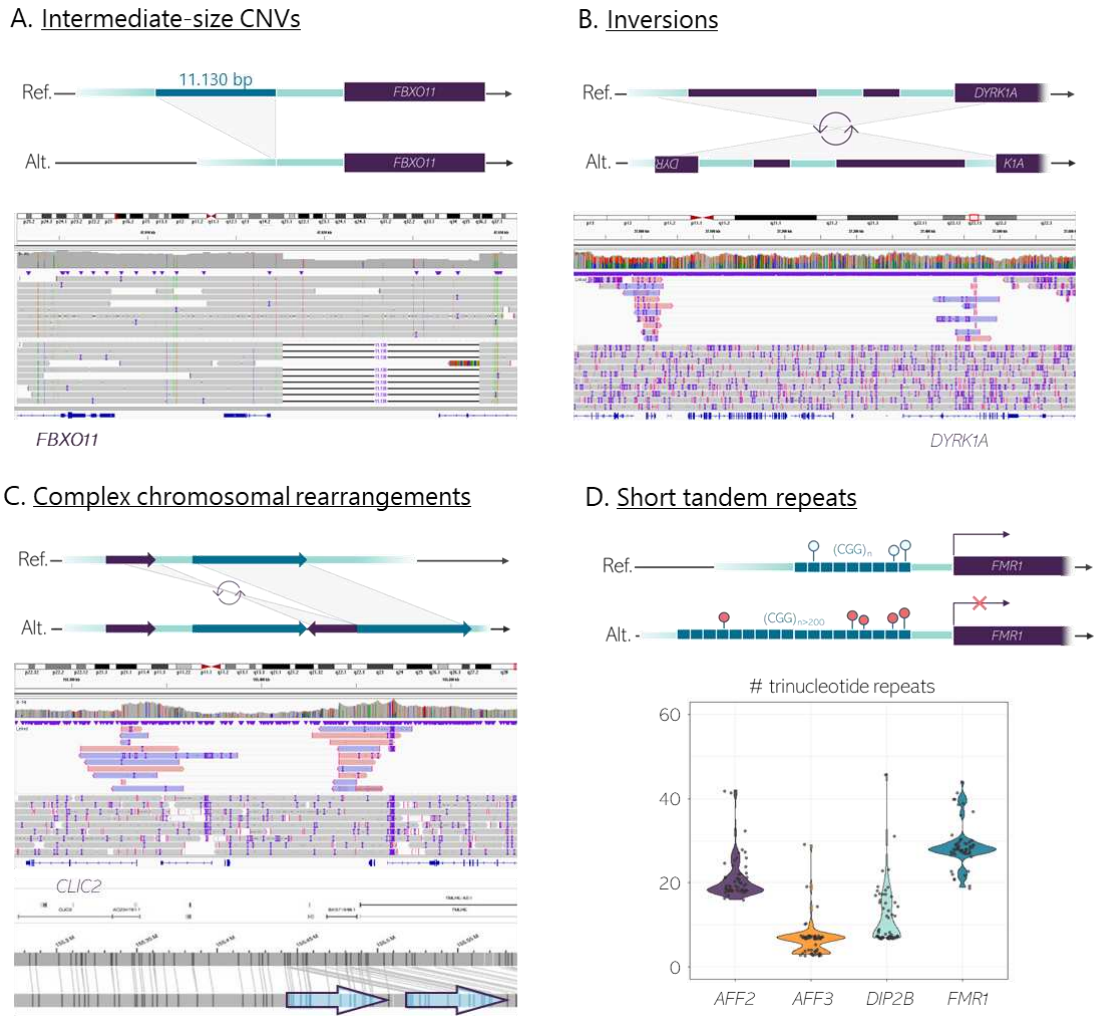
Figure 2 : Structural variation: SV types, reference genomes, inheritance and population filtering



A. Average number of SVs per sample across reference genomes and SV subtypes. Use of the T2T reference is associated with a reduction in the total number of detected SVs, primarily driven by fewer insertions, consistent with its more complete genomic representation. **B.** Impact of inheritance- and population-based filtering on SV counts. *De novo* SVs are variants absent in both parents. Rare *de novo* SVs are absent in both parents and in 50 unrelated healthy parents. “True” *de novo* SVs are those confirmed by manual curation in IGV. Rare-XL variants are maternally inherited SVs located on the X chromosome and absent in a cohort of 25 healthy fathers. True rare-XL variants are those retained after manual IGV curation.

DEL: deletion; INS: insertion, BND: translocations, DUP: duplications, INV : inversions, XL : X-linked.

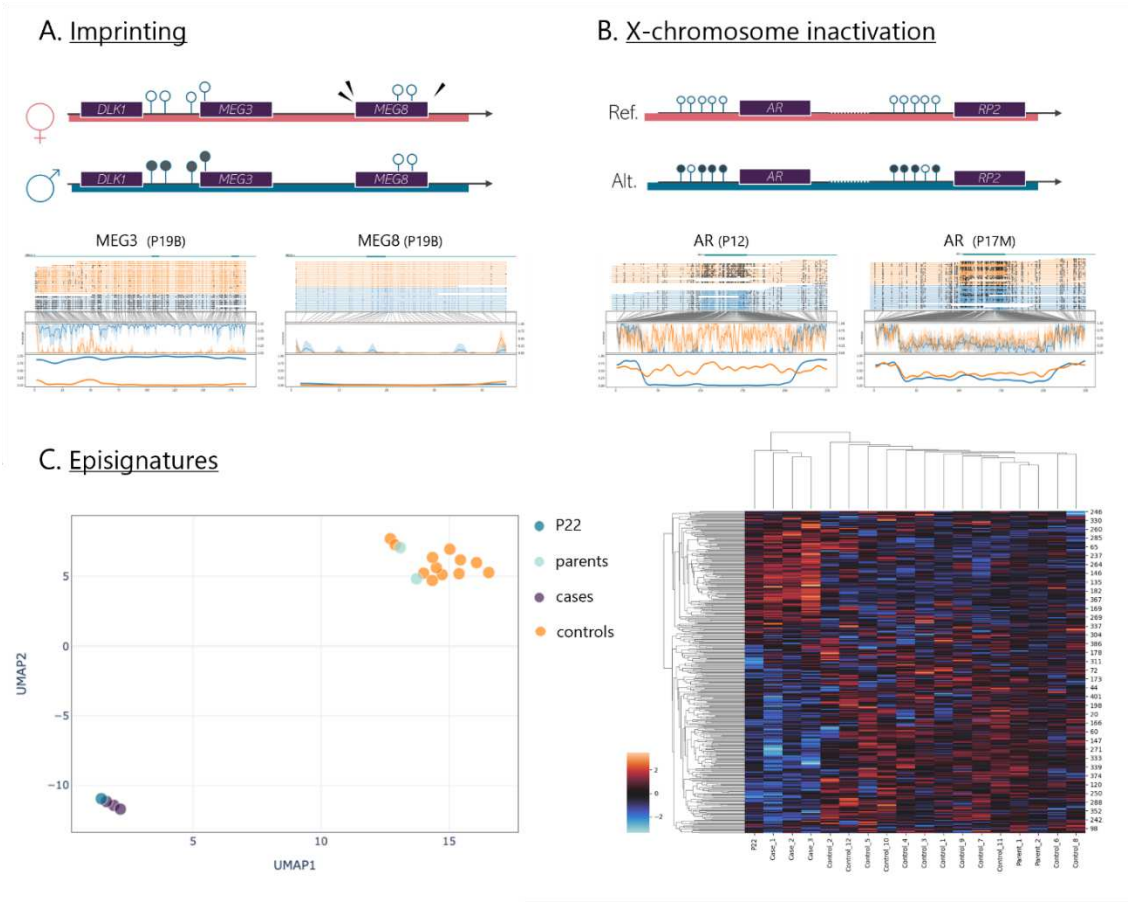
Figure 3 : Structural variants requiring LRS



A.-C. Schematic representation (top) of identified SVs and visualization of the raw data (down sampled to 20x) in IGV (bottom). The result of the optical genome mapping is also shown for the Xq28 CCR (**C**). **D.** Schematic representation of STRs (top) and repeat length for all probands for the *AFF2, AFF3, DIP2B* and *FMR1* genes (bottom).

CCR: complex chromosomal rearrangement; SVs: structural variants; STRs: short tandem repeats.

Figure 4. Epigenetic variation



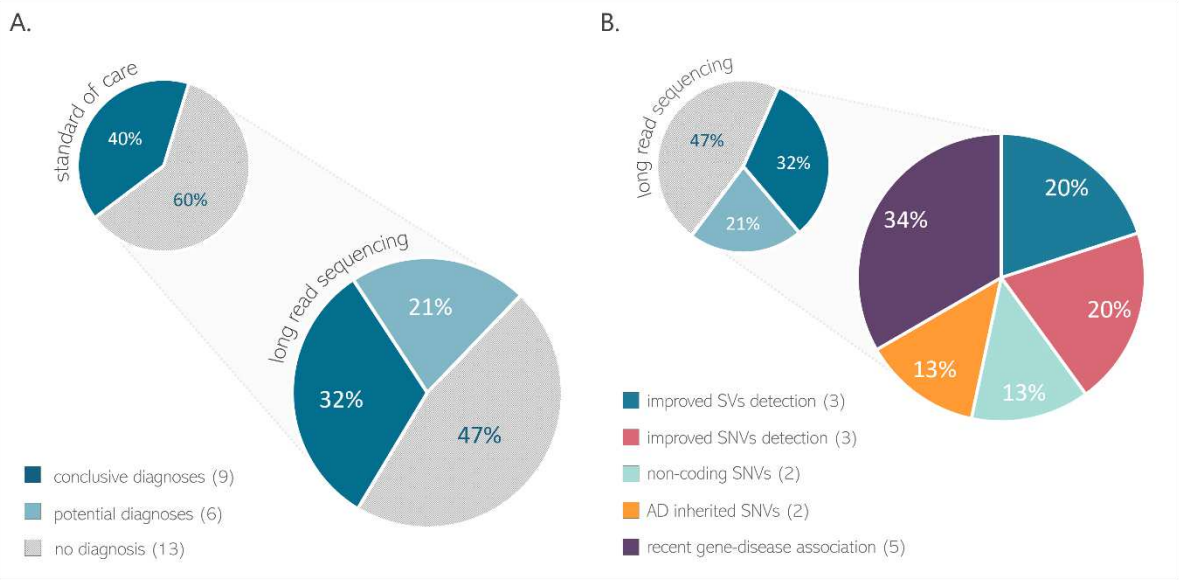
A.-B. Schematic representation (top) of the imprinting disturbance identified in P19B (A) and X-chromosome inactivation (B) as well as visualization of methylation levels using Methylartist (bottom). The upper part of each Methylartist[63] plot represents the reads, with methylated CpGs visualized as black circles. In the lower part of each plot, after translation from genome to CpG-only coordinate space, raw log-likelihood ratios are plotted above a smoothened graph of methylation fraction.

P19B presents a conserved imprinting of the *MEG3* DMR (chr14:100824185-100827640, hg38) but absence of methylation at the *MEG8* DMR (chr14:100904325-100905081, hg38).

P12 and P17M illustrate the occurrence of, respectively, skewed and random XCI at the *AR* locus (chrX:67538000-67552000, hg38). The increased variability in log-likelihood ratio of methylation in skewed XCI compared to imprinting, reflecting the higher variability in CpG methylation of methylated alleles, results in lower overall methylation fractions. **C.** UMAP scatter plot (left), hierarchical clustering and heatmap (right) evaluating methylation levels at *DYRK1A* syndrome episignature loci[35]. Nanopore sequencing data of P22 (dark blue), the two parents of P22 (light blue), three patients with pathogenic

variants in *DYRK1A* (purple) as well as twelve healthy controls (orange) is represented. P22 clusters in close proximity to the other individuals with *DYRK1A* syndrome.

Figure 5 : Diagnostic yield



A. Illustration of the diagnostic yield of the standard of care in the literature[7], and the yield for conclusive and potential diagnoses in present LRS reanalysis of unsolved cases after standard of care testing. **B.** Representation of the factors that enabled the identification of genetic variants in the present reanalysis. The number of diagnoses is provided in brackets, while the corresponding percentages are displayed in the pie charts.

AD: autosomal dominant, SNVs: single nucleotide variants, SVs: structural variants.

Availability of data and material

The datasets generated and analyzed during the current study will be available through the EGA repository (EGAC00001003120).

Ethics approval and consent to participate

This study was approved by the Ethical committee research UZ/KU Leuven (S65304 – S65991) and informed consent was obtained from all participants.

Consent for publication

Informed consent was obtained from all participants for publication.

Competing interests

The authors declare no conflicts of interest.

Funding

This work has been made possible by FWO-TBM grant T-003819N, FWO grant G0A2622N and KU Leuven grants, C1-C14/18/092 and C14/22/125 to JRV. BH was supported by a Collaborative Doctoral Partnership Agreement of the Joint Research Center F.7/KU LEUVEN B&G 35332.

Authors' contributions

Conceptualization: MG, BM, JRV, KVDB; Data curation: MG, BH, ES; Formal analysis: MG, BH, ES, GMc; Funding acquisition: KVDB, JRV; Investigation: MG, GP, NO, SM; Methodology: MG, BH; Project administration: MG, JRV, KVDB; Resources: JB, KDV, HP, GM, HVE, KDVB; Software: BH, ES; Supervision: JRV, KVDB; Validation: MG, BH, ES; Visualization: MG, BH, ES ; Writing – original draft: MG, BH ; Writing – review and editing: MG, BH, ES, NO, SM, GMc, GP, JB, KDV, HP, BM, GM, HVE, JRV, KVDB.

Acknowledgements

The authors thank the patients and their families who participated in this research.

REFERENCES

1. Vissers LELM, Gilissen C, Veltman JA. Genetic studies in intellectual disability and related disorders. *Nat Rev Genet*. *Nature Publishing Group*. 2016;17(1):9-18. doi:10.1038/nrg3999

2. Gilissen C, Hehir-Kwa JY, Tjwan Thung D, et al. Genome sequencing identifies major causes of severe intellectual disability. *Nature*. 2014;511:344-348. doi:10.1038/nature13394

3. Wright CF, Campbell P, Eberhardt RY, et al. Genomic Diagnosis of Rare Pediatric Disease in the United Kingdom and Ireland. *New England Journal of Medicine*. 2023;388(17):1559-1571. doi:10.1056/nejmoa2209046

4. Hayeems RZ, Ungar WJ, Marshall CR, et al. Comparing the performance of exome and genome sequencing for rare disease diagnostics: A randomized implementation effectiveness trial. *Genetics in Medicine*. Published online October 2025:101605. doi:10.1016/j.gim.2025.101605

5. Hamanaka K, Fujita A, Miyatake S, et al. Genome sequencing provides high diagnostic yield and new etiological insights for intellectual disability and developmental delay. *NPJ Genom Med*. 2025;10(1). doi:10.1038/s41525-025-00521-4

6. Wojcik MH, Lemire G, Berger E, et al. Genome Sequencing for Diagnosing Rare Diseases. *New England Journal of Medicine*. 2024;390(21):1985-1997. doi:10.1056/nejmoa2314761

7. Geysens M, Souche E, Baroni MC, et al. A nationwide prospective randomized trial for diagnosing developmental disorders demonstrates genome sequencing outperforms standard of care. Preprint posted online October 23, 2025. doi:10.1101/2025.10.18.25328208

8. Seaby EG, Ennis S. Challenges in the diagnosis and discovery of rare genetic disorders using contemporary sequencing technologies. *Brief Funct Genomics*. 2020;19(4):243-258. doi:10.1093/bfgp/elaa009

9. Eichler EE. Genetic Variation, Comparative Genomics, and the Diagnosis of Disease. *New England Journal of Medicine*. 2019;381(1):64-74. doi:10.1056/NEJMra1809315

10. Logsdon GA, Vollger MR, Eichler EE. Long-read human genome sequencing and its applications. *Nature Reviews*. 2020;21:597-614. doi:10.1038/s41576-020-0236-x

11. Höps W, Weiss MM, Derks R, et al. HiFi long-read genomes for difficult-to-detect, clinically relevant variants. *Am J Hum Genet*. 2025;112(2):450-456. doi:10.1016/j.ajhg.2024.12.013

12. Ebert P, Audano PA, Zhu Q, et al. Haplotype-resolved diverse human genomes and integrated analysis of structural variation. *Science (1979)*. 2021;372(6537). doi:10.1126/science.abf7117

13. Mostovoy Y, Boone PM, Huang Y, et al. Resolution of ring chromosomes, Robertsonian translocations, and complex structural variants from long-read sequencing and telomere-to-telomere assembly. *Am J Hum Genet*. 2024;111(12):2693-2706. doi:10.1016/j.ajhg.2024.10.006

14. Akbari V, Garant JM, O'neill K, et al. Genome-wide detection of imprinted differentially methylated regions using nanopore sequencing. *Elife*. 2022;5(11). doi:10.7554/eLife

15. Johansson J, Lidéus S, Höijer I, et al. A novel quantitative targeted analysis of X-chromosome inactivation (XCI) using nanopore sequencing. *Sci Rep*. 2023;13(1):12856. doi:10.1038/s41598-023-34413-3

16. Geysens M, Huremagic B, Souche E, et al. Clinical evaluation of long-read sequencing-based episignature detection in developmental disorders. *Genome Med*. 2025;17(1):1. doi:10.1186/s13073-024-01419-z

17. Nurk S, Koren S, Rhie A, et al. The complete sequence of a human genome. *Science (1979)*. 2022;376:44-53. https://www.science.org

18. Aganezov S, Yan SM, Soto DC, et al. A complete reference genome improves analysis of human genetic variation. *Science (1979)*. 2022;376(6588). doi:10.1126/science.abl3533

19. Miller DE, Sulovari A, Wang T, et al. Targeted long-read sequencing identifies missing disease-causing variation. *The American Journal of Human Genetics*. 2021;108(8):1436-1449. doi:10.1016/j.ajhg.2021.06.006

20. Hiatt SM, Lawlor JMJ, Handley LH, et al. Long-read genome sequencing for the molecular diagnosis of neurodevelopmental disorders. *Human Genetics and Genomics Advances*. 2021;2(2). doi:10.1016/j.xhgg.2021.100023

21. Pauper M, Kucuk E, Wenger AM, et al. Long-read trio sequencing of individuals with unsolved intellectual disability. *European Journal of Human Genetics*. 2021;29(4):637-648. doi:10.1038/s41431-020-00770-0

22. Kucuk E, van der Sanden BPGH, O'Gorman L, et al. Comprehensive de novo mutation discovery with HiFi long-read sequencing. *Genome Med*. 2023;15(1). doi:10.1186/s13073-023-01183-6

23. Steyaert W, Sagath L, Demidov G, et al. Unravelling undiagnosed rare disease cases by HiFi long-read genome sequencing. *Marwa Saeed Abd Elmaksoud*. 9. doi:10.1101/2024.05.03.24305331

24. Negi S, Stenton SL, Berger SI, et al. Advancing long-read nanopore genome assembly and accurate variant calling for rare disease detection. *Am J Hum Genet*. 2025;112(2):428-449. doi:10.1016/j.ajhg.2025.01.002

25. Sinha S, Rabea F, Ramaswamy S, et al. Long read sequencing enhances pathogenic and novel variation discovery in patients with rare diseases. *Nature Communications* . 2025;16(1). doi:10.1038/s41467-025-57695-9

26. Bladen CL, Salgado D, Monges S, et al. The TREAT-NMD DMD global database: Analysis of more than 7,000 duchenne muscular dystrophy mutations. *Hum Mutat*. 2015;36(4):395-402. doi:10.1002/humu.22758

27. Yuan D, Cai R, Mao A, et al. Improved Genetic Characterization of Congenital Adrenal Hyperplasia by Long-Read Sequencing Compared with Multiplex Ligation-Dependent Probe Amplification Plus Sanger Sequencing. *Journal of Molecular Diagnostics*. 2024;26(9):770-780. doi:10.1016/j.jmoldx.2024.05.009

28. Ek M, Kvarnung M, Ten Berk de Boer E, et al. Long-read genome sequencing enhances diagnostics of pediatric neurological disorders. *Genome Med*. Published online January 9, 2026. doi:10.1186/s13073-025-01596-5

29. Chen Y, Dawes R, Kim HC, et al. De novo variants in the RNU4-2 snRNA cause a frequent neurodevelopmental syndrome. *Nature*. 2024;632(8026):832-840. doi:10.1038/s41586-024-07773-7

30. Mo A, Paz-Ebstein E, Yanovsky-Dagan S, et al. A recurrent de novo variant in NUSAP1 escapes nonsense-mediated decay and leads to microcephaly, epilepsy, and developmental delay. *Clin Genet*. 2023;104(1):73-80. doi:10.1111/cge.14335

31. Jeffries L, Mis EK, McWalter K, et al. Biallelic CRELD1 variants cause a multisystem syndrome, including neurodevelopmental phenotypes, cardiac dysrhythmias, and frequent infections. *Genetics in Medicine*. 2024;26(2). doi:10.1016/j.gim.2023.101023

32. Sobreira N, Schiettecatte F, Valle D, Hamosh A. GeneMatcher: A Matching Tool for Connecting Investigators with an Interest in the Same Gene. *Hum Mutat*. 2015;36(10):928-930. doi:10.1002/humu.22844

33. Jansen S, van der Werf IM, Innes AM, et al. De novo variants in FBXO11 cause a syndromic form of intellectual disability with behavioral problems and dysmorphisms. *European Journal of Human Genetics*. 2019;27(5):738-746. doi:10.1038/s41431-018-0292-2

34. Spielmann M, Mundlos S. Structural variations, the regulatory landscape of the genome and their alteration in human disease. *BioEssays*. 2013;35(6):533-543. doi:10.1002/bies.201200178

35. Courraud J, Chater-Diehl E, Durand B, et al. Integrative approach to interpret DYRK1A variants, leading to a frequent neurodevelopmental disorder. *Genetics in Medicine*. 2021;23(11):2150-2159. doi:10.1038/s41436-021-01263-1

36. Andersen EF, Baldwin EE, Ellingwood S, Smith R, Lamb AN. Xq28 duplication overlapping the int22h-1/int22h-2 region and including RAB39B and CLIC2 in a family with intellectual and developmental disability. *Am J Med Genet A*. 2014;164(7):1795-1801. doi:10.1002/ajmg.a.36524

37. Billes A, Pujalte M, Jedraszak G, et al. Possible incomplete penetrance of Xq28 int22h-1/int22h-2 duplication. *Clin Genet*. Published online September 1, 2024. doi:10.1111/cge.14525

38. Levy M, Elron E, Shohat M, et al. Exploring inheritance, and clinical penetrance of distal Xq28 duplication syndrome: insights from 47 new unpublished cases. *J Hum Genet*. 2024;69(7):337-343. doi:10.1038/s10038-024-01252-7

39. Prasasya R, Grotheer K V., Siracusa LD, Bartolomei MS. Temple syndrome and Kagami-Ogata syndrome: Clinical presentations, genotypes, models and mechanisms. *Hum Mol Genet*. *Oxford University Press*. 2020;29:R107-R116. doi:10.1093/hmg/ddaa133

40. D'Angelo E, Pignata L, Cecere F, et al. Hypomethylation of the MEG8:Int2-DMR in patients with pathogenic PLAG1 variants suggests new role of the chr14q32 imprinting cluster in Silver-Russell syndrome. *Clin Epigenetics*. 2026;18(1). doi:10.1186/s13148-025-02024-6

41. Aref-Eshghi E, Kerkhof J, Pedro VP, et al. Evaluation of DNA Methylation Episignatures for Diagnosis and Phenotype Correlations in 42 Mendelian Neurodevelopmental Disorders. *The American Journal of Human Genetics*. 2020;106(3):356-370. doi:10.1016/j.ajhg.2020.01.019

42. Gehin C, Lone MA, Lee W, et al. CERT1 mutations perturb human development by disrupting sphingolipid homeostasis. *Journal of Clinical Investigation*. 2023;133(10). doi:10.1172/JCI165019

43. De Sainte Agathe JM, Pode-Shakked B, Naudion S, et al. ARF1 -related disorder: Phenotypic and molecular spectrum. *J Med Genet*. 2023;60(10):999-1005. doi:10.1136/jmg-2022-108803

44. Awamleh Z, Choufani S, Cytrynbaum C, et al. ANKRD11 pathogenic variants and 16q24.3 microdeletions share an altered DNA methylation signature in patients with KBG syndrome. *Hum Mol Genet*. 2023;32(9):1429-1438. doi:10.1093/hmg/ddac289

45. Ballout RA, Dickerson C, Wick MJ, et al. Int22h1/Int22h2-mediated Xq28 duplication syndrome: de novo duplications, prenatal diagnoses, and additional phenotypic features. *Hum Mutat*. 2020;41(7):1238-1249. doi:10.1002/humu.24009

46. Noyes MD, Harvey WT, Porubsky D, et al. Familial long-read sequencing increases yield of de novo mutations. *Am J Hum Genet*. 2022;109(4):631-646. doi:10.1016/j.ajhg.2022.02.014

47. Porubsky D, Dashnow H, Sasani TA, et al. Human de novo mutation rates from a four-generation pedigree reference. *Nature*. 2025;643(8071):427-436. doi:10.1038/s41586-025-08922-2

48. Kolmogorov M, Billingsley KJ, Mastoras M, et al. Scalable Nanopore sequencing of human genomes provides a comprehensive view of haplotype-resolved variation and methylation. *Nat Methods*. 2023;20(10):1483-1492. doi:10.1038/s41592-023-01993-x

49. Zhao Y, Tsuiko O, Jatsenko T, et al. Long-read whole-genome sequencing-based concurrent haplotyping and aneuploidy profiling of single cells. *Nucleic Acids Res*. 2025;53(6). doi:10.1093/nar/gkaf247

50. Ellingford JM, Ahn JW, Bagnall RD, et al. Recommendations for clinical interpretation of variants found in non-coding regions of the genome. *Genome Med*. 2022;14(1). doi:10.1186/s13073-022-01073-3

51. Gustafson JA, Gibson SB, Damaraju N, et al. High-coverage nanopore sequencing of samples from the 1000 Genomes Project to build a comprehensive catalog of human genetic variation. *Genome Res*. 2024;34(11):2061-2073. doi:10.1101/gr.279273.124

52. Garimella K V, Li Q, Wertz J, et al. Population-scale Long-read Sequencing in the All of Us Research Program. Preprint posted online October 5, 2025. doi:10.1101/2025.10.02.25336942

53. Sui Y, Lin J, Noyes MD, et al. Pangenome discovery of missing autism variants. Preprint posted online July 22, 2025. doi:10.1101/2025.07.21.25331932

54. Smits DJ, Debuy C, Brooks AS, et al. Clinical utility of DNA-methylation signatures in routine diagnostics for neurodevelopmental disorders. *European Journal of Human Genetics*. 2025;33(10):1281-1289. doi:10.1038/s41431-025-01919-5

55. Swierczek SI, Piterkova L, Jelinek J, et al. Methylation of AR locus does not always reflect X chromosome inactivation state. *Blood*. 2012;119(13):100-109. doi:10.1182/blood-2011-11

56. Fadra N, Schultz-Rogers LE, Chanana P, et al. Identification of skewed X chromosome inactivation using exome and transcriptome sequencing in patients with suspected rare genetic disease. *BMC Genomics*. 2024;25(1). doi:10.1186/s12864-024-10240-2

57. https://github.com/nanoporetech/dorado. Oxford Nanopore Technologies - Dorado.

58. Li H. Minimap2: Pairwise alignment for nucleotide sequences. *Bioinformatics*. 2018;34(18):3094-3100. doi:10.1093/bioinformatics/bty191

59. Sedlazeck FJ, Rescheneder P, Smolka M, et al. Accurate detection of complex structural variations using single-molecule sequencing. *Nat Methods*. 2018;15(6):461-468. doi:10.1038/s41592-018-0001-7

60. Scheinin I, Sie D, Bengtsson H, et al. DNA copy number analysis of fresh and formalin-fixed specimens by shallow whole-genome sequencing with identification and exclusion of problematic regions in the genome assembly. *Genome Res*. 2022;24. doi:10.1101/gr.175141.114

61. Dolzhenko E, English A, Dashnow H, et al. Characterization and visualization of tandem repeats at genome scale. *Nat Biotechnol*. 2024;42(10):1606-1614. doi:10.1038/s41587-023-02057-3

62. Zheng Zhenxian, Li Shumin, Su Junhao, Leung Amy Wing-Sze, Lam Tak-Wah, Luo Ruibang. Symphonizing pileup and full-alignment for deep learning-based long-read variant calling. *Nat Comput Sci*. 2022;2:797-803. doi:10.1038/s43588-022-00387-x

63. Cheetham SW, Kindlova M, Ewing AD. Methylartist: tools for visualizing modified bases from nanopore sequence data. *Bioinformatics*. 2022;38(11):3109-3112. doi:10.1093/bioinformatics/btac292

64. Robinson JT, Thorvaldsdóttir H, Winckler W, et al. Integrative genomics viewer. *Nature Biotechnology 2011 29:1*. 2011;29(1):24-26. doi:10.1038/nbt.1754

65. https://github.com/nanoporetech/modkit. Oxford Nanopore Technologies - Modkit.

66. 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. *Genetics in Medicine*. 2015;17(5):405-424. doi:10.1038/gim.2015.30

67. Riggs ER, Andersen EF, Cherry AM, 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). Published online 2020. doi:10.1038/s41436

68. Dahary D, Golan Y, Mazor Y, et al. Genome analysis and knowledge-driven variant interpretation with TGex. *BMC Med Genomics*. 2019;12(1). doi:10.1186/s12920-019-0647-8

69. McInnes L, Healy J, Saul N, Großberger L. UMAP: Uniform Manifold Approximation and Projection. *J Open Source Softw*. 2018;3(29):861. doi:10.21105/joss.00861

70. Waskom M. Seaborn: statistical data visualization. *J Open Source Softw*. 2021;6(60):3021. doi:10.21105/joss.03021

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- SupplementaryTable1.xlsx
- SupplementaryFigures.docx