

NGS platforms

Daniela Garcia Cobos
Comparative Genomics 2

What are “new-omics” sequencing platforms?

Sequencing platforms used in next generation sequencing (second and third generation)

Second generation sequencing: sequence an entire genome using short-read technology.

Third generation sequencing: long-read sequencing methodologies.

Aim to increase the sequencing throughput as well as the quality and length of the reads, while decreasing the time and cost of the process

Next- Generation Sequencing (NGS)

Methods: NGS platforms fragment DNA or RNA, then amplify and sequence these fragments simultaneously, enabling high-throughput **parallel sequencing** of short reads

Applications: Used for whole-genome sequencing, RNA-seq, exome sequencing, and targeted sequencing

Strengths: High accuracy, cost-effective for large-scale sequencing, and highly scalable, making it ideal for population studies

Limitations: Relies on short-read sequences (usually around 100–300 base pairs), which can make genome assembly challenging, especially for regions with high repetition

Platforms: Illumina, Ion Torrent, and BGI

Third generation sequencing (TGS)

Method: Produces long-read sequences (up to tens of thousands of base pairs per read) using a single-molecule sequencing, which doesn't require amplification.

Application: Ideal for resolving large structural variants or working with highly repetitive or complex genomes

Strengths: Genome assemblies are more accurate and easier.

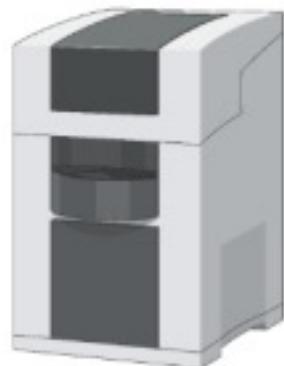
Limitations: Expensive and initially had high error rates (although this has been improved with recent technologies)

Platforms: Pacific Biosciences (PacBio) and Oxford Nanopore Technology (ONT)

DNA Sequencing Tools

1st Generation

Sanger Sequencing
500-1000bp



Advancing Genomics, Improving Life

2nd Generation

Illumina
50-500bp



3rd Generation

Oxford Nanopore Technology
Pacific Biosciences
10s of kb



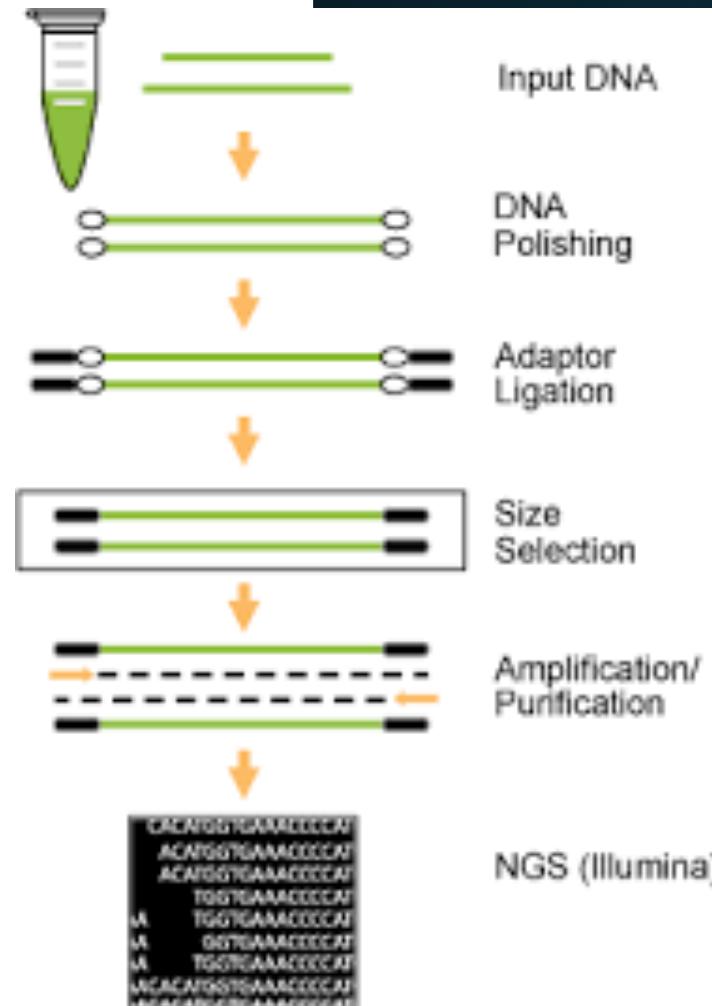
Novogene

NGS (short reads) preparation

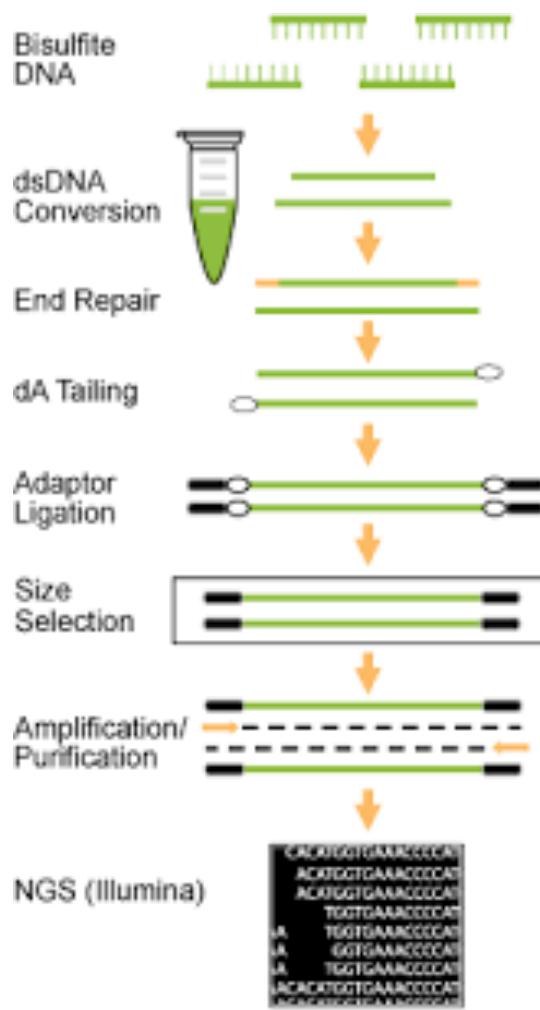
1) DNA extraction

2) Library preparation: Obtains templates corresponding to molecules of interest for sequencing (PCR and/or hybridization capture-based approaches) and prepares the fragments to make them compatible with the sequencing platform used.

- DNA fragmentation
- End repair
- Adaptor ligation
- Size selection



NGS (short reads) preparation (DNA)

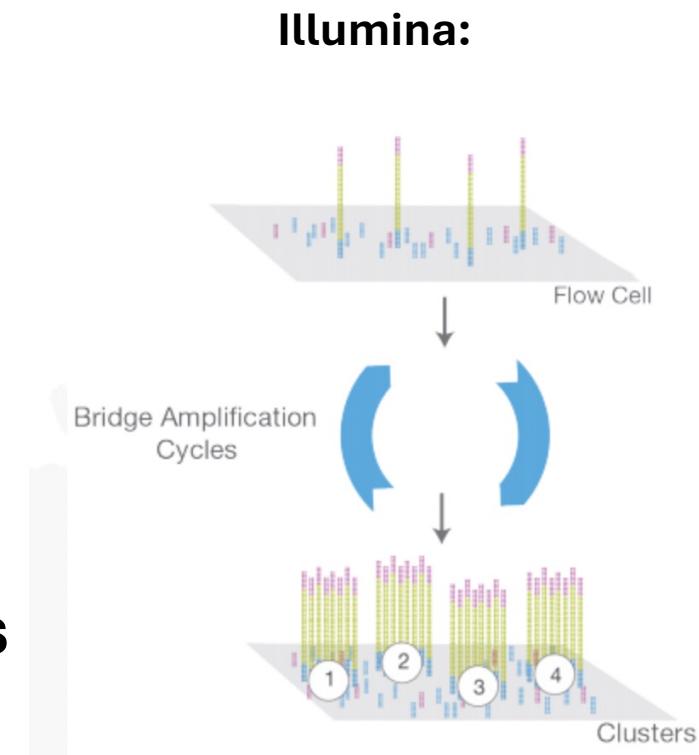
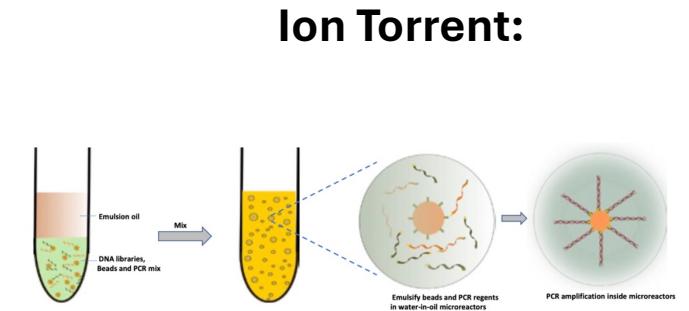


- **DNA fragmentation:** shears DNA into the optimal platform-specific size range (e.g. sonication, enzymatic or chemical)
- **End- repair:** prepares libraries for adaptor ligation by ensuring DNA fragment ends are free of overhangs that contain 5' phosphate and 3' hydroxyl groups.
- **Adaptor ligation:** include platform-specific sequences for fragment recognition by the sequencing instrument). Includes a barcode or index, to identify individual samples (pooling).
- **Size selection:** enriches DNA fragments within a defined size range and removes contaminants to improve sequencing efficiency (bead based or electrophoretic-based).

Short read-sequencing platforms

Include two consecutive general steps:

- 1) Clonal amplification:** amplifies DNA fragments to produce strong, detectable signals during sequencing. DNA fragments bind to a solid surface (e.g. beads, or flow cells)
- 2) Sequencing by synthesis**



Illumina sequencing

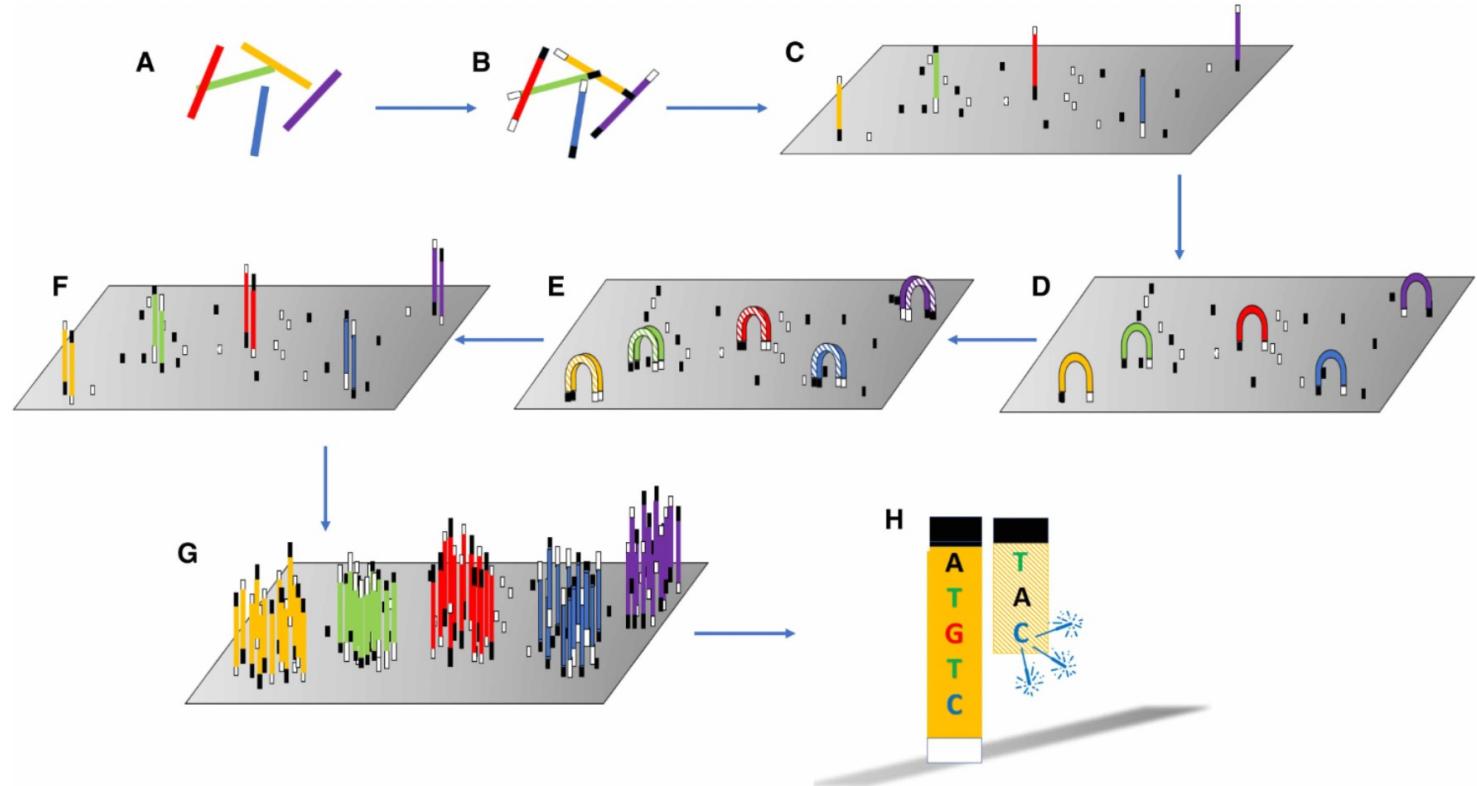


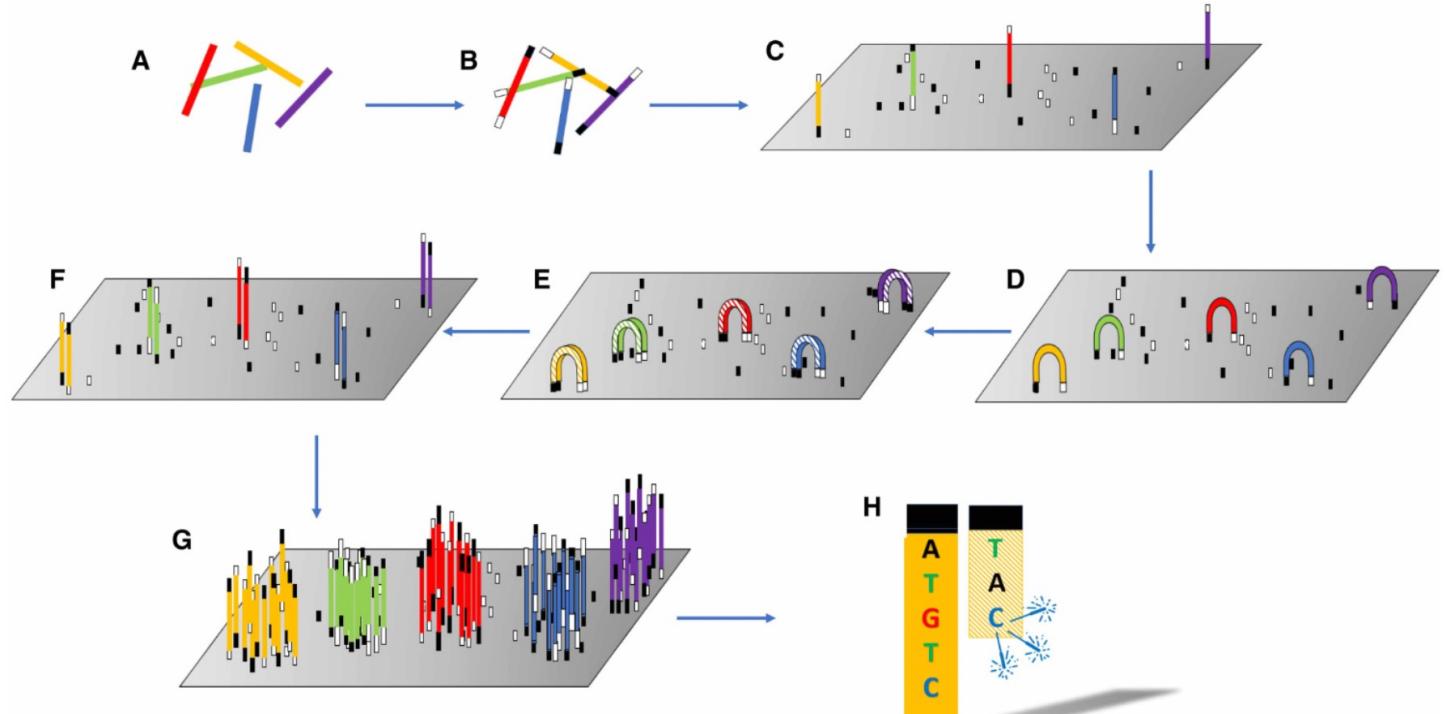
Figure 3. Principles of Illumina sequencing.

(A) fragmentation; (B) Adapters annealed; (C) Fragments bind to glass surface loaded with complementary primers; (D) Bridge formation; (E) Bridge amplification; (F) Dissociation; (G) Cluster formed from repeated bridge formation and amplification; (H) Sequencing by synthesis and fluorescence detection.

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

Incorporates fluorescently labeled nucleotides and detects each nucleotide as it's added to the DNA strand

Illumina sequencing



- The most used NGS platform currently used for medicine and research.
- Produces a forward and reverse read which increases the depth of sequencing
- Sequencing run quality on the Illumina instruments is assessed using three indicators: cluster density, percentage of clusters passing filters, and percentage of base calls with a quality score of at least Q30 (1 in 1000 probability of an incorrect base call).

Illumina sequencing

Illumina platform up to 2018:



MiniSeq System

Power and simplicity
for targeted sequencing.



MiSeq Series

Small genome and
targeted sequencing.



NextSeq Series

Everyday genome, exome
transcriptome sequencing,
and more.



HiSeq Series

Production-scale genome,
exome, transcriptome
sequencing, and more.



HiSeq X Series

Population- and production-
scale human whole-genome
sequencing.



NovaSeq Series

Population- and production-scale
genome, exome, transcriptome
sequencing, and more.

Ion Torrent sequencing

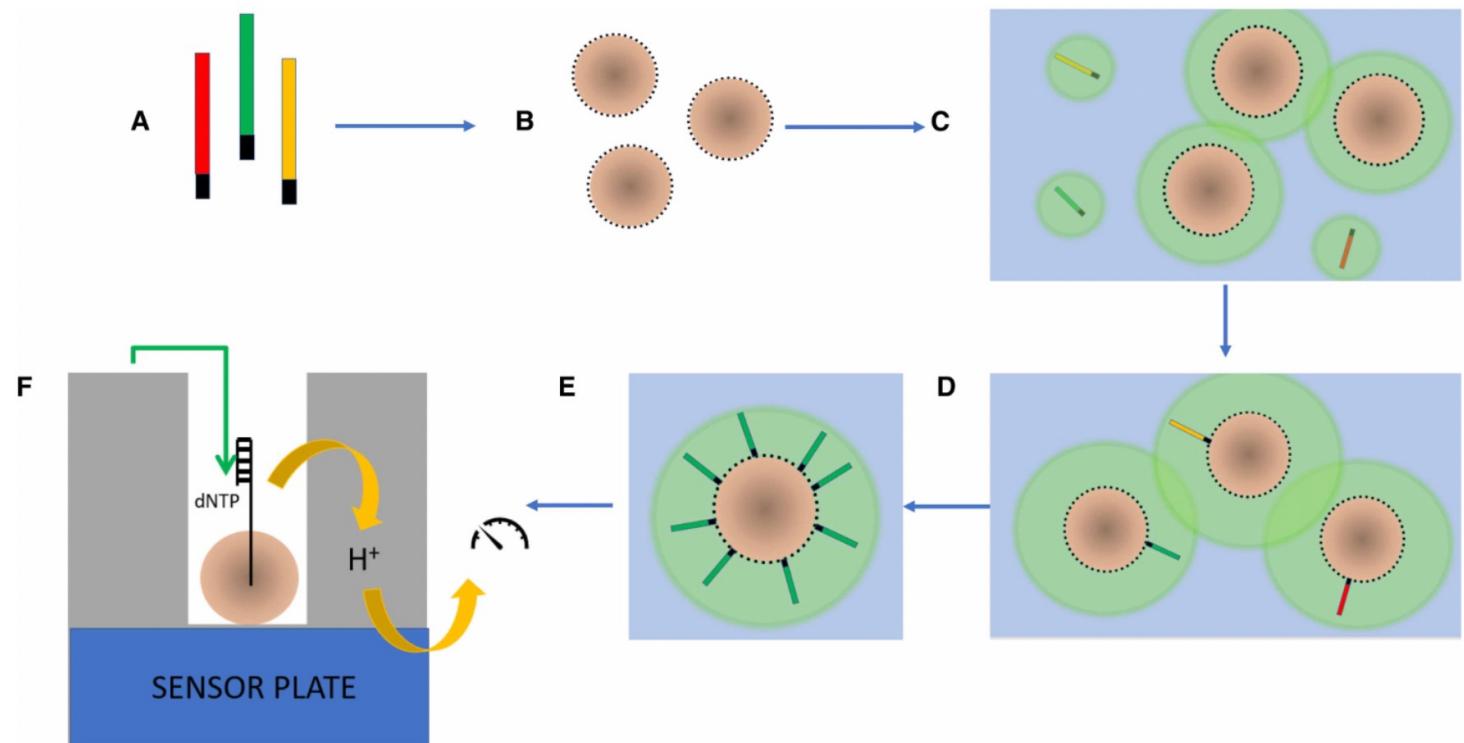
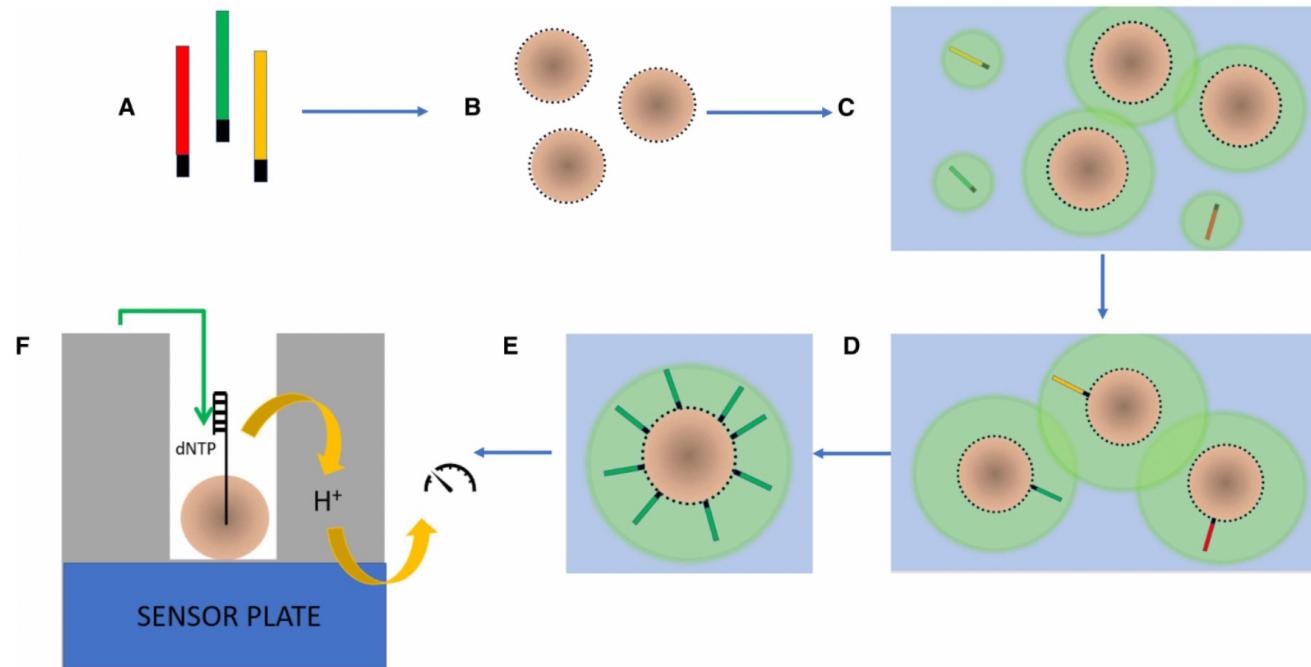


Figure 4. Principles of ThermoFisher Ion Torrent sequencing.

(A) Adaptors annealed to fragmented library; (B) Beads prepared with primers attached to surface; (C) Beads and library mixed in an aqueous/oil mixture; (D) Dilutions prepared so that only a single fragment anneals to a single bead; (E) Standard PCR cycling on the emulsion enables clones of amplified bead-bound oligonucleotides to be developed with up to 100 000 copies on each bead; (F) Detector chips with millions of wells, each holding a single bead, enables massively parallel sequencing of the bound oligonucleotides within the well. Different dNTPs are introduced to the chip in sequence and when one binds to a complementary nucleotide a hydrogen ion is released, resulting in a change of pH.

Measures changes in pH as each nucleotide is added to the growing DNA strand

Ion Torrent sequencing



- The amount of amplification is entirely dependent on the diameter of the beads:
- The 454 sequencing platform used beads that were 28 microns in diameter, and these were capable of generating up to 1 million copies.
- Ion Torrent uses beads of 5 microns amplifying the original DNA fragment into around 50,000 copies

Comparing Illumina with Ion Torrent

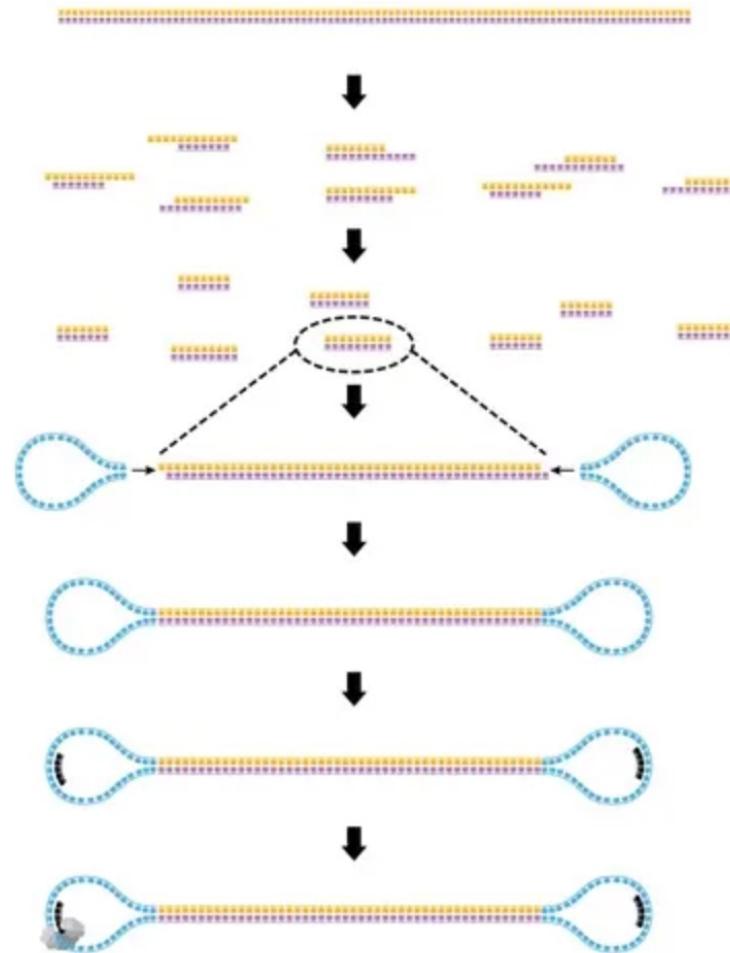
Short-read Sequencing Platforms and various characteristics.

Company	Illumina								ThermoFisher		
System Platform	iSeq	Miniseq	MiSeq	NextSeq550	NextSeq 1000&2000	NovaSeq 6000	MiSeqDx	NextSeq550 Dx	GeneStudio S5	Genexus	Ion PGM-Dx
Sequencing Principle	Sequence by Synthesis										
Detection	Fluorescent								Ion		
Applications	Small WGS, TS, Small RNA sequencing	Small WGS, TS, ChIP-Seq, Small RNA sequencing	Small WGS, TS, ChIP-Seq, Small RNA sequencing	TS, small WGS, exome and transcriptome sequencing		TS, WGS, WES, transcriptome and epigenome sequencing	TS, Small WGS	TS, exome and transcriptome sequencing	TS, epigenetic, exome, and transcriptome sequencing	TS	TS
Maximum Read length (bases)	2 × 150 bp		2 × 300 bp	2 × 150 bp	2 × 150 bp	2 × 250 bp	2× 300 bp	2× 150 bp	600 bp	400 bp	200 bp
Flow cells/device	1					2	1				
Output (per flow cell)	1.2 Gb	7.5 Gb	15 Gb	120 Gb	330 Gb	3000 Gb	≥5 Gb	≥90 Gb	15 Gb	24 Gb	1 Gb
Sequencing Run time	9.5-19 hr	5-24 hr	4-56 hr	11-29 hr	11-48 hr	13-44 hr	24 hr	≤35 hr	4.5-21.5 hr	14-31 hr	4.4 hr
Accuracy/Quality	Q30≥ 80% (2 × 150)		Q30≥ 70%	Q30≥ 75% (2 × 150 bp)		Q30≥ 75%	Accuracy	Accuracy	Accuracy	Accuracy	Accuracy
Score	bp)		(2 × 300 bp)			(2 × 250 bp)	>99.66%, Q30> 80%	≥99.98%, Q30≥75%	≥99%	≥99%	≥99%
Equipment Cost (USD)	\$19,900	\$49,500	\$99,000	\$275,000	\$335,000	on request					

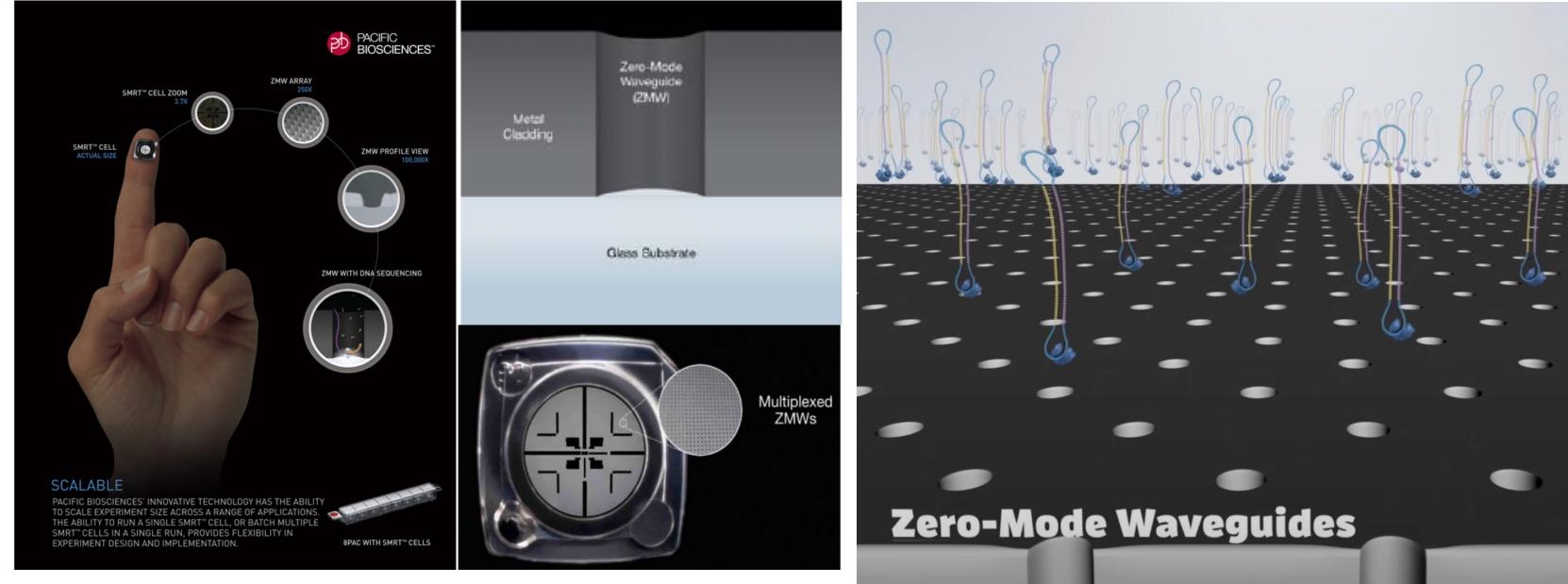
Illumina vs Ion Torrent

Platform	Advantages	Disadvantages
Illumina	<ul style="list-style-type: none">- High accuracy- High Throughput- Many platforms- Captures GC-Rich regions- Better for whole genomes	<ul style="list-style-type: none">- More expensive- Takes longer to run
Ion Torrent	<ul style="list-style-type: none">- Runs faster- Lower cost for operation- Adequate for target sequencing	<ul style="list-style-type: none">- Higher error rate in homopolymers (repeats)- Limited for whole genomes- Not effective in GC regions

PacBio Library preparation



PacBio Sequencing



Polymerase
incorporates
labeled
nucleotides light
is emitted

Nucleotide
incorporation is
measured in real
time

Table 2

Long-read Sequencing Platforms and various characteristics.

Company	Pacific Biosciences			Oxford Nanopore					
System Platform	Sequel	Sequel II	Sequel IIe	Flongle	MinION	GridION	PromethION		
Sequencing Principle	PacBio Single Molecule Sequencing			Nanopore Single molecule Sequencing					
Detection	Fluorescent			Electrical Conductivity					
Applications	Whole genome <i>de novo</i> assembly, variant detection, structural variation detection, full length transcript sequencing, targeted/amplicon sequencing, metagenomics sequencing			DNA, amplicons, cDNA, Direct RNA sequencing					
Maximum Read length (bases)	300 kb			Longest read so far: > 4 Mb					
Flow cells/device	12 SMRT Cells 1M can be used at a time, and 8 SMRT Cell 8M can be used serially			1 (126 channels per flow cell)	1 (512 channels per flow cell)	5 (512 channels per flow cell)	24 or 48 (3000 channels per flow cell)		
Output (per flow cell)	75 Gb	600 Gb	1 - 2 Gb ^a	10 - 30 - 50 Gb ^a		100 - 200 - 300 Gb ^a			
Sequencing Run time	Up to 20 hr	Up to 30 hr	1 min - 16 hr	1 min - 72 hr					
Accuracy/Quality Score	Number of HiFi Reads >99% Accuracy: Up to 5,000,000 reads	Number of HiFi Reads >99% Accuracy: Up to 4,000,000 reads	Single Molecule: R9 modal Accuracy >98.3%, R10 modal Accuracy >97.5%. New chemistry Accuracy >99% (coming soon) Consensus: R9.4.1: Current best Q45 (>99.99%) R10: Current Best Q50 (99.999%)						
Equipment Cost (USD)	approximately \$525,000			\$1,460 (12 flow cells included)	\$9,300	\$69,955	24 flow cells: \$335,455 48 flow cells: \$530,000		

Table 2

Long-read Sequencing Platforms and various characteristics.

Company	Pacific Biosciences			Oxford Nanopore				
System Platform	Sequel	Sequel II	Sequel IIe	Flongle	MinION	GridION	PromethION	
Advantages	Very long reads can help resolve ambiguities; no DNA amplification required, comparatively faster turnaround time				Fast-sequencing; Small instrument footprint; Portability; Real-time data analysis			
Disadvantages	Sequencing equipment is expensive, which could be cost-prohibitive for smaller clinical laboratories, large footprint of the equipment, historically higher error rate (continues to improve)				Historically higher error rate (continues to improve)			

How do I choose which sequencing approach/platform is optimum?

- Evolutionary time scale of the biological question?
- Resources available (terms of money, time and computational power)

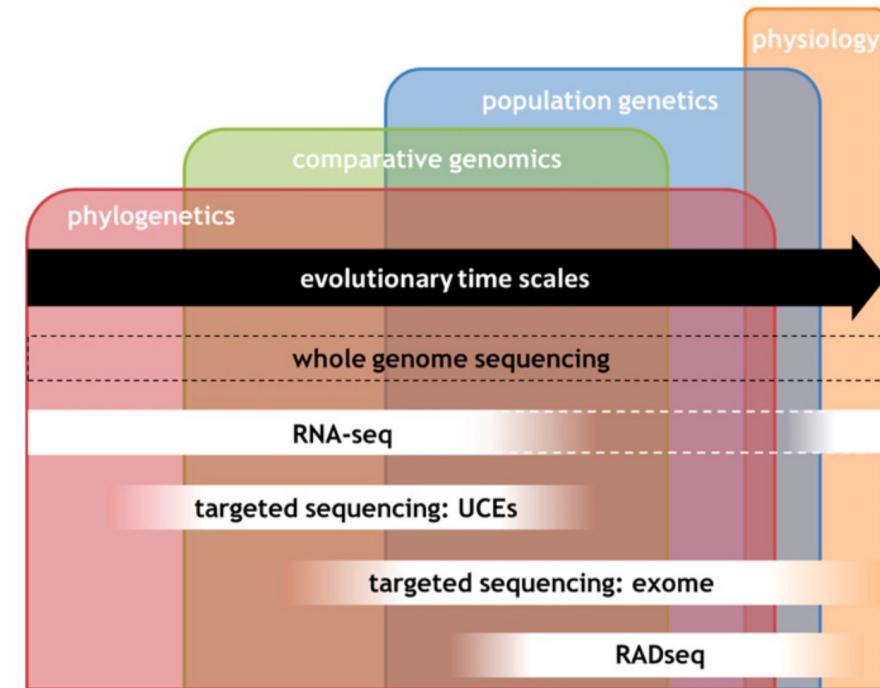


Fig. 1. Application of different high-throughput sequencing methods to different evolutionary time scales. Research applications related to evolutionary biology (colored