Sensitivity of HiFi long-read genome sequencing for difficult-to-detect pathogenic variants when applied to real-world clinical laboratory samples

Authors

Joseph M. Devaney, Jessica X. Chong, Patricia C. Lopes, ..., Flavia M. Facio, Michael J. Bamshad, Britt Johnson

Correspondence

britt.johnson@genedx.com

We applied HiFi long-read sequencing to 191 real-world clinical samples, which included buccal and low-molecular-weight DNA, and were enriched for difficult-to-detect variants. Overall, 99.6% (479/481) of variants were automatically detected with HiFi, showcasing its potential as a single approach for the diagnosis of rare disease.

Devaney et al., 2026, The American Journal of Human Genetics *113*, 1036–1048
May 7, 2026
https://doi.org/10.1016/j.ajhg.2026.04.001

CellPress

ARTICLE

Sensitivity of HiFi long-read genome sequencing for difficult-to-detect pathogenic variants when applied to real-world clinical laboratory samples

Joseph M. Devaney,[1] Jessica X. Chong,[2,3] Patricia C. Lopes,[1] Jessica Noya,[1] April S. Berlyoung,[1] Shamila Yusuff,[1] Solomon Lynch,[1] Rhonda Brandon,[1] Kathleen S. Hruska,[1] Lucas Lochovsky,[1] Julianna Spangler,[1] Kirsty McWalter,[1] Keith Nykamp,[1] Sarah R. Poll,[1] Andrew B. Stergachis,[3,4,5] John Greally,[6] Paul Kruszka,[1,7] Egor Dolzhenko,[8] Xiao Chen,[8] Alexander V. Robertson,[8] William J. Rowell,[8] Juniper A. Lake,[8] Andrew Carroll,[9] Robert Kueffner,[1] Michael A. Eberle,[8] Flavia M. Facio,[1] Michael J. Bamshad,[2,3] and Britt Johnson[1,*]

Summary

Leveraging new sequencing and omic technologies to enhance the detection of pathogenic variants in known disease genes is a key step toward increasing the likelihood of a precise genetic diagnosis for affected individuals. Short-read sequencing is widely used in clinical laboratories for multi-gene panels and exome and genome sequencing, but this technology has inherent limitations in detecting certain classes of genetic variation. As a result, diagnostic laboratories continue to offer complementary assays, often used sequentially, reducing efficiency and speed in providing a diagnosis. We applied PacBio long-read genome sequencing (HiFi) to samples from 191 probands previously tested with short-read sequencing alone and/or other diagnostic technologies and enriched for pathogenic variants difficult to detect (VDDs). HiFi's pipeline automatically detected 479 of 481 (99.6%) disease-causing variants, many of which were called in samples not optimized for long-read genome sequencing (such as buccal samples or low-molecular-weight DNA). The two variants not automatically detected were a mosaic trisomy 18 (23% mosaicism) and a 5,594-bp mosaic deletion (13% mosaicism). However, other mosaic variants were detected, indicating that HiFi at ~30× genome coverage is sensitive to the degree of mosaicism. Of 481 variants, 49 were suspected based on the clinical report but not confirmed molecularly prior to HiFi. Our findings demonstrate that HiFi sequencing detects a wide range of VDDs in real-world clinical laboratory samples, highlighting a key advantage of HiFi as a potential first-tier test over the myriad of complementary technologies currently used to detect VDDs.

Introduction

Advances in sequencing technologies have substantially improved our ability to detect pathogenic variants underlying rare genetic conditions. Short-read sequencing (SRS) has become a widely used approach in clinical diagnostics for multi-gene panels, exome sequencing (ES), and genome sequencing (GS). Short-read ES and GS are effective first-tier tests that yield diagnostic rates of ~18%–43% depending on factors such as age and indication for testing.[1] Nevertheless, SRS has well-known limitations for detecting certain classes of variants (i.e., variants difficult to detect [VDDs]), including structural variants (SVs) and variants involving tandem repeats, homopolymers, and regions of high pseudogene homology.[2,3] As a result, diagnostic laboratories continue to offer complementary assays to detect VDDs, including multiplex ligation-dependent probe amplification (MLPA), triplet-primed PCR (TPP), Sanger-based assays, fluorescence *in situ* hybridization (FISH), methods to detect DNA methylation, and karyotyping. The need for complementary assays is estimated to be up to 25% for all individuals referred for clinical testing, given the technical limitations of SRS.[4] The inability to detect VDDs using SRS alone extends the diagnostic odyssey for individuals harboring these variant types, resulting in delays or missed opportunities for therapeutic intervention.

Long-read GS (lrGS) offers improved performance in identifying VDDs, particularly in challenging genomic regions (e.g., Wenger et al.,[5] Cohen et al.,[6] AlAbdi et al.,[7] Hiatt et al.,[8] and Höps et al.[9]). The two leading lrGS methods are those of Pacific Biosciences (HiFi[5,10]) and Oxford Nanopore Technologies (ONT[11]). The value and superior sensitivity of HiFi and ONT lrGS over SRS for many VDDs have been illustrated in population-wide studies (e.g., Beyter et al.,[12] Mahmoud et al.,[13] and Gustafson et al.[14]), but these studies were performed on general population cohorts rather than on individuals with a clinical indication for testing.

Studies reporting the application of lrGS to individual subjects and to undiagnosed disease cohorts, following negative SRS, support a modest increase in diagnostic

[1]GeneDx LLC, Gaithersburg, MD 20877, USA; [2]Department of Pediatrics, University of Washington, Seattle, WA 98195, USA; [3]Brotman Baty Institute for Precision Medicine, Seattle, WA 98195, USA; [4]Department of Genome Sciences, University of Washington, Seattle, WA 98195, USA; [5]Division of Medical Genetics, Department of Medicine, University of Washington, Seattle, WA 98195, USA; [6]Department of Pediatrics, Division of Pediatric Genetic Medicine, Children's Hospital at Montefiore/Montefiore Medical Center/Albert Einstein College of Medicine, Bronx, NY 10467, USA; [7]Department of Pediatrics, Division of Pediatric Genetics, University of Virginia, Charlottesville, VA 22903, USA; [8]Pacific Biosciences, Menlo Park, CA 94025, USA; [9]Google LLC, Mountain View, CA 94043, USA
*Correspondence: britt.johnson@genedx.com
https://doi.org/10.1016/j.ajhg.2026.04.001.

yield by lrGS of rare Mendelian conditions over other diagnostic methods in the context of rare diseases (reviewed in Del Gobbo and Boycott[15]). Yet such studies, even collectively, have not assessed the full spectrum of pathogenic variants encountered in real-world clinical diagnostic settings. Additionally, while high-quality DNA of high molecular weight (HMW; the kind obtained from fresh blood using a HMW extraction method) is the ideal sample source for lrGS, a portion of samples currently received for clinical testing come from other sources, such as buccal swabs or banked DNA samples previously extracted by conventional methods not optimized to obtain HMW DNA. Accordingly, it is important to understand how lrGS performs when applied to these varied real-world clinical laboratory samples.

We applied HiFi GS to DNA samples from individuals with genetic conditions explained by pathogenic VDDs previously identified by SRS alone or in combination with one or more complementary diagnostic technologies. These variants represent the diversity of VDDs typically assessed in clinical laboratories, and the samples used reflect the variety of sample types found in a clinical laboratory, including blood and buccal samples, as well as DNA processed in various ways. By evaluating the performance of HiFi sequencing across this range of variant types and sample types, we assessed its sensitivity to detect VDDs and identified gaps and challenges that need to be addressed before lrGS can be considered as a potential first-line diagnostic tool.

Material and methods

Selection of samples with VDDs

We generated a list of genes/regions that tend to harbor VDDs by surveying loci in our clinical laboratory that, subsequent to SRS, required specialized assays (beyond standard SRS) to assess pathogenicity. The assays included MLPA, methylation-specific MLPA (msMLPA), gel/Sanger-based assays, TPP assays, methylation TPP, chromosomal microarray (CMA), karyotyping, FISH, optical genome mapping, Southern Blot, qPCR, and SRS of PCR-amplified mtDNA. A total of 27 different types of variants were identified. We also identified 27 affected individuals for whom the laboratory report contained variants difficult to map by SRS in known genes for Mendelian conditions.[5,16] Of these, three (*CFC1* [MIM: 605194], *SMN1* [MIM: 600354], and *SMN2* [MIM: 601627]) were considered not mappable in a previous study that employed HiFi[5] at the time the samples were selected. Samples were from individuals clinically tested at GeneDx ($n = 188$) or from individuals who had consented to participate in the University of Washington Repository for Mendelian Disorders ($n = 3$). In total, we identified 191 clinical samples (each from a separate individual) with previously detected variants that were VDDs, for which we retained sufficient residual sample (blood or buccal) or DNA for HiFi sequencing. These samples harbored 481 variants (Tables 1 and S1) that had been previously reported or suspected based on clinical findings. De-identified samples from individuals tested at GeneDx were used for the purposes of the test development and validation discussed in this publication, in accordance with the informed consent obtained from each such individual at the time of clinical testing. The University of Washington Repository for Mendelian Disorders was reviewed and approved by the University of Washington's institutional review board (IRB; study ID #STUDY00008810).

Samples used for assessing the accuracy of HiFi sequencing (expected vs. observed variants)

The VDDs assessed in this study are summarized in Table 1 and listed in detail in Table S1. Throughout the manuscript, the term VDD is used as an umbrella term encompassing events, including phases, that are not strictly variants but may be relevant to disease pathogenesis. The original diagnostic methods used per sample are provided in Table S2. The original diagnostic methods used per variant type are provided in Table S3.

Samples used for assessing the effect of sample source on HiFi sequencing quality metrics

The extraction chemistries and sample sources for each of the 191 samples (129 blood, 44 buccal, and 18 extracted DNA) used in this study can be found in Table S4. For 18 of the samples submitted to our lab as DNA (external DNA), the extraction chemistry is unknown. Of the 191 samples, the DNA for 58/129 blood samples and 39/44 buccal samples was extracted using the same kits (Applied Biosystems MagMax DNA Multi-Sample Ultra 2.0, catalog number A36570, or Qiagen QIASymphony DSP DNA Midi, catalog number 937255). We used these samples to assess the effect of sample sources on sequencing quality metrics.

Samples used for assessing the effect of DNA extraction chemistry on HiFi sequencing quality metrics

We used 129 of the 191 samples for which DNA was extracted from blood in our laboratory to test the effect of extraction chemistry on sequencing quality metrics. No buccal samples were used for the purposes of this analysis, as only 5 buccal samples would have been eligible for inclusion in the HMW group, and they could have introduced a confounding effect. For this analysis, we compared the results from samples extracted using kits designed to obtain HMW DNA with kits that were not designed for this purpose (henceforth, non-HMW). The following four HMW kits were used: New England Biolabs Monarch HMW kit (catalog number T3010), PacBio Nanobind CBB kit (catalog number 102-301-900), PacBio Nanobind HT CBB kit (catalog number 102-762-700), and Zymo Quick-DNA HMW MagBead Kit (catalog number D6060). All others were grouped into the non-HMW category. There were 61 samples in the HMW group and 68 in the non-HMW group.

HiFi lrGS (HiFi sequencing)

Genomic DNA was size selected for lrGS using either PacBio SRE (catalog number 102-208-300) or PacBio SRE XS (catalog number 102-208-200), based on input DNA quality, followed by shearing to a target size of 15–18 kb using either the Hamilton Microlab STAR (Hamilton, Franklin, MA) or Megaruptor 3 (Diagenode, Liège, Belgium) system. Libraries were prepared with the PacBio SMRTbell prep kit 3.0 (catalog number 102-141-700) and sequenced on a single SMRT Cell on a PacBio Revio system with 24-h movie times (targeted at generating 30× coverage). SMRT Link v.12.0–v.13.3 were used, along with corresponding updates to the Instrument Control Software (ICS), as instrument updates occurred throughout the study. Run settings were kept consistent

Table 1. Summary of expected and observed events for samples from the 191 individuals in the cohort

Event type by HiFi	Expected (*n*)	Observed (*n*)	Concordance (%)
Amplification	4	4	100
Balanced translocation	1	1	100
Deletion	74	71, 2,[a] 1[b]	100
Duplication	22	22	100
Indel	29	28, 1[b]	100
Insertion	2	2	100
Inversion	5	3, 1,[a] 1[b]	100
Mobile element insertion	3	3	100
Methylation	24	13, 11[b]	100
Mosaic deletion	2	1	50
Mosaic duplication	3	3	100
Mosaic indel	2	2	100
Mosaic SNV	2	2	100
Mosaic trisomy	2	1	50
mtDNA deletion	2	2	100
mtDNA SNV heteroplasmy	8	8	100
High pseudogene homology	4	4[a]	100
Phase	42	18, 24[b]	100
Rearrangement	1	1	100
Recombinant allele	1	1	100
Repeat contraction	2	2	100
Repeat expansion	77	70, 2,[a] 5[b]	100
Runs of homozygosity	45	44, 1[a]	100
Skewed X-inactivation	1	1[b]	100
SNV	120	115, 5[b]	100
Translocation	1	1	100
Trisomy	2	2	100
Total	481	479	99.6

"Expected" is defined by the results from the original clinical report or suspected based on clinical findings. "Observed" refers to the results of HiFi sequencing. Indel, insertion or deletion; SNV, single-nucleotide variant; mtDNA, mitochondrial DNA.
[a]Different variant called in the same genomic location as the expected result (e.g., an inversion instead of a deletion in the same region of a gene).
[b]Variant suspected based on the clinical findings but not confirmed molecularly prior to HiFi.

across versions where possible. Additionally, sample-specific details are provided in Table S4. Data were then processed using a bioinformatics pipeline based on the PacBio human whole-genome sequencing (WGS) WDL pipeline. Alignment to GRCh38 was performed with pbmm2 v.1.17.0. Reads were tagged by haplotype with WhatsHap v.2.4. Single-nucleotide variants (SNVs) and insertions or deletions (indels) were called with DeepVariant v.1.8.0, SVs with PBSV v.2.11.0 and Sawfish v.1.0.1, and copy-number variants (CNVs) with HifiCNV v.1.0.1. SNVs were phased with WhatsHap v.2.4. Loci at which pathogenic variants have been previously reported in short tandem repeats were called using Tandem Repeat Genotyping Tool (TRGT) v.1.5.1, and variants in regions known to have high homology due to pseudogenes and paralogs were called with Paraphase v.3.2.1. Assessment of CpG islands was performed with pb-CpG-tools v.2.3.2. Runs of homozygosity (ROHs) were identified with BCFtools v.1.14. Depth of coverage was assessed with MosDepth v.0.3.11, and variant calls were annotated using Slivar v.0.3.1. Command-line parameters and additional software information can be found in Table S5.

To measure concordance, the expected results were the variant classifications found in the clinical laboratory report. Previously detected variants that were present in the output of at least one caller in the HiFi pipeline were classified as concordant, while the absence of a call was classified as discordant. A different variant called in the same genomic location as the expected result (e.g., an inversion instead of a deletion in the same region of a gene) was classified as "concordant, revised variant" (marked by

footnote a in Table 1 and yes* in Table S1). Some variants were suspected based on the case information and discussed in the clinical laboratory report but were not confirmed molecularly prior to HiFi. These were classified as "concordant, new information" (marked by footnote b in Table 1 and yes^{+} in Table S1) in case the suspected variant was called.

Detection of imprinting-related methylation signatures using HiFi

HiFi data were analyzed for methylated CpGs (mCs) using the pb-CpG-tools v.3.0.0. HiFi reads with kinetics were generated from PacBio subread BAM files using ccs, and the resulting 5mC-tagged reads were aligned to the hg38 reference genome using pbmm2 v.1.17.0. Our cohort included samples with known imprinting disorders: fragile X syndrome (MIM: 300624); fragile X tremor/ataxia syndrome (MIM: 300623); diabetes mellitus, transient neonatal 1 (MIM: 601410); Prader-Willi syndrome (MIM: 176270); Angelman syndrome (MIM: 105830); Silver-Russell syndrome (MIM: 180860); Mulchandani-Bhoj-Conlin syndrome (MIM: 617352); Temple syndrome (MIM: 616222); Kagami-Ogata syndrome (MIM: 608149); and Desbuquois dysplasia-2 (Baratela-Scott syndrome; MIM: 615777). For each disorder, we examined the relevant CpG island(s) (Table S6) and compared average methylation values obtained from pb-CpG-tools between clinically positive samples and control samples. We expected control samples to have a methylation value of 50% within associated CpG islands for autosomal imprinting disorders and 50% (females) or 0% (males) for X-linked disorders. To validate these results, the BAM files containing methylation data were reviewed in the Integrative Genomics Viewer (IGV), and CpG islands of interest were inspected in positive and control samples. This review was completed independently by two scientists as a quality-control check.

Confirmation of mosaicism via digital droplet PCR or estimation via HiFi reads

Mosaicism for two trisomy samples was measured using digital droplet PCR (ddPCR) on an automated droplet generator and reader from Bio-Rad (QX200 Droplet Digital PCR, Bio-Rad, USA), which fractionates samples into ~20,000 droplets prior to PCR. For the mosaic chromosome 18 trisomy sample, a 70-bp amplicon for *RIT2* (MIM: 609592) was utilized (chr18:42759657 [GRCh38]; Hs06453395_cn; Thermo Fisher Scientific, USA), along with the *TERT* (MIM: 187270) copy-number reference (Thermo Fisher Scientific; catalog #4403316). For the mosaic chromosome 8 trisomy sample, a 77-bp amplicon in *DNAAF11* (MIM: 614930) (chr8:132675291; Hs05074776_cn; Thermo Fisher Scientific) was used, along with the *TERT* copy-number reference (Thermo Fisher Scientific; catalog #4403316). Each 22 μL reaction included 11 μL of 2× ddPCR Supermix for probes, 2.2 μL of TaqMan assay mix, and 12 ng of diluted DNA. Reactions were cycled as follows: 95°C for 10 min, 40 cycles of 94°C for 30 s and 60°C for 1 min, and a final extension at 98°C for 10 min. Fluorescence was measured using the QX200 droplet reader and analyzed with QX Manager Software to determine the copy number of each target. The mosaic level for the sample with two events in *SCN1A* (MIM: 182389) (deletion of exons 22–24 and deletion of exons 25–26; PBVAL_106) was estimated using read counts from the HiFi sequencing.

Statistical analyses

The effect of sample source and extraction chemistry on HiFi quality metrics was determined using Wilcoxon rank-sum tests. The 95% confidence interval (CI) for the overall sensitivity was obtained using the Agresti-Coull method. These analyses were done in R[17] v.4.4.2.

Results

Sensitivity of HiFi sequencing to detect pathogenic variants by event type

Of 481 VDDs within 27 event types expected based on the initial clinical report or suspected based on clinical findings, HiFi's pipeline of callers was able to automatically detect 479, or 99.6% (CI: 98.4%–100%) (Table 1). This included 49 VDDs that were suspected based on the clinical findings described during initial test ordering but for which molecular evidence was not available, including VDDs involving phase (24), methylation status (11), repeat expansions (5), SNVs (5), and four others (Table 1, events marked with footnote b). Additionally, 10 previously detected VDDs were further resolved or clarified by HiFi, including VDDs involving deletion (2), an inversion (1), pseudogene homology (4), repeat expansion (2), and a homozygosity run (1) (Table 1, events marked with footnote a).

HiFi provides additional information

Haplotype phasing and repeat expansion resolution

The use of HiFi sequencing allowed accurate determination of haplotype phase and repeat expansion sizing, which are often difficult to detect with SRS technologies. In our dataset, HiFi was able to set the phase for genetic variants separated by long molecular distances, including an indel in *GJB2* (MIM: 121011) and a deletion in *GJB6* (MIM: 604418) separated by 384 kb, two indels in *USH2A* (MIM: 608400) separated by 406 kb, an SNV and a CNV in *TANGO2* (MIM: 616830) separated by 270.5 kb, and two CNVs in *WWOX* (MIM: 605131) separated by 854 kb. Additionally, HiFi sequencing facilitated accurate sizing of repeat expansions across 16 distinct VDD scenarios, including a CTG expansion (up to 1,209 repeats) in *DMPK* (MIM: 605377), a CAGG expansion (up to 5,779 repeats) in *CNBP* (MIM: 116955), and GGC repeats (222 and 477 repeats) in *XYLT1* (MIM: 608124) (Table S1).

Resolving pseudogene homology and complex gene structures with Paraphase

To address mapping challenges posed by sequence similarity between genes and their paralogs or pseudogenes, Paraphase (v.3.3.2) was used to phase haplotypes and call variants for highly similar paralogous genes. Variants in genes homologous to pseudogenes can be particularly hard to detect using SRS technologies.[2,18] *NCF1* (MIM: 608512) exemplifies one such gene, as can be seen by the poor coverage in SRS (Figure 1A). *NCF1* has two pseudogenes with high sequence similarity[18] and poor coverage in both HiFi and SRS (Figures S1A and S1B), making variant calling challenging. Paraphase phases all haplotypes of the *NCF1* gene family together and assigns the

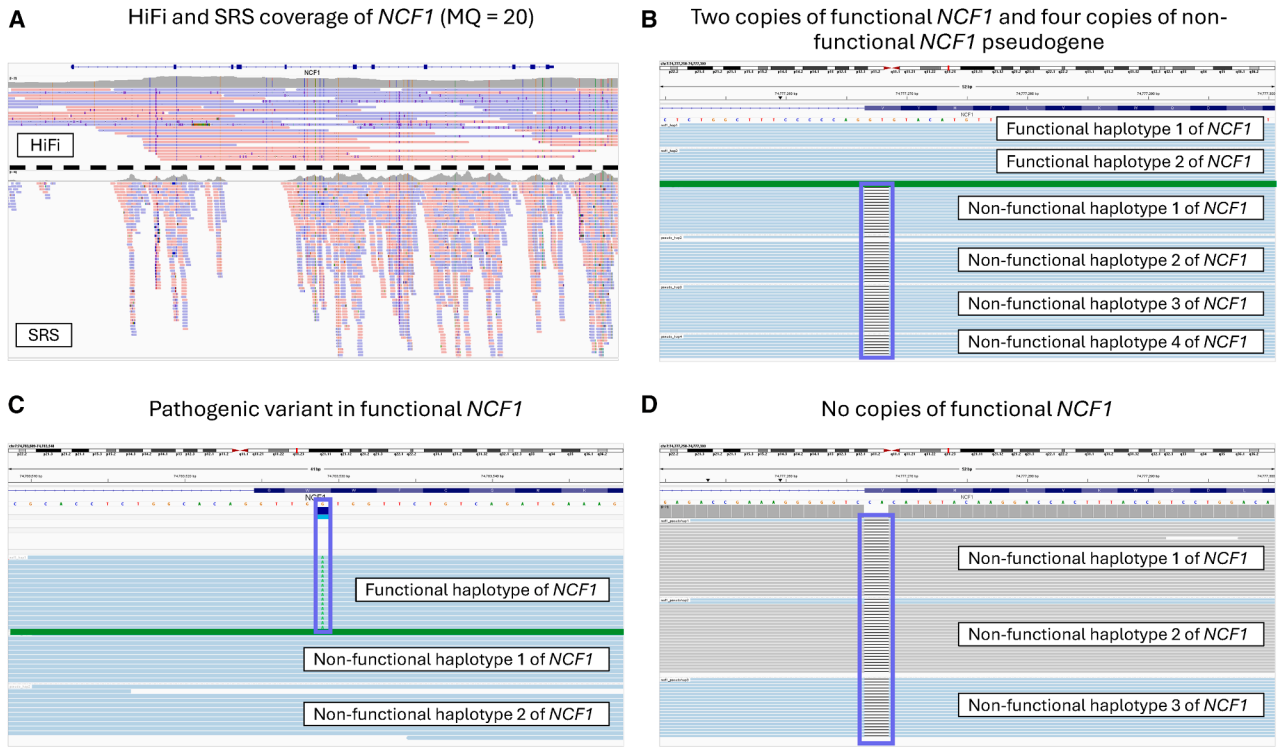

Figure 1. Integrative Genomics Viewer screenshots illustrating the ability of Paraphase to call *NCF1* genotypes

(A) Integrative Genomics Viewer (IGV) screenshot illustrating coverage of *NCF1* using HiFi and SRS (separated by a black dashed line; both mapping quality [MQ] = 20).

(B) IGV screenshot illustrating *NCF1* of a normal sample using a BAM file produced from Paraphase (MQ = 20). Paraphase phases haplotypes and assigns them into functional and non-functional (pseudogene) copies of *NCF1* based on the absence or presence of the ΔGT in exon 2 (blue box). The green line separates functional and non-functional copies of *NCF1*.

(C) IGV screenshot illustrating a homozygous stop gain variant (g.74783529G>A [GenBank: NC_000007.14]) in *NCF1* using HiFi (genome coverage: 21×; MQ = 20), analyzed with Paraphase (green line separates the functional and non-functional copies of NCF1). The variant (blue box) is present in the functional haplotype of *NCF1* that does not carry the ΔGT in exon 2. The three haplotypes are each present in two copies, consistent with the 95.2% ROH of this sample in this region.

(D) IGV screenshot illustrating a sample with no functional copies of *NCF1*. The results from Paraphase show three haplotypes for the non-functional (pseudogene) copies with ΔGT in exon 2 (blue box). Reads in gray are consistent with two haplotypes (first and second haplotypes, which are highly similar in sequence) and are randomly assigned to one haplotype. The three haplotypes are each present in two copies.

phased haplotypes to functional *NCF1* or non-functional pseudogene copies based on the status of the GT dinucleotide deletion that is characteristic of *NCF1* pseudogenes (Figure 1B). Using Paraphase, we were able to call pathogenic SNVs in functional copies of *NCF1* in two samples (Figures 1C and S1C). Finally, we detected samples with no copies of the functional gene, i.e., all haplotypes of the *NCF1* family carry the GT dinucleotide deletion (Figures 1D and S1D), suggesting possible gene conversion between *NCF1* and pseudogenes. Visualization of functional and non-functional copies of the *NCF1* gene family allows for simple calling of *NCF1* genotypes and can be used to identify pathogenic SNVs previously detected using a complex Sanger sequencing methodology.

Application of HiFi methylation profiling to diagnose imprinting disorders

HiFi sequencing demonstrated utility in resolving methylation disorders, particularly in the context of singleton testing. For example, ROHs in the 15q11.2q13.1 region could suggest either Prader-Willi or Angelman syndrome. Assessing methylation of the *SNURF-SNRPN* locus is key to distinguishing between these two conditions. We examined clinical samples with uniparental disomy (UPD) or deletions in the 15q11.2q13.1 region. Using methylation data for *SNURF-SNRPN* (MIM: 182279) (chr15:24954889–24955906; CpG island: 77), we were able to determine whether individuals were affected with Prader-Willi or Angelman syndrome (see examples in Figure 2). In addition, we used samples with amplification of the 15q11.2q13.1 region to test methylation calling for this region. All samples showed results similar to the clinical report. To further test the use of methylation data, we examined samples of known methylation disorders, such as transient neonatal diabetes mellitus. We demonstrated that a sample (PBVAL_042) with suspected transient neonatal diabetes mellitus (upd(6)paternal) showed hypomethylation in the promoter of *PLAGL1* (MIM: 603044) (chr6:144007780–144008709; CpG island: 111) and confirmed this disorder molecularly. We were able to use methylation data to confirm Silver-Russell (upd(7)maternal; CpG island: 120;

PBVAL_137 and PVAL_140], Temple (upd(14)mat; CpG islands: 45 and 18; PBVAL_091), Kagami-Ogata (upd(14)pat; CpG islands: 45 and 18; PBVAL_139), and Mulchandani-Bhoj-Conlin (upd(20)mat; CpG island: 190; PBVAL_138) syndromes as suspected from parental sequencing. All data for our methylation analysis are provided in Table S6.

Integrated detection of repeat expansions and methylation signatures

Long-read sequencing is useful for detecting repeat expansions, repeat contractions, and methylation associated with different disease modalities. HiFi detected repeat expansions in the *XYLT1* (MIM: 608124) promoter region, which is associated with hypermethylated alleles and is causative of Baratela-Scott syndrome.[19] HiFi sequencing enabled simultaneous detection of both repeat expansions and methylation status at the *XYLT1* locus in all affected individuals with suspected involvement (PBVAL_155 and PBVAL_168). Beyond providing this integrated and novel insight, HiFi also clarified a case initially reported as heterozygous, revealing that both alleles were, in fact, expanded (PBVAL_168). In addition, we examined samples with suspected *XYLT1* expansions (PBVAL_158, PBVAL_161, and PBVAL_162). Two of these samples showed an expansion of the *XYLT1* repeat with hypermethylation of the expanded allele (PBVAL_158 and PBVAL_161).

Similarly, long repeat expansions in the 5′ UTR of *FMR1* (MIM: 309550), which lead to hypermethylation of the promoter region and are associated with fragile X syndrome, can be detected by HiFi. In our dataset, HiFi detected methylation in all samples positive for the *FMR1* repeat expansion (PBVAL_113, PBVAL_166, and PBVAL_181; Figure S2). Additionally, HiFi was able to detect uneven methylation associated with carrier status and mosaicism, as seen in a female with fragile X tremor/ataxia syndrome and/or premature ovarian failure 1 (MIM: 311360) with an expanded allele (PBVAL_167; Figure S3) and in mosaicism for males (PBVAL_170 and PBVAL_171; Figure S4). For example, in one individual, hypermethylation was observed in alleles with >200 repeats and reduced methylation was observed in alleles with ∼40 repeats (within the typical allele size range). For a *DMPK* repeat expansion causative of myotonic dystrophy I (MIM: 160900), we analyzed a sample classified as mosaic. Again, mosaicism was visible for PBVAL_119 (Figure S5) compared to a positive sample for *DMPK* without mosaic alleles (PBVAL_141). Since some standard testing can miss mosaicism of *FMR1* and *DMPK* expanded repeats, the ability of HiFi to detect these events is an important feature.

HiFi limitations: Missed mosaic events

The events missed by HiFi consisted of two mosaic events: a deletion and a trisomy (Table 2). Two additional VDDs were also missed initially—multi-exon deletions in *PKD1* (MIM: 601313) (exons 31–33, 595 bp) and *MCOLN1* (MIM: 605248) (exons 1–7, 6,311 bp). While *PKD1* is known to have sequence homology with pseudogenes, no such confounding factors are known for *MCOLN1*. In both affected individuals, the VDDs were successfully detected following loading modifications in a subsequent run. The modification involved using a Hamilton Star system for shearing, which improved the size of the fragments loaded. Even if one considered these missed events, the concordance would still be >99%.

The mosaic trisomy 18 that was missed was measured by ddPCR and showed a level of mosaicism of 23% (30.2× coverage). In contrast, a trisomy 8 sample showed a level of mosaicism of 57% (26.2× coverage) and was called by our pipeline. Therefore, the level of mosaicism that can be called using HiFi is potentially limited by sequencing depth.

While these mosaic events were not called automatically, the missed *SCN1A* mosaic deletion of exons 22–24 (13% estimated mosaicism) was visible in IGV. Of note, this sample contained a second mosaic deletion of exons 25–26 in *SCN1A* (29% estimated mosaicism) that was successfully called with the HiFi pipeline. Figure S6 shows the genomic region with both deletions.

Potential utility of HiFi compared to previous diagnostic testing

Of the 191 individual samples included in this study, 103 (54%) harbored more than one VDD, spanning multiple event types. Prior to the application of HiFi sequencing, the six most frequently employed diagnostic technologies in this cohort were Sanger sequencing, ES/GS SRS, TPP assays, chromosomal microarrays, targeted gene panels, and MLPA (Figure 3A).

Two variants not detected automatically by the HiFi pipeline were previously identified either by microarray alone (mosaic trisomy) or by a combination of ES SRS and microarray (mosaic deletion), highlighting specific limitations of HiFi or existing analysis tools in detecting low-level mosaic events.

Among individuals in this cohort, an average of 1.62 diagnostic tests were performed prior to obtaining a definitive result (Figure 3B). The median time from sample receipt to completion of the clinical report was 27.5 days (Figure 3C), whereas the estimated time frame for generating results through HiFi in this study was 5–7 days.

The ability of HiFi to reduce the number of tests required to achieve a molecular diagnosis is illustrated by various affected individuals in this cohort. For instance, a newborn with hypotonia underwent four different test types: MLPA and Sanger sequencing for *SMN1*/*SMN2* dosage analysis, TPP for *DMPK* repeat expansion analysis, and both MLPA and msMLPA for Prader-Willi syndrome analysis. In this specific affected individual, all tests were ordered simultaneously, and the time to the result report was 13 days. HiFi alone would be able to detect all relevant VDDs and the methylation status captured by these individual tests within a predicted turnaround time shorter than the observed 13 days.

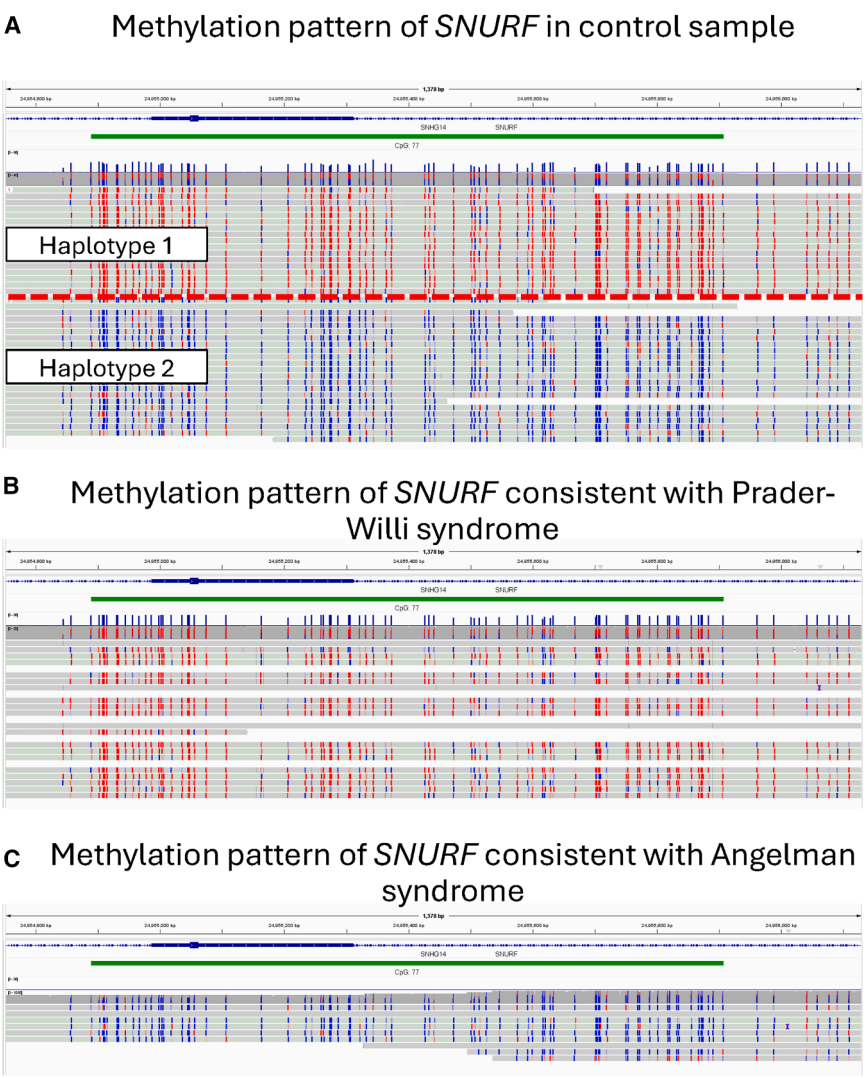

Figure 2. Integrative Genomics Viewer screenshots illustrating Prader-Willi syndrome and Angelman syndrome examples

Red indicates 100% and blue indicates 0% methylation.

(A) Integrative Genomics Viewer (IGV) screenshot illustrating methylation of *SNURF* (chr15:24954752–24956128; CpG: 77) in a control sample.

(B) IGV screenshot showing a sample with normal copy-number analysis and maternal UPD for the 15q11.2–15q13.1 genomic region. The sample showed 100% methylation of *SNURF* (chr15:24954752–24956128; CpG: 77). These results are consistent with Prader-Willi syndrome.

(C) IGV screenshot showing a sample with normal copy-number analysis and paternal UPD for the 15q11.2–15q13.1 genomic region. The sample showed 0% methylation of *SNURF* (chr15:24954752–24956128; CpG: 77). These results are consistent with Angelman syndrome.

Genome coverage and quality by DNA source and extraction chemistry

Across all sample types and extraction chemistries, we obtained a mean genome coverage of 24.7× and a mean N50 read length of 15,392 bp (Table 3). Blood samples yielded a higher mean coverage than buccal samples with either extraction chemistry (MagMax: Wilcoxon rank-sum test statistic or W = 502, $p = 0.0010$; QIASymphony: W = 246, $p = 0.0038$), with no significant difference in N50 read length for MagMax (W = 296, $p = 0.588$) but a larger value for QIASymphony (W = 236, $p = 0.0094$) (Figures 4A and 4B). HMW extraction methods yielded a significantly larger mean coverage (W = 3091, $p < 0.0001$), with no difference in N50 read length (W = 2285, $p = 0.3207$) (Figures 4C and 4D).

Discussion

We show that by using a standard HiFi sequencing workflow and variant-calling pipeline, we can automatically detect 99.6% (479/481) of VDDs without the need for complementary diagnostic methodologies, even when applied to real-world clinical laboratory DNA samples obtained from various sources and using various extraction methods. This workflow missed only two events, both of which were mosaic variants. Furthermore, HiFi was able to uncover additional clinically relevant variant information that was not available using previous laboratory diagnostic methods, highlighting its value in the context of rare genetic conditions associated with variants that are difficult to detect using current diagnostic methodologies. Together, these findings extend prior accuracy benchmarks by demonstrating robust performance under challenging conditions and support the potential use of HiFi as either a first-line diagnostic test or as a follow-up to SRS, particularly as the technology scales to high-throughput clinical diagnostic applications.

The subset of individuals in this study originally tested at GeneDx received their diagnosis after a median time of nearly 1 month, with their samples undergoing more

Table 2. Missed events using HiFi sequencing

Short-read/lab results	Event type	Original detection methods	Sample type; extraction	Mean coverage; % mosaicism
arr[GRCh37] (18)x3[0.3] mosaic gain	mosaic trisomy	microarray	blood; HMW	30.2×; 23%
arr[GRCh37] 2q24.3(166850968_166856561)x1.5 dn (chr2:165994458–166000051 hg38) *SCN1A* exons 22-24	mosaic deletion	exome SRS, microarray	buccal; non-HMW	17.9×; 13%

than one test on average to confirm the result. In contrast, in the context of this research study, we estimated a time frame of 5–7 days to generate results from HiFi. While this does not address the turnaround time for lrGS in a clinical diagnostic setting, it demonstrates the potential of HiFi to streamline diagnostic workflows, particularly since lrGS often does not require supplementary testing methodologies. A study comparing a group tested by HiFi to a standard-of-care control group (including SRS, karyotype, FISH, and CMA) found that the HiFi group obtained results significantly faster (27 vs. 62 days) and had a higher diagnostic yield.[20] Similar to our findings, that study reported that an average of 1.6 tests was needed in the standard-of-care group to obtain a diagnosis, compared to a single test with HiFi. Therefore, HiFi is likely to offer a substantially faster time to result than conventional testing, enabling quicker access to targeted treatments and changes in management and goals of care and, for many, potentially reducing the diagnostic odyssey.

The two variants not detected by our automated pipeline were a mosaic trisomy of chromosome 18 and a mosaic deletion spanning exons 22–24 of *SCN1A*. Similar limitations have been reported in other studies using HiFi. For instance, in a smaller study by Eisfeldt et al.,[21] one of the three events missed by lrGS (out of 16 total events) was a mosaic partial trisomy and tetrasomy on chr15. Our pipeline did, however, detect a mosaic trisomy for chr8 (PBVAL_001). Both the missed and the identified mosaic trisomies in our study were further investigated using ddPCR and found to differ markedly in terms of the level of mosaicism. The missed trisomy (30.2× coverage) had a much lower percentage of mosaicism (23%) compared to the detected trisomy 8 (57%; 26.2× coverage). For the undetected mosaic deletion (exons 22–24 in *SCN1A*), we note that the same sample also carried a mosaic deletion of exons 25–26 of *SCN1A* that was called by HiFi. The missed mosaic deletion in this sample had a lower estimated level of mosaicism (13%) than the detected one (29%). The DNA from this sample was obtained from a buccal sample, and the corresponding (genome) coverage depth was lower (17.9×) than the average cohort depth (24.7×). These findings suggest that the HiFi approach used in this study has modest sensitivity for detecting low levels of mosaicism and that a higher depth of coverage is still important for detecting mosaic variants. As HiFi potentially moves toward commercial adoption in diagnostic research laboratories, establishing minimum coverage standards, rather than relying on data from a single flow cell/run per sample, may be necessary for reliable mosaic variant detection.

A recent study by Höps et al.[9] evaluated the sensitivity of HiFi lrGS for clinically relevant VDDs in 100 samples. Their variant callers detected only 83% of expected variants, with another ∼10% visible upon manual review but not automatically called. The variants not detected automatically fell into four categories: skewed X-inactivation events ($n = 3$), GA-based repeat expansions in *RFC1* (MIM: 102579) ($n = 4$) or *FXN* (MIM: 606829) ($n = 1$), variants in segmental duplications ($n = 2$), and variants with breakpoints in highly repetitive regions ($n = 2$). In contrast, our study achieved much higher (99.6%) sensitivity, which we attribute largely to the use of different—and generally more recent[22]—analysis tools and variant callers (Table 4). For example, our pipeline automatically detected one skewed X-inactivation event using XiSkewAnalysis, whereas Höps et al.[9] used an in-house workflow to quantify X-inactivation. No well-validated methods for quantifying X-inactivation from lrGS currently exist, and there have yet to be robust studies correlating lrGS-based X-inactivation quantification with the conventional diagnostic assay for skewed X-inactivation, HUMARA, which measures methylation of CpG sites near a trinucleotide repeat in exon 1 of *AR*. As such, we noted at least one major difference between our pipelines that may explain the large difference: XiSkewAnalysis quantifies methylation at 5,642 CpG sites selected to be unmethylated in males and half-methylated in females, whereas Höps et al.'s workflow quantifies methylation at 16,109 sites unmethylated in males only. Our pipeline also detected eight variants involving GA-based repeat expansions in *FXN*, which we attribute to using an updated version of the TRGT software (v.3.0.0 vs. v.0.4.0). Höps et al.'s hypothesis that low coverage caused *FXN* repeat-calling failures seems less likely, as their mean sample coverage (30.92×) overlapped with ours (22.74×–36.34×). Both studies successfully detected variants in *SMN1/SMN2*.

Direct comparison between our study and that of Hops et al.[9] is limited by differences in variant representation: our dataset did not include *RFC1* repeat expansions, while theirs lacked *NCF1* variants—a challenging genomic region due to its homology to pseudogenes that are difficult to map.[5,16] Furthermore, the translocations tested in our study are not directly comparable to the ones tested by Höps et al.[9] Overall, these results underscore both the

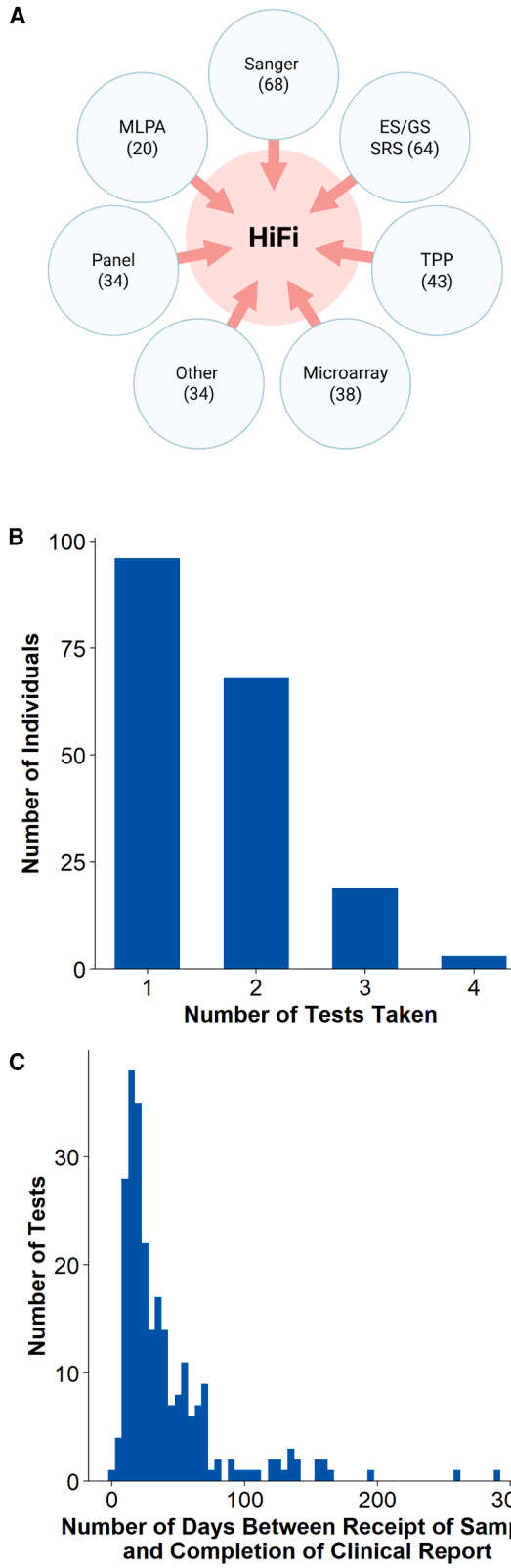

Figure 3. Characterization of clinical tests performed to reach a diagnosis for individuals in the cohort

(A) The six most frequently used technologies (out of 16) in the clinical laboratory to identify the 481 variants in the 191 individual samples. Numbers in parentheses indicate the number of individuals requiring the use of each specific technology for diagnosis. The "other" category includes FISH, gel-based assay, karyotyping, optical genome mapping, Southern blot, methylation-specific MLPA, methylation TPP, mitochondrial DNA sequencing, and qPCR. TPP, triplet-primed PCR; MLPA, multiplex ligation-dependent probe amplification.

(B) Distribution of the number of tests taken to achieve a diagnosis in 186 individuals (out of 191) with information on any previous testing. The mean number of tests was 1.62.

(C) Distribution of the number of days between the date a sample was received and the date the report was finished for samples tested at GeneDx. The median number of days was 27.5. Note that eight extreme values were removed (>365 days), reflecting samples for which re-analysis was triggered at a much later date after initial receipt of the sample.

rapid development of HiFi analysis tools and the substantial gains in automated detection performance, suggesting that automated calling of challenging variant types is increasingly within reach.

HiFi added information or clarification to 59 events, most commonly by providing phasing or methylation data. Phasing can improve diagnostic yield and refine variant interpretation. When the phase is known, some variants that might otherwise be reported as variants of uncertain significance can be excluded, and potential cases of compound heterozygosity can be assessed more confidently. Moreover, haplotype-specific variation may influence disease susceptibility or clinically relevant phenotypes, making the phase and methylation data obtainable from lrGS particularly valuable.[23] With SRS, phasing typically requires sequencing the proband plus family members (ideally both parents). Trio sequencing increases costs and is not always possible if parents are unavailable. The ability of lrGS to deliver read-backed phasing without trios or extra technologies represents a distinct advantage of this approach in the diagnostic landscape.

Ideally, lrGS would perform reliably across diverse sample sources and extraction methods, even when HMW DNA is not available. In our study, however, we found that extraction methods significantly affected sequencing depth and N50 read length, and buccal samples in particular showed reduced coverage, which may necessitate additional sequencing on a second SMRT Cell and lead to a higher overall sequencing cost. The optimal coverage depth for lrGS (HiFi or ONT) in a diagnostic setting has not yet been established. Of note, mosaic detection using HiFi was not a validation objective of this study, and the finding that mosaic variants with lower fractional abundance or reduced coverage were more likely to be missed underscores the importance of establishing minimum coverage thresholds. While 30× coverage is the standard for short-read clinical ES/GS,[24,25] the coverage depth required for diagnostic application of lrGS may differ depending on the specific intent of the assay. For example, a clinical assay designed solely to identify and characterize large SVs may only need 15× coverage, while a clinical lrGS assay that is intended to replace SR genomes as a first-tier test may demand 30× or higher coverage, with corresponding increased sequencing costs, in order to

Table 3. Summary of sample source, extraction chemistry, and resulting HiFi-lrGS sequence quality

Sample biological source	Sample extraction method	*n*	Mean genome coverage (x) ± SD	Mean mtDNA coverage (x) ± SD	Mean reads (M) ± SD	Mean yield (GB) ± SD	N50 read length (bp) ± SD
Blood	HMW	61	30.2 (4.71)	310 (220)	6.66 (0.90)	96.0 (14.5)	17,138 (2,157)
Blood	non-HMW	68	25.4 (5.68)	600 (786)	6.61 (1.32)	81.7 (18.1)	15,598 (4,419)
External DNA	unknown[a]	18	21.5 (6.96)	1,024 (1182)	5.64 (1.17)	68.5 (22.2)	14,539 (3,677)
Buccal	HMW	5	24.2 (7.21)	1,210 (568)	7.40 (1.73)	84.2 (17.2)	14,751 (1,709)
Buccal	non-HMW	39	16.8 (6.04)	1,780 (1156)	6.44 (1.77)	63.5 (22.6)	12,777 (3,899)
All samples	all	191	24.7 (7.39)	804 (964)	6.52 (1.34)	81.4 (22.0)	15,392 (3,892)

mtDNA, mitochondrial DNA.
[a]Extraction chemistry for samples submitted as DNA was not known.

ensure sufficient sensitivity to mosaic variation, including mitochondrial heteroplasmy.[9,26,27] Our data indicate that 20×–30× coverage may be sufficient for the detection of most variants. For lrGS to fulfill its potential as a single, best-in-class diagnostic test, technical challenges—such as automating HMW DNA extraction and defining minimum coverage thresholds—must be addressed.

While our study demonstrates the validity of HiFi sequencing for detecting VDDs currently diagnosed using a combination of methodologies employed by clinical labs, it was not designed to determine cutoff values for sensitivity and specificity of variant-calling workflows. Future studies are needed to understand scenarios in which short-read ES- or GS-based tests may perform better and be more advantageous than HiFi sequencing, with the understanding that this risk-benefit analysis will be ever evolving as lrGS technologies improve, allowing for higher throughput. Two applications for which there is increasing recognition of the significant benefit of lrGS are in the diagnosis of individuals with suspected rare genetic disorders that remain intractable despite extensive prior sequencing and analysis[8,28] and of critically ill individuals for whom the fastest time to diagnosis is especially important.

Finally, while our results support the potential of HiFi sequencing as a first-tier diagnostic test, broader implementation will require careful consideration of

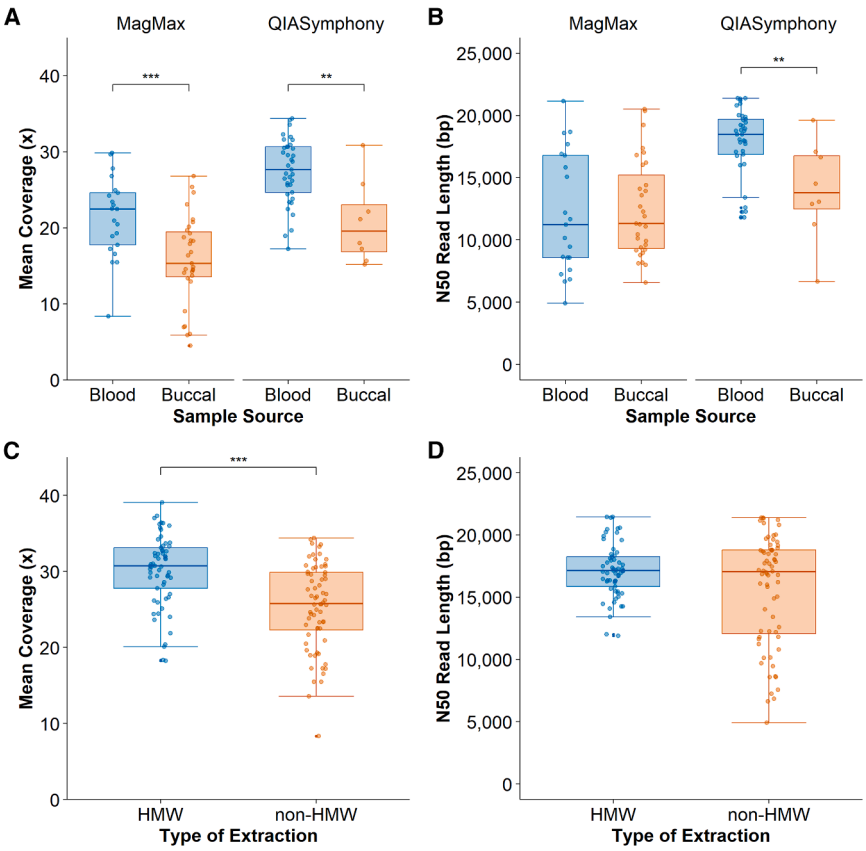

Figure 4. Quality measures of HiFi for different sample types and extraction chemistries

(A and B) Mean coverage (A) and N50 read length (B) for blood and buccal samples extracted using the same extraction chemistries.

(C and D) Mean coverage (C) and N50 read length (D) for blood samples extracted with kits designed to obtain high-molecular-weight (HMW) DNA and standard kits (non-HMW). Boxes show the interquartile range (IQR; Q1–Q3) with the median indicated by a central line. Whiskers extend to observations within 1.5 × IQR, and raw data points are displayed. Wilcoxon rank-sum test: ***$p \leq 0.001$ and **$p \leq 0.01$.

Table 4. Summary of tools used in Höps et al. and the current study

Tool used for	Höps et al.[9]	Current study
Alignment	pbmm2 v.1.10.0	pbmm2 v.1.17.0
Generating haplotigs	Hifiasm v.0.15.3	WhatsHap v.2.4
Structural variants	PBSV v.2.9.0	PBSV v.2.11.0; Sawfish v.1.0.1
Small variants	DeepVariant v.1.5.0	DeepVariant v.1.8.0
Phasing information	HiPhase v.0.10.1	HiPhase v.1.4.5
Copy-number variants	HifiCNV v.0.1.6	HifiCNV v.1.0.1
Short tandem repeats	TRGT v.0.4.0	TRGT v.1.5.1
Paralogs and pseudogenes	Paraphase v.2.2.3	Paraphase v.3.2.1
Methylation calls	pb-CpG-tools v.2.3.1	pb-CpG-tools v.2.3.2
Regions of homozygosity	visually	BCFtools v.1.14
XCI skewing	in-house script	XiSkewAnalysis

XCI, X chromosome inactivation.

throughput, cost, and operational feasibility. The Revio platform has the potential to process approximately 75–100 genomes per month per SMRT Cell at ~30× coverage; however, this throughput is much lower than the one obtained by SRS instruments (as much as 10× lower), highlighting that sequencing capacity is still a bottleneck for the implementation of lrGS in a high-volume clinical diagnostic environment. In contrast, HMW DNA extraction and library preparation can be automated, contributing to the operational feasibility of lrGS. Perhaps the most significant obstacle still includes the interpretation of lrGS, which will require increasing automation and storage capacity as the technology scales. These operational considerations must be balanced against the advantages of lrGS, which in our study enabled sensitive detection of diverse challenging variant types, improved automated calling compared with earlier reports, and provided additional diagnostic value through phasing, methylation analysis, and accurate characterization of repeat expansions and complex SVs—capabilities that are limited or require multiple complementary assays in standard SRS workflows. These represent practical trade-offs that laboratories must consider when determining how and when to integrate lrGS into routine diagnostic practice.

In conclusion, our findings demonstrate that HiFi sequencing detects a wide range of VDDs in real-world clinical laboratory samples, highlighting a key advantage of lrGS and demonstrating its capability to be deployed as a first-tier test over the myriad complementary technologies originally used to detect these VDDs.

Data and code availability

The datasets generated during this study, along with the code used to prepare figures not created in Excel or PowerPoint, are available in the supplemental information.

Acknowledgments

Three samples in this study were previously diagnosed by University of Washington Center for Rare Disease Research with support from NHGRI grants U01HG011744 and U24HG011746. We thank PacBio for providing some of the reagents used in this study.

Author contributions

Investigation, J.M.D., J.N., A.S.B., S.Y., S.L., and J.S.; writing – original draft, J.M.D., J.X.C., P.C.L., J.N., A.S.B., S.Y., F.M.F., and M.J.B.; writing – review & editing, K.McW., K.N., S.R.P., A.B.S., J.G., P.K., E.D., X.C., A.V.R., W.J.R., J.A.L., M.A.E., and B.J.; visualization, J.M.D. and P.C.L.; conceptualization, J.X.C., A.B.S., A.V.R., A.C., F.M.F., M.J.B., and B.J.; resources, J.X.C., M.J.B., and B.J.; formal analysis, P.C.L.; data curation, P.C.L. and S.R.P.; supervision, R.B., K.S.H., and F.M.F.; software, L.L., E.D., X.C., W.J.R., J.A.L., and R.K.

Declaration of interests

J.M.D., P.C.L., J.S., J.N., A.S.B., S.Y., S.L., R.B., K.S.H., L.L., K.McW., K.N., S.R.P., R.K., P.K., F.M.F., and B.J. are or were employees and may be shareholders of GeneDx. E.D., X.C., A.V.R., W.J.R., J.A.L., and M.A.E. are or were employees and may be shareholders of PacBio. PacBio provided a portion of the reagents used in this study. A.C. is an employee of Google LLC and owns Alphabet stock as part of the standard compensation package.

Supplemental information

Supplemental information can be found online at https://doi.org/10.1016/j.ajhg.2026.04.001.

Web resources

GenBank, https://www.ncbi.nlm.nih.gov/genbank/
OMIM, https://www.omim.org

Received: November 14, 2025
Accepted: April 1, 2026
Published: April 27, 2026

References

1. Chung, C.C.Y., Hue, S.P.Y., Ng, N.Y.T., Doong, P.H.L., Hong Kong Genome Project, Chu, A.T.W., and Chung, B.H.Y. (2023). Meta-analysis of the diagnostic and clinical utility of exome and genome sequencing in pediatric and adult patients with rare diseases across diverse populations. Genet. Med. *25*, 100896. https://doi.org/10.1016/j.gim.2023.100896.
2. Sedlazeck, F.J., Lee, H., Darby, C.A., and Schatz, M.C. (2018). Piercing the dark matter: bioinformatics of long-range sequencing and mapping. Nat. Rev. Genet. *19*, 329–346. https://doi.org/10.1038/s41576-018-0003-4.
3. Yang, Y., del Gaudio, D., Santani, A., and Scott, S.A. (2024). Applications of genome sequencing as a single platform for clinical constitutional genetic testing. Genet. Med. Open *2*, 101840. https://doi.org/10.1016/j.gimo.2024.101840.
4. Schobers, G., Derks, R., den Ouden, A., Swinkels, H., van Reeuwijk, J., Bosgoed, E., Lugtenberg, D., Sun, S.M., Corominas Galbany, J., Weiss, M., et al. (2024). Genome sequencing as a generic diagnostic strategy for rare disease. Genome Med. *16*, 32. https://doi.org/10.1186/s13073-024-01301-y.
5. Wenger, A.M., Peluso, P., Rowell, W.J., Chang, P.-C., Hall, R.J., Concepcion, G.T., Ebler, J., Fungtammasan, A., Kolesnikov, A., Olson, N.D., et al. (2019). Accurate circular consensus long-read sequencing improves variant detection and assembly of a human genome. Nat. Biotechnol. *37*, 1155–1162. https://doi.org/10.1038/s41587-019-0217-9.
6. Cohen, A.S.A., Farrow, E.G., Abdelmoity, A.T., Alaimo, J.T., Amudhavalli, S.M., Anderson, J.T., Bansal, L., Bartik, L., Baybayan, P., Belden, B., et al. (2022). Genomic answers for children: Dynamic analyses of >1000 pediatric rare disease genomes. Genet. Med. *24*, 1336–1348. https://doi.org/10.1016/j.gim.2022.02.007.
7. AlAbdi, L., Shamseldin, H.E., Khouj, E., Helaby, R., Aljamal, B., Alqahtani, M., Almulhim, A., Hamid, H., Hashem, M.O., Abdulwahab, F., et al. (2023). Beyond the exome: utility of long-read whole genome sequencing in exome-negative autosomal recessive diseases. Genome Med. *15*, 114. https://doi.org/10.1186/s13073-023-01270-8.
8. Hiatt, S.M., Lawlor, J.M.J., Handley, L.H., Latner, D.R., Bonnstetter, Z.T., Finnila, C.R., Thompson, M.L., Boston, L.B., Williams, M., Rodriguez Nunez, I., et al. (2024). Long-read genome sequencing and variant reanalysis increase diagnostic yield in neurodevelopmental disorders. Genome Res. *34*, 1747–1762. https://doi.org/10.1101/gr.279227.124.
9. Höps, W., Weiss, M.M., Derks, R., Galbany, J.C., Ouden, A.d., van den Heuvel, S., Timmermans, R., Smits, J., Mokveld, T., Dolzhenko, E., et al. (2025). HiFi long-read genomes for difficult-to-detect, clinically relevant variants. Am. J. Hum. Genet. *112*, 450–456. https://doi.org/10.1016/j.ajhg.2024.12.013.
10. Hon, T., Mars, K., Young, G., Tsai, Y.-C., Karalius, J.W., Landolin, J.M., Maurer, N., Kudrna, D., Hardigan, M.A., Steiner, C.C., et al. (2020). Highly accurate long-read HiFi sequencing data for five complex genomes. Sci. Data *7*, 399. https://doi.org/10.1038/s41597-020-00743-4.
11. Jain, M., Koren, S., Miga, K.H., Quick, J., Rand, A.C., Sasani, T.A., Tyson, J.R., Beggs, A.D., Dilthey, A.T., Fiddes, I.T., et al. (2018). Nanopore sequencing and assembly of a human genome with ultra-long reads. Nat. Biotechnol. *36*, 338–345. https://doi.org/10.1038/nbt.4060.
12. Beyter, D., Ingimundardottir, H., Oddsson, A., Eggertsson, H.P., Bjornsson, E., Jonsson, H., Atlason, B.A., Kristmundsdottir, S., Mehringer, S., Hardarson, M.T., et al. (2021). Long-read sequencing of 3,622 Icelanders provides insight into the role of structural variants in human diseases and other traits. Nat. Genet. *53*, 779–786. https://doi.org/10.1038/s41588-021-00865-4.
13. Mahmoud, M., Huang, Y., Garimella, K., Audano, P.A., Wan, W., Prasad, N., Handsaker, R.E., Hall, S., Pionzio, A., Schatz, M.C., et al. (2024). Utility of long-read sequencing for All of Us. Nat. Commun. *15*, 837. https://doi.org/10.1038/s41467-024-44804-3.
14. Gustafson, J.A., Gibson, S.B., Damaraju, N., Zalusky, M.P.G., Hoekzema, K., Twesigomwe, D., Yang, L., Snead, A.A., Richmond, P.A., De Coster, W., et al. (2024). High-coverage nanopore sequencing of samples from the 1000 Genomes Project to build a comprehensive catalog of human genetic variation. Genome Res. *34*, 2061–2073. https://doi.org/10.1101/gr.279273.124.
15. Del Gobbo, G.F., and Boycott, K.M. (2025). The additional diagnostic yield of long-read sequencing in undiagnosed rare diseases. Genome Res. *35*, 559–571. https://doi.org/10.1101/gr.279970.124.
16. Mandelker, D., Schmidt, R.J., Ankala, A., McDonald Gibson, K., Bowser, M., Sharma, H., Duffy, E., Hegde, M., Santani, A., Lebo, M., and Funke, B. (2016). Navigating highly homologous genes in a molecular diagnostic setting: a resource for clinical next-generation sequencing. Genet. Med. *18*, 1282–1289. https://doi.org/10.1038/gim.2016.58.
17. R Core Team (2024). R: A Language and Environment for Statistical Computing (R Foundation for Statistical Computing).
18. Chanock, S.J., Roesler, J., Zhan, S., Hopkins, P., Lee, P., Barrett, D.T., Christensen, B.L., Curnutte, J.T., and Görlach, A. (2000). Genomic Structure of the Human p47-phox (*NCF1*) Gene. Blood Cells. Mol. Dis. *26*, 37–46. https://doi.org/10.1006/bcmd.2000.0274.
19. LaCroix, A.J., Stabley, D., Sahraoui, R., Adam, M.P., Mehaffey, M., Kernan, K., Myers, C.T., Fagerstrom, C., Anadiotis, G., Akkari, Y.M., et al. (2019). GGC Repeat Expansion and Exon 1 Methylation of XYLT1 Is a Common Pathogenic Variant in Baratela-Scott Syndrome. Am. J. Hum. Genet. *104*, 35–44. https://doi.org/10.1016/j.ajhg.2018.11.005.
20. Thiffault, I., Farrow, E., Barrett, C., Scott, M., Ross, A., Means, J.C., Cheung, W.A., Johnson, A.F., Koseva, B., McLennan, R., et al. (2025). Clinical Long-Read Sequencing Test for Genetic Disease Diagnosis. JAMA Pediatr. *179*, 1355–1357. https://doi.org/10.1001/jamapediatrics.2025.3320.
21. Eisfeldt, J., Ameur, A., Lenner, F., Ten Berk de Boer, E., Ek, M., Wincent, J., Vaz, R., Ottosson, J., Jonson, T., Ivarsson, S., et al. (2024). A national long-read sequencing study on chromosomal rearrangements uncovers hidden complexities. Genome Res. *34*, 1774–1784. https://doi.org/10.1101/gr.279510.124.
22. Chen, X., Baker, D., Dolzhenko, E., Devaney, J.M., Noya, J., Berlyoung, A.S., Brandon, R., Hruska, K.S., Lochovsky, L., Kruszka, P., et al. (2025). Genome-wide profiling of highly similar paralogous genes using HiFi sequencing. Nat. Commun. *16*, 2340. https://doi.org/10.1038/s41467-025-57505-2.
23. Tewhey, R., Bansal, V., Torkamani, A., Topol, E.J., and Schork, N.J. (2011). The importance of phase information

for human genomics. Nat. Rev. Genet. *12*, 215–223. https://doi.org/10.1038/nrg2950.

24. Sims, D., Sudbery, I., Ilott, N.E., Heger, A., and Ponting, C.P. (2014). Sequencing depth and coverage: key considerations in genomic analyses. Nat. Rev. Genet. *15*, 121–132. https://doi.org/10.1038/nrg3642.
25. Kong, S.W., Lee, I.-H., Liu, X., Hirschhorn, J.N., and Mandl, K.D. (2018). Measuring Coverage and Accuracy of Whole Exome Sequencing in Clinical Context. Genet. Med. *20*, 1617–1626. https://doi.org/10.1038/gim.2018.51.
26. Sedlazeck, F.J., Rescheneder, P., Smolka, M., Fang, H., Nattestad, M., von Haeseler, A., and Schatz, M.C. (2018). Accurate detection of complex structural variations using single molecule sequencing. Nat. Methods *15*, 461–468. https://doi.org/10.1038/s41592-018-0001-7.
27. Hammond, N., Liao, L., Tong, P.W., Ng, Z., Nguyen, T.-M.P., Ho, C., Yang, Y., and Scott, S.A. (2025). Analytical validation of germline small variant detection using long-read HiFi genome sequencing. Genome Res. *35*, 1391–1399. https://doi.org/10.1101/gr.278836.123.
28. Steyaert, W., Sagath, L., Demidov, G., Yépez, V.A., Esteve-Codina, A., Gagneur, J., Ellwanger, K., Derks, R., Weiss, M., den Ouden, A., et al. (2025). Unraveling undiagnosed rare disease cases by HiFi long-read genome sequencing. Genome Res. *35*, 755–768. https://doi.org/10.1101/gr.279414.124.