STORAGE DEVELOPER CONFERENCE
SDC 22
Fremont, CA
September 12-15, 2022
BY Developers FOR Developers

A SNIA Event

Coding and Decoding, an Experience From a Brazilian Research Center

Joao Lucas Maehara Said dos Reis
Marília Santos Menossi Mortari

Lenovo | ipt

Who are we?

The Institute for Technological Research (IPT) has been contributing actively during 123 years.

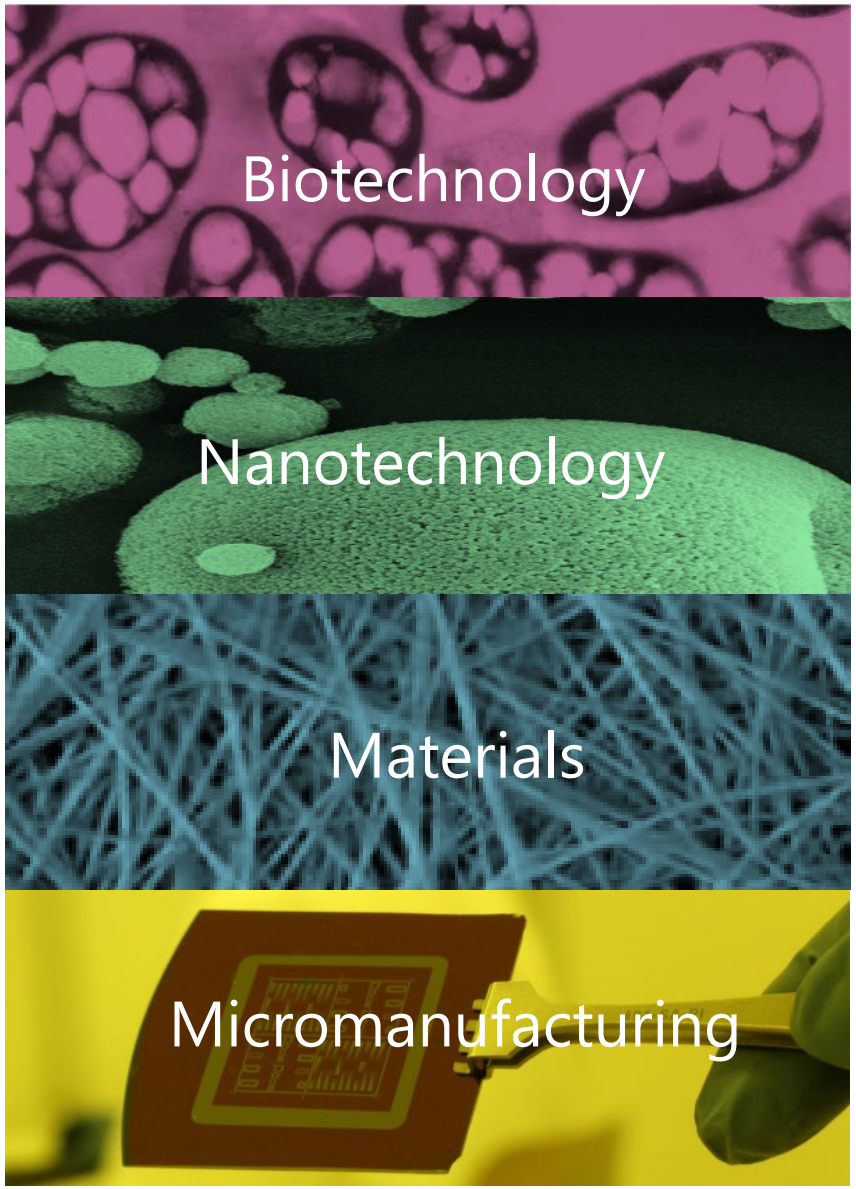
We provide technical solutions for industry, governments and society, enabling them to overcome the challenges of our time.

Lenovo | ipt

STORAGE DEVELOPER CONFERENCE SDC 22

Binanomanufacturing Center

Multidisciplinary team, with doctors, masters, and specialized technicians



Lenovo

And through this partnership, we are developing a DNA Data Storage project together

PROMETHEUS

Why DNA data storage?

Why DNA Data Storage

- DNA can be stable for thousands of years

- Potential to store Petabytes of data into 1 gram of DNA

Article | Open Access | Published: 26 May 2022

Bioarchaeological and palaeogenomic portrait of two Pompeians that died during the eruption of Vesuvius in 79 AD

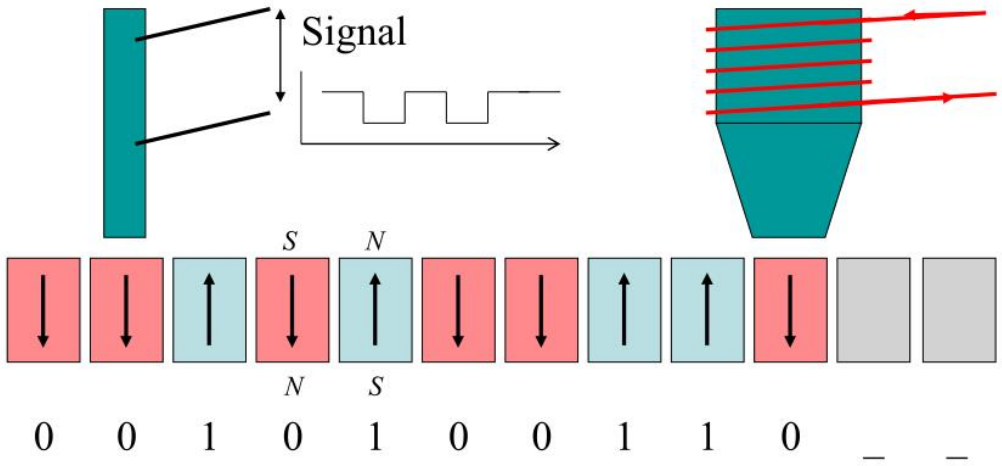
Gabriele Scorrano, Serena Viva, Thomaz Pinotti, Pier Francesco Fabbri, Olga Rickards & Fabio Macciardi

Scientific Reports **12**, Article number: 6468 (2022) | Cite this article

- DNA reading will never be outdated and will always be improved

How data is stored in DNA

LTO Tape



Binary code

0 

1 

Quaternary code

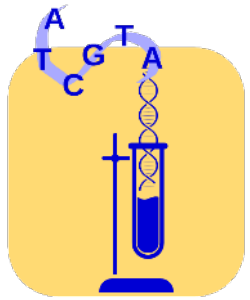
0 T
1 A
2 C
3 G

DNA data storage pipeline

1. Converting information to binary code

2. Converting binary code to DNA code

3. DNA Synthesis



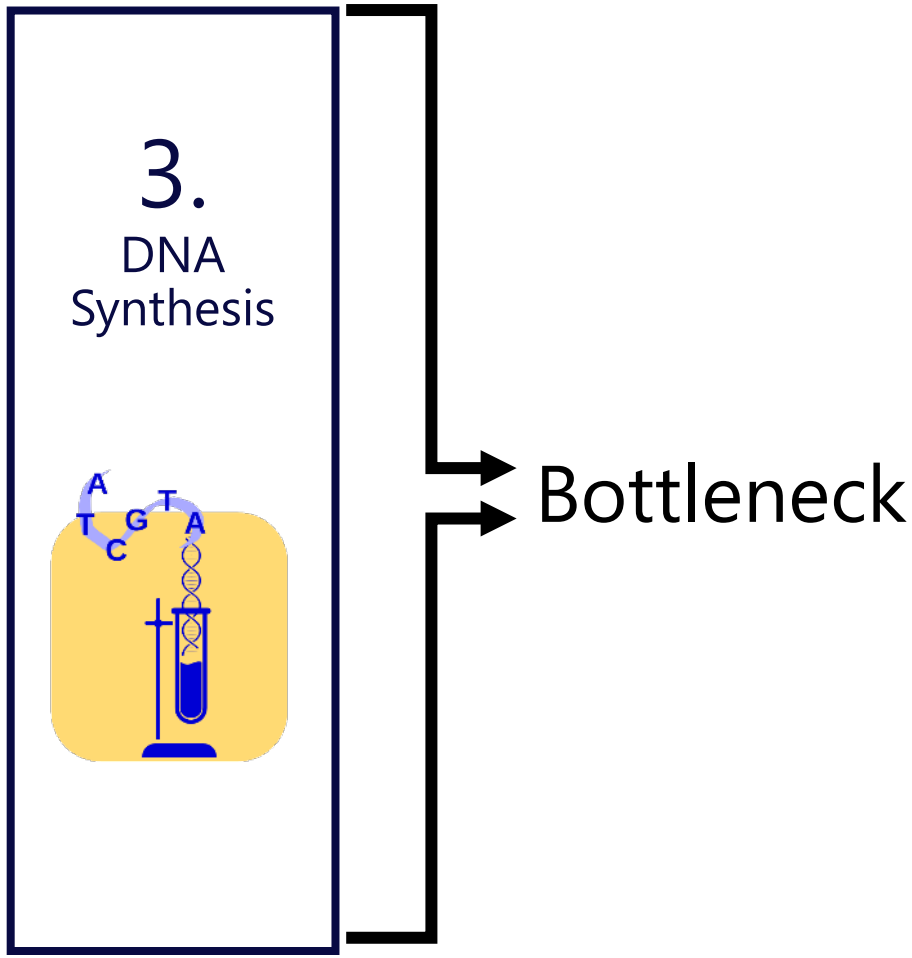
4. Storage

5. Recovery

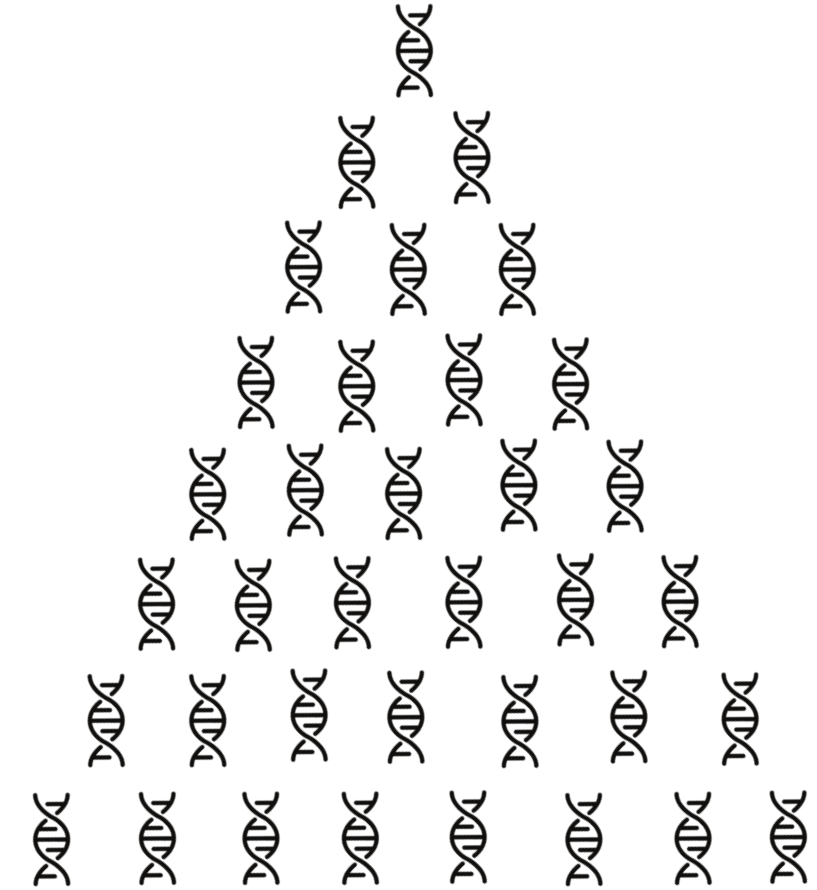
6. Sequencing

7. Decoding

DNA data storage pipeline

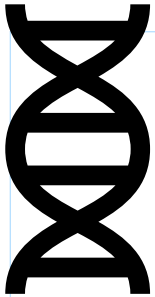


Bottleneck



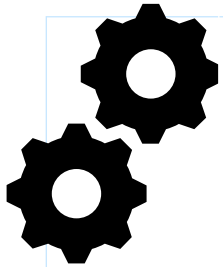
Billions of unique sequences
to store Petabytes of data!

Scope

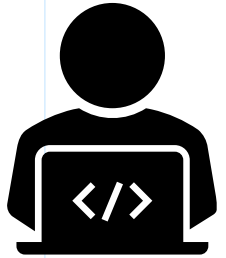


Enzymatic synthesis

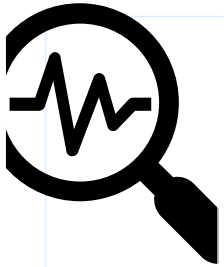
Chemical synthesis



Process miniaturization



Coding and decoding



Sequencing methods

Chemical DNA synthesis

Advantages:

- Standardized Technique -> Miniaturize the process;
- Fine synthesis control ;

Disadvantages:

- Hazardous Chemicals;

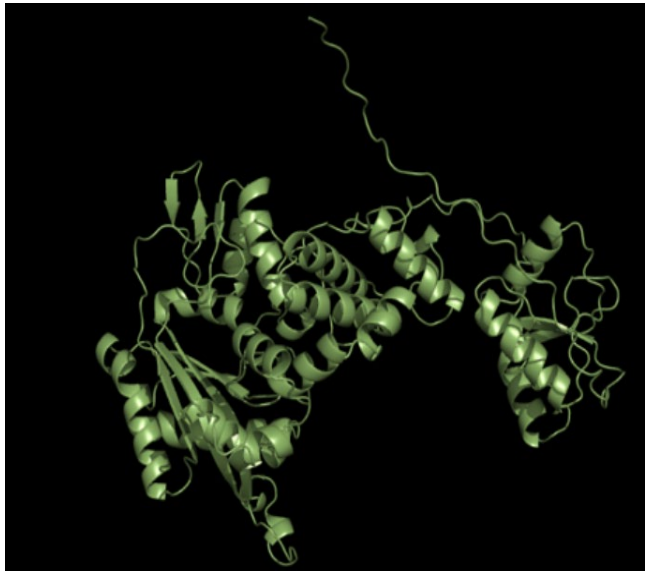
Enzymatic DNA synthesis

Advantages:

- Novel Technique -> Requires ;
- Encironmentally friendly process;

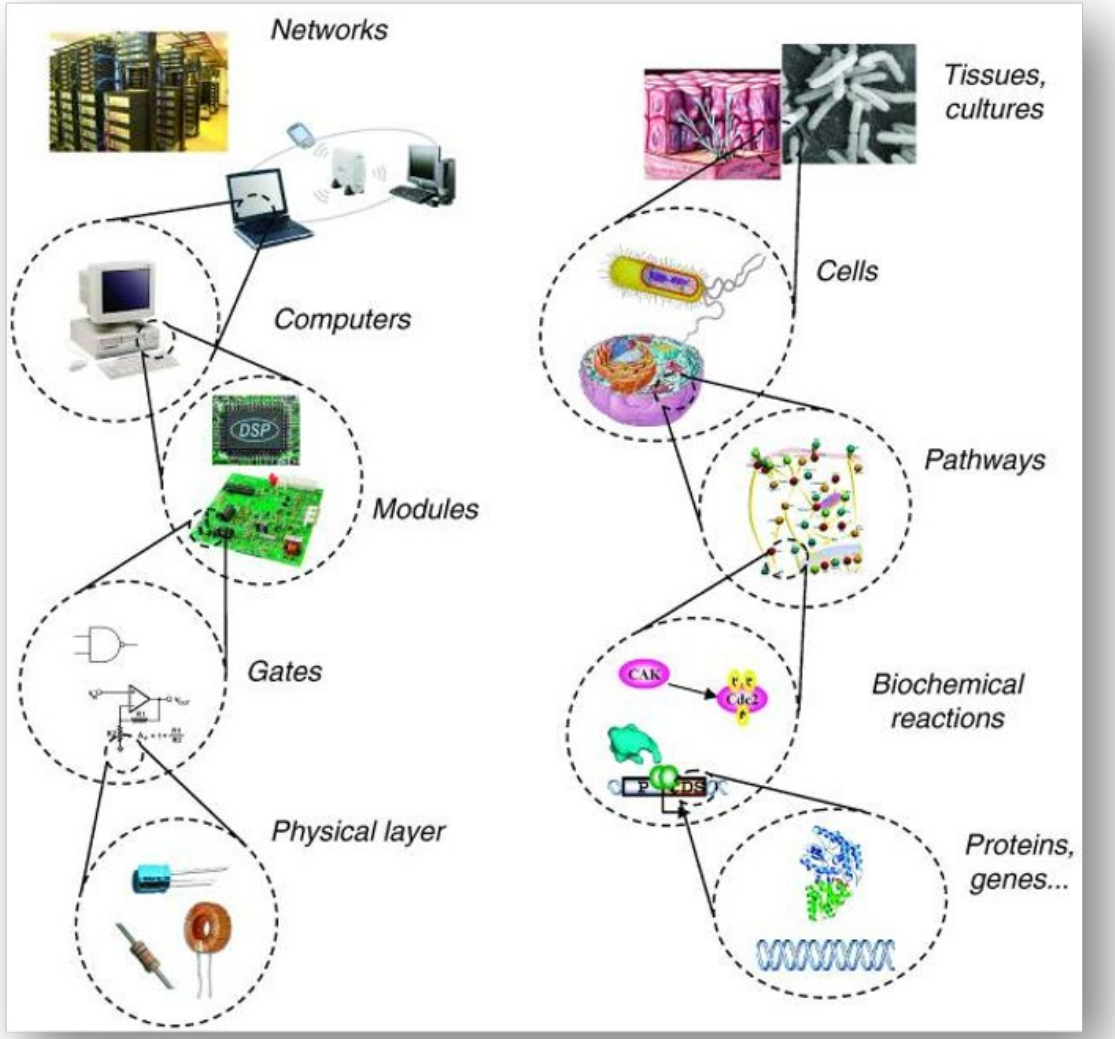
Disadvantages:

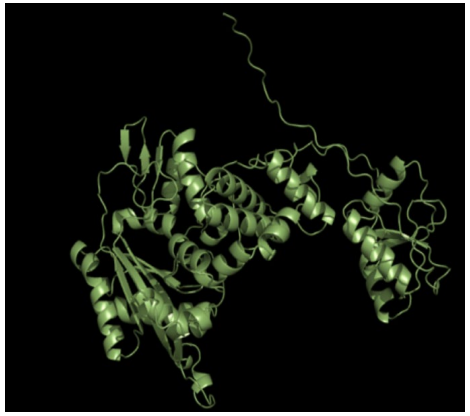
- Synthesis control is not simple;



TdT enzyme has been adopted for DNA Data Storage Applications

Biological Systems to manufacture DNA





Enzymes act like the head of a hard disk

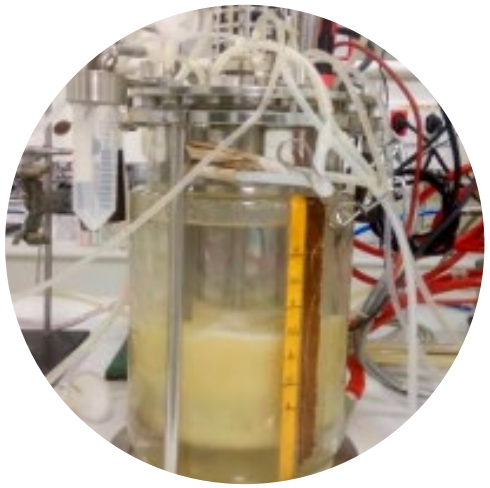
DNA is the physical medium for information being recorded

ANDRIANANTOANDRO, Ernesto et al. Synthetic biology: new engineering rules for an emerging discipline. Molecular systems biology, v. 2, n. 1, p. 2006.0028, 2006.

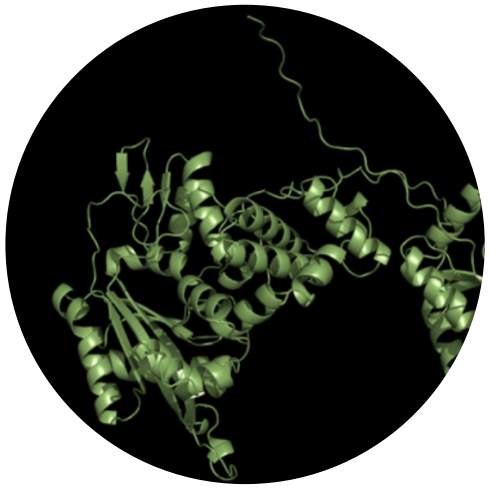
Development of new Enzymes

Mass production of TdT enzyme is important to reduce costs of DNA synthesis .

Programmed Bacteria



Large Scale Production



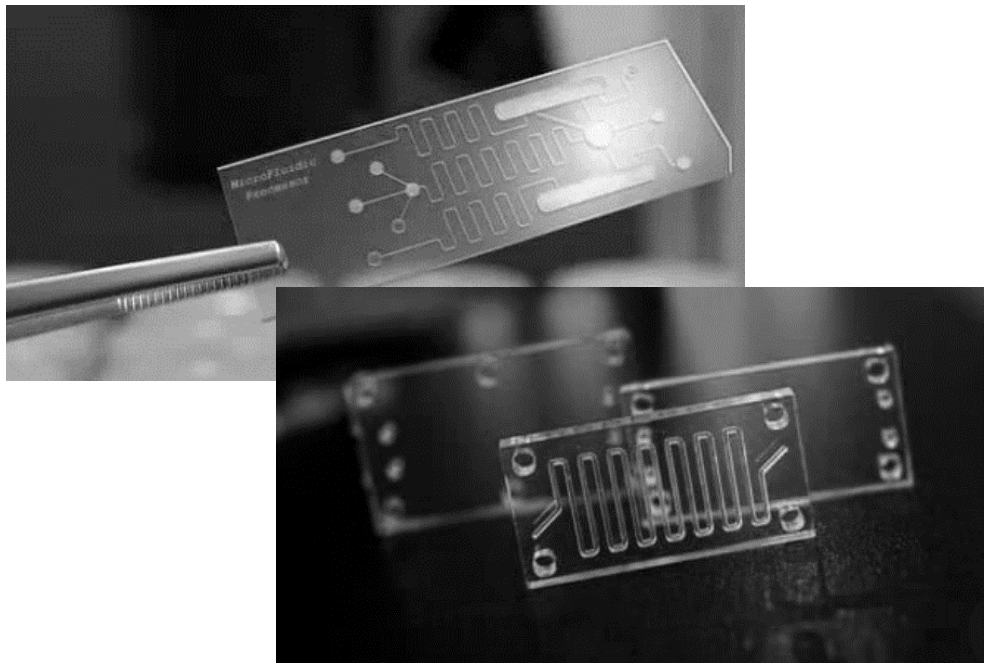
Enzyme Purification

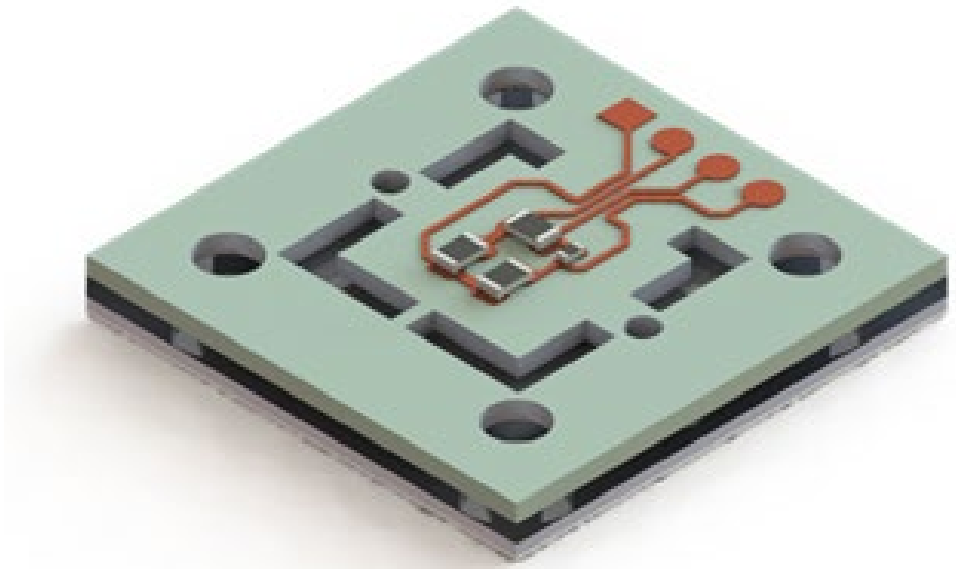
Application in DNA Synthesis

Developing microdevices

We have been developing a microdevice and working to improve the synthesis process.

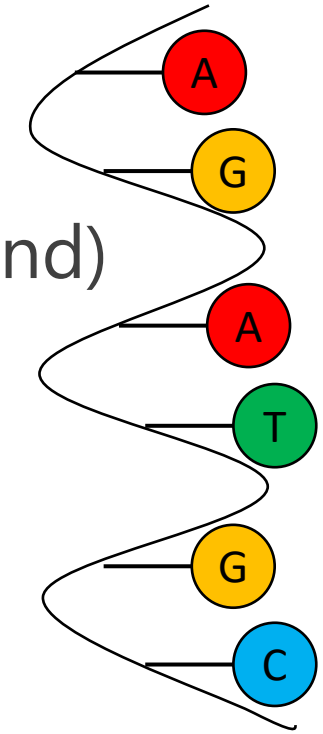
- Parallelization of DNA Synthesis
- Working with small volumes





DNA Chemical synthesis

DNA
(single strand)

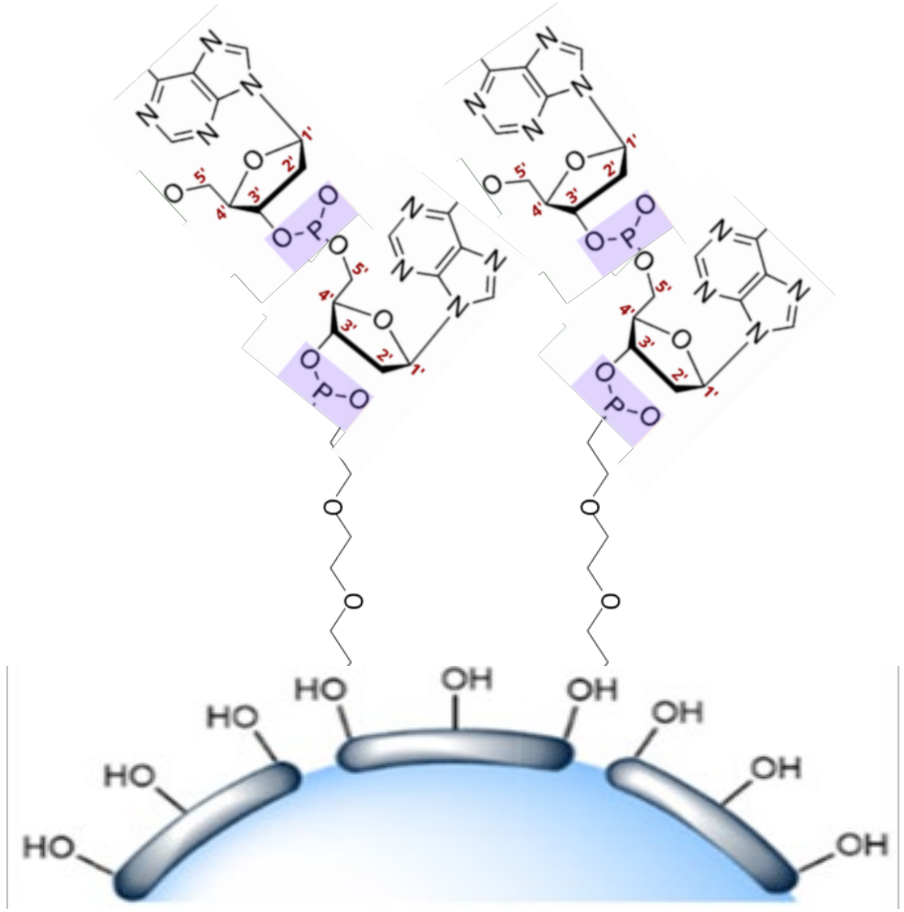


Phosphoramidite

Adenine
Thymine
Cytosine
Guanine

linker

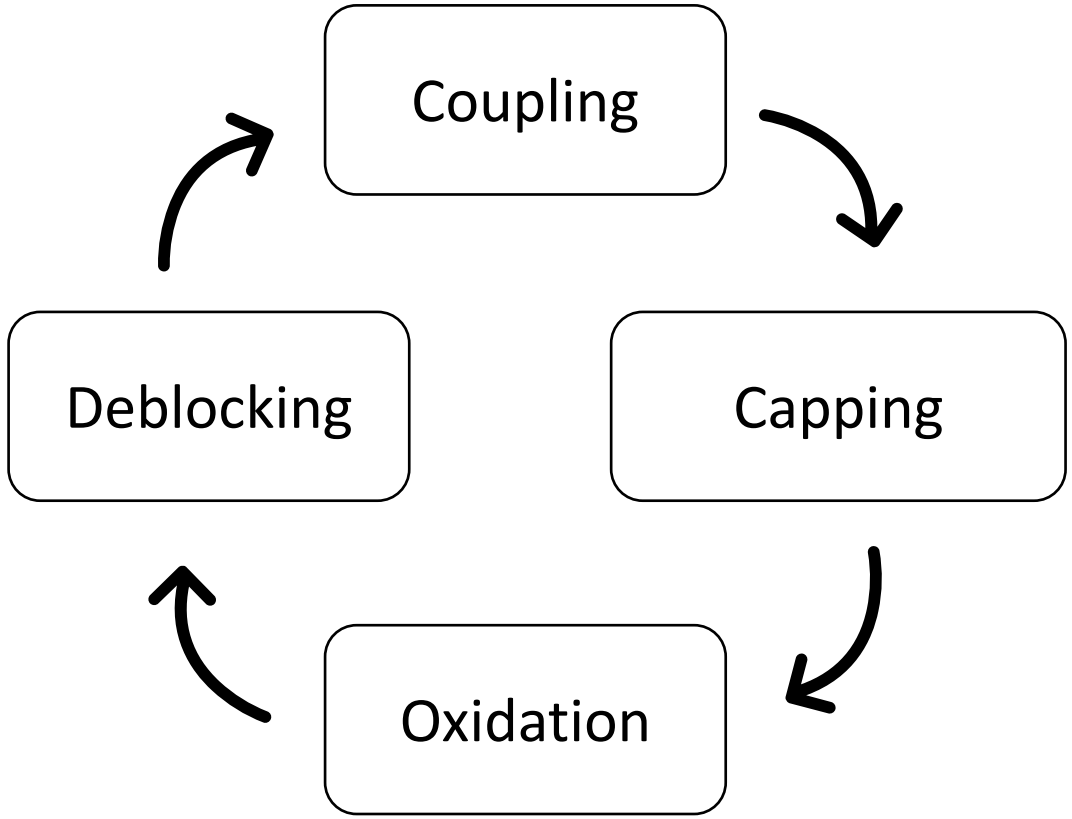
surface



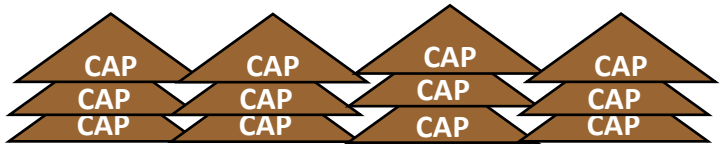
https://www.twistbioscience.com/blog/science/simple-guide-phosphoramidite-chemistry-and-how-it-fits-twist-biosciences-commercial

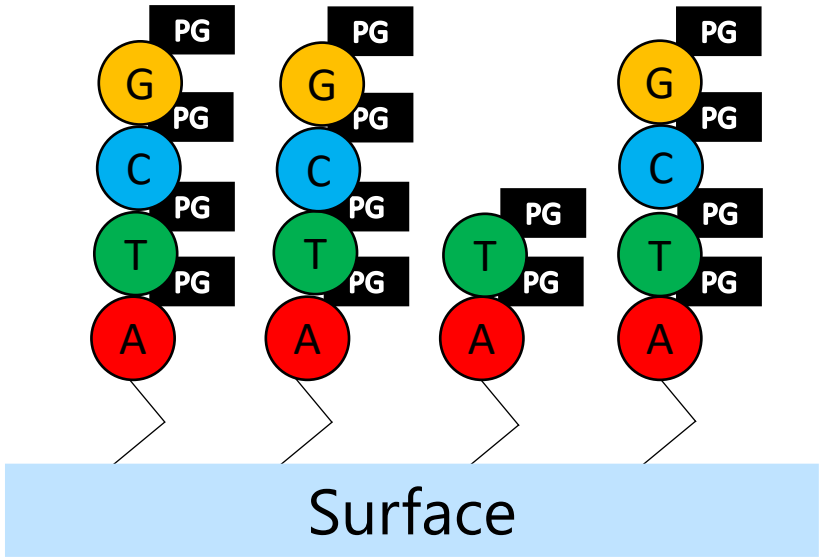
Chemical synthesis

A T C G



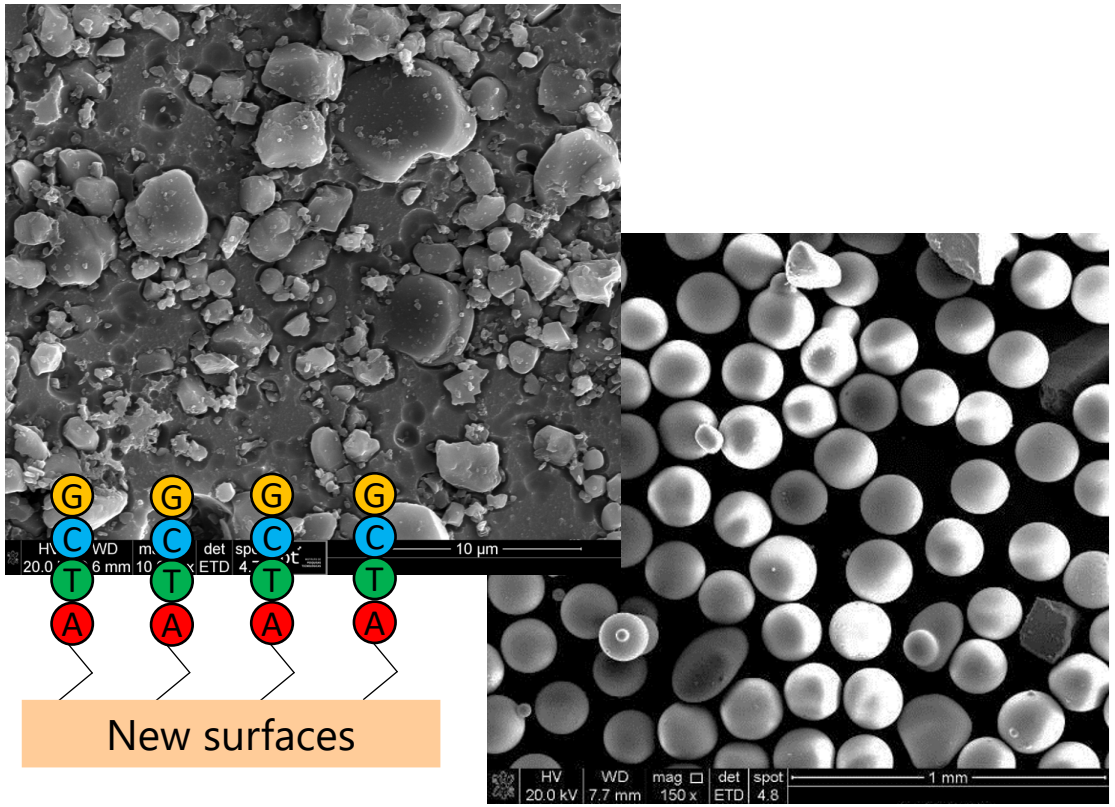
PG = protector group





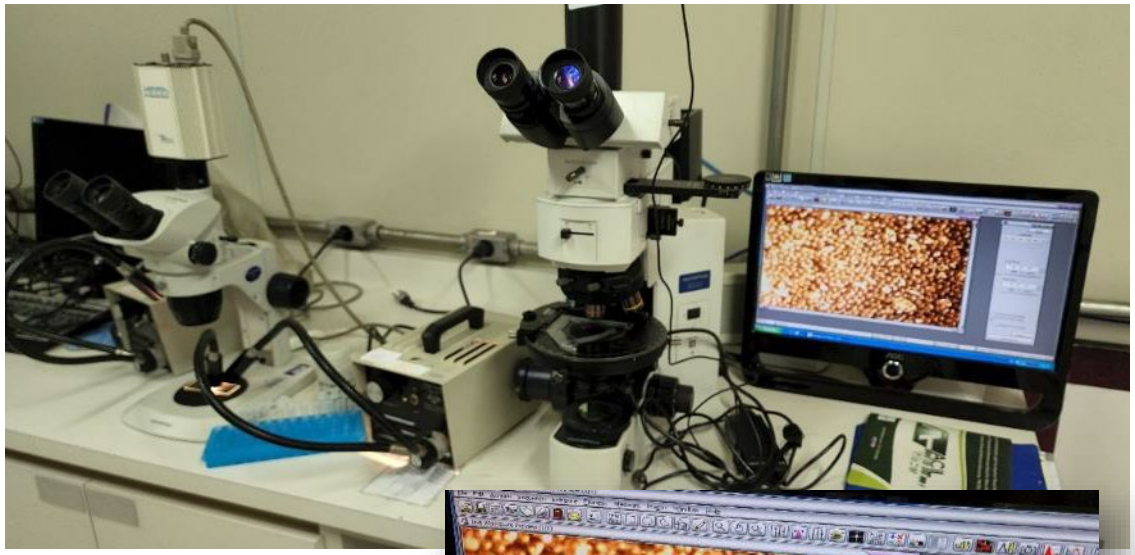
DNA synthesis

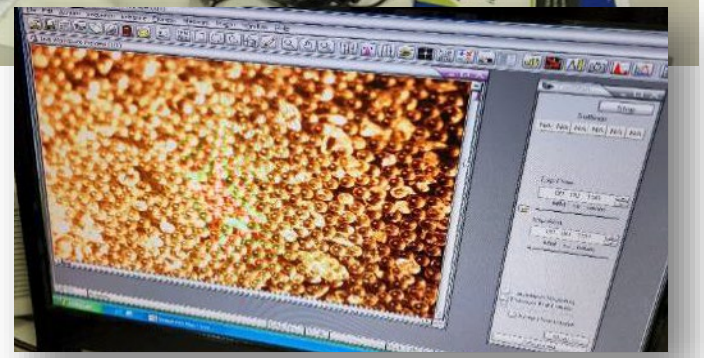
- On different surfaces



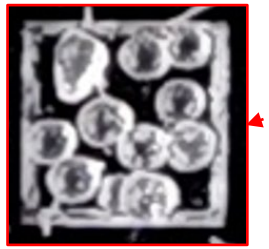
- In microreactors

New particles for DNA synthesis

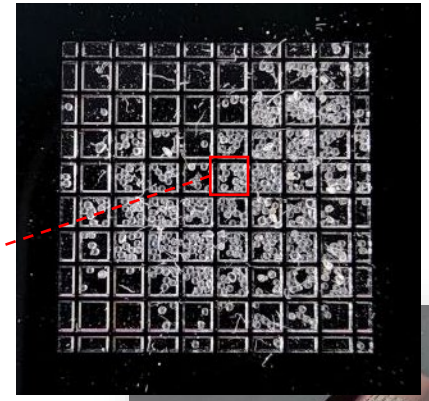


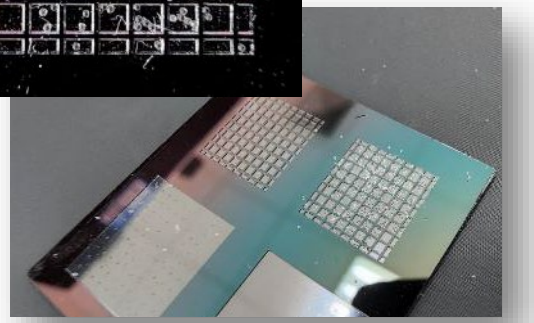


Macro appearance
Reference: SD Card

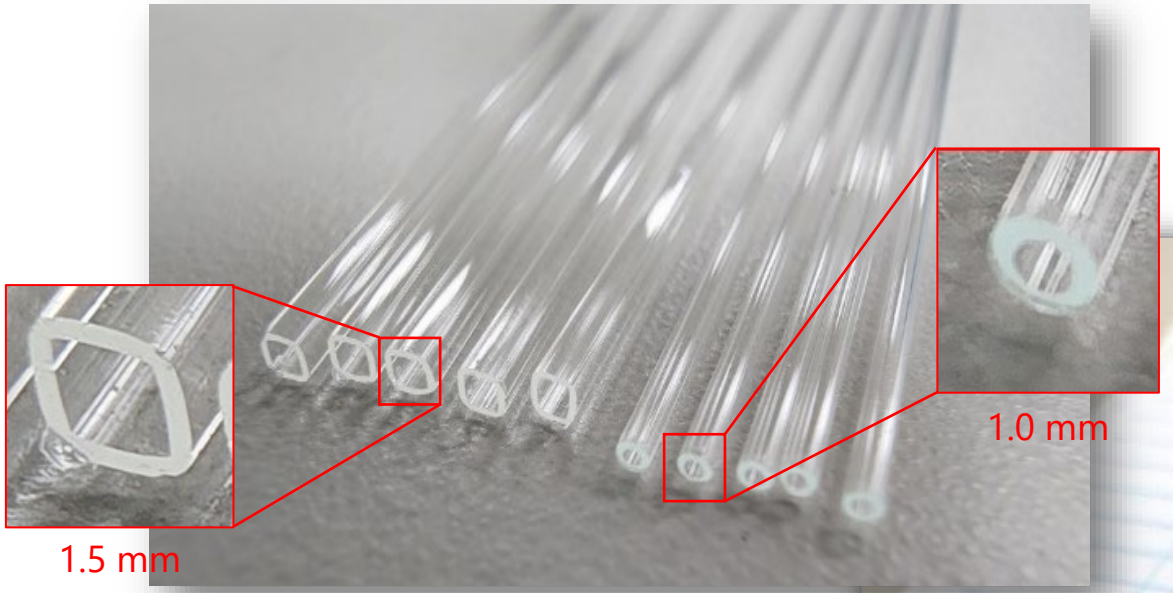


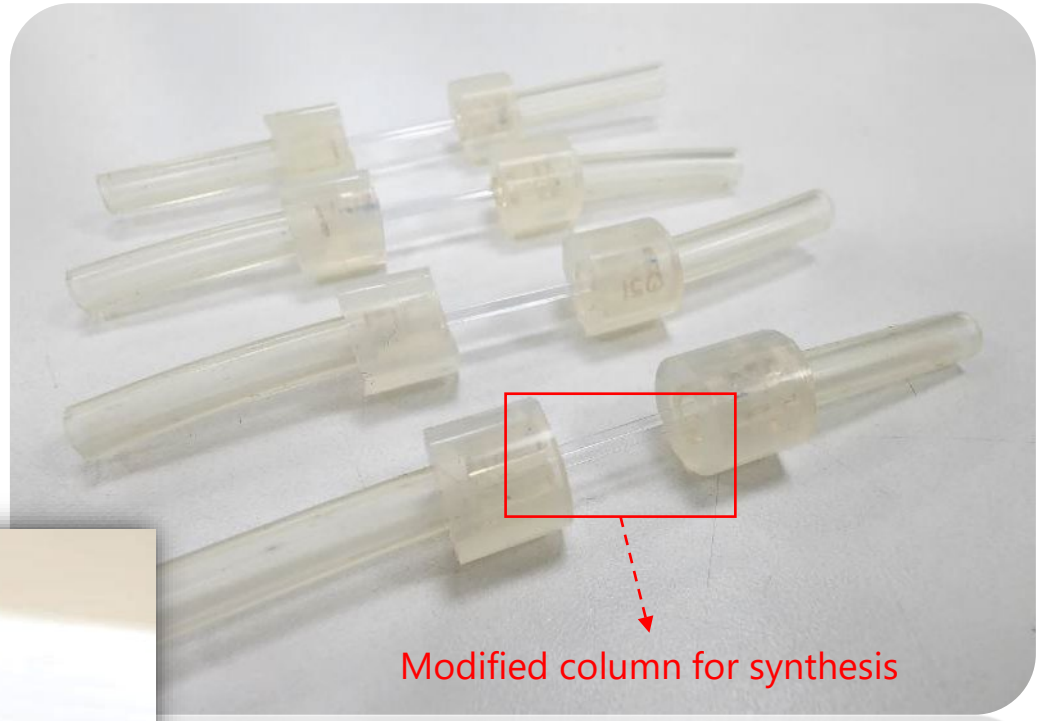
Square size
≈ 1 mm



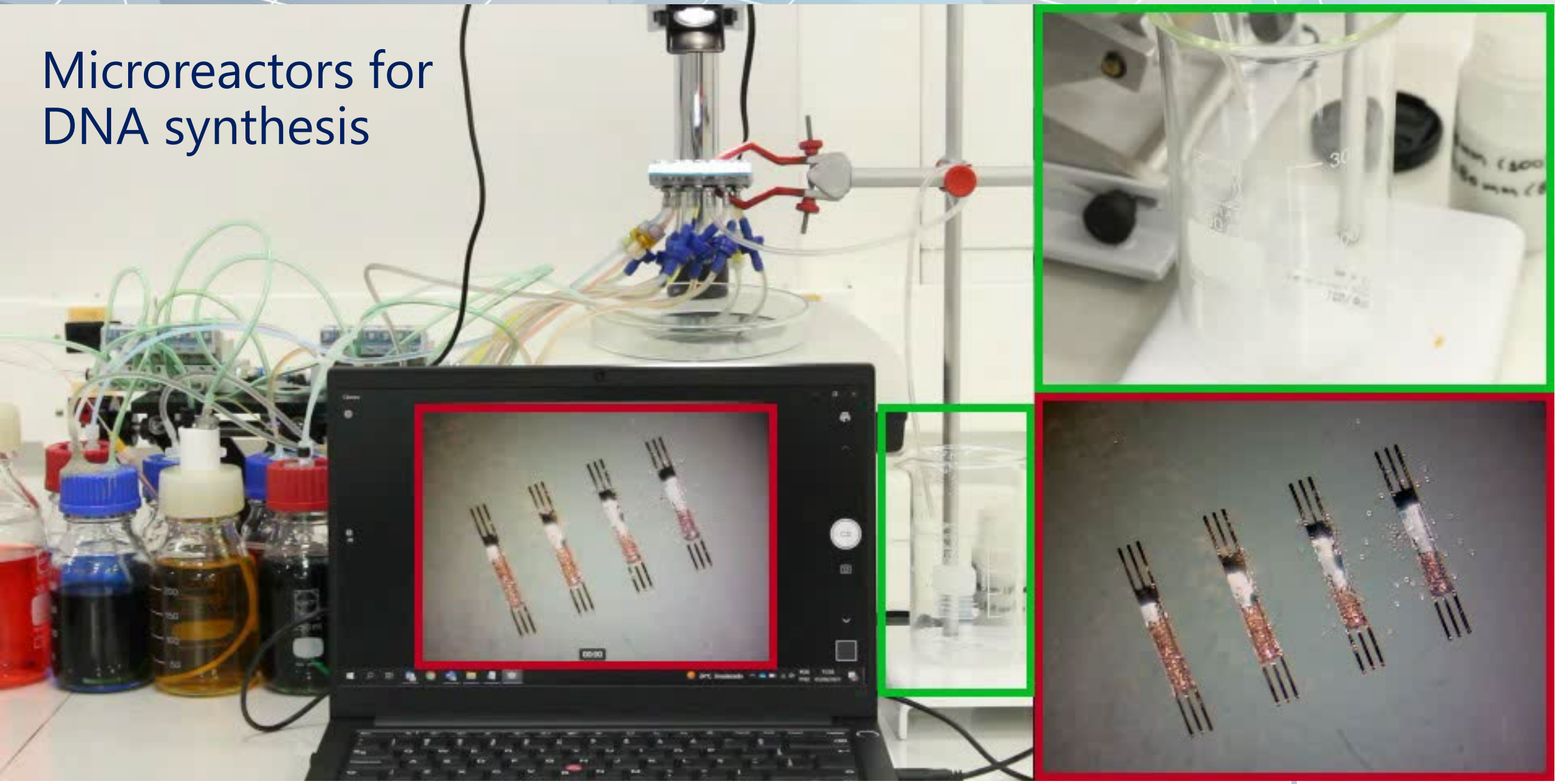


Columns for DNA synthesis

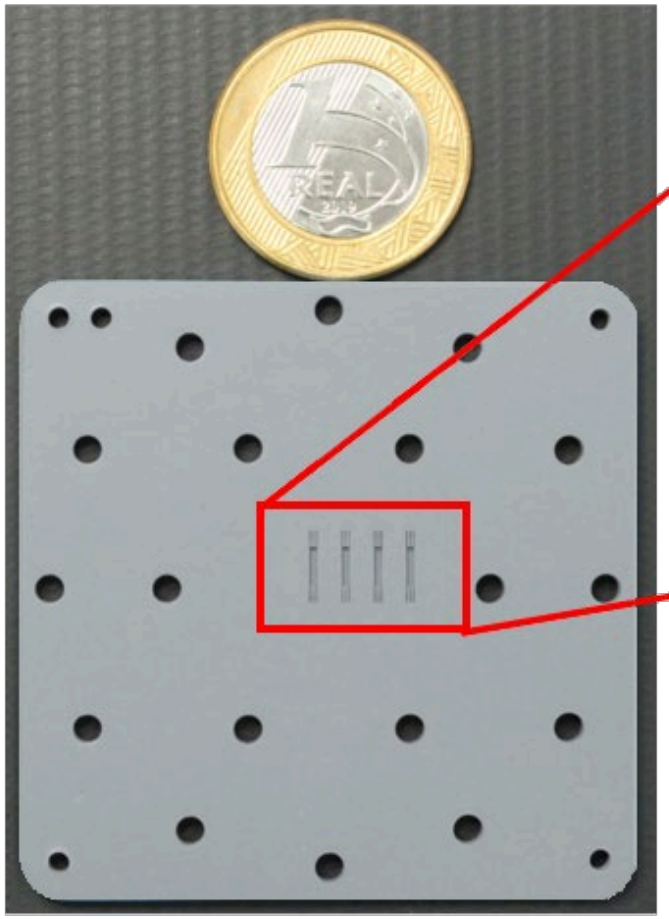


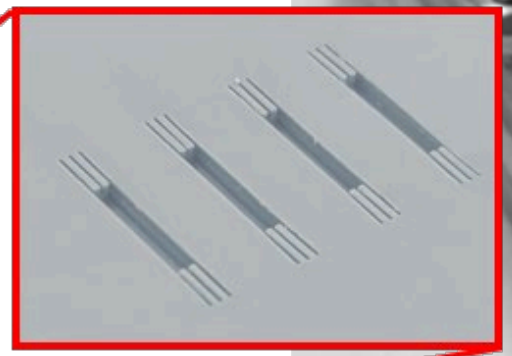


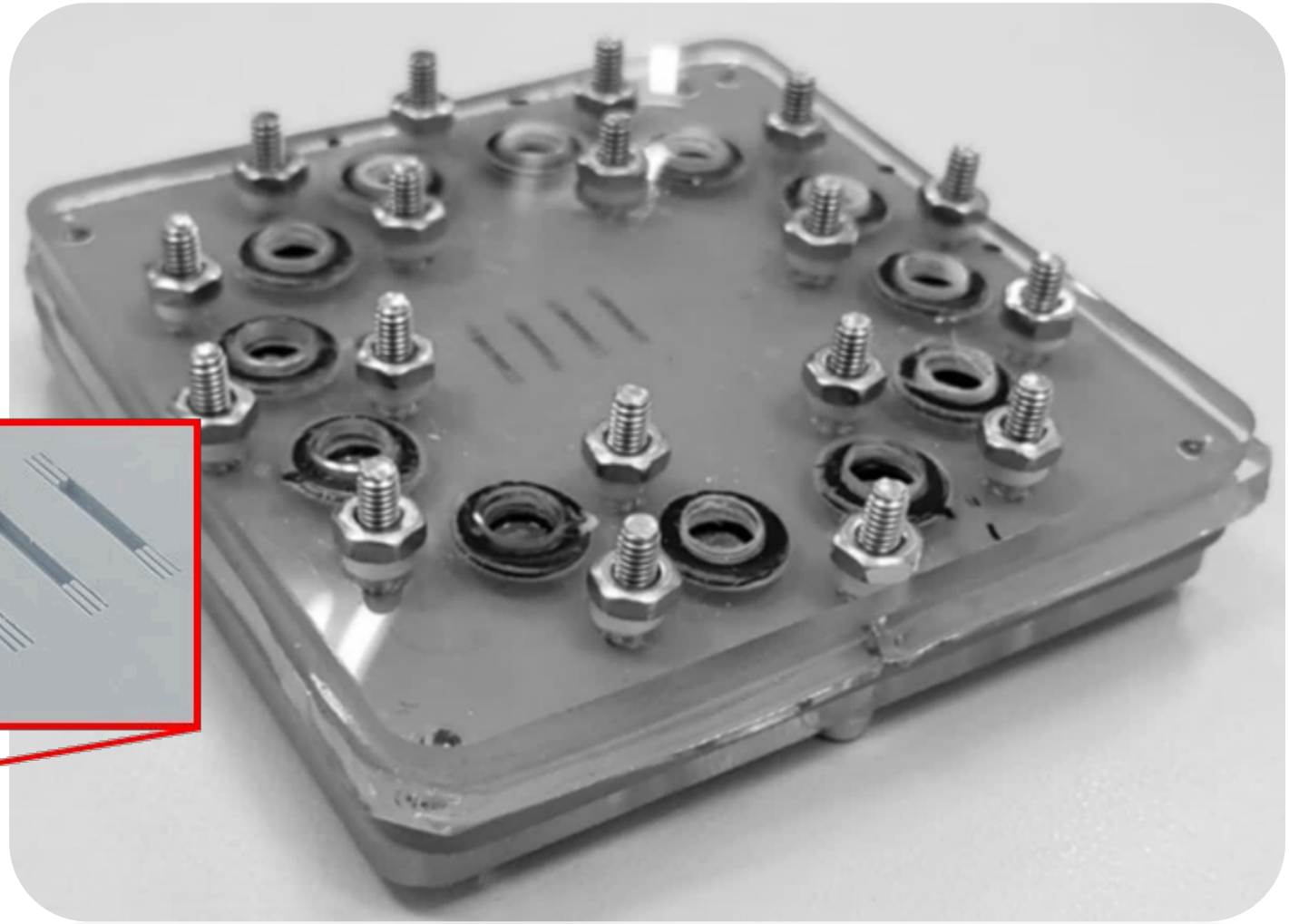
Microreactors for DNA synthesis



Microreactors for DNA synthesis





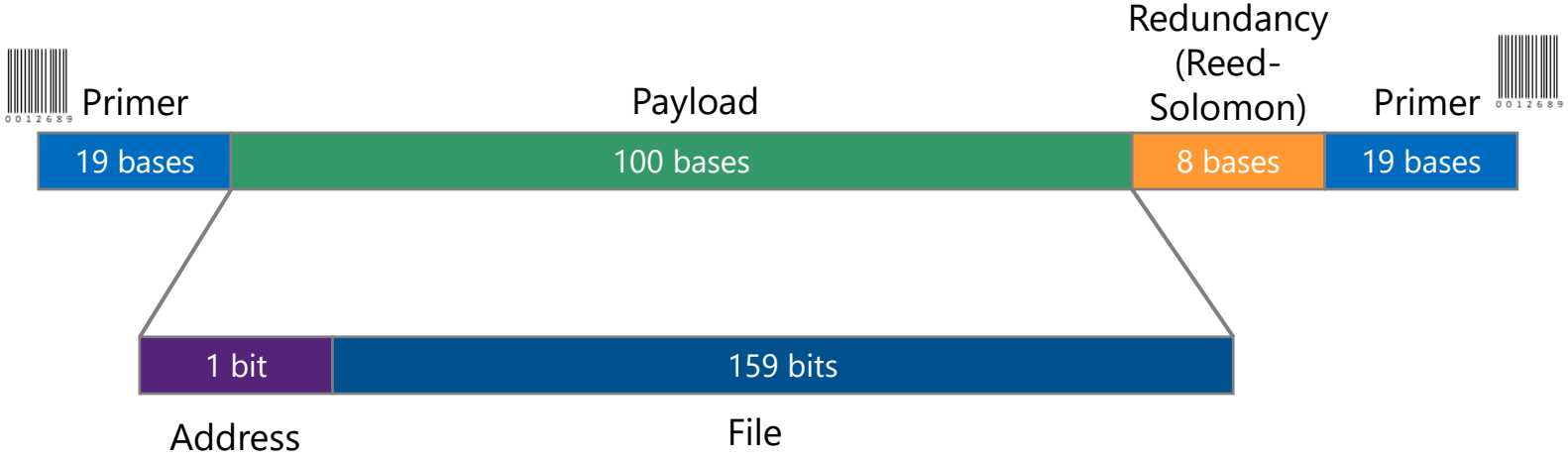


A little part of our results so far

Encoded file: 38byte .txt file containing

IPT and Lenovo, a successful partnership!

2 DNA sequences with 146 bases each



Bases = nitrogeneous bases = A T C G

DNA sequences

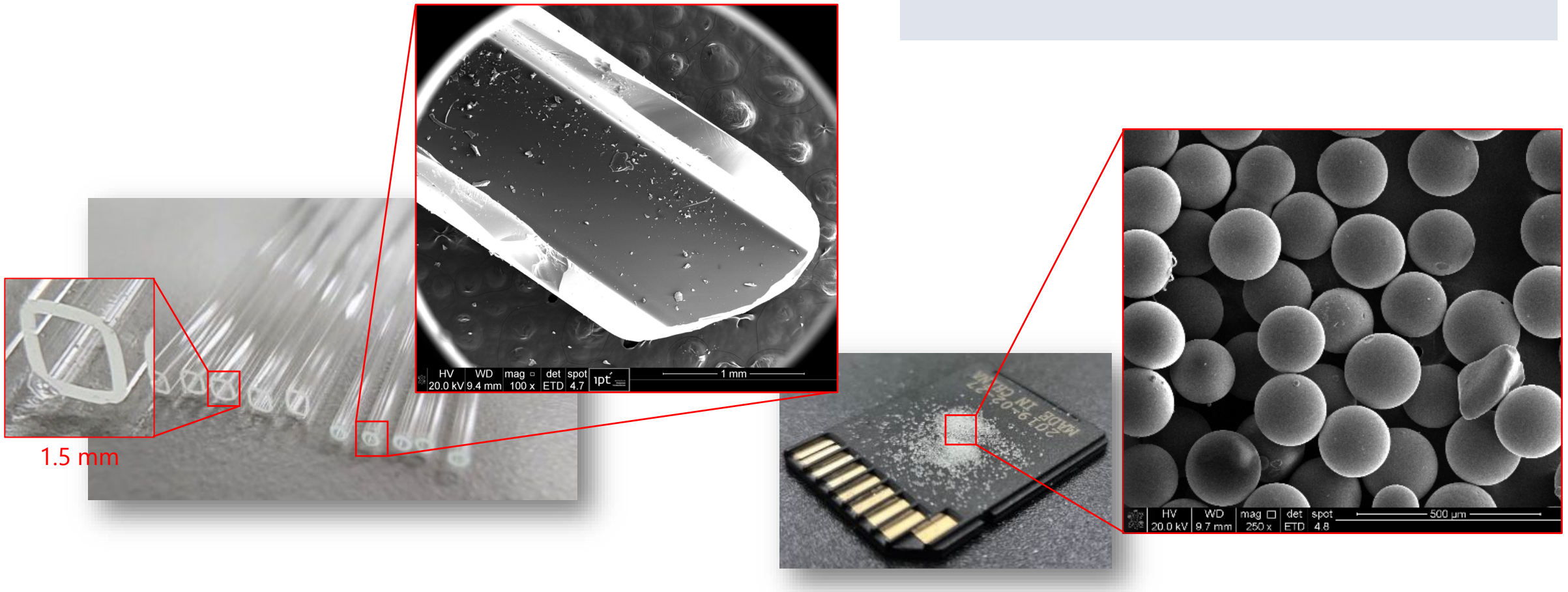
Forward Primer

5'CACGACGTTGTAAAACGACAGACAGGAGAAGCGTACTATATAAGGCCACAGACGATAA
GGTGCTATCCGGTAGCATGCTGCACGACTATATCGTGTACGGTCACGCTATATCGCATCACG
GGACGCCGGGTCATAGCTGTTTCCTG

5'CACGACGTTGTAAAACGACTCGTGGCAGATCAGTCCATAGCCGTCCAGACAAGAACAGT
ACGGCCAAGAACATATCGTCCCAGATCCGACCATATCCTCTCTGATACGCCATATAACACAG
AGGGATGGGTCATAGCTGTTTCCTG

Reverse Primer

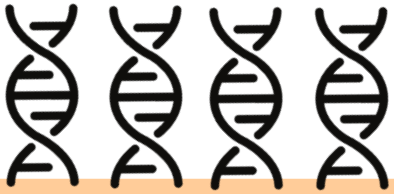
Chemical DNA synthesis on different surfaces



Lenovo | ipt

STORAGE DEVELOPER CONFERENCE SDC 22

Two options for storage of the synthesized DNA



Surface of the particles or columns

DNA storage

Surface storage with DNA

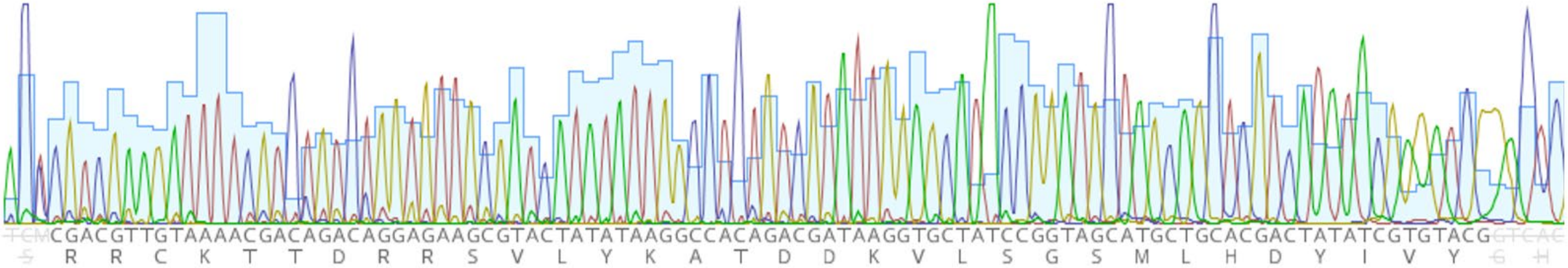
DNA storage

Surface storage with DNA

Creating copies by PCR methodology

PCR = Polymerase Chain Reactions

Sequencing



5'CACGACGTTGTAAAACGACAGACAGGAGAAGCGTACTATATAAGGCCACAGACGATAAGGTGCTATCCGGT
AGCATGCTGCACGACTATATCGTGTACGGTCACGCTATATCGCATCACGGGACGCCGGGTCATAGCTGTTTCCTG

5'CACGACGTTGTAAAACGACTCGTGGCAGATCAGTCCATAGCCGTCCAGACAAGAACAGTACGGCCAAGAAC
ATATCGTCCCAGATCCGACCATATCCTCTCTGATACGCCATATAACACAGAGGGATGGGTCATAGCTGTTTCCTG

ATCTGCA
ATCTGCG
ATCAGCA

ATCTGCA
ATCTGCG
ATCAGCA
ATCTGCA
ATCTGTA
ATATGCA

Alignment and consensus → Detach of primers → Check of errors via logical redundancy (Reed-Solomon)

```
primer      |-----------------|                                                                                                       |-----------------|
read1      TC*CGACGTTGTAAAACGACAGACAGGAGAAGCGTACTATATAAGGCCACAGACGATAAGGTGCTATCCGGTAGCATGCTGCACGACTATATCGTGTACGGTCACGC*--**C*---*ACG------------------------------
read2      ---------------------------G*A--********TAAGGC-*CAGACGATAAGGTGCTATCCGGTAGCATGCTGCACGACTATATCGTGTACGGTCACSCTATATCGCATCACGGGACGCCGGGTCATAGCTGTTTC**GA
consensus   C*CGACGTTGTAAAACGACAGACAGGAGAAGCGTACTATATAAGGCCACAGACGATAAGGTGCTATCCGGTAGCATGCTGCACGACTATATCGTGTACGGTCACGCTATATCGCATCACGGGACGCCGGGTCATAGCTGTTTC**G
original    CACGACGTTGTAAAACGACAGACAGGAGAAGCGTACTATATAAGGCCACAGACGATAAGGTGCTATCCGGTAGCATGCTGCACGACTATATCGTGTACGGTCACGCTATATCGCATCACGGGACGCCGGGTCATAGCTGTTTCCTG
errors       x                                                                                                                                              xx


primer      |-----------------|                                                                                                       |-----------------|
read1      TC*CGACGTTGTAAAACGACTCGTGGCAGATCAGTCCATAGCCGTCCAGACAAGAACAGTACGGCCAAGAACATATCGTCCCAGATCCGACCAT*-CCTCTC-*A**-GC******A------------------------------
read2      -----------------------------****G----T---CGT-CAGAC-AGAACAGTACGGCCAAGAACATATCGTCCCAGATCCGACCATATCCTCTCTGATACGCCATATAACACA*AGGGATGGGTCATAG*TGTTTC**GA
consensus   C*CGACGTTGTAAAACGACTCGTGGCAGATCAGTCCATAGCCGTCCAGACAAGAACAGTACGGCCAAGAACATATCGTCCCAGATCCGACCATATCCTCTCTGATACGCCATATAACACA*AGGGATGGGTCATAG*TGTTTC**G
original    CACGACGTTGTAAAACGACTCGTGGCAGATCAGTCCATAGCCGTCCAGACAAGAACAGTACGGCCAAGAACATATCGTCCCAGATCCGACCATATCCTCTCTGATACGCCATATAACACAGAGGGATGGGTCATAGCTGTTTCCTG
errors       x                                                                                                                     x             x      xx
```


Inverse mapping back to bit sequence

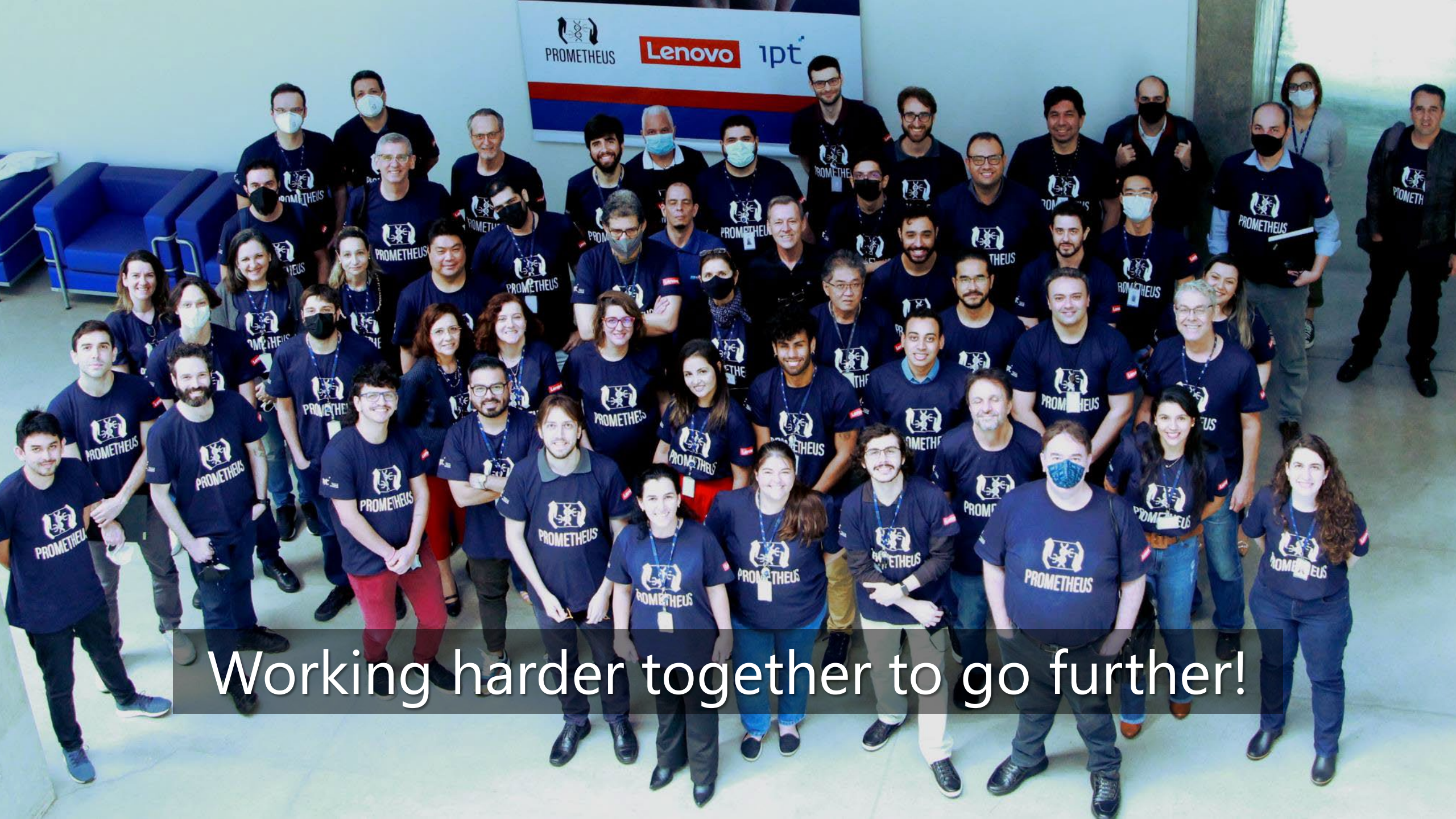
File reassembling according to segment addresses

IPT and Lenovo, a successful partnership!

Ongoing work

- Select the best surface for:
  - DNA synthesis
  - Recover of the synthesized DNA
- Develop a DNA storage protocol
- DNA synthesis inside the microfluidic device

DNA DATA STORAGE ALLIANCE
A SNIA TECHNOLOGY AFFILIATE


PROMETHEUS
Lenovo
ipt
Working harder together to go further!

Thank you!

Joao Lucas Maehara Said dos Reis (joaomaehara@ipt.br)
Marília Santos Menossi Mortari (mariliam@ipt.br)

