

Illumina

Illumina sfrutta lo stesso principio di Roche454, cioè si generano un ampio numero di colonie uniche generate da una polimerasi (*colonies*), che vengono poi sequenziate in parallelo.

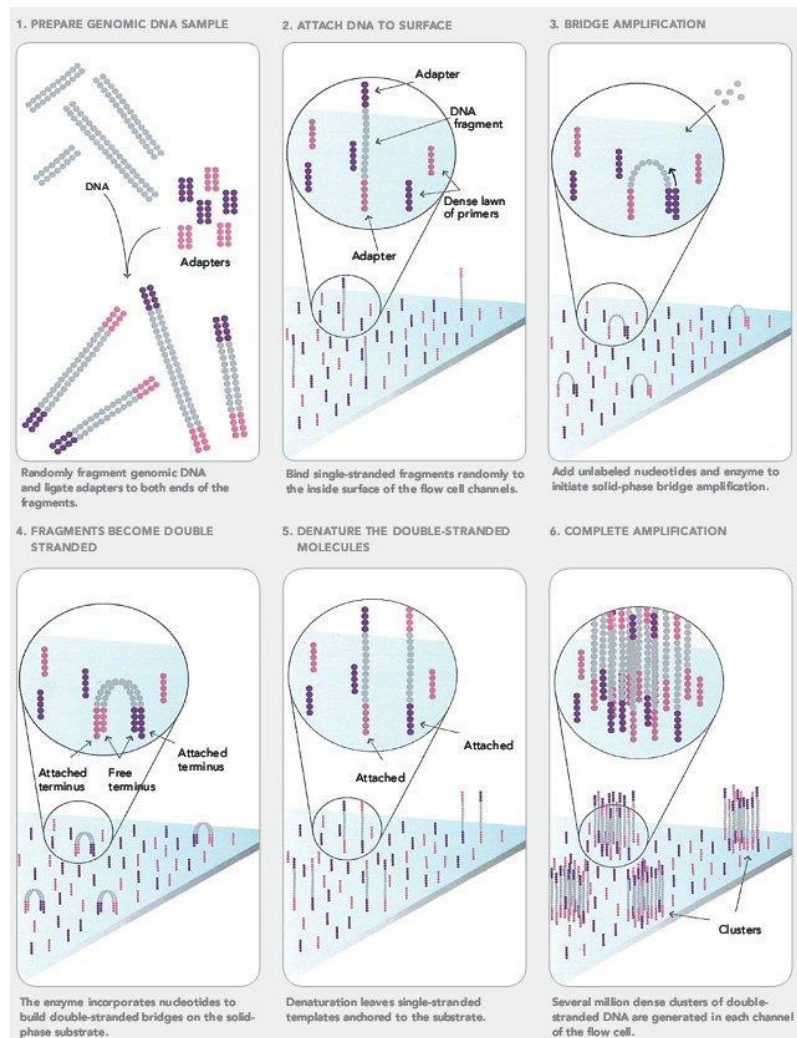
Queste reazioni avvengono sulla superficie di un *flow cell*, che presenta sulla superficie diverse lanes, ognuna delle quali può produrre milioni di sequenze di lunghezza compresa tra le 35 e le 300 coppie di basi.

Il metodo Illumina consta di 4 passaggi:

1. Preparazione della libreria di DNA
2. Bridge PCR (cluster generation)
3. Reazione di sequenziamento
4. Analisi dei dati

Preparazione della libreria

Il DNA di interesse viene frammentato in frammenti di 500 – 700 bp. È importante che la frammentazione sia random.

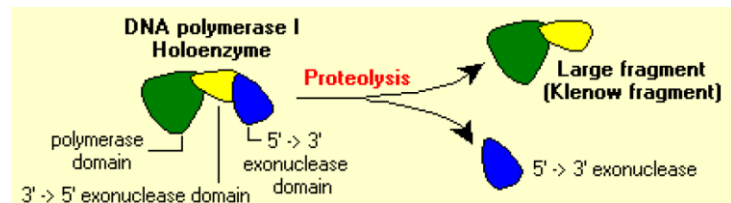


In seguito alla frammentazione del DNA potranno aversi dei frammenti le cui estremità non sono blunt, ma sporgenti. Queste devono essere rese blunt, e ciò si può fare utilizzando la T4 DNA polimerasi o il frammento di Klenow.

Frammento di Klenow

Il Frammento di Klenow è il frammento più grande proveniente dalla proteolisi (con subtilisina) della DNA polimerasi I di E.coli. Questo frammento mantiene l'attività polimerasica 5'-3' e l'attività esonucleasica 3'-5', ma perde l'attività esonucleasica 5'-3'.

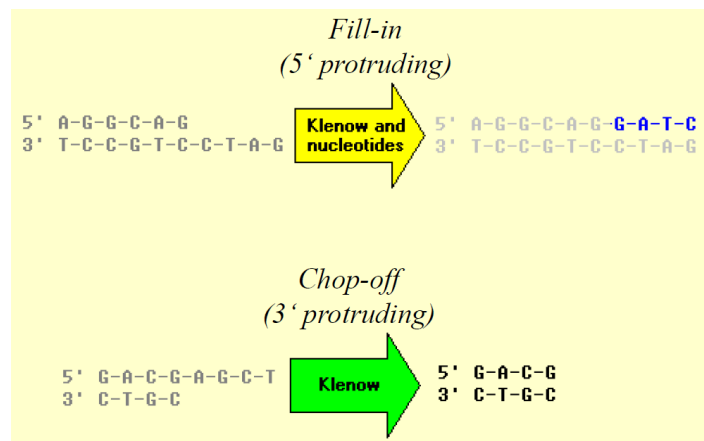
Per funzionare necessita di un template, un innesco 3'OH e dNTPs. Svolge una attività di **fill-in** rispetto alle estremità blunt, quindi rende le estremità blunt inserendo i nucleotidi mancanti nell'estremità corta.



T4 DNA polimerasi

Ha attività polimerasica 5'-3' ed attività esonucleasica 3'-5'. Non ha attività esonucleasica 5'-3'.

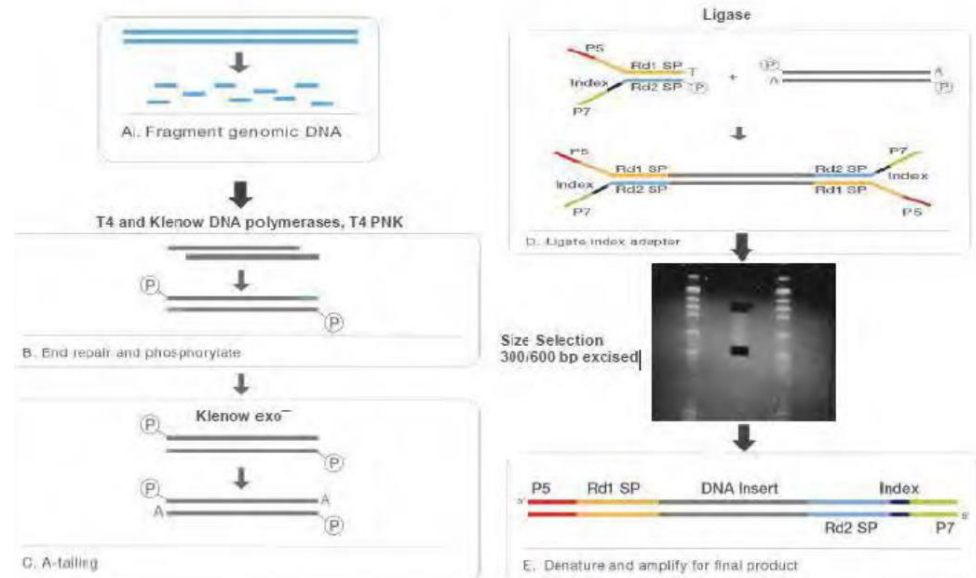
Anche questa polimerasi è in grado di rendere blunt le estremità sporgenti. Può svolgere sia attività di **fill-in** (in presenza di nucleotidi) che di **chop-off**.



Oltre a rendere blunt le estremità, si usa l'enzima T4 polinucleotide kinasi T4 PNK che catalizza la fosforilazione all'estremità 5' del DNA.

In base agli adattatori utilizzati può essere necessario aggiungere un nucleotide all'estremità 5'. Si può fare grazie all'enzima *deossinucleotidil transferasi terminale*, che lega appunto un nucleotide alle estremità 3' delle molecole di DNA.

A ciascun frammento vengono ligati 2 adattatori, creando così una libreria genomica.



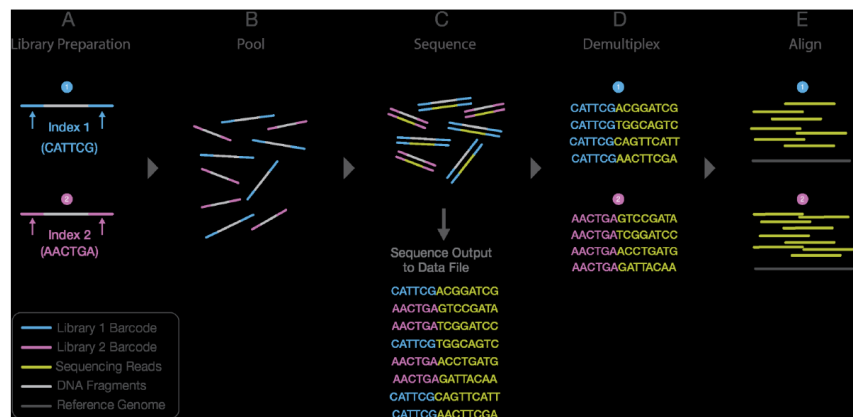
Gli adattatori presentano diverse regioni. Ogni adattatore possiede una sequenza tra la P5 e la P7. Queste sono sequenze riconosciute dai primer per effettuare la PCR, e questi primer sono situati sulla flow-cell.

Dopo aver legato gli adattatori vengono selezionati mediante elettroforesi i frammenti di dimensione adatta al sequenziamento, tra le 300 e le 600 bp.

Le molecole così ottenute sono così composte:

- Un inserto di DNA
 - Due adattatori:
 - Alle estremità ci sono le sequenze p5 e p7, riconosciute dai primer adesi alla flow cell.
 - A queste seguono delle sequenze riconosciute dai primer per il sequenziamento:
 - Alla regione P5 segue la sequenza Rd1 SP
 - Alla regione P7 segue la sequenza Rd2 SP
 - La regione P7 lega una sequenza *index*.
- Gli index sono dei barcode sotto forma di sequenza nucleotidiche, e permettono di sequenziare più librerie simultaneamente in un solo esperimento.

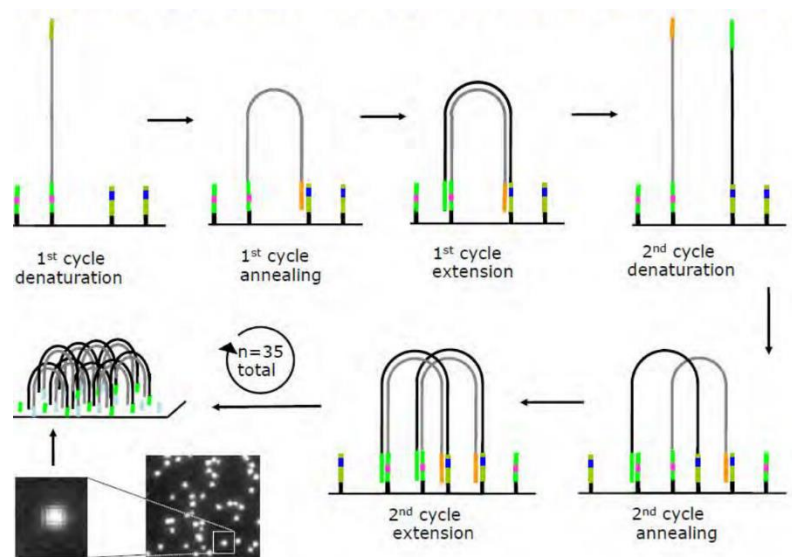
Esistono comunque metodiche più moderne che permettono di effettuare i passaggi di frammentazione in un unico step, che prende il nome di *tagmentation*, che aumenta l'efficienza del processo di preparazione della libreria.



Bridge PCR

La libreria viene denaturata e fatta fluire nella flow cell. La flow cell ha adesi dei primer complementari agli adattatori precedentemente legati al DNA. I frammenti di DNA della libreria si legheranno ai primer con una delle estremità.

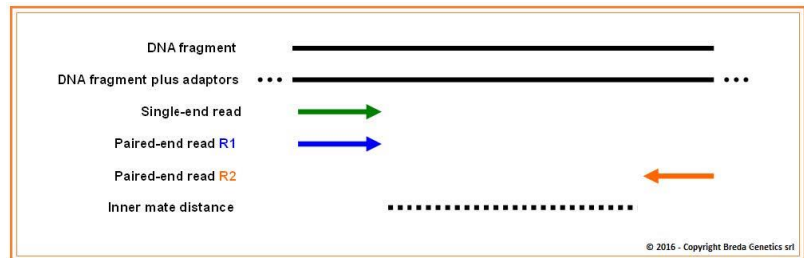
Avvenuta l'ibridazione col primer, il frammento di DNA può essere duplicato, e con esso anche l'adattatore alla estremità opposta. In seguito a denaturazione il frammento legato al primer sulla flow cell diventa un ssDNA e con la sua estremità non legata alla flow cell può legare un altro primer sulla flow cell, formando un ponte tra i due primer legati alla flow cell. A questo punto si può fare una seconda estensione del primer, con sintesi di un nuovo filamento.



In seguito a cicli di denaturazione, annealing con formazione di bridge ed estensione, si ottengono isole di cloni di una singola sequenza, le *polonies*.

Sequenziamento

La sequenza di DNA di sintesi che viene prodotta durante la reazione di sequenziamento prende il nome di **reads**. Il sistema illumina può sequenziare sia *single-end reads* che *paired-end reads*, cioè da una sola estremità o da entrambe le estremità della sequenza. Di solito, nel sequenziamento *paired-end reads*, le reads ottenute non sono né complementari né sovrapponibili, tranne che con il sistema **MiSeq** di Illumina.



Per effettuare la *paired-end reads*, dopo la prima amplificazione si usa un enzima che effettua un nick su una delle due sequenze di DNA, che in seguito a denaturazione si stacca dalla flow-cell e va in soluzione. Si bloccano le estremità 3'OH (CERCA!!!) così non ci può essere l'appaiamento con il filamento rimasto attaccato alla flow-cell. Sui filamenti che restano attaccate alla flow-cell viene fatto il sequenziamento.

Successivamente si liberano le estremità 3' e si effettua nuovamente la sintesi del filamento complementare. Si usa l'enzima di nick per tagliare sul filamento su cui già si è fatto il sequenziamento. Si fa così il sequenziamento sul filamento rimasto.

La prima sequenza letta prenderà il nome di R1, la seconda di R2, mentre la parte non letta tra le parti R1 ed R2 prende il nome di **inner mate distance**. La sequenza di quest'ultima porzione non è letta, ma si conosce all'incirca la sua dimensione, perché si conosce (più o meno) la lunghezza del frammento sequenziato e delle reads R1 ed R2. Quando si confronta la libreria sequenziata con un genoma di riferimento, l'informazione sulla lunghezza della inner mate distance può essere di aiuto per identificare delezioni ed inserzioni rispetto al genoma di riferimento.

La paired-end reads permette anche di mappare meglio delle ripetizioni nel genoma della libreria rispetto ad un genoma di riferimento.

La tecnica paired-end reads permette di identificare agevolmente la presenza di inserzioni, delezioni e duplicazioni.

Sequenziamento per reversione della terminazione

La metodica Illumina prende il nome di **sequenziamento per eversione della terminazione**. Si utilizzano nucleotidi modificati che portano in posizione 3' un gruppo rimovibile. Questi nucleotidi una volta incorporati bloccano l'allungamento del filamento nascente. Il gruppo rimovibile può essere clivato ripristinando il gruppo 3'OH. Questi nucleotidi portano legato alla base azotata un fluoroforo diverso per ciascuno dei quattro nucleotidi, per questo i quattro dNTP vengono aggiunti contemporaneamente alla soluzione.

L'incorporazione di un nucleotide è seguita nell'eccitazione da parte di un laser, che causa l'emissione della fluorescenza dal fluoroforo. La fluorescenza è rilevata da un rilevatore. Successivamente vengono clivati il fluoroforo ed il blocco in 3'.

Sequencing by synthesis

Dopo la PCR la flow-cell presenta dei cluster di molecole di DNA tutte uguali tra loro. Vengono aggiunti i quattro nucleotidi ciascuno legato al suo fluoroforo, il primer e la polimerasi.

Il primer si appaia ai filamenti. La polimerasi incorpora il primo nucleotide, e non può continuare oltre per via del blocco. Si ha l'eccitazione da parte di un laser, con emissione di luce dal fluoroforo, che viene rilevata e registrata. Vengono rimossi il blocco in 3' ed il fluoroforo, e vengono aggiunti nuovamente nucleotidi legati al fluoroforo e polimerasi, e ricomincia il ciclo di sintesi e rilevamento.

Analisi dati

I dati delle reads così ottenute vengono elaborati e vengono restituite delle sequenze, che possono poi ad esempio essere allineate con un genoma di riferimento.

Ogni frammento della library deve essere sequenziato più volte affinché la sequenza ottenuta sia attendibile. Si parla di DNA coverage indicando il numero di sequenze che si vuole ottenere per ciascun frammento della library.

