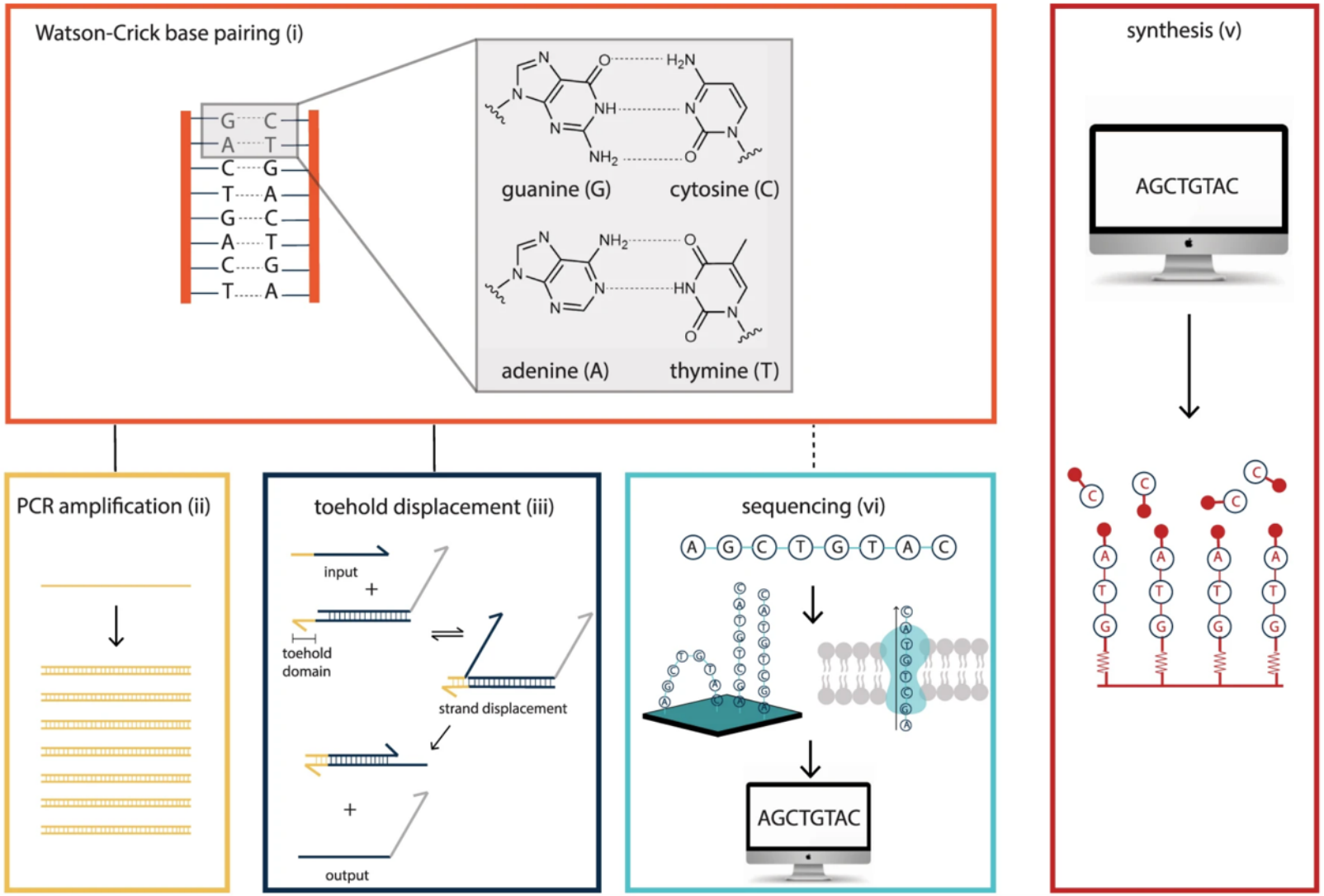
ETH zürich

Why is DNA so unique as a digital tool?

Why we love DNA

Prof. Dr. Robert Grass
ETH Zürich

DNA is a really special molecule



Meiser et al. *Nat. Commun.* **2022**.

Chemistry and information

Chemistry IN →

→ **Information OUT**

Analytical chemistry

Only with DNA: single molecule detection

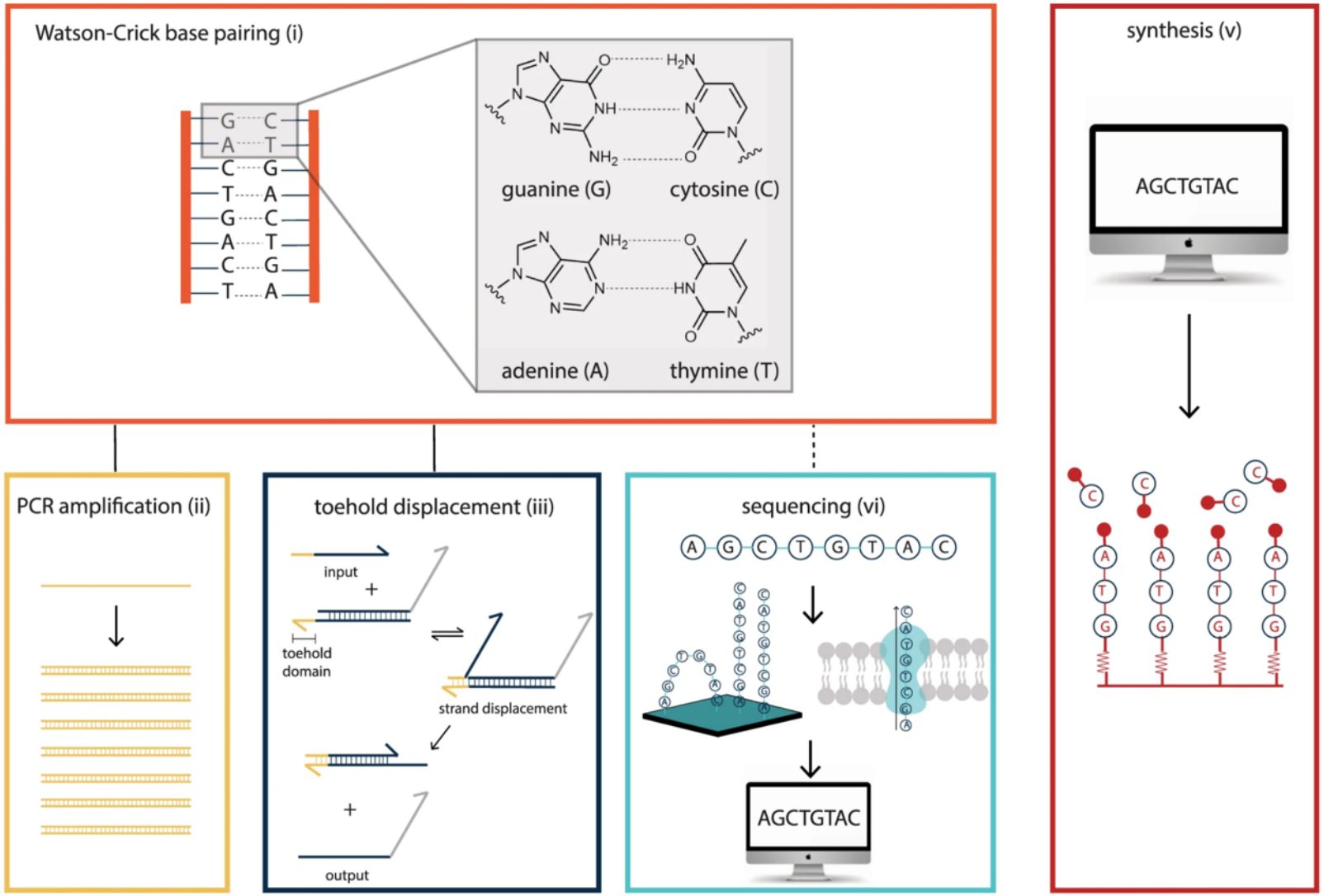
Chemistry IN → **Analytical (Bio-)Chemistry** → **Information OUT**

DNA (100 bp)

Detection limit: 1 molecule (PCR)
(0.1 attogram)

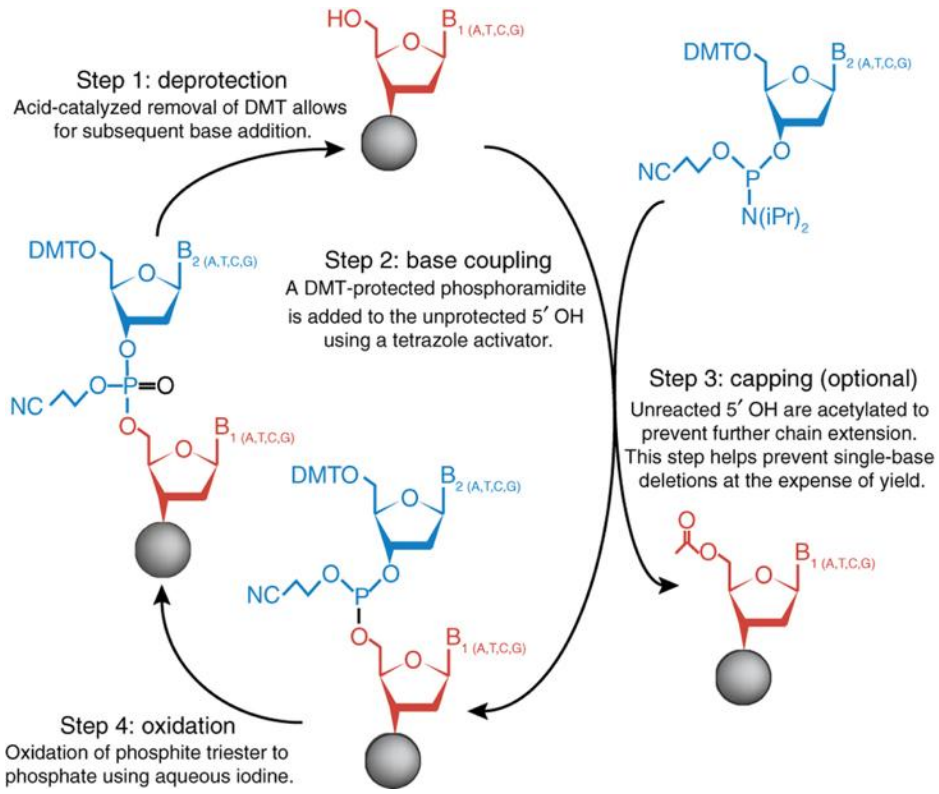
+ sequence (NGS sequencing)
200 bit Information per 0.1 attogram

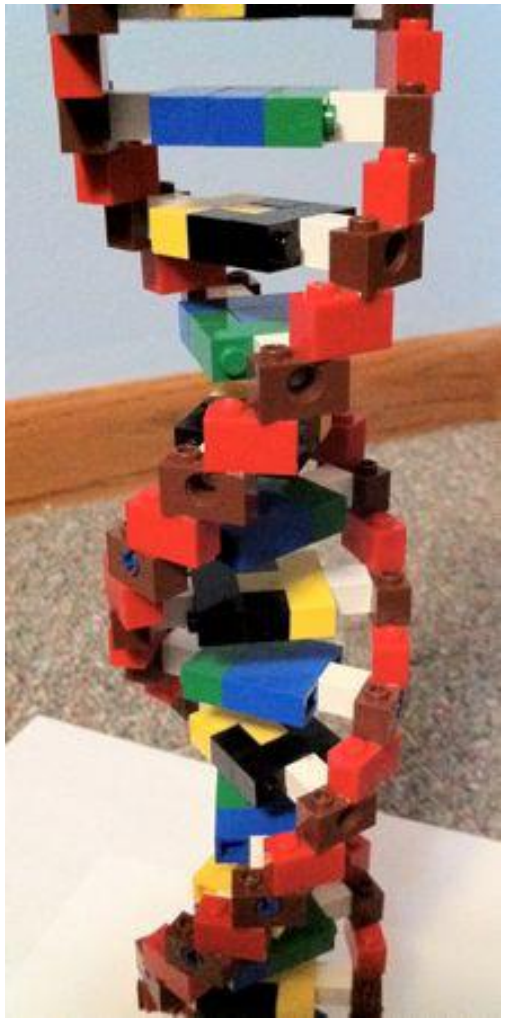
DNA is a really special molecule



Meiser et al. *Nat. Commun.* **2022**.

DNA Synthesis = Writing



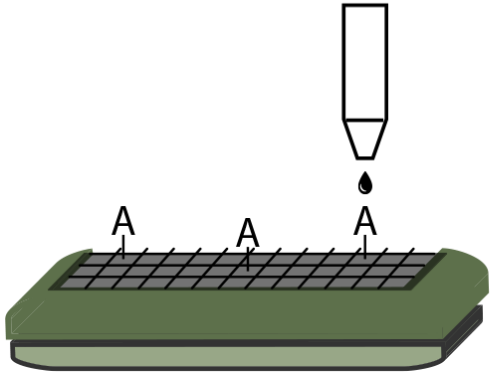


today: ~ 0.0005 USD / Base

Kosuri & Chruch, *Nat. Methods* **2014.**

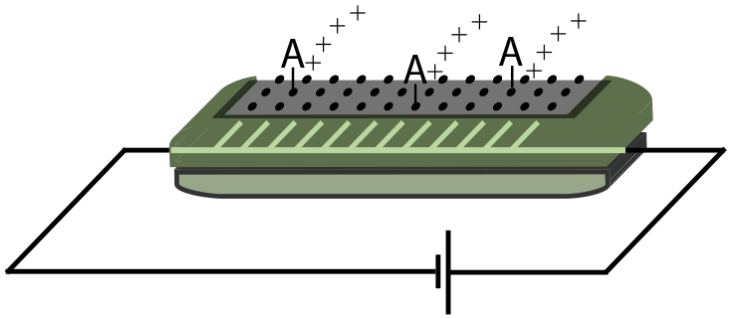
Only with DNA: Massive parallel synthesis

Chemical Synthesis via Ink-jet printing



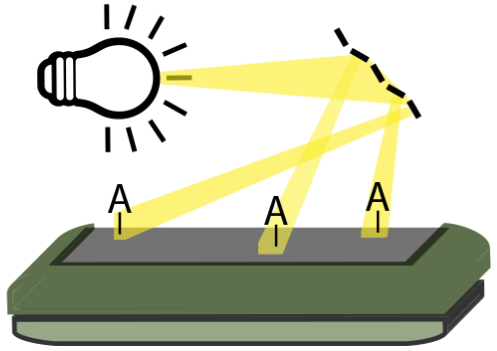
- Deposition of small volumes
- Complicated liquid handling
- High yield and low error rate

Electrochemical Synthesis



- Microfluidic device
- Electrochemical deprotection
- Medium quality DNA

Light-directed synthesis

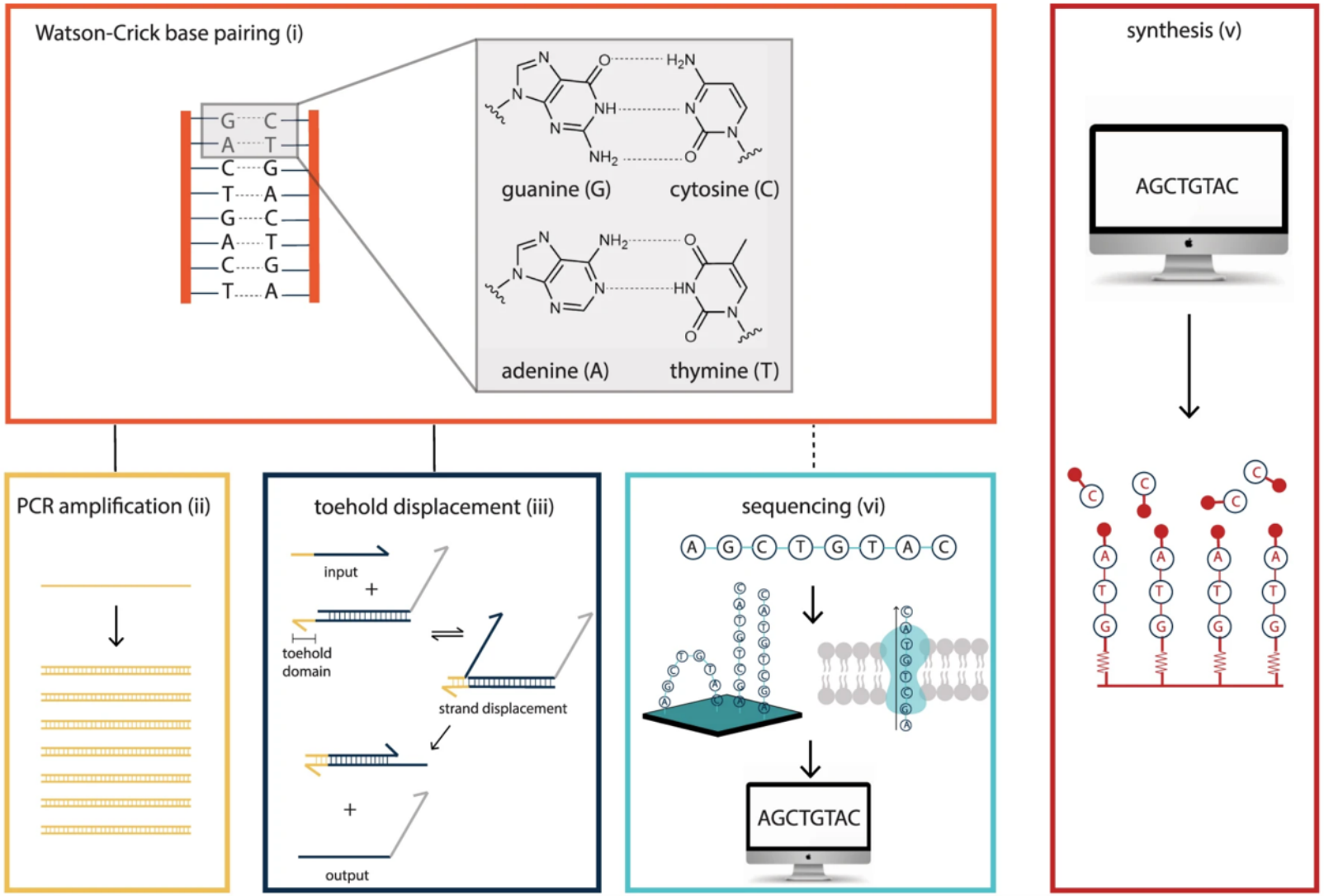


- Flow-cell
- Photochemical deprotection
- Low quality DNA

Synthesis with up to 8 million parallel reactions per step (Genscript)
Up to 300 nucleotides long (Twist Bioscience)
--> Theoretical 340 MBs per pool

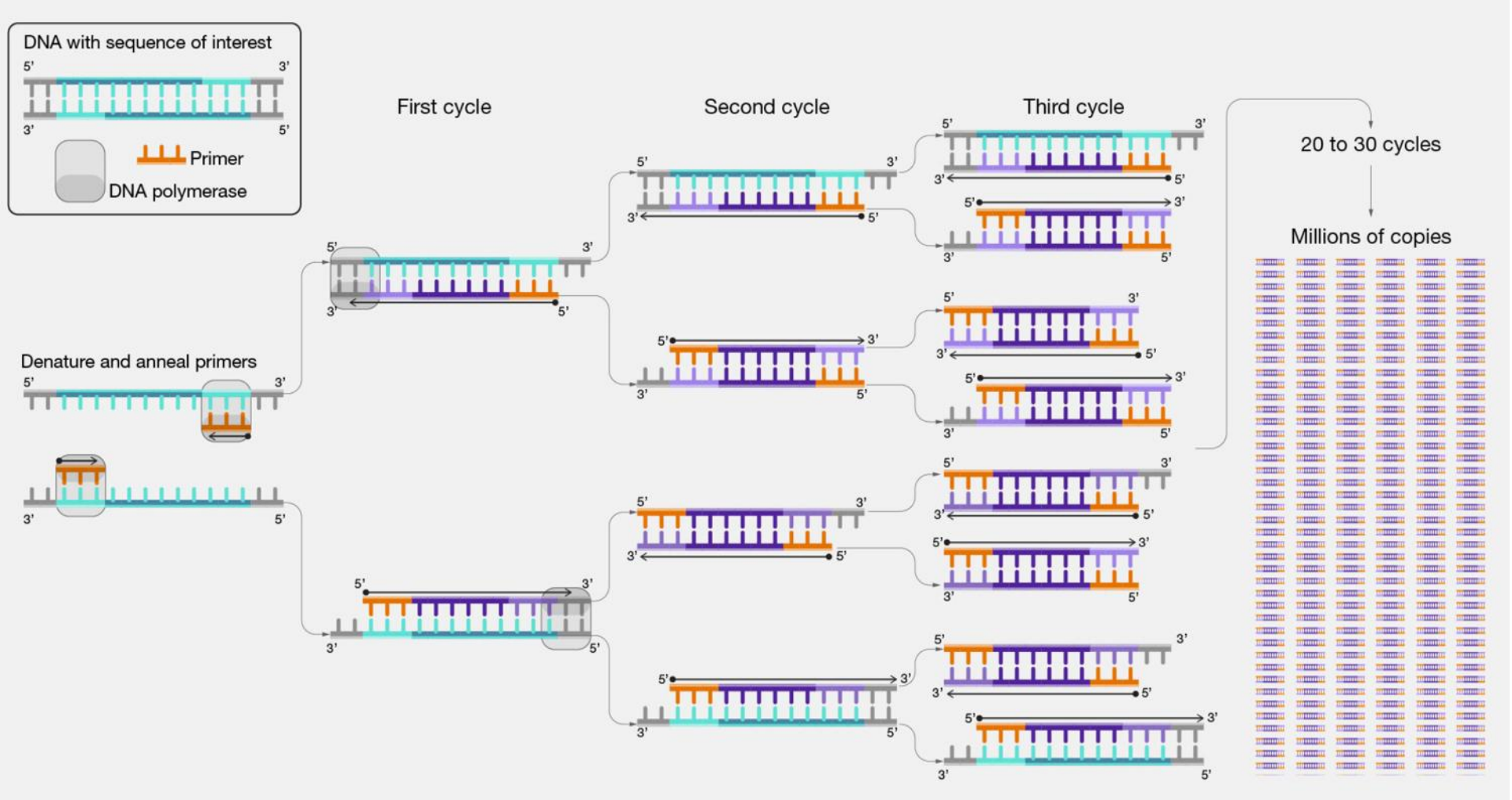
Antkowiak et al. *Nat. Commun.* **2020**.

DNA is a really special molecule



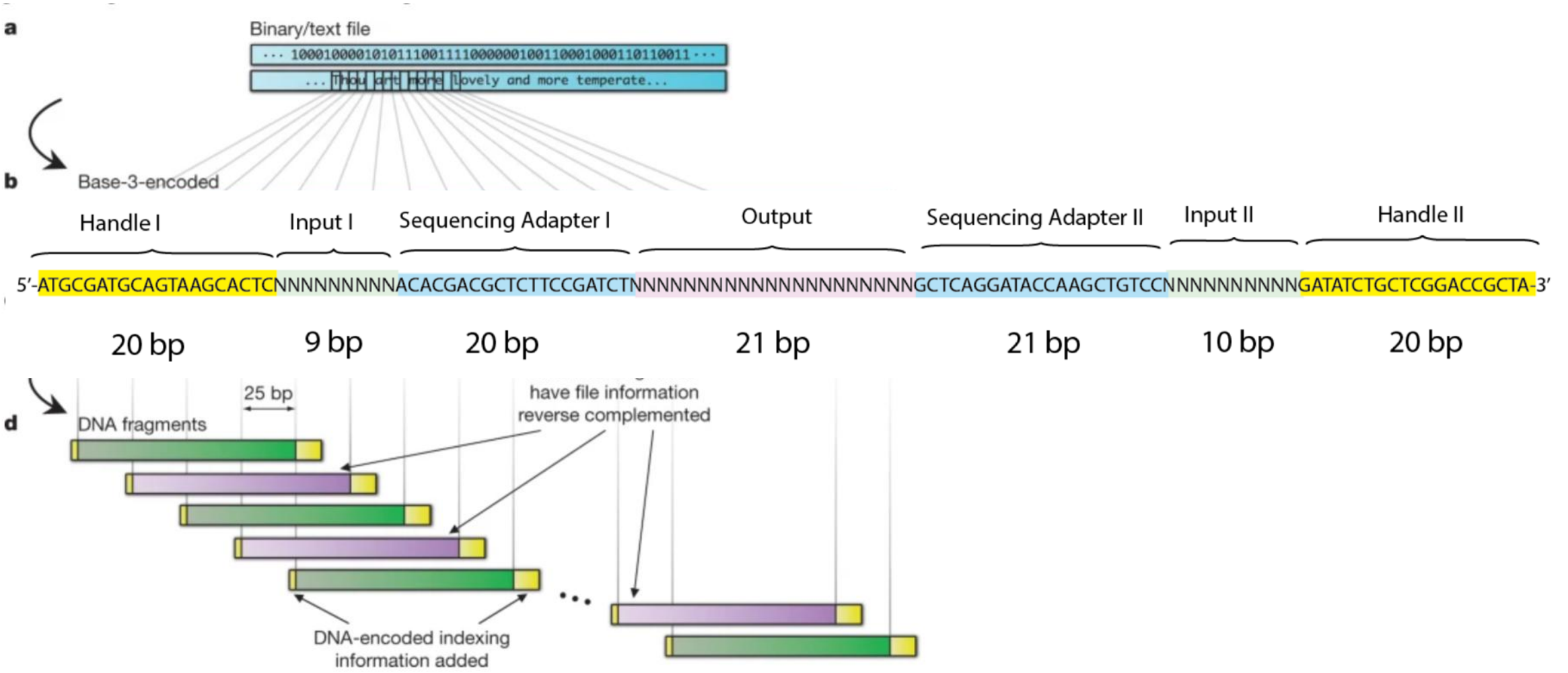
Meiser et al. *Nat. Commun.* **2022**.

Only with DNA: Making copies



Courtesy: National Human Genome Research Institute

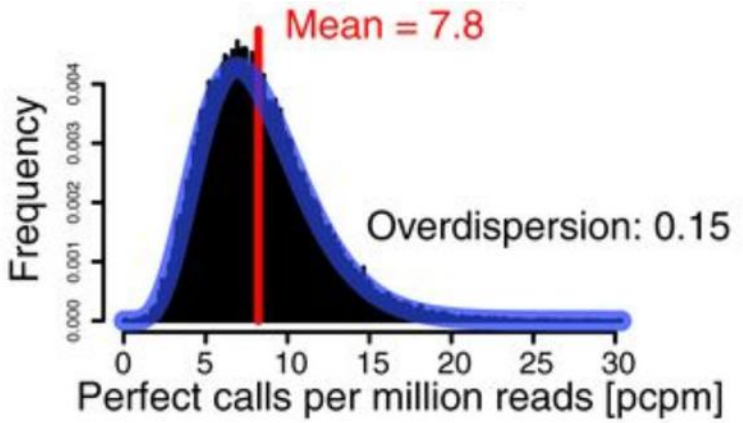
Making copies of DNA pools

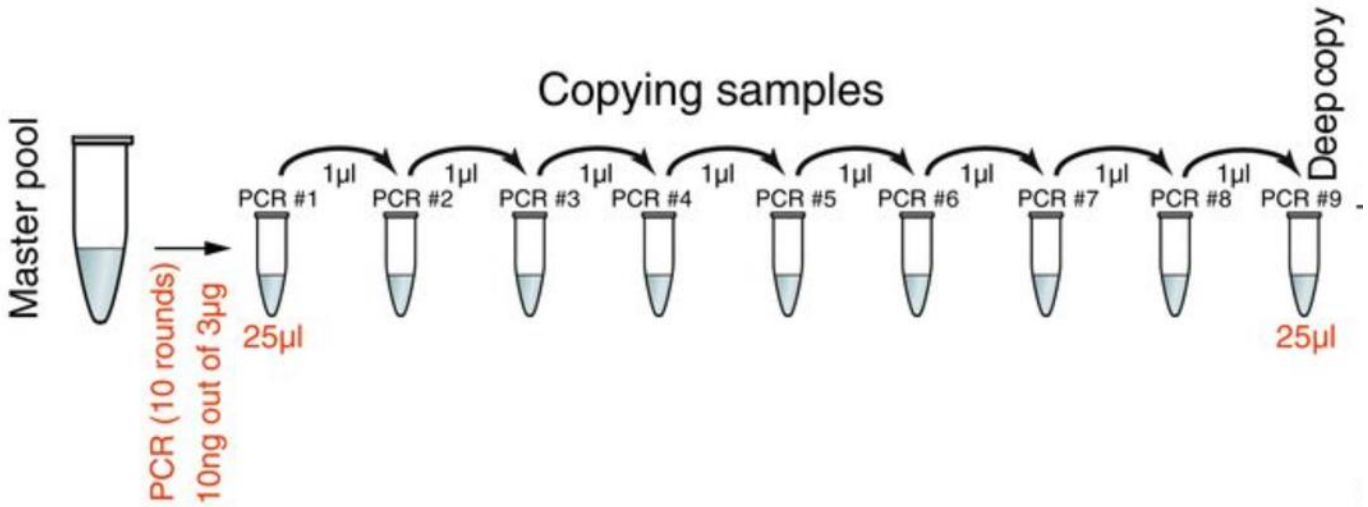


Goldman et al. *Nature*, **2013.**

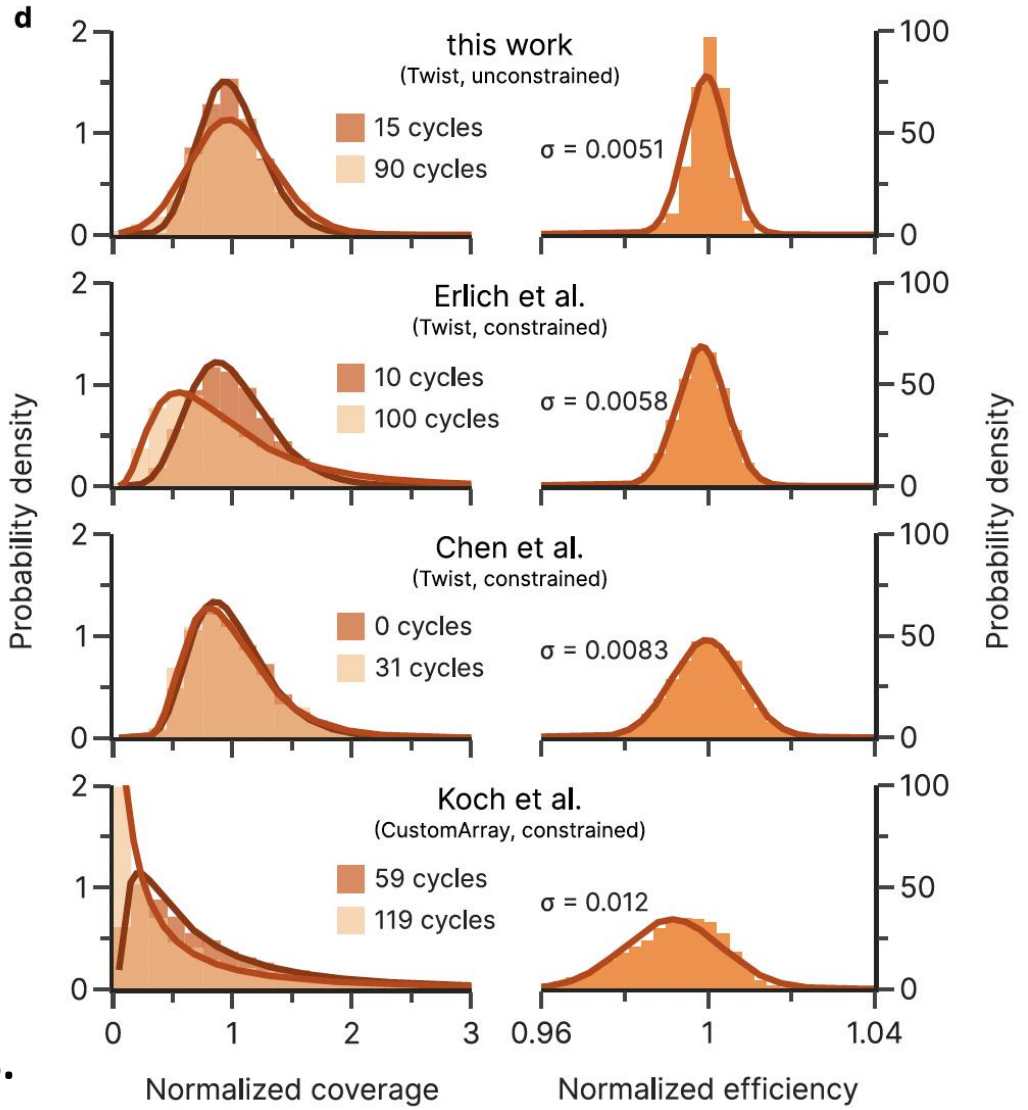
Luescher et al. *Nat. Commun.* **2023**.

Do all sequences amplify to the same extent?





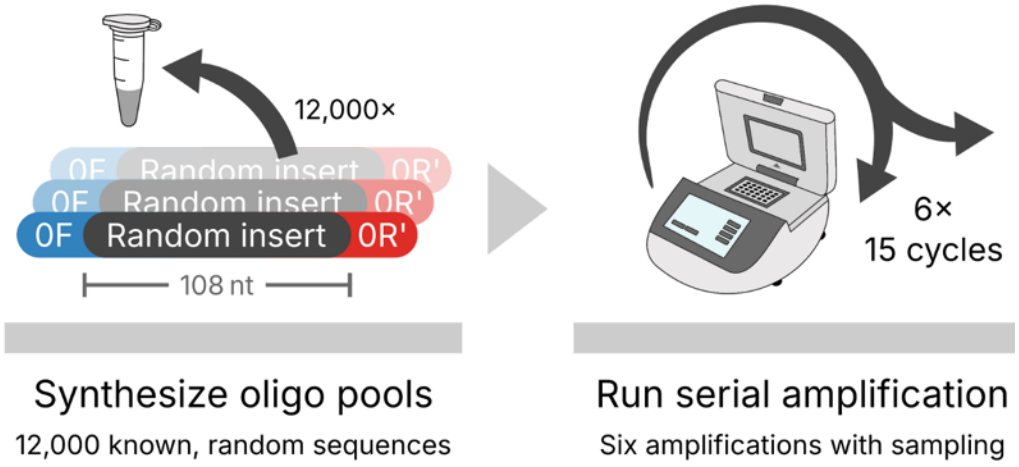
Erlich et al. *Science,* **2017.**

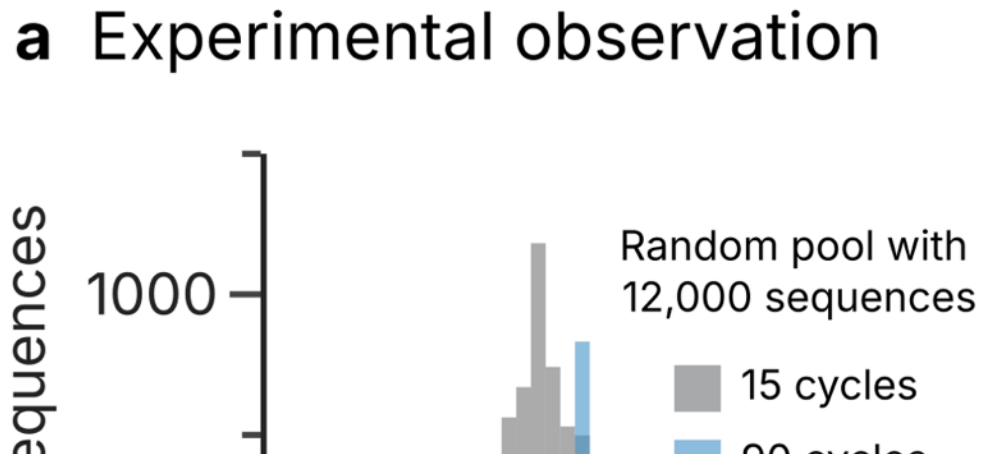


Gimpel et al. *Nat. Commun.,* **2023.**

Understanding bias in multi-template PCR

- Pools of DNA sequences with cons
- Measure sequence dependent am sequencing)

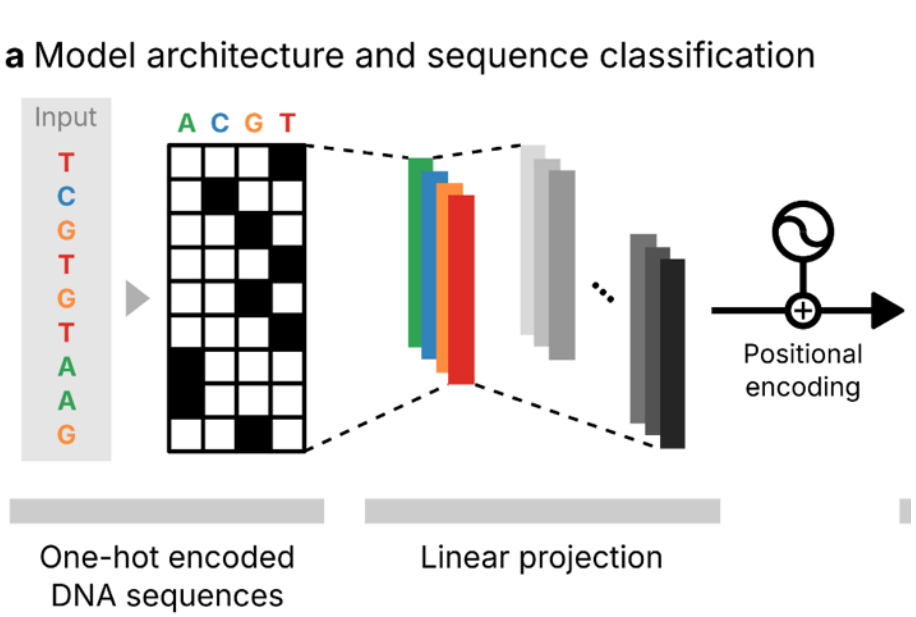


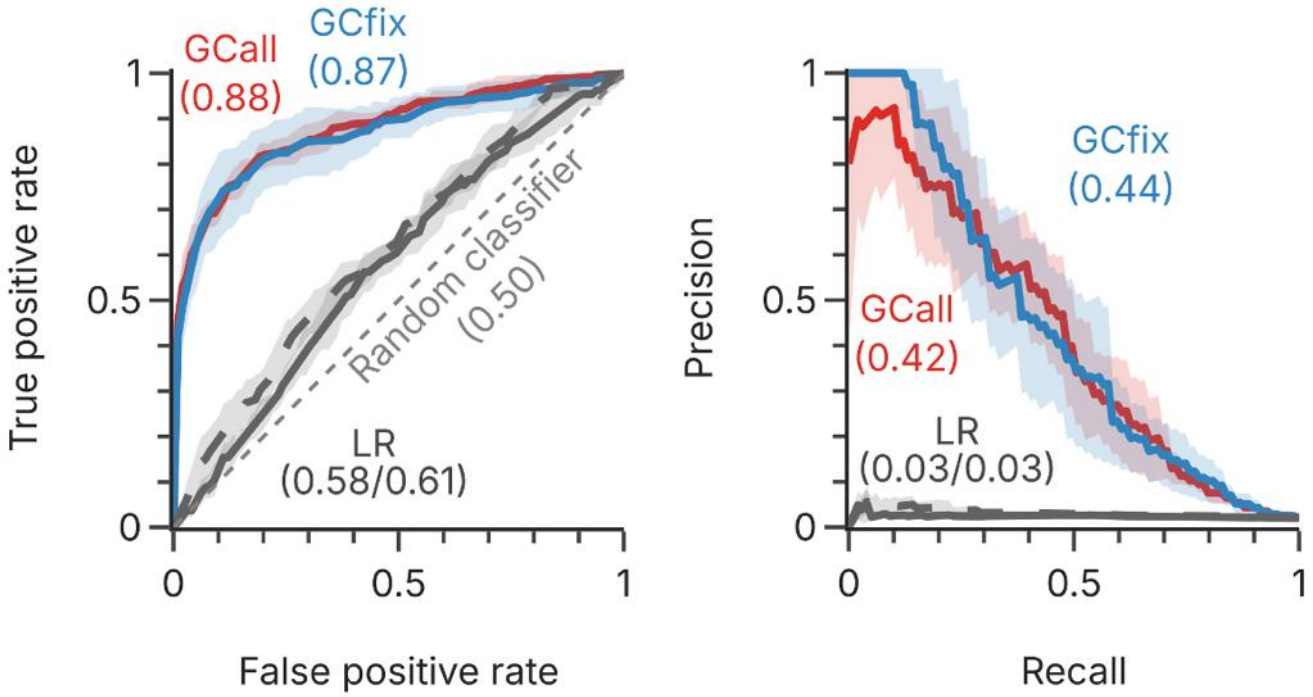


Gimpel et al. *bioRxiv,* 2024.09. 20.614030 **2024.,** collaboration with K. Borgwardt

Machine learning to connect sequence to amplification efficiency

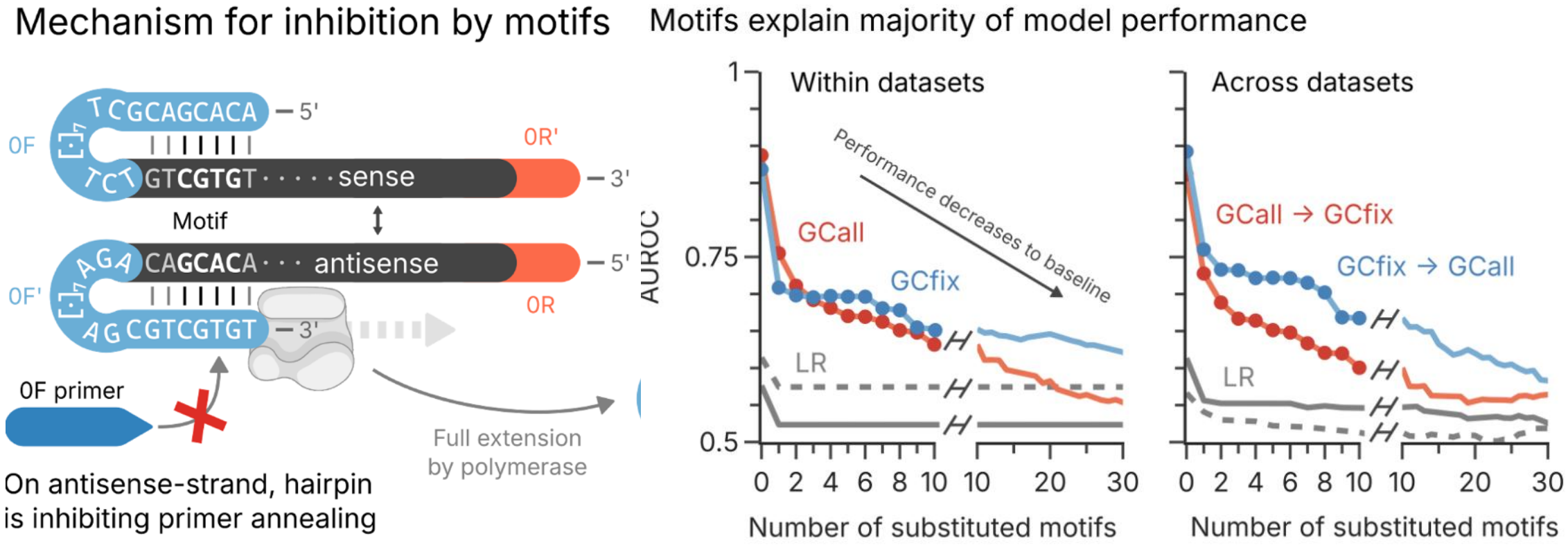
1D-CNN





Gimpel et al. *bioRxiv,* 2024.09. 20.614030 **2024.,** collaboration with K. Borgwardt

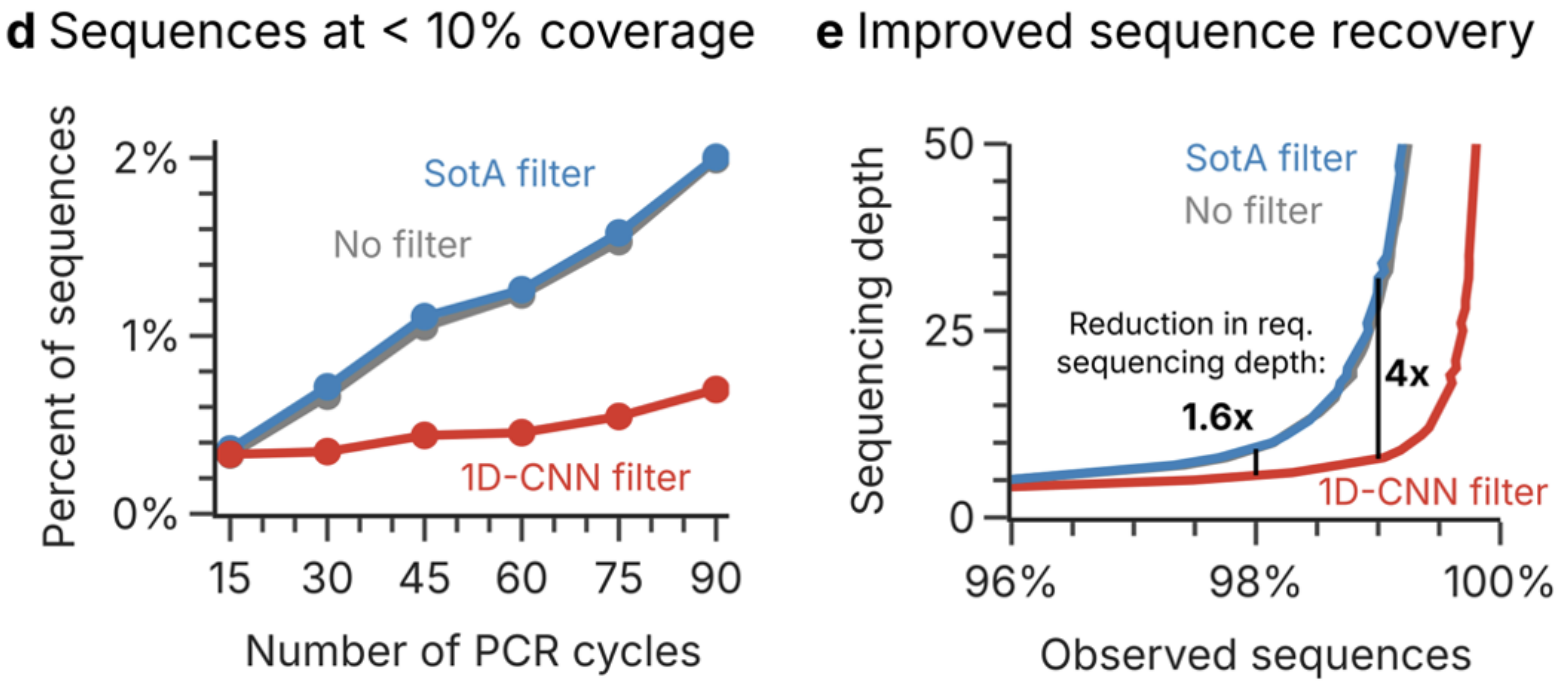
Motiv discovery



Intramolecular loop formation depends on distance – short distance motivs are more stable
→ Confirmed by thermodynamic calculations

Gimpel et al. *bioRxiv,* 2024.09. 20.614030 **2024.,** collaboration with K. Borgwardt

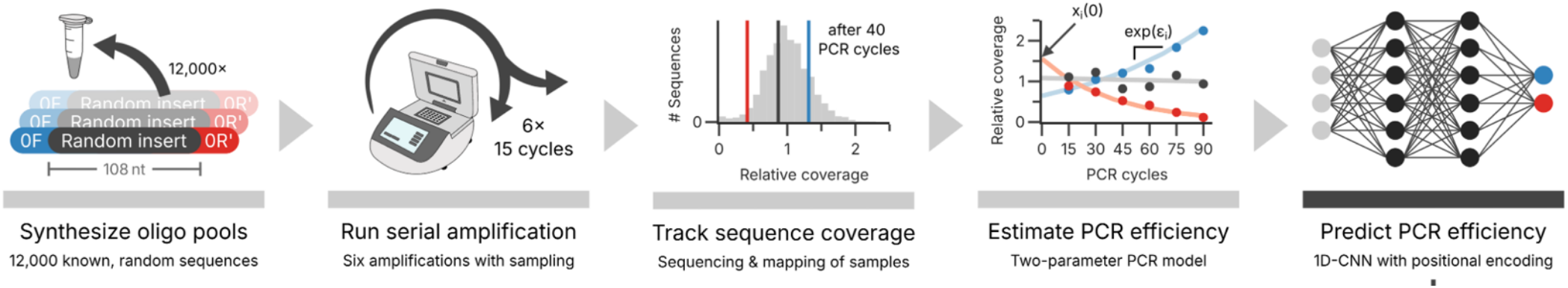
Avoiding sequences improves sequence recovery



It is not GC balance and polynucleotide repeats – but other features that impede amplification

Gimpel et al. *bioRxiv,* 2024.09. 20.614030 **2024.,** collaboration with K. Borgwardt

Proven workflow for identification of PCR bias



Synthesize oligo pools
12,000 known, random sequences

Run serial amplification
Six amplifications with sampling

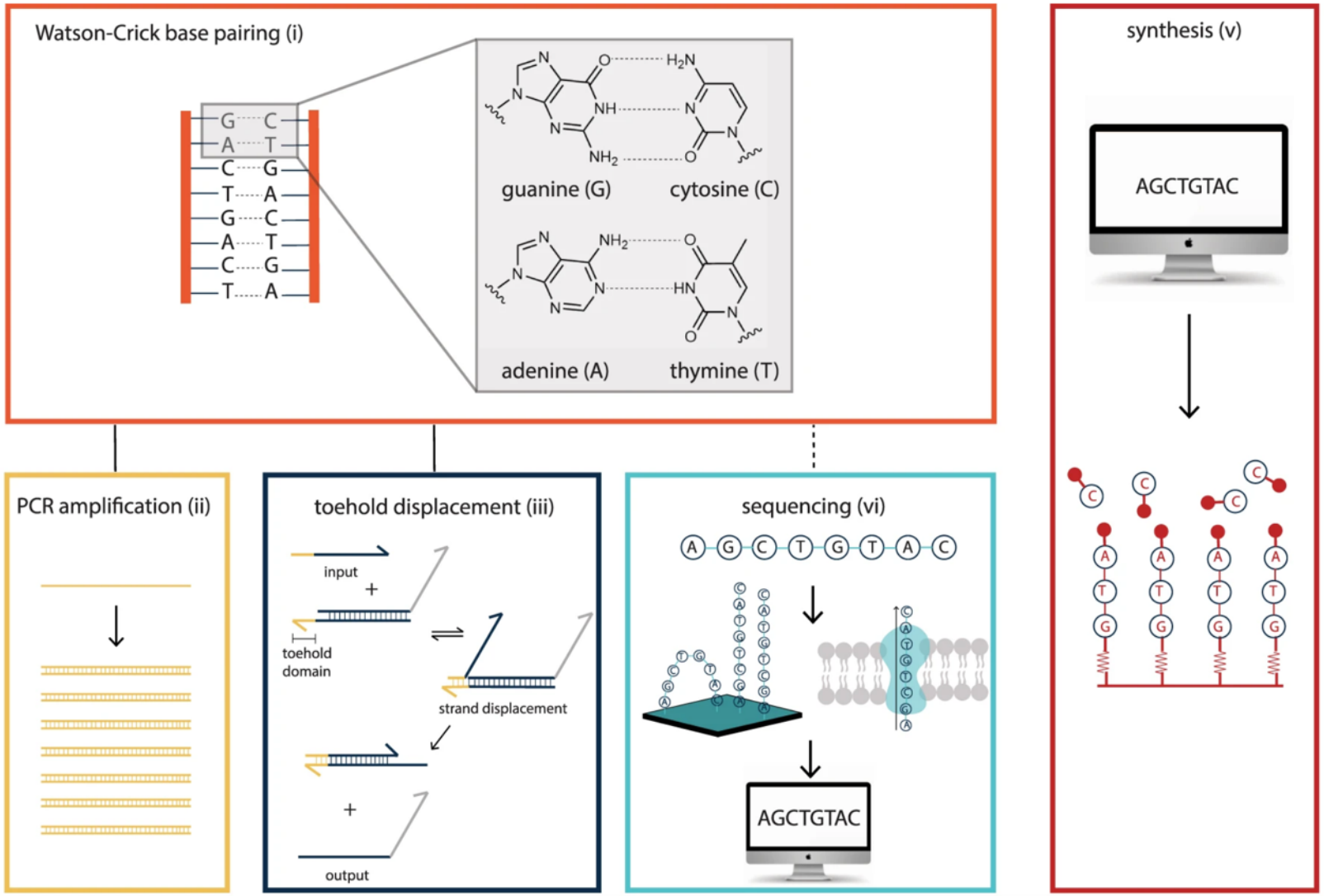
Track sequence coverage
Sequencing & mapping of samples

Estimate PCR efficiency
Two-parameter PCR model

Predict PCR efficiency
1D-CNN with positional encoding

Gimpel et al. *bioRxiv,* 2024.09. 20.614030 **2024.,** collaboration with K. Borgwardt

DNA is a really special molecule



Meiser et al. *Nat. Commun.* **2022**.

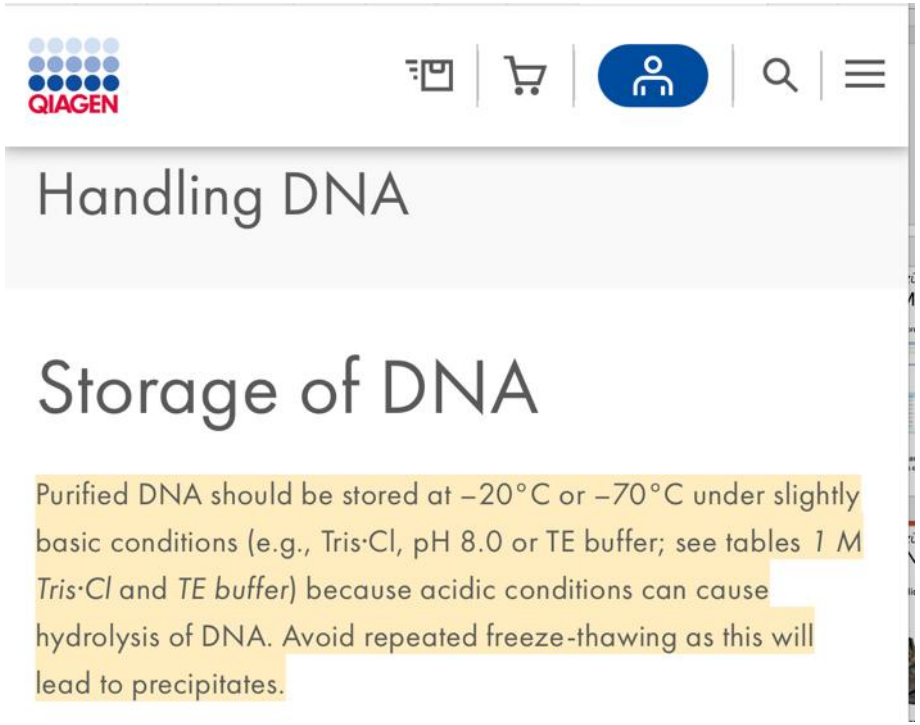
DNA weakest link: stability

> 1 Mio year old mammoth tooth

3’600 Mio base pairs after 1’000’000 years

van der Valk. *Nature* **2021**.

DNA storage recommondations



DNA only stable for extended storage at -20°C

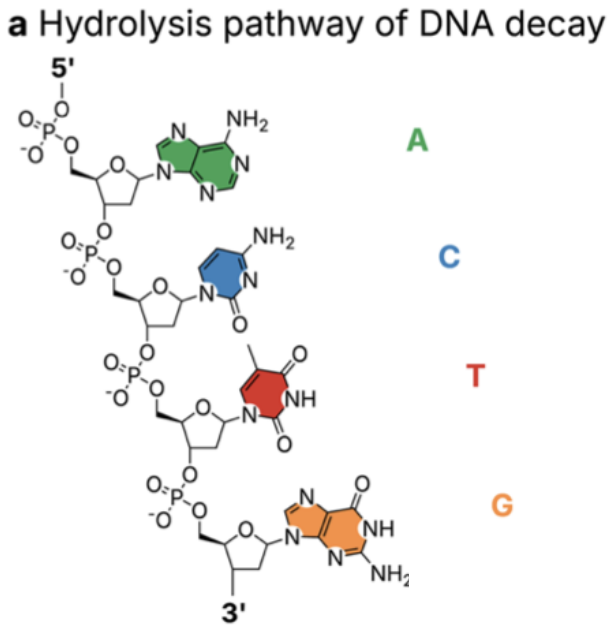
DNA stability

In our body: 60’000 errors per day – DNA repair

In environment: durability << 1 year

Nobel prize 2015

DNA decay



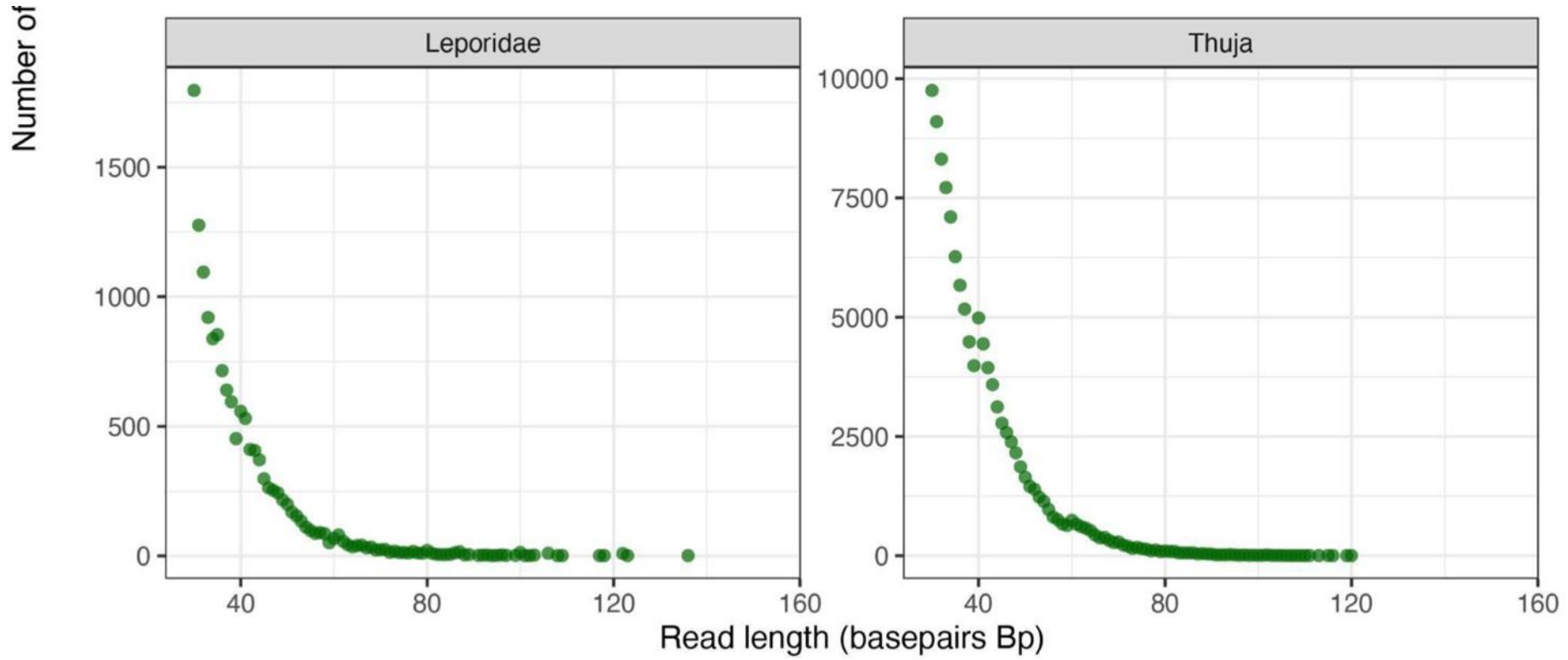
The decay rate is per nucleotide in the strand

→ Rate of having at least one break in a strand scales (linearly) with length of DNA

Lindahl, *Nature*, **1993**.
Gimpel et al. *Digital Discovey,* **2024**
Meiser et al. *Comms. Biol.* **2022.**

DNA weakest link: stability

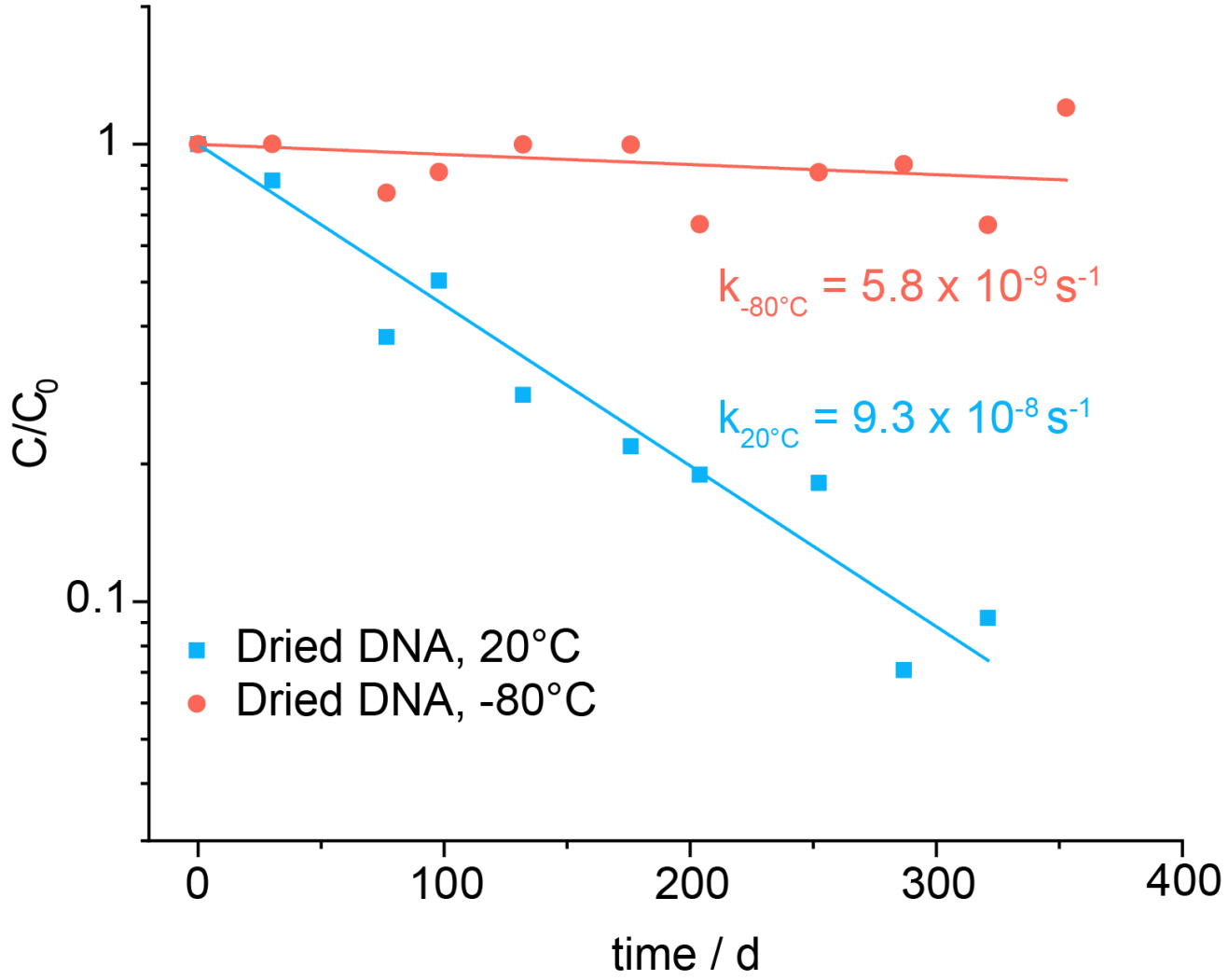
Ancient DNA (2 Mio years at -17°C)



Kjær et al. *Nature* **2022**.

How bad is it at room temperature?

Dry DNA
150nt length

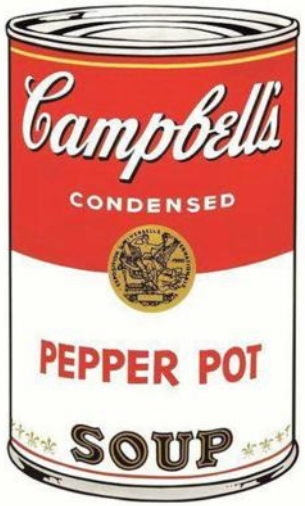


Antkowiak et al. *Small.* **2022.**

Preventing Decay

Cold

Tin-can

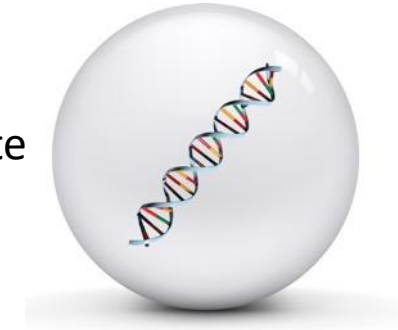


Imagene DNAShell©

Fossil

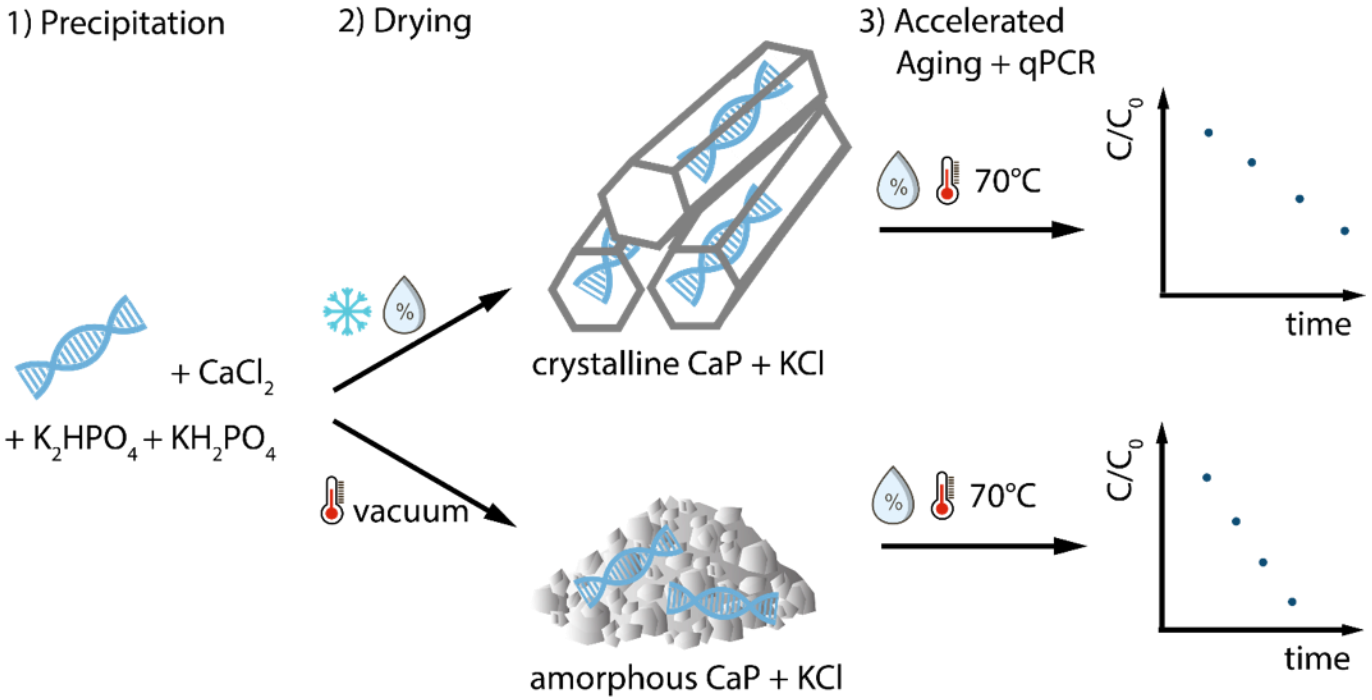
Glass capsules

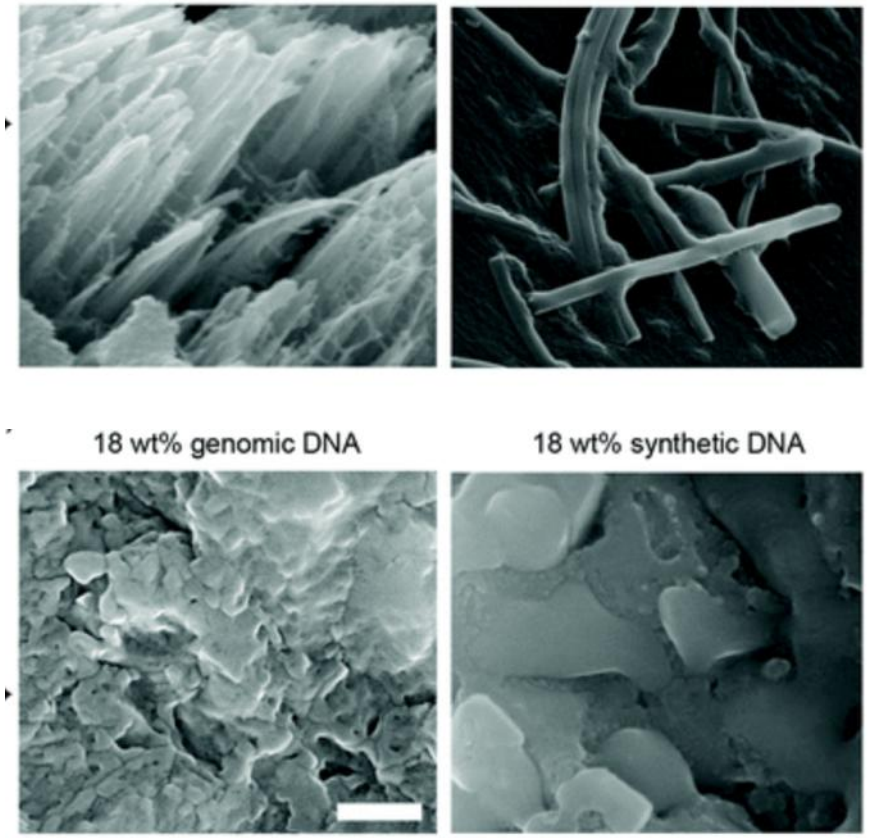
Protecte



here

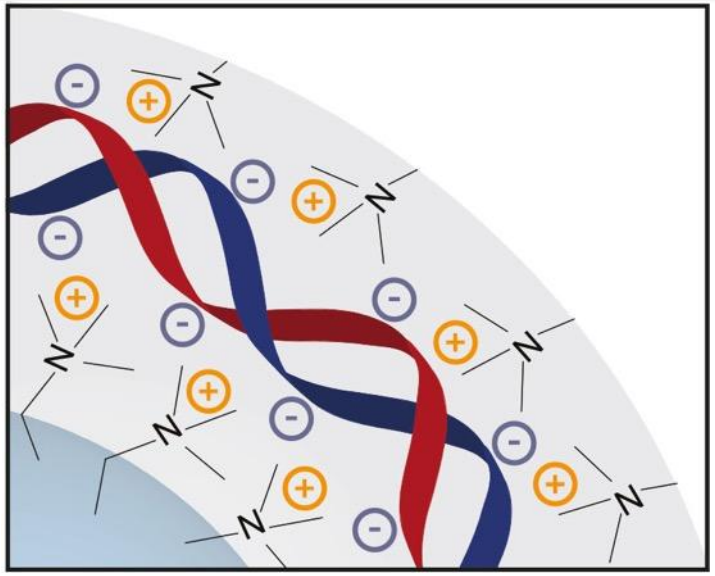
Storage in calcium phosphate

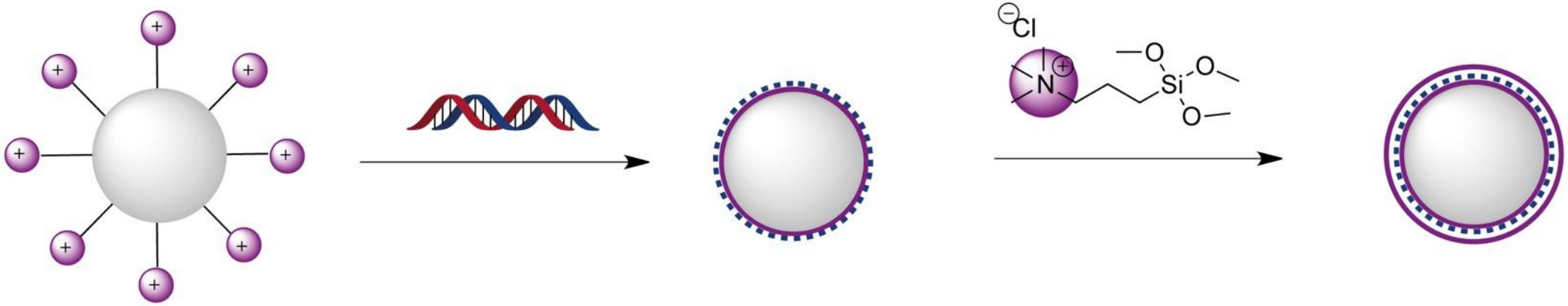




Antkowiak et al. *Chem. Commun.* **2022**

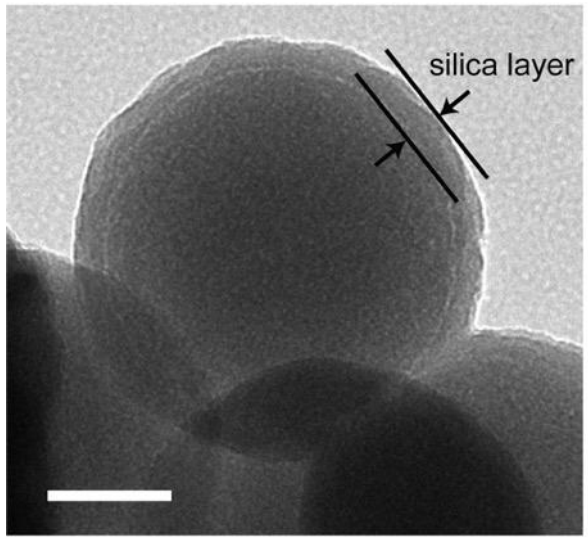
Synthetic fossil

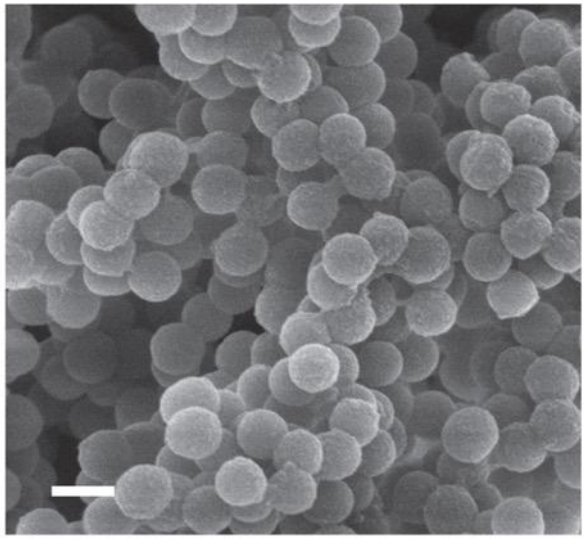


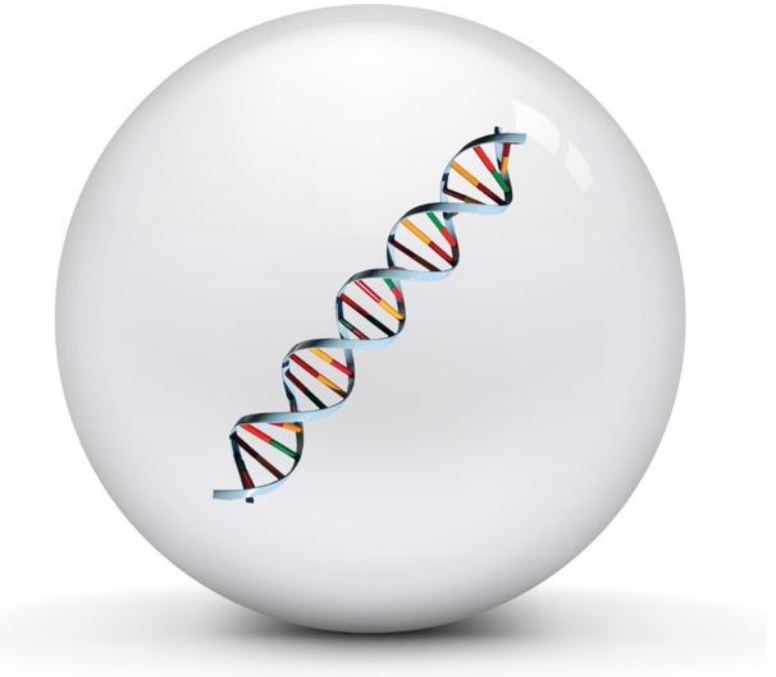


Paunescu et al. *Angew. Chem. Int. Ed.* **2013,** *52*, 4269.
Paunescu et al. *Nat. Protoc.* **2013**, *8*, 2440.

Synthetic fossil (made of glass)

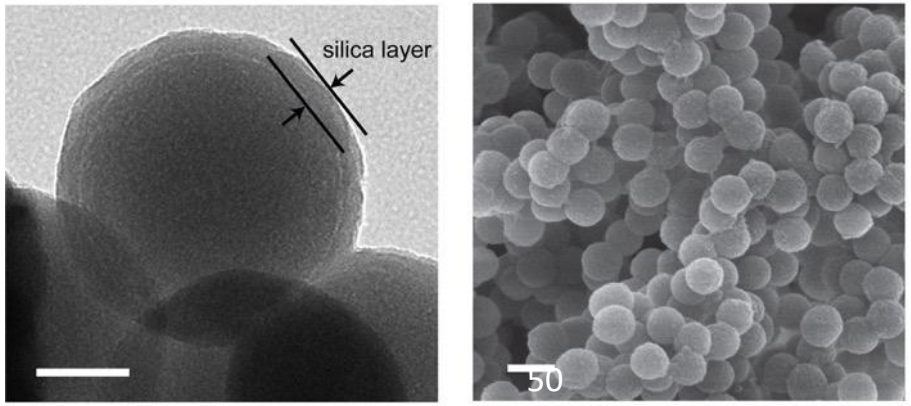


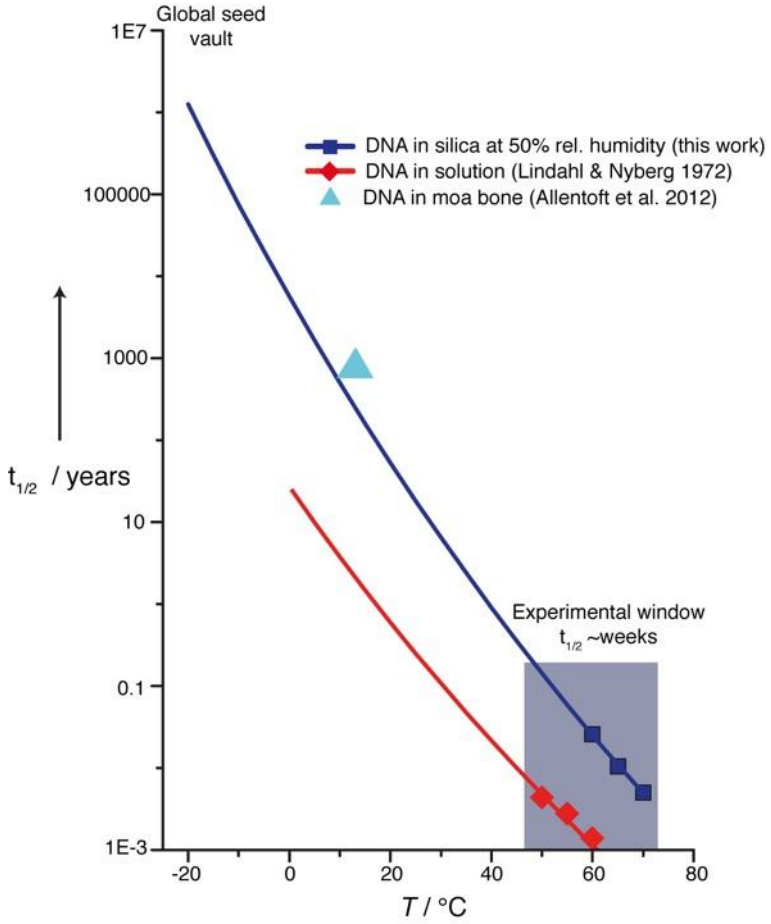




Paunescu, et al. *Angew. Chem. Int. Ed.* **2013,** *52*, 4269.
Paunescu et al. *Nat. Protoc.* **2013**, *8*, 2440.

If stored correctly, DNA is highly stable



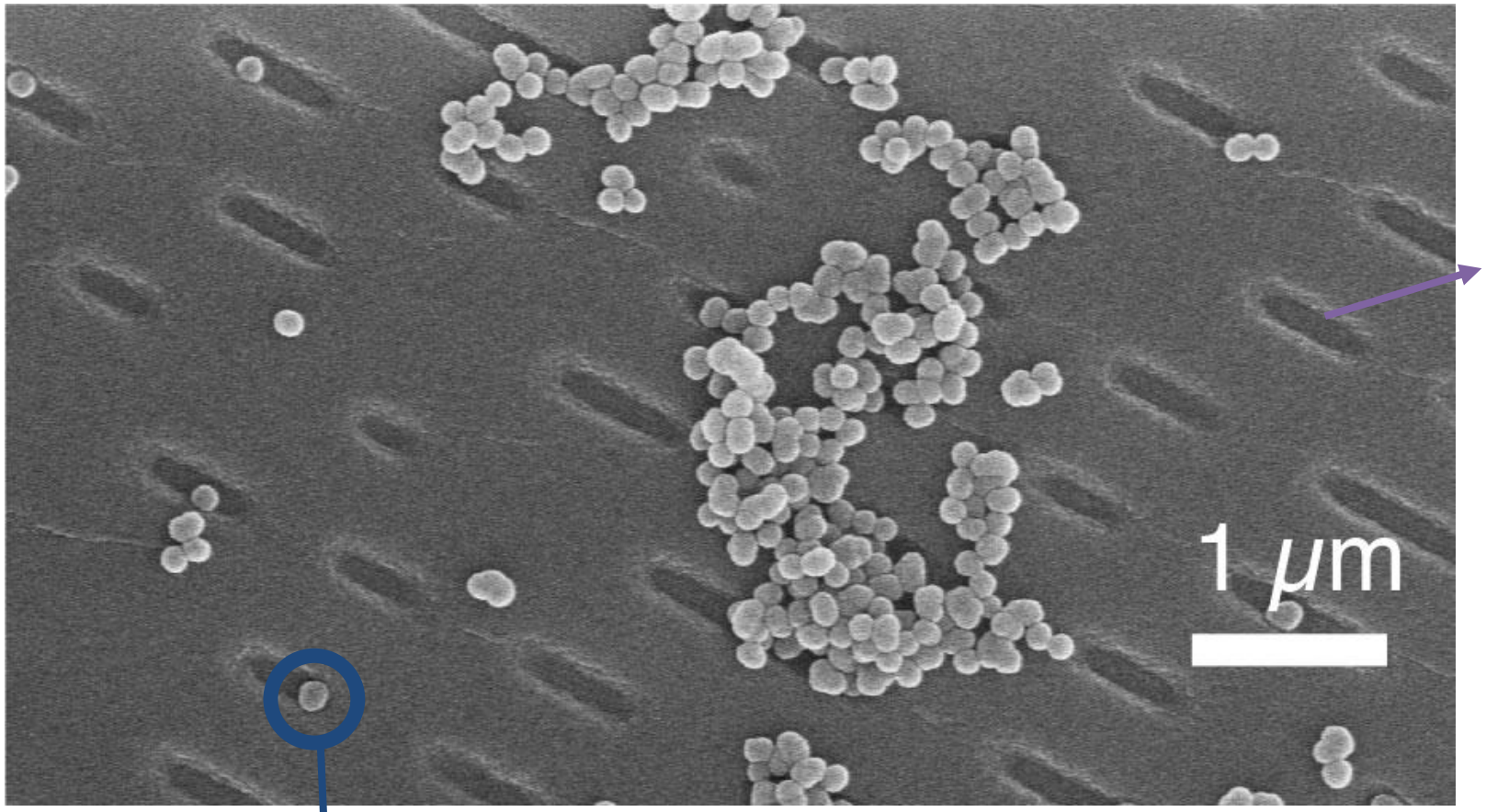


Paunescu et al., *Angew. Chem. Int. Ed.* **2013.**
Organick et al. *Small Methods* **2021**.

DNA long-term storage format

Antkowiak et al. *Small* **2022**.

Size difference: DNA Particles on DVD

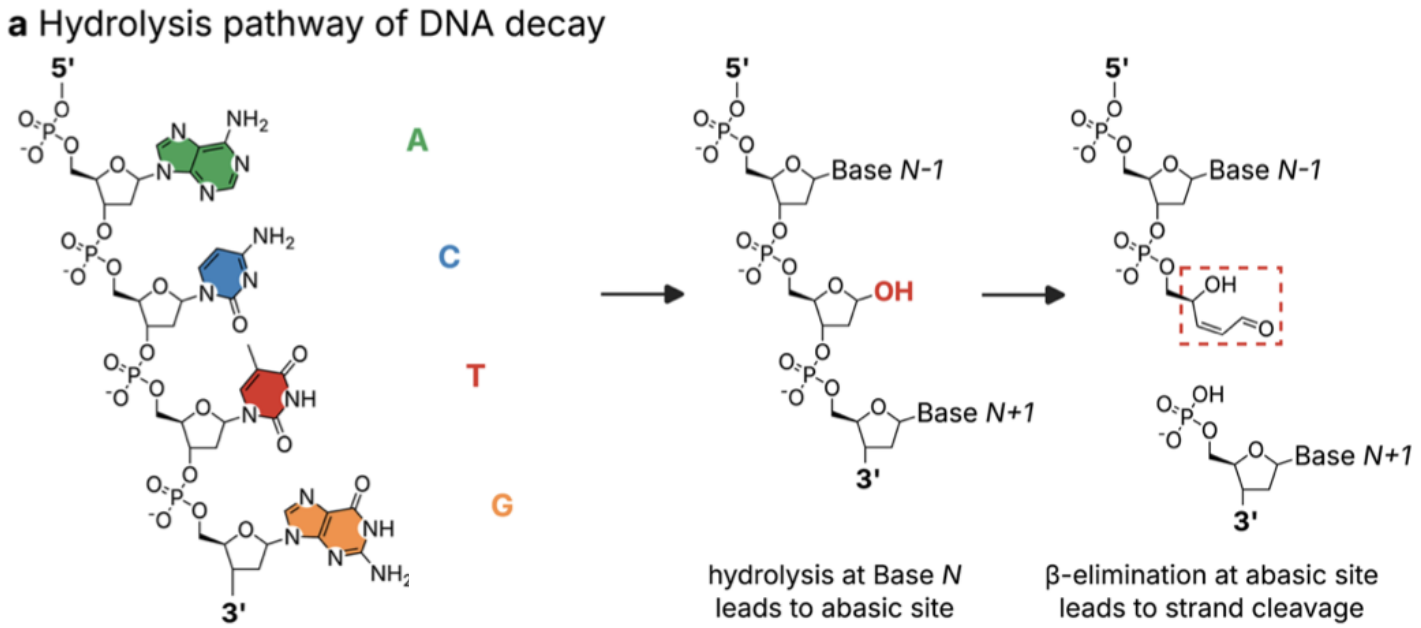


DVD
~ 3 bit

Encapsulated DNA: 10000 bit

On this image
~ 160 bit DVD
> 1 million bit DNA

DNA decay

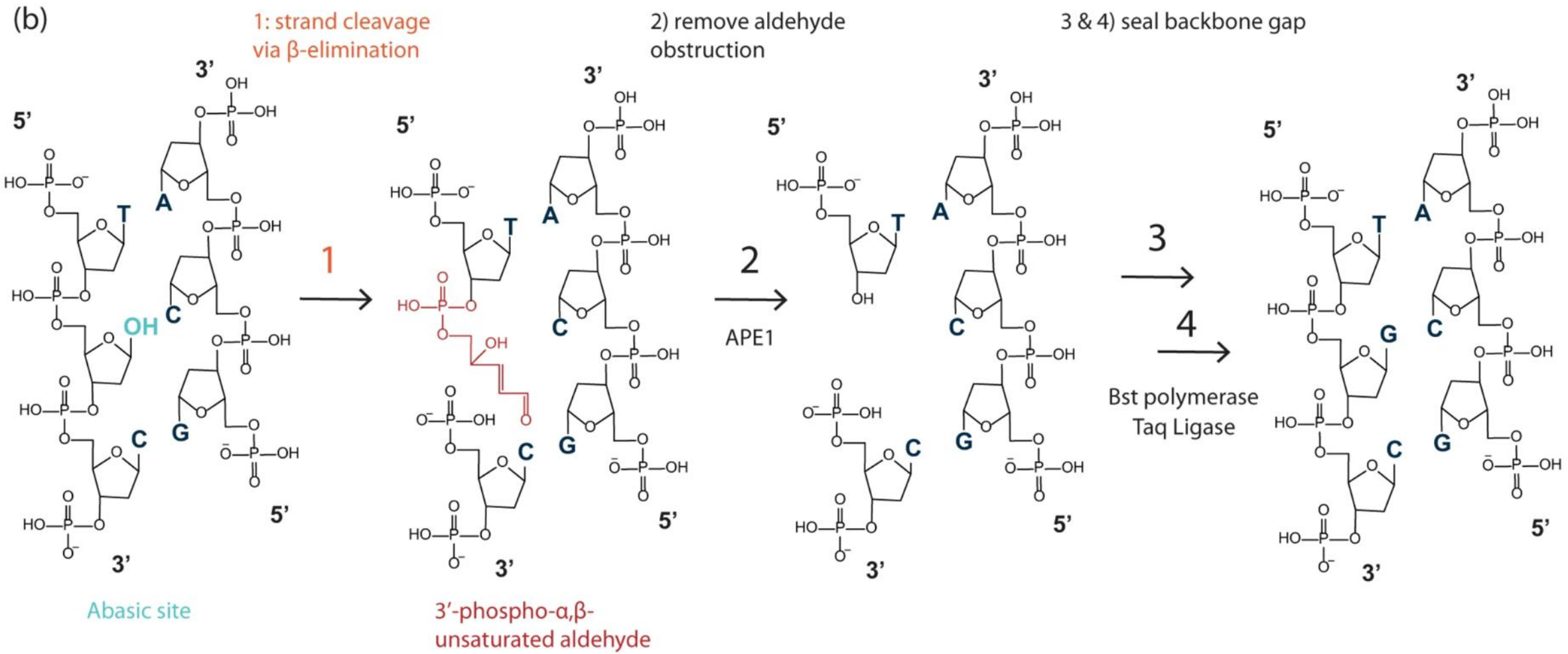


The decay rate is per nucleotide in the strand

→ Rate of having at least one break in a strand scales (linearly) with length of DNA

Lindahl, *Nature*, **1993**.
Gimpel et al. *Digital Discovey,* **2024**
Meiser et al. *Comms. Biol.* **2022.**

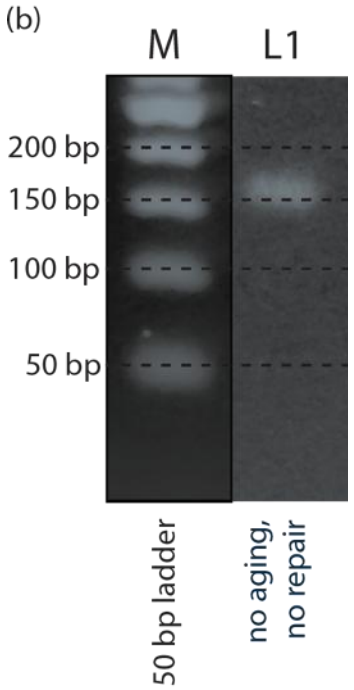
DNA decay and repair



Can only be repaired if unsaturated aldehyde is removed

Meiser et al. *Biol. Commun.* **2022.**

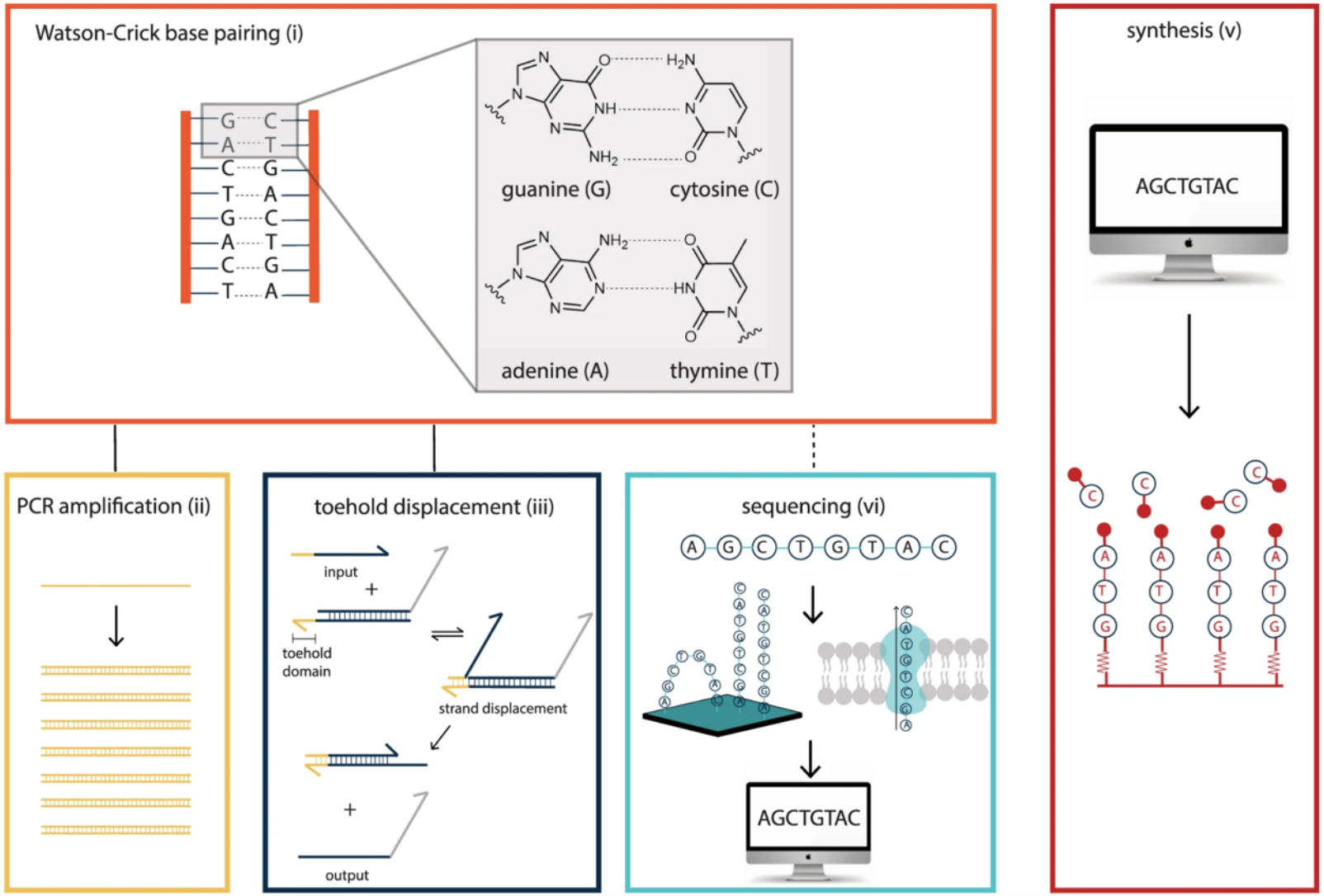
Only for DNA: Molecular repair



We can reverse some DNA decay by enz. reaction but DNA decay still has to be prevented

Meiser et al. *Comms. Biol.* **2022.**

DNA is a really special molecule



+ have technologies to prevent DNA decay

Meiser et al. *Nat. Commun.* **2022**.

What can we do with it?

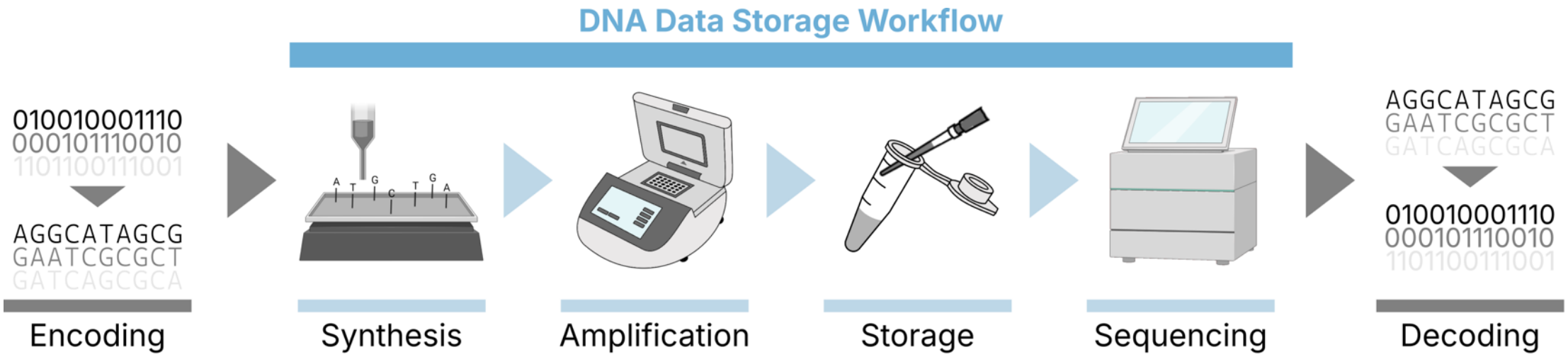
- DNA data storage

- DNA computing *(will not be discussed here)*

- DNA barcoding
- DNA of things

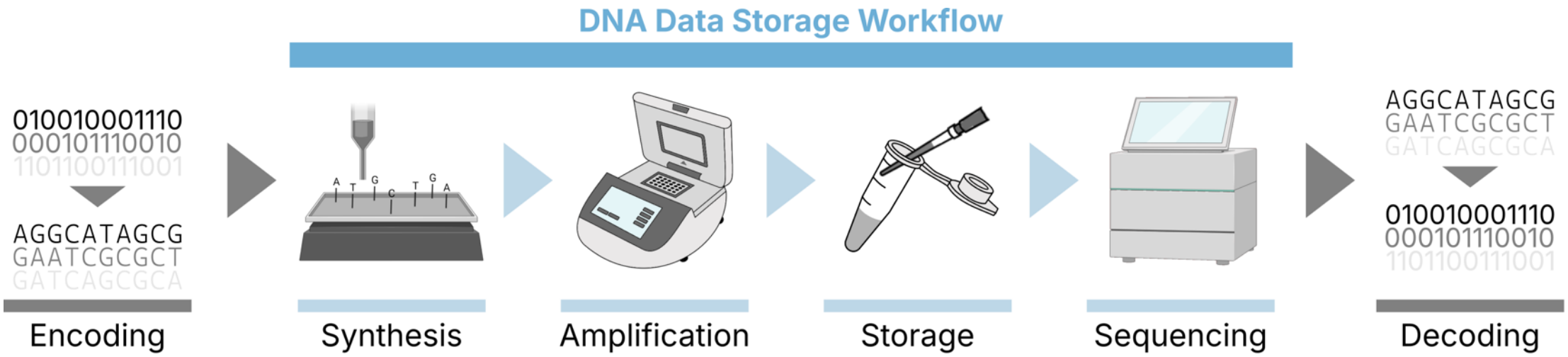
- DNA cryptography – chemical unique functions

DNA data storage



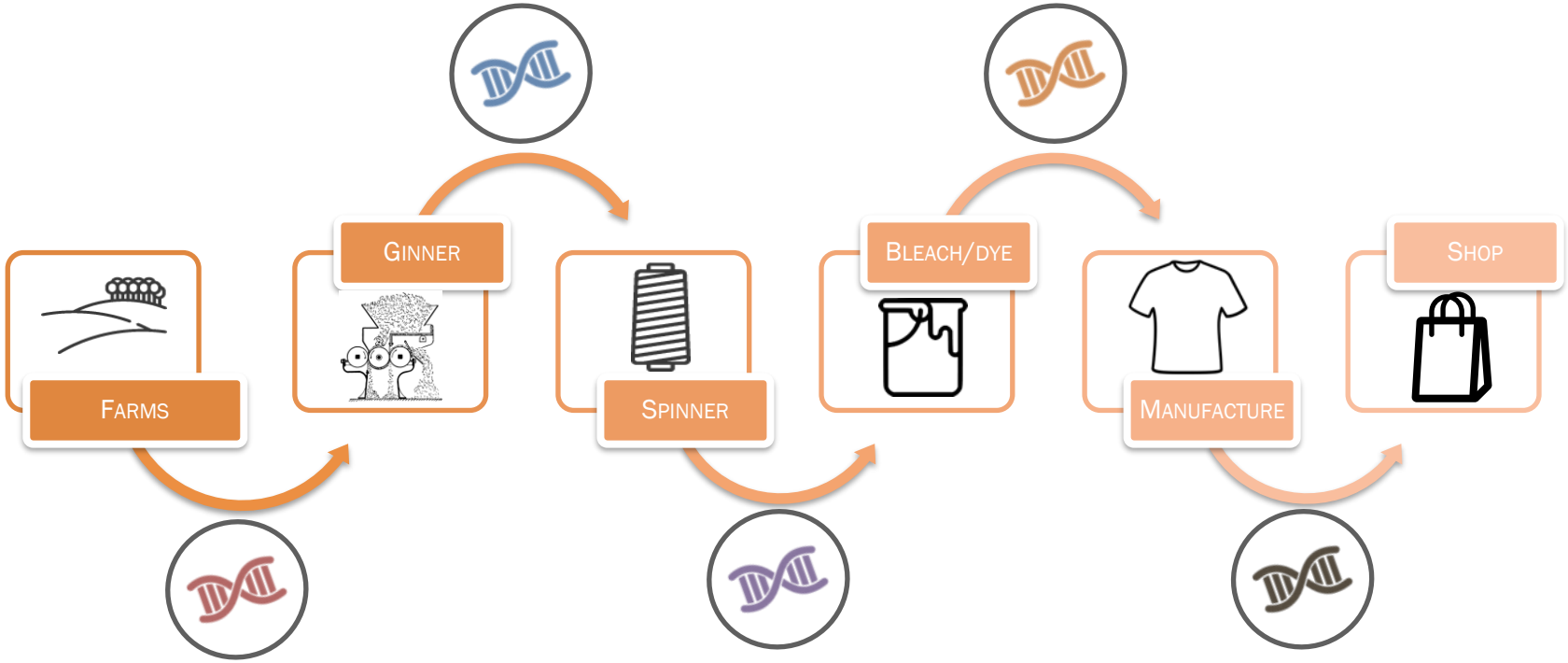
Gimpel et al. *Digital Discovery* **2024**.

DNA data storage



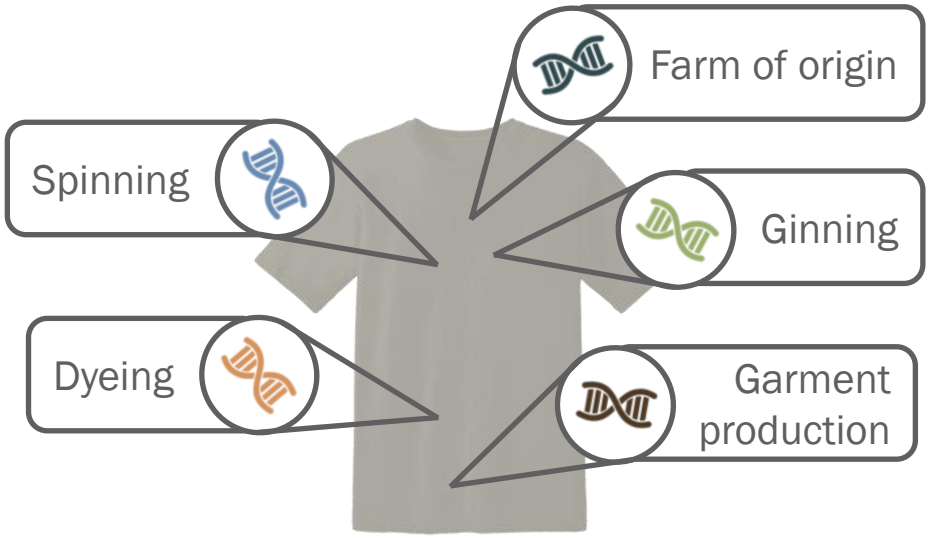
Gimpel et al. *Digital Discovery* **2024**.


DNA-containing capsules to directly label cotton
Haelixa
ETH zürich
Patented


FARMS
GINNER
SPINNER
BLEACH/DYE
MANUFACTURE
SHOP

IN-product labels

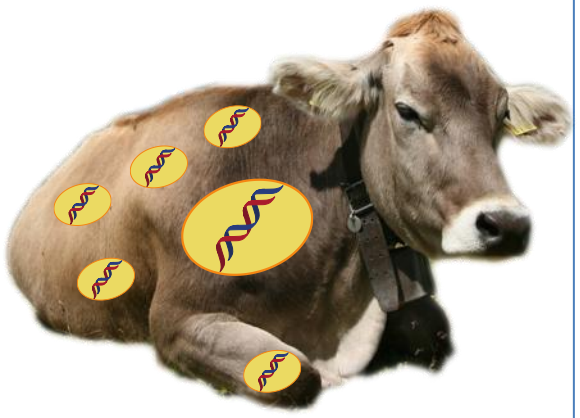
Each distinct marker represents a label of quality and unique identifier for each individual producer, manufacturer, location or batch. All of this information is stored in the product and further used to unravel its journey.



DNA marked cashmere in shops

Product information in products

NATURE



ca. 1.5 GB of genetic Info. per cell as DNA

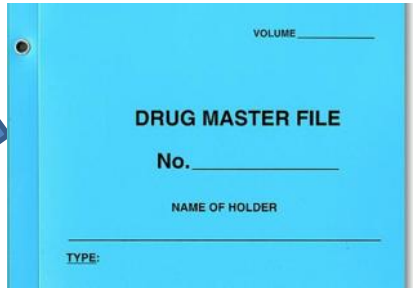
~ 10^{13} cells per body

Today

Product identification

Documentation
(Paper & Digital)

Future

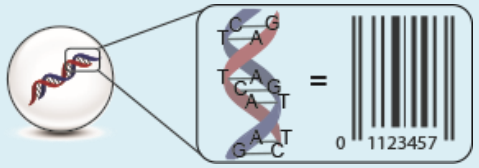


Photos: wikipedia.org

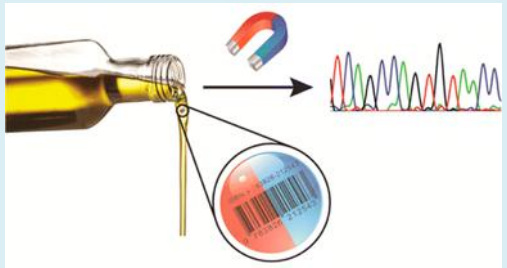
DNA of things

DNA =
Chemical Information

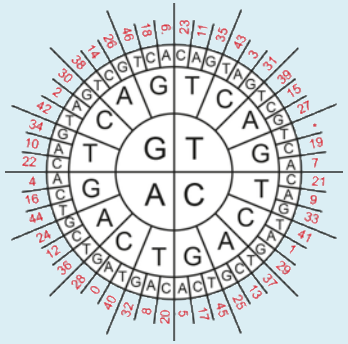
DNA Product barcoding



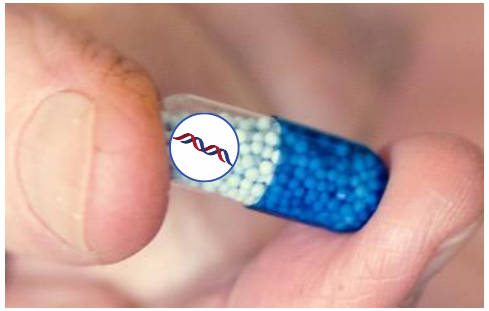
Individual strands



Digital data archive

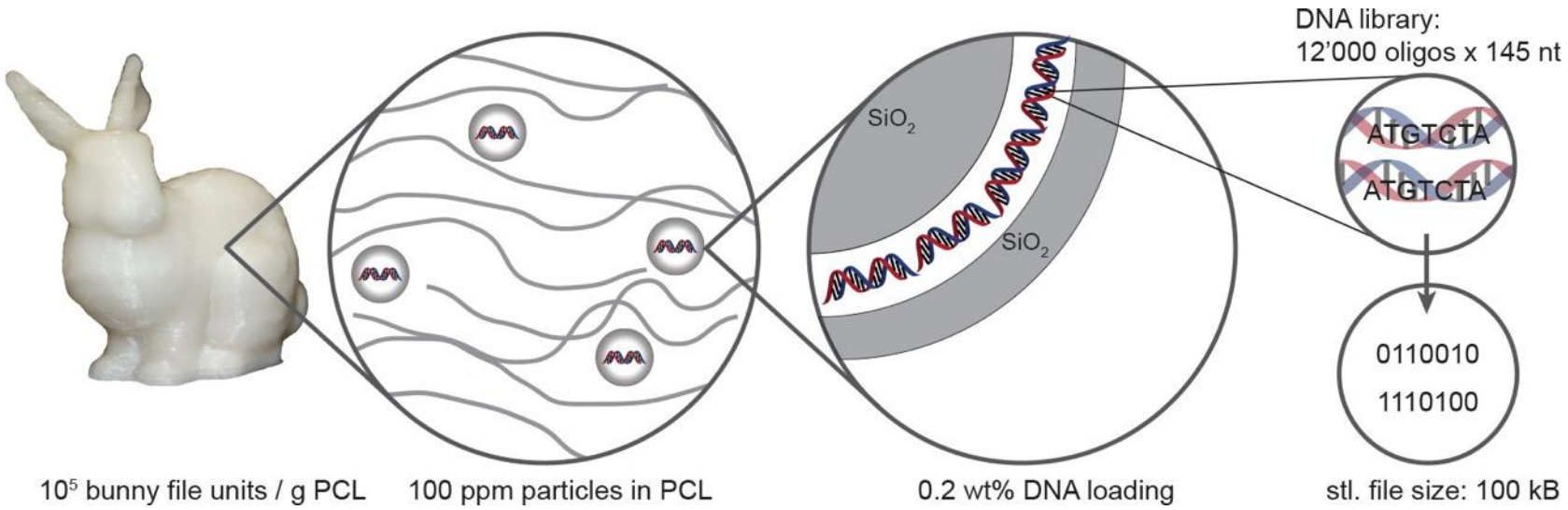


Millions of DNA strands (15 MB)

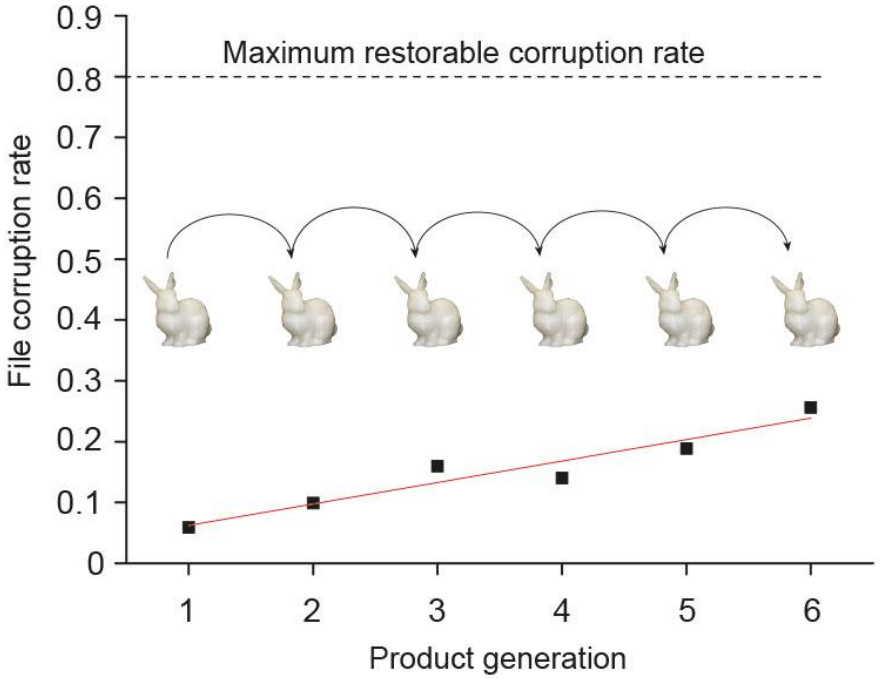


DNA of things

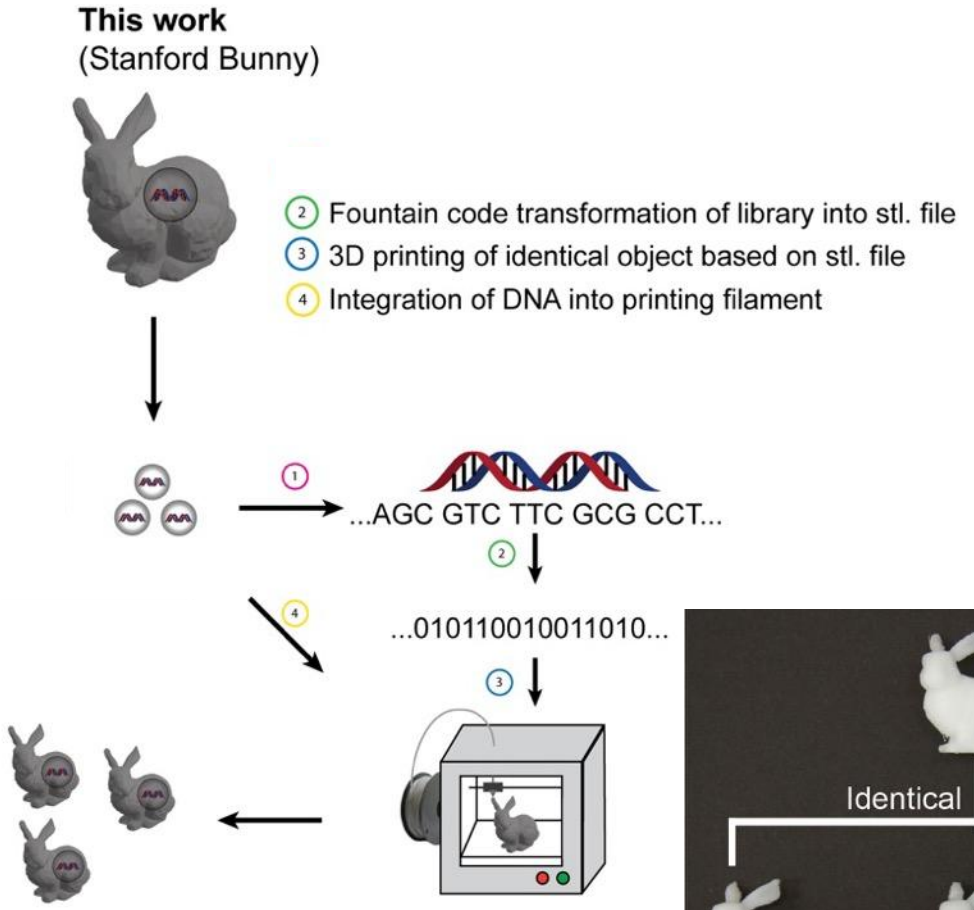
Application in 3D printing

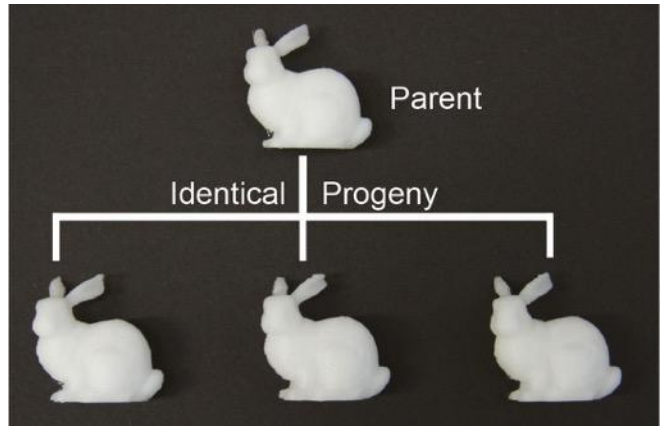


* collaboration Erlich & Grass
Koch et al. *Nat. Biotechnol.* **2020**



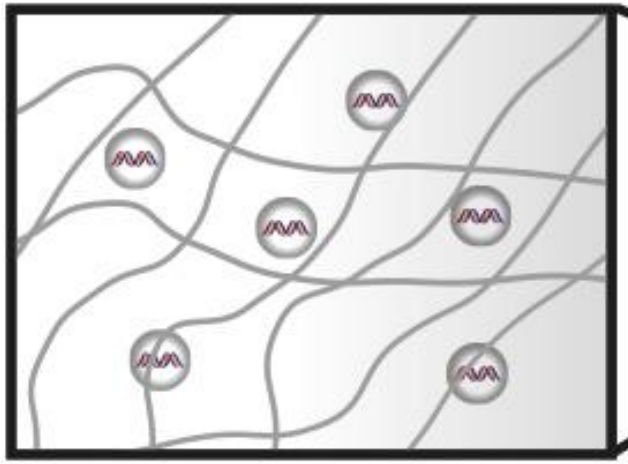
enables 8×10^{19} bunnies





* collaboration Erlich & Grass
Koch et al. *Nat. Biotechnol.* **2020**

Storing data in everyday objects



Digital Data file in DNA hidden in PMMA

600'000 copies of a 1.4 MB movie stored as DNA in the lens

* collaboration Erlich & Grass

Koch et al. *Nat. Biotechnol.* **2020**

DNA of democracy

- 75th anniversary of Germany's *Grundgesetz*
- Grundgesetz encoded into synthetic DNA
- Links biological heritage and cultural legacy
- DNA (in silica) added to ink of newspaper 400'000 edition, 18.05.2024

DNA der Demokratie

Wer diese SZ-Ausgabe besitzt, hält das Grundgesetz in seinen Händen. Es wurde als Erbgut in der Tinte gespeichert.

Diese Ausgabe der *Süddeutschen Zeitung* ist nicht nur eine Sonderedition zum 75. Jahrestag des Grundgesetzes. Sie ist in ihrer Papierform ein Symbol. Das Grundgesetz ist das Erbgut der Demokratie und der deutschen Gesellschaft. In der Titelseite dieser Papierzeitung steckt dieses Erbgut im wahrsten Sinne des Wortes. In der Tinte dieser Seite ist das Grundgesetz gespeichert. Das geht mit einem neuen Verfahren, das genauso funktioniert wie biologisches Erbgut. Science-Fiction? Ja. Aber synthetische DNA ist auch das Speichermedium der Zukunft.

Herkömmliche Geräte wie Festplatten oder USB-Sticks sind nach 40 Jahren kaputt. Nun haben Wissenschaftler herausgefunden, dass man Daten auf Molekülsträngen speichern kann, die der DNA des menschlichen Körpers entsprechen. Man kennt sie aus dem Biologieunterricht. Das ist die Doppel-Helix mit den vier Buchstaben A, C, G und T in immer neuen Kombinationen. Der Informatiker Reinhard Heckel von der TU München und der Chemiker Robert Grass von der ETH Zürich haben einen Weg gefunden, Computerdaten auf DNA zu speichern, die nicht von Menschen oder Tieren stammt, sondern künstlich hergestellt wird. In Zukunft sollen Daten so über Hunderte, wenn nicht Millionen Jahre erhalten werden.

Diese synthetische DNA, auf die Heckel und Grass für die *Süddeutsche Zeitung* eine Million Kopien des Grundgesetzes speicherten, ist mit bloßem Auge kaum erkennbar. Sie erinnert ein wenig an den feinen Sand in einer Sanduhr. Dieses Pulver wurde an den vier Druckstandorten der SZ in die Tinte für die Titelseite, die Anzeige auf Seite 7 und die Anzeige auf Seite 17 gemischt. Biologisch und chemisch ist das unbedenklich. Synthetische DNA hat keine biologische oder chemische Funktion. Sie dient auch jetzt schon als unsichtbare Markierung, in Europa für Textilien, in den USA für Lebensmittel. Das soll Markenfälschung verhindern. Menschen tragen sie also am Körper oder essen sie. Auslesen können Sie die DNA noch nicht. Für Heckel und Grass ist die Aktion ein Experiment. Sie werden Ausgaben dieser Zeitung in ihrem Labor auseinandernehmen und versuchen, das Grundgesetz wieder zu rekonstruieren. Die SZ hat für diese Seite mit Initiator Daniel Koller und dem Künstler Solimán López zusammengearbeitet, die diese Tinte für eine mehrmonatige Kampagne für die Demokratie verwenden werden.

Die SZ will mit dem Grundgesetz in der DNA-Tinte der Titelseite ein Zeichen setzen. Für die Demokratie. Für die Bürgerrechte. Für die Zukunft. Wir freuen uns, wenn Sie sie als Erinnerung daran behalten, dass nicht nur das Grundgesetz, sondern auch die freie Presse eine der Säulen unserer Gesellschaft ist. **Andrian Kreye**

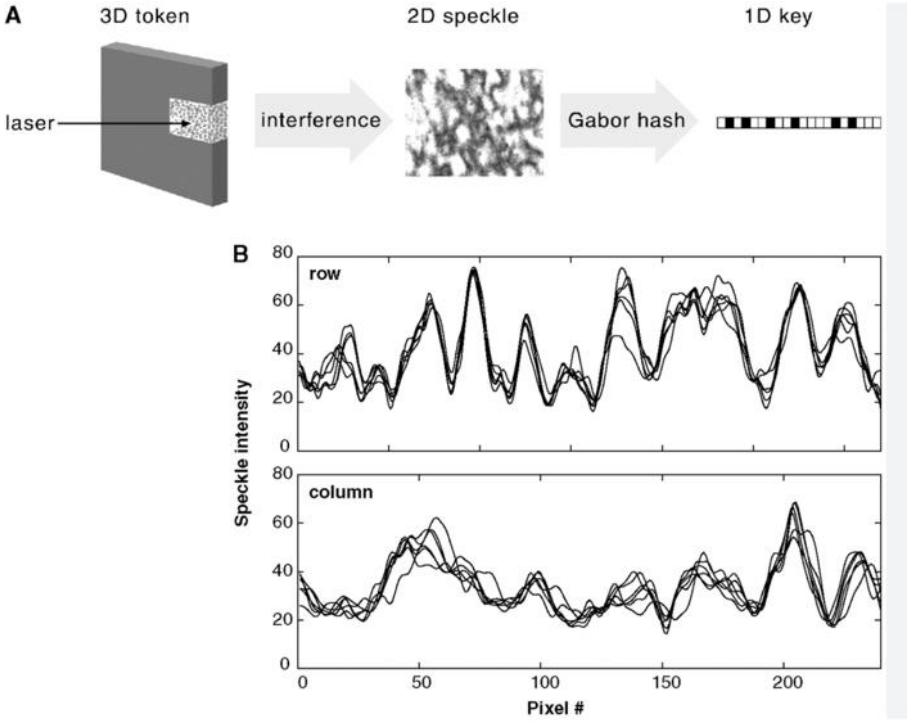
Massive Attack Mezzanine

TCTACTGGCCGCGAGATCGGAAGAGCACACGTCT
TCTCTGTGTTAAGAGATCGGAAGAGCACACGTCT
TCTGTAATTTACTAGATCGGAAGAGCACACGTCT
TCTTACAATTTGCAGATCGGAAGAGCACACGTCT
TCGATCAGCACTAAGATCGGAAGAGCACACGTCT
TCGCAGGGAAAGTAGATCGGAAGAGCACACGTCT
TCGGCCTTCTCCTAGATCGGAAGAGCACACGTCT
TCGTTGCCCAACAAGATCGGAAGAGCACACGTCT
TCCAAAATAGTTTAGATCGGAAGAGCACACGTCT
TCCCCTATATAAAAGATCGGAAGAGCACACGTCT
TCCGCTTTAAGCCAGATCGGAAGAGCACACGTCT
TCCTCCAAAGACCAGATCGGAAGAGCACACGTCT
TCAAGGCCGATCGAGATCGGAAGAGCACACGTCT
TCACGACTATCACAGATCGGAAGAGCACACGTCT
TCAGGGCTATAGGAGATCGGAAGAGCACACGTCT
TCATTGCGCACTGAGATCGGAAGAGCACACGTCT
TATACTGAATCATAGATCGGAAGAGCACACGTCT
TATCGGTTGAGACAGATCGGAAGAGCACACGTCT
TATGGTCTTTAACAGATCGGAAGAGCACACGTCT
TATTGCTTGAGTTAGATCGGAAGAGCACACGTCT
TAGAGCGCAGTGGAGATCGGAAGAGCACACGTCT
TAGCGGCAGCGTAAGATCGGAAGAGCACACGTCT
TAGGGAAACCGACAGATCGGAAGAGCACACGTCT
TAGTTGGCTAGTAAGATCGGAAGAGCACACGTCT
TACATCCAGCGAAAGATCGGAAGAGCACACGTCT
TACCGCAAGGCCGAGATCGGAAGAGCACACGTCT
TACGACCGTCAGCAGATCGGAAGAGCACACGTCT
TACTGCATATGAAAGATCGGAAGAGCACACGTCT
TAAATAGCCGGTAAGATCGGAAGAGCACACGTCT
TAACATATATGGTAGATCGGAAGAGCACACGTCT
TAAGCAGGTCCTGAGATCGGAAGAGCACACGTCT
TAATACAGAGCGTAGATCGGAAGAGCACACGTCT
TTTACGCCCCTATAGATCGGAAGAGCACACGTCT
TTTCGAAACTAGGAGATCGGAAGAGCACACGTCT
TTTGCCTCTTCTAAGATCGGAAGAGCACACGTCT
TTTTCCCGAGCGAAGATCGGAAGAGCACACGTCT
TTGACCGTCTCCAAGATCGGAAGAGCACACGTCT
TTGCTGCCTCTCGAGATCGGAAGAGCACACGTCT
TTGGCATAAGGGTAGATCGGAAGAGCACACGTCT
TTGTATCCCACCAAGATCGGAAGAGCACACGTCT
TTCAGTGCGTCGCAGATCGGAAGAGCACACGTCT
TTCCGCGCTGTCAAGATCGGAAGAGCACACGTCT
TTCGCACCACGCCAGATCGGAAGAGCACACGTCT
TTCTCCCCAATCGAGATCGGAAGAGCACACGTCT
TTAAGGACGTAAGAGATCGGAAGAGCACACGTCT
TTACAAGTTCGCAAGATCGGAAGAGCACACGTCT
TTAGCCAACCAACAGATCGGAAGAGCACACGTCT

Netflix Biohackers

- To promote Netflix German TV Series "Biohackers"
  - Store full episode (40 min) in DNA (ca. 65 MB, 360p)
  - 3.6 million DNA sequences (each 140 nt long)
  - Every vial contains 1 million copies of the whole episode

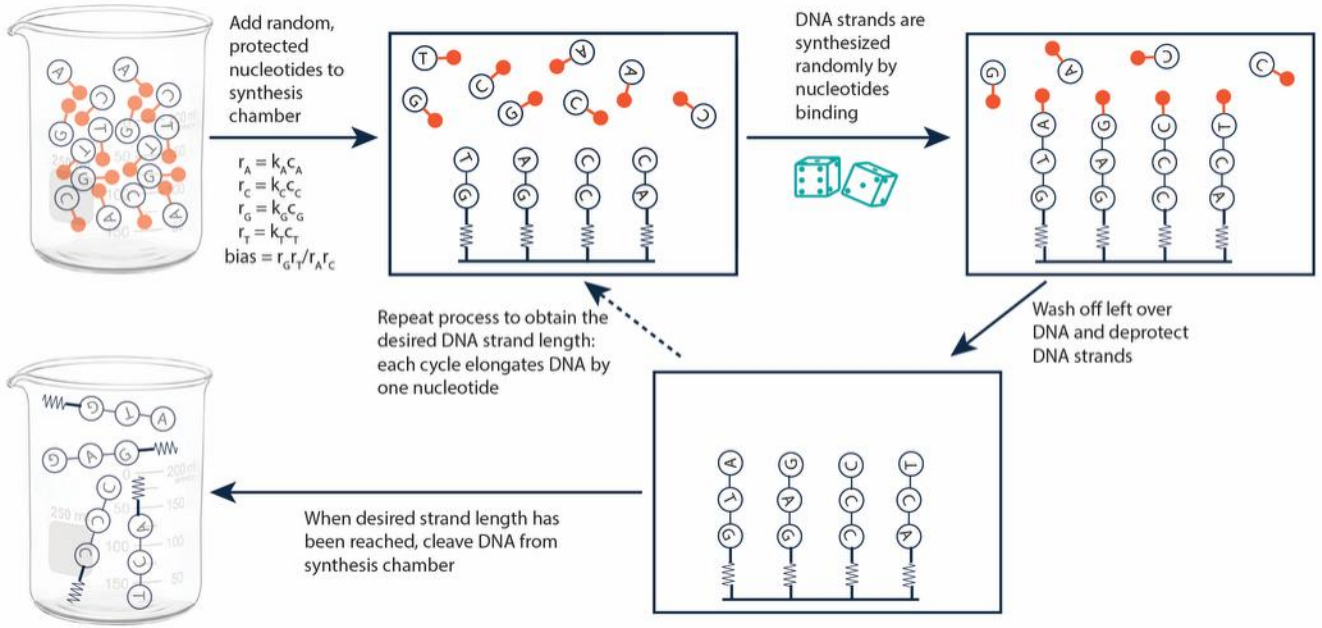
Cryptography using random objects



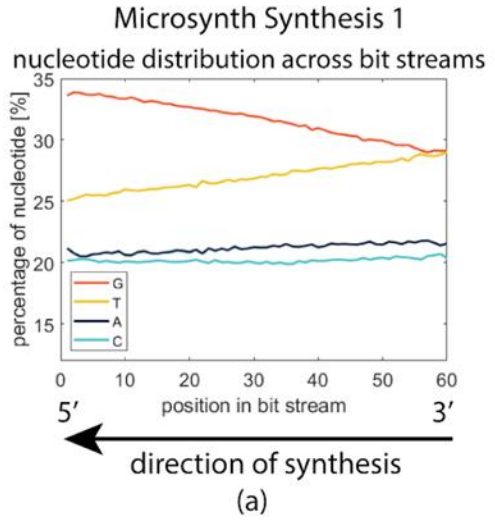
A physical one way function = Physical unclonable function

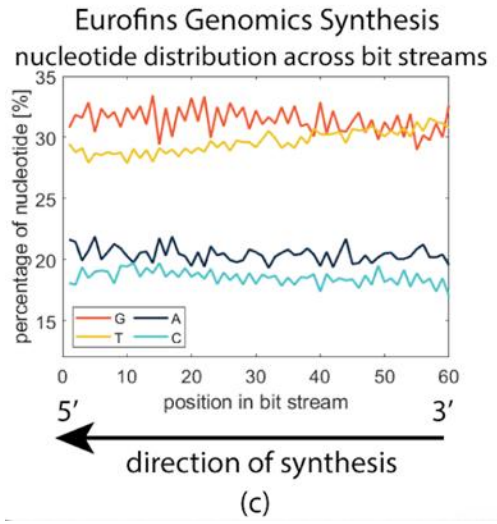
Pappu et al. *Science* **2002**.

Random DNA synthesis



Typical synthesis: 200 µg DNA = 10^{16} different DNA strands

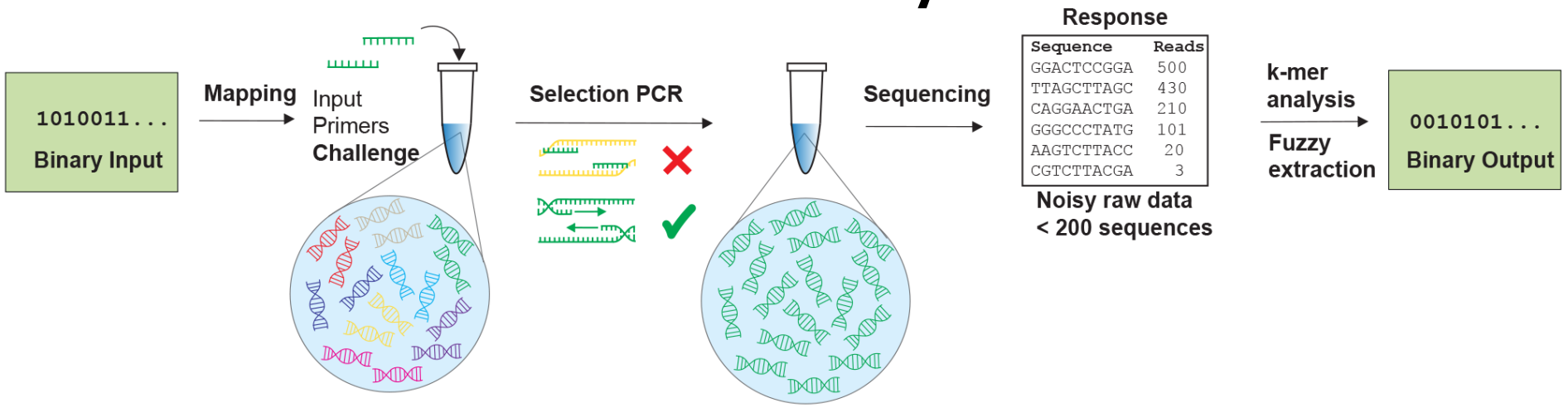


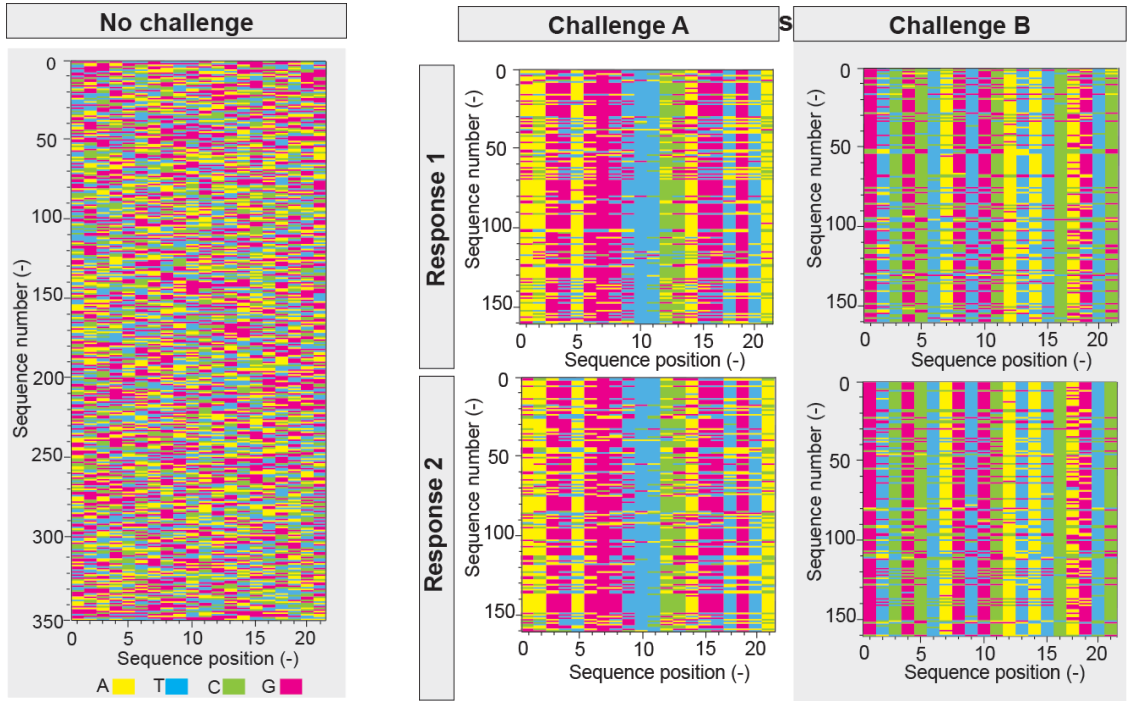


→ Bias from synthesis
→ Debiasing (e.g. von Neumann)
→ 7 million GB of randomness from one synthesis run (100 Usd)

Meiser et al. *Nat. Commun.* **2020.**

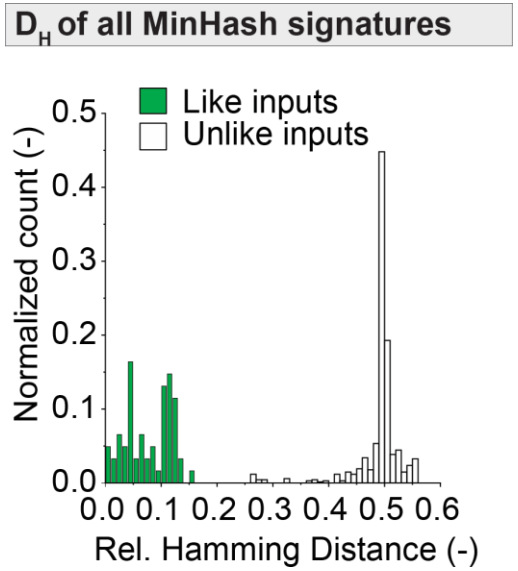
Chemcial one way functions

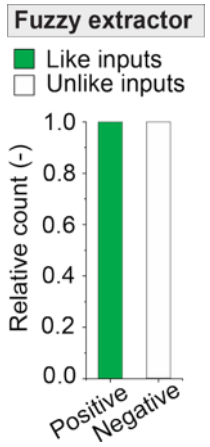




Luescher et al., *Nat. Commun.* **2024**

Same input and pool → same output

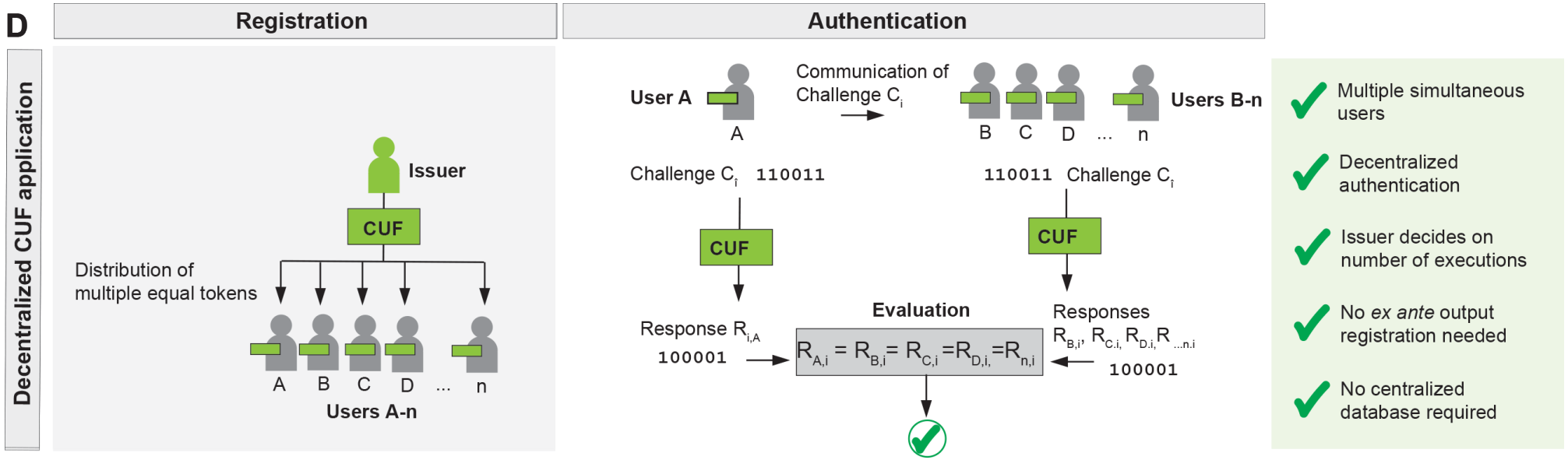




Cryptographic bio-chemical functions

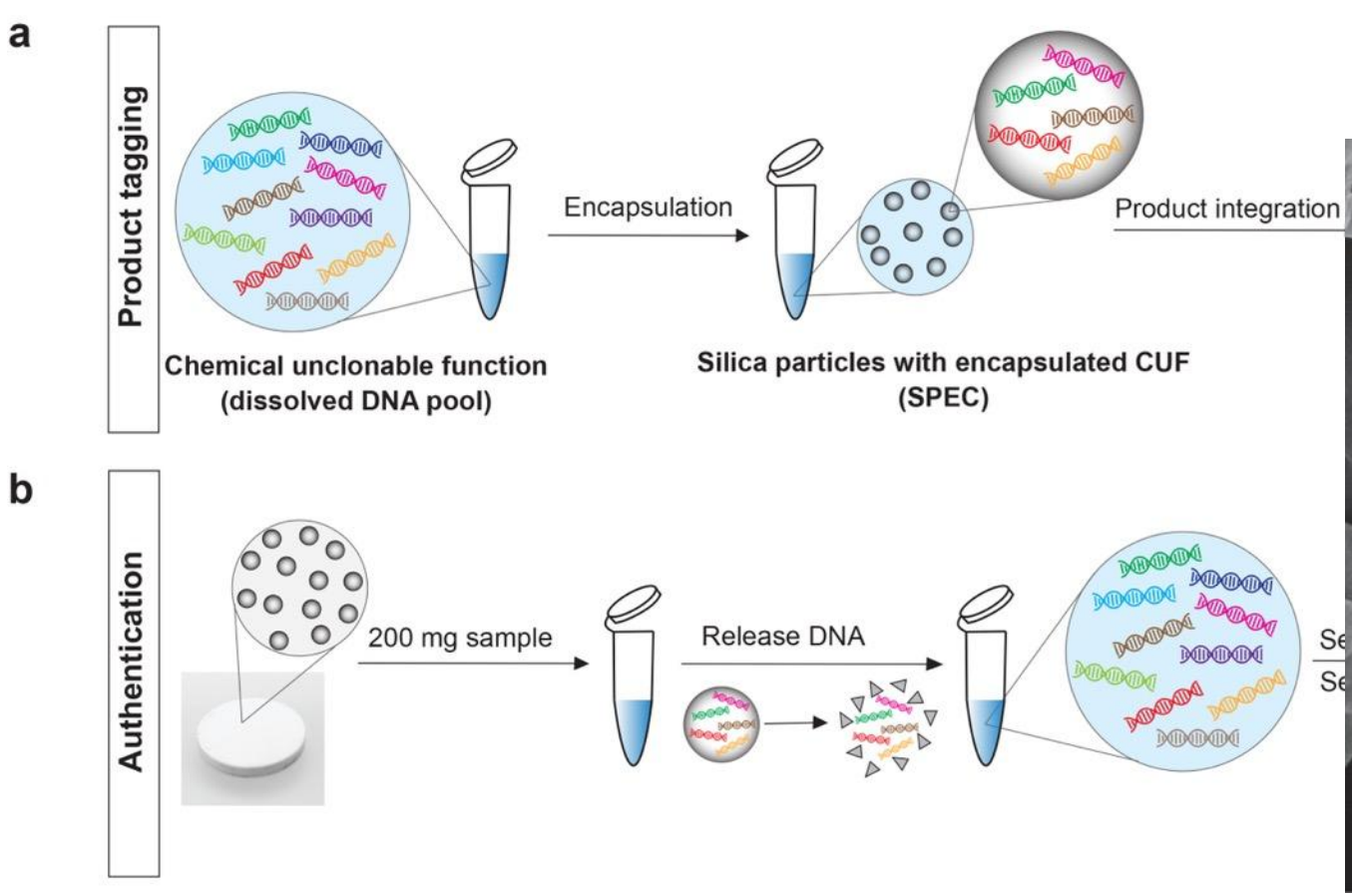
Luescher et al., *Nat. Commun.* **2024**

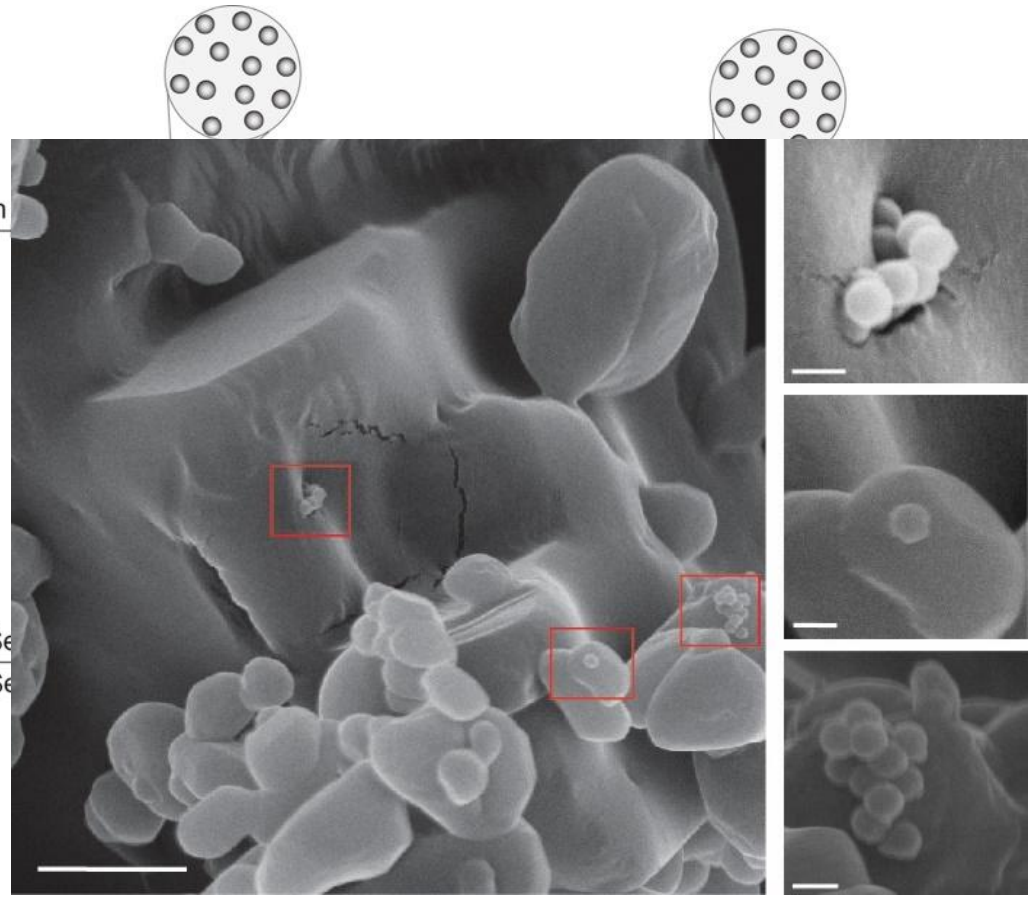
Decentralized authentication



Luescher et al., *Nat. Commun.* **2024**

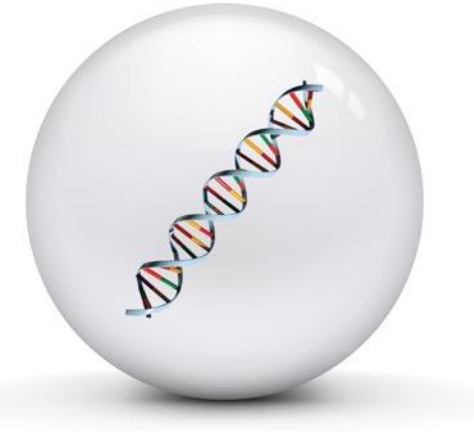
Application to paracetamol formulation





Luescher et al., *ACS Nano,* **2024**

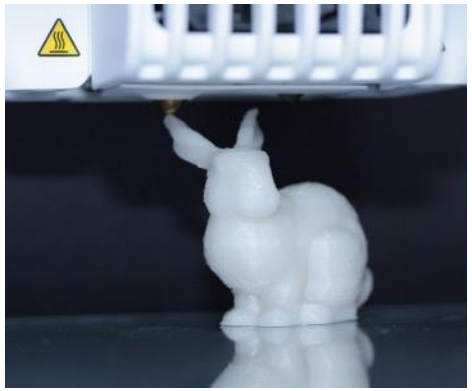
Why we love DNA



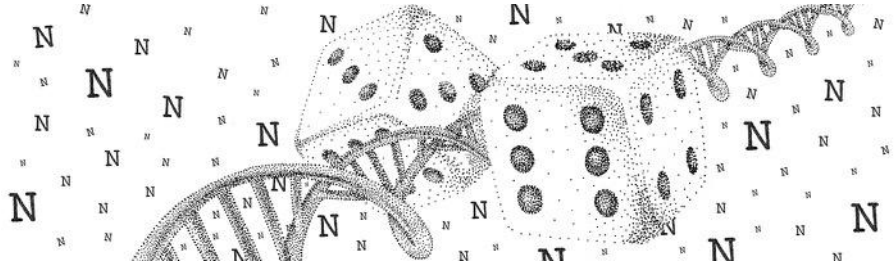
- Unique molecule
  - Self assembly, amplification, single molecule detection /sequencing
  - Can be stabilized by ”fossilization” and protection from water

- Sequences allow digital representation of
  - Information
  - Entropy

$$I_{max} = I + H = const$$



- Useful
  - for distribution/integration of information into objects
  - as a source of randomness & entropy



Acknowledgments

Functional Materials Laboratory

Daniela Paunescu, Linda Meiser, Francesca Granito
Philipp Antkowiak, Anne Lüscher, Andreas Gimpel

Prof. Reinhard Heckel (TUM)

Prof. Karsten Borgwardt (MPI)

Prof. Wendelin J. Stark (ETH)

Funding

FNSNF
Fonds national suisse
Schweizerischer Nationalfonds
Fondo nazionale svizzero
Swiss National Science Foundation

Marie Curie Actions

Seventh Framework Programme

Microsoft

FILTEX

Horizon 2020: DNA Fairylights

DNA FAIRYLIGHTS

European Innovation Counil: DIDAX

Funded by the European Union

DIDA
DNA DATA APPLICATIONS