TWIST
BIOSCIENCE

Synthetic DNA for Digital Storage

Brian W. Bramlett bbramlett@twistbioscience.com

October 2019

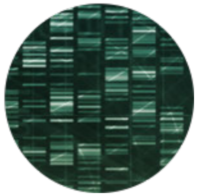
@TwistBioscience #WeMakeDNA

Writing Synthetic DNA on Silicon Platform

Fueling the Industrialization of Synthetic Biology

KEY ADVANTAGES OF WRITING DNA ON SILICON

MINIATURIZATION
10^{3-6} less volume of required reagents



THROUGHPUT
20M oligos/month

LOW COST
Driving adoption and new applications

VERSATILE PLATFORM
Broad applications

Multiple Large Market Opportunities

TWIST'S PLATFROM TECHNOLOGY ADDRESSES

INITIAL MARKETS

SYNTHETIC
BIOLOGY

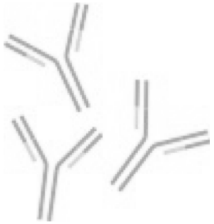
- Competitive Turnaround Time
- Lower Cost
- High Throughput
- High Quality

GENOMICS:
TARGETED NGS

- Fast Customization
- Performance
- Full Kit
- High Quality

SHORT TERM GOAL
Grow Revenue

LARGE MARKET OPPORTUNITIES



LARGE MARKET

DRUG DISCOVERY/ DEVELOPMENT

- High Quality Diversity Hits / Leads
- Shorter Time and Cost from Target to IND

MID TERM GOAL
Develop novel therapeutics

IT INDUSTRY

DATA STORAGE

- Permanence
- Density
- Ease of Copying
- Universal Format

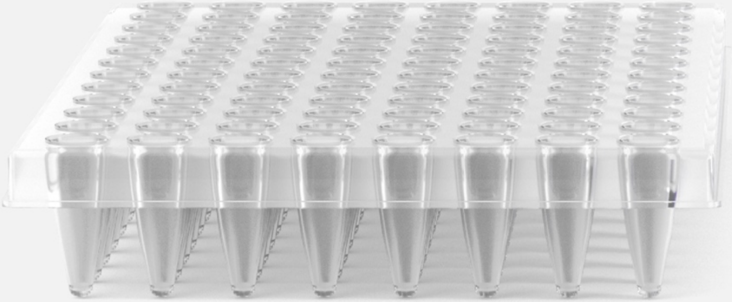
LONG TERM GOAL
Enter technology market

Synthetic Biology is Rapidly Growing

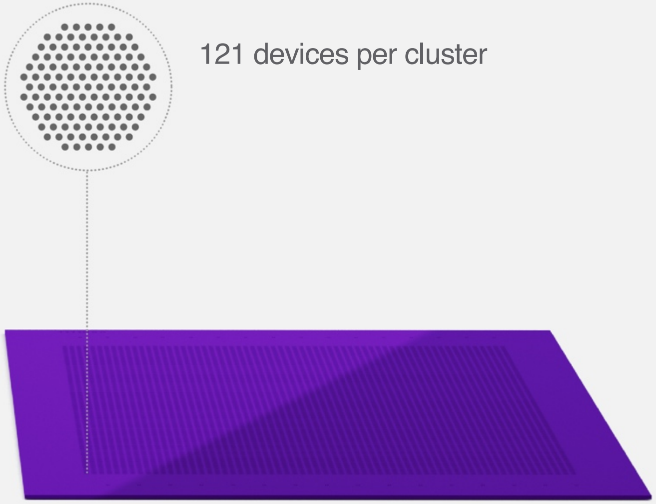
NEEDS	NEW APPLICATIONS FOR SYNTHETIC DNA
Healthcare • Better drug development tools to lessen time and lower costs • More effective diagnostic tools for DNA extraction to lower costs (i.e. NGS)	• Antibodies / TCR • Vaccines • Immuno and Cancer Therapies • Small Molecule Drug Manufacture
Industrial • Increased population growth impacting the sustainability of finite resources • Industrial production to address the needs of civilization	• Specialty Chemicals • Advanced Property Materials
Agriculture • Global population growing with decrease in per capita arable land • Food security and increased nutrition	• Self-fertilizing crops • Oil-Free Fertilizers • Drought Solutions • New Disease Protection

We need a new type of DNA supplier to meet demand

Rewriting DNA with the Power of Silicon



96 WELL PLATE
makes 1 gene



TWIST SILICON PLATFORM
can make 9,600 genes

Developing Game-Changing Throughput and Cost through Quality and Speed at Scale

Other Growth Verticals

TWIST'S PLATFROM EXTENDS TO

INITIAL MARKETS

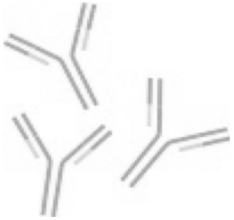
SYNTHETIC BIOLOGY

- Competitive Turnaround Time
- Lower Cost
- High Throughput
- High Quality

GENOMICS: TARGETED NGS

- Fast Customization
- Performance
- Full Kit
- High Quality

SHORT TERM GOAL
Grow Revenue



LARGE MARKET

DRUG DISCOVERY/ DEVELOPMENT

- High Quality Diversity Hits / Leads
- Shorter Time and Cost from Target to IND

MID TERM GOAL
Develop novel therapeutics

IT INDUSTRY

DATA STORAGE

- Permanence
- Density
- Ease of Copying
- Universal Format

LONG TERM GOAL
Enter technology market

WHY DNA?

DNA: Nature's Choice for Data Storage

**MAN-MADE,
NOT PERMANENT**

STABLE FOR 1000s of YEARS

**20,000
Years ago**

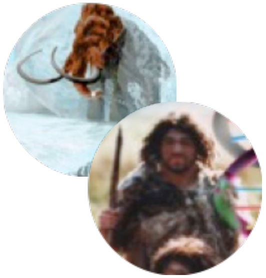
Sequencing the nuclear genome of the extinct woolly mammoth

Webb Miller[1], Daniela I. Drautz[1], Aakrosh Ratan[1], Barbara Pusey[1], Ji Qi[1], Arthur M. Lesk[1], Lynn P. Tomsho[1], Michael D. Packard[1], Fangqing Zhao[1], Andrei Sher[2]‡, Alexei Tikhonov[3], Brian Raney[4], Nick Patterson[5], Kerstin Lindblad-Toh[5], Eric S. Lander[5], James R. Knight[6], Gerard P. Irzyk[6], Karin M. Fredrikson[7], Timothy T. Harkins[7], Sharon Sheridan[7], Tom Pringle[8] & Stephan C. Schuster[1]

**40,000
Years ago**

A Draft Sequence of the Neandertal Genome

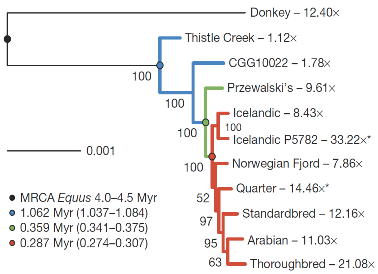
Richard E. Green,[1]*†‡ Johannes Krause,[1]†§ Adrian W. Briggs,[1]†§ Tomislav Maricic,[1]†§ Udo Stenzel,[1]†§ Martin Kircher,[1]†§ Nick Patterson,[2]†§ Heng Li,[2]† Weiwei Zhai,[3]†|| Markus Hsi-Yang Fritz,[4]† Nancy F. Hansen,[5]† Eric Y. Durand,[3]† Anna-Sapfo Malaspinas,[3]† Jeffrey D. Jensen,[6]† Tomas Marques-Bonet,[7,13]† Can Alkan,[7]† Kay Prüfer,[1]† Matthias Meyer,[1]† Hernán A. Burbano,[1]† Jeffrey M. Good,[1,8]† Rigo Schultz,[1] Ayinuer Aximu-Petri,[1] Anne Butthof,[1] Barbara Höber,[1] Barbara Höffner,[1] Madlen Siegemund,[1] Antje Weihmann,[1] Chad Nusbaum,[2] Eric S. Lander,[2] Carsten Russ,[2] Nathaniel Novod,[2] Jason Affourtit,[9] Michael Egholm,[9] Christine Verna,[21] Pavao Rudan,[10] Dejana Brajkovic,[11] Željko Kucan,[10] Ivan Gušic,[10] Vladimir B. Doronichev,[12] Liubov V. Golovanova,[12] Carles Lalueza-Fox,[13] Marco de la Rasilla,[14] Javier Fortea,[14]¶ Antonio Rosas,[15] Ralf W. Schmitz,[16,17] Philip L. F. Johnson,[18]† Evan E. Eichler,[7]† Daniel Falush,[19]† Ewan Birney,[4]† James C. Mullikin,[5]† Montgomery Slatkin,[3]† Rasmus Nielsen,[3]† Janet Kelso,[1]† Michael Lachmann,[1]† David Reich,[2,20]*† Svante Pääbo[1]*†



**560,000 - 780,000
Years ago**

Recalibrating *Equus* evolution using the genome sequence of an early Middle Pleistocene horse

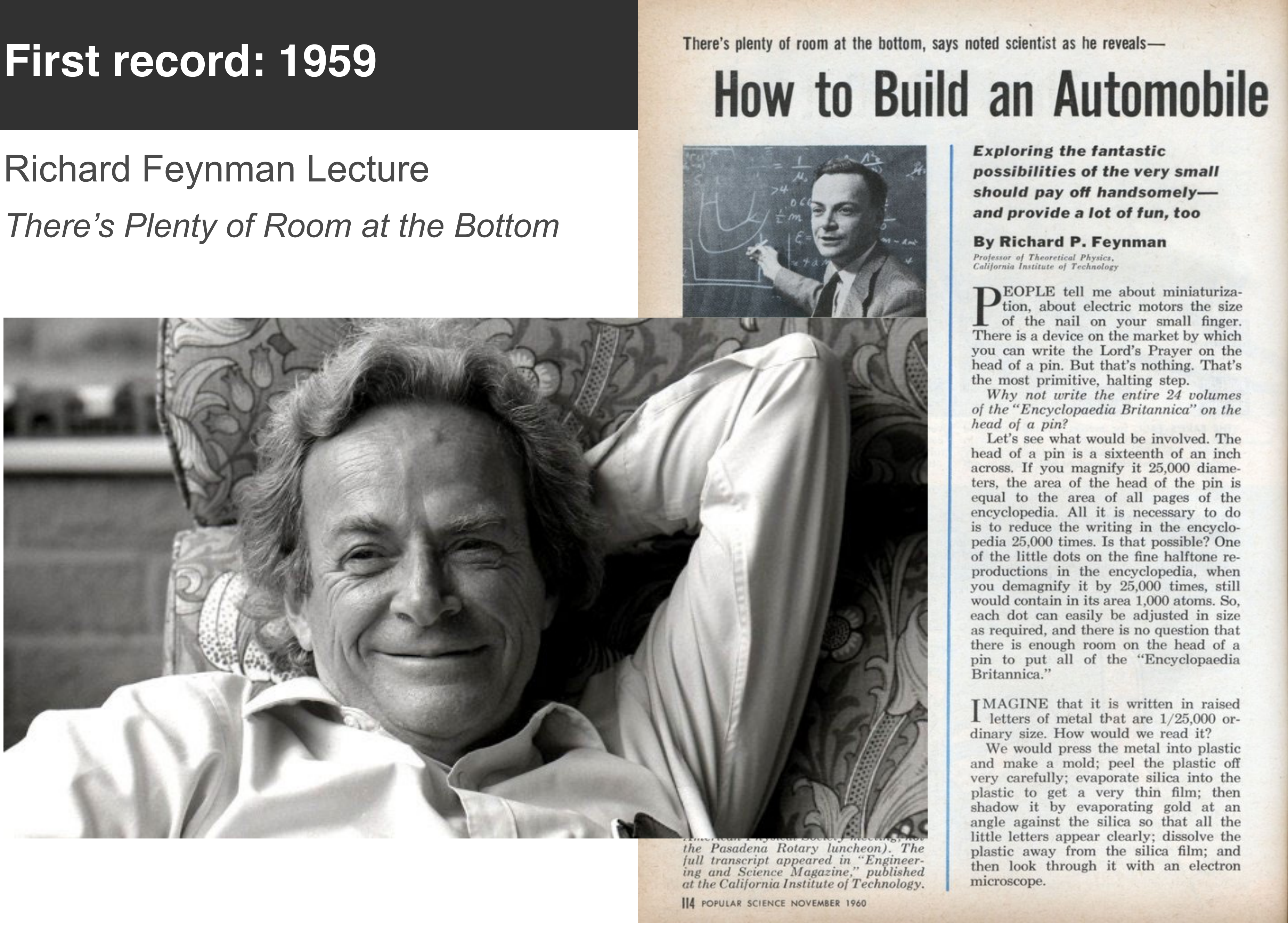
Ludovic Orlando[1]*, Aurélien Ginolhac[1]*, Guojie Zhang[2]*, Duane Froese[3], And
Enrico Cappellini[1], Bent Petersen[6], Ida Moltke[4,7], Philip L. F. Johnson[8], Matt
Thorfinn Korneliussen[1], Anna-Sapfo Malaspinas[1], Josef Vogt[9], Damian Szkla
Andrei Dolocan[12], Jesper Stenderup[1], Amhed M. V. Velazquez[1], James Cahill
Grant D. Zazula[13], Andaine Seguin-Orlando[1,14], Cecilie Mortensen[1,14], Kim
Jacobo Weinstock[16], Kristian Gregersen[1,17], Knut H. Røed[18], Véra Eisenman
Douglas F. Antczak[21], Mads F. Bertelsen[22], Søren Brunak[6,23], Khaled A. S. A
John Mundy[26], Anders Krogh[1,4], M. Thomas P. Gilbert[1], Kurt Kjær[1], Thomas
Jesper V. Olsen[10], Michael Hofreiter[27], Rasmus Nielsen[28], Beth Shapiro[5], Jun



First record: 1959

Richard Feynman Lecture

There's Plenty of Room at the Bottom



There's plenty of room at the bottom, says noted scientist as he reveals—

How to Build an Automobile

Exploring the fantastic possibilities of the very small should pay off handsomely—and provide a lot of fun, too

By Richard P. Feynman

Professor of Theoretical Physics, California Institute of Technology

PEOPLE tell me about miniaturization, about electric motors the size of the nail on your small finger. There is a device on the market by which you can write the Lord's Prayer on the head of a pin. But that's nothing. That's the most primitive, halting step.

Why not write the entire 24 volumes of the "Encyclopaedia Britannica" on the head of a pin?

Let's see what would be involved. The head of a pin is a sixteenth of an inch across. If you magnify it 25,000 diameters, the area of the head of the pin is equal to the area of all pages of the encyclopedia. All it is necessary to do is to reduce the writing in the encyclopedia 25,000 times. Is that possible? One of the little dots on the fine halftone reproductions in the encyclopedia, when you demagnify it by 25,000 times, still would contain in its area 1,000 atoms. So, each dot can easily be adjusted in size as required, and there is no question that there is enough room on the head of a pin to put all of the "Encyclopaedia Britannica."

IMAGINE that it is written in raised letters of metal that are 1/25,000 ordinary size. How would we read it?

We would press the metal into plastic and make a mold; peel the plastic off very carefully; evaporate silica into the plastic to get a very thin film; then shadow it by evaporating gold at an angle against the silica so that all the little letters appear clearly; dissolve the plastic away from the silica film; and then look through it with an electron microscope.

the Pasadena Rotary luncheon). The full transcript appeared in "Engineering and Science Magazine," published at the California Institute of Technology.

114 POPULAR SCIENCE NOVEMBER 1960

Why DNA?

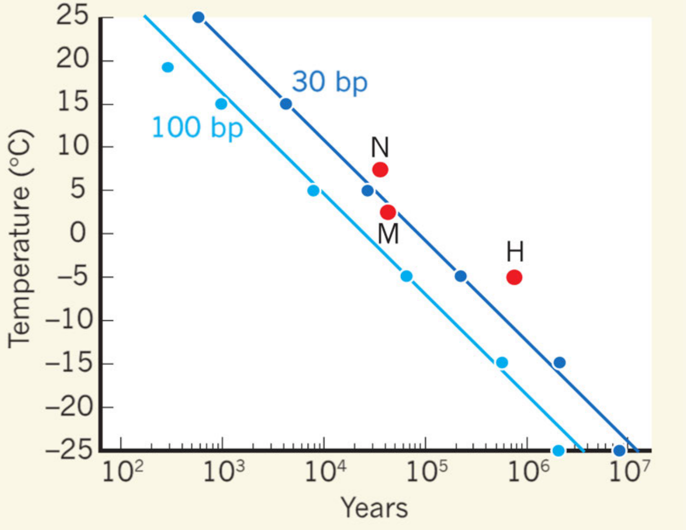
- Terminal media format,

  independent of read or write platform
- Permanence
- Density
- Copying
- Data Transfer

Towards a Million-Year-Old Genome

The rate of DNA decay varies with environmental conditions

Survival of the Coldest



N Neanderthal
M Woolly Mammoth
H Middle Pleistocene Horse

Millar and Lambert, *Nature* **499, p 34-35, 2013**

The Molecular Weight of Information

- **Compare Three Storage Media**
  - ***Hot Storage: Flash***
    - ***10^{10} Atoms per bit***
  - ***Cold Storage: Tape***
    - ***10^{10} Atoms per bit***
  - ***NAM: Nucleic Acid Memory***
    - ***14 Atoms per Bit***

The media is the information

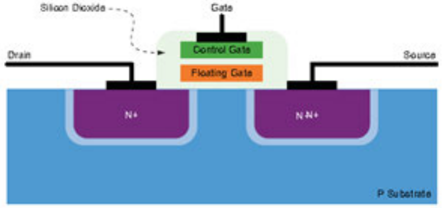
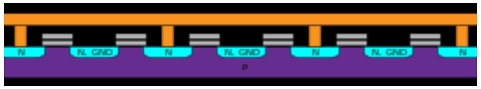
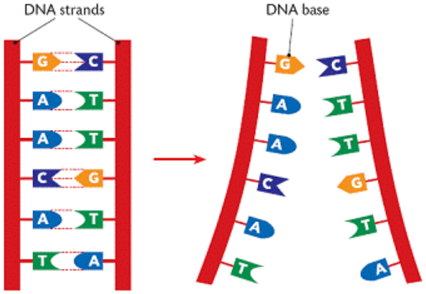


Figure 2. Flash Memory Cell





Data Transfer

Amazon Snowmobile — up to 100 PB

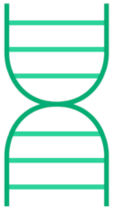
HOW?

Data Storage in DNA

1 Coding

00 → A
01 → G
10 → C
11 → T

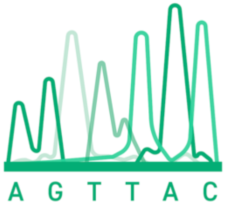
2 Synthesis



3 Storage

4 Retrieval

5 Sequencing



6 Decoding

A → 00
G → 01
C → 10
T → 11

Permanence • **Density** • **Random Access** • **Universal format**

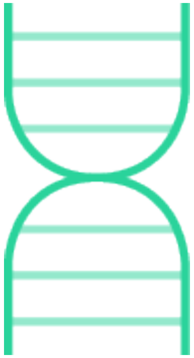
Coding

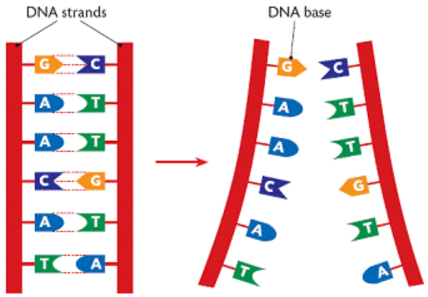
- Transform Binary information to Nature's code
- Information Theory ECC's adapted to correct for errors and provide redundancy

00 → A
01 → G
10 → C
11 → T

Synthesis (Write)

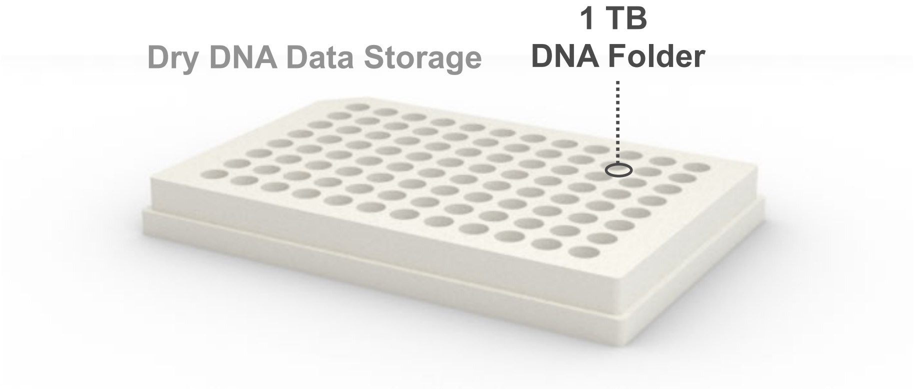
- Artificially assemble DNA from the chemical building blocks
- Molecule is identical to DNA occurring in living organisms but is created synthetically
- Similar to assembling Lego® blocks from 4 colors
- We exploit the molecules’ complementary nature





Storage

- Only requires a cool dry vessel to keep the data viable for hundreds of years
- Thousands of times more compact than current media



File Structures with DNA

GeopByte 10^{30}

1 Byte

kB

MB

GB

TB

PB - 10^{15}

EB

ZB

YB

BrontoByte 10^{27}

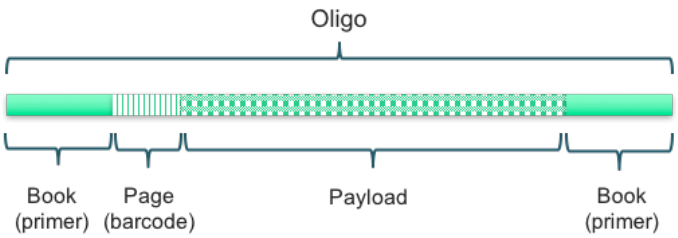
?

Molecular Addressing

Use Case Dependent

Hardware Addressing

Pools of DNA Molecules with Barcode



Oligo

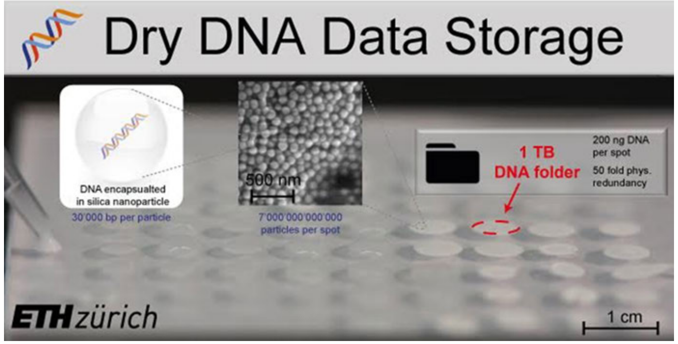
Book (primer)

Page (barcode)

Payload

Book (primer)

Addressable Objects for Storage



Dry DNA Data Storage

DNA encapsualted in silica nanoparticle

30'000 bp per particle

500 nm

7'000'000'000'000 particles per spot

1 TB DNA folder

200 ng DNA per spot

50 fold phys. redundancy

ETH zürich

1 cm

Sequencing (Read)

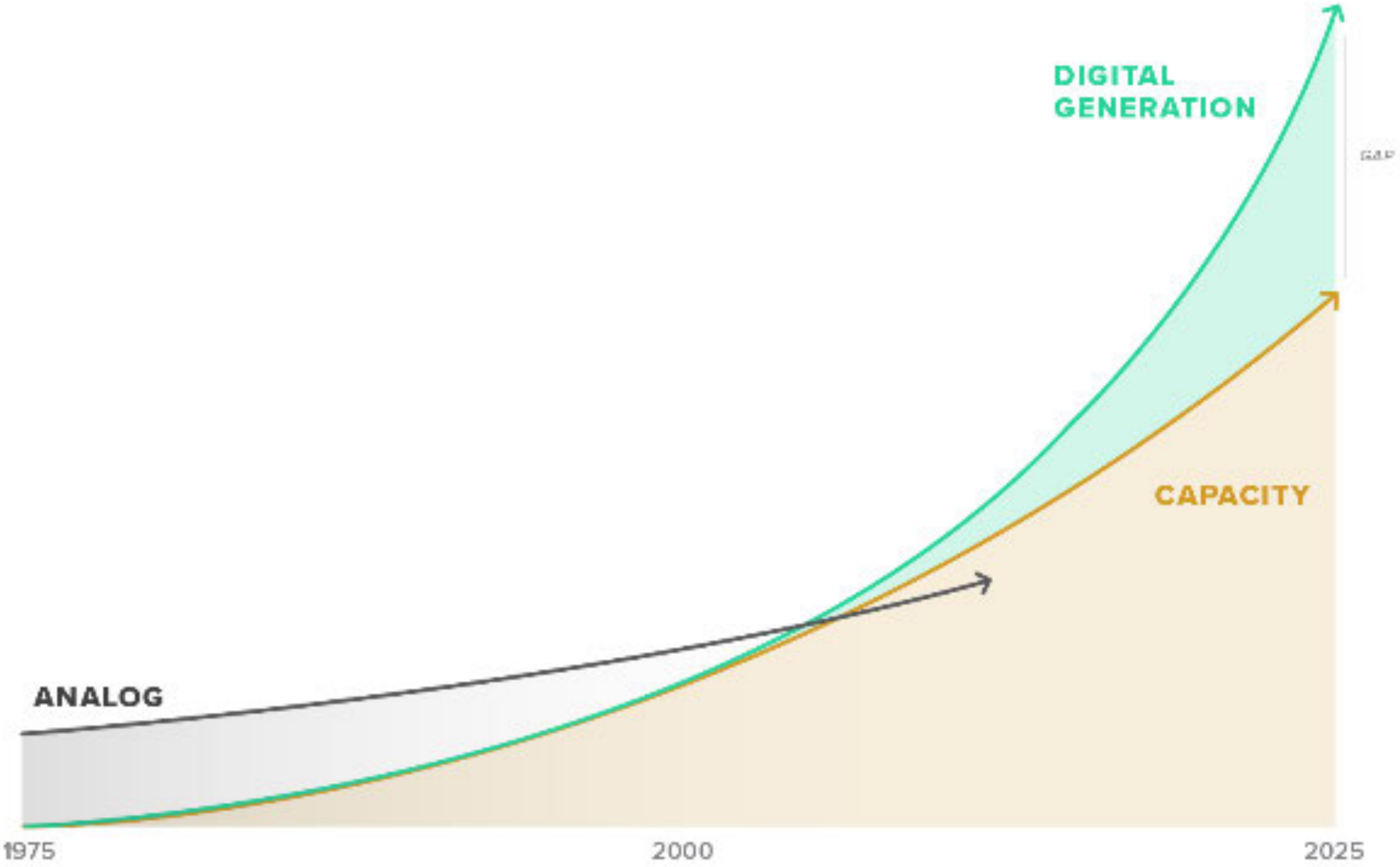
Decoding

- Reconstruct the digital file from sequences of DNA
- Computational methods remove errors that may be introduced in from the synthesis (write) and sequencing (read)

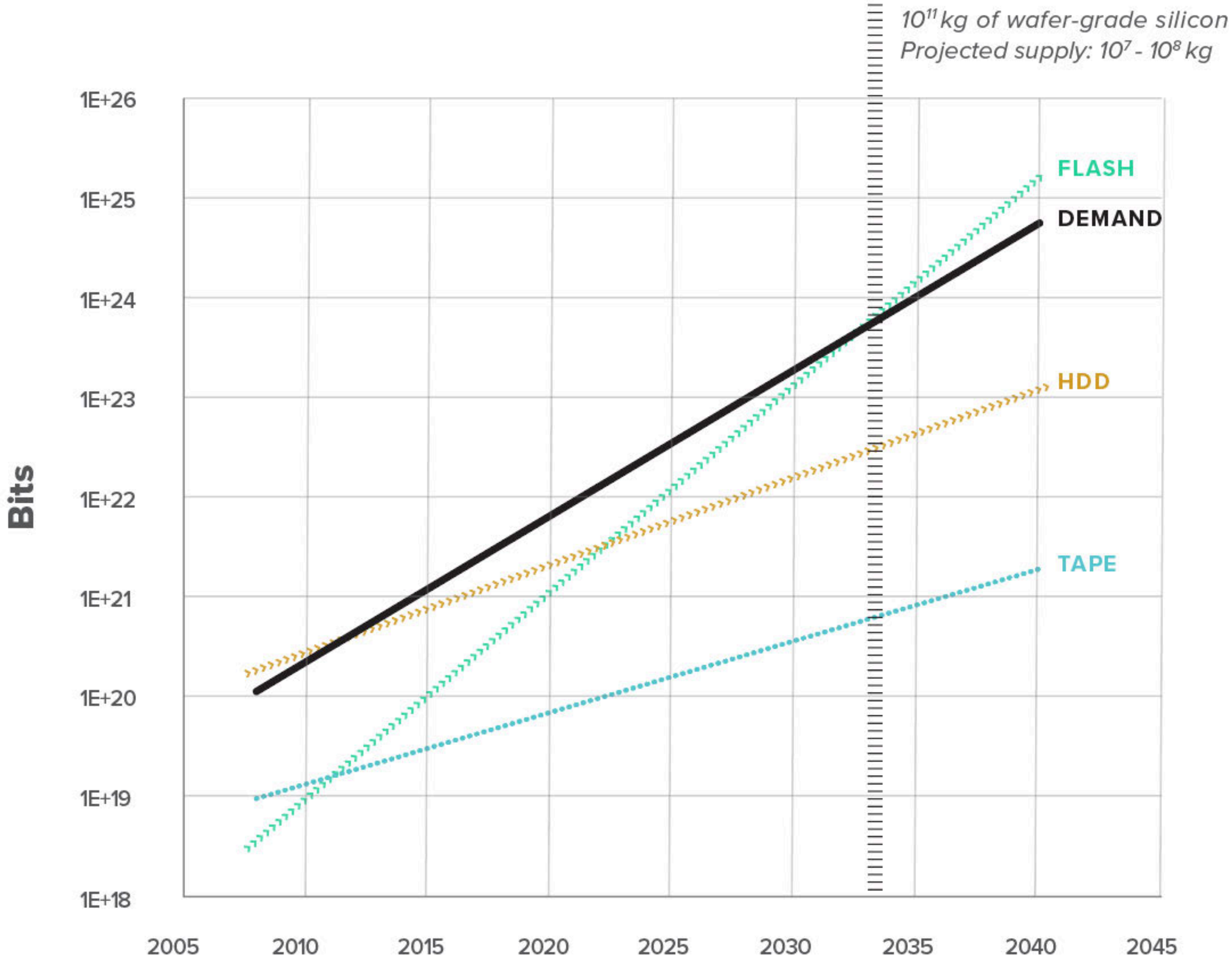
A → 00
G → 01
C → 10
T → 11

WHY NOW?

Storage Capacity

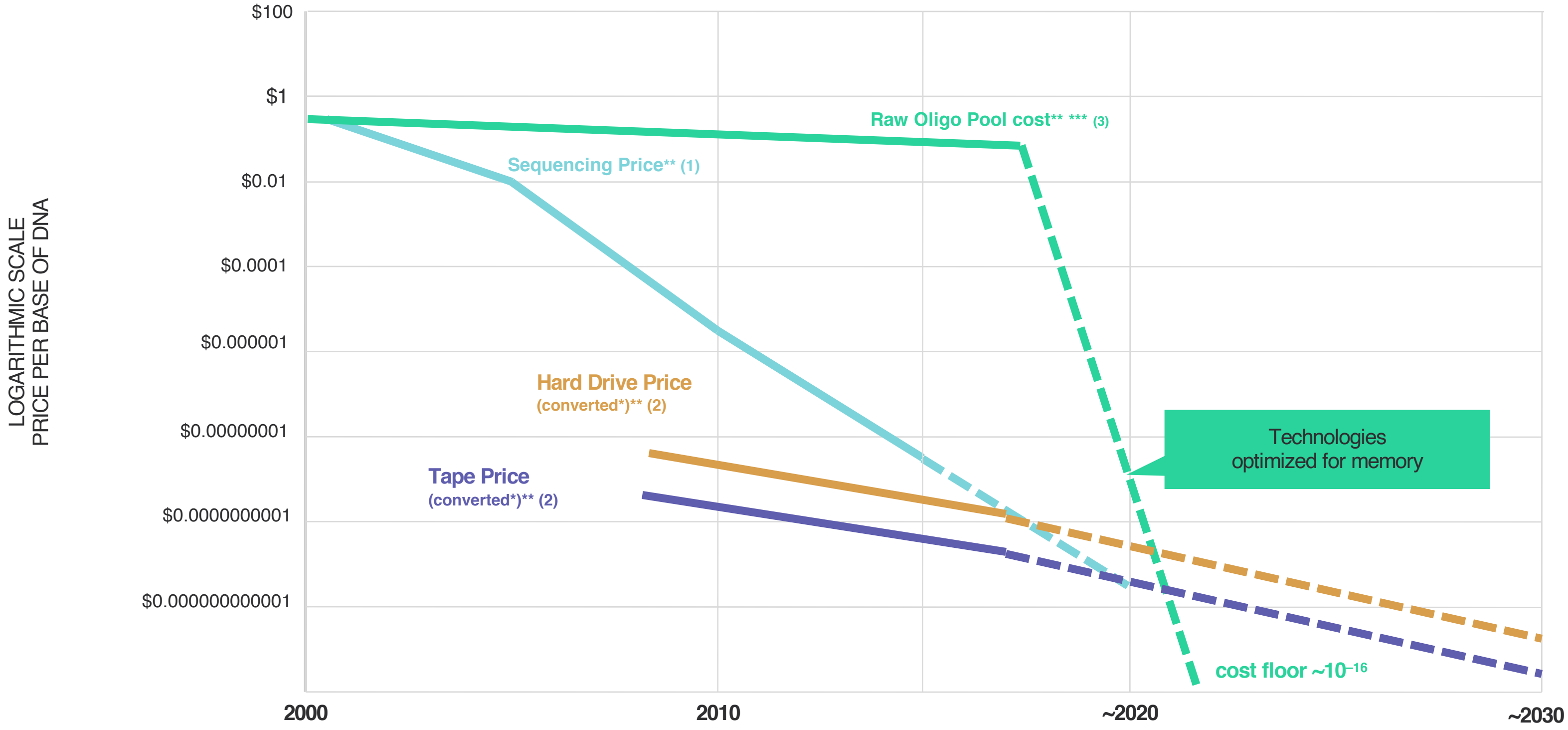


Storage Capacity



DNA Data Storage Trends and Projections

We believe new DNA technologies and cost efficiencies could surpass mature IT hardware solutions in 3–5 years



---

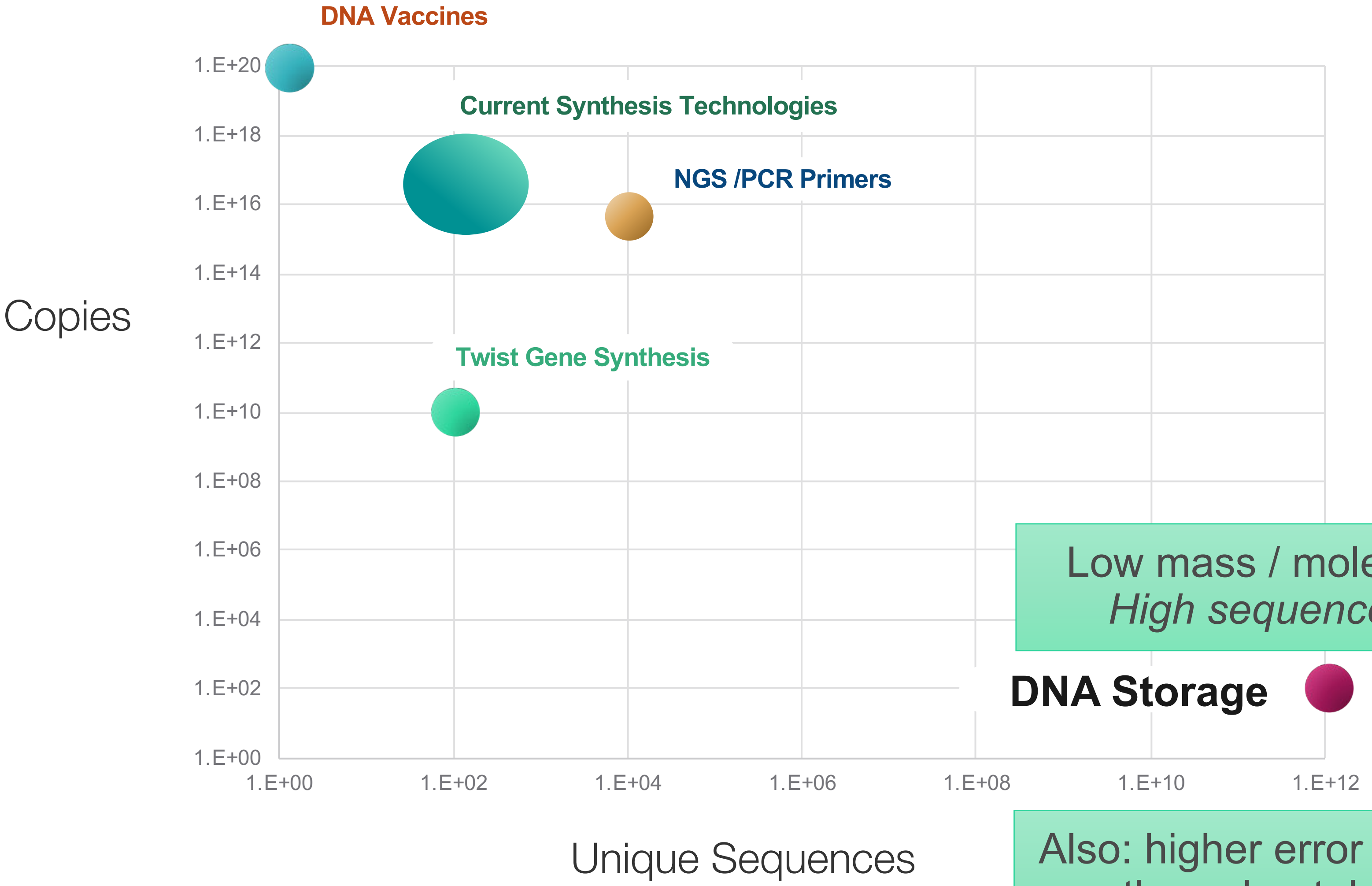
* DNA bases per byte for hard drive and tape shown at typical published encoding ranges from about 5:1 to 6.25:1

** All dotted lines represent extrapolations and assumes continued trajectory of historical trends, and that there will be continued decrease in price as technology improves.

*** Raw oligo pool cost extrapolation based on DARPA and another anticipated government-sponsored grant project objectives, both at specified time points

(1) www.genome. Gov (2) Bob Fontana, IBM Systems, Storage Media Overview, May 4,2016 (3) Bioeconomy Capital, Rob Carlson, January 20, 2018, www.synthesis.cc

How is DNA for Memory different from Genes



Low mass / molecule count
High sequence diversity

Also: higher error rate, higher throughput, lower cost

RECENT PROGRESS

Timeline of investment

2012–13	first papers & experiments on DNA memory (including our CEO)
2013	SRC establishes SemiSynBio (3 years, $2.5M) Twist Founded in April
2014-16	NIST-funded SemiSynBio Industry Roadmap ($0.5M)
2016	IARPA/SRC workshop on Exascale DNA Memory Twist engagement with Microsoft through SRC meeting
2017	DARPA program on Molecular Computing announced IARPA/SRC workshop to develop MIST program
2018	IARPA MIST program formally announced IARPA MIST proposal submission

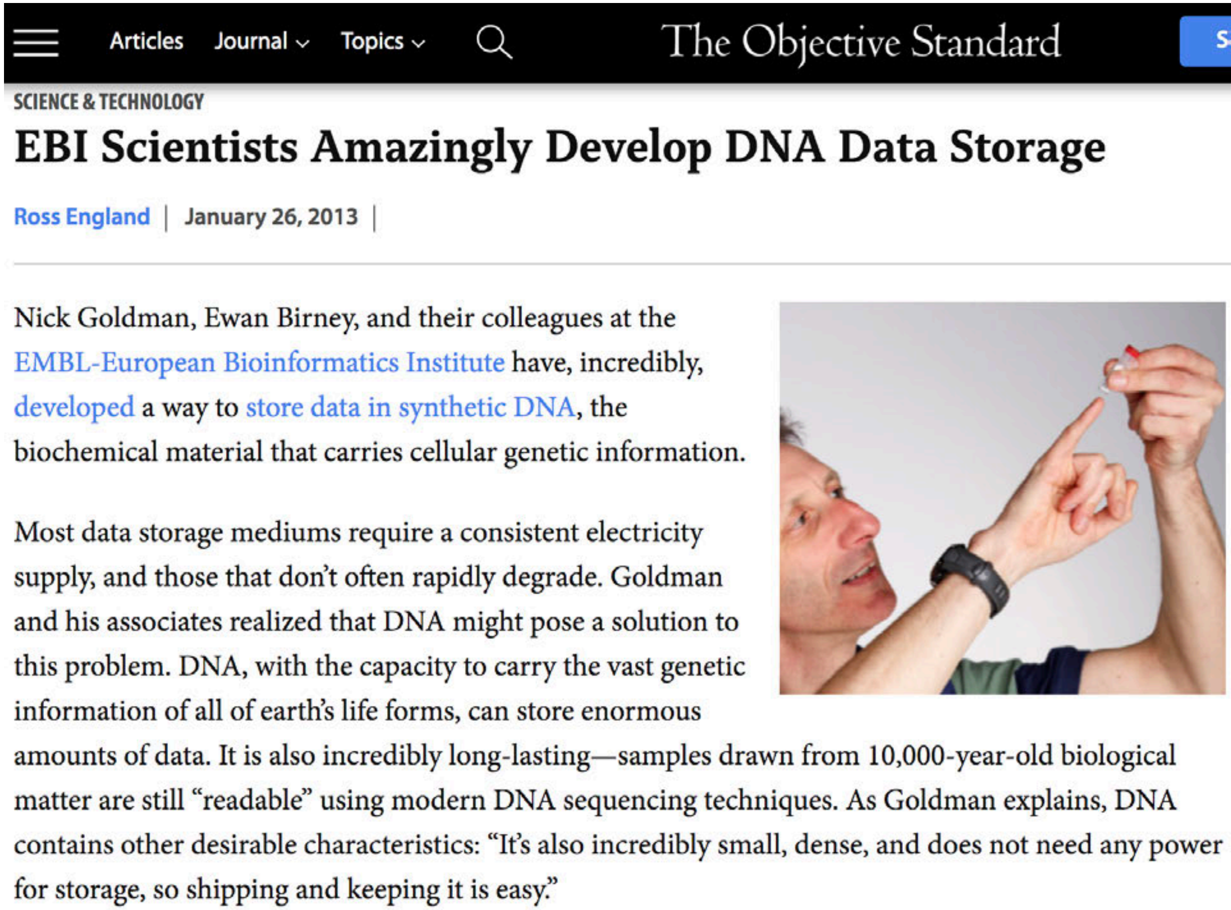
Twist Selected & Awarded for DARPA program $2.5M

IARPA subcontract in process

Encoding at Scale: first Papers (2012/2013)

Harvard & EBI

dealing with not just putting information into DNA, but techniques of recovery



SCIENCE & TECHNOLOGY

EBI Scientists Amazingly Develop DNA Data Storage

Ross England | January 26, 2013 |

Nick Goldman, Ewan Birney, and their colleagues at the EMBL-European Bioinformatics Institute have, incredibly, developed a way to store data in synthetic DNA, the biochemical material that carries cellular genetic information.

Most data storage mediums require a consistent electricity supply, and those that don't often rapidly degrade. Goldman and his associates realized that DNA might pose a solution to this problem. DNA, with the capacity to carry the vast genetic information of all of earth's life forms, can store enormous amounts of data. It is also incredibly long-lasting—samples drawn from 10,000-year-old biological matter are still "readable" using modern DNA sequencing techniques. As Goldman explains, DNA contains other desirable characteristics: "It's also incredibly small, dense, and does not need any power for storage, so shipping and keeping it is easy."

SRC & SemiSynbio (starting 2013)

February 2013

SRC establishes SemiSynBio (3 years, $2.5M)

Pre-competitive academic research, intersection of synthetic biology and semiconductors/computing

Five topics, including DNA memory

Second generation of funding with NSF started ($4.5M)

SRC Roadmap (2014–2016+)

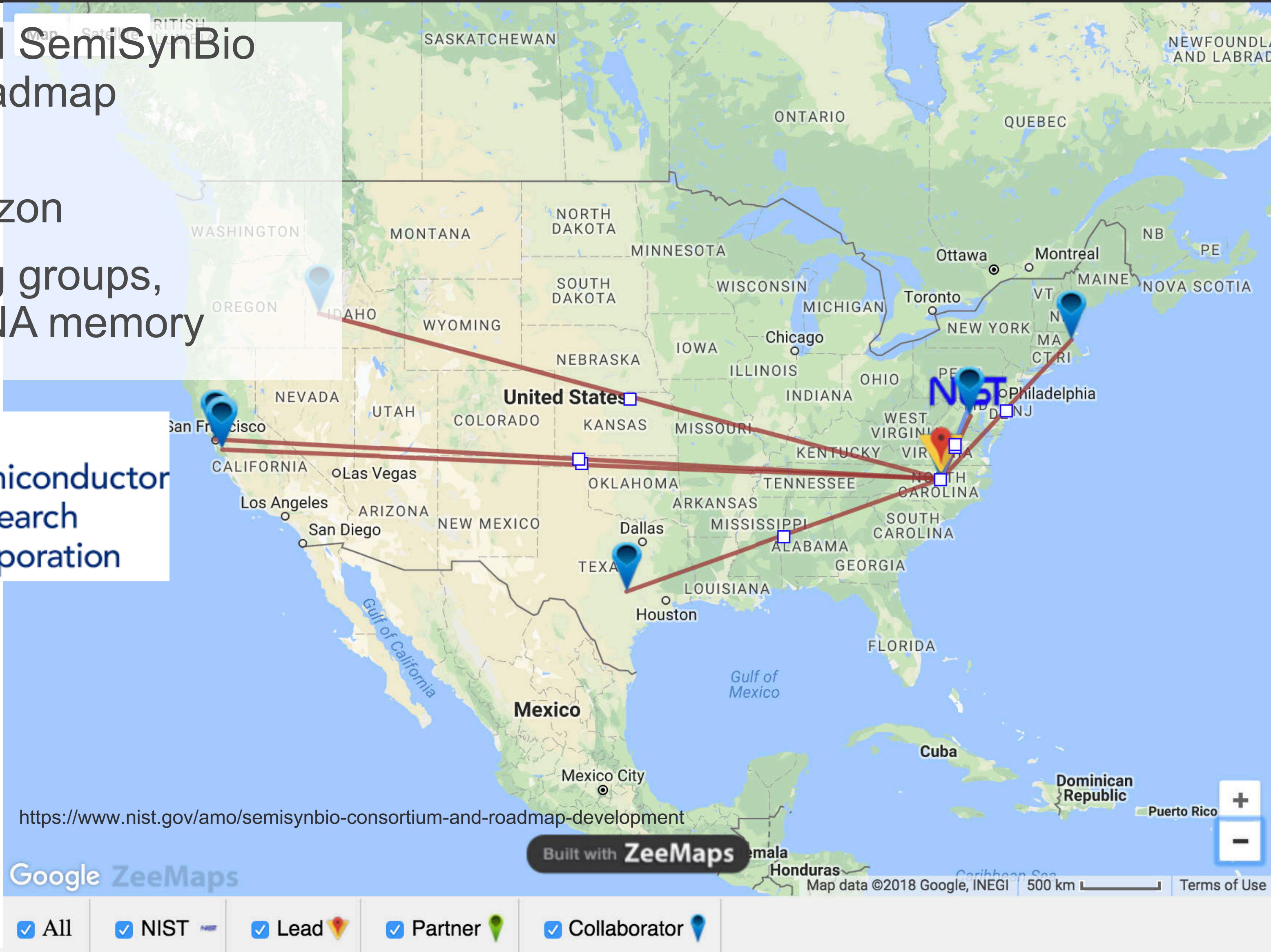
NIST-funded SemiSynBio Industry Roadmap

Planning for 15-year horizon

Five working groups, including DNA memory

Semiconductor Research Corporation

National Institute of Standards and Technology



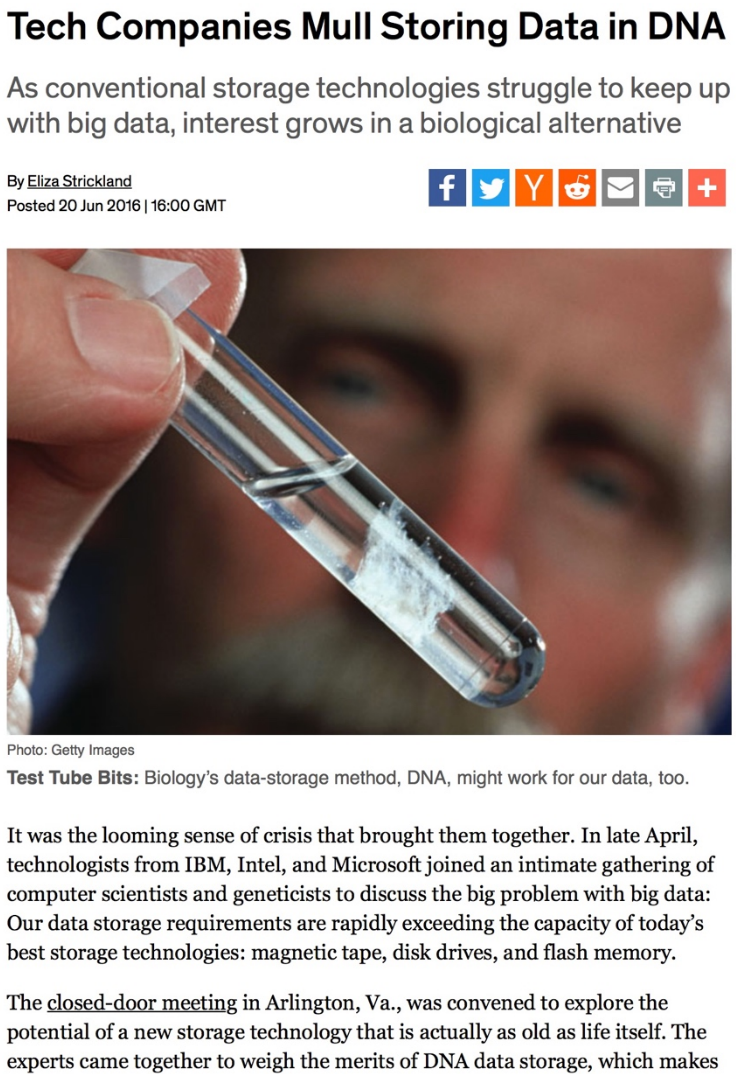
https://www.nist.gov/amo/semisynbio-consortium-and-roadmap-development

SRC/IARPA Workshop (April 2016)

Feasibility of Exascale storage system

Pre-cursor to MIST program (2018)

http://spectrum.ieee.org/biomedical/devices/tech-companies-mull-storing-data-in-dna



Tech Companies Mull Storing Data in DNA

As conventional storage technologies struggle to keep up with big data, interest grows in a biological alternative

By Eliza Strickland
Posted 20 Jun 2016 | 16:00 GMT

Photo: Getty Images

Test Tube Bits: Biology's data-storage method, DNA, might work for our data, too.

It was the looming sense of crisis that brought them together. In late April, technologists from IBM, Intel, and Microsoft joined an intimate gathering of computer scientists and geneticists to discuss the big problem with big data: Our data storage requirements are rapidly exceeding the capacity of today's best storage technologies: magnetic tape, disk drives, and flash memory.

The closed-door meeting in Arlington, Va., was convened to explore the potential of a new storage technology that is actually as old as life itself. The experts came together to weigh the merits of DNA data storage, which makes

DARPA Molecular Informatics

Announced April 2017, Awarded 2018

The Molecular Informatics program brings together a collaborative interdisciplinary community to explore completely new approaches to store and process information with molecules. Chemistry offers an untapped, rich palette of molecular diversity that may yield a vast design space to enable dense data representations and highly versatile computing concepts outside of traditional digital, logic-based approaches.

Given radical advances in tools and techniques to sense, separate, and manipulate at the molecular scale, what innovations can be injected into information technology, and what will the resulting systems be able to “compute”?

By addressing a series of mathematical and computational problems with molecule-based information encoding and processing, Molecular Informatics aims to discover and define future opportunities for molecules in information storage and processing.

#MemoriesInDNA

Program Goals
100 MB read/write in 1 day
from archive of 1GB

https://www.darpa.mil/program/molecular-informatics
https://www.darpa.mil/news-events/2017-03-23

IARPA Molecular Information Storage (MIST)

Office of the Director of National Intelligence

IARPA

BE THE FUTURE

The goal of the MIST program is to develop deployable storage technologies that can eventually scale into the exabyte regime and beyond with reduced physical footprint, power and cost requirements relative to conventional storage technologies.

MIST seeks to accomplish this by using sequence-controlled polymers as a data storage medium, and by building the necessary devices and information systems to interface with this medium.

Technologies are sought to optimize the writing and reading of information to/from polymer media at scale, and to support random access of information from polymer media archives at scale.

https://www.iarpa.gov/index.php/research-programs/mist

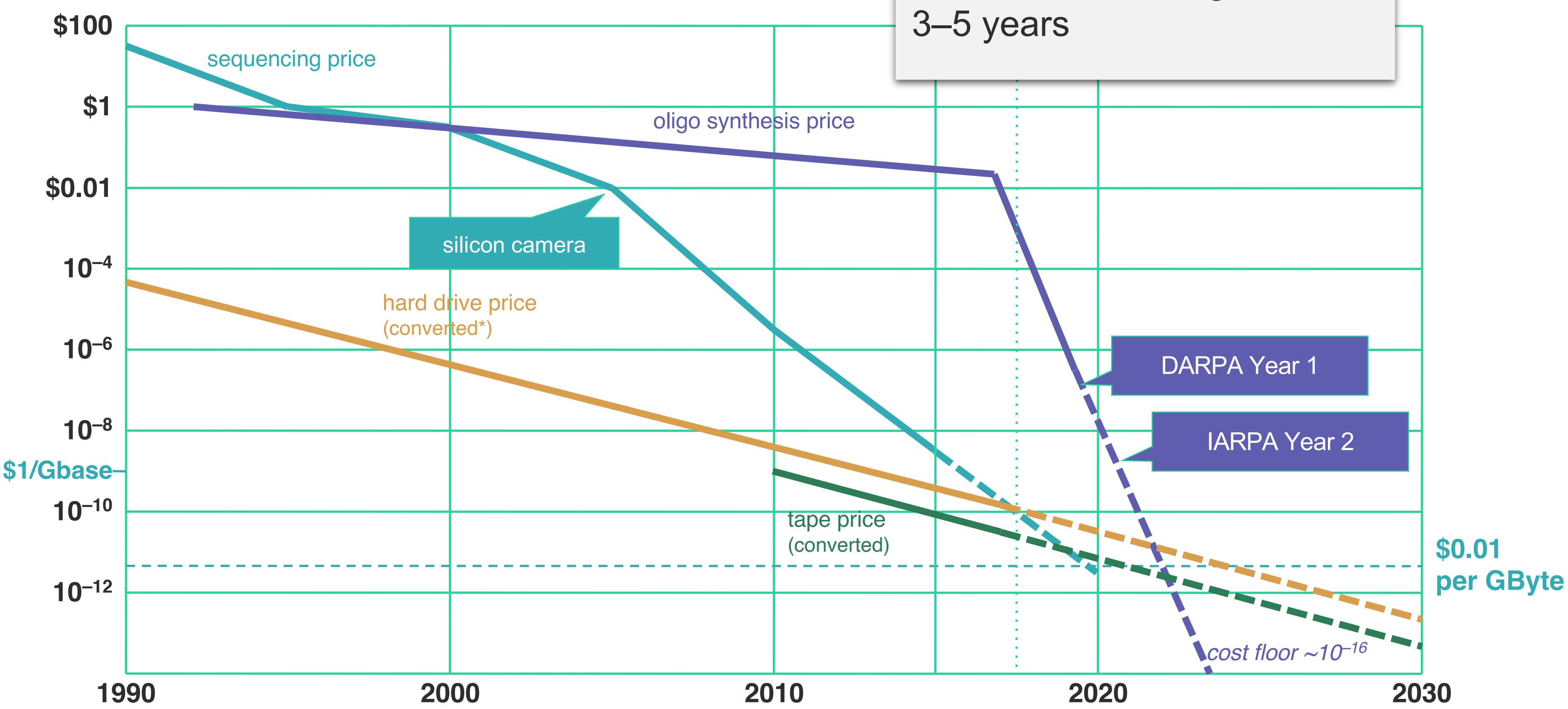
Program Goals
2 years: 10GB write, 1TB read
4 years: 1TB write system

+ PB/EB scalability plan

Trends in DNA Technology

Improvement of mature IT components intercepted by new DNA technologies in 3–5 years

Price of DNA read/write ($/base)



* DNA bases per byte conversion shown at typical published encoding ranges from about 5:1 to 6.25:1

PROJECT HIGHLIGHTS

Memory of the World

UNESCO

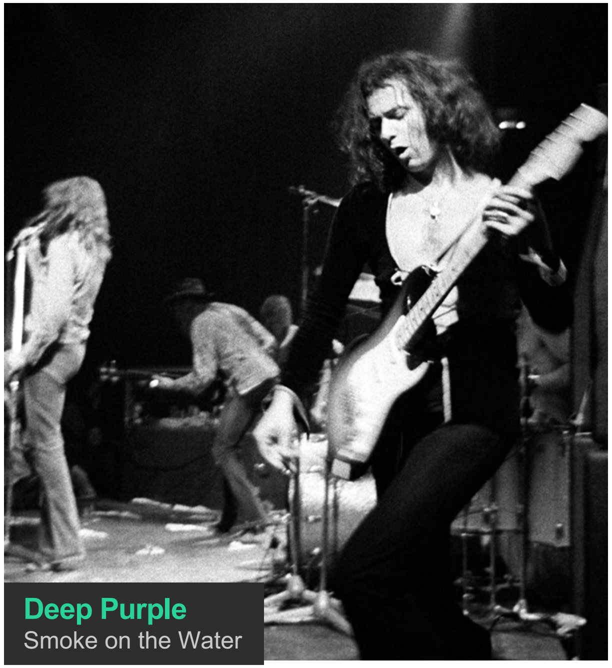
United Nations
Educational, Scientific and
Cultural Organization

Memory of
the World

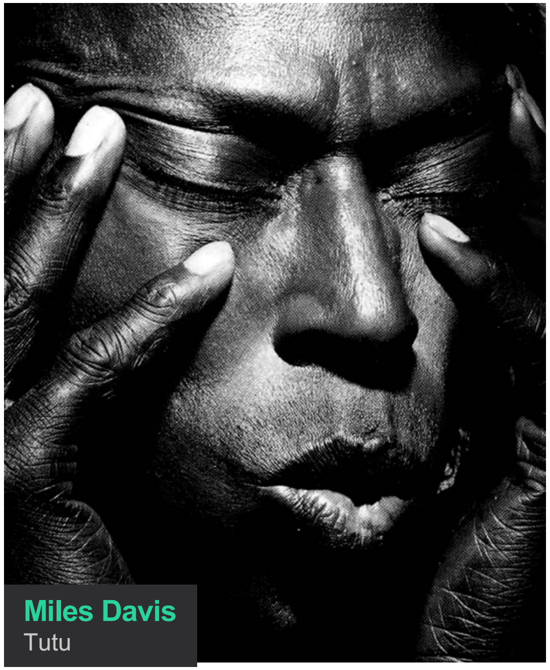
Montreux
Jazz Festival

The Treasures that Record our History from 1700 BC to present day

Our Cultural Heritage



Deep Purple
Smoke on the Water



Miles Davis
Tutu

W
UNIVERSITY of
WASHINGTON

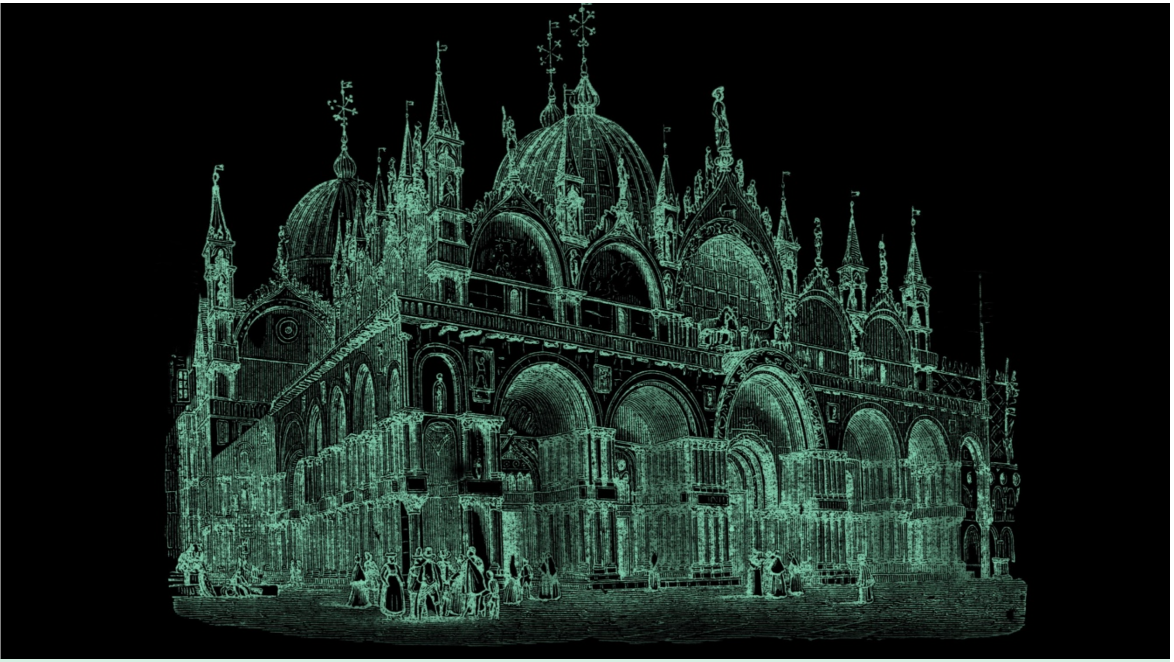
TWIST
BIOSCIENCE

Microsoft

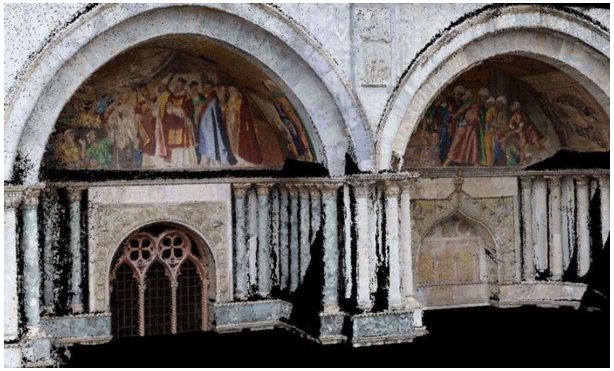
EPFL
ÉCOLE POLYTECHNIQUE
FÉDÉRALE DE LAUSANNE

Claude Nobs Foundation

Digitized Three-Dimensional Map of St Mark's Basilica Preserved in DNA



Time Machine
St Mark's Basilica



Microsoft

UNIVERSITY of WASHINGTON

TWIST BIOSCIENCE

Arch Mission Foundation Announces Digital Data Stored in DNA headed to the Lunar Library

ARCH MISSION

Microsoft

UNIVERSITY of WASHINGTON

BIOSCIENCE

TWIST BIOSCIENCE

Powering the Synthetic Biology Revolution