

NEXT GENERATION SEQUENCING:

Tecniche di sequenziamento di DNA e RNA condotte in parallelo su un grande numero di campioni.

Esistono circa 3 miliardi di basi nel DNA organizzate in un ordine particolare, con circa 20000 geni e altre sequenze non codificanti che però hanno un'importanza fondamentale.

Determinare la sequenza di DNA permette di ottenere informazioni sul tipo di mutazioni che incorrono in un individuo, sulla suscettibilità ad alcuni farmaci, sulla loro efficacia (farmacogenomica)

Il primo tentativo di sequenziamento è stato effettuato nel 1970, ma era incompleto e poco efficiente. Il metodo Sanger ha permesso di ottenere la sequenza di un frammento completo, anche se molto costoso (1975).

Il metodo sanger si basa sulla replicazione del DNA, utilizza una DNA polimerasi che catalizza la reazione di inserimento dei nucleotidi. Nella miscela di reazione vengono aggiunti dideossinucleotidi, ovvero nucleotidi privi di 3'-OH incapaci di allungare la catena una volta incorporati. Questi erano stati marcati in modo da poter essere distinti dai nucleotidi normali, e posti in provette diverse a seconda della base che portavano, in quantità 1:10 rispetto ai nucleotidi normali. Una volta fatta partire la reazione questa si arrestava in punti diversi a seconda del dideossinucleotide incorporato e, i frammenti ottenuti sono stati fatti correre su gel per poter leggere la sequenza in ordine dal basso (frammento più piccolo) all'alto (frammento più lungo) Un successivo avanzamento prevedeva l'utilizzo di fluorofori diversi per i ddNTP usati, in modo da effettuare l'elettroforesi in un gel capillare e colpire i frammenti con un laser che identifica il tipo di base incorporato.

Il progetto genoma umano aveva l'obiettivo di trovare il numero di geni umani e di sequenziare tutto il genoma (1990-2006).

Nel 2000 si sono affermate diverse tecnologie chiamate next generation sequencing, accumulate dal fatto di essere high-throughput ovvero il sequenziamento di molti campioni in parallelo, con una conseguente diminuzione dei costi.

LE NGS sfruttano un sistema di miniaturizzazione della replicazione del DNA all'interno di flowcell, ovvero camere in cui avvengono migliaia di reazioni contemporaneamente,

Le NGS richiedono una frammentazione del DNA da analizzare (fisica o enzimatica), riparazione delle estremità, che vengono rese blunt o con terminali A-T e un legame con adattatori specifici per ogni tecnologia NGS, così da ottenere una library di DNA genomico. La riparazione è importante per far lavorare gli enzimi, che non permetteranno l'allungamento finché i frammenti non sono parificati. La ligazione con un adattatore permette di legare gli oligo complementare nella flowcell, permettere l'arricchimento nella PCR di soli frammenti legati da adattatore, permette di indicizzare i campioni. L'arricchimento con PCR introduce alle molecole legate ad adattatore le sequenze richieste per l'ibridazione della flowcell, inoltre selettivamente arricchisce i frammenti di DNA che hanno adattatori ad entrambe le estremità per amplificare il totale di DNA nella libreria.

La frammentazione può avvenire per sonicazione, ovvero l'utilizzo di onde acustiche (ultrasoniche), condotta con un sonicatore che genera vibrazione meccaniche amplificate sfruttando corrente elettrica ad elevata frequenza, e trasmettendo gli ultrasuoni in una vasca contenente acqua. Successivamente si seleziona la dimensione dei frammenti mediante gel o beads magnetiche e si lega il DNA agli adattatori (es: illumina). Questi adattatori servono per ancorare il frammento alla cella dove avviene l'amplificazione e il sequenziamento e fungono da primer per la reazione di amplificazione,

Si può avere un sequenziamento single-read (una sola direzione) o paired-end (in entrambe le direzioni), con due adattatori perché parte da entrambe le estremità.

→ la caratteristica principale è il multiplexing, ovvero la possibilità di sequenziare contemporaneamente più campioni. Utile quando bisogna confrontare piccoli genomi o specifiche regioni genomiche

In ciascuna lane della flowcell vengono generati milioni di cluster di DNA a doppia elica.

Alcune NGS utilizzano nucleotidi bloccati in 3', marcati con fluo diversi, che una volta incorporati determinano l'arresto della sintesi. La differenza con il metodo sanger è che i nucleotidi vengono aggiunti contemporaneamente e che il blocco inserito è reversibile.

Il segnale luminoso ottenuto viene immediatamente letto dal computer, che registra il nucleotide inserito, e successivamente viene rimosso per permettere il proseguimento della sintesi.

I vantaggi sono throughput, costo per base inferiore, sequenza clonale: poiché si leggono così tante sequenze parallele si può osservare la % di mutazione per una certa regione di DNA.

VI sono 4 strategie:

-PCR emulsione: i frammenti di DNA sono legate a seq adattatori e inserite in micelle droplet. La PCR viene eseguita all'interno di questa micella caricata con primer, templat, Dna polimerasi e nucleotidi. Ciascuna micella viene coperta da migliaia di copie di sequenza di DNA (on-bead amplification)

-solid-phase bridge amplification (illumina): Il DNA frammentato viene legato ad adattatori e primer immobilizzati ad un supporto solido come le flowcell. Il templat si ibrida con adattatori, si ha l'estremità di DNA libera che può interagire con primer vicini e creare un bridge, amplificandosi. Questo porta a milioni di cluster clonali (migliaia di copie per ciascuno frammento da una amplificazione a ponte, con 30-50 milioni di clusters). Gli adattatori servono per legare il frammento o da sequenziare alla flow cell, dove sono presenti sequenze complementari che servono anche da primer.

-template walking in fase solida

-generazione di nanoball di Dna

La sequencing by ligation non utilizza una polimerasi ma una ligasi: vengono usati frammenti di 8 pb di DNA con due basi note e altre basi degenerate legate ad un fluoroforo (diverso a seconda del tipo di nucleotide), che vengono ibridate con un templat, la ligasi immobilizza il complesso e viene registrato il segnale del fluoroforo durante il legame. Successivamente le ultime basi della sonda vengono tagliate, così da spegnere il segnale, mentre 5 basi rimangono legate allo stampo (2 note e 3 no): si ottengono sequenze con basi note a due a due. Si ripete il ciclo diverse volte e dopo 10 round di ibridazione, ligazione, immagine e clivaggio, si identifica la sequenza

SOLiD è un acronimo che sta per **Sequencing Oligo Ligation Detection** che utilizza la tecnologia di **ligation and two-base coding**. Grazie al metodo del two-base calling il sistema SOLiD presenta un'accuratezza molto elevata (99,99%), che è per ora superiore a quella di tutti gli altri sistemi NGS disponibili. La lunghezza delle **reads** resta tuttavia un limite maggiore (si riescono ad ottenere reads della lunghezza massima di **85bp**, mentre con Roche 454 si può arrivare fino a 700 bp). L'**output è di 30 Gb a corsa**. Una corsa necessita **7 giorni** e genera un ammontare di dati pari a 4 TB

Sequencing by synthesis: si prepara una libreria di DNA, si amplifica su fase solida (illumina) mediante legame degli adattatori a primer e formazione di ponti durante l'amplificazione. Si generano cluster di 100-200 milioni di sequenze. Per il sequenziamento si usano primer terminatori reversibili, coniugati ad una fluoroforo. Quando incorporato blocca la sintesi e viene detectato, emettendo un colore corrispondente alla base incorporata, successivamente viene clivato e lavato dalla flowcell, mentre il gruppo 3'OH viene rigenerato e il ciclo può continuare.

→ 2 flowcell da 8 lane permettono di ottenere 375×10^6 cluster/lane con un totale di 6 Tb di DNA sequenziato (il genoma umano è 0,3 Tb)

Sequencing by single nucleotide addition: pirosequenziamento (Roche)

Si basa sulla polimerizzazione di un filamento a cui si aggiungono basi di un solo tipo per volta, queste vengono incorporate rilasciando pirofosfato che scatena una cascata enzimatica, che porta alla produzione di luce. Durante ogni ciclo è presente solo una specie di dNTP, i dNTP multipli identici possono essere incorporati durante un ciclo, aumentando la luce emessa. È stata dismessa dal 2015 perché meno efficiente degli altri metodi.

Ion Torrent (ThermoFisher): si utilizza un semiconduttore che quando viene incorporata una base rilascia uno ione H^+ che viene rilevato da un sensore CMOS-ISFET. Solo una specie nucleotidica è presente durante ogni ciclo, quando vengono incorporati più nucleotidi uguali di seguito si ha un'emissione di ioni maggiore

Gli NGS a corta lettura possono comportare errori, è importante aumentare l'accuratezza del metodo di sequenziamento per diagnosticare rare mutazioni subclonali

Per sequenziare frammenti più lunghi vi sono metodi di sequenziamento long read: permettono di risolvere sequenze più lunghe in un'unica lettura per eliminare errori dovuti a Dna altamente ripetuto e ambiguità di sequenze simili.

Questi metodi permettono di leggere interi RNA e fornire informazioni su esoni e particolari isoforme. Non usano cluster di Dna amplificato

Polimerasi immobilizzata su un supporto che permette la polimerizzazione di un filamento inserendo nucleotidi con fluorofori che emettono luce man mano che vengono incorporati.

Oxford nanopore: la sequenza lineare interagisce con il poro e una proteina motrice, una forcina permette un sequenziamento bidirezionale. Viene fatta passare attraverso un poro e si registra una corrente che viene modulata man mano che il Dna vi scorre.

Work flow:

- 1) si prepara la libreria per creare una collezione random di frammenti di RNA pronti per il sequenziamento.
- 2) si effettua un binding delle molecole di DNA con adattatore alla superficie della flowcell
- 3) amplificazione bridge: si amplifica la singola molecola per generare più segnale nei seguenti step di sequenziamento
- 4) per avere la certezza che tutte le sequenze in ogni colonia abbiano un orientamento, si rendono le sequenze single strand
- 5) Data analisi

Applicazioni:

-sequenziamento DNA genomico: permette di sequenziare l'intero genoma generando un'analisi completa di tutte le varianti rappresentate nel campione. È un processo ancora molto costoso, l'analisi è complessa e il coverage piuttosto limitato, il che significa che spesso si verificano dei buchi nel sequenziamento (regioni a basso o nullo coverage)

-sequenziamento whole-exome

-targeted sequencing: il DNA da sequenziare viene ibridato con sonde di DNA biotinilato, successivamente vengono inserite sfere magnetiche funzionalizzate con streptavidina, che

legano biotina. Queste vengono purificate, si separa il campione dalla sferetta, si amplifica e si sequenzia.

-deep sequencing: permette di sequenziare a coverage molto alto, favorendo l'identificazione di varianti molto rare, è molto economico. Tuttavia è applicabile solo a piccole porzioni di Dna (ampliconi di 100-5000 pb) ed è poco scalabile

-ChIP-Seq (chromatin ImmunoPrecipitation). Si effettua un cross-link DNA-proteine della cromatina, si lisa la cellula, si separa il nucleo per centrifugazione e si sonica si effettua una digestione enzimatica per ottenere i frammenti di cromatina. Successivamente si effettua una immunoprecipitazione con specifici anticorpi e si purifica il DNA per analizzare le porzioni legate mediante PCR, qPCR, microarray o sequenziamento

-Methylseq: si effettua per studiare le perturbazioni del pattern di metilazione. L'eterocromatina è di norma ipermetilata, se viene ipometilata determina instabilità genomica, con probabili patologie come x fragile, sindrome di Rett, autismo, schizofrenia, demenza, sindromi varie... Allo stesso modo, l'eucromatina deve essere ipometilata, se viene ipermetilata, soprattutto a livello dei promotori CpG, si ha repressione trascrizionale, con sviluppo di patologie.

Questa si effettua mediante conversione con bisolfito, amplificazione con PCR, tagmentazione, sequenziamento di library, ottenimento di sequenza del DNA con calcolo % di metilazione

→ per identificare mutazioni somatiche responsabili di insorgenza di un tumore raro, con pochi campioni disponibili e nessun dato in letteratura si utilizza la whole-genome o la whole-exome

→ per identificare mutazioni responsabili di un tumore frequente, con numerosi campioni disponibili e dati in letteratura che mostrano la presenza di mutazioni in 5 geni nel 90% dei casi, si effettua la sanger o la targeted sequencing.

→ per identificare una malattia minima residua di un paziente leucemico in remissione dopo chemioterapia, con un unico campione disponibile e mutazioni responsabili note, si effettua una deep.

-trascrittomica: RNA sequencing

si estrae l'RNA totale, si seleziona la specie desiderata (poliA, deplezione rRNA, selezione per dimensione), si frammenta e si retrotrascrive in cDNA, si legano adattatori specifici e si amplifica con PCR poi si esegue il sequenziamento. L'analisi di espressione RPM (reads per milioni mappate) è pari al rapporto tra reads mappate sul gene e le milioni di reads mappate sul genoma.

La trascrittomica a singola cellula può essere usata per studiare l'eterogeneità tumorale, mediante integrazione con analisi genetica.

miRNomica

-epigenomica

-genomica

SINGLE CELL BIOLOGY

Perché studiare le singole cellule:

Studiare popolazioni non permette di apprezzare le qualità delle singole componenti del sistema, consentendo di avere maggiori informazioni.

Si possono usare campioni con poche cellule, permettono di analizzare l'eterogeneità del genoma e del trascrittoma, che possono essere raccolti e analizzati grazie alla statistica, inoltre una promessa futura è quella di effettuare discipline omiche multimodali su singola cellula.

→ l'aumento di potenza di queste analisi è dovuto allo sviluppo di nuove tecnologie e strumenti: in particolare multiplexing, circuiti microfluidici, robotica integrata da microfluidica, nanodroplet, barcoding in situ, in cui si riesce ad analizzare centinaia di migliaia di cellule.

Data una popolazione costituita da 50% cellule wt e 50% cellule con il doppio dell'espressione, si ottiene una significativa eterogeneità in studi di popolazione, ma se le cellule che esprimono più geni sono solo poche copie all'interno di una popolazione, non è possibile identificarle se non si adottano studi a singola cellula.

Inoltre studiare le cellule singolarmente permette di superare il paradosso di Simpson, secondo il quale il trend della media può essere diverso dal trend di ciascun costituente del gruppo.

Infatti si può avere una reazione ad un farmaco nella maggior parte della popolazione, ma questo può risultare meno efficace su individui maschi piuttosto che femmine, rispetto ad un altro farmaco.

→ queste tecniche a singola cellula permettono di investigare le differenze individuali tra cellule, come variazioni di numero di copie, alterazioni strutturali, riarrangiamenti, mutazioni a singolo nucleotide, variabilità negli elementi trasponibili e integrazione virale.

→ inoltre permette di studiare le differenze nell'espressione genica e nella regolazione degli stessi: l'abbondanza dei trascritti, uso di siti alternativi di splicing o di regolazione

Potenziati implicazioni per la scoperta scientifica:

- 1) identificazione di nuovi/rari tipi cellulari, con successiva riclassificazione dei citotipi
- 2) variazioni da cellula a cellula in varie fasi dello sviluppo dell'organismo (variano i livelli di espressione nel giovane e nell'anziano)
- 3) dissezionare una popolazione complessa usando markers onco-genetici (leucemia mieloide cronica, associata ad una lesione singola, la traslocazione BCR-ABL: pur essendoci una terapia vi sono cellule staminali che resistono). Dopo un sequenziamento a singola cellula di queste cellule staminali mediante multiplexed, si è ottenuto un trascrittoma single-cell per predire la risposta alla terapia (cellule good responders e poor responders), monitorare la risposta, identificare nuovi target terapeutici.

In generale, gli esperimenti di biologia molecolare a singola cellula seguono il seguente work flow:

-ottenimento del campione primario

-individuazione e isolamento di cellule di interesse: si possono isolare mediante diluizione (sospensione cellulare: economica ma laboriosa, pozzetti negativi, condizioni mild sulle cellule), micromanipolazione (coltura cellulare: è richiesto un trasferimento con pipette, condizioni mild sulle cellule, laborioso, si ottengono cellule specifiche di interesse), FACS (efficiente nell'isolare numerose cellule, richiede un campione molto ampio, strumentazione costosa, il sorting può alterare il profilo di espressione di RNA), microfluidica (efficiente per isolare un grande numero di cellule, richiede un ampio campione iniziale, costoso), microdissezione laser-capture (diretto isolamento di singole o multiple cellule dal tessuto, perdita di informazione genomica per rottura dei nuclei, il montaggio potrebbe disturbare l'espressione genica), dissociazione tissutale (può portare alla distruzione meccanica, chimica o enzimatica, produce una sospensione da un tessuto solido, potenzialmente può alterare drasticamente l'espressione genica)

-lisi delle cellule

-amplificazione del whole-genome o trascrittoma (per studi del Dna/RNA) → bisogna amplificare perché da una singola cellula si ottiene poco materiale e questo può portare ad errori

-tecniche analitiche a scelta (NGS, gene panels, RT-PCR, microarray, Sanger)

-analisi dei dati e interpretazione

NB, fare attenzione ai **Batch effect**: si verifica quando fattori non biologici in un esperimento causano cambiamenti nei dati prodotti dall'esperimento. Tali effetti possono portare a conclusioni imprecise quando le loro cause sono correlate con uno o più risultati di interesse in un esperimento.

Inoltre bisogna distinguere un **bias di amplificazione** e quindi valutare quali dati siano biologicamente validi.

Si effettuano controlli di qualità e si normalizzano i dati per poter condurre analisi a livello cellulare (trascrittoma, espressione) e a livello di singolo gene (pathway, network regolatorio, interazione genica, profilo di metilazione di DNA)

DNA METHILATION

Avviene in posizione 5 della citosina soprattutto quando si trova in CpG islands da parte di DNA metiltransferasi (DNMT1 mantengono il pattern di metilazione del filamento parentale e DNMT3a/b metilano de novo) che usano SAM come donatore, possono avvenire soprattutto a livello del promotore (se non codificanti), nel 70% dei casi, mentre nell'1% dei promotori non si trovano (geni sempre espressi).

Queste modificazioni del Dna possono esercitare ruoli impo in diversi processi biologici: organizzazione e condensazione cromatina, silenziamento genico, imprinting genomico, inattivazione x, differenziamento cellulare, sviluppo embrionale.

Nelle cellule tumorale si può avere un pattern di metilazione aberrante, con promotori di tumor suppressor ipermetilate e geni che non dovrebbero essere espressi ipometilati.

Meccanicamente, la metilazione del DNA può contribuire al silenziamento dei geni sia impedendo che i fattori di trascrizione si associno ai promotori, sia reclutando proteine che modificano gli istoni e rimodellano la cromatina in uno stato trascrizionalmente non permissivo.

Ci sono diversi metodi per rilevare lo stato di metilazione del DNA.

Una tecnica, il saggio HELP, prevede due endonucleasi di restrizione chiamate HpaII e MspI, che rispettivamente scissione solo non metilato, o sia metilato e non metilato, sequenze CCGG. Confrontando i modelli di digestione prodotti da questi due enzimi, si può dedurre lo stato di metilazione del DNA.

Un altro metodo, chiamato immunoprecipitazione del DNA metilato o "MeDIP", utilizza anticorpi che si legano alle citosine metilate per arricchire le sequenze di DNA metilato.

Infine, l'analisi dei bisolfiti viene utilizzata per distinguere la citosina metilata da quella non metilata nel DNA, effettuando una reazione chimica che converte la citosina non metilata in uracile. A seguito di questa conversione, il DNA trattato con bisolfito può essere sottoposto a PCR, sequenziato e confrontato con un genoma di riferimento. Le citosine non metilate sono quelle presenti nel riferimento, ma sostituite con timine dopo l'analisi del bisolfito e la PCR.

Sottoponendo il DNA trattato con bisolfito alla spettrometria di massa, i ricercatori possono anche creare "epigrammi di metilazione", che rappresentano linearmente diversi CpG nel genoma e rappresentano il grado di metilazione in ciascuno di essi. Tali epigrammi sono particolarmente utili se i ricercatori desiderano confrontare i modelli di metilazione tra i diversi tipi di cellule.

Ci sono metodi per arricchire le sequenze metilate e non metilate basati su enzimi di restrizione che tagliano in modo specifico sequenze metilate, immunoprecipitazione con anticorpi specifici. Altri tipi di NGS permettono di distinguere altre modificazioni delle basi di DNA e sono più potenti. LA scelta della tecnica dipende dal tipo di applicazione che si deve fare e dalle dimensioni del genoma.

CHROMATINE ANALYSIS:

L'organizzazione della cromatina è fondamentale per numerosi processi biologici.

Un metodo per studiare le regioni di cromatina decondensate è il **saggio con DNAsI I** che identifica i siti sensibili all'idrolisi (più accessibili): si tratta il filamento con l'enzima e si ottengono una serie di frammenti che corrispondono alle aree non complessate ad istoni, questi vengono sequenziati, si costruiscono librerie e vengono mappati sul genoma per identificare le regioni più sensibili (sono correlate ad attività trascrizionale, promotori, enhancers).

FAIRE: isolamento assistito da formaldeide di elementi regolatori, In questo metodo, il DNA è legato covalentemente alle proteine della cromatina utilizzando la formaldeide come agente di cross-link e tagliato in piccoli pezzi mediante sonicazione. Il DNA libero viene poi purificato mediante l'estrazione in fenolo: cloroformio (in fase acquosa si trova il Dna non complessato con proteine, mentre in fase solita le proteine).

Il rapporto del DNA libero è determinato mediante qPCR o sequenziamento del DNA (DNA-seq) rispetto ad un campione di controllo che rappresenta il DNA totale. Le regioni con una struttura di cromatina più sciolta vengono arricchite nel campione di DNA libero, permettendo così l'identificazione delle regioni genomiche con minore compattazione della cromatina, probabilmente implicate nella regolazione genica

CHIP: chromatin immunoprecipitation, si possono targettare le modifiche specifiche degli istoni mediante anticorpi, che possono essere "fatti precipitare" insieme al DNA circostante. Con tecniche di PCR, microarray o il sequenziamento si possono poi identificare le regioni di DNA associate alle modificazioni istoniche di interesse. Variando gli anticorpi utilizzati durante ChIP, si possono anche individuare le regioni di DNA legate da fattori di trascrizione e altre proteine di regolazione. Per cominciare, le cellule di interesse vengono raccolte e trattate con formaldeide, che agiscono come reagenti di "cross-linking" e contribuiscono ad apporre proteine alle sequenze di DNA a cui si associano, facilitando la formazione di legami covalenti tra di loro. Bisogna fare attenzione a non "trattare eccessivamente" le cellule con la formaldeide, in quanto ciò può influire sulla capacità degli anticorpi di riconoscere le loro modifiche dell'istone bersaglio nelle fasi successive di ChIP. Per fermare il processo di cross-linking, si aggiunge glicina alla soluzione di formaldeide con cui le cellule vengono trattate. Le cellule vengono poi raccolte e lisate per rilasciare la cromatina. Per solubilizzare la cromatina e definire con precisione le regioni del DNA che si associano agli istoni modificati, la cromatina viene sonicata meccanicamente in pezzi più piccoli utilizzando onde sonore. In genere, gli scienziati mirano a creare frammenti di cromatina da 200 a 1000 coppie di base in lunghezza. Una volta che i frammenti di cromatina della dimensione desiderata sono generati, un anticorpo viene aggiunto alla soluzione, e la miscela viene incubata per dare all'anticorpo il tempo di riconoscere la modificazione dell'istone bersaglio.

Si introducono poi sferette magnetiche a cui gli anticorpi possono legarsi e si immobilizzano i complessi di cromatina associati agli anticorpi. Le sferette vengono raccolte attraverso l'uso di rack magnetici, e lavate più volte per sciacquare via la cromatina non legata o gli anticorpi.

Per liberare la cromatina da essi, le sferette vengono messe in una soluzione contenente il detergente SDS, e dopo aver raccolto le sferette con un magnete, si trattiene il surnatante, si aggiunge l'enzima proteinasi K per degradare tutte le proteine, compresi gli istoni, in modo che il DNA possa essere isolato e analizzato con PCR e sequenziamento (cipseq)

Chromosome conformation capture: sono diverse tecniche che studiano le interazioni intercromosomiche, si tratta il Dna con formaldeide per fissare struttura 3D cromatina, si digerisce con enzimi di restrizione e si effettua ligazione e amplificazione. Si ottiene una sequenza di Dna che riflette la struttura 3D originale nella cellula, questo permette di catturare la conformazione cromosomica da più punti.

Si differenziano in:

-3c (one vs one): si utilizza per confermare una particolare interazione tra promotore ed enhancers, si effettua PCR con primer specifici e richiede la conoscenza di entrambe le sequenze
-4C (one vs all), permette di visualizzare tutte le (interazioni non note partendo da un locus noto, si effettua una PCR inversa usando la sequenza nota per amplificare le sequenze non note che vi interagiscono, identifica fino ad un milione di interazioni, è molto precisa e riproducibile (si trovano nelle vicinanze nel nucleo)

-5C (many vs many), si usa quando nessuna sequenza è nota per identificare regioni che interagiscono, il coverage è relativamente basso, non si può applicare all'intero genoma perché il numero di primer è troppo basso

-Hi-C (all vs all) permette di studiare interi genomi e le interazioni tra tutti i frammenti a risoluzione più bassa della 4C, la metodica è sempre la stessa, si effettua una precipitazione con abiotinilati e un sequenziamento a cavallo delle giunzioni, si analizza attraverso una matrice heat map che permette di valutare l'interazione tra due frammenti di DNA.

-capture-C è un buon compromesso tra le tecniche, utilizza enzimi di restrizione che generano frammenti di 400 paia di basi (alta risoluzione), adattatori per sequenziamento che permettono di identificare i frammenti, si effettua arricchimento con oligo biotinilati specifici per le sequenze note che si vogliono studiare nelle interazioni, si effettuano amplificazione e sequenziamento.

Bisogna considerare il tipo di enzimi di restrizione considerati e la loro risoluzione, la sensibilità di varie tecniche (numero di letture per ciascun viewpoint: la capture C+ NGS è la più sensibile)