INDICE

INTRODUZIONE **1**

CAPITOLO 1: CARDIOPATIE E CARDIOMIOPATIE **4**

1.1 DEFINIZIONE 4
1.2 EPIDEMIOLOGIA 5
1.3 CANALOPATIE 7
1.3.1 Stratificazione del rischio *9*
1.3.2 Classificazione *11*
1.3.2 Diagnosi *11*
1.4 CARDIOMIOPATIE 28
1.4.1 Classificazione *32*
1.4.2 Diagnosi *33*

CAPITOLO 2: RUOLO DELLA GENETICA **46**

2.1 LA STRUTTURA DEL GENE 46
2.2 LE MUTAZIONI GENICHE 47
2.3 LA GENETICA APPLICATA ALLE PATOLOGIE CARDIACHE 51

CAPITOLO 3: METODI DI STUDIO DELLE VARIANTI **63**

3.1 TECNICA SANGER 64
3.2 NEXT GENERATION SEQUENCING 65
3.2.2 Preparazione della libreria (sequencing library) *66*
3.2.3 Amplificazione *67*
3.2.4 Sequenziamento *68*
3.2.5 Analisi *70*
3.2.6 Applicazione della Next Generation Sequencing *71*
3.2.7 Tipologie di test NGS *72*
3.2.8 Raccomandazioni operative per i test NGS *77*
3.2.9 Interpretazione dei risultati di un test genetico *78*

CAPITOLO 4: RAZIONALE DELLO STUDIO **84**

CAPITOLO 5: MATERIALI E METODI **86**

CAPITOLO 6: APPROCCIO ALLO STUDIO FUNZIONALE 89

CAPITOLO 7: RISULTATI 110

7.1 APPROFONDIMENTO CASO MEDICO-LEGALE 116

CAPITOLO 8: DISCUSSIONE 126

CAPITOLO 9: CONCLUSIONI 128

BIBLIOGRAFIA 130

RINGRAZIAMENTI 135

INTRODUZIONE

Tra le cardiopatie con correlazione genetica se ne riscontrano molteplici, tra cui le canalopatie e le cardiomiopatie. Si tratta di patologie con rilevanza clinica essenziale, dal momento che predispongono a morte cardiaca improvvisa, tachiaritmie e alterazioni della funzionalità cardiaca con riduzione della frazione di eiezione.

Dal punto di vista epidemiologico le cardiomiopatie sono sicuramente le più frequenti avendo un'incidenza sia per la forma dilatativa sia per la forma ipertrofica che si aggira attorno a 1:250-500. Molto meno frequenti, ma comunque di fondamentale importanza per la prognosi correlata, sono le canalopatie, le cui forme più comuni sono la sindrome di Brugada e la sindrome del QT lungo, patologie rare, con un'incidenza attorno a 1:2.000 ed una frequenza generalmente più elevata nei soggetti giovani di sesso maschile (Brugada).

Negli ultimi anni in seguito al sequenziamento del genoma umano si è molto sviluppato il campo dell'indagine genetica, portando ad enormi risvolti in termini di diagnosi e di gestione delle patologie. Questa ha permesso nel corso del tempo di identificare associazioni sempre più strette tra mutazioni geniche ed i diversi fenotipi associati ad ogni singolo gene mutato (associazioni genotipo-fenotipo).

Le mutazioni che possono interessare i geni sono numerosissime e molte risultano essere ancora oggi misconosciute, per questo non è disponibile un test genetico ufficialmente riconosciuto che con sicurezza possa attestare la presenza della patologia.

Oggi è noto che nonostante le moderne tecniche di diagnosi clinica ci sono casi in cui la patologia, pur presente, non viene rilevata con le metodiche strumentali, ed è quindi particolarmente importante ricercare l'effettiva correlazione genotipo-fenotipo per avere una miglior capacità discriminativa.

In alcuni casi, durante la ricerca genetica, si incorre nell'identificazione di nuove mutazioni, mai identificate prima, definite varianti di incerto significato (VUS), la cui patogenicità dovrà poi essere confermata tramite studi funzionali, oggetto dello studio sperimentale di seguito presentato.

Durante un test genetico capita che l'analisi identifichi mutazioni mai riscontrate e mai studiate prima che vengono definite di incerto significato; in questo caso si procede con studi volti ad identificare la benignità o la patogenicità della mutazione: inizialmente si utilizzano i test di segregazione per la ricerca della medesima mutazione nei familiari, successivamente si procede con una valutazione analizzando i dati in letteratura e, qualora non vi sia corrispondenza, si. prosegue utilizzando tools online per il confronto con varianti già identificate da altri laboratori.

A tal fine risulta essere di fondamentale importanza l'acquisizione del maggior numero possibile di informazioni anamnestiche personali e familiari, in modo da poter restringere il più possibile il campo di ricerca e poter essere più precisi nel categorizzare una nuova variante.

Se si evidenzia una variante mai identificata prima, seguono fasi successive degli studi funzionali: lo studio in vitro e lo studio in vivo, che richiedono risorse economiche e di tempo maggiori rispetto al primo step.

L'eterogeneità delle mutazioni rende l'analisi genetica non di facile e rapida effettuazione per la formulazione della diagnosi, che si basa ancora oggi principalmente su dati clinici, strumentali e sulla storia familiare dei pazienti.

Clinicamente il dato più frequentemente riscontrato sono episodi sincopali in soggetti giovani senza cause apparenti, o episodi di alterazioni elettriche cardiache, con pattern ECGgrafici tipici (canalopatie), o con alterazioni evidenti all'ecocardio (cardiomiopatie).

Lo scopo di questo studio è confermare o smentire quei reperti la cui diagnosi non può essere garantita con sicurezza dalle sole metodiche di indagine clinica e strumentale. Tale

processo di verifica si sviluppa attraverso una combinazione degli elementi oggi a disposizione utili a ridurre ai minimi termini il rischio di incorrere in errori, comunque ancora non trascurabile con le metodiche esistenti.

CAPITOLO 1: CARDIOPATIE E CARDIOMIOPATIE

1.1 Definizione

Con i termini di cardiopatia e cardiomiopatia si fa riferimento a patologie di interesse cardiaco, siano esse su base organica o su base funzionale.

Queste possono essere classificate in due macro-gruppi: congenite, se presenti sin dalla nascita, o acquisite. Le cause eziopatogeniche sono molteplici, come ad esempio l'invecchiamento e le infezioni.

All'interno del gruppo delle cardiopatie, le patologie più strettamente coinvolte dal punto di vista genetico sono sicuramente le aritmie che comprendono un gruppo ampio ed eterogeneo di anomalie elettriche del cuore, con o senza cardiopatia strutturale associata. Possono essere innocue, predisporre all'insorgenza di ictus o di embolie o possono essere potenzialmente fatali, tali da richiedere un trattamento di emergenza, poiché possono provocare la morte cardiaca improvvisa, una delle prime cause di morte per patologia cardiovascolare nei Paesi sviluppati.

Le sindromi aritmiche potenzialmente fatali sono comprese nella macrocategoria delle "canalopatie cardiache".[1]

Per quanto concerne la valutazione delle cardiomiopatie è nota l'importanza dello studio genetico, soprattutto grazie all'utilizzo di nuove tecniche di sequenziamento del DNA che permettono in maniera sempre più precisa di identificare mutazioni associate.

1.2 Epidemiologia

I dati epidemiologici più affidabili delle cardiomiopatie sono disponibili solo nei paesi sviluppati in cui i dati di prevalenza si basano sull'applicazione di valutazioni e criteri diagnostici stabiliti.

	Children (1–puberty)	Adults (19–64 y)
HCM	Uncommon	1:250/500*
DCM	Uncommon†	1:250/500‡
ARVC	Uncommon	1:2000/5000
RCM	Uncommon	Uncommon
LQT	1:2000	1:2000
Brugada (type 1 ECG)	Uncommon	1:2000/5000§
CPVT	1:5000/10 000	1:5000/10 000

Tabella 1. *Prevalenza delle più comuni condizioni cardiache ereditarie*

La diagnosi di cardiomiopatia dilatativa (DCM) è stata difficile da standardizzare a causa dell'influenza delle dimensioni del corpo, della differente massa muscolare interindividuale e dell'eterogeneità biologica sulle misurazioni del volume ventricolare sinistro e della funzione contrattile. I dati sulla prevalenza variano anche quando l'espressione della malattia è correlata all'età (cardiomiopatia ipertrofica, HCM), all'età e al sesso (sindrome di Brugada) o è presente alla nascita (sindrome di QT lungo). La tabella 1 presenta un riassunto della prevalenza della malattia nei bambini e negli adulti per HCM, DCM, ARVC (cardiomiopatia aritmogena del ventricolo destro), RCM (cardiomiopatia restrittiva) e canalopatie. Una prevalenza della malattia da 1:250 a 500 per l'HCM negli adulti sembra essere simile in tutte le etnie e l'espressione della malattia si verifica di solito negli adolescenti e nei giovani adulti, mentre è rara nei bambini. Mancano dati affidabili sull'epidemiologia della DCM, ma le stime suggeriscono una incidenza e prevalenza di circa il doppio di quanto osservato nella HCM. L'incidenza nei bambini è maggiore nel primo anno di vita (4,58/100.000) ed è molto meno comune da 1 a 18 anni (0,34/100.000). La prevalenza di ARVC non è stata studiata sistematicamente, ma è stimata tra 1/2.000 e 5.000; dati epidemiologici solidi mancano in parte a causa delle molteplici valutazioni cliniche richieste

per la diagnosi. L'ARVC nell'infanzia è rara e, come nell'HCM, l'espressione della malattia di solito si verifica durante l'adolescenza e la prima età adulta, ma può svilupparsi a qualsiasi età, in linea con il concetto di cardiomiopatia geneticamente determinata.

Per la sindrome di LQT, è stata stabilita una prevalenza di 1:2.000 nei neonati, e i dati suggeriscono che questa è simile per tutte le età e razze, sebbene le percentuali di eventi differiscano per LQT1, LQT2 e LQT3 per quanto riguarda l'età e il sesso.

La prevalenza della tachicardia ventricolare polimorfa catecolaminergica non è nota e le cifre riportate 1:5.000/10.000 possono essere una sottovalutazione di questa condizione che si presenta in genere con sincope correlata all'esercizio e morte improvvisa in bambini e adolescenti nonostante ECG ed ecocardiografia bidimensionale nella norma.[2]

La sindrome di Brugada è solitamente considerata una patologia dei giovani adulti di sesso maschile, apparentemente più comune nel Sud-Est Asiatico, in cui le aritmie esordiscono ad un'età media di 40 anni e la morte improvvisa avviene generalmente nel sonno.[3] La prevalenza mondiale è di circa 1-5:10.000, ma nel Sud-Est Asiatico raggiunge lo 0,5-1 per 1.000.[4]

1.3 Canalopatie

Le sindromi aritmiche ereditarie potenzialmente fatali comprese nella macrocategoria delle canalopatie, come la sindrome congenita del QT lungo (Long-QT Syndrome, LQTS), la sindrome di Brugada, la tachicardia ventricolare polimorfa catecolaminergica ed i disturbi correlati, determinano alterazioni elettriche con la propensione a causare aritmie fatali in soggetti con cuore strutturalmente normale.[5]

La **sindrome del QT lungo** è una patologia a trasmissione autosomica dominante e penetranza variabile descritta per la prima volta nel 1957 da Jervell e Lange-Nielsen, che può essere congenita (causata da un'alterazione genica) o acquisita (farmaco-indotta con farmaci quali antiaritmici di classe 1C, antipsicotici, eccetera; oppure secondaria ad anomalie elettriche come l'ipopotassiemia) e comprende un gruppo distinto di canalopatie cardiache caratterizzate da:

- Ripolarizzazione ritardata del miocardio;
- Allungamento del QT;
- Aumento del rischio di sincope, convulsioni e morte improvvisa in soggetti peraltro sani, con cuore strutturalmente normale.[6]

Questo difetto della ripolarizzazione è quasi sempre senza conseguenze: tuttavia, fattori scatenanti come l'attività fisica, il nuoto, le emozioni, stimoli uditivi (ad esempio la sveglia) ed il periodo postpartum possono, in rari casi, causare l'instabilità elettrica, scatenando aritmie potenzialmente fatali come la torsione di punta.

Anche se normalmente il ritmo cardiaco si normalizza spontaneamente dopo un episodio transitorio di sincope, nel 5% dei soggetti LQTS non trattati e non diagnosticati le aritmie fatali sono l'evento sentinella. Tuttavia, si stima che quasi la metà dei soggetti che muoiono di morte improvvisa secondaria a tale disturbo aritmogeno possa avere mostrato dei segni premonitori che non sono stati riconosciuti.

La **sindrome di Brugada**, invece, è una condizione aritmica ereditaria a trasmissione autosomica dominante, descritta per la prima volta nel 1992[7], caratterizzata da un tracciato ECG che comprende sopraslivellamento ST convesso con onda T negativa nelle derivazioni precordiali destre V1-V3 ed un aumento del rischio di morte improvvisa per tachiaritmia ventricolare polimorfa.[8,9] La penetranza e l'espressività della malattia sono molto variabili e vanno da soggetti asintomatici per tutta la vita a morte nel primo anno di vita.[9] La patologia può essere sporadica (60%) o familiare (40%).

La sindrome di Brugada si riscontra comunemente in soggetti di sesso maschile giovani, più frequentemente del Sud-Est asiatico, con eventi aritmici come primo episodio incorso prima dei 40 anni e morte cardiaca improvvisa tipicamente occorsa durante il sonno.[3,4] Tuttavia, la sindrome di Brugada è stata identificata anche in bambini piccoli e negli adolescenti.

Si tratta di una canalopatia che causa disfunzione cardiaca nei canali coinvolti nel potenziale d'azione. La disfunzione elettrica favorisce l'instaurarsi di aritmie e rappresenta il disordine principale responsabile delle complicanze aritmiche, in assenza di alterazioni strutturali.

Si ritiene che la sindrome di Brugada sia responsabile del 4-12% dei casi di morte cardiaca improvvisa e del 20-50% dei casi di morte cardiaca improvvisa in soggetti senza sottostanti alterazioni.[10]

1.3.1 Stratificazione del rischio

SINDROME DEL QT LUNGO

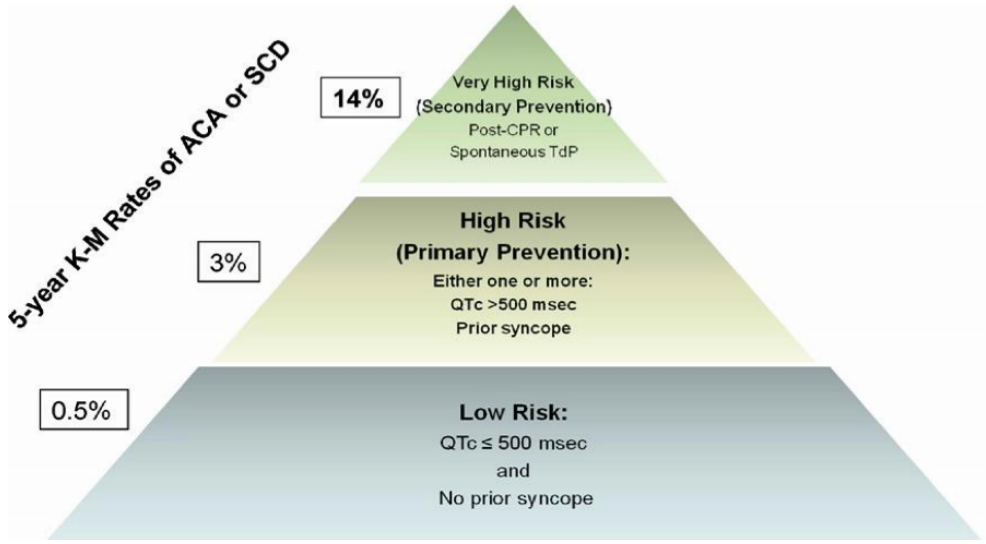


***Figura 1**. Stratificazione del rischio di LQTS*

I pazienti ad alto rischio sono coloro che hanno o un arresto resuscitato o una torsione di punta. Rappresentano il 14% del totale.

I pazienti con rischio intermedio sono coloro che hanno un QT corretto superiore a 500 ms ed episodi di sincope pregressi. Rappresentano il 3% del totale.

I pazienti a basso rischio sono coloro che hanno un QT corretto inferiore a 500 ms e assenza di sincope pregressa. [11]

SINDROME DI BRUGADA

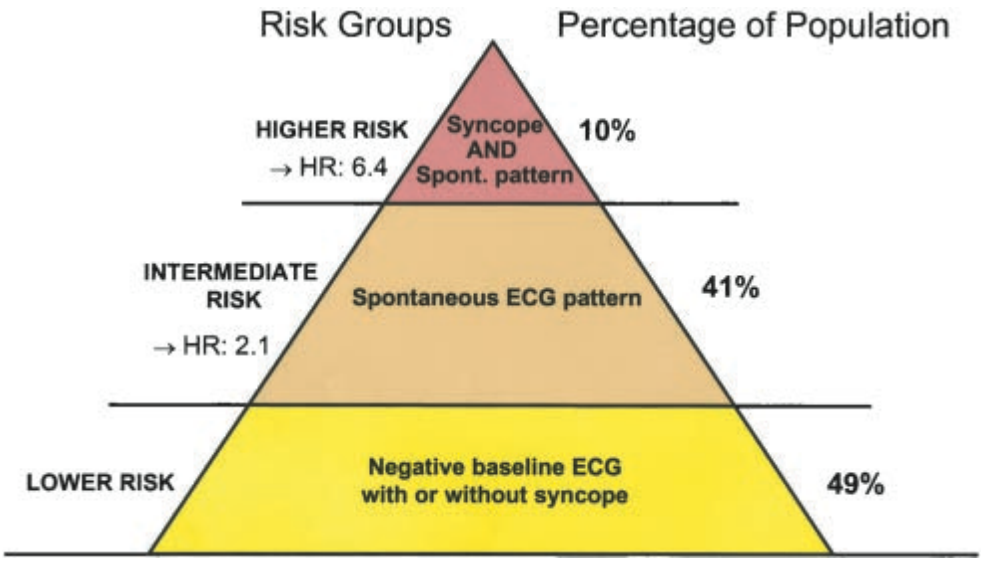


***Figura 2**. Stratificazione del rischio della sindrome di Brugada*[12]

Sono a più alto rischio di eventi aritmici e di morte cardiaca improvvisa i pazienti con S. di Brugada e:

- Tachiaritmie ventricolari documentate o arresto cardiaco resuscitato;
- Sincope;
- Pattern di tipo 1 spontaneo o indotto;
- Familiarità per morte cardiaca improvvisa.

1.3.2 Classificazione

Non è possibile effettuare una classificazione clinica né per la sindrome del QT lungo, né per la sindrome di Brugada, dal momento che, a parità di patologia, fenotipicamente non sembrano esserci grosse distinzioni tra una forma e l'altra.

Un aspetto sempre più rilevante, invece, è rappresentato dalla classificazione su base genetica, che per entrambe le patologie ci permette di fare una suddivisione identificando dei gruppi numerici in base ai geni interessati ed alle diverse funzioni svolte. Si tratta di una classificazione in continuo aggiornamento, perché il campo della genetica si sta oggi mostrando di fondamentale importanza ed è nel pieno del suo sviluppo e della sua progressione.

Man mano che nuove mutazioni vengono scoperte si ampliano le classificazioni conosciute.

Ad oggi è stato possibile identificare un totale di 15 mutazioni per la sindrome del QT lungo[6], dove le più frequenti in assoluto sono le prime 3 categorie, ed un insieme di 21 mutazioni per la sindrome di Brugada.

1.3.2 Diagnosi

Le canalopatie dovrebbero essere sospettate sulla base delle caratteristiche ECGgrafiche, della presentazione clinica e della storia familiare, ma è importante considerare che la presentazione è variabile, essendo patologie a penetranza incompleta.

Diagnosi clinica

Il sintomo tipico che porta alla diagnosi di **sindrome del QT lungo** è la sincope scatenata da attività fisica o da stress emotivo (paura, spavento, sorpresa, risveglio improvviso al suono di una sveglia), che nella maggior parte dei casi avviene in maniera immediata, senza sintomi premonitori. In alcuni casi la prima manifestazione della malattia è l'arresto cardiaco o morte

improvvisa: di fronte a questi episodi in un soggetto giovane è bene sospettare la patologia. Un aspetto fondamentale per la diagnosi clinica è la descrizione dell'episodio e quindi l'anamnesi al fine di poter fare un adeguato inquadramento ed una distinzione rispetto a patologie quali la sincope vaso-vagale o l'epilessia. Importante, soprattutto nelle forme eredo-familiari, inoltre, risulta essere la valutazione della storia clinica non solo del paziente, ma anche dei familiari.
Bisogna inoltre considerare che ci sono situazioni che possono indurre l'allungamento dell'intervallo QT:

- Farmaci;
- Ipokaliemia;
- Alcune condizioni neurologiche inclusa l'emorragia subaracnoidea;
- Alterazioni della struttura cardiaca.

Per quanto riguarda la **sindrome di Brugada**, è possibile identificare un numero maggiore di presentazioni cliniche:

- Fibrillazione ventricolare (FV) o morte cardiaca improvvisa resuscitata;
- Sincope;
- Respiro agonico notturno;
- Palpitazioni;
- Toracoalgia aspecifica;
- Assenza di sintomi.

Nella sindrome di Brugada, inoltre, i sintomi solitamente si verificano a riposo o nel sonno, durante uno stato febbrile superiore a 38 °C o in condizioni vagotoniche (per esempio in fase postprandiale), e solo raramente durante esercizio fisico.

Diagnosi strumentale

La diagnosi strumentale rimane senza ombra di dubbio un passaggio imprescindibile per giungere a diagnosi di questo tipo di patologie. La metodica diagnostica cardine è in questo ambito rappresentata dall'ECG a riposo, pur potendo esistere forme silenti (25% dei casi).[6]

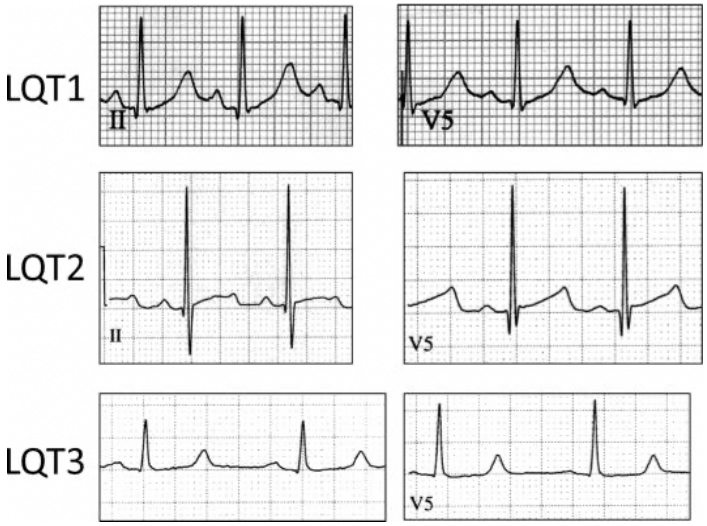


***Figura 3**. ECG pattern principali LQTS*

Nel sospetto di una **sindrome del QT lungo**, la valutazione dell'intervallo QT andrebbe fatta su almeno 5-6 cicli cardiaci, dall'inizio del primo complesso QRS fino alla fine dell'onda T successiva, a livello delle derivazioni II e V_5 o V_6.[6,13] In particolare, come mostrato in figura 3, nel caso del sottotipo LQTS1 si trova un'onda T con normale morfologia e base allargata; nel caso del LQTS2 l'onda T è di basso voltaggio e aspetto bifido; nel caso del LQTS3, invece, si riscontra un inizio tardivo dell'onda T.

Per una corretta valutazione, il referto ECGgrafico andrebbe corretto per la frequenza cardiaca usando la formula di Bazett dove l'intervallo RR è l'intervallo del ciclo cardiaco considerato dall'inizio di un complesso QRS al seguente, misurato in secondi (l'intervallo QT viene invece espresso in millisecondi):

$$QTc = \frac{\text{QT interval in seconds}}{\sqrt{\text{cardiac cycle in seconds}}} = \frac{QT}{\sqrt{RR}}$$

RATING	1-15 ANNI	UOMINI	DONNE
Normale	< 440	< 430	< 450
Borderline	440-460	430-450	450-470
Patologico	> 460	> 450	> 470

***Tabella 2**. Valori corretti con Bazett suggestivi per diagnosi prolungamento dell'intervallo QT*

Il QTc all'ECG a riposo non è né completamente sensibile né specifico per la diagnosi di LQTS. Circa il 25% degli individui con LQTS confermato dall'identificazione di una variante patogena in un gene associato a LQTS può avere, infatti, un intervallo QTc normale (*LQTS nascosto*).

Per questo motivo, si tende ad aggiungere altri elementi che possano aiutare a fare una valutazione il più possibile attendibile in soggetti con valori di QT "incerti":

- ECG sotto sforzo, che comunemente mostra l'incapacità del tratto QT di accorciarsi normalmente e il prolungamento dell'intervallo QT[14] e anomalie caratteristiche a carico dell'onda T;[15]
- Misurazione dell'intervallo QT nel passaggio da clinostatismo a ortostatismo;
- Test di provocazione farmacologica endovenosa (per esempio con epinefrina[16]), che può essere di aiuto nel dimostrare un allungamento inappropriato dell'intervallo QT. Il rischio di provocare un'aritmia è in questo caso accettabile se il test viene eseguito in ambienti protetti e con l'esperienza adeguata nella gestione e nel controllo dell'eventuale insorgenza.[17]

Per poter fare diagnosi di sindrome del QT lungo ci sono dei parametri diagnostici da rispettare, di seguito presentati (Tabella 3).

Criteria	Points
ECG Findings	
• QTc > 480	3
• QTc > 460 – 479	2
• QTc > 440 – 459 (male patients)	1
• Torsade de pointes	2
• T-wave alternans	1
• Notched T-waves in 3 leads	1
• Low heart rate for age	0.5
Clinical History	
• Syncope with stress	2
• Syncope without stress	1
• Congenital deafness	0.5
Family History	
• Family member with definite LQTS	1
• Unexplained sudden cardiac death in immediate family (age < 30 yr)	0.5

Punteggio:

≤ 1.0 = bassa probabilità

1.5-3.0 = probabilità intermedia

≥ 3.5 = Alta probabilità

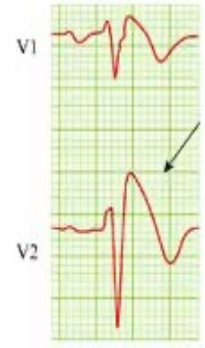
Tabella 3. *Criteri diagnostici per LQTS*[18]

La diagnosi di LQTS è stabilita nel probando con almeno uno dei seguenti criteri:

- Punteggio ≥ 3.5 in assenza di una causa secondaria del prolungamento QT;
- In presenza di un intervallo QT corretto ≥ 500 ms in ECG ripetuti in assenza di cause secondarie del prolungamento QT;
- Identificazione di una variante patogenica in uno dei geni noti per l'associazione con la patologia.

Quando viene identificato un paziente affetto da LQTS, l'ECG andrebbe fatto anche ai familiari di primo grado per determinare se ci sono altri soggetti affetti in famiglia.

La classificazione della **sindrome di Brugada** si basa sui differenti reperti ECGgrafici:



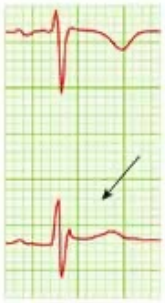
- **Tipo 1** (**Coved Type**): elevazione del segmento ST ≥ 2 mm seguita da un'onda T negativa in almeno una derivazione tra le derivazioni precordiali destre (V1-V3) con elettrodi posizionati al II, III, IV spazio intercostale.

***Figura 4**. Sindrome di Brugada Coved Type*

- **Tipo 2** (**Saddle-Back-Type**): elevazione del segmento ST ≥ 0,5 mm (generalmente concava tipo "sella") del segmento ST. Il segmento ST è seguito solitamente da un'onda T positiva in V2 e di variabile morfologia in V1.



***Figura 5**. Sindrome di Brugada Saddle-Back type*



- **Tipo 3**: elevazione del segmento ST di 0,1 mm in V1-V2 e con morfologia di tipo concavo (saddle-back) oppure con piccola impronta tendente alla morfologia "coved".

***Figura 6**. Sindrome di Brugada Tipo 3*

Il posizionamento delle derivazioni precordiali al II e III spazio intercostale oltre che al IV spazio intercostale aumenta la sensibilità diagnostica nella rilevazione del pattern ECG di tipo I.

I criteri diagnostici da considerare per la diagnosi di sindrome di Brugada sono di seguito riportati.

Sopraslivellamento del tratto ST di tipo 1 (tipo concava) ≥ 2 mm in più di una derivazione precordiale destra (V1-V3): ▪ **Spontaneamente** ▪ **O dopo esposizione a sodio-bloccante**	
E	
Uno dei seguenti:	
▪ **Fibrillazione ventricolare documentata** ▪ **Tachicardia ventricolare polimorfa (auto-terminante)** ▪ **Inducibilità di aritmie ventricolari con stimolazione elettrica programmata**	Aritmia ventricolare documentata
▪ **Storia familiare di morte improvvisa prima dei 45 anni** ▪ **Presenza di un ECG di tipo coved nei membri della famiglia**	Storia familiare
▪ **Sincope** ▪ **Respirazione agonica notturna**	Sintomi correlati ad aritmia
Esclusione altri fattori che potrebbero spiegare l'anomalia dell'ECG	

***Tabella 4**. Criteri diagnostici sindrome di Brugada*[8]

Diagnosi genetica

Grazie alle nuove tecnologie ed alle scoperte fatte negli ultimi anni, oggi assistiamo ad un notevole incremento di richieste di accertamenti genetici affianco alle diagnosi cliniche e strumentali, comunque ancora imprescindibili per giungere ad una diagnosi definitiva di canalopatia. In alcuni casi, infatti, come abbiamo già affermato in precedenza, nonostante l'utilizzo delle più moderne tecniche strumentali, non è possibile sbilanciarsi nel fare diagnosi e potrebbe quindi rendersi indispensabile un'indagine molecolare, in considerazione anche

dell'importanza dal punto di vista prognostico della patologia se non adeguatamente inquadrata e gestita.

Nel caso del **LQTS** il rendimento dei test genetici si è dimostrato essere il più alto (72%) tra gli individui testati con la più alta probabilità clinica. Pertanto, ci si può aspettare che l'attuale test genetico catturi circa tre quarti degli individui LQTS fenotipicamente affetti, mentre un test genetico negativo in un soggetto con LQTS clinico (cioè, LQTS genotipo negativo/fenotipo positivo) non fornisce alcuna base per una rivalutazione della diagnosi.[19] Nonostante ciò, un test genetico positivo può influenzare le decisioni terapeutiche e può fornire i mezzi per una precisa classificazione dello stato di "portatore" di parenti potenzialmente a rischio. Inoltre, i test genetici possono essere importanti nell'identificazione di LQTS nascosti, perché una minoranza significativa (dal 25% al 50%) di individui con LQTS geneticamente provata ha un QTc non diagnostico.[20]

I pazienti con sospetta **sindrome di Brugada** (BrS) potrebbero essere sottoposti a test genetici molecolari, purché si riconosca che il rendimento dei test genetici per BrS attualmente disponibili sia di circa il 20% - 30%. Inoltre, ad eccezione della recente associazione tra il tipo di mutazione SCN5A e la severità clinica, il ruolo primario dei test genetici per BrS è la conferma diagnostica della causa principale di un caso indice (*probando*). Questa procedura diagnostica è seguita da test genetici di conferma dei membri della famiglia (*test di segregazione*) per distinguere coloro che hanno bisogno di un continuo monitoraggio clinico e di misure preventive (come ad esempio evitare determinati farmaci), da coloro che possono essere classificati come clinicamente, elettrocardiograficamente e geneticamente non affetti.

La metodica principale oggi utilizzata per fare diagnosi genetica nel probando è la "*next generation sequencing*" (NGS), mentre per la ricerca nei familiari di primo grado si utilizza la tecnica Sanger.[21]

I geni del LQTS

L'allungamento del tratto QT può derivare sia da un decremento della ripolarizzazione delle correnti del potassio, sia da un ingresso tardivo inappropriato del sodio nei miociti. Più comunemente, il meccanismo patogenico alla base è rappresentato da un ritardo nella ripolarizzazione a causa di mutazioni a carico della subunità α dei canali ionici di sodio e potassio, in modo da delineare un numero considerevole di varianti possibili:

- LQT1: ripolarizzazione ritardata del canale del potassio a corrente lenta;
- LQT2: ripolarizzazione ritardata del canale del potassio a corrente rapida;
- LQT3: depolarizzazione prolungata del canale del sodio;
- LQT5: mutazione coinvolgente KCNQ1;
- LQT6: mutazione coinvolgente KCNH2.

Le alterazioni fin qui citate rappresentano la maggioranza delle varianti mutate ad oggi identificate ed è questo gruppo di 5 geni a caratterizzare LQTS come una canalopatia.[22]

Negli ultimi anni poi, grazie alle nuove tecnologie utilizzate ed agli studi effettuati, è stato possibile ampliare la classificazione con nuove mutazioni, fino ad arrivare oggi ad una casistica di 15 forme genetiche identificate (Tabella 5).

LOCUS	GENE	CROMOSOMA	CORRENTE	FREQUENZA
LQT1	KCNQ1	11p15.5	I_{KS}	57%
LQT2	KCNH2	7q36.1	I_{Kr}	23%
LQT3	SCN5A	3p21	I_{Na}	16%
LQT4	ANK2	4q25-q27	I_{Na}	raro
LQT5	KCNE1	21q22.1-q22.2	I_{Ks}/I_{Kr}	raro
LQT6	KCNE2	21q21.1	I_{Kr}/I_{Ks}	raro

LQT7	KCNJ2	17q23.1-q24.2		raro
LQT8	CACNA1C	12p13.3	$I_{Ca,L}$	raro
LQT9	CAV3	3p25	I_{Na}	raro
LQT10	SCN4B	11q23.3	I_{Na}	raro
LQT11	AKAP9	7q21-q22		raro
LQT12	SNTA1	20q11.2	I_{Na}	raro
LQT13	KCNJ5	11q24.3	$K_{ir}3.4$	raro
LQT14	CALM1	14q32.11		raro
LQT15	CALM2	2p21		raro

***Tabella 5**. Mutazioni identificate per LQTS aggiornate al 8 febbraio 2018* [23]

La struttura di ogni canale ionico consiste in una serie di aminoacidi con una regione N-terminale, 6 segmenti transmembrana con anse citoplasmatiche intracellulari e un'area dei pori, e una regione C-terminale.

- KCNQ1 e KCNH2 appartengono alla famiglia dei canali ionici del potassio voltaggio dipendenti, nei quali un tetramero di 4 subunità α costituisce ciascun canale.[24,25]
- SNC5A è un canale del sodio costituito da una singola proteina, un tetramero con 4 divisioni di subunità α.[26]

Due distinti meccanismi biofisici mediano la ridotta corrente ripolarizzante nei pazienti con mutazioni del canale del potassio:

1) **Difetti di coassemblaggio**, in cui le subunità mutate non vengono trasportate correttamente alla membrana cellulare e non riescono a incorporarsi nel canale tetramerico, con l'effetto netto di una riduzione del 50% della funzione del canale (***aploinsufficienza);***

2) **Formazione di canali difettosi** che coinvolgono subunità mutate, con la proteina del canale alterata trasportata alla membrana cellulare e risultante in un canale disfunzionale con una riduzione del 50% della corrente del canale (***effetto dominante negativo***).

Nei pazienti con mutazioni del canale del sodio SCN5A, il canale non si chiude correttamente dopo la depolarizzazione iniziale e la continua fuoriuscita di sodio nel canale determina un prolungamento del potenziale d'azione.

Quelli citati sono i geni maggiormente coinvolti ma, come rappresentato nella Tabella 5, ci sono altre mutazioni oggi identificate, ognuna responsabile di alterazioni differenti a seconda della funzione svolta:

- Ankirina B: adattatore della membrana citoscheletrica;[27]
- KCNJ2: riduzione della corrente $K_{ir}2.1$ e fenotipo dominato da alterazioni scheletriche (Andersen-Tawil syndrome);[28]
- CACNA1C: aumento nella corrente $Ca_V1.2$ e associazione con sindattilia in entrambe le mani ed i piedi (Timothy syndrome);[29]
- CAV3: aumento delle correnti tardive del sodio;[30]
- SCN4B: aumento delle correnti tardive del sodio.[31]

Si è visto che LQTS, associata con varianti patogeniche bialleliche o eterozigosi per varianti patogeniche in due geni differenti, si caratterizza per un fenotipo più severo.
Approssimativamente nel 20% delle famiglie che rispondono ai criteri clinici diagnostici per LQTS non ci sono varianti patogeniche nei geni noti fino ad oggi, a testimoniare che c'è ancora molto lavoro da fare per l'identificazione di tutte le strutture geniche coinvolte nella patologia in esame.

I geni della sindrome di Brugada

La sindrome di Brugada (BrS) è tipicamente ereditata in modo autosomico dominante anche se fino alla metà dei casi può essere di natura sporadica. Ad oggi sono stati identificati 21 geni di suscettibilità BrS (Tabella 6) tra cui la mutazione con perdita di funzione del gene SCN5A associato al canale del sodio, che rappresenta il substrato genetico più comune (15-30%). Oltre alla disfunzione del canale del sodio, le mutazioni coinvolgono la subunità α_1 del canale del Calcio di tipo L (CACNA1C32), β_2 (CACNB2B32) e $\alpha_2\delta$ (CACNA2D133) che possono causare dal 10% al 15% dei fenotipi, in particolare se accompagnati da brevi intervalli del tratto QT.

LOCUS	GENE	CROMOSOMA	PROTEINA	FUNZIONE	FREQUENZA
BrS1	SCN5A	3p21	$Na_v1.5$	Loss of function	15%-30%
BrS2	GPD1-L	3p24	G3PD1L	Loss of function	10%-15%
BrS3	CACNA1C	12p13.3	$Ca_v1.2$	Loss of function	10%-15%
BrS4	CACNB2b	10p12.33	$Ca_v\beta2$	Loss of function	10%-15%
BrS5	SCN1B	19q13.1	$Na_v\beta1$	Loss of function	Rare
BrS6	KCNE3	11q13-14	MiRP2, I_{to}/I_{ks}	Reduced pacemaker (I_f) current	Rare
BrS7	SCN3B	11q23.2	$Na_v\beta3$	Gain of function	Rare
BrS8	KCNH2	7q35	Kv11.1, Ikr	Gain of function	Rare
BrS9	KCNJ8	12p12.1	Kir6.1, IK-ATP	Gain of function	Rare
BrS10	CACNA2D1	7q21-22	Cavα2δ1	Gain of function	Rare
BrS11	RANGRF	17p13.1	MOG1	Increased IKATP current	Rare

BrS12	KCNE5	Xq22.3	MiRP4, Kv4.3, Ito	Loss of function	Rare
BrS13	KCND3	1p13.2	Kv4.3, Ito	Loss of function	Rare
BrS14	HCN4	15q24.1	If	Loss of function	Rare
BrS15	SLMAP	3p21.2-p14.3	SLMAP	Loss of function	Rare
BrS16	TRPM4	19q13.33	NSCCa	Loss of function	Rare
BrS17	SCN2B	11q23	Navβ2	Loss of function	Rare
BrS18	SCN10A	3p22.2	Nav1.8	Loss of function	Rare
BrS19	HEY2	6q22	Nav1.5	Loss of function	Rare
BrS20	PKP2	12p11.21	plakophilin-2	plakophilin-2 cause sodium current deficit	Rare
BrS21	ABCC9	12p12.1	SUR2A, IK-ATP	Gain of function	Rare

Tabella 6. *Mutazioni identificate per la Sindrome di Brugada*

SCN5A è responsabile del rapido aumento del potenziale d'azione cardiaco nella fase 0 e le variazioni patogene provocano una disfunzione del canale del sodio, che porta ad un rallentamento della conduzione nel cuore. Studi funzionali di queste mutazioni hanno rivelato una perdita di funzione nei canali del sodio che è stata influenzata attraverso diversi meccanismi tra cui: la diminuzione dell'espressione della proteina del canale del sodio ($Na_v1.5$) nel sarcolemma[32], l'espressione di canali non funzionali, o proprietà di gating alterate (attivazione ritardata, inattivazione precoce e più rapida, inattivazione lenta potenziata e recupero ritardato dall'inattivazione).[33]

Oltre ai geni SCN5A, molti geni correlati a BrS svolgono un ruolo nella regolazione della funzione del canale del sodio. Diverse varianti patogene in tre geni (SCN1B, SCN2B e SCN3B)

che codificano per le subunità β del canale del sodio $Na_v1.5$, si è scoperto essere responsabili dell'alterazione della funzione del canale.[34]

- SCN1B codifica le subunità β1 e β1b, subunità ausiliarie che modificano la funzione del canale del Na^+ cardiaco e diminuiscono la corrente I_{Na} influenzando l'attività del canale del Na^+; [35]
- SCN2B codifica per la subunità β2. La mutazione D211G in SCN2B riduce l'espressione di $Na_v1.5$ sulle cellule CHO;
- SCN3B codifica per la subunità β3 e porta ad una perdita di funzione nel canale cardiaco del Na^+;
- SCN10A, codifica per il canale neuronale del sodio $Na_v1.8$ e modula l'espressione di SCN5A e la funzione elettrica del cuore;
- Il gene RAN del fattore di rilascio del nucleotide della guanina (RANGRF), che codifica per MOG1, compromette il traffico di Nav1.5 verso la membrana, con conseguente riduzione di I_{Na};
- Mutazioni alla glicerina-3-fosfato deidrogenasi 1 (GPD1L) influenzano il traffico di canali cardiaci del Na^+ verso la superficie cellulare e possono causare una riduzione del 50% della corrente del Na^+ verso l'interno;
- Il gene per la proteina associata alla membrana sarcolemmale (SLMAP) che si trova nei tubuli a T e nel reticolo sarcoplasmatico, che ha una funzione sconosciuta, provoca BrS modulando il traffico intracellulare del canale $Na_v1.5$;
- Il gene per la plakofilina-2 (PKP2) è il gene principale responsabile della cardiomiopatia aritmogena ventricolare destra (ARVC);
- TRPM4 è una proteina non selettiva attivata dal canale cationico del calcio;

- CACNA1C e CACNB2b sono associati a casi di BrS, in cui i pazienti generalmente presentano intervalli QT più brevi del normale;
- Mutazioni in geni KCND3, che codificano per un canale voltaggio-dipendente Kv4.3 del potassio, portano ad un aumento della corrente I_{to} nel ventricolo destro;
- KCNE3 (codifica MiRP2) è una proteina che diminuisce il potassio (K^+), corrente transitoria verso l'esterno (I_{to}) interagendo con il canale $K_v4.3$;
- KCNE5 (KCNE1L) si trova sul cromosoma X e codifica per una subunità β ausiliaria del potassio.

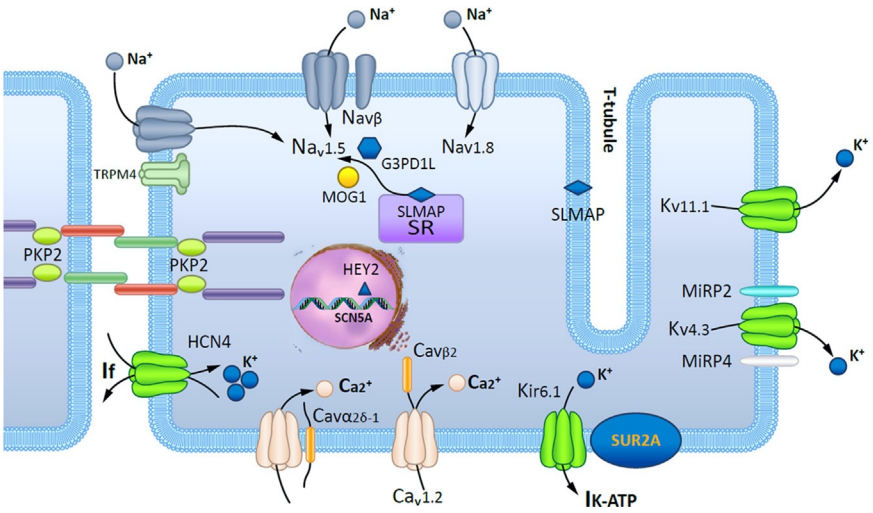


***Figura** 7. Proteine coinvolte nella patogenesi della sindrome di Brugada*

Indicazioni per l'analisi genetica

In accordo con le raccomandazioni dell'associazione HRS/EHRA[36], la diagnosi genetica di **LQTS** dovrebbe essere offerta con le seguenti modalità:

Il test genetico completo per LQTS o mirato per LQTS1-3 (KCNQ1, KCHN2, SCN5A) è raccomandato per qualsiasi paziente in cui un cardiologo ha stabilito un forte sospetto clinico per LQTS basato sull'esame della storia clinica del paziente, della storia familiare e del fenotipo elettrocardiografico (ECG a 12 derivazioni a riposo e/o stress test provocatorio con esercizio o infusione di catecolamine).

Il test genetico LQTS completo o mirato per LQT1-3 (KCNQ1, KCNH2 e SCN5A) è raccomandato per qualsiasi paziente asintomatico con prolungamento del QT in assenza di altre condizioni cliniche che potrebbero prolungare l'intervallo QT (come anomalie elettrolitiche, ipertrofia, fascio blocco di branca, ecc; cioè, altrimenti idiopatico) definito come QTc > 480 ms (nei bambini) o > 500 ms (negli adulti) su ECG seriali a 12 derivazioni.

Il test genetico LQTS completo o mirato a LQT1-3 (KCNQ1, KCNH2 e SCN5A) può essere preso in considerazione per qualsiasi paziente asintomatico con valori QTc altrimenti idiopatici > 460 ms (nei bambini) o > 480 ms (negli adulti) su ECG a 12 derivazioni.

Il test genetico specifico per la mutazione è raccomandato per i membri della famiglia a partire dai parenti di primo grado successivamente all'identificazione della mutazione causativa LQTS in un caso indice (probando).

Il test genetico sul probando ha il 75% di probabilità di identificare una mutazione associata a LQTS. Se il probando ha un test positivo, il genotipo deve essere ricercato anche negli altri membri della famiglia, questo perfezionerà ulteriormente la diagnosi genetica indipendentemente dai sintomi o dall'intervallo QT. Inoltre, pazienti con sincopi inspiegate, o con alterazioni dell'intervallo QT indotte da farmaci, che non rientrano nei criteri diagnostici

per la sindrome, ed hanno ricevuto una diagnosi di "possibile" LQTS, possono beneficiare dei test genetici. In questi casi, un test genetico positivo può aiutare nello spostare l'incertezza verso una diagnosi probabile. Al contrario, un test genetico negativo fornirebbe una misura indipendente per aiutare il medico ad allontanarsi da questa diagnosi, specialmente se il fenotipo clinico è debole.[37]

Le raccomandazioni di consenso degli esperti HRS/EHRA[36] per i test genetici nella **sindrome di Brugada** sono le seguenti:

Test genetici completi o BrS1 (SCN5A) mirati alla BrS possono essere utili per qualsiasi paziente in cui un cardiologo abbia stabilito un sospetto clinico per la BrS basato sull'esame della storia clinica del paziente, dell'anamnesi familiare e del fenotipo elettrocardiografico espresso (ECG a riposo a 12 derivazioni e/o test di provocazione farmacologica).

Test genetici non sono indicati nell'impostazione di un modello isolato di ECG di Brugada di tipo 2 o di tipo 3.

Test genetici specifici per la mutazione sono raccomandati per i membri della famiglia a partire dai parenti di primo grado in seguito all'identificazione della mutazione causale della BrS in un probando.

1.4 Cardiomiopatie

Il termine cardiomiopatia è stato introdotto per la prima volta nel 1957, ma ancora nel 1980 le conoscenze relative a questa particolare classe di malattie erano talmente vaghe che l'Organizzazione Mondiale della Sanità le definiva come "malattie del miocardio ad eziologia sconosciuta"; ancora oggi non esiste una definizione universale di cardiomiopatia. Sebbene vi sia consenso sul fatto che la cardiomiopatia secondaria a coronaropatia aterosclerotica, valvulopatie, cardiopatie congenite e ipertensione sistemica non vada classificata come cardiomiopatia, non vi è consenso sulla classificazione basata sulla morfologia e sull'inclusione delle malattie molecolari come le canalopatie. L'American Heart Association definisce le cardiomiopatie come "un gruppo eterogeneo di malattie del miocardio con disfunzione meccanica e/o elettrica che generalmente (ma non sempre) presenta ipertrofia ventricolare o dilatazione inappropriata e sono causate da vari fattori, spesso genetici. Le cardiomiopatie possono essere limitate al cuore o far parte di una malattia sistemica generalizzata, che spesso causa morte cardiovascolare o disabilità progressiva secondaria a insufficienza cardiaca".[1]

Si tratta di un gruppo diversificato di disturbi cardiaci che spesso causano insufficienza cardiaca e morte e sono ora riconosciuti con frequenza crescente. Vi sono stati progressi sostanziali nel riconoscimento clinico e nella comprensione della storia naturale di queste condizioni: a questo proposito sono utili anche tecniche ben consolidate e nuove di imaging cardiaco.[38]

Le prime valutazioni sulla correlazione tra l'insufficienza cardiaca ed i cambiamenti strutturali del cuore risalgono alla fine del '600 quando Lower scrisse che quando "il parenchima cardiaco soffre di infiammazione, ascesso o per una lesione" questa può renderlo incapace di provvedere "ad una costante circolazione sanguigna".[39] Lower riconobbe inoltre che il cuore è dilatato in

pazienti con insufficienza cardiaca. Da allora, lo studio delle cause di insufficienza cardiaca è stato sempre al centro di ogni ambito inerente alla cardiologia.

Nel XVIII secolo, l'insufficienza cardiaca fu attribuita in gran parte alla cardiopatia valvolare, all'epoca molto comune. Correlando le manifestazioni cliniche con i reperti autoptici, i medici hanno descritto un numero sempre crescente di insufficienze cardiache senza causa apparente.

Nel 1891 Krehl, medico di Lipsia, discusse delle malattie idiopatiche del muscolo cardiaco.

Una svolta avvenne nel 1957, quando Wallace Brigden - un cardiologo di Londra - pubblicò un articolo intitolato "Malattie miocardiche non comuni: le cardiomiopatie non coronariche" e affermò: "il termine cardiomiopatia viene qui usato per indicare una malattia miocardica non coronarica isolata". Brigden ha descritto le varie manifestazioni cliniche di queste condizioni e le loro possibili cause, tra cui forme familiari e processi patologici che hanno comportato infiammazione, necrosi e ipertrofia. Il termine cardiomiopatia fu ampiamente adottato e divenne presto evidente che questi disturbi non erano rari.[40]

La **cardiomiopatia dilatativa** (CMD) è caratterizzata da un ventricolo sinistro dilatato con disfunzione sistolica, non causata da malattia ischemica o valvolare. Prima di classificare la cardiomiopatia come idiopatica (termine che indica l'impossibilità di risalire ad una genesi specifica) si devono escludere le numerose cause genetiche, infatti nel 30-40% dei casi si presenta con caratteristiche di familiarità, e modello di ereditarietà autosomico dominante. Spesso, nei pazienti con CMD, lo sviluppo dei sintomi clinici è preceduto da un periodo di latenza con disfunzione sistolica asintomatica del ventricolo sinistro. I pazienti CMD sono a rischio di aritmie ventricolari e talora l'evento sentinella è la resuscitazione da un arresto cardiaco.[1]

Anche se la cardiomiopatia dilatativa (CMD) familiare è ormai considerata una malattia genetica, non è chiaro se le forme idiopatiche senza evidenze di familiarità abbiano una base genetica. Poiché la malattia ha una penetranza correlata all'età, la CMD è generalmente caratterizzata da una fase di latenza che dura per molti anni prima che si manifesti la fase sintomatica con:

- Insufficienza cardiaca;
- Aritmie;
- Eventi embolici.

A volte una CMD asintomatica ma clinicamente rilevabile viene diagnosticata per caso in occasione di esami clinici preoperatori o routinari, generalmente in seguito a minime anomalie ECGgrafiche che richiedono approfondimenti ecocardiografici. La lunga fase di latenza che precede l'insorgenza della malattia clinica mostra la notevole capacità del miocardio di mantenere per anni la gittata cardiaca e la pressione di riempimento nel range di normalità (o quasi), nonostante la presenza di una CMD asintomatica facilmente diagnosticabile. Questo principio è alla base dell'osservazione che la storia familiare è molto meno sensibile dello screening ecocardiografico nella diagnosi di patologia nei familiari dei soggetti con CMD idiopatica, sottolineando la necessità dello screening clinico dei geni responsabili di CMD familiare, classificati in base alla sede subcellulare (*ontologia genica*). La maggior parte dei geni responsabili codifica per le proteine sarcomeriche, dei dischi Z o del citoscheletro. L'elevato numero di geni diversi coinvolti mostra l'eterogeneità delle possibili vie metaboliche che possono portare al "fenotipo finale" di CMD. Presumibilmente, altri percorsi ancora sconosciuti potrebbero essere rilevanti nella patogenesi della malattia.

La **cardiomiopatia ipertrofica** (CMI), la più comune tra le patologie genetiche cardiovascolari, è causata da diverse mutazioni dei geni che codificano per le proteine del sarcomero cardiaco ed è caratterizzata da segni clinici omogenei e da una fisiopatologia specifica. Nelle forme familiari si tratta di una patologia a trasmissione autosomica dominante; le forme sporadiche possono essere dovute a cause non genetiche, oppure possono essere tali solo in apparenza per difetto di penetranza del gene mutato responsabile o scarsa espressività della malattia in altri membri della famiglia; la sporadicità, inoltre, può essere dovuta ad una mutazione de novo. Rappresenta la più comune causa di morte improvvisa nei giovani, tra cui gli atleti agonisti, soprattutto negli Stati Uniti, e può provocare disabilità da insufficienza cardiaca praticamente a qualsiasi età.

La CMI è caratterizzata da un ventricolo sinistro ispessito ma non dilatato in assenza di un'altra condizione cardiaca o sistemica (ad es. stenosi della valvola aortica, ipertensione sistemica, alcune espressioni fisiologiche del cuore d'atleta), con diminuito volume delle camere ventricolari. All'ipertrofia si associa una rigidità del muscolo che ne impedisce un completo rilassamento in diastole, ostacolando di conseguenza lo svuotamento degli atri che, a causa della sottigliezza delle loro pareti, vanno secondariamente incontro a dilatazione. Vari studi epidemiologici hanno evidenziato la prevalenza del fenotipo della CMI nella popolazione generale di approssimativamente lo 0,2%, per una frequenza di 1 su 500 persone, equivalente approssimativamente a 700.000 persone affette (negli Stati Uniti). Tale stima supera la frequenza relativamente bassa della CMI nella pratica clinica e ciò suggerisce che alla maggior parte delle persone coinvolte può non essere diagnosticata in quanto asintomatici per tutta la vita.[41]

1.4.1 Classificazione

Nel 1995 si distingueva tra le forme classiche ipertrofica e dilatativa e la cardiomiopatia restrittiva primaria. Tuttavia, questa classificazione presentava delle difficoltà legate ai criteri nosografici, basati in parte su caratteri anatomo-istologici e in parte su aspetti funzionali: oggi, invece, si possono utilizzare criteri differenti.

Nel 2006 l'American Heart Association ha fornito una nuova classificazione che include anche le canalopatie e altri difetti di conduzione responsabili delle cardiopatie aritmogene ereditarie[42] e oggi le divide in due gruppi: primarie e secondarie.

Nelle primarie il processo patologico è confinato principalmente al cuore mentre le secondarie descrivono condizioni nelle quali il coinvolgimento cardiaco è parte di una condizione sistemica.[38]

Acquired
Myocarditis
Peripartum
Tachycardia induced
Takotsubo (stress induced)
Genetic
Arrhythmogenic right ventricular dysplasia
Hypertrophic
Ion channel disorders
Left ventricular compaction
Mitochondrial myopathies
Mixed
Dilated
Restrictive
Information from reference 1.

***Figura 8**. Classificazione delle cardiomiopatie primarie*

Le cardiomiopatie primarie possono essere a loro volta suddivise in forme genetiche, acquisite, o miste in base all'eziologia, come mostrato in Figura 8.

Le forme genetiche sono causate da anomalie cromosomiche; le acquisite, da non confondere con le forme secondarie, sono cardiomiopatie su base non genetica che portano principalmente, o esclusivamente, a complicanze cardiache. Nelle forme miste, il fenotipo è dato comunemente da forme genetiche e acquisite insieme.[43]

Un secondo criterio di classificazione tiene conto, invece, delle modifiche funzionali e strutturali a carico del cuore, dividendo pertanto la patologia in 3 gruppi: cardiomiopatie dilatative, ipertrofiche e restrittive. Una quarta categoria, la displasia aritmogena del ventricolo destro, è stata aggiunta di recente.[44]

Tuttavia, nella maggioranza dei casi, ad oggi non è possibile ricondurre la cardiomiopatia ad una patogenesi certa, e vengono definite come cardiomiopatie idiopatiche.[40]

La cardiomiopatia ipertrofica, inclusa tra le forme primarie, è la più comune cardiomiopatia e può causare dispnea, lipotimie, dolore toracico atipico, insufficienza cardiaca, fino alla morte cardiaca improvvisa (MCI).

La cardiomiopatia dilatativa può essere genetica o acquisita e tipicamente si presenta con i classici sintomi dell'insufficienza cardiaca con frazione di eiezione ridotta (< 40%).[45]

La cardiomiopatia restrittiva è molto meno comune e spesso associata con patologie sistemiche.

Esistono, inoltre, varianti come la forma peripartum e la cardiomiopatia da stress.

1.4.2 Diagnosi

Il percorso diagnostico dei pazienti con **cardiomiopatia dilatativa** deve comprendere l'anamnesi completa, compresi i fattori di rischio per la coronaropatia: se l'anamnesi non viene raccolta con cura si può sottovalutare la durata dei sintomi. Si può manifestare angina, anche in assenza di coronaropatia epicardica, ma i sintomi anginosi devono indurre il sospetto diagnostico di coronaropatia. L'anamnesi familiare è essenziale, non solo per ricercare i sintomi suggestivi di insufficienza cardiaca, ma anche di morte improvvisa, che può essere riferita dal paziente come "morte per grave attacco cardiaco".

Il sospetto diagnostico di **cardiomiopatia ipertrofica** viene evocato dall'auscultazione di un soffio cardiaco mesotelesistolico e dalla sintomatologia. Il sospetto è aumentato se il paziente

ha una storia di sincope inspiegata o una storia familiare di morte improvvisa, la sincope inspiegabile nei giovani atleti, infatti, deve sempre destare il sospetto. La cardiomiopatia ipertrofica deve essere distinta dalla stenosi aortica e dalla coronaropatia, che producono una sintomatologia simile.

Diagnosi clinica

Nella **cardiomiopatia dilatativa** l'esame clinico mostra:

- Disfunzione biventricolare;
- Il 25% dei pazienti presenta dolore toracico;
- La disfunzione ventricolare sinistra si manifesta con dispnea e astenia da sforzo, dovute all'aumento della pressione diastolica del ventricolo sinistro e alla riduzione della gittata cardiaca;
- L'insufficienza del ventricolo destro invece si manifesta con edema periferico e distensione delle vene del collo. Di rado, nei pazienti più giovani, viene colpito prevalentemente il ventricolo destro; tale forma può essere spesso complicata da aritmie sopraventricolari e morte improvvisa dovuta a tachiaritmie ventricolari maligne.

Molti pazienti con cardiomiopatia ipertrofica sono asintomatici e vengono diagnosticati:

- Durante lo screening familiare;
- Mediante auscultazione di un soffio;
- Incidentalmente dopo un risultato anormale all'elettrocardiografia.

I segni e i sintomi più caratteristici della cardiomiopatia ipertrofica (HCM) includono:

- Dolore toracico atipico (che può essere associato a pasti, disidratazione o sforzo);
- Morte cardiaca improvvisa;[46,47]

- All'ispezione sono presenti itto presistolico intenso o doppio itto, espressione di una vigorosa contrazione ventricolare;
- All'auscultazione può essere presente un soffio mesotelesistolico al mesocardio, espressione dell'ostruzione dinamica.

I pazienti a cui viene diagnosticata la HCM possono avere una storia familiare di morte cardiaca improvvisa inspiegabile. All'esame, il medico può sentire un soffio sistolico che aumenta di intensità durante le manovre di Valsalva.

Diagnosi strumentale

L'**ecocardiogramma** è l'esame principale con cui è possibile valutare:

- Dimensioni del cuore: sia i volumi, sia i diametri in mono ed in bi-dimensionale;
- Frazione di eiezione: distingue forme a FE ridotta (<40%) e conservata (>50%);
- Insufficienze valvolari;
- Funzione diastolica;
- Pressioni ventricolari e la pressione polmonare (stimate).

A questo fanno poi seguito altre valutazioni, come l'elettrocardiogramma, la radiografia del torace, la scintigrafia miocardica e la risonanza magnetica cardiaca.

Cardiomiopatia ipertrofica

Elettrocardiogramma:

- Ipertrofia del ventricolo sinistro (IVS);
- Modifiche aspecifiche del tratto ST-T o blocco di branca;
- Inversione profonda e simmetrica dell'onda T nelle derivazioni I, aVL, V5 e V6;
- Possono essere presenti onde Q patologiche, ma la loro presenza aumenta la probabilità di cardiopatia aterosclerotica avanzata piuttosto che di cardiomiopatia primitiva;

- Nelle forme avanzate con fibrosi estesa, i tracciati degli elettrodi periferici possono mostrare bassi voltaggi.

Ecocardiogramma doppler:

- Distingue i pattern anatomici di cardiomiopatia ipertrofica (Figura 9);
- Quantifica la gravità dell'ipertrofia ed il grado di ostruzione del tratto di efflusso.

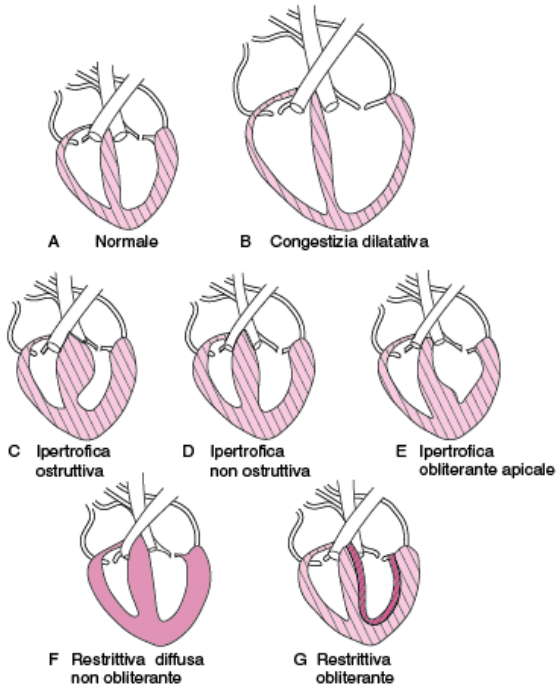


Figura 9. *Pattern cardiomiopatia ipertrofica*

A queste indagini si possono poi aggiungere ulteriori approfondimenti da riservare a casi particolari. Per esempio, l'ecocardiografia transesofagea dovrebbe essere considerata in caso di immagini ecocardiografiche transtoraciche non dirimenti.

L'utilizzo della biopsia miocardica, invece, viene impiegato sempre meno nella pratica clinica, ormai completamente sostituito da indagini non invasive.

Cardiomiopatia dilatativa

Ecocardiogramma:

- Variabile dilatazione biventricolare, a carico nella maggior parte dei casi del ventricolo di sinistra;
- Disfunzione sistolica del ventricolo sinistro, da lieve a grave;
- Spessore della parete del ventricolo sinistro è generalmente normale, ma la sua massa è quasi sempre aumentata;
- Ipocinesia globale del ventricolo sinistro, ma si possono osservare anche discinesie segmentarie, soprattutto discinesia settale nei soggetti con blocco di branca sinistra.

Le insufficienze mitralica e tricuspidale sono spesso presenti e possono essere gravi, anche in assenza di soffi cardiaci intensi all'esame obiettivo.

La funzione diastolica varia da normale a restrittiva; le forme restrittive sono più frequenti in presenza di insufficienza cardiaca "scompensata" con sovraccarico di volume e spesso migliorano dopo terapia con diuretici o vasodilatatori.

Rx torace:

Cardiomegalia, di solito di tutte le camere.

L'aumento della pressione venosa polmonare e l'edema interstiziale, inoltre, sono spesso accompagnati da versamento pleurico, soprattutto a destra.

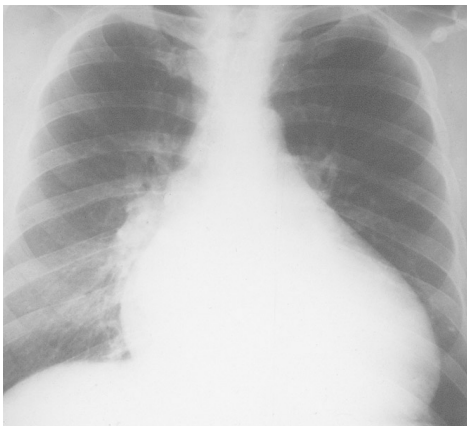


***Figura 10**. RX torace cardiomegalia in cardiomiopatia dilatativa*

La **coronarografia** è indicata in tutti i pazienti con fattori di rischio per coronaropatia o in una classe di età a rischio. In alternativa si può utilizzare l'angio-TC, che però non consente lo studio emodinamico che può essere utile in alcuni pazienti. Poiché la coronaropatia è comune, il significato funzionale di eventuali lesioni coronariche ostruttive deve essere valutato attentamente nella misura in cui la loro presenza può essere concomitante con la cardiomiopatia dilatativa.



Figura 11. *Risonanza magnetica cardiomiopatia*

La **risonanza magnetica cardiaca** può essere utile nella valutazione delle cardiomiopatie: l'enhancement tardivo non transmurale con gadolinio, con distribuzione non coronarica, in un ventricolo sinistro dilatato suggerisce una causa non ischemica. La RMN cardiaca consente di valutare il grado di fibrosi miocardica nella cardiomiopatia dilatativa.

Diagnosi genetica

La combinazione di dati genetici e fenotipici relativi alla struttura ed alla funzione del ventricolo sinistro (VS) è alla base della medicina genetica cardiovascolare. I test genetici molecolari per le cardiomiopatie non solo migliorano la cura delle forme sintomatiche, ma sono utili anche ai pazienti asintomatici e ai familiari per la valutazione del rischio. Inoltre, è probabile che in futuro i dati genomici consentiranno di predire la storia naturale della malattia e di guidare la terapia.

Le linee guida per la valutazione e test genetici clinici per la cardiomiopatia, applicabili a tutte le cardiomiopatie con possibile causa genetica, comprendono un'anamnesi familiare accurata, nell'arco di 3-4 generazioni, per ricercare evidenze di cardiomiopatia, distrofia muscolare o altri segni di malattie sindromiche comprensive di cardiomiopatia. Tuttavia, l'anamnesi familiare potrebbe essere negativa perché la cardiomiopatia può essere asintomatica nei familiari. Di conseguenza, è essenziale lo screening clinico di tutti i parenti di primo grado, con almeno: anamnesi, esame obiettivo, ECG, ecocardiografia. Se si trovano segni di patologia in un familiare, è indicato lo screening dei familiari di primo grado di quest'ultimo (*screening clinico graduale o a cascata*). Nell'ambito del counseling, i test genetici sono indicati in presenza di qualsiasi evidenza di malattia familiare, poiché l'identificazione di una mutazione patogena (in uno o più membri della famiglia con malattia evidente) consente di effettuare i test di genetica molecolare sugli altri soggetti a rischio, ma con malattia preclinica, contribuendo alla stratificazione dello stesso. Infatti, quelli che risultano negativi per la mutazione dovrebbero avere un rischio di cardiomiopatia significativamente ridotto, mentre quelli portatori della mutazione familiare dovrebbero sottoporsi a screening clinico intensivo per la diagnosi precoce poiché la terapia, se effettuata sin dalle prime fasi (ACE-inibitori o β-bloccanti), può ritardare o prevenire la progressione della malattia.

I test genetici si avvalgono oggi del sequenziamento di ultima generazione che comprende panel di 20-30 o più geni associati a cardiomiopatia e devono sempre essere condotti nell'ambito del counseling genetico, con l'obiettivo di rivedere le modalità di trasmissione ereditaria, gli aspetti clinicamente rilevanti delle forme idiopatica e familiare e garantire che l'anamnesi familiare sia stata raccolta in modo esaustivo e interpretata correttamente, identificando i parenti a rischio.

I geni della cardiomiopatia dilatativa

A fronte di una presentazione clinica relativamente uniforme, sono stati identificati oltre 30 geni che possono causare cardiomiopatia dilatativa (*eterogeneità dei loci*). La diversa localizzazione subcellulare di geni implicati nella CMD differenzia questa forma di cardiomiopatia dalla cardiomiopatia ipertrofica (HCM) e dalla cardiomiopatia aritmogena (ARCV), causate da varianti dei geni che codificano, rispettivamente, per le proteine sarcomeriche o desmosomiali. Recentemente, è stato suggerito che il 25% dei casi di CMD sia causato da varianti tronche della titina (TTN), una proteina gigante che forma impalcature cellulari. Tuttavia, anche una percentuale di controlli fino al 3% presenta varianti tronche di TTN e gli studi genome-wide che hanno dimostrato l'impatto e la rilevanza delle varianti tronche di TTN hanno anche varianti tronche non associate a CMD.

Si assiste ad una eterogeneità genetica e fenotipica attribuibile a diverse mutazioni che, pur a carico dello stesso gene, hanno però effetti fenotipici differenti.

Sono stati proposti diversi meccanismi implicati nella patogenesi della dilatazione del muscolo cardiaco, non necessariamente esclusivi l'un l'altro. Un'efficace contrazione muscolare richiede che il sarcomero generi una forza che deve essere trasmessa al citoscheletro e alla matrice extracellulare. Le mutazioni che causano cardiomiopatia dilatativa possono danneggiare l'uno o l'altro di questi meccanismi oppure possono agire attraverso una terza via, provocando danno e morte cellulare, e conseguente rimodellamento ventricolare. Questo rimodellamento comporta anche un certo grado di ipertrofia miocardica compensatoria, così come nei casi di cardiomiopatia ipertrofica si può osservare, in fase avanzata di malattia, dilatazione ventricolare.

Mutazioni sono state identificate in geni codificanti:

- Cysteine- and Glycine-rich protein 3 (CSRP3);

- Actina α2 (ACTN2);
- Desmina (DES);
- Metavinculina;
- Nexilina.

In ciascun caso sono state riscontrate delle anomalie istologiche peculiari a livello della fibra miocardica, sia nell'uomo sia in vari modelli animali. Alcune di queste proteine, inoltre, pur essendo espresse sia nel muscolo scheletrico sia nel miocardio, sembrano determinare alterazioni solo a carico di quest'ultimo: un fenomeno definito "cardio-tossicità selettiva".

LOCUS	GENE	CROMOSOMA	PROTEINA
CMD1A	LMNA	1q21.2	Lamina A/C
CMD1B		9q13	
CMD1C	LDB3	10q22-q23	LIM domain-binding 3
CMD1D	TNNT2	1q32	Troponina I cardiaca
CMD1E	SCN5A	3p21	Canale del sodio, tipo V, Subunità α
CMD1F		6q23	
CMD1G	TTN	2q24.3	Titina
CMD1H		2q14-q22	
CMD1I	DES	2q35	Desmina
CMD1J	EYA4	6q23-q24	Omologo di eyes absent nella Drosophila
CMD1K		6q12-q16	
CMD1L	SGCD	5q33	δ-sarcoglicano
CMD1M	CSRP3	11p15.1	Cystein- and Glycine-Rich Protein 3
CMD1N	TCAP	17q12	Titin-Cap (Telethonina)
CMD1O	ABCC9	12p12.1	ATP-binding cassette, subfamily C, n.9
CMD1P	PLN	6q22.1	Fosfolambano

CMD1Q		7q22.3-q31.1	
CMD1R	ACTC1	15q14	α-actina cardiaca
CMD1S	MYH7	14q12	Catena pesante β della miosina
CMD1T	TMPO	12q22	Timopoietina
CMD1U	PSEN1	14q24.3	Presenilina 1
CMD1V	PSEN2	1q31-q42	Presenilina 2
CMD1W	VCL	10q22-q23	Metavinculina
CMD1X	FCMD	9q31	Fukutina
CMD1Y	TPM1	15q22.1	Tropomiosina 1
CMD1Z	TNNC1	3p21.3-p14.3	Troponina C, isoforma cardiaca (o lenta)
CMD1AA	ACTN2	1q42-q43	Actina α2
CMD1BB	DSG2	18q12.1-q12.2	Desmogleina 2
CMD1CC	NEXN	1p32-p31	Nexilina (o nelina)
CMD1DD	RBM20	10q25.2	RNA-binding motif protein 20
CMD12A	TNNI3	19q13.4	Troponina I cardiaca
	MYBPC3	11p11.2	Myosin-binding protein C, isoforma cardiaca

***Tabella** 7. Mutazioni identificate per la cardiomiopatia dilatativa*

I geni della cardiomiopatia ipertrofica

Nella forma ipertrofica la malattia può essere causata da almeno 10 differenti geni codificanti altrettante proteine sarcomeriche, le cui funzioni sono strettamente interdipendenti.

- Il primo di questi geni, codificante la catena pesante β della miosina (MYH7), venne identificato dopo che l'analisi di linkage in una famiglia franco-canadese aveva assegnato il corrispondente locus alla regione cromosomica 14q2;
- Arg403Gln e Arg719Trp predispongono allo scompenso cardiaco ed alla morte improvvisa;
- Phe513Cys, Leu908Val e Gly2576Glu causano manifestazioni cliniche meno gravi.

Si tratta in tutti i casi di mutazioni missenso, da cui si può dedurre che esse conferiscano alla proteina proprietà alterate, con effetto dominante negativo.

Le mutazioni più frequentemente riscontrate sono quelle a carico di MYH7 e MYBCP3, ma è verosimile che ancora non siano stati identificati tutti i geni associati alla patologia.

LOCUS	GENE	CROMOSOMA	PROTEINA
CMH1	MYH7	14q12	Catena pesante β della miosina
CMH2	TNNT2	1q32	Troponina T cardiaca
CMH3	TPM1	15q22.1	α tropomiosina
CMH4	MYBPC3	11p11.2	Proteina C legante la miosina
CMH6		7q21-qter	
CMH7	TNNI3	19q13.4	Troponina I cardiaca
CMH8	MYL3	3p	Catena leggera della miosina, di tipo 3
CMH9	TTN	2q24.3	Titina
CMH10	MYL2	12q23	Catena leggera della miosina, di tipo 2
CMH11	ACT1	15q14	α-actina cardiaca
CMH12	CSRP3	11p15.1	Cysteine- e Glycine-Rich Proteine 3

***Tabella 8**. Mutazioni identificate per la cardiomiopatia ipertrofica*

Indicazioni per l'analisi genetica

Il discorso inerente alla consulenza genetica è univoco per la cardiomiopatia ipertrofica e per la cardiomiopatia dilatativa.

I casi familiari sono relativamente semplici, in quanto dalla posizione degli affetti nell'albero genealogico di solito non è difficile dedurre il modello di ereditarietà, che nella maggior parte dei casi è autosomico dominante. Una volta stabilito ciò, è facile calcolare i rischi di ricorrenza della malattia ed è anche ragionevole tentare di individuare il gene mutato causale. Data la molteplicità dei loci interessati, sia nelle forme ipertrofiche sia in quelle dilatative, può essere indicato procedere preliminarmente ad uno studio di linkage, per stabilire quale locus segreghi effettivamente con la malattia, dopo di che si potrà procedere ad un'analisi mutazionale del gene corrispondente al locus individuato. Naturalmente, nell'effettuare l'analisi di linkage sarà molto importante tenere conto di possibili casi con difetto di penetranza. Pertanto, sarà necessario procedere ad alcuni approfondimenti clinici (per esempio un'ecocardiografia) in quei soggetti, anche apparentemente sani, che sono tuttavia potenziali portatori della mutazione data la loro posizione nell'albero genealogico.

Se tutto ciò è teoricamente fattibile, ne va tuttavia soppesata l'utilità clinica. È infatti vero che l'individuazione del gene può, almeno in alcuni casi, fornire delle indicazioni prognostiche circa la gravità della malattia, ma purtroppo non fornisce spunti per una terapia specifica, salvo casi particolari.

Nella pratica della genetica clinica, la conoscenza di una mutazione causa di malattia viene di solito utilizzata nel contesto della diagnosi prenatale ma nel caso delle cardiomiopatie ipertrofiche e dilatative, il ricorso alla diagnosi prenatale è molto discutibile, se non improponibile, dato l'esordio generalmente tardivo della malattia e la perfetta normalità dello sviluppo generale del soggetto.

Ben diversa è la situazione quando si presenta un paziente sporadico, senza storia familiare positiva. In questo caso l'individuazione di un gene responsabile della malattia sarebbe importante per confermare che si tratta effettivamente di un caso di malattia su base genetica e non di una fenocopia. L'impresa appare però proibitiva, dato l'elevato numero di geni potenzialmente coinvolti, e fra l'altro andrebbe affrontata senza la certezza che porti ad un risultato chiaro e definitivo. Converrà allora prima di tutto verificare che il caso sia realmente sporadico, escludendo, su base ecocardiografica, che non vi siano nella famiglia altri affetti, ancora soggettivamente asintomatici. Accertata l'effettiva sporadicità del caso, si potrà procedere alla formulazione di una stima empirica del rischio di trasmissione della malattia dal soggetto affetto ai suoi figli, tenendo conto del fatto che non tutti i casi sono su base genetica e che, quando lo sono, l'ereditarietà è quasi sempre di tipo autosomico dominante.[42]

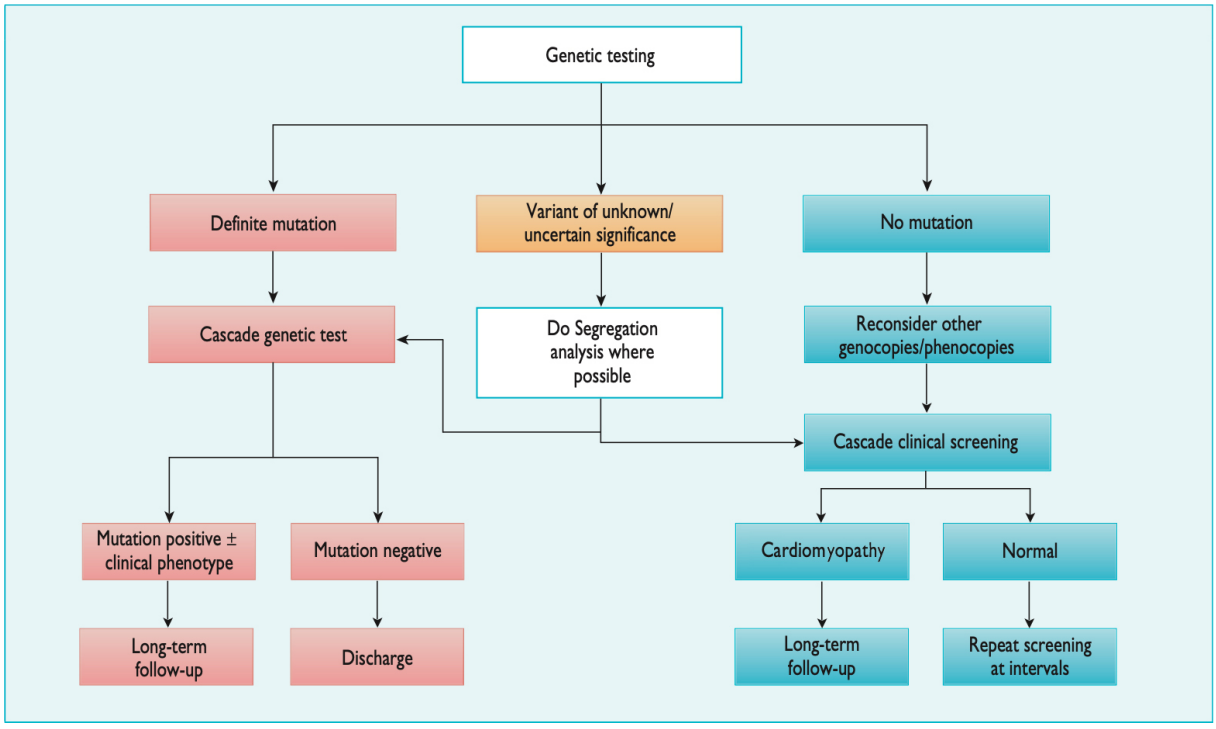


***Figura 12**. Iter test genetico*

CAPITOLO 2: RUOLO DELLA GENETICA

2.1 La struttura del gene

Dal punto di vista della genetica molecolare per gene si intende una sequenza di DNA potenzialmente trascrivibile in RNA funzionalmente attivo. Tale RNA può svolgere direttamente una funzione strutturale e/o catalitica oppure trasportare l'informazione per la sintesi di una proteina. Nel genoma umano si stima che siano presenti circa *23.000 geni codificanti proteine e 1.000-2.000 geni codificanti RNA strutturali*. Da recenti studi, inoltre, emerge l'esistenza di diverse migliaia di trascritti non codificanti che potrebbero non avere alcuna funzione o, viceversa, svolgere un ruolo fondamentale nella regolazione della conformazione della cromatina e della trascrizione dei geni codificanti proteine.

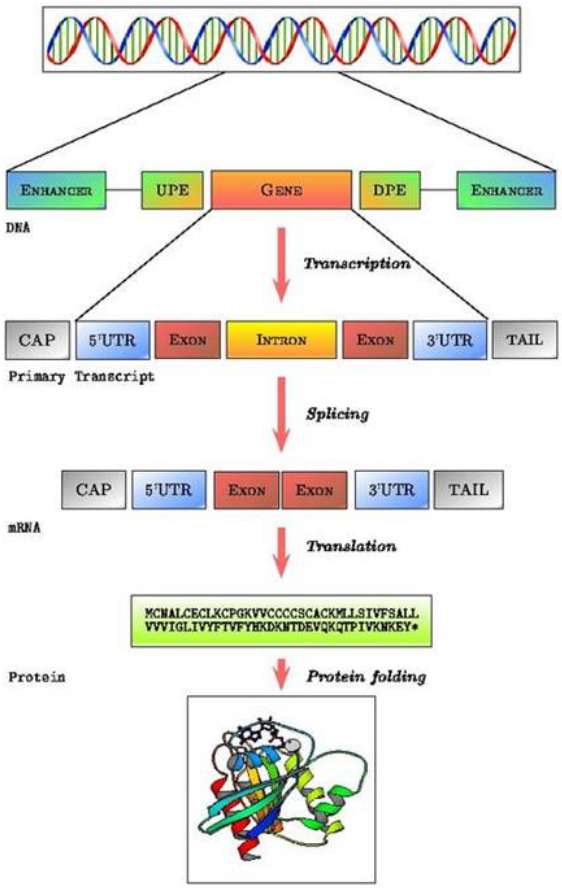


***Figura 13**. Struttura del gene*

I geni codificanti sono quelli che vengono trascritti in mRNA. Contengono una parte realmente codificante, che specifica la sequenza degli aminoacidi che costituiranno la proteina, ed una parte non codificante. A monte della sequenza che verrà trascritta in mRNA ci sono le *sequenze regolatrici*. La sequenza trascritta è costituita da due tipi di elementi, detti *esoni* ed *introni*. Solo gli esoni contengono informazioni per la sintesi della proteina, il primo

dei quali comincia al sito d'inizio della trascrizione, ma è un tratto non codificante.[42]

Gli esoni presenti all'estremità 5' e 3' del gene, non codificanti, prendono il nome di UTR (*UnTraslated region*).

La regione a monte del sito di inizio della trascrizione è detta *promotore*. In questa regione, di lunghezza variabile, si trova una serie di brevi sequenze (*elementi cis*) che viene riconosciuta e legata da fattori di trascrizione (*elementi trans*). I fattori di trascrizione favoriscono il legame dell'RNA polimerasi al sito giusto per iniziare la sintesi di RNA. I geni che presentano elevati livelli di trascrizione, come gli istoni o la beta-globina, hanno elementi del promotore che includono sempre un TATAbox o una sua variante. I promotori di molti altri geni, tra cui i cosiddetti geni housekeeping, ossia geni necessari per il funzionamento generale della cellula, non hanno TATAbox, ma presentano spesso altri elementi come i GCbox. Vi sono altre sequenze specifiche che vengono riconosciute da fattori di trascrizione quali gli *elementi di risposta,* localizzati nel promotore o nella regione 5' del gene, e gli elementi indicati come *enhancer*, che servono per aumentare i livelli basali della trascrizione e sono localizzati a distanza variabile dal gene, talvolta anche a valle del sito d'inizio della trascrizione, vale a dire all'interno della regione trascritta.

Tramite il *processo di splicing*, poi, gli introni, cioè le sequenze non tradotte, verranno eliminati dai trascritti primari.

2.2 Le mutazioni geniche

Il genoma umano, al pari di quello degli altri esseri viventi, è un'entità dinamica che nel corso dell'evoluzione è andata, e va tuttora, incontro a variazioni (*mutazioni*) della sequenza nucleotidica. Una mutazione consiste in un cambiamento permanente della sequenza normale del DNA (comunemente detta selvatica o *wild type*).

La costituzione genetica di un individuo viene definita **genotipo**, mentre le manifestazioni osservabili derivanti dall'interazione del genotipo con l'ambiente vengono definite **fenotipo**. I cambiamenti della sequenza del DNA possono essere privi di effetti fenotipici: in tal caso si tratta di varianti normali che vengono dette **polimorfismi** se si presentano nella popolazione in esame con una frequenza superiore all'1%; se una variante nucleotidica normale ha frequenza inferiore all'1% e viene riscontrata per esempio soltanto in una famiglia o in un piccolo numero di famiglie tra di loro correlate, viene classificata come **variante privata**.

Al contrario, molte mutazioni possono influenzare la funzione o la quantità del prodotto del gene e, quindi, avere un **effetto fenotipico**. Lo studio delle mutazioni e dei loro effetti è oggetto della patologia molecolare. I progressi esponenziali compiuti negli ultimi anni nel campo della genetica molecolare hanno consentito l'identificazione di centinaia di cause genetiche di malattia e una maggiore conoscenza dei meccanismi patogenetici che determinano l'insorgenza di un dato fenotipo. Oggi è possibile formulare la diagnosi molecolare pre- e post-natale per molte condizioni geneticamente determinate e fornire alle famiglie una più accurata consulenza genetica. Accanto alla definizione della diagnosi, l'altro obiettivo fondamentale della patologia molecolare è quello di stabilire le correlazioni genotipo-fenotipo di una determinata condizione, vale a dire prevedere le conseguenze fenotipiche di una specifica mutazione in un dato gene. Nella maggioranza delle malattie geneticamente determinate è tuttavia molto difficile stabilire correlazioni genotipo-fenotipo dato che appare ormai assodato che il fenotipo non dipenda esclusivamente dall'effetto di una mutazione, ma è piuttosto la conseguenza di una lunga cascata di eventi molecolari che coinvolgono un numero variabile di geni e di fattori ambientali. Le variazioni della sequenza di DNA possono essere classificate in diversi modi. Un primo criterio che si può adottare è quello dell'ampiezza del difetto genetico e della modalità di insorgenza. In particolare, vi sono mutazioni piccole e semplici che coinvolgono soltanto uno

o pochi nucleotidi e che sono comunemente indicate come **mutazioni puntiformi**. Accanto a queste, vi sono mutazioni complesse che originano da meccanismi differenti: tra esse le cosiddette **mutazioni dinamiche** ed i **riarrangiamenti genici strutturali**, che sono dovuti a fenomeni di vario tipo. Una classificazione alternativa delle mutazioni si basa, più che sul meccanismo mutazionale che le determina, sulla loro possibile o presunta patogenicità.

Gli effetti fenotipici delle mutazioni possono essere molteplici, tuttavia a grandi linee possono essere distinti nelle seguenti categorie:

- Mutazioni con perdita di funzione (*loss of function*);
- Mutazioni con guadagno di funzione (*gain of function*).

Le mutazioni che producono la perdita, o almeno una riduzione, della funzione di una proteina sono di solito responsabili di fenotipi recessivi. In questi casi l'effetto patogenetico della mutazione si manifesta solo quando entrambi gli alleli in un determinato locus sono mutati. Se uno solo dei due alleli presenta la mutazione, quello normale permette la sintesi di una quantità sufficiente di prodotto genico e l'espressione di un fenotipo normale. Causa perdita di funzione la gran parte delle mutazioni di geni responsabili di errori congeniti del metabolismo, generalmente geni che codificano enzimi che catalizzano varie tappe delle vie metaboliche.

Più raramente la perdita di funzione può determinare un fenotipo dominante: se il livello di proteina prodotto (circa 50%) è insufficiente a compensare la perdita dell'altro, si ha l'espressione di un fenotipo patologico con un fenomeno definito **aploinsufficienza**.

Si ha un effetto dominante anche quando il prodotto di un allele mutato non solo perde la propria funzione ma interferisce con quella dell'allele normale: si parla in questo caso di **effetto dominante negativo**. Appartengono a questa categoria mutazioni in geni che codificano proteine che formano dimeri o multimeri. Una mutazione con effetto dominante negativo può essere responsabile di un fenotipo grave, poiché una subunità alterata può formare dimeri o

oligomeri con subunità normali, impedendone il corretto funzionamento e determinando la perdita di funzione dell'intero complesso proteico.

Le mutazioni che determinano acquisto di funzione (gain of function) producono in genere un fenotipo dominante; i meccanismi che portano alla formazione di prodotti genici con funzioni nuove o alterate possono essere diversi.

Talvolta, le conseguenze di una mutazione possono essere limitate ad un solo organo o ad un solo tipo cellulare; altre volte le mutazioni possono avere effetti differenti a livello di diversi organi o tessuti (*effetto pleiotropico*). In effetti, uno stesso gene può essere espresso in numerosi processi biochimici che riguardano vari aspetti della funzione cellulare; ciò spiega perché una singola mutazione puntiforme a carico di un gene con effetto pleiotropico possa causare un ampio spettro di manifestazioni cliniche che coinvolgono più organi e apparati, definendo una sindrome.[42]

Una volta identificata una variante come mutazione andrà classificata in modo standardizzato per evitare designazioni ambigue e permettere una condivisione il più ampia possibile dei risultati. I referti clinici dovrebbero includere il riferimento alla sequenza per garantire una chiara denominazione della variante a livello del DNA, così come per garantire una corretta interpretazione funzionale.

Inoltre, i laboratori dovrebbero valutare l'impatto clinicamente rilevabile della variante, includendo trascrizioni alternative che contengano esoni aggiuntivi o estese regioni non trascritte, quando ci sono varianti note in queste regioni che sono clinicamente interpretabili.

2.3 La genetica applicata alle patologie cardiache

Notevoli progressi nel campo della genetica molecolare hanno fornito importanti strumenti per chiarire i substrati genetici di molte malattie genetiche che seguono i modelli ereditari mendeliani. Le basi genetiche delle cardiomiopatie ereditarie e potenzialmente letali e delle canalopatie cardiache, tra cui la cardiomiopatia ipertrofica (HCM), la cardiomiopatia dilatativa (DCM), la cardiomiopatia aritmogenica del ventricolo destro, la sindrome del QT lungo (LQTS), la tachicardia catecolaminergica polimorfica ventricolare (CPVT) e la sindrome di Brugada (BrS) sono state identificate e sono ora meglio comprese.

L'eterogeneità clinica e genetica è una caratteristica cardine di questi disturbi, con più geni alla base dei loro meccanismi patogeni. I test genetici per molte di queste cardiomiopatie ereditarie e canalopatie sono passati dall'essere utilizzati per il solo scopo di ricerca scientifica ad avere una vera e propria utilità clinica e sono ora disponibili in commercio come test genetici clinici.[37]

Anche se i test genetici per le patologie cardiache si stanno evolvendo rapidamente e sono disponibili in commercio, l'interpretazione dei risultati dei test è estremamente impegnativa a causa di un impatto clinico non chiaro, la necessità di effettuare test di conferma sui parenti di primo grado, rendimenti inferiori e tassi di rumore di fondo relativamente alti. Per questo tipo di patologie, però, l'utilità clinica dei test genetici è con grande probabilità attualmente la più elevata.[48,49]

Data la complessità delle informazioni genetiche che si ottengono con questi test, un elemento di fondamentale importanza è rappresentato dall'anamnesi familiare multigenerazionale completa e dalle informazioni sui sintomi.

L'emergere di piattaforme di sequenziamento di nuova generazione che hanno sostanzialmente diminuito il costo della genotipizzazione ha ora fornito nuovi strumenti per analizzare in modo efficiente l'intero genoma o l'esoma di un individuo (l'intera regione di codifica degli

aminoacidi del genoma) in una singola reazione. Questa tecnologia fornisce efficacemente una lista di ogni singola sostituzione nucleotidica e di ogni piccola inserzione/cancellazione (comune o rara, benigna o patogena) per ogni gene del genoma di un paziente ed ha permesso di identificare nuovi geni per diversi disturbi mendeliani.

L'approccio con test genetici può includere l'utilizzo di un pannello multigenico, un test per il singolo gene o test genomici più completi:

- **Pannello multigenico**. Include i geni appartenenti alla categoria di quelli riconosciuti fino ad oggi. Il pannello di geni e la sensibilità del test variano da laboratorio a laboratorio e nel corso del tempo. I medici devono determinare quale pannello multigenico è più probabile che identifichi la causa genetica della condizione al costo più ragionevole, limitando l'identificazione di varianti di significato incerto e varianti patogene in geni che non spiegano il fenotipo sottostante. In alcuni laboratori, le opzioni del pannello possono includere un pannello personalizzato progettato dal laboratorio e/o un'analisi dell'esoma incentrata sul fenotipo personalizzata che include geni specificati dal medico. I metodi utilizzati in un pannello possono includere analisi di sequenza, analisi di eliminazione/duplicazione e/o altri test non basati sul sequenziamento.
- **Test genetico singolo**. È possibile eseguire test genetici molecolari basati sul fenotipo dell'individuo (pattern dell'onda T e fattori scatenanti della sincope), che hanno dimostrato di predire il genotipo.[50] L'analisi della sequenza viene eseguita per prima; può essere seguita da un'analisi di delezione/duplicazione mirata al gene se non viene trovata alcuna variante patogena.

- **Test genomici più completi,** compreso il sequenziamento dell'esoma e il sequenziamento del genoma. Tali test possono fornire o suggerire una diagnosi non considerata in precedenza (ad esempio, mutazione di un gene o di geni diversi che si traduce in una presentazione clinica simile).



Figura 14. *Analisi genetica*

La recente commercializzazione di test diagnostici genetici a breve termine e la crescente disponibilità di test, stanno agevolando l'utilizzo di test genetici come strumento clinico. Tuttavia, prima dell'uso diffuso, la loro validità clinica deve essere stabilita più ampiamente.

Le varianti

Le mutazioni che coinvolgono i geni descritti sopra rivestono significati diversi in base alle modifiche che apportano a livello prima molecolare, poi proteico e, infine, funzionale. Per poterle classificare sono stati stabiliti due set di criteri: uno per identificare le mutazioni come patogenetiche o probabilmente patogenetiche (Tabella 10) e un altro per stabilire la benignità (Tabella 9), certa o probabile, di tali varianti. Ogni criterio patogenetico viene ulteriormente suddiviso in "molto forte", "forte", "moderato", o "di supporto", e ogni criterio di benignità, a sua volta, può assumere il significato di "stand-alone", "forte" o "di supporto". Per ogni variante che viene ritrovata si assegnano i criteri corrispondenti e, successivamente, si verifica lo stato della mutazione combinandoli (Tabelle 11).

Se una variante non rientra nei criteri benigni o maligni, o se i criteri assegnati sono in contrasto tra loro, ci troviamo di fronte a una variante di incerto significato (VUS).

Evidence of benign impact	Category
Stand-alone	BA1 Allele frequency is >5% in Exome Sequencing Project, 1000 Genomes Project, or Exome Aggregation Consortium
Strong	BS1 Allele frequency is greater than expected for disorder (see **Table 6**)
	BS2 Observed in a healthy adult individual for a recessive (homozygous), dominant (heterozygous), or X-linked (hemizygous) disorder, with full penetrance expected at an early age
	BS3 Well-established in vitro or in vivo functional studies show no damaging effect on protein function or splicing
	BS4 Lack of segregation in affected members of a family
	Caveat: The presence of phenocopies for common phenotypes (i.e., cancer, epilepsy) can mimic lack of segregation among affected individuals. Also, families may have more than one pathogenic variant contributing to an autosomal dominant disorder, further confounding an apparent lack of segregation.
Supporting	BP1 Missense variant in a gene for which primarily truncating variants are known to cause disease
	BP2 Observed in *trans* with a pathogenic variant for a fully penetrant dominant gene/disorder or observed in *cis* with a pathogenic variant in any inheritance pattern
	BP3 In-frame deletions/insertions in a repetitive region without a known function
	BP4 Multiple lines of computational evidence suggest no impact on gene or gene product (conservation, evolutionary, splicing impact, etc.)
	Caveat: Because many in silico algorithms use the same or very similar input for their predictions, each algorithm cannot be counted as an independent criterion. BP4 can be used only once in any evaluation of a variant.
	BP5 Variant found in a case with an alternate molecular basis for disease
	BP6 Reputable source recently reports variant as benign, but the evidence is not available to the laboratory to perform an independent evaluation
	BP7 A synonymous (silent) variant for which splicing prediction algorithms predict no impact to the splice consensus sequence nor the creation of a new splice site AND the nucleotide is not highly conserved

***Tabella 9**. Criteri per classificare le varianti benigne*

Evidence of pathogenicity	Category
Very strong	PVS1 null variant (nonsense, frameshift, canonical ±1 or 2 splice sites, initiation codon, single or multiexon deletion) in a gene where LOF is a known mechanism of disease Caveats: • Beware of genes where LOF is not a known disease mechanism (e.g., *GFAP*, *MYH7*) • Use caution interpreting LOF variants at the extreme 3′ end of a gene • Use caution with splice variants that are predicted to lead to exon skipping but leave the remainder of the protein intact • Use caution in the presence of multiple transcripts
Strong	PS1 Same amino acid change as a previously established pathogenic variant regardless of nucleotide change Example: Val→Leu caused by either G>C or G>T in the same codon Caveat: Beware of changes that impact splicing rather than at the amino acid/protein level PS2 De novo (both maternity and paternity confirmed) in a patient with the disease and no family history Note: Confirmation of paternity only is insufficient. Egg donation, surrogate motherhood, errors in embryo transfer, and so on, can contribute to nonmaternity. PS3 Well-established in vitro or in vivo functional studies supportive of a damaging effect on the gene or gene product Note: Functional studies that have been validated and shown to be reproducible and robust in a clinical diagnostic laboratory setting are considered the most well established. PS4 The prevalence of the variant in affected individuals is significantly increased compared with the prevalence in controls Note 1: Relative risk or OR, as obtained from case–control studies, is >5.0, and the confidence interval around the estimate of relative risk or OR does not include 1.0. See the article for detailed guidance. Note 2: In instances of very rare variants where case–control studies may not reach statistical significance, the prior observation of the variant in multiple unrelated patients with the same phenotype, and its absence in controls, may be used as moderate level of evidence.
Moderate	PM1 Located in a mutational hot spot and/or critical and well-established functional domain (e.g., active site of an enzyme) without benign variation PM2 Absent from controls (or at extremely low frequency if recessive) (Table 6) in Exome Sequencing Project, 1000 Genomes Project, or Exome Aggregation Consortium Caveat: Population data for insertions/deletions may be poorly called by next-generation sequencing. PM3 For recessive disorders, detected in *trans* with a pathogenic variant Note: This requires testing of parents (or offspring) to determine phase. PM4 Protein length changes as a result of in-frame deletions/insertions in a nonrepeat region or stop-loss variants PM5 Novel missense change at an amino acid residue where a different missense change determined to be pathogenic has been seen before Example: Arg156His is pathogenic; now you observe Arg156Cys Caveat: Beware of changes that impact splicing rather than at the amino acid/protein level. PM6 Assumed de novo, but without confirmation of paternity and maternity
Supporting	PP1 Cosegregation with disease in multiple affected family members in a gene definitively known to cause the disease Note: May be used as stronger evidence with increasing segregation data PP2 Missense variant in a gene that has a low rate of benign missense variation and in which missense variants are a common mechanism of disease PP3 Multiple lines of computational evidence support a deleterious effect on the gene or gene product (conservation, evolutionary, splicing impact, etc.) Caveat: Because many in silico algorithms use the same or very similar input for their predictions, each algorithm should not be counted as an independent criterion. PP3 can be used only once in any evaluation of a variant. PP4 Patient's phenotype or family history is highly specific for a disease with a single genetic etiology PP5 Reputable source recently reports variant as pathogenic, but the evidence is not available to the laboratory to perform an independent evaluation

***Tabella 10**. Criteri per classificare le varianti patogeniche*

Certamente patogeniche	▪ 1 criterio molto forte E o 1 criterio forte O o 2 criteri moderati O o 1 criterio moderato ed 1 di supporto O o 2 o più criteri di supporto – 2 o più criteri forti O – 1 criterio forte E o 3 o più criteri moderati O o 2 criteri moderati E 2 o più criteri di supporto O o 1 criterio moderato E 4 o più criteri di supporto
Probabilmente patogeniche	– 1 criterio molto forte E 1 criterio moderato O – 1 criterio forte E 1-2 criteri moderati O – 1 criterio forte E 2 o più criteri di supporto O – 1 criterio moderato O – 3 o più criteri moderati O – 2 criteri moderati E 2 o più criteri di supporto O – 1 criterio moderato E 4 o più criteri di supporto
Certamente Benigne	– 1 criterio stand-alone O – 2 o più criteri forti
Probabilmente benigne	– 1 criterio forte e 1 criterio di supporto O – 2 o più criteri di supporto
Di incerto significato	– Nessuno dei criteri sopra menzionati è presente O – I criteri per stabilire se la mutazione sia benigna o patogenetica sono contraddittori

***Tabella 11**. Valutazione e combinazione dei criteri di patogenicità e benignità per la classificazione delle varianti*

L'utilizzo di NGS con pannelli di tutti i geni conosciuti per essere associati alle cardiopatie facilita la diagnosi genetica, rendendola molto più rapida ed efficiente. Inoltre, grazie a questa metodica, sono aumentate notevolmente le rilevazioni di mutazioni, sia quelle già note come

patogenetiche, che quelle di nuova identificazione, di incerto significato (VUS). Quando le mutazioni di incerto significato emergono da un'analisi genetica di questo tipo, sono necessari ulteriori studi che permettono di determinarne l'entità: gli studi funzionali. Dal momento che un'approfondita caratterizzazione funzionale richiede molto tempo e risorse, prima di procedere con questo tipo di analisi si va a ricercare la mutazione interessata a livello familiare con studi di segregazione.

Per quanto concerne le patologie cardiache, non è raro identificare, in pazienti in cui la diagnosi clinica è sospettata ma non certa, varianti geniche di incerto significato. Quando questo si verifica la diagnosi viene messa in discussione e la patogenicità della mutazione deve essere verificata tramite studi funzionali. Nel caso in cui ci trovassimo di fronte a casi di nuove mutazioni è possibile procedere nella ricerca delle medesime nei familiari, con lo scopo di identificare eventuali informazioni utili alla formulazione di una diagnosi che sia il più possibile attendibile, valutando con un maggior numero di informazioni a propria disposizione quanto sia effettivamente plausibile che una determinata mutazione sia associata a patologia e non si tratti invece di un polimorfismo o di una variante benigna.

Ottenere un risultato negativo da un test genetico non consente comunque di escludere la presenza di mutazioni legate allo sviluppo della malattia. Tale esito potrebbe essere dovuto ad una base multifattoriale, ma è anche possibile che siano presenti mutazioni in altri geni con suscettibilità non ancora conosciute o in porzioni degli stessi geni non ordinariamente esplorate con le metodiche impiegate per i test genetici, per esempio le regioni non codificanti. Per questo motivo un simile risultato del test genetico viene oggi denominato "non informativo" e non semplicemente "negativo", dato che esso non modifica sostanzialmente la situazione rispetto a quella di partenza.

Approfondimento del significato delle mutazioni di incerto significato

Studi di segregazione

Con studi di segregazione intendiamo l'insieme delle metodiche che ci permettono di ricercare nei parenti più prossimi del paziente le mutazioni ritrovate in quest'ultimo. Tali studi hanno lo scopo di confermare il pattern di ereditarietà genetica della mutazione precedentemente individuata sulla base dei dati clinici.

Nonostante gli studi di segregazione siano molto utili nella valutazione delle mutazioni di incerto significato, un risultato positivo non è indicativo di patogenicità della mutazione stessa, ma serve sia come prova che la mutazione è stata ereditata, sia per fornirci informazioni sul pattern di ereditarietà che segue.[51,52] I modelli statistici tengono conto della penetranza correlata all'età e della fenocopia (ovvero la presenza di uno stesso fenotipo in due individui geneticamente diversi che compare a causa di esposizioni ambientali particolari). Grazie a sofisticati sistemi che combinano le predizioni in silico e la contemporanea presenza di una nota mutazione patogenetica nello stesso individuo, si può ricavare una misura quantitativa. I modelli statistici tengono conto della penetranza correlata singola della patogenicità.[53] Negli studi di segregazione è anche molto importante includere parenti non prossimi al paziente, perché è meno probabile che questi possano avere sia la stessa mutazione che la patologia per caso, rispetto ai componenti dello stesso nucleo familiare. Inoltre, è sempre fondamentale eseguire un sequenziamento completo dei geni (comprese le regioni non tradotte), dal momento che potrebbero essere presenti altre mutazioni non rilevate in un primo momento capaci di alterare il nostro giudizio sulla mutazione.

Ritrovare una mutazione e un fenotipo ricorrenti in più di un individuo all'interno della stessa famiglia e in famiglie diverse anche etnicamente supporta la teoria che la mutazione sia patogenetica e, in questi casi, invece che di "prova di supporto", si parla di "criterio di

patogenicità moderato" o "criterio di patogenicità forte". D'altra parte, la mancanza della segregazione di una variante e del fenotipo associato depone a sfavore della patogenicità, per cui non si può mai prescindere da un'accurata valutazione clinica che dovrebbe essere in grado di riconoscere, oltre ai sintomi classici, anche quelli più blandi ed eventualmente le fenocopie. Allo stesso modo saranno determinanti sia l'anamnesi del paziente che una valutazione attenta delle relazioni interpersonali all'interno della famiglia, poiché eventi come consanguineità, adozioni, fecondazione eterologa e altre relazioni non biologiche vanno ad influenzare lo studio di segregazione e rendono quindi il criterio di valutazione più o meno incisivo.[54]

Studi funzionali

Gli studi funzionali possono essere un potente strumento a sostegno della patogenicità; tuttavia, non tutti gli studi funzionali sono efficaci nel prevedere un impatto sulla funzione di un gene o di una proteina. Ad esempio, alcuni saggi enzimatici offrono approcci consolidati per valutare l'impatto di una variante missenso sulla funzione enzimatica in una via metabolica (ad esempio, la funzione dell'enzima α-galattosidasi). D'altra parte, alcuni saggi funzionali possono essere predittori meno coerenti dell'effetto delle varianti sulla funzione delle proteine. Per valutare la validità di un test funzionale, è necessario considerare quanto il test funzionale rifletta l'ambiente biologico. Ad esempio, l'analisi della funzione enzimatica direttamente dal tessuto prelevato dal paziente sottoposto a biopsia o da un modello animale fornisce prove più evidenti rispetto all'espressione della proteina in vitro. Allo stesso modo, l'evidenza è più forte se il test riflette la piena funzione biologica della proteina (ad esempio, la scomposizione del substrato da parte di un enzima) rispetto a un solo componente della funzione (ad esempio, l'idrolisi dell'adenosina trifosfato per una proteina con proprietà di legame aggiuntive). Validazione, riproducibilità e dati di robustezza che valutano le prestazioni analitiche del test e tengono conto

dell'integrità del campione, che può essere influenzata dal metodo e dal tempo di acquisizione, nonché dalla conservazione e dal trasporto, sono fattori importanti da considerare. Questi fattori sono attenuati nel caso di un test sviluppato in laboratorio da Clinical Laboratory Improvement Amendments o in un kit disponibile in commercio. I saggi che valutano l'impatto delle varianti a livello di RNA messaggero possono essere altamente informativi quando si valutano gli effetti delle varianti alle giunzioni di splicing e all'interno di sequenze codificanti e regioni non tradotte, nonché regioni introniche più profonde (p. Es., stabilità, elaborazione o traduzione dell'RNA messaggero).[54] Gli approcci tecnici includono:

- **Test in silico**. Consiste nell'analisi dei dati utilizzando dei tools online per la verifica di eventuali riscontri precedenti avvenuti in altri centri e per la previsione, in base ai dati registrati fino a quel momento, della potenziale patogenicità o meno della variante in esame.

  Non sopravvalutare le prove computazionali è importante, in particolare dato che diversi algoritmi possono fare affidamento sugli stessi (o simili) dati per supportare le previsioni e la maggior parte degli algoritmi non sono stati convalidati contro varianti patogene ben consolidate. Inoltre, gli algoritmi possono avere capacità predittive molto diverse per geni differenti. *Se tutti i programmi in silico testati concordano sulla previsione, questa prova può essere considerata a sostegno*. Se le previsioni in silico non sono d'accordo, tuttavia, questa prova non dovrebbe essere utilizzata per classificare una variante. La presenza di una variazione dell'amminoacido in più specie di mammiferi in una regione altrimenti ben conservata, suggerendo che la stessa non comprometta la funzione, può essere considerata una forte evidenza per un'interpretazione benigna. Bisogna, tuttavia, essere cauti nell'assumere un impatto

benigno in una regione non conservata se il gene si è evoluto di recente negli esseri umani.[54]

- **Test in vitro**. L'analisi diretta di RNA e/o derivati del DNA complementari e saggi di splicing del minigene in vitro basati sull'estrazione del DNA da sangue periferico; in casi selezionati sarà necessario anche effettuare il null-allele test.
- **Test in vivo**. Esperimenti effettuati su topi knock-out.

Varianti nulle

Alcuni tipi di varianti (per esempio Nonsense, frameshift, delezioni multiesone) possono portare a perdita totale di funzione del gene e quindi all'assenza del prodotto genico per mancanza di trascrizione. Tuttavia, è necessario prestare attenzione quando si classificano queste varianti come patogene considerando i seguenti principi:

- Quando si classificano tali varianti come patogene, è necessario assicurarsi che le varianti nulle abbiano un meccanismo noto di patogenicità coerente con il modello ereditario stabilito per la malattia. Ad esempio, ci sono geni per i quali solo le varianti missenso eterozigoti causano la malattia e le varianti nulle sono benigne in uno stato eterozigote (ad esempio, molti geni della cardiomiopatia ipertrofica). Una nuova variante eterozigote non-senso nel gene MYH7 non sarebbe considerata patogena per la cardiomiopatia ipertrofica dominante sulla base esclusivamente di questa evidenza, mentre una nuova variante eterozigote non-senso nel gene CFTR sarebbe probabilmente considerata una variante patogena recessiva.
- Bisogna anche essere cauti quando si interpretano varianti tronche: la lunghezza della proteina tronca prevista influirebbe sull'assegnazione della patogenicità e tali varianti non possono essere interpretate senza un test funzionale.

- Per quanto riguarda le varianti del sito di giunzione, queste possono comportare il salto dell'esone, l'accorciamento o l'inclusione di materiale intronico come risultato dell'utilizzo alternativo del sito donatore/accettore o della creazione di nuovi siti. Sebbene si preveda che le varianti del sito di giunzione portino ad un effetto nullo, la conferma dell'impatto richiede uno studio funzionale mediante analisi dell'RNA o delle proteine. Si deve anche considerare la possibilità di una delezione/inserimento in frame, che potrebbe mantenere i domini critici della proteina e quindi portare ad un effetto lieve o neutro con minore cambiamento o un effetto di guadagno di funzione.
- Considerare la presenza di trascritti genici alternativi e capire se sono biologicamente rilevanti e in quali tessuti sono espressi.
- Bisogna anche essere cauti nell'assumere che una variante nulla condurrà alla malattia se trovata in un esone in cui non sono state descritte altre varianti patogene. Ciò è particolarmente vero se la variante troncante prevista è identificata come un reperto accidentale (non correlato all'indicazione per il test), data la bassa probabilità a priori di trovare una variante patogena in tale contesto.

CAPITOLO 3: METODI DI STUDIO DELLE VARIANTI

Il primo approccio per lo studio delle varianti è rappresentato dall'utilizzo di database e dalla consultazione della letteratura.

I database da cui si possono attingere informazioni possono essere divisi in 3 categorie:

- Popolazione;
- Malattia;
- Sequenza.

Dalla prima categoria è possibile verificare la frequenza di una determinata variante su una larga coorte, ma non contiene molte informazioni riguardo gli effetti funzionali delle varianti o su possibili associazioni fenotipiche. Quando si utilizzano i database della popolazione, è necessario specificare se sono state utilizzate coorti sane o affette da malattie e, se possibile, se è stato incluso più di un individuo di una famiglia, nonché la fascia di età dei soggetti.

I database delle malattie contengono principalmente varianti trovate in pazienti ed una valutazione della patogenicità delle stesse.

Quando si utilizzano questi database, i laboratori clinici devono valutare con quale frequenza essi vengono aggiornati e quali metodi sono stati utilizzati per l'inserimento dei dati; confermare l'utilizzo della corretta nomenclatura HGVS e determinare la costruzione del genoma ed i riferimenti alla trascrizione utilizzati per nominare le varianti; determinare il grado di convalida dei dati per l'accuratezza analitica e valutare eventuali metriche di qualità fornite per valutare l'accuratezza dei dati, che potrebbe richiedere la lettura di pubblicazioni associate.[54]

3.1 TECNICA SANGER

Negli anni ‘60 del ‘900 Fred Sanger e colleghi hanno elaborato un metodo di sequenziamento in grado di sequenziare gli acidi nucleici, dai più semplici ai più complessi. La metodica, comunemente nota con il nome di **sequenziamento Sanger**, prende il nome di “metodo della terminazione della catena”.

La tecnica consiste nell’utilizzo di dideossinucleotidi (ddNTP) che, rispetto ai normali deossinucleotidi (dNTP), presentano una modificazione a livello dello zucchero: mancano del gruppo ossidrile anche al 3’. Questa loro particolare struttura impedisce la formazione di un legame fosfodiesterico e, quindi, il legame di un altro nucleotide. Di conseguenza una volta che legano una sequenza nucleotidica nascente, ne bloccano la sintesi.

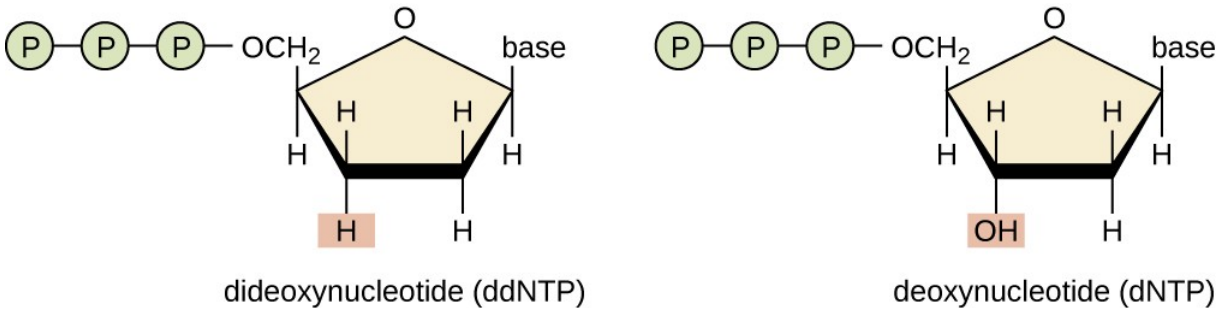


***Figura 15**. Confronto tra la struttura del dideossinucleotide (a sinistra) e quella del deossinucleotide (a destra)*

Il campione di partenza è un DNA a doppio filamento che viene denaturato, in modo tale da sfruttare l’azione della DNA polimerasi di utilizzare il filamento come stampo per sintetizzare il filamento complementare. Vengono effettuate in contemporanea quattro reazioni differenti, ciascuna delle quali contiene un primer di innesco, una DNA polimerasi ed i quattro dNTP. La differenza tra le quattro reazioni sta nel ddNTP che viene aggiunto – in quantità stechiometricamente inferiori rispetto ai dNTP – a ciascuna reazione, che è stato

opportunamente marcato. In questo modo quindi il filamento di nuova sintesi, complementare allo stampo, si allungherà incorporando i dNTP e si bloccherà quando verrà incorporato il ddNTP. Di conseguenza si creeranno frammenti di lunghezza differente, che saranno separati su un gel di poliacrilammide. Dall'osservazione delle differenti bande del gel si riuscirà a risalire alla sequenza nucleotidica.[55]

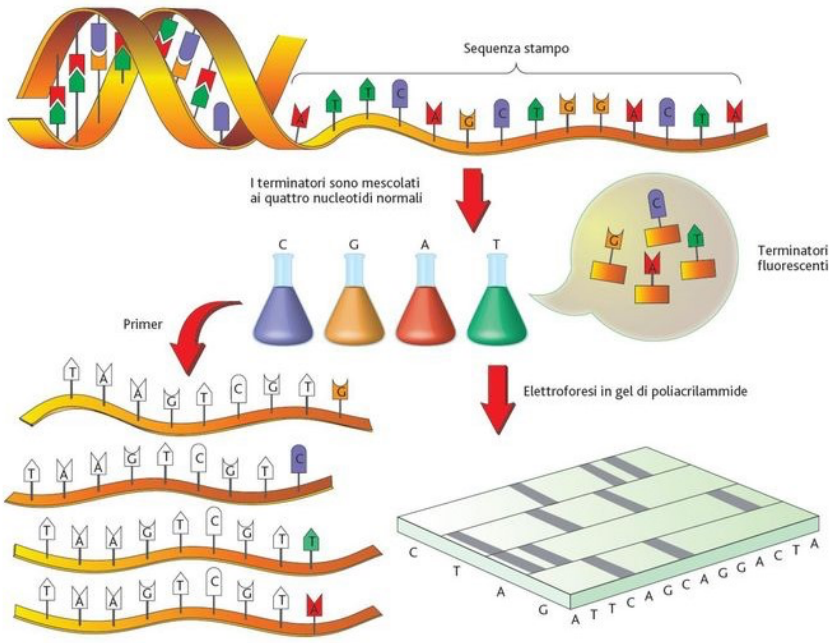


Figura 16. *Schema Tecnica di Sanger*

3.2 NEXT GENERATION SEQUENCING

3.2.1 L'evoluzione del sequenziamento: dalla tecnica Sanger alla tecnologia di terza generazione

La Next Generation Sequencing (NGS) rappresenta l'ultima evoluzione del sequenziamento: si basa su un principio analogo a quello della tecnica Sanger, ma la reazione può essere fatta per moltissimi frammenti di DNA in parallelo e non per uno soltanto. Tramite NGS è quindi possibile ottenere in modo più rapido ed economico una grandissima quantità di sequenze (in una singola corsa è possibile ottenere gigabasi, ed ormai quasi terabasi, di informazioni). Per

questo motivo la NGS è nota anche come high-throughput sequencing (sequenziamento ad alta resa).

Qualsiasi sistema NGS venga applicato tra quelli creati dalle diverse compagnie, ci sono dei passaggi comuni per tutti: la creazione di una libreria (sequencing library), l'amplificazione, il sequenziamento e l'analisi.

3.2.2 Preparazione della libreria (sequencing library)

Il campione di DNA genomico viene frammentato mediante taglio acustico, sonicazione o digestione enzimatica con DNAsi I o frammentasi e quindi etichettato con adattatori, tag, codici a barre e primer utilizzando metodi consolidati di legatura e PCR. In alternativa, la tecnologia di frammentazione di Illumina, chiamata Nextera Tagmentation, può essere implementata utilizzando un enzima trasposasi per frammentare e inserire simultaneamente sequenze di adattatori nel DNA ds e quindi ridurre la manipolazione del campione e il tempo di preparazione.[56] Per il sequenziamento mirato, gli esomi o le regioni di interesse all'interno del DNA frammentato possono essere catturati e arricchiti da kit di cattura dell'ibridazione della sonda o mediante amplificazione PCR con primer progettati su misura.

I frammenti di DNA così preparati tramite l'aggiunta di adattatori (o adaptors), necessari per ancorare e immobilizzare i frammenti al supporto su cui avverrà la reazione, costituiscono la cosiddetta *libreria di sequenziamento*.

Le librerie di DNA sono etichettate con tag di campioni di codici a barre, come l'identificatore multiplex (MID) per il sequenziamento Roche 454, per consentire il raggruppamento delle librerie e quindi massimizzare l'output della sequenza come fase di sequenziamento di ampliconi multiplex per ogni corsa di sequenziamento.[57]

Esistono almeno tre diversi tipi di adattatori e quindi tre diversi modi di preparare la sequencing library: adattatori lineari, circolari e a bolla.

3.2.3 Amplificazione

Dopo la costruzione della libreria, i frammenti di DNA vengono amplificati clonalmente mediante PCR in emulsione con microsfere[58,59] o mediante PCR in fase solida utilizzando primers attaccati ad una superficie solida[58,60,61] al fine di generare una quantità sufficiente di filamenti singoli, molecole di DNA e segnale rilevabile per produrre dati di sequenziamento sufficientemente affidabili.[62] Le piattaforme Roche 454, SOLiD di Life Technologies e Ion Torrent utilizzano la PCR in emulsione, mentre le piattaforme HiSeq/MiSeq di Illumina utilizzano la PCR in fase solida.[58]

Le copie clonali del frammento si legano poi all'enrichment bead ricoprendone la superficie sulla cosiddetta piastra Picotiter.

Un problema con la preparazione delle librerie di sequenziamento mediante amplificazione PCR è che la PCR introduce bias GC, una delle principali fonti di variazioni indesiderate ed errori nella copertura del sequenziamento.[63] L'utilizzo di metodi alternativi all'amplificazione PCR migliora la complessità della libreria, la copertura di regioni GC elevate e riduce il numero di letture duplicate.[64] In commercio sono disponibili diversi kit di preparazione delle librerie senza PCR, come NEXTflex PCR-Free di Bioo Scientific, kit di librerie Accel NGS 2S senza PCR di Swift Biosciences e il kit di preparazione dei campioni Illumina TruSeq DNA PCR-Free che utilizza la legatura amplificazione per Illumina e altri sistemi di piattaforme di sequenziamento.

3.2.4 Sequenziamento

Il sequenziamento avviene grazie a complessi meccanismi fluidici che, funzionando su scala microlitrica, regolano il flusso dei reagenti che vanno a cimentarsi col DNA immobilizzato.
Ogni ciclo di sequenziamento consiste nel cimentare il DNA immobilizzato con una soluzione contenente un nucleotide, nel successivo lavaggio e nella registrazione dell'evento molecolare appena verificatosi, tramite un sistema per immagini:

- **Roche 454 pyrosequencing by synthesis (SBS)** è stato il primo sistema di sequenziamento di seconda generazione di successo: questa tecnologia utilizza la chimica di sequenziamento, in base alla quale la luce visibile viene rilevata e misurata dopo essere stata prodotta da un sistema enzimatico ATP sulfurilasi, luciferasi, DNA polimerasi in proporzione alla quantità di pirofosfato che viene rilasciata durante l'incorporazione ripetuta di nucleotidi nella catena del DNA appena sintetizzata.[58,65,66]
  I principali svantaggi di questa tecnologia sono l'elevato costo dei reagenti e gli alti tassi di errore nelle ripetizioni di omopolimeri.
- **Illumina**. Oggi è il sistema di sequenziamento di maggior successo, in particolare con le piattaforme HiSeq e MiSeq. Il sequenziatore Illumina è diverso dal sequenziatore Roche 454 in quanto ha adottato la tecnologia del sequenziamento mediante sintesi utilizzando nucleotidi di terminazione della catena marcati in modo fluorescente rimovibili che sono in grado di produrre una maggiore produzione ad un costo inferiore del reagente.[58,66,67] Il DNA templato arricchito clonalmente per il sequenziamento viene generato mediante amplificazione a ponte PCR (nota anche come generazione di cluster) in colonie miniaturizzate chiamate polonie.[67]

- **Supported Oligonucleotide Ligation and Detection (SOLiD): sequenziamento mediante legatura e rilevamento degli oligonucleotidi.**

  Si basa sul sequenziamento di 2 nucleotidi mediante legatura (SBL).[58,66,67] Questa procedura prevede la ricottura sequenziale delle sonde al modello e la loro successiva legatura. Le reazioni di sequenziamento della serie SOLiD 5500 W utilizzano ancora sonde ottamer fluorescenti in cicli ripetuti di ricottura e legatura che vengono interrogate ed eventualmente decifrate in un complesso processo sottrattivo.[65,67,68]

- **Ion Torrent**. Le sequenze nucleotidiche vengono rilevate elettronicamente da cambiamenti nel pH della soluzione circostante proporzionali al numero di nucleotidi incorporati. Ogni volta che un nucleotide viene incorporato nel filamento in crescita, viene rilasciato un protone. Il supporto con il filamento in crescita viene cimentato con soluzioni contenenti un solo nucleotide per volta. Se il nucleotide cimentato non è complementare, questo non si lega, nessun protone viene rilasciato e non si ha quindi nessuna variazione di pH. Se per caso vengono incorporati due nucleotidi, vi sarà una variazione di pH "doppia", che verrà adeguatamente registrata.

  I due sequencers che usano questa tecnologia sono il sequencer Proton ad alto rendimento con oltre 165 milioni di sensori e la Ion Personal Genome Machine (PGM), con 11,1 milioni di sensori. Ion PGM funziona attraverso l'impiego di chip specifici: il chip Ion 314 fornisce letture più basse con 0,5 milioni di letture per chip, il chip 316 ha un output di 100 Mb, mentre il chip Ion 318 fornisce le letture più alte fino a 5,5 milioni di letture per chip. Ion PGM ha il vantaggio di essere molto veloce e con un'elevata accuratezza di analisi, che tende a restare molto stabile sia dopo molti cicli che in regioni tecnicamente difficili come le isole CpG.

La ion Proton è un po' più potente ed è infatti adatta all'exome e al genome sequencing e lavora con due tipi di chip: chip Proton I (160 milioni di pozzetti, adatto al sequenziamento di due esomi umani in parallelo) e chip Proton II (660 milioni di pozzetti, adatto al sequenziamento di un intero genoma umano).

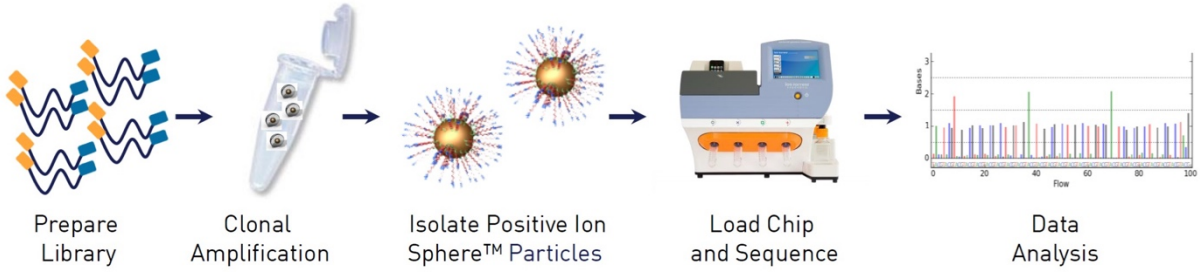


Figura 17. *Passaggi Ion Torrent*

Una volta completato il sequenziamento, i dati verranno analizzati nella fase computerizzata di analisi bioinformatica da cui si otterranno le varianti in tre fasi: alignment (allineamento), variant calling (selezione delle varianti) e filtering and annotation (filtraggio e annotazione).

3.2.5 Analisi

Alignment

Per analizzare il genoma (o l'esoma o anche solo pochi geni) di un individuo è necessario frammentare il suo DNA, così da poterlo amplificare prima e sequenziare poi. Al termine del sequenziamento, dunque, si ottengono tanti piccoli frammenti di DNA (*reads*) che devono essere ricongiunti a ricostruire le sequenze geniche del paziente. Questa operazione viene fatta allineando le reads alle sequenze di riferimento (reference sequences) del genoma umano, rese disponibili dal Progetto Genoma Umano e custodite in database come Genome Browser.

Variant calling

Le sequenze geniche e intergeniche mantengono un certo grado di omologia fra un individuo e l'altro, sia pur distinguendosi per la presenza di alcune varianti dette polimorfismi. Molti di questi polimorfismi sono noti e già caratterizzati, altri vengono identificati per la prima volta durante l'analisi del paziente. La quantità di dati ottenuta è enorme, poiché l'operazione di variant calling in un exome o genome sequencing porta all'identificazione di migliaia di varianti. La riduzione del numero di varianti da analizzare avviene nella terza fase, quella di filtering e annotation.

Filtering and annotation

Una volta identificate le varianti del paziente, il software procede in automatico alla loro caratterizzazione, filtrandole (selezionando cioè solo quelle che potrebbero avere un significato fenotipico) e annotandole (fornendo cioè tutti i dati disponibili nei database e/o in letteratura: se si tratti cioè di semplici polimorfismi, di varianti funzionali, di mutazioni patogene già descritte in letteratura o di varianti del tutto nuove e dunque dal possibile significato incerto).

3.2.6 Applicazione della Next Generation Sequencing

La possibilità di ottenere un'enorme quantità di informazioni sul patrimonio genetico individuale comporta notevoli vantaggi sia sul piano della ricerca, rendendo più facile l'individuazione di nuovi geni responsabili di tratti fenotipici, sia sul piano clinico, aumentando considerevolmente le potenzialità dei test genetici di raggiungere la diagnosi causale di malattia. In particolare, con i test di nuova generazione è possibile analizzare l'intero genoma/esoma anziché limitare l'indagine ai geni più frequentemente mutati nella patologia in esame.

In generale, si possono classificare 3 diversi tipi di varianti identificate con la NGS:

- **Varianti causative**. Sono quelle varianti a cui è possibile associare con certezza un ruolo patogenetico. Si tratta di varianti note in geni già associati a fenotipo-malattia, oppure varianti non descritte che, ad esempio, determinano una chiara perdita di funzione di un gene noto che si associa a malattia con meccanismo di aploinsufficienza.
- **Varianti responsabili di fenotipi non collegati al quesito clinico**. "Risultati incidentali" (IF).
- **Varianti di sequenza con effetti funzionali e clinici non definiti (VUS).** Tali varianti possono essere identificate in geni noti già implicati nel fenotipo per cui è stata richiesta l'indagine, oppure in altre regioni geniche non correlate ad esso. Il primo sottotipo di queste varianti (localizzate in geni responsabili della condizione clinica per cui è stato richiesto l'esame) è comune alle tecniche di sequenziamento tradizionale.[21]

3.2.7 Tipologie di test NGS

I test genetici basati sull'NGS possono essere suddivisi in base all'ampiezza della porzione di genoma analizzata.

Le principali categorie sono:

- Sequenziamento dell'intero esoma (Whole Exome Sequencing; WES);
- Sequenziamento dell'intero genoma (Whole Genome Sequencing; WGS);
- Sequenziamento mirato di pannelli di geni (Targeted NGS testing o Panel NGS testing);
- Analisi di tutti gli RNA prodotti da una cellula (Transcriptome Analysis).

Le applicazioni di WES e della target sequencing richiedono uno step aggiuntivo (il Target Enrichment) e possono essere eseguite anche per molti campioni in contemporanea, tramite il cosiddetto Multiplexing. I DNA di ciascun individuo possono infatti essere distinti attaccando

alle estremità di ogni frammento della sequencing library delle sequenze specifiche (dette *sequenze barcode*). Poiché nella sintesi delle reads vengono copiate anche le sequenze barcode, è possibile distinguere in un secondo momento le reads di ciascun individuo e separarle (de-multiplexing) prima di procedere all'allineamento con la reference sequence.

L'Enrichment si rende necessario per isolare le regioni genomiche di interesse ed ottenere da esse una quantità di DNA sufficiente all'analisi.

All'inizio si procede normalmente alla costituzione della libreria di sequenziamento.

Successivamente non si fa altro che ridurre la densità della libreria così ottenuta, limitando la sua consistenza alle regioni che si desidera analizzare. In alcuni sistemi (ad esempio, nel sistema Illumina) ciò viene realizzato con l'aiuto di sonde biotinilate e di granuli magnetici (magnetic beads). La sequencing library iniziale viene cioè cimentata con piccole sequenze di DNA specifiche (sonde biotinilate) che si legano alle regioni genomiche di interesse. In seguito, vengono aggiunti i magnetic beads che si legano a loro volta alle sonde biotinilate. È così possibile trattenere le regioni di DNA che si desidera analizzare su un supporto magnetico, mentre tutto il DNA restante viene lavato via. Le regioni di DNA così isolate possono quindi essere amplificate tramite PCR fino ad ottenere la quantità di DNA necessaria alla corsa di sequenziamento.

La scelta fra l'analisi di un singolo gene o di un pannello di geni dipende dal sospetto clinico e dalla genetica della malattia sospettata: per una patologia caratterizzata da ampia variabilità clinica, in particolare per quanto riguarda la gravità dei sintomi, e con un'ampia eterogeneità genetica è indicato procedere subito all'analisi del pannello comprendente tutti i geni noti fino ad ora, poiché mutazioni in questi geni sono identificabili solo in una minoranza di pazienti ed è verosimile che mutazioni anche in altri geni possano causare la malattia. Grazie all'analisi genetica è possibile delineare correlazioni genotipo-fenotipo; è tuttavia noto che la

sovrapposizione clinica fra i vari sottotipi genetici della malattia è molto estesa e non è quindi possibile fare a priori una scelta mirata per uno, due o tre geni in particolare da analizzare. Si ricorre quindi all'analisi di un pannello di geni, con due differenti opzioni:

- **Pannelli "pre-sequenziamento"** (*Targeted Resequencing*): permettono l'arricchimento di specifiche regioni genomiche prima del sequenziamento, così da analizzare un certo numero di pazienti in contemporanea ed abbattere i costi.
- **Pannelli "in silico"** (*Targeted Data Analysis*): si effettua un WES e successivamente si vanno ad analizzare solo quei geni già noti per essere associati alla patologia in esame. In questo caso è possibile analizzare solo pochi campioni per corsa, a seconda del tipo di strumentazione di cui si dispone, ed i costi sono maggiori.

Il vantaggio dei pannelli pre-sequenziamento è quello di permettere di mantenere costi più bassi ed utilizzare apparecchiature (*desktop sequencers*) meno dispendiose. Una volta disegnato, il pannello va poi testato per verificarne la resa sperimentale in termini di copertura ed accuratezza del dato ottenuto. Questo approccio ha inoltre il vantaggio di garantire una copertura elevata nelle regioni di interesse implementando così la probabilità di rilevare eventuali mosaicismi. Lo svantaggio di questa opzione consiste nella necessità di un aggiornamento continuo laddove nuovi geni vengano progressivamente associati ad una patologia di interesse. I pazienti che rimangono privi di diagnosi molecolare potranno successivamente essere rianalizzati mediante analisi dell'esoma/genoma.

Il pannello in silico, viceversa, ha costi maggiori. Per ottenere un coverage sufficiente mediante pannello in silico si richiede inoltre una buona pipeline di analisi che permetta, di volta in volta, di analizzare subset diversi di geni. Il vantaggio di questo approccio è che l'aggiornamento non richiede alcuna modifica al disegno ed alla piattaforma di arricchimento e permette, senza ulteriori costi, di analizzare "nuovi" geni malattia in pazienti già sequenziati. Inoltre, nel

momento in cui una rivalutazione clinica del paziente facesse emergere un sospetto differente da quello iniziale, può essere sufficiente cambiare il "target" dell'analisi.
Pertanto, per condizioni geneticamente eterogenee le cui basi genetiche siano ancora in larga parte sconosciute, l'uso di pannelli in silico può essere preferibile, in quanto permette di estendere l'analisi al resto dell'esoma in un contesto di ricerca, che può anche avere importanti ricadute cliniche; infatti ciò consente potenzialmente di riconoscere il coinvolgimento di geni malattia già noti in fenotipi "atipici" o di identificare nuovi geni malattia.[21]

Exome sequencing è un termine che identifica il sequenziamento dell'intera regione codificante del genoma di un individuo. Anche se la regione codificante rappresenta soltanto l'1% di tutto il genoma, si stima che fino all'85% di tutte le mutazioni patogene siano contenute in questa regione. L'exome sequencing è ad esempio utile nei casi di malattie geneticamente eterogenee, nelle quali alcuni geni devono ancora essere scoperti e per le quali i pazienti rimangono spesso senza la conferma genetica anche dopo l'esecuzione del test molecolare tradizionale.

Le applicazioni dell'exome sequencing sono principalmente tre:

- **Studio della patologia mendeliana**: l'approccio tradizionale alla ricerca delle malattie mendeliane è sempre stato lo studio di linkage seguito dalla tecnica del clonaggio posizionale. Questo metodo presenta tuttavia notevoli difficoltà, principalmente legate alla necessità di dover testare più di una famiglia (impresa non facile per malattie rare o rarissime) ed eterogeneità genetica, che talora caratterizza certe patologie mendeliane e che rende quasi impossibile identificare il gene chiave quando gli studi debbano limitarsi a poche famiglie. L'exome sequencing, invece, screenando a tappeto la

sequenza codificante, permette di identificare subito varianti patogene o verosimilmente patogene.

- **Studio della patologia multifattoriale**: negli ultimi anni l'approccio tradizionale allo studio genetico delle malattie multifattoriali (cioè quelle malattie che si suppone insorgano a seguito dell'interazione fra fattori ambientali e predisposizione genetica) è stato prevalentemente basato sul genome array, ossia sullo screening di un certo numero di varianti sparse su tutto il genoma. Questi studi, noti anche come GWAS (Genome Wide Association Study), si basano sull'ipotesi di fondo che la suscettibilità genetica ad una malattia multifattoriale sia dovuta all'interazione di varianti genetiche comuni. In effetti, le patologie multifattoriali sono patologie frequenti (si pensi solo al diabete di II tipo o alla cardiopatia ischemica) e l'ipotesi sottostante ad un GWAS è quindi "malattia frequente, variante frequente". L'applicazione della NGS ha tuttavia rivoluzionato anche questo settore: anche negli studi di associazione, cioè, si comincia a prendere in considerazione il genome o l'exome sequencing come alternativa al genome o exome array. L'exome sequencing, per via della sua economicità rispetto al genome sequencing, rappresenta in realtà l'alternativa di scelta principale. L'applicazione del genome/exome sequencing sottende una filosofia sostanzialmente opposta, e cioè "malattia frequente, variante rara". Si pensa cioè che la suscettibilità alla patologia multifattoriale sia dovuta all'interazione più fra varianti rare che fra varianti frequenti.
- **Diagnostica di routine**: sequenziare l'intero esoma, oltre ad essere in molti casi ormai economicamente più vantaggioso dell'analisi Sanger, offre una certa possibilità di identificare la causa genetica anche nei casi in cui manchi una conferma clinica precisa, dando per altro l'opportunità di caratterizzare geneticamente patologie ereditarie fino

ad ora poco conosciute. Il risultato di un semplice exome sequencing di routine può cioè rivelarsi non solo di interesse clinico, ma anche di interesse scientifico.

3.2.8 Raccomandazioni operative per i test NGS

- La caratterizzazione fenotipica è fondamentale per la scelta della tecnica di diagnosi molecolare (sequenziamento Sanger, Targeted sequencing, WES) e per la successiva analisi delle varianti identificate.
- Nel caso di fenotipi specifici, caratterizzati da bassa eterogeneità genetica e coinvolgenti geni di piccole dimensioni, può essere opportuno ricorrere ad analisi con metodiche molecolari convenzionali. Anche la lunghezza del gene può influenzare la scelta della tecnica analitica da utilizzare: in generale, è auspicabile che l'analisi molecolare di geni molto grandi (numerosi esoni) sia trasferita su piattaforme NGS per l'abbattimento dei costi e dei tempi di refertazione.
- Nel caso di condizioni con elevata eterogeneità genetica nelle quali mutazioni in un numero ridotto di geni sono responsabili della maggioranza dei casi è auspicabile l'uso di un targeted sequencing (pannelli di geni noti). Tale strategia è indicata anche in caso di sospetto mosaicismo.
- Nel caso di condizioni con eterogeneità genetica particolarmente marcata dove è coinvolto un numero sempre crescente di geni, ognuno dei quali è responsabile di una bassa percentuale di casi, è indicata l'esecuzione di un WES con un filtro limitato all'analisi di un pannello di geni noti (pannello "in silico"). In caso di negatività dell'analisi di un pannello di geni, i dati esomici restano quindi a disposizione per le eventuali indagini successive indirizzate alla ricerca di nuovi geni candidati o per l'analisi di geni causativi identificati in un secondo tempo.

- In tutti i casi in cui non può essere formulata un'ipotesi diagnostica su base clinica è preferibile l'analisi dell'intero esoma per l'individuazione del difetto genetico responsabile, che può riguardare mutazioni/geni già noti (ambito diagnostico) o nuovi geni candidati (ambito di ricerca) che necessitano di ulteriori conferme. È auspicabile che tali indagini siano eseguite in centri con comprovata esperienza nell'ambito dell'analisi di dati esomici per garantire la più alta probabilità di successo.

È utile confermare sempre la mutazione identificata tramite metodica NGS con sequenziamento Sanger e stabilirne la segregazione nella famiglia quando possibile.

3.2.9 Interpretazione dei risultati di un test genetico

In genetica molecolare i risultati di un test a volte possono essere di non semplice interpretazione: esistono numerose linee guida sull'argomento, sono tutte più o meno simili e si rifanno in gran parte alle linee stilate dall'American College of Medical Genetics (ACMG). Si possono dunque distinguere i seguenti tipi di mutazioni:

- **Mutazioni già descritte in letteratura come patogenetiche**

  Con alcune eccezioni, queste mutazioni possono essere tranquillamente interpretate come patogenetiche e la loro rilevazione nel DNA del paziente consente spesso di dare una conferma genetica della diagnosi clinica.

- **Mutazioni non precedentemente descritte ma che appartengono alla categoria di mutazioni verosimilmente patogenetiche**

  Rientrano in questa categoria tutte quelle mutazioni che si suppone generino un allele non funzionale (non-functional allele), di solito causa della traduzione di una proteina tronca o, più frequentemente, del decadimento dell'mRNA con conseguente perdita

totale della proteina. Tali varianti sono: le mutazioni nonsenso (che introducono un codone di stop), le piccole delezioni o inserzioni o delezioni/inserzioni che alterano il frame di lettura dell'mRNA (*frame-shifting*), le mutazioni che alterano i siti donatori e accettori di splicing, le grandi delezioni di uno o più esoni o dell'intero gene e, spesso, le grandi inserzioni o duplicazioni di uno o più esoni o dell'intero gene.

- **Mutazioni non precedentemente descritte, appartenenti alla categoria di mutazioni che potrebbero essere o non essere patogenetiche (mutazioni di incerto significato, VUS)**

  Fanno parte di questa categoria tutte le nuove mutazioni missenso, tutte le piccole delezioni o inserzioni o delezioni/inserzioni (*delins*) che non alterano il frame di splicing, lontano dalla giunzione esone/introne (tipicamente mutazioni localizzate all'interno dell'introne o mutazioni sinonime all'interno dell'esone) o che potrebbero influenzare significativamente la funzionalità dei siti donatori e accettori conservati alla giunzione esone/introne (tipicamente localizzate entro 10-20 posizioni nucleotidiche dalla giunzione esone/introne).

  Qualsiasi mutazione missenso non riportata precedentemente (*previously unreported*) va considerata di incerto significato, fintanto che non emergano da studi familiari o funzionali ulteriori evidenze sul loro possibile effetto.

- **Mutazioni non precedentemente descritte appartenenti alla categoria delle mutazioni verosimilmente non patogeniche**

  Questo gruppo contiene le mutazioni localizzate nella 5'UTR o 3'UTR, cioè tutte le mutazioni introniche e le mutazioni esoniche sinonime che verosimilmente non hanno un effetto sullo splicing.

- **Mutazioni precedentemente descritte come non patogenetiche**

  Un esempio è dato dalle mutazioni missenso note in letteratura per essere semplicemente dei polimorfismi con poca o nessuna influenza sulla funzionalità finale della proteina. Più raramente, possono far parte di questa categoria anche mutazioni di splicing o mutazioni nonsenso (evitate da meccanismi di salto di lettura o semplicemente localizzate alla fine del messaggero, laddove il prodotto proteico non risente della perdita di pochi residui amminoacidici alla fine della catena).

Per poter definire una variante come patogenica, VUS, o benigna, inoltre, si fa riferimento ai criteri riportati nelle Linee Guida dell'ACMG, secondo le quali è importante andare a verificare i seguenti parametri che possono concorrere a spostare la diagnosi verso una forma benigna o patogenica (Tabella 12):

- PVS1: variante nulla in un gene in cui LOF (loss of function) è un meccanismo noto di malattia;
- PS1: stessa variazione dell'aminoacido di una variante patogena stabilita in precedenza indipendentemente dalla variazione del nucleotide;
- PS2: mutazione De Novo in un paziente con patologia e assenza di storia familiare;
- PS3: studi funzionali in vitro o in vivo a supporto di un effetto dannoso sul gene o sul prodotto genico;
- PS4: prevalenza della variante negli individui affetti significativamente aumentata rispetto alla prevalenza nei controlli;
- PM1: variante in un punto caldo mutazionale e/o in un dominio funzionale critico e ben consolidato senza variazione benigna;

- PM2: variante assente dai controlli (o con frequenza estremamente bassa se recessiva) in Exome Sequencing Project, 1000 Genomes Project o Exome Aggregation Consortium;
- PM3: per disturbi recessivi, rilevati in trans con una variante patogena;
- PM4: lunghezza delle proteine cambia a seguito di delezioni/inserzioni nel frame in una regione non ripetitiva o varianti di stop loss;
- PM5: nuovo cambiamento di senso in un residuo amminoacidico in cui è stato visto prima un diverso cambiamento di senso determinato come patogeno;
- PM6: assunto de novo, ma senza conferme di paternità e maternità;
- PP1: cosegregazione con malattia in più membri della famiglia affetti in un gene noto in modo definitivo per causare la malattia;
- PP2: variante missenso in un gene che ha un basso tasso di variazione missenso benigna e in cui le varianti missenso sono un meccanismo comune di malattia;
- PP3: molteplici linee di prove computazionali supportano un effetto deleterio sul gene o sul prodotto genico (conservazione, evoluzione, impatto di splicing, ecc.);
- PP4: il fenotipo o la storia familiare del paziente è altamente specifico per una malattia con un'unica eziologia genetica;
- PP5: una fonte attendibile riporta recentemente la variante come patogena, ma le prove non sono disponibili al laboratorio per eseguire una valutazione indipendente;
- BA1: frequenza degli alleli > 5% in Exome Sequencing Project, 1000 Genomes Project o Exome Aggregation Consortium;
- BS1: frequenza degli alleli maggiore del previsto per il disturbo;

- BS2: osservato in un individuo adulto sano per un disturbo recessivo (omozigote), dominante (eterozigote) o legato all'X (emizigote), con penetranza completa prevista in tenera età;
- BS3: studi funzionali ben consolidati in vitro o in vivo non mostrano alcun effetto dannoso sulla funzione delle proteine o sullo splicing;
- BS4: mancanza di segregazione nei membri affetti di una famiglia;
- BP1: variante missenso in un gene per il quale è noto che principalmente varianti troncanti causano la malattia;
- BP2: osservato in trans con una variante patogena per un gene/disturbo dominante completamente penetrante o osservato in cis con una variante patogena in qualsiasi modello di ereditarietà;
- BP3: eliminazioni/inserimenti nel frame in una regione ripetitiva senza una funzione nota;
- BP4: molteplici linee di prove computazionali non suggeriscono alcun impatto sul gene o sul prodotto genico (conservazione, evoluzione, impatto di splicing, ecc.);
- BP5: variante trovata in un caso con una base molecolare alternativa per la malattia;
- BP6: una fonte attendibile riporta di recente la variante come benigna, ma le prove non sono a disposizione del laboratorio per eseguire una valutazione indipendente;
- BP7: una variante sinonima (silenziosa) per la quale gli algoritmi di predizione dello splicing non prevedono alcun impatto sulla sequenza di consenso della giunzione né la creazione di un nuovo sito di giunzione e il nucleotide non è altamente conservato.

	Benign		Pathogenic			
	Strong	**Supporting**	**Supporting**	**Moderate**	**Strong**	**Very strong**
Population data	MAF is too high for disorder BA1/BS1 **OR** observation in controls inconsistent with disease penetrance BS2			Absent in population databases PM2	Prevalence in affecteds statistically increased over controls PS4	
Computational and predictive data		Multiple lines of computational evidence suggest no impact on gene /gene product BP4 Missense in gene where only truncating cause disease BP1 Silent variant with non predicted splice impact BP7 In-frame indels in repeat w/out known function BP3	Multiple lines of computational evidence support a deleterious effect on the gene /gene product PP3	Novel missense change at an amino acid residue where a different pathogenic missense change has been seen before PM5 Protein length changing variant PM4	Same amino acid change as an established pathogenic variant PS1	Predicted null variant in a gene where LOF is a known mechanism of disease PVS1
Functional data	Well-established functional studies show no deleterious effect BS3		Missense in gene with low rate of benign missense variants and path. missenses common PP2	Mutational hot spot or well-studied functional domain without benign variation PM1	Well-established functional studies show a deleterious effect PS3	
Segregation data	Nonsegregation with disease BS4		Cosegregation with disease in multiple affected family members PP1	Increased segregation data →		
De novo data				De novo (without paternity & maternity confirmed) PM6	De novo (paternity and maternity confirmed) PS2	
Allelic data		Observed in *trans* with a dominant variant BP2 Observed in *cis* with a pathogenic variant BP2		For recessive disorders, detected in trans with a pathogenic variant PM3		
Other database		Reputable source w/out shared data = benign BP6	Reputable source = pathogenic PP5			
Other data		Found in case with an alternate cause BP5	Patient's phenotype or FH highly specific for gene PP4			

***Tabella 12**. Quadro delle prove a supporto della benignità o della malignità. BS: benign strong; BP: benign supporting; FH: family history; LOF: loss of function; MAF: minor allele frequency; path.: pathogenic; PM: pathogenic moderate; PP: pathogenic supporting; PS: pathogenic strong; PVS: pathogenic very strong.*

CAPITOLO 4: RAZIONALE DELLO STUDIO

Questo studio sperimentale nasce dalla necessità sempre più impellente di effettuare degli approfondimenti diagnostici in tutti quei casi in cui l'approccio clinico da solo non risulta essere sufficiente per giungere con relativa sicurezza ad una diagnosi in quelle patologie, soprattutto cardiache, in cui è ormai accertata l'esistenza di una base genetica, multifattoriale e con elevata eterogeneità.

Questo risulta essere ancor più importante quando nelle prime fasi di studio genetico si identificano delle mutazioni definite VUS, o di incerto significato, meritevoli di approfondimenti dal punto di vista funzionale, volti a determinare la patogenicità, probabile o certa, delle mutazioni in questione; devono essere in grado di associare un determinato genotipo al relativo fenotipo.

I casi presi in esame appartengono a tre categorie: casi segnalati dal reparto di cardiologia degli "Ospedali Riuniti" di Ancona, indagini richieste dalla medicina del territorio, o approfondimenti effettuati su campioni prelevati da soggetti deceduti per far luce sulle cause che hanno portato alla loro morte.

Delle patologie cardiache analizzate dai laboratori dell'Istituto di Medicina Legale dell'Azienda Ospedaliero-Universitaria "Ospedali Riuniti" di Ancona, sono risultate molteplici mutazioni geniche: ogni variante può portare ad alterazioni di tipo qualitativo o quantitativo, con una produzione aberrante o assente di proteine indispensabili per il corretto funzionamento cardiaco, come ampiamente esposto nelle pagine introduttive.

Importante, inoltre, è la distinzione tra polimorfismi e mutazioni associate a patologie, utile di fronte alla necessità di differenziare tra condizioni benigne o meno.

Gli studi funzionali, dopo la prima fase di valutazione con tools, devono proseguire con studi in vitro e, infine, con studi in vivo, utilizzando topi knockout, con lo scopo di rendere sempre più certi i risultati ottenuti.

Scopo primario di questa tesi, quindi, è quello di valutare se una variante è benigna o patogenica attraverso l'ausilio di tools computazionali di predizione e di saggiarne la sensibilità nonché la concordanza nel definire una variante.

CAPITOLO 5: MATERIALI E METODI

Sono stati presi in esame campioni di sangue periferico raccolto in EDTA, provenienti da pazienti di qualsiasi età, sia post-mortem sia su pazienti con sospetto clinico, o come conferma su familiari di primo grado di un probando, cioè di un paziente su cui è già stata confermata la patologia.

I campioni derivano da analisi effettuate nei laboratori dell'Istituto di Medicina Legale del presidio Ospedaliero-Universitario "Ospedali Riuniti di Ancona" nel corso degli ultimi 4 anni, da inizio 2016 fino ad agosto 2020.

Si è utilizzato un sistema di target sequencing, e con un On-Demand Panel personalizzato nell'apposito sito web del tool informatico Ion targeted AmpliSeqTM. Questo includeva le regioni codificanti di 101 geni utili per l'identificazione di alcune malattie rare ereditarie studiate nel nostro ospedale universitario, con campioni provenienti dal reparto di cardiologia, malattie rare, immunologia clinica, medicina del territorio e da reperti autoptici.

Il metodo utilizzato permette la sequenza delle regioni codificanti e parte dei siti di splicing con un'analisi sviluppata con tecnica NGS, su un pannello in-silico, filtrando:

- 26 geni coinvolti nella patogenesi della Sindrome di Brugada (BrS): ABCC9, AKAP9, CACNAC1, CACNA2D1, CACNB2, CALM2, FGF12, GPD1L, HCN4, HEY2, KCND2, KCND3, KCNE3, KCNE5, KCNH2, KCNJ8, PKP2, RANGRF, SCN10A, SCN1B, SCN2B, SCN3B, SCN5A, SEMA3A, SLMAP, TRPM4.

- 20 geni coinvolti nella patogenesi della Sindrome del QT lungo (LQTS): AKAP9, ANK2, CACNA1C, CALM1, CALM2, CALM3, CAV3, KCNE1, KCNE2, KCNH2, KCNJ2, KCNJ5, KCNQ1, NOS1AP, RYR2, SCN1B, SCN4B, SCN5A, SNTA1, TRDN.
- 9 geni coinvolti nella patogenesi di tachicardia ventricolare polimorfa catecolaminergica (CPVT): ANK2, CALM1, CALM2, CALM3, CASQ2, KCNJ2, RYR2, TECRL, TRDN.
- 18 geni coinvolti nella patogenesi della cardiomiopatia ipertrofica (HCM): ACTC1, ACTN2, CSRP3, DES, LAMP2, MYBPC3, MYH6, MYH7, MYL2, MYL3, MYOZ2, NEXN, PLN, TCAP, TNNC1, TNNI3, TNNT2, TPM1.
- 29 geni coinvolti nella patogenesi della cardiomiopatia dilatativa (DCM): ABCC9, ACTC1, ACTN2, ANKRD1, BAG3, CSRP3, DES, DSG2, DMD, EYA4, LDB3, LMNA, MYBPC3, MYH6, MYH7, NEXN, PLN, PSEN2, RBM20, SCN5A, SGCD, TAZ, TCAP, TMPO, TNNC1, TNNI3, TNNT2, TPM1, VCL.

Le librerie sono state preparate usando Ion Amplisequ Library kit 2.0 e i campioni sono stati rinominati con codice a barre usando Ion Xpress baracode Adapters: ogni libreria di campioni è stata unita e sottoposta a un'emulsione PCR da Ion PGM Hi-Q View OT2 kit. Il modello positivo Ion PGM Hi-Q View ISPs è stato arricchito su Ion One touh ES Instrument e sequenziato su Ion Torrent PGM instrument usando il kit di sequenziamento di Ion PGM Hi-Q View, 318 tipi di chip e modalità di lettura a 200 basi. Il test è stato effettuato tramite sequenziamento massivo parallelo (MPS) su strumento PGM ION TORRENT.

Si è eseguita l'analisi dei dati usando una pipeline bioinformatica personalizzata.

Su Torrent Suite 5.0.4 si sono allineate le letture e il genoma di riferimento umano, per poi proseguire con l'identificazione delle varianti rispetto al genoma di riferimento hg19. L'annotazione delle varianti è stata portata a termine utilizzando wANNOVAR e le varianti candidate sono state filtrate in base alla loro funzione (escludendo varianti introniche e sinonime) e in base alla frequenza dell'allele minore (MAF) < 0.1% nei database pubblici. Vengono quindi riportate solo le mutazioni esoniche missenso, nonsenso e frameshift e le mutazioni nei siti di splicing. Le regioni con un coverage (cioè il numero di volte che una data posizione di interesse viene sequenziata) minore di 30X sono escluse dall'analisi.

Le linee guida del "The American College of Medical Genetics and Genomics" (ACMG)[54] hanno fornito i mezzi per interpretare e dare la priorità alle varianti, con l'aiuto dei tools Varsome e InterVar e del sotfware Alamut Visual 2.1. La conferma delle varianti rare è testata con il sequenziamento di Sanger: i prodotti della PCR sono stati controllati dall'elettroforesi capillare Agilent Bioanalyzer, purificati con il mix di enzimi EXO-CIP, sequenziati in modo bidirezionale dal kit di sequenziamento Big Dye Terminator 1.1. I prodotti del sequenziamento sono stati rivelati da 3130 Genetic Analyzer e analizzati da SeqScape 3.0 allineando le sequenze al genoma di riferimento hg19.

Una volta eseguite le prime indagini, quindi, si può giungere a differenti risultati, evidenziando la presenza di un polimorfismo, una mutazione probabilmente o certamente benigna, una mutazione probabilmente o certamente patogenica, oppure una mutazione di incerto significato, che dovrà essere approfondita con successivi studi funzionali.

CAPITOLO 6: APPROCCIO ALLO STUDIO FUNZIONALE

All'interno di uno studio funzionale possiamo descrivere 3 parti:

- Tools computazionali;
- Studi in vitro;
- Studi in vivo.

Il primo approccio per lo studio delle varianti è quello della loro ricerca nei principali database ed il loro inserimento in tools informatici di predizione quali:

- **MutationTaster**

  MutationTaster utilizza un classificatore Bayesiano per predire il potenziale patogenetico di una mutazione. Il classificatore di Bayes è alimentato sia dai risultati di tutti i test che da quelli delle caratteristiche delle mutazioni. Da queste informazioni l'algoritmo calcola le probabilità che le mutazioni possano essere patogenetiche o semplici polimorfismi. La libreria è composta da > 390.000 mutazioni note come patogenetiche secondo HGMD Professional e > 6.800.000 SNPs benigne e Indel polimorfismi secondo il 1.000 Progetto Genoma (TGP).

  Modelli

  Sono stati previsti tre diversi modelli rivolti a differenti tipi di alterazioni: un modello rivolto verso varianti "silenti" (non-sinonimi o intronіche, senza il modello AAE – Amino Acid Exchange), un altro verso quelle mutazioni che portano a sostituzioni/inserzioni/delezioni di un singolo AA (modello AAE semplice) e un modello in cui avvengono cambi più complessi nella sequenza degli AA (per esempio mutazioni che introducono un codone di stop prematuro – modello AAE complesso). I modelli sono stati testati sia con tutti i comuni polimorfismi disponibili e compatibili,

che con tutte le mutazioni patogenetiche. MutationTaster determina automaticamente il modello corretto per ogni mutazione.

Risultato: valore di probabilità

Il valore di probabilità è la probabilità della predizione, per cui un valore vicino a 1 indica alta sicurezza della predizione mentre punteggi sotto 0.5 indicano che il nostro classificatore arriva a una diversa conclusione rispetto a quelle già determinate da altri classificatori.

MutationTaster inserisce le alterazioni in uno dei seguenti quattro gruppi:

- Patogenetiche;
- Patogenetiche automatiche – note per essere deleterie;
- Polimorfismi – probabilmente innocui;
- Polimorfismi automatici – sicuramente innocui.

MutationTaster tiene conto di numerosi fattori che rientrano nell'algoritmo del classificatore bayesiano e che vengono mostrati nei risultati una volta inserita la mutazione nell'apposito spazio nel sito di MutationTaster.

Tra questi fattori ricorrono:

- Nome e localizzazione dell'alterazione, sia a livello del gene che a livello del cromosoma dove si trova il gene mutato;
- Cambi di AA[69], posizione e stato di conservazione;
- Alterazioni frameshift;
- Presenza di polimorfismi noti;
- Conservazione dei nucleotidi tramite PhyloP e phastCons.[70] In phastCons si assegna uno score da 0 a 1, dove con 1 si intende la massima probabilità che il nucleotide sia conservato. Al contrario con PhyloP lo score va da -14 a +6. La

differenza tra i due è che in PhyloP l'alterazione viene confrontata anche con altre specie e inoltre viene fornita un'informazione sulla velocità con cui quel nucleotide è variato (un punteggio negativo significa un'accelerazione dell'evoluzione);

- Siti di splicing alterati con cambiamenti nella scelta degli esoni e degli introni[71] che poi andranno a comporre la proteina;
- Alterazione in una delle zone di Kozak;
- Caratteristiche perse o acquisite dalle proteine a causa della mutazione;
- Alterazione della lunghezza delle proteine;
- Anomalie nei codoni di stop e nelle code poliA;[72]
- Posizione di A nel codone iniziale ATG;
- Numero di nucleotidi dopo la zona NMD (nonsense-mediated mRNA decay) che normalmente dovrebbero essere più di 50;
- Lunghezza delle sequenze codificanti e non codificanti.

- **CADD**

Per simulare un numero di mutazioni e varianti proteiche equivalenti a quelle avvenute durante l'evoluzione da scimpanzé a uomo, è possibile utilizzare un modello evoluzionistico che analizza sia le sequenze in base alla frequenza di dimeri CpG che i tassi di mutazione localmente ridimensionati. I dettagli delle mutazioni sono stati estrapolati da Ensembl Variant Effect Predictor (VEP), dal progetto ENCODE e dal UCSC genome broswer track. Queste annotazioni riguardano un grande range di dati che includono: le metriche di conservazione come GERP, phastConse phyloP, i dati genomici funzionali con l'ipersensibilità alle DNA-asi e i transcriptor factor binding,

informazioni come la distanza tra esoni e introni o i livelli di espressione nelle linee cellulari comunemente studiate e i punteggi dei livelli proteici come Grantham, SIFT e PolyPhen. In CADD la matrice di calcolo contiene 29.4 milioni di mutazioni (metà osservate e metà simulate) e 63 diverse annotazioni. È stata poi sviluppata una macchina di supporto (SVM) che si occupa di calcolare le annotazioni dalle quali derivano le varie caratteristiche delle mutazioni, integrandole con un numero limitato di interazioni. Le stesse 63 annotazioni sono state ottenute per tutti gli 8.6 miliardi di possibili sostituzioni nel genoma di riferimento umano. Dopo che SVM ha osservato e simulato le mutazioni, il modello è stato applicato all'algoritmo per poter fornire un punteggio a tutte le possibili sostituzioni (Figura 18). Il punteggio finale "PHRED-scaled" (scaled C-score) va da 1 a 99 e si basa sul rango di ciascuna variante rispetto a tutte le possibili 8,6 miliardi di sostituzioni nel genoma di riferimento umano.[73] Il PHRED-score è un valore che indica la probabilità di patogenicità e tale valore aumenta consensualmente alla probabilità. Uno score di 30 è molto probabilmente patogenetico perché rappresenta lo 0.1% delle mutazioni con score 20, che sono a loro volta più probabilmente patogenetiche rispetto a quella con score 10.

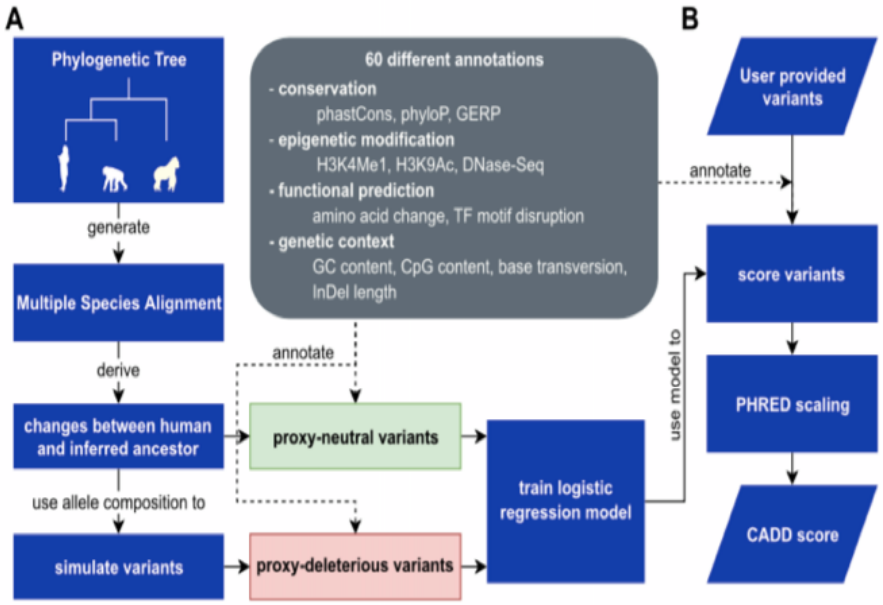


***Figura 18**. Algoritmo di CADD*

A. ***Per poter utilizzare CADD è richiesta l'identificazione delle varianti fisse o semifisse nella popolazione umana, ma assenti nel genoma ancestrale human-ape, "proxy-neutral variants".*** *La combinazione delle sequenze di queste mutazioni permette di determinare un set di varianti dette "proxy-deleterious". Usando più di 60 annotazioni diverse viene utilizzato un modello per classificare le varianti come "proxy-neutral" o "proxy- deleterious". Tutti i possibili polimorfismi del genoma umano di riferimento hanno le stesse caratteristiche e, pertanto, viene calcolato un punteggio CADD grezzo. La tabella di conversione PHRED deriva dalla classifica relativa del punteggio di questi modelli.*

B. *Viene inserito un set di varianti in formato VCF e l'algoritmo estrapola informazioni riguardanti il cromosoma, la posizione, l'allele di riferimento e l'allele alternativo. I punteggi vengono recuperati da file precaricati, oppure le varianti sono completamente annotate e viene calcolato il punteggio CADD. Il punteggio PHRED è poi ricercato nella tabella di conversione e entrambi i punteggi vengono riferiti nei risultati.*

- **FATHMM (Functional Analysis through Hidden Markov Models)**

  FATHMM è un software e un server che utilizza di base un metodo specie indipendente per la predizione delle funzioni di proteine mutate e che può adattarsi a determinate specie selezionando gli appositi pesi.

  Il software è formato da due algoritmi: uno basato sulla sequenza e sulla conservazione dei nucleotidi (non pesato) e l'altro che combina i primi due parametri con i pesi della patogenicità (pesato).

  I modelli probabilistici che compongono questo sistema di predizione possono essere usati per captare informazioni specifiche riguardo le mutazioni che ci permettono di confrontarle con sequenze omologhe. Per prima cosa, si presenta un metodo specie-indipendente, non pesato, nel quale sequenze omologhe vengono automaticamente raccolte e allineate usando una procedura di ricerca interattiva. Il risultante MSA (multiple sequence alignment) è poi utilizzato per costruire da zero il modello in cui analizzare la conservazione della sequenza, attraverso gli stati di unione interna del modello. Contemporaneamente viene analizzata la conservazione della sequenza all'interno dei modelli HMM curati manualmente, rappresentanti l'allineamento delle famiglie di domini proteici conservati (SUPERFAMILY e Pfam).

  Quest'ultima fase del procedimento è molto importante perché è in grado di individuare vincoli evoluzionistici e strutturali che passano spesso inosservati nell'allineamento automatico delle sequenze omologhe. Successivamente vengono introdotti i pesi specie-specifici. Nell'algoritmo pesato per specie e patogenicità i pesi derivano dalle frequenze relative della patologia e gli AASs funzionali e neutrali che mappano in domini conservati delle proteine. Usando un modello pesato per le mutazioni umane possiamo ottenere un'accuratezza che è paragonabile agli altri metodi di predizione tradizionali

come SIFT, PolyPhen e PANTHER.[74] Inoltre il software è in grado di predire le conseguenze molecolari e fenotipiche che emergono dalla mutazione, oltre che determinare se questa sia patogenetica o meno.

I punteggi approssimativamente uguali a 0 indicano che non c'è un cambio significativo nella sottostante struttura amminoacidica, al contrario valori inferiori a 0 indicano che la mutazione è probabilmente patogenetica, inoltre punteggi superiori a 0 indicano una mutazione probabilmente benigna e che potrebbe essere osservata più frequentemente nella popolazione wild type.[74]

Il punteggio finale è un numero che va da -16.13 a +10.64 con valori negativi patogenetici e valori positivi benigni.

- **FATHMM-MKL**

FATHMM-MLK predice le conseguenze funzionali di mutazioni a carico di singoli nucleotidi in zone codificanti e non. Le predizioni sono espresse come valori di *p* che vanno da 0 a 1: valori > 0.5 sono probabilmente patogenetici, mentre valori < 0.5 sono neutrali o benigni. Ovviamente, i valori che sono più vicini agli estremi hanno una maggiore affidabilità. Per caratterizzare le mutazioni sono stati individuati degli indicatori di posizione, sia per quelle che si trovano in regioni codificanti che per quelle localizzate in zone non codificanti del gene. Per le regioni codificanti sono stati selezionati 10 gruppi di caratteristiche, nominate da A a J, mentre per le regioni non codificanti i gruppi individuati sono 4 e sono nominati da A a D. Purtroppo, non tutti i gruppi hanno a disposizione informazioni tali da essere sufficienti, da un punto di vista statistico, a generare un valore di *p.* Per ovviare a questo problema si sono adeguati i pesi relativi alle categorie di cui si avevano dati sufficienti a quelle categorie che non li

avevano, anche se questo rende la predizione meno accurata, soprattutto se le caratteristiche mancanti riguardano la conservazione (gruppi A – E).

I database che vengono utilizzati dall'algoritmo sono due: uno basato su HGMD e l'altro, di controllo, utilizzante i polimorfismi a singolo nucleotide presenti in TGP (Thousand Genome Project) che potrebbe contenere anche diversi risultati positivi dato che al suo interno possiamo ritrovare mutazioni non annotate. Allo stesso modo, nel database sviluppato su HGMD potrebbero essere presenti dei risultati negativi. Pertanto, la classe positiva dovrebbe essere considerata come una serie di varianti associate alla malattia che sono arricchite da un impatto funzionale. Il database è stato poi modificato e suddiviso in base alla presenza di una sostituzione di AA o meno, dovuta alla mutazione.

I dieci gruppi di caratteristiche selezionati e nominati da A a J che sono utilizzati dal software per le mutazioni presenti in regioni codificanti sono i seguenti:

A. 46-Way Sequence conservation;
B. Modificazioni di istoni;
C. Trascrizione di Factor Binding Sites (TFBS PeakSeq);
D. Cromatina aperta basata su DNasi-Seq;
E. 100-Way Sequence Conservation;
F. Contenuto di GC;
G. Cromatina aperta basata su FAIRE (Formalehyde-assisted isolation of regulatory elements);
H. Trascrizione di Factor Binding Sites (TFBS SPP);
I. Segmentazione del genoma;
J. Impronta genetica.

Il database riguardante le mutazioni che avvengono in regioni non codificanti è poco bilanciato, dal momento che contiene più di 6,7 milioni di dati negativi e solamente 12.438 di quelli positivi. Per eliminare quanti più bias possibili, si è proceduto a filtrare un elevato numero di mutazioni[75] ma, nonostante ciò, sussisteva il problema di dati mancanti in alcune delle mutazioni, se si utilizzava il sistema a 10 gruppi di caratteristiche. La soluzione si è trovata andando ad utilizzare un sistema a 4 gruppi di caratteristiche che, comunque, permette di avere un database notevole, con un elevato numero di casi positivi che presentano tutte le caratteristiche (da A ad E).[76]

- **SIFT: Sorting Intolerant From Tolerant**

  SIFT utilizza un sistema di predizione che stabilisce se la sostituzione di un amminoacido determina o meno un'alterazione indiretta della funzione della proteina, andando ad alterare il suo fenotipo.[77,78] Il software si basa sia su sequenze proteiche omologhe, che sulle proprietà fisiche dei singoli aminoacidi e può essere applicato sia a polimorfismi non sinonimi sia a mutazioni missenso ottenute in laboratorio.

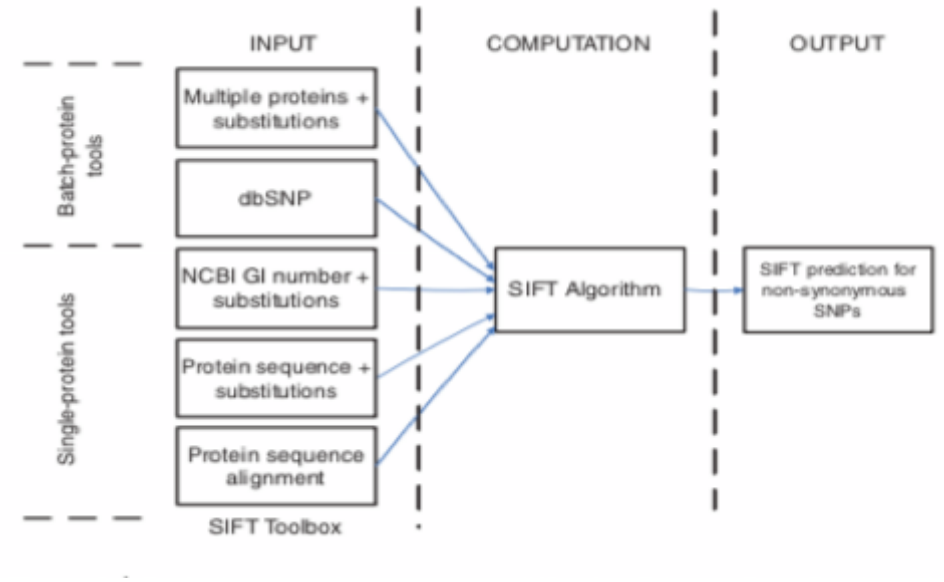


***Figura 19**. Algoritmo base di SIFT*

SIFT è stato applicato ai database di mutazioni umane ed è riuscito a distinguere mutazioni patogenetiche da polimorfismi neutrali (o benigni).[78] il punteggio finale che otteniamo come risultato varia da 0 a 1, con valori < 0,05 predetti come non tollerabili. L'algoritmo si basa, a differenza di altri tools, sulla sequenza degli aminoacidi per effettuare la predizione e non sulla struttura proteica. Nonostante ciò, i risultati che fornisce sono molto simili agli altri algoritmi[78-81] ed è dotato di una maggiore capacità analitica, dal momento che può studiare un più elevato numero di polimorfismi non sinonimi rispetti agli altri software.[77] SIFT si basa sul concetto che gli aminoacidi importanti per il funzionamento di una proteina rimangono conservati all'interno della famiglia della proteina stessa e, pertanto, certi cambiamenti in posizioni dette "ben conservate" tendono ad essere classificati come patogenetici. Assumendo che le mutazioni patogenetiche alterino la funzione proteica, si è applicato il software ad un database di mutazioni patogenetiche note[82] che vengono utilizzate come termine di paragone per le sequenze inserite da studiare. Il software considera la posizione nella quale è avvenuto il cambio amminoacidico e il tipo di aminoacido che viene cambiato. A questo punto avviene il confronto della sequenza con le proteine correlate, inserite nel database tramite l'allineamento delle sequenze. Le proteine da allineare vengono ricercate tramite un altro algoritmo, detto PSI-BLAST[83] (Figura 19). Successivamente, SIFT calcola la probabilità che il cambio di aminoacido non comporti variazioni per la funzionalità della proteina sulla base dell'allineamento precedentemente effettuato e la indica con un valore numerico. Se il valore, una volta normalizzato, è uguale o inferiore a 0,05 cut-off, la mutazione è predetta come patogenetica[77], anche se alcuni utenti hanno rilevato una migliore sensibilità del software alzando il cut-off a 0,1. Come in ogni test, avremo la possibilità di incorrere in risultati falsi positivi. In questo caso tali esiti sono

dovuti all'allineamento di proteine troppo simili tra loro che falsano l'algoritmo, inducendolo a rilevare come sequenze altamente conservate delle posizioni che in realtà sono neutre. Per ovviare all'elevato numero di falsi positivi che si otteneva con questa tecnica il programma calcola un valore di conservazione mediano che misura la diversità delle sequenze in allineamento e la conservazione di un sito va da 1/44,32, quando una posizione è completamente conservata, a 0, quando tutti i 20 amminoacidi sono osservati nella stessa posizione. Il valore di mediana di conservazione è settato di default a 3,0 e, aumentandolo, si avrà un maggior numero di falsi positivi: non è pertanto raccomandato l'uso di SIFT e degli altri tools informatici per la formulazione di diagnosi cliniche ma, piuttosto, il loro utilizzo è principalmente di supporto ad esperimenti futuri.[84]

Per poter utilizzare SIFT e introdurre una variante ci vengono presentate sei possibilità: due sono strumenti che consentono previsioni per più proteine e sostituzioni, mentre le restanti quattro offrono previsioni dettagliate per una singola proteina, con la possibilità di analizzare tutte le sostituzioni o solo alcune di esse.

- Tool A: Protein IDs with substitutions. Questo approccio dovrebbe essere utilizzato se si desidera avere un'analisi di più proteine che presentano polimorfismi sinonimi o non sinonimi oppure sostituzioni amminoacidiche. Per 1.000 proteine inserite il software impiegherà circa 10 minuti per elaborare i risultati ed è possibile impostare due parametri:
  - Si possono selezionare le sequenze da includere (BEST HITS/ALL HITS);
  - Si può scegliere di individuare la percentuale d'identità, cioè si vanno ad eliminare le sequenze identiche a quelle della richiesta in modo da

evitare che queste vengano analizzate e che falsino il punteggio, risultando come polimorfismi.

- Tool B: dbSNP identifiers. Anche questo tool analizza più proteine contemporaneamente e dovrebbe essere usato quando si è interessati nella predizione di polimorfismi noti, contenuti nel database. Per 1.000 proteine inserite i risultati verranno elaborati in 2-3 minuti.
- Tool C: GI number (Geneinfo Identifier number). Inserendo il numero GI o RefSeq (Reference Sequence database) è possibile ottenere predizioni dettagliate della proteina corrispondente. Questo metodo è molto più rapido rispetto al metodo D (protein sequence), perché grazie al numero identificativo SIFT rileva molto più velocemente le proteine associate da allineare con la richiesta. È un tool molto usato se si stanno ricercando tutte le sostituzioni amminoacidiche presenti in ogni posizione in quella proteina specifica. È inoltre possibile specificare quale mutazione si sta ricercando, come nel tool A.
- Tool D: Protein sequence. La ricerca della proteina secondo questo metodo è piuttosto lenta (circa 20 minuti), perché si va ad inserire la sequenza per analizzarla e BLAST deve ricercare le proteine per l'allineamento una per una. Per risparmiare tempo si consiglia di utilizzare il tool C se sono noti il numero GI oppure la RefSeq. Questo tool permette di selezionare tre parametri:
  - Database to search;
  - La conservazione mediana delle sequenze (da 0 a 1/44,32) con un valore di default di 3,0;
  - Percentuale di identità come nel tool A.

- Tool E: Protein sequence alignment. Questo tool viene utilizzato quando si hanno numerose proteine di allineamento alla richiesta, con sequenze omologhe.
- Tool F. Anche questo viene utilizzato quando si hanno numerose proteine in allineamento con la richiesta ma, a differenza del tool E, permette di inserire gli omologhi senza prima allinearli.[85]

- **MutationAssessor**

  Il server in questione è in grado di predire l'effetto di sostituzioni amminoacidiche e dei polimorfismi missenso sulle proteine e, inoltre, ha anche un utilizzo in campo oncologico nell'analizzare le mutazioni che colpiscono le cellule tumorali. Il meccanismo alla base della predizione è simile a quello utilizzato da SIFT: si parte dal presupposto che gli aminoacidi presenti in posizioni conservate abbiano un ruolo cruciale nel determinare la funzione delle proteine e, per questo, il server mette a confronto la sequenza della proteina da analizzare con altre proteine omologhe. Questo server fornisce un link all'analisi delle varianti, alle annotazioni, alla pagina html di allineamento ed alla pagina della struttura 3D predetta. Le sequenze inserite vengono associate alle sequenze di proteine Uniprot e Refseq (se possibile) ma, se il residuo di riferimento nella sequenza di proteine Uniprot è diverso da quello indicato nella variante da predire, l'analisi non verrà eseguita. Quello che differenzia MutationAssessor dagli altri sistemi di predizione è l'utilizzo della conservazione delle proteine in sottofamiglie. Tali sottofamiglie sono costituite dall'associarsi di sequenze di allineamento di sequenze omologhe che determinano una specifica funzione.[86,87] Il sistema di calcolo è molto complesso, tanto che comprende più di 30 caratteristiche biologiche da

confrontare, l'allineamento di mappe 3D multiple ed è inoltre in grado di analizzare migliaia di varianti attraverso WEBAPI.[88]

Spesso però i residui da identificare sono molto specifici, quasi da determinare una firma dei residui funzionali, e numerosi al tempo stesso, per cui l'algoritmo di MutationAssessor da solo non riesce sempre a rilevarli. Per questa ragione si è sviluppato un nuovo algoritmo chiamato Combinatorial Entropy Optimization (CEO) che ottimizza la suddivisione in sottofamiglie, e poi in clusters, in base alle diverse funzioni delle proteine. Questa firma è un insieme di residui chiave (posizioni di sequenza) che variano tipicamente tra sottofamiglie, ma che sono conservate all'interno di ogni sottofamiglia e la procedura classifica i residui chiave. Il parametro A influenza sia il valore ottimizzato della funzione di contrasto, sia il numero di sottofamiglie e l'insieme dei residui caratteristici che costituiscono la configurazione ottimale risultante. Per testare la solidità dei risultati si sono apportate modifiche ai vari parametri per poi verificare l'affidabilità del test che è stato confrontato ad altri software analoghi. Sebbene i vincoli evolutivi sui residui di specificità possano essere il risultato di qualsiasi tipo di interazione funzionale, i residui nelle interazioni proteina-proteina e nelle interazioni proteina-acido nucleico (NA) sono particolarmente ben definiti nelle strutture tridimensionali. In questi modelli è possibile evidenziare una forte coincidenza tra i residui di specificità ed i siti di legame, ad indicare che il metodo identifica correttamente i vincoli funzionali ponendoli sui residui di sito di legame.[71,72]

La predizione definitiva può appartenere a un particolare gruppo ("neutrale", "bassa", "media", "alta") e viene associata ad un punteggio chiamato "rank score" che va da 0 a 1, dove 1 indica una mutazione sicuramente patogenetica.

- **DANN: Deleterious Annotation of genetic variants using Neural Networks**

  È un tool informatico che ci permette di riconoscere varianti patogenetiche a partire da mutazioni genetiche, soprattutto non codificanti. DANN si basa su un algoritmo detto "neural network" che consiste in una banda di immissione della richiesta, una funzione sigmoide uscente come risultato, e tre strati da 1.000 nodi nascosti con funzione di attivazione tangente iperbolica. Questo tool predilige le presunte mutazioni patogenetiche, come quelle che derivano dallo studio di associazione Genome Wide (GWAS).

  L'SVM (macchina a vettori di supporto) di CADD può solo apprendere rappresentazioni lineari di dati, il che limita le sue prestazioni; per ovviare a questo, è stato implementato un algoritmo di rete neurale profonda (DNN) chiamato DANN. Un DNN è una rete neurale artificiale con diversi strati nascosti tra i livelli di input e output. Gli strati extra offrono un livello di astrazione aggiunto alla DNN, ma possono aumentare notevolmente il tempo di calcolo necessario per la formazione, per ogni variante sono state definite 949 funzioni. Il set di funzionalità è limitato e include un mix di numeri, numeri interi e valori binari reali.[89] I valori che otteniamo come risultato vanno da 0 a 1, con 1 che indica le varianti più pericolose.

- **PROVEAN: Protein Variation Effect Analyzer**

  PROVEAN è un software in grado di predire gli effetti di una mutazione, come la sostituzione, delezione o inserzione di un AA, sugli effetti biologici e le funzioni di una proteina. Un rapido calcolo ha permesso le previsioni per la sostituzione di 20 AA e per la delezione di un singolo AA, allo scopo di ottenere i punteggi di allineamento di sequenze a coppie per tutti gli AA di tutte le sequenze proteiche nell'uomo e nel topo.[90]

Il software genera una clusterizzazione con un parametro di identità di sequenza globale del 75%. I primi 30 cluster di sequenze strettamente correlate formano il set di sequenze di supporto, che verrà utilizzato per generare la previsione. Successivamente, viene calcolato un punteggio di allineamento delta per ciascuna sequenza di supporto. I punteggi vengono quindi calcolati come media all'interno e tra i cluster per generare il punteggio PROVEAN finale. Se il punteggio è uguale o inferiore ad una soglia predefinita, si prevede che la variante proteica abbia un effetto "deleterio", mentre se il punteggio è al di sopra della soglia, si prevede che la variante abbia un effetto "neutro". Il punteggio di allineamento delta si basa sulle varianti proteiche e sui riferimenti, tenendo conto degli omologhi, presenti sul database proteico di NCBI NR. Il punteggio soglia è generalmente di -2,5 ma può essere modificato per andare ad incrementare la specificità (abbassandolo per esempio a -4,1) o la sensibilità (alzandolo a -1,3) (Figura).

Il punteggio finale può essere un numero da -14 a +14, dove -14 indica una mutazione sicuramente patogenetica.

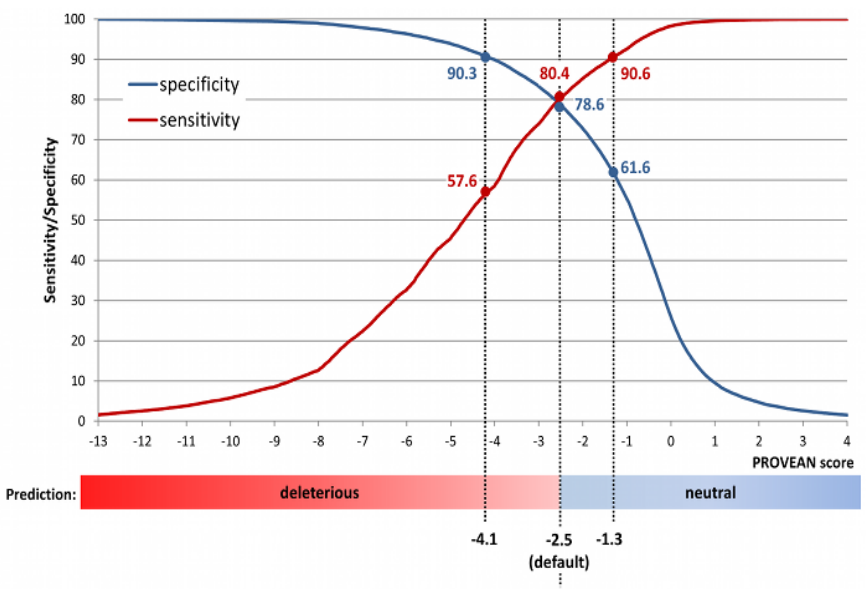


***Figura 20**. Selezione del cutoff in Provean*

- **LRT: Likelihood Ratio Test Query**

  Usando un set di dati genomici che mettono a confronto 32 specie di vertebrati si è dimostrato che un test del rapporto di verosomiglianza (LRT) può identificare un sottoinsieme di mutazioni che interrompono sequenze proteiche altamente conservate, che risulteranno probabilmente in proteine alterate e non funzionali.[91] Per distinguere le posizioni amminoacidiche funzionalmente obbligate o meno si è generato un database in cui vengono messe a confronto 32 specie di vertebrati: le enormi differenze tra le specie implicano che le posizioni non funzionalmente correlate non sono conservate mentre, al contrario, le mutazioni che incorrono in posizioni conservate sono più probabilmente patogenetiche.

  Per identificare le mutazioni deleterie presenti all'interno di un singolo genoma umano, sono state esaminate varianti non sinonime presenti in tre genomi[92-94] e per eliminare gli errori di sequenziamento, è stato usato solo un sottoinsieme di alleli di alta qualità.

  Stima del tasso di falsi negativi

  Il tasso di falsi negativi è stato stimato utizzando mutazioni missenso-patologiche note nel database OMIM. Delle 5.493 mutazioni, il 71,9% sono significative per LRT ($p<0,001$). Delle 1.546 mutazioni della malattia che non sono state identificate, il 42% si verifica in posizioni con dati genomici comparativi scarsi o inesistenti.

  Lo scopo principale di LRT è sia quello di portare alla luce le relazioni filogenetiche usando un quadro probabilistico sia quello di minimizzare i cambiamenti compensativi, che si ritiene siano comuni[95] e che possano causare falsi negativi se si usano solamente specie vertebrate strettamente correlate tra loro nell'analisi comparativa.

  Questo tool informatico differisce anche da altri metodi genomici comparativi in quanto tutti i cambiamenti di aminoacidi sono trattati allo stesso modo, piuttosto che applicando

i cosiddetti “pesi” e quindi considerando diversamente i cambiamenti di aminoacidi radicali e quelli conservativi.

L’LRT è stato usato per confrontare il modello nullo, in base al quale ciascun codone si sta evolvendo in modo neutro, con il modello alternativo, secondo cui il codone si è evoluto tramite un meccanismo di selezione negativa. Per rendere disponibili alla comunità le previsioni dell’LRT per ogni potenziale variante non sinonima del genoma umano, si è applicato questo software a 10.073.284 codoni, per i quali esisteva un numero sufficiente di specie allineate: un totale di 5.828.045 siti erano significativi ($p<0,001$ e $dN/dS<1$). SNP non sinonimi e sinonimi sono stati identificati usando geni noti e sono stati esclusi i SNP di codifica che erano presenti in frame di lettura ambigui, per esempio a causa della sovrapposizione di geni adiacenti o a frameshift causate da varianti di splicing note o pseudogeni noti.

Inoltre, per evitare errori di sequenziamento, sono stati applicati rigorosi filtri di qualità. Il risultato finale è un numero che va da 0 a 1, dove 0 indica una mutazione sicuramente patogenetica.

- **PolyPhen-2: Polymorphism Phenotyping 2**

  È uno strumento automatico per la previsione del possibile impatto di una sostituzione amminoacidica sulla struttura e sulla funzione di una proteina umana. Questa previsione si basa su semplici regole empiriche applicate alla sequenza ed alle informazioni filogenetiche e strutturali che caratterizzano la sostituzione. Il programma verifica se la mutazione ricade all’interno di un dominio proteico essenziale per la formazione della struttura secondaria e terziaria della proteina o per il legame con altre molecole. PolyPhen controlla anche l’insieme delle **proteine omologhe** per vedere se la stessa

mutazione sia stata osservata anche in altre proteine della stessa famiglia. Per una data sostituzione amminoacidica in una proteina, PolyPhen estrae varie sequenze e caratteristiche basate sulla struttura del sito di sostituzione e le invia a un classificatore probabilistico. Data una proteina mutata, PolyPhen cerca di individuare la voce corrispondente nel sottoinsieme delle proteine umane del database UniProtKB/Swiss-Prot. Quindi utilizza le informazioni a lui note per effettuare un confronto e verificare se la sostituzione dell'amminoacido avviene in un sito annotato come partecipante a un legame covalente, come sito di interesse o come parte di una regione di interesse. Sulla base di queste annotazioni, si determina se la sostituzione è in una regione transmembrana annotata o prevista. Per le sostituzioni nelle regioni transmembrana, PolyPhen-2 utilizza il punteggio della matrice specifica transmembrana PHAT[96] per valutare il possibile effetto funzionale; la sostituzione dell'amminoacido può essere incompatibile con lo spettro di sostituzioni osservato nella posizione nella famiglia delle proteine omologhe. PolyPhen identifica gli omologhi delle sequenze di input tramite la ricerca BLAST nel database UniRef100. L'insieme di hit BLAST viene filtrato per mantenere gli hit che hanno: identità della sequenza per la sequenza di input nell'intervallo dal 30% al 94%, incluso, e allineamento con la **sequenza di query** non inferiore a 75 residui di lunghezza. L'identità della sequenza è definita come il numero di corrispondenze diviso per la lunghezza dell'allineamento completo. L'allineamento multiplo risultante viene utilizzato dal software PSIC (Position Specific Independent Counts) per calcolare la cosiddetta matrice del profilo.[97] Gli elementi della matrice (punteggi del profilo) sono rapporti logaritmici tra la probabilità che un dato amminoacido si verifichi in una particolare posizione e la probabilità che questo amminoacido si verifichi in qualsiasi posizione (frequenza di fondo). PolyPhen calcola

la differenza tra i punteggi del profilo delle due varianti alleliche nella posizione polimorfica: grandi valori positivi di questa differenza possono indicare che la sostituzione studiata è raramente o mai osservata nella famiglia delle proteine. PolyPhen mostra anche il numero di sequenze allineate nella posizione della query e può essere utilizzato per valutare l'affidabilità dei calcoli del punteggio del profilo.[98]

L'output della pipeline di previsione PolyPhen è una previsione di probabilmente dannoso, potenzialmente dannoso o benigno, insieme a un punteggio numerico compreso tra 0,0 (benigno) e 1,0 (dannoso). Una previsione di probabile danno significa che si prevede che la sostituzione della sequenza di query sia dannosa con elevata confidenza, mentre una previsione di benigno significa che si prevede che la sostituzione sia benigna con elevata confidenza, inoltre una previsione di possibile danneggiamento significa che si prevede che la sostituzione della sequenza di query sarà dannosa, ma con scarsa probabilità. Sebbene il punteggio potenzialmente dannoso sia spesso interpretato come un'indicazione di un effetto lieve o di una penetranza bassa, in generale, è inteso come una misura della fiducia nella previsione piuttosto che della dimensione dell'effetto.

È difficile stimare l'accuratezza delle previsioni di PolyPhen in modo sistematico, ma in generale ci aspettiamo che siano ragionevolmente accurate. La stima attuale è che, per un tasso di falsi positivi del 20%, PolyPhen raggiunge tassi di previsione positivi veri del 92% sul set di dati HumDiv e del 73% sul set di dati HumVar.[99]

I tools informatici differiscono tra loro per la sensibilità e per i parametri studiati e, inoltre, in base agli algoritmi utilizzati dai singoli software, possono essere tra loro contradditori.

Per questo motivo, in tutti quei casi in cui le loro predizioni non si rivelino sufficientemente attendibili, si procede con gli step successivi dello studio funzionale: lo studio in vitro e lo studio in vivo.

Il primo viene eseguito mettendo in coltura cellule ed andando a valutare su di esse se vi sia una conseguenza per la presenza delle mutazioni identificate.

Il secondo, ben più complesso, necessita invece dell'utilizzo di animali, come topi knock-out, per comprendere appieno le conseguenze di un'alterazione a livello amminoacidico e quindi della singola proteina.

Questi step, inoltre, richiedono un maggior dispendio in termini economici, ma anche di personale altamente qualificato e preparato specificatamente per questo tipo di studi.

CAPITOLO 7: RISULTATI

Sono stati analizzati 93 campioni provenienti da pazienti della medicina del territorio, dei reparti di cardiologia, di malattie rare, di immunologia clinica e da riscontri autoptici, prendendo in considerazione soggetti senza limiti di età.

Alcune analisi sono state richieste per effettuare degli approfondimenti, dal momento che talvolta le indagini cliniche e strumentali non portano ad una diagnosi certa, che può però essere agevolata dall'utilizzo dei test genetici. In altri casi, invece, a partire da una diagnosi, si richiede un'analisi nei parenti di primo grado del paziente, definito probando, per capire meglio la natura della mutazione riscontrata e stabilire quindi se vi siano le condizioni per impostare un follow-up anche su altri membri della stessa famiglia.

Dei 93 campioni prelevati, si sono ottenuti risultati completi per 45 (il 48,39%) tramite l'utilizzo di indagini molecolari, analizzando un pannello di 101 geni; gli altri campioni sono ancora in fase di studio.

Lo scopo di queste indagini è stato quello di classificare una mutazione come benigna o come patogenica, ma in alcuni casi vi sono dei fattori che concorrono alla definizione di mutazioni come di incerto significato (VUS), per le quali verranno previsti poi ulteriori approfondimenti, come descritto nelle pagine precedenti.

SOSPETTO CLINICO	WILD TYPE o BENIGNE	VUS	CERTAMENTE PATOGENICHE
20 LQTS	6	9	5
8 BRUGADA	3	4	1
3 CPVT (tachicardia polimorfa)	2	1	
1 HCM (cardiomiopatia ipertrofica)			1
3 DCM (cardiomiopatia dilatativa)		1	2
2 DILATAZIONE AORTICA		2	
1 BLOCCO ATRIO-VENTRICOLARE		1	
3 MORTE IMPROVVISA	1	1	1
2 SINDROME SOTOS	2		
1 NON COMPATTAZIONE VENTRICOLARE			1
1 MIOCARDIO NON COMPATTO		1	
	14	**20**	**11**

Tabella 13. *Risultati delle indagini molecolari condotte nell'Istituto di Medicina Legale*

Come si evince dalla tabella, dei 45 campioni analizzati sono stati riscontrati:

- 24,45% casi patogenetici,
- 44,45% VUS,
- 31,10% wild type o mutazioni benigne.

L'aspetto interessante è la percentuale piuttosto elevata di varianti di incerto significato (VUS), quelle cioè che dovranno essere ulteriormente approfondite con l'utilizzo di studi funzionali prima di poter dare una risposta di diagnosi o di esclusione di patologia.

Di seguito si riportano degli esempi di varianti VUS tratti dai campioni analizzati al fine di spiegare come si arriva ad una simile definizione:

- **Esempio 1**. Sospetta cardiomiopatia primaria genetica in soggetto morto improvvisamente.

  Presenza della variante SCN5A:NM_198056:exon17:c.G2989A:p.A997T in eterozigosi.

  Interpretazione del risultato: si tratta di una sostituzione missenso, conosciuta in letteratura e considerata di incerto significato clinico (VUS) per sindrome di Brugada per i seguenti criteri:

  - PM1: presenza della variante in un hot-spot mutazionale di 61 bp del gene (canale del sodio) in cui il 63% delle varianti è ritenuta patogenetica;
  - PM5: due varianti alternative Ala997Ser sono classificate come patogenetiche da Uniprot e HGMD;
  - PP2: il 78% delle varianti missenso non-VUS del gene è patogenetico;
  - BP4: supporto all'ipotesi di effetto deleterio in 2 tools di predizione sull'effetto della variante (M-CAP, MVP), e 9 predicono l'effetto benigno (DANN, DEOGEN2, EIGEN, FAHTMM-MKL, MutationAssessor, MutationTaster, PrimateAl, REVEL, SIFT); CADD Phred = 0,001; posizione nucleotidica non conservata.

- **Esempio 2**. Sospetta canalopatia

  Presenza della variante CACNA1C:NM_199460:exon2:c:C101T:p.434V in eterozigosi.

  Interpretazione del risultato: si tratta di una sostituzione missenso, conosciuta in letteratura ma di incerto significato clinico (VUS) per i seguenti criteri:

  - PP2: variante missenso in un gene in cui questo tipo di mutazioni causa frequentemente la malattia (S. di Brugada, S. di Timothy);
  - PP3: supporto all'ipotesi di effetto deleterio in 4 tools di predizione sull'effetto della variante (DANN, FATHMM-MKL, M-CAP, SIFT);
  - BS2: variante osservata in individui sani adulti (freq. 0,0003 in GnomAD exomes SA).

- **Esempio 3**. Tachicardia ventricolare polimorfa catecolaminergica (CPVT)

  Presenza della variante RYR2:NM_001035:exon60:c.T8767G:p.Y2923D in eterozigosi.

  Interpretazione del risultato: si tratta di una sostituzione missenso, non conosciuta in letteratura e classificabile come variante di incerto significato clinico (VUS) per CPVT per i seguenti criteri:

  - PM2: variante non presente nei database GnomAD exomes/genomes;
  - PP3: 10 tools computazionali supportano la predizione di effetto deleterio della variante (DANN, DEOGEN2, EIGEN, FATHMM-MKL, M-CAP, MutationAssessor, MutationTaster, PrimateAl, REVEL, SIFT) e 1 supporta la predizione di effetto benigno (MVP); CADD Phred=29,7.

- **Esempio 4**. Sospetta sindrome del QT lungo

  Risultato dell'analisi

  1) Presenza della variante SCN5A:NM_198056:exon23:c.A41013G:p.E1368G in eterozigosi.

     Interpretazione del risultato: si tratta di una sostituzione missenso, non conosciuta in letteratura e classificabile come variante di incerto significato clinico (VUS) per LQTS per i seguenti criteri:

     - PP2: variante presente in un gene (subunità alfa del canale del sodio) in cui l'80% delle varianti missenso è causativo di canalopatie;
     - PM2: variante assente dai controlli nei database GnomAD exomes e genomes;
     - BP4: 5 tools computazionali supportano la predizione di effetto deleterio della variante (DANN, DEOGEN2, M-CAP, MutationTaster, REVEL) e 5 tools supportano la predizione di effetto benigno (EIGEN, FATHMM-MKL, MutationAssessor, PrimateAl, SIFT); CADD Phred=24,1.

  2) Presenza variante MYH11:NM_001040114:exon39:c.G5448C:p.K1816N in eterozigosi.

     Interpretazione del risultato: si tratta di una sostituzione missenso, non conosciuta in letteratura e classificabile come variante di incerto significato clinico (VUS) per LQTS per i seguenti criteri:

     - PM2: variante assente dai controlli nei database GnomAD exomes e genomes;
     - BP4: 8 tools computazionali supportano la predizione di effetto deleterio della variante (DANN, DEOGEN2, EIGEN, FATHMM-MKL, M-CAP, MutationAssessor, MutationTaster, REVEL) e 2 tools supportano la predizione di effetto benigno (PrimateAl, SIFT); CADD Phred=24,7;

- BP1: Presente in un gene (catena pesante della miosina) in cui l'81.2% delle varianti missenso è benigno.

- **Esempio 5**. Morte cardiaca improvvisa

  Presenza della variante MYBPC3:NM_000256:exon15:c.C1363T:p.L455F in eterozigosi.

  Interpretazione del risultato: si tratta di una sostituzione missenso nel gene della proteina sarcomerica che lega la miosina, isoforma cardiaca, le cui mutazioni sono considerate associate a cardiomiopatia ipertrofica, dilatativa e morte cardiaca improvvisa. La variante osservata è conosciuta in letteratura[100] e classificabile come probabilmente patogenetica per i seguenti criteri:

  - PM1: variante presente in un hot-spot in cui il 90% delle varianti è patogenetico;
  - PM2: variante rara nel database di frequenza GnomAD exomes (f = 0.0016%) e assente in GnomAD genomes;
  - PP2: l'82% delle varianti missenso non-VUS (variante di incerto significato) in questo gene è patogenetico;
  - PP3: 7 tools computazionali supportano la predizione di effetto deleterio della variante (DANN, FATHMM-MKL, M-CAP, MVP, MutationTaster, REVEL, SIFT) e 4 supportano la predizione di effetto benigno (DEOGEN2, EIGEN, MutationAssessor, PrimateAl).

Prestando attenzione all'esito dato dai tools computazionali, si può vedere come questi possano portare a sbilanciare alcune valutazioni: nel primo caso, se non fossero stati considerati, la condizione sarebbe stata valutata come patogenica tenendo conto dell'alta percentuale di varianti riscontrate nello stesso gene e ritenute patogeniche, ma 9 tools ne predicono la benignità, a scapito di soli 2 tools predittivi di patologia. Per questo motivo, non è possibile fare valutazione di condizione certamente patogenica e si è quindi classificata la mutazione come di incerto significato (VUS).

Nel secondo esempio, invece, i tools sono concordi nel predire un effetto deleterio della mutazione, che viene considerata VUS poiché la stessa mutazione è stata riscontrata in una coorte di soggetti adulti sani.

Il quarto esempio è emblematico di una situazione nella quale i tools non siano stati in grado di aiutare nello spostare la diagnosi verso la benignità o la malignità poiché un ugual numero di tools computazionali classificano la mutazione come benigna e come patogenica.

7.1 Approfondimento caso medico-legale

Per poter meglio comprendere il ruolo dei tools computazionali si è deciso di fare un approfondimento circa i risultati ottenuti nell'ultimo caso riportato, inerente un'indagine condotta dall'Istituto di Medicina Legale per accertamenti sulla morte cardiaca improvvisa di un uomo di 33 anni, senza storia clinica per patologie cardiache note.

Si tratta di un uomo rinvenuto cadavere nella sua auto dopo una serata trascorsa con amici, lo stesso giorno in cui, durante una partita di calcio, si era fatto sostituire poiché molto affaticato. Dai suoi conoscenti era poi emerso che da qualche tempo l'uomo riferiva dolore al petto e ad un braccio, ma senza che fossero state rilevate patologie agli accertamenti cui si era sottoposto.

Dalle indagini autoptiche è stato avanzato il sospetto di una condizione aritmica sottostante come causa del decesso, in associazione ad un'alterazione del ritmo e/o della conduzione cardiaca; si è pertanto ritenuto opportuno procedere con test genetici per meglio comprendere ed inquadrare il caso.

Dalle indagini molecolari è effettivamente emersa una mutazione meritevole di rivalutazione, a carico del gene MYBPC3, localizzato sul cromosoma 11.

Questo, come mostrato di seguito (Figura 21), codifica per una proteina fondamentale per il miocardio, coinvolta nel legame tra miosina ed actina, agendo con regolazione negativa della contrazione.

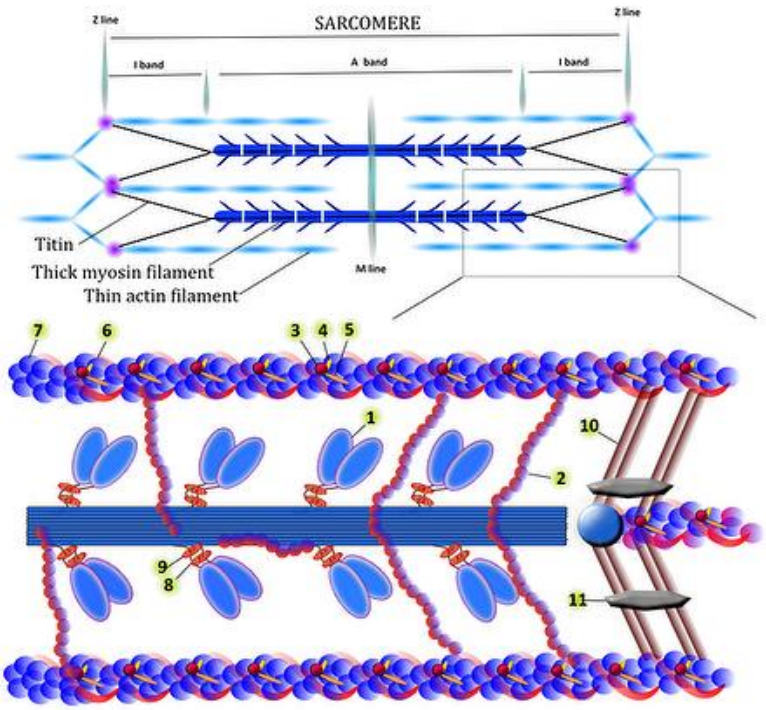


***Figura 21**. Presentazione schematica di un sarcomero con le mutazioni proteiche causa di HCM in accordo con la loro frequenza.*

1: MYH7: Beta Myosin Heavy Chain (40%); 2: MYBPC3: Myosin Binding protein C3 (40%); 3: TNNT2: Troponin T2 (5%); 4: TNNI: Troponin I 3 (5%); 5: TNNC1: Troponin C1 (raro); 6: TPM: Tropomyosin (2%); 7: ACTC: Alpha Actin (raro); 8: MYL2: Myosin Light Chain 2 (raro); 9: MYL 3: Myosin Light Chain 3 (1%); 10: ACTN2: Alpha 2 Actinin (raro); 11: MYOZ 2: Myozenin 2 (raro)

L'esito della variante studiata è quello di una mutazione missenso con formazione di una proteina tronca e, di conseguenza, con perdita di funzione, che rappresenta circa il 40% delle mutazioni proteiche responsabili di alterazioni a carico dei sarcomeri.

Le mutazioni riscontrate a carico di questo gene fino ad oggi note sono nella stragrande maggioranza dei casi associate a patologia ma, se andiamo a vedere l'esito dei tools computazionali, questi non sono tra loro concordi.

TOOL INFORMATICO DI PREDIZIONE	PUNTEGGIO O RISULTATO (per la mutazione studiata)	INTERPRETAZIONE DEI RISULTATI
Mutation Taster	Disease Causing	Il risultato è calcolato da un classificatore bayesiano e deriva dall'elaborazione di informazioni multiple (conservazione, sito di splicing, mRNA, caratteristiche della proteina e caratteristiche regolatorie)
SIFT	Damaging: 0.05	Viene utilizzato per predire il significato patogenetico di varianti non sinonime basandosi sull'omologia di sequenze omologhe raccolte tramite PSI-BLAST. Il punteggio varia da 0 a 1, con valori < 0.06 predetti come non tollerabili.
Mutation Assessor	Low Impact	Sfrutta il confronto tra le proteine omologhe per valutare la presenza di mutazioni patogenetiche. Il "rank-score" varia da 0 a 1, con 1 che indica varianti sicuramente patogenetiche.
Polyphen-VAR	Possibly Damaging	Il punteggio rappresenta la probabilità che una sostituzione sia dannosa. Il punteggio PolyPhen varia da 0,0 (tollerato) a 1,0 (deleterio). Il punteggio può essere interpretato come segue:

		Da 0,0 a 0,15 - Le varianti con punteggi in questo intervallo si prevede che siano benigne. Da 0,15 a 0,85 - Le varianti con punteggi in questo intervallo possono essere dannose. Da 0,85 a 1,0 - Si prevede che le varianti con punteggi in questo intervallo siano dannose con maggiore sicurezza.
CADD	23.9 (PHRED-score)	Il punteggio finale "PHRED-score" (scaled C-score) va da 1 a 99 e aumenta consensualmente alla probabilità che la mutazione sia patogenetica. Uno score di 30 è molto probabilmente patogenetico perché rappresenta lo 0.1% delle mutazioni con score 20 che sono a loro volta più probabilmente patogenetiche rispetto a quella con score 10.
FATHMM	Tolerated	I punteggi approssimativamente uguali a 0 indicano che non c'è cambio significativo nella sottostante struttura amminoacidica, al contrario valori inferiori a 0 indicano che una mutazione è probabilmente benigna e che potrebbe essere osservata frequentemente nella popolazione wild type.
Grantham	Conservative	Fa una previsione dell'effetto delle sostituzioni tra amminoacidi in base alle proprietà chimiche, tra cui polarità e volume molecolare, caratterizzate in classi di crescente diversità chimica: conservativa (0-50), moderatamente conservativa (51-100), moderatamente radicale (101- 150) o radicale (≥151).

Tabella 14. *Interpretazione dei risultati ottenuti con i tools informatici*

SUPPORTING EVIDENCE

Population data:

PM2 - Variant frequency in ExAC is < 0.000040.

ExAC - All (112128)	ExAC - American (10576)	1KG (~5000)	ESP (~13000)	Healthy volunteers - UK (625)	Healthy volunteers - Singapore (758)	HCM CGG (327)	HCM ACGV (6179)	DCM CGG (877)	DCM ACGV (1161)
4	1	-	-	-	-	0	-	0	-

Population controls (light green, allele counts), Healthy controls (dark green, sample counts), Disease cases (red, sample counts). See Appendix for details.

Computational data:

Functional data:

PP2 - Missense variants in MYBPC3, when found in a case, are likely to be pathogenic (EF*=0.82).

ADDITIONAL EVIDENCE

Computational algorithms have conflicting predictions on the effect of this mutation on protein function:

SIFT	*MutationTaster*	*MutationAssessor*	*Polyphen-VAR*	*LRT*	*FATHMM*	*Grantham*	*CADD*
Damaging	Disease-causing	Low impact	Possibly Damaging	-	Tolerated	Conservative	23.9

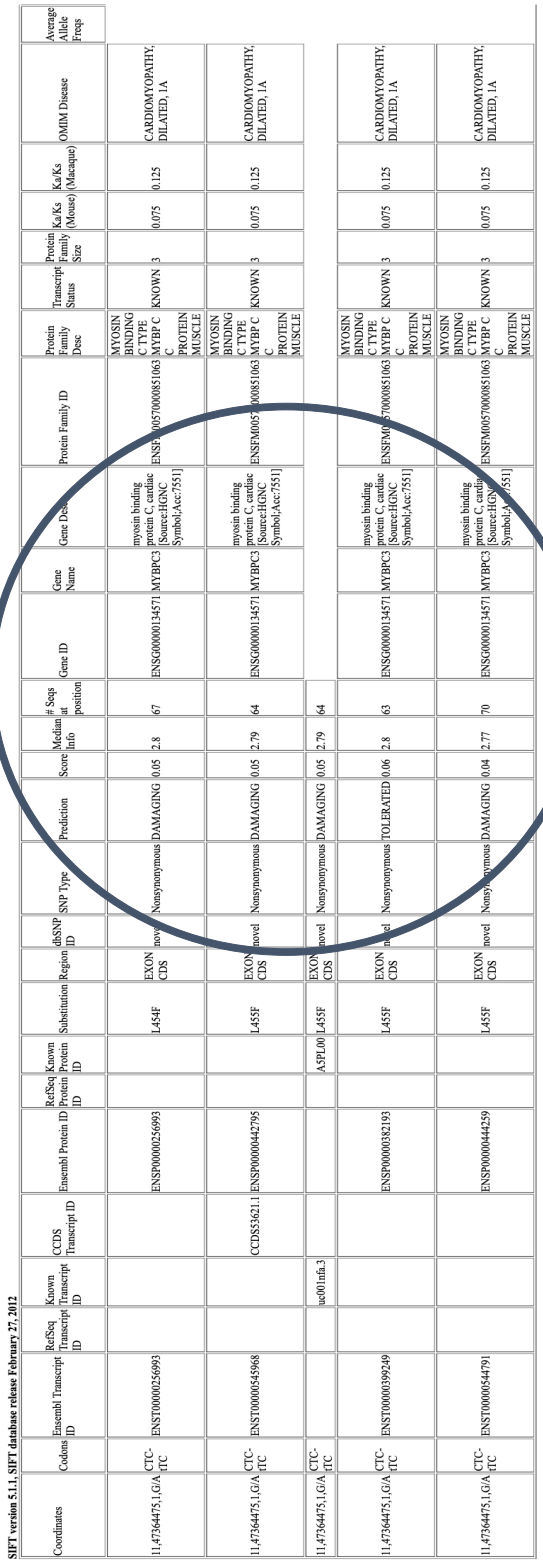
***Figura 22**. Evidenze a supporto della mutazione MYBPC3*

Amino acid comparison		
Trait	Leu (L)	Phe (F)
Amino acid name	leucine	phenylalanine
Polarity/charge	non-polar	non-polar
pH	neutral	neutral
Residue weight	113	147
Hydrophobicity score	3.8	2.8
Hydrophilicity score	-1.8	-2.5
Secondary structure propensity	strong α former β former	α former β former
Grantham difference	22	
MutPred likelihood of being deleterious	NONE CALCULATED	

dbNSFP3.5 predictions	
PolyPhen2 prediction	Possibly damaging
SIFT prediction	Damaging
LRT prediction	Not available
MutationTaster prediction	Disease causing
MutationAssessor prediction	Low impact
FATHMM	Tolerated
fathmm-MKL	Damaging
M-CAP	Damaging
CADD The larger the score the more likely the SNP is damaging (PHRED-like)	23.9
MetaSVM	Damaging
MetaLR	Damaging
PhyloP 20way The larger the score, the more conserved the site (max 1.199000).	1.038000
PhyloP 100way The larger the score, the more conserved the site (max 10.003000).	2.133000
GERP RS The larger the score, the more conserved the site (max 6.17).	3.48
ExAC	0.0000331/4
1000 Genomes	No data
gnomAD	0.0000124/3
Interpro domain	Immunoglobulin-like domain ↑↓

***Figura 23**. Sommario Tools predittivi applicati alla mutazione MYBPC3*

Uno dei tools più affidabili, rilasciato nel febbraio del 2012, è SIFT. Viene considerato tra i tools più attendibili perché al suo interno sono raccolte molteplici informazioni: il database, infatti, è in grado di confrontare tra loro numerosi criteri prima di effettuare la predizione.
Dalla data del suo rilascio, inoltre, sono stati effettuati continui aggiornamenti del software e questo ha permesso di raggiungere gradi di sensibilità e di predittività piuttosto elevati.
Il tool richiede di inserire la sequenza da analizzare in formato FASTA, dove, dopo aver scritto una descrizione a riga singola del gene mutato, seguono delle righe con i dati di sequenza. Dopo un'elaborazione di alcuni minuti, il software restituisce un pannello di informazioni di cui si riporta di seguito l'esito ricavato nell'esempio trattato.



SIFT version 5.1.1, SIFT database release February 27, 2012

Coordinates	Codons	Ensembl Transcript ID	RefSeq Transcript ID	Known Transcript ID	CCDS Transcript ID	Ensembl Protein ID	RefSeq Protein ID	Known Protein ID	Substitution	Region	dbSNP ID	SNP Type	Prediction	Score	Median Info	# Seqs at position	Gene ID	Gene Name	Gene Desc	Protein Family ID	Protein Family Desc	Transcript Status	Protein Family Size	Ka/Ks (Mouse)	Ka/Ks (Macaque)	OMIM Disease	Average Allele Freqs
11,47364475,1,G/A	CTC-TTC	ENST00000256993				ENSP00000256993			L454F	EXON CDS	novel	Nonsynonymous	DAMAGING	0.05	2.8	67	ENSG00000134571	MYBPC3	myosin binding protein C, cardiac [Source:HGNC Symbol;Acc:7551]	ENSFM00570000851063	MYOSIN BINDING C TYPE MYBP C C PROTEIN MUSCLE	KNOWN	3	0.075	0.125	CARDIOMYOPATHY, DILATED, 1A	
11,47364475,1,G/A	CTC-TTC	ENST00000545968			CCDS53621	ENSP00000442795			L455F	EXON CDS	novel	Nonsynonymous	DAMAGING	0.05	2.79	64	ENSG00000134571	MYBPC3	myosin binding protein C, cardiac [Source:HGNC Symbol;Acc:7551]	ENSFM00570000851063	MYOSIN BINDING C TYPE MYBP C C PROTEIN MUSCLE	KNOWN	3	0.075	0.125	CARDIOMYOPATHY, DILATED, 1A	
11,47364475,1,G/A	CTC-TTC			uc001nfa.3				A5PL00	L455F	EXON CDS	novel	Nonsynonymous	DAMAGING	0.05	2.79	64											
11,47364475,1,G/A	CTC-TTC	ENST00000399249				ENSP00000382193			L455F	EXON CDS	rs1	Nonsynonymous	TOLERATED	0.06	2.8	63	ENSG00000134571	MYBPC3	myosin binding protein C, cardiac [Source:HGNC Symbol;Acc:7551]	ENSFM00570000851063	MYOSIN BINDING C TYPE MYBP C C PROTEIN MUSCLE	KNOWN	3	0.075	0.125	CARDIOMYOPATHY, DILATED, 1A	
11,47364475,1,G/A	CTC-TTC	ENST00000544791				ENSP00000444259			L455F	EXON CDS	novel	Nonsynonymous	DAMAGING	0.04	2.77	70	ENSG00000134571	MYBPC3	myosin binding protein C, cardiac [Source:HGNC Symbol;Acc:7551]	ENSFM00570000851063	MYOSIN BINDING C TYPE MYBP C C PROTEIN MUSCLE	KNOWN	3	0.075	0.125	CARDIOMYOPATHY, DILATED, 1A	

SNP Type	Prediction	Score	Median Info	# Seqs at position	Gene ID	Gene Name
Nonsynonymous	DAMAGING	0.05	2.8	67	ENSG00000134571	MYBPC3
Nonsynonymous	DAMAGING	0.05	2.79	64	ENSG00000134571	MYBPC3
Nonsynonymous	DAMAGING	0.05	2.79	64		
Nonsynonymous	TOLERATED	0.06	2.8	63	ENSG00000134571	MYBPC3
Nonsynonymous	DAMAGING	0.04	2.77	70	ENSG00000134571	MYBPC3

***Figura 24**. Schermata risultati software SIFT*

Come si può vedere SIFT predice in definitiva un effetto dannoso della mutazione, ma non tutti i tools interpellati giungono alla stessa conclusione: l'assenza di una concordanza netta tra i tools utilizzati non permette di fare una diagnosi certa di mutazione patogenica pur essendo rispettati gli altri criteri richiesti dall'ACMG.

Per quanto riguarda gli altri criteri richiesti dall'ACMG, si è innanzitutto identificata la regione hotspot, cioè i minisatelliti della ricombinazione omologa, zone di DNA all'interno delle quali la frequenza di ricombinazione è più elevata rispetto alla media, ed è emerso che il 90% delle varianti è patogenetico (PM1).

Un secondo criterio prevede di consultare il Genome Aggregation Database (gnomAD), una risorsa sviluppata da un gruppo internazionale di ricercatori, con l'obiettivo di aggregare e armonizzare i dati di sequenziamento sia dell'esoma sia del genoma da un'ampia varietà di progetti di sequenziamento su larga scala e rendere disponibili dati di sintesi alla comunità scientifica[101] (PM2).

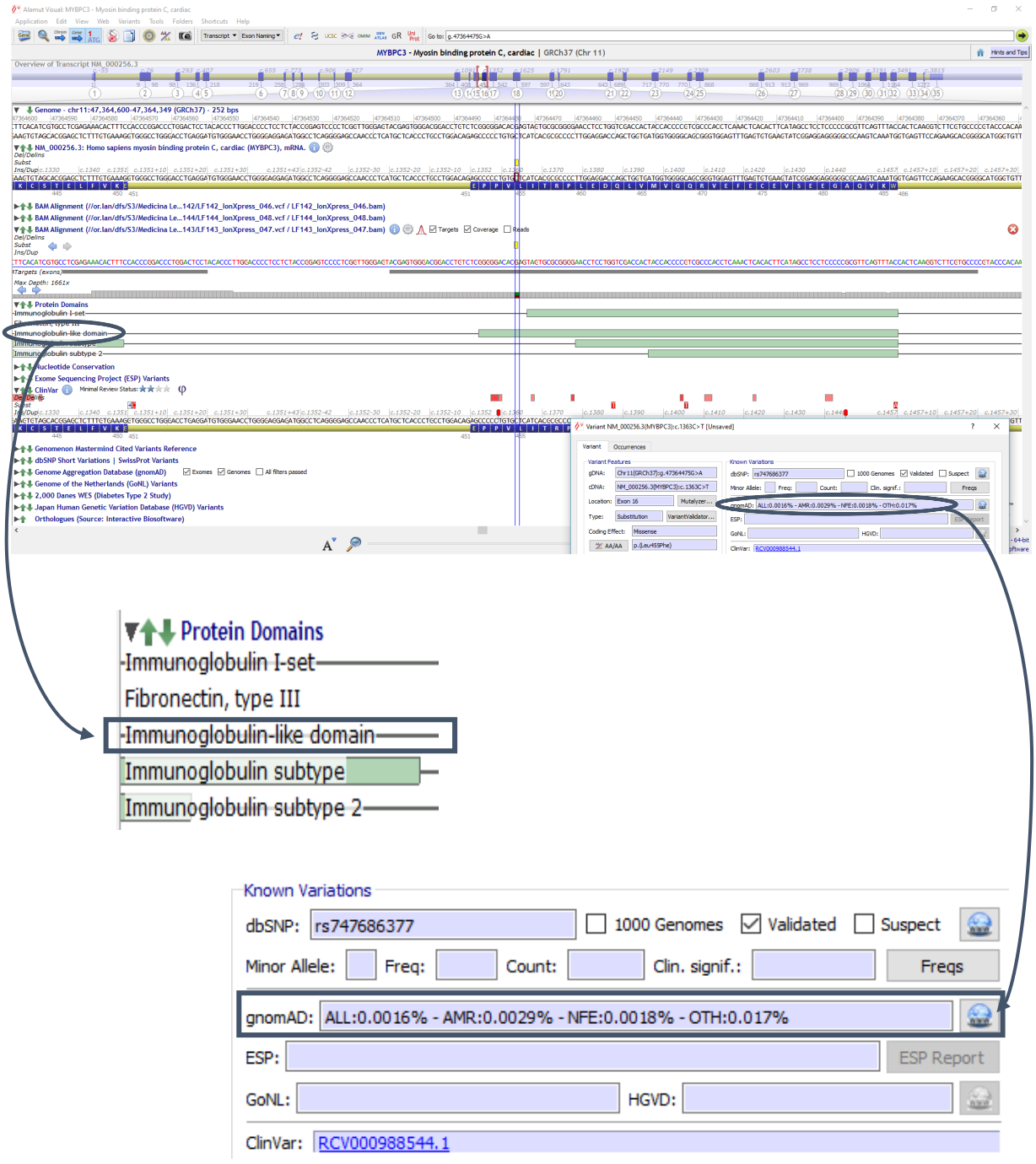
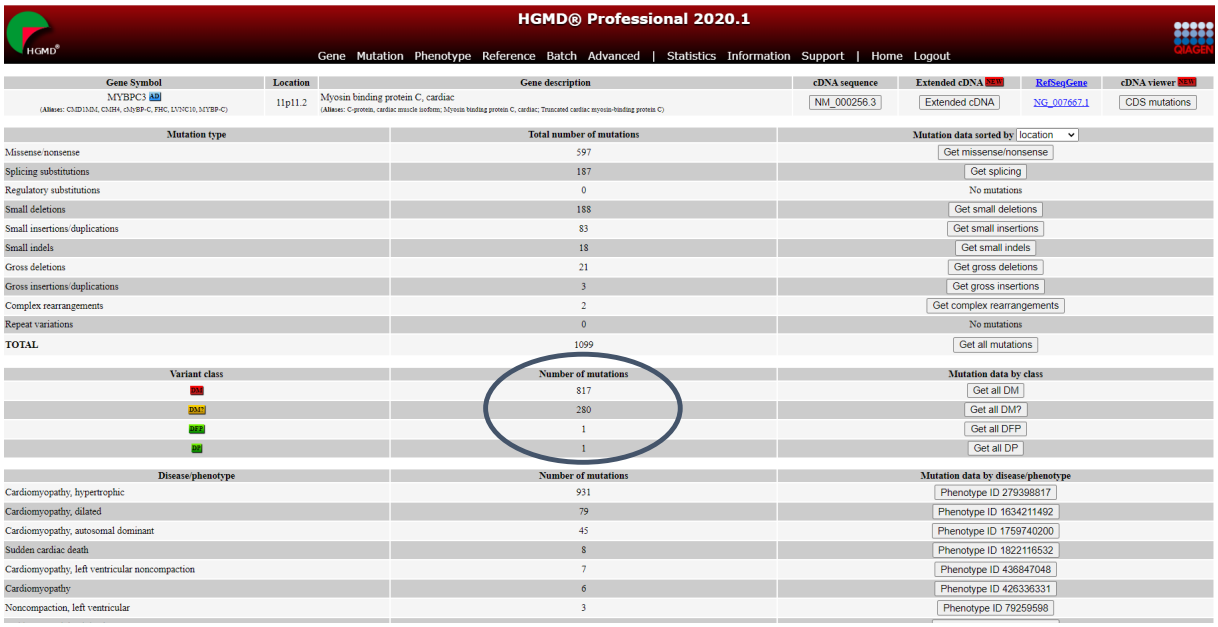


***Figura 25**. Alamut Visual applicato a MYBPC3*

Infine, è stato valutato il numero di varianti missenso non-VUS classificate come patogenetiche sul gene MYBPC3 verificando una positività in tal senso in 817 varianti su 1099 (75%) (PP2).



HGMD® Professional 2020.1

Gene Mutation Phenotype Reference Batch Advanced | Statistics Information Support | Home Logout

Gene Symbol	Location	Gene description	cDNA sequence	Extended cDNA	RefSeqGene	cDNA viewer
MYBPC3	11p11.2	Myosin binding protein C, cardiac	NM_000256.3	Extended cDNA	NG_007667.1	CDS mutations

Mutation type	Total number of mutations	Mutation data sorted by location
Missense/nonsense	597	Get missense/nonsense
Splicing substitutions	187	Get splicing
Regulatory substitutions	0	No mutations
Small deletions	188	Get small deletions
Small insertions/duplications	83	Get small insertions
Small indels	18	Get small indels
Gross deletions	21	Get gross deletions
Gross insertions/duplications	3	Get gross insertions
Complex rearrangements	2	Get complex rearrangements
Repeat variations	0	No mutations
TOTAL	1099	Get all mutations

Variant class	Number of mutations	Mutation data by class
DM	817	Get all DM
DM?	280	Get all DM?
DFP	1	Get all DFP
DP	1	Get all DP

Disease/phenotype	Number of mutations	Mutation data by disease/phenotype
Cardiomyopathy, hypertrophic	931	Phenotype ID 279398817
Cardiomyopathy, dilated	79	Phenotype ID 1634211492
Cardiomyopathy, autosomal dominant	45	Phenotype ID 1759740200
Sudden cardiac death	8	Phenotype ID 1822116532
Cardiomyopathy, left ventricular noncompaction	7	Phenotype ID 436847048
Cardiomyopathy	6	Phenotype ID 426336331
Noncompaction, left ventricular	3	Phenotype ID 79259598

***Figura 26**. Mutazioni note per essere patogeniche su MYBPC3*

In base alle informazioni ricavate, quindi, non è stato possibile sbilanciarsi nel fare una diagnosi e, come in tutti i casi in cui i tools da soli non siano dirimenti, si può procedere con test di segregazione sui familiari di primo grado e con test funzionali più avanzati.

CAPITOLO 8: DISCUSSIONE

L'interesse nell'utilizzo dei tools computazionali nasce dall'esigenza da parte dei clinici di ottenere risposte in tempi rapidi al fine di poter impostare il più precocemente possibile un follow up nei pazienti a rischio di morte cardiaca improvvisa, evento evitabile impostando le giuste terapie preventive.

Al tempo stesso, riscontrare un'alterazione mutazionale in un soggetto deceduto permette di effettuare test di segregazione ed approfondimenti sui familiari, permettendo di mettere in atto tutte quelle strategie che sono in grado di evitare una morte improvvisa.

Come è stato fin qui esposto, il ruolo cardine di questi software lo si trova in tutti quei casi in cui si ha il riscontro di mutazioni non ben inquadrabili clinicamente; i tools inoltre assumono un ruolo importante nello spostare l'ago della bilancia tra condizioni a più probabile natura benigna o patogenica.

Fermo restando il ruolo primario ad oggi svolto dalla pratica clinica e dalle metodiche strumentali nel fare diagnosi di patologie cardiache, i test genetici possono essere di grande ausilio quando la clinica da sola non sia dirimente, avendo una sensibilità diagnostica sempre più elevata, che si attesta oggi in un range variabile dal 20% nella cardiomiopatia ipertrofica, del 30% nella sindrome di Brugada, fino a superare il 70% nella sindrome del QT lungo.

Qualora vi sia una concordanza netta tra i tools consultati, il loro utilizzo risulta essere di fondamentale aiuto, rappresentando una prova a sostegno dell'una o dell'altra natura della mutazione, per effettuare diagnosi nel minor tempo possibile con una sensibilità che si è visto essere in aumento negli ultimi anni.

D'altra parte, però, la continua introduzione e l'impiego di tools di più recente creazione, con un numero minore di informazioni a cui possono attingere, riduce questa sensibilità, portando ad una minore concordanza e quindi molto spesso alla necessità di proseguire con lo studio

delle mutazioni tramite metodiche successive e quindi con le colture cellulari e con gli studi in vivo utilizzando topi knockout.

Tutto questo è ancor più vero tenendo conto della sensibilità ancora non troppo elevata soprattutto nei confronti di alcuni tipi di cardiomiopatie.

L'analisi genetica, ogni volta che sia possibile, andrebbe sempre suggerita a questo tipo di pazienti come metodo di conferma diagnostica, pur consapevoli del fatto che non è sempre facile accedere a questo metodo di indagine, dal momento che non ovunque sono presenti laboratori di Genetica molecolare.

Per superare il problema dell'attuale concordanza non sempre verificata con l'utilizzo dei tools computazionali, inoltre, sarebbe utile associarvi in maniera sistematica lo step dello studio in vitro utilizzando colture cellulari idonee al singolo caso analizzato.[102]

I test di segregazione andranno richiesti sia nel caso in cui rimangano dei sospetti dopo le prime indagini molecolari, sia di fronte ad una patogenicità appurata, al fine di andare ad identificare gli altri membri della famiglia affetti dalla medesima alterazione.

CAPITOLO 9: CONCLUSIONI

Le cardiopatie analizzate ed approfondite in questa tesi sono patologie per le quali il test genetico dovrebbe essere mandatorio quando sospettate, trattandosi di patologie per le quali si è visto dall'esperienza clinica che in una percentuale non trascurabile di casi, con la sola valutazione obiettiva e strumentale, la patologia rimarrebbe misconosciuta, esponendo il soggetto ad un rischio maggiore di eventi cardiovascolari importanti come la morte cardiaca improvvisa.

In questo studio è stato descritto in maniera approfondita l'iter che viene seguito per compiere l'analisi dei geni per i quali si conoscono relazioni con le suddette patologie, in modo da migliorare la sensibilità delle valutazioni cliniche.

Nonostante si tratti di un metodo oggi molto accurato, però, quello che è emerso dallo studio è che in una percentuale elevata di casi le mutazioni riscontrate nei singoli pazienti non sono presenti nei database, ovvero sono mutazioni spesso mai incontrate prima, e si procede in questi casi alla valutazione tramite tools computazionali. Non sempre, però, tali software sono esaustivi se utilizzati da soli. Quindi quando mostrano delle discordanze si renderà necessario passare a valutazioni più approfondite con gli step successivi degli studi funzionali.

Quello che si è visto essere di fondamentale importanza per aumentare la sensibilità dei test genetici è, dato un sospetto di patogenicità, richiedere un test Sanger sui parenti di primo grado, valutandone sempre la storia clinica ed eventuali sintomi presenti in altri soggetti della famiglia riconducibili alla medesima patologia. Trovare, infatti, una mutazione in altri membri della medesima famiglia, senza che vi sia associato alcun tipo di sintomo o di storia clinica, fa sempre propendere per una valutazione di mutazione di tipo benigno e può quindi essere utilizzato come ulteriore criterio di indagine.

Al maggior numero di informazioni raccolte tramite anamnesi, infatti, si associano una sensibilità e affidabilità sempre più elevate dei test genetici, comprese le predizioni effettuate dai tools computazionali.

In conclusione, è possibile affermare che ad oggi, nonostante i netti miglioramenti riscontrati nell'utilizzo delle tecniche predittive, per poter giungere ad una diagnosi di certezza di patogenicità o di benignità della mutazione l'unico metodo comprovato sembra ancora essere quello dell'utilizzo degli studi funzionali in vitro ed in vivo.

BIBLIOGRAFIA

1. Malattie del cuore di Braunwald.
2. McKenna, W. J., Maron, B. J. & Thiene, G. Classification, Epidemiology, and Global Burden of Cardiomyopathies. *Circ. Res.* **121**, 722–730 (2017).
3. Ruan, Y., Liu, N. & Priori, S. G. Sodium channel mutations and arrhythmias. *Nat. Rev. Cardiol.* **6**, 337–348 (2009).
4. Shimizu, W. Clinical Impact of Genetic Studies in Lethal Inherited Cardiac Arrhythmias. *Circ. J.* **72**, 1926–1936 (2008).
5. Schwartz, P. J. *et al.* Prevalence of the Congenital Long QT Syndrome. *Circulation* **120**, 1761–1767 (2009).
6. Ilan Goldenberg, MD, Arthur J. Moss, MD. Long QT Syndrome. *Journal of the American College of Cardiology*.
7. Brugada, P. & Brugada, J. Right bundle branch block, persistent ST segment elevation and sudden cardiac death: A distinct clinical and electrocardiographic syndrome. *J. Am. Coll. Cardiol.* **20**, 1391–1396 (1992).
8. Brugada, P., Benito, B., Brugada, R. & Brugada, J. Brugada Syndrome: Update 2009. 21 (2009).
9. Ali, M. Brugada Syndrome: Current Practices in Diagnosis, Prognosis, and Treatment. *Int. Cardiovasc. Forum J.* **9**, (2017).
10. Antzelevitch, C. *et al.* Brugada Syndrome: Report of the Second Consensus Conference: *Endorsed by the Heart Rhythm Society and the European Heart Rhythm Association*. *Circulation* **111**, 659–670 (2005).
11. Moss, A. J. *et al.* Effectiveness and Limitations of β-Blocker Therapy in Congenital Long-QT Syndrome. *Circulation* **101**, 616–623 (2000).
12. Priori, S. G., Napolitano, C. & Gasparini, M. Natural history of Brugada Syndrome. Insights for risk stratification and management. *ACC Curr. J. Rev.* **11**, 76 (2002).
13. Moss, A. J. *et al.* ECG T-Wave Patterns in Genetically Distinct Forms of the Hereditary Long QT Syndrome. *Circulation* **92**, 2929–2934 (1995).
14. Jervell & Lange-Nielsen 1957, Vincent et al 1991, Swan et al 1998, Horner et al 2011, Sy et al 2011.pdf.
15. Zhang, L. *et al.* Spectrum of ST-T–Wave Patterns and Repolarization Parameters in Congenital Long-QT Syndrome: ECG Findings Identify Genotypes. *Circulation* **102**, 2849–2855 (2000).
16. Epinephrine unmasks latent mutation carriers with LQT1 form of congenital long-QT syndrome | Elsevier Enhanced Reader. https://reader.elsevier.com/reader/sd/pii/S0735109702028504?token=4DCCD96B516AE91B6CC6BA0F76ECC29B6D54C2561C624A6A949F079FCE183AD58DD12FE6255477AB6DC5941B54564719 doi:10.1016/S0735-1097(02)02850-4.
17. Phenotype guided characterization and molecular analysis of Indian patients with long QT syndromes | Elsevier Enhanced Reader.
18. Schwartz, P. J., Moss, A. J., Vincent, G. M. & Crampton, R. S. Diagnostic criteria for the long QT syndrome. An update. *Circulation* **88**, 782–784 (1993).
19. Genetic testing for LQTS.pdf.
20. Priori, S. G., Napolitano, C. & Schwartz, P. J. Low Penetrance in the Long-QT Syndrome: Clinical Impact. *Circulation* **99**, 529–533 (1999).
21. Ferlini, A. *et al.* Il sequenziamento del DNA di nuova generazione: indicazioni per l'impiego clinico. 32 (2016).

22. Splawski, I. *et al.* Spectrum of Mutations in Long-QT Syndrome Genes: *KVLQT1* , *HERG* , *SCN5A* , *KCNE1* , and *KCNE2*. *Circulation* **102**, 1178–1185 (2000).
23. Vatta, M. Basi molecolari delle aritmie nei pazienti con cardiomiopatie genetiche: quando il citoscheletro incontra i canali ionici. **11**, 7 (2010).
24. Sanguinetti, M. C. Dysfunction of Delayed Rectifier Potassium Channels in an Inherited Cardiac Arrhythmia. *Ann. N. Y. Acad. Sci.* **868**, 406–413 (1999).
25. Abbott, G. W. *et al.* MiRP1 Forms I Kr Potassium Channels with HERG and Is Associated with Cardiac Arrhythmia. *Cell* **97**, 175–187 (1999).
26. Bennett, P. B., Yazawa, K., Makita, N. & George, A. L. Molecular mechanism for an inherited cardiac arrhythmia. *Nature* **376**, 683–685 (1995).
27. Yong, S., Tian, X. & Wang, Q. LQT4 Gene: The "Missing" Ankyrin. *Mol. Interv.* **3**, 131–136 (2003).
28. Plaster, N. M. *et al.* Mutations in Kir2.1 Cause the Developmental and Episodic Electrical Phenotypes of Andersen's Syndrome. *Cell* **105**, 511–519 (2001).
29. Splawski, I. *et al.* CaV1.2 Calcium Channel Dysfunction Causes a Multisystem Disorder Including Arrhythmia and Autism. *Cell* **119**, 19–31 (2004).
30. Vatta, M. *et al.* Mutant Caveolin-3 Induces Persistent Late Sodium Current and Is Associated With Long-QT Syndrome. *Circulation* **114**, 2104–2112 (2006).
31. Medeiros-Domingo, A. *et al.* SCN4B-Encoded Sodium Channel β4 Subunit in Congenital Long-QT Syndrome. *Circulation* **116**, 134–142 (2007).
32. Kyndt, F. *et al.* Novel *SCN5A* Mutation Leading Either to Isolated Cardiac Conduction Defect or Brugada Syndrome in a Large French Family. *Circulation* **104**, 3081–3086 (2001).
33. Akai, J. *et al.* A novel *SCN5A* mutation associated with idiopathic ventricular fibrillation without typical ECG findings of Brugada syndrome. *FEBS Lett.* **479**, 29–34 (2000).
34. Juang, J.-M. J. & Horie, M. Genetics of Brugada syndrome. *J. Arrhythmia* **32**, 418–425 (2016).
35. Valdivia, C. A trafficking defective, Brugada syndrome-causing SCN5A mutation rescued by drugs. *Cardiovasc. Res.* **62**, 53–62 (2004).
36. Ackerman, M. J. *et al.* HRS/EHRA Expert Consensus Statement on the State of Genetic Testing for the Channelopathies and Cardiomyopathies. *Heart Rhythm* **8**, 1308–1339 (2011).
37. Ackerman, M. J., Marcou, C. A. & Tester, D. J. Personalized Medicine: Genetic Diagnosis for Inherited Cardiomyopathies/Channelopathies. *Rev. Esp. Cardiol. Engl. Ed.* **66**, 298–307 (2013).
38. Brieler, J., Breeden, M. A., Tucker, J., University, S. L. & Louis, S. Cardiomyopathy: An Overview. **96**, 8 (2017).
39. Harvey, W. & Leake, C. D. *Exercitatio anatomica de motu cordis et sanguinis in animalibus*. (Springfield, Ill. : Thomas, 1628).
40. Braunwald, E. Cardiomyopathies: An Overview. *Circ. Res.* **121**, 711–721 (2017).
41. Maron, B. J. & Maron, M. S. Hypertrophic cardiomyopathy. *The Lancet* **381**, 242–255 (2013).
42. Neri, G. & Genuardi, M. *Genetica umana e medica*. (Elsevier).
43. Maron, B. J. *et al.* Contemporary Definitions and Classification of the Cardiomyopathies: An American Heart Association Scientific Statement From the Council on Clinical Cardiology, Heart Failure and Transplantation Committee; Quality of Care and Outcomes Research and Functional Genomics and Translational Biology Interdisciplinary Working Groups; and Council on Epidemiology and Prevention. *Circulation* **113**, 1807–1816

(2006).
44. Zaidi, A. & Sharma, S. Arrhythmogenic right ventricular remodelling in endurance athletes: Pandora's box or Achilles' heel?: Figure 1. *Eur. Heart J.* **36**, 1955–1957 (2015).
45. Gimbe, F. Linee guida per la diagnosi e la terapia dello scompenso. *Best Pract.* **11**, 7 (2019).
46. Spirito, P., Seidman, C. E., McKenna, W. J. & Maron, B. J. The Management of Hypertrophic Cardiomyopathy. *N. Engl. J. Med.* **336**, 775–785 (1997).
47. Fifer, M. A. & Vlahakes, G. J. Management of Symptoms in Hypertrophic Cardiomyopathy. *Circulation* **117**, 429–439 (2008).
48. Hershberger, R. E., Morales, A. & Siegfried, J. D. Clinical and genetic issues in dilated cardiomyopathy: A review for genetics professionals. *Genet. Med. Off. J. Am. Coll. Med. Genet.* **12**, 655–667 (2010).
49. Kapplinger, J. D. *et al.* Distinguishing Arrhythmogenic Right Ventricular Cardiomyopathy/Dysplasia-Associated Mutations from Background Genetic Noise. *J. Am. Coll. Cardiol.* **57**, 2317–2327 (2011).
50. Van Langen, I. M. The use of genotype-phenotype correlations in mutation analysis for the long QT syndrome. *J. Med. Genet.* **40**, 141–145 (2003).
51. Thompson, D., Easton, D. F. & Goldgar, D. E. A Full-Likelihood Method for the Evaluation of Causality of Sequence Variants from Family Data. *Am. J. Hum. Genet.* **73**, 652–655 (2003).
52. Bayrak-Toydemir, P. *et al.* Likelihood ratios to assess genetic evidence for clinical significance of uncertain variants: Hereditary hemorrhagic telangiectasia as a model. *Exp. Mol. Pathol.* **85**, 45–49 (2008).
53. Goldgar, D. E. *et al.* Integrated Evaluation of DNA Sequence Variants of Unknown Clinical Significance: Application to BRCA1 and BRCA2. *Am. J. Hum. Genet.* **75**, 535–544 (2004).
54.; on behalf of the ACMG Laboratory Quality Assurance Committee *et al.* Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. *Genet. Med.* **17**, 405–423 (2015).
55. James M. Heather ∗, Benjamin Chain. The sequence of sequencers: The history of sequencing DNA. *Elsevier*.
56. nextera-xt-library-prep-reference-guide-15031942-05.pdf.
57. Kulski, J. K. Next-Generation Sequencing — An Overview of the History, Tools, and "Omic" Applications. in *Next Generation Sequencing - Advances, Applications and Challenges* (ed. Kulski, J. K.) (InTech, 2016). doi:10.5772/61964.
58. Metzker, M. L. Sequencing technologies — the next generation. *Nat. Rev. Genet.* **11**, 31–46 (2010).
59. Rothberg, J. M. *et al.* An integrated semiconductor device enabling non-optical genome sequencing. *Nature* **475**, 348–352 (2011).
60. Bentley, D. R. *et al.* Accurate whole human genome sequencing using reversible terminator chemistry. *Nature* **456**, 53–59 (2008).
61. Ma, Z. *et al.* Isothermal amplification method for next-generation sequencing. *Proc. Natl. Acad. Sci. U. S. A.* **110**, 14320–14323 (2013).
62. *The Role of Bioinformatics in Agriculture*. (Apple Academic Press, 2014). doi:10.1201/b16568.
63. Aird, D. *et al.* Analyzing and minimizing PCR amplification bias in Illumina sequencing libraries. *Genome Biol.* **12**, R18 (2011).

64. Kozarewa, I. *et al.* Amplification-free Illumina sequencing-library preparation facilitates improved mapping and assembly of GC-biased genomes. *Nat. Methods* **6**, 291–295 (2009).
65. Mardis, E. R. Next-Generation DNA Sequencing Methods. *Annu. Rev. Genomics Hum. Genet.* **9**, 387–402 (2008).
66. Margulies, M. *et al.* Genome Sequencing in Open Microfabricated High Density Picoliter Reactors. *Nature* **437**, 376–380 (2005).
67. SuPrecision™ Platform - Creative Biolabs. https://www.creative-biolabs.com/suprecision/suprecision-platform.htm?gclid=EAIaIQobChMIvMTVq_aB7AIVE-R3Ch2nGAtvEAAYASAAEgLN1fD_BwE.
68. Gužvić, M. The History of DNA Sequencing / ISTORIJAT SEKVENCIRANJA DNK. *J. Med. Biochem.* **32**, 301–312 (2013).
69. grantham1974.pdf.
70. Pollard, K. S., Hubisz, M. J., Rosenbloom, K. R. & Siepel, A. Detection of nonneutral substitution rates on mammalian phylogenies. *Genome Res.* **20**, 110–121 (2010).
71. Reese, M. G., Eeckman, F. H., Kulp, D. & Haussler, D. Improved Splice Site Detection in Genie. *J. Comput. Biol.* **4**, 311–323 (1997).
72. Tabaska, J. E. & Zhang, M. Q. Detection of polyadenylation signals in human DNA sequences. *Gene* **231**, 77–86 (1999).
73. Rentzsch, P., Witten, D., Cooper, G. M., Shendure, J. & Kircher, M. CADD: predicting the deleteriousness of variants throughout the human genome. *Nucleic Acids Res.* **47**, D886–D894 (2019).
74. Shihab, H. A. *et al.* Predicting the Functional, Molecular, and Phenotypic Consequences of Amino Acid Substitutions using Hidden Markov Models. *Hum. Mutat.* **34**, 57–65 (2013).
75. Ritchie, G. R. S., Dunham, I., Zeggini, E. & Flicek, P. Functional annotation of noncoding sequence variants. *Nat. Methods* **11**, 294–296 (2014).
76. Shihab, H. A. *et al.* An integrative approach to predicting the functional effects of non-coding and coding sequence variation. *Bioinformatics* **31**, 1536–1543 (2015).
77. Ng, P. C. & Henikoff, S. Predicting Deleterious Amino Acid Substitutions. *Genome Res.* **11**, 863–874 (2001).
78. Ng, P. C. Accounting for Human Polymorphisms Predicted to Affect Protein Function. *Genome Res.* **12**, 436–446 (2002).
79. Chasman, D. & Adams, R. M. Predicting the functional consequences of non-synonymous single nucleotide polymorphisms: structure-based assessment of amino acid variation11Edited by F. Cohen. *J. Mol. Biol.* **307**, 683–706 (2001).
80. Saunders, C. T. & Baker, D. Evaluation of Structural and Evolutionary Contributions to Deleterious Mutation Prediction. *J. Mol. Biol.* **322**, 891–901 (2002).
81. Sunyaev, S. Prediction of deleterious human alleles. *Hum. Mol. Genet.* **10**, 591–597 (2001).
82. Bairoch, A. The SWISS-PROT protein sequence database and its supplement TrEMBL in 2000. *Nucleic Acids Res.* **28**, 45–48 (2000).
83. Altschul, S. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* **25**, 3389–3402 (1997).
84. Tchernitchko, D., Goossens, M. & Wajcman, H. In Silico Prediction of the Deleterious Effect of a Mutation: Proceed with Caution in Clinical Genetics. *Clin. Chem.* **50**, 1974–1978 (2004).
85. Kumar, P., Henikoff, S. & Ng, P. C. Predicting the effects of coding non-synonymous

variants on protein function using the SIFT algorithm. *Nat. Protoc.* **4**, 1073–1081 (2009).
86. Reva, B., Antipin, Y. & Sander, C. Determinants of protein function revealed by combinatorial entropy optimization. *Genome Biol.* **8**, R232 (2007).
87. Reva, B., Antipin, Y. & Sander, C. Predicting the functional impact of protein mutations: application to cancer genomics. *Nucleic Acids Res.* **39**, e118 (2011).
88. MutationAssessor_white_paper.pdf.
89. Quang, D., Chen, Y. & Xie, X. DANN: a deep learning approach for annotating the pathogenicity of genetic variants. *Bioinformatics* **31**, 761–763 (2015).
90. Choi, Y. A fast computation of pairwise sequence alignment scores between a protein and a set of single-locus variants of another protein. in *Proceedings of the ACM Conference on Bioinformatics, Computational Biology and Biomedicine - BCB '12* 414–417 (ACM Press, 2012). doi:10.1145/2382936.2382989.
91. Chun, S. & Fay, J. C. Identification of deleterious mutations within three human genomes. *Genome Res.* **19**, 1553–1561 (2009).
92. Levy, S. *et al.* The Diploid Genome Sequence of an Individual Human. *PLoS Biol.* **5**, e254 (2007).
93. Wang, X., Xu, H., Ju, D. & Xie, Y. Disruption of Rpn4-Induced Proteasome Expression in Saccharomyces cerevisiae Reduces Cell Viability Under Stressed Conditions. *Genetics* **180**, 1945–1953 (2008).
94. Wheeler, D. A. *et al.* The complete genome of an individual by massively parallel DNA sequencing. *Nature* **452**, 872–876 (2008).
95. Kondrashov, F. A., Rogozin, I. B., Wolf, Y. I. & Koonin, E. V. Selection in the evolution of gene duplications. *Genome Biol.* **3**, research0008.1-research0008.9 (2002).
96. Ng, P. C., Henikoff, J. G. & Henikoff, S. PHAT: a transmembrane-specific substitution matrix. *Bioinformatics* **16**, 760–766 (2000).
97. Sunyaev, S. R. *et al.* PSIC: profile extraction from sequence alignments with position-specific counts of independent observations. *Protein Eng. Des. Sel.* **12**, 387–394 (1999).
98. Adzhubei, I., Jordan, D. M. & Sunyaev, S. R. Predicting Functional Effect of Human Missense Mutations Using PolyPhen-2. *Curr. Protoc. Hum. Genet. Editor. Board Jonathan Haines Al* **0 7**, Unit7.20 (2013).
99. Adzhubei, I. A. *et al.* A method and server for predicting damaging missense mutations. *Nat. Methods* **7**, 248–249 (2010).
100. Berge, K. E. & Leren, T. P. Genetics of hypertrophic cardiomyopathy in Norway: Genetics of HCM in Norway. *Clin. Genet.* **86**, 355–360 (2014).
101. gnomAD. https://gnomad.broadinstitute.org/.
102. View of Combined use of in silico and in vitro splicing assays for interpretation of genomic variants of unknown significance in cardiomyopathies and channelopathies. https://www.pagepressjournals.org/index.php/cardiogen/article/view/cardiogenetics.2012.e6/pdf.

RINGRAZIAMENTI

Il primo ringraziamento va al mio relatore, il Professor Mauro Pesaresi, per avermi accolta nell'Istituto di Medicina Legale e avermi accompagnata negli ultimi due anni di questo lungo percorso, consigliandomi ed aiutandomi nei momenti di difficoltà. Lo ringrazio, inoltre, per avermi dato la possibilità di seguirlo nel suo lavoro, mostrandosi sempre disponibile a trasmettere nuovi insegnamenti, anche attraverso questa tesi.

Ringrazio, inoltre, il Dottor Valerio Onofri, senza il quale la stesura dei capitoli più specifici sulle procedure di laboratorio sarebbe stata per me ingestibile, e con lui anche tutti i medici e collaboratori dell'Istituto, in particolar modo gli specializzandi con cui mi sono interfacciata in questi mesi, che non mi hanno mai fatto mancare il loro sostegno.