








# Long and Accurate: How HiFi Sequencing is Transforming Genomics

Bo Wang (王博) <sup>1,2,#</sup>, Peng Jia (贾鹏) <sup>3,#</sup>, Shenghan Gao (高胜寒) <sup>1,2</sup>,  
Huanhuan Zhao (赵焕焕) <sup>3</sup>, Gaoyang Zheng (郑高洋) <sup>3</sup>, Linfeng Xu (许林峰) <sup>1,2</sup>,  
Kai Ye (叶凯) <sup>1,2,3,4,\*</sup>

<sup>1</sup>School of Automation Science and Engineering, Faculty of Electronic and Information Engineering, Xi'an Jiaotong University, Xi'an 710049, China

<sup>2</sup>MOE Key Lab for Intelligent Networks & Networks Security, Faculty of Electronic and Information Engineering, Xi'an Jiaotong University, Xi'an 710049, China

<sup>3</sup>Department of Gynecology and Obstetrics, Center for Mathematical Medical, The First Affiliated Hospital of Xi'an Jiaotong University, Xi'an 710061, China

<sup>4</sup>Faculty of Science, Leiden University, Leiden 2311 EZ, The Netherlands

\*Correspondence author: kaiye@xjtu.edu.cn (Ye K).

#Equal contribution.

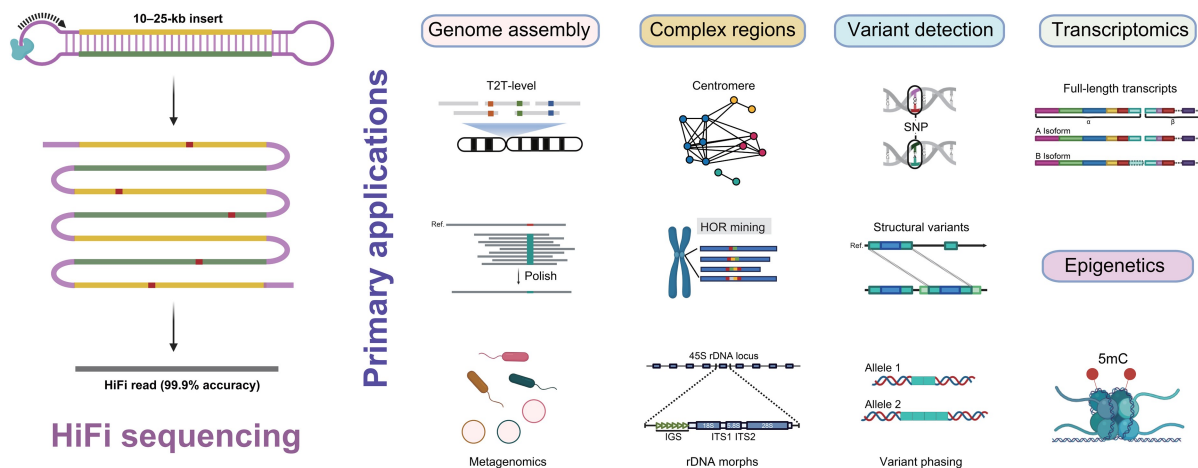
Handling Editor: Wanlu Liu

## Abstract

Recent developments in PacBio high-fidelity (HiFi) sequencing technologies have transformed genomic research, with circular consensus sequencing now achieving 99.9% accuracy for long (up to 25 kb) single-molecule reads. This method circumvents biases intrinsic to amplification-based approaches, enabling thorough analysis of complex genomic regions [including tandem repeats, segmental duplications, ribosomal DNA (rDNA) arrays, and centromeres] as well as direct detection of base modifications, furnishing both sequence and epigenetic data concurrently. This has streamlined a number of tasks including genome assembly, variant detection, and full-length transcript analysis. This review provides a comprehensive overview of the applications and challenges of HiFi sequencing across various fields, including genomics, transcriptomics, and epigenetics. By delineating the evolving landscape of HiFi sequencing in multi-omics research, we highlight its potential to deepen our understanding of genetic mechanisms and to advance precision medicine.

**Key words:** Long-read sequencing; Genome assembly; Complex genomic region; Variant detection; Centromere.

## Graphical abstract



## Introduction

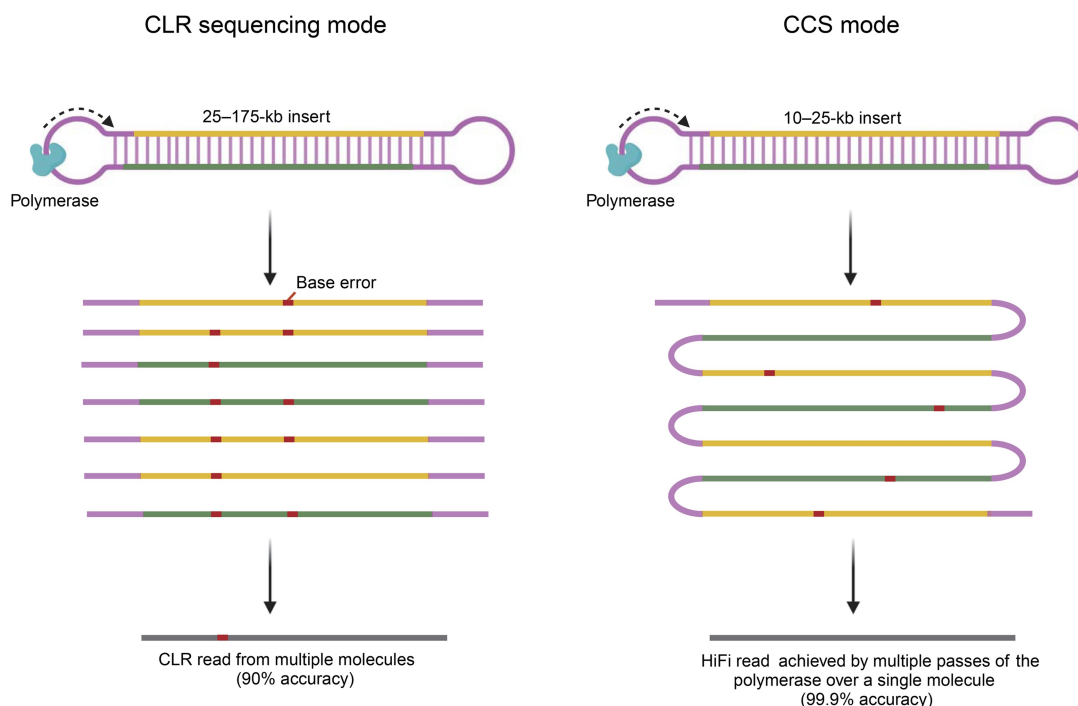
The complete genetic makeup of an organism, called its genome, holds crucial information for unraveling the origins and mechanisms of disease [1]. By enabling the identification of genomic regions associated with diseases, DNA sequencing technologies have provided invaluable insights into diagnosis,

prognosis, and treatment strategies [2]. In this respect, next-generation sequencing (NGS) platforms are widely used because they produce short reads with relatively high accuracy, making them well-suited for detecting both single-nucleotide variants (SNVs) and small insertions and deletions (indels). However, NGS platforms struggle with tasks requiring longer

Received: 20 May 2024; Revised: 16 December 2024; Accepted: 6 January 2025.

© The Author(s) 2025. Published by Oxford University Press and Science Press on behalf of the Beijing Institute of Genomics, Chinese Academy of Sciences / China National Center for Bioinformation and Genetics Society of China.

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted reuse, distribution, and reproduction in any medium, provided the original work is properly cited.



**Figure 1 PacBio HiFi sequencing modes**

The PacBio platform produces long reads with two types of sequencing modes: CLR and CCS. CLR, derived from 25–175-kb DNA fragment inserts, often exhibit an error rate of 8%–15% due to single-pass sequencing. By contrast, HiFi reads, generated from 10–25-kb inserts using CCS mode, have an error rate of  $\leq 1\%$  by leveraging multiple passes around the template. CLR, continuous long read; CCS, circular consensus sequencing; HiFi, high-fidelity.

stretches of DNA [3], such as *de novo* genome assembly, complex variant detection, and haplotype phasing. Long-read sequencing technologies, including PacBio continuous long read (CLR) sequencing [3] and Oxford Nanopore Technologies (ONT) sequencing [4], are better-suited for tackling these challenges as they routinely produce long reads exceeding 10 kb. These platforms work by directly sequencing single molecules, although compared to short-read NGS platforms they generally have reduced read accuracy (75% to 90%) [5]. Nevertheless, obtaining a completely accurate picture of the genome remains challenging, particularly for its most complex regions, such as centromeric sequences and tandem repeats [6,7].

In 2019, PacBio introduced high-fidelity (HiFi) sequencing technology, a significant advancement that addresses some of the limitations of previous technologies. HiFi implements circular consensus sequencing (CCS), which expands upon CLR sequencing by combining information from multiple readings of the same DNA molecule (discussed further below) [5]. CCS produces long, highly accurate reads from initially noisy individual subreads (Figure 1), and is currently implemented by the PacBio Sequel, Sequel II, and Sequel IIe systems. In 2023, PacBio released a new apparatus, Revio, which achieves a 15-fold increase in throughput which, coupled with a 4-fold decrease in cost, potentially enables the complete sequencing of approximately 1300 human genomes per year [8]. These advances have collectively been transformative for genomic research.

In this review, we first provide an overview of the mechanisms of HiFi sequencing, and then discuss its applications across several different fields of research (Figure 2), the challenges faced, and the future prospects for this technology.

## HiFi sequencing mechanism and performance

The accuracy of HiFi sequencing lies in its iterative approach to crafting consensus reads, achieved through the repetitive sequencing of identical DNA fragments (Figure 1). This iterative cycle entails multiple traversals around the template, facilitated by enzymes exceeding the length of the DNA insert. Consequently, each DNA fragment undergoes multiple sequencing cycles, resulting in the accrual of data conducive to building a consensus sequence. Assembly algorithms ingest the data from each sequencing iteration, correct errors, and then construct a consensus mirroring the authentic DNA sequence [5].

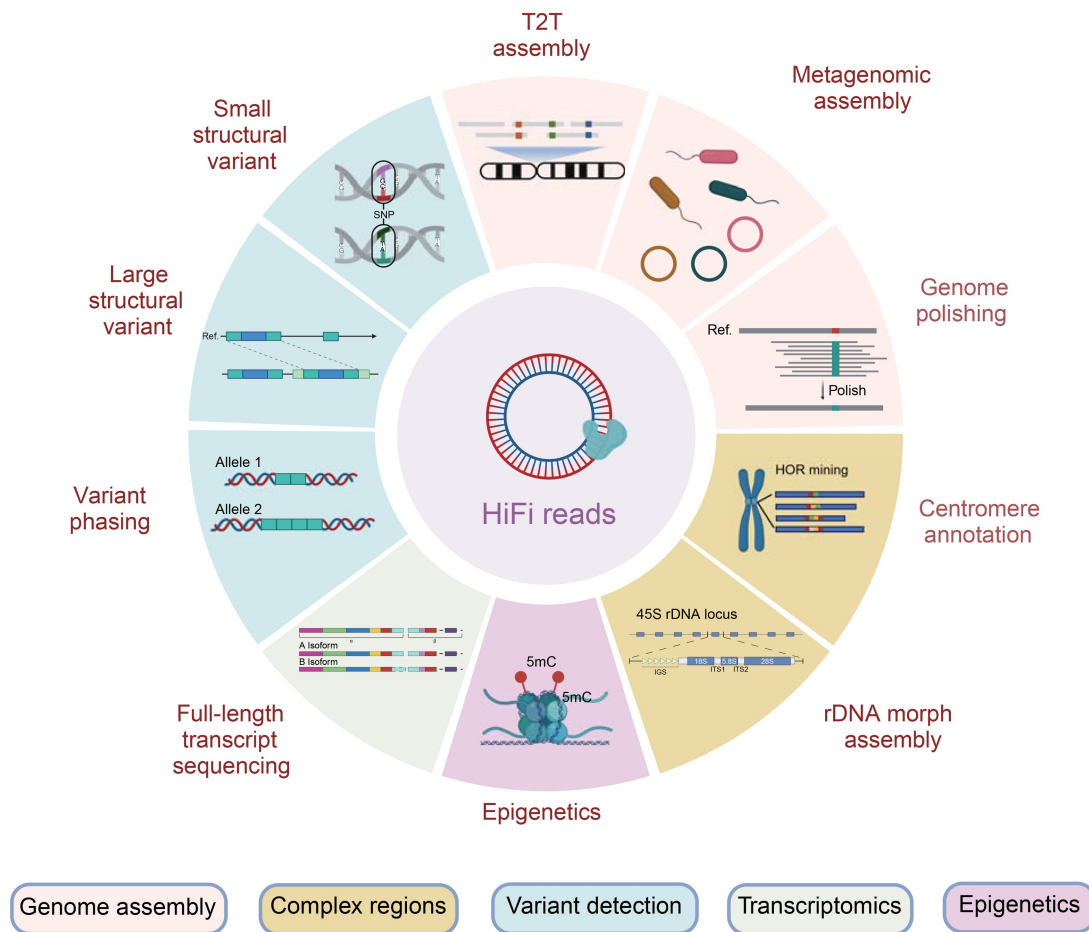
A standout characteristic of HiFi sequencing is its capacity to produce elongated read lengths (typically 10 to 25 kb). It is worth noting that shorter DNA fragments could also generate CCS reads, but this would lead to reduced data production. Users should primarily focus on the quality control of read length. Furthermore, it achieves remarkable precision, with a median accuracy of 99.9% and accurate resolution of over 99.5% of homopolymers that are 5 or more bases in length [9].

In striking a balance between read length and accuracy, HiFi reads may effectively cover shorter repetitive sequences while simultaneously aiding in the identification of longer, more complex repeats (Figure 2). This capability is crucial for tasks like assembling complete genomes and detecting complex genetic variants associated with diseases (Figure 2).

## Genome assembly

### Single-species telomere-to-telomere assembly

HiFi sequencing's ability to generate long, highly accurate reads has revolutionized the field of genome assembly. A major



**Figure 2 Primary applications of HiFi sequencing for genomic research**

The figure was created with [BioRender.com](https://www.biorender.com). T2T, telomere-to-telomere; rDNA, ribosomal DNA; Ref., reference; SNP, single-nucleotide polymorphism; HOR, higher-order repeat; IGS, intergenic spacer; ITS, internal transcribed spacer.

challenge in assembling complete genomes is resolving lengthy repetitive regions, such as centromeres. When combined with ONT ultra-long (> 100 kb) reads for hybrid assembly (discussed further below), HiFi can make a significant contribution toward producing completely gapless [telomere-to-telomere (T2T)] assemblies [10]. In 2020, HiFi sequencing played a key role in the first T2T assembly of a human X chromosome, primarily focusing on polishing, validating, and selecting unique anchors for the complex centromere region [11]. Two years later, the international T2T Consortium utilized both HiFi and ONT ultra-long reads to differentiate subtly diverged repeat copies, and ultimately assembled the 3.055 billion base pair sequence of a human genome [12], known as T2T-CHM13. This comprehensive sequence encompasses all centromeric satellite arrays, recent segmental duplications, and the short arms of all five acrocentric chromosomes. However, it is worth noting that CHM13 originates from a hydatidiform mole cell line with a nearly homozygous genome. This highlights the need for further advancements to assemble diploid genomes to the T2T level. We consider this imperative for clinically relevant samples and for enhancing our understanding of human genetic variation. To that end, Jarvis et al. evaluated methods for diploid human genome assembly and proposed an optimal combination of sequencing and assembly algorithms for accuracy and completeness of results with minimal manual curation [13]. They found that integrating both HiFi and ONT

ultra-long data into a diploid assembly graph, along with long-range phasing information from trios, high-throughput chromosome conformation capture (Hi-C), or Strand-seq, could soon enable fully automated T2T diploid genome assemblies [14]. In 2023, two research teams from China accomplished two fully-phased T2T diploid genomes of a male Han Chinese, T2T-CN1 [15] and T2T-Yao [16], respectively.

HiFi sequencing has also been employed in the assembly of a number of both model and non-model T2T (or near T2T) plant genomes, including *Arabidopsis thaliana* [17–19], the moss *Physcomitrium patens* [20], the green alga *Chlamydomonas reinhardtii* [21], rice [22], maize [23], soybean [24], kiwifruit [25], Chinese cork oak [26], sorghum [27], four species of poppy (genus *Papaver*) [28], and two species of algae (genus *Chlorella*) [29]. Recently, in the animal kingdom, chickens [30], Chinese sea bass [31], geese [32], and finless porpoises [33] have also relied on HiFi sequencing to accomplish T2T-level genome assemblies.

To the best of our knowledge, there are at present 11 *de novo* assembly tools designed for HiFi reads, namely HiCanu [34], hifiasm [35], Flye [36], hifiasm-meta [37], metaFlye [38], PEREGRINE [39], Shasta with HiFi-mode [40], Verkko [14], MECAT2 [41], miniasm [42], and NextDenovo [43]. Yu et al. conducted an evaluation of the aforementioned 11 assemblers by using quality assessment tools and benchmarking universal single-copy orthologs, alongside several

additional criteria [44]. The results indicate that hifiasm and hifiasm-meta emerge as the top choices for assembling both eukaryotic genomes and metagenomes with HiFi data. Verkko represents one of the initial attempts to automate T2T assembly utilizing HiFi data and ONT ultra-long reads (> 100 kb) [14]. However, Verkko demands significant computational resources and it lacks the capability to resolve haplotypes in polyploid samples. In contrast, Cheng et al. introduced hifiasm with ultra-long (UL) mode as an efficient solution for near T2T assembly of both diploid and polyploid samples [45]. hifiasm (UL) seamlessly integrates HiFi reads, ONT ultra-long reads, Hi-C reads, and trio data to generate high-quality assemblies in a single step. Notably, hifiasm (UL) combines two graphs — the first built from HiFi reads and the second from UL reads — resulting in a refined final assembly graph. Recently, Verkko v2.0 [14] or higher version was released, and it is significantly faster than the previous version. From the previous experience [15], for users with high-coverage HiFi reads, we suggest integrating Verkko with hifiasm for T2T assembly. This development marks a pivotal transition in T2T genomic research toward the era of pan-T2T genomics [45,46].

### Genome polishing

Although HiFi sequencing has facilitated T2T assembly, genome assemblies often still contain base-level errors, especially in homopolymer or low-complexity microsatellite regions, which are particularly susceptible to inaccuracies in HiFi reads [47]. Current genome polishing tools like Pilon [48], Racon [49], and NextPolish [50] often result in overcorrections and haplotype switch errors when rectifying errors in genomes assembled from HiFi reads [47]. To tackle this challenge, Hu et al. recently introduced NextPolish2 [47], a repeat-aware genome polishing tool proficient in correcting residual base errors in genomes assembled from HiFi reads. Compared to the modern polishing pipeline, Racon + Merfin [51], which was used to polish the CHM13 human T2T genome assembly, NextPolish2 minimizes excessive overcorrections, even in regions with highly repetitive elements [47]. The introduction of NextPolish2 holds significant promise for further enhancing the accuracy of T2T genomes.

### Metagenomic assembly

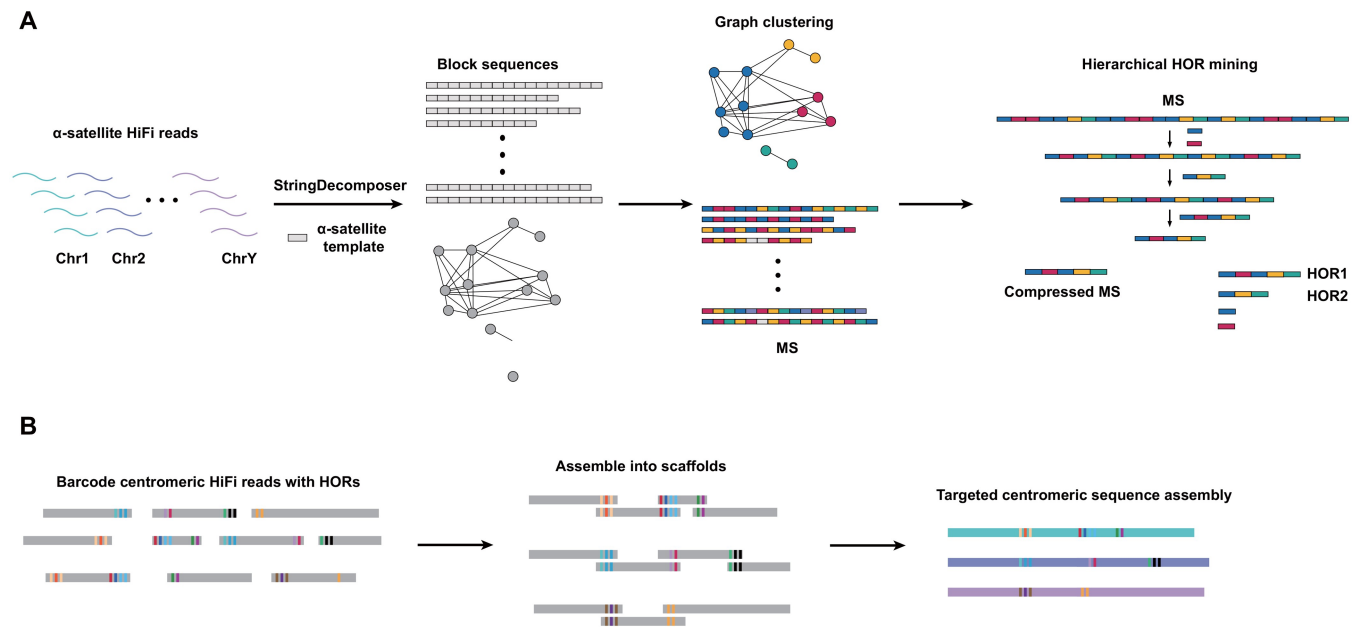
Metagenomic *de novo* assembly is a prevalent method for investigating microbial communities [52,53]. Prior to 2020, binning algorithms were frequently utilized to cluster short contigs for short-read metagenome assembly [54]. However, this method often led to highly fragmented and significant errors [55], thereby complicating or misleading downstream analyses. The limitations of short-read metagenome-assembled genomes (MAGs) prompted the development of (among other algorithms) hifiasm-meta, harnessing the potential of HiFi reads for this task [37]. Through HiFi metagenomic assembly, they successfully obtained 102 complete MAGs from five human fecal samples, potentially representing complete genomes of human gut prokaryotic species. Furthermore, Zhang et al. enhanced microbial genomes and gene catalogs of the chicken gut through metagenomic sequencing of HiFi reads, recovering a substantial portion of novel genomes and genes previously overlooked in short-read-based metagenome studies [56]. Utilizing GTDB-Tk, all 337 species-level genomes were accurately classified at the order level; however, over half ( $n = 189$ ) of these genomes failed to achieve

species-level classification. In another study, Bickhart et al. reported lineage-resolved high-quality MAGs at the strain level in a complex metagenome [57]. By employing Hi-C binning and MagPhase phasing techniques, they were able to produce strain-level MAGs. These lineage-resolved complete MAGs represent a significant advancement toward achieving isolate-quality genome assemblies for complex microbial organisms. More recently, Benoit et al. have introduced a novel metagenomics assembler named metaMDBG designed specifically for HiFi reads [55]. The authors demonstrated that for intricate microbial communities, metaMDBG outperformed existing methods by yielding up to twice the number of high-quality circularized prokaryotic genomes. Additionally, metaMDBG exhibited superior capabilities in the retrieval of viruses and plasmids compared to other approaches [55].

## Resolving complex genomic regions

### Centromere

Centromeres are crucial genomic regions and play a vital role in cell division. Understanding their structure is important for investigating how chromosomes are accurately copied and separated during cell division. Traditionally, analyzing centromeres has been challenging due to their repetitive DNA sequences, although HiFi reads have been instrumental in addressing the long-standing questions about their architecture and evolution [6]. Deep analysis of centromere architecture is critical for understanding genome stability, cell division, and disease development [58]. Centromere annotation refers to the process of partitioning centromeres into monomers and higher-order repeats (HORs). Annotation of HORs from HiFi reads is one direct way to obtain and validate centromere structures, for instance using Alpha-CENTAURI [59]. Due to the limitations of read-based annotation without genomic positional information for different centromeric HORs, researchers developed the first fully automated centromere annotation tool, HORmon, based on de Bruijn graphs and the T2T-CHM13 assembly [60]. In 2023, Gao et al. proposed a generalizable automatic centromere annotation tool named HiCAT [61], based on hierarchical tandem repeat mining. HiCAT employs a bottom-up iterative tandem repeat compression strategy to detect and represent locally-nested HORs. This approach significantly enhances annotation continuity and allows for the detection of fine structures and organization of HOR units, including length variations of HORs that were overlooked by HORmon. Gao et al. introduced an enhanced version of HiCAT, termed HiCAT-human [62], designed to automatically annotate centromere HOR patterns from HiFi reads (Figure 3A). HiCAT-human employs a combined approach integrating the “string decomposer”, “graph clustering”, and “hierarchical HOR mining” computational methods to annotate centromere sequences (Figure 3A). This analysis of HOR diversity can be utilized for tagging HiFi reads, aiding in further centromere targeted sequencing assembly (Figure 3B). Both HORmon and HiCAT series annotate centromeres based on known monomer sequences. In contrast, Wlodzimierz et al. introduced tandem repeat annotation and structural hierarchy (TRASH) [63], a tool capable of identifying and mapping tandem repeats in genome sequences without prior knowledge of repeat composition. TRASH proves particularly valuable for annotating putative centromeres in



**Figure 3 Applications of HiFi reads in centromere annotation and assembly**

**A.** Depicting a computational pipeline designed for annotating centromeric HOR patterns from HiFi reads. **B.** Barcoding the HiFi reads with centromeric HORs, followed by scaffolding and assembly of the centromeric sequences. MS, monomer sequence; Chr, chromosome.

non-model species, where centromeres are comprised of unknown tandem repeat monomers.

### Ribosomal DNA morph assembly

Ribosomal RNA (rRNA) is a crucial component of ribosomes, with the genes encoding rRNA organized in repetitive clusters called ribosomal DNA (rDNA) arrays. For example, in humans, these arrays are located on chromosomes 13, 14, 15, 21, and 22, comprising a few hundred copies of an approximately 45-kb repeat unit arranged in tandem repeats [12]. A morph represents the sequence of one complete repeat unit appearing once or more times within the rDNA arrays. Detailed analysis of morph types, including chromosome-specific morphs, aids in assembling rDNA across different chromosomes. Rautiainen et al. developed a tool called Ribotin [64], which utilizes a combination of HiFi reads and ONT ultra-long reads to resolve variations among rDNA morphs. Ribotin effectively identifies the most abundant morphs in both human and nonhuman genomes. Importantly, for species with short rDNA morph sizes (up to 10 kb per morph), accurate morphs can be obtained using only HiFi reads. When integrated with the assembly tool Verkko, Ribotin enables researchers to accurately assemble rDNA morphs for each chromosome, providing a more complete picture of these important genes [64].

### Variant detection

HiFi sequencing technology has been widely applied to both healthy and diseased human samples, catalyzing the development of variant detection methods (Table 1). HiFi reads not only demonstrate high proficiency in detecting small variants but also promise performance in identifying larger structural variants (SVs) and tandem repeats, which are challenging for short reads and noisier long reads, respectively. Furthermore, using HiFi sequencing technology, heterozygous variants in diploid samples can now routinely be phased.

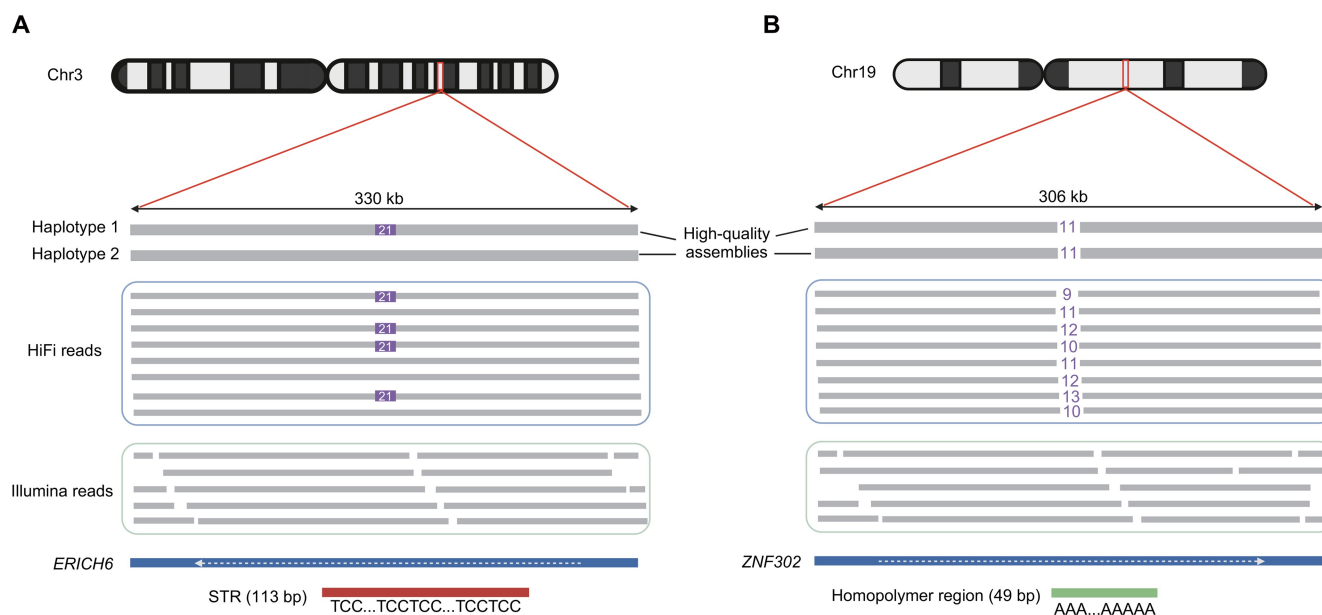
### SNVs and indels

SNVs and indels are the most abundant genetic variants in humans. In the last two decades, many small variants have been detected using short-read sequencing technologies, and genotypes associated with both pathogenic and other phenotypes have been identified through the efforts of many international consortia, including the 1000 Genomes Project [83], The Genome Aggregation Database (GnomAD) [84], and the Cancer Genome Atlas (TCGA) program [85]. However, these variants are mainly located at simple non-repetitive regions of the human genome, which account for approximately 85%–90% of the total sequence [83]. Prior to the release of T2T genomes, many SNVs and indels in complex regions had not been accurately resolved due to the inadequate length of sequencing reads [12]. Recently, HiFi reads achieved an F1 score of 0.998 for variant calling across all benchmark regions in the PrecisionFDA Truth Challenge, showing performance equivalent to short reads, which had an F1 score of 0.997 [86]. However, the longer length of HiFi reads enables more accurate alignments in genomic repetitive regions, where short reads may produce multiple aligned results [87]. For example, Jia et al. detected a 21-bp heterozygous insertion of a short tandem repeat (STR) region in *ERICH6* which was accurately identified by both haplotype-resolved assemblies and HiFi reads, but missed by Illumina short-read data [88] (Figure 4A). The authors highlighted an additional example involving a homozygous deletion within a homologous region (a 49-bp repeat of adenine) in *ZNF302*. This deletion was also overlooked by Illumina sequencing but identified as a homozygous deletion by HiFi sequencing and high-quality assemblies [88] (Figure 4B). Based on these results, haplotype-resolved assemblies identified the deletion as 11 bp in both haplotypes, while the deletion lengths detected by HiFi reads ranged from 9 to 13 bp. This suggests that haplotype-resolved assemblies might be more suitable for detecting small variants in homopolymer regions (Figure 4B). To that end, Vollger et al. systematically analyzed the small

**Table 1** List of variant detection methods for HiFi reads

Category	Method	Year	Link	Ref.
Small variant detection	DeepVariant	2018	<a href="https://github.com/google/deepvariant">https://github.com/google/deepvariant</a>	[65]
	PEPPER	2021	<a href="https://github.com/kishwarshafin/pepper">https://github.com/kishwarshafin/pepper</a>	[40]
	Clair3	2022	<a href="https://github.com/HKU-BAL/Clair3">https://github.com/HKU-BAL/Clair3</a>	[66]
SV detection	pbsv	2018	<a href="https://github.com/PacificBiosciences/pbsv">https://github.com/PacificBiosciences/pbsv</a>	Not available
	Sniffles	2018	<a href="https://github.com/fritzsedlazeck/Sniffles">https://github.com/fritzsedlazeck/Sniffles</a>	[67]
	Sniffles2	2024	<a href="https://github.com/fritzsedlazeck/Sniffles">https://github.com/fritzsedlazeck/Sniffles</a>	[68]
	SVIM	2019	<a href="https://github.com/eldariont/svim">https://github.com/eldariont/svim</a>	[69]
	cuteSV	2020	<a href="https://github.com/tjiangHIT/cuteSV">https://github.com/tjiangHIT/cuteSV</a>	[70]
	SVision	2022	<a href="https://github.com/xjtu-omics/SVision">https://github.com/xjtu-omics/SVision</a>	[71]
	SVision-pro	2024	<a href="https://github.com/songbowang125/SVision-pro">https://github.com/songbowang125/SVision-pro</a>	[72]
	DeBreak	2023	<a href="https://github.com/Maggi-Chen/DeBreak">https://github.com/Maggi-Chen/DeBreak</a>	[73]
	nanomonsv	2023	<a href="https://github.com/friend1ws/nanomonsv">https://github.com/friend1ws/nanomonsv</a>	[74]
	Tandem repeat genotyping	adVNTR	2018	<a href="https://github.com/mehrdadbakhtiari/adVNTR">https://github.com/mehrdadbakhtiari/adVNTR</a>
tandem-genotypes		2019	<a href="https://github.com/mcfrith/tandem-genotypes">https://github.com/mcfrith/tandem-genotypes</a>	[76]
Straglr		2021	<a href="https://github.com/bcgsc/straglr">https://github.com/bcgsc/straglr</a>	[77]
cTR		2023	<a href="https://github.com/morisUtokyo/cTR">https://github.com/morisUtokyo/cTR</a>	[78]
TRGT		2024	<a href="https://github.com/PacificBiosciences/trgt">https://github.com/PacificBiosciences/trgt</a>	[79]
Assembly calling		LongTR	2024	<a href="https://github.com/gymrek-lab/LongTR/">https://github.com/gymrek-lab/LongTR/</a>
	SVIM-asm	2020	<a href="https://github.com/eldariont/svim-asm">https://github.com/eldariont/svim-asm</a>	[81]
	PAV	2021	<a href="https://github.com/EichlerLab/pav">https://github.com/EichlerLab/pav</a>	[82]

Note: HiFi, high-fidelity; SV, structural variant.

**Figure 4** Variant detection of HiFi reads, short reads, and high-quality assembly in repeat regions

**A.** Heterozygous insertion at a STR region with a TCC repeat. **B.** Homozygous deletion at a homopolymer region. STR, short tandem repeat.

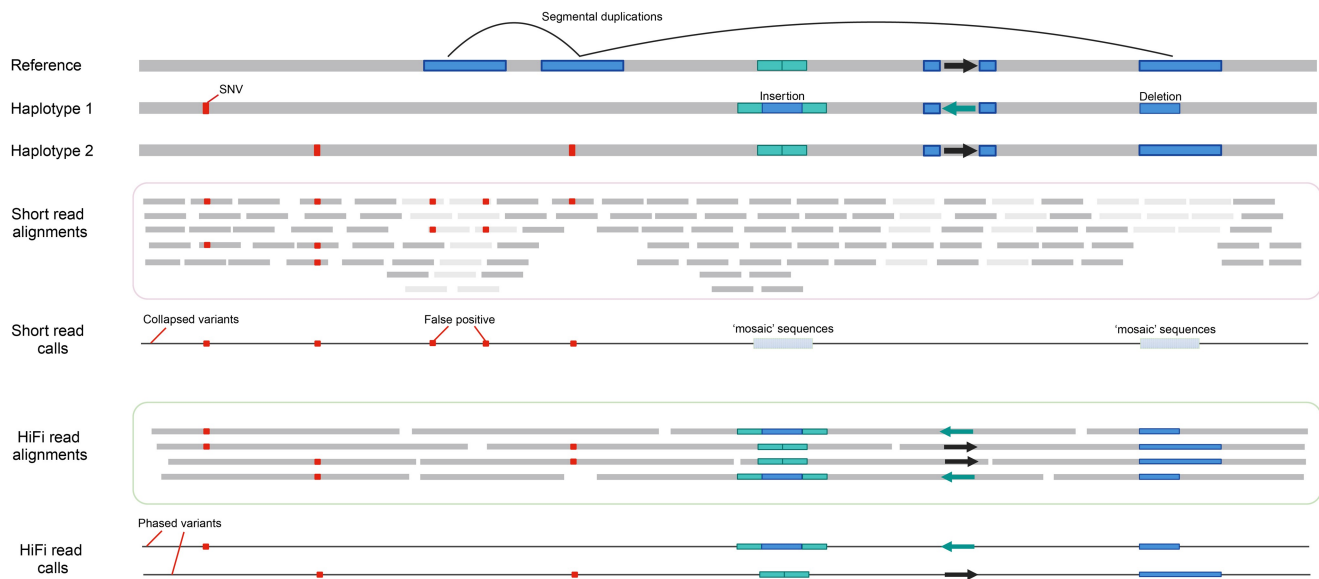
variants in segmental duplications (SDs) with HiFi assemblies, identifying a large number of SNVs in SDs that were previously considered largely inaccessible [89]. Furthermore, Chen et al. developed Paraphase to accurately identify variants of 160 long SDs in medically relevant genes which, also, were previously inaccessible [90].

Compared to germline variant calling, however, the detection of somatic variants in cancer samples requires much higher sequencing accuracy and depth due to their characteristically lower mutation frequency; this represents an ongoing challenge for HiFi sequencing.

### Large structural variants

SVs are typically described as gains, losses, and/or duplications of larger ( $\geq 50$  bp) stretches of genome sequence. Although

there are fewer SVs in a healthy human sample compared to small variants, the total length of the genome containing SVs is typically far greater [82]. Despite their large impact, SVs have been relatively understudied due to their complicated patterns compared to SNVs and indels. Nevertheless, long-read sequencing technology and algorithmic developments (Table 1) have facilitated the increased resolution of SVs in various samples. HiFi reads are able to cover most SVs and their high sequencing accuracy enables high-resolution identification of variant breakpoints, a capability that short reads and noisier long reads cannot achieve. Recent studies have demonstrated that HiFi reads can be used with deep learning models to analyze complex SVs involving multiple events [67,71]. Additionally, HiFi reads can detect mosaic and somatic variants in both mixed samples and cancer [68,72].



**Figure 5 Advantages of HiFi reads in variant phasing**

In comparison to short reads, HiFi reads excel in the read-backed variant phasing category. Haplotype1 and haplotype2 are two haplotypes. SNV, single-nucleotide variant.

### Tandem repeat detection

Tandem repeats, including STRs and variable number tandem repeats (VNTRs), contribute a substantial number of genetic variants in the human genome [91,92]. Expansions and contractions of these repeats are associated with a number of diseases including Huntington's disease and Fragile X syndrome [93]. Compared to short reads, HiFi reads can accurately resolve tandem repeat genotypes across the whole genome, particularly for VNTRs. Recently, advanced methods (Table 1), such as cTR [78], TRGT [79], and LongTR [80], have been developed for tandem repeat genotyping and visualization with HiFi sequencing reads, facilitating the application of this technology in tandem repeat-associated diseases [94,95]. Furthermore, HiFi reads are able to resolve some complex tandem repeats containing a mixture of multiple motifs, enabling deeper understanding of the relationship between the motif composition and phenotype [96,97].

### Variant phasing

Variant calling using HiFi sequencing technology achieves the highest precision and recall across all variant categories, according to the PrecisionFDA Truth Challenge V2 [86]. By analyzing HiFi reads that span heterozygous variants (Figure 5), these variants can be phased with haplotype-resolved resolution [98–101]. HiFi sequencing captures a large number of informative bases overlapping with heterozygous variants, making it highly effective for identifying both small variants and SVs. This is particularly true in SD regions, where HiFi reads offer substantial advantages over short reads (Figure 5). Moreover, HiFi reads could be used to generate high-quality haplotype-resolved assemblies for diploid samples by advanced assemblies, such as hifiasm [35,45] and Verkko [14]. Based on HiFi assemblies, all scales of genomic variant test samples can be systematically categorized including variants larger than HiFi read lengths [82].

### Epigenetics

HiFi sequencing enables simultaneous calling of the four DNA bases and 5-methylcytosine (5mC) from untreated genomic DNA, allowing for genome-wide detection and phasing of genetic and epigenetic variants using a single, standard HiFi library preparation [102]. With the capability of long-read sequencing, one can achieve epigenetic analysis without the need for bisulfite treatment [103]. Unlike methods requiring chemical conversion of DNA, HiFi sequencing detects modifications in native DNA through impacts on base incorporation kinetics, ensuring high accuracy of sequence and methylation. Methylation detection with HiFi sequencing is highly consistent with bisulfite sequencing (average Pearson correlation = 0.97; mean absolute difference = 0.06) [104]. Additionally, HiFi sequencing provides access to the entire genome, including challenging regions like repeats and centromeres beyond the reach of short-read sequencing [105]. Moreover, HiFi sequencing facilitates phasing, allowing identification of allele-specific methylation arising from parental imprinting, genetic variation, or repeat expansions.

Recently, ccsmeth, a deep learning tool, was developed to detect 5mC methylation from HiFi sequencing by utilizing kinetic features [inter-pulse durations (IPDs) and pulse widths (PWs)] of HiFi reads at both the read and genome-wide site levels [105]. Recent studies have applied HiFi sequencing to call 5mC methylation specifically in the centromeric regions of genomes. For instance, in 2024, Wang et al. utilized ccsmeth to analyze methylation in the centromeric regions of the green algae genome and found that the centromeric CENH3 signals of green algae display a pattern of hypomethylation [29], similar to humans and higher plants. The pocket of hypomethylated CpG DNA was called the centromere dip region (CDR). Mastroiosa et al. employed HiFi and ONT ultra-long sequencing to fully assemble and extract methylation tags of the centromeres from a parent–child trio where the child presented with trisomy 21 [106]. The affected individual carries three distinct chromosome 21 centromere haplotypes, differing in length by 11-fold, with both the

largest (H1) and smallest (H2) originating from the mother. The longest H1 allele exhibits a less clearly defined CDR based on CpG methylation and shows a significantly reduced signal in CENP-A chromatin immunoprecipitation sequencing (ChIP-seq) compared to H2 or paternal H3 centromeres. These epigenetic patterns suggest weaker kinetochore attachment for the maternally transmitted H1. Analysis of H1 in the mother indicates that the reduced CENP-A ChIP-seq signal, though not the CDR profile, existed prior to the meiotic nondisjunction event. These findings imply that recent differences in size and epigenetic characteristics of chromosome 21 centromeres may contribute to the risk of nondisjunction. HiFi sequencing has now become a comprehensive and accurate technology for 5mC detection and methylation phasing, particularly in repetitive genomic regions [104,105,107]. Its usage is expected to become even more widespread in epigenetics.

### Full-length transcript sequencing

HiFi sequencing holds the promise of significantly enhancing full-length transcript assembly. By generating long, accurate reads, it enables a more precise reconstruction of complex transcripts, thereby improving the identification of fusion genes and uncovering previously undetected isoforms with unprecedented accuracy [108,109]. This advancement not only aids in unraveling the intricate landscape of gene expression but also lays the groundwork for deeper insights into cellular functions and disease mechanisms.

Fusion genes, resulting from the fusion of two distinct pre-mRNAs via *trans*-splicing, play a pivotal role in cancer development and progression. Detecting fusion transcripts accurately is crucial for cancer diagnosis, prognosis, and the development of targeted therapies. Long-read transcriptome sequencing has emerged as a powerful tool for identifying fusion genes, offering the ability to capture full-length transcripts and detect complex rearrangements with high sensitivity and resolution. Among the various tools available for detecting fusion genes in long-read transcriptome sequencing data, several stand out for their effectiveness and applicability across different cancer types. LongGF [110], JAFFAL [111], FusionSeeker [112], pbfusion [113], and CTAT-LR-fusion [114] are examples of such tools, each offering unique features and capabilities. Specifically, pbfusion, designed for isoform sequencing (Iso-Seq) HiFi data, has demonstrated its utility in identifying both known and novel fusion events in sarcomas. Its ability to accurately detect driver events highlights the reliability of Iso-Seq HiFi sequencing in uncovering clinically relevant fusion transcripts. By leveraging Iso-Seq HiFi data, pbfusion enables comprehensive characterization of fusion events, providing valuable insights into the molecular mechanisms driving cancer progression. Similarly, CTAT-LR-fusion offers a versatile solution for detecting fusion transcripts from long-read transcriptomic data, applicable at both bulk and single-cell levels. Its flexibility and scalability make it a valuable tool for studying fusion events across different cancer types and experimental settings. By integrating advanced algorithms and HiFi long-reads, CTAT-LR-fusion enables precise identification and characterization of fusion transcripts, facilitating the discovery of novel biomarkers and therapeutic targets. Compared to Nanopore direct RNA sequencing (DRS), Iso-Seq HiFi demonstrates higher precision in both exonic and

intronic regions [112]. One drawback of HiFi sequencing, however, is that it typically produces fewer reads per run than ONT sequencing. Due to the limited throughput, multiplexed arrays isoform sequencing (MAS-Iso-Seq), a concatenation method designed to increase throughput by joining complementary DNA (cDNA) molecules into longer concatenated fragments, was introduced, increasing the throughput to nearly 40 million cDNA reads per run on the Sequel IIe sequencer [115]. Recently, PacBio has made significant advancements in yield, with the Revio system achieving approximately a 15-fold increase in throughput compared to the earlier Sequel IIe system. Additionally, tools like JAFFAL are proving valuable for enhancing full-length transcriptomic research [111]. Another limitation of HiFi sequencing is its primary application in detecting DNA modifications, whereas Nanopore DRS offers the advantage of direct, real-time sequencing of both DNA and RNA modifications [116].

Overall, the identification of fusion genes and novel isoforms through HiFi transcriptome sequencing data not only marks a substantial leap forward in cancer research but also holds immense promise for enhancing clinical diagnostics in the field.

### Single-cell sequencing

Understanding the intricate landscape of cancer requires comprehensive genotype–phenotype data at both the single-cell DNA and RNA levels. Single-cell NGS technologies emerged about a decade ago and have been commonly used for detecting variations through short-read sequencing protocols [117,118]. However, the limitations of short reads hinder their ability to capture the full range of genetic variation in individual cells.

To address this, Fan et al. developed single-molecule real-time sequencing of long fragments amplified through transposon insertion (SMOOTH-seq), a single-cell genome sequencing method based on HiFi long-read sequencing [119]. SMOOTH-seq reliably and effectively detects SVs and holds promise for *de novo* assembly of genomic DNA from individual cells [120]. However, SMOOTH-seq has limitations in detecting other types of genetic variations. To improve upon this, Hård et al. introduced a new method that uses an automated droplet-based multiple displacement amplification (dMDA) technique for single-cell genome assembly, combined with HiFi sequencing [121]. The inclusion of HiFi sequencing offers enhanced performance in detecting genetic variations in single cells, especially for haplotype phasing, complex SVs, and tandem repeats [121].

Single-cell transcriptome assembly may also benefit from full-length transcript sequencing. For example, HiFi long-read single-cell RNA sequencing (scRNA-seq) was conducted on clinical samples obtained from three ovarian cancer patients exhibiting omental metastasis [109]. Using HiFi sequencing depths of approximately 12,000 reads per cell, the authors captured a total of 152,000 isoforms, among which more than 30% were novel isoforms. In addition, cell type-specific isoforms and polyadenylation site utilization patterns could be identified in both tumor and mesothelial cells. Shi et al. developed HIT-scISOseq, a method designed to remove most artifactual cDNAs and concatenate multiple cDNAs for HiFi platforms, enabling high-throughput and high-accuracy scRNA-seq [122]. Building on this approach, Wang et al. adapted HIT-scISOseq for low-throughput cell analysis and

**Table 2 Summary of properties and applications of HiFi and ONT sequencing**

Feature	HiFi sequencing	ONT sequencing
Technology type	CCS	Nanopore sequencing
Read length	Up to 25 kb (typically)	Over hundreds of thousands of bases long or even exceeding a megabase
Accuracy	Q33	~ Q20 (up to Q30 for Duplex)
Read structure	Circular DNA template allows for multiple reads, generating a HiFi consensus	Linear DNA template read directly, producing single reads
Input	DNA	DNA, RNA
Typical runtime	24 h	72 h
Typical output file size (type)	55 GB (BAM)	1300 GB (fast5/pod5)
Detectable DNA modification	5mC	5mC, 5hmC, and 6mA
Variant calling	More accurate	Less accurate, inaccurately called indels
Chromosome-level genome assembly	Initial contig assembly combined with long-range data	Initial contig assembly with long-range data; polishing needed with HiFi or short-read data
T2T genome assembly	Initial assembly graph construction; phasing over heterozygous variants; limited in read length, need ultra-long reads	Resolving tangles; phasing through homozygous regions; need initial HiFi assembly

Note: ONT, Oxford Nanopore Technologies; CCS, circular consensus sequencing; T2T, telomere-to-telomere; BAM, Binary Alignment/Map; 5mC, 5-methylcytosine; 5hmC, 5-hydroxymethylcytosine; 6mA, N<sup>6</sup>-methyldeoxyadenosine.

applied it to sequence isoforms in single blastomeres of mouse preimplantation embryos [123]. Notably, the authors discovered 3894 transposable element loci that displayed dynamic changes during preimplantation development, which are likely involved in regulating the expression of neighboring genes.

Both ONT and PacBio technologies have been utilized in combination with scRNA-seq. Deng et al. conducted a systematic evaluation of scRNA-seq analysis performance using these two widely adopted long-read sequencing platforms [124]. In addition to enabling gene expression analysis, which is the primary application of NGS-based scRNA-seq, long-read scRNA-seq offers advanced capabilities such as gene splicing analysis and the identification of novel isoforms. Among long-read technologies, PacBio HiFi outperforms ONT in terms of sequencing quality, making it more effective for accurately identifying novel transcripts and detecting allele-specific gene and isoform expression [124].

## Comparative features and applications of HiFi and ONT sequencing

The comparative features and applications of HiFi sequencing and ONT sequencing highlight the strengths and trade-offs between these two state-of-the-art long-read sequencing platforms, such that users can make informed choices based on their individual needs (Table 2) [9,125]. HiFi sequencing utilizes CCS, achieving extremely high accuracy (Q33) through repeated reads of circular DNA templates. However, the accuracy of ONT (Q20) is lower, which may result in indels or other inaccuracies [126]. ONT Duplex sequencing reads have been reported to achieve an accuracy of up to Q30 [127,128]. However, it has not yet been commercialized, and we are looking forward to its development. HiFi offers moderate read lengths and excels in applications requiring high precision, such as initial assembly graph construction and phasing over heterozygous variants that are less than 10 kb apart [10]. In contrast, ONT sequencing is capable of generating ultra-long reads, often exceeding hundreds of thousands of bases or even a megabase, which helps compensate for the limitation in HiFi read length. ONT ultra-long reads are invaluable for resolving tangles and phasing

through homozygous regions over 100 kb in length [10,45]. As such, HiFi and ONT ultra-long sequencing reads should undoubtedly complement each other in T2T genome assembly. Two widely used tools for co-assembling these data are Verkko [14] and hifiasm [45], as discussed above. HiFi sequencing is also typically used for detecting DNA 5mC modifications, whereas ONT provides direct, real-time sequencing of both DNA and RNA, and can detect a broader range of DNA modifications including 5mC, 5-hydroxymethylcytosine (5hmC), and N<sup>6</sup>-methyldeoxyadenosine (6mA) (Table 2).

## Remaining challenges and future perspectives

### Read length of HiFi data

Long-read sequencing offers unique advantages compared to short-read sequencing, with increased read lengths greatly benefiting genome assembly and the sequencing of complex regions such as centromeres, telomeres, rDNA regions, or large structural rearrangements. Currently, the two leading platforms for generating long-read data are PacBio and ONT, each with its own strengths. In general, HiFi technology yields sequences of higher accuracy compared to ONT, although ONT stands out both for its ability to produce ultra-long reads (up to 4 Mb) [129] and its scalability, ranging from small portable sequencers to larger benchtop instruments. In the field of personalized cancer genomics, there is a pressing demand for a bioinformatics algorithm that seamlessly integrates data from both the ONT ultra-long and HiFi sequencing platforms to compensate for the relative shortcomings in HiFi read lengths.

### Relatively high cost of HiFi sequencing

HiFi sequencing (Revio) is approximately four times more expensive compared to short-read sequencing platform (Illumina NovaSeq 6000 S4) [8,130]. The challenge presented by the high cost of HiFi sequencing is multifaceted. First, the significant initial investment required for equipment, reagents, and personnel training can serve as a significant barrier to entry for numerous research institutions and organizations. Furthermore, ongoing expenses related to

maintenance, sample preparation, and data analysis compound the financial burden. Consequently, the steep price of HiFi sequencing limits access to cutting-edge genomic technologies, thereby constraining both the breadth and pace of scientific discovery and medical progress. Addressing this challenge involves exploring cost-saving strategies and enhancing efficiency in sequencing workflows, bolstering support for genomic research initiatives. One such strategy is targeted HiFi sequencing, which enables users to focus their sequencing efforts on specific genomic regions of interest. These regions may include intricate areas such as polymorphic human leucocyte antigen genes, centromeres, and VNTRs across the human genome. Another strategy to reduce costs is to increase throughput. PacBio users can attempt this by staying up to date with the latest single-molecule real-time (SMRT) cell versions, which offer substantial improvements. For example, the PacBio Revio system uses advanced SMRT cells with 25 million zero-mode waveguides, providing approximately a 15-fold increase in throughput compared to the earlier Sequel IIe system.

### Biases in library preparation

The throughput of HiFi sequencing is greatly influenced by molecular damage during library preparation. Therefore, it is crucial to develop appropriate protocols for extracting high-molecular-weight (HMW) DNA and methods for size selection. Accurate size selection of HMW DNA is vital for library preparation because long-read sequencing technology may exhibit bias toward sequencing smaller, more rapidly diffusing molecules [131]. This can have significant implications for downstream analyses, such as genome assembly and variant detection, where a comprehensive and unbiased view of the genome is essential. The bias toward smaller molecules not only affects the quality of the sequencing data but also the efficiency of the library preparation process. The presence of a bias can lead to an overrepresentation of certain genomic regions at the expense of others, which complicates the interpretation of sequencing results and potentially obscures biologically relevant information.

Given these challenges, the development of innovative techniques, such as droplet microfluidics, is a significant step forward in addressing the issue of bias in library preparation for HiFi sequencing. Droplet microfluidics utilizes a large number of microfabricated droplets (tens of thousands to millions) with picoliter–nanoliter volumes [132]. As illustrated by Figure 6, each droplet encapsulates, on average, less than one DNA molecule, effectively eliminating competition between HMW and low-molecular-weight DNA fragments

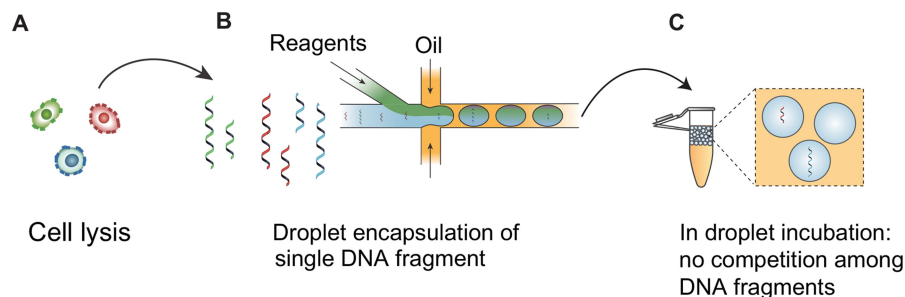
during critical library preparation steps such as ligation and polymerase chain reaction (PCR) [121,133]. Additionally, the increased DNA concentration within droplets compared to bulk solutions further enhances library preparation efficiency [132].

Droplet microfluidics further expands its utility in HiFi sequencing by enabling efficient targeted sequencing. Similar to digital droplet PCR (ddPCR) [134], each droplet can be compartmentalized with not only a single DNA molecule but also one or more fluorogenic probes specific to the target sequence. Following PCR amplification, droplets are analyzed for fluorescence, generating a binary positive or negative signal for target presence. Utilizing a fluorescent assisted droplet sorter (FADS) [135,136], positive droplets can be isolated with high throughput (> 90% accuracy) for downstream HiFi sequencing. Notably, to address potential sample limitations for HiFi sequencing, multiple displacement amplification (MDA) can be integrated within the droplet workflow to increase the total DNA content per droplet prior to sorting [121]. Furthermore, the utilization of clustered regularly interspaced short palindromic repeats/CRISPR-associated protein 9 (CRISPR/Cas9)-based [137] and hybridization-based enrichment strategies [138] represents progress toward enhanced and streamlined methodologies for targeted HiFi sequencing [8].

By overcoming these outstanding challenges, HiFi sequencing will gain even more value, providing the opportunity to delve deeper into genetic, epigenetic, and transcriptomic variations, along with their intricate connections to precision medicine and diagnosis.

### HiFi sequencing in precision medicine and diagnosis

HiFi long-read sequencing has emerged as a transformative technology in genomics due to its exceptional accuracy, long-read capabilities, and comprehensive coverage of complex regions in the genome. The fields of precision medicine and personalized diagnosis increasingly require highly accurate genomic data to identify mutations, SVs, and other genetic anomalies associated with an individual's response to disease [139]. By broadening the spectrum of detectable variants to more complex genomic regions, HiFi sequencing can improve the accuracy of diagnosing genetic diseases, identifying disease subtypes, and tailoring personalized treatment regimens for conditions like cancer, rare genetic disorders, and cardiovascular diseases [109,140]. Despite the diverse applications of HiFi sequencing technologies, their routine use in clinical settings nevertheless faces unique challenges [139]. One of



**Figure 6** A typical workflow of droplet-based microfluidics for single DNA fragment experiment

**A.** Cells are lysed and DNA fragments are collected. **B.** DNA fragments from cells and desired reagents are encapsulated into droplets with each droplet containing less than one DNA fragment. **C.** In droplet incubation to ensure there is no competition among DNA fragments.

the key obstacles is assembling complete cancer genomes to fully capture and resolve complex genetic variations, such as SVs and large rearrangements [141–143]. Additionally, creating personalized pangenome references that incorporate functional information, including cellular epigenetics and transcriptomic changes, remains an important goal for more comprehensive analysis [144–146].

To overcome these challenges, it is essential to improve the stability and reliability of cancer sequencing technologies. Furthermore, advancements in bioinformatics algorithms are needed to better analyze and interpret the vast amounts of data generated. Successfully integrating these developments into precision medicine and diagnostic laboratories will be critical for realizing the full potential of HiFi sequencing in clinical practice.

### CRedit author statement

**Bo Wang:** Investigation, Writing – original draft, Visualization, Funding acquisition. **Peng Jia:** Investigation, Writing – original draft, Visualization. **Shenghan Gao:** Investigation, Visualization. **Huanhuan Zhao:** Investigation, Visualization. **Gaoyang Zheng:** Visualization. **Lin Feng Xu:** Investigation, Writing – original draft, Visualization. **Kai Ye:** Conceptualization, Supervision, Writing – review & editing, Funding acquisition. All authors have read and approved the final manuscript.

### Competing interests

The authors have declared no competing interests.

### Acknowledgments

This study was supported by the National Key R&D Program of China (Grant Nos. 2022YFC3400300 and 2023YFF0613300) and the National Natural Science Foundation of China (Grant Nos. 32200510, 32400509, 32125009, 32070663, 32371516, and 32400520). We wish to extend our sincere gratitude to Professor Stephen J. Bush of Xi'an Jiaotong University for his invaluable assistance in language editing of this paper.

### ORCID

0000-0002-9041-878X (Bo Wang)

0000-0002-3429-919X (Peng Jia)

0000-0002-3810-6527 (Shenghan Gao)

0009-0003-8641-6487 (Huanhuan Zhao)

0000-0002-0854-5481 (Gaoyang Zheng)

0000-0002-6386-1815 (Lin Feng Xu)

0000-0002-2851-6741 (Kai Ye)

### References

- [1] Lappalainen T, Scott AJ, Brandt M, Hall IM. Genomic analysis in the age of human genome sequencing. *Cell* 2019;177:70–84.
- [2] Olson ND, Wagner J, Dwarshuis N, Miga KH, Sedlazeck FJ, Salit M, et al. Variant calling and benchmarking in an era of complete human genome sequences. *Nat Rev Genet* 2023; 24:464–83.
- [3] Rhoads A, Au KF. PacBio sequencing and its applications. *Genomics Proteomics Bioinformatics* 2015;13:278–89.
- [4] Lu H, Giordano F, Ning Z. Oxford nanopore MinION sequencing and genome assembly. *Genomics Proteomics Bioinformatics* 2016;14:265–79.
- [5] Wenger AM, Peluso P, Rowell WJ, Chang PC, Hall RJ, Concepcion GT, et al. Accurate circular consensus long-read sequencing improves variant detection and assembly of a human genome. *Nat Biotechnol* 2019;37:1155–62.
- [6] Miga KH. Centromere studies in the era of 'telomere-to-telomere' genomics. *Exp Cell Res* 2020;394:112127.
- [7] Logsdon GA, Vollger MR, Hsieh P, Mao Y, Liskovych MA, Koren S, et al. The structure, function and evolution of a complete human chromosome 8. *Nature* 2021;593:101–7.
- [8] Mastroiusta FK, Miller DE, Eichler EE. Applications of long-read sequencing to Mendelian genetics. *Genome Med* 2023; 15:42.
- [9] Logsdon GA, Vollger MR, Eichler EE. Long-read human genome sequencing and its applications. *Nat Rev Genet* 2020; 21:597–614.
- [10] Li H, Durbin R. Genome assembly in the telomere-to-telomere era. *Nat Rev Genet* 2024;25:658–70.
- [11] Miga KH, Koren S, Rhie A, Vollger MR, Gershman A, Bzikadze A, et al. Telomere-to-telomere assembly of a complete human X chromosome. *Nature* 2020;585:79–84.
- [12] Nurk S, Koren S, Rhie A, Rautiainen M, Bzikadze AV, Mikheenko A, et al. The complete sequence of a human genome. *Science* 2022;376:44–53.
- [13] Jarvis ED, Formenti G, Rhie A, Guarracino A, Yang C, Wood J, et al. Semi-automated assembly of high-quality diploid human reference genomes. *Nature* 2022;611:519–31.
- [14] Rautiainen M, Nurk S, Walenz BP, Logsdon GA, Porubsky D, Rhie A, et al. Telomere-to-telomere assembly of diploid chromosomes with Verkko. *Nat Biotechnol* 2023;41:1474–82.
- [15] Yang C, Zhou Y, Song Y, Wu D, Zeng Y, Nie L, et al. The complete and fully-phased diploid genome of a male Han Chinese. *Cell Res* 2023;33:745–61.
- [16] He Y, Chu Y, Guo S, Hu J, Li R, Zheng Y, et al. T2T-YAO: a telomere-to-telomere assembled diploid reference genome for Han Chinese. *Genomics Proteomics Bioinformatics* 2023; 21:1085–100.
- [17] Hou X, Wang D, Cheng Z, Wang Y, Jiao Y. A near-complete assembly of an *Arabidopsis thaliana* genome. *Mol Plant* 2022; 15:1247–50.
- [18] Wang B, Yang X, Jia Y, Xu Y, Jia P, Dang N, et al. High-quality *Arabidopsis thaliana* genome assembly with nanopore and HiFi long reads. *Genomics Proteomics Bioinformatics* 2022; 20:4–13.
- [19] Naish M, Alonge M, Wlodzimierz P, Tock AJ, Abramson BW, Schmücker A, et al. The genetic and epigenetic landscape of the *Arabidopsis* centromeres. *Science* 2021;374:eabi7489.
- [20] Bi G, Zhao S, Yao J, Wang H, Zhao M, Sun Y, et al. Near telomere-to-telomere genome of the model plant *Physcomitrium patens*. *Nat Plants* 2024;10:327–43.
- [21] Payne ZL, Penny GM, Turner TN, Dutcher SK. A gap-free genome assembly of *Chlamydomonas reinhardtii* and detection of translocations induced by CRISPR-mediated mutagenesis. *Plant Commun* 2023;4:100493.
- [22] Song JM, Xie WZ, Wang S, Guo YX, Koo DH, Kudrna D, et al. Two gap-free reference genomes and a global view of the centromere architecture in rice. *Mol Plant* 2021;14:1757–67.
- [23] Chen J, Wang Z, Tan K, Huang W, Shi J, Li T, et al. A complete telomere-to-telomere assembly of the maize genome. *Nat Genet* 2023;55:1221–31.
- [24] Wang L, Zhang M, Li M, Jiang X, Jiao W, Song Q. A telomere-to-telomere gap-free assembly of soybean genome. *Mol Plant* 2023;16:1711–4.
- [25] Yue J, Chen Q, Wang Y, Zhang L, Ye C, Wang X, et al. Telomere-to-telomere and gap-free reference genome assembly of the kiwifruit *Actinidia chinensis*. *Hortic Res* 2023; 10:uhac264.

- [26] Wang L, Li LL, Chen L, Zhang RG, Zhao SW, Yan H, et al. Telomere-to-telomere and haplotype-resolved genome assembly of the Chinese cork oak (*Quercus variabilis*). *Front Plant Sci* 2023;14:1290913.
- [27] Wei C, Gao L, Xiao R, Wang Y, Chen B, Zou W, et al. Complete telomere-to-telomere assemblies of two sorghum genomes to guide biological discovery. *Imeta* 2024;3:e193.
- [28] Gao S, Jia Y, Guo H, Xu T, Wang B, Bush SJ, et al. The centromere landscapes of four karyotypically diverse *Papaver* species provide insights into chromosome evolution and speciation. *Cell Genomics* 2024;4:100626.
- [29] Wang B, Jia Y, Dang N, Yu J, Bush SJ, Gao S, et al. Near telomere-to-telomere genome assemblies of two *Chlorella* species unveil the composition and evolution of centromeres in green algae. *BMC Genomics* 2024;25:356.
- [30] Huang Z, Xu Z, Bai H, Huang Y, Kang N, Ding X, et al. Evolutionary analysis of a complete chicken genome. *Proc Natl Acad Sci U S A* 2023;120:e2216641120.
- [31] Sun Z, Li S, Liu Y, Li W, Liu K, Cao X, et al. Telomere-to-telomere gapless genome assembly of the Chinese sea bass (*Lateolabrax maculatus*). *Sci Data* 2024;11:175.
- [32] Zhao H, Zhou H, Sun G, Dong B, Zhu W, Mu X, et al. Telomere-to-telomere genome assembly of the goose *Anser cygnoides*. *Sci Data* 2024;11:741.
- [33] Yin D, Chen C, Lin D, Hua Z, Ying C, Zhang J, et al. Telomere-to-telomere gap-free genome assembly of the endangered Yangtze finless porpoise and East Asian finless porpoise. *Gigascience* 2024;13:giae067.
- [34] Nurk S, Walenz BP, Rhie A, Vollger MR, Logsdon GA, Grothe R, et al. HiCanu: accurate assembly of segmental duplications, satellites, and allelic variants from high-fidelity long reads. *Genome Res* 2020;30:1291–305.
- [35] Cheng H, Concepcion GT, Feng X, Zhang H, Li H. Haplotype-resolved *de novo* assembly using phased assembly graphs with hifiasm. *Nat Methods* 2021;18:170–5.
- [36] Kolmogorov M, Yuan J, Lin Y, Pevzner PA. Assembly of long, error-prone reads using repeat graphs. *Nat Biotechnol* 2019;37:540–6.
- [37] Feng X, Cheng H, Portik D, Li H. Metagenome assembly of high-fidelity long reads with hifiasm-meta. *Nat Methods* 2022;19:671–4.
- [38] Kolmogorov M, Bickhart DM, Behsaz B, Gurevich A, Rayko M, Shin SB, et al. metaFlye: scalable long-read metagenome assembly using repeat graphs. *Nat Methods* 2020;17:1103–10.
- [39] Mills C, Muruganujan A, Ebert D, Marconett CN, Lewinger JP, Thomas PD, et al. PEREGRINE: a genome-wide prediction of enhancer to gene relationships supported by experimental evidence. *PLoS One* 2020;15:e0243791.
- [40] Shafin K, Pesout T, Chang PC, Nattestad M, Kolesnikov A, Goel S, et al. Haplotype-aware variant calling with PEPPER-Margin-DeepVariant enables high accuracy in nanopore long-reads. *Nat Methods* 2021;18:1322–32.
- [41] Xiao CL, Chen Y, Xie SQ, Chen KN, Wang Y, Han Y, et al. MECAT: fast mapping, error correction, and *de novo* assembly for single-molecule sequencing reads. *Nat Methods* 2017;14:1072–4.
- [42] Li H. Minimap and miniasm: fast mapping and *de novo* assembly for noisy long sequences. *Bioinformatics* 2016;32:2103–10.
- [43] Hu J, Wang Z, Sun Z, Hu B, Ayoola AO, Liang F, et al. NextDenovo: an efficient error correction and accurate assembly tool for noisy long reads. *Genome Biol* 2024;25:107.
- [44] Yu W, Luo H, Yang J, Zhang S, Jiang H, Zhao X, et al. Comprehensive assessment of 11 *de novo* HiFi assemblers on complex eukaryotic genomes and metagenomes. *Genome Res* 2024;34:326–40.
- [45] Cheng H, Asri M, Lucas J, Koren S, Li H. Scalable telomere-to-telomere assembly for diploid and polyploid genomes with double graph. *Nat Methods* 2024;21:967–70.
- [46] Wang B, Dang N, Yang X, Xu S, Ye K. The human pangenome reference: the beginning of a new era for genomics. *Sci Bull (Beijing)* 2023;68:1484–7.
- [47] Hu J, Wang Z, Liang F, Liu SL, Ye K, Wang DP. NextPolish2: a repeat-aware polishing tool for genomes assembled using HiFi long reads. *Genomics Proteomics Bioinformatics* 2024;22:qzad009.
- [48] Walker BJ, Abeel T, Shea T, Priest M, Abouelliel A, Sakthikumar S, et al. Pilon: an integrated tool for comprehensive microbial variant detection and genome assembly improvement. *PLoS One* 2014;9:e112963.
- [49] Vaser R, Sovic I, Nagarajan N, Sikic M. Fast and accurate *de novo* genome assembly from long uncorrected reads. *Genome Res* 2017;27:737–46.
- [50] Hu J, Fan J, Sun Z, Liu S. NextPolish: a fast and efficient genome polishing tool for long-read assembly. *Bioinformatics* 2020;36:2253–5.
- [51] Mc Cartney AM, Shafin K, Alonge M, Bzikadze AV, Formenti G, Fungtammasan A, et al. Chasing perfection: validation and polishing strategies for telomere-to-telomere genome assemblies. *Nat Methods* 2022;19:687–95.
- [52] Quince C, Walker AW, Simpson JT, Loman NJ, Segata N. Shotgun metagenomics, from sampling to analysis. *Nat Biotechnol* 2017;35:833–44.
- [53] Lapidus AL, Korobeynikov AI. Metagenomic data assembly — the way of decoding unknown microorganisms. *Front Microbiol* 2021;12:613791.
- [54] Yang C, Chowdhury D, Zhang Z, Cheung WK, Lu A, Bian Z, et al. A review of computational tools for generating metagenome-assembled genomes from metagenomic sequencing data. *Comput Struct Biotechnol J* 2021;19:6301–14.
- [55] Benoit G, Raguideau S, James R, Phillippy AM, Chikhi R, Quince C. High-quality metagenome assembly from long accurate reads with metaMDBG. *Nat Biotechnol* 2024;42:1378–83.
- [56] Zhang Y, Jiang F, Yang B, Wang S, Wang H, Wang A, et al. Improved microbial genomes and gene catalog of the chicken gut from metagenomic sequencing of high-fidelity long reads. *Gigascience* 2022;11:giac116.
- [57] Bickhart DM, Kolmogorov M, Tseng E, Portik DM, Korobeynikov A, Tolstoganov I, et al. Generating lineage-resolved, complete metagenome-assembled genomes from complex microbial communities. *Nat Biotechnol* 2022;40:711–9.
- [58] Talbert PB, Henikoff S. What makes a centromere? *Exp Cell Res* 2020;389:111895.
- [59] Sevim V, Bashir A, Chin CS, Miga KH. Alpha-CENTAURI: assessing novel centromeric repeat sequence variation with long read sequencing. *Bioinformatics* 2016;32:1921–4.
- [60] Kunyavskaya O, Dvorkina T, Bzikadze AV, Alexandrov IA, Pevzner PA. Automated annotation of human centromeres with HORmon. *Genome Res* 2022;32:1137–51.
- [61] Gao S, Yang X, Guo H, Zhao X, Wang B, Ye K. HiCAT: a tool for automatic annotation of centromere structure. *Genome Biol* 2023;24:58.
- [62] Gao S, Zhang Y, Bush SJ, Wang B, Yang X, Ye K. Centromere landscapes resolved from hundreds of human genomes. *Genomics Proteomics Bioinformatics* 2024;22:qzae071.
- [63] Wlodzimierz P, Hong M, Henderson IR. TRASH: tandem repeat annotation and structural hierarchy. *Bioinformatics* 2023;39:btad308.
- [64] Rautiainen M. Ribotin: automated assembly and phasing of rDNA morphs. *Bioinformatics* 2024;40:btac124.
- [65] Poplin R, Chang PC, Alexander D, Schwartz S, Colthurst T, Ku A, et al. A universal SNP and small-indel variant caller using deep neural networks. *Nat Biotechnol* 2018;36:983–7.
- [66] Zheng Z, Li S, Su J, Leung AW, Lam TW, Luo R. Symphonizing pileup and full-alignment for deep learning-based long-read variant calling. *Nat Comput Sci* 2022;2:797–803.
- [67] Sedlazeck FJ, Rescheneder P, Smolka M, Fang H, Nattestad M, von Haeseler A, et al. Accurate detection of complex structural

- variations using single-molecule sequencing. *Nat Methods* 2018;15:461–8.
- [68] Smolka M, Paulin LF, Grochowski CM, Horner DW, Mahmoud M, Behera S, et al. Detection of mosaic and population-level structural variants with Sniffles2. *Nat Biotechnol* 2024;42:1571–80.
- [69] Heller D, Vingron M. SVIM: structural variant identification using mapped long reads. *Bioinformatics* 2019;35:2907–15.
- [70] Jiang T, Liu Y, Jiang Y, Li J, Gao Y, Cui Z, et al. Long-read-based human genomic structural variation detection with cuteSV. *Genome Biol* 2020;21:189.
- [71] Lin J, Wang S, Audano PA, Meng D, Flores JJ, Kusters W, et al. SVision: a deep learning approach to resolve complex structural variants. *Nat Methods* 2022;19:1230–3.
- [72] Wang S, Lin J, Jia P, Xu T, Li X, Liu Y, et al. *De novo* and somatic structural variant discovery with SVision-pro. *Nat Biotechnol* 2025;43:181–5.
- [73] Chen Y, Wang AY, Barkley CA, Zhang Y, Zhao X, Gao M, et al. Deciphering the exact breakpoints of structural variations using long sequencing reads with DeBreak. *Nat Commun* 2023;14:283.
- [74] Shiraishi Y, Koya J, Chiba K, Okada A, Arai Y, Saito Y, et al. Precise characterization of somatic complex structural variations from tumor/control paired long-read sequencing data with nanomonsv. *Nucleic Acids Res* 2023;51:e74.
- [75] Bakhtiari M, Shleizer-Burko S, Gymrek M, Bansal V, Bafna V. Targeted genotyping of variable number tandem repeats with adVNTR. *Genome Res* 2018;28:1709–19.
- [76] Mitsuhashi S, Frith MC, Mizuguchi T, Miyatake S, Toyota T, Adachi H, et al. Tandem-genotypes: robust detection of tandem repeat expansions from long DNA reads. *Genome Biol* 2019;20:58.
- [77] Chiu R, Rajan-Babu IS, Friedman JM, Birol I. Straglr: discovering and genotyping tandem repeat expansions using whole genome long-read sequences. *Genome Biol* 2021;22:224.
- [78] Ichikawa K, Kawahara R, Asano T, Morishita S. A landscape of complex tandem repeats within individual human genomes. *Nat Commun* 2023;14:5530.
- [79] Dolzhenko E, English A, Dashnow H, De Sena Brandine G, Mokveld T, Rowell WJ, et al. Characterization and visualization of tandem repeats at genome scale. *Nat Biotechnol* 2024;42:1606–14.
- [80] Jam HZ, Zook JM, Javadzadeh S, Park J, Sehgal A, Gymrek M. Genome-wide profiling of genetic variation at tandem repeat from long reads. *bioRxiv* 2024;576266.
- [81] Heller D, Vingron M. SVIM-asm: structural variant detection from haploid and diploid genome assemblies. *Bioinformatics* 2021;36:5519–21.
- [82] Ebert P, Audano PA, Zhu Q, Rodriguez-Martin B, Porubsky D, Bonder MJ, et al. Haplotype-resolved diverse human genomes and integrated analysis of structural variation. *Science* 2021;372:eabf7117.
- [83] 1000 Genomes Project Consortium, Abecasis GR, Auton A, Brooks LD, DePristo MA, Durbin RM, et al. An integrated map of genetic variation from 1,092 human genomes. *Nature* 2012;491:56–65.
- [84] Chen S, Francioli LC, Goodrich JK, Collins RL, Kanai M, Wang Q, et al. A genomic mutational constraint map using variation in 76,156 human genomes. *Nature* 2024;625:92–100.
- [85] Cancer Genome Atlas Research Network, Weinstein JN, Collisson EA, Mills GB, Shaw KR, Ozenberger BA, et al. The Cancer Genome Atlas Pan-Cancer analysis project. *Nat Genet* 2013;45:1113–20.
- [86] Olson ND, Wagner J, McDaniel J, Stephens SH, Westreich ST, Prasanna AG, et al. PrecisionFDA Truth Challenge V2: calling variants from short and long reads in difficult-to-map regions. *Cell Genom* 2022;2:100129.
- [87] Conlin LK, Aref-Eshghi E, McClelland DA, Luo M, Rajagopalan R. Long-read sequencing for molecular diagnostics in constitutional genetic disorders. *Hum Mutat* 2022;43:1531–44.
- [88] Jia P, Dong L, Yang X, Wang B, Bush SJ, Wang T, et al. Haplotype-resolved assemblies and variant benchmark of a Chinese Quartet. *Genome Biol* 2023;24:277.
- [89] Vollger MR, Dishuck PC, Harvey WT, DeWitt WS, Guitart X, Goldberg ME, et al. Increased mutation and gene conversion within human segmental duplications. *Nature* 2023;617:325–34.
- [90] Chen X, Baker D, Dolzhenko E, Devaney JM, Noya J, Berlyoung AS, et al. Genome-wide profiling of highly similar paralogous genes using HiFi sequencing. *Nat Commun* 2025;16:2340.
- [91] Liao X, Zhu W, Zhou J, Li H, Xu X, Zhang B, et al. Repetitive DNA sequence detection and its role in the human genome. *Commun Biol* 2023;6:954.
- [92] Cui Y, Ye W, Li JS, Li JJ, Vilain E, Sallam T, et al. A genome-wide spectrum of tandem repeat expansions in 338,963 humans. *Cell* 2024;187:2336–41.e5.
- [93] Depienne C, Mandel JL. 30 years of repeat expansion disorders: what have we learned and what are the remaining challenges? *Am J Hum Genet* 2021;108:764–85.
- [94] Chen X, Harting J, Farrow E, Thiffault I, Kasperaviciute D, Genomics England Research Consortium, et al. Comprehensive SMN1 and SMN2 profiling for spinal muscular atrophy analysis using long-read PacBio HiFi sequencing. *Am J Hum Genet* 2023;110:240–50.
- [95] Hiatt SM, Lawlor JMJ, Handley LH, Latner DR, Bonnstetter ZT, Finnila CR, et al. Long-read genome sequencing and variant reanalysis increase diagnostic yield in neurodevelopmental disorders. *medRxiv* 2024;24304633.
- [96] Lu TY, Smaruj PN, Fudenberg G, Mancuso N, Chaisson MJP. The motif composition of variable number tandem repeats impacts gene expression. *Genome Res* 2023;33:511–24.
- [97] Rajan-Babu IS, Dolzhenko E, Eberle MA, Friedman JM. Sequence composition changes in short tandem repeats: heterogeneity, detection, mechanisms and clinical implications. *Nat Rev Genet* 2024;25:476–99.
- [98] Patterson M, Marschall T, Pisanti N, van Iersel L, Stougie L, Klau GW, et al. WhatsHap: weighted haplotype assembly for future-generation sequencing reads. *J Comput Biol* 2015;22:498–509.
- [99] Lin JH, Chen LC, Yu SC, Huang YT. LongPhase: an ultra-fast chromosome-scale phasing algorithm for small and large variants. *Bioinformatics* 2022;38:1816–22.
- [100] Holt JM, Saunders CT, Rowell WJ, Kronenberg Z, Wenger AM, Eberle M. HiPhase: jointly phasing small, structural, and tandem repeat variants from HiFi sequencing. *Bioinformatics* 2024;40:btac042.
- [101] Zhou Q, Ji F, Lin D, Liu X, Zhu Z, Ruan J. KSNP: a fast de Bruijn graph-based haplotyping tool approaching data-in time cost. *Nat Commun* 2024;15:3126.
- [102] Cheung WA, Johnson AF, Rowell WJ, Farrow E, Hall R, Cohen ASA, et al. Direct haplotype-resolved 5-base HiFi sequencing for genome-wide profiling of hypermethylation outliers in a rare disease cohort. *Nat Commun* 2023;14:3090.
- [103] Tse OYO, Jiang P, Cheng SH, Peng W, Shang H, Wong J, et al. Genome-wide detection of cytosine methylation by single molecule real-time sequencing. *Proc Natl Acad Sci U S A* 2021;118:e2019768118.
- [104] Sigurpalsdottir BD, Stefansson OA, Holley G, Beyter D, Zink F, Hardarson M, et al. A comparison of methods for detecting DNA methylation from long-read sequencing of human genomes. *Genome Biol* 2024;25:69.
- [105] Ni P, Nie F, Zhong Z, Xu J, Huang N, Zhang J, et al. DNA 5-methylcytosine detection and methylation phasing using PacBio circular consensus sequencing. *Nat Commun* 2023;14:4054.
- [106] Mastroianni FK, Rozanski AN, Harvey WT, Knuth J, Garcia G, Munson KM, et al. Complete chromosome 21 centromere sequences from a Down syndrome family reveal size asymmetry

- and differences in kinetochore attachment. *bioRxiv* 2024;581464.
- [107] Chen Y, Wu B, Ding Y, Niu L, Bai X, Lin Z, et al. High accuracy methylation identification tools on single molecular level for PacBio HiFi data. *bioRxiv* 2024;607879.
- [108] Mertens F, Johansson B, Fioretos T, Mitelman F. The emerging complexity of gene fusions in cancer. *Nat Rev Cancer* 2015; 15:371–81.
- [109] Dondi A, Lischetti U, Jacob F, Singer F, Borgsmüller N, Coelho R, et al. Detection of isoforms and genomic alterations by high-throughput full-length single-cell RNA sequencing in ovarian cancer. *Nat Commun* 2023;14:7780.
- [110] Liu Q, Hu Y, Stucky A, Fang L, Zhong JF, Wang K. LongGF: computational algorithm and software tool for fast and accurate detection of gene fusions by long-read transcriptome sequencing. *BMC Genomics* 2020;21:793.
- [111] Davidson NM, Chen Y, Sadras T, Ryland GL, Blombery P, Ekert PG, et al. JAFFAL: detecting fusion genes with long-read transcriptome sequencing. *Genome Biol* 2022;23:10.
- [112] Chen Y, Wang Y, Chen W, Tan Z, Song Y, Human Genome Structural Variation Consortium, et al. Gene fusion detection and characterization in long-read cancer transcriptome sequencing data with FusionSeeker. *Cancer Res* 2023;83:28–33.
- [113] Volden R, Kronenberg Z, Gillmor A, Verhey T, Monument M, Senger D, et al. Abstract LB078: pbfusion: Detecting gene-fusion and other transcriptional abnormalities using PacBio HiFi data. *Cancer Res* 2023;83:LB078.
- [114] Qin Q, Popic V, Yu H, White E, Khorgade A, Shin A, et al. CTAT-LR-fusion: accurate fusion transcript identification from long and short read isoform sequencing at bulk or single cell resolution. *bioRxiv* 2024;581862.
- [115] Al'Khafaji AM, Smith JT, Garimella KV, Babadi M, Popic V, Sade-Feldman M, et al. High-throughput RNA isoform sequencing using programmed cDNA concatenation. *Nat Biotechnol* 2024;42:582–6.
- [116] Jain M, Abu-Shumays R, Olsen HE, Akeson M. Advances in nanopore direct RNA sequencing. *Nat Methods* 2022; 19:1160–4.
- [117] Dong X, Zhang L, Milholland B, Lee M, Maslov AY, Wang T, et al. Accurate identification of single-nucleotide variants in whole-genome-amplified single cells. *Nat Methods* 2017; 14:491–3.
- [118] Zhang L, Dong X, Lee M, Maslov AY, Wang T, Vijg J. Single-cell whole-genome sequencing reveals the functional landscape of somatic mutations in B lymphocytes across the human lifespan. *Proc Natl Acad Sci U S A* 2019;116:9014–9.
- [119] Fan X, Yang C, Li W, Bai X, Zhou X, Xie H, et al. SMOOTH-seq: single-cell genome sequencing of human cells on a third-generation sequencing platform. *Genome Biol* 2021;22:195.
- [120] Xie H, Li W, Hu Y, Yang C, Lu J, Guo Y, et al. *De novo* assembly of human genome at single-cell levels. *Nucleic Acids Res* 2022;50:7479–92.
- [121] Hård J, Mold JE, Eisfeldt J, Tellgren-Roth C, Häggqvist S, Bunikis I, et al. Long-read whole-genome analysis of human single cells. *Nat Commun* 2023;14:5164.
- [122] Shi ZX, Chen ZC, Zhong JY, Hu KH, Zheng YF, Chen Y, et al. High-throughput and high-accuracy single-cell RNA isoform analysis using PacBio circular consensus sequencing. *Nat Commun* 2023;14:2631.
- [123] Wang C, Shi Z, Huang Q, Liu R, Su D, Chang L, et al. Single-cell analysis of isoform switching and transposable element expression during preimplantation embryonic development. *PLoS Biol* 2024;22:e3002505.
- [124] Deng E, Shen Q, Zhang J, Fang Y, Chang L, Luo G, et al. Systematic evaluation of single-cell RNA-seq analyses performance based on long-read sequencing platforms. *J Adv Res* 2025;71:141–53.
- [125] De Coster W, Weissensteiner MH, Sedlazeck FJ. Towards population-scale long-read sequencing. *Nat Rev Genet* 2021; 22:572–87.
- [126] Mahmoud M, Huang Y, Garimella K, Audano PA, Wan W, Prasad N, et al. Utility of long-read sequencing for All of Us. *Nat Commun* 2024;15:837.
- [127] Koren S, Bao Z, Guarracino A, Ou S, Goodwin S, Jenike KM, et al. Gapless assembly of complete human and plant chromosomes using only nanopore sequencing. *Genome Res* 2024;34:1919–30.
- [128] Kolesnikov A, Cook D, Nattestad M, Brambrink L, McNulty B, Gorzynski J, et al. Local read haplotagging enables accurate long-read small variant calling. *Nat Commun* 2024;15:5907.
- [129] Xu Y, Luo H, Wang Z, Lam HM, Huang C. Oxford Nanopore Technology: revolutionizing genomics research in plants. *Trends Plant Sci* 2022;27:510–1.
- [130] Eisenhofer R, Nesme J, Santos-Bay L, Koziol A, Sørensen SJ, Alberdi A, et al. A comparison of short-read, HiFi long-read, and hybrid strategies for genome-resolved metagenomics. *Microbiol Spectr* 2024;12:e0359023.
- [131] Warburton PE, Sebra RP. Long-read DNA sequencing: recent advances and remaining challenges. *Annu Rev Genomics Hum Genet* 2023;24:109–32.
- [132] Moragues T, Arguijo D, Beneyton T, Modavi C, Simutis K, Abate AR, et al. Droplet-based microfluidics. *Nat Rev Methods Primers* 2023;3:32.
- [133] Madsen EB, Höjjer I, Kvist T, Ameer A, Mikkelsen MJ. Xdrop: targeted sequencing of long DNA molecules from low input samples using droplet sorting. *Hum Mutat* 2020;41:1671–9.
- [134] Kiss MM, Ortoleva-Donnelly L, Beer NR, Warner J, Bailey CG, Colston BW, et al. High-throughput quantitative polymerase chain reaction in picoliter droplets. *Anal Chem* 2008; 80:8975–81.
- [135] Sciambi A, Abate AR. Accurate microfluidic sorting of droplets at 30 kHz. *Lab Chip* 2015;15:47–51.
- [136] Madrigal JL, Schoepp NG, Xu L, Powell CS, Delley CL, Siltanen CA, et al. Characterizing cell interactions at scale with made-to-order droplet ensembles (MODEs). *Proc Natl Acad Sci U S A* 2022;119:e2110867119.
- [137] Schultzhau Z, Wang Z, Stenger D. CRISPR-based enrichment strategies for targeted sequencing. *Biotechnol Adv* 2021; 46:107672.
- [138] Wang M, Beck CR, English AC, Meng Q, Buhay C, Han Y, et al. PacBio-LITS: a large-insert targeted sequencing method for characterization of human disease-associated chromosomal structural variations. *BMC Genomics* 2015;16:214.
- [139] Ermini L, Driguez P. The application of long-read sequencing to cancer. *Cancers* 2024;16:1275.
- [140] Penter L, Borji M, Nagler A, Lyu H, Lu WS, Cieri N, et al. Integrative genotyping of cancer and immune phenotypes by long-read sequencing. *Nat Commun* 2024;15:32.
- [141] Sakamoto Y, Sereewattanawoot S, Suzuki A. A new era of long-read sequencing for cancer genomics. *J Hum Genet* 2020; 65:3–10.
- [142] Erwin GS, Gürsoy G, Al-Abri R, Suriyaprakash A, Dolzhenko E, Zhu K, et al. Recurrent repeat expansions in human cancer genomes. *Nature* 2023;613:96–102.
- [143] Wang D, Liu B, Zhang Z. Accelerating the understanding of cancer biology through the lens of genomics. *Cell* 2023; 186:1755–71.
- [144] Kiri S, Ryba T. Cancer, metastasis, and the epigenome. *Mol Cancer* 2024;23:154.
- [145] Pradat Y, Viot J, Yurchenko AA, Gunbin K, Cerbone L, Deloger M, et al. Integrative pan-cancer genomic and transcriptomic analyses of refractory metastatic cancer. *Cancer Discov* 2023; 13:1116–43.
- [146] Sirén J, Eskandar P, Ungaro MT, Hickey G, Eizenga JM, Novak AM, et al. Personalized pangenome references. *Nat Methods* 2024;21:2017–23.

© The Author(s) 2025. Published by Oxford University Press and Science Press on behalf of the Beijing Institute of Genomics, Chinese Academy of Sciences / China National Center for Bioinformation and Genetics Society of China.  
This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted reuse, distribution, and reproduction in any medium, provided the original work is properly cited.  
Genomics, Proteomics & Bioinformatics, 2025, 23, 1–14  
<https://doi.org/10.1093/gpbjnl/qzaf003>  
Review