



Università degli Studi di Ferrara

DOTTORATO DI RICERCA IN
FARMACOLOGIA E ONCOLOGIA
MOLECOLARE
INDIRIZZO ONCOLOGICO

XXV CICLO

*STUDIO DELL'ESOMA MEDIANTE TECNOLOGIE DI
GENOTIPIZZAZIONE AD ALTA EFFICIENZA:
SEQUENZIAMENTO DI NUOVA GENERAZIONE (NGS)
e IBRIDAZIONE GENOMICA COMPARATIVA (CGH),
PER L'IDENTIFICAZIONE DI NUOVI GENI MALATTIA
IN PATOLOGIE NEUROMUSCOLARI.*

Settore Scientifico Disciplinare MED/03

Tutore

Prof.ssa FERLINI ALESSANDRA

Dottorando

Dott.ssa SCOTTON CHIARA

CoTutore

Dott.ssa GUALANDI FRANCESCA

Anni 2009/2012

SOMMARIO

ABSTRACT.....	4
RIASSUNTO.....	5
INTRODUZIONE	6
1. Sequenziamento di nuova generazione.....	6
1.1 Avvento della tecnologia di nuova generazione.....	6
1.2 Piattaforme di nuova generazione	7
1.3 Analisi dei dati NGS.....	14
1.4 Vantaggi delle tecnologie NGS	15
1.5 Limiti delle tecnologie NGS	16
1.6 Applicazioni delle tecnologie NGS	17
2. Miopatia di Bethlem.....	18
3. Miopatia miofibrillare	20
MATERIALI E METODI	22
1. Selezione dei campioni.....	22
2. Estrazione del DNA.....	23
3. Analisi di linkage	23
4. Analisi di sequenza dei geni noti	23
5. Array-CGH	24
6. Sequenziamento dell'esoma mediante tecnologia NGS.....	24
7. PIPELINE di analisi dei dati WES.....	25
8. Analisi di sequenza del gene candidato	26
RISULTATI.....	27
1. Famiglia Bethlem	27
1.1 Analisi di linkage.....	27
1.2 Array-CGH.....	28
1.3 Sequenziamento dell'esoma mediante tecnologia NGS	28
1.4 Analisi di sequenza del gene candidato.....	29
2. Miopatia miofibrillare	32
2.1 Analisi di sequenza dei geni noti	32
2.2 Array-CGH.....	32
2.3 Sequenziamento dell'esoma mediante tecnologia NGS	33

DISCUSSIONE.....40
CONCLUSIONI44
BIBLIOGRAFIA.....46

ABSTRACT

Over the years many different approaches and techniques have been employed to get insight genetic data of family and patients. The first approach for genetic studies and gene discovery was the linkage analysis, but to be efficient it required large family or large numbers of patients sharing the same disease phenotype.

The advent of sequencing technology made the genetic analysis more handy but still it was time consuming and not cost effective when a large number of genes needed to be screened , for example in case of diseases with a known genetic heterogeneity as the neuromuscular disorders (NMDs).

The high throughput molecular diagnostics tools such as Comparative Genomic Hybridization (CGH) and next Generation Sequencing (NGS) technology are changing medical genomics by accelerating new disease causing mutations discovery; these techniques could enable quick, reliable and cost-effective analysis of numerous NMD genes in parallel.

The NGS methods promise to speed up the discovery of the genetic causes of diseases both in the research and the clinical setting.

We performed whole exome sequencing analysis (WES) through NGS technology on a family with a Bethlem phenotype (BM) orphan of mutations in COLVI genes and a cohort of patients with a clinical diagnosis of myofibrillar myopathy (MFM).

We performed the linkage analysis on BM family; the linkage regions identified were used as filters in WES output data. We selected four components (two affected and two unaffected) of this family and performed Whole Exome Sequencing by Illumina GAIIE platform obtaining a few candidate genes.

Regarding the MFMs patients, we identified a large rearrangements in laminin alpha 2 (LAMA2) gene through CGH; while WES identified small variations in five patients: mutations in a known gene, and two variations in two novel genes previously unreported as involved in MFMs.

RIASSUNTO

Nel corso degli anni diversi approcci e tecniche sono stati applicati a famiglie o pazienti sporadici al fine di ottenere informazioni genetiche. L'analisi di linkage è stata la prima metodica utilizzata per identificare nuovi geni-malattia, tuttavia questa tecnica necessita di ampie famiglie o di un numero elevato di casi sporadici.

L'avvento della tecnologia di sequenziamento ha reso l'analisi genetica più maneggevole, ma tempi di analisi e costi rimanevano tuttavia elevati, soprattutto nell'indagine di patologie ad elevata eterogeneità genetica come le patologie neuromuscolari (NMD).

In questi ultimi anni si sta osservando una rivoluzione nella genetica medica sia dal punto di vista diagnostico che di ricerca scientifica applicata all'identificazione di nuovi geni-malattia, grazie alla scoperta di strumenti ad elevata efficienza quali l'ibridazione genomica comparativa (CGH) e il sequenziamento di nuova generazione (NGS); rapidità, affidabilità e costi limitati sono i vantaggi peculiari di queste tecniche che potrebbero consentire l'analisi in parallelo di numerosi geni associati a NMD.

In questa tesi viene descritta l'analisi dell'intero esoma (WES) mediante tecnologia NGS, applicata ad una famiglia con fenotipo Bethlem (BM), senza mutazione nei geni correlati alla patologia, e ad una coorte di pazienti con diagnosi clinica di miopatia miofibrillare (MFM).

La famiglia BM è stata studiata dapprima mediante analisi di linkage, usata in un secondo momento come filtro per l'analisi dei dati, e in seguito in quattro individui selezionati (due affetti e due sani) è stato analizzato tutto l'esoma tramite la piattaforma NGS Illumina GAIIE. Questa sequenza d'indagine ha permesso l'identificazione di un limitato numero di geni candidati.

Nei pazienti MFM, invece, un primo risultato eclatante concerne l'identificazione, mediante CGH, di un riarrangiamento del numero di copie nel gene LAMA2 (laminina alfa 2); mentre l'applicazione della tecnologia NGS ha permesso la determinazione di mutazioni a carico di un gene noto essere associato alla miopatia e di variazioni a carico di due geni coinvolti nei meccanismi di contrazione e rigenerazione muscolare, ma non precedentemente associati alla patologia.

INTRODUZIONE

1. SEQUENZIAMENTO DI NUOVA GENERAZIONE

1.1 AVVENTO DELLA TECNOLOGIA DI NUOVA GENERAZIONE

Negli ultimi sessanta anni si è osservato un notevole incremento delle conoscenze riguardanti il genoma umano e il codice genetico, che spaziano dalla scoperta della struttura del DNA nel 1953 fino alla pubblicazione della prima bozza del genoma umano nel 2001 (Lander et al. 2001, Venter et al. 2001).

Il sequenziamento Sanger (Sanger et al. 1977), noto anche come sequenziamento di prima generazione, è stato il metodo utilizzato per sequenziare il genoma all'interno del progetto "Genoma umano" portando all'intera sequenza genomica nel 2003, dopo 13 anni dall'inizio del progetto con un costo di 3 miliardi di dollari e il contributo di 6 diverse nazioni.

La realizzazione del progetto "Genoma umano" ha avuto esito positivo non solo grazie alle innovazioni scientifiche nel campo della biologia ma anche alle implementazioni informatiche che hanno portato allo sviluppo di calcolatori ad alta efficienza. L'evoluzione in parallelo di questi due settori ha consentito l'avvento di piattaforme per il sequenziamento di nuova generazione (NGS) più efficienti ed economiche.

Oggi con questa nuova tecnologia il genoma umano può essere sequenziato con soli dieci mila dollari in un tempo stimato di tre mesi (Bick and Dimmock 2011), tuttavia costi e tempi vengono fortemente ridotti analizzando solo le regioni codificanti. Nonostante l'esoma corrisponda solamente all'1% dell'intero genoma, nelle patologie monogeniche l'85% delle mutazioni causative si possono identificare sequenziando gli esoni e le regioni ad essi fiancheggiati (Botstein and Risch 2003).

Questo presupposto richiederebbe, ai fini di identificare un nuovo gene malattia, per un sequenziamento di tipo Sanger, la conoscenza di geni target per la diagnostica e uno stretto numero di geni candidati. La tecnologia di sequenziamento dell'esoma mediante NGS bypassa entrambe queste necessità catturando e permettendo il sequenziamento dell'intero esoma (Mamanova et al. 2010).

1.2 PIATTAFORME DI NUOVA GENERAZIONE

La prima piattaforma NGS ad entrare nel mercato nel 2005, il 454 Roche, deriva dalla convergenza di due metodiche già precedentemente descritte: un'amplificazione del DNA in micro-compartimenti (Tawfik and Griffiths, 1998) e il sequenziamento basato sulla misurazione del pirofosfato rilasciato durante la sintesi della molecola (Nyren et al. 1993).

Succeivamente l'aggiornamento o l'evoluzione di nuove metodiche ha introdotto altre piattaforme NGS, tutte caratterizzate da due importanti fasi: uno step "biologico", che va dalla preparazione del campione al vero e proprio sequenziamento che consente di svolgere più reazioni in contemporanea con una minore manualità rispetto alla metodica Sanger, e una fase informatica di analisi dei dati ottenuti.

La prima parte del processo si divide in tre fasi: la preparazione del campione, l'amplificazione e il sequenziamento.

La preparazione del campione, che è il passaggio comune a tutte le piattaforme, consiste nella frammentazione del DNA genomico in frammenti di piccole dimensioni che variano dalle 100 alle 500 paia di basi (bp) in relazione alla piattaforma utilizzata; ai frammenti vengono aggiunti degli adattori che serviranno per la fase successiva di amplificazione del frammento genomico.

L'amplificazione si basa su due esclusive metodiche: la PCR in emulsione o l'amplificazione su base solida. Nella PCR in emulsione, descritta per la prima volta da Tawfik a Griffiths, le singole molecole di DNA vengono amplificate clonalmente in micro-compartimenti costituiti da miscele di acqua e olio. Gli adattatori legati alle singole molecole di DNA si ibridano a sequenze complementari che rivestono la superficie di specifiche biglie in grado di legare covalentemente il DNA, e ad ogni biglia si legherà un singolo frammento di DNA che, mediante la cattura in un'emulsione di olio-acqua, viene amplificato in modo clonale (fig.1).

Anche nell'amplificazione su base solida i frammenti di DNA vengono immobilizzati tramite il legame con sequenze complementari all'adattatore, ma su una superficie solida simile ad un vetrino. I frammenti vengono amplificati mediante la formazione di un ponte dato dal ripiegamento dei filamenti di DNA che si ibridano ad un nucleotide di ancoraggio adiacente nel vetrino. Cicli successivi di amplificazione portano alla formazione di "cluster" di frammenti ripiegati amplificati clonalmente composti circa da 1000 ampliconi clonali (fig. 1).

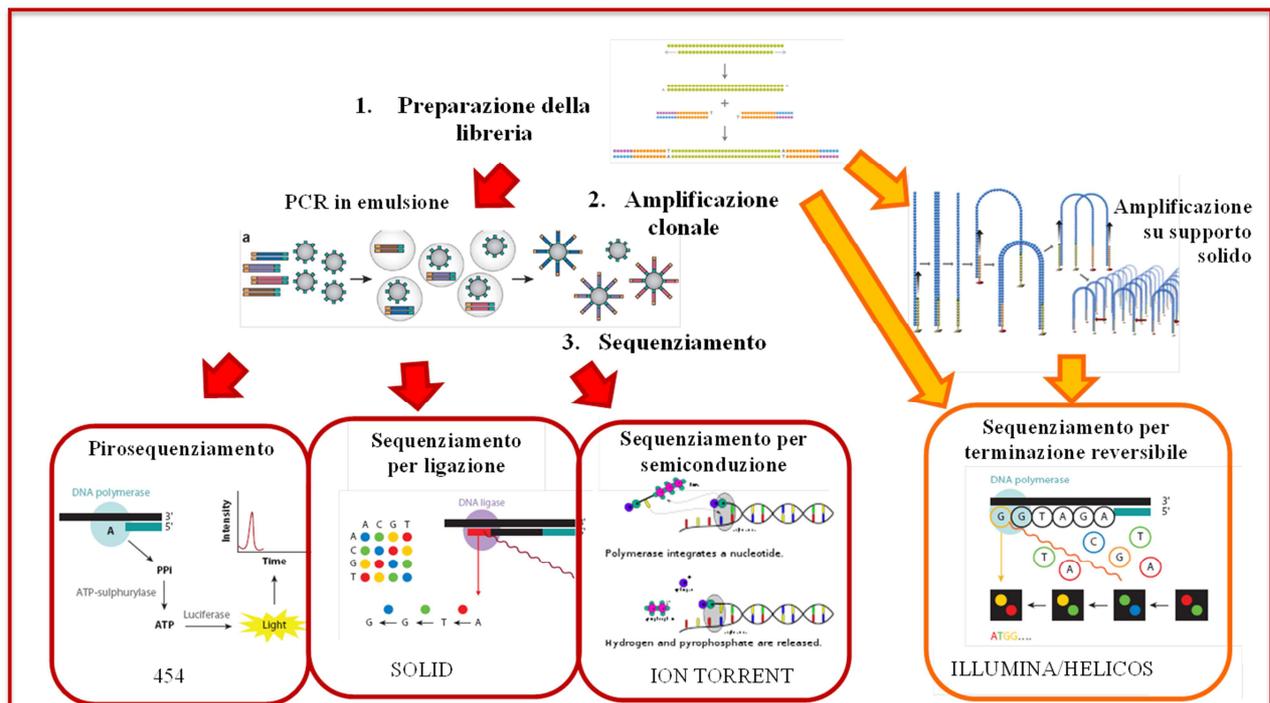


Fig. 1: schema illustrativo delle metodiche di sequenziamento.

Tutte le piattaforme NGS in commercio sono caratterizzate dalla capacità di sequenziare parallelamente e massivamente molecole di DNA amplificate in modo clonale o singole.

Diversamente dalla metodica Sanger, dove i frammenti di diverse dimensioni ottenuti da singole reazioni di sequenziamento venivano separati elettroforicamente, nelle tecnologie NGS il sequenziamento viene operato mediante la ripetizione di cicli di estensione nucleotidica o di ligazione di oligonucleotidi. Il principio su cui si basa il sequenziamento e l'acquisizione dell'immagine sono gli step peculiari che caratterizzano le diverse piattaforme in commercio.

Ad oggi sono noti quattro diversi metodi di sequenziamento:

1. Sequenziamento per terminazione ciclica reversibile (CRT). Questa tipologia di sequenziamento utilizza dei nucleotidi terminatori marcati reversibilmente (fig. 2) Ogni ciclo di sequenziamento comprende: l'incorporazione del nucleotide, l'acquisizione della fluorescenza e il taglio del nucleotide. La DNA polimerasi aggiunge alla sequenza di innesco un nucleotide modificato complementare al template e questa incorporazione provoca il termine della reazione di estensione. Il successivo step di lavaggio consente l'eliminazione di tutti i nucleotidi non incorporati. A questo punto viene catturata l'immagine per identificare il nucleotide incorporato, seguito da uno step di clivaggio che

rimuove il gruppo terminatore e il fluorocromo del nucleotide appena analizzato. Dopo l'eliminazione dei gruppi funzionali clivati, la polimerasi può continuare la reazione di estensione e legare il secondo nucleotide.

Questo processo è utilizzato da due tipi di piattaforme commerciali: Illumina ed Helicos, che differiscono per il template di partenza del sequenziamento. Mentre Illumina utilizza frammenti amplificati clonalmente su base solida, Helicos è attualmente l'unica piattaforma commerciale in grado di usare singole molecole di DNA non amplificate. Inoltre la piattaforma Illumina utilizza un'acquisizione a 4 colori (i quattro nucleotidi terminatori reversibili sono marcati con un diverso fluorocromo e vengono dispensati in contemporanea nel sequenziatore), mentre nella piattaforma Helicos tutti i nucleotidi sono marcati con lo stesso fluorocromo e vengono dispensati nel sequenziatore in un determinato ordine gerarchico.

(Fig. 3) (Metzker 2009).

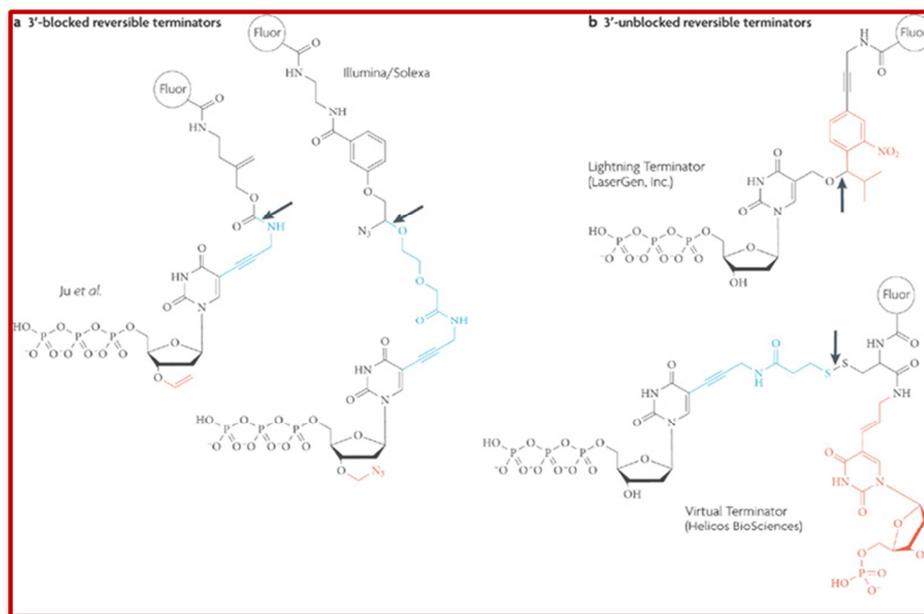


Fig. 2: rappresentazione dei nucleotidi marcati utilizzati nel sequenziamento per terminazione reversibile. La struttura in rosso evidenzia i gruppi funzionali responsabili della terminazione della sintesi. Le frecce indicano il sito di taglio che separa il fluorocromo dal resto del nucleotide. La struttura in blu indica il gruppo funzionale che dopo il taglio viene attaccato dal nucleotide successivo. (Metzker 2009).

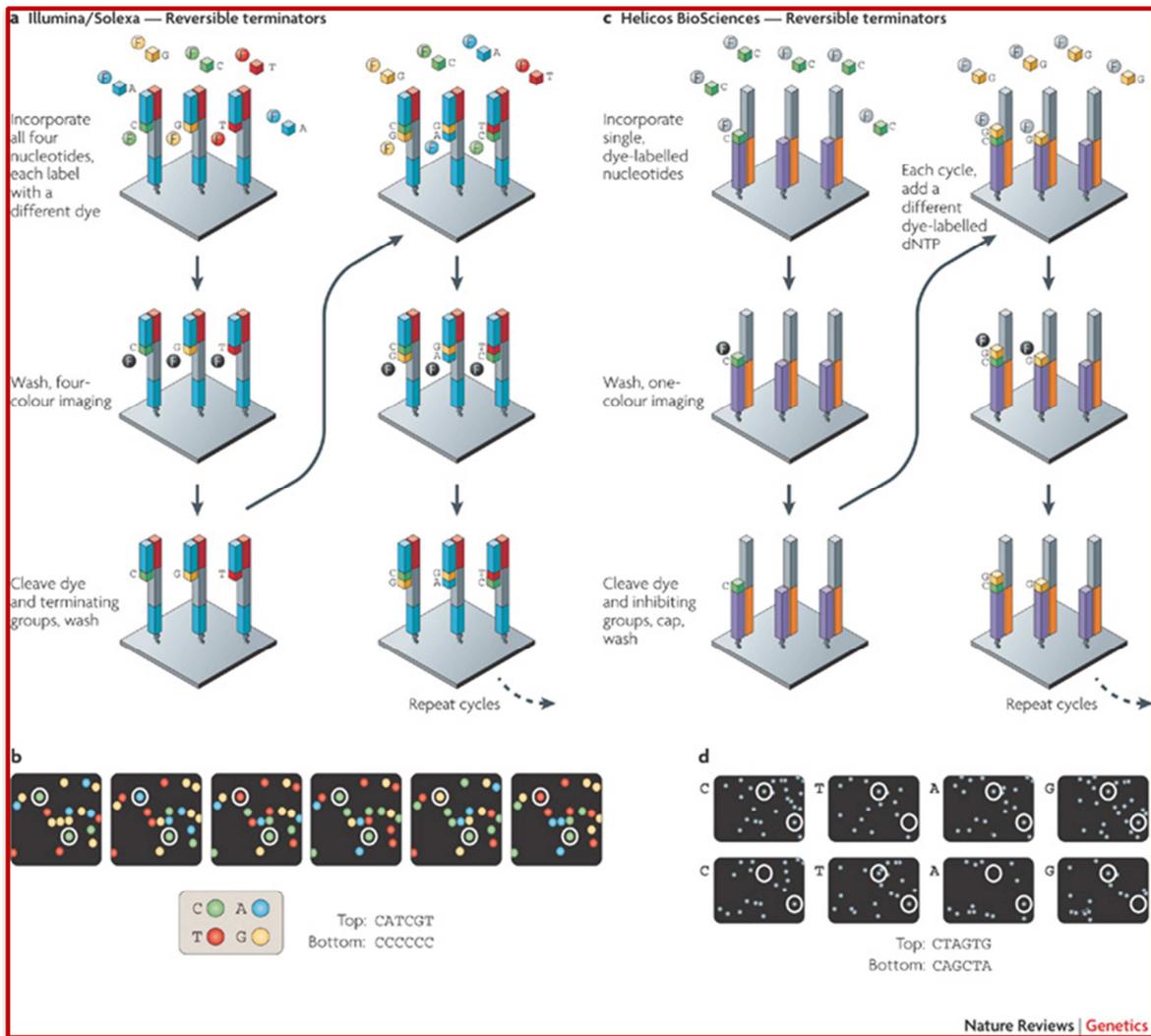


Fig. 3: sequenziamento per terminazione reversibile basato su acquisizione a 4 colori (Illumina) o ad un colore (Helicos) (Metzker 2009).

2. Sequenziamento per ligazione. Questo metodo differisce dal precedente per l'utilizzo della DNA ligasi e per l'uso di sonde codificanti 2 basi. Ogni sonda è formata da un ottamero costituito (in direzione 3'→5') da 2 basi specifiche seguite da 6 basi degenerate (nnnzzz) con uno dei 4 marcatori fluorescenti legato all'estremità 5'. Il primer che dovrà essere esteso è orientato in modo da esporre un gruppo fosfato al 5' a cui si legherà l'opportuna sonda per opera della DNA ligasi. Nel primo passaggio di ligazione vengono immesse: una ligasi termostabile e le 16 sonde che rappresentano tutte le combinazioni di 2 basi possibili; dopo il legame delle sonde e un lavaggio per

eliminare quelle non ibridate, viene rilevata la fluorescenza e clivata la porzione marcata delle sonde in modo da rigenerare un gruppo fosfato al 5' e ripetere il ciclo. Il filamento sintetizzato alla fine di questo ciclo viene successivamente denaturato e ibridato con primer sfasati di una base rispetto al ciclo precedente per un totale di 5 "round". Questo tipo di sequenziamento è caratteristico della piattaforma Solid dell' Applied Biosystem, che utilizza come metodo di amplificazione una PCR in emulsione. Questo tipo di approccio presenta il vantaggio di poter sequenziare due volte ogni nucleotide del DNA stampo. (Fig. 4) (Voelkerding 2009).

3.

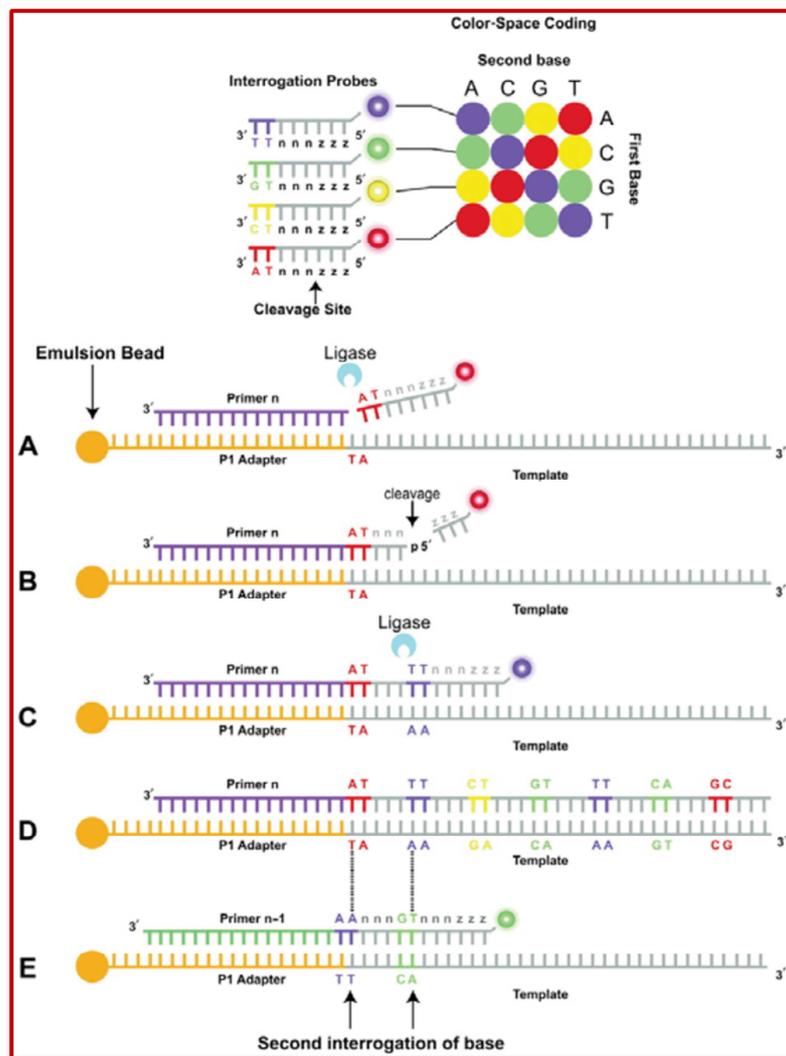


Fig. 4: sequenziamento per ligazione specifico della piattaforma Solid dell'Applied Biosystem.

4. Il pirosequenziamento. È una tecnica basata sulla rilevazione del pirofosfato (PPi) rilasciato durante la sintesi del DNA. Il PPi rilasciato dopo l'incorporazione di un nucleotide viene convertito in ATP per opera della ATP sulfurylasi. Utilizzando la fonte di energia così prodotta la luciferasi, presente nella reazione di sequenza, genera un segnale luminescente. Non vi è una differenza di segnale immesso tra i quattro nucleotidi ma questi vengono immessi nel sequenziatore in flussi separati seguiti da cicli di lavaggio per eliminare quello non incorporato (Fig. 5) (Ronaghi 2001).

Questa tecnologia, come già citato, è stata la prima ad essere immessa nel mercato accoppiando al pirosequenziamento un' amplificazione mediante PCR in emulsione ed è caratteristica della piattaforma 454 Roche.

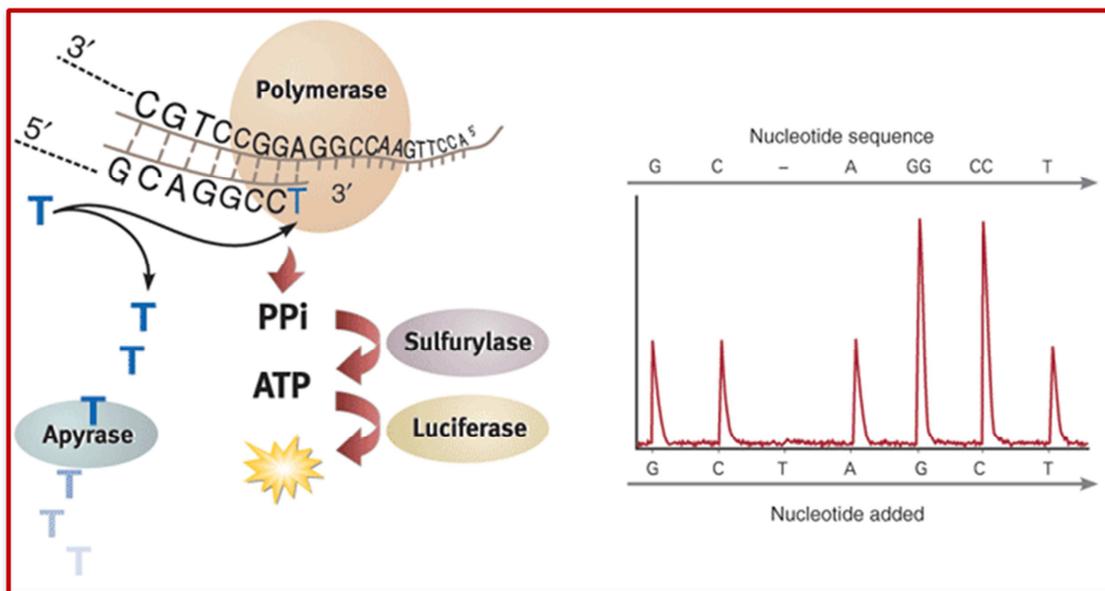


Fig. 5: Rappresentazione della metodica del pirosequenziamento.

5. Sequenziamento per semiconduzione. Questa tecnica, specifica della piattaforma Ion Torrent (sempre dell'Applied Biosystem), è definita tecnologia di terza generazione per l'abbattimento ulteriore dei costi da essa prodotta. Consiste nella trasformazione del segnale chimico in segnale digitale grazie alla presenza di un chip semiconduttore. Il principio peculiare riguarda la rilevazione di uno ione idrogeno rilasciato normalmente come sottoprodotto nel corso della replicazione del DNA, quando la polimerasi incorpora all'interno del nuovo filamento di DNA un nucleotide. Il rilascio del protone provoca un'alterazione del pH della soluzione che può essere misurata; anche in questo caso vi è l'emissione sequenziale dei flussi di nucleotidi seguiti da cicli di lavaggio. (Fig. 6) (Rothberg 2011)

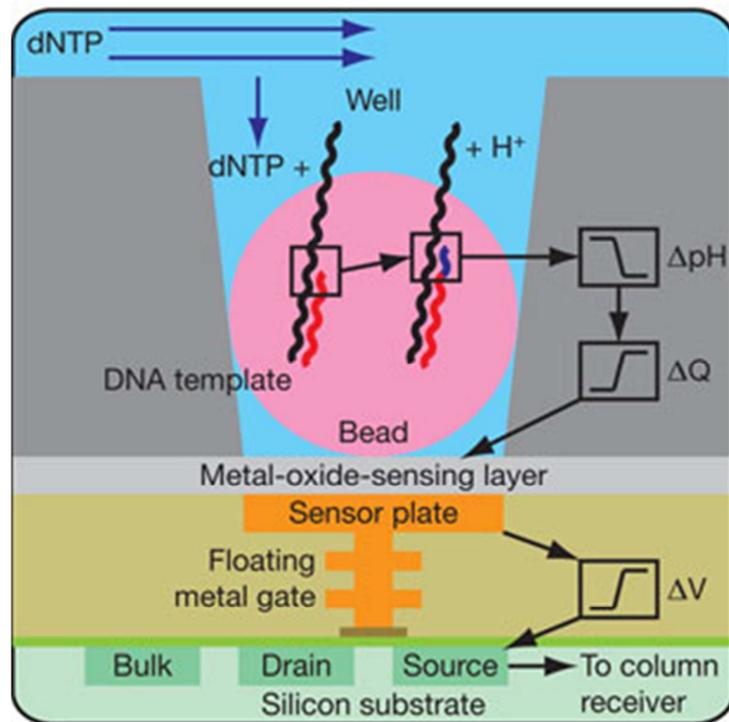


Fig. 6: rappresentazione grafica del sequenziamento per semiconduzione (Rothberg et al. 2011)

1.3 ANALISI DEI DATI NGS

La mole di dati prodotta dalle piattaforme NGS si colloca nell'ordine delle Terabyte (TB), rappresentando una difficoltà per l'analisi e lo stoccaggio dei dati.

I software di analisi differiscono in base alla tecnologia NGS utilizzata nel sequenziamento, ma seguono tutti un sistema di "pipeline" di analisi dei dati che ha il compito di convertire le immagini di luminescenza o fluorescenza acquisite in sequenze nucleotidiche ("reads"). A questo processo, definito "base calling", segue l'assegnazione di un punteggio di qualità ("quality score") a ciascun nucleotide, che indica la probabilità di errore ad esso associata.

I "quality score" costituiscono un importante strumento per eliminare dal processo di analisi basi o "reads" che non superano i parametri opportuni, migliorando in tal modo l'accuratezza della "pipeline" successiva, ovvero l'allineamento della sequenza ottenuta con sequenze di riferimento e l'annotazione di basi varianti ("variant calls") (Li et al. 2008).

Per un'adeguata efficienza la "pipeline" di allineamento necessita di "reads" superiori a 30 bp: infatti soltanto il 90% del genoma umano può essere allineato in modo univoco con reads di 30 bp (Whiteford et al. 2005). Un altro limite per l'allineamento è rappresentato dalle sequenze ripetute, la cui soluzione consta nell'assegnazione delle "reads" di sequenze ripetute a posizioni multiple nel genoma di riferimento oppure nella creazione di lacune ("gaps") nell'allineamento.

I tassi di errore associati alla tecnologia di nuova generazione sembrano maggiori rispetto alla tradizionale metodica Sanger; tuttavia l'accuratezza del sequenziamento nelle piattaforme NGS è assicurata da una lettura ripetuta e massiva di ogni frammento genico, che determina la "copertura" del genoma ("coverage"), dato appunto dalla somma delle "reads" sovrapposte che coprono il genoma.

Anche quest'ultimo parametro rappresenta un fattore essenziale nell'analisi, in quanto un "coverage" non adeguato (in generale indicato con valore inferiore a 20-50 volte in relazione alla piattaforma utilizzata) può determinare la presenza di falsi negativi, per campioni eterozigoti, nel rilevamento di una effettiva variazione nucleotidica (Wheeler et al. 2008).

Al termine del processo di annotazione delle varianti, la moltitudine di variazioni identificate sono soggette all'applicazione di vari filtri volti a ridurre il numero di candidati. I filtri più comunemente usati si basano su:

- modello di ereditarietà (autosomico/X-linked; dominante/recessivo),
- condivisione di variazioni uguali in famiglie ben caratterizzate
- condivisione di geni colpiti da diverse variazioni per casi non relati.

- rimozione o meno di varianti già note mediante l'uso di database pubblici (dbSNPs; 1000 genomi)
- focalizzazione in base alla potenziale dannosità della variazione (non senso, missenso, variazioni dei siti di splicing o inserzioni e delezioni che alterano la cornice di lettura)
- predizione degli effetti funzionali delle variazioni mediante l'uso di strumenti bioinformatici in silico come SIFT, POLYPHEN, ANNOVAR.

1.4 VANTAGGI DELLE TECNOLOGIE NGS

Nonostante nei decenni precedenti lo sviluppo della metodica Sanger abbia portato miglioramenti nella chimica, nell'automazione e nella miniaturizzazione del processo, non è stato e non sarà possibile un abbassamento dei costi e un aumento della produzione di dati di sequenziamento, per cui l'innovazione legata all'avvento delle tecnologie NGS ha offerto notevoli vantaggi.

L'incremento di volume di dati prodotti e la riduzione dei costi associati alle piattaforme NGS è la conseguenza di tre fattori:

- 1- Molte migliaia/milioni di reazioni di sequenza possono essere condotte in parallelo superando il limite delle 1-96 reazioni possibili con i tradizionali macchinari di sequenziamento.
- 2- Il clonaggio o l'amplificazione del frammento di DNA risultano, nelle nuove tecnologie, non necessari o completamente automatizzati all'interno delle piattaforme
- 3- Un altro vantaggio del sequenziamento massivo in parallelo è la capacità di rilevare con alta accuratezza l'allele minore, che si riflette nell'identificazione di una variazione anche in campioni a mosaico o di delezioni in eterozigosi. Infatti il numero di volte in cui un frammento di DNA viene amplificato e sequenziato è proporzionale all'abbondanza di tale segmento nel campione originale, per cui l'utilizzo di specifici algoritmi nell'analisi dei dati può portare all'identificazione anche di variazioni del numero di copie (Tucker et al. 2009).

1.5 LIMITI DELLE TECNOLOGIE NGS

La tecnologia NGS presenta tuttavia dei limiti legati più che altro alla magnitudine di dati prodotti; infatti nei risultati di NGS possono essere presenti sia falsi positivi che falsi negativi.

I falsi positivi possono derivare da:

- un non corretto allineamento con la sequenza genomica di riferimento. È possibile superare questo limite mediante l'applicazione ai dati di diversi software di allineamento.
- errori di sequenziamento sistematici. Questo tipo di errore può comunque essere identificato in quanto presente in tutti i campioni tra loro non relati e può quindi essere rimosso dalla lista finale delle variazioni identificate
- errori riconducibili al limite tecnico della macchina. Ad esempio, nel pirosequenziamento vi è un errore ricorrente nella lettura degli omopolimeri di lunghezza superiore alle 5-6 basi

I falsi negativi derivano invece da:

- la presenza di un basso coverage
- bassa copertura di arricchimento nelle regioni di interesse
- allineamento di regioni ripetute

La riduzione del numero di errori nelle piattaforme NGS può essere raggiunta mediante un incremento del coverage e quindi della qualità della corsa e la frammentazione del DNA in frammenti di maggiore estensione. Infatti le "reads" corte rendono difficoltoso l'allineamento e sono di difficile interpretazione nella determinazione della fase di lettura; l'implementazione del sequenziamento "paired-end", che è in grado di sequenziare frammenti di maggiore lunghezza da entrambe le estremità, permette l'analisi di frammenti con lunghezza di 5-10kb in relazione alla piattaforma utilizzata.

Il sequenziamento di tipo Sanger è comunque richiesto alla termine dell'analisi dei dati perché i risultati ottenuti mediante NGS necessitano di una validazione (Lin et al. 2012)

1.6 APPLICAZIONI DELLE TECNOLOGIE NGS

Le applicazioni ad oggi conosciute sono nello studio sia del genoma che del trascrittoma.

il genoma può essere valutato globalmente, solamente nelle regioni codificanti o mediante arricchimenti specifici solo in determinate regioni target. Il principale utilizzo è legato all'indagine delle regioni codificanti, in quanto il maggior numero delle patologie sono causate da mutazioni esoniche o di splicing che alterano la corretta sequenza amminoacidica delle proteine; infatti l'esoma, pur costituendo solamente l'1% dell'intero genoma umano, ospita l'85% delle mutazioni patogenetiche (Teer and Mullikin 2010).

Il numero di mutazioni note essere associate ad una patologia supera le 110.000 variazioni, in più di 3700 geni diversi, ed è stato stimato che solamente la metà delle patologie mendeliane nell'uomo hanno una base genetica nota (Bick and Dimmock 2011).

Per questi motivi la ricerca scientifica si concentra per lo più nell'esoma per l'identificazione di nuovi geni-malattia. L'approccio comunemente usato nel passato per identificare i loci trasmessi in associazione al fenotipo patologico era l'analisi di segregazione o di linkage, che aveva lo scopo di identificare le regioni geniche condivise negli individui affetti all'interno di famiglie con un ampio numero di membri.

L'avvento della tecnologia NGS con l'approccio improntato sul sequenziamento dell'intero esoma ("Whole Exome Sequencing", WES) ha innescato una rapida ascesa nell'identificazione di nuovi geni malattia, avendo il considerevole vantaggio di necessitare di un numero limitato di campioni.

La prima applicazione di successo di tale metodica ha portato all'identificazione del gene DHODH come causativo della sindrome di Miller (Ng et al. 2010). Dal 2010 ad oggi un'elevata quantità di studi ha identificato nuove mutazioni patogenetiche, tramite il sequenziamento dell'esoma in un numero ridotto di individui affetti, in vari tipi di patologie tra cui casi di neuropatia, poichilodermia associata a neutropenia, vitreoretinopatia essudativa familiare, disordini immunologici, disordini intellettivi, predisposizione tumorale (Lin et al. 2012).

L'utilizzo del WES ha portato a determinare mutazioni causative anche in patologie con elevata eterogeneità fenotipica dove il tradizionale approccio di analisi di linkage risulta essere più difficoltoso (Isidor et al. 2011; Wang et al. 2010).

Un nuovo settore di sviluppo della tecnologia NGS è legato all'identificazione di biomarcatori di severità fenotipica (Ley et al. 2010) o alla farmagenomica con lo sviluppo di terapie personalizzate (Wesolowska et al. 2011).

In questa tesi vengono approfonditi sia l'approccio volto alla ricerca di nuovi geni in una famiglia con miopatia di Bethlem, sia l'approccio diagnostico mediante lo studio di un gruppo di casi affetti da miopatia miofibrillare.

2. MIOPATIA DI BETHLEM

La Miopatia di Bethlem (BM), descritta per la prima volta nel 1976 da Bethlem e Wijngaarden in 28 individui appartenenti a tre famiglie tedesche (Bethlem and Wijngaarden 1976), è caratterizzata da moderata miopatia prossimale con atrofia dei muscoli estensori, più coinvolti rispetto ai muscoli flessori (Jobsis et al. 1999).

Normalmente i pazienti diventano sintomatici tra la prima e la seconda decade di vita. Tuttavia molti, già adulti, rimangono pauci o asintomatici e l'età di esordio può non essere stabilita con precisione (Merlini et al. 1994).

Una caratteristica peculiare della miopatia è la presenza di contratture che possono presentarsi in varie articolazioni specialmente a livello dei flessori delle dita, dei polsi, dei gomiti e delle caviglie, e questo, sommato alla debolezza, contribuisce alla disabilità. In molti casi la patologia è lentamente progressiva e più dei due terzi dei pazienti al di sopra dei 50 anni necessita di supporti per la deambulazione. Anche se il coinvolgimento cardiaco è solitamente assente (Bonnemann et al. 2003), il coinvolgimento dei muscoli respiratori in particolare quelli diaframmatici necessita di supporto per la ventilazione notturna (Haq et al. 1999).

A livello immunoistochimico, la biopsia muscolare solitamente non è alterata rispetto alla condizione fisiologica, ma talvolta può presentare variazioni nel diametro delle fibre muscolari con incremento del tessuto adiposo e occasionalmente fibre distrofiche o necrotiche (Merlini et al. 1994).

La prima correlazione tra fenotipo e genotipo, ad opera di Jobsis et al. nel 1996, identifica come gene causativo della miopatia il collagene di tipo VI (Jobsis et al. 1996)

Il Collagene VI è una proteina ubiquitaria espressa fortemente a livello della matrice extracellulare, dove forma un esteso network di microfilamenti che interagiscono con altri componenti della lamina basale come collagene IV, fibronectina, biglicano, decorina e perlecano, oltre a recettori transmembrana come integrine e proteoglicani che collegano lo spazio pericellulare allo spazio intracellulare; questo complesso pathway di interazioni ha l'importante compito di permettere il mantenimento dell'integrità cellulare. (Fig. 7)

Il collagene VI è una proteina costituita da tre catene peptidiche, alfa1(VI), alfa2(VI) e alfa3(VI), codificate rispettivamente dai geni COL6A1 e COL6A2, localizzati testa-coda nella regione subtelomerica del cromosoma 21q22.3, e dal gene COL6A3, situato nella regione subtelomerica del cromosoma 2q37.3

Mutazioni a carico dei geni del collagene VI possono essere associati, oltre che alla Miopatia di Bethlem, anche alla forma allelica più grave: la Distrofia Muscolare Congenita di Ullrich (UCMD), descritta per la prima volta da Ullrich (Ullrich et al. 1930) e caratterizzata da un esordio precoce con debolezza muscolare, contratture prossimali e iperlassità distale. Questa forma è velocemente progressiva e porta a un'insufficienza respiratoria nella prima o seconda decade di vita che è la causa di mortalità più frequente, a meno che i pazienti non vengano trattati con supporti notturni per la respirazione (Mercuri et al. 2002).

Il numero di variazioni identificate nei geni del collagene VI ha evidenziato una quota rilevante sia di polimorfismi che di mutazioni patogenetiche prevalentemente private con una difficile interpretazione della correlazioni genotipo-fenotipo, in quanto medesime mutazioni sono state ritrovate sia in pazienti BM che in UCMD. In ogni modo la "detection rate" non arriva a superare una sensibilità del 75%. Le possibili cause di tale sensibilità sono da ricondurre all'eterogeneità allelica nonché genetica (Lampe AK & Bushby KM 2005).

Considerando l'ipotesi dell'eterogeneità genetica associata a tale patologia, in questa tesi viene presentato il caso di un'ampia famiglia affetta da BM in cui mediante la tecnologia NGS si è ricercato un possibile nuovo gene malattia.

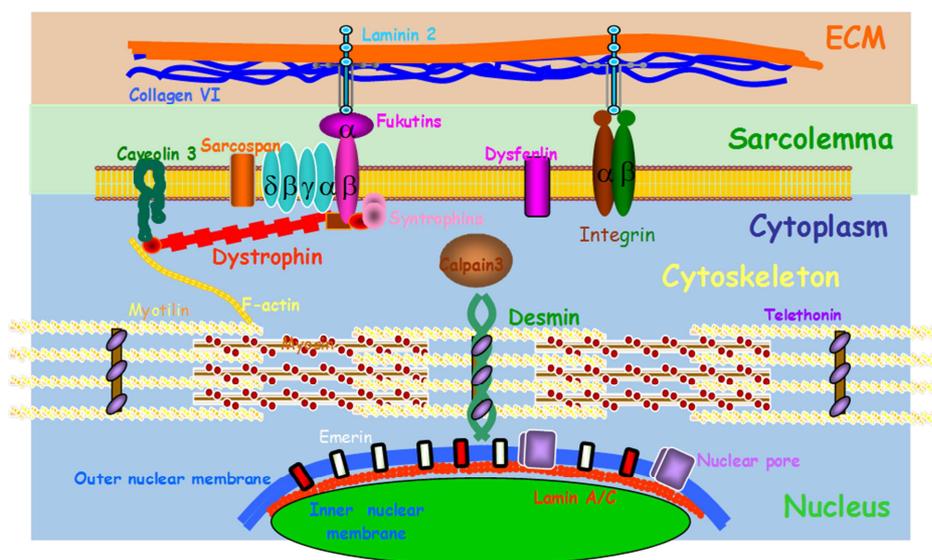


Fig. 7: rappresentazione della localizzazione del collagene VI e delle proteine con cui interagisce

3. MIOPATIA MIOFIBRILLARE

Il termine di miopatia miofibrillare (MFM) è stato coniato 15 anni fa per classificare dei disordini neuromuscolari geneticamente eterogenei con caratteristiche morfologiche peculiari.

L'incidenza complessiva della patologia nella popolazione è di difficile valutazione; è però noto come l'MFM, che provoca una disabilità cronica a lungo termine comportando un notevole onere ai pazienti, alle famiglie e alla sanità pubblica, colpisca sia adulti che bambini.

Dal punto di vista clinico gli affetti presentano un'eterogeneità di sintomi e segni clinici, dei quali il più comune è una lenta e progressiva debolezza muscolare, in cui la debolezza distale è più pronunciata rispetto a quella prossimale ed è avvertita dall'80% degli individui. Altri segni di minore entità sono sintomi sensoriali, rigidità muscolare, dolori, crampi e una neuropatia periferica presente in circa il 20% degli individui affetti; solo nel 15% -30% dei casi si riscontra una cardiomiopatia conclamata con morte prematura che può derivare da coinvolgimento del muscolo cardiaco e/o respiratorio (Schroder and Schoser 2009).

Attualmente la diagnosi di MFM è effettuata mediante la biopsia muscolare, con cui si valuta innanzitutto la presenza di fibre anormali con depositi amorfi, granulari o di sostanza ialina caratterizzati da forma e dimensioni irregolari e che provocano una riduzione o perdita dell'attività enzimatica ossidativa che può essere misurata (Claeys et al. 2008). Dal punto di vista immunostochimico, si valuta la varietà delle proteine accumulate all'interno dei depositi (Claeys et al. 2008). Infine esami ultrastrutturali identificano una alterazione delle miofibrille e dei dischi Z con materiale granulo-filamentoso intercalato tra le miofibrille (Nakano et al. 1996)

Il meccanismo mediante il quale le proteine formino degli aggregati non è ancora completamente chiaro nonostante si ipotizzi la partecipazione di diversi fattori tra i quali un sistema ubiquitina-proteosoma difettoso, presenza di proteine alterate, regolazione anomala di alcuni fattori di trascrizione e stress ossidativo (Ferrer and Olivè, 2008).

L'eterogeneità clinica è accompagnata da un'eterogeneità genetica; infatti attualmente sono noti sei geni causativi: desmina (OMIM #601419), $\alpha\beta$ -cristallina (#608810), miotilina (#609200), ZASP (#609452), filamina C (#609524) e BAG3 (#612954). La maggior parte delle mutazioni riscontrate sono de-novo e associate ad un modello ereditario di tipo autosomico dominante (Goldfarb and Dalakas, 2009).

Attualmente la diagnosi genetica con metodica Sanger è dispendiosa sia in termini di tempo che di costi; inoltre nel 50% dei casi non viene identificata la variazione patogenetica. In questo lavoro sono state ricercate mutazioni in una coorte di 19 pazienti affetti da MFM

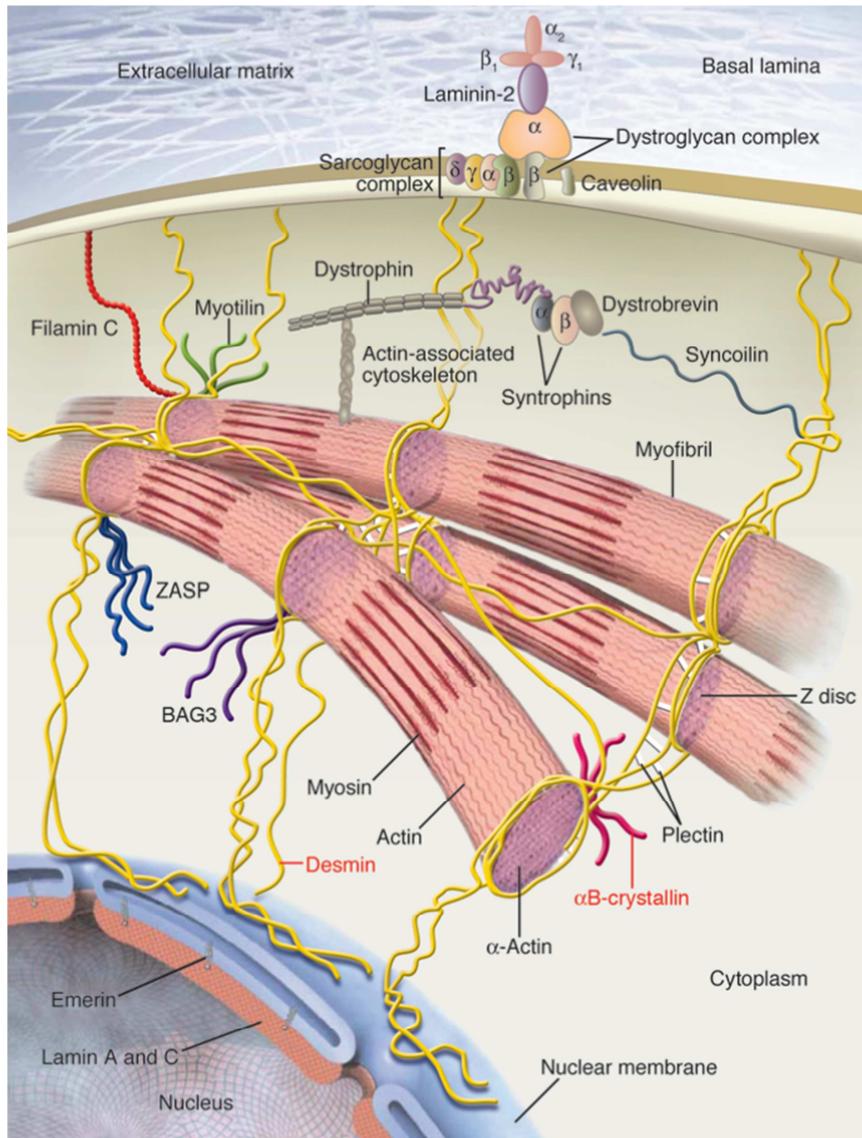


Fig. 8: rappresentazione delle miofibrille con le proteine coinvolte nelle miopatie miofibrillari. (Goldfarb and Dalakas 2009)

MATERIALI E METODI

1. Selezione dei campioni

Abbiamo selezionato una famiglia BM ben caratterizzata con 12 individui di cui 6 affetti (Fig. 9). Il pedigree suggerisce un modello di ereditarietà autosomico dominante. Nel probando della famiglia sono state escluse, nei tre geni del collagene (COL6A1, COL6A2, COL6A3) responsabili dei fenotipi BM e UCMD, mutazioni puntiformi mediante sequenziamento e grossi riarrangiamenti mediante un CGH-specifico (Bovolenta et al. 2010).

I 19 casi di MFM sono stati selezionati con un' accurata diagnosi clinica: tutti i pazienti mostravano la presenza di fibre alterate per forma e dimensione con la presenza di depositi granulari in immunostochimica, l'analisi di microscopia elettronica rivelava la frammentazione delle miofibrille e dei dischi Z con la presenza di materiale granulo-filamentoso intercalato tra le miofibrille.

Ad eccezione di due pazienti, in cui è riscontrabile una familiarità, tutti sono casi sporadici de novo.

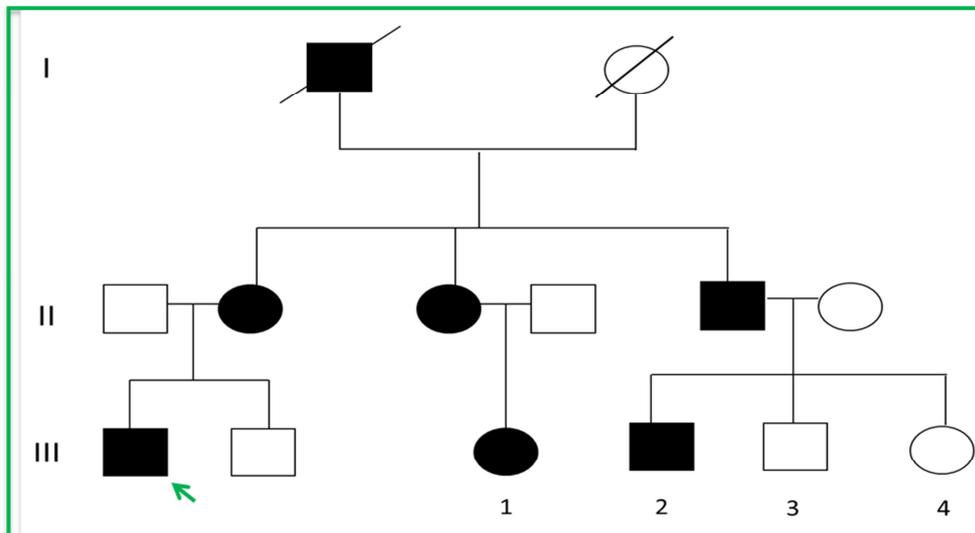


Fig. 9: rappresentazione della famiglia BM con 12 individui, 6 dei quali affetti.

La freccia indica il probando, mentre i numeri (1-4) indicano i campioni corsi mediante NGS.

2. Estrazione del DNA

Il DNA della famiglia BM e dei pazienti MFM è stato estratto da sangue intero mediante kit nucleon BACC3 (GE Healthcare Life Sciences, UK) in modo da ottenere DNA ad alto peso molecolare, non degradato e con rapporti 260/280 di circa 1.8 e 260/230 di circa 2.0. La qualità del DNA è stata valutata mediante Nanodrop ND-1000 (NanoDrop Technologies, Wilmington, DE).

3. Analisi di Linkage

Per tutta la famiglia BM è stata effettuata un'analisi di linkage mediante ABI PRISM® Linkage Mapping Set Version 2.5 MD10 (Applied Biosystems, Foster City, CA).

Il Linkage Mapping Set Version 2.5 MD10 contiene pannelli di coppie di primer marcati covalentemente con differenti fluorofori che generano prodotti di PCR che possono essere rilevati per iniezione capillare su sequenziatore automatico (ABIPRISM 3130, Applied Biosystems, Foster City, CA), definendo una risoluzione nella mappa dei cromosomi umani di 10 cM.

È stato escluso dall'analisi il cromosoma X in quanto il pedigree mostra un'evidente trasmissione autosomica.

4. Analisi di sequenza dei geni noti

In tutti i 19 campioni affetti da MFM è stata esaminata l'intera sequenza codificante, alla ricerca di mutazioni causative, nei seguenti geni correlati alla patologia: desmina (DES), miotilina (MYOT), $\alpha\beta$ -cristallina (CRYAB), ZASP (LIM domain binding 3, LDB3) e BAG3. Il gene della filamina C (FLNC) non è stato esaminato a causa della ampia dimensione genica associata ad una bassa "detection rate".

5. Array-CGH

Questa metodica si basa sulla competizione, per il legame a loci specifici, di due acidi nucleici genomici marcati con fluorocromi diversi (Cy3-Cy5): il DNA campione e un DNA genomico di riferimento. I due DNA vengono mescolati e ibridati contemporaneamente sul vetrino dell'array; per ogni locus sarà rilevata la fluorescenza corrispondente al DNA più rappresentato. Il risultato è espresso dal rapporto tra le due fluorescenze. L'intensità di fluorescenza emessa per ogni locus del chip è rilevata da analizzatori d'immagine computerizzati (scanner) e quantificata calcolando la deviazione dai valori attesi del rapporto tra i segnali emessi dal DNA campione e dal DNA di riferimento.

Sono stati ibridati nell'array tutti i 19 campioni MFM e il probando della famiglia BM.

Il disegno dell'array-CGH Roche-Nimblegen, con 135.000 sonde per subarray, che copre 50 geni coinvolti in patologie neuromuscolari (NMD), miopatie congenite e distrofie muscolari congenite (Kaplan 2011), è stato utilizzato all'interno del progetto europeo NMD-chip seguendo il protocollo standard descritto nel manuale versione 7.0 disponibile nel sito (<http://www.nimblegen.com/>).

Il campione è stato marcato con Cy3, mentre il DNA di controllo (maschio o femmina Promega) con Cy5 utilizzando il NimbleGen dual color labeling kit (Roche- NimbleGen). 31 µg di campione mescolato al controllo sono stati ibridati nel vetrino per 72 ore a 42°C utilizzando il kit NimbleGen Hybridization (Roche- NimbleGen).

Dopo i lavaggi eseguiti secondo il protocollo (NimbleGen Wash Buffer Kit) il vetrino è stato scannerizzato utilizzando lo scanner Agilent con una risoluzione di 5 µm e l'immagine acquisita è stata successivamente elaborata con il software di analisi NimbleScan v2.6 (Roche- NimbleGen).

6. Sequenziamento dell'esoma mediante tecnologia NGS

Sono stati selezionati in base alla qualità del DNA:

- Per la famiglia BM: due individui affetti e due sani

5 µg di DNA sono stati spezzettati mediante Covaris in frammenti di 150-200 bp.

Il DNA frammentato è stato catturato secondo il protocollo mediante SureSelect Human All Exon Kit (Agilent Technologies, Inc., Santa Clara, CA, USA). Questo kit è in grado di catturare 38Mb del genoma umano (pari al 1.22%) che corrisponde all'intera regione codificante del genoma umano riportata nel database NCBI.

La corsa della libreria di frammenti ottenuti è stata effettuata mediante Illumina Genome Analyzer IIe, utilizzando il protocollo per il sequenziamento 2x51.

- Per il gruppo di MFMs: cinque individui

5 µg di DNA sono stati spezzettati mediante Covaris in frammenti di 250-300 bp.

Il DNA frammentato è stato catturato secondo il protocollo mediante SureSelect Human All Exon Kit (Agilent Technologies, Inc., Santa Clara, CA, USA). Questo kit è in grado di catturare 38Mb del genoma umano (pari al 1.22%) che corrisponde all'intera regione codificante del genoma umano riportata nel database NCBI.

La corsa della libreria di frammenti ottenuti è stata effettuata mediante Illumina Genome Analyzer IIe, utilizzando il protocollo per il sequenziamento 2x101.

7. PIPELINE di analisi dei dati WES

Per l'ammontare dei dati ottenuti sia per i pazienti MFM che la famiglia BM ci si è avvalsi della collaborazione della "Telethon bioinformatics core facility", un centro di analisi di calcoli avanzato per la tecnologia NGS (<http://www.caspur.it/ngs>).

Il primo step dell'analisi è un controllo di qualità delle reads (FastQ format) effettuata mediante FastQC [<http://www.bioinformatics.babraham.ac.uk/projects/fastqc>] tramite cui le reads di bassa qualità vengono scartate. Le reads che al contrario passano il controllo di qualità sono state allineate con il genoma umano usando BWA (Li and Durbin 2009).

I file ottenuti in formato SAM vengono, quindi, convertiti in un formato binario di allineamento (BAM). La codifica delle variazioni che comprendono le variazioni di un singolo polipeptide (SNP) e le piccole inserzioni/duplicazioni (DIP) viene effettuata mediante l'utilizzo del software Genome Analysis Toolkit (GATK); infine le variazioni identificate vengono annotate tramite ANNOVAR e ne vengono definiti la presenza di un polimorfismo noto nella popolazione mediante dbSNPs o 1000genomi, la loro posizione genica e l'effetto provocato (senso, missenso, non-senso, frameshift); mentre la predizione dell'effetto patogenetico della mutazione viene stabilita con SIFT e POLYPHEN.

- Per i pazienti MFM: è stato adottato un tool bioinformatico (MedScan Pathway Studio) allo scopo di identificare una lista di geni candidati, sulla base della quale interrogare i dati di WES.

L'utilizzo di questo software ha permesso di identificare i geni e pathway funzionali correlati ai seguenti geni: DES, MYOT, CRYAB, FLNC, LDB3 e BAG3 e alle seguenti definizioni

caratterizzanti l'eziopatologia delle MFM: dischi Z, miofibrille, aggregazione proteica, anti-apoptotico, depositi granulari, amorfi e ialini.

Sono state ricercate le variazioni presenti in eterozigosi in accordo con un modello dominante di trasmissione della patologia

- Per la famiglia BM: sono stati usati come filtri la ricerca di variazioni
 - nelle regioni associate alle patologie identificate mediante l'analisi di linkage
 - esoniche in eterozigosi secondo il modello dominante
 - presenti nei due affetti e non presenti nei due individui sani

Tutte le variazioni identificate sono state validate con il sequenziamento Sanger.

8. Analisi di sequenza del gene candidato

È stata sequenziata l'intera regione codificante (20 esoni) del gene candidato identificato nella famiglia BM (PDE9A), in 20 casi affetti dalla medesima miopatia o esprimenti la forma allelica più grave (la miopatia di Ullrich).

RISULTATI

1. FAMIGLIA BETHLEM

1.1 Analisi di linkage

Nella famiglia BM sono state identificate 28 regioni in linkage (Fig. 10) per un totale di 50Mb che includono circa 5000 geni. Quest'analisi ha permesso l'esclusione, oltre al cromosoma X in accordo con il modello autosomico del pedigree, dei cromosomi 6, 9, 13, 14, e 15.

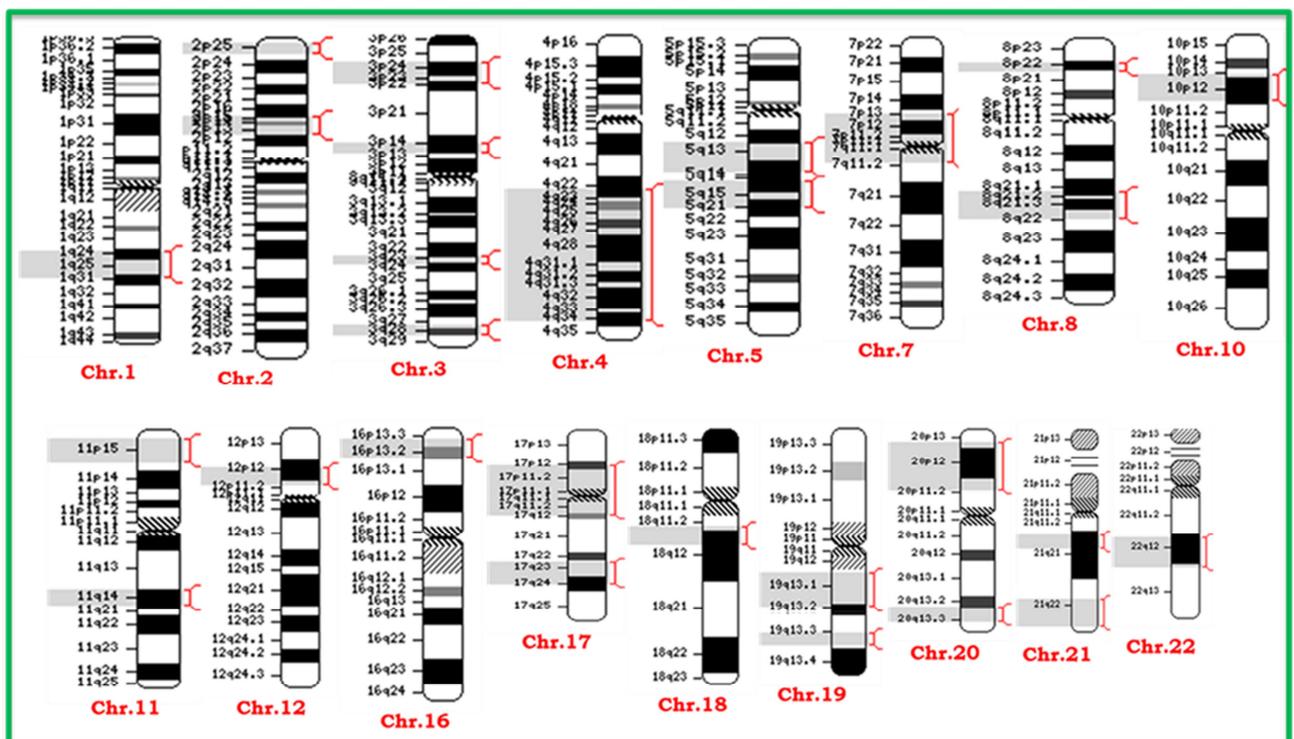


Fig. 10: nella famiglia BM l'analisi di linkage ha portato all'identificazione di 28 regioni in associazione con la patologia. Quest'analisi ha permesso l'esclusione dei cromosomi 6, 9, 13, 14, 15.

1.2 Array-CGH

L'analisi nel probando non ha identificato nessun riarrangiamento (delezione o duplicazione) in tutti i geni del collagene (COL6A1, COL6A2, COL6A3) e negli altri geni presenti nell'array correlati alle NMD.

1.3 Sequenziamento dell'esoma mediante tecnologia NGS

Sono state ottenute approssimativamente 18.181.000 milioni di paired-end reads, lunghe 51 bp. Dopo analisi di qualità e annotazione delle varianti, sono state identificate una media di 41360 SNP e 1160 DIP, nella tabella 1 è specificato l'ammontare delle variazioni individuate per paziente. I filtri applicati all'ammontare delle variazioni ottenute sono stati: innanzitutto si sono ricercate le varianti comuni ai due affetti ma non presenti nei due sani, successivamente sono state escluse variazioni presenti nel dbSNP e variazioni in altri pazienti non relati presenti nel nostro database interno, le regioni in linkage individuate hanno rappresentato un altro importante filtro, seguito infine dalla selezione dei SNP esonici o delle DIP (fig. 11).

Al termine di questo processo abbiamo ottenuto 15 variazioni esoniche, all'interno delle quali abbiamo escluso le variazioni sinonime ricavando quindi solo quattro variazioni missenso a carico dei geni SLC9A11, DYSF, TNS3 e PDE9A e una delezione in frame a carico del gene CEP68 (tabella 2).

La delezione in frame nel gene CEP68, oltre ad essere conservata nel genoma di riferimento di Rhesus, è presente in 1000 genomi con una frequenza del 10% nella popolazione caucasica; quindi questa variazione è difficilmente considerabile come patogenetica.

Una più ampia analisi di segregazione nell'intera famiglia BM delle rimanenti quattro variazioni missenso, ha permesso l'esclusione di due di esse (TNS3, DYSF).

La variazione SLC9A11 pur segregando nella famiglia non è espressa a livello muscolare indicando una difficile correlazione con la miopatia in esame.

Riguardo alla restante variazione nel gene PDE9A (NM_001001580 c.C55T p.P19S), è stata valutata per una possibile patogenicità mediante l'utilizzo di diversi tool bioinformatici: il risultato della predizione risulta discordante tra i diversi tool utilizzati, più precisamente Mutation Taster (<http://www.mutationtaster.org/>) associa una patogenicità alla variazione identificata, tale patogenicità è correlata, secondo predizione del tool ALAMUT, ad una alterazione di legame delle

proteine dello splicing, al contrario POLYPHEN2 (<http://genetics.bwh.harvard.edu/pph2/>) classifica la variazione come benigna (fig.12).

La variazione non è riportata nel dbSNP e in 1000 genomi, ma è presente nel Seattle Database (<http://evs.gs.washington.edu/EVS/>) seppure con una frequenza inferiore allo 0,5%.

1.4 Analisi di sequenza del gene candidato

La ricerca di mutazioni nella regione codificante del gene PDE9A in 20 pazienti affetti da miopia di Bethlem o dalla forma allelica più grave (miopia di Ullrich), ha evidenziato la presenza di polimorfismi noti ma non di mutazioni causative.

name	Variant calling	Filtered* SNPs	Filtered* DIPs (INDEL)
famiglia BM femmina affetta 1	Total: 64313 Filtered*: 40441	SNP: 39373 not in dbSNP: 2222 homozygous: 14558 heterozygous: 24815	DIP: 1068 not in dbSNP: 383 homozygous: 500 heterozygous: 568
famiglia BM maschio affetto 2	Total: 64187 Filtered*: 40400	SNP: 39507 not in dbSNP: 2551 homozygous: 14572 heterozygous: 24935	DIP: 893 not in dbSNP: 316 homozygous: 415 heterozygous: 478
famiglia BM maschio sano 3	Total: 76365 Filtered*: 48154	SNP: 46605 not in dbSNP: 2912 homozygous: 16935 heterozygous: 29670	DIP: 1549 not in dbSNP: 577 homozygous: 688 heterozygous: 861
famiglia BM femmina sana 4	Total: 66655 Filtered*: 41090	SNP: 39955 not in dbSNP: 2344 homozygous: 14249 heterozygous: 25706	DIP: 1135 not in dbSNP: 415 homozygous: 532 heterozygous: 603

Tabella 1: variazioni individuate nei 4 individui della famiglia BM corsi su piattaforma Illumina, con specificate SNP e DIP ottenute prima e dopo il processo di qualità, con l'applicazione di filtri come la presenza nel dbSNP e la ricerca di variazioni in omozigosi e in eterozigosi

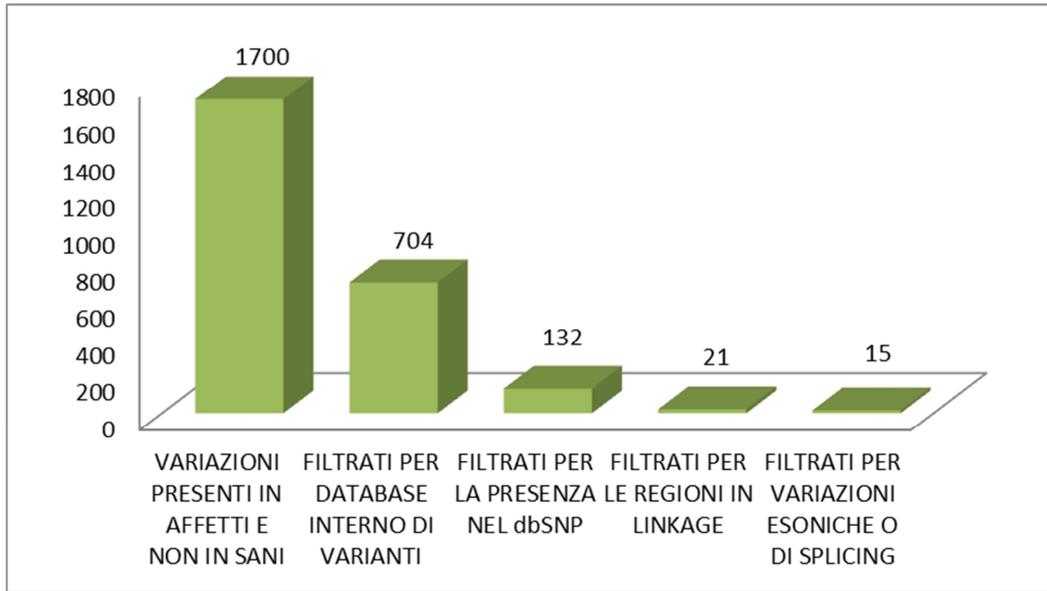


Fig. 11: rappresentazione dei filtri adottati in successione allo scopo di ridurre le variazioni candidate nella famiglia BM

Allele variation	Position	Var. type	Cove rage	state	dbSNP	gene	Gene location	Variant type	Functional annotation
G>A	Chr1:171759805	SNP	20	het	-	SLC9A11	exonic	missense	NM_178527 c.C2566T p.P856S
AAC>-	Chr2:65152151-65152153	DIP	69	het	- 1000g_ceu:0.10	CEP68	exonic	nonframeshift deletion	NM_015147 c.417_419del p.139_140del
G>T	Chr2:71745049	SNP	22	het	-	DYSF	exonic	missense	NM_001130976 c.G4988T p.R1663L
G>T	Chr7:47375320	SNP	15	het	-	TNS3	exonic	missense	NM_022748 c.C1448A p.P483H
C>T	Chr21:43044347	SNP	112	het	-	PDE9A	exonic	missense	NM_001001580 c.C55T p.P19S

Tabella 2: dettagli delle variazioni identificate dopo l'applicazione di diversi step di filtraggio.

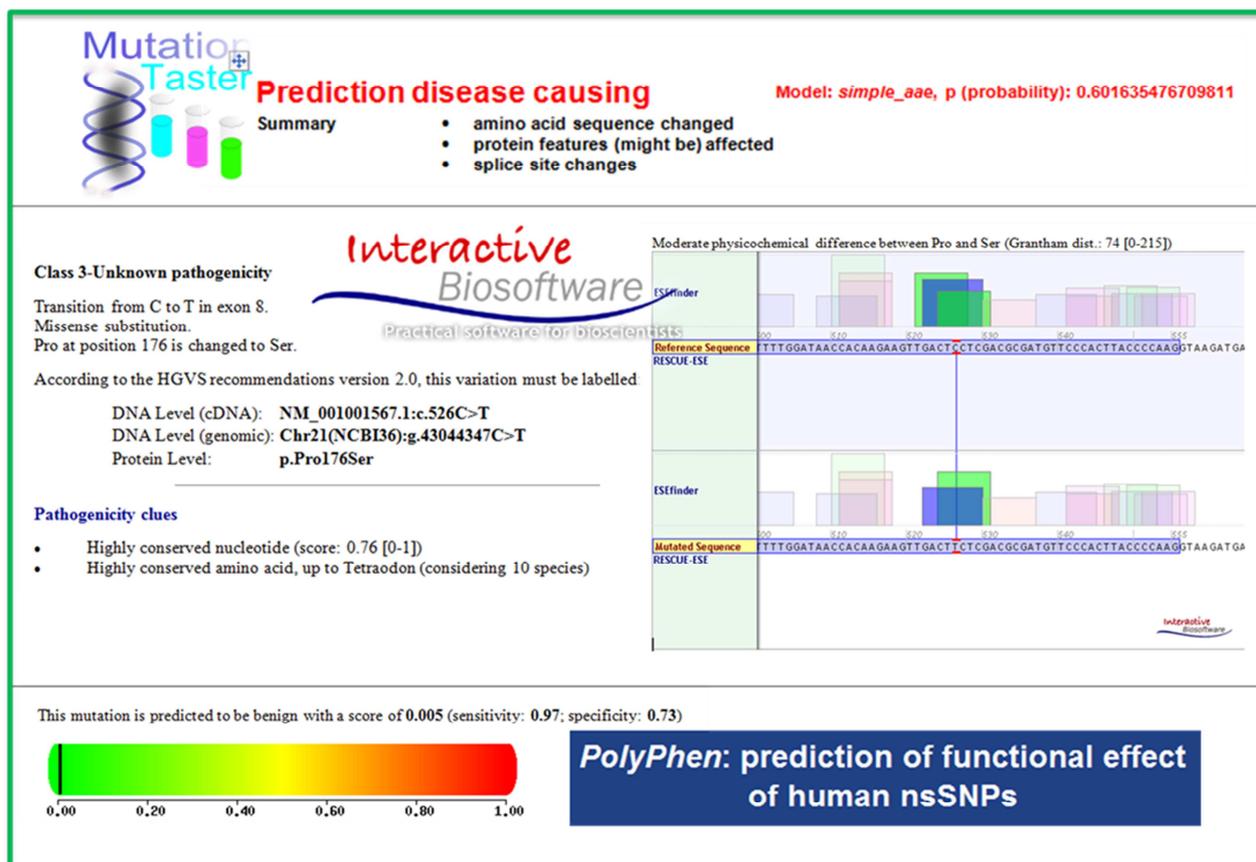


Fig.12: predizione della patogenicità della variazione sul gene PDE9A (NM_001001580 c.C55T p.P19S) mediante l'utilizzo di tre diversi tool bioinformatici.

2. MIOPATIA MIOFIBRILLARE

2.1 Analisi di sequenza dei geni noti

L'analisi dell'intera regione codificante dei geni noti per essere coinvolti in MFM: DESM, MYOT, CRYAB, LDB3 e BAG3 ha identificato la presenza di polimorfismi noti ma non di variazioni patogenetiche.

2.2 Array-CGH

L'analisi nei 19 pazienti MFM non ha identificato nessun riarrangiamento (delezione o duplicazione) in tutti i sei geni associati alla miopatia e negli altri geni correlati alle NMD ad eccezione di un evento duplicativo nel gene LAMA2 in due pazienti (paziente1 e paziente2) con uno score di 0,5. La regione interessata (Hg18 chr 2: 129.849.433-129.849.793; Hg19 chr2:130132963-130133323) coinvolge l'esone 56 e parte dell'introne 57 (Fig. 13).

L'analisi di segregazione nel paziente 1 con familiarità per la patologia ha mostrato la presenza di una duplicazione in eterozigosi anche nel padre affetto dalla patologia seppur con una lieve severità.

Allo scopo di identificare la presenza di mutazioni puntiformi nel secondo allele, è stato sequenziato l'intera regione codificante del gene LAMA2; tuttavia l'analisi non ha portato al riscontro di alcuna variazione causativa.

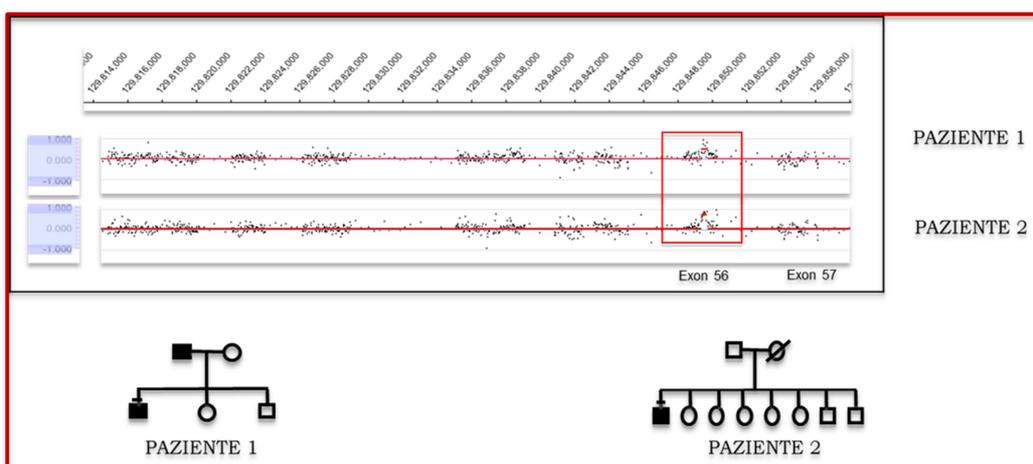


Fig. 13: rappresentazione dell'evento di duplicazione identificato nel gene LAMA2 mediante l'analisi di CGH-array in due pazienti. Il paziente 1 mostra una familiarità per la patologia, mentre il paziente 2 è un caso sporadico de novo.

2.3 Sequenziamento dell'esoma mediante tecnologia NGS

L'analisi WES è stata condotta su 5 pazienti MFM risultati negativi sia all'analisi di sequenza nei geni noti, (ad eccezione del gene FLNC che non è stato indagato), sia all' array-CGH.

Sono state ottenute approssimativamente 16.850.000 milioni di paired-end reads, lunghe 101 bp.

Dopo analisi di qualità e annotazione delle varianti, sono state identificate una media di 66640 SNP e 1840 DIP, nella tabella 3 è specificato l'ammontare delle variazioni individuate per paziente.

Il numero elevato di variazioni ci ha portato ad applicare ai risultati ottenuti una serie ulteriore di filtri, quali la selezione di variazioni esoniche e l'esclusione di variazioni presenti nei database dbSNP e 1000 genomi; un'ulteriore processo di esclusione si è basato sulle varianti presenti in un nostro database interno composto dalle variazioni identificate in tutti i precedenti esperimenti di WES, riconducibili ad un errore sistematico dello strumento. Inoltre i dati ottenuti sono stati interrogati sulla base di 880 geni candidati individuati mediante Medscan Pathway studio che comprendono, oltre ai geni MFM noti, geni coinvolti nel pathway dei dischi Z e della contrazione e rigenerazione muscolare.

Questo procedimento ha permesso l'identificazione in due pazienti di una mutazione nel gene noto FLNC, e nei restanti tre pazienti di due geni candidati (TTN e CAMK2D).

Di seguito l'analisi dettagliata nei cinque pazienti.

name	Variant calling	Filtered* SNPs	Filtered* DIPs (INDEL)
paziente MFM 3	Total: 124844 Filtered*: 76785	SNP: 74506 not in dbSNP: 4928 homozygous: 28252 heterozygous: 46254	DIP: 2279 not in dbSNP: 735 homozygous: 1048 heterozygous: 1231
paziente MFM 4	Total: 124099 Filtered*: 75279	SNP: 73352 not in dbSNP: 7698 homozygous: 27677 heterozygous: 45675	DIP: 1927 not in dbSNP: 629 homozygous: 905 heterozygous: 1022
paziente MFM 5	Total: 104849 Filtered*: 64667	SNP: 62846 not in dbSNP: 4542 homozygous: 23440 heterozygous: 39406	DIP: 1821 not in dbSNP: 577 homozygous: 769 heterozygous: 1052
paziente MFM 6	Total: 97637 Filtered*: 44022	SNP: 43144 not in dbSNP: 13270 homozygous: 8988 heterozygous: 34156	DIP: 878 not in dbSNP: 282 homozygous: 344 heterozygous: 534
paziente MFM 7	Total: 132843 Filtered*: 81646	SNP: 79352 not in dbSNP: 6157 homozygous: 25593 heterozygous: 53759	DIP: 2294 not in dbSNP: 780 homozygous: 901 heterozygous: 1393

Tabella 3: variazioni individuate nei 5 pazienti corsi su piattaforma Illumina, con specificate SNP e DIP ottenute prima e dopo il processo di qualità, con l'applicazione di filtri come la presenza nel dbSNP e la ricerca di variazioni in omozigosi e in eterozigosi.

- Paziente 3. (femmina affetta deceduta nel 2011 per arresto respiratorio). È stata identificata una mutazione nonsense in eterozigosi a carico dell'esone 48 del gene FLNC (NM_001127487 c.G8031A p.W2677X), mai precedentemente riportata. La mutazione è stata validata mediante Sanger ed è stata inoltre riscontrata in un fratello affetto con un fenotipo più lieve della patologia. Il pedigree e la variazione identificata sono mostrate nella figura 14.

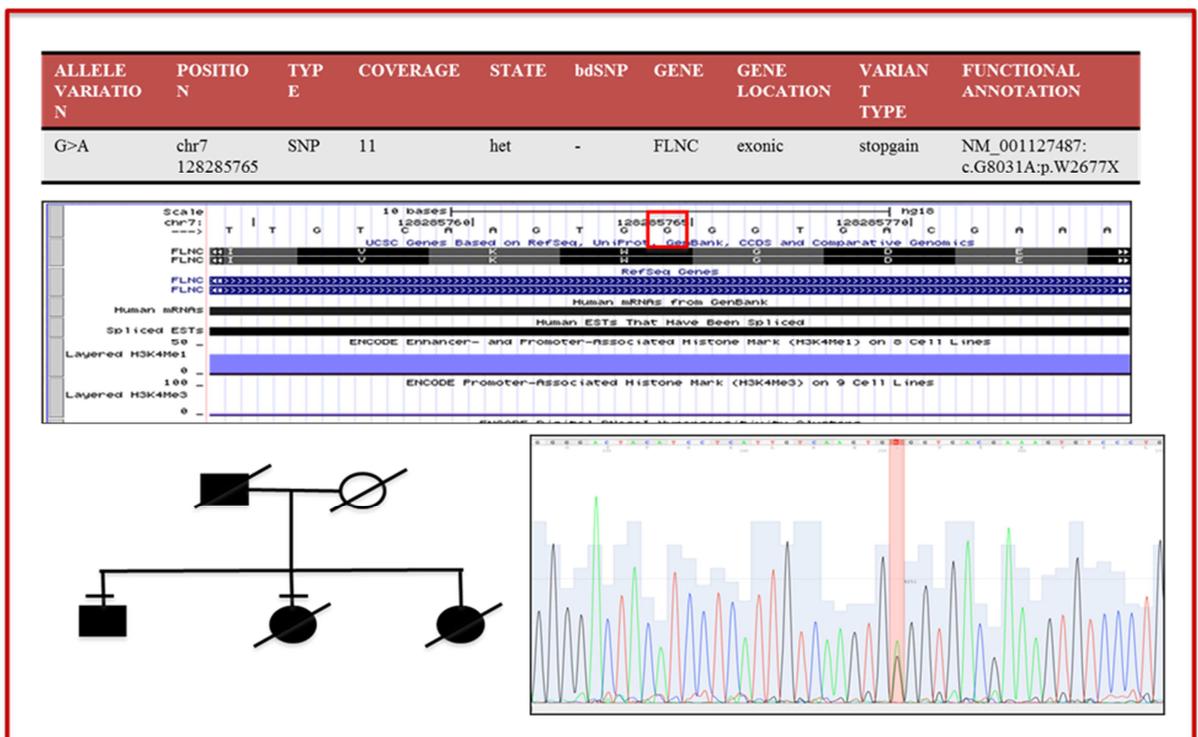


Fig. 14: in figura è rappresentato la specifica della mutazione nella FLNC del paziente MFM 3, il pedigree della famiglia da cui si deduce un modello di trasmissione autosomico dominante, e il cromatogramma della validazione mediante Sanger.

- Paziente 4. In questo paziente è stata identificata una nuova variazione presente nel gene FLNC in eterozigosi (NM_001127487:c.G4022A:p.R1341Q) che localizza all'interno dell'esone 24. Mediante il tool bioinformatico mutation taster (<http://www.mutationtaster.org/>) si è osservato il possibile effetto patogenetico della patologia. L'analisi di segregazione della variazione non è possibile in quanto il paziente è un caso sporadico (fig.15).

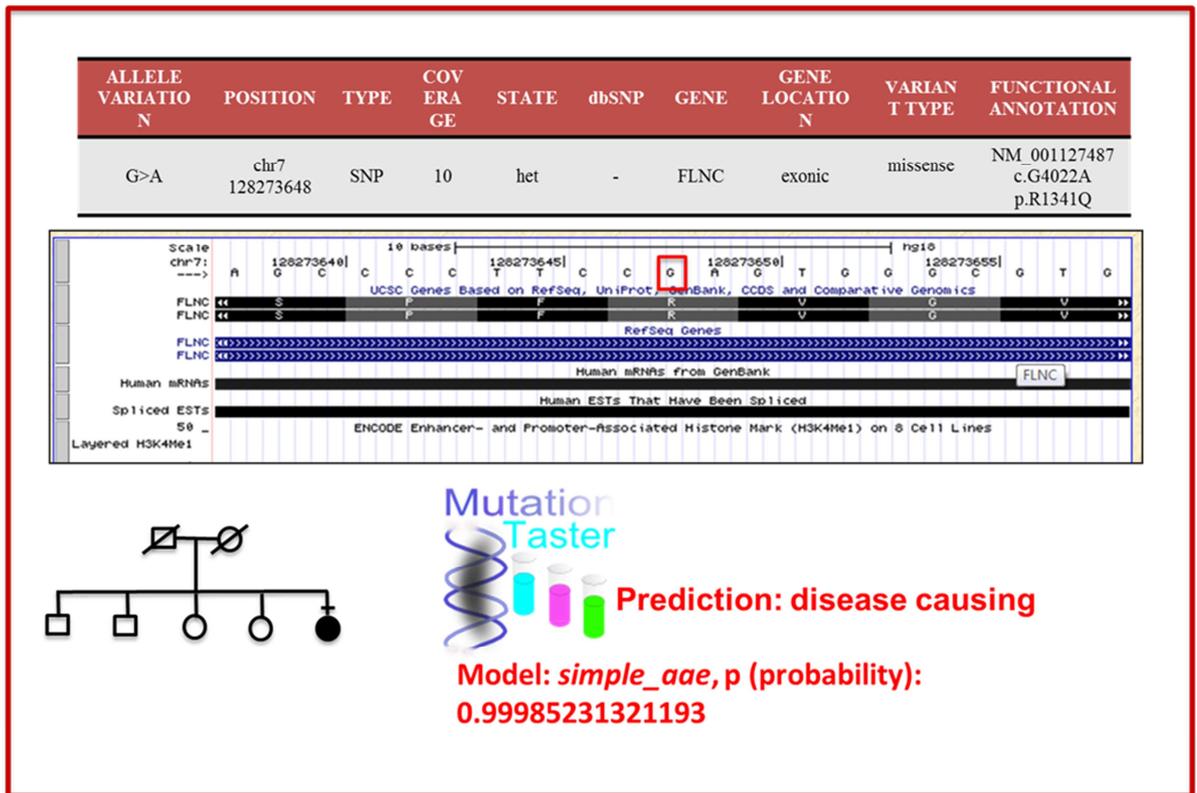


Fig. 15: in figura è rappresentata la specifica della mutazione nella FLNC del paziente MFM 4, il pedigree della famiglia e il risultato della predizione della patogenicità mediante il tool bioinformatico mutation taster (<http://www.mutationtaster.org/>).

- Paziente 5. È stata identificata una variazione missenso in eterozigosi nel gene TTN (NM_133379:c.C2585T:p.P862L) la cui predizione di patogenicità (<http://www.mutationtaster.org/>) indica una possibile correlazione al fenotipo MFM causata dall'alterazione amminoacidica e legata anche ad una elevata conservazione interspecie del residuo naturale verificata mediante polyphen2 (<http://genetics.bwh.harvard.edu/pph2/>) (fig.16).

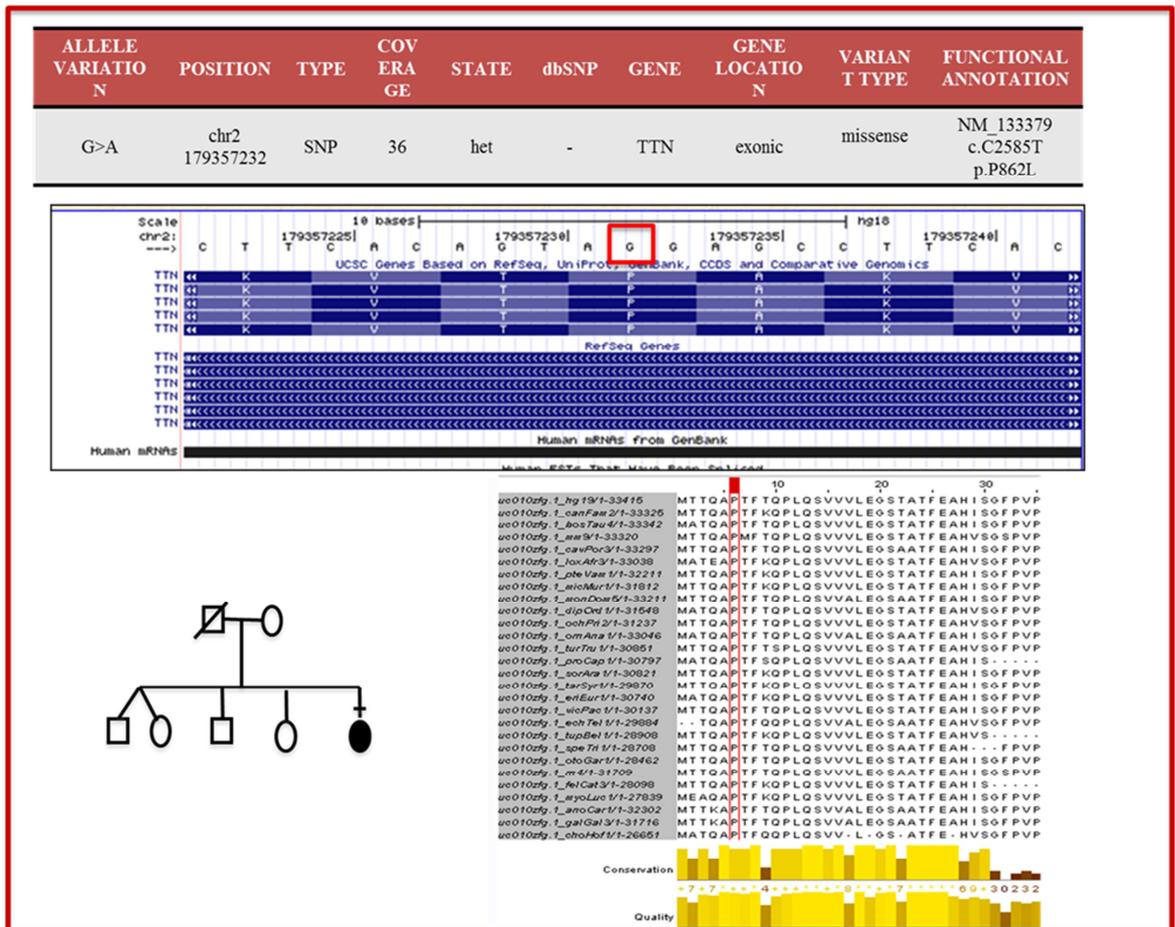


Fig. 16: in figura è rappresentato la specifica della mutazione nella TTN del paziente MFM 5, il pedigree della famiglia e la conservazione interspecie mediante polyphen2 (<http://genetics.bwh.harvard.edu/pph2/>)

- Paziente 6. Una seconda variazione in eterozigosi nel gene TTN è stata identificata nel paziente 6 (NM_133379:c.A13936G:p.K4646E). L'analisi mediante mutation taster (<http://www.mutationtaster.org/>) ha predetto un debole effetto patogenetico legato all'alterazione aminoacidica e ad una possibile alterazione del meccanismo di splicing. La disponibilità di due fratelli e una sorella senza segni clinici della patologia, ci ha permesso il sequenziamento dei medesimi. Il risultato mostra la presenza della variazione in due di questi (fig. 17).

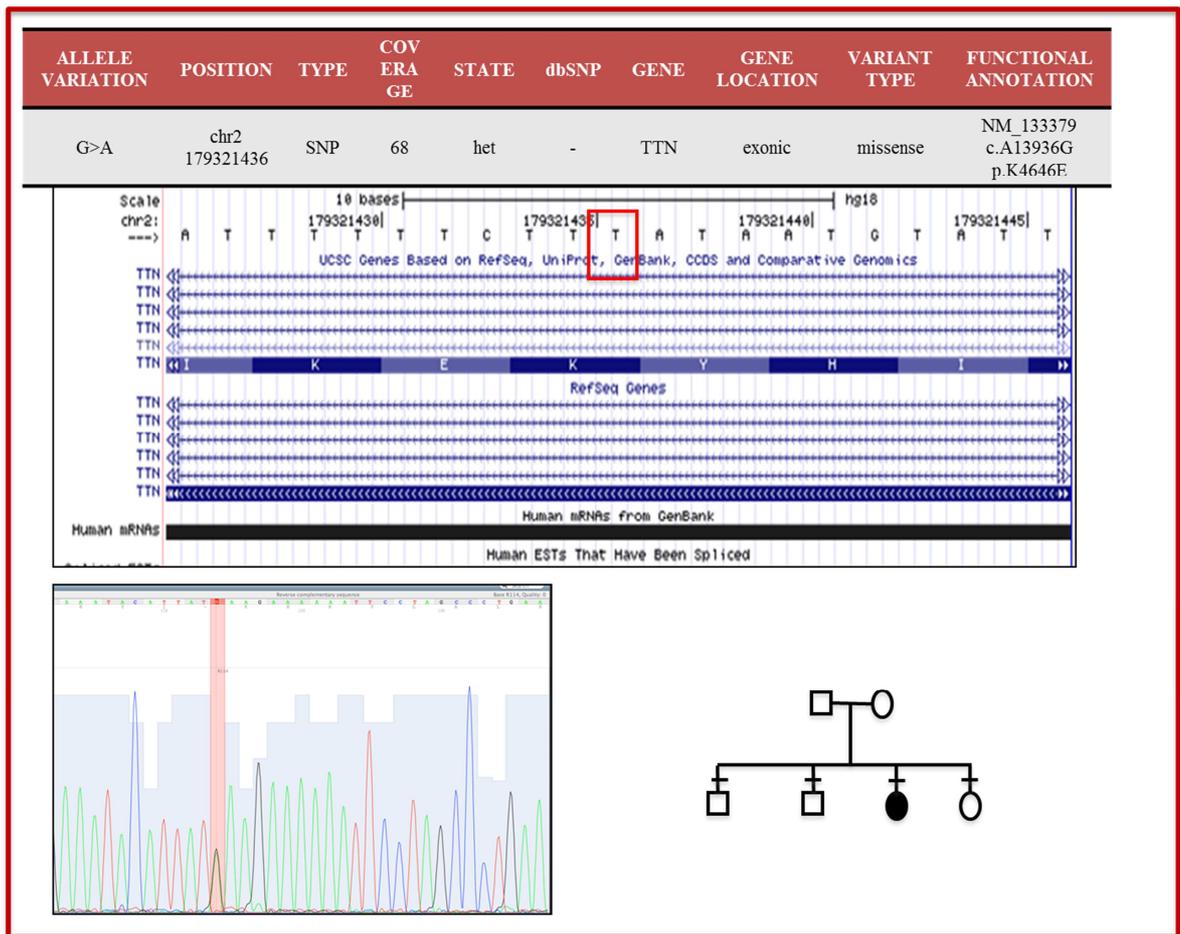


Fig. 17: in figura è rappresentata la specifica della mutazione nella TTN del paziente MFM 6, il pedigree della famiglia, il cromatogramma indicativo della presenza della variazione in due dei tre fratelli sani.

- Paziente 7. In questo paziente applicando il filtro per i 880 geni candidati ed escludendo le variazioni presenti in altri pazienti non MFM rilevate nel nostro database interno di variazioni WES, abbiamo ottenuto una variazione nonsense nel gene VIM che non è stata validata mediante Sanger e tre variazioni missenso a carico dei geni CD180, AKAP4 e CAMK2D, solo la variazione in quest'ultimo gene (NM_001221 c.C229T p.H77Y) è predetta come patogenetica (fig. 18).

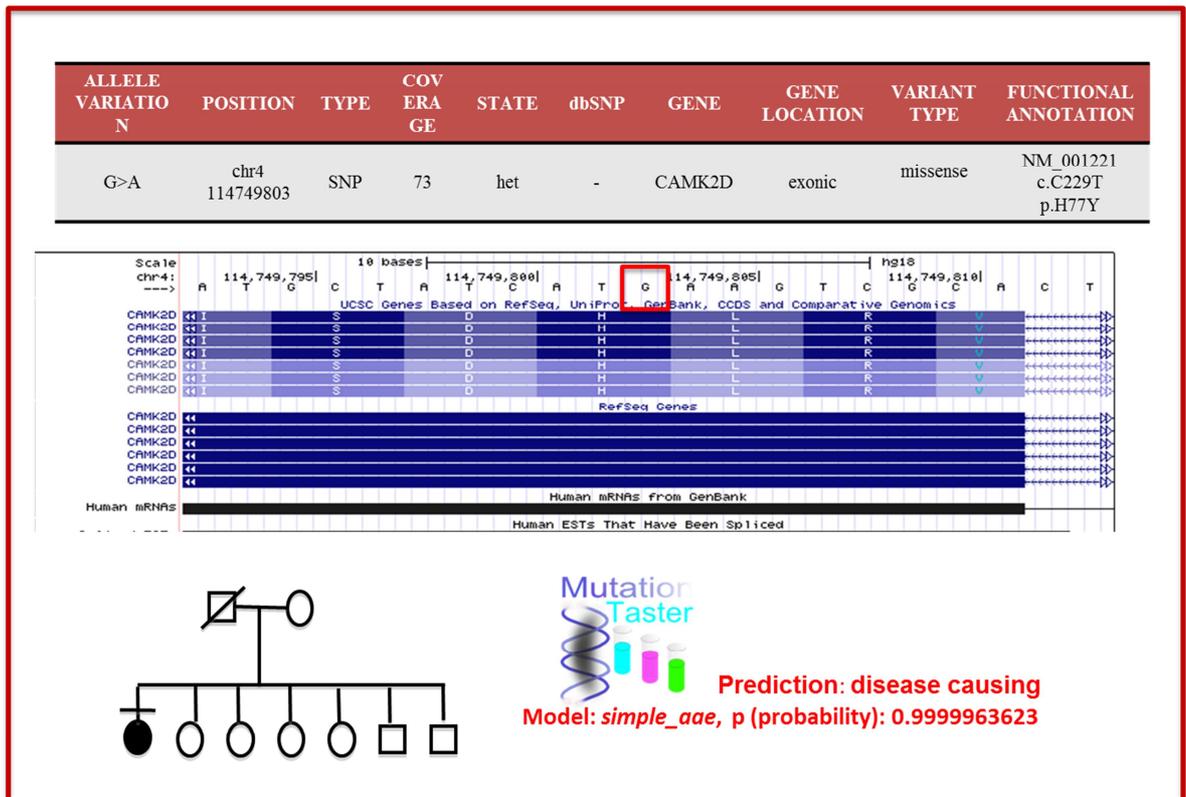


Fig. 18: in figura è rappresentato la specifica della mutazione nel gene CAMK2D del paziente MFM 7, il pedigree della famiglia e la predizione di patogenicità mediante mutation taster (<http://genetics.bwh.harvard.edu/pph2/>).

DISCUSSIONE

Il sequenziamento dell' esoma (WES) mediante la piattaforma di nuova generazione Illumina GaIIe è stato applicato allo scopo di identificare un nuovo gene malattia in una famiglia affetta da miopatia di Bethlem (BM) e allo scopo diagnostico in un gruppo di soggetti affetti da miopatia miofibrillare (MFM).

I geni noti essere responsabili della BM sono i COL6A1, COL6A2, COL6A3 che codificano per le tre catene che associandosi equimolarmente formano il collagene VI, una proteina localizzata nella matrice extracellulare che tramite il suo esteso network di legame ha il compito di mantenere l'integrità cellulare. Un' ampia casistica di pazienti (circa il 25%) rimangono orfani di una diagnosi molecolare, ciò suggerisce una possibile eterogeneità genetica (Lampe AK & Bushby KM 2005).

Allo scopo di identificare un nuovo gene malattia abbiamo analizzato quattro individui (due affetti e due sani) della famiglia BM con ereditarietà autosomica dominante; il risultato ottenuto dal sequenziamento massivo in parallelo ha indicato un solo possibile gene candidato (PDE9A) che viene espresso, seppur non ad elevati livelli, nel tessuto muscolare.

Il gene PDE9A fa parte di una classe di proteine, le fosfodiesterasi (PDE), composta da 11 membri che hanno diversa affinità per l'adenosina monofosfato ciclasi (cAMP), la guanosina monofosfato ciclasi (cGMP) oppure entrambe. Il compito di questi enzimi è di mediare regolare mediante idrolisi la concentrazione intracellulare dei nucleotidi ciclici implicati nella trasduzione del segnale di una varietà di pathway cellulari tra i quali regolazione di canali ionici, recettori, proteine del citoscheletro o fattori di trascrizione (Bender and Beavo 2006).

In particolare l'attività di idrolisi dell'enzima PDE9A è specifica per il cGMP che è implicata nel pathway dell'ossido nitrico (NO) (Lakics et al. 2010).

Alterazioni dell'attività delle fosfodiesterasi sono già note essere implicate nell'infertilità, nell' asma, nel diabete e nella depressione, ma non in patologie neuromuscolari (Bingham et al. 2006).

In un recente lavoro si può tuttavia trovare un punto di contatto tra le fosfodiesterasi e le distrofie congenite: Percival et al. allo scopo di amplificare il segnale dell' NO, solitamente attenuato nelle patologie neuromuscolari, ha trattato topi mdx (ovvero il modello murino della distrofia muscolare di duchenne) con un inibitore della fosfodiesterasi 5, anch'essa con attività specifica per il cGMP.

Il risultato ottenuto è stato una riduzione delle disfunzioni muscolari soprattutto a livello del diaframma e l'induzione di una corretta organizzazione della matrice extracellulare (Percival et al. 2012). Questi risultati potrebbero suggerire come mutazioni nelle fosfodiesterasi possano portare

all'insorgenza di miopatie mediante un meccanismo di "gain of function" ovvero con un'aumentata attività enzimatica che si ripercuote in un'elevata espressione intracellulare di cGMP.

Sfortunatamente il sequenziamento del gene PDE9A in altri 20 pazienti con quadro clinico comparabile ma esenti da mutazioni nei geni del collagene VI, non ha permesso l'identificazione di nessuna mutazione a carico del gene candidato.

Tuttavia il risultato dello screening potrebbe indicare come la mutazione identificata (NM_001001580 c.C55T p.P19S) sia solamente una variazione privata della famiglia, o potrebbe segnalare la presenza di un alto livello di eterogeneità genetica; infatti altre patologie neuromuscolari come ad esempio la Charcot-Marie-Tooth, dove sono noti 35 geni causativi, hanno un numero elevato numero di geni correlati alla malattia (Montenegro et al 2011).

Secondo un'ipotesi alternativa, rafforzata anche dalla presenza della variazione (NM_001001580 c.C55T p.P19S) nel Seattle database, la variante identificata potrebbe essere solamente un polimorfismo con bassa frequenza nella popolazione o un biomarcatore di severità fenotipica.

In tal caso l'identificazione del reale gene malattia non sarebbe riuscita per motivi legati alla complessità dell'analisi dei dati o per una bassa efficienza nella fase di cattura dell'esoma; infatti nei primi esperimenti rimaneva una quota pari all'8% che resisteva all'arricchimento, sebbene i sistemi di arricchimenti siano migliorati la catturata non arriva al 100% (Singleton 2011).

Inoltre uno studio approfondito dove si è comparata l'efficienza di tre diverse piattaforme ha riportato come siano state identificate solamente il 60-70% delle varianti riscontrate mediante sequenziamento Sanger (Hoppman-Chaney 2010).

Le MFM sono un gruppo di miopatie con una nota eterogeneità genetica ed una bassa "detection rate" pari al 50%. Lo studio della nostra coorte di pazienti MFM risultati negativi all'analisi di sequenza in cinque dei sei geni noti (DES, MYOT, CRYAB, LDB3 e BAG3) mediante tecniche ad alta efficienza come l'array CGH e l'analisi WES. Il CGH è in grado di identificare variazioni del numero di copie (delezioni o duplicazioni) che rappresentano il 5-6% delle mutazioni genetiche causative non determinabili mediante la tecnologia standard (Armour et al. 2002). Il CGH-array utilizzato in grado di coprire 50 geni noti coinvolti in NMD ci ha permesso di identificare un evento duplicativo in eterozigosi che coinvolge l'esone 56 del gene LAMA2 in due pazienti della nostra coorte.

Mutazioni in omozigosi o eterozigosi a carico del gene LAMA2 sono associate alla distrofia muscolare congenita di tipo 1A (MDC1A), un gruppo di patologie ad esordio neonatale o nei primi mesi di vita che si manifestano con ipotonia, debolezza muscolare degli arti e del tronco e contratture muscolari.

Mutazioni puntiformi in eterozigosi del gene LAMA2 sono anche state identificate in pazienti con distrofia muscolare prossimale (Rajakulendran et al 2011), mentre variazioni del numero di copie sempre in eterozigosi sono state associate a distrofia muscolare dei cingoli o ad una generica miopatia (Piluso et al. 2011). Dai dati di letteratura emerge un'elevata eterogeneità allelica associata a variazioni del gene LAMA2.

Nella determinazione delle variazioni causative nei cinque pazienti MFM corsi su piattaforma Illumina è risultato un utile strumento la lista dei geni candidati, individuati mediante il software MedScan Pathway Studio mediante il quale sono stati selezionati geni coinvolti nel meccanismo patofisiologico della MFM.

Nei pazienti 3 e 4 sono state identificate due nuove variazioni, rispettivamente nonsense e missenso, all'interno del gene FLNC noto coinvolto in MFM (Vorged et al. 2005) ma non analizzato in sequenza a causa della sua estensione genica. In particolare, la variazione nonsense identificata nel paziente 3 (W2677X) mappa nell'ultimo esone del gene, dove è stata descritta la variazione sempre nonsense più frequente (W2710X) della FLNC (Vorged et al. 2005). La patogenicità della variazione identificata è supportata dall'analisi di segregazione nel fratello affetto dove si riscontra la medesima variazione in eterozigosi.

L'analisi WES oltre ad uno scopo puramente diagnostico di indagine nei geni noti, ha permesso l'identificazione di un nuovo gene candidato: la titina (TTN). La titina è la proteina descritta più grande con 363 esoni differenzialmente splicati per originare cinque diverse isoforme, una singola molecola della proteina si estende da un disco Z alla linea M adiacente, la sua funzione consiste nello stabilizzare la posizione dei filamenti contrattili e consentire, grazie alla sua elasticità, al muscolo striato di tornare alla lunghezza di partenza.

Mutazioni a carico di questo gene sono associate ad un ampio spettro fenotipico e non è sempre possibile un'esatta correlazione genotipo-fenotipo: mutazioni puntiformi in eterozigosi che colpiscono gli esoni all'estremità carbossi-terminale sono associati a distrofia muscolare della tibia (Hackman et al. 2002), nella stessa regione mutazioni in omozigosi sono causa di una severa forma di miopatia scheletrica ad esordio precoce, LGMD2J (Udd et al. 2005). Mutazioni a carico del dominio chinasi nella banda M e nella banda A causano una miopatia ereditaria con precoce insufficienza respiratoria (Ohlsson et al. 2012), mentre eventi delettivi in omozigosi provocano cardiomiopatia ad esordio precoce (Carmignac et al. 2007). Le variazioni identificate in due pazienti (paziente 5: P862L; paziente 6: K4646E) cadono rispettivamente sui domini immunoglobulin-like 3 e 26. La variazione P862L è stata predetta mediante tool bioinformatici come patogenetica in quanto colpisce una zona altamente conservata

immediatamente vicina al sito di legame con il disco Z, tuttavia per una certa correlazione è necessaria un'analisi in altri pazienti affetti da MFM; ancora meno sicura è la correlazione tra la seconda variazione identificata K4646E e la patologia in quanto la medesima variazione è stata osservata in due fratelli sani suggerendo un significato benigno della variante oppure la necessità di una seconda mutazione all'interno del gene in base ad un modello di eterozigosi composta.

Nell'ultimo paziente indagato è stata riportata la variazione missenso H77Y all'interno del gene CAMK2D, questo gene appartiene alla famiglia delle protein chinasi calcio-calmodulina dipendenti di tipo II ed è coinvolto nella patogenesi della cardiomiopatia dilatativa e dell'insufficienza cardiaca (Toko et al. 2010). Ad oggi non è evidente nessun coinvolgimento del gene CAMK2D nei disordini neuromuscolari, solo la definizione genetica di altri casi di MFM o studi funzionali potranno attribuire una certa patogenicità.

In breve l'analisi WES si è dimostrata efficiente a scopi diagnostici ovvero nella ricerca di variazioni in geni noti (identificazione di mutazioni nel gene FLNC per i pazienti MFM 3 e 4); invece l'identificazione di nuovi geni-malattia si è constatata essere limitante nella ricerca di mutazioni in eterozigosi in base al modello dominante: infatti, ad oggi, sono stati identificati svariati geni causativi in famiglie con ereditarietà recessive spesso con consanguineità, ma pochi casi di famiglie dominanti (Zhi and Chen 2012). Un altro aspetto limitante è l'analisi di casi sporadici non relati (pazienti MFM), una possibile soluzione è da ricercarsi nell'ampliamento della casistica; infatti Zhi e Chen spiegano che anche nelle patologie che colpiscono il 5% della popolazione, il gene causativo può essere facilmente identificato solamente mediante il sequenziamento di 200 casi non relati (Zhi and Chen 2012).

CONCLUSIONI

L'analisi della famiglia BM tramite tecnologia NGS ha permesso, dopo importanti strumenti di filtro, l'identificazione di un solo gene candidato (PDE9A), purtroppo la valutazione dell'impatto della mutazione mediante diversi strumenti bioinformatici sottolinea un significato patogenetico discordante; inoltre la dubbia patogenicità è stata evidenziata dal riscontro della medesima variazione nel Seattle database seppur con bassa frequenza nella popolazione e dalla mancata identificazione di mutazioni in altri 20 pazienti con paragonabili segni clinici, suggerendo la presenza di una mutazione familiare o il fallimento dell'analisi informatica di WES.

Sarebbe quindi opportuna un'implementazione delle conoscenze riguardo alla frequenze alleliche dei polimorfismi in ampie popolazioni di riferimento, infatti la nostra variazione non riportata in 1000 genomi ha avuto un riscontro nel Seattle database che raccoglie informazioni derivanti da 10.000 individui; la costituzione di database per analisi WES con una moltitudine di dati di frequenza potrebbe riflettersi in una più accurata ricerca di nuovi geni con un numero inferiore di falsi positivi.

Le fasi maggiormente limitanti nella tecnologia WES si riscontrano: nello "step" iniziale di preparazione della libreria con l'arricchimento dell'esoma, infatti nonostante continui sviluppi rimane una percentuale di esoma, nei primi arricchimenti l'8%, che fallisce la cattura (Singleton 2011), e la fase finale di analisi dei dati, che rappresenta il passaggio più critico. Infatti si riscontrano falsi positivi, solitamente presenti in tutti i campioni analizzati in quanto associati ad errori sistematici, e falsi negativi conseguenti appunto ad una bassa cattura delle regioni di interesse, a un basso "coverage" delle letture o reads che ne implica l'eliminazione dall'analisi o alla presenza di regioni ripetute che vengono scartate determinando "gap" nell'allineamento (Lin et al. 2012).

Per questi motivi è in corso una ri-analisi dei dati della famiglia BM ai fini di aumentare il numero di geni candidati mediante aggiustamento di parametri informatici, sebbene una seconda analisi potrebbe non portare all'identificazione del gene-malattia in quanto tramite la tecnologia WES non possono essere individuate mutazioni introniche profonde ed estese variazioni del numero di copie (delezioni, duplicazioni); inoltre solamente il 60-70% delle varianti riscontrate mediante sequenziamento Sanger sono verificate mediante tecnologia NGS (Hoppman-Chaney 2010).

Un successo maggiore di questa tecnologia è stato evidenziato nell'applicazione diagnostica ovvero nella ricerca di mutazioni in geni noti, infatti nei cinque pazienti MFM corsi su piattaforma Illumina

sono state identificate due variazioni nel gene noto coinvolto nella miopatia, più precisamente nel paziente 3 si è individuata una mutazione non senso mentre nel paziente 4 una variazione missenso, entrambe non precedentemente descritte.

La ricerca di mutazioni nei rimanenti pazienti MFM, applicando una lista di geni candidati associati ai meccanismi di contrazione e rigenerazione muscolare, ha portato ad individuare due variazioni nel gene TTN associato ad un'elevata eterogeneità allelica e nel gene CAMK2D di cui non è attualmente noto un coinvolgimento in disordini neuromuscolari.

Per una determinazione certa di patogenicità si rendono necessari ulteriori studi volti ad un aumento del numero di pazienti indagati e alla valutazione dell' impatto delle mutazioni mediante studi funzionali.

BIBLIOGRAFIA

1. Armour JA, Barton DE, Cockburn DJ, Taylor GR. The detection of large deletions or duplications in genomic DNA. *Hum Mutat.* 2002 Nov;20(5):325-37.
2. Bender AT, Beavo JA. Cyclic nucleotide phosphodiesterases: molecular regulation to clinical use. *Pharmacol Rev.* 2006 Sep;58(3):488-520.
3. Bethlem J, Wijngaarden GK. Benign myopathy, with autosomal dominant inheritance. A report on three pedigrees. *Brain.* 1976 Mar;99(1):91-100.
4. Bick D and Dimmock D. Whole exome and whole genome sequencing. *Curr Opin Pediatr.* 2011 Dec;23(6):594-600. doi: 10.1097/MOP.0b013e32834b20ec.
5. Bingham J, Sudarsanam S, Srinivasan S. Profiling human phosphodiesterase genes and splice isoforms. *Biochem Biophys Res Commun.* 2006 Nov 10;350(1):25-32. Epub 2006 Sep 11.
6. Botstein D, Risch N. Discovering genotypes underlying human phenotypes: past successes for Mendelian disease, future approaches for complex disease. *Nat Genet.* 2003 Mar;33 Suppl:228-37.
7. Bovolenta M, Neri M, Martoni E, Urciuolo A, Sabatelli P, Fabris M, Grumati P, Mercuri E, Bertini E, Merlini L, Bonaldo P, Ferlini A, Gualandi F. Identification of a deep intronic mutation in the COL6A2 gene by a novel custom oligonucleotide CGH array designed to explore allelic and genetic heterogeneity in collagen VI-related myopathies. *BMC Med Genet.* 2010 Mar 19;11:44. doi: 10.1186/1471-2350-11-44.
8. Carmignac V, Salih MA, Quijano-Roy S, Marchand S, Al Rayess MM, Mukhtar MM, Urtizbera JA, Labeit S, Guicheney P, Leturcq F, Gautel M, Fardeau M, Campbell KP, Richard I, Estournet B, Ferreiro A. C-terminal titin deletions cause a novel early-onset myopathy with fatal cardiomyopathy. *Ann Neurol.* 2007 Apr;61(4):340-51.
9. Claeys KG, Fardeau M, Schröder R, Suominen T, Tolksdorf K, Behin A, Dubourg O, Eymard B, Maisonobe T, Stojkovic T, Faulkner G, Richard P, Vicart P, Udd B, Voit T, Stoltenberg G. Electron microscopy in myofibrillar myopathies reveals clues to the mutated gene. *Neuromuscul Disord.* 2008 Aug;18(8):656-66. doi: 10.1016/j.nmd.2008.06.367. Epub 2008 Jul 23.

10. Ferrer I, Olivé M. Molecular pathology of myofibrillar myopathies. *Expert Rev Mol Med*. 2008 Sep 3;10:e25. doi: 10.1017/S1462399408000793.
11. Goldfarb LG, Dalakas MC. Tragedy in a heartbeat: malfunctioning desmin causes skeletal and cardiac muscle disease. *J Clin Invest*. 2009 Jul;119(7):1806-13. doi: 10.1172/JCI38027. Epub 2009 Jul 1.
12. Hackman P, Vihola A, Haravuori H, Marchand S, Sarparanta J, De Seze J, Labeit S, Witt C, Peltonen L, Richard I, Udd B. Tibial muscular dystrophy is a titinopathy caused by mutations in TTN, the gene encoding the giant skeletal-muscle protein titin. *Am J Hum Genet*. 2002 Sep;71(3):492-500. Epub 2002 Jul 26.
13. Haq RU, Speer MC, Chu ML, Tandan R. Respiratory muscle involvement in Bethlem myopathy. *Neurology*. 1999 Jan 1;52(1):174-6.
14. Hoppman-Chaney N, Peterson LM, Klee EW, Middha S, Courteau LK, Ferber MJ. Evaluation of oligonucleotide sequence capture arrays and comparison of next-generation sequencing platforms for use in molecular diagnostics. *Clin Chem*. 2010 Aug;56(8):1297-306. doi: 10.1373/clinchem.2010.145441. Epub 2010 Jun 18.
15. Isidor B, Lindenbaum P, Pichon O, Bézieau S, Dina C, Jacquemont S, Martin-Coignard D, Thauvin-Robinet C, Le Merrer M, Mandel JL, David A, Faivre L, Cormier-Daire V, Redon R, Le Caignec C. Truncating mutations in the last exon of NOTCH2 cause a rare skeletal disorder with osteoporosis. *Nat Genet*. 2011 Mar 6;43(4):306-8. doi: 10.1038/ng.778.
16. Jöbsis GJ, Boers JM, Barth PG, de Visser M. Bethlem myopathy: a slowly progressive congenital muscular dystrophy with contractures. *Brain*. 1999 Apr;122 (Pt 4):649-55.
17. Jobsis GJ, Bolhuis PA, Boers JM, Baas F, Wolterman RA, Hensels GW, de Visser M. Genetic localization of Bethlem myopathy. *Neurology*. 1996 Mar;46(3):779-82.
18. Kaplan JC. The 2012 version of the gene table of monogenic neuromuscular disorders. *Neuromuscul Disord*. 2011 Dec;21(12):833-61.
19. Lakics V, Karran EH, Boess FG. Quantitative comparison of phosphodiesterase mRNA distribution in human brain and peripheral tissues. *Neuropharmacology*. 2010 Nov;59(6):367-74. doi: 10.1016/j.neuropharm.2010.05.004. Epub 2010 May 21.
20. Lander ES, Linton LM, Birren B, Nusbaum C, Zody MC, Baldwin J, Devon K, Dewar K, Doyle M, FitzHugh W, Funke R, Gage D, Harris K, Heaford A, Howland J, Kann L, Lehoczky J, LeVine R, McEwan P, McKernan K, Meldrim J, Mesirov JP, Miranda C, Morris W, Naylor J, Raymond C, Rosetti M, Santos R, Sheridan A, Sougnez C, Stange-Thomann N, Stojanovic N, Subramanian A, Wyman D, Rogers J, Sulston J, Ainscough R, Beck S, Bentley D, Burton

J, Clee C, Carter N, Coulson A, Deadman R, Deloukas P, Dunham A, Dunham I, Durbin R, French L, Grafham D, Gregory S, Hubbard T, Humphray S, Hunt A, Jones M, Lloyd C, McMurray A, Matthews L, Mercer S, Milne S, Mullikin JC, Mungall A, Plumb R, Ross M, Shownkeen R, Sims S, Waterston RH, Wilson RK, Hillier LW, McPherson JD, Marra MA, Mardis ER, Fulton LA, Chinwalla AT, Pepin KH, Gish WR, Chissole SL, Wendl MC, Delehaunty KD, Miner TL, Delehaunty A, Kramer JB, Cook LL, Fulton RS, Johnson DL, Minx PJ, Clifton SW, Hawkins T, Branscomb E, Predki P, Richardson P, Wenning S, Slezak T, Doggett N, Cheng JF, Olsen A, Lucas S, Elkin C, Uberbacher E, Frazier M, Gibbs RA, Muzny DM, Scherer SE, Bouck JB, Sodergren EJ, Worley KC, Rives CM, Gorrell JH, Metzker ML, Naylor SL, Kucherlapati RS, Nelson DL, Weinstock GM, Sakaki Y, Fujiyama A, Hattori M, Yada T, Toyoda A, Itoh T, Kawagoe C, Watanabe H, Totoki Y, Taylor T, Weissenbach J, Heilig R, Saurin W, Artiguenave F, Brottier P, Bruls T, Pelletier E, Robert C, Wincker P, Smith DR, Doucette-Stamm L, Rubenfield M, Weinstock K, Lee HM, Dubois J, Rosenthal A, Platzer M, Nyakatura G, Taudien S, Rump A, Yang H, Yu J, Wang J, Huang G, Gu J, Hood L, Rowen L, Madan A, Qin S, Davis RW, Federspiel NA, Abola AP, Proctor MJ, Myers RM, Schmutz J, Dickson M, Grimwood J, Cox DR, Olson MV, Kaul R, Raymond C, Shimizu N, Kawasaki K, Minoshima S, Evans GA, Athanasiou M, Schultz R, Roe BA, Chen F, Pan H, Ramser J, Lehrach H, Reinhardt R, McCombie WR, de la Bastide M, Dedhia N, Blöcker H, Hornischer K, Nordsiek G, Agarwala R, Aravind L, Bailey JA, Bateman A, Batzoglou S, Birney E, Bork P, Brown DG, Burge CB, Cerutti L, Chen HC, Church D, Clamp M, Copley RR, Doerks T, Eddy SR, Eichler EE, Furey TS, Galagan J, Gilbert JG, Harmon C, Hayashizaki Y, Haussler D, Hermjakob H, Hokamp K, Jang W, Johnson LS, Jones TA, Kasif S, Kasprzyk A, Kennedy S, Kent WJ, Kitts P, Koonin EV, Korf I, Kulp D, Lancet D, Lowe TM, McLysaght A, Mikkelsen T, Moran JV, Mulder N, Pollara VJ, Ponting CP, Schuler G, Schultz J, Slater G, Smit AF, Stupka E, Szustakowski J, Thierry-Mieg D, Thierry-Mieg J, Wagner L, Wallis J, Wheeler R, Williams A, Wolf YI, Wolfe KH, Yang SP, Yeh RF, Collins F, Guyer MS, Peterson J, Felsenfeld A, Wetterstrand KA, Patrinos A, Morgan MJ, de Jong P, Catanese JJ, Osoegawa K, Shizuya H, Choi S, Chen YJ; International Human Genome Sequencing Consortium. Initial sequencing and analysis of the human genome. *Nature*. 2001 Feb 15;409(6822):860-921.

21. Ley TJ, Ding L, Walter MJ, McLellan MD, Lamprecht T, Larson DE, Kandoth C, Payton JE, Baty J, Welch J, Harris CC, Lichti CF, Townsend RR, Fulton RS, Dooling DJ, Koboldt DC, Schmidt H, Zhang Q, Osborne JR, Lin L, O'Laughlin M, McMichael JF, Delehaunty KD,

- McGrath SD, Fulton LA, Magrini VJ, Vickery TL, Hundal J, Cook LL, Conyers JJ, Swift GW, Reed JP, Alldredge PA, Wylie T, Walker J, Kalicki J, Watson MA, Heath S, Shannon WD, Varghese N, Nagarajan R, Westervelt P, Tomasson MH, Link DC, Graubert TA, DiPersio JF, Mardis ER, Wilson RK. DNMT3A mutations in acute myeloid leukemia. *N Engl J Med*. 2010 Dec 16;363(25):2424-33. doi: 10.1056/NEJMoa1005143. Epub 2010 Nov 10.
22. Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics*. 2009 Jul 15;25(14):1754-60. doi: 10.1093/bioinformatics/btp324. Epub 2009 May 18.
23. Li H, Ruan J, Durbin R. Mapping short DNA sequencing reads and calling variants using mapping quality scores. *Genome Res*. 2008 Nov;18(11):1851-8. doi: 10.1101/gr.078212.108. Epub 2008 Aug 19.
24. Lin X, Tang W, Ahmad S, Lu J, Colby CC, Zhu J, Yu Q. Applications of targeted gene capture and next-generation sequencing technologies in studies of human deafness and other genetic disabilities. *Hear Res*. 2012 Jun;288(1-2):67-76. doi: 10.1016/j.heares.2012.01.004. Epub 2012 Jan 14.
25. Mamanova L, Coffey AJ, Scott CE, Kozarewa I, Turner EH, Kumar A, Howard E, Shendure J, Turner DJ. Target-enrichment strategies for next generation sequencing. *Nat Methods*. 2010 Feb;7(2):111-8. doi: 10.1038/nmeth.1419. Review.
26. Mercuri E, Yuva Y, Brown SC, Brockington M, Kinali M, Jungbluth H, Feng L, Sewry CA, Muntoni F. Collagen VI involvement in Ullrich syndrome: a clinical, genetic, and immunohistochemical study. *Neurology* 2002;58:1354–9.
27. Merlini L, Morandi L, Granata C, Ballestrazzi A. Bethlem myopathy: early-onset benign autosomal dominant myopathy with contractures. Description of two new families. *Neuromuscul Disord*. 1994 Sep-Nov;4(5-6):503-11.
28. Metzker ML. Sequencing technologies - the next generation. *Nat Rev Genet*. 2010 Jan;11(1):31-46. doi: 10.1038/nrg2626. Epub 2009 Dec 8.
29. Montenegro G, Powell E, Huang J, Speziani F, Edwards YJ, Beecham G, Hulme W, Siskind C, Vance J, Shy M, Züchner S. Exome sequencing allows for rapid gene identification in a Charcot-Marie-Tooth family. *Ann Neurol*. 2011 Mar;69(3):464-70. doi: 10.1002/ana.22235. Epub 2011 Jan 20.
30. Nakano S, Engel AG, Waclawik AJ, Emslie-Smith AM, Busis NA. Myofibrillar myopathy with abnormal foci of desmin positivity. I. Light and electron microscopy analysis of 10 cases. *J Neuropathol Exp Neurol*. 1996 May;55(5):549-62.

- Nat Methods. 2010 Feb;7(2):111-8. doi: 10.1038/nmeth.1419.
31. Ng SB, Buckingham KJ, Lee C, Bigham AW, Tabor HK, Dent KM, Huff CD, Shannon PT, Jabs EW, Nickerson DA, Shendure J, Bamshad MJ. Exome sequencing identifies the cause of a mendelian disorder. *Nat Genet.* 2010 Jan;42(1):30-5. doi: 10.1038/ng.499. Epub 2009 Nov 13.
 32. Nyren P, Pettersson B, Uhlen M. Solid phase DNA minisequencing by an enzymatic luminometric inorganic pyrophosphate detection assay. *Anal Biochem* 1993;208:171-5.
 33. Ohlsson M, Hedberg C, Brådvik B, Lindberg C, Tajsharghi H, Danielsson O, Melberg A, Udd B, Martinsson T, Oldfors A. Hereditary myopathy with early respiratory failure associated with a mutation in A-band titin. *Brain.* 2012 Jun;135(Pt 6):1682-94. doi: 10.1093/brain/aws103. Epub 2012 May 9.
 34. Percival JM, Whitehead NP, Adams ME, Adamo CM, Beavo JA, Froehner SC. Sildenafil reduces respiratory muscle weakness and fibrosis in the mdx mouse model of Duchenne muscular dystrophy. *J Pathol.* 2012 Sep;228(1):77-87. doi: 10.1002/path.4054. Epub 2012 Jul 18.
 35. Piluso G, Dionisi M, Del Vecchio Blanco F, Torella A, Aurino S, Savarese M, Giugliano T, Bertini E, Terracciano A, Vainzof M, Criscuolo C, Politano L, Casali C, Santorelli FM, Nigro V. Motor chip: a comparative genomic hybridization microarray for copy-number mutations in 245 neuromuscular disorders. *Clin Chem.* 2011 Nov;57(11):1584-96. doi: 10.1373/clinchem.2011.168898. Epub 2011 Sep 6.
 36. Rajakulendran S, Parton M, Holton JL, Hanna MG. Clinical and pathological heterogeneity in late-onset partial merosin deficiency. *Muscle Nerve.* 2011 Oct;44(4):590-3. doi: 10.1002/mus.22196.
 37. Ronaghi M. Pyrosequencing sheds light on DNA sequencing. *Genome Res.* 2001 Jan;11(1):3-11.
 38. Rothberg JM, Hinz W, Rearick TM, Schultz J, Mileski W, Davey M, Leamon JH, Johnson K, Milgrew MJ, Edwards M, Hoon J, Simons JF, Marran D, Myers JW, Davidson JF, Branting A, Nobile JR, Puc BP, Light D, Clark TA, Huber M, Branciforte JT, Stoner IB, Cawley SE, Lyons M, Fu Y, Homer N, Sedova M, Miao X, Reed B, Sabina J, Feierstein E, Schorn M, Alanjary M, Dimalanta E, Dressman D, Kasinskas R, Sokolsky T, Fidanza JA, Namsaraev E, McKernan KJ, Williams A, Roth GT, Bustillo J. An integrated semiconductor device enabling non-optical genome sequencing. *Nature.* 2011 Jul 20;475(7356):348-52. doi: 10.1038/nature10242.

39. Sanger F, Nicklen S, Coulson AR. DNA sequencing with chain-terminating inhibitors. *Proc Natl Acad Sci U S A*. 1977 Dec;74(12):5463-7.
40. Schröder R, Schoser B. Myofibrillar myopathies: a clinical and myopathological guide. *Brain Pathol*. 2009 Jul;19(3):483-92. doi: 10.1111/j.1750-3639.2009.00289.x.
41. Singleton AB. Exome sequencing: a transformative technology. *Lancet Neurol*. 2011 Oct;10(10):942-6. doi: 10.1016/S1474-4422(11)70196-X.
42. Tawfik DS, Griffiths AD. Man-made cell-like compartments for molecular evolution. *Nat Biotechnol* 1998;16:652-6.
43. Teer JK, Mullikin JC. Exome sequencing: the sweet spot before whole genomes. *Hum Mol Genet*. 2010 Oct 15;19(R2):R145-51. doi: 10.1093/hmg/ddq333. Epub 2010 Aug 12.
44. Tucker T, Marra M, Friedman JM. Massively parallel sequencing: the next big thing in genetic medicine. *Am J Hum Genet*. 2009 Aug;85(2):142-54. doi: 10.1016/j.ajhg.2009.06.022.
45. Udd B, Vihola A, Sarparanta J, Richard I, Hackman P. Titinopathies and extension of the M-line mutation phenotype beyond distal myopathy and LGMD2J. *Neurology*. 2005 Feb 22;64(4):636-42.
46. Ullrich O. Kongenitale, atonisch-sklerotische Muskeldystrophie. *Monatsschr Kinderheilkd* 1930;47:502–510.
47. Venter JC, Adams MD, Myers EW, Li PW, Mural RJ, Sutton GG, Smith HO, Yandell M, Evans CA, Holt RA, Gocayne JD, Amanatides P, Ballew RM, Huson DH, Wortman JR, Zhang Q, Kodira CD, Zheng XH, Chen L, Skupski M, Subramanian G, Thomas PD, Zhang J, Gabor Miklos GL, Nelson C, Broder S, Clark AG, Nadeau J, McKusick VA, Zinder N, Levine AJ, Roberts RJ, Simon M, Slayman C, Hunkapiller M, Bolanos R, Delcher A, Dew I, Fasulo D, Flanigan M, Florea L, Halpern A, Hannenhalli S, Kravitz S, Levy S, Mobarry C, Reinert K, Remington K, Abu-Threideh J, Beasley E, Biddick K, Bonazzi V, Brandon R, Cargill M, Chandramouliswaran I, Charlab R, Chaturvedi K, Deng Z, Di Francesco V, Dunn P, Eilbeck K, Evangelista C, Gabrielian AE, Gan W, Ge W, Gong F, Gu Z, Guan P, Heiman TJ, Higgins ME, Ji RR, Ke Z, Ketchum KA, Lai Z, Lei Y, Li Z, Li J, Liang Y, Lin X, Lu F, Merkulov GV, Milshina N, Moore HM, Naik AK, Narayan VA, Neelam B, Nusskern D, Rusch DB, Salzberg S, Shao W, Shue B, Sun J, Wang Z, Wang A, Wang X, Wang J, Wei M, Wides R, Xiao C, Yan C, Yao A, Ye J, Zhan M, Zhang W, Zhang H, Zhao Q, Zheng L, Zhong F, Zhong W, Zhu S, Zhao S, Gilbert D, Baumhueter S, Spier G, Carter C, Cravchik A, Woodage T, Ali F, An H, Awe A, Baldwin D, Baden H, Barnstead M, Barrow I, Beeson K, Busam D, Carver A, Center A,

- Cheng ML, Curry L, Danaher S, Davenport L, Desilets R, Dietz S, Dodson K, Doup L, Ferriera S, Garg N, Gluecksmann A, Hart B, Haynes J, Haynes C, Heiner C, Hladun S, Hostin D, Houck J, Howland T, Ibegwam C, Johnson J, Kalush F, Kline L, Koduru S, Love A, Mann F, May D, McCawley S, McIntosh T, McMullen I, Moy M, Moy L, Murphy B, Nelson K, Pfannkoch C, Pratts E, Puri V, Qureshi H, Reardon M, Rodriguez R, Rogers YH, Romblad D, Ruhfel B, Scott R, Sitter C, Smallwood M, Stewart E, Strong R, Suh E, Thomas R, Tint NN, Tse S, Vech C, Wang G, Wetter J, Williams S, Williams M, Windsor S, Winn-Deen E, Wolfe K, Zaveri J, Zaveri K, Abril JF, Guigó R, Campbell MJ, Sjolander KV, Karlak B, Kejariwal A, Mi H, Lazareva B, Hatton T, Narechania A, Diemer K, Muruganujan A, Guo N, Sato S, Bafna V, Istrail S, Lippert R, Schwartz R, Walenz B, Yooseph S, Allen D, Basu A, Baxendale J, Blick L, Caminha M, Carnes-Stine J, Caulk P, Chiang YH, Coyne M, Dahlke C, Mays A, Dombroski M, Donnelly M, Ely D, Esparham S, Fosler C, Gire H, Glanowski S, Glasser K, Glodek A, Gorokhov M, Graham K, Gropman B, Harris M, Heil J, Henderson S, Hoover J, Jennings D, Jordan C, Jordan J, Kasha J, Kagan L, Kraft C, Levitsky A, Lewis M, Liu X, Lopez J, Ma D, Majoros W, McDaniel J, Murphy S, Newman M, Nguyen T, Nguyen N, Nodell M, Pan S, Peck J, Peterson M, Rowe W, Sanders R, Scott J, Simpson M, Smith T, Sprague A, Stockwell T, Turner R, Venter E, Wang M, Wen M, Wu D, Wu M, Xia A, Zandieh A, Zhu X. The sequence of the human genome. *Science*. 2001 Feb 16;291(5507):1304-51.
48. Voelkerding KV, Dames SA, Durtschi JD. Next-generation sequencing: from basic research to diagnostics. *Clin Chem*. 2009 Apr;55(4):641-58. doi: 10.1373/clinchem.2008.112789. Epub 2009 Feb 26.
49. Wang JL, Yang X, Xia K, Hu ZM, Weng L, Jin X, Jiang H, Zhang P, Shen L, Guo JF, Li N, Li YR, Lei LF, Zhou J, Du J, Zhou YF, Pan Q, Wang J, Wang J, Li RQ, Tang BS. TGM6 identified as a novel causative gene of spinocerebellar ataxias using exome sequencing. *Brain*. 2010 Dec;133(Pt 12):3510-8. doi: 10.1093/brain/awq323. Epub 2010 Nov 23.
50. Wesolowska A, Dalgaard MD, Borst L, Gautier L, Bak M, Weinhold N, Nielsen BF, Helt LR, Audouze K, Nersting J, Tommerup N, Brunak S, Sicheritz-Ponten T, Leffers H, Schmiegelow K, Gupta R. Cost-effective multiplexing before capture allows screening of 25 000 clinically relevant SNPs in childhood acute lymphoblastic leukemia. *Leukemia*. 2011 Jun;25(6):1001-6. doi: 10.1038/leu.2011.32. Epub 2011 Mar 18.
51. Wheeler DA, Srinivasan M, Egholm M, Shen Y, Chen L, McGuire A, He W, Chen YJ, Makhijani V, Roth GT, Gomes X, Tartaro K, Niazi F, Turcotte CL, Irzyk GP, Lupski JR, Chinault C, Song XZ, Liu Y, Yuan Y, Nazareth L, Qin X, Muzny DM, Margulies M, Weinstock

- GM, Gibbs RA, Rothberg JM. The complete genome of an individual by massively parallel DNA sequencing. *Nature*. 2008 Apr 17;452(7189):872-6. doi: 10.1038/nature06884.
52. Whiteford N, Haslam N, Weber G, Prügél-Bennett A, Essex JW, Roach PL, Bradley M, Neylon C. An analysis of the feasibility of short read sequencing. *Nucleic Acids Res*. 2005 Nov 7;33(19):e171.
53. Zhi D, Chen R. Statistical guidance for experimental design and data analysis of mutation detection in rare monogenic mendelian diseases by exome sequencing. *PLoS One*. 2012;7(2):e31358. doi: 10.1371/journal.pone.0031358. Epub 2012 Feb 10.

RINGRAZIAMENTI

Innanzitutto grazie di cuore alla prof.ssa Ferlini che mi ha accolto nel suo laboratorio già per la tesi di laurea e mi ha nuovamente accompagnato come tutor nel dottorato di ricerca, non solo permettendomi di imparare tecniche innovative interagendo con altri gruppi di ricerca nazionali e internazionali ma anche per l'entusiasmo unico per la ricerca che sa trasmettere.

Ringrazio anche la dott.ssa Francesca Gualandi per avermi seguito passo passo in questi anni, insegnandomi con enorme disponibilità lo scibile della genetica dai concetti di base all'analisi critica dei risultati, grazie inoltre per tutti i consigli e le affettuose critiche che aiutano a crescere.

Tutto questo lavoro è opera di un grandioso team che ringrazio non solo per la collaborazione, i confronti ma anche per l'amicizia: in primis ringrazio la dott.ssa Marcella Neri, partner di lavori e viaggi, sempre disponibile per aiuti e consigli in ogni ambito; a seguire ringrazio la dott.ssa Simona Brioschi che ha condiviso con me il primo giorno di lavoro e da quel momento è diventata il mio punto di riferimento per ogni dubbio, idea o proposta lavorativa. Ringrazio anche la dott.ssa Chiara Passarelli collega di lavoro e di studio che nonostante il breve periodo trascorso insieme ha saputo insegnarmi molto. Ringrazio inoltre la dott.ssa Sofia Falzarano, probabilmente la persona in assoluto più disponibile dentro e fuori dal mondo del lavoro, e la dott.ssa Elena Martoni per i mille momenti di confronto e per avere un carattere che è di esempio per gli altri; infine ringrazio la dott.ssa Francesca Diraimo per il suo impegno e la sua collaborazione.

Ringrazio i collaboratori esterni: il centro di calcolo telethon, CASPUR, per l'analisi dei dati WES e il partner oltre-oceano dei progetti europei, ARIADNE, per le analisi dei pathway e la creazione della lista di geni candidati.

Infine ringrazio i finanziatori che hanno permesso la realizzazione di questo progetto: progetto europeo NMD-CHIP, progetto europeo BIO-NMD.