BIOTECH 201

How can Biotech...?

Week 1 - February 13, 2017

HUDSONALPHA INSTITUTE FOR BIOTECHNOLOGY

Our journey ahead

Feb 13	solve a crime? store information in DNA?
Feb 20	change my DNA? create “designer babies”?
Feb 27	help me lose weight? predict my death?
March 6	help us reach and colonize Mars?

How can Biotech...

solve a crime?

DNA based Forensic Testing

the identification of persons by matching DNA profiles of unknown evidence material with reference material from known persons

"the gold standard in forensic sciences"

utilizes segments of DNA called **STRs** (short tandem repeats)

6 repeats

```
...ACGTACATGCATGCATGCATGCATGCATGCATGAGTAACAGAT...
...TGCATGTACGTACGTACGTACGTACGTACGTACTCATTGTCTA...
```

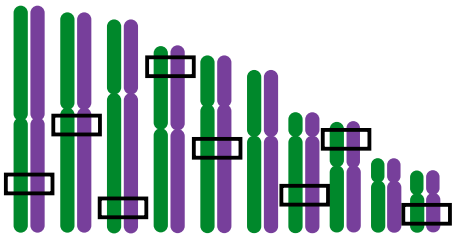
```
...ACGTACATGCATGCATGCATGCATGAGTAACAGAT...
...TGCATGTACGTACGTACGTACGTACTCATTGTCTA...
```

(4,6)

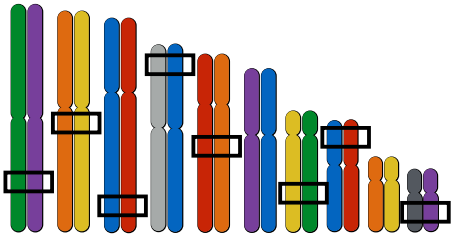
4 repeats

DNA based Forensic Testing

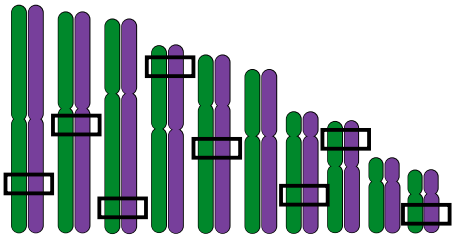
DNA analyzed at 20 STR regions across the genome - all are non-coding DNA



DNA from a crime sample



Suspect 1



Suspect 2

Genetic Marker	Crime Sample	Suspect 1	Suspect 2
D2S1338	8,12	8,12	8,12
D5S1358	11,16	10,12	11,16
D8S1179	4,14	8,16	4,14
D19S433	9,11	11,16	9,11

DNA based Forensic Testing

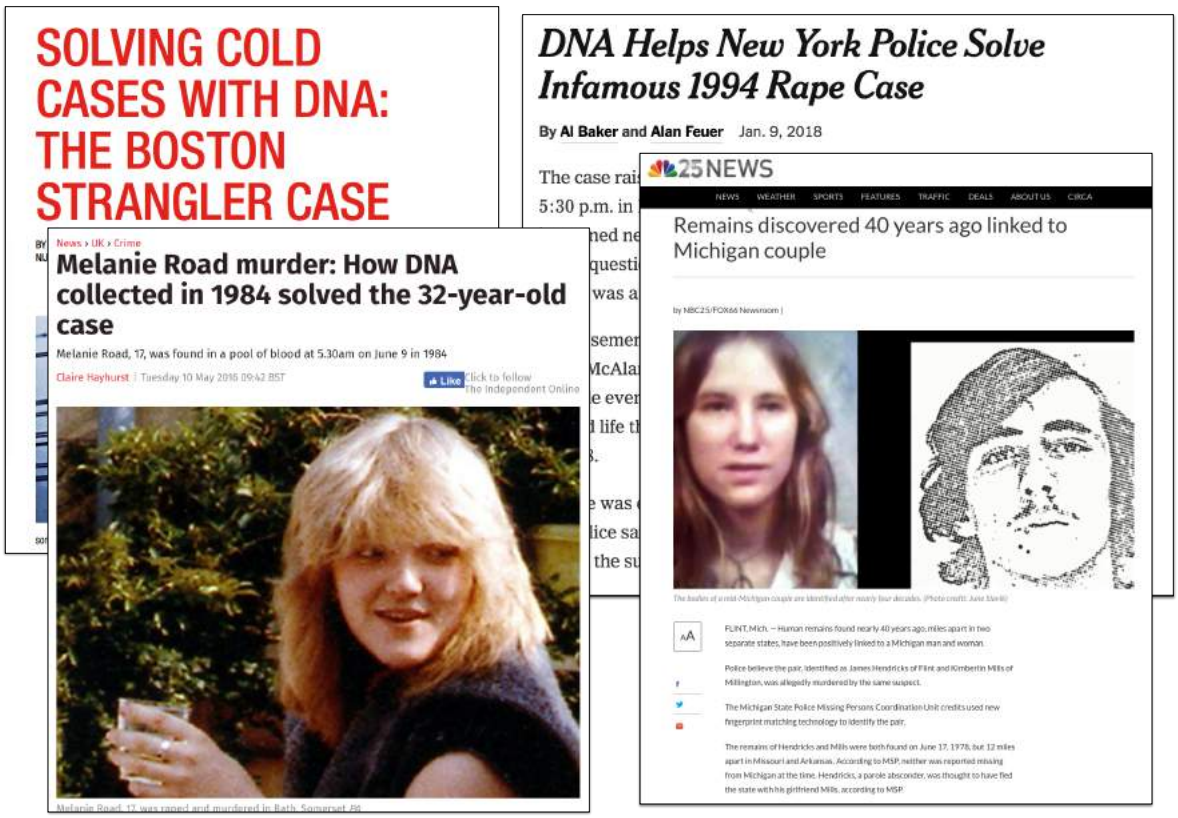
DNA profile uploaded to the Combined DNA Index System (CODIS)

collection of US national DNA databases:

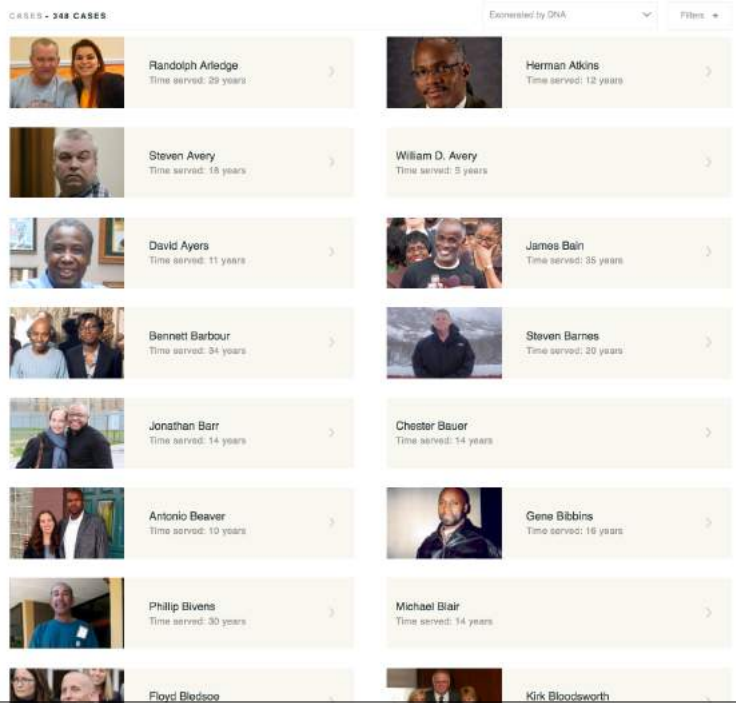
- missing persons
- convicted offenders
- forensic samples from crime scenes

as of Dec, 2017:

- > 13.1M offender profiles
- > 2.9M arrestee profiles
- > 823,000 forensic profiles
- other indexes also exist (unidentified human remains, missing persons, relatives of missing persons)
- aided in > 387,000 investigations



DNA based Forensic Testing



~200 individual have been exonerated using DNA testing

INNOCENCE PROJECT

The Innocence Project, founded in 1992 by Peter Neufeld and Barry Scheck at Cardozo School of Law, exonerates the wrongly convicted through DNA testing and reforms the criminal justice system to prevent future injustice.

The Innocence Project's mission is to free the staggering number of innocent people who remain incarcerated, and to bring reform to the system responsible for their unjust imprisonment.

DNA based Forensic Testing

weakness

- **minute or degraded samples**
- **contamination during sample collection/handling**
- **challenges with mixed samples**

  analysis tools based on “probabilistic genotyping”
  specific algorithms have come under increased scrutiny
- **unable to identify individuals whose profile is not already on file in the database**

  partially addressed with DNA “dragnets” & mass screenings
  a generally impractical approach

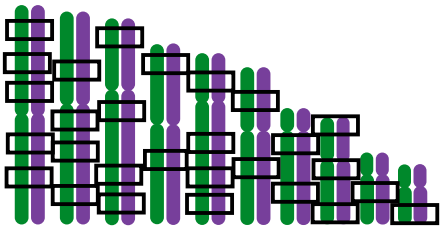
Forensic DNA Phenotyping

uses gene-based DNA information to infer **externally visible traits** such as:

decreasing scientific knowledge

- pigmentation (eyes, hair, skin)
- height
- hair loss/baldness
- age
- hair structure
- face shape and features

builds upon work of Drs. Greg Barsh & Devin Absher



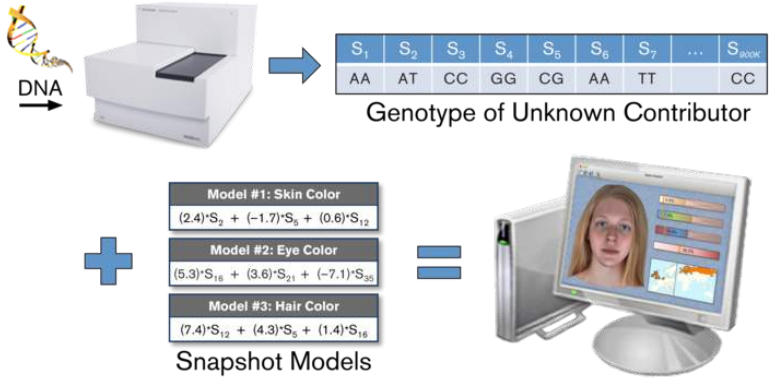
Forensic DNA Phenotyping

a “biological witness”

How Science is Putting a New Face on Crime Solving National Geographic July, 2016

Forensic DNA Phenotyping

Parabon NanoLabs



limitations of DNA Phenotyping

altered appearance

limitations of DNA Phenotyping

- **technical challenges**
- **ethical challenges**
  - ✦ privacy
  - ✦ disease risks
- **legal challenges**
- **limited scientific knowledge**

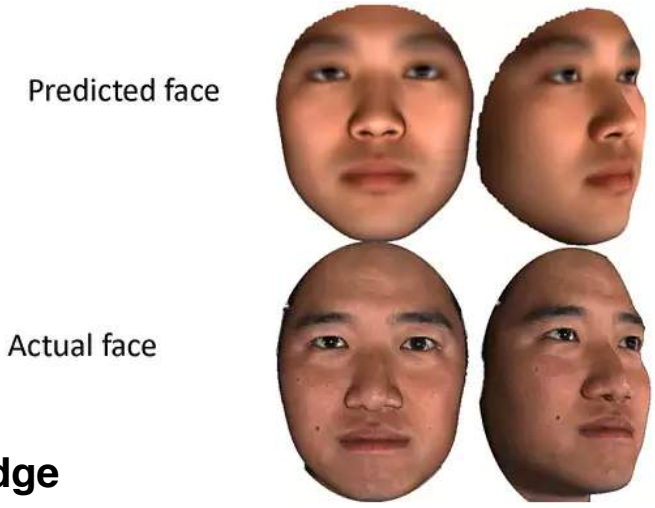


Photo Credit: Peter Claes

How can Biotech...

store information?

A few numbers for reference

1Kilobyte (KB=1,024 bytes)
2-3 paragraphs of text

1Megabyte (MB=1,024 KB)
800 pages of text
~25% of an audio MP3 song

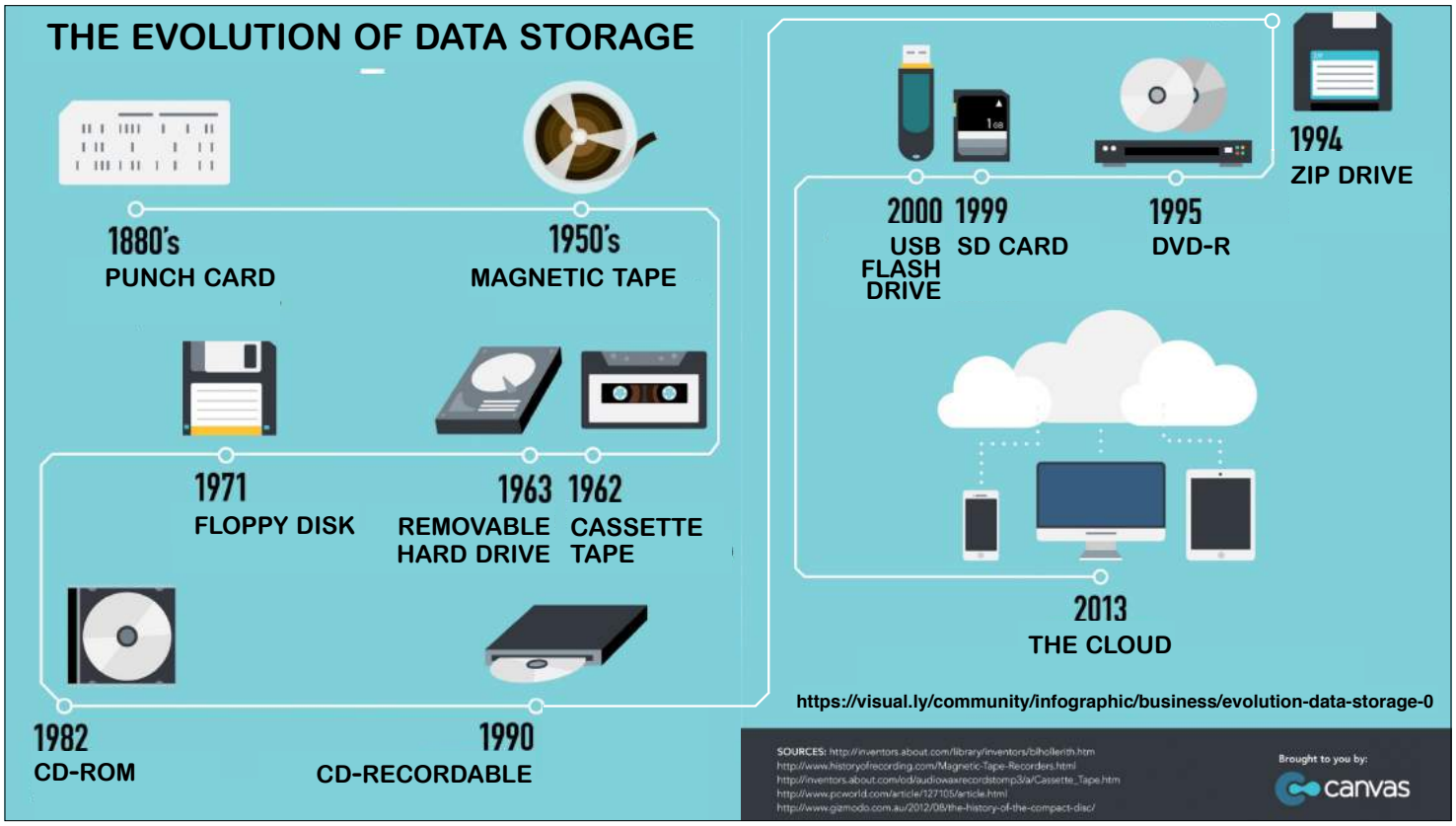
1 Gigabyte (GB=1,024 MB)
~900,000 pages of text
~350 digital photos (3MB avg)

1 Terabyte (TB=1,024 GB)
~916 million pages of text
~20 Blue-Ray (3 hr) discs

1 Petabyte (PB=1,024 TB)
~938 billion pages of text
~20 million 4-drawer filing cabinets

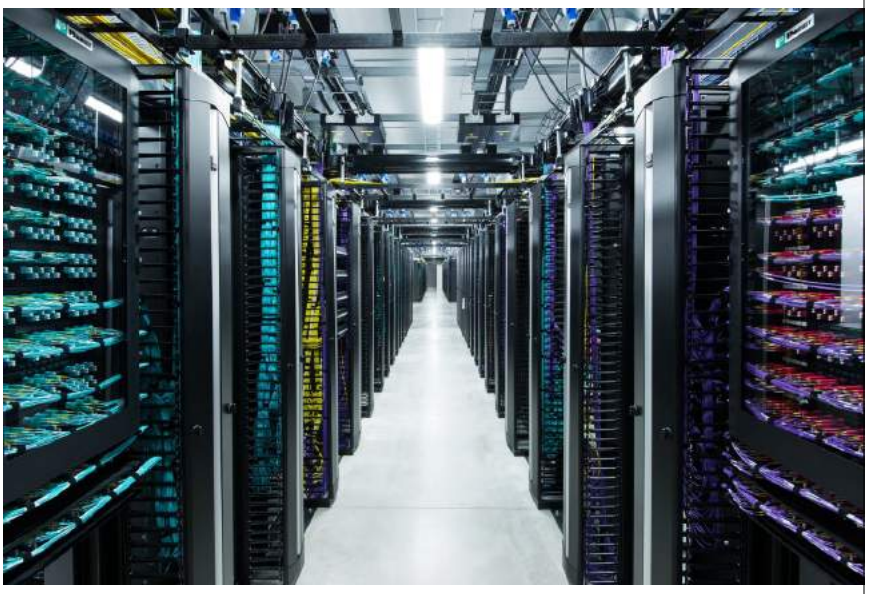
1 Exabyte (EB=1,024 PB)
~960 trillion pages of text

1 Zettabyte (ZB=1,024 EB)
~960 quadrillion pages of text



challenges ahead

- every 5 years - amount of data produced **increases 10X**
- more data created in last 2 years **than in all of preceding history**
- by 2040 ~ we'll have generated **three thousand zettabytes of data**
- capacity of existing storage media **isn't keeping pace with demand**

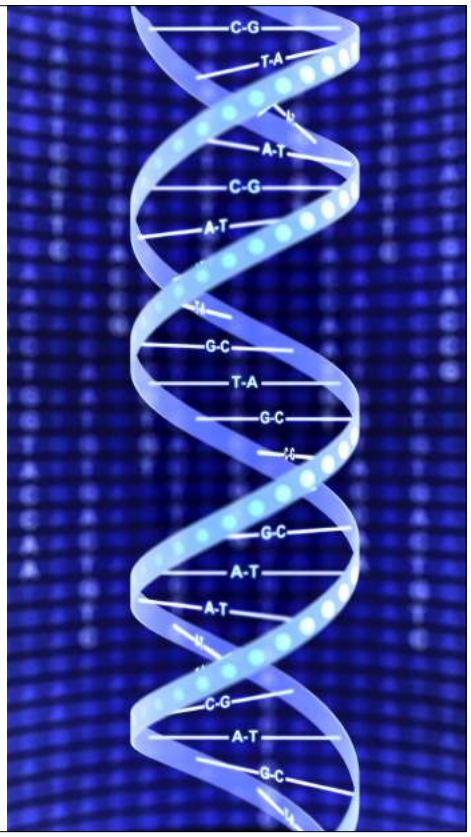


Facebook's data center in Lulea, Switzerland

DNA as a storage tool

Advantages

- tiny size
- high density
- low energy requirements for storage
- lasts thousands of years



DNA as a storage tool

A=00 T=01 C=10 G=11

David Copperfield

by Charles Dickens

Whether I shall turn out to be the hero of my own life, or whether that station will be held by anybody else, these pages must show. To begin my life with the beginning of my life, I record that I was born (as I have been informed and believe) on a Friday, at twelve o'clock at night. It was remarked that the clock began to strike, and I began to cry, simultaneously.

In consideration of the day and hour of my birth, it was declared by the nurse, and by some sage women in the neighborhood who had taken a lively interest in me several months before there was any possibility of our becoming personally acquainted, first, that I was destined to be unlucky in life; and secondly, that I was privileged to see ghosts and spirits; both these gifts inevitably attaching, as they believed, to all unlucky infants of either gender, born towards the small hours on a Friday night.

1. Letters are converted into binary code and then into DNA nucleotides

W h e t h e r

01010111 01101000 01100101 01110100 01101000 01100101 01110010

T T T G T C C A T C T T T G T A T C C A T C T T T G A C

2. Short fragments of DNA (plus locater codes) are artificially synthesized

DNA synthesizer

address locater

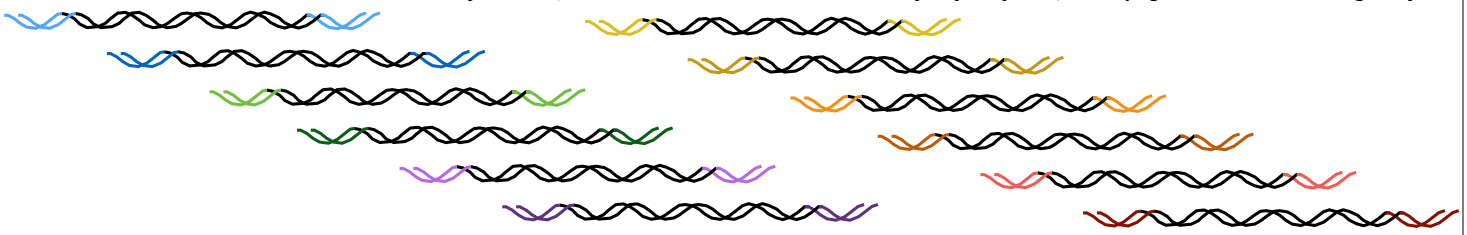
"W h e t h e r"

address locater

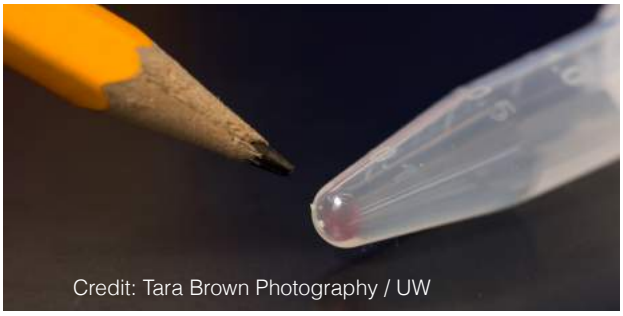
DNA as a storage tool

3. The information is coded into overlapping segments for redundancy

Whether I shall turn out to be the hero of my own life, or whether that station will be held by anybody else, these pages must show. To begin my

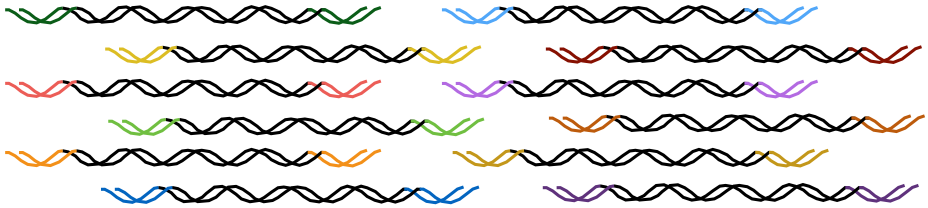


4. DNA fragments are dehydrated and frozen for long term storage

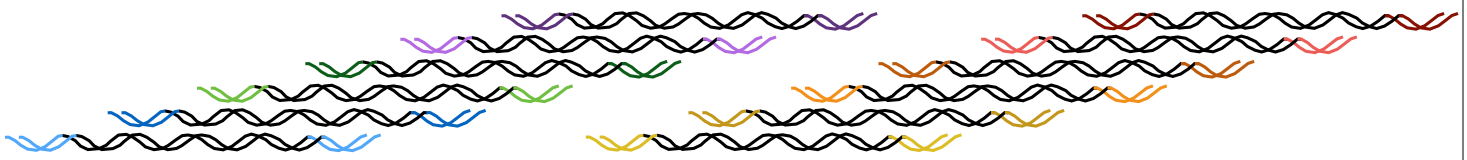


DNA as a storage tool

5. When needed, scientists rehydrate and sequence the DNA



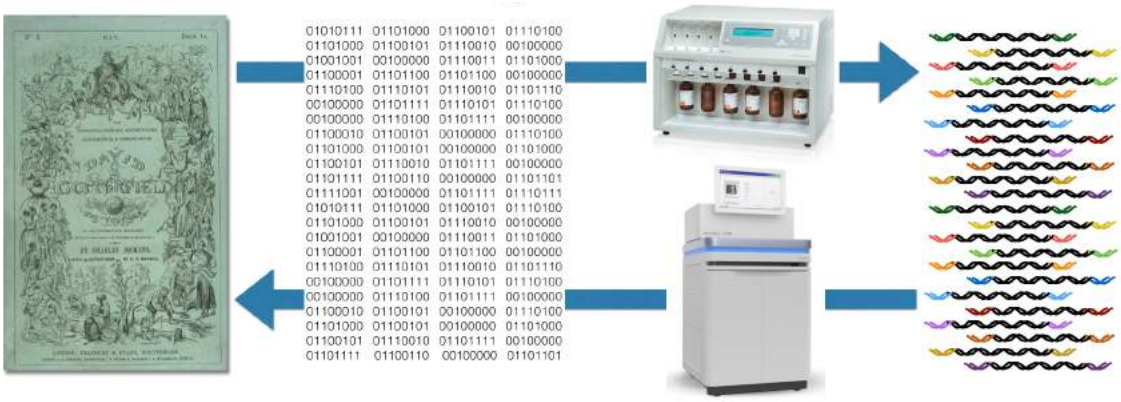
6. A computer reassembles the fragments (using the address locators), converts it to 0s and 1s, and then back to text



01010111 01101000 01100101 01110100 01101000 01100101 01110010 00100000 01001001 00100000 01110011 01101000 01100001 01101100 01101100 00100000 01110100 01110101 01110010 01101110 00100000 01101111 01110101 01110100 00100000 01110100 01101111 00100000 01100010 01100101 00100000 01110100 01101000 01100101 00100000 01101000 01100101 01110010 01101111 00100000 01101111 01100110 00100000 01101101 01111001 00100000 01101111 01110111 01101110 00100000 01101100

Whether I shall turn out to be the hero of my own life, or whether that station will be held by anybody else, these pages must show. To begin my

DNA as a storage tool



2012 - DNA 1st used to store digital data

2016 - 2 megabytes of data, stored across 72,000 DNA strands

2017 - technique developed capable of storing 215 petabytes per gram (~100 million movies)

DNA as a storage tool

a sugar cube's worth of DNA could store **1000X more data** than the largest archive facility today

it's estimated that **all the world's data** could be stored in an amount of DNA that would occupy the physical space of 2 pickup truck beds

challenges with DNA storage

- **cost of writing (synthesizing) DNA**
- **cost of reading (sequencing) DNA**
  - 1970s $600.00/nucleotide
  - today $ .10/nucleotide
  - "ideally" $ .000001/nucleotide
- **speed of storage**
- **errors in data**
  - data missing/corrupted
  - biochemically introduced errors
- **randomly accessing the data**

lessons from other data storage tools

modify tools that store, compress, & retrieve data in other mediums

error-correction

- Reed-Solomon Code - CDs & DVDs
- Fountain Code - video streaming

collaboration with molecular biologists, computer scientists & architects

MISL Microsoft TWIST BIOSCIENCE

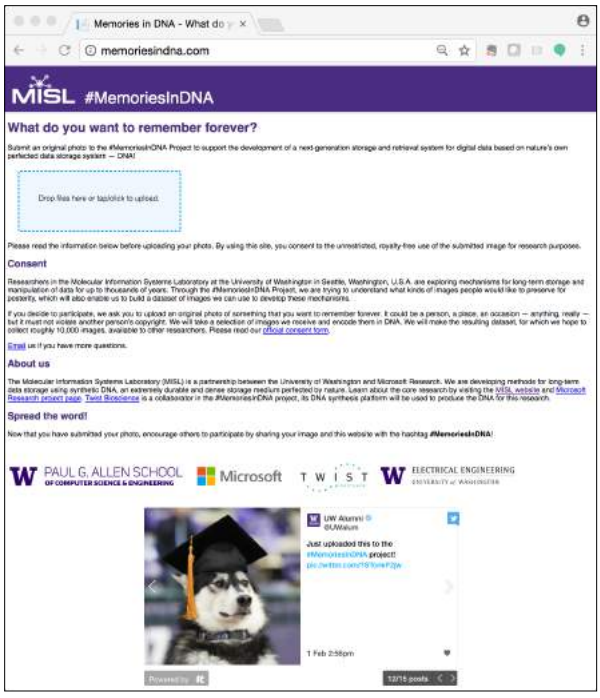
Future applications of DNA storage

storage backups for quantum computers

cells as living data recorders

image capture/recognition - search engine

Future applications of DNA storage



http://memoriesindna.com/

The project seeks to collect 10,000 original images from around the world and preserve them indefinitely in synthetic DNA. The public is invited to submit original photographs that they'd like to see preserved in DNA.

Want to know more?

The Basics

Making DNA Data Storage a Reality, Catherine Offord, The Scientist, Oct. 2017, (https://www.the-scientist.com/?articles.view/articleNo/50406/title/Making-DNA-Data-Storage-a-Reality/)

How DNA Could Store All the World's Data, Andy Extance, Nature, Sept 2016 (https://www.nature.com/news/how-dna-could-store-all-the-world-s-data-1.20496)

Storing data in DNA brings nature into the digital universe, Luis Ceze and Karin Strauss, The Conversation, July 27, 2017, (https://theconversation.com/storing-data-in-dna-brings-nature-into-the-digital-universe-78226)

Future Studies

DNA Data Storage - Setting the Density Record With DNA Fountain, Twist Bioscience Blog, Dec 1, 2017, (https://twistbioscience.com/company/blog/twistbiosiencednastoragefountain)

Articles from Wired Magazine

What if quantum computers used hard drives made of DNA? Sophia Chen, March 15, 2017 (https://www.wired.com/2017/03/quantum-computers-used-hard-drives-made-dna/)

Scientists upload a galloping horse GIF into bacteria with CRISPR, Megan Molteni, July 12, 2017 (https://www.wired.com/story/scientists-upload-a-galloping-horse-gif-into-bacteria-with-crispr/)

DARPA wants to build an image search engine out of DNA, Megan Molteni, Jan 24, 2018 (https://www.wired.com/story/darpa-wants-to-build-an-image-search-engine-out-of-dna/)

Next Week

6:30 – 8:00 pm CT

Feb 13	solve a crime? store information in DNA?
Feb 20	change my DNA? create "designer babies"?
Feb 27	help me lose weight? predict my death?
March 6	help us reach and colonize Mars?