



Università degli Studi di Cagliari

**DOTTORATO DI RICERCA**

**Terapia Pediatrica e Farmacologia dello Sviluppo**

**Ciclo XXVI**

**TITOLO TESI**

**NEXT GENERATION SEQUENCING NELL'ANALISI DEL  
DNA FETALE DA PLASMA MATERNO PER LA DIAGNOSI  
PRENATALE NON INVASIVA DI MALATTIE GENETICHE**

Settore scientifico disciplinare di afferenza

**BIO/11**

Presentata da:	Dott. Andrea Picciau
Coordinatore Dottorato	Prof. Paolo Moi
Relatore	Prof.ssa Maria Cristina Rosatelli

Esame finale anno accademico 2012 – 2013



## LA DIAGNOSI PRENATALE

Le recenti linee guida nazionali emanate dal Ministero della Salute forniscono precise indicazioni sulle procedure strumentali e di laboratorio per il monitoraggio della gravidanza fisiologica. L'insieme di queste procedure ha come obiettivo l'individuazione precoce di patologie genetiche, di malformazioni fetali riconducibili, ad esempio, a malattie infettive insorte nella madre durante la gravidanza e, in generale, la determinazione di fattori di rischio per il feto.

Le indagini prenatali possono essere eseguite sia attraverso procedure non invasive, come gli esami ultrasonografici e le valutazioni dei marcatori biochimici dosabili nel siero materno (screening prenatali); sia mediante tecniche diagnostiche invasive, come villocentesi e amniocentesi (diagnosi prenatali).

La diagnosi prenatale, necessaria per la definizione della costituzione genetica del feto, viene suggerita in tutti i casi in cui vi sia un'indicazione di rischio per una patologia genetica grave. I fattori che determinano la necessità dell'intervento diagnostico sono di diversi tipi:

- fattori genetici, quali familiarità per patologie genetiche o genitori portatori sani di mutazioni patologiche;
- fattori epidemiologici, come l'età materna avanzata, significativamente associata all'insorgenza di anomalie cromosomiche fetali;
- positività ai test di screening prenatali non invasivi.

La conoscenza in epoca prenatale della costituzione genetica del feto permette alla coppia, in caso di esclusione della patologia, di affrontare il periodo della gravidanza senza il carico psicologico della presenza di una eventuale patologia a carico del nascituro, o, viceversa, in presenza di fenotipo fetale patologico, valutare con piena consapevolezza se ricorrere all'interruzione volontaria della gravidanza o decidere di portarla a termine, preparandosi all'accoglienza di un neonato affetto.

L'esame e la valutazione dello stato di rischio per il feto vengono eseguiti dal consulente genetista che, in presenza di evidenze suggestive per una patologia,

ha l'incarico di esporre alla coppia quali indagini si rendano utili per la definizione dello stato di salute del feto, e quali rischi siano ad esse correlati.

#### TECNICHE DI SCREENING NON INVASIVE

Sono definiti metodi di indagine non invasiva i test di screening a cui tutte le donne sono invitate a sottoporsi in gravidanza, allo scopo di individuare un rischio aumentato per una data patologia a carico del feto, e suggerire un eventuale ricorso alle indagini invasive per una sua precisa definizione.

I test di screening non invasivi attualmente disponibili prevedono la valutazione della concentrazione di marcatori biochimici, come la beta gonadotropina corionica umana ( $\beta$ -hCG) e la proteina A plasmatica associata alla gravidanza (PAPP-A), che costituiscono nel loro insieme il Bi-test, a cui vengono associati la valutazione ecografica della translucenza nucale fetale e altri parametri ecografici, utili per la determinazione del rischio aumentato per patologie fetali. I risultati ottenuti vengono messi in relazione con l'età materna e l'epoca gestazionale. Questi metodi non possono essere annoverati tra i test diagnostici in quanto i loro risultati sono espressi, non come evidenze certe della presenza di caratteri patologici a carico del feto, ma come stima della probabilità di rischio. Tuttavia, poiché questi test di screening non presentano controindicazioni per la salute del feto e possono essere proposti in epoca gestazionale precoce, rivestono un ruolo importante nel monitoraggio della gravidanza e forniscono una precisa indicazione alla diagnosi prenatale.

Grazie al miglioramento e all'affidabilità raggiunta, le tecniche non invasive hanno assunto grande importanza, sia per le coppie a rischio, permettendo loro decidere con maggior consapevolezza se effettuare l'interruzione volontaria di gravidanza in caso di patologie fetali gravi o portare a termine la gravidanza, sia per il personale sanitario che, relativamente al momento del parto, potrà predisporre al meglio ad affrontare la nascita di un neonato con condizioni fisiologiche particolari.

## TECNICHE DI PRELIEVO INVASIVE

Ogni anno in Italia nascono circa 30.000 neonati affetti da malattie genetiche o congenite e molte di queste possono essere diagnosticate mediante indagini molecolari (1). Esse prevedono un prelievo di materiale fetale su cui svolgere le analisi, che può essere effettuato mediante amniocentesi o villocentesi.

L'amniocentesi è stata la prima tecnica di prelievo mirato alla diagnosi prenatale ad essere introdotta nella pratica clinica ed è tutt'oggi la più utilizzata. Essa consiste nel prelievo di 15-20ml di liquido amniotico attraverso un ago introdotto per via transaddominale nella cavità amniotica, sotto continuo controllo ecografico. Nel liquido amniotico sono presenti cellule derivate dallo sfaldamento dei tessuti e delle mucose del feto. Messe in coltura e opportunamente stimulate, esse iniziano a dividersi e permettono l'analisi dei cromosomi metafasici. L'epoca gestazionale ottimale per sottoporsi ad amniocentesi è compresa tra la 15<sup>a</sup> e la 17<sup>a</sup> settimana e il rischio di perdita fetale si attesta tra 0.5% e 1%. Prelievi effettuati in epoche precedenti a quelle citate mostrano rischio abortivo e di fallimento della tecnica aumentato.

La villocentesi prevede un prelievo di tessuto trofoblastico (villi coriali), di origine fetale, attraverso l'introduzione di un ago per via transvaginale o, preferibilmente, transaddominale, sotto controllo ecografico. Il periodo ottimale per la sua attuazione è intorno alla 11<sup>a</sup>-12<sup>a</sup> settimana di gestazione. Il rischio di perdita fetale è stimato tra 0.5% e 2%. Un aspetto da tenere in considerazione nell'applicazione di questa tecnica è che, nel prelievo dei villi coriali, vengono asportate anche cellule materne derivanti dalla decidua; è necessario quindi che queste vengano rimosse per evitare errori diagnostici derivanti da contaminazione materna.

Il ricorso alle tecniche di diagnosi invasiva deve essere preceduto da una comunicazione tra la gestante e il consulente genetista, in cui vengano esplicitate tutte le possibili implicazioni, rischi e benefici legati alle procedure di prelievo. Il consenso alla diagnosi deve essere redatto in forma scritta e firmato sia dal consulente che dalla gestante.

Ad oggi, il fattore che maggiormente influenza e limita l'accesso alla diagnosi prenatale è il suo carattere di invasività associato al rischio, seppur

minimo, di perdita fetale, a cui si aggiunge spesso lo stato d'ansia dovuto all'iter di ospedalizzazione a cui deve sottoporsi la donna in gravidanza. Questi limiti spiegano come la meta auspicata da coloro che operano in ambito diagnostico prenatale sia la possibilità di effettuare una diagnosi senza ricorrere a tecniche invasive, eliminando quindi i rischi a carico del feto.

# DNA FETALE

## CARATTERISTICHE DEL DNA FETALE

Negli ultimi 20 anni, l'obiettivo comune dei gruppi di ricerca operanti nell'ambito delle diagnosi prenatali è stato quello di sviluppare metodologie diagnostiche non invasive, finalizzate alla riduzione del rischio di complicità e all'anticipazione dell'epoca del prelievo. Uno degli approcci diagnostici che si è cercato di sviluppare è basato sull'analisi delle cellule fetali, presenti nel circolo periferico materno sin dalle prime settimane di gravidanza. Le cellule di derivazione fetale, e in particolar modo gli eritroblasti fetali che vengono designati come un target cellulare dalle caratteristiche ottimali, possono essere individuate grazie alla presenza di marcatori specifici intracellulari e di superficie, indispensabili per la distinzione delle cellule fetali da quelle materne. Tuttavia, la loro esigua presenza nel circolo periferico della madre rende molto difficoltosi la loro individuazione e il loro isolamento. Da ciò, scaturisce un livello di riproducibilità degli esperimenti non sufficiente perché le cellule fetali possano essere, ad oggi, considerate come una fonte sicura di materiale genetico fetale da impiegarsi nella diagnosi prenatale non invasiva.

La via che attualmente viene considerata più promettente per lo sviluppo di protocolli di diagnosi prenatale non invasiva è quella che prevede l'utilizzo del DNA fetale libero (cell free fetal DNA - cffDNA) come fonte di materiale genetico attraverso cui condurre le analisi del genotipo fetale. Nel 1996 furono pubblicati dei lavori in cui veniva descritta, nel plasma di pazienti oncologici, la presenza di frammenti di DNA derivanti da cellule tumorali, liberatosi nel circolo sanguigno in seguito ad apoptosi (2). Partendo da questa osservazione, il gruppo guidato da YMD Lo (3) valutò l'ipotesi che tale fenomeno potesse similmente verificarsi per il DNA fetale nel circolo materno. Con lo scopo di confermare questa ipotesi, fu condotto uno studio su 43 donne in gravidanza, 30 portatrici di un feto maschio e 13 di un feto femmina, e su ulteriori 10 donne non gravide come controlli negativi. Nell'80% delle donne in attesa di feto maschio, fu possibile amplificare sequenze mappanti sul cromosoma Y. Non fu riscontrata la presenza di queste sequenze né nelle donne portatrici di feto femmina né in quelle non in gravidanza. Attraverso

queste evidenze, fu semplicemente dimostrata la presenza di DNA fetale nel circolo materno. Questa scoperta ha avuto enormi risvolti in campo diagnostico tali che, ad oggi, l'analisi del DNA fetale libero è alla base di quasi tutti gli studi d'avanguardia nell'ambito della diagnosi prenatale non invasiva.

Il DNA libero nel circolo materno si presenta come una miscela di frammenti di acidi nucleici aventi due distinte origini: la madre e il feto. Verificare la concentrazione del DNA fetale rispetto a quello materno, è stato uno dei primi aspetti del cfDNA ad essere approfondito. Ancora oggi, la valutazione della quota fetale contenuta nei campioni di DNA libero rimane un dato di fondamentale importanza per ogni studio intrapreso, dato che dalla percentuale di quota fetale dipende l'effettiva possibilità di poter condurre delle analisi sul cfDNA (DNA circolante libero). In uno dei primi lavori sul DNA fetale libero, pubblicato dallo stesso gruppo di YMD Lo, veniva indicata come quantità di DNA fetale una frazione compresa tra 3% e 6% rispetto al DNA totale libero plasmatico (4), con concentrazioni più basse all'inizio della gravidanza e crescenti con il procedere della gestazione. In un lavoro del 2008, YMD Lo (5) ha pubblicato uno studio comparativo tra le diverse tecnologie disponibili per l'analisi del DNA fetale libero (Real-Time PCR, spettrometria di massa, Digital PCR). La Digital PCR ha ottenuto i migliori risultati proponendosi come metodica più affidabile per la quantificazione del DNA fetale. In questo stesso studio, YMD Lo affermava che la concentrazione del cfDNA nel plasma materno è notevolmente più alta rispetto ai primi dati proposti, attestandosi mediamente intorno al 10% durante il primo trimestre per poi crescere fino al 20-25% al terzo trimestre. Ulteriori studi, condotti in seguito mediante le tecniche più recenti di Next Generation Sequencing, descrivono casi con valori ancora superiori (6) (7) (8).

I dati finora descritti mettono in evidenza l'aspetto più critico legato all'analisi del cfDNA: nel plasma, esso è presente come miscela di frammenti di derivazione fetale e materna, e, poichè la quota materna è quantitativamente la più importante, evidenziare i caratteri fetali è un processo complicato, in quanto gli effetti attribuibili al DNA fetale vengono confusi e mascherati da quelli causati dalla controparte materna. Da ciò deriva che i segnali che più facilmente possono essere rilevati sono quelli derivanti da sequenze geniche non condivise tra feto e



madre, come le sequenze del cromosoma Y ereditate dal feto per via paterna e non presenti nel DNA materno.

Una peculiarità del DNA fetale, di fondamentale importanza per la diagnostica prenatale, è la sua rapida rimozione dal circolo materno dopo il parto: una sua persistenza non permetterebbe infatti analisi in gravidanze successive alla prima. Già nel 1999 veniva descritto da YMD Lo et al. (9) come il DNA fetale non fosse più evidenziabile dopo un giorno dal parto: analisi attraverso real-time PCR mirate alla identificazione di frammenti mappanti sul cromosoma Y, infatti, non davano evidenza alcuna dopo 24 ore (emivita media di 16 min). Lo stesso gruppo di lavoro, ha pubblicato recentemente un nuovo studio, realizzato con le tecnologie di sequenziamento massivo, confermando i dati già ottenuti e descrivendo come l'eliminazione del DNA fetale avvenga in due fasi differenti: la prima, che mostra una cinetica molto veloce, nella quale in un ora circa dal parto viene dimezzata la percentuale di cfDNA; la seconda, più lenta, porta ad un ulteriore dimezzamento entro 13 ore circa. Si può affermare che l'intera quota del DNA fetale venga rimossa entro 1-2 giorni (10).

Il DNA circolante libero, sia quello di origine fetale che quello materno, non è rappresentato da grandi molecole di acidi nucleici ma si presenta frammentato, con lunghezze che arrivano mediamente fino alle 200bp (11). Il grafico della distribuzione delle lunghezze dei frammenti del DNA libero mostra, per il DNA materno il picco più alto a 166bp, e per il DNA fetale a 143bp. I frammenti sia materni che fetali, inoltre, mostrano una periodicità nella misura delle loro lunghezze di 10bp: questa peculiarità è riconducibile all'azione di rimozione del DNA dalle proteine istoniche dei nucleosomi ad opera delle nucleasi, evidenza che permette inoltre di consolidare l'ipotesi secondo cui il DNA libero nel plasma derivi da apoptosi cellulare (12). Quale fosse la fonte di DNA fetale, è stata una delle prime domande a cui i ricercatori hanno cercato di dare risposta. Ad oggi, l'ipotesi più accreditata è che il DNA libero fetale derivi da cellule apoptotiche della placenta. Due evidenze possono essere citate a sostegno di questa tesi: la prima, come mostrato dai dati di Alberry nel 2007, è che gravidanze anembrioniche mostrano livelli di cfDNA pari a quelli di gravidanze normali in pari età gestazionale (13); la seconda, è che, dai diversi studi mirati alla

definizione del pattern epigenetico del DNA fetale, è emerso come le molecole di DNA fetale libero mostrino la stessa marcatura epigenetica delle cellule placentari (14) (15).

#### DETERMINAZIONE DEL SESSO FETALE

La prima applicazione di carattere diagnostico condotta attraverso l'utilizzo del cfDNA è stata la determinazione del sesso fetale. Lo stesso studio in cui per la prima volta veniva messa in evidenza la presenza del cfDNA nel circolo materno, descriveva, in ultima analisi, un protocollo di diagnosi di sesso fetale, attraverso la determinazione di frammenti mappanti sul cromosoma Y nel sangue della madre (3). Gli studi che si sono susseguiti successivamente hanno ampiamente messo in evidenza come la determinazione del sesso fetale attraverso cfDNA, oltre che essere di semplice e veloce esecuzione, mostri elevata accuratezza (16) (4). In alcuni di questi, in cui il sesso fetale è stato determinato valutando la concentrazione di frammenti del gene SRY mediante PCR quantitativa, sono stati evidenziati valori di sensibilità e specificità del 100% (17) (18). In particolare Rijnders in un lavoro pubblicato nel 2003 (18), indica come già dalla 5<sup>a</sup> settimana di gestazione sia possibile determinare il sesso del feto e specifica inoltre come tutti i campioni da lui osservati, con un limite minimo di età gestazionale di 10 settimane, siano stati definiti correttamente nel 100% dei casi. Dati di meta-analisi più recenti, scaturiti dalla valutazione globale dei dati pubblici presentati dal 1997 al 2011, indicano che la determinazione del sesso fetale attraverso il cfDNA mostra una sensibilità del 95.4% e una specificità del 98.6%, evidenziando come questi valori aumentino all'avanzare dell'epoca gestazionale (19).

Queste osservazioni trovano la loro più immediata applicazione diagnostica nelle patologie X-linked, per le quali la determinazione del sesso definisce la possibilità per il feto di presentare fenotipo patologico, rendendo possibile l'esclusione dalle procedure diagnostiche invasive a tutte le gestanti con feto non a rischio di malattia. In riferimento alle malattie X-linked recessive, è facilmente intuibile il vantaggio della definizione del sesso fetale, poichè in presenza di femmine è esclusa la possibilità di fenotipo patologico, rischio che invece permane per i maschi, per i quali saranno necessarie ulteriori indagini diagnostiche. Un

esempio dell'importanza della determinazione del sesso nel caso di una malattia X-linked recessiva è stato quello descritto dal gruppo di NBY Tsui in collaborazione con YMD Lo e R. Chiu (20), che, nell'ambito di uno studio sull'emofilia, pone alla base di tutto il protocollo diagnostico la definizione del sesso del nascituro consentendo l'esclusione da ulteriori indagini molecolari tutti i campioni di cffDNA risultati di sesso femminile, con evidente risparmio in termini di tempo e costi sostenuti, ma soprattutto annullando i rischi legati al prelievo invasivo.

Oltre alle malattie X-linked, sono presenti ulteriori patologie in cui l'identificazione del sesso fetale può essere di grande utilità. Rijinders nel 2001 ha descritto in una sua pubblicazione un caso di iperplasia surrenale congenita, malattia autosomica recessiva. In presenza di rischio per questa malattia, l'indicazione terapeutica consiste nella somministrazione alle madri in gravidanza di desametasone, farmaco che ha lo scopo di impedire la virilizzazione dei caratteri sessuali nei feti femmina. Poiché l'assunzione di questo farmaco però comporta gravi effetti collaterali, Rijinders sottolinea l'importanza della determinazione del sesso fetale in epoca precoce, allo scopo di risparmiare alle madri in attesa di feti maschi una terapia inutile e dannosa (17).

Le tante evidenze positive prodotte dalla letteratura in relazione alla determinazione del sesso fetale hanno fatto sì che, in alcuni Stati, essa sia già stata inserita nelle pratiche cliniche diagnostiche di routine. In altri paesi, come in Italia, ancora essa non ha ancora assunto un ruolo ufficiale tra le metodiche diagnostiche.

#### DETERMINAZIONE DEL FATTORE RH

E' stato già descritto come il DNA circolante libero sia costituito da una miscela di frammenti genomici di derivazione sia materna che fetale. Questo implica che i caratteri genetici fetali più facilmente evidenziabili siano quelli riconducibili ad una ereditarietà paterna e non condivisi con il genoma materno. Questa osservazione, sin dai primi anni successivi alla scoperta del DNA fetale libero, ha indirizzato diversi gruppi di ricerca verso l'utilizzo del cffDNA in un ulteriore ambito rispetto alla determinazione del sesso fetale: la genotipizzazione del fattore Rh fetale, informazione che, se in possesso dei clinici in epoca

prenatale, è in grado di offrire importanti vantaggi nella gestione della gravidanza (21).

Il fattore Rh è definito dalla presenza o assenza di una proteina di membrana codificata dal gene RHD. Nella popolazione caucasica questo gene si presenta spesso deletato, con conseguente assenza della proteina da esso codificata. Durante il parto, la presenza di un feto Rh positivo scatena in una madre Rh negativa il fisiologico processo di sensibilizzazione verso l'antigene D, riconosciuto come estraneo, con conseguente sviluppo di anticorpi IgG anti-D. Nelle gravidanze successive alla prima, in presenza di feto Rh+, poiché le immunoglobuline IgG anti-D presenti nella madre già immunizzata, capaci di attraversare la placenta, danno l'avvio ad una risposta immunitaria contro gli antigeni fetali, con conseguente insorgenza di malattia emolitica del neonato (MEN), un'emolisi massiva delle emazie fetali che nei casi più gravi è causa di morte intrauterina (eritroblastosi fetale). Per eliminare questo rischio, è prevista una adeguata terapia profilattica che consiste nella somministrazione alla madre, intorno alla 28ª settimana e entro 72 ore dalla nascita del neonato Rh positivo, di immunoglobuline anti-D, che hanno lo scopo di legare gli antigeni D fetali affinché vengano prontamente degradati e sia così inibita l'attivazione della risposta immunitaria materna.

L'introduzione nelle attuali pratiche cliniche della determinazione non invasiva del fattore Rh fetale presenterebbe dei vantaggi sia dal punto di vista economico che sanitario. Conoscere il fattore Rh del feto in epoca prenatale, permetterebbe anzitutto di ridurre in maniera notevole il numero delle donne sottoposte a terapia immunoprofilattica, escludendo tutte le madri con feto Rh negativo, quindi non a rischio di MEN. Inoltre, la diminuzione del numero delle donne da sottoporre a terapia implicherebbe un minor fabbisogno di immunoglobuline anti-D, la cui produzione comporta costi notevoli. Infine, nonostante la somministrazione delle IgG anti-D sia soggetta a stretto controllo, la produzione di queste molecole avviene a partire da sorgente umana, fatto che determina la presenza di un potenziale rischio infettivo.

I più recenti lavori presentati relativamente alla determinazione del fattore Rh da cfDNA, mostrano ampie casistiche e valori di sensibilità e specificità molto

elevati. Tra questi, Legler et al. (22) hanno pubblicato un lavoro di meta-analisi dei dati relativi alle casistiche presentate in letteratura negli anni 2006-2008, mettendo in evidenza come i risultati globali indichino accuratezza media pari al 97%. Riguardo poi alla efficacia del test in relazione all'epoca gestazionale, Akolekar (23) dichiara come, su un campione di circa 500 gravidanze comprese fra la 11<sup>a</sup> e la 13<sup>a</sup> settimana di gestazione, la metodica abbia mostrato una sensibilità pari al 100% (332 casi Rh+) e una specificità del 96,5% (4 falsi negativi Rh+ su 170 casi Rh- ). Questi dati dimostrano, quindi, come l'analisi del DNA fetale rappresenti uno strumento valido e dalla comprovata efficacia per l'analisi del fattore Rh fetale in epoca gestazionale precoce e, in virtù di ciò, in diversi paesi europei e degli Stati Uniti questa metodica è oramai presente nell'elenco dei test diagnostici disponibili nella pratica clinica ordinaria.

#### APPLICAZIONI DELL' ANALISI DEL CFDNA NELLE MALATTIE MONOGENICHE

Dagli inizi del loro utilizzo, le tecniche di diagnosi invasiva hanno manifestato rischio di perdita fetale, anche se con il progredire della tecnica e, soprattutto, dell'esperienza degli operatori, esso è stato fortemente ridotto raggiungendo attualmente valori molto bassi. Questo seppur minimo livello di rischio spiega il motivo per il quale, da sempre, i diversi gruppi di ricerca operanti nella diagnostica prenatale abbiano avuto come obiettivo lo sviluppo di protocolli di diagnosi non invasiva, e il perché il DNA fetale libero circolante nel sangue materno si sia affermato, sin dalla sua scoperta, come strumento di fondamentale importanza per poter acquisire, in epoca prenatale, le informazioni sul genoma fetale. La difficoltà maggiore che si riscontra nell'analisi del cfDNA è causata dalla costituzione stessa del DNA libero circolante: esso è una miscela di DNA materno e DNA fetale, in cui il primo, contenuto in quantità maggiore, determina un mascheramento degli effetti del secondo. Nei casi precedentemente citati riguardo la determinazione del sesso fetale e del fattore Rh, è stato messo in evidenza come sia possibile effettuare una diagnosi attraverso la ricerca di frammenti di DNA fetali ereditati per via paterna. E' relativamente semplice infatti effettuare delle valutazioni diagnostiche quando l'elemento oggetto della ricerca non è condiviso dal DNA materno, come nel caso dei frammenti del cromosoma Y o del gene

RHD nelle madri Rh negative, e la sua presenza nel DNA libero è chiaramente riconducibile al genoma fetale. Questo principio è valido, non solo nel caso del sesso fetale o del gene RHD, ma trova applicazione in diversi campi d'analisi, compreso quello delle malattie monogeniche recessive. Sono stati pubblicati, infatti, numerosi lavori in cui si descrive la possibilità di effettuare una diagnosi in epoca precoce valutando la presenza o l'assenza di frammenti fetali ereditati per via paterna e non condivisi con la madre. Tra i primi lavori, in cui vengono descritti casi di diagnosi prenatale non invasiva di patologie autosomiche recessive, è importante citare quello pubblicato nel 2002 da RWK Chiu relativamente alla iperplasia surrenale congenita. Questa patologia è caratterizzata da un deficit nella produzione dell'enzima 21-idrossilasi, codificato dal gene CYP21 e coinvolto nella sintesi di glucocorticoidi e mineralcorticoidi. Una marcata o totale assenza dell'enzima causa principalmente alterazioni nella produzione di cortisolo e aldosterone, con conseguente disfunzioni nell'omeostasi del sodio, e una abnorme produzione di androgeni, il cui effetto avverso si manifesta soprattutto nei feti di sesso femminile causando virilizzazione. Indicazione terapeutica intrauterina per questa patologia è la somministrazione di desametasone. Nel lavoro di Chiu viene descritto come sia stato possibile effettuare una diagnosi con esclusione di rischio per la malattia, attraverso l'amplificazione e l'analisi qualitativa di frammenti mappanti sul gene CYP21. Nella famiglia oggetto di studio, dall'osservazione dei genotipi parentali si evinceva come la madre presentasse una delezione in omozigosi del gene CYP21 ed il padre una mutazione non-senso in eterozigosi sull'esone 8. Inoltre, i genitori mostravano un quadro genotipico differente per una serie di marcatori polimorfici (microsatelliti) in stretto linkage con il sito di mutazione. L'analisi dei microsatelliti sul cfDNA aveva fatto emergere la presenza degli alleli ereditati per via paterna associati all'allele normale del sito di mutazione, permettendo di definire per il feto l'assenza della malattia e, di conseguenza, evitando alla madre la somministrazione di desametasone, farmaco di cui si è già riportata la presenza di pesanti effetti collaterali (24). Ulteriori studi, condotti dallo stesso gruppo del prof. Lo per la diagnosi non invasiva di  $\beta$ -talassemia, hanno mostrato risultati simili, confermando la possibilità, anche mediante l'impiego di tecniche e metodologie diverse, di verificare con elevata accuratezza la presenza o

l'assenza degli alleli paterni ereditati, non condivisi dal genoma materno (25) (26). E' da segnalare che, nel caso in cui si riscontri l'assenza del frammento oggetto di analisi, perché i risultati proposti possano essere ritenuti validi, deve essere sempre verificata la frazione di DNA fetale, solitamente attraverso amplificazioni selettive di sequenze di riferimento, per escludere che la negatività del test non sia dovuta alla bassa quota fetale piuttosto che alla reale assenza della variante genetica mutata.

#### ULTERIORI APPLICAZIONI NELLE MALATTIE MONOGENICHE

Gli approcci finora menzionati sono caratterizzati da due importanti limitazioni: primo, la conferma della presenza o assenza di alleli ereditati per via paterna, non permette di dedurre alcuna informazione riguardo quelli ereditati per via materna; secondo, essi non possono fornire alcuna informazione nel caso in cui padre e madre siano portatori degli stessi alleli (27). Queste osservazioni sono di notevole importanza per comprendere i progressi che si sono avuti nel tempo, sia per quanto riguarda le metodologie applicate allo studio del DNA fetale sia riguardo alle particolari tecnologie che si sono sviluppate e hanno fortemente favorito la crescita delle applicazioni sul cfDNA.

Nel settembre 2008 è stato pubblicato un lavoro sulla rivista PNAS dal gruppo di YMD Lo (28) in cui viene presentato un metodo efficace per la determinazione quantitativa degli alleli ereditati dal feto nel caso in cui entrambi i genitori siano portatori per la stessa mutazione. Questo metodo, definito Relative Mutation Dosage (RMD), consiste nel valutare se gli alleli normale e mutato di una variante osservata sono fra loro bilanciati o sbilanciati nel plasma materno. Attraverso la quantificazione di questo sbilanciamento si può successivamente definire il genotipo fetale. Nel lavoro in esame viene mostrato come attraverso questo metodo sia stato possibile determinare il genotipo fetale in dieci campioni di cfDNA di coppie a rischio per  $\beta$ -talassemia in cui entrambi i genitori erano portatori della stessa mutazione. Dal punto di vista pratico, il metodo si basa sulla valutazione quantitativa dei due alleli della variante osservata, ossia sulla conta nel cfDNA, dei frammenti che portano l'allele mutato e l'allele normale, stimando poi il rapporto tra questi valori. Condizione di partenza nell'applicazione del RMD è la

situazione di eterozigotità dei genitori per la mutazione da studiare, condizione che definisce lo stato di rischio per il feto nelle patologie monogeniche recessive. In assenza di DNA fetale nel plasma materno, gli alleli mutato e normale sono presenti nello stesso numero di copie, in un rapporto pari ad uno. La presenza di un feto con genotipo eterozigote per la mutazione dà luogo ad un rapporto ancora pari ad uno, dato che sia la madre che il feto contribuiscono con quantità equilibrate di alleli mutato e normale. Invece, nel caso in cui il feto abbia ereditato lo stesso allele da entrambi i genitori, nel plasma materno la quota dell'allele trasmesso sarà più elevata rispetto all'altro poichè tutto il contributo fetale sarà in favore di un solo allele. Il rapporto mutato/normale si mostrerà perciò in questo caso sbilanciato, e, più precisamente, risulterà maggiore di uno nel caso in cui il feto sia omozigote per la mutazione o minore di uno nel caso in cui erediti entrambi gli alleli normali. Questa variazione sarà tanto più evidente quanto più elevata è la quota del DNA fetale rispetto alla totalità del cfDNA.

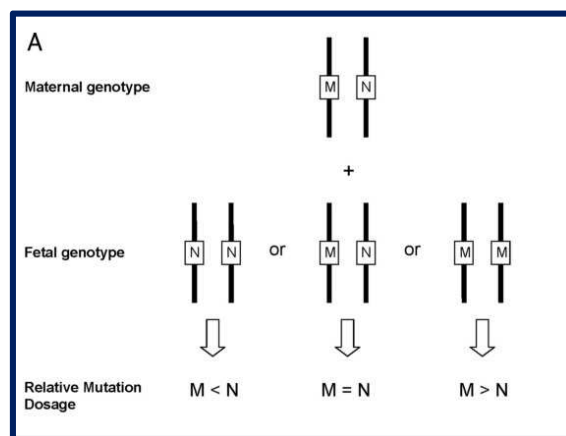


Figura 1: Illustrazione schematica del Relative Mutation Dosage (da Lun et al., PNAS 2008)

Questo approccio, descritto per le patologie monogeniche recessive, è valido e può essere applicato anche alle malattie dominanti e alle malattie X-linked. Nel caso di una madre eterozigote per una mutazione dominante (padre omozigote per l'allele normale) il rapporto fra l'allele mutato e quello normale può indicare: se pari ad uno, un genotipo fetale eterozigote che definisce lo stato di malattia a carico del feto; se inferiore ad uno, la trasmissione da parte materna dell'allele normale con esclusione del rischio per il nascituro. Similmente, nel caso



di una patologia X-linked, la trasmissione dell'allele normale o mutato potrà essere diagnosticata attraverso la valutazione dello sbilanciamento che potrà assumere, in questo caso, valori, rispettivamente, minore o maggiore di uno ( figura 2).

Disease type	Maternal genotype	Fetal genotype	Maternal plasma DNA						M/N ratio
			Maternally derived alleles		Fetally derived alleles		(Maternal + fetal) alleles		
			M (copies)	N (copies)	M (copies)	N (copies)	M (copies)	N (copies)	
Auto-somal diseases			90	90 <sup>†</sup>	0	20 <sup>†</sup>	90	110	0.82:1
			90	90	10	10	100	100	1:1
			90	90	20	0	110	90	1.22:1
X-linked diseases			90 <sup>†</sup>	90	0	10 <sup>†</sup>	90	100	0.90:1
			90	90	10	0	100	90	1.11:1

Figura 2: Schema RMD per le malattie autosomiche e X-linked

E' stato messo in evidenza come l'approccio dell'RMD, in precedenza descritto, si basi sulla possibilità di una valutazione quantitativa del numero di copie degli alleli normale e mutato relativi ad un sito polimorfo. Ciò presuppone, dal punto di vista tecnico, la disponibilità di una tecnologia ad alta risoluzione che consenta di rilevare un segnale, in questo caso rappresentato dall'allele fetale, anche quando la sua presenza è mascherata da un forte rumore di fondo, rappresentato dagli alleli materni predominanti nel cfDNA.

Nel lavoro in cui è stato descritto per la prima volta l'RMD le indagini sperimentali sono state effettuate attraverso l'utilizzo della tecnologia della Digital PCR. Questa tecnica, descritta nel 1999 da Vogelstein e Kinzler (29), era stata proposta dapprima come strumento utile in campo oncologico in quanto capace di mettere in evidenza, all'interno di un'ampia popolazione cellulare, la presenza di mutazioni somatiche a carico di un numero esiguo di cellule. Tuttavia, grazie alla sua efficacia e alla semplicità di applicazione, essa ha trovato immediatamente impiego in numerosi campi di ricerca e nella pratica clinica. La metodica della

Digital PCR prevede una prima fase in cui il campione del DNA in esame viene amplificato per il frammento contenente la mutazione da studiare. L'amplificato ottenuto viene diluito e portato a bassissime concentrazioni, tali che, attraverso una sua distribuzione su una piastra da amplificazione o su un analogo supporto a pozzetti, possa essere distribuita mediamente una molecola del frammento di DNA amplificato in ciascun pozzetto, dove vengono inoltre dispensati anche tutti i reagenti utili per la reazione di amplificazione. L'analisi quantitativa degli alleli avviene attraverso l'utilizzo di sonde fluorescenti allele-specifiche che, legandosi specificamente ai frammenti amplificati e opportunamente sollecitate, permettono di evidenziare quanti pozzetti contengano l'allele normale, l'allele mutato, entrambi, o quanti di questi siano vuoti. I dati così ottenuti permettono di ricavare il genotipo fetale attraverso la successiva applicazione dell'approccio RMD. Nei suoi primi lavori, Vogelstein proponeva, come supporto da utilizzare per l'amplificazione, piastre da 96-384 pozzetti: l'avanzamento tecnologico odierno consente la produzione di supporti contenenti fino a 20.000 micropozzetti, favorendo un maggior potere risolutivo e una più robusta valutazione statistica dei dati ottenuti, grazie all'elevato numero di reazioni attuabili in uno stesso esperimento.

In un lavoro del 2011, il gruppo di YMD Lo descrive un protocollo di diagnosi non invasiva di emofilia, patologia X-linked, attraverso l'utilizzo della Digital PCR e dell'approccio RMD (20). In un campione costituito da 12 cfDNA prelevati da madri in attesa di un feto maschio, è stata condotta l'analisi per determinare la presenza o l'assenza della mutazione causativa nel feto. I dati emersi hanno evidenziato un'accuratezza della metodica del 100% nella definizione del genotipo fetale. I risultati di questa ricerca sono una dimostrazione di come la tecnologia della Digital PCR possa essere favorevolmente applicata alle tecniche di NIPD (Noninvasive Prenatal Diagnosis).

Un'ulteriore esempio di applicazione di questa tecnologia relativamente alla diagnosi di malattie monogeniche, è stato messo in evidenza da A.N. Barrett e dal suo gruppo di ricerca in uno studio di fattibilità per la diagnosi non invasiva in coppie a rischio per anemia falciforme (30). In un campione totale di 65 cfDNA, sono stati determinati correttamente circa l'80% dei genotipi fetali attraverso Digital

PCR e successivo approccio RMD. Nonostante il dato possa apparire meno preciso rispetto al lavoro sull'emofilia, è importante mettere in risalto alcune differenze, che valorizzano la presente ricerca: in primo luogo, i campioni di DNA fetale analizzati da Barrett erano sia di sesso maschile che femminile; oltre a ciò, i genitori erano entrambi portatori di mutazioni identiche. In questo lavoro Barrett mette in evidenza inoltre due aspetti fondamentali da tenere in considerazione nelle analisi del DNA fetale: la necessità di un processamento immediato del campione (separazione del plasma entro le 8 ore dal prelievo), e l'importanza di una sufficiente frazione di DNA fetale nel cfDNA, necessaria per la corretta determinazione del genotipo fetale. Infatti, presi in considerazione solo i campioni con quota fetale superiore al 7%, Barrett descrive un'accuratezza del metodo pari al 100%, dimostrando sia la robustezza del protocollo proposto sia l'importanza della quantità del DNA fetale presente nel plasma materno, e indicando come l'analisi di campioni di cfDNA, prelevati in epoca gestazionale molto precoce, possa essere caratterizzata da minore accuratezza a causa della scarsa frazione fetale.

In sintesi, come già messo in evidenza nei casi precedentemente proposti, i due parametri che maggiormente influenzano l'efficacia del Relative Mutation Dosage sono: la percentuale di quota fetale contenuta nel plasma materno, e la capacità d'analisi della tecnologia impiegata nei processi di genotipizzazione. Maggiore è la profondità d'analisi dalla tecnologia utilizzata, minore sarà la quota fetale minima analizzabile.

## NEXT GENERATION SEQUENCING

### CARATTERISTICHE DEL SEQUENZIAMENTO NEXT GENERATION

Nel 1977 il fisico e biochimico Walter Gilbert nei laboratori di Biochimica e Biologia Molecolare della Harvard University di Cambridge nel Massachusetts, e in contemporanea Frederick Sanger, chimico del Dipartimento Biologia Molecolare del Medical Research Council Laboratory alla Cambridge University in Inghilterra, sono stati gli autori di due lavori, pubblicati dalla rivista PNAS a distanza di pochi mesi, che hanno meritato ad entrambi l'assegnazione condivisa del premio Nobel per la chimica, per le loro scoperte relative al sequenziamento del DNA (31) (32). Il sequenziamento del DNA, cioè la tecnica di lettura della sequenza delle basi che costituiscono l'acido nucleico, da quel momento ha occupato un posto di primaria importanza tra le tecniche di biologia molecolare, e tutt'oggi, lo stesso metodo proposto da Sanger, che già dopo pochi anni aveva assunto un ruolo preminente rispetto a quello di Gilbert, è alla base del funzionamento dei più moderni sequenziatori automatici a capillari. Per anni la tecnica proposta da Sanger è stata l'unica utilizzata per il sequenziamento del DNA. Solo nel 2005 si è affacciato sul mercato un sistema tecnologico alternativo, denominato Second-Generation DNA Sequencing (o alternativamente Next Generation Sequencing, NGS, o Massive Parallel Sequencing, MPS). In tempi brevissimi, diverse piattaforme tecnologiche basate su questo nuovo sistema di sequenziamento sono state sviluppate e via via perfezionate, dando origine ad una vera e propria rivoluzione nel campo del sequenziamento degli acidi nucleici, sia per quanto riguarda la quantità dei dati prodotti sia per la velocità con cui essi vengono generati. Inoltre, negli ultimi anni si è assistito ad una considerevole diminuzione dei costi di processamento dei campioni, fattore essenziale che ha favorito l'ampia diffusione di questi sistemi.

Un esempio esplicativo della differente capacità produttiva delle due tecnologie 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 (con la prima pubblicazione nel 2001), 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. 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.

Le tecnologie NGS presentano diversi vantaggi rispetto al sequenziamento di tipo Sanger. In primo luogo, un sistema NGS può produrre molte gigabasi di sequenze in minor tempo e con costi nettamente inferiori rispetto al metodo Sanger. Il grado di miniaturizzazione raggiungibile in queste tecnologie, infatti, è di gran lunga superiore rispetto a quello possibile nella tecnologia a capillari e ciò permette il processamento contemporaneo di una quantità di frammenti molto elevata; nella tecnologia a capillari è invece presente un limite fisico, dato dalla struttura stessa dei capillari, che non può essere superato. In secondo luogo, il passaggio di amplificazione clonale dei frammenti nella tecnologia NGS non prevede l'utilizzo di plasmidi per il clonaggio (come era avvenuto nel caso del sequenziamento "shotgun" nel Progetto Genoma) ma viene eseguito mediante un processo automatizzato inserito nel protocollo sperimentale (emulsion PCR, bridge PCR), il che favorisce un notevole risparmio economico e di tempo. Altra caratteristica che differenzia profondamente le tecnologie Next Generation da Sanger è la capacità di determinare in maniera accurata la presenza di alleli a bassa frequenza. Il dato prodotto da un sequenziatore a capillare deriva dalla osservazione dell'intensità e della lunghezza d'onda della fluorescenza emessa globalmente da tutti i frammenti di pari lunghezza che attraversano il capillare; per sua natura, la rilevazione del segnale dato dalla sovrapposizione delle emissioni di tutti i fluorocromi, fa sì che l'effetto del gruppo fluoroforo che marca la maggior

parte dei frammenti, sovrasti l'effetto di un eventuale gruppo alternativo meno rappresentato, causandone la mancata rilevazione da parte dello strumento. Nelle NGS ciascun frammento viene invece sequenziato all'interno di un microreattore in maniera indipendente rispetto agli altri: i dati così prodotti vengono quindi acquisiti non in maniera massiva, ma per singolo reattore. Questo sistema rende possibile la determinazione della sequenza specifica del frammento contenuto in ciascun reattore, e consente di mettere in evidenza le varianti presenti. In questo modo è possibile dedurre la frazione di ciascun allele relativa al campione di partenza, poiché il numero delle volte che esso viene sequenziato è proporzionale alla sua quantità relativa originaria. Un ulteriore vantaggio offerto dalle NGS, sia in termini di rapidità nella produzione dei dati che economici, è la possibilità di processare diversi campioni in uno stesso esperimento attraverso un metodo di marcatura delle sequenze (barcoding), che, nella fase di analisi informatica, permette di contraddistinguere e separare le sequenze proprie di ciascun campione.

A fronte di una serie di aspetti positivi e dalle proficue ricadute sugli attuali ambiti di ricerca, le NGS presentano delle caratteristiche limitanti che ne condizionano l'applicazione. Una prima limitazione è la lunghezza dei frammenti (reads) che possono essere sequenziati. Generalmente si aggirano intorno alle 200-300 bp, contro anche le 1000 bp di un sequenziatore a capillari. Questo aspetto complica notevolmente l'assemblaggio delle sequenze contigue, soprattutto nei casi di genomi non noti o con ampie zone ripetute o contenenti dei riarrangiamenti, rendendo difficoltosa la separazione in fasi delle varianti. Ulteriore fattore limitante è l'aumentata percentuale di errore (falsi positivi e falsi negativi) nelle chiamate delle basi, se paragonata al sistema Sanger. Tale limite, in gran parte, viene superato attraverso l'aumento del livello di coverage (quantità di sequenze analizzate relative ad un frammento di DNA) e attraverso l'ausilio di polimerasi ad altissima efficienza, ma non del tutto rimosso. E' questo uno dei principali motivi per cui, nella pratica clinico-diagnostica, il sequenziamento Sanger mantiene la sua preminenza. Strettamente legata a questo aspetto è la necessità di continuo sviluppo di algoritmi bioinformatici, che consentano di analizzare con efficacia i dati prodotti. L'aspetto informatico ha profonda rilevanza nelle tecnologie di Next

Generation. Da un lato si riscontra la necessità di continue implementazioni dei software di analisi per rendere sempre più stabili e efficaci le chiamate delle basi, dall'altro sono necessari molti software dedicati per adempiere alle diverse fasi del processo di analisi (gestione dei sequenziatori, analisi dei dati grezzi, filtraggi delle reads, allineamento, visualizzazione, chiamata delle varianti). Questa complessità di livelli di analisi informatica, ciascuno con varie opzioni fruibili, complica notevolmente il compito degli operatori, in particolar modo nei casi in cui sia necessario confrontare risultati ottenuti attraverso differenti piattaforme, poiché ciascuna tecnologia sfrutta programmi e parametri di analisi differenti. Inoltre, l'interpretazione dei dati e la gestione di un sistema informatico così complesso prevede competenze specialistiche e richiede la presenza di figure professionali che spesso non sono presenti nei laboratori clinici.

## TECNOLOGIE NGS

Negli ultimi anni lo sviluppo delle tecnologie NGS ha subito un fortissimo incremento. Diverse compagnie propongono sequenziatori e sistemi NGS, ognuno con caratteristiche specifiche, offrendo agli utilizzatori diverse possibilità di scelta in base alle esigenze sperimentali.

Le caratteristiche che differenziano le tecnologie di sequenziamento massivo fra loro sono principalmente due: la chimica utilizzata nel processo di sequenziamento e la tecnica di amplificazione clonale dei frammenti. Da questi due parametri principali ne derivano poi degli altri: lunghezza delle reads, tempi di corsa, megabasi sequenziate, accuratezza dei dati grezzi, costo per campione, costo iniziale della tecnologia.

Il primo sistema a fare la sua comparsa nel mercato è stato il 454 prodotto dalla Life Sciences - Roche nel 2005 (33). Questo sistema prevede l'amplificazione clonale dei frammenti attraverso una particolare tecnica chiamata Emulsion PCR. Essa consiste nella preparazione di un'emulsione di acqua e olio in cui le microbolle d'acqua svolgono l'azione di microreattori per le reazioni di PCR. In soluzione acquosa vengono preventivamente dispensati tutti i reagenti per la reazione di amplificazione unitamente a delle biglie di supporto (beads) su cui si legano i frammenti di DNA. Quando i reagenti sono presenti in quantità

bilanciate, tra loro e rispetto ai volumi di acqua e olio, secondo la legge della distribuzione di Poisson, essi si distribuiranno in maniera tale che ogni microreattore contenga un frammento di DNA, la polimerasi, i reagenti necessari per l'amplificazione e una biglia di supporto. Sottoposti a delle variazioni cicliche di temperatura, come avviene in una normale reazione di PCR, i microreattori daranno luogo alla produzione di migliaia di frammenti clonali che, grazie a delle sequenze specifiche complementari a delle corte sequenze adese alle biglie, andranno a ricoprirne l'intera superficie (rappresentazione schematica in figura 3).

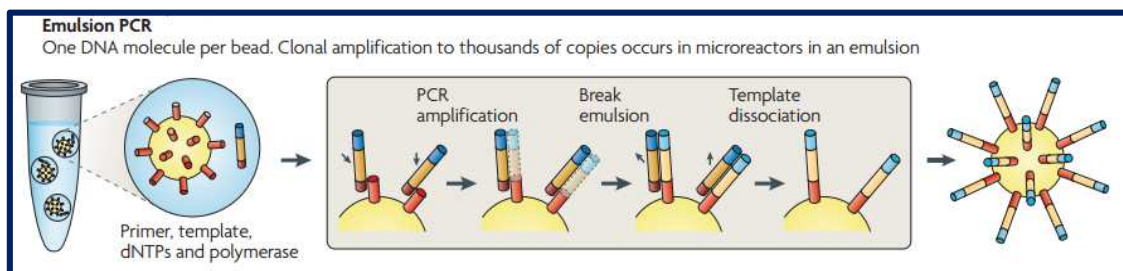


Figura 3: Schema Emulsion PCR (modificato da M. L. Metzker - Nature Review Genetics 2010)

Il sequenziamento nel 454 viene effettuato attraverso il metodo del pirosequenziamento (Fig. 4), che si basa sulla determinazione della presenza del pirofosfato che viene rilasciato al momento della formazione di un nuovo legame fosfodiesterico nella catena nascente del DNA. Il pirofosfato funge da substrato per una luciferasi causando, immediatamente dopo l'incorporazione di un nucleotide, una emissione di energia luminosa che viene rilevata da un sensore.

I punti di forza mostrati dal 454 Roche sono la velocità di processamento e la lunghezza delle reads prodotte, superiore rispetto agli altri sistemi di NGS (oltre 700bp).



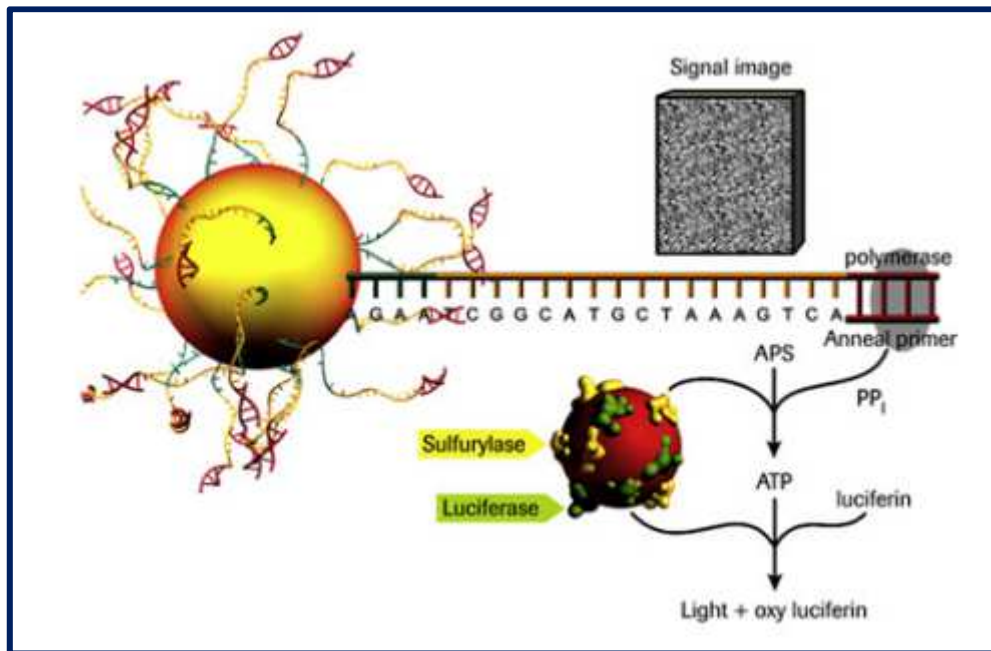


Figura 4: 454 Roche: Rilevazione del segnale nel pirosequenziamento (modificato da T. Tucker et al. 2009 - AJHG)

Un'altra tecnologia NGS è quella proposta dalla Illumina Inc. che ha sviluppato un sistema di sequenziamento basato sull'utilizzo di terminatori fluorescenti reversibili. L'amplificazione clonale avviene attraverso la metodica detta "bridge PCR". In una prima reazione, vengono ligate ai frammenti da amplificare delle sequenze di riconoscimento universali dette "adattatori". I frammenti successivamente vengono distribuiti su un supporto rigido a cui si legano grazie alla presenza di sequenze complementari agli adattatori, ancorate alla superficie. Ciascun frammento viene clonato attraverso la formazione delle caratteristiche "strutture a ponte", che danno luogo alla costituzione dei "cluster", microaree circoscritte del supporto nelle quali sono presenti i frammenti clonati da un unico template di partenza (Fig.5).

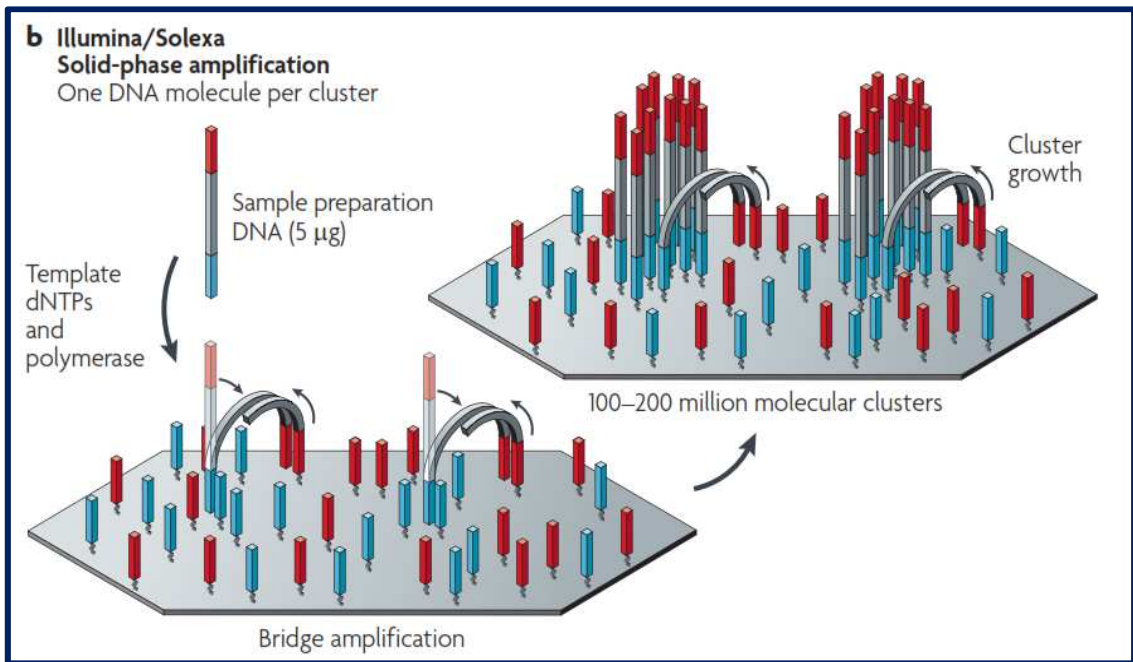


Figura 5: Amplificazione clonale attraverso Bridge PCR (modificato da M. L. Metzker - Nature Review Genetics 2010)

Il sequenziamento avviene attraverso l'utilizzo di nucleotidi con gruppi terminatori fluorescenti reversibili. Ad ogni ciclo del processo di sequenziamento (flows), sulla superficie del supporto su cui sono presenti i cluster viene riversata una miscela contenente i terminatori e una DNA polimerasi modificata. Ad ogni flow si ha l'incorporazione di uno dei dNTP fluorescenti, la cui successiva emissione luminosa viene rilevata e convertita dal sistema in immagini globali dell'intera superficie del vetrino di supporto (Fig. 6). In seguito, avviene il rilascio per via chimica dei gruppi terminatori che consente la prosecuzione del processo, con l'incorporamento di ulteriori nucleotidi e l'allungamento della sequenza .

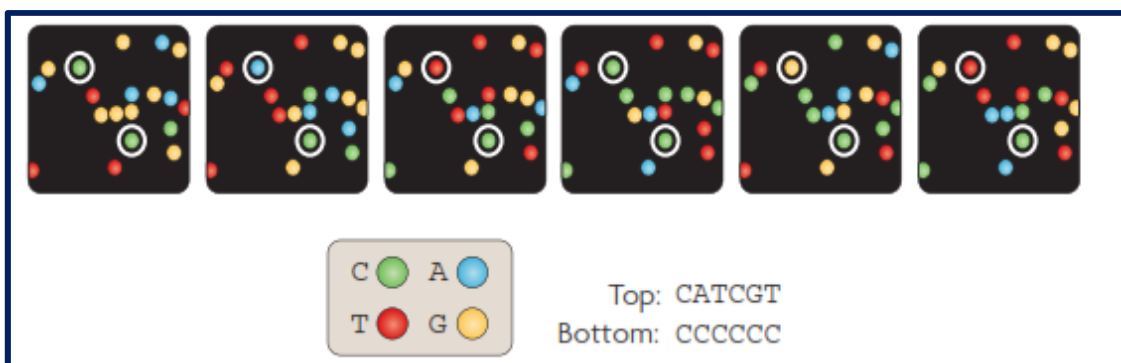


Figura 6: Incorporazione dei dNTP fluorescenti e lettura della sequenza (modificato da M. L. Metzker - Nature Review Genetics 2010)

Il sistema proposto da Illumina è ad oggi il più diffuso tra le tecnologie NGS; questo grazie, sia all'elevata accuratezza dei dati prodotti attraverso questa tecnologia, sia alla 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.

Un'altra tecnologia utilizzata fino a pochi anni fa è il SOLiD® Next-Generation Sequencing proposto dalla Life Technologies. Questo sistema prevede un'amplificazione clonale del campione attraverso emulsion PCR, analogamente al 454 Roche, mediante l'utilizzo di beads di supporto, le quali, a differenza del 454, dopo l'amplificazione vengono distribuite e immobilizzate su supporto solido. Il protocollo SOLiD si differenzia nel processo di sequenziamento per l'utilizzo di una ligasi e non di una polimerasi. Il sequenziamento, infatti, prevede l'utilizzo di ottameri marcati con fluorocromi, che si legano selettivamente alla sequenza stampo e il cui primo nucleotide viene legato all'innesco. L'emissione luminosa viene registrata e in seguito viene eseguito un taglio tra il 5° e 6° nucleotide, che elimina il marcatore fluorescente e permette l'inserimento di un nuovo ottamero. La valutazione complessiva dei frammenti prodotti permette di ricostruire la sequenza del template d'origine.

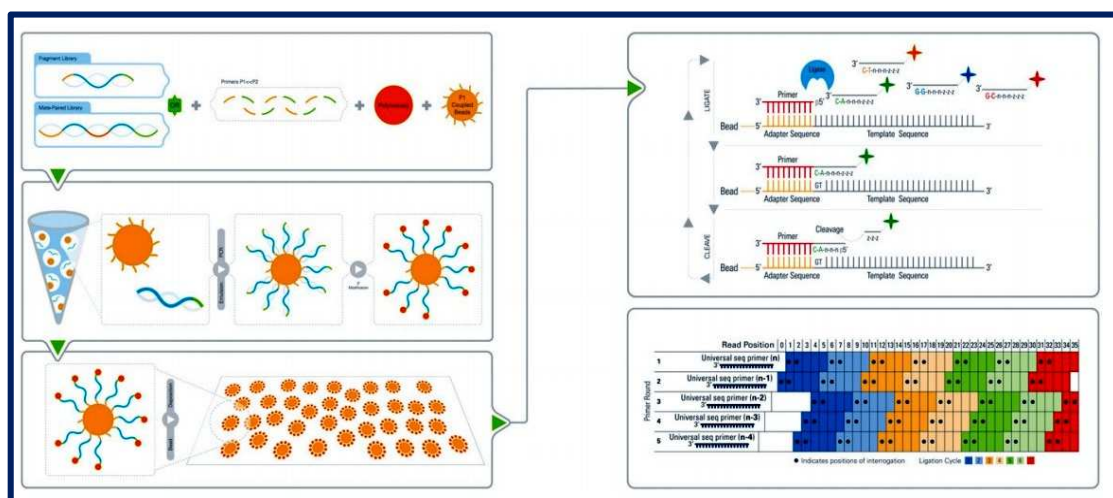


Figura 7: Schema del protocollo SOLiD (modificato da T. Tucker et al. 2009 - AJHG)

## ION TORRENT TECHNOLOGY

In tempi più recenti, e precisamente nel Luglio 2011 (34), è stato pubblicato un lavoro nel quale viene descritto un nuovo modello di tecnologia di sequenziamento massivo. Primo autore del lavoro è J.M. Rothberg, ingegnere chimico-biomedico, già noto nell'ambito delle tecnologie Next Generation, in quanto fondatore della 454 Life Sciences Corporation, e ideatore del primo sequenziatore NGS. Rothberg descrive in questo lavoro la Ion Torrent Technology (in particolare il sistema Ion PGM™ System - Personal Genome Machine, unico strumento prodotto dalla Ion Torrent Inc. al momento della pubblicazione, figura 8), una tecnologia di massive parallel sequencing con peculiarità tali da differenziarla dagli altri sistemi NGS.



Figura 8: Ion PGM™ System - Personal Genome Machine

Il PGM™ nasce come sequenziatore di tipo "benchtop", letteralmente da "bancone" con ridotte capacità di processività rispetto alle strumentazioni concorrenti, con l'intento di rendere più economica, quindi maggiormente fruibile, la tecnologia NGS per tutti i progetti (targeted DNA sequencing, targeted RNA sequencing, microbial sequencing) in cui la quantità di basi da sequenziare è notevolmente inferiore rispetto al sequenziamento di un intero genoma umano. Anche il protocollo Ion Torrent prevede l'amplificazione clonale del campione attraverso Emulsion PCR ma il processo di sequenziamento presenta due aspetti essenziali che lo caratterizzano rispetto alle altre tecnologie: a) l'utilizzo dei

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

Nucleo centrale dell'intero sistema di sequenziamento è l'Ion Chip (Fig. 9). Esso è costituito da uno strato superficiale superiore, in cui sono ricavati dei pozzetti appositamente strutturati per poter accogliere le beads e i reagenti per il sequenziamento. Ciascun pozzetto ha diametro di pochi micron, tale che in esso possa trovare spazio una sola biglia. Questo primo strato superficiale poggia su un secondo costituito da semiconduttori, che permette la trasmissione dei segnali allo strato ancora sottostante. Quest'ultimo è strutturato come una piastra di sensori, uno corrispondente a ciascun pozzetto, che hanno la capacità di registrare le piccolissime variazioni di pH che avvengono all'interno del pozzetto durante il sequenziamento, trasformandole in differenze di potenziale, ovvero, in dati digitali.

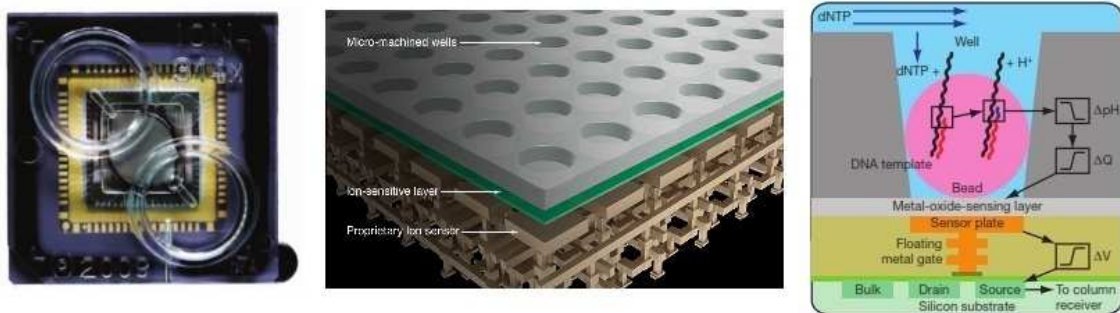


Figura 9 Particolari dell'Ion Chip. Nell'immagine centrale, sezione del Chip con evidenti i pozzetti superficiali. A destra, schema del processo di rilevazione del segnale.

Il processo di sequenziamento è simile a quelli precedentemente descritti, caratterizzato da un susseguirsi ciclico sulla superficie del chip di lavaggi e flussi (flows) di soluzioni, contenenti ciascuno dei quattro diversi nucleotidi. In questo modo, quando viene dispensato il nucleotide complementare alla prima base libera sul frammento stampo, esso è incorporato alla sequenza nascente. Poiché la bead, su cui è avvenuta l'amplificazione clonale, presenta sulla propria superficie frammenti tutti uguali, il legame della nuova base si verifica contemporaneamente su migliaia di frammenti. La formazione di ogni nuovo legame fosfodiesterico determina l'idrolisi del gruppo trifosfato del nucleotide entrante con la liberazione

netta di un protone ( $H^+$ ). 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 (Fig. 10). 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.

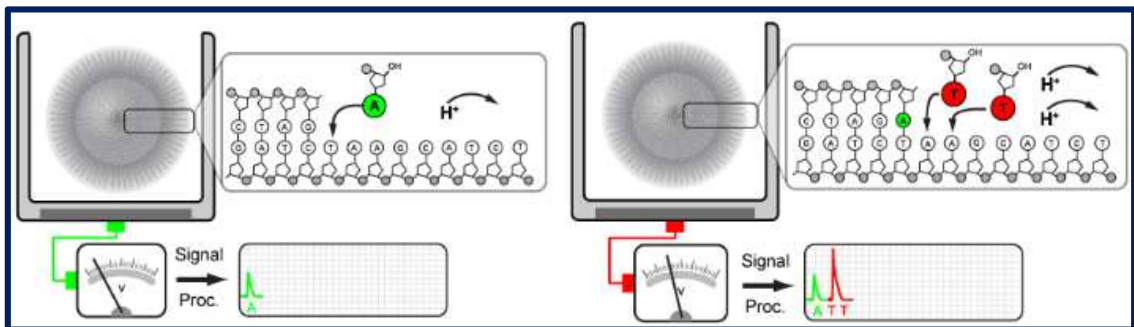


Figura 10: Schema di sequenziamento nell'Ion Chip. Primers, Dna polimerasi e beads a cui è legato il template sono contenuti all'interno di un pozzetto. Ciclicamente fluiscono sulla superficie del chip i quattro nucleotidi. La differenza di potenziale viene registrata per ciascun pozzetto del chip ad ogni flow e, nel caso si registri una sua variazione, questa è tradotta nella chiamata della base.

Questo dato viene rappresentato attraverso un grafico definito Ionogramma, rappresentato in figura 11, in cui è visualizzato il numero delle basi incorporate rispetto ai flows eseguiti dalla macchina. Il vantaggio offerto da questo sistema consiste nella possibilità di introduzione multipla di basi nella sequenza nascente, contrariamente a quanto avviene nel sistema Illumina in cui l'incorporazione per ogni flow è di una singola base. Tuttavia all'aumentare del numero delle basi identiche consecutive, il sistema Ion Torrent mostra una diminuzione di accuratezza nella chiamata delle basi stesse.

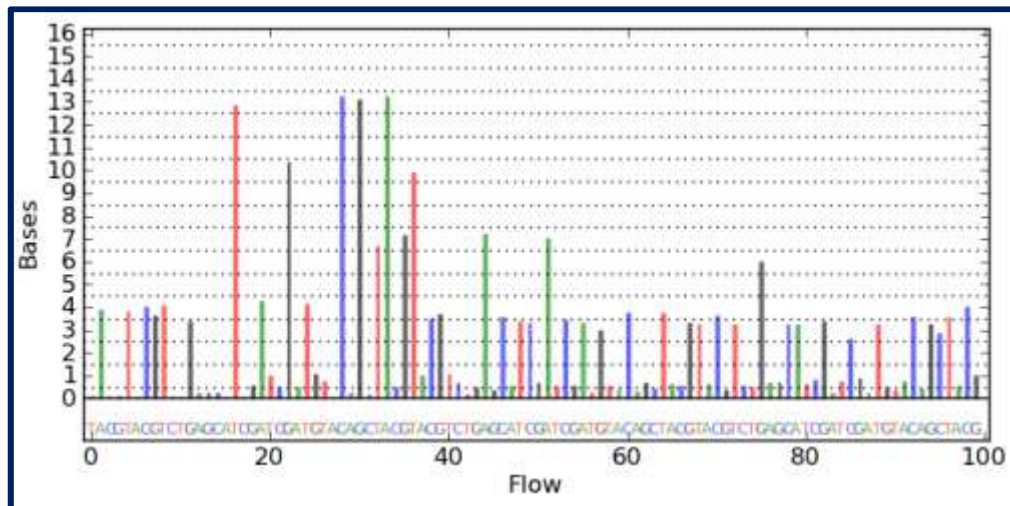


Figura 11: Esempio di ionogramma prodotto dal PGM™ Ion Torrent

E' stato in precedenza indicato come la variazione del pH, conseguente all'inserimento di nuovi nucleotidi, venga misurata dai sensori posizionati al di sotto dei pozzetti, venga convertita in differenza di potenziale e espressa come un dato digitale. Il processo totale di generazione e misura del segnale (da quando viene riversata la soluzione contenente i nucleotidi sulla superficie del Chip fino alla avvenuta misura del pH) dura poco più di 4 s. Successivamente, attraverso il fluire di una soluzione di lavaggio, necessaria per eliminare dai pozzetti i reagenti già utilizzati, il Chip viene preparato per un successivo flow che produce l'incorporazione del successivo nucleotide. Questo processo viene ripetuto ciclicamente per 500 volte consentendo la lettura di reads lunghe circa 200bp. A differenza delle tecnologie che sfruttano i segnali di fluorescenza, il sistema Ion Torrent non prevede le reazioni chimiche necessarie per la eliminazione dei gruppi fluorofori legati ai nucleotidi e ciò favorisce la rapidità del processo di sequenziamento che caratterizza questa tecnologia. Un aspetto vantaggioso offerto dalla tecnologia Ion Torrent è la disponibilità di chip con capacità scalari che consente di modulare il sistema in base alle necessità sperimentali. Nel più piccolo dei chip attualmente disponibili per PGM™ (Ion 314™ Chip v2) si contano circa  $1.2 \cdot 10^6$  pozzetti, ciascuno del diametro di  $3.5 \mu\text{m}$ , per una totale capacità di sequenziamento che oscilla tra le 30 e le 100 Mb, mentre i chip a più alta processività ( $\sim 12 \cdot 10^6$  pozzetti) consentono il sequenziamento dalle 600 Mb alle 2 Gb (Ion 318™ Chip v2). Ultimamente, in seguito agli sviluppi attuati per questa

tecnologia, oltre al "Benchtop PGM™ System", Life Technologies è presente sul mercato con l'Ion Proton™ Sequencer, strumento caratterizzato dalla medesima tecnica di rilevazione ma capace di maggior processività (10 Gb - Ion PI™ Chip v2), più appropriato per il sequenziamento massivo di esomi e genomi.

Platform	Illumina MiSeq	Ion Torrent PGM	Illumina GAIIx	Illumina HiSeq 2000
Instrument Cost*	\$128 K	\$80 K**	\$256 K	\$654 K
Sequence yield per run	1.5-2Gb	20-50 Mb on 314 chip, 100-200 Mb on 316 chip, 1Gb on 318 chip	30Gb	600Gb
Sequencing cost per Gb*	\$502	\$1000 (318 chip)	\$148	\$41
Run Time	27 hours***	2 hours	10 days	11 days
Reported Accuracy	Mostly > Q30	Mostly Q20	Mostly > Q30	Mostly > Q30
Observed Raw Error Rate	0.80 %	1.71 %	0.76 %	0.26 %
Read length	up to 150 bases	~200 bases	up to 150 bases	up to 150 bases
Paired reads	Yes	Yes	Yes	Yes
Insert size	up to 700 bases	up to 250 bases	up to 700 bases	up to 700 bases
Typical DNA requirements	50-1000 ng	100-1000 ng	50-1000 ng	50-1000 ng

Figura 12: Confronto tra diverse piattaforme Illumina e l'Ion Torrent PGM™ System (modificato da Quail et al. 2012)

Due ulteriori strumenti vengono impiegati nel sistema Ion Technology: Ion OneTouch DL, che svolge il compito di effettuare automaticamente l'Emulsion PCR, e il OneTouch ES, che viene utilizzato per l'arricchimento/purificazione del campione successivamente alla reazione di amplificazione clonale (Fig. 13).



Figura 13: Ion OneTouch ES e Ion OneTouch DL



## APPLICAZIONI DEL SEQUENZIAMENTO NEXT GENERATION

Negli ultimi 15 anni le tecniche di Next Generation Sequencing hanno subito un rapido incremento trovando applicazione nei diversi campi della biologia molecolare: dalla caratterizzazione di nuovi geni malattia al sequenziamento di nuove specie di organismi, dalla ricerca di mutazioni somatiche in ambito oncologico al sequenziamento di DNA negli studi di farmacogenomica. Le tecnologie MPS inoltre hanno giocato un ruolo fondamentale negli studi relativi al DNA plasmatico e tuttora svolgono un ruolo di primaria importanza nell'ambito della diagnosi prenatale non invasiva. E' stato già citato in precedenza come l'utilizzo della Digital PCR sia stato di grande aiuto per l'analisi del DNA fetale libero, grazie alle sue caratteristiche tecniche che ne consentono l'applicazione nell'analisi del cfDNA. Anche le tecnologie NGS permettono questo tipo di approccio, con l'ulteriore vantaggio di una maggior profondità di analisi rispetto a quella possibile con la Digital PCR. Il numero delle molecole analizzabili mediante NGS è infatti superiore rispetto alla Digital, e questo conferisce una maggior risoluzione d'analisi, caratteristica molto utile in tutti i casi in cui, come per il DNA fetale, l'elemento ricercato è presente in bassissime concentrazioni.

### TECNOLOGIE NGS NELLA DIAGNOSI DELLE ANEUPLOIDIE

Il settore della NIPT in cui l'utilizzo delle tecnologie di NGS ha ottenuto finora i maggiori successi è quello della diagnosi delle più frequenti aneuploidie fetali. Negli ultimi anni, diversi gruppi di ricerca si sono dedicati allo sviluppo di protocolli per la diagnosi delle aneuploidie attraverso sequenziamento massivo parallelo, con particolare riferimento per le aneuploidie dei cromosomi 21, 18, 13 e sessuali. I risultati emersi dai diversi trial hanno mostrato un'elevata efficacia della tecnologia nella definizione di queste cromosomopatie, in particolare nella determinazione della trisomia 21 (T21).

M. Ehrich e il suo gruppo di lavoro, nel 2011 hanno pubblicato uno studio condotto su 480 campioni di cfDNA con l'obiettivo di stilare un protocollo basato su MPS per la diagnosi della Sindrome di Down. Escludendo i campioni con concentrazioni di DNA fetale molto basse o con caratteristiche non ottimali, i

risultati proposti sono del 100% riguardo la sensibilità e 99.7% per la specificità (35). Parimenti, il gruppo di YMD Lo nel 2012, sfruttando un diverso approccio, un target sequencing per all'analisi di circa 3000 SNP distribuiti su diversi cromosomi, mirato alla valutazione di eventuali sovradosaggi cromosomici, conferma l'affidabilità dei protocolli d'analisi del cfDNA nella diagnosi di trisomia 21 (36), evidenziando anche la versatilità del DNA da plasma riguardo ai possibili protocolli utilizzabili per la sua analisi. I risultati positivi ottenuti per la trisomia 21, replicati e validati da diversi gruppi di ricerca attraverso approcci diagnostici differenti, hanno consentito l'inserimento di questa metodica tra le indagini di screening per le aneuploidie fetali nelle donne ad alto rischio, sin dal 2011, in diversi stati del Nord America, dell'Asia e dell'Europa (37). Essi vanno ad integrarsi ai test ecografici, alla valutazione dell'età materna e alle misurazioni dei marker molecolari per la determinazione del fattore di rischio, aumentandone in maniera sostanziale l'attendibilità, che si attesta su valori superiori al 99% (in particolare 98.6-100% riferita ai falsi negativi e 99.7-100% ai falsi positivi).

Più difficoltosa, rispetto al cromosoma 21, si è mostrata la determinazione della sovrarappresentazione dei cromosomi 13 e 18, come dimostra lo stesso gruppo di Lo, in un lavoro pubblicato su PLOS ONE (38). I valori di sensibilità e specificità presentati si sono rivelati inferiori rispetto a quelli ottenuti per la Sindrome di Down. L'autore identifica la causa di questo risultato con l'elevata presenza, nelle sequenze analizzate, di regioni ricche in GC, suggerendo l'applicazione di un algoritmo correttivo che potesse incrementare la precisione del processo di allineamento delle reads e rettificare, attraverso un approccio statistico, eventuali deviazioni dei risultati rispetto ai dati reali. Attraverso questo metodo i valori si innalzano rispettivamente a 100% di sensibilità e 98.5% di specificità, mostrando come, attraverso appropriati protocolli e algoritmi dedicati, anche la diagnosi delle trisomie dei cromosomi 13 e 18, mediante l'analisi del cfDNA e l'applicazione delle tecnologie NGS, si riveli efficace.

Ad oggi i valori di sensibilità si attestano al 97,4% per la trisomia del cromosoma 18 (di poco inferiore rispetto alla T21) e all'83,3% per la trisomia del 13. I dati ottenuti per le aneuploidie dei cromosomi sessuali sono meno consistenti (aneuploidie X 60-100%).

Tuttavia, l'introduzione nella pratica clinica di routine, di questi test non invasivi, è ancora difficoltosa. Tra i fattori che si possono annoverare tra le possibili cause troviamo i tempi di esecuzione dei protocolli delle tecniche NGS, più lunghi rispetto a quelli tradizionali, e i costi, sia degli strumenti che dei reagenti, che per essere abbattuti efficacemente richiedono elevati numeri di campioni da analizzare. Un'ulteriore difficoltà che si riscontra nelle indagini sul DNA libero, a causa della ridotta quantità del DNA fetale, è data dalla necessità di avere a disposizione una cospicua quantità di plasma da processare e una sufficiente frazione fetale in esso contenuta, condizioni non sempre possibili. A conferma di ciò, si riscontra in molti lavori l'esclusione dalle analisi di diversi campioni a causa dell'esigua quantità di plasma o della scarsa concentrazione del cfDNA in essi presente. Questo dato mette in evidenza un ulteriore fattore limitante di cui bisogna tenere conto nelle analisi del cfDNA, ovvero la presenza del rischio di fallimento, stimato attualmente intorno al 4-5%.

Ad oggi, esistono diverse aziende al mondo che offrono servizi di screening prenatale per le aneuploidie cromosomiche tra le quali possiamo citare: Sequenom, la prima in assoluto a proporre test di screening non invasivi attraverso il cfDNA, Verinata Health, Ariosa Diagnostics, Natera Inc. e BGI (NIFTY). I dati emersi dagli studi di validazione svolti finora da queste società, oltre ai più recenti dati presentati in letteratura, fanno intuire come, attraverso il continuo miglioramento e sviluppo dei già efficaci protocolli d'indagine, sarà possibile raggiungere elevati standard di accuratezza, sensibilità e specificità per l'analisi delle diverse aneuploidie mediante le tecnologie NGS. La combinazione di questi metodi con le tecniche ultrasonografiche, i marcatori sierici e la età materna potrà aumentare notevolmente l'accuratezza negli screening prenatali, abbattendo in maniera consistente soprattutto la presenza di falsa positività (39).

Recenti pubblicazioni hanno messo in evidenza come le indagini sul cfDNA basate su tecniche MPS si siano rivelate utili, non solo per le anomalie che coinvolgono interi cromosomi, ma anche per la diagnosi di traslocazioni, microdelezioni e microduplicazioni, con una definizione spesso maggiore rispetto alle tecniche tradizionali. A dimostrazione di questo può essere citato il recente

studio pubblicato da Srinivasan nel 2013 in cui si descrive l'identificazione di una microdelezione a carico del feto dell'ordine dei 300Kb (40).

I dati presenti in letteratura dimostrano come lo studio del cfDNA attraverso le tecnologie NGS possa essere considerato un efficace metodo di screening per le trisomie dei cromosomi 21, 18 e 13. Queste tecnologie inoltre, grazie al continuo sviluppo di cui sono oggetto, in un prossimo futuro potranno essere introdotte nella pratica clinica sia per lo screening delle aneuploidie cromosomiche, ad oggi ancora caratterizzate da bassi livelli di sensibilità, sia per l'analisi di anomalie del cariotipo più circoscritte come microduplicazioni e microdelezioni.

#### TECNOLOGIE NGS E DIAGNOSI DI MALATTIE MONOGENICHE

E' stato già messo in evidenza come la difficoltà principale nella diagnosi non invasiva attraverso DNA da plasma sia quella di far emergere, dalla globalità dei dati di prevalente origine materna, i caratteri genetici del feto.

La condizione più semplice è quella in cui i caratteri fetali non sono condivisi con la madre, ma sono riconducibili esclusivamente ad una origine paterna. Tra questi infatti possiamo includere i casi già descritti della definizione del fattore Rh fetale, della determinazione del sesso fetale, o della determinazione della ereditarietà della mutazione paterna nelle patologie monogeniche dominanti in cui la presenza della variante mutata nel plasma materno fornisce l'indicazione dello stato patologico del feto, che in questo caso viene quindi definito esattamente. Diversamente, nel caso di una patologia monogenica recessiva in cui padre e madre siano portatori di differenti mutazioni, la presenza della variante mutata paterna nel cfDNA permette di diagnosticare un possibile stato di portatore o affetto, ma non consente una definizione precisa del genotipo fetale a causa della mancata informazione riguardo alla mutazione materna.

Il caso in cui le indagini di NIPD sono maggiormente complicate, è quello in cui feto e madre condividono la stessa mutazione. **In questi casi**, poiché entrambi i DNA presentano lo stesso genotipo, è complicato metter in evidenza l'effettivo apporto del DNA fetale. Questa condizione si verifica, ad esempio, per le patologie monogeniche dominanti in cui l'allele mutato è portato dalla madre, o

le patologie monogeniche recessive in cui padre e madre condividono la stessa mutazione. In questi casi, il metodo del Relative Mutation Dosage si è mostrato uno strumento utile per la definizione del genotipo fetale, così come messo in evidenza negli studi già citati sull'emofilia (20) o sull'anemia falciforme (30). In questi lavori, gli sbilanciamenti allelici erano stati determinati attraverso Digital PCR. Poiché anche le tecnologie Next Generation, parimenti alla Digital PCR, permettono una analisi indipendente di ciascun frammento di DNA presente nel campione iniziale, esse si propongono come sistema di analisi altrettanto valido in questo tipo di studi; esse inoltre hanno la capacità di analizzare un numero di frammenti genomici superiore rispetto alla Digital PCR, e con una maggior profondità d'analisi.

Nei primi anni dopo la scoperta del DNA fetale nel plasma materno, diversi lavori di ricerca sono stati avviati con l'obiettivo di tracciare un profilo definito delle sue caratteristiche. Diverse informazioni si sono ottenute, attraverso l'utilizzo di metodiche tradizionali e Digital PCR, riguardo la lunghezza dei frammenti, la permanenza nel sangue materno, la frazione relativa al cfDNA totale. Non si avevano però certezze riguardo alla rappresentazione del cfDNA nel plasma materno, e fino a pochi anni fa, nonostante il cfDNA si fosse già mostrato un affidabile strumento nella determinazione del sesso fetale nelle malattie X-linked, nella definizione del genotipo Rh, nello studio di malattie de novo, nelle dominanti a trasmissione paterna (16) (24) (25), nessuna evidenza però dimostrava l'intera o parziale rappresentazione del genoma fetale nel plasma materno. Nel 2010 il gruppo di YMD Lo ha condotto uno studio, pubblicato poi su *Prenatal Diagnosis* (41) attraverso il quale viene dimostrato come il genoma fetale sia interamente rappresentato nel plasma materno. Dal punto di vista sperimentale, il risultato è stato ottenuto mettendo a confronto il DNA libero, sequenziato attraverso massive parallel sequencing, e i DNA genomici dei genitori e quello fetale estratto da villo coriale, genotipizzati per i 900.000 SNP del sistema SNP Array 6.0 Affymetrix. Dal confronto, emerge oltre alla conferma dell'intera rappresentazione del genoma fetale nel plasma materno, anche l'evidenza di come la concentrazione frazionaria del cfDNA sia equiparabile per tutti i cromosomi, attestandosi, nel campione in esame (12<sup>a</sup> settimana di gestazione), intorno all'11%.

Nel lavoro qui citato viene messo in risalto un ulteriore importante risultato, riguardo le possibili applicazioni diagnostiche delle NGS per le patologie monogeniche recessive. In questo studio, infatti, è stata coinvolta una coppia a rischio per  $\beta$ -talassemia, in cui il padre era portatore della delezione -CTTT sui codoni 41/42 del gene HBB, e la madre portatrice della mutazione A $\rightarrow$ G in posizione -28. La mutazione paterna è stata facilmente evidenziata perché l'allele mutato era assente nel DNA materno. Ma il dato più rilevante è stato la capacità di definire l'ereditarietà della mutazione per via materna, che ha confermato l'assenza dell'allele mutato -28. L'approccio utilizzato in questo studio è quello del Relative Haplotype Dosage (RHDO). Questo metodo consiste nel considerare gli effetti, non i singoli effetti degli sbilanciamenti di ciascuno SNP, come avviene per l'RMD, ma considerando l'effetto globale di tutti gli SNP in linkage fra loro, verificando l'eventuale sovrarappresentazione di uno dei due aplotipi materni. L'applicazione del metodo RHDO implica la conoscenza degli aplotipi parentali; in questo caso, essi erano stati precedentemente ricavati attraverso l'informazione ottenuta dal DNA estratto dai villi coriali. Il lavoro propone, inizialmente, una suddivisione di tutti gli SNP, classificati in base alla loro diversa informatività riguardo la determinazione della trasmissione materna o paterna. In particolare, quelli utilizzabili per la discriminazione dei due aplotipi materni sono quelli per i quali la madre è eterozigote e il padre omozigote. Questa categoria di SNP viene ulteriormente divisa in due gruppi:

- SNP tipo  $\alpha$ , per i quali l'allele paterno è lo stesso dell'aplotipo materno 1 (Hap1);
- SNP tipo  $\beta$ , per i quali l'allele paterno è lo stesso dell'aplotipo materno 2 (Hap2)

La trasmissione al feto dell'aplotipo 1 causerà una maggior rappresentazione degli SNP di tipo  $\alpha$ , mentre la trasmissione dell'aplotipo 2 quella degli SNP di tipo  $\beta$ . Considerato perciò un segmento di DNA abbastanza ampio da contenere diversi SNP, vengono effettuate le conte cumulative delle reads che mappano sugli SNP di tipo  $\alpha$  e successivamente quelle che mappano sugli SNP di tipo  $\beta$ , inserendo nella somma cumulativa uno SNP per volta, in maniera ricorsiva. Il risultato

ottenuto viene valutato attraverso il Sequential Probability Ratio Test, che determina la significatività statistica dello sbilanciamento tra i due aplotipi. Quando la somma determinata delle reads cumulate supera i valori soglia del test, viene confermata l'ereditarietà del frammento di DNA e dell'aplotipo in esame. Nel caso specifico di questo lavoro, i risultati hanno evidenziato una trasmissione al feto dell'HAP1; poiché l'aplotipo portatore della mutazione materna era l'HAP2, è stato possibile definire lo stato di portatore per il feto, escludendo la presenza della patologia (dati illustrati in figura 14).

I risultati emersi da questa ricerca sono stati di fondamentale importanza: da un lato, per la conferma della applicabilità delle tecnologie NGS ai protocolli di analisi del DNA fetale, dall'altro, per aver suggerito una modalità di indagine capace di definire l'ereditarietà sia paterna che materna attraverso i metodi RHDO e SPRT.

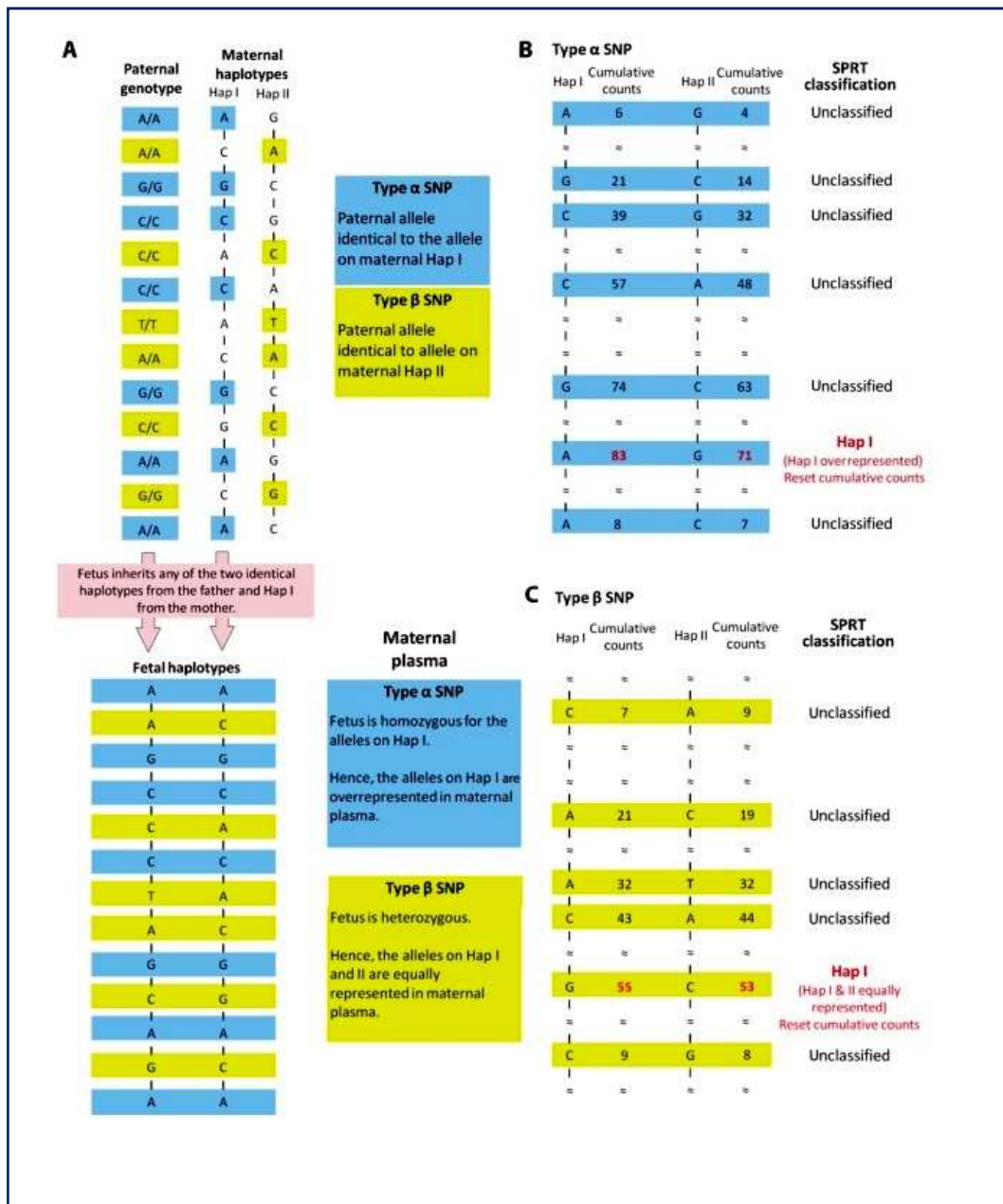


Figura 14: Relative Haplotype Dosage analysis e Sequential Probability Ratio Test ( da YMD Lo et al. 2010)

Ulteriore passo avanti nelle indagini delle malattie monogeniche può essere considerato il successivo lavoro pubblicato dallo stesso gruppo nel 2012 (42). In questo caso viene presentata l'analisi di due famiglie, entrambe a rischio per  $\beta$ -talassemia, in cui i genitori sono portatori di mutazioni differenti. Obiettivo primario di questo studio era verificare la fattibilità di una diagnosi di  $\beta$ -talassemia, non mediante un sequenziamento dell'intero genoma (come era stato fatto nello studio precedente), processo dispendioso sia dal punto di vista sperimentale che



economico, ma attraverso amplificazioni di sequenze selezionate del genoma, metodo conosciuto come "target sequencing". Scopo del lavoro era, quindi, comprendere se il processo di selezione e arricchimento dei target potesse in qualche modo modificare l'uniformità di distribuzione tra le sequenze arricchite e inficiare la robustezza del metodo RHDO.

Mediante l'utilizzo di 54.745 sonde sono stati selezionati frammenti per un totale di 5.1Mb, comprendenti le regioni localizzate tra le estremità 5' e 3' del gene HBB. Gli aplotipi dei genitori sono stati definiti attraverso Digital PCR. L'applicazione del metodo RHDO-SPRT è stata eseguita seguendo lo stesso protocollo applicato nel precedente studio. Nella prima famiglia analizzata, la mutazione paterna è stata messa facilmente in evidenza attraverso la presenza di sequenza mutate nel cfDNA non appartenenti alla madre. La mutazione materna è stata invece esclusa attraverso la valutazione di 51 SNP informativi (5 di tipo  $\alpha$  e 46 di tipo  $\beta$ ) che hanno indicato una maggior rappresentazione dell'aplotipo 2, normale, rispetto all'aplotipo 1 che portava l'allele mutato. Nella seconda famiglia, analogamente, è stato possibile escludere da subito la presenza della mutazione paterna, in virtù dell'assenza della variante mutata nel cfDNA. Tuttavia, maggiori difficoltà si sono riscontrate per la determinazione dell'allele di derivazione materna. In questa famiglia i genitori presentavano aplotipi molto simili (situazione ricorrente nei casi in cui i genitori condividano la stessa mutazione) e la quasi totalità degli SNP si presentavano in eterozigosi sia nella madre che nel padre. La suddivisione in  $\alpha$  e  $\beta$  ( che prevede madre eterozigote e padre omozigote) non aveva evidenziato SNP utili per l'analisi, rendendo di fatto impossibile la definizione dell'allele materno. Il superamento di questa problema è stato possibile mediante un'implementazione del metodo RHDO. Sono stati infatti inclusi nella suddivisione  $\alpha$  e  $\beta$  ulteriori SNP, e più precisamente, quelli con padre e madre entrambi eterozigoti, purchè le quattro fasi degli aplotipi parentali fossero note. In questo modo, è stato possibile identificare ulteriori 24 SNP informativi che hanno permesso di diagnosticare lo stato di portatore del feto. L'aplotipo materno maggiormente rappresentato era infatti quello associato all'allele mutato. Attraverso questi due lavori, Lam e Lo affermano, quindi, la fattibilità di una diagnosi di  $\beta$ -

talassemia attraverso l'analisi del cfDNA, anche quando i genitori condividono le stesse mutazioni, mediante l'applicazione del metodo RHDO/SPRT, e suggeriscono la possibilità di applicazione di questa metodica anche ad altre patologie monogeniche.

Si può affermare, quindi, che i metodi RMD e RHDO si siano dimostrati strumenti utili per l'analisi del DNA fetale: il primo, di più semplice applicazione sia dal punto di vista strumentale che d'analisi, è orientato all'osservazione di una singola mutazione; il secondo, che valuta la trasmissione di interi aplotipi, necessita di strumenti più evoluti (NGS) e indagini più articolate, ma è caratterizzato da maggior potere risolutivo.

Recentemente sono stati pubblicati due lavori dal gruppo di T.E. Papisavva del dipartimento di Genetica Molecolare per la Talassemia di Cipro (43). A Cipro circa il 12% della popolazione è portatrice di mutazioni per la  $\beta$ -talassemia, con l'IVSI-110 che rappresenta quasi l'80% dei casi. T.E. Papisavva propone un metodo di NIPD attraverso la selezione di un gruppo di SNP ad alta eterozigotà, e la loro analisi mediante tecnologie next generation con un approccio di target sequencing. Nel primo dei due lavori presentati, viene descritto uno studio preliminare mirato all'analisi degli SNP localizzati nell'intorno del gene HBB, per la selezione di quelli a più alta eterozigotà che potessero garantire una maggior informatività nella popolazione in esame. La seconda pubblicazione aveva come obiettivo finale mettere in evidenza la presenza della mutazione paterna nel cfDNA in dieci coppie a rischio per  $\beta$ -talassemia, attraverso l'analisi di quattro SNP altamente informativi, mediante un approccio NGS di target sequencing. Le indagini sono state condotte attraverso l'utilizzo del sequenziatore HiSeq 2000 e ogni campione analizzato è stato ripetuto in triplicato, con un coverage medio di circa due milioni di reads per campione. I risultati ottenuti, riguardo la definizione della trasmissione degli alleli paterni per ogni SNP analizzato, indicano una corretta determinazione in 27 casi su 34, (7 falsi tra positivi e negativi); in 6 casi non è stato possibile effettuare l'analisi, o per la non informatività dello SNP o a causa della bassa quota di DNA fetale riscontrata, il cui limite inferiore accettabile era stato imposto a 2,5%. La determinazione dell'allele paterno ereditato ha avuto quindi successo nel 67% dei test effettuati. Tuttavia, poiché gli SNP osservati e il

sito di mutazione sono in stretta associazione tra loro, è stato possibile considerare gli SNP come unico blocco e estendere l'analisi all'aplotipo che li comprende. I risultati emersi dall'analisi degli aplotipi hanno dimostrato il maggior potere di questo secondo approccio, consentendo in 8 famiglie su 10, la determinazione della presenza o dell'assenza della mutazione paterna nel feto. Nelle 2 famiglie restanti questo non è stato possibile a causa della mancata informatività degli SNP.

Il lavoro poc'anzi presentato dimostra, da un lato, la fattibilità di una diagnosi della trasmissione della mutazione paterna attraverso l'utilizzo di SNP e target sequencing, ma, dall'altro, mette in luce alcuni aspetti di questa metodica che necessitano di ulteriori approfondimenti e sviluppi. In primo luogo, nel caso in cui anche la madre sia portatrice dell'allele mutato, la sola analisi della mutazione paterna non può essere applicata e il genotipo fetale non può essere definito. Secondo, l'efficacia dell'approccio attraverso l'analisi degli SNP è tanto più robusto quanto maggiore è il numero degli stessi coinvolti; l'incremento del loro numero riduce, infatti, il rischio di scarsa informatività e incrementa la precisione nella stima della quota fetale. Terzo, la presenza dei falsi positivi, dovuta in gran parte a errati allineamenti delle reads, indica come particolare attenzione debba essere posta sia nella fase di preparazione del campione, evitando la formazione di frammenti aspecifici, sia nella fase di analisi bioinformatica, per la quale sono richiesti software altamente performanti che consentano di ridurre al minimo gli errori di chiamata delle basi e di allineamento.

## ATTIVITÀ DI RICERCA

### OBIETTIVI

L'obiettivo della presente ricerca è quello di testare l'applicabilità della piattaforma tecnologica Ion Personal Genome Machine (PGM) nella diagnosi della  $\beta$ -talassemia mediante l'utilizzo del cfDNA. In particolare, vorremmo sviluppare un protocollo che permetta la discriminazione nel DNA fetale delle più comuni mutazioni a carico del gene *HBB*, responsabili della  $\beta$ -talassemia, mediante l'amplificazione selettiva dei siti di mutazione e di altri siti polimorfi mappanti sul cluster  $\beta$ -globinico, sottoponendo al protocollo un campione di donne a rischio per  $\beta$ -talassemia. I dati ottenuti verranno analizzati attraverso due approcci:

- a) metodo RMD, considerando gli sbilanciamenti relativi alle singole mutazioni, per determinare la sensibilità del Relative Mutation Dosage nella determinazione del genotipo fetale;
- b) metodo RHDO, supportato dal test statistico dell'SPRT, valutando la trasmissione degli alleli parentali in relazione all'aplotipo a cui essi sono associati.

Gli aplotipi associati agli alleli mutato e normale potranno essere ricostruiti tramite l'ausilio del DNA fetale estratto da villi coriali, o, in assenza di villi coriali, tramite l'utilizzo di Long PCR allele selettive.

### MATERIALI E METODI

#### *CAMPIONI E ESTRAZIONE DEL DNA*

A tutte le coppie a rischio che richiedono diagnosi prenatale invasiva presso il nostro istituto, è proposta la partecipazione al presente progetto di ricerca. Il progetto è stato approvato dal Comitato Etico secondo le indicazioni nazionali vigenti e in ottemperanza al trattato di Helsinki (44). Tutte le donne che si rendono disponibili sottoscrivono un consenso informato e vengono sottoposte a un prelievo di sangue venoso di circa 20ml, dal quale viene separato il plasma ed estratto il DNA libero. Ad oggi, circa 200 donne, con età gestazionale compresa fra la 6<sup>a</sup> e 13<sup>a</sup>, hanno scelto di aderire al progetto. Il prelievo di sangue viene effettuato

sempre precedentemente all'intervento invasivo di villocentesi o amniocentesi: questo infatti causa un incremento della quantità di cfDNA nel plasma materno, con conseguente alterazione delle condizioni sperimentali. Ciascun campione di sangue viene processato entro 4 ore dal prelievo. L'intera quantità del sangue prelevato (circa 20ml) viene sottoposta a due successive centrifugazioni: la prima, a 3000g per 10 min, necessaria per la separazione del plasma dalla parte corpuscolata del sangue; la seconda, a 16000g per 45min, a cui viene sottoposto solo il plasma, mirata all'eliminazione delle eventuali cellule residue. Il surnatante (circa 10ml) viene successivamente separato in aliquote da 1ml e congelato a -80°C fino al momento dell'utilizzo. Per l'estrazione del DNA dal plasma viene utilizzato il kit Qiamp Circulating Nucleic Acid fornito dalla ditta Qiagen, particolarmente indicato per la selezione di frammenti corti, secondo le indicazioni del produttore.

Tra tutti i campioni a disposizione ne sono stati selezionati e studiati, finora, 18. Dal punto di vista genotipico, i genitori delle coppie selezionate sono o entrambi eterozigoti per la mutazione  $\beta^{039}$  o eterozigoti per mutazioni differenti (vedi tabella 10). Ogni esperimento prevede l'analisi contestuale del trio familiare e del DNA estratto da villi coriali, utilizzato come controllo e per la determinazione degli aplotipi parentali.

#### *AMPLIFICAZIONE*

Immediatamente dopo l'estrazione, il cfDNA viene amplificato per tutti i frammenti contenenti gli SNP e le mutazioni in esame. Alle stesse amplificazioni sono poi sottoposti i DNA parentali e quello estratto da villi coriali. Sono stati individuati e testati circa 130 frammenti. Tra questi ne sono stati selezionati in totale 93, così suddivisi:

- 23 siti di mutazioni sul gene **HBB**, comprendenti i siti oggetto di analisi nella pratica di diagnostica clinica (Figura 15b);
- 36 SNP ad alta eterozigosità nella popolazione CEU (dati dbSNP Build 137, NCBI; limits:  $0.3 < \text{MAF} < 0.5$  / validation: HapMap) distribuiti lungo il cluster  $\beta$ -globinico (Figura 15a);
- 2 loci mappanti su cromosoma Y (TSPY1, SRY), per la determinazione del sesso fetale;

- 1 frammento che mappa sui geni omologhi ZFX e ZFY. Questi geni dislocati uno sul cromosoma Y e l'altro sul cromosoma X mostrano regioni altamente omologhe. Il frammento studiato è un frammento di circa 80 bp in cui i due geni presentano due mismatch, a distanza di 6bp fra loro. La rilevazione di queste differenze nel frammento è indice della presenza del cromosoma Y nel sangue materno, e viene utilizzata sia per la determinazione del sesso che per la quantificazione della quota fetale;
- 31 SNP mappanti su vari cromosomi, utilizzati nella prima fase sperimentale (prime 5 famiglie), necessari esclusivamente per la valutazione dei dati di output dello strumento e non correlati alla diagnosi di talassemia.



Ogni reazione di amplificazione prevede volume finale di 20µl e reagenti secondo lo schema presentato in tabella 1:

Reagenti	Quantità (µl)
High Fidelity Platinum® Supermix	15
Spermina 0.5 mM +1mg/ml BSA	1
Primer Forward (25µMol)	0.5
Primer Reverse (25µMol)	0.5
Acqua Nuclease free	1
DNA (genomico: 20µg totali; da villo coriale: 40µg totali )	2

Tabella 1 Reagenti utilizzati nella prima reazione di amplificazione e selezione dei frammenti

Tutte le amplificazioni vengono eseguite in piastra MicroAmp® Fast Optical 96-Well Reaction Plate da 0.1 mL su Veriti® Thermal Cycler (Life Technologies) secondo il seguente protocollo di amplificazione (tabella 2):

Stage	Step	Temperature (°C)	Time (min)
Holding	Denature	95°	3:00
30 cicli	Denature	95°	0:30
	Anneal	55°/57°/60°	0:30
	Extend	68°	1:00
Holding	Extend	68°	5:00

Tabella 2 Schema dei tempi e delle temperature di amplificazione su Veriti Thermal Cycler

Unica differenza rispetto al precedente schema, il frammento relativo al gene SRY che sfrutta il protocollo in tabella 3:

Stage	Step	Temperature (°C)	Time (min)
Holding	Denature	95°	5:00
15 cicli (con decremento 0.5°C per ciclo)	Denature	95°	0:30
	Anneal	64°	0:30
	Extend	68°	1:00
	Denature	95°	0:30
25 cicli	Anneal	56°	0:30
	Extend	68°	1:00
	Holding	Extend	72°

Tabella 3 Schema dei tempi e delle temperature di amplificazione per il frammento sul gene SRY



Il controllo dei prodotti di PCR viene effettuato su gel d'agarosio al 2%.

#### LONG PCR ALLELE SPECIFICA

Nell'analisi di alcune famiglie non è stato possibile, anche in presenza del dato del DNA fetale da villi coriali, poter definire gli aplotipi parentali e determinare associazione l'allele normale o mutato e aplotipo. Per poter risolvere questi casi di indeterminazione abbiamo utilizzato un approccio basato sulla Long PCR allele-specifica: questo metodo consente, in fase di amplificazione, di discriminare i frammenti che portano gli alleli mutato o normale, amplificando selettivamente solo il filamento associato alla variante selezionata e con la conseguente ulteriore associazione fra tutti gli alleli dei siti polimorfi contenuti nel frammento amplificato. Questo tipo di amplificazione selettiva viene eseguito mediante l'utilizzo di primer allele-specifici, disegnati in modo tale che l'ultima base in 3' sia coincidente con il sito di mutazione e complementare a uno dei due alleli (tabella 4). A causa del mismatch creato rispetto all'allele opposto, in ciascuna reazione di amplificazione viene prodotto il frammento associato ad una sola variante allelica. Abbiamo processato mediante questo metodo i DNA parentali.

Mutazione	Primer ARMS	Primer	Frammento
<b>β<sup>39</sup></b>	M CCAAGGACTCAAAGAACCTGTG N CCCAAAGGACTCAAAGAACCTGTA	AGCCCATCATCACACTGATT (PRIMER D)	~7 kb
<b>β<sup>76</sup></b>	M CAAAGGTGCCCTTGAGGTTGTCCAGGTCAC N CAAAGGTGCCCTTGAGGTTGTCCAGGTCAG	AGCCCATCATCACACTGATT (PRIMER D)	~7 kb
<b>HbS</b>	M CCCACAGGGCAGTAACGGCAGCGTTGTGCA N CCCACAGGGCAGTAACGGCAGCGTTGTGCT	AGCCCATCATCACACTGATT (PRIMER D)	~7 kb
<b>β<sup>6</sup></b>	M CCCACAGGGCAGTAACGGCAGACTTCTGCA N CCCACAGGGCAGTAACGGCAGACTTCTGCA	AGCCCATCATCACACTGATT (PRIMER D)	~7 kb
<b>IVS I-110</b>	M ACCAGCAGCCTAAGGGTGGGAAAATACACT N ACCAGCAGCCTAAGGGTGGGAAAATACACC	AGCCCATCATCACACTGATT (PRIMER D)	~7 kb
<b>β<sup>39</sup></b>	M TTAGGCTGCTGGTGGTCTACCCTTGGAGCT N TTAGGCTGCTGGTGGTCTACCCTTGGAGCC	ACGTTGGATGTTGAACTCCTGACGTCGTGA (4417)	~4.4 kb
<b>β<sup>76</sup></b>	M TCGGTGCCTTTAGTGATGGCCTTGT N TCGGTGCCTTTAGTGATGGCCTTGC	ACGTTGGATGTTGAACTCCTGACGTCGTGA (4417)	~4.4 kb
<b>HbS</b>	M ACACCATGGTGCATCTGACTCCCGT N ACACCATGGTGCATCTGACTCCCGA	ACGTTGGATGTTGAACTCCTGACGTCGTGA (4417)	~4.4 kb
<b>β<sup>6</sup></b>	M ACACCATGGTGCATCTGACTCCCGG N ACACCATGGTGCATCTGACTCCCGA	ACGTTGGATGTTGAACTCCTGACGTCGTGA (4417)	~4.4 kb
<b>IVS I-110</b>	M AGGCACCTGACTCTCTGCTACTA N AGGCACCTGACTCTCTGCTACTG	ACGTTGGATGTTGAACTCCTGACGTCGTGA (4417)	~4.4 kb

Tabella 4 Elenco primer ARMS

Attraverso questa metodica è stato possibile produrre, per ciascun campione, attraverso l'utilizzo di Platinum® Taq polimerasi, due frammenti rispettivamente di circa 4,4Kb (in direzione 3' del gene HBB rispetto alla

mutazione in esame) e circa 7Kb (in direzione 5'). Ciascun amplificato allele specifico è stato poi utilizzato come frammento di partenza e sequenziato o attraverso il metodo di Sanger, nei casi in cui era necessario l'esame di pochi e selezionati polimorfismi, o sequenziato secondo protocollo Ion Torrent, nei casi in cui era necessario lo studio di tutti i polimorfismi mappanti sul frammento. Il controllo dei prodotti di PCR venne effettuato su gel d'agarosio all'1%.

Reagenti	Volume ( $\mu$ l)
High Fidelity Platinum® Supermix	15
Spermina 0.5 mM +1mg/ml BSA	1
Primer ARMS (10pMol)	1
Primer (10pMol)	1
DNA (genomico: 50pg totali)	2

Tabella 5 Elenco reagenti utilizzati nella reazione allele specifica

Stage	Step	Temperature ( $^{\circ}$ C)	Time (min)
Holding	Denature	94 $^{\circ}$	2:00
	Denature	94 $^{\circ}$	0:30
30 cicli	Anneal	63 $^{\circ}$	0:30
	Extend	68 $^{\circ}$	1 min x Kb

Tabella 6 Schema file delle temperature per l'amplificazione allele specifica

#### *PREPARAZIONE LIBRERIE ED EMULSION PCR*

Tutti gli ampliconi derivati da amplificazione target o Long PCR, vengono riuniti in un unico pool equimolare e purificati attraverso Agencourt AMPure XP - PCR Purification. Successivamente, i pool vengono quantificati attraverso fluorimetro Qubit® 2.0 (Life Technologies). 50ng di DNA vengono sottoposti ad una prima reazione di End Repair, che ricostituisce le estremità 3' blunt-end e 5' fosforilato, mediante il kit Ion Plus Fragment Library Kit, e successivamente purificati.

Poiché più campioni vengono sequenziati nella stessa corsa, occorre marcare ciascun frammento con delle sequenze di riconoscimento (barcode) che possano permettere, in fase di analisi informatica, l'attribuzione di ciascuno di essi al rispettivo campione di appartenenza. Inoltre, è necessario che i frammenti presentino anche delle sequenze standard (adattatori), necessarie nella successiva

fase di amplificazione clonale. A questo scopo le librerie dei campioni vengono sottoposte a reazione di ligazione dei barcode/adapters e nick repair (Ion Plus Fragment Library Kit) e successivamente purificate. Passaggio seguente è una breve reazione di PCR con 7 cicli di amplificazione, seguita da purificazione, secondo le indicazioni dei protocolli standard.

Un punto fondamentale del protocollo è l'utilizzo di una precisa concentrazione di template nella fase di amplificazione clonale attraverso Emulsion PCR. Una concentrazione troppo elevata genera policlonalità, che produce dati non analizzabili; al contrario, una concentrazione troppo bassa può causare la presenza di segnali troppo deboli e un'imprecisa rilevazione da parte dello strumento. Per poter calcolare la corretta concentrazione dei pool di librerie ci si avvale dell'ausilio del 2100 Bioanalyzer Instrument (Agilent). I pool vengono diluiti dapprima equimolarmente. Da questo punto in poi, possono essere riuniti e trattati come un unico campione. In questa fase di riunione dei pool, viene utilizzata una pari quantità di tutti i campioni, ad eccezione dei campioni derivanti da cfDNA, di cui viene utilizzato un volume 10 volte superiore rispetto ai DNA genomici, al fine di garantire una sufficiente presenza del DNA fetale nel campione in esame. Il pool così costituito viene successivamente quantificato e portato alla concentrazione 23pM; 20µl di questa soluzione vengono sottoposti ad amplificazione clonale attraverso Emulsion PCR. I reagenti e le relative quantità per la preparazione della miscela di reazione sono schematizzati nella tabella 7.

Reagente	Volume (µl)
Nuclease free water	280
Ion OneTouch 2x Reagent Mix	500
Ion OneTouch Enzyme Mix	100
Diluted Library	20
Ion One Touch 200 Ion Sphere Particles	100

Tabella 7 Miscela per la reazione di amplificazione clonale (Emulsion PCR)

L'intero volume (1 ml) di questa miscela, unitamente al Reaction Oil (1.5 ml), viene dispensato all'interno di un supporto (Reaction Filter) da installare poi sull'Ion OneTouch DL, nel quale viene eseguita l'Emulsion PCR. Nel passaggio

successivo, mediante l'utilizzo dell'Ion OneTouch ES, l'intero campione viene sottoposto a una speciale purificazione: questa favorisce l'arricchimento del campione attraverso la selezione delle Ion Sphere Particles su cui è avvenuta l'amplificazione, e l'eliminazione di quelle che invece non presentano sulla loro superficie frammenti amplificati. Il protocollo prosegue con un passaggio di controllo: una quantificazione, mediante fluorimetro Qubit, di 2µl di campione, opportunamente diluito, al fine di determinare l'effettivo arricchimento e l'eventuale rischio di elevata policlonalità.

#### *SEQUENZIAMENTO E ANALISI DEI DATI*

Ultimo passaggio della fase sperimentale è il sequenziamento. In una prima fase detta inizializzazione, lo strumento PGM automaticamente effettua dei lavaggi per la pulizia delle condutture interne e porta a pH ottimale (7.75) le soluzioni che intervengono nel sequenziamento. Poiché tutta la tecnologia è basata sulla misura di minime variazioni di pH, questa fase di inizializzazione è un processo delicato, che richiede la massima cura nella preparazione delle soluzioni di partenza, e da cui dipende il buon esito della corsa. Successivamente vengono preparate e installate nel PGM quattro ulteriori soluzioni contenenti i dNTP. Questi verranno fatti fluire uno per volta nel chip durante il sequenziamento determinando le sintesi della nuova catena nucleotidica.

Al campione vengono aggiunti, secondo il protocollo proposto, i reagenti per il sequenziamento, contenuti nell'Ion PGM 200 Sequencing Kit: Control Ion Sphere particles, Sequencing Primer, Annealing buffer e PGM 200 Sequencing Polymerase. Contemporaneamente, l'Ion Chip che dovrà essere utilizzato per il sequenziamento, viene inserito nel PGM e sottoposto ad un controllo di verifica dell'effettiva funzionalità. Concluso il Chip Check, il campione è pronto per essere caricato con il campione. Il caricamento prevede diverse fasi, tra cui figurano tre passaggi in apposita centrifuga, indispensabili per la ottimale distribuzione del campione su tutta la superficie del Chip e per l'eliminazione di eventuali bolle d'aria dal suo interno.

Il processo di sequenziamento prevede 500 flows e ha una durata proporzionale alla capienza del chip utilizzato, che varia da circa 2 ore per il Chip

314 fino alle 4.5 ore per il Chip 318. L'analisi dei dati grezzi comincia contestualmente al sequenziamento ma si estende oltre la fine della corsa; il tempo di analisi per un chip 318 è di circa 8 ore complessive. Il PGM è corredato di software specifico (Torrent Suite 4.0) che gestisce lo strumento in tutte le sue fasi operative. Lo stesso software incorpora gli algoritmi appositi per l'analisi e la valutazione dei dati grezzi fino alla produzione dei dati di output con gli allineamenti delle reads. I formati dei file di output sono quelli diventati oramai standard nelle tecnologie NGS (Sam, Bam, Bai, VCF, FASTQ, etc). I file Bam vengono visualizzati attraverso IGV Software Viewer e i dati, relativi alle reads di ciascun allele, annotati in tabelle mediante l'utilizzo del software Microsoft Excel. Il genoma di riferimento caricato su IGV è costituito dalle sequenze separate dei singoli ampliconi in esame. I dati vengono successivamente analizzati attraverso gli approcci RMD e RHDO e valutati statisticamente mediante il test SPRT.

Condizione necessaria per l'applicazione di RMD e RHDO è la determinazione delle fasi degli aplotipi paterni e materni. La genotipizzazione del DNA fetale estratto da villo coriale permette di ricostruire gli aplotipi parentali ad eccezione di quelle posizioni in cui madre, padre e villo presentano genotipo eterozigote. In questi casi, l'utilizzo di una Long PCR permette di ovviare a questo problema poiché consente di definire le singole fasi degli aplotipi parentali.

Come descritto precedentemente, attraverso IGV si visualizzano le reads relative a ciascuno SNP ed è possibile la conta dei frammenti che portano l'allele normale e l'allele mutato; sul confronto tra queste conte si basa tutta l'analisi del DNA fetale. La quota fetale viene ottenuta attraverso la formula  $f = \frac{2p}{p+q}$  (6), dove  $f$  è la quota fetale,  $p$  il numero delle reads dell'allele paterno e  $q$  il numero delle reads dell'allele materno.

## RISULTATI

Sono stati analizzati ad oggi 18 campioni di cffDNA provenienti da 16 famiglie, due di queste esaminate ciascuna per due gravidanze differenti. Gli SNP utilizzati nello studio, amplificati per tutte le famiglie in esame, sono stati circa 70, e sono così suddivisi:

- 23 SNP relativi alle mutazioni e i siti polimorfi del gene **HBB**;
- 36 SNP, 5 in posizione 3' rispetto ad **HBB** e 31 dislocati in direzione 5' fino al gene **HBE**, distribuiti in una regione di circa 40Kb;
- 2 ampliconi mappanti rispettivamente sui geni **SRY** e **TSPY** per la determinazione del sesso fetale;
- 1 amplicone mappante sui geni omologhi **ZFX** e **ZFY**.

La lunghezza delle sequenze coperte dai frammenti è di circa 6500bp.

Nelle prime cinque famiglie, inoltre, sono stati analizzati ulteriori 31 frammenti dislocati su diversi cromosomi. Questa quota addizionale è stata necessaria nei primi esperimenti per incrementare il numero di osservazioni, al fine di stimare con maggior cura l'efficacia del protocollo nella determinazione della frazione fetale.

Le prime osservazioni sono state rivolte alla verifica della capacità dello strumento di mettere in risalto la frazione fetale. Una prima necessità sorta durante le analisi è stata quella di suddividere gli SNP in base ai genotipi parentali. Ci siamo allineati alla classificazione proposta dal gruppo di YMD Lo nel 2010 (6) come descritto nella tabella 8:

SNP category	Parental genomic DNA analysis		Maternal plasma DNA sequencing and analysis		
	Paternal genotype	Maternal genotype	Analysis	Deduced fetal genotype	Additional information acquired
1.	A/A	C/C	Detection of the paternal allele and quality control	Heterozygous	1. Coverage of fetal genome by sequencing 2. Fractional concentration of fetal DNA 3. Sequencing error rate
2.	A/A	A/A	Quality control	Homozygous for the parental allele	Sequencing error rate
3.	A/C	A/A	Detection of the paternal-specific allele	1. Heterozygous if the paternal-specific allele is detected. 2. Homozygous if the paternal-specific allele is not detected.	Fractional concentration of fetal DNA
4.	A/A	A/C*	Relative haplotype dosage (RHDO) analysis*	1. Heterozygous if dosage imbalance is not detected. 2. Homozygous if haplotype dosage imbalance is detected	
5.	A/C	A/C	Not analyzed in this study		

Tabella 8 Suddivisione degli SNP in base all'informazione fornita (da YMD Lo et al. 2010)

Gli SNP utili per il controllo della quota fetale sono quelli per i quali la madre è omozigote e il padre, omozigote o eterozigote, trasmette al feto l'allele non presente nel genotipo materno (tipo 1 e 3). La totalità dei siti valutati in questo caso è di 91 (sono esclusi gli ampliconi sui geni TSPY e SRY). Oltre a questi, nel caso in cui il feto sia di sesso maschile, viene utilizzato per la determinazione della quota fetale anche il dato relativo all'amplicone ZFX/ZFY. La tabella 9 presenta il confronto fra i dati delle prime cinque famiglie analizzate, mirato a verificare se l'amplificazione degli SNP aggiuntivi distribuiti sui diversi cromosomi apportasse vantaggi nella determinazione della quota fetale. Messe a confronto, le quote fetali dedotte in presenza e assenza degli SNP supplementari non mostrano evidenti differenze. Il numero degli polimorfismi informativi, nel secondo gruppo, è evidentemente inferiore e nel caso della famiglia 1 è stata riscontrata assenza di SNP di tipo 1 e 3. Un ulteriore dato analizzato è stata la media di falsi negativi che risulta essere molto simile fra i due gruppi. Valutati questi risultati, e data la maggior onerosità dal punto di vista sperimentale ed economico del protocollo a

91 ampliconi, si è stabilito di procedere nei successivi esperimenti con il protocollo a 60 ampliconi.

Famiglia	91 SNP analizzati (cluster $\beta$ + cromosomi)				60 SNP analizzati (cluster $\beta$ )			
	SNP utili (tipo1 + tipo3)	falsi negativi	falsi negativi (%)	Quota fetale	SNP utili (tipo1 + tipo3)	falsi negativi	falsi negativi (%)	Quota fetale
1	8	2	25%	3,67	0	-	-	-
2	16	1	6,3%	5,25	9	0	0,0%	4,6
3	24	1	4,2%	8,25	15	1	6,7%	8,4
4	24	7	29,2%	9,17	18	5	27,8%	10,4
5	39	7	17,9%	4,25	25	5	20,0%	3,9

Tabella 9 Confronto dati relativi all'analisi delle prime cinque famiglie in presenza e assenza degli SNP distribuiti su cromosomi diversi dal cromosoma 11

La tabella 10 mostra i dati relativi alla quota fetale per tutte le famiglie prese in esame. I campioni 8 e 9 e i campioni 10 e 11 derivano da gravidanze singole differenti ma appartenenti agli stessi nuclei familiari. Il campione 11 è stato prelevato in epoca gestazionale molto precoce; esso è stato ugualmente analizzato, a scopo esplorativo, ma i dati, presentati per la completezza dello studio, confermano i limiti dei metodi attuali nell'analisi di frazioni fetali molto basse, come quelle presenti in 6<sup>a</sup> settimana di gravidanza. I dati relativi a questo campione, infatti, non godono di elevata consistenza, e, a conferma di ciò, si può notare l'elevato grado di errore nella determinazione degli alleli paterni. Il campione 14 ha presentato dei problemi sperimentali, riconducibili alla fase di preparazione della libreria.



Famiglia	SNP utili (tipo1 + tipo3)	Falsi negativi	Falsi negativi (%)	Quota fetale	Quota ZFX/ZFY	Epoca gestazionale (settimane)
1	0	-	-	-	2%	8
2	9	0	0,00%	4,6	2%	8+2
3	15	1	6,70%	8,4	-	7+3
4	18	5	27,80%	10,5	-	9+2
5	25	5	20,00%	3,9	10%	8+3
6	0	-	-	-	6%	9+3
7	8	1	12,50%	6,3	4%	8+4
8	0	-	-	-	2%	8
9	0	-	-	-	8%	9+1
10	19	0	0%	9,9	-	9+3
11*	19	7	36,84%	3,9	-	6+2
12	22	0	0%	8,7	12%	9+2
13	3	0	0%	5,3	10%	10+3
14*	17	5	29,41%	4,5	-	9
15	0	-	-	-	8%	9
16	7	0	0%	4,3	-	11+2
17	27	1	3,70%	6,4	4%	9
18	16	1	6,25%	8,4	-	8

Tabella 10 Quota fetale calcolata in base all'osservazione degli SNP di tipo 1 e 3

Il protocollo mostra una sensibilità nella determinazione dell'allele di derivazione paterna che varia dall'83,2% al 100% con una media del 93%. È importante notare come il numero degli SNP utilizzabili per la determinazione della quota fetale sia notevolmente differente tra le famiglie: la carenza, spesso riscontrata, di polimorfismi informativi è dovuta alla forte omogeneità della popolazione della Sardegna, cui appartiene la maggior parte delle famiglie analizzate, evidenza ancor più accentuata dalla ulteriore selezione effettuata sul nostro campione che include solo portatori di  $\beta$ -talassemia.

Si nota in tabella come al susseguirsi degli esperimenti, inseriti in ordine cronologico, si sia verificato un evidente miglioramento nella capacità di determinazione degli alleli fetali nel plasma materno. Ciò può essere riconducibile, da un lato, alla maggior esperienza acquisita nel corso del tempo dagli operatori

nel portare avanti un protocollo lungo e complesso, dall'altro, ai miglioramenti apportati alla tecnologia, sia in ambito strettamente sperimentale, con l'introduzione di reagenti chimici più efficaci, sia nell'analisi bioinformatica dei dati, con software e algoritmi via via più robusti ed affidabili.

Riguardo alla correlazione tra quota fetale e età di gestazione, dedotta attraverso l'analisi degli SNP informativi, i dati in nostro possesso sono ancora limitati dal punto di vista numerico e non consentono di poter evidenziare una correlazione diretta tra frazione fetale e epoca gestazionale, fenomeno ampiamente dimostrato dai dati in letteratura.

Con lo scopo di verificare l'effettiva possibilità di utilizzo della piattaforma Ion Torrent, e, in particolare lo strumento PGM, per l'analisi del DNA da plasma materno nella diagnosi prenatale non invasiva, abbiamo valutato per ogni famiglia gli sbilanciamenti nel plasma relativi alle mutazioni dei genitori (vedi tabella 11). Le famiglie selezionate comprendono sia coppie di portatori per la stessa mutazione ( $\beta^{\circ}39$ ) che coppie di portatori per mutazioni differenti, con feti normali, eterozigoti e affetti. Le famiglie sono state selezionate appositamente per poter disporre di un insieme eterogeneo di campioni, tale da valutare l'efficacia degli approcci RMD e RHDO in differenti condizioni sperimentali.

Famiglia	Genotipo paterno		Genotipo materno		Genotipo fetale (villo coriale)			Sesso fetale
	Aplotipo P1	Aplotipo P2	Aplotipo M1	Aplotipo M2	Aplotipo paterno	Aplotipo materno	Diagnosi (CVS)	
1	Norm	$\beta^{\circ}39$	$\beta^{\circ}39$	Norm	$\beta^{\circ}39$	$\beta^{\circ}39$	Affetto	XY
2	Norm	$\beta$ IVS II.745	Norm	$\beta^{\circ}39$	$\beta$ IVS II.745	Norm	Portatore	XX
3	Norm	$\beta$ IVS I.110	$\beta^{\circ}39$	Norm	$\beta$ IVS I.110	$\beta^{\circ}39$	Affetto	XX
4	$\beta^{\circ}39$	Norm	Norm	$\beta^{\circ}39$	Norm	$\beta^{\circ}39$	Portatore	XY
5	Norm	$\beta^{\circ}39$	$\beta^{\circ}39$	Norm	Norm	$\beta^{\circ}39$	Portatore	XX
6	Norm	$\beta^{\circ}39$	Norm	$\beta^{\circ}39$	$\beta^{\circ}39$	Norm	Portatore	XY
7	$\beta^{\circ}39$	Norm	Norm	$\beta^{\circ}39$	Norm	Norm	Normale	XY
8	Norm	$\beta^{\circ}39$	$\beta^{\circ}39$	Norm	$\beta^{\circ}39$	$\beta^{\circ}39$	Affetto	XY
9	Norm	$\beta^{\circ}39$	Norm	$\beta^{\circ}39$	$\beta^{\circ}39$	Norm	Portatore	XY
10	Norm	$\beta^{\circ}39$	$\beta^{\circ}39$	Norm	$\beta^{\circ}39$	$\beta^{\circ}39$	Affetto	XX
11	Norm	$\beta^{\circ}39$	$\beta^{\circ}39$	Norm	Norm	$\beta^{\circ}39$	Portatore	XX
12	$\beta^{\circ}76$	Norm	Norm	$\beta^{\circ}39$	Norm	Norm	Normale	XY
13	$\beta^{\circ}39$	Norm	Norm	$\beta^{\circ}39$	Norm	Norm	Normale	XY
14	$\beta^{\circ}39$	Norm	Norm	$\beta^{\circ}39$	Norm	Norm	Normale	XX
15	Norm	$\beta^{\circ}39$	$\beta^{\circ}39$	Norm	$\beta^{\circ}39$	$\beta^{\circ}39$	Affetto	XY
16	Norm	$\beta^{\circ}39$	$\beta^{\circ}6$ (-A)	Norm	$\beta^{\circ}39$	$\beta^{\circ}6$ (-A)	Affetto	XX
17	Norm	$\beta^{\circ}6$ (A>T)	$\beta^{\circ}39$	Norm	$\beta^{\circ}6$ (A>T)	$\beta^{\circ}39$	Affetto	XY
18	Norm	$\beta$ IVS I.110	$\beta^{\circ}39$	Norm	$\beta$ IVS I.110	$\beta^{\circ}39$	Affetto	XX

Tabella 11 Genotipi parentali e genotipo fetale dedotto da analisi su villo coriale

Mentre la determinazione dell'allele materno ereditato viene effettuata sempre attraverso il calcolo RMD, per l'allele paterno l'applicazione di questo approccio si rende necessaria solo nel caso in cui la mutazione paterna sia condivisa dalla madre; in caso contrario, cioè con genitori eterozigoti per siti di mutazione differenti, l'allele paterno viene facilmente messo in evidenza poiché assente nel genoma materno. I dati ottenuti sono sempre confrontati con i dati derivati dalla diagnosi invasiva. I risultati sono presentati in tabella 12

Famiglia	Genotipo fetale		Risultato RMD
16	$\beta^{39} \beta^6$	Affetto	Portatore
2	$\beta$ IVS II.745 WT	Portatore	Portatore
3	$\beta$ IVS I.110 $\beta^{39}$	Affetto	Affetto
18	$\beta$ IVS I.110 $\beta^{39}$	Affetto	Affetto
17	$\beta^6$ (A>T) $\beta^{39}$	Affetto	Portatore
12	WT WT	Normale	Normale
7	WT WT	Normale	Normale
13	WT WT	Normale	Normale
14*	WT WT	Normale	Portatore
15	$\beta^{39} \beta^{39}$	Affetto	Portatore
1	$\beta^{39} \beta^{39}$	Affetto	Affetto
8	$\beta^{39} \beta^{39}$	Affetto	Affetto
10	$\beta^{39} \beta^{39}$	Affetto	Affetto
4	WT $\beta^{39}$	Portatore	Normale
5	WT $\beta^{39}$	Portatore	Affetto
6	WT $\beta^{39}$	Portatore	Portatore
9	WT $\beta^{39}$	Portatore	Normale
11*	WT $\beta^{39}$	Portatore	Normale

Tabella 12 Risultati analisi famiglie attraverso RMD

Nella tabella 12, i dati sono ordinati in sottogruppi in base al genotipo fetale. L'analisi ha mostrato risultati coerenti con il dato atteso in 10 casi su 18 (55.5%) mentre in 8 casi il risultato si è rivelato errato. Il sottogruppo che ha mostrato il più alto tasso di errore è quello caratterizzato da feti e genitori eterozigoti per la mutazione  $\beta^{39}$ . Quando madre e feto sono entrambi eterozigoti per la stessa mutazione si riscontra una maggior imprecisione del metodo RMD e la valutazione degli sbilanciamenti dei due alleli del sito di mutazione risulta

particolarmente difficile: ci si attende infatti in questa situazione un rapporto paritario tra i due alleli che difficilmente viene riscontrato nei risultati. Contrariamente, nel caso in cui il feto sia omozigote, la presenza di un unico allele fetale, favorisce un più netto sbilanciamento nelle conte delle reads tra l'allele normale e l'allele mutato, e ciò determina una maggior efficacia del metodo RMD. Analizzando il sottogruppo caratterizzato da genitori eterozigoti per differenti mutazioni, si osserva che gli errori riscontrati sono tutti riconducibili alla errata definizione dell'allele fetale ereditato dalla madre, mentre la presenza dell'allele paterno viene correttamente evidenziata in tutti i casi. Un'osservazione generale di tutti gli SNP coinvolti nell'analisi, conferma quanto precedentemente osservato: gli scostamenti più evidenti rispetto al valore atteso emersi dall'analisi RMD, si presentano nei casi in cui madre e feto sono entrambi eterozigoti per le stesse varianti.

Un primo dato emerso dai risultati ottenuti attraverso l'approccio RMD, è la particolare difficoltà che esso mostra nella determinazione del genotipo fetale quando esso è identico a quello materno. Con lo scopo di aumentare il potere d'analisi nella definizione del genotipo fetale, soprattutto in riferimento alla ereditarietà da parte materna, in una seconda fase dello studio è stata valutata la trasmissione riferita agli aplotipi parentali piuttosto che ai singoli alleli relativi alla mutazione. Le fasi degli aplotipi parentali sono state determinate mediante l'ausilio del DNA fetale estratto dai villi coriali. La trasmissione dell'aplotipo viene valutata attraverso l'osservazione degli sbilanciamenti di tutti gli SNP informativi contenuti nell'aplotipo stesso: maggiore è il numero degli SNP che danno indicazione per uno stesso aplotipo, tanto più solido sarà il dato che lo descrive. Associando agli aplotipi parentali l'allele mutato e l'allele normale della mutazione in esame, potrà essere definito il genotipo fetale. Da questa analisi sono state inizialmente escluse le famiglie con genitori eterozigoti per la mutazione  $\beta^{39}$  e feto eterozigote, poiché in questi casi, in mancanza di ulteriori indicazioni, gli alleli delle varianti non possono essere associati agli aplotipi parentali.

Dall'analisi attraverso RMD è inoltre emersa la necessità di un metodo i cui dati ottenuti fossero supportati da una validazione statistica. Per soddisfare l'esigenza di un maggior potere d'analisi abbiamo testato il metodo RHDO,

descritto anch'esso dal gruppo di YMD Lo (41) (42), che, in diverse pubblicazioni, propone questo approccio come metodo utile per la determinazione della trasmissione dell'aplotipo materno al feto, in casi di famiglie a rischio per  $\beta$ -talassemia in cui i genitori erano portatori di diverse mutazioni. Questo metodo si basa sulla valutazione delle diverse probabilità di trasmissione dei due aplotipi materni stimate mediante l'impiego del test statistico SPRT.

Nella prima fase di questa analisi sono stati selezionati tutti gli SNP utili per l'applicazione dell'approccio l'RHDO e sono stati classificati in SNP di tipo  $\alpha$  e  $\beta$ , secondo allo schema indicato precedentemente, coerente con quello presentato da YMD Lo. In seguito, è stato applicato il test SPRT e valutata l'ipotesi di trasmissione per ciascuno dei due aplotipi materni contro l'ipotesi di non trasmissione. Anche in questo caso l'analisi risente in maniera considerevole del numero degli SNP informativi presenti nell'aplotipo. Per alcune famiglie infatti è stata riscontrata l'assenza di SNP informativi e non è stato possibile definire con esattezza il genotipo fetale.

famiglia	Genotipo fetale		Risultati analisi SPRT		
	Aplo. Paterno - Materno	Diagnosi	Aplo. paterno	Aplo. materno	Diagnosi
16	$\beta^{39} \beta^6$	Affetto	$\beta^{39}$	(no data)	Portatore/Affetto
2	$\beta$ IVS II.745 WT	Portatore	$\beta$ IVS II.745	WT	Portatore
3	$\beta$ IVS I.110 $\beta^{39}$	Affetto	$\beta$ IVS I.110	$\beta^{39}$	Affetto
18	$\beta$ IVS I.110 $\beta^{39}$	Affetto	$\beta$ IVS I.110	$\beta^{39}$	Affetto
17	$\beta^6 \beta^{39}$	Affetto	$\beta^6$	$\beta^{39}$	Affetto
12	WT WT	Normale	WT	WT	Normale
7	WT WT	Normale	WT	(no data)	Normale/Portatore
13	WT WT	Normale	WT	(no data)	Normale
14*	WT WT	Normale	WT	$\beta^{39}$	Portatore
15	$\beta^{39} \beta^{39}$	Affetto	$\beta^{39}$	(no data)	Portatore/Affetto
1	$\beta^{39} \beta^{39}$	Affetto	$\beta^{39}$	$\beta^{39}$	Affetto
8	$\beta^{39} \beta^{39}$	Affetto	(no data)	$\beta^{39}$	Portatore/Affetto
10	$\beta^{39} \beta^{39}$	Affetto	(no data)	$\beta^{39}$	Portatore/Affetto
4	WT $\beta^{39}$	Portatore	Hap P2	Hap M1	
5	WT $\beta^{39}$	Portatore	Hap P2	(no data)	
6	WT $\beta^{39}$	Portatore	Hap P2	Hap M1	
9	WT $\beta^{39}$	Portatore	(no data)	Hap M1	
11*	WT $\beta^{39}$	Portatore	(no data)	(no data)	(no data)

Tabella 13 Risultato dell'analisi delle famiglie attraverso il metodo RHDO-SPRT.

Per le famiglie 4, 5, 6, 9 e 11 (genitori e feto eterozigoti  $\beta^{39}$ ) è stato indicato, dove possibile, quale aplotipo paterno e materno sia stato ereditato dal feto. Non è stato specificato l'allele ereditato relativo alla mutazione poiché non è stato possibile associare gli aplotipi e gli alleli della variante. Per la famiglia 11 non è stato possibile ottenere alcuna informazione a causa della epoca gestazionale molto precoce, epoca in cui la frazione fetale è ancora molto bassa e perciò difficilmente rilevabile. Nei restanti casi, l'unico dato errato è stato riscontrato nella famiglia 14, di cui era già stata indicata la presenza di problemi sopraggiunti durante il processo sperimentale, anticipando la possibile inattendibilità dei risultati ottenuti. Nessuno SNP informativo è presente nelle famiglie 8 e 10, in riferimento all'aplotipo paterno; nelle famiglie 16, 7, 13 e 15 è stato possibile determinare il genotipo fetale solo al 50%, a causa dell'assenza di SNP informativi per l'aplotipo materno. In tutti i restanti casi la trasmissione degli aplotipi è stata definita con esattezza, e, quando nota l'associazione degli alleli del sito di mutazione con l'aplotipo, è stato possibile indicare il corretto genotipo fetale.

Al fine di rafforzare il segnale della quota fetale e mettere in risalto in maniera più efficace la componente fetale, i campioni di cfDNA sono stati sequenziati con coverage molto elevati (ordine delle decine di migliaia). Ciò ha determinato la necessità di alcune modifiche nel metodo SPRT (fig. 16) rispetto alla sua applicazione descritta nei lavori citati: gli SNP informativi, associati ad un determinato aplotipo, sono stati considerati nel loro insieme, senza tener conto del superamento delle soglie con i soli dati parziali, imponendo come risultato finale il dato espresso dalla somma cumulativa complessiva delle reads di tutti gli SNP costituenti l'aplotipo. Questo metodo ci permette di calcolare la probabilità di trasmissione dell'aplotipo tenendo conto del contributo complessivo di tutti gli alleli ad esso associati. Mediante l'applicazione di questo protocollo abbiamo ottenuto su 18 famiglie analizzate i seguenti risultati:

- 7 diagnosi corrette;
- 5 diagnosi determinate al 50% (il cui dato parziale ottenuto risulta comunque coerente con l'atteso);
- 1 diagnosi errata;

- 5 non analizzabili per assenza di informazioni.

Il test RHDO-SPRT si dimostra quindi un valido strumento per la valutazione della trasmissione degli aplotipi parentali. Esso ha offerto due principali vantaggi: da un lato, ha consentito la possibilità di avvalersi di un supporto statistico che rafforza i risultati ottenuti; dall'altro, l'applicazione di un metodo proposto e testato da gruppi di ricerca di riferimento nell'ambito dell'analisi del cfDNA, attribuisce maggior affidabilità ai dati prodotti. Nonostante i buoni risultati mostrati, l'approccio RHDO-SPRT mostra un lato debole da tenere in considerazione: il metodo soffre della scarsa numerosità degli SNP informativi e in loro assenza è del tutto inefficace. A conferma di ciò, in 6 delle nostre 18 famiglie, la sua applicazione non è stata possibile a causa della mancanza di SNP informativi e, di conseguenza, non si è potuta definire la trasmissione dell'aplotipo materno.

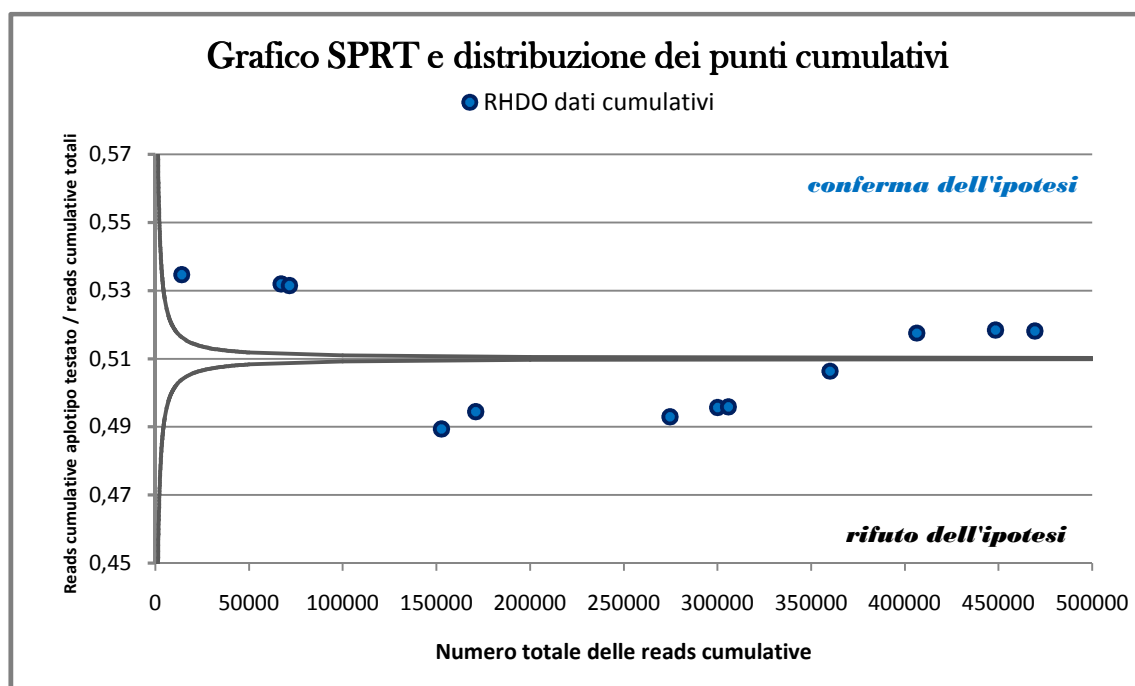


Figura 16 Valutazione dell'ipotesi di trasmissione dell'aplotipo. L'ultimo punto a destra nel grafico rappresenta il valore ottenuto sommando gli effetti di tutti gli SNP dell'aplotipo. Il punto giace al di sopra della curva e ciò conferma l'ipotesi di trasmissione

Un ulteriore problema di difficile risoluzione riguarda la definizione delle fasi nei casi delle famiglie che presentano genitori portatori e feto eterozigote per la

stessa mutazione, situazione che non consente la definizione del genotipo fetale attraverso gli approcci proposti, a causa dell'impossibilità di associare gli aplotipi parentali con gli alleli normale e mutato. Per poter definire gli aplotipi parentali e associare agli stessi gli alleli delle mutazioni in esame, abbiamo applicato la metodica molecolare dell'amplificazione allele-specifica, attraverso primer ARMS (Amplification Refractory Mutation System), per le amplificazioni di segmenti estesi di DNA parentali. Attraverso l'utilizzo di primer allele specifici è stato possibile amplificare dei frammenti lunghi tra le 4 e 7 kilobasi in direzione 3' e 5' rispetto alla posizione della mutazione in esame, consentendo, a differenza del protocollo basato sulla amplificazione selettiva degli ampliconi, di includere nell'analisi tutti gli SNP informativi presenti nei frammenti prodotti. Questo approccio allele selettivo consente la definizione immediata degli aplotipi e la loro associazione all'allele mutato o normale del sito di mutazione osservato. La selezione è possibile grazie all'impiego di coppie di primer differenti fra loro in posizione 3': ciascuno di essi in questa posizione presenta una base complementare o all'allele mutato o al normale. L'amplificazione perciò consentirà la amplificazione selettiva dell'unico filamento complementare al primer utilizzato nella reazione.

Durante questa fase sperimentale la principale difficoltà incontrata è stata l'ottimizzazione delle condizioni ideali di amplificazione: da un lato, è infatti necessario l'utilizzo di elevate temperature di melting richieste dai primer ARMS per un'efficace separazione degli alleli; dall'altro, condizioni di amplificazione più restrittive non favoriscono la sintesi di lunghi frammenti di DNA. Il primo amplicone ad essere prodotto è stato un frammento in direzione 3' rispetto al gene HBB, che si estende per circa 4Kb a partire dal sito di mutazione  $\beta^{39}$ , e comprende quattro loci polimorfi, già presenti nell'elenco iniziale degli SNP utilizzati nelle nostre analisi. Ciascuno dei due prodotti di Long PCR derivanti dalle due amplificazioni allele specifiche è stato sequenziato con metodo Sanger per la definizione degli alleli degli SNP. In questo modo è stato possibile ricostruire le fasi degli aplotipi nella zona 3' del gene HBB, e definire, per le 5 famiglie per le quali non era stato possibile, l'associazione fra aplotipo e allele della mutazione. Questi dati supplementari sono quindi stati integrati con i dati ottenuti



in precedenza. I risultati complessivi confermano l'efficacia del metodo proposto nelle determinazione delle mutazioni a carico del feto. I risultati ottenuti, schematizzati nella tabella 14, possono essere così descritti:

- 9 famiglie su 18, diagnosi corretta;
- 7 famiglie su 18, diagnosi indeterminata al 50% (i dati presenti sono coerenti con l'atteso);
- 1 famiglia su 18, diagnosi errata;
- 1 famiglia su 18 non analizzabile per assenza di informazioni.

Esclusi i casi in cui non è stato possibile applicare la metodica per assenza di SNP informativi, si può affermare che essa ha presentato affidabilità del 100% nella determinazione dell'aplotipo paterno e del 91.6% nella determinazione dell'aplotipo materno.

famiglia	Genotipo fetale		Risultati analisi SPRT		
	Aplo. Paterno - Materno	Diagnosi	Aplo. paterno	Aplo. materno	Diagnosi
16	$\beta^{\circ}39$ $\beta^{\circ}6$	Affetto	$\beta^{\circ}39$	(no data)	Portatore/Affetto
2	$\beta$ IVS II.745 WT	Portatore	$\beta$ IVS II.745	WT	Portatore
3	$\beta$ IVS I.110 $\beta^{\circ}39$	Affetto	$\beta$ IVS I.110	$\beta^{\circ}39$	Affetto
18	$\beta$ IVS I.110 $\beta^{\circ}39$	Affetto	$\beta$ IVS I.110	$\beta^{\circ}39$	Affetto
17	$\beta^{\circ}6$ $\beta^{\circ}39$	Affetto	$\beta^{\circ}6$	$\beta^{\circ}39$	Affetto
12	WT WT	Normale	WT	WT	Normale
7	WT WT	Normale	WT	(no data)	Normale/Portatore
13	WT WT	Normale	WT	(no data)	Normale
14*	WT WT	Normale	WT	$\beta^{\circ}39$	Portatore
15	$\beta^{\circ}39$ $\beta^{\circ}39$	Affetto	$\beta^{\circ}39$	(no data)	Portatore/Affetto
1	$\beta^{\circ}39$ $\beta^{\circ}39$	Affetto	$\beta^{\circ}39$	$\beta^{\circ}39$	Affetto
8	$\beta^{\circ}39$ $\beta^{\circ}39$	Affetto	(no data)	$\beta^{\circ}39$	Portatore/Affetto
10	$\beta^{\circ}39$ $\beta^{\circ}39$	Affetto	(no data)	$\beta^{\circ}39$	Portatore/Affetto
4	WT $\beta^{\circ}39$	Portatore	WT	$\beta^{\circ}39$	Portatore
5	WT $\beta^{\circ}39$	Portatore	WT	(no data)	Portatore/Affetto
6	WT $\beta^{\circ}39$	Portatore	WT	$\beta^{\circ}39$	Portatore
9	WT $\beta^{\circ}39$	Portatore	(no data)	$\beta^{\circ}39$	Normale/Portatore
11*	WT $\beta^{\circ}39$	Portatore	(no data)	(no data)	(no data)

Tabella 14 Risultato finale dell'analisi delle famiglie dopo l'utilizzo della metodica Long PCR

I risultati positivi ottenuti attraverso la sintesi dei frammenti di 4Kb in 3', sono serviti da impulso per la ulteriore messa a punto del metodo basato sulle

delle Long PCR allele specifiche. In successivi esperimenti è stato possibile ottenere l'amplificazione selettiva di frammenti lunghi circa 13Kb in direzione 5' rispetto al gene HBB, per una sequenza totale di circa 17Kb allele specifica. Negli esperimenti relativi alle ultime 4 famiglie studiate i DNA parentali sono stati analizzati attraverso preventiva amplificazione allele specifica, successiva frammentazione e sequenziamento massivo. Attraverso questo approccio è stata possibile, oltre l'immediata separazione delle fasi degli aplotipi, la valutazione di tutti gli SNP compresi nei frammenti amplificati, con l'inclusione di circa 50 ulteriori SNP, emersi dall'analisi dei genomi parentali, non presenti tra quelli selezionati inizialmente. Questo approccio utilizzato per l'amplificazione dei DNA parentali presenta due importanti aspetti positivi: da un lato esso contrasta efficacemente il rischio di scarsa informatività nella trasmissione degli aplotipi spesso dovuta all'esiguo numero di SNP osservati; dall'altro, poiché consente l'immediata definizione delle fasi degli aplotipi parentali, rende superfluo l'utilizzo del dato genotipico derivato dai villi coriali, condizione imprescindibile per lo sviluppo di un protocollo di diagnosi prenatale non invasiva.

## CONCLUSIONI

Nell'ultimo decennio si è assistito, nel campo della biologia molecolare, alla nascita e allo sviluppo di un nuovo settore tecnologico che ha determinato un profondo cambiamento nella metodologia di sequenziamento degli acidi nucleici, aprendo la via a nuove opportunità in tutti quei settori in cui l'analisi del DNA è alla base delle procedure di indagine. È la tecnologia chiamata “Next Generation Sequencing”, caratterizzata da strumenti e protocolli d'analisi ad altissima processività, ha favorito la pubblicazione di numerosi lavori scientifici e lo sviluppo di nuove procedure d'indagine clinica in molteplici settori della genetica molecolare. Come esempio delle potenzialità di questa tecnologia si può citare l'ambizioso progetto 1000 Genomi, portato avanti da Wellcome Trust Sanger Institute di Hinxton, l'Istituto di Genomica di Pechino, e il National Human Genome Research Institute che si propone di effettuare il sequenziamento di 1000 genomi umani mediante NGS, con lo scopo principale di definire nuove patologie genetiche o di determinare le varianti causative di quelle patologie per le quali ancora esse non sono note. Ancora, le tecnologie NGS hanno apportato un forte contributo nel campo dell'oncologia, permettendo per alcune patologie tumorali, di rilevare con notevole anticipo, rispetto al manifestarsi della malattia, la presenza di biomarcatori nel circolo venoso periferico aumentando la probabilità di guarigione. Un altro settore, in cui le tecnologie NGS hanno consentito lo sviluppo di nuovi ed efficaci protocolli d'analisi, alcuni dei quali già da tempo introdotti nelle pratiche clinico-diagnostiche, è quello della diagnosi prenatale non invasiva. In questo ambito, le potenzialità delle tecnologie NGS sono state sfruttate per l'analisi del DNA fetale libero presente nel circolo periferico materno determinando una vera e propria rivoluzione nell'ambito degli screening prenatali. In particolare, la loro applicazione ha riscontrato forte successo nella diagnosi non invasiva delle più comuni aneuploidie fetali.

Sin dalla sua scoperta, il DNA fetale libero ha rappresentato una valida sorgente di materiale genetico per la definizione del genotipo fetale in epoca prenatale precoce, con l'importante vantaggio che i protocolli diagnostici basati sulla sua analisi non comportano alcun rischio per il feto. La diagnosi prenatale,

fino alla scoperta del DNA fetale, era infatti basata esclusivamente su tecniche di prelievo invasive, alle quali ancora adesso è associato un rischio di perdita fetale stimato intorno allo 0.5-1%. La possibilità di ottenere materiale fetale, necessario per la definizione del genotipo del feto, senza dover sottoporre la donna a intervento invasivo, è stato da sempre l'obiettivo di numerosi gruppi di ricerca. Il DNA fetale libero presente nel plasma materno, grazie alla sua facile reperibilità, alla sua rapida eliminazione dal circolo materno poche ore dopo il parto, e alla semplicità e affidabilità degli approcci diagnostici, da subito ha mostrato la sua utilità nelle pratiche diagnostiche prenatali ed è attualmente considerato la principale fonte da cui poter ricavare informazioni sul genoma fetale in maniera non invasiva.

Nelle prime pubblicazioni che avevano per oggetto il DNA fetale veniva dimostrata la possibilità di evidenziare, nel DNA estratto da plasma, i caratteri fetali, in particolare quelli ereditati per via paterna e non presenti nel genoma della madre. Le prime applicazioni a carattere diagnostico per le quali è stato impiegato il cfDNA sono diverse: la determinazione del sesso fetale nelle malattie X-linked (17) (18) (19), la genotipizzazione del fattore Rh fetale (23), la definizione di patologie *de-novo*, la determinazione della trasmissione delle varianti patologiche nelle malattie autosomiche dominanti ad ereditarietà paterna, la trasmissione degli alleli mutati paterni nelle monogeniche recessive, tutti protocolli nei quali si ricerca la presenza di frammenti di derivazione paterna assenti nel genoma materno. Per il loro elevato grado di affidabilità questi protocolli sono stati i primi ad essere trasferiti dal campo della ricerca alla diagnostica clinica. Importanti risultati sono stati anche raggiunti nell'ambito delle patologie monogeniche, attraverso la ricerca dell'allele causativo paterno nel DNA libero. Dimostrando l'assenza di tale allele, infatti, può essere esclusa la presenza della patologia (23-27).

La prima, tra le tecnologie ad alta sensibilità, ad essere stata utilizzata nell'analisi del cfDNA è stata la Digital PCR. Essa consente, infatti, di poter determinare le frazioni relative agli alleli normale e mutato rappresentate nell'intero campione di cfDNA analizzato. Ai risultati ottenuti con la Digital vengono applicati dei metodi che consentano la definizione del genotipo fetale di un sito polimorfo, come il Relative Mutation Dosage, basato sul calcolo della

sovraespressione dell'allele normale rispetto al corrispondente mutato. Nonostante la Digital PCR sia una tecnologia efficace, gli studi basati sull'analisi del DNA fetale hanno subito un notevolissimo incremento con la disponibilità delle tecnologie di Next Generation Sequencing. Esse, analogamente alla Digital PCR, permettono di evidenziare l'effetto del DNA fetale nella totalità del DNA plasmatico, mostrando però, rispetto ad essa, un maggior potere d'analisi. Tra i vantaggi da esse apportati possiamo annoverare da un lato, grazie alla capacità di analisi contemporanea di numerosi siti polimorfi, la possibilità di sviluppare protocolli d'analisi per le malattie monogeniche, mediante il supporto di metodi statistici come l'*RHDO/SPRT*, utili per la valutazione della trasmissione degli aplotipi. Dall'altro, data la loro capacità di sequenziare ampie regioni genomiche, sono state efficacemente utilizzate per la determinazione delle aneuploidie fetali, con particolare riferimento alla trisomia 21, attraverso la valutazione dei sovradosaggi delle regioni cromosomiche. In relazione alle aneuploidie, le tecnologie NGS hanno mostrato efficienza e affidabilità tali che alcuni paesi Europei e degli Stati Uniti i protocolli d'analisi sono già inseriti nelle pratiche cliniche di screening per le trisomie 21, 18 e 13.

Più difficoltosa, si è rivelata l'applicazione di queste tecnologie alla diagnosi prenatale non invasiva delle malattie monogeniche. In questo caso, la definizione del genotipo fetale si rivela di più difficile determinazione, soprattutto nei casi in cui i genitori risultano portatori della stessa mutazione patogenetica, poiché in questo caso madre e feto condividono gli stessi alleli e la frazione di DNA di origine materna maschera l'effetto di quella fetale. Proprio in questo ambito di studio si è sviluppata la nostra ricerca, con l'intento di sviluppare un protocollo d'analisi utile per la diagnosi prenatale non invasiva di malattie monogeniche recessive, in particolare per la  $\beta$ -talassemia.

Pur essendo per definizione una malattia rara, la  $\beta$ -talassemia in Sardegna riveste un ruolo importante dal punto di vista sociale e sanitario: circa l'11% della popolazione sarda è portatore sano per la patologia, con il 95% degli alleli patologici rappresentati dalla mutazione per il codone 39 del gene **HBB**. Questi dati implicano che un elevato numero di coppie accedano alla diagnosi prenatale invasiva per ottenere informazioni circa il genotipo fetale relativamente alle

mutazioni per il gene **HBB**. La presenza di un protocollo di diagnosi non invasiva permetterebbe quindi di sottoporre le donne ad un iter diagnostico più semplice e in totale assenza di rischi per il feto.

A questo proposito abbiamo sviluppato un protocollo di analisi del cfDNA, mediante l'utilizzo della tecnologia Ion Torrent PGM, per la definizione del genotipo fetale nelle coppie a rischio per  $\beta$ -talassemia. Lo sviluppo di tale protocollo ha presentato due particolari difficoltà. La prima, dovuta alla elevata omogeneità genetica della popolazione, a causa della quale è possibile riscontrare nelle famiglie da analizzare l'assenza di SNP informativi, situazione che impedisce la determinazione degli alleli fetali nel cfDNA. La seconda, dovuta all'assenza di qualsiasi protocollo o algoritmo d'analisi appositamente costruito per l'utilizzo con l'Ion Torrent PGM, che potesse essere sfruttato per l'analisi del cfDNA. Inizialmente gli esperimenti sono stati infatti condotti senza possedere alcuna precisa informazione sull'effettiva capacità della piattaforma nel determinare la quota del DNA fetale.

Per gli studi dei trios familiari sono stati applicati i metodi RMD e RHDO/SPRT. I dati ottenuti hanno confermato l'utilità della tecnologia Ion Torrent unitamente alle potenzialità degli approcci RMD e RHDO nell'analisi del DNA fetale libero, mettendone però in evidenza anche alcuni punti deboli per i quali è necessario apportare dei miglioramenti.

In una prima fase del nostro studio abbiamo valutato l'attendibilità e l'efficacia dell'approccio RMD nella definizione del genotipo fetale. I risultati ottenuti hanno mostrato come questo metodo consenta solo parzialmente la corretta determinazione del genotipo fetale per la mutazione causativa. Su 18 famiglie infatti solo in 10 casi si è potuto evidenziare il corretto genotipo fetale, e la maggior parte dei risultati errati (precisamente in 4 famiglie su 5) sono stati riscontrati nelle famiglie con genitori e feto eterozigoti per la mutazione  $\beta^039$ . Nei trios con genitori portatori di mutazioni differenti, la mutazione paterna è stata sempre evidenziata correttamente (5 casi su 5) mentre quella materna è stata dedotta correttamente solo in 3 casi su 5. Dopo queste prime analisi è stato possibile affermare che, l'allele paterno, quando non è presente nel genoma materno viene sempre efficacemente rilevato; al contrario, il metodo RMD

applicato per la determinazione degli alleli materni, non si è dimostrato un approccio ottimale per la valutazione del genotipo del feto, e ha mostrato particolare perdita di efficacia nei casi in cui madre e feto siano entrambi eterozigoti. Inoltre, un ulteriore aspetto svantaggioso è dato dal fatto che, nell'approccio RMD, tutta la diagnosi è basata sull'analisi del solo sito di mutazione; un errore sperimentale casuale o uno sbilanciamento scorretto o non rilevabile nel sito di mutazione, inficiano l'intero risultato diagnostico.

Con lo scopo di ottenere risultati più consistenti, è stato successivamente proposto l'utilizzo di un metodo che si basasse sull'osservazione parallela di più SNP, associati fra loro in aplotipi: l'**RHDO** accompagnato dal test statistico **SPRT**. Questo approccio è caratterizzato da un maggior potere d'analisi rispetto all'**RMD**, in quanto, il suo risultato è definito, non da una, ma da un ampio numero di osservazioni, e questo conferisce minor suscettibilità agli errori. Per contro, esso è fortemente dipendente dal numero degli SNP informativi analizzabili, in assenza dei quali si rivela del tutto inapplicabile. Questa difficoltà è emersa anche nel corso dei nostri studi. Come mostrato nelle tabelle precedenti (12, 13) in 6 casi su 18 per gli aplotipi materni (33%), e in 4 per quelli paterni (22%) non è stato possibile riscontrare alcun polimorfismo informativo utile per la definizione dell'ereditarietà fetale. Per queste famiglie, 7 casi su 18, i genotipi fetali sono stati determinati al 50%, poiché definiti solo da un aplotipo parentale. In questi casi se l'allele di cui è nota la trasmissione è normale, questo risultato è sufficiente per affermare l'assenza della patologia; viceversa, se la trasmissione nota è relativa all'allele mutato, il fenotipo malato non può essere escluso. In sintesi, il **Relative Haplotype Dosage**, supportato dall'analisi **SPRT**, ha mostrato efficacia e attendibilità ampiamente superiori rispetto all'**RMD**. I dati analizzati attraverso questo metodo hanno presentato, infatti, un solo risultato errato contro gli otto dell'analisi **RMD**.

Un fattore limitante sia per l'analisi **RMD** che per l'**RHDO**, è la frazione della quota fetale, che in gravidanze in epoca gestazionale molto precoce si può rivelare insufficiente per poter determinare l'effetto del DNA fetale. Tra quelli da noi analizzati, è presente un campione, in 6<sup>a</sup> settimana di gestazione, la cui quota fetale si è rivelata troppo bassa per poter essere rilevata dallo strumento; per questo campione non è stata possibile l'estrapolazione di alcun dato relativo alle

fasi materna e paterna. Questo conferma ciò che in letteratura più volte viene messo in evidenza, cioè che l'analisi del DNA fetale deve essere effettuata in epoca superiore alle 9-10 settimane di gestazione affinché la frazione fetale raggiunga concentrazioni tali da poter essere rilevate con efficacia e i dati ricavati possano essere ritenuti attendibili.

Nelle analisi condotte attraverso il metodo RHDO, una delle difficoltà maggiori verificatesi è stata l'associazione degli aplotipi parentali agli alleli normale e mutato della variante patogenetica, in particolar modo in quelle famiglie in cui padre, madre e feto presentano tutti genotipo eterozigote per la stessa mutazione. Inizialmente, infatti, per cinque delle famiglie in esame era stato possibile definire la trasmissione degli aplotipi materno e paterno, senza però poter associare ad essi gli alleli normale e mutato della variante patogenetica e definire così il genotipo fetale. Questo ostacolo è stato superato attraverso un approccio molecolare di Long PCR allele-selettiva utilizzata nel processo di amplificazione dei DNA parentali; esse sono state eseguite mediante l'utilizzo di kit specifici per la sintesi di lunghi frammenti, congiuntamente all'impiego di primer ARMS allele specifici. È stato possibile costruire lunghi frammenti dei DNA parentali, selezionati in base alla presenza, sui filamenti amplificati, degli alleli mutato o normale. Questo approccio ha consentito il doppio vantaggio di ottenere, per ciascuna delle due reazioni allele specifiche, sia la fase degli aplotipi parentali sia la determinazione di quali aplotipi siano associati agli alleli mutato e normale. L'applicazione di questo protocollo ha permesso di colmare la mancanza di informazioni relative alle famiglie 4, 5, 6 e 9 consentendo, anche in questi casi, un'attribuzione del corretto genotipo fetale.

In sintesi, i risultati ottenuti ci consentono di affermare che il metodo RHDO / SPRT, associato a un protocollo di amplificazione basato su Long PCR allele-selettiva, ha dimostrato di essere un approccio valido per la definizione dei genotipi fetali nella diagnosi prenatale non invasiva di  $\beta$ -talassemia. Su 18 famiglie analizzate sono stati ottenuti i seguenti risultati:

- 9 famiglie in cui la diagnosi è stata eseguita correttamente;
- 1 famiglia per la quale si è riscontrato un risultato errato (per questa famiglia, già prima della analisi del genotipo fetale erano stati messi in



evidenza errori sperimentali tali da far supporre imprecisione nel risultato finale);

- 1 famiglia per la quale non è stato possibile effettuare la diagnosi a causa di assenza di SNP informativi;
- 7 famiglie per le quali è stato possibile eseguire soltanto una diagnosi al 50%, con la definizione della trasmissione di un solo alplotipo parentale.

Questi risultati permettono di affermare come il metodo da noi proposto presenti elevato potere diagnostico nella determinazione dei genotipi fetali mediante l'analisi del DNA fetale da plasma materno, in famiglie a rischio per  $\beta$ -talassemia, anche nei casi in cui entrambi i genitori siano portatori della medesima mutazione. Nonostante l'elevata omogeneità genetica che caratterizza la nostra popolazione e la scarsa informatività presente negli alplotipi parentali, anche nei casi di diagnosi al 50%, i risultati parziali si sono dimostrati corretti; l'analisi di un maggior numero di SNP avrebbe consentito la precisa definizione dei genotipi fetali anche in queste famiglie. L'unico dato non coerente con l'atteso è stato riscontrato in un'analisi inficiata da errori commessi durante il protocollo sperimentale.

L'inserimento del presente protocollo tra le attuali pratiche diagnostiche prenatali per la  $\beta$ -talassemia consentirebbe di evitare di sottoporre le donne alle indagini di carattere invasivo, con l'importante vantaggio dell'annullamento del rischio di perdita fetale e dell'abbattimento dei costi legati all'iter diagnostico.

Questo approccio, inoltre, benchè sviluppato per la  $\beta$ -talassemia, è facilmente trasferibile alla diagnosi non invasiva di altre patologie monogeniche.

Nel prossimo futuro sarà necessaria l'implementazione di due aspetti essenziali del protocollo proposto. Da un lato, sarà necessario l'ampliamento della casistica delle famiglie analizzate, fondamentale per la validazione del metodo e il suo inserimento tra i test diagnostici prenatali, e per la standardizzazione dell'intero flusso di lavoro. Dall'altro, considerata la grande quantità di dati che è necessario valutare per la determinazione della trasmissione degli alplotipi, e vista la necessità di introduzione di ulteriori polimorfismi fondamentali per una loro più precisa definizione, sarà di primaria importanza affiancare al protocollo sperimentale una struttura informatica che consenta l'acquisizione dei dati in maniera semplice ed

automatizzata, e che esegua una diretta valutazione del rischio per la patologia, escludendo la necessità di manipolazioni e modificazioni dei dati da parte dell'operatore.

## BIBLIOGRAFIA

1. *Diagnosi prenatale ed ecografica in gravidanza- linee guida 2013*. **Agenzia servizi sanitari**. s.l. : Agenzia nazionale per i servizi sanitari regionali, 2013.
2. *Microsatellite alterations in plasma DNA of small cell lung cancer patients*. **Chen, XQ**. s.l. : Nat. Med., 1996.
3. *Presence of fetal DNA in maternal plasma and serum*. **Lo, YMD and Corbetta, N**. s.l. : Lancet, 1997.
4. *Quantitative analysis of fetal DNA in maternal plasma and serum: implications for noninvasive prenatal diagnosis*. **Lo, YMD**. s.l. : Am. J. Hum. Genet., 1998.
5. *Microfluidics digital PCR reveals a higher than expected fraction of fetal DNA in maternal plasma*. **Lun, FMF, Chiu, RW and Lo, YMD**. s.l. : Clin. Chem., 2008.
6. *Maternal plasma DNA sequencing reveals the genome-wide genetic and mutational profile of the fetus*. **Lo, YMD, Allen Chan, KC and Chiu, Rossa WK**. s.l. : Prenatal Diagnosis, 2010.
7. *DNA sequencing of maternal plasma to detect Down syndrome: an international clinical validation study*. **Palomaki, GE and al**. s.l. : Genet. Med., 2011.
8. *DNA sequencing of maternal plasma reliably identifies trisomy 18 and trisomy 13 as well as Down syndrome: an international collaborative study*. **Palomaki, GE, Canick, JA and al**. s.l. : Genet. Med., 2012.
9. *Rapid clearance of fetal DNA from maternal plasma*. **Lo, YMD, Zahng, J and Hjelm, NM**. s.l. : Am. J. Hum. Genet., 1999.
10. *High-Resolution Profiling of Fetal DNA Clearance from Maternal Plasma by Massively Parallel Sequencing*. **Yu, Stephanie CY, Chiu, Rossa WK and Lo, YMD**. s.l. : Clinical Chemistry, 2013.
11. *Size separation of circulatory DNA in maternal plasma permits ready detection of fetal DNA polymorphisms*. **Li, Y., Zimmermann, B. and Hahn, S**. s.l. : Clinical Chemistry, 2004.
12. *Maternal Plasma DNA Sequencing Reveals the Genome-Wide Genetic and mutational profile of the fetus*. **Lo, YMD, Chan, KCA and Chiu, RWK**. s.l. : Prenatal Diagnosis, 2010.
13. *Free fetal DNA in maternal plasma in anembryonic pregnancies: confirmation that the origin is the trophoblast*. **Alberry, M., Maddocks, D. and Soothill, PW**. s.l. : Prenatal Diagnosis, 2007.

14. *Hypermethylated RASSF1A in maternal plasma: A universal fetal DNA marker that improves the reliability of noninvasive prenatal diagnosis.* **Chan, KC., et al.** s.l. : Clinical Chemistry, 2006.
15. *Fetal-specific DNA methylation ratio permits noninvasive prenatal diagnosis of trisomy 21.* **Papageorgiou, EA., Karagrigoriou, A. and Patsalis, PC.** s.l. : Nature Medicine, 2011.
16. *New Strategy for Prenatal Diagnosis of X-linked disorders.* **Costa, JM, Benachi, A. and Gautier, E.** s.l. : N. Eng. Journal of Medicine, 2002.
17. *Fetal sex determination from maternal plasma in pregnancies at risk for congenital adrenal hyperplasia.* **Rijinders, R.J. and Christiaens, G.C.** s.l. : Obstet. Gynecol., 2001.
18. *Earliest gestational age for fetal sexing in cell-free maternal plasma.* **Rijinders, R.J. and Christiaens, G.C.** s.l. : Prenatal Diagnosis, 2003.
19. *Noninvasive fetal sex determination using cell-free fetal DNA: a systematic review and meta-analysis.* **Devaney, S.A., Palomaki, G.E. and Bianchi, D.W.** s.l. : JAMA, 2011.
20. *Noninvasive prenatal diagnosis of hemophylia by microfluidics digital PCR analysis of maternal plasma.* **Tsui, NBY, et al.** s.l. : Blood, 2011.
21. *Prenatal Diagnosis of Fetal RhD Status by Molecular Analysis of Maternal Plasma.* **Lo, Y.M.D., Magnus, H. and Wainscoat, J.S.** s.l. : New England Journal of Medicine, 1998.
22. *Prenatal RhD Testing: A Review of Studies Published from 2006 to 2008.* **Legler, T.J., Muller, S.P. and Hahn, S.** s.l. : Transf. Medicine and Hemotherapy, 2009.
23. *Fetal RHD Genotyping in Maternal Plasma at 11-13 weeks of gestation.* **Akolekar, R., Finning, K. and Nicolaides, K.H.** s.l. : Fetal Diagnosis and Therapy, 2011.
24. *Noninvasive prenatal exclusion of congenital adrenal hyperplasia by maternal plasma analysis: a feasibility study.* **Chiu, RWK, Lau, TK and Lo, YMD.** s.l. : Clinical Chemistry, 2002.
25. *Prenatal exclusion of beta thalassaemia major by examination of maternal plasma.* **Chiu, RW, Lau, TK and Lo, YMD.** s.l. : Lancet, 2002.
26. *MS analysis of single-nucleotide differences in circulating nucleic acids: Application to noninvasive prenatal diagnosis.* **Ding, C, Chiu, RWK and Lo, YMD.** s.l. : PNAS, 2004.
27. *Genomic analysis of fetal nucleic acids in maternal blood.* **Lo, YMD and Chiu, RWK.** s.l. : Annu. Rev. Genomics Hum. Gen, 2012.
28. *Noninvasive prenatal diagnosis of mitonogenic diseases by digital size selection and Relative Mutation Dosage on DNA in maternal plasma.* **Lun, FMF, Tsui, NBY and Lo, YMD.** s.l. : PNAS, 2008.
29. *Digital PCR.* **Vogelstein, B and Kinzler, KW.** s.l. : PNAS, 1999.

30. *Digital PCR analysis of maternal plasma for noninvasive detection of sickle cell anemia.* **Barrett, A.N., Chan, KCA and Chitty, L.S.** s.l. : Clinical Chemistry, 2012.
31. *A new method for sequencing DNA.* **Maxam, A.M. and Gilbert, W.** s.l. : PNAS, February 1977.
32. *DNA sequencing with chain-terminating inhibitors.* **Sanger, F., Nicklen, S. and Coulson, R.A.** s.l. : PNAS, December 1977.
33. *Genome sequencing in microfabricated high-density picolitre reactors.* **Margulies, M.** s.l. : Nature, 2005, Vol. 437.
34. *An integrated semiconductor device enabling non-optical genome sequencing.* **Rothberg, J.M.** s.l. : Nature, 2011, Vol. 475.
35. *Noninvasive detection of fetal trisomy 21 by sequencing of DNA in maternal blood: a study in a clinical setting.* **Ehrich, MMD and al, et.** s.l. : Am. Journal Obstetrics Gynecology, 2011.
36. *noninvasive prenatal diagnosis of fetal trisomy 21 by allelic ratio analysis using targeted massively parallel sequencing of maternal plasma.* **Liao, GJW, Chiu, RWK and Lo, YMD.** s.l. : Plos ONE, 2012.
37. *Noninvasive prenatal testing using massively parallel sequencing of maternal plasma DNA: from molecular karyotyping to fetal whole-genome sequencing.* **Lo, YMD.** s.l. : Reproductive BioMedicine Online, 2013.
38. *Noninvasive prenatal diagnosis of fetal trisomy 18 and trisomy 13 by maternal plasma DNA sequencing.* **Chen, EZ, Nicolaides, KH and Lo, YMD.** s.l. : PLoS ONE, 2011.
39. *Non-invasive prenatal diagnostics of aneuploidy using next-generation sequencing technologies, and clinical considerations.* **Nepomnyashchaya, YN and Zhavoronkov, A.** s.l. : Clin. Chem. Lab. Med., 2013.
40. *Noninvasive detection of fetal subchromosome abnormalities via deep sequencing of maternal plasma.* **Srinivasan, A.** s.l. : Am. J. Human Gen., 2013.
41. *Maternal Plasma DNA Sequencing Reveals the Genome-Wide Genetic and mutational profile of the fetus.* **Lo, YMD, Chan, KCA and Chiu, RWK.** s.l. : Prenatal Diagnosis, 2010.
42. *Noninvasive Prenatal Diagnosis of monogenic diseases by target massively parallel sequencing of maternal plasma: application to Beta-thalassemia.* **Lam, KWG, Chiu, RWK and Lo, YMD.** s.l. : Clinical Chemistry - Molecular diagnostic end genetics, 2012.
43. *A minimal set of SNPs for the noninvasive prenatal diagnosis of Beta-Thalassaemia.* **Papasavva, TE and Kleanthous, M.** s.l. : Annals of Human Genetics, 2013.
44. *WMA Declaration of Helsinki - Ethical Principles for Medical Research Involving Human Subjects.*

# INDICE DELLE FIGURE

FIGURA 1: ILLUSTRAZIONE SCHEMATICA DEL RELATIVE MUTATION DOSAGE (DA LUN ET AL., PNAS 2008) .....	16
FIGURA 2: SCHEMA RMD PER LE MALATTIE AUTOSOMICHE E X-LINKED.....	17
FIGURA 3: SCHEMA EMULSION PCR (MODIFICATO DA M. L. METZKER - NATURE REVIEW GENETICS 2010) .....	24
FIGURA 4: 454 ROCHE: RILEVAZIONE DEL SEGNALE NEL PIROSEQUENZIAMENTO (MODIFICATO DA T. TUCKER ET AL. 2009 - AJHG).....	25
FIGURA 5: AMPLIFICAZIONE CLONALE ATTRAVERSO BRIDGE PCR (MODIFICATO DA M. L. METZKER - NATURE REVIEW GENETICS 2010) .....	26
FIGURA 6: INCORPORAZIONE DEI dNTP FLUORESCINATI E LETTURA DELLA SEQUENZA (MODIFICATO DA M. L. METZKER - NATURE REVIEW GENETICS 2010).....	26
FIGURA 7: SCHEMA DEL PROTOCOLLO SOLID (MODIFICATO DA T. TUCKER ET AL. 2009 - AJHG).....	27
FIGURA 8: ION PGM™ SYSTEM - PERSONAL GENOME MACHINE .....	28
FIGURA 9 PARTICOLARI DELL'ION CHIP. NELL'IMMAGINE CENTRALE, SEZIONE DEL CHIP CON EVIDENTI I POZZETTI SUPERFICIALI. A DESTRA, SCHEMA DEL PROCESSO DI RILEVAZIONE DEL SEGNALE. ....	29
FIGURA 10: SCHEMA DI SEQUENZIAMENTO NELL'ION CHIP. PRIMERS, DNA POLIMERASI E BEADS 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 È TRADOTTA NELLA CHIAMATA DELLA BASE. ....	30
FIGURA 11: ESEMPIO DI IONOGRAMMA PRODOTTO DAL PGM™ ION TORRENT .....	31
FIGURA 12: CONFRONTO TRA DIVERSE PIATTAFORME ILLUMINA E L'ION TORRENT PGM™ SYSTEM .....	32
FIGURA 13: ION ONE TOUCH ES E ION ONE TOUCH DL .....	32
FIGURA 14: RELATIVE HAPLOTYPE DOSAGE ANALYSIS E SEQUENTIAL PROBABILITY RATIO TEST ( DA YMD LO ET AL. 2010) .	40
FIGURA 15 SCHEMA CLUSTER β-GLOBINICO E SNP ANALIZZATI .....	47
FIGURA 16 VALUTAZIONE DELL'IPOTESI DI TRASMISSIONE DELL'APLOTIPO. L'ULTIMO PUNTO A DESTRA NEL GRAFICO RAPPRESENTA IL VALORE OTTENUTO SOMMANDO GLI EFFETTI DI TUTTI GLI SNP DELL'APLOTIPO. IL PUNTO GIACE AL DI SOPRA DELLA CURVA E CIÒ CONFERMA L'IPOTESI DI TRASMISSIONE .....	63

# INDICE DELLE TABELLE

TABELLA 1 REAGENTI UTILIZZATI NELLA PRIMA REAZIONE DI AMPLIFICAZIONE E SELEZIONE DEI FRAMMENTI.....	48
TABELLA 2 SCHEMA DEI TEMPI E DELLE TEMPERATURE DI AMPLIFICAZIONE SU VERITI THERMAL CYCLER .....	48
TABELLA 3 SCHEMA DEI TEMPI E DELLE TEMPERATURE DI AMPLIFICAZIONE PER IL FRAMMENTO SUL GENE SRY .....	48
TABELLA 4 ELENCO PRIMER ARMS .....	49
TABELLA 5 ELENCO REAGENTI UTILIZZATI NELLA REAZIONE ALLELE SPECIFICA .....	50
TABELLA 6 SCHEMA FILE DELLE TEMPERATURE PER L'AMPLIFICAZIONE ALLELE SPECIFICA .....	50
TABELLA 7 MISCELA PER LA REAZIONE DI AMPLIFICAZIONE CLONALE (EMULSION PCR) .....	51
TABELLA 8 SUDDIVISIONE DEGLI SNP IN BASE ALL'INFORMAZIONE FORNITA (DA YMD LO ET AL. 2010) .....	55
TABELLA 9 CONFRONTO DATI RELATIVI ALL'ANALISI DELLE PRIME CINQUE FAMIGLIE IN PRESENZA E ASSENZA DEGLI SNP DISTRIBUITI SU CROMOSOMI DIVERSI DAL CROMOSOMA 11 .....	56
TABELLA 10 QUOTA FETALE CALCOLATA IN BASE ALL'OSSERVAZIONE DEGLI SNP DI TIPO 1 E 3.....	57
TABELLA 11 GENOTIPI PARENTALI E GENOTIPO FETALE DEDOTTO DA ANALISI SU VILLO CORIALE .....	58
TABELLA 12 RISULTATI ANALISI FAMIGLIE ATTRAVERSO RMD .....	59
TABELLA 13 RISULTATO DELL'ANALISI DELLE FAMIGLIE ATTRAVERSO IL METODO RHDO-SPRT. ....	61
TABELLA 14 RISULTATO FINALE DELL'ANALISI DELLE FAMIGLIE DOPO L'UTILIZZO DELLA METODICA LONG PCR .....	65

# SOMMARIO

La diagnosi prenatale .....	3
Tecniche di screening non invasive .....	4
Tecniche di prelievo invasive .....	5
DNA fetale .....	7
Caratteristiche del DNA Fetale .....	7
Determinazione del sesso fetale .....	10
Determinazione del fattore Rh .....	11
Applicazioni dell' Analisi del cfDNA nelle malattie monogeniche.....	13
Ulteriori applicazioni nelle malattie monogeniche.....	15
Next generation sequencing .....	20
Caratteristiche del Sequenziamento Next Generation .....	20
Tecnologie NGS.....	23
Ion Torrent Technology.....	28
Applicazioni del Sequenziamento Next Generation .....	33
Tecnologie NGS nella diagnosi delle Aneuploidie.....	33
Tecnologie NGS e diagnosi di malattie monogeniche.....	36
Attività di ricerca .....	44
Obiettivi.....	44
Materiali e Metodi.....	44
Campioni e Estrazione del DNA.....	44
Amplificazione.....	45
Long PCR Allele Specifica.....	49
Preparazione librerie ed Emulsion PCR .....	50
Sequenziamento e Analisi dei dati .....	52
Risultati .....	54
Conclusioni .....	67
Bibliografia.....	75
Indice delle Figure .....	78
Indice delle Tabelle .....	79





## MATERNAL PLASMA DNA ANALYSIS BY NEXT GENERATION SEQUENCING FOR NON INVASIVE PRENATAL DIAGNOSIS OF GENETIC DISEASES

Risk for monogenic disease and aneuploidies are the most common reasons that prompt couples to opt for prenatal diagnosis (PD). Unfortunately, current procedures of prenatal diagnosis are invasive and carry a 0.5-1% risk of fetal mortality. The discovery of fetal DNA in maternal plasma had opened new opportunities for non invasive diagnosis and to date, cffDNA( cell-free fetal DNA) is considered the ideal target to conduct a noninvasive diagnosis (NIPD). Simultaneously, the large develop of next generation technologies (NGS) has provided a robust method to detect and analyze cfDNA (cell free DNA) from maternal plasma. Our intent is develop new protocols able to carry forward a noninvasive diagnosis starting from cffDNA. In this work, we propose a protocol that allow fetal genotype detection from ccfDNA through a target amplification of several SNP and mutation sites, through analysis by Ion Torrent PGM platform technology and supported by statistical approaches useful to discriminate fetal DNA contribution into mixture of fetal/maternal DNA (RHDO/SPRT). In order to define haplotypes we propose a long-range PCR method based that can support the detection of the parental haplotypes in association to normal and mutated alleles. Using these approaches, we have analyzed 18 cfDNA samples with these results: 9 samples defined correctly; 7 samples defined at 50%; 1 sample not defined correctly; 1 sample not analyzed. To date, obtained data show that this method is effective and reliable, though will required additional data to confirm these results. Data collected suggest that this method could be inserted, in the near future, in clinical diagnostic practices removing risks of fetal loss, facilitating the diagnostic process and providing significant economic advantage in clinical procedures.