



# Next Generation Sequencing 101

Molecular and Genomics Informatics Core

By Alexander Lemenze, PhD

# Our Journey Today

## MEETING AGENDA

### History of Sequencing

Sanger Sequencing

Next Generation  
Sequencing

### Illumina

Library Construct

Flow cell binding/Bridge  
amplification

SBS

Operational Machines

### Extended NGS

BGI/MGI

PacBio

Oxford Nanopore

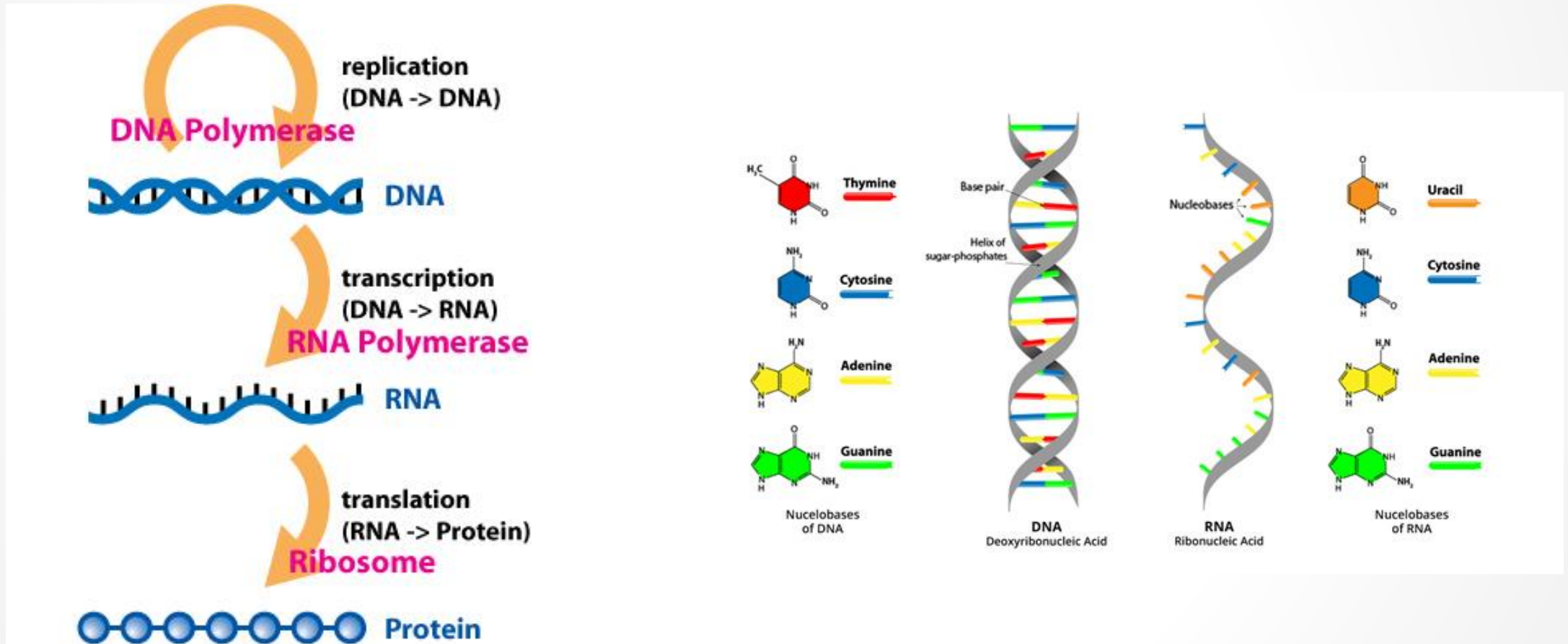
Single Cell Approaches

An abstract graphic in the top right corner of the slide. It consists of a complex network of small, dark blue circular nodes connected by thin, light blue lines. The nodes are arranged in a way that suggests a hierarchical or interconnected structure, with some nodes having many connections and others having fewer. The overall shape of the network is roughly triangular, pointing towards the top right corner.

# History of Sequencing



# Central Dogma of Molecular Biology

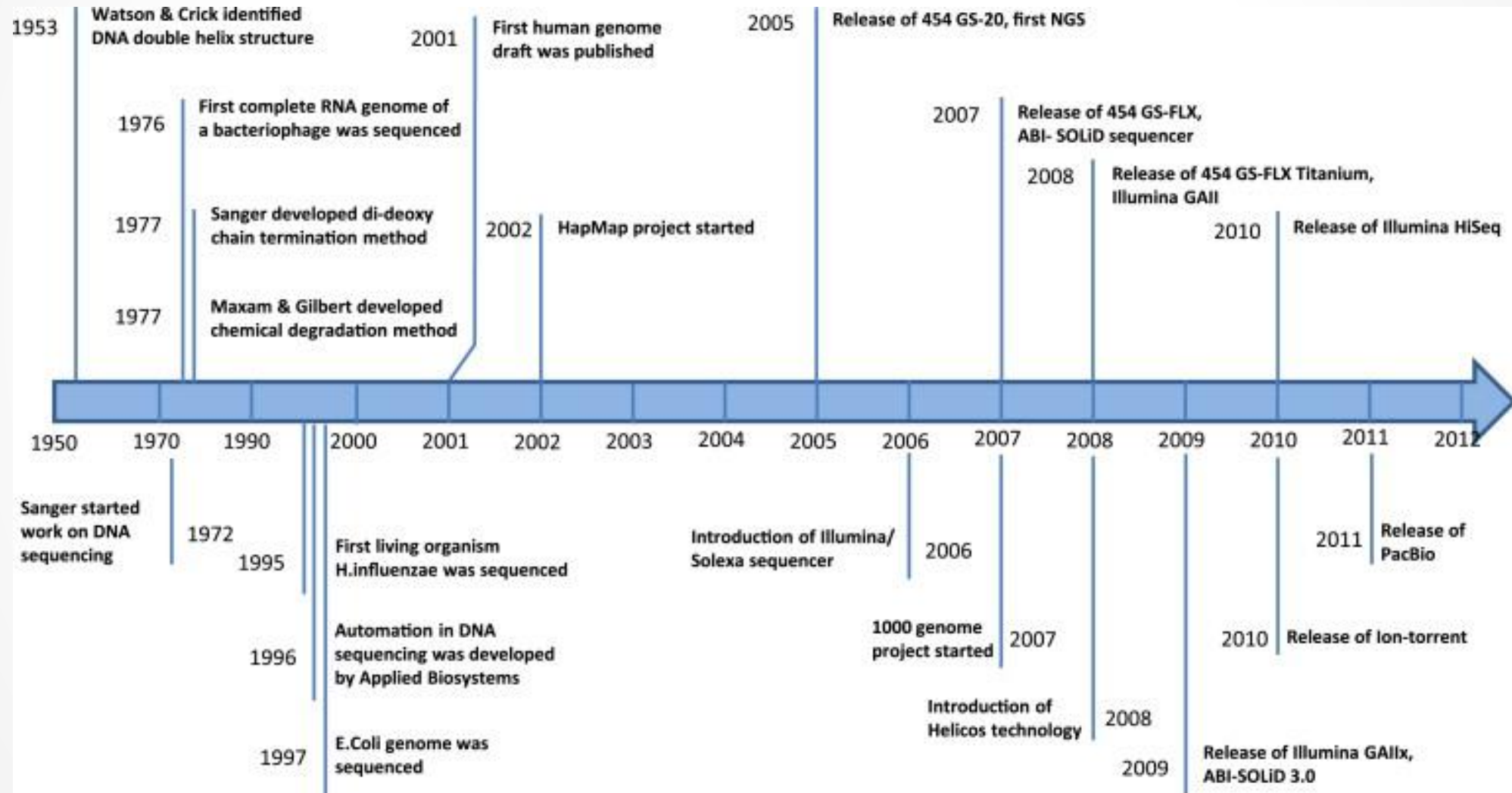


- Genetic mutations that lead to disease
- Response to drugs
- Cancer targeting

- Antibiotic resistance
- Evolution

- Infestation resistance for GMO
- Crop yield

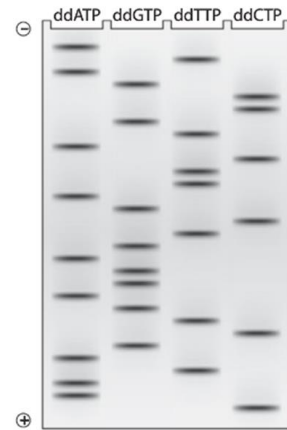
# Timeline for Sequencing



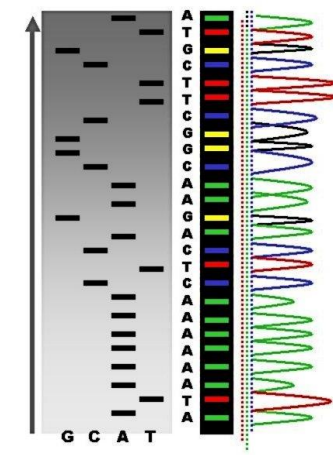
# First technologies for sequencing

## What becomes Sanger Sequencing

Gel sequencing with  
radioactive  
nucleotides

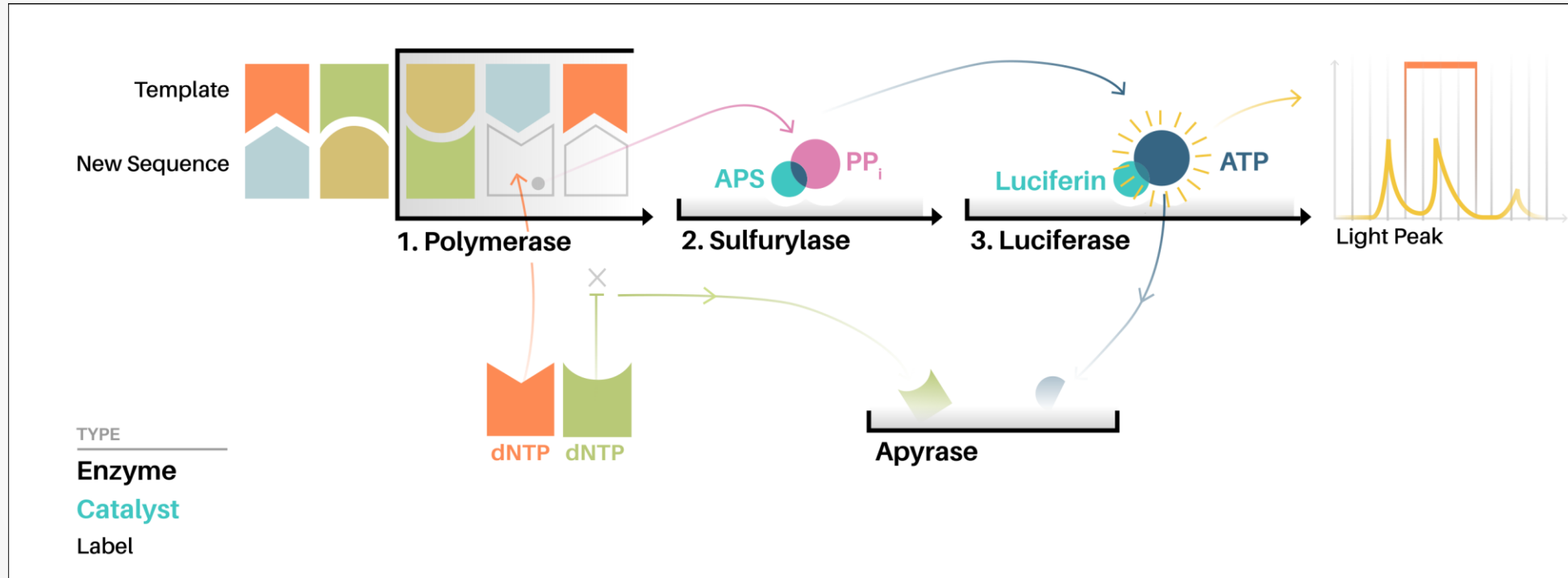


Automated with  
fluorescent  
nucleotides



# The start of NGS

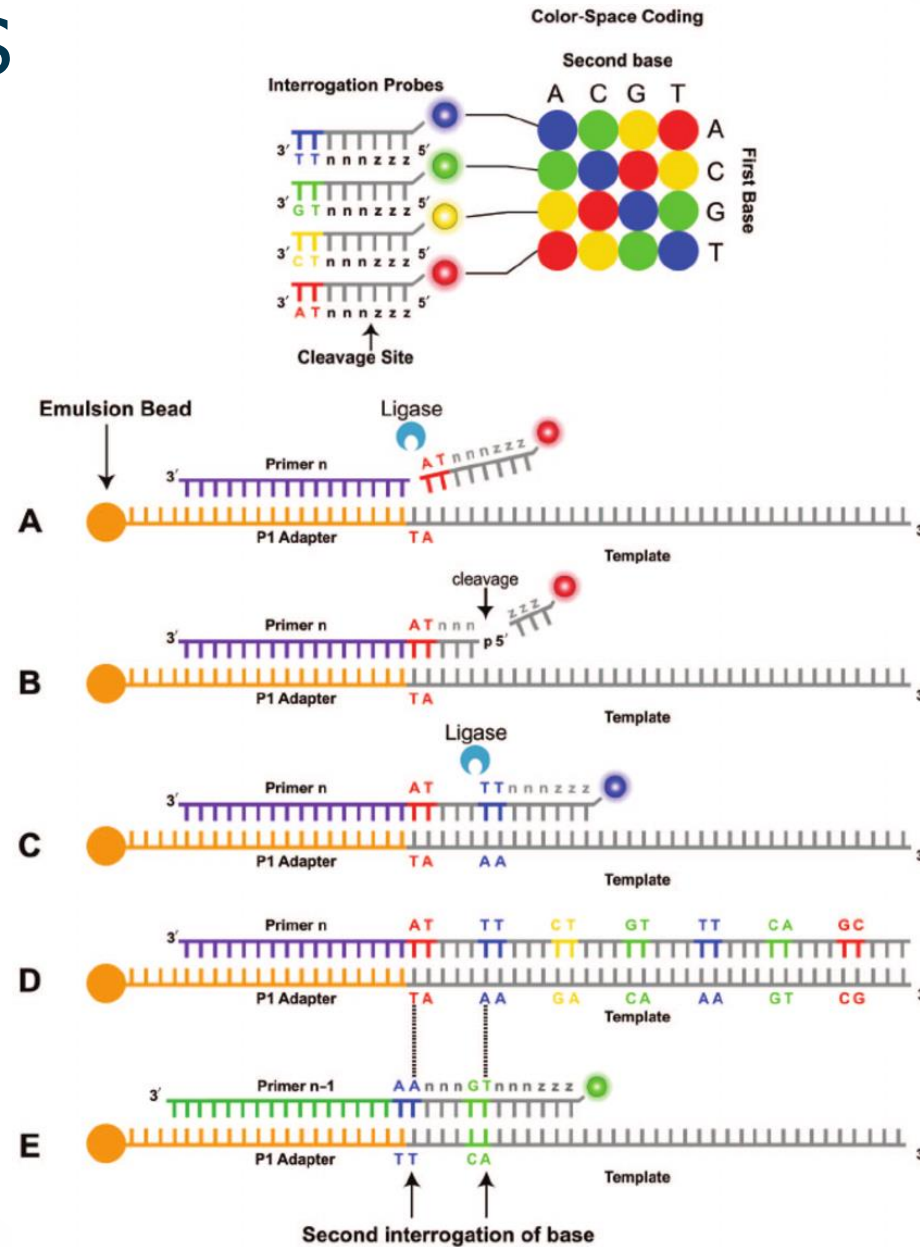
## Pyrosequencing





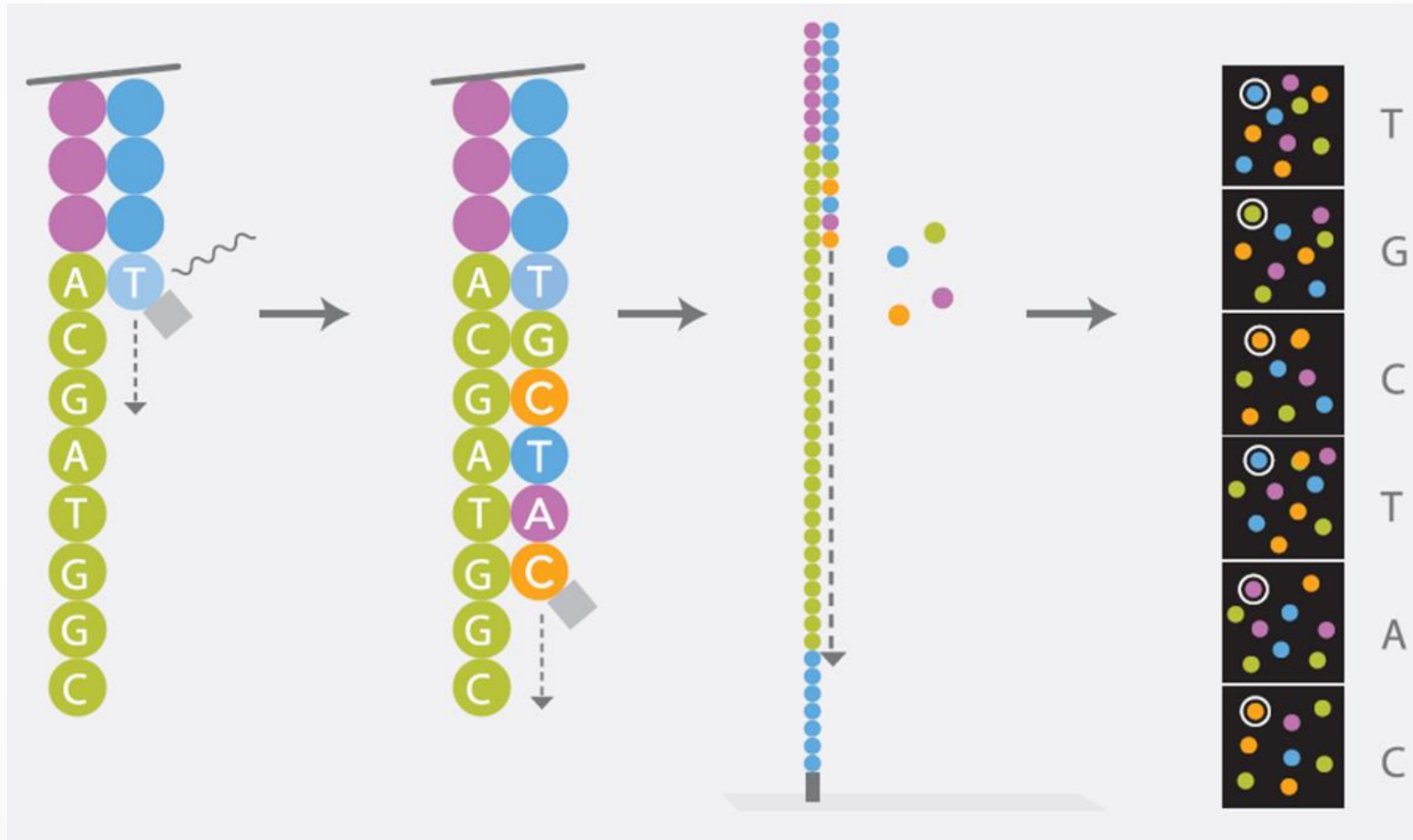
# The start of NGS

## Sequencing by Ligation

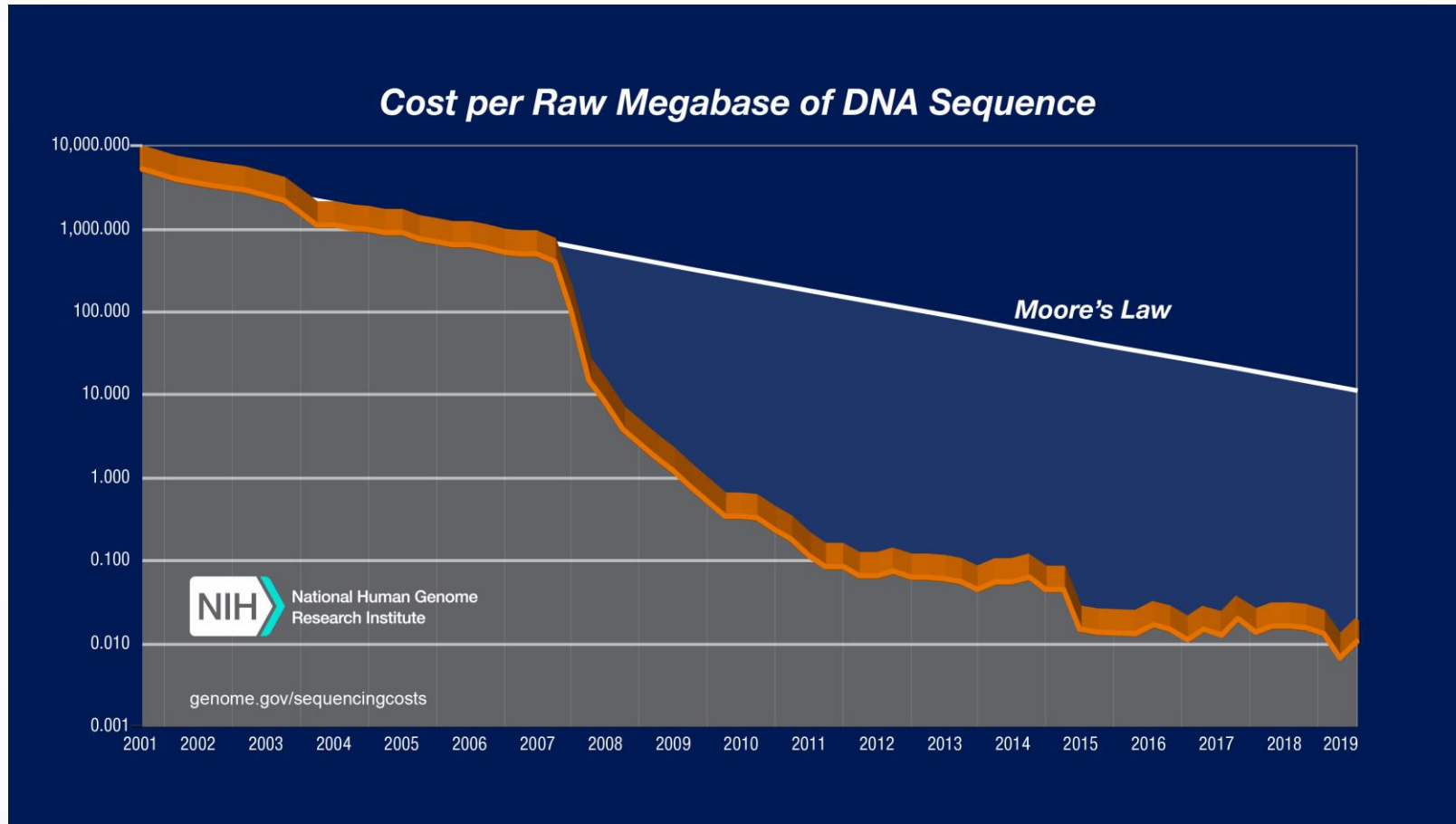


# The start of NGS

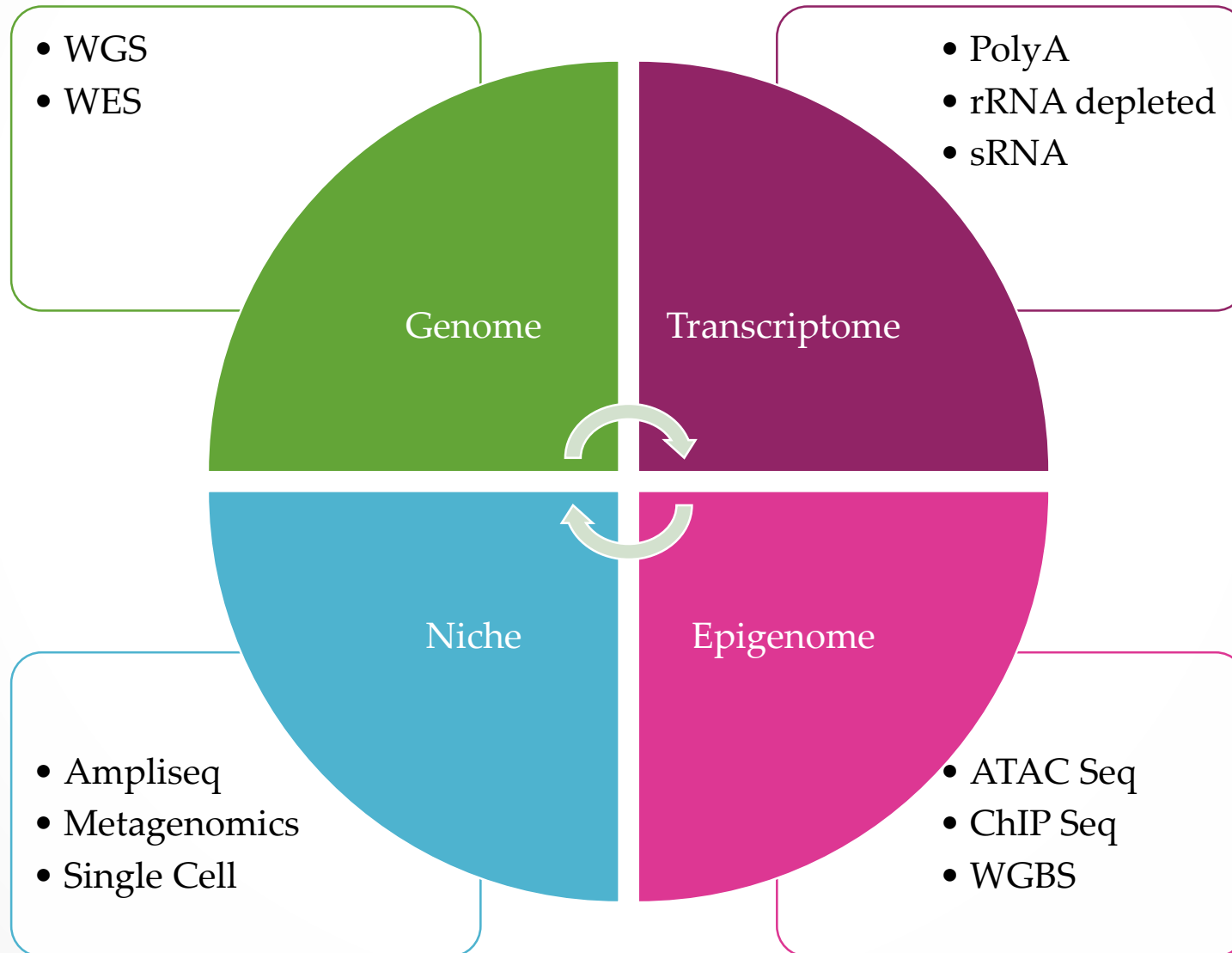
## Sequencing by Synthesis



# Cost over time



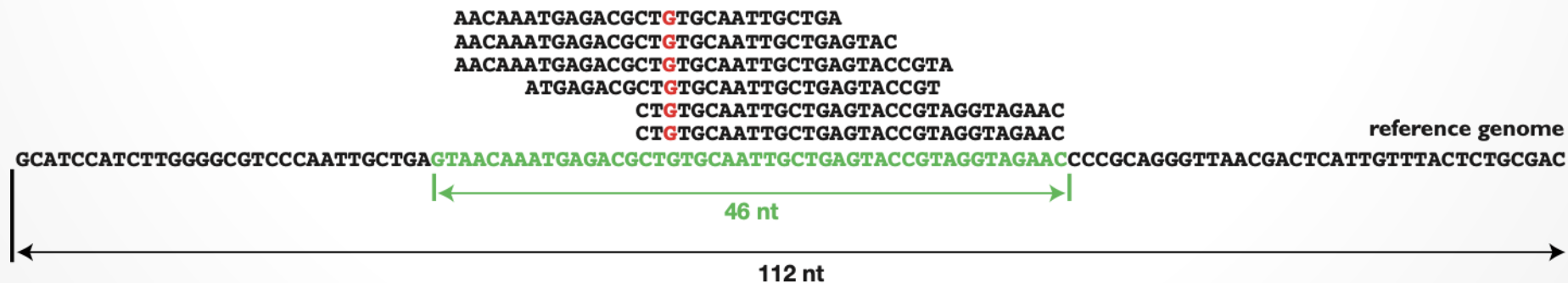
# Applications and Flavors





# Common Terms and Calculations

- Reads
  - The individual output string from a sequencer. Essentially the calls of “ATCG...”
- Coverage
  - How many times over a base has been sequenced. The more times a base has been sequenced, the more confident you can be in it
  - IE 1X coverage means you have each base in your sample once, whereas 30X coverage means you have each base 30 times.



# Calculating Coverage and Depth needed

- Coverage= Base pairs sequenced/ Genome size
  - You have the human genome (3Gb) and you have 350M reads of 150bp size
    - $Y = (350000000 * 150) / 3000000000$
    - Y= 17.5X coverage

Common Sequencing Target Depths	
WGS De Novo assembly	>100X coverage
WGS Variant Analysis	>30X coverage
RNA Seq- Eukaryotic	>20-30M reads/sample
RNA Seq- Prokaryotic	>1-5M reads/sample

An abstract graphic in the top right corner of the slide. It consists of a complex network of small, dark blue circular nodes connected by thin, light blue lines. The nodes are arranged in a way that suggests a hierarchical or interconnected structure, with some nodes having many connections and others having fewer. The overall shape of the network is roughly triangular, pointing towards the top right corner.

# Illumina Sequencing

# Library Construction

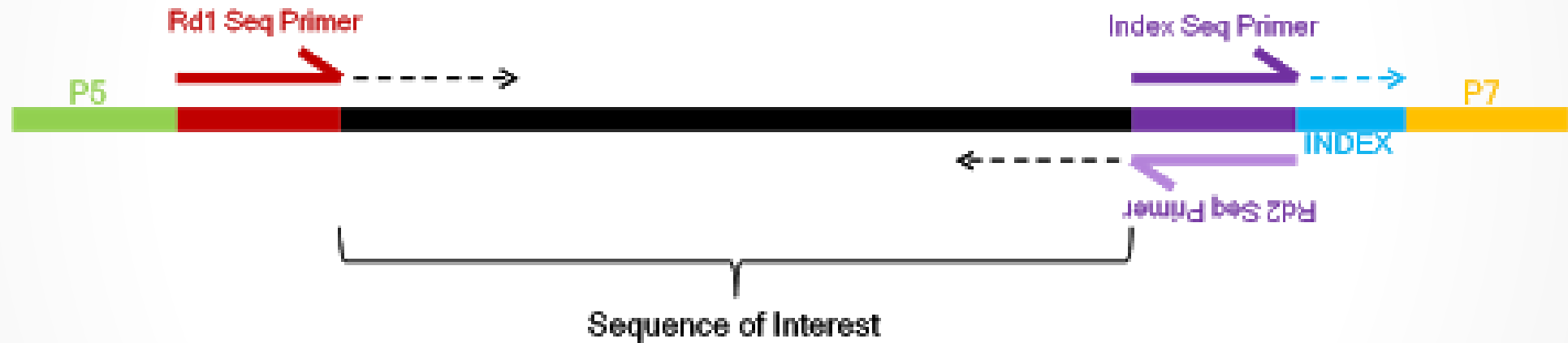
Getting your fragment of interest



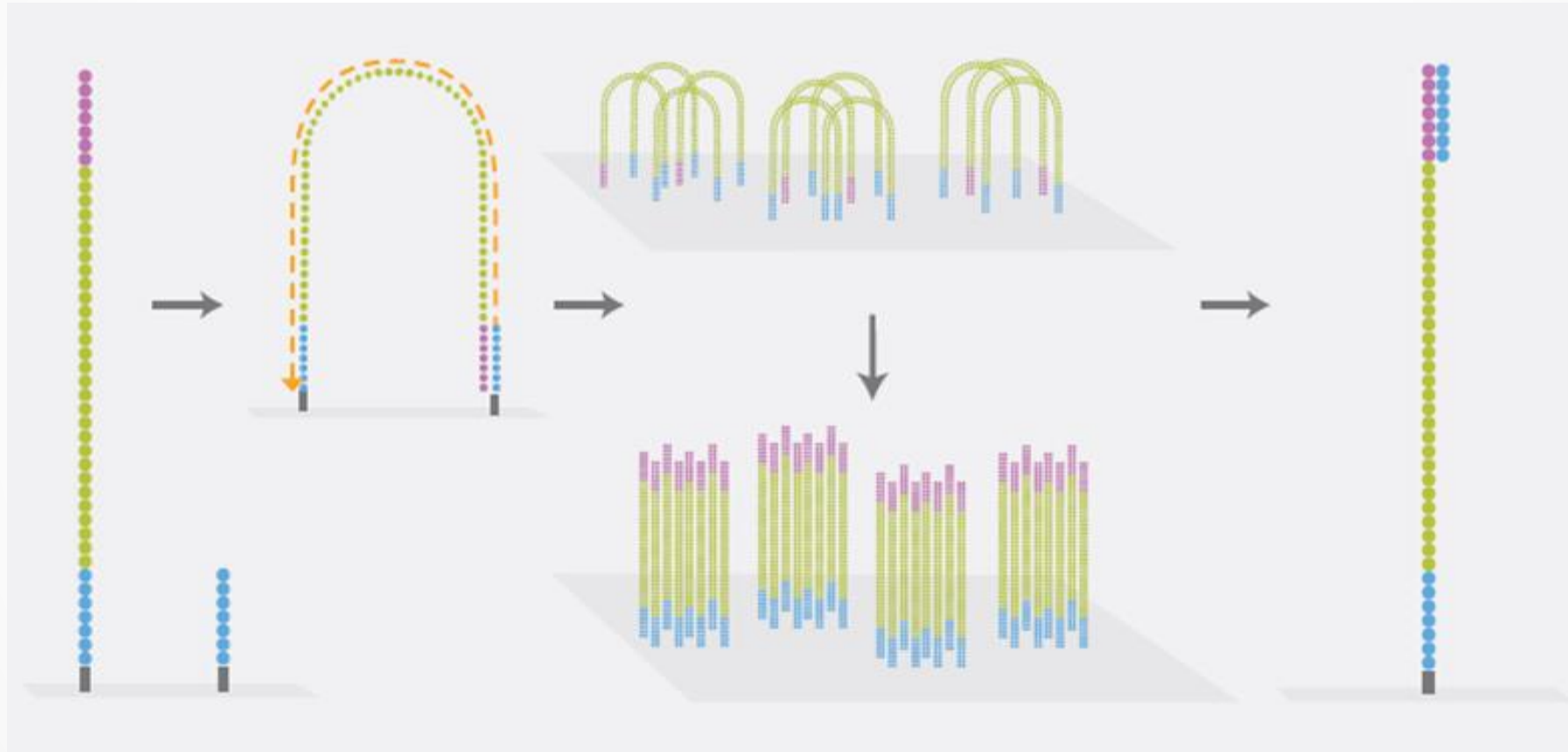


# Library Construction

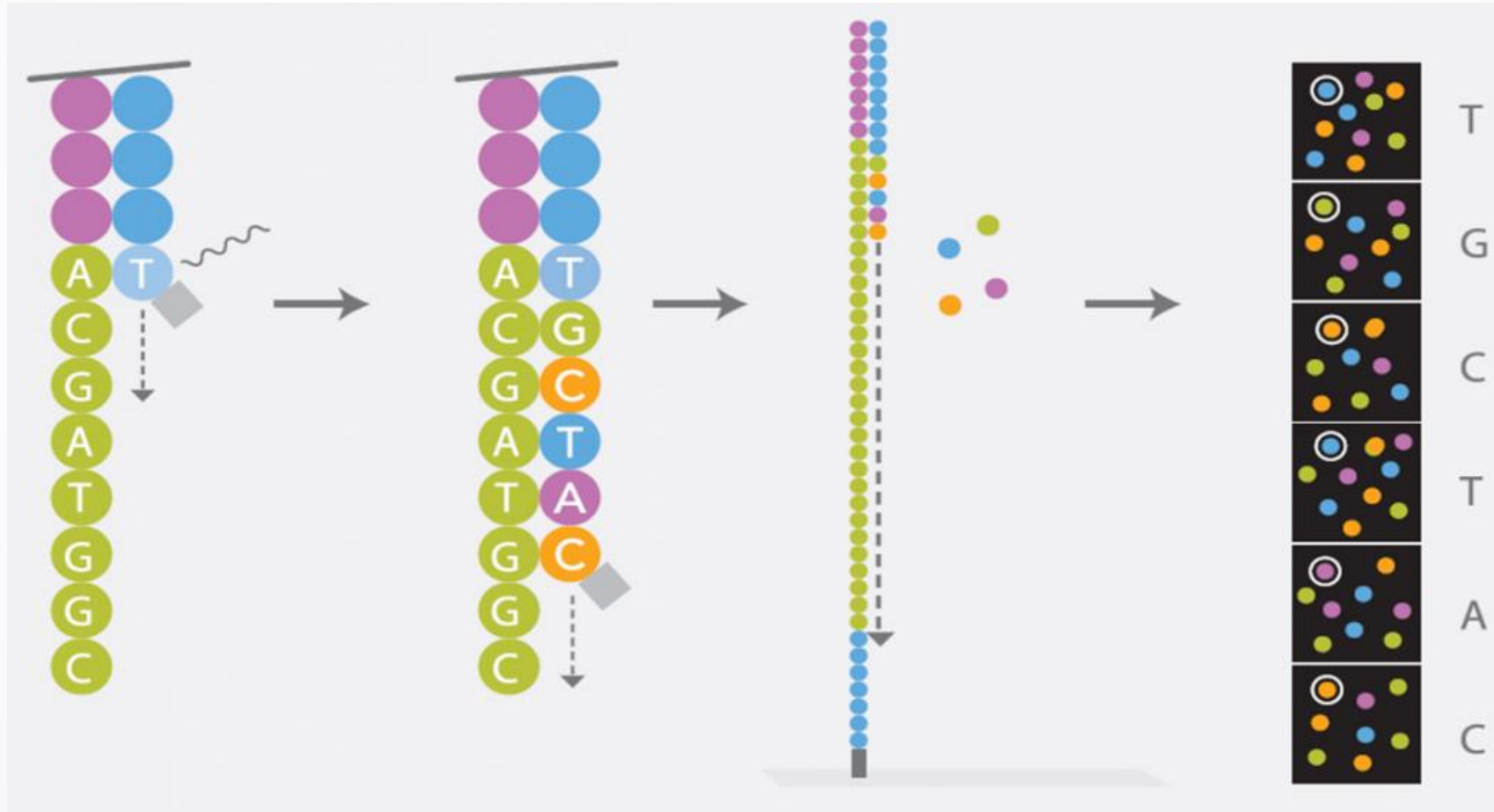
Example final construct



# Flow cell binding and Bridge amplification

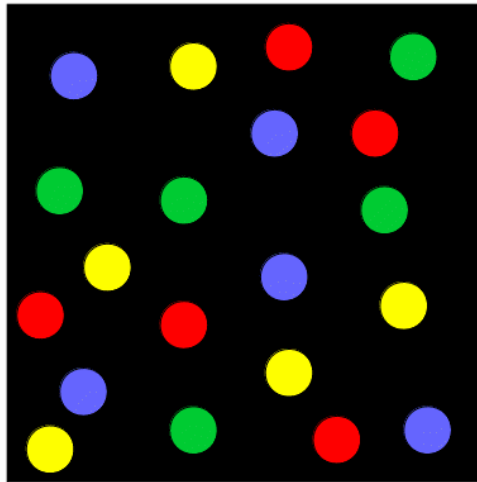


# Sequencing By Synthesis



## 2 channel vs 4 channel

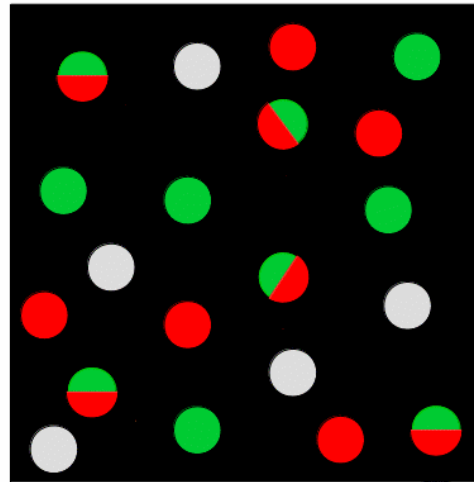
4-Channel system (4 dyes)



4 Filter channels



2-Channel system (2 dyes)



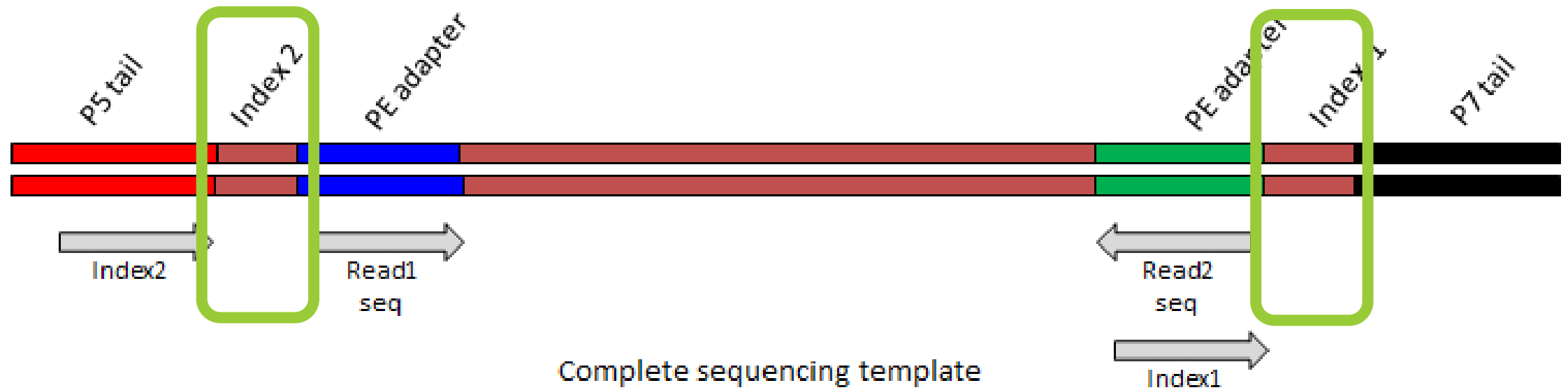
2 Filter channels



\*No detected dye



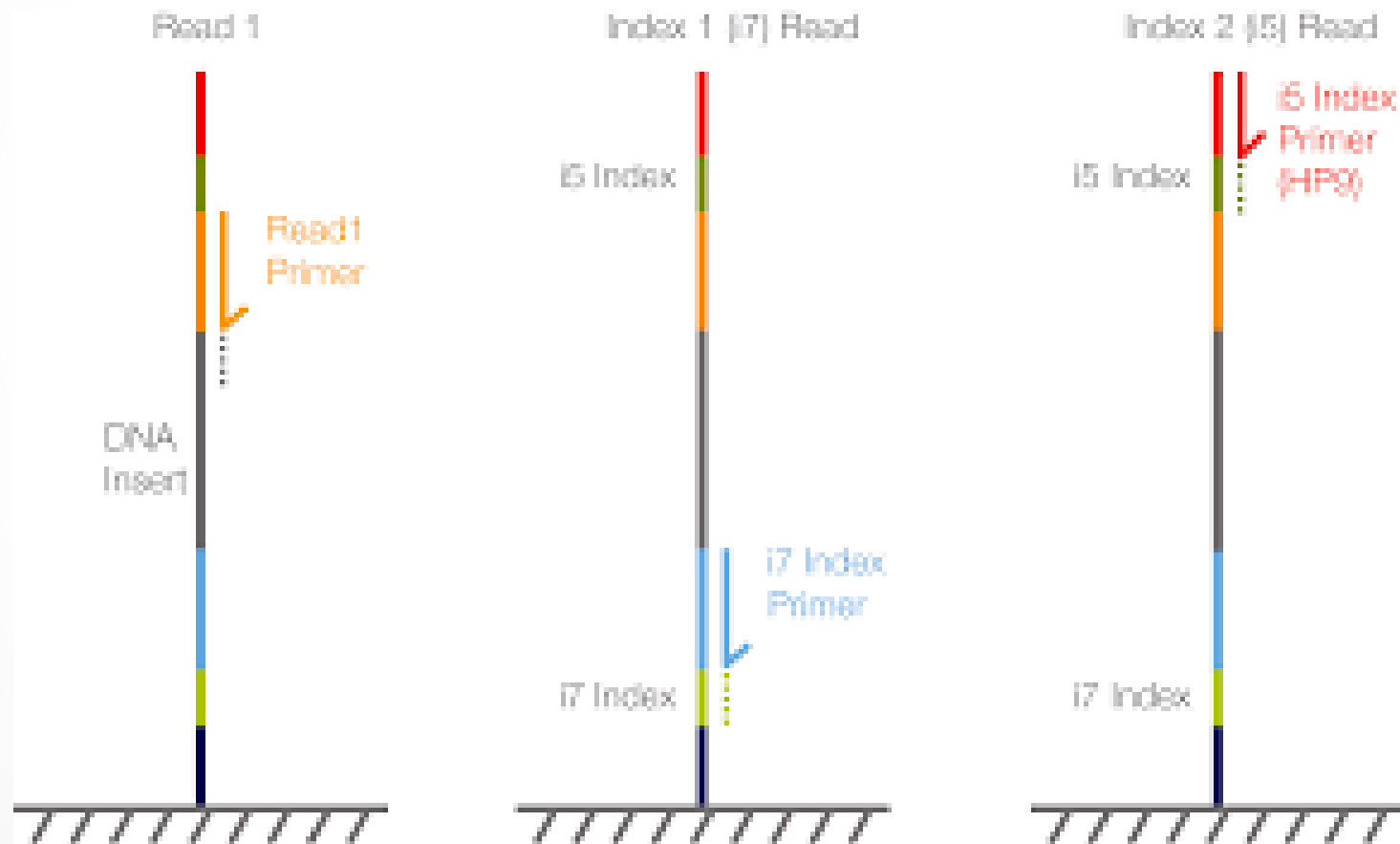
# Multiplexing



Pekka Ellonen FIMM 2013

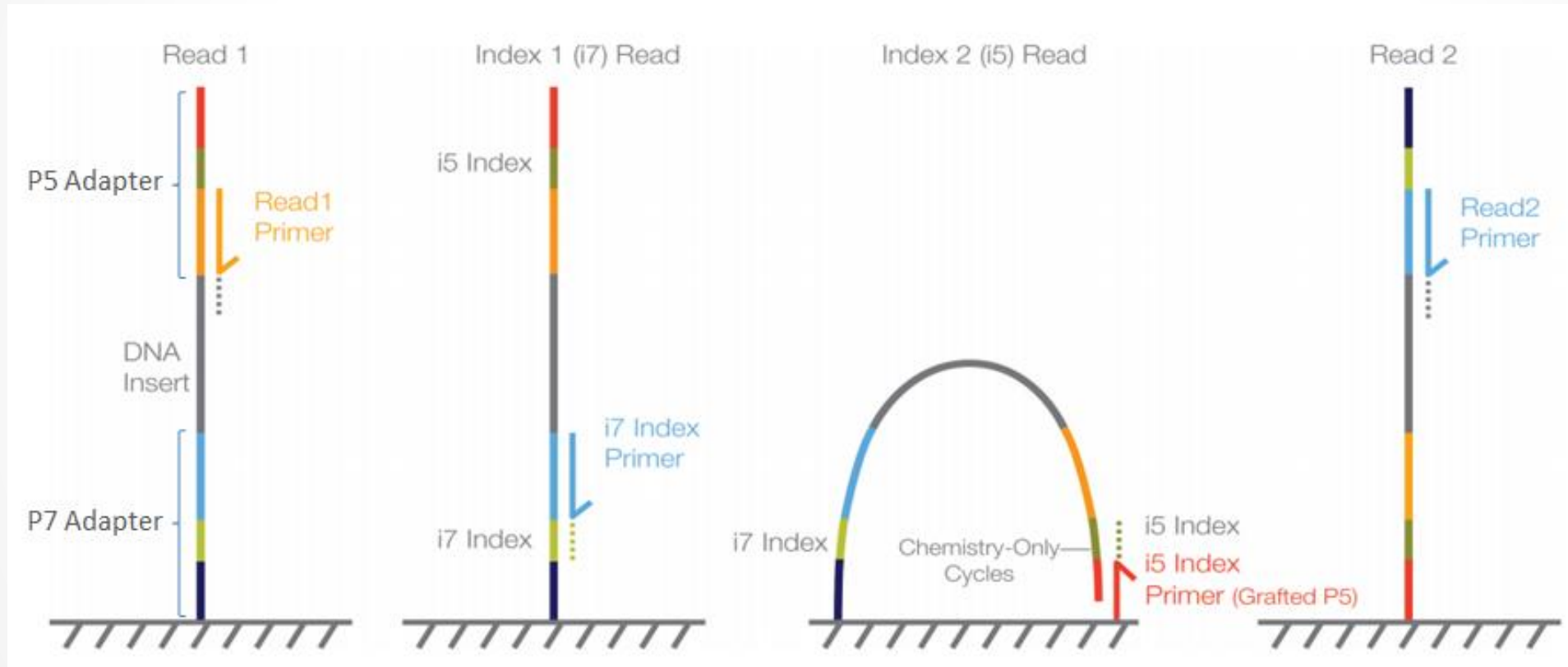
# Order of Sequencing

## Single end sequencing



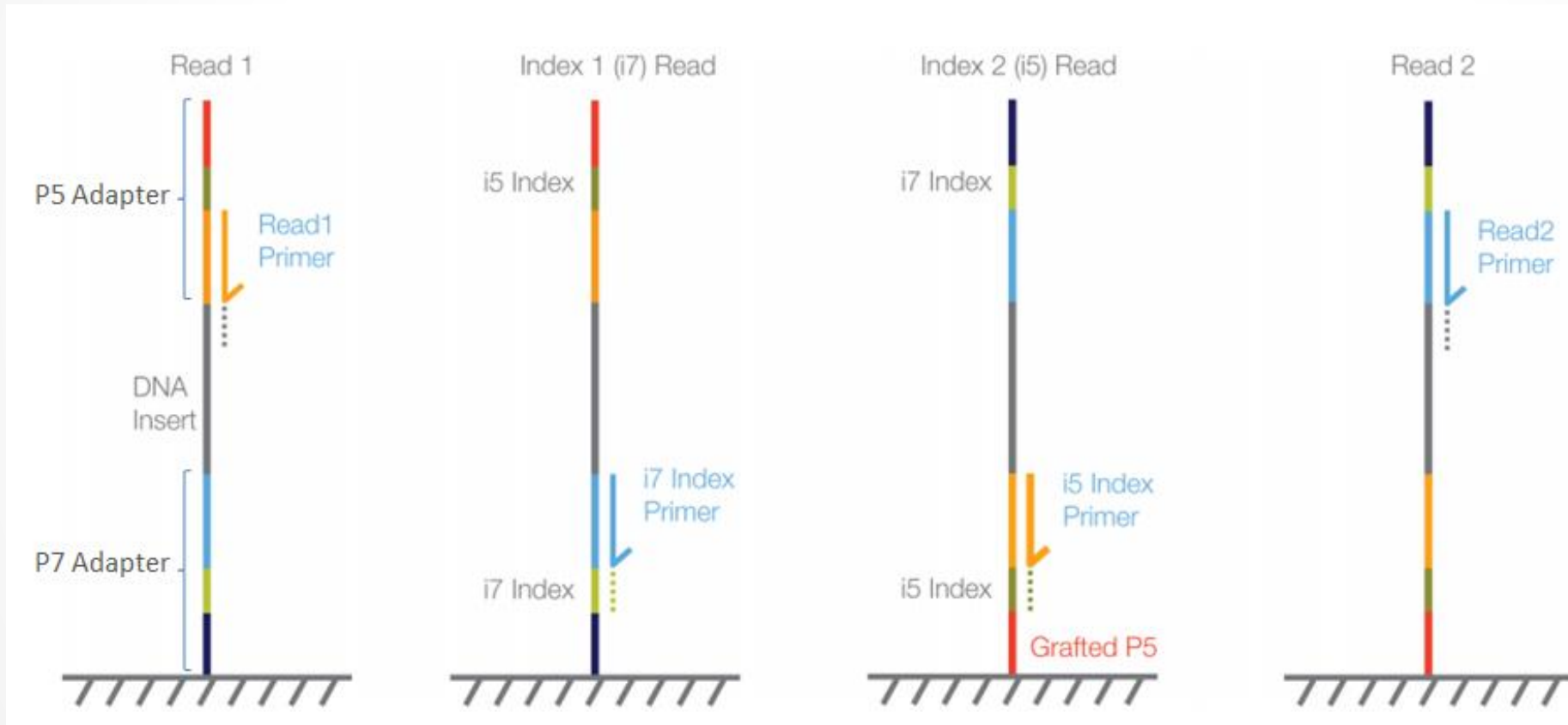
# Order of Sequencing

Paired end sequencing- for NovaSeq, MiSeq, HiSeq 2500



# Order of Sequencing

Paired end sequencing for HiSeq X, HiSeq 4000, NextSeq





# Illumina Sequencers and Capacity



# Illumina Sequencers and Capacity

Platform	Number of Cycles	Flow cell size	Data output (reads)	Data output (Gb)
MiSeq V2	600 cycle	1 flow cell with no divisible units	~10M paired end reads/flow cell	~4.5-7.5 Gb/flowcell
MiSeq V3	600 cycle	1 flow cell with no divisible units	~20M paired end reads/flow cell	~13 Gb/flowcell
NextSeq	300 cycle	1 flow cell with no divisible units	~400M paired end reads/flow cell	~XX Gb/flowcell
HiSeq X	300 cycle	1 flow cell with 8 divisible lanes	~350M paired end reads/flow cell	~106 Gb/flowcell
HiSeq 2500	500 cycle	1 flow cell with 2 divisible lanes	~150M paired end reads/flow cell	~75 Gb/flowcell
NovaSeq SP	300 cycle or 500 cycle	1 flow cell with no divisible lanes	~650M paired end reads/flow cell	~200-350 Gb/flowcell
NovaSeq S1	300 cycle	1 flow cell with no divisible lanes	~1.3B paired end reads/flow cell	~400 Gb/flowcell
NovaSeq S2	300 cycle	1 flow cell with 2 divisible lanes	~3.3B paired end reads/flow cell	~1 Tb/flow cell
NovaSeq S4	300 cycle	1 flow cell with 4 divisible lanes	~8B paired end reads/flow cell	~2.3 Tb/flow cell. ~600Gb/lane

# Extended NGS



# BGI/MGI

Beijing Genomics Institute



华大基因  
BGI

100%



# BGI Machines and outputs

Platform	Number of Cycles	Flow cell size	Data output (reads)	Data output (Gb)
BGISEQ-50	SE 50	2 lanes per flowcell	Up to 275M reads	Up to 225 Gb
BGISEQ-500	Up to PE 100	2 lanes per flowcell	Up to 1.3B reads	Up to 520 Gb
DNBSEQ-G50	Up to PE 100	1 lane per flowcell	Up to 300M reads	Up to 60 Gb
DNBSEQ-G400	Up to SE 400	2 or 4 lanes per flow cell	Up to 1.8B reads	Up to 1.4 Tb
DNBSEQ-T7	Up to PE 150	Up to 4 flow cells at a time	Up to 5B reads	Up to 6 Tb



# PacBio



*PacBio RS II System*



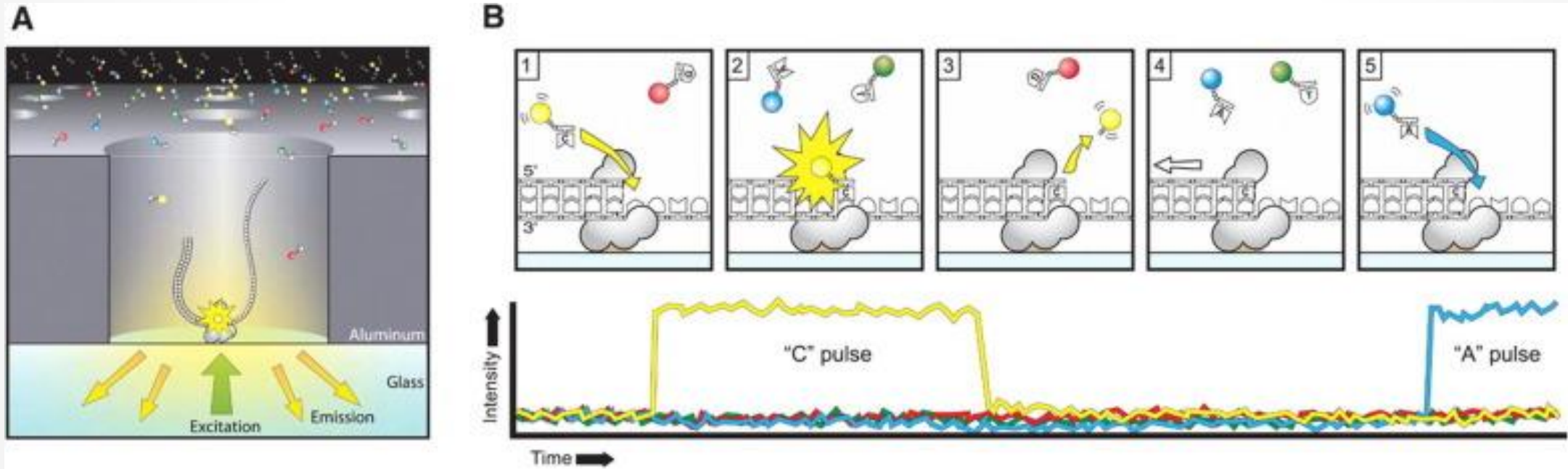
*Sequel System*



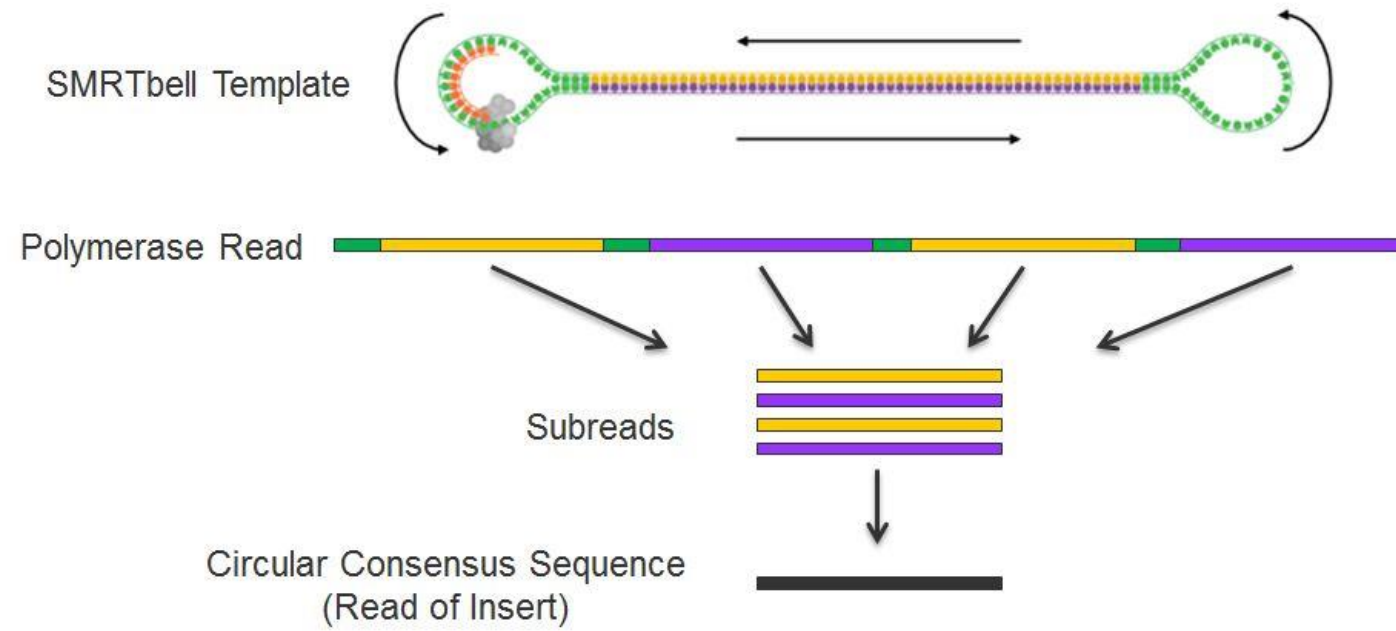
PACBIO®

# PacBio

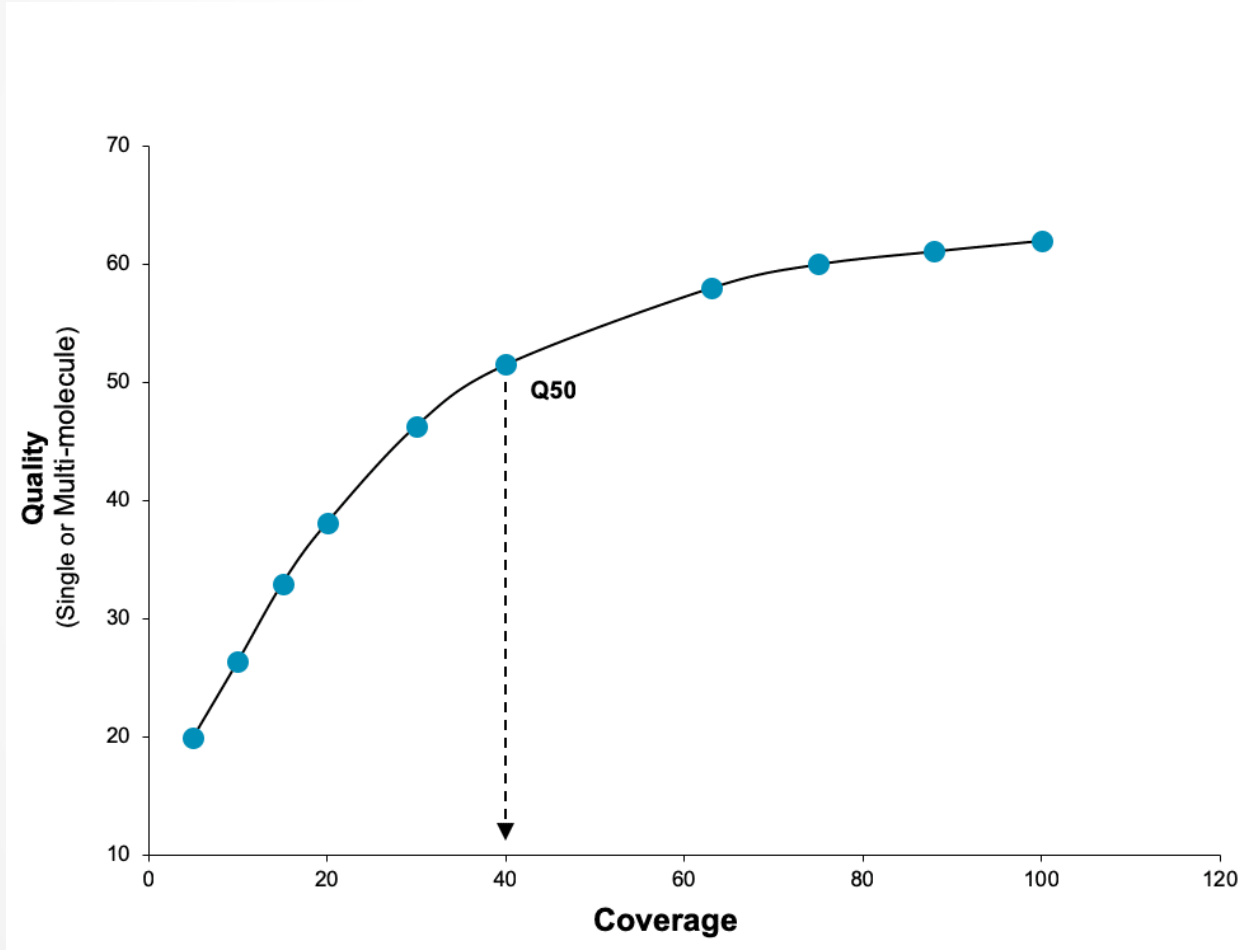
## SMRT Sequencing



# PacBio



# PacBio Error rate



11% Error rate but Washes out!

# PacBio

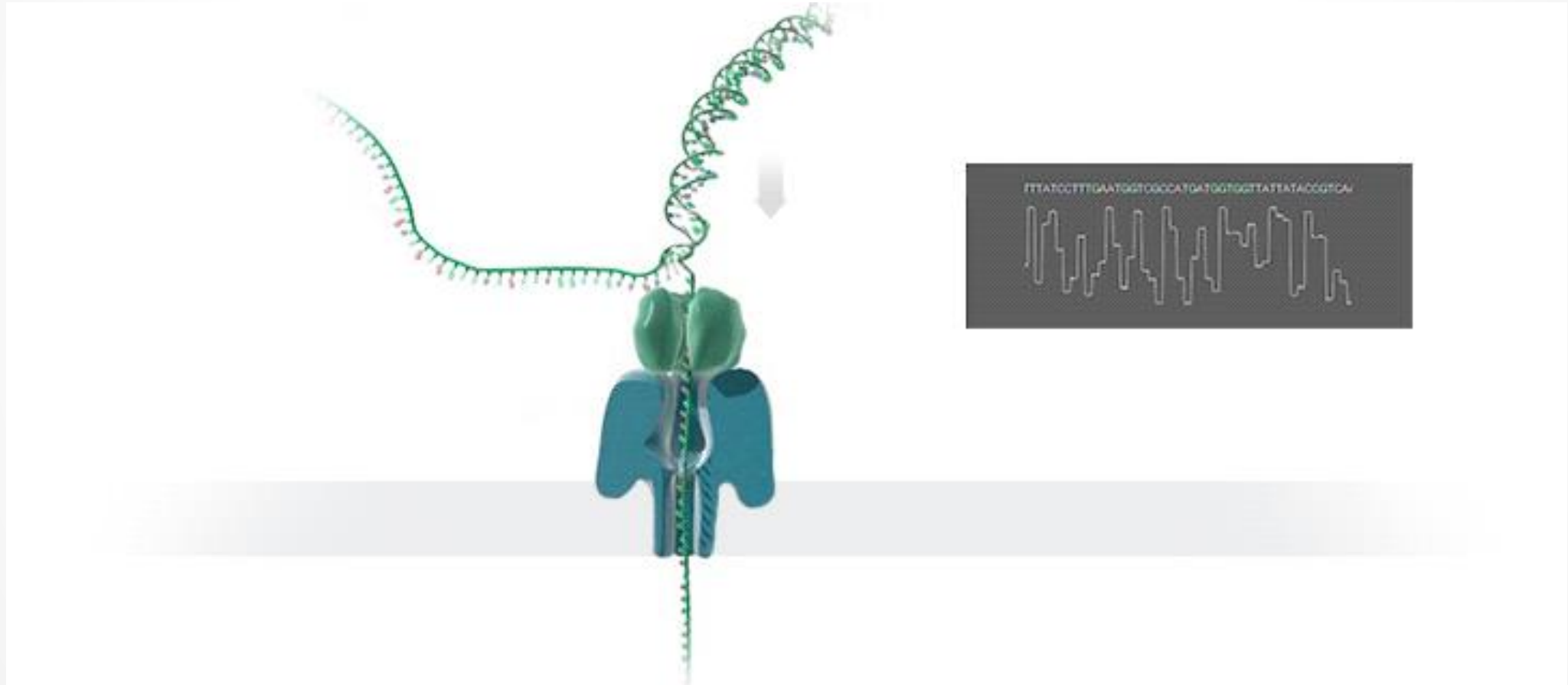
Platform	Read time	Flow cell size	Data output (reads)	Data output (Gb)
RS II	Up to 4 hours	1 flow cell with up to 16 SMRT cells	~55,000 reads	~1 Gb/SMRTcell
Sequel	Up to 20 hours	1 flow cell with up to 16 SMRT cells	~300,000-500,000 reads	~8-10 Gb/SMRTcell
Sequel II	Up to 30 hours	1 flow cell with up to 16 SMRT cells	~3-4M reads	~40-50 Gb/SMRTcell

# Oxford Nanopore





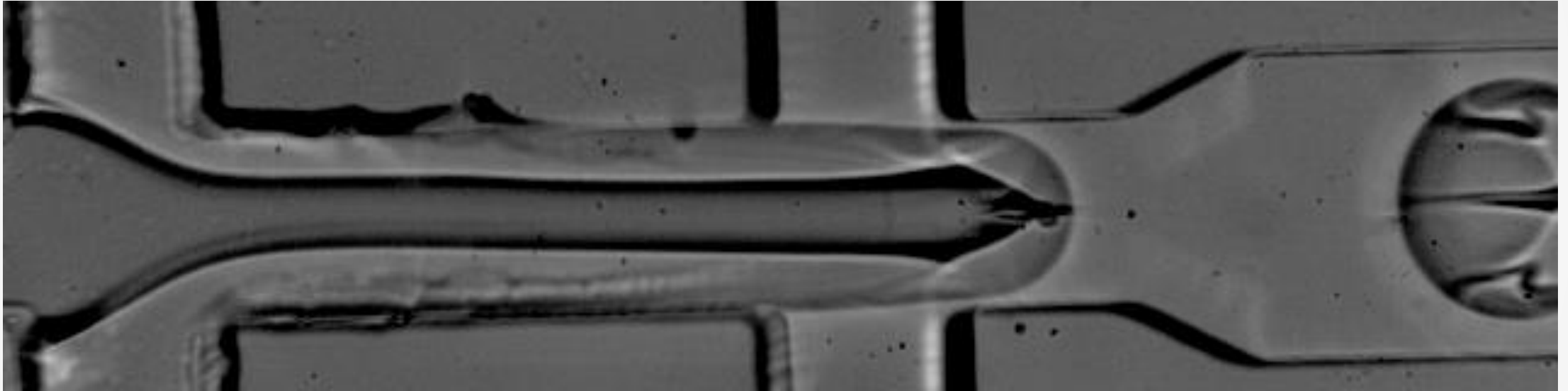
# Oxford Nanopore



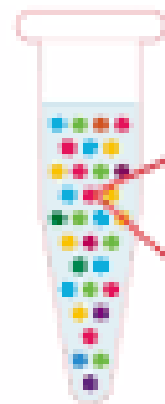
# Oxford Nanopore

<b>Platform</b>	<b>Number of channels per flow cell</b>	<b>Number of flow cells per device</b>	<b>Run time</b>	<b>Data output per device</b>
SmidgION	Not yet released	Not yet released	Not yet released	Not yet released
Flongle	126 channels	1 flowcell	Up to 16 hours	1-2 Gb
MinION Mk1b	512 channels	1 flowcell	Up to 48 hours	15-30 Gb
GridION Mk1	512 channels	5 flowcells	Up to 48 hours	75-150 Gb
PromethION 24	3000 channels	24 flowcells	Up to 72 hours	2.4-4.3 Tb
PromethION 48	3000 channels	24 flowcells	Up to 72 hours	4.8-8.6 Tb

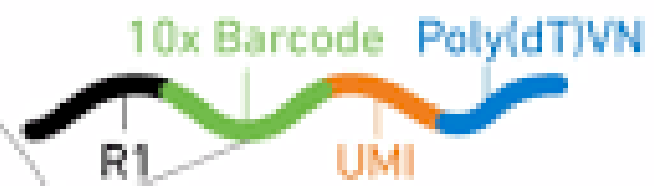
# Single Cell



Nombreuse GEM  
avec des Barcodes uniques.



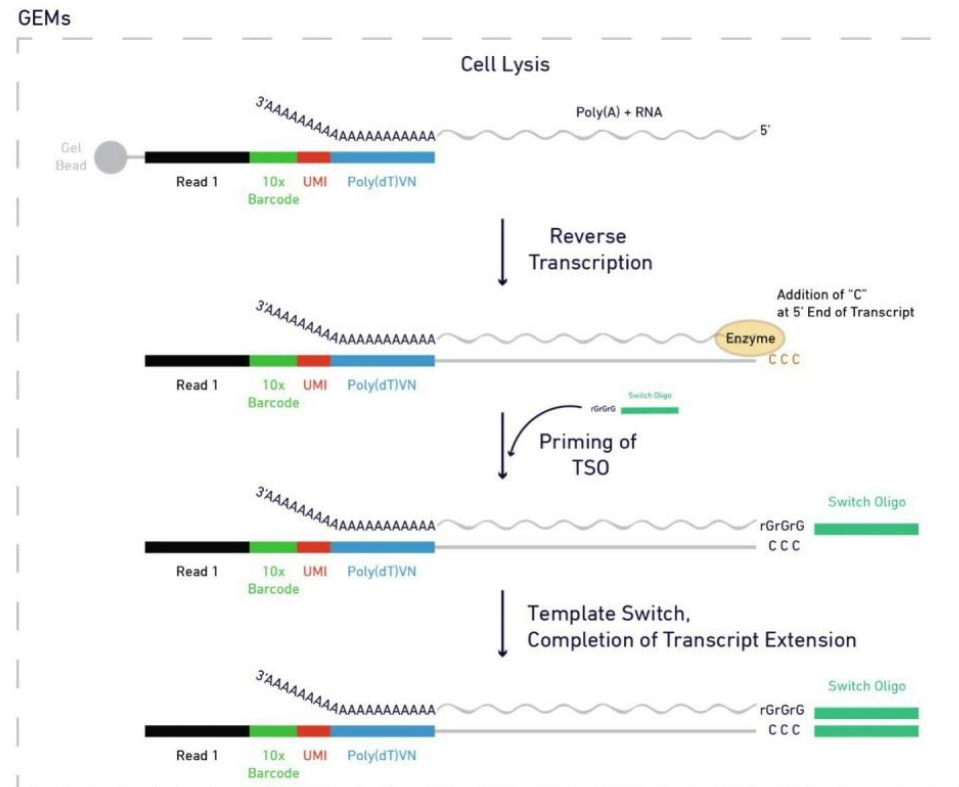
GEM



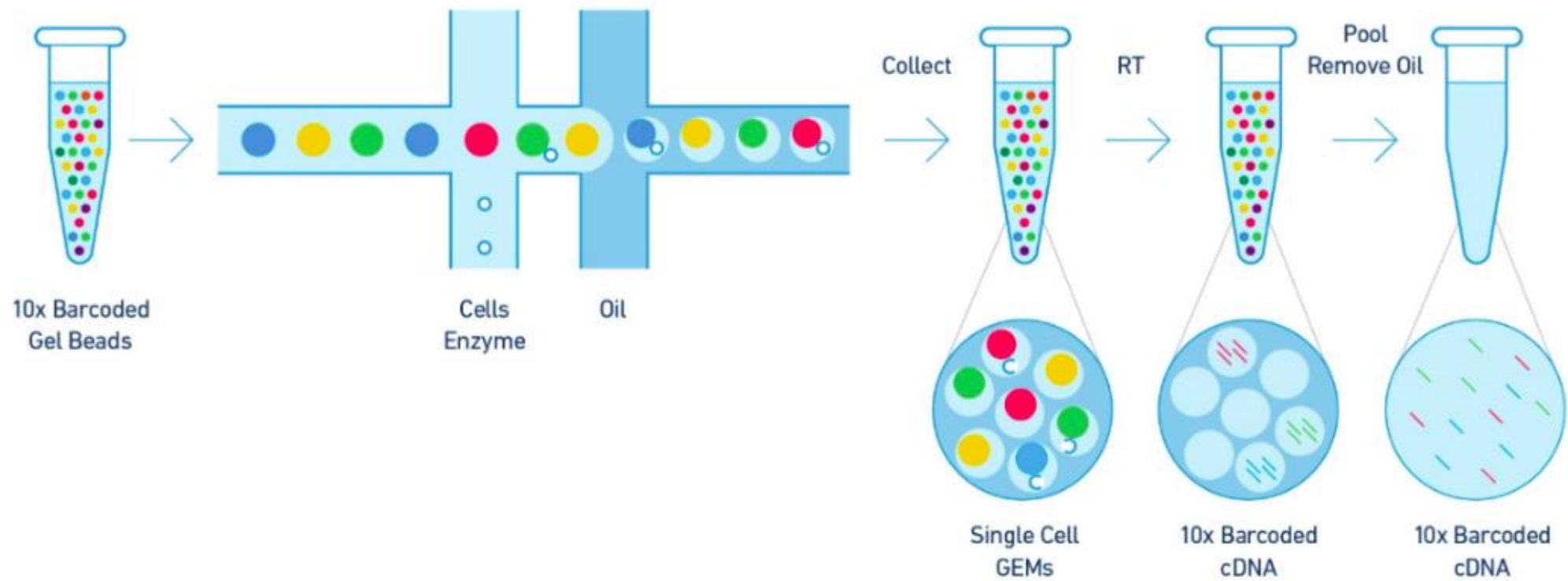
# Single Cell

## The Single Cell 3' v2 Assay Scheme

10X  
GENOMICS



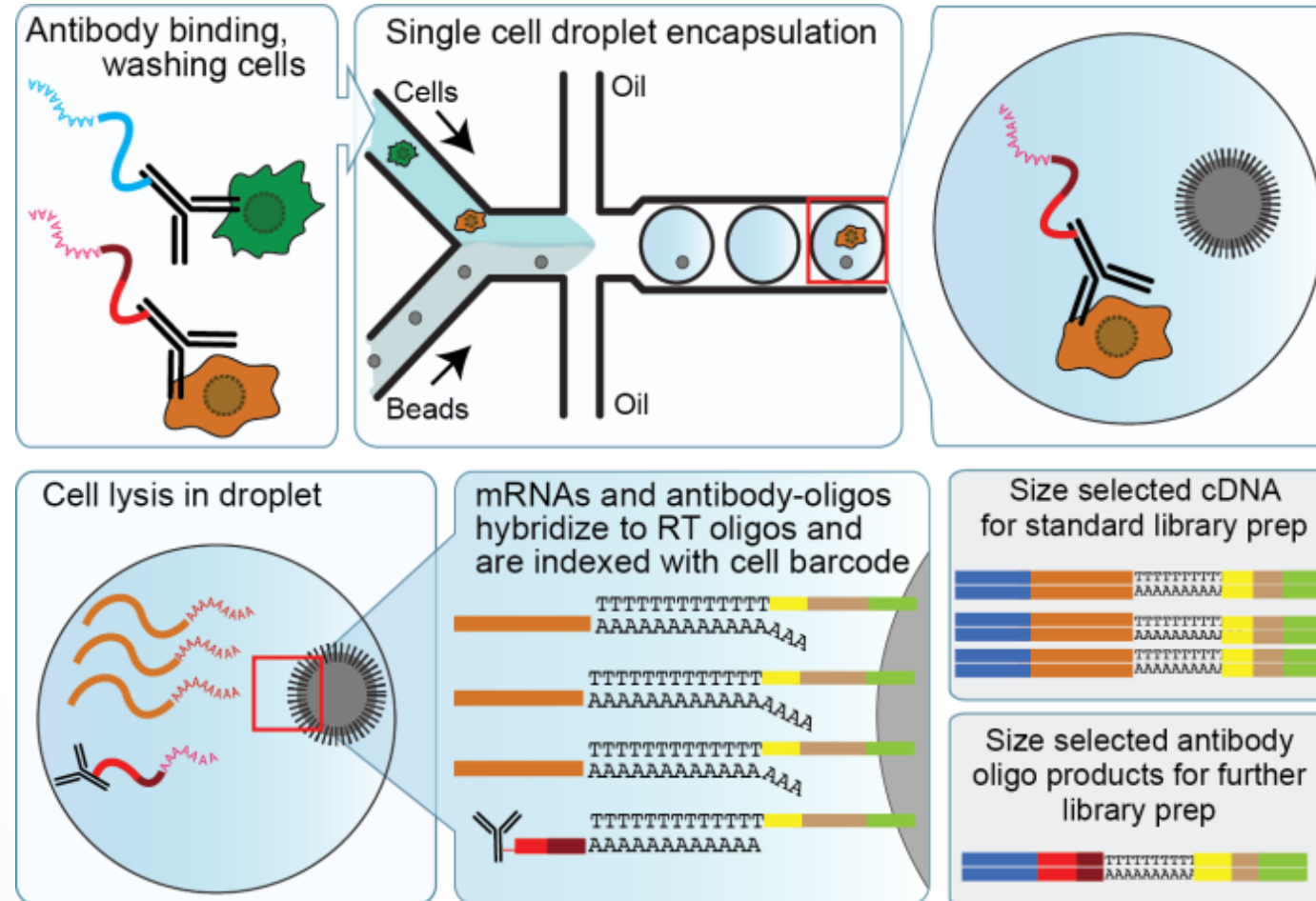
# Single Cell





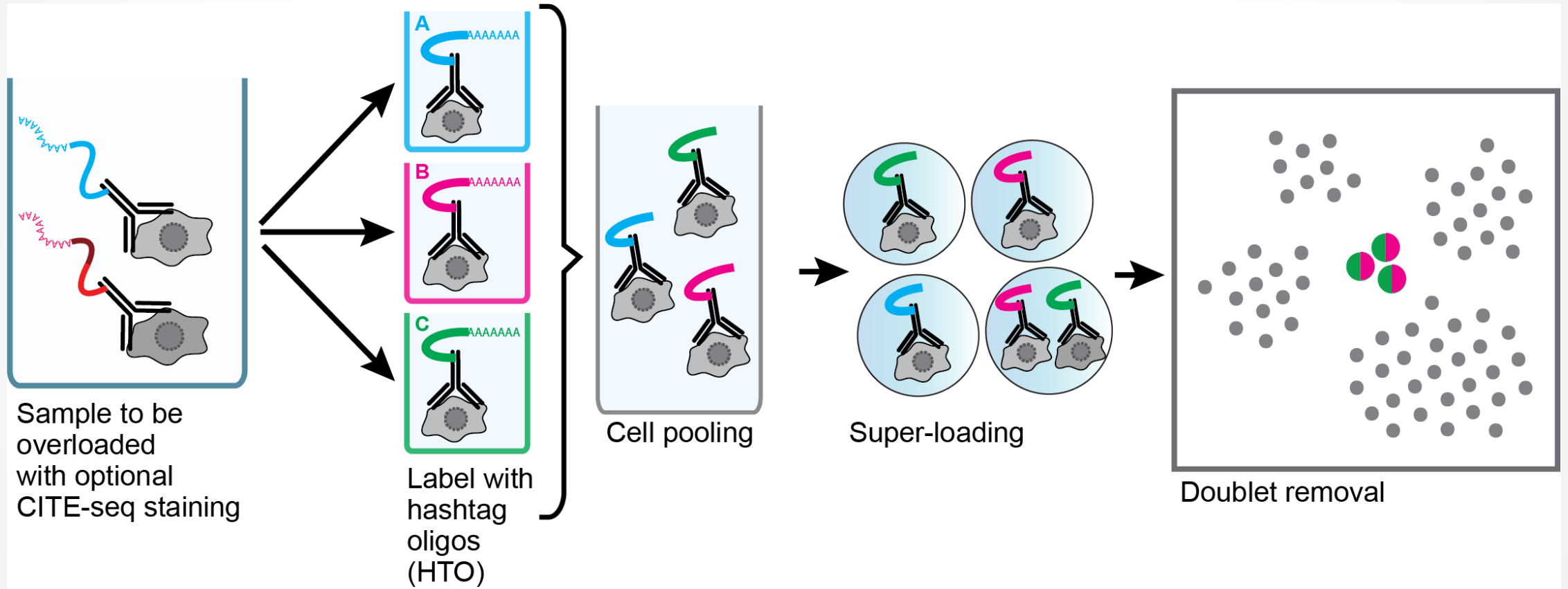
# Single Cell

CiteSeq/ReapSeq/FeatureBarcoding



# Single Cell

## HashTagging





THANK YOU