UNIVERSITA' DEGLI STUDI DI FOGGIA

DOTTORATO DI RICERCA IN:

“INNOVAZIONE E MANAGEMENT DEGLI ALIMENTI AD ELEVATO IMPATTO SALUTISTICO”

-CICLO XXIX-

POSSIBILE INTERAZIONE TRA I LIVELLI PLASMATICI DI OMOCISTEINA E ABITUDINI ALIMENTARI COME FATTORI DI RISCHIO PER LO SVILUPPO DI PATOLOGIE NELL’UOMO: IL RUOLO NUTRACEUTICO DEI SUPPLEMENTI VITAMINICI

Dottorando: GIOVANNI FAVUZZI

Tutor: Chiar.mo Prof. MAURIZIO MARGAGLIONE

Anno Accademico 2015/2016

INDICE

Riassunto 1
Introduzione 2
1. Il sequenziamento del DNA 2
1.1 Sequenziamento Sanger 3
1.2 Next Generation Sequencing - Sequenziamento di seconda generazione 5
1.3 Analisi dei dati 17
1.4 Vantaggi e limiti delle piattaforme NGS 21
1.4 Applicazioni 22
1.5 Sequenziamento del DNA di terza generazione 25
2. Le tecnologie NGS per la diagnosi clinica 28
3. Difetti Congeniti del Metabolismo 30
3.1 Metabolismo degli amminoacidi 32
3.2 Aciduria Metilmalonica con Omocistinuria 34
Scopo del progetto 45
Casi clinici 46
Materiali e metodi 48
Analisi dell'esoma mediante NGS 48
Strategia analitica per l'analisi molecolare (metodo Sanger) 52
Risultati e Discussione 66
Bibliografia 72

Riassunto

Le tecnologie di sequenziamento NGS (Next Generation Sequencing) hanno permesso il sequenziamento del DNA ad una velocità senza precedenti e a costi sempre più bassi determinando la produzione di sequenze con una processività parallela e massiva che continua a crescere in modo esponenziale. Questo fa sì che l'NGS estenda le proprie opportunità di investimento in ogni settore delle scienze della vita.

Anche per la genetica umana è diventata una tecnologia insostituibile. Infatti, essa permette di effettuare un'indagine genetica andando a sequenziare, per finalità di ricerca, l'intero genoma, oppure solo la porzione codificante (l'esoma) o un pannello molto ampio di geni senza dover necessariamente ipotizzare a priori un gene causativo della patologia.

In particolare in questo lavoro di tesi il sequenziamento dell'esoma ha permesso di formulare una diagnosi ben precisa di una determinata patologia (Acidemia Metilmalonica con Iperomocistineinemia) in un soggetto che non mostrava un fenotipo clinico del tutto ascrivibile alla patologia stessa e di individuare in altri tre soggetti alcune varianti genetiche potenzialmente causative dell'iperomocisteinemia plasmatica.

Lo studio è stato condotto utilizzando la piattaforma MiSeq dell'Illumina per identificare le potenziali varianti patogeniche e il metodo di sequenziamento tradizionale (Sanger) per confermare, successivamente, le mutazioni trovate.

Le NGS, infatti, offrono risultati con una validità tale da consentire una gestione clinica del paziente più mirata ed accurata. Infatti, dopo aver effettuato la diagnosi si è proceduto ad applicare una terapia con supplementi vitaminici del complesso della vitamina B al fine di correggere i valori legati all'iperomocisteinemia. Il presente lavoro di tesi ha quindi l'intento di mettere in evidenza l'importante ruolo nutraceutico dei alcuni particolari supplementi vitaminici.

Introduzione

1. Il sequenziamento del DNA

Il sequenziamento del DNA, il processo che permette di determinare l'esatta sequenza delle basi nucleotidiche all'interno di molecole di DNA, è un ambito di ricerca in piena evoluzione. Negli ultimi sessant'anni si è osservato un notevole incremento delle conoscenze riguardanti il genoma umano e il codice genetico, che spaziano dalla scoperta della struttura del DNA nel 1953 fino all'introduzione di nuove strumentazioni (i sequenziatori) capaci di determinare, anche per genomi molto complessi, l'ordine dei diversi nucleotidi che costituiscono l'acido nucleico.

A partire dagli anni settanta del secolo scorso sono state ideate diverse strategie per ottenere la sequenza nucleotidica del DNA: il metodo sviluppato da Allan Maxam e Walter Gilbert si basa su modificazioni chimiche del DNA e sul conseguente taglio in posizioni specifiche[(1)]. Questo metodo viene comunemente descritto come metodo chimico per differenziarlo dal metodo di Frederick Sanger che viene, invece, definito enzimatico per l'utilizzo dell'enzima DNA polimerasi responsabile della sintesi di un filamento di DNA[(2)].

Di questi due metodi di sequenziatori genomici di prima generazione il Maxam-Gilbert venne progressivamente accantonato a causa della complessità tecnica e dell'uso ripetuto di sostanze tossiche; mentre la tecnica proposta da Sanger ha rappresentato una svolta epocale nella biologia molecolare e ciò è stato dimostrato dal fatto che F. Sanger ricevette per questo motivo il premio Nobel per la Chimica nel 1980. Tuttavia, nonostante lo scopo iniziale fosse arrivare a sequenziare l'intero genoma umano, si sono riscontrate diverse limitazione legate al tempo necessario per sequenziare il DNA e l'alto costo di esecuzione che rendevano proibitivo l'uso di questi strumenti su sequenze troppo lunghe. Un esempio è ben rappresentato dall'esperienza relativa al sequenziamento dell'intero genoma umano. Nel 1990 coordinato dal National Institutes of Health degli Stati Uniti e dal U.S. Department of Energy (DOE) era stato avviato lo Human Genome Project, con l'intento di sequenziare il genoma umano e identificare e mappare i geni in esso contenuti. Il progetto, basato dal punto di vista tecnico sul sequenziamento di tipo Sanger, ha visto il suo compimento in circa tredici anni di lavoro, la maggior parte dei quali sono stati necessari per la parte sperimentale del progetto stesso, il sequenziamento. Oggi, un attuale sequenziatore NGS permette eseguire il sequenziamento di un

genoma umano in un tempo di circa 1-2 settimane con una richiesta di fondi chiaramente ridotta (1,5 milioni di dollari)[4].

Da questo confronto, si può quindi facilmente comprendere come le tecniche NGS abbiano dato origine ad un enorme impulso negli studi di biologia molecolare, aprendo nuovi fronti di ricerca e consentendo il raggiungimento di risultati fino a pochi anni fa insperati.

Nonostante ciò, ad oggi il sequenziamento di Sanger, per la sua elevatissima affidabilità, mantiene ancora la sua preminenza nella pratica clinica. Infatti è tutt'oggi utilizzato in alcuni ambiti come test di paternità e la biologia forense; inoltre, le potenziali varianti patogeniche identificate mediante NGS vengono confermate con il metodo Sanger.

1.1 Sequenziamento Sanger

Il protocollo prevede inizialmente l'amplificazione delle molecole di DNA da sequenziare. L'amplificazione è l'operazione con cui si ottengono molte copie identiche delle molecole originali La PCR è una tecnica con cui una breve regione di una molecola di DNA è copiata molte volte da una DNA polimerasi, un enzima presente nelle cellule. Ad una estremità della sequenza da amplificare viene ibridata una breve sequenza di oligonucleotidi, primer, che funge da innesco per DNA polimerasi, definendo il punto di inizio della regione da amplificare. La fase di adesione del primer alla molecola è detta primer annealing. Per iniziare un'amplificazione mediante PCR, si aggiunge DNA polimerasi allo stampo di DNA legato ai primer e si incuba la miscela in modo che l'enzima sintetizzi nuovi filamenti complementari. La miscela viene quindi scaldata per separare i nuovi filamenti sintetizzati (denaturazione termica) e raffreddata per permettere l'ibridazione di ulteriori primer nelle loro rispettive posizioni, fra cui posizioni sui filamenti appena sintetizzati. La DNA polimerasi esegue quindi un secondo ciclo di sintesi di DNA. L'intero ciclo di denaturazione-ibridazione-sintesi viene ripetuto, di solito 25-30 volte, portando alla sintesi finale di parecchi milioni di copie del frammento di DNA originale.

La successiva fase di sequenziamento inizia con la denaturazione delle molecole in input, che porta alla separazione dei due singoli filamenti di DNA. Il successivo passaggio è il primer annealing, che crea un punto di partenza per la sintesi del filamento complementare, catalizzata da una DNA polimerasi. La reazione di sintesi

del filamento è iniziata aggiungendo alla soluzione contenente i filamenti di DNA l'enzima DNA polimerasi e i quattro nucleotidi dNTP (dATP, dTTP, dCTP, dGTP). Alla miscela vengono inoltre aggiunti dei dideossinucleotidi (ddNTP) in quantità inferiore rispetto ai dNTP e marcati con un fluorescente. I ddNTP possono essere incorporati nel filamento sintetizzato dalla DNA polimerasi esattamente come un nucleotide normale, ma comportano la terminazione dell'elongazione del filamento generato. Ciò accade perché il ddNTP è privo del gruppo ossidrilico nella posizione 3' dello zucchero che lo compone (Figura 1) e non fornisce quindi il sito di legame per il dNTP successivo: ogni volta che un ddNTP viene incorporato dalla DNA polimerasi si ha terminazione della catena.

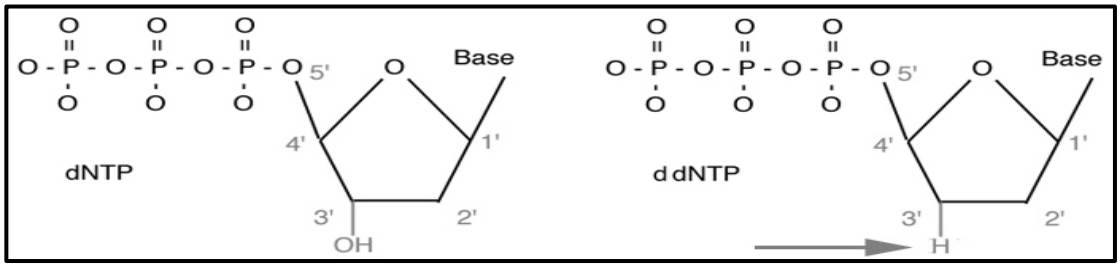

Figura 1: Struttura di un dNTP e di un ddNTP. La freccia indica la posizione in cui il gruppo –OH viene sostituito da –H.

Si ottengono così quattro famiglie distinte di nuovi filamenti: una che contiene filamenti che terminano tutti in dATP, una di filamenti che terminano in dTTP e così via. Ogni famiglia comprende filamenti di lunghezza diversa, secondo la posizione in cui è avvenuto l'evento di terminazione.

I ddNTP sono marcati per fluorescenza, cioè legati covalentemente a un fluorocromo il cui colore identifica univocamente lo specifico ddNTP. Le quattro famiglie di filamenti saranno perciò distinguibili fra loro, anche se unite in un'unica miscela. Il tipo di ddNTP che costituisce l'ultima base del filamento è quindi noto, rimane incognita la lunghezza del filamento, cioè la posizione del ddNTP. Questa informazione si ottiene applicando l'elettroforesi capillare, un'evoluzione della tecnica originaria che utilizzava invece l'elettroforesi su gel. Si fanno passare i filamenti di DNA all'interno di un capillare, ai capi del quale è applicata una differenza di potenziale; il DNA è carico negativamente e i frammenti migrano quindi verso il polo positivo del capillare. Questo è riempito di un gel che ostacola la migrazione dei frammenti: quelli di lunghezza inferiore saranno meno ostacolati nel

percorso e percorreranno il capillare in un tempo minore permettendo di separare i filamenti in base alla loro lunghezza. All'uscita del capillare, un laser eccita i marcatori fluorescenti provocando emissioni elettromagnetiche, captate da un rilevatore che ne riconosce la lunghezza d'onda ottenendo la "traccia" Sanger, poi elaborata dal software in picchi di colore diverso (elettroferogramma). Note le lunghezze dei frammenti e il ddNTP con cui terminano è immediato leggere la sequenza: se un frammento è lungo 20 nt e il marcatore è del colore che identifica una G, in posizione 20 della read ci sarà una G.

La tecnologia Sanger permette di ottenere sequenze di lunghezza tra i 600 e i 1000 nucleotidi sequenziando comunemente fino a 96 sequenze, con un throughput giornaliero di 6 Mb. Gli errori di sequenziamento sono imputabili principalmente all'amplificazione, a contaminazioni del campione e a errato riconoscimento dei ddNTP e della lunghezza delle sequenze, a causa di errori nella separazione elettroforetica. Mediamente, il tasso di errore medio (media valutata su tutte le basi della sequenza) è comunque piuttosto basso, compreso tra 10^{-5} e 10^{-6}.

1.2 Next Generation Sequencing - Sequenziamento di seconda generazione

Le piattaforme di NGS permettono la lettura multipla e parallela di migliaia di singoli frammenti di DNA di lunghezza variabile, consentendo l'analisi di un elevatissimo numero di basi in tempi ridotti.

La possibilità di sequenziare un gran numero di frammenti in parallelo permette di ottenere una grossa mole di dati, interpretabile solo mediante avanzati software bioinformatici. È quindi necessario lo sviluppo di algoritmi sempre nuovi per una migliore analisi ed interpretazione della grossa quantità di informazioni che si ottengono[(5)].

Le piattaforme di sequenziamento NGS comprendono il Roche/454 Life Science, il Genome Analyzer Illumina/Solexa, l'Applied BioSystem/SOLiD.

Accanto a queste, esistono le piattaforme di terza generazione: Ion Torrent, Pacific BioSciences, Oxford Nanopore.

Sebbene si differenzino per il tipo di biochimica alla base e per i metodi di acquisizione e di elaborazione dei dati, le piattaforme NGS sono accomunate da un work-flow operativo simile che prevede tre fasi principali: preparazione e

amplificazione del campione di DNA; sequenziamento e imaging; analisi bioinformatica dei dati.

ROCHE/454 LIFE SCIENCES

Il primo sistema introdotto nel mercato nel 2005 è stato il sequenziatore Genome Sequencer commercializzato da 454 Life Sciences (ora proprietà di Roche Diagnostic)[(6)]. Un passaggio comune a tutte le piattaforme NGS è la preparazione del campione. Questo step consiste nella frammentazione chimica (enzimi), meccanica (nebulizzazione) o fisica (ultrasuoni) del DNA genomico in frammenti di piccole dimensioni che variano dalle 100 alle 500 paia di basi (bp) in relazione alla piattaforma utilizzata; ai frammenti denaturati vengono aggiunti alle estremità degli adattatori noti che serviranno per la fase successiva di amplificazione del frammento genomico. Gli adattatori sono delle sequenze di riconoscimento universali, complementari ai primer che verranno usati per il sequenziamento.

Mediante un sistema di arricchimento basato su avidina e biotina si possono selezionare solo i frammenti che contengono gli adattatori permettendo quindi la creazione di una libreria di frammenti.

Nella fase di amplificazione vengono utilizzate delle biglie dal diametro di 28µm sulla cui superficie sono presenti oligonucleotidi complementari ad uno degli adattatori presenti. I frammenti della libreria vengono mescolati con un eccesso di biglie in modo tale che ogni biglia si leghi ad un solo frammento di DNA a singolo filamento; pertanto vengono posti nella fase acquosa di un'emulsione acqua in olio per far avvenire l'amplificazione clonale tramite Emulsion PCR del frammento legato alla biglia. Al termine della reazione, sulla superficie della microsfera saranno presenti circa 10 milioni di copie del frammento[(8)].

A questo punto le biglie contenenti la libreria di DNA a singolo filamento vengono deposte in una piastra Pico Titer costituita da circa 1,6 milioni di pozzetti aventi un diametro tale, 44 µm, da poter contenere un'unica microsfera legante il DNA che viene miscelata con la DNA polimerasi e uno strato di sferette di dimensioni più piccole contenenti gli enzimi sulfurilasi, luciferasi, apirasi e luciferina necessari per la reazione di pirosequenziamento. Questa metodica consiste nella preparazione della libreria di DNA, nell'amplificazione clonale dei frammenti attraverso una particolare tecnica chiamata Emulsion PCR e infine si realizza la reazione di sequenziamento mediante pirosequenziamento (Figura 2). Questa tecnica si basa sulla rilevazione del

pirofosfato (PPi) rilasciato dopo l'incorporazione di un nucleotide, reazione catalizzata dalla DNA polimerasi. Il PPi viene, dunque, convertito in ATP per opera della ATP sulfurilasi in presenza di adenosina 5' fosfosolfato (APS) e, utilizzando la fonte di energia così prodotta, la luciferasi, presente nella reazione di sequenza, converte la luciferina in ossiluciferina con conseguente produzione di luce di intensità proporzionale alla quantità di ATP. La luce prodotta è rilevata da una fotocamera ad alta risoluzione (CCD camera) e visualizzata come picco in un pirogramma[7]. La Figura 2 riassume graficamente un singolo ciclo di pirosequenziamento.

Successivamente la piastra viene, dunque, posizionata nello strumento di fronte ad una CCD camera. Nello strumento è presente una camera di flusso che fa fluire i deossinucleotidi (dATP, dCTP, dGTP, dTTP) perpendicolarmente ai pozzetti della piastra uno alla volta nell'ordine TACG.

Ogni volta che avviene l'incorporazione di un nucleotide in un pozzetto si ha un'emissione luminosa che passa attraverso il fondo dei pozzetti, viene catturata dalla CCD camera e registrata all'interno del pirogramma. Ciascun flusso viene tradotto in un'immagine dal software che le elabora generando un pirogramma in cui l'intensità del segnale sarà proporzionale all'ATP prodotto e quindi al nucleotide inglobato. Questo significa che un picco di intensità doppia, ad esempio, rileva che nello stesso ciclo sono stati inglobati 2 dNTP; un picco di intensità tripla rileva 3 dNTP inglobati e così via.

I dNTPs vengono aggiunti uno alla volta e poiché il dATP è un substrato naturale della luciferasi al suo posto viene utilizzata la deossiadenina α-tio-trifosfato (dATPS) che viene utilizzata efficacemente dalla DNA polimerasi ma non viene riconosciuta dalla luciferasi.

Inizialmente questo sistema arrivava a reads di 100-150bp, producendo circa 200.000 reads, con un throughput per run di 20Mb. Nel 2008 il sequenziatore 454 GS FLX Titanium arrivava a read lunghe 700bp, con una precisione del 99.9% dopo il filtraggio, con un output di 0.7Gb per run in 24 ore. Nel 2009, la combinazione del metodo GS Junior con sistema 454 GS portò l'output a 14Gb per run. Ulteriori evoluzioni hanno portato al GS FLX+, in grado di sequenziare read lunghe fino a 1kb[9].

La grande velocità, unita alla lunghezza delle reads prodotte, sono i punti forti di questo sistema. Tuttavia il costo dei reagenti resta un problema da risolvere.

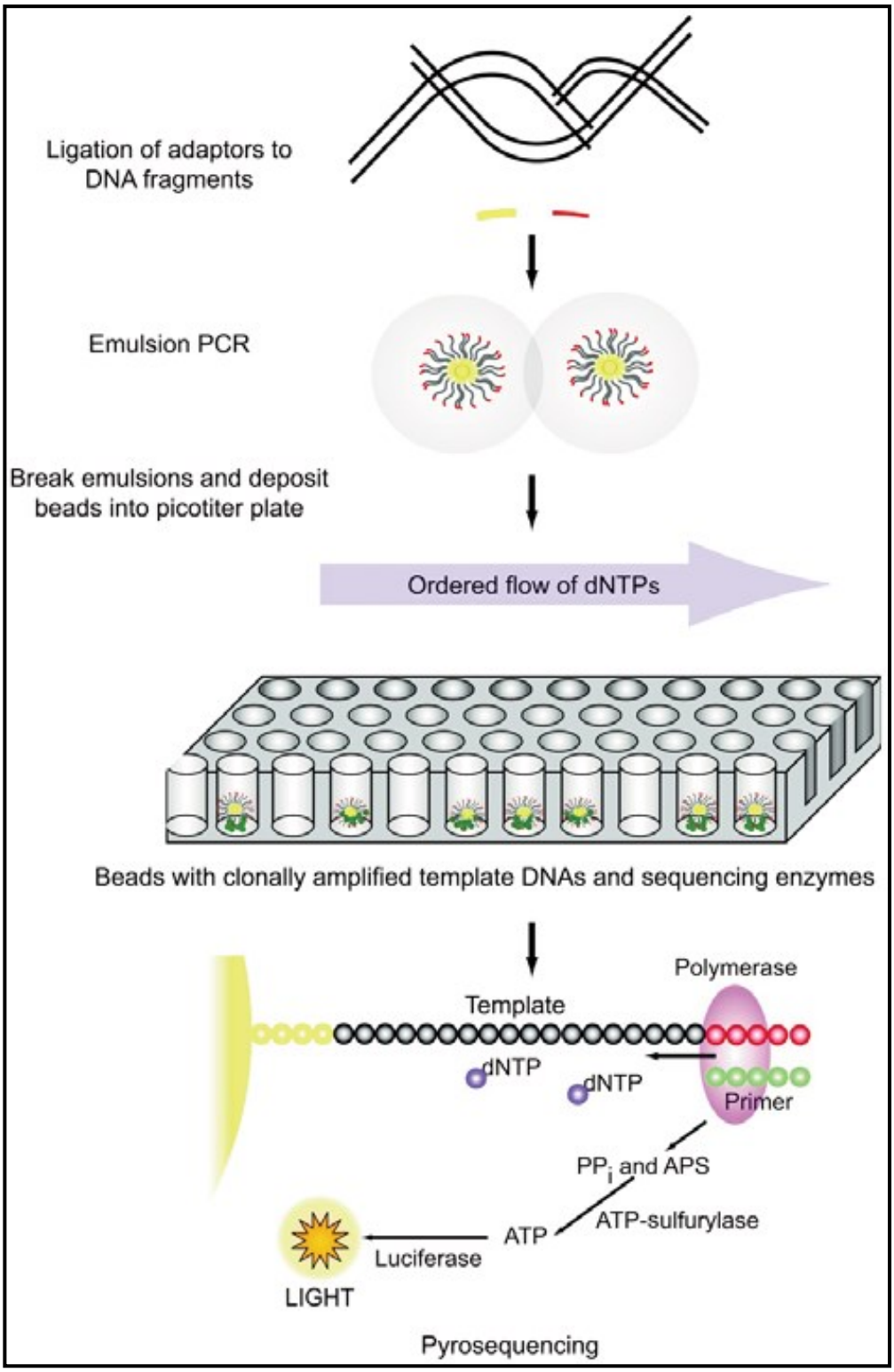

Figura 2: Sequenziamento mediante piattaforma Roche 454 GS FLX. Il DNA stampo è frammentato, legato a oligonucleotidi adattatori e amplificato clonalmente mediante PCR in emulsione. Dopo amplificazione, le biglie sono depositate in pozzetti di una "slide" a fibre ottiche (piastra "Pico Titer") insieme agli enzimi necessari per la reazione di sequenziamento. La piastra Pico Titer agisce da cella a flusso in cui sono effettuati cicli iterativi di pirosequenziamento. In ogni pozzetto, ad ogni incorporazione nucleotidica, si ha la produzione di pirofosfato (PPi) e la generazione di un segnale luminescente. APS, adenosina 5' fosfosolfato.

ILLUMINA/SOLEXA

Introdotta nel mercato nel 2006 da Solexa e in seguito rinominata Genome Analyzer da Illumina, questa piattaforma si basa sul sequenziamento a terminazione reversibile, mentre l'amplificazione clonale avviene attraverso la metodica detta "bridge PCR"[10].

La fase di preparazione dei campioni, finalizzata ad ottenere un filamento stampo della sequenza da sequenziare, prevede la denaturazione della doppia elica e l'aggiunta anche in questo caso, di adattatori, complementari ai primer che verranno usati per il sequenziamento (Figura 3).

I frammenti successivamente vengono immobilizzati su una superficie solida simile ad un vetrino, flow cell di Genome Analyzer, composta da otto comparti (lane) indipendenti, sulle cui superfici sono fissati degli oligonucleotidi di ancoraggio complementari agli adattatori.

L'ibridazione del filamento del DNA alla sequenza di ancoraggio determina la formazione di una struttura "a ponte" data dal ripiegamento dei filamenti[11]. Una volta immobilizzati, avviene l'amplificazione dei frammenti ad opera della DNA polimerasi che sintetizza il filamento complementare.

La molecola ottenuta viene denaturata e i due filamenti rimangono legati alla piastra a una sola estremità.

L'operazione si ripete ciclicamente: i due frammenti si "piegano" fino a ibridare la loro estremità rimasta libera con un complementare oligonucleotide della flow cell, riassumendo nuovamente la forma a ponte e permettendo la sintesi di un nuovo filamento complementare. I passaggi di immobilizzazione-sintesi-denaturazione continuano fino a ottenere un cluster di migliaia di frammenti, legati ad una estremità al substrato e raccolti in uno spazio molto limitato. Il processo è definito amplificazione bridge-PCR, proprio in virtù della forma "a ponte" che i filamenti assumono quando entrambi gli adattatori sono legati al substrato[12].

I cluster ottenuti sono composti sia di filamenti uguali al filamento originale sia di filamenti inversi; questi ultimi vengono rimossi, ottenendo cluster di circa 1000 copie identiche della stessa sequenza. Ai frammenti di ogni cluster viene eseguito l'annealing del primer di sequenziamento che permette l'avvio della reazione di sequenziamento vera e propria.

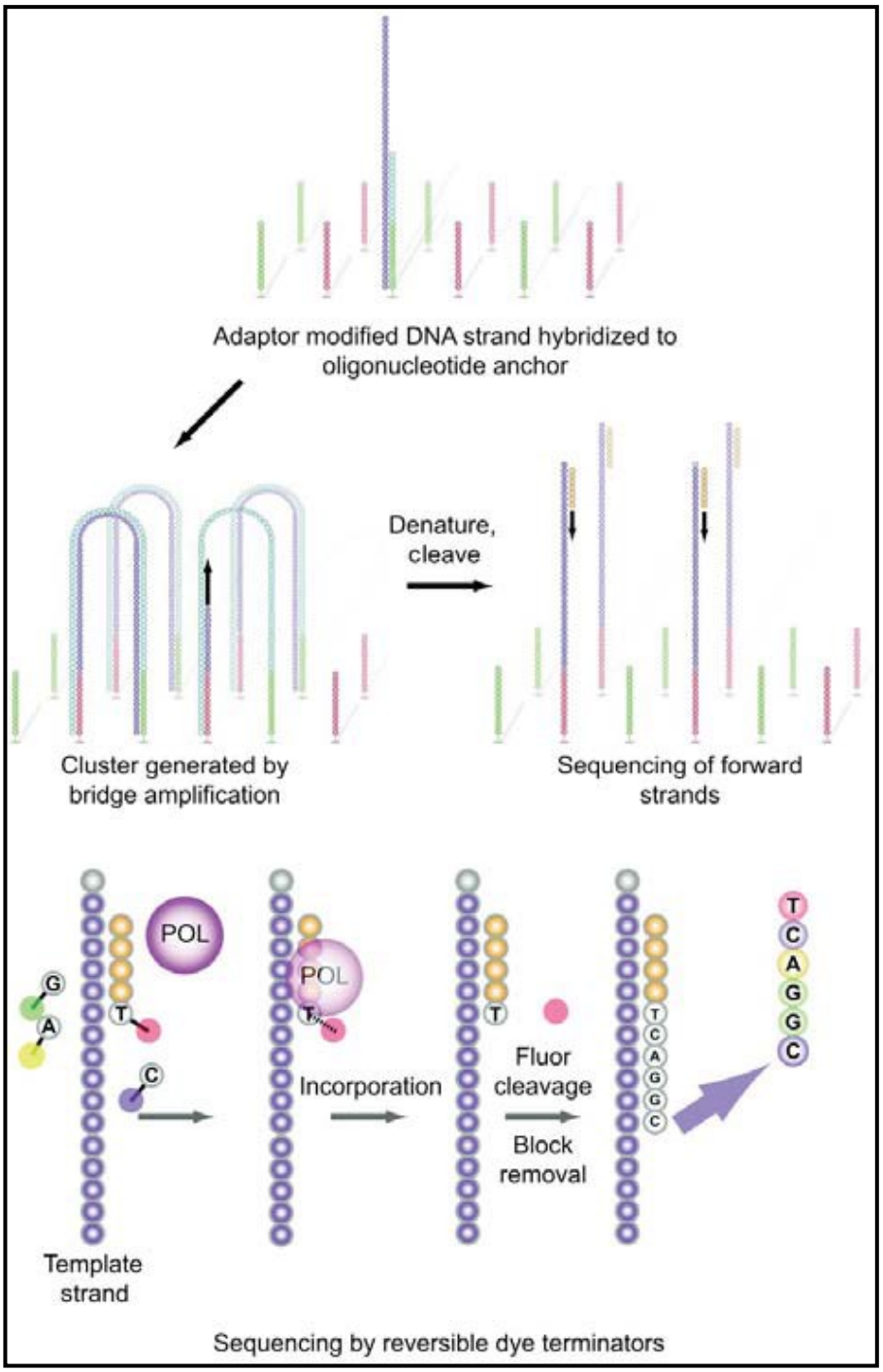

Figura 3: Sequenziamento mediante piattaforma Illumina Genome Analyzer. Il DNA a singolo filamento modificato con oligonucleotidi adattatori è aggiunto alla cella a flusso e immobilizzato mediante ibridazione. L'amplificazione "a ponte" genera "cluster" di frammenti amplificati clonalmente. I "cluster" sono denaturati e recuperati; il sequenziamento ha inizio con l'aggiunta di "primer", DNA polimerasi e 4 nucleotidi terminatori reversibili marcati con composti fluorescenti. Dopo l'incorporazione delle basi, la fluorescenza è rilevata e registrata. La fluorescenza e il blocco sono rimossi prima del successivo ciclo di sintesi.

Ogni ciclo di sequenziamento comprende l'incorporazione del nucleotide, l'acquisizione della fluorescenza e il taglio del nucleotide con il coinvolgimento di una DNA polimerasi e i quattro dNTP a cui sono state apportate due modificazioni. La prima consiste nell'incorporazione di un marcatore fluorescente (diverso per ogni dNTP) che ne permette l'identificazione; la seconda nell'aggiunta di un terminatore reversibile. Il terminatore è una molecola che blocca il gruppo ossidrile impedendo l'ulteriore sintesi, in modo da garantire l'incorporazione di una sola base.

Questa limitazione è necessaria poiché in ogni ciclo sono forniti tutti i dNTP e quindi la sintesi potrebbe riguardare più basi in contemporanea; il terminatore permette invece di arrestare la sintesi dopo ogni singola incorporazione. Il terminatore è detto reversibile in quanto può essere dissociato chimicamente, riattivando la sintesi.

In particolare, la DNA polimerasi aggiunge alla sequenza di innesco un nucleotide modificato complementare al templato e questa incorporazione provoca il termine della reazione di estensione. Il successivo step di lavaggio consente l'eliminazione di tutti i nucleotidi non incorporati.

Dopo ogni incorporazione, un laser eccita il fluorescente del dNTP generando un'emissione luminosa che viene rilevata e convertita dal sistema in immagini globali dell'intera superficie del vetrino di supporto permettendo l'identificazione del dNTP (Figura 4).

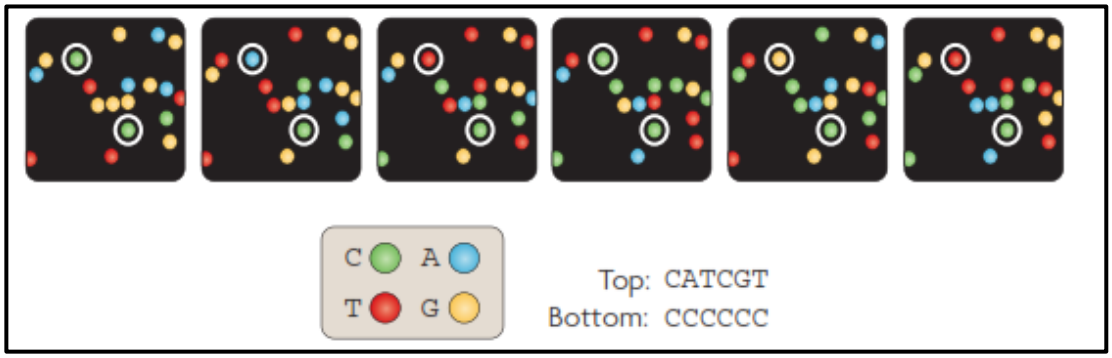

Figura 4: Incorporazione dei dNTP fluorescenti e lettura della sequenza.

In seguito, avviene il rilascio per via chimica del terminatore e del fluorocromo del nucleotide appena analizzato. Dopo l'eliminazione dei gruppi funzionali clivati vengono incorporati ulteriori nucleotidi e la polimerasi può legare il secondo nucleotide per continuare la reazione di estensione della sequenza.

Il sistema proposto da Illumina è ad oggi il più diffuso tra le tecnologie NGS sia per l'elevata accuratezza dei dati prodotti che per la disponibilità commerciale di sequenziatori dalle diverse caratteristiche, adattabili alle molteplici esigenze dei laboratori di ricerca. Per contro, questo sistema presenta costi, relativi all'acquisto iniziale, più elevati rispetto alle altre tecnologie disponibili.

Le prime versioni di Illumina Genome Analyzer erano caratterizzate da un tasso di errore non trascurabile e da reads di lunghezza molto inferiore rispetto a 454/Roche (35-50 nt) ma erano in grado di ottenere dati con un throughput di oltre 10000 Mb al giorno (oltre 10 volte più di 454/Roche) a costi molto contenuti. Le modifiche apportate alla tecnologia e ai protocolli sperimentali hanno sensibilmente ridotto gli errori di sequenziamento, seppure siano ancora presenti dei bias, e aumentato la lunghezza delle reads a 100 nt e oltre.

Nel 2010 viene rilasciato l'HiSeq 2000, che utilizza le stesse strategie del predecessore, arrivando a 600Gb per run, ottenibile in 8 giorni. Nel prossimo futuro dovrebbe poter arrivare addirittura ad 1Tb per run[13]. Attualmente è anche presente sul mercato il MiSeq che offre un throughput di dati minore, da 120 Mb a 1,5Gb.

Per le loro performances HiSeq e MiSeq sono principalmente usati nei progetti di sequenziamento genomico e nell'ambito clinico-diagnostico per la diagnosi molecolare di malattie genetiche.

Il costo per operazione è tra i più bassi, rispetto ai vari modelli di sequenziatori presenti sul mercato. L'unica pecca è rappresentata da un valore di accuratezza di circa il 98%; quindi la percentuale di errore è leggermente superiore rispetto a quella del Sanger che presenta un'accuratezza del 99.999%.

APPLIED BIOSYSTEM/SOLiD

Sviluppata sul metodo sequencing-by-ligation (a differenza dei due metodi precedentemente descritti che possono essere definiti sequencing-by-synthesis) proposto dal laboratorio George Church nel 2005[14], la piattaforma SOLiD viene introdotta nel mercato a partire dal 2007 da Applied Biosystems. Il protocollo SOLiD (Supported Oligonucleotide Ligation and Detection) si differenzia, infatti, nel processo di sequenziamento per l'utilizzo di una ligasi e non di una polimerasi. La DNA ligasi è un enzima in grado di legare covalemtemente due frammenti di DNA.

La preparazione dei campioni è simile quella effettuata per la tecnologia 454, in cui si realizza la denaturazione della doppia elica e i frammenti di DNA sono legati a

adattatori noti. I frammenti sono, poi, ibridati a delle sfere la cui superficie è ricoperta da oligonucleotidi complementari a uno degli adattatori. Le sfere sono messe in emulsione assieme ai reagenti e ai primer della PCR per ottenere l'amplificazione clonale dei filamenti mediante "Emulsion PCR" in modo del tutto simile a quanto visto per 454/Roche. Nella tecnologia SOLiD, però, le biglie sono molto più piccole (diametro 0,75 μm invece di 28 μm), permettendo una maggiore densità di biglie sequenziabili per run.

Le sferette recanti i prodotti di amplificazione clonale vengono immobilizzate sulla superficie di vetro funzionalizzata di una cella a flusso, sulla quale ha luogo la reazione di sequenziamento che viene innescata tramite l'ibridazione di un primer complementare all'adattatore a livello della giunzione adattatore-DNA stampo (Figura 5). Invece di fornire un gruppo ossidrile al 3' per l'estensione mediata dalla DNA polimerasi, il "primer" è orientato in modo da fornire un gruppo fosfato al 5' per consentire la ligazione alle sonde durante il primo passaggio di sequenziamento mediante ligazione ("ligation sequencing").

Ogni ciclo di sequenziamento coinvolge la sfera, un primer di sequenziamento, la ligasi e quattro sonde di dNTP. Ogni sonda è composta da otto basi più un marcatore fluorescente e presentano un sito di rottura fra i nucleotidi 5 e 6. L'ottamero è costituito (in direzione 3'→5') da 2 basi complementari a due nucleotidi da sequenziare, le basi da 3 a 5 sono degeneri, cioè capaci di appaiarsi a qualsiasi base della sequenza stampo, le basi 6, 7 e 8 sono anch'esse degeneri e possono essere rimosse, assieme al marcatore fluorescente, tramite la rottura del legame con la base in posizione 5. Le 2 basi specifiche corrispondono a una delle 16 possibili combinazioni di doppiette di basi (per es. TT, GT e così via).

Nel primo passaggio di "ligation sequencing" sono presenti una ligasi termostabile e le sonde rappresentanti le 16 possibili combinazioni (figura 6), che competono tra loro per l'ibridazione alla sequenza stampo immediatamente adiacente al "primer".

Dopo il legame delle sonde, viene effettuata una reazione di ligazione a cui segue una fase di lavaggio per eliminare le sonde non ibridate.

Una volta che una sonda si è appaiata alla sequenza stampo, avviene l'eccitazione del marcatore legato alla sonda e la conseguente emissione di fluorescenza. Contemporaneamente, viene rotto il legame tra le posizioni 5 e 6, lasciando libero l'estremità 5' della quinta base della sonda e permettendo il legame con una successiva sonda.

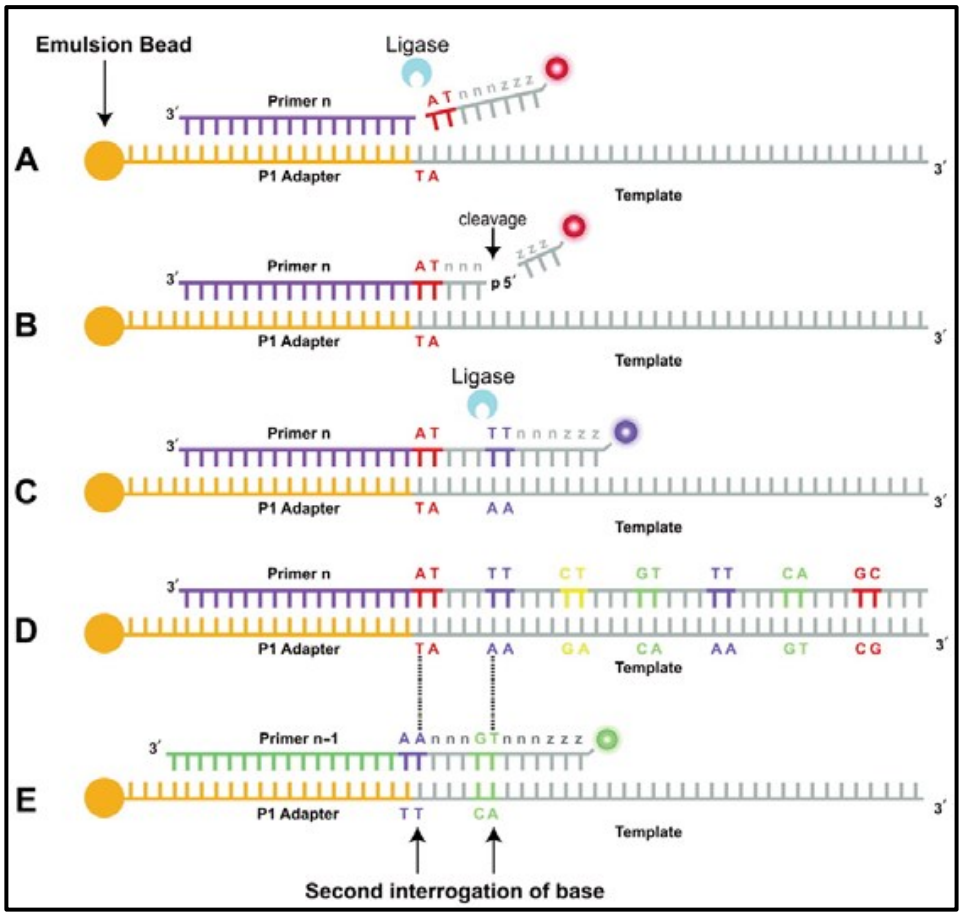

Figura 5: Sequenziamento mediante piattaforma Applied Biosystems SOLiD.
(A) Il DNA stampo legato all'oligonucleotide adattatore P1, con il primer (n) ibridato, è interrogato dalle sonde che rappresentano le 16 combinazioni possibili di doppiette di basi. In questo esempio, le 2 basi specifiche, complementari al DNA stampo, sono AT; (B) dopo l'ibridazione e la ligazione della sonda il segnale di fluorescenza viene rilevato, la porzione della sonda corrispondente alle ultime 3 basi degenerate viene tagliata e prima del secondo passaggio di sequenziamento, l'estremità 5' della sonda viene fosforilata; (C) ibridazione e ligazione della sonda successiva; (D) estensione completa del primer (n) durante il primo round costituito da 7 cicli di ligazione; (E) il prodotto ottenuto a seguito dell'estensione del primer (n) è denaturato dall'oligonucleotide adattatore legato al DNA stampo e viene effettuato un secondo round di sequenziamento con il primer (n-1). Grazie all'utilizzo progressivo di primer sfasati rispetto al primer iniziale (in questo esempio i primer sono sfasati di una base: n-1), vengono sequenziate le basi dell'oligonucleotide adattatore, che funge da sequenza di riferimento di partenza, che in combinazione con il sistema di codificazione a 2 basi (in cui ogni coppia di basi adiacenti è correlata ad un fluoroforo specifico) permette di decifrare algoritmicamente la sequenza stampo. Mediante tale piattaforma, ogni nucleotide del DNA stampo viene sequenziato due volte.

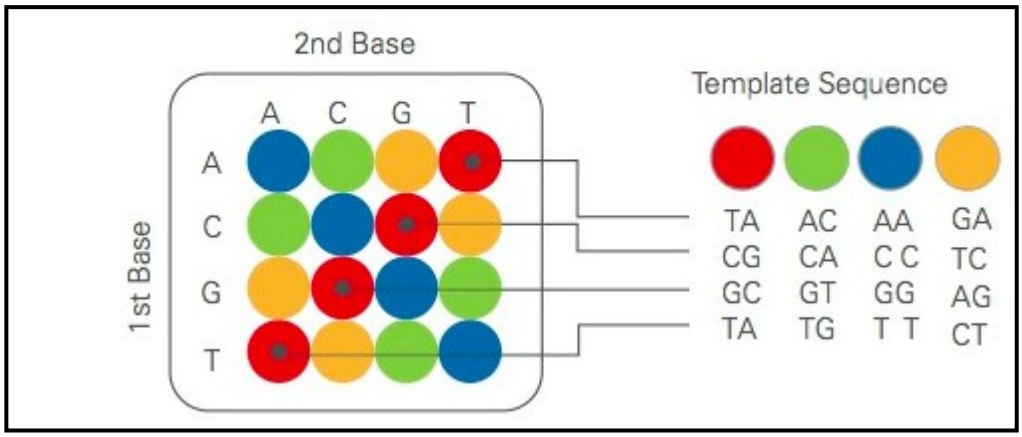

Figura 6: Le possibili combinazioni di nucleotidi che i quattro colori dei marcatori fluorescenti rappresentano. Ciascun colore non rappresenta univocamente una singola coppia di basi, ma 4 combinazioni sulle 16 possibili.

L'emissione luminosa è rilevata dalla strumentazione in modo da identificare il colore associato alla coppia delle prime due basi. Ogni marcatore produce una specifica emissione luminosa che identifica 4 di 16 possibili coppie di nucleotidi (4 possibili nucleotidi su due posizioni).

Nei cicli successivi di sequenziamento, le sonde vengono ligate al gruppo fosfato all'estremità 5' del pentamero precedente. Vengono effettuati 7 cicli di ligazione successivi, definiti come un "round", per estendere il primo primer. Successivamente, il filamento sintetizzato viene denaturato e viene ibridato un nuovo primer di sequenziamento, sfasato di una base (n-1) rispetto al primer usato nel passaggio precedente. In totale sono realizzati 5 round, utilizzando ogni volta nuovi primer sfasati di basi successive (n-2, n-3 e così via), e questo tipo di approccio permette che ogni nucleotide del DNA stampo venga sequenziato due volte permettendo di migliorare la bontà del sequenziamento.

Si veda ad esempio la Figura 7: la base nella posizione di lettura 5 viene coinvolta dal primer numero 2 nel secondo ciclo (indicato in azzurro) e dal primer numero 3 nel primo ciclo (indicato in blu).

Il colore dell'emissione da parte del fluorescente permette di identificare la coppia delle prime basi, ma non è sufficiente per distinguere un nucleotide dall'alto. Oltre alla specifica emissione per ogni coppia è infatti necessario che una delle basi della sequenzia sia nota. La base nota è incorporata nell'ultimo (il quinto) ciclo di sequenziamento e corrisponde all'ultimo nucleotide del primer. Quindi, dato che ciascun colore rappresenta 4 possibili coppie di dNTP (Figura 7) e per ognuna di esse

il secondo nucleotide coincide con la prima base della coppia successiva, conoscere una base della sequenza permette di interpretare tutta la sequenza nella sua interezza[15].

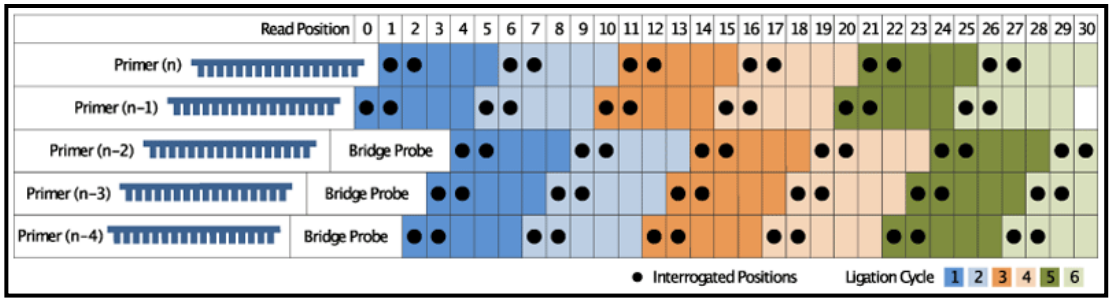

Figura 7: Design del sequenziamento con tecnologia SOLiD. Sono rappresentati i cinque cicli (uno per riga) di sequenziamento, ciascuno dei quali composto da sei passaggi di aggiunta di nuove sonde ad opera della ligasi. I punti neri indicano le coppie di basi interrogate in ogni passaggio.

La Figura 8 propone un esempio di decodifica dei colori rilevati, nota la prima base della sequenza.

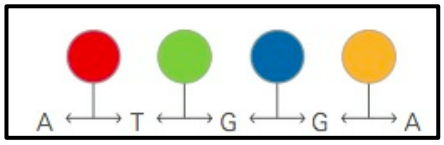

Figura 8: Esempio di decodifica dei colori per determinare la sequenza. Si suppongono note la prima base (A) e l'ordine dei colori rilevati. Dallo schema di codifica (Figura 7) è evidente che se la coppia è identificata dal colore rosso e la prima base è una A allora sicuramente la seconda base sarà un T. Nella seconda coppia, la prima base è la T appena identificata e il colore della coppia è verde: la seconda base sicuramente sarà una G.

Grazie al sequenziamento a due basi, l'accuratezza dell'applicazione arrivava al 99.85%. Nel 2010 viene rilasciato il SOLiD 5500x1, con reads length di 85bp, precisione del 99.99% e output di 30Gb per run. Un run completo è possibile eseguirlo in 7 giorni[5].

Le fonti possibili di errore del sequenziatore SOLiD sono diverse: errore di background nella fase di amplificazione, letture di bassa qualità a causa di presenze di sequenze diverse su una stessa sfera, oltre ad artefatti introdotti nel processo ciclico di ligasi e rimozione del marcatore. Il tasso medio di errore vari da 10^{-2} a 10^{-4}. Le prestazioni in termini di throughput (20 Gb/giorno) e di costo di sequenziamento sono superiori a quelle degli altri sequenziatori NGS.

Il problema principale di questo metodo è la lunghezza delle reads, infatti nonostante la tecnologia SOLID sequenzi oltre 300 milioni di sfere in parallelo, genera read molto corte di lunghezza pari a 50 nt, che per molte applicazioni non risulta essere sufficiente.

1.3 Analisi dei dati

Gli esperimenti con piattaforme NGS generano una quantità di informazioni senza precedenti e questo rappresenta una sfida per la gestione, l'archiviazione e, soprattutto, l'analisi dei dati[16]. Come primo tipo di informazione si ottiene un'enorme quantità di immagini relative ai segnali di luminescenza o di fluorescenza acquisiti dalla superficie della cella a flusso, registrate dopo ogni passaggio di sequenziamento iterativo. La dimensione delle informazioni generata durante una singola seduta analitica effettuata con le tecnologie 454 GS FLX, Illumina e SOLiD è rispettivamente di circa 15 GB, 1 TB e 15 TB[5].

Attraverso questa serie di immagini, le single biglie o i relativi "clusters" vengono identificati e localizzati. Mediante l'impiego di algoritmi specifici per ogni piattaforma vengono valutati una serie di parametri di immagine quali intensità, rumore di fondo e presenza di eventuali segnali aspecifici per generare sequenze nucleotidiche e assegnare a ciascuna base punteggi di qualità ("quality scores"), correlati alla probabilità di errore, processo definito "base calling". Gli algoritmi di base calling consentono di convertire l'informazione di sequenza in una informazione di immagine. Inoltre, vengono corretti artefatti come il "crosstalk" e il "phasing". Il crosstalk è causato dalle sovrapposizioni di emissioni di fluorescenza di nucleotidi diversi. Il phasing è dovuto alla dispersione e alla diffusione del segnale tra i cicli, che si verifica quando le molecole di un cluster non incorporano un nucleotide oppure quando la sintesi non si arresta dopo l'incorporamento. Questo fa sì che, se per esempio una guanina non è inserita nel ciclo 2, si rilevi il segnale aspecifico di quella base nei cicli 1 e 3.

I dati di output sono principalmente in formato testuale, il FASTQ che include, quindi, sia la sequenza (in genere nucleotidica) sia la qualità di ogni base. Sebbene la valutazione della qualità vari a seconda del tipo di piattaforma, i calcoli sono tutti effettuati in base allo storico "phred score", introdotto nel 1998 per i dati di sequenza di Sanger[17-18]. Il valore di qualità del phred score (q) utilizza una scala matematica per convertire a una scala logaritmica la probabilità stimata per l'identificazione non

corretta di una base (e): $q = -10 \cdot \log10(e)$. Le probabilità di identificare una base in modo non corretto pari a 0,1 (10%), 0,01 (1%) e 0,001 (0,1%) producono, rispettivamente, un valore di phred score di 10, 20 e 30.

Alcune piattaforme assegnano quality scores che sono specifici dello strumento ma che comunque possono essere convertiti nel quality score standard. In genere i software del sequenziamento considerano buoni quei dati con phred score pari o superiore a 30.

A questo punto, l'analisi bioinformatica si divide in tre passi:

1. Allineamento, che ricerca corrispondenze tra le reads e il genoma di riferimento,
2. Variant Calling, che tenta di separare differenze dovute ad errori genetici dagli errori strumentali compiuti nell'analisi,
3. Filtraggio e annotazione[19].

1. La fase di allineamento comprende anche una fase di assemblaggio in cui le sequenze lette ed allineate vengono composte tra loro per formare la sequenza genetica originaria del campione oggetto di analisi. La sequenza allineata viene memorizzata in formato SAM (Sequence Alignment/Map) e al termine dell'allineamento il formato viene convertito in BAM (SAM binario, file compresso e più adatto ad essere utilizzato dai programmi di analisi che operano i passaggi successivi.

Per avere un'adeguata efficienza nell'allineamento è necessario disporre di reads superiori a 30 bp: infatti soltanto il 90% del genoma umano può essere allineato in modo univoco con reads di 30 bp[20]. Un altro limite per l'allineamento è rappresentato dalle sequenze ripetute, la cui soluzione consta nell'assegnazione delle reads di sequenze ripetute a posizioni multiple nel genoma di riferimento oppure nella creazione di lacune (gaps) nell'allineamento. I tassi di errore associati alla tecnologia di nuova generazione sembrano maggiori rispetto alla tradizionale metodica Sanger; tuttavia l'accuratezza del sequenziamento nelle piattaforme NGS è assicurata da una lettura ripetuta e massiva di ogni frammento genico, che determina la "copertura" del genoma, "coverage", dato appunto dalla somma delle reads sovrapposte che coprono il genoma. Anche quest'ultimo parametro rappresenta un fattore essenziale nell'analisi, in quanto un coverage non adeguato (in generale indicato con valore inferiore a 20-50 volte in relazione alla piattaforma utilizzata)

può determinare la presenza di falsi negativi, per campioni eterozigoti, nel rilevamento di una effettiva variazione nucleotidica[21].
Per la piattaforma 454 diversi studi hanno dimostrato che coverage minori di 20-30 volte sono sufficienti a ridurre l'accuratezza nell'identificazione di un polimorfismo[22]. Per il formato Illumina sono state utilizzate intensità di coverage 50-60 volte più elevate in modo da migliorare l'allineamento di short reads, l'assemblaggio e l'accuratezza, sebbene un "coverage" di 20-30 volte possa essere sufficiente per alcune applicazioni di risequenziamento[23].
Sono stati sviluppati e resi disponibili per la comunità scientifica un grande numero di programmi per l'allineamento e l'assemblaggio, la maggior parte dei quali utilizza il sistema operativo Linux, mentre alcuni sono disponibili per Windows. Anche se sta crescendo il numero di algoritmi disponibili per l'allineamento e l'assemblaggio, occorre valutare se sia preferibile la velocità o l'accuratezza con cui sono valutati molti ma non tutti i possibili allineamenti e cercare di trovare un equilibrio tra allineamento ideale e efficienza computazionale.

2. Dopo l'allineamento delle reads, il DNA in analisi viene confrontato con il genoma di riferimento per individuare le eventuali variazioni presenti (variant calling) e, in base alla frequenza della variante, è possibile comprendere se si tratta di una reale mutazione presente nel materiale genetico di partenza o se si tratta di un errore di incorporazione durante il sequenziamento. Lo strumento MiSeq dell'Illumina identifica le differenze rispetto alla sequenza di riferimento automaticamente e produce il file .vcf (variant call format) che può memorizzare SNPs, inserzioni, delezioni.

3. Il passaggio successivo è quindi quello di determinare quali di queste variazioni, non dovute ad errori di sequenziamento hanno un possibile significato patogeno, pertanto è necessario procedere ad un'attenta selezione eliminando tutti i polimorfismi clinicamente neutrali. La fase di filtraggio, infatti, riduce notevolmente il numero di varianti su cui indirizzare l'attenzione utilizzando criteri di selezione piuttosto ampi, ma efficaci come la frequenza allelica, in modo da escludere automaticamente tutte le varianti con frequenza uguale o superiore all'1%, poiché si tratta di polimorfismi neutrali. Un altro criterio interessante è rappresentato dalla presenza delle varianti in tessuto normale e tessuto tumorale, per escludere le varianti

che sono presenti sia nell'uno e nell'altro in studi in cui si vogliono individuare le mutazioni somatiche sviluppatesi nelle sole cellule tumorali. Ancora, un altro parametro su cui focalizzare l'attenzione è la modalità di segregazione delle varianti nell'albero genealogico per selezionare tutte le varianti compatibili con un dato modello di trasmissione: ad esempio, nell'ipotesi che si stia cercando di individuare la mutazione patogena in una famiglia con una malattia a chiara trasmissione autosomica recessiva, il sistema selezionerà automaticamente tutte le varianti che sono eterozigoti nei genitori e omozigoti nel figlio affetto.

In linea di massima, nella diagnostica delle malattie a trasmissione mendeliana, la frequenza allelica è il criterio più usato.
Il tool più utilizzato al momento per l'identificazione e l'analisi delle varianti è fornito nel software GATK (Genome Analysis Toolkit), utilizzato anche nel progetto *1000 Genome Project* e nel *The Cancer Genome Browser*. Nella valutazione delle varianti identificate è di fondamentale importanza basarsi sulle varianti prefiltrate, rese accessibili dai progetti HAPMap Project e dbSNP, come training data per il clustering dei dati da analizzare.
Il formato di file per la memorizzazione delle varianti è il Variant Call Format (VCF), un formato testuale che può memorizzare SNP, DIP e varianti strutturali più lunghe. Il formato prevede anche uno standard per l'annotazione di informazioni come genotipo, allele ancestrale, profondità di lettura, qualità della mappatura.
Terminata la fase di filtraggio, si passa alla fase di annotazione, che può essere automatica, ma non di rado è ancora parzialmente o totalmente manuale. In questa fase si cercano di prendere in considerazione parametri multipli: risultato delle predizioni in silico, possibile influsso del contesto genomico circostante, eventuali effetti modulatori sul fenotipo pur in assenza di un effetto patogeno diretto e, se esistenti, dati già pubblicati in letteratura.
Il MiSeq dell'Illumina dispone di un programma "Illumina Variant Studio" che, dopo aver caricato il .vcf file, fornisce un output in cui è riportato il gene, la variante, le coordinate cromosomiche, il cromosoma, il tipo di variazione, il genotipo e se ha superato gli indici di qualità del dato.

1.4 Vantaggi e limiti delle piattaforme NGS

In base ai dati raccolti dal "National Human Genome Research Institute" è possibile affermare che i costi di sequenziamento del genoma si sono notevolmente ridotti negli anni. La Tabella 1 riporta le performances relative al sequenziamento Sanger e alle piattaforme di nuova generazione mostrando, oltre ai costi, una riduzione dei tempi, aumento del throughput, migliore qualità delle reads e riduzione della lunghezza delle reads (ad esclusione della tecnologia Roche 454).

Table 5 Sequencing statistics of six individual human genomes

platform	individual	No. of reads (millions)	read length (bases)	read coverage	genome coverage (%)	SNPs (millions)	No. of runs	estimated cost (US$)	references
Sanger	J. Craig Venter	31.9	800	7.5 ×	N/A	3.21	>340,000	70,000,000	Levy et al., 2007
Roche 454	James D. Watson	93.2	250	7.4 ×	95	3.32	234	1,000,000	Wheeler et al., 2008
SOLiD	James R. Lupski	238	35	29.6 ×	99.8	3.42	3	75,000	Lupski et al., 2010
Illumina Solexa	Yoruba male (NA18507)	3681	35	40.6 ×	99.9	4	40	250,000	Pushkarev 2009
	Han Chinese male (YH)	2950	35	36 ×	99.9	3.07	35	500,000	Wang et al., 2008
	Korean male (SJK)	1647	35, 74	29.0 ×	99.9	3.44	15	250,000	Ahn et al., 2009
	Korean male (AK1)	1910	36, 88, 106	27.8 ×	99.8	3.45	30	200,000	Kim et al., 2009
Helicos	Stephen R. Quake	2725	32	28 ×	90	2.81	4	48,000	Pushkarev et al., 2009

Tabella 1: Confronto tra le diverse piattaforme NGS.

Tutti questi vantaggi sono dovuti al fatto che le NGS hanno permesso di condurre in parallelo migliaia/milioni di reazioni di sequenza superando il limite delle 1-96 reazioni possibili con i tradizionali macchinari di sequenziamento; inoltre il clonaggio o l'amplificazione del frammento di DNA risultano non necessari o completamente automatizzati all'interno delle piattaforme. Però, la minore estensione delle reads rende più difficoltoso recuperare la sequenza intera di partenza, pertanto, per ovviare a questo problema, è possibile sequenziare read paired-end o mate-pair, ovvero coppie di read per cui è nota la distanza (alcune decine di migliaia di basi) sulla sequenza di DNA.

Avere reads corte dà problemi anche in altre applicazioni, come il de novo assembly, che si occupa di ricomporre il genoma senza utilizzarne uno di riferimento. Questo sistema risulta di difficile applicazione con le NGS per la brevità dei frammenti da unire, brevità che causa la valutazione errata dei frammenti, da cui si ha una

ricostruzione errata del genoma. Oltre alla lunghezza ridotta delle reads, un altro problema è sorto col progredire di questi sequenziatori.
Ottenere una notevole quantità di dati dalle tecnologie di sequenziamento di nuova generazione rappresenta senza dubbio un considerevole vantaggio, tuttavia essa risulta essere una capacità superiore alla capacità di gestione, analisi e stoccaggio dei dati. I limiti tecnologici dell'hardware sono di fatto diventati il collo di bottiglia per questi strumenti, sia come limiti di calcolo, sia come limiti di immagazzinamento dati.

1.4 Applicazioni

I sequenziatori di nuova generazione sono diventati strumenti d'elezione per analisi approfondite nel campo della biologia e della medicina. Questi strumenti consentono di approcciarsi in maniera nuova e più globale ad una serie di applicazioni come sequenziamento de novo, resequencing, studi di metagenomica, di epigenomica e sequenziamento del trascrittoma.
In particolare, il sequenziamento de novo si propone di determinare genomi di organismi non ancora sequenziati, per i quali quindi non si dispone di genomi di riferimento. Questo è l'ambito che maggiormente ha beneficiato dell'aumento della processività e dell'abbattimento dei costi raggiunti con l'NGS. Basti pensare che queste tecniche hanno bisogno di una decina di giorni per sequenziare un intero genoma, contro i dieci anni impiegati la prima volta col metodo Sanger.
Un'altra applicazione consiste nel risequenziamento di un intero genoma già sequenziato allo scopo di identificare mutazioni di un singolo nucleotide (SNP), inserzioni e delezioni di sequenze più o meno lunghe in determinate posizioni del genoma e variazioni nel numero di copie di determinati geni (CNV). Questo è utile per comprendere i meccanismi di sviluppo di alcune patologie in modo da individuare le direzioni per un futuro trattamento clinico.
Oltre all'intero genoma, le piattaforme NGS possono essere utilizzate per sequenziare selettivamente determinate regioni genomiche o un pannello di geni (mediante il targeted resequencing), oltre ad ottenere informazioni sull'intero esoma (tramite il whole exome-sequencing). Questo può essere molto utile per quei casi di eterogeneità genetica, cioè più geni che danno lo stesso fenotipo patologico. In questo ambito si sta lavorando per produrre pannelli genici certificati, in modo da semplificare notevolmente il lavoro.

Le tecnologie NGS hanno permesso l'avvento su larga scala della meta-genomica, cioè lo studio dell'insieme del materiale genetico di tutti gli organismi viventi di un determinato campione ambientale.

L'impatto di queste nuove metodiche di analisi ha permesso di ottenere l'accesso ad informazioni maggiori di quelle ottenibili in precedenza. Infatti, lo studio di comunità microbiche direttamente nel loro ambiente naturale consente di superare il problema del prelevamento e della coltivazione in laboratorio. In molti casi, risulta difficile riprodurre le reali condizioni di un determinato ambiente e, molto spesso, alcuni organismi hanno particolari esigenze (temperature elevatissime, pressioni pari a quelle dei fondali oceanici, concentrazioni saline alte, ecc.). In questo caso ciò che viene sequenziato è la regione variabile dell'RNA ribosomiale 16S, una regione altamente conservata nel gruppo tassonomico ma che presenta una variabilità interspecifica sufficiente a discriminare tra le varie specie batteriche.

La caratterizzazione delle comunità microbiche umane e ambientali risulta utile nell'identificazione di geni (e specie codificanti) impiegabili per scopi commerciali quali la produzione di biocarburanti o di altri composti a carattere farmaceutico, agrochimico o enzimatico[24-25]. In relazione alle applicazioni in ambito clinico, va menzionato il recente utilizzo delle tecnologie NGS per lo studio della variabilità genomica di virus patogeni, quali HIV e HCV[26].

L'epi-genomica rappresenta un'altra applicazione delle piattaforme NGS e riguarda lo studio dei meccanismi di regolazione dell'espressione genica.

Le proprietà epigenetiche sono tipiche dei cromosomi eucariotici perché solo i cromosomi eucariotici sono organizzati in cromatina. Un carattere epigenetico è ereditabile, ma indipendente dalla sequenza del DNA, che può determinare una variazione dell'espressione genica senza alterare la sequenza del DNA[27].

Per esempio, è possibile fare attraverso una tecnica di immunoprecipitazione della cromatica (ChIP-Seq) uno studio di interazione tra DNA e fattori di trascrizione o tra DNA e proteine istoniche. Un altro meccanismo che può essere studiato è la metilazione del DNA.

La ChIP-Seq della cromatina permette di individuare i punti di DNA a cui si legano i fattori di trascrizione o gli istoni, o proteine che controllano la replicazione del DNA stesso e il successivo sequenziamento di questi frammenti di DNA permette, quindi, di ottenere un quadro complessivo dell'attività regolativa della proteina analizzata. Con tale approccio, le proteine associate al DNA genomico sono ibridate

chimicamente (tipicamente tramite trattamento blando con formaldeide) ai loro siti di legame e il DNA è frammentato mediante sonicazione o digestione con nucleasi micrococcica[5]. Le proteine ibridate con il DNA sono immunoprecipitate con anticorpi specifici per le proteine di interesse. Il DNA nell'immunoprecipitato viene convertito in una libreria che viene analizzata mediante piattaforme NGS. Le reads ottenute sono mappate al genoma di interesse di riferimento per generare una mappa sull'intero genoma (genome-wide) dei siti di riconoscimento delle DNA-binding proteins[28-29]

Un importante vantaggio caratteristico della ChIP-Seq è rappresentato dal fatto di non richiedere la conoscenza a priori delle posizioni genomiche di legame proteico.

Le tecnologie di nuova generazione hanno fornito un nuovo e potente approccio per la mappatura e la quantificazione dei trascritti nei campioni biologici, denominato RNA-Seq. il Sequenziamento del trascrittoma o dei microRNA (piccoli RNA non codificanti ad attività regolativa) si può considerare un altro formidabile strumento d'indagine per lo studio delle funzioni biologiche di una cellula, in quanto fornisce informazioni quantitative sulle differenze nei livelli di espressione dei geni in tessuti diversi o in uno stesso tessuto analizzato in diverse condizioni sperimentali (ad esempio sottoposto a diversi trattamenti) o in differenti stadi di sviluppo[30].

Un protocollo tipico prevede una prima fase di retrotrascrizione dell'RNA di partenza in modo da ottenere un filamento di cDNA che viene successivamente frammentato e legato ad adattatori per l'analisi mediante le tecnologie NGS. Dopo il sequenziamento, le reads sono allineate a un genoma di riferimento, comparate a sequenze di trascritti noti o utilizzate per un de novo assembly per costruire una mappa di trascrizione.

Sebbene la tecnologia di RNA-Seq sia ai primi stadi di sviluppo, essa ha già mostrato alcuni vantaggi: consente la caratterizzazione dei trascritti senza una conoscenza a priori dei siti genomici di inizio trascrizione; permette di ottenere una risoluzione a livello di singole basi e mostra una maggior capacità di distinguere isoforme di RNA; inoltre permette di determinare l'espressione allelica e di rilevare le variazioni di sequenza. I livelli di espressione sono dedotti dal numero totale di reads che mappano per gli esoni di un gene, normalizzate per la lunghezza degli esoni che possono essere mappati in modo univoco[5].

Un ulteriore ambito di applicazione delle NGS può rientrare nel campo della diagnosi prenatale al fine di studiare casi di aneuploidie usando DNA fetale che si

riscontra nel sangue materno durante la gravidanza evitando, così, l'utilizzo delle attuali metodiche invasive come per esempio l'amniocentesi[31].

1.5 Sequenziamento del DNA di terza generazione

Negli ultimi anni molti sforzi sono stati dedicati allo sviluppo di una nuova generazione di tecnologie più sensibili che, grazie all'impiego di nanotecnologie, sono capaci di riconoscere fisicamente i singoli nucleotidi. Battezzate third-generation sequencing, queste rappresentano un'evoluzione della generazione precedente in quanto offrono reads più lunghe, più alto throughput e maggior accuratezza oltre a una riduzione di costi e tempo impegnato. Le principali protagoniste di questa nuova fase, sono Ion Torrent, Pacific Biosciences e Oxford Nanopore.

La Ion Torrent sfrutta un processo analogo al pirosequenziamento, ma invece di rivelare il pirofosfato, misura gli ioni idrogeno rilasciati durante la reazione di elongazione della molecola di DNA ad opera della DNA polimerasi[32]. Il rilascio dei protoni causa l'abbassamento del pH della soluzione presente all'interno del pozzetto; questa variazione viene trasmessa e registrata dal sensore sottostante e digitalizzata (Figura 9).

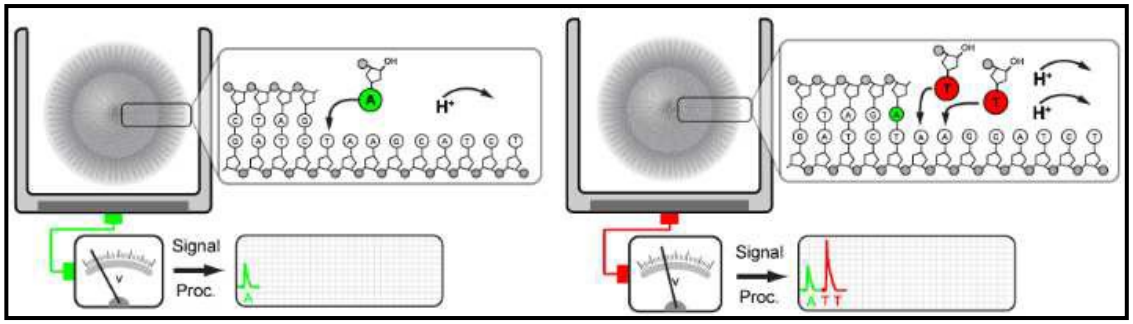

Figura 9: Schema di sequenziamento nell'Ion Chip. Primers, Dna polimerasi e biglie a cui è legato il templato sono contenuti all'interno di un pozzetto. Ciclicamente fluiscono sulla superficie del chip i quattro nucleotidi. La differenza di potenziale viene registrata per ciascun pozzetto del chip ad ogni flow e, nel caso si registri una sua variazione, questa e tradotta nella chiamata della base.

Durante ciascun flow, la variazione di pH registrata è direttamente proporzionale al numero delle basi incorporate nella sequenza nascente e, nel momento in cui la sequenza stampo presenta una zona omopolimerica, l'intensità della variazione di potenziale registrata è tanto superiore quante più sono le basi identiche consecutive.

Questo dato viene rappresentato attraverso uno grafico definito Ionogramma, rappresentato in Figura 10, in cui è visualizzato il numero delle basi incorporate rispetto ai flows eseguiti dalla macchina.

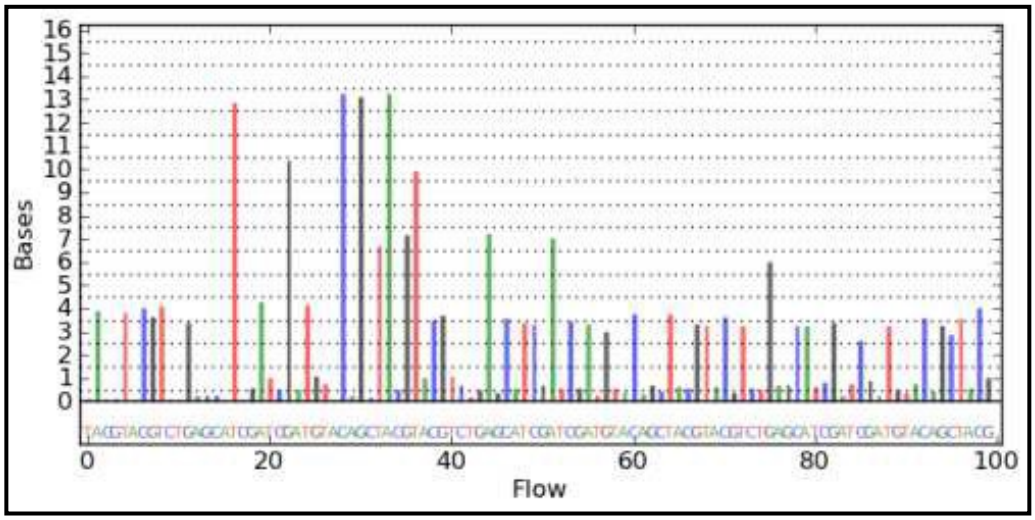

Figura 10: Esempio di ionogramma prodotto dal PGM™ Ion Torrent

Un aspetto vantaggioso offerto dalla tecnologia Ion Torrent è la disponibilità di chip con throughput differente che consente di modulare il sistema in base alle necessità sperimentali, inoltre è un sistema semplice ed economico in quanto si basa su una rilevazione di pH che non necessita di costosi sistemi di rilevazione ottica per l'acquisizione delle immagini.

Un sistema davvero innovativo è rappresentato da Pacific BioScience che utilizza la DNA polimerasi e nucleotidi marcati con la produzione di reads lunghe 1000-1500 basi in tempi brevi (15-20 minuti per corsa). L'enzima viene immobilizzato su un chip, posto a contatto con i diversi frammenti di DNA e il sistema permette di seguire in tempo reale il processo di replicazione emettendo un segnale fluorescente diverso a seconda della base incorporata. Purtroppo è una macchina molto costosa con un tasso di errore molto alto (15-20%) dovuto proprio al sistema di rivelazione non all'attività dell'enzima. Per questo motivo si sta pensando di accoppiare questa tecnologia con la piattaforma Illumina che garantisce il coverage, ma la lunghezza delle reads prodotte con Pac-Bio ne facilita l'allineamento.

L'ultima innovazione tecnologica genomica, presentata durante il meeting AGBT (Advances in Genome Biology & Technology) di Marco Island, in Florida, si chiama MiniION. Presentato dall'azienda inglese Oxford Nanopore Technologies, il dispositivo è un sequenziatore di DNA USB che collegato a un computer può sequenziare del materiale genetico e in tempo reale trasmettere i dati a un software

installato sul PC. Il sistema è costituito da una matrice composta da tanti piccoli pori che vengono attraversati dall'acido nucleico ed è capace di riconoscere le basi durante questo passaggio. Dal momento che il riconoscimento si basa sulle caratteristiche chimico-fisiche della molecola che sta attraversando il poro, la tecnologia permette di rilevare in tempo reale anche le modificazioni sul DNA, come per esempio una metilazione, ampliando le conoscenze del campo dell'epi-genomica.

2. Le tecnologie NGS per la diagnosi clinica

La Next Generation Sequencing è molto più che una nuova tecnica per leggere le sequenze di DNA. Produce sequenze con una processività parallela immensa che continua a crescere in modo esponenziale e oggi è nell'ordine di centinaia di miliardi di basi di DNA per analisi. Anche per la genetica umana è diventata una tecnologia insostituibile. L'NGS ha consentito di effettuare o rivalutare la diagnosi clinica di alcune malattie genetiche rivoluzionando i test genetici diagnostici, sostituendo l'approccio "gene per gene" con una strategia a pannelli di geni o dell'intero esoma. L'introduzione delle tecnologie di Next Generation Sequencing è particolarmente promettente per la diagnosi di malattie pediatriche neuromuscolari, endocrinologiche, metaboliche ecc, che sono caratterizzate da una forte eterogeneità clinica e genetica. Con la NGS è possibile effettuare un'indagine genetica andando a sequenziare, per finalità di ricerca, l'intero genoma, oppure solo la porzione codificante (l'esoma) o un pannello molto ampio di geni senza dover necessariamente ipotizzare a priori un gene causativo della patologia. Infatti, approcci WGS (Whole Genome Sequencing, ovvero sequenziamento dell'intero genoma) vengono utilizzati per scoprire variazioni genetiche che contribuiscono a malattie rare o comuni. Nonostante la diminuzione nei costi di sequenziamento questi approcci restano molto dispendiosi nella gestione e nell'analisi di un numero grande di campioni. Al contrario, approcci WES (Whole Exome Sequencing, sequenziamento dell'intero esoma) sono ritenuti una valida alternativa al WGS in quanto possono identificare le varianti genetiche potenzialmente associate alle patologie concentrandosi solo sull'esoma. L'esoma costituisce circa l'1% dell'intero genoma e rappresenta la porzione codificante.

Nonostante le difficoltà interpretative dovute all'alta densità dei dati NGS, le informazioni generate dall'applicazione di queste tecnologie creano enormi opportunità per una maggiore comprensione scientifica di numerose patologie con importanti ricadute a livello clinico-diagnostico-terapeutico:

- Per l'identificazione dei geni-malattia nelle malattie mendeliane,
- per chiarire la patogenesi delle malattie multigeniche e multifattoriali,
- per la diagnosi differenziale,
- per il trattamento personalizzato,
- il monitoraggio dell'evoluzione clinica e la valutazione della prognosi.

L’adozione delle NGS nella routine diagnostica determina la stesura di linee guida che sarebbe opportuno seguire prima di effettuare ogni procedura WGS/WES:

- il fenotipo o i dati familiari sono indicativi di una patologia genetica ma il fenotipo non è riconducibile ad una specifica ipotesi diagnostica;
- il paziente è chiaramente affetto da una patologia genetica con elevata eterogeneità, per studiare la quale l’approccio WES risulta più conveniente;
- il paziente è probabilmente affetto da una malattia genetica ma i test specifici non hanno permesso di raggiungere una diagnosi.

In questo quadro si inserisce il mio lavoro di tesi dove è stato effettuato il sequenziamento dell’esoma di una paziente con un quadro clinico non ascrivibile ad una patologia ben definita. La sintomatologia riscontrata portava ad ipotizzare la presenza di una patologia che rientrava nei difetti congeniti del metabolismo, un gruppo eterogeneo di patologie causate dall'alterato funzionamento di una specifica via metabolica.

3. Difetti Congeniti del Metabolismo

Con il termine metabolismo si indica l'insieme dei processi che l'organismo mette in atto per trasformare le proteine, i grassi e gli zuccheri contenuti negli alimenti, in altri composti utili, oppure in sostanze più semplici allo scopo di ricavarne energia.

Le malattie metaboliche sono un gruppo eterogeneo di patologie rare, genetiche e non tutte di natura ereditaria, molte delle quali non hanno ancora una diagnosi certa né una terapia. Esse sono causate da un difetto nella catena di funzionamento delle reazioni metaboliche determinando un'alterazione di quei processi biochimici che consentono alla cellula di utilizzare e scomporre sostanze come i carboidrati (zuccheri), le proteine e i lipidi, assunti con la dieta, in amminoacidi, acidi grassi e glucosio. Questi composti sono poi ulteriormente ridotti ad acqua ed anidride carbonica all'interno delle cellule attraverso la rottura enzimatica dei loro legami chimici. Quest'ultimo processo libera l'energia necessaria alle cellule per tutte le funzioni biologiche: crescita, contrazione muscolare cardiaca e scheletrica, sintesi di mielina per il sistema nervoso centrale, detossicazione dell'organismo, secrezione ed assorbimento, ecc.

Alcune malattie metaboliche sono dovute principalmente alla carenza di un prodotto importante che non viene più sintetizzato; per altre, invece, il danno è provocato dall'accumulo di metaboliti che risultano tossici; in altre ancora si verificano entrambi i meccanismi.

Le malattie metaboliche sono più di 500, alcune di esse ancora di difficile diagnosi. Per più di 40 patologie metaboliche esiste però oggi la possibilità dello screening neonatale esteso che permette di riconoscere malattie che possono comportare disabilità gravissima quando non diagnosticate alla nascita e immediatamente trattate. Ciò è importante per minimizzare gli effetti di tali patologie, che spesso si manifestano nei primissimi mesi con:

- ritardo nella crescita;
- scarso sviluppo psicomotorio;
- deficit nel linguaggio e nell'apprendimento;
- disturbi gastrointestinali;
- predisposizione a patologie secondarie.

Lo screening neonatale esteso è previsto per legge, ma non ancora attuato in tutte le regioni italiane.

Il mancato riconoscimento della patologia o il suo trattamento in centri medici non qualificati per queste patologie si traduce in un peggioramento della prognosi e della qualità di vita dei pazienti oltre che in un alto costo sociale per i gravi danni neurologici che ne derivano. La necessità per questi malati di un laboratorio specializzato e di una assistenza multi disciplinare disponibile 24 ore al giorno (pediatra, esperto in malattie metaboliche ereditarie, neurologo, anestesista, rianimatore, nefrologo, dietista, fisioterapista, servizio di assistenza sociale) ha limitato in Italia la loro assistenza a pochissimi centri qualificati.

Molte di queste patologie sono curabili attraverso l'eliminazione dalla dieta delle fonti alimentari che causano i metaboliti interessati dal blocco enzimatico e con l'utilizzazione di farmaci e cofattori enzimatici in grado di facilitare la depurazione dell'organismo dai prodotti tossici. Molte spesso comportano l'uso di farmaci orfani, prodotti farmaceutici non presenti sul mercato per il loro scarso rendimento economico; altre rispondono bene a terapie enzimatiche sostitutive, a trapianti d'organo, di epatociti, ecc. Per molte, però, non esiste al momento nessuna cura.

Da un punto di vista clinico, tutti gli errori congeniti del metabolismo possono essere divisi in due grandi categorie:

1. Malattie che coinvolgono soltanto un sistema funzionale oppure un organo o sistema anatomico. Presentano sintomi clinici uniformi, e la diagnosi è di solito facile anche quando il danno di base biochimico determina conseguenze sistemiche.

2. Malattie nelle quali il difetto biochimico di base può compromettere una via metabolica comune ad un grande numero di cellule o organi oppure è limitato ad un organo ma provoca conseguenze umorali e sistemiche. I sintomi clinici che si presentano sono molto diversi. Il sistema nervoso centrale è frequentemente coinvolto nella malattia avanzata, e la diagnosi potrebbe essere resa difficile dalla presenza di anormalità secondarie. Questa categoria include la maggior parte degli errori congeniti del metabolismo intermedio, disordini del traffico intracellulare, e disordini lisosomiali.

La seconda categoria può essere suddivisa secondo la base patofisiologica in tre gruppi, utili a fini diagnostici.

- Gruppo 1: Disturbi della sintesi o del catabolismo delle molecole complesse, come i disordini lisosomiali, perossisomiali e i disturbi del traffico intracellulare. I sintomi clinici sono permanenti, progressivi, indipendenti da eventi intercorrenti, e non correlati all'alimentazione.

- Gruppo 2**:** Errori congeniti del metabolismo intermedio che conducono ad intossicazione acuta o progressiva causata dall'accumulo di metaboliti tossici in prossimità del blocco metabolico. Questo gruppo include le aminoacidopatie, la maggior parte delle acidurie organiche, i difetti congeniti del ciclo dell'urea, e le intolleranze ereditarie agli zuccheri. In queste condizioni, un intervallo libero dai sintomi è seguito da segni clinici di intossicazione acuta o cronica e da ricorrenti crisi metaboliche. La diagnosi si basa principalmente sulla cromatografia di aminoacidi ed acidi organici nel plasma e nelle urine. Il trattamento consiste nella rimozione delle tossine.
- Gruppo 3**:** Disordini legati a "deficit energetico", nei quali i sintomi sono causati almeno in parte da una deficienza nella produzione o utilizzazione dell'energia conseguente ad un errore congenito del metabolismo intermedio nel fegato, miocardio, muscoli o cervello. Questo gruppo include glicogenosi, difetti della gluconeogenesi, acidemie lattiche congenite, difetti di ossidazione degli acidi grassi, disordini della catena respiratoria mitocondriale. Queste malattie presentano uno spettro clinico sovrapponibile, e le manifestazioni risultano talora da accumulo di composti tossici così come da un deficit di produzione dell'energia. Comunemente si riscontrano ipoglicemia, iperlattacidemia, severa ipotonia generalizzata, miopatia, cardiomiopatia, scarsa crescita, insufficienza cardiaca, sindrome della morte improvvisa infantile, e malformazioni congenite, suggerendo che il disturbo metabolico ha interessato le vie energetiche fetali[33].

Nei capitoli successivi verrà trattata, in particolare, la patologia oggetto di studio del presente lavoro di tesi, ovvero l'Aciduria Metlmalonica con Omocistinuria, causata da un difetto nel metabolismo degli amminoacidi.

3.1 Metabolismo degli amminoacidi

Gli amminoacidi, derivati in gran parte dalla degradazione delle proteine della dieta o intracellulari, sono una classe di biomolecole la cui ossidazione porta un contributo significativo alla generazione di energia metabolica nella cellula. La degradazione ossidativa comporta, per gli amminoacidi, la perdita del gruppo amminico con conseguente generazione degli α-chetoacidi i quali vengono ossidati a CO_2 e acqua. Accade spesso che lo scheletro carbonioso degli amminoacidi produca unità a tre o

quattro atomi di carbonio che possono essere utilizzate per la sintesi di glucosio necessario per le funzioni del cervello, del muscolo scheletrico e di altri tessuti.
I 20 processi metabolici a carico degli amminoacidi convergono verso la formazione di sei prodotti principali, tutti in grado di entrare nel ciclo dell'acido citrico (Figura 11). Gli scheletri carboniosi sono quindi indirizzati alla gluconeogenesi o alla chetogenesi o sono completamente ossidati a CO_2 e acqua.

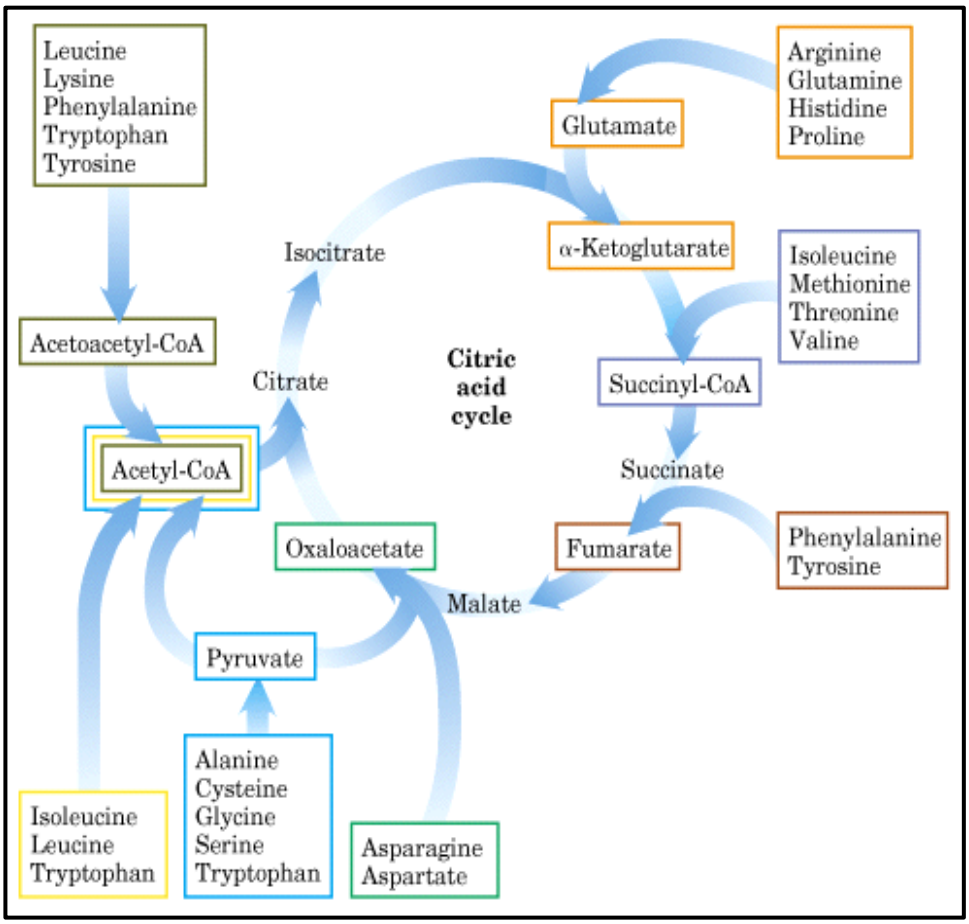

Figura 11: Riepilogo del catabolismo degli amminoacidi. Gli amminoacidi sono qui raggruppati in funzione del loro principale prodotto finale di degradazione. Alcuni amminoacidi sono elencati più volte in quanto parti del loro scheletro carbonioso sono degradate in prodotti finali differenti.

In particolare, alanina, triptofano, cisteina, serina e glicina vengono convertiti in parte o completamente in piruvato che può essere ulteriormente convertito in acetil-CoA (un precursore dei corpi chetonici) o in ossalacetato (precursore per la gluconeogenesi); vengono degradati ad acetil-CoA anche triptofano, lisina, fenilalanina, tirosina, leucina, isoleucina e treonina; attraverso la formazione dell'α-chetoglutarato, prolina, glutammato, glutammina, arginina e istidina entrano nel ciclo

dell'acido citrico; asparagina e aspartato entrano nel ciclo dell'acido citrico sottoforma di ossalacetato; mentre gli scheletri carboniosi di metionina, isoleucina, treonina, e valina sono degradati da vie metaboliche che portano alla formazione del propionil-CoA che viene carbossilato a metilmalonil-CoA, da parte dell'enzima propionil-CoA carbossilasi e, successivamente, convertito in succinil-CoA, un intermedio del ciclo dell'acido citrico, da parte dell'enzima metilmalonil-CoA mutasi che utilizza il coenzima B_{12} (Figura 12).

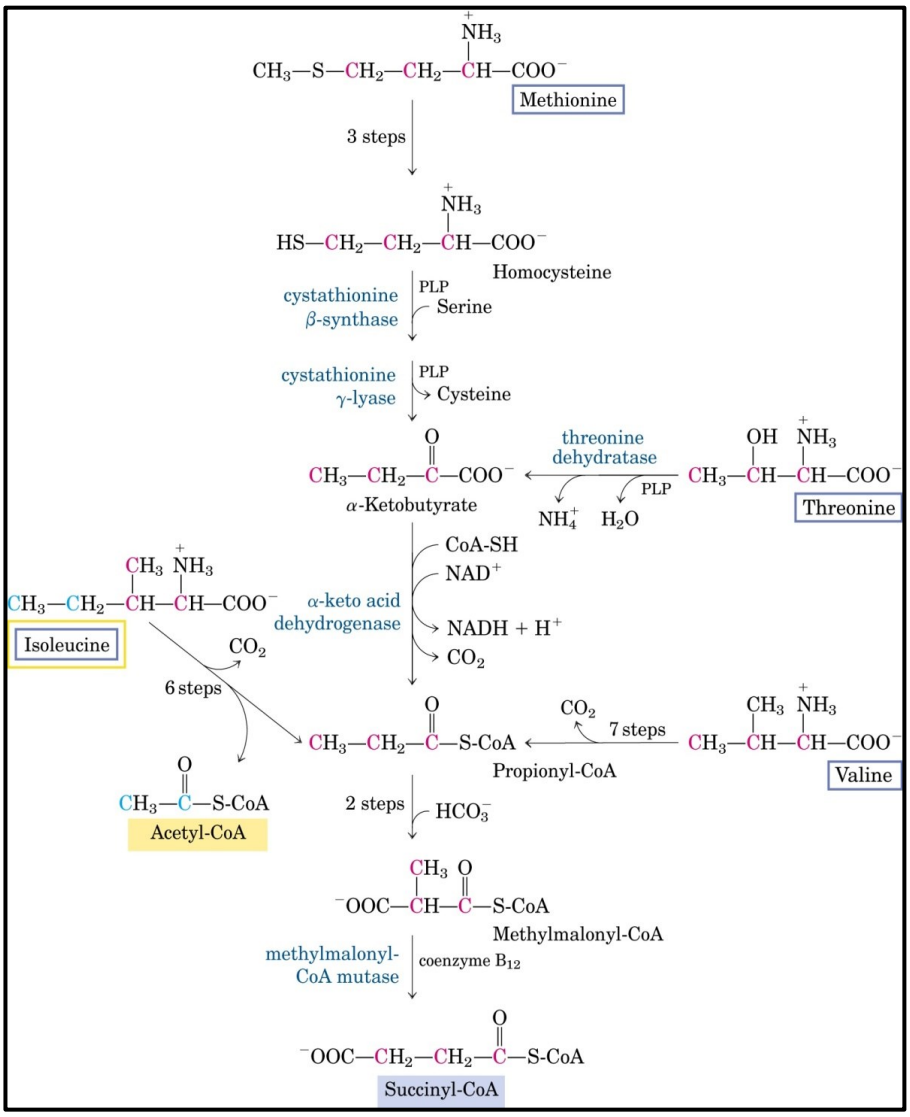

Figura 12: Vie cataboliche di metionina, isoleucina, treonina e valina.

3.2 Aciduria Metilmalonica con Omocistinuria

Un difetto genetico a carico della metilmalonil-CoA mutasi causa una grave malattia, chiamata Acidemia Metilmalonica che conduce ad un accumulo di acido metilmalonico nel plasma e nelle urine causando serie complicanze metaboliche.

Sono state identificate diverse forme di deficit di attività dell'enzima metilmalonil-CoA mutasi, dovuti a mutazioni a livello del locus MUT, che codifica per la mutasi, designate "mut0" (attività enzimatica non evidenziabile) e "mut–" (presenza di attività residua della mutasi, sebbene anomala); oppure il deficit enzimatico può essere dovuto a mutazioni a livello dei due loci genici MMAA e MMAB richiesti per la biosintesi del suo cofattore, 5-deossiadenosilcobalamina (AdoCbl), forma metabolicamente attiva (insieme alle metilcobalamina) della vitamina idrosolubile cobalamina (vitamina B_{12})[33].

Come metilcobalamina (MeCbl), agisce come cofattore dell'enzima metionina sintasi che interviene nella reazione di rimetilazione dell'omocisteina nella metionina; come adenosilcobalamina (AdoCbl), agisce, come già detto, come cofattore dell'enzima mitocondriale metilmalonil-CoA mutasi.

Per essere assorbita nell'ileo terminale, la vitamina B12 alimentare richiede un fattore intrinseco, una glicoproteina secreta dalle cellule parietali gastriche. Tale proteina è trasportata nel sangue da aptocorrina (TCI) e transcobalamina II (TCII). Il complesso transcobalamina II-cobalamina (TCII-Cbl) è riconosciuto da uno specifico recettore sulla membrana cellulare e penetra nella cellula mediante endocitosi. In seguito alla idrolisi nel lisosoma del complesso TCII-Cbl, la cobalamina libera è rilasciata nel citosol. Il cobalto della molecola è ridotto da tre valenze a due prima di entrare nei mitocondri, dove si verifica un'ulteriore riduzione a cobalamina con una valenza. Quest'ultimo composto reagisce con l'adenosina, per formare adenosilcobalamina). La cobalamina libera nel citosol subisce anche una serie di reazioni enzimatiche che conducono alla formazione di metilcobalamina (Figura 13).

Sono stati identificati almeno otto difetti diversi del metabolismo intracellulare della cobalamina. I difetti cblA, cblH e cblB causano esclusivamente acidemia metilmalonica; cblB è provocato da un deficit di adenosilcobalamina trasferasi. Nei pazienti con cblC, cblD e cblF, risultano compromesse sia la sintesi di adenosilcobalamina sia quella di metilcobalamina, con conseguente omocistinuria in associazione all'acidemia metilmalonica. I difetti cblE e cblG coivolgono solo la sintesi di metilcobalamina, causando omocistinuria senza aciduria metilmalonica (ma solitamente è presente anemia megaloblastica).

Le manifestazioni cliniche dei pazienti con acidemia metilmalonica dovuta a mut0, mut–, cblA, cblB e cblH sono analoghe. A prescindere dalla natura del difetto

enzimatico o dall'anomalia biochimica, la presentazione clinica è estremamente variabile, dal neonato gravemente malato all'adulto asintomatico.

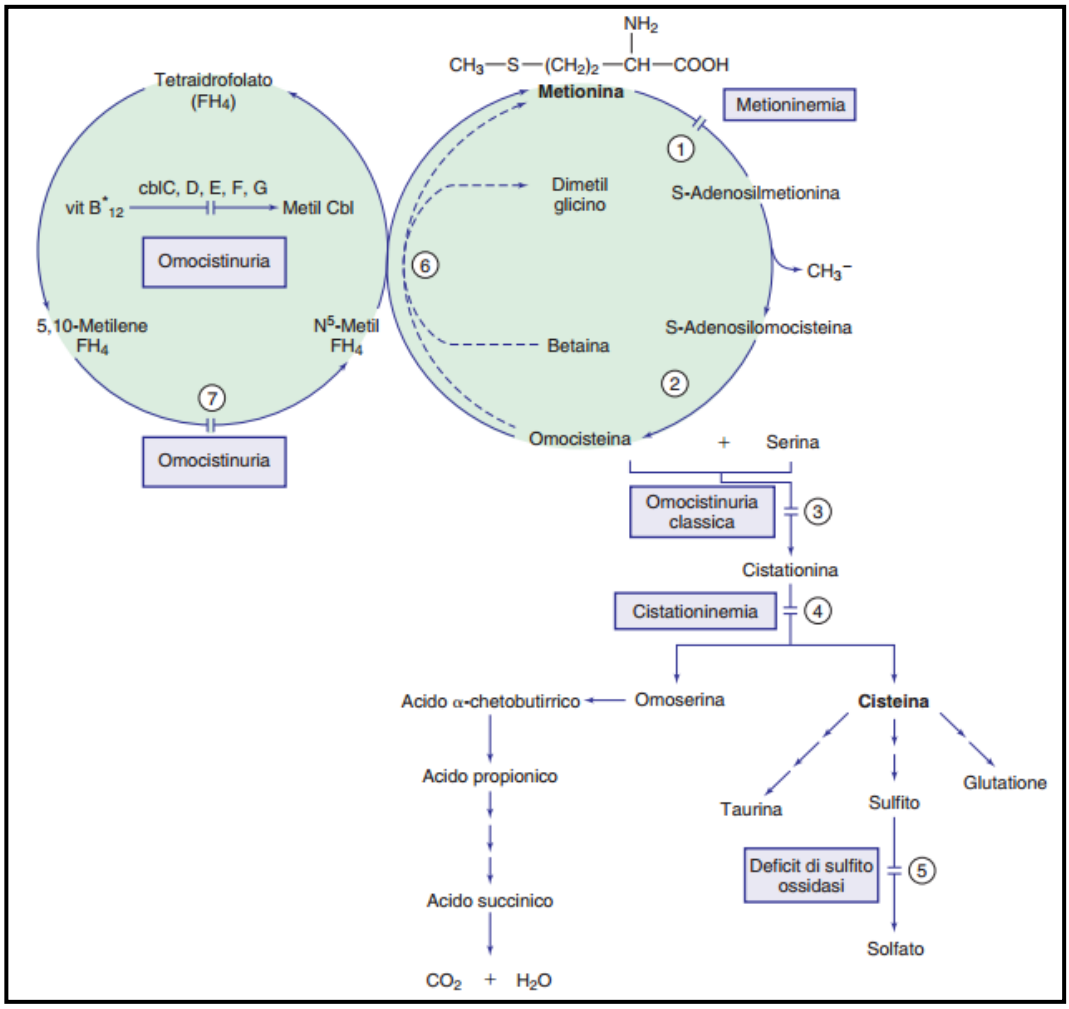

Figura 13: Vie metaboliche degli amminoacidi contenenti zolfo. Enzimi: (1) metionina adenosiltrasferasi, (2) adenosilomocisteina idrolasi, (3) cistationina sintetasi, (4) cistationasi, (5) sulfito ossidasi, (6) betaina omocisteina metiltrasferasi, (7) metilene tetraidrofolato reduttasi.

Nelle forme più gravi si sviluppano nei primi giorni di vita letargia, disturbi dell'allattamento, vomito, tachipnea (dovuta all'acidosi) e ipotonia e, se non trattate, possono progredire fino al coma e alla morte del paziente. I bambini che sopravvivono al primo attacco possono andare incontro a nuovi episodi metabolici acuti durante stati catabolici (come infezioni) o in seguito a una dieta a elevato contenuto proteico. Tra un attacco e l'altro, solitamente il paziente continua a manifestare ipotonia e disturbi dell'allattamento, con deficit della crescita staturo-ponderale.

Nelle forme più lievi, sintomi come ipotonia, deficit della crescita staturo-ponderale e ritardo dello sviluppo possono presentarsi anche durante l'infanzia. Sono stati segnalati anche casi di individui asintomatici con le anomalie biochimiche tipiche dell'acidemia metilmalonica. È importante sottolineare che il QI e lo sviluppo mentale dei pazienti con acidemia metilmalonica possono rientrare in un range di normalità, malgrado ripetuti attacchi acuti, a prescindere dal tipo di deficit enzimatico.

I reperti di laboratorio includono chetosi, acidosi, anemia, neutropenia, trombocitopenia, iperglicinemia, iperammoniemia e la presenza di elevate quantità di acido metilmalonico nei liquidi corporei.

La diagnosi può essere confermata misurando l'incorporazione di propionato o l'attività della mutasi, eseguendo test di complementazione su colture di fibroblasti o identificando il gene mutante.

Il trattamento degli attacchi acuti viene attuato somministrando nei pazienti un'elevata dose di vitamina B12 (1 mg/die, quantità che può essere ridotta sulla base della risposta clinica). Il trattamento a lungo termine prevede una dieta a basso contenuto calorico e proteico (1,0-1,5 mg/kg/die).

Una terapia alcalinizzante cronica è solitamente necessaria per correggere l'acidosi cronica, in particolare durante la prima infanzia. I livelli ematici di ammoniaca di solito si normalizzano tra un attacco e l'altro, per cui il trattamento cronico dell'iperammoniemia è richiesto solo in rari casi. La costipazione o le situazioni di stress (come le infezioni) che rischiano di scatenare un attacco acuto devono essere prevenute o trattate tempestivamente.

Una complicanza frequente e preoccupante della gestione a lungo termine di questi pazienti consiste in un'inadeguata alimentazione orale, conseguenza dello scarso appetito. L'alimentazione enterale (mediante sonda nasogastrica o gastrotomia) dovrebbe essere presa in considerazione nella fase iniziale del trattamento. Per garantire una dieta equilibrata, è necessario un attento monitoraggio del pH, dei livelli degli amminoacidi, delle concentrazioni ematiche e urinarie di metilmalonato e dei parametri di crescita. Il trapianto di fegato (e di fegato e reni combinato) è stato tentato con successo variabile.

La prognosi dipende dalla gravità dei sintomi e dall'insorgenza di complicanze. In generale, i pazienti con deficit dell'apoenzima mutasi (mut0, mut–) hanno una prognosi meno favorevole; il difetto cblA ha un esito migliore rispetto al cblB. Nei

sopravvissuti sono state notate un certo numero di complicanze. In alcuni pazienti nel corso di un attacco acuto di scompenso metabolico si sono verificati infarti cerebrali con conseguenti manifestazioni extrapiramidali (tremori, distonia) e piramidali (paraplegia). La patogenesi di questa complicanza resta poco chiara.
In alcuni pazienti adulti con questo disturbo è stata riscontrata un'insufficienza renale che ha reso necessario il trapianto. Questa complicanza può essere osservata in tutte le forme genetiche del deficit e si ritiene che la sua principale causa sia la nefrite tubulo-interstiziale, la cui patogenesi resta da definire. Episodi di pancreatite acuta e ricorrente sono stati osservati già in lattanti di 13 mesi. Questa complicanza è una delle principali cause di ricovero in ospedale.
La prevalenza della malattia è stimata a 1/48 000 e tutti i difetti responsabili di acidemia metilmalonica sono ereditati con modalità autosomica recessiva. Lo screening neonatale di massa è realizzato mediante spettrometria di tandem massa.
Il gene per la mutasi è localizzato sul braccio corto del cromosoma 6; sono state identificate almeno 160 differenti mutazioni patogenetiche del gene MUT, incluso un certo numero di mutazioni specifiche per determinate etnie. Sono stati segnalati neonati con acidemia metilmalonica e grave diabete dovuto all'assenza di cellule che presentano grave isodisomia paterna uniparentale del cromosoma 6 (i cromosomi ereditati dallo stesso genitore sono identici dal punto di vista allelico). In un certo numero di pazienti sono state identificate mutazioni del gene per cblA (MMAB , localizzato sul cromosoma 4q31-q31.2). Il gene per cblH non è stato ancora mappato. Sono state descritte diverse gravidanze con esito positivo sia per la mamma sia per il bambino.
Nella maggior parte dei pazienti affetti da CblA e CblB, supplementi farmacologici di cianocobalamina o idrossicobalamina comportano una riduzione dell'accumulo di metilmalonato, offrendo un valido supporto alla terapia dietetica basata sulla restrizione di proteine.

Per quanto riguarda pazienti con acidemia metilmalonica dovuta a difetti cblC, cblD, cblF risultano compromesse sia la sintesi di adenosilcobalamina sia quella di metilcobalamina, con conseguente omocistinuria in associazione all'acidemia metilmalonica. L'aciduria metilmalonica in questi pazienti è caratteristicamente meno severa di quella sviluppata nei pazienti con deficienza isolata dell'enzima mutasi.

Il difetto cblC è attualmente suddiviso in due varianti, la prima con disfunzione della metionina sintasi e la seconda della metilmanolil-CoA mutasi.
Il difetto nei disordini CblC e CblD coinvolge un passaggio iniziale del metabolismo intracellulare della Cbl, ossia la riduzione citosolica dello stato di ossidazione +3 del cobalto centrale della Cbl, dopo efflusso della Cbl dai lisosomi. Come risultato, viene ostacolata la formazione sia di AdoCbl sia di MeCbl, e quindi il legame della Cbl ai due enzimi intracellulari.
I reperti neurologici sono più evidenti in queste forme cblC e cblD. I pazienti con cblC sviluppano nei primi mesi di vita un deficit staturo-ponderale, letargia, disturbi dell'allattamento, ritardo mentale, convulsioni, coma, e anormalità ematologiche come anemia megaloblastica, trombocitopenia, leucopenia e neutropenia; talora può anche manifestarsi come patologia multisistemica con insufficienza renale, miocardiopatia, retinopatia e sindrome emolitica-uremica.
L'anemia megaloblastica è di comune riscontro. Nei liquidi corporei si nota un aumento da lieve a moderato della concentrazione di acido metilmalonico e omocisteina. A differenza dei pazienti con omocistinuria classica, i livelli plasmatici di metionina risultano ridotti o normali. Non è presente iperammoniemia, né iperglicinemia.
Clinicamente, i pazienti con difetto CblC risultano più gravemente affetti del gruppo CblD, sebbene siano stati riportati casi di CblC, con esordio più tardivo della malattia, durante l'infanzia o l'adolescenza, con manifestazioni neurologiche che includono confusione, disorientamento, riduzione delle performance cognitive, mentre le anormalità ematologiche sono state riscontrate soltanto in metà di questi pazienti.

Il disordine CblF è, invece, dovuto a un deficit del rilascio lisosomiale della Cbl, dopo il suo legame con la transcobalamina II.
La sintomatologia, che prevede disturbo dell'attattamento, ritardo di crescita e dello sviluppo, stomatite persistente, si può manifestare nelle prime settimane di vita. Nei pazienti con difetto cblF è stato notato un malassorbimento di vitamina B12.
L'esperienza nel trattamento di pazienti con i difetti cblC, cblD e cblF è ancora piuttosto limitata.

La terapia, caratterizzata da dieta ipoproteica, dosi farmacologiche di idrossicobalamina (1-2 mg/die) e supplementazione di betaina (6-9 g/die) sembra produrre un miglioramento biochimico, ma l'effetto clinico è minimo.

Le complicanze più gravi nei pazienti con difetto cblC consistono in anemia emolitica inspiegabile, idrocefalo e insufficienza cardiaca congestiva. Il gene per cblC è localizzato sul cromosoma 1.

Infine, pazienti con difetti cblE e cblG non presentano acidemia metilmalonica in quanto queste alterazioni coinvolgono solo la sintesi di metilcobalamina, causando omocistinuria senza aciduria metilmalonica (ma solitamente è presente anemia megaloblastica).

Sono state identificate tre forme di omocistinuria: omocistinuria dovuta a deficit di cistationina β-sintetasi, (il più comune errore congenito del metabolismo della metionina), omocistinuria da difetti nella formazione di metilcobalamina e omocistinuria da deficit di metilenetetraidrofolato reduttasi (MTHFR).

Il deficit di cistationina β-sintasi (CBS) che, con la vitamina B6 come cofattore, converte l'omocisteina in cistationina nella via metabolica della trans-sulfurazione del ciclo della metionina, oltre all'aumento dell'omocisteina (>100 m/L) determina anche aumento della metionina nel sangue ed omocistinuria; mentre la cisteina è assente o molto ridotta nel plasma.

La presenza di omocisteina andrebbe verificata nell'urina fresca, perché questo composto è instabile e tende a scomparire in quella conservata.

In base ai dati epidemiologici l'incidenza della malattia è compresa tra 1/200 000 e 1/350 000, ma in alcunei paesi, come Irlanda o Australia, il disturbo sembra più comune (1/60 000).

I bambini affetti da questa malattia sono normali alla nascita, hanno frequentemente pigmentazione chiara, occhi azzurri ed un peculiare arrossamento degli zigomi. La diagnosi viene formulata di solito dopo i 3 anni, quando compare la sublussazione del cristallino (ectopia lentis) che provoca grave miopia e successivamente astigmatismo, glaucoma, cataratta, distacco retinico e atrofia ottica. Durante l'infanzia, le manifestazioni cliniche sono aspecifiche e includono deficit della crescita staturo-ponderale e ritardo dello sviluppo.

Frequente è un progressivo ritardo mentale, convulsioni (20% circa dei casi), turbe psichiatriche (50%): è stata però riscontrata un'intelligenza normale in pazienti responsivi alla vitamina B6. L'osteoporosi generalizzata, in particolare a livello della

colonna vertebrale, costituisce il reperto radiografico principale. I pazienti con omocistinuria manifestano anomalie scheletriche analoghe a quelle della sindrome di Marfan: solitamente alti e magri, hanno arti allungati e presentano aracnodattilia. Sono frequenti gli episodi tromboembolici che interessano sia i grandi che i piccoli vasi, specialmente quelli cerebrali che possono comparire a qualsiasi età. Atrofia ottica, paralisi, convulsioni, cuore polmonare ed ipertensione grave (secondaria a infarti renali) sono tra le gravi conseguenze della tromboembolia[34-35]. Questa è dovuta a modificazioni delle pareti vascolari e ad un'aumentata attivazione piastrinica secondaria ad elevati livelli di omocisteina[36]. Il rischio di tromboembolia aumenta in seguito ad interventi chirurgici.

Accanto alla forma omozigote rarissima e grave, ha un impatto di un certo rilievo l'eterozigosi per il deficit di CBS che si riscontra in circa lo 0.5% della popolazione generale[35] e l'eterozigosi per la mutazione della CBS 844ins68, presente nel 10-15% della popolazione generale[37-38]. Quest'ultima però si associa ad iperomocisteinemia se coesistono altre alterazioni genetiche o acquisite. La diagnosi si può stabilire mediante il dosaggio dell'enzima in campioni di biopsia epatica, in fibroblasti coltivati o linfociti stimolati con fitoemoagglutinina mentre la diagnosi prenatale è attuabile eseguendo un dosaggio enzimatico sulle colture di cellule amniotiche o su villi coriali o mediante analisi del DNA. Diverse mutazioni patogenetiche sono state identificate in varie famiglie.

L'omocistinuria è ereditata con modalità autosomica recessiva. Il gene della cistationina -sintetasi è localizzato sul cromosoma 21q22.3. Nella maggior parte dei casi, i pazienti colpiti sono doppi eterozigoti per due differenti alleli. I portatori sani eterozigoti sono solitamente asintomatici e gli eventi tromboembolici insieme alla malattia coronarica sono più comuni in questi soggetti rispetto alla popolazione generale.

Il trattamento con dosi elevate di vitamina B6 (200-1000 mg/ die) consente un notevole miglioramento nella maggior parte dei pazienti che rispondono a questa terapia. Il grado di risposta alla vitamina B6 può variare a seconda delle famiglie. Alcuni pazienti possono risultare non responsivi a causa di una deplezione di folato, per cui prima di considerare un paziente non responsivo è opportuno provare ad aggiungere acido folico (1-5 mg/die) alla terapia. In questi pazienti non responsivi è consigliata la restrizione dell'assunzione di metionina, associata alla somministrazione di un supplemento di cisteina. L'opportunità delle restrizioni

dietetiche nei soggetti responsivi è una questione controversa; in alcuni di essi l'aggiunta di betaina può evitare il ricorso alla dieta. La betaina (trimetilglicina, 6-9 g/die negli adulti o 200-250 mg/kg/die nei bambini) riduce i livelli di omocisteina nei liquidi corporei, mediante rimetilazione dell'omocisteina a metionina (Figura 13); ciò può determinare una ulteriore elevazione dei livelli di metionina plasmatica. Questo trattamento ha consentito il miglioramento clinico (prevenzione degli episodi vascolari) in pazienti non responsivi alla vitamina B6.

Sono stati descritti oltre 100 casi di gravidanze di donne con la forma classica di omocistinuria, con esito positivo sia per la mamma sia per il bambino. Nella maggior parte dei casi, i bambini erano sani e nati a termine.

Il secondo tipo di omocistinuria è causato da difetti nella formazione di metilcobalamina

La metilcobalamina è il cofattore per l'enzima metionina sintetasi, che catalizza la rimetilazione dell'omocisteina a metionina.

Esistono per lo meno cinque difetti diversi del metabolismo intracellulare della cobalamina in grado di interferire con la formazione di metilcobalamina e sono indicati come cblC, cblD, cblE (metionina sintetasi reduttasi), cblF e cblG (metionina sintetatsi).

I reperti di laboratorio includono anemia megaloblastica, omocistinuria e ipometioninemia. La presenza dell'anemia megaloblastica consente di distinguere questi difetti dall'omocitinuria dovuta al deficit di metilenetetraidrofolato reduttasi, il terzo tipo di omocistinuria;. la presenza di ipometioninemia aiuta a differenziare entrambi questi disturbi dal deficit di cistationina β-sintetasi.

Il quadro clinico è caratterizzato spesso da vomito, inappetenza, letargia, ipotonia e ritardo dello sviluppo psicomotorio che possono comparire nei primissimi mesi di vita.

La diagnosi si effettua mediante adeguati studi su fibroblasti coltivati. La diagnosi prenatale è realizzata mediante esami su colture di cellule amniotiche.

Il gene per cblE (metionina sintetasi reduttasi, MTRR) è stato localizzato sul cromosoma 5p15.3-p15.2, mentre quello per cblG (metionina sintetasi, MTR) è mappato sul cromosoma 1q43; sono state descritte diverse mutazioni patogenetiche, inclusa una mutazione comune missense (P1173L) nel gene MTR. Il trattamento con vitamina B12, in forma di idrossicobalamina (1-2 mg/die), è utilizzato per correggere le anomalie biochimiche e cliniche con risultati che variano a seconda del tipo di

difetto[39-40]. A volte può risultare di una qualche utilità la somministrazione di betaina.

Il terzo tipo di omocistinuria è causata dal deficit di metilenetetraidrofolato reduttasi (MTHFR). Questo enzima riduce il 5-10 metilenetetraidrofolato per formare 5-metiltetraidrofolato, che fornisce il gruppo metilico necessario alla cobalamina, la quale, una volta metilata, funge da cofattore cedendo il gruppo metilico per la rimetilazione dell'omocisteina a metionina (Figura 13).

Il gene per l'enzima è stato localizzato sul cromosoma 1p36.3. Il disturbo si trasmette con modalità autosomica recessiva e costituisce il 5-10% di tutti i casi di omocistinuria. Allo stato omozigote, (1 caso ogni 3.355.000 nati) la gravità del difetto enzimatico e delle manifestazioni cliniche mutano considerevolmente nelle diverse famiglie. I reperti clinici variano da apnea, convulsioni, microcefalia, coma e decesso, a ritardo dello sviluppo, atassia, anomalie motorie e disturbi psichiatrici. Sono stati segnalati casi di malattia vascolare prematura o neuropatia periferica come uniche manifestazioni del deficit enzimatico. Malgrado la gravità del deficit, gli adulti possono risultare del tutto asintomatici. L'esposizione all'ossido nitrico utilizzato come anestetico (che inibisce la metionina sintasi) in pazienti con deficit di MTHFR può causare deterioramento neurologico, con conseguente decesso del paziente.

Il deficit parziale interessa lo 0,5% della popolazione generale e può esprimersi con un quadro clinico molto variabile.

Inoltre recentemente sono stati individuati un certo numero di polimorfismi del gene MTHR fra cui la mutazione 677C > T e 1298A > C che possono incidere sui livelli plasmatici di omocisteina e sono stati presi in considerazione come potenziali fattori di rischio per un'ampia gamma di patologie, tra cui i difetti congeniti, la malattia vascolare, il cancro, il morbo di Alzheimer. Tali polimorfismi sembrano influire anche sulla probabilità di sopravvivenza in caso di leucemia. I dati più attendibili attualmente a disposizione suggeriscono un ruolo del polimorfismo 677C >T come fattore di rischio nei difetti del tubo neurale. Benché il test clinico per il polimorfismo sia disponibile, il suo valore predittivo nei singoli individui non è ancora stato determinato[38-41]

Il quadro ematologico è caratterizzato da bassi livelli di metionina, ai limiti inferiori della norma e un moderato aumento della omocisteina nel sangue e nelle urine. Questo dato differenzia la malattia dall'omocistinuria classica dovuta a deficit di

cistationina β-sintetasi mentre l'assenza di anemia megaloblastica distingue questa condizione dall' omocistinuria causata dalla formazione di metilcobalamina.

La conferma diagnostica si ottiene però solo con il dosaggio dell'enzima su biopsia epatica, fibroblasti coltivati e leucociti. La diagnosi prenatale è realizzata misurando l'attività dell'enzima MTHFR in colture di villi coriali o amniociti, mediante analisi di linkage delle famiglie informative o analisi del DNA della mutazione.

I trattamenti proposti sono un'associazione di folato, nella forma di 5-metiltetraidrofolato ove disponibile, cioè nella forma ridotta utilizzata dall'organismo, vitamina B6 e vitamina B12 con supplementazione di metionina e betaina, e il trattamento precoce con betaina che sembra dare i migliori risultati [39-40].

Scopo del progetto

Con questo lavoro , si è voluto valutare il ruolo della supplementazione di vitamine nelle patologie umane associate a disturbi metabolici dell'omocisteina.

L' Omocisteina è un aminoacido che si forma nel corpo a partire dalla metionina e il metabolismo dell'omocisteina è regolata da vitamine B6, B12, acido folico e betaina. I fattori determinanti l'iperomocisteinemia sono: genetici, dieta, disturbi renali e altre condizioni patologiche.

Il sequenziamento del DNA tramite NGS ha cambiato l'approccio al percorso diagnostico delle malattie genetiche e sta entrando anche in molti altri settori della medicina. Con la NGS, infatti, è possibile effettuare un'indagine genetica andando a sequenziare, per finalità di ricerca, l'intero genoma, oppure solo la porzione codificante (l'esoma) in quei pazienti con malattia genetica che rimangono senza una diagnosi, così da arrivare ad identificare il difetto molecolare alla base della loro patologia.

Il presente lavoro di tesi descrive 4 soggetti italiani con iperomocisteina e combinando i dati clinici con i risultati del sequenziamento si sono identificate le variante geniche potenzialmente causative.

La screening genetico ha permesso una corretta diagnosi e terapia con integratori vitaminici del complesso della vitamina B al fine di correggere le anomalie metaboliche dell'omocisteina e i conseguenti aspetti clinici.

Casi clinici

Presso il laboratorio di Genetica Medica dell'azienda Ospedaliero Universitaria Ospedali Riuniti di Foggia sono giunti due fratelli (C. C. e C. G.) rispettivamente di 26 e 24 anni con un fenotipo clinico tipico dell'iperomocisteinemia plasmatica.

C.C. infatti presentava la dislocazione bilaterale del cristallino e un habitus marfanoide mentre C.G. ha una storia anamnestica caratterizzata da un ictus ischemico giovanile e un habitus marfanoide.

Successivamente si è studiato un uomo (T.M.) di 29 anni con iperomocisteinemia plasmatica e trombosi venosa superficiale giovanile.

Oggetto di studio è stata anche una donna (B. L.) di 34 anni con diagnosi di proteinuria (presenza di proteine nelle urine) e microematuria (presenza di tracce di sangue nelle urine). Questa diagnosi è stata effettuata dopo essersi rivolta a diversi centri specialistici in Italia attraverso la valutazione di biopsie renali, la cui analisi ha portato alla conclusione che la donna fosse affetta da una forma di nefropatia non meglio identificata (glomerulonefrite mesangioproliferativa con immunofluorescenza negativa).

Trattata con cortisonici, ha mostrato una parziale remissione della sintomatologia renale.

In seguito ad una interruzione di gravidanza nelle ultime settimane di gestazione per cause sconosciute, si reca dal medico ginecologo che le prescrive una serie di analisi dalle quali si riscontrano elevati livelli di omocistinemia (112 μmol/L; v.n.: <11.1) con omocistinuria e normali livellli plasmatici e urinari di metionina. I risultati delle analisi inducono, quindi a sospettare che la donna fosse affetta da una malattia metabolica.

La sintomatologia della paziente, però, non offriva la possibilità di poter effettuare una diagnosi certa di Aciduria Metilmalonica con Omocistinuria in quanto la signora non era affetta da nessuna forma di ritardo mentale, caratteristica peculiare della patologia sopracitata. È bene precisare che la donna ha conseguito la laurea magistrale in ingegneria.

Relativamente ai primi tre casi clinici (C.C., C.G. e T.M.), avendo un fenotipo clinico tipico dell'ipeomocisteinemia, si è confermata la diagnosi attraverso l'indagine genetica.

Mentre, il caso B.L. passa al medico genetista il quale, di fronte all'impossibilità di formulare una diagnosi con i dati a disposizione, decide di ricorrere alle tecnologie di

sequenziamento di nuova generazione per studiare l'intero esoma al fine di trovare mutazioni nei geni coinvolti nel metabolismo dell'omocisteina.

Materiali e metodi

Analisi dell'esoma mediante NGS

Il metodo ad oggi più efficiente ed efficace per stabilire la mutazione causativa di una patologia mendeliana è sicuramente il sequenziamento tramite NGS dell'esoma. L'esoma costituisce circa l'1% del genoma (circa 30 milioni di basi) ed è costituito dall'insieme di tutte le sequenze esoniche (circa 200.000) che ne costituiscono la parte codificante. Il fatto che il sequenziamento sia svolto solo su tale regione facilita enormemente la successiva analisi informatica escludendo tutta quell'enorme parte di DNA genomico che al momento non sembra avere rilevanza clinica. Si stima infatti che più del 90% delle patologie genetiche siano dovute a mutazioni puntiformi nelle regioni codificanti. All'atto pratico, per evitare di analizzare la parte di DNA non codificante al fine di mantenere un alto numero di letture e un'elevata copertura delle zone di interesse, è necessario prima di procedere al sequenziamento vero e proprio un'ulteriore fase di "arricchimento" o "cattura". È quindi necessario indirizzare il sequenziamento verso specifici frammenti di DNA costruendo una libreria arricchita o costituita esclusivamente dalle regioni di interesse: nel caso dell'exome sequencing sono i circa 200.000 frammenti accennati prima, ma possono anche essere in numero minore qualora si vogliano analizzare solo alcuni geni di interesse.

L'arricchimento procede prima con la denaturazione dei frammenti totali di DNA a cui successivamente vengono aggiunte di sonde biotinilate che si ibridizzano solo ai frammenti di interesse. È così possibile "catturare" tali frammenti tramite delle biglie coniugate con streptavidina ed eluire in ultimo le sequenze di interesse scartando tutto il resto.

L'analisi prevede i seguenti passaggi:

- Estrazione del DNA genomico con successiva determinazione quantitativa del DNA purificato mediante metodo fluorimetrico;
- Sequenziamento dei campioni amplificati e dosati mediante l'impiego di piattaforme di sequenziamento di nuova generazione.

<u>ESTRAZIONE E DOSAGGIO</u>

Il DNA della paziente è stato estratto da campioni di sangue periferico e dosato mediante un metodo fluorimetrico, QUBIT (Life technologies, Invitrogen).

Il prelievo è stato eseguito presso il Laboratorio di Genetica Medica degli OO.RR di Foggia.

Il protocollo di estrazione del kit QIAamp DNA Blood Mini Kit (Qiagen) (Figura14) utilizzato per la purificazione del DNA genomico del campione di sangue verrà descritto nel dettaglio nel paragrafo successivo.

Il saggio di quantificazione al QUBIT è basato sulla rilevazione della fluorescenza emessa da uno specifico fluorocromo che si lega in modo aspecifico al dsDNA generando una fluorescenza rilevabile attraverso il fluorimetro Qubit® (life technologies™, Invitrogen). Il kit che è stato utilizzato è il Qubit® dsDNA BR Assay con cui è possibile rilevare la quantità di DNA genomico estratto.

Questo strumento permette di effettuare misure di fluorescenza tramite l'utilizzo di fluorofori, ovvero molecole contenenti gruppi flurescenti in grado di legare il DNA a doppia elica (dsDNA). La fluorescenza è un fenomeno di emissione in cui si misura la trasmissione energetica all'interno della molecola da analizzare, da un livello energetico più alto ad uno inferiore, rilevando la radiazione emessa piuttosto che quella assorbita. Affinché vi sia fluorescenza la molecola deve essere "eccitata" mediante l'assorbimento di luce di lunghezza d'onda minore (cioè a maggiore energia) rispetto a quella emessa (a minore energia).

Il principale vantaggio dell'utilizzo della fluorescenza rispetto ad altre misure di assorbimento è la maggiore sensibilità e accuratezza in quanto il segnale di fluorescenza parte da un background pari a zero.

Inizialmente è stata preparata la soluzione contenente l'intercalante fluorescente (Qubit® dsDNA BR reagent, 1µl per ogni campione) e il Qubit® dsDNA BR buffer (199µl per ogni campione). Da ogni campione di DNA genomico precedentemente estratto sono stati prelevati 5µl e posti nelle Qubit® assay tubes, microtubi forniti con il kit in grado di far passare il fascio luminoso del fluorimetro e permettere in questo modo al fluoroforo legato al DNA di eccitarsi ed emettere la luminescenza nel caso di ibridazione.

I campioni sono stati preparati nel modo seguente:

- 5µl di campione contenente il DNA genomico estratto + 195µl di soluzione contenente il buffer e l'intercalante;
- 10µl di Qubit® dsDNA BR standard #1 (0ng/µl) + 190µl di soluzione contenente buffer e l'intercalante;
- 10µl di Qubit® dsDNA BR standard #2 (100ng/µl) + 190µl di soluzione contenente buffer e l'intercalante.

Il fluorimetro utilizzato nella presente analisi contiene una sorgente d'eccitazione costituita da due LED (Light Emitting Diode), una cella per il campione e un rilevatore costituito da fotodiodi in grado di misurare lunghezze d'onda emesse che vanno da 300 a 1000 nm. Alla soluzione di DNA estratto viene aggiunta una soluzione (Quant-iTTM Working solution) contenente la sonda fluorescente in modo tale da ottenere, per ciascun campione da analizzare, un volume finale pari a 200ml. Le molecole di DNA estratto, legate alla sonda fluorescente, vengono eccitate da raggi UV e la luce proveniente dalla sorgente d'eccitazione passa attraverso un monocromatore. La fluorescenza viene dispersa da un'altro monocromatore e rivelata dai fotodiodi.

L'intensità della fluorescenza è linearmente proporzionale alla concentrazione dell'analita.

Per determinare la concentrazione dalla fluorescenza emessa da ogni campione analizzato occorre effettuare la calibrazione dello strumento con due standard forniti dalla casa produttrice in modo tale da ottenere una curva di lavoro.

La quantificazione del DNA estratto ha permesso di 'normalizzare' ogni campione ad una concentrazione standard di utilizzo pari a 10 ng/ml, necessaria per attuare le successive analisi.

SEQUENZIAMENTO DELL'ESOMA

L'analisi NGS, a partire dai campioni di DNA estratti e dosati, è stata effettuata presso GATC Biotech, un'azienda tedesca leader in Europa nel campo del sequenziamento del DNA e della bioinformatica per l'industria e la ricerca accademica. Lo studio è stato condotto mediante l'utilizzo della piattaforma di nuova generazione "Illumina HiSeq 2500 system".

Per effettuare l'analisi NGS sono stati utilizzati 500ng di DNA. Successivamente la procedura sperimentale prevede l'ibridazione di oligonucleotidi specifici per le regioni d'interesse ai campioni di DNA e rimozione degli oligonucleotidi non ibridati. Mediante una reazione di PCR, effettuata per mezzo di primers che aggiungono al campione sequenze indice e sequenze adattatrici necessarie per la generazione dei cluster, vengono amplificati i complessi contenenti le regioni genomiche d'interesse. I prodotti di amplificazione vengono, poi, purificati e la library così ottenuta viene normalizzata e caricata sul HiSeq dell'Illumina.

Concluso il sequenziamento, lo strumento fornisce una serie di informazioni come Q-score, cluster density, cluster passing filter riportate tutte in un foglio excel inviato successivamente al nostro laboratorio per essere analizzato.

Il dispositivo dell'illumina genera files in formato .BAM (in cui le reads sono comparate al genoma umano di riferimento (Hg19) e in formato .vcf che permettono di effettuare la procedura di allineamento per l'identificazione e l'analisi delle varianti. Quest'analisi è stata effettuata mediante il software Illumina VariantStudio V.2.2 che descrive accuratamente ogni variante in termini di localizzazione, tipologia, eventuali conseguenze generate a livello proteico dalla mutazione e una predizione degli effetti funzionali che questa mutazione può causare nell'individuo che ne è dotato.

Il software ha anche la capacità di filtrare le varianti di interesse in modo da identificare varianti presumibilmente patogenetiche selezionandole da un background comprendente polimorfismi o errori di sequenziamento.

Per capire se le varianti individuate mediante il sequenziamento dell'esoma della paziente attraverso l'utilizzo della piattaforma HiSeq siano patogene o meno si può ricorrere alla cosiddetta analisi in silico, ossia all'analisi tramite l'utilizzo di programmi computerizzati (software). I software più comunemente utilizzati per l'analisi in silico delle mutazioni missenso sono Poly-Phen-2, SIFT, Mutation Tastere Align-GVGD. Poly-Phen-2 e Mutation Taster sono forse i più usati, poichè mostrano elevati livelli di sensibilità e specificità (questi software sembrano cioè predire il vero effetto della mutazione con un buon livello di approssimazione).

Questo sistema di filtraggio permette di ridurre notevolmente il numero di geni potenzialmente patogeni coinvolti dalle 50.000 varianti riscontrate mediante questo tipo di analisi. In questa fase si cercano di prendere in considerazione parametri multipli: risultato delle predizioni in silico, possibile influsso del contesto genomico circostante, eventuali effetti modulatori sul fenotipo pur in assenza di un effetto patogeno diretto e, se esistenti, dati già pubblicati in letteratura.

Nello specifico, avendo a disposizione sia i dati clinici relativi all'anamnesi della paziente, che presentava iperomocistinemia e valori di acido metilmalonico aumentati, che il risultato del sequenziamento è stato possibile combinare queste due tipologie di dati per la formulazione di una prima ipotesi di gene coinvolto nella patologia non ancora ben diagnosticata, procedendo ad analizzare e a verificare la

presenza di mutazioni solo nella una famiglia di geni coinvolta nel metabolismo dell'omocisteina fino ad identificare la mutazione che spiega il fenotipo.
Da un punto di vista pratico la cosa più importante da sottolineare è che la maggior parte delle varianti rilevabili sono semplici polimorfismi senza alcun significato patogeno.

Strategia analitica per l'analisi molecolare (metodo Sanger)

La strategia prevede i seguenti passaggi:

- Estrazione del DNA genomico con successiva determinazione quantitativa del DNA purificato mediante spettrofotometro;
- Reazione di amplificazione del DNA mediante Polymerase Chain Reaction (PCR) ai fini di amplificare in vitro la regione codificante completa e parte della regione intronica per ciascun esone del gene con purificazione degli amplificati su sistema a vuoto MILLIPORE;
- Reazione di sequenziamento dell'amplificato, precedentemente purificato, con Big-Dye X-Terminator (Applied Byosistem) mediante la tecnica di terminazione della sequenza; purificazione dei prodotti mediante kit di purificazione BigDye ®X-Terminator™ Purification Kit per la rimozione dei BigDye Terminators non incorporati e di sali, successiva elettroforesi capillare dei frammenti utilizzando il sequenziamento automatico a tecnologia fluorescente ABI PRISM 3100 Genetic Analyzer (Applied Biosystem);

ESTRAZIONE DNA GENOMICO

Il materiale biologico di partenza di cui ci siamo serviti è stato un prelievo di sangue periferico, prelevato dal soggetto analizzato, contenente sodio citrato 3.8% in quanto questo tipo di anticoagulante non interferisce con i kit a disposizione per la fase di purificazione degli acidi nucleici né con la successiva procedura di amplificazione.
Il prelievo è stato eseguito presso il Laboratorio di Genetica Medica degli OO.RR di Foggia.
Il protocollo di estrazione del kit QIAamp DNA Blood Mini Kit (Qiagen) (Figura14) utilizzato per la purificazione del DNA genomico del campione di sangue è il seguente:

- preparare un'aliquota di sangue di 200 µl;

- aggiungere 30 µl di proteinasi K fornita dal kit per la degradazione delle proteine;
- aggiungere 200 µl di Buffer AL (lysis Buffer);
- vortexare per 15 sec e incubare per 10 minuti a 56 °C;
- centrifugare a 8000 g per 1 minuto in modo da eliminare il materiale adeso alle pareti del tubo;
- aggiungere 200 µl di etanolo puro per la solubilizzazione dei lipidi e precipitazione del DNA, vortexare per 15 sec e centrifugare per 1 minuto a 8000g;
- aspirare il materiale e inserirlo nelle colonnine fornite dal kit composte da una colonna con membrana in silice che ha proprietà di ritenzione del DNA, e da un tubo di raccolta del filtrato;
- centrifugare a 8000 g per 1 minuto, eliminare il tubo con il materiale filtrato dalla colonnina ed inserirla in un tubo nuovo;
- aggiungere 500 µl di Buffer AW1 di lavaggio, centrifugare per 8000 g per 1 minuto e gettare il tubo con il filtrato;
- aggiungere 500 µl di Buffer AW2 di lavaggio, centrifugare per 13000 g per 4 minuti e gettare il tubo con il filtrato;
- ripetere il precedente punto, centrifugando per 1 minuto senza aggiungere, però, il Buffer AW2, per eliminare eventuali residui di Buffer presenti nella membrana della colonnina;
- aggiungere dai 150 ai 200 µl di Eluition Buffer AE, attendere 1 minuto a temperatura ambiente e centrifugare a 8000g per 1 minuto;
- raccogliere il materiale eluito contenente DNA e conservare a -20°C.
- mettere la figura: fasi del processo di purificazione di DNA genomico da sangue intero secondo il protocollo fornito dal kit QIAamp DNA Blood Mini Kit (Qiagen).

Tale metodica ha permesso di ottenere fino a 12 µg di DNA a partire da un volume iniziale di 200 µl di sangue intero.

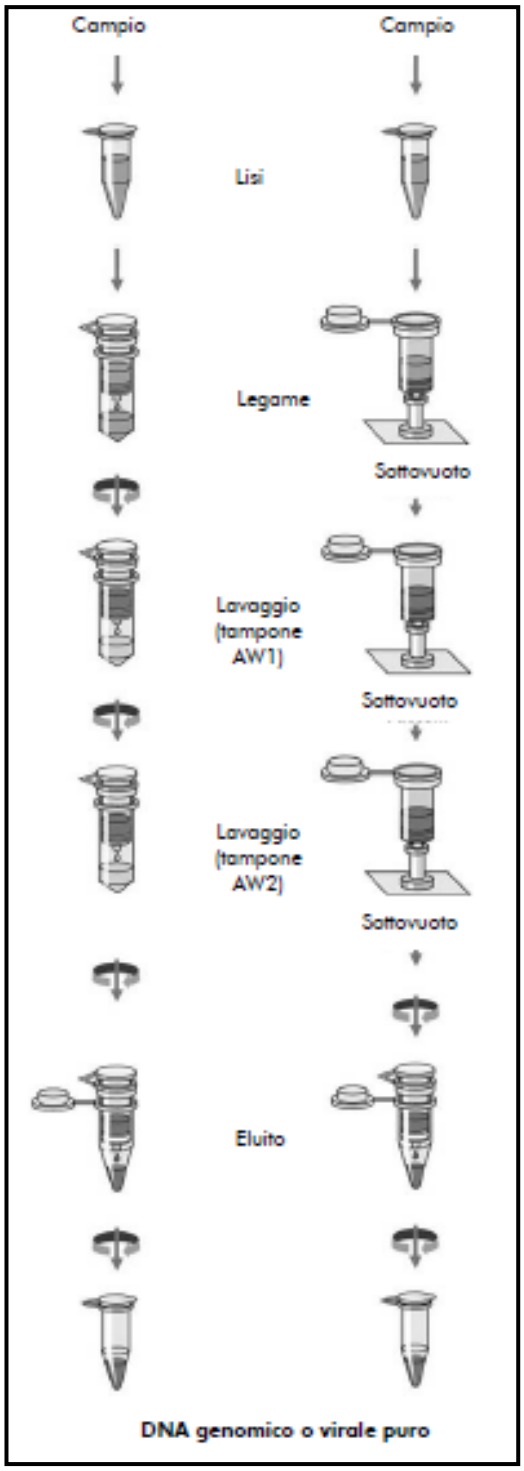

Figura 14: Schema di estrazione di DNA.

DETERMINAZIONE DELLA CONCENTRAZIONE DEL DNA

Dopo l'estrazione del DNA, i campioni vengono sottoposti ad un'analisi spettrofotometrica per calcolare la concentrazione e la purezza del DNA presente.

Sono state effettuate 3 letture allo spettrofotometro perché è fondamentale accertarsi della purezza del DNA:

- Lettura a 260 nm per gli acidi nucleici
- Lettura a 280 nm per le proteine
- Lettura a 230 nm per eventuali residui

Il rapporto tra l'assorbanza a 260 nm e quella a 280 nm è indice di purezza del DNA quando è compreso tra 1.7 e 2.0; rapporti inferiori a 1.7 indicano una eccessiva presenza di proteine contaminanti, mentre rapporti maggiori a 2.0 evidenziano la degradazione del DNA.

L'assorbanza a 230nm riflette la contaminazione del campione dovuta a sostanze come carboidrati, fenoli, peptidi o composti aromatici. per campioni puri il rapporto tra l'assorbanza a 260 nm e quella a 230 nm dovrebbe essere di circa 2,2.

La concentrazione del DNA ottenuto è stata poi misurata sempre tramite lo spettrofotometro, che permette di rilevare l'assorbimento della radiazione a 260 nm direttamente proporzionale alla quantità di DNA in sospensione, dopo aver opportunamente diluito il campione (1 µl di DNA + 99 µl di acqua sterile) e dopo averlo posto in una cuvetta di quarzo per la lettura.

La lettura allo spettrofotometro si basa sulla legge di Lambert-Beer ($A = \varepsilon\ c\ l$), una relazione empirica che correla la quantità di luce assorbita (A) da un mezzo alla natura chimica (ε), alla concentrazione (c) ed allo spessore (l) del mezzo attraversato. Sapendo che una soluzione di DNA a doppio filamento, di concentrazione pari a 50µg/ml ha un assorbimento pari a 1 a 260 nm, la concentrazione del DNA sarà calcolata mediante la seguente proporzione:

$$C(\mu g/ml) : A_{260} = 50\ \mu g/ml : 1 \text{ unità di assorbimento}$$

$$\downarrow$$

$$C(\mu g/ml) = A_{260} \times \frac{50\ \mu g/ml}{1 \text{ unità di assrobimento}}$$

Per risalire alla concentrazione iniziale della soluzione bisogna moltiplicare il valore ottenuto per il fattore di diluizione (nel caso specifico 100).

REAZIONE DI AMPLIFICAZIONE (PCR)

I campioni di DNA sono stati sottoposti ad amplificazione genica relativa al gene MMACHC al fine di individuare le mutazioni presenti.

La reazione a catena della polimerasi (PCR, Polimerase Chain Reaction) è una tecnica di amplificazione in vitro che consente la sintesi esponenziale di un frammento di DNA a partire da uno stampo di DNA. L'amplificazione avviene mediante cicli ripetuti di polimerizzazione che utilizzano come stampo i due filamenti complementari di DNA.

Per la reazione fungono da innesco due oligonucleotidi sintetici (primer) che delimitano la regione da amplificare. I cicli di amplificazione avvengono in condizioni controllate di tempo e di temperatura e possono essere schematizzati nella ripetizione di tre fasi fondamentali: denaturazione, appaiamento (annealing) dei primer e polimerizzazione.

In una prima fase il DNA è incubato a 98 °C in modo da causare la denaturazione dei due filamenti. In questo modo si rendono disponibili i siti d'ibridazione per i primers. La seconda fase permette "l'annealing" o ibridazione dei primers alla sequenza target e si realizza portando la miscela di reazione ad una temperatura inferiore rispetto a quella di denaturazione, detta Tm (melting temperature), che dipende dalla composizione in basi dell'oligonucleotide (Figura 15).

La Tm può essere calcolata in base alla formula: **2(A+T) + 4(G+C) - 1(CT) + 1(AG)**

Dove:

- (A+T) indica la somma del numero di adenine e timine presenti nella sequenza del primer;
- (G+C) indica la somma del numero di guanina e citosina presenti nella sequenza del primer;
- (CT) indica la coppia adiacente di pirimidine;
- (AG) indica la coppia adiacente di purine.

Tale temperatura può influenzare la specificità della reazione, nel senso che temperature di "annealing" basse favoriscono la formazione di aspecifici, mentre temperature molto alte rendono selettiva la reazione di ibridazione primer-stampo. Per calcolare la temperatura di annealing è necessario conoscere il contenuto in GC e AT dei primer utilizzati per la reazione e applicare questa formula: **4(C+G)+2(A+T)** sia per il forward che per il reverse. Tra i due risultati si sceglie la temperatura più bassa diminuita di un grado.

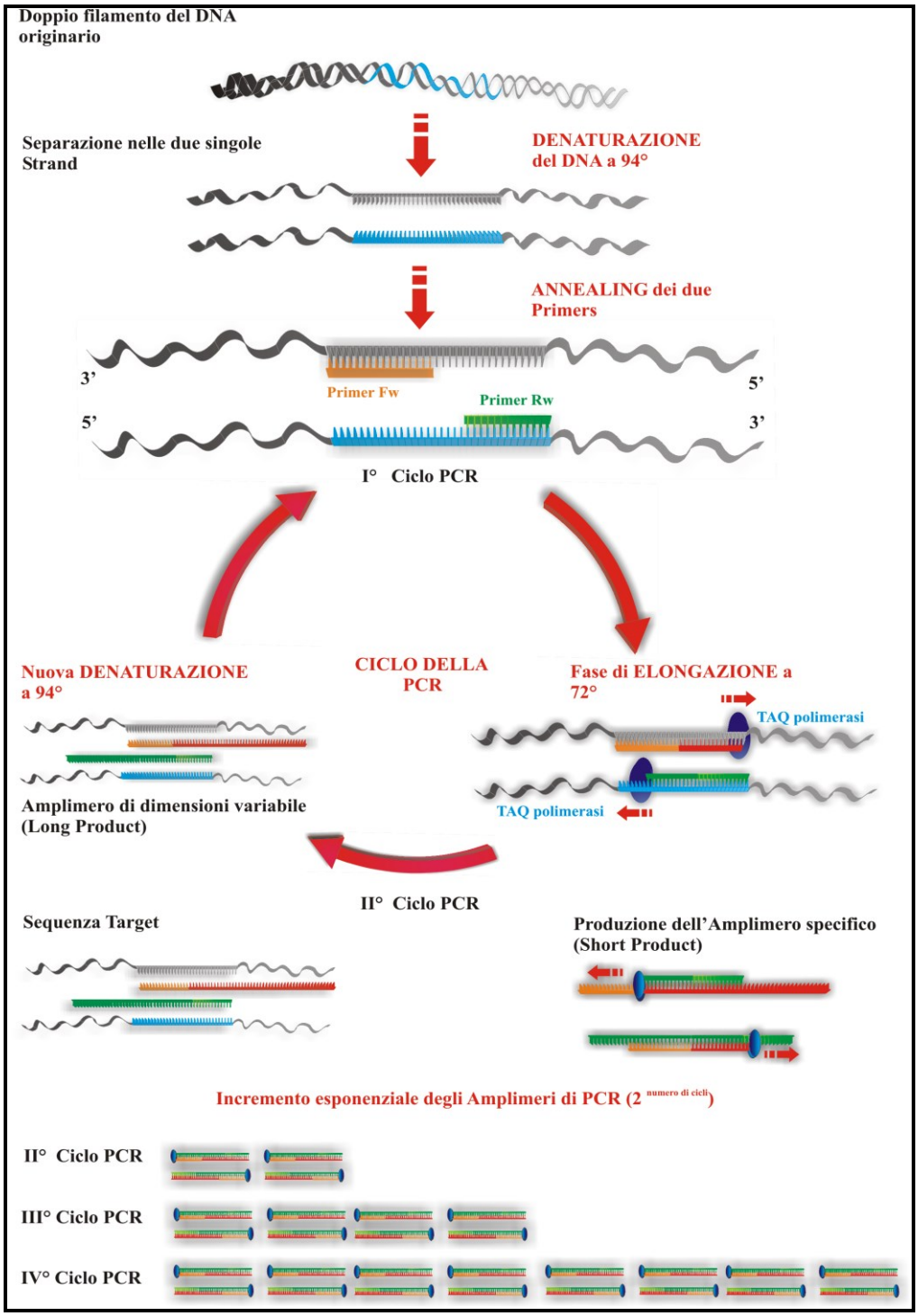

Figura 15: Rappresentazione schematica delle fasi relative alla reazione di PCR

Nel caso specifico di questo studio sono stati utilizzati i seguenti primer:

- MMACHC Ex3Fw: 5'-TCA TGT TTT CCC TTC TGA GGA-3'
- MMACHC Ex3Rw: 5'-CAA AGC TAA TTT GTT CTG GGT TG-3'
- MMACHC Ex4Fw: 5'-AGG CCT AGC TTG CAA TGA TG-3'
- MMACHC Ex4Rw: 5'-GAA GGC AGA TGG GAA TTC TG-3'

I primers sono stati disegnati utilizzando il programma “Primer 3” disponibile in rete. La scelta di una corretta coppia di primer viene effettuata basandosi su informazioni fornite dal programma circa la lunghezza e il contenuto in GC. È importante che sia il primer Forward che il Reverse abbiano una lunghezze di circa 20pb e il contenuto in GC sia di circa il 50% affinché si abbiano buoni risultati di amplificazione.

La terza fase è quella di elongazione, in cui si ha la replicazione del frammento in presenza dei deossinucleotiditrifosfato (dNTP) e della DNA polimerasi che determina l'allungamento dei primer in direzione 5'→3'. Il prodotto che si ottiene dopo il primo ciclo è di lunghezza indeterminata ed è definito “long product”. Nei cicli successivi i prodotti che si accumulano derivano dall'amplificazione di frammenti a loro volta amplificati nei cicli precedenti e non dalla sequenza stampo. Tali prodotti sono frammenti di DNA che terminano con le sequenze specifiche dei primer utilizzati e sono definiti “short-product”. Nella fase d’allungamento è utilizzata una Taq polimerasi termostabile che questo permette all’enzima di resistere alle alte temperature imposte dalla denaturazione con il vantaggio di non doverlo aggiungere ad ogni ciclo. Inoltre, tale enzima conserva un'attività esonucleasica 5'→3' mentre manca di quella 3'→5'.

Il processo d’amplificazione è reso automatico mediante l’uso di un thermal cycler (Figura 16) la cui temperatura può essere variata secondo le necessità e per i tempi opportuni.

Figura 16:Esempio di termociclatore

Nella pratica di laboratorio è necessario preparare 48µl di mix di reazione a cui si aggiungono 2µl del campione di DNA per un volume finale di 50µl. La mix di reazione contiene i seguenti reagenti:

	Concentrazioni	Volumi (µl)
Buffer	10X	5
$MgCl_2$	15µM	2,5
dNTPs	5mM	2
Primer Fw	100 µM	0,2
Primer Rw	100 µM	0,2
Taq	5U/ µl	0,3
H_2O		37,8

Tabella 2: Reagenti per la rezione di amplificazione.

Successivamente, al termociclatore vengono impostate le condizioni per l'amplificazione secondo il seguente schema:

	Temperatura	*Tempo*	
Denaturazione	95 °C	3 minuti	
Denaturazione	95 °C	30 secondi	x 35 volte
Annealing	57 °C	30 secondi	
Estensione	72 °C	45 secondi	
Elongazione	72 °C	3 minuti	

ANALISI ELETTROFORETICA SU GEL DI AGAROSIO

Dopo la PCR, la presenza del prodotto di amplificazione viene verificata attraverso una elettroforesi in gel di agarosio, una tecnica classicamente utilizzata per analizzare e separare acidi nucleici. Questa tecnica sfrutta le cariche presenti nelle molecole di DNA o RNA (caricate negativamente) per farle migrare, in un campo elettrico, attraverso un gel di agarosio che funge da setaccio, essendo costituito da una rete di pori, i quali consentono di separare le molecole in base alla loro grandezza.

È stato preparato un gel al 2% in un V_f di 200ml di tampone TAE 1X, preparato a partire da una soluzione madre 50X. Per calcolare il volume da prelevare si fa uso della seguente formula: **C.i. x V.i. = C.f. x V.f.**

Pertanto: $[TAE]_i = 50X$; $[TAE]_f = 1X$; V.f. = 2 litri

$$V.i. = (C.f. \times V.f.)/C.i. = 40 \text{ ml}$$

Il TAE è composto da una soluzione di Tris-acetato, generalmente a pH 8.0, e EDTA che sequestra i cationi bivalenti. Il buffer TAE ha un potere tamponante inferiore rispetto al buffer TBE (Tris-Borato-EDTA) ma è importante sottolineare che il DNA a doppio filamento corre più velocemente nel TAE. Il TBE, invece, è particolarmente indicato per la separazione di piccoli frammenti di DNA (PM<1000) ad esempio piccoli prodotti risultanti dalla digestione con enzimi di restrizione. Generalmente più costoso del TAE, il TBE ed inibisce la DNA ligasi che può causare problemi nel caso si volesse procedere con la purificazione o la ligazione del DNA.

Per la preparazione del gel di agarosio occorre pesare 4g di agarosio, versarlo nella beuta e aggiungere 200ml di TAE 1X. Si porta ad ebollizione per sciogliere l'agarosio. Nel frattempo si accende la cappa a flusso laminare e si prepara il supporto per la formazione del gel. Sciolto l'agarosio si lavora sotto cappa e nella beuta si aggiungono 20µl di etidio bromuro e successivamente il contenuto della beuta viene versato sul supporto per consentirne la solidificazione. Una volta polimerizzato, il gel viene inserito nella cameretta all'interno della quale avviene la corsa elettroforetica, coperto dal tampone TAE 1X. Nel primo pozzetto viene caricato il marker di peso molecolare noto che permetterà di stimare la grandezza dei frammenti amplificati e, a seguire, negli altri pozzetti i campioni di interesse e infine il bianco (controllo negativo, costituito dalla stessa mix di reazione senza il DNA). Prima del caricamento vengono mescolati con il colorante (Gel loading buffer 6X, costituito da blu di bromo fenolo 0,25%, xilenecianolo 0,25% e glicerolo 50%).

Chiusa la cameretta si fa partire la corsa elettroforetica per circa 15 minuti a 160V.

Terminata la corsa, il gel viene visualizzato al trans illuminatore dove grazie alla presenza dell'etidio bromuro è possibile visualizzare gli acidi nucleici. Infatti, l'etidio bromuto è una molecola intercalante, planare, che si si intercala, appunto, tra le basi dell'acido nucleico a doppio filamento emettendo luce fluorescente quando irradiata con luce ultravioletta.

Il confronto tra l'altezza delle bande relative al campione amplificato e l'altezza delle bande del marker permette di valutare la corretta riuscita di amplificazione, l'assenza di amplificazione aspecifica mediante il confronto del peso molecolare dell'amplificato ottenuto con quello atteso; per escludere la presenza di DNA contaminante estraneo, nella corsia del controllo negativo non si deve osservare alcuna banda.

SEQUENZIAMENTO DIRETTO

La procedura di sequenziamento è costituita da 3 step:

1. Prima purificazione: eliminazione dei reagenti contenuti nella miscela di PCR;
2. Reazione di marcatura: ibridazione del DNA con dNTP fluorescenti;
3. Seconda purificazione: eliminazione dei reagenti residui utilizzati nella marcatura per ridurre contaminazioni di fluorescenza in eccesso.

1. I prodotti di PCR vengono sottoposti ad un processo di purificazione mediante l'utilizzo di una piastra Millipore (Montage), composta da 96 pozzetti. La piastra ha sul fondo dei pozzetti una membrana in gel di silice in grado di trattenere i frammenti amplificati e di isolarli dalle impurità di PCR di dimensioni inferiori mediante la creazione di una pressione che consente la filtrazione. L'apparato di purificazione è composto da una pompa da vuoto, da una valvola di controllo della pressione, dal supporto per le piastre, dai tubi di collegamento e da una beuta di raccolta delle impurità (Figura 17).

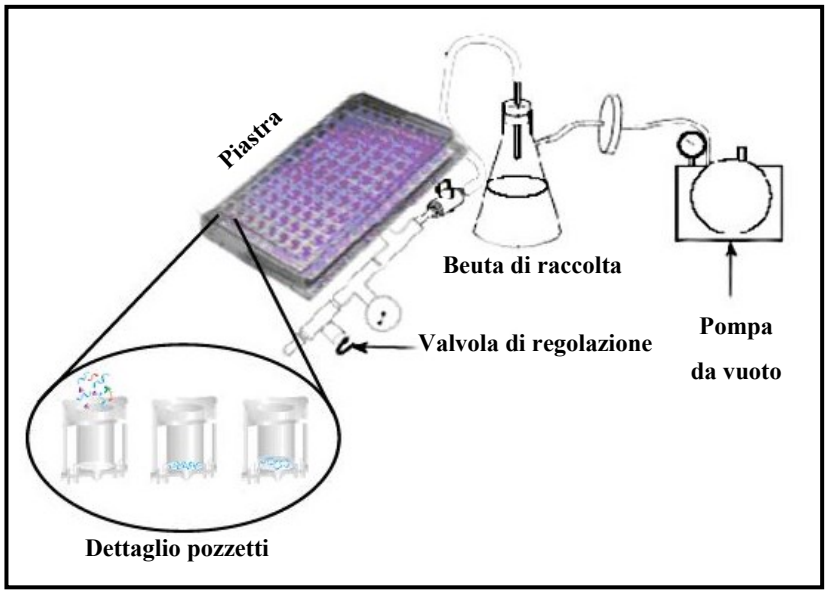

Figura 17: Schema del sistema di purificazione.

Le fasi del protocollo di prima purificazione da noi utilizzato in laboratorio sono le seguenti:

- Portare il volume di PCR a 100 µl con TE 1X (oppure con H_2O milliQ).
- Pipettare il campione all'interno del pozzetto della piastra.

- Aspirare il volume applicando una depressione di 20atm fino a completa disidratazione (~ 8') per permettere il passaggio attraverso le maglie del gel di primer, basi, sali, buffer e dimeri di primer mentre il prodotto di PCR resta nel pozzetto avendo dimensioni minori.
- Far adsorbire su carta l'eventuale condensa creatasi sulla base della piastra.
- Aggiungere 25 µl di TE 1X (oppure H_2O milliQ) all'interno del pozzetto.
- Mettere la piastra ad agitare per 10' a 500 Revs/min sull'apposito Orbital shaker.
- Aspirare il prodotto e metterlo in una provetta da 0,2 µl.

2. Il metodo di sequenziamento automatico si basa sulla tecnica di sequenziamento enzimatico di Sanger messa a punto nel 1977. Questa tecnica prevede l'uso di uno stampo di DNA a singola elica, un solo primer (Fw o Rw), una DNA polimerasi, deossinucleotidi trifosfato (dNTPs) e dideossinucleotidi (ddNTP). Analoghi alle quattro basi del DNA, i dideossi non possiedono il gruppo ossidrilico (OH) nel carbonio in posizione 3' dello scheletro di zucchero legato alla base, necessario per formare il legame con il nucleotide successivo e far procedere la neosintesi della catena di DNA.

La doppia elica di DNA è denaturata per consentire al primer di ibridarsi alla regione da sequenziare. La polimerasi utilizzando il primer come innesco della reazione incorpora i dNTP o i ddNTP complementari, secondo un principio di competizione. Ogni qualvolta l'enzima inserisce il ddNTP, essendo questo privo del gruppo 3' OH, provoca uno stop della reazione di polimerizzazione.

Nelle tecniche che prevedono l'uso di sequenziatori automatici, a questi dideossinucletidi è inoltre legato un diverso colorante fluorescente, ciascuno caratterizzante la base a cui è coniugato.

Il rapporto di concentrazione tra i dideossinucleotidi e i normali nucleotidi è tale che l'enzima DNA polimerasi statisticamente terminerà la catena nascente in tutte le posizioni possibili in cui un dideossint può essere inserito. la DNA polimerasi produce cioè una serie di frammenti che condividono tutti la stessa estremità iniziale, ma ciascuno di essi riuslterà sfasato. al termine del processo di marcatura il prodotto ottenuto è sottoposto ad una ulteriore purificazione.

Nelle tecniche che prevedono l'uso di sequenziatori automatici, a questi dideossinucletidi è inoltre legato un diverso colorante fluorescente. Questa molecola

produce uno spettro d'emissione caratteristico per ogni ddNTP cui è legato ed ha la caratteristica di produrre un segnale omogeneo e con bassissimo rumore di fondo.
La DNA polimerasi usata nella reazione di sequenziamento è la T7 DNA polimerasi ed è caratterizzata da:

- elevata processività, che riduce il distacco precoce della polimerasi dal filamento stampo;
- attività esonucleasica 3'→5' bassa o, nulla necessaria per ridurre la capacità di discriminazione tra il dNTP e i suoi analoghi ddNTP;
- attività esonucleasica 5'→3' bassa o, nulla che evita l'inserimento di nucleotidi all'estremità 5' che deve restare invariata;

In questo caso sono state preparate due mix di reazione per ogni esone contenente ciascuna il primer forward e il primer reverse del corrispettivo esone. La mix di reazione preparata con 2µl di Buffer (10X); 1,6µl di primer (forward e reverse ciascuno in una diversa mix); 1µl di Mix di marcatura; 3,4µl di acqua per raggiungere un volume finale di 8µl a cui vanno aggiunti 2µl di DNA derivante dalla prima purifica.
Questa doppia marcatura del frammento permette di leggere la sequenza in tutta la sua lunghezza e di confermare direttamente la presenza di varianti nella sequenza.
Le provette vengono, dunque, trasferite in un termociclatore su cui vengono impostate le seguenti condizioni per la reazione di marcatura: 98°C per 10 secondi, 98°C per 10 secondi, 60°C per 2 minuti, ripetuto per 30 volte.

3. Al termine del processo di sequenziamento mediante dideossinucleotidi, il prodotto ottenuto è sottoposto ad un'ulteriore purificazione. In questa fase sono stati utilizzati il BigDye ®X-Terminator™ Purification Kit (Applied Biosystem) che a differenza di altri metodi di purificazione richiede solo l'aggiunta di 2 reagenti che possono essere dispensati assieme in un'unica mix di purificazione o sequenzialmente per ogni singolo campione. La purificazione è stata eseguita in conformità al protocollo fornito dal produttore (Tabella 3) utilizzando i seguenti reagenti:

- XTerminator™ Solution per l'eliminazione dei dideossinucleotidi non incorporati e sali liberi:
- SAM™ Solution per aumentare l'efficienza di purificazione.

Volume della reazione di sequenziamento	10µl
SAM solution (µl/campione)	45µl
Big Dye X Terminator Solution	10µl

Tabella 3: Quantità di reagenti di purificazione per campione secondo il protocollo di purificazione BigDye ®X-Terminator™ (Applied Biosystem).

CORSA ELETTROFORETICA CAPILLARE

I purificati così ottenuti sono inseriti in una piastra da 96 pozzetti che viene collocata nel sequenziatore automatico ABI 3100 Avant Automated Capillary DNA Sequencer (Applied Biosystem) che si basa su una corsa elettroforetica capillare in grado di separare ed analizzare automaticamente 4 campioni di DNA marcati con fluorocromi. I capillari sono riempiti con un polimero che ha la funzione di separare i frammenti durante la corsa elettroforetica.

Applicando una piccola differenza di potenziale, una quantità di campione sufficiente per l'analisi, è automaticamente aspirata all'interno dei capillari (Figura 12). I capillari si immergono quindi all'interno della vaschetta contenente il buffer di corsa 1X ottenuto dalla diluizione della soluzione madre 10X contenente acido etilendiamminotetracetico (EDTA). Questo permette la chiusura del circuito e la creazione di una differenza di potenziale pari a 15kV che permette la separazione elettroforetica dei campioni.

Il capillare usato ha una lunghezza di 36 cm, e permette la lettura in modo attendibile di circa 400 bp. Il capillare di silice è rivestito lungo tutta la sua lunghezza da una guaina. A livello della finestra di lettura è presente un'interruzione della guaina che consente il passaggio di una luce laser a ioni di Argon. Il prodotto di marcatura, a livello della finestra di lettura, è colpito dalla luce laser che eccita il fluoroforo. Le emissioni vengono raccolte e analizzate da una fotocamera digitale CCD che riporta la sequenza delle emissioni in un grafico chiamato elettroferogramma caratterizzato da una successione di picchi di 4 colori diversi (adenina in verde, citosina in blu, guanina in nero e timina in rosso), corrispondenti alle emissioni fluorescenti dei diversi marcatori.

L'identificazione di una mutazione è ritenuta affidabile quando l'elettroferogramma è composto da picchi alti e risolti con un rumore di fondo basso.

Le sequenze ottenute sono analizzate mediante il software "Sequencher" che permette l'allineamento multiplo delle sequenze aumentando sensibilmente la velocità di analisi e consentendo un confronto diretto tra le sequenze dei diversi campioni in esame.

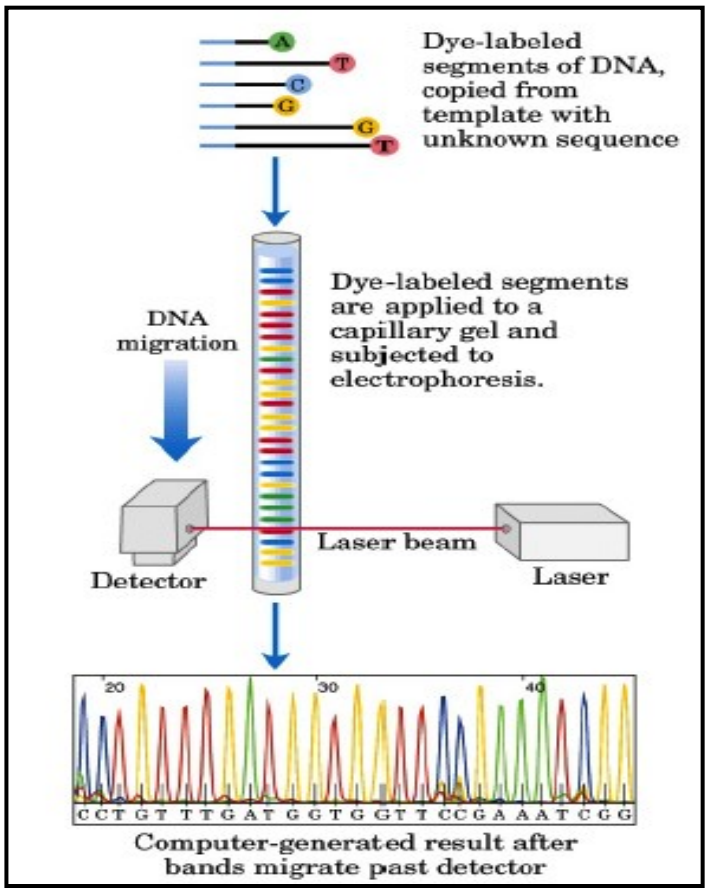

Figura 18: Schema dell'elettroforesi capillare alla base del sequenziamento automatico e risultante elettroferogramma.

Risultati e Discussione

I dati che normalmente risultano dal sequenziamento di un intero esoma mostrano un altissimo numero di mutazioni, inserzioni e delezioni. La stragrande maggioranza di queste alterazioni non ha alcun significato clinico ed è quindi necessario filtrarle secondo criteri stabiliti analizzando la popolazione generale in modo da escludere tutti i polimorfismi noti, insieme a ovviamente tutte le variazioni che non alterano la sequenza amminoacidica, che hanno minime conseguenze sulla struttura proteica, che non rispecchiano la via di trasmissione ereditaria della patologia o che non segregano con la patologia nelle famiglie affette.

Pertanto, dapprima sono stati esclusi dai possibili candidati i geni chiaramente non implicati nelle patologie costituite da difetti del metabolismo e sicuramente non compromettenti. In seguito, sono stati eliminati tutti quei geni che presentavano un fortissimo grado di mutazione nella popolazione generale, il cui significato clinico può quindi essere escluso.

Combinando i dati clinici relativi all'anamnesi con il report risultante dal sequenziamento dell'esoma della paziente, si sono individuati i geni e le mutazioni, tutte precedentemente descritte, potenzialmente causative il fenotipo clinico e biochimico dei soggetti studiati.

Per i primi 3 soggetti studiati (C.C., C.G. e T.M.) il gene CBS, che codifica per la cistationina β-sintasi, è parso essere quello coinvolto nell'alterazione dei livelli plasmatici dell'omocisteina.

Il saquenziamento di Sanger ha confermato la presenza di mutazioni in eterozigosi, infatti i fratelli C.C. e C.G. sono risultati essere portatori di due mutazioni in eterozigote con conseguente cambio degli amminoacidi in posizione "p.Ala114Val" e "p.Arg336His".

Mentre il soggetto T.M. è risultato essere portatore, sempre a livello del gene CBS, di due mutazioni in eterozigote con alterazione degli amminoacidi in posizione "p.Thr262Met" e "p.Ile278Thr".

Per questi tre soggetti, il trattamento con supplementi di vitamina B6, B12 e acido folico hanno consentito una normalizzazione del fenotipo biochimico legato ai livelli elevati di omocisteina plasmatica con conseguente miglioramento del fenotipo clinico.

Lo studio del soggetto B.L. ha evidenziato, come candidato ideale, essere il gene MMACHC gene responsabile della Metilmalonico Aciduria di tipo cblC[43].

Il sequenziamento Sanger ha confermato la presenza di mutazioni in eterozigosi a livello del gene MMACHC. In particolare, dall'analisi dell'elettroferogramma" le due mutazioni trovate sono "p.Tyr130His" e "p.Tyr222Stop".

La prima indica una variazione amminoacidica, relativa al codone 388, che vede una sostituzione del nucleotide T con una C, determinando la formazione di una variante missenso del codone wild type TAC nel codone CAC causando un cambiamento nella traduzione che risulta nella comparsa di una istidina al posto della tirosina (Figura 19).

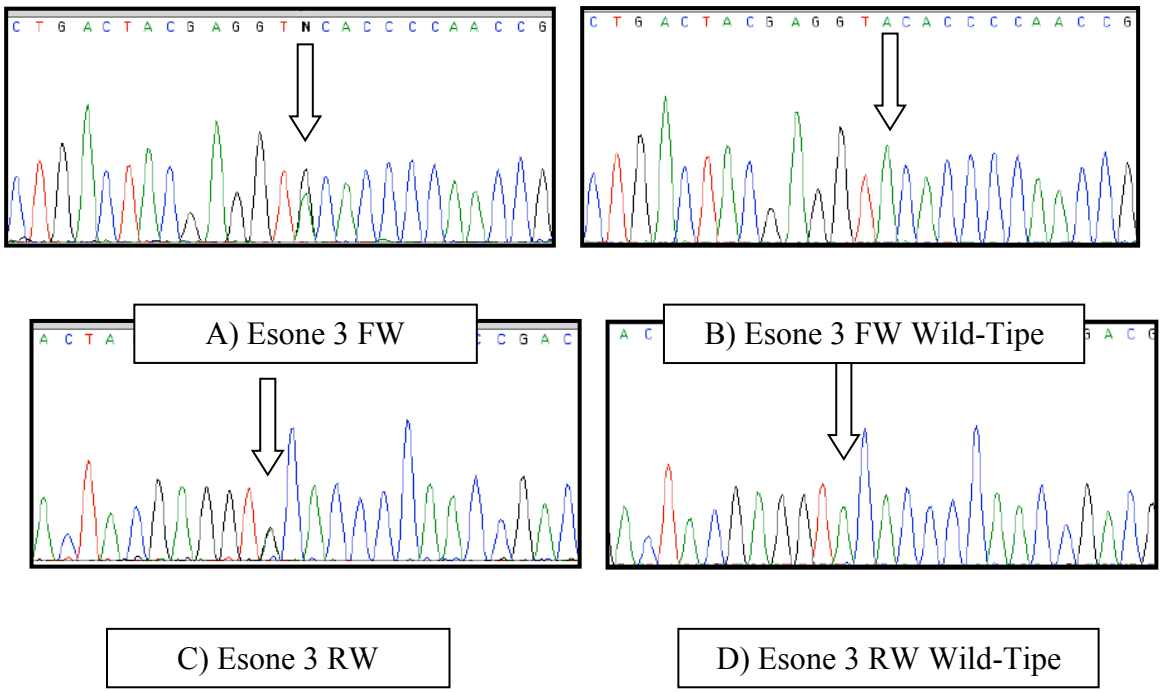

Figura 19: Elettroferogrammi relativi all'esone 3Fw e 3Rw comparati con un controllo sano. Le frecce indicano la presenza della mutazioni in (A) e (C) rispetto alla sequenza normale in (B) e (D).

Anche la seconda mutazione presenta una sostituzione amminoacidica, relativa al codone 666 che vede la nucleotide T con una C, determinando il cambiamento nella tripletta da codone wild-type TAC nel codone TAA, (UAA) che rappresenta un codone di stop al posto dell'amminoacido tirosina (Figura 20).

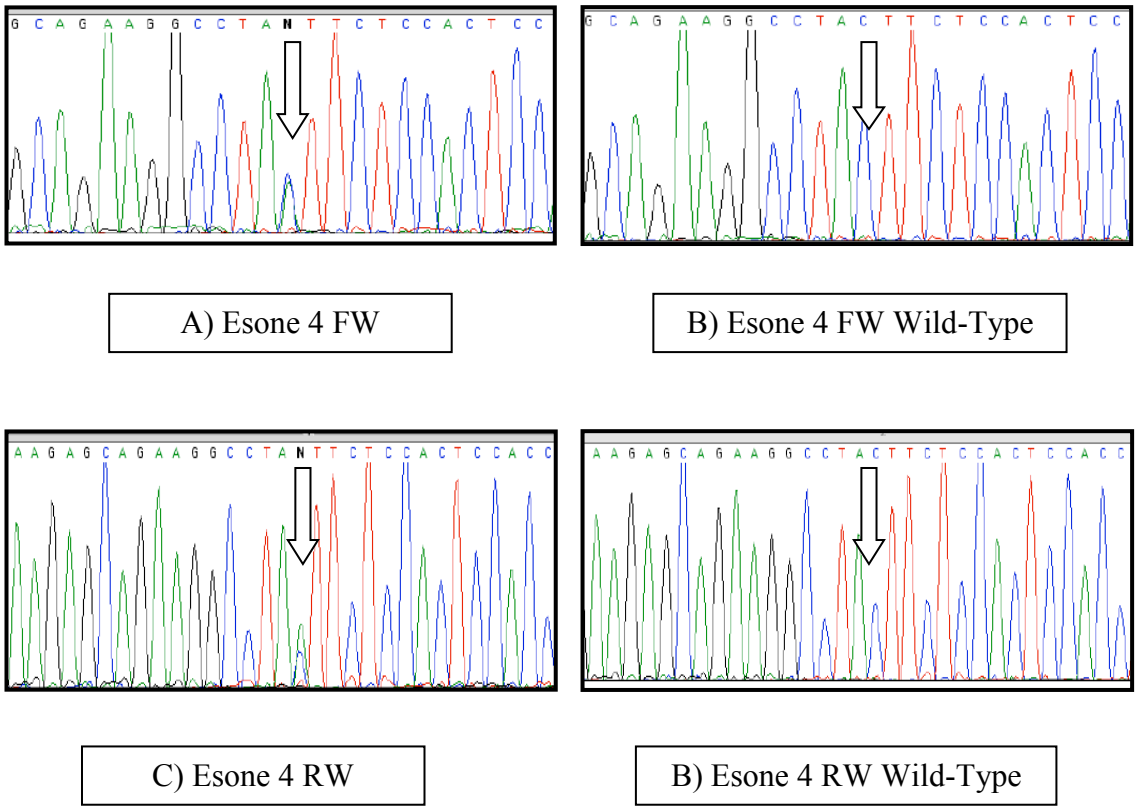

Figura 20: Elettroferogrammi relativi all'esone 3Fw e 3Rw comparati con un controllo sano. Le frecce indicano la presenza della mutazioni in (A) e (C) rispetto alla sequenza normale in (B)e in (D).

Queste due particolari mutazioni sono poco descritte in letteratura ma diversi studi hanno dimostrato l'associazione di mutazioni nel gene MMACHC associate alla patologia, come il lavoro pubblicato su Nature da Lerner-Ellis et.al nel 2006[43], in cui in 204 individui sono state individuate 42 differenti mutazioni la maggior parte delle quali caratterizzate da "loss-of-function" ovvero si verifica una ridotta, o addirittura, assenza della funzione biologica del prodotto genico. Nel 40% dei casi si riscontra la mutazione 271dupA, mentre in questo lavoro, delle due mutazioni riscontrate nella paziente oggetto di studio, la mutazione c.388T>C è stata diagnosticata in un soggetto in età infantile in associazione ad un'altra mutazione in eterozigosi; mentre la mutazione c.666C>A è stata individuata in 3 soggetti differenti, di cui due in età neonatale e uno durante l'adolescenza, anche in questo caso in associazione ad altre mutazioni.

Un altro studio in cui sono descritte le due mutazioni precedentemente citate è stato pubblicato nel 2008 da Nogueira[44] in cui oggetto di studio sono stati 24 pazienti italiani e 17 portoghesi con difetto cblC. Anche in questo lavoro la mutazione più frequente risulta essere 271dupA (55%) mentre la mutazione c.388T>C è stata

individuata in un soggetto di nazionalità italiana di sesso femminile all'età di 17 anni e la mutazione c.666C>A è stata osservata in due bambini italiani di sesso maschile e femminile rispettivamente in eterozigosi e in omozigosi.

Recentemente identificato sul cromosoma 1p34.1, il gene MMACHC codifica per un polipeptide di 282 aminoacidi di probabile localizzazione citosolica, la cui funzione non è ancora conosciuta; tuttavia, esperimenti di modellistica molecolare mostrano che la sua regione C-terminale presenta una somiglianza con TonB, una proteina batterica coinvolta nella trasduzione dell'energia generata dalla forza proton- motrice al trasporto di ferro e cobalamina attraverso la membrana batterica esterna. Per questo motivo è stato ipotizzato che il dominio C-terminale della proteina codificata da MMACHC funzioni in maniera simile al dominio C-terminale di tonB in E.coli, per cui anch'essa risulta avere un ruolo nel traffico intracellulare della cobalamina[43].

Vista la possibile diagnosi di Aciduria Metilmalonica si è passati al dosaggio, nel plasma e nelle urine della paziente, della metionina, dell'acido metilmalonico e dell'omocisteina.

I risultati mostravano valori bassi per la metionina e alti per gli altri due, confermando la diagnosi di Aciduria Metilmalonica con Omocistinuria.

Le tecniche d'indagine comunemente utilizzate non avrebbero mai permesso di identificare una specifica anomalia genetica causativa. Infatti, nonostante il quadro clinico della paziente indicasse sicuramente una malattia monogenica, l'assenza nel fenotipo di ritardo mentale, caratteristica peculiare della patologia poi riscontrata, non avrebbe mai portato ad ipotizzare che il gene mutato fosse quello responsabile della MMA di tipo cblC.

L'avvento della tecnologia NGS ha aperto importanti prospettive nella diagnostica molecolare e rappresenta un potente mezzo per l'analisi simultanea di un gran numero di regioni codificanti. L'applicazione di questa tecnica nella analisi molecolare di malattie dovute a grossi geni e di malattie caratterizzate da eterogeneità genetica riduce notevolmente i lunghi tempi di analisi, soprattutto in quei casi, come quello oggetto di studio nel presente lavoro di tesi, in cui non si hanno abbastanza informazioni cliniche.

Tutto questo non può che confermare come per la genetica umana i nuovi metodi di sequenziamento siano molto più che una nuova tecnica per leggere le sequenze di DNA, ma piuttosto sono diventati una tecnologia insostituibile per la diagnosi.

In conclusione, l'indagine genetica ha permesso di fare una diagnosi certa di alterazione del metabolismo dell'omocisteina e il sequenziamento dell'esoma della paziente B.L., che aveva un fenotipo non del tutto ascrivibile ad una patologia ben definita e con un quadro clinico che induceva ad ipotizzare il coinvolgimento di geni implicati in patologie metaboliche, ha permesso di soffermare la nostra attenzione sul gene MMACHC.

Dal momento che si trattava essenzialmente di un'ipotesi di partenza l'approccio NGS ha permesso di poter analizzare in un'unica seduta analitica l'intera regione codificante, costituita dall'esoma consentendo di poter effettuare ulteriori tentativi di ricerca di mutazioni in altri geni qualora l'ipotesi non si fosse rivelata corretta. eventualità resa possibile dall'assenza nel fenotipo della caratteristica peculiare della patologia sospetta, ovvero una forma di ritardo mentale.

In questo caso il sospetto del coinvolgimento del gene MMACHC è stato confermato mediante il sequenziamento tradizionale (Sanger) permettendo di formulare una diagnosi ben precisa di una determinata patologia (Acidemia Metilmalonica con Iperomocistineimia) in un soggetto che non mostrava un fenotipo clinico del tutto ascrivibile alla patologia stessa.

Dunque, partendo dall'analisi di un numero enorme di varianti individuate tramite sequenziamento dell'esoma si è giunti all'identificazione di due mutazioni in eterozigosi nel gene MMACHC che spiegano il fenotipo clinico. Pertanto, mentre la sfida per il bioinformatico è ottenere dati di qualità col minor numero di artefatti possibile, la sfida del genetista è quella di essere in grado di interpretare propriamente il significato delle varianti.

Il risultato ottenuto ha, quindi, una validità notevole in quanto consente una gestione clinica del paziente più mirata ed accurata. Infatti, il trattamento con idrossicobalammina sintetica ha portato ad una remissione della sintomatologia con normalizzazione pressocchè completa dell'iperomocisteinemia conferendo alla paziente la possibilità di condurre un normale stile di vita e di portare a termine una seconda gravidanza con la nascita di un bambino completamente sano.

Quindi, questo studio ci permette di dire che la terapia con supplementazione del complesso delle vitamina B è stata efficace in molte forme di iperomocisteinemia nonché in queste forme di grave iperomocisteinemia. Oggigiorno la terapia con vitamina B12 consiste nella supplementazione per via parenterale. Nel prossimo futuro, sarebbe auspicabile avere una formulazione utile per la terapia orale sia di

cianocobalamina e idrossicobalamina. Inoltre, sarebbe auspicabilmente avere un pool unico di vitamine del complesso B (B6, B12, acido folico, betaina) complessato con il cibo, per una terapia orale facile per i pazienti.

Bibliografia

1. W. Gilbert and A. Maxam (1973) The Nucleotide Sequence of the lac Operator. Proc Natl Acad Sci U S A. 70(12 Pt 1-2): 3581-3584.
2. F. Sanger, A.R. Coulson (1975) A rapid method for determining sequences in DNA by primed synthesis with DNA polymerase. J Mol Biol. 25;94(3): 441-448.
3. F. Sanger, S. Nicklen, and A.R. Coulson (1977) DNA sequencing with chain-terminating inhibitors. Proc Natl Acad Sci U S A. 74(12): 5463–5467.
4. D.A. Wheeler, M. Srinivasan, M. Egholm et al. (2008) The complete genome of an individual by massively parallel DNA sequencing. Nature. 452:872-6
5. K.V. Voelkerding, S.A. Dames, J.D. Durtschi (2009) Next-generation sequencing: from basic research to diagnostics. Clinical chemistry. 55(4): 641-658
6. M. Margulies, (2005) Genome sequencing in microfabricated hig-density picolitre reactors. Nature. 437: 376-80.
7. M. Ronaghi (2001) Pyrosequencing sheds light on DNA sequencing. Genome Res.11(1):3-11.
8. D.S. Tawfik, A.D. Griffiths (1998) Man-made cell-like compartments for molecular evolution. Nat Biotechnol.16(7): 652-6.
9. S.M. Huse, J.A. Huber, H.G. Morrison et al. (2007) Accuracy and quality of massively parallel DNA pyrosequencing. Genome Biology. 8(7):R143.
10. G. Turcatti, A. Romieu, M. Fedurco, et al. (2008) A new class of cleavable fluorescent nucleotides: synthesis and optimization as reversibile terminators for DNA sequencing by synthesis. Nucleic Acids Research. 36(4): e25.
11. J. Shendure and Hanlee Ji (2008) Next-generation DNA sequencing. Nature Biotechnology. 26, 1135 – 1145.
12. M.L. Metzker (2010) Nature Reviews Genetics 11, 31-46 doi:10.1038/nrg2626
13. K. Nakamura, T. Oshima, T. Morimoto et al. (2011) Sequence--- specific error profile of Illumina sequencers. Nucleic Acids Research Advance Access, 9(13):e90. doi: 10.1093/nar/gkr344.
14. J. Shendure, G.J. Porreca, N.B. Reppas et al. (2005) Accurate multiplex polony sequencing of an evolved bacterial genome. Church Science 309, 1728 DOI: 10.1126/science.1117389

15. T. Tucker , M. Marra, J.M. Friedman (2009) Massively parallel sequencing: the next big thing in genetic medicine. Am J Hum Genet. 85(2):142-54. doi: 10.1016/j.ajhg.2009.06.022.
16. . T.D. Harris, P.R. Buzby, H. Babcock et al. (2008). Single-molecule DNA sequencing of a viral genome. Science. 320:106-9.
17. B. Ewing, P. Green. (1998) Base-calling of automated sequencer traces using phred. II. Error probabilities. Genome Res. 8:186-94.
18. B .Ewing, L. Hillier, M.C. Wendl et al. (1998) Base-calling of automated sequencer traces using phred. I. Accuracy assessment. Genome Res. 8:175-85
19. K.H. Taylor, R.S. Kramer, J.W. Davis, J. Guo et al. (2007) Ultradeep bisulfite sequencing analysis of DNA methylation patterns in multiple gene promoters by 454 sequencing. Cancer Res. 15;67(18):8511-8.
20. N. Whiteford, N. Haslam, G. Weber et al. (2005)An analysis of the feasibility of short read sequencing. Nucleic Acids Res. 7;33(19):e171
21. D.A. Wheeler, M. Srinivasan, M. Egholm,et al. (2008) The complete genome of an individual by massively parallel DNA sequencing. Nature. 17;452(7189):872-6. doi: 10.1038/nature06884.
22. W. Brockman, P. Alvarez, S. Young et al. (2008) Quality scores and SNP detection in sequencing-by-synthesis systems. Genome Res;18:763-70.
23. J.C. Dohm, C. Lottaz, T. Borodina, et al. (2008) Substantial biases in ultra-short read data sets from high-throughput DNA sequencing. Nucleic Acids Res. 36:e105.
24. S. Jaenicke, C. Ander, T. Bekel et al. (2011) Comparative and joint analysis of two metagenomic datasets from a biogas fermenter obtained by 454-pyrosequencing. PLoS One. 6(1): e14519.
25. P.D. Schloss and Handelsman J. (2003) Biotechnological prospects from metagenomics. Curr Opin Biotechnol. 14(3): 303-10.
26. P. Parameswaran, P. Charlebois, Y. Tellez et al. (2012) Genome-wide patterns of intrahuman dengue virus diversity reveal associations with viral phylogenetic clade and interhost diversity. J Virol. 86(16): 8546-58.
27. V.E.A. Russo, R.A. Martienssen e A.D. Riggs. (1996) Epigenetic mechanisms of gene regulation. Cold Spring Harbor Laboratory Press.
28. A. Barski, S. Cuddapah, K. Cui et al. (2007) High-resolution profiling of histone methylations in the human genome. Cell. 129:823-37.

29. A. Valouev, J. Ichikawa, T. Tonthat et al. (2008) A high-resolution, nucleosome position map of C. elegans reveals a lack of universal sequence-dictated positioning. Genome Res. 18:1051-63.
30. A.N. Massa, K.L. Childs, L. Haining et al. (2011) The transcriptome of the reference potato genome Solanum tuberosum Group Phureja clone DM1-3 516R44. PLoS One. 6(10): e26801.
31. H.C. Fan, Y-J. Blumenfeld, U. Chitkara et al. (2008) Noninvasive diagnosis of fetal aneuploidy by shotgun sequencing DNA from maternal blood. Proc Natl Acad Sci U S A. Oct 21; 105(42): 16266–16271.
32. J.M. Rothberg, W. Hinz, T.M. Rearick, J. Schultz et al. (2011) An integrated semiconductor device enabling non-optical genome sequencing. Nature. 20;475(7356):348-52. doi: 10.1038/nature10242.
33. C.R. Scriver, A.L. Beaudet, W.S. Sly et al. (2001) The Metabolic and Molecular Bases of Inherited Disease, Eighth Edition McGraw-Hill.
34. K.S. McCully (1969) Vascular pathology of homocysteinemia: implications for the pathogenesis of arteriosclerosis. Am I Pathol. 561 11-28.
35. S.H. Mudd, R. Skovby, H.L. Levy et al. (1985) The natural history of homocystinuria due to cystathionine-Beta-synthase deficiency. Am I Hum Genet 37:1-31.
36. G. Davì, G. Di Minno, A. Coppola, et al.(2001) Oxidative Stress and Platelet Activation in Homozygous Homocystinuria. Circulation. 104:1 124
37. G. Sebastio, M.T. Sperandeo, M. Panico et al. (1995) The molecular basis of homocystinuria due to cysthationine-Beta-synthase deficiency in Italian families and report of four novel mutations. Am J Hum Genet 56:1324-33.
38. A. Coppola, C. De Simone, D. Coppola et al. (2004) Omocisteina, folati e patologia tromboembolica. Riv.It.Ost.Gin. 1: 13-19.
39. I. Rezvani (2002) Metionina. in Nelson: Trattato di Pediatria. Edizioni Minerva Medica. XVI Ed.; X-82.3: 350-52.
40. S.H. Mudd, H.L. Levy, J. Kraus (2001) Disorders of Transsulfuration in the metabolic and molecular bases of inherited Diseases. VIII Ed. Mc Grow Hill; 88: 2008-48
41. K. Meyer, A. Fredriksen, P.M. Ueland (2004) High-level multiplex genotyping of polymorphisms involved in folate or homocysteine metabolism

by matrix-assisted laser desorption/ionization mass spectrometry. Clin Chem. 50:391-402.

42. D.J. Hedges, D. Burges, E. Powell et al. (2009) Exome Sequencing of a Multigenerational Human Pedigree. PLoS ONE, 4;4(12):e8232.
43. J.P. Lerner-Ellis, J.C. Tirone, P.D. Pawelek et al. (2006) Identification of the gene responsible for methylmalonic aciduria and homocystinuria, cblC type. Nat Genet, 38(1):93-100.
44. C. Nogueira, C.Aiello, R.Cerone et al. (2008) Spectrum of MMACHC mutations in Italian and Portuguese patients with combined methylmalonic aciduria and homocystinuria, cblC type. Mol. Genet. Metab, 93(4):475–480. (Epub 2008/01/01)