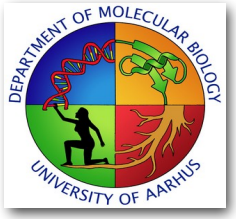
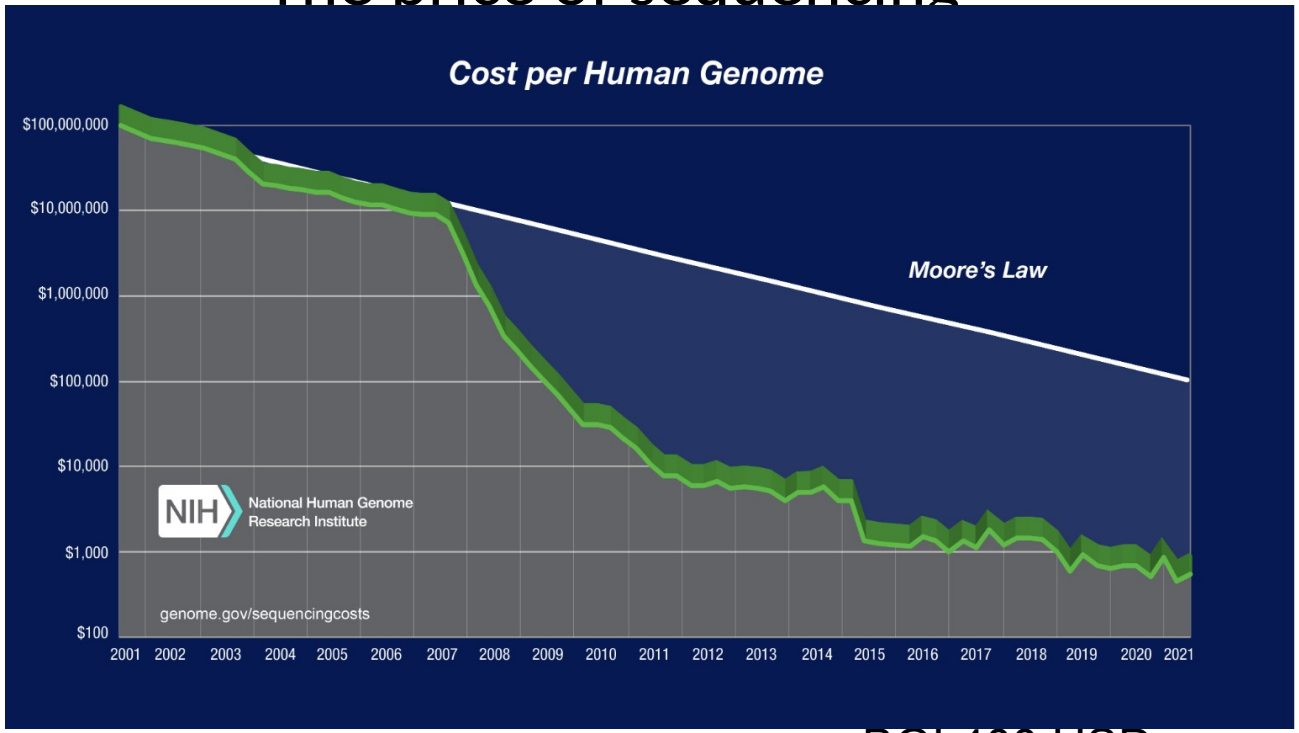
Sequencing technologies



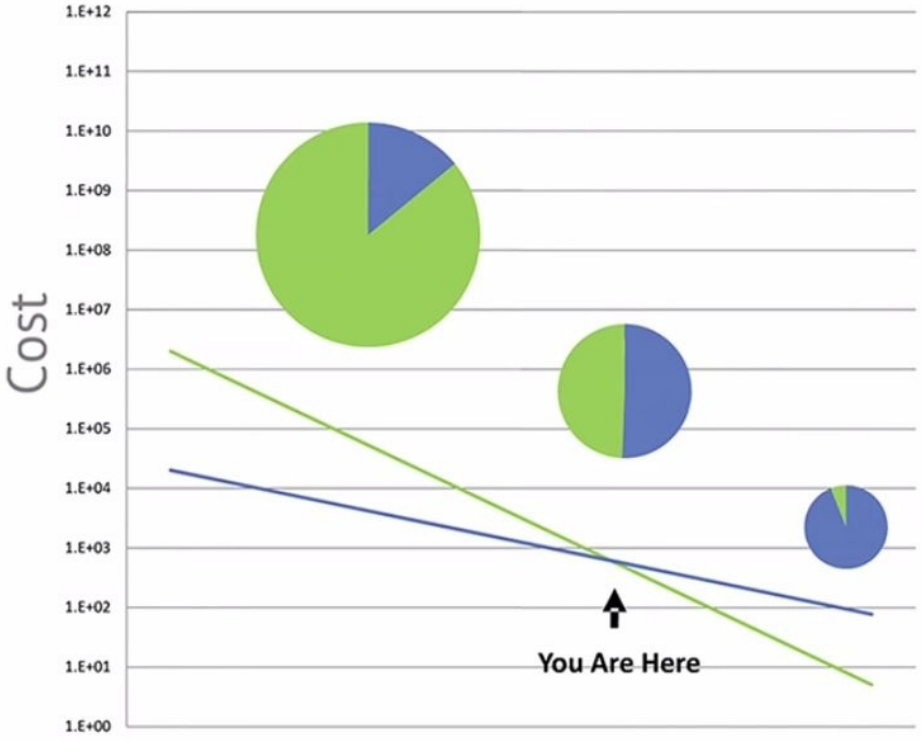
The price of sequencing



Reagents
Machine cost
Labour to run instrument
Bioinformatics
High performance computing

BGI 400 USD
Pacbio CLS: 2000-3000 USD
Pacbio HIFI: 8000 USD
Nanopore: 800 USD

DNA Sequencing Economics



Sequencing cost dropping ~5x per year

Informatics cost dropping ~2x per year

Informatics is now the bottleneck

Sequencing, sequencing, sequencing

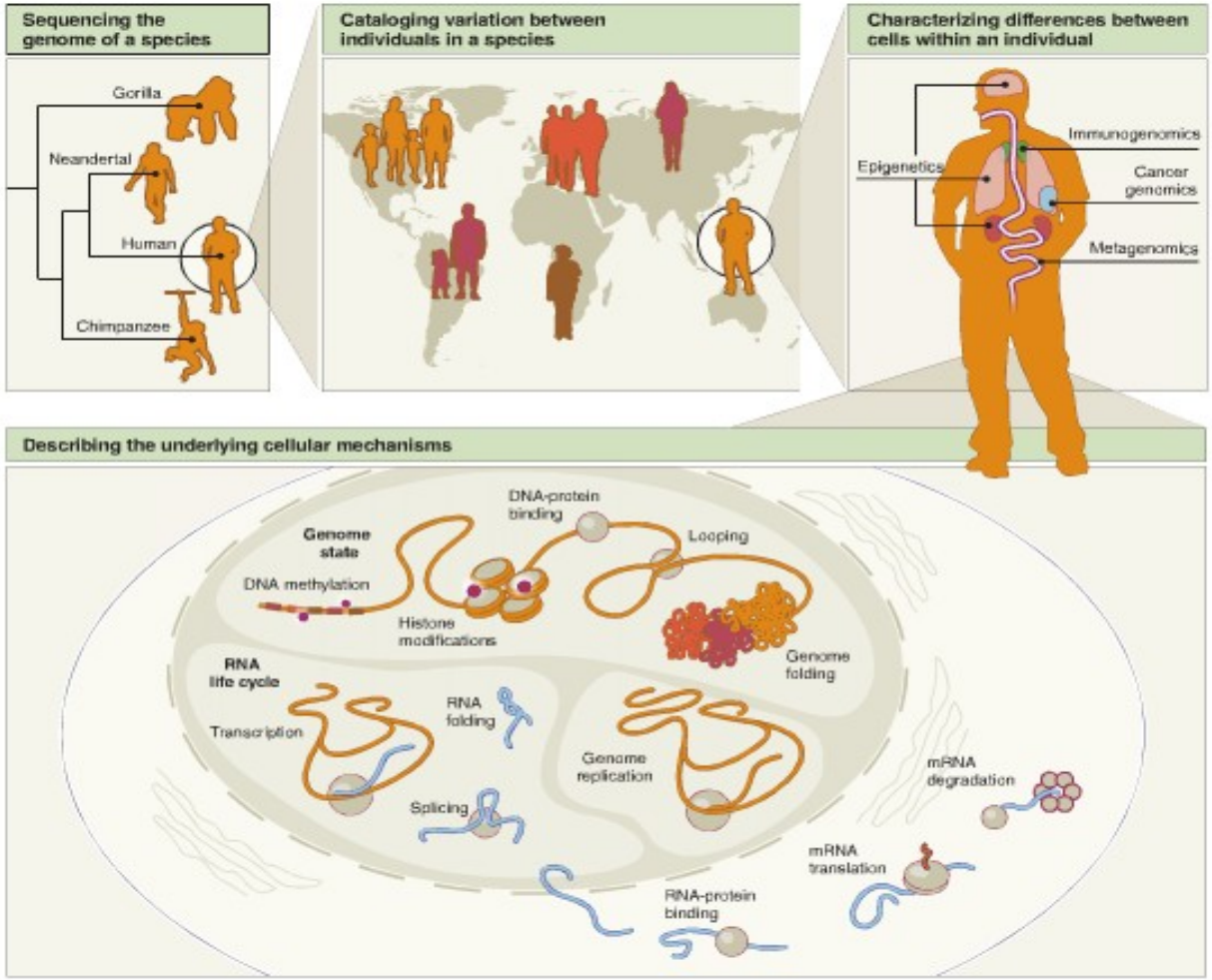
All eucaryotic species on Earth within the next 5 years

2-3 million known (10-15 million expected)

All birds (10,000 species) – 500 done

All primates (500 species) – 251 done

When is a sequence final?


Sequencing the genome of a species
Gorilla
Neandertal
Human
Chimpanzee
Cataloging variation between individuals in a species
Characterizing differences between cells within an individual
Immunogenomics
Epigenetics
Cancer genomics
Metagenomics
Describing the underlying cellular mechanisms
DNA-protein binding
Looping
Genome state
DNA methylation
Histone modifications
Genome folding
RNA life cycle
RNA folding
Transcription
Genome replication
mRNA degradation
Splicing
mRNA translation
RNA-protein binding

Technologies

First generation

- Sanger sequencing (from 1977)

Second generation

- 454 (2005)
- Illumina (2006)
- SOLID (2006)
- Qiagen gene reader (2016)
- Complete Genomics (2008) – BGI revamp (2015) – BGI revamp (2019)
- Ultima (2022)

Third generation competitors

- Helicos (2008)
- Ion torrent (2010)
- Pacific Bioscience (2013)

Fourth generation

- Oxford Nanopore (2014)

Technologies

First generation

- **Sanger sequencing (from 1977) – almost forgotten**

Second generation

- ~~454 (2005)~~
- **Illumina (2006)**
- ~~SOLID (2006)~~
- ~~Qiagen gene reader (2016) – clinical~~
- Complete Genomics (2008) – BGI revamp (2015) – **BGI revamp (2019)**
- **Ultima (2022)????**

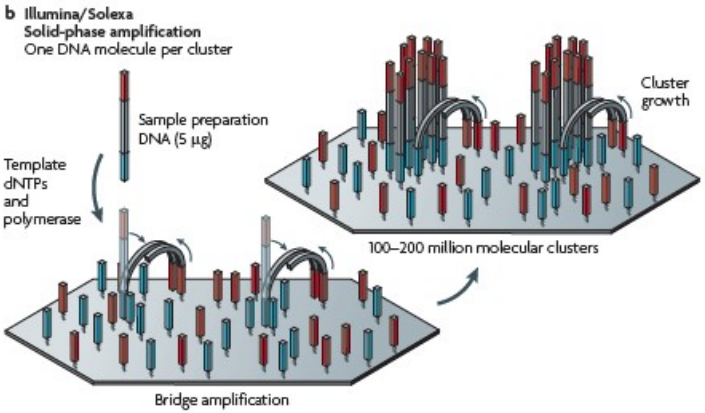
Third generation competitors

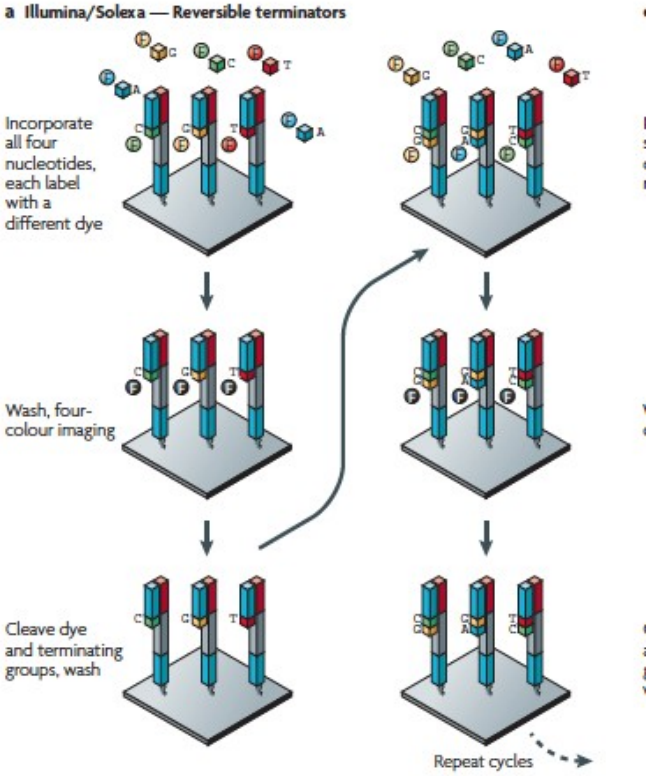
- ~~Helicos (2008)~~
- Ion torrent (2010)
- **Pacific Bioscience (2013)**

Fourth generation

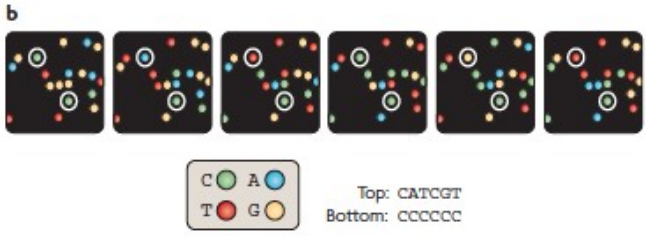
- **Oxford Nanopore (2014)**

Illumina technology

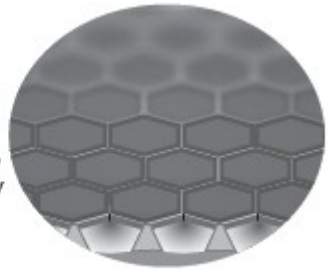




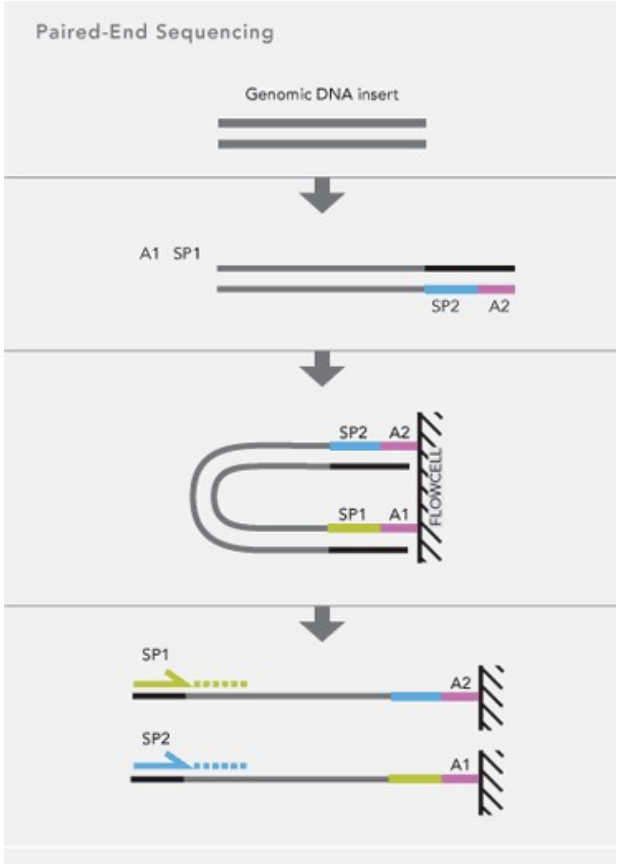
All four colours at once, thus all sequences equally long, max 250 bp



Patterned flow cell
Microwells on flow cell direct cluster generation, increasing cluster density



Illumina paired end sequencing



Illumina throughput




	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		●	●	●

How to Choose a Benchtop Sequencer

This Benchtop Sequencing Buyer's Guide has tips to help you make a smooth transition to next-generation sequencing and select the best benchtop sequencing system to achieve your research objectives.

Read Benchtop Buyer's Guide ›

Run Time	9–17.5 hours	4–24 hours	4–55 hours	12–30 hours
Maximum Output	1.2 Gb	7.5 Gb	15 Gb	120 Gb
Maximum Reads Per Run	4 million	25 million	25 million †	400 million
Maximum Read Length	2 × 150 bp	2 × 150 bp	2 × 300 bp	2 × 150 bp
	Explore iSeq	Explore MiniSeq	Compare MiSeq	Compare NextSeq

Next sequencer, NovaSeq 600, 100 USD human genome




	NextSeq Series	HiSeq Series	HiSeq X Series‡	NovaSeq 6000 System
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	●	●		●
miRNA & Small RNA Analysis	●	●		●
DNA-Protein Interaction Analysis	●	●		●
Methylation Sequencing	●	●		●
Shotgun Metagenomics	●	●		●

Optimized NGS Sample Tracking and Workflows

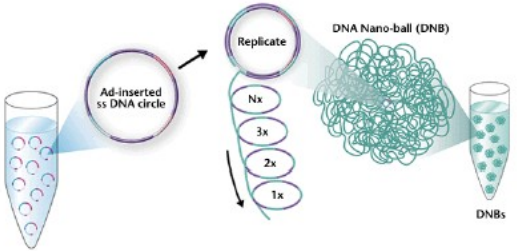
See how BaseSpace Clarity LIMS (Laboratory Information Management System) enabled this large genomics lab to standardize lab procedures and cope with increasing sample volumes from diverse clients.

Read Case Study ›

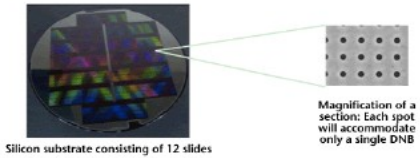
Run Time	12–30 hours	< 1–3.5 days (HiSeq 3000/HiSeq 4000) 7 hours–6 days (HiSeq 2500)	< 3 days	16–36 hours (Dual S2 flow cells) 44 hours (Dual S2 flow cells)
Maximum Output	120 Gb	1500 Gb	1800 Gb	6000 Gb
Maximum Reads Per Run	400 million	5 billion	6 billion	20 billion
Maximum Read Length	2 × 150 bp	2 × 150 bp	2 × 150 bp	2 × 150 bp

BGI-seq based on nanoballs

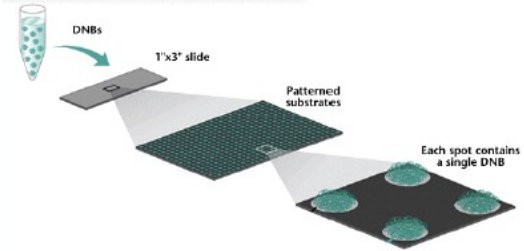
ssDNA -> DNA nanoballs



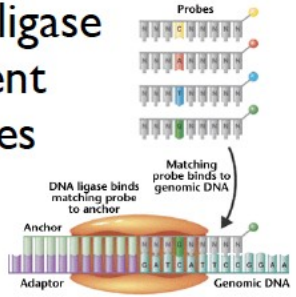
Use silicon chips with sticky spots



Place DNBs into each spot



Sequence using ligase and flourescent labeled probes

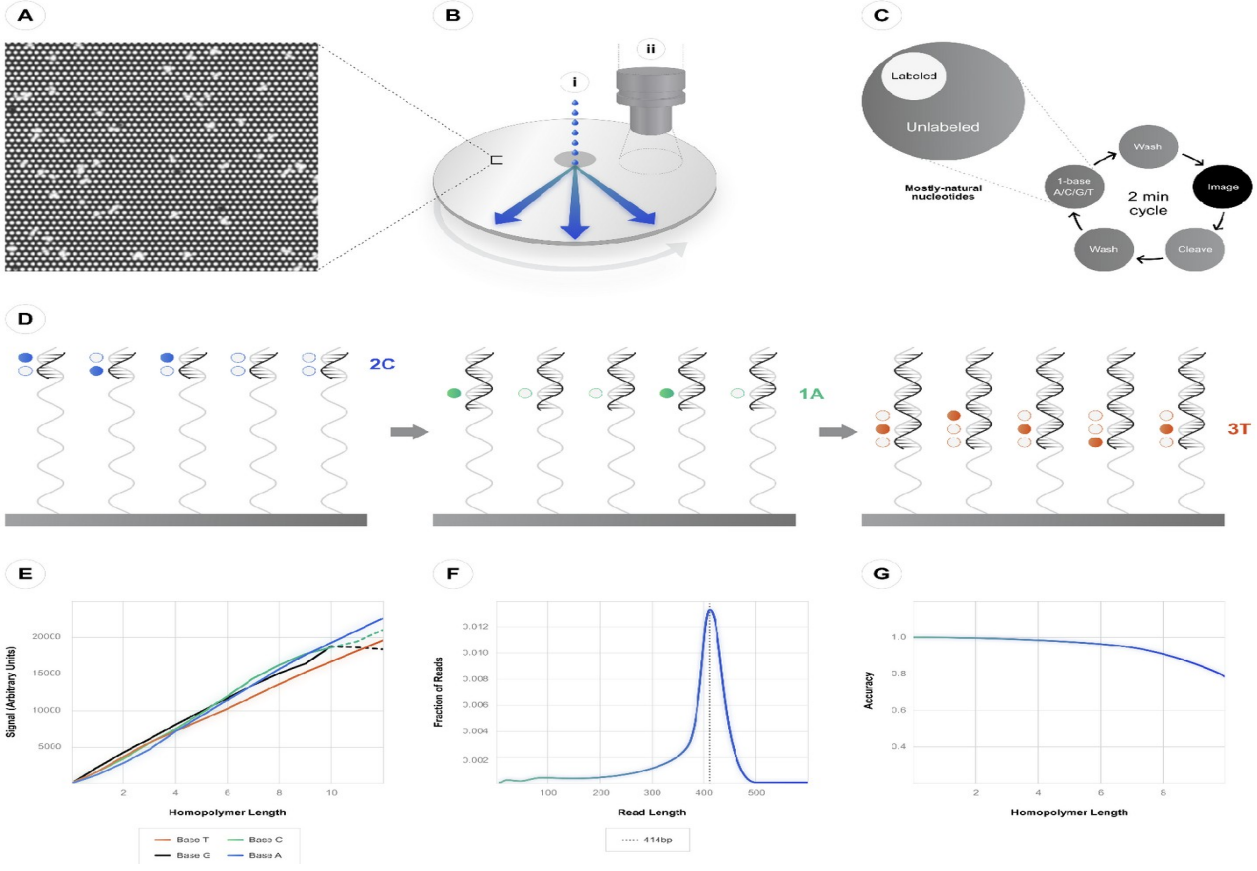


Credit Simon Rasmussen

Illumina versus BGIseq

PARAMETER	DNBSEQ (average)	HiSeq X Ten (average)
Clean reads (M)	1,001	732
Clean data amount (Gb)	100	110
Clean read Q20 (%)	95	97.01
Clean read Q30 (%)	89.9	90.47
GC content (%)	41.71	40.94
Mapping rate (%)	99.47	96.52
Unique rate (%)	94.33	85.14
Duplicate rate (%)	1.77	11.76
Mismatch rate (%)	0.53	0.56
Average sequencing depth	33	31
Coverage (%)	99.1	98.95
Coverage at least 4X (%)	98.62	98.43
Coverage at least 10X (%)	97.68	97.24
Coverage at least 20X (%)	93.09	91.45

ULTIMA sequencing (100 USD genome)

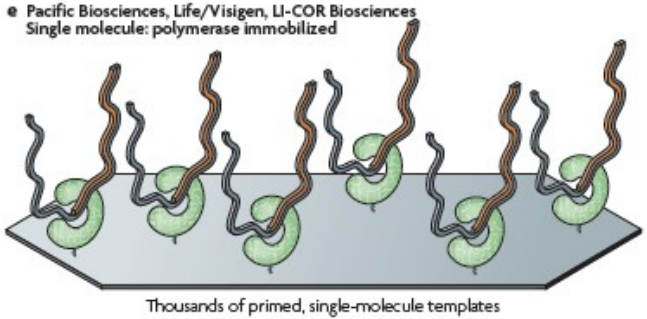


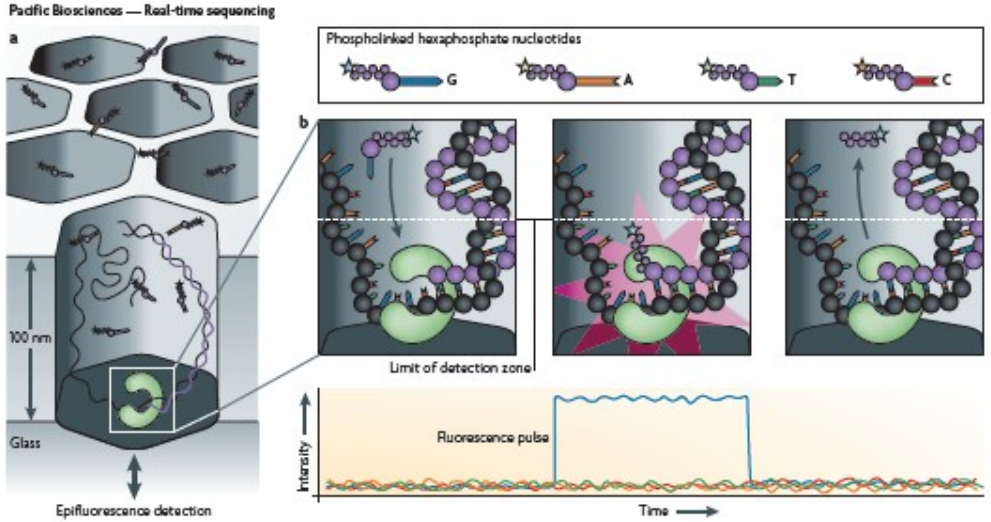
Third generation sequencing

- Simpler chemistry
- Longer reads
  - Easier assembly
  - Easier phasing of chromosomes
- Direct measurement of DNA modifications

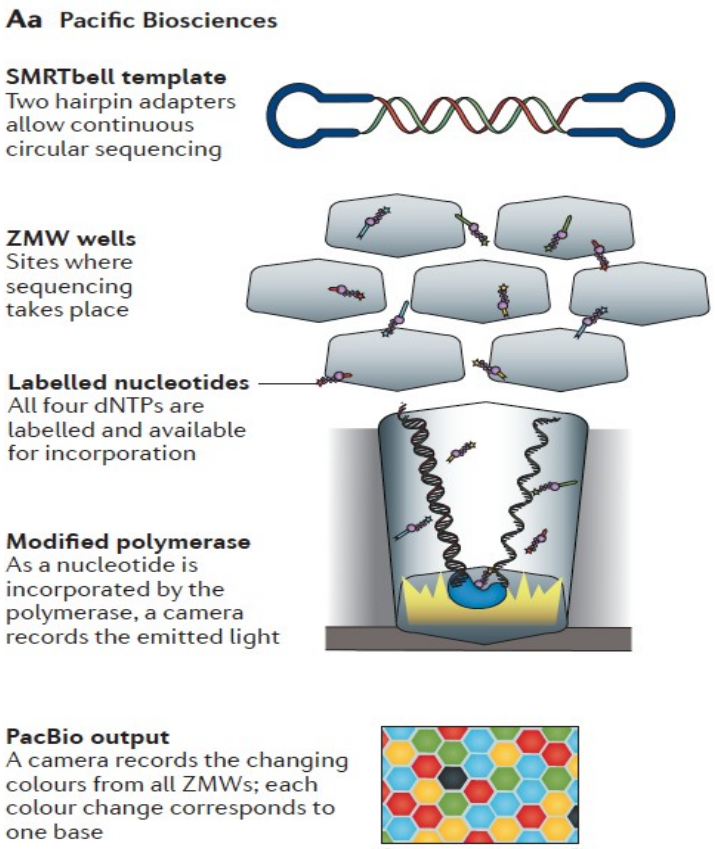
Pacific biosciences

Real time sequencing





https://www.pacb.com/smrt-science/smrt-sequencing/



PacBio accuracy and capacity

System Performance

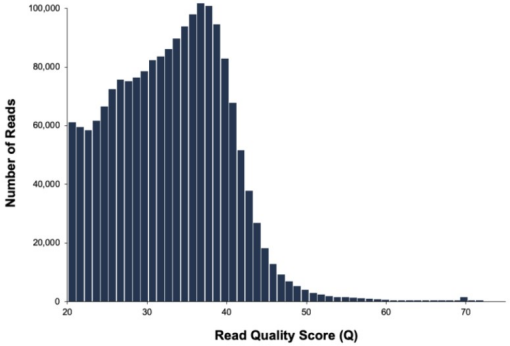
Example data from genomic libraries generated using the continuous long read (CLR) and HiFi read modes of sequencing on the Sequel II System.

Highly Accurate Long Reads

HiFi Sequencing

Number of >99% (Q20) 9-13 kb Reads:

Up to 2 million



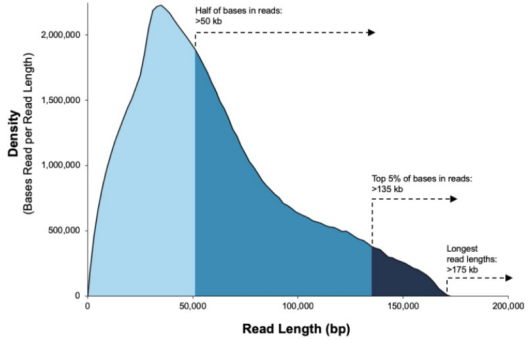
Data from a 11 kb size-selected human library using the SMRTbell Template Prep Kit 1.0 on a Sequel II System (1.0 Chemistry, Sequel II System Software v7.0, 30-hour movie)*.

Long Read Lengths

CLR Sequencing

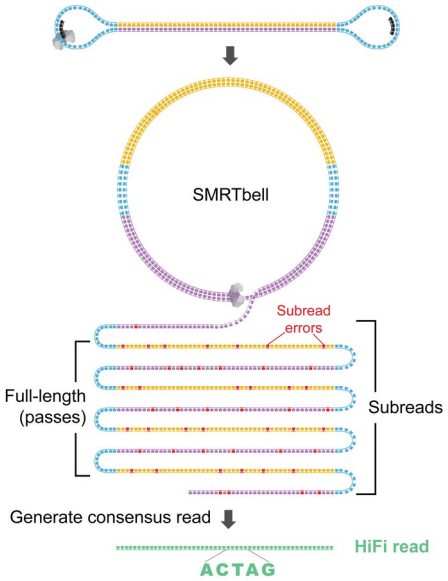
Half the Data in Reads: >50 kb

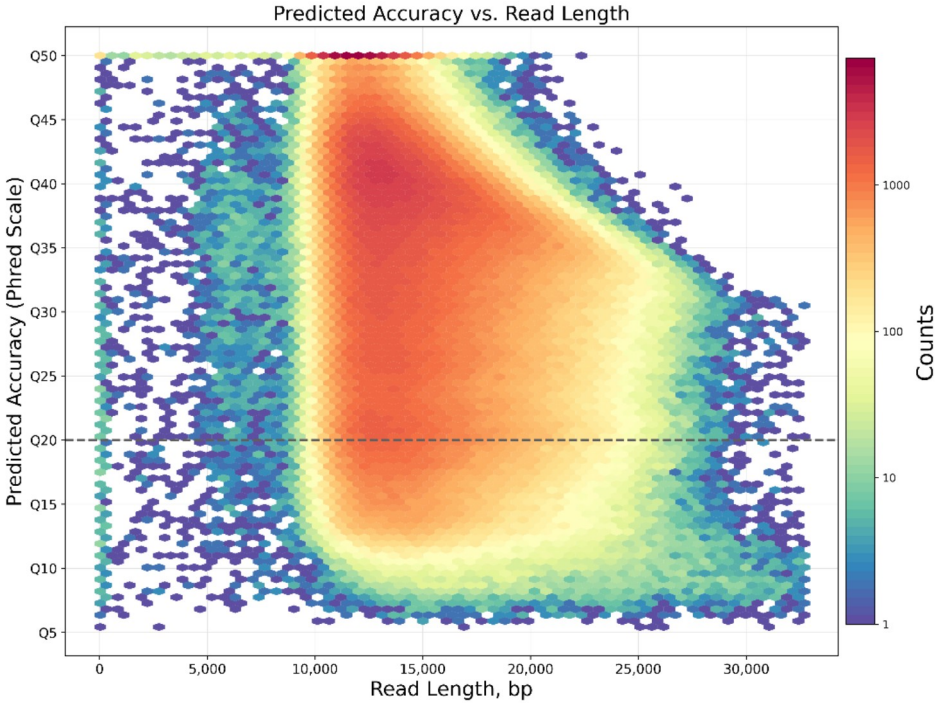
Data per SMRT Cell: Up to 160 Gb



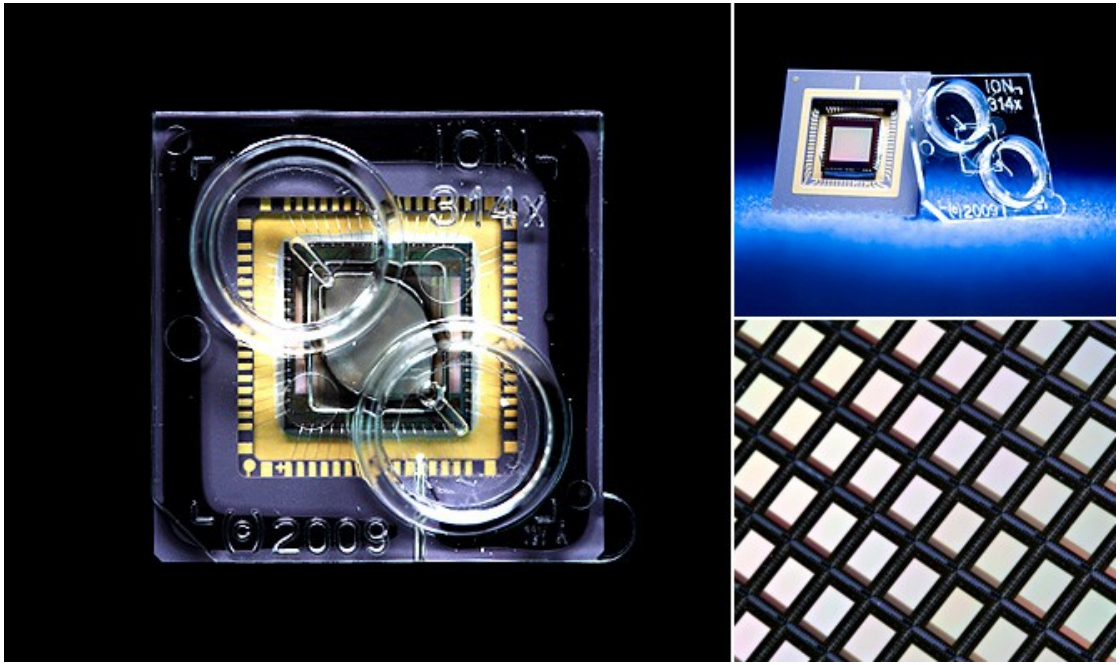
Data from a 35 kb size-selected *E. coli* library using the SMRTbell Express Template Prep Kit 2.0 on a Sequel II System (1.0 Chemistry, Sequel II System Software v7.0, 15-hour movie)*.

Pacbio long read sequencing



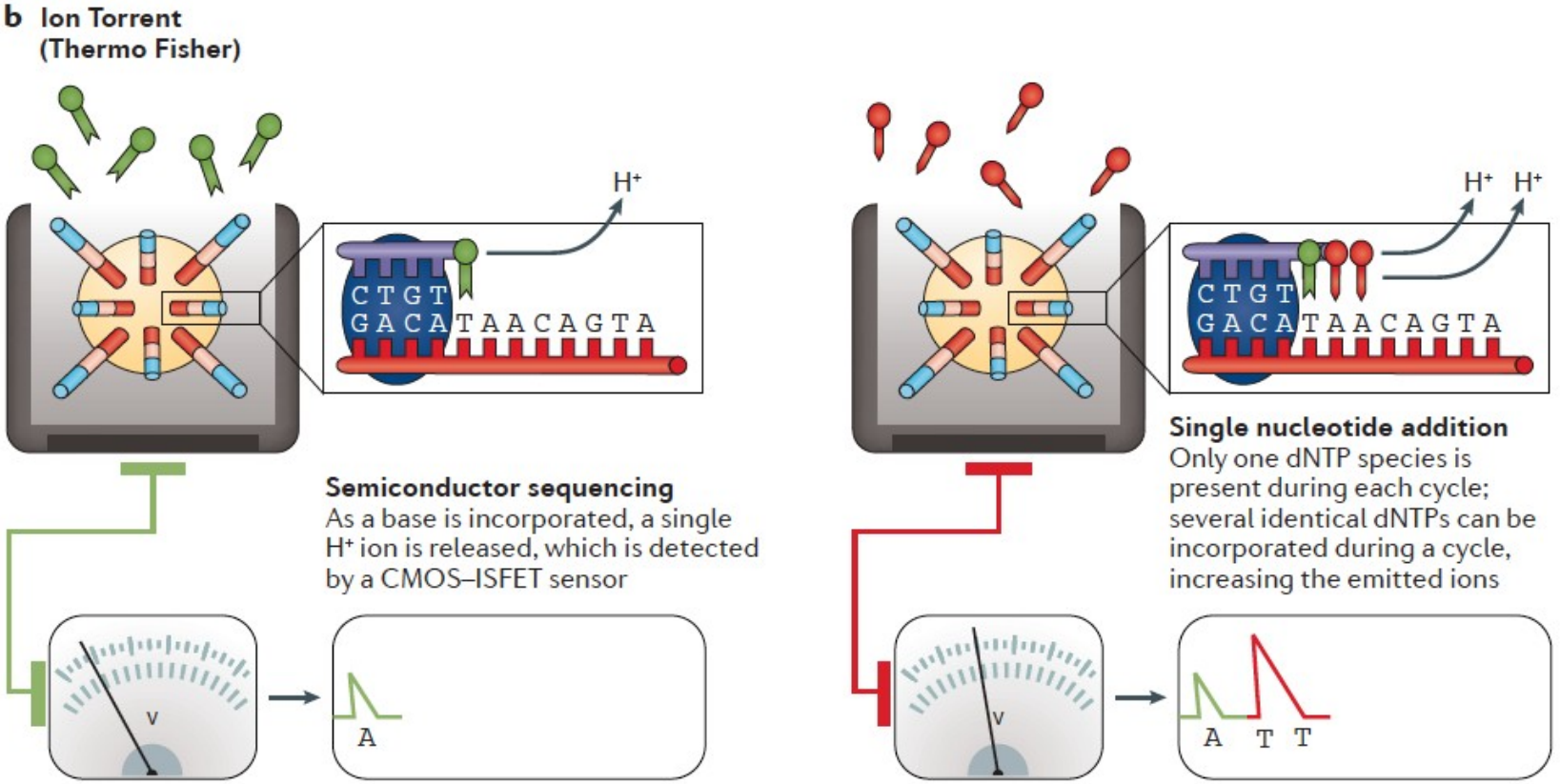


Ion torrent

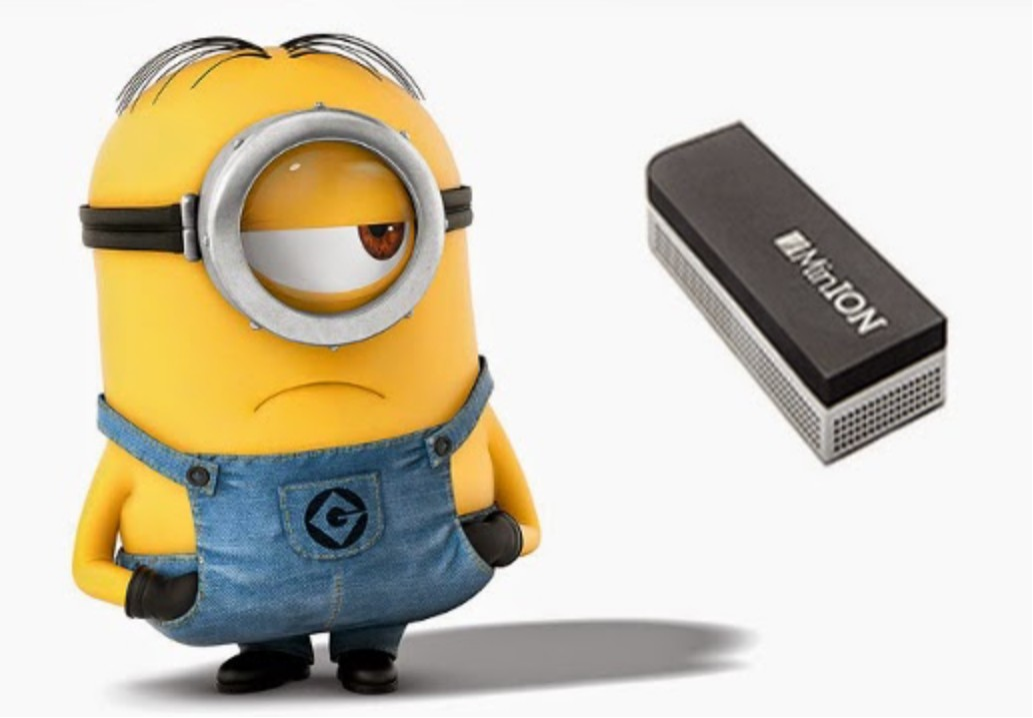


Sequencing on a semiconductor chip which is a very precise pH-meter

Ion torrent

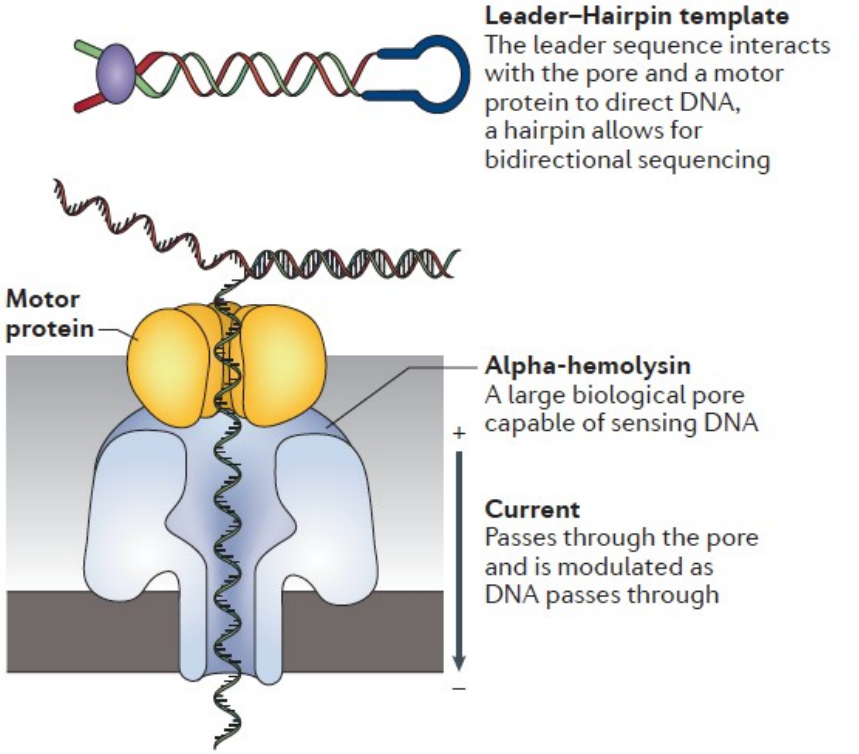

b Ion Torrent
(Thermo Fisher)
H+
C T G T
G A C A T A A C A G T A
Semiconductor sequencing
As a base is incorporated, a single H+ ion is released, which is detected by a CMOS–ISFET sensor
V
A
H+ H+
C T G T
G A C A T A A C A G T A
Single nucleotide addition
Only one dNTP species is present during each cycle; several identical dNTPs can be incorporated during a cycle, increasing the emitted ions
V
A T T

Oxford Nanopore

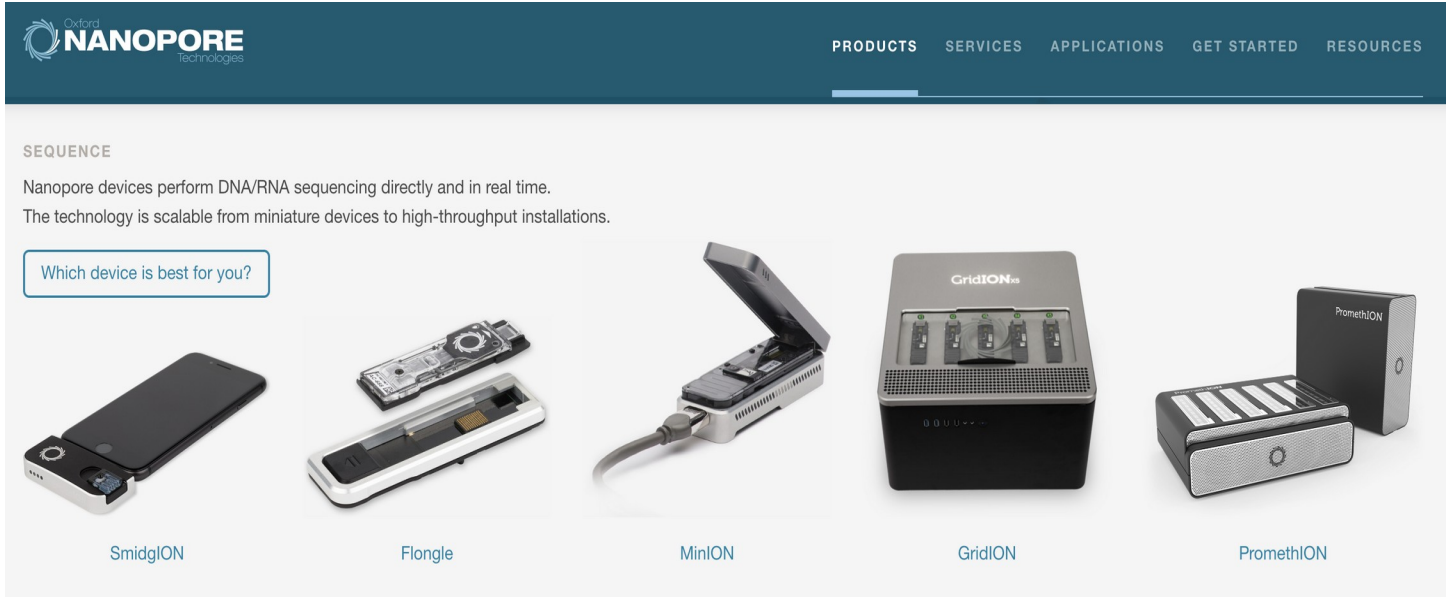


Long pieces, only DNA cleavage, voltage measurement the key, scalable

Ab Oxford Nanopore Technologies



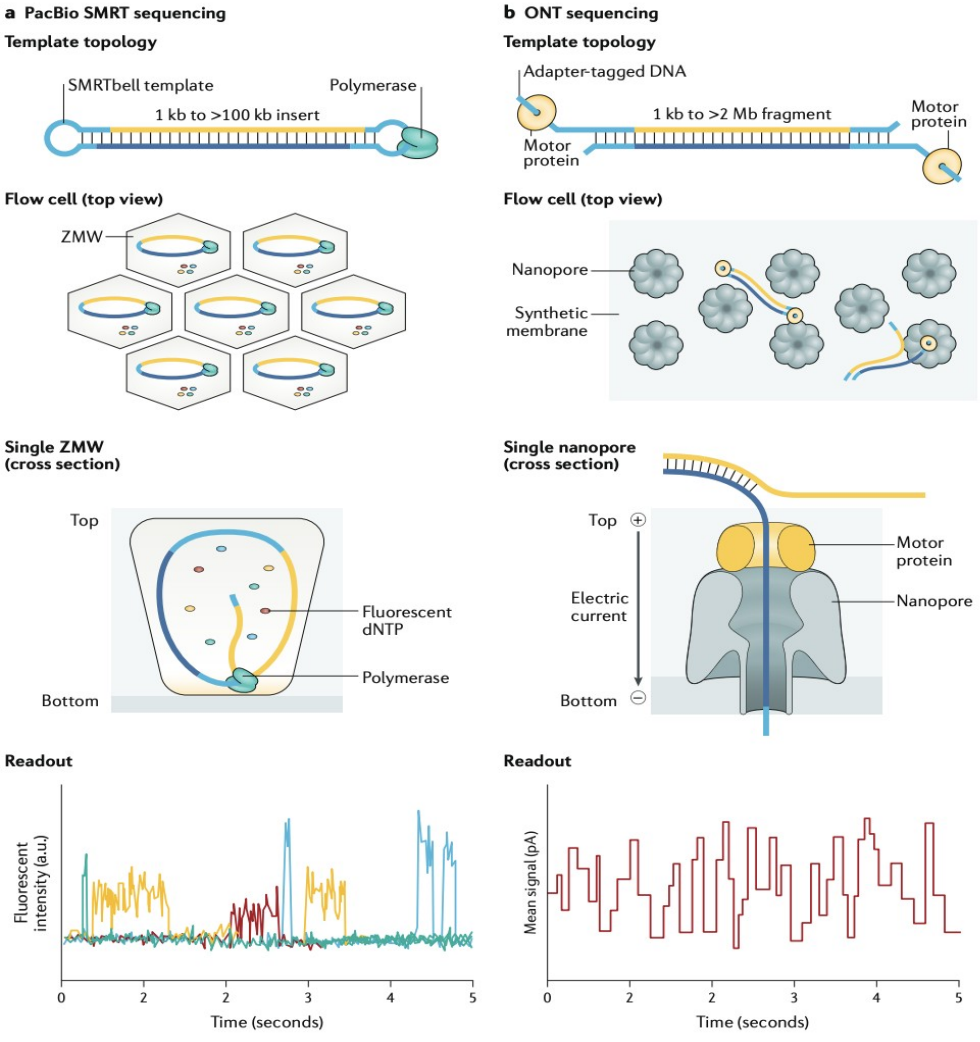
The minION, gridION, promethION, smidgION



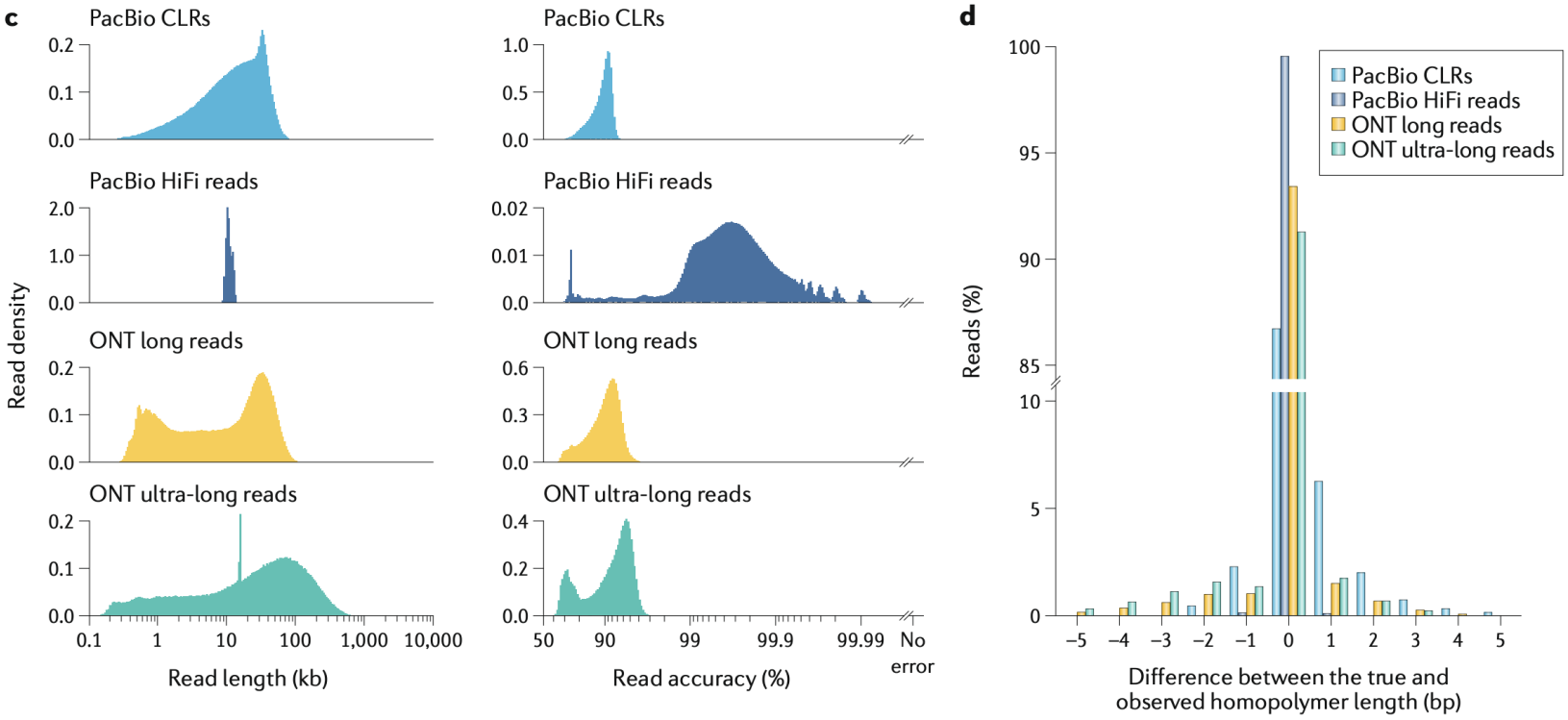
Accuracy versus throughput

Sequencing technology	Platform	Data type	Read length (kb)		Read accuracy (%)	Throughput per flow cell (Gb)		Estimated cost per Gb (US$)	Maximum throughput per year (Gb)[a]
			N50	Maximum		Mean	Maximum		
Pacific Biosciences (PacBio)	RS II[b]	CLR	5–15	>60	87–92	0.75–1.5	2	333–933[c]	4,380
	Sequel	CLR	25–50	>100		5–10	20	98–195[d]	17,520
	Sequel II	CLR	30–60	>200		50–100	160	13–26[e]	93,440
		HiFi	10–20	>20	>99	15–30	35	43–86[e]	10,220
Oxford Nanopore Technologies (ONT)	MinION/ GridION	Long	10–60	>1,000	87–98	2-20	30	50–500[f]	21,900 (MinION) 109,500 (GridION)
		Ultra-long	100–200	>1,500		0.5–2	2.5	500–2,000[f]	913 (MinION) 4,563 (GridION)
	PromethION	Long	10–60	>1,000		50–100	180	21–42[f]	3,153,600
Illumina	NextSeq 550	Single-end	0.075–0.15	0.15	>99.9	16–30	>30	50–63[g]	>47,782
		Paired-end	0.075–0.15 (×2)	0.15 (×2)		32–120	>120	40–60[g]	>70,080
	NovaSeq 6000	Single-end	0.05–0.25	0.25		65–3,000	>3,000	10–35[h]	>1,194,545
		Paired-end	0.05–0.25 (×2)	0.25 (×2)					

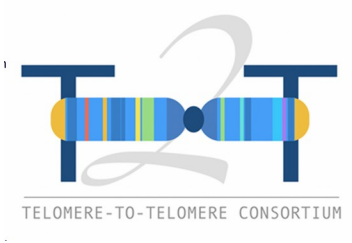
Technical differences



Comparison of long read methods



A complete human genome



Sequencing a hydatidiform mole ie a haploid genome

Using Nanopore and Pacbio HIFI

STATISTICS	GRCH38	T2T-CHM13	DIFFERENCE (±%)
Summary			
Assembled bases (Gbp)	2.92	3.05	+4.5
Unplaced bases (Mbp)	11.42	0	−100.0
Gap bases (Mbp)	120.31	0	−100.0
Number of contigs	949	24	−97.5
Contig NG50 (Mbp)	56.41	154.26	+173.5
Number of issues	230	46	−80.0
Issues (Mbp)	230.43	8.18	−96.5
Gene annotation			
Number of genes	60,090	63,494	+5.7
Protein coding	19,890	19,969	+0.4
Number of exclusive genes	263	3,604	
Protein coding	63	140	
Number of transcripts	228,597	233,615	+2.2
Protein coding	84,277	86,245	+2.3
Number of exclusive transcripts	1,708	6,693	
Protein coding	829	2,780	