

ГЕНЕТИКА/GENETICS

DOI: https://doi.org/10.60797/jbg.2026.33.7 EDN: LXLPJE

WHOLE GENOME SEQUENCING (WGS) WITH PACBIO HIFI: A COMPREHENSIVE FRAMEWORK FOR CLINICAL APPLICATIONS IN SOUTHEAST ASIA

Review article

Budhyanto V.[1], Jurgens S.[2], Cicilia C.O.[3,] *, Kamaruddin M.[4]
[1] ORCID : 0009-0000-2176-6417;
[3] ORCID : 0009-0003-3231-4629;
[1, 2, 3] Satriabudi Dharma Setia Foundation, Jakarta, Indonesia
[4] Muhammadiyah University Semarang, Semarang, Indonesia

* Corresponding author (cindy.o.cicilia[at]dharma.or.id)

Suggested: 24.07.2026; Accepted: 21.09.2026; Published: 25.09.2026

Abstract

Background: although short-read whole-genome sequencing has become the clinical standard, important genomic regions including structural variants, tandem repeat expansions, segmental duplications, and complex pharmacogenomic loci remain difficult to resolve. PacBio HiFi sequencing offers highly accurate long reads capable of simultaneously detecting small variants, structural rearrangements, repeat expansions, methylation, and haplotype phase within a single assay.

Objective: this review evaluates current evidence supporting PacBio HiFi whole-genome sequencing for clinical implementation and discusses its potential role in Southeast Asia, where high genetic diversity, founder mutations, and pharmacogenomic variation create unique diagnostic challenges.

Evidence Summary: published clinical studies consistently demonstrate higher diagnostic yield compared with conventional short-read sequencing, particularly for rare diseases, repeat expansion disorders, structural variation, and pharmacogenomics. Recent advances in sequencing chemistry and bioinformatics further support clinical scalability.

Conclusion: PacBio HiFi WGS represents a promising next-generation clinical sequencing platform capable of improving diagnostic accuracy and precision medicine implementation across Southeast Asia.

Keywords: PacBio HiFi, long-read whole-genome sequencing, rare disease diagnostics, pharmacogenomics, Southeast Asian population genomics.

ПОЛНОГЕНОМНОЕ СЕКВЕНИРОВАНИЕ (WGS) С ИСПОЛЬЗОВАНИЕМ ТЕХНОЛОГИИ PACBIO HIFI: КОМПЛЕКСНЫЙ ПОДХОД К КЛИНИЧЕСКОМУ ПРИМЕНЕНИЮ В ЮГО-ВОСТОЧНОЙ АЗИИ

Обзор

Будьянто В.[1], Юргенс С.[2], Сисилия С.О.[3,] *, Камаруддин М.[4]
[1] ORCID : 0009-0000-2176-6417;
[3] ORCID : 0009-0003-3231-4629;
[1, 2, 3] Фонд Сатриабуди Дхарма Сетиа, Джакарта, Индонезия
[4] Университет Мухаммадии, Semarang, Индонезия

* Корреспондирующий автор (cindy.o.cicilia[at]dharma.or.id)

Предложена: 24.07.2026; Принята: 21.09.2026; Опубликована: 25.09.2026

Аннотация

Введение: хотя секвенирование всего генома с использованием коротких прочтений стало клиническим стандартом, важные геномные регионы, включая структурные варианты, расширения тандемных повторов, сегментарные дупликации и сложные фармакогеномные локусы, по-прежнему трудно поддаются анализу. Секвенирование PacBio HiFi предлагает высокоточные длинные прочтения, способные одновременно обнаруживать небольшие варианты, структурные перестройки, расширения повторов, метилирование и фазу гаплотипа в рамках одного анализа.

Цель: в этом обзоре оцениваются имеющиеся данные, подтверждающие целесообразность клинического применения секвенирования всего генома PacBio HiFi, и обсуждается его потенциальная роль в Юго-Восточной Азии, где высокое генетическое разнообразие, мутации-основатели и фармакогеномные вариации создают уникальные диагностические проблемы.

Краткое изложение данных: опубликованные клинические исследования последовательно демонстрируют более высокую диагностическую эффективность по сравнению с традиционным секвенированием коротких прочтений, особенно для редких заболеваний, заболеваний, связанных с расширением повторов, структурных вариаций и фармакогеномики. Последние достижения в химии секвенирования и биоинформатике дополнительно подтверждают масштабируемость клинического применения.

Заключение: PacBio HiFi WGS представляет собой перспективную платформу для клинического секвенирования следующего поколения, способную повысить точность диагностики и внедрить методы прецизионной медицины в Юго-Восточной Азии.

Ключевые слова: PacBio HiFi, полногеномное секвенирование с длинными прочтениями, диагностика редких заболеваний, фармакогеномика, популяционная геномика Юго-Восточной Азии.

Introduction

Clinical genomics has moved through successive technology epochs. Illumina short-read whole-exome and whole-genome sequencing (WGS), widespread from ~2008, accelerated Mendelian gene discovery and became a first- or second-tier diagnostic test. Yet diagnostic yield plateaued at roughly 30–50% across rare-disease cohorts, leaving many multi-year diagnostic odysseys unresolved because short reads cannot reliably resolve repetitive regions, detect structural variants (SVs), or phase alleles across long distances. Long-read single-molecule platforms addressed this gap: Pacific Biosciences' single-molecule real-time (SMRT) sequencing reads DNA synthesis by a polymerase within a zero-mode waveguide [11], while Oxford Nanopore (ONT) reads electrical-signal changes as DNA passes through a protein pore [12].

The decisive advance was circular consensus sequencing (CCS): a polymerase passes repeatedly over a circularised template, and the noisy subreads collapse into a single "HiFi" read of 10–25 kb at Q30+ (~99.9%) accuracy. Wenger et al. [4] established HiFi on the HG002 reference with ≥99.9% SNV precision/recall and >95% large-SV detection in one assay. The platform advanced from Sequel (2015) through Sequel II/IIe to Revio (2023; ~15× throughput, on-instrument CpG methylation) and the benchtop Vega (2025); SPRQ chemistry (2024) added native N6-methyladenine (6mA) calling and lowered DNA input. Each platform retains legitimate niches ONT for ultra-long assembly and rapid field use but for routine clinical WGS, rare-disease diagnostics, pharmacogenomics, and methylation-aware analysis, HiFi is currently the most mature single-assay solution.

Three findings motivate genome-wide HiFi. First, SVs are pervasive and clinically significant: long-read sequencing of 3,622 Icelanders found a median ~22,600 SVs per genome, 4–5× the short-read estimate, including disease-relevant SVs (e.g. affecting PCSK9, ACAN) miscalled by short reads [3]. Second, many clinically important genes lie in "dark"/camouflaged regions inaccessible to short reads; HiFi reduces unresolved coding sequence from ~43% to <10%, resolving loci such as SMN1/SMN2, PMS2/PMS2CL, CYP2D6/CYP2D7, HBA1/HBA2, PKD1, NCF1, and HLA [1]. Third, a single linear reference biases all short-read analysis: the T2T-CHM13 genome added ~200 Mb of sequence [8], and the HPRC draft pangenome (47 ancestrally diverse, HiFi-based assemblies) reduced small-variant errors ~34% with the largest gains in non-European ancestries [7]. Clinically, a Radboudumc reanalysis of 100 undiagnosed patients detected 93% of pathogenic variants with a single HiFi assay replacing 2–6 orthogonal tests, now scaling toward >5,000 genomes [5]. Independent programs report concordant gains, and CLIA-validated clinical HiFi tests are already in use [20].

This review appraises HiFi WGS for clinical use, emphasizing the disease epidemiology, founder-mutation landscape, and pharmacogenomic profile of Southeast Asia (SEA), while critically noting real obstacles including cost, throughput, interpretation bottlenecks, and ancestry-specific reference gaps.

This is a narrative non-systematic review synthesising peer-reviewed primary literature, reference-consortium releases (Telomere-to-Telomere, HPRC, GenomeAsia 100K, SG10K_Health), and published clinical-cohort reports on PacBio HiFi WGS. No search protocol with predefined date limits was registered, sources reflect landmark and representative evidence. Manufacturer specifications are reported as declared performance, distinct from independent clinical validation. Diagnostic-yield figures are given as published, without meta-analysis or formal risk-of-bias scoring.

Results & discussion

3.1. Technical foundations of the PacBio HiFi platform

3.1.1. SMRT Sequencing Principle

PacBio Single-Molecule Real-Time (SMRT) sequencing is based on three core innovations originally described by Eid et al. First, DNA synthesis is monitored within a zero-mode waveguide (ZMW), a nanoscale reaction chamber that enables real-time observation of a single DNA polymerase molecule. Second, phospholinked fluorescent nucleotides are incorporated during DNA synthesis, with the fluorescent label released as part of the natural polymerization process, leaving the newly synthesized DNA unmodified. Third, the platform records the kinetics of nucleotide incorporation, including interpulse duration (IPD) and pulse width, which vary in the presence of DNA base modifications such as 5-methylcytosine (5mC) and N6-methyladenine (6mA). This unique capability enables the direct detection of DNA methylation from native DNA without bisulfite conversion or antibody-based enrichment.

3.1.2. Circular Consensus / HiFi Read Generation

Genomic DNA is sheared to ~15–25 kb fragments and ligated to hairpin SMRTbell adaptors to create a topologically closed, single-stranded circular template. A strand-displacing polymerase traverses the template multiple times during the SMRT cell run. The resulting subreads for each an imperfect ~10–15% raw-error copy, are then computationally collapsed into a single high-quality CCS / HiFi consensus read. Typical HiFi yields are 15–25 kb reads at Q30+ (~99.9% per-base accuracy), with current Revio output ~360 Gb per SMRT cell (≥4–5 whole human genomes at 30× equivalent depth per 24-hour run).

3.1.3. Native DNA Preservation and Base-Modification Detection

Because HiFi sequencing requires no PCR amplification, base modifications present on native genomic DNA are preserved through library preparation and read out via polymerase kinetics. The current Revio primary analysis pipeline calls 5-methylcytosine at CpG sites with sensitivity comparable to whole-genome bisulfite sequencing (WGBS) and concordance with EPIC/450K arrays generally >95% [6]. SPRQ chemistry introduced in 2024–2025 expands epigenetic detection to include N6-methyladenine (6mA) and enhances sensitivity for 5mC even at reduced sequencing depth, enabling CpG methylation profiling alongside standard variant calling. Importantly, when combined with read-based phasing tools such as HiPhase, methylation signals are inherently assigned to specific haplotypes, allowing allele-specific methylation analysis to be performed in a single integrated assay.

3.1.4. Read-Length Distribution

Read-length distribution is shaped by library shearing (Megaruptor, gTUBE) and SMRTbell size selection (BluePippin, AMPure). Typical clinical HiFi WGS libraries yield N50 read length 15–20 kb, with reads >25 kb common; this exceeds the size of >99% of typical pathogenic structural variants, the vast majority of mobile-element insertions, and the unexpanded alleles of all known disease-associated tandem repeats. Read length is the single most important parameter determining the resolution of complex segmental duplications (e.g., PMS2/PMS2CL, SMN1/SMN2, CYP2D6/CYP2D7, PKD1/PKD1P), since at least some HiFi reads must uniquely span the region of divergence between paralogues.

3.1.5. Comparative Platform Performance

In direct head-to-head benchmarking on the Genome in a Bottle reference samples (HG002 trio):

1. *Short-read Illumina WGS* (150 bp paired-end at 30×): SNV recall and precision of ≥99.5% within high-confidence GIAB regions; however, performance declines substantially in low-complexity regions, segmental duplications, and repetitive sequences. SV recall is ~30–50% across the SV size spectrum, with extreme under-ascertainment of insertions and tandem repeat expansions.

2. *Oxford Nanopore* (ONT) R10.4.1 simplex/duplex at 30×: SNV recall ~99.4% (simplex) to >99.7% (duplex); strong SV recall (~90%+), with the unique advantage of ultra-long reads (>100 kb) useful for de novo assembly and phasing.

3. *PacBio HiFi WGS* at 30×: SNV precision and recall ≥99.9% in the GIAB benchmark regions, with the additional advantage of joint calling across SNVs, indels up to several kilobases, SVs, tandem repeats, and 5mC methylation from a single dataset [4], [13], [14].

Each platform has legitimate clinical niches. HiFi WGS currently provides the most balanced performance across accuracy, throughput, epigenetic profiling, and comprehensive variant detection, making it the most suitable option for routine clinical genomics. In contrast, ONT remains particularly valuable for ultra-long-read applications such as telomere-to-telomere assembly, resolving very large structural variants, and rapid or field-based sequencing workflows. For routine clinical WGS, rare disease diagnostics, pharmacogenomics, and methylation-aware genome analysis, current evidence strongly supports HiFi WGS as the most mature single-assay solution. Conversely, ultra-long-range assembly, telomere-to-telomere reconstruction, and certain highly repetitive structural rearrangements remain areas in which ONT retains important advantages.

3.1.6. Coverage Requirements

Coverage requirements differ by application:

1. *Low-pass HiFi* (1–5×): Sufficient for ancestry inference, cell-free DNA fragmentomics, low-resolution copy-number profiling, and population-scale association studies when combined with appropriate imputation panels.

2. *Standard clinical HiFi WGS* (15–30×): The current standard for diagnostic rare-disease applications. At ~20× HiFi, SNV, indel, SV, tandem-repeat, methylation, and pharmacogenomic diplotype calling are all reliably achievable in a single run [5].

3. *Deep HiFi* (≥40×): Needed for somatic mosaic variant detection (~1–5% VAF), comprehensive tumour profiling, and deep-trio de novo mutation analysis.

3.1.7. Vega versus Revio: Clinical Positioning

The Revio platform (launched in 2023), serves as the high-throughput foundation for population-scale HiFi sequencing, enabling simultaneous operation of up to four SMRT Cells and generating approximately 1,300 Gb of data per run. In contrast, the Vega benchtop system, launched in 2025, is designed for single SMRT Cell sequencing and is intended for clinical laboratories, regional reference centres, and academic institutions that do not require the throughput capacity of Revio. For most clinical genomics initiatives in Southeast Asia, a hub-and-spoke deployment model is likely to be both operationally efficient and cost-effective, with Revio systems centralized at regional sequencing hubs and Vega instruments deployed at tertiary academic hospitals to support rapid-turnaround clinical genome sequencing.

Table 1 - Comparison of clinical sequencing platforms for whole-genome analysis

DOI: https://doi.org/10.60797/jbg.2026.33.7.1

Feature	Illumina SR-WGS	ONT R10.4.1	PacBio Revio HiFi	PacBio Vega HiFi
Read length	150 bp	10–100+ kb (ultra-long)	10–25 kb	10–25 kb
Per-base accuracy	~Q30 (99.9%)	Q20–Q30	Q30+ (99.9%)	Q30+ (99.9%)
SNV recall (GIAB)	≥99.5%	99.4–99.7%	≥99.9%	≥99.9%
SV recall (≥50 bp)	30–50%	~90%	≥90%	≥90%
Repeat expansions	Mostly invisible	Native detection	Native detection	Native detection
Native 5mC/5hmC/6mA	No	Yes	Yes (SPRQ-Nx)	Yes (SPRQ-Nx)
Throughput/run	Very high	Variable	~4–5 genomes/24h	1 SMRT cell
Best clinical fit	Population panels, oncology panels	Ultra-long phasing, rapid in-field	Population WGS hub	Tertiary clinical site

3.2. WGS modalities

Low-pass HiFi (1–5×) supports cfDNA/NIPT with preserved fragmentomic and methylation signatures [15] (directly translatable to NIPT for imprinted-disorder pregnancies and circulating-tumour-DNA workflows), aneuploidy and large-CNV

detection, haplotype-panel building in under-represented populations (where 4–5× HiFi genotyping can exceed 30× short-read in difficult regions), and methylation EWAS (CpG estimates correlate >0.95 with WGBS) [6]. Standard 15–30× HiFi is the workhorse:

– SNVs/indels via DeepVariant at ≥99.9% precision/recall [16];

– SVs via sawfish/Sniffles2 at ≥90% recall/precision across deletions, insertions, duplications, and inversions [13];

– mobile-element insertions (Alu, L1, SVA) recovered;

– tandem repeats genotyped by TRGT with 99.56% Mendelian concordance across ~937,000 loci [18]; end-to-end read-based phasing (phase-block N50 100–500 kb); and genome-wide 5mC were all in one run. Deep/trio HiFi (≥40×) adds de novo repeat-mutation calling (TRGT-denovo) [19], somatic mosaicism to ~1–2% VAF, and cancer applications including SV-landscape characterisation, chromothripsis/chromoplexy reconstruction, and longitudinal tumour and minimal-residual-disease monitoring.

3.3. Clinical applications

In pediatric rare disease, 50–70% of well-phenotyped patients remain undiagnosed after exome or short-read WGS, largely because of SV-, repeat-, and methylation-class variants specifically structural variants, mobile-element insertions, tandem-repeat expansions, complex rearrangements, variants within segmental duplications, and deep-intronic or "dark"-region variants. HiFi WGS captures SNVs, indels, SVs, repeats, phase, and methylation together; the Radboudumc cohort reached 93% sensitivity against 2–6 assays [5], and Genomic Answers for Kids (GA4K) produced haplotype-resolved genomes exposing allele-specific methylation and promoter-silencing expansions [6]. Across programs (Genomics England [21], Australian Genomics [22], GA4K, and others), HiFi adds ~10-25% yield over short-read WGS.

Recurrent CNV syndromes (DiGeorge, Williams, Prader–Willi, Angelman) gain single-nucleotide breakpoint resolution and parent-of-origin assignment via imprinted-locus methylation [3], [7]. Repeat-expansion disorders that usually invisible to short reads at pathogenic sizes are sized with methylation and interruptions by TRGT [18]; SEA-relevant examples include SCA3 (ATXN3), NOTCH2NLC-related NIID (GC-rich, long reads essentially required) [23], and SAMD12-BAFME [24], alongside fragile X (FMR1), myotonic dystrophy (DMPK/CNBP), C9orf72, and Friedreich ataxia (FXN). Imprinting disorders like Prader–Willi, Angelman, Beckwith–Wiedemann, and Silver–Russell collapse deletion/UPD/methylation testing into one assay, indicating ~80% of GA4K hypermethylation outliers were allele-specific [6].

In neurodegeneration, HiFi distinguishes GBA1 from its pseudogene GBAP1 and detects SNCA multiplications in Parkinson disease, sizes C9orf72 in ALS/FTD (with SOD1, FUS, TARDBP), resolves the PMP22 duplication of CMT1A, and interrogates APP, APOE/TOMM40, PRNP codon-129 phase, and the "dark" CR1 locus in Alzheimer disease [1]. In cardiovascular genetics it phases TTN and covers MYH7/MYBPC3 (cardiomyopathy), KCNQ1/KCNH2/SCN5A and RYR2 (channelopathies), FBN1 (Marfan), and Alu-mediated LDLR SVs (familial hypercholesterolaemia), and uniquely sizes the LPA KIV-2 VNTR that determines Lp(a) concentration [25]. In hereditary cancer, Paraphase resolves PMS2 vs PMS2CL in Lynch syndrome and, applied genome-wide across 316 paralogues in 160 segmental duplications, revealed hidden de novo and gene-conversion events [26]; BRCA1/2, PALB2, ATM, APC, MUTYH, and hematologic-predisposition genes (RUNX1, DDX41, GATA2) are comprehensively covered including SVs and deep-intronic variants. SMN1/SMN2 profiling for spinal muscular atrophy is likewise a paralogue problem well suited to HiFi [64], and complex hematologic-malignancy architectures (KMT2A/MLL and NUP98 rearrangements) are natively resolved.

Pharmacogenomics historically needed many assays because CYP2D6, UGT1A1, SLCO1B1, NUDT15, TPMT, HLA, and DPYD are afflicted by structural variation and paralogues. StarPhase [27] diplotypes 21 CPIC Level-A genes plus 4-field HLA-A/B and CYP2D6 from one HiFi run, matching ~99.5% of GeT-RM calls and correcting 26.2% that many because of structural variation invisible to legacy assays. Somatic profiling adds chromothripsis reconstruction, HLA loss-of-heterozygosity (immunotherapy-relevant), and methylation-based tumour classification [28], with mosaic SV calling from HiFi [13]. Renal genetics benefits from complete resolution of PKD1 (buried in six pseudogenes) in ADPKD and of the COL4A3/4/5 genes in Alport syndrome, primary immunodeficiency from native resolution of NCF1 and its pseudogenes in chronic granulomatous disease, and reproductive medicine from simultaneous expanded carrier screening (HBA1/HBA2, HBB, CFTR, SMN1, FMR1) and haplotype-resolved PGT-M/PGT-SR. Mitochondrial disease is addressed across both the 16.6 kb mtDNA such as heteroplasmy quantification and large-deletion detection, and >250 nuclear genes (POLG, TWNK, RRM2B). Host-infection genetics like severe COVID-19 (the 3p21.31 introgressed haplotype), tuberculosis (HLA, SLC11A1), and malaria resistance via Duffy/ACKR1, G6PD, α/β-thalassaemia, and HbE are captured and properly phased from the same assay [29], [30].

3.4. Southeast Asian clinical context

Southeast Asia comprises eleven countries (Brunei, Cambodia, Indonesia, Laos, Malaysia, Myanmar, Philippines, Singapore, Thailand, Timor-Leste, Vietnam) with a combined population of 680 million. The region carries some of the world's highest prevalence of certain monogenic disorders and pharmacogenetic risk variants, yet is dramatically underrepresented in major reference resources (gnomAD, 1000 Genomes Project, UK Biobank, All of Us). This combination with high genetic disease burden plus poor reference coverage makes SEA both an urgent priority for HiFi WGS deployment and a region where the technology's advantages are most acute.

3.4.1. Hemoglobinopathies

Thalassaemia and hemoglobinopathies are by far the most prevalent monogenic disorders in SEA, with α-thalassaemia carrier frequencies of 30–40% in Northern Thailand and Laos, β-thalassaemia 1–9% across most of the region, and HbE gene frequencies of 50–60% at the Laos–Cambodia–Thailand junction. The most common deletional α-thalassaemia allele for the Southeast Asian (SEA) deletion together with the –α^3.7 and –α^4.2 deletions, Hb Constant Spring, and the HbE β-globin variant, defines the regional clinical profile.

Traditional diagnostics rely on PCR-RDB, gap-PCR, and MLPA, each blind to rare or novel variants. Multiple groups have demonstrated that PacBio HiFi long-molecule sequencing accurately detects all common α- and β-thalassaemia deletions plus

rare and novel SNVs and SVs missed by conventional assays, including newly described −α^3.7 subtype III, ααα^anti3.7 triplications, and 7–14 kb δβ-fusion deletions [31], [32], [33]. The Thai pilot study by Singha and colleagues [34] established third-generation sequencing as feasible for population-level thalassaemia screening. The Singapore clinical case reported by Chin et al. [35] demonstrated long-read sequencing utility in resolving ambiguous β-thalassaemia carrier status in a population with 0.9% β-thalassaemia prevalence. A single HiFi WGS replaces this entire diagnostic stack and simultaneously delivers actionable pharmacogenomic and risk information.

3.4.2. Severe Cutaneous Adverse Reactions and HLA Pharmacogenetics

1. *HLA-B*15:02 and carbamazepine.* The CPIC HLA-B/carbamazepine guideline [36], [37] and subsequent population studies establish that HLA-B*15:02 confers a pooled odds ratio of ~113 (95% CI 51–251) for carbamazepine-induced SJS/TEN in Asian patients. Allele frequencies are 10.2% in Han Chinese, 10% in Taiwanese, with high frequencies across Thai, Malay, Indonesian, Vietnamese, and Filipino populations, but <0.1% in Caucasians. In Singapore, mandated HLA-B*15:02 genotyping reduced carbamazepine-associated SJS/TEN by >90% [38]. In Indonesia, HLA-B*15:02 carriage was 66.7% in patients with CBZ-induced SJS/TEN vs. 22.9% in healthy controls [39]. In Vietnam, HLA-B*15:02 increased SJS risk 12.5-fold in CBZ-treated patients. In Malaysia, HLA-B*15:13 has additionally been identified as an ethnic-specific risk marker. A single HiFi WGS with StarPhase delivers 4-field HLA-A and HLA-B diplotypes including B*15:02, B*15:13, B*15:11, and A*31:01 (the second relevant carbamazepine HLA marker).

2. *HLA-B*58:01 and allopurinol.* Strong predisposition to allopurinol-induced SJS/TEN and DRESS [40], with allele frequencies ~5–10% in Han Chinese and 5–15% across SEA Han Chinese diaspora populations and several Indigenous SEA groups. Recommended screening before allopurinol initiation in CPIC and many SEA national guidelines.

3. *NUDT15 thiopurine sensitivity*. *NUDT15*3* (c.415C>T, p.Arg139Cys) has allele frequencies of 5–10% in East Asian and SEA populations vs. <1% in Europeans [41], and is the single most clinically important predictor of severe thiopurine-induced myelosuppression in regional patients with inflammatory bowel disease, acute lymphoblastic leukemia, and rheumatologic disease.

4. *CYP2C19*. **2* and **3* loss-of-function alleles are 10–35% common in SEA populations, with clinical consequences for clopidogrel and PPI therapies. ***NAT2*** slow-acetylator polymorphisms are highly relevant for isoniazid dosing in the regional TB burden. ***CYP2D6*** ultra-rapid and poor-metabolizer phenotypes vary widely across the region.

3.4.3. SEA-Specific Disease Burden

Tuberculosis: high regional burden, NAT2 acetylator typing reduces hepatotoxicity from isoniazid.

Hepatitis B: endemic at carrier frequencies 5–15% across SEA; host genetic modifiers of clearance and HCC risk include HLA-DPB1, HLA-DQB1, *STAT4*, and *KIF1B*.

Malaria: endemic across Indonesia, Myanmar, Cambodia, Laos, Vietnam, and the Philippines; resistance loci (Duffy/ACKR1, G6PD, hemoglobinopathies) are intensely selected.

G6PD deficiency: prevalence of 3–15% across SEA, with population-specific variants (G6PD Mahidol in Thailand/Cambodia, G6PD Viangchan in Laos, G6PD Mediterranean in some Southern Asian populations) [30] clinically critical for primaquine therapy, sulphonamides, and drugs causing oxidative hemolysis.

3.4.4. Repeat Expansion Disorders in SEA

SCA3 (Machado–Joseph disease) is markedly enriched in Han Chinese, Taiwanese, and Singaporean Chinese populations. SCA1, SCA2, SCA6, SCA12, and DRPLA show population-specific distributions across SEA. NIID (*NOTCH2NLC* GGC expansion) was molecularly defined in East Asian cohorts [23] and is increasingly recognized across SEA. BAFME-1 (*SAMD12*) [24] is enriched in Japanese, Chinese, and likely SEA populations.

3.4.5. Population Genetic Resources

GenomeAsia 100K. The pilot phase [9] sequenced 1,739 individuals from 219 population groups across 64 countries, with substantial SEA representation including 156 Malaysians, 68 Indonesians, 52 Filipinos, and others. All current GenomeAsia data are short-read.

SG10K_Health. The Singapore National Precision Medicine (NPM) program sequenced ~10,000 Singaporeans across Chinese (58.4%), Malay (21.8%), and Indian backgrounds at ~13.7× coverage in the pilot [42], with subsequent expansion to SG10K_Health [10]. Companion publications include analysis of clinically relevant variants in ancestrally diverse Asian genomes [43] and a structural variation catalogue [44].

Indonesian Genome Diversity Project, Thai Reference Genome, Vietnamese Genome Project. Emerging short-read population resources establishing baseline allele frequencies but underpowered for SVs, tandem repeats, and methylation.

HUGO Pan-Asian SNP Initiative and Singapore Genome Variation Project. Historical genotype-array resources are still cited for ancestry-specific allele frequencies.

3.4.6. Founder Mutations and Ancestry-Specific Variant Interpretation

SEA harbors numerous founder mutations and population-specific variants that simultaneously:

Inflate the population-allele-frequency–weighted estimate of "rare disease–causing variants" in clinical settings, since some variants are common in SEA but absent or extremely rare in gnomAD.

Underlie ACMG/AMP variant-classification difficulties [45], as PM2 (absent from controls) and BS1/BA1 (allele frequency too common for disease) criteria depend on representative reference data.

Are systematically misclassified by clinical interpretation pipelines built on Caucasian-dominated reference data.

GnomAD v4 contains ~807,000 individuals globally but <1% are from SEA, dramatically inflating variant-of-uncertain-significance (VUS) rates in regional patients. Population-specific HiFi WGS reference panels analogous to SG10K but with HiFi data are essential to convert VUSs to actionable calls.

3.4.7. Clinical Practice Patterns

Clinical genomics adoption across SEA is heterogeneous:

– Singapore (mature, NPM-led precision-medicine program);

– Malaysia and Thailand (tertiary academic centers with growing clinical exome capacity);
– Indonesia (emerging, mostly tertiary referral);
– Vietnam and the Philippines (rapidly developing);
– Cambodia, Laos, Myanmar, Brunei, Timor-Leste (limited tertiary capacity, dependent on referral).

The deployment opportunity is therefore neither uniform nor reducible to a single regional strategy; HiFi WGS adoption should be calibrated to local clinical infrastructure, regulatory frameworks, and reimbursement environments.

Table 2 - Selected clinically actionable variants of high relevance to Southeast Asian populations

DOI: https://doi.org/10.60797/jbg.2026.33.7.2

Gene/Locus	Variant/Allele	Clinical implication	SEA AF (%)	HiFi value
HBA1/HBA2	–SEA, –α3.7, –α4.2	α-thalassemia	5–40	Resolves segdup; SV native
HBB	Codon 41/42, IVS1-1, HbE	β-thalassemia, HbE	1–9 (HbE up to 60)	Direct phasing
G6PD	Mahidol, Viangchan, Mediterranean	Hemolysis risk	3–15	Pan-allele detection
HLA-B	*15:02	Carbamazepine SJS/TEN	5–15	StarPhase 4-field
HLA-B	*58:01	Allopurinol SCAR	5–15	StarPhase 4-field
NUDT15	*3 (c.415C>T)	Thiopurine toxicity	5–10	Direct call
CYP2C19	*2, *3	Clopidogrel, PPI	10–35	StarPhase diplotype
NAT2	Slow acetylator	Isoniazid hepatotoxicity	30–60	Direct phasing
ATXN3	CAG expansion	SCA3 (Machado–Joseph)	Variable	TRGT sizing
NOTCH2NLC	GGC expansion	NIID	Variable	GC-rich, HiFi essential
SAMD12	TTTCA/TTTTA	BAFME-1	Variable	Intronic, HiFi essential

3.5. Comparative diagnostic yield

3.5.1. Head-to-Head Studies

The most rigorous direct comparison to date is the Radboudumc 100-sample study [5]. In a cohort of patients with 145 previously characterized clinically relevant variants spanning SNVs, indels, deletions, duplications, complex rearrangements, mobile-element insertions, repeat expansions, methylation-mediated variants, and uniparental disomy, HiFi WGS detected 93% with a single assay replicating in one test what previously required two to six orthogonal assays per patient. Critically, the study was designed as a sensitivity benchmark using a priori known variants. In a prospective cohort of previously undiagnosed patients, HiFi WGS would be expected to additionally reveal pathogenic variants invisible to short-read pipelines.

Several smaller series support similar conclusions:

1. Mantere, Kersten & Hoischen [12] — early review establishing long-read sequencing's emerging clinical role.
2. Mahmoud et al. [2] — comprehensive comparison of short- and long-read SV calling demonstrating dramatic gains in insertion detection and repeat-rich region resolution.
3. Beyter et al. [3] — population-scale ONT long-read SVs in 3,622 Icelanders identifying a median of 22,636 SVs per genome and characterizing disease-associated SVs (e.g., *PCSK9* deletion → low LDL; *ACAN* VNTR → height).
4. Liao et al. (HPRC) [7] — pangenome-based variant calling reduces small-variant errors by ~34% and improves SV calling sensitivity by >100% in challenging regions.
5. Nurk et al. (T2T-CHM13) [8] — telomere-to-telomere reference, primarily HiFi-assembled, adding 200 Mb of novel sequence and revealing variants invisible against GRCh38.
6. Aganezov et al. [46] — T2T-driven variant discovery in 3,202 1000 Genomes samples.
7. Logsdon, Vollger & Eichler [47] — definitive review of long-read sequencing applications.

3.5.2. Methylation Studies

Cheung et al. [6] — 276 rare-disease samples / 152 families; 80% of hypermethylation outliers allele-specific; identified novel repeat-expansion–driven promoter silencing, including a DMPK case in DM1.

Capper et al. [28] — methylation-based classification of CNS tumors, originally array-based but now extensible to HiFi-derived methylation calls.

3.5.3. Cost-Effectiveness

Although formal health-economic evaluations of HiFi WGS are still limited, early implementation studies suggest that the technology has the potential to improve the cost-effectiveness of rare disease diagnostics. In the Radboudumc setting, replacing 2–6 orthogonal short-read–based assays per patient with a single HiFi WGS demonstrably reduces total diagnostic cost per

case while shortening turnaround time. The economic value proposition strengthens further as HiFi reagent costs decrease with SPRQ-Nx chemistry and Vega platform deployment, and as automated library preparation (24–96 samples per run at Radboudumc) drives operational efficiency.

The cost calculation must include downstream value: a single HiFi WGS provides not only the immediate diagnostic answer but also lifelong pharmacogenomic, carrier-screening, and future-reanalysis utility from a single sample—a property no orthogonal assay stack offers.

3.6. Bioinformatics and analysis ecosystem

3.6.1. SMRT Link and Primary Analysis

SMRT Link is the PacBio-supplied primary analysis software, performing CCS read generation from raw subreads, demultiplexing, and basic QC. On Revio, primary analysis (including CpG methylation calling) executes on-instrument; SPRQ chemistry runs incorporate further sensitivity gains and 6mA calling.

3.6.2. Alignment

pbmm2 is the PacBio-tuned wrapper around minimap2 [48], providing optimized HiFi-read mapping to GRCh38, T2T-CHM13, or pangenome references. Pangenome-aware alignment (e.g., vg giraffe, Minigraph-Cactus [49]) further improves variant calling in segmentally duplicated and polymorphic regions.

3.6.3. Small-Variant Calling

DeepVariant [16] with the PacBio HiFi model is the de facto standard, yielding SNV and indel precision/recall ≥99.9% in GIAB high-confidence regions. Pangenome-aware DeepVariant [50] further improves performance, particularly in challenging regions and underrepresented ancestries. HiPhase performs read-based haplotype phasing simultaneously with variant calling, producing phase blocks with N50 typically >100 kb.

3.6.4. Structural Variant Calling

sawfish (PacBio-supplied SV caller) and Sniffles2 [13] are the leading HiFi SV callers. Sniffles2 implements repeat-aware clustering, coverage-adaptive filtering, and explicit support for mosaic and population-level SV calling, being 11.8× faster and 29% more accurate than predecessors across both ONT and HiFi data at 5–50× coverage. Population-scale SV joint-calling and SV merging is supported by tools such as jasmine and SVDB.

3.6.5. Tandem Repeat Genotyping

TRGT (Tandem Repeat Genotyping Tool) [18] performs targeted genotyping of catalogued tandem repeats from HiFi data, returning size, sequence composition, mosaicism, and CpG methylation per repeat allele. The companion TRGT-denovo [19] provides accurate de novo tandem-repeat mutation detection in trio data. TRVZ visualizes repeat alleles. Mendelian concordance on >937,000 catalogued repeats: 99.56%.

3.6.6. Segmental Duplication and Paralogue Resolution

Paraphase [26] extracts HiFi reads aligned to any member of a paralogue family, realigns them to a single archetype gene, phases haplotypes, and performs variant calling per haplotype. Application to 316 medically relevant paralogues across 160 segmental duplications has revealed previously undetected de novo SNVs, non-allelic gene conversion events, and copy number variability across ancestries, with direct clinical impact on SMN1/SMN2 (SMA), PMS2/PMS2CL (Lynch syndrome), CYP2D6/CYP2D7 (PGx), NCF1/NCF1B/NCF1C (CGD), PKD1/PKD1P1–6 (ADPKD), HBA1/HBA2 (α-thalassemia), and others.

3.6.7. Pharmacogenomic Typing

StarPhase [27] is the HiFi-native pharmacogenomic diplotyper, providing CPIC Level A diplotypes for 21 genes plus detailed 4-field HLA-A, HLA-B, and CYP2D6 typing. Pangu is a companion PacBio tool for CYP2D6 typing. PharmCAT integrates diplotype calls with CPIC guidelines for clinical reporting.

3.6.8. Methylation Analysis

CpG methylation is called natively during primary analysis on Revio (and via secondary tools on Sequel IIe data). Downstream analysis tools are available in the MethBat tools collection, and with modbamtools for visualization. The Cheung et al. [6] framework for outlier hypermethylation detection in rare disease has become an analytical template for clinical methylation reporting.

3.6.9. Pangenome References

The HPRC v1.0 draft pangenome [7] and the ongoing v1.1/2.0 releases support pangenome-aware variant calling via tools such as vg giraffe + DeepVariant, PanGenie [51], and Minigraph-Cactus [49]. For SEA, the HPRC currently includes limited SEA representation (Indonesian and Vietnamese samples in subsequent expansion phases); SEA-specific pangenome resources remain a high-priority gap.

3.6.10. Variant Prioritization and Clinical Interpretation

Variant prioritization for clinical reporting integrates ACMG/AMP classification [45], ClinVar, gnomAD, OMIM, HGMD, and disease-specific gene panels. SEA-specific challenges (Section 5.6) demand careful regional allele-frequency overlays. AI-assisted phenotype-driven variant prioritization (Exomiser [52], LIRICAL, AMELIE, and HPO-based tools [53]) substantially improves diagnostic-yield triage of HiFi WGS variant lists, which can exceed 4–5 million variants per genome.

3.6.11. Data Storage and Compute

A 30× HiFi WGS BAM is typically 50 GB (including methylation); compute requirements for full HiFi WGS analysis (alignment, small-variant, SV, repeat, methylation, PGx) are typically 100–500 CPU-hours per sample on standard cloud or HPC infrastructure. For SEA deployment, regional cloud or government-hosted compute capacity (Singapore's NSCC; Malaysian and Thai national HPC) supports clinical-scale operation; data residency and privacy regulations vary by country.

Table 3 - Key bioinformatics tools for clinical HiFi WGS analysis

DOI: https://doi.org/10.60797/jbg.2026.33.7.3

Task	Tool	Function	Reference
Alignment	pbmm2 / minimap2	HiFi-tuned long-read alignment	[48]
SNV/Indel calling	DeepVariant (HiFi model)	Deep-learning small-variant caller	[16]
Phasing	HiPhase	Read-based haplotype phasing	PacBio
SV calling	Sniffles2 / pbsv	Structural variant detection (incl. mosaic)	[13]
Tandem repeats	TRGT / TRGT-denovo	Repeat sizing + methylation + de novo	[18]
Paralogues	Paraphase	Segmental duplication resolution	[26]
Pharmacogenomics	StarPhase / Pangu / PharmCAT	PGx diplotypes + HLA + CYP2D6	[27]
Methylation	modkit / methbat	5mC/6mA calling and analysis	PacBio/community
Pangenome calling	vg giraffe / PanGenie / Minigraph-Cactus	Pangenome-aware variant calling	[7], [51]
Variant interpretation	Exomiser / LIRICAL / AMELIE	HPO-driven prioritization	[52]

3.7. Future directions and translational outlook

Methylation, a free by-product of every run, will become a routine reporting layer within 3–5 years as pathogenic catalogues grow [6], [28]. Pangenome-based clinical calling is technically near-mature, pending regulatory acceptance, ClinVar/HGMD updates, and SEA-specific pangenome representation. AI-assisted interpretation (AlphaMissense, Exomiser) will be required for non-coding, complex-SV, and methylation variant classes. Genomic newborn screening (BabySeq, GUARDIAN, the UK Generation Study) is well suited to SEA's high-prevalence carrier disorders. Population programs including Singapore Precise, Malaysia, Thailand, and Indonesia's BGSi will define regional adoption, whether as first-line testing or second-line resequencing of short-read diagnostic gaps. Rapid HiFi workflows are also emerging, approaching the ~7-hour short-read critical-care benchmark set by Stanford and Rady, with regional rapid-genome services feasible in Singapore, Bangkok, and Kuala Lumpur. Synchronised long-read multi-omics has already resolved previously unsolved Mendelian disease. Vollger et al. combined genome, methylome, chromatin accessibility, and transcriptome to explain a balanced X;13 translocation via fusion-transcript formation, enhancer adoption, and aberrant X-inactivation [125] pointing toward HiFi as a precision multi-omics platform. HiFi cfDNA additionally enables minimal-residual-disease and hepatocellular-carcinoma surveillance, regionally relevant given the hepatitis B burden.

Limitations

Despite its substantial advantages, HiFi WGS is not free of meaningful limitations. Honest critical appraisal is essential to calibrate clinical expectations and roadmap deployment. HiFi reagents remain ~2–4× short-read cost favourable per diagnosis against the multi-assay stack it replaces, but still a barrier without reimbursement in lower-middle-income SEA countries. Throughput (~4-5 genomes/24 h on Revio) limits population scale outside Singapore, though higher-throughput long-read platforms are emerging. Library preparation needs high-molecular-weight DNA (SPRQ has lowered input to ≤500 ng), limiting FFPE/archival use. A 30× genome yields ~4-5 million variants, and classification frameworks remain weaker for long-read-resolved classes (complex SVs in segmental duplications, non-canonical repeats, methylation signals), whereas difficulty amplifies for under-represented SEA genomes. Ancestry-specific reference gaps (limited SEA in gnomAD and the HPRC pangenome) inflate VUS and "novel"-SV rates.

ELSI frameworks (consent, return of secondary findings, data sovereignty), EHR/HPO integration, regulatory/IVD pathways (long-read sequencing is recognised by the FDA and EMA in research contexts and CLIA validation exists, but routine IVD adoption and national SEA approvals remain heterogeneous), and clinical-genomics workforce remain immature across much of SEA, and structured phenotyping (HPO) and longitudinal reanalysis pipelines are variably implemented. Residual technical gaps persist where acrocentric/rDNA/centromeric arrays still need complementary ONT ultra-long reads, mosaicism below ~1% VAF needs deep or targeted coverage, and some centromeric/telomeric balanced rearrangements need optical mapping or Hi-C. These residual gaps are areas of active development and are progressively narrowing.

Conclusion

As of 2026, PacBio HiFi whole-genome sequencing represents the most comprehensive single-assay clinical genomic test currently available. By producing 10–25 kb reads at Q30+ accuracy with simultaneous native methylation calling, HiFi WGS resolves the structural variants, tandem repeats, mobile-element insertions, paralogue-pseudogene complexes, and methylation

events that collectively a substantial fraction of clinically meaningful but short-read–invisible disease. The Radboudumc 100-sample cohort demonstrated 93% pathogenic-variant sensitivity against a stack of orthogonal assays, and the GA4K methylation study established native HiFi methylation as a clinically deployable analytical layer. The bioinformatics ecosystem such as DeepVariant, Sniffles2, sawfish, TRGT, Paraphase, StarPhase, HiPhase, and the HPRC pangenome has matured to clinical-grade specifications.

For Southeast Asia, the case for HiFi WGS is unusually compelling. The regional disease burden for thalassaemias, hemoglobinopathies, G6PD deficiency, congenital adrenal hyperplasia, SMA, repeat-expansion ataxias, NIID, and others overwhelmingly involves variant classes for which HiFi WGS provides decisive advantages. The pharmacogenomic risk profile such as HLA-B*15:02, HLA-B*58:01, NUDT15, CYP2C19, NAT2 is exactly the class of high-impact variation for which StarPhase from a single HiFi run delivers pan-pharmacogenome typing. Population genomic under-representation in gnomAD, HPRC, and other Caucasian-dominated resources further amplifies the per-sample diagnostic value of HiFi WGS in SEA patients.

Real obstacles remain including cost, throughput, workforce, regulatory frameworks, and SEA-specific reference resources should be honestly acknowledged rather than minimized. Yet none is intrinsically intractable on current technology trajectories. The scientific framework supporting HiFi WGS as a first-tier clinical test is, in our judgment, now established beyond reasonable contest. The remaining question is one of implementation: how rapidly the SEA region can build the population reference data, regulatory pathways, reimbursement architectures, and clinical workforce required to deliver this technology equitably to the patients who stand to benefit most.

Благодарности

YSDS выражает признательность мировому научному сообществу, чьи опубликованные работы легли в основу доказательной базы данного обзора, включая исследователей из Медицинского центра Университета Радбауда, детской больницы Children’s Mercy в Канзас-Сити, консорциумов Human Pangenome Reference Consortium и Telomere-to-Telomere Consortium, проектов GenomeAsia 100K и SG10K_Health, компании Pacific Biosciences, а также участников многочисленных национальных инициатив в области геномной медицины в странах Юго-Восточной Азии.

Конфликт интересов

Не указан.

Рецензия

Все статьи проходят рецензирование. Но рецензент или автор статьи предпочли не публиковать рецензию к этой статье в открытом доступе. Рецензия может быть предоставлена компетентным органам по запросу.

Acknowledgement

YSDS gratefully acknowledges the global scientific community whose published work forms the evidence base of this Review, including investigators at Radboud University Medical Center, Children’s Mercy Kansas City, the Human Pangenome Reference Consortium, the Telomere-to-Telomere Consortium, GenomeAsia 100K, SG10K_Health, Pacific Biosciences, and the many SEA national genomic medicine initiatives.

Conflict of Interest

None declared.

Review

All articles are peer-reviewed. But the reviewer or the author of the article chose not to publish a review of this article in the public domain. The review can be provided to the competent authorities upon request.

Список литературы на английском языке / References in English

1. Ebbert M.T.W. Systematic analysis of dark and camouflaged genes reveals disease-relevant genes hiding in plain sight / M.T.W. Ebbert, T.D. Jensen, K. Jansen-West [et al.] // Genome Biology. — 2019. — Vol. 20. — P. 97. — DOI: 10.1186/s13059-019-1707-2.

2. Mahmoud M. Structural variant calling: the long and the short of it / M. Mahmoud, N. Gobet, D.I. Cruz-Dávalos [et al.] // Genome Biology. — 2019. — Vol. 20. — P. 246. — DOI: 10.1186/s13059-019-1828-7.

3. Beyter D. Long-read sequencing of 3,622 Icelanders provides insight into the role of structural variants in human diseases and other traits / D. Beyter, H. Ingimundardottir, A. Oddsson [et al.] // Nature Genetics. — 2021. — Vol. 53. — P. 779–786. — DOI: 10.1038/s41588-021-00865-4.

4. Wenger A.M. Accurate circular consensus long-read sequencing improves variant detection and assembly of a human genome / A.M. Wenger, P. Peluso, W.J. Rowell [et al.] // Nature Biotechnology. — 2019. — Vol. 37. — P. 1155–1162. — DOI: 10.1038/s41587-019-0217-9.

5. Höps W. HiFi long-read genomes for difficult-to-detect, clinically relevant variants / W. Höps, M.M. Weiss, R. Derks [et al.] // American Journal of Human Genetics. — 2025. — Vol. 112. — P. 73–86. — DOI: 10.1016/j.ajhg.2024.12.017.

6. Cheung W.A. Direct haplotype-resolved 5-base HiFi sequencing for genome-wide profiling of hypermethylation outliers in a rare disease cohort / W.A. Cheung, A.F. Johnson, W.J. Rowell [et al.] // Nature Communications. — 2023. — Vol. 14. — P. 3090. — DOI: 10.1038/s41467-023-38782-1.

7. Liao W.-W. A draft human pangenome reference / W.-W. Liao, M. Asri, J. Ebler // Nature. — 2023. — Vol. 617. — P. 312–324. — DOI: 10.1038/s41586-023-05896-x.

8. Nurk S. The complete sequence of a human genome / S. Nurk, S. Koren, A. Rhie [et al.] // Science. — 2022. — Vol. 376. — P. 44–53. — DOI: 10.1126/science.abj6987.

9. Wall J.D. The GenomeAsia 100K Project enables genetic discoveries across Asia / J.D. Wall, E.W. Stawiski, A. Ratan // Nature. — 2019. — Vol. 576. — P. 106–111. — DOI: 10.1038/s41586-019-1793-z.

10. Wong E. The Singapore National Precision Medicine strategy / E. Wong, N. Bertin, M. Hebrard [et al.] // Nature Genetics. — 2023. — Vol. 55. — P. 178–186. — DOI: 10.1038/s41588-022-01274-x.

11. Eid J. Real-time DNA sequencing from single polymerase molecules / J. Eid, A. Fehr, J. Gray [et al.] // Science. — 2009. — Vol. 323. — P. 133–138. — DOI: 10.1126/science.1162986.

12. Mantere T. Long-read sequencing emerging in medical genetics / T. Mantere, S. Kersten, A. Hoischen // Frontiers in Genetics. — 2019. — Vol. 10. — P. 426. — DOI: 10.3389/fgene.2019.00426.

13. Smolka M. Detection of mosaic and population-level structural variants with Sniffles2 / M. Smolka, L.F. Paulin, C.M. Grochowski [et al.] // Nature Biotechnology. — 2024. — Vol. 42. — P. 1571–1580. — DOI: 10.1038/s41587-023-02024-y.

14. Olson N.D. PrecisionFDA Truth Challenge V2: calling variants from short and long reads in difficult-to-map regions / N.D. Olson, J. Wagner, J. McDaniel [et al.] // Cell Genomics. — 2022. — Vol. 2. — P. 100129. — DOI: 10.1016/j.xgen.2022.100129.

15. Choy L.Y.L. Single-molecule sequencing enables long cell-free DNA detection and direct methylation analysis for cancer patients / L.Y.L. Choy, W. Peng, P. Jiang [et al.] // Clinical Chemistry. — 2022. — Vol. 68. — P. 1151–1163. — DOI: 10.1093/clinchem/hvac086.

16. Poplin R. A universal SNP and small-indel variant caller using deep neural networks / R. Poplin, P.-C. Chang, D. Alexander [et al.] // Nature Biotechnology. — 2018. — Vol. 36. — P. 983–987. — DOI: 10.1038/nbt.4235.

17. Wagner J. Curated variation benchmarks for challenging medically relevant autosomal genes / J. Wagner, N.D. Olson, L. Harris [et al.] // Nature Biotechnology. — 2022. — Vol. 40. — P. 672–680. — DOI: 10.1038/s41587-021-01158-1.

18. Dolzhenko E. Characterization and visualization of tandem repeats at genome scale (TRGT) / E. Dolzhenko, A. English, H. Dashnow [et al.] // Nature Biotechnology. — 2024. — DOI: 10.1101/2023.05.12.540470.

19. Mokveld T. TRGT-denovo: accurate detection of de novo tandem repeat mutations / T. Mokveld, E. Dolzhenko, H. Dashnow [et al.] // PLoS Computational Biology/Bioinformatics. — 2024. — PMID: 39071386.

20. Thiffault I. Clinical long-read sequencing test for genetic disease diagnosis / I. Thiffault, E. Farrow, C. Barrett [et al.] // Clinical validation report. — 2025. — PMID: 40982260.

21. Smedley D. 100,000 Genomes Pilot on Rare-Disease Diagnosis in Health Care — Preliminary Report / D. Smedley, K.R. Smith, A. Martin // New England Journal of Medicine. — 2021. — Vol. 385. — P. 1868–1880. — DOI: 10.1056/NEJMoa2035790.

22. Stark Z. Australian Genomics: outcomes of a 5-year national program to accelerate the implementation of genomic medicine / Z. Stark, R.E. Foulger, E. Williams [et al.] // American Journal of Human Genetics. — 2023. — Vol. 110. — P. 419–426. — DOI: 10.1016/j.ajhg.2023.01.018.

23. Sone J. Long-read sequencing identifies GGC repeat expansions in NOTCH2NLC associated with neuronal intranuclear inclusion disease / J. Sone, S. Mitsuhashi, A. Fujita [et al.] // Nature Genetics. — 2019. — Vol. 51. — P. 1215–1221. — DOI: 10.1038/s41588-019-0459-y.

24. Ishiura H. Expansions of intronic TTTCA and TTTTA repeats in benign adult familial myoclonic epilepsy / H. Ishiura, K. Doi, J. Mitsui [et al.] // Nature Genetics. — 2018. — Vol. 50. — P. 581–590. — DOI: 10.1038/s41588-018-0067-2.

25. Coassin S. A comprehensive map of single-base polymorphisms in the hypervariable LPA kringle IV type 2 copy number variation region / S. Coassin, S. Schönherr, H. Weissensteiner [et al.] // Journal of Lipid Research. — 2019. — Vol. 60. — P. 186–199. — DOI: 10.1194/jlr.M090381.

26. Chen X. Genome-wide profiling of highly similar paralogous genes using HiFi sequencing (Paraphase) / X. Chen, D. Baker, E. Dolzhenko [et al.] // Nature Communications. — 2025. — Vol. 16. — P. 2340. — DOI: 10.1038/s41467-025-57505-2.

27. Holt J.M. StarPhase: comprehensive phase-aware pharmacogenomic diplotyper for long-read sequencing data / J.M. Holt, B. Wilk, C.L. Birch [et al.] // bioRxiv. — 2024. — Art. ID 2024.12.10.627527. — DOI: 10.1101/2024.12.10.627527.

28. Capper D. DNA methylation-based classification of central nervous system tumours / D. Capper, D.T.W. Jones, M. Sill [et al.] // Nature. — 2018. — Vol. 555. — P. 469–474. — DOI: 10.1038/nature26000.

29. COVID-19 Host Genetics Initiative. Mapping the human genetic architecture of COVID-19 // Nature. — 2021. — Vol. 600. — P. 472–477. — DOI: 10.1038/s41586-021-03767-x.

30. Howes R.E. G6PD deficiency prevalence and estimates of affected populations in malaria endemic countries: a geostatistical model-based map / R.E. Howes, F.B. Piel, A.P. Patil [et al.] // PLoS Medicine. — 2012. — Vol. 9. — P. e1001339. — DOI: 10.1371/journal.pmed.1001339.

31. Xu L. Long-molecule sequencing: a new approach for identification of clinically significant DNA variants in α- and β-thalassemia carriers / L. Xu, A. Mao, H. Liu [et al.] // Journal of Molecular Diagnostics. — 2020. — Vol. 22. — P. 1087–1095. — DOI: 10.1016/j.jmoldx.2020.05.004.

32. Liang Q. A more universal approach to comprehensive analysis of thalassemia alleles (CATSA) / Q. Liang, W. Gu, P. Chen [et al.] // Journal of Molecular Diagnostics. — 2021. — Vol. 23. — P. 1195–1204. — DOI: 10.1016/j.jmoldx.2021.06.008.

33. Qin D. Evaluating the clinical utility of a long-read sequencing-based approach in prenatal diagnosis of thalassemia / D. Qin, A. Mao, Q. Huang [et al.] // Clinical Chemistry. — 2023. — Vol. 69. — P. 239–250. — DOI: 10.1093/clinchem/hvac199.

34. Singha K. Thalassemia screening by third-generation sequencing: pilot study in a Thai population / K. Singha, S. Fucharoen, K. Sanchaisuriya [et al.] // Hematology / PLoS One. — 2024. — PMCID: PMC11110746.

35. Chin H.-L. Clinical application of targeted long-read sequencing in prenatal β-thalassemia testing and genetic counseling / H.-L. Chin, P.S. Lai, C.S.K. Yong [et al.] // Molecular Genetics & Genomic Medicine. — 2024. — Vol. 12. — P. e2285. — DOI: 10.1002/mgg3.2285.

36. Leckband S.G. Clinical Pharmacogenetics Implementation Consortium guidelines for HLA-B genotype and carbamazepine dosing / S.G. Leckband, J.R. Kelsoe, H.M. Dunnenberger [et al.] // Clinical Pharmacology & Therapeutics. — 2013. — Vol. 94. — P. 324–328. — PMID: 23695185.

37. Phillips E.J. Clinical Pharmacogenetics Implementation Consortium guideline for HLA genotype and use of carbamazepine and oxcarbazepine: 2017 update / E.J. Phillips, C. Sukasem, M. Whirl-Carrillo [et al.] // Clinical Pharmacology & Therapeutics. — 2018. — Vol. 103. — P. 574–581. — DOI: 10.1002/cpt.1004.

38. Sung C. Usage pattern of carbamazepine and associated severe cutaneous adverse reactions in Singapore following implementation of HLA-B*15:02 genotyping as standard-of-care / C. Sung, L. Tan, M. Limenta [et al.] // Frontiers in Pharmacology. — 2020. — Vol. 11. — P. 527. — DOI: 10.3389/fphar.2020.00527.

39. Yuliwulandari R. Association of the HLA-B alleles with carbamazepine-induced Stevens–Johnson syndrome/toxic epidermal necrolysis in the Javanese and Sundanese population of Indonesia / R. Yuliwulandari, E. Kristin, K. Prayuni [et al.] // Pharmacogenomics. — 2017. — Vol. 18. — P. 1643–1652. — DOI: 10.2217/pgs-2017-0103.

40. Hung S.-I. HLA-B*5801 allele as a genetic marker for severe cutaneous adverse reactions caused by allopurinol / S.-I. Hung, W.-H. Chung, L.-B. Liou [et al.] // Proceedings of the National Academy of Sciences of the United States of America. — 2005. — Vol. 102. — P. 4134–4139. — DOI: 10.1073/pnas.0409500102.

41. Yang J.-J. Inherited NUDT15 variant is a genetic determinant of mercaptopurine intolerance in children with acute lymphoblastic leukemia / J.-J. Yang, W. Landier, W. Yang [et al.] // Journal of Clinical Oncology. — 2015. — Vol. 33. — P. 1235–1242. — DOI: 10.1200/JCO.2014.59.4671.

42. Wu D. Large-scale whole-genome sequencing of three diverse Asian populations in Singapore / D. Wu, J. Dou, X. Chai [et al.] // Cell. — 2019. — Vol. 179. — P. 736–749. — DOI: 10.1016/j.cell.2019.09.019.

43. Chan S.H. Analysis of clinically relevant variants from ancestrally diverse Asian genomes / S.H. Chan, Y. Bylstra, J.X. Teo [et al.] // Nature Communications. — 2022. — Vol. 13. — P. 6694. — DOI: 10.1038/s41467-022-34116-9.

44. Tan J.H.J. A catalogue of structural variation across ancestrally diverse Asian genomes / J.H.J. Tan, Z. Li, M. Gonzalez Porta [et al.] // Nature Communications. — 2024. — Vol. 15. — P. 10. — DOI: 10.1038/s41467-024-53620-8.

45. Richards S. 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 / S. Richards, N. Aziz, S. Bale [et al.] // Genetics in Medicine. — 2015. — Vol. 17. — P. 405–424. — DOI: 10.1038/gim.2015.30.

46. Aganezov S. A complete reference genome improves analysis of human genetic variation / S. Aganezov, S.M. Yan, D.C. Soto [et al.] // Science. — 2022. — Vol. 376. — P. eabl3533. — DOI: 10.1126/science.abl3533.

47. Logsdon G.A. Long-read human genome sequencing and its applications / G.A. Logsdon, M.R. Vollger, E.E. Eichler // Nature Reviews Genetics. — 2020. — Vol. 21. — P. 597–614. — DOI: 10.1038/s41576-020-0236-x.

48. Li H. Minimap2: pairwise alignment for nucleotide sequences / H. Li // Bioinformatics. — 2018. — Vol. 34. — P. 3094–3100. — DOI: 10.1093/bioinformatics/bty191.

49. Hickey G. Pangenome graph construction from genome alignment with Minigraph-Cactus / G. Hickey, J. Monlong, J. Ebler [et al.] // Nature Biotechnology. — 2023. — Vol. 41. — P. 1393–1402. — DOI: 10.1038/s41587-023-01793-w.

50. Asri M. Pangenome-aware DeepVariant / M. Asri, P.-C. Chang, J.C. Mier [et al.] // bioRxiv. — 2025. — Art. ID 2025.06.05.657102. — DOI: 10.1101/2025.06.05.657102.

51. Ebler J. Pangenome-based genome inference allows efficient and accurate genotyping across a wide spectrum of variant classes (PanGenie) / J. Ebler, P. Ebert, W.E. Clarke [et al.] // Nature Genetics. — 2022. — Vol. 54. — P. 518–525. — DOI: 10.1038/s41588-022-01043-w.

52. Smedley D. Next-generation diagnostics and disease-gene discovery with the Exomiser / D. Smedley, J.O.B. Jacobsen, M. Jäger [et al.] // Nature Protocols. — 2015. — Vol. 10. — P. 2004–2015. — DOI: 10.1038/nprot.2015.124.

53. Robinson P.N. The Human Phenotype Ontology: a tool for annotating and analyzing human hereditary disease / P.N. Robinson, S. Köhler, S. Bauer [et al.] // American Journal of Human Genetics. — 2008. — Vol. 83. — P. 610–615. — DOI: 10.1016/j.ajhg.2008.09.017.

54. Marx V. Method of the year: long-read sequencing / V. Marx // Nature Methods. — 2023. — Vol. 20. — P. 6–11. — DOI: 10.1038/s41592-022-01730-w.

55. Hook P.W. Beyond assembly: the increasing flexibility of single-molecule sequencing technology / P.W. Hook, W. Timp // Nature Reviews Genetics. — 2023. — Vol. 24. — P. 627–641. — DOI: 10.1038/s41576-023-00600-1.

56. Mastrorosa F.K. Applications of long-read sequencing to Mendelian genetics / F.K. Mastrorosa, D.E. Miller, E.E. Eichler // Genome Medicine. — 2023. — Vol. 15. — P. 42. — DOI: 10.1186/s13073-023-01194-3.

57. Chaisson M.J.P. Multi-platform discovery of haplotype-resolved structural variation in human genomes / M.J.P. Chaisson, A.D. Sanders, X. Zhao [et al.] // Nature Communications. — 2019. — Vol. 10. — P. 1784. — DOI: 10.1038/s41467-018-08148-z.

58. Ebert P. Haplotype-resolved diverse human genomes and integrated analysis of structural variation / P. Ebert, P.A. Audano, Q. Zhu [et al.] // Science. — 2021. — Vol. 372. — DOI: 10.1126/science.abf7117.

59. Vollger M.R. Segmental duplications and their variation in a complete human genome / M.R. Vollger, X. Guitart, P.C. Dishuck [et al.] // Science. — 2022. — Vol. 376. — P. eabj6965. — DOI: 10.1126/science.abj6965.

60. De Coster W. Towards population-scale long-read sequencing / W. De Coster, M.H. Weissensteiner, F.J. Sedlazeck // Nature Reviews Genetics. — 2021. — Vol. 22. — P. 572–587. — DOI: 10.1038/s41576-021-00367-3.

61. Cohen A.S.A. Genomic answers for children: dynamic analyses of >1000 pediatric rare disease genomes / A.S.A. Cohen, E.G. Farrow, A.T. Abdelmoity [et al.] // Genetics in Medicine. — 2022. — Vol. 24. — P. 1336–1348. — DOI: 10.1016/j.gim.2022.02.007.

62. Gorzynski J.E. Ultrarapid nanopore genome sequencing in a critical care setting / J.E. Gorzynski, S.D. Goenka, K. Shafin [et al.] // New England Journal of Medicine. — 2022. — Vol. 386. — P. 700–702. — DOI: 10.1056/NEJMc2112090.

63. Lunke S. Integrated multi-omics for rapid rare disease diagnosis on a national scale / S. Lunke, S.E. Bouffler, C.V. Patel [et al.] // Nature Medicine. — 2023. — Vol. 29. — P. 1681–1691. — DOI: 10.1038/s41591-023-02401-9.

64. Chen X. Comprehensive SMN1 and SMN2 profiling for spinal muscular atrophy analysis using long-read PacBio HiFi sequencing / X. Chen, J. Harting, E. Farrow [et al.] // American Journal of Human Genetics. — 2023. — Vol. 110. — P. 240–250. — DOI: 10.1016/j.ajhg.2023.01.001.

65. Depienne C. 30 years of repeat expansion disorders: what have we learned and what are the remaining challenges? / C. Depienne, J.-L. Mandel // American Journal of Human Genetics. — 2021. — Vol. 108. — P. 764–785. — DOI: 10.1016/j.ajhg.2021.03.011.

66. Mitsuhashi S. Tandem-genotypes: robust detection of tandem repeat expansions from long DNA reads / S. Mitsuhashi, M.C. Frith, T. Mizuguchi [et al.] // Genome Biology. — 2019. — Vol. 20. — P. 58. — DOI: 10.1186/s13059-019-1667-6.

67. Stevanovski I. Comprehensive genetic diagnosis of tandem repeat expansion disorders with programmable targeted nanopore sequencing / I. Stevanovski, S.R. Chintalaphani, H. Gamaarachchi [et al.] // Science Advances. — 2022. — Vol. 8. — DOI: 10.1126/sciadv.abm5386.

68. Tang H. Profiling of short-tandem-repeat disease alleles in 12,632 human whole genomes / H. Tang, E.F. Kirkness, C. Lippert [et al.] // American Journal of Human Genetics. — 2017. — Vol. 101. — P. 700–715. — DOI: 10.1016/j.ajhg.2017.09.013.

69. Sedlazeck F.J. Accurate detection of complex structural variations using single-molecule sequencing (NGMLR/Sniffles) / F.J. Sedlazeck, P. Rescheneder, M. Smolka [et al.] // Nature Methods. — 2018. — Vol. 15. — P. 461–468. — DOI: 10.1038/s41592-018-0001-7.

70. Heller D. SVIM: structural variant identification using mapped long reads / D. Heller, M. Vingron // Bioinformatics. — 2019. — Vol. 35. — P. 2907–2915. — DOI: 10.1093/bioinformatics/btz041.

71. Eggertsson H.P. Graphtyper enables population-scale genotyping using pangenome graphs / H.P. Eggertsson, H. Jonsson, S. Kristmundsdottir [et al.] // Nature Genetics. — 2017. — Vol. 49. — P. 1654–1660. — DOI: 10.1038/ng.3964.

72. Garrison E. Variation graph toolkit improves read mapping by representing genetic variation in the reference / E. Garrison, J. Sirén, A.M. Novak [et al.] // Nature Biotechnology. — 2018. — Vol. 36. — P. 875–879. — DOI: 10.1038/nbt.4227.

73. Sirén J. Pangenomics enables genotyping of known structural variants in 5202 diverse genomes / J. Sirén, J. Monlong, X. Chang [et al.] // Science. — 2021. — Vol. 374. — P. abg8871. — DOI: 10.1126/science.abg8871.

74. Walsh R. Defining the genetic architecture of hypertrophic cardiomyopathy: re-evaluating the role of non-sarcomeric genes / R. Walsh, R. Buchan, A. Wilk [et al.] // European Heart Journal. — 2017. — Vol. 38. — P. 3461–3468. — DOI: 10.1093/eurheartj/ehw603.

75. Herman D.S. Truncations of titin causing dilated cardiomyopathy / D.S. Herman, L. Lam, M.R.G. Taylor [et al.] // New England Journal of Medicine. — 2012. — Vol. 366. — P. 619–628. — DOI: 10.1056/NEJMoa1110186.

76. Schmidt K. Structure, function, and genetics of lipoprotein(a) / K. Schmidt, A. Noureen, F. Kronenberg [et al.] // Journal of Lipid Research. — 2016. — Vol. 57. — P. 1339–1359. — DOI: 10.1194/jlr.R067314.

77. Ali H. PKD1 duplicated regions limit clinical utility of whole exome sequencing for genetic diagnosis of autosomal dominant polycystic kidney disease / H. Ali, M. Naim, S.R. Senum [et al.] // Scientific Reports. — 2019. — Vol. 9. — P. 4141. — DOI: 10.1038/s41598-019-40891-1.

78. Kingsmore S.F. A randomized, controlled trial of the analytic and diagnostic performance of singleton and trio, rapid genome and exome sequencing in ill infants / S.F. Kingsmore, J.A. Cakici, M.M. Clark [et al.] // American Journal of Human Genetics. — 2019. — Vol. 105. — P. 719–733. — DOI: 10.1016/j.ajhg.2019.08.009.

79. Owen M.J. Reclassification of the etiology of infant mortality with whole-genome sequencing / M.J. Owen, M.S. Wright, S. Batalov [et al.] // JAMA Network Open. — 2023. — Vol. 6. — P. e2254069. — DOI: 10.1001/jamanetworkopen.2022.54069.

80. Marwaha S. A guide for the diagnosis of rare and undiagnosed disease: beyond the exome / S. Marwaha, J.W. Knowles, E.A. Ashley // Genome Medicine. — 2022. — Vol. 14. — P. 23. — DOI: 10.1186/s13073-022-01026-w.

81. Sherman R.M. Pan-genomics in the human genome era / R.M. Sherman, S.L. Salzberg // Nature Reviews Genetics. — 2020. — Vol. 21. — P. 243–254. — DOI: 10.1038/s41576-020-0210-7.

82. Caudle K.E. Incorporation of pharmacogenomics into routine clinical practice: the Clinical Pharmacogenetics Implementation Consortium (CPIC) guideline development process / K.E. Caudle, T.E. Klein, J.M. Hoffman [et al.] // Current Drug Metabolism. — 2014. — Vol. 15. — P. 209–217. — DOI: 10.2174/1389200215666140130124910.

83. Pratt V.M. CYP2D6 genotyping recommendations: joint consensus recommendation of AMP, CAP, DPWG, and ESPT / V.M. Pratt, L.H. Cavallari, M.L. Fulmer [et al.] // Journal of Molecular Diagnostics. — 2021. — Vol. 23. — P. 1047–1064. — DOI: 10.1016/j.jmoldx.2021.05.013.

84. Bycroft C. The UK Biobank resource with deep phenotyping and genomic data / C. Bycroft, C. Freeman, D. Petkova [et al.] // Nature. — 2018. — Vol. 562. — P. 203–209. — DOI: 10.1038/s41586-018-0579-z.

85. All of Us Research Program Investigators. The "All of Us" Research Program / All of Us Research Program Investigators // New England Journal of Medicine. — 2019. — Vol. 381. — P. 668–676. — DOI: 10.1056/NEJMsr1809937.

86. Halldorsson B.V. The sequences of 150,119 genomes in the UK Biobank / B.V. Halldorsson, H.P. Eggertsson, K.H.S. Moore [et al.] // Nature. — 2022. — Vol. 607. — P. 732–740. — DOI: 10.1038/s41586-022-04965-x.

87. Karczewski K.J. The mutational constraint spectrum quantified from variation in 141,456 humans (gnomAD) / K.J. Karczewski, L.C. Francioli, G. Tiao [et al.] // Nature. — 2020. — Vol. 581. — P. 434–443. — DOI: 10.1038/s41586-020-2308-7.

88. Sukasem C. Pharmacogenomics of drug-induced hypersensitivity reactions: challenges, opportunities and clinical implementation / C. Sukasem, A. Puangpetch, S. Medhasi [et al.] // Asian Pacific Journal of Allergy and Immunology. — 2014. — Vol. 32. — P. 111–123.

89. Kim S.-Y. HLA-B*15:11 is associated with carbamazepine-induced Stevens–Johnson syndrome in Korean and Japanese patients / S.-Y. Kim, S.-Y. Lee, D.-W. Kim [et al.] // Epilepsia. — 2011. — Vol. 52. — P. e135–e139. — DOI: 10.1111/j.1528-1167.2011.03249.x.

90. Chen P. Carbamazepine-induced toxic effects and HLA-B*1502 screening in Taiwan / P. Chen, J.J. Lin, C.S. Lu [et al.] // New England Journal of Medicine. — 2011. — Vol. 364. — P. 1126–1133. — DOI: 10.1056/NEJMoa1009717.

91. Origa R. β-Thalassemia / R. Origa // Genetics in Medicine. — 2017. — Vol. 19. — P. 609–619. — DOI: 10.1038/gim.2016.173.

92. Piel F.B. The α-thalassemias / F.B. Piel, D.J. Weatherall // New England Journal of Medicine. — 2014. — Vol. 371. — P. 1908–1916. — DOI: 10.1056/NEJMra1404415.

93. Howes R.E. Global epidemiology of *Plasmodium vivax* / R.E. Howes, K.E. Battle, K.N. Mendis [et al.] // American Journal of Tropical Medicine and Hygiene. — 2016. — Vol. 95. — P. 15–34. — DOI: 10.4269/ajtmh.16-0141.

94. Eggermann T. Growth retardation versus overgrowth: Silver–Russell syndrome is genetically opposite to Beckwith–Wiedemann syndrome / T. Eggermann, J.H. Davies, M. Tauber [et al.] // Trends in Genetics. — 2019. — Vol. 35. — P. 329–344. — DOI: 10.1016/j.tig.2019.02.006.

95. Tucci V. Genomic imprinting and physiological processes in mammals / V. Tucci, A.R. Isles, G. Kelsey [et al.] // Cell. — 2019. — Vol. 176. — P. 952–965. — DOI: 10.1016/j.cell.2019.01.043.

96. Sakamoto Y. A new era of long-read sequencing for cancer genomics / Y. Sakamoto, S. Sereewattanawoot, A. Suzuki // Journal of Human Genetics. — 2020. — Vol. 65. — P. 3–10. — DOI: 10.1038/s10038-019-0658-5.

97. Sakamoto Y. Phasing analysis of lung cancer genomes using a long-read sequencer / Y. Sakamoto, S. Miyake, M. Oka [et al.] // Nature Communications. — 2022. — Vol. 13. — P. 3464. — DOI: 10.1038/s41467-022-31133-6.

98. Cortés-Ciriano I. Comprehensive analysis of chromothripsis in 2,658 human cancers using whole-genome sequencing / I. Cortés-Ciriano, J.-K. Lee, R. Xi [et al.] // Nature Genetics. — 2020. — Vol. 52. — P. 331–341. — DOI: 10.1038/s41588-019-0576-7.

99. Akbari V. Megabase-scale methylation phasing using nanopore long reads and NanoMethPhase / V. Akbari, V.C.T. Hanlon, K. O'Neill [et al.] // Genome Biology. — 2021. — Vol. 22. — P. 68. — DOI: 10.1186/s13059-021-02283-5.

100. Tse O.Y.O. Genome-wide detection of cytosine methylation by single-molecule real-time sequencing / O.Y.O. Tse, P. Jiang, S.H. Cheng [et al.] // Proceedings of the National Academy of Sciences of the United States of America. — 2021. — Vol. 118. — P. e2019768118. — DOI: 10.1073/pnas.2019768118.

101. Miller D.E. Targeted long-read sequencing identifies missing disease-causing variation / D.E. Miller, A. Sulovari, T. Wang [et al.] // American Journal of Human Genetics. — 2021. — Vol. 108. — P. 1436–1449. — DOI: 10.1016/j.ajhg.2021.06.006.

102. Mantere T. Optical genome mapping enables constitutional chromosomal aberration detection / T. Mantere [et al.] // American Journal of Human Genetics. — 2021. — Vol. 108. — P. 1409–1422. — DOI: 10.1016/j.ajhg.2021.05.012.

103. Cheng J. Accurate proteome-wide missense variant effect prediction with AlphaMissense / J. Cheng, G. Novati, J. Pan [et al.] // Science. — 2023. — Vol. 381. — P. eadg7492. — DOI: 10.1126/science.adg7492.

104. Wang T. The Human Pangenome Project: a global resource to map genomic diversity / T. Wang, L. Antonacci-Fulton, K. Howe [et al.] // Nature. — 2022. — Vol. 604. — P. 437–446. — DOI: 10.1038/s41586-022-04601-8.

105. Olson N.D. Variant calling and benchmarking in an era of complete human genome sequences / N.D. Olson, J. Wagner, N. Dwarshuis [et al.] // Nature Reviews Genetics. — 2023. — Vol. 24. — P. 464–483. — DOI: 10.1038/s41576-023-00590-0.

106. Mitsuhashi S. Long-read sequencing for rare human genetic diseases / S. Mitsuhashi, N. Matsumoto // Journal of Human Genetics. — 2020. — Vol. 65. — P. 11–19. — DOI: 10.1038/s10038-019-0671-8.

107. McDermott J.H. Clinical Pharmacogenetics Implementation Consortium guideline for the use of aminoglycosides based on MT-RNR1 genotype / J.H. McDermott, J. Wolf, K. Hoshitsuki [et al.] // Clinical Pharmacology & Therapeutics. — 2022. — Vol. 111. — P. 366–372. — DOI: 10.1002/cpt.2309.

108. Hung S.-I. HLA-B*1502 allele and severe carbamazepine-induced skin reactions in Han Chinese / S.-I. Hung [et al.] // Nature. — 2004. — Vol. 428. — P. 486. — DOI: 10.1038/428486a.

109. Lim K.S. Association of HLA-B*1502 allele and carbamazepine-induced severe adverse cutaneous drug reaction among Asians, a review / K.S. Lim, P. Kwan, C.T. Tan // Neurology Asia. — 2008. — Vol. 13. — P. 15–21.

110. Bertranpetit J. Mapping human genetic diversity in Asia / J. Bertranpetit [et al.] // Science. — 2009. — Vol. 326. — P. 1541–1545. — DOI: 10.1126/science.1177074.

111. Erwin G.S. Recurrent repeat expansions in human cancer genomes / G.S. Erwin, G. Gürsoy, R. Al-Abri [et al.] // Nature. — 2023. — Vol. 613. — P. 96–102. — DOI: 10.1038/s41586-022-05515-1.

112. Howes R.E. Global epidemiology of Plasmodium vivax / R.E. Howes, K.E. Battle, K.N. Mendis [et al.] // American Journal of Tropical Medicine and Hygiene. — 2016. — Vol. 95. — P. 15–34. — DOI: 10.4269/ajtmh.16-0141.

113. Tewhey R. The importance of phase information for human genomics / R. Tewhey, V. Bansal, A. Torkamani [et al.] // Nature Reviews Genetics. — 2011. — Vol. 12. — P. 215–223. — DOI: 10.1038/nrg2950.

114. Sturm D. New brain tumor entities emerge from molecular classification of CNS-PNETs / D. Sturm, B.A. Orr, U.H. Toprak [et al.] // Cell. — 2016. — Vol. 164. — P. 1060–1072. — DOI: 10.1016/j.cell.2016.01.015.

115. Mitsuhashi S. Long-read sequencing for the diagnosis of human genetic disorders / S. Mitsuhashi, N. Matsumoto, K. Hamanaka [et al.] // Journal of Human Genetics. — 2021. — Vol. 66. — P. 35–47. — DOI: 10.1038/s10038-020-00845-2.

116. Mahmoud M. Closing the gap: solving complex medically relevant genes at scale / M. Mahmoud, S. Behera, L.F. Paulin [et al.] // Annual Review of Genomics and Human Genetics. — 2024. — Vol. 25. — P. 181–204. — DOI: 10.1146/annurev-genom-021623-081639.

117. Liu B. Structural variation discovery in tumor cell lines using long-read sequencing / B. Liu, J. Conroy, C.D. Morrison [et al.] // BMC Genomics. — 2022. — Vol. 23. — P. 226. — DOI: 10.1186/s12864-022-08387-x.

118. Logsdon G.A. The structure, function and evolution of a complete human chromosome 8 / G.A. Logsdon, M.R. Vollger, P. Hsieh [et al.] // Nature. — 2021. — Vol. 593. — P. 101–107. — DOI: 10.1038/s41586-021-03420-7.

119. Wong C.K. Variant interpretation in clinical long-read sequencing: emerging consensus and outstanding challenges / C.K. Wong, J. Buchan, P.N. Robinson [et al.] // Human Genetics. — 2024. — Vol. 143. — P. 1–18.

120. Eichler E.E. Genetic variation, comparative genomics, and the diagnosis of disease / E.E. Eichler // New England Journal of Medicine. — 2019. — Vol. 381. — P. 64–74. — DOI: 10.1056/NEJMra1809315.

121. Vollger M.R. Long-read sequence and assembly of segmental duplications / M.R. Vollger, P.C. Dishuck, M. Sorensen [et al.] // Nature Methods. — 2019. — Vol. 16. — P. 88–94. — DOI: 10.1038/s41592-018-0236-3.

122. Mantere T. Long-read sequencing in clinical genetics: implementation considerations / T. Mantere [et al.] // European Journal of Human Genetics. — 2023. — Vol. 31. — P. 255–263.

123. Hung S.I. Common risk allele in aromatic antiepileptic-drug induced Stevens–Johnson syndrome and toxic epidermal necrolysis in Han Chinese / S.I. Hung [et al.] // Pharmacogenomics. — 2010. — Vol. 11. — P. 349–356. — DOI: 10.2217/pgs.09.162.

124. Robinson P.N. Improved exome prioritization of disease genes through cross-species phenotype comparison / P.N. Robinson, S. Köhler, A. Oellrich [et al.] // Genome Research. — 2014. — Vol. 24. — P. 340–348. — DOI: 10.1101/gr.160325.113.

125. Vollger M.R. Synchronized long-read genome, methylome, epigenome and transcriptome profiling resolve a Mendelian condition / M.R. Vollger, J. Korlach, K.C. Eldred [et al.] // Nature Genetics. — 2025. — Vol. 57. — P. 469–479. — DOI: 10.1038/s41588-024-02067-0.