Nanopore Detection Assisted DNA Information Processing

Zichen Song 1,†, Yuan Liang 2,† and Jing Yang 1,*

1 School of Control and Computer Engineering, North China Electric Power University, Beijing 102206, China
2 Department of Computer Science and Technology, School of Electronics Engineering and Computer Science, Peking University, Beijing 100871, China
* Correspondence: yjzcd_2000@ncepu.edu.cn
† These authors contributed equally to this work.

Abstract: The deoxyribonucleotide (DNA) molecule is a stable carrier for large amounts of genetic information and provides an ideal storage medium for next-generation information processing technologies. Technologies that process DNA information, representing a cross-disciplinary integration of biology and computer techniques, have become attractive substitutes for technologies that process electronic information alone. The detailed applications of DNA technologies can be divided into three components: storage, computing, and self-assembly. The quality of DNA information processing relies on the accuracy of DNA reading. Nanopore detection allows researchers to accurately sequence nucleotides and is thus widely used to read DNA. In this paper, we introduce the principles and development history of nanopore detection and conduct a systematic review of recent developments and specific applications in DNA information processing involving nanopore detection and nanopore-based storage. We also discuss the potential of artificial intelligence in nanopore detection and DNA information processing. This work not only provides new avenues for future nanopore detection development, but also offers a foundation for the construction of more advanced DNA information processing technologies.

Keywords: nanopore detection; DNA storage; ONT nanopores; artificial intelligence; DNA information processing

1. Introduction

A storage system for electronic information is a fundamental component of modern information technology. However, with the advent of the era of big data, the storage capacity typically becomes inadequate for the system requirements. The surge in electronic information [1] has led to the development of long-term storage systems for big data that are based on magnets and semiconductors [2]. These systems come with unsustainable disadvantages, including limited lifespans [3], high infrastructure costs, and huge power consumption [4]. As a biological storage carrier of genetic information, DNA is an ideal medium for a next-generation storage system, offering the following advantages: ultra-high information density (455 EB data per gram [5]); long lifespan (half-life > 500 years [6,7]); relatively low energy consumption; programmability; and addressability [8]. Therefore, technologies that process DNA information have benefited from the cross-disciplinary integration between biology and information technology [9]. The three basic components of such technologies—DNA computation, DNA storage, and DNA self-assembly—provide a novel approach for the processing of large amounts of information.

The decoding of molecular information is essential to DNA information processing, where DNA storage is a typical example. In DNA storage, digital information can be encoded into DNA sequences according to specific algorithms, then stored in DNA strands by nucleotide synthesis technology and read out using DNA sequencing methods. The quality of DNA storage depends on DNA synthesis and sequencing [10,11]. Large-scale DNA can be synthesized by rapid low-cost solid-phase synthesis [12,13]. Traditional
methods of DNA sequencing, such as Sanger sequencing [14] and Illumina sequencing, are costly and can result in the failure of large-scale DNA storage. Therefore, a highly efficient method for DNA sequencing is necessary to support the expansion of commercial applications of DNA information processing.

Nanopore detection is a single-molecule detection technique, and is useful for single-molecule chemistry studies [15], peptide and protein folding investigation [16,17], analysis of the mechanical aspects related to unzipping of nucleic acids or nucleic acids–protein complexes [18–20]. In the field of DNA information processing, nanopore detection is an efficient detection technique with two main applications: DNA nanopore sequencing and single-molecule sensing [21]. Nanopore sequencing technology with single nucleotide resolution is used to read information from DNA strands, thus assisting in error-free data recovery of large-scale DNA storage systems [22,23]. Additionally, nanopores can be assembled that act as single-molecule sensors for molecular identification. By modifying biological nanopores, the specificity and sensitivity of nanopores can be improved, thus expanding the detection range of molecules. Currently, a wide variety of molecules can be detected by nanopores, including DNA [24–27], DNA and RNA with lesions and nucleotide modification [28–33], whole cell nucleic acid preparation [34–38], unfolded protein peptides [39–46], and bacterial toxins [47]. Because of the efficient molecular identification capabilities of nanopores, DNA databases have more options for information storage. Take DNA modification as a typical example. DNA modifications are biochemically processed to bind DNA strands to other molecules. For the same DNA modification, the presence or absence of a modified nucleotide can be considered as a digital bit “1” and “0” along the oligonucleotide strand. Therefore, DNA modification sequences can be logically regarded as binary sequences. On the other hand, nanopores can precisely identify and distinguish multiple DNA modifications, which expands the alphabet of DNA storage and directly enhances potential storage density of DNA storage. At present, nanopore detection technique has been widely used in DNA storage systems.

This review was intended to cover recent advancements in nanopore detection technology and its application in DNA information processing. Therefore, we conducted a systematic investigation of the following: (1) principles and development history of nanopore detection technology; (2) developments in nanopore detection; (3) the two types of DNA storage that use nanopore detection; and (4) applications of artificial intelligence (AI) in nanopore data processing and DNA information technology. We envision this review article to advance the development of nanopore detection in the field of DNA information processing.

2. Principles and History of Single-Molecule Detection

Nanopore detection is a single-molecule detection technique that originated from patch clamp technology in 1976 [48]. Single-molecule detection features real-time monitoring and can be used to obtain the function, structure, dynamics, and other information of molecules [49]. According to the principles of detection, there are three methodological categories of single-molecule technology: optical, mechanical, and electrochemistry. Optical methods add fluorescent biological labels to molecules and observe the excitation or quenching of fluorescence through microscopy [50]. However, these methods change the structure of molecules, which can lead to inaccurate observations. Mechanical methods use atomic force microscopy [51] or scanning tunneling microscopy [52] to photograph biological molecular structures, and can provide direct images of biological molecules without fluorescent labels. Unfortunately, mechanical methods have several disadvantages that make them unsuitable for large-scale molecular detection: namely, bulky observation equipment, high cost and complex operations. Electrochemical-based single-molecule detection techniques generally rely on sensitive current monitors to detect molecules by analyzing ionic current signals [53]; the nanopore detection technique is one of its representatives.
2.1. General Principles of Biological Nanopore Detection Technology

In recent years, nanopore detection, whose advantageous features include no fluorescent-label requirement, portable equipment (Minion is only 90 g [54]) and low detection cost [55], has attracted much attention from researchers. The working principle of the experiment instrument for nanopore detection is similar to that of a Coulter counter [50], a single charged molecule is driven through the tiny pore embedded in a membrane by an electric field force, generating a transient current blocking signal. The general process of biological nanopore detection [56,57] begins by embedding a biological nanopore in an electrically resistant polymer membrane. A tank filled with an electrolyte solution is separated into two chambers by the membrane, and the nanopore becomes the only channel between the two chambers. Two electrodes are added to the chambers, and a constant voltage is applied to the electrodes; the negative voltage side is defined as “cis”, and the opposite side is defined as “trans”. During the detection process, the molecule of interest is driven along the nanopore under the influence of the applied electric field or/and the ensuing electroosmotic flow manifested inside ion-selective nanopores [58]. Charged biomolecules may be driven through the nanopores from “cis” to “trans” or in the opposite direction. Because of the nanoscale size internal diameter of the nanopore, biomolecules passing through the nanopore will hinder the flow of ionic current, producing discrete signal changes. Analysis of the signal characteristics of the ionic current through the use of a computer algorithm yields a variety of direct information about the target biomolecule, such as its species, structure and function.

2.2. Development of Nanopore Detection Technology

In the development process of nanopore detection, sequencing of DNA or RNA was initially focused on achieving high-quality base calling. The concept of DNA nanopore sequencing specifically originated in the late 1980s. In 1989, Hagan Bayley’s team began exploring the structure and function of oligomeric transmembrane protein pores such as α-hemolysin [59], and they later hypothesized that channel proteins could act as biosensors for molecules [60]. In 1996, Kasianowicz et al. was the first group to capture ionic current signals generated from single-stranded DNA (ssDNA) translocation through α-hemolysin nanopores [61], and, in 1999, Akeson et al. utilized α-hemolysin as a biosensor to achieve rapid discrimination between pyrimidine and purine segments along an RNA molecule [62]. More than a decade later, Gundlach et al. used Mycobacterium smegmatis porin A (MspA) [63] as a biological nanopore in combination with bacteriophage phi29 DNA polymerase to increase the number of identifiable DNA bases to ~30 [64]. Subsequently, in 2014, Oxford Nanopore Technologies (ONT) launched the first commercial nanopore sequencer, MinION, which has the advantages of single-molecule detection, long read length, fast sequencing speed, and portability [54].

2.3. Categories of Nanopores

Currently, there are two categories of nanopores based on raw material composition: protein nanopores and solid nanopores (made from solid materials). Due to their relatively small internal diameters, biological nanopores have a high signal-to-noise ratio and resolution. Moreover, with the advent of site-directed protein modification, biological nanopores can be used to detect a wide range of molecules, including DNA [65], RNA [66], protein, and metallic ions [67]. Biological nanopores include α-hemolysin (α-HL), MspA, Escherichia coli cytolysin (ClyA) [68], and aerolysin [69]. However, the structural stability of protein nanopores is easily affected by environmental conditions.

Compared to the biological nanopores, solid-state nanopores possess the advantages of excellent geometry flexibility, mechanical and chemical stability as well as compatible properties with modern semiconductor and microfluidics fabrication techniques [70–75]. Materials used to prepare solid-state nanopores include inorganic silicon [76], glass capillaries [77], and graphene [78]. Among these, silicon nitride is the most widely used material because of its low mechanical stress and excellent chemical stability. Although the excellent
properties of solid-state nanopores hold great potential in biomolecular detection at single molecule level, they are still plagued by a number of particular drawbacks including high cost [79], poor reproducibility among different nanopores, large inner diameter, and lack of atomic-resolution functionalization [80].

3. Nanopores Detect Specific Modifications Carried by DNA Molecules

Nanopores can be used to simultaneously detect information about multiple changes or damage to single/double-stranded DNA that occurs during protein binding or modification by inorganic chemicals [81,82].

3.1. Detection of DNA Lesions

The accumulation of unrepaired DNA lesions may lead to premature cellular senescence, cancer, and some neurodegenerative diseases [83,84]. Currently, nanopores are being used as biosensors for DNA lesion detection. In 2019, Ma et al. proposed a nanopore detection method to identify cisplatin lesions on DNA [85] (Figure 1a). Cisplatin binds to $N^7$ atoms on purines to form lesions on DNA, inhibiting the normal replication and transcription of DNA in cancer cells. Ma et al. applied MspA for the accurate detection of cisplatin-induced DNA lesions, demonstrating that the nanopore sequencing technique could identify cisplatin lesions in an input sequencing library of less than 10 ng. Furthermore, discrimination of multiple DNA lesions was achieved by observing the speed of DNA translocation through the nanopores. In 2022, Zhang et al. achieved direct identification of O6-carboxymethylguanine (O6-CMG), O6-methylguanine (O6-MG), and base-free (AP) sites through observation of the kinetics of enzymatic deceleration [86] (Figure 1b). They observed that progressive movement of phi29 DNA polymerase was hindered by DNA lesions such as O6-CMG, and recorded enzymatic arrest, suggesting that kinetic information generated by the interaction between a motor enzyme and DNA lesions can be used to identify multiple DNA lesions.

3.2. Detection of Nucleic Acids

The current nanopore sequencing techniques enable the detection of heterogenic nucleic acid. Heterogenic nucleic acids, also known as xeno-nucleic acids (XNAs), are a class of nucleic acid molecules with a non-natural backbone or nucleobase [87–89]. In 2019, Yan et al. presented a method for direct sequencing of 2-deoxy-2-fluoroarabinoic acid (FANA) through nanopore-induced phase shift sequencing (NIPSS) [90] (Figure 1c). By ligating FANA with a DNA drive-strand, they showed that the direct sequencing of FANA could be achieved by NIPSS with phi29 DNA polymerase. Their contribution led to the development of a universal identification method based on nanopore sequencing that can be used to clearly distinguish between DNA, RNA and XNA nucleotides.

3.3. Detection of Peptides and Proteins

Protein detection can also be achieved using nanopore sequencing techniques [20,91,92]. In 2021, Wang et al. observed single molecule ratcheting of peptides with MspA [93] (Figure 1d). By constructing peptide–oligonucleotide conjugates (POCs) and using NIPSS for measurements, they directly observed the discrete steps of the ratcheting motion of a peptide. Experimentally, the current signal results of peptides generated from NIPSS measurements show a highly consistent pattern, with a clear correlation to the amino acid sequence of the peptide. Subsequently, in November 2021, Brinkerhoff et al. demonstrated a nanopore-based peptide sequencing method with single-amino acid resolution [94]. They designed an experiment in which a POC was driven through an MspA nanopore by a DNA helicase, and they were able to observe steplike current signals generated by a peptide ratcheting through the nanopore. Additionally, they changed amino acids at fixed sites on the peptide sequence and observed the resulting changes in the ionic current sequence, enabling protein sequencing.
3.4. Detection of Inorganic Chemical Molecules

In addition to detecting biomolecules, nanopores can be used to detect inorganic chemical molecules. In 2021, Jia et al. proposed a programmable nanoreactor for random stochastic sensing (PNRSS) based on a nanopore sequencing technique, enabling real-time monitoring of chemical reactions at the single-molecule level \[95\] (Figure 1e). A wide range of single-molecule reactions of metal ions, simple organic compounds such as lactic acid, and nucleoside analogues have been directly observed through PNRSS. Moreover, they used AI tools to enhance the sensing resolution of PNRSS, which ultimately allowed them to detect a total of 20 types of chemical reactions.

![Figure 1](image-url)

**Figure 1.** The schematic diagrams of nanopore-based molecular detection systems. (a) Detection of cisplatin lesions on DNA. Reproduced with permission from Fubo Ma, ACS Sensors; published by American Chemical Society, 2021. (b) Stalling kinetics readout during nanopore sequencing using an MspA nanopore (blue) and a phi29 DNAP (yellow). Reproduced with permission from Jinyue Zhang, Nano Letters; published by American Chemical Society, 2022. (c) Sequencing of chimeric DNA (grey) -FANA (cyan) with an abasic spacer (red). Reprinted from \[90\]. (d) Ratcheting motion of a POC using NIPSS. Reproduced with permission from Shuanghong Yan, Nano Letters; American Chemical Society, 2021. (e) Detection of inorganic chemical molecules through PNRSS. Reprinted from \[95\].

4. DNA Storage Based on Nanopore Sequencing Technology

Although the nucleotide strands carrying digital information are fragile, which may be interfered with and destroyed by external environmental factors during the preservation, such as ultraviolet rays, extreme temperature changes, and biological contamination such as bacteria and viruses. However, with properties of ultra-high information density and long lifespan, DNA is expected to become a novel data storage medium for the next generation of information processing systems. We have reason to believe that with the advancement of DNA preservation techniques and preservation equipment \[96–99\], the capacity, duration and storage quality of DNA storage will be greatly improved. Currently, DNA nanopores can be used to identify sequences of DNA/RNA and molecular or chemical modifications of DNA/RNA, all of which can be considered as bit sites, thereby providing more options for DNA storage systems. There are two categories of DNA as a carrier for digital data storage: (1) synthetic DNA base sequences; and (2) DNA nanostructures/modifications.

4.1. DNA Storage Based on Synthetic DNA Sequences

As illustrated in Figure 2a, the entire procedure of DNA storage is commonly divided into six steps \[100\] (Figure 2a): (1) encoding digital information into DNA sequences; (2) designing and synthesizing DNA sequences; (3) preserving the DNA in vivo or in vitro;
(4) random access of specific DNA sequences; (5) reading of the specific DNA sequences; and (6) decoding and recovering DNA sequences into digital information. At present, nanopore sequencing technique is widely used in steps 4 and 5.

Figure 2. The workflows of DNA storage systems based on synthetic nucleotide sequences. (a) The major steps of digital data storage in DNA corresponding to conventional hard-disk storage. Reproduced with permission from Yaya Hao, SMALL STRUCTURES; published by John Wiley and Sons, 2020. (b) An overview of a DNA data storage workflow using ONT nanopores as tools to sequence long double-stranded DNA strands obtained by random access. Reprinted from [101]. (c) The workflow of an in vivo DNA storage system using an artificial yeast chromosome. Reprinted from [22].

Nanopore sequencing technology is advantageous for the data decoding of DNA storage systems. In 2018, Lee et al. designed and validated a large primer library using over 13 million oligonucleotides stored in 35 files and totaling 200 MB of data, and they were able to achieve error-free data recovery of all DNA files using random access methods and nanopore sequencing [23]. In 2019, Lopez et al. demonstrated a method for DNA storage that combines random access, DNA assembly, and nanopore sequencing [101] (Figure 2b). They employed the MinION sequencer to successfully recover digital information stored in 111,499 oligonucleotides and totaling 1.67 terabytes of data. This method allows for an approximately 100-fold increase in sequencing and decoding capacity compared with previous reports using nanopores in a DNA storage system.

Nanopore sequencing technology can be applied to both in vivo and in vitro storage systems. In 2021, Chen et al. designed and synthesized an in vivo system using an artificial yeast chromosome of 254,886 bp [22] (Figure 2c). The chromosome was written into 37,782 bits of data using sparse low-density parity-check (LDPC) codes and pseudorandom sequences, comprising a total of two images and a video clip. During the DNA information reading stage, they used a nanopore sequencer to achieve accurate base calling, achieving reliable data recovery at an error rate of 10.79%.

4.2. DNA Storage Using DNA Nanostructures and Modifications as Information Carriers

Nanopore sequencing techniques that provide accurate identification on DNA modifications or nanostructures offer new solutions for DNA storage. Highly programmable DNA nanostructures offer a variety of address sites for storage of digital data [102,103]. In 2018, Chen and Kong et al. proposed a DNA storage scheme that considered DNA hairpins as bit sites [104] (Figure 3a). In their work, DNA hairpins of different lengths were regarded as digital bits and used to develop a high-resolution solid-state nanopore
sequencing method. DNA hairpins of 8 bp and 16 bp in length could be clearly distinguished using quartz nanopores with an internal diameter of ~5 nm. Thus the 8-bp and 16-bp hairpins were assigned bit-0 and bit-1, respectively, and used to attach 56 hairpins to a 7228-bp long oligonucleotide, thereby forming a 56-bit storage segment. Using a similar idea, Bell and Keyser [105] designed a DNA nanostructure library based on the principles of DNA origami in which each member has a unique barcode, and each bit on a barcode is represented by the presence or absence of a DNA dumbbell hairpin. They eventually confirmed that a 3-bit barcode could be recognized with 94% accuracy through solid-state nanopore sequencing.

An alternative DNA storage system of nanopores is being used for identification of biopolymer sequences. In 2020, Cao et al. used specially tailored biopolymer sequences as bit information storage carriers [106] (Figure 3b). The biopolymer sequences are biohybrid macromolecules comprising two different-sized monomers (n-propyl phosphate and [2,2-diynyl]-propyl phosphate) and natural nucleotides, where the monomers are mapped as bit-0 and bit-1, respectively. The study used bioengineered nanopores of the aerolysin toxin to successfully achieve bit-site recognition of the customized biopolymers with single-base resolution. Additionally, deep learning was applied to achieve high-precision encoding and decoding of up to 4-bit digital sequences. This unique system provides inspiration for the development of new DNA storage systems.

5. Applications of AI in Nanopore Data Processing and DNA Information Technology

The integration of AI with DNA information processing has led to unexpectedly good outcomes in recent years. Deep learning is a branch of AI that mainly uses multi-layer
neural networks to learn from data. Deep learning models can be used to automatically learn from and extract features of raw data, and they have powerful capabilities for mining the potential rules of big data. Models of deep learning such as convolutional neural networks (CNNs), deep confidence networks, and recurrent neural networks (RNNs) have been developed that perform well on multiple tasks, including computer vision, speech recognition, and natural language processing [107–109]. According to previous reports [110,111], the huge amount of sequence data generated from nanopore detection has led to the design and training of multiple deep learning models that are being widely used in DNA information processing tasks, such as base calling, biomolecular detection, and DNA storage.

5.1. Base Calling

Base calling is the process of inferring the order of nucleotides in a DNA segment during sequencing [112]. Because nanopore sequencing generates current signals, base calling requires computer algorithms to process the sequence data. To date, many researchers, including the team at ONT, have designed a variety of software programs based on deep learning models for base calling. These software programs can be categorized by the two types of input data: segmentation events and raw current signals.

Early base calling software relies on the analysis of segmented events. In 2016, David et al.'s research and development team developed Nanocall, the first open-source, offline basecaller for Oxford nanopore sequencing data [113]. It uses a hidden Markov-based model (HHM) for base sequence identification. Using the R7.3 version of the MiniION, Nanocall analyzed data from two *Escherichia coli* and two human genetic samples and found reads with 68% identity. Because HHM is not suitable for long homogeneous polymer detection, RNNs are applied to segmented event sequences. In 2016, Boza et al. proposed DeepNano, an open-source DNA basecaller with deep RNNs [114]. Using R7.3 test datasets for *E. coli* and *Klebsiella pneumoniae* to evaluate the basecaller accuracy of DeepNano, they found that, for 2D reads, DeepNano achieved the base recognition accuracies of 88.5% and 86.7%, respectively. In terms of the speed of base calling, DeepNano is 5 to 20 times faster than Nanocall.

The direct conversion of raw current signals into base sequences by deep learning models is convenient and accurate. BasecRAWller, a base-calling software for nanopore data based on raw current signals, was proposed by Stoiber et al. in 2017 [115]. BasecRAWller uses a pair of unidirectional RNNs to make real-time DNA base calls directly from raw nanopore reads and has been evaluated on the basis of its performance with two data sets: *E. coli* and human. Reportedly, BasecRAWller reads have 81.7% and 72.5% identities on the *E. coli* and human datasets. Soon thereafter, Teng et al. reported on Chiron, the first deep learning model to implement end-to-end base calls and convert raw current signals directly into nucleotide sequences [116]. Chiron combines a CNN with an RNN and a connectionist temporal classification decoder, which allows it to learn directly from the raw signal data without using event segmentation. Chiron achieves 90.57% and 81.54% identities on *E. coli* and human datasets, respectively, which are higher than those achieved by the other three software programs mentioned above. In terms of speed on a central processing unit processor, Chiron is slower than BasecRAWller (21 bp/s vs. 81 bp/sec, respectively). Moreover, Chiron is fully open source, allowing users to train their own neural networks and develop specialized base-calling applications.

5.2. Biomolecule Detection

Deep learning is a powerful tool for improving the nanopore identification accuracy of multiplex biomolecules, which can be effectively used in the detection of biomolecular modifications. In 2019, Liu et al. designed and trained DeepMod [117] (Figure 4a), a bidirectional RNN model with short long-term memory (LSTM) that is suited for high-precision DNA modification detection from raw current signal extracted by the ONT nanopore sequencer. DeepMod was evaluated in the nanopore readouts of the *E. coli*,
Chlamydomonas reinhardtii, and human genomes. The results showed that DeepMod can detect methylation-modified DNA nucleotides with high accuracy; for example, 5-methylcytosine (5 mC) and 6-methyladenine (6 mA) exhibited average detection accuracies of 0.99 and 0.9, respectively. Similarly, in 2021, Ni et al. designed a deep learning tool called DeepSignal-plant that delivers genome-wide detection of all three sequence contexts of cytosine methylations that are naturally occurring in plants using nanopore reads [118] (Figure 4b). With an architecture of multilayer bidirectional RNN followed by a full connection layer, DeepSignal-plant can automatically extract and learn both sequence features and signal features from nanopore data, and is one-eighth the size of its predecessor, DeepSignal [119].

Figure 4. The applications of AI in DNA information storage. (a) Architecture of the neural network model DeepMod, which is used to capture the time-series characteristics of nanopore signals and detect DNA modifications. Reprinted from [117]. (b) Architecture of DeepSignal-plant, in which sequence and signal features can be extracted using bidirectional LSTM (biLSTM). Reprinted from [118]. (c) Evolutionary model workflow for hybrid prediction DNA coding. Reprinted from [120]. (d) Schematic diagram of CNNs used to quickly read molecular labels. Reprinted from [121].
5.3. DNA Storage

AI can be integrated into multiple steps of DNA storage, potentially accelerating the reading speed of oligonucleotides, providing efficient and accurate random-access methods, and promoting the further realization of commercialized large-scale DNA storage systems.

Deep learning holds great potential in the rapid analysis of nanopore reads. In 2020, Nivala et al. proposed a method for tagging physical objects using DNA or other molecules in situations where traditional methods such as radiofrequency identification tags and quick response codes do not apply [120] (Figure 4c). They developed the Porcupine system, an end-user molecular tagging system with the ability to read DNA-based tags in seconds using a portable nanopore device. Its digital bits are represented by the presence or absence of different DNA strands, called molecular bits (molbits), which are classified by a CNN directly from the raw nanopore signal. This method avoids the need for base calling using DNA sequences and thus greatly reduces the time requirement and complexity.

In summary, deep learning can be applied to rapidly increase the base recognition accuracy of DNA sequencing (identities increased from 68% to 90.57%), strongly expand the range of molecular types that can be detected by nanopores, and provide the possibility of high-speed DNA reading, all of which are crucial factors for the large-scale commercialization of DNA storage and nanopore technologies.

Deep learning can also be used to scale up the capacity of DNA storage systems with computer simulation, which provides helpful suggestions to guide research. Having access to a large-scale DNA storage system is equivalent to being able to design complex DNA primer sequences, which are unaffordable in most research settings. Therefore, precise control and prediction of the DNA hybridization process are critical for the design of large-scale DNA storage systems. In 2021, David Buterez was the first to present a comprehensive study of a machine learning technique for DNA hybridization prediction [121] (Figure 4d). As a baseline, he conducted performance evaluations of multiple machine learning models on an in silico-generated hybridization dataset containing more than 2.5 million DNA sequence pairs. Next, he evaluated this dataset using CNN, RNN and RoBERTa models and found that the deep learning models delivered more accurate DNA hybridization prediction and reduced the running time by one to two orders of magnitude compared with the baseline models.

6. Conclusions

DNA information processing technologies utilize the DNA molecule as a data storage medium and data calculation unit and have the potential to store big data. However, the development of these technologies has been hampered by the high cost, low speed, and relatively low accuracy of traditional DNA sequencing. Nanopore detection is a new single-molecule detection technique that enables molecular identification through the analysis of ionic current signals generated by molecules passing through nanopores. Compared with other methods, rapid nanopore detection offers the advantages of being label-free, low-cost, and convenient, and satisfies the requirements of DNA information processing. At present, nanopore detection is widely used in DNA information processing tasks, such as biomolecular detection and DNA storage. However, the development of nanopore detection is encountering obstacles that may hinder further improvements in DNA information processing.

Nanopores are frequently employed in biomolecular detection, due to improvements in specificity and sensitivity, and are useful in the detection of a variety of molecules, including nucleic acids, proteins, and inorganic ions. Nevertheless, a nanopore detection platform with excellent performance has not been systematically unified to eliminate repetitions of experiments in different labs. In fact, different biological nanopores have been independently designed by several research teams to detect molecules. Due to a lack of agreed standards, there is a range of reported detection results for the same molecules, potentially hindering the development of biomolecular detection.
In DNA storage systems, digital information can be stored in DNA sequences, sequences containing modified DNA, or other biomolecules. Nanopore technology can optimize the reading process of DNA information, providing high-speed, accurate base calling in the field of DNA storage. However, DNA nanopore sequencing remains limited in two aspects. First, although the base-calling accuracy of nanopore sequencing has improved greatly, currently ranging 90–95%, it is still lower than the 99% accuracy of next-generation sequencing [122]. Second, the throughput of nanopore detection is relatively low. Nanopore detection relies on microcurrents from molecules passing through nanopores, and simultaneous detection of multiple nanopores can lead to signal distortion due to superposition of electrical signals. The current read speed for a single nanopore is ~10 ms/base, which would take approximately 20 years to sequence the human genome at a depth of 10× coverage [123]. Therefore, higher sequencing accuracy and detection throughput could promote wider commercial applications of large-scale DNA storage.

The application of AI in nanopore technology can be expected to overcome the above obstacles. In biomolecular detection, the powerful function of pattern recognition within deep learning models can allow for simultaneous detection of a wide range of molecules. Additionally, AI offers high-precision sequence prediction, thereby improving the speed and accuracy of base calling for DNA storage. Finally, AI can also efficiently predict molecular folding of proteins, which is useful to modify structures of biological nanopores, potentially providing a new avenue to improve detection throughput. With the help of more advanced nanotechnologies and AI, we anticipate that DNA nanopores will continue to provide new applications in the area of DNA information processing.

Author Contributions: Z.S. wrote the manuscript. J.Y. and Y.L. revised and supervised the manuscript. All authors have read and agreed to the published version of the manuscript.

Funding: National Key Research and Development Program of China [2017YFE0130600 and 2021YFF1200103]; National Natural Science Foundation of China [62273008, 62073133, 61872007]; Beijing Natural Science Foundation (Z201100008320002), CAS Interdisciplinary Innovation Team (JCTD-2020-04); GBA Research Innovation Institute for Nanotechnology, Guangzhou, 510530, China.

Conflicts of Interest: The authors declare no conflict of interest.

References
1. Zhirnov, V.; Zadegan, R.M.; Sandhu, G.S.; Church, G.M.; Hughes, W.L. Nucleic acid memory. Nat. Mater. 2016, 15, 366–370. [PubMed]
2. Goda, K.; Kitsuregawa, M. The History of Storage Systems. Proc. IEEE Inst. Electr. Electron. Eng. 2012, 100, 1433–1440.
3. Anžel, A.; Heider, D.; Hattab, G. The visual story of data storage: From storage properties to user interfaces. Comput. Struct. Biotechnol. J. 2021, 19, 4904–4918. [PubMed]
4. Andrae, A.S.G.; Edler, T. On Global Electricity Usage of Communication Technology: Trends to 2030. Challenges 2015, 6, 117–157.
5. Church, G.M.; Gao, Y.; Kosuri, S. Next-generation digital information storage in DNA. Science 2012, 337, 1628. [CrossRef]
6. Allentoft, M.E.; Collins, M.; Harker, D.; Haile, J.; Oskam, C.L.; Hale, M.L.; Campos, P.F.; Samaniego, J.A.; Gilbert, M.T.; Willerslev, E.; et al. The half-life of DNA in bone: Measuring decay kinetics in 158 dated fossils. Proc. Biol. Sci. 2012, 279, 4724–4733.
7. Bhat, W.A. Bridging data-capacity gap in big data storage. Future Gener. Comput. Syst. 2018, 87, 538–548.
8. Jaeger, L.; Chworos, A. The architecture of programmable RNA and DNA nanostructures. Curr. Opin. Struct. Biol. 2006, 16, 531–543.
9. Ishmukhametov, I.; Batasheva, S.; Rozhina, E.; Akhatova, F.; Mingaleeva, R.; Rozhin, A.; Fakhrrullin, R. DNA/Magnetic Nanoparticles Composite to Attenuate Glass Surface Nanotopography for Enhanced Mesenchymal Stem Cell Differentiation. Polymers 2022, 14, 344. [CrossRef]
10. Dong, Y.; Sun, F.; Ping, Z.; Ouyang, Q.; Qian, L. DNA storage: Research landscape and future prospects. Natl. Sci. Rev. 2020, 7, 1092–1107.
11. Heckel, R.; Shomorony, I.; Ramchandran, K.; David, N. Fundamental limits of DNA storage systems. In Proceedings of the 2017 IEEE International Symposium on Information Theory (ISIT), Aachen, Germany, 25–30 June 2017; pp. 3130–3134.
12. Merrifield, B. Solid phase synthesis. Science 1986, 232, 341–347. [PubMed]
13. Belitsky, J.M.; Nguyen, D.H.; Wurtz, N.R.; Dervan, P.B. Solid-phase synthesis of DNA binding polymides on oxime resin. Bioorg. Med. Chem. 2002, 10, 2767–2774. [PubMed]
14. Sikkema-Raddatz, B.; Johansson, L.F.; de Boer, E.N.; Almomani, R.; Boven, L.G.; van den Berg, M.P.; van Spaendonck-Zwarts, K.Y.; van Tintelen, J.P.; Sijmons, R.H.J.; Jongbloed, J.D. Targeted next-generation sequencing can replace Sanger sequencing in clinical diagnostics. *Hum. Mutat.* 2013, 34, 1035–1042. [PubMed]

15. Cao, C.; Long, Y.-T. Biological Nanopores: Confined Spaces for Electrochemical Single-Molecule Analysis. *Acc. Chem. Res.* 2018, 51, 331–341.

16. Oukhaled, A.; Bacri, L.; Pastoriza-Gallego, M.; Betton, J.-M.; Pelta, J. Sensing Proteins through Nanopores: Fundamental to Applications. *ACS Chem. Biol.* 2012, 7, 1935–1949.

17. Mereuta, L.; Asandei, A.; Seo, C.H.; Park, Y.; Luchian, T. Quantitative Understanding of pH- and Salt-Mediated Conformational Folding of Histidine-Containing, β-Hairpin-like Peptides, through Single-Molecule Probing with Protein Nanopores. *ACS Appl. Mater. Interfaces* 2014, 6, 13242–13256.

18. Hornblower, B.; Coombs, A.; Whitaiker, R.D.; Kolomeisky, A.; Picone, S.J.; Mellor, A.; Akeson, M. Single-molecule analysis of DNA-protein complexes using nanopores. *Nat. Methods* 2007, 4, 315–317.

19. Mathé, J.; Visram, H.; Viasnoff, V.; Rabin, Y.; Mellor, A. Nanopore unzipping of individual DNA hairpin molecules. *Biophys. J.* 2004, 87, 3205–3215. [CrossRef]

20. Luchian, T.; Park, Y.; Asandei, A.; Schiopu, I.; Mereuta, L.; Apetrei, A. Nanoscale Probing of Informational Polymers with Nanopores. Applications to Amyloidogenic Fragments, Peptides, and DNA–PNA Hybrids. *Acc. Chem. Res.* 2019, 52, 267–276. [CrossRef]

21. Desai, T.A.; Hansford, D.J.; Kulinsky, L.; Nashat, A.H.; Rasi, G.; Tu, J.; Wang, Y.; Zhang, M.; Ferrari, M. Nanopore technology for biomedical applications. *Biond. Microdevices* 1999, 2, 11–40. [CrossRef]

22. Chen, W.; Han, M.; Zhou, J.; Ge, Q.; Wang, P.; Zhang, X.; Zhu, S.; Song, L.; Yuan, Y. An artificial chromosome for data storage. *Nat. Sci. Rev.* 2021, 8, nwab028. [CrossRef]

23. Organick, L.; Ang, S.D.; Chen, Y.-J.; Lopez, R.; Yekhanin, S.; Makarychev, K.; Racz, M.Z.; Kamath, G.; Gopalan, P.; Nguyen, B.; et al. Random access in large-scale DNA data storage. *Nat. Biotechnol.* 2018, 36, 242–248. [CrossRef]

24. Schneider, G.F.; Dekker, C. DNA sequencing with nanopores. *Nat. Biotechnol.* 2012, 30, 326–328. [CrossRef] [PubMed]

25. Goto, Y.; Akahori, R.; Yanagi, I.; Takeda, K.-i. Solid-state nanopores towards single-molecule DNA sequencing. *J. Hum. Genet.* 2020, 65, 69–77. [CrossRef] [PubMed]

26. Clarke, J.; Wu, H.-C.; Jayasinghe, L.; Patel, A.; Reid, S.; Bayley, H. Continuous base identification for single-molecule nanopore DNA sequencing. *Nat. Nanotechnol.* 2009, 4, 265–270. [CrossRef] [PubMed]

27. Traversi, F.; Raillon, C.; Benameur, S.M.; Liu, K.; Khlybov, S.; Tosun, M.; Krasnozhon, D.; Kis, A.; Radenovic, A. Detecting the translocation of DNA through a nanopore using graphene nanoribbons. *Nat. Nanotechnol.* 2013, 8, 939–945. [CrossRef] [PubMed]

28. Riedl, J.; Ding, Y.; Fleming, A.M.; Burrows, C.J. Identification of DNA lesions using a third base pair for amplification and nanopore sequencing. *Nat. Commun.* 2015, 6, 8807. [CrossRef]

29. Fleming, A.M.; Mathewson, N.J.; Howpay Manage, S.A.; Burrows, C.J. Nanopore Dwell Time Analysis Permits Sequencing and Conformational Assignment of Pseudouridine in SARS-CoV-2. *ACS Cent. Sci.* 2021, 7, 1707–1717. [CrossRef]

30. Yan, S.; Wang, L.; Wang, Y.; Cao, Z.; Zhang, S.; Du, X.; Fan, P.; Zhang, P.; Chen, H.Y.; Huang, S. Non-binary Encoded Nucleic Acid Barcodes Directly Readable by a Nanopore. *Angew. Chem. Int. Ed.* 2022, 61, e202116482. [CrossRef]

31. Jia, W.; Hu, C.; Wang, Y.; Liu, Y.; Wang, L.; Zhang, S.; Zhu, Q.; Gu, Y.; Zhang, P.; Ma, J.; et al. Identification of Single-Molecule Catecholamine Enantiomers Using a Programmable Nanopore. *ACS Nano* 2020, 16, 6615–6624. [CrossRef]

32. Liu, P.; Kawano, R. Recognition of Single-Point Mutation Using a Biological Nanopore. *Small Methods* 2020, 4, 2000101. [CrossRef]

33. Wang, F.; Zahid, O.K.; Swain, B.E.; Parsonage, D.; Hollis, T.; Harvey, S.; Perrino, F.W.; Kohli, R.M.; Taylor, E.W.; Hall, A.R. Solid-State Nanopore Analysis of Diverse DNA Base Modifications Using a Modular Enzymatic Labeling Process. *Nano Lett.* 2017, 17, 7110–7116. [CrossRef] [PubMed]

34. Ying, Y.L.; Zhang, J.; Gao, R.; Long, Y.T. Nanopore-based sequencing and detection of nucleic acids. *Angew Chem. Int. Ed. Engl.* 2013, 52, 13154–13161. [CrossRef] [PubMed]

35. Wang, L.; Chen, X.; Zhou, S.; Roobabani, G.M.; Zhang, Y.; Wang, D.; Guan, X. Displacement chemistry-based nanopore analysis of nucleic acids in complicated matrices. *ChemComm* 2018, 54, 13977–13980. [CrossRef] [PubMed]

36. Workman, R.E.; Tang, A.D.; Tang, P.S.; Jain, M.; Tyson, J.R.; Razaghi, R.; Zuzarte, P.C.; Palstra, T.; Payne, A.; Quick, J.; et al. Nanopore native RNA sequencing of a human poly(A) transcriptome. *Nat. Methods* 2019, 16, 1297–1305. [CrossRef] [PubMed]

37. Nakane, J.J.; Akeson, M.; Marziali, A. Nanopore sensors for nucleic acid analysis. *J. Phys. Condens. Matter* 2003, 15, R1365. [CrossRef]

38. Deamer, D.W.; Branton, D. Characterization of nucleic acids by nanopore analysis. *Acc. Chem. Res.* 2002, 35, 817–825. [CrossRef]

39. Derrington, I.M.; Craig, J.M.; Stava, E.; Laszlo, A.H.; Ross, B.C.; Brinkerhoff, H.; Nova, I.C.; Doering, K.; Tickman, B.I.; Ronaghi, M.; et al. Subangstrom single-molecule measurements of motor proteins using a nanopore. *Nat. Biotechnol.* 2015, 33, 1073–1075. [CrossRef] [PubMed]

40. Yusko, E.C.; Bruhn, B.R.; Eggenberger, O.M.; Houghtaling, J.; Rollings, R.C.; Walsh, N.C.; Nandivada, S.; Pindrus, M.; Hall, A.R.; Sept, D.; et al. Real-time shape approximation and fingerprinting of single proteins using a nanopore. *Nat. Nanotechnol.* 2017, 12, 360–367. [CrossRef] [PubMed]

41. Waduge, P.; Hu, R.; Bandarkar, P.; Yamazaki, H.; Cressiot, B.; Zhao, Q.; Whitford, P.C.; Wanunu, M. Nanopore-Based Measurements of Protein Size, Fluctuations, and Conformational Changes. *ACS Nano* 2017, 11, 5706–5716. [CrossRef]
42. Wei, X.; Ma, D.; Zhang, Z.; Wang, L.Y.; Gray, J.L.; Zhang, L.; Zhu, T.; Wang, X.; Lenhart, B.J.; Yin, Y.; et al. N-Terminal Derivatization-Assisted Identification of Individual Amino Acids Using a Biological Nanopore Sensor. *ACS Sens.* 2020, 5, 1707–1716. [CrossRef] [PubMed]

43. Wloka, C.; Van Meervelt, V.; van Gelder, D.; Danda, N.; Jager, N.; Williams, C.P.; Maglia, G. Label-Free and Real-Time Detection of Protein Ubiquitination with a Biological Nanopore. *ACS Nano* 2017, 11, 4387–4394. [CrossRef]

44. Afshar Bakshloo, M.; Kasiansomwicicz, J.J.; Pastori-Gallo, M.; Mathé, J.; Daniel, R.; Piquet, F.; Oukhaled, A. Nanopore-Based Protein Identification. *J. Am. Chem. Soc.* 2022, 144, 2716–2725. [CrossRef] [PubMed]

45. Shimizu, K.; Mijiddorj, B.; Usami, M.; Mizoguchi, I.; Yoshida, S.; Hamada, Y.; Ohyama, A.; Usui, K.; Kawamura, I.; et al. De novo design of a nanopore for single-molecule detection that incorporates a β-hairpin peptide. *Nat. Nanotechnol.* 2022, 17, 67–75. [CrossRef] [PubMed]

46. Zhou, S.; Wang, H.; Chen, X.; Wang, Y.; Zhou, D.; Liang, L.; Wang, L.; Wang, D.; Guan, X. Single-molecule Study on the Interactions between Cyclic Nonribosomal Peptides and Protein Nanopore. *ACS Appl. Bio Mater.* 2020, 3, 554–560. [CrossRef] [PubMed]

47. Reiner, J.E.; Kasianowicz, J.J.; Nablo, B.J.; Robertson, J.W. Theory for polymer analysis using nanopore-based single-molecule mass spectrometry. *Proc. Natl. Acad. Sci. USA* 2010, 107, 12080–12085. [CrossRef]

48. Neher, E.; Sakmann, B. Single-channel currents recorded from membrane of denervated frog muscle fibres. *Nature* 1976, 260, 799–802. [CrossRef]

49. Arroyo, J.O.; Kukura, P. Non-fluorescent schemes for single-molecule detection, imaging and spectroscopy. *Nat. Photonics* 2016, 10, 11–17. [CrossRef]

50. Yuana, Y.; Oosterkamp, T.H.; Bahatyrova, S.; Ashcroft, B.; Garcia Rodriguez, P.; Bertina, R.M.; Osanto, S. Atomic force microscopy: A novel approach to the detection of nanosized blood microparticles. *J. Thromb. Haemost.* 2010, 8, 315–323. [CrossRef] [PubMed]

51. Li, Y.; Zhao, L.; Yao, Y.; Guo, X. Single-Molecule Nanotechnologies: An Evolution in Biological Dynamics Detection. *ACS Appl. Bio Mater.* 2020, 3, 68–85. [CrossRef] [PubMed]

52. Jain, M.; Olsen, H.E.; Paten, B.; Akeson, M. The Oxford Nanopore MinION: Delivery of nanopore sequencing to the genomics community. *Genome Biol.* 2016, 17, 239. [CrossRef] [PubMed]

53. Li, Y.; Zhao, L.; Yao, Y.; Guo, X. Single-Molecule Nanotechnologies: An Evolution in Biological Dynamics Detection. *ACS Appl. Bio Mater.* 2020, 3, 68–85. [CrossRef] [PubMed]

54. Huo, W.; Ling, W.; Wang, Z.; Ya, L.; Zhou, M.; Ren, M.; Li, X.; Li, J.; Xia, Z.; Liu, X.; et al. Miniaturized DNA Sequencers for Personal Use: Unreachable Dreams or Achievable Goals. *Front. Nanotechnol.* 2021, 3, 628861. [CrossRef]

55. Maglia, G.; Heron, A.J.; Stoddart, D.; Japun, D.; Bayley, H. Analysis of single nucleic acid molecules with protein nanotubes. *Meth. Enzymol.* 2010, 475, 591–623.

56. Deamer, D.; Akeson, M.; Branton, D. Three decades of nanopore sequencing. *Nano Lett.* 2016, 16, 518–524. [CrossRef] [PubMed]

57. Finkes, M.; Pedone, D.; Knez, T.; Döblinger, M.; Rant, U. Electrically facilitated translocations of proteins through silicon nitride nanopores: Conjoint and competitive action of diffusion, electrophoresis, and electroosmosis. *Nano Lett.* 2010, 10, 2162–2167. [CrossRef] [PubMed]

58. Song, L.; Hobaugh, M.R.; Shustak, C.; Chelsey, S.; Bayley, H.; Gouaux, J.E. Structure of staphylococcal alpha-hemolysin, a heptamer transmembrane pore. *Science* 1996, 274, 1859–1866. [CrossRef]

59. Walker, B.W.; Kasianowicz, J.J.; Krishnaswasy, M.V.; Bayley, H. A pore-forming protein with a metal-actuated switch. *Proc. Natl. Acad. Sci. USA* 2008, 105, 20647–20652. [CrossRef]

60. Akeson, M.; Branton, D.; Kasianowicz, J.J.; Brandin, E.; Deamer, D. Characterization of Individual Polynucleotide Molecules Using a Membrane Channel. *Proc. Natl. Acad. Sci. USA* 1996, 93, 13770–13773. [CrossRef]

61. Akeson, M.; Branton, D.; Kasianowicz, J.J.; Brandin, E.; Deamer, D.W. Microsecond time-scale discrimination among polycytidylic acid, polyadenylic acid, and polyuridylic acid as homopolymers or as segments within single RNA molecules. *Biophys. J.* 1999, 77, 3227–3233. [CrossRef]

62. Butler, T.Z.; Pavlenok, M.; Derrington, I.M.; Niederweis, M.; Gundlach, J.H. Single-molecule DNA detection with an engineered MspA pore on paramagnetic nanoparticles. *Proc. Natl. Acad. Sci. USA* 2008, 105, 20647–20652. [CrossRef] [PubMed]

63. Manroo, E.A.; Derrington, I.M.; Laszlo, A.H.; Langford, K.W.; Hopper, M.K.; Gillgren, N.; Pavlenok, M.; Niederweis, M.; Gundlach, J.H. Reading DNA at single-nucleotide resolution with a mutant MspA nanopore and phi29 DNA polymerase. *Nat. Biotechnol.* 2012, 30, 349–353. [CrossRef] [PubMed]

64. Manroo, E.A.; Derrington, I.M.; Pavlenok, M.; Niederweis, M.; Gundlach, J.H. Nucleotide discrimination with DNA immobilized in the MspA nanopore. *PLoS ONE* 2011, 6, e25723. [CrossRef] [PubMed]

65. Wang, Y.; Guan, X.; Zhang, S.; Liu, Y.; Wang, S.; Fan, P.; Du, X.; Yan, S.; Zhang, P.; Chen, H.-Y. Structural-profiling of low molecular weight RNAs by nanopore trapping/translocation using Mycobacterium smegmatis porin A. *Nat. Commun.* 2021, 12, 1–14. [CrossRef] [PubMed]

66. Roozbahani, G.M.; Chen, X.; Zhang, Y.; Wang, L.; Guan, X. Nanopore detection of metal ions: Current status and future directions. *Small Methods* 2020, 4, 2000266. [CrossRef] [PubMed]
68. Soskine, M.; Biesemans, A.; Moeyaert, B.; Cheley, S.; Bayley, H.; Maglia, G. An engineered ClyA nanopore detects folded target proteins by selective external association and pore entry. *Nano Lett.* 2012, 12, 4895–4900. [CrossRef]

69. Wang, Y.; Gu, L.Q.; Tian, K. The aerolysin nanopore: From peptidomim to genomic applications. *Nanoscale* 2018, 10, 13857–13866. [CrossRef]

70. Long, Z.; Zhan, S.; Gao, P.; Wang, Y.; Lou, X.; Xia, F. Recent Advances in Solid Nanopore/Channel Analysis. *Anal. Chem.* 2018, 90, 577–588. [CrossRef]

71. Howorka, S.; Siwy, Z. Nanopore analytics: Sensing of single molecules. *Chem. Soc. Rev.* 2009, 38, 2360–2384. [CrossRef]

72. van den Berg, A.; Wessling, M. Silicon for the perfect membrane. *Nature* 2007, 445, 726. [CrossRef] [PubMed]

73. Hu, R.; Tong, X.; Zhao, Q. Four Aspects about Solid-State Nanopores for Protein Sensing: Fabrication, Sensitivity, Selectivity, and Durability. *Adv. Healthc. Mater.* 2020, 9, 2000933. [CrossRef] [PubMed]

74. Kleefen, A.; Pedone, D.; Grunwald, C.; Wei, R.; Firnkes, M.; Abstreiter, G.; Rant, U.; Tampé, R. Multiplexed Parallel Single Transport Recordings on Nanopore Arrays. *Nano Lett.* 2010, 10, 5080–5087. [CrossRef] [PubMed]

75. Roman, J.; Jarroux, N.; Patriarche, G.; Français, O.; Pelta, J.; Le Picoule, B.; Bacrì, L. Functionalized Solid-State Nanopore Integrated in a Reusable Microfluidic Device for a Better Stability and Nanoparticle Detection. *ACS Appl. Mater. Interfaces* 2017, 9, 41634–41640. [CrossRef] [PubMed]

76. Storm, A.J.; Chen, J.H.; Ling, X.S.; Zandbergen, H.W.; Dekker, C. Fabrication of solid-state nanopores with single-nanometre precision. *Nat. Mater.* 2003, 2, 537–540. [CrossRef] [PubMed]

77. Freedman, K.J.; Otto, I.M.; Ivanov, A.P.; Barik, A.; Oh, S.-H.; Edel, J.B. Nanopore sensing at ultra-low concentrations using single-molecule dielectrophoretic trapping. *Nat. Commun.* 2016, 7, 10217. [CrossRef]

78. Garaj, S.; Huber, W.; Reina, A.; Kong, J.; Branton, D.; Golovchenko, J.A. Graphene as a subnanometre trans-electrode membrane. *Nature* 2010, 467, 190–193. [CrossRef]

79. Gasparyan, L.; Mazo, I.; Simonyan, V.; Gasparyan, F. DNA Sequencing: Current State and Prospects of Development. *Biophys. J.* 2019, 09, 169–197. [CrossRef]

80. Shi, W.; Friedman, A.K.; Baker, L.A. Nanopore Sensing. *Anal. Chem.* 2017, 89, 157–188. [CrossRef]

81. Schibel, A.E.; An, N.; Jin, Q.; Fleming, A.M.; Burrows, C.J.; White, H.S. Nanopore detection of 8-oxo-7,8-dihydro-2′-deoxyguanosine in immobilized single-stranded DNA via adduct formation to the DNA damage site. *J. Am. Chem. Soc.* 2010, 132, 17992–17995. [CrossRef]

82. Perera, R.T.; Fleming, A.M.; Johnson, R.P.; Burrows, C.J.; White, H.S. Detection of benz[a]pyrene-guanine adducts in single-stranded DNA using the α-hemolysin nanopore. *Nanotechnology* 2015, 26, 074002. [CrossRef] [PubMed]

83. Rouse, J.; Jackson, S.P. Interfaces between the detection, signaling, and repair of DNA damage. *Science* 2002, 297, 547–551. [CrossRef] [PubMed]

84. Zhou, B.B.; Elledge, S.J. The DNA damage response: Putting checkpoints in perspective. *Nature* 2000, 408, 433–439. [CrossRef] [PubMed]

85. Ma, F.; Yan, S.; Zhang, J.; Wang, Y.; Wang, L.; Zhang, S.; Du, X.; Zhang, P.; Chen, H.-Y.; et al. Nanopore Sequencing Accurately Identifies the Cisplatin Adduct on DNA. *ACS Sens.* 2021, 6, 3082–3092. [CrossRef] [PubMed]

86. Zhang, J.; Wang, Y.; Zhang, P.; Chen, H.-Y.; Huang, S. Discrimination between Different DNA Lesions by Monitoring Single-Molecule Polymerease Stalling Kinetics during Nanopore Sequencing. *Nano Lett.* 2022, 22, 5561–5569. [CrossRef]

87. Taylor, A.I.; Arangundy-Franklin, S.; Holliger, P. Towards applications of synthetic genetic polymers in diagnosis and therapy. *Curr. Opin. Chem. Biol.* 2014, 22, 79–84. [CrossRef]

88. Feldman, A.W.; Romesberg, F.E. Expansion of the Genetic Alphabet: A Chemist’s Approach to Synthetic Biology. *Acc. Chem. Res.* 2018, 51, 394–403. [CrossRef]

89. Pinheiro, V.B.; Arangundy-Franklin, S.; Holliger, P. Compartmentalized Self-Tagging for In Vitro-Directed Evolution of XNA Polymers. *Curr. Protoc. Nucleic Acid Chem.* 2014, 57, 9. [CrossRef]

90. Yan, S.; Li, X.; Zhang, P.; Wang, Y.; Chen, H.Y.; Huang, S.; Yu, H. Direct sequencing of 2′-deoxy-2′-fluoroarabinonucleic acid (FANA) using nanopore-induced phase-shift sequencing (NIPSS). *Chem. Sci.* 2019, 10, 3110–3117. [CrossRef]

91. Piguet, F.; Ouldali, H.; Pastoriza-Gallego, M.; Manivet, P.; Pelta, J.; Ouahmed, A. Identification of single amino acid differences in uniformly charged homopolymeric peptides with aerolysin nanopore. *Nat. Commun.* 2018, 9, 966. [CrossRef]

92. Movileanu, L.; Schmittschmitt, J.P.; Scholtz, J.M.; Bayley, H. Interactions of peptides with a protein pore. *Biophys. J.* 2005, 89, 1030–1045. [CrossRef] [PubMed]

93. Yan, S.; Zhang, J.; Wang, Y.; Guo, W.; Zhang, S.; Liu, Y.; Cao, J.; Wang, Y.; Wang, L.; Ma, F.; et al. Single Molecule Ratcheting Motion of Peptides in a Mycobacterium smegmatis Porin A (MspA) Nanopore. *Nano Lett.* 2021, 21, 6703–6710. [CrossRef] [PubMed]

94. Brinkerhoff, H.; Kang, A.S.; Liu, J.; Aksimentiev, A.; Dekker, C. Multiple rereads of single proteins at single–amino acid resolution using nanopores. *Science* 2021, 374, 1509–1513. [CrossRef] [PubMed]

95. Jia, W.; Hu, C.; Wang, Y.; Gu, Y.; Qian, G.; Du, X.; Wang, L.; Liu, Y.; Cao, J.; Zhang, S.; et al. Programmable nano-reactors for stochastic sensing. *Nat. Commun.* 2021, 12, 5811. [CrossRef] [PubMed]

96. Guo, J.; Amini, S.; Lei, Q.; Ping, Y.; Agola, J.O.; Wang, L.; Zhou, L.; Cao, J.; Franco, S.; Noureddine, A.; et al. Robust and Long-Term Cellular Protein and Enzymatic Activity Preservation in Biomineralized Mammalian Cells. *ACS Nano* 2022, 16, 2164–2175. [CrossRef] [PubMed]
Small Struct. 100. Hao, Y.; Li, Q.; Fan, C.; Wang, F. Data Storage Based on DNA.

Clin. Chem. 123. Bayley, H. Nanopore sequencing: From imagination to reality.

Liu, Q.; Fang, L.; Yu, G.; Wang, D.; Xiao, C.-L.; Wang, K. Detection of DNA base modifications by deep recurrent neural network for long-term data storage with earth alkaline salts. ChemComm 2020, 56, 3613–3616. [CrossRef]

Hao, Y.; Li, Q.; Fan, C.; Wang, F. Data Storage Based on DNA. Small Struct. 2021, 2, 2000046. [CrossRef]

Lopez, R.; Chen, Y.; Dumas Ang, S.; Yekhanin, S.; Makarychev, K.; Racz, M.Z.; Seelig, G.; Strauss, K.; Ceze, L. DNA assembly for nanopore data storage readout. Nat. Commun. 2019, 10, 2933. [CrossRef]

Liu, H.; Wang, J.; Song, S.; Fan, C.; Gothelf, K.V. A DNA-based system for selecting and displaying the combined result of two input variables. Nat. Commun. 2015, 6, 10089. [CrossRef] [PubMed]

Ge, Z.; Gu, H.; Li, Q.; Fan, C. Concept and Development of Framework Nucleic Acids. J. Am. Chem. Soc. 2018, 140, 17808–17819. [CrossRef] [PubMed]

Chen, K.; Kong, J.; Zhu, J.; Ermann, N.; Predki, P.; Keyser, U.F. Digital Data Storage Using DNA Nanostructures and Solid-State Nanopores. Nano Lett. 2019, 19, 1210–1215. [CrossRef] [PubMed]

Bell, N.A.W.; Keyser, U.F. Digitally encoded DNA nanostructures for multiplexed, single-molecule protein sensing with nanopores. Nat. Nanotechnol. 2016, 11, 645–651. [CrossRef]

Cao, C.; Krapp, L.F.; Al Ouahabi, A.; König, N.F.; Ciraquii, N.; Radenovic, A.; Lutz, J.F.; Peraro, M.D. Aerolysin nanopores decode digital information stored in tailored macromolecular analytes. Sci. Adv. 2020, 6, eabc2661. [CrossRef]

LeCun, Y.; Boser, B.E.; Denker, J.S.; Henderson, D.; Howard, R.E.; Hubbard, W.E.; Jackel, L.D. Backpropagation Applied to Handwritten Zip Code Recognition. Neural Comput 1989, 1, 541–551. [CrossRef]

Zaremba, W.; Sutskever, I.; Vinyals, O. Recurrent Neural Network Regularization. arXiv 2014, arXiv:1409.2329.

Hinton, G.E.; Salakhutdinov, R.R. Reducing the dimensionality of data with neural networks. Science 2006, 313, 504–507. [CrossRef]

Taniyagi, M.; Minami, S.; Ono, C.; Hamajima, R.; Morimura, A.; Hamaguchi, S.; Akeda, Y.; Kanai, Y.; Kobayashi, T.; Kamitani, W.; et al. Combining machine learning and nanopore construction creates an artificial intelligence nanopore for coronavirus detection. Nat. Commun. 2021, 12, 3726. [CrossRef]

Arima, A.; Tsutsui, M.; Washio, T.; Baba, Y.; Kawai, T. Solid-state nanopore platform integrated with machine learning for digital diagnosis of virus infection. Anal. Chem. 2020, 93, 215–227. [CrossRef]

Ledergerber, C.; Dessimoz, C. Base-calling for next-generation sequencing platforms. Brief. Bioinform. 2011, 12, 489–497. [CrossRef] [PubMed]

David, M.; Dursi, L.J.; Yao, D.; Boutros, P.C.; Simpson, J.T. Nanocall: An open source basecaller for Oxford Nanopore sequencing data. Bioinformatics 2017, 33, 49–55. [CrossRef]

Boža, V.; Brejová, B.; Vinař, T. DeepNano: Deep recurrent neural networks for base calling in MinION nanopore reads. PLoS ONE 2017, 12, e0178751. [CrossRef] [PubMed]

Stoiber, M.; Brown, J. BasecallRAWller: Streaming Nanopore Basecalling Directly from Raw Signal. bioRxiv 2017, 133058. [CrossRef]

Teng, H.; Cao, M.D.; Hall, M.B.; Duarte, T.; Wang, S.; Coin, L.J.M. Chiron: Translating nanopore raw signal directly into nucleotide sequence using deep learning. GigaScience 2018, 7, gix037. [CrossRef]

Liu, Q.; Fang, L.; Yu, G.; Wang, D.; Xiao, C.-L.; Wang, K. Detection of DNA base modifications by deep recurrent neural network on Oxford Nanopore sequencing data. Nat. Commun. 2019, 10, 2449. [CrossRef]

Ni, P.; Huang, N.; Nie, F.; Zhang, J.; Zhang, Z.; Wu, B.; Bai, L.; Liu, W.; Xiao, C.-L.; Luo, F.; et al. Genome-wide detection of cytosine methylations in plant from Nanopore data using deep learning. Nat. Commun. 2021, 12, 5976. [CrossRef]

Ni, P.; Huang, N.; Zhang, Z.; Wang, D.P.; Liang, F.; Miao, Y.; Xiao, C.L.; Luo, F.; Wang, J. DeepSignal: Detecting DNA methylation state from Nanopore sequencing reads using deep-learning. Bioinformatics 2019, 35, 4586–4595. [CrossRef]

Doroschak, K.; Zhang, K.; Queen, M.; Mandyam, A.; Strauss, K.; Ceze, L.; Nivala, J. Rapid and robust assembly and decoding of molecular tags with DNA-based nanopore signatures. Nat. Commun. 2020, 11, 5454. [CrossRef]

Buterez, D. Scaling up DNA digital data storage by efficiently predicting DNA hybridisation using deep learning. Sci. Rep. 2021, 11, 20517. [CrossRef]

Rang, F.J.; Kloosterman, W.P.; de Ridder, J. From squiggle to basepair: Computational approaches for improving nanopore sequencing read accuracy. Genome Biol. 2018, 19, 90. [CrossRef] [PubMed]

Bayley, H. Nanopore sequencing: From imagination to reality. Clin. Chem. 2015, 61, 25–31. [CrossRef] [PubMed]