

Module 1: Introduction to Sequencing Technologies

Presented by:

Marcela Sjöberg Herrera

Assistant professor

Laboratory of Epigenetic Regulation

Faculty of Biological Sciences

Pontificia Universidad Católica de Chile

msjoberg@bio.puc.cl

Next Generation Sequencing Bioinformatics Course
23-27 January 2023 - Santiago - Chile



FACULTAD DE
CIENCIAS BIOLÓGICAS
PONTIFICIA UNIVERSIDAD
CATÓLICA DE CHILE



WELLCOME GENOME CAMPUS

CONNECTING SCIENCE
ADVANCED COURSES +
SCIENTIFIC CONFERENCES

Outline

- ❖ **History of DNA sequencing**
- ❖ **Next generation sequencing technologies (NGS)**
2005-present
 - **Short-read NGS**
 - **Long-read NGS**

Outline

❖ History of DNA sequencing

❖ Next generation sequencing technologies (NGS) 2005-present

- Short-read NGS
- Long-read NGS

DNA sequencing

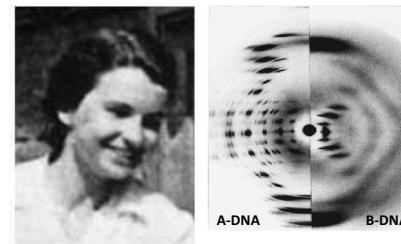
Discovery of the
DNA structure

Complexity and diversity
of genomes

Two techniques:

- Enzymatic: Sanger sequencing (1975)
- Chemical: Maxam & Gilbert sequencing (1977)

Rosalind Franklin (1952)



James Watson

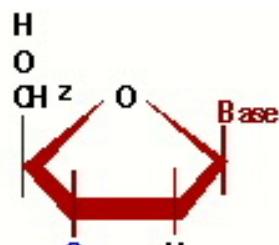


Francis Crick (1953)

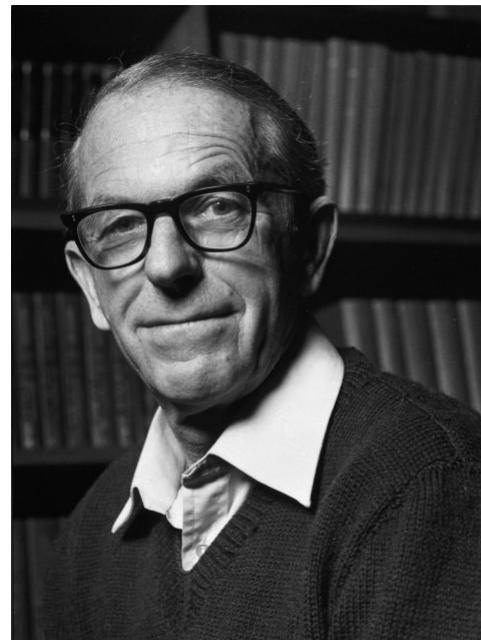
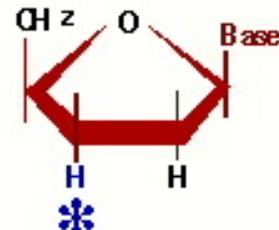
Sanger sequencing

- Cloning
- Primer hybridization
- DNA polymerase
- Radioactive dNTPs
- **Chain terminations (ddNTP):** the synthesis reaction is split in 4 tubes, each one with a different terminator to stop the synthesis at random positions.

Normal nucleotides:



Dideoxy Chain Terminators:



Frederick Sanger (1980)

Fragment to be sequenced, cloned in M13 phage

3' --- AG --- CT **GCTCGCAT** --- 5'

TC --- GA

Primer

↓
DNA polymerase
4 dNTPs (radioactive)
ddGTP

Synthesis of complementary second strands:

5' TC --- GA**CddG** 3'

5' TC --- GA**CGA ddG** 3'

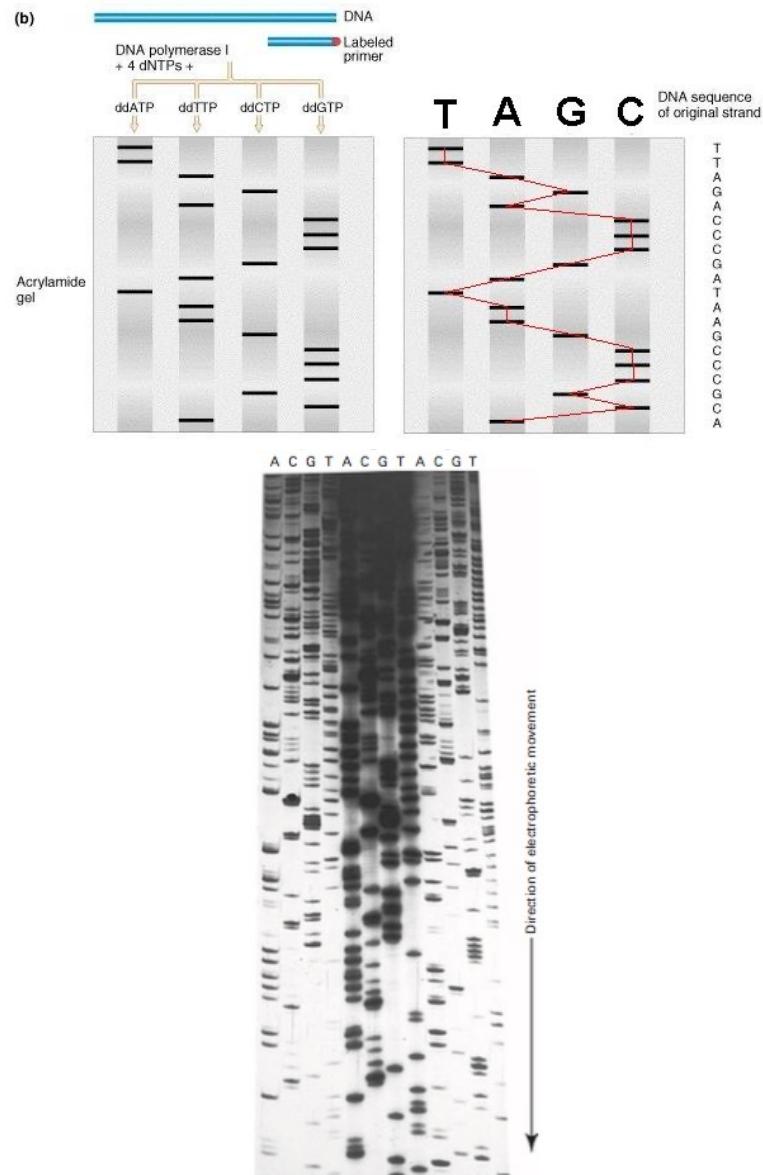
5' TC --- GA**CGAG CddG** 3'

Denature to give single strands

Run on sequencing gel alongside products of
ddCTP, ddATP and ddTTP reactions

Sanger sequencing

- **Random incorporation of ddNTPs**
-> DNA products \neq sizes
- **Denaturing polyacrylamide gels**
to separate the products according
to size
- **Autoradiography:** to determine
the DNA sequence based on
positions of radioactive bands
- The lengths of fragments are ~300
bases



Capillary electrophoresis and automated fluorescent gel sequencing

AB applied
biosystems™



ABI 3100



ABI 3130



ABI 3730XL

Capillary electrophoresis and automated fluorescent gel sequencing

- No radioactive dNTPS
- Fluorochrome labelled ddNTPs
- All in one reaction (not 4)
- Capillary electrophoresis (no gel)
- Light emission is interpreted by a computer
- The lengths of fragments are ~500 bp

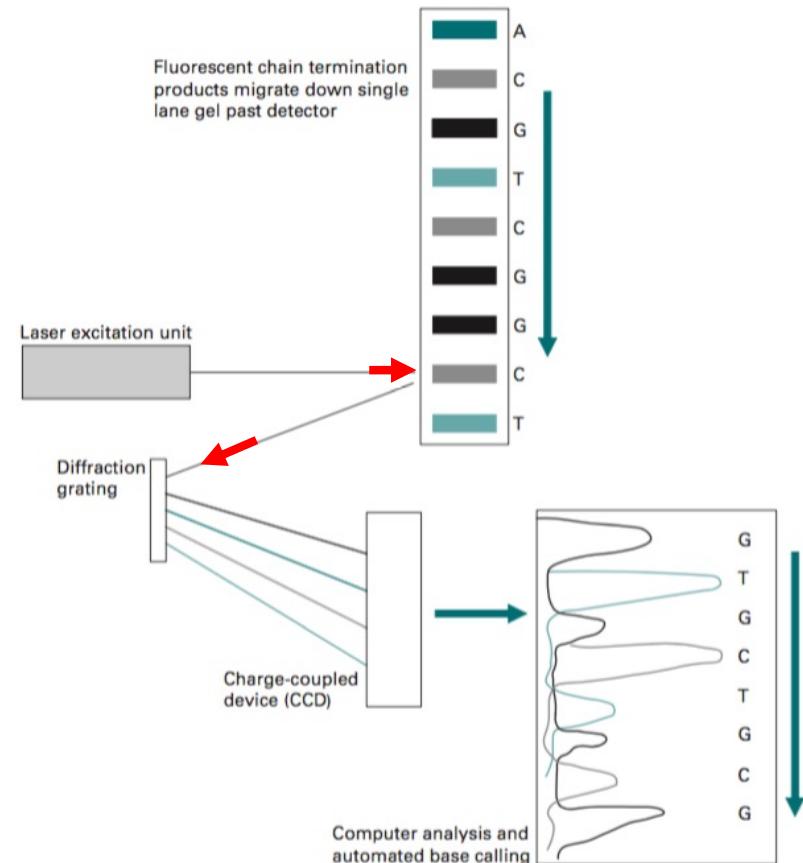
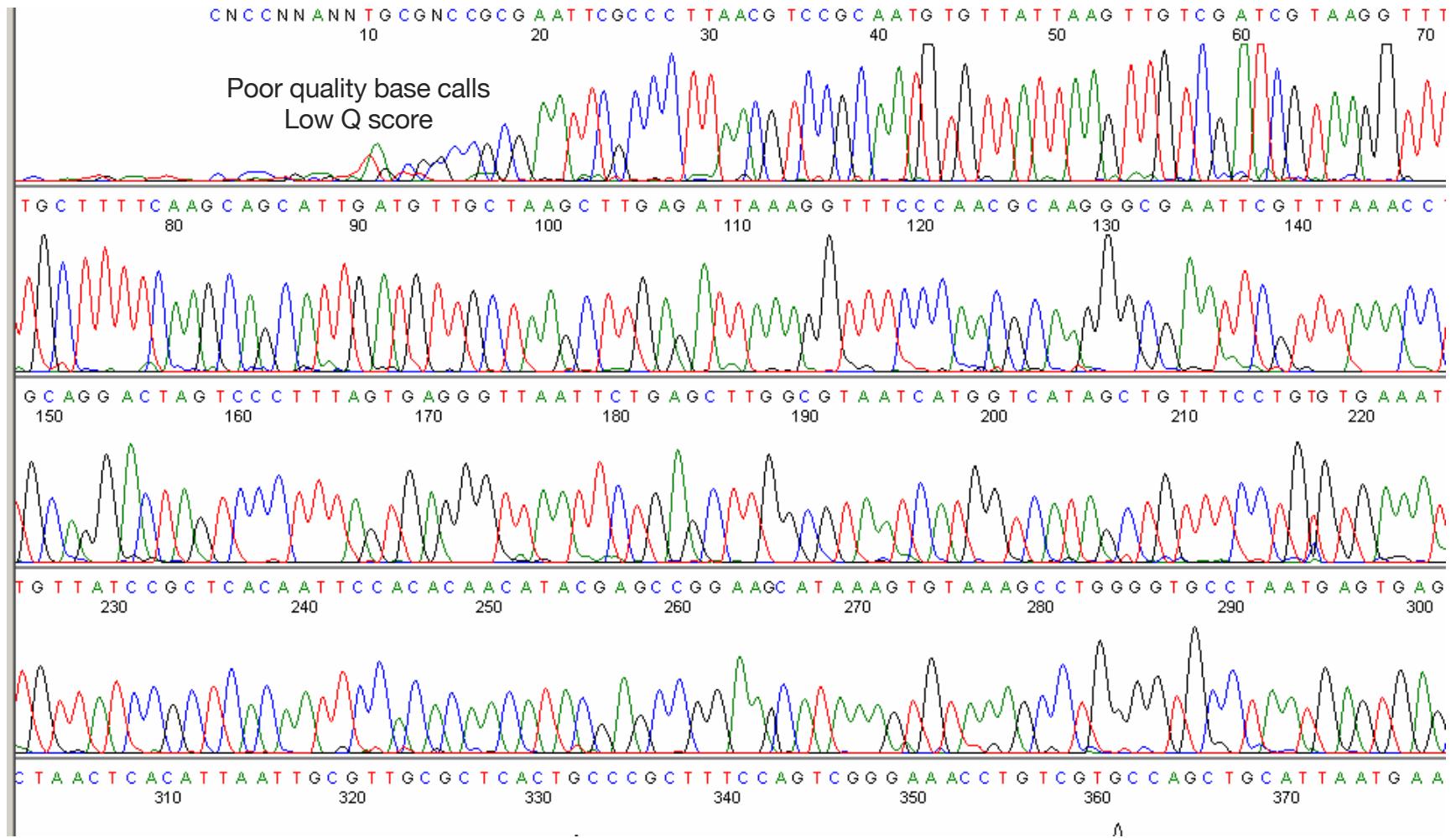


Fig. 5.40 Automated fluorescent sequencing detection using single-lane gel and charge-coupled device.

Capillary electrophoresis DNA sequence files



Capillary electrophoresis ABI 3730

- Individual reactions -> up to 96-capillary array
- Fast + easy for individual samples
- Read length: up to 900 bases
- Run time: 1-2 hour
- Accurate sequence calls per base, Q30 (Q score is reported on a log scale, so Q10 =1 error in 10, Q20 =1 error in 100, Q30=1 error in 1000, ...Q50)
- Robust technology and cheapest for specific targets (clinical diagnostic)
- Capillary sample prep: PCR (errors) or Miniprep (Cloning bias)
- Limited yield: 100 kb/run
- **\$1,000,000/Gb**
- \$200k per instrument

Outline

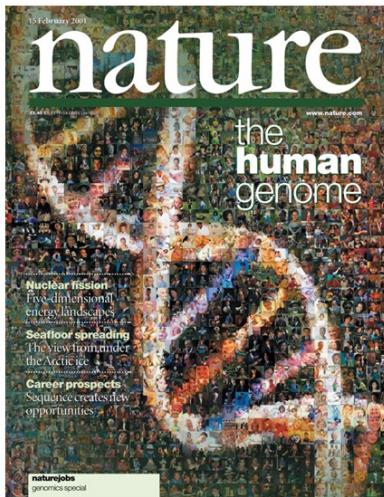
- ❖ History of DNA sequencing
- ❖ Next generation sequencing technologies (NGS)
2005-present
 - Short-read NGS
 - Long-read NGS

Next-generation sequencing (NGS)

Considerable technical innovations supported the Human Genome Project (1998-2003)

More advanced sequencing technologies were needed to answer new biological questions

NGS is massively parallel, not limited to few reactions per run

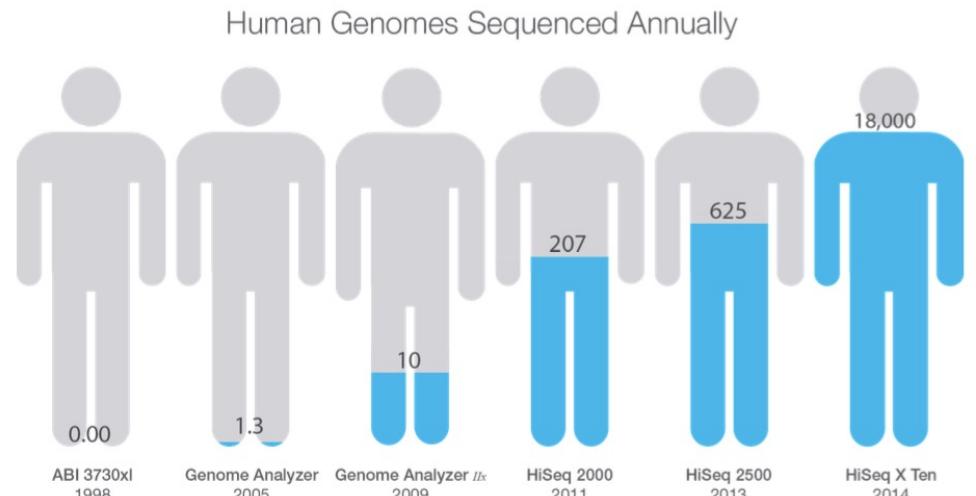
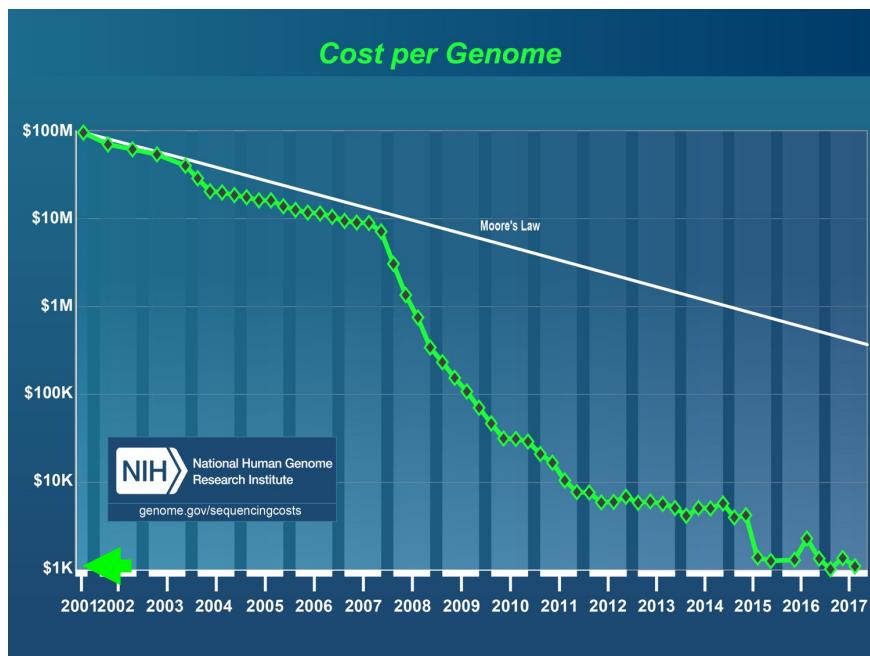


Next-generation sequencing (NGS)

Different modern sequencing technologies, developed from the mid 2000s, that allow DNA and RNA sequencing much more quickly and cheaply, which are revolutionizing the study of genomics, transcriptomics and epigenomics.

Solexa/Illumina Genome Analyzer (2006): first high-throughput sequencing platform

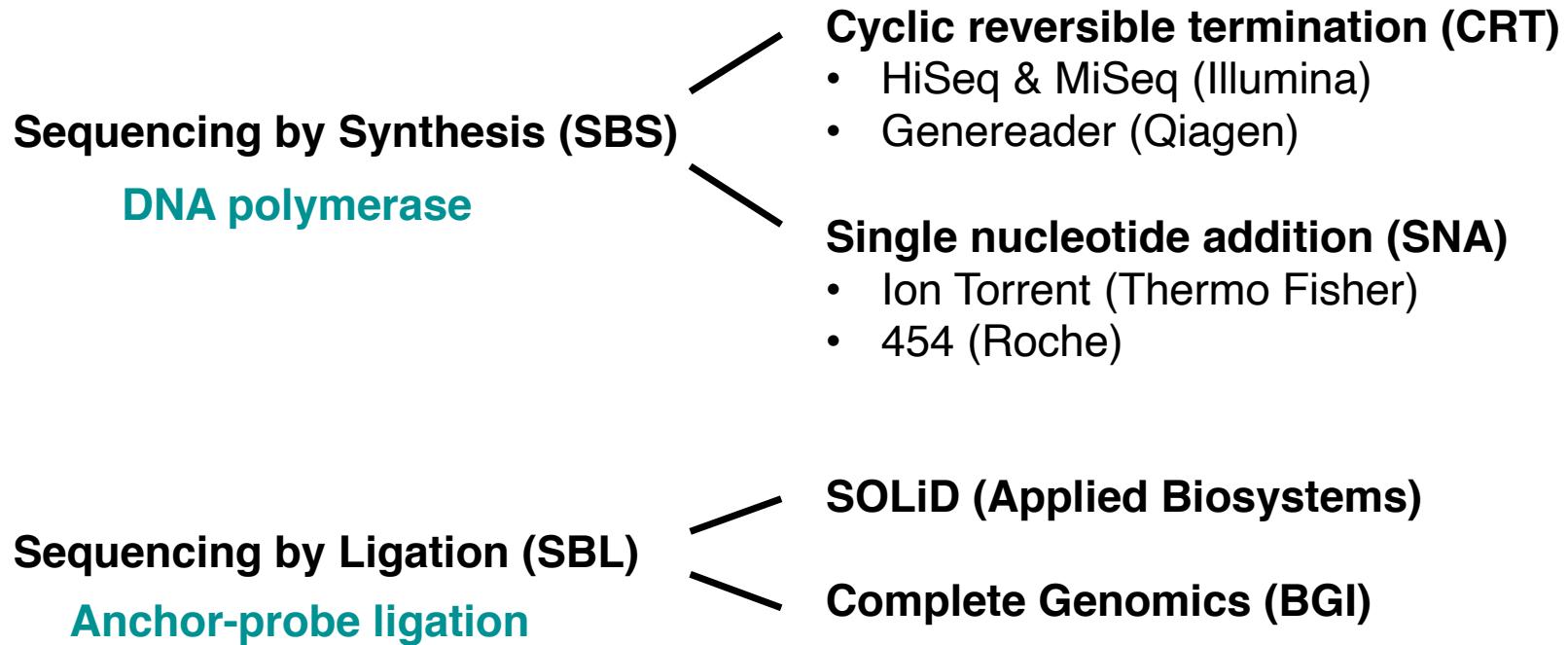
Illumina HiSeq X® Ten System (2014): ultra-high-throughput ~ 45 human genomes a day



Outline

- ❖ History of DNA sequencing
- ❖ Next generation sequencing technologies (NGS)
 - **Short-read NGS**
 - Long-read NGS

Short read NGS



Short read NGS

Sequencing by Synthesis (SBS)

DNA polymerase

Cyclic reversible termination (CRT)

- HiSeq & MiSeq (Solexa/Illumina)
- Genereader (Qiagen)

Single nucleotide addition (SNA)

- Ion Torrent (Thermo Fisher)
- 454 (Roche)

Sequencing by Ligation (SBL)

Anchor-probe ligation

SOLiD (Applied Biosystems)

Complete Genomics (BGI)

Solexa

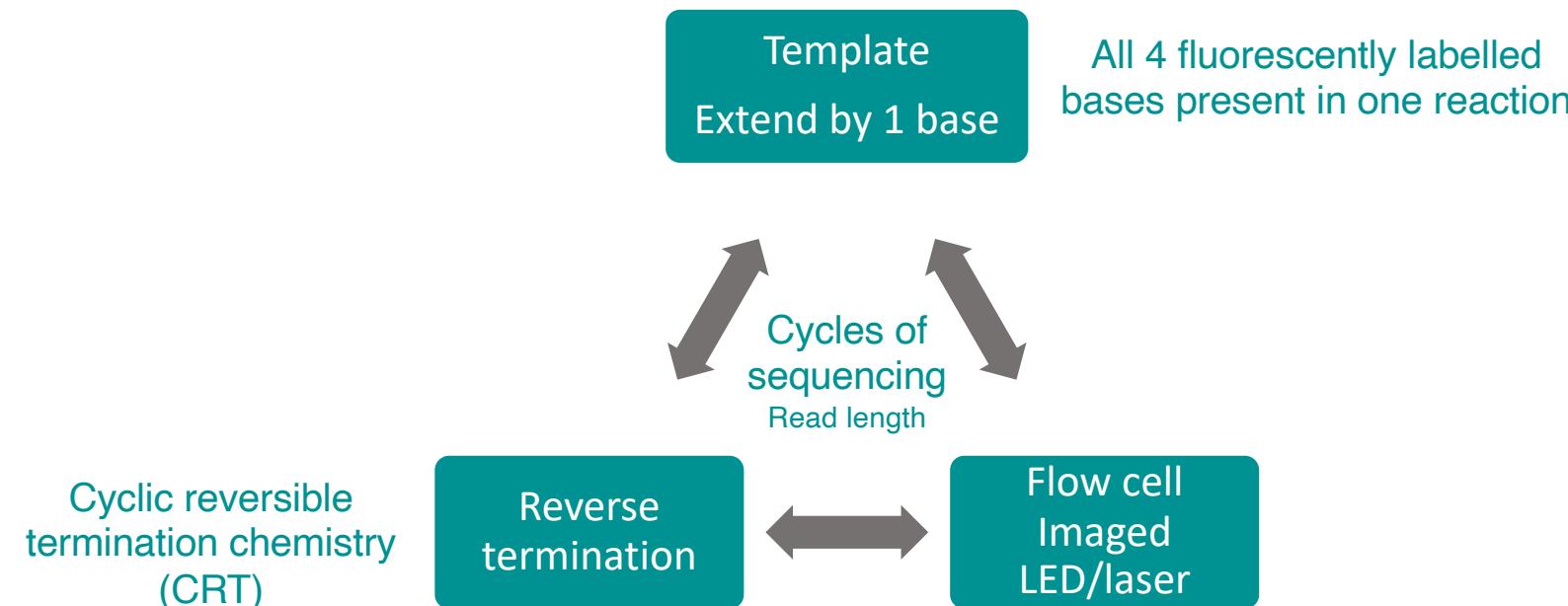
- Spinout from Cambridge University in 2000
- Launched their Genome Analyzer in 2006
- Genome Analyzer at launch: 1Gb/run. Now up to 6Tb
- Acquired by Illumina in 2007 -> range of **Illumina Sequencing Technology Platforms**

illumina®

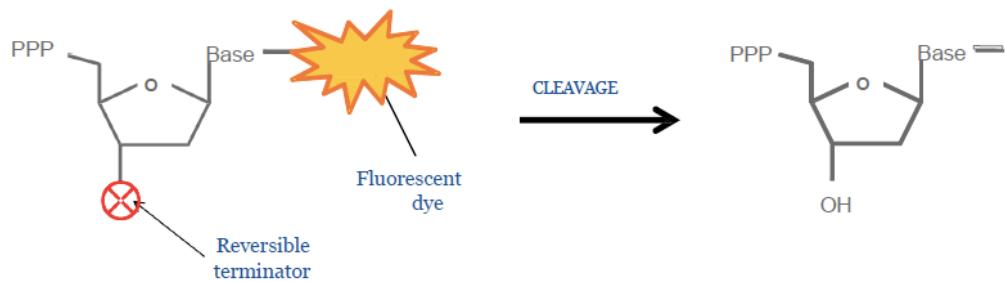


<https://www.youtube.com/watch?v=fCd6B5HRaZ8&feature=youtu.be>

Illumina Sequencing by Synthesis (SBS)

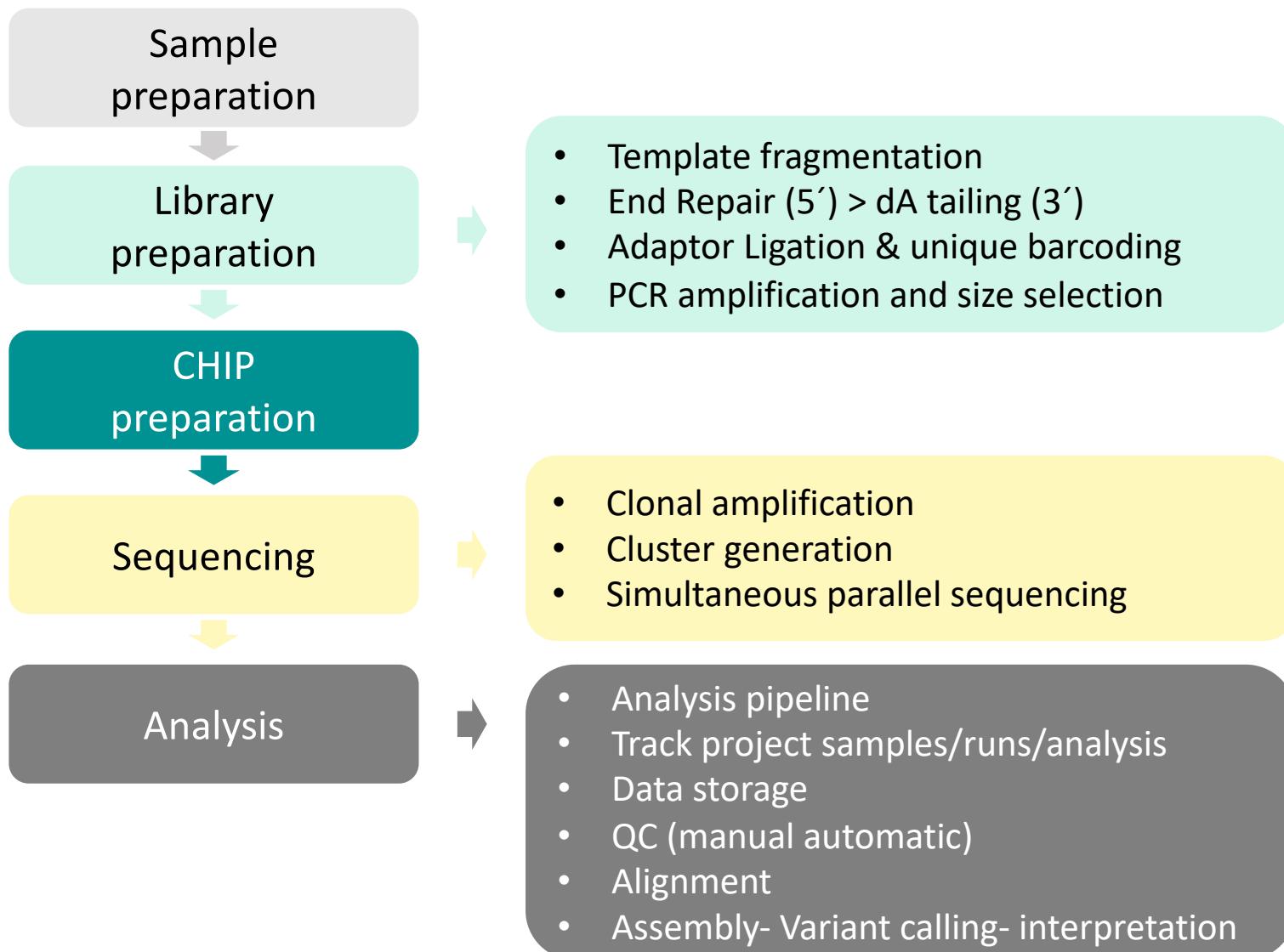


Illumina modified NTP:



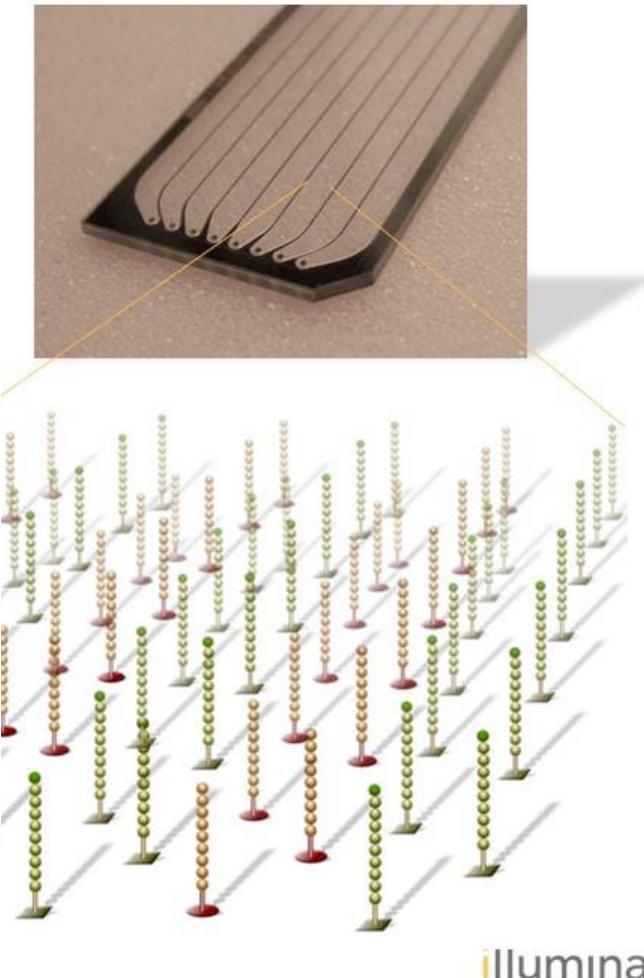
- ✓ Faster
- ✓ Accurate
- ✓ No problems with homopolymers

Illumina Workflow

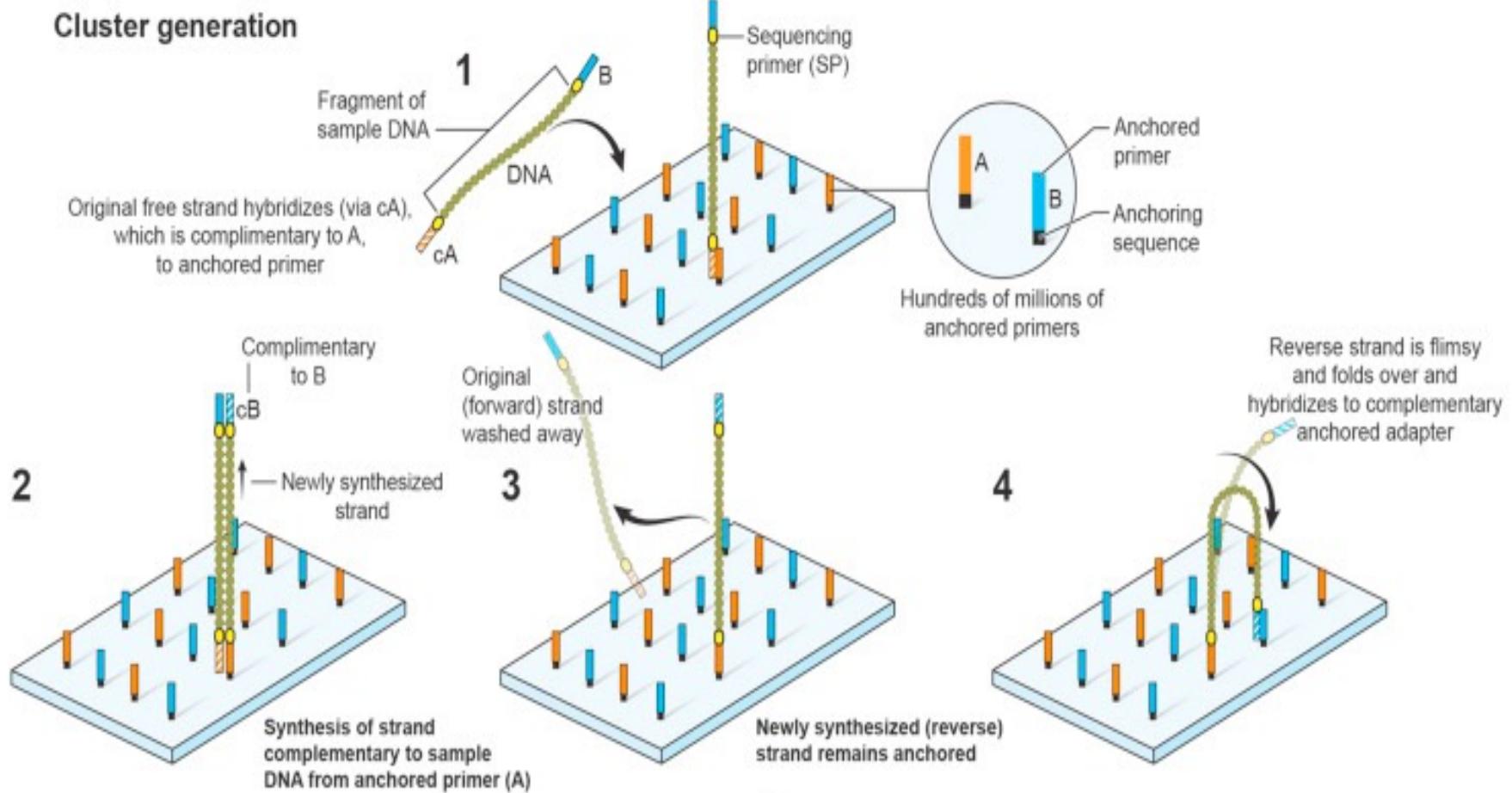


Illumina Flow Cell

- Is where a cluster generation occurs
- A thick glass slide with lanes
- Each lane coated with oligos that are complementary to library adapters



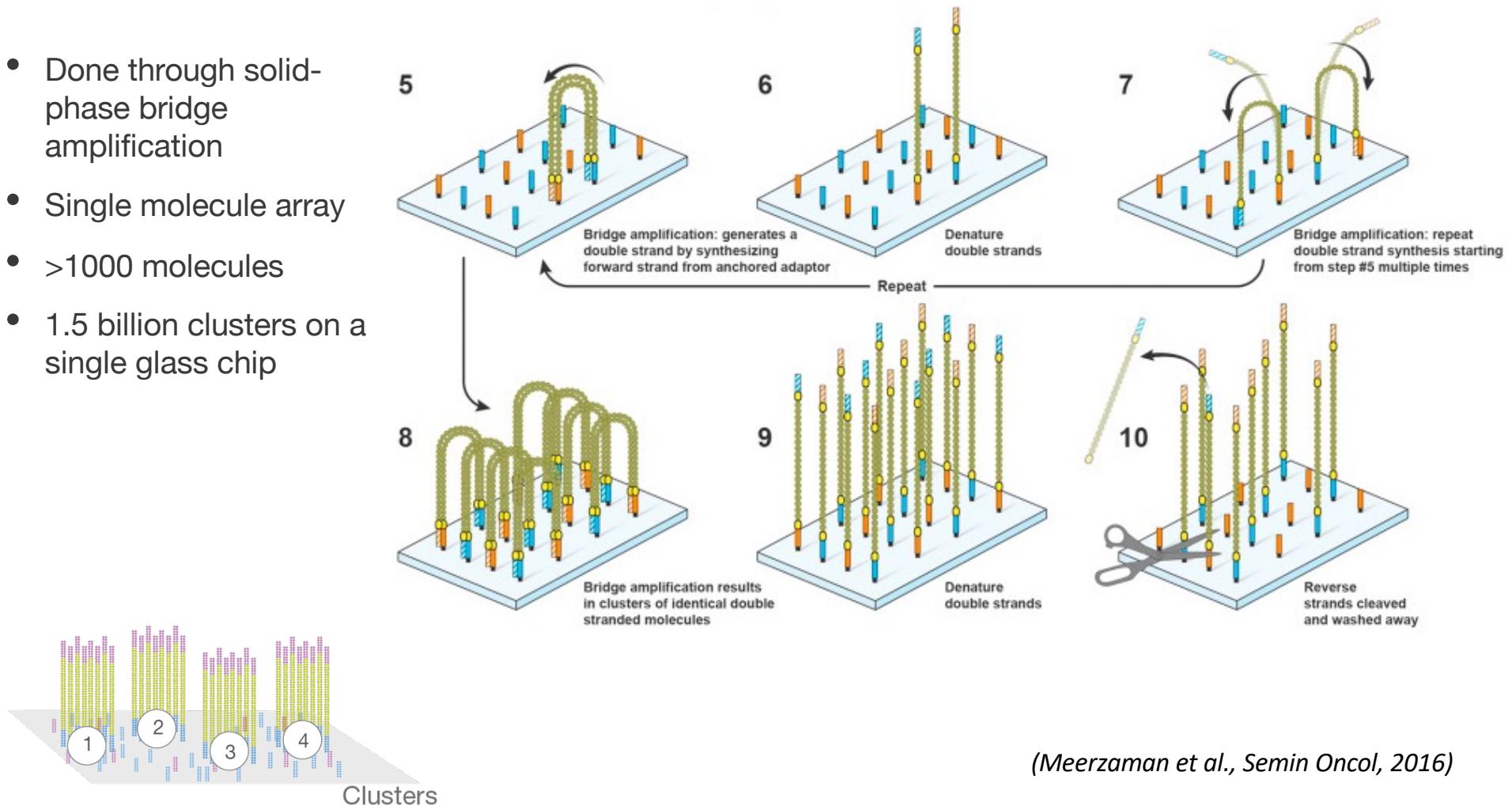
Illumina Cluster Generation



(Meerzaman et al., Semin Oncol, 2016)

Illumina Clonal Amplification

- Done through solid-phase bridge amplification
- Single molecule array
- >1000 molecules
- 1.5 billion clusters on a single glass chip



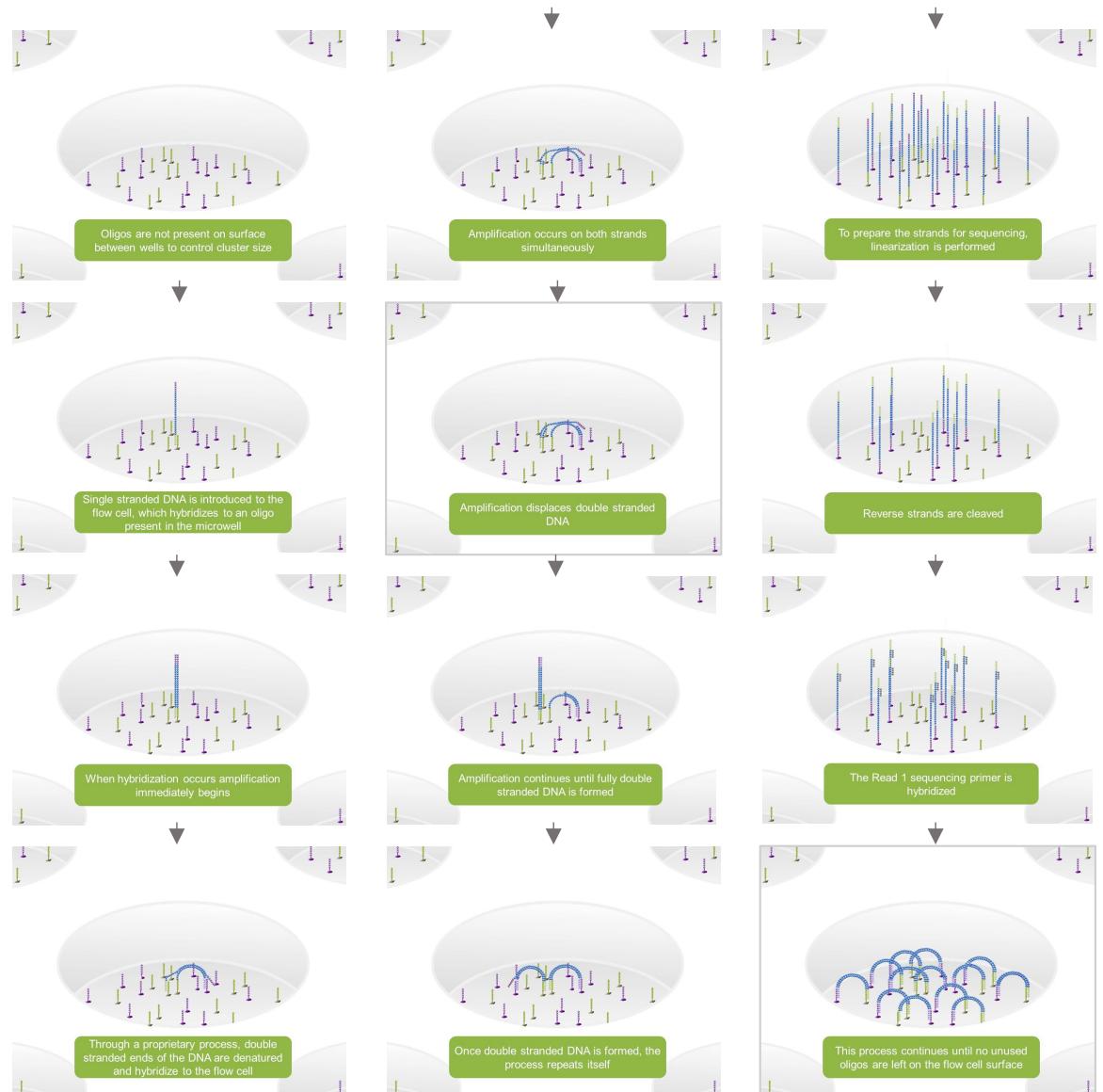
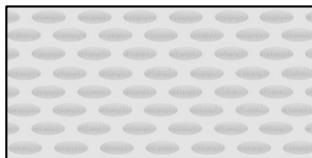
Illumina Patterned Flow Cell Technology

- iSeq100, HiSeq4000, Xten, NovaSeq

- Oligos are present inside patterned microwells to control cluster size

- Clusters of defined size and spacing are formed within ordered microwells

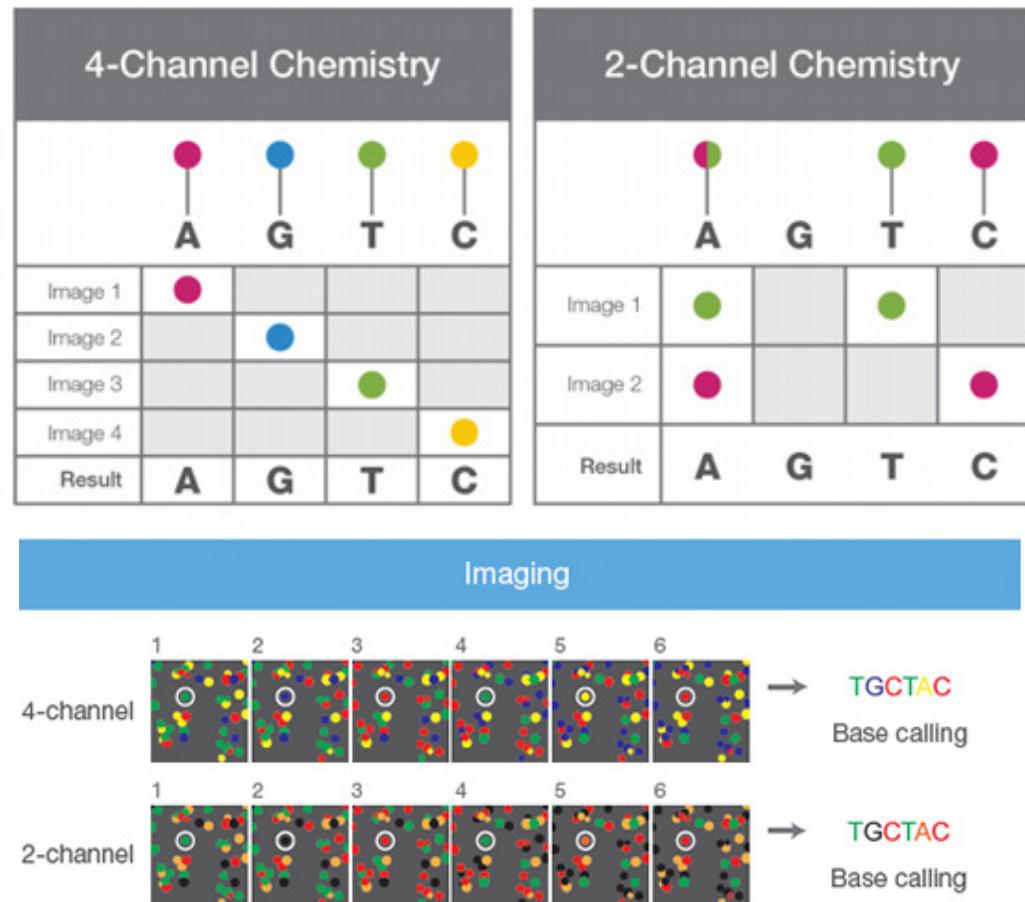
- Higher cluster density, simplified imaging, faster registration due to skipping template generation



Illumina sequencing technology

4 colour vs 2 colour chemistry

- **4-channel system:** GA, HiSeq, MiSeq, NovaSeq...
- **2-channel system:** iSeq100, NextSeq 500b and MiniSeq
- Faster DNA sequencing
- Shorter processing times
- Slightly higher error profile (poly G artifacts)
- Underperformance for low-diversity samples owing to more ambiguous base discrimination

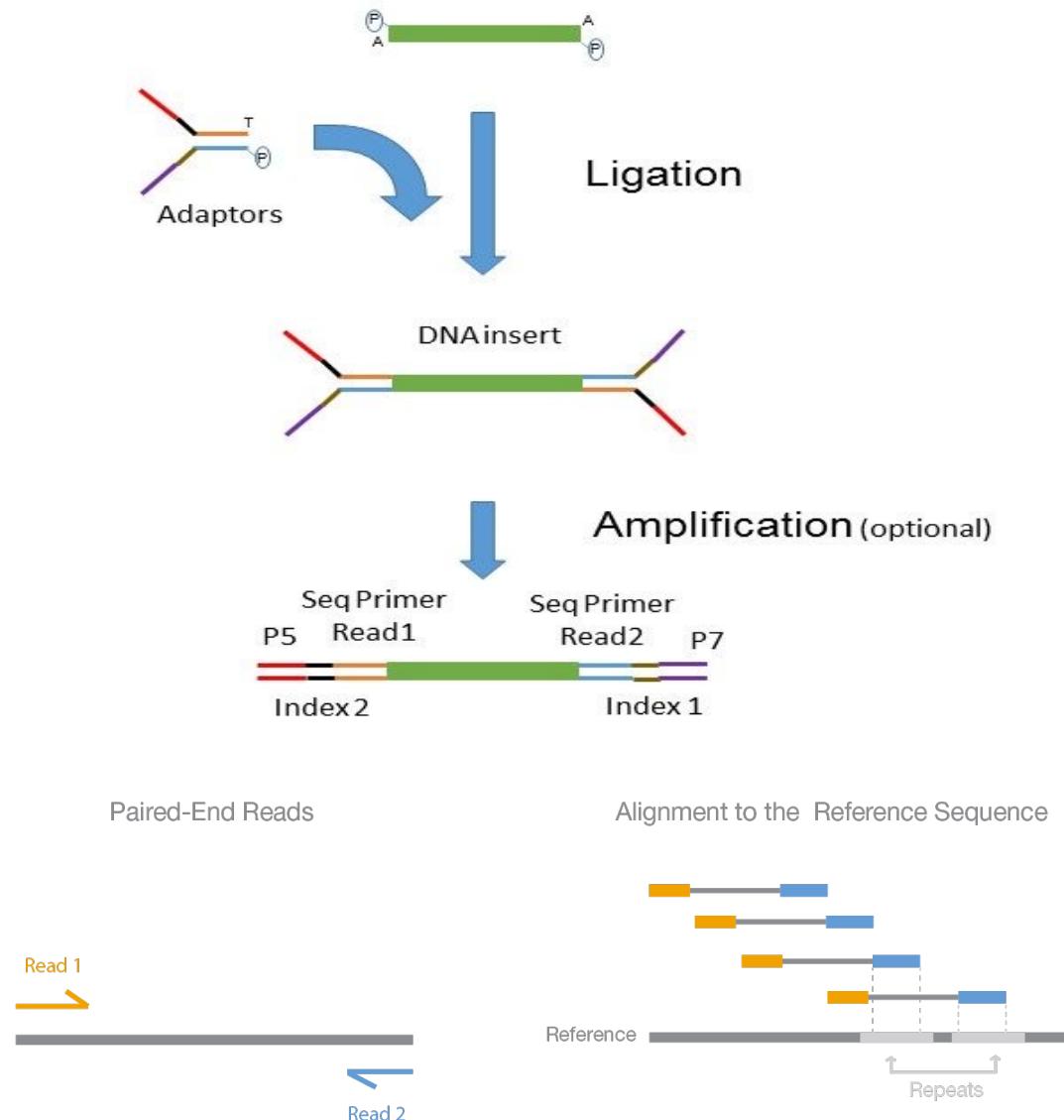


<https://www.illumina.com/2-channel-sbs.html>

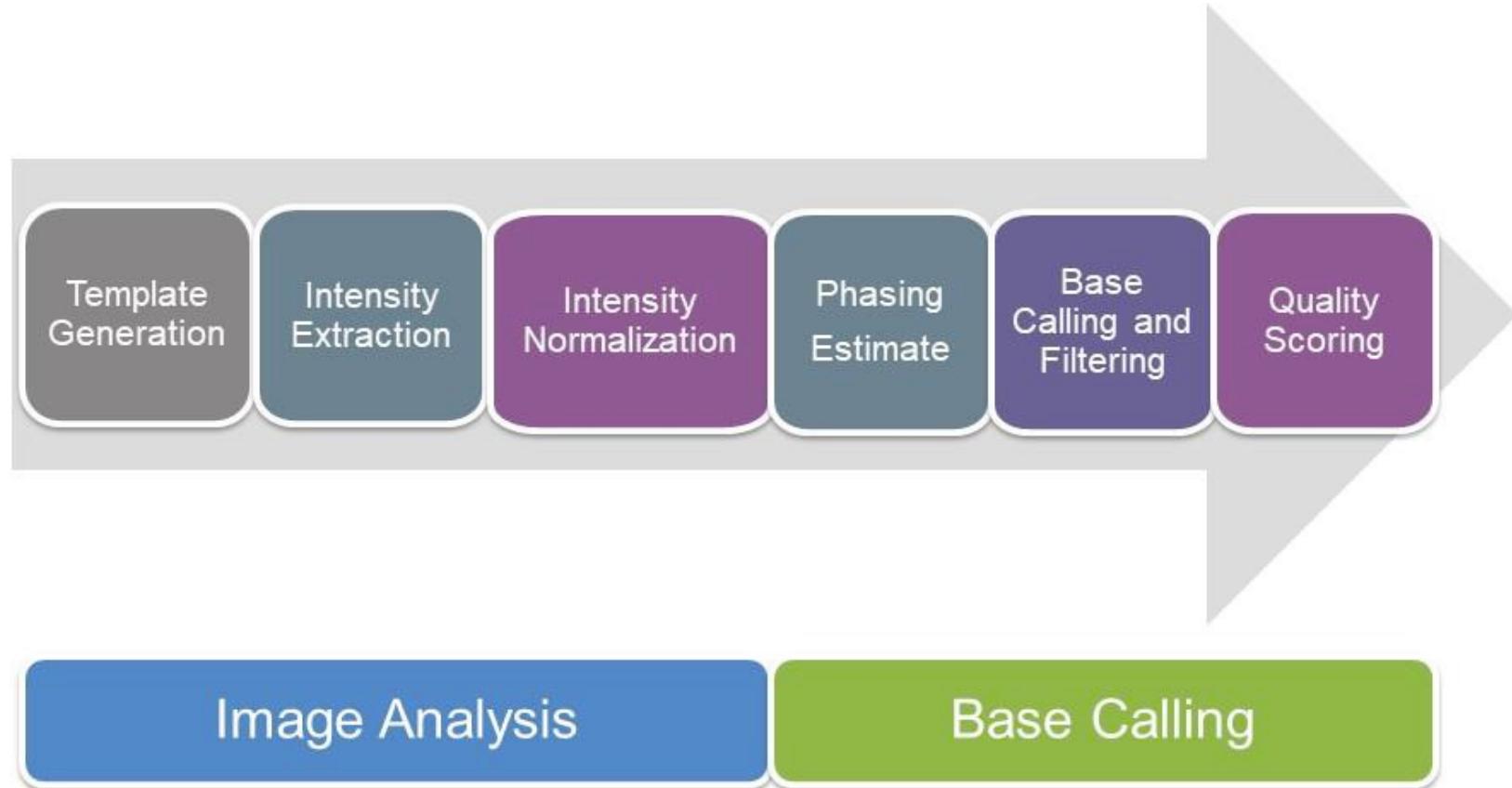
Illumina Paired-End Library Prep with Dual- Indexing

Paired-end sequencing (PE)

- Clusters are regenerated for the synthesis of the same strand in the opposite direction using a second sequencing primer complementary to the adaptor.
- PE reads are easier to align (particularly to repetitive regions) and allow the detection of insertions, deletions, and inversions.
- PE RNA sequencing helps to detect gene fusions and splice variants.



Illumina Primary Data Analysis Workflow



Illumina Sequence Quality Score

Quality typically declines at the end of the read

Error limits read length
(is reported in log scale)

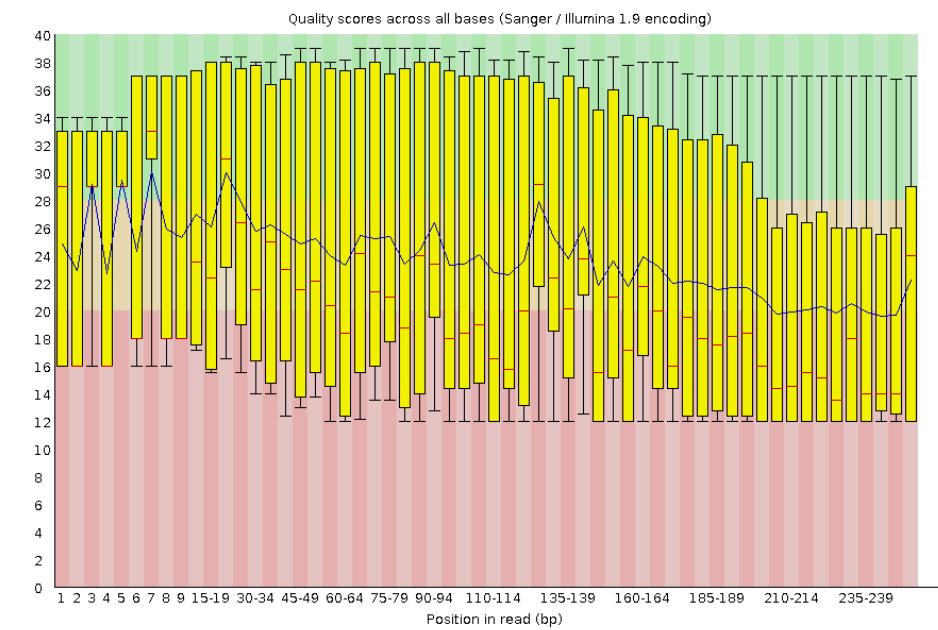
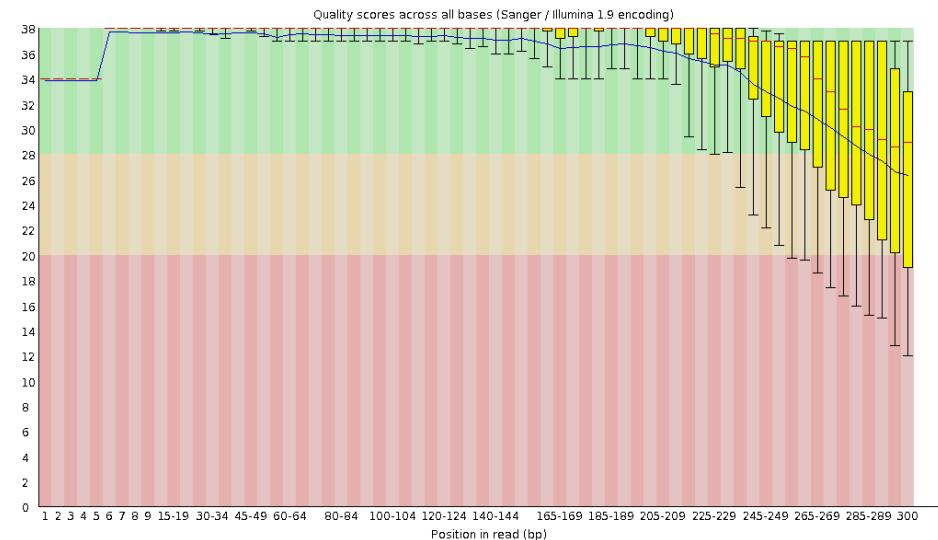
Q10: 1 error in 10

Q20: 1 error in 100

Q30: 1 error in 1.000

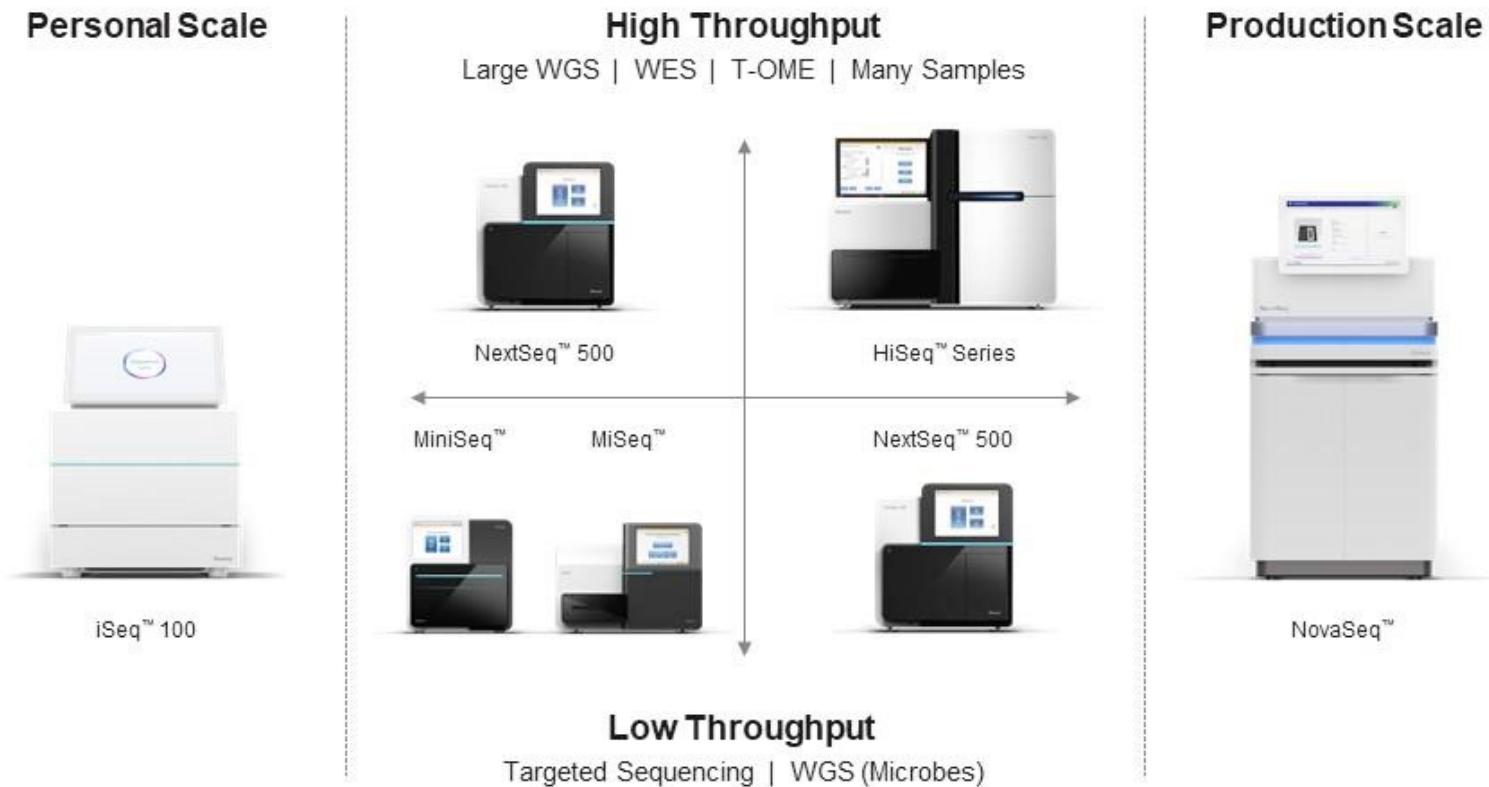
Q40: 1 error in 10.000

Q50: 1 error in 100.000



Illumina sequencers portfolio

Instruments for short-read sequencing, range from small, **low-throughput** benchtop units to large ultra-**high-throughput** instruments dedicated to population-level whole-genome sequencing (WGS).



Illumina sequencers portfolio

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			●	●	●

Illumina sequencers portfolio

Benchtop Sequencers		Production-Scale Sequencers		
		NextSeq Series +	HiSeq Series +	
				
Popular Applications & Methods	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				

Illumina sequencers portfolio



iSeq™ 100 System

	iSeq 100	\$19,990 USD	1 Dye Chemistry	9-17 hours run time		
Number of Flow Cells	1					
Lanes per flow cell	1					
Sequencing length	PE 150bp	PE 75bp	SE 75bp	SE 50bp	SE 36bp	
Output	1.2 Gb	600 Mb	300 Mb	200 Mb	144 Mb	
Reads output	4 millions	4 millions	4 millions	4 millions	4 millions	

*1 Dye Chemistry (nucleotides are labelled with a single dye, with the exception of the G nucleotide)



MiniSeq System

	MiniSeq	\$49,500 USD	2 colour chemistry	24 hour run time	
Number of Flow Cells	1				
Lanes per flow cell	1				
Sequencing length	PE 150bp	PE 75bp	SE 75bp	PE 150bp	
Output	6.6-7.5 Gb	3.3-3.75 Gb	1.65-1.875 Gb	2.1-2.4 Gb	
Kit type	High-Output	High-Output	High-Output	Mid-Output	
Single Reads output	22-25 millions			7-8 millions	
Paired-end Reads output	44-50 millions			14-16 millions	

Illumina sequencers



MiSeq Series

	MiSeq	\$99,000 USD	1-2 days run time	\$100-\$240/Gb	Error rate 0.1%				
Number of Flow Cells	1								
Lanes per flow cell	1								
Sequencing length	PE 300bp	PE 75bp	PE 250bp	PE 150bp	PE 25bp	SE 36bp	PE 150bp	PE 250bp	PE 150bp
Output	13.2-15 Gb	3.3-3.8 Gb	7.5-8.5 Gb	4.5-5.1 Gb	750-850 Mb	540-610 Mb	1.2 Gb	500Mb	300 Mb
Kit type	Reagent kit v3	Reagent kit v3	Reagent kit v2	Reagent kit v2	Reagent kit v2	Reagent kit v2	Reagent kit v2 Micro	Reagent kit v2 Nano	Reagent kit v2 Nano
Single Reads output	22-25 millions		7-8 millions				4 millions	1 million	
Paired-end Reads output	44-50 millions		14-16 millions				8 millions	2 millions	



NextSeq Series

	Nextseq 550	\$250,000 USD	2 colour chemistry	\$30-\$45/Gb	1.25 days run time
Number of Flow Cells	1				
Lanes per flow cell	1 (technically 4, but not physically isolated)				
Sequencing length	PE 150bp	PE 75bp	SE 75bp	PE 150bp	PE 75bp
Output	100–120 Gb	50–60 Gb	25–30 Gb	32.5–39 Gb	16.25–19.5 Gb
Kit type	High-Output	High-Output	High-Output	Mid-Output	Mid-Output
Single Reads output	400 millions			130 millions	
Paired-end Reads output	800 millions			260 millions	

Illumina sequencers



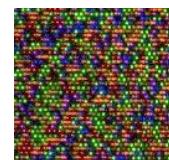
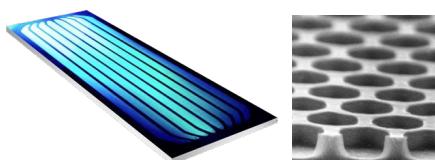
HiSeq Series

	HiSeq 2500	\$740,000 USD					
Number of Flow Cells	2						
Lanes per flow cell	8 (High Output run mode)						
Sequencing length	PE 125bp	PE 100bp	PE 50bp	SE 36bp	PE 100bp	PE 50bp	SE 36bp
Dual flow cell Output	900 Gb - 1 Tb	720–800 Gb	360–400 Gb	128–144 Gb	540–600 Gb	270–300 Gb	95–105 Gb
Single flow cell Output	450 - 500 Gb	360–400 Gb	180–200 Gb	64–72 Gb	270–300 Gb	135–150 Gb	47–52 Gb
Kit type	HiSeq SBS v4	HiSeq SBS v4	HiSeq SBS v4	HiSeq SBS v4	TruSeq SBS v3	TruSeq SBS v3	TruSeq SBS v3
Single Reads output (Single flow cell)	Up to 4 billions				Up to 2 billions		
Paired-end Reads output (Single flow cell)	Up to 8 billions				Up to 4 billions		

	HiSeq 2500	\$740,000 USD				
Number of Flow Cells	2					
Lanes per flow cell	2 (Rapid run mode)					
Sequencing length	PE 250bp	PE 150bp	PE 100bp	PE 50bp	SE 36bp	
Dual flow cell Output	250–300 Gb	150–180 Gb	100–120 Gb	50–60 Gb	18–22 Gb	
Single flow cell Output	125–150 Gb	75–90 Gb	50–60 Gb	25–30 Gb	9–11 Gb	
Kit type	HiSeq Rapid SBS v2	HiSeq Rapid SBS v2	HiSeq Rapid SBS v2	HiSeq Rapid SBS v2	HiSeq Rapid SBS v2	
Single Reads output (Single flow cell)	Up to 300 millions					
Paired-end Reads output (Single flow cell)	Up to 600 millions					

	HiSeq 3000	\$740,000 USD	\$24/Gb
Number of Flow Cells	1		
Lanes per flow cell	8		
Sequencing length	PE 150bp	PE 75bp	SE 50bp
Output	650–750 Gb	325–375 Gb	105–125 Gb
Single Reads output	Up to 2.5 billions		
Paired-end Reads output	Up to 5 billions		

	HiSeq 4000	\$900,000 USD	
Number of Flow Cells	2 (Dual or Single flow cell run)		
Lanes per flow cell	8		
Sequencing length	PE 150bp	PE 75bp	SE 50bp
Dual flow cell Output	1300–1500 Gb	650–750 Gb	210–250 Gb
Single flow cell Output	650–750 Gb	325–375 Gb	105–125 Gb
Single Reads output (Single flow cell)	Up to 2.5 billions		
Paired-end Reads output (Single flow cell)	Up to 5 billions		



Illumina sequencers ultra-high-throughput



HiSeq X Series

	HiSeq X	\$1M USD
Number of Flow Cells	2 (Dual or Single flow cell run)	
Lanes per flow cell	8	\$8/Gb
		\$1000 genome
Sequencing length	PE 150bp	3 days run time
Dual flow cell Output	1.6-1.8 Tb/run	
Single flow cell Output	8-9 Tb	
Single Reads output (Single flow cell)	5.3-6 billions	
Paired-end Reads output (Single flow cell)	10.6-12 billions	



NovaSeq Series

	NovaSeq 6000	\$985,000 USD		NovaSeq 5000	\$850,000 USD		
Number of Flow Cells	2 (Dual or Single flow cell run)						
Lanes per flow cell	2 (S1 flow cell)			2 (S2 flow cell)			4 (S4 flow cell)
Sequencing length	PE 150bp	PE 100bp	PE 50bp	PE 150bp	PE 100bp	PE 50bp	PE 150bp
Dual flow cell Output	800 - 1000 Tb	466 - 666 Gb	268 - 334 Gb	1700-2000 Gb	1100 - 1300 Gb	560 - 666 Gb	4800-6000 Gb
Single flow cell Output	400-500 Gb	266-333 Gb	134-167 Gb	850-1000 Gb	560-667 Gb	280-333 Gb	2400-3000 Gb
Single Reads output (Single flow cell)	1.3-1.6 billions			2.8 - 3.3 billions			8 - 10 billions
Paired-end Reads output (Single flow cell)	2.6 - 3.2 billions			5.6 - 6.4 billions			16 - 20 billions

Illumina sequencing technology

- ✓ **Cheap \$6-30/Gb** (\$1000/genome on Xten; \$800/genome on NovaSeq)
- ✓ **Accurate, mostly Q30**
- ✓ **Faster and higher throughput. Massively parallel. Millions of reads**
- ✓ **High concordance rate at SNPs** (Illumina vs microarrays)

- ✗ **Short fragment sequencing** does not give long-range information
(use long range sample prep approach HiC, Tell-Seq, Haplotagging –Meier et al., BioRxIV 2020)
- ✗ **Tendency towards substitution errors**, some systematic errors make difficult to spot rare variants (<1%) (see Duplex seq by Schmitt et al.)
- ✗ **Low complexity templates**
- ✗ **Index Hopping** -wrong barcode to wrong sample (See Sinha et al BioRXIV 2017)

Future of Illumina Sequencing

- ✓ Lower cost
- ✓ Faster & higher HT
- ✓ Single molecule sensing
- ✓ Sample to answer applications
- ✓ Clinical solutions and diagnostics
- ✓ Long reads?
- ✓ Cloud computing for sharing data

GeneReader (CRT) Platform

- A device for a clinical set up with an explicit focus on gene panels
- Intended to be an **all-in-one NGS platform**, from sample preparation to analysis
- Continuous loading of multiple flow cells
- The system is bundled with the **QIAcube** sample preparation system and the **Qiagen Clinical Insight** platform for variant analysis, making it a *de-facto* all-in-one NGS platform



<https://www.youtube.com/watch?v=-YTL1FVbMZs>

Qiagen GeneReader (CRT)

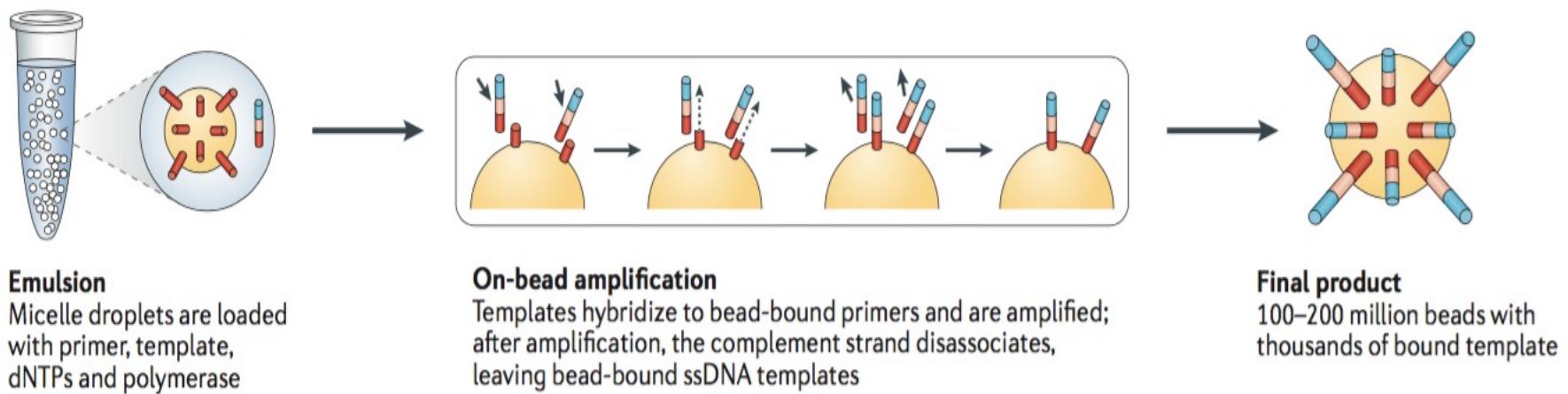
Cluster generation

- **Bead-based emulsion PCR**

Adapter-ligated DNA sequences are captured in an aqueous droplet (micelle)

Beads are loaded with primers

Each bead becomes a cluster of clonally amplified fragments

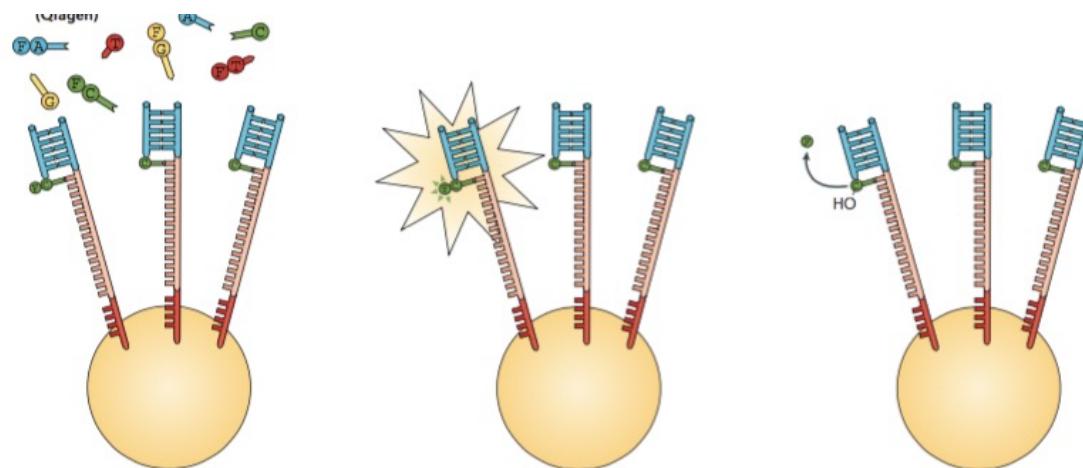


(Goodwin et al., Nat Rev Genet, 2016)

Qiagen GeneReader (CRT)

Sequencing

- Like Illumina **4-channel system**
- **Just enough fluorescently-labelled dNTPs** to achieve identification
- **Cleavable fluorophore**



(Goodwin et al., Nat Rev Genet, 2016)

Short read NGS

Sequencing by Synthesis (SBS)

DNA polymerase

Cyclic reversible termination (CRT)

- HiSeq & MiSeq (Illumina)
- Genereader (Qiagen)

Single nucleotide addition (SNA)

- Ion Torrent (Thermo Fisher)
- 454 (Roche)

Sequencing by Ligation (SBL)

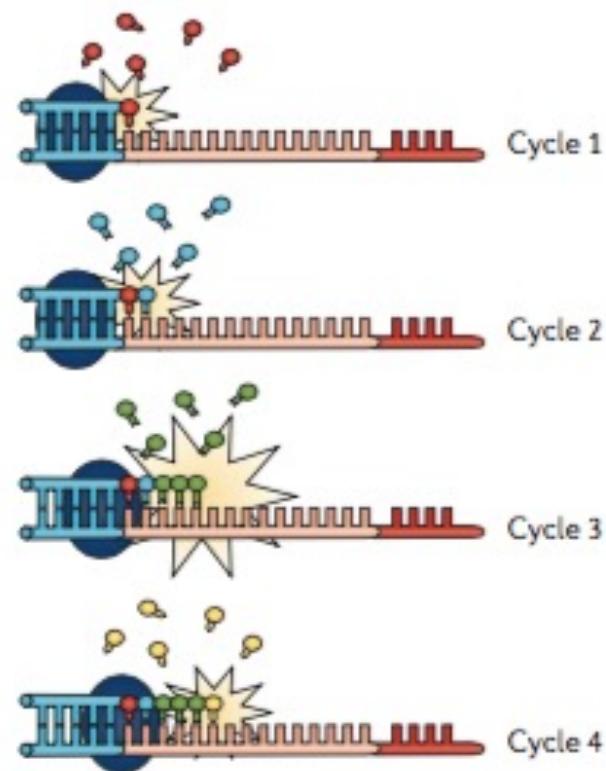
Anchor-probe ligation

SOLiD (Applied Biosystems)

Complete Genomics (BGI)

Principles of Single-nucleotide addition (SNA)

- A **single signal** marks the incorporation of dNTPs into the elongating strand
- dNTPs added **iteratively**
- dNTPs are **unblocked**, absence prevents elongation
- **Homopolymers** (e.g. “TTT”) are identified by a proportional increase in the signal.
- **Cluster generation: Bead-based emulsion PCR**



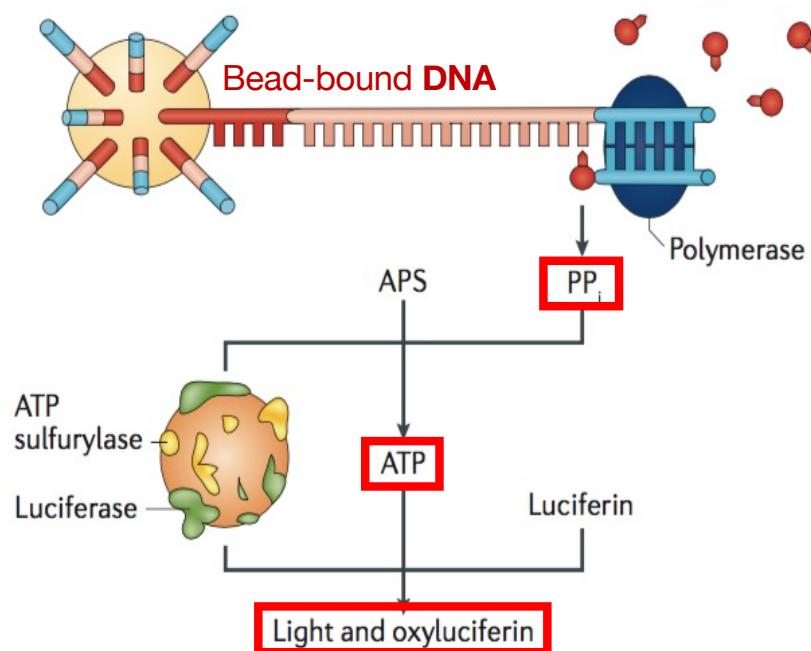
454 Pyrosequencer platform (SNA)

(started NGS 2005)
(discontinued in 2016)



454 Pyrosequencer platform (SNA)

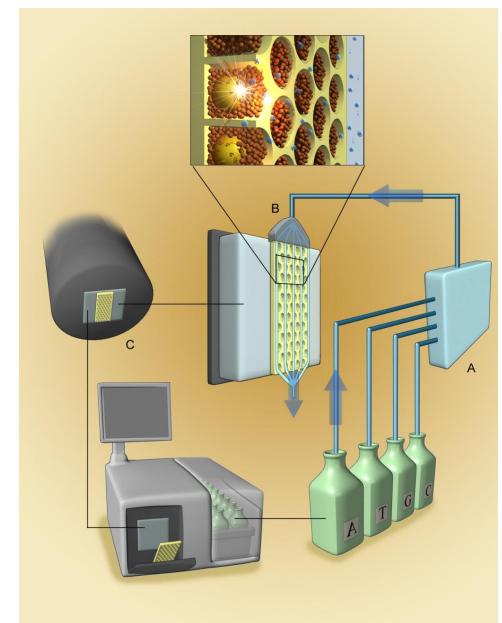
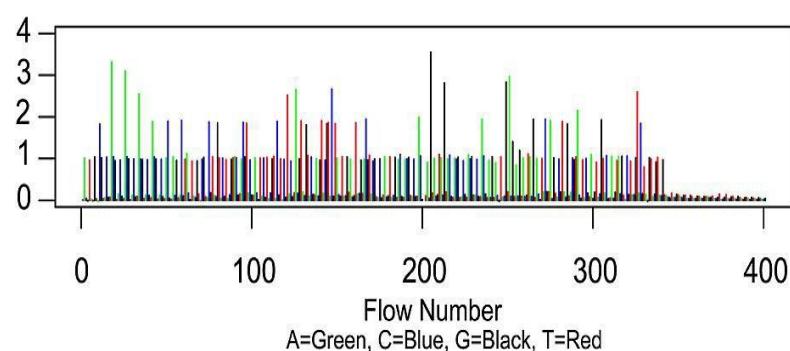
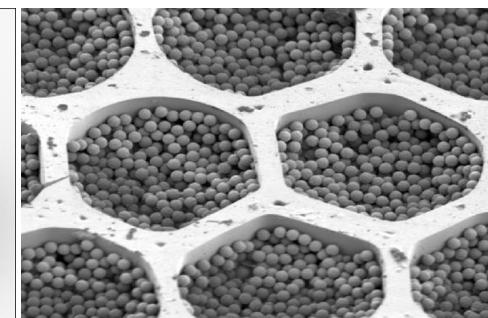
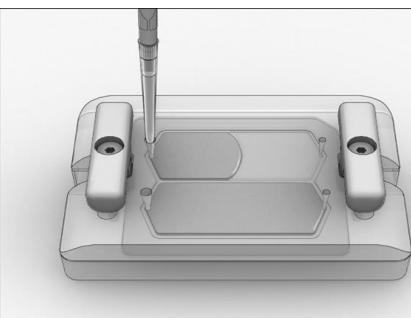
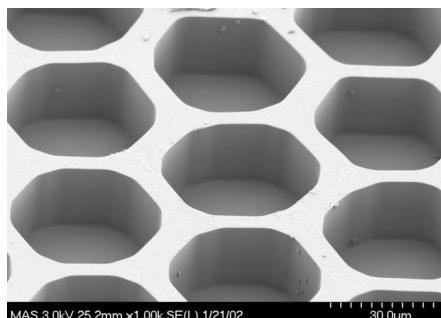
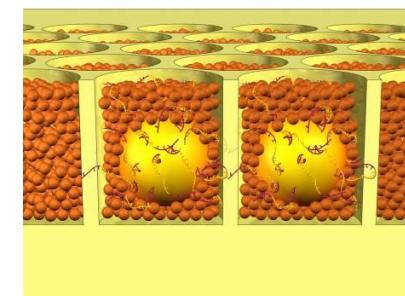
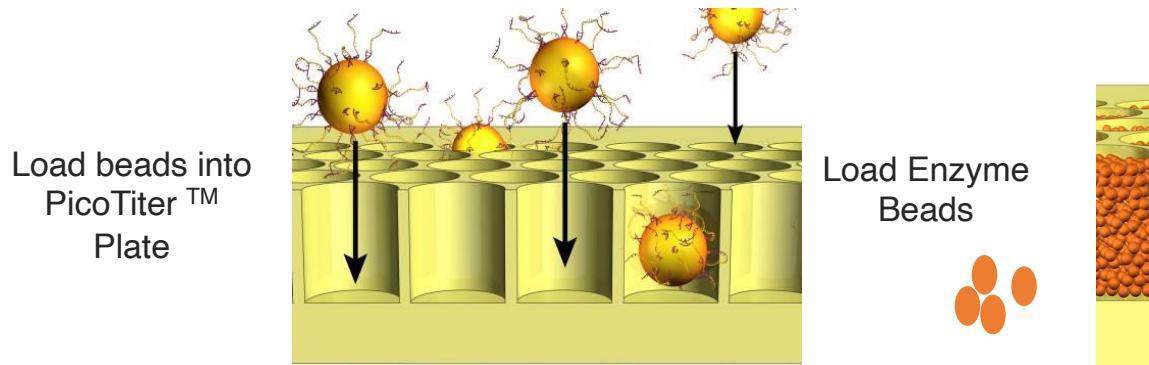
- Bead-bound **DNA** is placed in a micro-well along with beads coated with **ATP sulfurylase** and **luciferase**. In solution is the **DNA polymerase**, **adenosine phosphosulfate (APS)** and **luciferin**



(Goodwin et al., Nat Rev Genet, 2016)

454 Emulsion PCR

Depositing DNA Beads into the PicoTiter™Plate



454 Pyrosequencer platform: pros and cons

- ✓ **Superior read lengths** up ~700 bp compared to Illumina sequencers
- ✓ Good for amplicons and de-novo sequencing
- ✓ Resolves better repetitive DNA sequences
- ✓ Fast turnaround <24 hour run time
- ✗ Low throughput 0.7 Gb/run - high running costs \$7000/Gb
- ✗ High error rate near homopolymers
- ✗ Single end only

Platform	Read length (bp)	Throughput	Reads	Runtime	Error profile	Instrument cost (US\$)	Cost per Gb (US\$, approx.)
454 GS Junior	Up to 600; 400 average (SE, PE)*	35 Mb*	~0.1 M*	10 h*	1%, indel†	NA§	\$40,000‡
454 GS Junior+	Up to 1,000; 700 average (SE, PE)*	70 Mb*	~0.1 M*	18 h*	1%, indel†	\$108,000‡	\$19,500‡
454 GS FLX Titanium XLR70	Up to 600; 450 mode (SE, PE)*	450 Mb*	~1 M*	10 h*	1%, indel†	NA§	\$15,500‡
454 GS FLX Titanium XL+	Up to 1,000; 700 mode (SE, PE)*	700 Mb*	~1 M*	23 h*	1%, indel†	\$450,000‡	\$9,500‡

(Goodwin *et al.*, *Nat Rev Genet*, 2016)

Ion Torrent PGM platforms (SNA)

ThermoFisher
SCIENTIFIC



Ion Personal Genome Machine (PGM)



Ion Proton

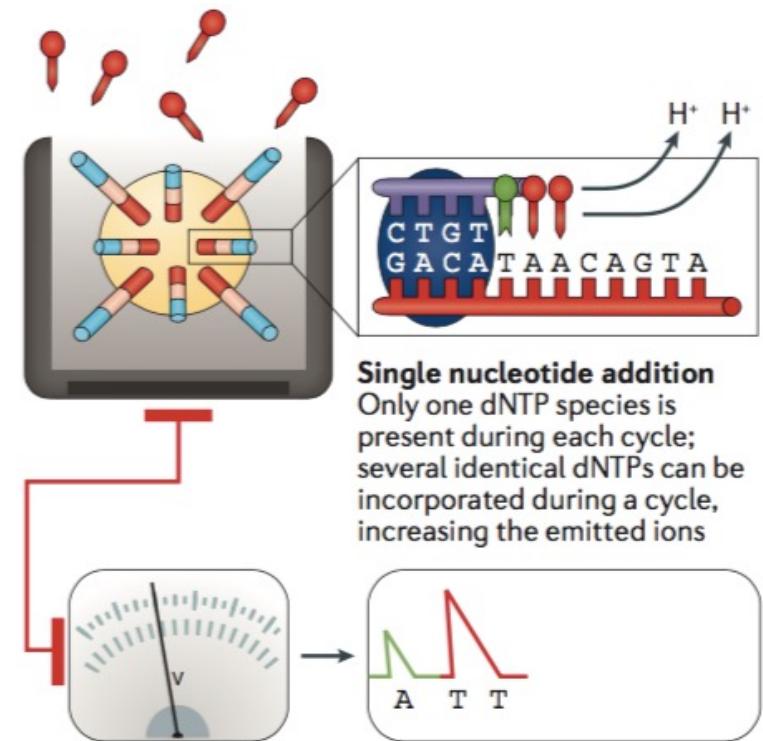


Ion S5 series

<https://www.youtube.com/watch?v=zBPKj0mMcDg>

Ion Torrent PGM platform (SNA)

- dNTP incorporation releases H⁺ ions resulting in a pH change
- pH change is detected by a semiconductor (CMOS) and an ion sensitive field-effect transistor (ISFET)
- The pH change detected is proportional to the number of nucleotides detected, allowing for **limited accuracy in measuring homopolymers > 6–8 bp**
- Library prep like original 454
- Amplification on beads by emPCR
- CMOS chip detection
- Cyclic addition sequencing
- Not single-molecule

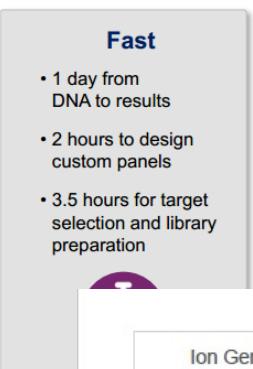
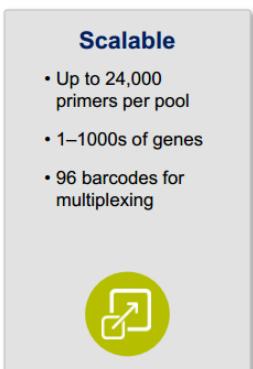
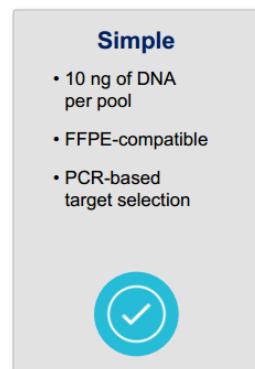


(Goodwin et al., Nat Rev Genet, 2016)

Ion GeneStudio S5 Series | Flexible Portfolio

Ion AmpliSeq™ technology: As Simple As PCR
Your Targets, Your Genome, Your Panel

The most comprehensive gene coverage
with the lowest amount of DNA or RNA Input



Ion GeneStudio™ S5



Fast.

Ion GeneStudio™ S5 Plus



New

Ion GeneStudio™ S5 Prime



Powerful



Ion 510™
Chip
2–3 M reads
Up to 400 bp



Ion 520™
Chip
3–6 M reads
Up to 600 bp



Ion 530™ Chip
15–20 M reads
Up to 600 bp



Ion 540™ Chip
60–80 M reads
Up to 200 bp

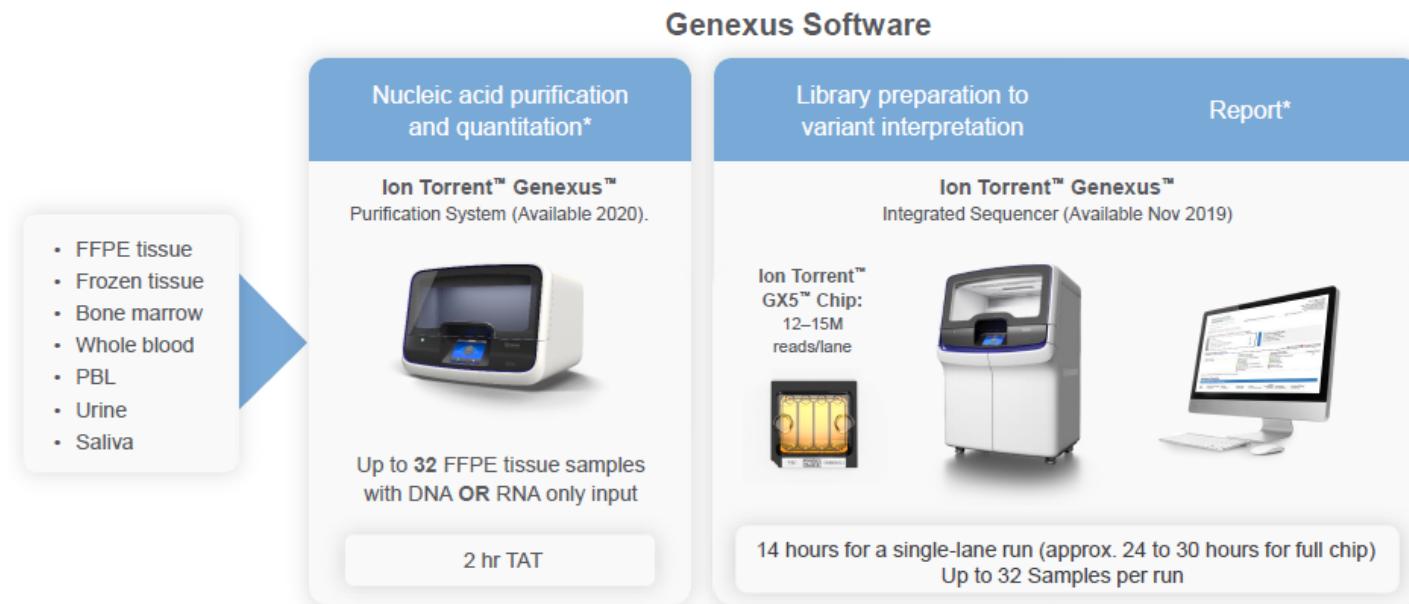


New
Ion 550™ Chip
100–130 M reads
Up to 200 bp

For Research Use Only. Not for use in diagnostic procedures. * Throughputs based on 200bp sequencing

New Ion Torrent platform

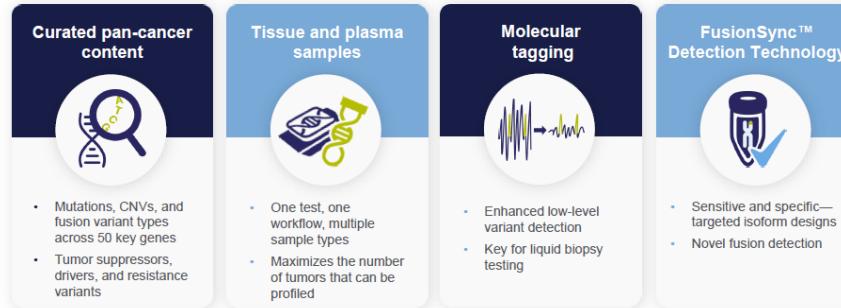
Genexus System—Tomorrow's Specimen-to-Report NGS Workflow



- * Specimen-to-report workflow available after Ion Torrent Genexus Purification System launches in 2020.
The content provided herein may relate to products that have not been officially released and is subject to change without notice.

Oncomine Precision Assay on Ion Torrent Genexus System

Maximizes your ability to detect relevant variants



The content provided herein may relate to products that have not been officially released and is subject to change without notice.

Ion Torrent platform: pros and cons

- ✓ Capital cost is cheaper than Illumina
 - ✓ Simple platform to operate, paired to Ion Chef library preparation. Good for targeted sequencing
Used a lot in clinical settings for disease panel sequencing.
 - ✓ Faster than other platforms (short runtimes)
 - ✓ Offers different types of chips (~50 Mb to 15 Gb and longer reads 200-400bp)
-
- X Not good for WGS.
 - X Ion Proton and S5 devices can not support PE sequencing limiting their utility for elucidating long-range genomic or transcriptomic structure (the exception is the Ion PGM Dx sequencer)
 - X Prone to errors near homopolymers - Error rate 1-2%

Platform	Read length (bp)	Throughput	Reads	Runtime	Error profile	Instrument cost (US\$)	Cost per Gb (US\$, approx.)
Ion PGM 314	200 (SE)	30–50	400,000–550,000*	23 h	1%, indel [‡]	\$49 [‡]	\$25–3,500 [‡]
	400 (SE)	60–100 Mb*		3.7 h*			
Ion PGM 316	200 (SE)	300–500 Mb	2–3 M*	3 h	1%, indel [‡]	\$49 [‡]	\$700–1,000 [‡]
	400 (SE)*	600 Mb–1 Gb*		4.9 h*			
Ion PGM 318	200 (SE)	600 Mb–1 Gb	4–5.5 M*	4 h	1%, indel [‡]	\$49 [‡]	\$450–800 [‡]
	400 (SE)*	1–2 Gb*		7.3 h*			
Ion Proton	Up to 200 (SE)	Up to 10 Gb*	60–80 M*	2–4 h*	1%, indel [‡]	\$224 [‡]	\$80 [‡]
Ion S5 520	200 (SE)	600 Mb–1 Gb	3–5 M*	2.5 h	1%, indel [‡]	\$65 (REF. 158)	\$2,400*
	400 (SE)*	1.2–2 Gb*		4 h*			
Ion S5 530	200 (SE)	3–4 Gb	15–20 M*	2.5 h	1%, indel [‡]	\$65 (REF. 158)	\$950*
	400 (SE)*	6–8 Gb*		4 h*			
Ion S5 540	200 (SE)*	10–15 Gb*	60–80 M*	2.5 h*	1%, indel [‡]	\$65 (REF. 158)	\$300*

(Goodwin et al., Nat Rev Genet, 2016)

Short read NGS

Sequencing by Synthesis (SBS)

DNA polymerase

Cyclic reversible termination (CRT)

- HiSeq & MiSeq (Illumina)
- Genereader (Qiagen)

Single nucleotide addition (SNA)

- Ion Torrent (Thermo Fisher)
- 454 (Roche)

Sequencing by Ligation (SBL)

Anchor-probe ligation

SOLiD (Applied Biosystems)

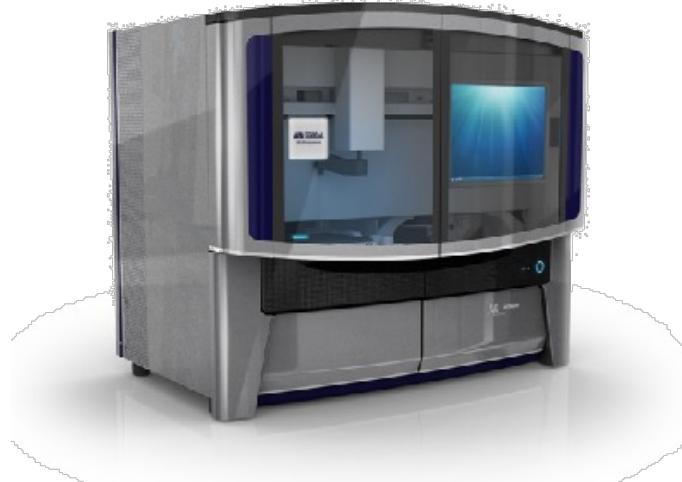
Complete Genomics (BGI)

SOLiD system (SBL)



SOLiD 5500 xl

Cluster generation: Bead-based emulsion PCR

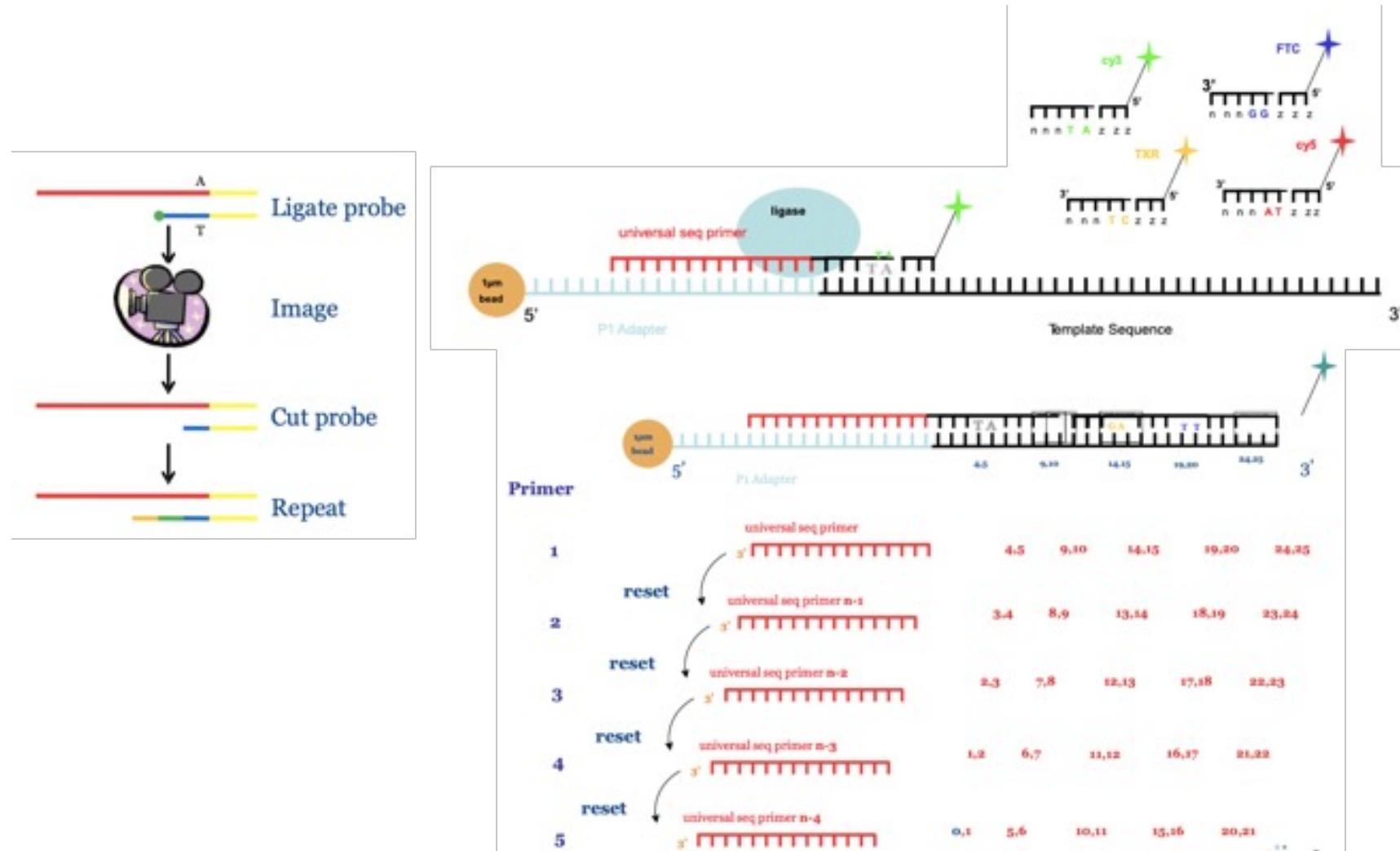


SOLiD 5500 wildfire

Solid-phase template walking

SOLiD system

Sequencing by Ligation (SBL)



<https://youtu.be/Otw8tUmuGtc>

ABI SOLiD system: pros and cons

- ✓ **High accuracy**, each base is probed multiple times – low error rate <0.1%
- ✓ **High throughput** >100 Gb/run + **low cost** for genomes \$30/Gb
- ✗ **Low sensitivity and specificity**, resulting in the loss of some true variants
- ✗ **Under-representation of AT-rich regions**, some substitution errors
- ✗ **Short read length up to 75 bp**, limited use for genome assembly and structural variant detection
- ✗ **Long sequencing times** 6-10 days/run

Platform	Read length (bp)	Throughput	Reads	Runtime	Error profile	Instrument cost (US\$)	Cost per Gb (US\$, approx.)
Sequencing by ligation							
SOLiD 5500 Wildfire	50 (SE)	80 Gb	~700 M*	6 d*	★ ≤0.1%, AT bias [‡]	NA [§]	\$130 [‡]
	75 (SE)	120 Gb					
	50 (SE)*	160 Gb*					
SOLiD 5500xl	50 (SE)	160 Gb	~1.4 B*	10 d*	★ ≤0.1%, AT bias [‡]	\$251,000 [‡]	\$70 [‡]
	75 (SE)	240 Gb					
	50 (SE)*	320 Gb*					

(Goodwin et al., Nat Rev Genet, 2016)

Short read NGS

Sequencing by Synthesis (SBS)

DNA polymerase

Cyclic reversible termination (CRT)

- HiSeq & MiSeq (Illumina)
- Genereader (Qiagen)

Single nucleotide addition (SNA)

- Ion Torrent (Thermo Fisher)
- 454 (Roche)

Sequencing by Ligation (SBL)

Anchor-probe ligation

SOLiD (Applied Biosystems)

Complete Genomics (BGI)

Complete Genomics (SBL)



BGISEQ-500

- » 210 Gb / slide (18 slides)
- » 35-base paired reads
- » 1 week run time
- » \$50 / Gb (finished)
- » Raw error rate 0.5%
- » Consensus accuracy 10⁻⁵



BGISEQ-50



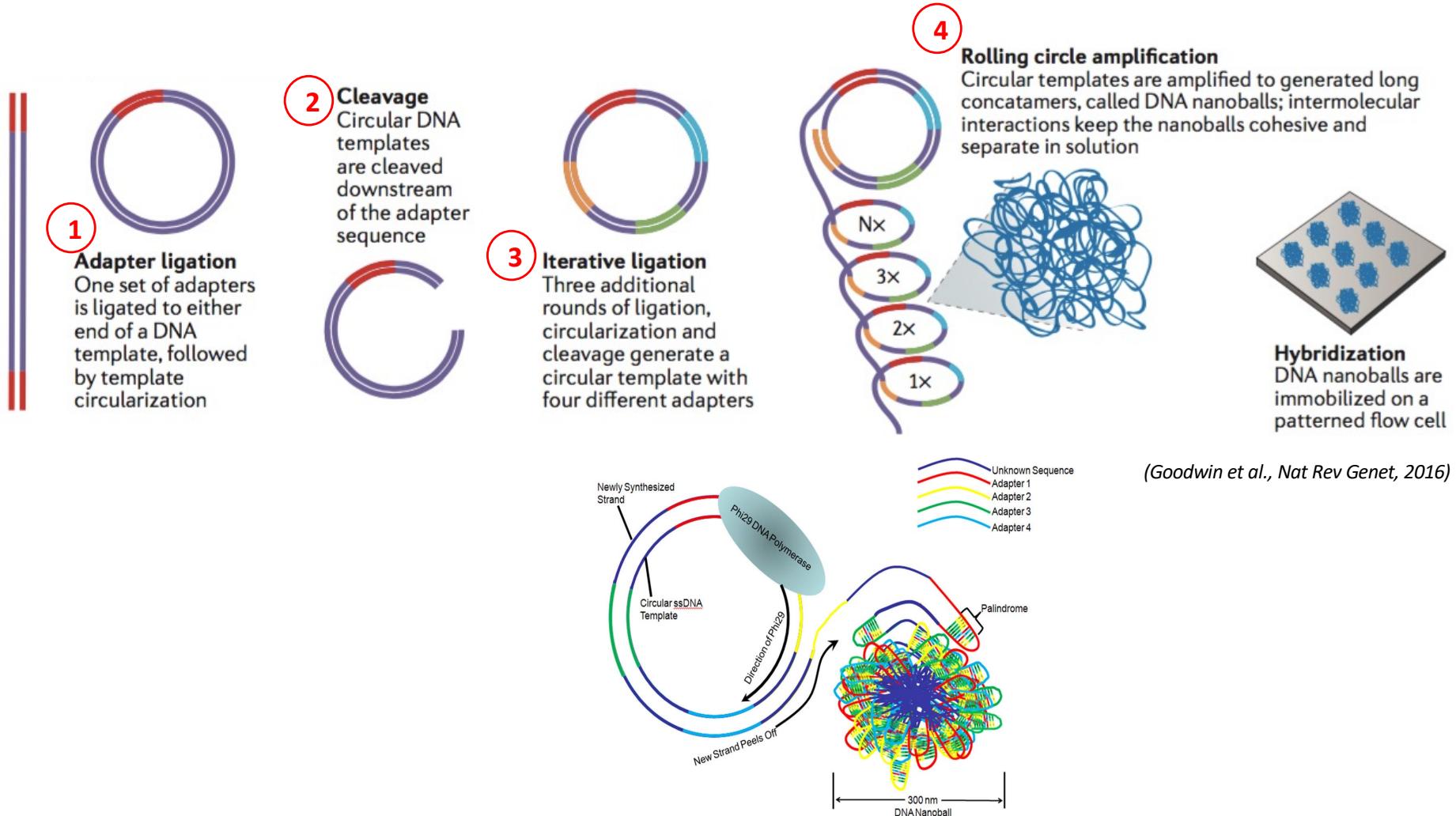
Revology

- » 10,000 genomes/year
- » \$12M
- » 10-120 samples at a time
- » 96% genome coverage
- » 300 bp insert. 2 x 28bp reads
- » Instrument cheaper than Illumina

Complete Genomics (SBL)

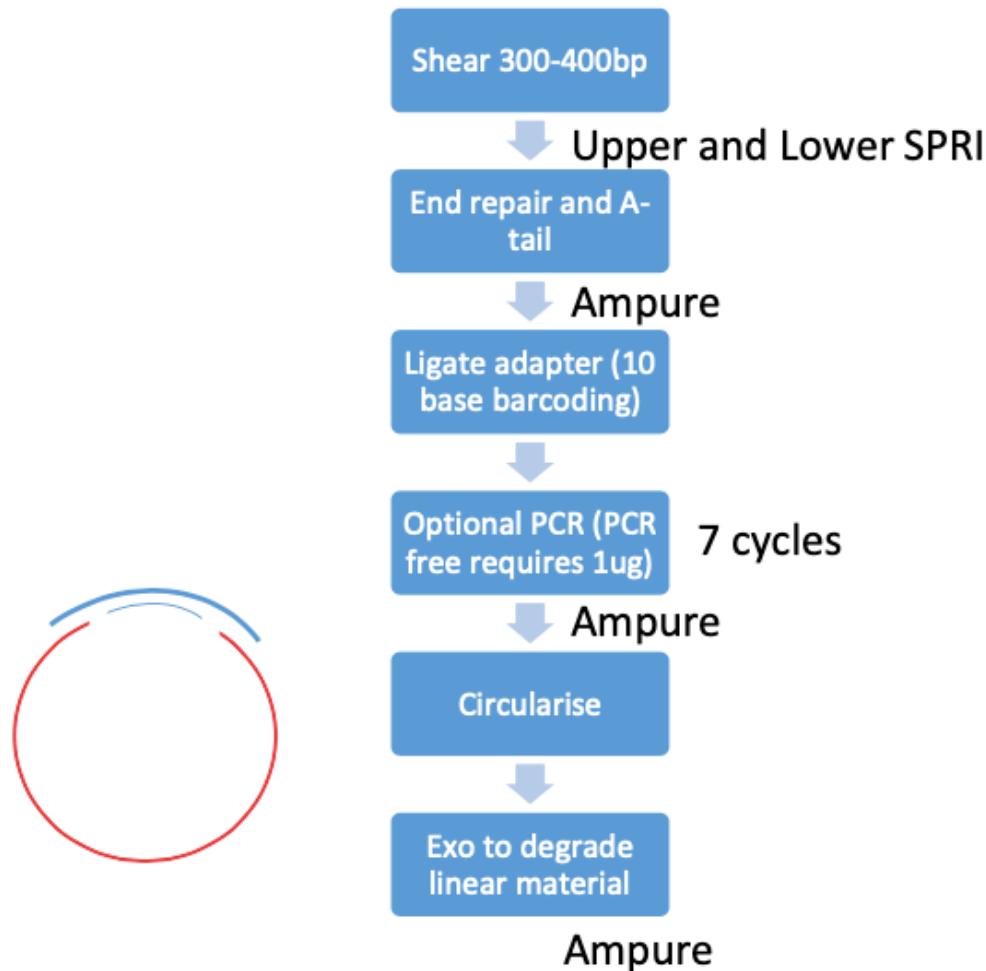
Cluster generation:

- In-solution DNA nano-balls ordered array



Complete Genomics

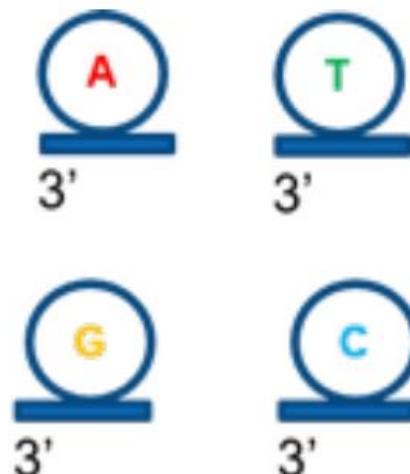
BGI (MGI) Library Prep



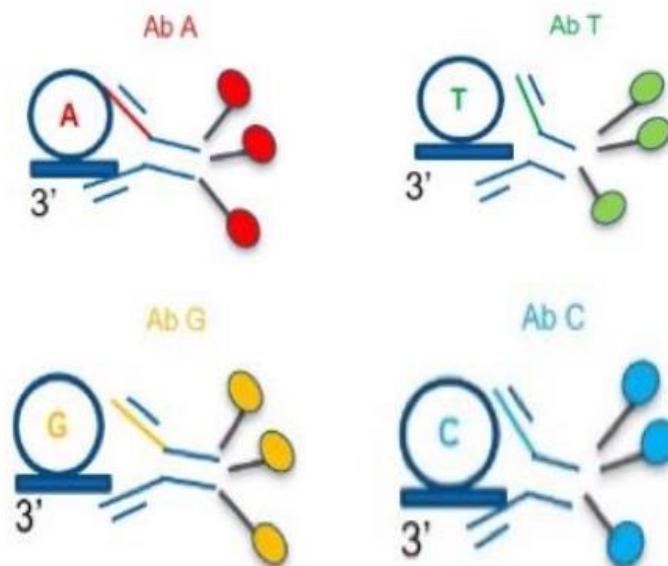
Complete Genomics acquired by BGI (MGI)

MGI CoolMPS Chemistry
Novel Antibody-based Sequencing

Natural Nucleobase with 3' Extension Blocks



Base Specific Block Dependent Antibodies



Complete Genomics

Cleaner, Brighter, Longer Sequencing



DNBSEQ-T7 Update



Flexible:

- 1 to 4 independent flow cells per run
- 15 - 60 WGS/day

Cost Effective:

- \$500/genome with just 10-15 genomes per flow cell

✓ 20 installed in China and Asia Pacific since September 2019

✓ Assembled **COVID-19 sequence** using Deep Sequencing on 2 T7s in Wuhan

✓ Fastrack **approved** for clinical use in China NMPA

Complete Genomics

Cleaner, Brighter, Longer Sequencing
DNBSEQ-Tx Sequencer – Making \$100 Genome Real



An integrated system
with 4 Imagers

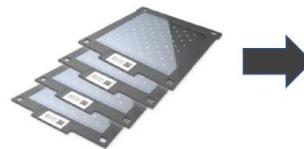


Tx – Extreme Throughput

T7 Output: 6 Tb/Day

T7 Flow Cell

7,000,000,000 (7B)
DNB binding spots

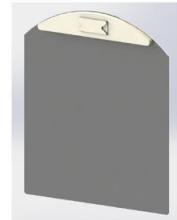


Up to 15 Genomes/Flow Cell

Tx Output: 20 Tb/Day

Tx Slide (4x Bigger)

68,526,000,000 (~70B)
DNB binding spots



Up to 150 Genomes/Slide

171mm*143mm

Sequencer	DNB/FOV Million	Imager/Set	Flow cell or Slide/Set	Output/Run (Tb)	Run time (day)	Throughput (Tb/day)*
DNBSEQ-T7	2	1	4	6	1	6
DNBSEQ-Tx	8.5	4	8	70	3.5	20

*Calculated using 70%ESR, 96% barcode split rate for PE100

Complete Genomics: pros and cons

- ✓ cPAL high accuracy levels (~99.99%) – raw error rate 0.2%
- ✓ High throughput 210 Gb/slide + low cost for genomes \$50/Gb
- ✓ No index hopping. No fundamental Bias. No fragment size dependent representation bias.
- ✗ Only available as a service for human WGS at BGI (40x human genome at \$600 per genome)
- ✗ cPAS platform is limited to mainland China – Instrument 33% less than cost of a HiSeq

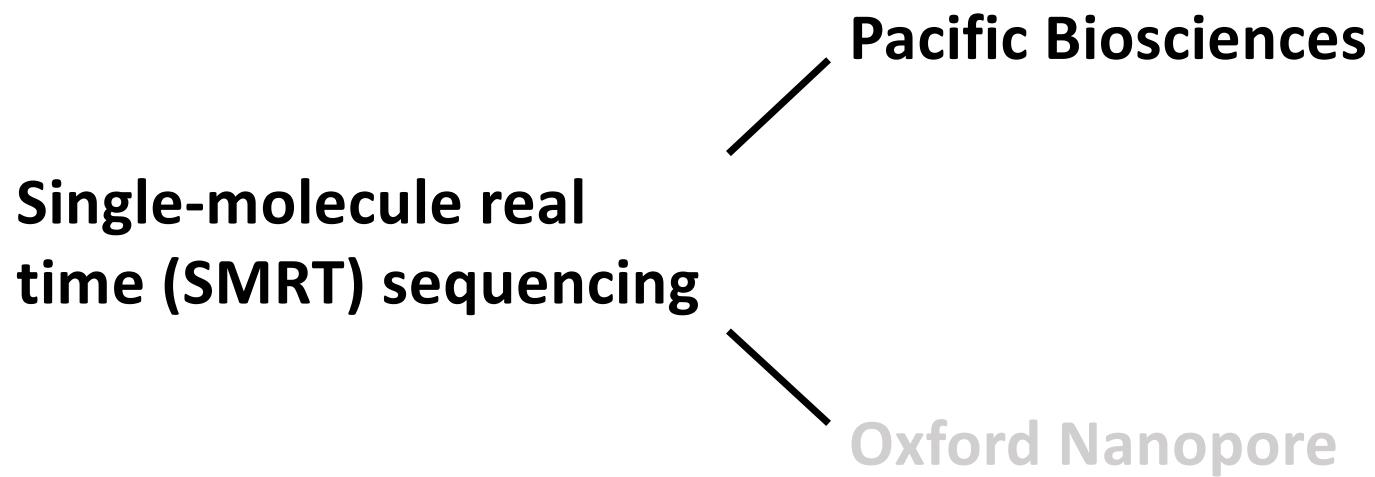
Platform	Read length (bp)	Throughput	Reads	Runtime	Error profile	Instrument cost (US\$)	Cost per Gb (US\$, approx.)
BGISEQ-500 FCS ¹⁵⁵	50–100 (SE/PE)*	8–40 Gb*	NA [¶]	24 h*	≤0.1%, AT bias [‡]	\$250 (REF. 155)	NA [¶]
BGISEQ-500 FCL ¹⁵⁵	50–100 (SE/PE)*	40–200 Gb*	NA [¶]	24 h*	≤0.1%, AT bias [‡]	\$250,000 (REF. 155)	NA [¶]

(Goodwin *et al.*, *Nat Rev Genet*, 2016)

Outline

- ❖ History of DNA sequencing
- ❖ Next generation sequencing technologies (NGS)
 - Short-read NGS
 - **Long-read NGS**

Long-read NGS



Long-read NGS

Pacific Biosciences SMRT sequencing



PacBio RS II



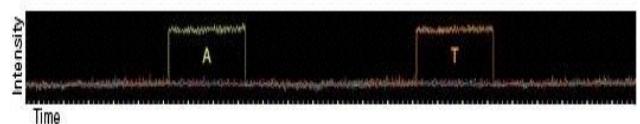
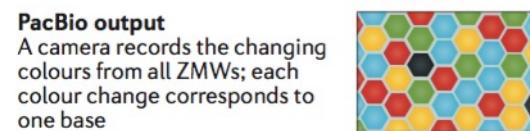
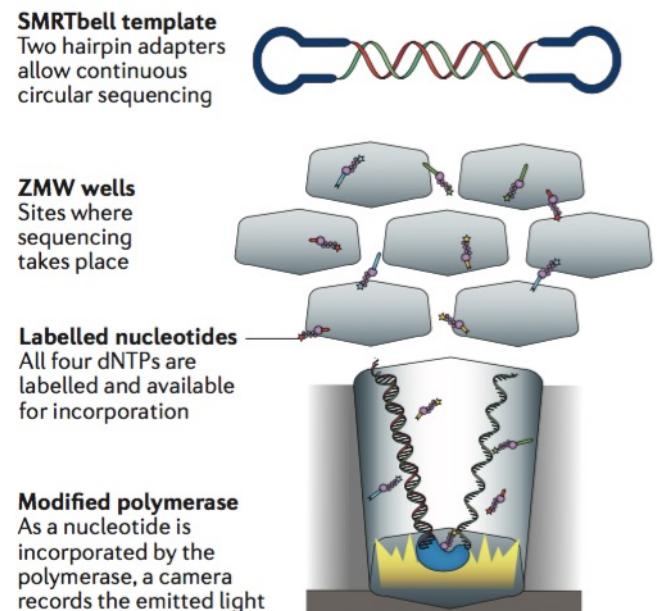
Sequel ii system

- » 2019 release
- » 8 million ZMWs/SMRT
- » 100Gb/SMRT cell
- » ~\$9/Gb
- » \$475K instrument

<https://www.youtube.com/watch?v=v8p4ph2MAvI&feature=youtu.be>

Pacific Biosciences SMRT sequencing

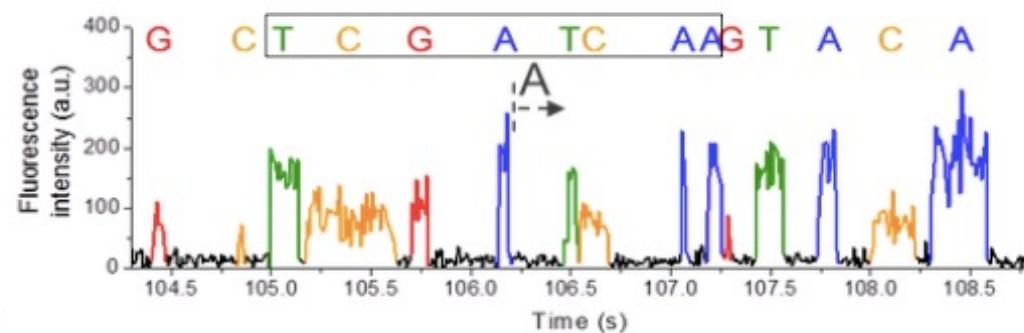
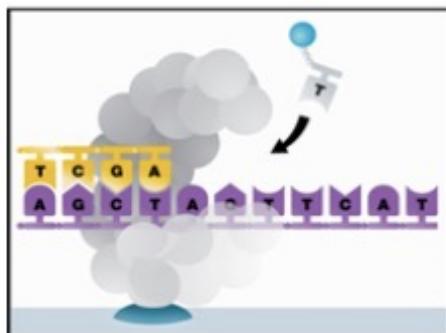
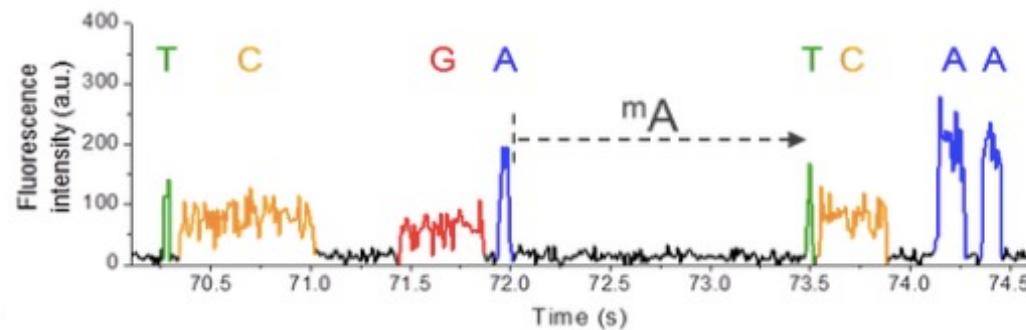
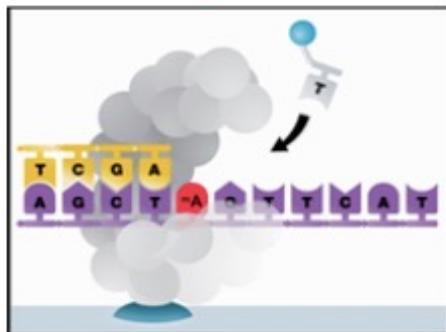
- **SMRTbell:** Hairpin adapter ligation results in a circular DNA molecule with constant ssDNA regions at each end
- **Zero-mode waveguides (ZMW):** SMRTbells - 1 DNA molecule and 1 polymerase are added to each picolitre wells with transparent bottoms
 - ZMWs allow light to illuminate the bottom of a well. 4 colours flash in realtime as polymerase acts
 - Methylated bases have distinct pattern
- **Fluorescently-labelled dNTP incorporation** recorded by a camera system that register the colour and duration of the light emitted.
- Long-insert genome sequence and Circular Consensus Sequencing



(Goodwin et al., Nat Rev Genet, 2016)

Pacific Biosciences - Base Modifications

Example: N⁶-methyladenine



- SMRT Sequencing uses kinetic information from each nucleotide addition to call bases
- This same information can be used to distinguish modified and native bases by comparing results of SMRT Sequencing to an *in silico* kinetic reference for incorporation dynamics without modifications.

PacBio sequencing: pros and cons

- ✓ Long-read sequencing (>21 Kb read length) with few systematic errors
- ✓ 200-500 Mb yield, 10 Gb per SMRT Cell
- ✓ Ideal for *de-novo* genome assembly, insertions, long-range genomic structures and full-length transcript, haplotyping
- ✓ Can detect some base modifications – DNA modifications studies
- ✓ ~\$9/Gb

- ✗ High error rate (~15% indels), addressed by increasing coverage
- ✗ Downside of increasing cost (\$400/Gb RSII and \$100/Gb Sequel) and time
- ✗ Equipment cost \$750K

Platform	Read length (bp)	Throughput	Reads	Runtime	Error profile	Instrument cost (US\$)	Cost per Gb (US\$, approx.)
Pacific BioSciences RS II	~20Kb	500 Mb–1 Gb*	~55,000*	4 h*	13% single pass, ≤1% circular consensus read, indel [#]	\$695 [‡]	\$1,000 [‡]
Pacific Biosciences Sequel	8–12 Kb ⁶⁹	3.5–7 Gb*	~350,000*	0.5–6 h*	NA	\$350 (REF. 69)	NA

(Goodwin *et al.*, Nat Rev Genet, 2016)

Long-read NGS

**Single-molecule real
time (SMRT) sequencing**

Pacific Biosciences

Oxford Nanopore

Long-read NGS



MinION

GridION X5

PromethION

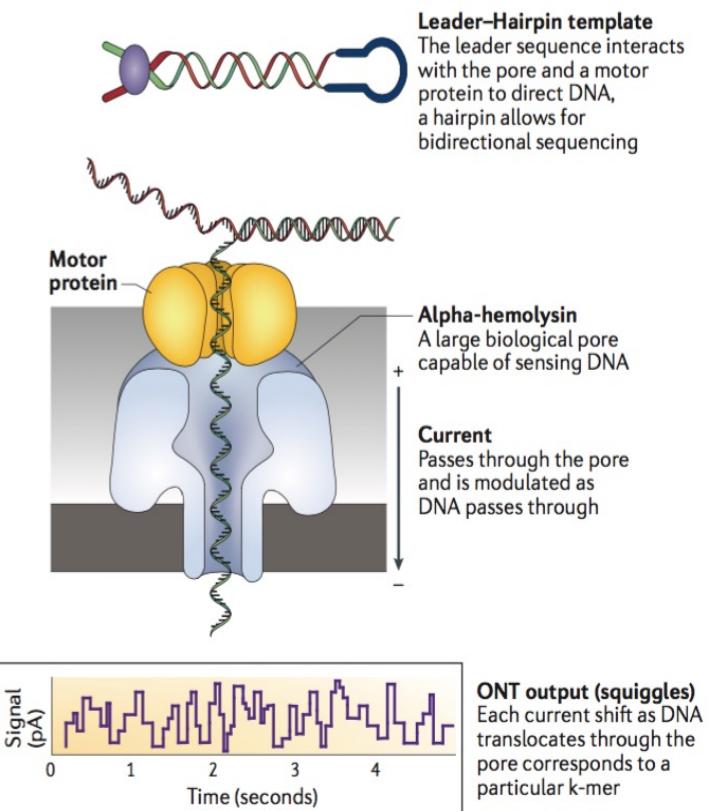


SmidgION

<https://www.youtube.com/watch?v=GUb1TZvMWsw&feature=youtu.be>

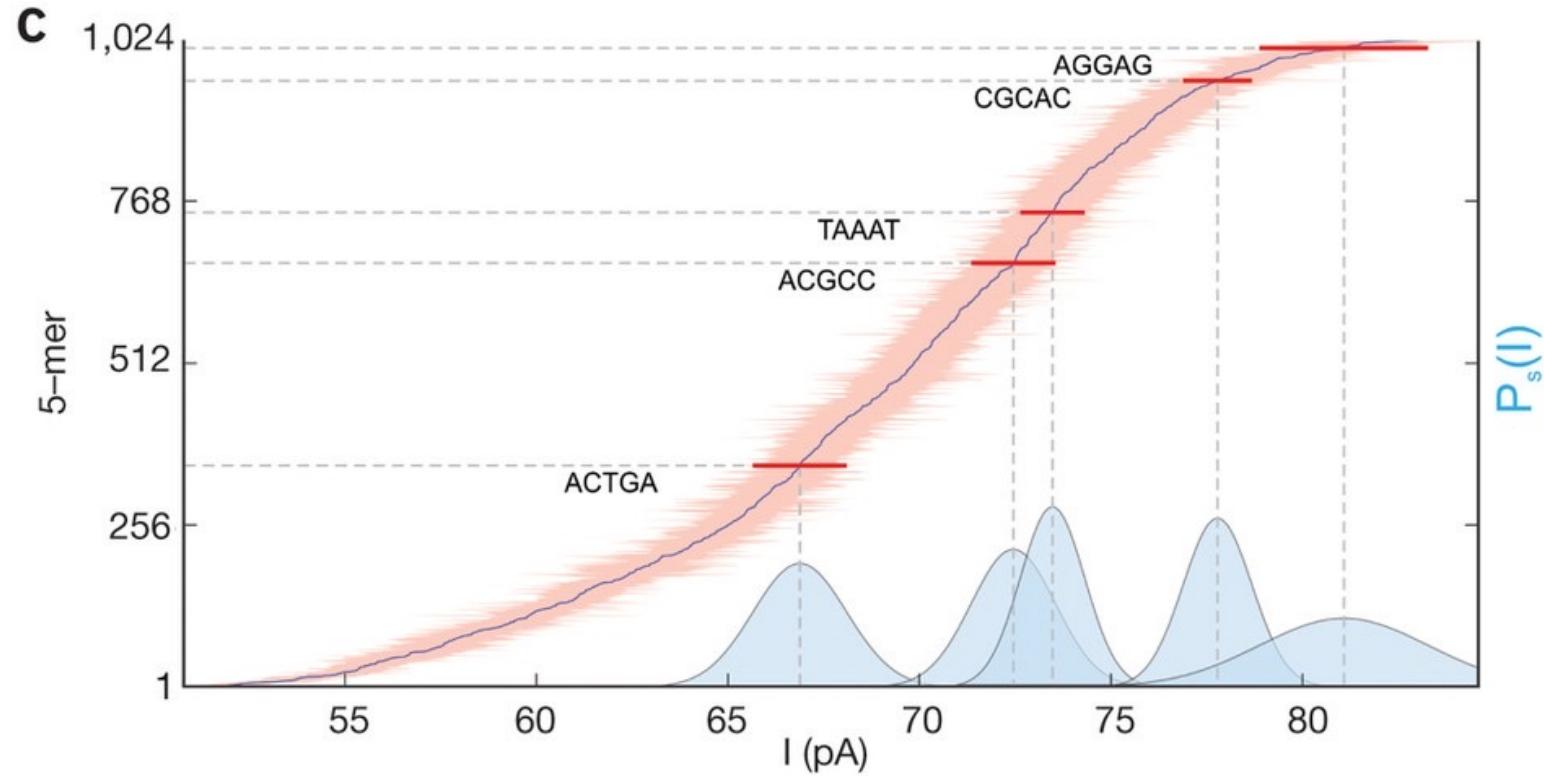
Oxford Nanopore sequencing

- **Protein pore** (**alfa-hemolysin**), **a membrane** and a **motor protein** that translocate the DNA molecule at a defined rate (~500 bases/sec)
- Ionic gradient > current passes through the pore
- **K-mer:** a substring of 3-6 bases within the hairpin ssDNA template
- **Squiggle space:** the tracing of the shift in current mediated by a k-mer passing through the pore
- Each **k-mer** (>1,000) generates a unique **squiggle space**, which allows base calling in real-time (e.g. R9.4.1 pore: 5-mer x 4 bases = 1024 potential current levels)



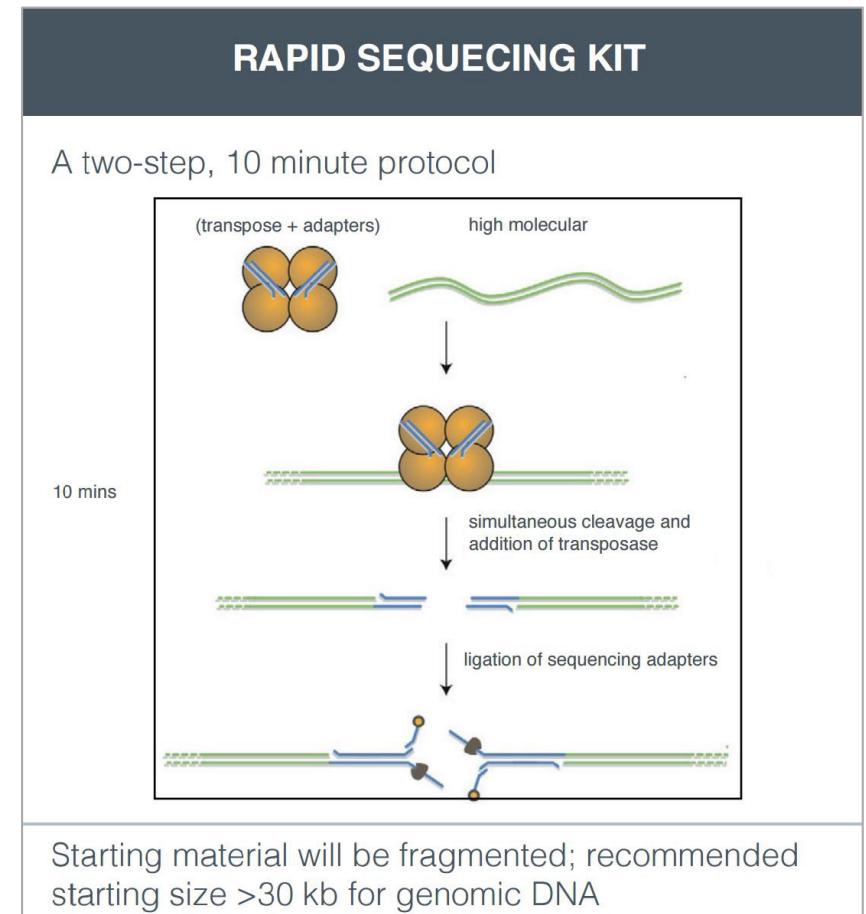
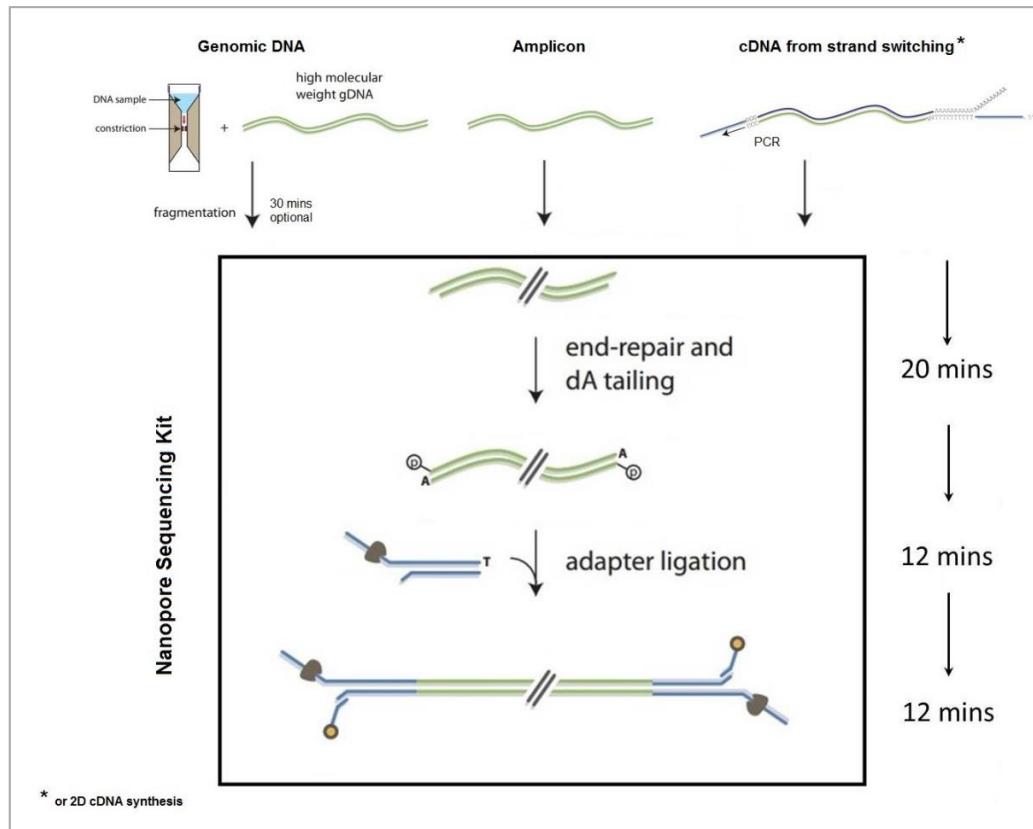
(Goodwin et al., Nat Rev Genet, 2016)

Oxford Nanopore base calling



Szalay & Golovchenko, Nat. Biotech., 2015.

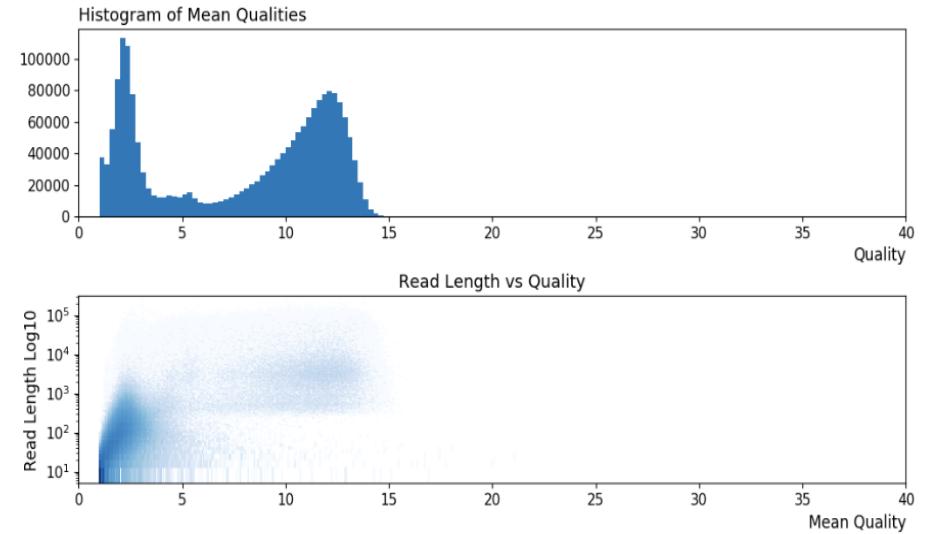
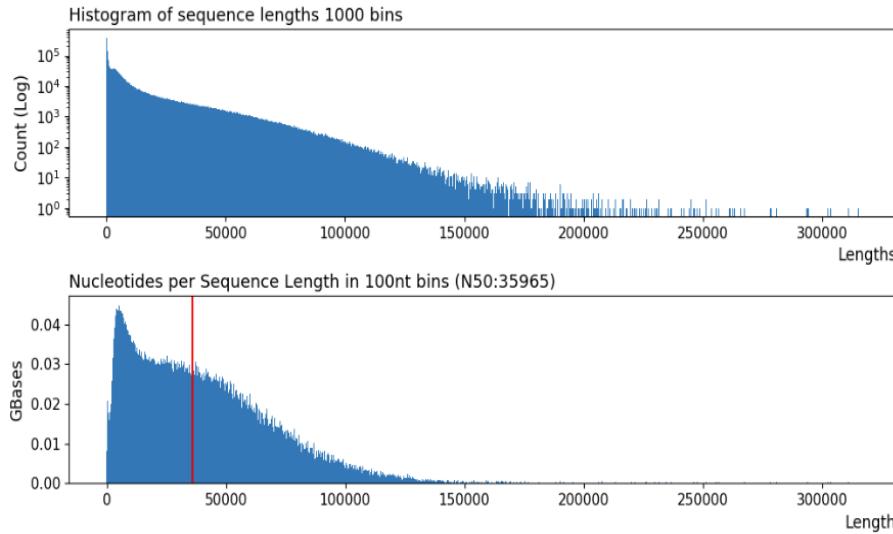
ONT 1D library preparation



1. Ligate an adaptor that contains a motor protein and a cholesterol molecule that attracts the sequence template to the region of the pore for sequencing.

2. Transposase based method that simultaneously cleaves and tags the adaptor sequence to your template (<10 min but fragments your DNA).

ONT performance graphs & data quality



- ✓ Single read raw error rate 2-3%
- ✓ Read lengths are template dependent.
- ✓ Record >2Mb but shorter fragments give higher yield
- ✓ Yield per flowcell depends on DNA quality

Flongle up to 3Gb

MinION 1-30Gb

PromethION 10-200Gb

- ✓ R9.4 pore gives errors around homopolymers >5 bases
- ✓ R10 pore deals with homopolymers much better (Van der Verren et al., Nature Biotech 2020)

Oxford Nanopore platform MinION

MinION



- A USB-based device that runs off a personal computer
- Direct RNA or DNA sequencing
- >10kbp read length,
- 10-20 Gb yield (500USD/flow cell)
- Portability for rapid clinical analysis and **on-field use**

MinION Mk1c



	Basic	Enhanced	Development
MinION ⓘ	Select 1	Select 1	Select Up to 2*
Flow cells ⓘ	2	4	16
Sequencing kits ⓘ	1	2	4
Wash kits ⓘ	1	1	1
Community Support ⓘ	Included	Included	Included
Enhanced Support ⓘ	Optional	8 weeks included	8 weeks included
Rapid Start Day ⓘ	Optional	Optional	Optional
	\$1,000.00	\$4,999.00	\$15,677.00

MinION

Ebola outbreak in Western Africa (2014), a team led by the European Mobile Laboratories in Guinea was able to monitor the transmission history and evolution of the Ebola virus as it unfolded (Quick et al., Nature, 2016).

A. Equipment

Item	Number	Model
Thermocycler	1-3	MasterCycler Personal (Eppendorf)
Fluorometer	1	Qubit 3.0 (Life Technologies)
Laptop	2-3	NT310-H (Stone)
MinION	2-3	-
Pipettes	6	P2, 10, 20, 100, 200, 1000 (Gilson)
Microfuge	1-2	
Dry bath	1	Mini Dry Bath Incubator (Starlab)
Magnetic rack	1	MagnaRack (Life Technologies)
Power strip	1	Dependent on country

B. Consumables

Item	Supplier
DNA LoBind Tubes (2 ml)	Eppendorf
Protein LoBind Tubes (2 ml)	Eppendorf
Qubit Assay Tubes	Life Technologies
PCR Tubes with Flat Caps (0.2 ml)	Starlab
Pipette Tips (10 µl, 20 µl, 100 µl, 200 µl, 1000 µl)	Sarstedt
Nitrile Gloves	Kimberly Clark Professional

C. Reagents

Reagent	Shipping Condition	Supplier
Nuclease-Free Water	Ambient	Qiagen
Ethanol 100%	Ambient	-
HighPrep PCR	Chilled	MAGBIO
Dynabeads His-Tag Isolation and Pulldown	Chilled	Life Technologies
Oligos	Chilled	Sigma
Qubit dsDNA HS Assay Kit	Chilled	Life Technologies
MinION Flowcells	Chilled	Oxford Nanopore Technologies
NEBNext End-Repair Module	Frozen	New England Biolabs
NEBNext dA-Tailing Module	Frozen	New England Biolabs
Blunt/T4 Ligase Master Mix	Frozen	New England Biolabs
SuperScript III One-Step RT-PCR System with Platinum Taq DNA Polymerase	Frozen	Life Technologies
SQK-MAP005	Frozen	Oxford Nanopore Technologies



Figure 1 | Deployment of the portable genome surveillance system in Guinea. **a**, We were able to pack all instruments, reagents and disposable consumables within aircraft baggage. **b**, We initially established the genomic surveillance laboratory in Donka Hospital, Conakry, Guinea. **c**, Later we moved the laboratory to a dedicated sequencing laboratory in Coyah prefecture. **d**, Within this laboratory we separated the sequencing instruments (on the left) from the PCR bench (to the right). An uninterruptible power supply can be seen in the middle that provides power to the thermocycler. (Photographs taken by J.Q. and S.D.)

Quick et al. *Genome Biology* (2015) 16:114
DOI 10.1186/s13059-015-0677-2



RESEARCH

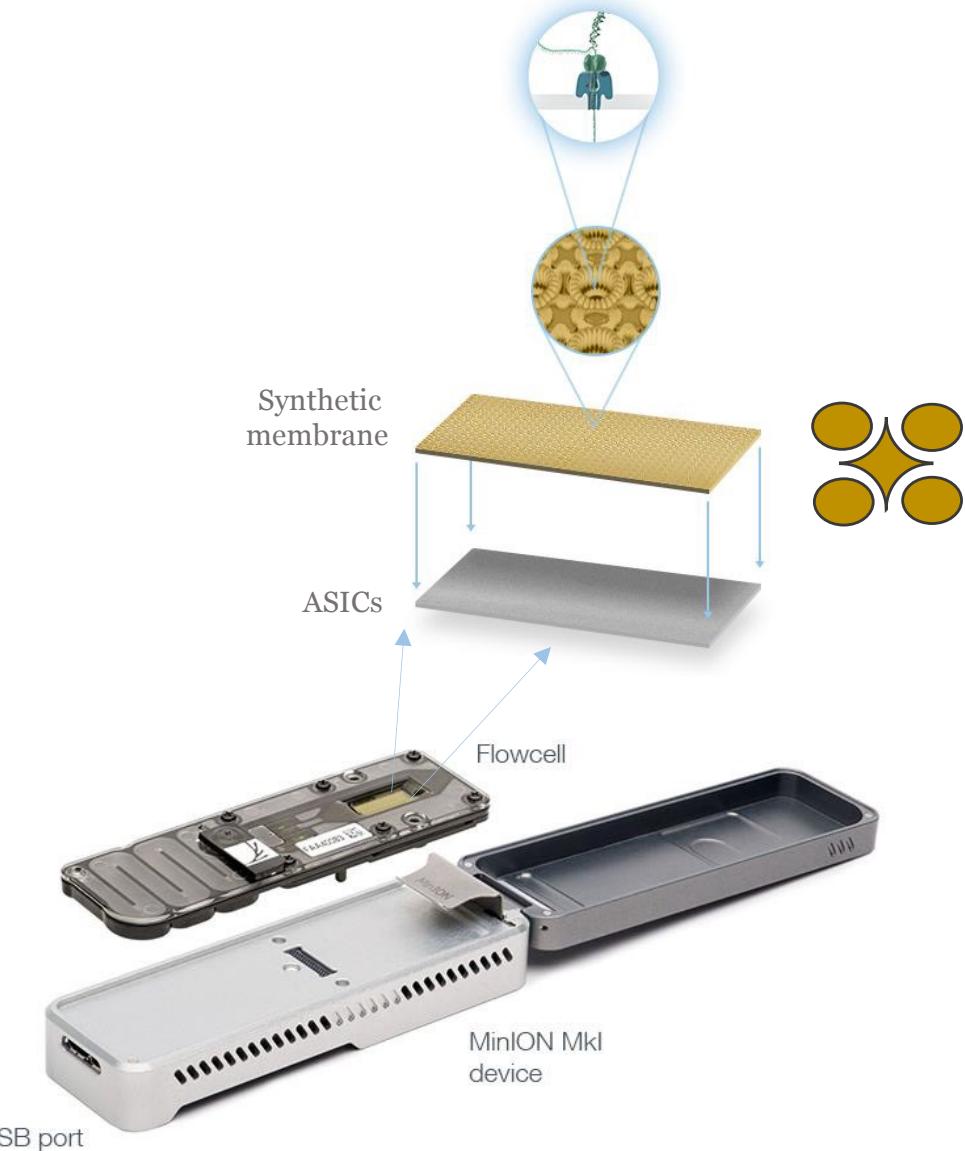
Open Access



Rapid draft sequencing and real-time nanopore sequencing in a hospital outbreak of *Salmonella*

Flow cell design MinION

- ✓ Application-Specific Integrated Circuits (ASICs) contains 512 channels
- ✓ Each channel is surrounded by **4 pores** & records **only 1** at the time
- ✓ 512 pores max recorded at the time
- ✓ Scan for “fresh” active pores automatically every 24h or when manual restarted



SmidgION & Flongle



SmidgION

Uses the same core as MinION
but is designed for use with
smartphones for field-based
applications



Flongle

A flow cell adapter
\$90 for 2-3 Gb
Disposable cartridge device
Reusable electronics
Lowest cost NGS solution!

GridION X5 and PromethION

ONT platforms for service sector

- A benchtop system designed to run and analyze up to 5 MinION Flow Cells
- **P24, P48:** >10kbp read length, 24 or 48 flow cells, 30-90GB yield per flow cell, 144,000 nanopores at a time.



GridION X5

- Multiple sequencing devices, one compute module
- Use up to five MinION Flow Cells at a time
- Benchtop processor capable of handling high data volumes in real time
- Rapid, real-time applications such as *Read Until...*

[About GridION](#)[Get in touch](#)

Choose GridION X5 if you:

- would like to offer nanopore sequencing as a service
- want the choice to invest from a CapEx or consumable budget
- work on larger sequencing projects (50–100Gb per 48 hours)
- would like on-device basecalling – no local infrastructure requirement.



PromethION

- High-throughput, high-sample number benchtop system
- Modular: Up to 48 flow cells, each with up to 3,000 nanopore channels (total up to 144,000)
- Flow cells may be run individually or concurrently
- Same workflow as MinION at larger scale

[About PromethION](#)[Early Access](#)

Choose PromethION if you:

- would like to offer nanopore sequencing as a service
- are interested in very large data volumes projects (Tb)
- are seeking on-demand sequencing for large numbers of samples
- would like to avoid CapEx investments.

	GridION X5 Starter Pack	GridION X5 – CapEX	GridION X5 – OpEX
GridION <small>1</small>	1	1	1
Flow cells included <small>1</small>	60	0	300
Price per flow cell <small>1</small>	-	\$299.00	\$475.00
Sequencing kits included <small>1</small>	10	Bought separately	Bought separately
Enhanced support <small>1</small>	Included	Included	Included
Device maintenance and service <small>1</small>	Included (6 months)*	Included (12 months)*	Included (12 months)*
Shipping <small>1</small>	Included	Included	Included
	\$49,955.00	\$126,000.00	\$158,500.00

Purchasing the PEAP bundle will give you:

PEAP
PromethION Instrument
No charge
Starter Pack (12 PromethION Flow Cells and kits)
\$28,575
Compute service charge***
\$15,000 p.a.
12 MinION Flow Cells**
\$8,100
48 PromethION Flow Cells*
\$69,600
Shipping and Installation
\$5,269
Reagent kits 10 ligation, 3 rapid, 4 wash
\$8,456
Bundle \$135,000

Summary Nanopore platforms

Nanopore Live

NANOPORE SYSTEMS

Summary



	MinION	GridION X5	PromethION
Sequencer type	Mobile	Benchtop	Benchtop
System Price	Starter pack of \$1000	\$0 when ordering 300 flow cells	Starter pack of \$135,000
Data produced by starter pack <i>(based on internal test Mar 17)</i>	Up to 40GB	Up to 6TB	Coming soon
Fee For Service available	No	Yes	Yes
Specifications based on internal test Mar 2017			
Run Time	1 min – 48 hours	1 min – 48 hours	1 min – 48 hours
Yield per flow cell	20GB	20GB	50GB*
Yield per Instrument run	20GB	100GB	2.4TB*

* PromethION yield still in development through the PromethION Early Access Programme

Oxford

ONT platforms can be bought for no upfront capital cost

Capital free devices Starter packs can be bought with consumable budget						LC Life Sciences
						
Flongle	MinION Mk 1B	MinION Mk 1C	GridION Mk 1	P24	P48	
Starter pack: \$1,870	Starter pack: \$1,000	Starter pack: \$4,900	Starter pack: \$49,955	Starter pack: \$165,000	Starter pack: \$287,000	
Includes: 12 flow cells	Includes: 2 flow cells + kit	Includes: 6 flow cells + kit <small>12 month software licence & warranty*</small>	Includes: 60 flow cells + kit <small>12 month software licence & warranty**</small>	Includes: 60 flow cells + kit <small>12 month software licence & warranty**</small>	Includes: 120 flow cells + kit <small>12 month software licence & warranty**</small>	

ONT applications

- “Read until” done. W.I.M.P
- Selective reads
- Cas9 mediated enrichment of target sequences
- Mobile sequencing (artic, jungle)
- Direct methylation detection probability
- Direct RNA sequencing

Oxford Nanopore: pros and cons

Advantages:

- Long read sequencing, typically 5-100Kb (>1Mb possible). Run until done
- Portability and low capital cost
- Single molecule, real-time base-calling and data analysis
- Detection of modified bases: shift in the typical voltage for a given k-mer
- Direct RNA sequencing
- Zero capital cost price instruments available – fast moving technology

Disadvantages:

- Pay more for sequencing reagents
- High error rate (**1%**) dominated by indels and transition errors
- 99.3% accuracy rate single
- 99.9% accuracy rate duplex
- Systematic errors around homopolymers: it can be difficult to identify when one k-mer leaves the pore and another k-mer enters

Other long-range NGS approaches

Table 1 | Long-range sequencing and mapping platforms

Platform	General characteristics and costs	Major applications	Bioinformatics challenges
PacBio SMRT sequencing	Single-molecule long reads averaging ~10 kb with some approaching 100 kb; several fold more expensive than short reads	De novo genome assembly, structural variant detection, gene isoform resolution and epigenetic modifications	Raw reads have high error rates dominated by false insertions; requires new alignment and error correction algorithms
Oxford Nanopore sequencing	Single-molecule long reads averaging ~10 kb with some >1 Mb; several fold more expensive than short reads	De novo genome assembly, structural variant detection, gene isoform resolution and epigenetic modifications	Raw reads have high error rates dominated by false deletions and homopolymer errors; requires new alignment and error correction algorithms
10X Genomics Chromium	Linked reads spanning ~100 kb derived from a collection of short-read sequences; moderately more expensive than short reads	De novo genome assembly and scaffolding, phasing, detection of large structural variants (>10 kb) and single-cell gene expression	Sparse sequencing rather than true long reads; more complicated to align, with poorer resolution of locally repetitive sequences
Hi-C-based analysis	Pairs of short reads (<100 bp) formed from crosslinking chromatin interactions; moderately more expensive than short reads	Genome scaffolding and phasing	Sparse sequencing with highly variable genomic distance between pairs (1 kb to 1 Mb or longer)
BioNano Genomics optical mapping	Optical mapping of long DNA molecules (~250 kb or longer) labelled with fluorescent probes; less expensive than short reads	Genome scaffolding and detection of large structural variants (>10 kb)	Limited algorithms to discover high-confidence alignment between an optical map and a sequence assembly

PacBio SMRT, Pacific Biosciences single-molecule real time.

(Sedlazeck et al., *Nat Rev Genet*, 2018)

The GenapSys™ Sequencing Platform

Sequencing Without Compromise

Exceptional Accuracy

Highly accurate data validated by numerous applications in experienced third party labs

Unrivaled Scalability

A range of outputs tuned to your sample throughput needs

Amazing Affordability

Modest run and instrument pricing enable operational flexibility



GenapSys Sequencer

GenapSys Sequencing
Prep System

A novel and robust Next-Generation Sequencing technology is here, with advantages for labs of all types. The compact GenapSys™ Sequencer combines electronic data detection, CMOS chip technology (Complementary Metal Oxide Semiconductor), and proven sequencing by synthesis (SBS) chemistry to deliver high accuracy data. The sequencer with the 16M chip generates 1.2 - 2.0 Gb of data per run and delivers the high resolution and analytical sensitivity needed for detection of rare variants and transcripts.

- Fast, cost-effective runs that match your sample throughput needs and applications
- Perform runs on your own schedule, without having to wait for additional sample batching
- Control the sequencing process from beginning to end for higher confidence in sample integrity and data analysis results
- Affordable list price puts sequencing within the reach of virtually any lab

Thank you very
much for your
attention!