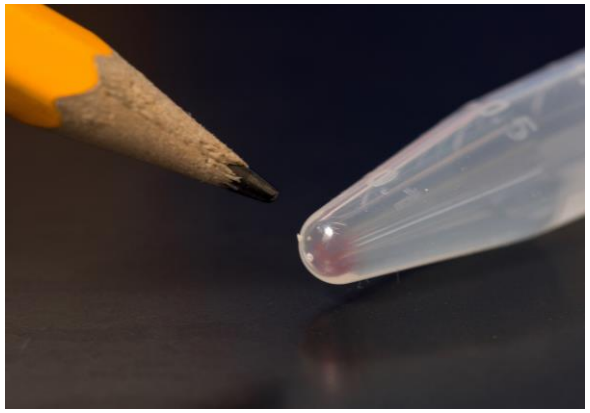
DNA data storage and computation

Karin Strauss, Microsoft
Luis Ceze, University of Washington

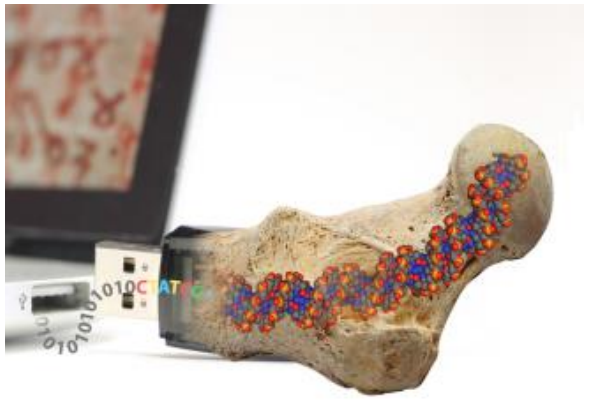
DNA data storage advantages

Density



Credit: Tara Brown Photography/University of Washington

Durability



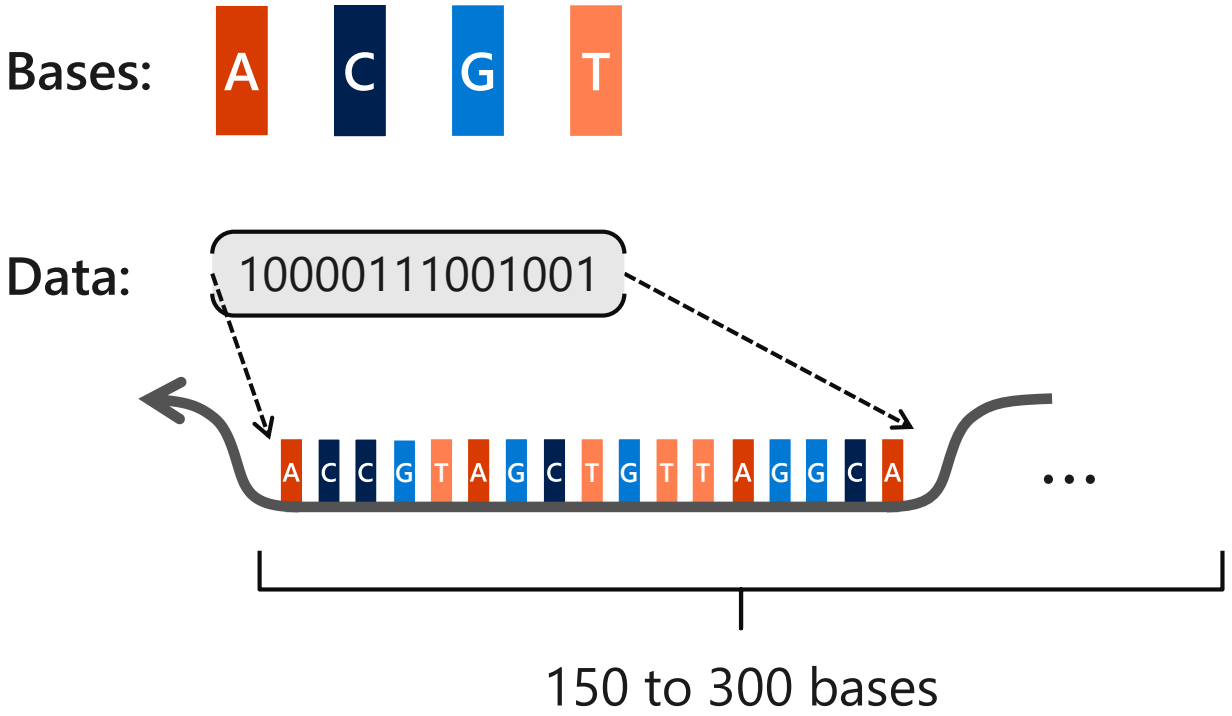
Credit: Philipp Stössel/ETH Zurich

No obsolescence issue

Credit: Illumina

Ability to perform computation

DNA data storage basics

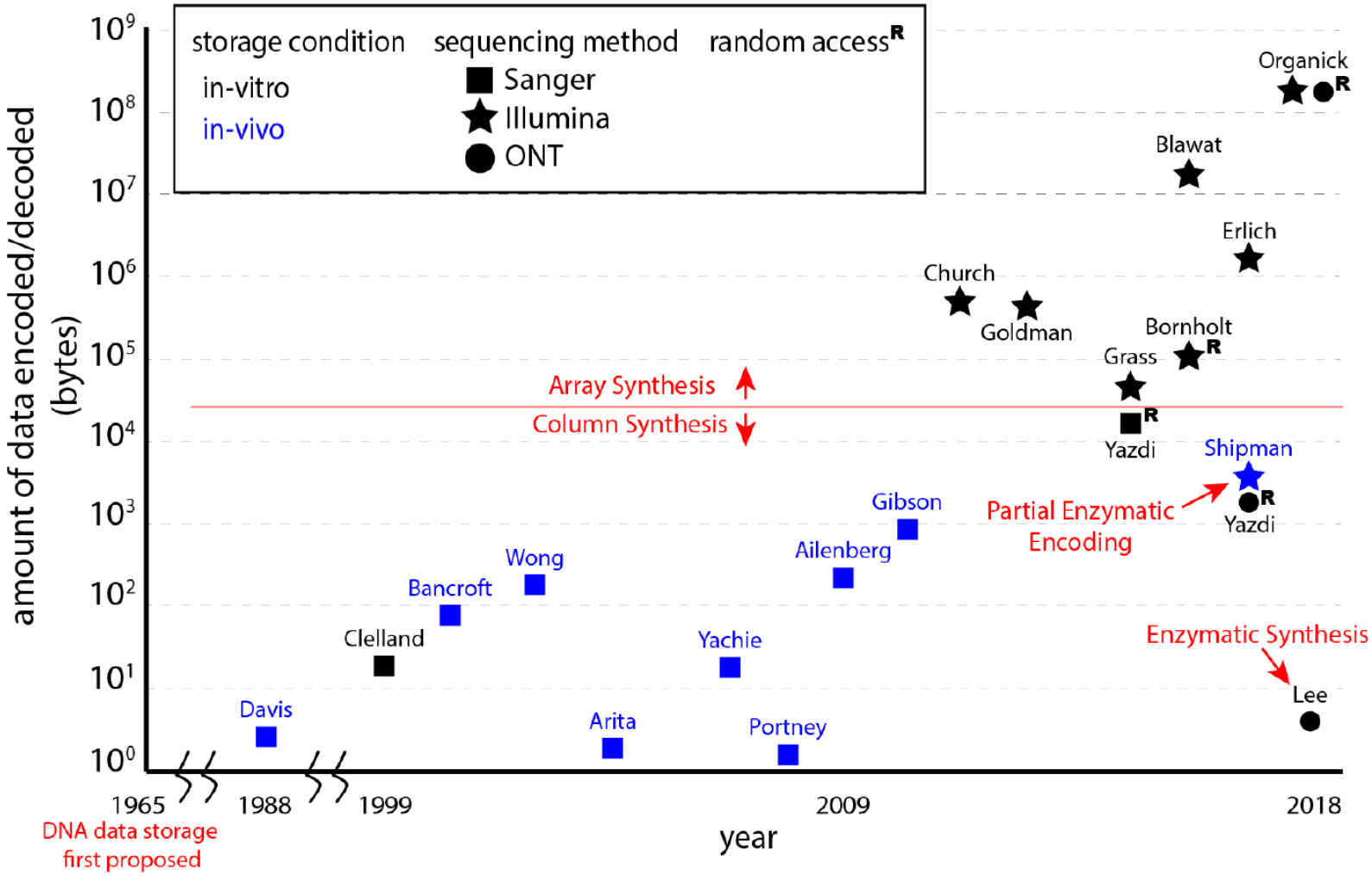


Simple mapping:

Bits	Base
00	A
01	C
10	G
11	T

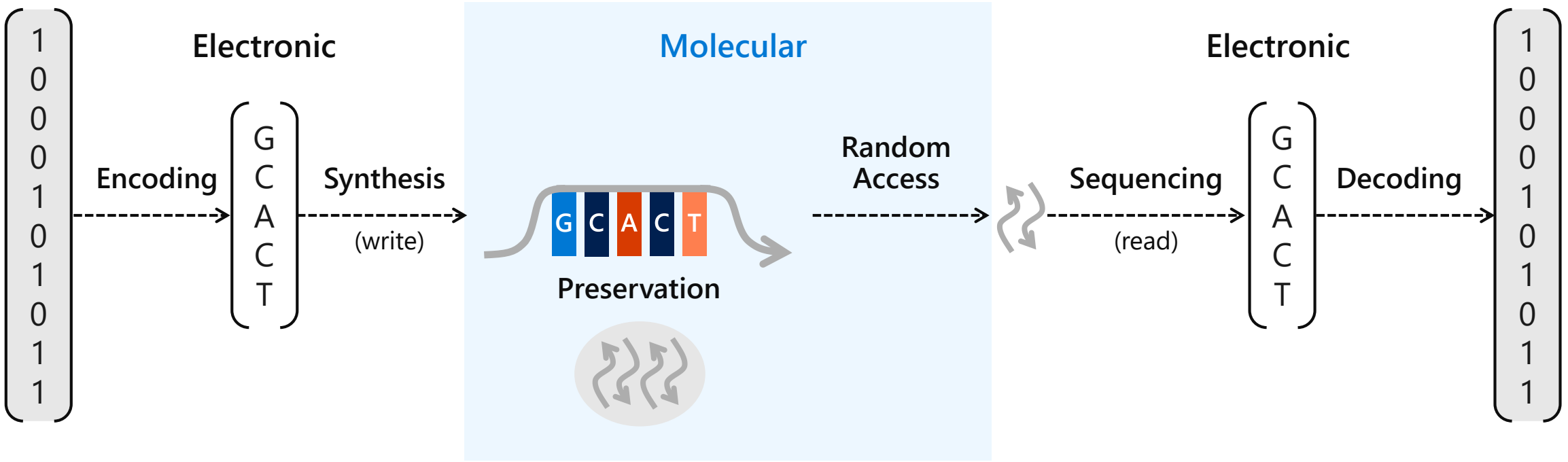
Store data in synthetic DNA strands

Improvements in DNA data storage

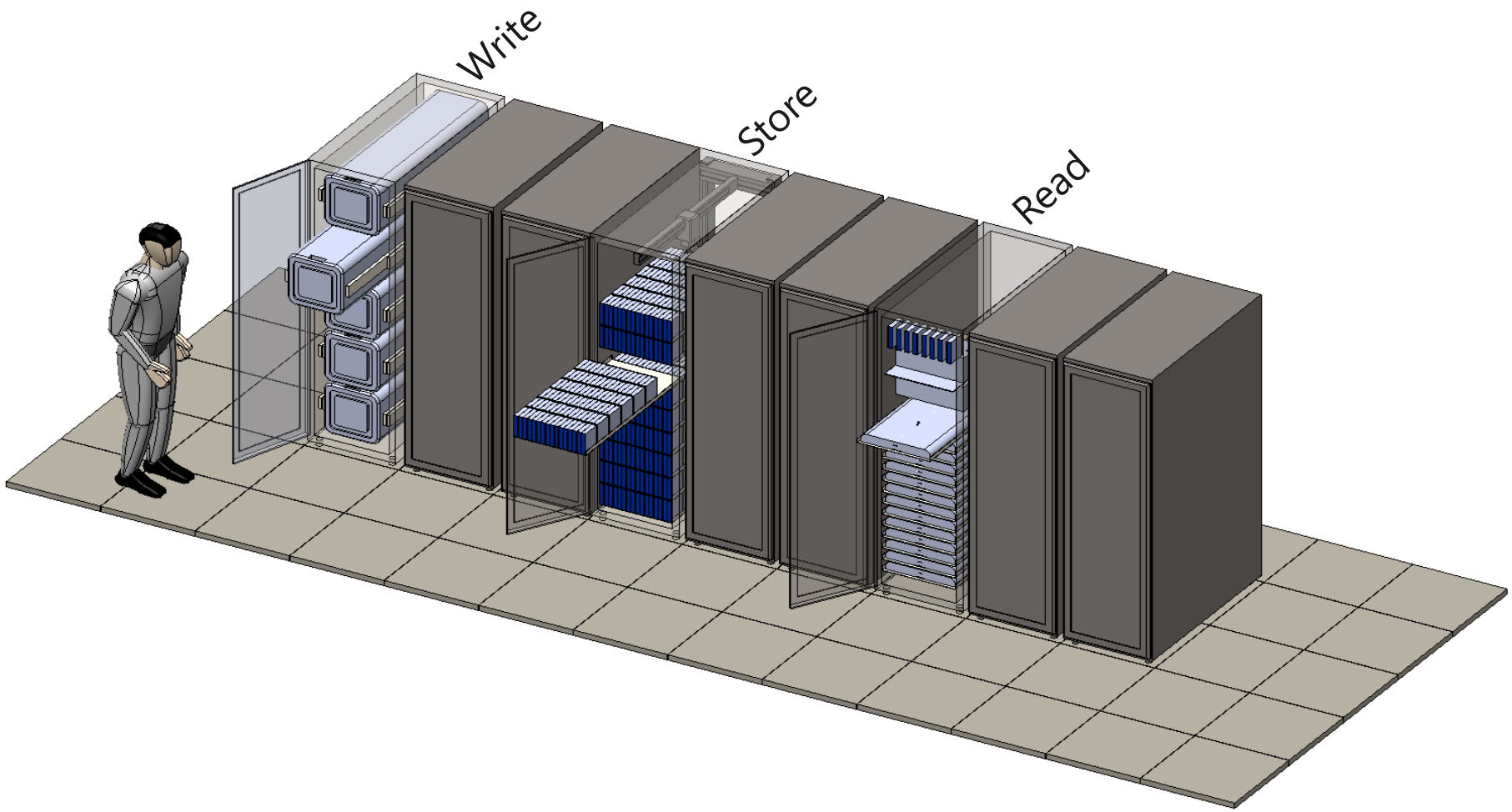


Ceze, Nivala and Strauss, Nature Reviews Genetics, 2019

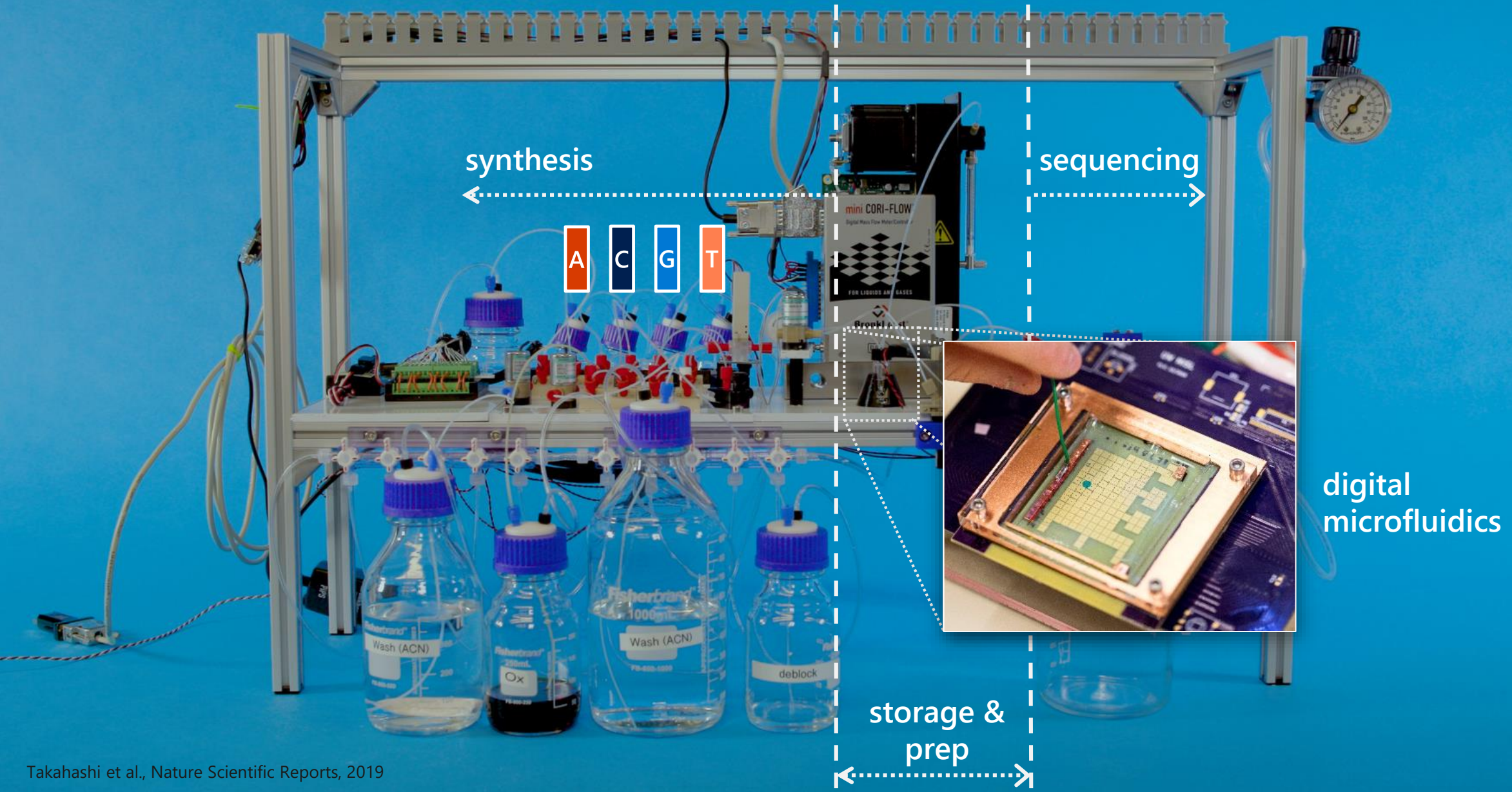
DNA storage end-to-end system



End-to-end system in a datacenter



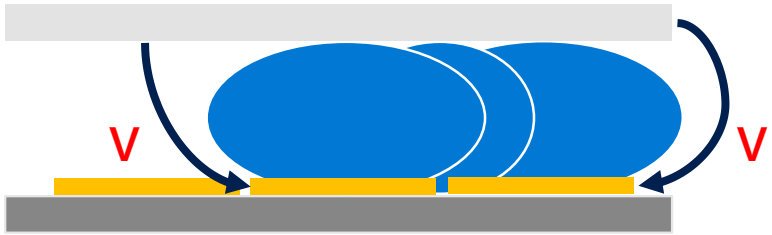
First fully automated DNA data storage system

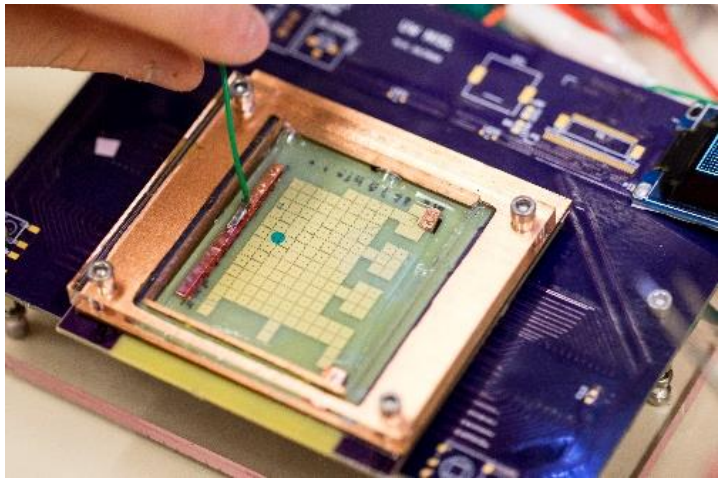


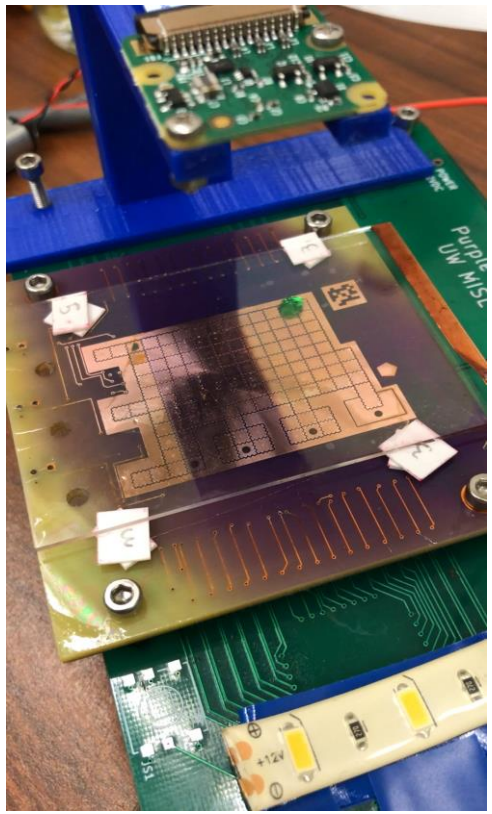
Takahashi et al., Nature Scientific Reports, 2019

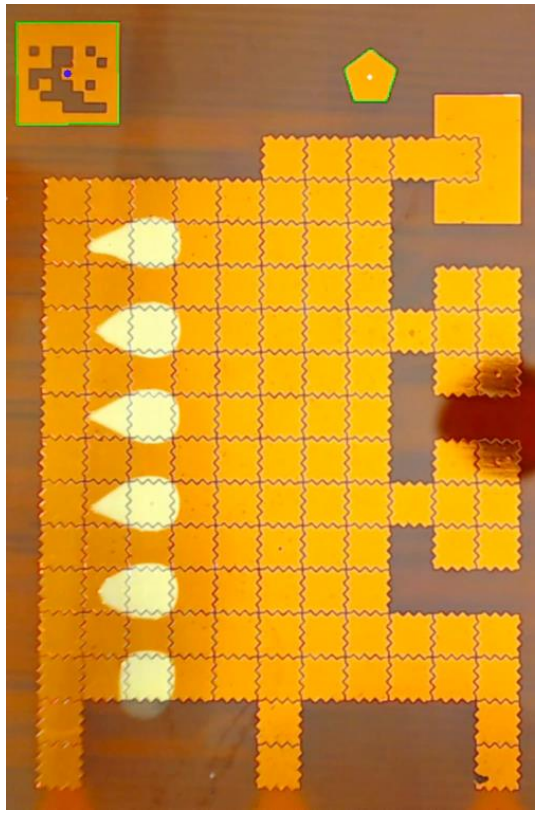
Digital microfluidics

Versatile platform to implement wet lab preparation protocols









Newman et al., Nature Communications, 2019; Willsey, ASPLOS, 2019

Affordable full-stack SW/HW digital microfluidics platform

High-level programming with *Puddle*

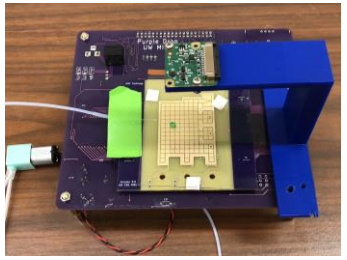
```
def thermocycle(droplet, temps_and_times):
    for temp, time in temps_and_times:
        heat(droplet, temp, time)
    if droplet.volume < MIN_VOLUME:
        droplet += input("water", min_volume)

def pcr(droplet, n_iter):
    thermocycle(droplet, n_iter * [
        (95,  3 * minutes),
        (62, 30 * seconds),
        (72, 20 * seconds),
    ])
```

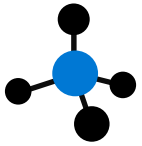
"Assembly code"

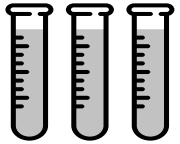
```
activate(3,0)
activate(3,1)
activate(3,2)
...
```

Hardware



Willsey et al., ASPLOS, 2019





Molecular Computing	Experiment	Medical Diagnostics
Synthetic DNA	Chemistry	Medicine

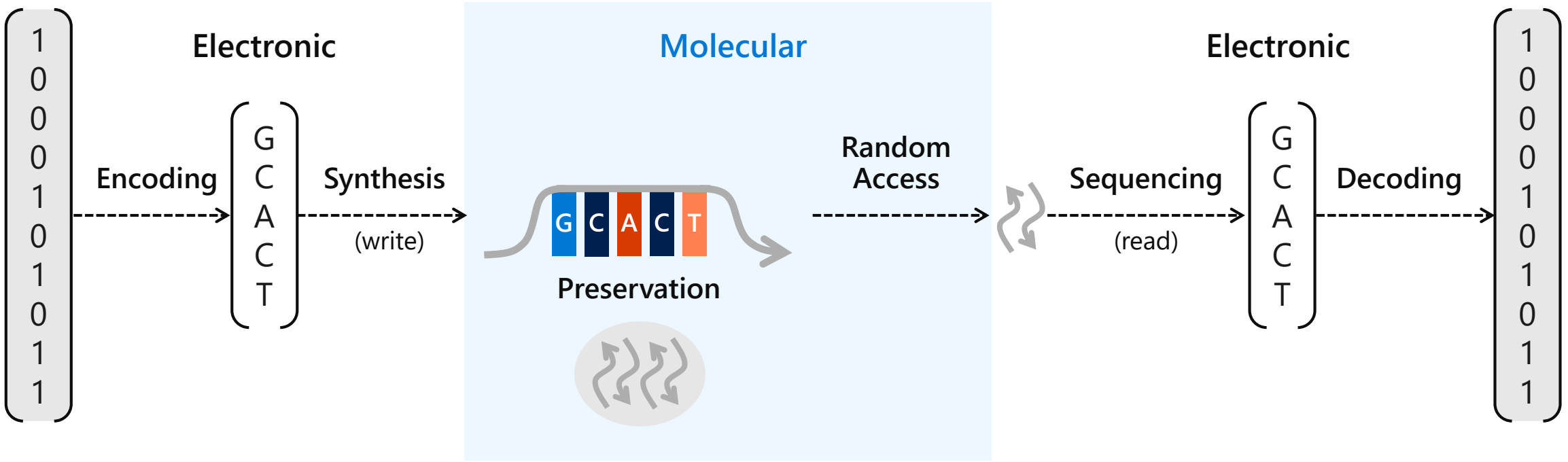
Domain Specific Reasoning

Core Fluidic Semantics

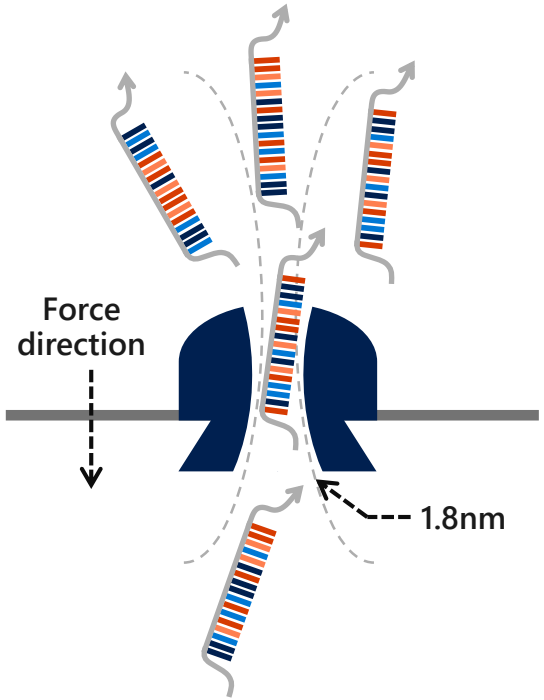
Hardware Abstraction

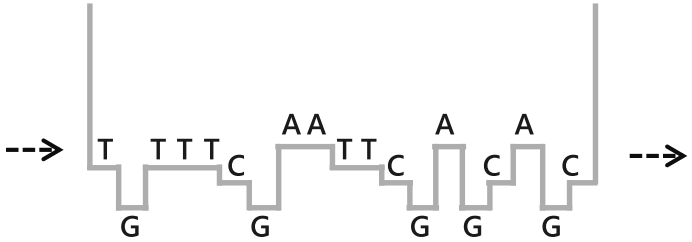
Microfluidic Chips

DNA storage end-to-end system



Reading DNA with nanopores





ACGT AC GTCTCTGA TGCA

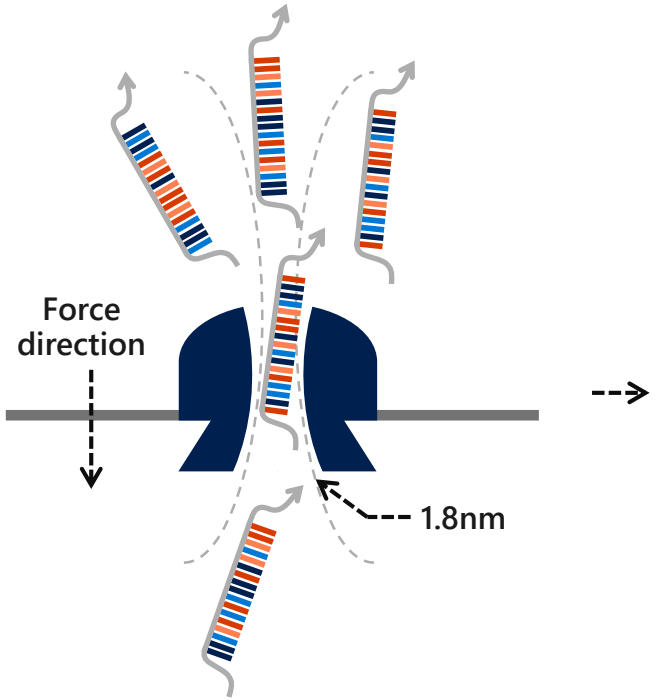
ACGT AC CTCTCAGA TGCA

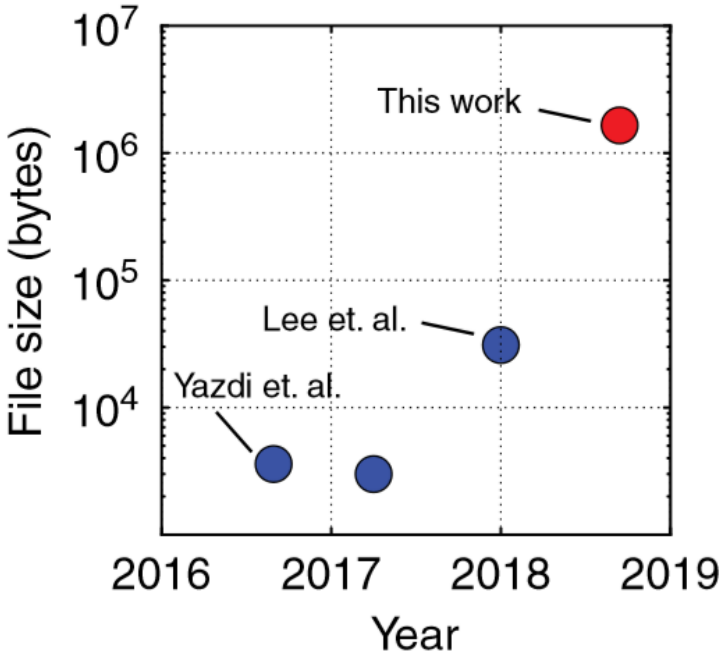
ACGT AC CTCTGTGA TGCA

ACGT AC GACTCTGA TGCA

Yazdi et al., Scientific Reports, 2017; Organick et al., Nature Biotechnology, 2018; Lopez et al., Nature Communications, 2019

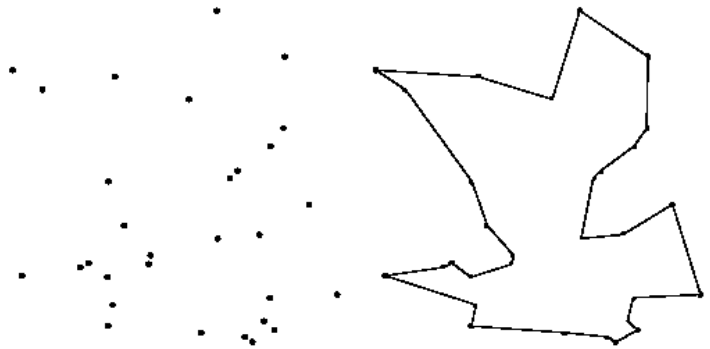
Reading DNA with nanopores



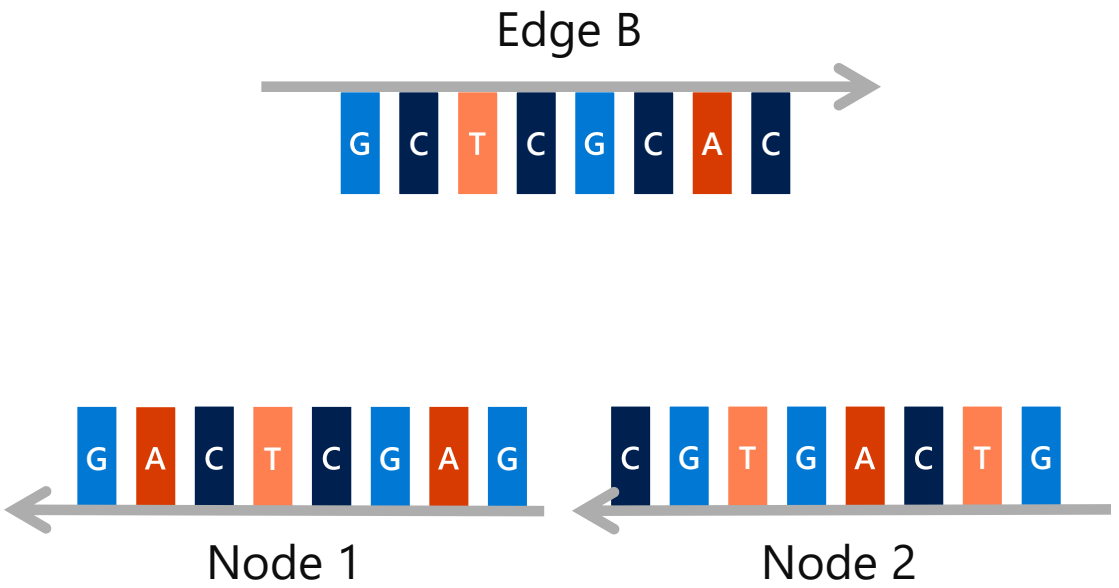


Yazdi et al., Scientific Reports, 2017; Organick et al., Nature Biotechnology, 2018; Lopez et al., Nature Communications, 2019

DNA computing in the 80s



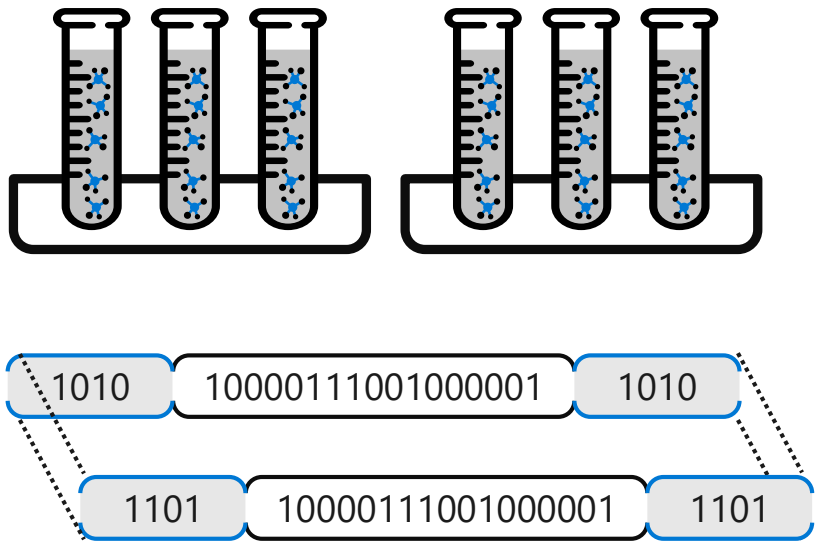
Hamiltonian path problem



Problem: shifts complexity from time to amount of material

Adleman, DNA1, 1994

DNA "computing" in the age of big data



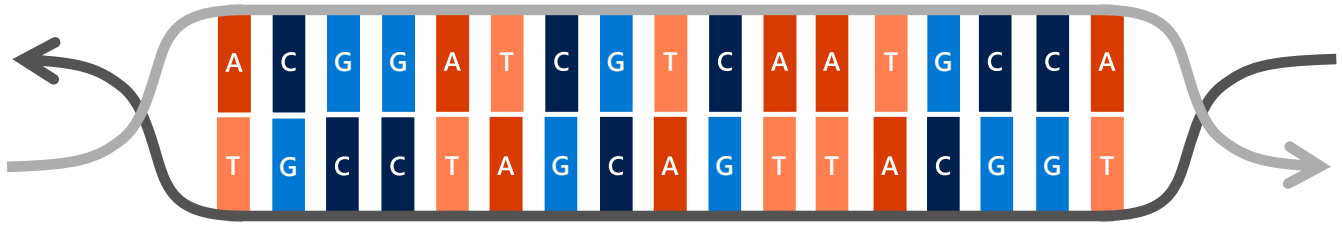
Operate over data already stored in DNA

Target polynomial time algorithms

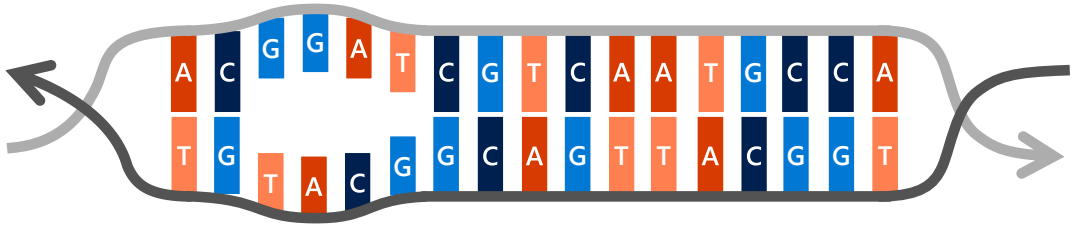
Extremely parallel and energy efficient

Exploiting matches for exact and approximate search

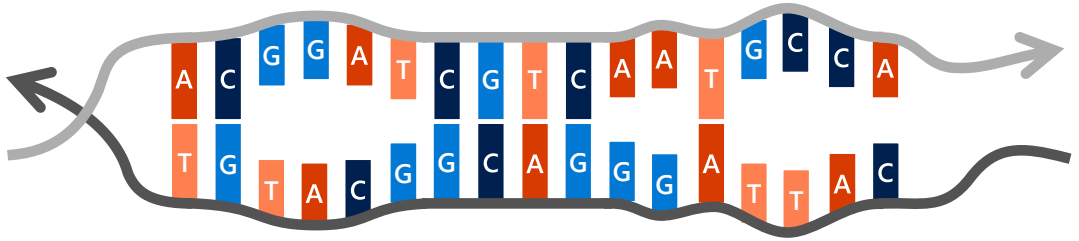
Double helix: complete match



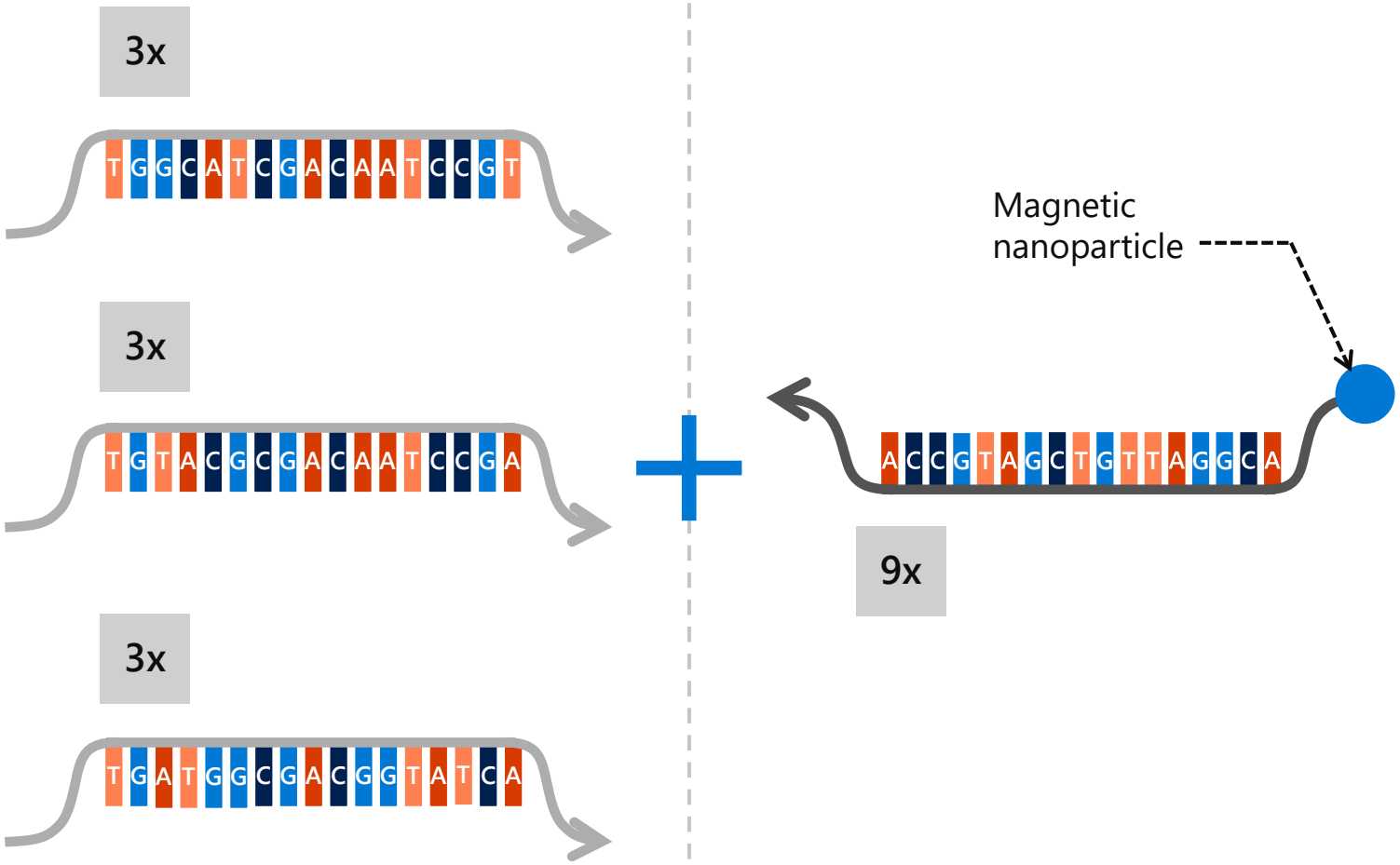
Good partial match



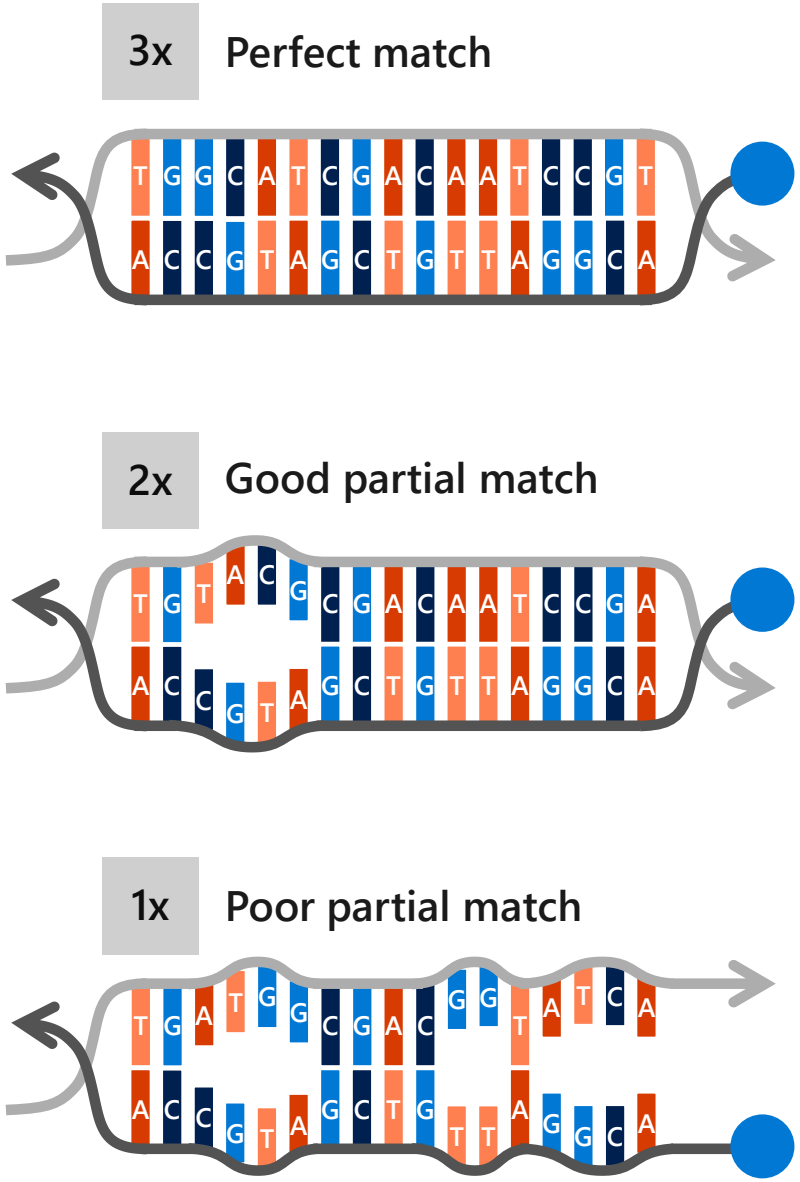
Poor partial match



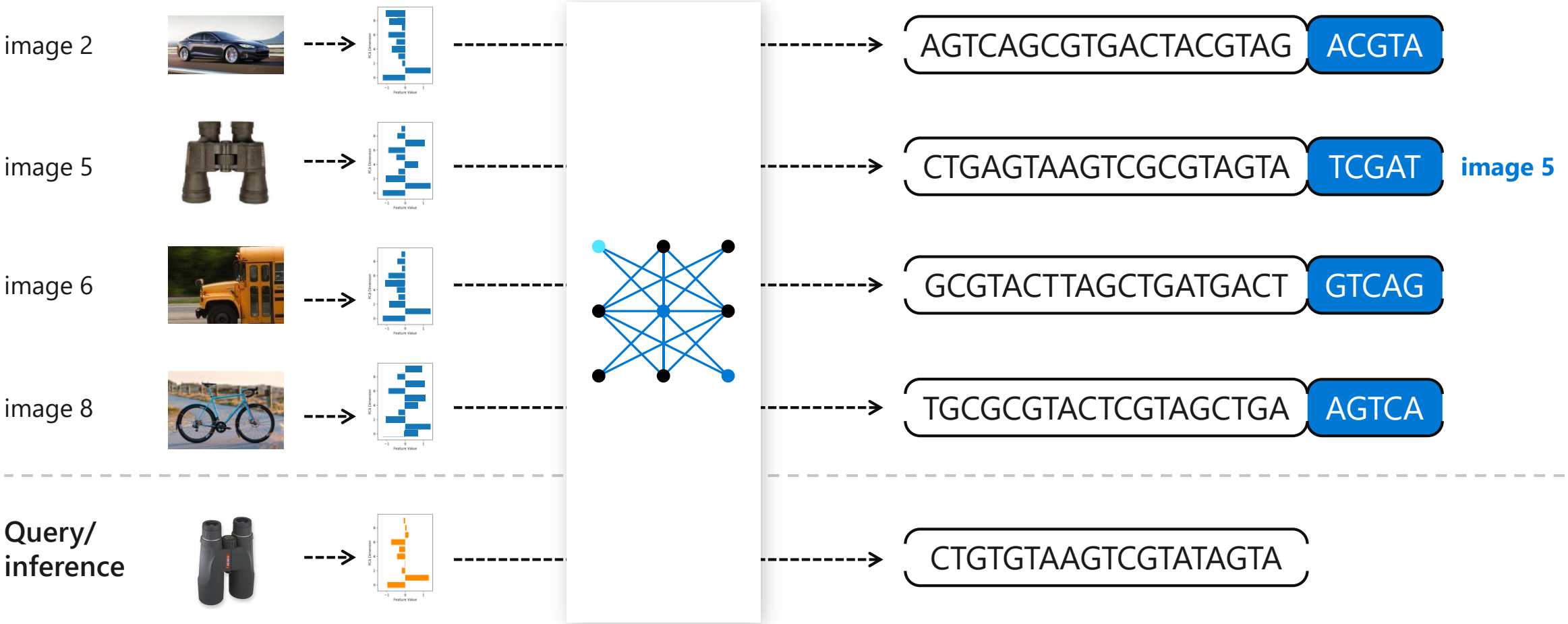
Searching with DNA



Match-dependent yield



Content based media search



Stewart et al., DNA24, 2018

Content based media search

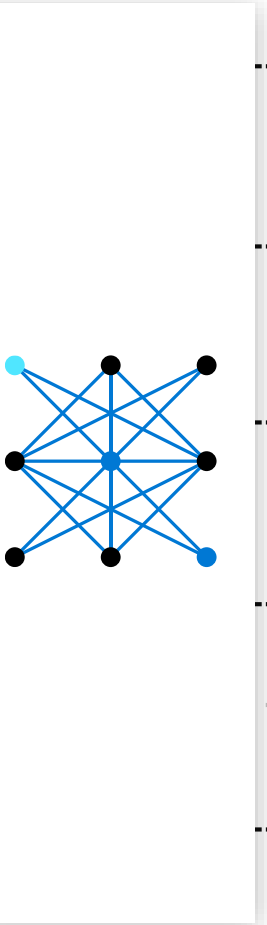
Database/ training

image 2

image 5

image 6

image 8



AGTCAGCGTGACTACGTAG ACGTA

CTGAGTAAGTCGCGTAGTA TCGAT image 5

GCGTACTTAGCTGATGACT GTCAG

TGCGCGTACTCGTAGCTGA AGTCA

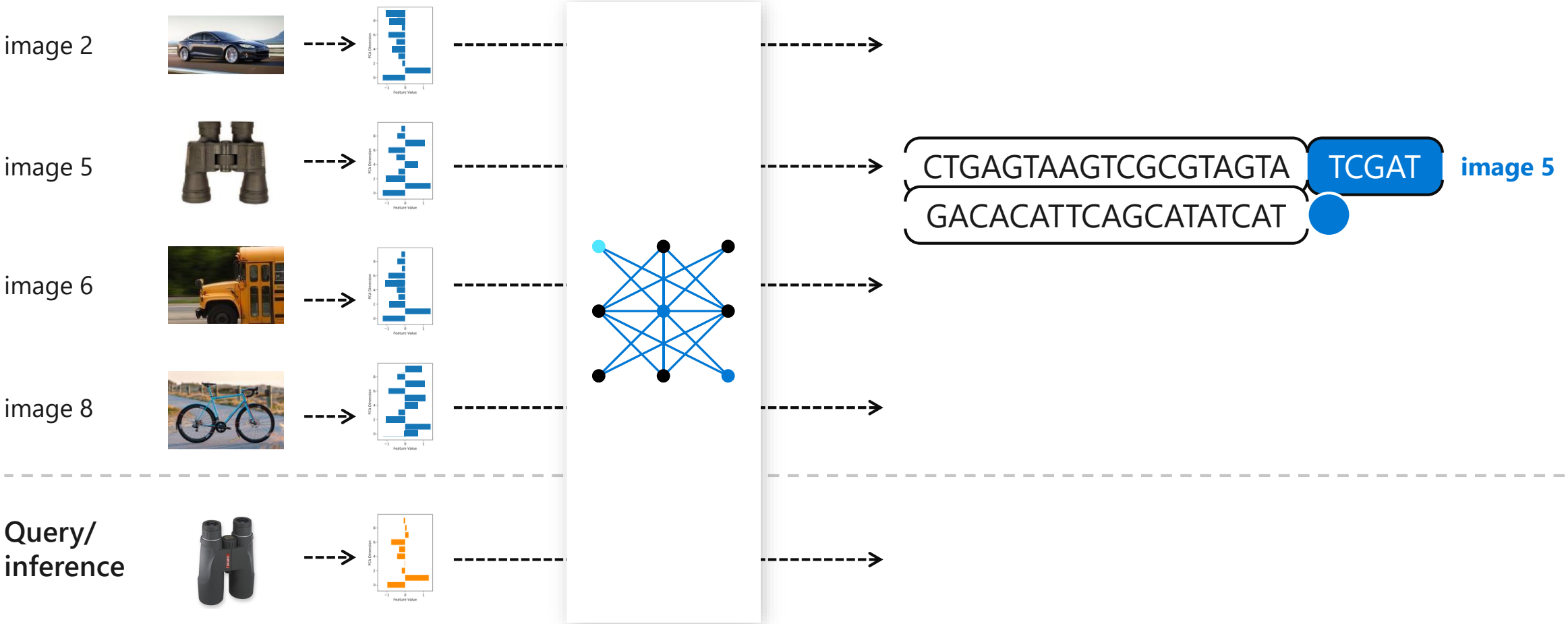
Query/ inference

CTGTGTAAGTCGTATAGTA

GACACATTCAGCATATCAT

Stewart et al., DNA24, 2018

Content based media search



Stewart et al., DNA24, 2018

Microsoft MISL

Questions?