npj | genomic medicine

Article

Published in partnership with CEGMR, King Abdulaziz University

https://doi.org/10.1038/s41525-024-00445-5

Benchmarking nanopore sequencing and rapid genomics feasibility: validation at a quaternary hospital in New Zealand

Denis M. Nyaga[1,8], Peter Tsai[1,2,8], Clare Gebbie[1,8], Hui Hui Phua[1,8], Patrick Yap[3,8], Polona Le Quesne Stabej[1,2], Sophie Farrow[1], Jing Rong[1], Gergely Toldi[1,4], Eric Thorstensen[1], Zornitza Stark[5,6], Sebastian Lunke[5,6], Kimberley Gamet[3], Jodi Van Dyk[1], Mark Greenslade[7] & Justin M. O'Sullivan[1] ✉

Approximately 200 critically ill infants and children in New Zealand are in high-dependency care, many suspected of having genetic conditions, requiring scalable genomic testing. We adopted an acute care genomics protocol from an accredited laboratory and established a clinical pipeline using Oxford Nanopore Technologies PromethION 2 solo system and Fabric GEM™ software. Benchmarking of the pipeline was performed using Global Alliance for Genomics and Health benchmarking tools and Genome in a Bottle samples (HG002-HG007). Evaluation of single nucleotide variants resulted in a precision and recall of 0.997 and 0.992, respectively. Small indel identification approached a precision of 0.922 and recall of 0.838. Large genomic variations from Coriell Copy Number Variation Reference Panel 1 were reliably detected with ~2 M long reads. Finally, we present results obtained from fourteen trio samples, ten of which were processed in parallel with a clinically accredited short-read rapid genomic testing pipeline (Newborn Genomics Programme; NCT06081075; 2023-10-12).

The rise of clinical genomic testing, utilizing exome and whole genome sequencing, has enabled the detection of genomic changes (i.e. single nucleotide variants [SNVs], small insertions and deletions [indels], copy number variants [CNVs], structural variants [SVs]), and elucidated the underlying genetic basis of rare Mendelian disorders and cancers[1–7]. Over the past decade, the increasing adoption of genomic testing has generated substantial evidence supporting precision medicine[2,4,6]. Such approaches have enabled molecular diagnoses for genetic disorders, guiding tailored medical interventions[6–9].

Genomic testing has been established using cost-effective, short-read (150 bp in fragments) sequencing platforms from Illumina (i.e. NovaSeq, HiSeq, MiSeq, NextSeq), Thermo Fisher (Ion Torrent sequencer), and BGI (i.e. BGISEQ and MGISEQ). However, short-read sequencing has well recognised limitations[10,11]. Firstly, it is difficult to uniquely align short reads to complex repetitive genomic regions involved in short tandem repeat (STR) expansion disorders (e.g. Fragile X syndrome and Huntington's disease)[10,12]. Secondly, the requirement for PCR amplification in short-read sequencing may contribute artefacts and hinder the identification of native base modifications[10]. Thirdly, short read lengths hinder the identification and precise phasing of alleles in large SVs[10].

The recently developed long-read sequencing (LRS) platforms (i.e. Oxford Nanopore Technologies [ONT] and Pacific Biosciences [PacBio]) employ direct inspection of single molecules during DNA synthesis, yielding long phaseable reads (>10 kb) in real-time[12–15]. Consequently, long reads generate highly reliable complete genome assemblies[15], which can serve as benchmarks for short-read data. The utilization of ONT long reads as a standalone sequencing platform in clinical diagnosis has been demonstrated[16–18]. In the research setting, LRS has been used to: a) identify and fine-map structural variations at single-nucleotide resolution; and b) resolve the haplotypes of heterozygous SVs[13,14,19]. Novel pathogenic variants have been uncovered by LRS technology in human diseases with a previously unknown underlying genetic cause[20]. Additionally, long reads facilitate the characterization of pathogenic repeat expansions in genomic regions that are challenging to sequence using short-read sequencing technology[21,22].

The clinical application of LRS[13–22] requires confidence in the accuracy of variant calling for SVs, CNVs, STRs, and SNVs/indels. However, high

[1]Liggins Institute, The University of Auckland, Auckland, New Zealand. [2]Molecular Medicine and Pathology, The University of Auckland, Auckland, New Zealand. [3]Genetic Health Service New Zealand-Northern Hub, Te Toka Tumai, Auckland, New Zealand. [4]Starship Child Health, Te Whatu Ora Te Toka Tumai, Auckland, New Zealand. [5]Victorian Clinical Genetics Services, Murdoch Children's Research Institute, Parkville, Melbourne, Australia. [6]Faculty of Medicine, Dentistry and Health Sciences, The University of Melbourne, Melbourne, Australia. [7]Diagnostic Genetics, Department of Pathology and Laboratory Medicine, Te Toka Tumai, Auckland, New Zealand. [8]These authors contributed equally: Denis M. Nyaga, Peter Tsai, Clare Gebbie, Hui Hui Phua, Patrick Yap. ✉e-mail: justin.osullivan@auckland.ac.nz

per-base error rates in low-complexity and homopolymer sequences[12,23], and other issues have led to concerns about the application of ONT in clinical settings. Thus, there is a need for comprehensive benchmarking to: 1) confirm precision relative to the routinely used short-read technologies; and 2) illustrate the benefits and limitations of LRS technology for application to CGS.

The Global Alliance for Genomics and Health (GA4GH) developed benchmarking protocols to evaluate the performance of sequencing platforms and variant-calling methods before their integration into clinical practice[24–26]. Benchmarking is essential to ensure adherence to standards and relies upon datasets where the relationship between input and output is known. This facilitates testing of consistency between the expected and observed outcomes (true positives)[24]. The Genome in a Bottle (GIAB) consortium offers reference samples (e.g. HG001 from the HapMap project and trios of Ashkenazi Jewish and Han Chinese ancestry from the Personal Genome Project) with ground-truth calls for SNVs, small indels, and SVs[27,28]. Notably, GIAB recently provided a curated benchmark of challenging medically relevant genes through haplotype-resolved whole-genome assembly[29]. The GA4GH resources enable performance assessment, optimization, and analytical validation of CGS assays and workflows for detecting genomic variations[24,25]. Indeed, GIAB datasets and benchmarks are considered the gold standard for evaluating sequencing technologies and variant calling pipelines[27].

The ONT platform generates sequencing data in real-time, allowing samples to be distributed across flow cells to reduce the sequencing time, where each additional flow cell reduces the sequencing time needed on a sample by 1/n (where n is the number of flow cells). Notably, this has been demonstrated in a recent study that sequenced a single human genome across 48 flow cells, generating high-depth genome-wide data (200 Gigabases) and candidate variant identification in less than eight hours[30]. The ONT platform is also capable of targeted sequencing through adaptive sampling, which removes the need to design custom probes to capture genes or regions of interest through a dynamic and modifiable process during the sequencing run[31]. DNA and RNA base modifications, including 5-methylcytosine (5mC), 5-hydroxymethylcytosine (5hmC), and N^6-methyladenine (6 mA), can also be detected computationally on raw ONT data without the need to perform special library preparations such as bisulphite conversion[32], which is known to cause DNA damage and can lead to overestimation of the 5mC level[33].

We have established an expandable rapid genomic testing pipeline based around the ONT PromethION2 (P2) solo system connected to AI-driven genomics analysis and interpretation software (i.e. Fabric GEM™ software) for tertiary analysis. In the establishment phase, we benchmarked our pipeline using GAG4H tools and GIAB reference cell lines HG002 - HG007 for SNVs and small indels analysis. In addition, we used CNVPANEL01 (Coriell Institute) to measure our ability to detect large-scale chromosomal abnormalities. Finally, we present the results of the pipeline validation phase, performed in parallel with a clinically accredited short-read rapid genomic testing service.

Methods

Benchmarking samples and truth sets

We acquired CNVPANEL01 as 3 µg genomic DNA (at 100 µg/ml) per sample and GIAB reference samples (i.e. HG002 - HG007) with available truth sets, from the Coriell repository (Coriell Institute for Medical Research, 403 Haddon Avenue Camden, NJ 08103, USA).

Library preparation and nanopore sequencing

DNA samples (1500 ng) were sheared to 10–15 kb using Covaris g-TUBES (Covaris) in a bench-top centrifuge for 1 min at 2000 RCF (room temperature). Nanopore sequencing libraries were prepared according to the genomic DNA Ligation Sequencing Kit V14 (SQK-LSK114) protocol (ONT, Oxford Science Park, OX4 4DQ, UK). Prepared libraries were loaded on PromethION flow cells (R10.4) and sequenced (i.e. depth of between 24-42X) with the PromethION 2 (P2) solo device using Kit 14 chemistry and MinKNOW v23.07.8 (Oxford Nanopore Technologies [ONT], Oxford Science Park, OX4 4DQ, UK).

Base calling of nanopore reads and variant calling

Base calling of raw ONT signal data was completed using Dorado v0.3.3 (https://github.com/nanoporetech/dorado) with the high accuracy (hac) model (dna_r10.4.1_e8.2_400bps_hac@v4.2.0). In addition, base calling of the HG002 sample was also completed with the super accuracy (sup) model (dna_r10.4.1_e8.2_400bps_sup@v4.2.0). The resulting FASTQ files, with a Phred quality score (Q score) > 9, in the fastq_pass folder, were processed with EPI2ME Labs' wf-alignment pipeline (https://github.com/epi2me-labs/wf-alignment; v0.5.2). Briefly, FASTQ files were aligned to the GRCh38 reference genome using minimap2 (v2.26)[34]. EPI2ME Labs' wf-human-variation pipeline (https://github.com/epi2me-labs/wf-human-variation; v1.7.0) was subsequently employed for genomic variant processing, including SNV and small indel calling with Clair3 (v1.0.4)[35], SV calling with Sniffles2 (v2.2)[36], and CNV calling with QDNAseq (v1.38)[37] using default parameters, with a VNTR annotation file provided for accurate SV identification. Repeat expansions were genotyped using Straglr (https://github.com/philres/straglr)[38] as implemented in EPI2ME Labs' wf-human-variation pipeline v1.7.0.

Benchmarking of variant calling

Variant comparison tools (https://github.com/ga4gh/benchmarking-tools)[24] are integral to genomic benchmarking as they identify shared variations between ground-truth calls and comparison results (i.e., true positives [TP]), along with variants unique to each set (i.e., false negatives [FN]), and additional variants (i.e., false positives [FP]). We compared called SNVs and small indels with GIAB ground-truth variants (benchmark version v4.2.1)[24] using hap.py v0.3.15 (https://github.com/Illumina/hap.py), and each variant was labelled as TP, FP, or FN. Hap.py also provides precision (positive predictive value [PPV]), recall (sensitivity) and F1 scores (harmonic mean of precision and recall) calculated as follows:

$$Precision = True\ Positives/(True\ Positives + False\ Positives) \quad (1)$$

$$Recall = True\ Positives/(True\ Positives + False\ Negatives) \quad (2)$$

$$F1\ score = \frac{2\,x(Precisionx\ Recall)}{(Precision + Recall)} \quad (3)$$

For SVs, we employed Truvari v4.1.0 (https://github.com/ACEnglish/truvari)[39] to benchmark variants with GIAB ground-truth SVs. Each variant was categorized as TP, FP, or FN based on this comparison.

Rarefaction benchmarking analysis

Rarefaction was performed to evaluate the sensitivity and reliability of long-read variant calling across different sequencing depths. Subsampling of the Binary Alignment Map (BAM) files was performed using Samtools[40], by randomly selecting subsets of reads from the original alignment files. The subsampled BAMs were then subjected to variant calling analysis as described in the variant calling section. Benchmarking for SNVs and small indels was conducted as detailed in the benchmarking section. Rarefaction curves were generated using python v3.10.8 and the seaborn v0.12.2 library to illustrate the relationship between sequencing depth and the called variants, enabling the evaluation of variant calling performance and reliability across varied sequencing depths.

Benchmarking analysis for challenging clinically relevant genes

We called SNVs and small indels across genomic regions overlapping challenging clinically relevant genes[29] using the original BAM files and pipeline outlined in the variant calling section. Benchmarking for SNVs and small indels was conducted as detailed the benchmarking section.

Methylation analysis

Raw ONT signal data in POD5 files (https://github.com/nanoporetech/pod5-file-format) was base called (Dorado v0.5.0) using the high accuracy (hac) DNA base modification model (dna_r10.4.1_e8.2_400bps_hac@v4.2.0_5mCG_5hmCG@v2) to detect modified bases (i.e. 5-methylcytosine [5mC], 5-hydroxymethylcytosine [5hmC]). The modified BAM files (modBAMs) were aligned to the GRCh38 reference genome and modkit v0.2.3 (https://github.com/nanoporetech/modkit) employed to generate genome-wide summary counts of modified and unmodified bases into bedMethyl files. Haplotype-specific 5mC differentially methylated regions (DMRs) in the HG002 genome were identified using ont-methylDMR-kit (https://github.com/NyagaM/ont-methylDMR-kit), which utilizes the Bioconductor DSS (Dispersion Shrinkage for Sequencing) package[41]. We used the following DSS[41] parameters for calling DMRs: delta (threshold for defining DMRs) at 10%, *p*-value at < 0.01; minimum DMR length of 100 bps, and at least 10 CpG sites per DMR (https://github.com/NyagaM/ont-methylDMR-kit). The pipeline supports haplotype-specific analysis, DMR detection between two samples, group methyl analysis, as well as genomic annotation of significant DMRs (https://github.com/NyagaM/ont-methylDMR-kit).

Benchmarking results visualization

Plots were generated using the seaborn v0.12.2 and matplotlib v3.7.1, and python v3.10.8.

Newborn Genomics Programme (NBG) study design

This is a research study to determine the medical and economic impacts of rapid whole genome sequencing (rWGS) within the New Zealand health care landscape. Ethics approval was obtained from the Northern B Health and Disability Ethics Committee for the study entitled: *Newborn Genomics – Te Ira oo Te Arai* (Ethics reference: 2023 FULL 15542). Locality approval was obtained from the Research Review Committee Te Toka Tumai Auckland for the project entitled: *Newborn Genomics – Te Ira oo Te Arai* (Reference A + 9855 [FULL 15542]). This study is registered in ClinicalTrials.gov (Newborn Genomics Programme; NCT06081075; 2023-10-12). The clinical protocol was adopted and modified as per Lunke et al., 2023[42].

NBG study participants and recruitment criteria

Children with suspected genetic conditions and their families were recruited into the study from the neonatal and paediatric intensive care units (i.e. NICU and PICU, respectively) and the National Metabolic Service at Te Toka Tumai | Auckland City Hospital (New Zealand) between November 2023 and August 2024. Within NICU, participation was limited to proband-parent trios of critically sick neonates with evidence of a suspected genetic condition, without a clear non-genetic aetiology, or who developed an abnormal response to standard therapy for an underlying condition within the preceding seven days. For infants within PICU or under the care of Metabolic Services, participation was limited to proband/parent trios of children with an acute or chronic illness with evidence of a suspected genetic condition without a clear non-genetic aetiology.

All participants continued to receive the standard of care, irrespective of whether they were included in the study.

Potential participants were referred to the geneticist on-call (by telephone) for a formal genetic review, mainly by a neonatologist or a paediatric intensivist or the lead paediatric subspecialist for the patient when a genetic condition was suspected, or when the aetiology of a condition was unclear and a genetic cause needed to be ruled out to guide further clinical management.

Inclusion and exclusion criteria were modified from Dimmock et al.[43] and McKeown et al.[44].

The inclusion criteria was:

- acutely ill inpatient
- admitted to NICU or PICU between April 2023 – March 2026
- under the care of the National Metabolic Service between April 2023 and March 2026
- within 1 week of hospitalization or within 1 week of developing abnormal response to standard therapy for an underlying condition
- suspected genetic condition, without a clear non-genetic aetiology

The exclusion criteria was:

- patients whose clinical course is entirely explained by
  - isolated prematurity
  - isolated unconjugated hyperbilirubinemia
  - infection or sepsis with expected response to therapy
  - a previously confirmed genetic diagnosis that explains the clinical condition
  - isolated transient neonatal tachypnoea
  - meconium aspiration syndrome
  - trauma
- inability to source blood or buccal samples for DNA extraction from at least the mother and child

Of note, participants were only considered for the study if they were referred to clinical genetics as a part of their standard of care workup.

Following the referral of potential participants to the study, a multidisciplinary meeting (MDM) was convened via video conference to evaluate the eligibility of the referral based on the study's inclusion and exclusion criteria. At a minimum, the MDM was comprised of a clinical geneticist, genetic counsellor, principal investigator, project manager, representative from the genomic analytical team (bioinformatician, variant curator) and the referring clinician. After agreeing to participate, the patients were registered on RedCap, and a study reference was generated. Subsequently, the clinical geneticists and genetic counsellors completed the clinical information, including phenotypic characterization using HPO terms, and facilitated informed consent.

Parents or guardians of the proposed probands were informed of the details of the study using the HDEC-approved Newborn Genomics Programme Participant Information Sheet (Supplementary Note 1), and had the opportunity to ask questions to an on-call geneticist and genetic counsellor from the Genetic Health Service New Zealand. Written informed consent was obtained from parents or guardians before any study-specific processes were undertaken (Newborn Genomics Programme consent form; Supplementary Note 2).

Clinical geneticists and subspecialists performed clinical phenotyping, which was recorded on RedCap using the Human Phenotype Ontology (HPO) terms (https://hpo.jax.org/app/) to optimize phenotypic data exchange during the curation stages of the analysis. At the same time, a phenotype-focus gene list was generated using PanelApp (Australia [https://panelapp.agha.umccr.org/] and the UK [https://panelapp.genomicsengland.co.uk/]) and shared with the genomic analytical team for inclusion in the Bayesian AI-based clinical decision support tool (Fabric GEM™ software).

NBG sample collection, DNA extraction, library preparation, sequencing, and variant calling

After obtaining consent, duplicate blood samples were collected: 4 mL EDTA blood samples from the mother and father, and 500 µL EDTA blood samples from the child. One set of samples was sent to the Liggins Institute newborn genomics laboratory for sequencing and variant analysis, while the second set was sent to the clinical laboratory, Victorian Clinical Genetics Services (VCGS) in Melbourne, Australia, for a concurrent, independent, short-read-based analysis as described in Lunke et al. 2023[42].

High molecular weight DNA was extracted from 300 µl of the whole blood using the Puregene DNA extraction Kit (Qiagen) following the manufacturer's protocol, and the extracted DNA eluted in nuclease-free water (Thermo Fisher Scientific). The quantification and purity assessment of the DNA samples were performed using the Qubit system (Thermo Fisher Scientific) and a spectrophotometer (Implen NanoPhotometer). The library preparation and sequencing procedures were carried out as detailed in the library preparation section. Finally, the base calling of sequenced reads and variant calling analysis was conducted following the methods described

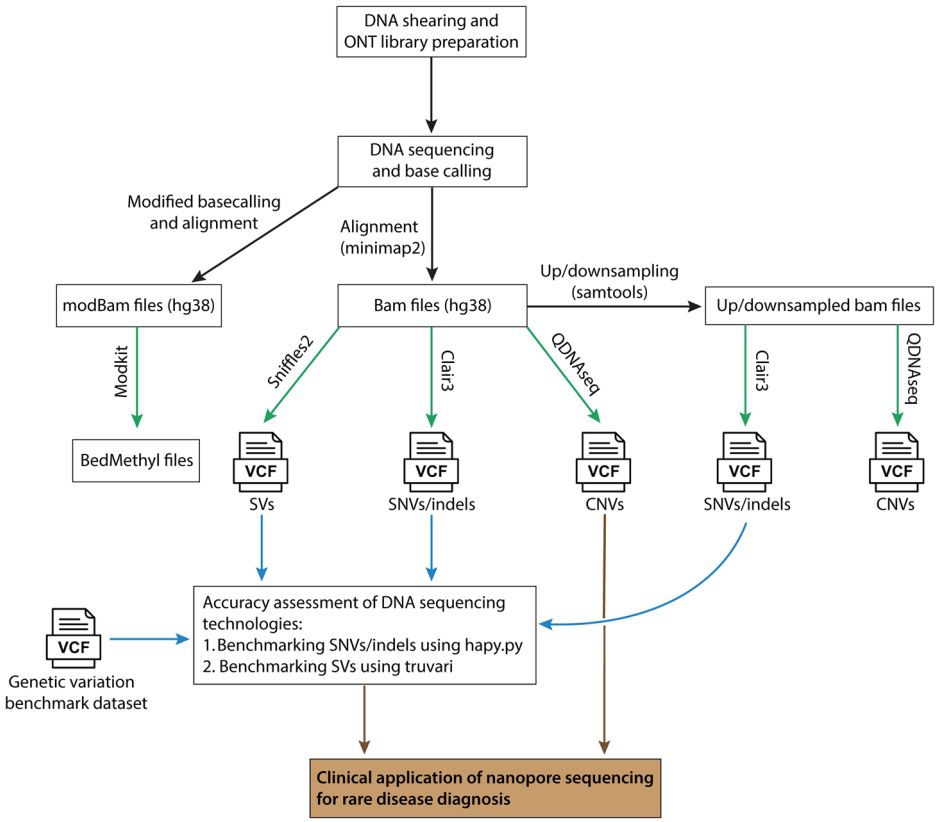

Fig. 1 | An overview of the genomic benchmarking for the acute care clinical pipeline. Genomic DNA sequencing and variant calling was performed using Genome in a Bottle cell lines (HG002 - HG007) and Coriell Institute for Medical Research CNVPA-NEL01 cell lines. Nanopore sequencing libraries were prepared and sequenced using PromethION flow cells (R10.4). Variant calling was performed using EPI2ME pipeline that includes: **a** clair3 for single nucleotide variants and small indels analysis, **b** sniffles2 for structural variant calling; and (**c**) QDNAseq for copy number variant analysis. Single nucleotide variants and small indels were benchmarked against Genome in a Bottle ground-truth variants using hap.py v0.3.15 (NIST v4.2.1), and Truvari v4.1.0 was employed to benchmark structural variants (NIST v0.6). Modified base calling and alignment to GRCh38 reference genome was performed using dorado v.5.0 and genome-wide summary counts of 5-methylcytosine (5mC) and 5-hydroxymethylcytosine (5hmC) were generated using modkit v0.2.3. The figure was created using Adobe Illustrator.

in the base calling and variant calling sections. Of note, we have developed a Standard Operating Procedure (SOP) to lock pipelines to specific versions and outline procedures for updates and upgrades since the software for base calling and variant analysis is frequently updated (Supplementary Note 3).

NBG candidate variant prioritization and genomic results reporting

For variant filtering and prioritisation, we used Fabric GEM™ as the primary interpretation platform and QCI Clinical Insights Interpret-translational (QIAGEN Inc., https://digitalinsights.qiagen.com/) as the secondary 'research-confirmatory' platform. We retained variants with ≥ 10 reads, a Variant Allele Frequency (VAF) ≥ 20% in the proband, a frequency ≤ 1% in gnomAD v3.1 and those located in the exonic regions or within +/−20 bases of exon/intron boundaries. Additionally, intronic variants beyond ±20 bases from exon start/end predicted to affect splicing by MaxEntScan were also retained.

In the initial analysis, we focused on variants in candidate genes, HPOs and panels (PanelApp Australia[45]) suggested by clinicians. If no pathogenic or likely pathogenic variants were identified (based on the American College of Medical Genetics and Genomics [ACMG] guidelines[46]), we expanded the analysis to all genes. Genes in the incidentalome (PanelApp Australia Version 0.308[45]) were excluded, unless they were relevant to the patient's phenotype as indicated by the clinician. In so doing, we adopt a judicious approach to the reporting of variants of uncertain significance (VUS) in the acute care setting, only including those that are deemed related to the patient's phenotype and are typically very close to being classified as 'likely pathogenic' according to ACMG criteria[46]. This stratification of VUS is recommended by the Association for Clinical Genomic Science (ACGS) in the UK.

A multidisciplinary review meeting (MDM) was then held to evaluate the results. The review MDM comprised the same clinicians and study representatives who attended the recruitment MDM. During the meeting, the genomic data analysts presented the quality control report and discussed the prioritized variants, and the evidence for pathogenic or likely pathogenic variants, for genotype-phenotype correlation. The VCGS results were not shared with the NBG team, ensuring they remained blinded to the clinically validated results until the variant review MDM. Finally, the clinical geneticist and genetic counsellor disclosed and discussed the molecular diagnosis based on the accredited acute care genomics service (VCGS) results.

Upon completing the variant review meeting, the NBG study team generated a research report. Simultaneously, the clinical laboratory (VCGS) produced its validated clinical report, which was directly returned to the clinical team for disclosure to the families. Finally, the genetic counsellor communicated the study report findings to the study participants and addressed any discrepancies identified in the reports.

Ethics statement

This study was performed in line with the principles of the Declaration of Helsinki. Ethics approval was obtained from the Northern B Health and Disability Ethics Committee for the study entitled: *Newborn Genomics – Te Ira oo Te Arai* (Ethics reference: 2023 FULL 15542). Locality approval was obtained from the Research Review Committee Te Toka Tumai Auckland for the project entitled: *Newborn Genomics – Te Ira oo Te Arai* (Reference A + 9855 [FULL 15542]).

Patient consent statement

Parents of the participating newborns provided written informed consent.

Results

Overview of genomic benchmarking workflow for acute care clinical pipeline

We have performed genomic variant benchmarking of an expandable acute care clinical pipeline, using the set standards and guidelines provided by GAG4H[24]. DNA samples from the Coriell repository, including the characterized GIAB reference samples (HG002 - HG007) with available truth sets, were used for benchmarking variant calling of SNVs (i.e. single base substitutions) and small indels (i.e., insertions and deletions < 50 bps) (Fig. 1). Sample HG002 was used to benchmark SVs (i.e., genomic alteration >50 bps encompassing insertions, deletions, duplications, inversions, and

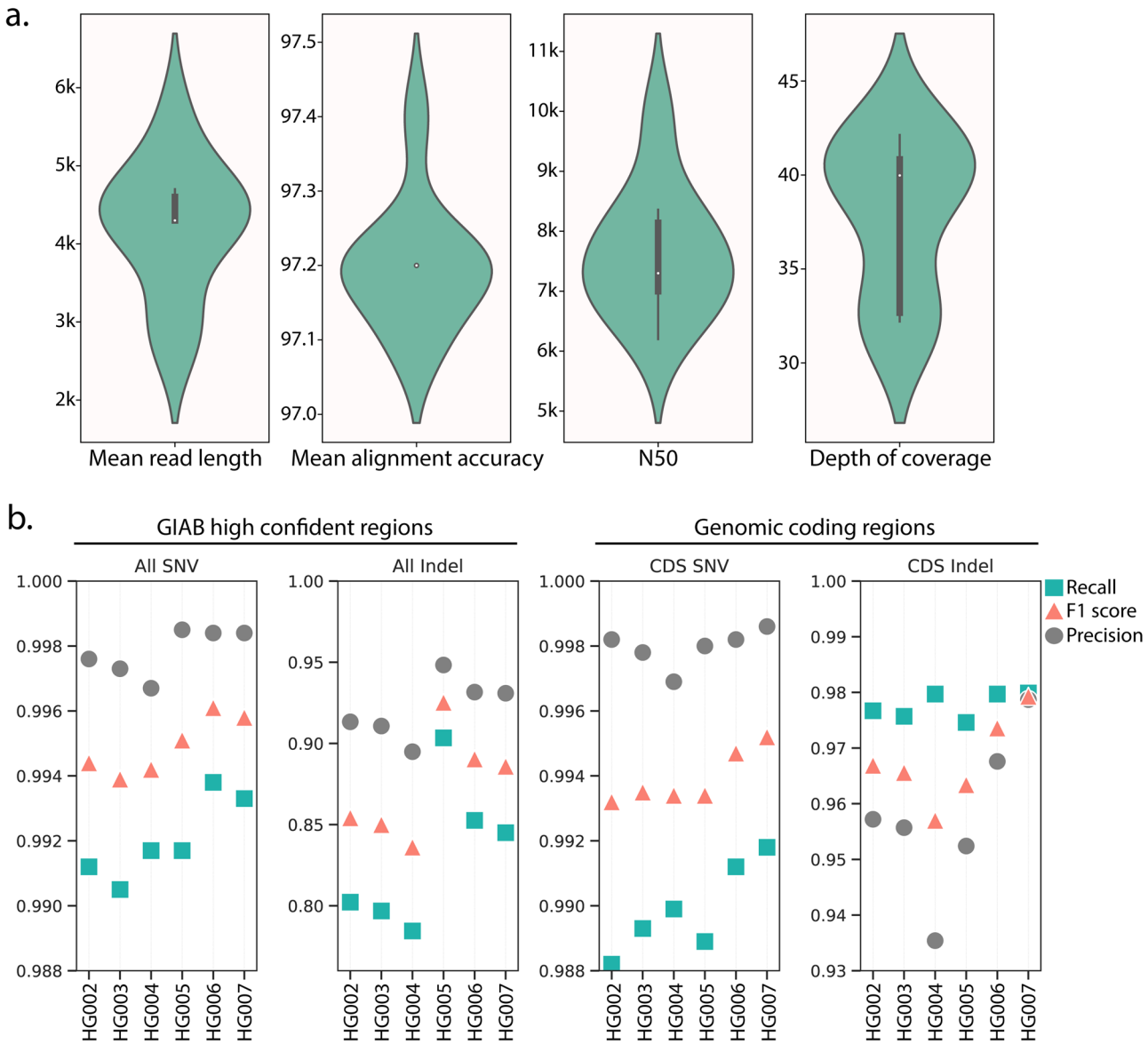

Fig. 2 | Sequencing metrics and variant calling accuracy ONT summary statistics. **a** Violin plots showing mean read length (kb), mean alignment accuracy (%), N50 (kb), and mean depth of coverage for Genome in a Bottle samples. Plots showing the relationship between read length and average per sequence quality scores (Q scores) for these samples are available as Supplementary Figs. 1–12. **b** Comparisons of precision (positive predictive value), recall (sensitivity) and F1 scores (harmonic mean of precision and recall) for single nucleotide variants and small indels called from nanopore sequencing data compared to Genome in a Bottle high-confidence regions (left) and coding regions excluding homopolymers and difficult-to-map genomic regions (right) for Genome in a Bottle samples HG002-HG007. A summary of benchmarking metrics across two separate sequencing runs for HG002-HG007 samples is available in Supplementary Table 1. CDS: coding sequence; N50: the length of the shortest read among the longest sequences, encompassing ~50% of the total nucleotides in a set of sequences.

translocations). Coriell samples carrying pathogenic variants (i.e., GM06936, GM06870, GM01416, GM20556, GM09367, GM05966, GM05067, GM09216) were used to evaluate the performance of long reads in the identification of large-scale chromosomal abnormalities (Fig. 1).

ONT long reads provide high precision for small variant calls

Following sequencing using the R10.4 pore, base calling, and variant calling, we obtained mean read lengths of 4.2 kb, mean alignment accuracy of 97.2%, read N50 of 7.6 kb, and average depth of coverage of 37.9X for six samples (Fig. 2a). Benchmarking was performed to evaluate the general performance of ONT reads on SNVs and small indels (up to 50 bps) calling from GIAB samples HG002 – HG007 (see Methods). Precision, recall, and F1 scores were computed against truth sets (National Institute of Standards and Technology [NIST)] benchmark v4.2.1; https://github.com/ga4gh/benchmarking-tools/blob/master/resources/high-confidence-sets/giab.md), for: 1) high-confidence regions excluding homopolymers, defined as four or more consecutive identical nucleotides ±1 base pair on each side; 2) genome-wide coding regions; and 3) 273 challenging medically relevant genes for the HG002 genome (CMRG v1.0; https://data.nist.gov/od/id/mds2-2475).

Across two separate sequencing runs, the average SNV precision and recall were 0.998 and 0.992), respectively, while small indel precision and recall were 0.922 and 0.831, respectively, within GIAB high-confidence regions (Fig. 2b; Supplementary Table 1; Supplementary Table 2). When assessing variants exclusively within coding regions and regions excluding homopolymers and difficult-to-map genomic regions, SNVs achieved an F1 score of 0.994, and small indels reached 0.968 (Fig. 2b; Supplementary Table 3; Supplementary Table 4). Furthermore, we assessed the performance of long reads on identifying variants in CMRG. ONT LRS demonstrated precision and recall scores > 0.967 and > 0.978, respectively, for SNVs and

Table 1 | Benchmarking metrics for SNVs and small indels within CMRG

GIAB Sample	Run	SNV				INDEL			
		Truth Total	Recall	Precision	F1	Truth Total	Recall	Precision	F1
HG002	1	17,559	0.979	0.967	0.973	3605	0.704	0.84	0.766
	2	17,559	0.978	0.969	0.974	3605	0.701	0.836	0.763

TP True positive, *FP* False positive, *FN* False negative. Precision (positive predictive value), recall (sensitivity) and F1 scores (harmonic mean of precision and recall).

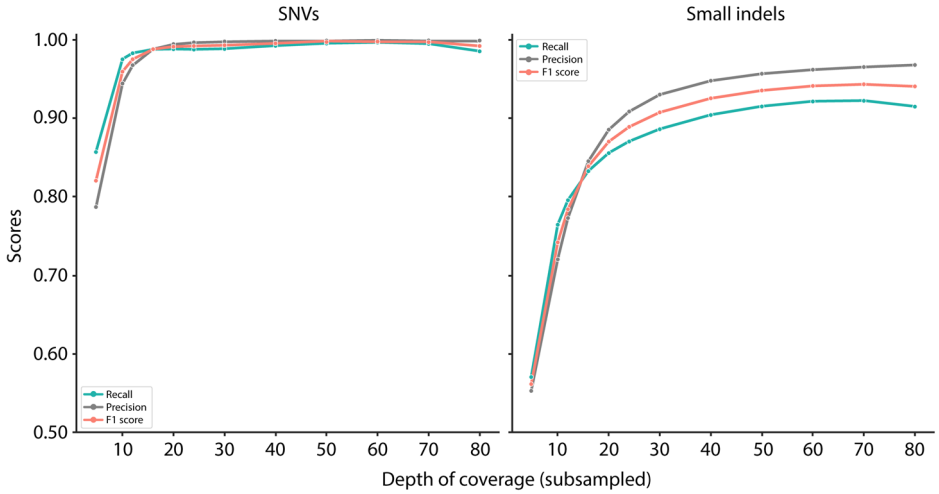

Fig. 3 | Sequencing depth exceeding 40X does not improve single nucleotide variant or small indels detection. Rarefaction analysis of subsampled HG005 BAM files reveal consistent F1 scores for single nucleotide variant at 40X and 80X, with scores at 0.995 and 0.992, respectively (left panel). Similarly, F1 scores for small indels (<50 bp) vary by <2% (i.e. 0.925 and 0.941) at depths of 40X and 80X, respectively (right panel).

Table 2 | Performance metrics for identification and genotyping of high-confident SVs

Sample	SV					
	TP	FN	FP	Recall	Precision	F1 score
HG002 run 1	9094	547	455	0.943	0.952	0.948
HG002 run 2	9234	407	474	0.958	0.951	0.954
ONT public data	9396	245	486	0.975	0.951	0.963

Comparison of HG002 genome-wide structural variants (SVs) with high-confidence Genome in a Bottle ground-truth SVs across two separate sequencing runs indicate that our established pipeline reliably and accurately identifies SVs. For this comparison, the HG002 genome was aligned to the GRCh37. ONT public data was accessed from https://labs.epi2me.io/giab-2023.05.

>0.836 and >0.701, respectively, for small indels within the 273 genes in the CMRG set (Table 1). We observed slightly improved F1 scores for small indels called from the super accuracy (sup) base called HG002 genome (i.e., F1 score = +0.01) for high-confidence regions; and an F1 score of 0.776 for the CMRG set (Supplementary Table 5; Supplementary Table 6). Overall, these results are consistent with previous benchmarking reports[47] and validate the efficacy of the EPI2ME Labs' implementation of Clair3[35] in generating high-quality small variant calls comparable to gold-standard results.

Increasing sequencing depth beyond 40X does not improve small variant detection

Sequencing depth has been identified as being critical for the accurate identification of variants for the diagnosis of genetic diseases[48]. We determined the optimal genomic depth for precise small variant identification from ONT reads. We downsampled the HG005 alignment file (~ 40X) by randomly extracting sets of reads (i.e. at proportions of 0.12, 0.25, 0.3, 0.4, 0.5, 0.6, and 0.75 of the total set) to simulate sequence data of the same sample at sequencing depths of 4.8X, 10X, 12X, 16X, 20X, 24X, and 30X. Upsampling was performed at proportions of 1.25, 1.5, 1.75, and 2 to mimic depths of 50X, 60X, 70X, and 80X. Our analysis revealed that beyond a sequencing depth of 40X, there were no significant improvements in the detection of SNVs and small indels (Fig. 3). These findings indicate that ~40X is the optimal depth for accurate small variant discovery from ONT reads, and additional depth beyond this threshold does not enhance the accuracy or sensitivity of small variant detection.

ONT reads accurately identify structural variations and haplotype-specific tandem repeats

SVs were identified from the HG002 genome using Sniffles2 with a tandem repeat bed file provided to improve SV calling in repetitive regions (Methods)[36]. SV calling performance from two separate sequencing runs, together with publicly available ONT SVs (https://labs.epi2me.io/giab-2023.05/), were benchmarked against the publicly available high-confidence GIAB ground-truth SVs (from GRCh37 reference genome) and SVs called within the CMRG genes using Truvari v4.1.0. ONT reads demonstrated high precision (> 0.951) and recall (> 0.943) for high-confident regions (Table 2). Additionally, we observed precision and recall metrics of > 0.889 and 0.946, respectively, for SVs within CMRG genes (Table 3), consistent with published benchmarks on ONT long reads[47].

We profiled genomic regions associated with repeat expansions since these regions contribute to the development of numerous neurodevelopmental disorders (NDDs) (e.g. congenital and childhood-onset myotonic dystrophy type 1[49]). Using Straglr[38], we genotyped and quantified 37

clinically relevant tandem repeat regions (including *DMPK* and *NOTCH2NLC*) in HG001, HG002, HG003, and HG004 genomes. Our findings indicate that long reads enable the genotyping of these regions (Supplementary Figs. 13–16), while also providing haplotype-specific information for the repeat elements (Supplementary Fig. 17).

Accurate identification of copy number variation at 2X sequencing depth

CNVs occur in neonatal disorders (e.g., 22q11.2 deletion syndrome[50]). In clinical testing, whole genome sequencing is poised to replace chromosomal microarray for the detection of CNVs[48]. As such, the clinical utility of long-read sequencing in identifying CNVs has been demonstrated[17]. We assessed whether our variant calling workflow accurately detects clinically relevant pathogenic large chromosomal aberrations from sixteen Coriell samples (Table 4; Supplementary Figs. 18–33). Benchmarking results confirmed the reliable and accurate detection of these pathogenic CNVs using long-read sequencing (Table 4). Notably, we successfully identified two CNVs at ~2X coverage by downsampling the BAM files to ~2.6 M reads: 1) the isodicentric chromosome CNV (i.e., 47,XY,+idic(15)(q13).ish idic(15)(q13)(D15Z1 + +,D15S11 + +,GABRB3 + +).arr Yq11.223q11.23(23920264-27079691)x2,15q11.1q13.3(18276329-30557740)x4); and 2) the XXXX syndrome CNV (i.e., 48,XXXX) (Supplementary Fig. 34). Collectively, these results underscore the potential for long-read genomic testing in neonatal intensive care for the diagnosis of suspected genetic conditions resulting from large chromosomal events.

Table 3 | Performance metrics for identification and genotyping SVs within CMRG genes

Sample	SV					
	TP	FN	FP	Recall	Precision	F1 score
HG002 run 1	196	10	11	0.907	0.947	0.927
HG002 run 2	176	24	11	0.889	0.946	0.916
ONT public data	197	19	9	0.912	0.956	0.934

Comparison of HG002 structural variants (SVs) with SVs called within challenging medically relevant genes across two separate sequencing runs indicate that our established pipeline reliably and accurately identifies SVs in clinically relevant difficult regions of the genome. ONT public data was accessed from https://labs.epi2me.io/giab-2023.05.

High concordance between ONT methylation calls and bisulphite sequencing

DNA methylation (5mC) is implicated in the pathogenic mechanism of *FMR1*-related disorders (e.g. fragile X syndrome), with expanded alleles typically exhibiting promoter hypermethylation and silencing of *FMR1*[51]. As such, DNA methylation profiling is essential for complete genetic diagnosis of *FMR1*-related disorders[51]. Notably, ONT sequencing facilitates concurrent profiling and quantification of DNA methylation (5-methylcytosine [5mC], and 5-hydroxymethylcytosine [5hmC]). Genome-wide 5mC characterization of the HG002 genome using ONT sequencing identified 28.8 million CpG sites (98% of total GRCh38 CpG sites). We did not detect any high levels of 5hmC methylation across the genome. Comparing ONT 5mC calls to standard whole-genome bisulphite sequencing (WGBS) of the HG002 genome, acquired from the ONT open data repository (https://labs.epi2me.io/gm24385-5mc), identified a strong correlation ($r = 0.949$; Fig. 4a), consistent with highly accurate methylation calling. Notably, using the ont-methylDMR-kit pipeline (https://github.com/NyagaM/ont-methylDMR-kit), we identified haplotype-specific 5mC DMRs within gene promoters for imprinted genes (Fig. 4b, c), as well as in novel DMRs (Supplementary Fig. 35; Supplementary Data 1). This illustrates the potential utility of the haplotype-level resolution offered by ONT-based sequencing reads.

Application of the long-read pipeline in acute care genomic diagnosis

Our goal was to develop a scalable acute care genetic diagnostic pipeline by harnessing the capabilities of the ONT PromethION 2 solo system integrated with Fabric GEM™ (an AI-driven genomics analysis and inter-

Table 4 | Long reads accurately detect pathogenic CNVs from CNVPANEL01 samples

Coriell Catalog ID	Chr	Truth CNV variant (ISCN)	Truth CNV detected
GM20556	15	47,XY,+idic(15)(q13).ish idic(15)(q13)(D15Z1 + +,D15S11 + +,GABRB3 + +).arr Yq11.223q11.23(23920264-27079691)x2,15q11.1q13.3(18276329-30557740)x4	Yes
GM01416	X	48,XXXX	Yes
GM06870	18	47,XX,+i(18)(p10).arr[hg19] 18p11.32p11.1(11542-15401751)x4	Yes
GM06936	10	46,XX,del(10)(p13)[20]	Yes
GM05966	14	46,XY,dup(14)(q22q24).arr[hg19]14q22.2q24.3(54,953,370-76,136,883)x3	Yes
GM09367	6	46,XX,dup(6)(q21q24).ish dup(6)(q21q24)(wcp6 +).arr 6q21q24.2(107861056-143105847)x3	Yes
GM05067	9	47,XY, + 9,del(9)(q11)[20]	Yes
GM09216	2	46,XY,del(2)(pter>p25.1::p23.3>qter).ish del(2)(D2S447 +)	Yes
GM17867	X	47,XXY	Yes
GM20027	X	45,X.arr[hg19](1-22)x2,(X)x1	Yes
GM07945	20	46,XY,del(20)(pter>q13.1::q13.3>qter).ish del(20)(20QTEL14 +).arr cgh 20q13.1q13.3(CN_874692,SNP_A_2089662)x1	Yes
GM10925	7	46,XY,del(7)(p14p12).arr 7p14.1p11.2(38598541-54681998)x1	Yes
GM17942	22	46,XY[20].arr[GRCh37] 22q11.21(18748427_21611337)x1	Yes
GM21887	15	46,XX,del(15)(q11q13).ish del(15)(q11q13)(D15Z1 + ,SNRPN-,[D15S10/UBE]-,GABRB3-,PML +).arr 15q11.2q13.1(20224751-26500067)x1	Yes
GM20022	3	46,XY,dup(3)(q21q29).ish dup(3)(q21q29)(wcp3 + ,D3S4560 +).arr 3q22.2q29(136044785-197137370)x3	Yes
GM10636	X	46,X,dup(X)(p11.1p11.3)[17].arr[GRCh37] Xp11.4p11.1(39822172_58561918)x3	Yes

Pathogenic large chromosomal aberrations in sixteen CNVPANEL01 samples were identified using QDNAseq (v1.38)37 from the wf-human-variation pipeline (v1.7.0; https://github.com/epi2me-labs/wf-human-variation) with default parameters. QDNAseq plots of the CNVs are available in Supplementary Figs. 18–33. *Chr* Chromosome, *ISCN* International System for Human Cytogenomic Nomenclature.

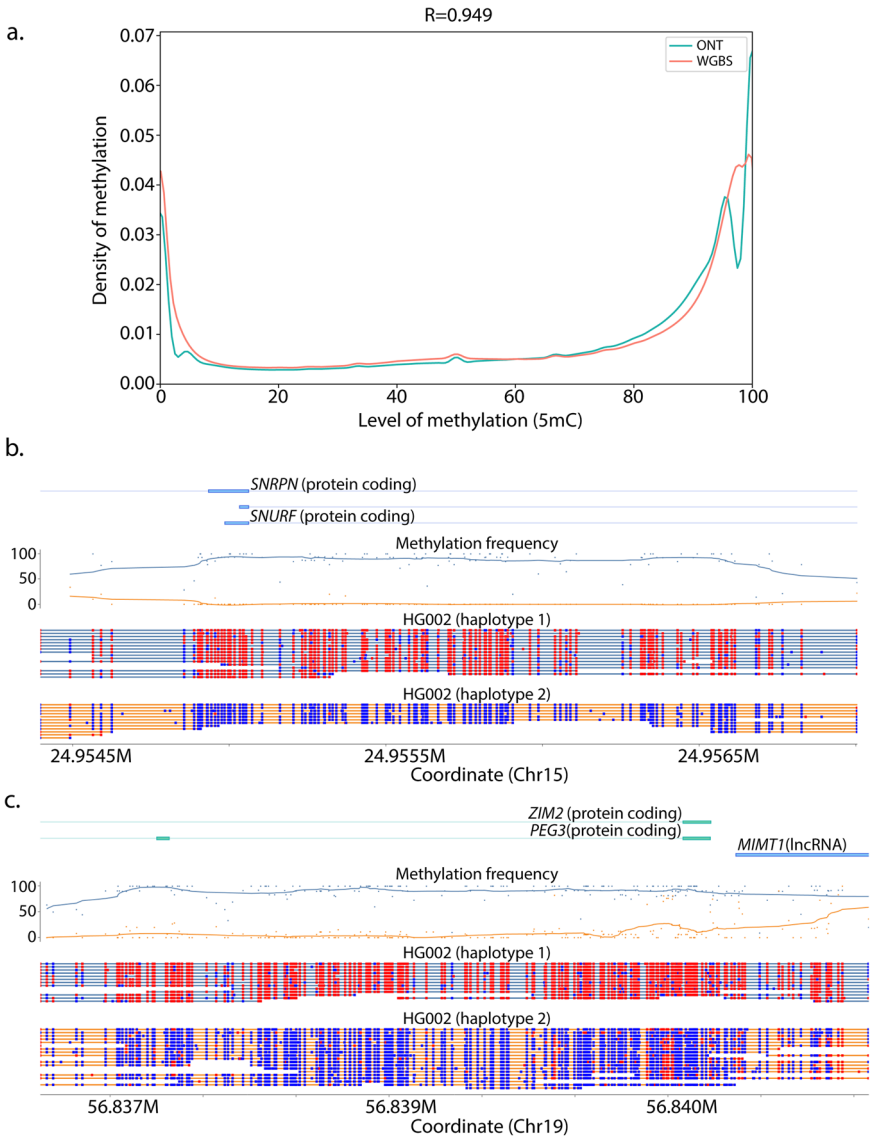

Fig. 4 | Strong correlation between 5mC methylation calls from nanopore sequencing and whole-genome bisulphite sequencing. a Density plots of the level of 5mC base modification across CpG sites detected through nanopore sequencing (green) and bisulphite sequencing (orange) from the HG002 sample. Pearson correlation analysis identified a strong correlation for methylation levels across CpG sites between the nanopore sequencing and whole genome bisulphite sequencing technologies. **b** *SNURF-SNRPN* and (**c**) *PEG3* loci exhibits haplotype-specific differential methylation, linked to maternal imprinting (paternal expressed allele) as detected by ont-methylDMR-kit pipeline (https://github.com/NyagaM/ont-methylDMR-kit) and visualized using modbamtools (https://github.com/rrazaghi/modbamtools). Methylated CpG sites are denoted in red, while unmethylated sites are represented in blue. A methylation frequency plot and gene loci are visualized above the haplotype reads. A summary of genome-wide haplotype-specific 5mC differentially methylated regions and their genomic annotations is available in Supplementary Data 1.

pretation software; https://fabricgenomics.com/). This integration was designed to facilitate precise genome annotation with rapid variant interpretation and prioritisation. The ultimate goal was to provide clinicians with access to actionable information pertaining to SNVs, small indels, and SVs (Fig. 5).

During the establishment phase of this pipeline, ten critically sick children in the neonatal and paediatric intensive care (NICU and PICU) at Te Toka Tumai/Starship Child Health were referred for rapid long-read genomic sequencing (Ethics: Approval from Health and Disability Ethics Committee [Reference 2023 FULL 15542]; Locality approval: A + 9855 [FULL15542]). In parallel, samples were provided to a clinically accredited genomics laboratory (VCGS; Melbourne, Australia) for rapid short-read Illumina genomic sequencing as described in Lunke et al. 2023[42]. Sequencing, genome variant curation and analysis were independently undertaken at each site. For the 30 samples, we obtained a mean read length of 8.2 kb, mean alignment accuracy of 96.8%, read N50 of 7.6 kb, and an average depth of coverage of 36X, with probands reaching 43.9X (Supplementary Fig. 36). Plots showing the relationship between read length and average per sequence quality scores (Q scores) for the 10 proband samples are available in Supplementary Figs. 37–46).

Genomic results from the accredited laboratory were independently provided to the clinician so that the long-read provider was unaware of the accredited results until after their results were presented to the MDM. Identical results were obtained across the ten proband-parent trios recruited into the programme. Six received a genomic diagnosis, while four did not (Table 5). The concordance of genomic findings in both positive and negative cases demonstrates the applicability and reliability of our pipeline for acute clinical care. Finally, four additional trios were sequenced (Supplementary Table 7). In one of these trios, we identified complex heterozygous genotypes in the *GBA* gene, consisting of a 55 bp deletion on one haplotype and a missense variant on the other (Supplementary Fig. 47). This highlights the potential of the long-read system to accurately assign variants to the correct haplotype and improve diagnostic yield, a benefit we anticipate will become apparent with larger sample sizes.

Discussion

Rapid genomic diagnosis offers potential benefits for critically sick patients that include guiding clinical management and improving prognosis[16,30,52]. Although long-read sequencing technology is opening new opportunities for rapid diagnosis and treatment of rare genetic disorders, its adoption in clinical settings has been limited. This is despite evidence suggesting that integrating long reads increases genetic testing capabilities beyond SNVs and small indels to include SVs, CNVs, and STRs[13–22], as well as the possibility of moving testing closer to point-of care[53]. The underutilization of this

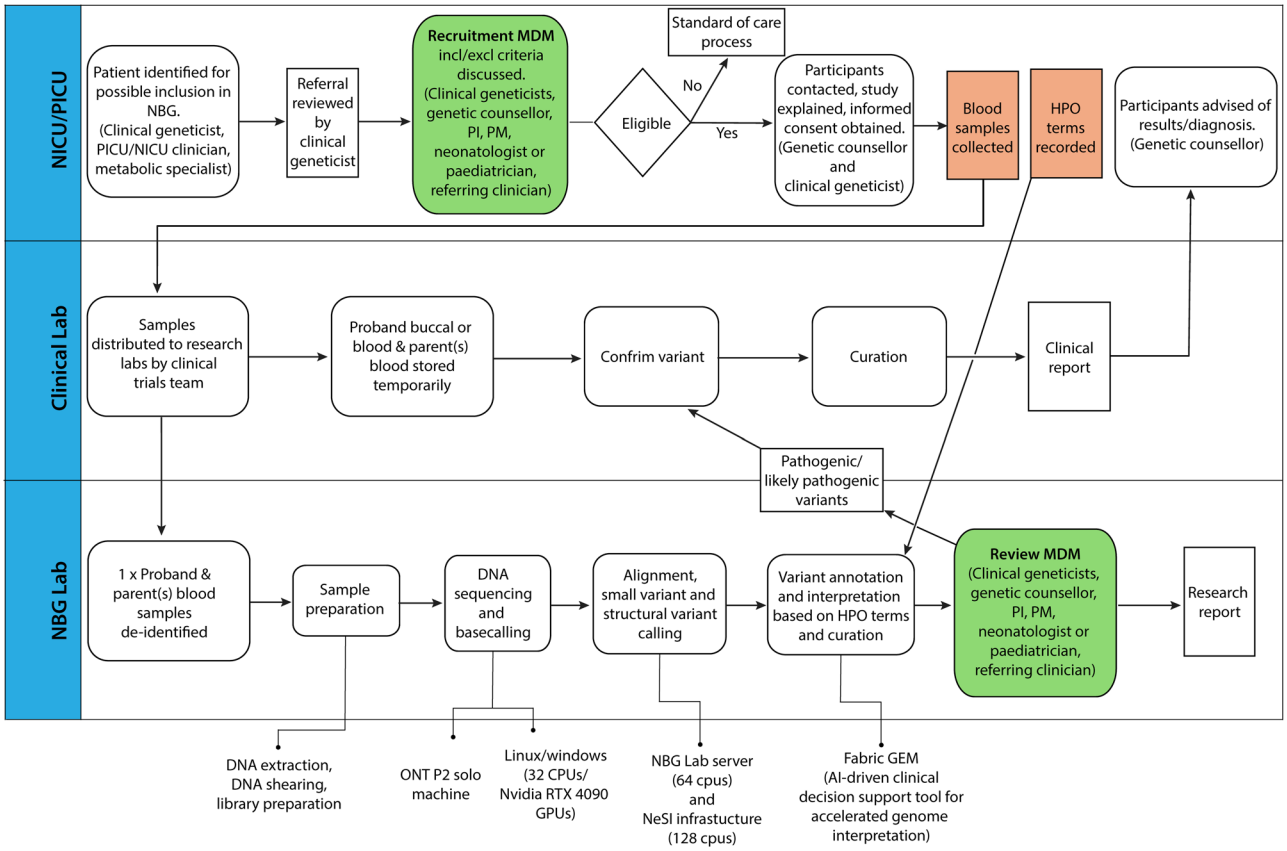

Fig. 5 | Flowchart for a scalable acute care clinical pipeline for rapid genome sequencing. The process begins with the recruitment of critically ill newborns, infants, and children suspected to have genetic disorders admitted to the neonatal and paediatric intensive care (NICU and PICU) (in this instance at Starship Child Health) into the Newborn Genomics programme (NBG) as described in the Methods. Once consented, blood samples (from mother, father) and a buccal or blood sample (from the child) are collected in duplicate. One set of samples is sent to the NBG laboratory for sequencing, and the second to the clinical laboratory (in this instance at the Victorian Clinical Genetics Services [VCGS] in Melbourne, Australia) for variant confirmation. Samples sent to the NBG lab are sequenced using the nanopore PromethION 2 solo system and analysed as described in the Methods. Subsequently, precise genome annotation and rapid variant interpretation is conducted using Fabric GEM™. A further variant review multi-disciplinary meeting is convened as described in the Methods, and orthogonal confirmation (i.e. Sanger Sequencing) of the candidate genetic variants performed by accredited genetic testing facilities (in this instance at VCGS). Finally, clinical (from accredited genetic testing facilities) and research (from NBG) reports are generated, summarizing the evidence for the identified variants, with the clinical report forwarded to the genetic counsellor for communication of the genetic diagnosis to the participants. The figure was created using Adobe Illustrator.

technology in the acute care of patients admitted in neonatal and paediatric intensive care units represents a gap, given the potential benefits it could offer for patient care and disease management.

The incorporation of nanopore sequencing into clinical practice presents challenges given perceptions about the stability and performance of LRS technology[12,23,54,55]. For example, the 97% alignment accuracy can lead to false positives. We contend that these issues are partly due to the inherent limitations of the GRCh38 reference genome. While accurate and comprehensive, it contains gaps and repetitive sequences, such as transposable elements, homopolymer regions, satellite DNA, and segmental duplications, making unique alignment difficult[56]. Additionally, since the reference genome is a composite of multiple individuals, it may not perfectly match any single genome, leading to alignment discrepancies[56]. To address these concerns, we have extensively benchmarked and validated our pipeline, focusing on ensuring consistent and stable performance. Furthermore, we contend that a sequencing depth of 40X, particularly for the sick probands, is sufficient to enhance statistical confidence in genotyping variant calls, even in homopolymer regions, as previously reported[57]. Moreover, in line with clinical and technical recommendations[45,58,59], we have developed standardized protocols and quality control measures for long-read data to facilitate the integration of nanopore sequencing into clinical practice. These protocols ensure data consistency and reliability across different laboratory personnel, analytical teams, and platforms. Implementing such standardization measures is crucial for enabling the broader adoption of LRS technology in clinical settings.

Our established acute care clinical workflow integrates a highly scalable genome sequencing platform (i.e. PromethION 2 solo) with an AI-driven genomics analysis and interpretation software (Fabric GEM™; https://fabricgenomics.com/) to facilitate precise genome annotation, rapid variant interpretation, and prioritization. This pipeline has been established to provide clinicians with actionable information regarding SNVs, small indels, and SVs. In the initial benchmarking phase, we employed GA4GH benchmarking tools and GIAB samples sequenced to >30X coverage, demonstrating highly accurate variant calling metrics, particularly in high-confidence and coding regions of the human genome. Additionally, our workflow accurately detected large-scale CNVs in all sixteen samples from the Coriell Copy Number Variation Reference Panel 1 (CNVPANEL01), showcasing the capability of genotyping STRs and reliably profiling DNA methylation genome-wide. We believe that future updates to the latest ONT library preparation Kit 14 chemistry, along with R10.4.1 flow cells and continuous updates to the base calling models, will further improve precision and recall of SNVs, small indels, and SVs[60].

Finally, in the validation phase, the clinical utility of long-read trio sequencing was demonstrated through the concordance of genomic findings with an established Acute Care Genomics service (VCGS; Melbourne, Australia), which utilized Illumina short-read sequencing technology[42]. Notably, this concordance was achieved in all ten acute cases examined.

https://doi.org/10.1038/s41525-024-00445-5

Table 5 | Clinical characterisation and genomic results from ten proband-parent trio genome sequencing using Illumina short reads and nanopore long reads

Family	Genomic results (Liggins NBG - Acute care genomic research)		Genomic results (VCGS –Acute care genomic service)		Variant type/ mode of inheritance	Clinical diagnosis (phenotypic characterisation in probands)	Clinical indication & Outcome	Turn-around time (TAT; days)*
	Gene	Variant (classification)	Gene	Variant (classification)				
1	*KLHL40*	NM_152393.4: c.1516 A > C p.(Thr506Pro) (Pathogenic)	*KLHL40*	NM_152393.4: c.1516 A > C p.(Thr506Pro) (Pathogenic)	Missense/ AR	Nemaline myopathy 8 (IUGR, microcephaly, multiple contractures, ventilator dependent)	Unifying monogenic diagnosis achieved	7
2	ND	NcV	ND	NcV	ND	Neonatal alloimmune liver disease - non-genetic (confirmed on post-mortem examination)	No unifying monogenic diagnosis. Final diagnosis achieved through PM (non-genetic)	6
3	*VIPAS39*	NM_001400335.1:c.1048-1 G > A (Pathogenic)	*VIPAS39*	NM_001400335.1:c.1048-1 G > A (Pathogenic)	Splice site/ AR	*VIPAS39*-related ARCS2 (Arthrogryposis, renal dysfunction, and cholestasis)	Unifying monogenic diagnosis achieved	5
4	*FAM111A*	NM_001374866.1:c.1451 C > A p.(Ala484Asp) (VUS-3A)	*FAM111A*	NM_001374866.1:c.1451 C > A p.(Ala484Asp) (VUS-3A)	Missense/ de novo AD	*FAM111A*-related craniosteostenosis (DCDA twin, abnormally shaped and thin cranium, slender long bones, fracture, platyspondyly and hypoplastic ilia, hydrops fetalis, ventilator dependent)	Unifying monogenic diagnosis achieved	7
5	*NLRP1*	NM_033004.4: c.3641 C > G p.(Pro1214Arg) (Pathogenic)	*NLRP1*	NM_033004.4: c.3641 C > G p.(Pro1214Arg) (Pathogenic)	Missense/ de novo AD	Autoinflammation with arthritis and dyskeratosis (severe CLD of uncertain aetiology, thrombocytopaenia possibly related to NAIT, conjugated hyperbilirubinemia with raised transaminases, fibrotic pulmonary vein stenosis)	Pathogenic NLRP1 may be contributory but unclear if other monogenic factors involved	4
6	*PTPN11*	NM_002834.5: c.1492 C > T p.Arg498Trp (Pathogenic)	*PTPN11*	(NM_002834.5): c.1492 C > T p.Arg498Trp (Pathogenic)	Missense/ de novo AD	Noonan syndrome 1 (IUGR, complex congenital heart defect, including DORV, septal defects, mod PS, bilateral congenital diaphragmatic hernia, renal pelvis dilatation)	Unifying monogenic diagnosis achieved	4
7	ND	NcV	ND	NcV	ND	No unifying monogenic diagnosis (complex congenital heart defects, persistently raised methionine and homocysteine on NBS)	No unifying monogenic diagnosis for phenotype. Data reanalysis indicated and planned.	6
8	ND	NcV	ND	NcV	ND	No unifying monogenic diagnosis, DDx includes Fanconi anaemia and VACTERL (complex CHD – hypoplastic arch with coarctation and VSD, HLHS), duodenal atresia, sacral segmentation anomaly, anorectal malformation)	No unifying monogenic diagnosis, no pathogenic variants in genes associated with Fanconi anaemia	5
9	ND	NcV	ND	NcV	ND	No monogenic cause found for clinically unexplained neonatal liver failure and associated complication from disseminated intravascular coagulopathy)	No monogenic diagnosis for phenotype. Data reanalysis indicated and planned	4
10	*PCSK1*	NM_000439.5: c.928 G > A p.(Gly310Arg) (VUS-3A)	*PCSK1*	NM_000439.5: c.928 G > A p.(Gly310Arg) (VUS-3A)	Missense/ AR	*PCSK1*-related neonatal severe generalized malabsorptive diarrhoea and failure to thrive (profuse diarrhoea, faltering growth, no endocrinopathy)	Unifying monogenic diagnosis achieved	4

NBG Newborn Genomic Programme, *VCGS* Victorian Clinical Genetic Services, *AR* Autosomal recessive, *AD* autosomal dominant Solved: a clinical diagnosis and therapeutic management consistent with the HPO was made and implemented for the patient. *ND* Not determined, *NcV* no clinically relevant variant, *PM* post-mortem. Age of the probands at consent was between 1 and 60 days. * TAT is working days only and excludes weekends and public holidays.

However, we did not validate the limit of detection of mosaicism for SNV, small indels, CNVs and SVs. Nonetheless, our findings reinforce the potential of long-read sequencing for comprehensive genomic analysis and its applicability in clinical diagnostics.

The implementation of nanopore sequencing in the clinical care pathway is crucial[16–18,42,48,53]. This study demonstrates the feasibility of such a pathway with the availability of funding, technology, skilled laboratory personnel, and researchers supporting rapid genomic testing for critically ill patients. The specialist multidisciplinary team (MDT) model is ideal for complex cases and provides clinicians with input for rare diagnoses that often lack established clinical management guidelines[61,62]. In our opinion, acute care genetics requires a turnaround-driven approach to optimize care and offer economic benefits. In our experience to date, the 5-day average turnaround time from our study seems to meet clinical requirements[63]. In the evaluation of the cohort reported in this study, some cases remained undiagnosed, with at least one being non-genetic. In one case, the diagnosis of biallelic *PCSK1* congenital diarrhoea[64] informed management and surveillance, with the family being reassured about the self-limiting nature of the diarrhoea and surveillance initiated through paediatric endocrine services. Three families used the results for reproductive risk assessment, while reanalysis of the negative trios will continue in the research setting. As expected in critically ill patients, many babies died. However, obtaining a diagnosis offers closure, enabling families to plan subsequent pregnancies.

Deployment of a system like this in New Zealand, a country that arguably has a challenging geography and population distribution, can be managed in two ways. One option is to use the Oxford Nanopore P2 machines as part of a distributed system. In theory, the distributed system can comprise of fully independent, or interdependent nodes that incorporate machines in different locations to optimize return times to a central analytical facility. However, both models could place significant strain on the workforce, requiring highly skilled workers who are already in short supply. The alternative is to centralize facilities in the main population centres in the North and South of the country. Despite New Zealand's challenging geography, we have systems in place to transport samples to these centralized location within hours of collection. These rapid transit programs ensure that turnaround time is not significantly impacted, and the model reduces workforce impacts. We believe this approach will be effective in practice.

In conclusion, we have successfully implemented a scalable clinical pipeline for rapid trio long-read whole-genome sequencing in an acute care setting, aiming to provide prompt and actionable genomic information to clinicians. Through this effort, we have demonstrated the feasibility of achieving rapid precision medicine for critically sick children on a national scale using long-read technology.

Data availability

The genomic data (i.e., bam files) for the Genome in a Bottle (GIAB) samples is publicly available from NCBI Sequence Read Archive (SRA) under BioProject accession number: PRJNA1117929. The genomic and phenotypic data from families analysed in this study cannot be shared publicly due to privacy and ethical restrictions.

Code availability

Pipelines and software used for the analyses reported in this manuscript are publicly available. EPI2ME workflows for aligning FASTQ files sequences to the reference genome is available on https://github.com/epi2me-labs/wf-alignment and the human variation pipeline for variant calling (i.e. SNVs/indels, SVs, CNVs, STRs), and modified bases analysis can be accessed on https://github.com/epi2me-labs/wf-human-variation. Hap.py is available on https://github.com/Illumina/hap.py and Truvari on https://github.com/ACEnglish/truvari. Dorado is available on https://github.com/nanoporetech/dorado. Modkit is available on https://github.com/nanoporetech/modkit. Samtools is available on https://github.com/samtools/samtools. Modbamtools is available on https://github.com/rrazaghi/modbamtools. A docker image containing seaborn, matplotlib, and python libraries used to generate plots is available as docker://nyagam/seaborn:latest. Ont-methylDMR-kit, a workflow for differential methylation analysis, together with a docker image containing the Bioconductor DSS (Dispersion Shrinkage for Sequencing) package, is available on https://github.com/NyagaM/ont-methylDMR-kit.

Received: 17 June 2024; Accepted: 29 October 2024;
Published online: 08 November 2024

References

1. Lappalainen, T., Scott, A. J., Brandt, M. & Hall, I. M. Genomic analysis in the age of human genome sequencing. *Cell* **177**, 70–84 (2019).
2. Bick, D., Jones, M., Taylor, S. L., Taft, R. J. & Belmont, J. Case for genome sequencing in infants and children with rare, undiagnosed or genetic diseases. *J. Med. Genet.* **56**, 783–791 (2019).
3. Gahl, W. A. et al. The National Institutes of Health Undiagnosed Diseases Program: Insights into rare diseases. *Genet. Med.* **14**, 51–59 (2012).
4. Isik, E. et al. Clinical utility of a targeted next generation sequencing panel in severe and pediatric onset Mendelian Diseases. *Eur. J. Med. Genet.* **62**, 103725 (2019).
5. Lupski, J. R. et al. Exome sequencing resolves apparent incidental findings and reveals further complexity of *SH3TC2* variant alleles causing Charcot-Marie-Tooth neuropathy. *Gen. Med.* **5**, 1–17 (2013).
6. Yang, Y. et al. Clinical whole-exome sequencing for the diagnosis of Mendelian Disorders. *N. Engl. J. Med.* **369**, 1502–1511 (2013).
7. Lincoln, S. E. et al. A systematic comparison of traditional and multigene panel testing for hereditary breast and ovarian cancer genes in more than 1000 patients. *J. Mol. Diagnostics* **17**, 533–544 (2015).
8. Ramakrishnan, K. A. et al. Precision molecular diagnosis defines specific therapy in combined immunodeficiency with megaloblastic anemia secondary to MTHFD1 deficiency. *J. Allerg. Clin. Immunol.: In Practice* **4**, 1160–1166 (2016).
9. Costain, G. et al. Genome sequencing as a diagnostic test in children with unexplained medical complexity. *JAMA Netw. Open* **3**, 1–15 (2020).
10. Goodwin, S., McPherson, J. D. & McCombie, W. R. Coming of age: Ten Years of next-generation Sequencing Technologies. *Nat. Rev. Genet.* **17**, 333–351 (2016).
11. Quail, M. et al. A tale of three next generation sequencing platforms: Comparison of Ion Torrent, Pacific Biosciences and Illumina MiSeq sequencers. *BMC Genomics* **13**, 341 (2012).
12. Mantere, T., Kersten, S. & Hoischen, A. Long-read sequencing emerging in Medical Genetics. *Front. Genet.* **10**, 1–14 (2019).
13. Cretu Stancu, M. et al. Mapping and phasing of structural variation in patient genomes using nanopore sequencing. *Nat. Commun.* **8**, 1–13 (2017).
14. de la Morena-Barrio, B. et al. Long-read sequencing identifies the first retrotransposon insertion and resolves structural variants causing antithrombin deficiency. *Thrombosis Haemost.* **122**, 1369–1378 (2022).
15. Chaisson, M. J. et al. Resolving the complexity of the human genome using single-molecule sequencing. *Nature* **517**, 608–611 (2014).
16. Gorzynski, J. E. et al. Ultrarapid Nanopore genome sequencing in a critical care setting. *N. Engl. J. Med.* **386**, 700–702 (2022).
17. Greer, S. U. et al. Implementation of nanopore sequencing as a pragmatic workflow for copy number variant confirmation in the Clinic. *J. Transl. Med.* **21**, 1–11 (2023).
18. Zalusky, M. P. G. et al. 3-Hour genome sequencing and targeted analysis to rapidly assess genetic risk. *Genetics in Medicine Open.* **2**, 1–9 (2022).
19. Reiner, J. et al. Cytogenomic identification and long-read single molecule real-time (SMRT) sequencing of a bardet–biedl syndrome 9 (BBS9) deletion. *npj Genom. Med.* **3**, (2018).
20. Miller, D. E. et al. Targeted long-read sequencing identifies missing disease-causing variation. *Am. J. Hum. Genet.* **108**, 1436–1449 (2021).

21. Zeng, S. et al. Long-read sequencing identified intronic repeat expansions in *samd12* from Chinese pedigrees affected with familial cortical myoclonic tremor with epilepsy. *J. Med. Genet.* **56**, 265–270 (2018).
22. Ebbert, M. T. et al. Long-read sequencing across the c9orf72 'GGGGCC' repeat expansion: Implications for clinical use and genetic discovery efforts in human disease. *Mol. Neurodegeneration* **13**, 1–17 (2018).
23. Amarasinghe, S. L. et al. Opportunities and challenges in long-read sequencing data analysis. *Genome Biol.* **21**, 1–16 (2020).
24. Krusche, P. et al. Best practices for benchmarking germline small-variant calls in human genomes. *Nat. Biotechnol.* **37**, 555–560 (2019).
25. Jennings, L., Van Deerlin, V. M. & Gulley, M. L. Recommended principles and practices for validating clinical molecular pathology tests. *Arch. Pathol. Lab. Med.* **133**, 743–755 (2009).
26. Aziz, N. et al. College of American pathologists' laboratory standards for next-generation sequencing clinical tests. *Arch. Pathol. Lab. Med.* **139**, 481–493 (2015).
27. Zook, J. M. et al. Integrating human sequence data sets provides a resource of benchmark SNP and Indel Genotype calls. *Nat. Biotechnol.* **32**, 246–251 (2014).
28. Zook, J. M. et al. Extensive sequencing of seven human genomes to characterize benchmark reference materials. *Scientific Data* **3**, 1–26 (2016).
29. Wagner, J. et al. Curated variation benchmarks for challenging medically relevant autosomal genes. *Nat. Biotechnol.* **40**, 672–680 (2022).
30. Goenka, S. D. et al. Accelerated identification of disease-causing variants with ultra-rapid nanopore genome sequencing. *Nat. Biotechnol.* **40**, 1035–1041 (2022).
31. Payne, A. et al. Readfish enables targeted nanopore sequencing of gigabase-sized genomes. *Nat. Biotechnol.* **39**, 442–450 (2020).
32. Yue, X. et al. Simultaneous profiling of histone modifications and DNA methylation via nanopore sequencing. *Nat. Commun.* **13**, 1–14 (2022).
33. Olova, N. et al. Comparison of whole-genome bisulfite sequencing library preparation strategies identifies sources of biases affecting DNA methylation data. *Genome Biol.* **19**, 1–19 (2018).
34. Li, H. Minimap2: Pairwise alignment for nucleotide sequences. *Bioinformatics* **34**, 3094–3100 (2018).
35. Zheng, Z. et al. Symphonizing pileup and full-alignment for deep learning-based long-read variant calling. *Nat. Comput. Sci.* **2**, 797–803 (2022).
36. Smolka, M. et al. Detection of mosaic and population-level structural variants with Sniffles2. *Nat. Biotechnol.* https://doi.org/10.1038/s41587-023-02024-y (2024).
37. Scheinin, I. 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.* **24**, 2022–2032 (2014).
38. Chiu, R., Rajan-Babu, I.-S., Friedman, J. M. & Birol, I. Straglr: Discovering and genotyping tandem repeat expansions using whole genome long-read sequences. *Genome Biol.* **22**, 1–20 (2021).
39. English, A. C., Menon, V. K., Gibbs, R. A., Metcalf, G. A. & Sedlazeck, F. J. Truvari: Refined structural variant comparison preserves allelic diversity. *Genome Biol.* **23**, 1–20 (2022).
40. Li, H. et al. The sequence alignment/map format and SAMtools. *Bioinformatics* **25**, 2078–2079 (2009).
41. Feng, H., Conneely, K. N. & Wu, H. A bayesian hierarchical model to detect differentially methylated loci from single nucleotide resolution sequencing data. *Nucleic Acids Res.* **42**, 1–11 (2014).
42. Lunke, S. et al. Integrated multi-omics for rapid rare disease diagnosis on a national scale. *Nat. Med.* **29**, 1681–1691 (2023).
43. Dimmock, D. et al. Project Baby Bear: Rapid Precision Care incorporating rwgs in 5 California Children's hospitals demonstrates improved clinical outcomes and reduced costs of care. *Am. J. Hum. Genet.* **108**, 1231–1238 (2021).
44. McKeown, C. et al. A pilot study of exome sequencing in a diverse New Zealand cohort with undiagnosed disorders and cancer. *J. R. Soc. N.Z.* **48**, 262–279 (2018).
45. Martin, A. R. et al. Panelapp crowdsources expert knowledge to establish consensus diagnostic gene panels. *Nat. Genet.* **51**, 1560–1565 (2019).
46. Durkie, M. et al. ACGS best practice guidelines for variant classification in rare disease 2024. *Assoc. Clin. Genomic Sci.* 1–54 (2024).
47. Gustafson, J. A. et al. Nanopore sequencing of 1000 genomes project samples to build a comprehensive catalog of human genetic variation. https://doi.org/10.1101/2024.03.05.24303792 (2024).
48. Marshall, C. R. et al. Best practices for the analytical validation of clinical whole-genome sequencing intended for the diagnosis of germline disease. *npj Genomic Med.* **5**, 1–12 (2020).
49. Johnson, N. E. et al. Consensus-based care recommendations for congenital and childhood-onset myotonic dystrophy type 1. *Neurol. Clin. Pract.* **9**, 443–454 (2019).
50. Cortés-Martín, J. et al. Deletion syndrome 22q11.2: A systematic review. *Children* **9**, 1168 (2022).
51. Nobile, V., Pucci, C., Chiurazzi, P., Neri, G. & Tabolacci, E. DNA methylation, mechanisms of *FMR1* inactivation and therapeutic perspectives for Fragile X Syndrome. *Biomolecules* **11**, 296 (2021).
52. Priest, J. R. et al. Molecular diagnosis of long QT syndrome at 10 days of life by rapid whole genome sequencing. *Heart Rhythm* **11**, 1707–1713 (2014).
53. Lunke, S. & Stark, Z. Can rapid nanopore sequencing bring genomic testing to the bedside? *Clin. Chem.* **68**, 1484–1485 (2022).
54. Carbo, E. C. et al. A comparison of five Illumina, Ion Torrent, and nanopore sequencing technology-based approaches for whole genome sequencing of SARS-COV-2. *SSRN Electron. J.* **42**, 701–703 (2023).
55. Oehler, J. B., Wright, H., Stark, Z., Mallett, A. J. & Schmitz, U. The application of long-read sequencing in clinical settings. *Human Genomics* **17**, 1–13 (2023).
56. Nurk, S. et al. The complete sequence of a human genome. *Science* **376**, 44–53 (2022).
57. Bowden, R. et al. Sequencing of human genomes with Nanopore Technology. *Nat. Commun.* **10**, 1–9 (2019).
58. Richards, S. et al. Standards and guidelines for the interpretation of sequence variants: A joint consensus recommendation of the American College of Medical Genetics and genomics and the Association for Molecular Pathology. *Genet. Med.* **17**, 405–424 (2015).
59. Fukasawa, Y., Ermini, L., Wang, H., Carty, K. & Cheung, M.-S. LongQC: A Quality Control tool for third generation Sequencing long read data. *G3 Genes|Genomes|Genet.* **10**, 1193–1196 (2020).
60. Ni, Y., Liu, X., Simeneh, Z. M., Yang, M. & Li, R. Benchmarking of nanopore R10.4 and R9.4.1 flow cells in single-cell whole-genome amplification and whole-genome shotgun sequencing. *Computational Struct. Biotechnol. J.* **21**, 2352–2364 (2023).
61. Macken, W. L. et al. Specialist multidisciplinary input maximises rare disease diagnoses from whole genome sequencing. *Nat. Commun.* **13**, 1–9 (2022).
62. Ma, A. et al. What is the power of a genomic multidisciplinary team approach? A systematic review of implementation and Sustainability. *Eur. J. Hum. Genet.* **32**, 381–391 (2024).
63. Goranitis, I. et al. Is faster better? an economic evaluation of rapid and ultra-rapid genomic testing in critically ill infants and children. *Genet. Med.* **24**, 1037–1044 (2022).
64. Aerts, L. et al. Novel homozygous inactivating mutation in the *PCSK1* gene in an infant with congenital malabsorptive diarrhea. *Genes* **12**, 710 (2021).

Acknowledgements

The authors would like to express our thanks to the families who participated in the sequencing programme during times of great stress. The authors would like to thank: Anita Lee and the Genetic Counsellors and Clinical Geneticists at Genetics Health Services New Zealand; the Neonatologists, Paediatricians, Clinical specialists, nurses and support staff at the Neonatal Intensive Care and Paediatric Intensive Care Units at Auckland City Hospital; the staff at Diagnostic Genetics, LabPLUS for their discussions, and the clinical care of the patients involved in this study. We gratefully acknowledge the generous donations from the Dines Family Trust, The Tautoku Trust, and The Kelliher Trust. The authors wish to acknowledge the use of New Zealand eScience Infrastructure (NeSI) high performance computing facilities, consulting support and/or training services as part of this research. New Zealand's national facilities are provided by NeSI and funded jointly by NeSI's collaborator institutions and through the Ministry of Business, Innovation & Employment's Research Infrastructure programme (https://www.nesi.org.nz).

Author contributions

Conceptualization: J.M.O.; Study design: J.M.O., D.M.N., P.T., C.G., H.H.P., P.Q.S., S.F., G.T., E.T., K.G., J.V.D., P.Y., M.G; Bioinformatics analysis: D.M.N., P.T.; Writing-original draft: D.M.N., P.T.; Visualization: D.M.N., P.T.; Formal analysis: D.M.N., P.T.; Methodology: C.G., H.H.P., J.R., S.F; Ethics: C.G., J.V.D; Investigation: J.M.O., D.M.N., P.T., C.G., H.H.P., P.Q.S., S.F., J.V.D., P.Y., M.G; Funding acquisition: J.M.O; Resources: J.M.O., J.V.D; Patient recruitment: P.Y., K.G. Writing-review & editing: J.M.O., D.M.N., P.T., C.G., H.H.P., P.Q.S., S.F., G.T., E.T., S.L., Z.S., K.G., J.V.D., P.Y., M.G; Principal investigator: J.M.O. All authors approved the submission of this manuscript.

Competing interests

Prof Zornitza Stark is an Associate Editor of npj Genomic Medicine and is not involved in the peer review process or decision-making for this manuscript. The remaining authors have no conflict of interest to disclose.

Additional information

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41525-024-00445-5.

Correspondence and requests for materials should be addressed to Justin M. O'Sullivan.