



# Introduction to sequencing technologies

Martha Luka

[marthaluka.owlstown.net](http://marthaluka.owlstown.net)

01 Nov 2021

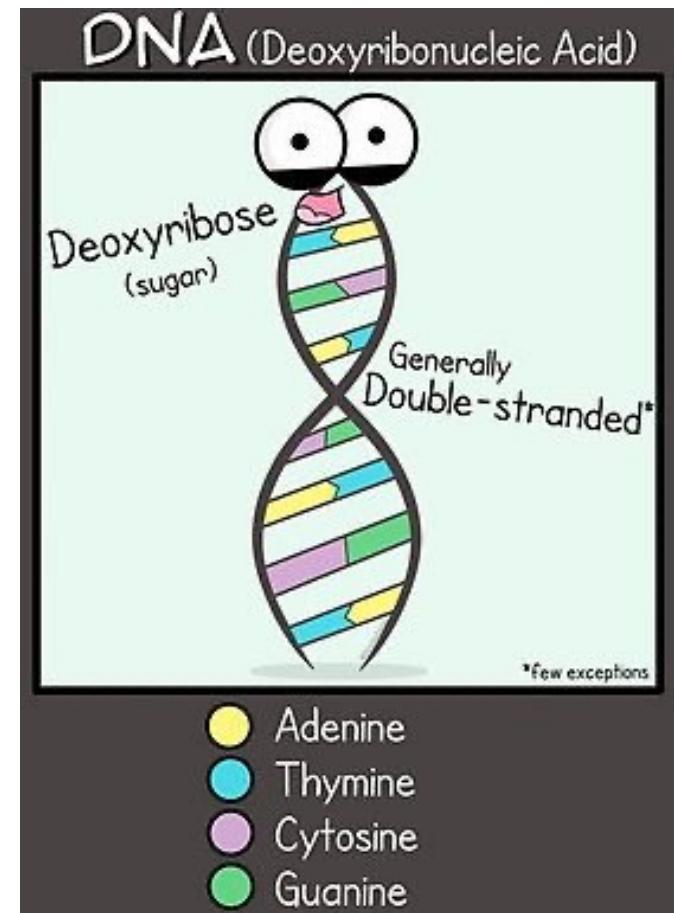
# Objectives

- Define DNA sequencing
- Describe 1st, 2nd and 3rd generation sequencing
- Applications of different sequencing techniques
- Template preparation
- How do sequencing techniques compare?
- NGS data analysis – an overview

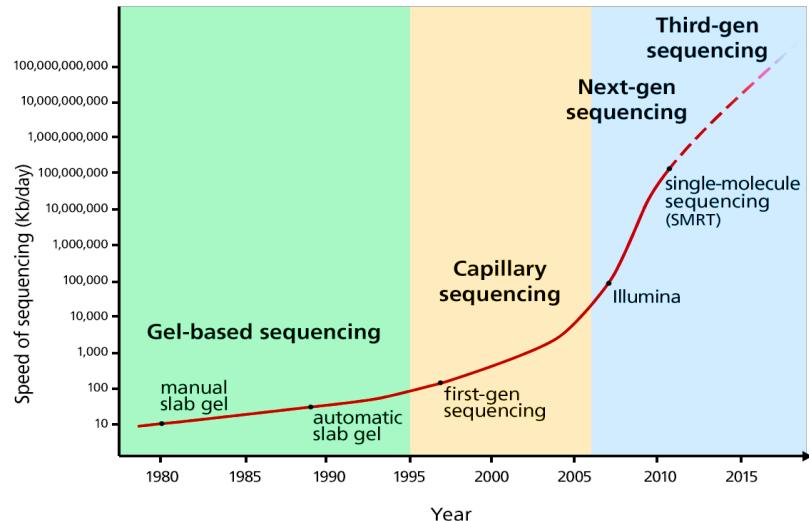


# DNA sequencing

- The process of determining the order of nucleotides, or bases (A,T,C,G) in DNA
- Applications
  - Clinical microbiology
  - Immunology & Drug discovery
  - Cancer research
  - Agrigenomics
  - Complex disease genomics
  - Human identification
  - Evolutionary biology



# History



*“... [A] knowledge of sequences could contribute much to our understanding of living matter.”*

Frederick Sanger

**1953**

3D DNA  
structure

*Watson & Crick*

**1977**

Sanger  
sequencing

*Frederick  
Sanger*

**1987**

First  
automated  
sequencer

*ABI 370*

PCR not yet invented!

1983

**1996**

Pyrosequencing

*Pål Nyrén & Mostafa  
Ronaghi*

**2001**

Draft human  
genome

**2006**

Era of NGS

# History

## First Generation



Sanger Sequencing  
Maxam and Gilbert  
Sanger Chain-termination

- Infer nucleotide identity using dNTPs then visualize with electrophoresis
- 500-1000 bp fragments

## Second Generation (Next Generation Sequencing)



454, Solexa,  
Ion Torrent  
Illumina

- High throughput from the parallelization of sequencing reactions
- ~50-500 bp fragments

## Third Generation



PacBio  
Oxford Nanopore

- Sequence native DNA in real time with single-molecule resolution
- Tens of kb fragments, on average

Short-read sequencing

Long-read sequencing



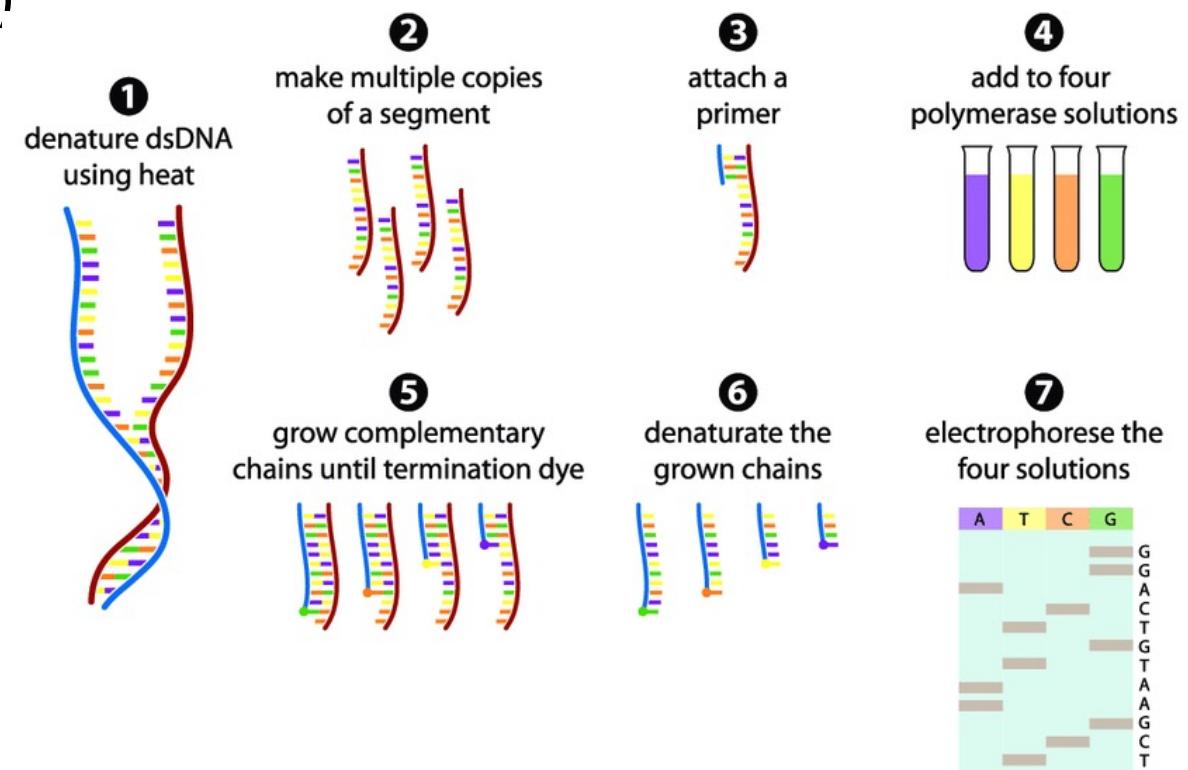
# First generation sequencing

- Maxam-Gilbert
- Sanger/Di-deoxy



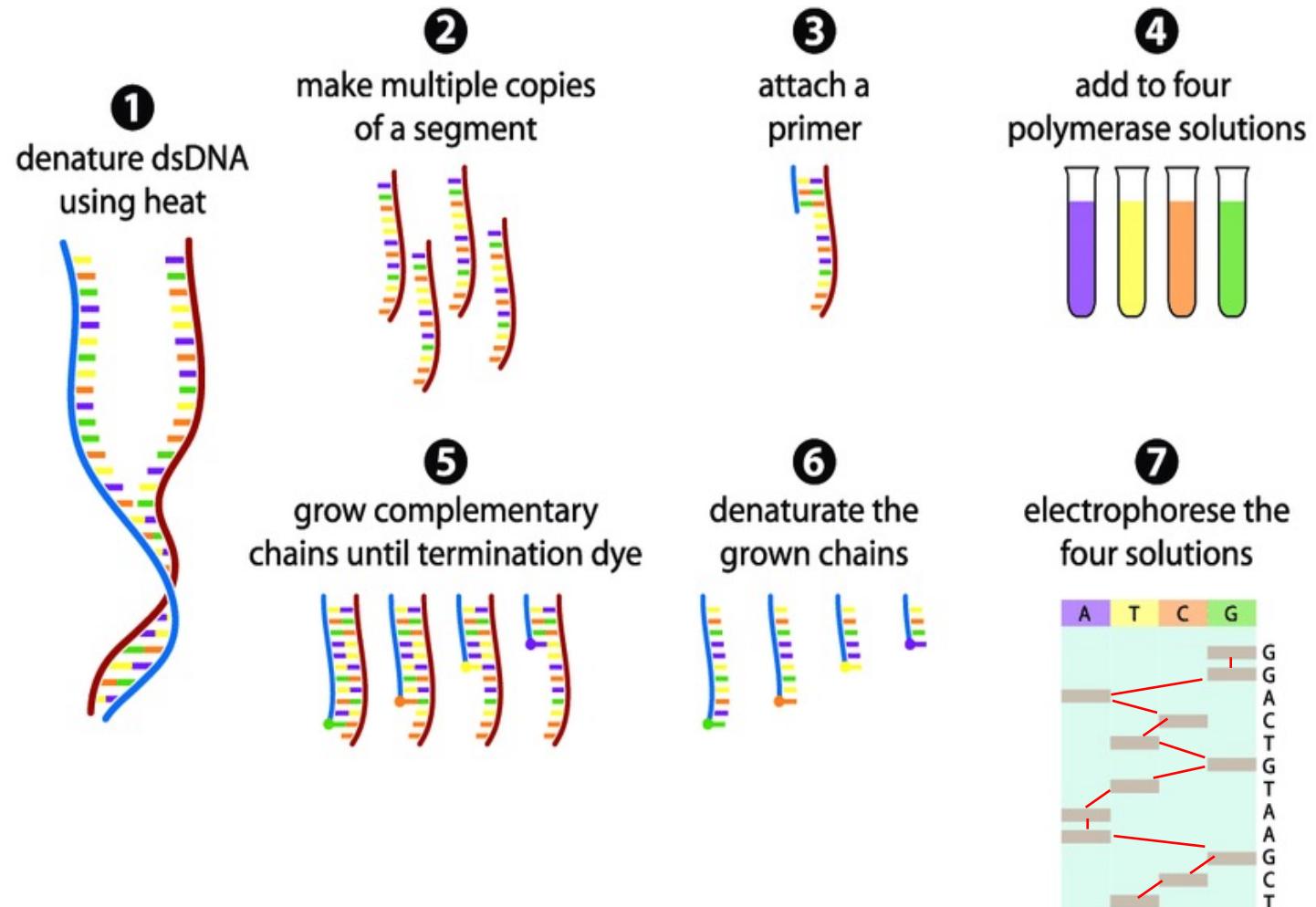
# Sanger sequencing

1. Clone DNA fragment into vectors and **amplify** (also achieved using PCR)
2. Attach **primers** to fragments
3. Four tubes:
  1. Four standard dNTPs
  2. Only one type of ddNTPs
4. DNA **polymerase** to initiate the synthesis
5. Gel electrophoresis
6. Read sequence from bands



# Sanger sequencing

1. Clone DNA fragment into vectors and amplify (also achieved using PCR)
2. Attach primers to fragments
3. Four tubes:
  1. Four standard dNTPs
  2. Only one type of ddNTPs
4. DNA polymerase to initiate synthesis
5. Gel electrophoresis
6. Read sequence from bands

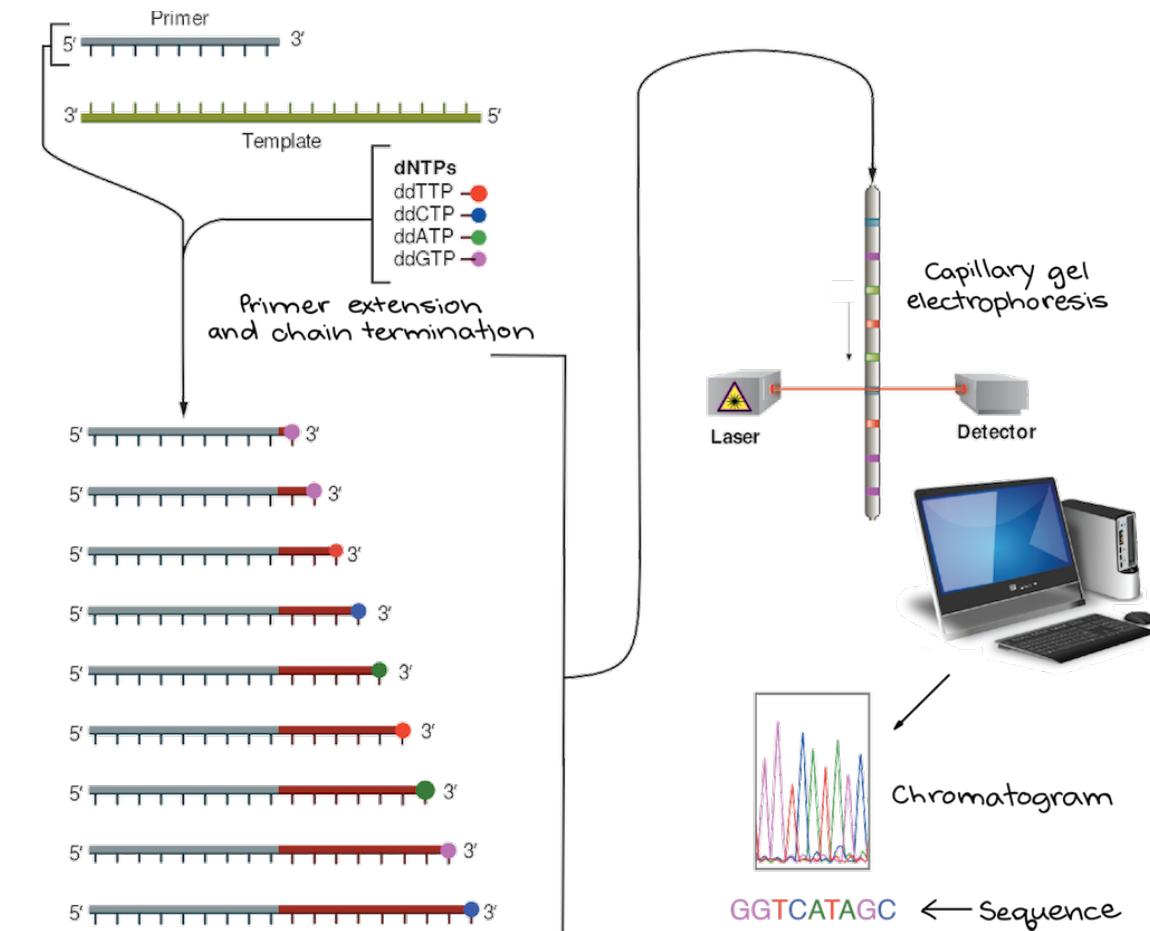


GGACTGTAAAGCT



# Automated Sanger sequencing

- Same process but dNTPs are fluorescently labelled
- High resolution capillary electrophoresis coupled to four color emission spectra
- Advantages
  - High accuracy
  - Long individual reads
- Disadvantages
  - More expensive and impractical for larger sequencing projects



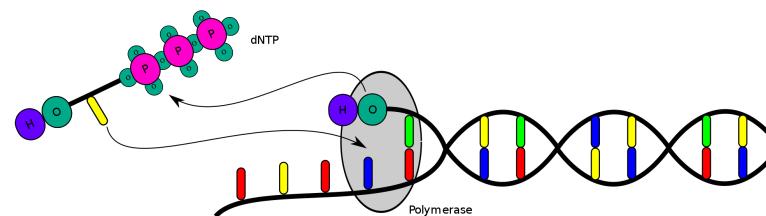
# Second generation sequencing

- Roche/ 454 sequencing
- Ion torrent sequencing
- ABI/ SOLiD sequencing
- Illumina/ Solexa sequencing

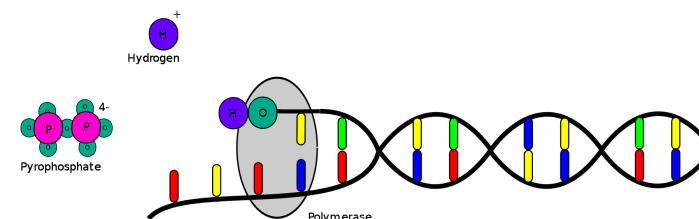


# Ion torrent sequencing

- Based on the detection of hydrogen ions that are released during DNA polymerization i.e. sequencing by synthesis
- A complementary strand is built based on the sequence of a template strand



Polymerase integrates a nucleotide.

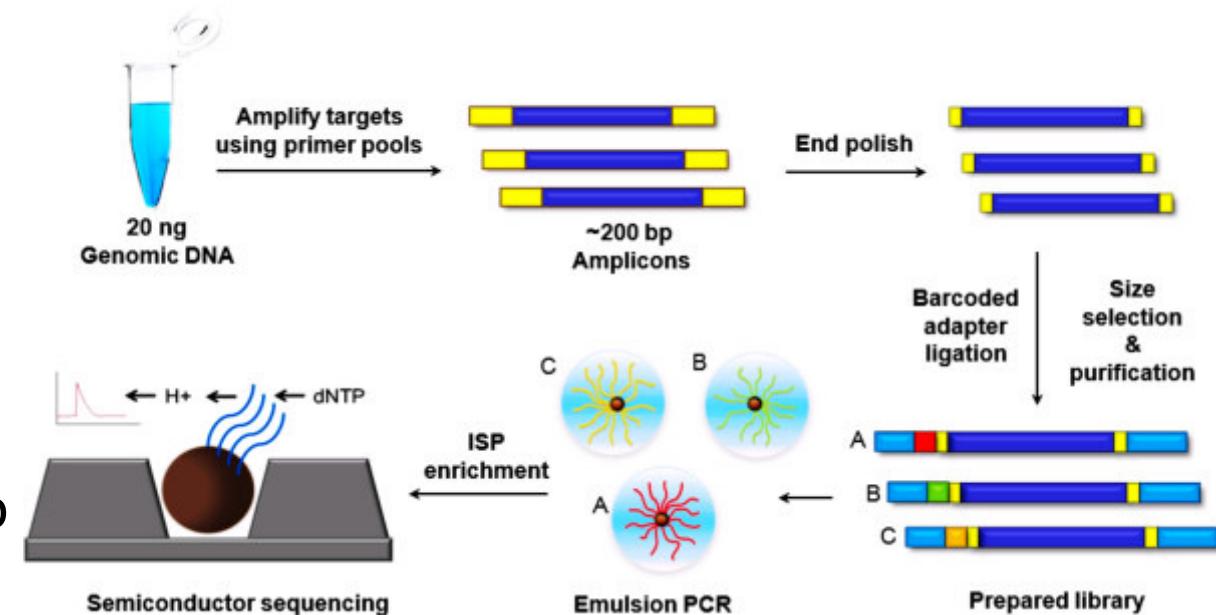


Hydrogen and pyrophosphate are released.



# Ion torrent sequencing

- DNA is fragmented
- Each fragment attached to bead, copied till covers bead
- Ion chip with millions of wells. Beads each deposit to a well
- Hydrogen ion changes pH of well.
- Wells transform nt info into digital info – base calling
- Ion sensitive layer beneath well



# Ion torrent sequencing

## ***Advantages***

- Rapid sequencing speed and low upfront and operating costs
- Sequencing can occur in real-time (by synthesis)

## ***Disadvantages***

- Difficult to enumerate long repeats (homopolymers)
- Short read length (~400 nt per read)

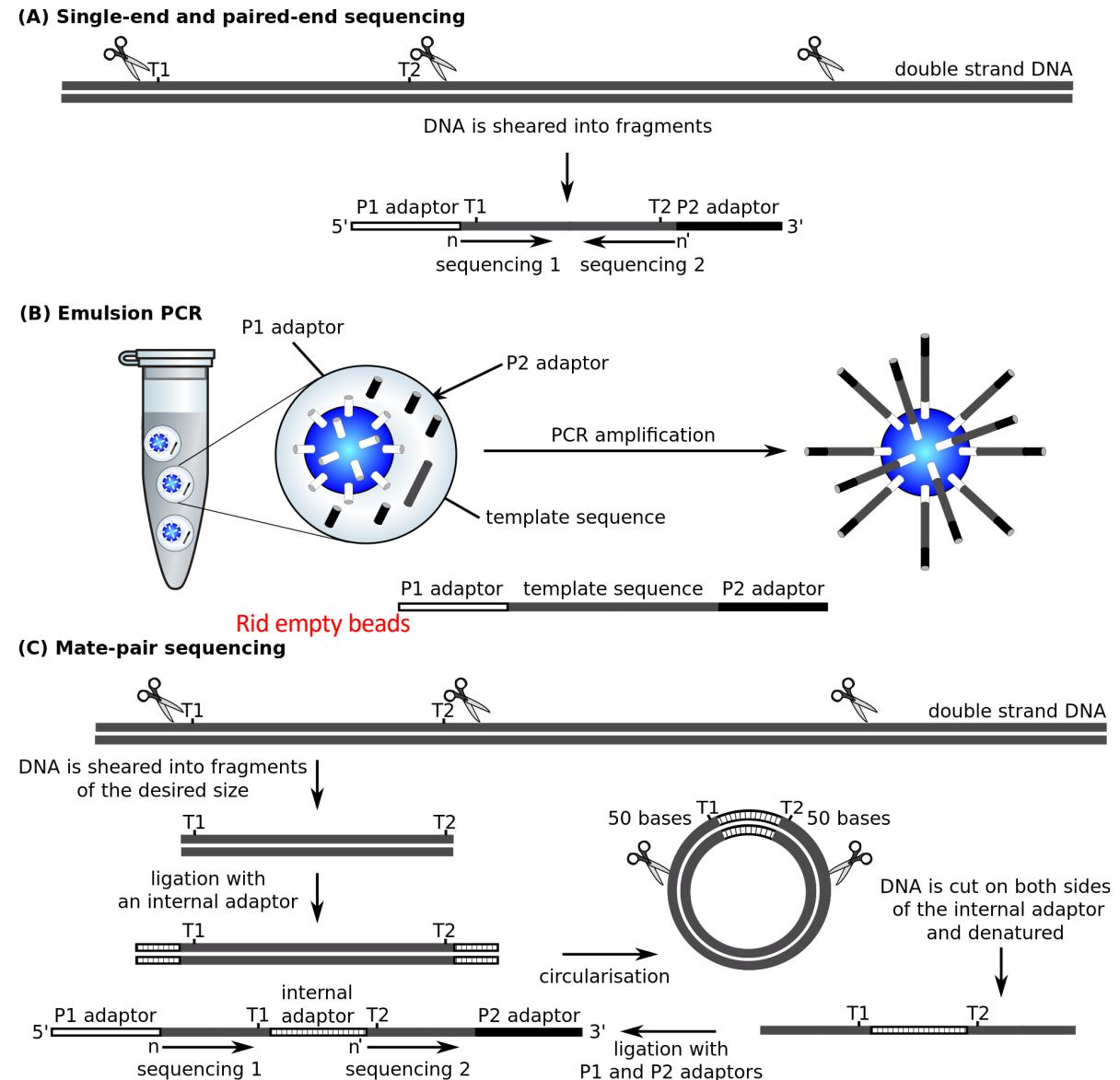
## ***Applications***

- A rapid, compact and economical sequencer. Works well in smaller labs.
- Small scale applications eg microbial genome, transcriptome and targeted sequencing



# ABI/ SOLiD sequencing

- Sequencing by Oligonucleotide Ligation and Detection(SOLID)
- Primer binds to template strand
- Probe hybridization and ligation to primer
- Measure fluorescence



# ABI/ SOLiD sequencing

## ***Advantages***

- Low cost per run
- Not hindered by homopolymers

## ***Disadvantages***

- Slower than other methods
- Trouble with palindromic sequences, (DNA is inaccessible as in loops)

## ***Applications***

- Transcriptomics and epigenomics

5' - ATAT - 3'

3' - TATA - 5'

Palindromic

5' - ATTA - 3'

3' - TAAT - 5'

NOT Palindromic



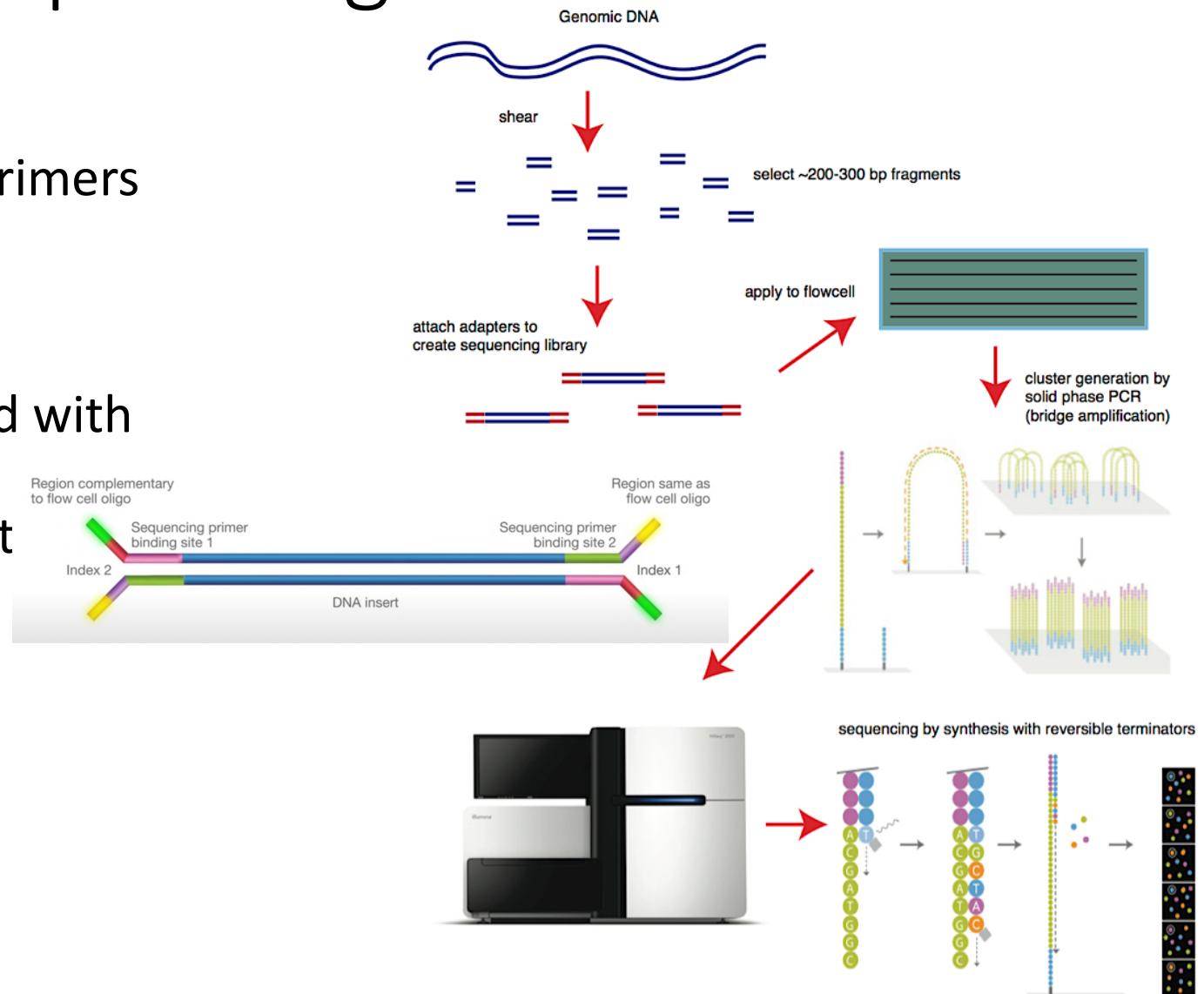
# Illumina/ Solexa sequencing

## Sample prep

- Tagmentation – adapters, primers

## Cluster generation

- Amplification - flowcell lined with two types of oligos, complementary to fragment strands
- Reverse strands washed off

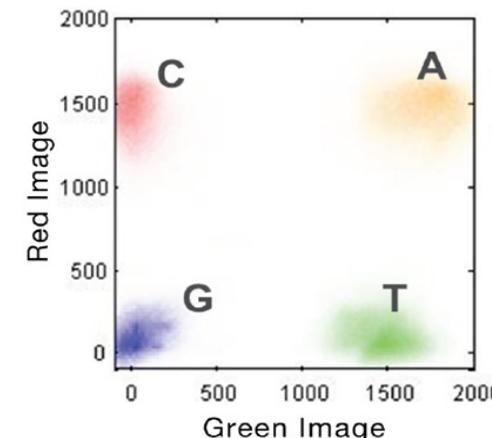
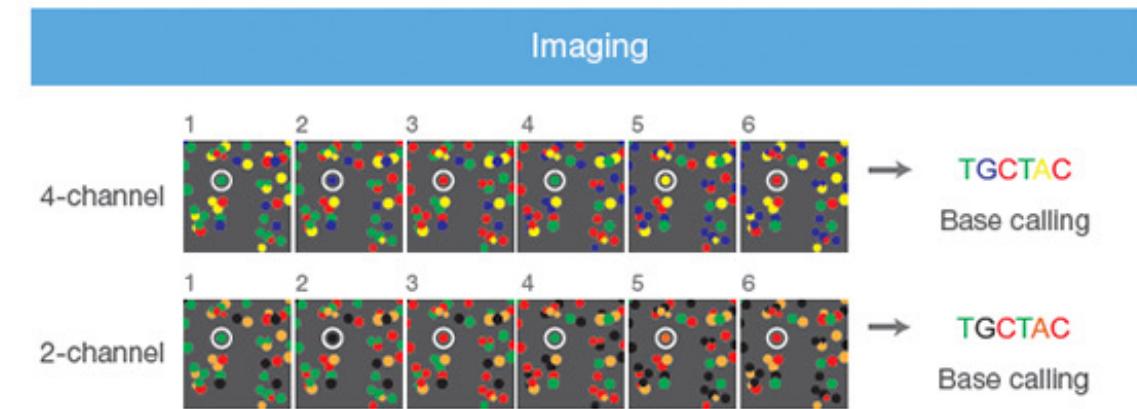


# Illumina/ Solexa sequencing

## Sequencing

- Sequencing primer
- 4 fluorescently nucleotides.
- Identical strands read simultaneously
- Read 1 washed away, template folds over
- Read 2 sequenced (PE sequencing)

- Characteristically fluorescent signal
- Sequences separated based on indicies at library prep



# Illumina/ Solexa sequencing

A detailed animation of this

<https://www.illumina.com/science/technology/next-generation-sequencing/sequencing-technology/2-channel-sbs.html>



# Illumina/ Solexa sequencing

## ***Advantages***

- Proven base calling accuracy
- High quality pairwise alignments
- Cost effective
- Deep coverage making it easier to perform de novo assemblies for genome
- Scalable



## ***Limitations***

- High start up cost
- Less cost-effective for sequencing low numbers of targets (1–20 targets)
- Short reads

## ***Applications***

- Whole-genome sequencing
- Gene expression profiling
- Methylation profiling
- Metagenomics



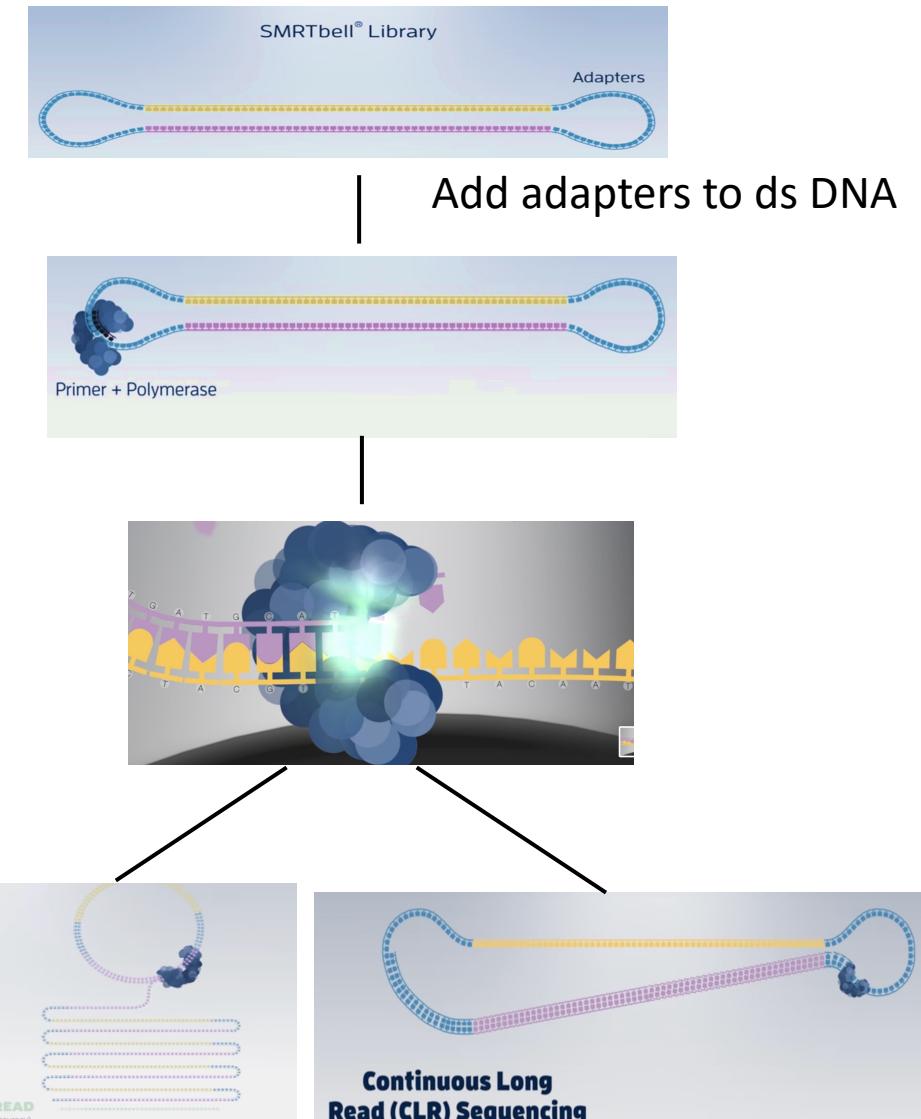
# Third generation sequencing

- Pacific biosciences SMRT sequencing
- Oxford MinION (Nanopore) sequencing



# Pacific biosciences SMRT

- Single Molecule, Real-Time (SMRT) Sequencing
- Adapters added to dsDNA creating a loop
- Polymerase incorporates tagged dNTPs. Emission of fluorescence
- Nucleotide incorporation measured in real-time
- Two sequencing modes:
  - Longest possible read (CLR)
  - Circular (High quality HIFI reads)



# Pacific biosciences SMRT

## ***Advantages***

- Long reads (tens of kilobases in length)
- Accuracy (>99.999% consensus accuracy)
- Epigenetics. With no PCR amplification step, base modifications are directly detected during sequencing.

## ***Disadvantages***

- Expensive equipment
- Low yield at high accuracy



### Whole Genome Sequencing

For humans, plants, animals and microbes including *de novo* assembly and variant detection



### Complex Populations

Understand variants among bacterial, viral and cancer cell populations



### RNA Sequencing

In-depth analysis of cDNA sequences across the entire transcriptome or targeted genes



### Epigenetics

Detect DNA modifications in your samples while you sequence on the PacBio platform

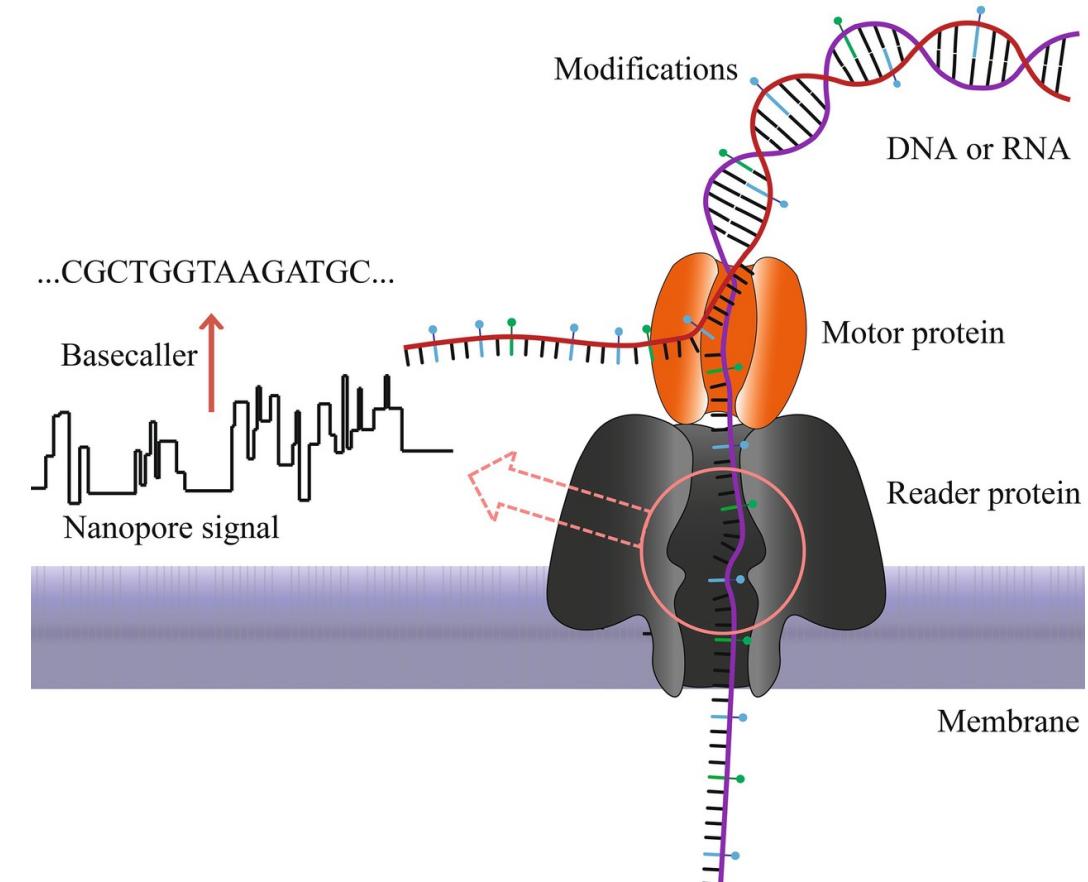


### Targeted Sequencing

Study relevant genome targets across any regions of interest

# Oxford nanopore sequencing

- Uses protein nanopores bathed in an electrophysiological solution
- An ionic current is passed through the nanopores
- Molecules moving through the nanopores, disrupt the current
- Signal analysed in real time
- Read length directly related to template length presented (up to 2 Mb)
- Direct sequencing allows modification information e.g. methylation



# Oxford nanopore sequencing

## ***Advantages***

- Sequence any fragment *length* (short to ultra-long)
- *Direct sequencing* of native DNA/RNA - epigenetics
- Data available in *real time* – analyse as soon as you start sequencing
- *Scalable* - Sequence anything, anywhere
- *Cost effective*
- *Simple & rapid library preparation*

## ***Disadvantages***

- Error rates can be as high as 15%
- Frequent modification of kits and reagents
- High levels of interrun-variability

## **Applications**

- Whole genome sequencing
- Epigenetics
- Targeted sequencing
- ....



# Comparison of different technologies

	Sanger	Ion torrent (sequencing by synthesis)	Illumina (sequencing by synthesis)	SOLiD (sequencing by ligation)	Oxford nanopore	Single Molecule Real-Time (SMRT)
Read length	400-900bp	200bp	50-500bp	50+50 bp	10 kb – 2.3Mb	2900bp avg, upto 20kb
Accuracy	99.9%	98%	98%	99.9%	99.1%	87% (read length mode), 99% accuracy
Reads per run	NA	<5 mn	<3 bn	1.2-1.4 bn	>2 mn (24 flowcell)	35-75kb
Time per run	<3 hours	2 hours	1-10 days	1-2 weeks	Scalable	0.5-2 hours
Cost (1million bases)	\$2400	\$1	\$0.15	\$0.13	<\$0.05	\$2



# THE EVOLUTION OF SEQUENCING

## First Generation

Sanger Sequencing  
Maxam and Gilbert  
Sanger Chain-termination

- Infer nucleotide identity using dNTPs then visualize with electrophoresis
- 500-1000 bp fragments
- Relatively slow and expensive

## Second Generation

Next Generation Sequencing

454, Solexa, Ion Torrent,  
Illumina

- High throughput from the parallelization of sequencing reactions
- High accuracy
- ~50-500 bp fragments
- Faster and more affordable

## Third Generation

PacBio, Oxford Nanopore

- Sequence native DNA in real time with single-molecule resolution
- Traditionally lower accuracy than NGS
- Tens of kb fragments, on average



Short-read sequencing

Long-read sequencing

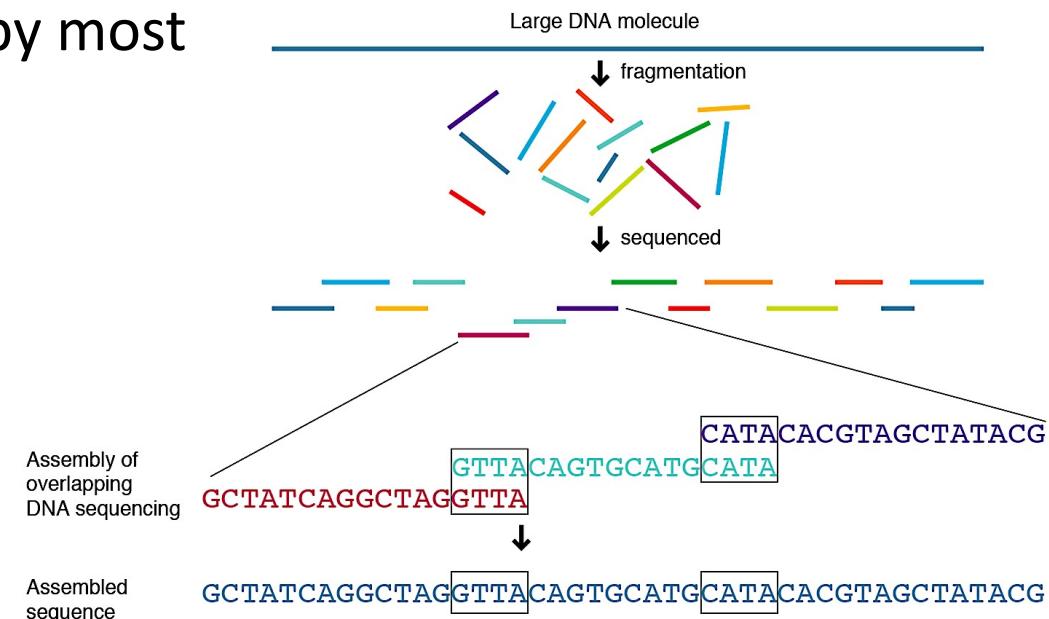
# Metagenomics vs targeted enrichment

- Target Enrichment - pre-sequencing step where DNA sequences are either:
  - Directly amplified (amplicon or multiplex PCR-based) or
  - Captured (hybrid capture-based)
- *Agnostic/metagenomics* - all the nucleic acid (DNA and/ or RNA) of a specimen is sequenced in parallel.
  - Unbiased (in theory)
  - Expensive



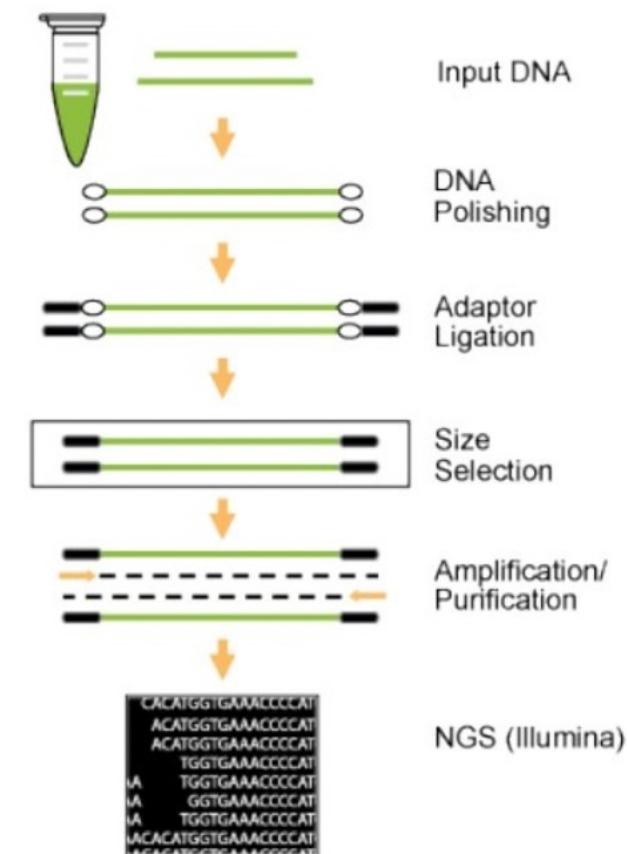
# Library preparation

- Similar principle across sequencers:
  - Fragmentation
  - Size selection
  - Adapter ligation
- Long fragments cannot be directly sequenced by most sequencers.
  - Nebulization (compressed nitrogen)
  - Sonication (vibration)
  - Digestion (restriction enzymes)



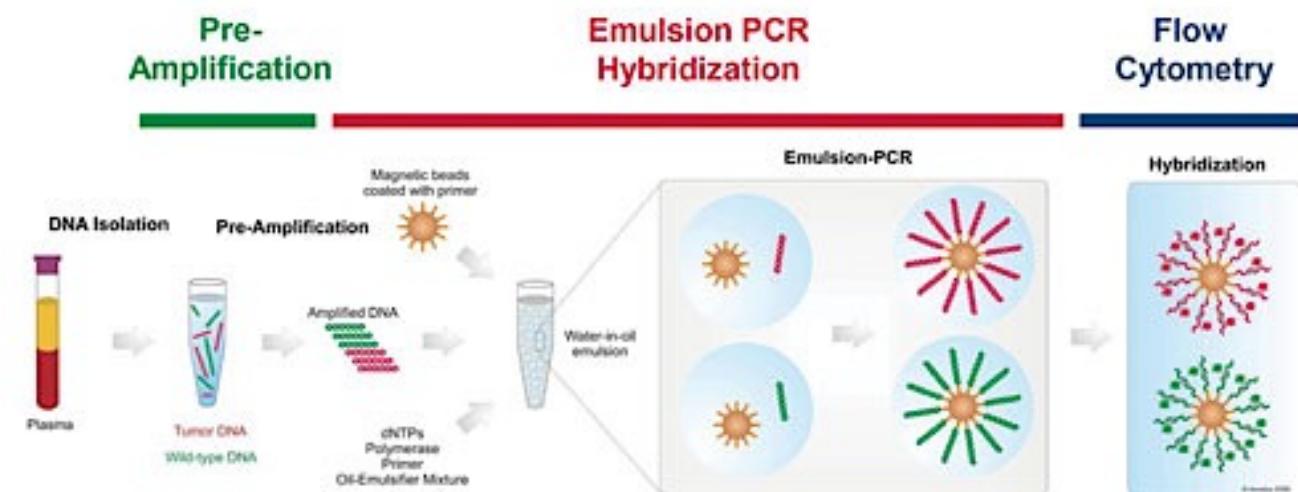
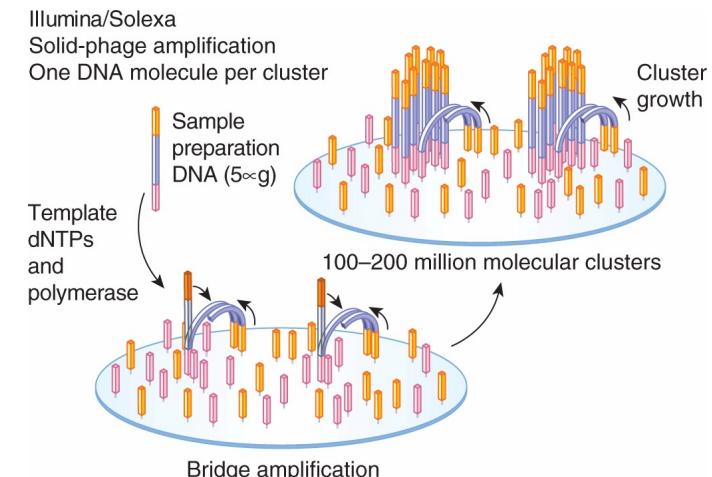
# Library preparation

- Size selection:
  - Many fragments end up with overhanging sticky ends
  - Transformed into blunt ends using a polymerase
  - After blunting, 5' ends are phosphorylated for ligation and 3' dA tailing for Illumina
- Adapter ligation – both ends using DNA ligase
  - Universal PCR primers for amplification
  - Hybridization sequences
  - Molecular barcodes for multiplex pooled sequencing



# Amplification/ Enrichment

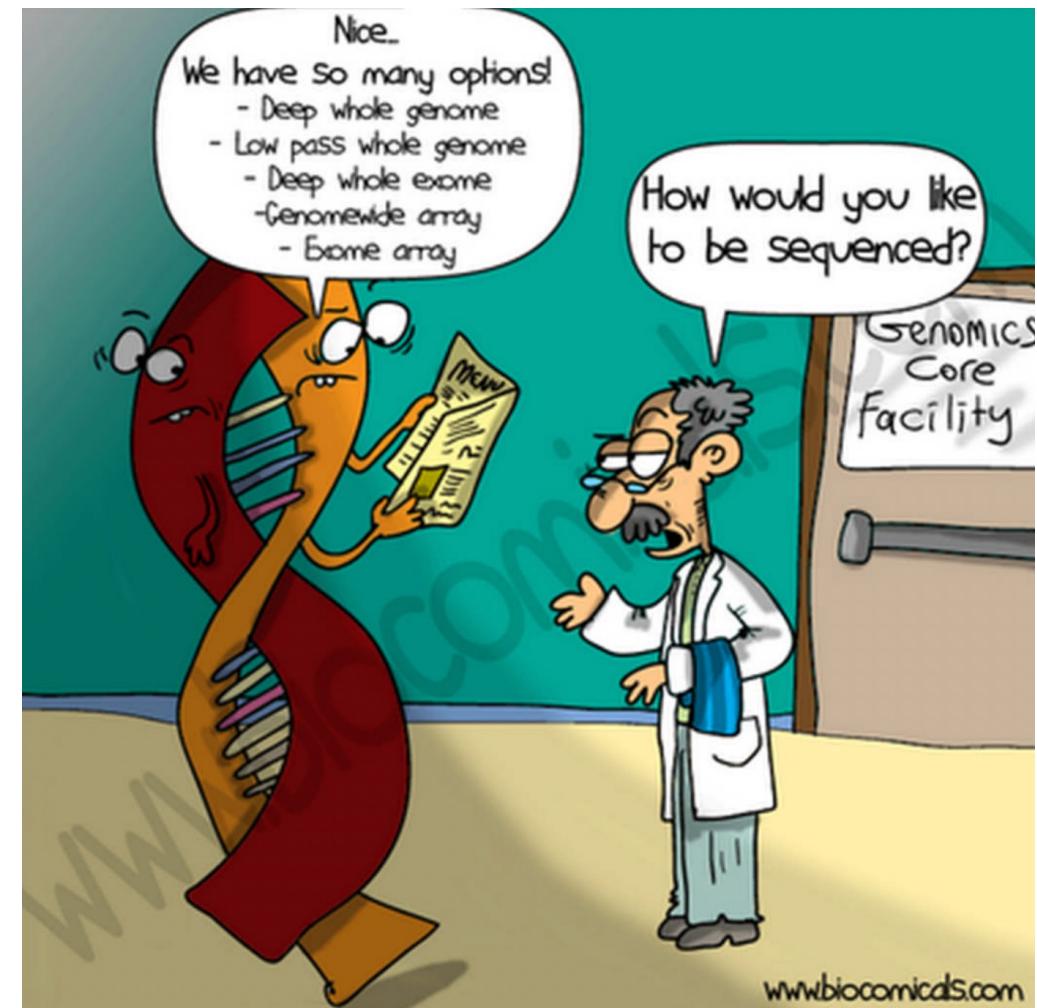
- DNA libraries are clonally amplified
- Amplification necessary before sequencing as most sequencing platforms are not designed to sequence one molecule
- Technique depends on platform, but most common are
  - Emulsion PCR (SOLiD and Ion Torrent)
  - Solid-phase amplification (Illumina)



# Applications

How do I then choose a sequencing technology?

- Why I'm I sequencing?
- What I'm I sequencing?
  - Template
  - Number of samples
- Do I have budget limitations?
- Other logistics



# Applications

## Question 1

- Michael aims to understand the genetic diversity of Cassava within East Africa using whole genome sequencing. The genome size of Cassava is 770 Mb. Which sequencing technique would you advise him to use?



# Applications

## Question 2

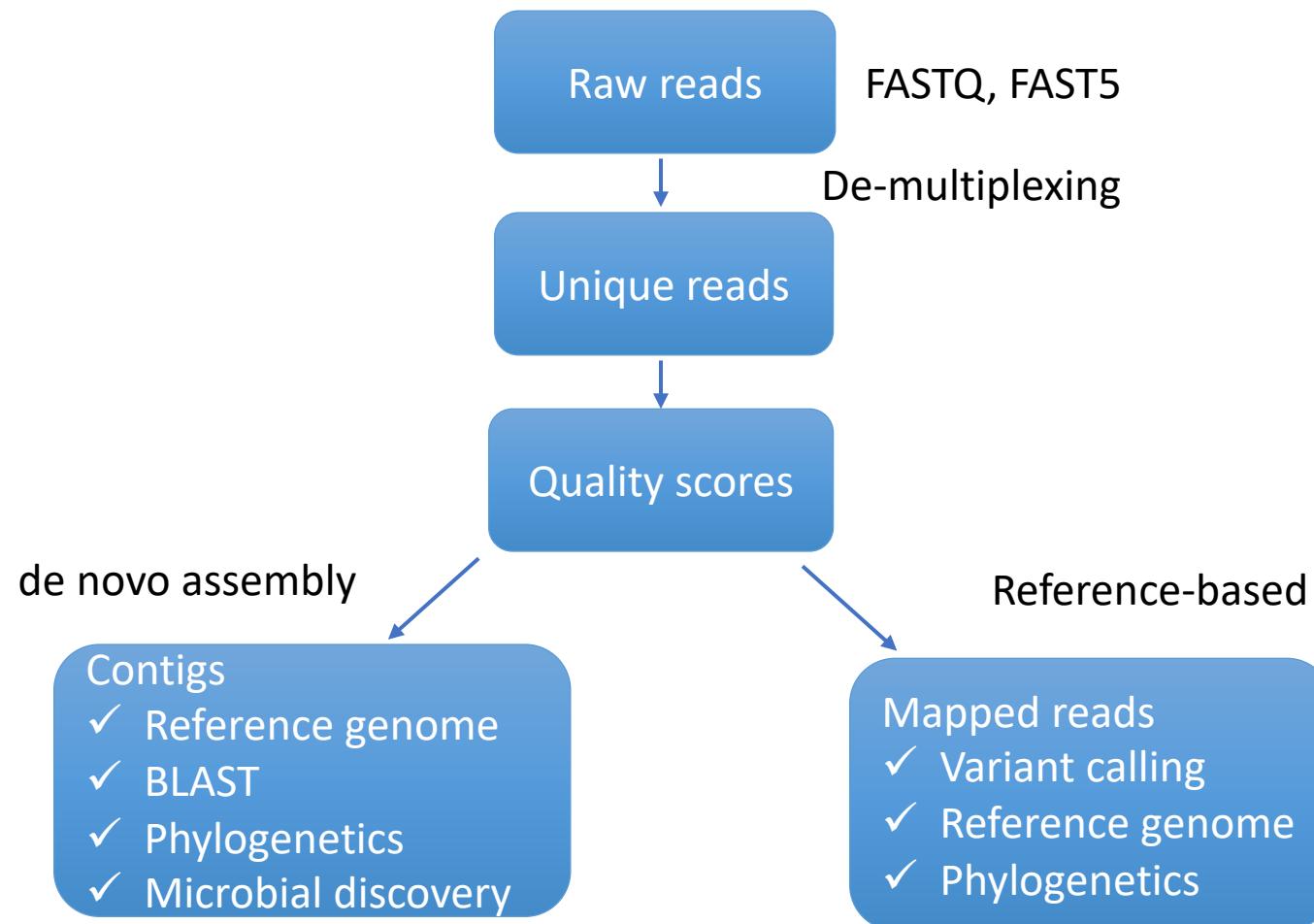
- Public health officials have raised an alarm of a new diarrheal disease within a certain sub-urban residence. Samples from affected individuals have tested negative for common diarrheal pathogens in the region. What sequencing approach would you use to identify the causative organism?



# Data analysis - Sanger



# Data analysis -NGS

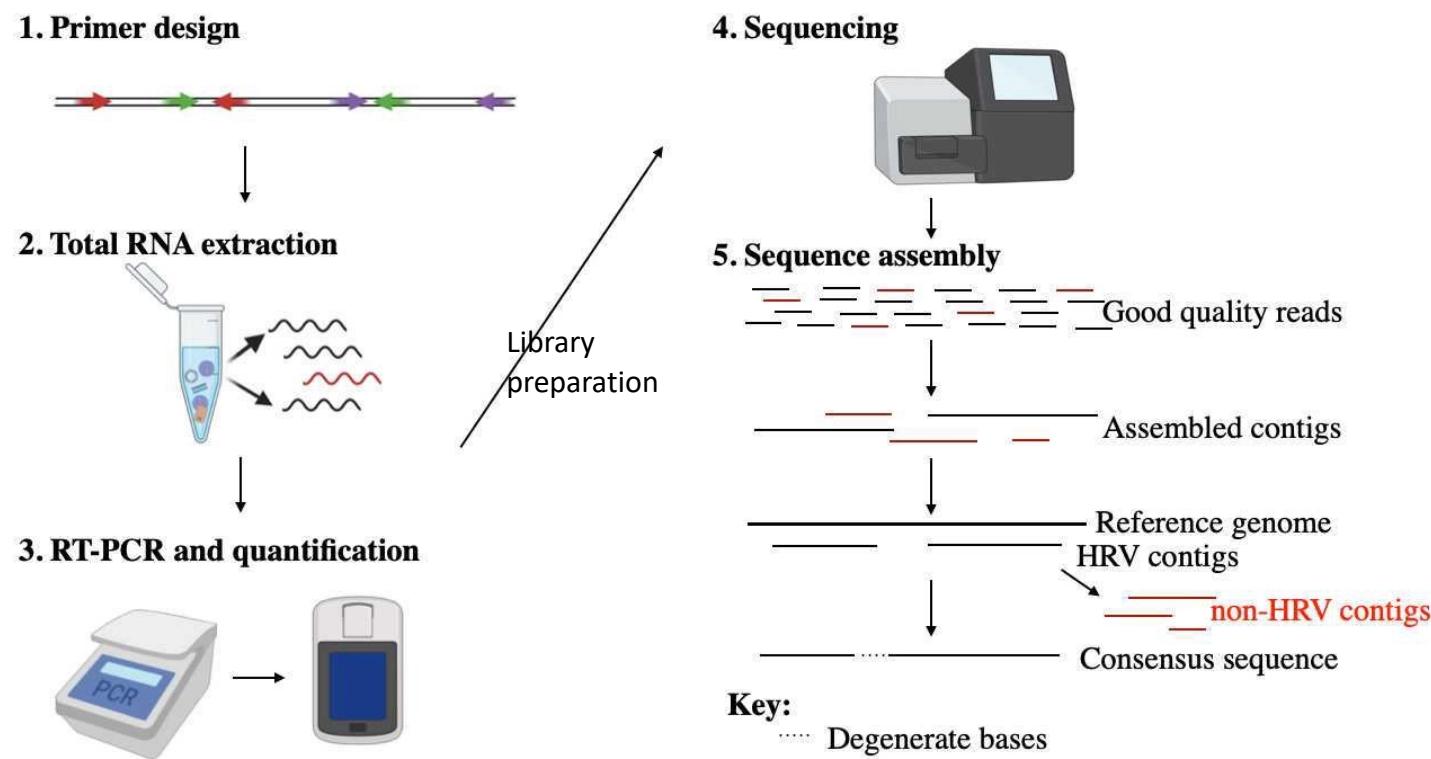


QC: Read depth

**Remember:** Even the best bioinformatics cannot save bad sequencing!



# A microbial NGS workflow



**Human rhinovirus sequencing workflow.** (1) Genotype-specific primer design, (2) extraction of total RNA, (3) RT-PCR using HRV primers, (4) short-read sequencing on the Illumina MiSeq and (5) reference-based sequence assembly, incorporating degenerate bases in regions not sequenced.



# Let's recap

- Intro to DNA sequencing
- First, second and third generation sequencing techniques
- Applications
- Template preparation
- How do they compare?
- NGS data analysis – an overview



*Questions?*

