



NGS Technologies and Workflow

Dr Luria Founou
Head of Research, CEDBCAM-RI
Advisory Member, WHO FERG
Honorary Senior Lecturer, UKZN

Learning objectives

- Recognize the importance of the Human Genome Project (sequencing & databases)
- Recognize that sequencing greatly impacts today's research and its applications
- Summarize the evolution of sequencing technologies
- Describe the general workflow of NGS

Agenda

Introduction

Why is HT sequencing important?

From Sanger Sequencing to NGS

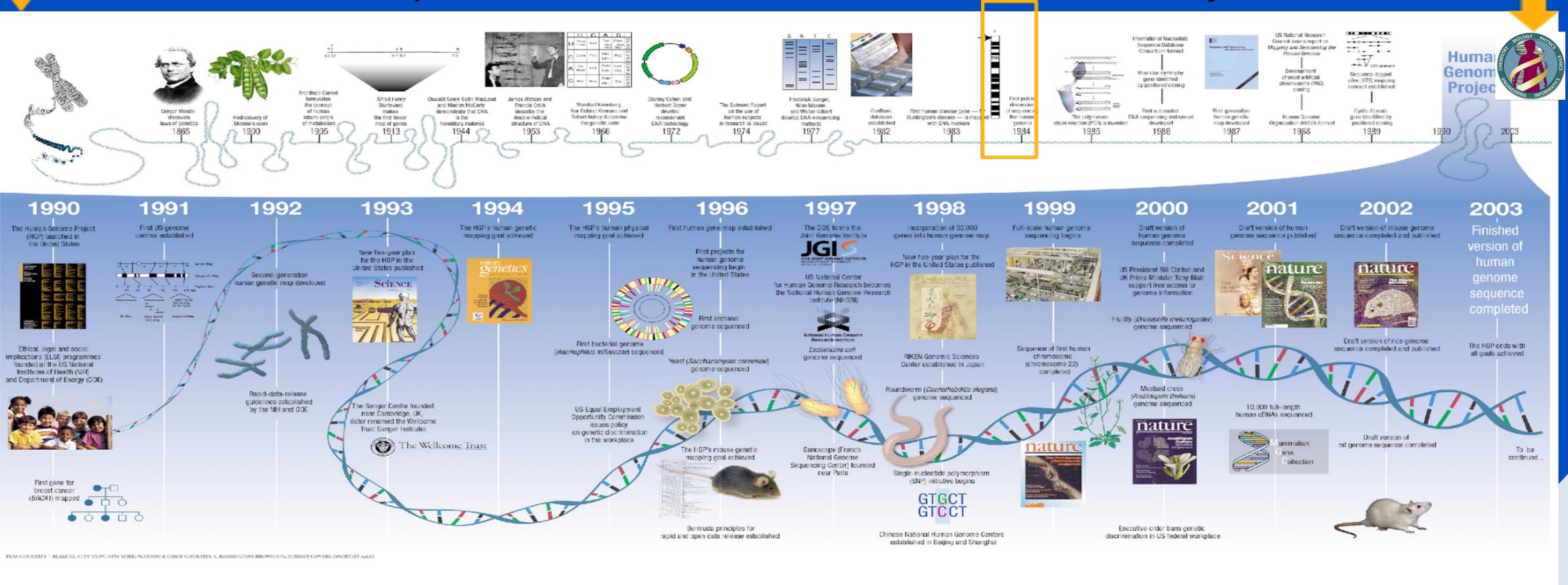
NGS Sequencing technologies

Challenges of NGS

NGS Data formats

Introduction

- 1st and 2nd Revolutions in Genomics
- The Discovery of DNA and its double helix structure
- Unexpected shift with the “Human Genome Project”



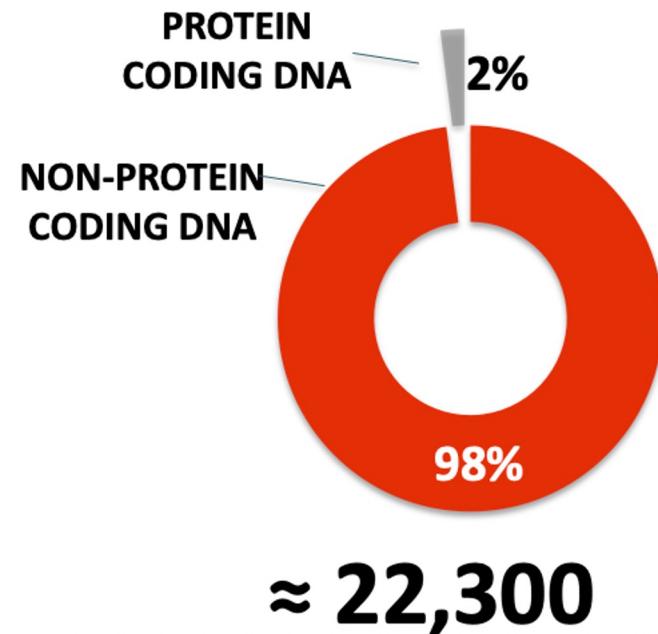
HGP: The legacy

Goals



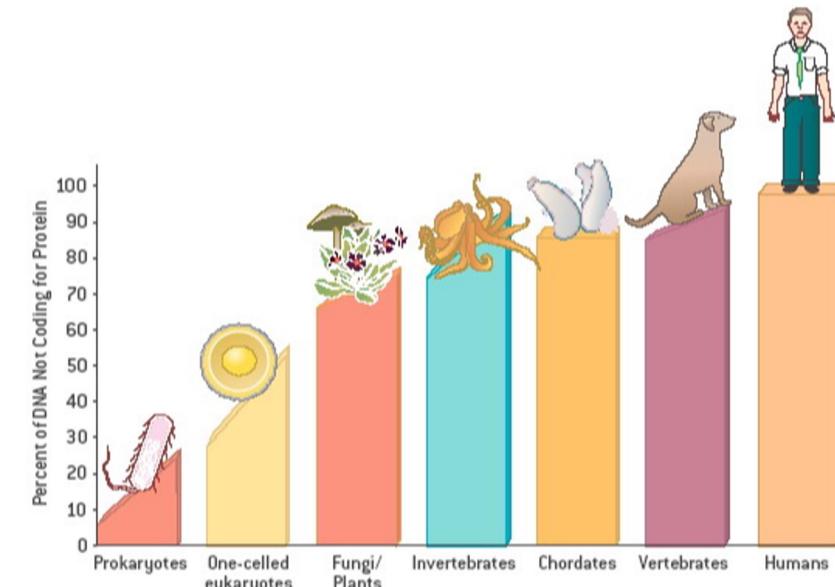
- Sequencing the nucleotides that constitute human DNA
- Identify and map all human genes at both a physical and a functional levels
- Identify the genetic variants that are related to diseases

HGP: The Legacy



The human genome contains only about 22,300 protein-coding genes : sequence alone is not enough to explain the whole complexity !

The proportion of non-coding DNA increases with organism complexity



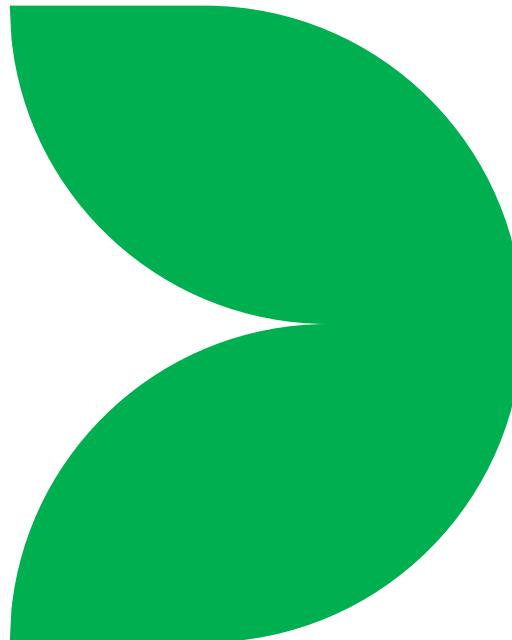


HGP : The legacy

New Challenges !

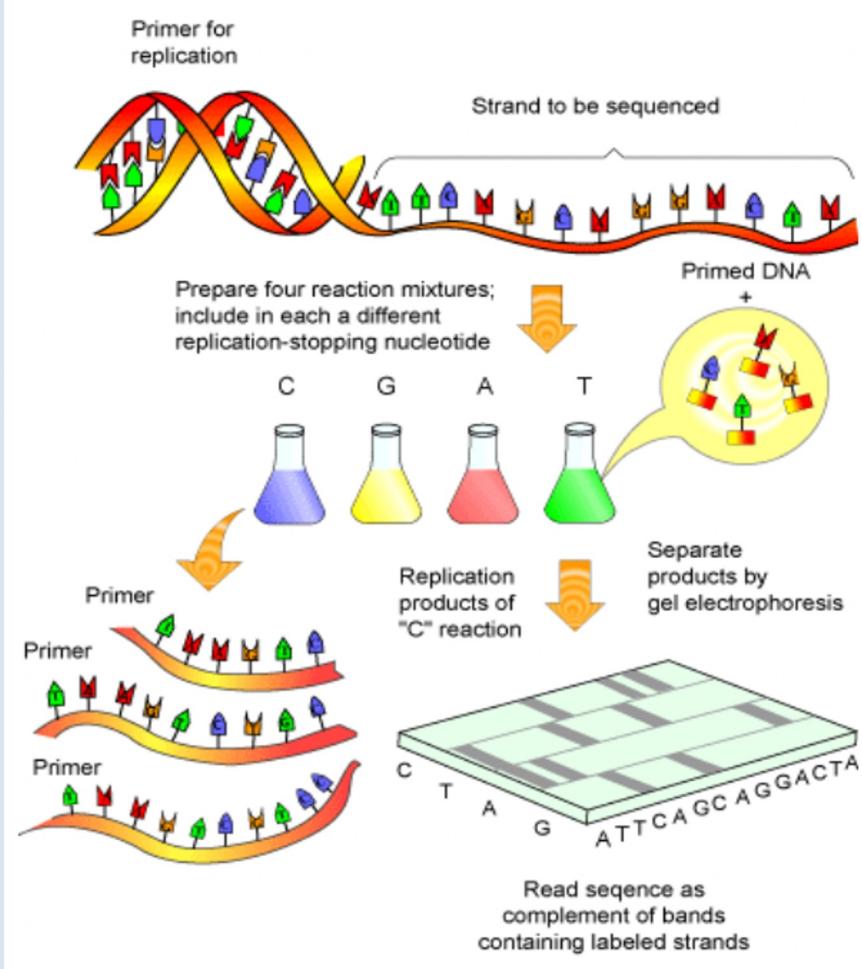
- ❑ **Exploitation:** Large amount of information ↗ major impact in medical science
 - ❑ **Global Analysis Technologies:** information on the sequence of an entire genome does not answer all of our questions - need a full understanding of the function of genes and related regulation from other regions of the genome
 - ❑ **Output:** discover how sets of genes and their products work together under normal and “abnormal” conditions (diseases).
- ❑ **One of the main requirements of these studies is the appropriate development / use of High-Throughput (HT) Sequencing technologies**

From Sanger Sequencing to NGS



First Generation

Sanger Sequencing

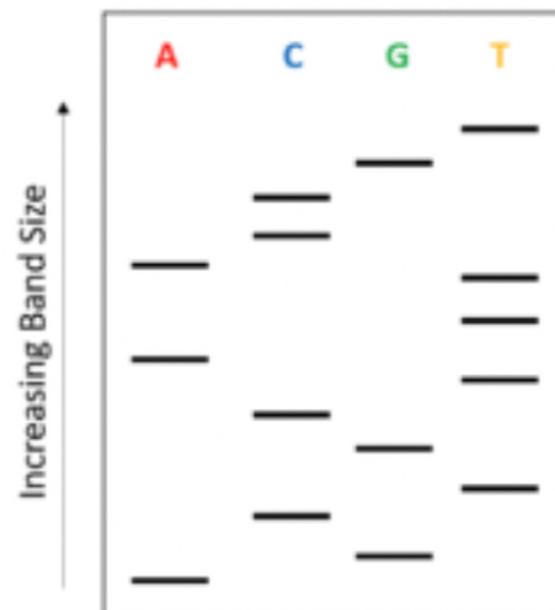


Sequence: AGCTGCTATTACCGT

Chain-terminating
PCR:

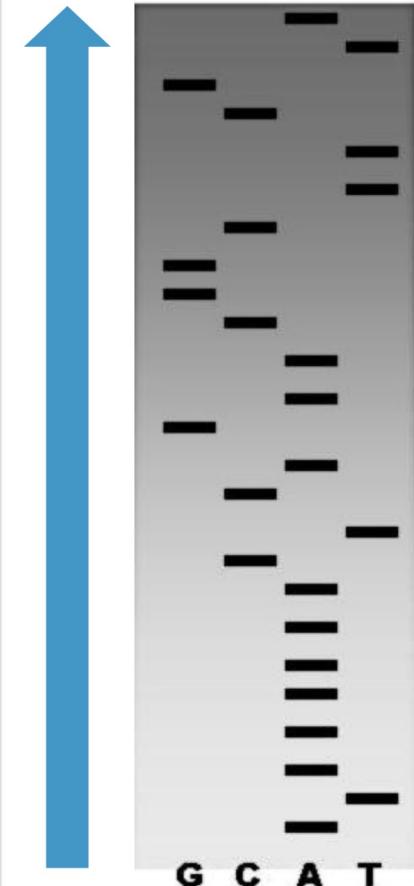


Gel Electrophoresis:

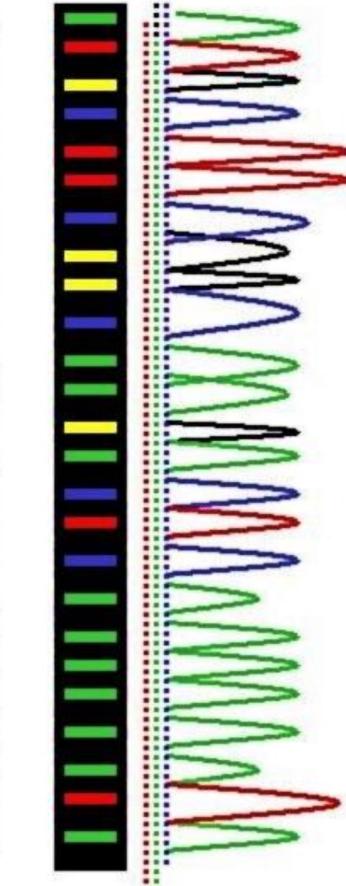


"Single" Read System/Run (i.e. 1 DNA Fragment)

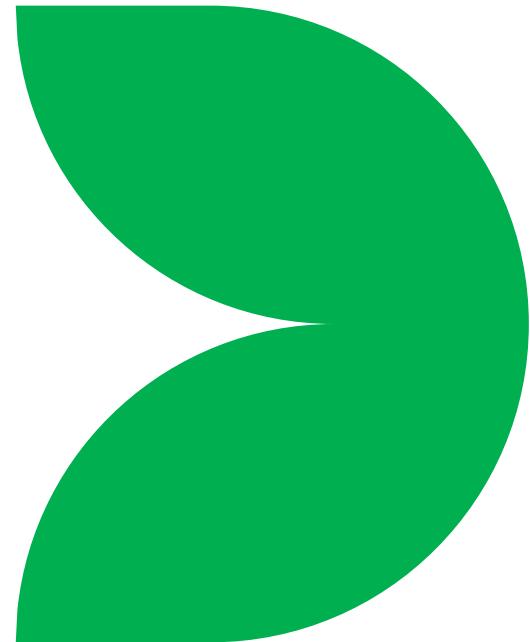
Radio-Labeled Nucleotides



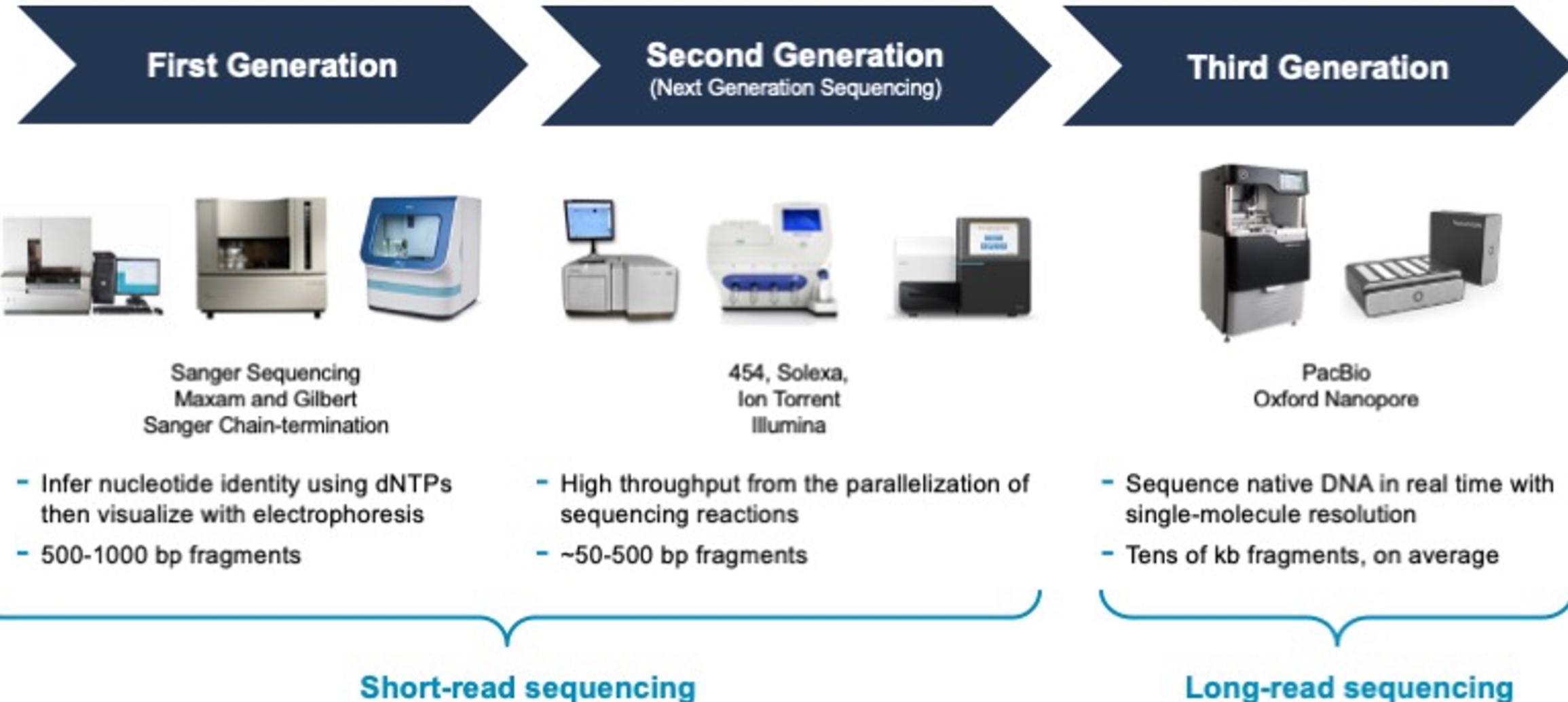
Fluorescently Labeled Nucleotides



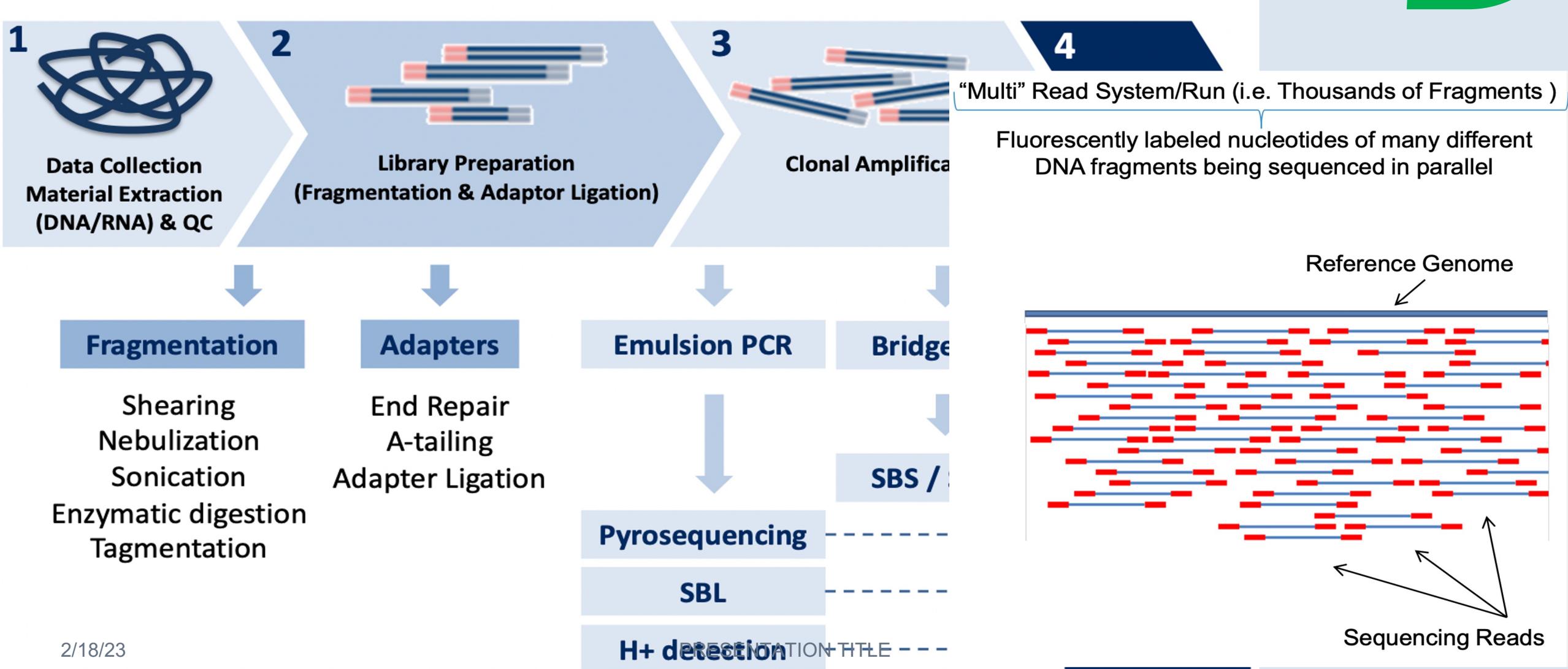
NGS Technologies



Generations of Sequencing Technologies



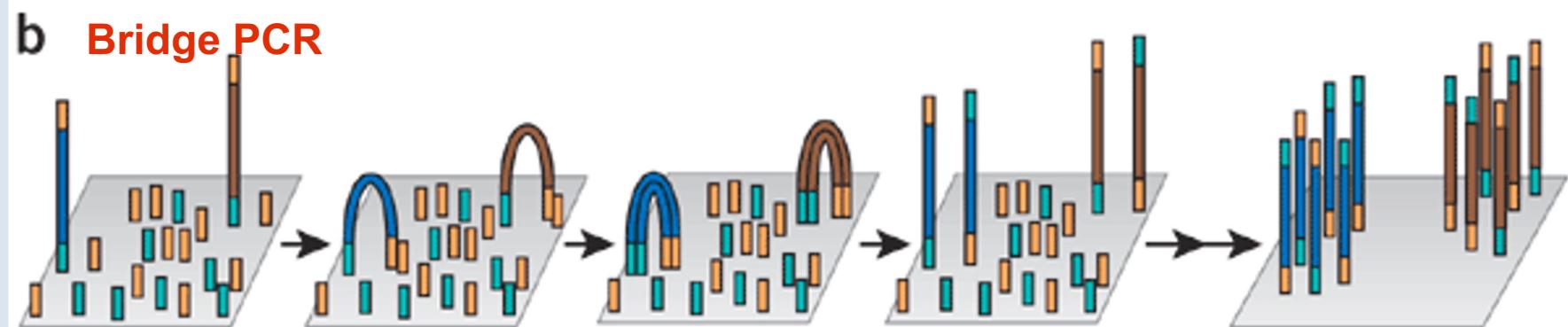
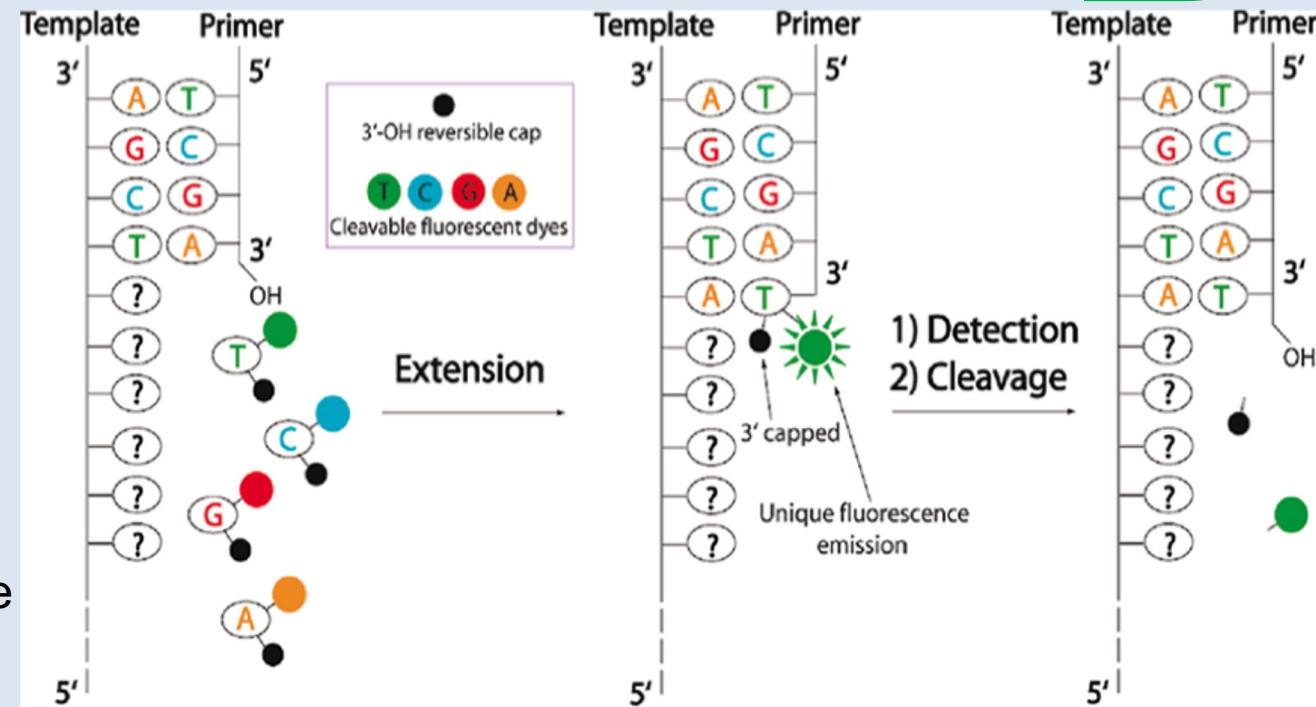
General Principle and workflow of NGS



Second generation

Principle of "Sequencing By Synthesis" (commonly **SBS**) = tracking the addition of fluorescently labelled nucleotides as the DNA chain is copied.

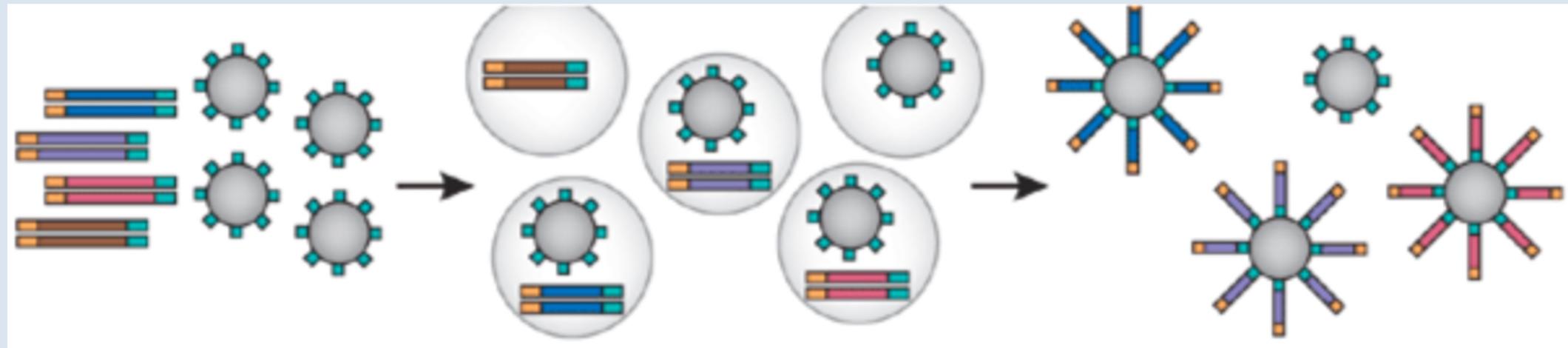
- The DNA template is immobilized.
- Solutions of A, nucleotides C, G and T added and removed.
- Fluorescence is emitted when a nucleotide complements the unpaired base.
- A chemiluminescent signal was detected to determine the sequence.



Second Generation

Principle of "emPCR"

- An adaptor-flanked shotgun library is PCR amplified in the context of a water-in-oil emulsion.
- PCR primer is 5'-attached on micron-scale beads.
- 1 bead-containing compartments = 0 or 1 template DNA.
- PCR amplicons are captured to the surface of the bead.
- 1 clonally amplified bead = PCR products corresponding to amplification of a single molecule from the library (Shendure & Ji, 2008)

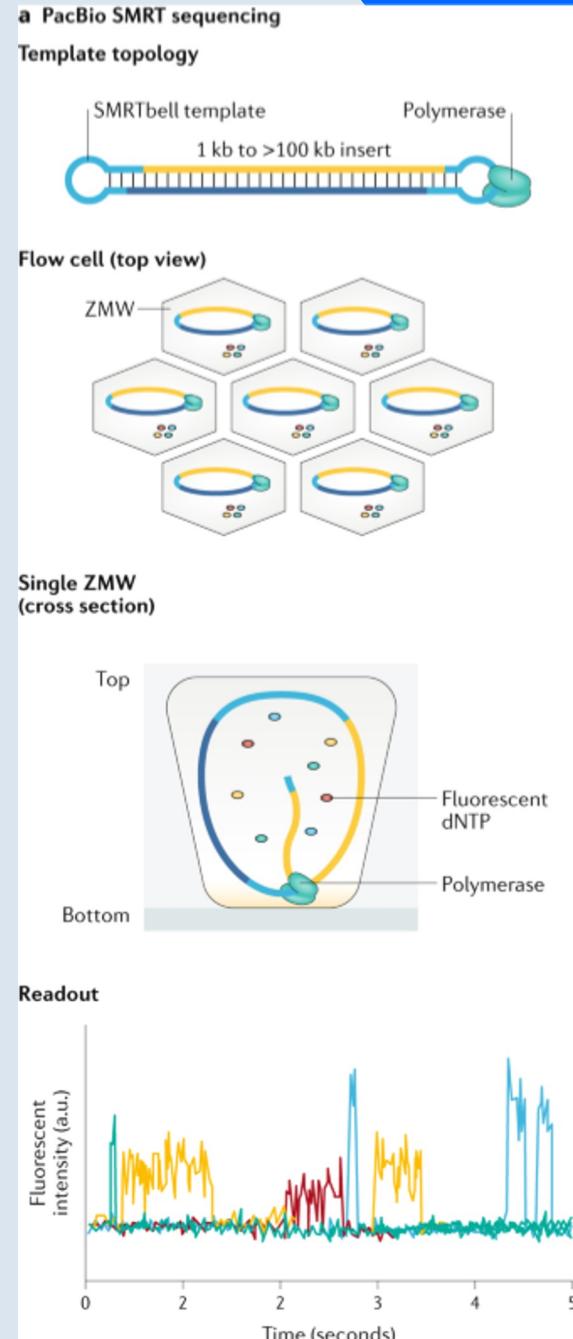
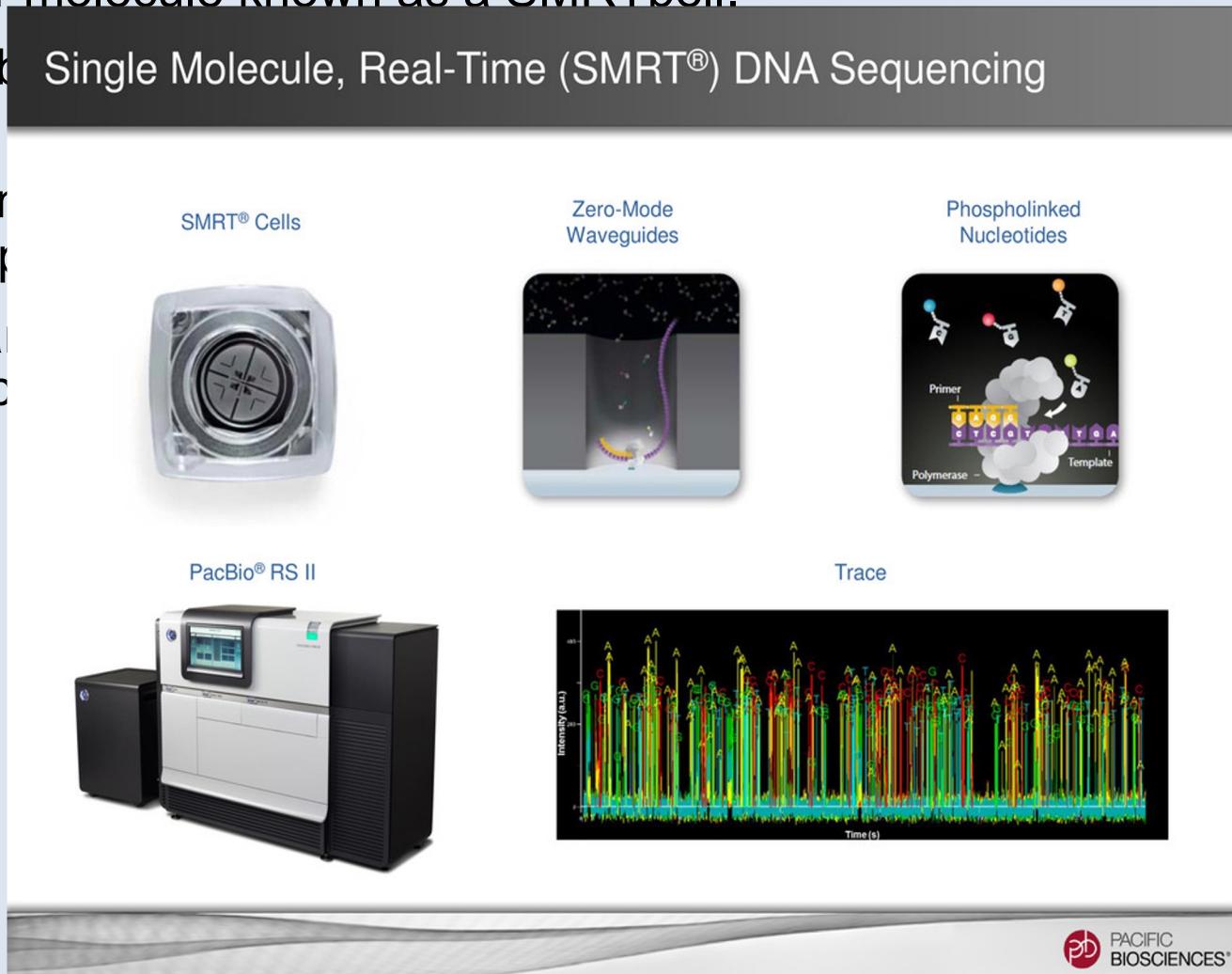


Third Generation

- DNA is fragmented and ligated to hairpin adapters (light blue) to form a topologically circular molecule known as a SMRTbell.
- SMRTbell is bound by a polymerase to a template strand during sequencing.
- Each SMRT Cell contains a series of flow cells with individual chambers that hold polymerase and template.
- Fluorescent dNTP are added to the reaction mixture along with a phosphate that will bind to the growing strand.

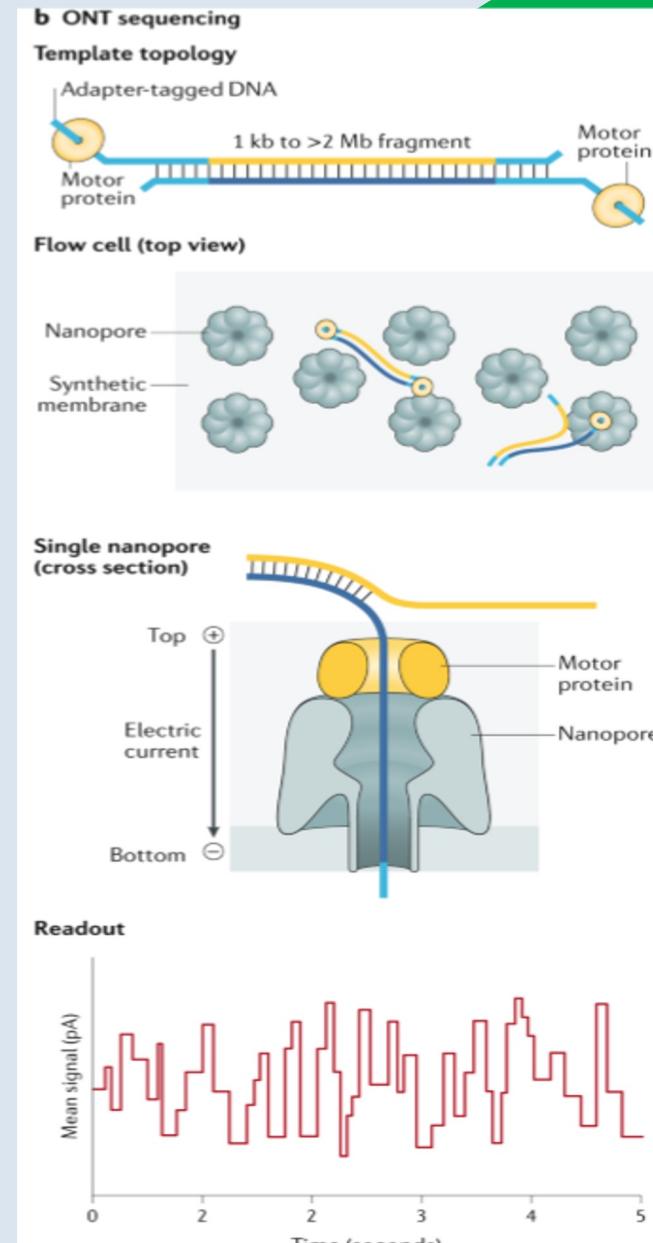
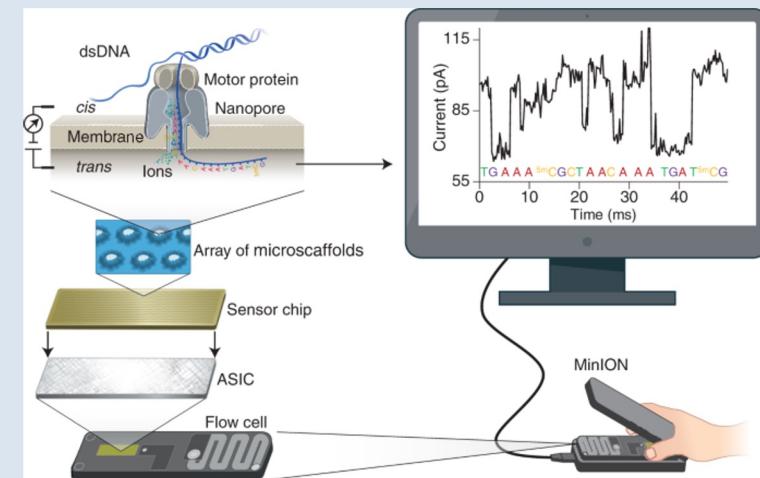
(Logsdon et al., 2020)

2/18/23



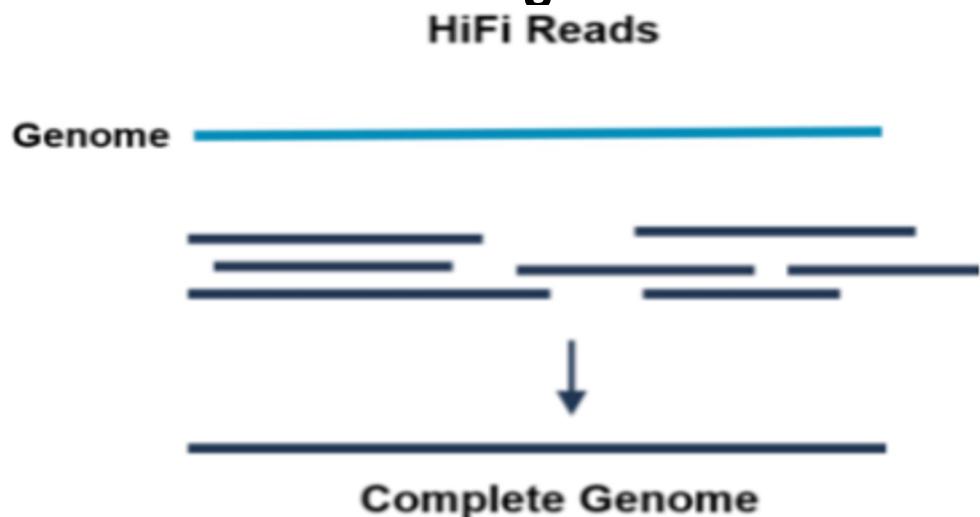
Third Generation

- DNA is tagged with sequencing adapters preloaded with a motor protein.
- The flow cell contains hundreds to thousands of protein nanopores in a synthetic membrane
- Sequencing adapter inserts into the nanopore, and the motor protein unwinds the DNA.
- An electric current is applied driving the negatively charged DNA through the pore
- As the DNA moves through the pore, it causes characteristic disruptions to the current, generating a readout known as a 'squiggle'.



Long vs short reads

- ❑ PacBio and ONT both are considered long-read sequencing technologies
- ❑ They generate different types of long reads that differ both in length and accuracy
- ❑ **The sequencing of « long reads » produces reads of 10s of kb in length**



A comprehensive structural, functional and organizational picture of the genome

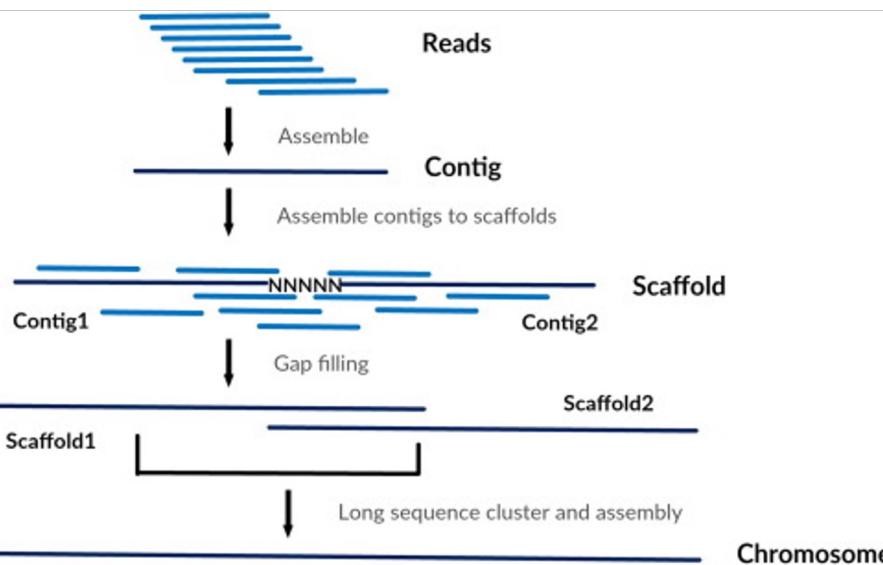
The sequencing of « short reads » produces reads of 50-500 bp



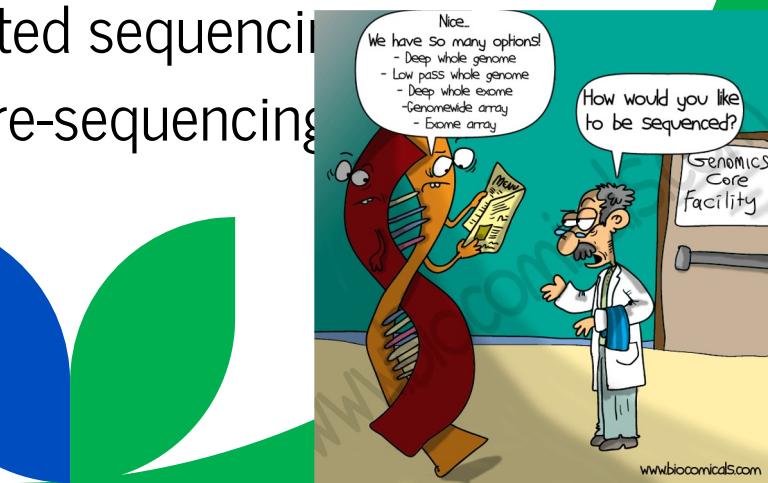
Missing sequencing leads to missed genes and limits biological interpretation

Long vs short reads

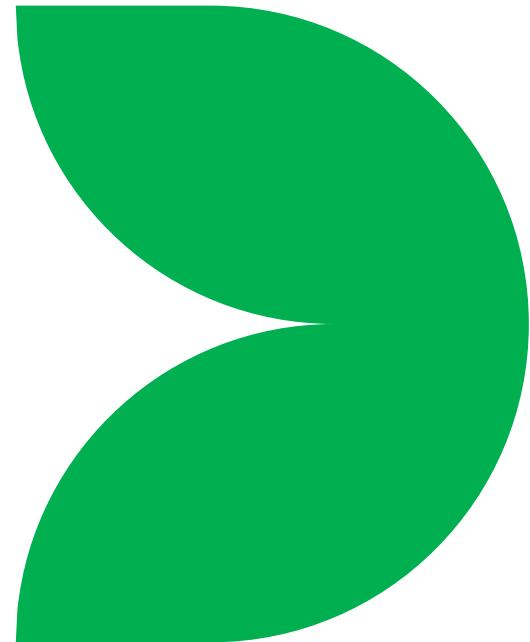
- Long reads are used to improve genome assembly
- Short & Long reads can complement each other



- The choice of one or (combination of) WGS platform will depend on your application:
 - Species identification
 - Strain/clone identification
 - Mutation identification (variant calling)
 - Accuracy: High quality finished/closed genomes (including plasmids)
 - Whole-genome vs. targeted sequencing
 - *de novo* sequencing vs. re-sequencing
 - High or low throughput



Why is HT Sequencing important?



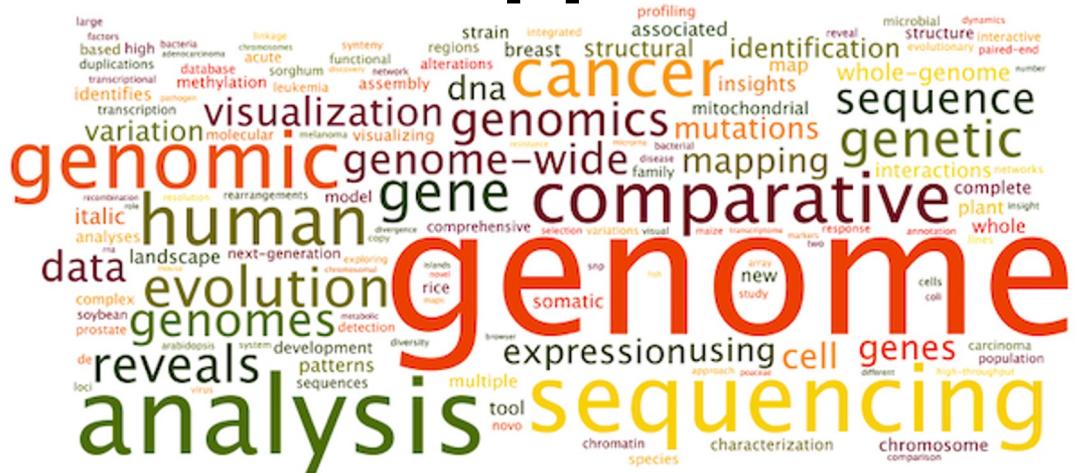
Importance of HT Sequencing

Advances

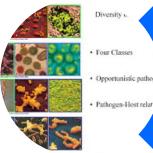
- Reduction of costs
 - Established sequencing workflows
 - Improvement of platforms
 - Sequencing performance (output size and rapidity)

Today, HT Sequencing offers a broad range of applications

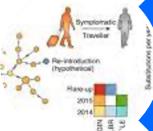
WGS applications



Applications of HT Sequencing



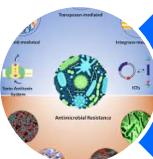
Pathogen diversity & evolution



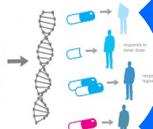
Transmission dynamics



Outbreak investigation

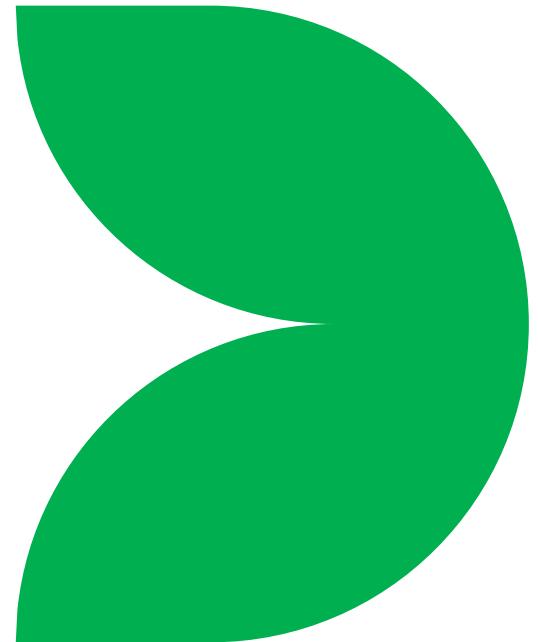


AMR surveillance



Personalized medicine

Challenges of HT Sequencing?



The Challenges of HT Sequencing

□ Data Collection

- Extensive use of technical and human resources (Data size, Data security, etc)
- Sampling accuracy (collection or heterogeneity can bias later analytical stages)

□ Data Analysis and Interpretation

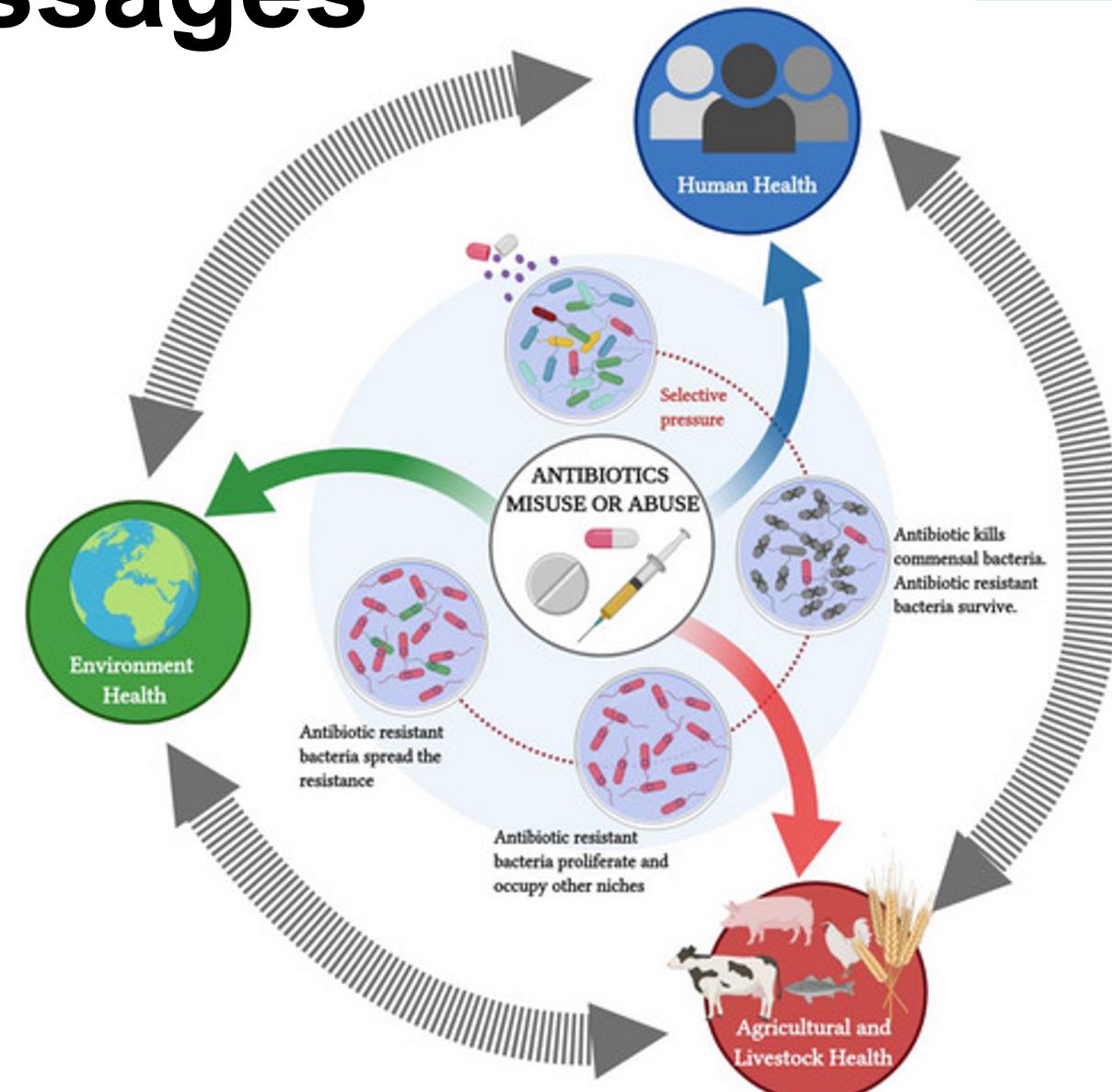
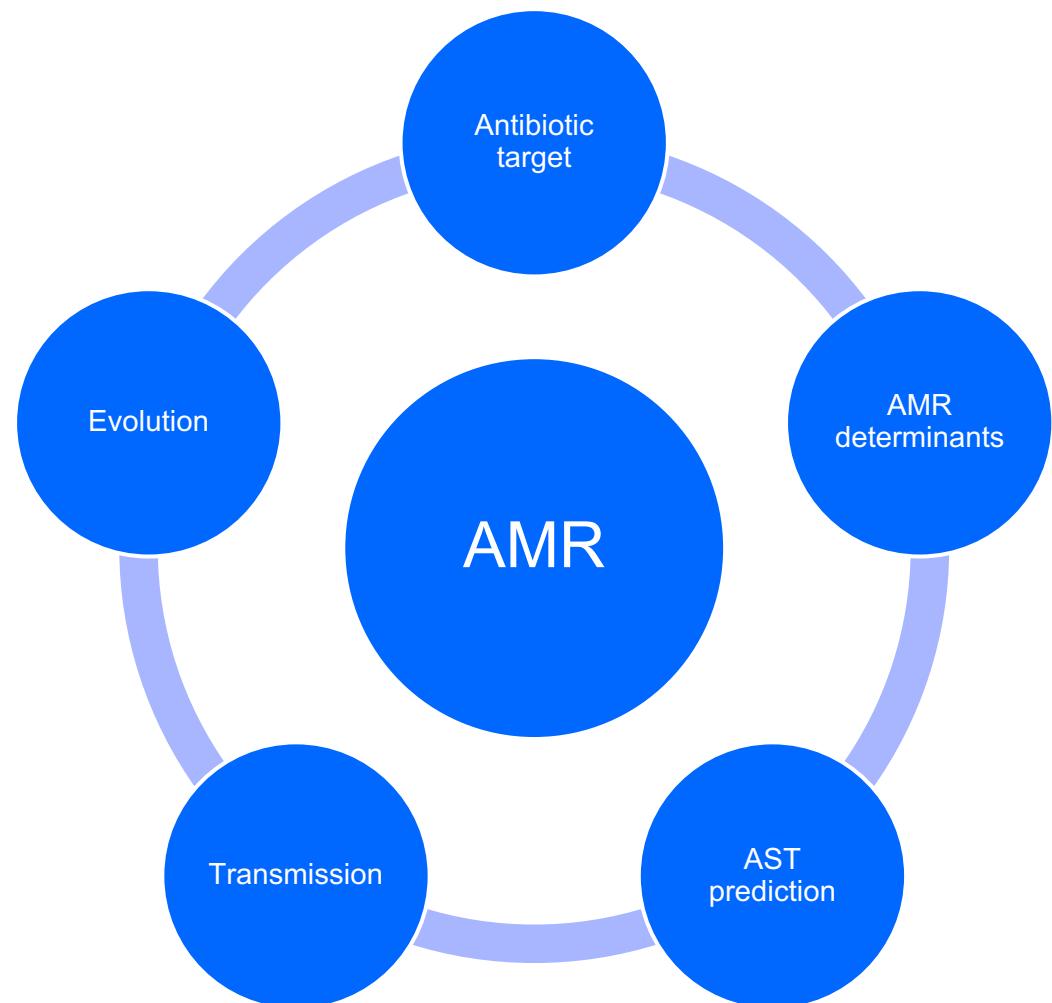
- Development and management of tools, analysis workflows, etc
- Multidisciplinary environment is crucial (biology, biostatistics, bioinformatics)

□ Analytical/computational challenges

- HT Sequence data are both high-dimensional and complex in structure both algorithmically and computationally challenging to integrate with other data sets, platforms or technologies, to obtain a complete disease profile.
- We need specific algorithmic approaches including:
- Meta-omics (integration of independent data sets at the same omics level)
- Poly-omics (integration of different omics types)

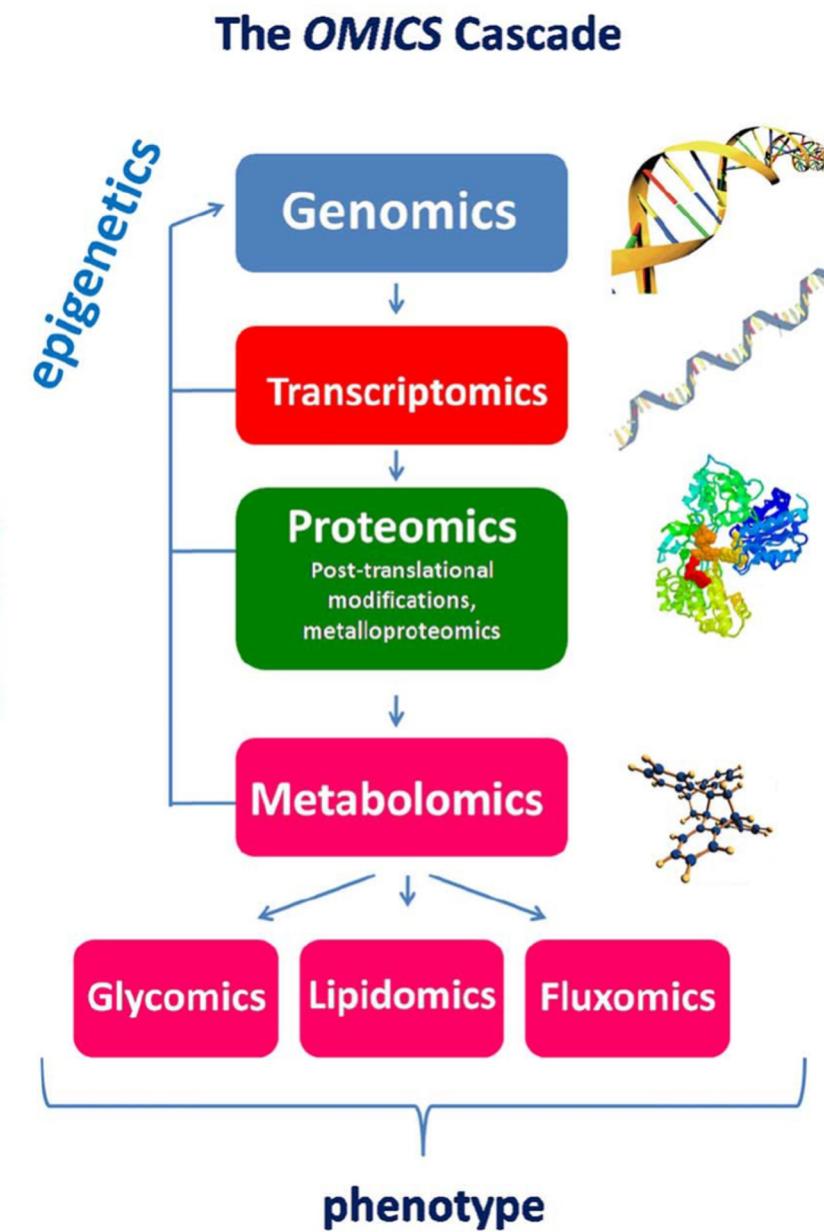
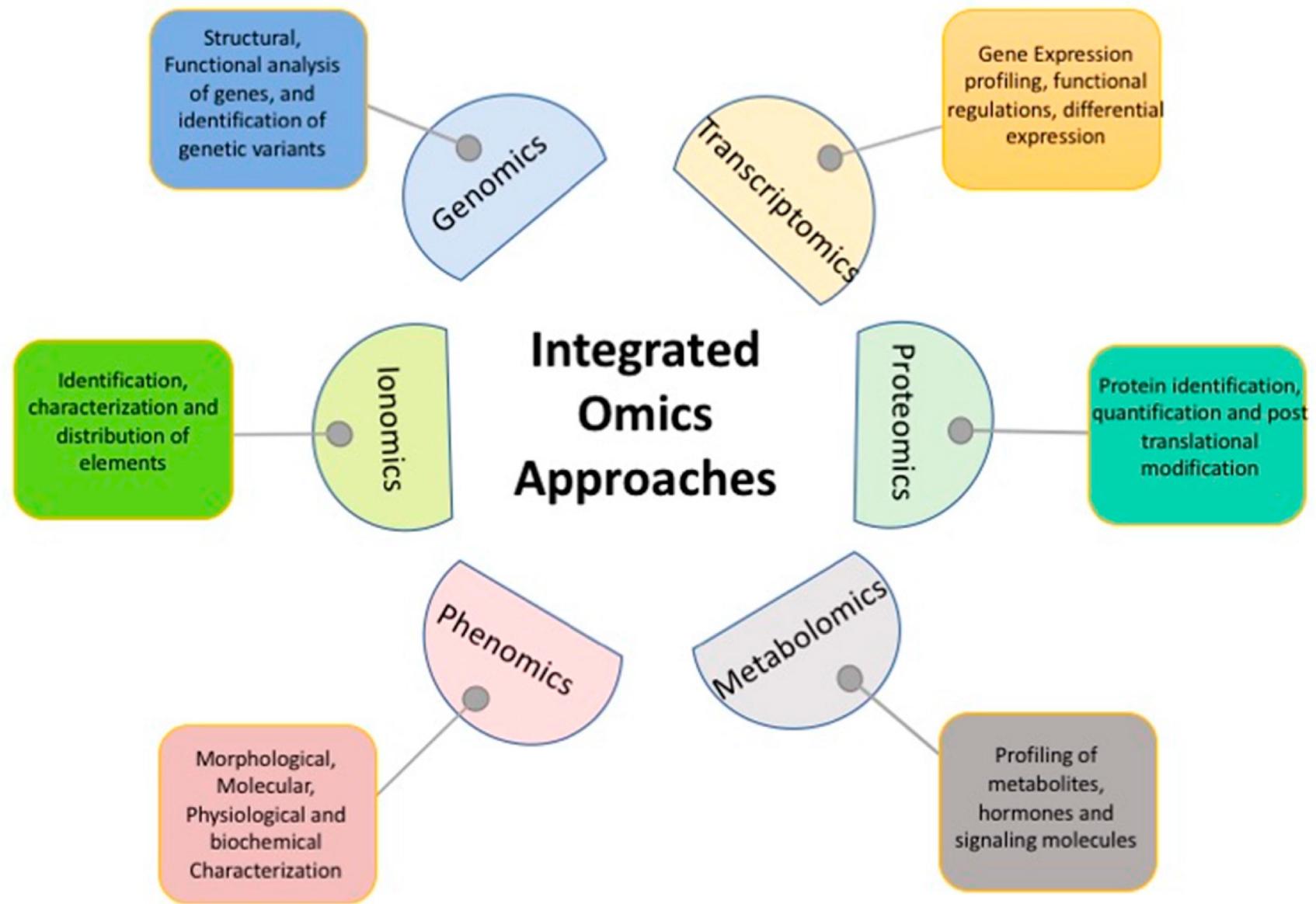
Take home messages

WGS



Leading the way through science and research with impact

Take-home messages



Take home messages

- ❑ There is no single method better than another one, all depends on your question of interest!
- ❑ One can envision a powerful symbiosis between old and new technologies (microarrays vs NGS, etc.)
- ❑ Or a combination of short vs long reads sequencing!

Great Places to Learn More About Sequencing Platforms!

1. PubMed – There are TONS of papers out there that review/compare NGS sequencing platforms
2. Some Favorite Websites:

www.youtube.com/watch?v=PMIF6zUeKko : Fantastic seminar of NGS technologies presented by Elaine Mardis, Ph.D., Genome Institute at Washington University in St. Louis.

www.SeqAnswers.com : Great message board for learning about & troubleshooting sequencing related topics.

Thank you

Dr Luria Leslie Founou
luriafounou@gmail.com

Creative Commons

This work is licensed under a

[Creative Commons Attribution-Share
Alike Licence \(CC BY-SA 4.0\).](#)



**Attribution-ShareAlike 4.0 International
(CC BY-SA 4.0)**

