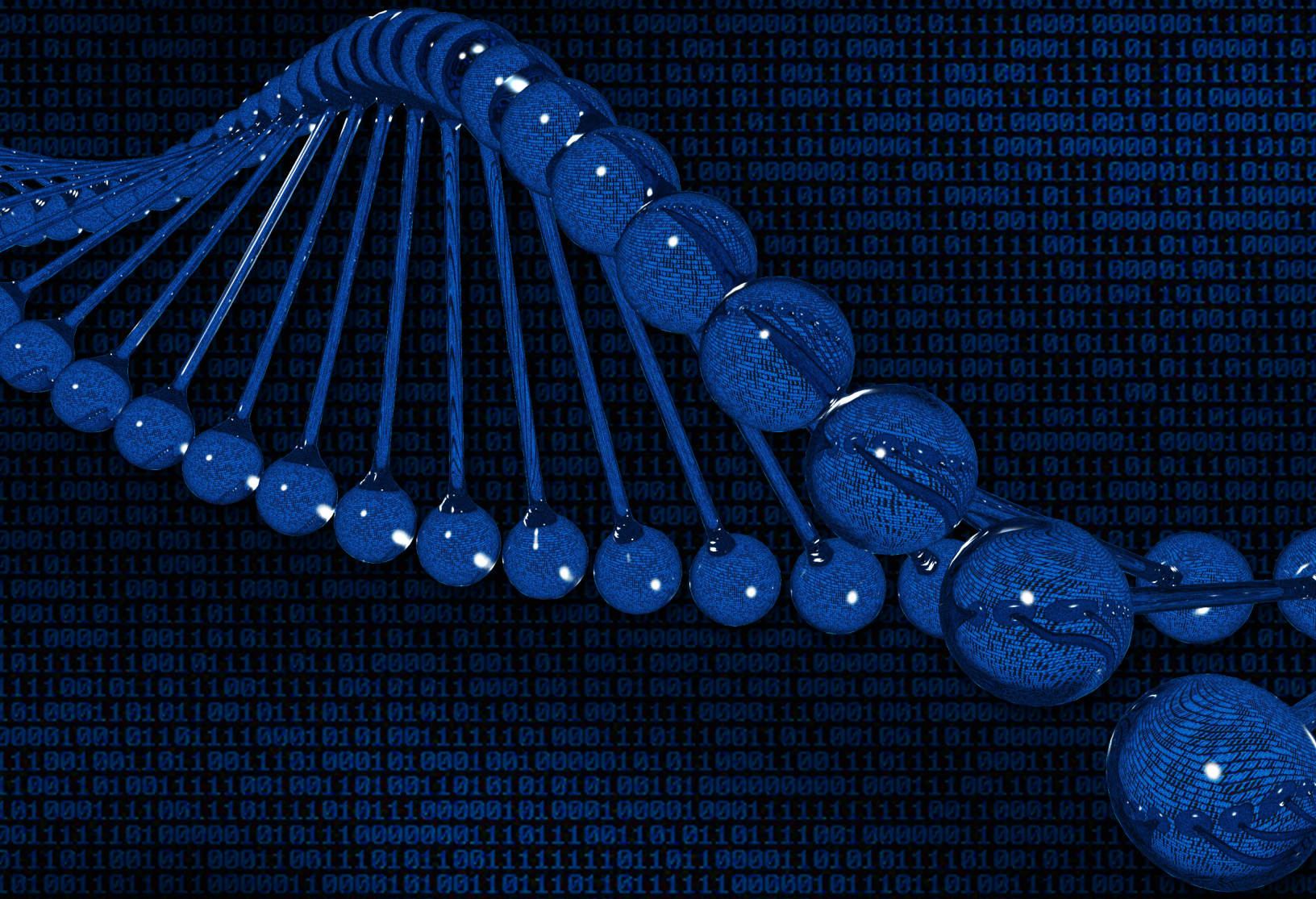
The Future of DNA Data Storage



POTOMAC INSTITUTE FOR POLICY STUDIES

September 2018

A POTOMAC INSTITUTE FOR POLICY STUDIES REPORT

The Future of DNA Data Storage

September 2018

NOTICE: This report is a product of the Potomac Institute for Policy Studies. The conclusions of this report are our own, and do not necessarily represent the views of our sponsors or participants. Many thanks to the Potomac Institute staff and experts who reviewed and provided comments on this report.

Cover image: Alex Taliesen

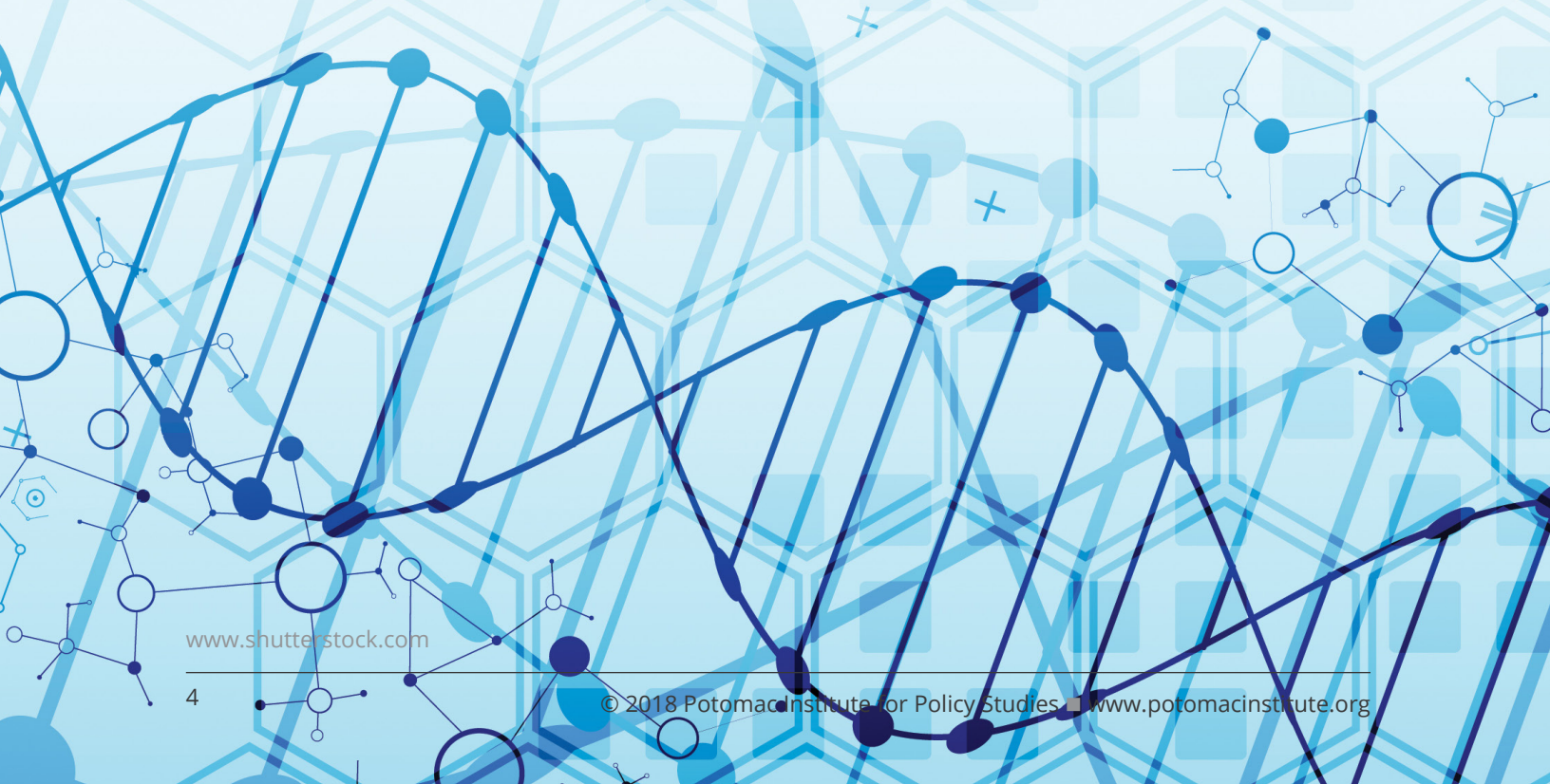
POTOMAC INSTITUTE FOR POLICY STUDIES
901 North Stuart St., Suite 1200 | Arlington, VA 22203 | 703-525-0770 | www.potomacinstitute.org

CONTENTS

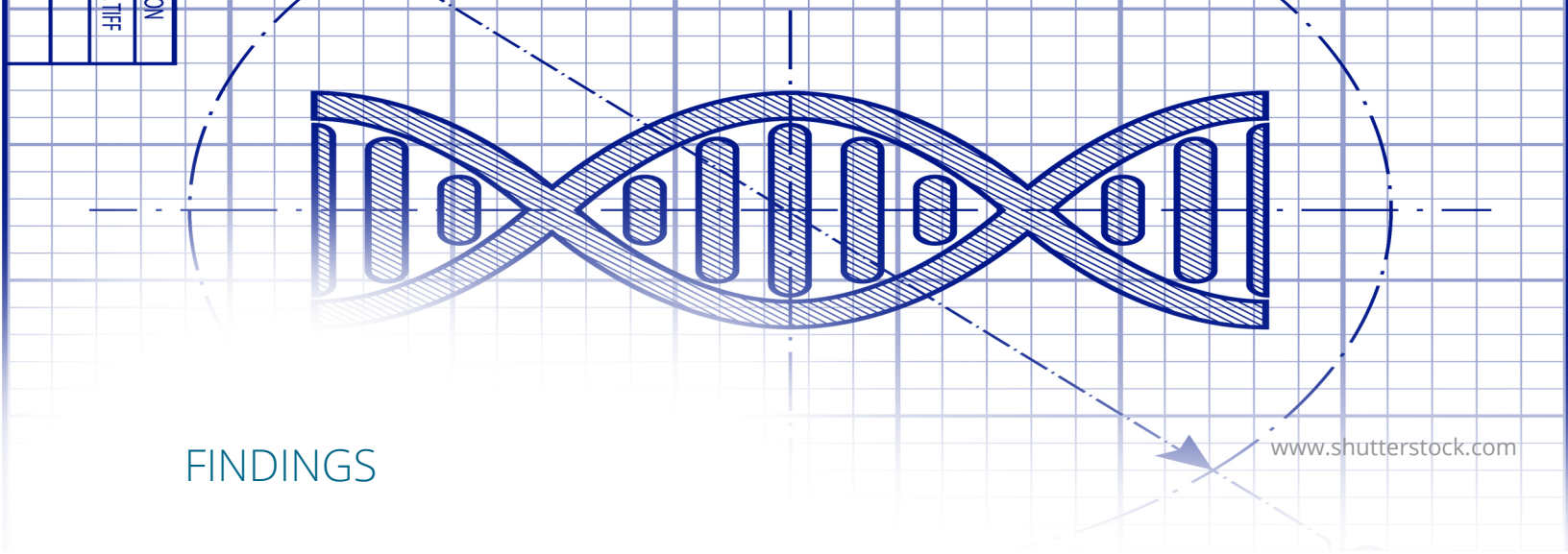
Executive Summary

The demand for digital data storage is currently outpacing the world's storage capabilities, and the gap is widening as the amount of digital data produced grows exponentially. Current technologies will be insufficient to address these challenges. DNA offers an abundant, sustainable, and stable data storage solution, with storage density orders of magnitude better than today's best methods. As a reference, all of the world's data today could theoretically be stored in 1 kg of DNA. Although the science behind DNA data storage has been proven, its commercial viability is currently limited. This is because DNA technology has been developed to support applications in the life sciences industry and not for data storage purposes.

In this study report, we provide findings on the current state of the DNA Data Storage field and provide forecasts for key determinants that will advance this field. The Potomac Institute identified the key technologies required for storing data in DNA, analyzed the state-of-the-art and emerging technologies, and determined necessary steps for advancing the field forward.



www.shutterstock.com



www.shutterstock.com

FINDINGS

1. The science behind storing data in DNA has been proven.

Researchers have demonstrated that DNA is a scalable, random-access and error-free data storage system. DNA is also stable for thousands of years and offers utility in long-term data storage. Advancements in next generation sequencing have enabled rapid and error-free readout of data stored in DNA. As the data storage crisis worsens in the coming years, DNA will be utilized to store vast amounts of data in a highly dense medium.

---

2. Research and development efforts within this field are underway in academia, private industry and the public sector with several cross-sector partnerships.

Table 1 highlights leaders in this field and their primary roles.

TABLE 1. LEADERS IN THE FIELD OF DNA DATA STORAGE.

Sector	Organization	Primary Effort
Private	Microsoft	R&D with the eventual goal of a proto-commercial DNA data storage system
Private	Semiconductor Research Corporation	R&D in advanced alternative data storage solutions
Private	Catalog	Commercialization of DNA data storage technology
Private	Iridia	Commercialization of DNA data storage technology
Public	NSF, IARPA, DARPA, NIH	Funding support to key players in the DNA data storage field
Academia	University of Washington	Research that is pushing towards increasing the volume of information stored in DNA
Academia	Harvard University	R&D of DNA synthesis technology and novel mechanisms of encoding and retrieving information from DNA
Academia	ETH Zurich	Research on storing varying types of files in DNA

3. The first commercial DNA storage company called Catalog is poised to take orders in 2019.

Catalog is building a proprietary DNA data storage machine in partnership with Cambridge Consultants that will synthesize 1TB of data per day at a cost of a few thousand dollars. This will revolutionize our approach to archival data storage and pave the way forward towards further advancements in the field.

4. Currently, the cost of DNA synthesis is limiting advancement within this field.

DNA synthesis methods are still reliant on methods based on organic chemistry, that are three decades old. Our ability to synthesize novel DNA sequences that store data is significantly limited by the inefficiencies of non-biological DNA synthesis methods. These inefficiencies have limited DNA data storage to the domains of research laboratories and significantly limit the data file size that can be stored in DNA.

5. Technology that utilizes engineered biological enzymes to synthesize DNA fragments will radically decrease costs and propel the field forward.

Biology-inspired engineering approaches to synthesizing DNA will be the catalyst that drives down cost of DNA synthesis. Several independent groups are developing DNA synthesis technologies that utilize enzymes to construct novel DNA sequences. This 2nd generation synthesis technology already has commercial developers and is also actively being developed in academia. This technology is the first major breakthrough in DNA synthesis in decades and should lead to significant reductions in costs and facilitate development of new technologies that support storing data in DNA.

6. New technologies for faster reading of data stored in DNA are also needed to advance the field forward.

In conjunction with the development of cheap DNA synthesis technologies, technologies that are compact, fast, and efficient must be developed to allow for easy read-out of data stored in DNA. Development of advanced coding schemes and operating systems tailored to DNA storage devices are also necessary. Random access retrieval of data stored in DNA facilitated by *biologically-inspired* mechanisms must be developed. Furthermore, the density of data stored in DNA can be exponentially increased by exploring the utility of DNA modification techniques.

7. As storing data in DNA becomes cost-effective, this field's technological advantages will revolutionize our approach to data access and computing.

DNA offers compelling advantages over today's best methods for data storage such as orders of magnitude greater storage density, and long-term stability. These are significant competitive advantages that should facilitate the commercial usage of DNA as a data storage medium, especially as DNA synthesizing technology becomes cost-effective. New supporting technologies will be developed in conjunction, such as computers with operating systems tailored for random access retrieval of data stored in DNA. These advancements will usher in a new paradigm for computing with little to no limitations on the volume of data that we can produce, store, and access.

Background

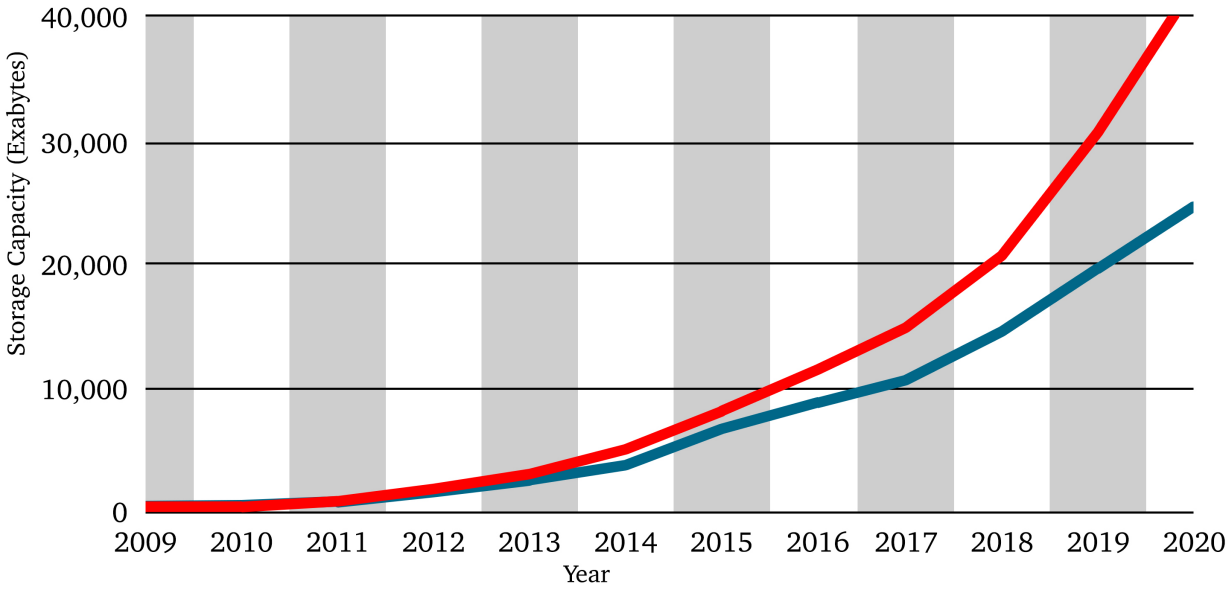


Figure 1
Data storage capacity supply versus demand worldwide (in exabytes), estimated through 2020.[2]

DATA STORAGE CRISIS

The rise of the internet age, and its associated technologies and platforms, has led to an explosion in the amount of digital data being produced. By 2025, humans are expected to produce 160 zettabytes of data *each year*.[1] The demand for digital data storage is currently outpacing our storage capabilities, and the gap is widening as the amount of digital data produced grows exponentially (Figure 1). Modern digital information storage technologies (e.g., flash memory) depend on microelectronics made from silicon. Analysts estimate that storing the world's data in flash in 2040 would require more than 1,000 kilograms (kg) of wafer-grade silicon. The projected supply of single-crystal wafer grade silicon in 2040 is 108 kg.[3] New, sustainable materials will be required to support the world's information technology base and storage of digital data.

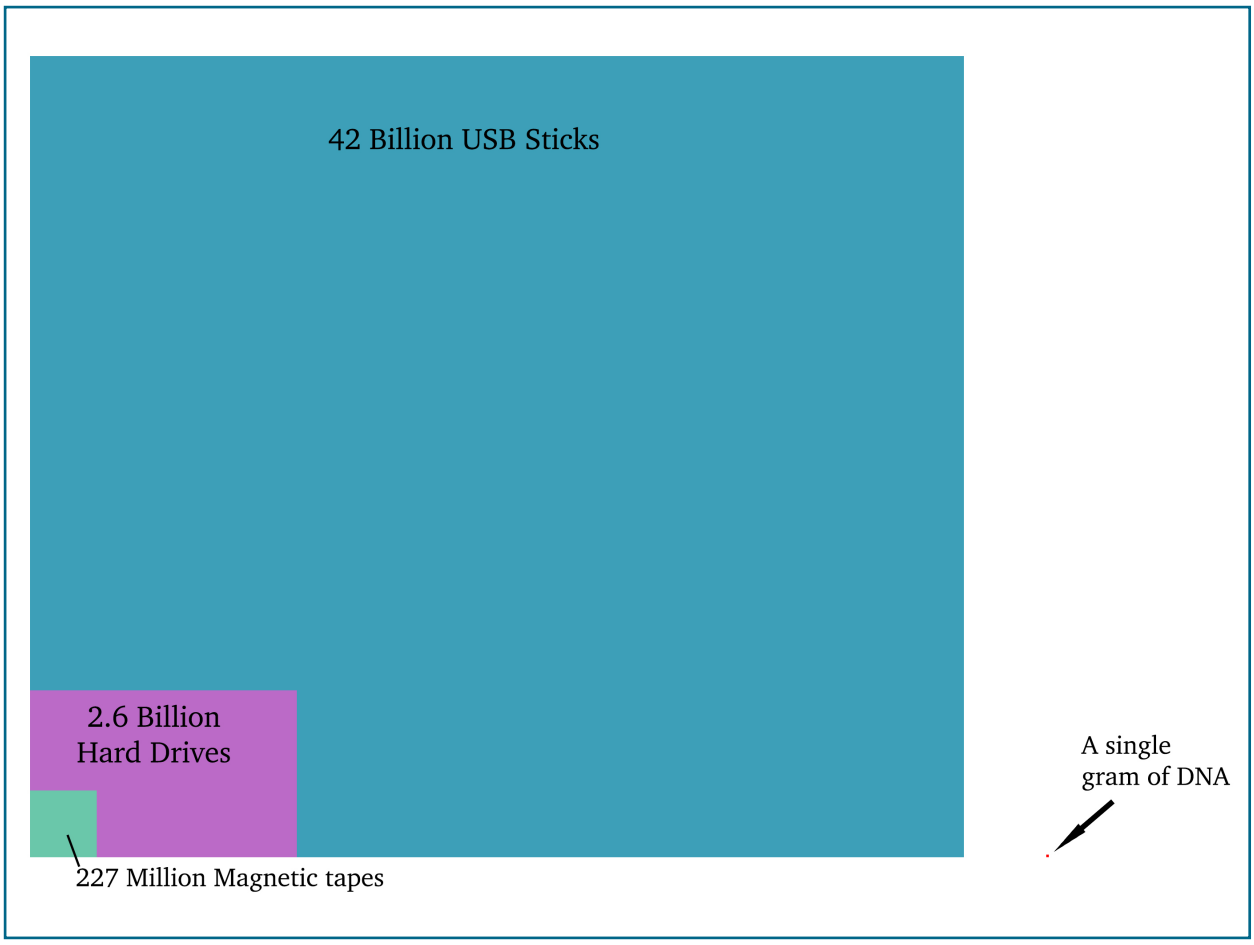


Figure 2
The amount of traditional storage media needed to store 40 zettabytes of data (estimated to be the world's data output in 2020) versus DNA.[5]

To meet the demand of big data storage, companies like Microsoft, IBM, Facebook, and Apple are looking beyond silicon for solutions. The next-generation data storage market is expected to be valued at $144.76B by 2022.[4]

Nature has provided us with a solution to this storage problem: DNA (deoxyribonucleic acid). DNA is the storage system for all instructions/code that governs biological life. Not only is DNA abundant and sustainable, but it offers storage density far greater than our current data storage media (Figure 2) and can be kept and accessed for at least hundreds of thousands of years.

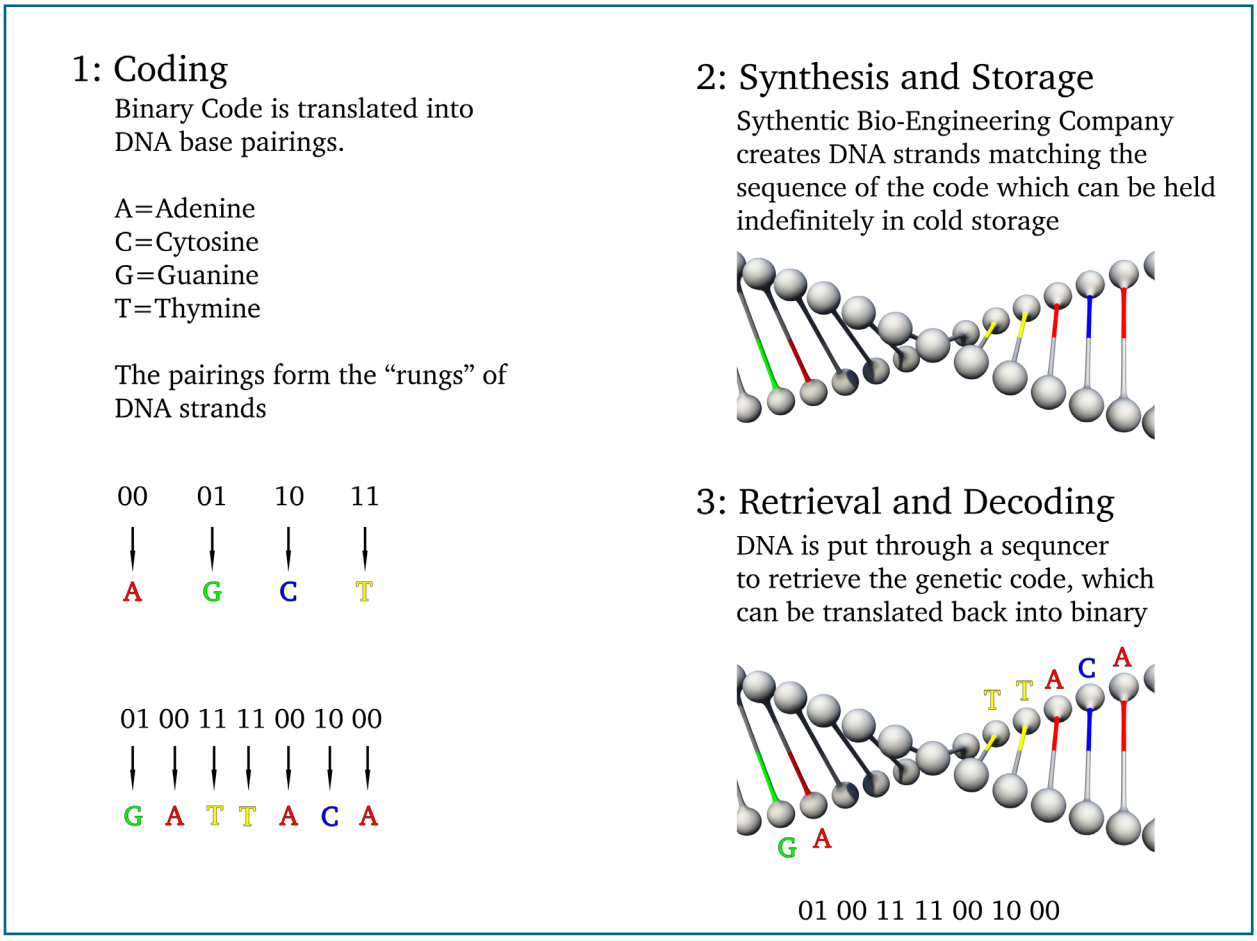


Figure 3
Overview process of encoding and decoding digital data into and from DNA.[6]

DNA AS A DATA STORAGE MEDIUM

DNA data storage is the process of encoding binary data into synthetic, man-made strands of DNA. To store a binary digital file in DNA, the bits (binary digits) are converted from 1s and 0s into the letters A, C, G, and T. These letters represent the four unique nucleotides that make up DNA: adenine, cytosine, guanine, and thymine.

The physical storage medium is a synthesized chain of DNA containing the As, Cs, Gs, and Ts in a sequence corresponding to the order of the bits in the digital file. To recover the data, the chain of DNA is sequenced and the order of As, Cs, Gs, and Ts are decoded back into the original digital sequence (Figure 3).

ADVANTAGES

Storing digital data in DNA has two main advantages over storage in traditional media: storage density and stability.

Storage Density

DNA's information storage density is several orders of magnitude higher than any other known storage technology (Figure 4). For example, flash memory is capable of storing 1 bit of data in ~10 nanometers (nm), at best. DNA is capable of storing 2 bits per 0.34 nm. One kilogram of DNA can store 2×1024 bits; the same amount would require >109 kg of silicon for flash memory. A few tens of kilograms of DNA could meet the world's storage needs for centuries to come.[7]

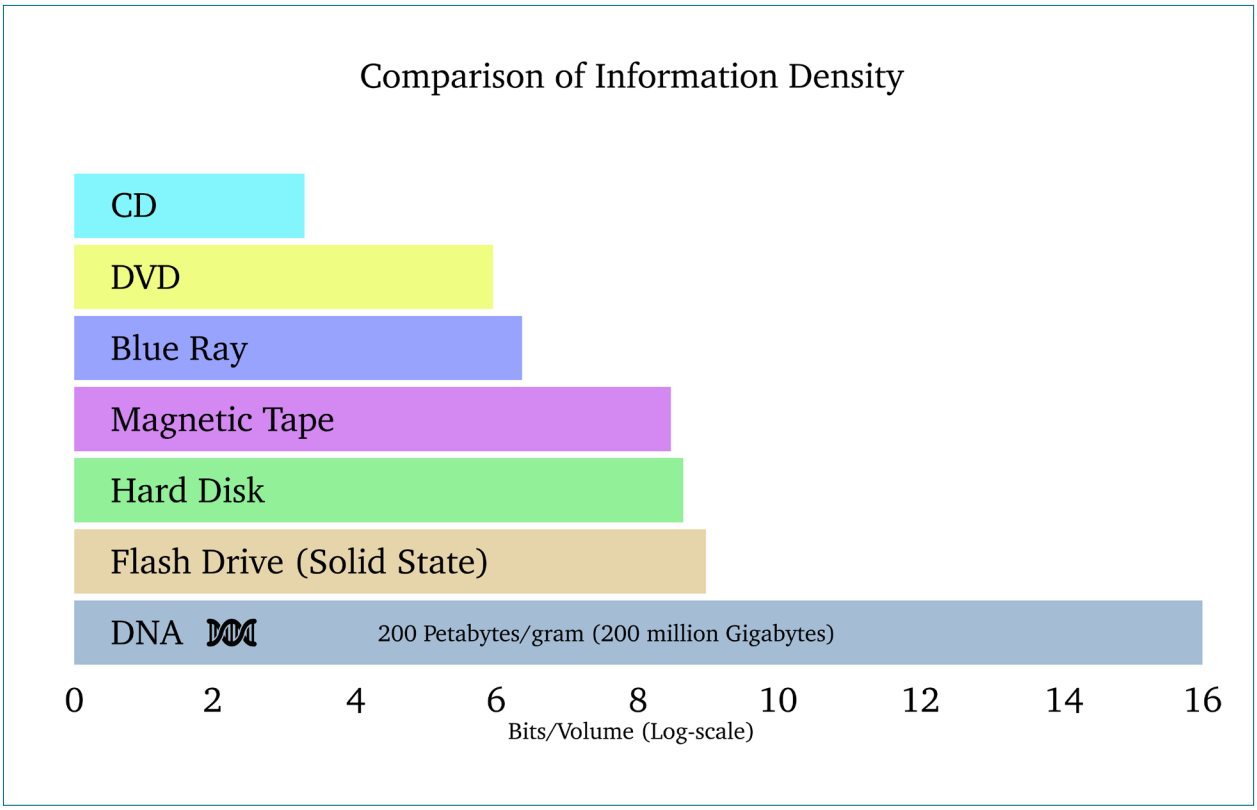


Figure 4
Information storage density of DNA versus traditional media.[8,9]

Stability

Traditional forms of data storage, including magnetic (cassette tapes, floppy disks, hard drives), optical (CDs, DVDs, Blu-ray), and flash (memory sticks, solid state drives), are limited in their long-term usage due to degradation over time and obsolescence of the technology needed to read the data.

DNA is an extremely stable molecule with a half-life of over 500 years.[10] If stored in cold conditions, DNA is capable of remaining intact for hundreds of thousands of years. 700,000-year old horse's DNA, stored in the permafrost, was sequenced in 2013.[11] The error rates of synthetic DNA information storage are on par with those of hard disk drives, with the capacity to be improved further.[12] As compared with other proposed next-generation data storage media, DNA is abundant and scalable (Figure 5).

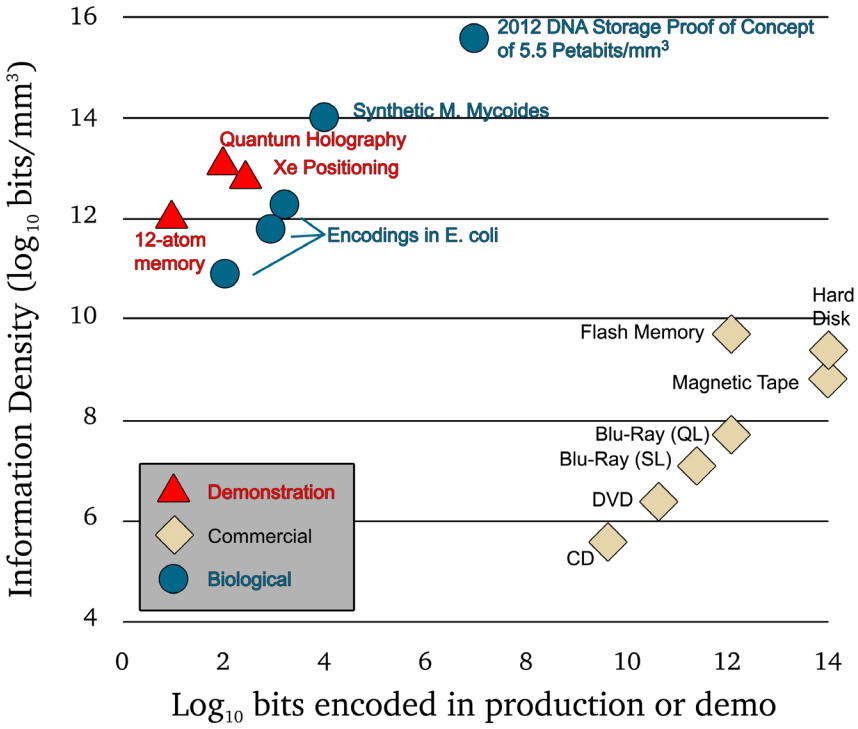


Figure 5
DNA-based information storage density as compared to other technologies, plotted as information density (log10 of bits/mm3) versus scalability, as of 2012.[13]

HISTORY

The proof of principle of storing digital data in DNA was first shown in 1988 by artist Joe Davis in collaboration with Harvard researchers. Davis stored the image of a runic symbol representing 35 bits of data by representing light and dark pixels as binary 1s and 0s, and encoding it in 28 base pairs within the DNA of *E. coli*.[14] As of February 2018, researchers have stored 0.4GB (~3,200,000,000 bits) of data in DNA.[15] The complexity of data stored in DNA has also increased from simple text to high-definition music videos, entire databases, MPEG, JPG and PDF files, and malware (Table 2. History of Data Storage in DNA).

TABLE 2. HISTORY OF DATA STORAGE IN DNA.

Date	Size (Megabytes)	Group	Description
1988[16]	.000004 MB (35 bits)	Harvard University	Encoded image
1999[17]	.00009 MB	Ars Electronica	Encoded text from Genesis
2003[18]	.0001 MB	Pacific Northwest National Laboratory	Part of "It's a Small World"
2005[19]	.0001 MB	DNA2.0 (Now ATUM)	Poem "Tomten"
2009[20]	.0002 MB	University of Toronto	Text, music, image
2010[21]	.0009 MB	The J. Craig Venter Institute	Watermarking of synthetic genome
August 2012[22]	.66 MB	Harvard University	Book (53,426 words, 11 JPG images) and JavaScript program
February 2013[23]	.74 MB	European Bioinformatics Institute	Shakespeare's sonnets, 26-second audio clip of an MLK speech, Watson and Crick's paper on the structure of DNA
February 2015[24]	.08 MB	ETH Zurich	Swiss Federal Charter of 1291, Archimedes Palimpsest
April 2016[25]	.15 MB	Microsoft, University of Washington	Image files
June 2016[26]	22 MB	Harvard University, Technicolor	MPEG compressed movie sequence
March 2017[27]	2.14 MB	New York Genome Center, Columbia University	Graphical operating system, movie, PDF, image, text, and malware
March 2017[28]	200 MB	Microsoft, University of Washington	Universal Declaration of Human Rights (in 100 languages), high-definition music video, database of seeds stored in Svalbard Global Seed Vault
February 2018[29]	400 MB	Microsoft, University of Washington	Unspecified
April 2018[30]	15 MB	ETH Zurich, Rice University	Music album

Current State of DNA Data Storage

Interest in DNA as a data storage solution has garnered interest since the late 1980s. Only recently however have researchers made significant gains in this field, storing a maximum of 400 MB of data in DNA. The major limitation on progress are the staggeringly high costs of DNA synthesis, as the primary mode of DNA synthesis technology has not advanced significantly in the past few decades. Advancements in DNA synthesis technologies that lower the cost of gene synthesis are necessary before DNA can serve as an option for data storage.

TECHNOLOGY OF DNA DATA STORAGE

DNA data storage involves two main processes: *writing* the code via DNA synthesis and *reading* the code via DNA sequencing. Trends and technical milestones in these two processes will determine our ability to use DNA as a storage solution in the short and long term.

Writing Data to DNA

The primary methodology in use today for synthesizing DNA is based on phosphoramidite chemistry and was initially developed more than three decades ago. New tools and technologies in support of chemistry-based DNA synthesis have been developed over the years, such as microfluidic systems, ink-jet printing technology, digital photolithography, and electrochemistry. These methods have tradeoffs in terms of the length of the sequence that can be generated, error rates, yield/volume of product, production time/speed, and cost.

Figure 6
Applied Biosystems' ABI 380A Three Column DNA Synthesizer. Launched in 1983. Image courtesy of American Laboratory Trading.[31]

The first commercial DNA synthesizer employing phosphoramidite chemistry was introduced in 1983 by Applied Biosystems, Incorporated (ABI) (Figure 6).[32] Synthetic DNA synthesis has traditionally been based on the solid-phase synthesis of short fragments of nucleic acids sequences called oligonucleotides. The chemical synthesis of oligonucleotides is carried out by the stepwise addition of a nucleotide to the 5'-terminus of the growing chain until the desired sequence is assembled. Oligonucleotide sequences are then connected by annealing and ligation or polymerase reactions. This approach allows for the *de novo* synthesis of DNA chains, without the need for template DNA.

DNA synthesis today is primarily done using microarray-based oligonucleotide synthesis technology, which was initially developed to synthesize oligonucleotides attached to a microchip surface using a modified phosphoramidite synthesis process. DNA synthesis on microarrays produces large volumes of short oligo sequences that are subsequently stitched together to produce longer gene sequences (Figure 7). Microarray-based oligonucleotide synthesis is an efficient method that generates large quantities of unique sequences in parallel. Compared to column-based oligos, microarray oligos are two-to-four orders of magnitude cheaper.

Affymetrix, an early pioneer of this technology, used standard mask-based photolithography techniques to control the spatially-separated synthesis of oligonucleotides on the microchip surface. This technology later evolved beyond the use of photolithography masks, with NimbleGen and LC Sciences' use of programmable micromirror devices. The current industry leader in this field is Twist Biosciences. Their proprietary technology is a high-throughput silicon platform that miniaturizes the chemistry necessary for DNA synthesis, reducing reaction volumes by a factor of 1,000,000 while increasing throughput by a factor of 1,000. Twist Bioscience's method is capable of synthesizing 9,600 genes on a single silicon chip at full scale.[33] In comparison, traditional microarray-based synthesis methods produce a single gene in the same physical space using a 96-well plate (Figures 7-8).

The main limitation of microarray-based synthesizing technology is the inefficiency of the chemical synthesis methods. These methods require synthesis of small DNA fragments in parallel that must be assembled in a separate step, making the process tedious, costly and prone to error as the DNA strands get longer. There can also be issues with the quality of synthesized oligonucleotides as spontaneous depurination of oligonucleotides can occur on-chip. Misalignment of droplets, which facilitate reactions on the silica chip, can also result in low quality synthesis.

On the whole, there haven't been sufficient market pressures to drive innovation within this space that drives down costs by orders of magnitude. However, there are several indications that activity in the DNA synthesis space is ramping up, driven primarily by the nascent field of Synthetic Biology. Twist Bioscience, an industry leader in producing synthetic DNA, has signed several large-volume synthetic DNA purchase agreements within the last three years with companies like Gingko Bioworks and Microsoft (Table 3). The April 2016 agreement between Twist Bioscience and Microsoft resulted in a record 200 MB of data stored using DNA, and the April 2017 order has brought the record to 400 MB.

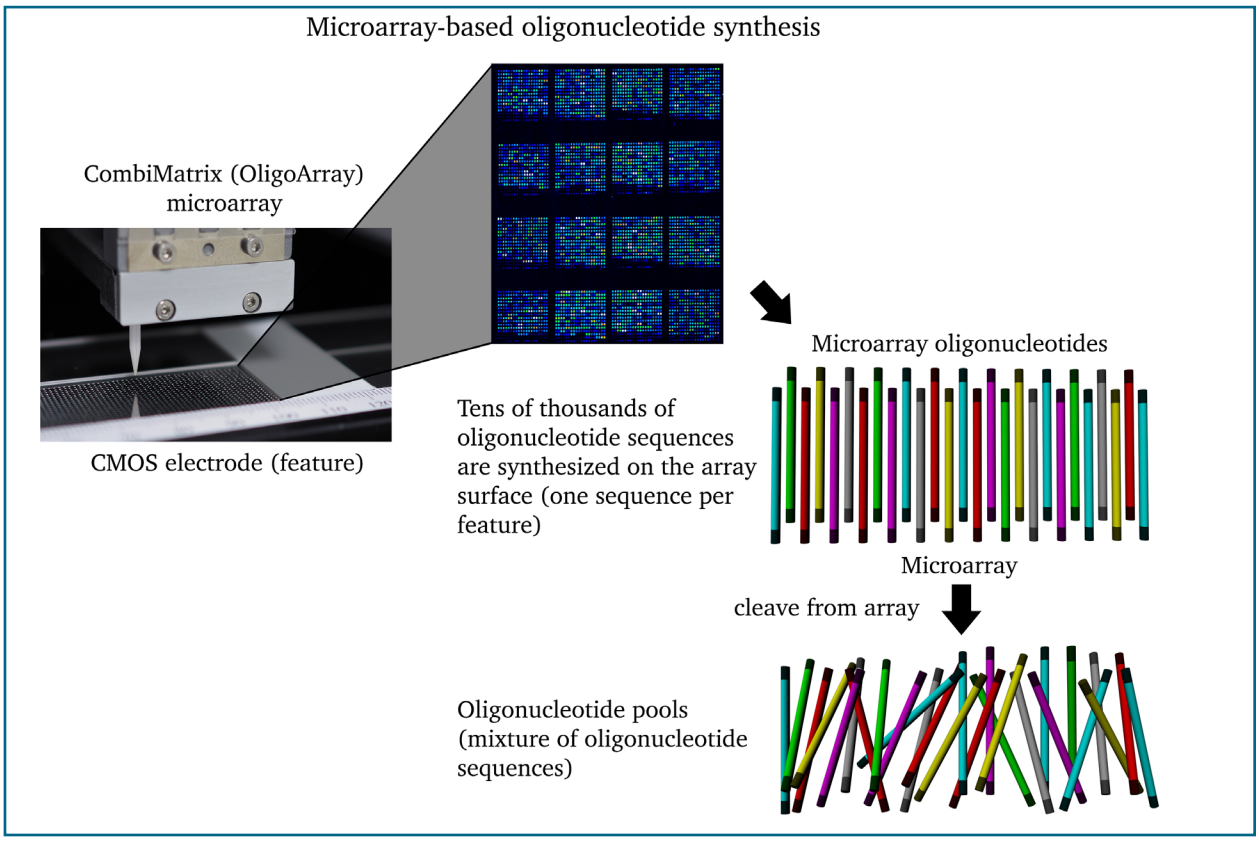


Figure 7
Generalized scheme of microarray-based oligonucleotide synthesis.[34,35]

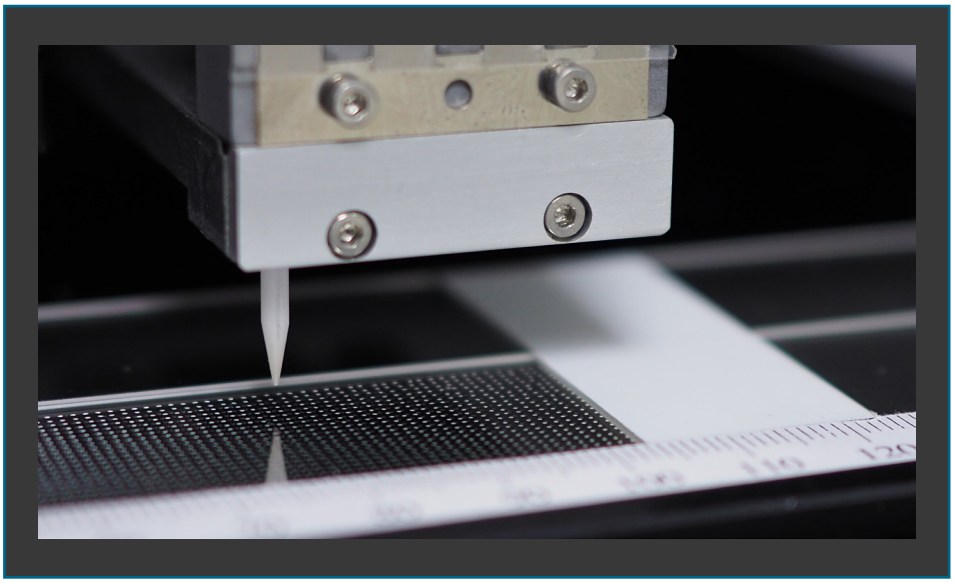


Figure 8
Microarray printing on glass slide, non-contact printing system, selective focus on tip and spots on glass slide. (courtesy of Shutterstock).

TABLE 3. RECENT HISTORY OF TWIST BIOSCIENCE'S SUPPLY AGREEMENTS FOR SYNTHETIC DNA.

Date	Buyer	Purchase Volume	Purpose
November 2015	Gingko Bioworks[36]	100 million base pairs (bp)	Synthetic Biology
April 2016	Microsoft, University of Washington [37]	10 million bp	Data Storage
April 2017	Microsoft, University of Washington	10 million bp	Data Storage
June 2017	BioBricks Foundation[38]	10,000 synthetic genes (~10 million to 30 million bp)	Synthetic Biology
October 2017	Gingko Bioworks[39]	1 billion bp	Synthetic Biology

Rob Carlson, a leading voice on the economics of synthetic biology, has suggested that the utility of DNA as a storage medium could serve as a catalyst for demand and innovation. There is ample research showing the utility and feasibility of using DNA as a dense data storage medium. As the data storage crisis intensifies, so will industry demand for a dense, long-term solution. Carlson states:

> The scale of the demand for DNA storage, and the price at which it must operate, will completely alter the economics of reading and writing genetic information, in the process marginalizing the use by existing multibillion-dollar biotech markets while at the same time massively expanding capabilities to reprogram life.[40]

For now, Microsoft is continuing to push forward with its research on DNA storage. The company plans to have a proto-commercial system storing data in DNA at one of its data centers by the end of this decade.[41] The device is expected to be the size of a 1970s Xerox copier and will have a boutique application. The cost of encoding data has also decreased over the years. In 2012, researchers encoded 0.74 MB of data in DNA at a cost of $12,400 per MB. In 2017, the price of encoding data to DNA had fallen to $3,500 per MB.[42]

Growth of the synthetic biology market, which will increasingly require greater quantities of synthetic DNA, is likely to bring down costs further. The number of global synthetic biology companies has more than doubled this decade, from 194 in 2010 to 411 as of 2016 (Figure 9).[43] Investment in this industry has also been increasing over the last decade, with over $1B invested in synthetic biology companies in 2016 (Figure 10).[44]

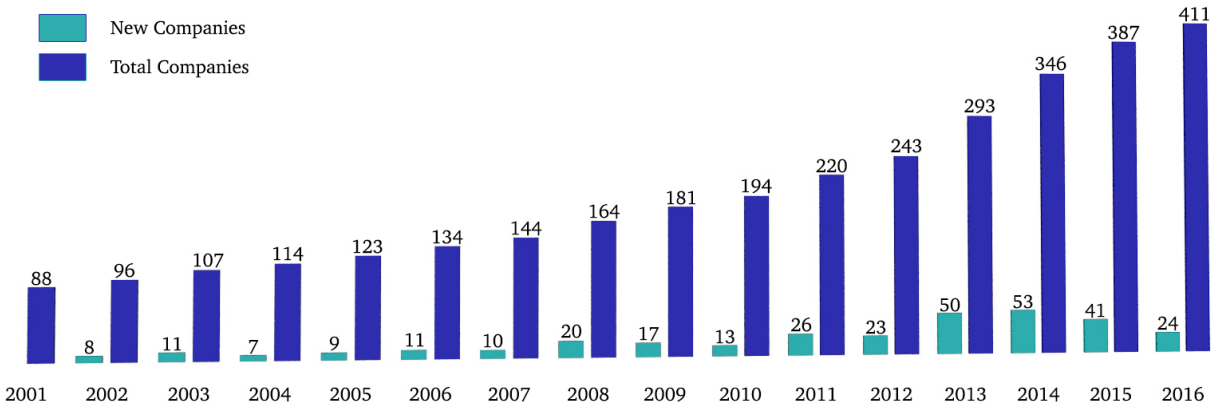


Figure 9
Number of global synthetic biology companies, 2001–2016.[45]

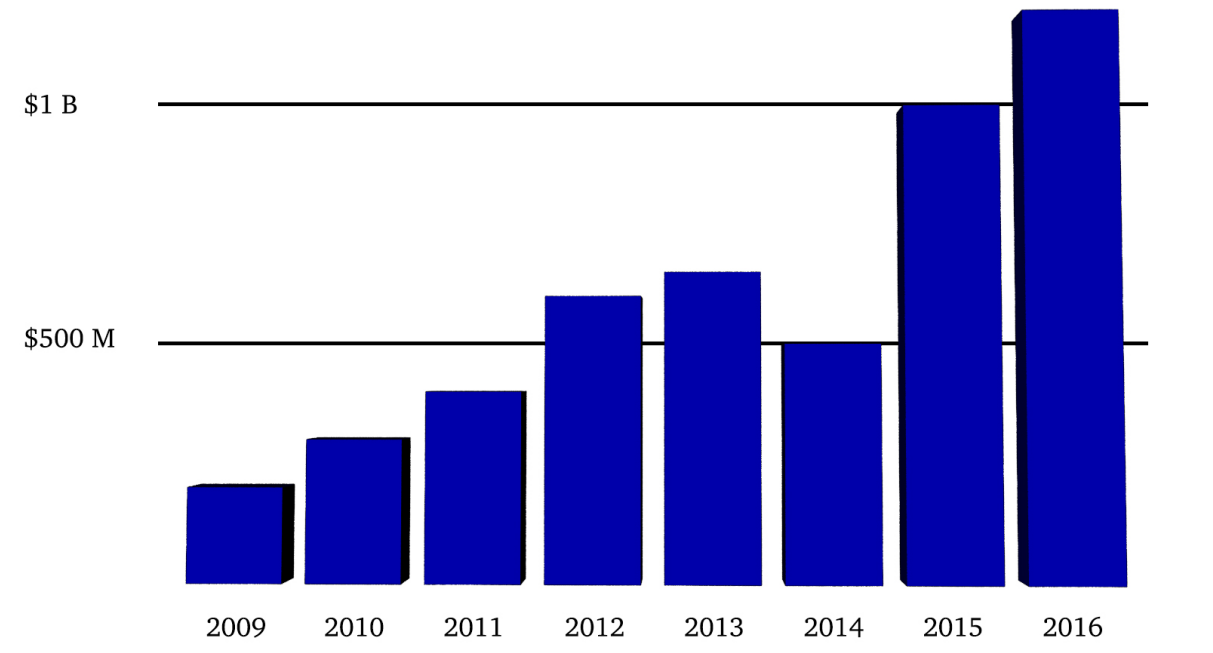


Figure 10
Global synthetic biology investment, 2009-2016.[46]

Reading Data from DNA

At the completion of the Human Genome Project in 2001, the cost to sequence a single genome was roughly ~$100M.[47] By the end of 2007, a major technological shift in sequencing technology triggered an exponential decline in sequencing costs. Previously, sequencing had been done using the "Sanger" methodology, which was costly and limited in its ability to sequence long strands of DNA. The drastic post-2007 cost reductions in genome sequencing can be attributed to advancements that are broadly defined under the label of "next generation sequencing" (NGS).[48] NGS revolutionized DNA sequencing technology as it created a cell-free system of cloning DNA fragments, allowing millions of sequencing reactions to run in parallel and the identification of base pairs in real-time. By 2017, the cost to sequence an entire genome had dropped to $1K.

There are currently numerous biotechnology companies offering sequencing services using various NGS technologies which vary in their methodologies and practical usage. Below are descriptions of three industry leaders who use the most cutting-edge sequencing technologies: Illumina, Pacific Biosciences and Oxford Nanopore.

1. Illumina is the industry leader in DNA sequencing technology, offering the most accurate NGS sequencing platform available. Illumina sequencers perform what is called "short read sequencing."[49] Small clusters of denatured (single-stranded) DNA (200-600 bp) are sequenced in parallel through a constant denaturation/synthesis process. A complementary sequence is synthesized on denatured DNA using base pairs that are given a unique fluorescent tag. A polymerase (DNA duplication enzyme) is used to create the complementary sequence on the single-stranded DNA fragments. The order of fluorescent labeling can be detected by the sequencing machine to deduce the sequence of DNA. Data analysis tools are used post-sequencing to arrange the short DNA fragments in the correct order.

2. Pacific Biosciences (PacBio) sequencers perform "long read sequencing" using proprietary technology called single molecule real-time sequencing.[50] PacBio technology can sequence large fragments of DNA, scaling up to 14,000 base pairs in a single read. The sequencing is conducted inside a "nano-well" using optical technology. It works by denaturing the double-stranded DNA and using a polymerase to rebuild the double-strand using nucleotides with unique fluorescent labels. As each fluorescent nucleotide is attached to the DNA, the emitted light signature is detected in the nano-well, providing a read out of the DNA sequence.

3. Oxford Nanopore's sequencing technologies also performs long read sequencing, with the added benefit of real-time results. Their sequencers utilize protein "nanopores" that are inserted into an electrically-resistant membrane made from synthetic polymers. A potential is applied across the membrane, resulting in a current flowing through the pore. As DNA is fed through the nanopore, each nucleotide uniquely disrupts the flow of current. Deviations in the nanopore current are detected, allowing

the DNA sequence to be deduced in real-time. Oxford Nanopore's also offers the only portable DNA sequencing device. The MinION is a USB thumb drive-sized sequencer that can be plugged into the USB port of computers and laptops (Figure 11). The sequencer has been used outside the lab in environments such as mountains, jungles, the arctic, and the International Space Station. However, a current drawback of this sequencing technology is that it offers less accurate reads than its competitors. Oxford Nanopore is also developing the SmidgION, which will allow real-time DNA sequencing using smartphones and other mobile, low power devices.[51]

Figure 11
MinION is the only portable real-time device for DNA and RNA sequencing. Each consumable flow cell can now generate 10–20 Gb of DNA sequence data. Image courtesy of NASA.[52]

KEY PLAYERS IN DNA DATA STORAGE

Research into DNA data storage is largely performed in academic labs, with the work being funded by the National Science Foundation (NSF), the National Institutes of Health (NIH), the Intelligence Advanced Research Projects Activity (IARPA), and the Defense Advanced Research Projects Agency (DARPA).[53-56] Although some news articles claim major industry players, such as Google, Facebook, and Apple, are interested in pursuing DNA as a viable storage media, only Microsoft is leading a known R&D effort in this area. Start-ups offering DNA storage services have also begun to surface this year. Below, the top players from each sector are highlighted.

Academia

University of Washington

The Molecular Information Systems Lab (MISL) at the University of Washington (UW), led by Luis Ceze, is a partnership between UW's Departments of Computer Science and Electrical Engineering, and Microsoft Research. The program brings together faculty, students and research scientists with expertise in computer architecture, programming languages, synthetic biology, and biochemistry. The UW/Microsoft team holds the current record for the amount of data stored in DNA (~400 MB).[57]

Harvard University

Efforts at Harvard University's Wyss Institute are led by geneticist George Church. The Church lab made a huge breakthrough in the field of DNA data storage in 2012, by developing a novel encoding scheme that allowed for synthesis of thousands of oligonucleotides on a single DNA microchip, and storing 0.66 MB of data (a book, JPG images, and a JavaScript program).[58] Prior to this work, the largest DNA data storage project (by the J. Craig Venter Institute) had encoded 0.0009 MB of data.[59] George Church and his collaborators have also used CRISPR to encode data sequentially into the genomes of living bacteria.[60]

Columbia University

Research at Columbia University on DNA storage was led by Yaniv Erlich, an associate professor of computer science. Erlich pioneered a technology, "DNA Fountain," that optimized the storage density of DNA within 80% of its theoretical limit.[61] In 2017, Erlich took a leave of absence from Columbia University to join MyHeritage as Chief Science Officer.[62]

University of Illinois, Urbana-Champaign

Olgica Milenkovic, professor of Electrical and Computer Engineering, leads a $1.5M effort to produce new DNA-based storage nanoscale devices using chimeric DNA, a hybrid molecule made from two different sources. SemiSynBio, a three-year research project funded by a partnership between NSF and the Semiconductor Research Corporation (SRC), aims to design a method to read, write, and store data in a more cost-effective way than current DNA storage techniques. Milenkovic also recently received a three-year, $2.5M grant from DARPA to combine synthetic DNA with computing.[63]

ETH Zurich

ETH Zurich's DNA data storage research team, led by Robert Grass, has pioneered a technique for storing DNA *in silica*, creating a "synthetic fossil" that can preserve DNA and the data it stores for thousands of years, even in extreme conditions.[64]

Research Consortium

Semiconductor Research Corporation

SRC is a North Carolina-based, nonprofit research consortium focused on microelectronics and semiconductor research. Concerned with the upcoming age of zettabyte data generation, it is leading research efforts focused on the advancement and use of alternative storage technologies such as DNA storage, 5D optical storage, magnetic storage, and cryogenics.[65] SRC conducted a workshop co-hosted with IARPA on DNA-based Massive Information Storage in April 2016.[66]

Industry

Microsoft Research

The most active and visible industry leader in DNA storage technology is Microsoft Research. As a research partner in the University of Washington's Molecular Information Systems Lab (MISL), Microsoft is pushing the boundaries of DNA data storage capabilities and has demonstrated significant leaps in the past several years in overall storage capacity.[67]

Micron Technology

Top memory manufacturer Micron Technology is also funding DNA digital storage research in collaboration with researchers at Boise State University, as well as through research consortia, to explore the viability of DNA for future storage needs.

Other

News reports have suggested that Apple, Facebook, Google, Intel, and IBM, are also exploring DNA as a data storage medium.[68,69] However, their research efforts or technological advancements within this field have not been reported to date.

Start-ups

Catalog

Catalog is a Boston-based start-up that raised $9M in funding in June 2018 in the hopes of providing commercial DNA data storage services. Catalog intends to bypass the DNA synthesis process. Their methodology utilizes a large collection of premade molecules to encode data in DNA. Using their technology, they have stored around 1 KB of data which includes literary works from Douglas Adams and Robert Frost in DNA. Catalog's Chief Science Officer, Devin Leake, was most recently Head of DNA Synthesis at Ginkgo Bioworks. Prior to Ginkgo, he held positions at Gen9, Thermo Fisher and Dharmacon.[70]

Iridia

Formerly known as Dodo OmniData, Iridia is a start-up based out of San Diego working on developing new methods for storing data in DNA. Their technology looks to combine DNA polymer synthesis technology, electronic nano-switches and semiconductor fabrication technologies, towards a highly-parallel format that enables an array

of nanomodules to store data at exceptionally high density.[71] An investor for the start-up is Jay Flately, the founder and executive chairman of Illumina, the world leader in DNA sequencing technology.[72]

Helixworks

Founded in 2015, this Irish startup took in an undisclosed amount of seed funding last year to turn their proprietary DNA data storage technology into a commercial product. Helixworks is set to offer the world's first commercially available DNA storage drive, which will be available to purchase on Amazon for $199, shortly. Helixworks' DNADrive offers "512 KB of data storage in specially encoded DNA, encapsulated specially in a custom gold pill, with a guaranteed lifetime (under normal conditions) of 10 years, and a potential shelf-life of thousands."[73]

Government

U.S.

DARPA

DARPA's Molecular Informatics Program, announced in March 2017, aims to create a new paradigm for data storage, retrieval, and processing using encoded molecules. The program awarded ~$15M in funding to researchers at Harvard University, Brown University, the University of Illinois, and the University of Washington.[74]

IARPA

IARPA's Molecular Information Storage Technology (MIST) program aims to develop technology that can write 1 TB of data and read 10 TB of data per day using DNA.[75] Furthermore, it aims to scale this technology to the exabyte scale with a viable path to commercialization within 10 years. The program is slated to begin in October 2018.

NIH

NIH funds individual researchers working on DNA data storage. The recent demonstration by Harvard's George Church of encoding a small video file in the bacterial genome was funded by NIH.[76]

NSF

The National Science Foundation solicited proposals for DNA information storage research in 2017. Available grants from the NSF totaled up to $4M.56.

Foreign

European Bioinformatics Institute

The European Union has provided funding to the European Bioinformatics Institute to conduct research on DNA as a data storage medium.[77] EBI's efforts, led by Nick Goldman, date back to 2013 and have pioneered error-correction methods.[78]

Forecast of DNA Data Storage

The science behind DNA data storage is settled; DNA has been proven capable of providing a scalable, random-access, and error-free data storage system. However, technical advancements in encoding data in DNA and reading it back out, decreased costs in synthesizing DNA and new technologies designed for storing data in DNA are required for the field to exit the lab and become viable storage system by industry or government actors.

Below, two forecasts are provided: 1) a forecast of DNA synthesis costs, and 2) a forecast that highlights the steps necessary for DNA-based data storage to become technologically competitive with traditional methods used today.

DNA SYNTHESIS COST FORECAST

DNA synthesis costs are currently the primary factor limiting wide adoption of DNA as a data storage system. The cost of synthetically producing DNA oligos has decreased by 96% since 2004, down to the current price of USD $0.07 per base pair.[79-82] This decrease has been driven by the development of microarray-based synthesis methods and use of oligo pools in recent years.

While costs for DNA synthesis have been decreasing over the years, they have not followed a similar trend as DNA sequencing that exponentially decreased in cost and increased in efficiency with the development of next generation sequencing. The lack of significant cost reduction for DNA synthesis is a consequence of stagnant technological innovation in the field. The standard methods of DNA synthesis still functionally rely on 30-year old chemistry-based methods.

Stagnant innovation in DNA synthesis can be pinpointed to the lack-luster market demand for synthetic DNA (Figure 12). Small-scale, academic users make up the bulk of orders in the synthetic DNA space (75% as of August 2017).[83] Furthermore, academic orders are typically for synthesis supplies such as enzymes and primers, rather than synthetic genes. Large-scale, commercial users who place orders for synthetic genes make up only a quarter of the market. For most industrial applications, synthetic DNA is a tool that

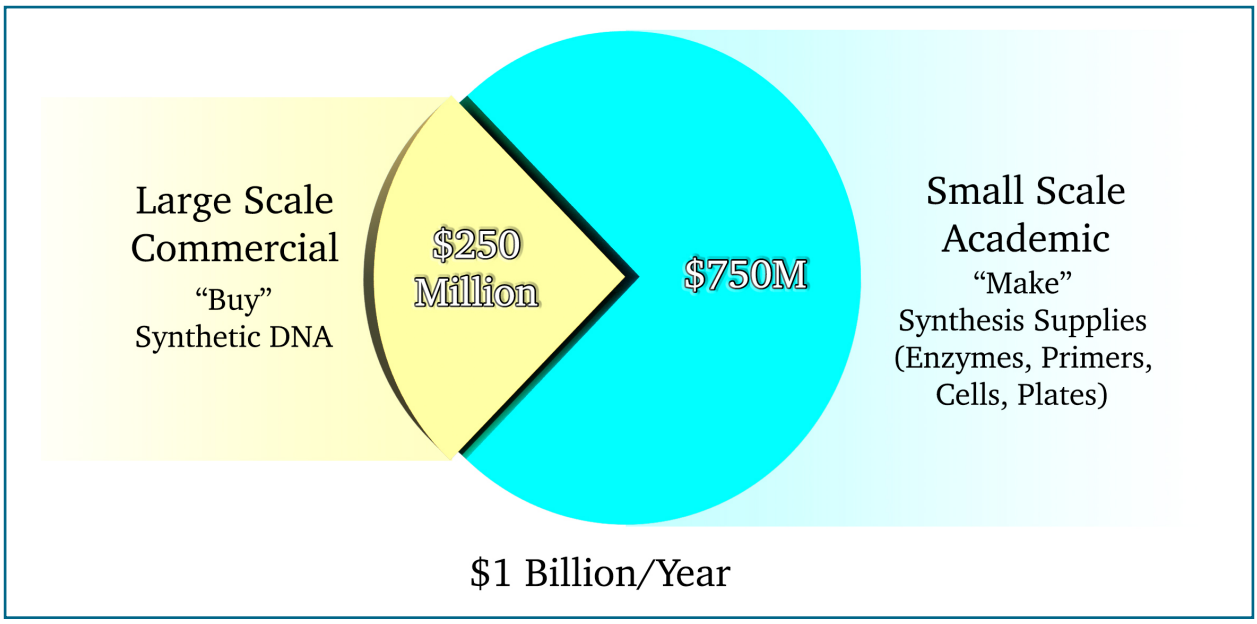


Figure 12
Commercial versus academic market for synthetic DNA (as of August 2017).[84]

is likely ordered only once and used to create a product. Once obtained, there is no requirement to repeat the order as the gene can be replicated cheaply in-house. Furthermore, most commercial or academic users are not currently designing complex synthetic gene circuits, which would place a demand for bulk orders of synthetic DNA.

The ideal DNA synthesis platform that drives down costs and improves efficiency would move beyond phosphoramidite chemical synthesis. It would perform *de novo*, uninterrupted synthesis of large sequences. Several start-up companies and academic institutes are developing novel enzyme-based DNA synthesis technologies. This technology is the first major breakthrough in DNA synthesis technology in ~40 years and will significantly cut costs and production time. The method's potential to disrupt the DNA synthesis field has been compared to NGS's 2007 disruption of the DNA sequencing field. Researchers have stated that enzyme-based approach will lead to ubiquitous "DNA printers."[85]

The following research groups are actively developing enzymatic DNA synthesis technology:

- In July 2018, researchers at the Department of Energy's Joint BioEnergy Institute (JBEI) published a DNA synthesis method that uses an enzyme called terminal deoxynucleotidyl transferase (TdT)—an enzyme found in immune cells that is capable of extending DNA by 200 bases per minute without the need for a template.[86] Their enzyme-based technology could produce DNA strands 10 times longer than today's methods, with greater accuracy and significantly faster speeds.

- The Paris-based DNA Script, founded in 2014, has received $25M in public and private financing from Illumina Ventures, Merck Ventures, Sofinnova, Kurma, Idinvest, Bpifrance, and the European Commission. The 30-person team has reached the following milestones: synthesis of a codon (3-letter strand of DNA) in 2015, enzymatic synthesis of a 10-nucleotide-long strand of DNA in 2016, enzymatic synthesis of a 30-nucleotide strand of DNA in 2017, and the enzymatic synthesis of a 50-nucleotide strand and a 150-nucleotide strand of DNA in 2018. The company has said that its technology will be released to early adopters in 2019, indicating that commercialization is right around the corner.[87]

- San Diego-based Molecular Assemblies was founded in 2013 by J. William Efcavitch and Curt Becker, two early founders of Applied Biosystems Incorporated—the company that first commercialized solid-phase DNA synthesizers in the early 1980s.[88] In 2016, Molecular Assemblies raised $2.3M in seed-round financing from investors Agilent Technologies, Cavendish Impact Capital Fund, Eleven Two Capital, Keshif Ventures, Genomics Investment Syndicate, Newport Holdings, LP, and Alexandria Venture Investments.[89]

- UK-based Evonetix is developing two sets of proprietary technology: an enzymatic synthesis technology that can be integrated into a new a silicon array DNA synthesis platform capable of independent control over 10,000 parallel reactions.[90] The company claims that the merger of the two technologies will radically decrease DNA synthesis costs and product turnover times. The company secured a £ 1.3M grant for this project from Innovate UK, which is the UK's Innovation Agency.[91]

- Harvard Medical School and Technicolor Research and Innovation Lab are actively researching enzymatic synthesis technology, tailored for the purpose of DNA data storage. They have a pre-print publication that describes the use of enzymatic DNA synthesis to encode data in DNA.[92]

- A UK-based start-up company called Nuclera is also developing enzymatic DNA synthesis technology.[93] The novelty of their work is unclear, and much is unknown about their R&D efforts so far. Similar to Evonetix, they have also received funding from Innovate UK.

DNA synthesis costs have steadily decreased since the start of this century. In 2001, DNA synthesis costs were $15 per base pair, equating to ~$63M/MB of data (Figure 13). Today, DNA synthesis costs are ~$0.07 per base pair, equating to ~$294,000/MB of data (Figure 13). While costs have gone down significantly, current prices on storing data in DNA are still exorbitantly high. As mentioned in this report, this is due to old, inefficient, organic chemistry-based methods of DNA synthesis. Even with steady improvements in this first-generation synthesis technology, major cost reductions are yet to happen.

By projecting current costs of DNA synthesis forward, assuming no development of a 2nd generation synthesis technology, the cost to synthesize DNA will reach $0.005 per base pair by 2027 (Figure 13). This equates to ~$21,000/MB of data. To put this in perspective, it would cost $14M to synthesize

a human genome, or about 715 MB of data. This cost is prohibitive for adoption of DNA as a storage medium. It is indicative of the need for a 2nd generation DNA synthesis technology that offers massive cost reductions, akin to what 2nd generation sequencing technologies brought to the DNA sequencing world in 2007.

The exponential drop in price for DNA sequencing in 2007 was driven by the commercialization of "second-generation" sequencing platforms. The first academic paper showing the ability to sequence DNA using a second-generation method was published in 1998.[94] A 2005 publication demonstrated the use of this methodology to sequence a full genome.[95] By 2007, the technology was commercialized by Roche, initiating the exponential price drop in DNA sequencing costs.

A similar "second-generation" platform has yet to be commercialized in the field of DNA synthesis. As mentioned previously, several independent groups are developing enzymes that perform *de novo* DNA synthesis. Results have been promising and enzymatic synthesis projects to be the "2nd generation" synthesis technology, alongside advancements in microarray platforms, that will significantly drive down costs in the coming years.

This technology has reached significant milestones recently. DNA Script first demonstrated the enzymatic synthesis of a 3nt sequence in 2015 and have since followed this up by synthesizing a 50nt sequence. JBEI demonstrated this year the synthesis of a 10nt sequence. Molecular Assemblies announced in August 2018 the first demonstration of enzymatic synthesis of DNA that was used to store and retrieve data, marking a significant milestone in this field. DNA Script demonstrated the synthesis of a 150nt sequence with a 99.5% efficacy for each nucleotide added in 2018, with commercialization plans starting in 2019.

Using the NGS model for cost decrease and DNA Script's proprietary synthesis technology being released to early adopters next year, enzymatic synthesis technology is expected to hit the commercial market by 2020, leading to significant price drops in 2021and beyond (Figure 13).

A price drop in DNA synthesis, mimicking the post-2007 exponential decrease in DNA sequencing costs, would result in drastic reductions in cost of using DNA for data storage. In this model, by 2025, the cost to synthesize DNA will reach $0.00001 per base pair (Figure 13). This equates to ~$42/MB of data encoding costs. By 2030, DNA synthesis costs would fall to $0.0000001 per base pair equating to ~$1/MB of data (Figure 13).

It should be noted that IARPA's MIST program has set much more aggressive goals for reductions in DNA synthesis costs (Figure 14). These cost reduction targets require that DNA synthesis costs decrease at a significantly faster rate than the cost reductions that happened in the DNA sequencing field. If cost targets for DNA synthesis as outlined in the MIST program are achieved, storing 1 MB of data in DNA will cost ~$0.00008 by 2023. A workshop hosted by Semiconductor Research Corporation on the future of DNA synthesis technologies highlighted several areas of advancement that can produce significant cost reductions. Notable areas of advancement include improvements in microarray density/resolution and advancing towards enzyme-only approaches.[96]

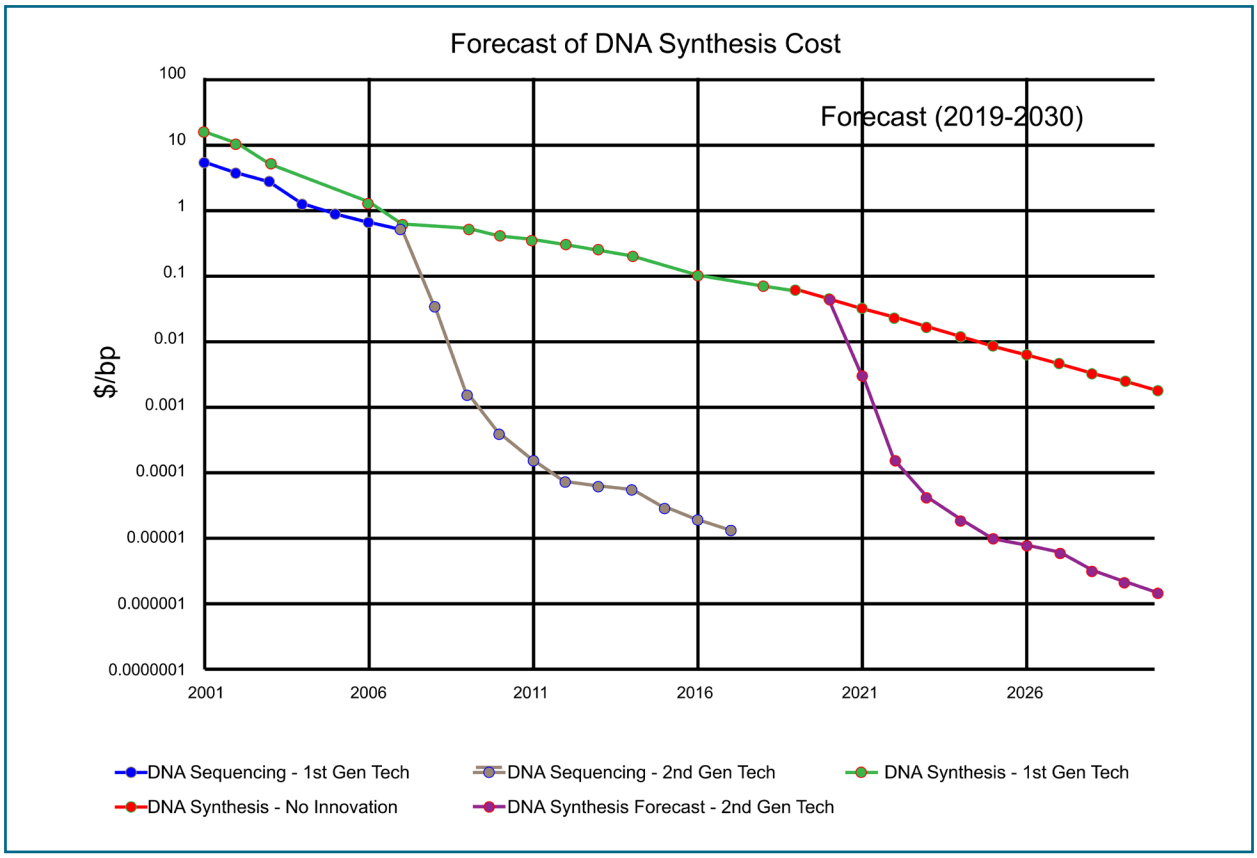


Figure 13
Projected costs of DNA synthesis costs per base pair.

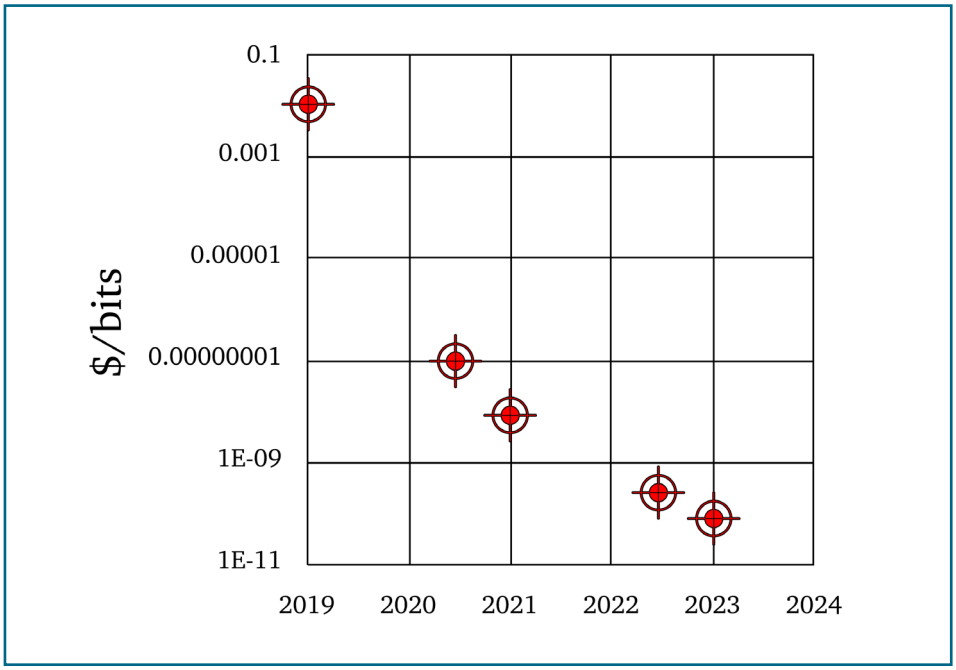


Figure 14
IARPA's cost targets for DNA Synthesis technology by 2023.[97]

FORECAST FOR DNA DATA STORAGE TECH ADVANCEMENT

While the cost of storing data in DNA is primarily reliant on synthesis technology, innovations in other areas will also lead to significant cost reductions and improvements in functional capability. These areas include our ability to increase the storage density of DNA, improving the coding architecture underlying DNA data storage, faster DNA reading speeds, and developing new random-access data retrieval methods.

www.shutterstock.com

Increasing Data Storage Density in DNA

The theoretical storage density of DNA is 2 bits of data per nucleotide (i.e., ~14 atoms per bit), far greater than any traditionally-used storage medium. To date, the best storage density in DNA was performed by Yaniv Erlich's team at Columbia in 2017, at 1.6 bits of data per nucleotide (80% of the theoretical limit). Erlich's DNA Fountain technique, based on recent developments in computer coding theory, is capable of packing up to 215 petabytes of data in a single gram of DNA—a 10X improvement over previous DNA storage methods.[98,99]

Another theoretical approach could be chemically modifying DNA bases to increase data storage density. The base cytosine can be chemically modified to uracil for instance. This could theoretically enable the storage of two layers of information within the same DNA sequence by modifying C to U, with each representing a different digital value. Other approaches could look at modification of specific base pairs through the addition of "tags." An example of this would be the addition of a methyl group to either cytosine or adenine.

Advanced Coding Schemes

Coding schemes can be designed to optimize how data is stored within DNA. Two companies, Helixworks and Catalog are taking radically different approaches to this.

- Helixworks is working on a coding system called Molecular Storage System (MOSS). The MOSS coding architecture breaks from conventional DNA storage architectures by only using the A (represented '0') and T (representing '1') bases to encode information. The C and G bases are used to extend sequences and stabilize DNA. Such a system radically decreases the number of combinations and operational blocks necessary to encode information in DNA. Another benefit of this approach would be in the readout of sequences, as only the A and T bases would require sequencing.
- Catalog is developing a different coding system that relies on pre-synthesized blocks of DNA (20-30 bases), that can be stitched together to encode data in DNA. This approach can radically reduce the need to synthesize novel DNA fragments continuously from scratch and allow for faster and cheaper encoding of information. Catalog says that it aims to encode 1 TB of data per day in DNA by 2019.[100] To achieve this goal, they have recently entered into an agreement with the UK-firm Cambridge Consultants to build the world's first DNA storage device, that will approximately be the size of a school bus (Figure 15).

These approaches are creative solutions to the prohibitive costs of DNA synthesis, and if scalable, will work towards bringing down data storage costs for DNA. When examined within the context of enzymatic-DNA synthesis breakthroughs and significant cost reductions in DNA synthesis, it is likely that these approaches will radically transform our ability to synthesize DNA sequences encoding bits in a rapid and cost-effective manner.

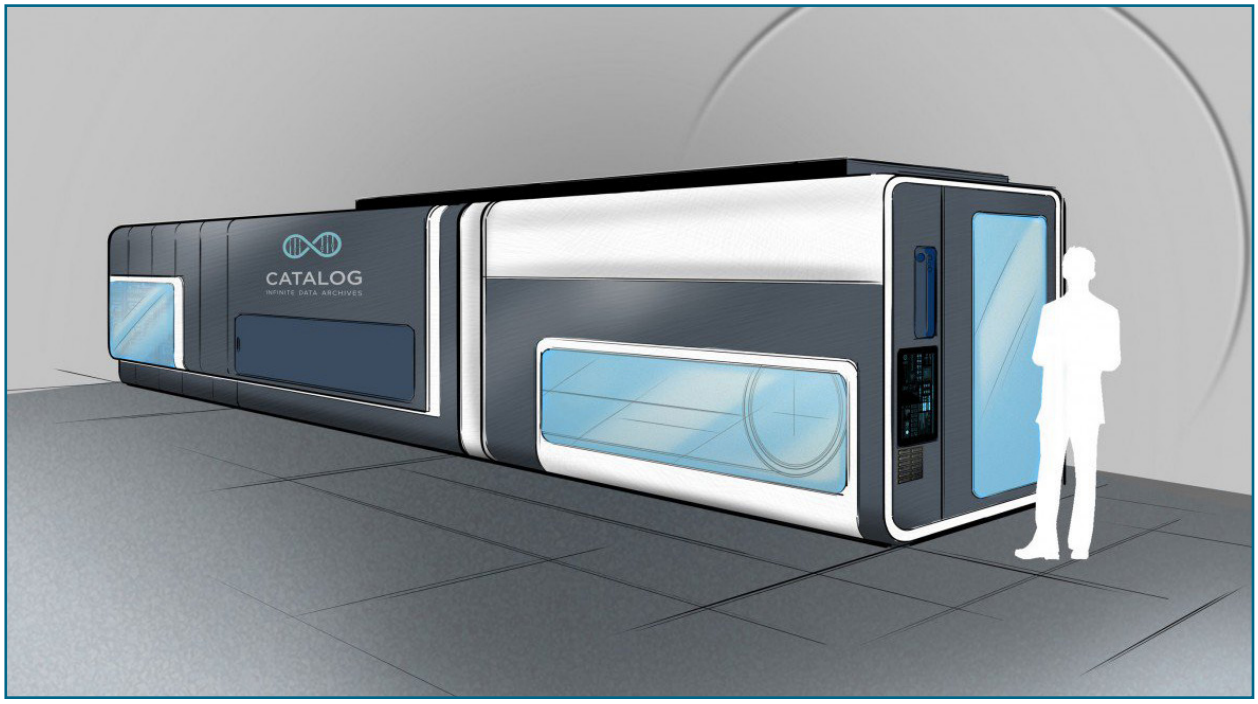


Figure 15
Artistic rendering of Catalog's DNA storage device. Image courtesy of Catalog Technologies Inc.

DNA Sequencing Methods

Although DNA sequencing technology has progressed exponentially over the last decade, the speed of sequencing would still require great improvement for routine and fast retrieval of data from DNA. Sequencers capable of incredibly fast reading speeds are bulky devices, whereas portable sequencers are significantly slower. The bulky but fast sequencers can be utilized for accessing archived data but would have little utility as a DNA "hard drive." The portable sequencers that might function as a "hard drive" are too slow, preventing fast data retrieval.

- The fastest sequencing technology on the market belongs to Illumina, which announced in 2017 that its new NovaSeq products can sequence a human genome in one hour.[101] Since the human genome is roughly ~3 billion bases long, this equates to reading 833,333 bases per second. However, Illumina devices are not portable and quite expensive.

Olgica Milenkovic's group at University of Illinois at Urbana-Champaign used Oxford Nanopore's portable MinION device to demonstrate the error-free read-out from DNA of 3.6 KB of binary data coding for two compressed images.[102] However, the sequencing speed was just 75 bases per second.

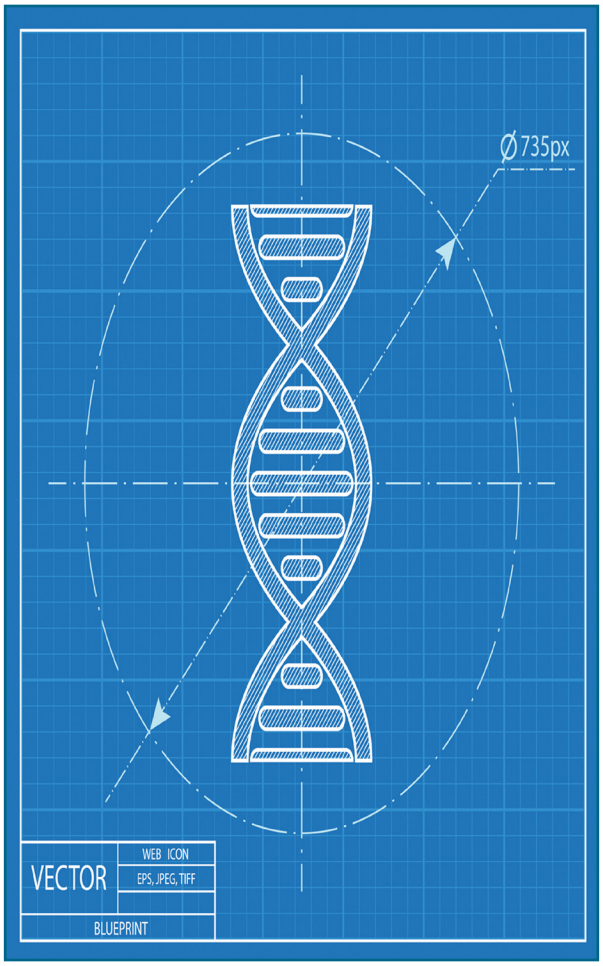
Sequencing speeds in portable machines must be improved exponentially for DNA storage to have broad applications beyond archiving old data that is infrequently accessed. Moving forward, development of DNA sequencers for the specific purpose of reading data stored in DNA will become commonplace. Such devices would rely on research that has shown that accuracy of sequencing is not a prerequisite to retrieving data accurately from DNA. Error-correcting codes can be deployed to account for common mistakes made by sequencers, allowing for accurate data retrieval. Initiatives such as IARPA's MIST program will facilitate the development of such technologies. One of the Technical Areas of focus for MIST is the development of "a table-top device capable of randomly accessing information from molecular media with a target throughput and resource utilization budget."

DNA Data Retrieval

Random access retrieval of data stored in DNA has been demonstrated.[103] The method employs a commonly-used genetics technique called Polymerase Chain Reaction (PCR), which can selectively detect and amplify a targeted region of DNA. This method has applications for accessing archived data stored in DNA, that wouldn't require rapid read-out. However, it would have little to no application in a portable DNA hard drive as the process of PCR is time-intensive and requires a machine capable of thermocycling.

Future technology that could rapidly retrieve data from DNA, using a random-access approach, should be inspired by biology. DNA in cells is read using a variety of mechanisms such as epigenetic tags, promoter sequences that are bound by specialized proteins that turn genes "on" or "off" and enzymes that read sequences to generate messengerRNAs. For fast and efficient retrieval of stored data, research should explore how such biological mechanisms can be exploited for random-access data retrieval purposes.

Development of Operating Software compatible with and optimized for DNA reading devices is another capability that would lead to improvements in data retrieval. Development of an OS compatible with DNA reading devices is currently a focus of IARPA's MIST program.



www.shutterstock.com

Conclusions

The ability to store data in DNA was initially proven in the late 1980s and research within this field has significantly ramped up this decade. Within the last year, major leaps have been made in the volume of data stored (400 MB) and maximization of storage density (80% of theoretical limit). However, using DNA to store data is still solely within the domains of laboratory science. This is primarily caused by the exorbitant price of DNA synthesis, which is currently reliant on decades old chemical-based methodologies. Another factor is the lack of dedicated technologies for DNA data storage, as researchers have to use life sciences DNA technology for data storage purposes.

Recent breakthroughs in enzymatic DNA synthesis and the novel approach to DNA data storage undertaken by Catalog show that storing data within DNA is poised to move outside of laboratory science and into practical use by 2019-2020. Catalog's proprietary DNA storage machine, which will complete development by early 2019, is expected to store 1TB of data in DNA per day at a cost of a few thousand dollars. This technology will encode enormous volumes of data, which is currently stored in large data centers, into miniature test tubes containing invisible pellets of DNA.

The initial utility of DNA data storage will be for "cold" archival storage of data that is infrequently used or large volumes of data that must be transported physically across distant locations. Looking further into the future, this field presents an exciting opportunity to innovate in the broader computer science field. New technologies that support DNA data storage will be developed through initiatives such as IARPA's MIST programs. These technologies will include DNA reading devices that are significantly faster than today's best DNA sequencers, random access retrieval methods utilize novel molecular approaches at scanning DNA sequences and operating software that are tailor-made for working with DNA storage devices. As these technologies are developed, DNA data storage will become part of a broader ecosystem of new computing technology that is based on the merger of synthetic biology and the semiconductor industry.

Endnotes

1. Molteni, M. June 26, 2018. The Rise of DNA Data Storage. Wired. Retrievable at: https://www.wired.com/story/the-rise-of-dna-data-storage/

2. Data Storage Supply and Demand Worldwide, from 2009 to 2020 (in exabytes). Statista. Retrievable at: https://www.statista.com/statistics/751749/worldwide-data-storage-capacity-and-demand/

3. Zhirnov, V; Zadegan R, Sandhu, G; Church. G; Hughes, W. March 23, 2016. Nucleic acid memory. Nature Materials. Retrievable at: https://www.nature.com/articles/nmat4594

4. November, 2016. Next-Generation Data Storage Market by System (Direct-Attached, Network-Attached, Storage Area Network, Software-defined, Cloud, Unified), Architecture (File, Object, Block), Technology (Magnetic, Solid-State), and Industry - Global Forecast to 2022. Markets and Markets. Retrievable at: https://www.marketsandmarkets.com/Market-Reports/next-generation-data-storage-market-12592401.html

5. The Next Generation of Information Storage. iGEM. Retrievable at: http://2016.igem.org/Team:Edinburgh_UG

6. Cornish, C. February 5, 2018. How DNA could store all the world's data in a semi-trailer. Financial Times. Retrievable at: https://www.ft.com/content/45ea22b0-cec2-11e7-947e-f1ea5435bcc7

7. New Research Topics. Semiconductor Research Corporation. Retrievable at: https://www.src.org/src/new-research-topics/

8. Dormehl, L. July 8, 2018. When we run out of room for data, scientists want to store it in DNA. Digital Trends. Retrievable at: https://www.digitaltrends.com/cool-tech/dna-data-catalog-startup/

9. Bishop, B; McCorkle, N; Zhirnov, V. September 14, 2017. Technology Working Group Meeting on future DNA synthesis technologies. Retrievable at: https://www.src.org/program/grc/semisynbio/semisynbio-consortium-roadmap/6043-full-report-dna-based-storage-final-twg1-4-18.pdf

10. Molecular Information Systems Lab. University of Washington. Retrievable at: http://misl.cs.washington.edu/

11. July 4, 2013. Recalibrating Equus evolution using the genome sequence of an early Middle Pleistocene horse. Nature. Retrievable at: https://www.ncbi.nlm.nih.gov/pubmed/23803765

12. Blawat, M; Gaedke, K; Hutter, I; Chen, X.M.; Turczyk, B; Inverso, S; Pruitt, B.W.; Church, G.M. June 1, 2016. Forward Error Correction for DNA Data Storage. Procedia Computer Science. Retrievable at: https://www.sciencedirect.com/science/article/pii/S1877050916308742

13. Church, G.M; Gao, Y; Kosuri, S. September 28, 2012. Next-Generation Digital Ifnromation Storage in DNA. Science. Retrievable at: http://science.sciencemag.org/content/337/6102/1628.full?ijkey=ziOBjbBreAezY&keytype=ref&siteid=sci

14. Extance, A. August 31, 2016. How DNA could storage all the world's data. Nature. Retrievable at: https://www.nature.com/news/how-dna-could-store-all-the-world-s-data-1.20496

15. February 19, 2018. UW and Microsoft researchers achieve random access in large-scale DNA data storage. University of Washington. Retrievable at: https://news.cs.washington.edu/2018/02/19/uw-and-microsoft-researchers-achieve-random-access-in-large-scale-dna-data-storage/

16. Extance, A. August 31, 2016. How DNA could storage all the world's data. Nature. Retrievable at: https://www.nature.com/news/how-dna-could-store-all-the-world-s-data-1.20496

17. Eduardo Kac. September, 2019. Genesis. Ars Electronica. Retrievable at: http://www.ekac.org/geninfo.html

18. McDowell, N. January 8, 2013. Data stored in multiplying bacteria. New Scientist. Retrievable at: https://www.newscientist.com/article/dn3243-data-stored-in-multiplying-bacteria/

19. Gustafsson, C. April 9, 2018. For anyone who ever said there's no such as a poetic gene. Nature. Retrievable at: https://www.nature.com/articles/458703a

20. Rotstein, O.D.; Ailnberg, M. April 25, 2018. An improved Huffman coding method for archiving text, images, and music characters in DNA. Future Science. Retrievable at: https://www.future-science.com/doi/full/10.2144/000113218

21. July 2, 2010. Creation of a Bacterial Cell Controlled by a Chemically Synthesized Genome. Science. Retrievable at: http://science.sciencemag.org/content/329/5987/52

22. Church, G.M; Gao, Y; Kosuri, S. September 28, 2012. Next-Generation Digital Ifnromation Storage in DNA. Science. Retrievable at: http://science.sciencemag.org/content/337/6102/1628.full?ijkey=ziOBjbBreAezY&keytype=ref&siteid=sci

23. Goldman, N; Bertone, P; Chen S; Dessimoz, C; LeProust, E; Sipos, B; Birney, E. February 7, 2013. Towards practical, high-capacity, low-maintenance information storage in synthesized DNA. Nature. Retrievable at: https://www.nature.com/articles/nature11875

24. Grass, R.N.; Hecker, R; Puddu, M; Paunescu, D; Stark, J.W. February 4, 2015. Robust Chemical Preserviation of Digital Information on DNA in Silica with Error-Correcting Codes. Angewandte Chemie International Edition. Retrievable at: https://onlinelibrary.wiley.com/doi/abs/10.1002/anie.201411378

25. Bornholt, J; Lopez, R; Carmean, D.M.; Ceze, L; Seelig, G; Strauss, K. 2016. A DNA-Based Archival Storage System. ASPLOS. Retrievable at: https://homes.cs.washington.edu/~bornholt/dnastorage-asplos16/

26. Blawat, M; Gaedke, K; Hutter, I; Chen, X.M.; Turczyk, B; Inverso, S; Pruitt, B.W.; Church, G.M. June 1, 2016. Forward Error Correction for DNA Data Storage. Procedia Computer Science. Retrievable at: https://www.sciencedirect.com/science/article/pii/S1877050916308742

27. Erlich, Y; Zielinski, D. March 3, 2017. DNA Fountain enables a robust and efficient storage architecture. Science. Retrievable at: http://science.sciencemag.org/content/355/6328/950

28. March 2018. Scaling up DNA data storage and random access retrieval. Nature Biotechnology. Retrievable at: https://www.microsoft.com/en-us/research/publication/scaling-dna-data-storage-random-access-retrieval/

29. February 19, 2018. UW and Microsoft researchers achieve random access in large-scale DNA data storage. University of Washington. Retrievable at: https://news.cs.washington.edu/2018/02/19/uw-and-microsoft-researchers-achieve-random-access-in-large-scale-dna-data-storage/

30. Bergamin, F. April 20, 2018. Entire music album to be stored on DNA. ETH Zurich. Retrievable at: https://www.ethz.ch/en/news-and-events/eth-news/news/2018/04/entire-music-album-to-be-stored-on-DNA.html

31. Kling, J. June, 1997. Peptide, Oligonucleotide Synthesis: Key In Molecular Studies. The scientist. Retrievable at: https://www.the-scientist.com/technology/peptide-oligonucleotide-synthesis-key-in-molecular-studies-57449

32. Hogrefe, R. A short history of Oligonucleotide synthesis. TriLink Biotechnologies. Retrievable at: https://www.trilinkbiotech.com/tech/oligo_history.asp

33. Technology. Twist Bioscience. Retrievable at: https://twistbioscience.com/technology

34. Frequently Asked Questions. GenScript. Retrievable at: http://www.customarrayinc.com/faq_main.htm

35. Technology. Twist Bioscience. Retrievable at: https://twistbioscience.com/technology

36. November 4, 2015. Ginkgo Bioworks and Twist Bioscience Just Made a 100 Million Base Pair Deal. SynBioBeta. Retrievable at: https://synbiobeta.com/news/ginkgo-bioworks-and-twist-bioscience-just-made-a-100-million-base-pair-deal/

37. April 17, 2017. Twist Bioscience Expands Agreement to Pursue Higher Density Digital Data Storage on DNA with Microsoft and University of Washington. Twist Bioscience. Retrievable at: https://twistbioscience.com/company/press/twist-bioscience-expands-agreement-pursue-higher-density-digital-data-storage-dna-microsoft-university-washington

38. June 14, 2017. Twist Bioscience and the BioBricks Foundation Announce Partnership to Provide 10,000 Public-Benefit Genes to the Synthetic Biology Community. Twist Bioscience. Retrievable at: https://twistbioscience.com/company/press/twist-biobricks-foundation-partnership

39. October 3, 2017. Twist Bioscience to Provide One Billion Base Pairs of Synthetic DNA to Ginkgo Bioworks to Support Expansion into New Industries. Twist Bioscience. Retrievable at: https://twistbioscience.com/company/press/supplies-1b-base-pairs-dna-ginkgo-bioworks

40. Carlson, R. May 13, 2016. Late Night, Unedited Musings on Synthesizing Secret Genomes. Synthesis. Retrievable at: https://www.synthesis.cc/synthesis/2016/05/synthesizing_secret_genomes

41. Regaldo, A. May 22, 2017. Microsoft Has a Plan to Add DNA Data Storage to Its Cloud. Wired. Retrievable at: https://www.technologyreview.com/s/607880/microsoft-has-a-plan-to-add-dna-data-storage-to-its-cloud/

42. Yong, E. Mar 2, 2017. This Speck of DNA Contains a Movie, a Computer Virus, and an Amazon Gift Card. The Atlantic. Retrievable at: https://www.theatlantic.com/science/archive/2017/03/this-speck-of-dna-contains-a-movie-a-computer-virus-and-an-amazon-gift-card/518373/

43. The Synthetic Biology Industry: Annual Industry Growth Update. SynBioBeta. Retrievable at: https://synbiobeta.com/wp-content/uploads/sites/4/2017/03/Synthetic-Biology-Industry-Annual-Growth-Update-2017.pdf

44. Ibid.

45. Ibid.

46. Ibid.

47. DNA Sequencing Costs: Data. National Human Genome Research Institute. Retrievable at: https://www.genome.gov/sequencingcostsdata/

48. Heather, J; Chain, B. January 2016. The sequence of sequencers: The history of sequencing DNA. Genomics. Retrievable at: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4727787/

49. Illumina Sequencing by Synthesis. Illumina. Retrievable at: https://www.youtube.com/watch?v=fCd6B5HRaZ8

50. Single Molecular Real Time Sequencing. Pacific Biosciences. Retrievable at: https://www.youtube.com/watch?v=v8p4ph2MAvI

51. SmidgION. Oxford Nanopore Technologies. Retrievable at: https://nanoporetech.com/products/smidgion

52. Oxford Nanopore MinION. KRISP.Retrievable at: http://www.krisp.org.za/toys.php?id=14

53. March 23, 2017. Turning Chemistry for New "Computing" Concepts. Defense Advanced Research Projects Agency. Retrievable at: https://www.darpa.mil/news-events/2017-03-23

54. Intelligence Community Wants to Use DNA to Store Exabytes of Data. NextGov. Retrievable at: https://www.nextgov.com/emerging-tech/2018/06/intelligence-community-wants-use-dna-store-exabytes-data/148908/

55. NSF Program Solicitation 17-557. National Science Foundation. Retrievable at: https://www.nsf.gov/pubs/2017/nsf17557/nsf17557.htm

56. NIH to Fund DNA Data Sample Storage. GenomeWeb. Retrievable at: https://www.genomeweb.com/neurological-psychological-disorders/nih-fund-dna-data-sample-storage#.W0d5BC2ZMWo

57. February 19, 2018. UW and Microsoft researchers achieve random access in large-scale DNA data storage. University of Washington. Retrievable at: https://news.cs.washington.edu/2018/02/19/uw-and-microsoft-researchers-achieve-random-access-in-large-scale-dna-data-storage/

58. Church, G.M; Gao, Y; Kosuri, S. August 16, 2012. Next-Generation Digital Information Storage in DNA. Sciencexpress. Retrievable at: http://arep.med.harvard.edu/pdf/Church_Science_12.pdf

59. July 2, 2010. Creation of a Bacterial Cell Controlled by a Chemically Synthesized Genome. Science. Retrievable at: http://science.sciencemag.org/content/329/5987/52

60. Mullin, E. July 12, 2017. Scientists used CRISPR to Put A GIF Inside a Living Organism's DNA. MIT Technology Review. Retrievable at: https://www.technologyreview.com/s/608268/scientists-used-crispr-to-put-a-gif-inside-living-dna/

61. Capacity-approaching DNA storage. Erlich Lab. Retrievable at: http://dnafountain.teamerlich.org/

62. Erlich Lab. Retrievable at: http://teamerlich.org/website/home.html

63. Gudeman, K. July 17, 2018. Illinois receives $1.5 million to lower cost, improve viability of DNA data storage. Retrievable at: https://csl.illinois.edu/news/illinois-receives-15-million-lower-cost-improve-viability-dna-data-storage

64. Stock, M. March 22, 2016. DNA data storage could last thousands of years. Reuters. Retrievable at: https://www.reuters.com/article/us-dna-storage/dna-data-storage-could-last-thousands-of-years-idUSKCN0WO1DX

65. Semiconductor Research Corporation. Retrievable at: https://www.src.org/

66. 2018. SemiSynBio Roadmap. Semiconductor Research Corporation. Retrievable at: https://www.src.org/program/grc/semisynbio/

67. February 15, 2018. Sandhu, G. Micron Named Among Top 100 Global Innovators for Sixth Straight Year. Micron. Retrievable at: https://www.micron.com/about/blogs/2018/february/micron-makes-list-of-top-100-innovators

68. Sandle, T. July 16, 2017. Major tech companies see DNA storage as the future. Digital Journal. Retrievable at: http://www.digitaljournal.com/tech-and-science/science/microsoft-exploring-dna-storage-solution/article/497735

69. Dreir, T. March 9, 2017. Need Some Extra Storage? Try DNA. PCMag. Retrievable at: https://www.pcmag.com/article/352193/need-some-extra-storage-try-dna

70. Keown, A. June 26, 2018. Boston's Catalog Secures $9 Million in Funding to Advance DNA Data Storage Technology. Biospace. Retrievable at: https://www.biospace.com/article/boston-s-catalog-secures-9-million-in-funding-to-advance-dna-data-storage-technology/

71. Iridia. Retrievable at: http://iridia.com/about/

72. Data Storage Solution May Be in The DNA. Iridia. Retrievable at: http://iridia.com/data-storage-solution-may-be-in-the-dna/

73. DNADrive. HelixWorks. Retrievable at: https://www.amazon.com/Helixworks-dsDNA300-750hw-DNADrive/dp/B01IVSH4OM/

74. Molteni, M. January 24, 2018. DARPA wants to build an image search engine out of DNA. Wired. Retrievable at: https://www.wired.com/story/darpa-wants-to-build-an-image-search-engine-out-of-dna/

75. Intelligence Community Wants to Use DNA to Store Exabytes of Data. NextGov. Retrievable at: https://www.nextgov.com/emerging-tech/2018/06/intelligence-community-wants-use-dna-store-exabytes-data/148908/

76. Collins, F. July 18, 2017. DNA-Encoded Movie Points Way to "Molecular Recorder". NIH Director's Blog. Retrievable at: https://directorsblog.nih.gov/2017/07/18/

77. Forsyth, D. April 22, 2013. It's in the genes – data storage turns to DNA. Horizon. Retrievable at: https://horizon-magazine.eu/article/it-s-genes-data-storage-turns-dna.html

78. Goldman, N; Bertone, P; Chen S; Dessimoz, C; LeProust, E; Sipos, B; Birney, E. February 7, 2013. Towards practical, high-capacity, low-maintenance information storage in synthesized DNA. Nature. Retrievable at: https://www.nature.com/articles/nature11875Retrievable at:

79. Genes. Twist Bioscience. Retrievable at: https://www.twistbioscience.com/products/genes

80. Leproust, E. 2015. Beyond the $1K Genome: DNA 'Writing' Comes Next. Retrievable at: https://techcrunch.com/2015/09/18/beyond-the-1k-genome-dna-writing-comes-next-2/

81. Synthetic Biology FAQ. OpenWetWare. Retrievable at: https://openwetware.org/wiki/Synthetic_Biology:FAQ

82. DNA Sequencing Costs: Data. National Human Genome Research Institute. Retrievable at: https://www.genome.gov/sequencingcostsdata/

83. Leproust, E. July, 2017. Rewriting DNA Synthesis. Twist Bioscience. Retrievable at: https://www.bio.org/sites/default/files/1030AM-Emily%20Leproust.pdf

84. Ibid.

85. Sanders, R. June 18, 2018. New DNA synthesis technique promises rapid, high-fidelity DNA printing. Berkley News.Retrievable at: https://news.berkeley.edu/2018/06/18/new-dna-synthesis-technique-promises-rapid-high-fidelity-dna-printing/

86. Palluk, S; Arlow, D.H.; de Rond, T; Barthel, S; Kang, J.S.; Brector, R; Baghdassarian, H.M.; Troung, A; Kim, P.W.; Singh, A.K.; Hillson, N.J.; Keasling, J.D. June 18, 2018. De novo DNA synthesis using polymerase-nucleotide conjugates. Nature Biotechnology. Retrievable at: https://www.nature.com/articles/nbt.4173

87. October 2, 2018. DNA Script Announces World's First Enzymatic Synthesis of a High-Purity 150-Nucleotide Strand. DNAScript. Retrievable at: http://www.dnascript.co/pr9

88. History. Molecular Assemblies. Retrievable at: http://molecularassemblies.com/about/story/

89. December 12, 2016. Molecular Assemblies Raises 2.3 Million in Seed Round. Molecular Assemblies. Retrievable at: http://molecularassemblies.com/molecular-assemblies-raises-2-3-million-in-seed-round/

90. Technology. Evonetix. Retrievable at: https://www.evonetix.com/technology/

91. July 2, 2018. Evonetix awarded Innovate UK funding for £1.3 million gene synthesis project. Evonetix. Retrievable at: https://www.evonetix.com/evonetix-business-weekly-awards-2/

92. Lee, H.H.; Kalhor, R; Goela, N; Bolot, J. Church, G.M. June 16, 2018. Enzymatic Dna Synthesis for digital information storage. BioRXiv. Retrievable at: https://www.biorxiv.org/content/early/2018/06/16/348987

93. Nuclera. Retrievable at: https://www.nuclera.com

94. M, Ronaghi; M, Uhlen; Nyren, P. July 17, 2018. A sequencing method based on real-time pyrophosphate. Science. Retrievable at: https://www.ncbi.nlm.nih.gov/pubmed/9705713

95. September 15, 2005. Genome sequencing in microfabricated high-density picolitre reactors. Nature. Retrievable at: https://www.ncbi.nlm.nih.gov/pubmed/16056220

96. 2018. SemiSynBio Roadmap. Semiconductor Research Corporation. Retrievable at: https://www.src.org/program/grc/semisynbio/

97. Ibid.

98. Choi, C. March 2, 2017. Pushing the Theoretical Limits of DNA Data Storage. Discover Magazine. Retrievable at: http://blogs.discovermagazine.com/d-brief/2017/03/02/dna-data-storage-limits/

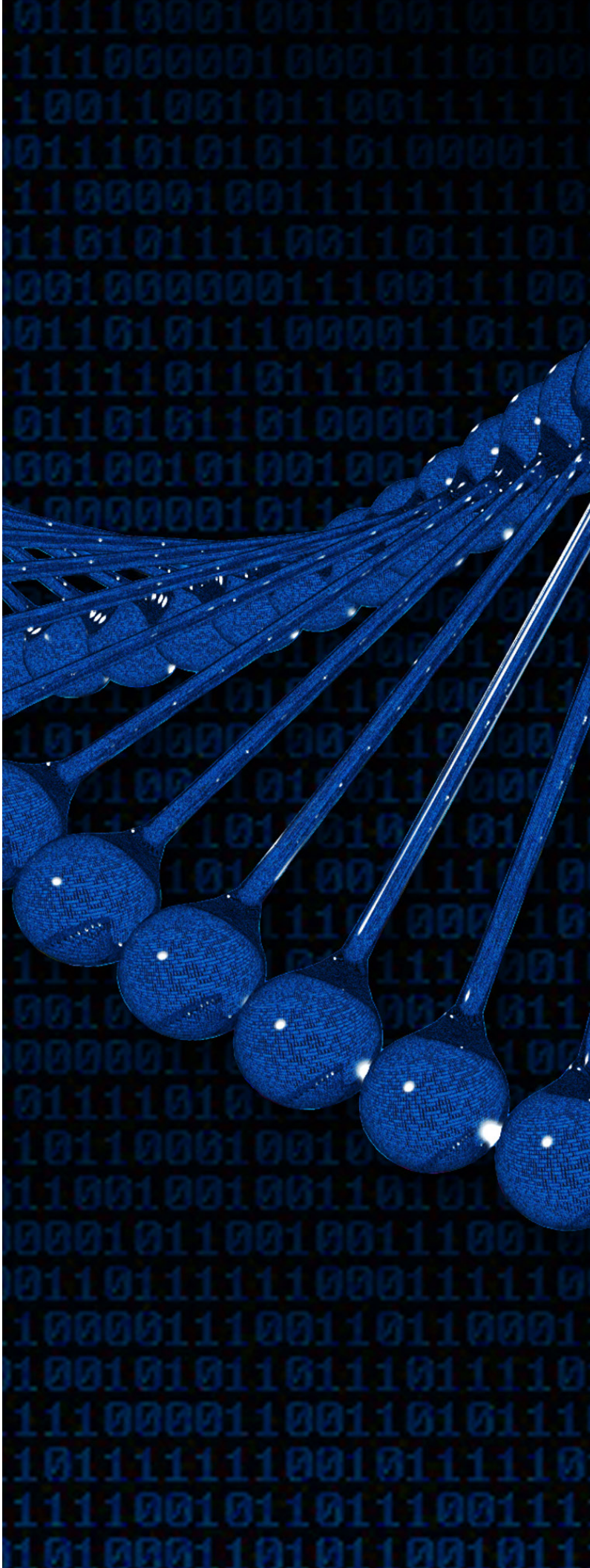
99. Erlich, Y; Zielinski, D. March 3, 2017. DNA Fountain enables a robust and efficient storage architecture. Science. Retrievable at: http://science.sciencemag.org/content/355/6328/950

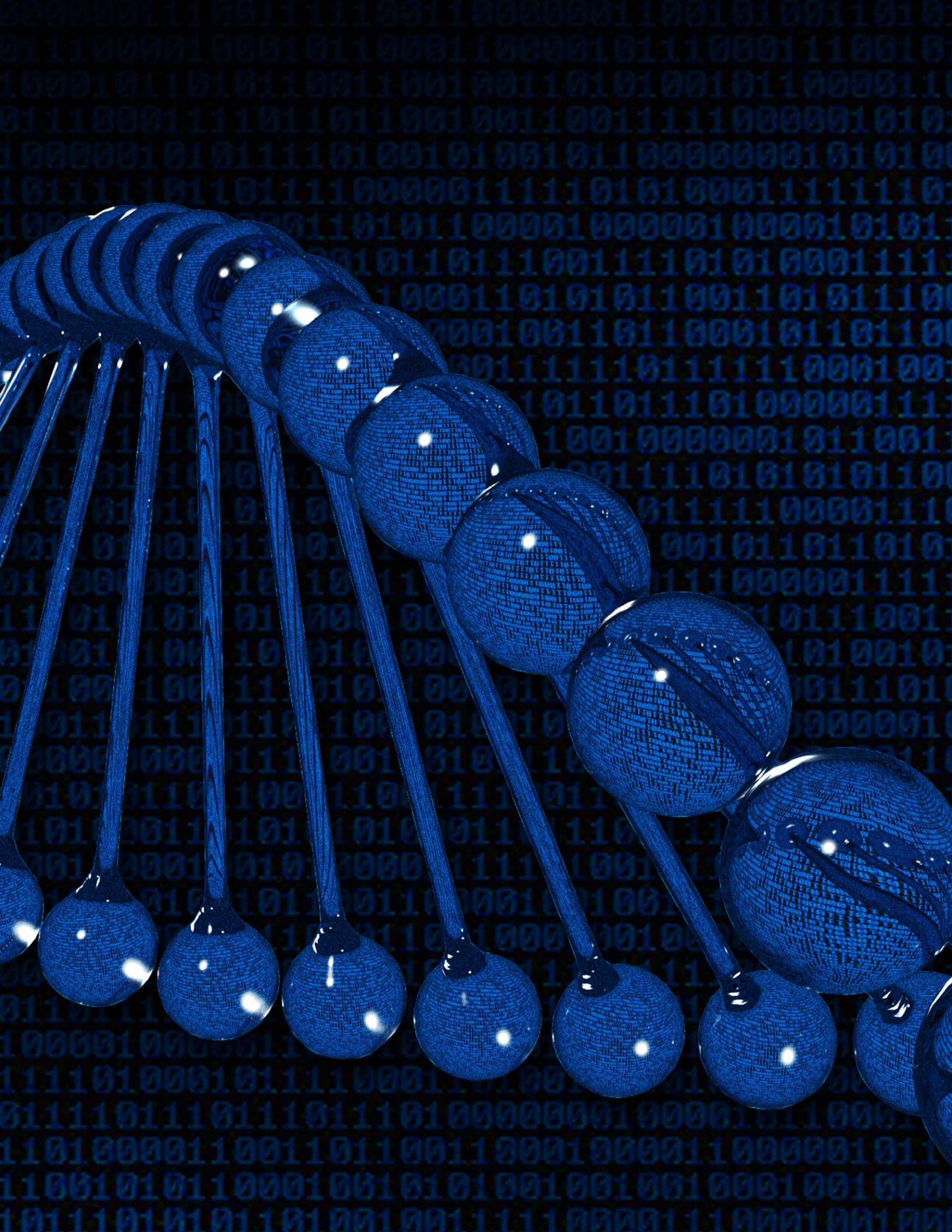
100. Regaldo, A. October 2, 2018. The DNA data storage machine that's the size of a school bus. MIT Technology Review. Retrievable at: https://www.technologyreview.com/s/612225/the-dna-data-storage-machine-thats-the-size-of-a-school-bus/

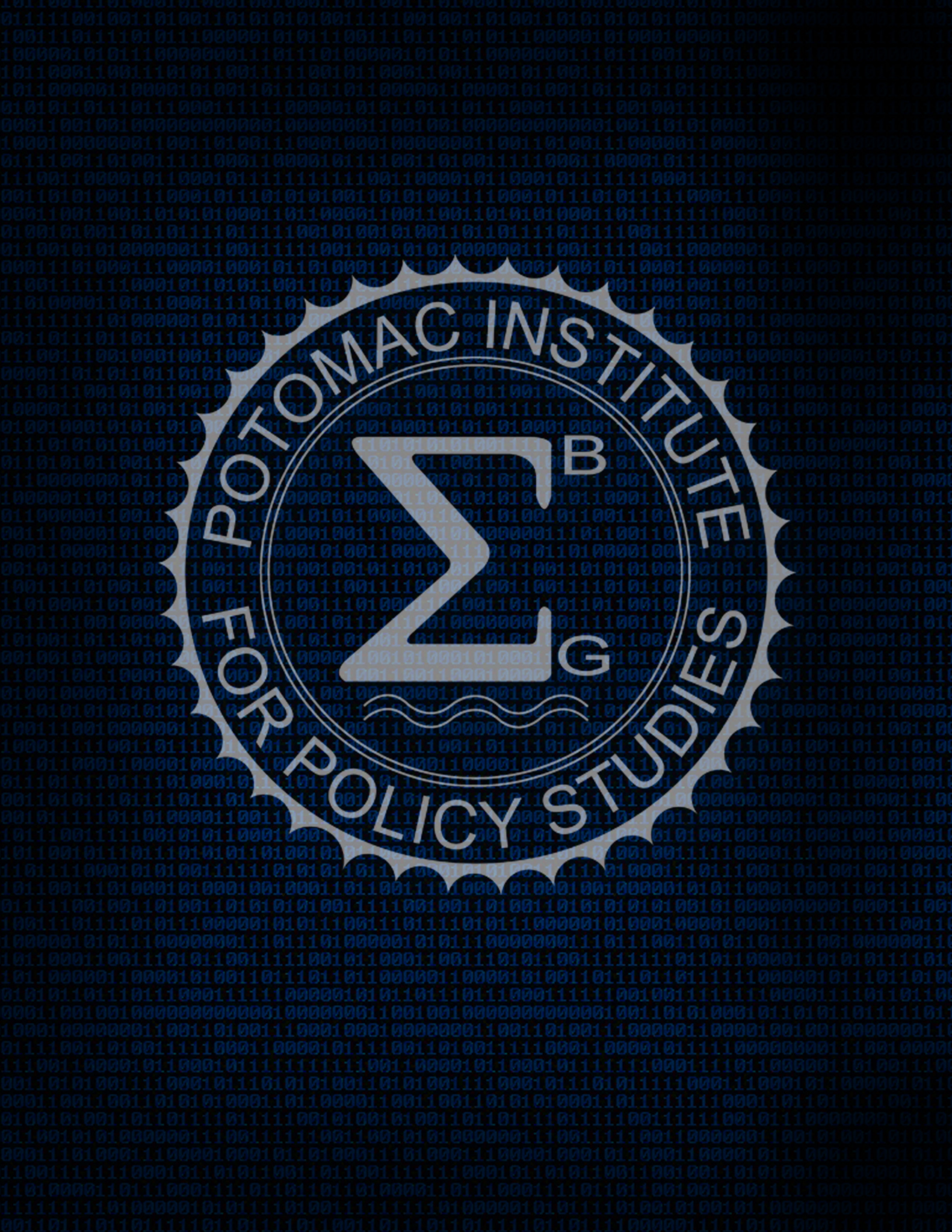
101. Fikes, B. January 9, 2017. New machines can sequence human genome in one hour, Illumina announces. The San Diego Union-Tribune. Retrievable at: http://www.sandiegouniontribune.com/business/biotech/sd-me-illumina-novaseq-20170109-story.html

102. Yazdi, S.M.; Gabrys, R; Milenkovic, O. July 10, 2017. Portable and Error-Free DNA-Based Data Storage. Nature Scientific Reports. Retrievable at: https://www.nature.com/articles/s41598-017-05188-1

103. February 19, 2018. Random access in large-scale DNA data storage. Nature Biotechnology. Retrievable at: https://www.nature.com/articles/nbt.4079






POTOMAC INSTITUTE
FOR POLICY STUDIES