APPLIED GENOMICS

DNA SEQUENCING

Prof. Alberto Pallavicini
pallavic@units.it

GENOMICS ANALYSIS

- Genomics is highly technology driven. The enormous impact of genomics research on medical and agro-technological sciences has inspired commercial life-science companies to develop innovative genomic tools at a tremendously high speed.

Gene sequencing has proved its usefulness as a diagnostic and prognostic tool.
Its use in the identification of BRCA1 mutations is already a gold standard in cancer research.

whole genome sequencing (WGS) is turning into a common practice faster than one could have originally expected.

Pricing by popular demand

- last year (2017) it was possible to get your genome sequenced for around €900 in a few days.

- The next generation sequencing (NGS) market, including but not limited to WGS, was valued at €4.6Bn in 2015 and is expected to reach €19Bn by 2020.

Pricing by popular demand

- Illumina is one of the biggest players in the sequencing industry and at the start of 2017 announced a new NovaSeq range of sequencers that “one day will enable the $100 genome”.
- While the new range has proved extremely popular with customers, exact predictions on when the $100 genome will happen are less clear.
- Illumina’s more popular sequencers range from €800,000 to almost €1M.

Drowning in data

- Everybody talks about the $1,000 genome, but they don’t talk about the $2,000 mapping problem behind the $1,000 genome.
- In 2014, The National Cancer Institute said that it would pay €18M to move copies of the 2.6 petabyte Cancer Genome Atlas into the cloud.
- Amazon and Google understand this need and already offer to keep a copy of any genome for €25 a year, which translates to roughly €0.02/GB per month, since a file is commonly between 100 and 400GB.

LINK

https://www.genomicsengland.co.uk/taking-part/genomic-medicine-centres/

http://www.phgfoundation.org/news/plans-for-an-australian-100000-genomes-project

https://www.geenivaramu.ee/en/news/estonia-offers-100000-residents-free-genetic-testing-effort-aims-develop-personalized-medicine

http://www.sciencemag.org/news/2016/02/nih-s-1-million-volunteer-precision-medicine-study-announces-first-pilot-projects

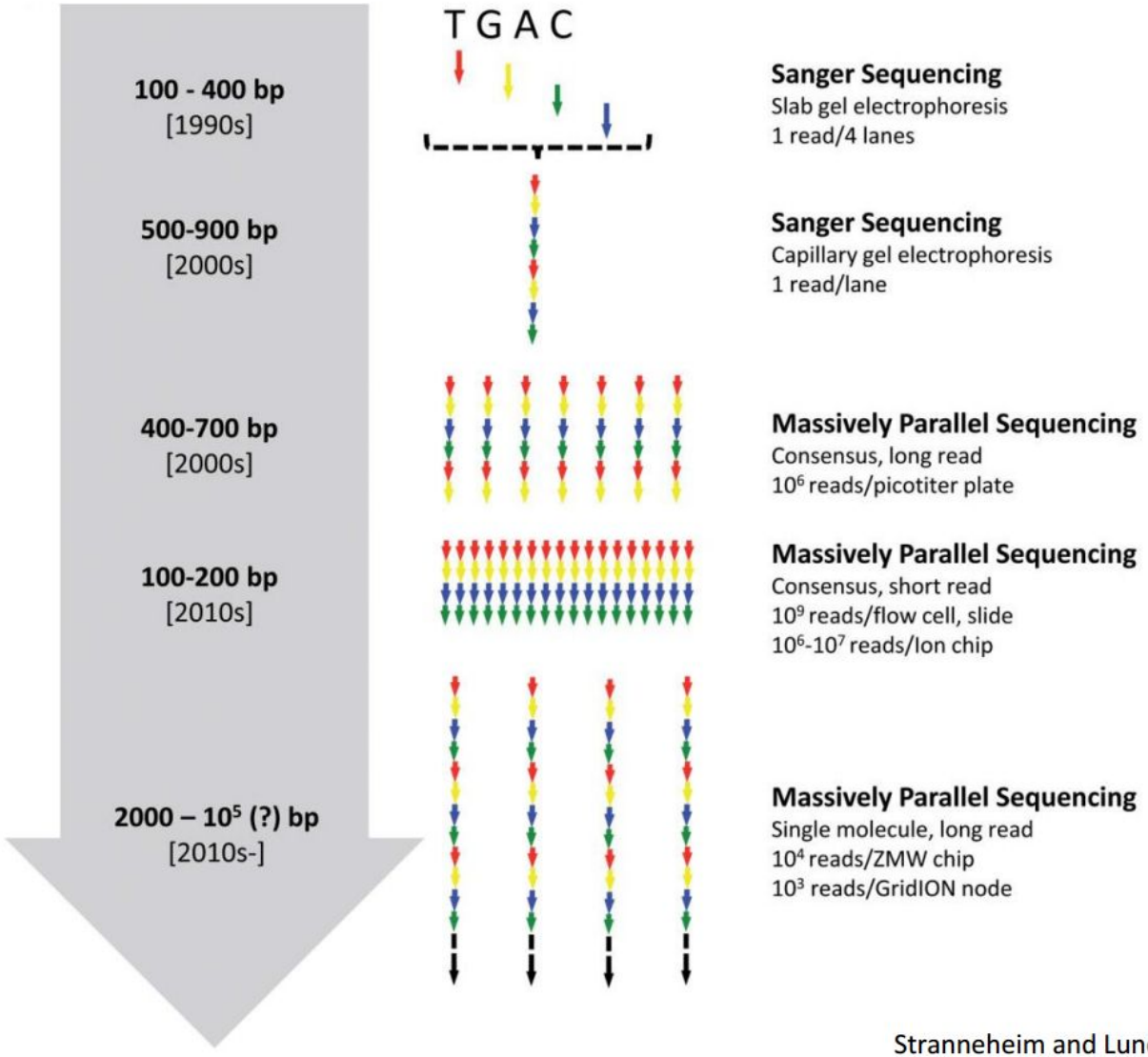
https://www.gouvernement.fr/sites/default/files/document/document/2016/06/22.06.2016_remise_du_rapport_dyves_levy_-_france_medecine_genomique_2025.pdf

http://www.genomeasia100k.com/

https://www.nature.com/articles/hgv201616

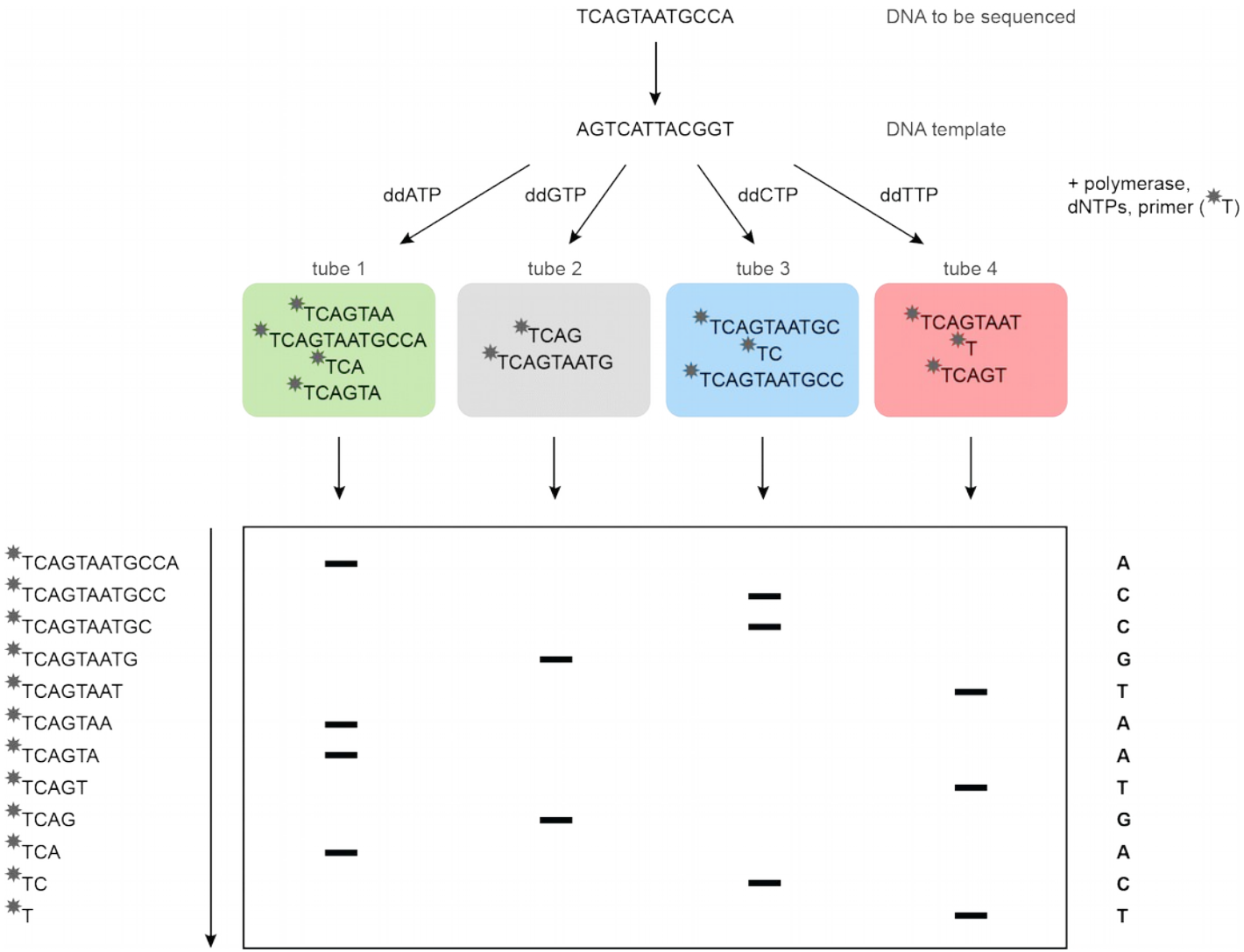
https://twitter.com/GenomicsEngland

GENOMICS ANALYSIS

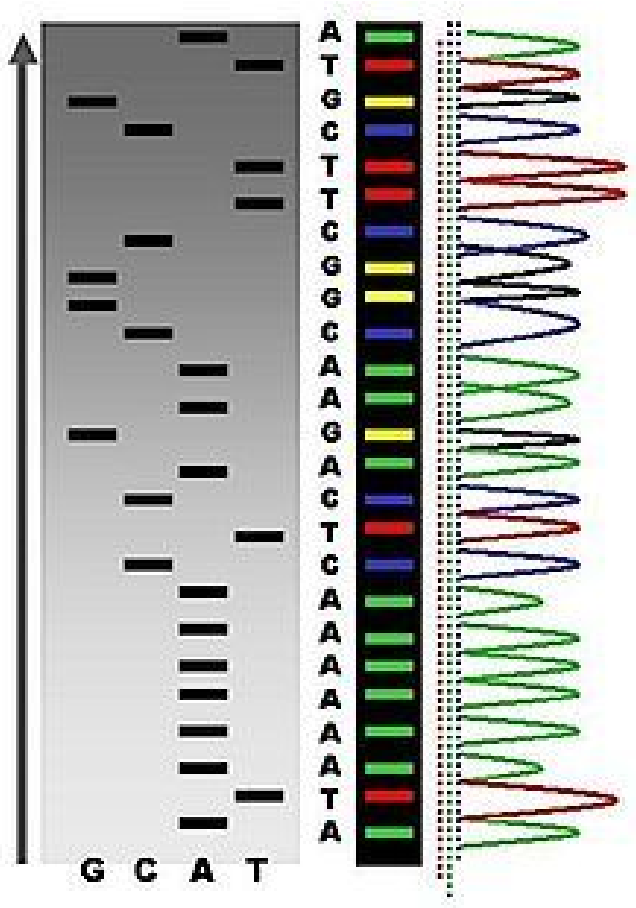


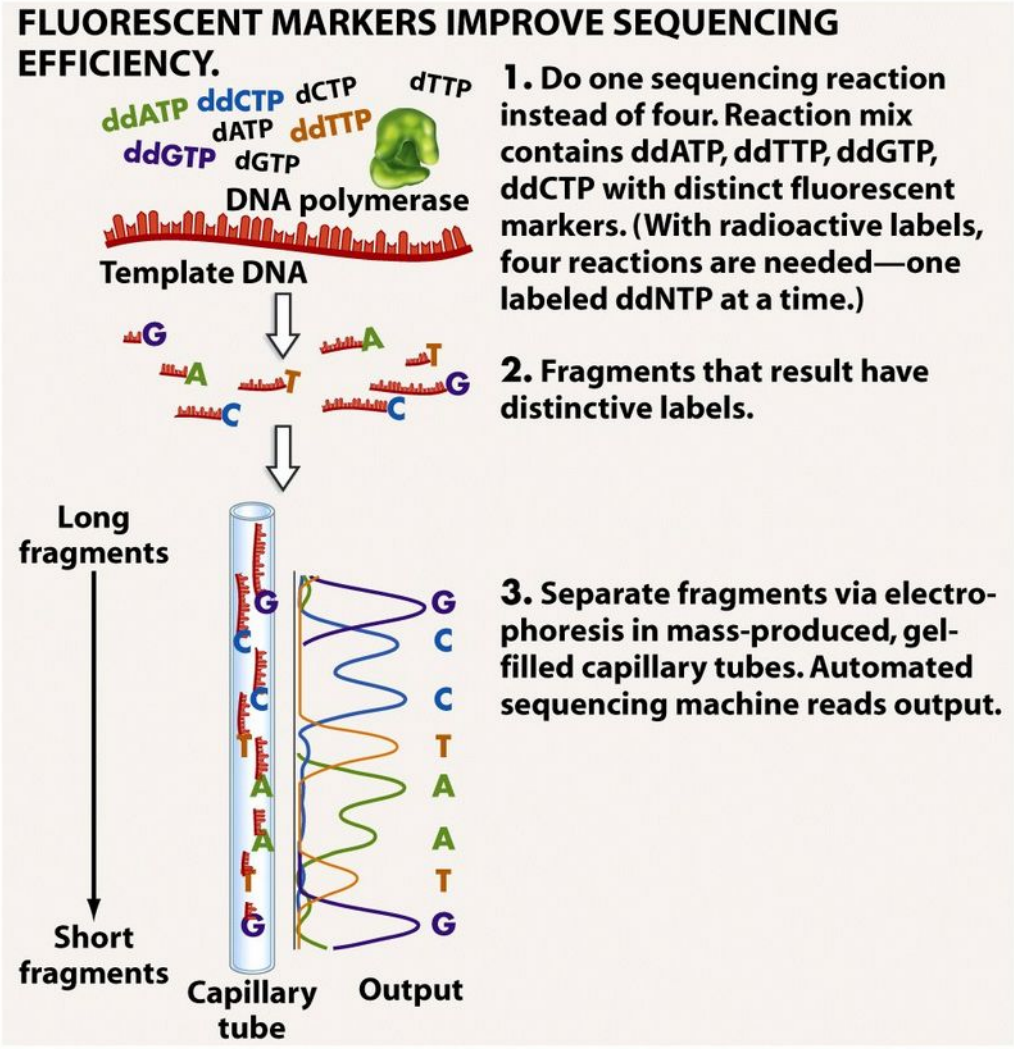
Stranneheim and Lundeberg 2012

…BUT LET'S START FROM THE BEGINNIN: SANGER SEQUENCING



SANGER SEQUENCING





SANGER SEQUENCING

- Advantages
  - Long reads (~900bps)
  - Suitable for small projects

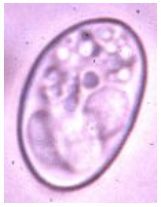
- Disadvantages
  - Low throughput
  - Expensive

SANGER SEQUENCING

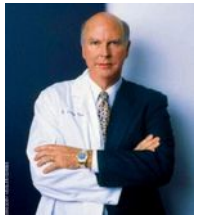
1994: *H. Influenzae*
1.8 Mbp
(Fleischmann et al.

2007: *Global Ocean Sampling Expedition*
~3,000 organisms,
7G r et al.)

1980 1990 2000



1982: *lambda virus*
DNA stretches up to
30-40Kbp
(Sanger et al.)



2001: *H. Sapiens, D. Melanogaster*
3 Gbp
(Venter et al.)

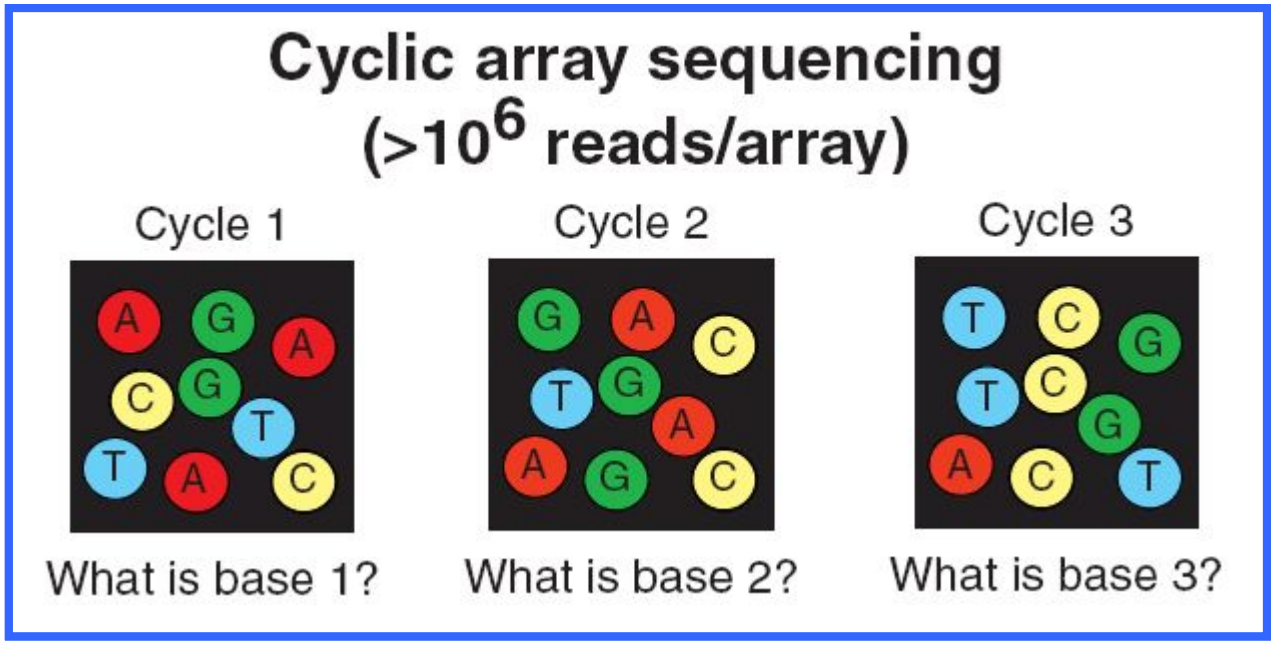
NEXT GENERATION SEQUENCING: WHY NOW?

- **Motivation:** HGP and its derivatives, personalized medicine
- **Short reads applications:** (re-)sequencing, other methods (e.g. gene expression)
- Advancements in technology

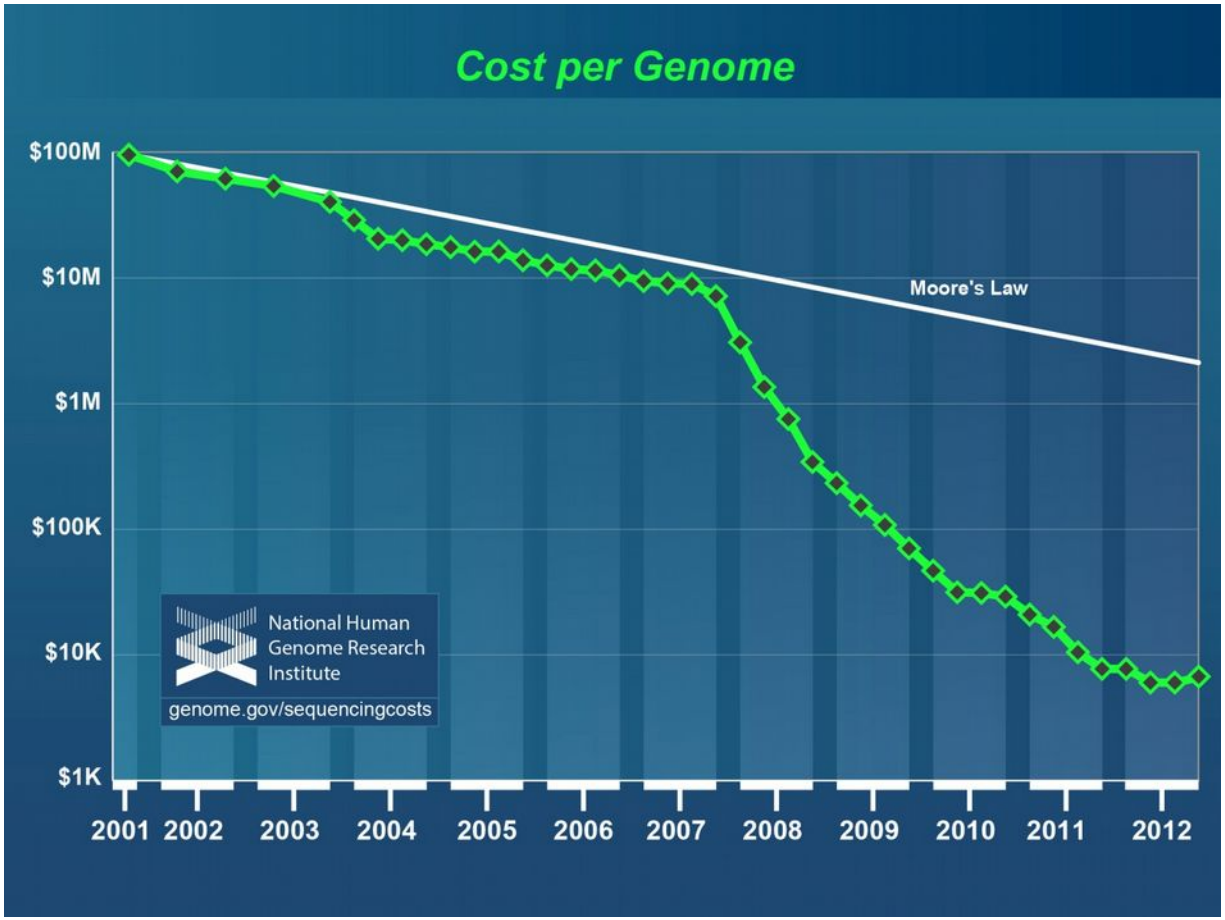
HIGH PARALLELISM IS ACHIEVED IN POLONY SEQUENCING

Sanger

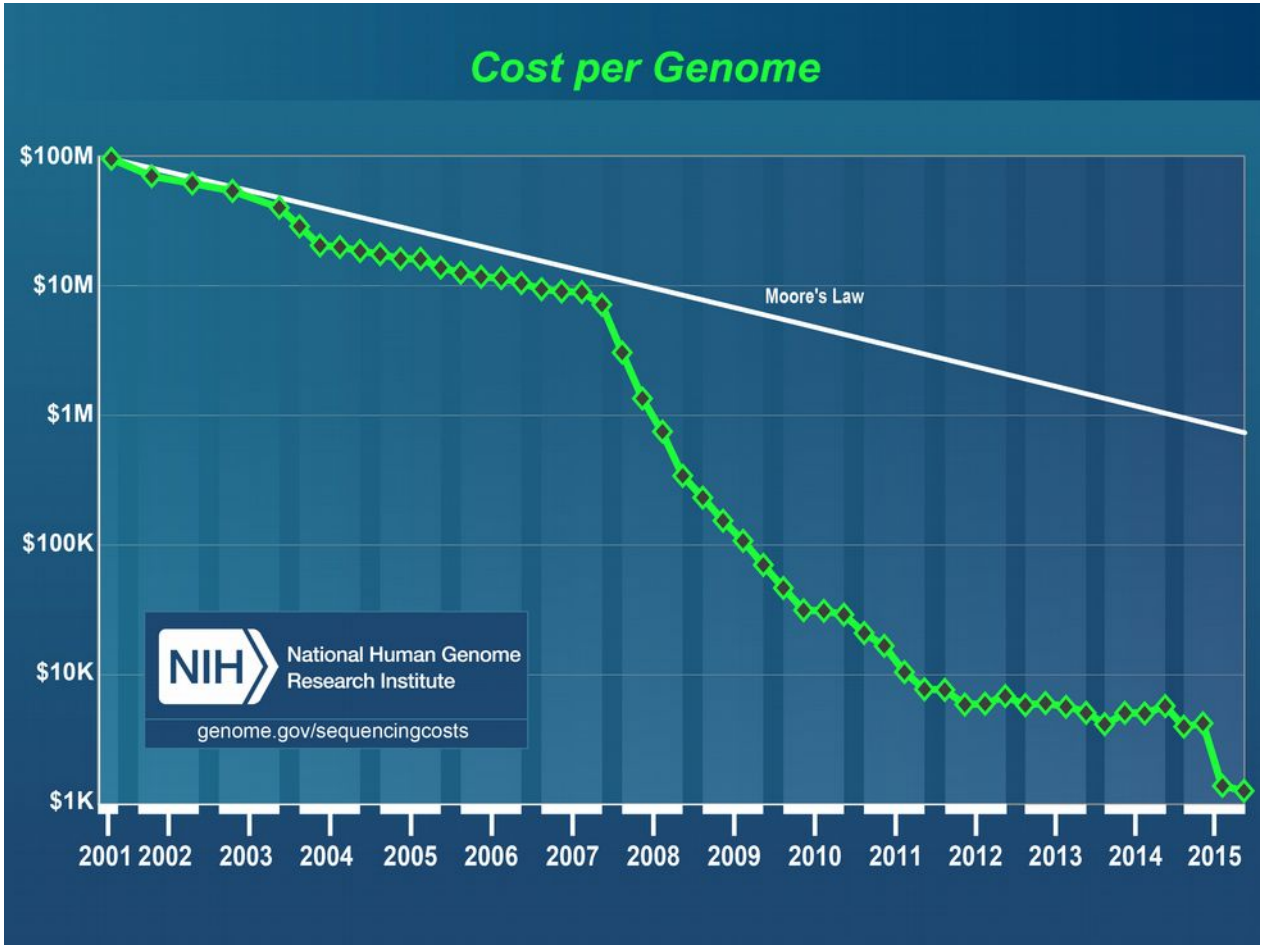
Polony



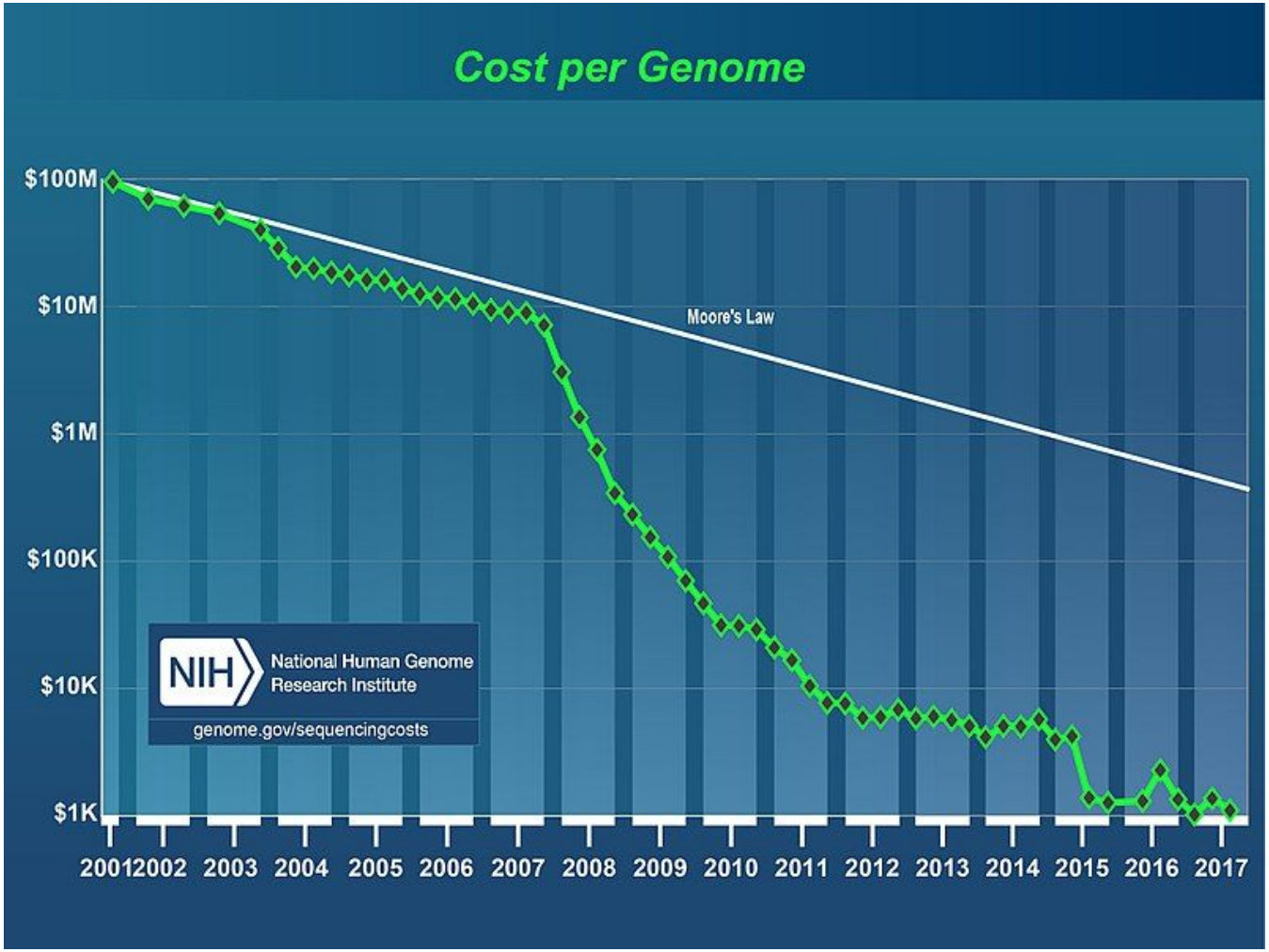
Sequencing costs have fallen



Sequencing costs have fallen



Sequencing costs have fallen



HIGH PARALLELISM IS ACHIEVED IN POLONY SEQUENCING

Perchè Next Generation Sequencing

Si possono generare centinaia di milioni di corte sequenze (35bp-250bp) in una sola corsa in un tempo breve con un basso prezzo per base sequenziata.

2000

- Illumina HiSeq 2500, MiSeq, Next seq 500
- Life Technologies Ion Proton/Ion PGM
- Applied Biosystems SOLiD e Roche/454 FLX, Titanium

Reviews: Michael Metzker *(2010) Nature Reviews Genetics 11:31*
Quail et al (2012) BMC Genomics Jul 24;13:341.

ILLUMINA MACHINES

	 MiniSeq System	 MiSeq Series	 NextSeq Series	 HiSeq Series	 HiSeq X Series*
Key Methods	Amplicon, targeted RNA, small RNA, and targeted gene panel sequencing.	Small genome, amplicon, and targeted gene panel sequencing.	Everyday exome, transcriptome, and targeted resequencing.	Production-scale genome, exome, transcriptome sequencing, and more.	Population- and production-scale whole-genome sequencing.
Maximum Output	7.5 Gb	15 Gb	120 Gb	1500 Gb	1800 Gb
Maximum Reads per Run	25 million	25 million†	400 million	5 billion	6 billion
Maximum Read Length	2 × 150 bp	2 × 300 bp	2 × 150 bp	2 × 150 bp	2 × 150 bp
Run Time	4–24 hours	4–55 hours	12–30 hours	<1–3.5 days (HiSeq 3000/HiSeq 4000) 7 hours–6 days (HiSeq 2500)	<3 days
Benchtop Sequencer	Yes	Yes	Yes	No	No
System Versions	• MiniSeq System for low-throughput targeted DNA and RNA sequencing	• MiSeq System for targeted and small genome sequencing • MiSeq FGx System for forensic genomics • MiSeqDx System for molecular diagnostics	• NextSeq 500 System for everyday genomics • NextSeq 550 System for both sequencing and cytogenomic arrays	• HiSeq 3000/HiSeq 4000 Systems for production-scale genomics • HiSeq 2500 Systems for large-scale genomics	• HiSeq X Five System for production-scale whole-genome sequencing • HiSeq X Ten System for population-scale whole-genome sequencing

http://www.illumina.com/systems/sequencing.html

ION TORRENT MACHINES

Ion S5 Systems

Ion PGM System

https://www.thermofisher.com/it/en/home/brands/ion-torrent.html

Next generation sequencing

	Run Time	Read Length	Quality	Total nucleotides sequenced	Cost /MB
454 Pyrosequencing	24h	700 bp	Q20-Q30	1 GB	$10
Illumina Miseq	27h	2x300bp	> Q30	15 GB	$0.15
Illumina Hiseq 2500	1 - 10days	2x250bp	>Q30	3000 GB	$0.05
Ion torrent	2h	400bp	>Q20	50MB-1GB	$1
Pacific Biosciences	30m - 4h	10kb - >40kb	>Q50 consensus >Q10 single	500 - 1000MB /SMRT cell	$0.13 - $0.60

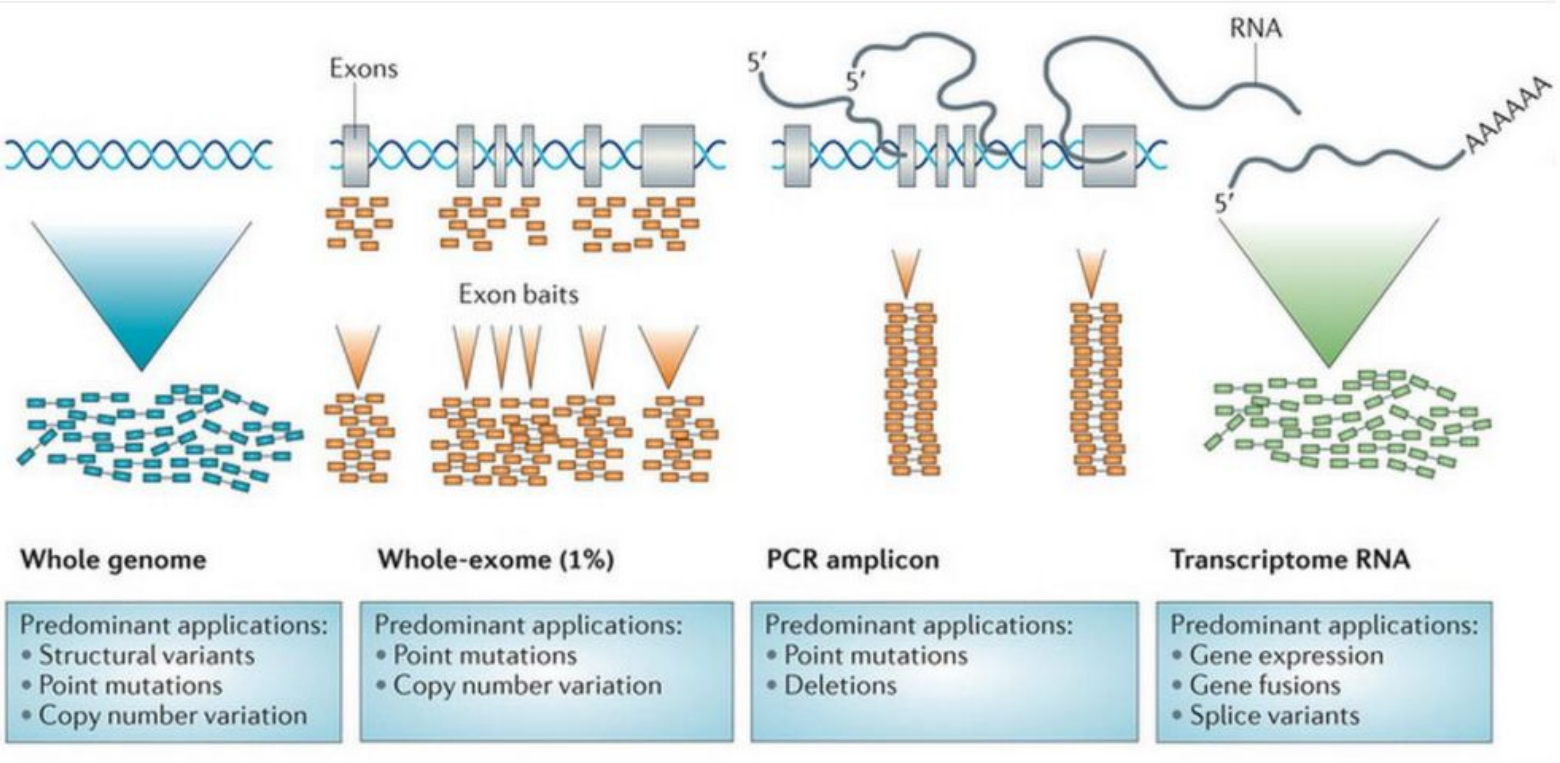
http://www.hindawi.com/journals/bmri/2012/251364/
http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3431227

circa 2015

Terminology

- **Coverage (depth):** the number of times a nucleotide is read during the sequencing process
- **Quality Score:** Each called base comes with a quality score which measures the probability of base call error.
- **Paired-End Sequencing:** Both end of the DNA fragment is sequenced, allowing highly precise alignment.
- **Multiplex Sequencing:** "barcode" sequences are added to each sample so they can be distinguished in order to sequence large number of samples on one lane.
- **Mapping:** Align reads to reference to identify their origin.
- **Assembly:** Merging of fragments of DNA in order to reconstruct the original sequence.
- **Duplicate reads:** Reads that are identical.
- **Multi-reads:** Reads that can be mapped to multiple locations equally well.

Applications: genomes, exomes, transcriptomes



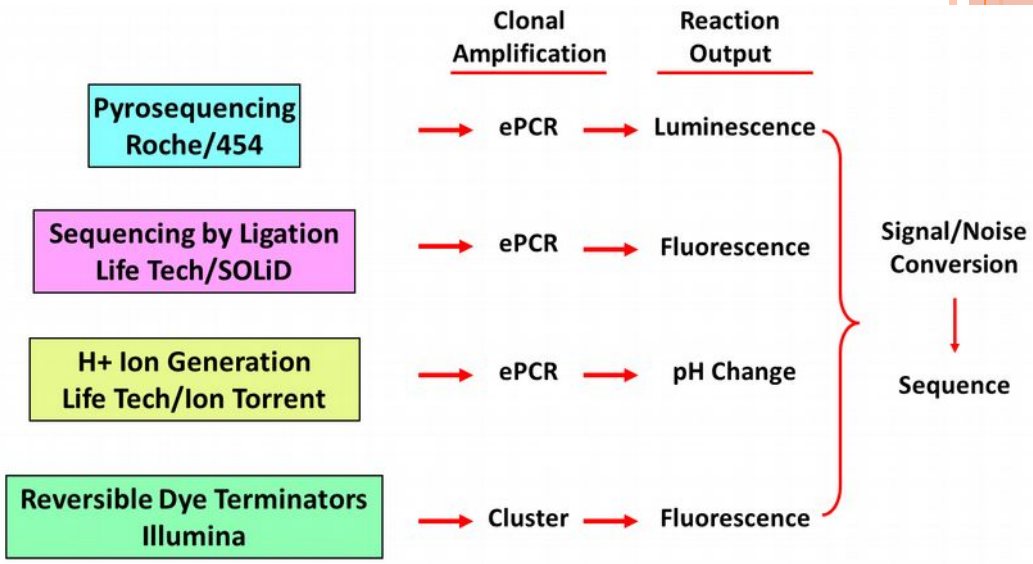
Implementing personalized cancer genomics in clinical trials

Richard Simon & Sameek Roychowdhury

Nature Reviews Drug Discovery **12**, 358–369 (2013) | doi:10.1038/nrd3979

NGS PLATFORMS OVERVIEW

- Differ in design and chemistries
- Fundamentally related-sequencing of thousands to millions of clonally amplified molecules in a massively parallel manner
- Orders of magnitude more information-will continue to evolve



Pacific Biosciences
Helicos Biosciences
NABsys
VisiGen Biotechnologies
Complete Genomics
Oxford Nanophore Technologies

Si basano sul principio del sequenziamento di '*cluster*' clonali

Il processo, che incomincia con una singola molecola target, prevede la creazione di targets clonali durante un processo intermedio di amplificazione. Copie multiple identiche sono infatti necessarie per avere un alto rapporto segnale-rumore

SEQUENZIAMENTO SANGER AD ALTA PROCESSIVITÀ

PREPARAZIONE DELLA LIBRERIA

Frammentazione casuale del DNA genomico
clonazione e trasformazione in batteri

7-10 giorni
assumendo di possedere una piattaforma robotica per alta processività

↓

Raccolta delle colonie

↓

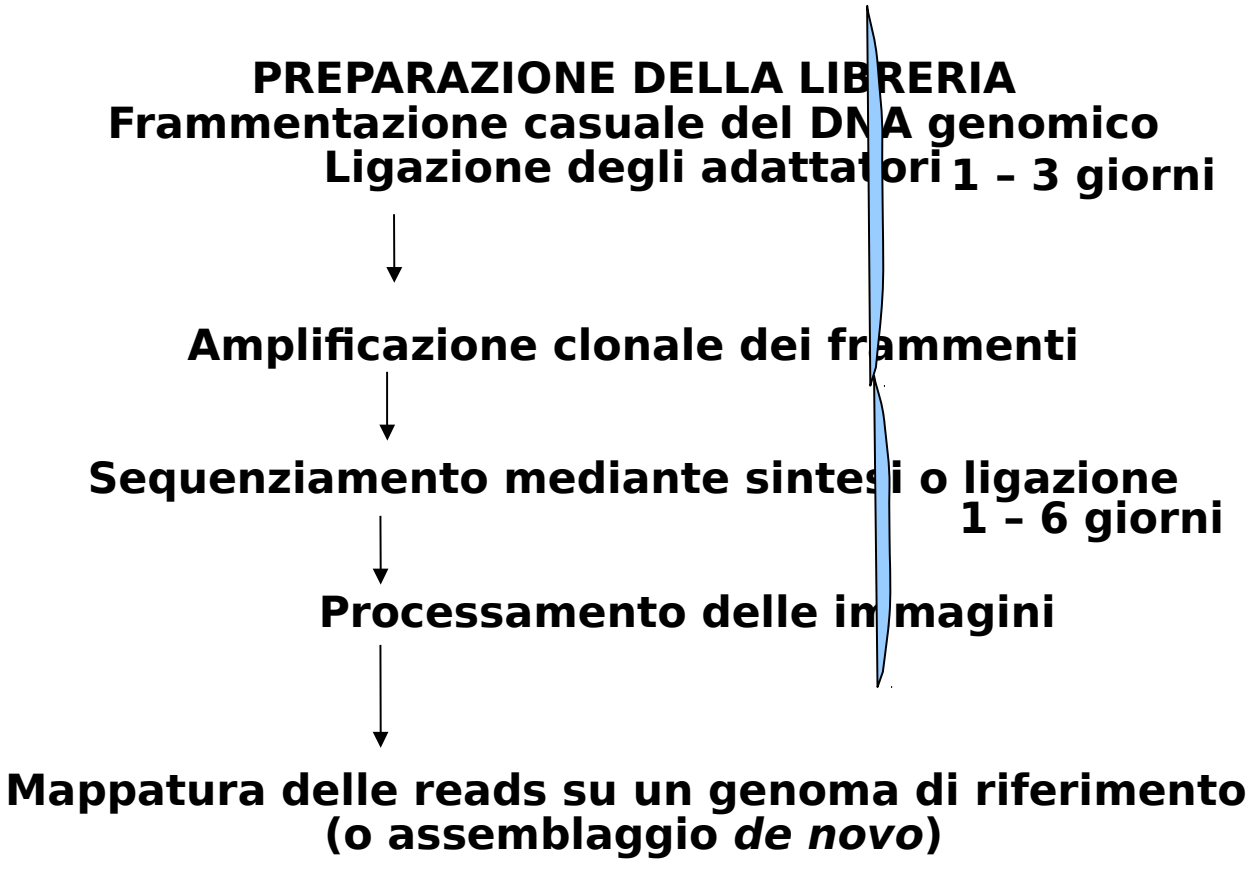
Purificazione del DNA dalle colonie
Sequenziamento Sanger
Elettroforesi capillare

Settimane-anni (!),
dipendentemente dalla dimensione del genoma (e copertura richiesta), dal numero di sequenziatori capillari

↓

Whole genome *de novo* assembly or mapping to a reference (re-sequencing)

SEQUENZIAMENTO DI NUOVA GENERAZIONE



Vantaggi delle piattaforme di nuova generazione

- Non sub-clonazione, non utilizzo di cellule batteriche *E. coli*
  - abolizione di *bias* di clonazione
  - rapidità nel preparare le librerie (non c'e' colony picking!)
- Ciascuna sequenza proviene da una molecola di DNA unica.
  - quantificazione attraverso 'conta' digitale
  - aumento del range dinamico
  - rilevazione di varianti rare
- Fornisce una eccezionale risoluzione per molti tipi di esperimenti (es. analisi di espressione, sequenziamento di DNA immunoprecipitato, di RNA piccoli, analisi di medie/grandi inserzioni-delezioni nei genomi….)
- Rivoluzionaria diminuzione del costo e del tempo per generare dati di sequenza (lavorano in multi-parallelo)
- Richiesta meno robotica nelle fasi precedenti al caricamento sul sequenziatore

Svantaggi delle piattaforme next-gen

- Sono prodotte sequenze più corte

  - relativamente alle sequenze da sequenziatori capillari (metodo Sanger)

  - è necessario ri-parametrizzare l'accuratezza della procedura di chiamata delle basi

  - enorme difficoltà nell'analisi dei dati; richiesto un grande sforzo di programmazione per costruire nuovi algoritmi.

- La mole enorme di dati 'traumatizza' le infrastrutture informatiche.

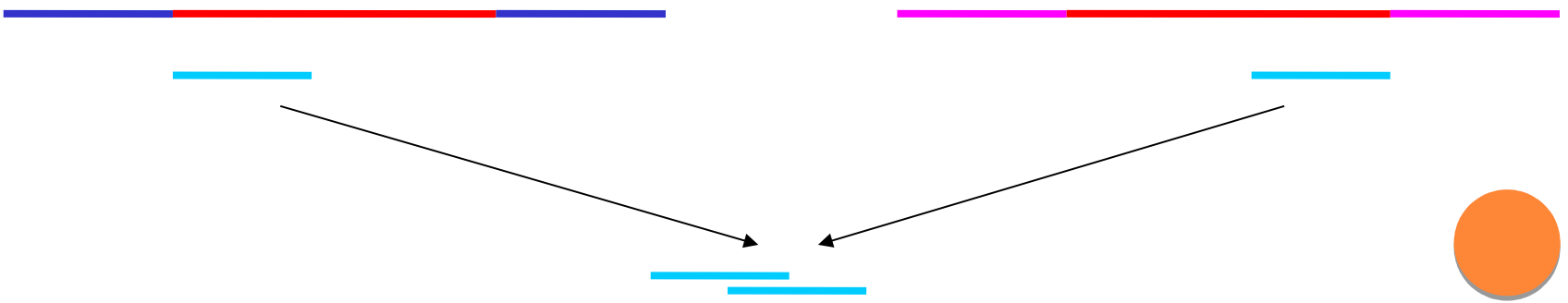
  - da 50-100 Gb a diversi centinaia di Tb di dati grezzi prodotti per corsa (dipende dalla piattaforma)

  - il processamento delle *read* tramite *pipeline* informatiche richiede molta capacità di calcolo (CPU)

  - è necessario prendere accurate decisioni su cosa salvare e cosa cancellare

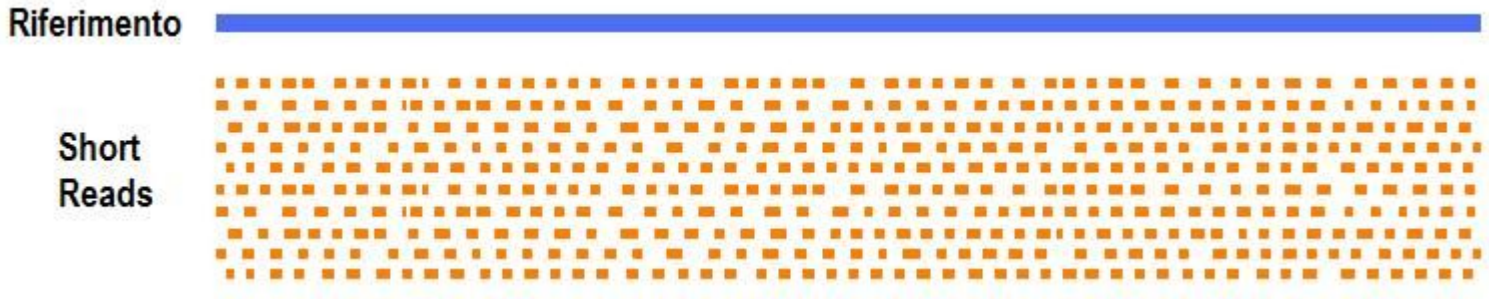
SEQUENZE CORTE

- Difficoltà di assemblare sequenze corte *de novo*, soprattutto per il problema delle sequenze ripetute complicato ancora di più rispetto a Sanger (lunghezza media 700-900bp)

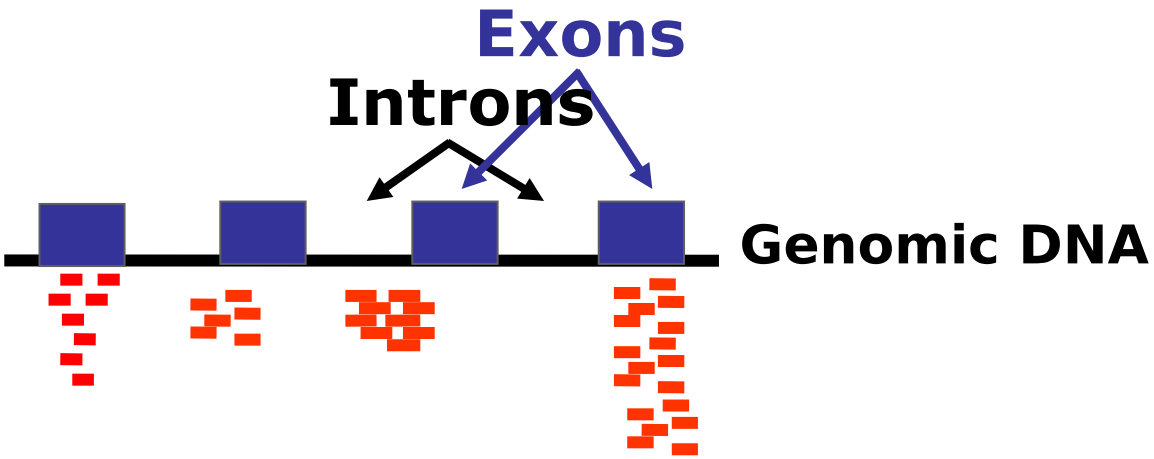


RISEQUENZIAMENTO

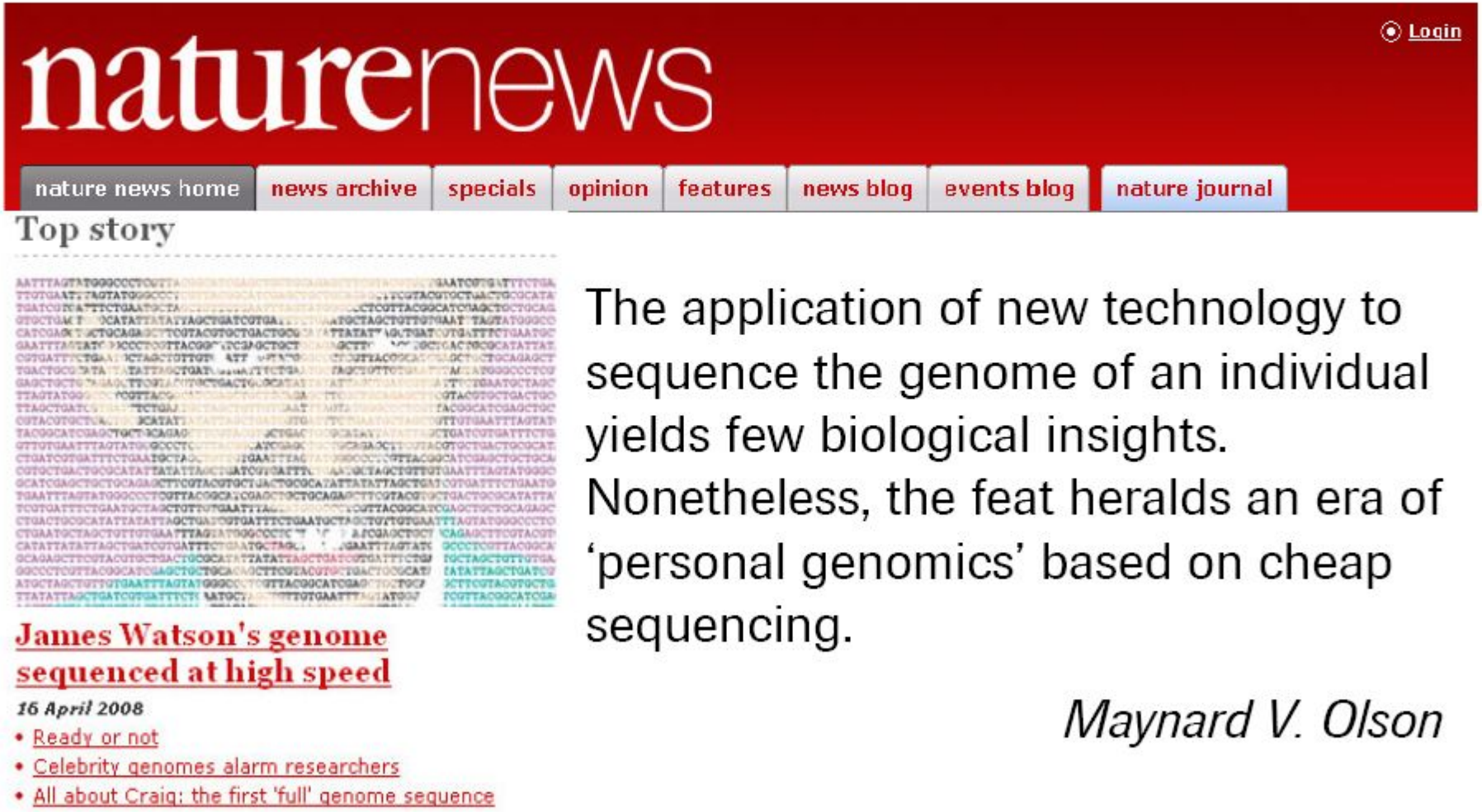
- In presenza di un genoma di riferimento di buona qualità posso effettuare un ri-sequenziamento e allineare tutte le reads ottenute:



- Non solo del genoma, ma anche del trascrittoma



The First of The Rest of Us

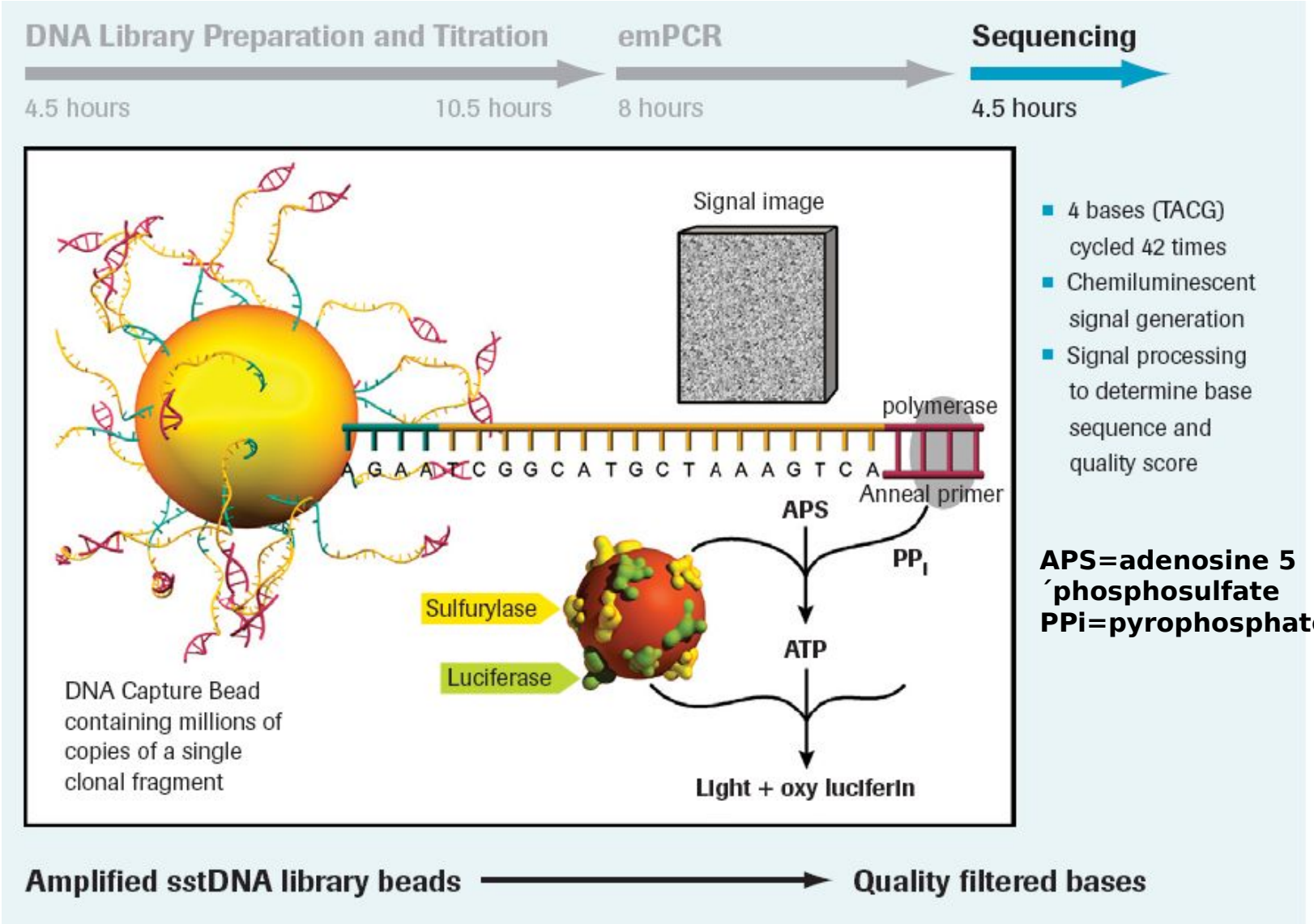


The application of new technology to sequence the genome of an individual yields few biological insights. Nonetheless, the feat heralds an era of 'personal genomics' based on cheap sequencing.

Maynard V. Olson

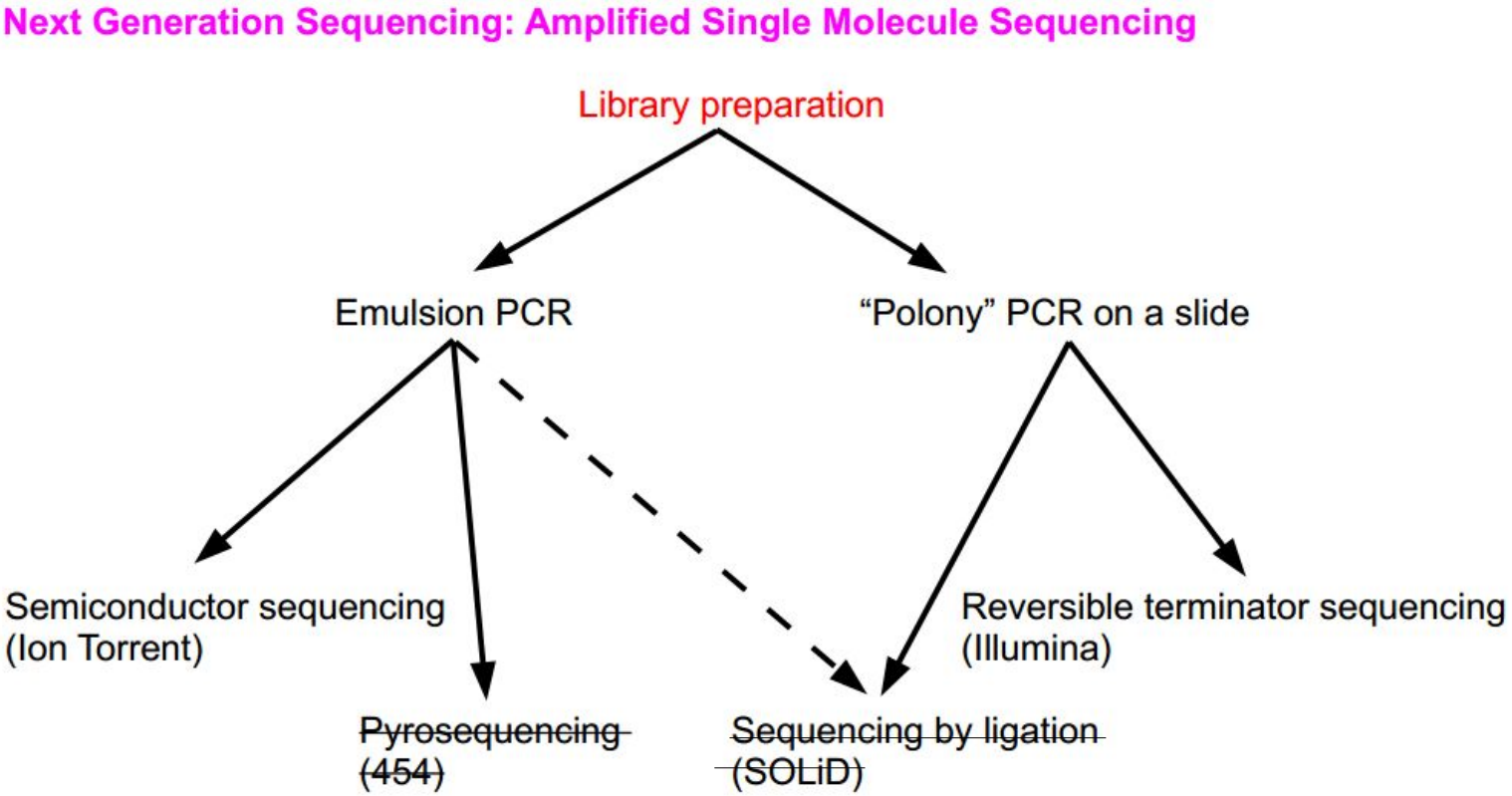
SEQUENZIAMENTO CON LA TECNOLOGIA 454

Tecnologia 454

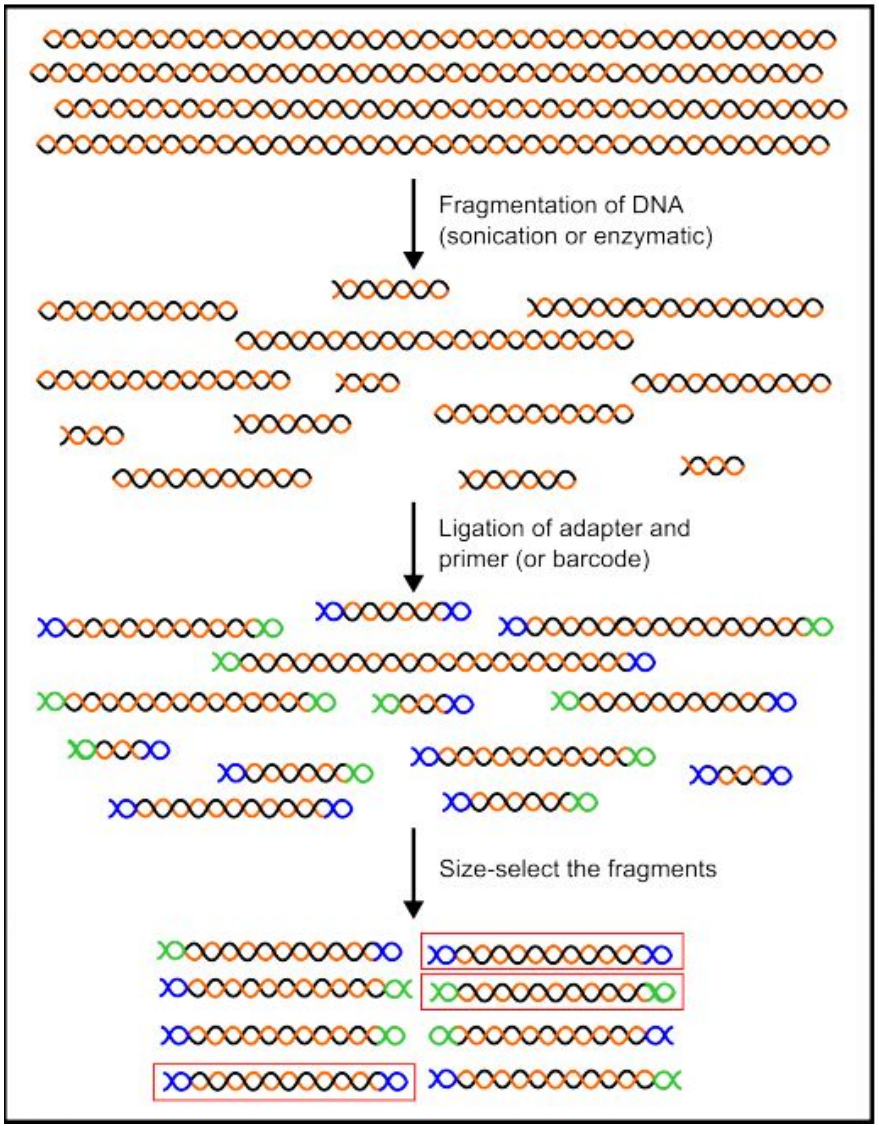


APS=adenosine 5´phosphosulfate
PPi=pyrophosphate

WORKFLOW

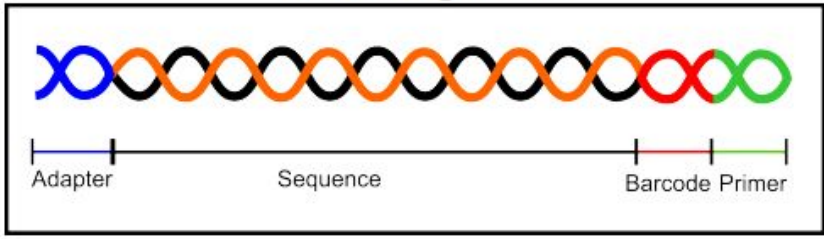

Next Generation Sequencing: Amplified Single Molecule Sequencing
Library preparation
Emulsion PCR
"Polony" PCR on a slide
Semiconductor sequencing
(Ion Torrent)
Pyrosequencing
(454)
Sequencing by ligation
(SOLiD)
Reversible terminator sequencing
(Illumina)

Next Generation Sequencing: Amplified Single Molecule Sequencing

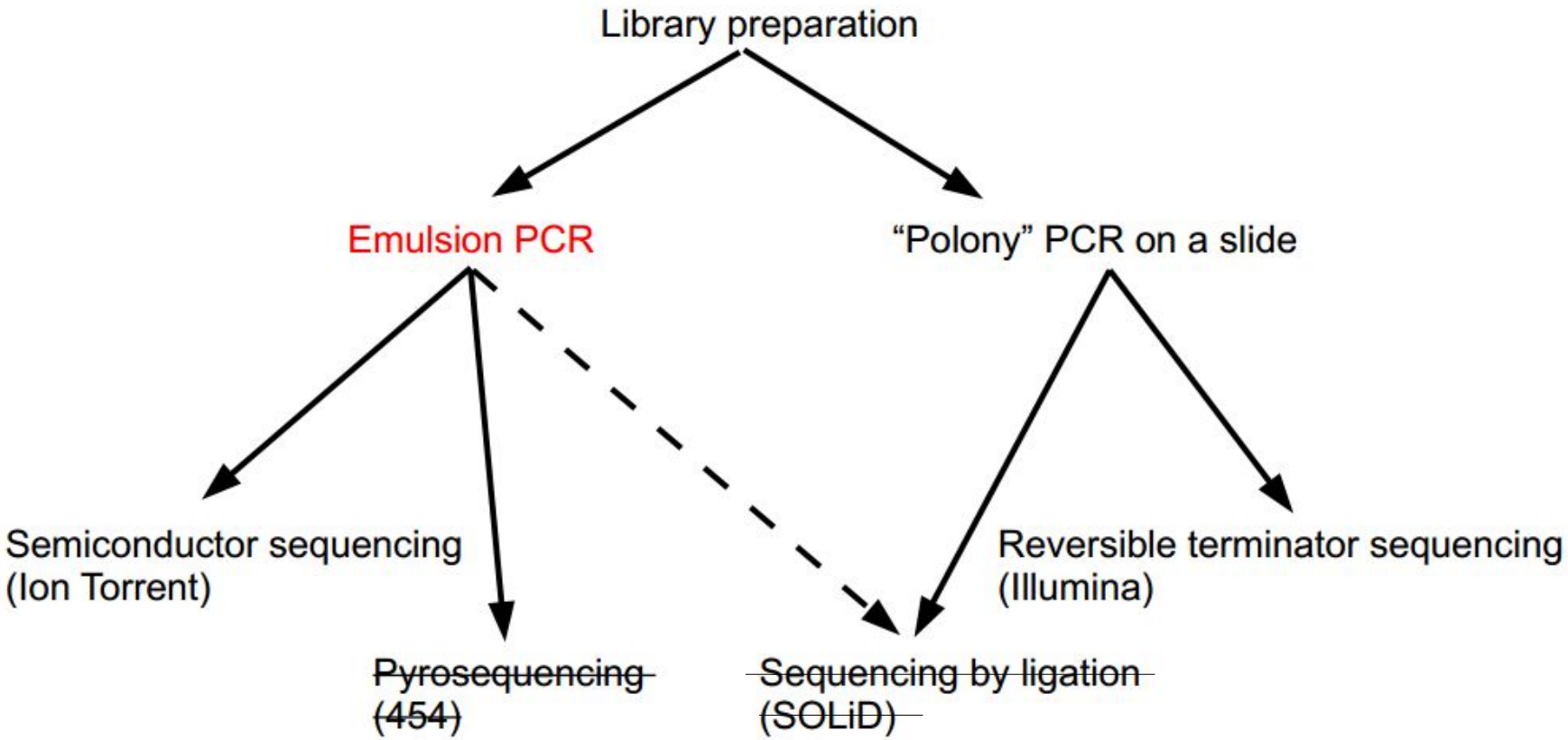


Library preparation

Good fragments:

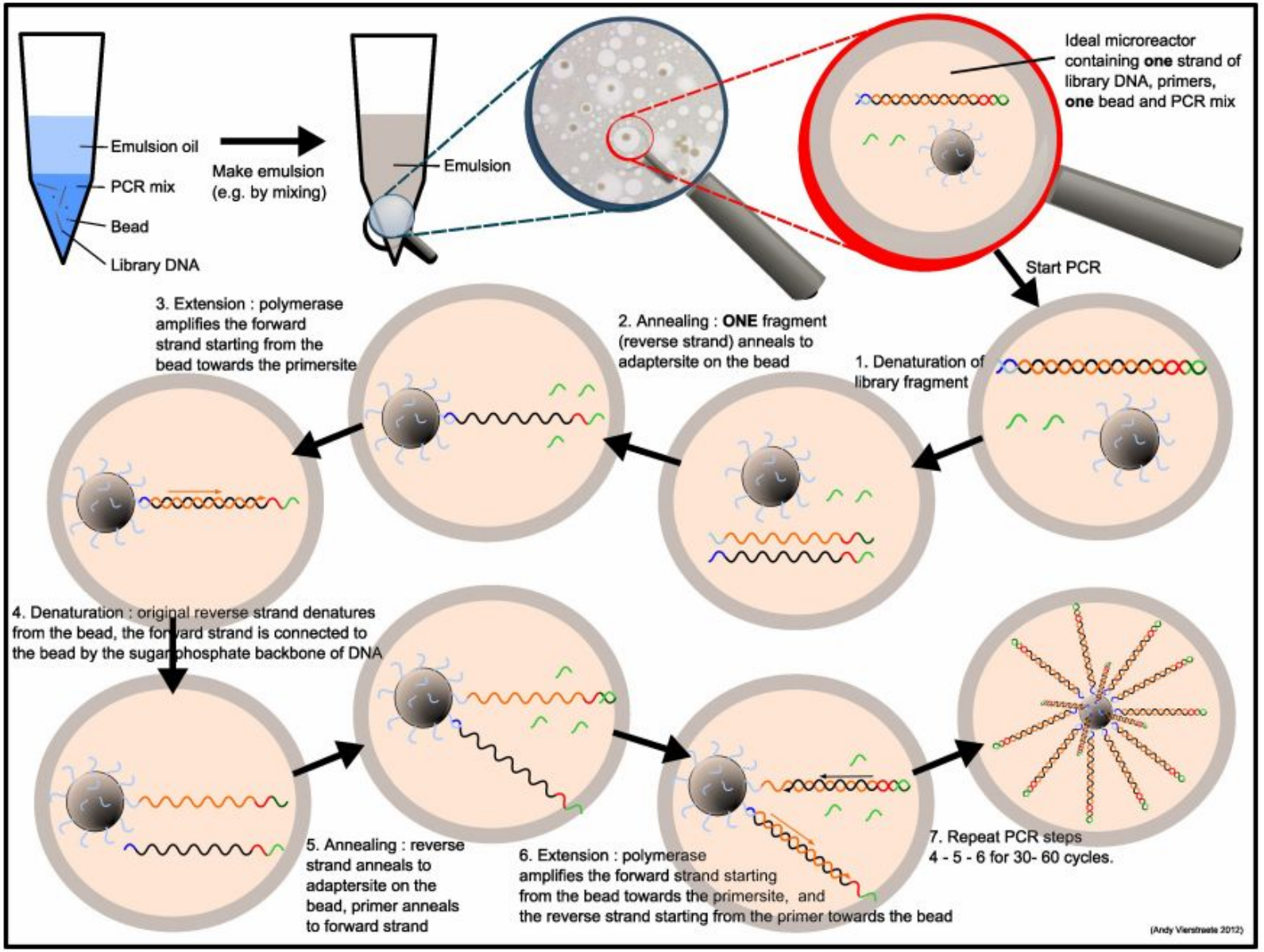


Next Generation Sequencing: Amplified Single Molecule Sequencing



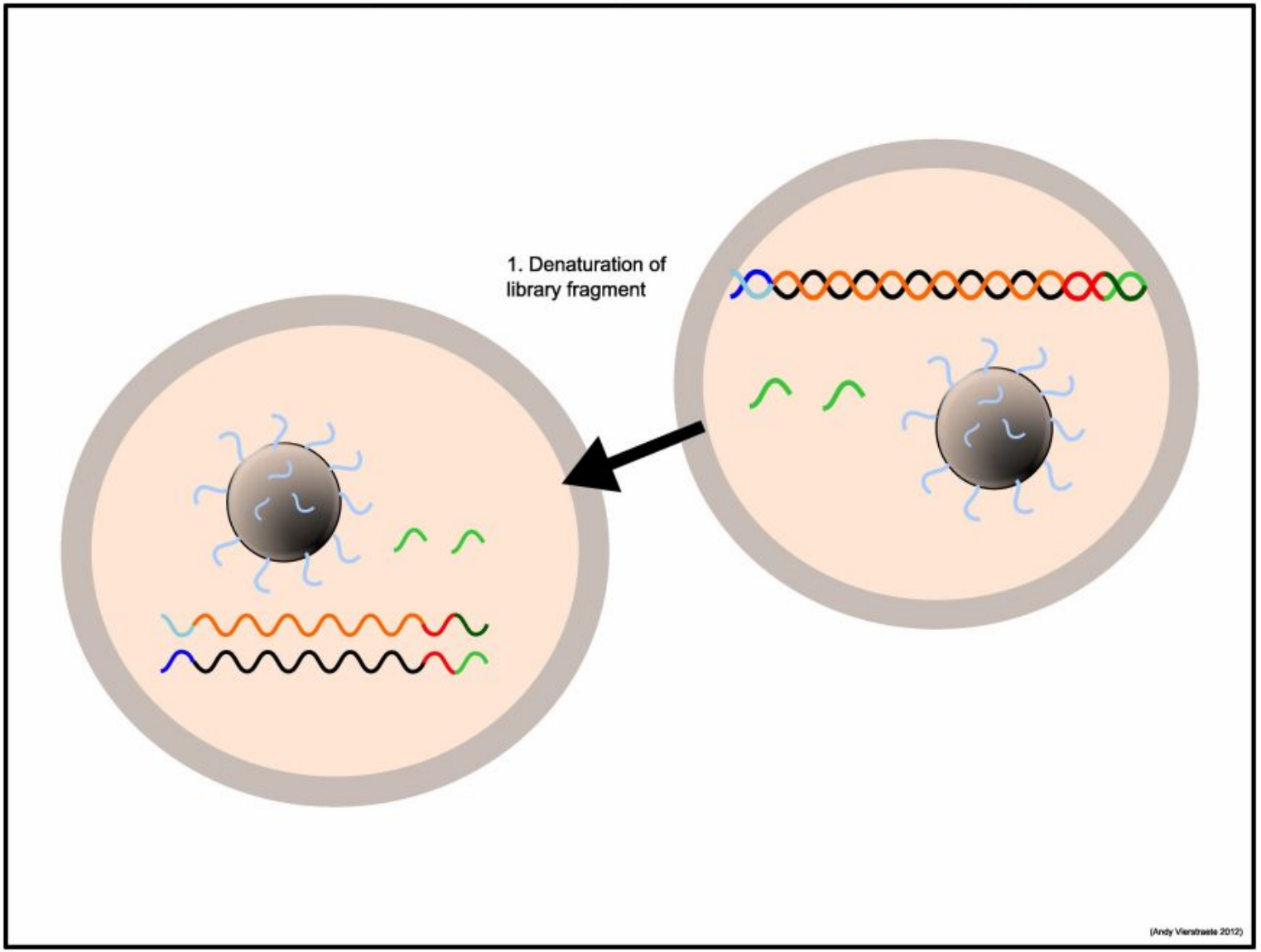
Next Generation Sequencing: Amplified Single Molecule Sequencing

Emulsion PCR



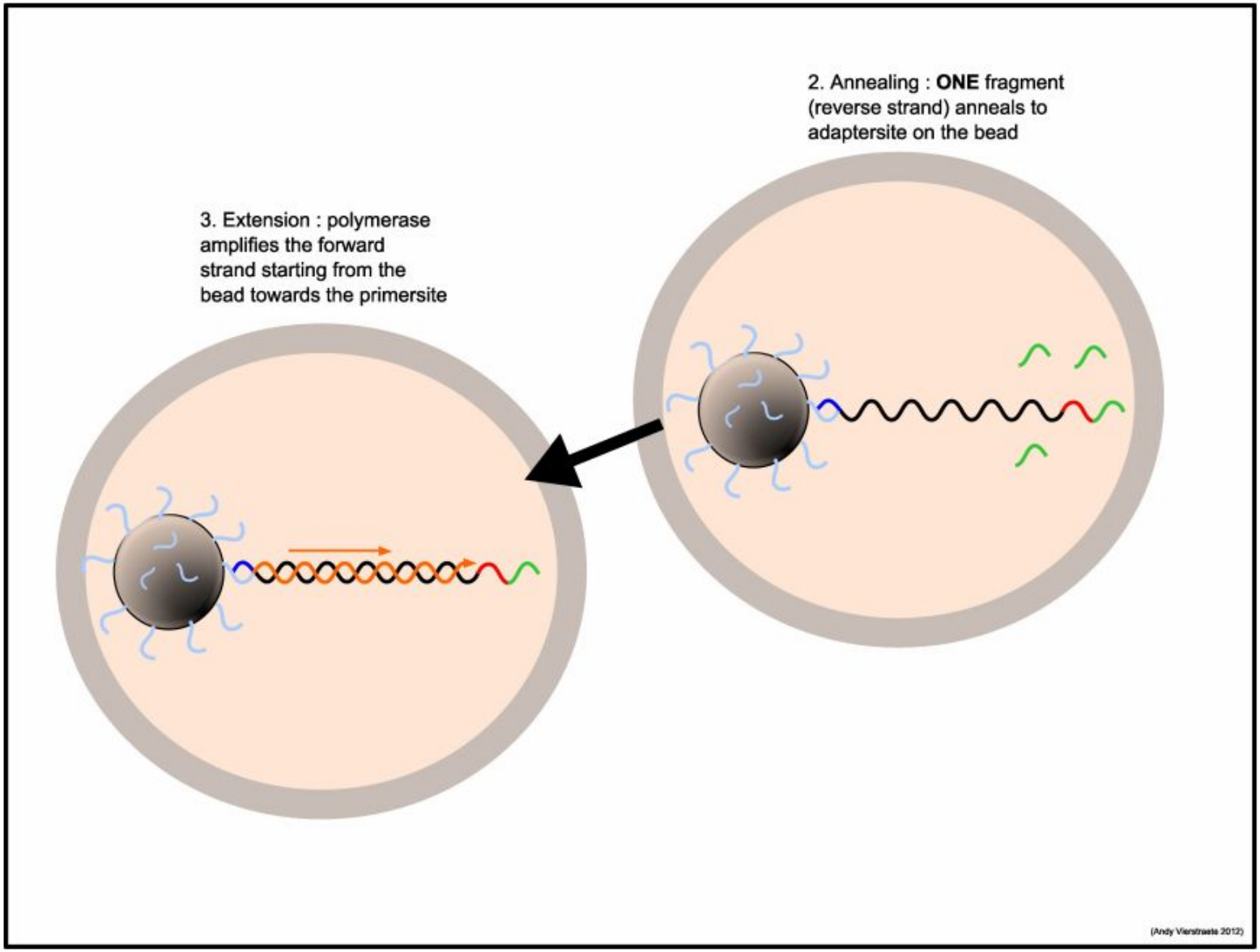
Next Generation Sequencing: Amplified Single Molecule Sequencing

Emulsion PCR



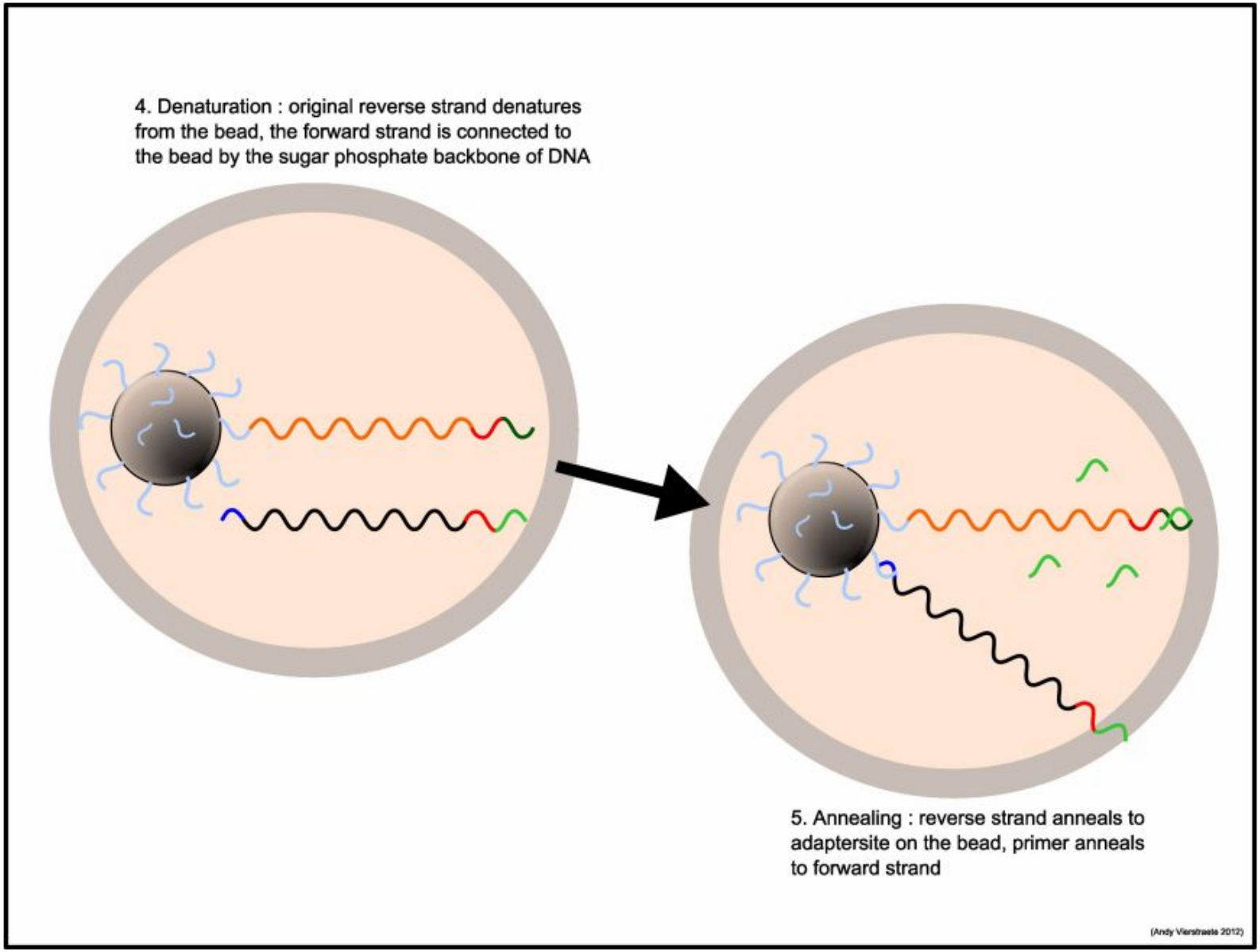
Next Generation Sequencing: Amplified Single Molecule Sequencing

Emulsion PCR



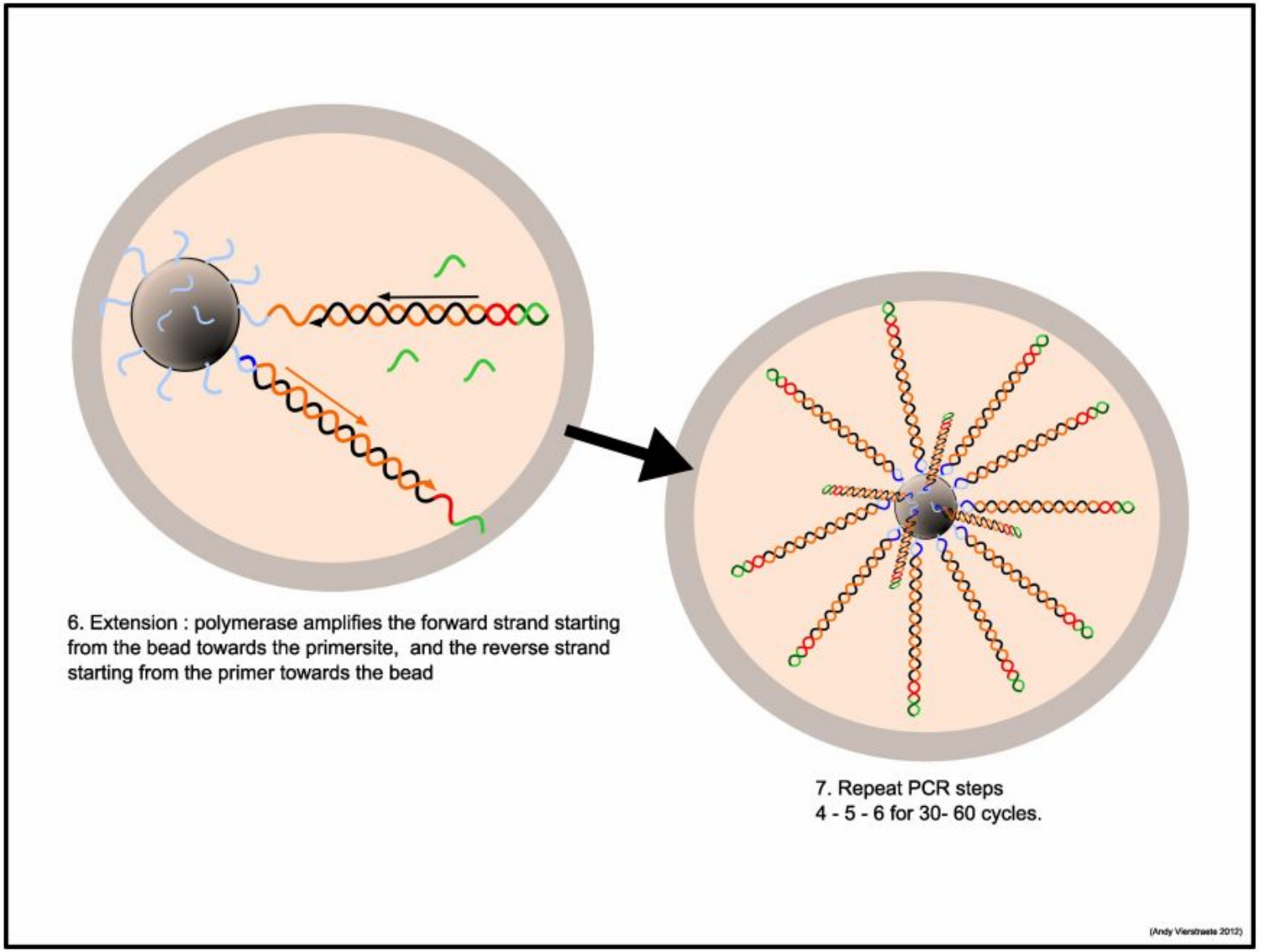
Next Generation Sequencing: Amplified Single Molecule Sequencing

Emulsion PCR



Next Generation Sequencing: Amplified Single Molecule Sequencing

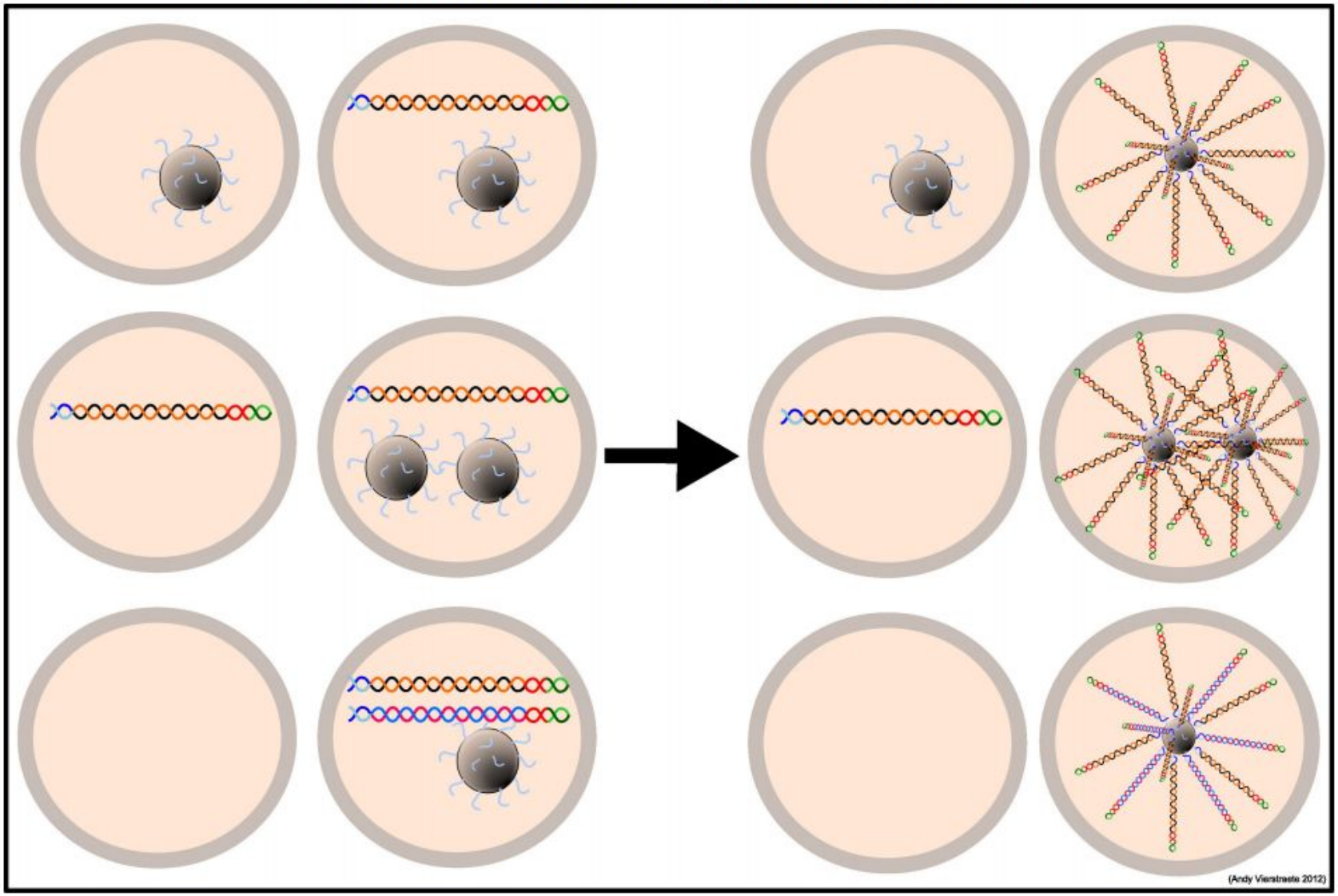
Emulsion PCR



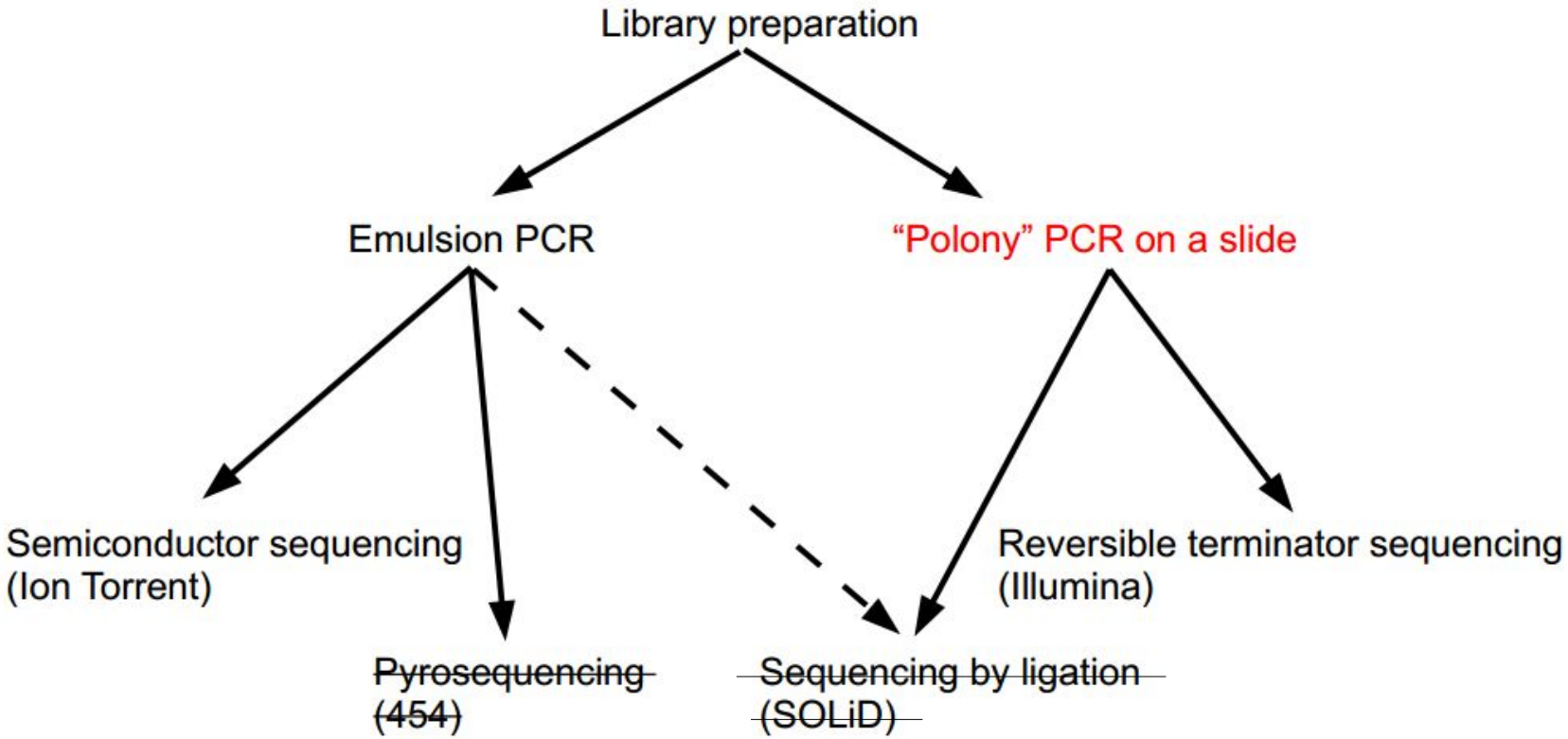
Next Generation Sequencing: Amplified Single Molecule Sequencing

Emulsion PCR

different micro reactors: only 15 % are good ones

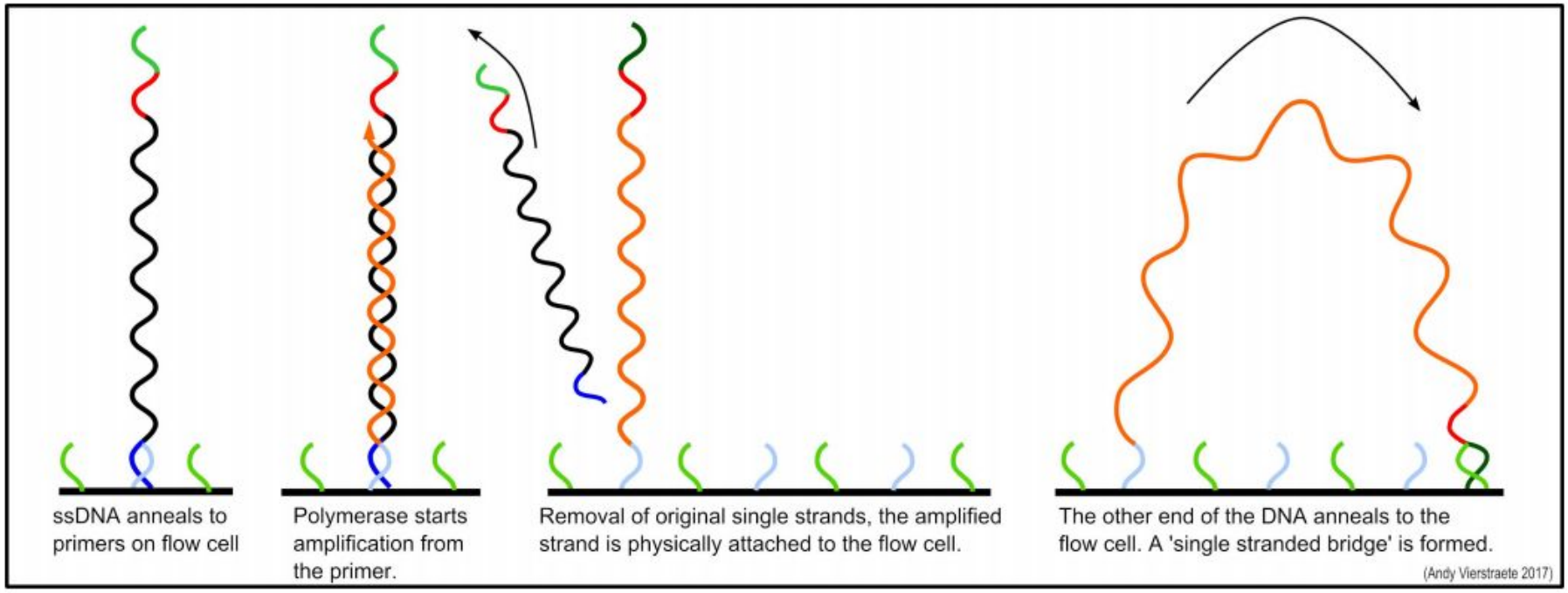


Next Generation Sequencing: Amplified Single Molecule Sequencing



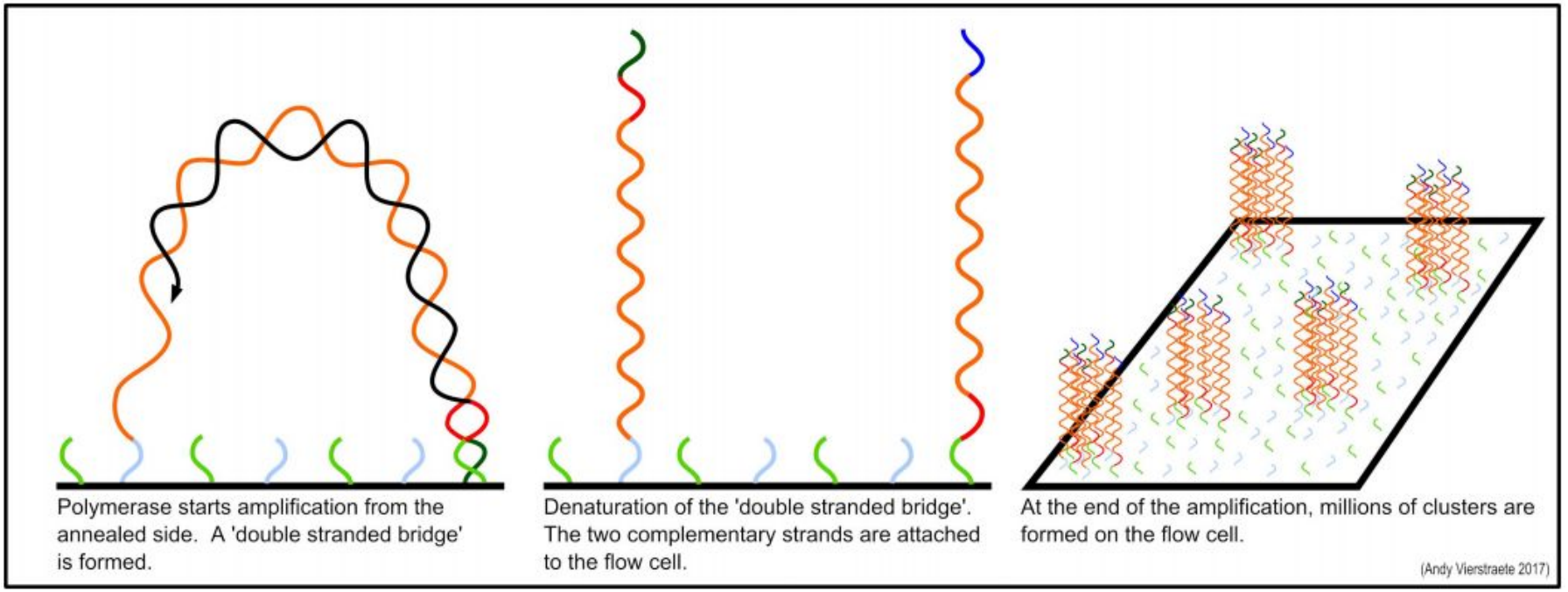
Next Generation Sequencing: Amplified Single Molecule Sequencing "Polony" PCR

Bridge amplification: Illumina



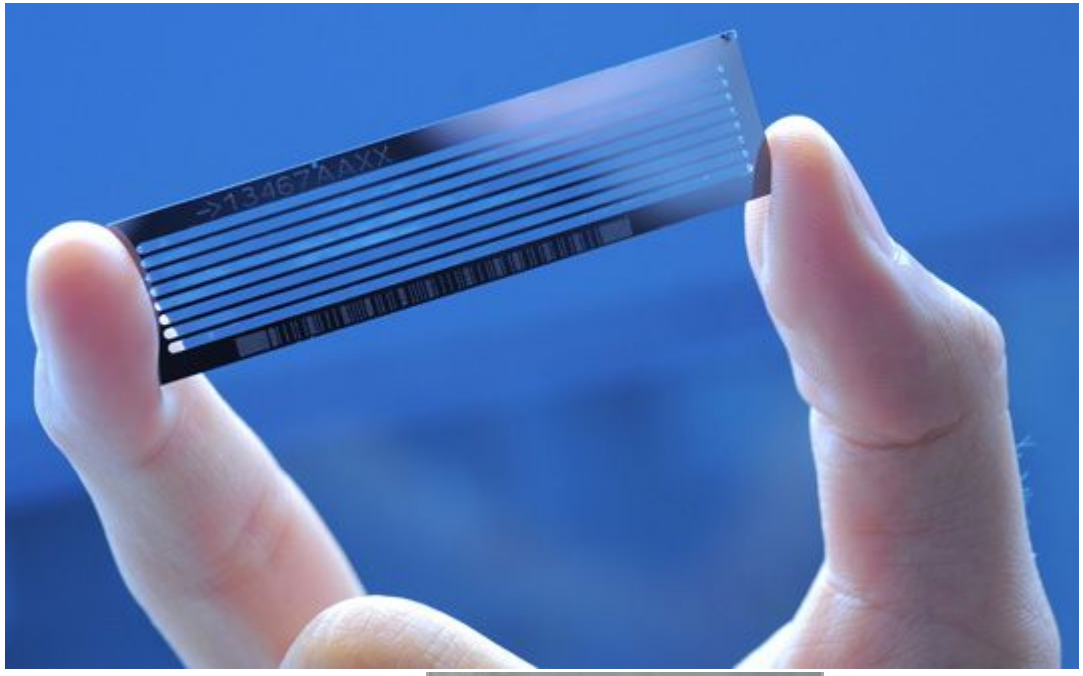
Next Generation Sequencing: Amplified Single Molecule Sequencing "Polony" PCR

Bridge amplification: Illumina



Cluster Station

Strumento che permette di preparare la *flow-cell* (=supporto di vetro su cui i frammenti della libreria verranno sequenziati in parallelo)



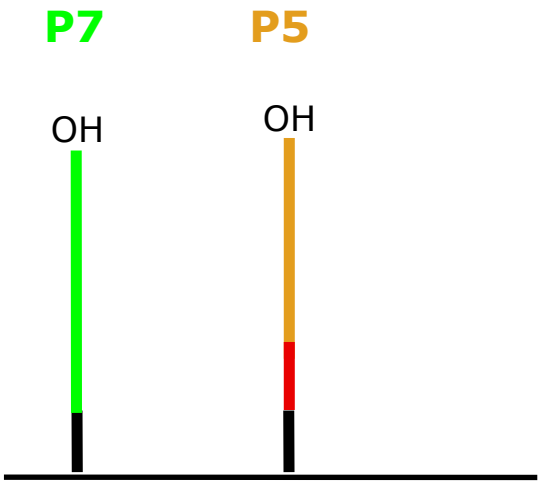

illumina


illumina®

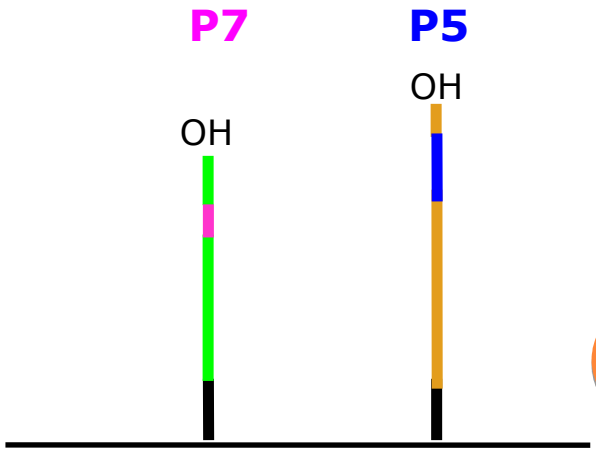
Cluster Generation

Grafted Flow Cells

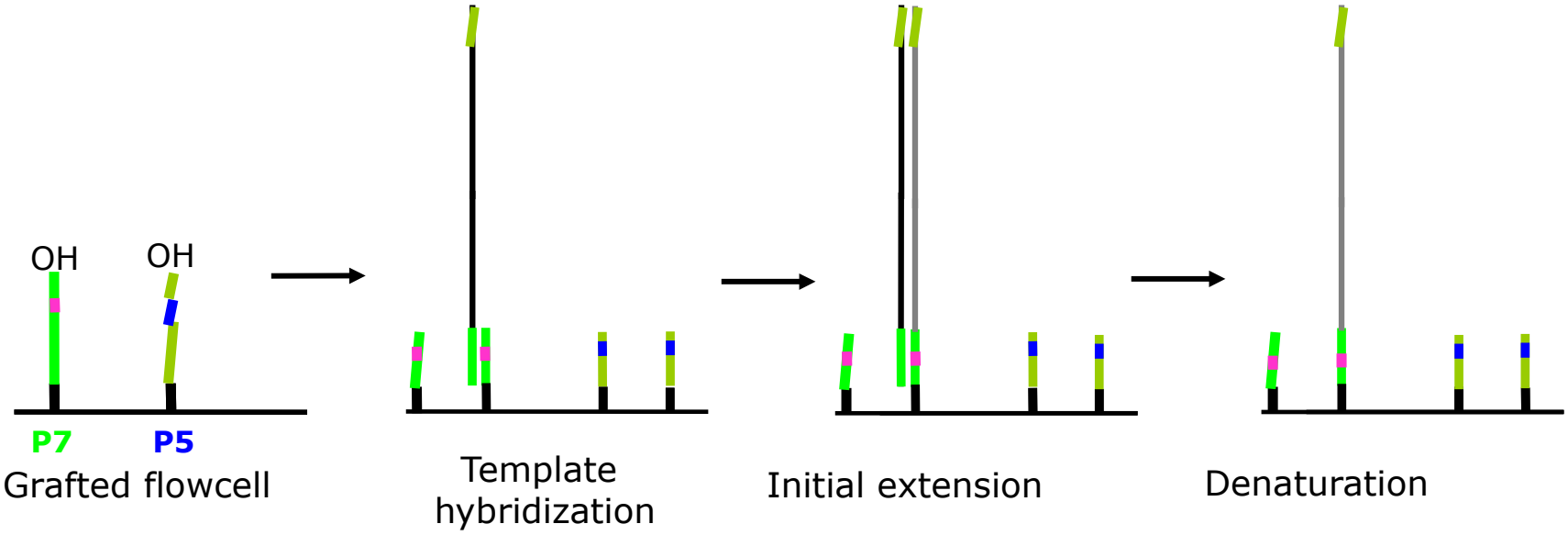
Single pass



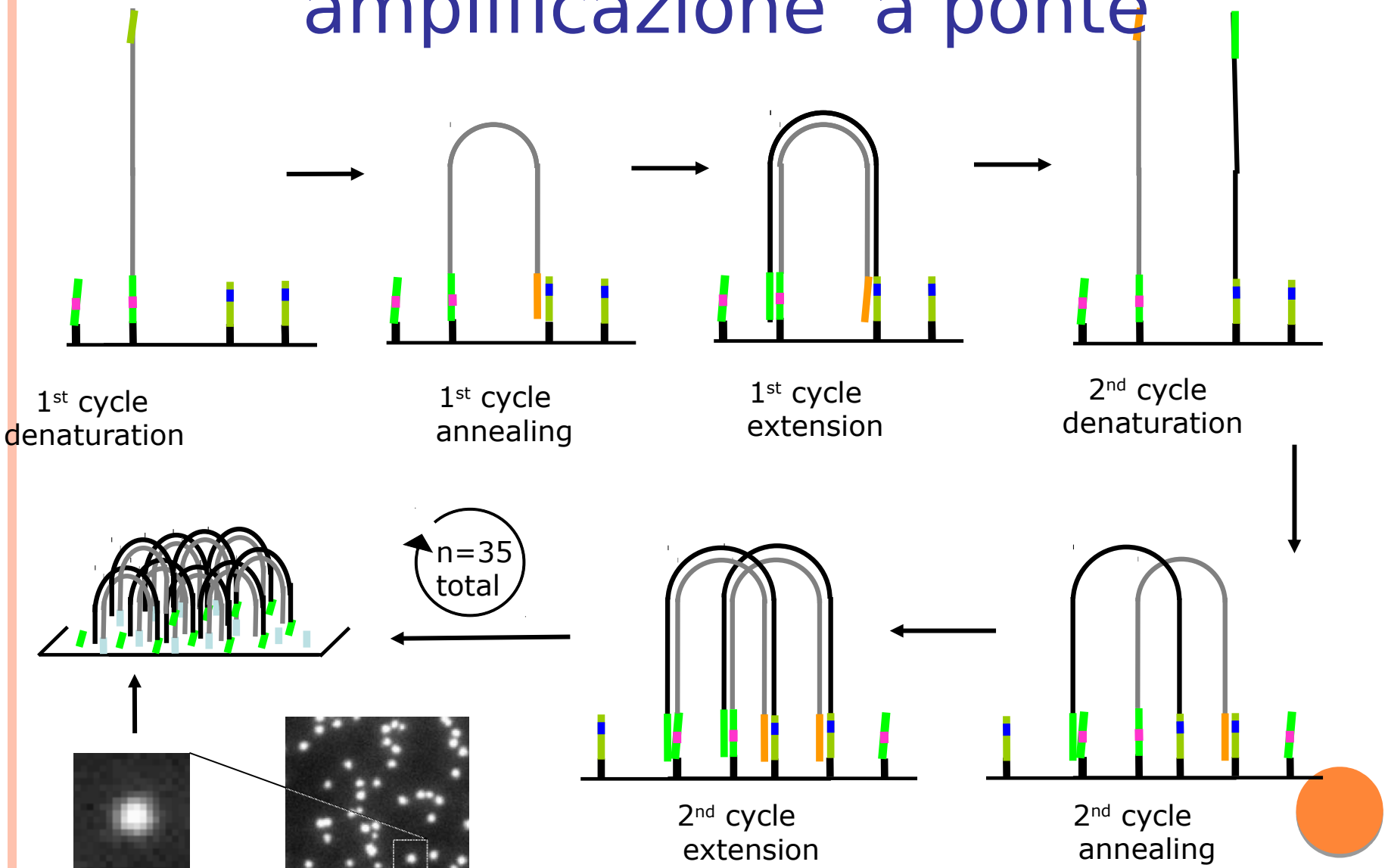
Paired end



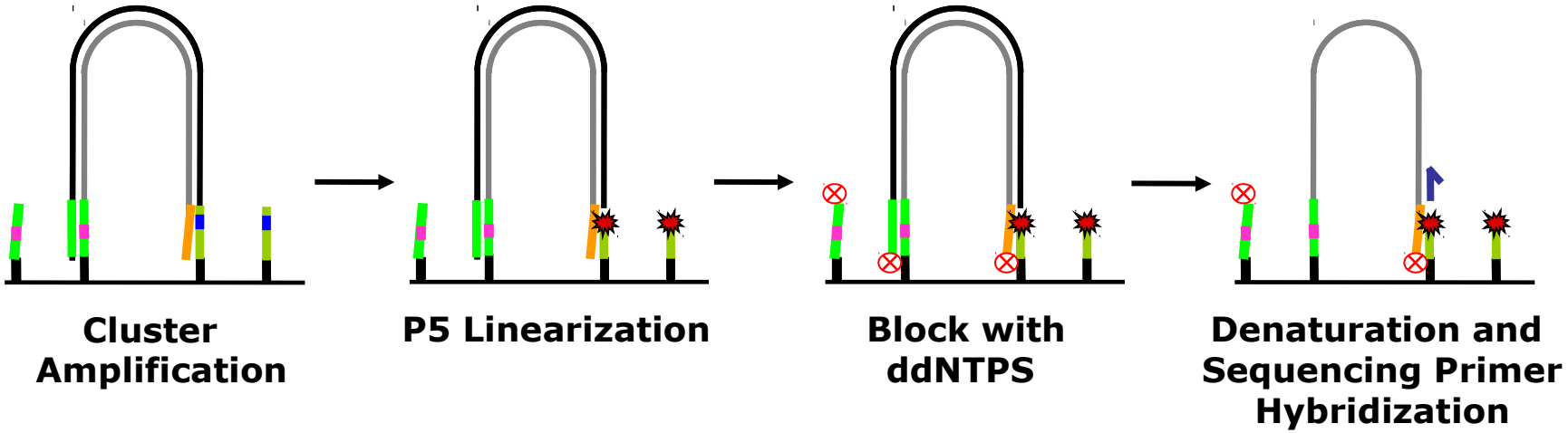
Cluster Generation



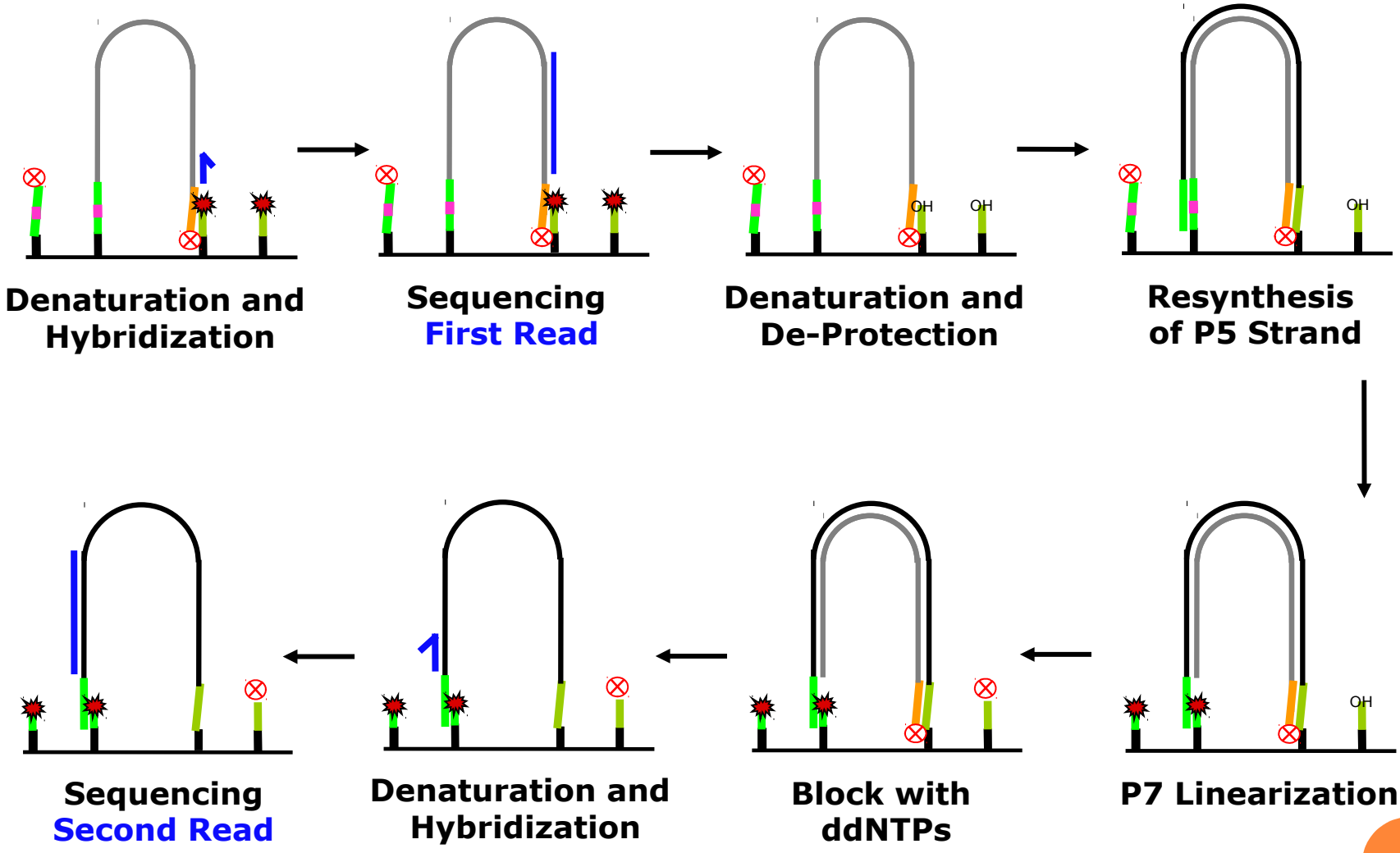
La generazione dei cluster: amplificazione a ponte



La generazione dei cluster

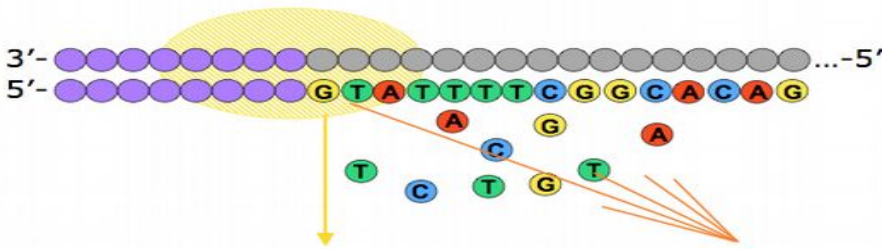


Sequenziamento pair-end



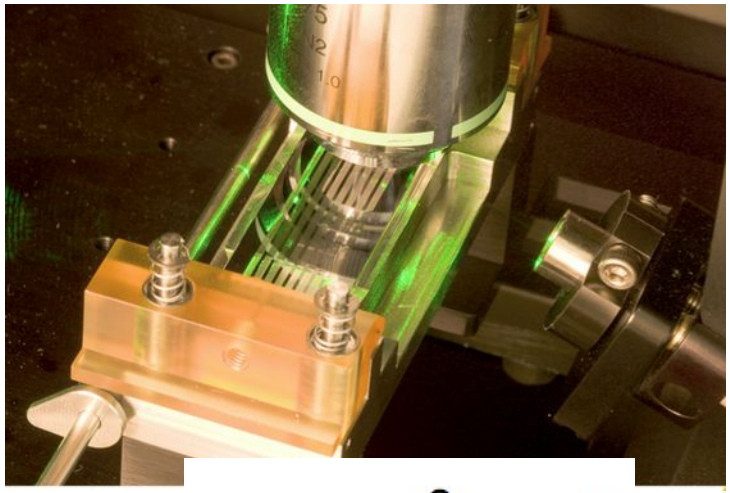
Sequencing Chemistry

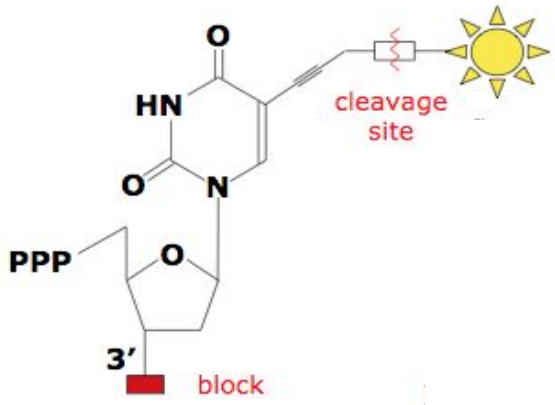
Sequencing by Synthesis



Cycle 1: Add sequencing reagents
First base incorporated
Remove unincorporated bases
Detect signal

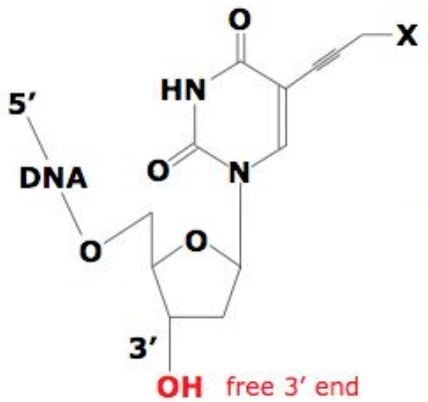
Cycle 2-n: Add sequencing reagents and repeat



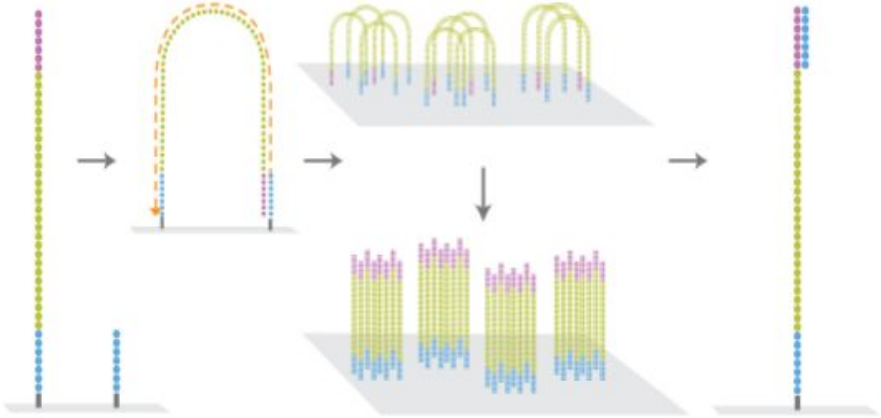


Cleave fluorophore

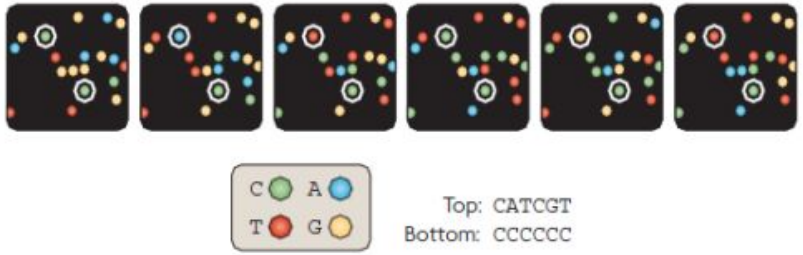
De-block 3′ terminus

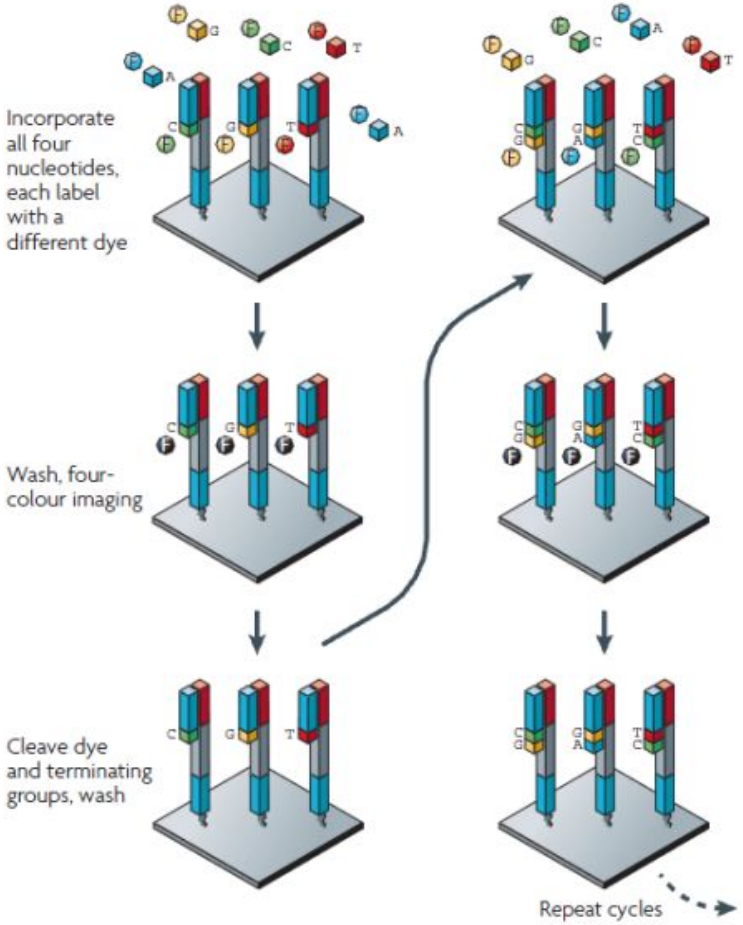


Illumina: Reversible termination sequencing

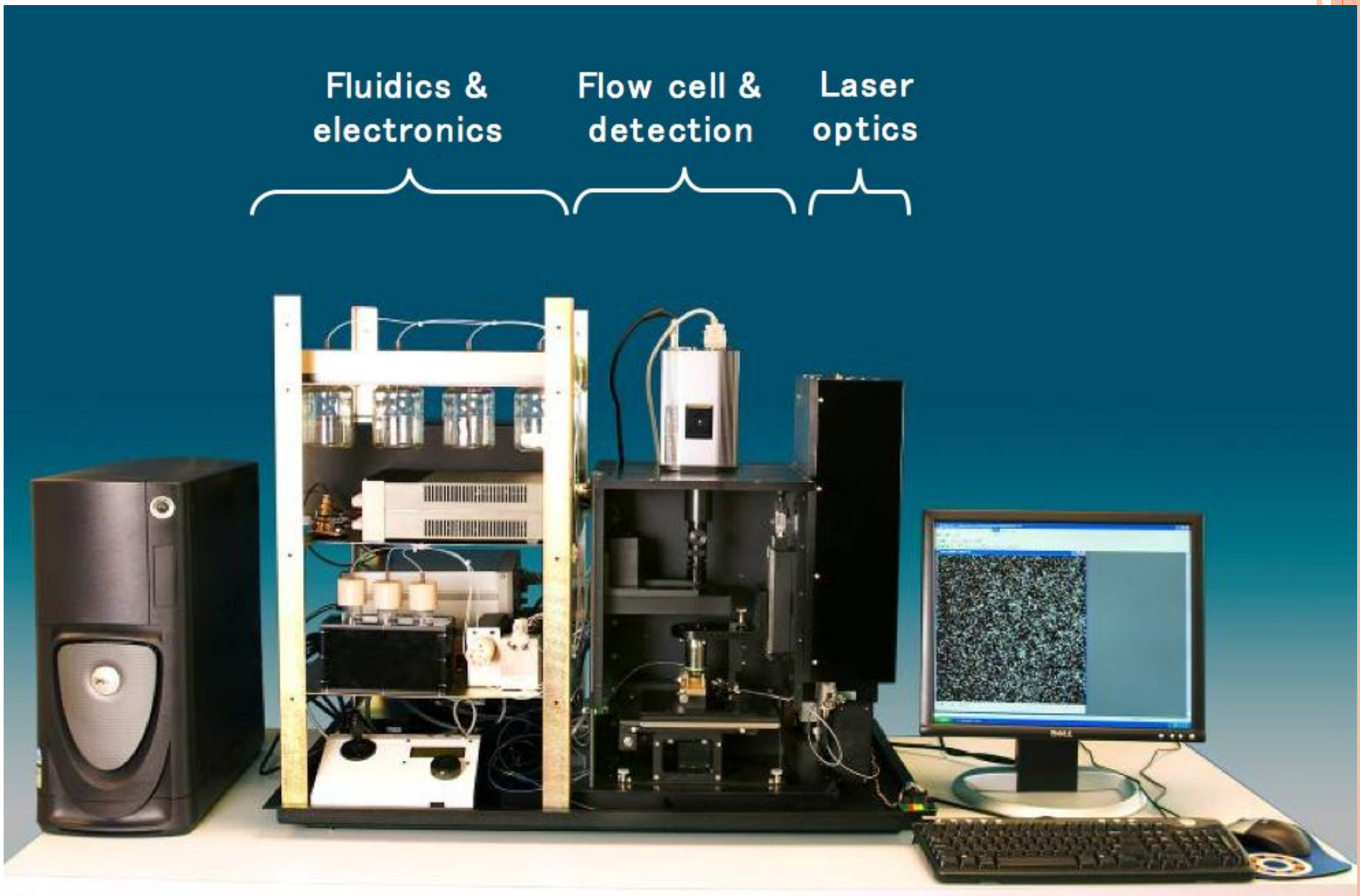


4 nucleotides with different dye flow simultaneous

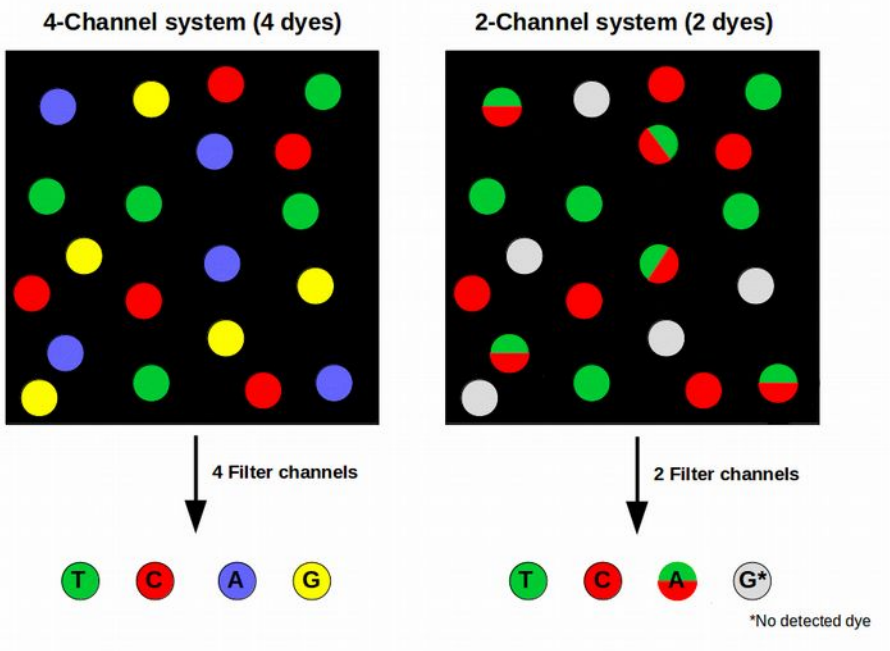


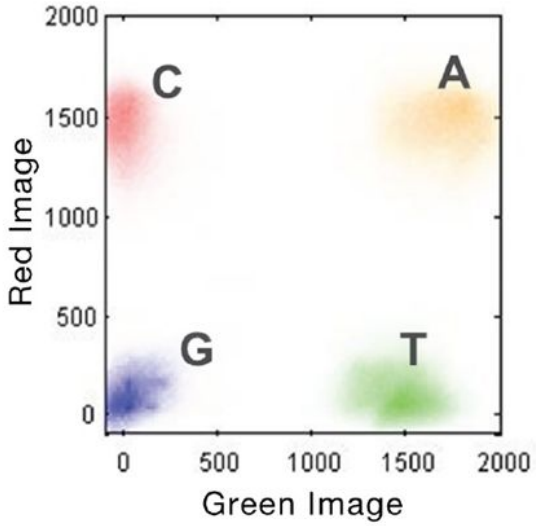


Illumina 2- and 4-channel SBS (sequencing by synthesis) sequencing technology

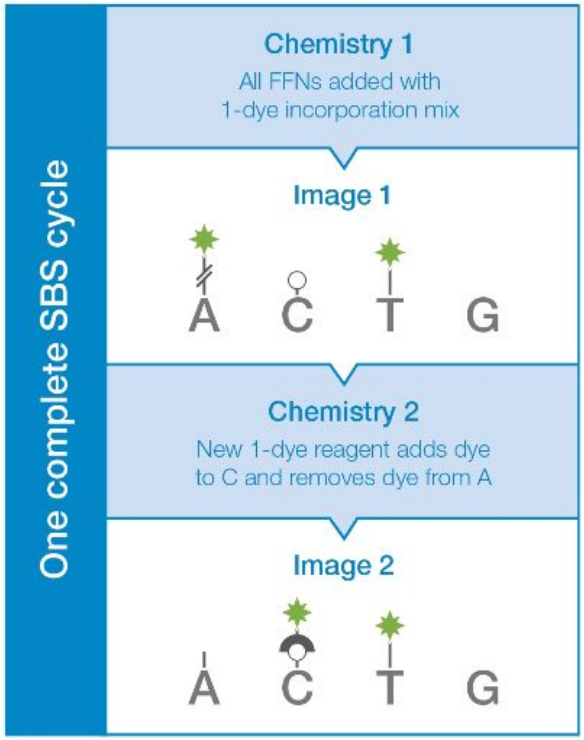

Fluidics &
electronics
Flow cell &
detection
Laser
optics

ILLUMINA DETECTION CHANGES





A.



B.

Image 1	Image 2	Result
ON	OFF	A
OFF	ON	C
ON	ON	T
OFF	OFF	G

Figure 5: One-Channel SBS Chemistry—(A) One-channel SBS chemistry features two chemistry steps and two imaging steps per sequencing cycle using nucleotides that can be labeled or unlabeled depending on the chemistry step. (B) The base call is determined by the signal pattern across both images.

NovaSeq 5000 / 6000

HiSeq X

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		●	●	●
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

Benchtop Sequencers	Production-Scale Sequencers		

NextSeq Series

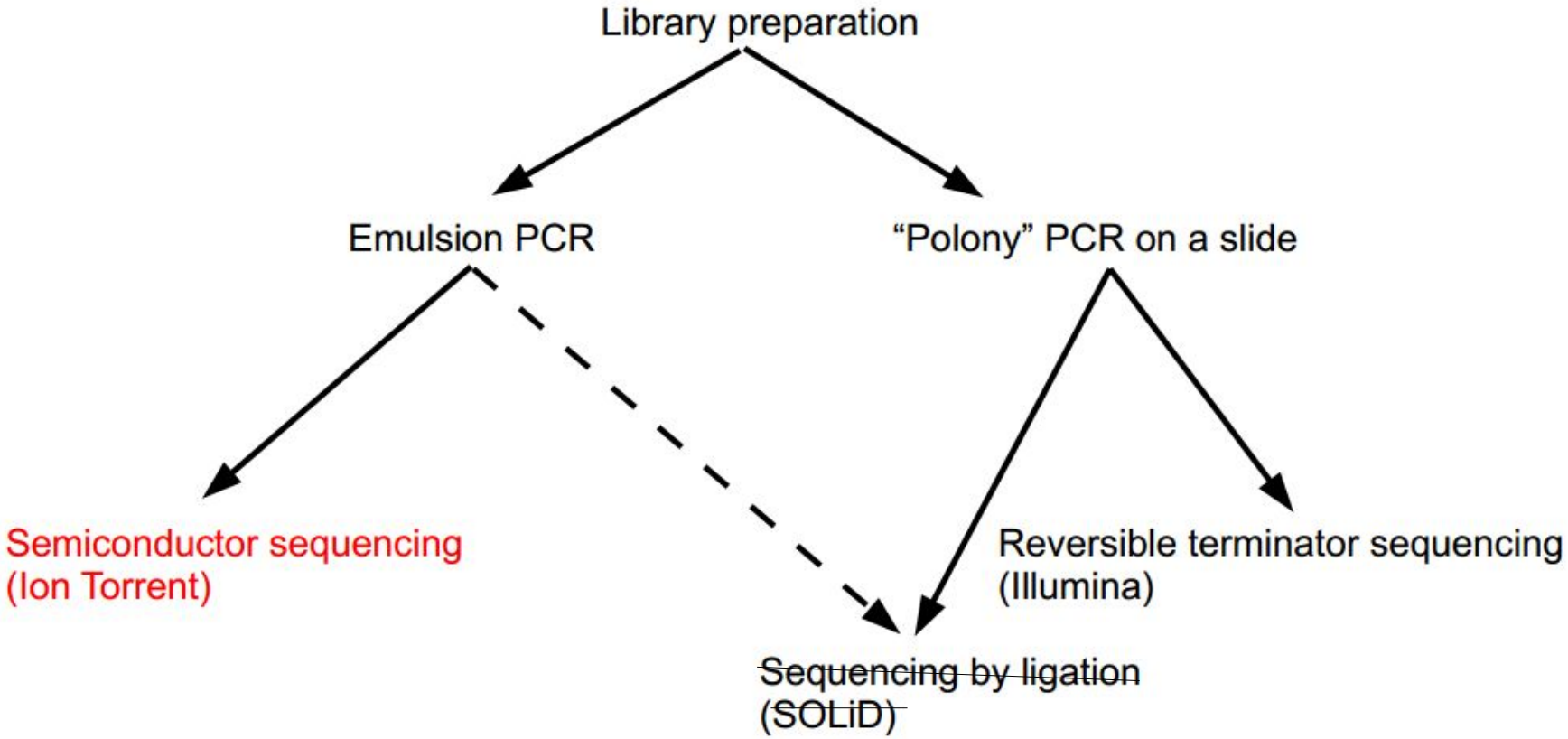
HiSeq Series

HiSeq X Series‡

NovaSeq 6000 System

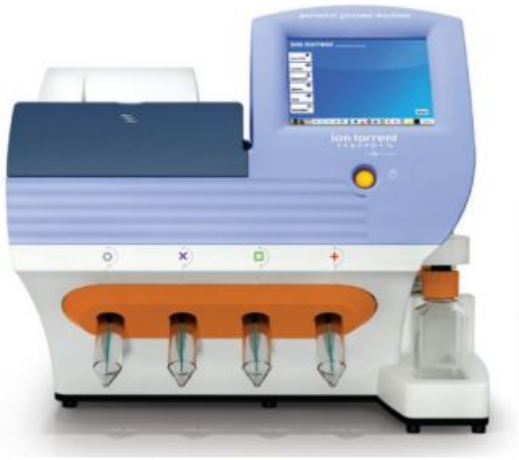
Popular Applications & Methods	Key Application	Key Application	Key Application	Key Application
Large Whole-Genome Sequencing (human, plant, animal)	●	●	●	●
Small Whole-Genome Sequencing (microbe, virus)	●	●		●
Exome Sequencing	●	●		●
Targeted Gene Sequencing (amplicon, gene panel)	●	●		●
Whole-Transcriptome Sequencing	●	●		●
Gene Expression Profiling with mRNA-Seq	●	●		●
miRNA & Small RNA Analysis	●	●		●
DNA-Protein Interaction Analysis	●	●		●
Methylation Sequencing	●	●		●
Shotgun Metagenomics	●	●		●
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

Next Generation Sequencing: Amplified Single Molecule Sequencing



Ion Torrent

PGM
(Personal Genome Machine)



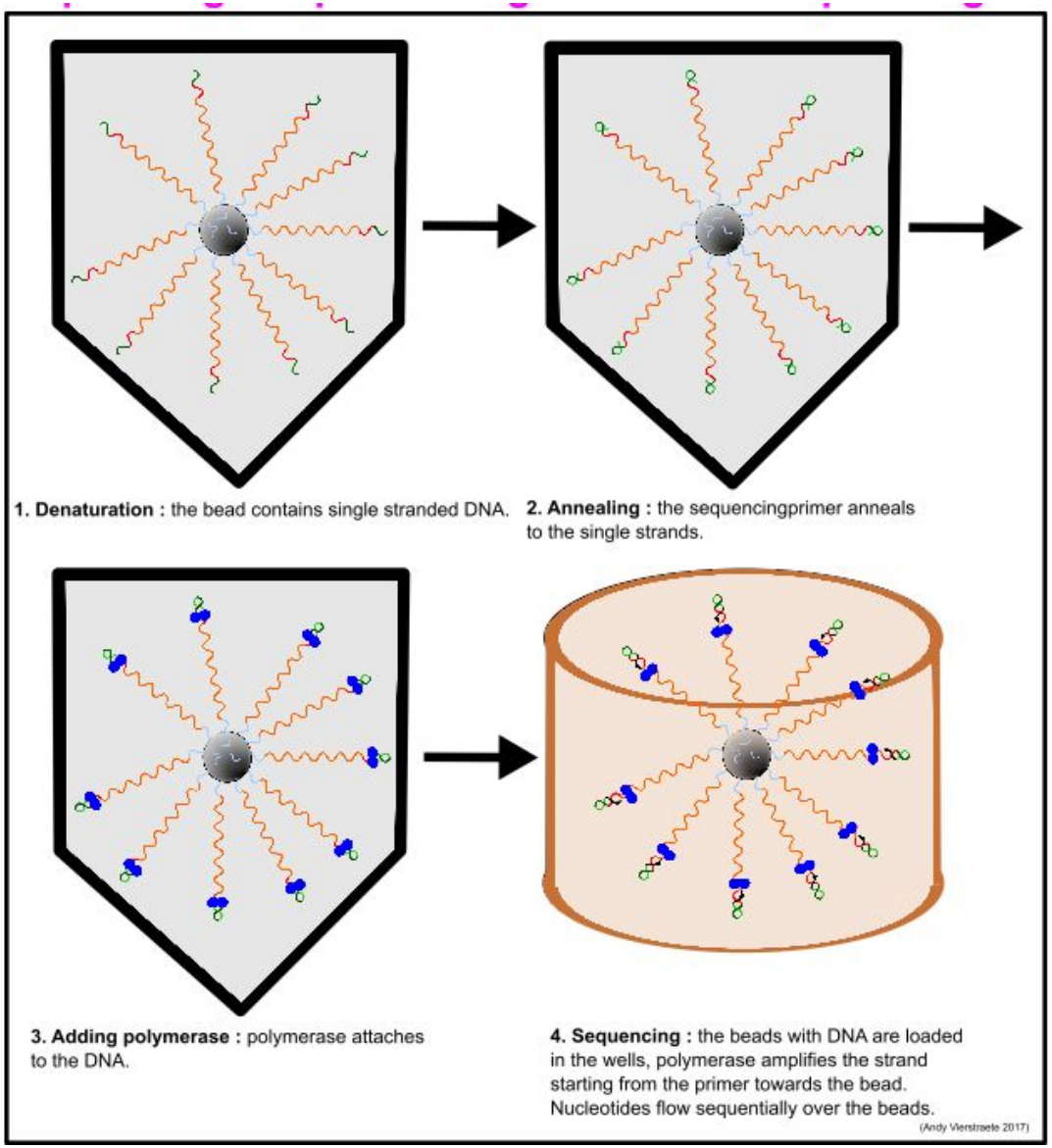
S5 / S5 XL

Proton

	PGM	S5 / S5 XL	Proton
Chip	314 – 316 - 318	520 – 530 – 540	PI – PII
Read length	400 bp	400/600 – 400/600 – 200	200 bp - ?
Throughput	0,1 – 0,6 – 2 Gb	2 – 8 - 15 Gb	10 -100 Gb
Reads per run	0,5 – 3 – 6 million	5 – 20 – 80 million	80 - 250 million
Accuracy	99 % (raw read)	99 % (raw read)	99 % (raw read)
Run Time	4 – 5 – 7 hours	8 – 17 - 17 hours (4 times faster for XL)	4 hours

Ion Torrent

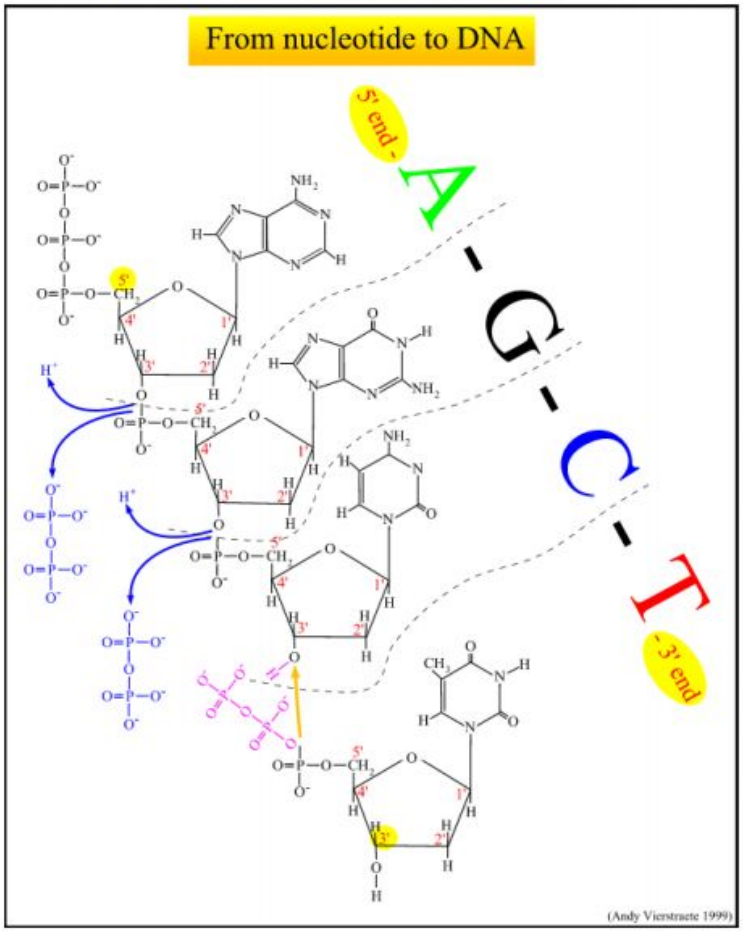
Sequencing

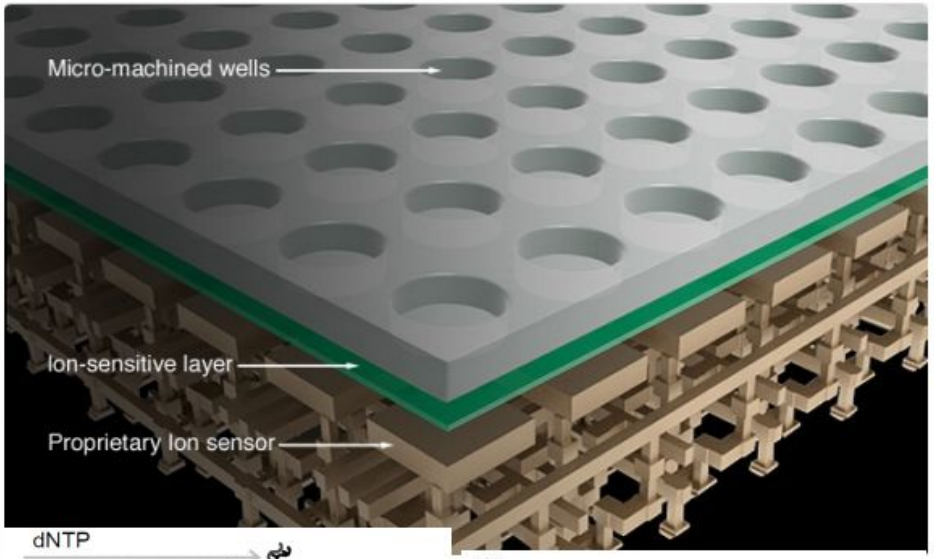


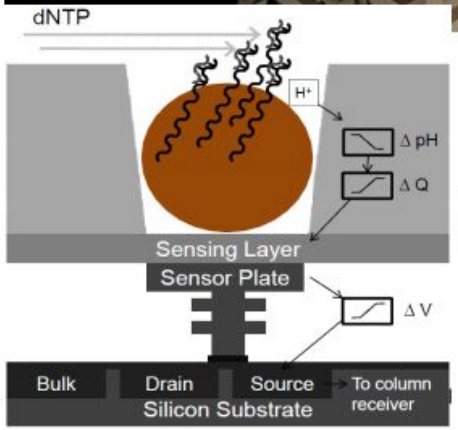
Next Generation Sequencing: Amplified Single Molecule Sequencing

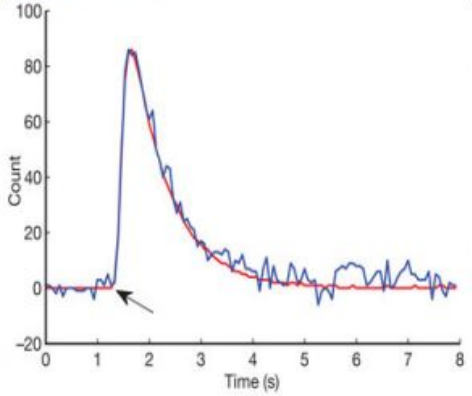
Ion Torrent

Workflow: Library preparation → Emulsion PCR → Semiconductor Sequencing





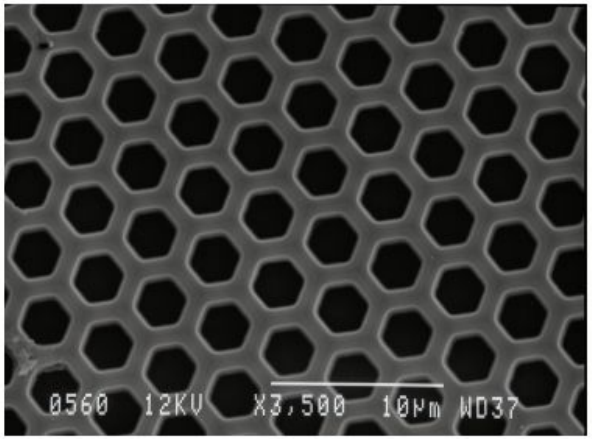


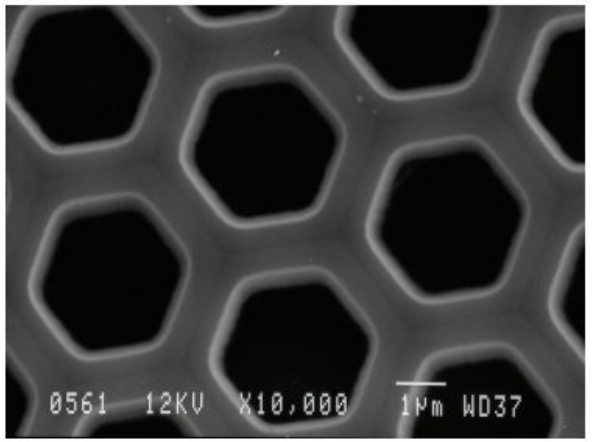


Next Generation Sequencing: Amplified Single Molecule Sequencing

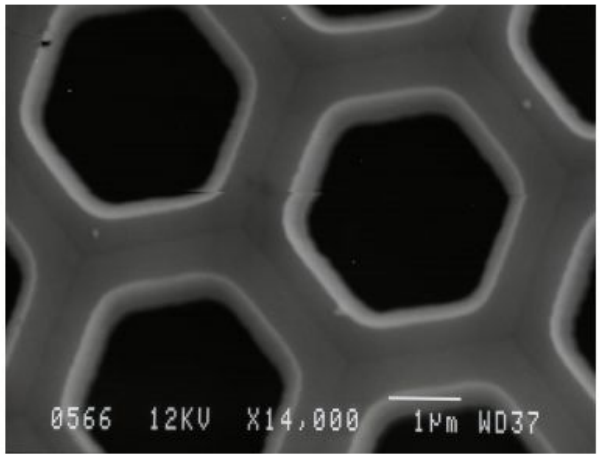
Ion Torrent

Workflow: Library preparation → Emulsion PCR → Semiconductor Sequencing





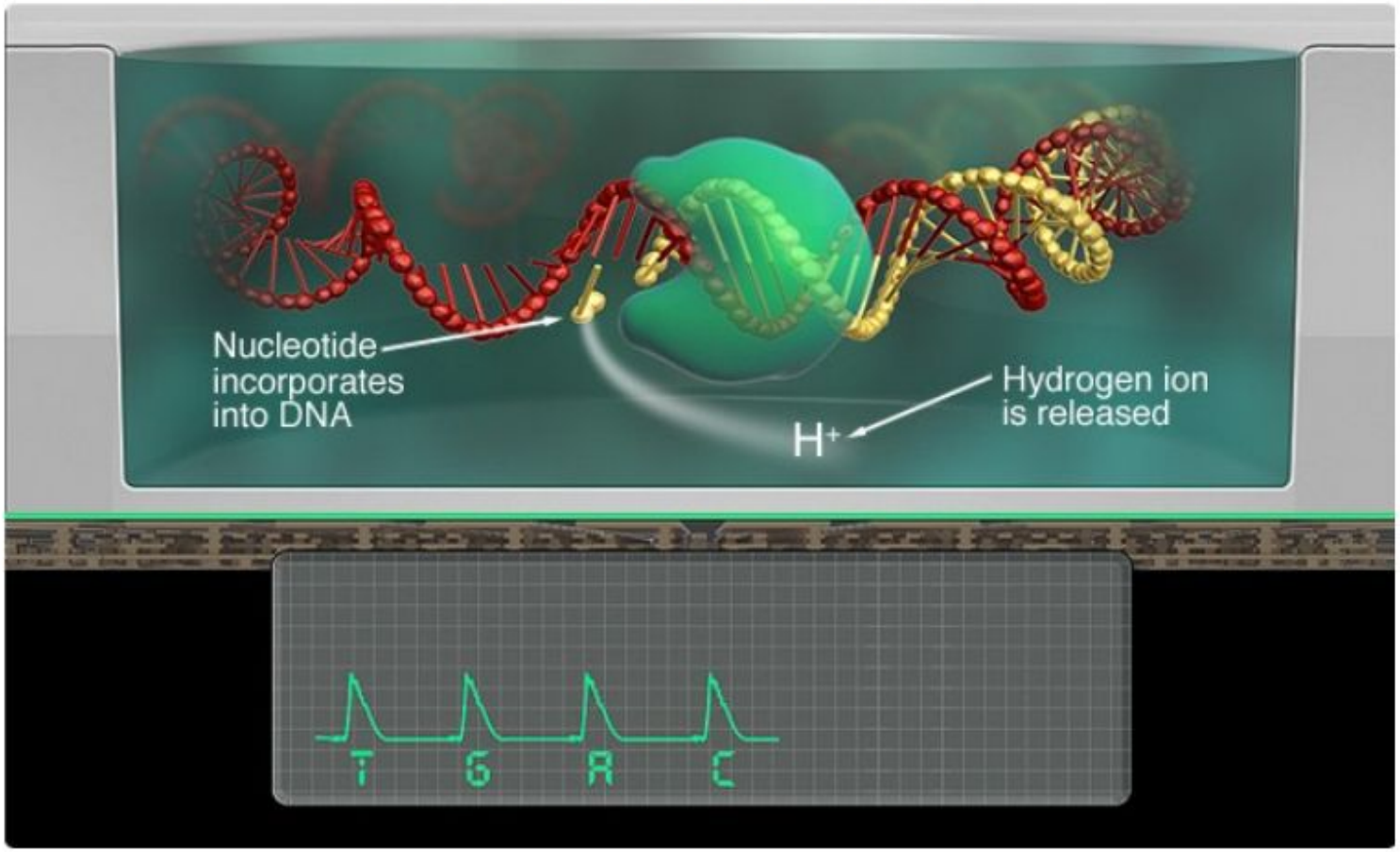
Picture of the wells in a 318 chip



Next Generation Sequencing: Amplified Single Molecule Sequencing

Ion Torrent

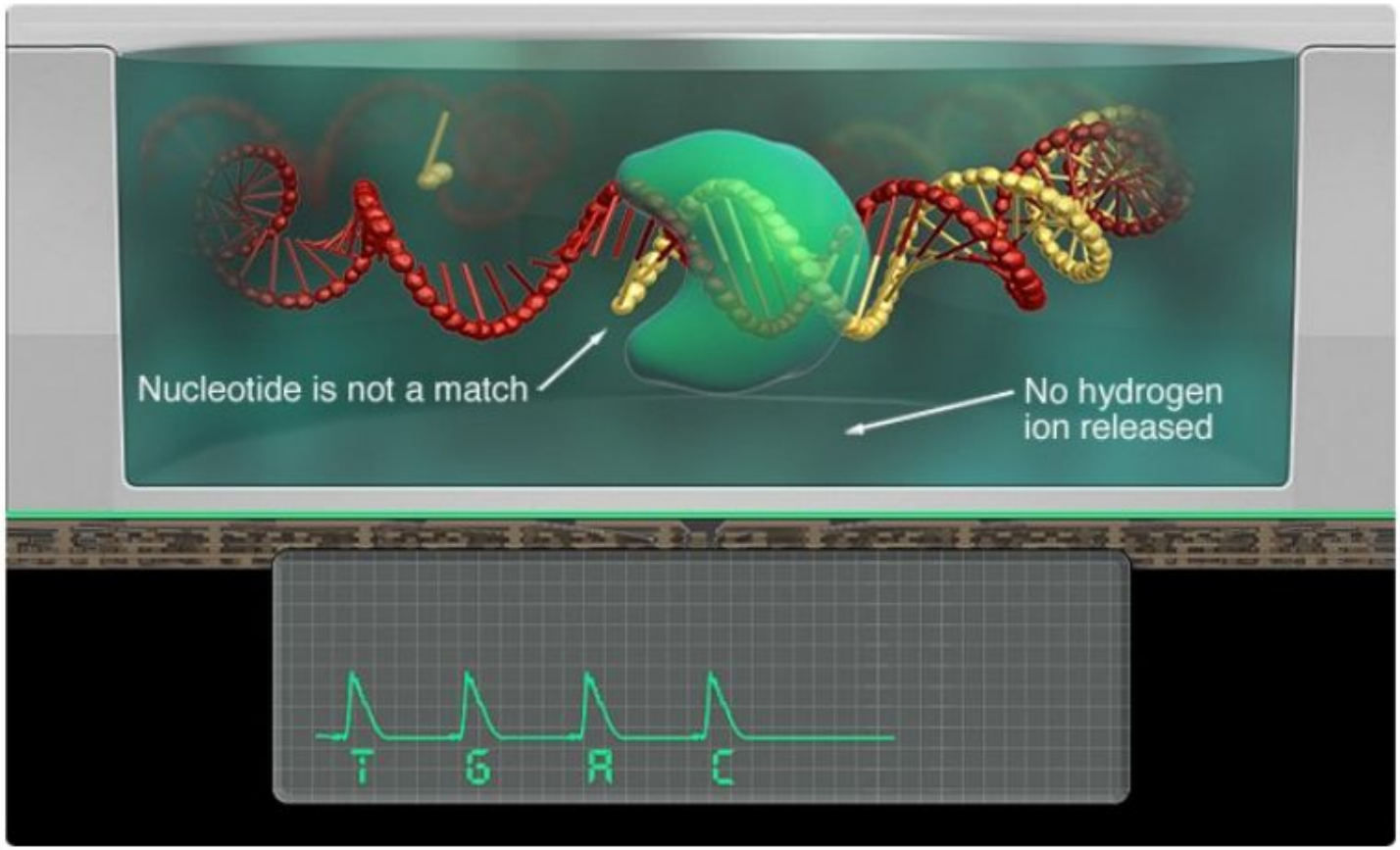
4 nucleotides flow sequentially



No camera, just a pH sensor

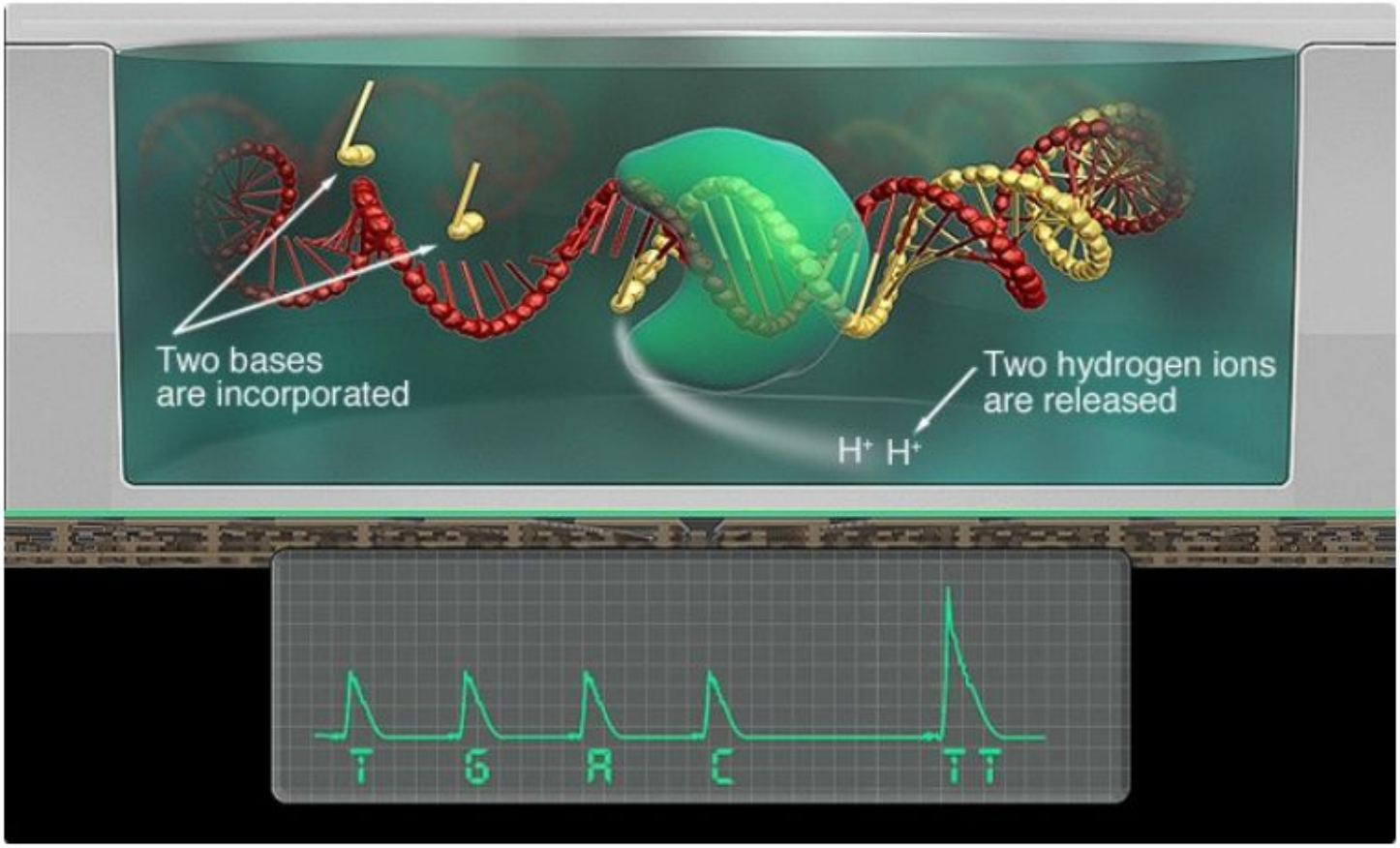
Next Generation Sequencing: Amplified Single Molecule Sequencing

Ion Torrent



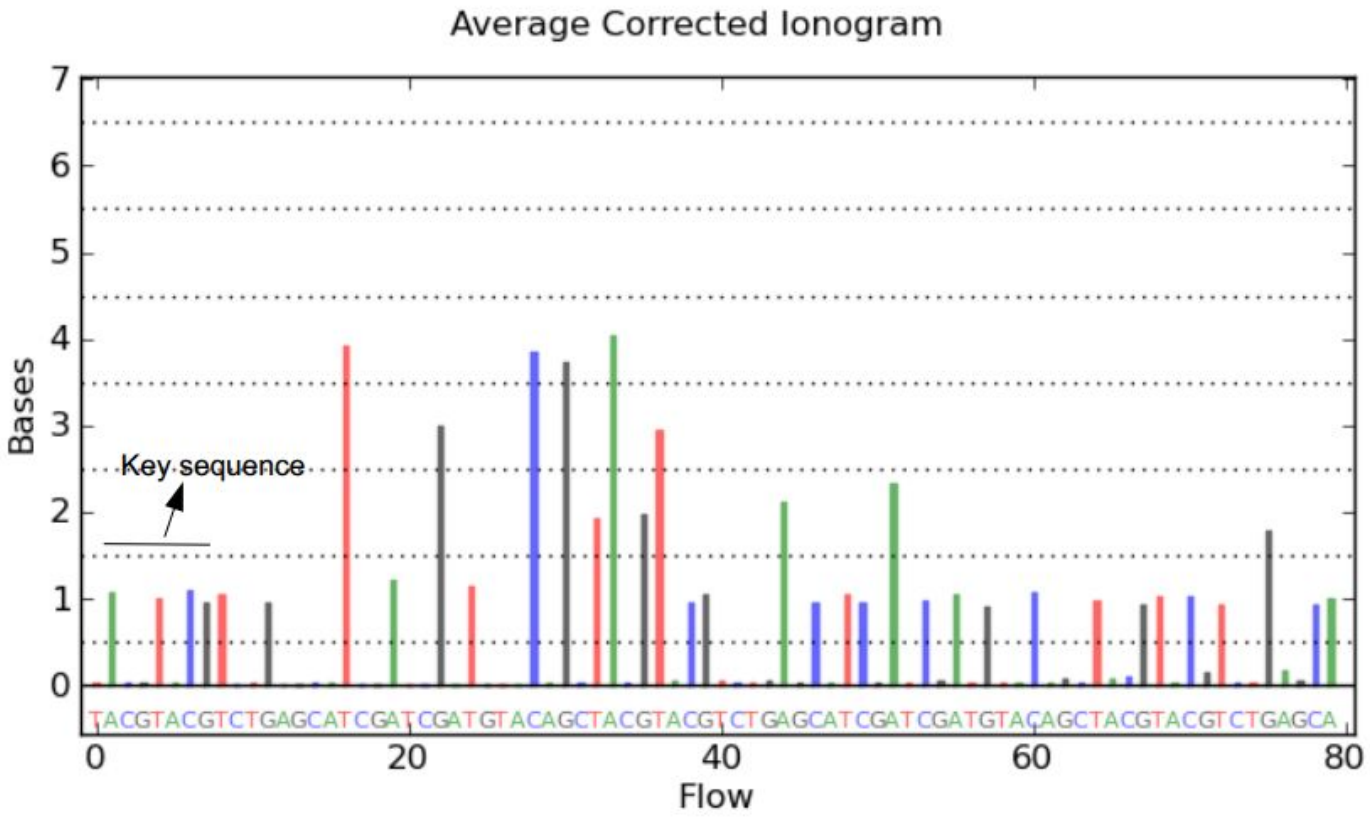
Next Generation Sequencing: Amplified Single Molecule Sequencing

Ion Torrent



Next Generation Sequencing: Amplified Single Molecule Sequencing

Ion Torrent

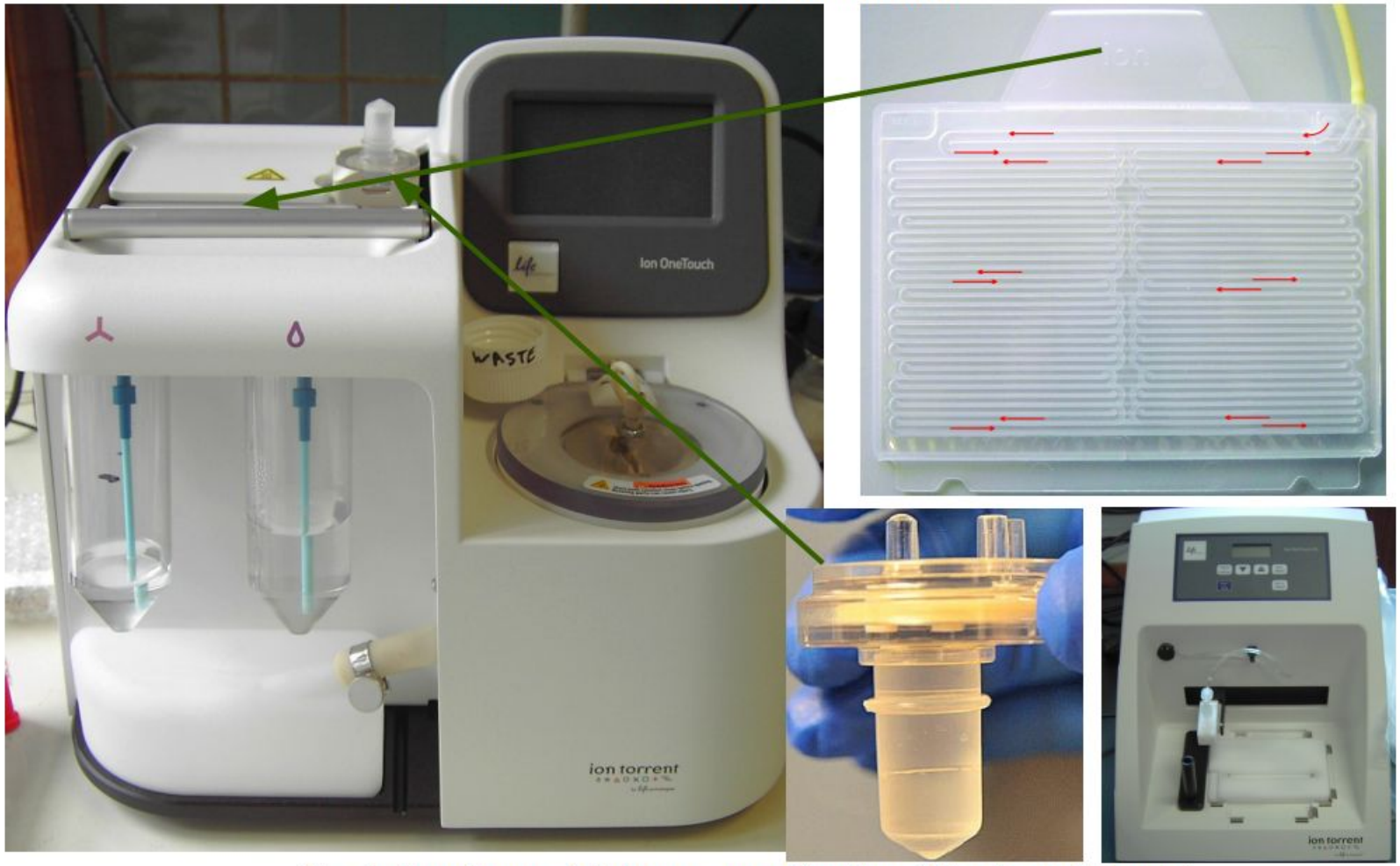


ATCGTGTTTTAGGGTCCCCGGGGTT...

Next Generation Sequencing: Amplified Single Molecule Sequencing

Ion Torrent

Emulsion PCR: OneTouch Instrument

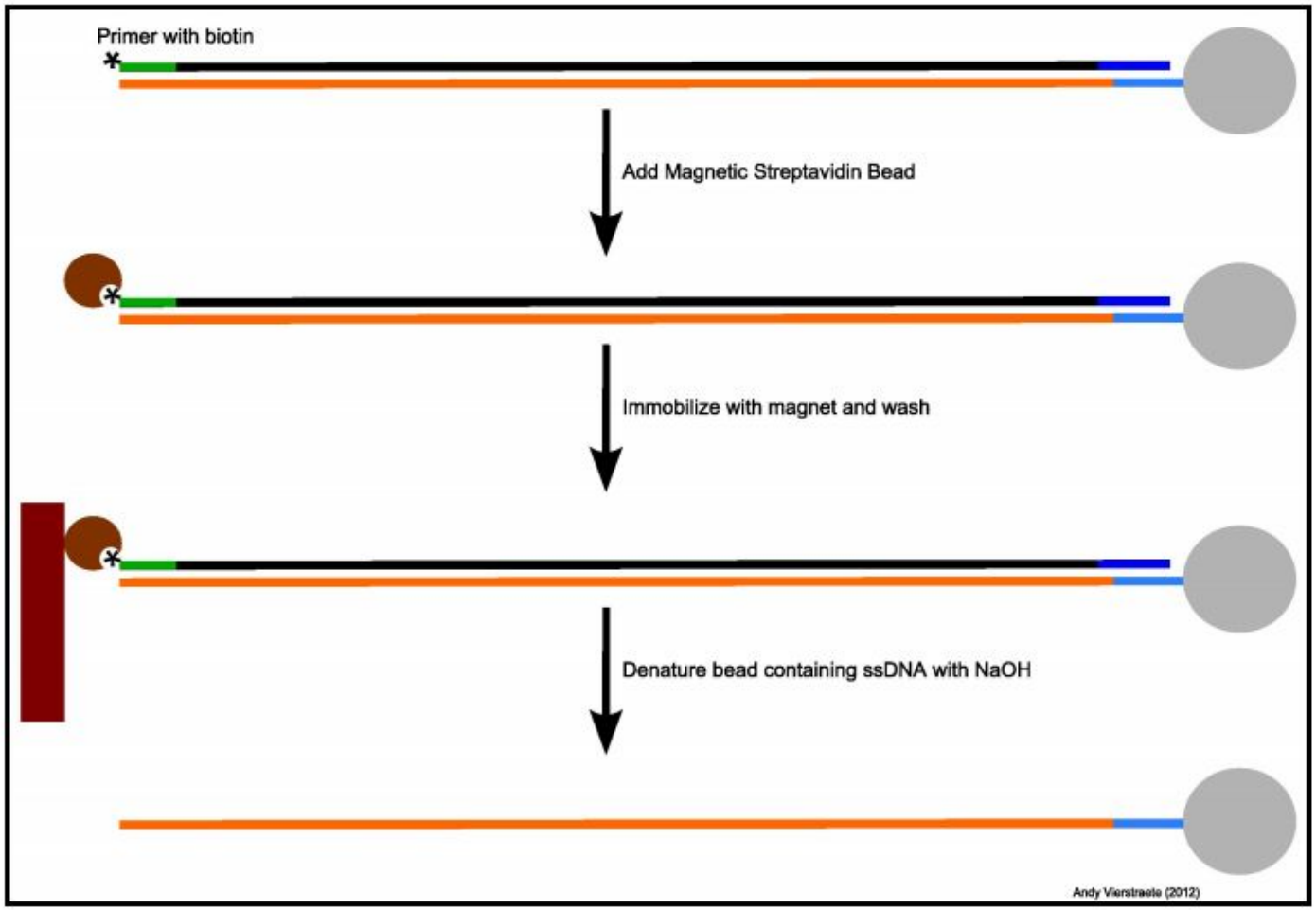


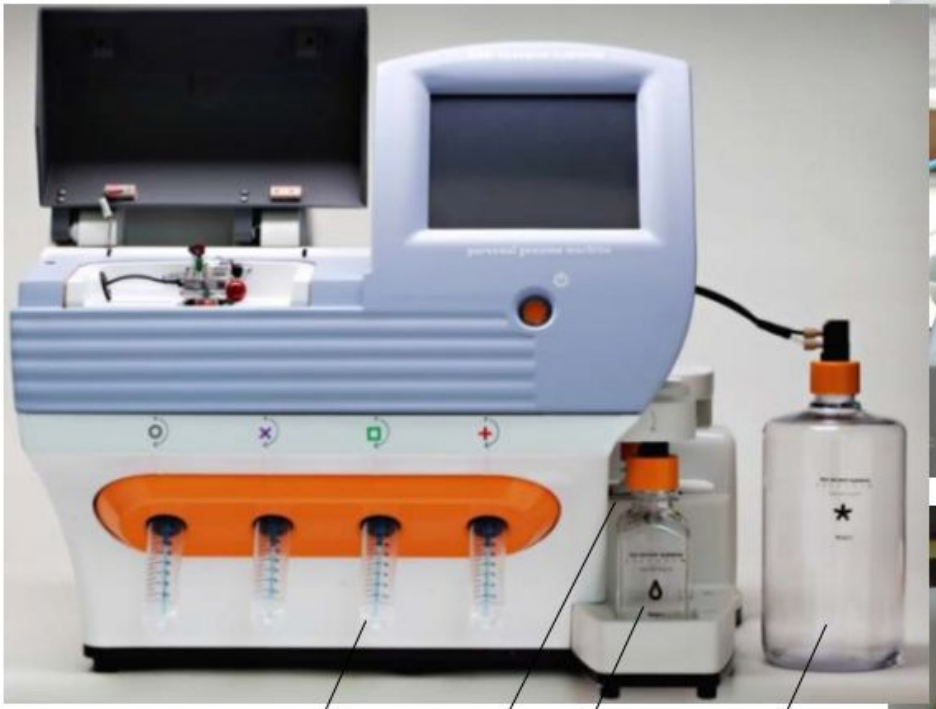
15 min hands-on; 4-8 hours amplification; 35 min enrichment

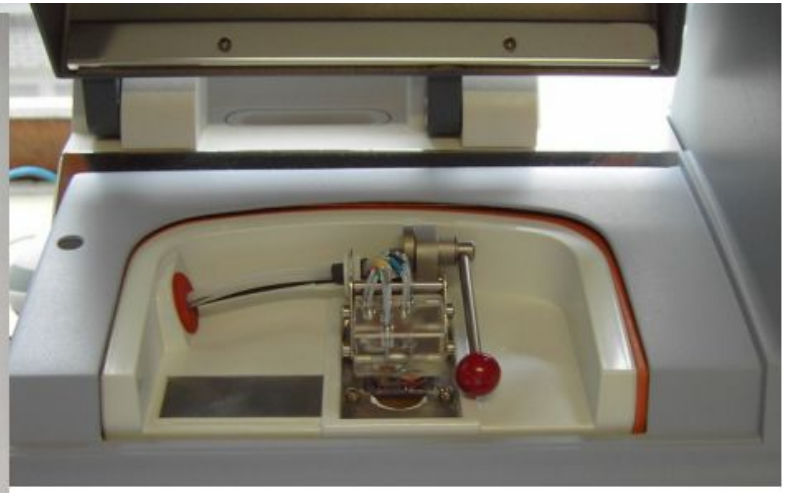
Next Generation Sequencing: Amplified Single Molecule Sequencing

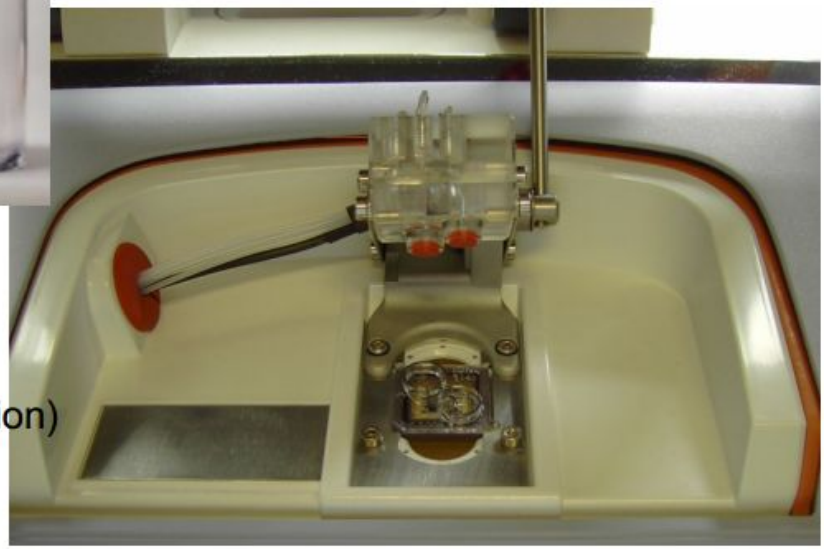
Ion Torrent

Enrichment: select only the beads that contain DNA
-> maximizing sequencing yield









W1 bottle: 350µl 100 mM NaOH
W2 bottle: 2 liter W2 solution (contains pcr solution)
W3 bottle: 50 ml W3 (= buffer with known pH)
4 tubes with 20 µl of the different dNTP's

Reports : Auto_ionadmin_SN2-129-Chip1-_pollini_e_funghi_ITS_4-10-18_373 (183)

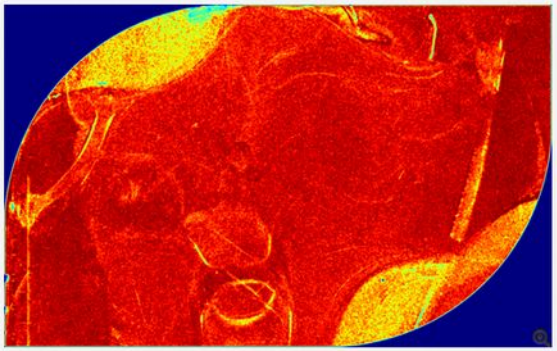
Run Summary: R_2018_10_04_12_57_15_ionadmin_SN2-129-Chip1-_pollini_e_funghi_ITS_4-10-18

Read Summary: Unaligned

1.18 G	54
Total Bases	Key Signal

89 %
ISP Loading

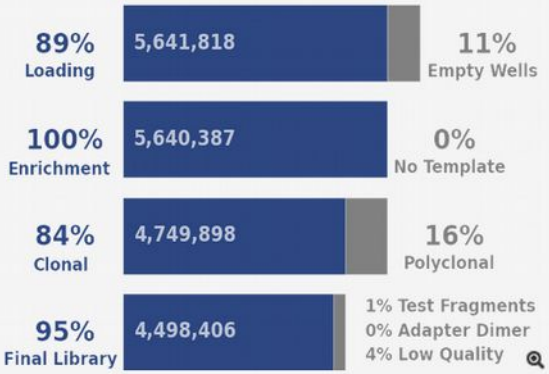
ISP Density



4,498,406
Total Reads

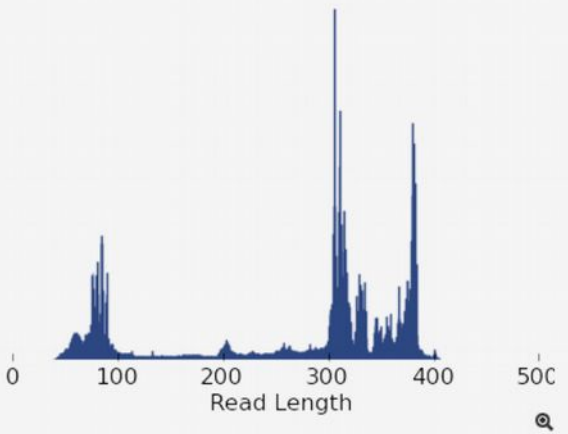
81 %
Usable Reads

ISP Summary



262 bp	308 bp	304 bp
Mean	Median	Mode

Read Length



Pacific Biosciences: il futuro, il sequenziamento massivo di singole molecole

Published online before print January 23, 2008, 10.1073/pnas.0710982105
PNAS | **January 29, 2008** | vol. 105 | no. 4 | **1176-1181**

◂ Previous Article | Table of Contents | Next Article ▸

BIOLOGICAL SCIENCES / BIOPHYSICS

Selective aluminum passivation for targeted immobilization of single DNA polymerase molecules in zero-mode waveguide nanostructures

Jonas Korlach, Patrick J. Marks, Ronald L. Cicero, Jeremy J. Gray, Devon L. Murphy, Daniel B. Roitman, Thang T. Pham, Geoff A. Otto, Mathieu Foquet, and Stephen W. Turner*

Pacific Biosciences, 1505 Adams Drive, Menlo Park, CA 94025

Communicated by Watt W. Webb, Cornell University, Ithaca, NY, November 20, 2007 (received for review August 6, 2007)

Optical nanostructures have enabled the creation of subdiffraction detection volumes for single-molecule fluorescence microscopy. Their applicability is extended by the ability to place molecules in the confined observation volume without interfering with their biological function. Here, we demonstrate that processive DNA synthesis thousands of bases in length was carried out by individual DNA polymerase molecules immobilized in the observation volumes of zero-mode waveguides (ZMWs) in high-density arrays.

Third Generation Sequencing: Single Molecule Sequencing

Pacific Biosciences

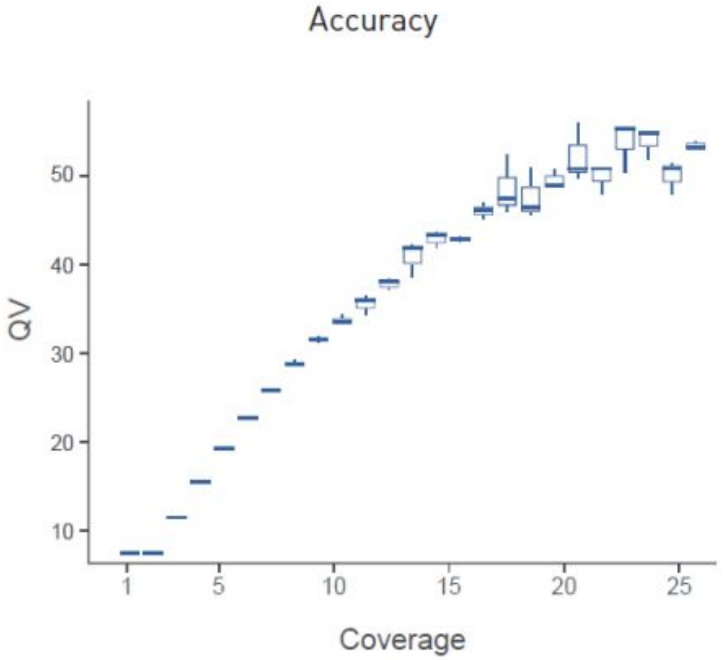
Pacbio RS

Sequel System

	Pacbio RS	Sequel System
Read Length	50 % > 20 kb (max > 60 kb)	50 % > 20 kb (max > 60 kb)
Throughput	1 Gb/SMRT cell (max 16/run)	5-8 Gb/SMRT cell (max 16/run)
Reads per run	55,000	365,000
Accuracy	86 %	86 %
Run Time	30 minutes – 6 hours/ SMRT cell	30 minutes – 10 hours/ SMRT cell

Third Generation Sequencing: Single Molecule Sequencing

Pacific Biosciences

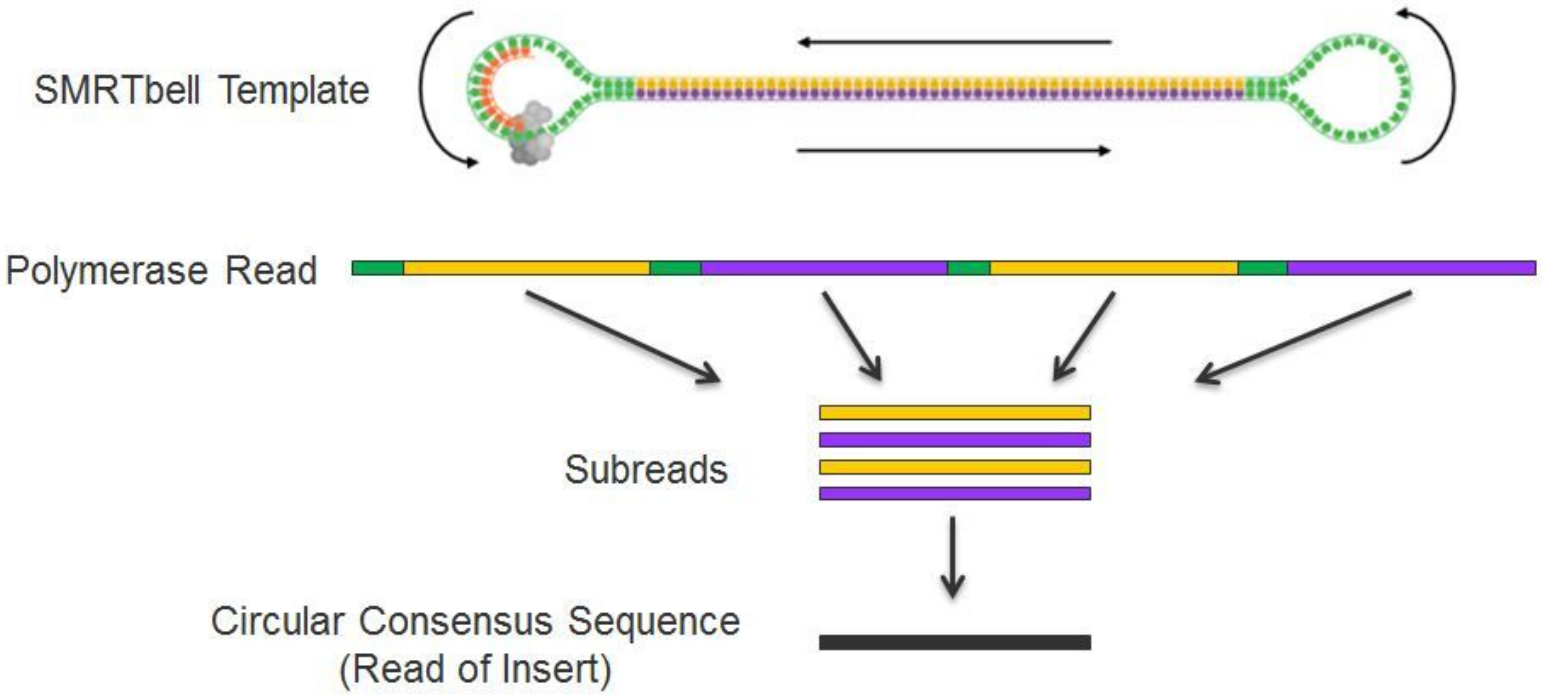


Quality scores in sequencing:
Q17, Q20, Q30, ...

Quality score	Probability of incorrect bases	Base call accuracy
10	1 in 10	90 %
17	1 in 50	98 %
20	1 in 100	99 %
30	1 in 1000	99,9 %
40	1 in 10.000	99,99 %
50	1 in 100.000	99,999 %
60	1 in 1.000.000	99,9999%

- Circular Consensus Sequencing (CCS reads)
- Consensus by sequencing many reads

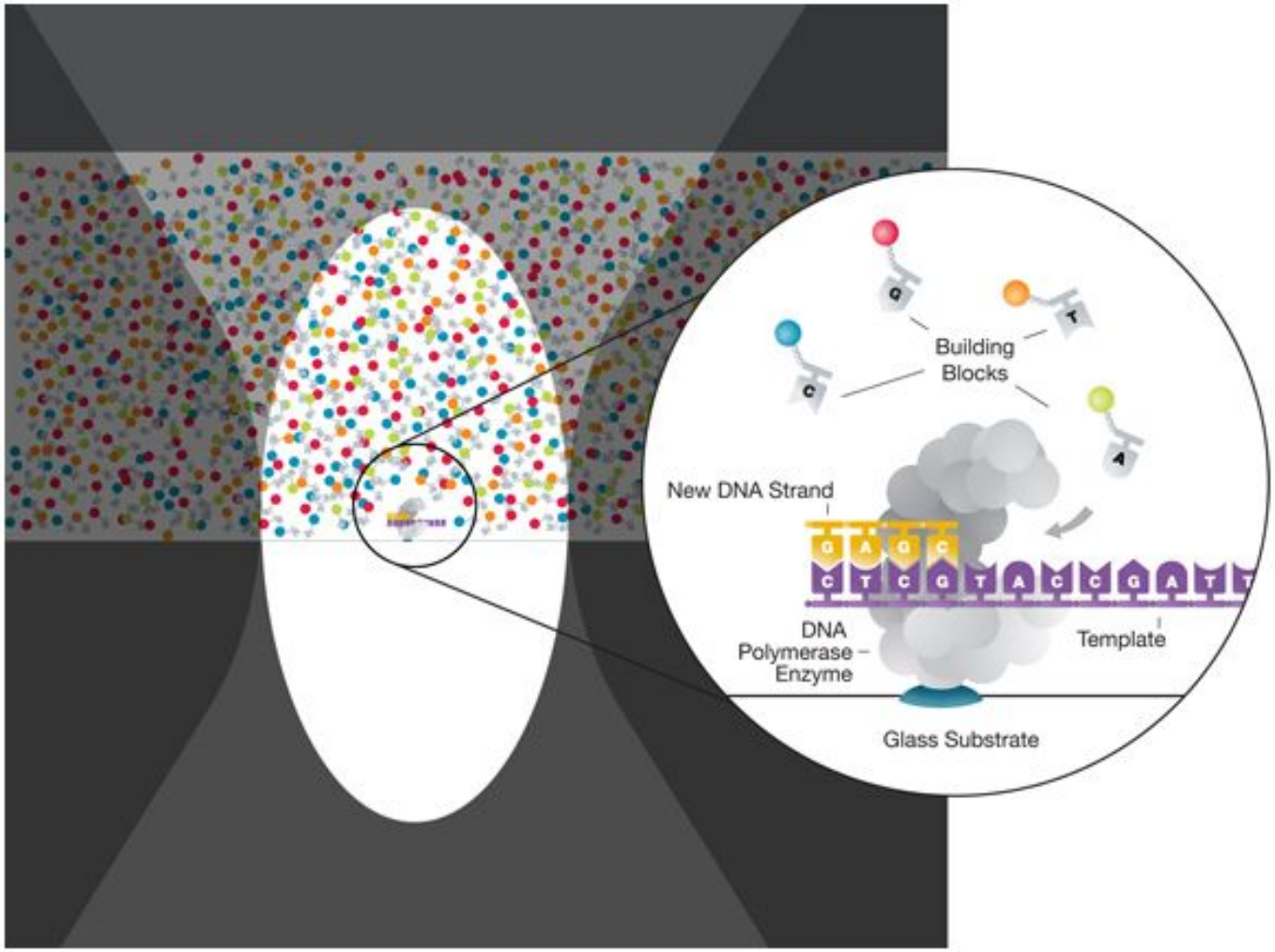
```
TGCAGATCATTACT-AAACAACGC-TCC-AC-TATCAAAT-CCGGGTGCG-CTTGTTGTATAACACAAAC-AGG-CGAAAAAACATA-TCG-AGTT
TGCAGATCATTACT-AAACAACGC-TCC-AC-TATCAAAT-CCGGGTGCG-CTTGTTGTATAACACAAAC-AGG-CGAAAAAACATA-TCGAAGTT
TGCAGATCATTACTAAAACAACGC-TCC-AC-TATCAAAT-CCGGGTGCGACTTGTTGTATAACACAACT-AGG-CGAAAAAACATA-TCG-AG-T
TGCAGATCATTACT-AAACAACGC-TCC-AC-TATCAAAT-CCGGGTGCG-CTTGTTGTATAAC
TGCAGATCATTACT-AAACAACGC-TCC-AC-TATCAAAT-CCGGGTGCG-CTTGTTGTATAACACAAAC-AGG-CGAAAAAACATA-TCG-AGTT
TGCAGATCATTACT-AAACAACGC-TCC-AC-TATCAAAT-CCGGGTGCG-CTTGTTGT
TGCAGATCATTACT-AAACAACGC-TCC-AC-TATCAAAT-CCGGGTGCG-CTTGTTGTA
TGCAGATCATTACT-AAACAACGC-TCC-ACGTATCAAAT-CCGGGTGCG-CTTGTTGTATAACACAAAC-AGG-CG-AAAAACATA-TCG-AGTT
                ACGCGTCCTAC-TATC-AAT-CCGGGTGCG-C-TGTTGTATAACACAAACTAGG-CGAAAAAACATA-TCG-AGTT
```


SMRTbell Template
Polymerase Read
Subreads
Circular Consensus Sequence
(Read of Insert)

SMRT SEQUENCING

- **SMRT=Single Molecule Real-Time**
- **una ZMW è un foro, di diametro di qualche decina di nanometri, fabbricato in un film di metallo di 100nm depositato su un substrato di diossido di silicone**
- **Ciascun ZMW diventa una camera di visualizzazione nanofotonica che fornisce un volume di rilevazione di solo 20 zeptolitri (10^{-21} litri).**
- **In un tale volume, l'attività di una singola molecola può essere rilevata in un background di migliaia di nucleotidi marcati in fluorescenza**

Pacific Biosciences ZMW



ZMW='Zero Mode Waveguide'

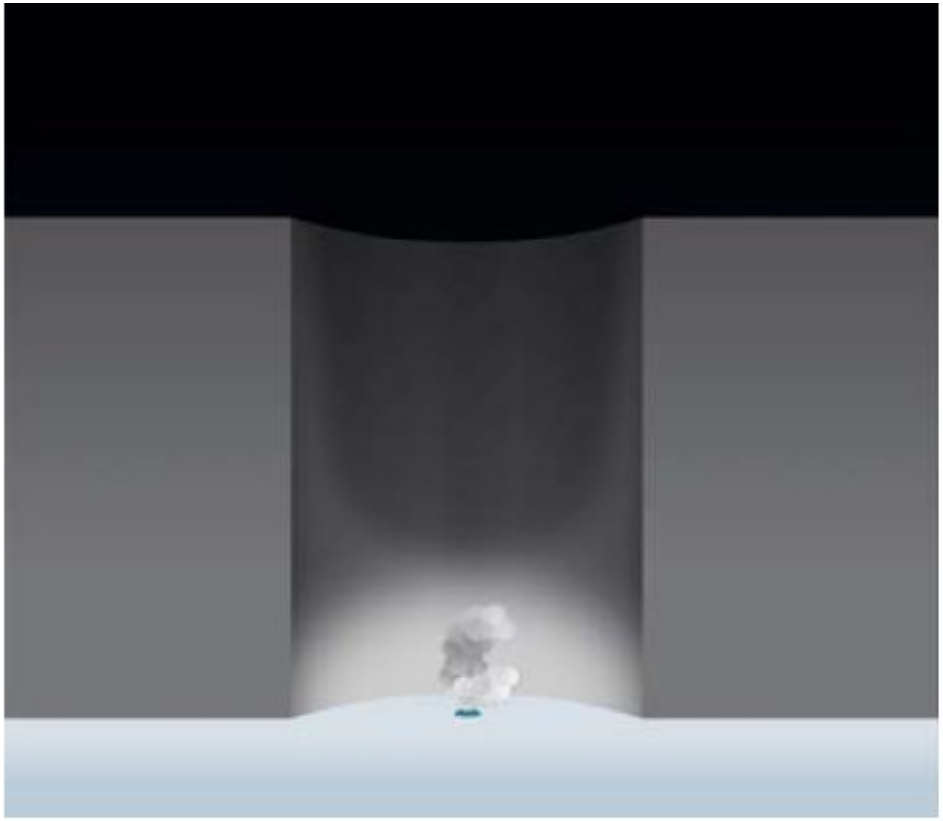
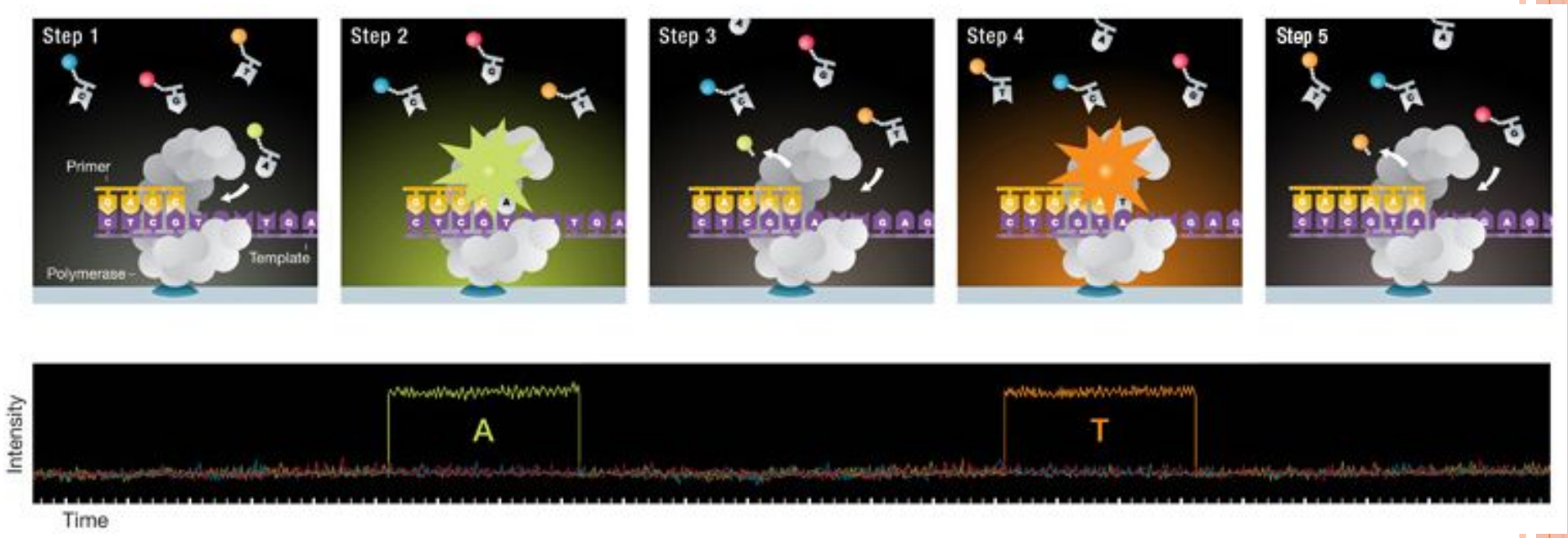


Figure 5. ***ZMW with DNA polymerase***

A single DNA polymerase molecule is attached to the bottom of the ZMW using a proprietary biased immobilization process.

Pacific Biosciences



- ✓ **Sequenziamento di singole molecole**
- ✓ **Incorporazione di molecole fluorescenti (sequenziamento mediante sintesi)**
- ✓ **Monitoraggio in tempo reale dell'attività della polimerasi**
  - **(eccitazione/rilevazione)**

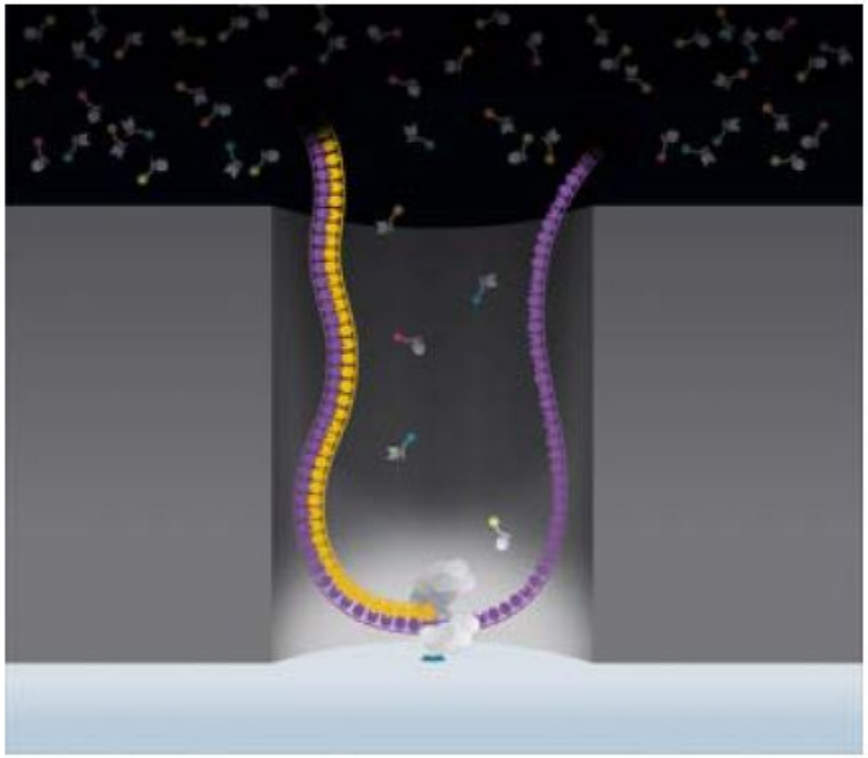
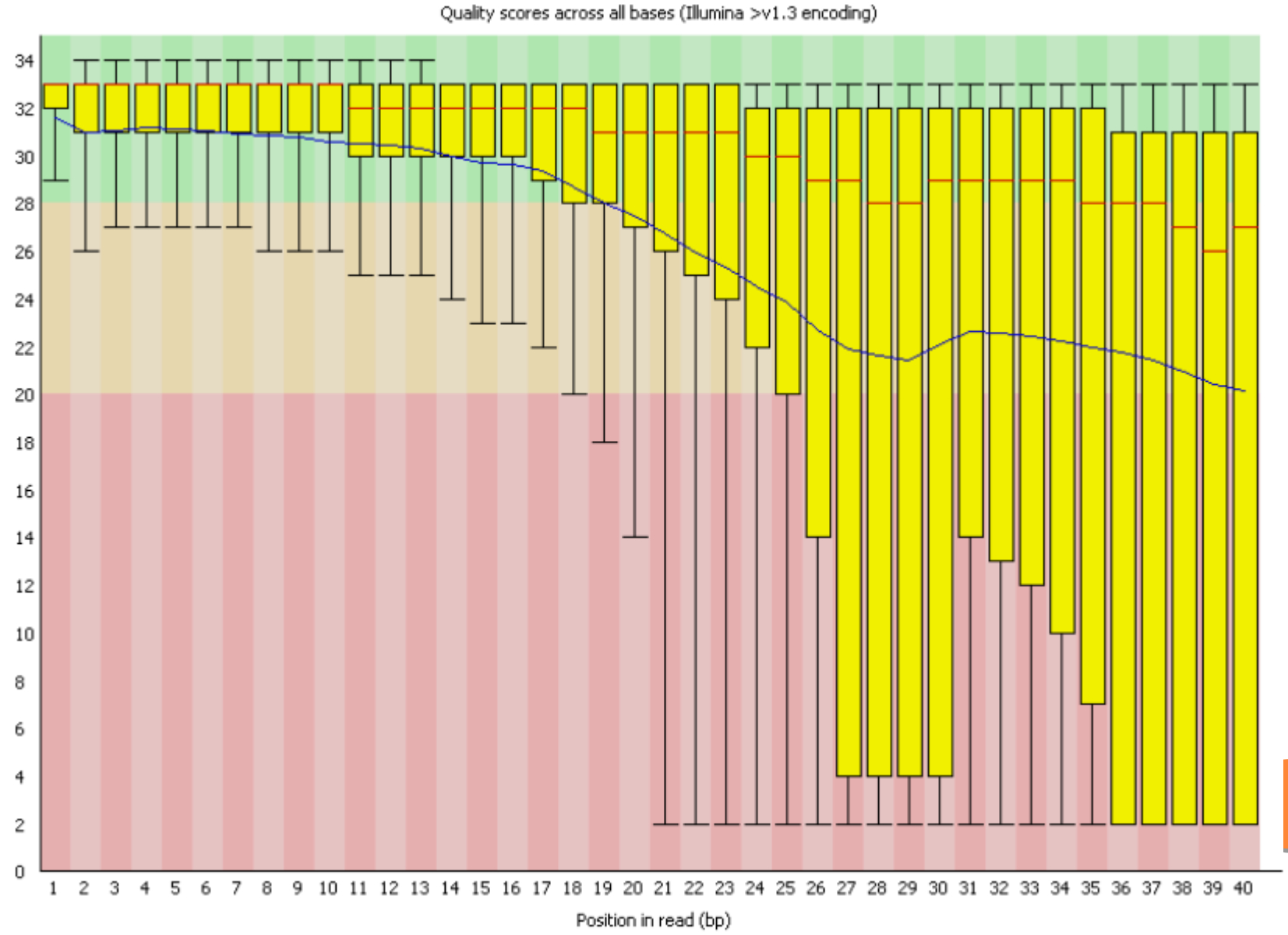


Figure 11. Synthesis of long DNA.

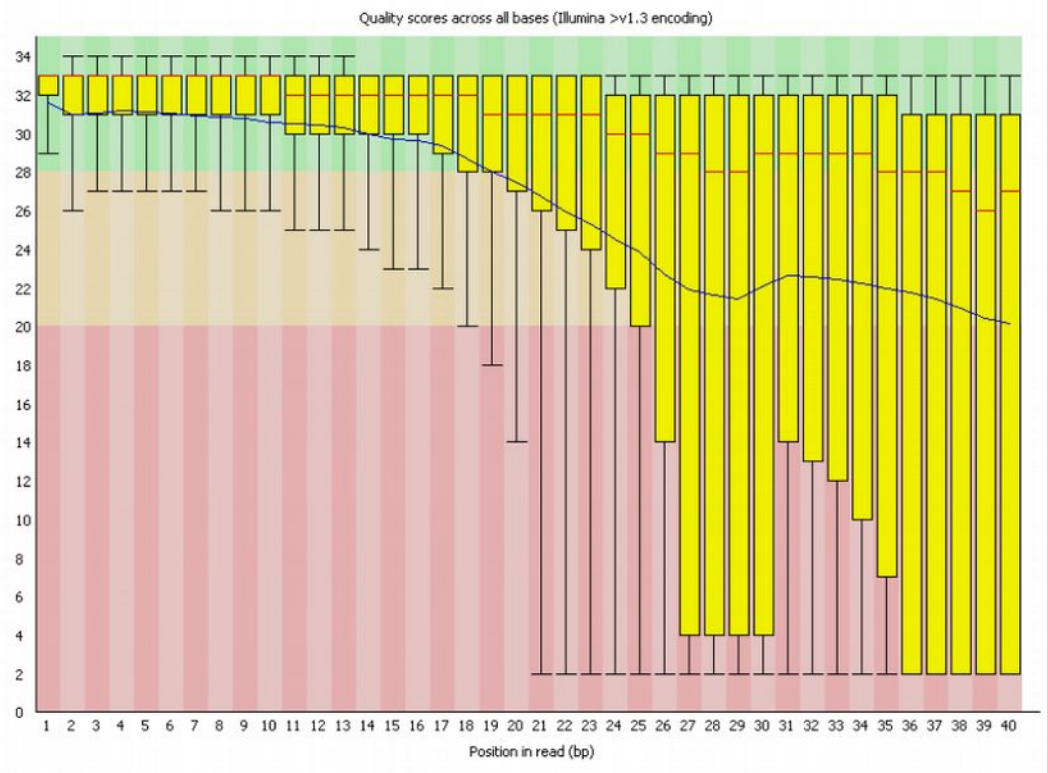
DNA polymerase processively incorporates nucleotides producing long, natural DNA.

https://www.bioinformatics.babraham.ac.uk/projects/fastqc/

FastQC



FastQC

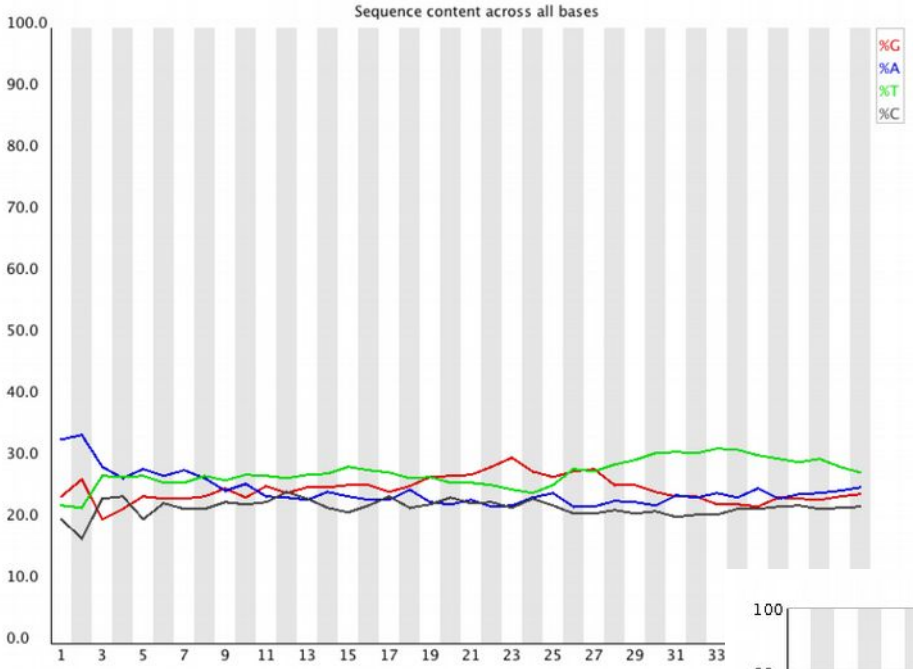


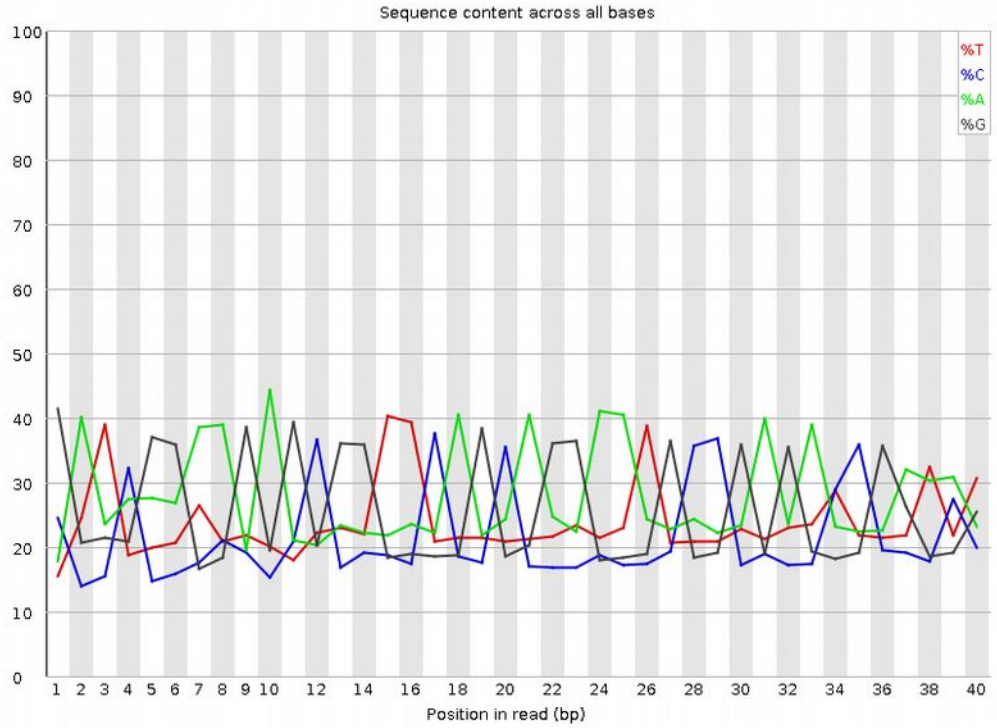
The central red line is the median value

The yellow box represents the inter-quartile range (25-75%)

The upper and lower whiskers represent the 10% and 90% points

The blue line represents the mean quality


Sequence content across all bases
%G
%A
%T
%C
100.0
90.0
80.0
70.0
60.0
50.0
40.0
30.0
20.0
10.0
0.0
1 3 5 7 9 11 13 15 17 19 21 23 25 27 29 31 33


Sequence content across all bases
%T
%C
%A
%G
100
90
80
70
60
50
40
30
20
10
0
1 2 3 4 5 6 7 8 9 10 12 14 16 18 20 22 24 26 28 30 32 34 36 38 40
Position in read (bp)

Third Generation Sequencing: Single Molecule Sequencing

Oxford Nanopore

SmidgION

MinION

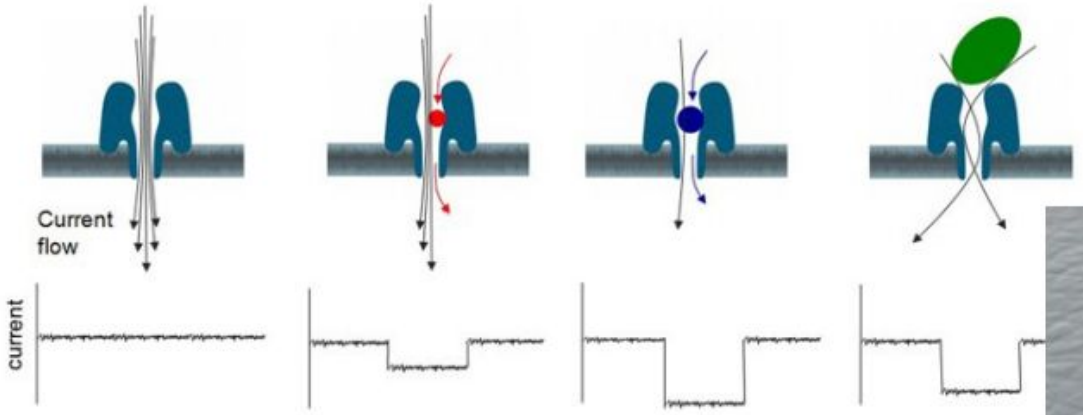
GridION X5

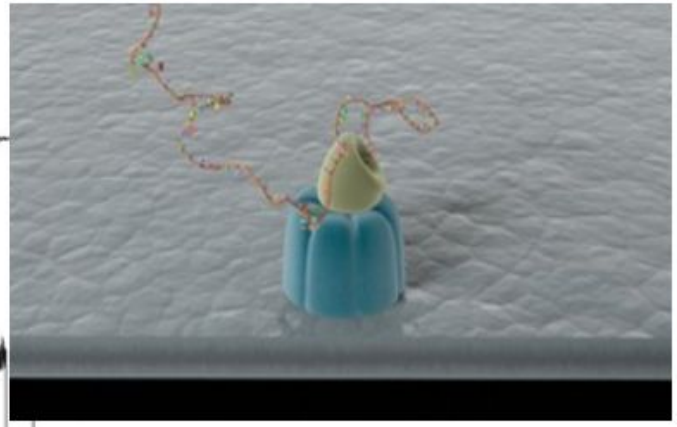
PromethION

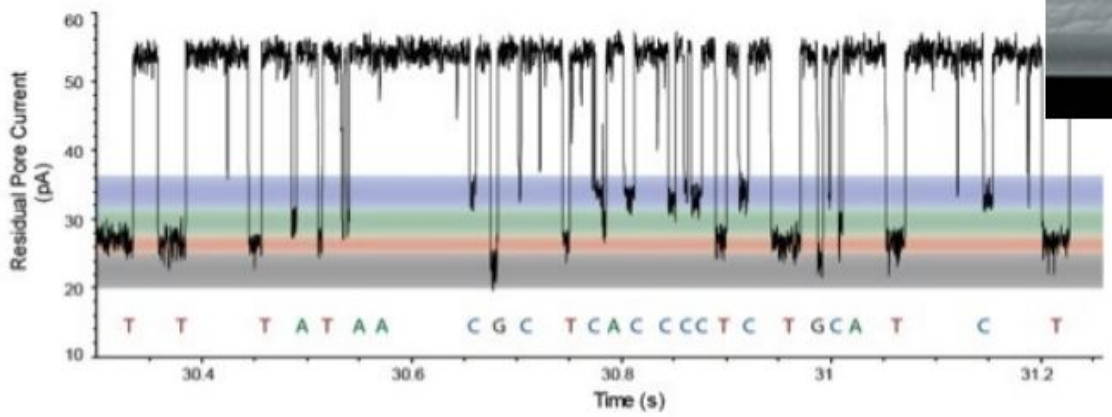
	SmidgION	MinION	GridION X5	PromethION
Read Length	?	> 200 kb		
Throughput	1 Gb (1 flow cell with 256 pores) ?	10-20 Gb (1 flow cell with 512 pores)	100 Gb (5 flow cells with 512 pores/cell) 2560 pores	50 – 250 Gb per flow cell/48hours ? (48 flow cells, 3000 pores/cell) 144,000 pores
Reads per run	?	10,000 – >300,000		
Accuracy		90 % (1D) – 96 %($1D^2$)		
Run Time	1 – 4 hours	1 - 48 (70) hours	1 – 48 hours	1 - 48 hours

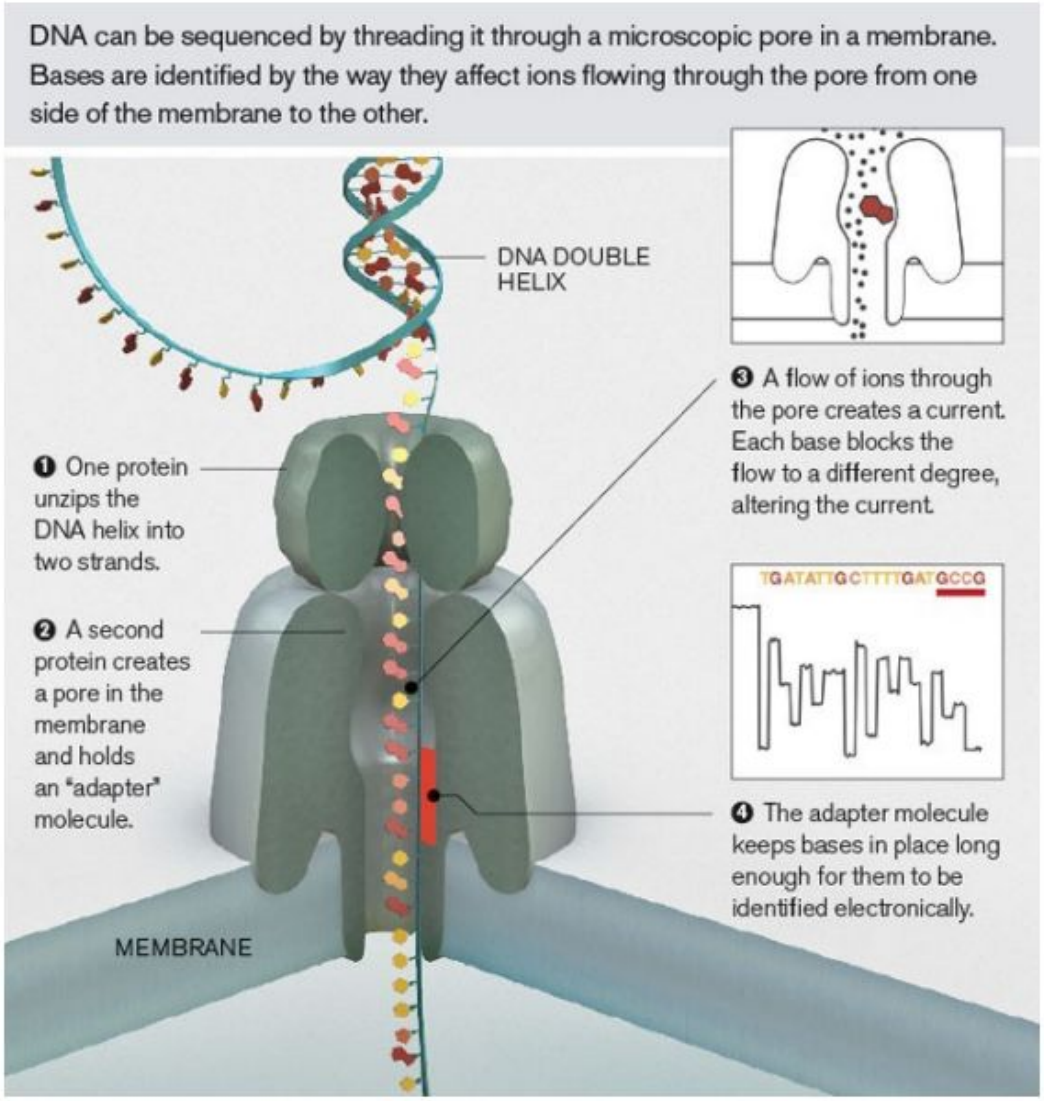
consensus accuracy improved to 99.5% at 30× coverage

Oxford Nanopore









R9.4: 450 nucleotides/second

VolTRAX

Rapid, programmable, portable, disposable sample processor

The Journal of Molecular Diagnostics, Vol. 20, No. 4, July 2018

the Journal of Molecular Diagnostics
jmd.amjpathol.org

Accurate Typing of Human Leukocyte Antigen Class I Genes by Oxford Nanopore Sequencing

Chang Liu,* Fangzhou Xiao,† Jessica Hoisington-Lopez,‡ Kathrin Lang,§ Philipp Quenzel,§ Brian Duffy,¶ and Robi D. Mitra‡

From the Department of Pathology and Immunology, Division of Laboratory and Genomic Medicine, and the Department of Genetics,‡ Center for Genome Sciences and Systems Biology, Washington University School of Medicine, St. Louis, Missouri; the Division of Biology and Biological Engineering,† California Institute of Technology, Pasadena, California; the DKMS Life Science Lab GmbH,§ Dresden, Germany; and the HLA Laboratory,¶ Barnes-Jewish Hospital, St. Louis, Missouri

Accepted for publication February 13, 2018.

Address correspondence to Robi D. Mitra, Ph.D., Department of Genetics, Center for Genome Sciences & Systems Biology, or Chang Liu, M.D., Ph.D., Division of Laboratory and Genomic Medicine, Department of Pathology and Immunology, Washington University School of Medicine, St. Louis, MO 63110. E-mail: rmitra@genetics.wustl.edu or cliu32@wustl.edu.

Oxford Nanopore Technologies' MinION has expanded the current DNA sequencing toolkit by delivering long read lengths and extreme portability. The MinION has the potential to enable expedited point-of-care human leukocyte antigen (HLA) typing, an assay routinely used to assess the immunologic compatibility between organ donors and recipients, but the platform's high error rate makes it challenging to type alleles with accuracy. We developed and validated accurate typing of HLA by Oxford nanopore (Athlon), a bioinformatic pipeline that i) maps nanopore reads to a database of known HLA alleles, ii) identifies candidate alleles with the highest read coverage at different resolution levels that are represented as branching nodes and leaves of a tree structure, iii) generates consensus sequences by remapping the reads to the candidate alleles, and iv) calls the final diploid genotype by blasting consensus sequences against the reference database. Using two independent data sets generated on the R9.4 flow cell chemistry, Athlon achieved a 100% accuracy in class I HLA typing at the two-field resolution. *(J Mol Diagn 2018, 20: 428–435; https://doi.org/10.1016/j.jmoldx.2018.02.006)*