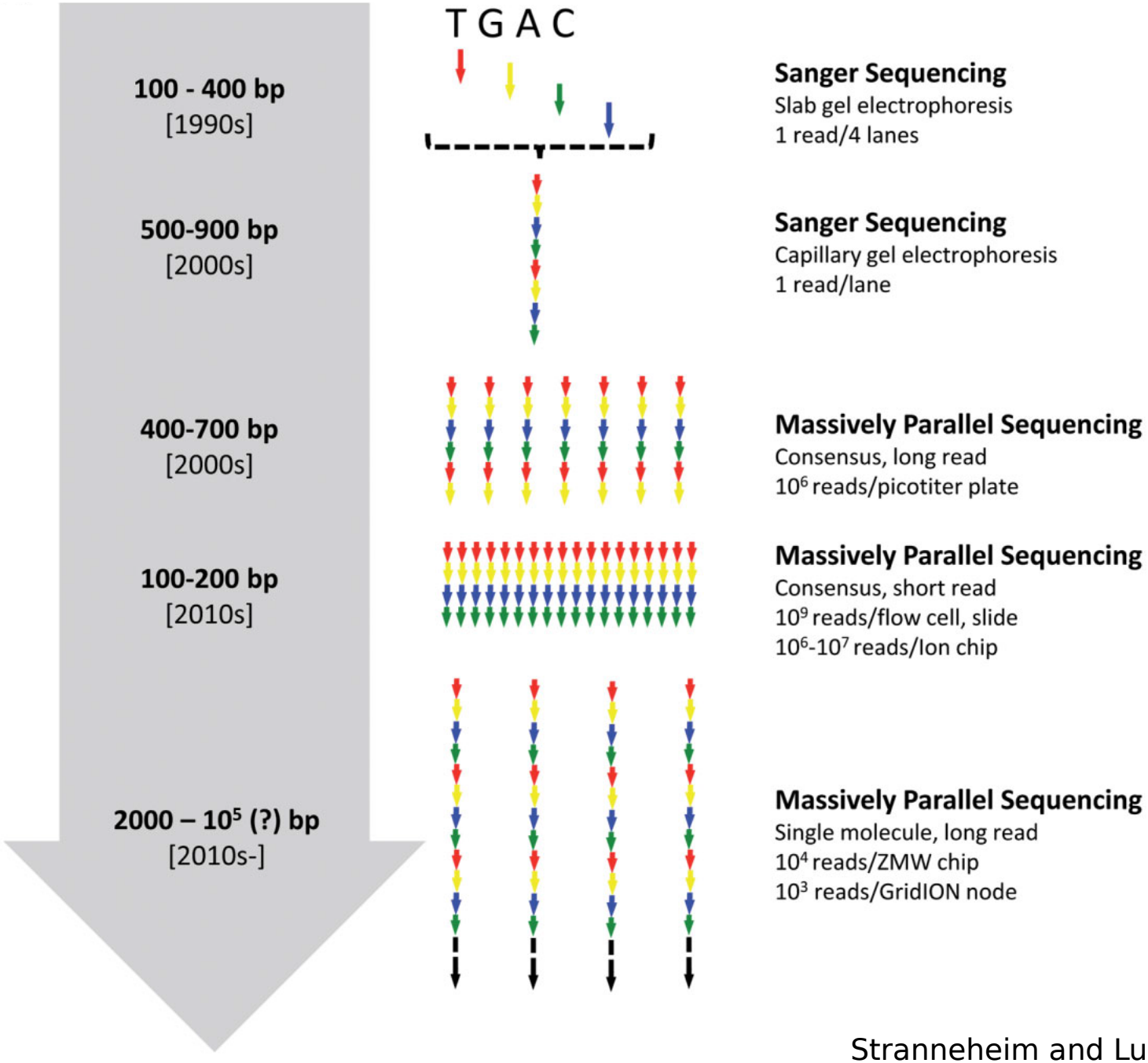
Lezione 4: Metodologie di sequenziamento di nuova generazione (Illumina)



Stranneheim and Lundeberg 201

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.

(4000)

- 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.

Perchè bioinformatica

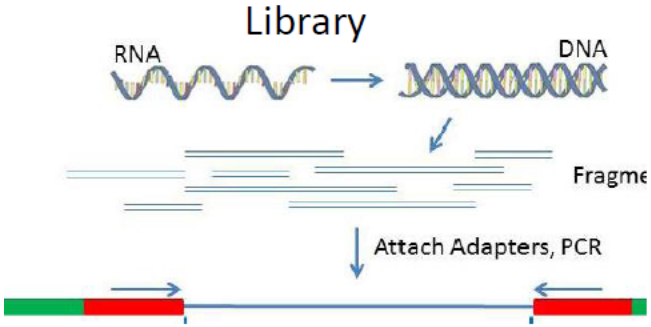
(wall.hms.harvard.edu)

Bioinformatica: sfide in NGS Data Analysis

- File di testo MOLTO GRANDI (migliaia di milioni di righe)
  - Non si possono usare gli strumenti “soliti”
  - Enorme utilizzo di memoria e tempi di corsa
  - Gestire, analizzare, accumulare, trasferire ed archiviare file giganteschi
- Necessità di computer potenti e di competenze
  - Computer clusters
  - Necessità di nuovi algoritmi e software spesso open source Unix/Linux based.
  - Collaborazione tra chi sviluppa la tecnologia, i bioinformatici e i biologi

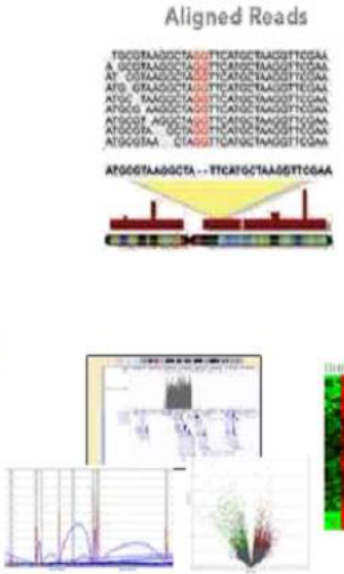
Basic NGS Workflow

Samples preparation



Application specific

Sequencing



Data Analysis

Application Specific

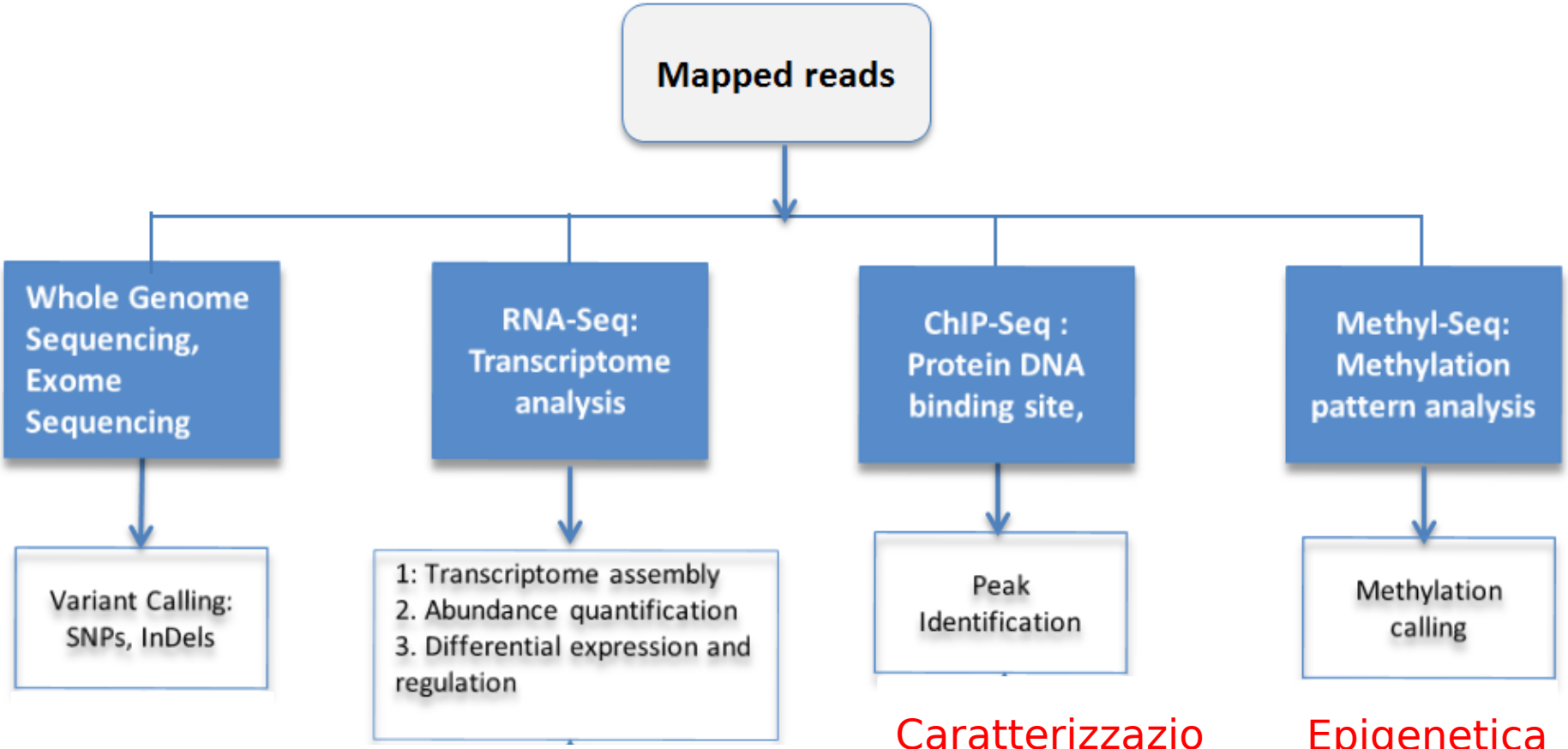
https://www.youtube.com/watch?v=jFCD8Q6qSTM

Olson *et al.*

Terminology

- Coverage (depth): the number of times during the sequencing process ide is read
- 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

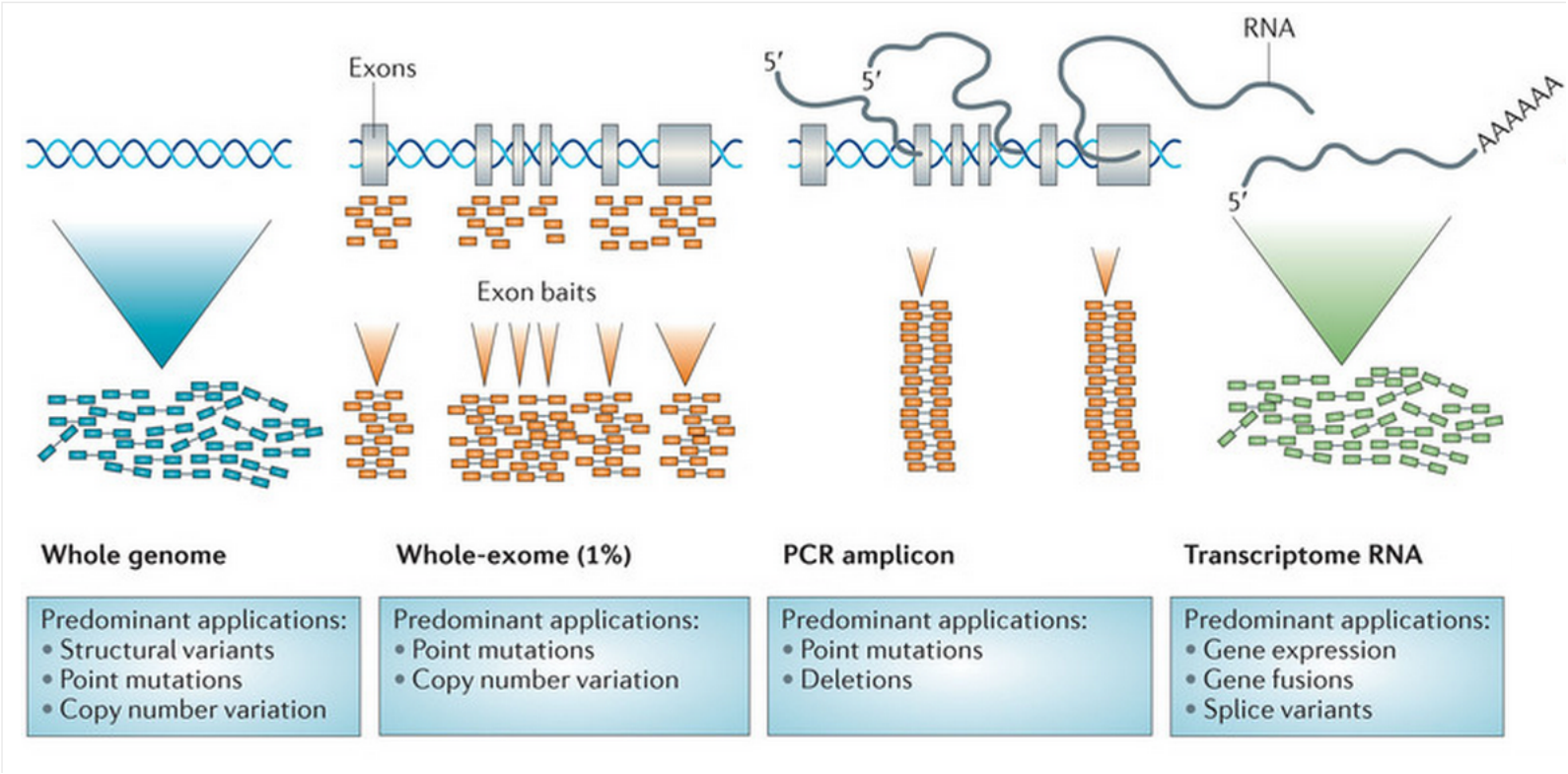


Espressione genica

Caratterizzazione regioni di interazione DNA-proteine

Epigenetica

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

library	• Frammentazione • Legame adattatori
Template preparation	• Serve a reggiungere una quantità di DNA stampo sufficiente per la lettura del sequenziamento
sequencing	• Sequencing by synthesis • Lettura del segnale

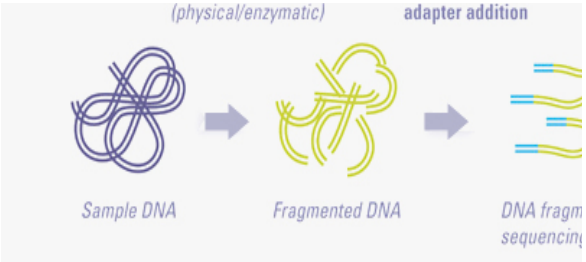
Preparazione della *library*

Preparazione della *library*

Il DNA va pre-processato prima di essere sequenziato

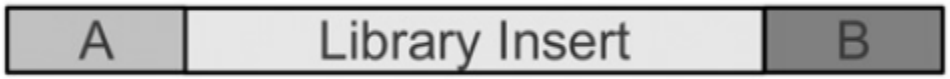
Passaggi:

1. Frammentazione (fisica o enzimatica)



2. Le estremità dei frammenti vengono riparate e rese «blunt» o con specifici nucleotidi terminali (es A o T)

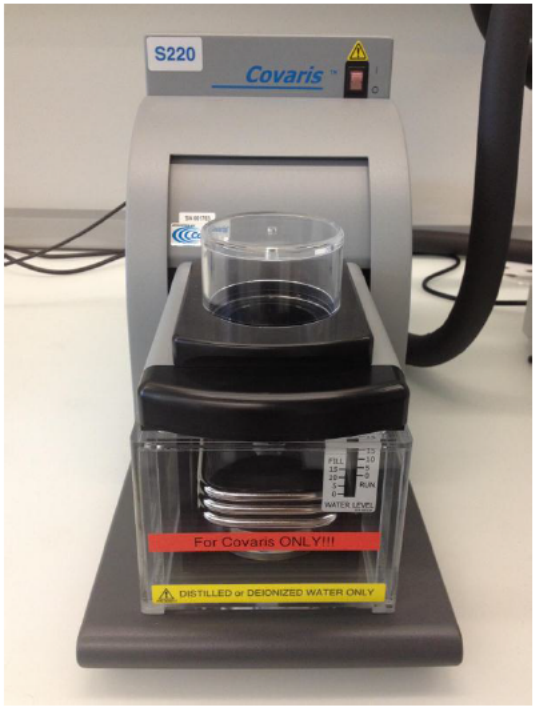
3. Alle estremità vengono legati degli adattatori specifici per ogni NGS technology



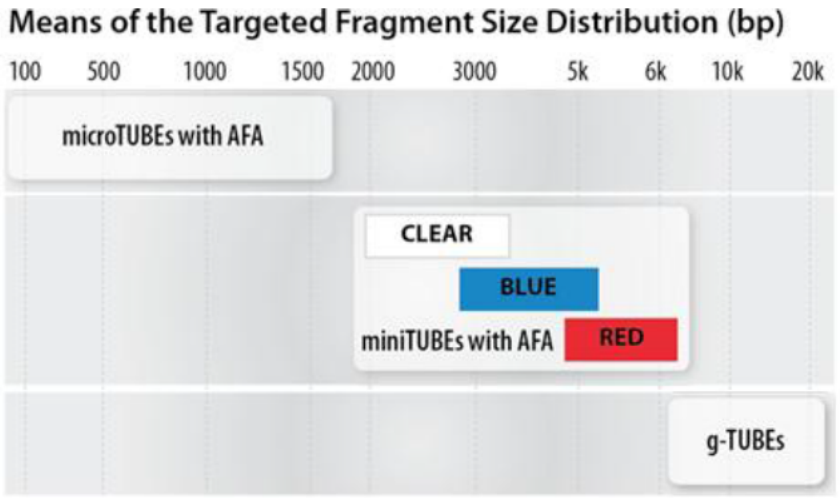
Preparazione della *library*

DNA fragmentation: ultra sonication (COVARIS)

Sonicazione: l'utilizzo di onde acustiche, in particolare ultrasoniche, condotta con l'ausilio di un sonificatore, un apparecchio che genera vibrazioni meccaniche amplificate sfruttando corrente elettrica ad elevata frequenza prodotta da un generatore. Gli ultrasuoni vengono trasmessi in una vasca contenente acqua.

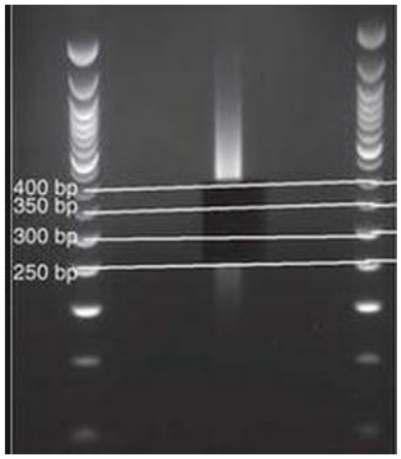


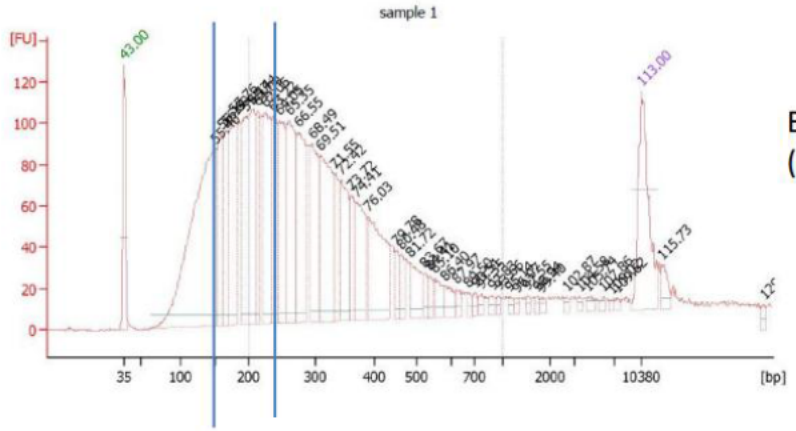
The Covaris process is conducted under isothermal conditions, ensuring the integrity of the nucleic acid sample is maintained and providing high recovery of double-stranded DNA. Combined with the specifically engineered AFA Tubes it is possible to precisely and accurately fragment DNA and RNA to the **100 – 1500bp** range (microTUBE), or **2 – 5kb** range (miniTUBE).



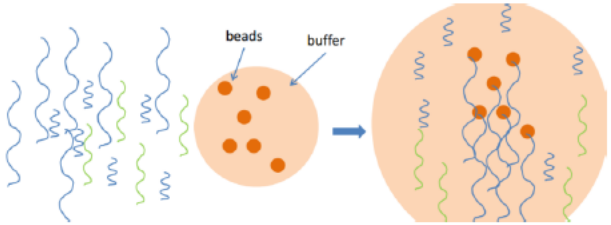
Preparazione della *library*

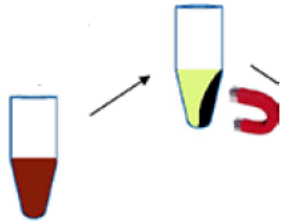
Size selection: gel o beads magnetiche

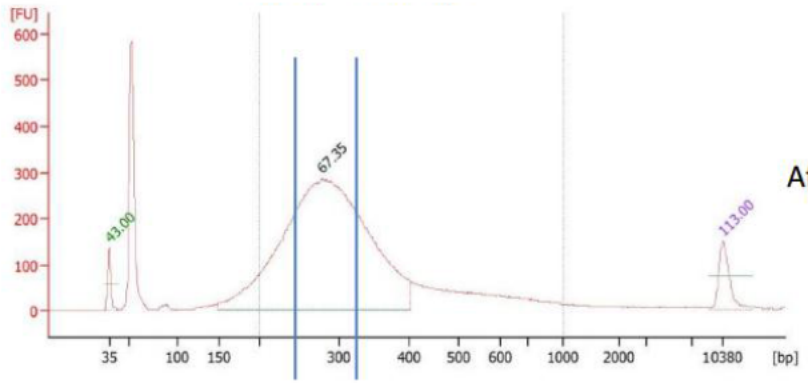




Before size selection
(only shearing)





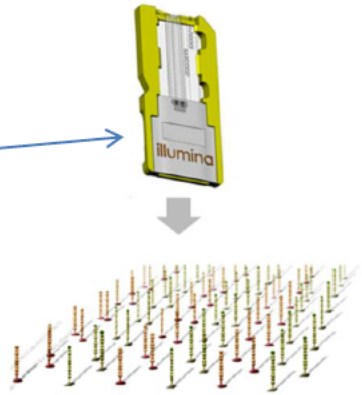


After size selection I

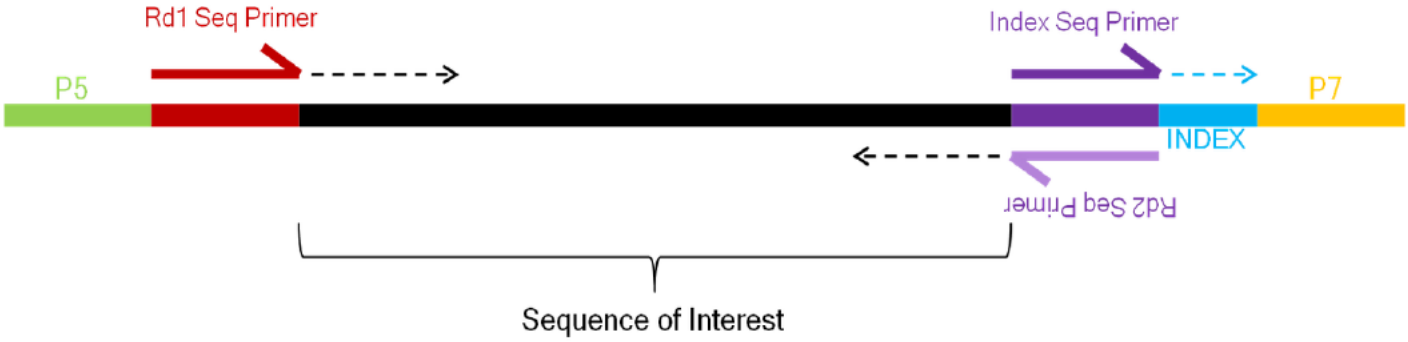
Preparazione della *library:* Illumina

Illumina adaptors

(gli adattatori servono a legare il frammento alla cella dove avviene l'amplificazione e il sequenziamento (P5/P7) e fungono da primer per la reazione di amplificazione prima e di sequenziamento poi)
(Rd1 e Rd2 servono per il paired-end sequencing = sequenziamento a partire da entrambe le estremità di un frammento)



P5/ P7: binding sites to the flow cell
Rd 1 SP: read1 sequencing primer
Rd 2 SP: read2 sequencing primer



http://nextgen.mgh.harvard.edu/IlluminaChemistry.html

Preparazione della *library:* Illumina

Tagmentation (nextera): non rottura meccanica+ligazione adattatori, ma 'tagmentazione' enzimatica + inserzione adattatori

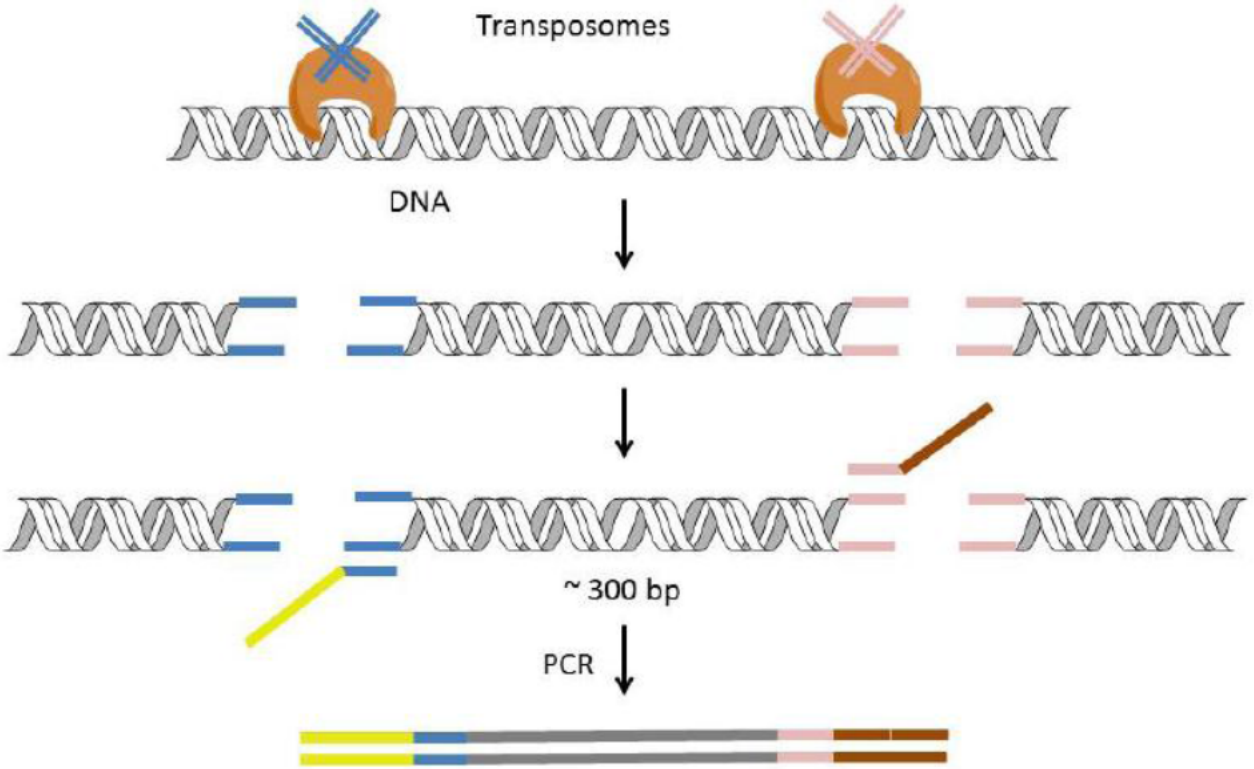


Figure 2. DNA library preparation using a transposase-based method (Nextera) developed by Illumina. The transpososome complex comprises an engineered transposase pre-loaded with two double-stranded sequencing adapters. The transpososome simultaneously fragments the DNA and inserts the adapters. The full Illumina adapter sequences are completed during subsequent PCR cycling, after which the library is ready for quantitation and loading onto the flow cell.

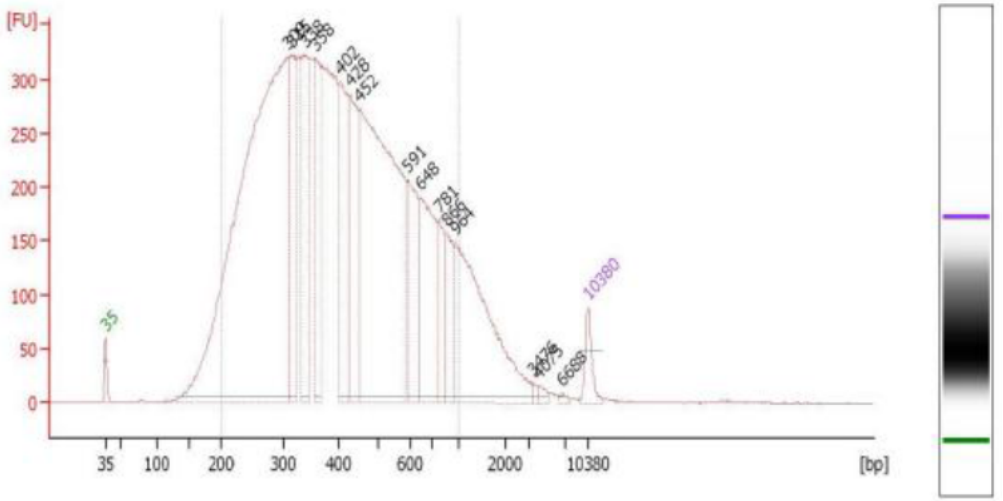
Library quality control and quantitation

Prima del sequenziamento dobbiamo sapere se i frammenti hanno un certo range di lunghezze, se il DNA è di buona qualità e la concentrazione

- Fluorometer (Qubit)

- qPCR (real time PCR usando primer che riconoscono gli adattatori all'estremità dei frammenti della library)

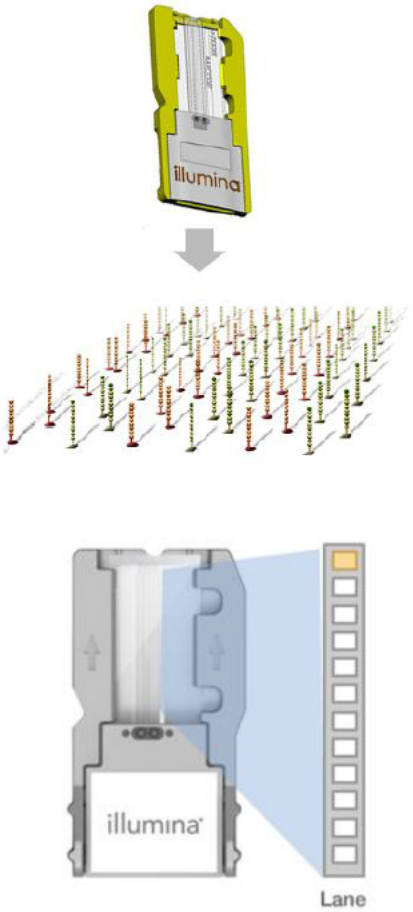
- Agilent Bioanalyser

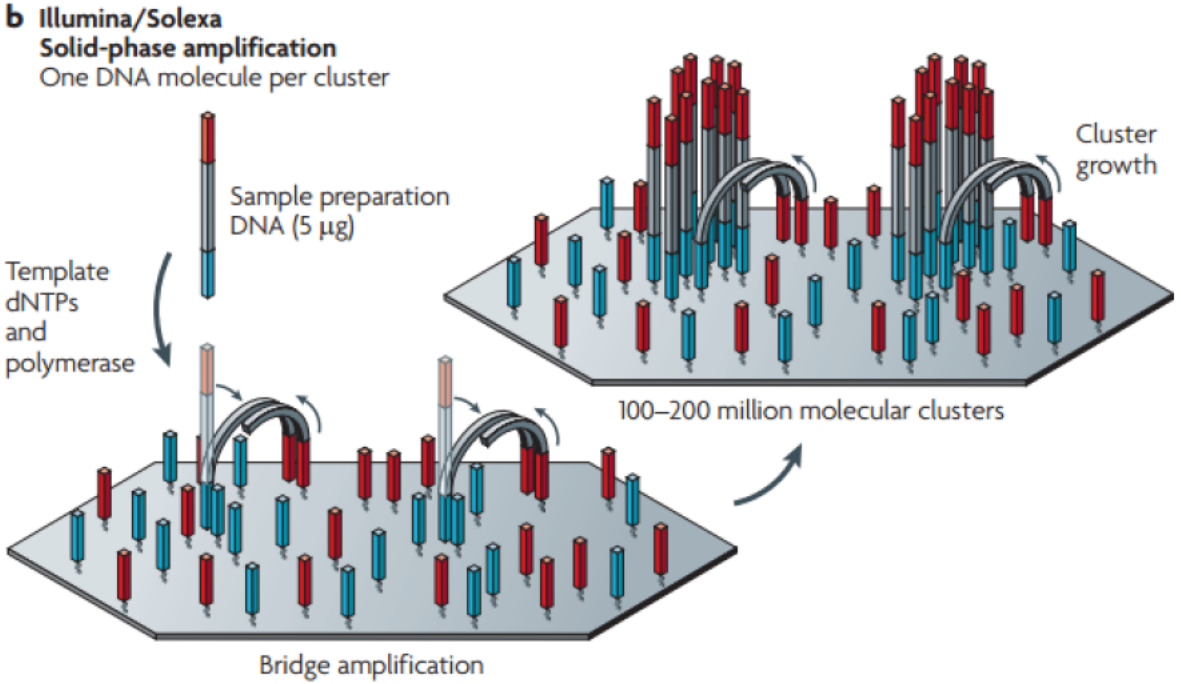


Amplificazione clonale della library
(già nello strumento per NGS!)

Metodologie principali

- Solid-phase cluster generation (Illumina)

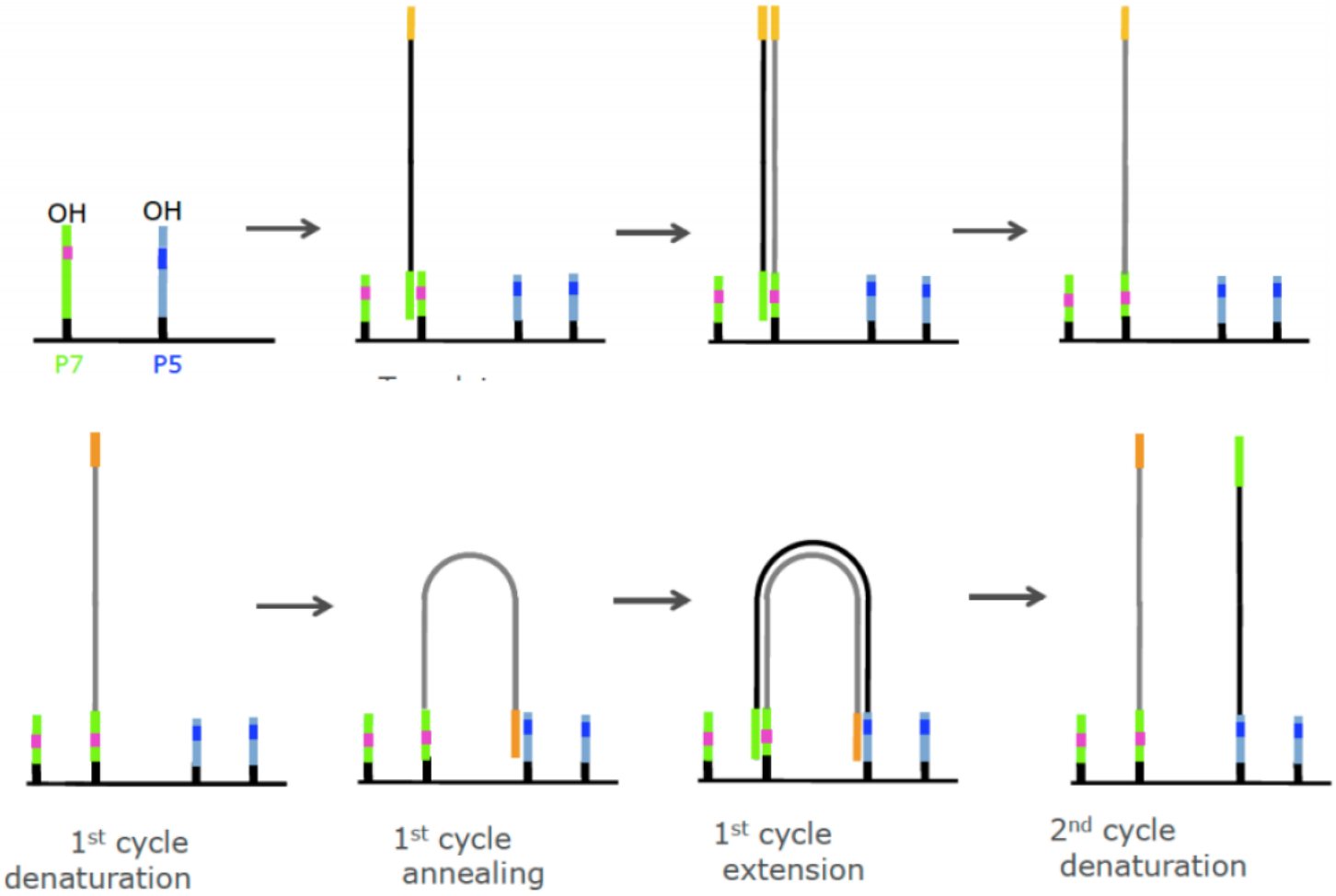


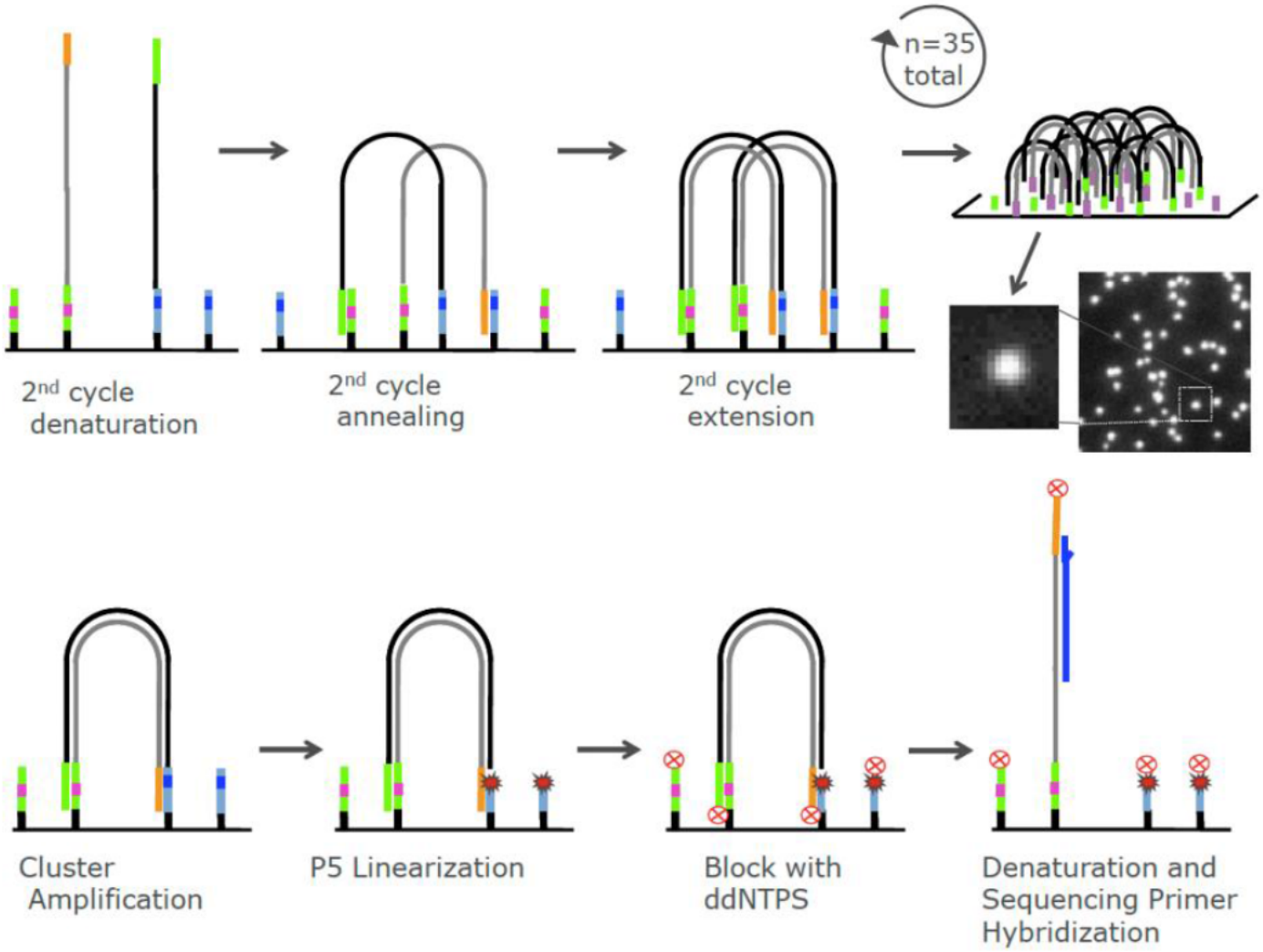


1,000 copies of each fragment are generated by bridge amplification, creating clusters (30-50 million of clusters).

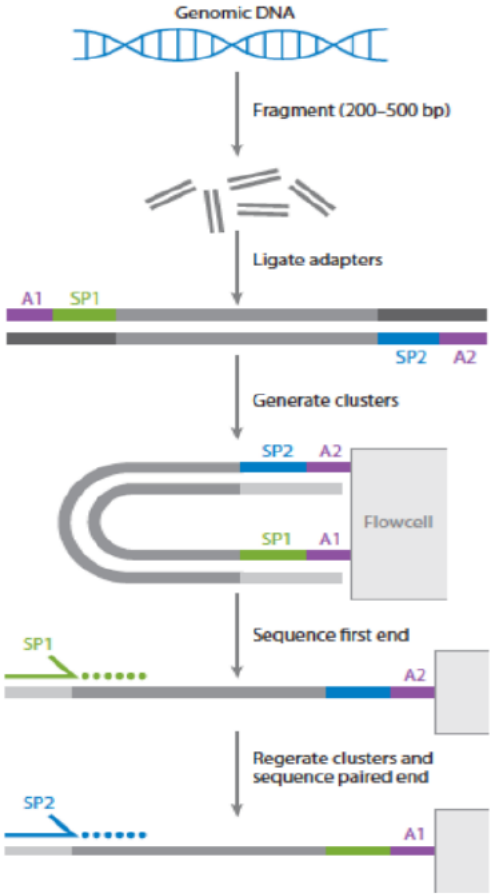
https://www.youtube.com/watch?v=ZlPH9FzUOq8

Gli adattatori P5 e P7 servono per legare il frammento da sequenziare alla flow cell dove ci sono delle sequenze complementari e servono anche da primer




n=35
total
2nd cycle
denaturation
2nd cycle
annealing
2nd cycle
extension
Cluster
Amplification
P5 Linearization
Block with
ddNTPS
Denaturation and
Sequencing Primer
Hybridization

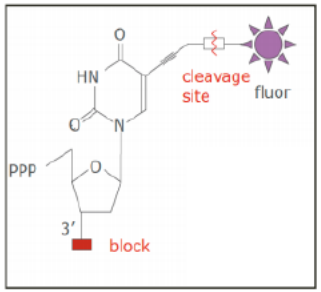
Illumina: summary



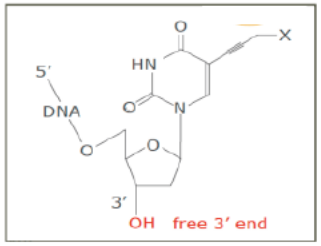
- Shear high molecular weight DNA with sonication
- Polish ends
- Ligate synthetic DNA adapters
- Produce size fractions
- Quantitate
- Amplify library fragments on flow cell surface
- Denature clusters to single-stranded
- Hybridize sequencing primer to linearized ss cluster DNAs
- Proceed to sequencing or hybrid capture

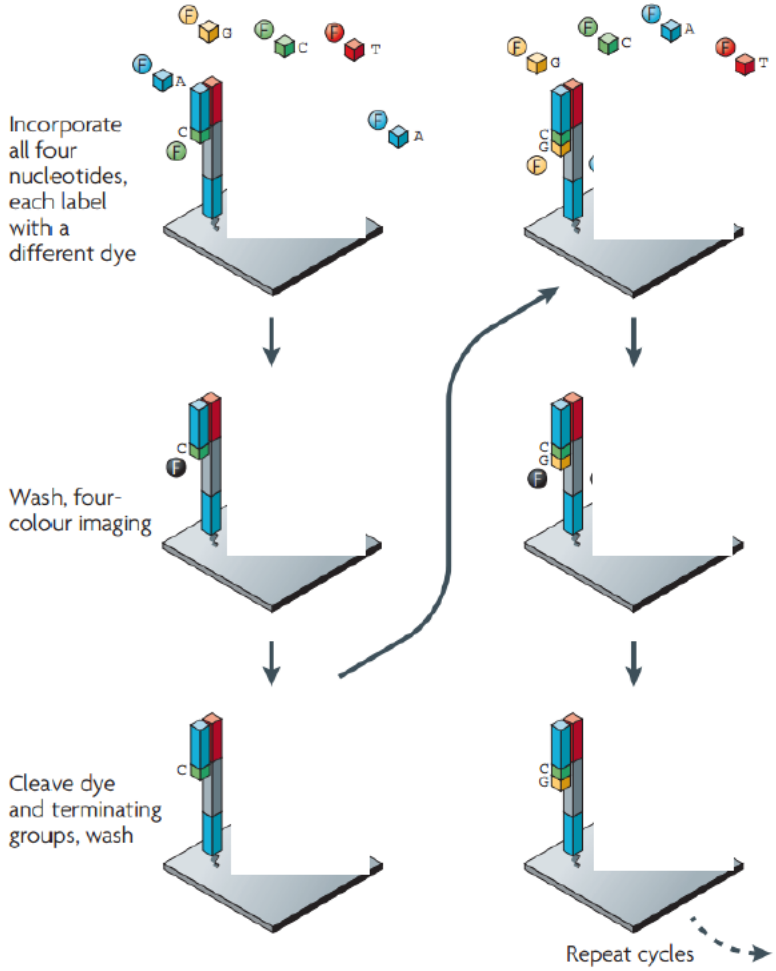
Sequenziamento e rilevazione del segnale

Reversible terminators (Illumina)



Incorporate
Detect
De-block
Cleave fluor





Mod. from Metzker 2010

Platforms: Illumina

MiSeq

NextSeq 500

HiSeq 2500

HiSeq X*

Key applications	Small genome, amplicon, and targeted gene panel sequencing.	Everyday genome, exome, transcriptome sequencing, and more.		Production-scale genome, exome, transcriptome sequencing, and more.		Population-scale human whole-genome sequencing.
Run mode	N/A	Mid-Output	High-Output	Rapid Run	High-Output	N/A
Flow cells processed per run	1	1	1	1 or 2	1 or 2	1 or 2
Output range	0.3-15 Gb	20-39 Gb	30-120 Gb	10-180 Gb	50-1000 Gb	1.6-1.8 Tb
Run time	5-65 hours	15-26 hours	12-30 hours	7-40 hours	< 1 day - 6 days	< 3 days
Reads per flow cell†	25 Million‡	130 Million	400 Million	300 Million	2 Billion	3 Billion
Maximum read length	2 × 300 bp	2 × 150 bp	2 × 150 bp	2 × 150 bp	2 × 125 bp	2 × 150 bp

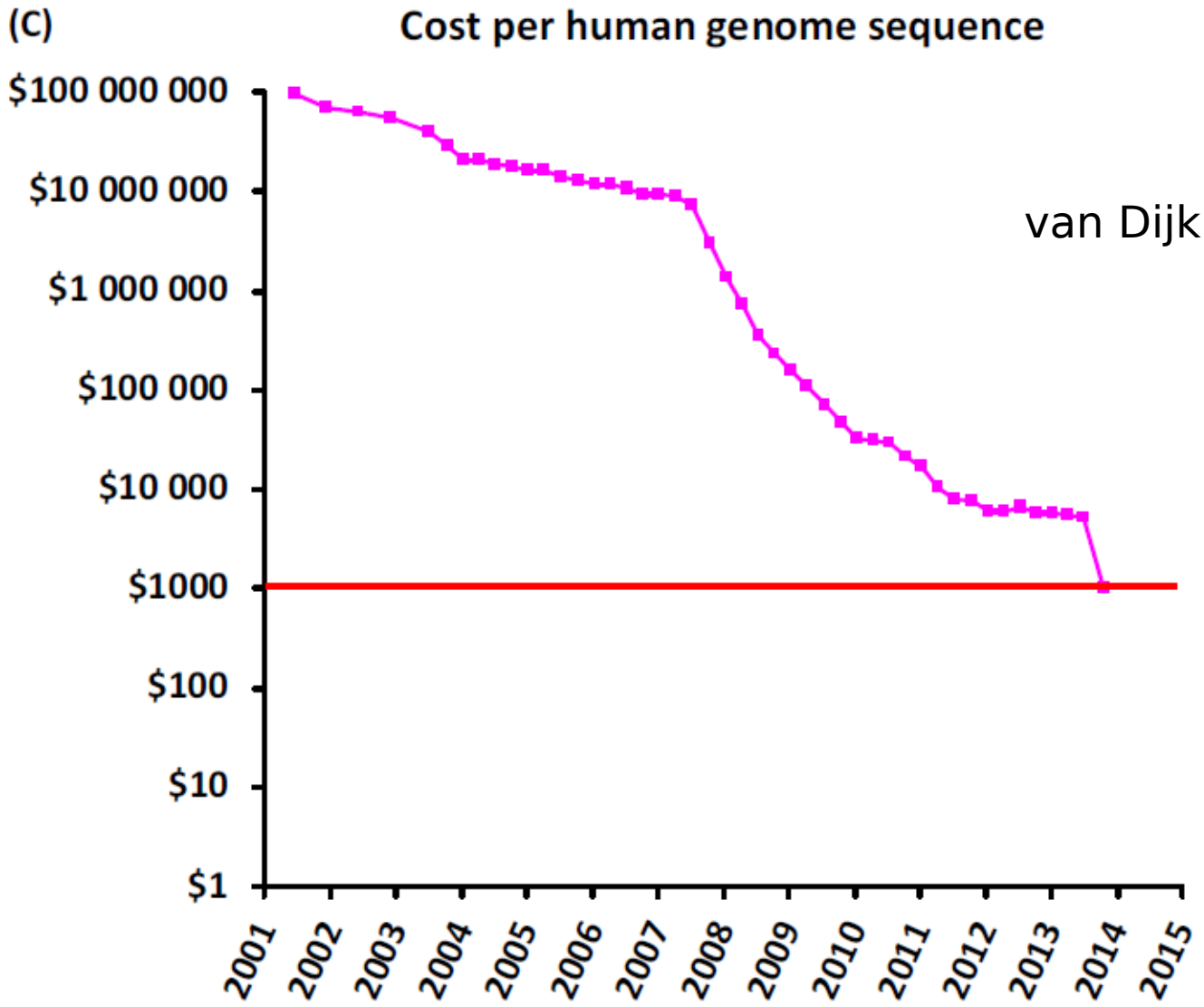
- High accuracy, range of capacity and throughput
- Longer read lengths on some platforms (MiSeq)
- Improved kits, improved software pipeline and capabilities, cloud compute

Summary

(NB: i numeri sono indicativi, dipendono molto dalle condizioni che si usano per il sequenziamento)

		Detection	Run time	Read lenght (bp)	Reads per run	Output per run
Roche	FLX	Pyrophos release	23 h	700	1 million	700 Mb
	Junior	Pyrophos release	10 h	400	0.1 million	40 Mb
Life Technologies	Ion Torrent	H+ release	4 h	200-400	4 million	1.5-2 Gb
	Proton	H+ release	4 h	125	60-80 million	8-10 Gb
Illumina/Solexa	Hiseq2000/2500	Fluorescence: reversible terminators	12 days	2 x 100	3 billion	600 Gb
	MiSeq		65 h	2 x 300	25 million	15 Gb

tp://www.molecularecologist.com/next-gen-fieldguide-2014/



van Dijk et al. TIG 2014

Graph showing the evolution of the cost of sequencing a human genome from 2001 until today. Costs have sharply decreased over the recent years thanks to the appearance of next-generation sequencing (NGS) technologies and their subsequent upgrades. Very recently, the milestone of the US$1000 genome has been reached with Illumina's HiSeq X Ten system

Per una guida comparata alle piattaforme

Home How to … NGS Field Guide 2014 About Legal Info News

2014 NGS Field Guide: Overview

Search

- http://www.molecularecologist.com/next-gen-fieldguide-2014/