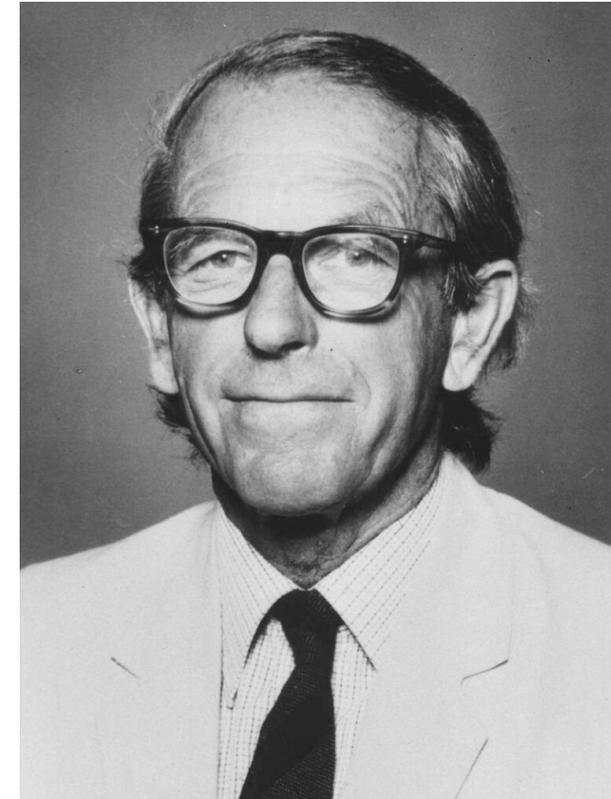
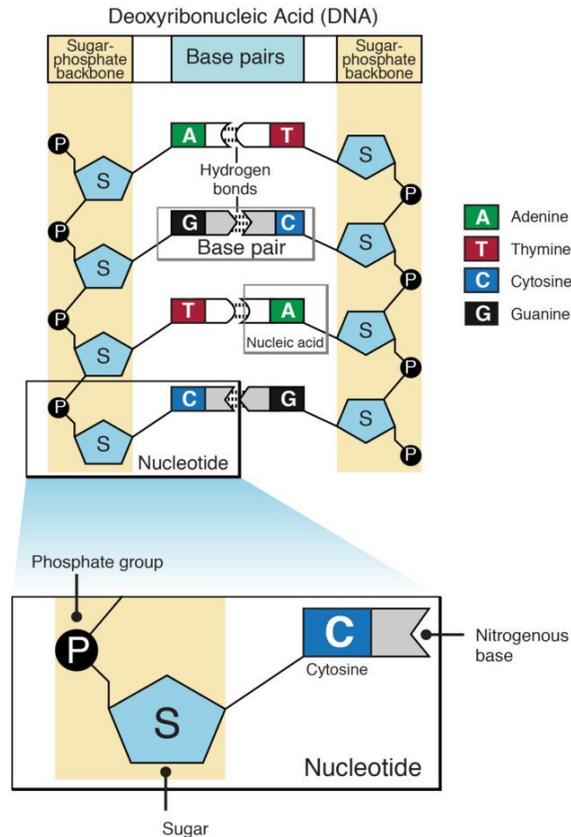


**Sequenziamento  
massivo degli acidi  
nucleici**

# Cos'è il sequenziamento ?

Sequenziare significa determinare l'ordine esatto dei monomeri che formano una biomolecola.

Nel caso del DNA, il sequenziamento permette di ricostruire l'ordine dei nucleotidi che formano una molecola di acido desossiribonucleico.

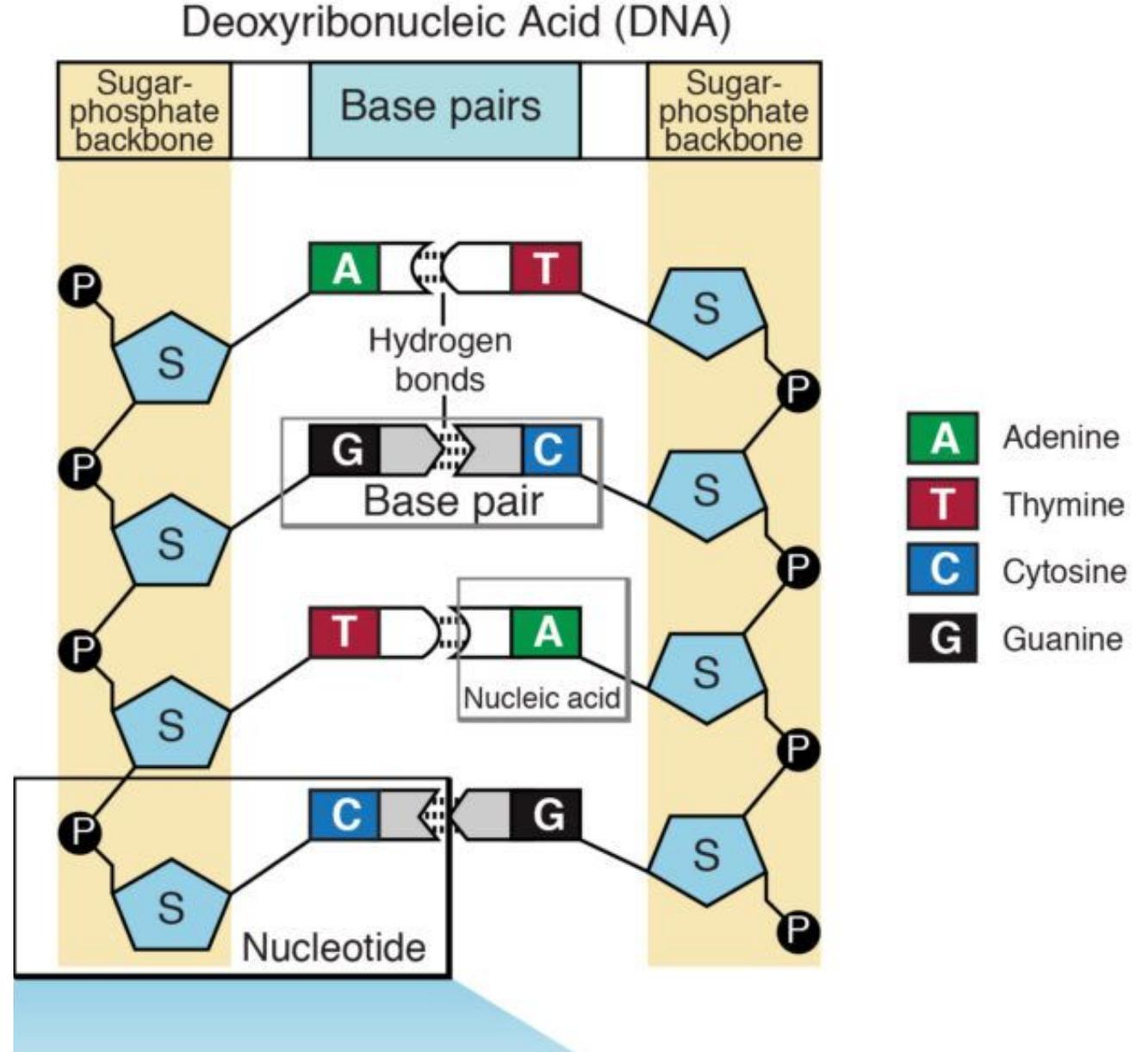
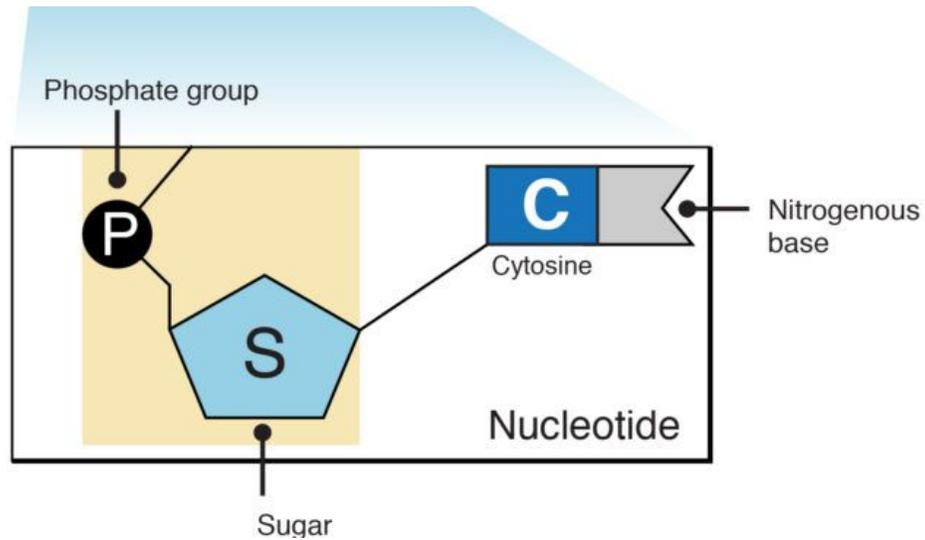


Frederick Sanger ha ideato un metodo di sequenziamento in grado di sequenziare gli acidi nucleici, dai più semplici ai più complessi. La tecnica è conosciuta come **sequenziamento Sanger** (o *metodo della terminazione della catena*).

# Metodi di prima generazione: metodo Sanger

## Caratteristiche generali

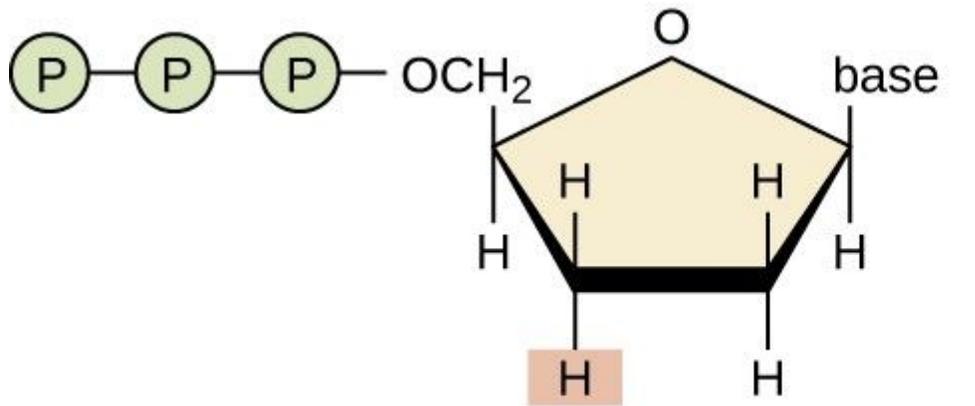
Per poter comprendere al meglio il funzionamento della tecnica di Sanger è necessario ripassare in breve qualche concetto di biochimica sulla struttura del DNA.



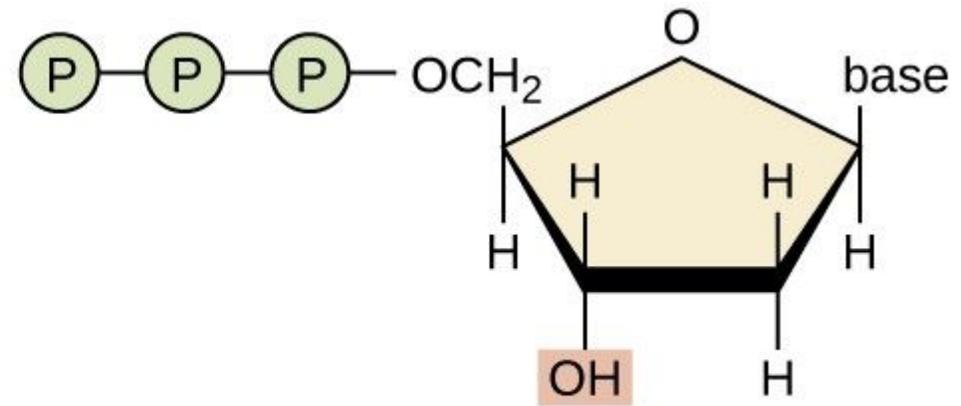
# Metodi di prima generazione: metodo Sanger

## Funzionamento

Il metodo di Sanger o **metodo della terminazione della catena** consiste nell'utilizzo di **dideossiribonucleotidi**; quest'ultimi sono privi del **gruppo ossidrilico** in posizione 3'.



dideoxynucleotide (ddNTP)



deoxynucleotide (dNTP)

# Metodi di prima generazione: metodo Sanger

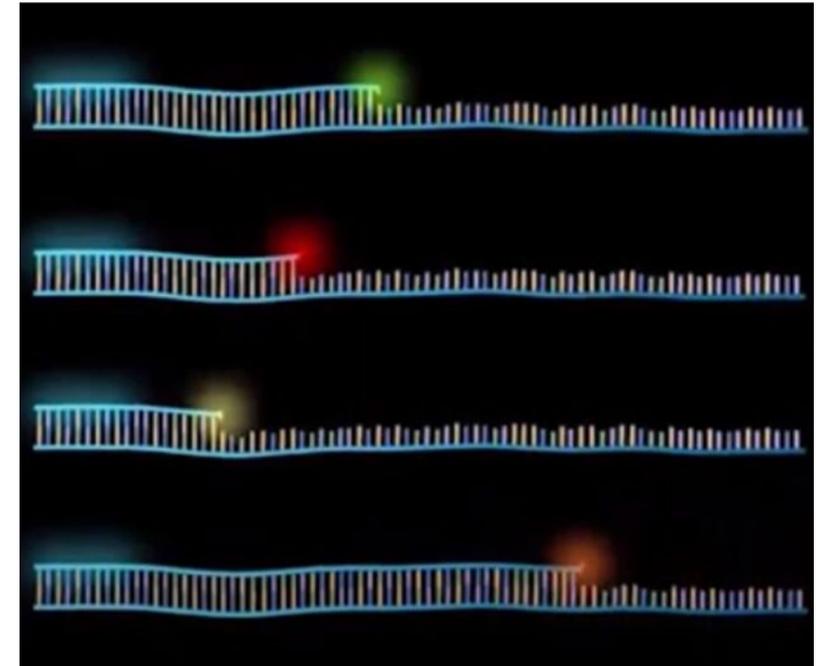
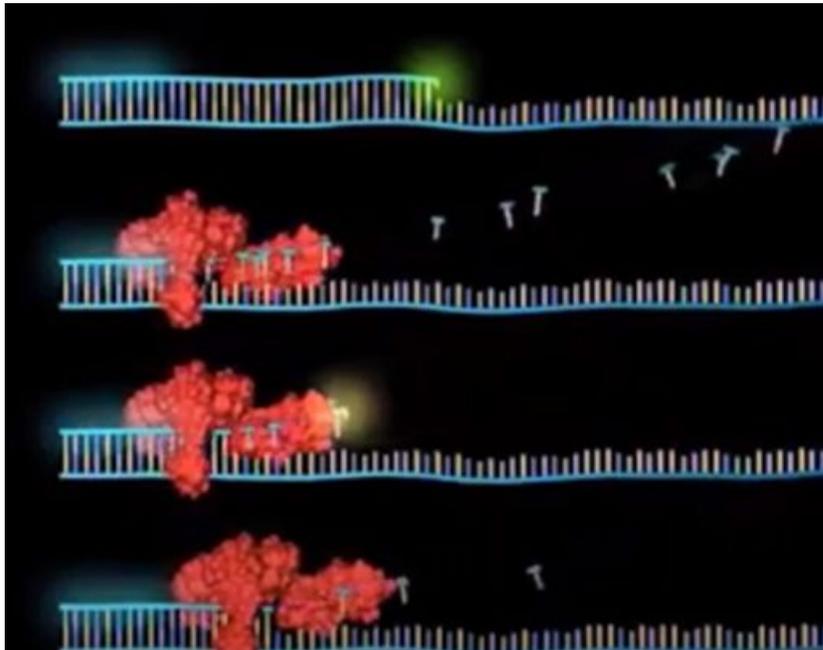
## Procedura

Preparazione del campione

1. Amplificazione del campione -> PCR
2. Denaturazione delle copie amplificate -> mediante aumento della temperatura
3. Aggiunta di DNA polimerasi, Primer, dNTP e ddNTP

Reazione di sequenziamento

4. Sintesi di un nuovo filamento di DNA

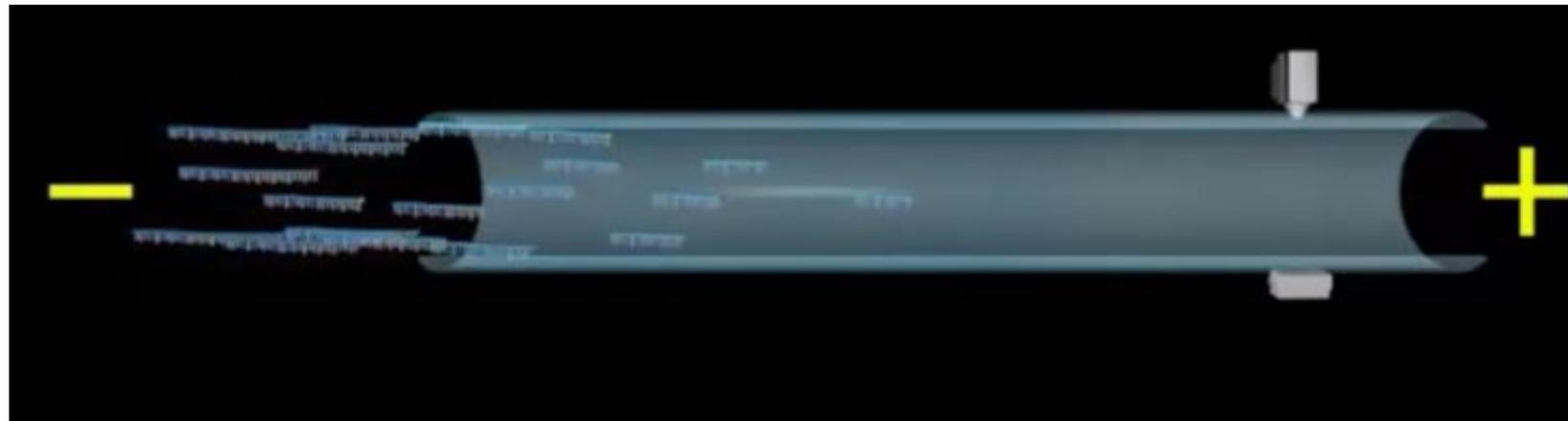
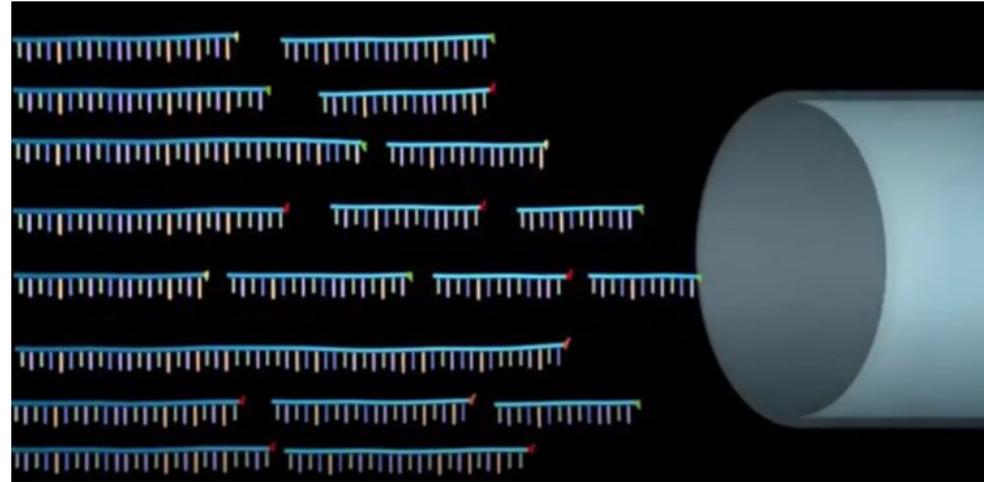


## Metodi di prima generazione: metodo Sanger

Elettroforesi capillare:

5. Si denatura nuovamente il DNA

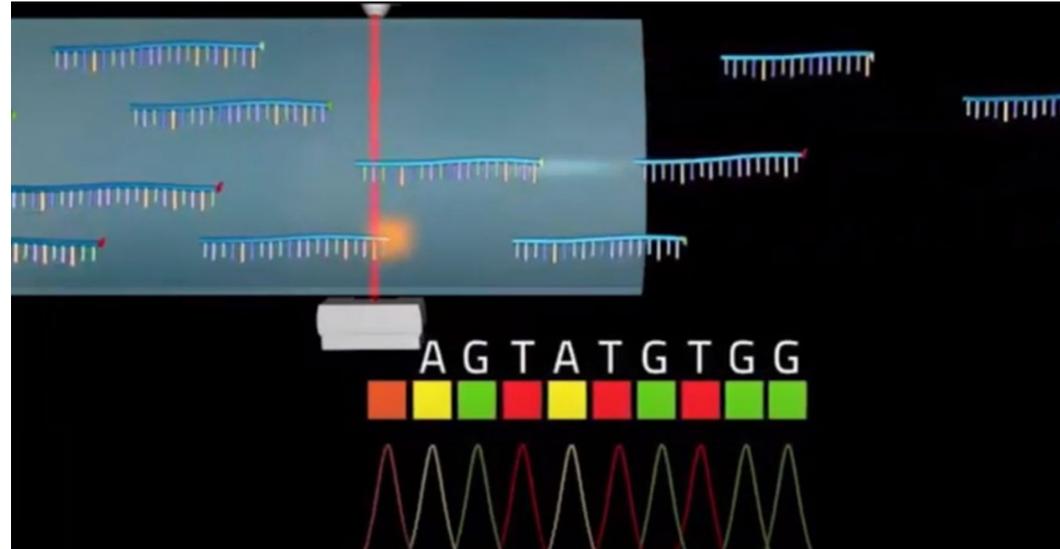
6. I singoli filamenti vengono sottoposti ad elettroforesi capillare



## Metodi di prima generazione: metodo Sanger

Lettura della corsa elettroforetica:

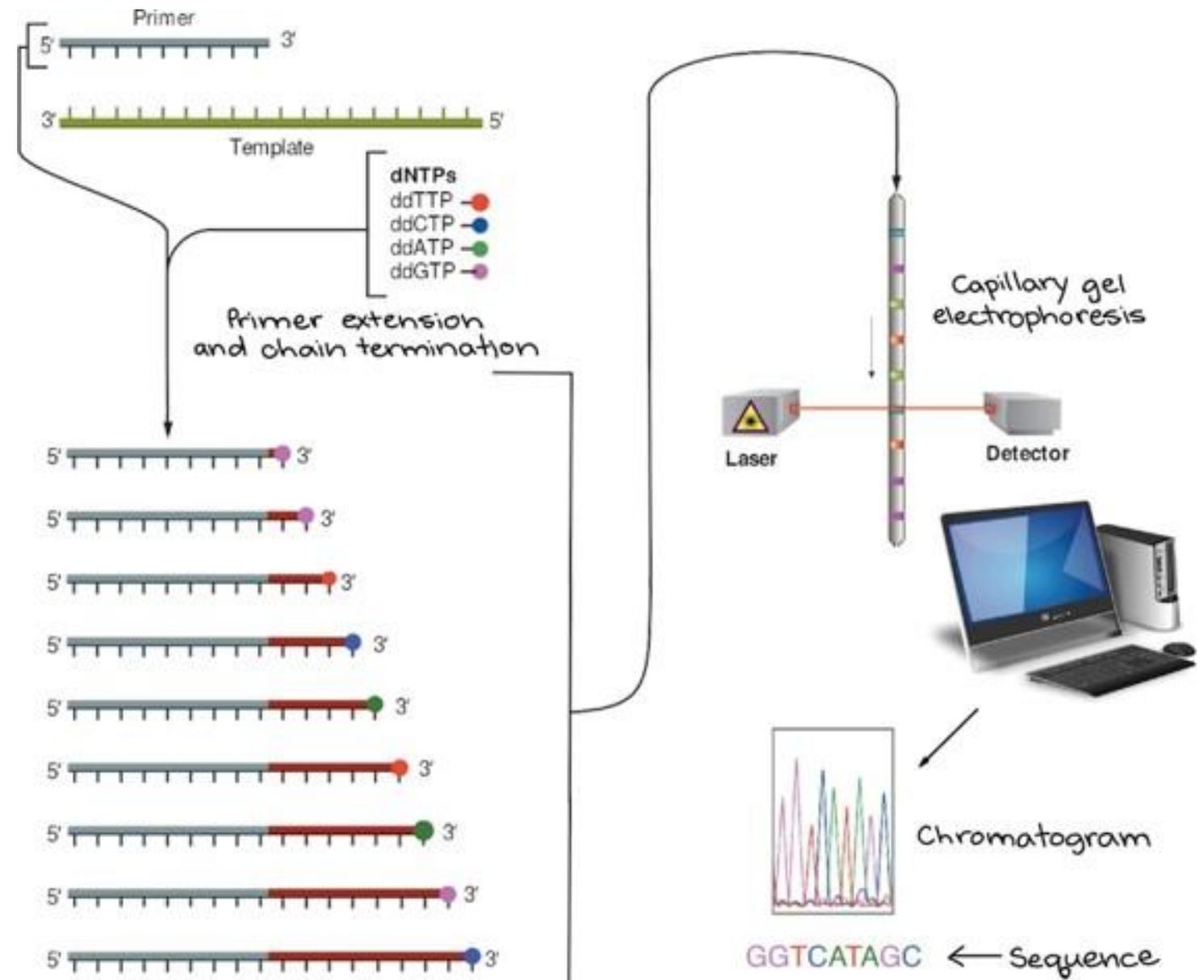
7. Al termine della corsa elettroforetica, un laser eccita i fluorocromi e le emissioni di luce vengono poi rilevate da un detector



# Metodi di prima generazione: metodo Sanger

Elaborazione dei dati:

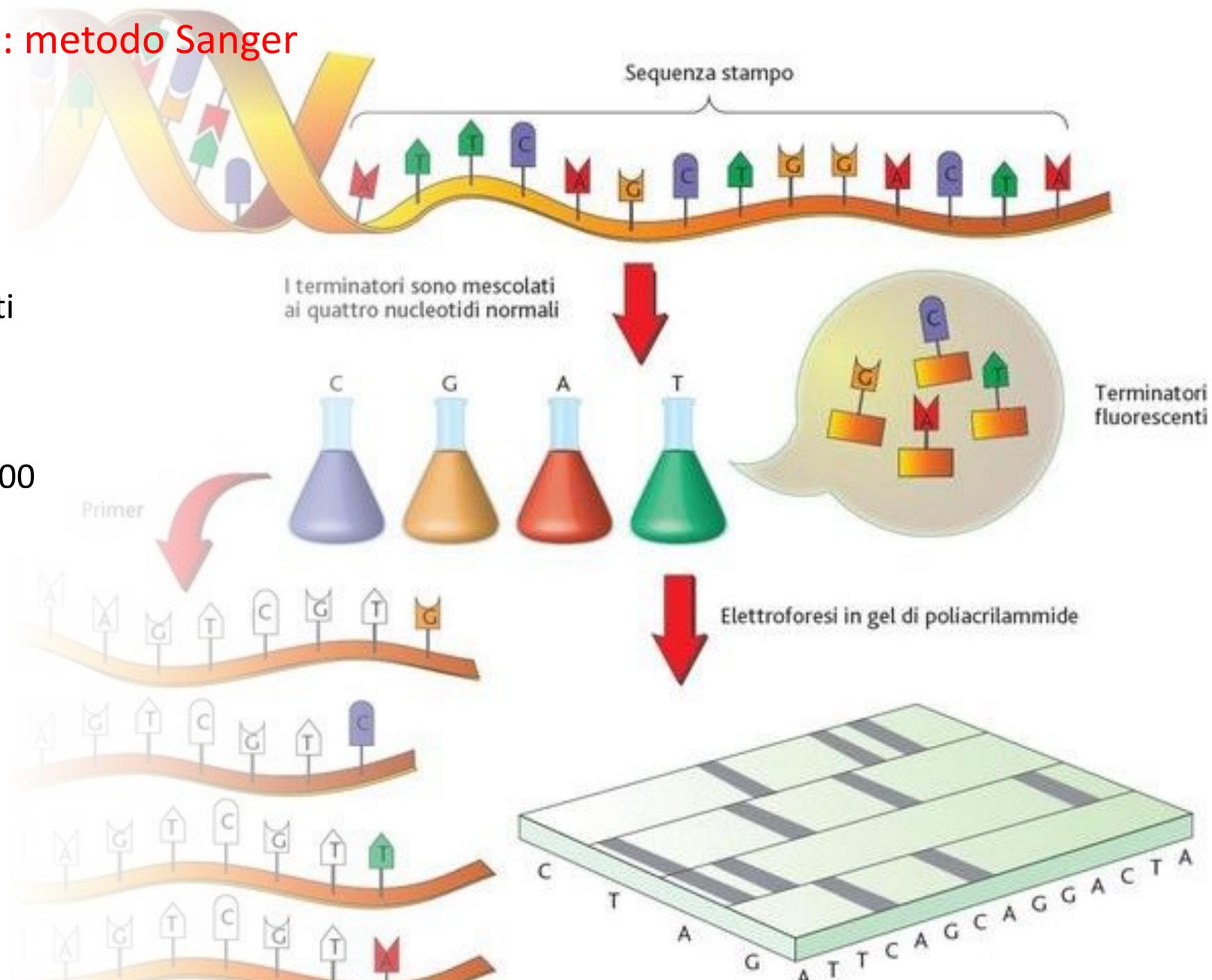
8. I risultati sono elaborati da un software.



# Metodi di prima generazione: metodo Sanger

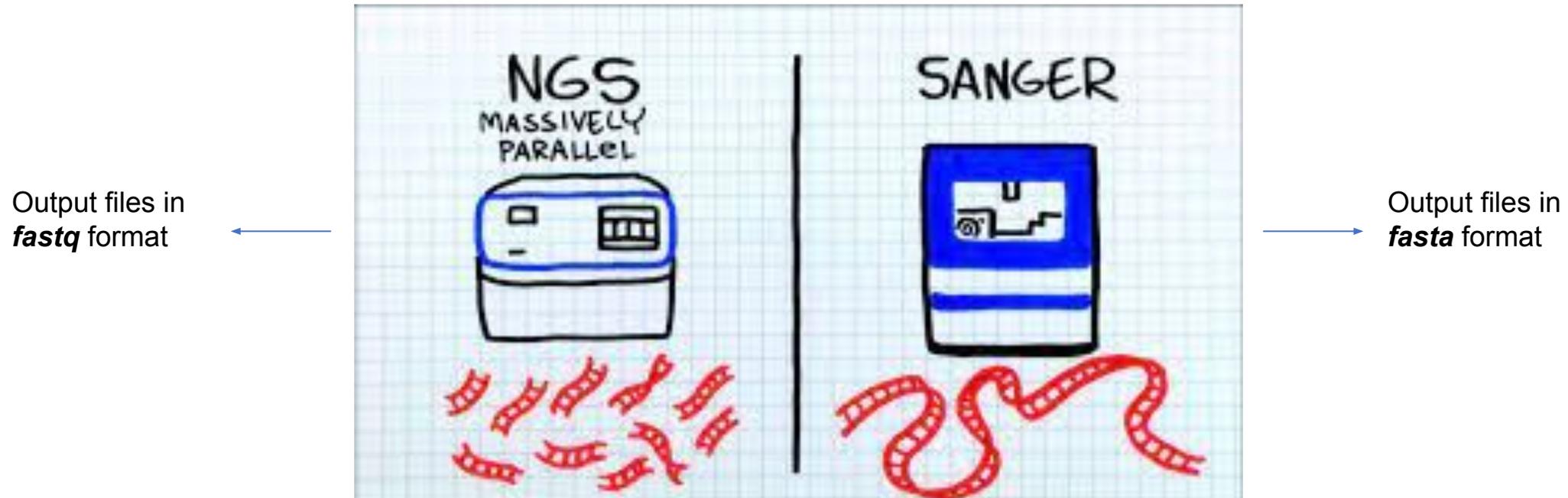
## LIMITI

- Necessità di clonare i frammenti da sequenziare (tramite la PCR)
- Limitazione massima di sequenziamento per ogni corsa elettroforetica è di massimo 1000 bp per ciascuna elettroforesi capillare
- I costi elevati



## Tecnologie innovative per il sequenziamento

Nel 2007, l'introduzione delle tecniche di nuova generazione o **NGS** (*Next-Generation Sequencing*) ha completamente rivoluzionato l'approccio al sequenziamento del DNA, dopo anni in cui il Metodo Sanger rappresentava ormai un punto fermo per gli scienziati e i ricercatori di tutto il mondo.



Per comprendere la portata dell'avanzamento tecnologico in termini di costi, basti pensare che il Progetto Genoma Umano impiegò 10 anni e costò \$3.000.000.000, mentre oggi un macchinario NGS è in grado di sequenziare un singolo genoma umano in un giorno e per meno di \$5.000. In termini quantitativi invece, nei primi anni '80 una persona avrebbe potuto sequenziare un genoma di 30 000 paia di basi di DNA (circa le dimensioni del genoma del SARS-CoV-2) lavorando ininterrottamente per oltre 60 giorni.

# Tecnologie innovative per il sequenziamento

Le diverse piattaforme NGS si differenziano per le modalità con cui effettuano le varie fasi. Si suddividono in 2 categorie

## **N**EXT **G**ENERATION **S**EQUENCING

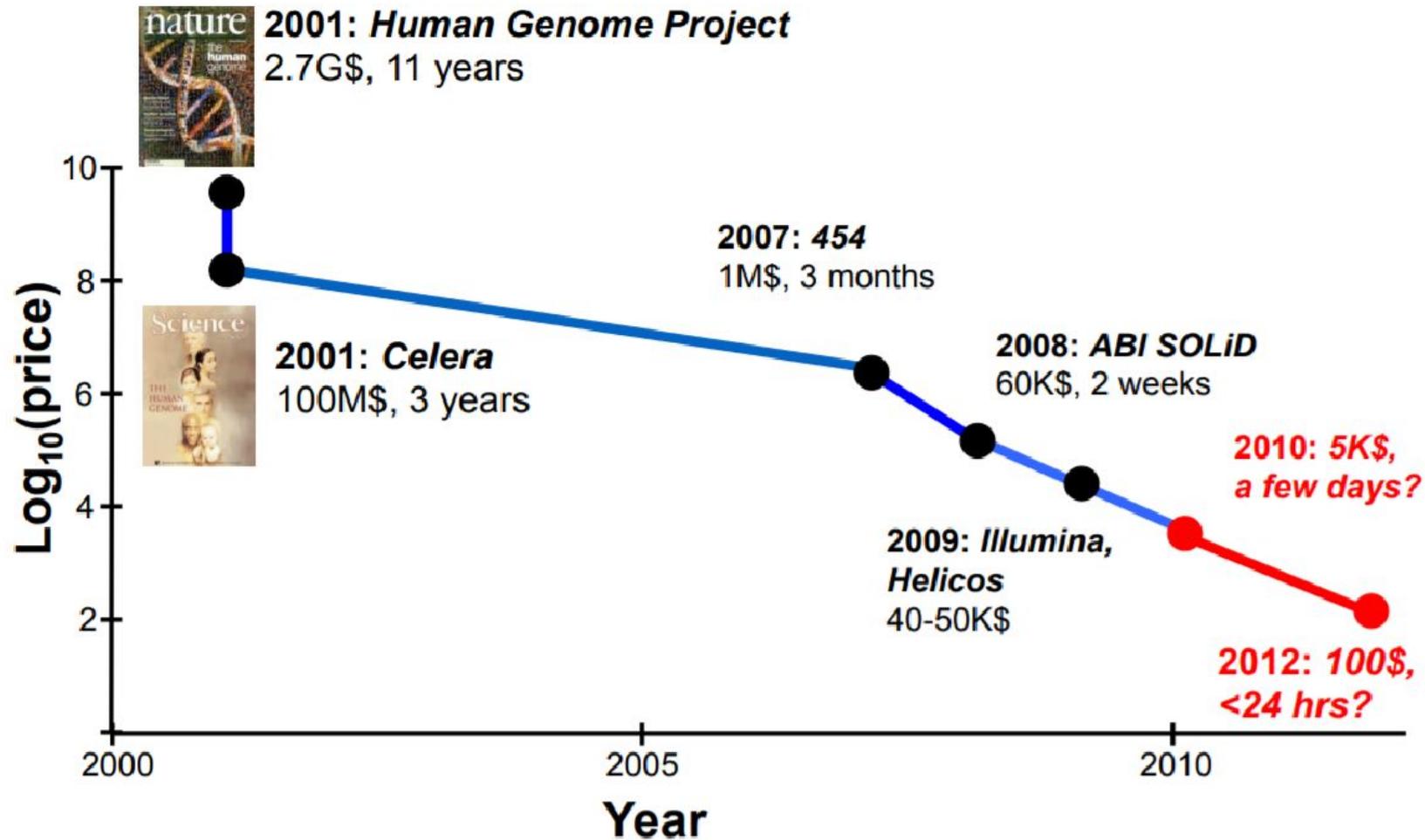
Sequenziatori NGS di  
**Seconda generazione**  
☐ necessitano di  
amplificazione clonale:

- **ROCHE/454 (pyrosequencing)**
- **ILLUMINA/SOLEXA (modified Sanger)**
- **APPLIED BIOSYSTEMS SOLID (sequencing by ligation)**
- **ION TORRENT**

Sequenziatori NGS di  
**terza generazione**  
☐ non necessitano di  
amplificazione clonale:

- **PACIFIC BIOSCENCE**
- **OXFORD NANOPORE**

# Sequencing the Human Genome

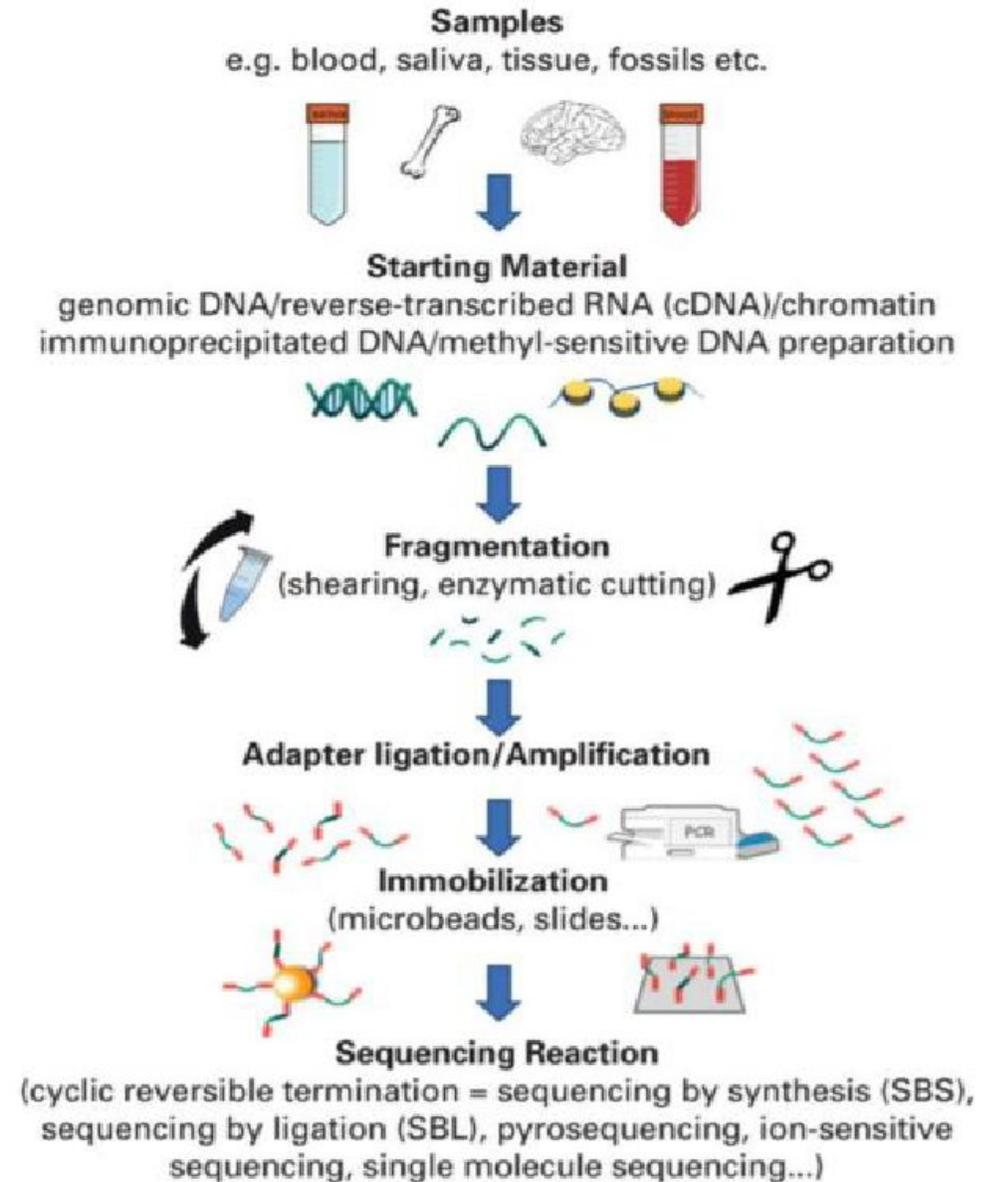


# NGS (Next Generation Sequencing)

Prevedono diverse fasi:

- Estrazione acidi nucleici (dna o rna)
- Frammentazione acidi nucleici e selezione frammenti di dimensione omogenea
- Aggiunta di adattatori alle estremità dei frammenti
- Immobilizzazione su un supporto
- Determinazione sequenza

# Next Generation Sequencing



## Roche 454

La tecnologia di amplificazione Roche 454 è una metodologia di sequenziamento di nuova generazione (NGS) che ha rappresentato un passo avanti significativo rispetto al metodo di sequenziamento tradizionale Sanger.

I passi principali di questa tecnologia sono:

1. **Pirometria:** Il cuore della tecnica 454 è la pirometria. Durante il sequenziamento, ogni inserzione di nucleotide produce una reazione enzimatica che rilascia una piccola quantità di luce. La quantità di luce rilasciata è proporzionale al numero di nucleotidi incorporati, permettendo la rilevazione di omopolimeri (serie di nucleotidi identici in sequenza).
2. **Amplificazione su Sferette (Beads):** Il DNA viene prima amplificato su piccole sferette in gocce d'olio in un processo chiamato emulsione PCR. Ogni sferetta porta molteplici copie di una singola molecola di DNA.
3. **Sequenziamento parallelo:** Dopo l'amplificazione, le sferette vengono caricate in una piastra contenente molti piccoli pozzi. Ogni pozzetto contiene una singola sferetta, e il sequenziamento avviene in parallelo in migliaia di pozzi contemporaneamente.
4. **Reads Lunghe:** Un vantaggio distintivo del 454 era la sua capacità di produrre letture relativamente lunghe rispetto ad altre piattaforme NGS del suo tempo.
5. **Flusso di Lavoro Integrato:** La piattaforma 454 comprendeva l'amplificazione, il sequenziamento e l'analisi dei dati, il tutto in un sistema integrato.

## Roche 454

Il pyrosequencing è la metodologia di sequenziamento del DNA su cui è basata la piattaforma Roche 454. Essa è incentrata sulla rilevazione della liberazione di pirofosfato (PPi) durante la sintesi del DNA. Come funziona il pyrosequencing:

1. **Inizio:** Un singolo filamento di DNA da sequenziare viene usato come stampo su cui verrà sintetizzato un nuovo filamento. Si aggiunge un primer (un breve pezzo di DNA) che si lega (si anneals) allo stampo in una posizione specifica, marcando l'inizio del sequenziamento.
2. **Sintesi del DNA e Rilevazione:** Si aggiungono nucleotidi (A, T, C o G) uno alla volta. Quando un nucleotide si lega al filamento in crescita, viene rilasciato un pirofosfato (PPi).
3. **Generazione di Luce:** La presenza di PPi innesca una serie di reazioni enzimatiche che risultano nella produzione di luce visibile. Questa luce viene rilevata da un sensore e indica che il nucleotide è stato incorporato nel filamento in crescita. L'intensità della luce è proporzionale al numero di nucleotidi consecutivi incorporati (per esempio, se ci sono quattro adenine (A) consecutive nella sequenza del DNA e si aggiunge l'adenina come nucleotide, si otterrà una luce quattro volte più intensa di quella che si avrebbe con una singola adenina).
4. **Lettura Sequenziale:** Dopo ogni aggiunta di nucleotide, gli eventuali nucleotidi non incorporati vengono lavati via e il successivo nucleotide viene aggiunto. La sequenza viene determinata osservando l'ordine in cui la luce viene prodotta durante l'aggiunta di ciascun nucleotide.
5. **Analisi dei Dati:** La sequenza di luci rilevate viene poi tradotta in una sequenza di nucleotidi, fornendo la sequenza del DNA.

# Roche 454

Frammentazione tramite nebulizzazione



Frammenti trasferiti in nano sfere e immobilizzate su vetrino



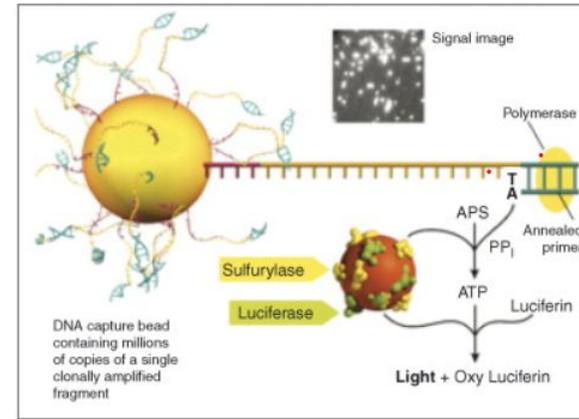
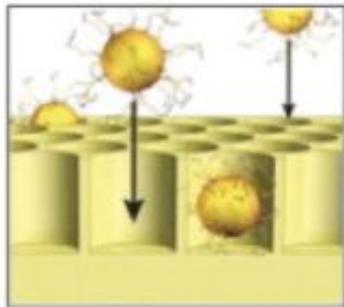
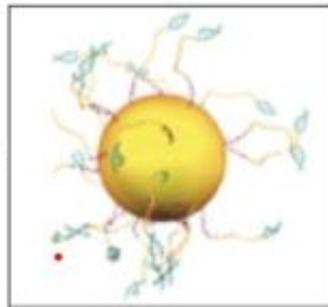
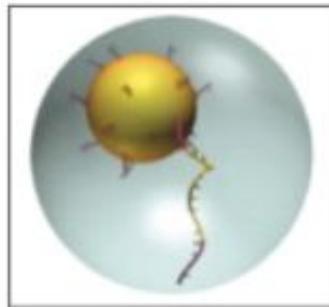
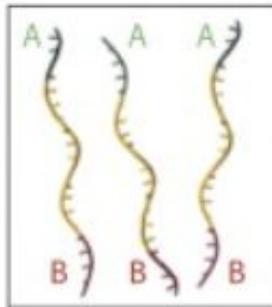
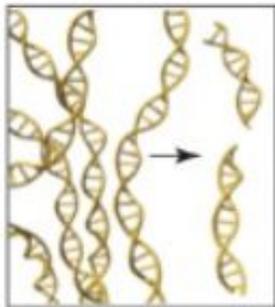
Ogni sfera viene amplificata utilizzando l'emulsione PCR



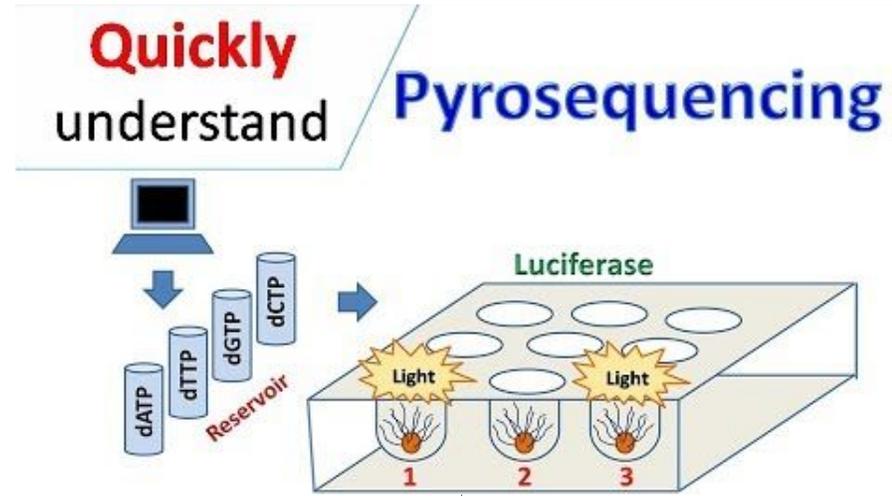
## Pirosequenziamento

Roche (454) GSFLX Workflow:

Library construction



Pyrosequencing reaction



Emulsion PCR

PTP loading

# Roche 454

Frammentazione tramite nebulizzazione



Frammenti trasferiti in nano sfere e immobilizzate su vetrino



Ogni sfera viene amplificata utilizzando l'emulsione PCR



**Pirosequenziamento**

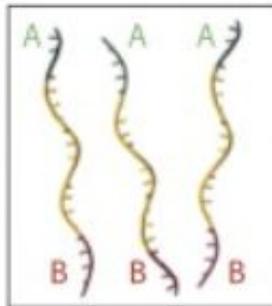
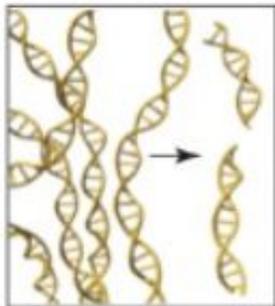
La **nebulizzazione** è un processo fisico in cui il DNA viene sottoposto a forze di taglio per essere frammentato in piccoli pezzi. Questo avviene forzando una soluzione contenente il DNA attraverso un piccolo ugello (nebulizzatore) ad alta pressione, creando uno spray o aerosol. Durante questo processo, le molecole di DNA si rompono in frammenti di dimensioni variabili a causa delle forze meccaniche generate.

Dopo la frammentazione, vengono attaccati degli adattatori ai frammenti, permettendo il loro ancoraggio alle superfici solide durante il processo di sequenziamento e facilitando l'amplificazione.

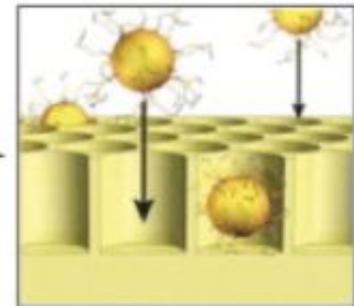
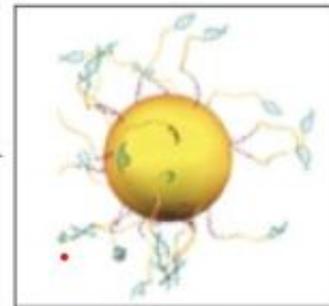
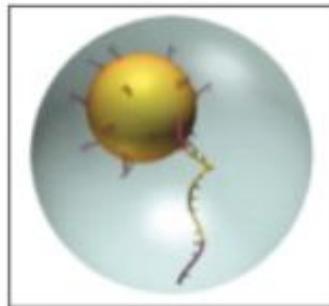
L'emulsione PCR è una forma di **PCR in compartimenti isolati**. In questo caso, si crea una **emulsione** mescolando la soluzione contenente le sfere con olio. Questa emulsione forma migliaia di goccioline acquose microscopiche, ciascuna contenente una singola sfera con un frammento di DNA legato. Ogni gocciolina agisce come un micro-reattore separato, dove avviene la reazione di PCR. Ciò significa che **ogni singolo frammento di DNA attaccato alla sfera viene replicato** molte volte, generando migliaia di copie di quel frammento specifico.

Roche (454) GSFLX Workflow:

Library construction



Emulsion PCR



# Roche 454: Pirosequencing

Il principio di base del pirosequenziamento è che, quando un nucleotide viene incorporato in una nuova catena di DNA durante la sintesi, viene rilasciato un pirofosfato (PPi). Questo pirofosfato rilasciato innesca una serie di reazioni enzimatiche che producono un segnale luminoso. L'intensità di questa luce può essere misurata e correlata all'incorporazione di un nucleotide specifico.

## Sintesi del DNA:

- Si inizia con un singolo filamento di DNA a cui viene aggiunto un primer complementare per iniziare la sintesi. Viene poi fornito un pool di nucleotidi (A, T, C, G) uno alla volta, in modo sequenziale.

## Incorporazione del Nucleotide:

- Quando il nucleotide complementare corretto viene aggiunto alla nuova catena di DNA, la DNA polimerasi lo incorpora e rilascia una molecola di **pirofosfato (PPi)**.

# Roche 454: Pirosequencing

## Reazione Enzimatica:

- Il **pirofosfato** rilasciato avvia una serie di reazioni enzimatiche. In particolare:
  - Il pirofosfato viene convertito in **ATP** dall'enzima ATP sulfurilasi.
  - L'**ATP** serve da substrato per l'enzima luciferasi, che converte il luciferin in ossiluciferin, producendo un segnale luminoso.

## Rilevazione del Segnale Luminoso:

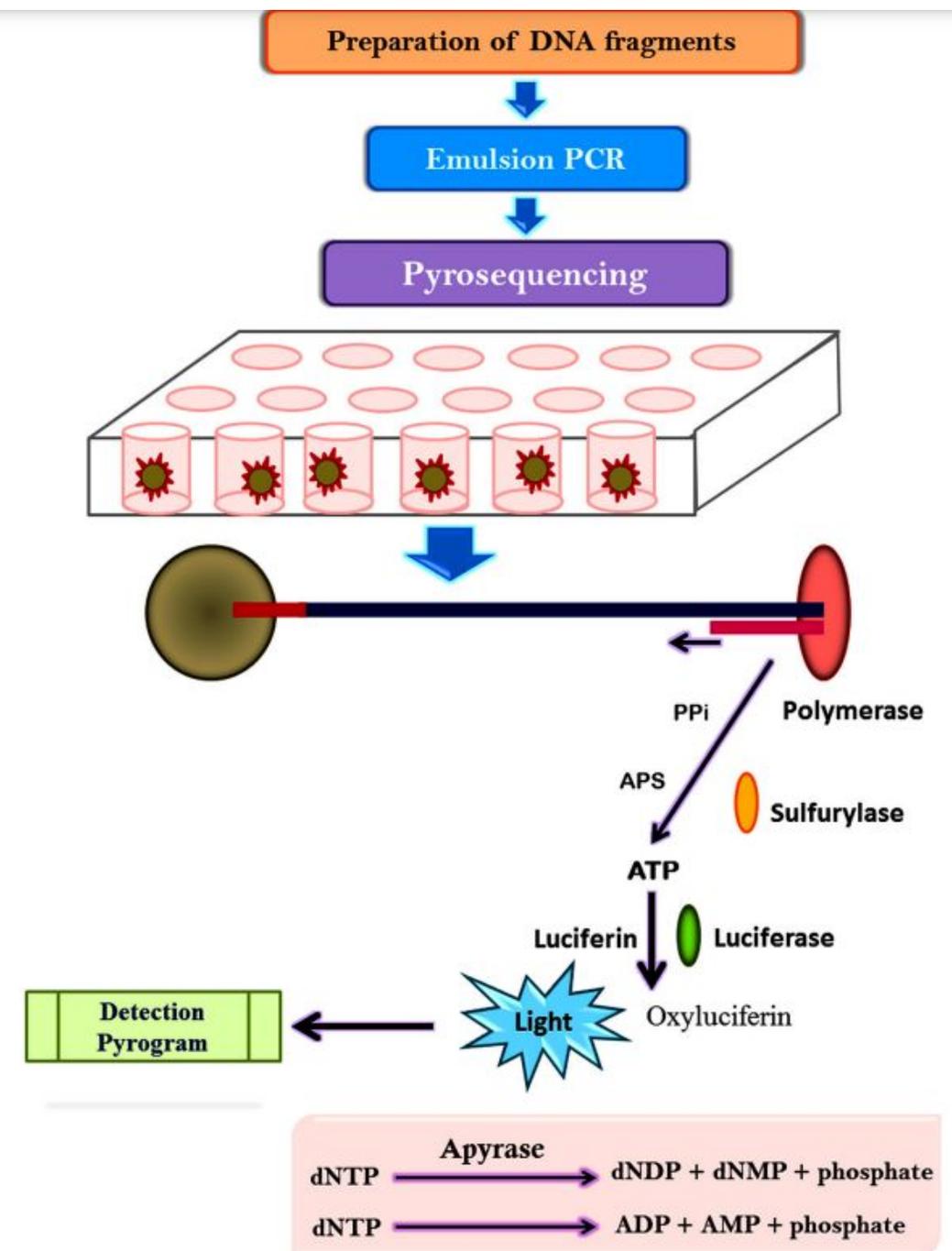
- La luce prodotta dalla reazione della luciferasi è proporzionale alla quantità di nucleotide incorporato. Questo segnale viene rilevato da un rilevatore di luce (fotometro). Se il nucleotide somministrato non è complementare, non avviene alcuna reazione, quindi non viene emesso alcun segnale luminoso.

## Ciclo Ripetuto:

- Ogni nucleotide viene aggiunto sequenzialmente (A, T, C, G), e la quantità di luce emessa durante ogni fase indica quale nucleotide è stato incorporato e in quale quantità (ad esempio, se due nucleotidi uguali si trovano consecutivamente, la luce sarà più intensa).

# Roche 454 - vantaggi

- 1. Read Lunghe:** Uno dei principali vantaggi del Roche 454 è stata (fino al 2013) la sua capacità di produrre letture relativamente lunghe rispetto ad altre piattaforme NGS dell'epoca. Questo lo ha reso ideale per applicazioni come il sequenziamento di nuovi genomi e la risoluzione di regioni genomiche complesse.
- 2. Velocità:** Rispetto ai metodi tradizionali, come il sequenziamento Sanger, il 454 è molto più veloce, consentendo il sequenziamento di interi genomi in tempi molto più brevi.
- 3. Flessibilità:** Il sistema può essere utilizzato per una serie di applicazioni, dal sequenziamento del genoma completo, alla metagenomica, al sequenziamento mirato.
- 4. Senza Clonazione:** A differenza del sequenziamento Sanger, il 454 non richiede la clonazione di DNA in vettori batterici, il che riduce il tempo e la complessità del workflow.

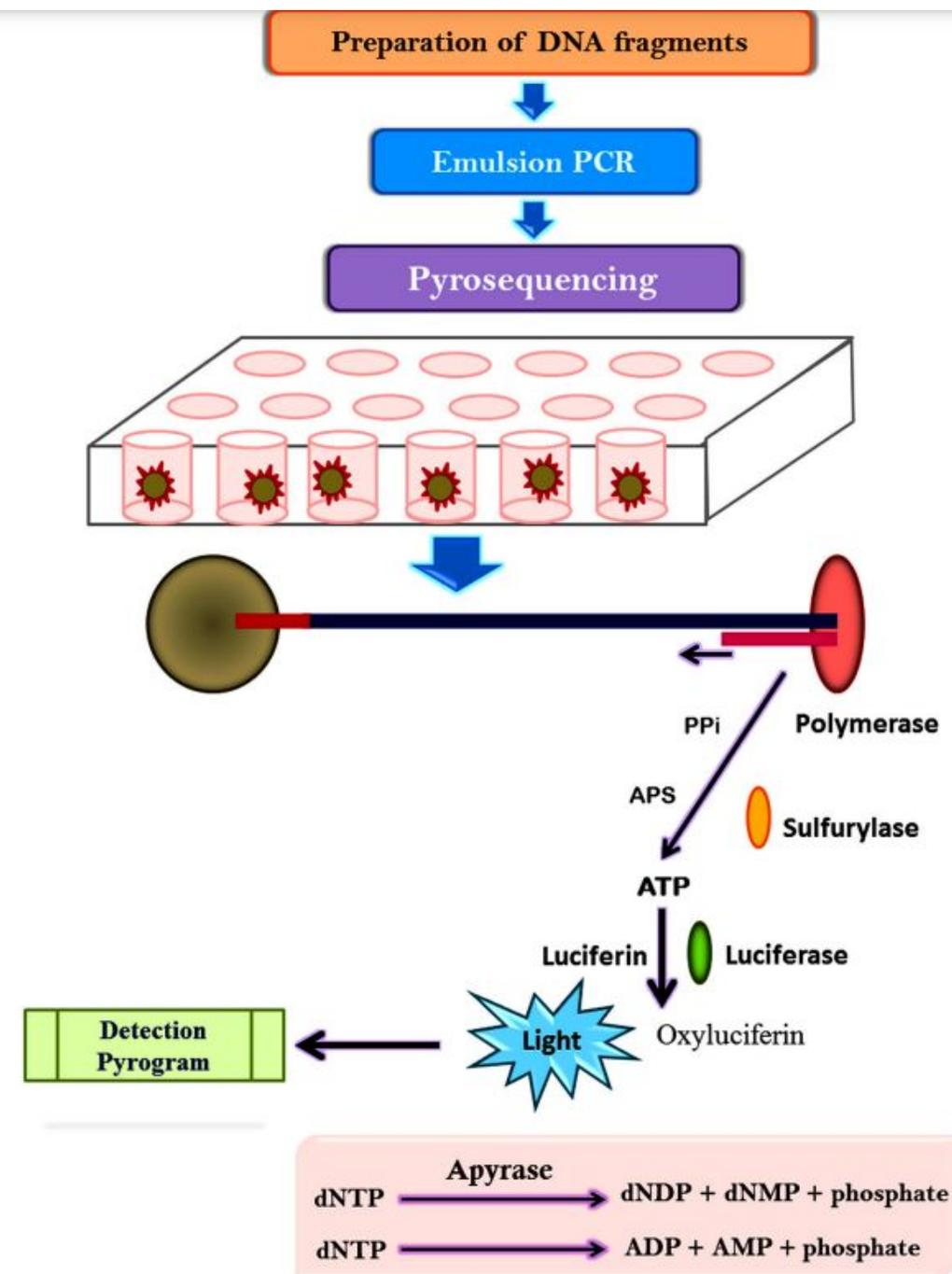


# Roche 454 - limitazioni

1. **Errori in Omopolimeri:** Come il sequenziamento Ion Torrent, il 454 ha difficoltà con le sequenze di omopolimeri. Questo può portare a errori nel determinare il numero di basi ripetute in sequenza.
2. **Costo per Lettura:** Anche se il costo totale per run potrebbe non essere troppo alto, il costo per singola lettura tende ad essere più alto rispetto ad altre piattaforme NGS, come Illumina.
3. **Bassa Capacità rispetto alle Piattaforme più Nuove:** Sebbene il 454 fosse rivoluzionario al suo debutto, la sua capacità è stata superata da altre tecnologie NGS che possono produrre molti più dati per run.
4. **Fine della Produzione:** Roche ha interrotto il sistema 454 nel 2013, il che significa che non ci sono più aggiornamenti e il supporto potrebbe essere limitato. Questo rende meno attraente l'uso del sistema per nuovi progetti.
5. **Lunghezza Variabile delle Reads:** Sebbene le letture siano generalmente lunghe, possono variare in lunghezza, il che può rendere alcune applicazioni, come l'assemblaggio del genoma, più complesse.

In conclusione, mentre il sequenziamento Roche 454 ha avuto un impatto significativo sulla genetica e la genomica, l'emergere di nuove tecnologie con capacità superiore e costi inferiori per lettura ha gradualmente spostato la preferenza della comunità scientifica verso altre piattaforme.

Tuttavia, per il suo tempo, è stato un passo avanti rivoluzionario.



# Sequenziamento Illumina

La tecnologia di sequenziamento Illumina, spesso denominata "sequenziamento di nuova generazione" o NGS, è attualmente una delle più utilizzate al mondo per il sequenziamento di DNA ad alta throughput. Di seguito una sintesi del funzionamento:

**1. Preparazione della Libreria:** Il DNA di interesse viene frammentato in piccoli pezzi e ad ogni frammento vengono legati degli adattatori specifici. Questi adattatori servono per immobilizzare i frammenti su una superficie solida e per fornire siti di legame per i primer utilizzati nel sequenziamento.

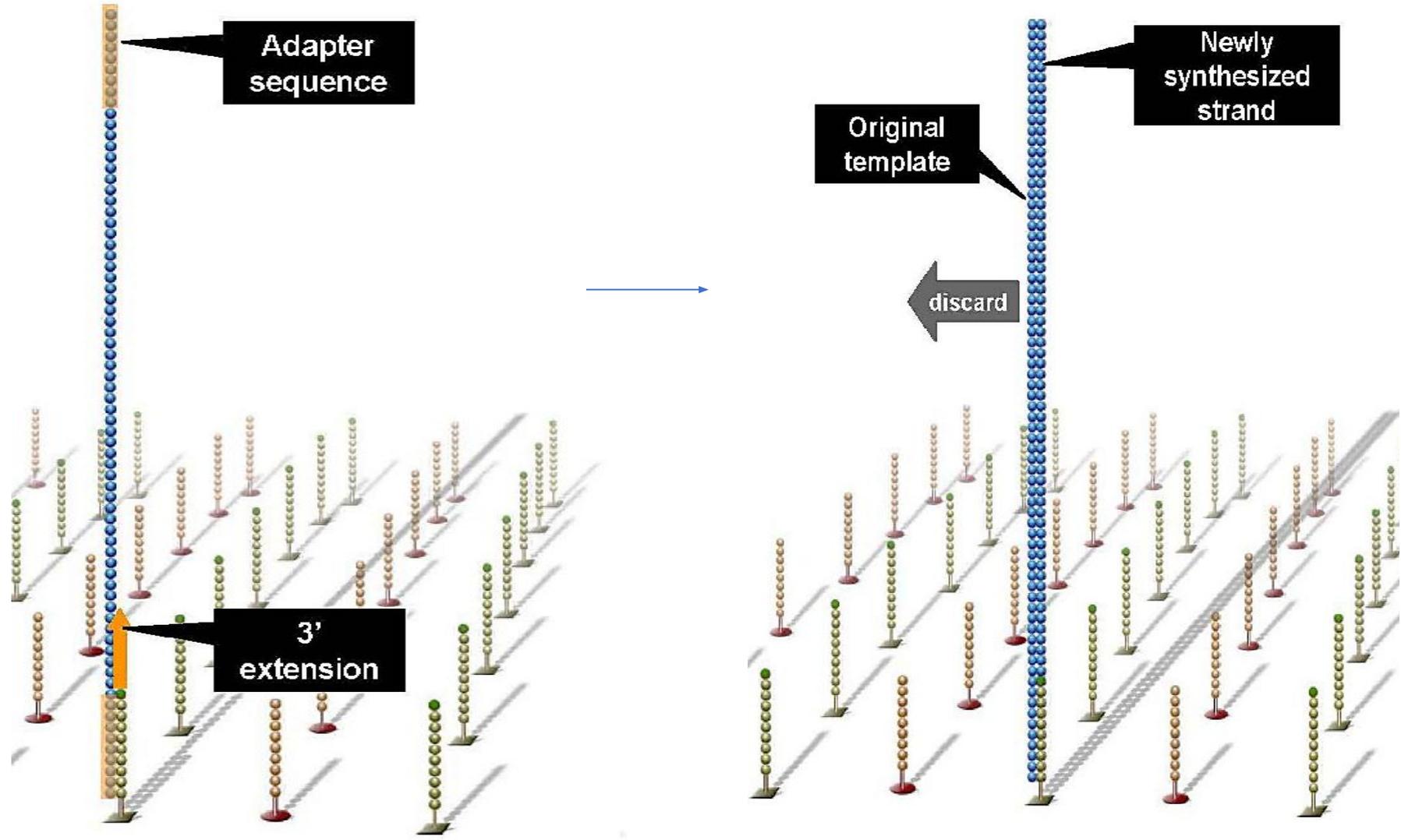
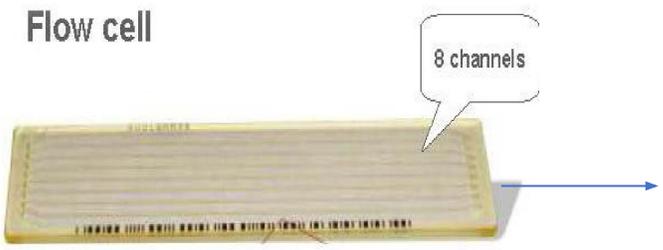
**2. Ponteggio del DNA:** I frammenti di DNA preparati vengono caricati su una cella di flusso e poi riversati su una superficie rivestita con degli oligonucleotidi complementari agli adattatori. Una volta che i frammenti si sono legati alla superficie, iniziano a formare "ponti", piegandosi e legandosi ai siti vicini, creando molteplici copie del medesimo frammento. Questo processo viene chiamato amplificazione di ponte e genera cluster di copie identiche del frammento di DNA.

# Sequenziamento Illumina

3. **Sequenziamento:** Una volta formati i cluster, inizia il sequenziamento vero e proprio. Si tratta di un sequenziamento "sintetico" basato sull'aggiunta sequenziale di nucleotidi fluorescentemente marcati. Ad ogni ciclo, viene aggiunto un tipo di nucleotide marcato. Se il nucleotide è complementare al prossimo nucleotide del filamento stampo, viene incorporato dal polimerasi. Successivamente, la cella di flusso viene sottoposta a scansione da una macchina fotografica, registrando la fluorescenza emessa da ogni cluster. Dopo ogni ciclo di lettura, la fluorescenza del nucleotide marcato viene rimosso, e il ciclo ripete con il nucleotide successivo.

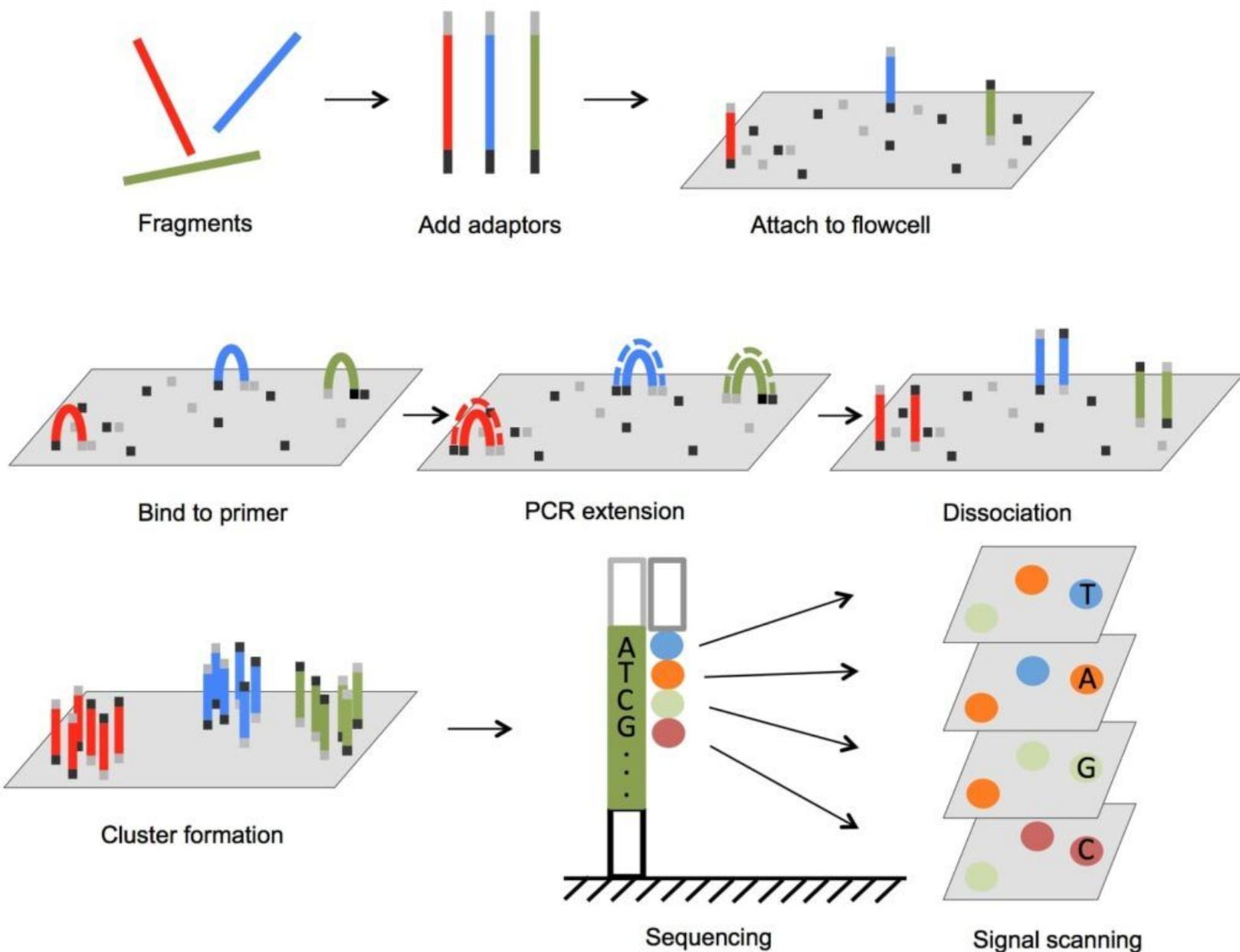
Alla fine del processo di sequenziamento, si ottiene una serie di "read" o letture, che rappresentano le sequenze di DNA dei frammenti originari. Utilizzando software specializzati, queste letture possono essere allineate, assemblate e analizzate per ottenere informazioni sul DNA di partenza.

# ILLUMINA - Amplification

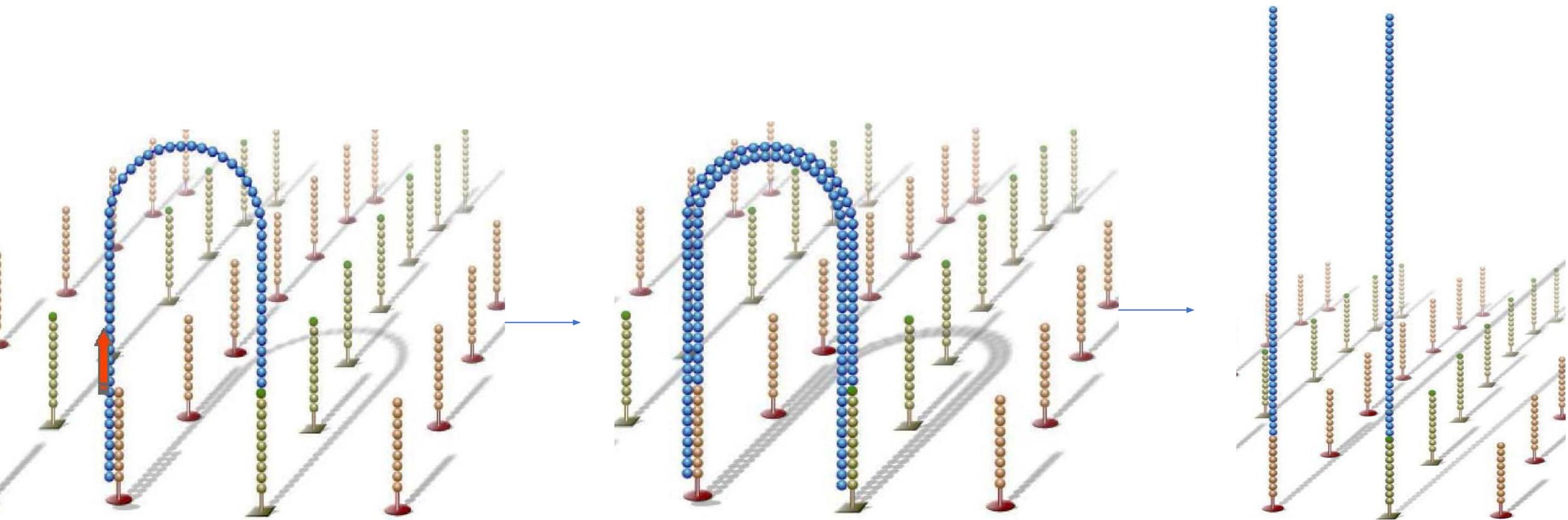


In figura riportiamo la fase iniziale (immagine a sinistra) e finale (immagine a destra) dell'amplicazione Illumina. Il dettaglio del processo di amplificazione lo vediamo con le prossime slide .

# Illumina Technology

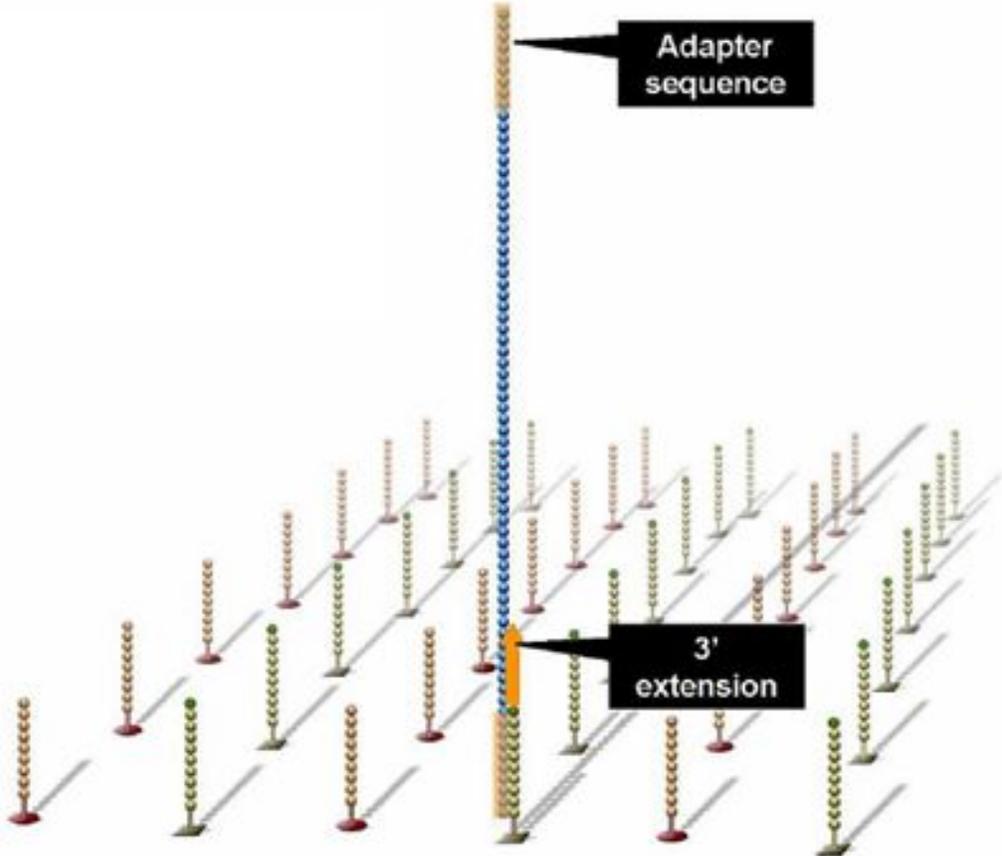


# Illumina - Amplification



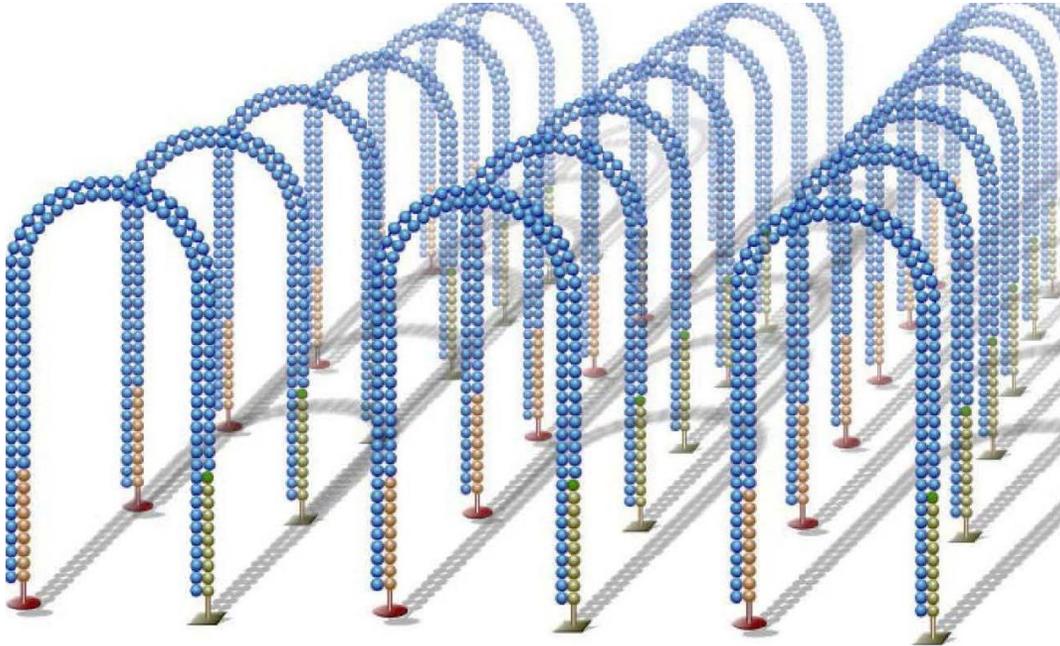
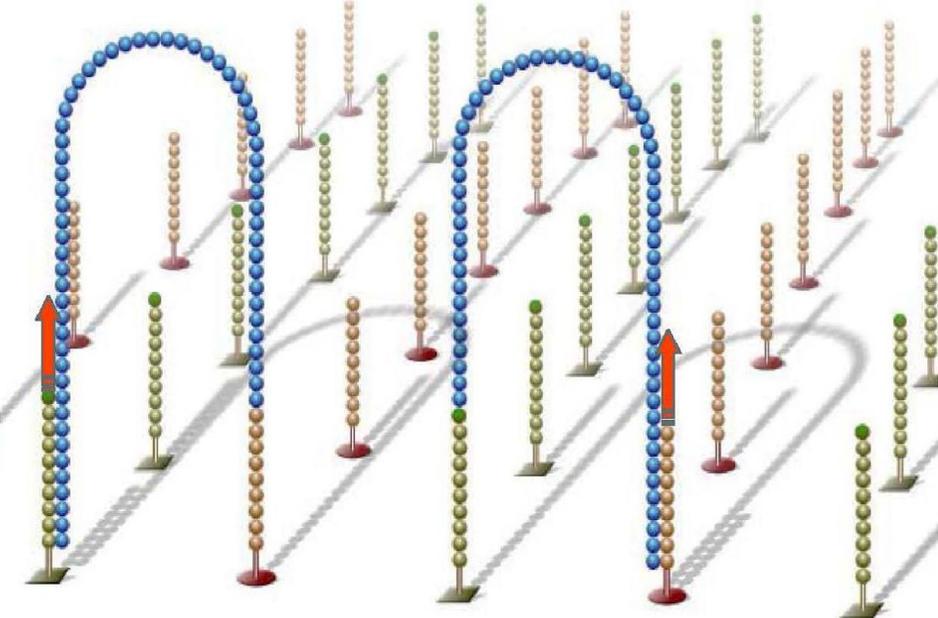
# Illumina - Amplification

## Cluster Generation *Hybridize Fragment & Extend*



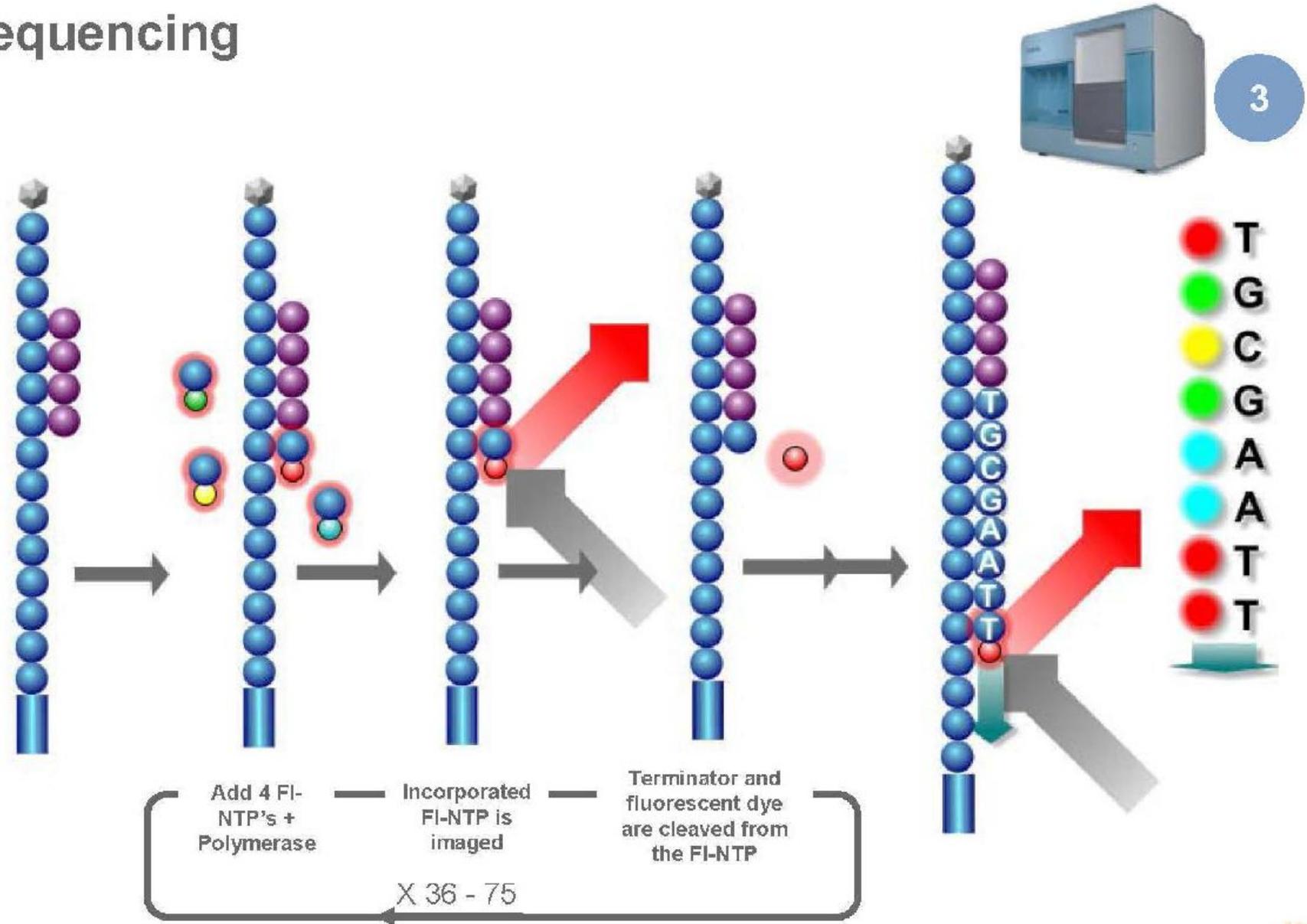
Input requirement: 0.1–1.0  $\mu\text{g}$  (c)DNA

# Illumina - Amplification



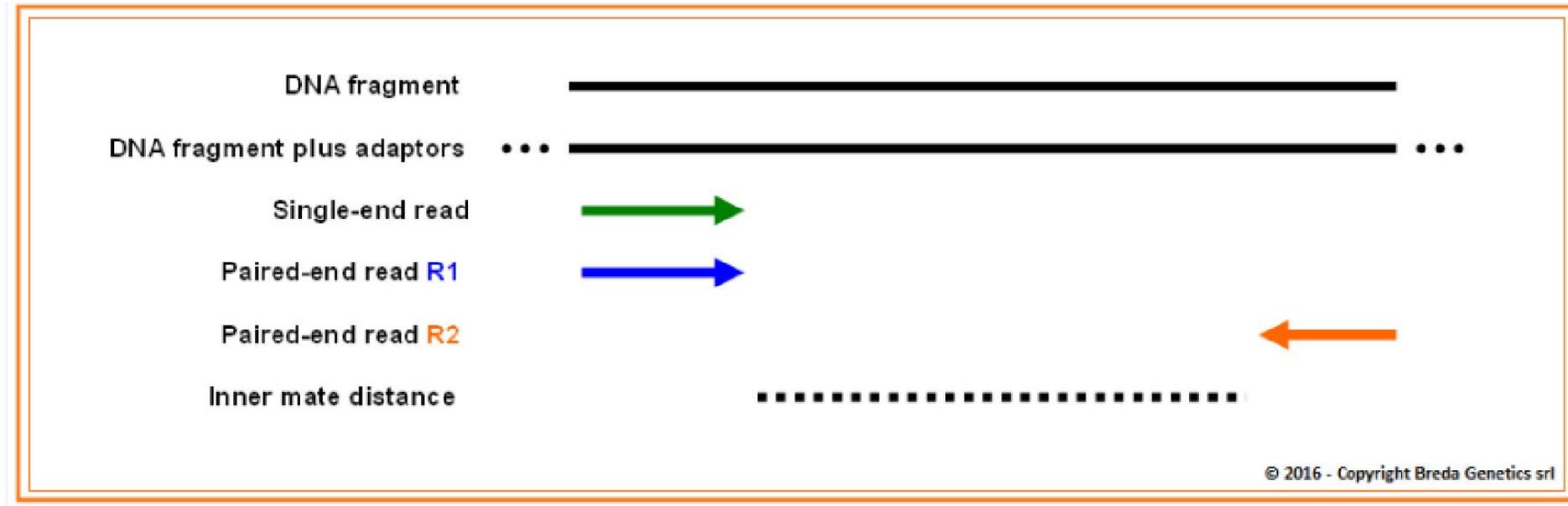
# Illumina - Sequencing

## Sequencing



Una volta amplificati i frammenti, verranno sottoposti al sequenziamento. La tecnica di sequenziamento e' analoga a quella vista con il Sanger, ovvero tramite l'utilizzo di deoxinucleotidi modificati con fluorocromi.

# Illumina - Sequencing



Il sequenziamento può essere fatto a partire da una sola estremità del frammento (**sequenziamento con single-end reads**) o partendo da entrambe le estremità e proseguendo in direzioni opposte (**sequenziamento con paired-end reads**).

➔ **L' inner mate distance** è la lunghezza della sequenza interposta tra R1 ed R2. (Ad esempio, in un sistema nel quale si utilizzi una sequencing library con frammenti da 500bp e nel quale si producano reads di 100bp, la distanza intermedia fra R1 e R2 sarà di circa 300bp.) Grazie alla quale è possibile individuare mutazione genomiche. Quali?

# Illumina sequencing: vantaggi e limitazioni

## Vantaggi:

- Alto throughput: Può sequenziare miliardi di basi in un singolo run.
- Precisione: Ha un tasso di errore molto basso.
- Flessibilità: Può essere utilizzato per varie applicazioni, dal sequenziamento del genoma intero a panel di geni *targeted*.

## Limitazioni:

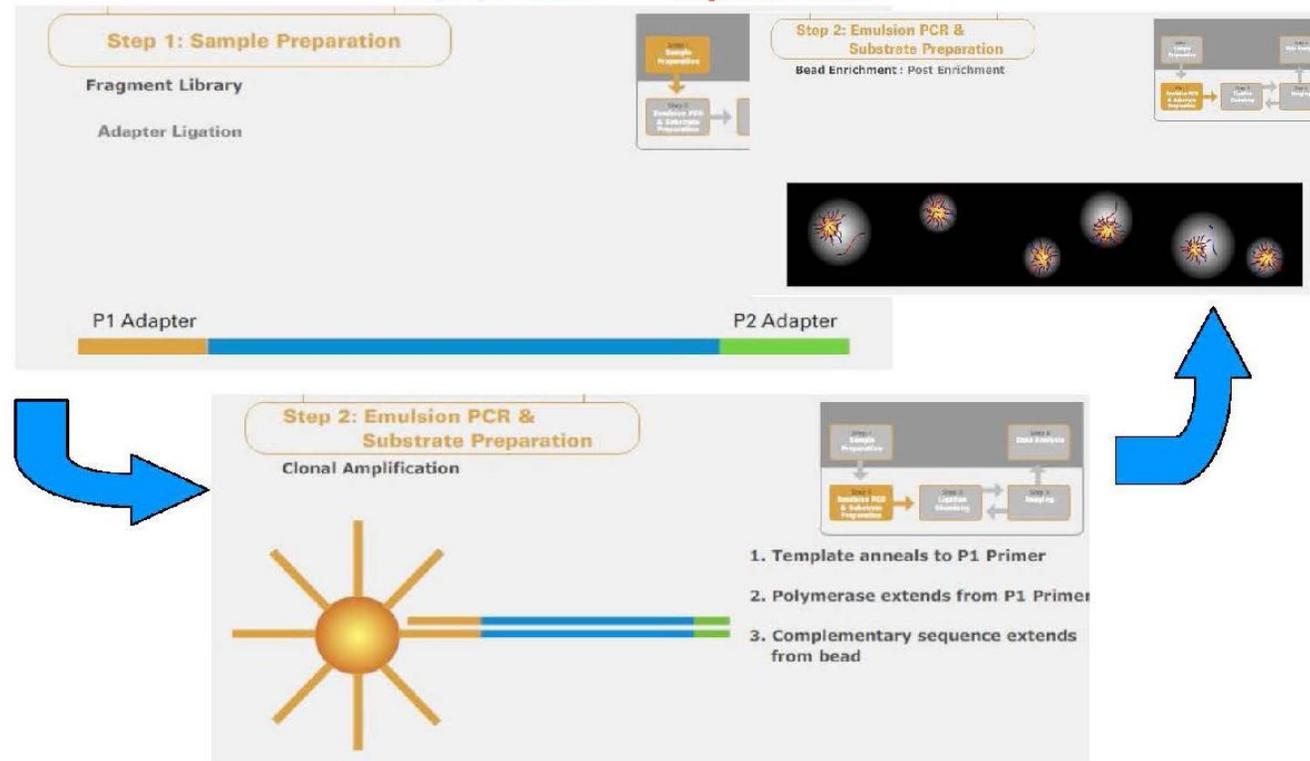
- La lunghezza dei read è tipicamente più corta rispetto ad altre tecniche come il sequenziamento Sanger.

Il sequenziamento Illumina ha rivoluzionato la genetica e la genomica, rendendo possibile sequenziare interi genomi in modo rapido e conveniente e contribuendo significativamente all'era della genomica personale e della medicina personalizzata.

# Sequenziamento Solid

Il sistema **SOLID** prodotto da Applied Biosystem è una piattaforma per il sequenziamento in parallelo di segmenti di DNA amplificati in modo clonale e legati a sferette magnetiche. La metodologia di sequenziamento è basata sulla "ligazione sequenziale" di oligonucleotidi marcati con fluorocromi (SBL, sequencing by ligation).

## Sequencing by Ligation SOLID™ system



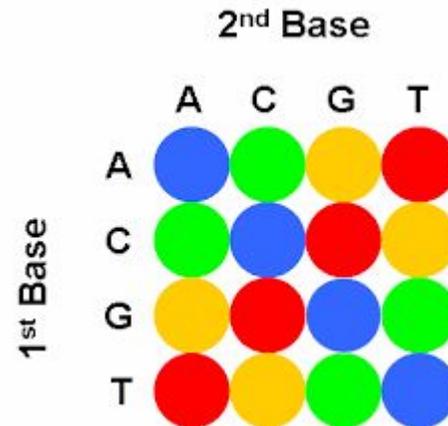
Il sistema SOLID può generare fino a 3 Gbp di dati di sequenza con un'accuratezza superiore al 99%.

# Solid

Questa tecnologia di nuova generazione genera  $10^8 - 10^9$  letture di *short reads* contemporaneamente. Quindi è capace di produrre milioni di sequenze brevi in un singolo processo, rendendo la tecnica molto efficiente e ad alto rendimento.

Utilizza la codifica a 2 basi (**2 Base Encoding**) per decodificare i dati grezzi generati dalla piattaforma di sequenziamento in dati di sequenza.

Nel sequenziamento tradizionale, una base viene letta alla volta, e ciascuna base (A, T, C o G) viene identificata da una specifica fluorescenza o un segnale. Tuttavia, nel sistema SOLiD, invece di identificare singolarmente ogni base, vengono identificate coppie di basi consecutive.



## **Solid: reazione di ligazione**

Una **reazione di ligazione** è un processo biochimico in cui due frammenti di DNA vengono uniti insieme.

Questo processo è mediato da un enzima chiamato ligasi, che facilita la formazione di un legame covalente tra i gruppi fosfato e i gruppi ossidrile dei frammenti di DNA, risultando in un singolo pezzo di DNA.

Sequenziamento SOLiD: la reazione di ligazione può essere utilizzata per attaccare sonde fluorescenti specifiche al DNA durante il processo di sequenziamento.

# Solid

Come funziona la codifica basata su due basi:

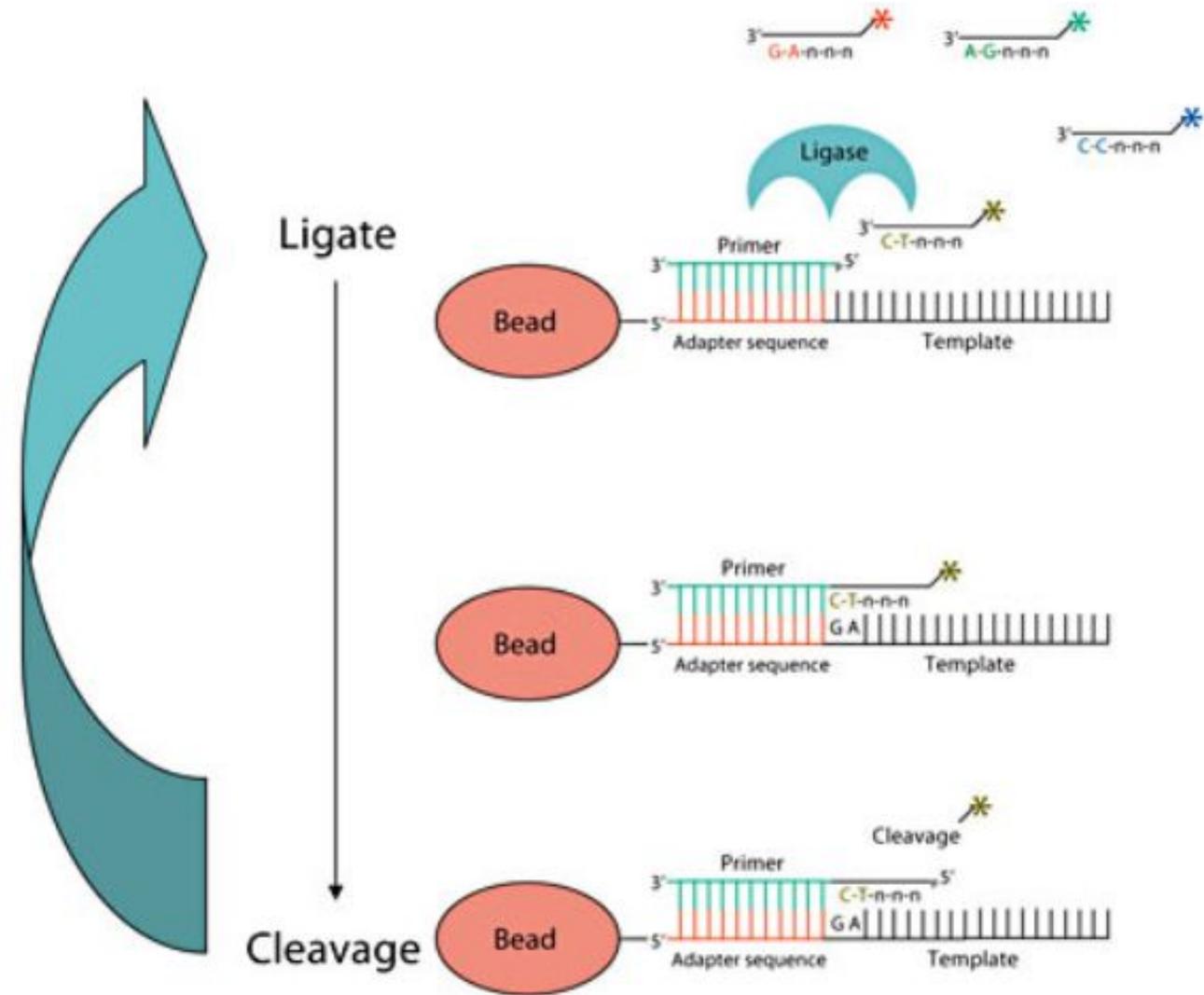
1. **Sonde Oligonucleotidiche:** SOLiD utilizza sonde fluorescenti a breve oligonucleotide. Ogni sonda è progettata per riconoscere una coppia specifica di basi e possiede un colorante fluorescente unico.
2. **16 Possibili Combinazioni:** Poiché ci sono quattro basi nel DNA (A, T, C, G), ci sono 16 possibili combinazioni di coppie di basi (per esempio, AA, AT, AC, AG, TA, TT, ecc.). Ogni combinazione ha un colore associato in base alla sonda che la riconosce.
3. **Sequenziamento:** Durante il processo di sequenziamento, le sonde competono per legarsi al frammento di DNA target. Quando una sonda si lega, emette una fluorescenza che viene rilevata e registrata. Poiché ogni coppia di basi ha un colore unico, la sequenza del DNA viene determinata in base alla serie di colori rilevati.
4. **Duplica Rilevazione:** Una delle principali forze della codifica basata su due basi è che ogni base nel frammento di DNA target viene letta due volte (una volta in ogni coppia). Ciò aumenta l'accuratezza del sequenziamento poiché eventuali errori in una lettura possono essere corretti o confermati dalla lettura successiva.

## Perché è vantaggioso?

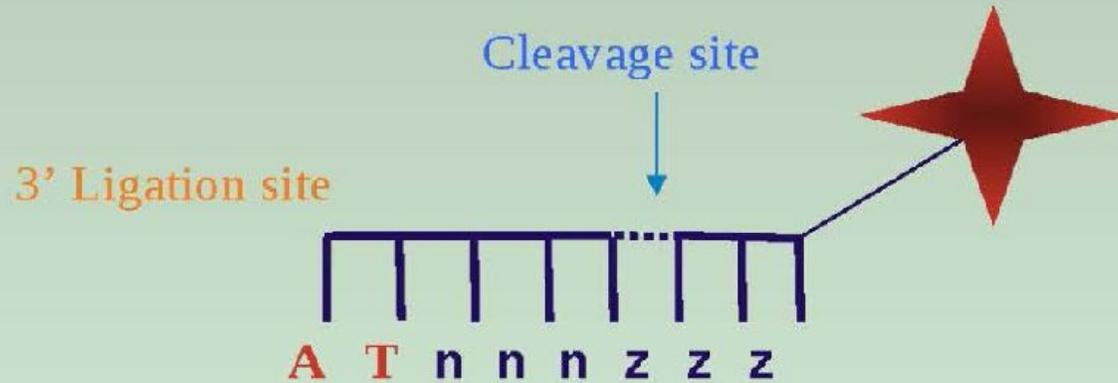
L'approccio di codifica basato su due basi offre una maggiore accuratezza rispetto ad altri metodi. L'errore in una lettura potrebbe non manifestarsi nella lettura successiva, offrendo una forma di correzione dell'errore integrata nel processo. Questa ridondanza nel rilevamento rende il sistema SOLiD particolarmente robusto contro gli errori di sequenziamento.

## Solid: reazione di ligazione

Rappresentazione schematica del sequenziamento SOLiD™ tramite ligazione. I primer si ibridano all'adattatore all'interno del modello di libreria. Un set di 4 sonde a due-basi fluorescenti competono per la ligazione al primer di sequenziamento. Queste sonde hanno una sequenza di DNA parzialmente degenerata (indicata da n). La specificità della sonda a due-basi viene raggiunta interrogando ogni 1<sup>a</sup> e 2<sup>a</sup> base in ciascuna reazione di ligazione. Vengono eseguiti cicli multipli di ligazione, rilevamento e scissione, con il numero di cicli che determina la lunghezza finale della lettura.



# Solid – Ligation Probe



Fluorescent dye interrogates  
base on  
1<sup>st</sup> + 2<sup>nd</sup> position

- **Ligation Probes are Octamers**
  - N=degenerate bases, Z=universal bases
  - $4^5 = 1024$  probes (256 probes per color)

2<sup>nd</sup> Base

	A	C	G	T
A				
C				
G				
T				

es4 B

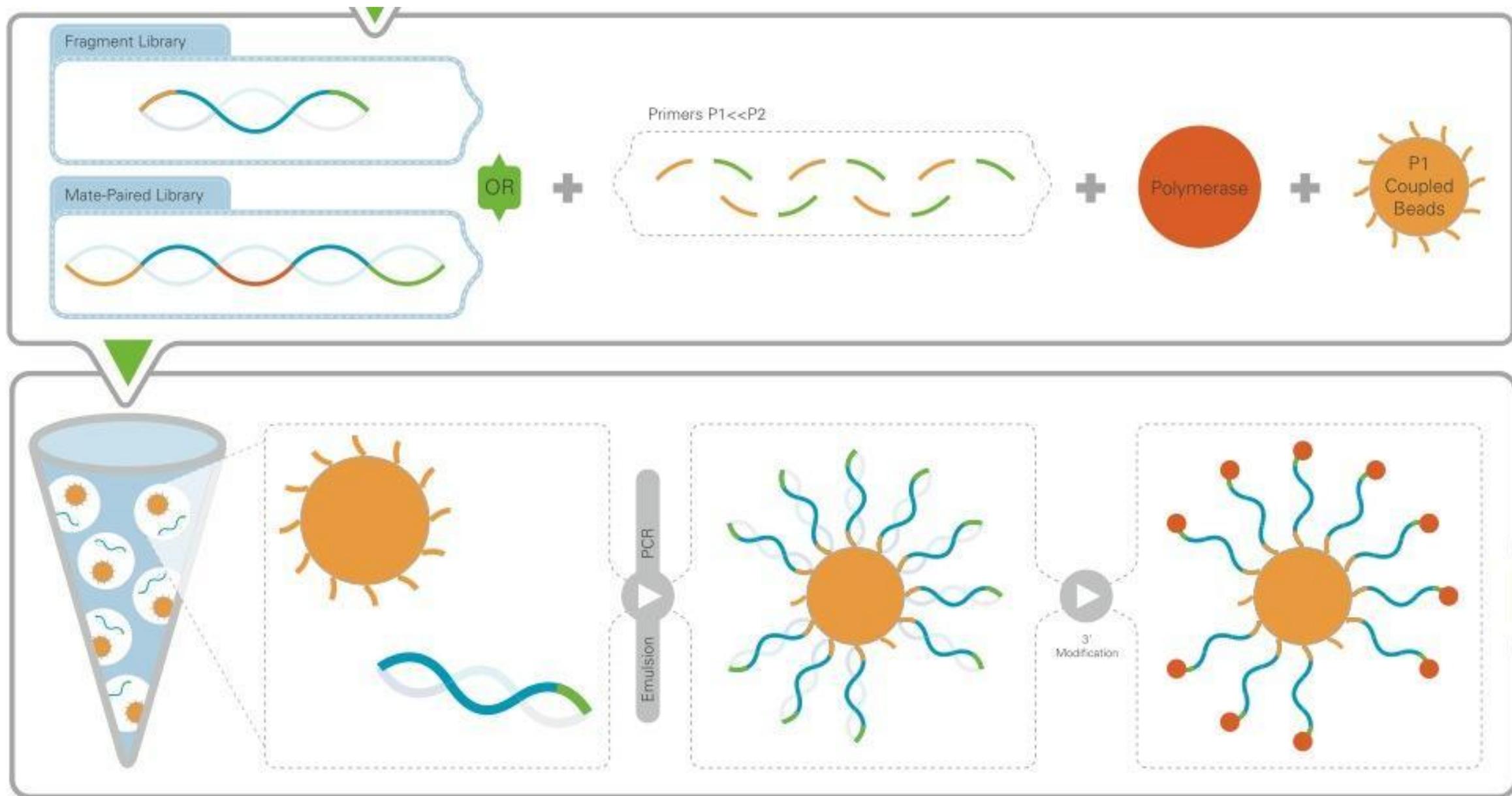
# Solid amplification

Processo Dettagliato:

- Una libreria di frammenti di DNA viene preparata e ogni frammento viene attaccato a una perla magnetica con una sequenza unica.
- Due adattatori vengono poi attaccati alla libreria di frammenti.
- I frammenti vengono amplificati e depositati su una lama di vetro.
- Vengono utilizzati degli inneschi e sonde fluorescentemente etichettate per competere nella ligazione.
- Le sonde funzionano analizzando le basi in ogni reazione di ligazione per determinare la sequenza del DNA.
- Dopo vari cicli di ligazione, vengono rimossi i prodotti estesi e il template viene reimpostato per ulteriori cicli di ligazione.
- Ogni base nel template viene sequenziata due volte e i dati risultanti vengono decodificati usando un sistema di codifica a due basi, dove ciascuna coppia di basi sul 3' della sonda viene assegnata a uno dei quattro possibili colori.

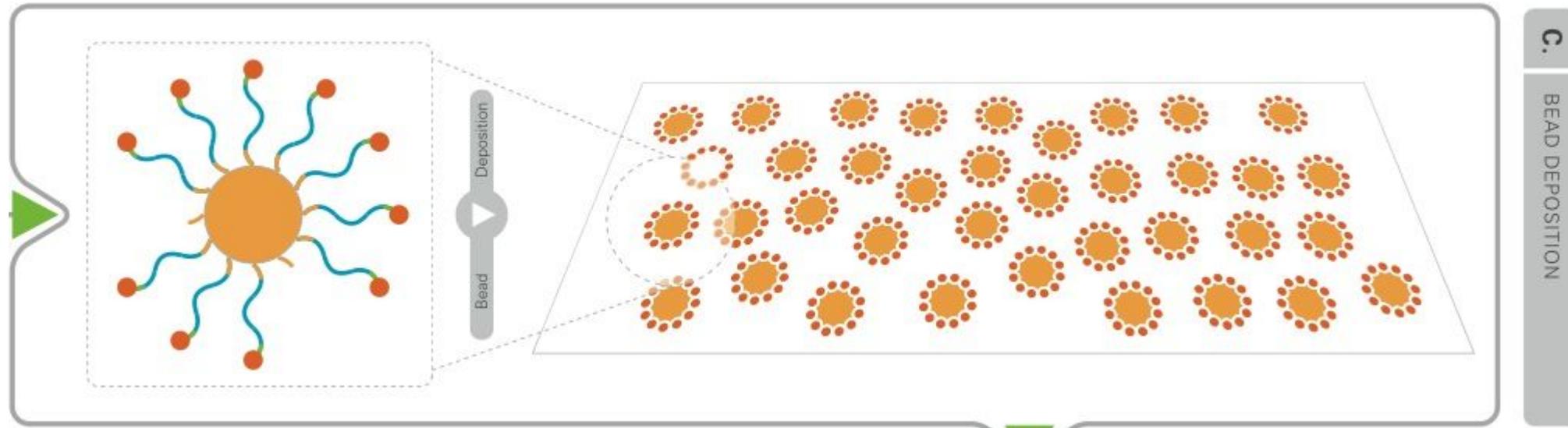
# Solid – Emulsion PCR

B. EMULSION PCR / BEAD ENRICHMENT



## Solid – Emulsion PCR

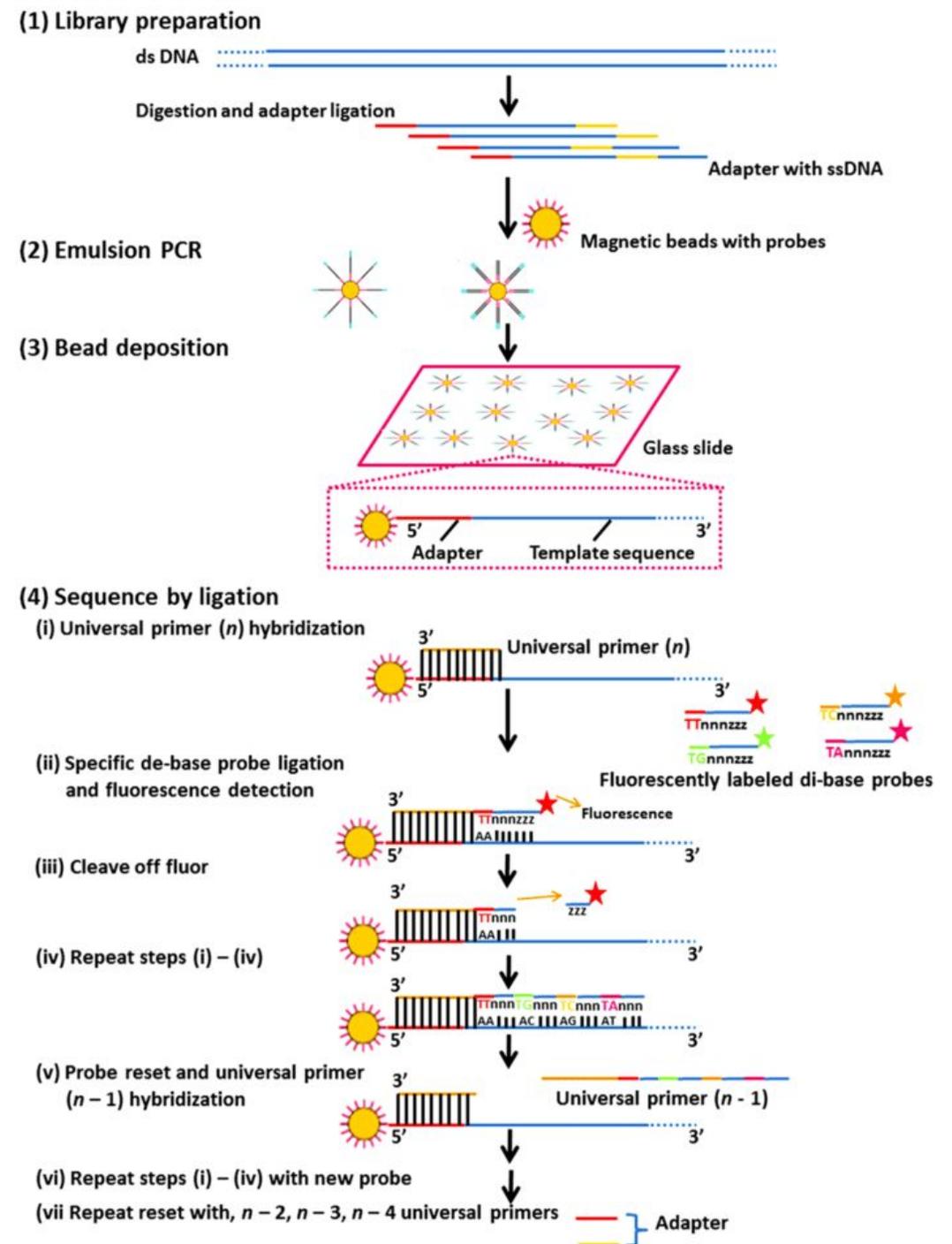
Ciascun sfera viene quindi attaccata alla superficie di una cella a flusso tramite modifiche in 3' ai filamenti di DNA.



A questo punto, abbiamo una cella a flusso la cui superficie è ricoperta da migliaia di perline ciascuna contenente una singola specie di DNA genomico, con adattatori unici su entrambe le estremità. Ogni microsfera può essere considerata una reazione di sequenziamento separata che viene monitorata simultaneamente tramite imaging digitale sequenziale.

# Solid – Emulsion PCR

- (1) **Preparazione della libreria:** due adattatori distintivi vengono legati al DNA genomico frammentato.
- (2) **PCR in emulsione:** la PCR in emulsione è gestita con perline magnetiche per generare “cloni su perline”, in cui ciascuno comprende una singola specie di acido nucleico.
- (3) **Deposizione delle perline:** le perline vengono poi associate alla superficie esterna di una lastra di vetro.
- (4) **Sequenziamento tramite ligazione:** il sequenziamento mediato dalla ligasi inizia con l'annealing di un primer universale alle sequenze adattatore comuni su ciascun frammento amplificato (i) e successivamente viene fornita la DNA ligasi accompagnata da specifici 8-mers fluorescentemente etichettati, nei quali le due basi all'estremità 3' della sonda sono codificate dal cluster fluorescente annesso. Ogni passo di ligazione è monitorato dal riconoscimento della fluorescenza

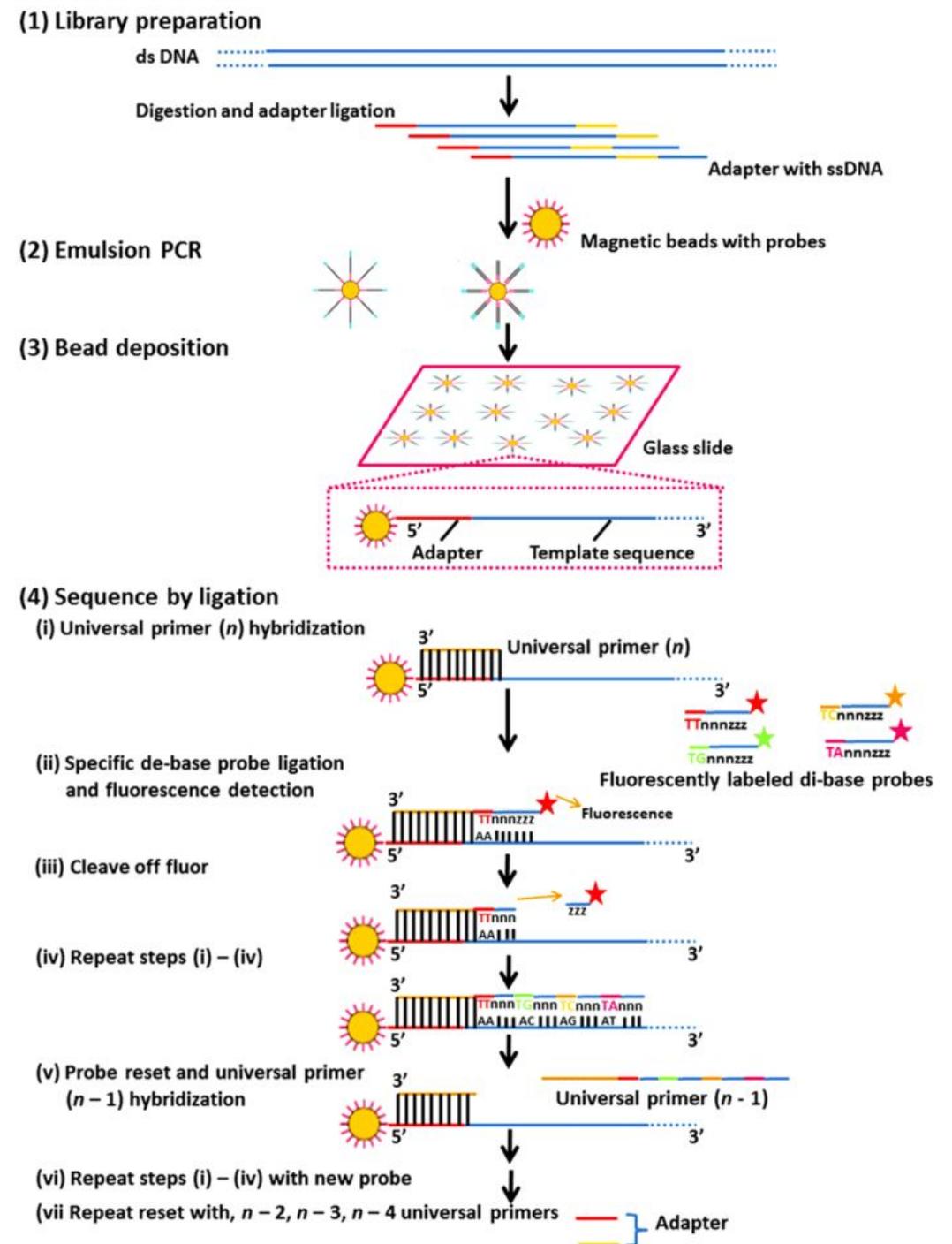


# Solid – Emulsion PCR

## (4) Sequenziamento tramite ligazione:

Ogni passo di ligazione è monitorato dal riconoscimento della fluorescenza (ii), dopodiché un passo di recupero elimina le basi dall'8-mer ligato (includendo il cluster fluorescente) (iii), e in modo corrispondente prepara la sonda estesa per un ulteriore ciclo di ligazione (iv–vii).

Successivamente, ogni cluster fluorescente su un 8-mer ligato distingue una combinazione di due basi, le letture di sequenza risultanti possono essere analizzate per errori nella chiamata delle basi rispetto a corretti polimorfismi o eliminazioni di singole basi, confrontando le sequenze ottenute con una sequenza di riferimento nota e ottimale.

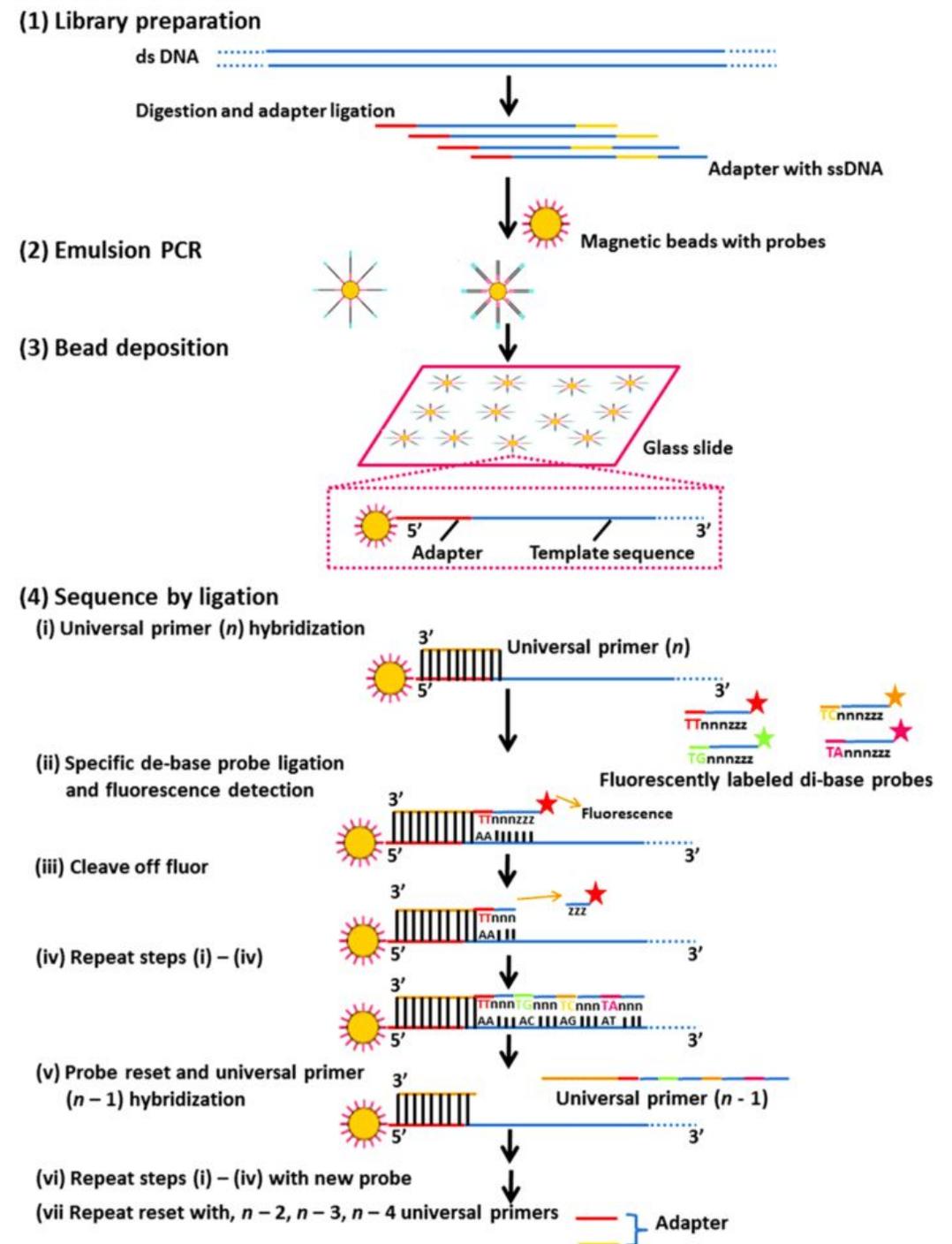


# Solid – vantaggi

## Perché è vantaggiosa la tecnologia Solid?

L'approccio di codifica basato su due basi offre una maggiore accuratezza rispetto ad altri metodi. L'errore in una lettura potrebbe non manifestarsi nella lettura successiva, offrendo una forma di correzione dell'errore integrata nel processo. Questa ridondanza nel rilevamento rende il sistema SOLiD particolarmente robusto contro gli errori di sequenziamento.

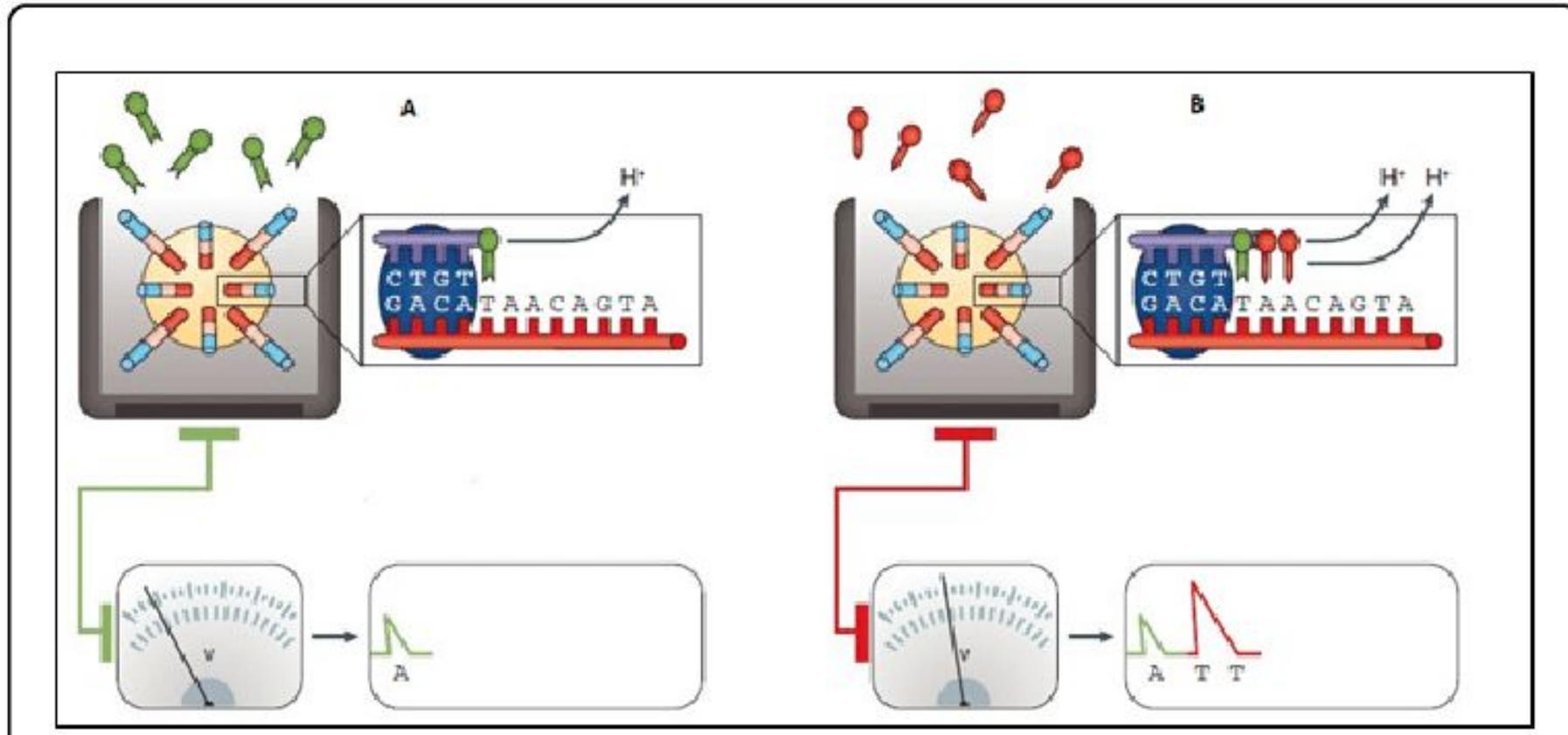
In sintesi, la codifica basata su due basi nel sistema SOLiD è una metodologia di sequenziamento in cui coppie consecutive di basi vengono rilevate simultaneamente piuttosto che singolarmente, offrendo un'accuratezza superiore nel rilevamento della sequenza.



# Ion torrent

Il sequenziamento Ion Torrent, conosciuto anche come sequenziamento a **semiconduttore**, è un metodo di sequenziamento del DNA che **non fa uso di marcatori fluorescenti**, a differenza di molte altre tecniche di sequenziamento di nuova generazione.

Questa tecnica di amplificazione si basa sulla rilevazione di ioni idrogeno rilasciati durante la sintesi del DNA.



# Ion torrent: processo di sequencing

- 1. Preparazione della Libreria:** Come molti altri metodi, il sequenziamento Ion Torrent inizia con la preparazione di una libreria di frammenti di DNA, ciascuno dei quali viene legato a una perla distinta in un processo noto come emulsione PCR.
- 2. Caricamento delle Perle:** Le perle cariche con singole molecole di DNA vengono poi caricate su una piastra a pozzetti, con una perla per pozzetto.
- 3. Sintesi del DNA e Rilevazione:** Si aggiungono nucleotidi uno alla volta nella piastra. Quando un nucleotide si incorpora in una catena di DNA in crescita, rilascia un protone (ione idrogeno). Questo rilascio provoca una piccola variazione del pH nell'ambiente circostante.
- 4. Rilevazione con Chip a Semiconduttore:** La piastra è in realtà un chip a semiconduttore. Rileva i cambiamenti di pH (che corrispondono al rilascio di ioni idrogeno) e li traduce in un segnale elettrico. Questo segnale indica che un particolare nucleotide è stato incorporato nel frammento di DNA in quel pozzetto.
- 5. Traduzione in Sequenza:** Ripetendo il processo di aggiunta di nucleotidi e rilevazione del segnale elettrico, si ottiene una sequenza di segnali che può essere tradotta in una sequenza di nucleotidi, fornendo la sequenza del frammento di DNA.

# Ion torrent: vantaggi

1. **Rapidità:** Una delle principali attrattive del sequenziamento Ion Torrent è la sua velocità. Può produrre risultati in poche ore, a differenza di altri metodi che possono richiedere giorni.
2. **Semplicità del Rilevamento:** Poiché il sistema rileva ioni idrogeno (cambiamenti di pH) invece di segnali fluorescenti, non sono necessari laser o camere specializzate, il che semplifica la strumentazione.
3. **Costi:** L'assenza di reagenti fluorescenti e la semplificazione della strumentazione possono portare a una riduzione dei costi operativi. Inoltre, il costo iniziale del sistema è in genere inferiore rispetto ad altre piattaforme di sequenziamento di nuova generazione.
4. **Flessibilità:** Il sistema Ion Torrent offre flessibilità nella dimensione dei chip, consentendo agli utenti di scegliere la capacità di sequenziamento più appropriata per il loro progetto.

# Ion torrent: limitazioni

1. **Lunghezza delle Reads:** Sebbene la lunghezza delle Reads dell' Ion Torrent sia superiore a quella di alcune piattaforme, come la tecnologia Illumina, è ancora inferiore rispetto ad altre tecnologie, come il sequenziamento di terza generazione offerto da PacBio.
2. **Errori di Omissione:** Una delle principali limitazioni del sequenziamento Ion Torrent è la tendenza agli errori in regioni di omopolimeri (cioè sequenze in cui uno stesso nucleotide si ripete molte volte di seguito). Questo perché il sistema può avere difficoltà a determinare il numero esatto di nucleotidi identici incorporati in una sequenza di omopolimeri.
3. **Capacità di Sequenziamento:** Anche se i chip più grandi possono sequenziare grandi quantità di DNA, la capacità complessiva è generalmente inferiore rispetto a piattaforme come Illumina.
4. **Affidabilità:** Alcuni utenti hanno segnalato una variabilità nella qualità dei dati tra i chip o tra le corsie sullo stesso chip, il che può influenzare la riproducibilità.

# Tecnologie innovative per il sequenziamento

Le diverse piattaforme NGS si differenziano per le modalità con cui effettuano le varie fasi. Si suddividono in 2 categorie

## **N**EXT **G**ENERATION **S**EQUENCING

Sequenziatori NGS di  
**Seconda generazione**  
☐ necessitano di  
amplificazione clonale:

- **ROCHE/454 (pyrosequencing)**
- **ILLUMINA/SOLEXA (modified Sanger)**
- **APPLIED BIOSYSTEMS SOLID (sequencing by ligation)**
- **ION TORRENT**

Sequenziatori NGS di  
**terza generazione**  
☐ non necessitano di  
amplificazione clonale:

- **PACIFIC BIOSCENCE**
- **OXFORD NANOPORE**

## Sequenziamento PACIFIC BIOSCENCE

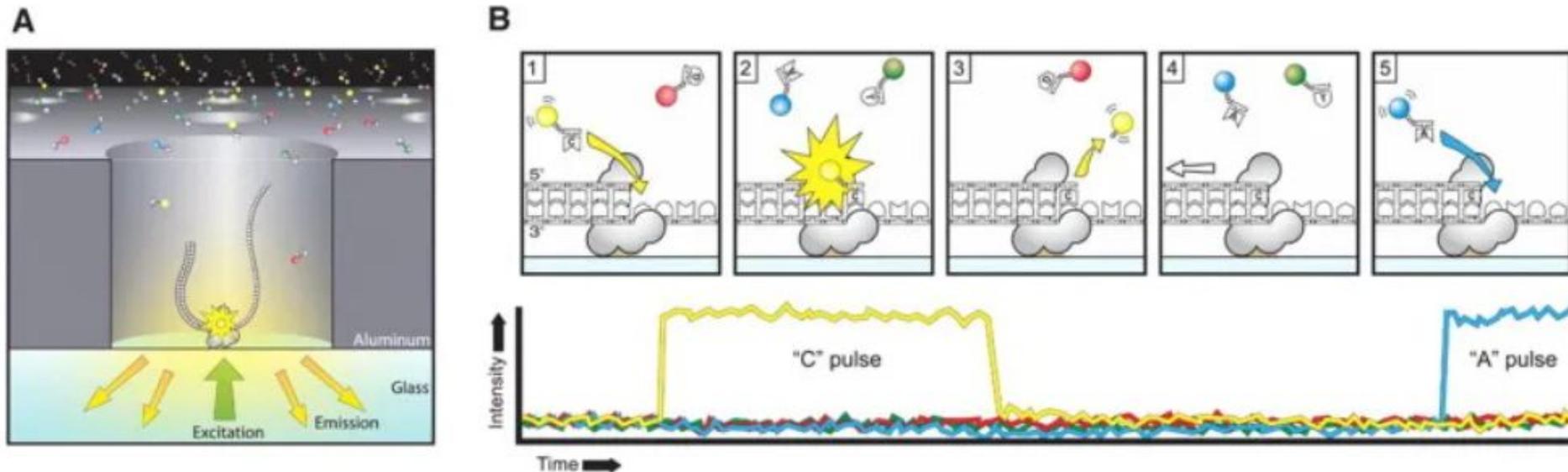
La piattaforma di *sequencing Pacific Biosciences* (spesso abbreviato come ***PacBio***) si basa su una tecnologia di sequenziamento noto come "Single Molecule Real-Time" (SMRT) sequencing. E' il metodo di sequenziamento di terza generazione più utilizzato, principalmente per il fatto che le letture ottenute con essa sono molto lunghe, anche superiori a 20.000 bp. Pensate che con questa tecnica è possibile, con una sola corsa di sequenziamento, sequenziare un intero genoma batterico.

Anche in questo caso è necessario costruire una libreria prima di procedere con il sequenziamento vero e proprio. Per la costruzione della libreria è necessario frammentare il DNA estratto da sequenziare, generalmente mediante sonicazione e, successivamente, aggiungere adattatori circolari alle estremità dei frammenti di DNA ottenuti. Questi sono chiamati ***DNA SMRTbell***.

# PACIFIC BIOSCIENCE

Caratteristiche uniche che distinguono questo metodo da altre piattaforme di sequenziamento:

1. **Tecnologia Single Molecule:** A differenza di molte tecnologie di sequenziamento che richiedono l'amplificazione del DNA, la piattaforma **SMRT** di PacBio sequenzia singole molecole di DNA in tempo reale. Questo elimina potenziali errori e bias introdotti dall'amplificazione.
2. **Zerodrome:** Il cuore della tecnologia **SMRT** è una struttura chiamata "**Zerodrome**". Si tratta di una piccola camera a forma di pozzo in cui si svolge la reazione di sequenziamento. Una sola molecola di polimerasi DNA viene ancorata al fondo del Zerodrome e una singola molecola di DNA da sequenziare viene fatta passare attraverso essa.



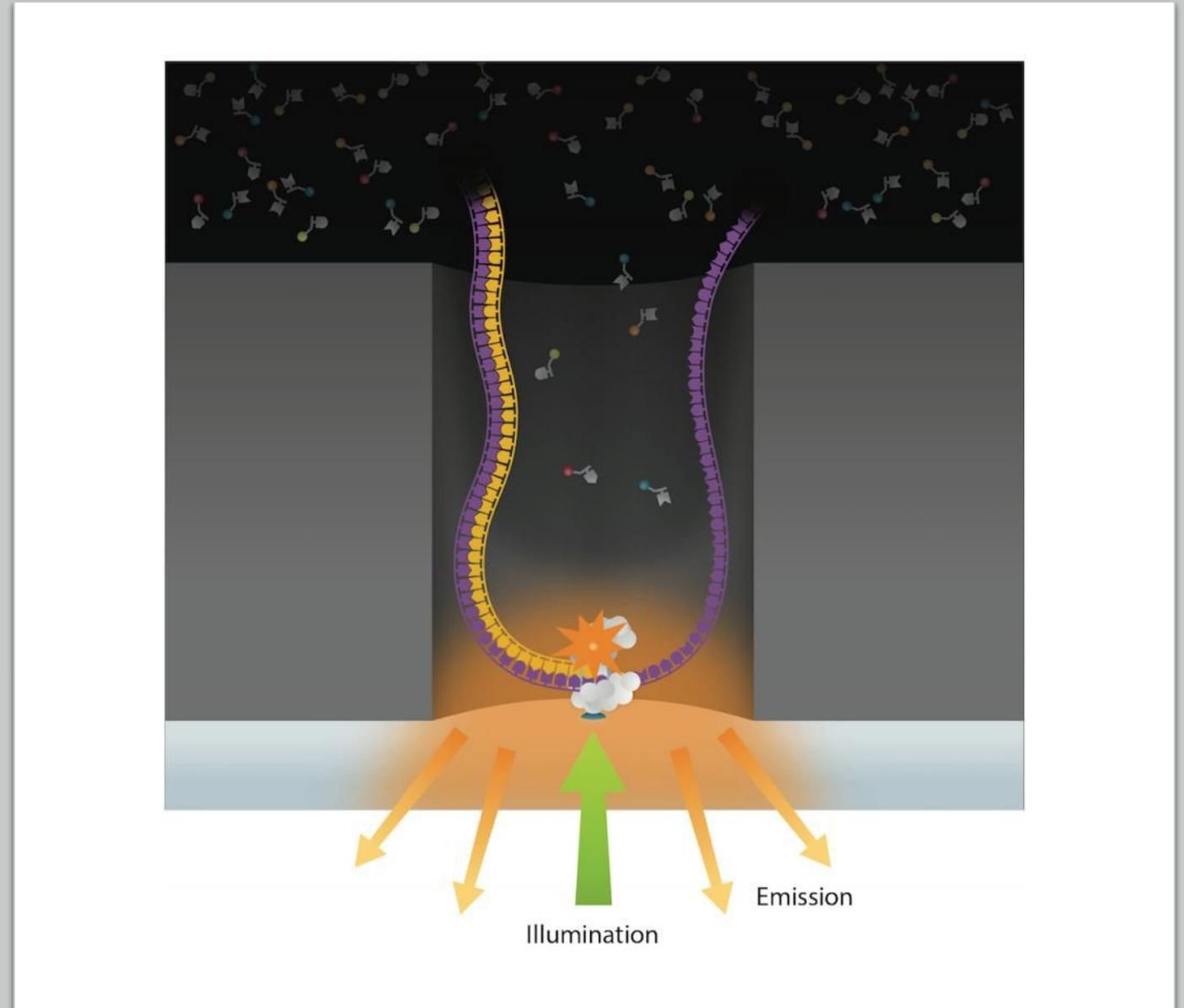
## PACIFIC BIOSCENCE

La reazione di sequenziamento avviene su di un chip noto con il nome di **SMRT cell**. Questo supporto è costituito da piccolissime unità di sequenziamento chiamate “ZeroMode Wave Guide” (ZMW). Ogni **SMRT cell** contiene all’incirca 150.000 **ZMW**



# PACIFIC BIOSCIENCE

- All'interno di ogni unità è presente una DNA polimerasi ancorata al fondo, a questo vengono aggiunti dei deossiribonucleotidi marcati con fluorofori legati ai gruppi fosfato.
- Durante l'incorporazione dei deossiribonucleotidi al nostro campione, il fosfato viene tagliato e con esso anche il fluoroforo che rilascia un impulso di luce. Questo viene rilevato da uno specifico detector e il software le rielabora convertendole nella sequenza nucleotidica rispettiva.

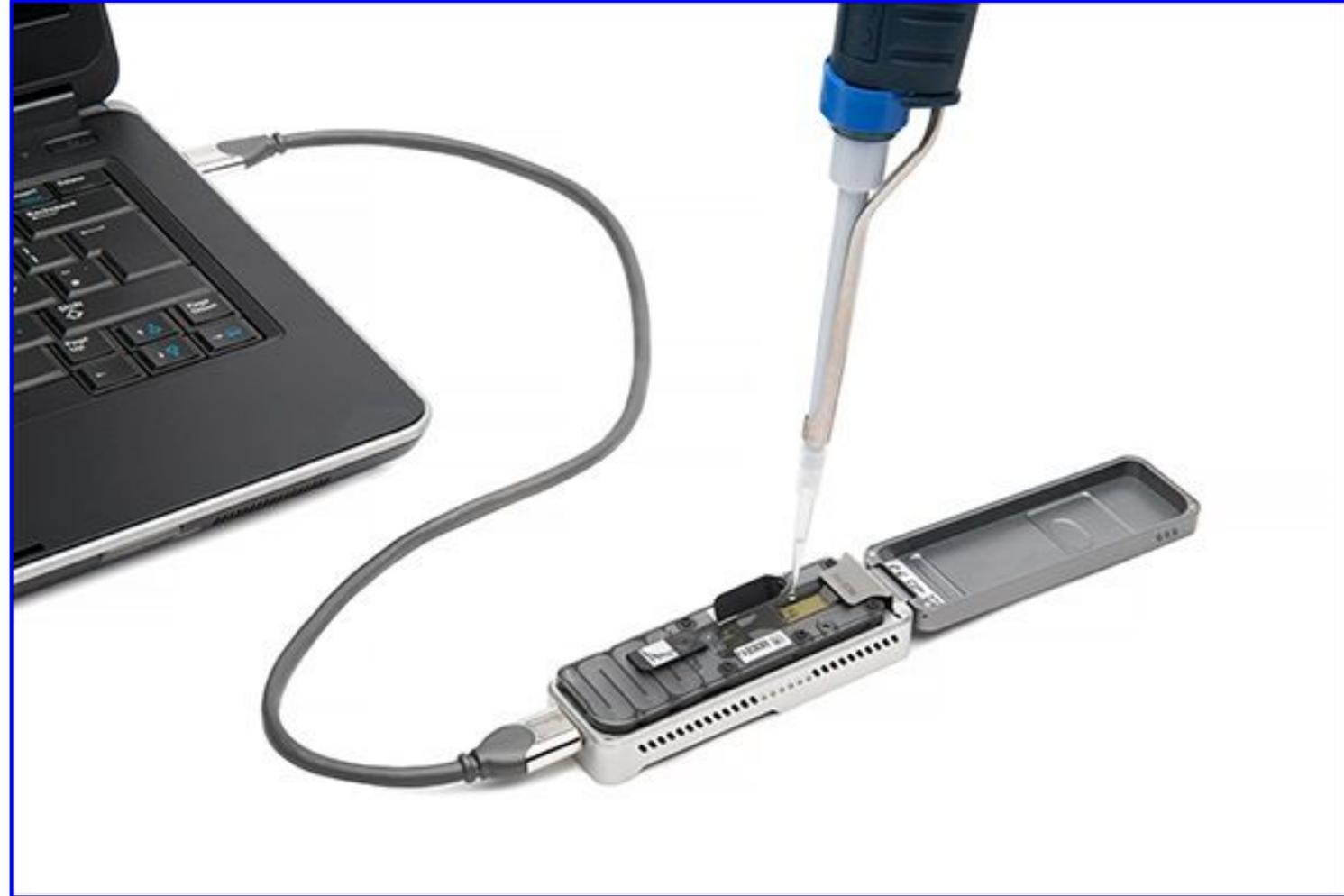


- 3. Sequenziamento in Tempo Reale:** Durante il processo di sequenziamento, ogni nucleotide che viene incorporato da una molecola di DNA polimerasi viene rilevato in tempo reale. Ogni nucleotide è legato a una molecola fluorescente chiamata fluoroforo. Quando un nucleotide viene incorporato, il fluoroforo emette una luminescenza che viene rilevata e registrata.
- 4. Long reads:** Una delle caratteristiche distintive della piattaforma PacBio è la sua capacità di produrre letture estremamente lunghe, spesso superiori a 20.000 basi, rendendola ideale per applicazioni come l'assemblaggio del genoma o lo studio di regioni genomiche complesse.
- 5. Rilevazione Diretta di Modifiche Epigenetiche:** Poiché PacBio sequenzia singole molecole di DNA in tempo reale, è possibile rilevare modifiche epigenetiche come la metilazione direttamente dal processo di sequenziamento, senza necessità di trattamenti chimici o procedimenti aggiuntivi.
- 6. Errore Casuale:** Sebbene il tasso di errore intrinseco per singola lettura possa essere superiore rispetto ad altre tecnologie di sequenziamento, l'errore è in gran parte casuale, il che significa che può essere corretto facilmente attraverso il sequenziamento profondo (*deep sequencing*).

## PACIFIC BIOSCIENCE

Si tratta di un sequenziatore portatile, di piccole dimensioni, alimentato tramite una semplice porta USB e caratterizzato da un'elevatissima processività.

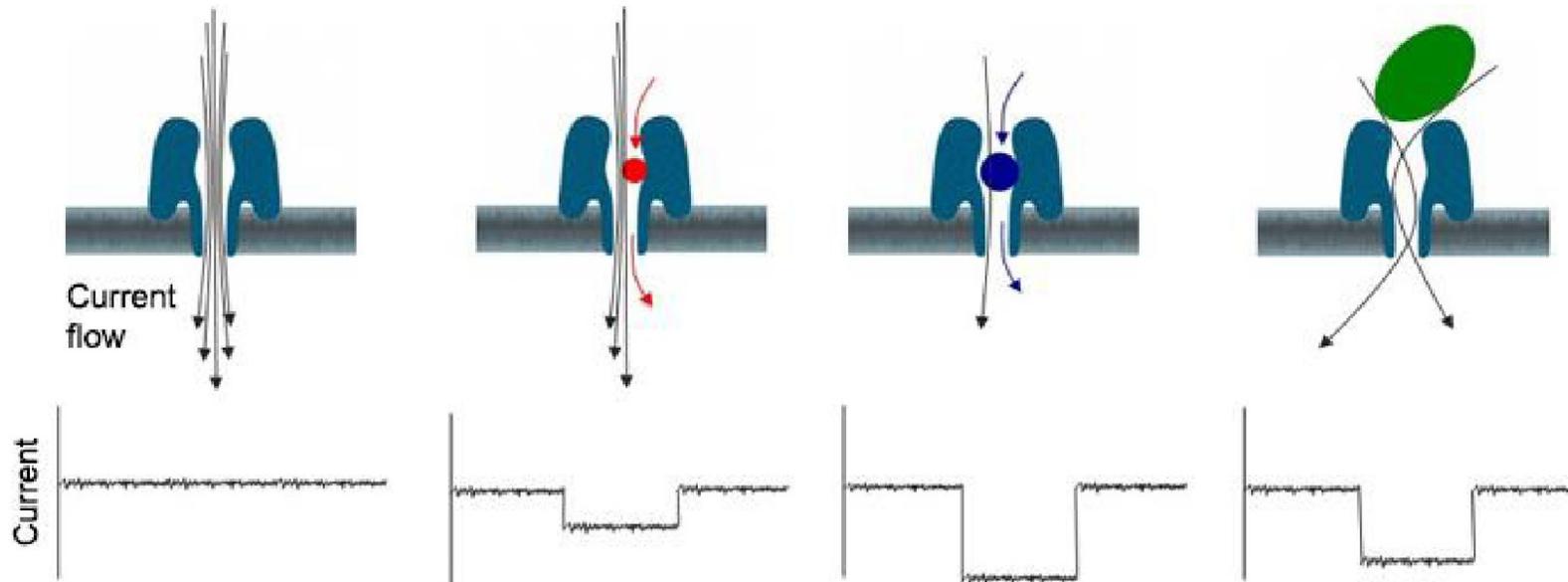
In sintesi, la tecnologia *SMRT* di Pacific Biosciences offre vantaggi unici, tra cui letture lunghe (*long reads*) e rilevazione diretta di modifiche epigenetiche. Sebbene abbia un tasso di errore intrinseco più elevato per singola lettura rispetto ad altre piattaforme, la natura casuale degli errori e la lunghezza delle letture la rendono particolarmente potente per certe applicazioni genomiche.



# Oxford Nanopore Technology

Alla base del MinION vi è il **sequenziamento a nanopori**. In questa tecnica si utilizzano dei nanopori inseriti in una membrana lipidica o sintetica, dove formano dei canali attraverso cui le molecole possono passare. Ai capi della membrana è applicata una differenza di potenziale che determina la creazione di una corrente elettrica.

Durante il transito all'interno del nanoporo, la molecola blocca il passaggio degli ioni causando una **perturbazione di corrente**.

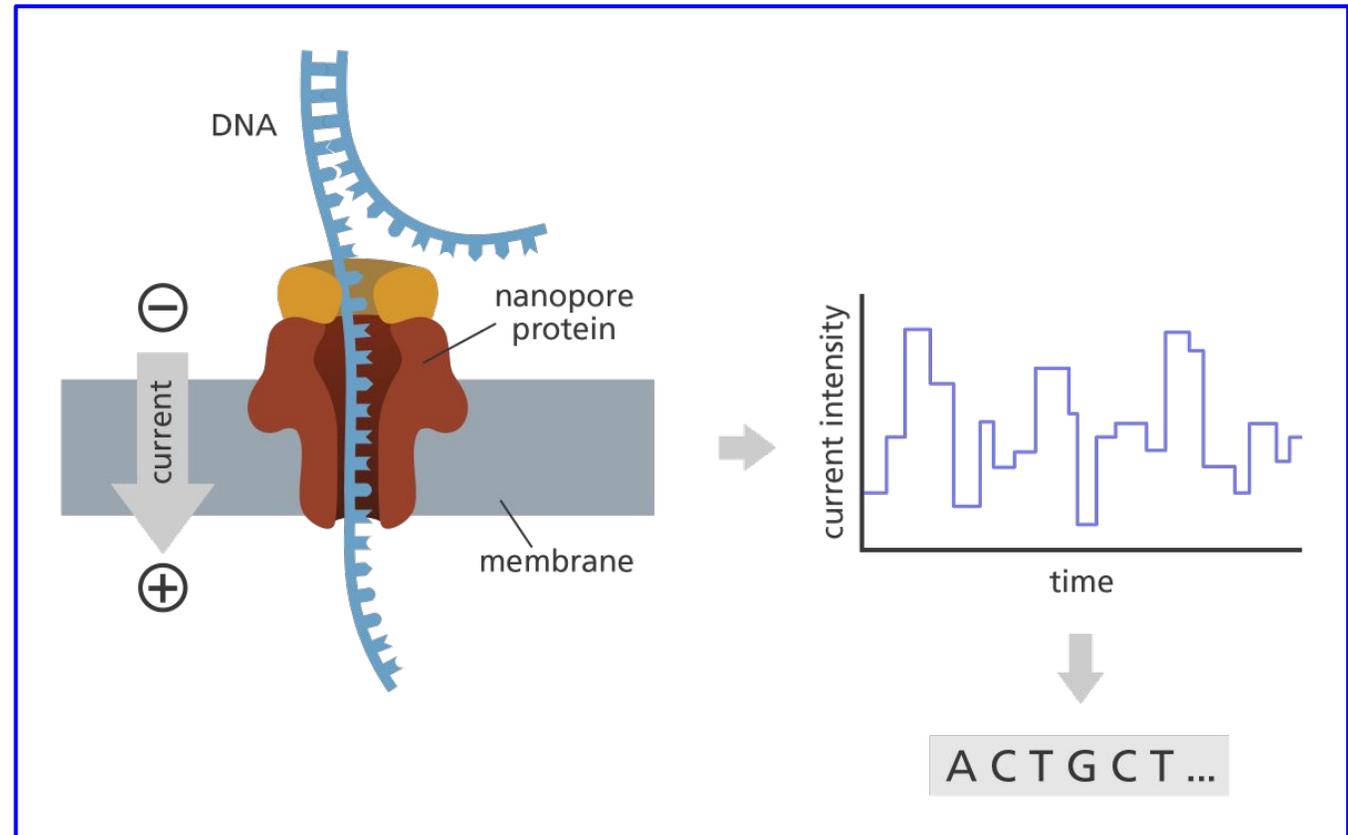


<https://www.youtube.com/watch?v=RcP85JHLmnl&list=PLxpxXZj-IgXyR0it4QjoBETC5JeEanW-o&t=1s>

# Oxford Nanopore Technology

**Cos'è un Nanoporo?:** Un nanoporo è un piccolo foro, con un diametro dell'ordine dei nanometri, che viene creato in una membrana sottile. Nel contesto del sequenziamento, una singola molecola di DNA o RNA passa attraverso questo nanoporo.

**Rilevazione Elettrica:** Mentre una molecola di DNA o RNA passa attraverso il nanoporo, modifica le proprietà elettriche del nanoporo. Questi cambiamenti nelle correnti elettriche possono essere misurati e registrati in tempo reale. Ogni nucleotide (A, T, C o G nel DNA) produce un segnale elettrico distintivo.



# Oxford Nanopore Technology

**Senza Bisogno di Amplificazione:** Una delle grandi caratteristiche della tecnologia nanopore è che può sequenziare direttamente molecole di DNA o RNA senza la necessità di un'ampia fase di amplificazione precedente, il che riduce il bias (distorsioni sistematiche nella rappresentazione dei dati sequenziati rispetto al genoma o al campione originale) e gli errori associati all'amplificazione.

**Long reads:** Come la tecnologia SMRT di Pacific Biosciences, anche il sequenziamento nanopore è noto per produrre letture estremamente lunghe, che possono estendersi per dec

**Portabilità:** Un grande vantaggio del sequenziamento nanopore è la sua portabilità. Il MinION, uno dei dispositivi di Oxford Nanopore, è delle dimensioni di una chiavetta USB e può essere collegato a un laptop per il sequenziamento sul campo.

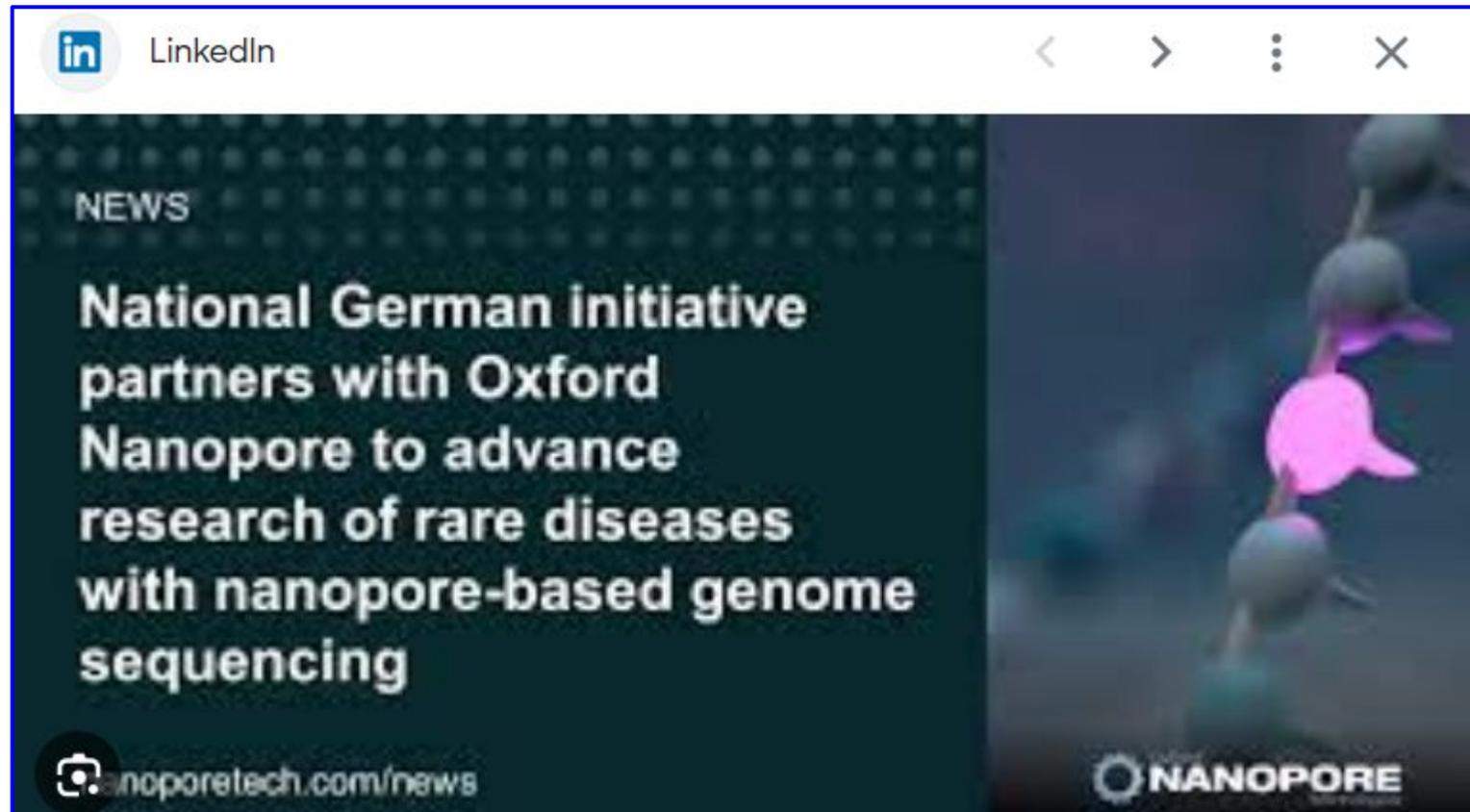


# Oxford Nanopore Technology

**Applicazioni Versatili:** Oltre al sequenziamento del DNA, la piattaforma nanopore può essere utilizzata per sequenziare l'RNA, identificare modifiche epigenetiche come la metilazione del DNA e analizzare proteine.

**Errori:** Come molte tecnologie di sequenziamento di terza generazione, il sequenziamento nanopore ha un tasso di errore intrinseco più elevato per lettura singola rispetto alle piattaforme di sequenziamento di seconda generazione. Tuttavia, poiché gli errori sono in gran parte casuali, possono essere corretti attraverso il sequenziamento profondo e l'uso di algoritmi appropriati.

In sintesi, la tecnologia di sequenziamento nanopore di Oxford offre la capacità di sequenziare direttamente DNA o RNA con dispositivi portatili e produce letture lunghe, rendendola adatta a una vasta gamma di applicazioni, dalle ricerche sul campo alla genetica clinica.



# Sequenziamento di terza generazione

I sequenziatori di terza generazione, come quelli basati sulla tecnologia Nanopore (Oxford Nanopore Technologies) e Pacific Biosciences (PacBio), hanno introdotto importanti innovazioni rispetto ai sequenziatori di seconda generazione (come Illumina, Ion Torrent e Roche 454). Ecco alcune delle differenze chiave e le caratteristiche distintive:

## Lunghezza delle reads:

- **Terza Generazione:** Entrambe le piattaforme Nanopore e PacBio possono produrre letture molto lunghe, spesso superiori a decine di migliaia di basi. Questo è particolarmente utile per risolvere regioni complesse del genoma e per assemblare genomi *de novo*.
- **Seconda Generazione:** Producono *reads* più corte, tipicamente da alcune decine a qualche centinaio di basi.

## Velocità del Sequenziamento:

- **Terza Generazione:** Entrambe Nanopore e PacBio offrono sequenziamento in tempo reale. In particolare, Nanopore permette la rilevazione delle basi mentre il DNA passa attraverso la nanopore in tempo reale.
- **Seconda Generazione:** La rilevazione delle basi avviene in cicli, e non è in tempo reale.

# Sequenziamento di terza generazione

## Complessità e costi:

- **Terza Generazione:** Gli apparecchi come il MinION di Oxford Nanopore sono portatili, del formato di una chiavetta USB, e relativamente economici, rendendo il sequenziamento accessibile anche in ambienti fuori dal laboratorio tradizionale.
- **Seconda Generazione:** Gli strumenti sono in genere più grandi, richiedono una struttura di laboratorio e possono avere costi operativi più elevati.

## Errori:

- **Terza Generazione:** Pur offrendo *reads* lunghe, le tecnologie di terza generazione hanno tassi di errore intrinsecamente più alti rispetto ai sequenziatori di seconda generazione. Tuttavia, poiché gli errori sono spesso casuali, possono essere mitigati con una copertura sufficientemente alta.
- **Seconda Generazione:** Hanno un tasso di errore per base generalmente più basso.

# Sequenziamento di terza generazione

## Applicazioni:

- **Terza Generazione:** Particolarmente utili per l'analisi di regioni genomiche complesse, strutturazioni di genomi, rilevazione di modifiche epigenetiche direttamente durante il sequenziamento (come la metilazione del DNA nel caso della tecnologia Nanopore).
- **Seconda Generazione:** Ampiamente utilizzati per un'ampia varietà di applicazioni, tra cui il sequenziamento di genoma intero, la trascrittomica, e studi di espressione genica.

In conclusione, mentre i sequenziatori di seconda generazione hanno rivoluzionato la genomica rendendo possibile l'analisi ad alta throughput di campioni multipli, i sequenziatori di terza generazione stanno ulteriormente ampliando le capacità del sequenziamento del DNA, consentendo una maggiore risoluzione di regioni complesse e la rilevazione di eventi epigenetici.

# Vantaggi delle piattaforme next-gen

- ❖ **Non c'è necessità di sub-clonazione, né di utilizzo di cellule batteriche**
  - abolizione di *bias* di clonazione
  - rapidità nel preparare le librerie (non c'è colony picking)
- ❖ **Ciascuna sequenza proviene da una molecola di DNA unica**
  - rilevazione di varianti rare
- ❖ **Fornisce una eccezionale risoluzione per molti tipi esperimenti**  
(analisi di espressione(RNAseq), sequenziamento di DNA immunoprecipitato (ChIPseq), analisi di medi/grandi genomi....)
- ❖ **Rivoluzionaria diminuzione del costo e del tempo per generare dati di sequenza (lavorano in multi-parallelo)**

# Svantaggi delle piattaforme next-gen

- ❖ **Sono prodotte sequenze più corte relativamente alle sequenze da sequenziatori capillari (metodo Sanger)**
  - enorme difficoltà nell'analisi dei dati; richiesto un grande sforzo di programmazione per costruire nuovi algoritmi
  - è necessario ri-parametrizzare l'accuratezza della procedura di chiamata delle basi
- ❖ **La mole enorme di dati 'traumatizza' le infrastrutture informatiche**
  - da 10 Gb a diversi Tb di dati grezzi prodotti per corsa
  - il processamento delle *read* tramite *pipeline* informatiche richiede molta capacità di calcolo (CPU)
  - è necessario prendere accurate decisioni su cosa salvare e cosa cancellare

## Il formato FASTQ

Formato standard di output di strumenti di sequenziamento NGS. **FASTQ format** formato di testo che include sia la sequenza (in genere nucleotidica) che la qualità di ogni base (score).

**FASTQ Format (Illumina Example)**

The diagram shows a FASTQ file with four records. Callouts identify the following fields:

- Read Record Header:** The first line of each record, containing metadata like Flow Cell ID, Lane, Tile, and Barcode.
- Flow Cell ID:** The first part of the header (e.g., @DJG84KN1).
- Lane:** The lane number (e.g., :2).
- Tile:** The tile coordinates (e.g., :1101:12432:5554).
- Barcode:** The barcode number (e.g., 1:N:0).
- Read Bases:** The sequence of nucleotides (e.g., AGTCAA).
- Read Quality Scores:** The quality scores for each base, represented by a '+' sign followed by a string of characters (e.g., BCCFFFDH...).
- Separator (with optional repeated header):** A '+' sign separating the bases from the quality scores.

```
@DJG84KN1:272:D17DBACXX:2:1101:12432:5554 1:N:0:AGTCAA
CAGGAGTCTTCGTACTGCTTCTCGGCCTCAGCCTGATCAGTCACACCGTT
+
BCCFFFDH...
@DJG84KN1:272:D17DBACXX:2:1101:12454:5610 1:N:0:AG
AAACTCTTACTACATCAGTATGGCTTTTAAAACCTCTGTT...
+
@@@DD?DDHFDHFHEHIIHIIIIIBBGEBHIEDH=EEHI>FDABHHFGH2
@DJG84KN1:272:D17DBACXX:2:1101:12438:5704 1:N:0:AG
CCTCCTGCTTAAAACCCAAAAGGTCAGAAGGATCGTGAGGCCCCCGCTTTC
+
CCCFHHGH...
@HGIJJJJIIJGIGIHIJJJIIIIJJ
```

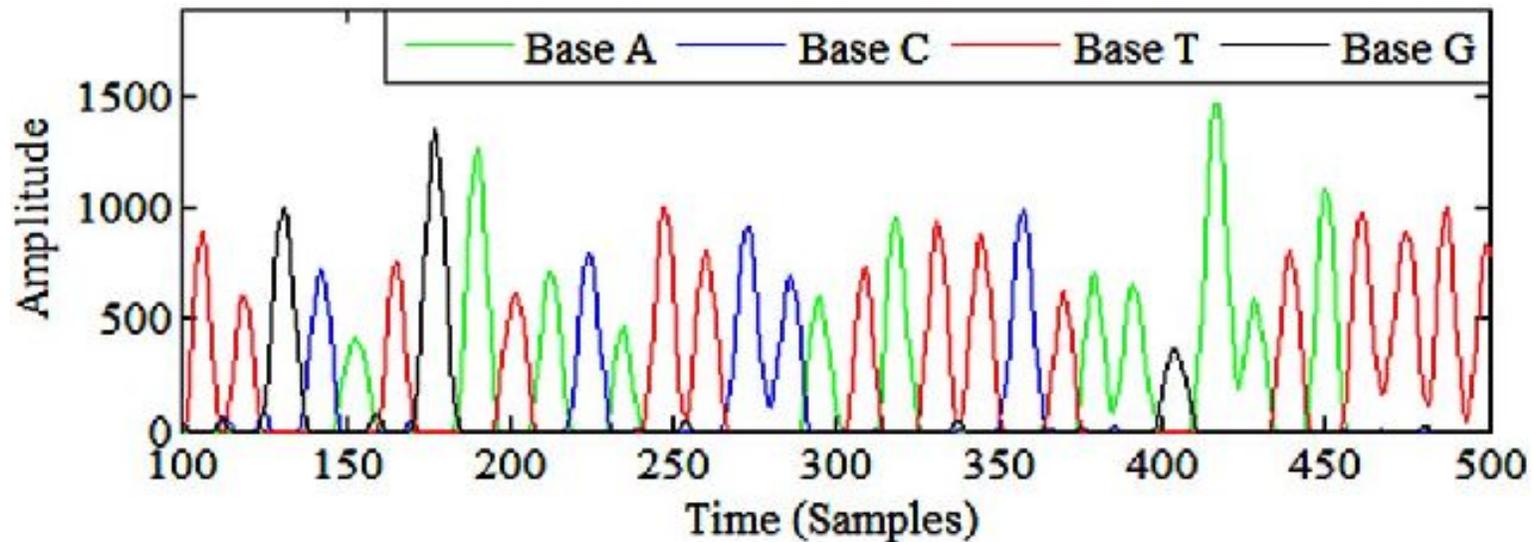
NOTE: for paired-end runs, there is a second file with one-to-one corresponding headers and reads

- È un formato di puro testo
- disegnato da Wellcome Trust Sanger Institute per associare ad una sequenza la qualità di ogni sua singola base
- un file in formato FASTQ ha estensione: **\*.fq** oppure **\*.fastq**



## Base calling

Un esperimento di sequenziamento produce in generale un cromatogramma da cui viene derivata la sequenza delle basi tramite il cosiddetto processo (software) di base calling.

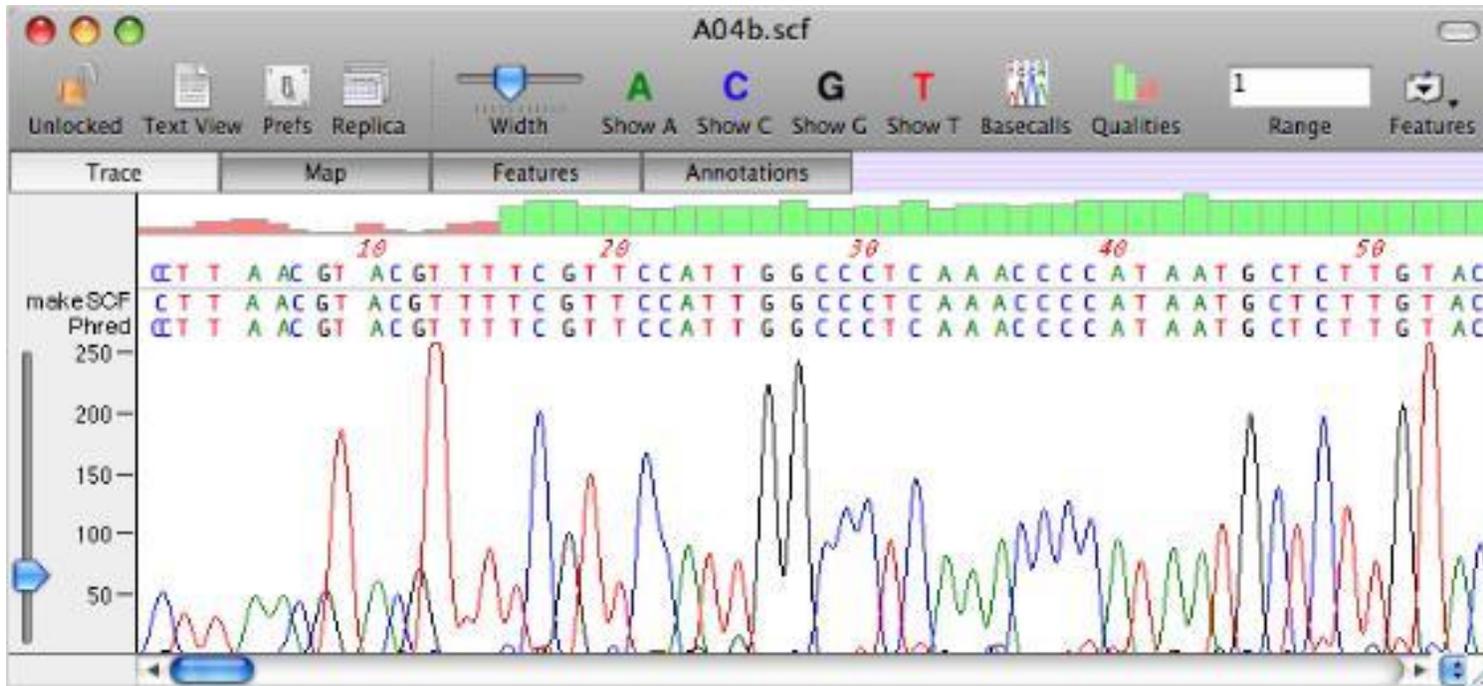


La lettura di un cromatogramma permette l'associazione di un indice di qualità ad ogni base chiamata. Oltre alla sequenza di basi, si ottiene dunque anche la corrispondente sequenza di indici di qualità. L'indice di qualità più usato attualmente è il **Phred Quality Score**.

# Phred Quality Score

Sviluppato con il software di base calling Phred (primi anni Novanta - gruppo di Phil Green della Washington University).

- Phred è un algoritmo che prende le informazioni del cromatogramma da una corsa di sequenziamento automatizzato e rivaluta i picchi per produrre una "identificazione di base" che di solito è significativamente più accurata della chiamata originale.



Qualità  $q$  della base  $b$  in una sequenza:

$$q = -10 \log_{10} (p)$$

$p$  = probabilità che  $b$  sia errata  
( $0 \leq p \leq 1$ )

NB:  $q$  calcolato viene arrotondato all'intero più vicino

- Oltre a ricalcolare i residui, phred aggiunge anche informazioni sul punteggio di qualità a ciascun residuo. Questo è un valore logaritmico da 0 a 99 dove un valore di 10 indica che c'è una possibilità su 10 che la chiamata sia in errore, un punteggio di 20 indica una possibilità su 1 su 100 che la chiamata sia in errore, un punteggio di 30 indica 1 possibilità su 1.000 di errore, ecc.

## Phred Quality Score

Il "quality score" o punteggio di qualità, espresso come valore *Phred*, rappresenta la qualità delle basi sequenziate. La sua formula è:

$$Q = -10 \log_{10} P.$$

dove *P* è la probabilità di errore.

Il range teorico del punteggio di qualità (Q) può iniziare da 0 (che indica una probabilità di errore del 100% o una certezza di 0% per la base sequenziata) a infinito (che indica una probabilità di errore di 0% o una certezza del 100% per la base sequenziata).

Tuttavia, nella pratica, i valori *Phred* sono solitamente compresi in un intervallo più ristretto.

Ad esempio, per il formato FASTQ che usa il codice ASCII:

- Con il sistema di codifica ASCII 33 (spesso associato al formato Sanger FASTQ), il range di Q va da 0 a 93.
- Con altri sistemi di codifica (ad esempio, Illumina 1.3+), il range di Q potrebbe iniziare da 0 e avere un valore massimo di circa 41-42.

## Phred Quality Score

La relazione del Phred score può anche essere scritta così:

$$P = 10^{\frac{-Q}{10}} .$$

Ad esempio, se Phred assegna un punteggio di qualità pari a 30 a una base, le probabilità che questa base venga chiamata in modo errato sono 1 su 1000.

# Phred Quality Score

## Approfondimento



Esercizio 1:

calcolare il **phred value q** di una base che ha una probabilità dell'1% di essere errata (cioè corretta al 99%)

Esercizio 2:

calcolare il **phred value q** di una base che ha una probabilità del 100% di essere errata

Esercizio 3:

calcolare il **phred value q** di una base che ha una probabilità dello 0% di essere errata (cioè corretta al 100%)

Esercizio 4:

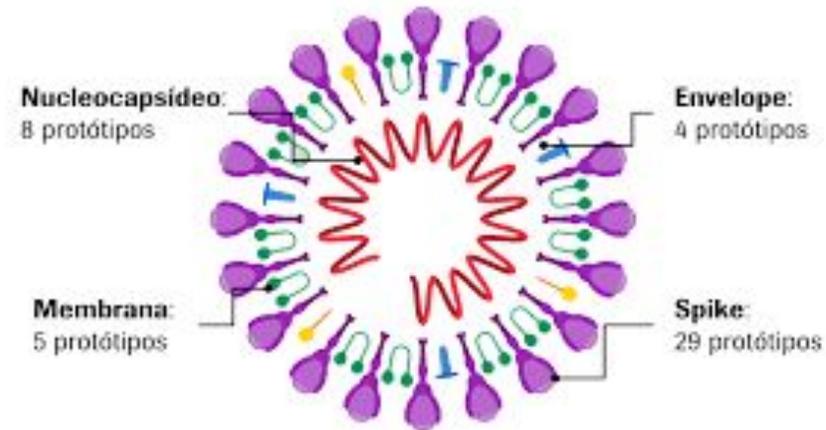
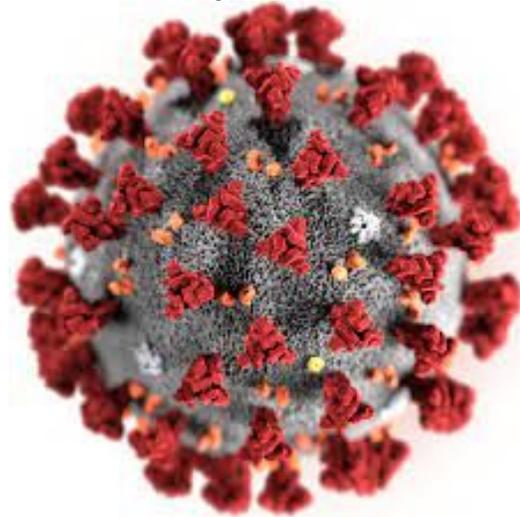
calcolare il **phred value q** di una base che ha una probabilità del 75% di essere corretta

NOTA BENE: una base con  $q$  maggiore o uguale a 50 è considerata praticamente corretta

## Sequenziamento e SARS-CoV-2

Evidenziamo l'utilità degli approcci basati sul sequenziamento del genoma di un patogeno nella risposta in tempo (quasi) reale a una gravissima crisi di salute pubblica.

A cavallo tra il 2019 e il 2020, un gruppo di scienziati, ha applicato una particolare tecnica di sequenziamento, che consente di determinare la sequenza dell'RNA di tutte le specie viventi presenti in un campione biologico/ambientale (sequenziamento metagenomico del RNA totale) per analizzare un campione di fluido di lavaggio broncoalveolare (BALF) ottenuto da un singolo paziente a **Wuhan**, in **Cina**, dove erano stati segnalati diversi casi di gravi infezioni respiratorie



Tramite l'applicazione di strumenti bioinformatici, **Wu e colleghi** sono riusciti a ricostruire dai dati di sequenziamento il genoma del potenziale agente patogeno, un nuovo coronavirus successivamente denominato **SARS-CoV-2**. Nei primi giorni del gennaio 2020, la **sequenza completa del genoma virale è stata depositata in banca dati** e resa disponibile all'intera comunità scientifica. La ricostruzione della sequenza del genoma del nuovo agente patogeno ha facilitato lo sviluppo di test diagnostici molecolari rapidi in tutto il mondo.