Academic Journal of Science and Technology
ISSN: 2771-3032 | Vol. 12, No. 2, 2024

Progress in the Application of CRISPR/Cas Family Mediated Third-generation Sequencing Technology

Xizhen Chen

College of Life Sciences, Lanzhou University, Lanzhou 730030, China

Abstract: The first generation sequencing technology based on the dideoxynucleotide (ddNTP) chain termination method proposed by Sanger was gradually eliminated due to its high cost, low sequencing read length and cumbersome process. The second generation of sequencing technology, called High-throughput sequencing (HTS), which was then developed, still has the problem of fixed read length. In recent years, the third generation sequencing technology represented by SMRT technology and Nanopore sequencing technology has gradually become popular. Compared with the previous two generations of sequencing technology, the most significant advantage of the third generation sequencing technology is its ability to carry out single molecule sequencing. In this process, the infinite length of nucleic acid sequence can be determined theoretically without the help of PCR amplification. This paper first introduces the basic principles, advantages and disadvantages of third-generation sequencing, and introduces in detail the CRISPR/Cas family-mediated SMRT technology and Nanopore sequencing technology in the third-generation sequencing technology. Finally, the research progress and prospects of the combination of the third generation sequencing technology and gene editing technology in the future are analyzed.

Keywords: Sequencing Technology; SMRT Technology; Nanopore Sequencing Technology; The CRISPR/Cas Family.

1. Introduction

As human understanding of the genome deepens, genomics has become the key to comprehending the complexity of living organisms. The diversity of microbial communities, the origins of genetic diseases, and the paths of biological evolution [1] are all directly related to the genome sequence. in order to accurately read the target gene sequences inside and outside the body as well as the changes in epigenetic modifications of genes [2], researchers have been exploring new high-tech sequencing technologies with higher accuracy, broader sequencing range, and fewer interfering factors.

The first-generation sequencing technology invented by Sanger and others, despite its milestone significance in history, was limited by its high cost, low efficiency, short read lengths, and low throughput. Although it was a sensation at the time, it could not be widely applied, thus heralding the development of high-throughput sequencing technologies. The emergence of second-generation sequencing technology (HTS) greatly reduced sequencing costs and increased data output, but it still did not solve the problem of short read lengths, and therefore still had certain limitations, unable to measure high CG regions and some complex target areas. In recent years, with single-molecule sequencing (SMRT sequencing) and nanopore sequencing leading the way, the third-generation sequencing technology has emerged. This technology has changed the previous limitations of sequencing technology that was bound to PCR amplification, and its high throughput, long read length, and no amplification are its highlights, quickly promoting it in the field of molecular biology. However, the downside is that due to its higher error rate compared to the previous two generations of sequencing technologies, it is far less practical than the second-generation sequencing technology when high-precision measurement is required.

At this time, the discovery of the CRISPR/Cas system, especially the proteins Cas9, Cas12, and Cas13, has brought revolutionary tools to gene editing technology. They are highly target-specific with a low off-target rate, easy to use, and when combined with third-generation sequencing technology to enhance their targeting, the application range of this system has been expanded like never before, from basic gene function research to complex disease treatment. This study aims to explore the potential and applications of the combination of the CRISPR/Cas system with third-generation sequencing technology.

2. An Overview of CRISPR/Cas Technology

2.1. Types of CRISPR/Cas Gene Editing Tools

The types of CRISPR/Cas systems found in nature are rich and diverse, and they can be broadly divided into two classes based on their protein effectors, with six types and various subtypes [2]. Class I includes Types I, III, and IV, which need multiple Cas proteins to form a multi-protein effector complex for coordinated work, accounting for 90% of known CRISPR types and widely present in bacteria and archaea. Class II systems are characterized by the usage of a single Cas protein to perform the cut, including Types II, V, and VI, which make up 10% of known types and are preferred for biotechnological applications due to their simplicity and thorough research. Among them, Cas9 from Streptococcus pyogenes (S. pyogenes Cas9, SpCas9) is widely applied in the realm of gene editing due to its simple and rapid characteristics. This system is guided by a dual RNA, including the tracrRNA that targets DNA and the crRNA that forms a complex with Cas9, with its recognition of the NGG PAM sequence facilitating its rapid development. In addition, researchers are also exploring other CRISPR systems with different characteristics, such as Cas12 and Cas13, as well as more than 70 Cas9 homologs, which provide different PAM sequence recognition and target selection, either enhancing the range of sequence recognition or improving the efficiency of gene editing. In summary, they have enriched the gene editing toolkit [3].

2.2. The Working Principle of the CRISPR/Cas Gene Editing Tool

The CRISPR/Cas system, an acquired bacterial and archaeal-originated adaptive immune system, consists of CRISPR RNA (crRNA) and Cas proteins. Resist foreign DNA such as viruses or plasmids by forming specific effector complexes. Among them, the CRISPR/Cas9 system, which includes crRNA, transactivating crRNA (tracrRNA), and Cas9 proteins, has been widely used in genome editing. crRNA can specifically identify target sites on the genome through complementary base pairing, and cooperate with tracrRNA to recruit Cas9 protein. In subsequent studies, the scientists simplified the technique by fusing crRNA and tracrRNA into a single guide RNA (sgRNA) [4]. The activity of Cas9 is also regulated by protospacer-adjacent motif,PAM. After the Cas9 protein scans the sequence, CRRNA-specific binding to the target site is required to create double-strand breaks (DSBs), allowing the target gene to complete cut.

After cutting, cells use endogenous repair mechanisms to repair DSBs, completing insertion, deletion, or base replacement. The major mechanisms include nonhomologous end joining (NHEJ) and homology directed repair (HDR), which usually result in random insertion or deletion of nucleotides silencing genes or targeted disruption of regulatory elements silencing downstream genes and the latter allows precise gene editing by inserting foreign DNA sequences [2].

Unlike Cas9, which requires two separate short Rnas, the Cas12a and Cas13 proteins use only one gRNA and cut at the target DNA site to form sticky ends, rather than producing flat ends as Cas9 does [3].

3. Third Generation Sequencing Technology

Third-generation sequencing technology, also known as single-molecule sequencing, represents a significant breakthrough in genomics research. This innovation enables the simultaneous sequencing of individual DNA molecules in real time, by passing the requirement for PCR amplification. It is characterized by high read length and real-time data output, providing a more complete view of the genome and more accurate detection of structural variations.

The main third-generation sequencing platforms currently available on the market include PacBio's SMRT technology and Oxford Nanopore Technologies' nanopore sequencing technology. These platforms' high throughput and long read capabilities give them unique advantages in genome assembly, the resolution of complex genomic regions, and metagenomics research. With the continuous advancement of the technology, third-generation sequencing technology is expected to play a more critical role in future biomedical research and clinical diagnostics.

3.1. SMRT Technology

At present, Single Molecule Real-Time (SMRT) Sequencing, developed by Pacific Biosciences (referred to as PacBio), is one of the mainstream third-generation sequencing platforms. It is important to note that before carrying out SMRT sequencing, a library must first be prepared from double-stranded DNA. The hairpin adapters is connected to the DNA molecule to form a closed-loop SMRTbell. Next, A primers and a polymerase are bound to the adapter, and the library is loaded into an SMRT Cell containing a large number of nanoscale observation chambers (ZMWs). The RSII system can load up to 150,000 ZMWs, while the Sequel platform can load up to 1 million[5].In distinction to Next-Generation Sequencing (NGS) techniques, Single Molecule Real-Time (SMRT) sequencing detects nucleotide incorporation in real-time on a per-molecule basis. This method also leverages incorporation timing to infer base modification profiles. To mitigate the error rate, Pacific Biosciences (PacBio) offers an enhanced protocol known as Circular Consensus Sequencing (CCS). The DNA templates are circularized, allowing the polymerase to read through them multiple times, thereby significantly improving the precision from around 90% to 99.8%[6].

Due to its advantages such as ultra-long read length, no need for template amplification, shorter run times, and real-time monitoring of epigenetic modification sites [7], its use is gaining broader acceptance, extending its reach beyond the realm of fundamental science to encompass practical domains like agriculture, environmental studies, and medical investigation.

In basic science, SMRT sequencing can study the molecular mechanisms of living cells with new resolution [8]. This makes it possible to construct dynamic genomes and monitor real-time gene synthesis and origin in organisms. At the same time, due to its PacBio circular consensus sequencing (CCS) with high-throughput and high accuracy, it plays an important role in epigenetics, such as DNA methylation modifications [9, 10], as well as in the analysis of RNA isoform sequencing [11].

3.2. Nanopore Sequencing Technology

Oxford Nanopore Technologies (ONT) secured the licensing rights for its foundational nanopore sequencing patents in 2007 and initiated its strand sequencing operations in 2010. And in June 2014, Oxford Nanopore Technologies (ONT) introduced individual MinION devices as part of a comprehensive initiative known as the MinION Access Program (MAP)[12]. Nanopores are categorized into two principal classes: biological and solid-state. Biological nanopores encompass entities like the α-hemolysin (α-HL) pore, Mycobacterium smegmatis porin A (MspA), and the phi29 connector. In contrast, solid-state nanopores are predominantly crafted from silicon-based substrates, including silicon and silicon nitride. The field has also witnessed the emergence of novel materials for nanopore fabrication in recent times, expanding the scope of materials science in this domain[13]. Its core is a nanoscale protein pore, acting as a biosensor, set within an electrically insulating polymer membrane. This system functions by maintaining a steady voltage across the nanopore immersed in an electrolytic solution, propelling the negatively charged single-stranded DNA or RNA molecules through the pore from the negatively charged 'cis' compartment to the positively charged 'trans' compartment[14].

As the nucleic acid molecule traverses the pore in a stepwise fashion, fluctuations in the ionic current reflect the sequence of nucleotide within the detection region. These changes are captured in real-time and decoded by computational algorithms to determine the genetic sequence of the single molecule.

This "sequencing by synthesis" approach allows for the real-time, single-molecule sequencing, distinguishing ONT's technology through its ability to measure the unique electrical signatures produced by each base as DNA molecules pass

through the nanopore. The distinct current signatures are decoded to identify the specific bases, offering a direct and label-free method for genomic analysis[15].

Similarly, nanopore sequencing technology is a disruptive and innovative third generation sequencing technology. Its application range includes DNA and RNA molecules, so it has been widely used in environmental science [16], criminal investigation [17], agriculture [18], basic research [19], medicine [20,21] and food [22]. However, due to its certain error rate, it will not completely replace other sequencing technologies in the next few years, but a comprehensive and systematic understanding and improvement of this technology is still of great significance for future development.

4. Enrichment Improvement based on CRISPR/Cas Family

4.1. Basic Advantage

In traditional sequencing methods, PCR amplification is usually required. However, this process can lead to a very high error rate, especially when amplifying (tandem) repeats, and removes all epigenetic marks. Therefore, the use of amplification methods hinders obtaining a comprehensive genetic and epigenetic profile of the tandem repeats [5]. At the same time, there are certain limitations when the sample DNA has characteristics such as being repetitive, low complexity, or rich in AT/CG. As a result, in recent years, a large number of researchers have developed targeted sequencing technologies, including but not limited to the Xdrop system [23], adaptive sampling [24], and CRISPR/Cas family-based third-generation sequencing technologies, which can perform targeted sequencing on specific DNA sites and have a stronger ability to resolve SVs and other complex regions in the human genome [25]. In CRISPR-mediated third-generation sequencing technology, it not only enables researchers to fully characterize the amplification of homozygotes for the first time, providing a deeper insight into the instability of repeat length and chimera that traditional methods cannot accurately estimate [26]; it can also determine the results of genome editing at any site (including "dark" regions), including the frequency of non-homologous end joining (NHEJ) and homology-directed repair (HDR) [27]. At the same time, compared with STR analysis, the sequence-level data generated by CRISPR-guided SMRT sequencing technology provides higher genotyping accuracy, thanks to its ability to detect all instabilities amplified in a particular allele [27], thereby more accurately characterizing the dynamic nature of the genome. This is of great significance for assessing the effectiveness and safety of gene-editing technologies. The following will introduce the general methods and progress of two types of sequencing mediated by the CRISPR/Cas family.

4.2. CRISPR-mediated SMRT Sequencing

4.2.1. Construction of CRISPR/Cas9 System

The process primarily involves the design and synthesis of guide RNA (gRNA) and its assembly with the Cas nuclease. Prior to the research, it is essential to identify the target gene site and craft the gRNA based on the sequence of the target gene. The gRNAs are created by annealing the CRISPR RNA (crRNA) with the trans-activating crRNA (tracrRNA), and it is crucial to ensure complementarity with the target sequence [28]. We recommend utilizing an online CRISPR RNA configurator accessible on the Dharmacon website to generate the anticipated target sequence and verify the specificity of the target against the human genome reference sequence to guarantee recognition by Cas9 without binding to other proteins [28]. Ultimately, the CRISPR/Cas9 complex is synthesized using chemical methods, and further elaboration on this point is unnecessary.

4.2.2. Preparation of SMRTbell Template

For the target gene locus, an amplification-free targeted enrichment method using Cas9 is utilized. Initially, DNA is fragmented to simplify the genome. The DNA fragments are then connected to site-specific hairpin adapters that bear the used cutting sites, and these are ligated into SMRTbell template libraries by a ligase. Other DNA fragments, incompatible with the SMRTbell adapters, are digested by nucleases. Subsequently, the nucleases are combined with the SMRTbell templates that have been prepared with the specific gRNA digestion. After the reaction, PolyA hairpin capture adapters are ligated to the digested, targeted DNA; this structure allows for effective RNA binding and selective capture and binding [28], resulting in the final asymmetric SMRTbell template molecule library, which is enriched using Pacific Biosciences' Magbead system.

4.2.3. Targeted SMRT Sequencing

To prepare for sequencing, it is essential to first remove excess DNA and various enzymes. To ready the target fragments for SMRT sequencing, a standard PacBio primer that does not contain a PolyA sequence is annealed, followed by an incubation process. If possible, utilize Circular Consensus Sequencing (CCS) reads to enhance the accuracy of sequencing. Ultimately, this process generates a genome coverage map and facilitates mapping.

4.3. CRISPR-mediated Nanopore Sequencing

4.3.1. Cas9-mediated Nanopore Sequencing

4.3.1.1 Simple Mediated Method

First, the subsequent guide CRISPR-RNAs need to be designed, assembled from synthetic crRNA (IDT, custom designed) and tracrRNA (IDT, 1072532) into a duplex. The crRNA can be designed using CRISPick, the IDT tool, and CHOPCHOP. You can further evaluate and select the crRNA with the highest predicted target performance and the lowest off-target activity without common single nucleotide polymorphisms (SNPs) using the IDT CRISPR-Cas9 crRNA checker (https://sg.idtdna. com/site/ order/ designtool/ index/ CRISPR_ SEQUENCE) [29].

Based on the ONT (Oxford Nanopore Technologies) protocol, which can be obtained from the ONT community, one can acquire a sequencing library. Alternatively, a method involving initial dephosphorylation, followed by the addition of the designed Cas9/gRNA complex, Taq polymerase, and then proceeding with A-tailing of DNA ends can also achieve enrichment. After completing double or multiple cutting of the double strands, sequencing is carried out.

4.3.1.2 Data Processing and Analysis

Raw sequencing signals were base-called and transformed into FASTQ format reads, and we can utilize GUPPY for performing the basecalling. Minimap2 can be employed to align the reads with the GRCh38 human reference genome, and then categorize them using SAMtools or the 'bincov' script from the SURVIVOR software package. Alignments can be visualized on the Integrative Genomics Viewer (IGV) [29,30]. Specifically, during diagnostic sequencing for

familial hypercholesterolemia [30], considering the off-target effects of the Cas9 complex's nonspecific binding and the physical shearing forces during the experiment that may damage gDNA molecules outside the region of interest (ROI), a unique method for calculating coverage was devised. In this context, "off-targets" are defined as regions outside the regions of interest (ROIs) that exhibit a spatial deviation exceeding 1000 base pairs (bp) and possess a sequencing depth or coverage greater than 25-fold. To identify these "off-target" candidates, we can use the SURVIVOR tool. Subsequently, to gain insights into the enrichment efficiency of the intended targets, the coverage levels within ROIs, "off-target" sites, as well as across the entire genome, were quantified utilizing the SAMtools software package. Specifically, the on-target ratio, was derived by dividing the number of on-target reads by the number of total reads generated in the experiment [30].

4.3.1.3 Cas9 Tiling Related Methods

When cutting certain target DNA or RNA sequences, there is a risk of losing or depleting the fragments due to their small size after cutting. Research has shown that without the enrichment using the ONT (Oxford Nanopore Technologies) platform, a method known as Cas9 tiling can be used to enrich longer targets, which involves targeted de novo assembly to accurately reconstruct the genome region of interest [31]. Unlike conventional Cas9-mediated nanopore sequencing, the Cas9 tiling method has been modified for sequencing extremely lengthy targets and was utilized for the reconstruction of this specific area in the non-shattering P. vulgaris cultivar Midas. In the research on the common bean's genome [32], the precision of the region of interest (ROI) assembly, which was produced using the Cas9 tiling technique, was validated against the traditional de novo assembly of the entire genome. The Cas9 tiling technique, by assembling a genomic region into a continuous sequence, achieves high consistency compared to traditional long-read WGS methods while improving the accuracy of variant detection.

Utilizing the regions reconstructed by Cas9 tiling, it is possible to more precisely locate read sequences within specific varieties, facilitating the accurate calling of variations, which is suitable for GWAS (Genome-Wide Association Studies) and fine-mapping studies related to given phenotypes. Additionally, integrating these reconstructed sequences into the reference genome can enhance the genotyping capability for short-read sequencing data of inbred line individuals, providing new strategies for genomic research and molecular breeding.

4.3.2. Cas12a-mediated Nanopore Sequencing

As previously mentioned, in current research on CRISPR/Cas9-mediated nanopore sequencing, there are often sequencing errors caused by off-target effects, leading to inaccuracies in the results. However, evidence suggests that Cas12a, previously known as Cpf1, may have enhanced features, including a lower incidence of off-target effects [33-35], although this is still a matter of debate [36]. Moreover, the nucleic acid recognition capabilities of Cas12a have also been utilized for diagnostic applications, notably for the detection of SARS-CoV-2 genomic RNA [37]. Specifically, the realization of the collateral cleavage activity of Cas proteins, including Cas12 and Cas13, has facilitated the targeted detection of specific transcript sequences, yielding observable signals [38]. The Cas12a protein is structured as a bi-lobed monomer, comprising distinct recognition and nuclease domains. When a Cas12a enzyme complexes with a CRISPR RNA (crRNA), it scans DNA for a complementary target sequence. The interaction with the target is facilitated by a flexible pocket that is formed by the convergence of the wedge (WED), the REC1 subdomain, and the PAM-interacting (PI) domain. This pocket specifically binds to a protospacer adjacent motif (PAM) on the DNA, which for Cas12a is typically a TTTV sequence, where V represents A, C, or G[39], so its binding has high specificity and stability. In current research, a well-established method is the solid-state CRISPR-Cas12a-assisted nanopores (SCAN), which has been used for the specific detection of HIV-1 DNA and has optimized the buffer conditions for nanopore sensing [38,40]. However, in its early stages, this technology could not detect target DNA at concentrations less than 10 nM. Therefore, a reverse transcription amplification coupled with the modified nanopore sensing technique method for SCAN devices was developed to enhance sensitivity, thereby allowing the examination of reporter size distributions and their relative abundance [40].

Due to its highly sensitive collateral cleavage activity, the intact circular single-stranded DNA (ssDNA) reporters' cleavage ratio can be quantified using the SCAN method, which serves as a benchmark for distinguishing between positive and negative test outcomes. The Cas12a protein's collateral cleavage activity, where it cuts non-target ssDNA after binding to dsDNA via RNA guidance, has been repurposed for diagnostic applications. In the SCAN method, a specific circular ssDNA is used as a reporter and when the target DNA, such as HIV-1, is present, it activates the Cas12a/crRNA complex, leading to the digestion of the reporter[40]. The degree of reporter degradation reflects the target DNA's concentration. In the absence of the target, the reporter remains undigested. The SCAN technique uses nanopores for electronic quantification of the reporter, allowing for sensitive detection of specific DNA sequences[40]. The specific binding of the Cas12a protein quantifies the nanopore sequencing technology, thereby elevating this technique to a higher level.

4.3.3. Cas13a/b-mediated Nanopore Sequencing

Similarly, for Cas13a (formerly known as C2c2) or Cas13b proteins, they exhibit a "collateral effect" of mixed-up ribonuclease activity upon the recognition of targets. Scientists have integrated this phenomenon with isothermal amplification techniques to devise a system for detecting nucleic acids that leverages these proteins, termed SHERLOCK technology, which stands for Specific High-Sensitivity Enzymatic Reporter UnLOCKing. It has been utilized for various sensitive detections, including identifying specific strains of Zika and Dengue viruses, differentiating harmful bacteria, genotyping human DNA, recognizing mutations in circulating tumor DNA [41]. Moreover, SHERLOCK reagents could be freeze-dried to eliminate the need for cold storage and can be easily restored when needed, making them suitable for use in paper-based formats that are ideal for on-site testing and applications in resource-limited settings [41]. The collateral cleavage significantly enhances the sensitivity of RNA detection [42]. Regrettably, the author is not aware of any research cases combining such Cas13 proteins with third-generation sequencing, hence will not elaborate further on this topic.

5. Conclusion and Prospect

In this article, we have discussed research related to the CRISPR/Cas system, providing a preliminary overview of its principles. We introduced third-generation sequencing technology and explored the enrichment and improvements of CRISPR/Cas system-mediated third-generation sequencing technology. We also briefly introduced methods for data analysis and processing, revealing the immense potential of this cutting-edge technology in genomic research. It can not only provide higher target accuracy but also reduce interference in the target sequence caused by factors such as high CG content or overly complex structural variations through an amplification-free approach, showing a promising prospect for development and application.

Despite the application of the CRISPR/Cas system and its favor for its efficiency and accuracy, challenges and limitations persist. Off-target effects and the difficulty of achieving more than 99% accuracy have always been core issues in the field of gene editing, potentially leading to unexpected mutations at non-target sites, thereby affecting gene function and cell stability, and ultimately causing sequencing results and analysis errors. In addition, improving editing efficiency is also key to further development, especially when dealing with complex or hard-to-edit genomic regions. Faced with these challenges, for the CRISPR/Cas system, one can start by optimizing gRNA design, selecting Cas proteins, and enhancing the specificity of the DSBs repair pathway. Through such improvements, the system can play a central role in gene function research, disease model construction, and precise gene editing and therapeutics.

For sequencing technologies, there are multiple directions for improvement. Firstly, technological innovation can further enhance the throughput and data output of sequencing, such as increasing the number of ZMWs (Zero-Mode Waveguides) in a single sequencing reaction or optimizing sequencing conditions, thereby obtaining more high-quality data in one sequencing run. Secondly, enhancing the read length and quality of sequencing can increase the scope and precision of sequencing. Of course, with these advancements, improving data processing and analysis tools also becomes essential to handle more complex and vast genomes.

At the same time, the development of third-generation sequencing technologies, especially those mediated by the CRISPR/Cas family such as SMRT (Single Molecule Real-Time) and the nanopore sequencing, provides a fresh perspective for genomic research through different Cas proteins and cutting methods. These technologies, characterized by their single-molecule sequencing capability, long read length, and real-time base monitoring, have broken through the limitations of traditional sequencing methods, providing powerful tools for deciphering complex genomic regions (structural variants, SVs), understanding gene expression regulation, and discovering new genetic variations.

References

[1] Man J C, Cheng J, Zhao L. Effect of Knocking Out HOXA5 Gene by CRISPR-Cas9-Mediated Gene Editing Technique on Proliferation of Acute Myeloid Leukemia Cells[J]. Zhongguo shi yan xue ye xue za zhi, 2024, 32(1): 52-56.

[2] Hu Sihui, Liu Qianyi, Xie Dongchun, et al. Clinical research progress of CRISPR/Cas genome editing in the treatment of human genetic disorders[J]. Life Sciences, 2022,34(10):1250-1263. DOI:10.13376/j.cbls/2022139.

[3] Yingying CHEN, Yang LIU, Junjie SHI, Junying MA, Jianhua JU. CRISPR/Cas systems and their applications in gene editing with filamentous fungi[J]. Synthetic Biology Journal, 2024, 5(3): 672-693.

[4] Maruyama T, Dougan SK, Truttmann MC, Bilate AM, Ingram JR, Ploegh HL. Increasing the efficiency of precise genome editing with CRISPR-Cas9 by inhibition of nonhomologous end joining. Nat Biotechnol. 2015 May;33(5):538-42. doi: 10.1038/nbt.3190. Epub 2015 Mar 23. Erratum in: Nat Biotechnol. 2016 Feb;34(2):210. doi: 10.1038/nbt0216-210c. PMID: 25798939; PMCID: PMC4618510.

[5] Ardui S, Ameur A, Vermeesch JR, Hestand MS. Single molecule real-time (SMRT) sequencing comes of age: applications and utilities for medical diagnostics. Nucleic Acids Res. 2018 Mar 16;46(5):2159-2168. doi: 10.1093/ nar/ gky066. PMID: 29401301; PMCID: PMC5861413.

[6] Kraft F, Kurth I. Long-read sequencing to understand genome biology and cell function. Int J Biochem Cell Biol. 2020 Sep;126:105799. doi: 10.1016/j.biocel.2020.105799. Epub 2020 Jul 3. PMID: 32629027.

[7] Liu YH, Wang L, Yu L. The principle and application of the single-molecule real-time sequencing technology. Yi Chuan = Hereditas. 2015 Mar;37(3):259-268. DOI: 10.16288/j.yczz.14-323. PMID: 25787000.

[8] Hestand MS, Ameur A. The Versatility of SMRT Sequencing. Genes (Basel). 2019 Jan 4;10(1):24. doi: 10.3390/genes 1001 0024. PMID: 30621217; PMCID: PMC6357146.

[9] Feng L, Lou J. DNA Methylation Analysis. Methods Mol Biol. 2019;1894:181-227. doi: 10.1007/978-1-4939-8916-4_12. PMID: 30547463.

[10] Searle B, Müller M, Carell T, Kellett A. Third-Generation Sequencing of Epigenetic DNA. Angew Chem Int Ed Engl. 2023 Mar 27;62(14):e202215704. doi: 10.1002/anie. 20221 5704. Epub 2023 Jan 25. PMID: 36524852.

[11] Shi ZX, Chen ZC, Zhong JY, Hu KH, Zheng YF, Chen Y, Xie SQ, Bo XC, Luo F, Tang C, Xiao CL, Liu YZ. High-throughput and high-accuracy single-cell RNA isoform analysis using PacBio circular consensus sequencing. Nat Commun. 2023 May 6;14(1):2631. doi: 10.1038/s41467-023-38324-9. PMID: 37149708; PMCID: PMC10164132.

[12] Deamer, D., Akeson, M. & Branton, D. Three decades of nanopore sequencing. Nat Biotechnol 34, 518–524 (2016). https://doi.org/10.1038/nbt.3423.

[13] Bing-Yuan Guo, Tao Zeng, Hai-Chen Wu, Recent advances of DNA sequencing via nanopore-based technologies, Science Bulletin, Volume 60, Issue 3, 2015, Pages 287-295, ISSN 2095-9273, https://doi.org/10.1007/s11434-014-0707-6.

[14] Wang Y, Zhao Y, Bollas A, Wang Y, Au KF. Nanopore sequencing technology, bioinformatics and applications. Nat Biotechnol. 2021 Nov;39(11):1348-1365. doi: 10.1038/ s41587-021-01108-x. Epub 2021 Nov 8. PMID: 34750572; PMCID: PMC8988251.

[15] Li Z, DU C, Lin Y, et al. [Application of nanopore sequencing in environmental microbiology research]. Sheng wu Gong Cheng xue bao = Chinese Journal of Biotechnology. 2022 Jan;38(1):5-13. DOI: 10.13345/j.cjb.210085. PMID: 35142114.

[16] Zhao X, Liu Y, Chen X, Mi Z, Li W, Wang P, Shan X, Lu X. Detection and Characterization of Single Cisplatin Adducts on DNA by Nanopore Sequencing. ACS Omega. 2021 Jun 22;6 (26):17027-17034. doi: 10.1021/acsomega.1c02106. PMID: 34250360; PMCID: PMC8264939.

[17] Niamh Nic Daeid, Lucina Hackman, Rob Ogden, Nina Vasiljevic, Stefan Prost; Nanopore sequencing in non-human

forensic genetics. Emerg Top Life Sci 24 September 2021; 5 (3): 465–473. doi: https://doi.org/10.1042/ETLS20200287.

[18] Bu Xuan, Xiao Guiqing. Application of single molecule sequencing technology in agriculture and forestry related plants [J]. Biochemistry, 2023,43(06):881-887.DOI:10. 13488/j. smhx. 20230063.

[19] Namkung S, Tran NT, Manokaran S, He R, Su Q, Xie J, Gao G, Tai PWL. Direct ITR-to-ITR Nanopore Sequencing of AAV Vector Genomes. Hum Gene Ther. 2022 Nov;33(21-22):1187-1196. doi: 10.1089/hum.2022.143. PMID: 36178359; PMCID: PMC9700346.

[20] Lewandowski K, Xu Y, Pullan ST, Lumley SF, Foster D, Sanderson N, Vaughan A, Morgan M, Bright N, Kavanagh J, Vipond R, Carroll M, Marriott AC, Gooch KE, Andersson M, Jeffery K, Peto TEA, Crook DW, Walker AS, Matthews PC. Metagenomic Nanopore Sequencing of Influenza Virus Direct from Clinical Respiratory Samples. J Clin Microbiol. 2019 Dec 23;58(1):e00963-19. doi: 10.1128/JCM.00963-19. PMID: 31666364; PMCID: PMC6935926.

[21] Chen J, Xu F. Application of Nanopore Sequencing in the Diagnosis and Treatment of Pulmonary Infections. Mol Diagn Ther. 2023 Nov;27(6):685-701. doi: 10.1007/s40291-023-00669-8. Epub 2023 Aug 11. PMID: 37563539; PMCID: PMC 10590290.

[22] Solcova M, Demnerova K, Purkrtova S. Application of Nanopore Sequencing (MinION) for the Analysis of Bacteriome and Resistome of Bean Sprouts. Microorganisms. 2021 Apr 27;9(5):937. doi: 10.3390/microorganisms9050937. PMID: 33925711; PMCID: PMC8146283.

[23] Madsen EB, Höijer I, Kvist T, Ameur A, Mikkelsen MJ. Xdrop: Targeted sequencing of long DNA molecules from low input samples using droplet sorting. Hum Mutat. 2020 Sep;41 (9): 1671-1679. doi: 10.1002/humu.24063. Epub 2020 Jun 29. PMID: 32516842; PMCID: PMC7496172.

[24] Miller DE, Sulovari A, Wang T, Loucks H, Hoekzema K, Munson KM, Lewis AP, Fuerte EPA, Paschal CR, Walsh T, Thies J, Bennett JT, Glass I, Dipple KM, Patterson K, Bonkowski ES, Nelson Z, Squire A, Sikes M, Beckman E, Bennett RL, Earl D, Lee W, Allikmets R, Perlman SJ, Chow P, Hing AV, Wenger TL, Adam MP, Sun A, Lam C, Chang I, Zou X, Austin SL, Huggins E, Safi A, Iyengar AK, Reddy TE, Majoros WH, Allen AS, Crawford GE, Kishnani PS; University of Washington Center for Mendelian Genomics; King MC, Cherry T, Chong JX, Bamshad MJ, Nickerson DA, Mefford HC, Doherty D, Eichler EE. Targeted long-read sequencing identifies missing disease-causing variation. Am J Hum Genet. 2021 Aug 5;108(8):1436-1449. doi: 10.1016/ j.ajhg. 2021.06.006. Epub 2021 Jul 2. PMID: 34216551; PM CID: PMC8387463.

[25] Phan, M.; Gomes, M.A.; Stinnett, V.; Morsberger, L.; Hoppman, N.L.; Pearce, K.E.; Smith, K.; Phan, B.; Jiang, L.; Zou, Y.S. An Integrated Approach Including CRISPR/Cas9-Mediated Nanopore Sequencing, Mate Pair Sequencing, and Cytogenomic Methods to Characterize Complex Structural Rearrangements in Acute Myeloid Leukemia. Biomedicines 2024, 12, 598. https://doi.org/10.3390/biomedicines12030598.

[26] Hafford-Tear NJ, Tsai YC, Sadan AN, Sanchez-Pintado B, Zarouchlioti C, Maher GJ, Liskova P, Tuft SJ, Hardcastle AJ, Clark TA, Davidson AE. CRISPR/Cas9-targeted enrichment and long-read sequencing of the Fuchs endothelial corneal dystrophy-associated TCF4 triplet repeat. Genet Med. 2019 Sep;21(9):2092-2102. doi: 10.1038/s41436-019-0453-x. Epub 2019 Feb 8. PMID: 30733599; PMCID: PMC6752322.

[27] Höijer I, Johansson J, Gudmundsson S, Chin CS, Bunikis I, Häggqvist S, Emmanouilidou A, Wilbe M, den Hoed M, Bondeson ML, Feuk L, Gyllensten U, Ameur A. Amplification-free long-read sequencing reveals unforeseen CRISPR-Cas9 off-target activity. Genome Biol. 2020 Dec 1;21(1):290. doi: 10.1186/s13059-020-02206-w. PMID: 33261 648; PMCID: PMC7706270.

[28] Yang Liu, Yousry A. El-Kassaby, Landscape of Fluid Sets of Hairpin-Derived 21-/24-nt-Long Small RNAs at Seed Set Uncovers Special Epigenetic Features in Picea glauca, Genome Biology and Evolution, Volume 9, Issue 1, January 2017, Pages 82–92, https://doi.org/ 10.1093/ gbe/ evw283.

[29] Gilpatrick T, Lee I, Graham JE, Raimondeau E, Bowen R, Heron A, Downs B, Sukumar S, Sedlazeck FJ, Timp W. Targeted nanopore sequencing with Cas9-guided adapter ligation. Nat Biotechnol. 2020 Apr;38(4):433-438. doi: 10.1038/ s41587-020-0407-5. Epub 2020 Feb 10. PMID: 32042167; PMCID: PMC7145730.

[30] Xu S, Shiomi H, Yamashita Y, Koyama S, Horie T, Baba O, Kimura M, Nakashima Y, Sowa N, Hasegawa K, Suzuki A, Suzuki Y, Kimura T, Ono K. CRISPR-Cas9-guided amplification-free genomic diagnosis for familial hypercholesterolemia using nanopore sequencing. PLoS One. 2024 Mar 20;19(3):e0297231. doi: 10.1371/ journal. pone. 0297231. PMID: 38507394; PMCID: PMC10954175.

[31] Geng K, Merino LG, Wedemann L, Martens A, Sobota M, Sanchez YP, Søndergaard JN, White RJ, Kutter C. Target-enriched nanopore sequencing and de novo assembly reveals co-occurrences of complex on-target genomic rearrangements induced by CRISPR-Cas9 in human cells. Genome Res. 2022 Oct;32 (10):1876-1891. doi: 10.1101/gr.276901.122. Epub 2022 Sep 30. PMID: 36180232; PMCID: PMC9712622.

[32] Lopatriello G, Maestri S, Alfano M, Papa R, Di Vittori V, De Antoni L, Bellucci E, Pieri A, Bitocchi E, Delledonne M, Rossato M. CRISPR/Cas9-Mediated Enrichment Coupled to Nanopore Sequencing Provides a Valuable Tool for the Precise Reconstruction of Large Genomic Target Regions. Int J Mol Sci. 2023 Jan 5;24(2):1076. doi: 10.3390/ijms24021076. PMID: 36674592; PMCID: PMC9863143.

[33] Kim D, Kim J, Hur JK, Been KW, Yoon SH, Kim JS. Genome-wide analysis reveals specificities of Cpf1 endonucleases in human cells. Nat Biotechnol. 2016 Aug;34(8):863-8. doi: 10.1038/ nbt.3609. Epub 2016 Jun 6. Erratum in: Nat Biotechnol. 2016 Aug 9;34(8):888. doi: 10.1038/nbt0816-888a. PMID: 27272384.

[34] Kim Y, Cheong SA, Lee JG, Lee SW, Lee MS, Baek IJ, Sung YH. Generation of knockout mice by Cpf1-mediated gene targeting. Nat Biotechnol. 2016 Aug;34(8):808-10. doi: 10.1038/nbt.3614. Epub 2016 Jun 6. PMID: 27272387.

[35] Kleinstiver BP, Tsai SQ, Prew MS, Nguyen NT, Welch MM, Lopez JM, McCaw ZR, Aryee MJ, Joung JK. Genome-wide specificities of CRISPR-Cas Cpf1 nucleases in human cells. Nat Biotechnol. 2016 Aug;34(8):869-74. doi: 10.1038/nbt. 3620. Epub 2016 Jun 27. PMID: 27347757; PMCID: PMC 498 0201.

[36] Murugan K, Seetharam AS, Severin AJ, Sashital DG. CRISPR-Cas12a has widespread off-target and dsDNA-nicking effects. J Biol Chem. 2020 Apr 24;295(17):5538-5553. doi: 10.1074/ jbc. RA120.012933. Epub 2020 Mar 11. PMID: 32161115; PMCID: PMC7186167.

[37] Broughton JP, Deng X, Yu G, Fasching CL, Servellita V, Singh J, Miao X, Streithorst JA, Granados A, Sotomayor-Gonzalez A, Zorn K, Gopez A, Hsu E, Gu W, Miller S, Pan CY, Guevara H, Wadford DA, Chen JS, Chiu CY. CRISPR-Cas12-based detection of SARS-CoV-2. Nat Biotechnol. 2020 Jul;38(7): 870-874. doi: 10.1038/s41587-020-0513-4. Epub 2020 Apr 16. PMID: 32300245; PMCID: PMC9107629.

[38] Nouri R, Jiang Y, Lian XL, Guan W. Sequence-Specific Recognition of HIV-1 DNA with Solid-State CRISPR-Cas12a-Assisted Nanopores (SCAN). ACS Sens. 2020 May 22;5 (5): 1273-1280. doi: 10.1021/acssensors.0c00497. Epub 2020 May 8. PMID: 32370494.

[39] Naqvi, M.M., Lee, L., Montaguth, O.E.T. et al. CRISPR–Cas12a-mediated DNA clamping triggers target-strand cleavage. Nat Chem Biol 18, 1014–1022 (2022). https://doi.org/ 10.1038/s41589-022-01082-8.

[40] Nouri R, Jiang Y, Tang Z, Lian XL, Guan W. Detection of SARS-CoV-2 with Solid-State CRISPR-Cas12a-Assisted Nanopores. Nano Lett. 2021 Oct 13;21(19):8393-8400. doi: 10.1021/ acs.nanolett.1c02974. Epub 2021 Sep 20. PMID: 34542296.

[41] Gootenberg JS, Abudayyeh OO, Lee JW, Essletzbichler P, Dy AJ, Joung J, Verdine V, Donghia N, Daringer NM, Freije CA, Myhrvold C, Bhattacharyya RP, Livny J, Regev A, Koonin EV, Hung DT, Sabeti PC, Collins JJ, Zhang F. Nucleic acid detection with CRISPR-Cas13a/C2c2. Science. 2017 Apr 28;356(6336):438-442. doi: 10.1126/science. aam9321. Epub 2017 Apr 13. PMID: 28408723; PMCID: PMC5526198.

[42] Sanger F, Nicklen S, Coulson AR. DNA sequencing with chain-terminating inhibitors. Proc Natl Acad Sci U S A. 1977 Dec; 74(12):5463-7. doi: 10.1073/pnas.74.12.5463. PMID: 271968; PMCID: PMC431765.