***Sequenziamento del genoma***

Il sequenziamento del DNA è il processo di determinare l'ordine dei nucleotidi (A, C, G e T) che compongono una molecola di acido desossiribonucleico. È un passo fondamentale per comprendere la funzione di un gene e per identificare regioni regolatrici e codificanti, oltre a individuare mutazioni e stabilire correlazioni filogenetiche tra specie diverse.

Esistono tre generazioni di sequenziamento che si distinguono tra loro:

1. Sequenziamento di prima generazione o "Sanger Sequencing": Questo tipo di sequenziamento, a basso throughput, permette di sequenziare due filamenti distinti (forward e reverse) in una reazione di sequenziamento. Le sequenze ottenute, chiamate reads, sono brevi frammenti di DNA sintetizzato.
2. Sequenziamento di seconda generazione o "Next Generation Sequencing" (NGS): Questo tipo di sequenziamento ad alto throughput consente di ottenere milioni di reads contemporaneamente, permettendo il sequenziamento di più reads da ogni campione. L'NGS è un approccio di alto volume e può generare una grande quantità di dati.
3. Sequenziamento di terza generazione o "Single Molecule Sequencing": Questa tecnica, sviluppata negli ultimi otto anni, consente il sequenziamento del DNA di una singola molecola. Sebbene sia costosa e precisa, non è facilmente accessibile a tutti.

Nel contesto della medicina personalizzata e di precisione, le informazioni ottenute dal sequenziamento di seconda generazione (NGS) sono le più utilizzate. Tuttavia, l'interpretazione dei dati può essere complessa poiché si tratta di sequenziare una "macedonia di frutti", ossia ottenere informazioni da molteplici frammenti di DNA e successivamente interpretarle. Questo problema non si presenta nel sequenziamento di terza generazione, poiché si estrae il DNA da un singolo "frutto", ad esempio una sola fragola.

Per la diagnosi e la terapia di malattie complesse, vengono utilizzati modelli di intelligenza artificiale basati sui dati provenienti dal NGS. In generale, le nuove tecnologie di sequenziamento (seconda e terza generazione) si basano sullo sviluppo di nanotecnologie e permettono l'analisi simultanea e parallela di molte molecole di DNA.

***Metodo Sanger***

Il metodo Sanger viene definito come primo sequenziatore automatico (1987), che permette di ottenere un’analisi di un singolo frammento amplificato mediante PCR e il risultato è un elettroferogramma. Più sono grandi i geni più è necessario fare frammenti di amplificato che poi verranno sequenziati. La sequenza viene allineata al riferimento e analizzata per identificare varianti nucleotidiche.

*Differenze tra Metodo Sanger e NGS*

Nel metodo Sanger, si sequenzia un singolo esone e si ottiene un'unica sequenza relativa a quel target specifico. Quindi, da un singolo campione, si ottiene una singola sequenza. Nel caso della NGS, è possibile sequenziare contemporaneamente e parallelamente più di un singolo target, ovvero molteplici esoni per molteplici campioni. Questo è reso possibile tramite un meccanismo chiamato PCR clonale, in cui vengono amplificati e sequenziati frammenti identici.

Nella NGS, si parte da un campione di riferimento e si generano le reads. Durante l'analisi NGS, vengono contati i cambiamenti nelle basi, ad esempio contando le citosine rispetto alle guanine. Tuttavia, non sempre si ottiene un allineamento preciso tra le reads, ovvero le sequenze non vengono posizionate perfettamente una sopra l'altra. Ad esempio, nel metodo Sanger, potrebbe essere possibile Immagine che contiene testo, schermata, design

Descrizione generata automaticamenteleggere la T ma non la C, a causa di reads incomplete prodotte dalla NGS, che possono influenzare l'affidabilità dei dati.

Oggi ci troviamo di fronte a una situazione molto complessa in cui abbiamo a disposizione molte più informazioni che devono essere interpretate correttamente. Grazie alle nuove tecnologie, possiamo avere una visione più completa della complessità delle diverse patologie, che possono essere associate a varianti a basso impatto, come ad esempio i polimorfismi. Spesso, per identificare una variante, non è sufficiente una singola tecnica, ma è necessario ricorrere a diverse analisi genetiche. Attraverso queste analisi genetiche, sono stati identificati i geni responsabili delle malattie mendeliane.

La NGS ci ha permesso di individuare malattie molto rare, analizzando le varianti genetiche che predispongono a una determinata patologia e che possono essere caratterizzate da geni di suscettibilità. Queste varianti, a bassa frequenza e basso impatto, sono state identificate grazie alla NGS. Al contrario, le varianti più comuni, che sono di basso impatto e di natura benigna, sono state individuate attraverso un tipo di analisi che permette di identificare varianti molto frequenti.

L'approccio attuale alla diagnosi e alla terapia è diverso. In passato, si cercava la causa di una malattia, mentre oggi si parte da un'indagine genetica per individuare la causa biologica e poi correlarla con l'aspetto fenotipico. Si parla di geno-terapia e geno-diagnostica. Dal 2009 ad oggi, le tecnologie si sono evolute notevolmente: i costi sono diminuiti e le capacità di processamento sono aumentate (è possibile analizzare molti campioni su un singolo microchip). Negli ultimi anni, gli studiosi hanno cercato di ottenere dati di qualità sempre maggiore e di sviluppare sistemi di validazione affidabili.

*Affidabilità dell’NGS*

La NGS può presentare notevoli errori, ma esistono score che valutano la qualità e l'affidabilità dei dati (il metodo Sanger ha un'affidabilità del 90%).

Nel contesto della NGS:

1. Ricerca: La NGS contribuisce all'identificazione di nuovi geni responsabili di tratti fenotipici.
2. Diagnosi: La NGS aumenta le potenzialità dei test genetici nell'individuazione della causa molecolare di una patologia.

Esistono diverse piattaforme NGS, che si distinguono per concentrazione e chimica, ma condividono la capacità di generare informazioni di sequenza ad alta produttività, ossia leggere decine di migliaia di sequenze in parallelo.

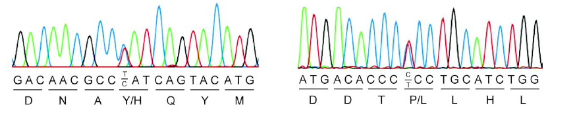
Le piattaforme NGS possono essere classificate in:

* Piattaforme High throughput: Sono progettate per sequenziare ampie regioni genomiche in una singola analisi.
* Piattaforme middle throughput: Sono progettate per sequenziare regioni genomiche di dimensioni minori in una singola analisi.

*Il disegno sperimentale dell’NGS:*

Nel contesto del sequenziamento NGS, il processo di analisi inizia con un'indagine anamnestica familiare per risalire alle informazioni sulla famiglia.

1. Preparazione del campione: Il DNA genomico viene frammentato per essere pronto per il sequenziamento.
2. Sequenziamento: Il campione frammentato viene sequenziato utilizzando la tecnologia NGS.
3. Analisi dettagliata: Vengono condotte analisi approfondite sui dati ottenuti dal sequenziamento.
4. Validazione dei risultati: I risultati vengono successivamente convalidati utilizzando il metodo di sequenziamento Sanger.

Per comprendere meglio il processo, possiamo immaginare una biblioteca rappresentante il nostro genoma, dove ogni libro rappresenta una parte del genoma. Dei linker sono posizionati all'inizio e alla fine di ogni frammento del DNA, consentendo di distinguere il DNA specifico di ogni individuo e di applicare il sequenziamento.

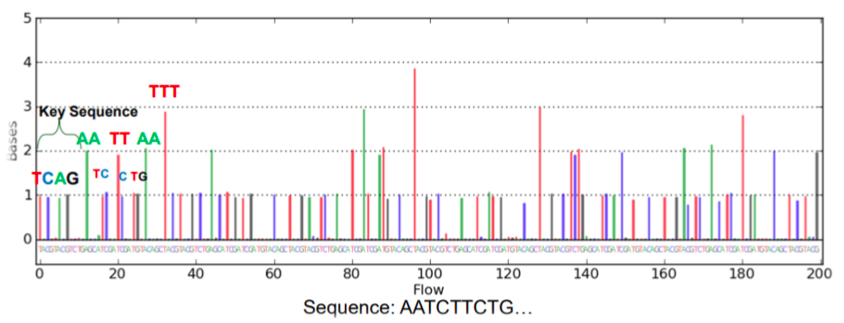
Gli ultimi passaggi del processo sono di natura bioinformatica e coinvolgono tre fasi:

1. Analisi informatica primaria: Questa fase include il "base calling", che consiste nella chiamata delle basi sequenziate. In seguito, avviene l'allineamento dei frammenti a un genoma di riferimento.
2. Analisi informatica secondaria: Questa fase, chiamata "variant calling", verifica la presenza di varianti in posizioni specifiche all'interno del genoma di riferimento. La sensibilità e la specificità del sequenziamento dipendono dai parametri di analisi scelti in base alla natura del sequenziamento stesso.
3. Analisi informatica terziaria: Questa fase viene svolta da un genetista utilizzando software dedicati. Consente l'interpretazione delle varianti, considerando parametri come la localizzazione, la veridicità e le possibili cause e conseguenze. Inoltre, valuta se la variante è presente o meno nella popolazione. Se viene individuata una variante mai vista prima, sarà necessaria un'ulteriore valutazione in vitro e in vivo.

Quindi, il flusso di lavoro globale consiste nella preparazione della libreria, frammentazione del DNA, preparazione dei filamenti singoli e successivo sequenziamento. Nel caso del metodo Sanger, il sequenziamento avviene su un singolo filamento di DNA. Dopo il sequenziamento, vengono eseguite analisi bioinformatiche per allineare i frammenti e identificare le basi sequenziate.

Passando alle nanotecnologie, in questo caso si lavora con un microchip che contiene fili di ingresso e di uscita. Una determinata quantità di libreria genomica viene posizionata sul microchip, con sequenze legate a braccetti che corrispondono ai linker presenti all'inizio e alla fine dei frammenti. Il DNA viene dosato correttamente per occupare i pozzetti del microchip. Questi pozzetti funzionano come sensori di voltaggio, in quanto il sequenziatore rileva le variazioni di voltaggio causate dalla produzione di ioni protoni durante l'incorporazione dei nucleotidi. Questi segnali vengono tradotti in un grafico a ionogramma, che consente la sequenziazione dei frammenti e l'individuazione dei nucleotidi.

Dopo qualche ora dal processo di sequenziamento, il server genera le "reads", che sono piccole sequenze che vengono successivamente allineate al genoma di riferimento (analisi primaria). Successivamente, viene eseguito il "variant calling" (analisi secondaria), che permette di identificare le varianti nelle posizioni specifiche. Il report generato dalla macchina può essere difficile da interpretare; quindi, vengono utilizzati software per annotare queste varianti tramite sistemi di annotazione. Questi sistemi forniscono informazioni sulle mutazioni funzionali e strutturali delle varianti, inclusa la localizzazione e il significato delle stesse.

In conclusione, l'analisi terziaria viene eseguita da un genetista utilizzando software specializzati per interpretare e annotare le mutazioni, valutando la loro localizzazione, veridicità, cause, conseguenze e la loro presenza nella popolazione. L'annotazione è fondamentale per comprendere i risultati del sequenziamento NGS, che altrimenti fornirebbe solo una serie di numeri e lettere senza significato chiaro.

Quali sono le maggiori computazioni performate con i dati NGS?

* *Base calling,* cioè la chiamata delle basi, attraverso la formazione delle reads;
* *Allineamento* ad un genoma di riferimento;
* *Variant calling*, cioè la presenza di picchi in dui si deve verificare se in una determinata posizione ci sia o meno una variante.

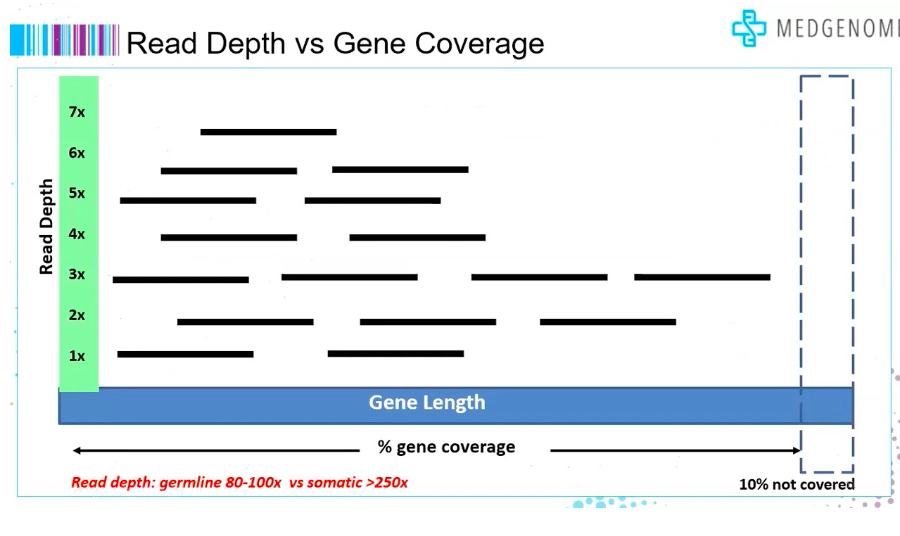
Questi sono gli aspetti più importanti di un’analisi di NGS.

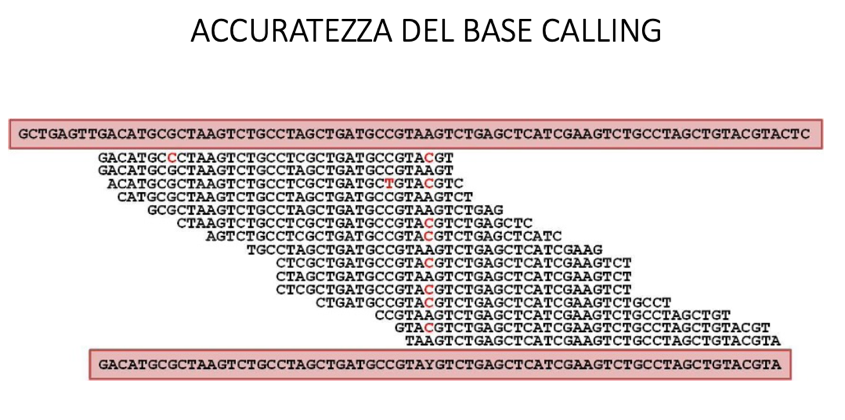
Vediamo ora alcune definizioni:

* La "reads" è una piccola sequenza di circa 30 nucleotidi ottenuta da un frammento di DNA durante il processo di sequenziamento NGS. La lunghezza di 30 nucleotidi è scelta in modo da coprire l'intero genoma e consentire il sequenziamento anche di sequenze più piccole come gli SINE.
* La lunghezza della "reads" è determinata dai nucleotidi contigui prodotti durante il sequenziamento.
* Il "coverage" o "read depth" rappresenta il numero medio di volte che una base viene analizzata. Indica il numero di "reads" che mappano su una specifica base del genoma.
* La "call" è il processo di determinazione di una specifica sequenza di basi durante il sequenziamento.
* La "call quality" rappresenta l'accuratezza con cui viene effettuata una specifica chiamata di base. Indica la qualità della determinazione della sequenza di base.

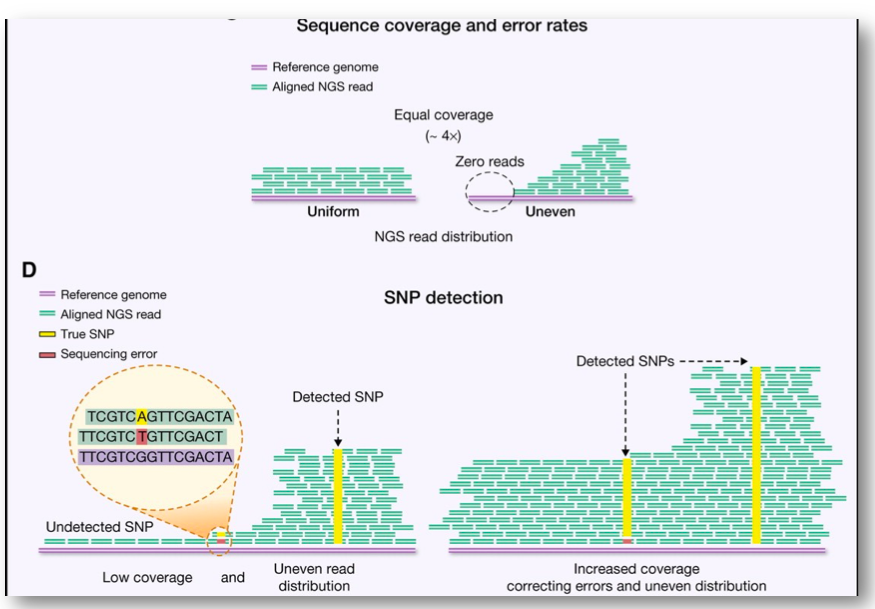
Il sequenziamento NGS può presentare diversi errori, ma fortunatamente ci sono sistemi che valutano la qualità delle basi, il "coverage" e la profondità adeguata all’analisi. Ad esempio, il "Q score" è un punteggio numerico (Q) che indica la probabilità che una specifica base sia chiamata erroneamente dal sequenziatore.

**Differenza tra Read Depth e Gene Coverage**

Nell'immagine accanto sono rappresentati sei geni, dove le "reads" (trattini neri) si formano durante il processo di sequenziamento. I numeri presenti nella banda verde indicano il numero di volte che una base viene letta in quel momento. La banda blu rappresenta la lunghezza del gene. Il "gene coverage" indica quanto del genoma di riferimento è coperto durante il sequenziamento. Nell'immagine, ad esempio, il 10% del gene non viene coperto.

Successivamente, viene mostrato un grafico che indica quante volte una specifica base viene chiamata in una determinata posizione rispetto al genoma di riferimento. L'obiettivo è determinare se una base è stata letta correttamente o se è stato commesso un errore tecnico nella lettura. Questa valutazione di accuratezza viene spesso misurata utilizzando il Phred Quality Score (Q score), un punteggio numerico che indica la probabilità che una specifica base sia stata chiamata erroneamente dal sequenziatore.

La "read depth" è un altro concetto importante nel sequenziamento del DNA. Per le mutazioni germinali, che sono malattie ereditarie, è necessario avere un "read depth" di circa 800-100x, ovvero bisogna avere almeno 80-100 letture per ottenere risultati affidabili. Nel caso delle mutazioni somatiche, che si verificano nei tumori, è necessario avere un "read depth" ancora maggiore, superiore a 250 letture (anche 500), a causa dell'aumento della variabilità. Le analisi somatiche richiedono quindi una maggiore copertura e approfondimento rispetto alle analisi germinali.

Nel caso dell'immagine, viene considerato un "coverage" di 4x, che indica il numero di letture sovrapposte su una specifica regione del genoma. Tuttavia, il 4x può essere ottenuto con diverse distribuzioni delle letture. È importante analizzare la corretta distribuzione delle letture e verificare se ci sono parti del genoma in cui la copertura è 0. Un "coverage" ampio, ma non uniforme, può portare alla presenza di falsi negativi a causa della presenza di varianti genetiche. Aumentando la profondità del sequenziamento, è possibile correggere gli errori e ottenere dati più affidabili.

Sebbene la tecnologia di nuova generazione abbia tassi di errore apparentemente più elevati rispetto alla metodica tradizionale di Sanger, l'accuratezza nel sequenziamento NGS è garantita dalla lettura massiva e ripetuta di ogni frammento genico. Questo assicura una buona copertura del genoma, ma è importante raggiungere un "coverage" adeguato (superiore a 20/50x) per evitare errori e falsi negativi.

In conclusione, il "coverage" rappresenta una delle misure di profondità e completezza del sequenziamento del DNA. È importante valutare sia l'uniformità del "coverage", che indica quanto del genoma è stato coperto, sia la "Coverage Depth", ovvero il numero di letture per nucleotide durante.

***ANNOTAZIONE***

L’annotazione dei geni codificanti proteine viene suddivisa in:

• ***Annotazione funzionale***, consiste nel caratterizzare ogni singolo gene, assegnando una funzione biologica a ogni proteina codificata dal gene stesso.

• ***Annotazione genica*** o semplicemente annotazione, consiste nel definire

all’interno del genoma:

• ***La localizzazione di ciascun gene***.

• ***La struttura di ciascun gene*** (esoni,CDS, UTR).

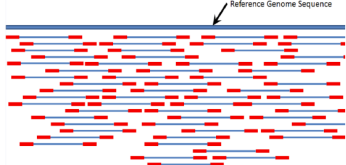
• ***Gli eventuali trascritti alternativi***.

Queste sequenze da sole non forniscono informazioni utili.

Informazioni aggiuntive che possono essere inserite in queste banche dati riguardano la posizione all’interno della sequenza del cromosoma di geni (con i loro introni, esoni, regioni 3’ e 5’ non tradotte e promotori), regioni variabili, mutazioni, SNP etc…

Tutte queste informazioni vengono definite annotazioni.

Quello che fuoriesce da un’analisi NGS non è utile se non viene associato ad un’analisi terziaria, che permette di capire quale sia la funzione di una determinata variante.

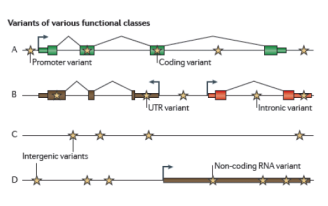
***OUTPUT DELL’ESPERIMENTO*** 

-Allineamento delle letture ottenute con sequenza di riferimento.

-Identificazione di TUTTE le varianti

Dall'esperimento si ottiene una quantità di dati molto elevata che deve essere interpretata.

La mole di dati prodotta dalle piattaforme NGS si colloca nell’ordine delle Terabyte

(TB…1.024.000 megabyte…pari a 1024 gigabyte (GB), cioè (1024) ⋅ (1024)3 byte =

1.099.511.627.776 byte…), rappresentando una difficoltà per l’analisi e lo

stoccaggio dei dati.

Nonostante la macchina impieghi poche ore a fornire un risultato, è necessario tenere in conto il tempo utilizzato per interpretare tale risultato, che può variare da ore a settimane. Il tempo subisce un ulteriore aumento quando il risultato non è riportato all’interno del database, soprattutto quando viene effettuata un’analisi genomica.

*Problemi etici legati all’utilizzo delle nuove tecnologie:*

Il DNA del paziente viene conservato, ma questa tipologia di dato deve essere protetta ed utilizzata esclusivamente previa autorizzazione del paziente; quest’ultimo, infatti, deve essere a conoscenza delle analisi che vengono effettuate sul campione.

***TIPOLOGIE DEI TEST NGS***

•Sequenziamento dell’intero genoma (*Whole genome sequencing, WGS*)

•Sequenziamento dell’intero esoma (*Whole exome sequencing, WES)*

•Sequenziamento mirato di pannelli di geni (*Target NGS testing o Panel NGS testing)*

***NGS: applicazioni***

(I rettangolini rossi rappresentano le Reads)

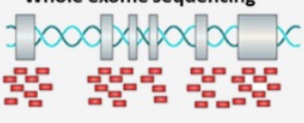
1. **Whole genome sequencing**

Regione del sequenziamento: intero genoma 

Profondità richiesta: >30X

Copre qualunque cosa - può identificare tutti i tipi di varianti, incluse: SNPs, INDELs and SV.

Il sequenziamento è lungo l’**intero genoma**, copre tutta la lunghezza del DNA, che siano introne, esone, promotore e sequenza ripetuta.



1. **Whole exome sequencing**

Regione del sequenziamento: tutte le parti esoniche,

quindi le parti codificanti.

Profondità richiesta: >50X ~ 100X

Identifica tutti i tipi di varianti, incluse SNPs, INDELs and SV in coding region.

Meno costoso del WGS.

1. **Targeted sequencing**

Permette la copertura più profonda di esoni di particolari geni.

Regione del sequenziamento: regioni specifiche.

Profondità richiesta: >500X

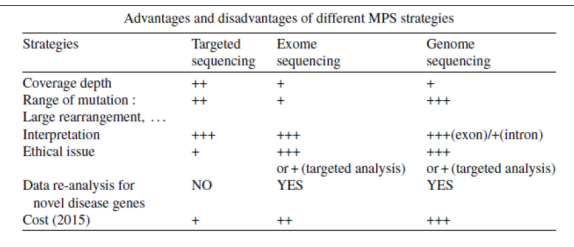
Identifica tutti i tipi di varianti, incluse SNPs, INDELs in specifiche regioni.

Costo più basso dei precedenti.

***NGS nella diagnosi delle malattie rare***

La scelta tra WGS, WES e Targeted sequencing varia in base a diversi parametri.

Può partire ad esempio dall’analisi dei costi o dall’analisi del parametro che si vuole valutare.

*Ad esempio*: il costo di un targeted sequencing di 150 geni è equivalente al costo dell’analisi di 3-4 geni tramite il Sanger sequencing.

Quest’ultimo, paradossalmente, è molto più costoso rispetto ad un next generation sequencing.

A parità di campione, si utilizza il Sanger sequencing, poiché utilizzare un next generation sequencing per un solo campione è uno spreco.

Quando si hanno più campioni invece si opta per un NGS.

La **copertura delle reads** sarà maggiore in un targeted sequencing rispetto ad un exome sequencing, poiché nel targeted sequencing si metteranno meno geni che di conseguenza avranno una copertura più adeguata.

L**’interpretazione** può essere difficile in tutti i casi, ma i problemi etici che si sollevano con il targeted sequencing sono molto minori rispetto all’ exome e al genome sequencing.

I **problemi etici** che si sollevano con la targeted sequencing sono minori rispetto alle altre due, poiché i risultati ottenuti saranno relativi ai geni che sono già stati correlati alla specifica malattia.

*Ad esempio*: Nel targeted sequencing viene richiesto di trovare geni specifici correlati ad una malattia (il paziente è dunque a conoscenza di questa ricerca e della possibilità di risultare positivo), mentre con il genome sequencing si possono trovare geni, correlati a tantissime malattie, che non sono stati richiesti e di conseguenza si avrebbe l’obbligo di doverli comunicare a qualcuno competente.

Il **costo** è molto più basso per il targeted sequencing.

***FATTORI DETERMINANTI NELLA SCELTA DELL’APPROCCIO***

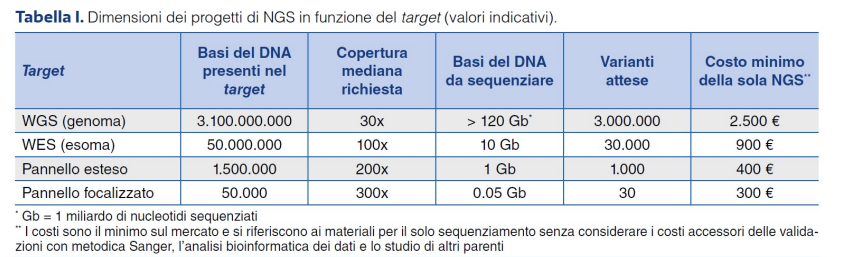
***Costi:*** il costo del sequenziamento di un esoma può oggi essere minore del sequenziamento di un pannello di geni.

***Finalità:*** malattie ad elevata eterogeneità genetica/malattie mendeliane (o sospette tali) con gene ancora non noto.

***Sensibilità:*** è in funzione del coverage delle sequenze indagate (n. letture “reads” dei singoli tratti del DNA esaminati e del grado di sovrapposizione tra queste).

***Probabilità di ottenere IF e VUS:*** dipende dall’ampiezza della porzione di genoma analizzata e interrogata.

***Archiviazione dei dati:*** necessità di disporre di piattaforme adeguate all’archiviazione di un gran volume di dati.



Il livello di coverage viene scelto in base al metodo che si utilizza, nonché ad altri fattori come la dimensione del genoma di riferimento, i livelli di espressione genica, l'applicazione specifica di interesse, la letteratura pubblicata e le migliori pratiche della comunità scientifica.

Di seguito sono elencati esempi di raccomandazioni sulla copertura del sequenziamento per alcuni metodi comuni.

***WES & WGS***

Importante miglioramento per la prospettiva di ottenere una diagnosi genetica perché geneticamente e fenotipicamente agnostico nell'approccio.

“Agnostico” si riferisce al fatto che si parte dal nulla, da qualcosa di non noto.

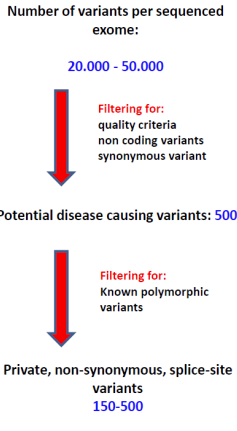
Un gene già correlato ad una malattia potrebbe essere correlato a qualcosa che ancora non è stato scoperto.

***SEQUENZIAMENTO DELL’INTERO GENOMA (WGS)***

Perché non viene effettuato direttamento un WGS?

Perché l’analisi mediante WGS presenta ancora molte limitazioni e non viene considerata come applicabile a livello diagnostico per le difficoltà interpretative e per i costi elevati.

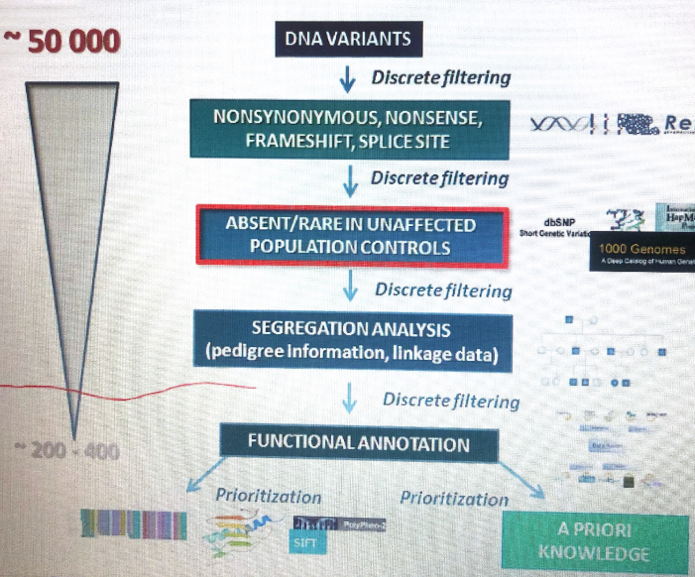
***WHOLE EXOME SEQUENCING (WES)***

il Whole esoma sequencing è una tecnica che permette di sequenziare tutti i tratti del genoma umano che codificano proteine, rappresentando meno del 2% del genoma totale di un individuo. Si stima che l'esoma contenga circa l'85% delle mutazioni associate a patologie genetiche. I primi successi di questa tecnica risalgono al 2009, quando sono stati diagnosticati tre pazienti affetti da malattie rare. Nel 25% dei casi analizzati è stato possibile identificare il gene responsabile della malattia.

Il Whole esoma sequencing viene principalmente utilizzato per le patologie le cui basi molecolari sono ancora sconosciute. Un vantaggio significativo è la possibilità di analizzare e identificare nuovi geni malattia in pazienti già sequenziati, nonché di individuare mutazioni in geni malattia noti in fenotipi "atipici". Tuttavia, uno dei limiti di questa tecnica è la necessità di un adeguato sistema di archiviazione dei dati, in quanto l'enorme quantità di informazioni generate richiede una gestione efficiente.

Quando un pannello di geni specifici correlati a una determinata malattia non produce risultati utili, si perde la possibilità di identificare nuovi geni correlati. Molti dei nuovi geni associati a malattie sono stati scoperti proprio attraverso l'utilizzo del Whole esoma sequencing.

Durante il whole esoma sequencing, vengono identificate un numero significativo di varianti, che può essere compreso tra 20.000 e 50.000. Successivamente, è necessario applicare un sistema di filtraggio per discriminare quali varianti siano rilevanti. Questo processo coinvolge la valutazione se la variante si trova nella regione codificante del gene, se provoca un cambiamento proteico, se il cambiamento avviene in una zona funzionale e se ha un impatto significativo sulla proteina. Infine, si valuta se la variante è già presente nei database e con quale frequenza. Queste azioni costituiscono il processo di "filtro" per identificare le varianti più rilevanti e significative.

***FILTRAZIONE DELLE VARIANTI E PRIORITIZZAZIONE***

Facendo l’analisi di un “trio”, ovvero padre-madre-figlio, se il figlio risulta l’unico ad essere affetto, l’analisi di filtraggio sarà vedere la presenza delle varianti in omozigosi provenienti dalla madre e dal padre. Quindi la segregazione della variante sarà un elemento importante!

***CLINICAL EXOME SEQUENCING***

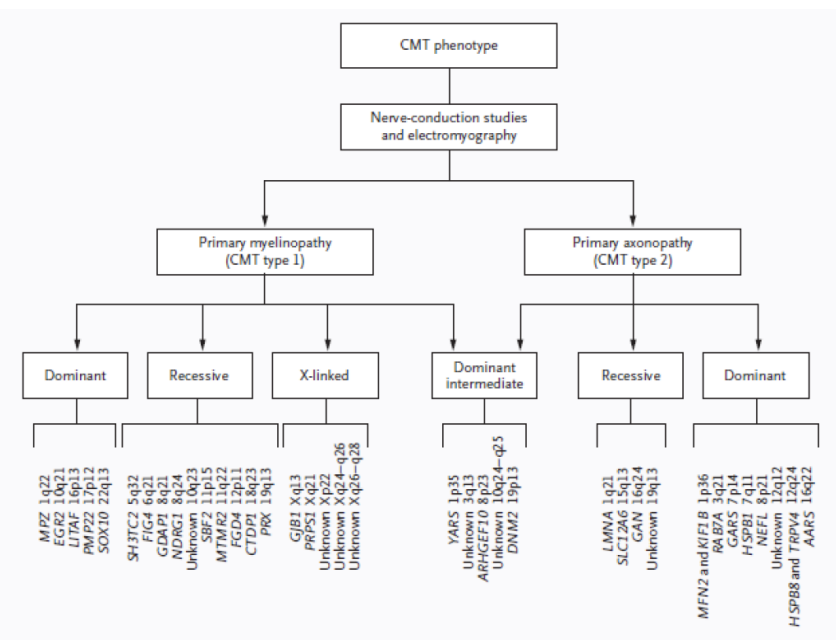
L’esoma clinico è un esoma limitato ai geni mendeliani, cioè i geni associati alle malattie genetiche.

E’ dunque il sequenziamento delle regioni codificanti dei geni noti per essere associati a malattia (OMIM).

***PANNELLO GENICO***

I geni inclusi nei panel clinici possono variare a seconda del laboratorio di analisi. Esistono due tipi principali di pannelli: il pannello fisico e il pannello virtuale.

Il pannello fisico è designato da un laboratorio e include geni associati a un ampio fenotipo (ad esempio, cardiomiopatia) o a sindromi riconoscibili con eterogeneità genetica (ad esempio, la Sindrome di Noonan). In questo caso, vengono analizzati solo i geni specifici per una determinata patologia.

D'altra parte, il pannello virtuale è creato dai clinici in base a sospette diagnosi o fenotipi. È come un guardaroba pieno di vestiti, in cui è possibile selezionare diversi "capi" (geni) da analizzare.

Un esempio di malattia ad elevata eterogeneità genetica analizzata con un pannello fisico è la Charcot-Marie-Tooth (CMT), una patologia demielinizzante causata da mutazioni genetiche. La CMT può essere suddivisa in due forme principali: CMT1 e CMT2, che presentano diversi modelli di ereditarietà (autosomica recessiva, autosomica dominante e legata all'X). Poiché coinvolge molti geni, la scelta tra esoma e pannello è importante. Di solito, viene preferito un pannello clinico perché consente di selezionare solo i geni rilevanti per la patologia in questione, ottimizzando così i tempi e i costi dell'analisi. Tuttavia, se il pannello risulta negativo, è possibile ampliare l'analisi a ulteriori geni correlati alla patologia mediante l'utilizzo di un esoma clinico "virtuale".

La scelta tra pannello e esoma dipende anche dal fenotipo del paziente, dai costi e dai tempi di analisi. In alcuni casi, i fenotipi clinici sono così eterogenei che non possono essere associati a un singolo gene. Di solito, il primo step è l'esecuzione di un pannello, poiché è un'opzione più immediata e accessibile, che consente di fornire rapidamente una risposta al paziente. Se il pannello risulta negativo, si può passare all'esoma clinico. In alcuni casi, se il laboratorio è ben attrezzato, è possibile eseguire due esomi clinici.

In conclusione, la scelta tra pannello e esoma dipende dalla specificità della patologia, dal fenotipo del paziente e dalle risorse disponibili, e viene valutata caso per caso.

***TARGETED SEQUENCING***

*Indicazioni principali*:

patologie dove mutazioni in geni noti sono responsabili della maggior parte dei casi (es. Teleangectasia Emorragica Ereditaria - mutazioni geni ENG, ALK1 e SMDA4 nell’80-87% dei casi).

*Vantaggi:*

Arricchimento specifiche regioni genomiche e di non trovare le incidental findings (IF);

Abbattimento dei costi;

Analisi di un numero elevato di pazienti contemporaneamente;

*Limiti:*

Necessità di un ragionamento del pannello con l’identificazione di nuovi geni associati alla malattia di interesse (*es. un pannello comprato 5 anni fa deve essere riacquistato poiché non è aggiornato*).

***ESOMA VS TARGETED***

L’esoma ha il vantaggio di fare un'analisi imparziale (unbiased) riguardo a quale set di geni viene analizzato, consentendo l'interrogazione parallela della maggior parte dei geni nel genoma umano.

Questo porterà di conseguenza a più problemi interpretativi.

A differenza del pannello, che esplora regioni selezionate, l’esoma clinico può permettere di scoprire nuovi geni correlati a malattie.

***TRIO-WES “GENE-AGNOSTIC”***

Una strategia che consente di condurre un'indagine più accurata nel sequenziamento del pool esome è l'approccio del trio, che coinvolge la madre, il padre e il figlio. Questo approccio può rivelare varianti incidentali, nonché varianti di significato sconosciuto (variants of uncertain significance, VUS), che sono varianti di sequenza con effetti funzionali e clinici non definiti. L'interpretazione dei dati rappresenta una delle sfide nell'utilizzo di piattaforme ad alta tecnologia.

L'analisi del trio WES (Whole Exome Sequencing) è particolarmente adatta per l'analisi delle malattie pediatriche. Ciò è dovuto a diversi motivi:

Le malattie pediatriche si manifestano precocemente e presentano varianti causali di natura e frequenza diverse.

Queste malattie sono correlate a varianti a completa penetranza.

L'analisi del genitore e del bambino consente di ottenere informazioni più complete.

Il filtraggio del trio genitori-figli è utile per l'identificazione di mutazioni de novo nelle condizioni sporadiche o nell'ereditarietà recessiva. Questo approccio consente di fenotipizzare il bambino e analizzare l'esoma sia del bambino che dei genitori. Ciò permette di identificare varianti ereditate dal padre e dalla madre, confermando la patogenicità e la correlazione attraverso l'omozigosi (due varianti dello stesso gene trasmesse al bambino). Tuttavia, nel caso in cui i genitori non siano i genitori biologici del bambino, si richiedono ulteriori colloqui per ottenere informazioni supplementari, e in alcuni casi può essere necessario ricorrere al reverse phenotyping.

Il deep phenotyping è una descrizione clinica accurata del paziente. L'approccio del reverse phenotyping, invece, è un metodo di ricerca nella genomica clinica che ha il potenziale per trasformare la comprensione degli effetti delle variazioni genetiche sulla salute e sulla malattia umana. Ad esempio, nel caso di un paziente, l'analisi genomica tramite NGS fornisce dati genomici correlati a caratteristiche cliniche che il paziente non presenta al momento dell'analisi. Pertanto, si richiede al medico di eseguire un'analisi clinica approfondita del paziente al fine di individuare segni della malattia identificata attraverso l'analisi genomica.