



Università degli Studi di Cagliari

## **DOTTORATO DI RICERCA**

Terapia Pediatrica e Farmacologia dello Sviluppo

Ciclo XXVIII

## **TITOLO TESI**

**SVILUPPO DI UNA PIATTAFORMA PER LA DIAGNOSI PRENATALE  
NON INVASIVA DI MALATTIE GENETICHE IN EPOCA GESTAZIONALE  
PRECOCE**

Settore scientifico disciplinare di afferenza

BIO/11

|                        |                                   |
|------------------------|-----------------------------------|
| Presentata da:         | Valentina Capponi                 |
| Coordinatore Dottorato | Prof. Paolo Moi                   |
| Tutor                  | Prof.ssa Maria Cristina Rosatelli |

Esame finale anno accademico 2014 – 2015

|                                                                    |    |
|--------------------------------------------------------------------|----|
| 1. Introduzione                                                    | 2  |
| 1.1 La diagnosi prenatale                                          | 2  |
| 1.2 Il DNA fetale                                                  | 5  |
| 1.3 Determinazione non invasiva del sesso fetale                   | 14 |
| 1.4. Genotipizzazione non invasiva del fattore RHD                 | 16 |
| 1.6 Diagnosi prenatale non invasiva (NIPD) di malattie monogeniche | 17 |
| 1.5 Non invasive prenatal testing (NIPT) delle aneuploidie         | 26 |
| 2. Obiettivo dello studio                                          | 27 |
| 3. Materiali e metodi                                              | 29 |
| 3.1 Pazienti e raccolta dei campioni                               | 29 |
| 3.2 Disegno sperimentale                                           | 29 |
| 3.3 Amplificazione dei DNA del trio                                | 32 |
| 3.4 Preparazione della libreria di DNA                             | 33 |
| 3.5 Emulsion PCR                                                   | 35 |
| 3.6 Sequenziamento e analisi dei dati                              | 36 |
| 3.7 Estrazione del cffDNA                                          | 38 |
| 3.8 Amplificazione del cffDNA                                      | 39 |
| 3.9 Preparazione della DNA library, emulsion PCR e sequenziamento. | 40 |
| 3.10 Costruzione del pannello di aplotipi di riferimento           | 41 |
| 3.11 Analisi dei dati                                              | 42 |
| 4. Risultati                                                       | 48 |
| 5. Conclusioni                                                     | 64 |
| Bibliografia                                                       | 70 |

# 1. Introduzione

## ***1.1 La diagnosi prenatale***

La diagnostica prenatale può essere definita come un complesso di indagini strumentali e di laboratorio finalizzate alla valutazione e al monitoraggio dello stato di salute embrio-fetale durante la gravidanza.

L'obiettivo della diagnosi prenatale è quello di preparare le coppie a rischio alla eventuale nascita di un figlio affetto, permettere loro di compiere una scelta consapevole sul proseguo della gravidanza ed avviare, nei casi in cui sia possibile, un trattamento farmacologico e/o chirurgico nel feto malato.

I test prenatali possono essere divisi in due categorie: test di screening e test diagnostici.

I test di screening prenatali, sviluppati nel corso degli ultimi 30 anni, hanno lo scopo di individuare le gravidanze a rischio per alcune patologie fetali. Tali test si basano essenzialmente sul dosaggio di alcuni marcatori biochimici presenti nel siero materno e sull'esecuzione di accertamenti ecografici. Un esame ultrasonografico eseguito nel primo trimestre è la misura della translucenza nucale, ovvero dello spessore dello spazio sottocutaneo retronucale, il cui aumento è spesso correlato ad un rischio maggiore di feto affetto da trisomia 21 o altre anomalie genetiche. I marcatori biochimici maggiormente utilizzati nei test di screening sono la proteina plasmatica A associata alla gravidanza (PPAP-A) e la frazione libera della gonadotropina corionica umana (free  $\beta$ -hCG). La PPAP-A è prodotta dal trofoblasto placentare e, generalmente, la sua concentrazione aumenta rapidamente nel sangue materno durante il primo trimestre mentre si riduce sensibilmente nel caso di feti affetti da trisomia 21. La free  $\beta$ -hCG raggiunge un picco di concentrazione tra l'8<sup>a</sup>

e la 10<sup>a</sup> settimana di gravidanza ed è presente in concentrazioni maggiori nel sangue di madri con feto affetto da sindrome di Down. La misura della translucenza nucale, combinata con la quantificazione di questi due marcatori biochimici e con l'età materna, costituiscono nell'insieme il "test combinato", che viene eseguito tra la 11<sup>a</sup> e la 13<sup>a</sup> settimana di gestazione. Questo test di screening non invasivo ha una detection rate di circa il 90% per la predizione della Trisomia 21, con una percentuale di falsi positivi pari a circa il 5% (Wald et al. 2003, Spencer et al. 2007). Ulteriori marcatori ultrasonografici, noti per essere associati con la Trisomia 21, quali l'ipoplasia dell'osso nasale e l'insufficienza tricuspide, possono aumentare l'accuratezza del test combinato del primo trimestre. Attraverso tali test di screening è quindi possibile individuare le donne nelle quali è evidente un aumentato rischio di anomalie cromosomiche fetali e per le quali può essere indicato un test diagnostico prenatale invasivo.

Per le coppie a rischio per le più frequenti malattie monogeniche (talassemia, fibrosi cistica, sindrome dell'X-fragile, etc) non sono disponibili test di screening di questo tipo, ma viene richiesta la diagnosi prenatale invasiva.

I test di screening forniscono esclusivamente una stima della probabilità di avere un figlio affetto da malattia genetica, mentre la diagnosi certa può essere eseguita soltanto attraverso l'analisi del cariotipo fetale o lo studio molecolare dei geni-malattia.

Le Linee Guida per i Test Genetici raccomandano il ricorso alla diagnosi prenatale qualora si verifichi una delle seguenti condizioni:

- presenza di rischio procreativo prevedibile "a priori", ad esempio: età materna avanzata (maggiore o uguale ai 35 anni); genitore portatore di anomalie cromosomiche strutturali; genitori portatori di mutazioni geniche.

- presenza di un rischio fetale reso evidente nel corso della gestazione, ne sono esempi: le malformazioni evidenziate all'ecografia; positività dei test di screening; malattie infettive materne insorte in gravidanza.

I test diagnostici comprendono gli esami genetici eseguiti su campioni biologici prelevati mediante tecniche invasive, quali la villocentesi e l'amniocentesi.

L'amniocentesi consiste nel prelievo ecoguidato di liquido amniotico contenente cellule fetali (amniociti) derivate dallo sfaldamento cutaneo, dall'epitelio del tratto gastroenterico, urogenitale, respiratorio e delle membrane amniotiche; tale procedura è generalmente effettuata attorno alla 15<sup>a</sup>-16<sup>a</sup> settimana di gestazione.

Sugli amniociti possono essere eseguite indagini citogenetiche per la determinazione del cariotipo fetale e in alternativa analisi molecolari finalizzate all'identificazione di mutazioni genetiche responsabili di malattie monogeniche.

Il prelievo dei villi coriali (CVS) viene eseguito intorno alla 11<sup>a</sup>-12<sup>a</sup> settimana e consiste nel prelievo transaddominale del "*chorion fondosum*".

Il tessuto trofoblastico, preventivamente separato dalla decidua materna, può essere utilizzato per l'esecuzione di indagini citogenetiche eseguibili sulle metafasi spontanee del citotrofoblasto oppure sulle metafasi ottenute da cellule in coltura.

In alternativa il DNA fetale estratto dai villi coriali può essere sottoposto a tecniche di biologia molecolare per la diagnosi di malattie monogeniche. L'indicazione per la diagnosi prenatale in questo caso è limitata alle coppie portatrici della mutazione causativa della malattia in esame.

La villocentesi ha il vantaggio di poter essere eseguita più precocemente; tuttavia, a causa di mosaicismi confinati alla placenta, nell'1% dei casi si può osservare una discordanza di risultato tra l'assetto cromosomico stabilito durante le indagini

citogenetiche e il reale cariotipo del feto. In questi casi si rende necessaria la ripetizione del test prenatale attraverso l'amniocentesi.

Le tecniche di prelievo sopra descritte comportano un rischio di perdita fetale che viene stimata intorno allo 0.5-2% (Mujezinovic e Alfrenic 2007), percentuale che è soggetta a variabilità in relazione all'esperienza dell'operatore che esegue il prelievo.

Queste percentuali di rischio per la salute del feto, seppur basse, hanno indotto per decenni i ricercatori a sviluppare dei protocolli di diagnosi prenatale non invasiva.

Per anni diversi gruppi hanno orientato i loro studi sulla ricerca di cellule fetali nucleate (eritroblasti) nel circolo periferico materno. Tuttavia la rarità di queste cellule, la mancanza di anticorpi specifici e la difficoltà nella riproducibilità degli esperimenti ha fortemente limitato l'applicazione di questo approccio.

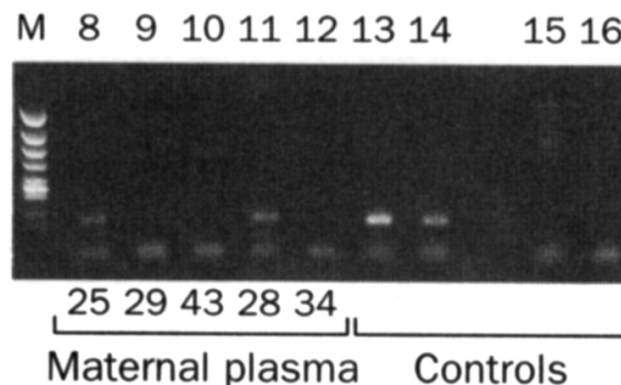
Questi limiti hanno spinto i ricercatori allo studio di fonti alternative di materiale genetico fetale e la scoperta del DNA fetale libero nel plasma materno (Lo et al. 1997) ha permesso, negli ultimi 18 anni, di studiare nuovi approcci utili ad analizzare le caratteristiche genetiche del feto.

## ***1.2 Il DNA fetale***

La presenza nel plasma di acidi nucleici liberi da cellule è stata descritta per la prima volta nel 1948 (Mandel et al. 1948). In seguito alla pubblicazione di diversi studi nei quali veniva descritta la presenza di DNA tumorale libero da cellule nel plasma di pazienti affetti da neoplasie (Chen et al. 1996, Nawroz et al. 1996), nel 1997 il gruppo di YMD Lo valutò l'ipotesi che anche il feto potesse rilasciare DNA libero da cellule nel plasma materno nel corso della gravidanza.

Per confermare tale ipotesi, condusse uno studio preliminare su un campione di 43 donne in epoca gestazionale compresa tra la 12<sup>a</sup> e la 40<sup>a</sup> settimana. Di queste, 30

erano in attesa di un feto di sesso maschile e 13 di un feto di sesso femminile. I campioni di DNA estratto dal plasma e dal siero materno vennero sottoposti a reazione di amplificazione per l'individuazione di sequenze specifiche del cromosoma Y (*DSY14*). Un segnale di amplificazione venne rilevato correttamente in 24/30 campioni di plasma ottenuti da donne in attesa di un feto di sesso maschile. Al contrario non venne osservato nessun segnale di amplificazione nel plasma delle 13 donne in attesa di un feto di sesso femminile (Lo et al. 1997) (Fig.1). I risultati così ottenuti hanno consentito di dimostrare per la prima volta che nel plasma delle donne in gravidanza è presente materiale genetico di origine fetale.



**Figura 1.** Amplificazione di sequenze del cromosoma Y fetali nel plasma materno (Lo et al. 1997)

Studi successivi hanno consentito di definire più in dettaglio le caratteristiche fisiche e biologiche del cffDNA (cell-free fetal DNA).

Tali studi hanno dimostrato che l'apoptosi del sinciziotrofoblasto della placenta, derivante dal turnover fisiologico delle cellule placentari, è il principale meccanismo di rilascio del cffDNA nel plasma materno durante la gravidanza (Gupta et al. 2004).

A supporto di tali evidenze è stato dimostrato che le molecole del cffDNA presentano caratteristiche epigenetiche tipiche del tessuto placentare (Chan et al. 2006, Chim et al. 2005) e la presenza del cffDNA nel plasma materno è stata

osservata anche nelle gravidanze anembrioniche (Alberry et al. 2007). Inoltre in caso di mosaicismo confinato alla placenta, le caratteristiche genetiche del cffDNA sono coerenti con quelle riscontrate a livello placentare (Flori et al. 2004, Masuzaki et al. 2004).

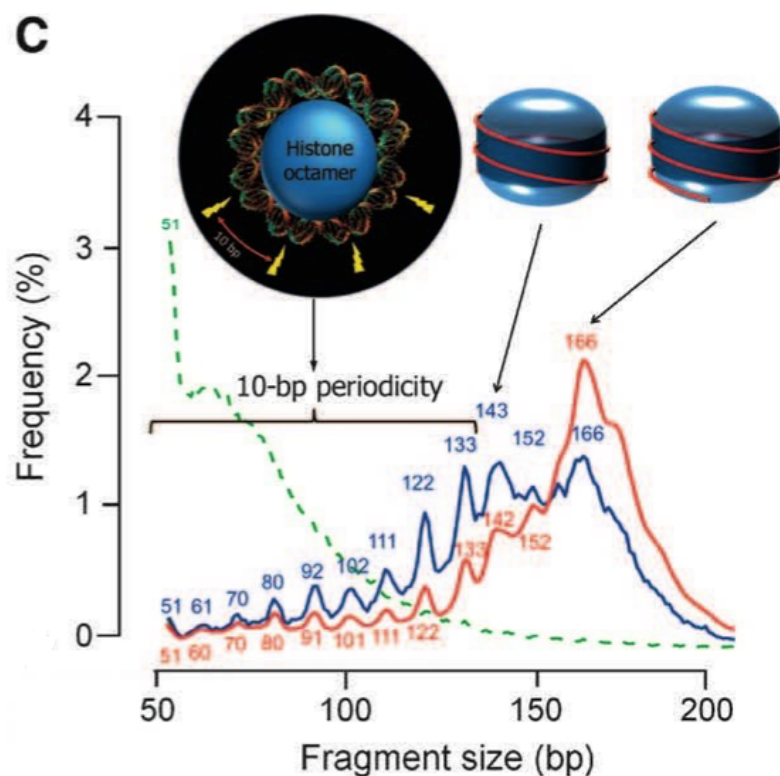
L'apoptosi delle cellule ematopoietiche (Lui et al. 2002) e/o del tessuto adiposo (Haghiac et al. 2012) è al contempo responsabile del rilascio nel plasma di DNA libero materno.

L'origine apoptotica del DNA libero è responsabile anche della sua natura frammentaria. Uno studio eseguito utilizzando due saggi di PCR quantitativa (uno specifico per il gene *leptin*, utilizzato come marcatore del DNA materno, e l'altro per il gene *SRY*, utilizzato come marcatore del DNA fetale) ha dimostrato che le molecole di DNA materno presenti nel plasma di donne in gravidanza hanno una lunghezza maggiore rispetto a quelle individuate nel plasma di donne non in gravidanza. Inoltre è stato osservato che il DNA di origine materna ha dimensioni maggiori rispetto al DNA fetale, la cui lunghezza media è inferiore alle 313 bp (Chan et al. 2004).

L'utilizzo di tecnologie d'analisi più sensibili come il sequenziamento massivo parallelo (MPS) ha permesso di definire più in dettaglio la struttura del cffDNA (Lo et al. 2010). Attraverso l'impiego del MPS, è stato possibile dimostrare per la prima volta che l'intero corredo cromosomico fetale è rappresentato nel cffDNA e infine che le molecole di DNA libero materno e fetale hanno due diversi profili di frammentazione (Fig.2).

Il DNA libero di origine materna è infatti costituito da frammenti di lunghezza media pari a 166 bp, corrispondenti al DNA avvolto attorno al core nucleosomale (circa 146 bp) e da un frammento di DNA linker (20 bp). Il DNA fetale mostra invece una prevalenza di frammenti di 143 bp e di frammenti di dimensioni inferiori che

mostrano una periodicità di 10 bp dovuta al clivaggio del DNA ad opera delle nucleasi. Tali osservazioni hanno ulteriormente confermato che l'apoptosi rappresenta il principale meccanismo fisiologico alla base dell'origine del DNA libero circolante.



**Figura 2.** Distribuzione delle dimensioni del DNA fetale (curva blu), DNA totale (curva rossa) e DNA mitocondriale (curva verde) (Lo et al. 2010).

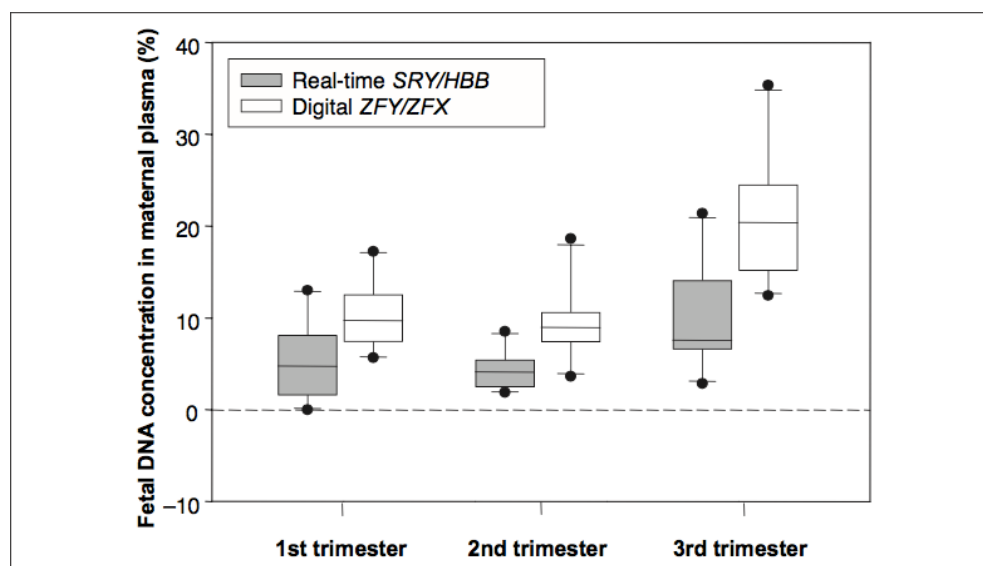
La concentrazione del cffDNA varia in funzione di diversi fattori fisiologici e/o patologici.

Le prime stime sulla concentrazione del cffDNA furono ottenute nel 1998 da Lo et al. nell'ambito di uno studio in cui vennero amplificati tramite real-time PCR il gene *SRY* (*sex determining region Y*), utilizzato come marcatore del DNA fetale, contestualmente ad un frammento del gene  $\beta$ -globinico come marcatore del DNA libero totale. Nei campioni analizzati venne rilevata una concentrazione del DNA

fetale che varia dal 3.4% in epoca gestazionale precoce (11<sup>a</sup>-17<sup>a</sup> settimana di gestazione) al 6.2% in epoca gestazionale tardiva (37<sup>a</sup>-43<sup>a</sup> settimana), evidenziandone il progressivo incremento al progredire della gestazione.

Lo sviluppo di tecnologie più sensibili ha consentito di ottenere una stima più precisa della concentrazione del cfDNA. Infatti in uno studio del 2008 è stata messa a confronto la sensibilità della real-time PCR con quella della digital PCR nella quantificazione del DNA fetale isolato dal plasma di 3 gruppi di donne in attesa di un feto di sesso maschile al primo, secondo e terzo trimestre di gravidanza (Fig.3).

La concentrazione media rilevata con la digital PCR è stata del 9.7%, 9.0% e 20.4% rispettivamente nel primo, secondo e terzo trimestre di gravidanza; i corrispondenti valori ottenuti con la real time PCR sono stati rispettivamente pari al 4.8%, 4.1% e 7.6% (Lun et al. 2008a).



**Figura 3.** Percentuale del DNA fetale maschile misurato con un saggio di Real Time PCR *SRY/HBB* e con un saggio *ZFY/ZFX* eseguito con Digital PCR nel primo, secondo e terzo trimestre nel plasma materno (Lun et al. 2008a).

A differenza della real-time PCR tradizionale, la digital PCR fornisce una quantificazione assoluta basata sulla conta di ogni singola molecola target e senza l'utilizzo di curve standard di riferimento o di controlli endogeni. I risultati ottenuti da questo studio hanno dimostrato che la concentrazione del cffDNA è superiore a quella sino ad allora stimata attraverso l'applicazione di tecniche d'analisi meno sensibili.

Poiché la concentrazione del cffDNA è un criterio importante per l'accuratezza del risultato di un test genetico non invasivo, la sua determinazione risulta essere fondamentale nell'ambito di tale esame diagnostico.

La preponderanza di DNA materno nella miscela DNA materno+DNA fetale e la condivisione di metà del patrimonio genetico tra feto e madre rappresentano i principali ostacoli che si incontrano nell'analisi del cffDNA. La determinazione della frazione fetale si basa quindi sulla ricerca, nel plasma materno, di marcatori specifici del DNA fetale non condivisi con il genoma materno. La mancata rilevazione di tali marcatori può essere indicativa dell'assenza reale del cffDNA o, in alternativa, essere correlata ad un errore metodologico.

L'amplificazione selettiva di sequenze specifiche del cromosoma Y, come il gene multicopia *TSPY1* (*Testis Specific Protein, Y-Linked 1*) e il gene *SRY*, può essere utilizzata per confermare la presenza del cffDNA e definirne la concentrazione, ma solo nelle gravidanze in cui il feto è di sesso maschile.

Un approccio alternativo, indipendente dal sesso, consiste nel ricercare, nel plasma materno, la presenza di polimorfismi a singolo nucleotide (SNPs), di inserzioni/delezioni o di comuni Short Tandem Repeats (STRs) polimorfici che sono ereditati dal feto per via paterna ma assenti nel genoma materno (Pertl et al. 2000, Scheffer

et al. 2010, Page-Christiaens et al. 2006, Tynan et al. 2011, Barrett et al. 2012, Lo et al. 2010).

Diversi studi hanno individuato alcuni marcatori fetali universali, ovvero marcatori indipendenti dal sesso o da polimorfismi ereditati per via paterna, che si basano sulla presenza di differenze epigenetiche tra il DNA materno e quello fetale. Tra questi la metilazione del DNA rappresenta uno degli esempi di marcatori fetali universali meglio caratterizzati. Nel 2002 vennero dimostrate per la prima volta le differenze di metilazione tra il DNA materno e quello fetale in una regione compresa tra i geni *IGF2* (*Insulin-like growth factor 2*) e *H19* (Poon et al. 2002). Un altro esempio di marcatore fetale universale è il promotore del gene *SERPINB5* (*Serpin Peptidase Inhibitor, Clade B (Ovalbumin), Member 5*) che si presenta ipometilato nella placenta e ipermetilato nel DNA materno (Chim et al. 2005). Al contrario il gene oncosoppressore *RASSF1A* (*Ras-Association Domain Family Member 1, transcript variante A*) è ipermetilato nel cffDNA e ipometilato nel DNA materno (Chan et al. 2006, White et al. 2012, Kolialexi et al. 2012, Zejskova et al. 2010). Attraverso l'utilizzo di enzimi di restrizione che riconoscono sequenze ipometilate, è possibile sottoporre a digestione enzimatica soltanto il DNA libero di origine materna amplificando selettivamente il DNA di origine fetale.

Un metodo alternativo recentemente descritto per calcolare la frazione fetale è basato sul rapporto tra i frammenti più corti e i frammenti più lunghi presenti nel plasma materno, considerando che le molecole di DNA di derivazione fetale sono più corte di quelle materne (Yu et al. 2014).

La quantificazione del cffDNA, attraverso l'analisi di marcatori fetali, può essere eseguita con la real-time PCR o in alternativa con tecnologie più sensibili come la digital PCR e il sequenziamento Next Generation Sequencing.

Nella quantificazione della frazione fetale è importante considerare che essa può essere influenzata da diversi fattori fisio-patologici quali l'epoca gestazionale, l'indice di massa corporea della madre, la presenza di aneuploidie fetali, di patologie materne legate alla gravidanza, ed infine di eventuali gravidanze gemellari.

L'epoca gestazionale è il principale fattore fisiologico che influenza la frazione fetale in quanto, come precedentemente citato, essa aumenta con il procedere della gravidanza. Un recente studio eseguito su 22384 donne in differenti epoche gestazionali (10-40 settimane di gestazione) ha evidenziato un iniziale incremento del DNA fetale pari allo 0.1% a settimana tra la 10<sup>a</sup> e la 20<sup>a</sup> settimana di gestazione, seguito da un aumento dell'1% a settimana a partire dalla 21<sup>a</sup> settimana fino al parto (Wang et al. 2013).

Recenti studi hanno dimostrato che il peso e/o l'indice di massa corporea (BMI) materno possono influenzare la concentrazione del cffDNA nel plasma materno. L'obesità, ad esempio, è associata alla presenza di frazioni fetali inferiori (Ashoor et al. 2012, Ashoor et al. 2013, Canick et al. 2013, Wang et al. 2013). Questo fenomeno potrebbe essere correlato all'aumentato volume della circolazione materna che si traduce nella maggiore diluizione del DNA fetale (Canick et al. 2013). In alternativa, la ridotta frazione fetale potrebbe essere causata da un aumento degli eventi apoptotici o di necrosi a carico della frazione stromo-vascolare e del tessuto adiposo in donne obese in gravidanza (Haghiac et al. 2012).

Poiché il rilascio di cffDNA è strettamente legato alla morfogenesi della placenta, le patologie legate alla gravidanza possono influire sulla sua concentrazione nel plasma materno. La preeclampsia è uno degli esempi meglio studiati (Hahn et al. 2011). Infatti lo stress ossidativo ad essa correlato determina un aumento dell'apoptosi dei trofoblasti e del rilascio di microparticelle del sinciziotrofoblasto, che

si traduce in un incremento del cffDNA liberato nel sangue materno (Hahn et al. 2011, Knight et al. 1998). Elevati livelli di cffDNA possono essere rilevati ancora prima dell'insorgenza dei sintomi clinici (Levine et al. 2004). Per tale motivo la quantificazione del cffDNA può rappresentare un valido marcatore biologico per il monitorare l'esordio e la progressione della patologia. Un aumento del DNA fetale circolante può essere indicativo del manifestarsi di altre condizioni patologiche quali, ad esempio, l'aumentato rischio di travaglio pretermine (Farina et al. 2005), o, ancora, di un ritardo di crescita intrauterino (IUGR), più frequente in pazienti con IUGR dovuta ad insufficienza placentare (Al Nakib et al. 2009).

Eventi patologici a carico del feto possono indurre anch'essi un incremento della frazione fetale. Nelle gravidanze con trisomia del cromosoma 21, ad esempio, si osserva un'aumentata frazione fetale che potrebbe essere correlata con l'aumento dell'apoptosi dei trofoblasti; al contrario in caso di trisomia 13, 18 e monosomia del cromosoma X essa risulta diminuita (Rava et al. 2014).

Per quanto concerne le gravidanze gemellari, la frazione fetale risulta aumentata di circa un terzo rispetto alle gravidanze singole (Canick et al. 2012).

Diversi studi hanno dimostrato che la permanenza del cffDNA nel sangue materno è limitata nel tempo poiché ha un'emivita ridotta. Nel 1999 in seguito ad esperimenti eseguiti con la real-time PCR, utilizzando il gene *SRY* come marcatore fetale, e prendendo in esame 12 donne in tempi diversi dopo il parto, Lo et al. dimostrò che il cffDNA ha un'emivita di 16 minuti e viene rapidamente rimosso dopo il parto.

Uno studio recente eseguito mediante la tecnica del MPS ha chiarito che la clearance del cffDNA dopo il parto prevede due fasi, con differenti cinetiche: una fase iniziale rapida in cui la concentrazione totale viene dimezzata nell'arco di un'ora, seguita da una fase più lenta in cui il cffDNA viene totalmente rimosso entro

13 ore dal parto (Yu et al. 2013). La rapida rimozione del cffDNA dopo il parto fa sì che il cffDNA non persista nel plasma materno dopo il termine della gestazione e di conseguenza non influenzi l'esito di test genetici in gravidanze successive.

L'evoluzione delle tecnologie impiegate per l'esecuzione dei test molecolari ha consentito di approfondire le conoscenze sulle caratteristiche biologiche del cffDNA, permettendo lo sviluppo di test diagnostici non invasivi alcuni dei quali sono già stati trasferiti nella pratica clinica di diversi Stati sia all'interno di strutture pubbliche che private.

### ***1.3 Determinazione non invasiva del sesso fetale***

La prima applicazione clinica basata sull'analisi del cffDNA è stata la determinazione non invasiva del sesso fetale.

Questo test può essere effettuato a partire dalla 7a settimana di gestazione attraverso real-time PCR (RT-qPCR) per identificare la presenza o assenza di sequenze specifiche del cromosoma Y (*SRY*, *DYS14*) nel DNA circolante nel plasma materno (Devaney et al. 2011). Questo esame può essere utile nell'ambito delle gravidanze a rischio per patologie X-linked (Distrofia Muscolare di Duchenne (DMD), Adrenoleucodistrofia, etc.) poiché consente di selezionare le gravidanze che dovranno essere indirizzate alla diagnosi prenatale invasiva, evitando questo inutile percorso diagnostico alle donne in attesa di un feto di sesso femminile. Questo test non invasivo è stato inserito da diversi anni nella pratica clinica di numerosi Paesi europei, come il Regno Unito (Chitty et al. 2012, Hill et al. 2011, ), consentendo di ridurre del 50% il numero delle diagnosi prenatali invasive eseguite per patologie X-linked (Hill et al. 2011).

La determinazione non invasiva del sesso fetale assume particolare rilevanza anche nella gestione delle gravidanze a rischio per iperplasia surrenale congenita (CAH), una malattia autosomica recessiva causata da un deficit dell'enzima 21-idrossilasi e che determina uno sviluppo ambiguo dei genitali esterni femminili.

La somministrazione di desametasone a partire dalla nona settimana di gestazione impedisce la virilizzazione nei feti di sesso femminile. La determinazione del sesso fetale attraverso diagnosi prenatale non invasiva può consentire di interrompere l'inutile trattamento farmacologico nei feti di sesso maschile. Al contrario nelle donne in attesa di un feto di sesso femminile è necessario proseguire la terapia con desametasone sino all'accertamento del genotipo CAH fetale eseguito con esame invasivo.

Una meta-analisi (Devaney et al. 2011) ha raccolto i dati ottenuti da 57 studi pubblicati dalla fine degli anni '90 finalizzati alla determinazione non invasiva del sesso fetale in 3524 gravidanze con feto di sesso maschile e 3017 con feto di sesso femminile. Complessivamente il test è risultato affidabile quando eseguito tra la 7<sup>a</sup> e la 12<sup>a</sup> settimana di gravidanza, con una sensibilità del 94.8% e una specificità del 98.9%, percentuali che aumentano rispettivamente al 99% e 99.6% quando il test viene effettuato dopo la 20<sup>a</sup> settimana di gestazione (Devaney et al. 2011).

La presenza di falsi negativi può essere attribuita alla mancata rilevazione di sequenze del cromosoma Y che può essere causata dalla ridotta concentrazione di cffDNA nel campione esaminato o imputabile ad un errore metodologico compiuto durante l'esecuzione del test. Questo rischio può essere annullato confermando la presenza del cffDNA attraverso l'utilizzo di marcatori fetali indipendenti dal sesso, come per esempio le sequenze diversamente metilate.

Un risultato falso positivo può invece essere dovuto ad un evento fisiologico noto come “*vanishing twin*” (Masala et al. 2015, Vlková et al. 2014). Questo fenomeno è il risultato del riassorbimento di uno degli embrioni presenti nell’utero nel corso di una gravidanza gemellare. E’ stato infatti dimostrato che l’embrione non più vitale continua a rilasciare cfDNA nel sangue materno anche per 9 settimane dopo la sua morte intrauterina. Per questo motivo prima di eseguire un test non invasivo è necessario escludere la presenza di un *vanishing twin* mediante esami ultrasonografici in epoca gestazionale precoce.

#### **1.4. Genotipizzazione non invasiva del fattore RHD**

Un altro test non invasivo, disponibile già da tempo nella pratica clinica di numerosi paesi Europei e non, è la genotipizzazione del fattore RhD fetale in donne RhD negative con partner RhD positivo.

Gli antigeni del gruppo sanguigno Rh sono codificati da due geni posizionati sul cromosoma 1, *RHD* e *RHCE*, che codificano per due polipeptidi di 416 amminoacidi, noti come RhD e RhCE (Daniels et al. 2004).

Nella popolazione caucasica il fenotipo RhD negativo è comunemente causato dalla delezione del gene *RHD* (Colin et al. 1991).

L’incompatibilità RhD materno-fetale si verifica quando una donna RhD negativa, immunizzata in seguito ad una precedente gravidanza RHD positiva, affronta una seconda gravidanza con feto RhD positivo. Infatti gli anticorpi anti-RhD presenti nel sangue materno attraversano la placenta e causano l’emolisi dei globuli rossi fetali, fenomeno noto come malattia emolitica feto-neonatale (MEFN).

La genotipizzazione prenatale non invasiva del fattore RhD viene eseguita amplificando, tramite real-time PCR, sequenze specifiche del gene *RHD* ereditate

dal feto per via paterna. La performance del test non invasivo è stata valutata attraverso diversi trial clinici su larga scala, che hanno dimostrato un'elevata accuratezza diagnostica, con valori di sensibilità pari al 99.5-99.8 % e di specificità del 94.0-99.5% (Legler et al. 2009, Chitty et al. 2014, Wikman et al. 2012).

La diffusione del test non invasivo ha consentito di razionalizzare la somministrazione delle immunoglobuline anti-D alle sole donne in attesa di un feto RhD positivo (Finning et al. 2004, Ferres et al. 2014, Clausen et al. 2014).

### ***1.6 Diagnosi prenatale non invasiva (NIPD) di malattie monogeniche***

Nell'ambito delle malattie monogeniche, il test non invasivo ha trovato applicazione soprattutto nella diagnosi delle patologie autosomiche dominanti ereditate per via paterna, come la malattia di Huntington (Gonzalez-Gonzalez et al. 2003a, Gonzalez-Gonzalez et al. 2003b, Bustamante-Aragones et al. 2008c), la distrofia miotonica (Amicucci et al. 2000), la distonia primaria ad esordio precoce (Meaney et al. 2009) e la craniosinostosi (Galbiati et al. 2014) o causate da mutazioni *de novo* (acondrodisplasia). In questi casi il test molecolare è finalizzato alla ricerca di tratti genetici fetali che sono assenti nel genoma materno.

La presenza/assenza dell'allele mutato nel plasma materno può essere associata nelle malattie autosomiche dominanti rispettivamente con lo stato affetto/non affetto del feto. Questi test genetici sono raccomandati non solo alle coppie per le quali è stata accertata una storia familiare, ma anche nei casi in cui siano state riscontrate anomalie morfologiche tramite ultrasonografia, come ad esempio nel caso dell'acondroplasia.

Questa patologia è causata nella maggior parte dei casi da mutazioni *de novo* a carico del gene *FGFR3* (*Fibroblast growth factor receptor 3*). Diversi studi hanno dimostrato la fattibilità del test non invasivo per la diagnosi prenatale di acondrodisplasia (Lim et al. 2011, Saito et al. 2000, Chitty et al. 2011).

Nel Regno Unito l'identificazione nel cffDNA delle mutazioni a carico dei geni *FGFR3* e *FGFR2* (*Fibroblast growth factor receptor 2*) è diventata parte integrante del percorso clinico per le gravidanze a rischio per acondroplasia, displasia tanatofora e Sindrome di Apert, determinando una notevole riduzione del numero dei test invasivi (Everett & Chitty 2015).

La diagnosi prenatale non invasiva delle malattie autosomiche recessive o dominanti ereditate dal feto per via materna si presenta più complessa a causa della difficoltà di differenziare, nel plasma materno, il DNA fetale da quello di origine materna.

Nei vari studi riportati in letteratura l'approccio maggiormente utilizzato si basa sull'accertamento della presenza/assenza della mutazione paterna nelle malattie autosomiche recessive in cui i genitori sono portatori di differenti mutazioni. Nel caso in cui venga identificata la presenza dell'allele mutato ereditato per via paterna, il feto avrà un rischio del 50% di essere affetto dalla patologia in esame e, in tal caso, sarà raccomandato il ricorso al test diagnostico invasivo. In alternativa, l'assenza dell'allele mutato paterno sarà indicativa di un feto con genotipo normale o eterozigote, e pertanto non sarà suggerita la diagnosi prenatale invasiva. Uno dei primi esempi di applicazione di questo approccio è stato condotto su una coppia a rischio per iperplasia surrenale congenita (CAH) in cui l'ereditarietà nel feto dell'allele paterno mutato è stata valutata utilizzando un metodo indiretto basato sull'analisi di marcatori polimorfi (Chiu et al. 2002).

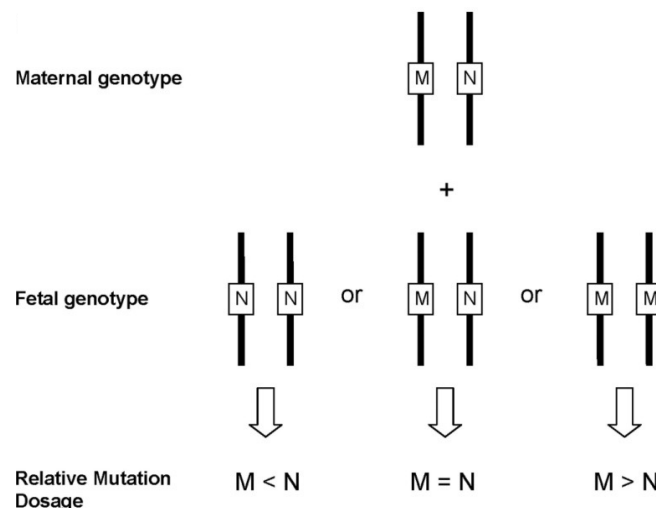
Altri studi sono stati eseguiti utilizzando tecnologie differenti per la diagnosi prenatale non invasiva di altre in malattie autosomiche recessive come la fibrosi cistica (Gonzalez-Gonzalez et al. 2002, Bustamante-Aragones et al. 2008a, Hill et al. 2015), l'acidemia propionica (Bustamante-Aragones et al. 2008b) e la  $\beta$ -talassemia (Galbiati et al. 2011).

Nel caso delle malattie autosomiche recessive, nelle quali invece i genitori sono portatori della stessa mutazione, vi è una maggiore difficoltà ad effettuare correttamente una diagnosi molecolare a causa della preponderante percentuale di DNA materno che maschera gli alleli ereditati dal feto per via materna. Questo limite metodologico ha richiesto non soltanto lo sviluppo di approcci diagnostici non tradizionali, ma anche l'impiego di tecnologie più sensibili e di algoritmi bioinformatici dedicati. Lun et al. nel 2008 hanno descritto per primi un approccio di tipo quantitativo, basato sull'utilizzo della Digital PCR, una tecnologia che consente, utilizzando saggi di discriminazione allelica, di effettuare una quantificazione assoluta degli alleli normali e/o mutati presenti nel campione di cfDNA isolato dal plasma materno a livello del sito di mutazione. Successivamente viene valutata la presenza di una maggiore rappresentazione dell'allele normale o mutato che sarà indicativa rispettivamente di un feto sano o affetto.

Tale approccio, noto come Relative Mutation Dosage (RMD), consente di stabilire il genotipo fetale attraverso la valutazione dell'eventuale sbilanciamento allelico osservato nel sito della mutazione patogenetica. Il contributo degli alleli di origine fetale andrà infatti a modificare il rapporto 50:50 osservato tra gli alleli normale e mutato in una donna con genotipo eterozigote. In assenza di sbilanciamento di tale rapporto il genotipo fetale atteso sarà identico a quello materno (eterozigote). Al contrario nel caso in cui vi sia una sotto- o sovra-rappresentazione dell'allele mutato

rispetto a quello wild-type si potrà dedurre rispettivamente un genotipo fetale omozigote normale e mutato (Fig.4). Questo approccio metodologico è stato utilizzato per la diagnosi non invasiva di  $\beta$ -talassemia e di emoglobina E in 10 coppie a rischio: il genotipo fetale è stato correttamente determinato nella metà dei casi esaminati (Lun et al. 2008b).

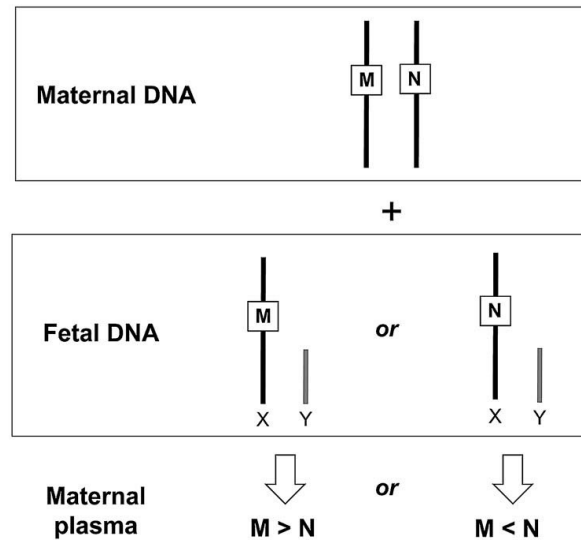
Lo stesso approccio è stato utilizzato anche per la diagnosi prenatale di anemia falciforme (Barrett et al. 2012), consentendo di determinare il genotipo corretto nell'80% (52/65) dei casi.



**Figura 4.** Illustrazione schematica del principio del Relative Mutation Dosage (Lun et al. 2008b). N, allele wild type; M, allele mutato.

Studi successivi hanno descritto l'utilizzo della digital PCR e dell'RMD per la diagnosi prenatale di una malattia X-linked, l'emofilia, in sette donne in attesa di un feto di sesso maschile. In queste patologie, poiché il contributo fetale deriva esclusivamente dal cromosoma X ereditato per via materna, si può osservare soltanto una sovra-rappresentazione dell'allele normale o in alternativa di quello

mutato indice, rispettivamente, di un feto sano o affetto (Fig.5). In questo studio il genotipo è stato predetto correttamente in tutti i casi esaminati.



**Figura 5.** Rappresentazione schematica dell'RMD per le malattie X-linked (Tsui et al. 2011).

La rapida diffusione delle tecnologie per il sequenziamento massivo parallelo (MPS) ha dato ulteriore impulso allo sviluppo di nuovi protocolli per la NIPD di malattie monogeniche. Queste tecnologie consentono infatti di sequenziare diverse gigabasi di DNA in tempi ridotti e di ottenere informazioni anche di tipo quantitativo non rilevabili con il tradizionale metodo di Sanger. Inoltre, diversamente dalla digital PCR, che limita l'indagine al solo sito di mutazione, l'MPS consente di estendere l'analisi molecolare a polimorfismi a singolo nucleotide (SNPs) presenti nella regione genomica in cui è presente il gene target. Nel 2010 è stato dimostrato da Lo et al. per la prima volta l'applicabilità della tecnologia MPS per il sequenziamento dell'intero genoma fetale e contestualmente per la diagnosi prenatale non invasiva di  $\beta$ -talassemia. In questo caso le mutazioni parentali ricercate sul campione di DNA

circolante isolato dal plasma materno erano: la delezione di origine paterna di 4 paia di basi -CTTT dei codoni 41/42 e la sostituzione A→G del nucleotide -28 di origine materna. Attraverso lo studio di SNPs informativi presenti nella regione circostante il gene *HBB* e l'estensione dell'approccio RMD a ciascuno di questi siti polimorfi, è stato possibile identificare gli aplotipi parentali ereditati dal feto. Questo nuovo metodo d'analisi è stato denominato RHDO (Relative Haplotype Dosage) poiché si basa sulla valutazione dello sbilanciamento osservato in blocchi di SNPs vicini che consente di dedurre gli aplotipi parentali. Questo approccio prevede la conoscenza a priori degli aplotipi parentali, che in questo studio sono stati ricostruiti attraverso le informazioni ottenute dal DNA estratto dai villi coriali.

Gli SNPs informativi nei genitori e nel feto, precedentemente caratterizzati mediante piattaforma Affimetrix, sono stati suddivisi in cinque categorie (Fig.6).

| 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<br>2. Fractional concentration of fetal DNA<br>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.<br>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.<br>2. Homozygous if haplotype dosage imbalance is detected                |                                                                                                                   |
| 5.           | A/C                           | A/C               | Not analyzed in this study                           |                                                                                                                                |                                                                                                                   |

**Figura 6:** Suddivisione degli SNPs in 5 categorie (Lo et al. 2010).

Gli SNPs per i quali i genitori sono omozigoti per alleli identici (categoria 2) sono stati utilizzati per valutare il controllo della qualità ed il tasso di errore nel sequenziamento.

Gli SNPs per i quali i genitori sono entrambi omozigoti, ma per alleli differenti e quelli per i quali il padre è eterozigote e la madre omozigote (rispettivamente categoria 1 e 3) sono stati utilizzati per calcolare la frazione fetale attraverso l'applicazione della seguente formula:

$$f = \frac{2p}{p+q}$$

dove  $p$  è il numero di reads (frammenti di DNA) sequenziate dell'allele specifico del feto e  $q$  rappresenta le conte delle reads dell'altro allele, che è condiviso dai genomi materno e fetale.

Per valutare l'ereditarietà dell'aplotipo paterno sono stati utilizzati gli SNPs in cui la madre è omozigote e il padre eterozigote (categoria 3) ovvero rilevando la presenza di alleli che non sono presenti nel genoma materno. Al contrario dedurre l'ereditarietà dell'aplotipo materno risulta più complicato, sia a causa della condivisione di alleli tra la madre e il feto, sia perché il DNA di origine fetale è presente in bassa percentuale nel plasma materno. Per questo scopo sono stati utilizzati gli SNPs in cui il padre è omozigote e la madre eterozigote (categoria 4).

Gli SNPs per i quali i genitori sono entrambi eterozigoti (categoria 5) non sono stati presi in esame in questo studio.

La significatività statistica degli sbilanciamenti osservati è stata determinata attraverso il sequential probability ratio test (SPRT).

In questo primo esempio di NIPD di  $\beta$ -talassemia, il sequenziamento del cffDNA ha consentito di dedurre lo stato di portatore sano del feto, poiché è stata identificata la presenza della mutazione paterna e, attraverso l'RHDO, è stato possibile stabilire l'ereditarietà nel feto dell'aplotipo materno associato all'allele  $\beta$ -globinico normale (Lo et al. 2010).

La predittività di questo approccio è stata ulteriormente implementata attraverso il "target sequencing" ovvero circoscrivendo il sequenziamento ad una regione limitata del genoma contenente il gene target. Questo approccio è stato applicato con successo alla NIPD di  $\beta$ -talassemia nel 2012 da Lam et al. che ha eseguito il target-sequencing del cffDNA su una regione di 288-kb del cluster  $\beta$ -globinico, ottenendo una maggiore copertura della regione sequenziata (coverage). L'analisi bioinformatica è stata eseguita utilizzando un algoritmo che prendeva in considerazione una nuova categoria di SNPs (padre e madre eterozigoti). Inoltre la deduzione degli aplotipi parentali è stata ottenuta con un approccio molecolare (digital PCR) e senza l'ausilio del DNA estratto da villi coriali. Nello studio sono state analizzate due famiglie a rischio per  $\beta$ -talassemia in cui i genitori erano portatori di mutazioni differenti. Nella prima famiglia il padre era eterozigote per la delezione -CTTT dei codoni 41/42 del gene *HBB* mentre la madre era eterozigote per la sostituzione nucleotidica A→G del nucleotide -28. La seconda famiglia invece era composta da un padre eterozigote per la mutazione A→T nel codone 17 del gene *HBB*, e una madre eterozigote per la delezione -CTTT. In entrambe le famiglie la diagnosi prenatale invasiva ha evidenziato un feto portatore sano di  $\beta$ -talassemia. Questo risultato è stato confermato in entrambi i casi dalla NIPD.

Lo studio degli aplotipi si è rivelato uno strumento fondamentale per la determinazione non invasiva del genotipo fetale in 15 coppie a rischio per CAH in cui

la presenza di uno pseudogene adiacente al gene-malattia rende particolarmente difficile l'indagine molecolare circoscritta al solo sito di mutazione. Questi 15 casi fanno riferimento a due lavori recentemente pubblicati. In uno di essi il target-sequencing di una regione di circa 6 Mb, contenente il gene *CYP21A2*, e l'applicazione dell'RHDO/SPRT hanno consentito di dedurre correttamente entrambi gli aplotipi parentali nei 14 campioni di cffDNA esaminati. In tutte le famiglie la ricostruzione degli aplotipi parentali è stata eseguita attraverso la genotipizzazione del DNA di un precedente figlio affetto (New et al. 2014).

Nel secondo studio è stata analizzata una sola famiglia con un probando che, anche in questo caso, ha consentito la ricostruzione degli aplotipi parentali. L'arricchimento delle regioni target è stato esteso all'intero esoma ed a regioni sparse nel genoma contenenti un milione di SNPs potenzialmente informativi sia per la ricostruzione degli aplotipi ereditati dal feto che per la determinazione della frazione fetale. Gli SNPs considerati informativi nei genitori sono stati suddivisi in tre categorie: la categoria 1 comprende gli SNPs per i quali la madre è eterozigote e il padre omozigote, la categoria 2 si riferisce agli SNPs per i quali il padre è eterozigote e la madre omozigote, ed infine la categoria 3 è costituita da SNPs per i quali i genitori sono entrambi eterozigoti. La frazione fetale è stata calcolata analizzando gli SNPs in cui i genitori sono omozigoti per alleli diversi. Basandosi sugli aplotipi parentali e sui dati del sequenziamento del cffDNA, gli aplotipi fetali sono stati ricostruiti utilizzando 12577 SNPs informativi sul cromosoma 6 sfruttando modello statistico chiamato "hidden Markov model" (HMM), che consente di stabilire gli aplotipi parentali ereditati dal feto con maggiore probabilità tra le quattro possibili combinazioni di aplotipi. L'accuratezza degli alleli materni e degli alleli paterni inferiti sul cromosoma 6 è stata rispettivamente del 95,01% e 98,42%. Considerando

invece l'intero genoma, sono stati analizzati un totale di 179431 SNPs nelle 22 coppie di autosomi, con un'accuratezza del 96,41% e del 97,81% per l'inferenza rispettivamente degli alleli materni e degli alleli paterni. Attraverso questo metodo, è stato possibile stabilire che il feto ha ereditato l'aplotipo paterno associato all'allele mutato e l'aplotipo materno associato all'allele normale, predicendo lo stato di portatore; il risultato è stato confermato dall'amniocentesi (Ma et al. 2014). Un approccio simile, basato sullo studio degli aplotipi e sul modello statistico HMM, è stato descritto per la diagnosi prenatale non invasiva della sordità congenita (Meng et al. 2014) e di una patologia X-linked, la distrofia muscolare di Duchenne (Xu et al. 2015).

### **1.5 Non invasive prenatal testing (NIPT) delle aneuploidie**

L'analisi del cfDNA per lo screening delle più comuni aneuploidie fetali costituisce una delle applicazioni di maggior successo ed è stato oggetto di ricerca da parte di diversi gruppi che, negli ultimi anni, hanno contribuito allo sviluppo di test prenatali non invasivi (NIPT).

Il principio generale di questi test si basa sulla comparazione tra il numero dei frammenti del cromosoma coinvolto nell'aneuploidia con il numero dei frammenti di un altro cromosoma atteso in una condizione di disomia.

I primi studi volti ad eseguire una diagnosi prenatale non invasiva di Trisomia 21 risalgono al 2007, quando è stato descritto un approccio che permette di rilevare nel plasma materno un incremento dei frammenti di DNA appartenenti al cromosoma 21 utilizzando la tecnica della digital PCR (Lo et al. 2007, Fan et al. 2007).

Con l'avvento della Next Generation Sequencing (NGS) sono stati sviluppati diversi approcci per la NIPT: il whole-genome NGS, che prevede il sequenziamento

dell'intero genoma; il targeted NGS, che consiste nell'amplificazione selettiva di specifiche regioni sul cromosoma di interesse; infine l'approccio basato sull'amplificazione di un elevato alto numero di SNPs sul cromosoma in esame. Ciascuno di questi approcci descritti richiede l'ausilio di algoritmi bioinformatici dedicati per l'analisi dei dati e mostra elevati valori di sensibilità e di specificità nella detection delle più comuni aneuploidie (Boon et al. 2013).

Oggi sono disponibili in commercio molti test NIPT che utilizzano diverse piattaforme NGS e differenti algoritmi bioinformatici; ne sono esempi: MaterniT21 PLUS (Sequenom Laboratories), Verifi prenatal test (Verinata Health, oggi Illumina), NIFTY Test (BGI), IONA Test (Premaitha Health), Harmony prenatal (Ariosa Diagnostics, oggi Roche), Panorama (Natera inc.).

Questi test prevedono lo screening delle più comuni aneuploidie autosomiche (Trisomia 13, 18 e 21) e alcuni di essi anche delle anomalie nel numero dei cromosomi sessuali. È importante sottolineare che NIPT, nonostante la sua elevata sensibilità, è considerato un test di screening, piuttosto che un esame diagnostico; pertanto è raccomandata la conferma di ogni risultato positivo attraverso le procedure di diagnosi invasive.

## **2. Obiettivo dello studio**

Questo studio è stato svolto nell'ambito di un progetto di ricerca più ampio, approvato dal Comitato Etico, condotto da diversi anni presso il laboratorio di Genetica Molecolare dell'Ospedale Pediatrico Microcitemico di Cagliari e finalizzato allo sviluppo di protocolli per la diagnosi prenatale non invasiva di  $\beta$ -talassemia.

In particolare, in questo studio, è stato utilizzato un approccio basato sul target sequencing del DNA libero fetale isolato dal plasma materno mediante piattaforma NGS Ion Torrent PGM. Parallelamente, è stata sviluppata una pipeline per la gestione e l'analisi dei dati di sequenziamento che ha consentito di ottenere la predizione del genotipo fetale in tempi rapidi e con un workflow totalmente automatizzato.

A causa della elevata omogeneità genetica della popolazione sarda, sono state prese in esame 37 coppie a rischio per beta talassemia in cui entrambi i genitori erano eterozigoti per la mutazione  $\beta^{39}$ , che, in Sardegna, rappresenta il 95,7% degli alleli patologici (Rosatelli et al. 1992).

La determinazione del genotipo fetale è stata eseguita indirettamente attraverso l'inferenza degli aplotipi parentali più probabilmente ereditati dal feto, in una regione che si estende per circa 62 Kb nel cluster dei geni beta globinici. Poiché questo approccio implica la conoscenza a priori degli aplotipi parentali, il target sequencing del cfDNA è stato preceduto dal sequenziamento, nella stessa regione, del DNA del corrispondente Trio (padre, madre, trofoblasto).

I dati del sequenziamento sono stati analizzati attraverso l'utilizzo di una pipeline dedicata che ha consentito, da un lato, la ricostruzione degli aplotipi parentali e, dall'altro, l'inferenza dei due aplotipi ereditati con maggiore probabilità dal feto.

I risultati così ottenuti sono stati validati attraverso il confronto con l'esito della diagnosi prenatale eseguita con approccio invasivo.

### **3. Materiali e metodi**

#### ***3.1 Pazienti e raccolta dei campioni***

Lo studio è stato condotto in 37 coppie a rischio per  $\beta$ -talassemia eterozigoti per la mutazione  $\beta^{\circ}39$ , pervenute presso il Servizio di Screening e Consulenza Genetica dell' Ospedale Microcitemico per sottoporsi a diagnosi prenatale tramite prelievo dei villi coriali.

A ciascuna coppia sono stati illustrati gli obiettivi del progetto e, dopo firma di un consenso informato approvato dal Comitato Etico, ogni donna è stata sottoposta ad un prelievo di sangue periferico di circa 20 ml. Le pazienti sono state reclutate, prima della villocentesi, in un'epoca gestazionale compresa tra la 7<sup>a</sup> e la 14<sup>a</sup>+3 settimana di gravidanza.

#### ***3.2 Disegno sperimentale***

Il principio su cui si basa il protocollo è quello di dedurre, nel cffDNA, il genotipo fetale determinando gli aplotipi parentali ereditati con maggiore probabilità dal feto ed associati all'allele normale e/o mutato della variante  $\beta^{\circ}39$ . Per questo motivo, lo studio del cffDNA è stato preceduto dal target sequencing del DNA genomico dei genitori e del DNA fetale estratto dai villi coriali (Trio).

Una regione del cluster  $\beta$ -globinico che si estende per circa 62,7 kilobasi e che contiene, oltre il sito di mutazione  $\beta^{\circ}39$ , circa 1232 SNPs, è stata inizialmente selezionata nel Trio attraverso cinque reazioni di long range PCR. Il sequenziamento di questa regione ha consentito di identificare e selezionare, nel Trio, gli SNPs ritenuti informativi, ossia quelli per i quali i genitori sono omozigoti per alleli differenti e quelli in cui almeno uno di essi ha genotipo eterozigote.

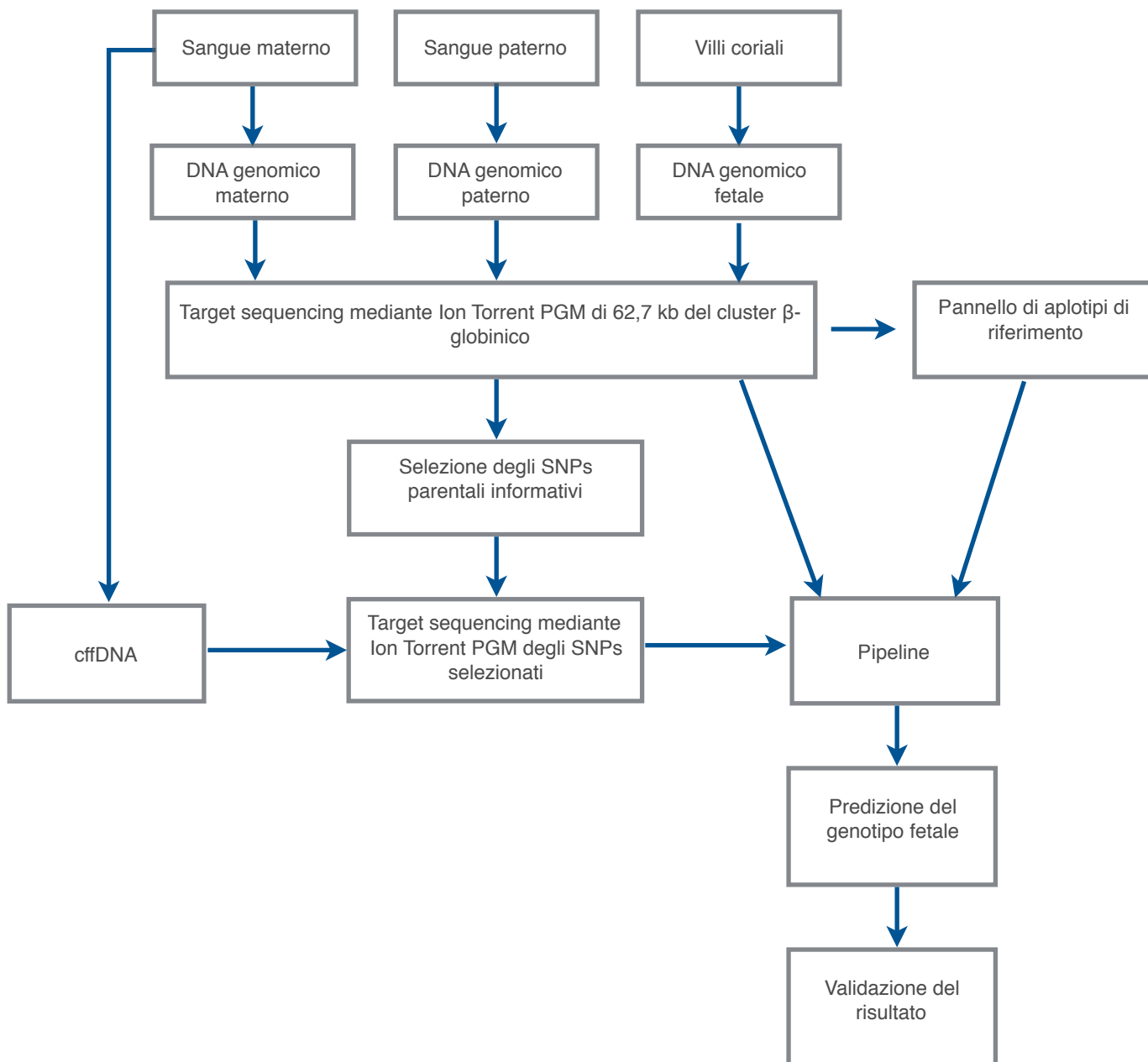
Tali SNPs sono stati successivamente sequenziati nel cffDNA mediante l'amplificazione di frammenti lunghi 100-150 bp, gli unici compatibili con la natura frammentaria del DNA libero isolato dal plasma.

Il processamento dei dati ottenuti dal sequenziamento avviene con modalità rapida e totalmente automatizzata grazie all'impiego di una pipeline dedicata che si basa su una iniziale ricostruzione degli aplotipi parentali e una successiva inferenza del genotipo  $\beta$ -globinico fetale.

La costruzione degli aplotipi parentali è stata eseguita considerando i genitori indipendentemente, senza le informazioni provenienti dal DNA estratto dal trofoblasto, e utilizzando un pannello di aplotipi di riferimento, precedentemente realizzato implementando gli aplotipi del progetto 1000 genomi, con quelli di 39 Trio sardi sequenziati con la piattaforma Ion Torrent PGM. La presenza di aplotipi di individui sardi, all'interno del pannello di riferimento, ha avuto l'obiettivo di incrementare la rappresentazione della mutazione  $\beta^{s39}$ , considerata una variante rara nelle popolazioni incluse nel progetto 1000 genomi.

La ricostruzione nel cffDNA degli aplotipi ereditati dal feto viene eseguita valutando, per l'aplotipo paterno, la presenza/assenza degli alleli non condivisi con il genoma materno. La determinazione dell'aplotipo materno ereditato si basa, invece, sulla quantificazione dell'eventuale sbilanciamento allelico osservato nei siti con genotipo eterozigote. Il genotipo fetale così stabilito viene in ultimo comparato con quello determinato attraverso la diagnosi prenatale invasiva.

Maggiori dettagli sul protocollo sopra descritto sono illustrati nella figura 7.



**Figura 7.** Rappresentazione schematica del workflow del protocollo di NIPD di β-talassemia

### 3.3 Amplificazione dei DNA del trio

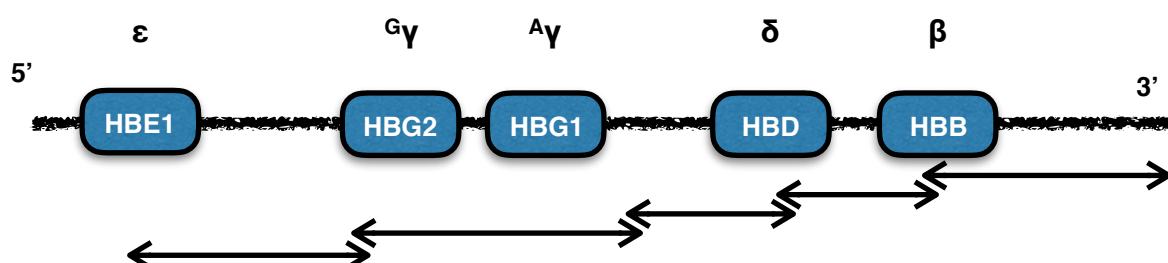
I DNA parentali sono stati estratti da 500 µl di sangue periferico attraverso l'utilizzo del kit di estrazione DiaSorin Blood DNA extraction e il sistema robotizzato NorDiag Arrow. I corrispondenti campioni di DNA fetale sono stati estratti dalle cellule del trofoblasto mediante la tecnica del *salting out*.

I DNA dei trio sono stati analizzati per una regione di 62,7 kilobasi del cluster  $\beta$ -globinico (5209078-5271956) mediante l'amplificazione di cinque lunghi frammenti parzialmente sovrapposti (tab.1.e fig.8) utilizzando due differenti kit:

- Platinum® PCR SuperMix High Fidelity (Thermo Fisher Scientific);
- SequalPrep™ Long PCR Kit with dNTPs (Thermo Fisher Scientific).

|             | coordinate      | dimensioni (bp) |
|-------------|-----------------|-----------------|
| amplicone 1 | 5209078–5226842 | 17765           |
| amplicone 2 | 5226667–5233471 | 6805            |
| amplicone 3 | 5232619–5243798 | 11180           |
| amplicone 4 | 5243662–5257959 | 14298           |
| amplicone 5 | 5255544–5271956 | 16413           |

**Tabella 1.** Coordinate e dimensioni degli ampliconi prodotti mediante Long Range PCR.



**Figura 8.** Rappresentazione del cluster  $\beta$ -globinico. Le frecce sottostanti rappresentano i 5 frammenti amplificati.

Ciascuna reazione di PCR è stata verificata attraverso l'elettroforesi su gel d'agarosio all'1%.

Per ciascun campione, i 5 lunghi ampliconi sono stati riuniti in un unico "pool", sottoposto prima a purificazione mediante biglie magnetiche Agencourt AMPure XP (Beckman Coulter) e successivamente a quantificazione con il kit Qubit dsDNA BR (broad range) Assay Kit e lo strumento Qubit® 2.0 Fluorometer (Thermo Fisher).

### **3.4 Preparazione della libreria di DNA**

Le librerie di DNA sono state prodotte utilizzando il pool dei prodotti di PCR ed il kit "Ion Plus Fragment Library kit" (Thermo scientific), secondo il protocollo previsto dalla piattaforma di sequenziamento Ion Torrent PGM.

Il pool di ampliconi è stato inizialmente sottoposto a frammentazione enzimatica con l'obiettivo di ottenere frammenti lunghi 200-300 paia di basi con estremità "piatte".

La reazione di frammentazione è stata eseguita a 37°C per 15 minuti utilizzando i seguenti componenti:

- 100 ng Long Amplicons, e acqua nuclease-free per un volume finale di 35 µl;
- Ion Shear Plus 10x reaction buffer, 5 µl;
- Ion Shear Plus Enzyme Mix II, 10 µl.

Il prodotto della digestione enzimatica è stato quindi purificato con il sistema Agencourt AMPure XP ed eluito in 25 µl di Low TE.

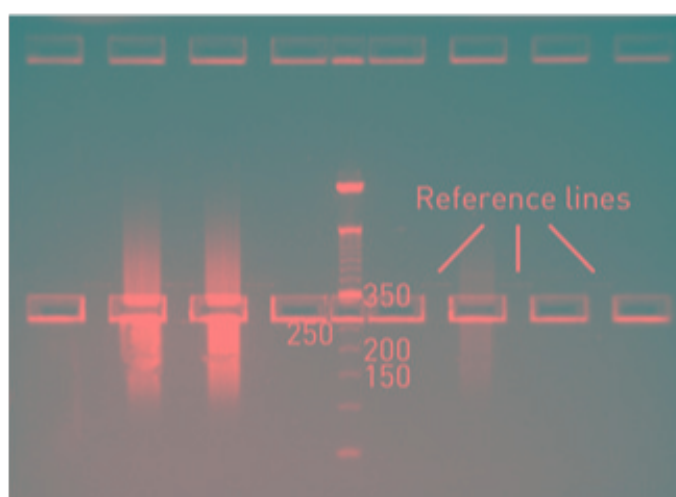
Le singole librerie di DNA così generate sono state sottoposte al legame di sequenze di riconoscimento univoche (barcode), un processo che ha consentito di creare un *pool* di librerie destinato ad essere sequenziato utilizzando un unico chip semiconduttore. La reazione di "barcoding" è stata eseguita a 25°C per 15 minuti, seguiti da 5 minuti a 72°C, utilizzando i seguenti reagenti:

- 10x ligase buffer, 10 µl;
- Ion P1 adapter, 2 µl;
- Ion Xpress Barcode X\* (\*barcode scelto per ciascun campione), 2 µl;
- dNTP mix, 2 µl;
- Acqua nuclease-free, 49 µl;
- DNA ligase, 2 µl
- Nick Repair Polymerase, 8 µl.

Al completamento della reazione, e dopo la purificazione con le biglie magnetiche, per ciascun pool sono stati selezionati i frammenti lunghi circa 330 bp attraverso il sistema E-Gel Size Select Agarose Gels, 2% (Fig.9) e poi amplificati, secondo le condizioni mostrate nella tabella 2, utilizzando una miscela composta da Platinum PCR SuperMix High Fidelity (100 µl) e Library Amplification Primer Mix (5 µl).

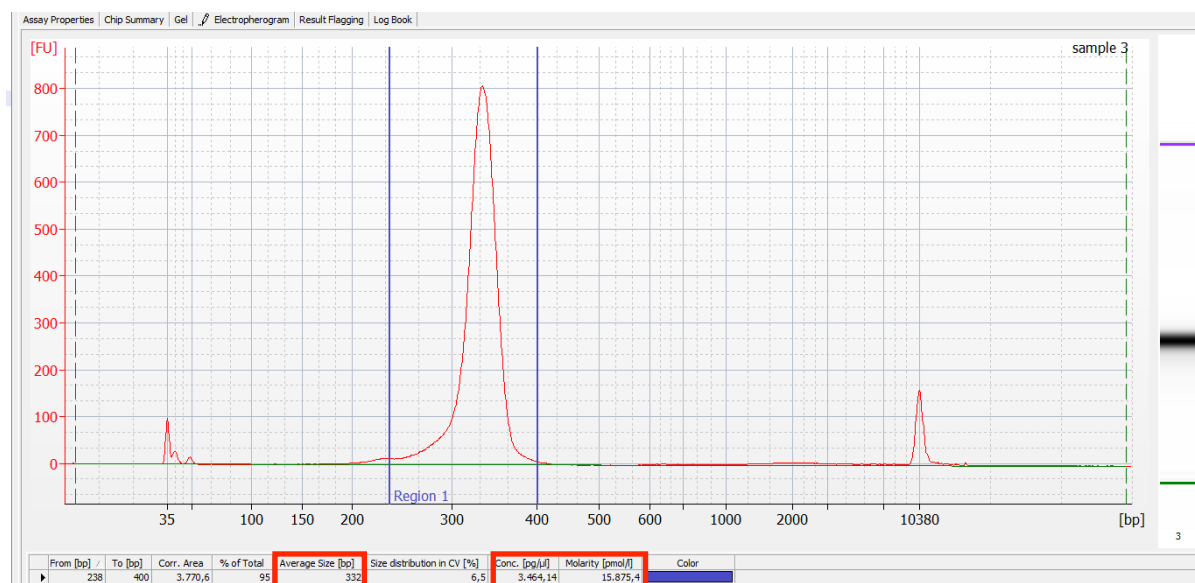
| Stage   | Step          | Temperatura | Tempo  |
|---------|---------------|-------------|--------|
| Holding | Denaturazione | 95°C        | 5 min  |
| 8 cicli | Denaturazione | 95°C        | 15 sec |
|         | Annealing     | 58°C        | 15 sec |
|         | Estensione    | 70°C        | 1 min  |
| Holding | -             | 4°C         | ∞      |

**Tabella 2.** Protocollo di amplificazione della libreria.



**Figura 9.** 200-base-read library gel.

Infine, le librerie amplificate sono state quantificate attraverso lo strumento Agilent 2100 Bioanalyzer (Fig.10).



**Figura 10.** Quantificazione della libreria, frammentata ed amplificata, mediante lo strumento Agilent 2100 Bioanalyzer. Il grafico fornisce informazioni sia sulle dimensioni medie dei frammenti sia sulla concentrazione della libreria.

### 3.5 Emulsion PCR

Le librerie sono state diluite alla concentrazione di 100 pM ciascuna e riunite in un unico pool equimolare che, a sua volta, è stato diluito alla concentrazione finale di 26 pM. Ad esso sono stati infine aggiunti i seguenti reagenti:

- Ion PGM Template OT2 200 Reagent Mix, 500 μl;
- Ion PGM Template OT2 200 Reagent B, 300 μl;
- Ion PGM Template OT2 200 Enzyme Mix, 50 μl;
- Ion PGM Template OT2 200 Ion Sphere Particles, 100 μl.

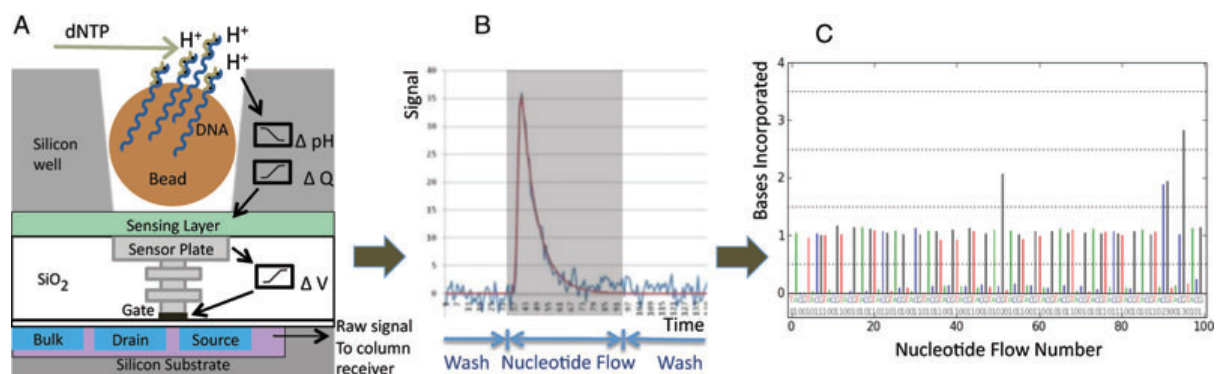
L'amplificazione clonale delle librerie (Emulsion PCR, e-PCR) è stata eseguita nello strumento Ion OneTouch 2 (Thermo Scientific) in microreattori, costituiti da

microbolle di acqua disperse in olio, all'interno delle quali si posizionano le Ion Sphere Particles (ISPs) sulla cui superficie si legano le molecole di DNA amplificate. Le ISPs su cui è avvenuta la reazione di amplificazione sono state arricchite mediante lo strumento Ion OneTouch ES.

### 3.6 Sequenziamento e analisi dei dati

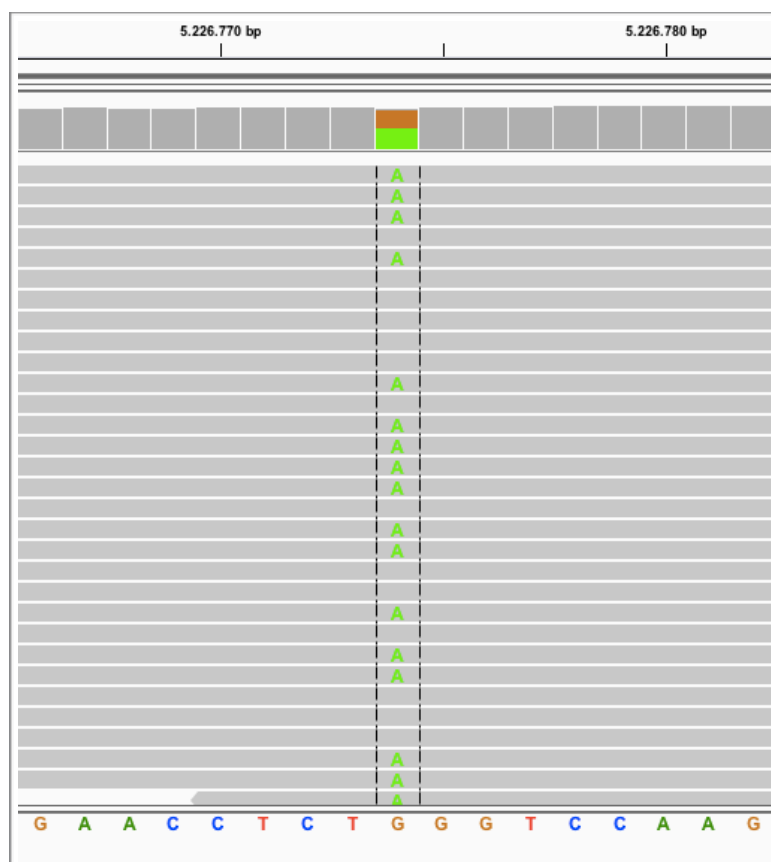
Il sequenziamento è avvenuto utilizzando la piattaforma Ion Torrent PGM che si basa sull'impiego di chip semiconduttori in cui le variazioni di pH registrate in seguito all'incorporazione di un nucleotide nella catena nascente sono rilevate da un sensore e convertite in un dato bioinformatico.

La reazione di sequenziamento è stata preceduta dalla denaturazione dei frammenti di DNA (95°C per 2 minuti), dall'ibridazione dei primers (37°C per 2 minuti) e, infine, dall'attivazione della DNA polimerasi (5 minuti a temperatura ambiente). Il campione è stato quindi caricato in un chip (Ion 314/316 chip), che è costituito da migliaia di micropozzetti, in ognuno dei quali alloggia una ISP ed i reagenti necessari per il sequenziamento. All'incorporazione di un nucleotide corrisponde il rilascio di uno ione idrogeno ( $H^+$ ) che determina una variazione di pH rilevata da un sensore. Nelle sequenze omopolimere il rilascio degli ioni  $H^+$  è proporzionale al numero di nucleotidi incorporati (Fig.11).



**Figura 11.** Overview della chimica di sequenziamento Ion Torrent (Merriman et al. 2012).

L'analisi dei file VCF di ciascun trio ha consentito di selezionare gli SNPs informativi, che sono stati successivamente amplificati e sequenziati nel corrispondente campione di cffDNA.



37

### **3.7 Estrazione del cffDNA**

Il campione di sangue periferico materno è stato sottoposto a due centrifugazioni successive: una prima a 3000g per 10 minuti a 4°C, per separare il plasma dalla parte corpuscolata e una seconda, a 16000g per 10 minuti a 4°C, per rimuovere le cellule materne residue.

Il plasma ottenuto (circa 10 ml) è stato suddiviso in aliquote da 1 ml e conservato a -80°C fino al momento dell'estrazione.

Il cffDNA è stato estratto da 8 ml di plasma materno utilizzando il kit QIAamp Circulating Nucleic Acid kit della ditta Qiagen, specifico per l'estrazione degli acidi nucleici circolanti nel plasma, nel siero e nelle urine.

Il protocollo di estrazione prevede una prima fase di lisi che avviene in condizioni altamente denaturanti, ad elevate temperature ed in presenza di proteinasi K e del Buffer ACL, che assicurano l'inattivazione delle DNasi e RNasi e completano la dissociazione degli acidi nucleici dalle proteine, lipidi e vescicole a cui è complessato.

Il campione lisato, a cui è stato aggiunto il Buffer ACB, è stato trasferito all'interno di colonnine (QIAamp Mini column) in cui il DNA viene adsorbito ad una membrana di silice in esse contenuta. Al contrario, le proteine e gli altri agenti contaminanti non vengono trattiene dalla membrana, ma eluiti con opportune soluzioni di lavaggio. Il DNA, così purificato, è stato infine eluito dalla colonnina utilizzando 150 µl di Buffer AVE.

### 3.8 Amplificazione del cffDNA

I frammenti contenenti gli SNPs informativi selezionati nei DNA parentali, sono stati successivamente amplificati nel cffDNA utilizzando il kit Platinum® PCR SuperMix High Fidelity (Thermo Fisher Scientific) combinando i seguenti reagenti:

- High Fidelity Platinum® Supermix;
- Spermina 0.5 mM + 1mg/ml BSA;
- Primer Forward (25pMol);
- Primer Reverse (25pMol);
- cell free DNA (2 µl).
- Le condizioni di PCR utilizzate sono riportate nella seguente tabella (tab.3):

| Stage    | Step          | Temperatura             | Tempo  |
|----------|---------------|-------------------------|--------|
| Holding  | Denaturazione | 95°C                    | 3 min  |
| 30 cicli | Denaturazione | 95°C                    | 30 sec |
|          | Annealing     | 55°/57°/60°/62°/65°/67° | 30 sec |
|          | Estensione    | 68°C                    | 1 min  |
| Holding  | Estensione    | 68°C                    | 5 min  |

**Tabella 3.** Condizioni di PCR per l'amplificazione del DNA estratto del plasma.

In ciascun campione di cffDNA, sono state amplificate alcune regioni localizzate nei cromosomi X e Y che hanno consentito di eseguire la diagnosi non invasiva del sesso fetale. Tali regioni sono localizzate all'interno dei geni *TSPY1*, *SRY* e del gene omologo *DZFX/Y*. L'amplificazione del gene *TSPY* è stata eseguita con lo stesso protocollo utilizzato per gli ampliconi del cluster  $\beta$ -globinico mentre per i frammenti che mappano nei geni *ZFX/ZFY* e *SRY* sono state utilizzate le condizioni di PCR riportate nella tabella 4.

| Stage                                        | Step          | Temperatura | Tempo  |
|----------------------------------------------|---------------|-------------|--------|
| Holding                                      | Denaturazione | 95°C        | 5 min  |
| 15 cicli (con decremento di 0.5°C per ciclo) | Denaturazione | 95°C        | 30 sec |
|                                              | Annealing     | 64°C        | 30 sec |
|                                              | Estensione    | 68°C        | 1 min  |
| 25 cicli                                     | Denaturazione | 95°C        | 30 sec |
|                                              | Annealing     | 56°C        | 30 sec |
|                                              | Estensione    | 68°C        | 1 min  |
| Holding                                      | Estensione    | 72°C        | 10 min |

**Tabella 4.** Condizioni di PCR per l'amplificazione dei frammenti che mappano sui geni ZFX/ZFY e SRY.

Il controllo dell'avvenuta reazione di PCR è stata eseguito attraverso elettroforesi su gel di agarosio al 2%.

I prodotti di PCR ottenuti da ciascun campione di cffDNA sono stati riuniti in un unico pool, che è stato successivamente purificato mediante biglie magnetiche Agecourt AMPure XP e quantificato al Qubit utilizzando il kit Qubit® dsDNA HS (High Sensitivity) Assay.

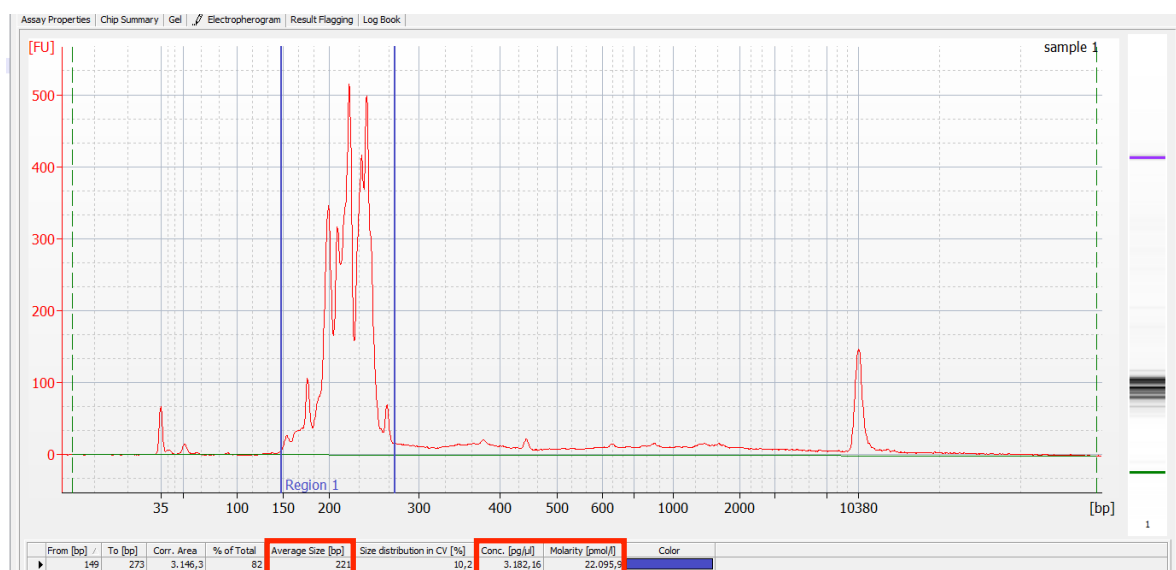
### ***3.9 Preparazione della DNA library, emulsion PCR e sequenziamento.***

Il pool di ampliconi è stato sottoposto alla reazione di “end-repair” per la preparazione delle estremità al successivo legame con i “barcode” molecolari. La reazione è stata eseguita a temperatura ambiente, per 20 minuti, in presenza dei seguenti reagenti:

- Pool di ampliconi, 50 ng, diluiti in 79 µl di acqua nuclease-free;
- 5x End Repair Buffer, 20 µl;
- End Repair Enzyme, 1 µl.

Al termine dell'incubazione la libreria è stata purificata con le biglie magnetiche, eluita in 25 µl di Low TE e sottoposta alla reazione di “barcoding”.

La libreria è stata quindi amplificata, quantificata attraverso lo strumento Agilent 2100 Bioanalyzer (Fig.13), sottoposta a emulsion PCR e sequenziata utilizzando il protocollo precedentemente descritto.



**Figura 13.** Quantificazione mediante lo Strumento 2100 Bioanalyzer di una libreria di cffDNA amplificata.

### 3.10 Costruzione del pannello di aplotipi di riferimento

Il pannello di aplotipi di riferimento della regione di interesse del cromosoma 11 (5209078-5271956), è stato costruito integrando gli aplotipi presenti nel database del progetto 1000 Genomi con gli aplotipi dei 39 Trio sardi, sequenziati mediante piattaforma Ion Torrent PGM, allo scopo di incrementare la rappresentazione della mutazione  $\beta^{*39}$ . I file BAM di ciascun Trio sono stati inizialmente allineati attraverso il programma BWA ed analizzati mediante il programma Samtools, al fine di ottenere tre file VCF in cui gli aplotipi dei campioni non sono ancora attribuiti. Tali file sono

stati filtrati per qualità e per posizioni (considerando solo le varianti presenti in 1000 Genomi) e uniti al set di aplotipi di 1000 Genomi. Infine, sono stati generati gli aplotipi con il programma Shapeit, utilizzando le informazioni provenienti da PIRs (phase informative reads) e considerando i DNA parentali ed il DNA fetale come trio.

### ***3.11 Analisi dei dati***

I dati ottenuti dal sequenziamento del cffDNA e dei corrispondenti DNA parentali sono stati analizzati mediante l'utilizzo di una pipeline automatizzata, sviluppata al fine di effettuare la predizione del genotipo fetale, sfruttando un approccio basato sull'inferenza nel cffDNA degli aplotipi parentali.

Tale pipeline è stata sviluppata sulla base di altri modelli descritti in letteratura (Kitzman et al. 2012, Chen et al. 2013) ed adattata alla diagnosi non invasiva di una malattia monogenica, la  $\beta$ -talassemia.

Essa è in grado di eseguire la predizione degli aplotipi parentali più probabilmente ereditati dal feto; il risultato viene poi comparato con i genotipi determinati attraverso il sequenziamento del DNA fetale ottenuto mediante villocentesi.

La pipeline è scritta in linguaggio di programmazione Python 2.7, è eseguibile mediante terminale Linux e utilizza alcuni software bioinformatici opensource (BWA, SAMtools, BCFtools, SHAPEiT, VCFtools).

Gli aplotipi materno e paterno ereditati dal feto sono stati determinati nel cffDNA utilizzando un approccio di predizione "sito per sito", che valuta i singoli siti indipendentemente, e applicando un modello di Hidden Markov (HMM) (algoritmo di Viterbi), per la correzione di eventuali errori di "switch".

L'analisi dei dati di sequenziamento dei DNA parentali e del cffDNA è avvenuta secondo il seguente workflow:

1. ricostruzione degli aplotipi paterni e materni;
2. suddivisione degli SNPs informativi in 4 categorie;
3. conta dei siti del cffDNA;
4. quantificazione della frazione fetale;
5. predizione dell'aplotipo paterno più probabilmente ereditato;
6. predizione dell'aplotipo materno più probabilmente ereditato;
7. predizione del genotipo fetale;
8. validazione del risultato.

Nel primo passaggio la pipeline utilizza come input i file BAM relativi al sequenziamento del DNA del padre e della madre, li allinea alla sequenza di riferimento del cromosoma 11, con il programma BWA e, mediante SAMtools, produce un file VCF (Variant Call Format), nel quale sono riportate informazioni quali il genotipo, il coverage, la qualità delle reads utilizzate.

Il programma SHAPEiT analizza i genitori e ne ricostruisce gli aplotipi utilizzando un Pannello di Aplotipi di Riferimento (1000 Genomi + 39 trio), senza l'ausilio delle informazioni provenienti dal DNA estratto dai villi coriali.

Il file VCF, contenente gli aplotipi parentali, è stato analizzato per individuare gli SNPs informativi appartenenti alle seguenti categorie (tab.5):

- Homozygous only → SNPs per i quali il padre e la madre sono omozigoti per alleli diversi; vengono usati nel calcolo della frazione fetale;
- Paternal only → SNPs per i quali il padre è eterozigote e la madre omozigote; vengono utilizzati per il calcolo della frazione fetale e per la predizione dell'ereditarietà paterna;
- Maternal only → SNPs per i quali il padre è omozigote e la madre eterozigote; utilizzati per la predizione dell'ereditarietà materna;

- Heterozygous only → SNPs per i quali i genitori sono entrambi eterozigoti; partecipano anch'essi alla predizione dell'ereditarietà materna.

| HOMOZYGOUS-ONLY |       |                      | PATERNAL-ONLY     |       |                                  |
|-----------------|-------|----------------------|-------------------|-------|----------------------------------|
| Padre           | Madre | Possibile predizione | Padre             | Madre | Possibile predizione             |
| 0   0           | 1   1 | 0/1                  | 0   1             | 0   0 | 0/0      0/1                     |
| 1   1           | 0   0 |                      | 0   1             | 1   1 | 1/1      0/1                     |
| MATERNAL-ONLY   |       |                      | HETEROZYGOUS-ONLY |       |                                  |
| Padre           | Madre | Possibile predizione | Padre             | Madre | Possibile predizione             |
| 0   0           | 0   1 | 0/0      0/1         | 0   1             | 0   1 | 0/0      0/1<br><br>1/1      0/1 |
| 1   1           | 0   1 | 1/1      0/1         | 1   0             | 1   0 |                                  |

**Tabella 5.** Suddivisione degli SNPs in 4 categorie.

Successivamente la pipeline analizza il file BAM del cffDNA. Mediante BWA lo allinea alla sequenza di riferimento e attraverso SAMtools esegue la chiamata delle varianti e produce un file che contiene il numero di reads per ogni sito analizzato.

La pipeline procede con il calcolo della frazione fetale (FF) media che viene determinata prendendo in esame i siti *homozygous only* e *paternal only* per i quali il contributo dell'allele paterno sia superiore all'1,4% ma inferiore al 11%. I campioni nei quali venga determinata una  $FF < 2\%$  e  $> 20\%$  vengono automaticamente considerati non processabili ed esclusi dalle analisi successive.

La predizione dell'aplotipo paterno viene eseguita utilizzando i siti paternal only, per i quali viene applicata la formula  $N \cdot FF / 2$  (Kitzman et al. 2012) dove N equivale al numero di reads osservate nell'allele paterno non condiviso con il genoma materno. Il risultato è pari a 0 quando il feto non eredita l'allele paterno che differisce da quello materno ed è, al contrario, diverso da 0 se tale allele viene ereditato. Per escludere eventuali risultati falsi positivi, non vengono presi in esame i siti il cui

coverage sia inferiore a 1000 e viene inoltre imposto un cut-off che varia in funzione della frazione fetale. In particolare, la presenza dell'allele paterno viene inferita solo quando la percentuale delle reads in cui esso è presente risulta superiore al 10° percentile di una distribuzione di 100 parti identiche, con valore iniziale pari a 0 e valore finale equivalente alla frazione fetale.

La predizione sito per sito eseguita dalla pipeline produce blocchi di aplotipi contigui; poiché assumiamo che in una regione genomica relativamente piccola, come quella sequenziata, gli eventi di ricombinazione siano estremamente rari, è stato utilizzato l'algoritmo di Viterbi, come modello HMM, per correggere gli eventuali errori di *switch* tra i blocchi aplotipici e per inferire l'aplotipo paterno più probabilmente ereditato dal feto.

La predizione dell'aplotipo materno viene effettuata utilizzando tutti i siti in cui la madre è eterozigote (*maternal only* e *heterozygous only*). Poiché la determinazione dell'aplotipo materno è strettamente legata alla predizione dell'aplotipo paterno ereditato, l'assenza di siti *paternal only* impedisce la determinazione di entrambi gli aplotipi parentali ereditati dal feto. Conoscendo l'aplotipo ereditato per via paterna, l'inferenza dell'allele materno più probabilmente ereditato viene eseguita misurando gli sbilanciamenti allelici osservati per ciascun sito e calcolando la probabilità che l'allele ereditato per via materna sia identico o diverso da quello ereditato per via paterna. Tale probabilità ( $k$ ) viene calcolata utilizzando le seguenti formule (Kitzman et al.), dove  $N$  rappresenta il numero di reads totali e  $FF$  la frazione fetale:

- $Pr(K=k|N,FF)=Bin(N, (1-FF)/2 + FF/2 + FF/2)$  per calcolare la probabilità che il genotipo sia omozigote a livello di un determinato sito;

- $Pr(K=k|N, FF) = \text{Bin}(N, (1-FF)/2 + FF/2) = \text{Bin}(N, 0.5)$  per calcolare invece la probabilità che il genotipo sia eterozigote.

Al fine di escludere la presenza di falsi positivi vengono analizzati solo quei siti che mostrano uno sbilanciamento allelico inferiore o uguale al 54% ed un coverage maggiore o uguale a 1500. Come descritto per la predizione dell'aplotipo paterno, anche l'aplotipo materno più probabilmente ereditato dal feto viene inferito utilizzando il modello HMM e l'algoritmo di Viterbi.

I genotipi fetali determinati nel cffDNA sono infine confrontati con quelli ottenuti dal sequenziamento del DNA fetale estratto dai villi coriali.

I passaggi dell'analisi dei dati sono rappresentati nella figura 14.

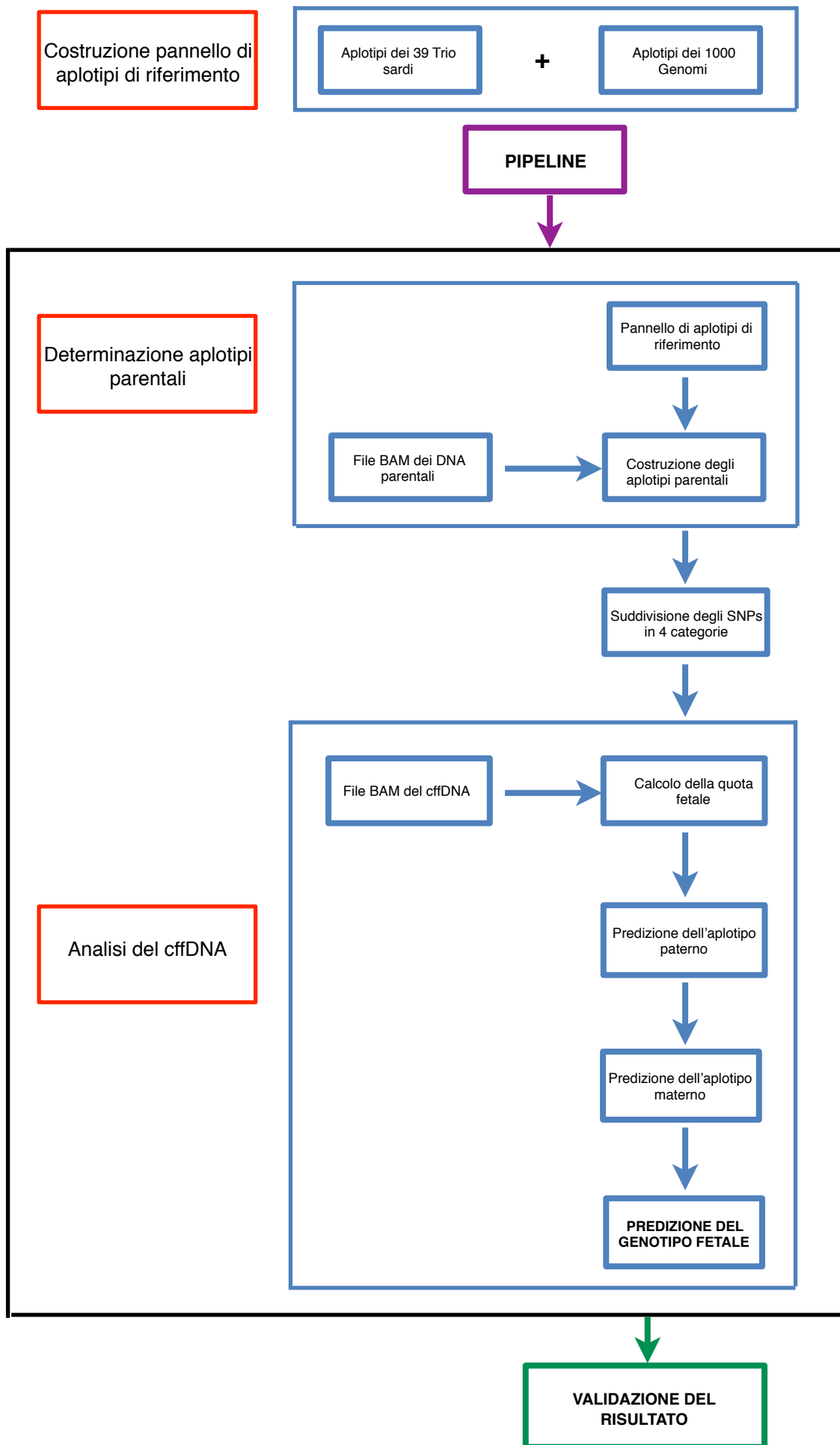
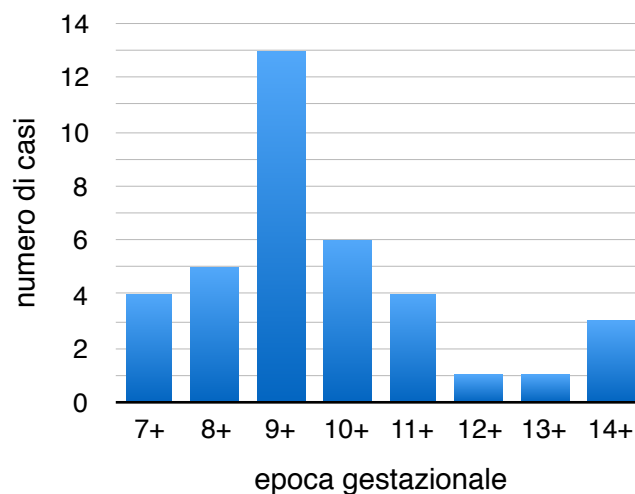


Figura.14 Workflow della pipeline.

## 4. Risultati

In questo studio, finalizzato a sviluppare un protocollo di diagnosi prenatale non invasiva di  $\beta$ -talassemia, sono stati sequenziati 37 campioni di DNA estratto dal plasma di donne in epoca gestazionale compresa tra la 7<sup>a</sup> e la 14<sup>a</sup>+3 settimana di gestazione (media di 9+6 settimane, fig.15).

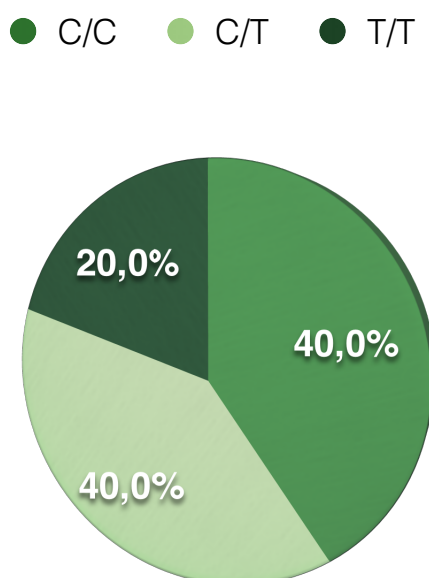


**Figura 15.** Distribuzione dei campioni nelle diverse epoche gestazionali.

In 7/37 campioni di cffDNA sequenziati non è stato possibile predire il genotipo fetale. In particolare, in 2/7 campioni l'assenza di siti "paternal only" non ha consentito di eseguire la predizione dell'aplotipo paterno e, di conseguenza, neanche quella dell'aplotipo ereditato dal feto per via materna. Negli altri 5 campioni non è stato possibile calcolare la frazione fetale per assenza di SNPs informativi appartenenti alle categorie paternal only e homozygous only.

La pipeline ha completato l'analisi in 30/37 campioni nei quali i genotipi determinati attraverso la villocentesi e attesi nei corrispondenti campioni di cffDNA erano i

seguenti: omozigoti per l'allele wild type C/C (12/30), eterozigoti C/T (12/30) e omozigoti per l'allele mutato T/T (6/30) (fig.16).



**Figura 16.** Classificazione dei genotipi attesi dei campioni analizzati.

Il sequenziamento dei DNA parentali ha evidenziato, in media, la presenza di 115 SNPs informativi per ciascuna famiglia. Di questi SNPs, tuttavia, è stato possibile analizzarne solo il 53% (media di 61/115 SNPs) nei corrispondenti campioni di cffDNA (tab.8). Numerosi SNPs esclusi dal sequenziamento sono localizzati nei geni  $\gamma$ -globinici, che si caratterizzano per una elevata omologia di sequenza, altri, invece, risiedono nella sequenza ripetuta LINE1, che si estende per circa 6 kb, ad una distanza di 4,7 kb dal 3' del gene *HBB*. Per tali SNPs non è stato possibile disegnare i relativi saggi di amplificazione, che consentissero di selezionare frammenti di dimensioni inferiori alle 150 bp, gli unici compatibili con la struttura frammentata del cffDNA.

Dei siti sequenziati nei campioni di cffDNA, soltanto il 49,2%, in media, è stato analizzato attraverso la pipeline automatizzata per la predizione degli aplotipi parentali (tab. 8 e 9). In particolare, di tali siti è stato utilizzato, in media, il 21,3% (13/61) per la predizione dell'aplotipo paterno e il 27,9% (17/61) per la predizione dell'aplotipo materno (tab.9). Il 50,8% degli SNPs è stato escluso dall'analisi poiché non soddisfavano i cut-off stabiliti (coverage inferiore a 1000x per gli SNPs *paternal only*, coverage inferiore a 1500x e sbilanciamento maggiore del 54% per gli SNPs *maternal only* ed *heterozygous only*).

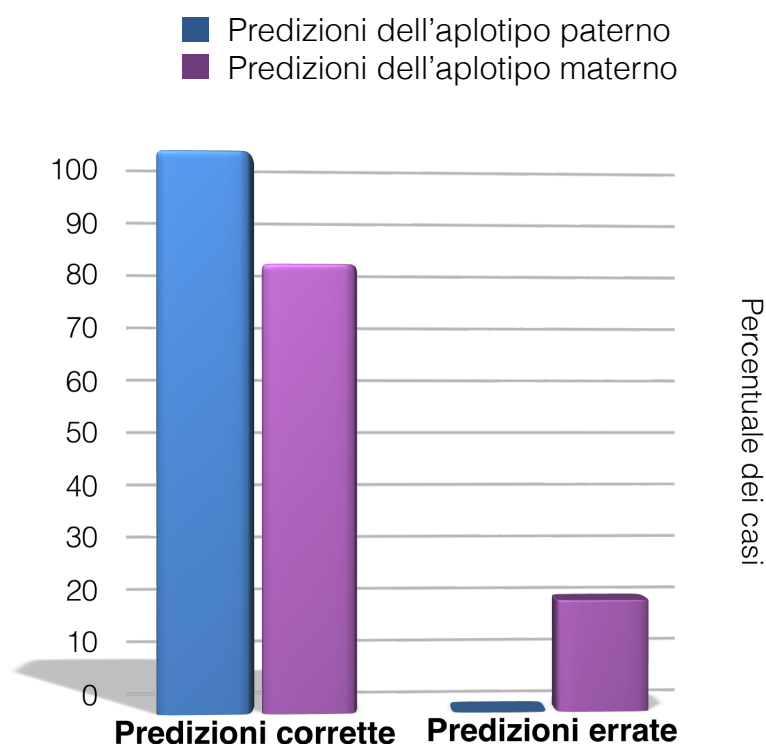
Il sequenziamento dei campioni di cffDNA ha prodotto una media di 442689 reads per campione, con un coverage medio di 7186x, mentre la frazione fetale media è risultata essere pari al 7%, con un range compreso tra il 3,7% e il 12,6%. (tab.10).

Poiché la predizione del genotipo fetale è correlata all'inferenza nel cffDNA degli aplotipi parentali ereditati dal feto, è stato necessario sviluppare un approccio indipendente dall'analisi del trio, che consentisse di ricostruire gli aplotipi parentali senza considerare le informazioni provenienti dal DNA estratto dai villi coriali. Per questo motivo è stata inizialmente creata una referenza di aplotipi implementando il pannello di 1000 Genomi con gli aplotipi definiti in 39 trio sardi.

Gli aplotipi parentali ricostruiti attraverso la pipeline sono stati quindi messi a confronto con quelli prodotti attraverso l'analisi del trio, con l'ausilio di Shapeit e della referenza precedentemente generata.

L'accuratezza media di ricostruzione degli aplotipi è stata del 99,7% per gli aplotipi paterni e del 99,6% per quelli materni.

La predizione dell'aplotipo paterno è stata eseguita correttamente in 30/30 casi (100%); la predizione dell'aplotipo materno è risultata corretta in 24/30 casi (80%) ed errata in 6/30 casi (20%) (fig.17 e tab.10).



**Figura 17.** Detection dell'aplotipo paterno e materno.

In generale, l'attribuzione dell'aplotipo materno errato è stata causata dalla presenza nel cfDNA di uno sbilanciamento allelico errato, esteso a più siti, che, dopo correzione con HMM, ha portato alla predizione dell'aplotipo non ereditato dal feto. Abbiamo inoltre osservato che un'altra possibile causa d'errore può essere rappresentata dalla non accurata ricostruzione, da parte della pipeline, degli aplotipi parentali, che costituiscono il termine di riferimento per l'inferenza degli aplotipi più probabilmente ereditati dal feto. In particolare, abbiamo riscontrato questo tipo di errore in un campione nel quale la percentuale di accuratezza nella costruzione dell'aplotipo materno è risultata pari al 94.5%, un valore notevolmente inferiore rispetto alla media (99,6%) osservata negli altri campioni.

| Campione | SNPs informativi totali | SNPs sequenziati (% sul totale) | SNPs utilizzati per la predizione (% sul totale) |
|----------|-------------------------|---------------------------------|--------------------------------------------------|
| 1        | 133                     | 69 (51,9%)                      | 44 (33,1%)                                       |
| 2        | 121                     | 58 (47,9%)                      | 27 (22,3%)                                       |
| 3        | 124                     | 60 (48,4%)                      | 26 (21,0%)                                       |
| 4        | 107                     | 57 (53,3%)                      | 18 (16,8%)                                       |
| 5        | 109                     | 58 (53,2%)                      | 31 (28,4%)                                       |
| 6        | 96                      | 58 (60,4%)                      | 34 (35,4%)                                       |
| 7        | 89                      | 55 (61,8%)                      | 34 (38,2%)                                       |
| 8        | 125                     | 66 (52,8%)                      | 45 (36,0%)                                       |
| 9        | 116                     | 76 (65,5%)                      | 37 (31,9%)                                       |
| 10       | 136                     | 75 (55,1%)                      | 34 (25,0%)                                       |
| 11       | 134                     | 61 (45,5%)                      | 24 (17,9%)                                       |
| 12       | 165                     | 80 (48,5%)                      | 38 (23,0%)                                       |
| 13       | 109                     | 58 (53,2%)                      | 27 (24,8%)                                       |
| 14       | 137                     | 64 (46,7%)                      | 48 (35,0%)                                       |
| 15       | 40                      | 13 (32,5%)                      | 12 (30,0%)                                       |
| 16       | 126                     | 66 (52,4%)                      | 36 (28,6%)                                       |
| 17       | 143                     | 69 (48,3%)                      | 38 (26,6%)                                       |
| 18       | 132                     | 67 (50,8%)                      | 27 (20,5%)                                       |
| 19       | 94                      | 50 (53,2%)                      | 33 (35,1%)                                       |
| 20       | 100                     | 44 (44,0%)                      | 32 (32,0%)                                       |
| 21       | 111                     | 61 (55,0%)                      | 30 (27,0%)                                       |
| 22       | 113                     | 45 (39,8%)                      | 19 (16,8%)                                       |
| 23       | 105                     | 56 (53,3%)                      | 38 (36,2%)                                       |
| 24       | 136                     | 67 (49,3%)                      | 37 (27,2%)                                       |
| 25       | 72                      | 47 (65,3%)                      | 26 (36,1%)                                       |
| 26       | 109                     | 56 (51,4%)                      | 25 (22,9%)                                       |
| 27       | 111                     | 58 (52,3%)                      | 26 (23,4%)                                       |
| 28       | 95                      | 56 (58,9%)                      | 10 (10,5%)                                       |
| 29       | 137                     | 81 (59,1%)                      | 21 (15,3%)                                       |
| 30       | 138                     | 86 (62,3%)                      | 37 (26,8%)                                       |

**Tabella 8.** SNPs informativi totali, SNPs sequenziati e SNPs utilizzati per le predizioni del genotipo fetale.

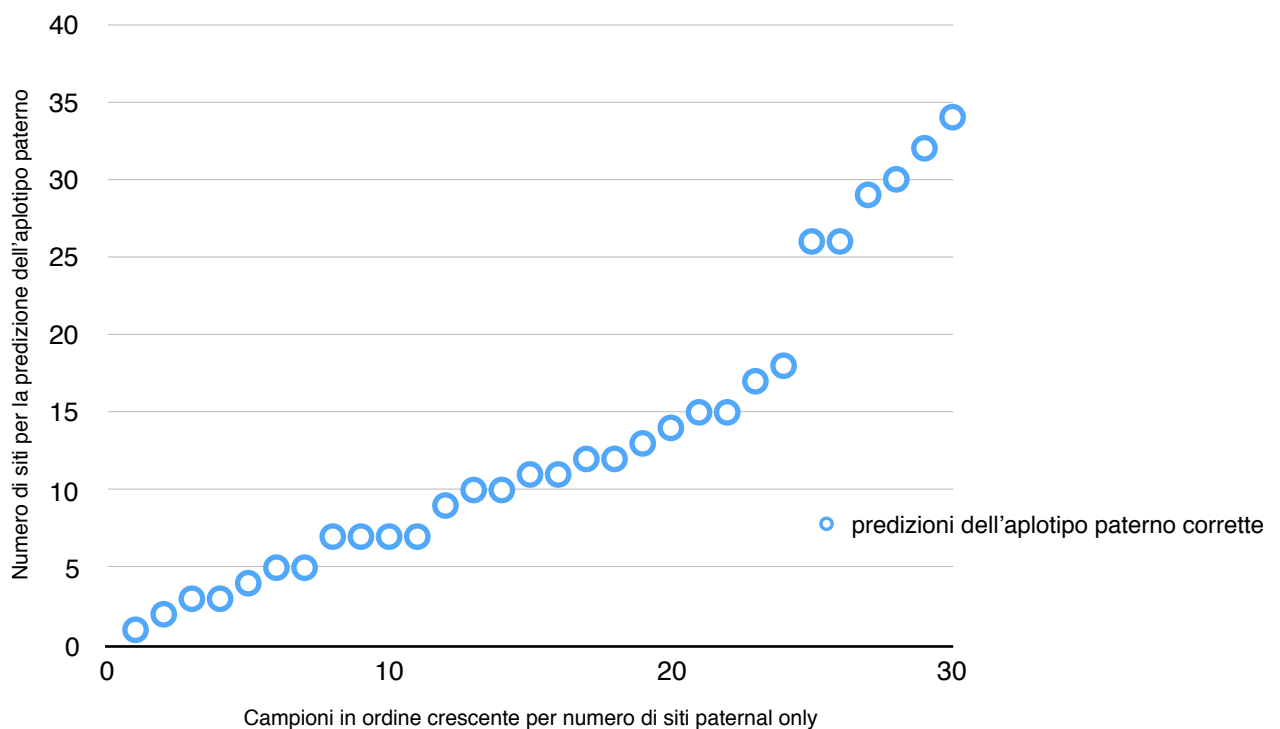
| Campione | SNPs paterni<br>(% sui siti<br>sequenziati) | SNPs materni<br>(% sui siti<br>sequenziati) | SNPs esclusi<br>(% sui siti<br>sequenziati) |
|----------|---------------------------------------------|---------------------------------------------|---------------------------------------------|
| 1        | 32 (46,4%)                                  | 12 (17,4%)                                  | 25 (36,2%)                                  |
| 2        | 11 (19,0%)                                  | 16 (27,6%)                                  | 31 (53,4%)                                  |
| 3        | 1 (1,7%)                                    | 25 (41,7%)                                  | 34 (56,7%)                                  |
| 4        | 3 (5,3%)                                    | 15 (26,3%)                                  | 39 (68,4%)                                  |
| 5        | 5 (8,6%)                                    | 26 (44,8%)                                  | 27 (46,6%)                                  |
| 6        | 14 (24,1%)                                  | 20 (34,5%)                                  | 24 (41,4%)                                  |
| 7        | 7 (12,7%)                                   | 27 (49,1%)                                  | 21 (38,2%)                                  |
| 8        | 30 (45,5%)                                  | 15 (22,7%)                                  | 21 (31,8%)                                  |
| 9        | 26 (34,2%)                                  | 11 (14,5%)                                  | 39 (51,3%)                                  |
| 10       | 15 (20,0%)                                  | 19 (25,3%)                                  | 41 (54,7%)                                  |
| 11       | 10 (16,4%)                                  | 14 (23,0%)                                  | 37 (60,7%)                                  |
| 12       | 17 (21,3%)                                  | 21 (26,3%)                                  | 42 (52,5%)                                  |
| 13       | 10 (17,2%)                                  | 17 (29,3%)                                  | 31 (53,4%)                                  |
| 14       | 34 (53,1%)                                  | 14 (21,9%)                                  | 16 (25,0%)                                  |
| 15       | 11 (84,6%)                                  | 1 (7,7%)                                    | 1 (7,7%)                                    |
| 16       | 26 (39,4%)                                  | 10 (15,2%)                                  | 30 (45,5%)                                  |
| 17       | 15 (21,7%)                                  | 23 (33,3%)                                  | 31 (44,9%)                                  |
| 18       | 7 (10,4%)                                   | 20 (29,9%)                                  | 40 (59,7%)                                  |
| 19       | 7 (14,0%)                                   | 26 (52,0%)                                  | 17 (34,0%)                                  |
| 20       | 29 (65,9%)                                  | 3 (6,8%)                                    | 12 (27,3%)                                  |
| 21       | 13 (21,3%)                                  | 17 (27,9%)                                  | 31 (50,8%)                                  |
| 22       | 3 (6,7%)                                    | 16 (35,6%)                                  | 26 (57,8%)                                  |
| 23       | 12 (21,4%)                                  | 26 (46,4%)                                  | 18 (32,1%)                                  |
| 24       | 4 (6,0%)                                    | 33 (49,3%)                                  | 30 (44,8%)                                  |
| 25       | 5 (10,6%)                                   | 21 (44,7%)                                  | 21 (44,7%)                                  |
| 26       | 7 (12,5%)                                   | 18 (32,1%)                                  | 31 (55,4%)                                  |
| 27       | 18 (31,0%)                                  | 8 (13,8%)                                   | 32 (55,2%)                                  |
| 28       | 2 (3,6%)                                    | 8 (14,3%)                                   | 46 (82,1%)                                  |
| 29       | 9 (11,1%)                                   | 12 (14,8%)                                  | 60 (74,1%)                                  |
| 30       | 12 (14,0%)                                  | 25 (29,1%)                                  | 49 (57,0%)                                  |

**Tabella 9.** SNPs utilizzati per le predizioni dell'aplotipo paterno e materno e SNPs esclusi dall'analisi.

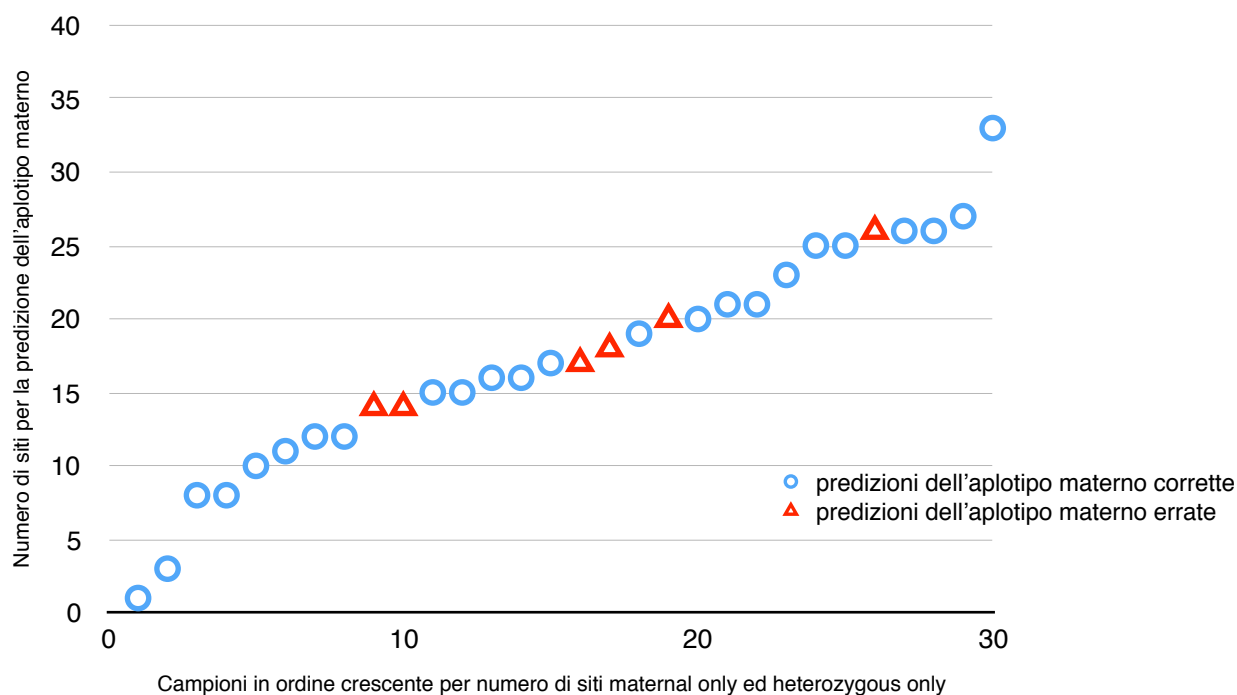
| Campione | Predizione genotipo fetale | Diagnosi  | Predizione $\beta^*39$ paterna | Predizione $\beta^*39$ materna | % genotipi corretti | Epoca gestazionale | Frazione fetale | Coverage | Numero di reads |
|----------|----------------------------|-----------|--------------------------------|--------------------------------|---------------------|--------------------|-----------------|----------|-----------------|
| 1        | Normale                    | Normale   | WT                             | WT                             | 99,7                | 8+5                | 4,6             | 5319     | 364874          |
| 2        | Portatore                  | Portatore | WT                             | $\beta^*39$                    | 99,9                | 9+5                | 7,4             | 7727     | 456359          |
| 3        | Portatore                  | Portatore | WT                             | $\beta^*39$                    | 100,0               | 14                 | 4               | 6005     | 395070          |
| 4        | Affetto                    | Affetto   | $\beta^*39$                    | $\beta^*39$                    | 100,0               | 9+3                | 3,7             | 3948     | 265576          |
| 5        | Normale                    | Normale   | WT                             | WT                             | 100,0               | 7                  | 5,6             | 8473     | 523614          |
| 6        | Affetto                    | Affetto   | $\beta^*39$                    | $\beta^*39$                    | 99,9                | 9+4                | 8,6             | 11014    | 653000          |
| 7        | Normale                    | Normale   | WT                             | WT                             | 99,9                | 14+3               | 4,2             | 8596     | 483198          |
| 8        | Normale                    | Normale   | WT                             | WT                             | 99,8                | 9                  | 4,9             | 8152     | 546375          |
| 9        | Portatore                  | Portatore | WT                             | $\beta^*39$                    | 99,9                | 10                 | 7               | 5781     | 444691          |
| 10       | Portatore                  | Portatore | WT                             | $\beta^*39$                    | 100,0               | 11+3               | 5,5             | 6258     | 483118          |
| 11       | Normale                    | Portatore | WT                             | WT                             | 98,0                | 9+4                | 4,9             | 2246     | 156862          |
| 12       | Portatore                  | Portatore | $\beta^*39$                    | WT                             | 99,8                | 9                  | 9,7             | 5217     | 424340          |
| 13       | Normale                    | Normale   | WT                             | WT                             | 100,0               | 9+3                | 7,4             | 5242     | 325080          |
| 14       | Normale                    | Portatore | WT                             | WT                             | 95,2                | 8+5                | 4,6             | 7328     | 472053          |
| 15       | Portatore                  | Portatore | WT                             | $\beta^*39$                    | 100,0               | 9+3                | 5,8             | 11861    | 214823          |
| 16       | Normale                    | Normale   | WT                             | WT                             | 98,6                | 10                 | 7,1             | 4721     | 317601          |
| 17       | Portatore                  | Portatore | WT                             | $\beta^*39$                    | 99,8                | 10                 | 7,9             | 9945     | 758990          |
| 18       | Portatore                  | Normale   | WT                             | $\beta^*39$                    | 90,6                | 11                 | 10,9            | 7019     | 482577          |
| 19       | Affetto                    | Portatore | $\beta^*39$                    | $\beta^*39$                    | 94,0                | 8                  | 5,9             | 6741     | 366668          |
| 20       | Affetto                    | Affetto   | $\beta^*39$                    | $\beta^*39$                    | 99,5                | 9+2                | 8,3             | 5383     | 246203          |
| 21       | Portatore                  | Affetto   | $\beta^*39$                    | WT                             | 94,3                | 7                  | 11,4            | 6742     | 432449          |
| 22       | Normale                    | Normale   | WT                             | WT                             | 99,9                | 9+4                | 12,5            | 11876    | 600393          |
| 23       | Portatore                  | Portatore | WT                             | $\beta^*39$                    | 100,0               | 8                  | 7,8             | 8087     | 533372          |
| 24       | Affetto                    | Affetto   | $\beta^*39$                    | $\beta^*39$                    | 99,6                | 10                 | 12,6            | 7356     | 497791          |
| 25       | Normale                    | Normale   | WT                             | WT                             | 99,9                | 9                  | 8,7             | 12327    | 621664          |
| 26       | Portatore                  | Normale   | WT                             | $\beta^*39$                    | 95,1                | 14+3               | 4,6             | 6627     | 384619          |
| 27       | Normale                    | Normale   | WT                             | WT                             | 99,8                | 7+2                | 7,4             | 7778     | 466341          |
| 28       | Portatore                  | Portatore | $\beta^*39$                    | WT                             | 99,9                | 11+4               | 4,6             | 5505     | 337961          |
| 29       | Affetto                    | Affetto   | $\beta^*39$                    | $\beta^*39$                    | 99,8                | 12+3               | 5,6             | 6111     | 516133          |
| 30       | Normale                    | Normale   | WT                             | WT                             | 99,7                | 13                 | 5,8             | 6197     | 508886          |
| Media    |                            |           |                                |                                |                     | 9+6                | 7,0             | 7186     | 442689          |

**Tabella 10.** Risultati generali. In blu le predizioni corrette, in rosso le predizioni errate.

I grafici riportati nelle figure 18 e 19 mostrano la distribuzione del numero di siti utilizzati, rispettivamente, per la predizione paterna e per quella materna. In generale, in entrambi i grafici, non si osserva una evidente correlazione tra il numero di siti analizzati e il risultato ottenuto nella predizione degli aplotipi fetali. Tuttavia, la maggiore concentrazione di una errata predizione dell'aplotipo materno tra i campioni con numero di siti analizzati medio-alto (compreso tra 14 e 26), induce a pensare che altre cause, quali ad esempio una minore frazione fetale, l'interferenza tra diversi ampliconi, una più scarsa performance delle reazioni di amplificazione o la presenza di un eventuale vanishing twin, possano aver generato sbilanciamenti allelici errati nei siti analizzati.



**Figura 18.** Distribuzione del numero di siti analizzati per la predizione dell'aplotipo paterno.

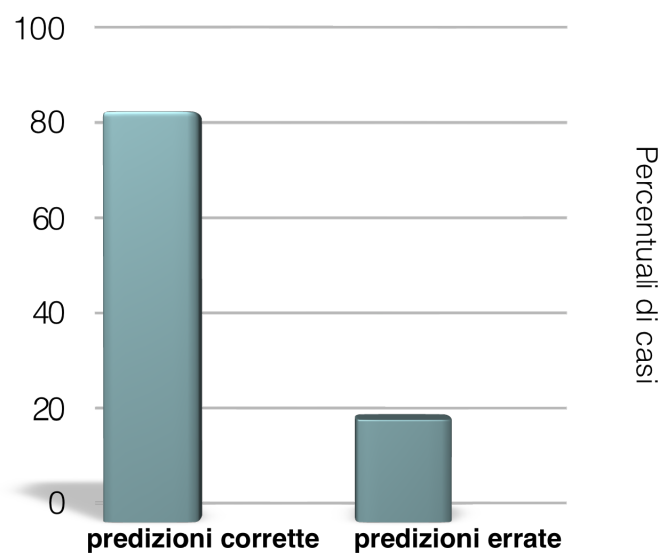


**Figura 19.** Distribuzione del numero di siti analizzati per la predizione dell'aplotipo materno.

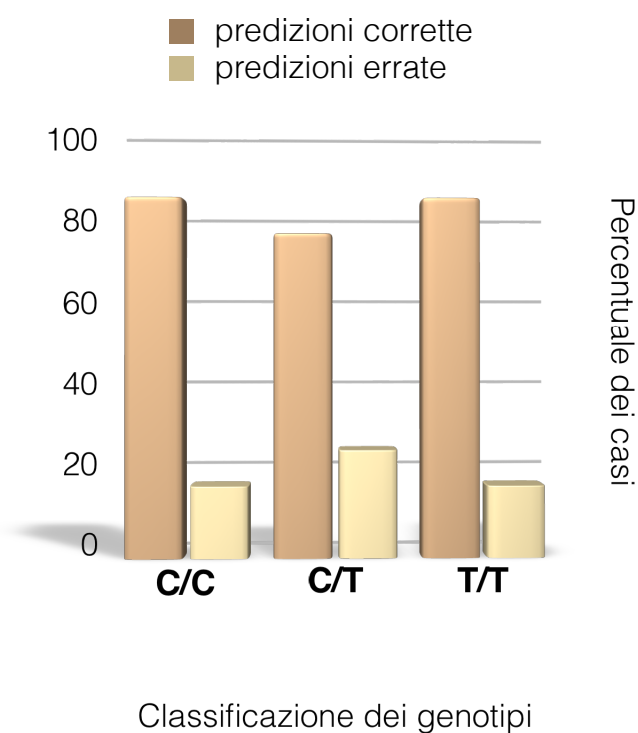
L'accuratezza nella genotipizzazione degli SNP, ottenuta attraverso il sequenziamento e l'inferenza con HMM, è risultata essere pari, in media, al 99.8% e al 94.5%, nei campioni aventi, rispettivamente, una predizione del genotipo  $\beta$  talassemico fetale corretta ed errata. (tab. 10).

Nella figura 20 sono riportati i risultati complessivi della diagnosi prenatale non invasiva che è risultata essere corretta nell'80% (24/30) ed errata nel 20% (6/30) dei casi.

Il grafico della figura 21, che mostra la distribuzione dei risultati nelle tre categorie di genotipo fetale atteso, non evidenzia la presenza di differenze significative nell'esito della predizione; infatti, l'errata predizione dell'aplotipo materno è stata equamente riscontrata nei tre gruppi di campioni.



**Figura 20.** Percentuali di predizioni del genotipo  $\beta$ -talassemico fetale corrette ed errate.



**Figura 21.** Distribuzione dei risultati della predizione del genotipo fetale nelle tre categorie.

Inoltre, da tali risultati, si evince che l'elevata profondità di sequenziamento, ottenuta attraverso l'amplificazione degli SNPs informativi, potrebbe essere uno dei parametri

che condiziona in misura maggiore la rilevazione di varianti alleliche ereditate dal feto per via paterna e non condivise con il genoma materno. Inoltre, che la valutazione della presenza/assenza anche di una sola di queste varianti, può essere sufficiente per attribuire al feto l'ereditarietà dell'aplotipo paterno corretto.

Al contrario, la predizione dell'aplotipo materno è resa più complessa a causa della preponderante quantità di DNA materno presente nel campione di cffDNA sottoposto a sequenziamento. Per questo motivo, l'inferenza degli alleli ereditati dal feto per via materna, deve essere eseguita applicando dei modelli d'analisi più complessi che valutano la significatività statistica degli sbilanciamenti allelici osservati, nel cffDNA, a livello dei siti materni informativi. La presenza di sbilanciamenti allelici diversi da quelli attesi, che in 6 campioni ha portato alla predizione errata, potrebbe essere stata causata, come accennato in precedenza, da un bias del processo di amplificazione, dalla presenza di una minore frazione fetale o, ancora, dall'errata ricostruzione degli aplotipi materni di riferimento.

La pipeline automatizzata ha consentito di gestire rapidamente l'analisi dei dati di sequenziamento e di ottenere l'esito di ciascuna diagnosi nell'arco di poche ore (3-4). Nella figura 14 è illustrato il workflow d'analisi che viene avviato dall'operatore nel momento in cui inserisce nella pipeline i file BAM prodotti dal sequenziamento con la piattaforma Ion Torrent PGM. L'ottimizzazione delle performance della pipeline è stata ottenuta attraverso l'esecuzione di cicli d'analisi ripetuti che hanno consentito di individuare i parametri e i *cut-off* migliori per l'analisi dei dati e la creazione dei report finali, dei quali viene dato un esempio nelle figure 22 e 23 e 24.

```

PATERNAL PREDICTION...
FETAL FRACTION: 0.074

-----
PATERNAL SITES EXIST!!!
rs11549407 not present in paternal site!!!
-----
10 percentile cutoff for selection paternal: 0.992618629174
-----
POS    wild    variants    mother    left    right    fraction    DEPTH    hap_predict
4    5215621    63    1644    1    0    1    0.963    1718    0
9    5228140    4058    2    0    0    1    1.000    4132    0
11   5236808    11    4876    1    1    0    0.998    4893    0
15   5237362    7    6235    1    1    0    0.999    6250    0
16   5237597    37    14453    1    1    0    0.997    14528    0
17   5237622    26    14510    1    1    0    0.998    14546    0
18   5237626    43    14492    1    1    0    0.997    14548    0
19   5237759    89    9848    1    1    0    0.991    9926    1
21   5238497    14    12745    1    1    0    0.999    12919    0
22   5239228    25    16929    1    1    0    0.999    16957    0
23   5239346    1    10089    1    1    0    1.000    10096    0
25   5241552    44    7970    1    1    0    0.995    8020    0
26   5242347    5    4885    1    1    0    0.999    4906    0
27   5242541    16446    1    0    0    1    1.000    16457    0
28   5243699    146    10898    1    1    0    0.987    11075    1
29   5245498    214    20799    1    1    0    0.990    20923    1
30   5245731    8    9175    1    1    0    0.999    9184    0
32   5247567    23    10743    1    1    0    0.998    10851    0
-----
START HMM!!!
Observed: [0 0 0 0 0 0 0 1 0 0 0 0 0 0 1 1 0 0]
Best path:[0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0]
Perfect match for Haplo A!!
NUMBER of PATERNAL site:18
-----
POS    REF    ALT    padre    madre    paternal_prediction    info
4    5215621    C    T    0|1    1|1    0    left
9    5228140    C    T    0|1    0|0    0    left
11   5236808    T    A    1|0    1|1    1    left
15   5237362    C    T    1|0    1|1    1    left
16   5237597    G    C    1|0    1|1    1    left
17   5237622    G    A    1|0    1|1    1    left
18   5237626    T    A    1|0    1|1    1    left
19   5237759    C    A    1|0    1|1    1    left
21   5238497    G    C    1|0    1|1    1    left
22   5239228    T    C    1|0    1|1    1    left
23   5239346    T    C    1|0    1|1    1    left
25   5241552    A    C    1|0    1|1    1    left
26   5242347    C    G    1|0    1|1    1    left
27   5242541    G    C    0|1    0|0    0    left
28   5243699    A    G    1|0    1|1    1    left
29   5245498    C    T    1|0    1|1    1    left
30   5245731    T    C    1|0    1|1    1    left
32   5247567    G    A    1|0    1|1    1    left
-----
-----
-----

```

**Figura 22.** Predizione sito per sito degli alleli paterni e inferenza dell'aplotipo paterno.

```

MATERNAL PREDICTION...
FETAL FRACTION: 0.074
-----
MATERNAL SITES EXIST!!!
HETEROZIGOUS SITES EXIST!!!
-----
POS      ID REF ALT padre madre DEPTH wild variants wild% var%
38 5226774 rs11549407_b39 G A 0|1 0|1 11189 5522 5632 49.5 50.5
-----

PROBS: P1 0.53690685413
PROBS: P2 0.5
-----

POS pater left right var_wild var_var left_ right_ DEPTH
0 5224973 1 0 1 49.8 50.2 6e-03 5e-16 12889
1 5225126 0 1 0 48.4 51.6 6e-03 2e-03 2195
2 5226774 0 0 1 49.5 50.5 9e-13 4e-03 11189
3 5247910 1 0 1 49.6 50.4 8e-03 3e-09 6679
4 5251917 1 0 1 48.2 51.8 5e-05 4e-05 7816
5 5252457 1 0 1 47.6 52.4 8e-04 7e-03 2496
6 5258792 1 0 1 53.6 46.4 2e-10 1e-02 6833
7 5263748 1 0 1 53.6 46.4 2e-14 8e-03 10126
-----

START HMM!!!
Observed: [0 0 1 0 0 1 1 1]
Best path:[0 0 0 0 0 0 0 0]
-----

Perfect match for Haplo A!!
NUMBER of MATERNAL site:8
-----
POS REF ALT padre madre maternal_prediction info
0 5224973 A C 1|1 0|1 0 left
1 5225126 T A 0|0 1|0 1 left
2 5226774 G A 0|1 0|1 0 left
3 5247910 C A 1|0 0|1 0 left
4 5251917 C G 1|0 0|1 0 left
5 5252457 T G 1|0 0|1 0 left
6 5258792 G C 1|0 0|1 0 left
7 5263748 A G 1|0 0|1 0 left
-----

```

**Figura 23.** Predizione sito per sito degli alleli materni e inferenza dell'aplotipo materno.

```

-----
sample name: 27
FETAL FRACTION:7.4%
NUMBER of PATERNAL sites:18
NUMBER of MATERNAL sites:8

PATERNAL HAPLOTYPE A - ACCURACY: 100.0% || - linked to beta_39 allele 0
PATERNAL HAPLOTYPE B - ACCURACY: 100.0% || - linked to beta_39 allele 1
MATERNAL HAPLOTYPE A - ACCURACY: 99.8% || - linked to beta_39 allele 0
MATERNAL HAPLOTYPE B - ACCURACY: 99.8% || - linked to beta_39 allele 1

PIPELINE PREDICTION:
FETAL HAPLOTYPE PREDICTED: paternal 0 maternal 0
FETAL GENOTYPE PREDICTED: G/G
STATUS: NORMAL

VALIDATION:
FETAL GENOTYPE: G/G
PATERNAL:0 MATERNAL: 0
STATUS: NORMAL

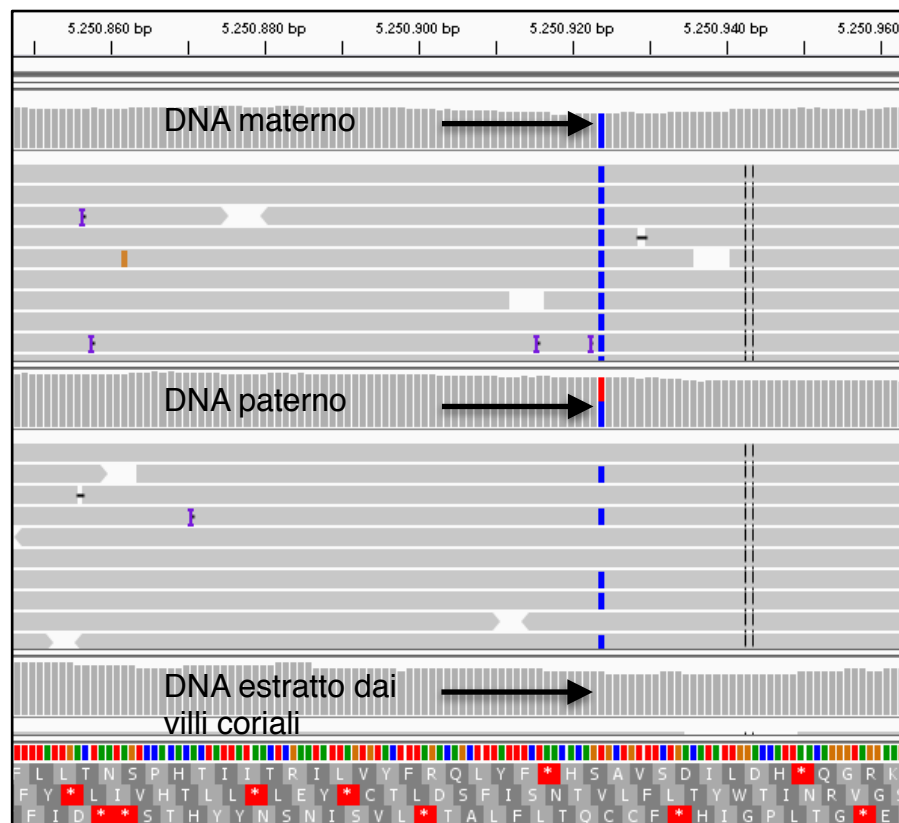
DIAGNOSI CORRETTA
-----

```

**Figura 24.** Esempio di report finale.

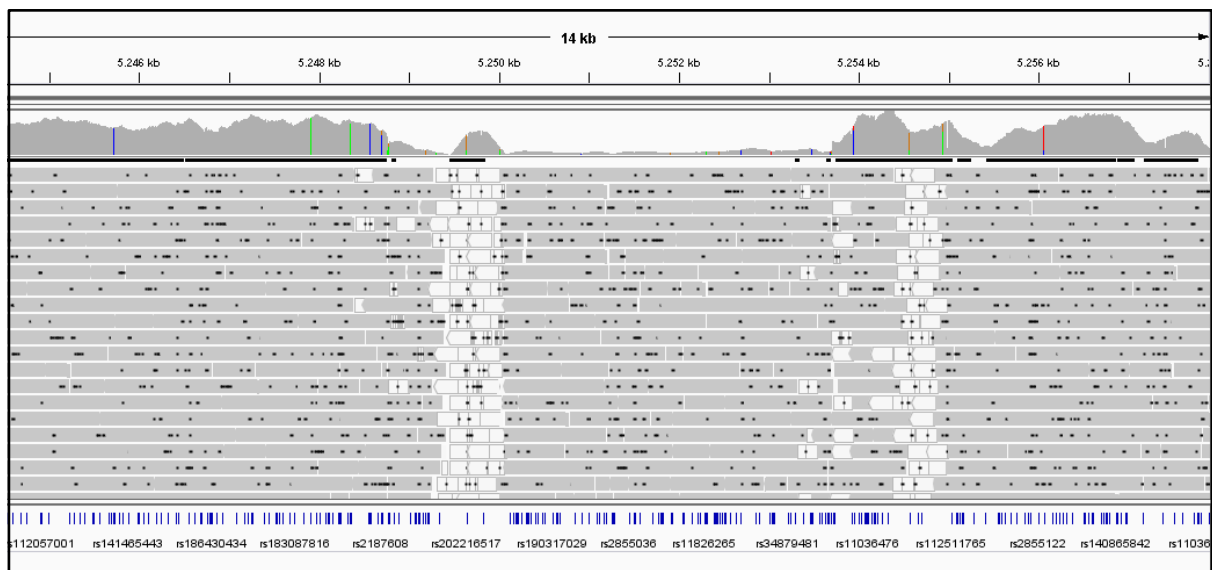
Nell'ambito di questo studio sono stati sequenziati, attraverso piattaforma Ion Torrent PGM, 39 trio sardi per la creazione di un pannello di aplotipi di riferimento utilizzato dalla pipeline per la ricostruzione degli aplotipi parentali.

I sistemi di sequenziamento NGS consentono di individuare la presenza sia di variazioni di un singolo nucleotide che di estese delezioni e complessi riarrangiamenti genici. Questo è quanto accaduto per uno dei campioni da noi sequenziati. Analizzando, infatti, i file VCF di uno dei 39 trio è stata osservata, a livello di alcuni SNPs localizzati tra i geni *HBG1* (gene  $\text{A}\gamma$ -globinico) e *HBG2* (gene  $\text{G}\gamma$ -globinico), un' incompatibilità tra il genotipo dei genitori e quello del DNA estratto dal trofoblasto. Tale incompatibilità è stata facilmente evidenziata anche analizzando i corrispondenti file BAM del trio con il software IGV (fig.25).



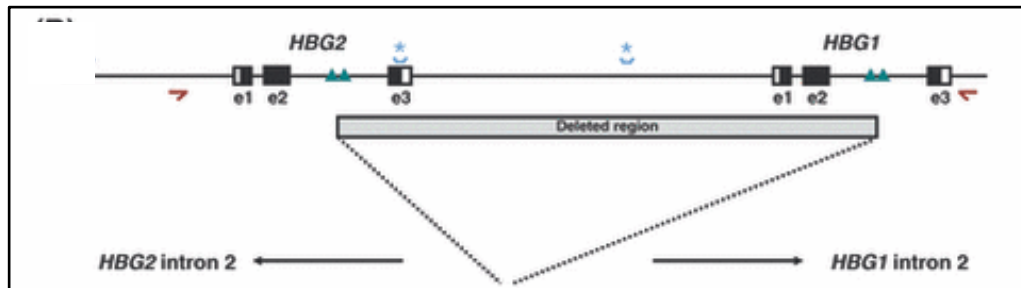
**Figura 25.** Visualizzazione su IGV dell'incompatibilità di genotipo a livello di uno SNP.

Visualizzando tali file ad una risoluzione inferiore, che consente di osservare una regione genica più estesa, si osserva la presenza, nel campione materno e in quello corrispondente al trofoblasto fetale, di una notevole diminuzione del coverage in un'area compresa tra i geni  $\gamma$ -globinici (fig.26). Tali evidenze hanno suggerito la probabile presenza di una delezione in eterozigosi trasmessa al feto per via materna.



**Figura 26.** Visualizzazione su IGV ad una risoluzione inferiore che evidenzia il calo del coverage a livello di un'ampia regione compresa tra i geni  $\gamma$ -globinici.

Per confermare tale ipotesi è stata eseguita una gap-PCR utilizzando una coppia di primers proposta in un precedente lavoro (Lee et al. 2009) nel quale viene descritta una delezione di circa 4,9 kb tra l'IVS2 di *HBG2* e l'IVS2 di *HBG1* che determina la formazione di un gene di fusione *HBG2-HBG1* costituito dall'esone 1 e 2 di *HBG2* e dall'esone 3 di *HBG1* (fig.27). L'elevata omologia di sequenza tra i due geni gamma globinici rende particolarmente complicato individuare l'esatto breakpoint della delezione. Tuttavia si ritiene che la fusione tra i due geni si verifichi in corrispondenza delle repeat polimorfiche (TG)<sub>n</sub>(CG)<sub>m</sub> all'interno dell'IVS2.



**Figura 27.** Rappresentazione della delezione a livello dei geni  $\gamma$ -globinici (da Lee et al. 2009).

La gap-PCR eseguita sui nostri campioni, in cui è stata ipotizzata la presenza della delezione, ha consentito di ottenere un frammento di circa 2,2 kb poi sequenziato con la metodica di Sanger. La sequenza ottenuta ha confermato la presenza della stessa delezione descritta nel lavoro di Lee et al. Tuttavia, mentre il caso descritto in questo studio presenta un'alterazione dei parametri ematologici ed, in particolare, una concentrazione elevata di emoglobina fetale, nel nostro caso tali parametri risultano nella norma.

Il cffDNA corrispondente a questo campione è stato analizzato con successo dalla pipeline suggerendo che la presenza della delezione ereditata dal feto non impedisce di eseguire una corretta predizione del genotipo fetale.

## 5. Conclusioni

La diagnosi prenatale di malattie genetiche viene attualmente eseguita nelle gravidanze a rischio mediante procedure invasive, a cui è associata una percentuale di rischio di aborto stimata intorno allo 0,5-2%.

In seguito all'identificazione, nel 1997, della presenza di DNA fetale libero (cffDNA) nella circolazione periferica materna, sono stati condotti numerosi studi volti allo sviluppo di protocolli di diagnosi prenatale non invasiva. Successivamente a questa scoperta alcuni gruppi di ricerca hanno focalizzato i loro studi sulla descrizione delle caratteristiche specifiche del cffDNA. Tali ricerche hanno consentito di scoprire che il DNA fetale libero ha origine da processi apoptotici del sinciziotrofoblasto e che la sua concentrazione nel plasma materno aumenta al progredire dell'epoca gestazionale e in relazione ad alcune patologie (preeclampsia, aneuploidie fetali etc.); inoltre il cffDNA viene rapidamente rimosso dal circolo materno subito dopo il parto, evitando di inficiare una eventuale diagnosi in una gravidanza successiva.

Data la sua natura frammentaria e la sua ridotta concentrazione rispetto al DNA libero di origine materna, l'analisi del cffDNA richiede l'utilizzo di approcci altamente sensibili ed accurati.

Inizialmente le applicazioni erano limitate unicamente alla valutazione della presenza/assenza di alleli non condivisi col genoma materno, come i tratti genetici ereditati per via paterna. Pertanto le prime applicazioni hanno riguardato la determinazione non invasiva del sesso fetale e la genotipizzazione del fattore RhD, che sono entrate a far parte da alcuni anni della pratica clinica in diversi Paesi europei, come il Regno Unito.

L'evoluzione delle tecnologie di Next Generation Sequencing ha consentito il rapido sviluppo dei test non invasivi per l'identificazione delle più comuni aneuploidie fetali, oggi disponibili in commercio da parte di numerose company private.

Al contrario l'analisi del cfDNA per l'identificazione delle malattie monogeniche risulta essere più complicata.

L'applicazione più semplice riguarda la detection di mutazioni *de novo*, come nel caso dell'acondroplasia, o della mutazione paterna nelle malattie autosomiche dominanti ereditate per via paterna. La diagnosi non invasiva di acondroplasia e displasia tanatofora fanno attualmente parte della pratica clinica nel Regno Unito. L'esclusione della mutazione paterna risulta essere una delle applicazioni più semplici anche per la NIPD di malattie monogeniche a trasmissione autosomica recessiva nel caso di coppie a rischio in cui i genitori siano portatori di mutazioni differenti. In questo caso, l'assenza della mutazione paterna consente di escludere la patologia, mentre la sua presenza si traduce in una probabilità del 50% che il feto sia affetto.

Nonostante le difficoltà nella determinazione dell'allele materno, negli ultimi anni sono stati pubblicati alcuni studi relativi alla diagnosi prenatale non invasiva di malattie autosomiche recessive in cui esso viene correttamente identificato, utilizzando approcci basati sullo studio degli aplotipi, come nel caso della  $\beta$ -talassemia (Lo et al. 2010, Lo et al. 2012) e dell'iperplasia surrenale congenita (Lo et al. 2014, Ma et al. 2014).

L'obiettivo del nostro studio è stato quello di sviluppare un protocollo per effettuare una predizione del genotipo fetale analizzando il cfDNA in coppie in cui i genitori sono entrambi portatori della  $\beta^{39}$ , mutazione che, a causa dell'elevata

dell'omogeneità genetica della popolazione sarda, costituisce il 95,7% degli alleli patologici.

In questo caso non è quindi sufficiente limitare l'analisi al singolo sito di mutazione, ma essa deve essere estesa ad un elevato numero di SNPs al fine di poter determinare gli aplotipi parentali ereditati dal feto.

A tale scopo lo studio si è basato sul sequenziamento NGS, nel cffDNA, di regioni target del cluster  $\beta$ -globinico, e sull'inferenza degli aplotipi parentali più probabilmente ereditati dal feto attraverso l'utilizzo di una pipeline automatizzata.

Nella prima fase dello studio è stato costruito un pannello di aplotipi di riferimento, costituito dagli aplotipi di 1000 Genomi e dagli aplotipi di 39 Trio sardi (ciascuno costituito dai DNA parentali e dal DNA estratto dai villi coriali) con l'obiettivo di incrementare la rappresentazione della  $\beta^{*39}$ , variante considerata rara tra i campioni di 1000 Genomi. A tal fine sono stati sequenziati, mediante piattaforma Ion Torrent PGM, i DNA di 39 Trio sardi in una regione di 62,7 kb del cluster  $\beta$ -globinico.

Analizzando i dati di sequenziamento sono stati selezionati gli SNPs informativi nei DNA parentali, che sono stati successivamente amplificati e sequenziati in 37 campioni di cffDNA, attraverso lo strumento Ion Torrent PGM. I DNA parentali sono stati analizzati mediante una pipeline automatizzata che in una prima fase ricostruisce gli aplotipi parentali analizzando i genitori sulla base del pannello di aplotipi di riferimento precedentemente costruito e considerandoli indipendentemente, senza le informazioni del DNA estratto dal trofoblasto. Successivamente essa analizza i dati del sequenziamento del cffDNA effettuando una predizione sito per sito degli alleli ereditati, seguita dall'inferenza degli aplotipi parentali trasmessi al feto con maggiore probabilità, sfruttando il modello statistico HMM. Dei 37 campioni di cffDNA, ne sono stati analizzati dalla pipeline 30,

escludendo quei campioni in cui l'impossibilità di calcolare la quota fetale o l'assenza di SNPs informativi ne ha impedito l'analisi.

L'approccio utilizzato ha permesso di ottenere una corretta predizione del genotipo  $\beta$ -talassemico fetale nel 80% dei campioni analizzati. La detection dell'aplotipo paterno è stata correttamente effettuata nel 100% dei casi, mentre l'aplotipo materno è stato identificato correttamente nell'80% dei casi. Infatti nel 20% dei campioni, in cui la predizione del genotipo fetale è risultata non corretta, la causa dell'errore è stata rintracciata nella errata determinazione dell'aplotipo materno.

La detection dell'aplotipo paterno si basa sulla valutazione della presenza/assenza dell'allele paterno in ciascun sito analizzato nel campione di cffDNA. Il protocollo da noi sviluppato si è rivelato un sistema altamente sensibile ed efficiente nella determinazione dell'aplotipo paterno in quanto, grazie all'elevato coverage, è possibile rilevare la presenza dell'allele non condiviso con la madre anche in presenza di una bassa quota fetale e di pochi SNPs.

La detection dell'aplotipo materno risulta invece più complicata in quanto legata alla valutazione dello sbilanciamento allelico a livello di siti in cui la madre ha genotipo eterozigote, ed è basata sul calcolo della probabilità che l'allele ereditato per via materna sia identico o diverso da quello ereditato per via paterna. Non sempre lo sbilanciamento osservato è coerente con quello atteso e l'accumulo di sbilanciamenti erronei porta l'HMM ad inferire l'aplotipo non realmente ereditato, come è accaduto in 5 dei campioni in cui la predizione dell'aplotipo materno è risultata errata.

Tali sbilanciamenti errati potrebbero essere riconducibili ad eventi legati alla PCR, come una minore efficienza nell'amplificazione, l'interferenza tra diversi ampliconi o ancora all'amplificazione preferenziale di un allele rispetto all'altro. Inoltre non è stato

possibile escludere la presenza, nei campioni analizzati, di un eventuale *vanishing twin* che potrebbe giustificare gli sbilanciamenti errati che hanno portato ad una predizione del genotipo fetale non corretta.

Una possibile causa di errore nella predizione può anche essere attribuita ad una costruzione degli aplotipi parentali non accurata, che può avere come conseguenza la predizione di un aplotipo associato all'allele non ereditato. Per esempio in uno dei campioni con predizione errata, la ricostruzione non corretta dell'aplotipo materno ha causato un fallimento nella diagnosi. Incrementare il pannello di aplotipi di riferimento inserendo un maggiore numero di campioni consentirebbe una ricostruzione più accurata delle fasi parentali.

Inoltre il potere predittivo e statistico può essere incrementato, soprattutto nella determinazione dell'aplotipo materno, includendo un più alto numero di SNPs analizzabili per ciascuna famiglia. Purtroppo nel nostro studio è stato necessario escludere dall'amplificazione numerosi SNPs localizzati in regioni ad elevata omologia del cluster  $\beta$ -globinico.

In questo studio è stata descritta una piattaforma in grado di predire in modo rapido ed accurato il genotipo fetale utilizzando dei modelli già descritti in letteratura ed applicati al sequenziamento dell'intero genoma fetale. Tali modelli sono stati da noi adattati allo studio di una malattia monogenica sfruttando il target sequencing della regione di interesse in una casistica di campioni piuttosto ampia non ancora descritta per la NIPD della  $\beta$ -talassemia. L'analisi dei dati di sequenziamento mediante la pipeline ha consentito di rendere più rapida ed automatizzata la gestione dei dati, limitando l'intervento dell'operatore all'esecuzione di un unico comando. Per migliorare ulteriormente l'efficienza della piattaforma sarà necessario ampliare il pannello di aplotipi di riferimento con un più alto numero di campioni sardi,

aumentare il numero di SNPs analizzabili e standardizzare maggiormente i passaggi sperimentali del flusso di lavoro. Inoltre sarà necessario ricorrere ad un sistema che permetta di superare i bias dovuti all'amplificazione iniziale che, come precedentemente descritto, può portare a sbilanciamenti erranei. Un sistema basato sulla selezione dei frammenti di interesse attraverso un processo di ibridazione con sonde molecolari altamente specifiche, in alternativa alla selezione di tali regioni attraverso la PCR, consentirebbe di superare questo limite della piattaforma. Nel prossimo futuro sarà inoltre fondamentale estendere l'analisi a un maggior numero di campioni di cffDNA, ampliando la casistica, in modo da verificare la riproducibilità dei risultati ottenuti e la validazione del metodo.

In conclusione questo approccio sviluppato per la  $\beta$ -talassemia, basato sul target sequencing attraverso piattaforma Ion Torrent PGM e sull'analisi statistica mediante HMM, risulta essere un valido strumento che può essere potenzialmente applicato anche alla diagnosi di altre malattie monogeniche.

## Bibliografia

Alberry M, Maddocks D, Jones M, Abdel Hadi M, Abdel-Fattah S, Avent N, Soothill PW. Free fetal DNA in maternal plasma in anembryonic pregnancies: confirmation that the origin is the trophoblast. *Prenat Diagn.* 2007;27:415-8.

Al Nakib M, Desbrière R, Bonello N, Bretelle F, Boubli L, Gabert J, Levy-Mozziconacci A. Total and Fetal Cell-Free DNA Analysis in Maternal Blood as Markers of Placental Insufficiency in Intrauterine Growth Restriction. *Fetal Diagn Ther.* 2009;26:24–28.

Amicucci P, Gennarelli M, Novelli G, Dallapiccola B. Prenatal diagnosis of myotonic dystrophy using fetal DNA obtained from maternal plasma. *Clin Chem* 2000;46:301-2.

Ashoor G, Poon L, Syngelaki A, Mosimann B, Nicolaides KH. Fetal fraction in maternal plasma cell-free DNA at 11-13 weeks' gestation: effect of maternal and fetal factors. *Fetal Diagn Ther* 2012;31:237-43.

Ashoor G, Syngelaki A, Poon LC, Rezende JC, Nicolaides KH. Fetal fraction in maternal plasma cell-free DNA at 11-13 weeks' gestation: relation to maternal and fetal characteristics. *Ultrasound Obstet Gynecol* 2013;41:26-32.

Barrett AN, McDonnell TCR, Allan Chan KC, Chitty LS. Digital PCR analysis of maternal plasma for non-invasive detection of sickle cell anemia. *Clin Chem* 2012; 58: 1026-32.

Boon EM, Faas BH. Benefits and limitations of whole genome versus targeted approaches for noninvasive prenatal testing for fetal aneuploidies. *Prenat Diagn* 2013;33:563-8.

Bustamante-Aragones A, Gallego-Merlo J, Trujillo-Tiebas MJ, de Alba MR, Gonzalez-Gonzalez C, Glover G, Diego-Alvarez D, Ayuso C, Ramos C. New strategy for the prenatal detection/exclusion of paternal cystic fibrosis mutations in maternal plasma. *J Cyst Fibros*. 2008a;7:505-10.

Bustamante- Aragonés A, Pérez-Cerdà C, Pérez B, de Alba MR, Ugarte M, Ramos C. Prenatal diagnosis in maternal plasma of a fetal mutation causing propionic acidemia. *Mol Genet Metab*. 2008b;95:101-3.

Bustamante-Aragones A, Trujillo-Tiebas MJ, Gallego-Merlo J, Rodriguez de Alba M, Gonzalez-Gonzalez C, Cantalapiedra D, Ayuso C, Ramos C. Prenatal diagnosis of Huntington disease in maternal plasma: direct and indirect study. *Eur J Neurol*. 2008c;15:1338-44.

Canick JA, Kloza EM, Lambert-Messerlian GM, Haddow JE, Ehrich M, van den Boom D, Bombard AT, Deciu C, Palomaki GE. DNA sequencing of maternal plasma to identify Down syndrome and other trisomies in multiple gestations. *Prenat Diagn* 2012;32:730-4.

Canick JA, Palomaki GE, Kloza EM, Lambert-Messerlian GM, Haddow JE. The impact of maternal plasma DNA fetal fraction on next generation sequencing tests for common fetal aneuploidies. *Prenat Diagn* 2013;33:667-74.

Chan KC, Ding C, Gerovassili A, Yeung SW, Chiu RW, Leung TN, Lau TK, Chim SS, Chung GT, Nicolaides KH, Lo YM. Hypermethylated RASSF1A in maternal plasma: A universal fetal DNA marker that improves the reliability of noninvasive prenatal diagnosis. *Clin Chem*. 2006;52:2211-8.

Chan KC, Zhang J, Huy AB, Wong N, Lau TK, Leung TN, Lo KW, Huang DWS, Lo YM. Size distributions of maternal and fetal DNA in maternal plasma. *Clin Chem*. 2004;50:88-92.

Chen S, Ge H, Wang X, Pan X, Yao X, Li X, Zhang C, Chen F, Jiang F, Li P, Jiang H, Zheng H, Zhang L, Zhao L, Wang W, Li S, Wang J, Wang J, Yang H, Li Y, Zhang X. Haplotype-assisted accurate non-invasive fetal whole genome recovery through maternal plasma sequencing. *Genome Med.* 2013;5:18

Chen XQ, Stroun M, Magnenat JL, Nicod LP, Kurt AM, Lyautey J, Lederrey C, Anker P. Microsatellite alterations in plasma DNA of small cell lung cancer patients. *Nat. Med.* 1996; 2:1033-35.

Chim SS, Tong YK, Chiu RW, Lau TK, Leung TN, Chan LY, Oudejans CB, Ding C, Lo YM. Detection of the placental epigenetic signature of the maspin gene in maternal plasma. *Proc Natl Acad Sci U S A.* 2005;102:14753-8.

Chitty LS, Chatelain P, Wolffenbuttel KP, Aigrain Y. Prenatal management of disorders of sex development. *J Pediatr Urol.* 2012;8:576-584.

Chitty LS, Finning K, Wade A, Soothill P, Martin B, Oxenford K, Daniels G, Massey E. Diagnostic accuracy of routine antenatal determination of fetal RHD status across gestation: population based cohort study. *BMJ.* 2014;349:g5243.

Chitty LS, Griffin DR, Meaney C, Barrett A, Khalil A, Pajkrt E, Cole TJ. New aids for the non-invasive prenatal diagnosis of achondroplasia: dysmorphic features, charts of fetal size and molecular confirmation using cell-free fetal DNA in maternal plasma. *Ultrasound Obstet Gynecol.* 2011 Mar;37:283-9.

Chiu RW, Lau TK, Cheung PT, Gong ZQ, Leung TN, Lo YM. Noninvasive prenatal exclusion of congenital adrenal hyperplasia by maternal plasma: a feasibility study. *Clin Chem* 2002; 48:778-780.

Clausen FB. Integration of noninvasive prenatal prediction of fetal blood group into clinical prenatal care. *Prenat Diagn.* 2014;34:409-15.

Colin Y, Chérif-Zahar B, Le Van Kim C, Raynal V, Van Huffel V, Cartron JP. Genetic basis of the RhD-positive and RhD-negative blood group polymorphism as determined by Southern analysis. *Blood*. 1991;78:2747-52.

Daniels G, Finning K, Martin P, Soothill P. Fetal blood group genotyping from DNA from maternal plasma: an important advance in the management and prevention of haemolytic disease of the fetus and newborn. *Vox Sang* 2004;87:225–32.

Devaney SA, Palomaki GE, Scott JA, Bianchi DW. Noninvasive fetal sex determination using cell-free fetal DNA: a systematic review and meta-analysis. *JAMA*. 2011; 306:627-636.

Everett TR and Chitty LS. Cell-free fetal DNA: the new tool in fetal medicine. *Ultrasound Obstet Gynecol* 2015; 45: 499–507

Fan HC, Quake SR. Detection of aneuploidy with digital polymerase chain reaction. *Anal Chem*. 2007;79:7576-79.

Farina A, LeShane ES, Romero R, Gomez R, Chaiworapongsa T, Rizzo N, Bianchi DW. High levels of fetal cell-free DNA in maternal serum: a risk factor for spontaneous preterm delivery. *Am J Obstet Gynecol*. 2005;193:421-5.

Ferres MA, Hui L, Bianchi DW. Antenatal non-invasive DNA testing: clinical experience and impact. *Am J Perinatol*. 2014;31:577-82.

Finning K, Martin P, Daniels G. A clinical service in the UK to predict fetal Rh (Rhesus) D blood group using free fetal DNA in maternal plasma. *Ann N Y Acad Sci* 2004;1022:119-23.

Flori E, Doray B, Gautier E, Kohler M, Ernault P, Flori J, Costa JM. Circulating cell-free fetal DNA in maternal serum appears to originate from cyto- and syncytiotrophoblastic cells. Case report. *Hum Reprod*. 2004;19:723-4.

Galbiati S, Brisci A, Lalatta F, Seia M, Makrigiorgos GM, Ferrari M, Cremonesi L. Full COLD-PCR protocol for noninvasive prenatal diagnosis of genetic diseases. *Clin Chem*. 2011;57:136-8.

Galbiati S, Stenirri S, Sbaiz L, Barberis M, Cremonesi L, Restagno G, Ferrari M. Identification of an 18 bp deletion in the TWIST1 gene by CO-amplification at lower denaturation temperature-PCR (COLD-PCR) for non-invasive prenatal diagnosis of craniosynostosis: first case report. *Clin Chem Lab Med*. 2014 Apr;52:505-9.

González-González MC, García-Hoyos M, Trujillo MJ, Rodríguez de Alba M, Lorda-Sánchez I, Díaz-Recasens J, Gallardo E, Ayuso C, Ramos C. Prenatal detection of a cystic fibrosis mutation in fetal DNA from maternal plasma. *Prenat Diagn*. 2002;22:946-8.

González-González MC, Trujillo MJ, Rodríguez de Alba M, García-Hoyos M, Lorda-Sánchez I, Díaz-Recasens J, Ayuso C, Ramos C. Huntington disease-unaffected fetus diagnosed from maternal plasma using QF-PCR. *Prenat Diagn*. 2003a; 23:232-4.

González-González MC, Trujillo MJ, Rodríguez de Alba M, Ramos C. Early Huntington disease prenatal diagnosis by maternal semiquantitative fluorescent-PCR. *Neurology*. 2003b;60:1214-5.

Gupta AK, Holzgreve W, Huppertz B, Malek A, Schneider H, Hahn S. Detection of fetal DNA and RNA in placenta-derived syncytiotrophoblast microparticles generated in vitro. *Clin Chem*. 2004;50:2187-90.

Haghiac M, Vora NL, Basu S, Johnson KL, Presley L, Bianchi DW, Hauguel De Mouzon S. Increased death of adipose cells, a path to release cell-free DNA into systemic circulation of obese women. *Obesity (Silver Spring)* 2012; 20:2213-9.

Hahn S, Rusterholz C, Hosli I, Lapaire O. Cell-free nucleic acids as potential markers for preeclampsia. *Placenta*. 2011;32(Suppl.):S17-20.

Hill M, Finning K, Martin P, Hogg J, Meaney C, Norbury G, Daniels G, Chitty LS. Non-invasive prenatal determination of fetal sex: translating research into clinical practice. *Clin Genet*. 2011; 80:68-75.

Hill M, Twiss P, Verhoef TI, Drury S, McKay F, Mason S, Jenkins L, Morris S, Chitty LS. Non-invasive prenatal diagnosis for cystic fibrosis: detection of paternal mutations, exploration of patient preferences and cost analysis. *Prenat Diagn*. 2015;35:950-8.

Kitzman JO, Snyder MW, Ventura M, Lewis AP, Qiu R, Simmons LE, Gammill HS, Rubens CE, Santillan DA, Murray JC, Tabor HK, Bamshad MJ, Eichler EE, Shendure J. Noninvasive Whole-Genome Sequencing of a Human Fetus. *Sci Transl Med*. 2012;4:137ra76.

Knight M, Redman CW, Linton EA, Sargent IL. Shedding of syncytiotrophoblast microvilli into the maternal circulation in pre-eclamptic pregnancies. *Br J Obstet Gynaecol*. 1998; 105: 632-40.

Kolialexi A, Tounta G, Apostolou P, Vrettou C, Papantoniou N, Kanavakis E, Antsaklis A, Mavrou A. Early non-invasive detection of fetal Y chromosome sequences in maternal plasma using multiplex PCR. *Eur J Obstet Gynecol Reprod Biol*. 2012;161:34-7.

Kotsopoulou I, Tsoplou P, Mavrommatis K, Kroupis C. Non-invasive prenatal testing (NIPT): limitations on the way to become diagnosis. *Diagnosis* 2015; 2: 141–158.

Lam KW, Jiang P, Liao GJ, Chan KC, Leung TY, Chiu RW, Lo YM. Noninvasive prenatal diagnosis of monogenic diseases by targeted massively parallel sequencing of maternal plasma: application to  $\beta$ -thalassemia. *Clin Chem*. 2012;58:1467-75.

Legler TJ, Müller SP, Haverkamp A, et al. Prenatal RhD testing: a review of studies published from 2006 to 2008. *Transfus Med Hemother* 2009;36:189–98.

Levine RJ, Qian C, Leshane ES, Yu KF, England LJ, Schisterman EF, Wataganara T, Romero R, Bianchi DW. Two stage elevation of cell-free fetal DNA in maternal sera before onset of preeclampsia. *Am J Obstet Gynecol.*2004;190:707-13.

Lim JH, Kim MJ, Kim SY, Kim HO, Song MJ, Kim MH, Park SY, Yang JH, Ryu HM. Non-invasive prenatal detection of achondroplasia using circulating fetal DNA in maternal plasma. *J Assist Reprod Genet.* 2011 28:167-72.

Lo YM, Chan KC, Sun H, Chen EZ, Jiang P, Lun FM, Zheng YW, Leung TY, Lau TK, Cantor CR, Chiu RW. Maternal plasma DNA sequencing reveals the genome-wide genetic and mutational profile of the fetus. *Sci Transl Med.* 2010;2:61ra91.

Lo YM, Corbetta N, Chamberlain PF, Rai V, Sargent IL, Redman CW, Wainscoat JS. Presence of fetal DNA in maternal plasma and serum. *Lancet* .1997; 350:485-7.

Lo YM, Lun FM, Chan KC, Tsui NB, Chong KC, Lau TK, Leung TY, Zee BC, Cantor CR, Chiu RW. Digital PCR for the molecular detection of fetal chromosomal aneuploidy. *Proc Natl Acad Sci U S A.* 2007;104:13116-21.

Lo YM, Tein MS, Lau TK, Haines CJ, Leung TN, Poon PM, Wainscoat JS, Johnson PJ, Chang AM, Hjelm NM. Quantitative analysis of fetal DNA in maternal plasma and serum: implications for noninvasive prenatal diagnosis. *Am J Hum Genet.* 1998;62:768-75.

Lo YM, Zhang J, Leung TN, Lau TK, Chang AM, Hjelm NM. Rapid clearance of fetal DNA from maternal plasma. *Am J Hum Genet.* 1999; 64:218-24.

Lui YY, Chik KW, Chiu RW, Ho CY, Lam CW, Lo YM. Predominant hematopoietic origin of cell-free DNA in plasma and serum after sex-mismatched bone marrow transplantation. *Clin Chem* 2002; 48:421-7.

Lun FMF, Chiu RWK, Chan KCA, Leung TY, Lau TK, Lo YMD. Microfluidics Digital PCR reveals a higher than expected fraction of fetal DNA in maternal plasma. *Clin Chem*. 2008a; 54:1664-72

Lun FM, Tsui NB, Chan KC, Leung TY, Lau TK, Charoenkwan P, Chow KC, Lo WY, Wanapirak C, Sanguansermisri T, Cantor CR, Chiu RW, Lo YM. Noninvasive prenatal diagnosis of monogenic diseases by digital size selection and relative mutation dosage on DNA in maternal plasma. *Proc Natl Acad Sci U S A*. 2008b 105:19920-5.

Ma D, Ge H, Li X, Jiang T, Chen F, Zhang Y, Hu P, Chen S, Zhang J, Ji X, Xu X, Jiang H, Chen M, Wang W, Xu Z. Haplotype-based approach for noninvasive prenatal diagnosis of congenital adrenal hyperplasia by maternal plasma DNA sequencing. *Gene*. 2014;544:252-8.

Mandel P, Métais P. Les acides nucléiques du plasma sanguin chez l'homme. *C R Acad Sci Paris*. 1948; 142: 241-243.

Meng M, Li X, Ge H, Chen F, Han M, Zhang Y, Kang D, Xie W, Gao Z, Pan X, Dai P, Chi F, Chen S, Liu P, Zhang C, Cao J, Jiang H, Xu X, Wang W, Duan T. Noninvasive prenatal testing for autosomal recessive conditions by maternal plasma sequencing in a case of congenital deafness. *Genet Med*. 2014;16:972-6.

Merriman B; Ion Torrent R&D Team, Rothberg JM. Progress in ion torrent semiconductor chip based sequencing. *Electrophoresis*. 2012;33:3397-417.

Masala M, Saba L, Zoppi MA, Puddu R, Picciau A, Capponi V, Iuculano A, Monni G, Rosatelli MC. Pitfalls in noninvasive fetal RhD and sex determination due to a vanishing twin. *Prenat Diagn*. 2015;35:506-8.

Masuzaki H, Miura K, Yoshiura KI, Yoshimura S, Niikawa N, Ishimaru T. Detection of cell free placental DNA in maternal plasma: direct evidence from three cases of confined placental mosaicism. *J Med Genet*. 2004;41:289-92.

Meaney C, Norbury G. Non-invasive prenatal diagnosis of early onset primary dystonia I in maternal plasma. *Prenat Diagn* 2009; 29:1218-21.

Mujezinovic F, Alfirovic Z. Procedure-related complications of amniocentesis and chorionic villous sampling: a systematic review. *Obstet Gynecol.* 2007;110:687-694.

Nawroz H, Koch W, Anker P, Stroun M, Sidransky D. Microsatellite alterations in serum DNA of head and neck cancer patients. *Nat Med.* 1996; 2:1035-37.

New MI, Tong YK, Yuen T, Jiang P, Pina C, Chan KC, Khattab A, Liao GJ, Yau M, Kim SM, Chiu RW, Sun L, Zaidi M, Lo YM. Noninvasive Prenatal Diagnosis of Congenital Adrenal Hyperplasia Using Cell-Free Fetal DNA in Maternal Plasma. *J Clin Endocrinol Metab.* 2014;99:E1022-30.

Page-Christiaens GCML. Bossers B, van der Schoot CE, de Haas M. Use of bi-allelic insertion/deletion polymorphisms as a positive control for fetal genotyping in maternal blood: first clinical experience. *Ann N Y Acad Sci* 2006; 1075:123-9.

Pertl B1, Sekizawa A, Samura O, Orescovic I, Rahaim PT, Bianchi DW. Detection of male and female fetal DNA in maternal plasma by multiplex fluorescent polymerase chain reaction amplification of short tandem repeats. *Hum Genet.* 2000;106:45-9.

Poon LL, Leung TN, Lau TK, Chow KC, Lo YM. Differential DNA methylation between fetus and mother as a strategy for detecting fetal DNA in maternal plasma. *Clin Chem.* 2002; 48: 35–41.

Rava RP, Sehnert AJ, Bianchi DW. Circulating fetal cell-free DNA fractions in autosomal aneuploides and monosomy X. *Clin Chem.* 2014;60:243-50.

Rosatelli MC, Dozy A, Faa V, Meloni A, Sardu R, Saba L, Kan YW, Cao A. Molecular Characterization of  $\beta$ -Thalassemia in the Sardinian Population. *Am J Hum Genet* 1992. 50:422-426.

Saito H, Sekizawa A, Morimoto T, Suzuki M, Yanaihara T. Prenatal DNA diagnosis of a single-gene disorder from maternal plasma. *Lancet* 2000;356:1170.

Scheffer PG, van der Schoot CE, Page-Christiaens GC, Bossers B, van Erp F, de Haas M. Reliability of fetal sex determination using maternal plasma. *Obstet Gynecol* 2010; 115:117-26.

Spencer K. Aneuploidy screening in the first trimester. *Am J Med Genet C Semin Med Genet* 2007;145C:18-32.

Tsui NB, Kadir RA, Chan KC, Chi C, Mellars G, Tuddenham EG, Leung TY, Lau TK, Chiu RW, Lo YM. Noninvasive prenatal diagnosis of hemophilia by microfluidics digital PCR analysis of maternal plasma DNA. *Blood*. 2011;117:3684-91.

Tynan Ja, Mahboudi P, Cagasan LL, van den Boom D, Ehrich M, Oeth P. Restriction enzyme-mediated enhanced detection of circulating cell-free fetal DNA in maternal plasma. *J Mol Diagn*. 2011;13:382-9.

Vlková B, Hodosy J. Vanishing twin as a potential source of bias in noninvasive fetal sex determination: a case report. *J Obstet Gynaecol Res* 2014;40:1128–31.

Wald NJ, Rodeck C, Hackshaw AK, Walters J, Chitty L, Mackinson AM. First and second trimester antenatal screening for Down's syndrome: the results of the Serum, Urine and Ultrasound Screening Study (SURUSS). *J Med Screen* 2003;10:56–104.

Wang E, Batey A, Struble C, Musci T, Song K, Oliphant A. Gestational age and maternal weight effects on fetal cell-free DNA in maternal plasma. *Prenat Diagn* 2013;33:662-6.

White HE, Dent CL, Hall VJ, Crolla JA, Chitty LS. Evaluation of a novel assay for detection of the fetal marker RASSF1A: facilitating improved diagnostic reliability of non-invasive prenatal diagnosis. *PLoS One* 2012;7:e45073.

Wikman AT, Tiblad E, Karlsson A, Olsson ML, Westgren M, Reilly M. Noninvasive single-exon fetal RHD determination in a routine screening program in early pregnancy. *Obstet Gynecol.* 2012;120(2 Pt 1):227-34.

Xu Y, Li X, Ge HJ, Xiao B, Zhang YY, Ying XM, Pan XY, Wang L, Xie WW, Ni L, Chen SP, Jiang WT, Liu P, Ye H, Cao Y, Zhang JM, Liu Y, Yang ZJ, Chen YW, Chen F, Jiang H, Ji X. Haplotype-based approach for noninvasive prenatal tests of Duchenne muscular dystrophy using cell-free fetal DNA in maternal plasma. *Genet Med.* 2015;17:889-96.

Yu SC, Chan KC, Zheng YW, Jiang P, Liao GJ, Sun H, Akolekar R, Leung TY, Go AT, van Vugt JM, Minekawa R, Oudejans CB, Nicolaides KH, Chiu RW, Lo YM. Size-based molecular diagnostics using plasma DNA for noninvasive prenatal testing. *Proc Natl Acad Sci U S A.* 2014;111:8583-8.

Yu SC, Lee SW, Jiang P, Leung TY, Chan KC, Chiu RW, Lo YM. High-resolution profiling of fetal DNA clearance from maternal plasma by massively parallel sequencing. *Clin Chem.* 2013; 59:1228-37.

Zejskova L, Jancuskova T, Kotlabova K, Doucha J, Hromadnikova I. Feasibility of fetal-derived hypermethylated RASSF1A sequence quantification in maternal plasma--next step toward reliable non-invasive prenatal diagnostics. *Exp Mol Pathol.* 2010;89:241-7.