



**UNIVERSITÀ DEGLI STUDI DI CAGLIARI**

**Facoltà di Medicina e Chirurgia**

Dipartimento di Sanità Pubblica, Medicina Clinica e Molecolare

Sezione di Scienze Biomediche e Biotecnologie

**Dottorato di Ricerca**

**“Terapia Pediatrica e Farmacologia dello Sviluppo”**

Ciclo XXVI

**Diagnosi molecolare della Distrofia Muscolare di  
Duchenne e della Neurofibromatosi tipo 1 con la  
tecnologia di sequenziamento *Next Generation***

Settore scientifico disciplinare di afferenza

MED/03

Tesi di Dottorato della: Dott.<sup>SSa</sup> Roberta Chessa

Coordinatore del corso: Prof.<sup>RE</sup> Paolo Moi

Tutor: Dott.<sup>SSa</sup> Maria Addis

A.A. 2012-2013

La presente tesi è stata prodotta durante la frequenza del corso di dottorato in Terapia Pediatrica e Farmacologia dello Sviluppo dell'Università degli Studi di Cagliari, A.A. 2012/2013 - XVI ciclo, con il supporto di una borsa di studio finanziata con le risorse del P.O.R. SARDEGNA F.S.E. 2007-2013 - Obiettivo competitività regionale e occupazione, Asse IV Capitale umano, Linea di Attività 1.3.1

# Indice:

<b>Introduzione</b> .....	pag. 4
<b>Le Distrofinopatie</b> .....	pag. 4
<b>La Distrofia Muscolare di Duchenne</b> .....	pag. 5
<i>Manifestazioni cliniche DMD</i> .....	pag. 5
<i>Il gene DMD</i> .....	pag. 6
<i>La Distrofina</i> .....	pag. 7
<i>Criteri diagnostici</i> .....	pag. 9
<i>Patologia molecolare</i> .....	pag. 9
<i>Diagnosi molecolare</i> .....	pag. 10
<b>Le Neurofibromatosi</b> .....	pag. 11
<b>Neurofibromatosi tipo 1</b> .....	pag. 12
<i>Criteri diagnostici</i> .....	pag. 13
<i>Il gene NF1</i> .....	pag. 13
<i>La Neurofibromina</i> .....	pag. 14
<i>Patologia molecolare</i> .....	pag. 15
<i>Diagnosi molecolare</i> .....	pag. 16
<b>Il sequenziamento del DNA</b> .....	pag. 17
<i>Il sequenziamento di Sanger</i> .....	pag. 17
<i>Tecnologie di seconda generazione Next Generation</i> .....	pag. 18
<i>Tecnologie di Terza Generazione</i> .....	pag. 22
<i>Le applicazioni della NGS</i> .....	pag. 23

<b>Obiettivi della tesi</b> .....	pag. 24
<b>Materiali &amp; Metodi</b> .....	pag. 25
<i>Pazienti</i> .....	pag. 25
<i>Estrazione del DNA</i> .....	pag. 25
<i>Sequenziamento con il sistema Ion PGM</i> .....	pag. 26
<i>Disegno dei primers</i> .....	pag. 26
<i>Preparazione della libreria</i> .....	pag. 27
<i>Amplificazione del target</i> .....	pag. 28
<i>Digestione parziale delle sequenze dei primers</i> .....	pag. 29
<i>Ligazione degli adattatori agli ampliconi e purificazione</i> .....	pag. 29
<i>Purificazione della libreria non amplificata</i> .....	pag. 30
<i>Quantificazione della libreria amplificata: Bioanalyzer 2100 (Agilent)</i> .....	pag. 30
<i>Equalizzazione</i> .....	pag. 32
<i>Preparazione del Template</i> .....	pag. 32
<i>Sequenziamento</i> .....	pag. 34
<i>Analisi dei dati</i> .....	pag. 36
<i>Strumenti bioinformatici per l'analisi delle varianti</i> .....	pag. 40
<i>Sequenziamento con metodo di Sanger</i> .....	pag. 44
<b>Risultati</b> .....	pag. 45
<b>Discussione</b> .....	pag. 57
<b>Bibliografia</b> .....	pag. 62



## **Introduzione:**

Le malattie neuromuscolari, come anche le neuropatie e le miopatie, sono malattie ereditarie rare e spesso gravi che colpiscono muscoli e nervi in età pediatrica od insorgere in età adulta .

L'eterogeneità genetica e clinica, caratteristiche di queste patologie, è dovuta al coinvolgimento di molti geni strutturali, spesso di grandi dimensioni, che necessitano di essere studiati con diversi metodi di biologia molecolare con conseguente aumento dei tempi della diagnosi. Attualmente l'applicazione del sequenziamento massivo-parallelo, chiamato anche sequenziamento di nuova generazione, potrebbe essere di aiuto nella identificazione di mutazioni nelle malattie genetiche rare come la Distrofia muscolare di Duchenne o la Neurofibromatosi di tipo 1 consentendo una strategia diagnostica rapida e veloce.

## **Le Distrofinopatie:**

Le distrofinopatie sono un gruppo di malattie muscolari legate all'X recessive, che presentano fenotipi che possono variare dall'aumento asintomatico della concentrazione della creatinfosfochinasi (CK) ad alterazioni muscolari progressive gravi che sono classificate come Distrofie muscolari di Duchenne e Becker (DMD e BMD) quando sono coinvolti i muscoli scheletrici, e, come Cardiomiopatia

Dilatativa (XLDC) quando è coinvolto primariamente il cuore.

Sono patologie, dovute alla mancanza o riduzione della distrofina nei muscoli scheletrici e cardiaco, causate da mutazioni del gene DMD.

### **La Distrofia Muscolare di Duchenne:**

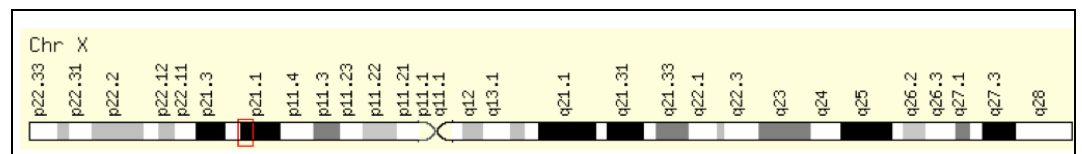
La DMD è la distrofinopatia clinicamente più grave con un esordio precoce, la cui frequenza è di 1/3.300 neonati maschi con una prevalenza di 150-200 casi-anno.

#### ***Manifestazioni Cliniche DMD:***

I primi segni clinici si manifestano intorno al terzo anno di vita con difficoltà nel camminare, correre, salire le scale od alzarsi da terra (manovra di Gower). L'evoluzione della DMD è grave ed infausta, con perdita della deambulazione intorno alla seconda decade di vita e la manifestazione di una grave insufficienza cardiorespiratoria con esito letale nel giovane adulto.

## ***Il gene DMD:***

Il gene della distrofina è, il più grande gene finora descritto, lungo 2,9 Mb in posizione Xp21.2. È costituito da 79 esoni e produce un trascritto primario full-length di 14.3 Kb [Mandel, 1989; Manole, 1995] (*Fig. 1*).



*Fig. 1* Localizzazione del gene DMD sul cromosoma X (Xp21.2);

Il gene DMD produce almeno 7 trascritti alternativi, generati da 3 tipi di processi molecolari:

- Utilizzo di almeno 8 diversi promotori situati a monte ed all'interno del gene, spesso tessuto-specifici;
- Splicing alternativo;
- Utilizzo di diversi segnali di poli-adenilazione;

L'espressione dei trascritti full-length è controllata da 3 promotori regolati in modo indipendente: Brain (B), Muscle (M) e Purkinje (P), i cui nomi riflettono i maggiori siti d'espressione della distrofina. Trascritti più corti, che producono isoforme tronche della distrofina sono regolati da 4 promotori interni indicati come Retinal (R), Brain-3 (B3), Schwann cell (S), e general.

Studi funzionali hanno permesso di ipotizzare che, a livello muscolare, il gene della distrofina umana venga regolato da

almeno 4 enhancers (*Dystrophin Muscle Enhancers*, DME 1, 2, 3 e 4). DME 1, 2, 3 sono situati in una regione di 14 Kb; 7 Kb a monte e 7 Kb a valle dell'esone 1 muscolare, mentre è stato ipotizzato che il DME 4 sia localizzato al di fuori di questa regione di 14 KbG.

### ***La Distrofina:***

La distrofina è una proteina citoscheletrica che funge da ponte tra l'apparato contrattile intracellulare e la matrice extra-cellulare.

È localizzata sul versante citoplasmatico della membrana plasmatica e svolge la funzione di mediare l'ancoraggio del citoscheletro di actina delle fibre muscolari striate della membrana basale, grazie al complesso proteico di membrana associato alla distrofina (*Dystrophin-Associated Protein Complex*, DAPC) (*Fig. 2*).

Uno dei ruoli principali di questo complesso proteico è quello di stabilizzare il sarcolemma e proteggere le fibre muscolari dai danni a lungo termine indotti dalla contrazione [Michalak *et al.*, 1997].

La sua mancanza comporta un difetto al supporto strutturale della membrana, rendendola più suscettibile a rotture durante l'attività contrattile.

La distrofina rappresenta il 3% delle proteine di membrana e lo 0.002% del totale delle proteine muscolari.

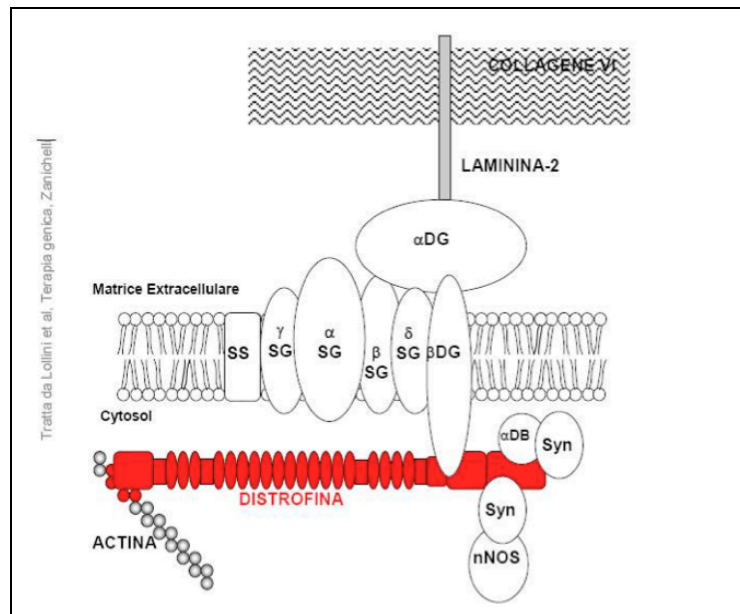


Fig. 2 Complesso glicoproteico associato alla distrofina (DAPC);

La Distrofina è costituita da 4 domini (Fig. 2):

- Il dominio ammino terminale [*Actin Binding Domain* (ABD)] (N-terminale);
- Il dominio centrale (*Central Rod Domain*), un'ampia regione bastoncellare a tripla elica, è formato da 24 ripetizioni, disposte in tandem, ciascuna di circa 109 aminoacidi;
- Il dominio ricco in cisteine (*Cysteine-rich domain*) di 280 aminoacidi, che contiene un sito di legame per il  $Ca^{2+}$  e siti di legame per i diversi componenti del complesso proteico associato alla distrofina;
- Il dominio carbossi terminale (*Carboxy-terminal domain*) di 420 aminoacidi, la cui struttura secondaria è una  $\alpha$ -elica avvolta a spirale (Blake *et al.*, 1995) e per questo chiamata regione CC (*coiled coil*). Il dominio CC è responsabile del legame della distrofina con la distrobrevina e la sintrofina [Blake *et al.*, 1995];

### ***Criteri diagnostici:***

La diagnosi di distrofinopatie si basa su:

- **Dosaggio della concentrazione di CK (Creatin chinasi) sierica:** nei pazienti con DMD è 10-20 volte la norma;
- **Biopsia muscolare:** che consente di appurare il quadro istopatologico, tipico di un processo distrofico con degenerazione, rigenerazione fibrale e sostituzione fibro-adiposa. Su questo tessuto è possibile studiare con metodiche immunohistochimiche e biochimiche la presenza od assenza delle distrofina;
- **Analisi molecolare:** del DNA che consente di individuare mutazioni nel gene DMD;

### ***Patologia molecolare:***

Le mutazioni sono rappresentate nel 70% dei casi da delezioni, di uno o più esoni, nel 5-10% da duplicazioni e nel restante 20-25% da piccole delezioni, mutazioni puntiformi, mutazioni dei siti di splicing e mutazioni introniche. Delezioni di estensione variabile che alterano il quadro di lettura (out of frame) dell'RNA determinano l'assenza di distrofina con conseguente quadro clinico grave, mentre delezioni che mantengono intatto il quadro di lettura (in frame) producono distrofine più corte e meno funzionali od in quantità ridotta e danno origine a fenotipi clinici più lievi.

Le mutazioni nel gene della distrofina sono responsabili oltre che della DMD e della BMD anche della Cardiomiopatia Dilatativa X linked (XLDC), una malattia del muscolo cardiaco caratterizzata dalla dilatazione ventricolare e dalla funzione sistolica ridotta.

Molte mutazioni, causanti tale patologia, sono delezioni che coinvolgono il promotore M e l'esone 1.

### ***Diagnosi Molecolare:***

La diagnosi molecolare si basa sull'utilizzo di due metodiche:

- MLPA;
- Sequenziamento DNA e RNA.

## **Le Neurofibromatosi:**

Le neurofibromatosi sono patologie neurocutanee, ereditate con modalità autosomica dominante. Sono caratterizzate dalla presenza di tumori multipli del tessuto nervoso. Sulla base dei criteri stabiliti dal *National Institute of Health* si definiscono due forme principali di neurofibromatosi:

**La Neurofibromatosi tipo 1:** (NF1) (MIM#162200) caratterizzata dalla presenza di numerose macchie caffè-latte, lentiginosi in aree ben determinate del corpo, noduli iridei di Lisch e presenza di tumori dei nervi periferici (Neurofibromi) che, a seconda della sede, si distinguono in cutanei, sottocutanei o nodulari. Il gene NF1, responsabile della malattia, è localizzato sul cromosoma 17 (17q11.2).

**La Neurofibromatosi tipo 2:** (NF2) (MIM#101000) è una forma piuttosto rara la cui frequenza è di 1:33000 nati, caratterizzata principalmente dalla presenza di tumori bilaterali del nervo acustico, detti schwannomi vestibolari. Il gene NF2, responsabile della patologia, è localizzato sul cromosoma 22 (22q12.2).



### ***Neurofibromatosi tipo 1:***

È una malattia neurocutanea multisistemica ereditaria, caratterizzata dalla predisposizione allo sviluppo di tumori benigni e maligni.

È una patologia autosomica dominante con frequenza di 1 su 4.000-5.000 nati vivi in tutte le popolazioni [Orphanet].

La penetranza è completa e la maggior parte dei segni clinici sono presenti nel 90% dei pazienti entro la pubertà [Mattocks *et al.*, 2004]. La variabilità fenotipica è intrafamiliarmente, infatti l'espressione fenotipica è estremamente variabile anche all'interno dello stesso nucleo familiare.

Le cause della variabilità sono dovute a diversi fattori quali:

- Background genetico: polimorfismi che modulano l'entità delle manifestazioni;
- Tipo di mutazione;
- Presenza di mosaicismi;
- Elevato tasso di mutazione spontanea. Il 50% dei soggetti affetti da NF1 non presenta una forma familiare della malattia ma una mutazione *de novo*.

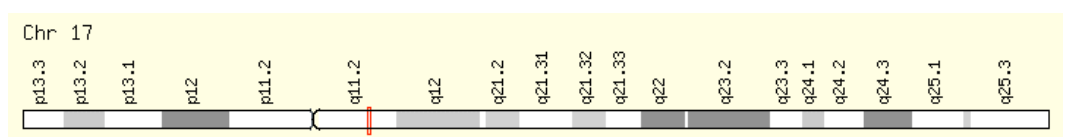
### ***Criteria diagnostici:***

La diagnosi della NF1 si basa sulla presenza di almeno tre delle seguenti manifestazioni cliniche:

- Presenza di sei o più macchie caffè-latte (*Cafè au lait spot*, CLS);
- Presenza di neurofibromi, ovvero tumori della guaina dei nervi periferici che si presentano sotto forma di lesioni cutanee, sottocutanee o plessiformi;
- Lentiggini nelle pieghe cutanee;
- Due o più noduli iridei di Lisch;
- Un glioma delle vie ottiche;
- Difetti scheletrici specifici (assottigliamento della corteccia delle ossa lunghe, displasia delle ali dello sfenoide);
- Un familiare di primo grado affetto.

### ***Il gene NF1:***

Il gene NF1 fu mappato mediante clonaggio posizionale, sul braccio lungo del cromosoma 17 (17q11.2) (*Fig. 3*).



*Fig. 3* Localizzazione del gene NF1 sul cromosoma 17 (17q11.2);

Il gene NF1 si estende su 350 kb ed è costituito da 58 esoni che codificano per un trascritto di 11-13 kb. Quest'ultimo comprende un *Open Reading Frame* (ORF) di 8.457 bp tradotto in una proteina, la neurofibromina.

### ***La Neurofibromina:***

La Neurofibromina è una proteina idrofilica citoplasmatica con elevata omologia di sequenza con GTPase activating proteins (GAP). L'espressione del mRNA risulta esser ubiquitario ma la proteina viene espressa soprattutto nel Sistema Nervoso Centrale e nelle cellule cromaffini del surrene. La proteina presenta una sequenza primaria di 2.818 amminoacidi ed una massa molecolare stimata in 327 kDa [Shen *et al.*, 1996] e mostra una attività GTPasi dovuta alla guanosina trifosfatasi in grado di legarsi a Ras p21 e modularne positivamente la conversione di guanosina trifosfato (GTP) a guanosina difosfato (GDP) (*Fig. 4*).

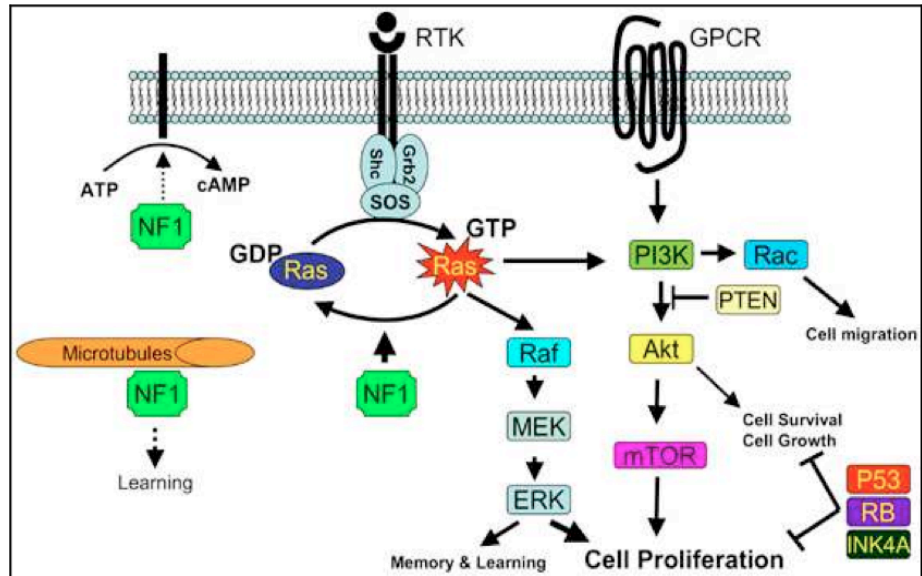


Fig. 4 Schema dell'attività GAP della Neurofibromina e delle diverse vie cellulari correlate;

La Neurofibromina, essendo una GAP-like protein è essenziale per la regolazione negativa di Ras, agendo, pertanto, da oncosoppressore; fisiologicamente esercita un controllo negativo sulla disponibilità di ras attivato e quindi sulla proliferazione cellulare.

### ***Patologia molecolare:***

Le mutazioni causanti la malattia sono rappresentate per il 90% da mutazioni puntiformi e per il 5 - 10% da delezioni del gene.

### ***Diagnosi molecolare:***

- Sequenziamento di DNA;
- MLPA.

Nelle due patologie prese in considerazione per questo studio le delezioni e le duplicazioni rappresentano il 70% dei difetti molecolari nella DMD/BMD e circa il 10 % nella NF1. In entrambi i casi, la metodica di elezione, è la MLPA (*Multiplex Ligation Probe Amplification*).

MLPA è una tecnica che si basa sull'ibridazione di più sonde specifiche amplificate contemporaneamente in un'unica PCR analizzata con elettroforesi capillare e analisi computerizzata dei picchi.

Nel 30% delle DMD e nel 90% delle NF1 il difetto molecolare è rappresentato da mutazioni puntiformi. La metodica utilizzata per la definizione di questo tipo di difetto è il sequenziamento del DNA/RNA.

## **Il sequenziamento del DNA:**

Le tecniche di sequenziamento del genoma umano nel corso di questi ultimi venti anni sono state sempre più perfezionate, standardizzate e velocizzate, rendendo il sequenziamento del DNA il gold standard per la ricerca delle mutazioni del genoma umano e per l'identificazione del gene-malattia. Il sequenziamento del DNA applicato alla diagnostica, consente di identificare le mutazioni responsabili delle malattie ereditarie e di sviluppare nuovi trattamenti mirati a compensare ed intervenire per cercare di correggere il difetto genetico.

I metodi di sequenziamento del DNA che si sono sviluppati dal 1973 [Maxam e Gilbert] a tutt'oggi hanno contribuito significativamente nel determinare la sequenza del DNA.

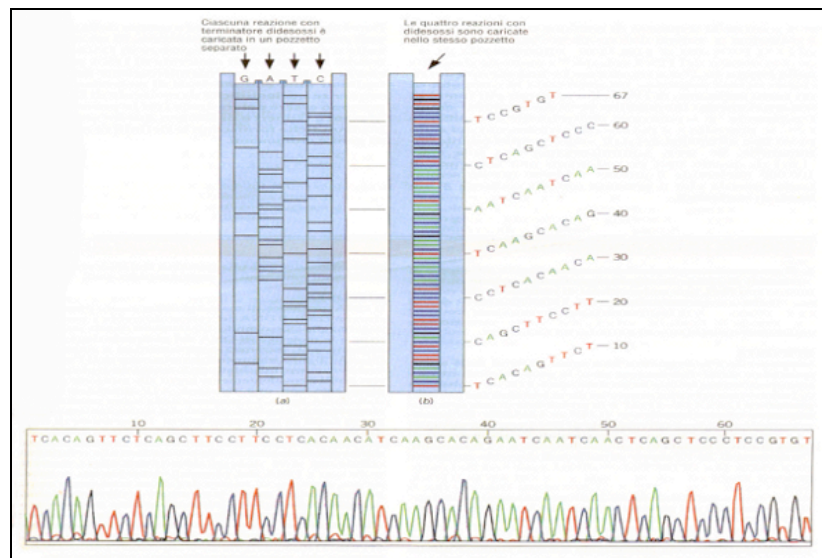
La rapidità del processo di sequenziamento è stato di grande aiuto per il sequenziamento su larga scala del genoma umano, consentendoci oggi di utilizzare tecniche di sequenziamento di terza generazione.

### ***Il sequenziamento di Sanger:***

Il metodo di sequenziamento Sanger ha segnato una svolta nel campo della biologia molecolare e Frederick Sanger, per questi suoi studi, ricevette il premio Nobel per la Chimica nel 1980.

La tecnica di sequenziamento ideata da di Frederick Sanger nel 1977 si basa sull'uso di dideossinucleotidi trifosfati (ddNTP) al posto dei normali deossinucleotidi trifosfati (dNTP), ciascuno dei quali marcato con un fluorocromo, per arrestare la sintesi del DNA in posizione 3'.

I frammenti così ottenuti si fanno migrare mediante elettroforesi capillare vengono separati e analizzati automaticamente dopo rivelazione laser (*Fig. 5*).



*Fig. 5* Elettroforesi capillare ed elettroferogramma;

### ***Tecnologie di Seconda Generazione Next Generation:***

Le tecniche di sequenziamento ad alta processività di seconda generazione permettono di generare e sequenziare migliaia o milioni di piccoli frammenti (*reads*) con alta velocità e migliore accuratezza.

Il target di sequenziamento può essere l'intero genoma, l'esoma, o porzioni di DNA più ristrette. Quest'ultimo approccio definito *targeted analysis* (analisi mirata) consiste nell'analisi di uno o più geni responsabili di una determinata patologia ad eterogeneità genetica.

Attualmente numerose società bio-tecnologiche hanno investito ingenti risorse per produrre sequenziatori sempre più potenti e sempre meno costosi con l'obiettivo di rendere più accessibile l'analisi del DNA (*Tab. 1*).

Platform	Illumina MiSeq	Ion Torrent PGM	PacBio RS	Illumina GAIIx	Illumina HiSeq 2000
Instrument Cost*	\$128 K	\$80 K**	\$695 K	\$256 K	\$654 K
Sequence yield per run	1.5-2Gb	20-50 Mb on 314 chip, 100-200 Mb on 316 chip, 1Gb on 318 chip	100 Mb	30Gb	600Gb
Sequencing cost per Gb*	\$502	\$1000 (318 chip)	\$2000	\$148	\$41
Run Time	27 hours***	2 hours	2 hours	10 days	11 days
Reported Accuracy	Mostly > Q30	Mostly Q20	<Q10	Mostly > Q30	Mostly > Q30
Observed Raw Error Rate	0.80 %	1.71 %	12.86 %	0.76 %	0.26 %
Read length	up to 150 bases	~200 bases	Average 1500 bases**** (C1 chemistry)	up to 150 bases	up to 150 bases
Paired reads	Yes	Yes	No	Yes	Yes
Insert size	up to 700 bases	up to 250 bases	up to 10 kb	up to 700 bases	up to 700 bases
Typical DNA requirements	50-1000 ng	100-1000 ng	~1 µg	50-1000 ng	50-1000 ng

\* All cost calculations are based on list price quotations obtained from the manufacturer and assume expected sequence yield stated.  
 \*\* System price including PGM, server, OneTouch and OneTouch ES.  
 \*\*\* Includes two hours of cluster generation.  
 \*\*\*\* Mean mapped read length includes adapter and reverse strand sequences. Subread lengths, i.e. the individual stretches of sequence originating from the sequenced fragment, are significantly shorter.

*Tab. 1* Specifiche tecniche delle piattaforme NGS;

Tra queste le più utilizzate sono:

- *Illumina;*
- *Life Technologies;*
- *Pacific Biosciences;*
- *Oxford Nanopore.*



Nonostante ciascuna delle piattaforme sia unica nella modalità di sequenziamento, alcune di loro (Illumina ed Ion Torrent) utilizzano una metodologia di base simile che comprende la preparazione del "*template*", sequenziamento, visualizzazione ed analisi bioinformatica dei dati.

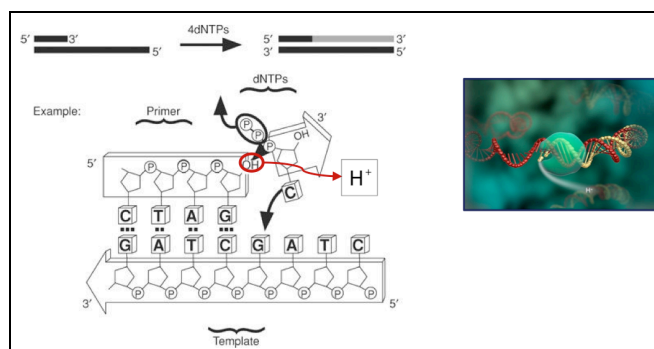
La piattaforma *Illumina/Solexa* utilizza un supporto fisso per immobilizzare i frammenti di DNA da sequenziare, incollandoli a un vetrino tramite degli adattatori. Questo supporto consente di leggere milioni di frammenti di DNA contemporaneamente e poiché ogni frammento si trova in un punto preciso e non può essere confuso con altri. La novità di questo metodo è l'uso di terminatori reversibili che possono essere riattivati, grazie all'azione di un enzima che taglia via la parte di molecola che blocca il lavoro della DNA polimerasi [Liu *et al.*, 2012; Quail *et al.*, 2012].

La piattaforma *AB SOLiD* immobilizza le molecole di DNA non su un supporto solido, ma su piccole sfere poste su un vetrino ed utilizza una DNA ligasi che unisce frammenti di DNA anziché sintetizzarne di nuovi come la DNA polimerasi. Inoltre, ha elaborato un sistema di sequenziamento che consente di leggere due volte lo stesso nucleotide riducendo di molto la possibilità di commettere errori [Liu *et al.*, 2012].

La piattaforma *Roche/454* non utilizza nucleotidi fluorescenti, ma il pirofosfato inorganico (Ppi) che si ottiene come prodotto di scarto della DNA polimerasi, che genera dei flash di luce ogni volta che un nucleotide nuovo viene aggiunto. Il protocollo consiste nel render disponibile

il PPI ad una sulfurilasi che lo utilizza per generare una molecola di ATP. L'ATP viene prelevato da una luciferasi e lo utilizza per ossidare una luciferina e produrre un segnale luminoso. Poiché questo segnale è sempre uguale per tutti i nucleotidi, essi devono essere immessi nel sistema uno alla volta in successione; prima le A, poi le C e così via [Liu *et al.*, 2012].

La piattaforma **Ion Torrent** della Life Technologies, come gli strumenti di nuova generazione precedentemente descritti, utilizza la DNA polimerasi, ma non misura i segnali luminosi, ma gli ioni idrogeno rilasciati durante la reazione di elongazione della molecola di DNA. La novità della Life Technologies consiste nel non avvalersi dell'ausilio di nucleotidi marcati e, pertanto, non necessita di costosi sistemi di rilevazione ottica. Le macchine della Ion Torrent utilizzano come supporto un materiale semiconduttore e sono in grado di leggere le sequenze di DNA misurando le variazioni di pH che vengono generate quando un nuovo nucleotide viene aggiunto al filamento sintetizzato (*Fig. 6*).



*Fig.6* Chimica della rilevazione del corretto nucleotide;

Questa tecnica di sequenziamento, ad opera del Personal Genome Machine (PGM), ha consentito una notevole riduzione dei costi dello strumento nel mercato [ioncommunity; Liu *et al.*, 2012; Quail *et al.*, 2012].

### ***Tecnologie di Terza Generazione:***

Le piattaforme ***Pacific Biosciences*** ed ***Oxford Nanopore*** appartengono alle tecniche di sequenziamento di terza generazione e rappresentano un vero e proprio salto di qualità e di potenzialità pur essendo attualmente poco più che dei prototipi. Sicuramente in un prossimo futuro ci si potrà avvalere di strumenti “*low-cost*” di piccole dimensioni, capaci di entrare con più facilità nella pratica medica di routine.

La ***Pacific Biosciences*** sta mettendo a punto un sequenziatore capace di sequenziare singole molecole di DNA. Riuscire a leggere una singola molecola di DNA monitorando in tempo reale l’aggiunta dei nucleotidi, significa poter usare piccole quantità di DNA ed avere anche un vantaggio economico, risparmiando sui reagenti.

La piattaforma ***Oxford Nanopore*** è in grado di leggere una sequenza di DNA basandosi sui cambiamenti nel flusso di corrente che attraversa dei fori di pochi nanometri, che sono caratteristici dello specifico nucleotide che viene aggiunto alla catena.

### ***Le applicazioni della NGS:***

Le applicazioni della NGS hanno consentito un rapido progresso in diversi settori della biologia molecolare. Nel campo della diagnostica clinica molecolare il sequenziamento massivo parallelo è stato estremamente efficace nella diagnosi di numerose patologie neuromuscolari che coinvolgono diversi geni (Vasli *et al.*, 2012) che possono essere studiati contemporaneamente in tempi rapidi.

Il risequenziamento del genoma umano ha consentito di identificare i geni e gli elementi regolatori coinvolti nei processi patologici. La NGS grazie al sequenziamento dell'intero genoma di diversi organismi oggi è applicabile anche nei campi della Salute Pubblica e dell'epidemiologia, perché consente di identificare nuovi fattori di virulenza attraverso il sequenziamento di specie batteriche e virali. Studi di espressione genica che utilizzano l'RNA-Seq (NGS-RNA) hanno incominciato a sostituire l'uso dell'analisi dei microarray con la visualizzazione dell'espressione dell'RNA sotto forma di sequenze.

## **Obiettivi della tesi:**

Oggi, differenti tecniche vengono utilizzate nella routine diagnostica per identificare mutazioni causative in geni responsabili di malattie genetiche. Il metodo più comune e che rappresenta il gold standard per l'identificazione di mutazioni puntiformi, è basato sull'amplificazione del DNA genomico tramite PCR delle sequenze codificanti del gene candidato seguito dal sequenziamento tramite metodo Sanger. Tuttavia, se il gene è di grandi dimensioni, come nel caso dei geni DMD e NF1, e le mutazioni sono distribuite sull'intero gene, il metodo Sanger è laborioso e richiede tempi lunghi.

Lo scopo di questo studio è stato quello di validare un protocollo per l'analisi delle sequenze dei geni DMD e NF1 con la tecnologia di sequenziamento di nuova generazione (NGS). Per il raggiungimento di questo obiettivo è stato sviluppato un protocollo di sequenziamento basato sulla tecnologia IonTorrent (Life Technologies) e applicato a soggetti a rischio per mutazioni puntiformi nei geni DMD o NF1.

Questo nuovo metodo di sequenziamento, rispetto al metodo Sanger, se accurato e preciso, permetterebbe di ridurre notevolmente i tempi di analisi e consente lo studio di più campioni contemporaneamente.

Queste caratteristiche rendono tale metodo utile per la sua applicazione nella diagnostica di routine in quanto consente di abbreviare i tempi di refertazione.

## **Materiali & Metodi:**

### ***Pazienti:***

Il *targeted Next generation Sequencing* dei geni DMD e NF1 è stato effettuato in:

- 10 pazienti di cui 6 maschi con sospetto clinico di distrofia muscolare di Duchenne/Becker (D/BMD) e 4 femmine con valori elevati di CPK (Gene DMD);
- 7 pazienti affetti da Neurofibromatosi di tipo 1 (Gene NF1);

In tutti i soggetti l'analisi attraverso Multiplex Ligation probe Amplification (MLPA), con Salsa kit P034 e P035 per il gene DMD e con Salsa Kit P081 e P082 per il gene NF1 (MRC-Holland) non ha messo in evidenza alcuna delezione o duplicazione.

In 11 campioni (8 DMD e 3 NF1) l'analisi è stata retrospettiva, infatti tali campioni erano stati precedentemente sequenziati con metodo Sanger.

### ***Estrazione del DNA:***

Il DNA genomico è stato estratto da campioni di sangue periferico con metodi convenzionali: purificazione con tecnologia delle particelle magnetiche (Kit Arrow Blood

DNA 500, NorDiag), estrazione salina, purificazione su colonna di silice (Perfect Pure DNA Blood kit, 5Prime).

### ***Sequenziamento con il sistema Ion PGM (Ion Torrent):***

Fasi del sequenziamento con il sistema Ion PGM:

- ✓ Disegno dei primers;
- ✓ Preparazione della libreria;
- ✓ Preparazione del Template;
- ✓ Sequenziamento;
- ✓ Analisi dei dati;

### ***Disegno dei primers:***

I primers per il sequenziamento dei geni DMD e NF1 sono stati disegnati e ordinati utilizzando il software Ion AmpliSeq Designer disponibile sul sito [www.ampliseq.com](http://www.ampliseq.com) (Life Technologies).

Ion AmpliSeq Designer fornisce il miglior set di primers possibili per ottenere il più alto coverage con il minor numero di ampliconi.

Ciascun Ion Ampliseq Custom primers contiene 2 pool di coppie di primers pre-aliquotati alla concentrazione 2X da utilizzare nella preparazione della libreria ed una piastra contenente i primers individuali, in TE alla concentrazione di 307  $\mu\text{M}$ . I primers contengono delle specifiche

modificazioni che ne permettono la rimozione durante la preparazione della libreria.

Per lo studio delle regioni esoniche, 5' UTR e 3' UTR (16,22Bb) del gene DMD sono state disegnate 169 coppie di primers, 85 contenute nel pool 1 e 84 nel pool 2. Il coverage era del 98,16%.

Per lo studio delle stesse regioni (13,34 Kb) del gene NF1 sono state disegnate 143 coppie di primers, 72 contenute nel pool 1 e 71 nel pool 2, con un coverage di 93,61%.

### ***Preparazione della libreria:***

La preparazione delle librerie è stata effettuata con il kit IonAmpliSeq Library Kit 2.0 (Life Technologies) e comprende i seguenti step (*Fig. 7*):

- Amplificazione del target;
- Digestione parziale delle sequenze dei primers;
- Ligazione degli adattatori agli ampliconi e purificazione;
- Purificazione della libreria non amplificata;
- Quantificazione della libreria: Bioanalyzer 2100 (Agilent);
- Equalizzazione;



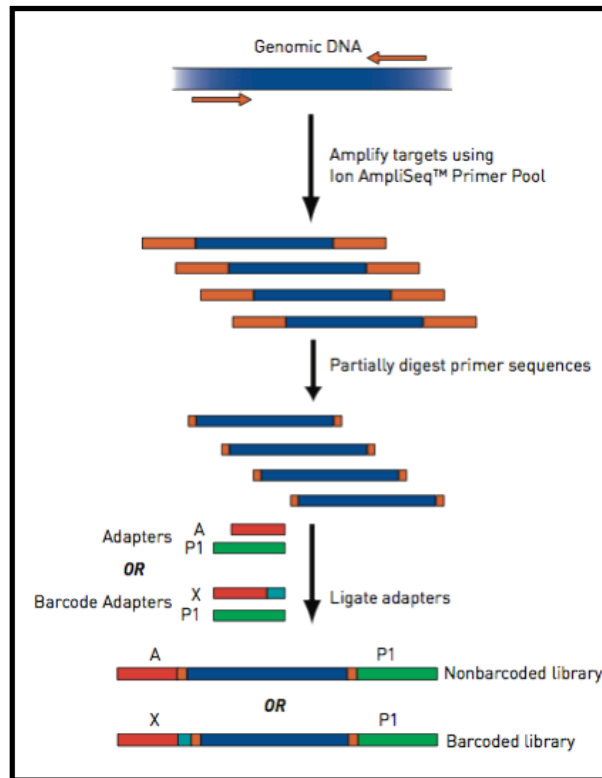


Fig. 7 Workflow della preparazione della libreria;

### ***Amplificazione del target:***

Come da protocollo, le regioni target sono state amplificate utilizzando 10 ng di DNA genomico. La concentrazione del DNA è stata determinata con lo strumento Qubit 2.0 Fluorometer e il kit Qubit dsDNA HS Assay (Life Technologies). Per ciascuna combinazione di Primer Pool e DNA si aggiungono, per reazione, i reagenti, riportati in tabella (Tab. 2). L'amplificazione è stata effettuata su termociclature GeneAmp PCR System 9700 alle seguenti condizioni: denaturazione a 99° per 15 sec, 19 cicli di *annealing* e *extension* a 60° per 4 min.

2X Primer Pools (Cat. nos. 4477685, 4477686, and Custom and Community Panels)		
Component	Volume without Sample ID Panel	Volume with Sample ID Panel
5X Ion AmpliSeq™ HiFi Master Mix (red cap)	4 µL	4 µL
2X Ion AmpliSeq™ Primer Pool	10 µL	10 µL
Optional: 20X Ion AmpliSeq™ Sample ID Panel	—	1 µL
gDNA, 10 ng	Y µL	Y µL
Nuclease-free Water	(6 - Y) µL	(5 - Y) µL
<b>Total</b>	<b>20 µL</b>	<b>20 µL</b>

Tab. 2 reagenti per l'amplificazione del target;

### ***Digestione parziale delle sequenze dei primers:***

Gli ampliconi ottenuti sono stati trattati con 2 µL di reagente FuPa per consentire la parziale digestione dei primers e la fosforilazione degli ampliconi. La reazione è stata condotta su termociclatore GeneAmp® PCR System 9700 alle seguenti: 50° per 10 min, 55° per 10 min e 60° per 20 min.

### ***Ligazione degli adattatori agli ampliconi e purificazione:***

Successivamente gli ampliconi sono stati ligati agli adattatori contenenti i barcode e purificati. I barcode permettono di assegnare a ciascun campione un tag di riconoscimento e mantenerne l'identità durante tutta la procedura, quando si analizzano più campioni contemporaneamente. Il kit utilizzato per il barcoding è Ion Xpress Barcode Adapters 1-16 (Life Technologies). Per ciascun barcode X scelto è stata preparata una mix composta da Ion P1 Adapter ed

Ion Xpress Barcode X nel rapporto di 1:4 per ciascun adattatore.

La reazione di ligazione viene effettuata combinando insieme la SwitchSolution, la mix dell'adattatore barcodato e 2 $\mu$ L di DNA Ligase. La reazione è stata eseguita su termociclatore GeneAmp PCR System 9700 alle seguenti condizioni: 22° per 30 min e 72° per 10 min.

### ***Purificazione della libreria non amplificata:***

La purificazione della libreria non amplificata è stata eseguita con il kit AMPure XP (Agencourt; Beckman) ed Etanolo al 70% (Carlo Erba). La purificazione permette di legare selettivamente gli ampliconi di dimensioni maggiori di 100bp, rimuovendo i primers, i nucleotidi, i sali e gli enzimi.

### ***Quantificazione della libreria: Bioanalyzer 2100 (Agilent):***

Immediatamente dopo lo step precedente, per 8 (4 DMD e 4 NF1) dei 20 campioni esaminati, abbiamo eseguito la quantificazione della libreria con lo strumento Bioanalyzer 2100 (Agilent). Prima della quantificazione le librerie sono state amplificate al fine di ottenere materiale sufficiente per una accurata quantificazione. L'amplificazione è stata effettuata aggiungendo al pellet ottenuto allo step precedente 50 $\mu$ L of Platinum PCR SuperMix High Fidelity

and 2 $\mu$ L of Library Amplification Primer Mix ed eseguendo la reazione nel termociclatore con il seguente programma: 98° per 2 min, 5 cicli a 98° per 5 sec e 64 ° per 1 min. Le librerie così amplificate sono state purificate con il kit AMPure XP. 1 $\mu$ L delle librerie amplificate sono state analizzate con il kit Agilent High Sensitivity DNA e lo strumento Bioanalyzer 2100 (Agilent). Le librerie degli ampliconi possiedono picchi multipli nel range tra le 125 e le 300 bp. Con l'ausilio del Programma 2100 Expert Software version B.02.08 si è determinato la concentrazione molare della libreria. La molarità viene calcolata automaticamente in pmol/mL (Fig. 8). Successivamente le librerie sono state diluite con TE in maniera tale da ottenere una concentrazione di  $\sim$ 100 pM.

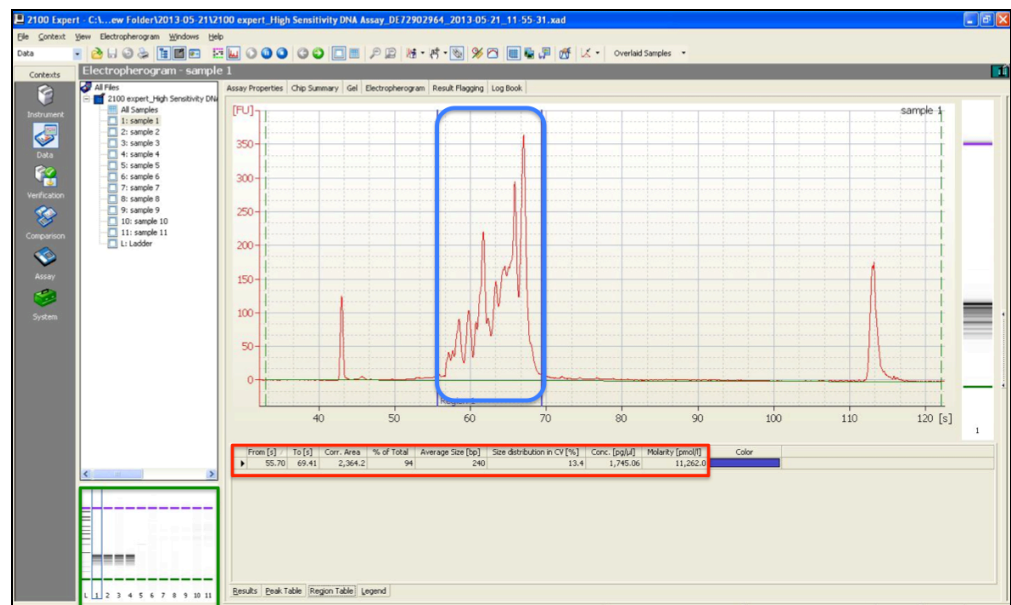


Fig. 8 Schermata del Bioanalyzer:

In celeste: picchi corrispondenti all'amplificazione del target;

In verde: Corsa del campione come se fosse in un gel;

In rosso: concentrazione della libreria;

### ***Equalizzazione:***

L'equalizzazione, eseguita sui restanti 12 campioni (6 DMD e 6 NF1), è stata effettuata con il kit Ion Library Equalizer (Life Technologies). Questo metodo permette di normalizzare le librerie ad una concentrazione di ~100 pM senza richiedere alcuna quantificazione. Il protocollo prevede l'amplificazione della libreria in termociclatore a 98° C per 2 min, 7 cicli di 98° C per 15 sec e 60 per 1 min, aggiungendo al pellet 50µL di Platinum PCR Super Mix High Fidelity e 2µL di Equalizer Primers. Alla libreria amplificata è stato aggiunto l'Equalizer Capture e successivamente l'Equalizer Beads. Dopo il lavaggio delle beads, la libreria equalizzata è stata eluita in 100 µL di Equalizer Elution Buffer per ottenere una concentrazione finale di 100pM.

### ***Preparazione del Template:***

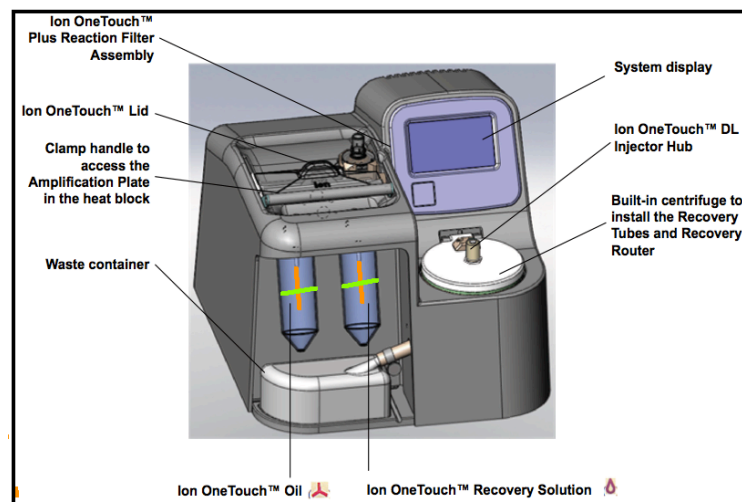
Le librerie quantificate sono state combinate per poter essere sequenziate in un unico chip e si è proceduto con la preparazione del *template*.

La preparazione del *template* avviene attraverso l'emulsion PCR, un'emulsione di acqua ed olio. I frammenti di DNA a singolo filamento si attaccano alla superficie delle biglie grazie agli adattatori. Le biglie sono compartimentate in un'emulsione di H<sub>2</sub>O ed olio che viene fatta sgocciolare in

questa emulsione; ciascuna goccia cattura una biglia e la PCR produce copie del singolo *template* di DNA.

Il *template* è stato ottenuto utilizzando il kit Ion OneTouch 200 Template v2. Il protocollo prevede la diluizione di 2 $\mu$ L della libreria totale avente una concentrazione di ~100 pM con 18  $\mu$ L acqua Nuclease Free. In tal modo la concentrazione della libreria di partenza sarà nel range stabilito dal protocollo pari a 8-10 pM o 1,2 ng/mL.

Per preparare il *template* è stato utilizzato lo strumento Ion One Touch (*Fig. 9*), nel quale l'emulsione di H<sub>2</sub>O e olio viene fatta passare in una piastra di amplificazione seguendo un percorso nel quale avviene l'emulsion PCR.



*Fig. 9* Ion OneTouch per la preparazione del *template*;

Lo strumento impiega 3 ore e 52 minuti per preparare il *template*. Al termine, il campione non arricchito è stato quantificato con il Qubit e il kit Ion Sphere Quality Control (life Technologies). Il range del campione non arricchito, stabilito dal protocollo deve essere inferiore al 30% (*Fig. 10*). Questo step non è richiesto per le librerie equalizzate.

Qubit Calibration Factor Calculation						
Calibration Standard	RFU	Calibration Factor				
Alexa Fluor® 488 Calibration Standard	17201,6	0,26				
Alexa Fluor® 647 Calibration Standard	59217,1					
Percent Templated ISPs						
Sample ID	Raw RFU Value		(Negative Control Tube)		Conversion Factor*	Percent Templated ISPs
	AF 488	AF 647	AF 488	AF 647		
campione non arricchito	273,3	177,7	26,9	11,9	1,27	22,57%

Fig. 10 Foglio di lavoro per la lettura del valore del campione non arricchito;

Dopo il controllo della qualità del campione non arricchito si è proceduto con l'arricchimento del campione utilizzando lo strumento Ion One Touch ES.

Una volta completato l'arricchimento, il *template* arricchito è stato quantificato attraverso il Qubit e il kit Ion Sphere™ Quality Control (Life Technologies), come per il campione non arricchito. In questo caso, il range del campione arricchito deve essere superiore al 70%.

### **Sequenziamento:**

Per il sequenziamento è stato utilizzato il Kit Ion PGM Sequencing 200 e il chip Ion 314.

Il workflow (Fig. 11) ha previsto i seguenti step: la pianificazione della corsa, il lavaggio, l'inizializzazione dello strumento Ion PGM [Personal Genome Machine, Life Technologies (Fig. 12)] ed il caricamento del Chip.

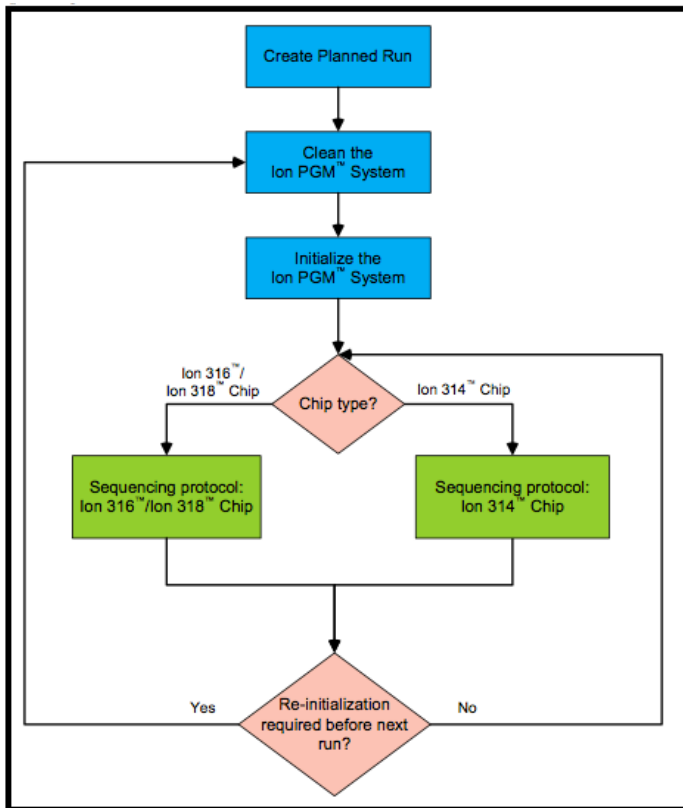


Fig. 11 Workflow per il sequenziamento con il PGM;

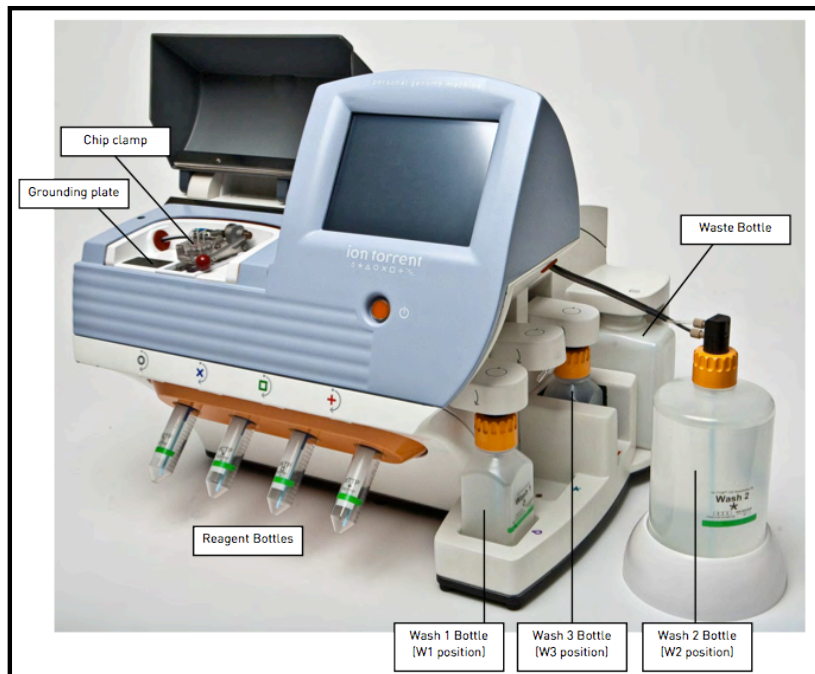


Fig.12 PGM;



La pianificazione della corsa è stata effettuata tramite il *Torrent Browser*, nel quale sono state inserite tutte le informazioni necessarie alla corsa (kit utilizzato, Chip, *barcode*, ecc.) e i plugin d'analisi da eseguire a fine corsa (coverage, variant caller, ecc);

### ***Analisi dei dati:***

A fine corsa (2h e 20 min), i risultati del sequenziamento sono stati esaminati nel *Torrent Browser* (*Figg. 13a, 13b*).

Il formato dei dati di output è in FASTQ, un formato molto comune per la condivisione di dati genetici di sequenziamento, in grado di combinare sia la sequenza di basi che un quality score associato ad ogni base nucleica, ovvero un punteggio di attendibilità sulla lettura di quella base all'interno della sequenza. Il formato FASTQ nasce come un'estensione del formato FASTA, ma, in aggiunta si hanno le informazioni sull'attendibilità della lettura, rappresentando così la sequenza di sequenziamento con un livello di dettaglio maggiore, ma senza pesare sulla dimensione dei dati.

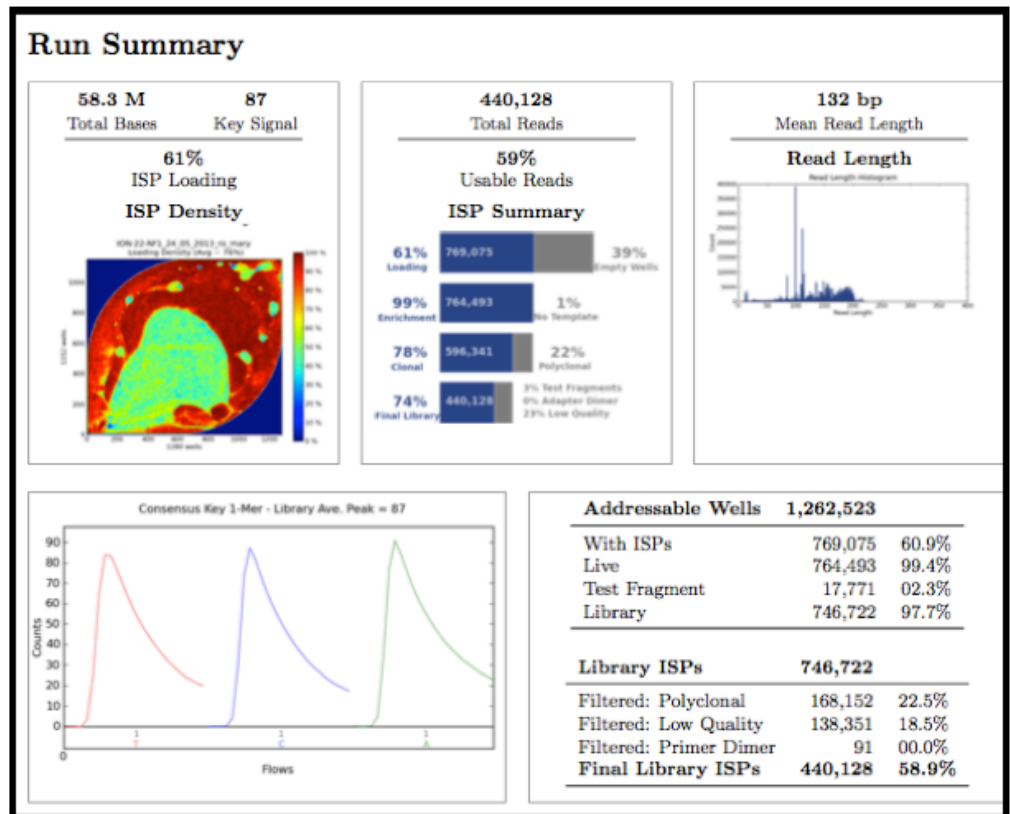


Fig. 13 a Run Summary/Report della corsa;

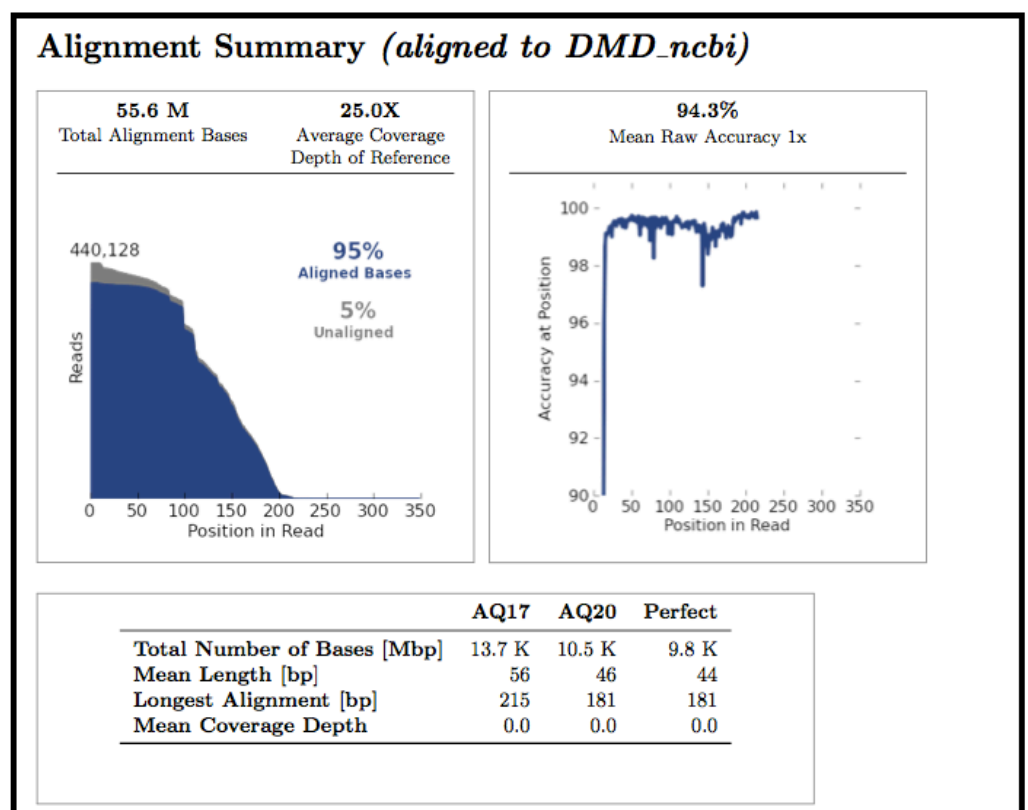


Fig. 13 b Run Summary/Report della corsa;

Dall'analisi si ottengono tutte le informazioni relative all'esperimento:

- ❖ Qualità del caricamento del chip;
- ❖ Percentuale di *template* caricato, percentuale di Ion Sphere Particles (ISP), sfere a cui sono legati i frammenti di DNA amplificati;
- ❖ *Clonality*, un ISP risulta clonale se tutti i frammenti di DNA originano da un singolo *template*. Tutti i frammenti contenenti le biglie sono identici e rispondono in modo uniforme durante l'incorporazione dei vari nucleotidi attraverso il chip;
- ❖ *Policlonality*, le biglie contengono dei cloni di due o più *template*; Range <30%;
- ❖ *Total bases*, numero di coppie di basi filtrate e corrette riportate nel file di output BAM (formato binario per archiviare i dati delle sequenze);
- ❖ *Key Signal*, percentuale di LiveISPs con un segnale chiave che risulta identico al segnale chiave della libreria;
- ❖ *Bead Loading*, percentuale di pozzetti del chip (*Chip wells*) contenenti LiveISPs;
- ❖ *Total Reads*, numero totale di sequenze utilizzabili, dopo la filtrazione indipendentemente dalla lunghezza riportata nel file di output BAM;
- ❖ *Usable Sequence*, la percentuale di sequenza utilizzabile che supera i requisiti richiesti post analisi di policlonalità, bassa qualità ed i diversi filtri riguardanti i dimero di primers;

- ❖ *Loading*, percentuale di *wells* nel chip che contengono LiveISPs;
- ❖ *Empty Wells*, percentuale di *wells* nel chip che non contengono un ISP;
- ❖ *Enrichment*, numero di *LiveISPs* che hanno un segnale chiave identico ai parametri del segnale chiave della *Library*, dopo filtraggio secondo il *Test Fragment* ISPs;
- ❖ *No Template*, percentuale di *wells* nel chip che non contengono uno stampo di DNA;
- ❖ *Final Library*, percentuale di letture che passano tutti i filtri e che sono registrate nel file di output BAM. Tale valore può essere diverso dal *Total Reads* per cause tecniche in quanto i filtri sono maggiori rispetto a quelli usati per le *Total Reads*;
- ❖ *% Test Fragment*, percentuale di LiveISPs con un segnale chiave identico al segnale chiave del *test Fragment*;
- ❖ *% Adapter Dimer*, percentuale di ISPs, con una lunghezza degli inserti inferiore alle 8 bp;
- ❖ *% Low Quality*, percentuale di ISPs con un segnale basso o non associabile;
- ❖ *Addressable Wells*, numero totale di *wells* indirizzabili;
- ❖ *With ISPs*, numero e percentuale di *wells* utilizzabili, cioè che hanno superato i vari filtri e che sono stati determinati come "positivi" per la presenza di un ISP all'interno dei *wells*. La positività viene determinata misurando la velocità di diffusione di un flusso con un pH diverso. I *wells* contenenti ISPs hanno un ritardo nella variazione di pH dovuto alla presenza di un ISP

che rallenta la rilevazione della variazione di pH della soluzione;

- ❖ *Live*, numero e percentuale di *wells* con ISPs, cioè *wells* contenenti un ISP con un segnale chiave e con i requisiti per essere associato alla libreria od al *test fragment*;
- ❖ *Test Fragment*, numero e percentuale di *LiveISPs* con un segnale chiave identico al segnale chiave che deve avere il *test fragment*;
- ❖ *Library*, numero e percentuale di *LiveISP* con un segnale chiave identico al segnale chiave;
- ❖ *Library ISPs*, numero di *LiveISPs* che hanno un segnale chiave identico all'atteso segnale chiave;
- ❖ *Filtered Polyclonal*: *ISPs* che contengono cloni di due o più modelli;
- ❖ *Filtered Low Quality*: Segnale basso od irriconoscibile;
- ❖ *Filtered Primer dimer*: Frammento inferior alle 8 bp;
- ❖ *Final Library ISPs*: Numero e percentuale di *ISPs* che superano tutti i filtri e che vengono registrati nel file di output BAM.

### ***Strumenti bioinformatici per l'analisi delle varianti:***

Il Browser, attraverso il plugin VariantCall (*Fig. 14*), elenca le varianti derivate dall'allineamento della sequenza del campione con la sequenza di riferimento. Tali varianti vengo visualizzate con il programma Integrative Genomics Viewer (IGV).

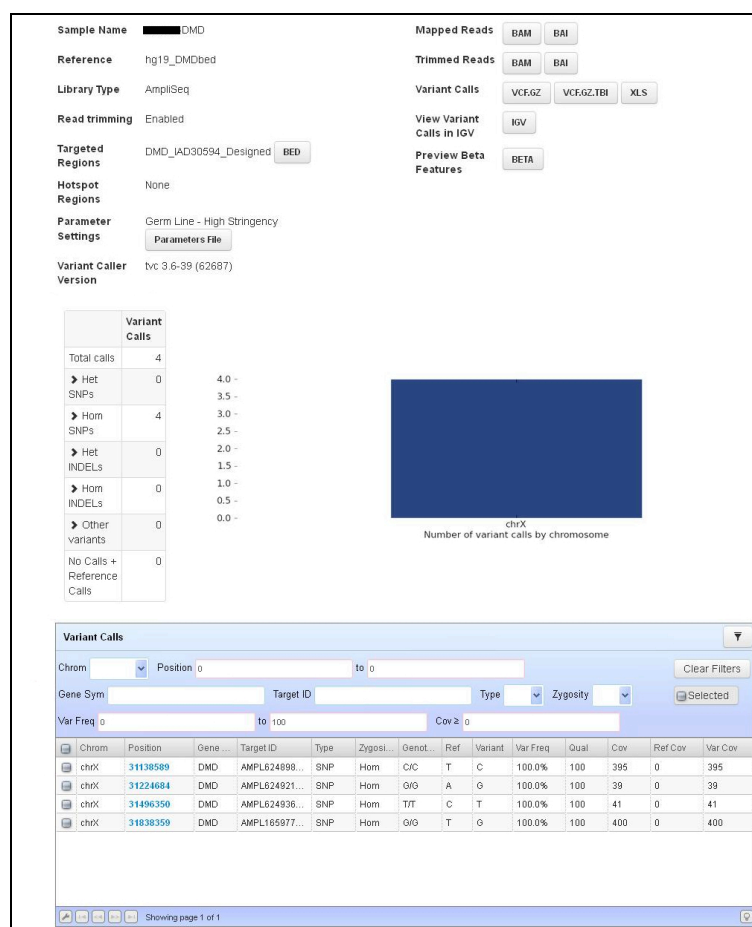


Fig. 14 Schermata VariantCall;

Tale software, fruibile on line (<http://www.broadinstitute.org/igv>) è un *high performance viewer* che consente di gestire in modo efficiente grandi quantità di dati eterogenei, fornisce una visualizzazione semplice e intuitiva della sequenza del campione che si vuole analizzare allineata al genoma (Figg. 15a, 15b).

La sequenza di riferimento utilizzata per gli esperimenti è hg19 in associazione ai BED files relativi agli ampliconi sequenziati.

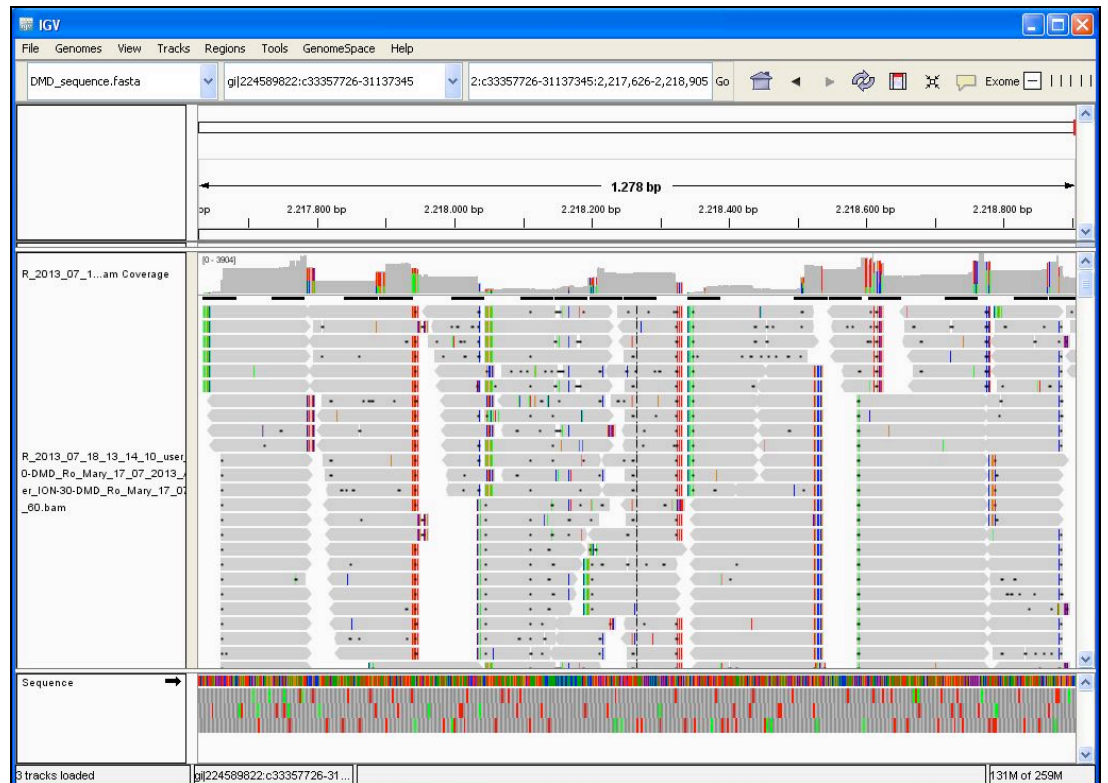


Fig. 15a Schermata IGV;

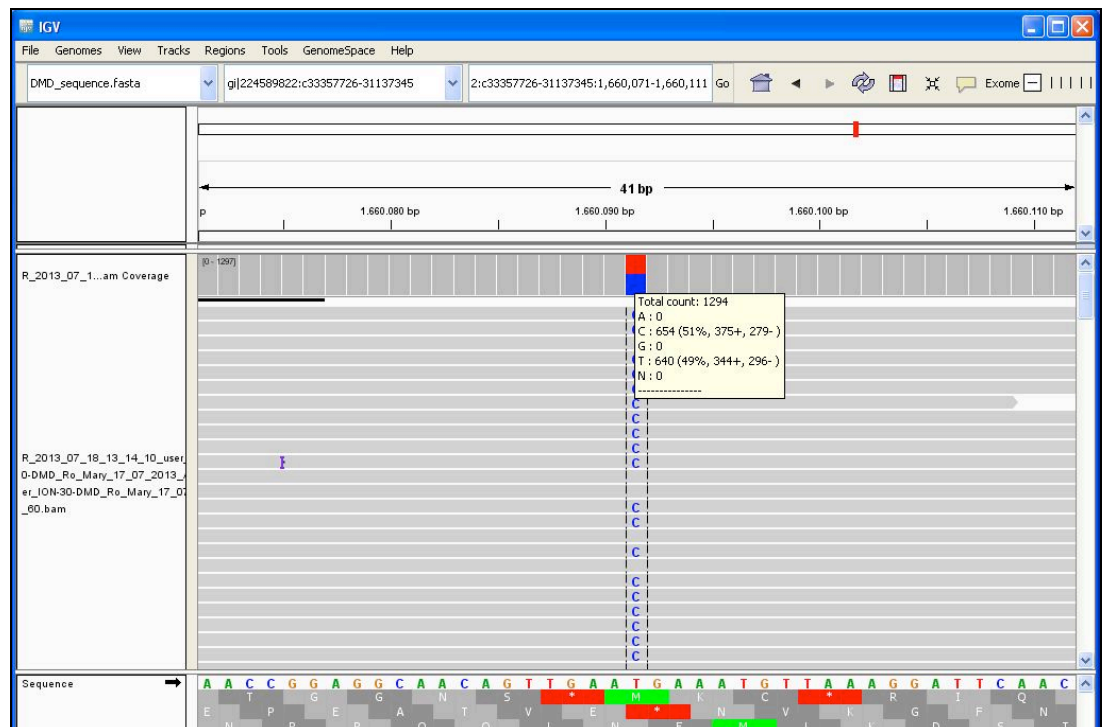


Fig. 15b Schermata IGV zoom-in, in cui si evince la presenza di un eterozigote C/T;

Il software ANNOVAR (<http://www.openbioinformatics.org>; <http://www.biobase-international.com/product/annovar> Rev. del 23\_08\_2013) è un efficiente strumento per annotare le varianti genetiche rilevate. Infatti è in grado di eseguire:

- Annotazioni genetiche: identificazione di SNPs o CNV (Copy Number Variation);
- Annotazioni Regione Base: identificare le varianti in specifiche regioni genomiche;
- Annotazioni filtrate: identificare le varianti che sono riportate nei data base (dbSNP), o identificare il sottoinsieme di SNPs comuni (*MAF, Minor Allele Frequency*>1%);
- Altre funzioni: consente di recuperare la sequenza nucleotidica in tutte le posizioni genomiche specifiche perciò l'identificazione, in elenco di geni candidati per una determinata patologia mendeliana dai dati di exome presenti nel web, come *RefSeq genes, UCSC genes, ENSEMBL genes, GENCODE genes* etc. etc..

Il software Galaxy è un open source che gira su server basati su Linux/Unix (<http://usegalaxy.org/toolshed> o <http://toolshed.g2.bx.psu.edu>).



### ***Sequenziamento con metodo Sanger:***

Le varianti definite causative dai software bioinformatici sono state valutate con metodo Sanger e strumento ABI PRISM 3130XL Genetic Analyzer (Life Technologies) utilizzando i Kit Big Dye Terminator v3.1 Cycle Sequencing (Life Technologies), AmPure XP e CleanSeq (Agencourt; Beckman) su piattaforma robotizzata Biomek NX (Beckman).

## Risultati:

Al fine di valutare l'applicabilità del metodo di sequenziamento di nuova generazione con la piattaforma Ion Torrent sono stati eseguiti 9 esperimenti, nei quali sono stati complessivamente esaminati 17 campioni: 10 per il gene DMD e 7 per il gene NF1. Dieci campioni (6 DMD e 4 NF1) erano già stati sequenziati con il metodo Sanger, pertanto l'indagine è stata retrospettiva.

La qualità del sequenziamento cioè la quantità di sequenze utilizzabili è stata determinata dalla percentuale di caricamento del chip, che rappresenta la porzione dei pozzetti del chip che sono stati riempiti fisicamente con delle sfere (ISP), dalla percentuale di sfere arricchite con librerie monoclonali e dai parametri "AQ17", "AQ20" e "perfect" che forniscono informazioni sulla qualità dell'allineamento in scala *Phred*. Tali valori raggruppano il numero di basi sequenziate ed allineate, in base alla probabilità di un incorretto "base calling". Le basi della categoria AQ17 hanno una accuratezza del 98%, quelle della categoria AQ20 del 99% (1 errore in 100bp) e quelle della categoria Perfect una accuratezza >99%.

Tramite il plug-in "Coverage Analysis", disponibile nel Torrent Server, è possibile avere informazioni sulla copertura delle regioni bersaglio di ciascun campione. Tutti questi fattori sono stati valutati per ciascun campione analizzato. In tutti gli esperimenti l'allineamento è stato effettuato contro il

genoma umano “*hg19*” e le regioni codificanti dei geni DMD e NF1 contenute nei rispettivi BED file degli ampliconi. Attraverso la selezione del plug-in *VariantCaller* dal Torrent Server è stata effettuata la chiamata delle varianti. Come mostrato in tabella 3 negli esperimenti 1 e 3 l’*ISP density* superiore al 70% indica un buon caricamento del Chip 314, ma la bassa percentuale di clonalità, rispettivamente del 22 e 39% (valore ottimale maggiore del 70%) e l’alta percentuale di policlonalità (78% e 61%, rispettivamente), hanno reso utilizzabili solo il 10% e 13% delle reads. La clonalità e policlonalità delle sfere è legata alla fase dell’*emulsion PCR*. La sfera è clonale quando ad essa è legato e amplificato un singolo amplicone, mentre è policlonale quando sono presenti più ampliconi diversi. Le sfere policlonali vengono automaticamente escluse, poiché di bassa qualità. In questi due esperimenti la percentuale degli ISPs del campione non arricchito era al di fuori del range ottimale, (range ottimale 10-30%): rispettivamente di 86% e 90%.

Nella Tabella 3 vengono mostrati i dati ed i risultati degli esperimenti:

Esperimento	Reads utilizzabili	ISP density caricamento Chip	Clonalita	Campioni	Studio retrospettivo SI/No	Mutazione con Sanger	Mutazione con IonTorrent PGM	Mutazione con IonTorrent VariantCaller
1*	10%	76%	22%	NF1_2392	Si	c.4831C>T	c.4831C>T	c.4831C>T
2*	59%	61%	78%	DMD_2368	Si	no mut	no mut	no mut
3*	13%	77%	39%	NF1_2253	Si	c.4819_4820ins	c.4819_4820insT	NO MUT
				DMD_2486	Si	c.6526delA	c.6526delA	c.6526delA
4*				NF1_2329*				
5*	60%	39%	78%	DMD_2458	Si	no mut	no mut	no mut
				DMD_2461*				
6*	59%	53%	78%	NF1	Si	no mut	no mut	no mut
				NF1_2396	No	c.3916C>T	c.3916C>T	c.3916C>T
7*	73%	90%	80%	DMD_2354	Si	no mut	no mut	no mut
				DMD_2414	No		no mut	no mut
				DMD_2509	No		no mut	no mut
8*	66%	90%	71%	DMD_2251	Si	no mut	no mut	no mut
				DMD_2190	Si	c.10087-1G>C	c.10087-1G>C	no mut
				DMD_2177	Si	c.3940C>T	c.3940C>T	c.3940C>T
9*	78%	88%	83%	NF1_2349	No	c.3233C>G	c.3233C>G	c.3233C>G
				NF1_2	No		no mut	no mut
				NF1_2357	No	c.1672_1675del	c.1672_1675delinsTGA	c.1672_1675delinsT

Tab. 3 Dati e risultati degli esperimenti; i campioni che mostrano un asterisco avevano librerie non idonee, perciò non sono state analizzate al PGM;

Nelle figure 16 e 17 sono visualizzati i risultati della quantizzazione degli ISPs del campione non arricchito e del campione arricchito, mentre nelle figure 18 e 19 sono visualizzati i sommari delle corse.

Qubit Calibration Factor Calculation						
Calibration Standard	RFU	Calibration Factor				
Alexa Fluor® 488 Calibration Standard	17568,6	0,28				
Alexa Fluor® 647 Calibration Standard	56410,9					
Percent Templated ISPs						
Sample ID	Raw RFU Value		(Negative Control Tube)		Conversion Factor*	Percent Templated ISPs
	AF 488	AF 647	AF 488	AF 647		
campione non arricchito 22_05_2013	1264,9	2958,9	32,7	13,4	1,27	85,95%
campione arricchito 22_05_2013	150,3	303,9	36,4	11,9	1,27	92,18%

Fig. 16 Quantizzazione degli ISPs della libreria dell'esperimento 1;

Qubit Calibration Factor Calculation						
Calibration Standard	RFU	Calibration Factor				
Alexa Fluor® 488 Calibration Standard	17256,3	0,27				
Alexa Fluor® 647 Calibration Standard	59153,7					
Percent Templated ISPs						
Sample ID	Raw RFU Value		(Negative Control Tube)		Conversion Factor*	Percent Templated ISPs
	AF 488	AF 647	AF 488	AF 647		
campione non arricchito 21_06_2013	619,1	1591,5	29,9	12,7	1,27	90,25%
campione arricchito 21_06_2013	174,8	475,1	31,6	11,1	1,27	109,13%

Fig. 17 Quantizzazione degli ISPs della libreria dell'esperimento 3;

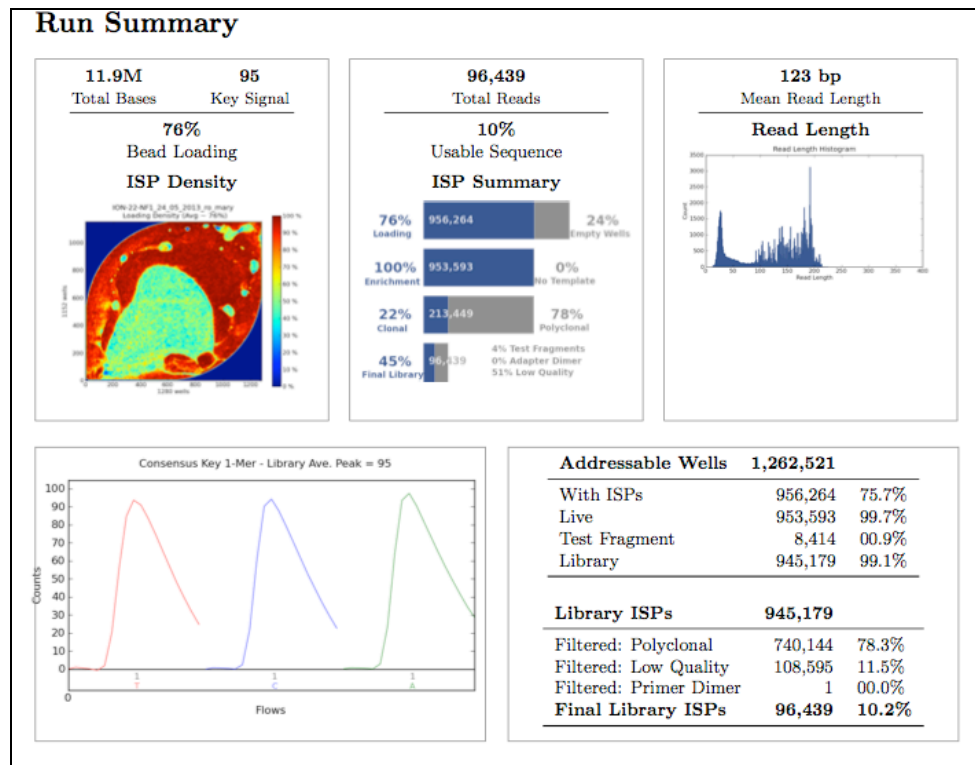


Fig. 18 Sommario della corsa dell'esperimento 1;

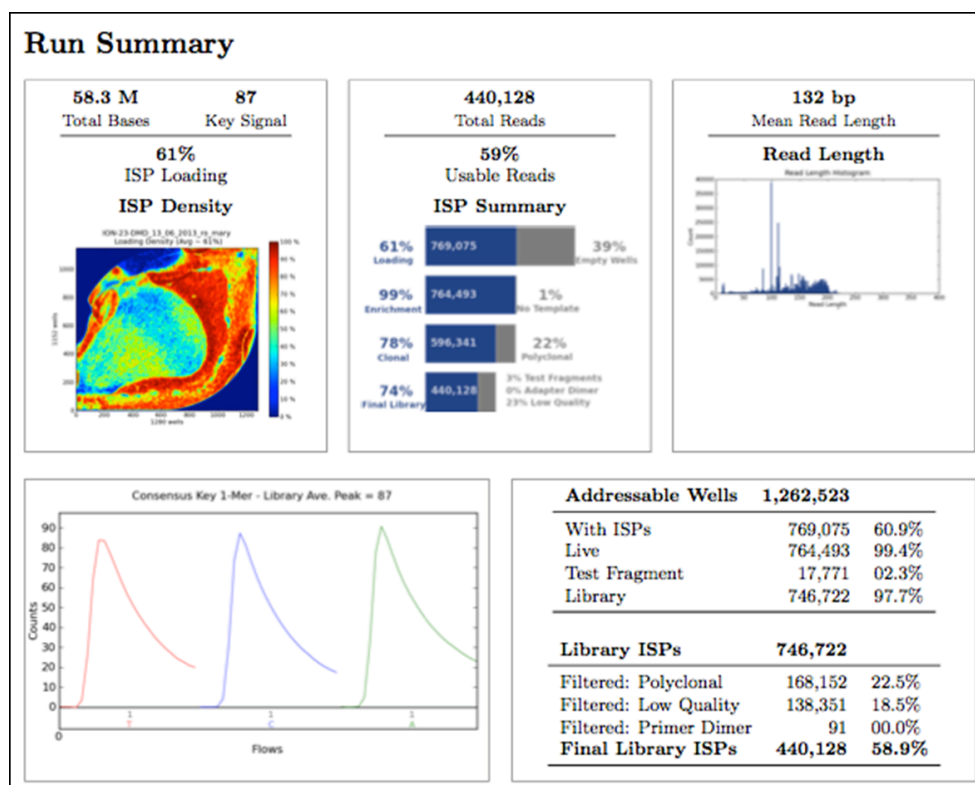
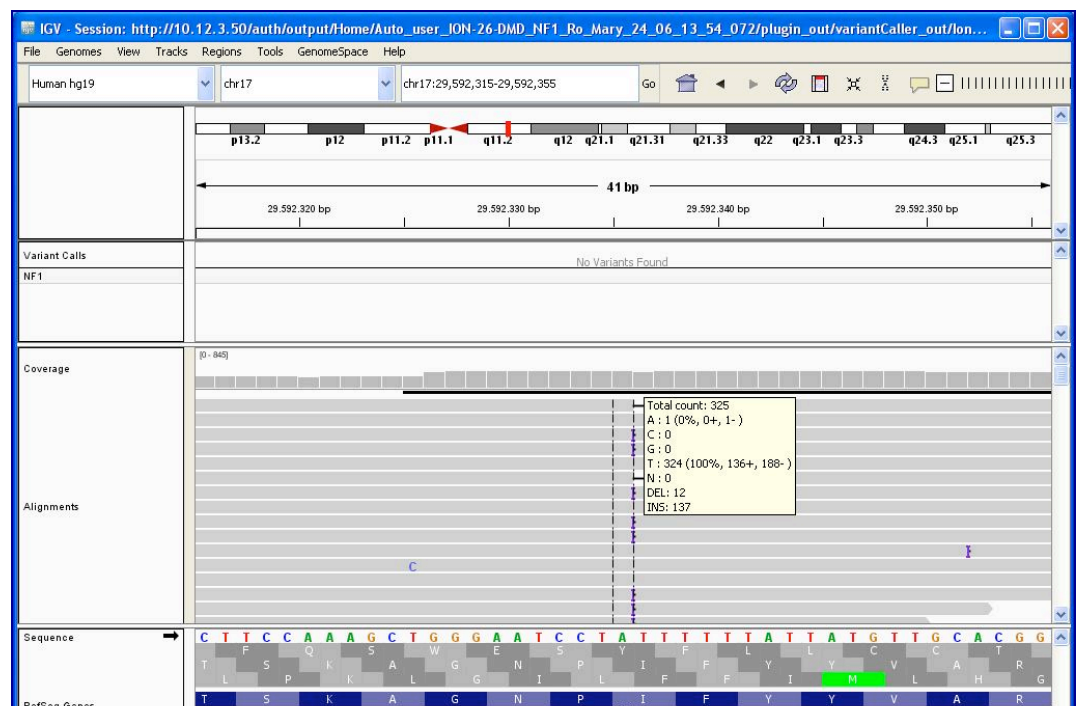


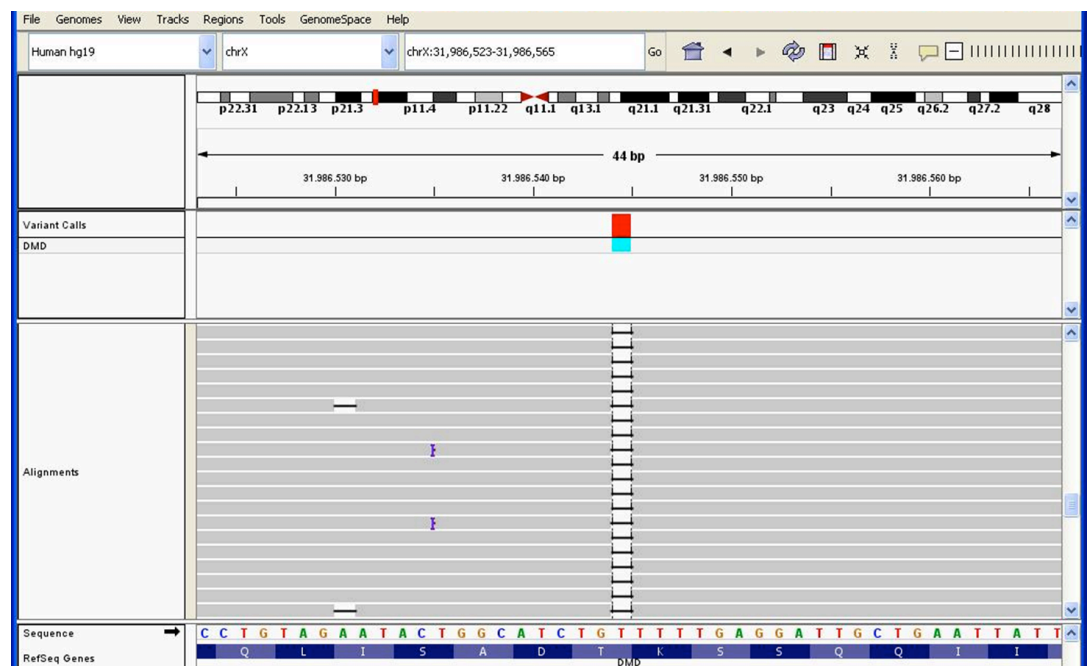
Fig. 19 Sommario della corsa dell'esperimento 3;

L'analisi con *VariantCaller* non ha correttamente chiamato la mutazione patogenetica identificata con il sequenziamento Sanger in un campione su tre. Infatti, per il campione NF1\_2253 in cui è presente la mutazione c.4819\_4820insT (esone 36 del gene NF1) situata in corrispondenza di un stretch di T, il sequenziamento è avvenuto correttamente perché la mutazione è visibile su IGV (*Fig. 20*) ma il software *VariantCaller*, sia in modalità alta stringenza che in bassa stringenza, non l'ha evidenziata.



*Fig. 20* Immagine IGV del campione NF1\_2253;

Nel campione DMD\_2486, in cui è presente la mutazione patogenetica c.6526delA dell'esone 45 correttamente identificata e chiamata (*Fig. 21*), sono state evidenziate alcune incoerenze tra i risultati del sequenziamento Ion Torrent e quelli del sequenziamento di Sanger. Infatti, i polimorfismi IVS 41 delAA e la sostituzione C>T (rs2270672) osservati con il Sanger, sono stati sequenziati correttamente dal PGM ma non chiamati da VariantCaller, mentre l'inversione AT>TA non è stata correttamente sequenziata dal PGM (*Figg. 22 e 23*).



*Fig. 21* Immagine IGV del campione DMD\_2486;



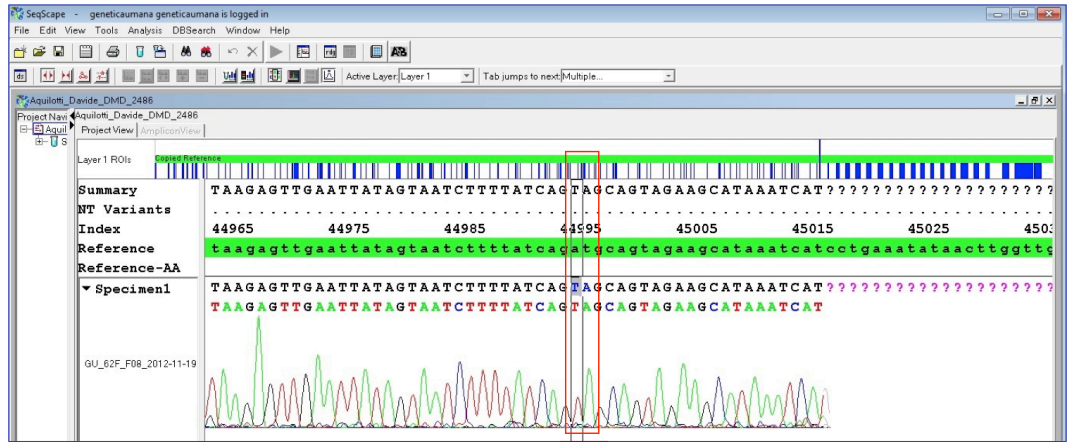


Fig. 22 Schermata dell'inversione AT>TA con il programma SeqScape;

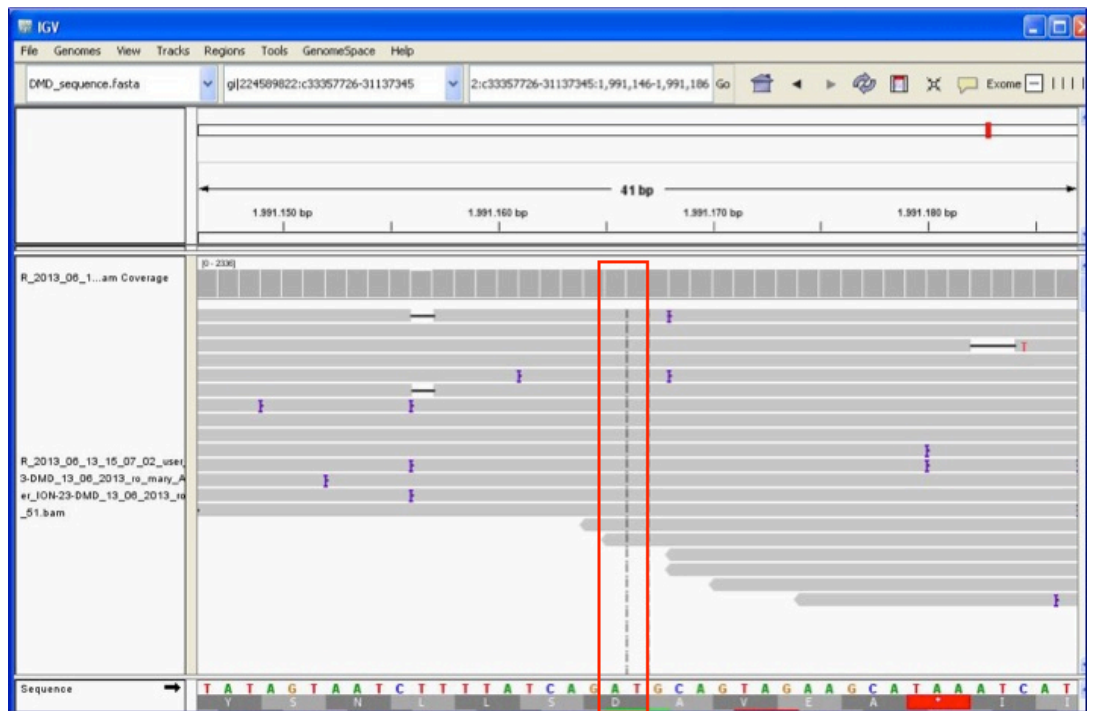


Fig. 23 Screenshot programma IGV che non conferma l'inversione;

Poiché i risultati degli esperimenti 1 e 3 non sono stati soddisfacenti si è deciso di non effettuare il sequenziamento della libreria ottenuta dall'esperimento 4, in cui, come mostrato in figura 24, la percentuale del campione non arricchito era del 117%.

Qubit Calibration Factor Calculation						
Calibration Standard	RFU	Calibration Factor				
Alexa Fluor® 488 Calibration Standard	17369,6	0,27				
Alexa Fluor® 647 Calibration Standard	58463,8					
Percent Templated ISPs						
Sample ID	Raw RFU Value		(Negative Control Tube)		Conversion Factor*	Percent Templated ISPs
	AF 488	AF 647	AF 488	AF 647		
campione non arricchito 03_07_2013	208	619,7	30,2	12,6	1,27	117,12%
campione arricchito						#DIV/0!

Fig. 24 Quantizzazione degli ISPs della libreria dell'esperimento 4;

Negli esperimenti 2, 5, 6, 7, 8 e 9 i valori relativi alla percentuale delle ISPs dei campioni non arricchito e arricchito, della clonalità e delle reads utilizzabili erano soddisfacenti (Figg. 25 e 26).

Qubit Calibration Factor Calculation						
Calibration Standard	RFU	Calibration Factor				
Alexa Fluor® 488 Calibration Standard	17201,6	0,26				
Alexa Fluor® 647 Calibration Standard	59217,1					
Percent Templated ISPs						
Sample ID	Raw RFU Value		(Negative Control Tube)		Conversion Factor*	Percent Templated ISPs
	AF 488	AF 647	AF 488	AF 647		
campione non arricchito 17_07_2013	273,3	177,7	26,9	11,9	1,27	22,57%
campione arricchito 17_07_2013	60,7	122	26,9	11,9	1,27	109,25%

Fig. 25 Quantizzazione degli ISPs della libreria dell'esperimento 5;

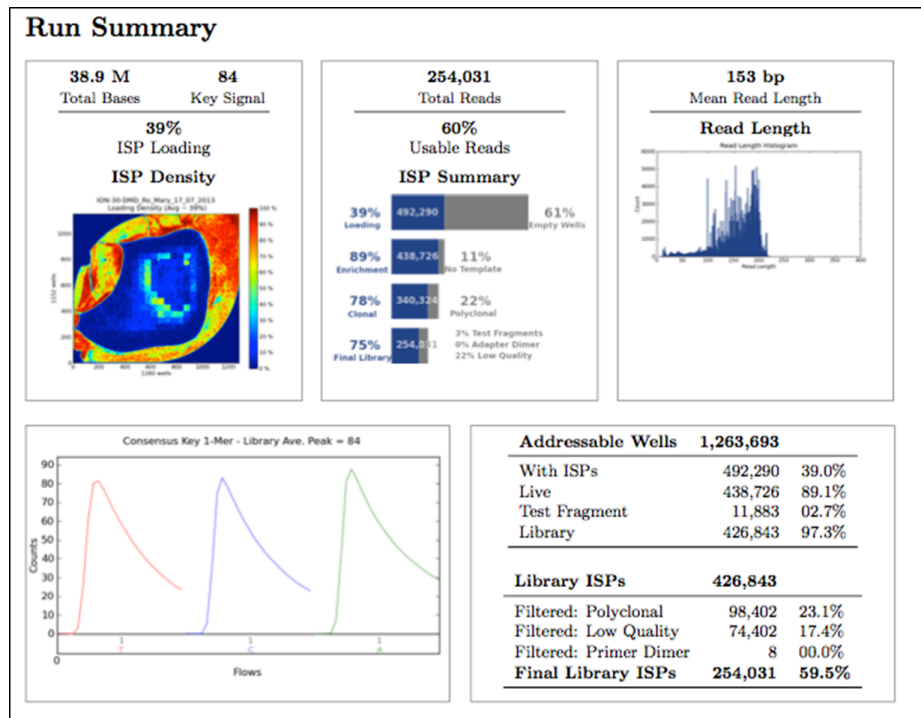


Fig. 26 Sommario della corsa dell'esperimento 5;

In 6 dei 10 campioni studiati retrospettivamente, l'analisi con NGS ha dato risultati concordanti rispetto al sequenziamento Sanger. Nel campione DMD\_2190, la mutazione patogenetica c.10087-1G>C evidenziata dal Sanger, è stata sequenziata dal PGM, ma non è stata chiamata da *VariantCaller*. Tuttavia in questo caso, come si vede dalla figura 27, la sequenza di riferimento relativa all'esone 66, contenuto nel BED file del gene DMD, non si estende sino alla regione interessata dalla mutazione. L'analisi ripetuta effettuando l'allineamento con una sequenza di riferimento più estesa ha messo in evidenza la mutazione.

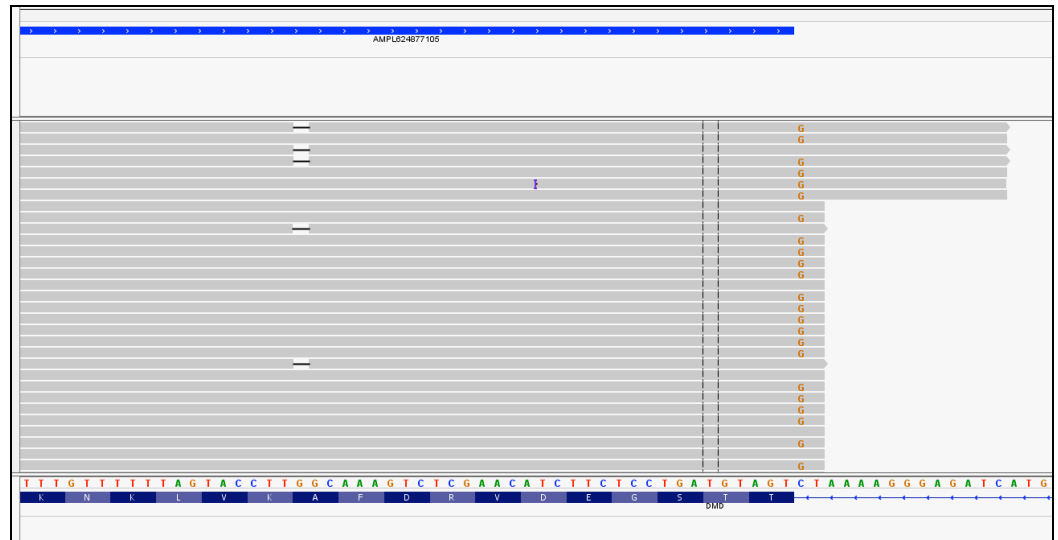


Fig. 27 Screenshot del campione DMD\_2190;

Nel campione DMD\_2458 sono state osservate le seguenti incoerenze tra i risultati del sequenziamento Ion Torrent e quelli del sequenziamento di Sanger: IVS42 delT e IVS66 C>T osservate al PGM e non confermate al Sanger.

Negli esperimenti 6, 7, 8 e 9 sono stati analizzati 6 campioni (4 NF1 e 2 DMD) non preventivamente studiati con metodo Sanger. In 3 campioni NF1 il sequenziamento PGM e il *VariantCaller* hanno correttamente identificato la mutazione che è stata successivamente confermata con il metodo Sanger. Nei restanti 3 soggetti (1 NF1 e 2 DMD) il sequenziamento tramite PGM non ha evidenziato alcuna mutazione. Nell'esperimento 5 sono state preparate le librerie di due campioni DMD, ma poiché la lettura del campione DMD\_2461 con il Bioanalyzer ha mostrato che il campione non si era amplificato, pertanto è stato eliminato dalle fasi successive del protocollo (Fig. 28).

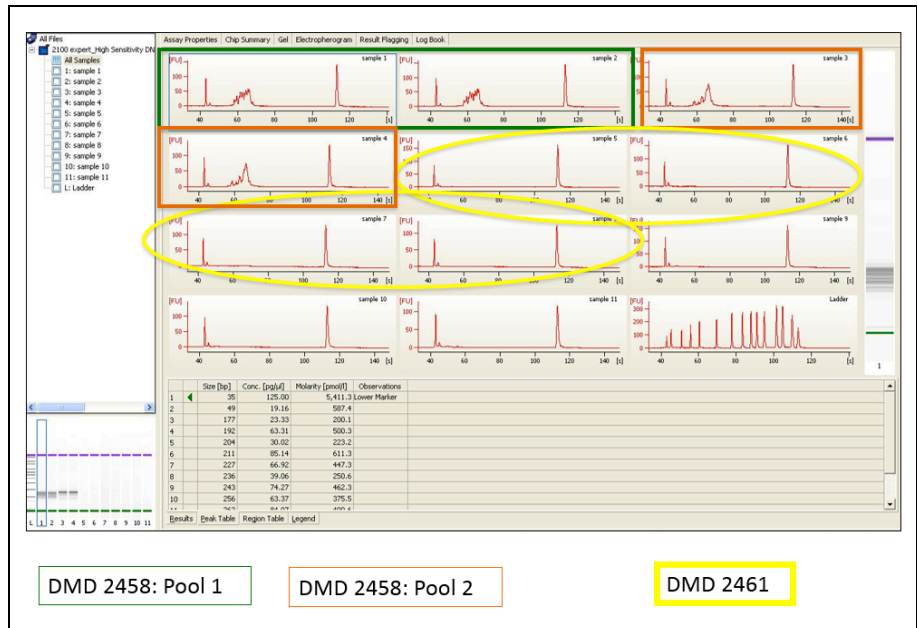


Fig. 28 In giallo i risultati del Bioanalyzer del campione DMD\_2461;

In tutti gli esperimenti e per ciascun campione la copertura delle regioni sequenziate è stata superiore a 200X.

## Discussione:

L'identificazione delle mutazioni causative di malattie genetiche permette di confermare il sospetto clinico e di fornire una diagnosi accurata. La conoscenza della mutazione, solitamente, consente di effettuare una correlazione genotipo-fenotipo, migliora la gestione del paziente, permette l'inclusione nei trials terapeutici, rende possibile la consulenza genetica, la determinazione dello stato di portatori e la diagnosi prenatale.

Le strategie per l'analisi delle mutazioni da utilizzare nei laboratori di diagnostica devono essere semplici, rapide, avere un'alta accuratezza e una bassa percentuale di errore.

Il sequenziamento attraverso il metodo Sanger rappresenta il gold standard per l'identificazione delle mutazioni puntiformi e delle piccole inserzioni/delezioni. Tuttavia, se il gene è molto grande e le mutazioni sono distribuite su l'intera sequenza del gene, questo approccio è laborioso e richiede tempi lunghi.

La tecnologia di sequenziamento Next Generation è un sistema "*high-throughput*" in grado di generare in breve tempo milioni di sequenze genetiche e consiste nella preparazione della libreria, fase in cui è possibile incorporare dei barcode che permettono l'analisi simultanea di più pazienti, nella preparazione del *template*,

nel sequenziamento e nell'analisi dei dati.

L'obiettivo di questo studio è stato quello di valutare l'applicabilità di questa tecnologia alla pratica routinaria della ricerca di mutazioni puntiformi nei geni DMD e NF1 e di validare un protocollo d'analisi.

Per entrambi i geni l'identificazione delle mutazioni puntiformi è un compito impegnativo. Infatti, il gene DMD, uno dei più grandi geni sinora identificati, e il gene NF1 contengono rispettivamente 79 e 57 esoni e non sono presenti hot spots mutazionali.

In questo studio è stata utilizzata la piattaforma IonTorrent che sfrutta la tecnologia dei semiconduttori ed è incentrata sull'impiego di Chip ad alta densità. Attraverso l'utilizzo del protocollo AmpliSeq con custom primers e del Chip 314, sono stati analizzati, in 8 esperimenti, 10 soggetti per il gene DMD e 7 per il gene NF1. I risultati ottenuti hanno permesso di confrontare la metodica di sequenziamento Next Generation IonTorrent con quella classica di Sanger e di mettere in evidenza sia i punti di forza sia i punti di debolezza.

I punti di forza sono rappresentati dalla possibilità di esaminare in breve tempo più campioni simultaneamente (in un chip 314: 3 campioni per gene DMD o 4 campioni per gene NF1, in 5 giorni), dalla disponibilità di kit commerciali e di un protocollo robusto, dal supporto fornito dallo strumento Ion One Touch che automatizza la fase della preparazione del *template*, dalla disponibilità di

software accessibili anche ai non bioinformatici e da costi paragonabili o inferiori al metodo Sanger.

L'accuratezza nell'identificare la mutazione patogenetica è stata del 90%, infatti la mutazione è stata correttamente evidenziata in 9 pazienti su 10 esaminati retrospettivamente. Il falso negativo, un campione NF1 con l'inserzione di una T in corrispondenza di uno stretch di T, potrebbe essere imputabile ai bassi indici di qualità dell'esperimento, infatti lo strumento ha correttamente sequenziato la regione, ma *VariantCaller* non l'ha chiamata.

I punti di debolezza del sistema sono rappresentati dai falsi negativi e falsi positivi. Dallo studio retrospettivo e relativamente alle sole varianti polimorfiche sono stati evidenziati 2 falsi positivi e 3 falsi negativi. I falsi negativi sono rappresentati da errori in corrispondenza degli omopolimeri. Gli errori legati al corretto sequenziamento degli omopolimeri, sono una caratteristica dominante dei dati provenienti dal PGM [Bragg *et al.*, 2013; Yeo *et al.*, 2012; Vasli e Laporte, 2013], ma non esclusiva di questa tecnologia [Quail *et al.*, 2012]. Questi errori sono la conseguenza di una inaccurata registrazione, in più o in meno, delle basi contenute negli omopolimeri e non sempre è associata ad omopolimeri lunghi 2 o più basi.

Un altro inconveniente della tecnologia NGS è rappresentato dall'alta percentuale di errore, che per IonTorrent è stata quantizzata nella misura del 1.78% [Quail *et al.*, 2012].

I falsi positivi vengono prodotti durante l'amplificazione del



*template* o durante il sequenziamento e possono essere ridotti con l'utilizzo di protocolli di filtraggio dei dati di output più selettivi (SAMtools, Dindel, GATK's UnifiedGenotyper ed UnifiedGenotyper e Filtered (*BAF, B-Allele Frequency e VARW Variation of the Width of gaps and insert thresholds*) [Yeo *et al.*, 2012]. A causa dell'elevata percentuale di errore il sequenziamento tramite Sanger rappresenta uno step di validazione essenziale delle varianti interessanti. Tale test, però, comporta un aumento sia dei costi sia dei tempi di consegna dei risultati.

La fine caratterizzazione molecolare della distrofia muscolare di Duchenne/Becker e della Neurofibromatosi tipo 1 è diventata obbligatoria per selezionare i pazienti da includere nei trials di medicina personalizzata in cui ciascuna specifica alterazione richiede un farmaco molecolare diverso. Negli ultimi anni numerose strategie terapeutiche aventi per bersaglio il difetto genetico primario del gene della distrofina sono state inserite in trials clinici. Tra questi, l'exon-skipping con oligonucleotidi antisenso (AON) in grado di rettificare i reading-frame aberranti e by-passare i codoni di stop e il PTC124 (ataluren) che trasforma un codone di stop in un codone nonsense.

Sebbene attualmente la terapia dei pazienti con Neurofibromatosi tipo 1 sia la rimozione chirurgica dei neurofibromi ed il trattamento delle caratteristiche cliniche associate alla malattia, sono stati riportati risultati sperimentali con fattori di crescita, tirosina chinasi ed

inibitori delle chinasi oncogeniche che potrebbero beneficiare dalla conoscenza esatta e dettagliata del background genetico.

Grazie allo studio oggetto di questa tesi è possibile affermare che la tecnologia di sequenziamento high-throughput Ion Torrent rappresenta uno strumento di indagine molecolare del gene DMD e NF1 applicabile in un contesto di diagnosi routinaria poiché permette di ottenere risultati accurati in breve tempo.

## Bibliografia:

- Acharyya, S., S. A. Villalta, et al. "Interplay of IKK/NF-kappaB signaling in macrophages and myofibers promotes muscle degeneration in Duchenne muscular dystrophy". *J Clin Invest* 117(4): 889-901. (2007).
- Allsop, K. G. and F. A. Ziter "Loss of strength and functional decline in Duchenne's dystrophy". *Arch Neurol* 38(7): 406-11. (1981).
- Alman, B. A. "Duchenne muscular dystrophy and steroids: pharmacologic treatment in the absence of effective gene therapy". *J Pediatr Orthop* 25(4): 554-6. (2005).
- Anderson, J. T., R. P. Rogers, et al. "Ca<sup>2+</sup>-calmodulin binds to the carboxyl-terminal domain of dystrophin". *J Biol Chem* 271(12): 6605-10. (1996).
- Angelini, C., E. Pegoraro, et al. "Deflazacort in Duchenne dystrophy: study of long-term effect". *Muscle Nerve* 17(4): 386-91. (1994).
- Aravind L, Neuwald AF, Ponting CP. "Sec14p-like domains in NF1 and Dbl-like proteins indicate lipid regulation of Ras and Rho signaling". *Curr. Biol*, 9 (6): R 195-7. (1999).
- Ars E, Kruyer H, Morell M, Pros E, Serra E, Ravella A, Estivill X, Lázaro C., "Recurrent mutations in the NF1 gene are common among neurofibromatosis type 1 patients". *J Med Genet*; 40: e82. (2003).
- Ars E, Serra E, Garcia J, Kruyer H, Gaona A, Lazaro C, Estivill X., "Mutations affecting mRNA splicing are the most common molecular

- defects in patients with neurofibromatosis type 1". *Human Molecular Genetics*, 9(2): 237-247. (2000).
- Baralle D, Baralle M., "Splicing in action-assessing disease causing sequence variations". *JMed Genet* 42:737-748. (2005).
- Barnea, E., D. Zuk, et al. "Specificity of expression of the muscle and brain dystrophin gene promoters in muscle and brain cells". *Neuron* 5(6): 881-8. (1990).
- Beroud C, Carrie A, Beldjorf C, Deburgrave N, Llense S, Carelle N, Peccate C, Cuisset JM, Pandit F, Carre-Pigeon F, Mayer M, Bellance R, Recan D, Chelly J, Kaplan JC, Leturcq F., "Dystrophinopathy caused by mid-intronic substitutions activating cryptic exons in the DMD gene". *Neuromuscul Disord* 14(1):10-8. (2004).
- Blake, D. J., A. Weir, et al. "Function and genetics of dystrophin and dystrophin-related proteins in muscle". *Physiol Rev* 82(2): 291-329. (2002).
- Blake, D. J., J. M. Tinsley, et al. "Coiled-coil regions in the carboxy-terminal domains of dystrophin and related proteins: potentials for protein-protein interactions". *Trends Biochem Sci* 20(4): 133-5. (1995).
- Bollag G, McCormick F, Clark R., "Characterization of full-length neurofibromin: tubulin inhibits Ras GAP activity". *EMBO J*; 12: 1923-1927. (1993).
- Bonifati, M. D., G. Ruzza, et al., "A multicenter, double-blind, randomized trial of deflazacort versus prednisone in Duchenne muscular dystrophy". *Muscle Nerve* 23(9): 1344-7. (2000).

- Bonilla, E., C. E. Samitt, et al. "Duchenne muscular dystrophy: deficiency of dystrophin at the muscle cell surface". *Cell* 54(4): 447-52. (1988).
- Bottillo I, De Luca A, Colapietro P, Schirinzi A, Guida V, Pizzetti A, Larizza L, Dalla Piccola B., "Mutazioni di splicing nell'esone 7 del gene NF1 associate all'eliminazione o all'introduzione di siti ESE: evidenze per la presenza di un'isoforma alternativa NF1-E7 della neurofibromina". Poster, SIGU (2005).
- Bragg LM, Stone G, Butler MK, Hugenholtz P, Tyson GW., "Shining a light on dark sequencing: characterising errors in Ion Torrent PGM data". *PLoS Comput Biol.* Apr;9(4) Epub Apr 11. (2013).
- Brooke, M. H., G. M. Fenichel, et al., "Clinical investigation of Duchenne muscular dystrophy. Interesting results in a trial of prednisone." *Arch Neurol* 44(8): 812-7. (1987).
- Bruno Francesco Gavassini, "Identificazione di geni differenzialmente espressi nella distrofia muscolare di Duchenne e loro ruolo nella progressione di malattia", Tesi di Dottorato XXI Ciclo;
- Caceres JF and Kornblihtt AR., "Alternative splicing: multiple control mechanisms and involvement in human disease". *Trends Genet.* 18, 186-193. (2002).
- Caputi M, Casari G, Guenzi S, Tagliabue R, Sidoli A, Melo CA, Baralle FE., "A novel bipartite splicing enhancer modulates the differential processing of human fibronectin EDA exon". *Nucleic Acids Research* 22:1018-1022. (1994).

- Cartegni L, Chew SL, and Krainer AR., "Listening to silence and understanding nonsense: exonic mutations that affect splicing". *Nat Rev Genet* 3, 285-298. (2002).
- Cawthon RM, O'Connell P, Buchberg AM, et al., "Identification and characterization of transcripts from the neurofibromatosis 1 region: the sequence and genomic structure of EVI2 and mapping of other transcripts". *Genomics*; 7: 555-65. (1990).
- Cawthon RM, Weiss R, Xu G, Viskochil D, Culver M, Stevens J, Robertson M, Dunn D, Gesteland R, O'Connell P, White R., "A major segment of the neurofibromatosis type 1 gene: cDNA sequence, genomic structure, and point mutations". *Cell*; 62(1): 193-201. (1990).
- Cheadle, C., M. P. Vawter, et al. "Analysis of microarray data using Z score transformation". *J Mol Diagn* 5(2): 73-81. (2003).
- Chelly, J., G. Hamard, et al. "Dystrophin gene transcribed from different promoters in neuronal and glial cells". *Nature* 344(6261): 64-5. (1990).
- Chen, Y. J., R. Kodell, et al. "Normalization methods for analysis of microarray gene-expression data." *J Biopharm Stat* 13(1): 57-74. (2003).
- Clarkson, P. M., J. M. Devaney, et al. "ACTN3 genotype is associated with increases in muscle strength in response to resistance training in women." *J Appl Physiol* 99(1): 154-63. (2005).
- Clementi M, Barbujani G, Turolla L, Tenconi R. "Neurofibromatosis-1: a maximum likelihood estimation of mutation rate". *Hum. Genet.*; 84: 116-118. (1990).

- Collins S., M. Morgan, and A. Patrinos, "The Human Genome Project: lessons from large-scale biology," *Science*, vol. 300, no. 5617, pp. 286–290, (2003).
- Crosbie, R. H., J. Heighway, et al. "Sarcospan, the 25-kDa transmembrane component of the dystrophin-glycoprotein complex." *J Biol Chem* 272(50): 31221-4. (1997).
- Culligan, K. and K. Ohlendieck "Diversity of the Brain Dystrophin-Glycoprotein Complex." *J Biomed Biotechnol* 2(1): 31-36. (2002).
- Cummings LM., Trent JM, Marchuck DA. Identification and mapping of type 1 neurofibromatosis (NF1) homologous loci. *Cytogenet Cell Genet*; 73: 334-340. (1996).
- Dasgupta B, Dugan LL, Gutmann DH., "The neurofibromatosis 1 gene product neurofibromin regulates pituitary adenylate cyclase-activating polypeptide-mediated signaling in astrocytes". *The J of Neuroscience*; 23: 8949- 8954. (2003).
- Daston MM, Scrabble H, Nordlund M, Sturbaum AK, Nissen LM, Ratner N. "The protein product of neurofibromatosis type 1 gene is expressed at the highest abundance in neurons, Schwann cells and oligodendrocytes". *Neuron* 8: 415-428. (1992).
- Davis, M. H., R. Cappel, et al. "Creatine kinase activity in normal and Duchenne muscular dystrophy fibroblasts." *Muscle Nerve* 5(1): 1-6. (1982).
- De Luca A, Buccino A, Gianni D, Mangino M, Giustizi S, Richetta A, Divona L, Calvieri S, Mingarelli R, Dalla Piccola B. "NF1 gene analysis based on DHPLC". *Hum Mutat.* 21(2): 171-2. (2003).

- Dogra, C., D. S. Srivastava, et al. "Protein-DNA array-based identification of transcription factor activities differentially regulated in skeletal muscle of normal and dystrophin-deficient mdx mice." *Mol Cell Biochem* 312(1-2): 17-24. (2008).
- Dubowitz, V. "Prednisone for Duchenne muscular dystrophy." *Lancet Neurol* 4(5): 264. (2005).
- Dubowitz, V. and M. Fardeau "Proceedings of the 27th ENMC sponsored workshop on congenital muscular dystrophy. 22-24 April 1994, The Netherlands." *Neuromuscul Disord* 5(3): 253-8. (1995).
- Dudoit, S. and T. P. Speed "A score test for the linkage analysis of qualitative and quantitative traits based on identity by descent data from sib-pairs." *Biostatistics* 1(1): 1-26. (2000).
- Dugan LL, Kim JS, Zhang Y, Bart RD, Sun Y, Holtzman DM, Gutmann DH. "Differential effects of cAMP in neurons and astrocytes". *Role of B-raf*. *J Biol Chem*; 274: 25842-25848. (1999).
- Emery, A. E. "Duchenne muscular dystrophy. Genetic aspects, carrier detection and antenatal diagnosis." *Br Med Bull* 36(2): 117-22. (1980).
- Engel, A. G. "Congenital myasthenic syndromes." *Neurol Clin* 12(2): 401-37. (1994).
- Ervasti, J. M. "Dystrophin, its interactions with other proteins, and implications for muscular dystrophy." *Biochim Biophys Acta* 1772(2): 108-17. (2007).



- Fadic, R. "Cell surface and gene expression regulation molecules in dystrophinopathy: mdx vs. Duchenne." *Biol Res* 38(4): 375-80. (2005).
- Fashold R, Hoffmeyer S, Mischung C, Gille C, Ehlers C, Küçükceylan N, Abdel- Nour M, Gewies A, Peters H, Kauffman D, Buske A, Tinschert S, Nürnberg P., "Minor lesion mutational spectrum of the entire NF1 gene does not explain its high mutability but points to a functional domain upstream of the GAP- related domain". *Am J Hum Genet.*; 66: 790-818. (2000).
- Fehniger T. A., T. Wylie, E. Germino et al., "Next-generation sequencing identifies the natural killer cell microRNA transcriptome," *Genome Research*, vol. 20, no. 11, pp. 1590–1604, (2010);
- Fenichel, G. M., J. M. Florence, et al. "Long-term benefit from prednisone therapy in Duchenne muscular dystrophy." *Neurology* 41(12): 1874-7. (1991).
- Ferner RE, Gutmann DH. "International consensus statement on malignant peripheral nerve sheath tumors in neurofibromatosis 1". *Cancer Res* 62:1573-1577. (2002).
- Gargioli, C., M. Coletta, et al. "PlGF-MMP-9-expressing cells restore microcirculation and efficacy of cell therapy in aged dystrophic muscle." *Nat Med* 14(9): 973-8. (2008).
- Gilbert W, Maxam A., "The nucleotide sequence of the lac operator". *Proc Natl Acad Sci U S A*. Dec;70(12):3581-4. (1973).
- Goedde, H. W., I. Christ, et al. "Creatine kinase isoenzyme patterns in Duchenne muscular dystrophy." *Clin Genet* 14(5): 257-60. (1978).

Golan David e Paul Medvedev, "Using state machines to model the Ion Torrent sequencing process and to improve read error rates". *Bioinformatics*, Vol. 29 ISMB/ECCB, pages i344–i351; (2013).

Gorecki, D. C., A. P. Monaco, et al. "Expression of four alternative dystrophin transcripts in brain regions regulated by different promoters." *Hum Mol Genet* 1(7): 505-10. (1992).

Goswami, H. K. and B. D. Chaurasia "Mutation rate in Duchenne muscular dystrophy." *J Med Genet* 8(3): 310-1. (1971).

Griggs, R. C., R. T. Moxley, 3rd, et al. "Duchenne dystrophy: randomized, controlled trial of prednisone (18 months) and azathioprine (12 months)." *Neurology* 43(3 Pt 1): 520-7. (1993).

Griggs, R. C., R. T. Moxley, 3rd, et al. "Prednisone in Duchenne dystrophy. A randomized, controlled trial defining the time course and dose response. Clinical Investigation of Duchenne Dystrophy Group." *Arch Neurol* 48(4): 383-8. (1991).

Guo HF, The I, Hannan F, Bernards A, Zhong Y. Requirement of *Drosophila* NF1 for activation of adenylyl cyclase by PACAP-like neuropeptides. *Science*; 276: 795-798. (1997).

Gutmann DH, Geist RT, Douglas E, Wright E, Snider WD. Expression of the neurofibromatosis type 1 (NF1) isoforms in developing and adult rat tissues. *Cell growth and Differentiation*; 6:315-323. (1995).

Han S, Cooper D, Upadhyaya M. "Evaluation of denaturing high performance liquid chromatography (DHPLC) for the mutational analysis of the neurofibromatosis type 1 (NF1) gene". *Hum. Genet.* 109:487-497. (2001).

Hattori S, Maekawa M, Nakamura S. "Identification of neurofibromatosis type 1 gene product as an insoluble GTP-ase-activating protein toward ras p21". *Oncogene*; 7:481-5. (1992).

Hefferon TW, Groman JD, Yurk CE and Cutting GR. "A variable dinucleotide repeat in the CFTR gene contributes to phenotype diversity by forming RNA secondary structures that alter splicing". *Proc Natl Acad Sci USA* 101, 3504–3509. (2004).

Hirst, R. C., K. J. McCullagh, et al. "Utrophin upregulation in Duchenne muscular dystrophy." *Acta Myol* 24(3): 209-16. (2005).

Hoffman, E. P. "Genotype/phenotype correlations in Duchenne/Becker dystrophy." *Mol Cell Biol Hum Dis Ser* 3: 12-36. (1993).

Hoffman, E. P. and L. Schwartz "Dystrophin and disease." *Mol Aspects Med* 12(3): 175-94. (1991).

Hoffman, E. P., R. H. Brown, Jr., et al. "Dystrophin: the protein product of the Duchenne muscular dystrophy locus." *Cell* 51(6): 919-28. (1987).

Howard J. Jacob, Ph.D. "Next-Generation Sequencing for Clinical Diagnostics". *The New England Journal of Medicine* Downloaded from [nejm.org](http://nejm.org) on October 7, (2013).

[http://en.wikipedia.org/wiki/DNA\\_sequencing/;](http://en.wikipedia.org/wiki/DNA_sequencing/)

[http://en.wikipedia.org/wiki/DNA\\_sequencing/;](http://en.wikipedia.org/wiki/DNA_sequencing/)

<http://mygenomix.wordpress.com;>

<http://mygenomix.wordpress.com;>

[http://www.gsjunior.com/instrument-workflow.php;](http://www.gsjunior.com/instrument-workflow.php)

Huber CG, Oefner PJ, Bonn GK. "High resolution liquid chromatography of oligonucleotides on nonporous alkylated styrene-divinylbenzene copolymers". *Anal. Biochem.*; 212: 351-358. (1993).

Ion AmpliSeq™ Library Preparation Protocol Catalog Numbers 4475345, 4480441, 4480442, 4479790, and 4482298 Publication Part Number MAN0006735 Revision 5.0;

Ion OneTouch™ 200 Template Kit v2 DL Catalog Number 4480285 Publication Number MAN0006957 Revision 6.0;

Ion PGMTM Sequencing 200 Kit Protocol Catalog Number 4474004 Publication Number 4474246 Revision H;

Izawa I, Tamaki N, Saya H. "Phosphorylation of neurofibromatosis type 1 gene product (neurofibromin) by cAMP-dependent protein kinase". *FEBS Lett*; 382: 53-59. (1996).

Jellett, L.B., Kennedy MC., et al. "Duchenne pseudohypertrophic muscular dystrophy: a clinical and electrocardiographic study of patients and female carriers." *Aust N Z J Med* 4(1): 41-7. (1974).

Jones, G. E. and J. A. Witkowski "Analysis of skin fibroblast aggregation in Duchenne muscular dystrophy." *J Cell Sci* 48: 291-300. (1981).

Judge, L. M. and J. S. Chamberlain "Gene therapy for Duchenne muscular dystrophy: AAV leads the way." *Acta Myol* 24(3): 184-93. (2005).

Kai Wang, Mingyao Li and Hakon Hakonarson "ANNOVAR: functional annotation of genetic variants from high-throughput sequencing. *Nucleic Acids Res.* September, 38(16) e164; (2010).

Kayes LM, Burke W, Riccardi VM, Bennett R, Ehrlich P, Rubenstein A, Stephens K. "Deletions spanning the neurofibromatosis 1 gene:

- identification and phenotype of five patients". *Am J Hum Genet.*; 54: 424-436. (1994).
- Kim HA, Ratner N, Roberts TM, Stiles CD., "Schwann cell proliferative responses to cAMP and NF1 are mediated by Cyclin D1". *J Neurosci*; 21: 1110- 16. (2001).
- Knudson AG Jr. "Mutation and cancer: statistical study of retinoblastoma". *Proc Natl Acad Sci USA*; 68: 820-823. (1971).
- Kobayashi, T., S. Mawatari, et al. "Lipids and proteins of erythrocyte membrane in Duchenne muscular dystrophy." *Clin Chim Acta* 85(3): 259-66. (1978).
- Koshimizu E, Satoko M, Nobuhiko O, Mitsuko N, Yoshinori T, Noriko M, Hiroto S, Naomichi M., "Performance Comparison of Bench-Top Next Generation Sequencers Using Microdroplet PCR-Based Enrichment for Targeted Sequencing in Patients with Autism Spectrum Disorder". *PLOS ONE*, September, Volume 8; (2013).
- Kunkel, L. M. and E. P. Hoffman "Duchenne/Becker muscular dystrophy: a short overview of the gene, the protein, and current diagnostics." *Br Med Bull* 45(3): 630-43. (1989).
- Kunkel, L., G. Burns, et al. "Genetic analysis of Duchenne dystrophy." *Adv Exp Med Biol* 182: 287-94. (1985).
- Lagrota-Candido, J., R. Vasconcellos, et al. "Resolution of skeletal muscle inflammation in mdx dystrophic mouse is accompanied by increased immunoglobulin and interferon-gamma production." *Int J Exp Pathol* 83(3): 121-32. (2002).

Lakkis MM, Tennekoon GI, "Neurofibromatosis type 1". I. General overview. *Journal of Neuroscience Research*; 62: 755-763. (2000).

Ledbetter DH, Rich DC, O'Connell P, Leppert M, Carey JC. "Precise localization of NF1 to 17q11.2 by balanced translocation". *Am J Hum Genet*. Jan; 44(1): 20-24. (1989).

Li Y, O'Connell P, Breidenbach HH, Cawthon R, Stevens J, Xu G, Neil S, Robertson M, White R, Viskochil D. Genomic organization of the neurofibromatosis 1 gene (NF1). *Genomics*; 25: 9-18. (1995).

Lin L, Yinhu L, Siliang L, Ni H, Yimin H, Ray P, Danni L, Lihua L, and Law L, "Comparison of Next-Generation Sequencing Systems". Review Article, *Journal of Biomedicine and Biotechnology*, Article ID 251364, 2 April. (2012).

Lin Z, Thomas NJ, Wang Y, Guo X, Seifart C, Shakoor H, Floros J. "Deletions within a CA-Repeat-rich region of intron 4 of the human SP-B gene affect mRNA splicing". *Biochem J*. 389:403-12. (2005).

Love, D. R., S. M. Forrest, et al. "Molecular analysis of Duchenne and Becker muscular dystrophies." *Br Med Bull* 45(3): 659-80. (1989).

MacArthur, D. G. and K. N. North "A gene for speed? The evolution and function of alpha-actinin-3." *Bioessays* 26(7): 786-95. (2004).

Mandel JL, "Dystrophin. The gene and its product". *Nature*. Jun 22;339(6226):584-6. (1989).

Manole E., "The dystrophin gene and its product a view". *Rom J Neurol Psychiatry*. Apr-Jun;33(2):109-19. (1995).

Mardis E. R., "The impact of next-generation sequencing technology on genetics," *Trends in Genetics*, vol. 24, no. 3, pp. 133-141, 2008;

- Markham, L. W., R. L. Spicer, et al.. "Steroid therapy and cardiac function in Duchenne muscular dystrophy." *Pediatr Cardiol* 26(6): 768-71. (2005).
- Matlin AJ, Clark F, Smith CW. "Understanding alternative splicing: towards a cellular code". *Nat Rev Mol Cell Biol* 6:386-398. (2005).
- Mattocks C, Baralle D, Tarpey P, French-Constant C, Bobrow M, Whittaker J. "Automated comparative sequence analysis identifies mutations in 89% of NF1 patients and confirms a mutation cluster in exons 11-17 distinct from the GAP related domain". *J Med Genet*; 41: e48. (2004).
- McCormick F. "Ras GTPase activating protein: signal transmitter and signal terminator". *Cell*; 56: 5-8. (1989).
- McGuire, S. A. "Diagnosing Duchenne Muscular Dystrophy." *West J Med* 150(5): 575. (1989).
- Messiaen LM, Callens T, Mortier G, Beysen D, Vandebroucke I, Van Roy N, Speleman F, De Paepe A. Exhaustive mutation analysis of the NF1 gene allows identification of 95% of mutations and reveals a high frequency of unusual splicing defects. *Human Mutation*. 15: 541-555. (2000).
- Metherell LA, Akker SA, Munroe PB, Rose SJ, Caulfield M, Savage MO, Chew SL, Clark AJ. Pseudoexon activation as a novel mechanism for disease resulting in atypical growth-hormone insensitivity. *Am. J. Hum. Genet*. 69, 641-646. (2001).
- Michael A. Quail, Miriam Smith, Paul Coupland, Thomas D Otto, Simon R Harris, Thomas R Connor, Anna Bertoni, Harold P Swerdlow and

- Yong Gu, "A tale of three next generation sequencing platforms: comparison of Ion Torrent, Pacific Biosciences and Illumina MiSeq sequencers". *BMC Genomics*, 13:341. (2012).
- Michalak M, Opas M., "Functions of dystrophin and dystrophin associated proteins". *Curr Opin Neurol*. Oct;10(5):436-42. (1997).
- Mills R.E.,K.Walter,C.Stewartetal.,"Mappingcopynumber variation by population-scale genome sequencing," *Nature*, vol. 470, no. 7332, pp. 59-65. (2011).
- Molecular analysis of two uncharacterized sequence variants of the VHL gene. *J Hum Genet* 51, 964-968. (2006).
- Monaco, A. P. "Dystrophin, the protein product of the Duchenne/Becker muscular dystrophy gene." *Trends Biochem Sci* 14(10): 412-5. (1989).
- Monica Forzan, "Functional Splicing Assay mediante l'utilizzo di minigeni plasmidici nel gene NF1", Tesi di Dottorato, XX Ciclo;
- Nakamura Kensuke, Taku Oshima, Takuya Morimoto, Shun Ikeda, Hirofumi Yoshikawa, Yuh Shiwa, Shu Ishikawa, Margaret C. Linak, Aki Hirai, Hiroki Takahashi, Md. Altaf-Ul-Amin, Naotake Ogasawara and Shigehiko Kanaya<sup>1</sup> "Sequence-specific error profile of Illumina sequencers". *Nucleic Acids Research*, 2011, Vol. 39, No. 13 e90, 16 May (2011).
- Newcombe RG. Two-sided confidence intervals for the single proportion: comparison of seven methods. *Stat Med* 17, 857-872. (1998).



- Norman, A. M., H. E. Hughes, et al. "Dystrophin analysis in the diagnosis of muscular dystrophy." *Arch Dis Child* 64(10): 1501-3. (1989).
- O'Donovan MC, Oefner PJ, Roberts SC, Austin J, Hoogendoorn B, Guy C, Speight G, Upadhyaya M, Sommer SS, McGuffin P. Blind analysis of denaturing high-performance liquid chromatography as a tool for mutation detection. *Genomics*; 52: 44-49. (1998).
- Pagani F, Baralle FE. Genomic variants in exons and introns: identifying the splicing spoilers. *Nat Rev Genet.* May;5(5):389-96. Review. (2004).
- Pagani F, Buratti E, Stuani C, and Baralle FE. Missense, nonsense, and neutral mutations define juxtaposed regulatory elements of splicing in cystic fibrosis transmembrane regulator exon 9. *J Biol Chem* 278: 26580-26588. (2003).
- Pagani F, Buratti E, Stuani C, Bendix R, Dork T, and Baralle FE. A new type of mutation causes a splicing defect in ATM. *Nat Genet* 30: 426-429. (2002).
- Pagani F, Buratti E, Stuani C, Romano M, Zuccato E, Niksic M, Giglio L, Faraguna D, Baralle FE. Splicing factors induce cystic fibrosis transmembrane regulator exon 9 skipping through a nonevolutionary conserved intronic element. *J Biol Chem.* 275:21041-7. (2000).
- Pagani F, Stuani C, Zuccato E, Kornblihtt AR and Baralle FE. Promoter architecture modulates CFTR exon 9 skipping. *J Biol Chem* 278, 1511-1517. (2003).

- Pearce, P. H., R. D. Johnsen, et al. "Muscle lipids in Duchenne muscular dystrophy." *Aust J Exp Biol Med Sci* 59(1): 77-90. (1981).
- Pena, D. J. "Is Duchenne muscular dystrophy a simple genic disorder." *Lancet* 2(8090): 630. (1978).
- Perry, S. V. "Duchenne muscular dystrophy." *Biochem Soc Trans* 12(3): 362-5. (1984).
- Pfaffl, M. W., G. W. Horgan, et al. "Relative expression software tool (REST) for group-wise comparison and statistical analysis of relative expression results in real-time PCR." *Nucleic Acids Res* 30(9): e36. (2002).
- Pierno, S., B. Nico, et al. "Role of tumour necrosis factor alpha, but not of cyclo-oxygenase-2-derived eicosanoids, on functional and morphological indices of dystrophic progression in mdx mice: a pharmacological approach." *Neuropathol Appl Neurobiol* 33(3): 344-59. (2007).
- Pros E, Gómez C, Martín T, Fàbregas P, Serra E, Làzaro C. Nature and mRNA effect of 282 different NF1 point mutation: focus on splicing alteration. *Human Mutation*; 26: E173-E193. (2008).
- Quackenbush, J. "Microarray data normalization and transformation." *Nat Genet* 32 Suppl: 496-501. (2002).
- Radley, H. G., M. J. Davies, et al. "Reduced muscle necrosis and long-term benefits in dystrophic mdx mice after cV1q (blockade of TNF) treatment." *Neuromuscul Disord* 18(3): 227-38. (2008).

Raponi M, Upadhyaya M and Baralle D. Functional splicing assay show a pathogenetic intronic mutation in Neurofibromatosis type 1 (NF1) due to intronic sequence exonization. *Hum. Mut.* 27:294-295. (2006).

Reed GH, Wittwer CT. Sensitivity and Specificity of Single-Nucleotide Polymorphism Scanning by High-Resolution Melting Analysis. *Clinical Chemistry* 50:10; 1748–1754. (2004).

Rizzo Stefano Giovanni “Una base dati per il knowledge discovery in genetica medica”, A.A. 2012-2013.

Ruggieri M e Tenconi R. Le Neurofibromatosi. Associazione Linfa, Lottiamo Insieme per la Neurofibromatosi- ONLUS- (2001).

Sacconi S, Salviati L, Nishigaki Y, Walker WF, Hernandez-Rosa E, Trevisson E et al. A functionally dominant mitochondrial DNA mutation. *Hum Mol Genet* 17, 1814–1820. (2008).

Sacheck, J. M., A. Ohtsuka, et al. "IGF-I stimulates muscle growth by suppressing protein breakdown and expression of atrophy-related ubiquitin ligases, atrogin-1 and MuRF1." *Am J Physiol Endocrinol Metab* 287(4): E591-601. (2004).

Samaha, F. J., B. Davis, et al. "Duchenne muscular dystrophy: adenosine triphosphate and creatine phosphate content in muscle." *Neurology* 31(7): 916-9. (1981).

Sandri, M., C. Sandri, et al. "Foxo transcription factors induce the atrophy- related ubiquitin ligase atrogin-1 and cause skeletal muscle atrophy." *Cell* 117(3): 399- 412. (2004).

- Schmidt MA, Michels VV, Dewald GW, Opitz JM, Reynolds JF. Cases of neurofibromatosis with rearrangements of chromosome 17 involving band 17q11.2. *Am J Med Genet*; 28: 771-777. (1987).
- Seizinger BR, Rouleau GA, Ozelius LJ, Lane AH, Faryniarz AG, Chao MV, Huson S, Korf BR, Parry DM et al. Genetic linkage of von Recklinghausen neurofibromatosis to the nerve growth factor receptor gene. *Cell*; 49(5): 589-594. (1987).
- Shen MH, Harper PS, Upadhyaya M. Molecular genetics of neurofibromatosis type 1 (NF1). *J Med Genet*; 33: 2-17. (1996).
- Skovgaard O., M. Bak, A. Løbner-Olesen et al., "Genome- wide detection of chromosomal rearrangements, indels, and mutations in circular chromosomes by short read sequencing," *Genome Research*, vol. 21, no. 8, pp. 1388–1393. (2011).
- Stern, L. M., J. D. Fewings, et al. "The progression of Duchenne muscular dystrophy: clinical trial of allopurinol therapy." *Neurology* 31(4): 422-6. (1981).
- Sun H and Chasin LA. Multiple splicing defect in an intronic false exon. *Mol. Cell. Biol.* 20: 6414–6425. (2000).
- Takagi, A. and I. Nonaka "Duchenne muscular dystrophy: unusual activation of single fibers in vitro." *Muscle Nerve* 4(1): 10-5. (1981).
- Tews, D. S. "Tumour necrosis factor-mediated cell death pathways do not contribute to muscle fibre death in dystrophinopathies." *Acta Neuropathol* 109(2): 217- 25. (2005).

- Thorvaldsdottir H, Robinson J.T. and Mesirov J. P., Integrative Genomics Viewer (IGV): high-performance genomics data. RIEFINGS IN BIOINFORMATICS. VOL 14. NO 2. 178-192, 14 March (2012).
- Tomkins, J. K. "The molecular defect in Duchenne muscular dystrophy." Muscle Nerve 3(6): 529-30. (1980).
- Treisman R, Orkin SH, Maniatis T. Specific transcription and RNA splicing defects in five cloned beta-thalassemia genes. Nature 302, 591-596. (1983)
- Tzetis M, Efthymiadou A, Doudounakis S and Kanavakis E. Qualitative and quantitative analysis of mRNA associated with four putative splicing mutations (621+3A-4G, 2751+2T-4A, 296+1G-4C, 1717-9T-4C-D565G) and one nonsense mutation (E822X) in the CFTR gene. Hum Genet 109, 592-601. (2001).
- Upadhyaya M, Osborn MJ, Maynard J, Kim MR, Tamanoi F, Cooper DN. Mutational and functional analysis of the neurofibromatosis type 1 (NF1) gene. Hum Genet; 99: 88-92. (1997).
- Upadhyaya M, Shaw DJ and Harper PS. Molecular basis of neurofibromatosis type 1 (NF1): mutation analysis and polymorphisms in the NF1 gene. Human Mutation 4:83-101. (1994).
- Van Gelder, R. N., M. E. von Zastrow, et al. "Amplified RNA synthesized from limited quantities of heterogeneous cDNA." Proc Natl Acad Sci U S A 87(5): 1663-7. (1990).
- Vandenbroucke I, Van Oostveldt P, Coene E, De Paepe A, Messiaen L. Neurofibromin is actively transported to the nucleus. FEBS Lett.; 560 (1-3): 98-102. (2004).

- Vasli N, Böhm J, Le Gras S, Muller J, Pizot C, Jost B, Echaniz-Laguna A, Laugel V, Tranchant C, Bernard R, Plewniak F, Vicaire S, Levy N, Chelly J, Mandel JL, Biancalana V, Laporte J. "Next generation sequencing for molecular diagnosis of neuromuscular diseases". *Acta Neuropathol.* Aug;124(2):273-83. (2012).
- Viskochil D, White R, Cawthon R. The Neurofibromatosis type 1 gene. *Annu. Rev. Neurosci.* 16:183-205. (1993).
- White H, Potts G. Mutation scanning by high resolution melt analysis. Evaluation of RotorGene™ 6000 (Corbett Life Science), HR1™ and 384 well LightScanner™ (Idaho Technology). NGRLW\_HRM\_1.0 (2006).
- White R, Nakamura Y, O'Connell P, Leppert M, Lalouet JM, Barker D, Golgar D, Skolnick M, Carey J, Wallis CE. Tightly linked markers for the neurofibromatosis type 1 gene. *Genomics* 1:364-7. (1987).
- Wiest V, Eisenbarth I, Schmegner C, Krone W, Assum G. Somatic NF1 mutation spectra in a family with Neurofibromatosis type 1: toward a theory of genetic modifiers. *Hum Mutat.*; 22: 423-427. (2003).
- Yan Denise, Mustafa Tekin, Susan H. Blanton, and Xue Zhong Liu Next-Generation Sequencing in Genetic Hearing Loss GENETIC TESTING AND MOLECULAR BIOMARKERS Volume 17, Number 8, a Mary Ann Liebert, Inc. Pp. 581–587. (2013).
- Yang, N., D. G. MacArthur, et al. "ACTN3 genotype is associated with human elite athletic performance." *Am J Hum Genet* 73(3): 627-31. (2003).

Yaping Yang, Ph.D., Donna M. Muzny, M.Sc., Jeffrey G. Reid, Ph.D., Matthew N. Bainbridge, Ph.D., Alecia Willis, Ph.D., Patricia A. Ward, M.S., Alicia Braxton, M.S., Joke Beuten, Ph.D., Fan Xia, Ph.D., Zhiyv Niu, Ph.D., Matthew Hardison, Ph.D., Richard Person, Ph.D., Mir Reza Bekheirnia, M.D., Magalie S. Leduc, Ph.D., Amelia Kirby, M.D., Peter Pham, M.Sc., Jennifer Scull, Ph.D., Min Wang, Ph.D., Yan Ding, M.D., Sharon E. Plon, M.D., Ph.D., James R. Lupski, M.D., Ph.D., Arthur L. Beaudet, M.D., Richard A. Gibbs, Ph.D., and Christine M. Eng, M.D. Clinical Whole-Exome Sequencing for the Diagnosis of Mendelian Disorders *The new england journal of medicine* *The New England Journal of Medicine* Downloaded from [nejm.org](http://nejm.org) on October 7, (2013).

Zhen Xuan Yeo, Maurice Chan, Yoon Sim Yap, Peter Ang, Steve Rozen, Ann Siew Gek Lee Improving Indel Detection Specificity of the Ion Torrent PGM Benchtop Sequencer *PLOS ONE* September, Volume 7; (2012).

Zhu, Y. Q. and X. D. Tan "TFF3 modulates NF- $\kappa$ B and a novel negative regulatory molecule of NF- $\kappa$ B in intestinal epithelial cells via a mechanism distinct from TNF- $\alpha$ ." *Am J Physiol Cell Physiol* 289(5): C1085-93. (2005).

Zielenski J, Bozon D, Markiewicz D, Aubin G, Simard F, Rommens JM et al. "Analysis of CFTR transcripts in nasal epithelial cells and lymphoblasts of a cystic fibrosis patient with 621+1G-4T and 711+1G-4T mutations". *Hum Mol Genet* 2, 683-687. (1993).

Zuccato E, Buratti E, Stuani C, Baralle FE and Pagani F. "An intronic polypyrimidine-rich element downstream of the donor site modulates cystic fibrosis transmembrane conductance regulator exon 9 alternative splicing". *J Biol Chem* 279, 16980–16988. (2004).



## **Ringraziamenti:**

La prima persona che sento di dover ringraziare è il Prof.<sup>re</sup> Renzo Galanello che ho avuto l'onore ed il piacere di conoscere e dal quale spero d'aver imparato l'amore e la dedizione per il lavoro e la ricerca.

La Prof.<sup>ssa</sup> Maria Antonietta Melis che mi ha permesso di muovere i primi passi verso la strada estera della ricerca.

Il Prof.<sup>re</sup> Egbert Bakker, i Dott.<sup>ri</sup> Cornelis L. Harteveld, Piero C. Giordano e Rolf H. A. M. Vossen del Leiden University Medical Center che mi hanno consentito di apprendere un nuovo approccio alla ricerca ed anche un valido insegnamento sul profilo umano.

La Dott.<sup>ssa</sup> Milena Cau che mi ha seguita lungo il percorso dei miei tre anni di dottorato.

Ringrazio il Prof.<sup>re</sup> Paolo Moi che ha avuto fiducia in me e mi ha dato, insieme alla Dott.<sup>ssa</sup> Marisella Addis, l'occasione di terminare il mio dottorato al Weill Cornell Medical College, presso il laboratorio del Prof.<sup>re</sup> Stefano Rivella, in cui sto avendo la possibilità di conoscere validi colleghi e di iniziare a muovere i primi passi in quello che potrebbe esser il mio cammino futuro.

Ringrazio tutti i colleghi e gli amici, dei Laboratori di Genetica Umana, Ematologia e Citologia dell'Ospedale Microcitemico, incontrati durante questi tre anni di studi e da ciascuno dei quali ho imparato qualcosa.

Inoltre ringrazio la mia famiglia per avermi sempre supportata, per essermi stata vicino e per avermi sostenuta nelle mie scelte.

Vi ho tutti nel mio cuore.