CellPress

Trends in Biotechnology

Review

# Novel Modalities in DNA Data Storage

Cheng Kai Lim [1,5] Saurabh Nirantar,[2] Wen Shan Yew,[3,5] and Chueh Loo Poh [4,5,*]

**The field of storing information in DNA has expanded exponentially. Most common modalities involve encoding information from bits into synthesized nucleotides, storage in liquid or dry media, and decoding via sequencing. However, limitations to this paradigm include the cost of DNA synthesis and sequencing, along with low throughput. Further unresolved questions include the appropriate media of storage and the scalability of such approaches for commercial viability. In this review, we examine various storage modalities involving the use of DNA from a systems-level perspective. We compare novel methods that draw inspiration from molecular biology techniques that have been devised to overcome the difficulties posed by standard workflows and conceptualize potential applications that can arise from these advances.**

## Highlights

Viable information storage in DNA is largely limited by cost and throughput.

Advances in synthesis and sequencing are key in driving adoption.

Different novel methods of storing information outside of nucleotide conversion are being explored.

The key workflows are not established, giving significant room for exploration.

The integration of molecular biology, engineering, and computing will drive further innovation.

## DNA as a Novel Information Storage Medium

Alternative methods of storing data are a burgeoning area of research due to the immense amount of data being generated every year. Existing storage formats are insufficient to accommodate this growth. Among various potential storage formats, ranging from storage in quantum bits to the metabolome [1–4], DNA has emerged as a promising material due to its immense storage density, its longevity, and the surfeit of tools developed for manipulating its properties.

The basic information encoding workflow into DNA is well established [5–7]. It involves the **encoding** (see Glossary) of digital information into nucleotide sequences and synthesizing them (writing), organizing and storing them in an appropriate medium (storage), accessing desired sequences (random access), and finally sequencing and **decoding** (reading) them using a sequencer and appropriate decoding schemes (Figure 1). Error-correction methods are applied throughout to ensure data fidelity.

Although this workflow is elegant and facile, it currently faces certain limitations in its practical applications. First, its practical implementation is heavily dependent on the costs of DNA synthesis and sequencing. Although developments in both fields, such as the advent of enzymatic synthesis and new next-generation sequencing techniques, will likely continue to lead to a drop in overall costs, these are still projected to be orders of magnitude more than existing long-term data storage solutions [7]. Second, the present throughput of the workflow is low with long turnaround times: Synthesis takes weeks, whereas sequencing still requires days. Third, although existing error-correction mechanisms allow accurate retrieval of encoded data, inherent limitations in the synthesis and sequencing processes still generate significant errors, which complicates the encoding design. The various key aspects of the workflow and the work being done to address these limitations have been extensively reviewed [7–11].

Although the previously-mentioned limitations are being actively worked on, there is great interest in developing alternative or complementary novel methods that can circumvent these problems, offering fresh perspectives and directions to achieve practical DNA data storage solutions. These recently reported novel methods and workflows have not been covered in existing reviews, which have largely focused on various encoding techniques from an information theory perspective. In

[1]NUS Graduate School of Integrative Sciences and Engineering, National University of Singapore, Singapore 119077, Singapore
[2]Illumina Inc, Singapore 757716, Singapore
[3]Department of Biochemistry, Faculty of Medicine, National University of Singapore, Singapore 117597, Singapore
[4]Department of Biomedical Engineering, Faculty of Engineering, National University of Singapore, Singapore 117583, Singapore
[5]NUS Synthetic Biology for Clinical and Technological Innovation (SynCTI), Centre for Life Sciences, National University of Singapore, Singapore 117456, Singapore

*Correspondence:
poh.chuehloo@nus.edu.sg (C.L. Poh).

 https://doi.org/10.1016/j.tibtech.2020.12.008

this review, we detail these new methods and systems that have been devised recently and compare their unique characteristics with the existing workflow (Table 1). We posit that the current paradigm of research does not sufficiently leverage the rich history of molecular biology techniques and understanding and that the increasing cross-pollination of ideas from molecular biology and engineering, which recent work has begun to explore, will lead to new ways of storing data in DNA and advance DNA data storage toward practical applications. It is also clear that there is currently no single method capable of addressing all issues associated with DNA storage and that different workflows will fulfill different applications, which are also detailed in this review.

## Writing

### Encoding within Nucleotides

Most DNA storage methods store information in the form of nucleotide sequences (Figure 2A). The theoretical **Shannon information capacity** [12] of DNA is 2 bits/nt, although this limit is difficult to reach due to inherent errors in the DNA storage channel in the form of sequencing and synthesis error rates and biochemical constraints in sequence space [9,13]. As such, most work has been on encoding methods to optimize information redundancy. Other methods have explored ways such as using composite, degenerate base encoding, a way of representing positions in oligo sequences such that they represent a ratio of unique nucleotides in a mixed pool, thereby allowing increased logical density and reducing the number of synthesis cycles required [14,15]. Further possible advances in expanding the genetic code include incorporating unnatural nucleotides, such as **Hachimoji DNA** and unnatural base pairs [16,17], which results in an increase in the Shannon information capacity of DNA due to the expansion of usable alphabets. Another method involves the use of a fixed set of DNA oligomers that can be continuously reused to generate novel sequences, much like a printing press; this method was pioneered by the company Catalog Technologies, Inc. [18]. Although this reduces the overall information density, it decreases the costs required for synthesizing novel oligonucleotide sequences. Other variants include the use of **recombinase**-based methods to write information via the switching or excision of specific nucleotide sequences, which, despite its relative simplicity, is mostly restricted to *in vivo* applications due to their limited information capacity (Figure 2B)

### Glossary

**DNA computation:** a field of research involving implementing computational processes on DNA strands.
**DNA nanostructures:** the concept of generating unique shapes and structures made solely from DNA by complementary base pairing of defined DNA strands.
**Encoding and decoding:** the algorithm used to convert binary information into another form, usually DNA nucleotide sequences.
**Hachimoji DNA:** a synthetic nucleic analog that comprises four novel nucleotides with unique base-pairing abilities.
**Nanopores:** nanometer-sized pores, typically either a pore-forming protein or as a hole in a material such as silicon or graphene.
**Recombinase:** an enzyme that alters genetic sequences in a predefined manner.
**Shannon information capacity:** a measurement of the amount of information able to be stored in a defined material.
**Steganography:** a technique involving the concealment of secret data within existing materials.

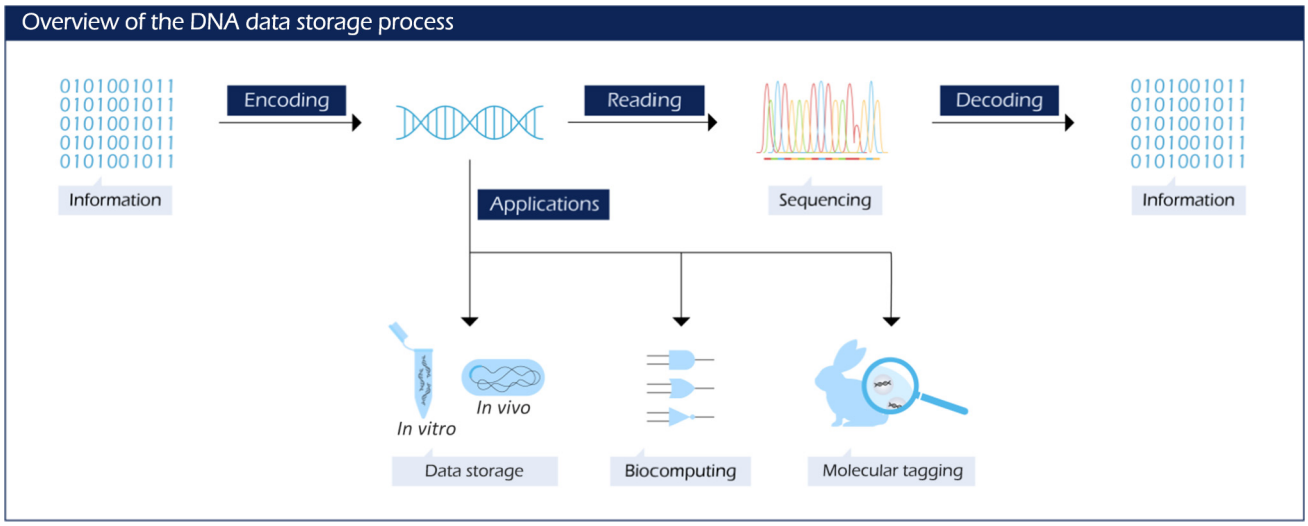


Figure 1. Overview of the DNA Data Storage Process. Information is encoded into DNA, which is subsequently stored in various storage media. These encoded DNA can be used for various applications, such as molecular tagging and computation. Information is retrieved from stored DNA via sequencing and subsequent decoding.

Table 1. Notable Proposed DNA Data Storage Schemes

| Writing | Storage | Retrieval | Notable advances | Amount of data stored | Refs |
|---|---|---|---|---|---|
| Sequence | Liquid | SBS | First notable demonstrations of DNA data storage | 650 KB/630 KB | [5,6] |
| Sequence | Silica beads | SBS | Storage in silica beads | 80 KB | [56] |
| Sequence | Liquid | SBS | Large-scale random access | 200 MB | [44] |
| Sequence | Living cells | SBS | Storage of digital information in living cells | 2.6 KB | [65] |
| Sequence | Chip | SBS | Storage on chips | 2.7 MB | [60] |
| Sequence | Liquid | SBS | Enzymatic synthesis of DNA with transition encoding | 110 bits | [99] |
| Structural | Liquid | Nanopore | Demonstration of structural information storage | 12 bits | [35] |
| Epigenetic | Liquid | SBS (with modifications) | Use of epigenetic modifications | NA | [38] |
| Recombinase | Living cells | NA | Recombinase writing of information in DNA | 1.375 bytes | [19] |

Abbreviations: KB, kilobytes; MB, megabytes; NA, not applicable; SBS, sequencing by synthesis.

[19,20]. Other methods of encoding and error correction, such as using Huffman codes or Reed-Solomon codes, have been reviewed extensively by others [7,8,21].

DNA synthesis, being the predominant method of encoding information into DNA, composes most of the costs associated with DNA data storage. Current synthesis costs have been estimated to be approximately $3500 per megabyte of data [22–24], depending on the scale of synthesis conducted, with large-scale processes such as silicon chip deposition platforms having lowered costs per base compared with smaller-scale, column-based synthesis [25]. The predominant method for synthesizing a large number of oligos suitable for DNA storage databases currently is via chemical synthesis on silicon chip platforms [26]. Other methods of DNA synthesis have been proposed, such as one involving the mass synthesis of random, unique oligo sequences, which are then sequenced. Desired sequences are then selectively retrieved from this verified pool of oligos and stored, thereby providing specified sequences [27–29]. A notable paradigm shift involves transitioning from chemical to enzymatic methods, which promises to lower costs due to the increased lengths that could be generated along with the potentially higher throughput and lower error rates (Box 1).

The previously-mentioned methods all deal with the synthesis of specific DNA strands that can encode information by directly mapping information in bits to nucleotides. However, the high costs involved in this process have prompted the exploration of alternative methods for writing information into DNA.

### Structural Encoding

One potential method is that of encoding information in **DNA nanostructures**, which involve the spontaneous folding of single-stranded DNA (ssDNA) using specified short ‘staple strands’ that bind with existing scaffolds (Figure 2C) [30–32]. Storing information in such nanostructures entails the formation of hairpin loops that can mimic bits via the mixture of a specific composition of nucleotides [33,34]. This concept was extended by forming a string of DNA hairpins, representing bits via their presence or absence, on a scaffold that is then read with **nanopores** [35,36]. Another similar concept uses enzymes to nick positions of DNA, thereby encoding information

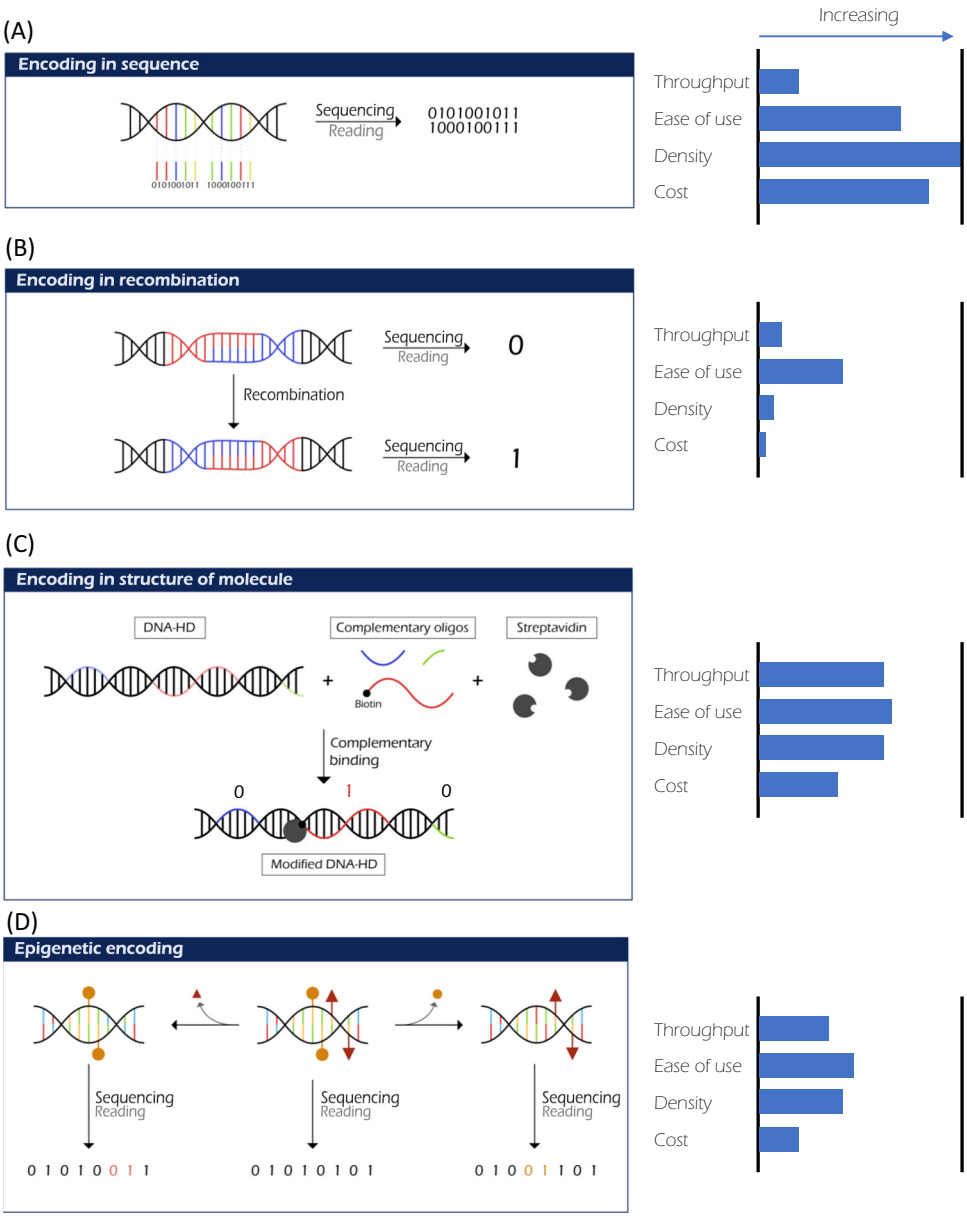


Figure 2. Various Information Encoding Methods into DNA. (A–D) Comparisons of characteristics of different information-encoding methods proposed thus far: sequence (A), recombination (B), structure (C), and epigenetic (D) [5,6,19,34,38]. 'Throughput' refers to the amount of information that can be encoded and decoded in a defined span of time. 'Ease of use' refers to the ease with which information can be encoded and decoded. 'Density' refers to the amount of information stored per gram of DNA, whereas 'cost' refers to the relative cost required to encode information for each method. 'DNA-HD' refers to the concept of a 'DNA hard-drive', whereby information can be written, erased, and rewritten by nanostructural changes. Bar charts on the right are qualitative representations of information based on references in the main text.

via the absence or presence of nicks [37]. Starting with a DNA template with '0' baseline state, bit transitions occur by the addition of appropriate oligos corresponding to 1s or 0s at different sites. The entire template can be 'reset' to the baseline unwritten state simply by heating, ready for

Box 1. Enzymatic DNA Synthesis Methods

Most common enzymatic DNA synthesis methods use the terminal deoxynucleotidyl transferase (TdT) enzyme, which is able to catalyze the template-independent addition of nucleotides on the 3′ end of a DNA strand, paving the way for synthesis of novel DNA strands by controlling the composition of nucleotide substrates. Various reviews of the potential for enzymatic DNA synthesis using TdT have been published [97,98]. Notably, Lee and colleagues proposed a method for using TdT in DNA synthesis for data storage purposes by encoding information in base transitions rather than the individual bases themselves, obviating the necessity for controlling individual nucleotide additions [99]. Improvements to this method have also been made by coupling synthesis on a chip with digital micromirror device (DMD) platforms, allowing spot-controlled activity of the TdT enzyme [100]. Further improvements can be made in resolution, along with expanding the available nucleotide pool, by using advances in DMD technology for nucleic acid photolithography [25,101]. Another notable advancement involved the use of TdT–nucleotide conjugates, as detailed by Palluk and colleagues [102]. In this method, TdT–nucleotide conjugates were developed with a unique conjugate for each base pair. This allows the controlled addition of unique base pairs to extend a DNA strand due to the inability of additional TdT–nucleotide conjugates to bind to the existing strand after the first incorporation event. The paper demonstrated synthesis of oligos up to 10-mer, with yields ranging from 93% to 98% [102]. Another unique method to generate unique ssDNA is via the use of primer exchange reaction cascades, which enables the production of ssDNA up to 60 nucleotides in length. This is done via the synthesis of user-specified sequences, which then form into hairpins that can be sequentially added to an existing DNA primer via polymerase strand displacement reactions, allowing the generation of longer sequences [103,104]. Another recently published paper detailed a novel way of synthesizing DNA using cyclic reversible termination enzymatic synthesis. This involves the use of single-stranded oligos on solid surfaces, whereby transient hybridization of oligos to another oligo allows extension by engineered DNA polymerases. In this manner, reversible nucleotides can be added to enable controlled synthesis of unique DNA strands using a universal templating strand as a template for transient hybridization. The authors demonstrated the production of a 20-mer ssDNA oligonucleotide with a stepwise efficiency of over 98% [105].

rewriting. Advantages of nanostructural information include cost savings because staple sequences can be premade and reused, parallelism of encoding, and ease of rewritability. Synthesizing long DNA oligos is thus obviated because existing DNA such as genomic DNA can serve as appropriate templates for sequence encoding, thereby saving costs. However, the information density is lowered, albeit still surpassing that of existing storage schemes, and such methods are also limited in scalability due to the necessity of synthesizing multiple individual staples coupled with the need to have multiple template DNAs suitable for 'writing' on. Furthermore, one advantage of DNA data storage, which is that of the ease of copying information via PCR, is not directly applicable in this scenario. Thus, this method likely would not replace the existing paradigm, but rather complement it – potentially as a way to store metadata or indexing information in a quick-access format to reduce the necessity of reading the entire pool of DNA.

### Epigenetic-Inspired Encoding

Another method of encoding information into DNA involves the use of epigenetics (Figure 2D) [38]. In living organisms, epigenetic alterations to DNA molecules act as an additional layer of information. This concept was used in DNA data storage by mapping information to modified nucleobases. The authors demonstrated retrieval of multiple layers of data from a single strand of DNA, depending on the chemical conversions conducted, with most conversions attaining >90% efficiency. Some drawbacks include increased latency, complexity, and costs due to the addition of conversion steps and specialized nucleotides. This method, however, shows ways in which chemical modifications and unique nucleotides, along with existing chemical or biological manipulations, can lead to different ways of storing and processing information in DNA.

## Storage and Access

The process of storing information in DNA involves consideration of the environment in which the DNA generated is stored. The resilience of DNA is highly dependent on the storage medium in which it was kept and its inherent redundancy. Key traits of an appropriate storage medium include the ability to preserve the existing information, additional post-processing costs and steps, and physical space requirements. The ability of random access is also an important consideration in developing the storage media.

Existing DNA data storage media are usually aqueous solutions (Figure 3A). Although this is relatively stable, DNA can still be damaged by hydrolysis and environmental factors [39,40]. Consequently, more redundancy in the form of additional copies is required. The main advantage of liquid storage is the ease of retrieval, with random access of DNA having previously been demonstrated [41–44]. This was achieved by the design of specific primer-binding sites in

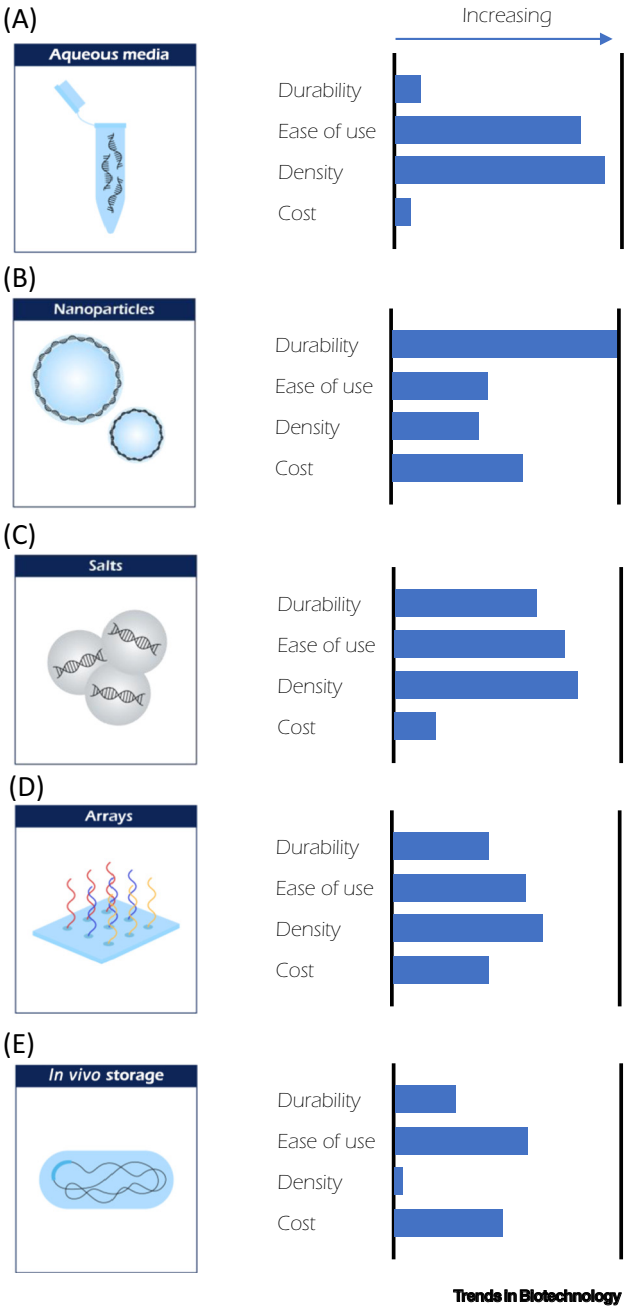


Figure 3. Comparison of Storage Formats. (A–E) DNA that is generated after encoding can be stored in multiple ways. Different storage formats that balance between varying characteristics such as density and accessibility have been proposed [56,58,60,65]. 'Durability' refers to the amount of time that information fidelity of DNA can be preserved. 'Ease of use,' 'density,' and 'cost' are as previously described. Bar charts are qualitative representations of information based on references in the main text.

every DNA oligo that can then be selectively retrieved via the addition of corresponding primers and the use of PCR to amplify targeted strands that can then be sequenced and decoded [43,45]. Although this allows selective access, the process is destructive: Strands that are not amplified will be lost, resulting in loss of information over time. This process is also error-prone, requiring an adequate amount of starting material, and retrieval can be difficult at lower concentrations. The lower limit of PCR retrieval was recently demonstrated to be approximately 10 copies per DNA strand, providing a theoretical capacity of 17EB $g^{-1}$ compared with the theoretical limit of 455EB $g^{-1}$ [5,13]. Selection via PCR also requires portions in the oligos to be allocated for random access primer binding, reducing information density. Thermocycling steps involved in the process can also result in damage of the DNA and also require specific equipment. Methods have recently been developed to overcome these limitations, such as the use of nested PCR for multidimensional access [46] or selective enrichment of DNA strands via use of biotin-labeled primers that are then captured by magnetic streptavidin beads [47]. Isothermal amplification techniques developed for rapid diagnostic testing can also be used to alleviate concerns of heat-induced damage [48].

A recent method proposed circumventing the use of direct PCR amplification by using biotin enrichment coupled with transcription into RNA for selection. This involves the selection of desired DNA sequences via specific primer binding that can be retrieved with biotin magnetic separation. The selected sequence is transcribed to RNA and reverse transcribed into DNA for information retrieval. The original sequence is then returned to the original pool. This method is enabled by the construction of ssDNA overhangs to form toeholds in their storage architecture [49]. The advantages include the preservation of the original sequences, the isothermal nature of the procedure, and the potential to conduct operations such as locking and editing of selected sequences. However, the long-term stability of such toeholds is a question that remains to be answered because exposed regions are more susceptible to secondary structure formation and enzymatic or chemical degradation than double-stranded DNA (dsDNA) oligos, which would likely obviate their use for long-term data storage.

### Storage in Alternative Media

Developing novel preservation media is another intensive area of research. Common DNA preservation techniques involve drying DNA in specific media compositions aimed at preventing the deleterious effects of hydrolysis or UV damage [50]. DNA has also been encapsulated in various substrates for preservation [51,52]. One encapsulation method involves the use of silica shells (Figure 3B) [53–56]. This entails the binding of DNA to functionalized silica beads, which are subsequently coated with a layer of silica, encapsulating the DNA. DNA stored in this manner has a projected shelf-life of approximately 2000 years while stored at 9.4°C [56]. The main limitation of such methods involves the lowered loading capacity and thus density of the medium, along with the increased complexity of the procedure itself. As such, further improvements, such as designing an alternating DNA–silica layer [57] or using salts (Figure 3C) [58], have been developed. One other disadvantage is the additional steps required for reading information due to the isolation of encapsulated DNA molecules, requiring the retrieval of DNA from its silica shell before selection can occur. To enable a level of access without removing the silica layer, another work proposed functionalizing encapsulated DNA particles with unique ssDNA strands, which act as barcodes for selection [59]. User-specified fluorescent ssDNA, with sequences complementary to the desired barcode, is then added to enable flow cytometric sorting and enrichment of desired beads, thus obviating the use of the entire encapsulated library for retrieval of selective information. Further development of such methods by integrating with writing formats as detailed in the previous section can potentially lead to ever simpler storage procedures while maximizing access and minimizing cost.

### Chip-Based Storage

Although most methods mentioned previously have generally assumed the resulting DNA to be stored in test tubes, such a storage method limits automation and accessibility. As such, chip-based storage, analogous to silicon wafers, has been explored. DNA has been stored in spatially segregated, dried spots on glass slides [60], allowing random access via selective rehydration of dried DNA spots using digital microfluidics [61] and reducing the necessity for PCR primer selection. Another alternative is by retention on microchip arrays (Figure 3D). There are many chip modalities that are functionalized with DNA, ranging from microarrays to sequencing chips. These can enable facile synthesis, processing, and access because the DNA is conjugated on the surface and will not be destroyed or lost upon replication. A recent work involved the creation of a 'DNA microdisk' whereby DNA strands are conjugated to microchips made of a polymer mesh engraved with a QR code containing metadata of the DNA bound to it [62]. Information on the composition of the disk is obtained by scanning the QR code, and nondestructive readout of the stored DNA information is enabled via denaturation of the dsDNA and resynthesis, showcasing a write-once, read-many approach. These chip-based modalities might possess lowered density due to the inherent spatial separation present and might require more fluidics or equipment for accessing said information, but the ability to provide another layer for random access and retrieval further opens avenues for increased manipulability and rapid access. Potential advancements include reducing the space required for effective spatial separation, such as through changing of chip substrates to enable closer packing. An interesting application enabled by spatial separation is that of proximity-based reactions, which was demonstrated via the use of spatially separated DNA hairpins on a DNA origami that allowed the creation of defined signal propagation and logic gates [63]. Similar advances could also open the door for DNA-on-a-chip modalities to have programmable computational and information-processing capabilities.

### *In Vivo* Storage

Last, one other alternative is storing information *in vivo* (Figure 3E), with a prominent demonstration being the insertion of an animated image via clustered regularly interspaced short palindromic repeats (CRISPR) spacer acquisition into a population of living cells [64,65]. One main advantage of this is the easy propagation of information due to the inherent ability of cellular reproduction to replicate information. More interestingly, this also comes along with the potential to use the properties of cellular processing to act as molecular recorders [66,67], functioning as living, decentralized smart devices that can sense and respond to varying environmental cues, storing this external information in an internal DNA 'hard drive' (Box 2). However, the available storage

#### Box 2. *In Vivo* DNA Recording and Writing

A number of methods have been developed to enable the writing of environmental signals or cues into DNA itself, allowing the archiving and subsequent retrieval of such events via reading of DNA in the appropriate manner [67,106,107]. Initial examples of such molecular recorders used recombinases that recorded information via flipping or excision of specified DNA sequences, which can then be determined via sequencing, DNA separation methods such as chromatography or gel electrophoresis, or phenotypic changes that can be observed. Such recombinases have been shown to record over 1 byte of combinatorial events [20]. Another notable development in the recording of cellular events is via the use of the genetic modification tool, CRISPR. Cas9 nucleases and base editors were used in two systems, known as CAMERA 1 and CAMERA 2, respectively, which were able to record cellular exposure to multiple types of stimuli such as antibiotics, viruses, and light. Recordings could also be read in the order of stimulus events and were shown to function in both bacterial and mammalian cells [108]. This concept of using base editors as write heads was extended in the DOMINO schematic, whereby base editing was expanded with multiple guide RNA recording to generate logic gate sequential DNA writing in specified DNA locations [109]. Another notable concept using CRISPR leveraged the acquisition of spacer sequences, which are then incorporated into the CRISPR array in a unidirectional manner, analogous to a tape recorder. This concept has been used to record DNA that was externally introduced either via a transformed plasmid producing intracellular DNA [66] or simply via electroporation of oligonucleotide sequences [64,65]. An extension of this concept further allowed the recording of intracellular transcriptional events via the incorporation of intracellular RNA abundance as CRISPR spacer sequences [110,111]. Further advances in this domain promise to expand the recording capacity of individual cells, as well as the variety and complexity of signals and inputs that can be recorded and processed.

density is much lower because existing *in vivo* writing methods are unable to fully use the available genomic space within a cell, and there are risks of information loss via mutations, although durability of information was demonstrated to potentially last for 12 000 years in recent work [68]. Costs of storage in living cells vary, depending on whether the information to be stored is synthesized from scratch or if it is written via recombinase- or gene editing–based methods, with synthesis from scratch being significantly more costly.

## Reading

Advances in sequencing, the predominant form of information retrieval or 'reading,' drive the viability of DNA data storage. Predominant methods of reading DNA include sequencing by synthesis (SBS), along with third-generation sequencing methods involving single-molecule sequencing via nanopores or enzymatic well-based reactions, which have been reviewed extensively [69,70]. This section thus covers specific considerations of sequencing in DNA data storage and highlights advances made that can potentially lead to novel methods of reading information from DNA.

### SBS

Most DNA data storage methods use SBS for information retrieval. This is due to the immense throughput provided, allowing reads ranging from 4 million to 20 billion in number (Illumina Sequencing Platforms [https://www.illumina.com/systems/sequencing-platforms.html]). Sequencing runs are never error-free, with various substitution, insertion, and deletion errors occurring in every run [9,71]. Thus, a portion of the potential data storage pool is used as overhead to increase sequence redundancy and allow error correction. Current storage methods do not subject the entire database of information to sequencing; sequences to be read are usually first amplified via PCR. It was previously demonstrated that a mean copy number of 10 initial DNA molecules required a sequencing coverage mean of 35× for accurate retrieval of data [13,44], albeit without controlling for variation in the DNA storage workflow itself in the form of differences in DNA synthesis, library preparation, and sequencing methods.

The necessity of prior amplification of DNA via PCR is one limitation of SBS-based methods because PCR amplification has detrimental effects on the fidelity of a sequence and results in significant bias due to inherent stochasticity [72]. SBS is also potentially destructive because DNA sequences are generally not recovered after sequencing. The information is also not obtained in 'real time'; only after completion of the run can information be retrieved and decoded. SBS is also limited to reading bases; any modifications to the sequence, such as addition of epigenetic modifiers or structural changes, must undergo various forms of processing before being sequenced. As such, it is likely that DNA storage methods using SBS as their 'read head' would likely be more suited for large-scale, long-term data storage, whereby a large amount of information is accessed infrequently.

### Single-Molecule Sequencing

Single-molecule sequencing, particularly that of nanopore sequencing, has also been explored. Nanopore sequencing involves the use of small protein or solid-state nanopores in an insulating membrane, which allows a stream of ionic particles to pass through, resulting in a measurable current. These pores are only wide enough to allow a single molecule of DNA to pass through, whereby the different bases composing the DNA strand block the pore to different extents as they pass through the nanopore, resulting in a sequence-defined current trace, from which the nucleotide sequence can be extracted. Although nanopore sequencing is more error-prone than SBS methods, with an average error rate of 5–15% [73], the relative speed of the process, coupled with the lower costs, its real-time nature, and the longer read lengths available, have led it to be the prominent alternative method for sequencing for both genomics and DNA data storage

methodologies [41,45,74,75]. Potential advances in enzymatic synthesis also promise to generate longer strands, playing to the advantages of long-read sequencing. The relative lack of preprocessing and ease of use have also opened up opportunities for automation, as demonstrated by a customized setup built for automation of the DNA data storage workflow [10]. The density of information stored for nanopore readouts is inherently less due to the increased error rates necessitating more redundancy and error correction, but developments in accuracy with methods ranging from bioinformatics algorithms to multiplexing primers show promise in resolving these issues [73,76–78].

Another advantage of nanopores is their ability to 'read' various properties of molecules, such as size differences and modifications. Thus, the previously-mentioned novel methods of storage, such as via nanostructures, can only be read by single-molecule technologies because these methods require the preservation of higher-order structures. The range of material properties and molecules that can be detected by nanopores provides more flexibility for them to be used as a potential 'read head' for future DNA storage modalities [79,80].

Other sequencing technologies are also in the pipeline. Notably, there have been efforts to develop semiconductor-based sequencers. This was first demonstrated using polymerases bound to carbon nanotubes [81], with companies such as Roswell Biotechnologies developing similar technologies [82]. These advances promise to provide a much faster rate of reading than current methods and at a cheaper cost, which could accelerate adoption of DNA data storage. The further development of bioelectronic interfaces also showcase possible ways in which DNA can be directly read electronically, with one notable demonstration of CRISPR-Chip™ technology from Cardea Bio using a CRISPR–Cas9 complex conjugated to a graphene chip. This allows the biosensing of specific DNA molecules complementary to specified guide RNAs, which results in a change of conformation in the CRISPR complex upon binding that can be detected as a shift in electric current through the graphene layer of the chip [83]. Another potential advance demonstrates an interface between biochemical and electronic signals, whereby DNA is used as an information transfer channel to convert chemical inputs into electronic signals [84]. These single-molecule efforts, coupled with advances in molecular interface developments, demonstrate a promising way forward for DNA to act as a 'hot' storage medium, whereby small amounts of information in DNA is rapidly converted into electronic bits that can then be manipulated with existing computational architectures.

## Novel Applications of DNA Data Storage

The advent of new modalities reviewed in this paper has also led to the development of novel applications for DNA information storage beyond information archiving (Figure 1). One application is in the field of item authenticity and verification, whereby DNA that contains information on the veracity of an item is tagged onto the item itself. Multiple groups have developed ways of implementing DNA-based authentication, with tags being used in barcoding of oils [55], and for use in environmental tracing for characterizing aquifers [85]. This was extended with the development of the DNA of Things [86], which entailed the storage of information required to synthesize a product within the molecules of the product itself, similar to how individual cells in an animal contain the genetic code necessary to 'synthesize' the entire animal. This novel storage architecture was achieved by encapsulating DNA using silica beads and mixing the beads with the material used to shape the target object. This thus allows the ability to essentially store information as DNA in any physical object and potentially opens up new areas for DNA storage, molecular identification, and DNA **steganography**.

A notable recent development involves the generation of molecular DNA tags that are then mixed together in specific amounts, allowing tagging to happen in a rapid and similarly implemented in a

concept termed 'MolBits,' whereby a pool of known DNA strands are mixed in specific amounts to form a 96-bit tag that is applied to objects and can be read to determine object provenance and identity [87]. These applications would require rapid on-site sequencing capabilities, which would be more amenable to long-read, single-molecule sequencing efforts as listed previously. Companies such as Aanika Biosciences, which uses unique microbial combinations that encode for information, have used such forms of molecular DNA barcoding.

Further exciting applications of DNA data storage can be seen in the rise of **DNA computation**. Nucleic acids have been increasingly proposed as a medium of computing, with multiple advances such as logic gates and neural networks developed recently [31,63,88–90]. However, there has not been a direct link between storage of data in DNA and computing with that very same data. Recent work has proposed the development of DNA circuitry, whereby writing and reading of eight combinations of three-bit data could be done rapidly and robustly [91]. This opens the way toward electronic circuits that can directly sense biological reactions and circuit behavior, which would be necessary for truly rapid, digital-like programming and responses. Another notable development proposed the implementation of a similarity search in a DNA database, leveraging the parallelism of DNA hybridization processes. This was done by representing document similarities as geometric space, which is preserved in DNA. This preservation entails the formation of stable, hybridized structures of single-stranded target and queries when both are neighboring and unstable structures when both are not. This thus allows a similarity search to occur simply via the binding and retrieval of stable structures upon addition of a query to the single-stranded database [92]. The encoding of information as topological alterations on DNA also further opens up computational possibilities, as detailed in the SIMD|| DNA (single instruction, multiple data computation with DNA strand displacement cascades) concept, whereby DNA strand displacement *in vitro* allows alterations of DNA topologies, thereby enabling computation [93].

Further development of such concepts would require designing algorithms that can convert binary information into DNA sequences that can also fulfill logic operations or to explore ways to edit and alter DNA sequences to emulate logic operations. These advances pave the way toward a hybrid molecular-electronic computational framework, as proposed in another review [94], combining the inherent parallelism afforded by DNA with the rapid serialism provided by silicon-based storage media.

## Concluding Remarks and Future Prospects

DNA has been the information store for biological information for millennia and has been the foundation of much work in modern biology. Now, its potential as a manipulatable material, coupled with the advent of technologies that can assist in this manipulation, has greatly expanded its applications. Data storage is but one of these applications. Novel advances in biotechnology have led to innovative approaches to storing data in DNA, and further innovations will only continue to increase as our understanding of how to engineer DNA grows over time. Increasing developments from both engineering and computing perspectives, in addition to biotechnological advancements, are necessary for the continued development of this fertile area of research.

The early stage of research has thus far shown that the inherent concept of DNA data storage is entirely feasible. However, more work is required to develop a biological data storage device that can seamlessly interface with DNA (see Outstanding Questions). This can be adapted from how biological systems, which are a natural demonstration of how information can be stored in DNA and subsequently manipulated to produce a multitude of outcomes, are designed. New systems can be created to leverage the unique traits of biological storage, such as dimensionality,

### Outstanding Questions

How can we increase the speed and throughput of writing and reading DNA?

What are the requirements for a robust DNA storage format?

How should the DNA be stored to facilitate preservation and retrieval?

How will emerging biotechnological advances be integrated with this workflow?

What does a DNA storage device designed from first principles look like?

What novel applications will arise from advances in this field?

parallelism, and biological significance, by including spatial interactions inspired by chromosomal packing, parallel computation with protein–DNA interactions, and even communication between packets of protocells or bacterial networks for information processing [95,96]. Further integration of various molecular biology tools, such as novel PCR techniques, CRISPR, DNA nanotechnology, and more, also has potential to create novel advances in the field. The design of such systems, coupled with advances in the speed and throughput of writing and reading DNA, will be key to the viability of DNA as a data storage medium.

## Acknowledgments

The authors thank Wen Xuan Er for her help in creation of the figures used in this paper. The authors are grateful for the support provided by the Synthetic Biology Initiative of the National University of Singapore (NUS) (DPRT/943/09/14), the Summit Research Program of the National University Health System (NUHSRO/2016/053/SRP/05), and an NUS startup grant (R-397-000-257-133).

## References

1. Kalff, F.E. *et al.* (2016) A kilobyte rewritable atomic memory. *Nat. Nanotechnol.* 11, 926–929
2. Cafferty, B.J. *et al.* (2019) Storage of information using small organic molecules. *ACS Cent. Sci.* 5, 911–916
3. Arcadia, C.E. *et al.* (2020) Multicomponent molecular memory. *Nat. Commun.* 11, 691
4. Rosenstein, J.K. *et al.* (2020) Principles of information storage in small-molecule mixtures. *IEEE Trans. Nanobiosci.* 19, 378–384
5. Church, G.M. *et al.* (2012) Next-generation digital information storage in DNA. *Science* 337, 1628
6. Goldman, N. *et al.* (2013) Towards practical, high-capacity, low-maintenance information storage in synthesized DNA. *Nature* 494, 77–80
7. Ceze, L. *et al.* (2019) Molecular digital data storage using DNA. *Nat. Rev. Genet.* 20, 456–466
8. Meiser, L.C. *et al.* (2020) Reading and writing digital data in DNA. *Nat. Protoc.* 15, 86–101
9. Heckel, R. *et al.* (2019) A characterization of the DNA data storage channel. *Sci. Rep.* 9, 9663
10. Takahashi, C.N. *et al.* (2019) Demonstration of end-to-end automation of DNA data storage. *Sci. Rep.* 9, 4998
11. Dong, Y. *et al.* (2020) DNA storage: research landscape and future prospects. *Natl. Sci. Rev.* 7, 1092–1107
12. Shannon, C.E. (1948) A mathematical theory of communication. *Bell Syst. Tech. J.* 27, 379–423
13. Organick, L. *et al.* (2020) Probing the physical limits of reliable DNA data retrieval [published correction appears in *Nat. Commun.* (2020) 11, 1080]. *Nat. Commun.* 11, 616
14. Choi, Y. *et al.* (2019) High information capacity DNA-based data storage with augmented encoding characters using degenerate bases. *Sci. Rep.* 9, 6582
15. Anavy, L. *et al.* (2019) Data storage in DNA with fewer synthesis cycles using composite DNA letters. *Nat. Biotechnol.* 37, 1229–1236
16. Malyshev, D.A. *et al.* (2012) Efficient and sequence-independent replication of DNA containing a third base pair establishes a functional six-letter genetic alphabet. *Proc. Natl. Acad. Sci. U. S. A.* 109, 12005–12010
17. Hoshika, S. *et al.* (2019) Hachimoji DNA and RNA: a genetic system with eight building blocks. *Science* 363, 884–887
18. Roquet, N. *et al.* (2020) Catalog Technologies, Inc. In *Nucleic acid-based data storage*
19. Bonnet, J. *et al.* (2012) Rewritable digital data storage in live cells via engineered control of recombination directionality. *Proc. Natl. Acad. Sci. U. S. A.* 109, 8884–8889
20. Yang, L. *et al.* (2014) Permanent genetic memory with >1-byte capacity. *Nat. Methods* 11, 1261–1266
21. Rutten, M.G.T.A. *et al.* (2018) Encoding information into polymers. *Nat. Rev. Chem.* 2, 365–381
22. Kosuri, S. and Church, G.M. (2014) Large-scale de novo DNA synthesis: technologies and applications. *Nat. Methods* 11, 499–507
23. Zhirnov, V. *et al.* (2016) Nucleic acid memory. *Nat. Mater.* 15, 366–370
24. Hughes, R.A. and Ellington, A.D. (2017) Synthetic DNA synthesis and assembly: putting the synthetic in synthetic biology. *Cold Spring Harb. Perspect. Biol.* 9, a023812
25. Antkowiak, P.L. *et al.* (2020) Low cost DNA data storage using photolithographic synthesis and advanced information reconstruction and error correction. *Nat. Commun.* 11, 5345
26. LeProust, E.M. *et al.* (2010) Synthesis of high-quality libraries of long (150mer) oligonucleotides by a novel depurination controlled process. *Nucleic Acids Res.* 38, 2522–2540
27. Lee, H. *et al.* (2015) A high-throughput optomechanical retrieval method for sequence-verified clonal DNA from the NGS platform. *Nat. Commun.* 6, 6073
28. Hwang, B. and Bang, D. (2016) Toward a new paradigm of DNA writing using a massively parallel sequencing platform and degenerate oligonucleotide. *Sci. Rep.* 6, 37176
29. Lim, H. *et al.* (2018) Highly selective retrieval of accurate DNA utilizing a pool of in situ-replicated DNA from multiple next-generation sequencing platforms. *Nucleic Acids Res.* 46, e40
30. Seeman, N.C. and Sleiman, H.F. (2018) DNA nanotechnology. *Nat. Rev. Mater.* 3, 17068
31. Li, J. *et al.* (2017) Engineering nucleic acid structures for programmable molecular circuitry and intracellular biocomputation. *Nat. Chem.* 9, 1056–1067
32. Hunter, P. (2018) Nucleic acid-based nanotechnology. *EMBO Rep.* 19, 13–17
33. Halvorsen, K. and Wong, W.P. (2012) Binary DNA nanostructures for data encryption. *PLoS One* 7, e44212
34. Chandrasekaran, A.R. *et al.* (2017) Addressable configurations of DNA nanostructures for rewritable memory. *Nucleic Acids Res.* 45, 11459–11465
35. Chen, K. *et al.* (2019) Digital data storage using DNA nanostructures and solid-state nanopores. *Nano Lett.* 19, 1210–1215
36. Chen, K. *et al.* (2020) Nanopore-based DNA hard drives for rewritable and secure data storage. *Nano Lett.* 20, 3754–3760
37. Tabatabaei, S.K. *et al.* (2020) DNA punch cards for storing data on native DNA sequences via enzymatic nicking. *Nat. Commun.* 11, 1742
38. Mayer, C. *et al.* (2016) An epigenetics-inspired DNA-based data storage system. *Angew. Chem. Int. Ed.* 55, 11144–11148
39. Lindahl, T. and Nyberg, B. (1972) Rate of depurination of native deoxyribonucleic acid. *Biochemistry* 11, 3610–3618
40. Pfeifer, G.P. *et al.* (2005) Mutations induced by ultraviolet light. *Mutat. Res. Mol. Mech. Mutagen.* 571, 19–31
41. Yazdi, S.M.H.T. *et al.* (2015) A rewritable, random-access DNA-based storage system. *Sci. Rep.* 5, 14138
42. Bornholt, J. *et al.* (2016) A DNA-based archival storage system. In *ASPLOS '16: Proceedings of the Twenty-First International Conference on Architectural Support for Programming Languages and Operating Systems, Atlanta, GA, USA*, pp. 637–649, Association for Computing Machinery

43. Yazdi, S.M.H.T. *et al.* (2017) Portable and error-free DNA-based data storage. *Sci. Rep.* 7, 5011
44. Organick, L. *et al.* (2018) Random access in large-scale DNA data storage. *Nat. Biotechnol.* 36, 242–248
45. Lopez, R. *et al.* (2019) DNA assembly for nanopore data storage readout. *Nat. Commun.* 10, 2933
46. Song, X. *et al.* (2019) Multidimensional data organization and random access in large-scale DNA storage systems. *bioRxiv* Published online August 22, 2019. https://doi.org/10.1101/743369
47. Tomek, K.J. *et al.* (2019) Driving the scalability of DNA-based information storage systems. *ACS Synth. Biol.* 8, 1241–1248
48. Gill, P. and Ghaemi, A. (2008) Nucleic acid isothermal amplification technologies – a review. *Nucleosides Nucleotides Nucleic Acids* 27, 224–243
49. Lin, K.N. *et al.* (2020) Dynamic and scalable DNA-based information storage. *Nat. Commun.* 11, 2981
50. Wan, E. *et al.* (2010) Green technologies for room temperature nucleic acid storage. *Curr. Issues Mol. Biol.* 12, 135–142
51. Clermont, D. *et al.* (2014) Assessment of DNA encapsulation, a new room-temperature DNA storage method. *Biopreserv. Biobank.* 12, 176–183
52. Howlett, S.E. *et al.* (2014) Evaluation of DNAstable™ for DNA storage at ambient temperature. *Forensic Sci. Int. Genet.* 8, 170–178
53. Paunescu, D. *et al.* (2013) Protection and deprotection of DNA – high-temperature stability of nucleic acid barcodes for polymer labeling. *Angew. Chem. Int. Ed.* 52, 4269–4272
54. Paunescu, D. *et al.* (2013) Reversible DNA encapsulation in silica to produce ROS-resistant and heat-resistant synthetic DNA 'fossils. *Nat. Protoc.* 8, 2440–2448
55. Puddu, M. *et al.* (2014) Magnetically recoverable, thermostable, hydrophobic DNA/silica encapsulates and their application as invisible oil tags. *ACS Nano* 8, 2677–2685
56. Grass, R.N. *et al.* (2015) Robust chemical preservation of digital information on DNA in silica with error-correcting codes. *Angew. Chem. Int. Ed.* 54, 2552–2555
57. Chen, W.D. *et al.* (2019) Combining data longevity with high storage capacity – layer-by-layer DNA encapsulated in magnetic nanoparticles. *Adv. Funct. Mater.* 29, 1901672
58. Kohll, A.X. *et al.* (2020) Stabilizing synthetic DNA for long-term data storage with earth alkaline salts. *Chem. Commun. (Camb.)* 56, 3613–3616
59. Banal, J.L. *et al.* (2020) Random access DNA memory in a scalable, archival file storage system. *bioRxiv* Published online April 6, 2020. https://doi.org/10.1101/2020.02.05.936369
60. Newman, S. *et al.* (2019) High density DNA data storage library via dehydration with digital microfluidic retrieval. *Nat. Commun.* 10, 1706
61. Willsey, M. *et al.* (2019) Scaling microfluidics to complex, dynamic protocols: invited paper. In *2019 IEEE/ACM International Conference on Computer-Aided Design (ICCAD), Westminster, CO, USA*, pp. 1–6
62. Choi, Y. *et al.* (2020) DNA micro-disks for the management of DNA-based data storage with index and write-once–read-many (WORM) memory features. *Adv. Mater.* 32, 2001249
63. Chatterjee, G. *et al.* (2017) A spatially localized architecture for fast and modular DNA computing. *Nat. Nanotechnol.* 12, 920–927
64. Shipman, S.L. *et al.* (2016) Molecular recordings by directed CRISPR spacer acquisition. *Science* 353, aaf1175
65. Shipman, S.L. *et al.* (2017) CRISPR–Cas encoding of a digital movie into the genomes of a population of living bacteria. *Nature* 547, 345–349
66. Sheth, R.U. *et al.* (2017) Multiplex recording of cellular events over time on CRISPR biological tape. *Science* 358, 1457–1461
67. Sheth, R.U. and Wang, H.H. (2018) DNA-based memory devices for recording cellular events. *Nat. Rev. Genet.* 19, 718–732
68. Song, L. and Zeng, A.-P. (2018) Orthogonal information encoding in living cells with high error-tolerance, safety, and fidelity. *ACS Synth. Biol.* 7, 866–874
69. Deamer, D. *et al.* (2016) Three decades of nanopore sequencing. *Nat. Biotechnol.* 34, 518–524
70. Shendure, J. *et al.* (2017) DNA sequencing at 40: past, present and future. *Nature* 550, 345–353
71. Schirmer, M. *et al.* (2016) Illumina error profiles: resolving fine-scale variation in metagenomic sequencing data. *BMC Bioinforma.* 17, 125
72. Chen, Y.-J. *et al.* (2020) Quantifying molecular bias in DNA data storage. *Nat. Commun.* 11, 3264
73. Rang, F.J. *et al.* (2018) From squiggle to basepair: computational approaches for improving nanopore sequencing read accuracy. *Genome Biol.* 19, 90
74. Loose, M. *et al.* (2016) Real-time selective sequencing using nanopore technology. *Nat. Methods* 13, 751–754
75. Fuller, C.W. *et al.* (2016) Real-time single-molecule electronic DNA sequencing by synthesis using polymer-tagged nucleotides on a nanopore array. *Proc. Natl. Acad. Sci. U. S. A.* 113, 5233–5238
76. Karst, S.M. *et al.* (2020) Enabling high-accuracy long-read amplicon sequences using unique molecular identifiers with Nanopore or PacBio sequencing. *bioRxiv* Published online January 11, 2020. https://doi.org/10.1101/645903
77. Currin, A. *et al.* (2019) Highly multiplexed, fast and accurate nanopore sequencing for verification of synthetic DNA constructs and sequence libraries. *Synth. Biol.* 4, ysz025
78. Chandak, S. *et al.* (2020) Overcoming high nanopore basecaller error rates for DNA storage via basecaller-decoder integration and convolutional codes. *bioRxiv* Published online April 6, 2020. https://doi.org/10.1101/2019.12.20.871939
79. Xu, L. and Seki, M. (2020) Recent advances in the detection of base modifications using the Nanopore sequencer. *J. Hum. Genet.* 65, 25–33
80. Ouldali, H. *et al.* (2020) Electrical recognition of the twenty proteinogenic amino acids using an aerolysin nanopore. *Nat. Biotechnol.* 38, 176–181
81. Pugliese, K.M. *et al.* (2015) Processive incorporation of deoxynucleoside triphosphate analogs by single-molecule DNA polymerase I (Klenow fragment) nanocircuits. *J. Am. Chem. Soc.* 137, 9587–9594
82. Merriman B.L. and Mola P.W. Roswell Biotechnologies, Inc. Biomolecular sensors and methods, WO2016210386A1
83. Hajian, R. *et al.* (2019) Detection of unamplified target genes via CRISPR–Cas9 immobilized on a graphene field-effect transistor. *Nat. Biomed. Eng.* 3, 427–437
84. Mousavi, P.S. *et al.* (2020) A multiplexed, electrochemical interface for gene-circuit-based sensors. *Nat. Chem.* 12, 48–55
85. Mikutis, G. *et al.* (2018) Silica-encapsulated DNA-based tracers for aquifer characterization. *Environ. Sci. Technol.* 52, 12142–12152
86. Koch, J. *et al.* (2020) A DNA-of-things storage architecture to create materials with embedded memory. *Nat. Biotechnol.* 38, 39–43
87. Doroschak, K. *et al.* (2020) Rapid and robust assembly and decoding of molecular tags with DNA-based nanopore signatures. *Nat. Commun.* 11, 5454
88. Qian, L. *et al.* (2011) Neural network computation with DNA strand displacement cascades. *Nature* 475, 368–372
89. Currin, A. *et al.* (2017) Computing exponentially faster: implementing a non-deterministic universal Turing machine using DNA. *J. R. Soc. Interface* 14, 20160990
90. Song, X. and Reif, J. (2019) Nucleic acid databases and molecular-scale computing. *ACS Nano* 13, 6256–6268
91. Song, Y. *et al.* (2018) DNA multi-bit non-volatile memory and bit-shifting operations using addressable electrode arrays and electric field-induced hybridization. *Nat. Commun.* 9, 281
92. Bee, C. *et al.* (2020) Content-based similarity search in large-scale DNA data storage systems. *bioRxiv* Published online April 6, 2020. https://doi.org/10.1101/2020.05.25.115477
93. Wang, B. *et al.* (2019) SIMDIIDNA: single instruction, multiple data computation with DNA strand displacement cascades. In *DNA Computing and Molecular Programming. DNA 2019. Lecture Notes in Computer Science, Vol. 11648* (Thachuk, C. and Liu, Y., eds), pp. 219–235, Springer
94. Carmean, D. *et al.* (2019) DNA data storage and hybrid molecular–electronic computing. *Proc. IEEE* 107, 63–72
95. Joesaar, A. *et al.* (2019) DNA-based communication in populations of synthetic protocells. *Nat. Nanotechnol.* 14, 369–378

96. Tavella, F. *et al.* (2019) DNA molecular storage system: transferring digitally encoded information through bacterial nanonetworks. *IEEE Trans. Emerg. Top. Comput.* Published online August 8, 2019. https://doi.org/10.1109/TETC.2019.2932685
97. Jensen, M.A. and Davis, R.W. (2018) Template-independent enzymatic oligonucleotide synthesis (TiEOS): its history, prospects, and challenges. *Biochemistry* 57, 1821–1832
98. Sarac, I. and Hollenstein, M. (2019) Terminal deoxynucleotidyl transferase in the synthesis and modification of nucleic acids. *Chembiochem* 20, 860–871
99. Lee, H.H. *et al.* (2019) Terminator-free template-independent enzymatic DNA synthesis for digital information storage. *Nat. Commun.* 10, 2383
100. Lee, H. *et al.* (2020) Photon-directed multiplexed enzymatic DNA synthesis for molecular digital data storage. *Nat. Commun.* 11, 5246
101. Hölz, K. *et al.* (2019) Multi-level patterning nucleic acid photolithography. *Nat. Commun.* 10, 3805
102. Palluk, S. *et al.* (2018) De novo DNA synthesis using polymerase-nucleotide conjugates. *Nat. Biotechnol.* 36, 645–650
103. Kishi, J.Y. *et al.* (2018) Programmable autonomous synthesis of single-stranded DNA. *Nat. Chem.* 10, 155–164
104. Minev, D. *et al.* (2019) Rapid in vitro production of single-stranded DNA. *Nucleic Acids Res.* 47, 11956–11962
105. Hoff, K. *et al.* (2020) Enzymatic synthesis of designer DNA using cyclic reversible termination and a universal template. *ACS Synth. Biol.* 9, 283–293
106. Farzadfard, F. and Lu, T.K. (2018) Emerging applications for DNA writers and molecular recorders. *Science* 361, 870–875
107. Ishiguro, S. *et al.* (2019) DNA event recorders send past information of cells to the time of observation. *Curr. Opin. Chem. Biol.* 52, 54–62
108. Tang, W. and Liu, D.R. (2018) Rewritable multi-event analog recording in bacterial and mammalian cells. *Science* 360, eaap8992
109. Farzadfard, F. *et al.* (2019) Single-nucleotide-resolution computing and memory in living cells. *Mol. Cell* 75, 769–780 e4
110. Schmidt, F. *et al.* (2018) Transcriptional recording by CRISPR spacer acquisition from RNA. *Nature* 562, 380–385
111. Tanna, T. *et al.* (2020) Recording transcriptional histories using Record-seq. *Nat. Protoc.* 15, 513–539