

Nuove metodologie di ricerca sulle sordità ereditarie Next Generation Sequencing (NGS)

Ambrosetti Umberto

Università degli Studi di Milano, Dipartimento di Scienze Cliniche e di Comunità,
U.O.C. Audiologia, Fondazione I.R.C.C.S. "Ca' Granda", Ospedale Maggiore Policlinico, Milano.

INTRODUZIONE

La sordità neurosensoriale (NSHL) ad eziologia genetica è il deficit uditivo più diffuso [1].

A volte la NSHL è parte di un complesso sindromico: ad oggi, sono state identificate oltre 400 sindromi in cui la sordità è correlata ad altre specifiche alterazioni patologiche.

Le cause ereditarie non sindromiche rappresentano almeno il 50% delle ipoacusie neurosensoriali profonde [2]. Esse vengono trasmesse con modalità mendeliana semplice e si suddividono in forma autosomiche recessive (75-80%), dominanti (20%), X-linked (2-5%) e mitocondriali (1%) *vedi figura 1.*

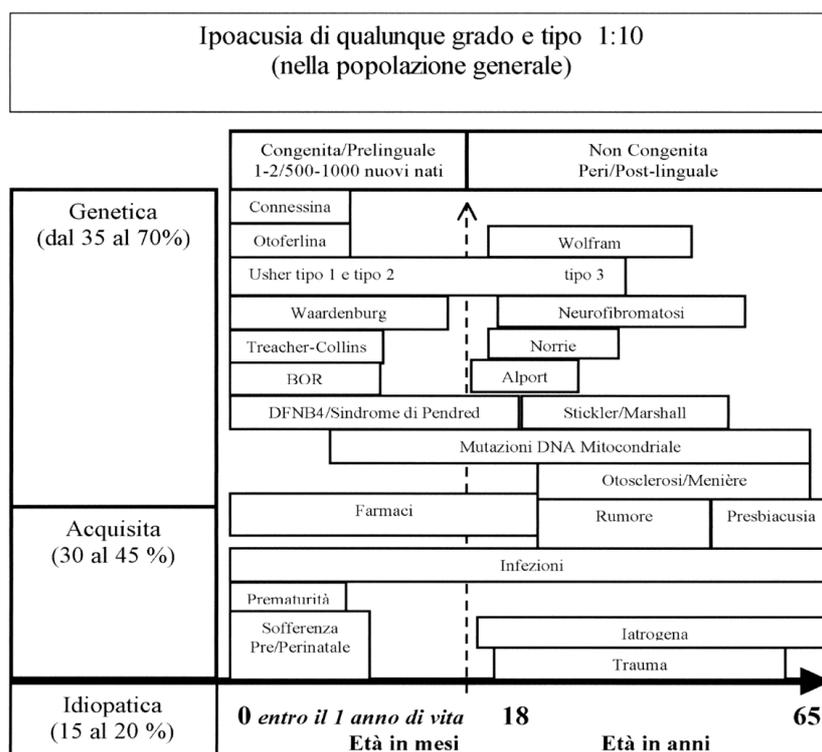


Figura 1-Distribuzione per eziologia ed età d'insorgenza dell'ipoacusia (da A. Martini e A. Castiglione, 2011) [2].

La sordità autosomica recessiva insorge solitamente in epoca pre-linguale, è di entità grave o profonda, non progressiva [2], mentre la forma autosomica dominante è prevalentemente

caratterizzata da insorgenza post-linguale, esordendo tra la II e la V decade di vita, ed è progressiva [3].

Nella NSHL ereditaria la correlazione genotipo/fenotipo è estremamente variabile rispetto alle altre patologie di tipo ereditario e la comprensione di queste correlazioni è stata ed è tuttora una grande sfida scientifica [4].

La prevalenza della NSHL nei bambini in epoca pre-linguale è di circa 1,3 per ogni 1000 nati e si raddoppia approssimativamente nei bambini di 5 anni di età [5].

L'identificazione dei geni causa di sordità è stata una grande scoperta: dal 1995 vedi figura 2 ad oggi sono stati scoperti 100 geni correlati alla sordità [6].

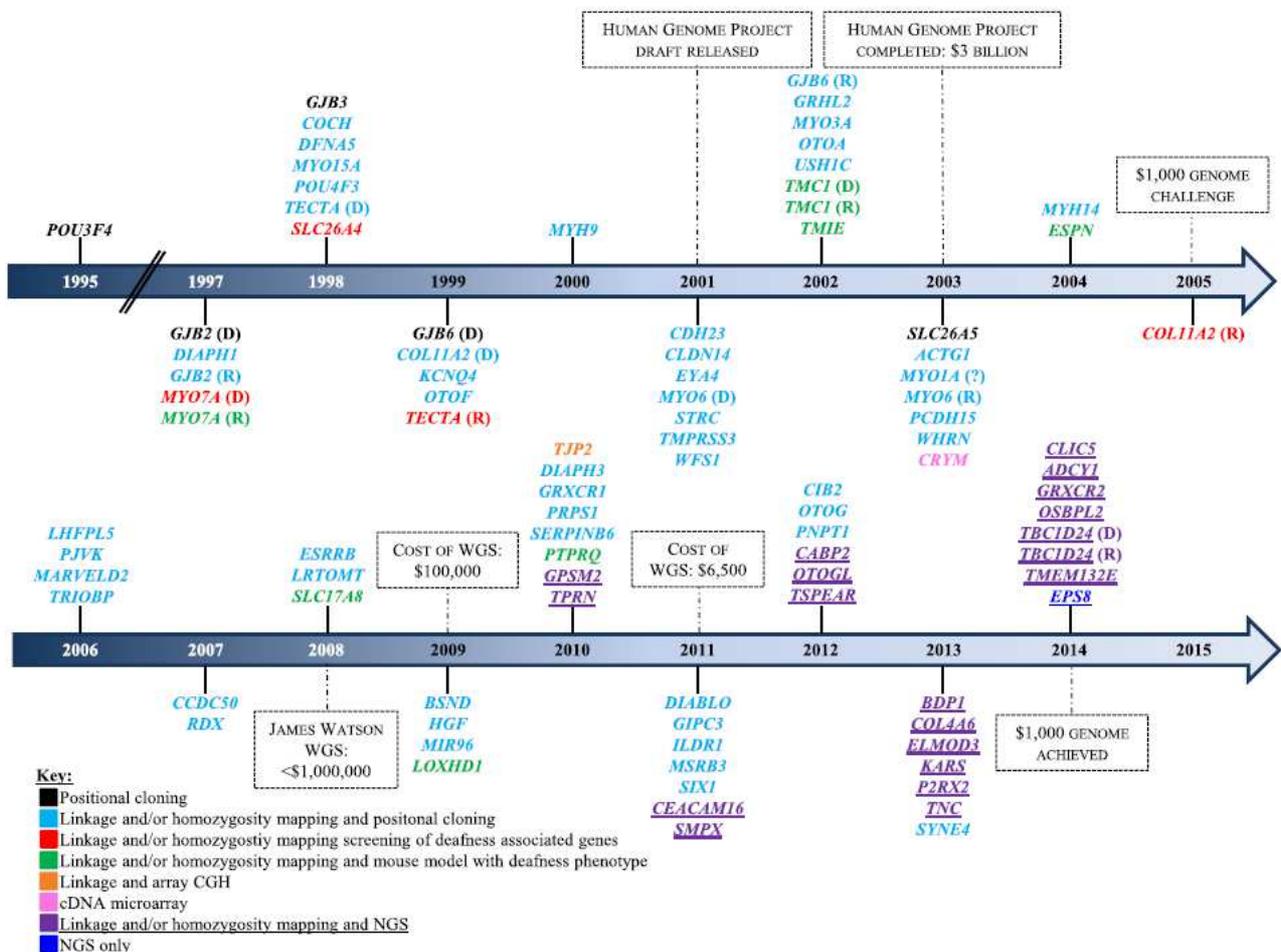


Figura 2-Cronologia dei geni NSHL identificati secondo il metodo di scoperta. *Abbreviazioni: D forma dominante di HL; R forma recessiva di HL; WES sequenziamento dell'esoma; WGS sequenziamento dell'intero genoma* [6].

La scoperta di nuovi geni correlati alla sordità è stata effettuata utilizzando la metodica di “Massive Parallele Sequencing” (MPS) o la “Next Generation Sequencing” (NGS) metodica più recente,

sviluppata successivamente alla tecnologia di sequenziamento di Sanger.¹ Il “*Sanger sequencing*” realizzato nel 1975, ha rappresentato il gold standard della diagnostica genetica molecolare fino all’introduzione nel 2008 della tecnica NGS, detta anche “*Second Generation Sequencing*” o “*High-throughput Sequencing*” [10-12].

“High-throughput Sequencing” è un termine molto diffuso per identificare le moderne metodiche, che consentono di analizzare un’elevatissima quantità di sequenze in breve tempo e a costi relativamente contenuti.

Il “*Sanger sequencing*” è il metodo di sequenziamento più tradizionale che a volte è più preciso, ma è decisamente più lento e molto più costoso.

Mentre un sequenziatore capillare con tecnica Sanger ogni giorno consente di valutare un campione dell’ordine di migliaia di basi, le macchine “*High-throughput sequencing*” analizzano campioni dell’ordine di gigabasi. Si parla di sequenziamento ad alta resa perché, a differenza del sequenziamento tradizionale effettuato con il metodo Sanger, è possibile sequenziare moltissimi frammenti in parallelo. I frammenti di queste reazioni vengono fatti correre su un gel che permette la separazione dei vari frammenti con la risoluzione di un nucleotide. La reazione viene letta direttamente sul gel. Attualmente è possibile effettuare una sola reazione utilizzando i 4 deossiribonucleotidi marcati con fluorocromi e il sequenziamento è divenuto molto più veloce utilizzando appropriati lettori ottici.

Con la NGS è possibile analizzare sia il DNA che l’RNA.

IDENTIFICAZIONE DELLA CAUSA DELLA SORDITÀ GENETICA NON SINDROMICA

Nel 1988 fu identificato il primo locus sul cromosoma Xq applicando un approccio di linkage² e questo è stato il primo passo di un’analisi sviluppata in 7 anni che ha portato all’identificazione del

¹ Il **sistema Sanger**, dal nome del suo inventore, premio Nobel per questa scoperta, messo a punto nel 1975 è uno dei primi metodi per ottenere la sequenza nucleotidica del DNA. E’ un metodo enzimatico che si basa sull’utilizzo di nucleotidi modificati per interrompere la reazione di sintesi in posizioni specifiche. Il protocollo classico richiede un template di DNA a singolo filamento, un primer per iniziare la reazione di polimerizzazione, una DNA polimerasi per terminare la reazione. I nucleotidi modificati o il primer devono essere marcati per poter visualizzare le bande di frammenti di DNA neosintetizzato dopo aver effettuato l’elettroforesi. Il campione di DNA da sequenziare è diviso in 4 reazioni, separate ognuna delle quali contiene la DNA polimerasi e tutti i 4 deossiribonucleotidi. [7-9].

² Con il termine **linkage** si intende l’associazione tra geni. Due geni sono associati se risultano vicini su un cromosoma. Questa vicinanza fa sì che i geni tendano ad essere ereditati assieme (perché sarebbe raro un crossing over che porti proprio alla separazione di quei due geni vicini). I genetisti fanno delle analisi basate sull’associazione tra i geni, dette appunto analisi di linkage, che permettono ad esempio di arrivare a scoprire il gene responsabile di un fenotipo basandosi sull’associazione (quindi sul linkage) del gene con un marcatore (una sequenza di DNA di cui si conosce la posizione).

gene **POU3F4**³ [15]; cinque anni dopo con la stessa metodica è stato identificato il gene **DIAPH1**⁴ localizzato sul cromosoma 5q31 [16].

I geni citati sono stati identificati in un'epoca in cui la genetica molecolare era ancora agli albori e con elevato rischio di errore. Questo lavoro è servito tuttavia per i successivi studi di identificazione dei geni candidati nella NSHL.

I nuovi loci identificati come causa di NSHL sono stati classificati con un numero progressivo, procedura che continua tutt'oggi.

I loci che hanno una trasmissione di tipo dominante sono denominati con l'abbreviazione di **DFNA**, mentre i loci con ereditarietà recessiva sono abbreviati con **DFNB**, quelli X-linked con **DFNX**.

Ad oggi sono stati identificati e pubblicati su riviste peer-reviewed 141 loci. E' stato dimostrato inoltre che alcuni geni possono essere responsabili sia di forme dominanti che recessive, in rapporto alla mutazione.

Studi effettuati su popolazioni caratterizzate da elevati tassi di matrimonio tra consanguinei hanno consentito di scoprire un numero significativo di geni correlati alla sordità recessiva [17].

Un dato di notevole rilievo riguarda l'evidenza che la distribuzione dei geni della sordità può variare molto nelle differenti etnie e popolazioni, ad esempio: il gene GJB2, uno dei più importanti geni causa di sordità in Europa (28-63 % della sordità) [18] è stato per molti anni poco ricercato nella diagnostica della sordità perché in altre popolazioni (Arabia Saudita, Africa, Afro-americani, Ispanici dei Caraibi, Marocchini, Pakistani) questa mutazione è poco rilevante [19- 21]. Questi dati evidenziano come non tutte le popolazioni presentano lo stesso carico di mutazioni dei geni della

³ Il gene **POU3F4** del cromosoma X fornisce istruzioni per produrre una proteina che aiuta a regolare l'attività di altri geni. Sulla base di questo ruolo, la proteina è chiamata fattore di trascrizione. Il gene POU3F4 è parte di una grande famiglia di geni chiamati geni dominio POU, che producono fattori di trascrizione. I geni POU giocano un ruolo nel determinare i tipi di cellule del sistema nervoso centrale durante lo sviluppo precoce. Il gene POU3F4 è coinvolto nello sviluppo dell'orecchio medio e interno, ed è attivo anche in alcune regioni del cervello prima della nascita. Le mutazioni del gene POU3F4 determinano una forma di sordità non sindromica chiamata **DFN3**. Questa forma di sordità di solito comporta anomalie anatomiche sia dell'orecchio interno e del medio (perdita uditiva mista). La sordità nei maschi affetti è prelinguale, progressiva, mista con fissazione stapediale; i soggetti che vengono operati di stapedotomia per questa forma di sordità sono ad alto rischio della complicanza denominata "*gusher perilinfatico*". Sono state identificate più di 15 mutazioni POU3F4. La maggior parte di questi cambiamenti genetici altera blocchi singoli di proteine (aminoacidi) nella proteina POU3F4 o eliminare una piccola quantità di materiale genetico del gene. Le mutazioni impediscono alle cellule di produrre qualsiasi proteina POU3F4 o altera regioni della proteina che sono fondamentali per il legame al DNA. La mancanza della proteina POU3F4, probabilmente, sovrverte il normale sviluppo delle strutture dell'orecchio medio e interno, con conseguente perdita dell'udito. [13, 14]

⁴ Il gene **DIAPH1** è un omologo del gene diafana Drosophila, ed è causa di una sordità neurosensoriale autosomica dominante, non sindromica progressiva interessante le basse frequenze. E' coinvolta nella polimerizzazione dell'actina. Si è ipotizzato che questo gene possa avere un ruolo nella regolazione della polimerizzazione nelle cellule cigliate dell'orecchio interno.

sordità e per questo il genetista che opera in una società multietnica come ormai è divenuta l'Italia, deve disporre di dati relativi alle mutazioni più ricorrenti nelle diverse etnie di provenienza.

DNA E IL “PROGETTO GENOMA UMANO”

Il sequenziamento del DNA è avvenuto nel 1968, circa 15 anni dopo la scoperta dell'acido desossiribonucleico caratterizzato dalla struttura a doppia elica (1953) [22].

Il primo sequenziamento completo del DNA di un organismo, da parte di Nicklen e Coulson, riguardò il fago $\Phi X174$. Da quando le tecniche di sequenziamento sono migliorate, sono state sequenziate molecole maggiori di 200 kb (es: citomegalovirus umano) ed è nata l'analisi computerizzata e la bioinformatica.

Gli sforzi del sequenziamento hanno raggiunto nuove vette con il “*Progetto genoma umano*” iniziato nel 1990 e terminato nel 2000, circa 50 anni dopo la scoperta del DNA, permettendo un fondamentale impulso nella conoscenza e consentendo il sequenziamento massivo di bilioni di coppie di basi.

La durata del Progetto Genoma Umano coordinato dal Dipartimento per l'Energia degli USA (DOE) e dal Centro Nazionale della Ricerca sul genoma umano che fa parte del National Institutes of Health (NIAH) si era stimato essere di circa 15 anni e costare circa 3 bilioni di dollari.

Il grande impegno ha portato ad una rete di collaborazione tra diversi centri di sequenziamento negli USA, Europa e Giappone.

Nel 1994 è stata pubblicata una mappa genetica dettagliata del genoma umano che comprendeva 5840 loci mappati [23] con un numero di geni inferiore al previsto, circa 30.000.

Nel 1999 il Progetto Genoma Umano ha superato la determinazione di un bilione di coppie di basi ed è stata riportata per la prima volta la sequenza completa di un cromosoma umano, il 22 [24].

Nel febbraio 2001 sono stati pubblicati nella stessa settimana su Science e Nature le sequenze del genoma umano realizzate dalla società privata CELERA Genomix e dal Consorzio Internazionale pubblico sotto l'egida dei NIH (National Institutes of Health) [25, 26].

Attualmente il Progetto Genoma è in fase di prosecuzione, con l'approccio anche al mappaggio di numerosi organismi modello come: *Escherichia coli*, *Saccaromyces cerevisiae*, *Caenorhabditis elegans*, *Drosophila melanogaster*, mammiferi e pesci.

Va precisato che la complessità di un organismo non è solo correlata al numero di geni, ma anche al modo in cui i differenti geni controllano la formazione delle proteine: ad esempio, un gene spesso determina la formazione di più di una proteina, oppure più geni possono concorrere alla formazione di una sola proteina funzionante.

Nel luglio 2015, è stata pubblicata su Nature la mappa dell'attività dei geni di 18 differenti organi umani, ovvero una mappa epigenetica. Tale ricerca porterà ad una comprensione più approfondita

del funzionamento e dell'intima organizzazione dei tessuti umani per meglio comprendere le varie patologie [27].

TECNOLOGIA NGS

Nel 2005 la Roche ha messo in commercio un'apparecchiatura basata sulla creazione di piattaforme NGS che permettono un rapido sequenziamento. Ne sono seguite altre con l'obiettivo di ridurre i costi e i tempi di lavoro [28].

Questo sviluppo tecnologico ha permesso di trovare velocemente varianti per tutti i disordini mendeliani compresa la sordità e il sequenziamento a cominciare da piccoli pannelli genetici per tutti gli esoni codificanti fino al sequenziamento dell'esoma (WES) o al sequenziamento completo del DNA genomico "*Whole Genome Sequencing*" (WGS) o "*Genome Sequencing*".

Con la realizzazione di questi obiettivi è stato rivoluzionato l'approccio a nuovi geni, e loci correlati che contengono centinaia di geni da screenare che attualmente possono essere sequenziati in parallelo.

Utilizzando questo approccio nel periodo 2010-2015 sono stati identificati 21 geni correlati alla sordità [29, 30].

Questa tecnologia, inizialmente impiegata solo a scopo di ricerca, si è in breve tempo estesa come strumento indispensabile per la diagnosi.

Nella sordità vi sono centinaia di geni correlati a fenotipi simili ed è molto difficile correlare le ipotesi pre-diagnostiche con i dati clinici ed audiometrici: per esempio, l'insorgenza di alcune forme di sordità dominanti e recessive non sono differenziabili nonostante vi siano differenti eziologie genetiche sottostanti. Lo screening di centinaia di geni clinicamente rilevanti con una singola prova è una strategia efficace sia in termini di costo che di tempo per l'approccio diagnostico delle sordità genetiche eterogenee [31, 32].

Attualmente sono stati identificati 49 loci causa di sordità senza identificazione di un gene specifico, pubblicati su riviste peer-reviewed, il che corrisponde al 35% [15] dei loci totali e probabilmente ci sono circa 49 geni della sordità in attesa di scoperta.

La variabilità intra-familiare della sordità anche nelle forme dominanti, es. nei gemelli, dimostra che esistono dei fattori ambientali o genetici secondari (geni modificatori) che determinano una differente espressività con un'incompleta o ridotta penetranza.

Alla luce dei dati clinici, audiometrici, vestibolari e radiologici che non possono essere correlati univocamente ad una specifica eziologia dovuta ad una mutazione genetica, la NGS permette lo studio non più del singolo gene ma contemporaneamente dei numerosi geni validati per la sordità.

I test genetici, un tempo applicati alla fine dell'iter diagnostico-clinico per la sordità, attualmente hanno invece un ruolo preminente subito dopo l'esame obiettivo, l'esame audiometrico e

l'anamnesi familiare. La disponibilità dei test NGS è diventata una parte indispensabile del lavoro clinico. E'imprescindibile che una struttura audiologica di III livello all'avanguardia abbia nel suo organico un qualificato specialista in Genetica Clinica e disponga di un laboratorio di Genetica Clinica attrezzato e aggiornato sulla moderna tecnologia NGS.

Tale tecnologia richiede un aggiornamento continuo realizzando sempre nuovi e più estesi o specifici pannelli in accordo con le più recenti scoperte.

Attualmente i costi della NGS si sono notevolmente ridotti, e nel futuro continueranno a diminuire con lo sviluppo tecnologico.

Esistono diversi sistemi NGS, sviluppati da differenti aziende. Ogni piattaforma NGS si distingue per la propria peculiare combinazione di protocolli di preparazione del template, sequenziamento con avanzati sistemi di imaging e analisi del dato. Tutti questi sistemi condividono almeno tre fasi fondamentali:

- 1) Preparazione e immobilizzazione del DNA, cioè la preparazione della cosiddetta "*sequencing library*";
- 2) Reazione di amplificazione;
- 3) Reazione di sequenziamento.

Esistono diversi metodi per preparare una sequencing library, a seconda del tipo di piattaforma che si desidera utilizzare (le piattaforme NGS più diffuse sono Life Technologies, Illumina, Roche e Pacific Biosciences, Ion Torrent).

I tre step fondamentali nella preparazione di una sequencing library di DNA sono:

- frammentazione e dimensionamento del campione ad una lunghezza predefinita conosciuta;
- aggancio degli adattatori (adapters) alle estremità dei frammenti ottenuti;
- quantificazione della libreria ottenuta.

Nella preparazione delle sequencing library a partire da RNA è presente un'ulteriore step, quello della conversione dell'RNA in una molecola di cDNA; in questo caso, la frammentazione può avvenire prima o dopo la sintesi del cDNA.

Frammentazione e dimensionamento del campione

La frammentazione dell'acido nucleico può essere fatta con metodi fisici (frammentazione acustica e sonicazione), enzimatici (tramite l'uso di endonucleasi aspecifiche come la DNase I o la Fragmentase o di kit enzimatici commerciali come il Nextera tagmentation kit di Illumina, il quale, fra l'altro, non solo frammenta il DNA, ma procede all'immediato aggancio degli adattatori tramite una trasposasi) o chimici. Per la frammentazione del DNA si utilizza solitamente il metodo fisico o enzimatico (Covaris, ad esempio, è uno strumento utilizzato per la frammentazione acustica).

La dimensione dei frammenti da ottenere è un parametro chiave. Essa dipende sia dalla piattaforma che si intende utilizzare sia dallo scopo dell'analisi. Ad esempio, nel sistema Illumina si possono ottenere buoni risultati con frammenti fino a 1500 paia di basi (1500bp) di lunghezza. Tuttavia, nel caso in cui si desideri procedere al sequenziamento di un esoma, è raccomandabile creare delle sequencing library con inserti (così si chiama tecnicamente il frammento di DNA una volta che gli adattori sono stati agganciati alle sue estremità) non più lunghi di 200-250 bp, semplicemente perché la dimensione media di un esone umano è di 200 bp.

Aggancio degli adattatori

Subito dopo la fase di frammentazione si passa alla fase di aggiunta degli “*adattori*”, seguita da un ulteriore step di dimensionamento (sizing) nel quale si rimuovono tutti i frammenti di lunghezza non desiderata e tutti i dimeri di adattori (cioè i dimeri che gli adattori in eccesso hanno formato legandosi fra di loro). È bene ricordare che i dimeri di adattori possono ridurre sensibilmente la resa della reazione di sequenziamento e che sono particolarmente abbondanti laddove la quantità di DNA iniziale sia particolarmente bassa.

Preparazione e immobilizzazione del DNA (ovvero creazione della sequencing library)

Il campione di DNA viene preparato attraverso un processo di frammentazione casuale. Ai frammenti casuali così ottenuti vengono aggiunte delle sequenze predefinite (note come “adattatori” o adaptors) che sono necessarie per ancorare e immobilizzare i frammenti al supporto sul quale avrà luogo la reazione di sequenziamento. I frammenti di DNA così preparati tramite l'aggiunta degli adattatori costituiscono la cosiddetta libreria di sequenziamento (“sequencing library”). Esistono almeno tre diversi tipi di adattatori e quindi tre diversi modi di preparare la “sequencing library”: adattatori lineari, circolari e a bolla.

Esistono naturalmente anche tipi di supporto di ancoraggio diversi: ad esempio, nel sistema SOLiD i frammenti vengono ancorati ad una lastra di vetro.

Amplificazione

L'amplificazione può essere fatta in emulsione (sistema Roche e sistema SOLiD) o in soluzione. Ad esempio, nel sistema GS FLX (Roche) un frammento della sequencing library viene incorporato in una microscopica bolla di acqua assieme a dei cosiddetti “*enrichment beads*”, che sono praticamente delle piccole sfere a cui gli adattatori si possono legare. La reazione di amplificazione (PCR) viene fatta in questa microbolla acquosa, all'interno della quale il frammento di DNA viene amplificato numerose volte. Le copie clonali del frammento si legano poi all'enrichment bead

ricoprendone la superficie. Gli enrichment beads così ottenuti vengono poi depositati sulla cosiddetta piastra Picotiter.

Sequenziamento

Il sequenziamento avviene grazie a complessi meccanismi fluidici che, funzionando su scala microlitrica, regolano il flusso dei reagenti che vanno a cimentarsi col DNA immobilizzato. Ogni ciclo di sequenziamento consiste nel cimentare il DNA immobilizzato con una soluzione contenente un nucleotide (che, se complementare alla sequenza, viene incorporato), un successivo lavaggio e la registrazione dell'evento molecolare appena verificatosi.

La registrazione dell'evento molecolare avviene con un sistema per immagini (nel sistema GS FLX Roche, ad esempio, la reazione di incorporazione del nucleotide nel filamento complementare crescente avviene in concomitanza dell'emissione di un lampo di luce, da ciò il nome di “*pyrosequencing*”, dal greco antico *pyros*, che significa fuoco). Fa eccezione al sistema di registrazione per immagini la macchina Ion Torrent (Life Technologies) che, basata su un sistema di rilevazione con semiconduttore, riconosce l'emissione di ioni idrogeno che vengono rilasciati durante la reazione di polimerizzazione del DNA.

Come detto all'inizio, i sistemi NGS sono anche detti high-throughput (ad alta resa) perché tutti i frammenti immobilizzati sul supporto di sequenziamento vengono sequenziati in parallelo (a differenza del Sanger sequencing tradizionale, in cui si può sequenziare un solo frammento per volta).

Una volta completato il sequenziamento, i dati vengono analizzati nella fase computerizzata di analisi bioinformatica.

Analisi bioinformatica

Al termine dei processi di amplificazione e sequenziamento la macchina NGS raccoglie ed elabora le immagini acquisite e fornisce file di sequenze suddivise per pazienti pronte per essere inserite nel processo di analisi.

L'analisi dei dati NGS può essere schematizzata in due fasi principali: la fase di allineamento e la successiva fase d'identificazione di varianti.

Allineamento

Le sequenze generate vengono allineate sulla sequenza di riferimento nota. I maggiori limiti descritti nel processo di allineamento sono rappresentati dal posizionamento di sequenze in regioni

ripetute e dalla presenza di varianti strutturali quali grosse inserzioni o delezioni, inversioni, duplicazioni segmentali e altre differenze nel numero di copie. Queste problematiche sono in parte ridotte, ma non del tutto eliminate, dall'utilizzo di template di sequenza sovrapposti tra loro. A seguito dell'allineamento delle sequenze del genoma di riferimento è possibile osservare l'effettiva copertura del pannello, ovvero andare a osservare da quanti template è coperta una specifica regione genica di interesse.

Identificazioni di varianti

Successivamente all'allineamento delle sequenze al genoma di riferimento è possibile procedere con l'identificazione delle varianti. Questo passaggio si compone dell'applicazione di diversi algoritmi informatici che possono essere di volta in volta personalizzati a seconda del target ricercato. L'obiettivo finale dell'uso degli algoritmi è quello di effettuare uno screening di tutte le varianti osservate e fornire un report finale contenente solo le varianti su cui focalizzare l'attenzione. Lo screening effettuato da questi algoritmi è volto in primis all'eliminazione di tutte quelle varianti che possono, con elevata probabilità, essere correlate a errori di amplificazione: per esempio, tramite algoritmo si può richiedere che non vengano prese in considerazione le varianti rilevate in sequenze che non risultano essere coperte da un sufficiente numero di letture oppure, come altro esempio, chiedere che vengano eliminate le varianti rilevate in una frazione minima di letture sul totale delle letture di una data sequenza. Altre tipologie di algoritmi sono invece volti al filtraggio di varianti di sequenze note, come per esempio i polimorfismi a elevata frequenza nella popolazione.

Le procedure di allineamento e di identificazione delle varianti qui citate comportano un massiccio utilizzo di risorse informatiche e di lavoro, in termini di tempo e di specifiche competenze tecniche. La quantità di dati ottenuta è enorme, infatti vengono identificate migliaia di varianti. La riduzione del numero di varianti da analizzare avviene nella terza fase, quella di "*filtering e annotation*".

Filtering e annotation

Una volta identificate le varianti del paziente (che in un'analisi di esone o "*genome sequencing*" sono nell'ordine di migliaia), il software procede in automatico alla loro caratterizzazione, filtrandole (selezionando cioè solo quelle che potrebbero avere un significato fenotipico) e annotandole (fornendo cioè tutti i dati disponibili nei database e/o in letteratura per capire se si tratti di semplici polimorfismi, di varianti funzionali, di mutazioni patogene già descritte in letteratura o di varianti del tutto nuove e dunque dal possibile significato incerto).

Una volta che il software ha completato le tre fasi, i dati devono essere valutati dal genetista.

SEQUENZIATORI NGS



Fig. 3 Genome Sequencer FLX Roche 454

Roche454

Il 454 Roche è stato il primo sequenziatore NGS lanciato sul mercato.

Il sistema Roche 454 si basa sulla tecnologia del “*pyrosequencing*” (così chiamato per via dei pirofosfati che vengono sfruttati nella reazione di sequenziamento). Il 454 riconosce la sequenza grazie al susseguirsi di segnali luminosi causati dal distacco di un pirofosfato ogni qualvolta un nucleotide viene incorporato nel filamento in crescita. I frammenti che compongono la libreria di DNA vengono denaturati per ottenere filamenti singoli che vengono catturati dagli “*amplification beads*”, piccole sfere sulla cui superficie sono ancorati oligonucleotidi che riconoscono e legano gli adattatori (adaptors), che erano stati opportunamente aggiunti alle estremità di ogni frammento (non si dimentichi che, per definizione, la DNA library o sequencing library è l’insieme dei frammenti di DNA preparati tramite l’aggiunta degli adattatori) [28].

Quello che segue è un processo cosiddetto di “*PCR in emulsione (emulsion PCR)*”. Gli amplificati ottenuti vengono posti sulla “*picotiter plate*”, un supporto sul quale avviene la reazione di sequenziamento vera e propria. Gli amplificati vengono qui cimentati con una soluzione contenente un solo dNTP per volta (insieme ad altri reagenti che hanno lo scopo di catalizzare l’eventuale reazione luminosa). Se il dNTP somministrato (dATP, dGTP, dCTP o dTTP) è complementare alla prima base da copiare, esso viene incorporato nel filamento in crescita scatenando una reazione luminosa causata dal rilascio del pirofosfato. Il dNTP in eccesso viene poi degradato dall’enzima apirasi e lavato via. Gli amplificati vengono quindi cimentati con il dNTP successivo e così via. Come detto sopra, le reazioni luminose vengono registrate ed elaborate dalla macchina, il cui software ricostruisce la sequenza. Questo sistema, nel quale i dNTP vengono aggiunti uno alla volta ad ogni ciclo successivo, si dice “*sistema a interrogazione di singolo nucleotide*”.

Illumina

Illumina ha realizzato tre nuove piattaforme NGS: HiSeq 2000, NextSeq 500 e MiSeq.

- **HiSeq 2000** è dotato di due laser e quattro filtri di rilevazione di nucleotidi e di un hard-disk da 3 TB e può produrre fino a 600 Gb di dati per corsa (una corsa si completa in 8 giorni). Tramite il multiplexing (realizzabile con l'utilizzo dei kit specifici di primer e adattori P5/P7), HiSeq 2000 può processare anche migliaia di campioni in parallelo. Recentemente è stato lanciato anche il modello HiSeq 2500, in grado di sequenziare un intero genoma umano in un solo giorno.
- **NextSeq 500** è un sequenziatore di portata intermedia, utile per il sequenziamento di esoni in parallelo (idealmente da 12 fino a 100 esoni umani per volta).
- **MiSeq**, macchina dalle dimensioni ridotte è particolarmente adatta al sequenziamento di PCR e di campioni microbiologici. MiSeq può raggiungere un output di circa 15 Gb a corsa. Il tempo che intercorre dalla preparazione del campione al risultato finale è di circa 8 ore.

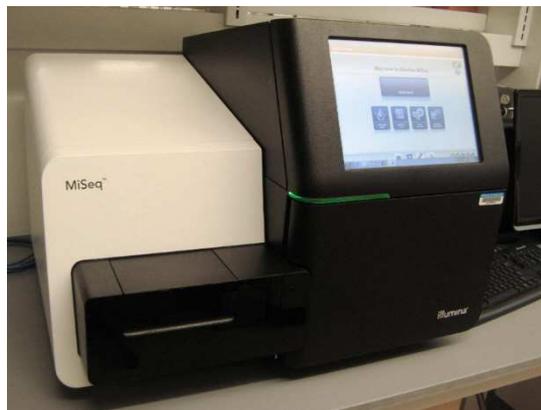


Fig. 4 Genome Sequencer MiSeq Illumina

Ion Torrent Personal Genome di Life Tecnology

Il PGM™ nasce come sequenziatore di tipo "benchtop", letteralmente da "bancone" con ridotte capacità di processamento rispetto alle strumentazioni concorrenti, con l'intento di rendere più economica e quindi maggiormente fruibile, la tecnologia NGS per tutti i progetti (targeted DNA sequencing, targeted RNA sequencing, microbial sequencing) in cui la quantità di basi da sequenziare è notevolmente inferiore rispetto al sequenziamento di un intero genoma umano.

Anche il protocollo Ion Torrent prevede l'amplificazione clonale del campione attraverso Emulsion PCR, ma il processo di sequenziamento presenta due aspetti particolari che lo caratterizzano rispetto alle altre tecnologie:

- a) utilizzo di semiconduttori come elemento strutturale dei supporti in cui viene dispensato il campione per il sequenziamento (Ion Chips);
- b) impiego di un sistema di rilevamento non basato su reazioni luminescenti ma su variazioni di potenziale.

Nucleo centrale dell'intero sistema di sequenziamento è l'Ion Chip. Esso è costituito da uno strato superficiale superiore, in cui sono ricavati dei pozzetti appositamente strutturati per poter accogliere

le beads e i reagenti per il sequenziamento. Ciascun pozzetto ha diametro di pochi micron, tale che in esso possa trovare spazio una sola biglia. Questo primo strato superficiale poggia su un secondo costituito da semiconduttori, che permette la trasmissione dei segnali allo strato ancora sottostante. Quest'ultimo è strutturato come una piastra di sensori, uno corrispondente a ciascun pozzetto, che hanno la capacità di registrare le piccolissime variazioni di pH che avvengono all'interno del pozzetto durante il sequenziamento, trasformandole in differenze di potenziale, ovvero, in dati digitali.



Fig. 5 Genome Sequencer Ion Torrent

MULTIPLEXING: UN'APPLICAZIONE FORMIDABILE DELLA NGS

Le moderne macchine di NGS consentono di processare un elevatissimo numero di sequenze in parallelo. Ciò ha reso possibile analizzare in tempi rapidi e con costi accettabili l'intero genoma di un individuo “*whole genome sequencing*” o più comunemente “*genome sequencing*” o parti di genoma di diversi individui in parallelo. In altre parole, a seconda delle dimensioni delle sequenze che si desidera analizzare, è possibile eseguire l'analisi di più campioni alla volta.

Non è infrequente, ad esempio, sequenziare l'intero esoma (exome sequencing) di almeno un paio di soggetti in un'unica corsa.

L'analisi in parallelo delle sequenze di più di un campione è detta in gergo multiplexing. Per poter procedere al multiplexing, è necessario aggiungere ai frammenti di ciascun campione una sequenza specifica, detta “etichetta” o “sequenza barcode”. La sequenza barcode consente di identificare in modo univoco tutti i frammenti di DNA di uno stesso individuo (o di uno stesso microorganismo).

Le librerie di DNA dei diversi individui così preparate possono poi essere unite in un'unica provetta per la reazione di sequenziamento in multiplexing. Tutte le reads ottenute durante la reazione conterranno anche le sequenze barcode (che vengono effettivamente amplificate e sequenziate né

più né meno come le sequenze di DNA a cui sono attaccate). Identificando le sequenze barcode è poi possibile procedere al de-multiplexing delle reads, cioè alla loro separazione sulla base dell'appartenenza ai diversi individui. Le reads così separate vengono poi allineate con le reference sequences dei database fino ad ottenere la sequenza reale di ciascun campione analizzato.

Tutto questo, che un tempo sarebbe stato possibile solo tramite Sanger sequencing a costi esorbitanti e in parecchie settimane, è ora fattibile in un paio di giorni al massimo. Le macchine NGS, inoltre, riducono al minimo l'intervento umano nell'analisi, poiché non forniscono un elettroferogramma da interpretare, ma la sequenza vera e propria dei nucleotidi e la lista delle varianti identificate. Il processo di alignment con le reference sequences e l'operazione di variant calling (cioè di identificazione delle varianti) sono infatti completamente automatizzate.

Anche la quantità di DNA necessaria è molto bassa: basti pensare che soli 30 ng di DNA possono essere sufficienti ad ottenere un'intero esoma!

Le tecnologie NGS presentano diversi vantaggi rispetto al sequenziamento di tipo Sanger. In primo luogo, un sistema NGS può produrre molte gigabasi di sequenze in minor tempo e con costi nettamente inferiori rispetto al metodo Sanger. Il grado di miniaturizzazione raggiungibile con queste tecnologie, infatti, è di gran lunga superiore rispetto a quello possibile nella tecnologia a capillari e ciò permette il processamento contemporaneo di una quantità di frammenti molto elevata; nella tecnologia a capillari è invece presente un limite fisico, dato dalla struttura stessa dei capillari, che non può essere superato. In secondo luogo, il passaggio di amplificazione clonale dei frammenti nella tecnologia NGS non prevede l'utilizzo di plasmidi per il clonaggio (come era avvenuto nel caso del sequenziamento "shotgun" nel Progetto Genoma) ma viene eseguito mediante un processo automatizzato inserito nel protocollo sperimentale (emulsion PCR, bridge PCR), il che favorisce un notevole risparmio economico e di tempo. Altra caratteristica che differenzia profondamente le tecnologie NGS dalla Sanger è la capacità di determinare in maniera accurata la presenza di alleli a bassa frequenza. Il dato prodotto da un sequenziatore a capillare deriva dall'osservazione dell'intensità e della lunghezza d'onda della fluorescenza emessa globalmente da tutti i frammenti di pari lunghezza che attraversano il capillare; per sua natura, la rilevazione del segnale dato dalla sovrapposizione delle emissioni di tutti i fluorocromi, fa sì che l'effetto del gruppo fluoroforo che marca la maggior parte dei frammenti, sovrasti l'effetto di un eventuale gruppo alternativo meno rappresentato, causandone la mancata rilevazione da parte dello strumento. Nelle NGS ciascun frammento viene invece sequenziato all'interno di un microreattore in maniera indipendente rispetto agli altri: i dati così prodotti vengono quindi acquisiti non in maniera massiva, ma per singolo reattore. Questo sistema rende possibile la determinazione della sequenza specifica del frammento contenuto in ciascun reattore, e consente di mettere in evidenza le varianti presenti.

In questo modo è possibile dedurre la frazione di ciascun allele relativa al campione di partenza, poiché il numero delle volte che esso viene sequenziato è proporzionale alla sua quantità relativa originaria.

A fronte di una serie di aspetti positivi e dalle proficue ricadute sugli attuali ambiti di ricerca, le NGS presentano delle caratteristiche limitanti che ne condizionano l'applicazione.

Una prima limitazione è la lunghezza dei frammenti (reads) che possono essere sequenziati. Generalmente si aggirano intorno alle 200-300 bp, contro le 1000 bp di un sequenziatore a capillari. Questo aspetto complica notevolmente l'assemblaggio delle sequenze contigue, soprattutto nei casi di genomi non noti o con ampie zone ripetute o contenenti dei riarrangiamenti, rendendo difficoltosa la separazione in fasi delle varianti.

Ulteriore fattore limitante è l'aumentata percentuale di errore (falsi positivi e falsi negativi) nelle chiamate delle basi, se paragonata al sistema Sanger. Tale limite, in gran parte, viene superato attraverso l'aumento del livello di coverage (quantità di sequenze analizzate relative ad un frammento di DNA) e attraverso l'ausilio di polimerasi ad altissima efficienza.

È questo uno dei principali motivi per cui, nella pratica clinico-diagnostica, il sequenziamento Sanger mantiene tuttora la sua importanza. Strettamente legata a questo aspetto, è la necessità di continuo sviluppo di algoritmi bioinformatici, che consentano di analizzare con elevata efficacia i dati prodotti. L'aspetto informatico ha profonda rilevanza nelle tecnologie NGS. Da un lato si riscontra la necessità di continue implementazioni dei software di analisi per rendere sempre più stabili e efficaci le chiamate delle basi, dall'altro sono necessari molti software dedicati per adempiere alle diverse fasi del processo di analisi (gestione dei sequenziatori, analisi dei dati grezzi, filtraggi delle reads, allineamento, visualizzazione, chiamata delle varianti). Questa complessità di livelli di analisi informatica, ciascuno con varie opzioni fruibili, complica notevolmente l'impegno degli operatori, in particolar modo nei casi in cui sia necessario confrontare risultati ottenuti attraverso differenti piattaforme, poiché ciascuna tecnologia sfrutta programmi e parametri di analisi differenti. Inoltre, l'interpretazione dei dati e la gestione di un sistema informatico così complesso prevede competenze specialistiche e richiede la presenza di figure professionali che spesso non sono presenti nei laboratori clinici.

PANNELLO NGS NELLA GESTIONE DELLA SORDITÀ NEUROSENSORIALE

Nell'epoca pre-NGS, l'utilizzo del sequenziamento con il metodo Sanger per l'identificazione della mutazione patogena comportava notevole lavoro, tempo ed elevato costo.

La grande eterogeneità delle forme di sordità neurosensoriali ereditaria era oltremodo complessa e la loro specifica diagnosi è stata una sfida alla diagnostica molecolare [33].

Tuttora il primo approccio alla ricerca di mutazioni genetiche nella sordità, che possiamo chiamare “screening di routine”, viene in genere avviato con la ricerca delle mutazioni GJB2 e GJB6 perché il 30 – 40% delle NSHL con ascendenza europea presentano una mutazione in questo gene [5], a meno che sintomi clinici caratteristici suggeriscano la ricerca di specifici geni (il gozzo indirizza la ricerca del gene SLC26A4 o la neuropatia uditiva il gene OTOF). Seguendo solo questo percorso la stragrande maggioranza delle mutazioni che non sono GJB2 o GJB6 rimangono senza diagnosi genetiche.

Lo sviluppo e l’ottimizzazione dei pannelli NGS specifici per la sordità ampliano lo spettro di geni rilevanti per la sordità permettendo di ottenere i migliori risultati ed evidenziando rare mutazioni responsabili della patologia un tempo di difficile determinazione.

Riportiamo le linee guida dell’American College of Medical Genetics and Genomics (2014) [34] per la valutazione della sordità: *“Tutti i neonati e i bambini con perdita di udito andrebbero sottoposti ad una valutazione globale in cui venga effettuata un’anamnesi completa di tipo familiare ed un esame clinico focalizzato alla ricerca di eventuali tratti dismorfici; negli adulti la valutazione deve ricercare l’età di insorgenza e precisare altre caratteristiche della perdita uditiva. Alla nascita l’anamnesi può essere utile per differenziare la sordità da causa ereditaria od acquisita; in ambito prenatale bisogna indagare le infezioni materne come CMV, rosolia, sifilide, uso di farmaci o droghe; in ambito neonatale occorre conoscere se vi è stato un parto prematuro, un basso peso alla nascita, un’ipossia neonatale, iperbilirubinemia, sepsi o esposizione a farmaci ototossici.*

In ambito postnatale occorre indagare eventuali infezioni virali, batteriche, meningite, trauma cranico, esposizione a rumore, o farmaci ototossici.

In base all’esame audiometrico devono essere definite l’entità e le caratteristiche della perdita uditiva, se progressiva o meno, fluttuante, se simmetrica o meno; inoltre va indagata la presenza o meno di una disfunzione vestibolare.

Per quanto concerne l’anamnesi familiare vanno identificati i parenti di I o II grado con sordità oppure che abbiano caratteristiche associate a sordità come retinite pigmentosa, anomalie renali, morte improvvisa, anomalie degli archi branchiali. Va presa in considerazione l’etnia e il paese di origine, specie se il paziente proviene da aree geograficamente isolate; altro parametro da considerare è la presenza o meno di consanguineità.

All’esame obiettivo occorre rilevare la presenza di note dismorfiche come eventuali asimmetrie, aspetto insolito del viso, anomalie pigmentarie della cute, eventuali anomalie neurologiche, disturbi dell’equilibrio, anomalie scheletriche o altri reperti patologici.

Se viene sospettata una perdita uditiva sindromica deve essere fornita la consulenza genetica e dopo il consenso informato devono essere effettuati i test genetici specifici (ricerca di un singolo gene, pannelli di sequenziamento per la sordità, WES, WGS, analisi cromosomica, analisi di microarray). I test per le mutazioni del DNA mitocondriale sono indicati quando sono stati utilizzati farmaci aminoglicosidici.

In assenza di indicazioni cliniche specifiche o nei casi di apparente sordità ereditaria autosomica recessiva il primo passo è l'effettuazione del test per la ricerca della mutazione GJB2 o GJB6.

Se tale test risulta negativo, bisogna applicare la tecnica NGS con pannelli di sequenziamento più ampi mirati per la sordità, oppure eseguire WES e WGS.

Il clinico audiologo e il genetista devono essere consapevoli dei geni inclusi nel pannello prescelto ed essere informati sulle caratteristiche di prestazione della piattaforma scelta, compresa la copertura, la sensibilità analitica e quali tipi di mutazioni potranno essere rilevate.

Se il test genetico rileva una mutazione di un gene correlato alla sordità, deve essere fornita la consulenza specifica per la mutazione.

Se il test genetico non riesce ad identificare mutazioni occorre considerare che la perdita uditiva sia di tipo acquisito, in questo caso bisogna procedere con indagini radiologiche mirate all'orecchio interno che sono necessarie anche per un eventuale impianto cocleare.

Pari importanza deve essere data all'esecuzione del test CMV nei bambini con perdita d'udito congenita.

Nel caso in cui non venga identificata una causa genetica, sono raccomandati incontri di follow-up con il genetista ogni 3 anni poiché alcune forme sindromiche di sordità possono non essere evidenti alla nascita o nella prima infanzia, inoltre le visite di controllo permettono di informare i pazienti sui nuovi test genetici divenuti disponibili o di aggiornare l'interpretazione dei risultati dei test precedenti in base alle conoscenze mediche aggiornate.

Indipendentemente dal fatto che i risultati dei test genetici siano positivi, negativi o inconclusivi i risultati devono essere comunicati attraverso il processo di consulenza genetica e non dal solo audiologo” vedi figura 6.

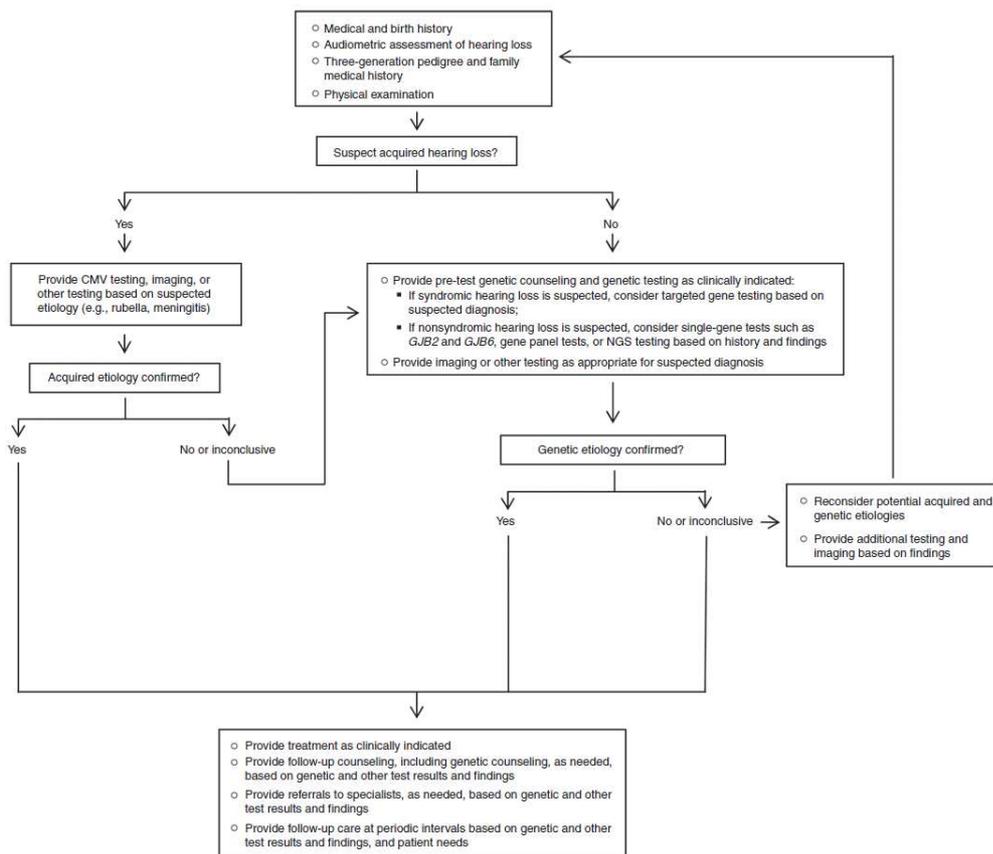


Figura 6-Flow-chart relativa alla valutazione clinica della diagnosi eziologica della sordità. *CMV*, citomegalovirus; *NGS*, sequenziamento next generation [34].

Con l'introduzione della tecnica NGS si è modificato l'iter diagnostico aggiungendo in caso di negatività all'analisi di primo livello un'analisi mediante NGS, e rinviando gli approfondimenti clinico diagnostici (ECG, Eco reni, Es. urine, RMN/TAC rocche petrose, Elettroretinogramma) al termine della definizione genetico-molecolare, se il risultato risulta ancora negativo.

Si è concordato di modificare il percorso diagnostico del paziente affetto da sordità apparentemente isolata a seguito della disponibilità in ambito genetico di tecniche di nuova generazione in quanto si è ritenuto che il programma di approfondimento clinico precedentemente utilizzato risulta più costoso, necessita di tempi di esecuzione lunghi (l'Elettroretinogramma ha una attendibilità dopo il 6 anno di vita), sottopone il paziente ad esami invasivi (RMN/TAC in anestesia generale nel bambino) e non sempre porta all'identificazione di segni clinici indicativi per una specifica analisi genetica.

Si segnala che la gestione del paziente secondo questi nuovi percorsi trova corrispondenza con le indicazioni internazionali.

La tecnica NGS in Regione Lombardia ha una codifica nel nomenclatore tariffario (cod. 29.29.7 con rimborso alla struttura sanitaria di circa 2000 euro ad oggi), per cui il paziente esegue le analisi tramite il SSN pagando, in assenza di esenzioni (ID23, E11) un ticket pari a 66 euro.

In breve tempo i costi si sono già ridotti e la progressiva evoluzione della tecnologia di analisi potrà ad ulteriori riduzioni economiche in un prossimo futuro. Inoltre, la strategia di centralizzazione dei Laboratori d'analisi che sta adottando Regione Lombardia porterà verosimilmente al contenimento della spesa sanitaria e quindi anche ai rimborsi delle metodiche NGS.

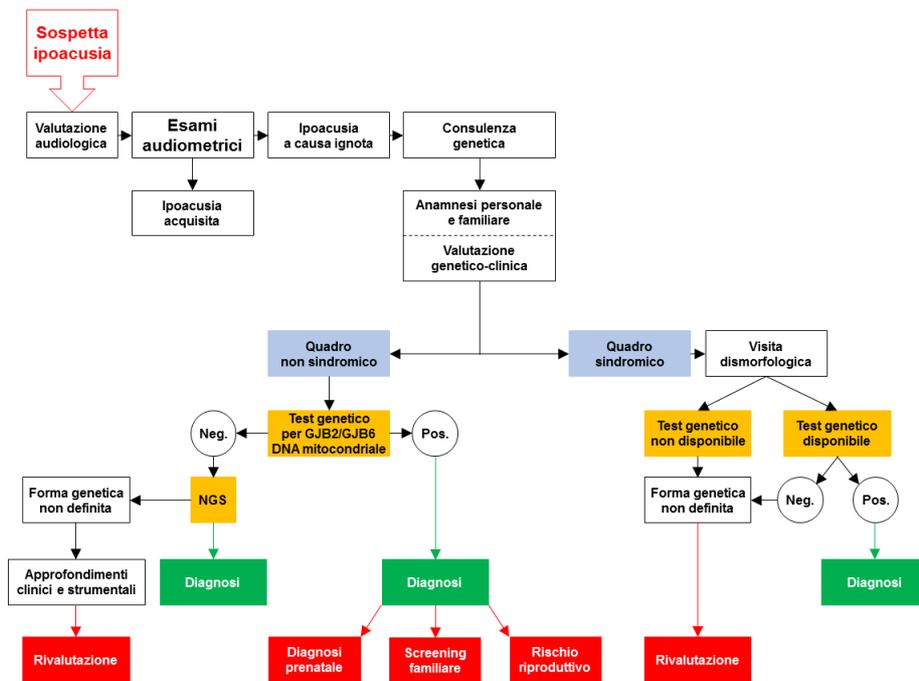


Figura 7-Flow-chart relativa alla valutazione clinica della diagnosi eziologica della sordità utilizzata dall'U.O.C. di Audiologia di Milano dal 2000.

Per sottolineare lo straordinario apporto dell'NGS nell'inquadramento della sordità non sindromica riportiamo i dati riportati da Vona e coll. pubblicata nel 2014 [33]: in una casistica di 23 soggetti sordi sono stati applicati due pannelli rispettivamente di 80 e 129 geni, riportando un tasso di successo nella determinazione della causa genetica della sordità pari al 57% (13 soggetti), individuando 8 forme dominanti e 5 recessive.

La tecnologia NGS apre le porte all'amplificazione massiva e completa del genoma; per un altissimo numero di patologie non è comunque necessario effettuare un dispendioso screening completo del genoma ma si può raggiungere un risultato esaustivo indagando limitate regioni genomiche rendendo, tra le altre cose, questo tipo di analisi economicamente sostenibile. La maggior parte delle piattaforme NGS a ora sviluppate sono state costruite specificatamente per effettuare analisi "targettate", ovvero non per effettuare la piena visualizzazione del genoma ma per indagare un limitato numero di set genici patologia specifici. Le metodiche applicative esistenti consistono di diverse tipologie di piattaforme analitiche ognuna caratterizzata da specifici vantaggi

e limiti tecnologici da tenere in considerazione nella stesura dei cosiddetti pannelli “targettati” - ovvero set analitici comprendenti specifici geni da analizzare – in ambito diagnostico e/o di ricerca. Per quanto riguarda nello specifico le sordità neurosensoriali la scelta dei geni da inserire in questi pannelli analitici, che garantiscono una sostenibile applicazione delle metodiche NGS, rappresenta sicuramente una sfida ambiziosa, in quanto nella letteratura vengono continuamente rilevate alterazioni e mutazioni causative in un numero sempre maggiore di geni, a trasmissione recessiva e/o dominante. Il lavoro di selezione dei possibili geni candidati all'analisi deve essere concertato tra: l'aspetto “tecnico” specifico della piattaforma analitica in dotazione, la fine revisione della letteratura esistente e l'esperienza applicativa di pannelli sviluppati e già operativi presso altri enti. Nel nostro specifico caso, tra gli aspetti sopra citati, si è voluto maggiormente concentrare l'attenzione sul primo, ovvero quello tecnico, col fine di poter ottenere una quanto più efficiente copertura delle regioni geniche indagate; per copertura del pannello si intende la possibilità tecnica della piattaforma di visualizzare in modo completo le regioni interessate dall'analisi garantendo quindi la piena rilevazione di varianti e mutazioni nucleotidiche associate.

Il Molecular Otolaryngology and Renal Reserche Laboratories (MORL) University of Iowa ha sviluppato la piattaforma OtoSCOPE® comprendente 109 geni causa di sordità e quelli che causano le sindromi più frequenti associate alla sordità (Usher, Pendred, Alström, Bor, Jervell e Lange Nielsen, Stickler, Perrault) [35]

OtoSCOPE® presenta una sensibilità e specificità del 99%, il suo costo è di circa 1500 dollari, questo approccio globale permette una riduzione dei costi d'analisi e dei tempi di definizione della diagnosi. Non è indicato nella diagnosi delle sordità monolaterali.

La Società Fulgen Diagnostic (4978 Santa Anita Ave Ste 205 Temple City, CA, 91780) fornisce al mercato ben 170 pannelli preimpostati per svariate patologie con la possibilità di effettuare personalizzazioni.

Per la sordità è attualmente in commercio un pannello che studia 103 mutazioni *vedi tabella 1*.

ACTB, ACTG1, ADGRV1, ATP2B2, ATP6V1B1, BCS1L, BSND, CATSPER2, CCDC50, CDH23, CLDN14, CLRN1, COCH, COL11A2, COL9A3, CRYM, DFNA5, DFNB31, DFNB59, DIAPH1, DSPP, EDN3, EDNRB, ERCC2, ERCC3, ESPN, ESRRB, EYA1, EYA4, FGF3, FOXI1, GATA3, GIPC3, GJA1, GJB2, GJB3, GJB6, GPSM2, GRHL2, GRXCR1, GSTP1, HGF, ILDR1, JAG1, KCNE1, KCNJ10, KCNQ1, KCNQ4, LHFPL5, LHX3, LOXHD1, LRTOMT, MARVELD2, MIR96, MITF, MSRB3, MTAP, MYH14, MYH9, MYO15A, MYO1A, MYO1C, MYO1F, MYO3A, MYO6, MYO7A, OTOA, OTOF, PAX3, PCDH15, PDZD7, PMP22, POU3F4, POU4F3, PRPS1, PTPRQ, RDX, SERPINB6, SIX1, SLC17A8, SLC26A4, SLC26A5, SLC4A11, SMPX,

SNAI2, SOX10, SPINK5, TBL1X, TCF21, TECTA, TIMM8A, TJP2, TMC1, TMIE, TMPRSS3, TMPRSS5, TPRN, TRIOBP, TRMU, USH1C, USH1G, USH2A, WFS1

Tabella 1-Elenco dei 106 geni correlati alla sordità testati dal pannello della Fulgen Diagnostics.

CONCLUSIONI

L'NGS sta trovando numerose applicazioni in medicina, biologia, alimentazione, criminologia.

Nel campo della genetica medica è diventato, ormai, uno strumento indispensabile, consentendo di studiare contemporaneamente molti geni e permettendo una visione d'insieme delle malattie genetiche.

Il vantaggio dell'NGS rispetto alle tecniche basate sul tradizionale metodo Sanger è che si possono analizzare quantità di DNA, in termini di basi sequenziate, di gran lunga superiori con una drastica riduzione dei tempi e dei costi: ne risulta una grande quantità di informazioni da analizzare con un notevole aumento di probabilità di trovare la mutazione in uno dei geni sottoposti al sequenziamento.

Per le patologie ad elevata eterogeneità genetica come per la sordità, dove più di 100 geni possono causare la stessa malattia, i sistemi di lettura di nuova generazione rappresentano uno strumento diagnostico ad alta processività che consente diagnosi molecolari altrimenti non percorribili.

Questa tipologia di lettura, infatti, applicata a casi clinici per i quali non si sia trovata una causa genetica, peraltro fortemente sospettata, permette di analizzare contemporaneamente nello stesso individuo estesi pannelli di geni.

I primi risultati sono promettenti. Anche se l'analisi e l'interpretazione dei dati è piuttosto complessa, questo approccio porta alla soluzione di circa il 30% dei casi senza una specifica diagnosi.

L'analisi e la refertazione dei dati ottenuti mediante NGS può portare oltre all'identificazione di mutazioni note e descritte anche a difficoltà interpretative: più mutazioni presenti in geni diversi, modalità di trasmissione complessa (digenica, triallelica), geni modificatori.

In questi casi si possono rendere necessari studi di segregazione familiare, approfondimenti molecolari a diversi livelli (DNA, RNA, Proteine), vedi esempi di refertazione NGS in differenti sordità (Referti A-B-C-D).

Paragonando le varie piattaforme NGS per la sordità quella con microgocce basate sulla PCR sembra dimostrare standards superiori e sembra essere la tecnica più promettente in termini di combinazione di alta sensibilità, velocità ed efficienza dei costi ed utilizzabilità per la diagnosi clinica, anche se l'arricchimento basato sull'ibridazione ha il vantaggio di ridurre i costi e il lavoro.

L'applicazione dell'NGS nei laboratori clinici potrà migliorare molto la precisione della diagnosi di sordità permettendo di effettuare interventi precoci e mirati per il paziente affetto da perdita uditiva. La tecnica NGS sta aprendo nuove prospettive nella diagnostica molecolare delle malattie rare, ma per la sua complessità richiede sia una notevole **esperienza professionale** sia l'impiego di **tecnologie strumentali avanzate**. Di conseguenza oltre alla necessità di Laboratori altamente specializzati e personale altrettanto qualificato si sta verificando un progressivo cambiamento anche nell'atteggiamento del genetista clinico che ha il compito di trasferire a livello clinico l'enorme quantità di informazioni provenienti da tali analisi. Un compito complesso che vede la Genetica Clinica come il "nucleo" di questa nuova realtà e il genetista clinico come una figura che necessita di una evoluzione verso un "*clinico di nuova generazione*".

L'attuale inquadramento genetico-audiologico del paziente con sordità non può prescindere pertanto dall'applicazione di tecniche NGS secondo un programma di analisi impostato a diversi livelli.

Bibliografia

1. Smith RJ, Bale Jr JF, White KR. Sensorineural hearing loss in children. Lancet 2005; 365: 879-90.
2. Martini A, Castiglione A. Le ipoacusie su base genetica. In G. Paludetti: Ipoacusia infantile, dalla diagnosi alla terapia. Ed. Omega Torino 2011, 69-93.
3. Bitner-Glindzicz M. Hereditary deafness and phenotyping in humans. Br Med Bull 2002; 63: 73-94.
4. Petersen MB. Non-syndromic autosomal-dominant deafness. Clin Genet 2002; 62: 1-13.
5. Morton CC, Nance WE. Newborn hearing screening: a silent revolution. N Engl J Med 2006; 354: 2151-64.
6. Vona B, Nanda I, Hofrichter MAH, Shehata-Dieler W. Non-syndromic hearing loss gene identification: A brief history and glimpse into the future. Mol Cell Probes 2015; pii:S08990-8508(15)00035-3. Doi.10.1016/j.mcp.2015.3.008.

7. Sanger F, Coulson AR. A rapid method for determining sequences in DNA by primed synthesis with DNA polymerase. *J Mol Biol* 1975; 94 (3): 441-8. . Doi:10.1016/0022-2836(75)90213-2.
8. Sanger F, Nicklen S, Coulson AR. DNA sequencing with chain-terminating inhibitors. *Proc Natl Acad Sci USA* 1977; 74 (12): 5463-7. Doi:10.1073/pnas.74.12.5463. PMC431765.
9. Smith LM, Sanders JZ, Kaiser RJ. et al. Fluorescence detection in automated DNA sequence analysis. *Nature* 1986; 321 674-679. Doi:10.1038/321674a0.
10. Mardis E R. The impact of next-generation sequencing technology on genetics. *Trends Genet.* 2008; 24,133-144. Doi: 10.1016/j.tig2007.12.007.
11. Shendure JA, Ji H. Next-generation DNA sequencing. *Nat Biotechnol* 2008; 26: 1135-1145. Doi:10.1038/nbt1486
12. Mamanova L, Coffey AJ, Scott CE, et al. Target-enrichment strategies for next-generation sequencing. *Nat Methods* 2010; 7: 111-118. Doi:10.1038/nmeth.1419.
13. Friedman RA, Bykhovskaya Y, Tu G et al. Molecular analysis of the POU3F4 gene in patients with clinical and radiographic evidence of X-linked mixed deafness with perilymphatic gusher. *Ann Otol Rhinol Laryngol.* 1997; 106(4): 320-5.
14. Wang QJ, Li QZ, Rao SQ. A novel mutation of POU3F4 causes congenital profound sensorineural hearing loss in a large Chinese family. *Laryngoscope.* 2006 ;116(6): 944-50.
15. De Kok YJ, Van der Maarel SM, Bitner-Glindzicz M, et al. Association between X-linked mixed deafness and mutations in the POU domain gene POU3F4. *Science* 1995; 267: 685-8.
16. Leon PE, Raventos H, Lynch E, et al. The gene for an inherited form of deafness maps to chromosome 5q31. *Proc Natl Acad Sci U S A* 1992; 89: 5181-4.
17. Leon PE, Raventos H, Lynch E, et al. The gene for an inherited form of deafness maps to chromosome 5q31. *Proc Natl Acad Sci U S A* 1992; 89: 5181-4.
18. Avraham KB, Kanaan M. Genomic advances for gene discovery in hereditary hearing loss. *J Basic Clin Physiol Pharmacol* 2012; 23(3): 93-7. Doi: 10.1515/jbcpp-2012-0033.
19. Mahdiah N, Rabbani B. Statistical study of 35delG mutation of GJB2 gene: a meta-analysis of carrier frequency. *Int J Audiol* 2009; 48(6): 363-370. Doi: 10.1080/14992020802607449.
20. Imtiaz F, Taibah K, Ramzan K, et al. A comprehensive introduction to the genetic basis of non-syndromic hearing loss in the Saudi Arabian population. *BMC Med Genet* 2011; 12:91. Doi: 10.1186/1471-2350-12-91.
21. Nahili H, Ridal M, Boulouiz R et al. Absence of GJB3 and GJB6 mutations in moroccan familial and sporadic patients with autosomal recessive non-syndromic deafness. *Int J Pediatr Otorhinolaryngol* 2008; 72: 1633-6.

22. Trotta L, Iacona E, Primignani P, et al. GJB2 and MRTNR1 contributions in children with hearing impairment from Northern Cameroon. *Int J Audiol* 2011; 1(50): 133-138. Doi: 10.3109/14992027.2010.537377.
23. Hutchison CA. DNA sequencing: bench to bedside and beyond. *Nucleic Acid Res* 2007; 35: 6227. Doi: 10.1093/nar/gkm688.
24. Murrey JC, Buetow KH, Weber JL et al. A comprehensive human linkage map with centimorgan density. Cooperative Human Linkage Center (CHLC). *Science* 1994; 265: 2049-2054.
25. Dunham I, Shimizu N, Roe BA, et al. The DNA sequence of human chromosome 22. *Nature* 1999; 402: 489-495. Doi: 10.1038/990031.
26. Lander ES, Linton LM, Birren B et al. Initial sequencing and analysis of the human genome. *Nature* 2001; 409: 860-921. Doi: 10.1038/35057062.
27. Venter JC, Adams MD, Myers EW et al. The Sequence of the Human Genome. *Science* 2001; 291 n. 5507: 1304-1351. Doi: 10.1126/science.1058040.
28. Schultz MD, He Y, Whitaker JW et al. Human body epigenome maps reveal non canonical DNA methylation variation. *Nature* 2015; 523: 209-216. Doi: 10.1038/nature 14465.
29. Liu L, Li Y, Li S, et al. Comparison of next-generation sequencing systems. *J Biomed Biotechnol* 2012(2012): 251364. Doi: 10.1155/2012/251364.
30. Rehman AU, Morell RJ, Belyantseva IA, et al. Targeted capture and next-generation sequencing identifies C9orf75, encoding taperin, as the mutated gene in non syndromic deafness DFNB79. *Am J Hum Genet* 2010; 86: 378-388. Doi: 10.1016/j.ajhg.2010.01.039.
31. Walsh T, Shahin H, Elkan-Miller T, et al. Whole exome sequencing and homozygosity mapping identify mutation in the cell polarity protein GPSM2 as the cause of non syndromic hearing loss DFNB82. *Am J Hum Genet* 2010; 87: 90-94. Doi: 10.1016/j.ajhg.2010.05.010.
32. Hoefsloot LH, Feenstra I, Kunst HP, Kremer H. Genotype phenotype correlations for hearing impairment: approaches to management. *Clin Genet* 2014; 85(6): 514-523. Doi: 10.1111/cge.12339.
33. Shearer AE, Smith RJ. Genetics: advances in genetic testing for deafness. *Curr Opin Pediatr* 2012; 24: 679-686. Doi: 10.1097/MOP.0b013e3283588f5e.
34. Vona B, Müller T, Nanda I, et al. Targeted next-generation sequencing of deafness genes in hearing-impaired individuals uncovers informative mutations. *Genet Med* 2014; 16(12): 945-953. Doi:10.1038/gim.2014.65.
35. Mutai H, Suzuki N, Shimizu A, et al. Diverse spectrum of rare deafness genes underlies early-childhood hearing loss in Japanese patients: a cross-sectional, multi-center next-generation sequencing study. *Orphanet J Rare Dis* 2013; 8: 172. Doi: 10.1186/1750-1172-8-172.

36. Shearer AE, De Luca AP, Hildebrand MS, et al. Comprehensive genetic testing for hereditary hearing loss using massively parallel sequencing. Proc Natl Acad Sci USA 2010; 107: 21104-21109. Doi. 10.1073/pnas.1012989107.
37. Valencia CA, Pervaiz MA, Husami A, et al. Next generation sequencing technologies in medical genetics. Springer Briefs in Genetics. Doi: 10.1007/978-1-4614-9032-6.

TERMINOLOGIA GENETICA

Alleli: le differenti forme di un gene. **Y** e **y** sono la coppia di alleli che determinano il colore del seme. Gli alleli occupano loci corrispondenti, o posizioni, sui cromosomi omologhi.

Autosomico: un locus su qualsiasi cromosoma eccetto i cromosomi sessuali. Non sex-linked.

Allelico-dominanti: due diversi alleli in un locus sono responsabili per diversi fenotipi, ed entrambi influiscono sul fenotipo dell'eterozigote. Per esempio, si consideri la situazione dove vi sono tre alleli (allelia multipla) **A**, **B**, e **O** che determinano il gruppo sanguigno umano. Vi sono tre possibili genotipi **AA**, **BB**, **OO** che corrispondono ai fenotipi del sangue tipo **A**, **B**, e **O** rispettivamente; Due altri genotipi sono **AO** e **BO** che corrispondono al sangue tipo **A** e **B** rispettivamente, perché l'allele **O** é recessivo, L'ultimo genotipo é **AB**, che corrisponde al sangue di tipo **AB**. Entrambi gli alleli **A** e **B** contribuiscono al fenotipo dell'eterozigote. Perciò gli alleli **A** e **B** vengono chiamati co-dominanti.

Linkage completo (associazione): il linkage completo descrive il comportamento ereditario di 2 geni sullo stesso cromosoma quando si osserva una frequenza di crossing over fra i loci uguale a zero.

Dioico: organismo che produce solo un tipo di gamete; es. l'uomo.

Carattere dominante: un carattere espresso prevalentemente rispetto a l'altro.

Drosophila melanogaster: il moscerino della frutta, (organismo preferito nelle analisi genetiche).

Epistasi: un gene maschera l'espressione di un gene per un diverso carattere.

Generazione F1: discendenza di un incrocio fra razze pure, omozigote per il carattere in esame.

Generazione F2: discendenza di un incrocio che riguarda individui della F1.

Genotipo: l'assetto genetico di un organismo in riferimento al carattere. Per un singolo carattere autosomico, un individuo può essere omozigote per un carattere dominante, eterozigote, o omozigote per un carattere recessivo. I semi gialli sono dominanti, Ma le piante a semi gialli possono avere genotipo sia **YY** che **Yy**.

Emizigote: se vi è una sola copia (allele) di un gene per un particolare carattere, in un organismo diploide, l'individuo è emizigote per quel carattere e mostrerà il fenotipo recessivo. I geni X-linked nella drosophila e nell'uomo maschio sono in emizigosi.

Eterozigote: differenti alleli per un carattere in un individuo, come Yy.

Cromosomi omologhi: coppia di cromosomi in un individuo diploide che contengono la stessa sequenza di geni. Ciascun cromosoma omologo della coppia viene ereditato da uno dei genitori.

Omozigote: entrambi gli alleli per un carattere sono uguali nello stesso individuo. Può essere omozigote dominante (YY), o omozigote recessivo (yy).

Ibrido: eterozigote; normalmente si riferisce alla discendenza dell'incrocio fra due individui appartenenti a razze pure (omozigoti) che differiscono per i caratteri in esame.

Dominanza incompleta: Fenotipi intermedi in F1, i fenotipi parentali ricompaiono in F2. I fiori di della Bella di notte possono essere rossi, rosa o bianchi. Il colore è determinato da un singolo locus. Il genotipo RR presenta fiori rossi mentre rr presenta fiori bianchi. L'eterozigote Rr presenta fiori rosa. Quando un eterozigote presenta un fenotipo intermedio rispetto all'individuo omozigote dominante o omozigote recessivo, si parla di dominanza incompleta.

Alleli letali: geni mutati che sono capaci di causare la morte (normalmente quando in omozigosi).

Linkage: geni che vengono ereditati insieme sullo stesso cromosoma. Sono possibili tre modalità di eredità: non-linkage, Linkage parziale, e linkage completo.

Legge di Mendel dell'assortimento indipendente degli alleli: gli Alleli di differenti geni assortiscono indipendentemente gli uni dagli altri durante la formazione dei gameti.

Legge di Mendel della segregazione: gli Alleli segregano gli uni dagli altri durante la formazione dei gameti.

Monoecio: Organismo che produce sia gameti maschili che femminili.

Incrocio monoibrido: incrocio che coinvolge parenti che differiscono per un solo carattere.

Mutazione: variazione della sequenza del DNA di un gene in una nuova forma ereditabile. Generalmente, ma non sempre risulta in un allele recessivo.

Non-linkage: non-linkage descrive il modello di eredità per 2 geni sullo stesso cromosoma, quando la frequenza attesa per il crossing over fra i loci è almeno 1. Il modello di eredità osservato per i geni non-linked sullo stesso cromosoma è lo stesso che per due geni su 2 differenti cromosomi (indipendenza).

Linkage parziale: il linkage parziale è un modello di eredità per 2 geni sullo stesso cromosoma, quando la frequenza attesa per il crossing over fra i loci è maggiore di zero ma inferiore a uno. Per mezzo dell'analisi del linkage parziale si possono avere informazione sull'ordine lineare e la distanza dei geni sullo stesso cromosoma.

Fenotipo: è la manifestazione fisica di un organismo rispetto a un carattere, i.e. semi gialli (Y) o verdi (y) nelle piante di pisello. Il carattere dominante si rappresenta normalmente con la lettera maiuscole, mentre il carattere recessivo con la stessa lettera minuscola.

Pleiotropia: quando un solo gene influisce su più di un fenotipo in un organismo.

Carattere recessivo: il contrario di dominante. Un carattere che è normalmente mascherato.

Incrocio reciproco: quando si usano gameti maschili e femminili per due diversi caratteri, alternando la fonte dei gameti.

Cromosomi sessuali: cromosomi che determinano il sesso

Eredità sex-linked: geni presenti sui cromosomi sessuali, come i geni presenti sul cromosoma (X-linked) nei moscerini e nell'uomo.

Test cross o back cross: generalmente un incrocio che utilizza un individuo omozigote recessivo. Quando si studia un singolo carattere, un test cross è un incrocio fra un individuo con fenotipo dominante ma con genotipo ignoto (omozigote o eterozigote) con un individuo omozigote recessivo. Se l'individuo in esame è eterozigote, approssimativamente il 50% della discendenza mostrerà il fenotipo recessivo.

Razza pura: omozigote per il carattere in esame.

Allele selvatico: forma del gene non mutato che codifica per una funzione genetica normale. Generalmente, ma non sempre è un allele dominante.

Ricavata da:

<http://www.biology.arizona.edu>

All contents copyright © 1996. All rights reserved. Traduzione Prof. Erminio Terio © 1999