

A red-tinted microscopic image of several DNA double helix molecules. Superimposed over the DNA are numerous binary digits (0s and 1s) in white and light red, appearing as if they are floating or falling through the space between the DNA molecules.

Introduction to NGS

Walid Gharib



Swiss Institute
of
Bioinformatics

At the End of the course, you will/should be able to...

- Identify the functioning of 2nd / 3rd generation sequencing platforms Pros/Cons
- Familiarize with the lexicology
- Retreive data from NCBI databases using command line tools
- Identify and understand the content of different file formats used in the field
- Filter raw data for better downstream analysis
- Align processed raw data to a reference genome

Platforms



Swiss Institute of
Bioinformatics

Major technologies

illumina®

MiSeq
NextSeq 500
HiSeq 2500
HiSeq X

life
technologies™

SOLID
Ion PGM
Ion Proton



PACIFIC
BIOSCIENCES™

PacBio RS II

Oxford
NANOPORE
Technologies™

MinION
PromethION
GridION

Roche

454
SEQUENCING

GS Junior
GS FLX+

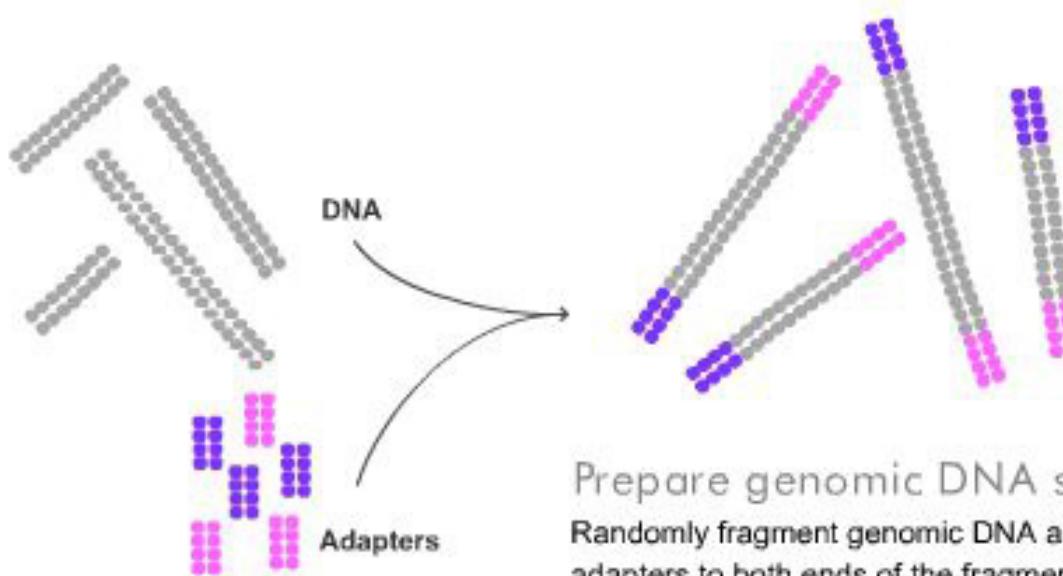
illumina® - Sequencing by synthesis



Swiss Institute of
Bioinformatics

Sequencing-By-Synthesis Demo

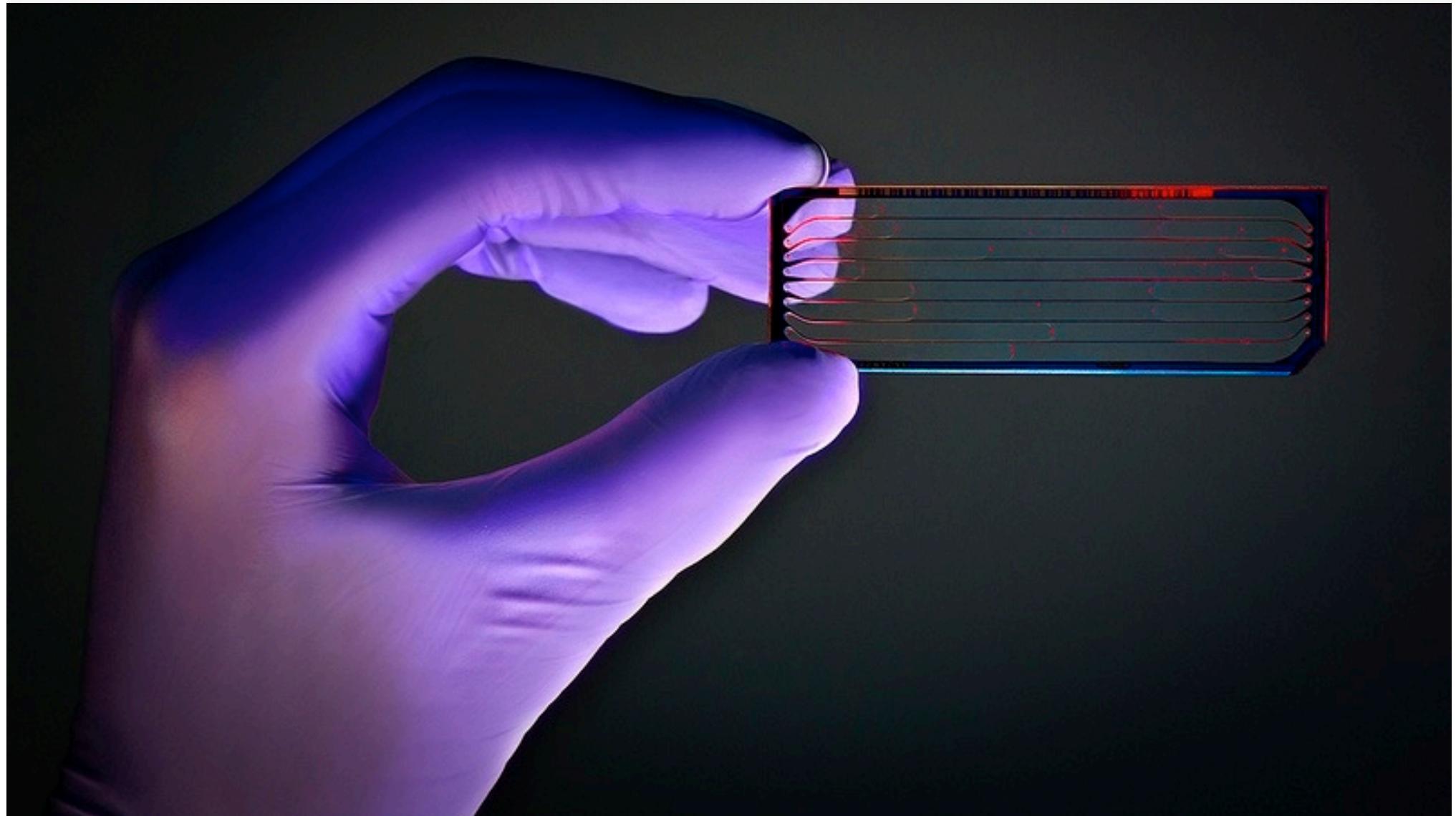
1



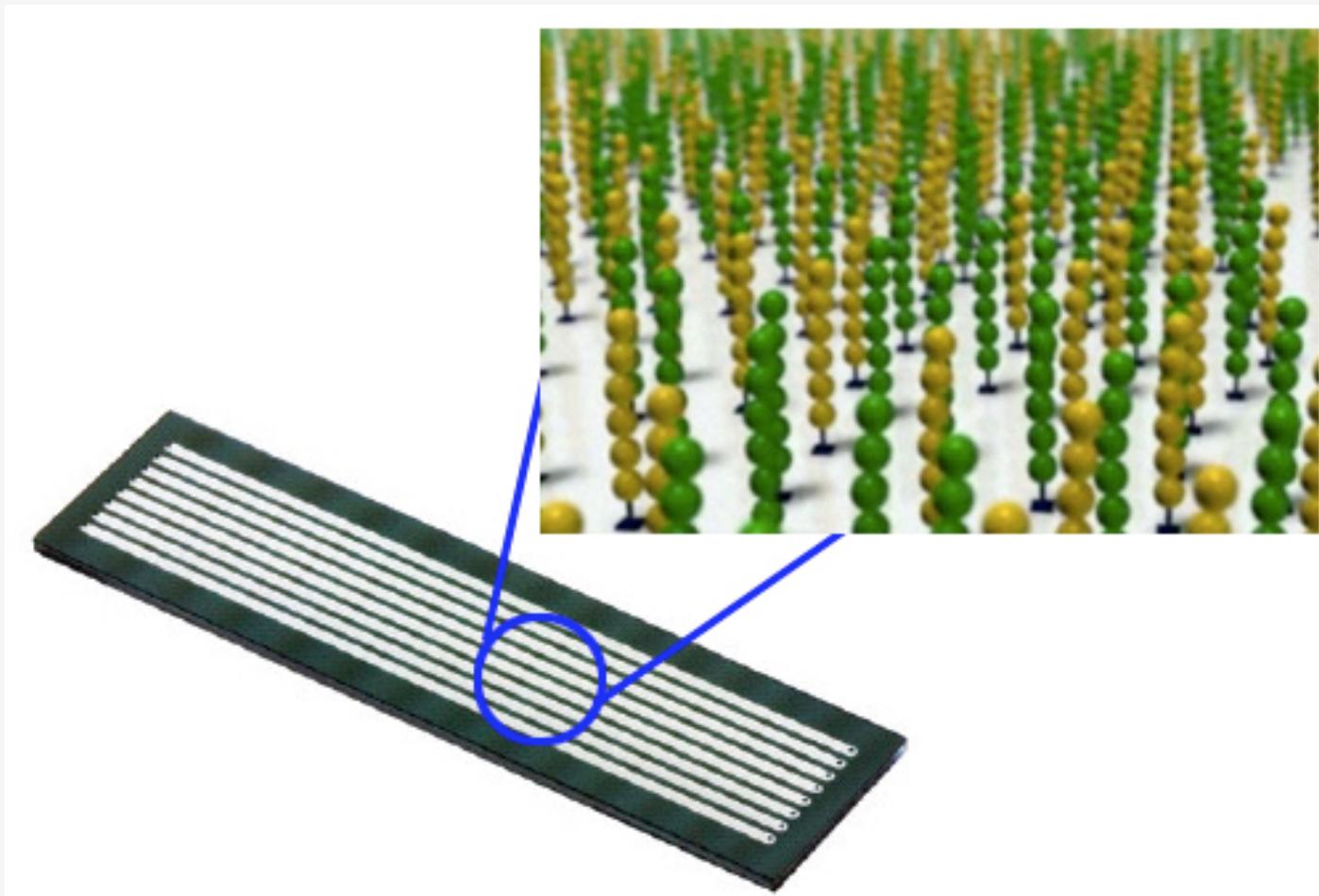
← PREV

NEXT →

Flow cell - 8 lanes



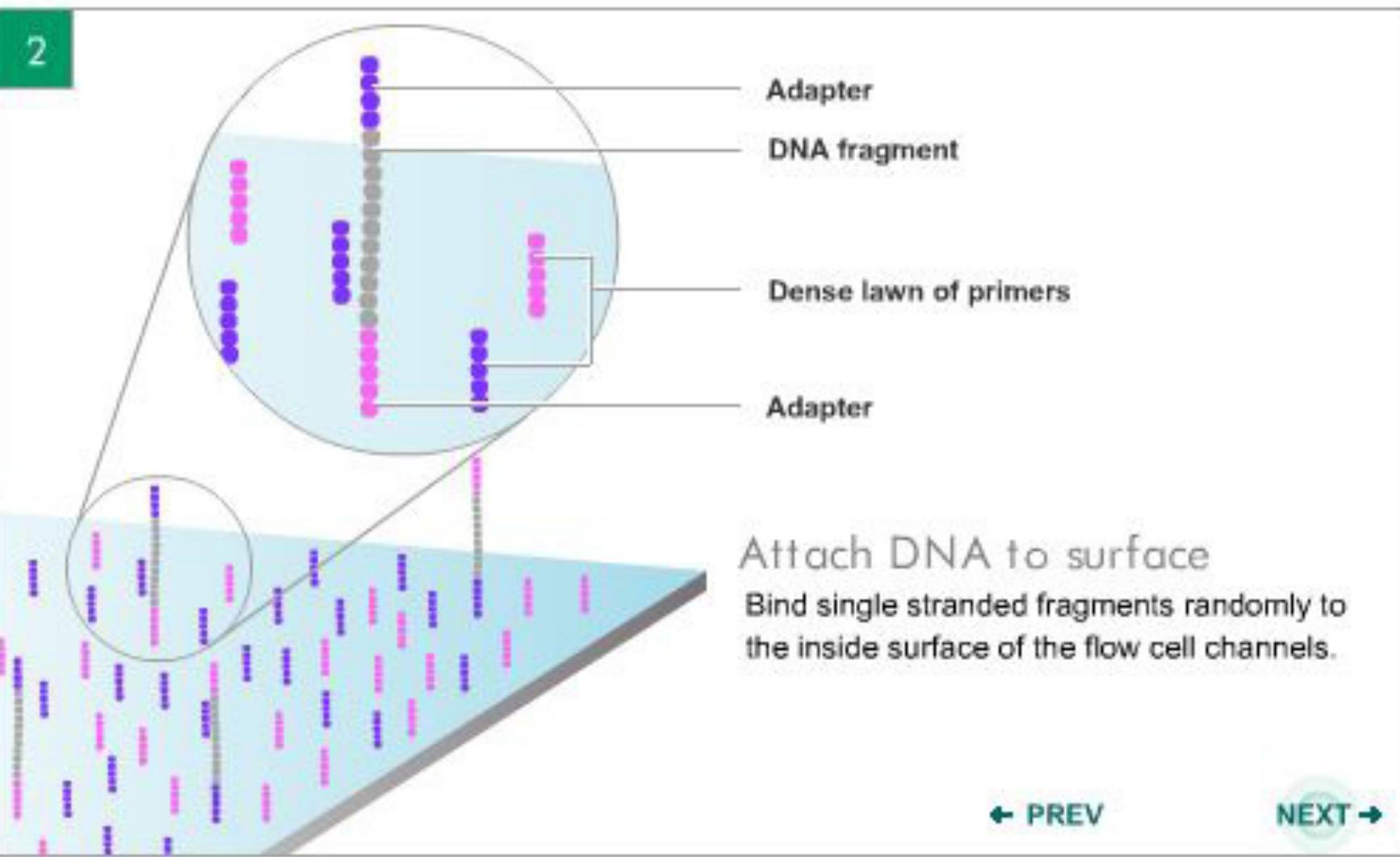
Adapters to the flow cell



Outline – Day2

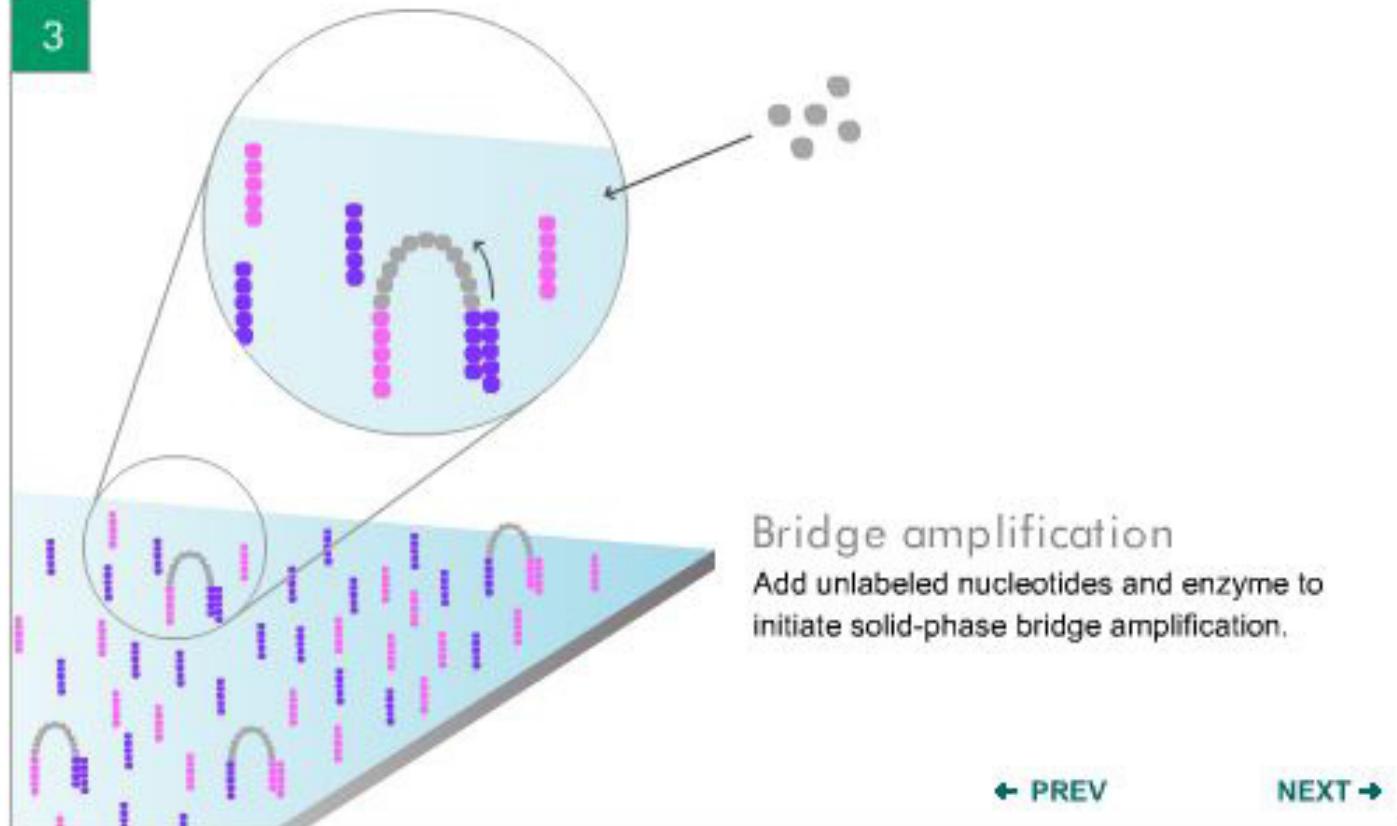
File Formats, Quality Assessment

Sequencing-By-Synthesis Demo



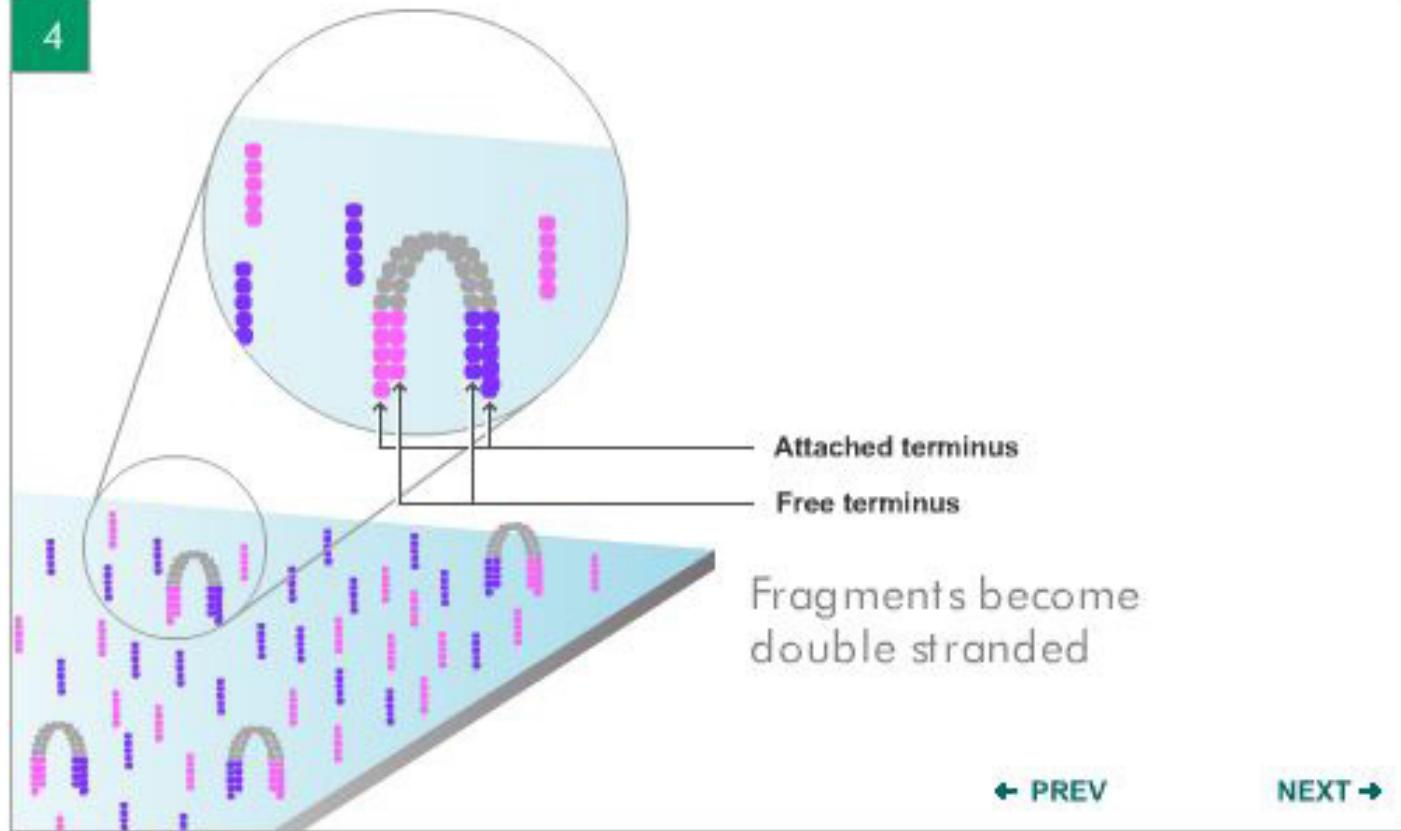
Sequencing-By-Synthesis Demo

3



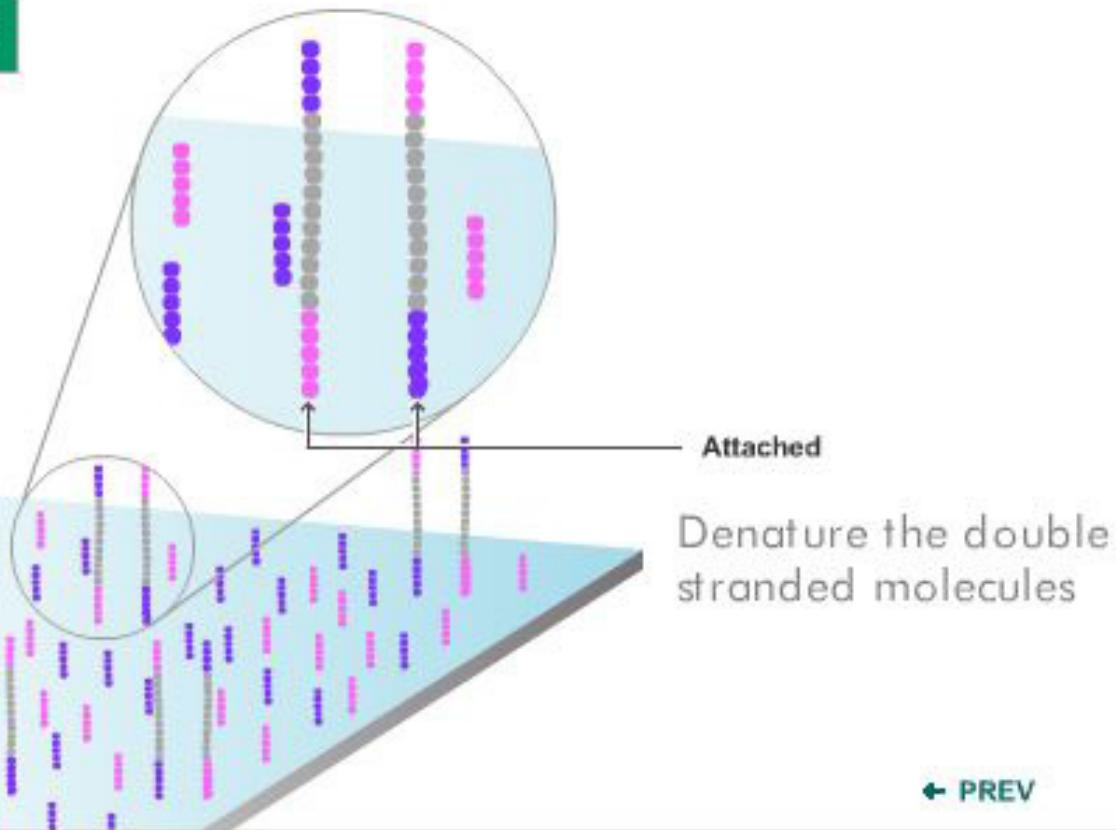
Sequencing-By-Synthesis Demo

4



Sequencing-By-Synthesis Demo

5

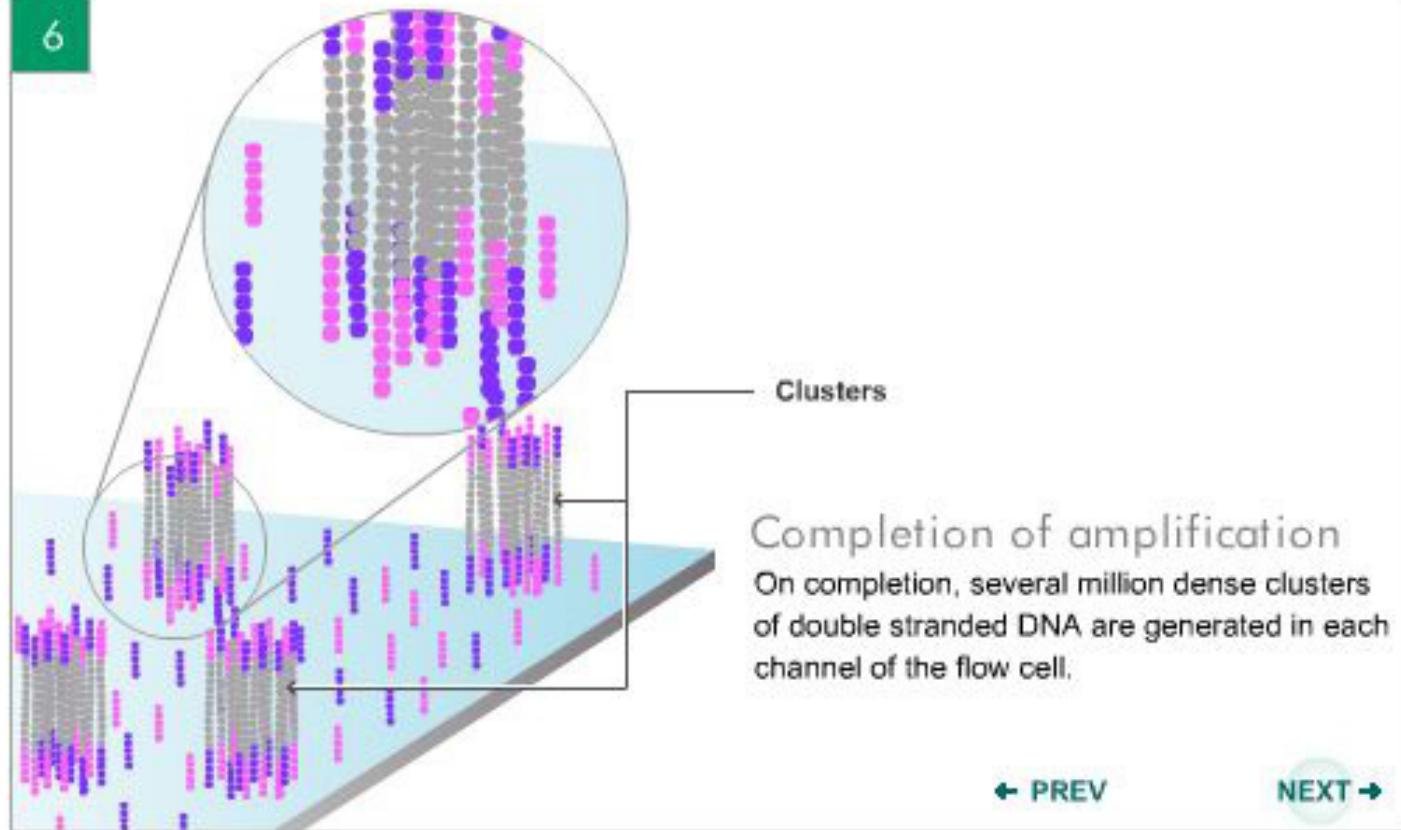


← PREV

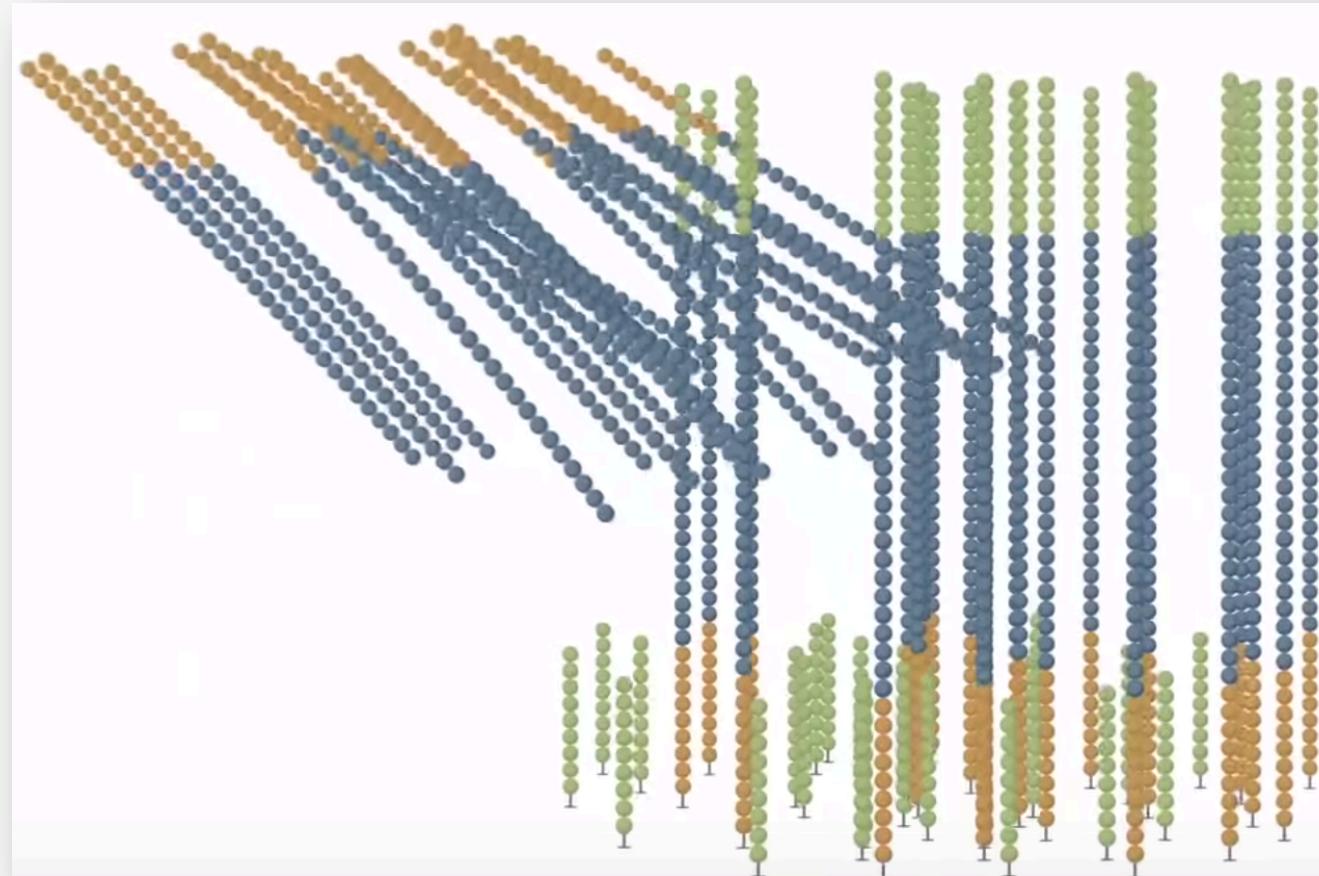
NEXT →

Sequencing-By-Synthesis Demo

6

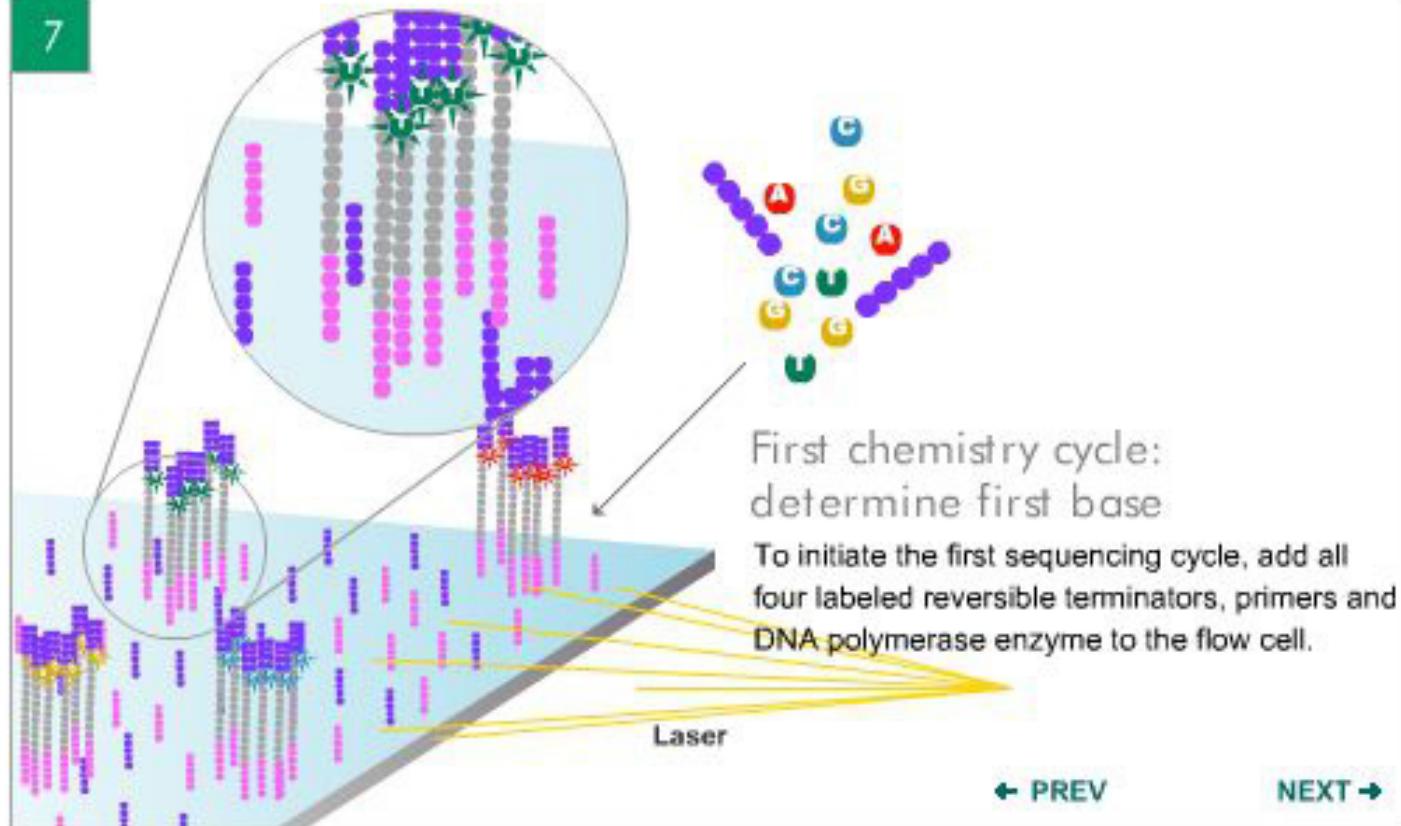


Cleavage of one strand



Sequencing-By-Synthesis Demo

7



Sequencing-By-Synthesis Demo

8

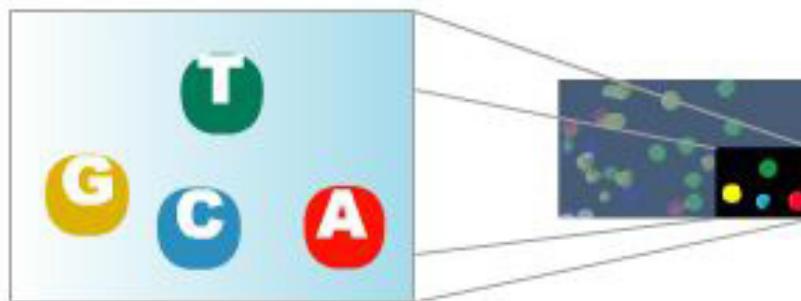


Image of first chemistry cycle

After laser excitation, capture the image of emitted fluorescence from each cluster on the flow cell. Record the identity of the first base for each cluster.

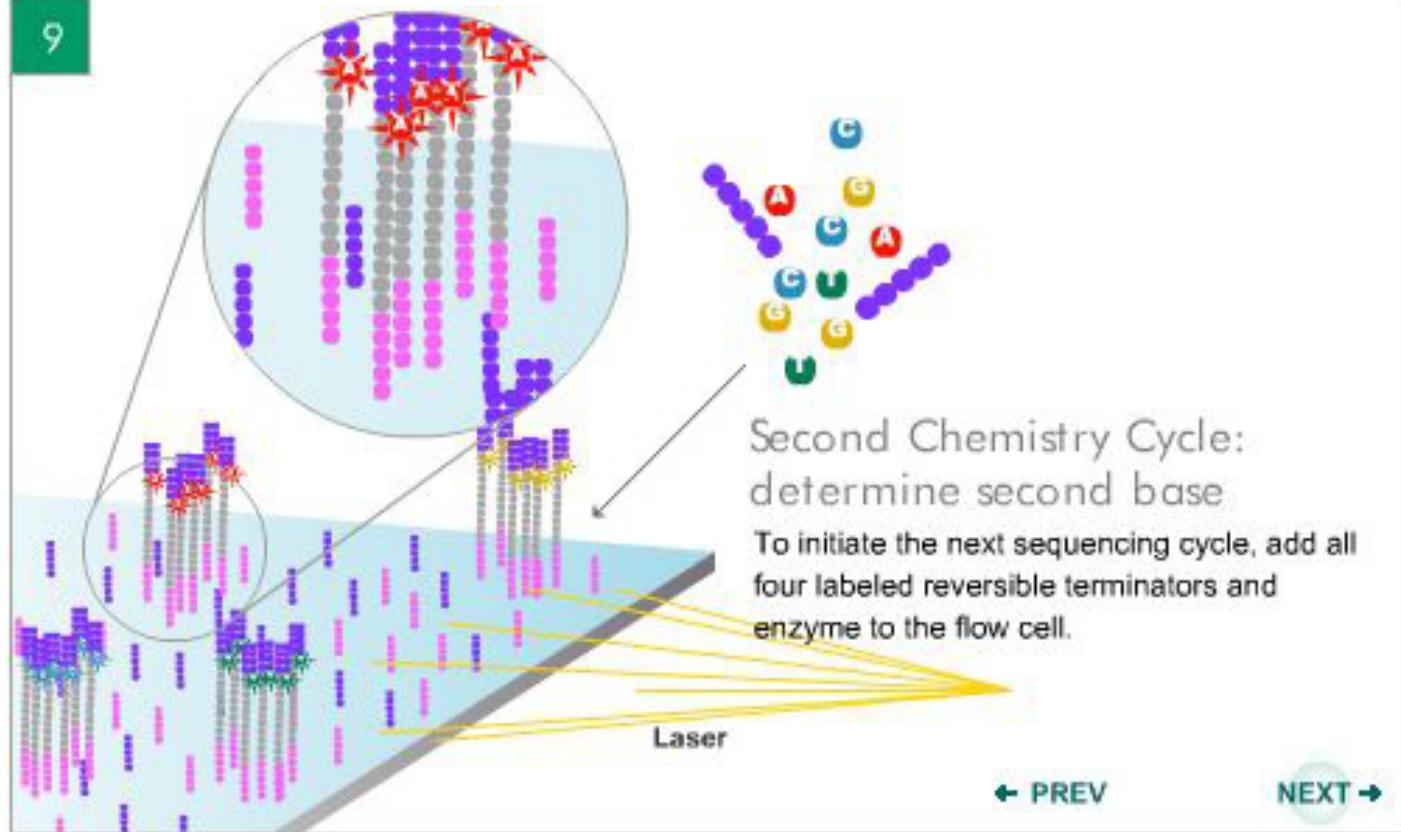
Before initiating the next chemistry cycle
The blocked 3' terminus and the fluorophore from each incorporated base are removed.

◀ PREV

NEXT ▶

Sequencing-By-Synthesis Demo

9



Sequencing-By-Synthesis Demo

10

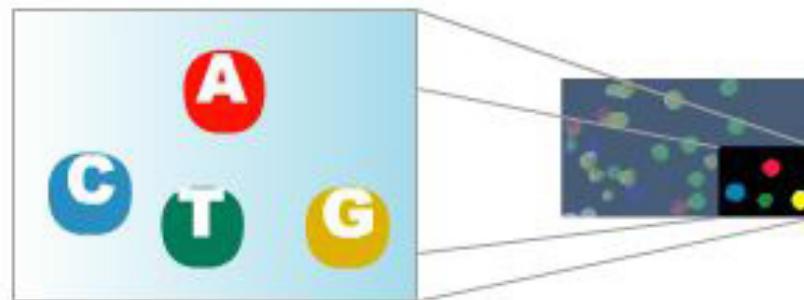


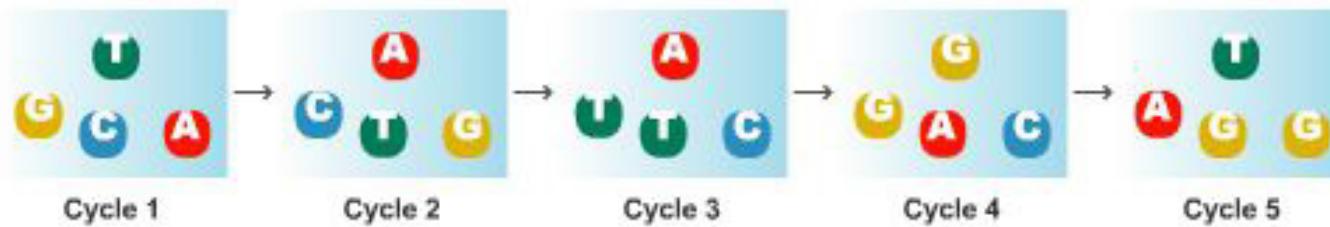
Image of second chemistry cycle
is captured by the instrument
After laser excitation, collect the image data as
before. Record the identity of the second base
for each cluster.

◀ PREV

NEXT ▶

Sequencing-By-Synthesis Demo

11



Cycle 1

Cycle 2

Cycle 3

Cycle 4

Cycle 5

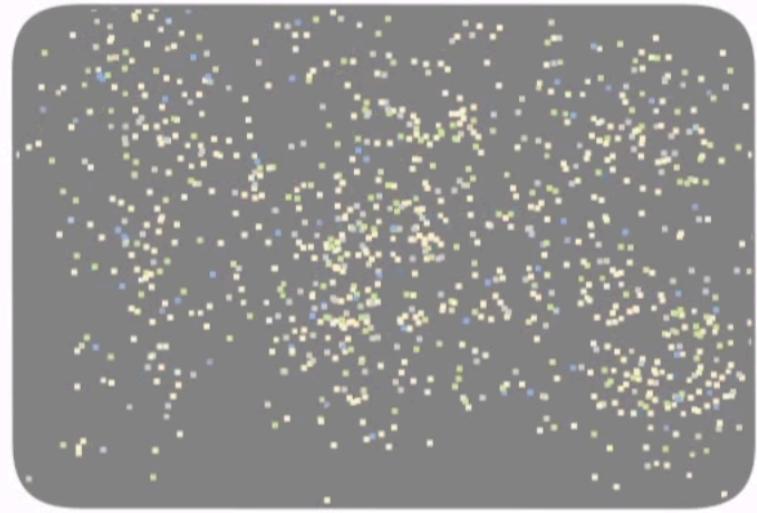
GCTGA....

Sequence read over multiple chemistry cycles

Repeat cycles of sequencing to determine the sequence of bases in a given fragment a single base at a time.

← PREV

NEXT →



Flow cell

AAACCTCCTTCTTATTTCAG
AGCAGTAGTAAGAAACAAA/
TTGACAAACCTCCTTCTTAT
AGTAGTAAGAAACAAA TGG
ACAAAAAGCAATTGACAAAC
AAACCTCCTTCTTATTCTT G
AGCAGTAGTAAGAAACAAA
TTGACAAACCTCCTTACTAC
AGTAGTAAGAAACAAAAA GG
ACAAAAAGCAATTGA CTTAC
AAACCTCCTTCTTATTCT AG
AGCAGTAGTAAGAAA TCAA
TTGACAAACCTCCTTCTTAC
AGTAGTAAGAAACAAAAGC
ACAAAAAGCAATTGACA TAC
AAACCTCCTTCTTATTCTTAC
AGCAGTAGTAAGAA CTCAA
TTGACAAACCTCCTTCTTAT
AGTAGTAAGAAACA AATGG



Benchtop Sequencers

Production-Scale Sequencers

	iSeq 100 System	MiniSeq System	MiSeq Series	NextSeq Series
Popular Applications & Methods	Key Application	Key Application	Key Application	Key Application
Large Whole-Genome Sequencing (human, plant, animal)				●
Small Whole-Genome Sequencing (microbe, virus)	●	●	●	●
Exome Sequencing				●
Targeted Gene Sequencing (amplicon, gene panel)	●	●	●	●
Whole-Transcriptome Sequencing				●
Gene Expression Profiling with mRNA-Seq				●
Targeted Gene Expression Profiling	●	●	●	
Long-Range Amplicon Sequencing*	●	●	●	
miRNA & Small RNA Analysis	●	●	●	●
DNA-Protein Interaction Analysis			●	●
Methylation Sequencing				●
16S Metagenomic Sequencing		●	●	●

How to Choose a Benchtop Sequencer

This Benchtop Sequencing Buyer's Guide has tips to help you make a smooth transition to next-generation sequencing and select the best benchtop sequencing system to achieve your research objectives.

[Read Benchtop Buyer's Guide >](#)

Run Time	9–17.5 hours	4–24 hours	4–55 hours	12–30 hours
Maximum Output	1.2 Gb	7.5 Gb	15 Gb	120 Gb
Maximum Reads Per Run	4 million	25 million	25 million †	400 million
Maximum Read Length	2 × 150 bp	2 × 150 bp	2 × 300 bp	2 × 150 bp

[Explore iSeq](#) [Explore MiniSeq](#) [Compare MiSeq](#) [Compare NextSeq](#)



Benchtop Sequencers

Production-Scale Sequencers

Popular Applications & Methods	Key Application	Key Application	Key Application	Key Application
Large Whole-Genome Sequencing (human, plant, animal)	●	●	●	●
Small Whole-Genome Sequencing (microbe, virus)	●	●		●
Exome Sequencing	●	●		●
Targeted Gene Sequencing (amplicon, gene panel)	●	●		●
Whole-Transcriptome Sequencing	●	●		●
Gene Expression Profiling with mRNA-Seq	●	●		●
miRNA & Small RNA Analysis	●	●		●
DNA-Protein Interaction Analysis	●	●		●
Methylation Sequencing	●	●		●
Shotgun Metagenomics	●	●		●

Optimized NGS Sample Tracking and Workflows

See how BaseSpace Clarity LIMS (Laboratory Information Management System) enabled this large genomics lab to standardize lab procedures and cope with increasing sample volumes from diverse clients.

[Read Case Study >](#)

Run Time	12–30 hours	< 1–3.5 days (HiSeq 3000/HiSeq 4000) 7 hours–6 days (HiSeq 2500)	< 3 days	16–36 hours (Dual S2 flow cells) 44 hours (Dual S2 flow cells)
Maximum Output	120 Gb	1500 Gb	1800 Gb	6000 Gb
Maximum Reads Per Run	400 million	5 billion	6 billion	20 billion
Maximum Read Length	2 × 150 bp	2 × 150 bp	2 × 150 bp	2 × 150 bp

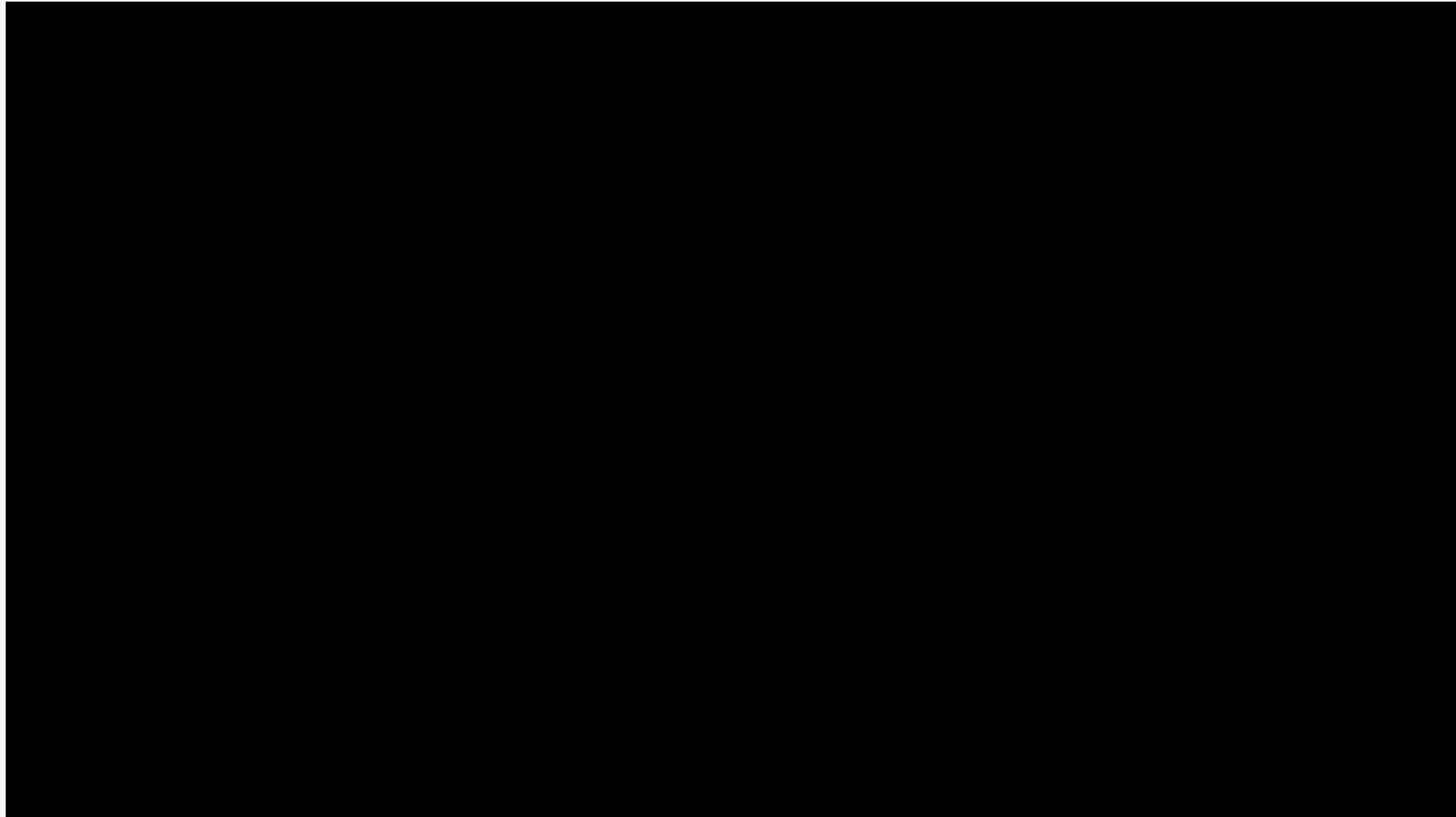
[Compare NextSeq](#) [Compare HiSeq](#) [Explore HiSeq X](#) [Explore NovaSeq](#)

[Request Pricing >](#) [Request Pricing >](#) [Request Pricing >](#) [Request Pricing >](#)

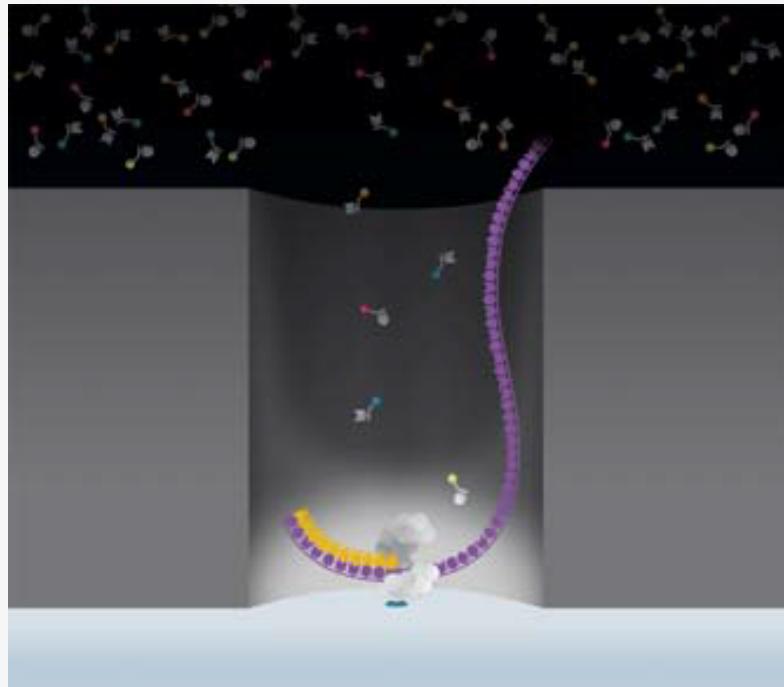
Pacific Biosciences SMRT



PACIFIC
BIOSCIENCES®

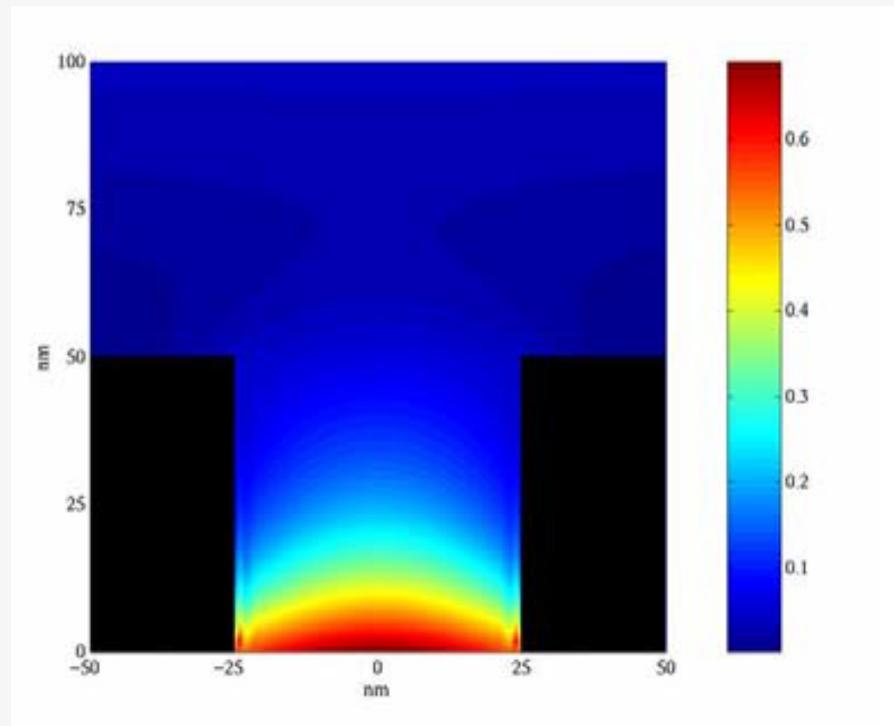


Pacific Biosciences SMRT Sequencing



Individual ZMW

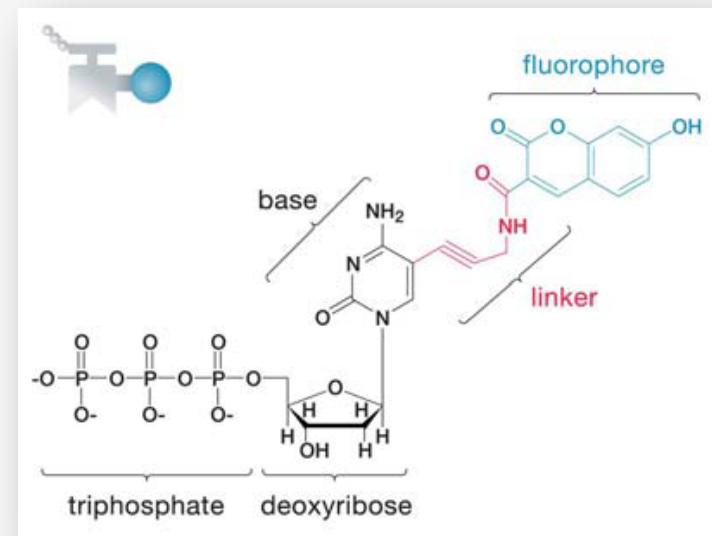
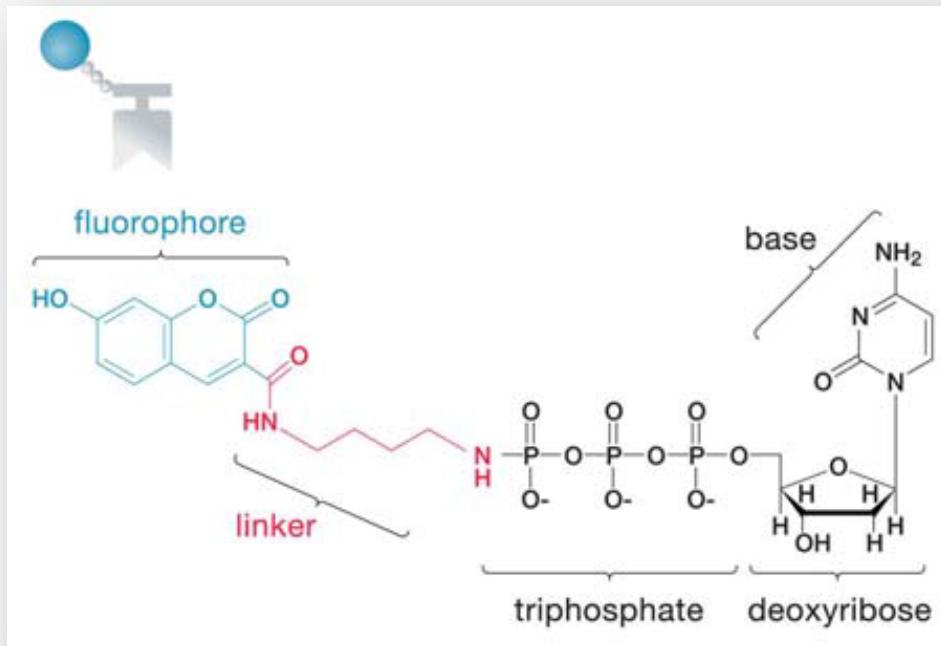
Each ZMW is a cylindrical hole tens of nanometers in diameter, perforating a thin metal film supported by a transparent



Detection volume

Attenuated light from the excitation beam penetrates only the lower 20-30 nm of each waveguide, creating a detection volume of 20 zeptoliters (10-21 liters).

Phospholinked Fluorophores

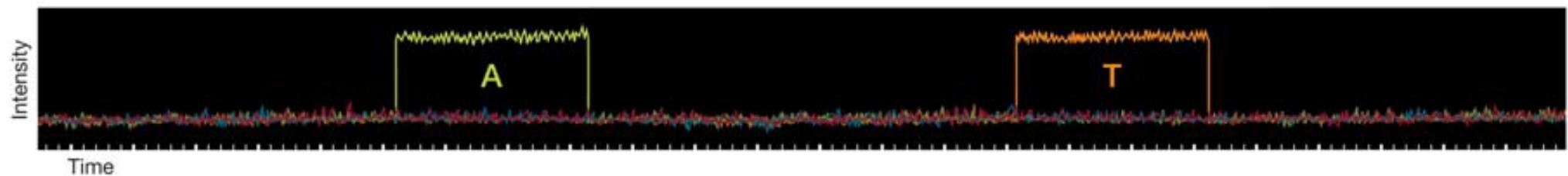


Phospholinked nucleotides

Phospholinked nucleotides have fluorophores attached to the triphosphate chain, which is naturally cleaved when the nucleotide is incorporated.

Base-labeled nucleotide. Base-labeled nucleotides have fluorophores chemically attached directly to the base.

Processive Synthesis



Processive Synthesis with Phospholinked Nucleotides.

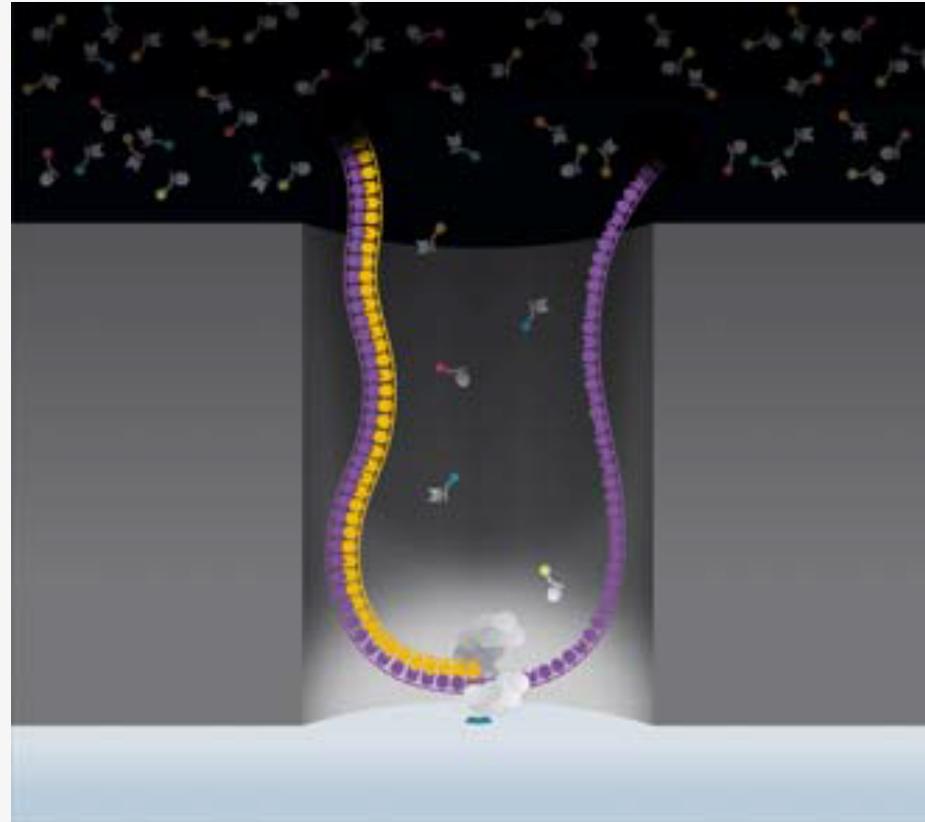
Step 1: Fluorescent phospholinked labeled nucleotides are introduced into the ZMW.

Step 2: The base being incorporated is held in the detection volume for tens of milliseconds, producing a bright flash of light.

Step 3: The phosphate chain is cleaved, releasing the attached dye molecule.

Step 4-5: The process repeats.

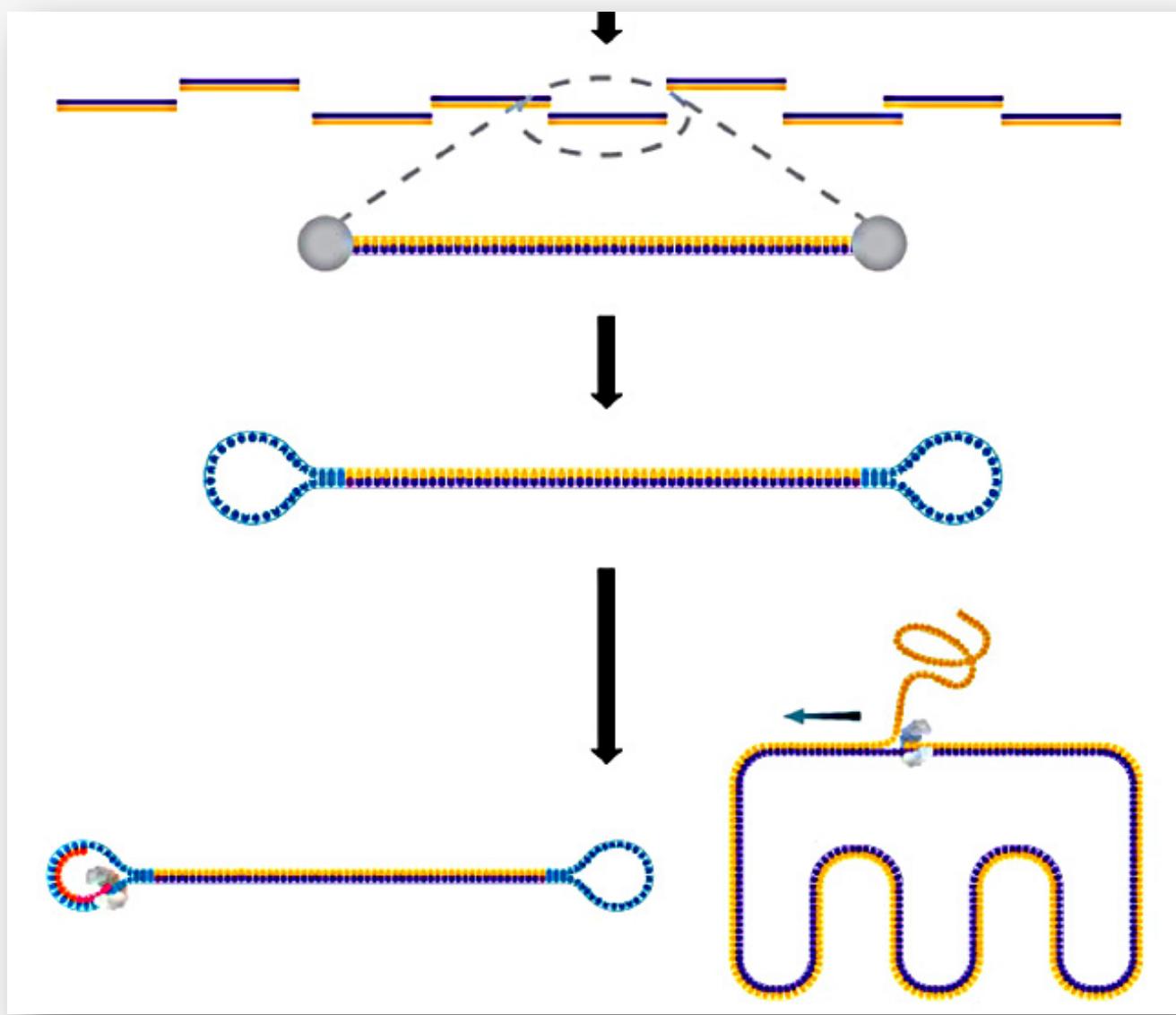
Synthesis of Long Duplex DNA



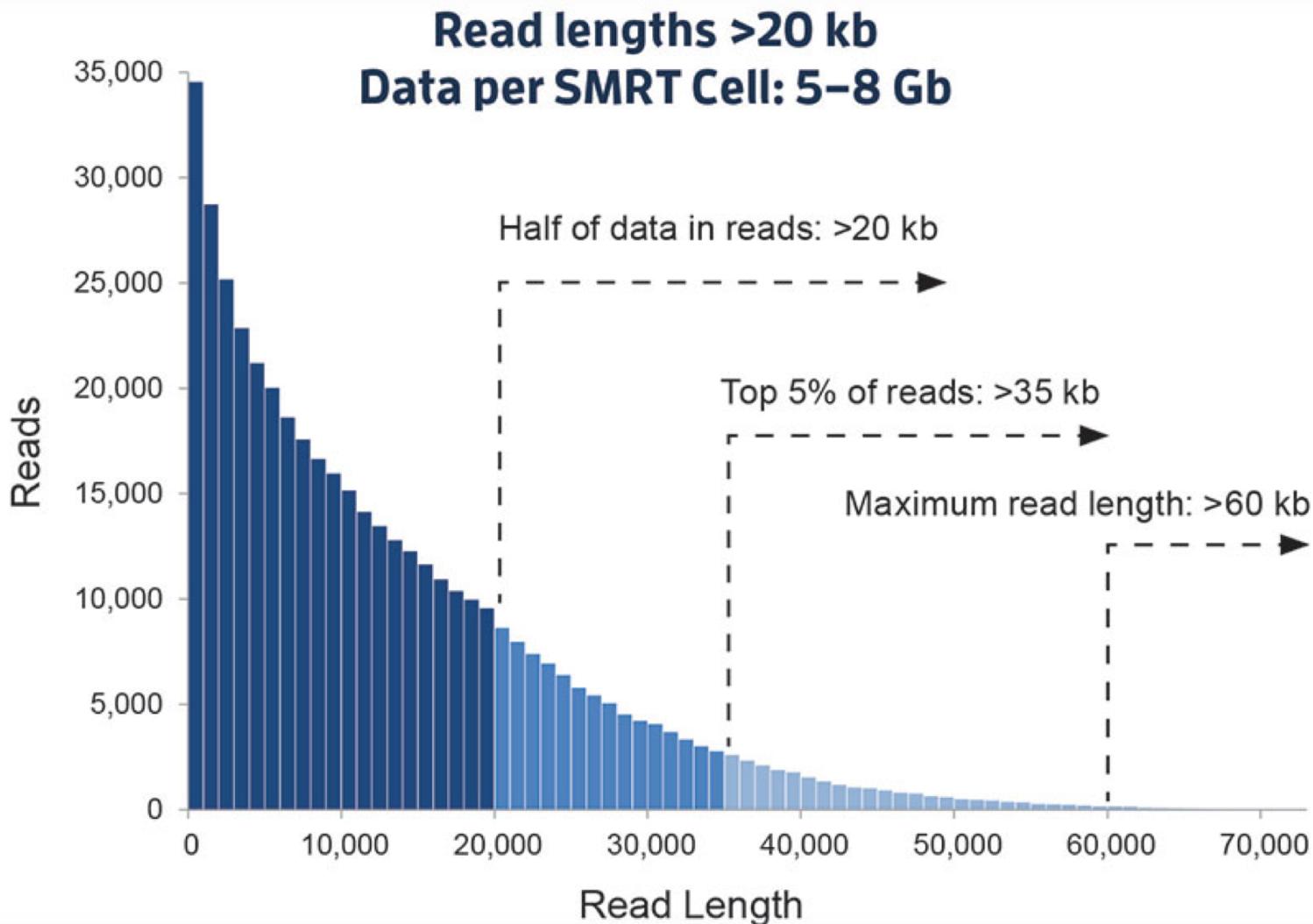
Synthesis of long DNA.

DNA polymerase processively incorporates nucleotides producing long, natural DNA.

Circular Templates Gives Redundant Sequencing and Accuracy







RSII flowcell has 150,000 ZMWs, Sequel flowcell has 1,000,000 ZMWs;
Typical RSII flowcell yield in ~1G base pairs of sequenced nucleotides; Sequel 5 - 8 Gbp.
Optimal sequencing run yields in half or reads longer than 20k and some reads longer than 60k

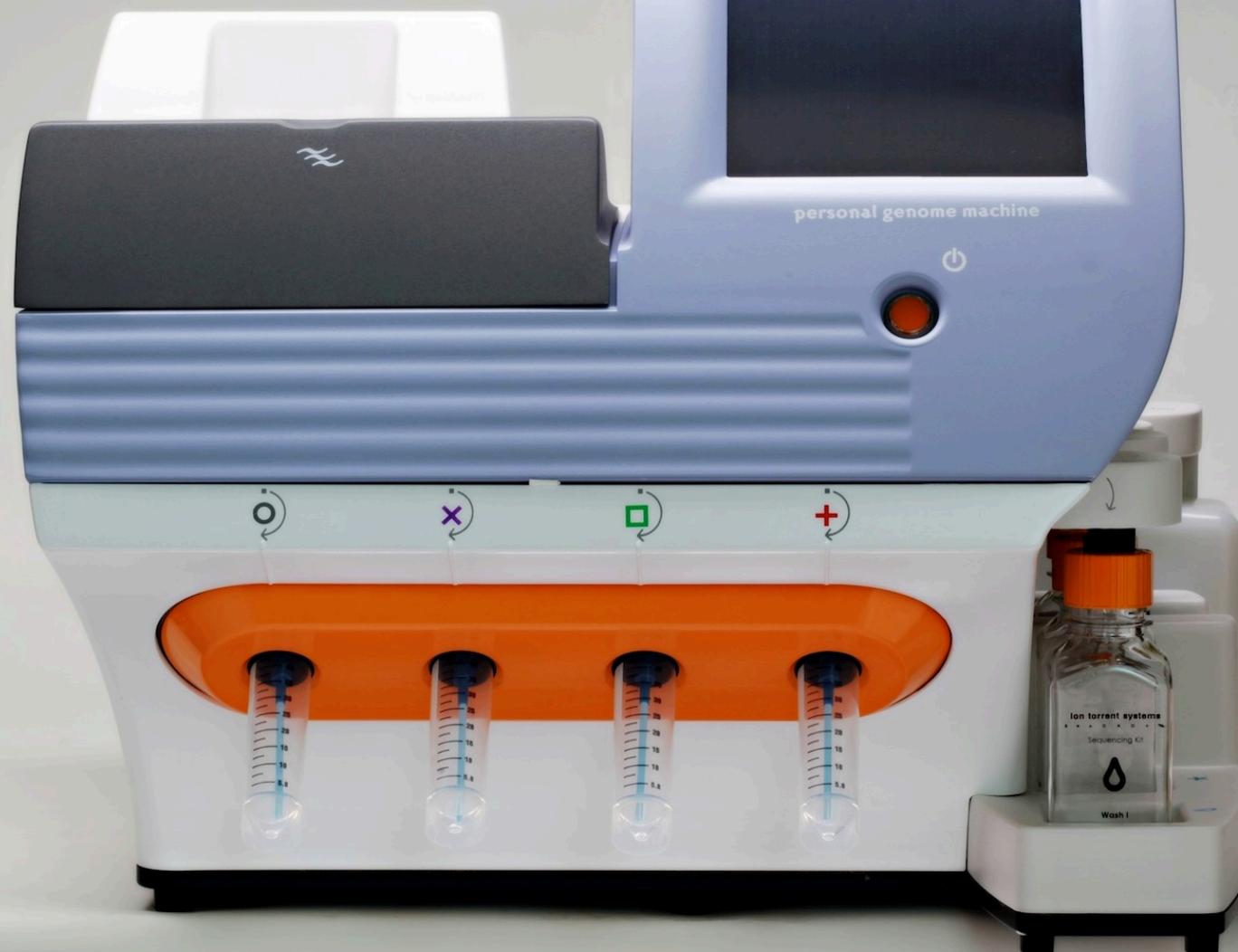
ion torrent



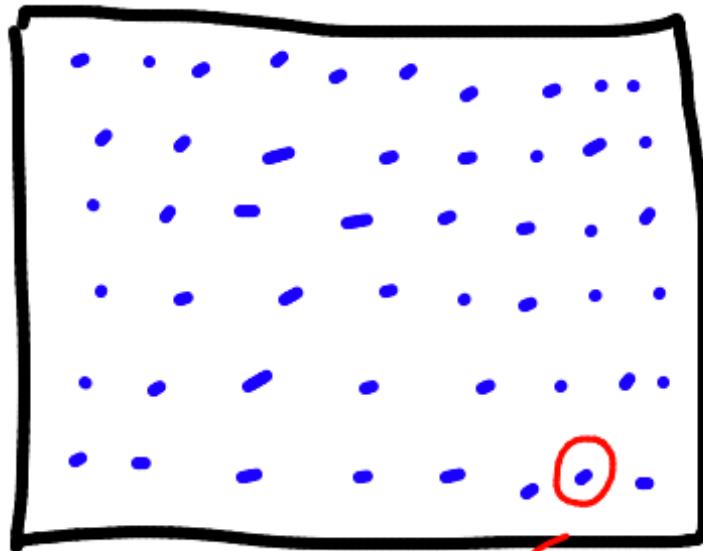
by *life* technologies™

ion torrent systems

personal genome machine



Ion Torrent



1e8 wells

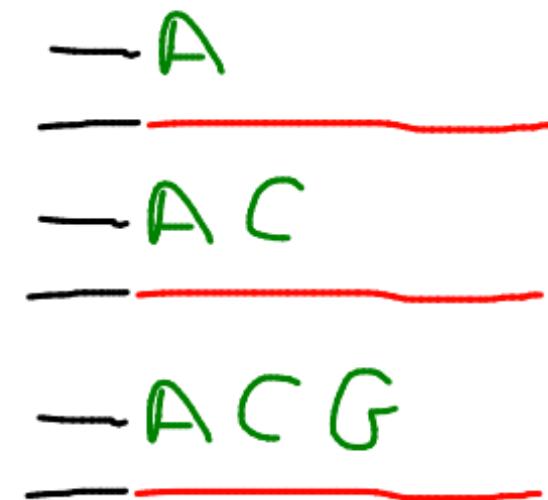
Each one is a mini pH meter



Add A

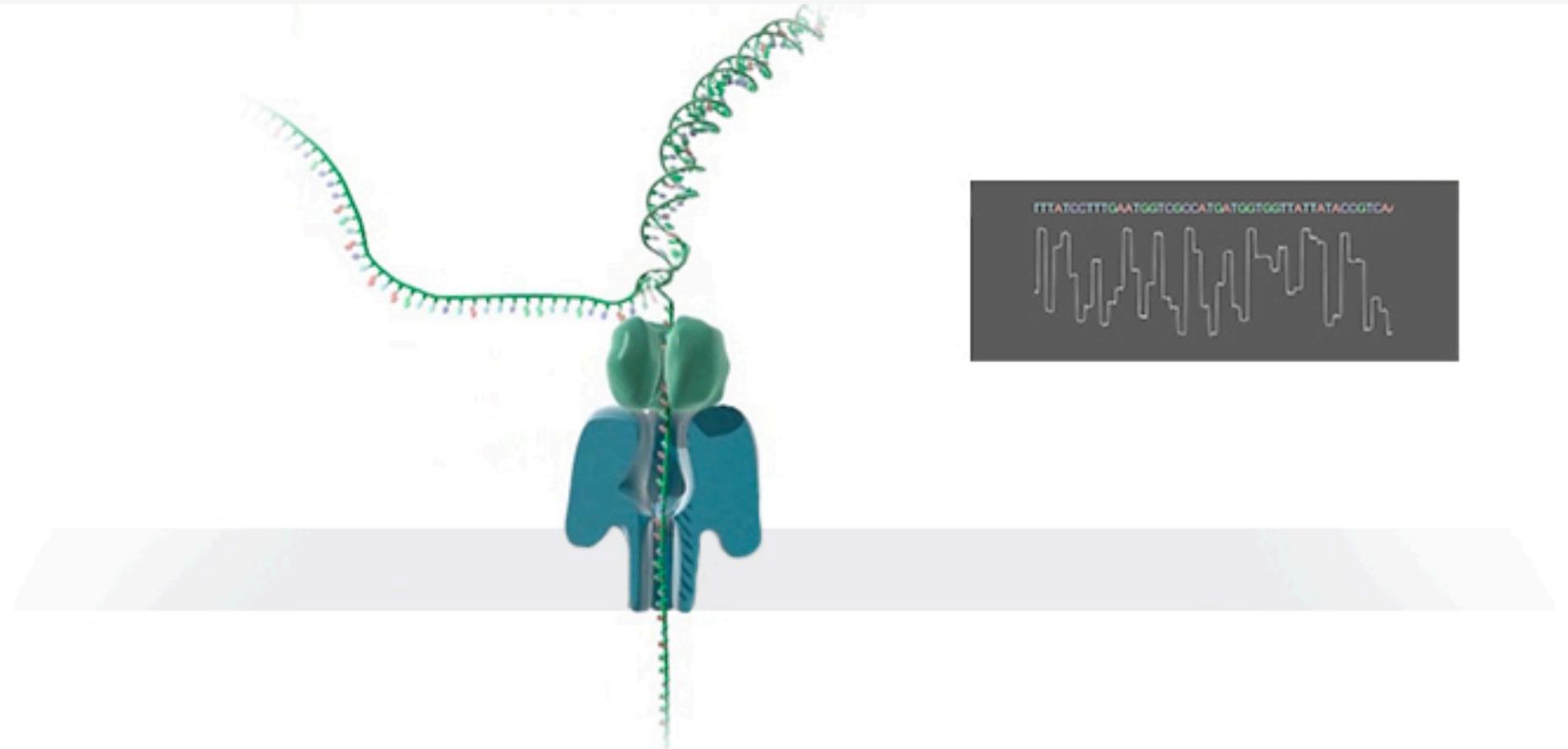
Did H get released
for this well?

Yes? Then next
base was A.

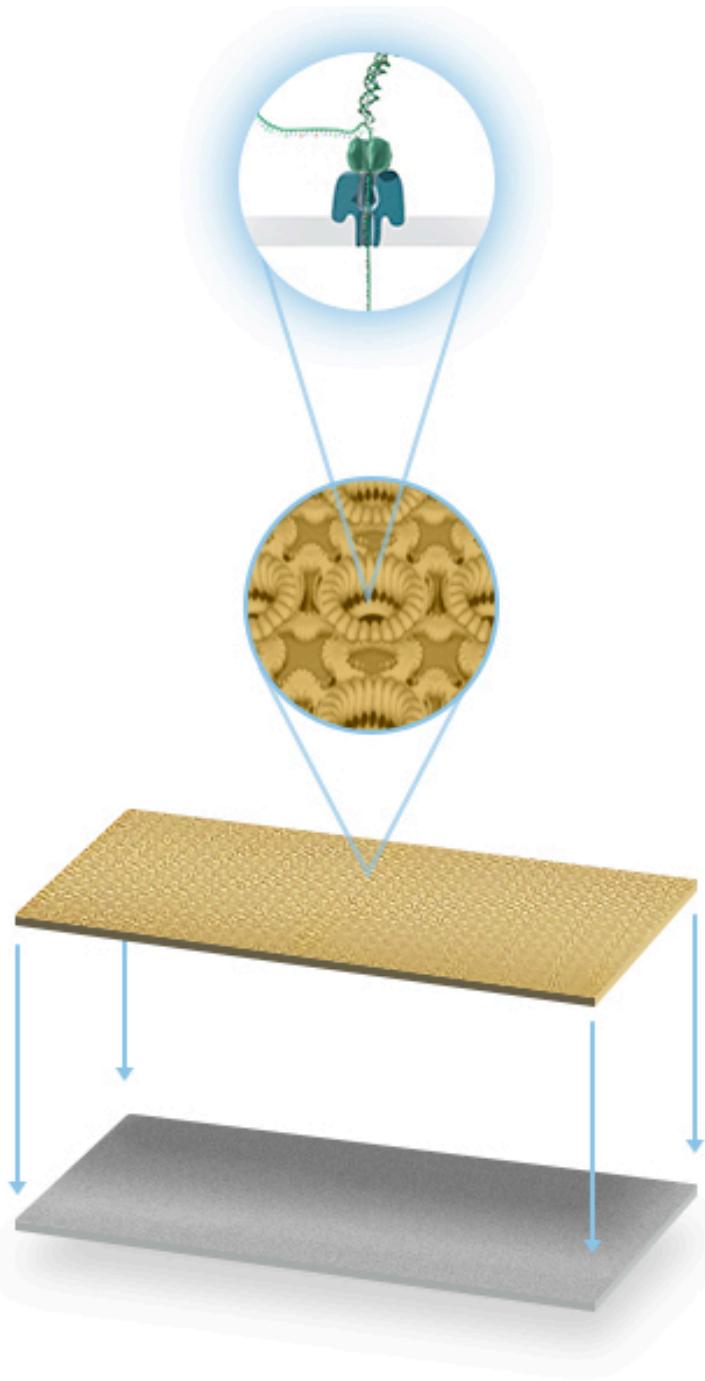


~ 6 hrs for sample
prep plus run => data

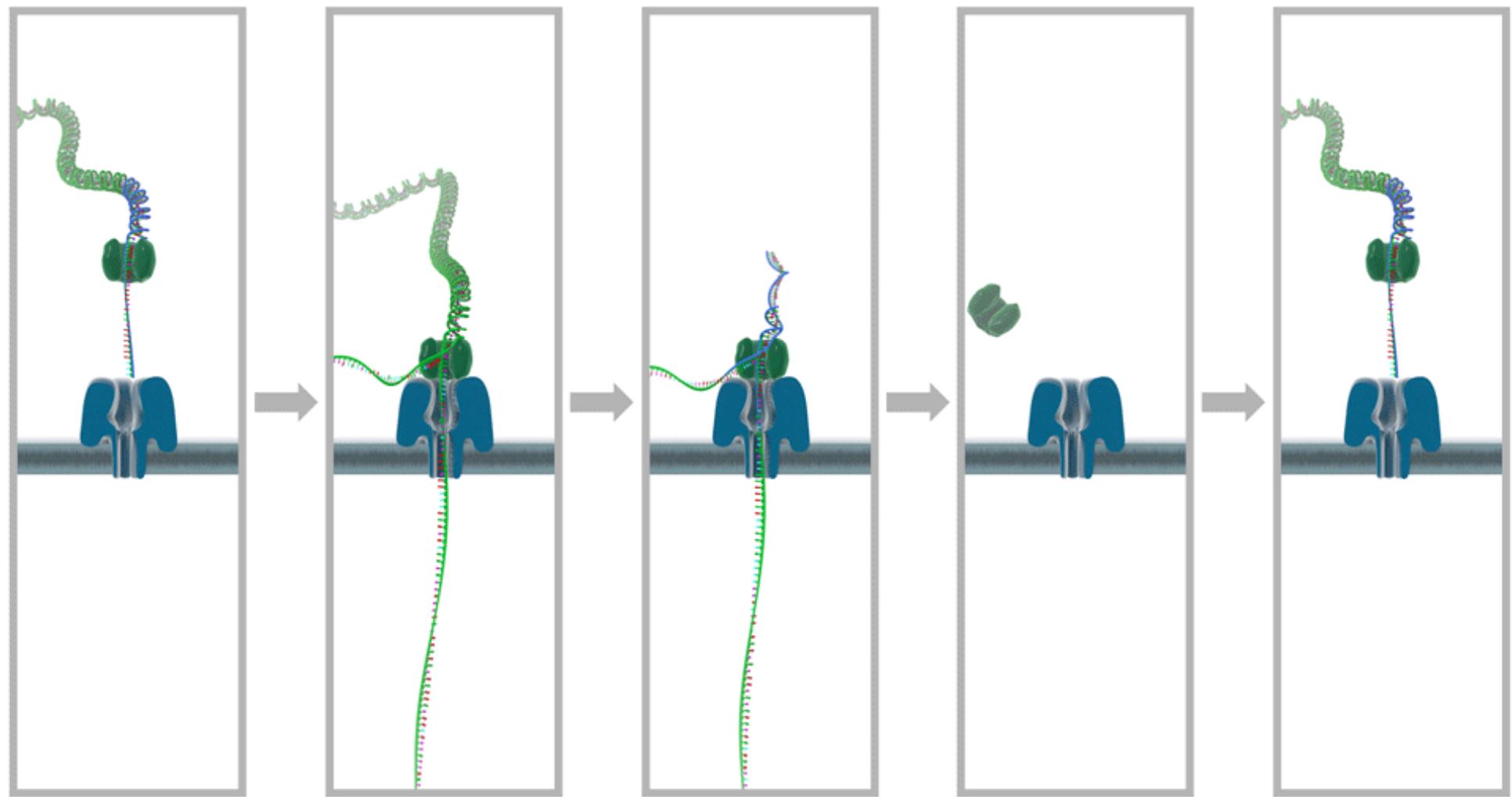
\$500 or so.

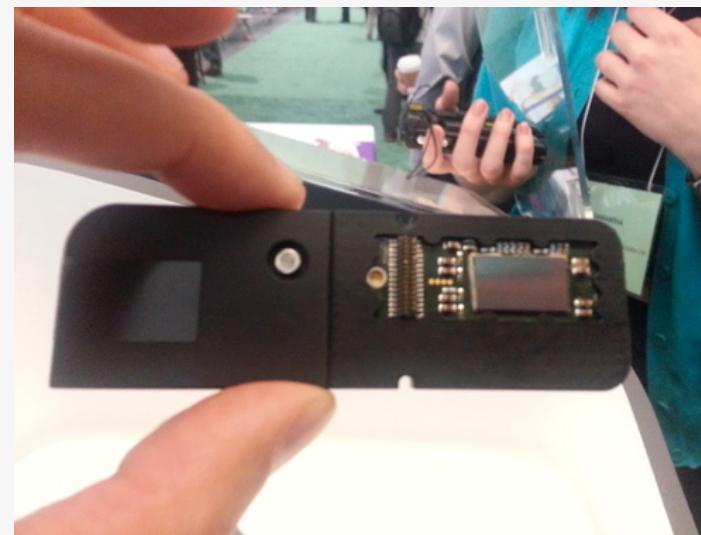
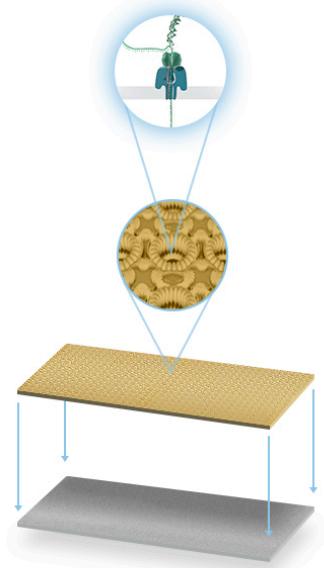


<http://erlichya.tumblr.com/post/64875644178/speaking-with-oxford-nanopore-representatives>



- Each consumable flow cell can now generate 10–20 Gb of DNA sequence data.
- Ultra-long read lengths are possible (hundreds of kb) as you can choose your fragment length.





PromethION



- Choose your read length: 5kb up to 200Kb
- Real time
- Low cost material and library preps



Biomolecular Detection and Quantification

Volume 3, March 2015, Pages 1–8



Open Access

Original Article

Assessing the performance of the Oxford Nanopore Technologies MinION

T. Laver^{a, 1}, , , J. Harrison^{a, 1}, , P.A. O'Neill^{a, b}, , K. Moore^{a, b}, , A. Farbos^{a, b}, , K. Paszkiewicz^{a, b}, , D.J. Studholme^a,

Open Access funded by Biotechnology and Biological Sciences Research Council

Under a Creative Commons license

Show more

doi:10.1016/j.bdq.2015.02.001

[Get rights and content](#)



PRODUCTS HOW IT WORKS

Human Genome on a MinION

Thu 20th October 2016

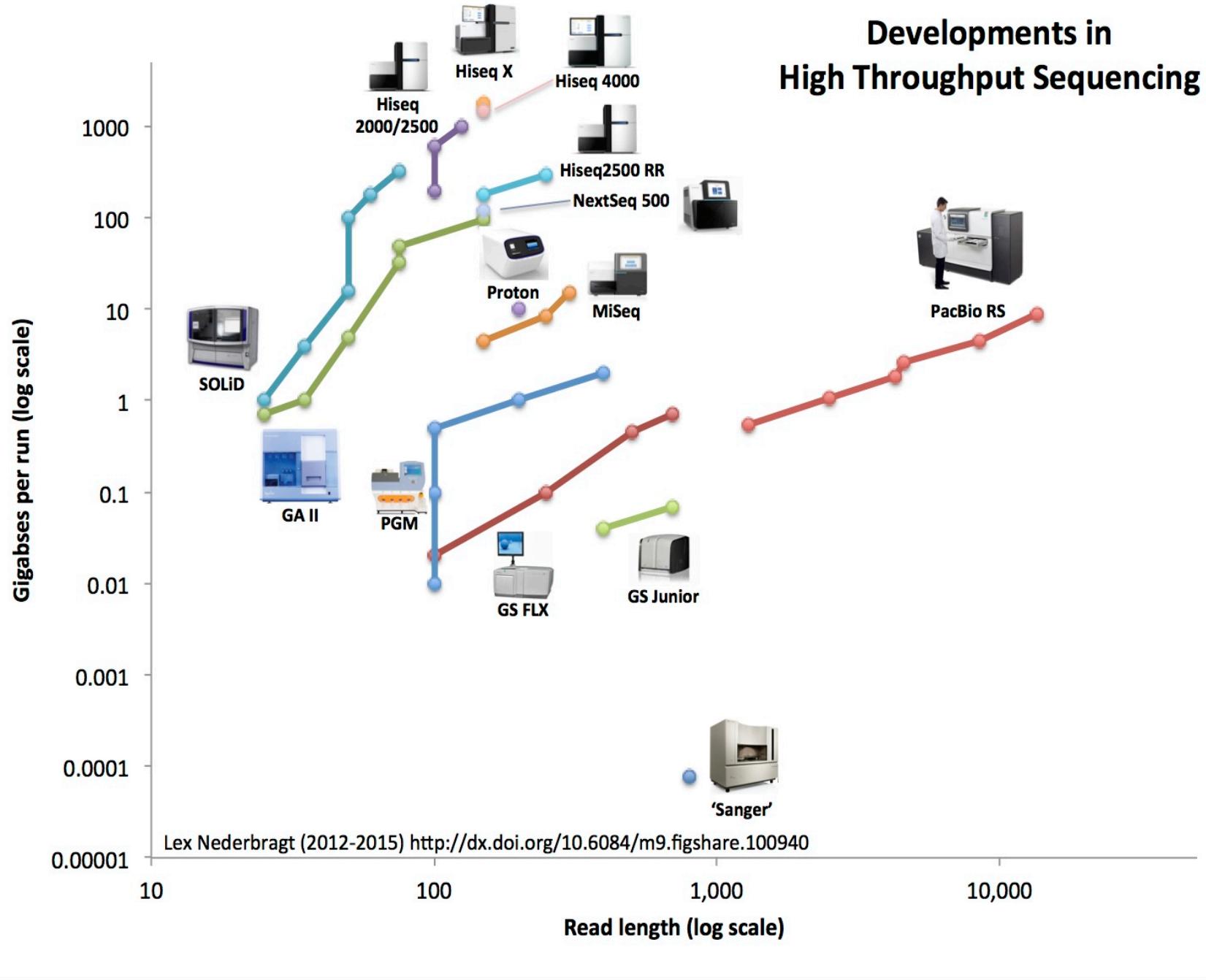
Oxford Nanopore announces today the result of our first attempt to sequence a human genome using the portable MinION device.

Human Genome Data using nanopore technology will be released soon – leave your details [here](#) if you would like to be notified when it is made available.

In the meantime, preliminary analysis of this first genome gives the following metrics:

<https://nanoporetech.com/about-us/news/human-genome-minion>

Developments in High Throughput Sequencing



Read size

illumina®

50-600

10K

life
technologies™

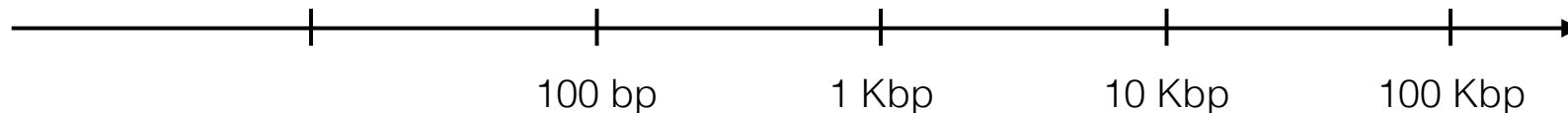
200-400

5 to 10 Kb

Oxford
NANOPORE
Technologies™

PB
PACIFIC
BIOSCIENCES™

250 bp to 20 Kb



Major technologies



Short reads
High throughput
Paired reads



Short reads
Medium throughput
Short run time
Library prep difficult



PACIFIC
BIOSCIENCES™

Long reads
Individual molecules
Base modifications
High, random error rate
Low throughput



Long reads
Individual molecules
Error rate!



454

SEQUENCING

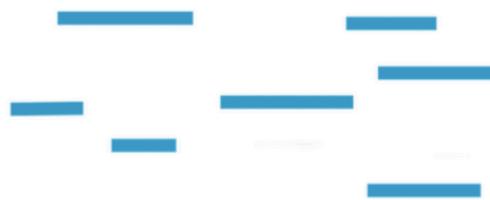
retired

Paired end Sequencing

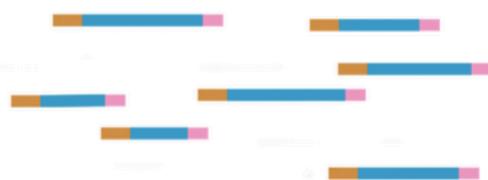
- Direct sequencing method: Input genomic DNA is fragmented by methods such as nebulization, hydrodynamic shearing, sonication...



Fragments of <800pb are selected



The ends are repaired, 3' ends are adenylated and paired-end adapters are added



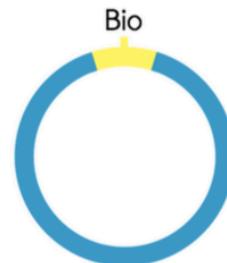
The product is amplified by PCR, purified and sequenced from both ends

Mate pair Sequencing

- For mate pair sequencing, the DNA is fragmented into 2-5kb segments



The labeled fragments are circularized



and fragmented again into 400-600bp pieces



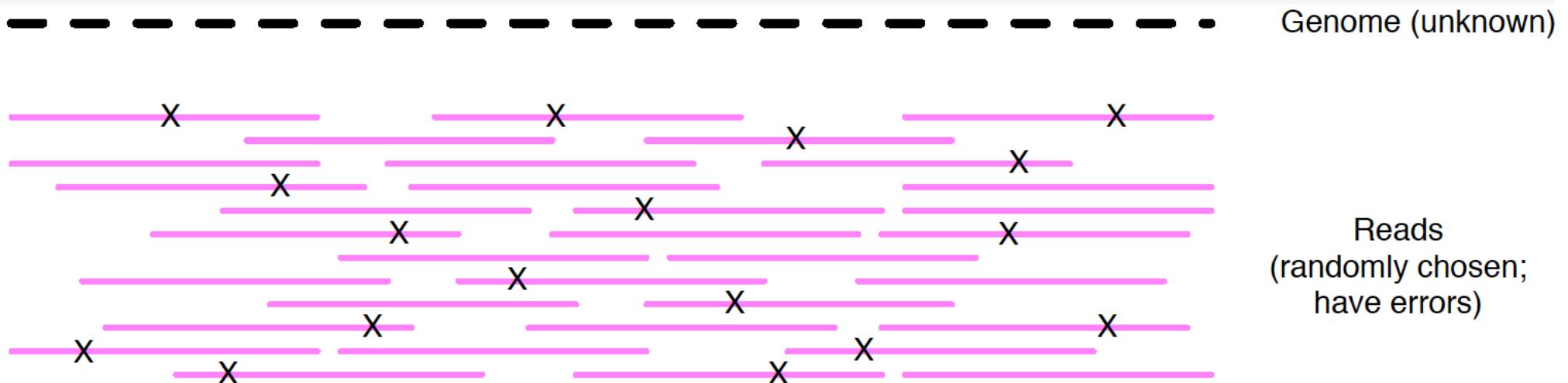
Fragments with the biotin labels are captured with streptavidin, they are enriched, end-repaired, and ligated with adapters



These fragments can now be sequenced from both ends

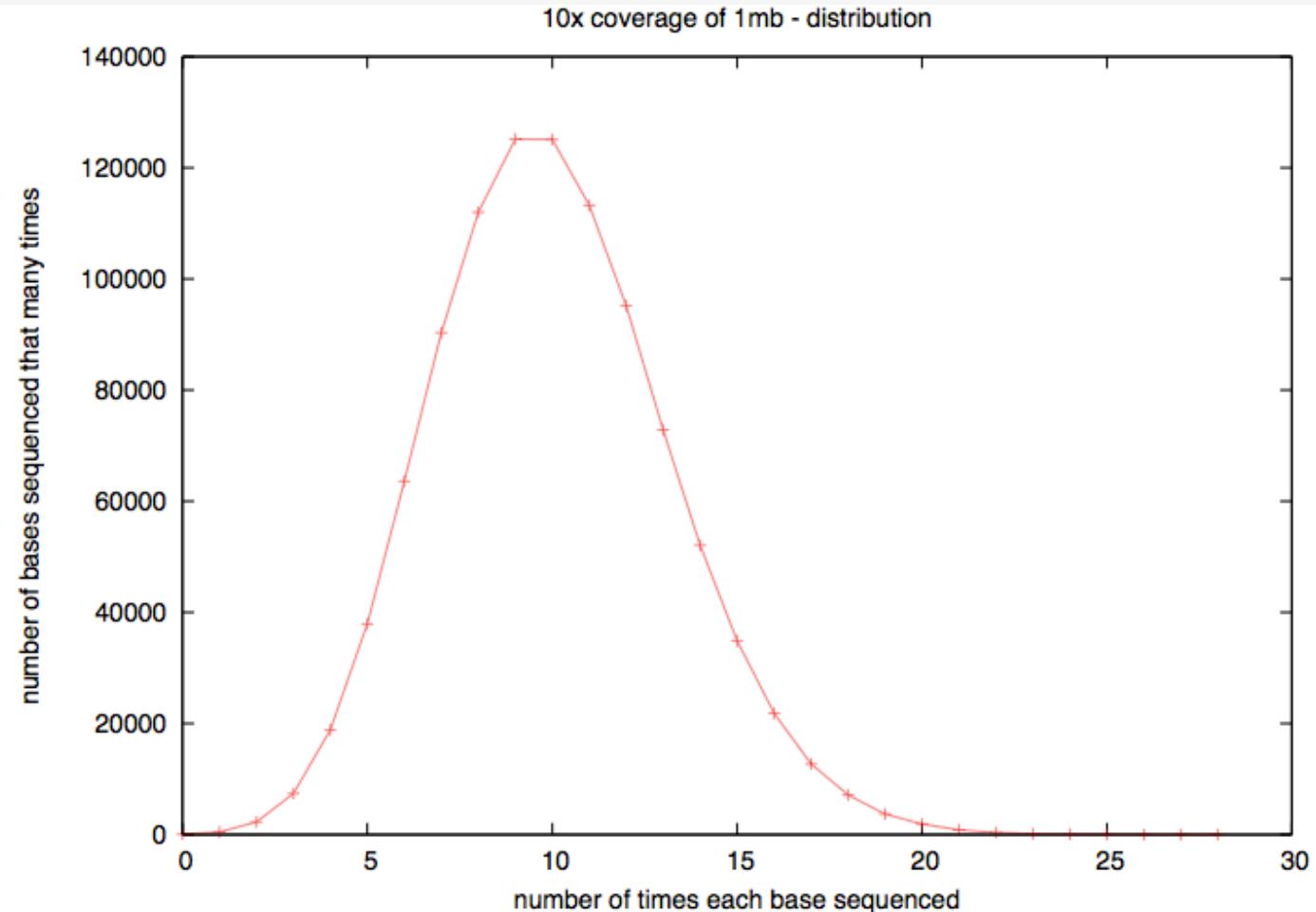


Coverage



“Coverage” is simply the average number of reads that overlap a base in the genome

Coverage



- “1x” doesn’t mean every DNA sequence is read once.
- It means that, if sampling were *systematic*, it would be.
- Sampling isn’t systematic, it’s random!
- To have an idea on the coverage you need for your experiment you can for example refer to other protocols:

[http://genome.ucsc.edu/ENCODE/protocols/dataStandards/
sequencing coverage calculator](http://genome.ucsc.edu/ENCODE/protocols/dataStandards/sequencing_coverage_calculator)
[Assembly coverage](#)

What you should worry about...

- How to reach your biological question? Or close enough
- Which technology or combination?
- Starting material & library preparation
- Data analysis: parameters, statistics, ...
- Rapid change of the field