



# DNA Data Storage Annual Reality Check

LOC Conference March 13<sup>th</sup>, 2022

Daniel Chadash





# DNA Data Storage Introduction & Status



# Introducing DNA's benefits / What DNA is and what it is not

- DNA is a new, complementary cold layer in the storage pyramid
- Initially, DNA can be the ideal medium for a 3<sup>rd</sup> copy (offline, secured, durable)
- It can last thousands of years with the right protection / packaging
- As natural DNA, It will always be readable
- No migration and minimal maintenance costs = attractive TCO
- DNA is a great way to diversify your media, mitigating the risk of a single tape supplier
- DNA is not a tape replacement (Sorry about that!)
- DNA is not going to store all the world's data in a shoebox
- DNA is not a hot/warm storage medium
- DNA is not cheap (yet)
- DNA is not coming to your nearby data center in the next 2 years



# What we need to build

Codec	Synthesis	Storage / Retrieval	Sequencing	System
No vendor lock-in	Synthesize MB	No cooling requirements	Sequence MB	Integration with storage management software
Error detection / correction	Synthesize GB	Safe & durable medium	Sequence GB	Support object storage and S3 APIs
	Synthesize TB	Easily copied for geographical distribution	Sequence TB	Random Access
	TB < \$1000	Compact footprint	TB < \$1000	On-Premise deployment
	Throughput of TBs/day	No manual handling	Non-destructive sequencing	Automated storage system (e.g., Tape Library)
	Always accessible (Natural DNA)	Adjustable "Automation ready" storage container	TBs/day throughput	"Fixity checks"
	Enzymatic Synthesis			Data Center "friendly" deployment



# What we have built

Codec	Synthesis	Storage / Retrieval	Sequencing	System
No vendor lock-in	Synthesize MB	No cooling requirements	Sequence MB	Integration with storage management software
Error detection / correction	Synthesize GB*	Safe & durable medium	Sequence GB	Support object storage and S3 APIs
	Synthesize TB	Easily copied for geographical distribution	Sequence TB	Random Access
	TB < \$1000	Compact footprint	TB < \$1000	On-Premise deployment
	Throughput of TBs/day	No manual handling	Non-destructive sequencing	Automated storage system (e.g., Tape Library)
	Always accessible (Natural DNA)	Adjustable "Automation ready" storage container	TBs/day throughput	"Fixity checks"
	Enzymatic Synthesis*			Data Center "friendly" deployment

\* Proof of concept



# Solving The Rest of The Puzzle

**DNA Data Storage  
Technology Progress**



# Goal is to develop a chip that produces 1 TB of coded DNA

- DNA is synthesized on a chip
  - Use a 2D array of electrochemical reactors to synthesize strands of DNA
  - After synthesis, the DNA is washed into a tube, then amplified, purified, and packaged
- Chip capacity is limited by the array pitch and chip size
  - There is a scaling limit; each reactor needs to produce enough DNA to practically store
  - Given the scaling limit, 1 TB from a chip is the practical limit – otherwise the chip becomes too large
- Twist's chip capacity roadmap
  - 62.5 GB → 250 GB → 1 TB
  - *We are working on the 62.5 GB chip*
- Synthesis cost drivers
  - At scale, the fixed cost is averaged over many units
  - Variable cost is driven by reagent prices and reaction volume – high density chips yield lowest cost



# Packaging

- DNA degrades by oxidation
  - Hermetically packaging DNA leads to a long shelf life
  - The package can be checked periodically for leaks – no leaks, no degradation
- DNA is dense, but packaging needs to be practical
  - Industrial automation required for process steps
  - And tubes that can be laser welded shut
- Barcoded tubes can be packed in arrays
  - Arrays are configurable
  - Array sizes: 96 TB, 384 TB, or 1,536 TB per bio automation spec

Imagene's DNASHELL



7.5mm x 18mm



96 TB Array





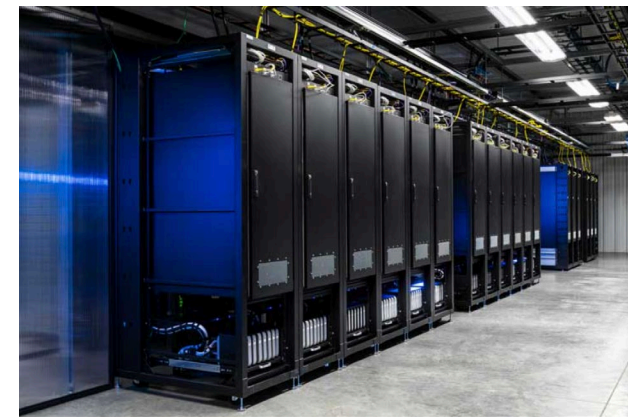
# Sequencing

- When synthesis is solved, sequencing becomes the most pressing challenge
- Overall sequencing cost depends on reading frequency
- Genomic sequencers are a non-starter; orders of magnitude too expensive
- Multiple groups working on molecular electronics sequencing



# System

- So far, the focus has been on the components as enablers
- We have phenomenal technologies, but not yet a product
- The product would have to integrate seamlessly into the customer workflow
  - Operated by IT team / Offered as a service
  - Integrated with storage management software
  - Data in/out
  - Data Center Ready – Monitoring, Debugging, Alerts, etc.
  - Implement common APIs and File Systems
  - Fixity Check solution





# Building the DNA Data Storage ecosystem

## History

- Formed on October 12th, 2020 by Illumina, Microsoft, Twist and Western Digital
- More than 50 member organizations (Including leading storage vendors: Seagate, WD, Kioxia, Dell, Quantum, Fujifilm, Fujitsu, Spectra Logic and Microsoft)

## Mission

- Create and promote an interoperable storage ecosystem based on DNA as a data storage medium

## Scope

- Educate the DNA data storage market to create awareness and adoption
- Identify use cases in various markets/industries for the use of DNA data storage
- Develop an industry technology roadmap for DNA data storage
- Foster standards or specifications as needed by ecosystem





CAAGCAAGATACGATACACGAGCATCGCATGGACTACAGGCATC  
CAGCAGCTACGACTAGATATATCTACACGAGCAGAAATCATAGATC  
GAGAGAGCGGATGAGGGATTACTAGCATGATAGATAGCTAGC  
TAGCAGCACACTATGAGCGGGAACGGGCAGACTAAGAGAGGTA  
CGAGAGAGAGACGAATCGATCCGAGCTAGCTAGGAGTGAGTGG  
ATATACGATATGGCTACTACGATCGACTAGTATCAGCTAGATC  
GAGAGAGCGCGAAGACGGATTACTAGCATCAGCTGATAGCTAG  
AGCCAGGACACTATCAGCGCTTACAGCACTATCATCCGAGAGGC  
ATAGCATGATATCGAGGGCGGATGAGCAGCTAAGCTAATAAT  
ATCCGAGAGATCATCGGCTGATCAGCAGTCTACTAGTCTAGACAG  
GATATCATGGAGATCTACAGCTATATATATATCCGCCATAGAGC  
GAGAGAGGGCGCATGACGGATTAGTAGCATGATCGATAGCTAGC