UNIVERSITÀ POLITECNICA DELLE MARCHE

DIPARTIMENTO SCIENZE DELLA VITA E DELL'AMBIENTE

Corso di Laurea Magistrale in Biologia molecolare e applicata

Il ruolo del gene desmoplachina nelle cardiomiopatie

Pathogenic role for desmoplakin in cardiomyopathy

Tesi di Laurea Magistrale di:

Iman Aboutabit

Relatore Chiar.mo Prof.

Mauro Pesaresi

Sessione di laurea di Luglio

Anno Accademico 2020/2021

INDICE

INTRODUZIONE 6
Capitolo 1 9
IL GENE DSP 9
1.1 Il gene: accenni generali sulla struttura del gene ed espressione genica 9
1.1.1 Definizione e struttura genica 10
1.1.2 Il meccanismo di trascrizione 15
1.1.3 Il meccanismo di splicing 22
1.1.4 Il meccanismo di traduzione 25
1.1.5 Folding proteico: dalla sequenza amminoacidica alla proteina funzionale 28
1.2 Il gene DSP 30
1.3 I desmosomi 37
1.3.1 Le caderine desmosomiali 40
1.3.2 La famiglia delle proteine armadillo 41
1.3.3 La famiglia delle plachine 42
1.3.4 Modulazione e interazione delle proteine desmosomiali 44
1.4 La desmoplachina 46
Capitolo 2 54
LE CARDIOMIOPTIE CORRELATE A VARIANTI DELLA DESMOPLACHINA 54
2.1 Varianti del gene e malattie correlate 54
2.2 Generalità sulle cardiomiopatie 58
2.3 Il ruolo della genetica nel quadro delle cardiomiopatie 63
2.4 Varianti sul gene DSP correlate a cardiomiopatie 66
2.4.1 Mutazioni: cenni generali 66
2.4.2 Varianti sul gene DSP 68
Capitolo 3 70

MATERIALI E METODI 70
3.1 Tecniche standard di studio delle varianti 70
3.1.1 Il sequenziamento di Sanger 70
3.1.2 Next Generation Sequencing 72
3.2 Interpretazione dei risultati di un test genetico 75
3.3 Materiali e Metodi 80
3.3.1 Tecnica NGS 80
3.3.1.1 Estrazione 80
3.3.1.2 Quantificazione 81
3.3.1.3 Preparazione delle librerie 84
3.3.1.4 Quantificazione delle librerie 87
3.3.1.5 Sequenziamento 87
3.3.2 Conferma variante con Sanger 89
3.3.2.1 Amplificazione 89
3.3.2.2 Quantificazione dei prodotti di amplificazione 90
3.3.2.3 Purificazione del DNA amplificato 90
3.3.2.4 Reazioni di sequenza 91
3.3.2.5 Rimozione dei ddNTPs non incorporati 92
Capitolo 4 95
RISULTATI E DISCUSSIONE 95
4.1 Risultati 95
4.1.1 Paziente 1 95
4.1.2 Paziente 2 101
4.2 Discussione 109
4.2.1 Paziente 1 109
4.2.2 Paziente 2 114
CONCLUSIONI 118
BIBLIOGRAFIA 121

INTRODUZIONE

Le cardiomiopatie sono un gruppo diversificato di disturbi cardiaci che spesso causano insufficienza cardiaca e morte, caratterizzate da un'alterazione anatomo-funzionale che riduce la capacità contrattile del cuore. Esse sembrano costituire una patologia emergente per il concorrere di diversi fattori come più precisi criteri diagnostici e maggiore disponibilità di mezzi strumentali che rendono possibile una diagnosi precoce.

I test genetici per molte cardiomiopatie 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.

Nel corso degli anni, sono stati effettuati numerosi studi di correlazione genotipo-fenotipo che hanno portato all'identificazione di numerose mutazioni nel gene della desmoplachina associate a cardiomiopatie, in particolare cardiomiopatia dilatativa e cardiomiopatia aritmogena.

Lo studio si propone di eseguire una correlazione genotipo-fenotipo in soggetti affetti da cardiomiopatia, portatori di mutazioni di geni che codificano per le proteine desmosomiali, utilizzando un pannello "in silico", ovvero analizzando solo quei geni già noti per essere associati alla patologia in esame.

L’individuazione delle mutazioni del gene della desmoplachina che risultano patogene è di fondamentale importanza per comprendere meglio le relazioni funzionali delle diverse parti della proteina e per tentare di stabilire correlazioni gene-malattia.

Capitolo 1

IL GENE DSP

1.1 Il gene: accenni generali sulla struttura del gene ed espressione genica

L'esistenza dei geni fu ipotizzata per la prima volta da Gregor Mendel, da molti riconosciuto come il padre della genetica, definendoli *elementi ereditari*. Il termine *gene* venne coniato per la prima volta dal botanico danese Wilhelm Ludvig Johannsen nel 1909 e negli anni successivi numerosi studiosi condussero diversi esperimenti grazie ai quali ottennero varie informazioni riguardo la struttura dei geni che portarono ad attribuire a questo termine diverse definizioni[1].

Con il progetto *Genoma Umano*, cominciato nel 1990 e terminato nel 2003, è stato possibile ottenere la sequenza completa del genoma umano e i suoi geni, circa 21.000[2].

Parallelamente agli sviluppi tecnologici che hanno consentito l'accesso diretto al materiale genetico, si è evoluto anche il concetto di gene, considerato inizialmente come un fattore genetico indivisibile con una certa posizione lungo il genoma e in grado di influenzare una caratteristica dell'organismo, successivamente indicato come unità funzionale e divisibile.

Con lo sviluppo della genetica molecolare e delle tecniche di sequenziamento del DNA, il gene venne riconosciuto come porzione del DNA che codifica per una proteina attraverso la produzione di un intermedio, l'RNA messaggero (mRNA). Con la decifrazione del codice genetico, ovvero le regole in base alle quali l'informazione contenuta nel DNA viene tradotta in proteina, e le tecniche di sequenziamento, nella definizione di gene vennero comprese anche regioni che vanno a far parte dell'mRNA senza per questo codificare per una proteina e regioni che inizialmente vengono trascritte in RNA per essere successivamente eliminate, ossia gli introni[3].

Ben presto però ci si rese conto che, oltre alla porzione genomica che determina il prodotto di un gene, si dovevano comprendere le cosiddette regioni regolatorie, ovvero tutte quelle regioni che stabiliscono quanto, quando e dove il gene viene espresso, le quali possono trovarsi a monte, a valle e all'interno del gene, ma che, in maggioranza, non si è ancora in grado di identificare e localizzare con precisione lungo la sequenza di DNA.

1.1.1 Definizione e struttura genica

In generale, il gene viene definito come unità funzionale del genoma, responsabile della trasmissione dei caratteri ereditari degli organismi viventi, costituito da una sequenza lineare di nucleotidi. Di tutto il genoma solo circa il 2% è costituito da geni codificanti proteine, nella restante porzione vi sono tutti

gli elementi regolatori importanti nel determinare come ogni gene viene espresso (tra queste regioni figurano i promotori, i terminatori e gli introni)[2] (Figura 1).

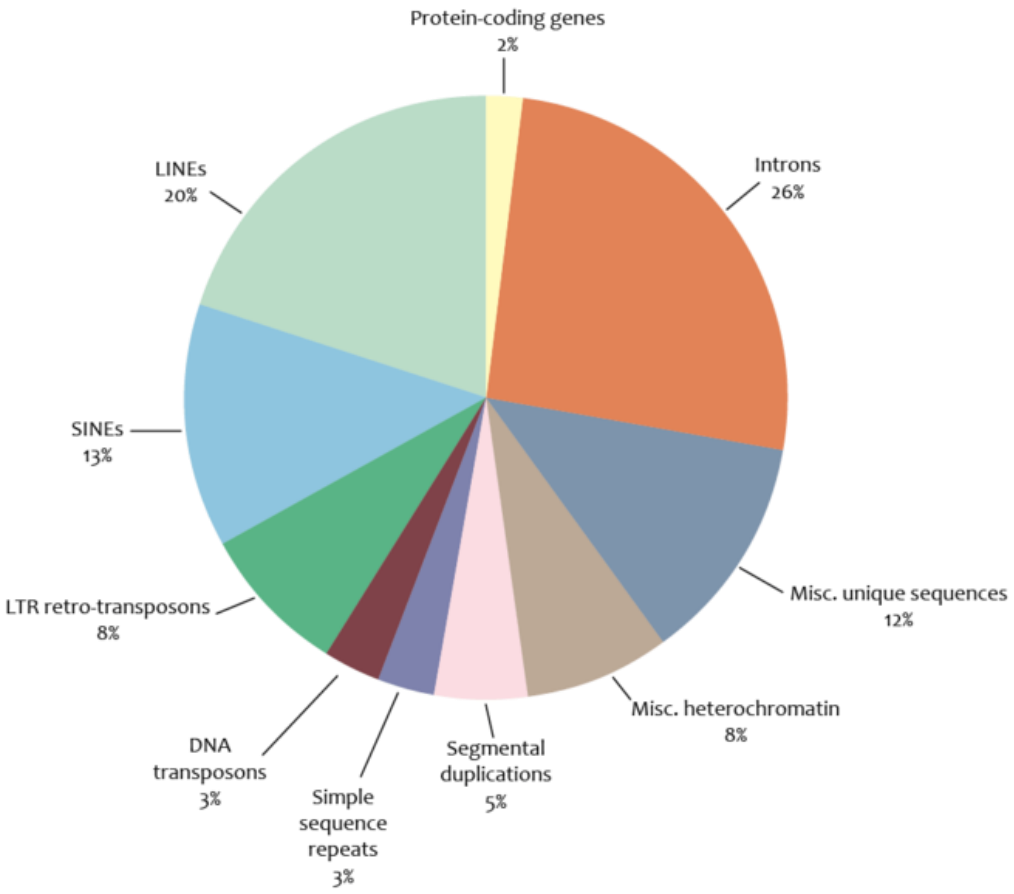


Figura 1. Diagramma dei componenti del genoma stimato nel 2004 dalla National Genetics and Genomics Education Centre[4]. Nell'immagine la porzione dei geni codificanti per proteine è il 2% dell'intero genoma (colorato in giallo), mentre tutte le atre regioni del diagramma rappresentano sequenze non codificanti come gli introni (26% in arancio), i retrotrasposoni (LINE, SINE e LTR, che in totale costituiscono il 41%, rispettivamente in verde chiaro in alto, in celeste e in verde scuro) e tutte le regioni regolatorie e DNA ripetuto.

I geni presentano caratteristiche ben precise e vengono distinti in tre classi principali, che differiscono in base al tipo di RNA che viene prodotto con la loro espressione. Distinguiamo, dunque, i geni della *classe I*, che codificano per RNA ribosomiali (rRNA 28S, 18S, 5.8S); i geni della *classe II*, che codificano per RNA messaggeri (mRNA), molti small nuclear RNA (snRNA) e per micro-RNA (miRNA); i geni della *classe III*, che codificano per RNA transfer (tRNA), piccoli RNA nucleolari (snoRNA), piccoli RNA citoplasmatici (scRNA) e per alcuni micro-RNA (miRNA).

I geni codificanti sono quelli che vengono trascritti in mRNA attraverso il processo di *trascrizione* e contengono una parte realmente codificante che specifica la sequenza degli amminoacidi che costituiranno la proteina. I geni che codificano proteine contengono cornici di lettura aperte, chiamate *ORF* (Open Reading Frame), riconoscibili in base ad una serie di triplette di nucleotidi, dette *codoni*, di inizio e di stop della traduzione.

Negli eucarioti i geni presentano una struttura complessa (Figura 2) che comprende:

- *promotore*, ovvero una regione coinvolta nel processo di trascrizione e presenta nella sua sequenza il sito di legame per la RNA polimerasi. Di solito è un tratto di 100-200 basi posto immediatamente a monte (cioè in posizione 5´) rispetto al sito di inizio della trascrizione. Molto spesso

comprende sequenze del tipo "TATATAT" (TATA box), “CAAT” e “GC” (elementi regolatori prossimali per determinare quando e quanto frequentemente un gene è espresso), che vengono riconosciute da proteine che attivano la trascrizione da parte dell’RNA polimerasi. L’integrità del promotore è richiesta per l’assemblaggio del macchinario generale della trascrizione in risposta ad attivatori della trascrizione.

- *enhancer*, ovvero una regione che può aumentare di molto l’efficienza del promotore, legando altre proteine che favoriscono la trascrizione (gli attivatori). A differenza delle sequenze del promotore, le sequenze "enhancer" possono trovarsi a monte o a valle della regione trascritta (anche a distanza di migliaia di basi da essa) o anche all’interno della regione stessa.
- *sequenza trascritta*, sequenza che codifica per il pre-mRNA, da cui poi si otterrà come prodotto finale la proteina. È costituito da porzioni realmente codificanti, dette *esoni*, intervallate da regioni non codificanti, chiamate *introni*. All’estremità 5’ e 3’ della sequenza trascritta vi sono due regioni di regolazione della trascrizione e traduzione, che prendono il nome di *UTR* (*UnTraslated Region*).

  La dimensione media di un esone è di circa 280 bp, mentre quella di un introne è di 6000 bp.

- *sequenza termine o terminatore*, ossia una sequenza di DNA posta all'estremità opposta del gene dopo la sequenza codificante, che segnala il punto di arresto della trascrizione. Il terminatore non deve essere confuso con il codone di stop (che fa parte del gene), infatti, la sequenza del terminatore si trova fuori dal tratto codificante, di regola dopo il codone di stop segnalando la fine della trascrizione a opera dell'RNA polimerasi.

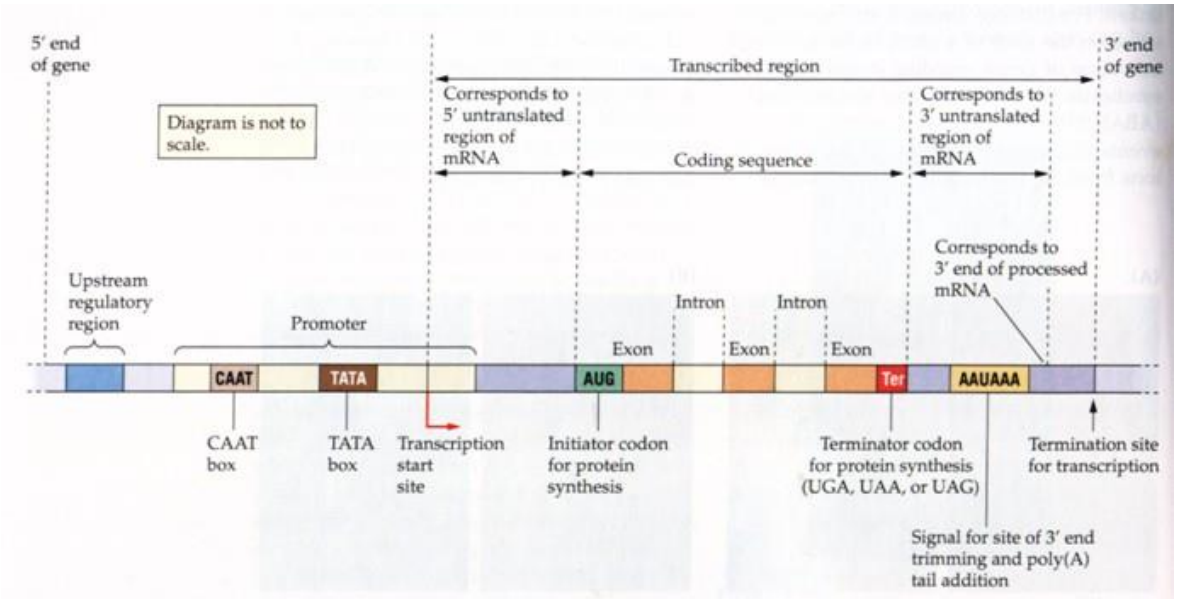


Figura 2. Struttura e organizzazione del gene eucariotico. Un gene si divide in più sezioni. La regione trascritta funge da modello per la sintesi dell'RNA, che viene quindi modificato e tradotto nel prodotto proteico. La regione trascritta è intervallata da sequenze non codificanti che suddividono la regione in sezioni codificanti (esoni) e sezioni non codificanti (introne). La regione trascritta è affiancata su entrambi i lati da sequenze non codificanti che svolgono un ruolo nella regolazione del gene (UnTrasleted Region, UTR). La maggior parte degli elementi della sequenza regolatoria si trovano nella regione fiancheggiante 5', indicata come promotore del gene, poiché contiene sequenze

importanti per la "promozione" della trascrizione. Queste sequenze sono chiamate "elementi cis-acting". L'elemento cis più altamente conservato è il TATA box, che di solito si trova circa 50 bp a monte del sito di inizio della trascrizione.

1.1.2 Il meccanismo di trascrizione

Il processo di trascrizione è il processo che porta nel caso in cui una sequenza di DNA sia codificante, attraverso la formazione di un intermedio mRNA, alla sintesi di peptidi o proteine funzionanti.

La trascrizione inizia con lo svolgimento di una piccola porzione della doppia elica del DNA e uno dei due filamenti agisce da stampo per la sintesi di una molecola di RNA.

Le cellule eucariote presentano tre distinti enzimi di trascrizione ciascuno dei quali è responsabile della sintesi di gruppi diversi di RNA, chiamati *RNA polimerasi*. Distinguiamo, dunque, la *RNA polimerasi I* che sintetizza i grandi RNA ribosomiali; la *RNA polimerasi II* che sintetizza gli RNA messaggeri e piccoli RNA nucleari; ed infine, la *RNA polimerasi III* che sintetizza i vari RNA di trasferimento e l'RNA ribosomiale 5S (Figura 3).

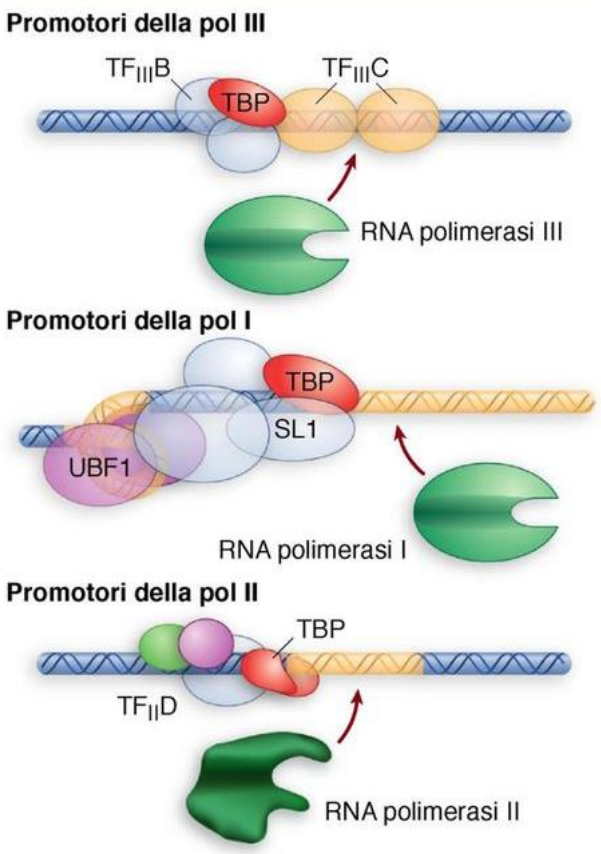


Figura 3. Le RNA polimerasi degli eucarioti svolgono la loro funzione in associazione di una propria serie di fattori di trascrizione. Ogni RNA polimerasi è assistita nel legame al DNA da un proprio fattore di posizionamento rappresentato dalla piccola proteina TBP associata ad altre componenti proteiche, chiamate fattori di trascrizione[5].

I geni che codificano per proteine sono tutti trascritti dalla RNA polimerasi II con l'ausilio di un certo numero di proteine, definite *fattori generali di trascrizione*. Questi ultimi aiutano a posizionare la RNA polimerasi correttamente sul promotore, a separare i due filamenti di DNA per permettere l'inizio della trascrizione e rilasciano l'RNA polimerasi dal promotore nella

modalità di allungamento una volta che la trascrizione è cominciata. Vengono definiti fattori "generali" poiché si assemblano su tutti i promotori usati dalla RNA polimerasi II.

Il processo di assemblaggio inizia con l'attacco della subunità TBP (*TATA Binding Protein*) del fattore di trascrizione per la polimerasi II, *TFIID*, alla sequenza TATA box del promotore. Dopo aver formato il complesso di inizio della trascrizione sul DNA del promotore, la RNA polimerasi II accede al filamento stampo al punto di inizio della trascrizione grazie alla proteina ausiliaria *TFIIH*, che contiene una DNA elicasi, rendendo possibile lo svolgimento del DNA ed esponendo così il filamento stampo (Figura 4).

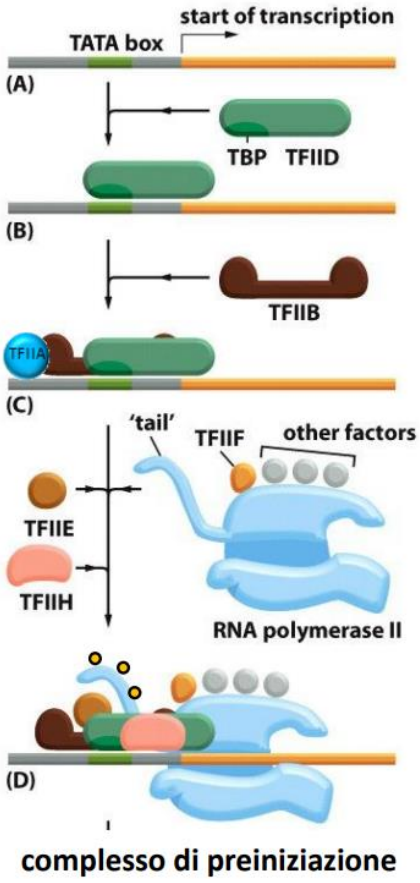


Figura 4. Rappresentazione della formazione del complesso di pre-inizio della trascrizione[6]. Il fattore di trascrizione TFIID si lega alla regione TATA box grazie alla subunità TBP, successivamente si associano anche altri fattori così come la stessa RNA polimerasi II. Infine, si legano i fattori FTIIE ed FTIIH per creare il cosiddetto «complesso chiuso». Questo completa la fase di organizzazione con la formazione del complesso chiuso di pre-inizio. FTIIH è composto da subunità che svolgono tre attività essenziali per la trascrizione: attività di elicasi; attività di chinasi; infine, partecipa anche alla riparazione di dsDNA danneggiato.

Una volta che la polimerasi II inizia ad allungare il trascritto di RNA, la maggior parte dei fattori generali di trascrizione viene rilasciata dal DNA.

L'RNA polimerasi II assembla, dunque, un trascritto primario, *pre-mRNA*, che è complementare al DNA dell'intera unità di trascrizione. Il pre-mRNA viene elaborato nel nucleo per formare l'mRNA maturo che è trasportato poi a livello del citoplasma, le prime tappe di questo processo di conversione includono l'aggiunta del cappuccio all'estremità 5' e della coda di poli(A) all'estremità 3'. L'estremità 5' di tutti gli mRNA possiedono inizialmente un trifosfato derivato dal primo nucleoside trifosfato che è stato incorporato nel sito di inizio della sintesi di RNA. Una volta che l'estremità 5' di un pre-mRNA si forma, alcuni enzimi modificano questa estremità della molecola, inizialmente l'ultimo dei tre fosfati è rimosso, convertendo il terminale 5' in un difosfato. Successivamente viene aggiunto un GMP con orientamento invertito, in modo tale che l'estremità 5' di una guanosina viene a trovarsi di fronte all'estremità 5' della catena di RNA, ottenendo come risultato un'unione degli ultimi due nucleosidi mediante un ponte trifosfato 5'-5'. Infine, la guanosina è metilata nella posizione 7 sulla base di guanina, mentre il nucleotide sul lato interno del ponte trifosfato è metilato nella posizione 2 del ribosio, si ottiene così il *cappuccio di metilguanosina* all'estremità 5' (Figura 5). Si ritiene che il cappuccio all'estremità 5' abbia diverse funzioni, tra le quali: impedire la digestione dell'estremità 5' del mRNA da parte di nucleasi; intervenire nel trasporto dell'mRNA fuori dal nucleo; ed infine avere un ruolo importante per

l'inizio del processo di *traduzione* (processo attraverso il quale dall'mRNA si ottiene la sequenza amminoacidica e quindi il prodotto proteico).

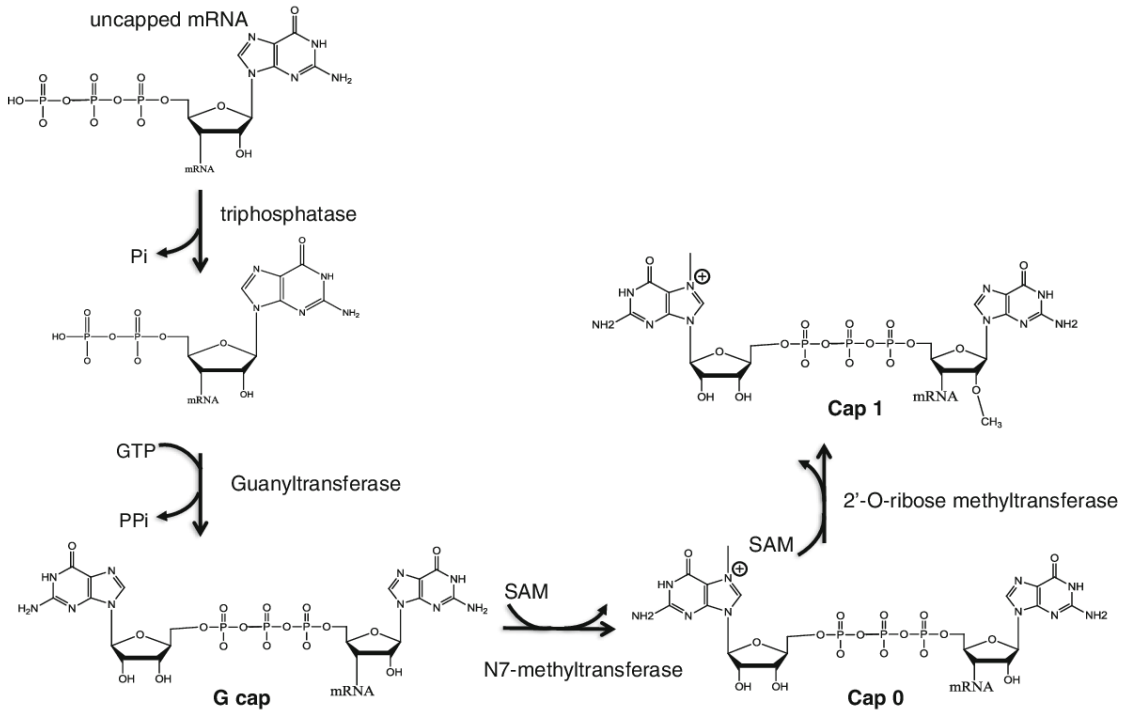


Figura 5. Percorso enzimatico coinvolto nel processo di capping[7]. La formazione del cappuccio a livello post-trascrizionale richiede una sequenza di tre funzioni enzimatiche. Una volta prodotta una struttura Cap 0, questa può essere ulteriormente modificata da 2'-O-ribosio metiltransferasi per ottenere una struttura Cap 1.

All'estremità 3' di un mRNA si trova invece una successione di residui di adenosina che forma la cosiddetta *coda di poli(A)*, la quale si trova approssimativamente 15 nucleotidi a valle di una sequenza AAUAAA, sequenza utilizzata come sito di riconoscimento per una nucleasi che taglia il pre-mRNA appena a valle di questo sito. In seguito alla rottura da parte della

nucleasi, un enzima chiamato *poli(A) polimerasi* aggiunge le adenosine che formano la coda di poli(A) (Figura 6). Questa è implicata nella stabilizzazione dell'mRNA nel citoplasma, proteggendolo da una prematura degradazione.

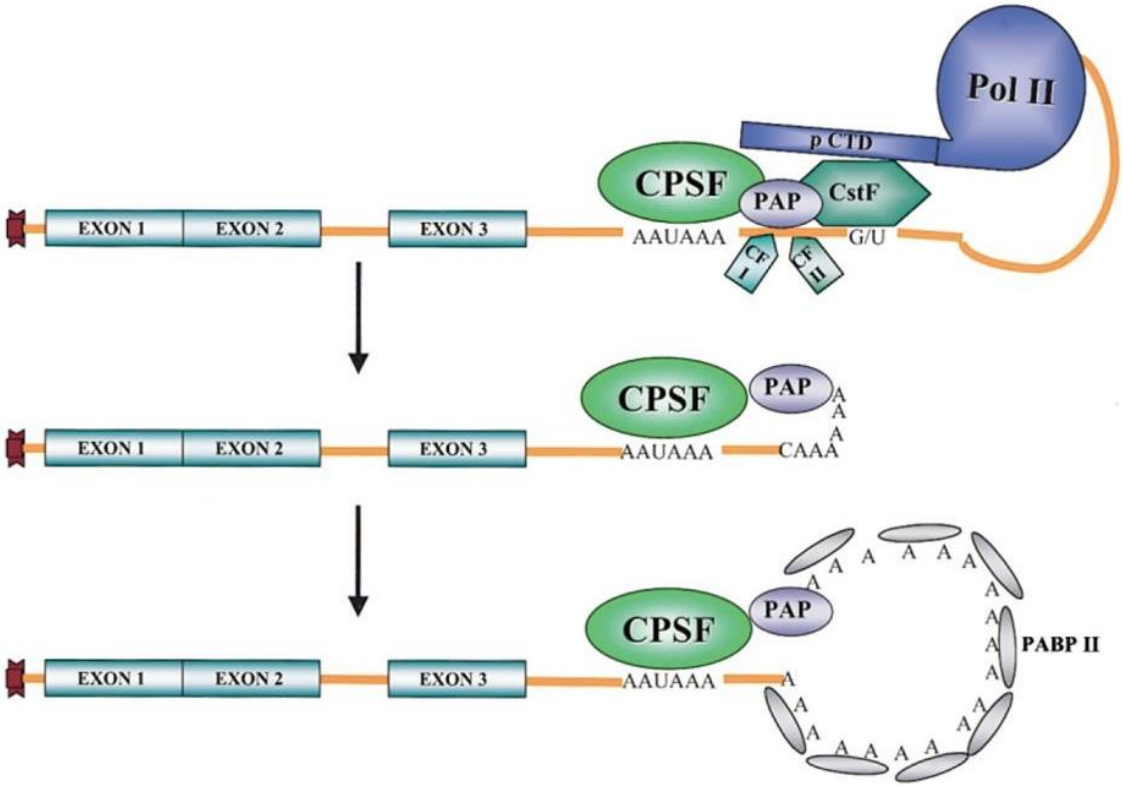


Figura 6. Durante la fine della trascrizione, l'RNA polimerasi II trascrive l'estremità 3′ che contiene segnali riconosciuti da proteine note come CstF (fattore di stimolazione del taglio F) e CPSF (fattore di specificità del taglio e della poliadenilazione) che si trovano già pronte sulla coda della polimerasi II. L'enzima coinvolto è la poli-A polimerasi (PAP) che, a differenza della polimerasi II, non ha bisogno di uno stampo ma crea la coda di poli-A con l'aiuto di specifiche proteine che legano il poli-A, il quale ruolo è quello di determinare la lunghezza finale della coda stessa. In questa fase, l'RNA viene tagliato e l'enzima (PAP) aggiunge, uno alla volta, 200 nucleotidi A all'estremità 3′. Dopodiché, alcune proteine che legano il poli-A si staccano dall'mRNA, mentre altre rimangono attaccate per poter guidare l'mRNA dal nucleo al citosol, fino al ribosoma[8].

1.1.3 Il meccanismo di splicing

Come detto precedentemente i geni sono costituiti da sequenze che sono realmente codificanti, gli esoni, intervallati da sequenze non codificanti, gli introni e sia gli esoni che gli introni sono trascritti in RNA. Tuttavia, le sequenze introniche vengono rimosse in un processo complesso chiamato *splicing dell'RNA* eseguito da un macchinario molecolare complesso e dinamico definito *spliceosoma*. La prima tappa dello splicing, dunque, è l'assemblaggio dello spliceosoma, costituito da piccole ribonucleoproteine nucleari (*snRNP*) complessate a piccoli RNA nucleari (*snRNA*) sul pre-mRNA. L'assemblaggio dello spliceosoma è seguito dalla rottura del sito 5' di splicing, formando un esone libero in 5' e un intermedio a cappio introne-3' esone. La prima reazione di rottura al sito 5' di splicing è seguita rapidamente da una seconda rottura al sito 3' di splicing, che asporta l'introne a forma di cappio e unisce simultaneamente le estremità dei due esoni adiacenti. In seguito al processo di splicing gli snRNP sono probabilmente rilasciati dal pre-mRNA come singole particelle, che vengono riassemblate a livello dei siti di altri introni (Figura 7).

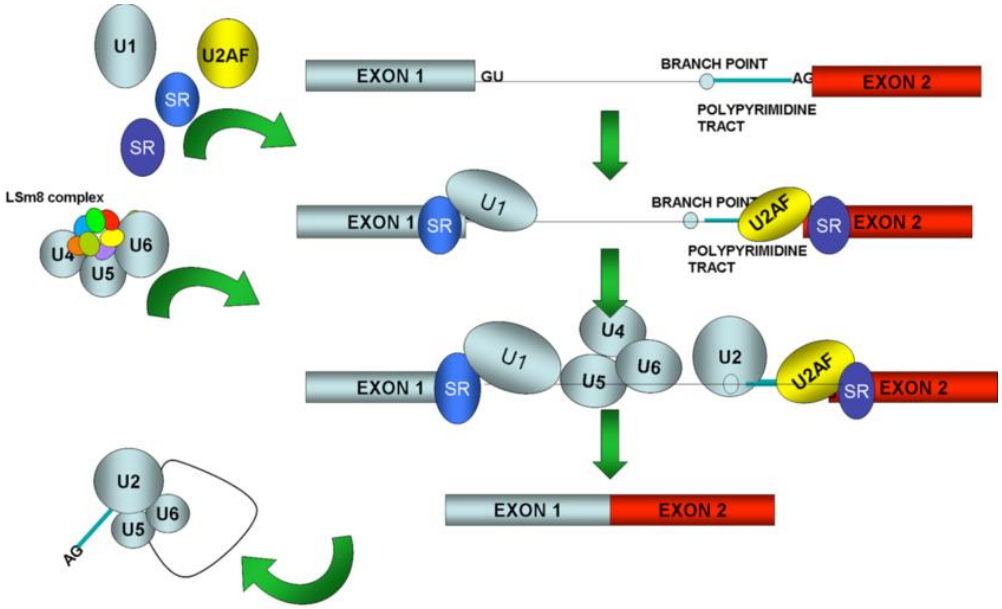


Figura 7. Diagramma schematico degli steps di assemblaggio dello spliceosoma e della rimozione dell'introne[9]. In alto viene mostrata la struttura relativa di esone e introne, che mostrana la sequenza AG alla fine 3' (accettore di giunzione), tratto di polipirimidina con il branch point e GU all'estremità 5' (donatore di giunzione). Al centro viene mostrato come U1 snRNP si lega al sito di giunzione 5'; U2AF si lega al sito 3' e al tratto di polipirimidina; e le proteine SR legano le sequenze enhancer della giunzione dell'esone. In basso è rappresentato il legame di U2 al sito di giunzione 3' a livello del branch point. Le snRNP U4, U6 e U5 sono prima legate al complesso LSm2-8, che coordina l'assemblaggio del tri-snRNP che poi unisce il complesso spliceosoma; successivamente a seguito di cambiamenti conformazionali, U1 e U4 si dissociano dal complesso e l'introne viene rimosso.

Vi è un processo che regola l'espressione genica a livello della maturazione dell'RNA chiamato *splicing alternativo*, che è un meccanismo diffuso mediante il quale un singolo gene può codificare per due o più proteine fra loro correlate. Lo splicing alternativo può diventare anche molto complesso, in

modo da permettere alle cellule di "scegliere" fra un'ampia varietà di combinazioni possibili di esoni differenti. Il meccanismo mediante il quale la cellula seleziona l'inclusione o l'escissione di esoni alternativi dipende primariamente dal fatto che uno specifico sito di splicing al 5' o al 3' venga riconosciuto dall'apparato di splicing come sito da tagliare. Nella maggior parte dei casi, le proteine prodotte da un certo gene mediante splicing alternativo sono identiche per gran parte della loro lunghezza, ma differiscono per regioni chiave che possono influire su proprietà importanti come la loro localizzazione cellulare, il tipo di substrati che esse possono legare o la cinetica delle loro attività catalitiche. Il fatto che venga prodotta una o l'altra variante della proteina regolatoria può cambiare la via differenziativa intrapresa dalla cellula (Figura 8).

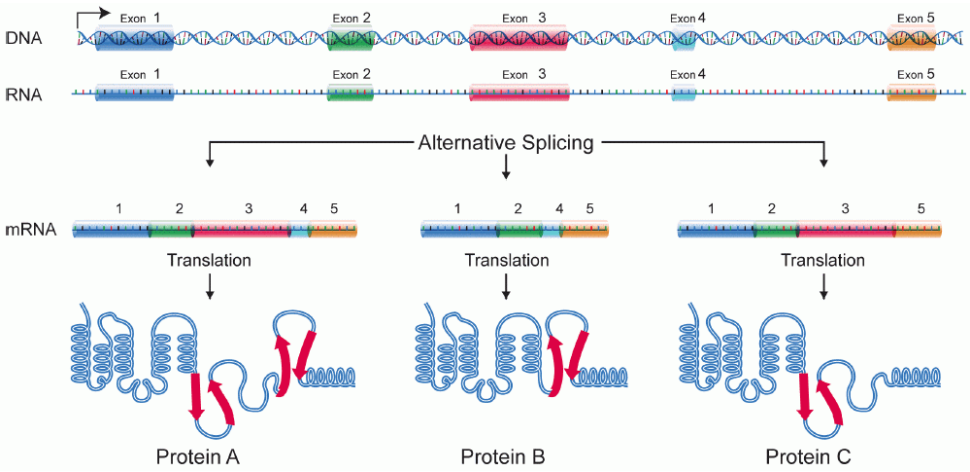


Figura 8. Rappresentazione schematica del meccanismo di splicing alternativo con la produzione di mRNA maturi, che differiscono tra loro per gli esoni presenti. Questi mRNA "alternativi" produrranno, in seguito alla traduzione, proteine differenti tra loro[10].

1.1.4 Il meccanismo di traduzione

Una volta che un mRNA è stato prodotto per trascrizione e modificazione, l'informazione presente nella sua sequenza nucleotidica è usata per sintetizzare una proteina, meccanismo definito *traduzione*. La conversione dell'informazione da RNA a proteine avviene seguendo regole che nel loro insieme sono note come *codice genetico*. La sequenza dei nucleotidi nella molecola di mRNA è letta consecutivamente in gruppi da tre, chiamati *codoni*, e ciascun codone specifica o un amminoacido o un segnale di stop al processo di traduzione. Il gruppo di tre nucleotidi non si lega direttamente all'amminoacido, ma il processo di traduzione dipende da molecole adattatrici che possono riconoscere e legare sia il codone che, in un altro sito sulla loro superficie, l'amminoacido. Questi adattatori consistono di una serie di piccole molecole di RNA note come *RNA transfer (tRNA)*. Per mantenere il corretto quadro di lettura e per assicurare accuratezza, la sintesi proteica viene eseguita nel *ribosoma*, una complessa macchina catalitica composta da più di 50 proteine diverse e da alcune molecole di RNA, gli *RNA ribosomiali (rRNA)*. La sintesi di una catena polipeptidica può essere divisa in tre fasi distinte: l'inizio della catena, l'allungamento della catena e la terminazione. Nella fase di inizio è essenziale che il ribosoma si attacchi ad un punto preciso sull'mRNA in modo tale che l'RNA messaggero possa venire letto secondo un corretto modulo di

lettura. Ciò si realizza in quanto il ribosoma inizia la traduzione in un sito specifico a livello dell'mRNA, chiamato *codone di inizio*. Negli eucarioti, la subunità minore del ribosoma riconosce prima l'estremità 5' dell'mRNA che porta il cappuccio di metilguanosina, scorre poi lungo l'mRNA fino a quando non raggiunge una sequenza di nucleotidi che contiene la tripletta AUG riconosciuta come codone di inizio. Alcune tappe della fase di inizio richiedono l'aiuto di proteine solubili, chiamate *fattori di inizio*.

Ciascun ribosoma ha due siti in grado di legare le molecole di tRNA, *sito A* e *sito P* che svolgono ruoli diversi nella traduzione. Il primo è generalmente quello nel quale gli aminoacil-tRNA entrano nel complesso RNA-ribosoma, mentre il secondo è quello nel quale il tRNA dona amminoacidi alla catena nascente. Quando il tRNA iniziatore è localizzato nel sito P, il ribosoma è pronto per accettare il secondo aminoacil-tRNA e questo è fondamentale affinchè inizi la fase di allungamento. Questo processo necessita l'associazione dei tRNA con i *fattori di allungamento*. La terminazione invece richiede la presenza di *fattori di rilascio*, che interagiscono direttamente con i codoni di stop, la cui funzione è di interrompere la formazione del polipeptide. Quando il ribosoma raggiunge uno di questi codoni, il segnale viene letto in modo da bloccare ogni ulteriore allungamento e liberare il polipeptide legato all'ultimo

tRNA. Dal processo di traduzione si ottiene la sequenza amminoacidica che determinerà il prodotto proteico (Figura 9).

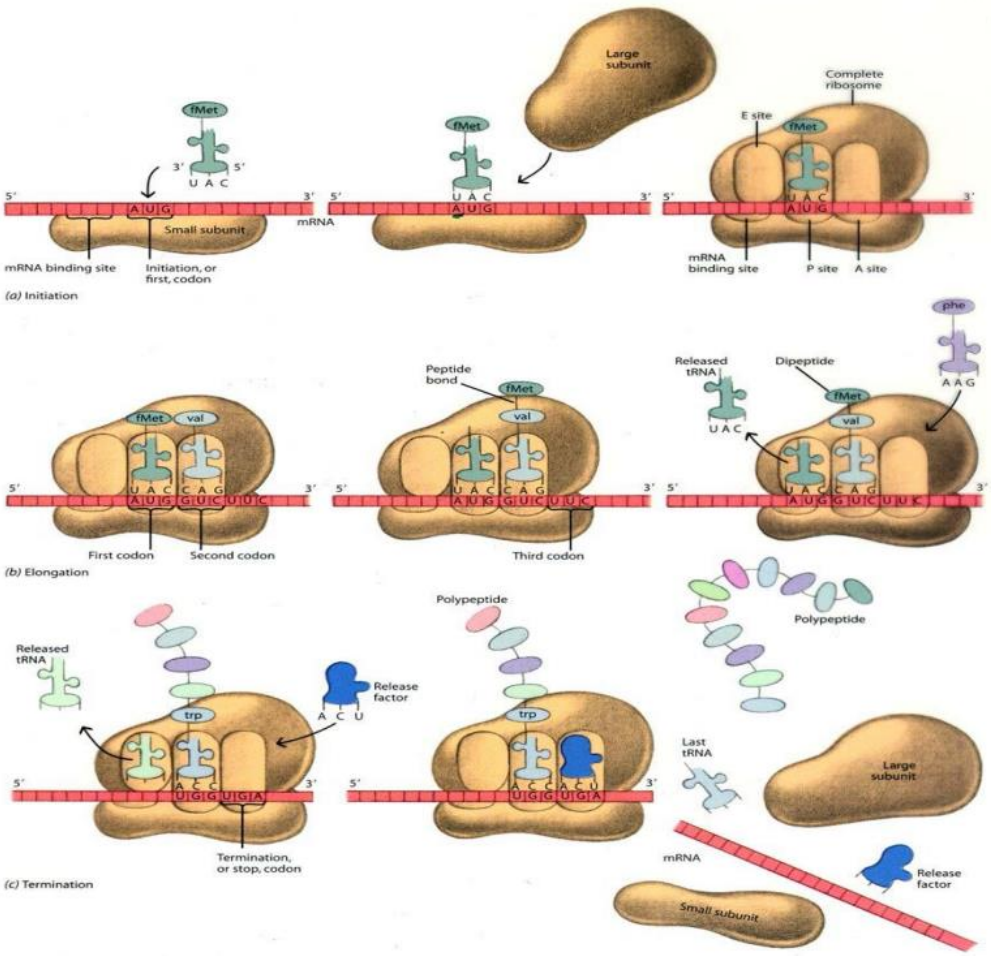


Figura 9. Raffigurazione schematica del processo di traduzione. In figura viene mostrata la fase di inizio, in alto, con la formazione del complesso di traduzione; nella parte centrale è raffigurata la fase di elongazione con la formazione del primo legame peptidico; infine, in basso, è mostrata la fase di terminazione con il rilascio del polipeptide e dei fattori che costituiscono il complesso di traduzione.

1.1.5 Folding proteico: dalla sequenza amminoacidica alla proteina funzionale

La sequenza amminoacidica specifica, dunque, per una determinata proteina che consiste di un'ossatura polipeptidica con catene laterali attaccate. Ciascun tipo di proteina differisce nella sequenza e nel numero di amminoacidi e sono perciò le sequenze delle catene laterali chimicamente diverse che contraddistinguono ciascuna proteina. La maggior parte delle proteine ha una struttura tridimensionale particolare che è determinata dall'ordine degli amminoacidi nella sua catena. Ciascuna proteina normalmente si ripiega in un'unica struttura finale ripiegata, o conformazione, stabile grazie all'aiuto di proteine speciali chiamate *chaperoni molecolari*. Tuttavia, la conformazione cambia leggermente quando la proteina interagisce con altre molecole nella cellula, cambiamento spesso cruciale per la funzione della proteina. Il ripiegamento di un polipeptide inizia con la formazione di un piccolo frammento di struttura secondaria, come un α-elica o un foglietto β, che può agire da "impalcatura" per la stabilizzazione di altre regioni ordinate della proteina; successivamente si hanno interazioni tra le diverse strutture secondarie che portano alla formazione di un globulo fuso, ovvero una struttura compatta nella quale è avvenuto gran parte del ripiegamento della struttura secondaria e terziaria, ma nella quale i residui idrofobici racchiusi all'interno

non si sono ancora assestati nella loro conformazione definitiva. Attraverso una serie di piccoli aggiustamenti conformazionali, la catena polipeptidica raggiunge una struttura terziaria più compatta, una conformazione nativa di una proteina con una sola subunità; successivamente le subunità si uniscono tramite legami chimicamente deboli e si hanno ulteriori piccoli aggiustamenti conformazionali che generano la struttura finale e funzionale della proteina (Figura 10).

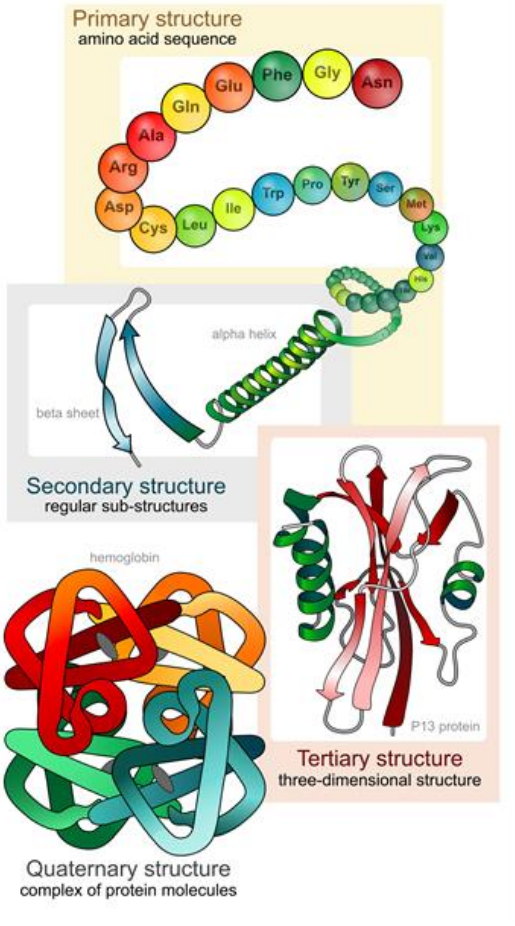


Figura 10. Rappresentazione del processo di folding proteico dell'emoglobina. Si parte dalla sequenza amminoacidica che si assembla a formare la struttura secondaria costituita da α-eliche e

foglietti β. Successivamente si passa dalla struttura terziaria ed infine alla formazione della struttura quaternaria che determina la molecola proteica funzionale.

1.2 Il gene DSP

Il gene della desmoplachina, *DSP*, è un gene codificante proteina localizzato sul cromosoma 6p24.3 forward strand (Figura 11) e si estende per ~ 45 kb del DNA genomico[11] ed è mappato con le coordinate 7.541.850-7.586.947 nel riferimento GRCh37.

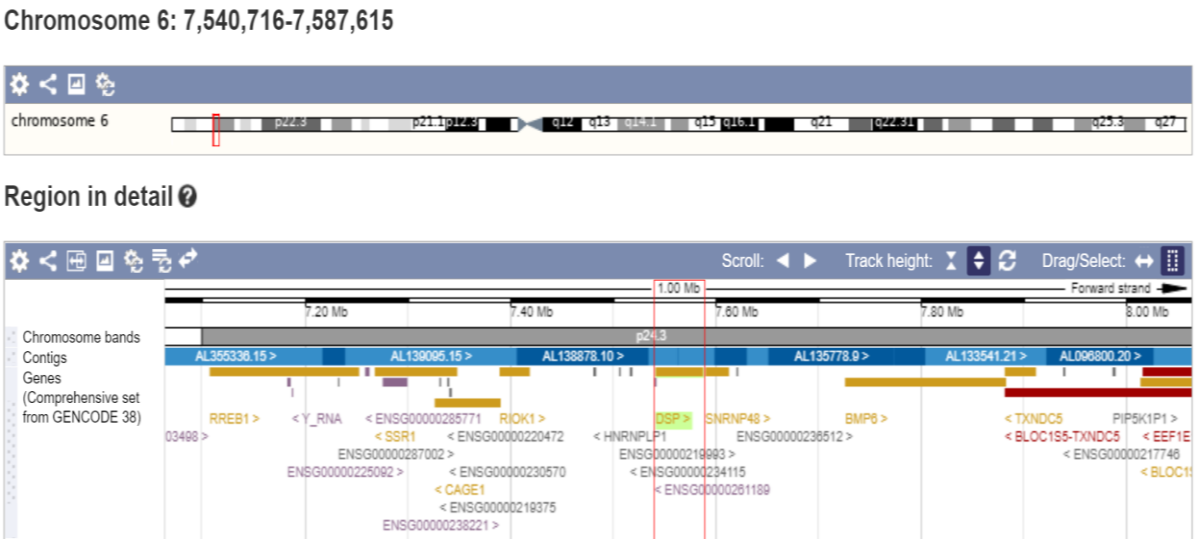


Figura 11. Immagine della localizzazione del gene *DSP* presente su Ensembl. In alto è posta la raffigurazione schematica del cromosoma 6, dove la posizione del gene è evidenziata dal rettangolo rosso. In basso viene raffigurata la regione a livello cromosomico in dettaglio in cui il gene in esame è evidenziato in verde.

Il gene comprende 24 esoni intervallati da 23 introni che presentano una lunghezza variabile. Gli esoni hanno una lunghezza che varia da 51 bp a 3922 bp, mentre la lunghezza degli introni varia da 0.3 kb a 14 kb (Figura 12).

Struttura genomica della desmoplachina				
N°	Dimensione esoni (bp)	Dimensione introni (kb)	Sequenza del donatore	Sequenza dell'accettore
1	449	14.0	GTGGGTACCT	Non Determinata
2	103	2.3	GTAAGCTTTC	GCCTCCTTAG
3	149	1.0	GTATTGTCCA	TTTCTTTCAG
4	175	3.3	GTGGTGGAGG	TTCTTTGCAG
5	129	1.0	GTACTTGTCT	GTGTGCGCAG
6	51	1.6	GTAAGCTGA	TGTCTTCTAG
7	162	0.9	GTAAGTTCAC	CTCTGTGCAG
8	105	0.9	GTAAGCTTCA	TCATTCACAG
9	96	0.3	GTTTTATAT	CTTGTTTCAG
10	126	0.5	GTATCGTCTC	TCTTTCACAG
11	153	0.6	GTGTGTTCTC	ACCATTGCAG
12	155	1.1	GTAAGTCATC	GCTCTTACAG
13	127	0.8	GTATGTCCTC	TGCTTCTTAG
14	202	0.3	GTCGGCTTGG	CTCCTTTCAG
15	227	2.0	GTAGGTATCT	CTTTTACAG
16	167	0.4	GTAAACTCAG	TTTTTGACAG
17	139	0.5	GTAACTTGAA	CTCTTTCCAG
18	194	0.8	GTATGCCAGC	CATCTTGCAG
19	163	0.5	GTATAAGCCA	TTCCCCCCAG
20	84	0.7	GTATTGTTGGT	TTTTGTATAG
21	108	0.6	GTATATATGT	TATGCTGCAG

22	99	0.7	GTAATTTATA	TTTCTCCAAG
23	2295	1.1	GTATTCACAA	CATTCCACAG
24	3922	-	Non Determinata	TCAATTCCAG

Figura 12. Nella seguente tabella[11] sono elencati gli esoni e gli introni che costituiscono il gene *DSP* con le rispettive dimensioni e le sequenze del sito donatore e accettore. Tali sequenze, chiamate *sequenze consenso dello splicing*, vengono riconosciute da snRNP che tagliano in questi punti l'mRNA per l'escissione delle sequenze introniche nel meccanismo dello splicing.

Il gene *DSP* presenta sette varianti di splicing, non tutte codificanti proteine, schematizzate nella Figura 13:

1. DSP-201: ha una lunghezza di 9697 bp e presenta 24 esoni; codifica per una proteina di 2871 aa ed è associata a 15197 varianti alleliche.
2. DSP-202: ha una lunghezza di 7812 bp, il trascritto presenta 24 esoni; codifica per una proteina di 2272 aa ed è associata a 15200 varianti alleliche.
3. DSP-206: ha una lunghezza di 457 bp ed è costituito da 2 esoni; codifica per una proteina di 77 aa ed è associata a 4178 varianti alleliche.
4. DSP-205: ha una lunghezza di 1910 bp e presenta 4 esoni; è un trascritto che non codifica per una proteina, ma è un introne trattenuto ed è associato a 5478 varianti alleliche.

5. DSP-207: ha una lunghezza di 1634 bp e presenta 5 esoni, anche questo trascritto è un introne trattenuto ed è associato a 1784 varianti alleliche.
6. DSP-204: ha una lunghezza di 1508 bp, presenta 3 esoni; è un introne trattenuto ed è associato a 1065 varianti alleliche.
7. DSP-203: ha una lunghezza di 754 bp, presenta 4 esoni; non codifica per nessuna proteina in quanto è un introne trattenuto ed è associato a 1885 varianti alleliche.

DSP-205, DSP-207, DSP-204 e DSP-203 sono tutti introni trattenuti, ovvero trascritti alternativi di splicing ritenuti contenere una sequenza intronica rispetto agli altri trascritti codificati dallo stesso gene. Si parla di introne trattenuto quando i siti di taglio di un introne possono non essere riconosciuti, in questo caso l'introne non viene eliminato dal trascritto di mRNA. In genere il fenomeno degli introni trattenuti interessa introni più corti rispetto agli introni che vengono esclusi dai trascritti mediante lo splicing. Se l'introne trattenuto si trovasse nella regione codificante potrebbe alterare la cornice di lettura degli esoni, se così fosse si potrebbe generare una proteina tronca o non funzionale.

Show/hide columns (1 hidden)

Name	Transcript ID	bp	Protein	Biotype
DSP-201	ENST00000379802.8	9697	2871aa	Protein coding
DSP-202	ENST00000418664.2	7812	2272aa	Protein coding
DSP-206	ENST00000683682.1	457	77aa	Protein coding
DSP-205	ENST00000683563.1	1910	No protein	Retained intron
DSP-207	ENST00000684395.1	1634	No protein	Retained intron
DSP-204	ENST00000682228.1	1508	No protein	Retained intron
DSP-203	ENST00000506617.1	754	No protein	Retained intron

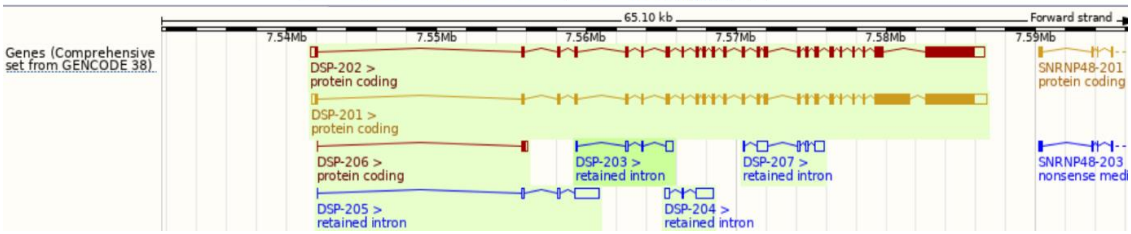


Figura 13. Rappresentazione delle varianti del gene *DSP*, mostrato su Ensembl . Nella tabella in alto sono elencate le varianti con rispettiva lunghezza della sequenza nucleotidica e della sequenza amminoacidica e il biotipo trascrizionale. Nella rappresentazione in basso sono raffigurate in modo schematico, evidenziate in verde, le sequenze dei diversi trascritti, in giallo e in rosso sono raffigurati i trascritti codificanti proteine, mentre in blu vengo rappresentati gli introni trattenuti.

Sono state osservate e descritte tre isoforme principali generate attraverso il meccanismo di splicing alternativo (Figura 14):

- La variante di trascrizione 1, nota anche come *DPI*, scelta come sequenza canonica, presenta una lunghezza di 9697 nt, la regione codificante (CDS, *coding DNA sequence*) è composta da 8616 nt. Da questo trascritto si ottiene come prodotto proteico la isoforma I (DSPI). La massa della proteina non processata è di 331,774 Da ed è costituita da 2871 aa.

- La variante di trascrizione 2, nota anche come *DPII*, presenta una lunghezza di 7812 nt e una CDS lunga 6819 nt. Codifica per la isoforma II (DSPII) costituita da 2272 aa e una massa della proteina non processata di 260,119 Da, DSPII differisce da DSPI per le dimensioni del dominio alfa-elicoidale centrale *rod*, in quanto DSPII presenta due terzi in meno di amminoacidi all'interno di questo dominio[12], in particolare manca della porzione che si estende dall'amminoacido in posizione 1195 all'amminoacido in posizione 1793, presente invece nella sequenza amminoacidica dell'isoforma canonica.
- La variante di trascrizione 3, individuata recentemente, si presenta con una lunghezza di 8368 nt e una CDS costituita da 7287 nt. Questa variante è espressa come risultato di un sito donatore di giunzione alternativo in-frame all'interno del gene DSPI. Codifica per la isoforma Ia (DSPIa) con una massa della proteina non processata di 278,916 Da e comprende 2428 aa, identica a DSPI e DSPII ad eccezione del dominio centrale dell'asta alfa-elicoilica, che è di dimensioni intermedie, essendo 156 amminoacidi più grande del DSPII e 443 aminoacidi più piccola del DSPI[13], infatti, manca della porzione che si estende dall'amminoacido in posizione 1351 all'amminoacido in posizione 1793 rispetto all'isoforma canonica DPI.

Figura 14. Nella figura, tratta dal Genome Browser dell'NCBI, sono rappresentate le tre principali varianti di splicing sopra descritte. In verde sono raffigurate in modo schematico le sequenze genomiche, in viola sono raffigurati gli mRNA messaggeri codificanti per proteina, infine, in rosso è rappresentato il prodotto proteico ottenuto dalla variante. Dall'alto verso il basso, il primo gruppo di sequenza genomica, mRNA e prodotto proteico è relativo alla variante di splicing canonica 1 che codifica per l'isoforma canonica DPI, il secondo gruppo è relativo alla variante di splicing 3 che codifica per l'isoforma DPIa; infine, il terzo gruppo è relativo alla variante di splicing 2 che codifica per l'isoforma DPII.

L'espressione dell'mRNA a livello dei tessuti mostra come i livelli di mRNA sono molto alti a livello del muscolo cardiaco e della pelle, oltre che a livello della lingua e dell'esofago, per quanto riguarda la pelle, l'esofago e la lingua per rilevare i livelli di espressione dell'mRNA sono stati utilizzati tessuti potenziati (Figura 15). La panoramica dell'espressione dell'RNA mostra i dati provenienti da tre diverse fonti, rispettivamente i dati RNA-seq generati internamente del Human Protein Atlas (HPA), i dati RNA-seq dal progetto Genotype-Tissue Expression (GTEx) e i dati CAGE dal progetto FANTOM5, il set di dati si basa su una combinazione di tutte e tre le fonti.

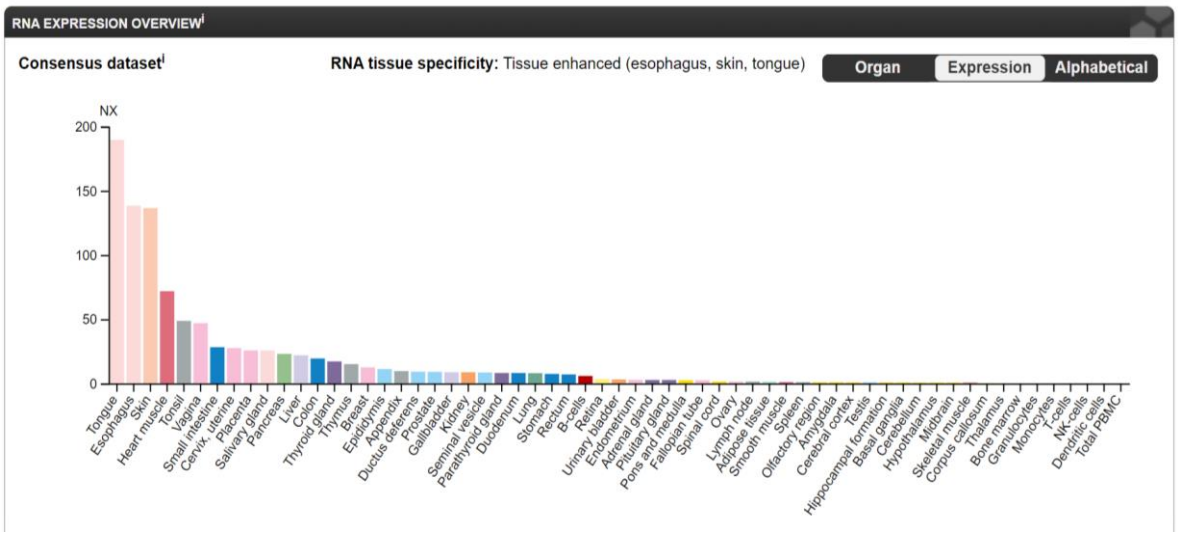


Figura 15. I dati di espressione normalizzati, estrapolati dal The Human Protein Atlas, sono stati ottenuti da 55 tipi di tessuti e 6 tipi di cellule del sangue, combinando i dati dei tre set di dati di trascrizione (HPA,GTEx e FANTOM5) utilizzando un iter di normalizzazione interno. La codifica a colori si basa su gruppi tissutali, ognuno costituito da tessuti con caratteristiche funzionali in comune.

1.3 I desmosomi

I desmosomi sono giunzioni intercellulari adesive, multiproteiche e simmetriche a forma di disco[14]. Sono espressi in particolar modo nei tessuti sottoposti a significativi stress meccanici in senso laterale o da stiramento come gli epiteli e il miocardio. Hanno una funzione principale nell'adesione cellulare mediante la connessione degli elementi del citoscheletro di una cellula con l'altra, permettendo così al tessuto di funzionare come unità strutturale. I desmosomi formano un reticolo senza soluzione di continuità che si dirama nel tessuto, con la principale funzione di fornire la forza tensile, ammortizzando le

sollecitazioni meccaniche e ponendo resistenza alle forze che agiscono per interrompere l'adesività cellula-cellula (Figura 16).

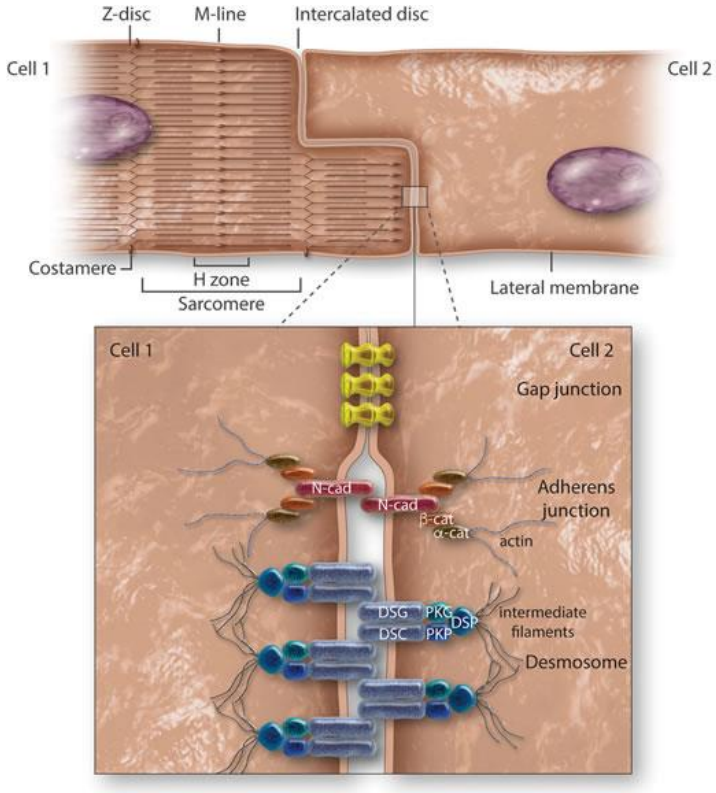


Figura 16. Raffigurazione della struttura desmosomiale tra due cellule[15]. Nell'immagine è possibile osservare in basso anche le varie proteine che sono coinvolte nell'adesione cellulare.

Da un punto di vista ultrastrutturale, i desmosomi presentano un aspetto peculiare; essi, infatti, sono costituiti da due regioni principali: la regione centrale extracellulare (ECD), detta anche "desmoglea", attraversata da una linea intermedia elettrondensa, e la placca citoplasmatica, densa e simmetrica,

che si sviluppa parallelamente alla membrana plasmatica. Ciascuna placca a sua volta consiste di due regioni: la placca densa esterna (ODP) e la placca interna un po' meno densa (IDP), sulla quale si inseriscono i filamenti intermedi (IF). La regione compresa tra i filamenti intermedi e il lato interno di ODP è chiamata "regione satellite"[16] (Figura 17).

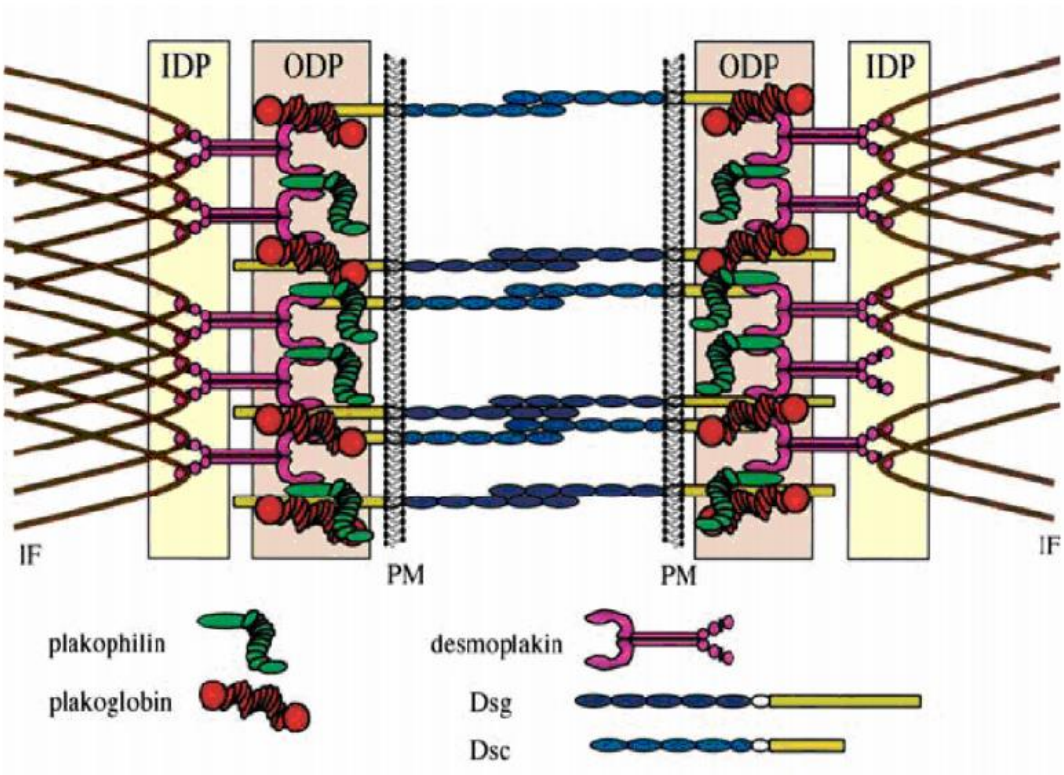


Figura 17. Modello molecolare dei desmosomi. Sono illustrate le interazioni tra le proteine desmosomali e la loro localizzazione all'interno della placca. (ODP: placca densa esterna; IDP: placca densa interna; IF: filamenti intermedi; PM: membrana plasmatica)[17].

Le proteine che si assemblano per formare i desmosomi appartengono a tre famiglie distinte e separate (Figura 18): le plachine (desmolplachina), le proteine armadillo (placofillina e placoglobina) e le caderine desmosomiali (desmogleina e desmocollina)[17].

1.3.1 Le caderine desmosomiali

Le caderine desmosomiali sono glicoproteine transmembrana calciodipendenti composte dalle desmogleine (DSG1-4) e dalle desmocolline (DSC1-3), ambedue codificate da una piccola famiglia multigenica situata nel cromosoma 18. Nelle caderine si riconoscono tre domini funzionali:

- Un dominio extracellulare aminoterminale, calcio-inducibile, che contiene quattro subdomini extracellulari altamente conservati (EC1-EC4), importante per interazioni intercellulari omofiliche, seguito da un altro dominio extracellulare variabile d'ancoraggio (EA). Il legame del calcio ai domini EC è importante per la stabilizzazione della loro struttura e funzione. Al primo dominio extracellulare delle caderine è presente una sequenza altamente conservata (CAR: cell adhesion-recognition) importante per le interazioni eterofiliche tra le cellule.
- Un singolo dominio transmembrana.
- Un subdominio intracellulare, che ancora il citoscheletro, tramite interazioni complesse con le varie proteine intracitoplasmatiche[18].

Le caderine desmosomiali si interfacciano con la loro componente extracellulare in maniera diretta con la rispettiva controparte delle cellule vicine. Con la loro porzione intracellulare invece interagiscono con le proteine armadillo placoglobina e la placofillina-2, che a loro volta interagiscono con la desmoplachina, che crea i siti di legame per i filamenti intermedi. Queste interazioni sono essenziali per il processo d'adesione. Le caderine, non hanno solamente un ruolo di adesione, ma anche un ruolo regolatorio sulla morfogenesi tessutale e possono inoltre partecipare in processi intracellulari di segnalazione[19]. Si ritiene che possano agire come sensori molecolari regolando il comportamento cellulare e traducendo dei segnali cellulari: DSG2 e DSC2 sono espresse ubiquitariamente nei tessuti contenenti desmosomi[19,20].

1.3.2 La famiglia delle proteine armadillo

La famiglia delle proteine armadillo è caratterizzata da un dominio centrale con un numero variabile di motivi ripetuti composti da 42 aminoacidi ciascuno, ognuno dei quali serve per l'interazione con le varie proteine[21]. Alla famiglia delle proteine armadillo appartengono la placofillina-2, la placoglobina, la β-catenina ed altre proteine. La placofillina-2 e la placoglobina si trovano nella parte esterna della placca densa dei desmosomi ed hanno funzioni simili, legando l'estremità N-terminale della desmoplachina e C-terminale delle caderine[22]. La placofillina-2 si trova in due isomorfe generate da splicing

alternativo (2a e 2b) ed è la placofillina prevalente nel cuore[23]. La placofillina lega le caderine desmosomiali con le desmoplachine e i filamenti intermedi, conferendo integrità strutturale e funzionale alle cellule adiacenti. Interagisce, inoltre, con molte altre proteine di adesione cellulare. La placoglobina (o γ-catenina) si trova sia nei desmosomi che nelle giunzioni aderenti ed è l'unica proteina conosciuta comune alle due giunzioni. Sembrerebbe avere un ruolo particolarmente importante nello stabilizzare questi complessi, nelle funzioni adesive tra le cellule e nelle interazioni con il citoscheletro: la sua funzione a livello dei desmosomi è quella di fungere da collegamento tra la desmoplachina e le caderine, si viene così a creare un "bridge" indiretto tra caderine e filamenti di desmina[23]. Nelle giunzioni aderenti invece la placoglobina interagisce con le caderine trasmettendo segnali di adesione ed ancora le giunzioni direttamente al citoscheletro di actina[24]. La placoglobina, inoltre, è coinvolta nelle vie dell'apoptosi.

1.3.3 La famiglia delle plachine

Le plachine sono una famiglia di proteine "linker", che si associano agli elementi citoscheletrici e ai complessi giunzionali "orchestrando" così lo sviluppo cellulare e mantenendo l'integrità cellulare. Numerose plachine sono espresse in tessuti che sono sottoposti a stress meccanici (come epiteli e muscoli), dove hanno un importante ruolo per l'integrità tessutale, legando i

filamenti citoscheletrici e ancorandoli alla membrana plasmatica. Alla famiglia appartengono la desmoplachina, l'epiplachina, la periplachina, l'envoplachina, la plectina e la MACF che si accomunano strutturalmente per il fatto di avere in comune due domini[25]. La desmoplachina si trova nella parte interna della placca densa ed è la proteina più abbondante dei desmosomi e le sue due estremità hanno anche un ruolo nella regolazione dell'assemblaggio e organizzazione del complesso desmosomiale[25,26].

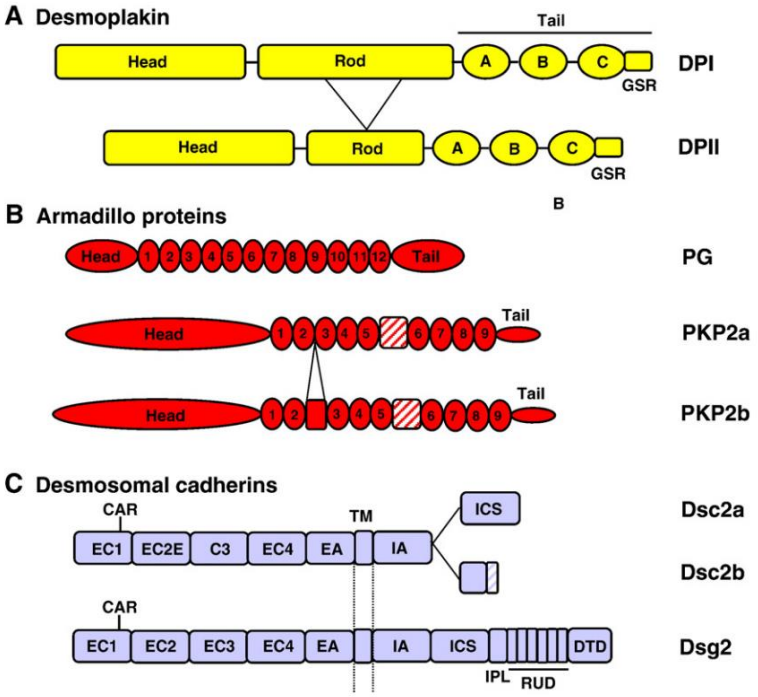


Figura 18. Struttura delle proteine desmosomiali[27]. (A) Vengono mostrate le due isoforme principali della desmoplachina, con i domini A, B e C che sono i domini ripetuti delle plachine e al C-terminale il dominio ricco in glicine-serine-arginine. (B) Sono mostrate le proteine armadillo pacoglobina (PG) e le due isoforme della placofillina (PKP2a e PKP2b). (C) Vengono rappresentate le caderine desmosomiali desmogleina (Dsg2) e le due isoforme della desmocollina (Dsc2a e Dsc2b).

1.3.4 Modulazione e interazione delle proteine desmosomiali

I geni che codificano per le proteine desmosomiali sono modulati da citochine. Il fattore di crescita trasformante TGFβ3 è un membro della famiglia dei fattori di crescita TGF, vasto gruppo di citochine regolatorie con effetti pleiotropici, che giocano un ruolo importante nello sviluppo e omeostasi tessutale. I TGFβ stimolano le cellule mesenchimali a proliferare e promuovono l'espressione di geni per la matrice extracellulare e sopprimono l'attività di geni che codificano per le metalloproteinasi della matrice, coinvolte alla degradazione della matrice extracellulare. È stato dimostrato che i TGFβ modulano l'espressione di geni, che codificano le proteine desmosomiali in vari tipi cellulari, potendo quindi influire sulla stabilità delle giunzioni tra cellula e cellula. I TGFβ3 quindi modula l'adesione cellulare[28]. Quest'ultima avviene mediante una serie di interazioni tra le varie proteine che portano all'assemblaggio dei desmosomi (Figura 19).

Le proteine desmoglleina e desmocollina interagiscono con i siti distinti nel dominio ripetuto centrale *arm* della placoglobina, attraverso i loro domini ICS. La placoglobina interagisce a sua volta con il dominio N-terminale delle plachine della desmoplachina. Il domio C-terminale della desmoplachina interagisce con i filamenti intermedi e questo completa la catena di molecole

che lega la membrana al citoscheletro. In questa catena la pacoglobina può agire come linker tra le caderine desmosomiali e la desmoplachina, ma nel caso in cui le caderine desmosomiali interagiscano direttamente con la desmoplachina, la pacoglobina viene coinvolta nella regolazione delle interazioni laterali. La pacofillina interagisce con partners multipli e numerosi studi hanno evidenziato che la testa del dominio della pacofillina è richiesto per il reclutamento della desmoplachina[29].

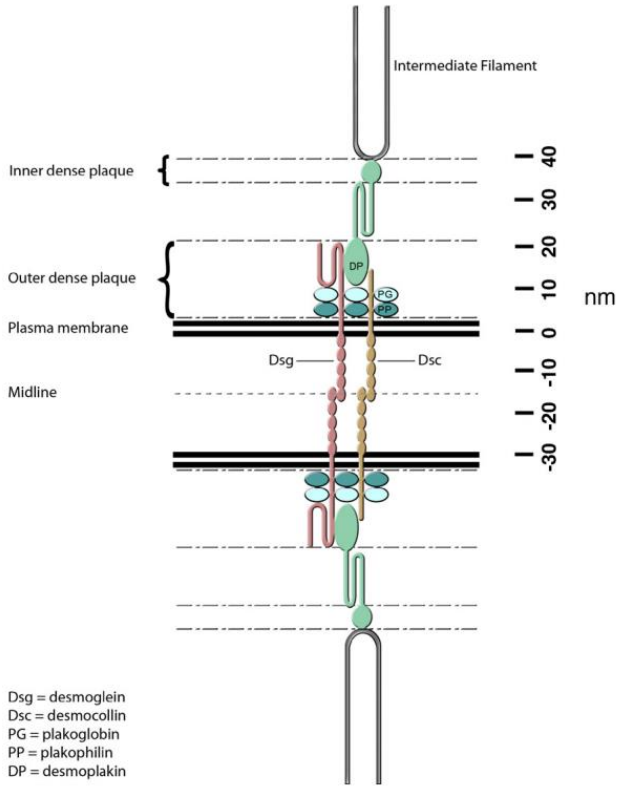


Figura 19. Modello schematico dei desmosomi mostrando le rispettive posizioni delle principali componenti desmosomiali[30]. La scala a destra indica la distanza in nm.

1.4 La desmoplachina

La desmoplachina è una proteina codificata dal gene *DSP* e appartenente alla super famiglia delle plachine.

È una componente obbligata dei desmosomi funzionali nei muscoli cardiaci e nelle cellule epidermiche, che funzionano per mantenere l'integrità strutturale nei contatti cellulari adiacenti.

La desmoplachina, assieme alla placoglobina, è una delle principali proteine della placca desmosomica che omodimerizza e adotta una conformazione a forma di manubrio[31] (Figura 20).

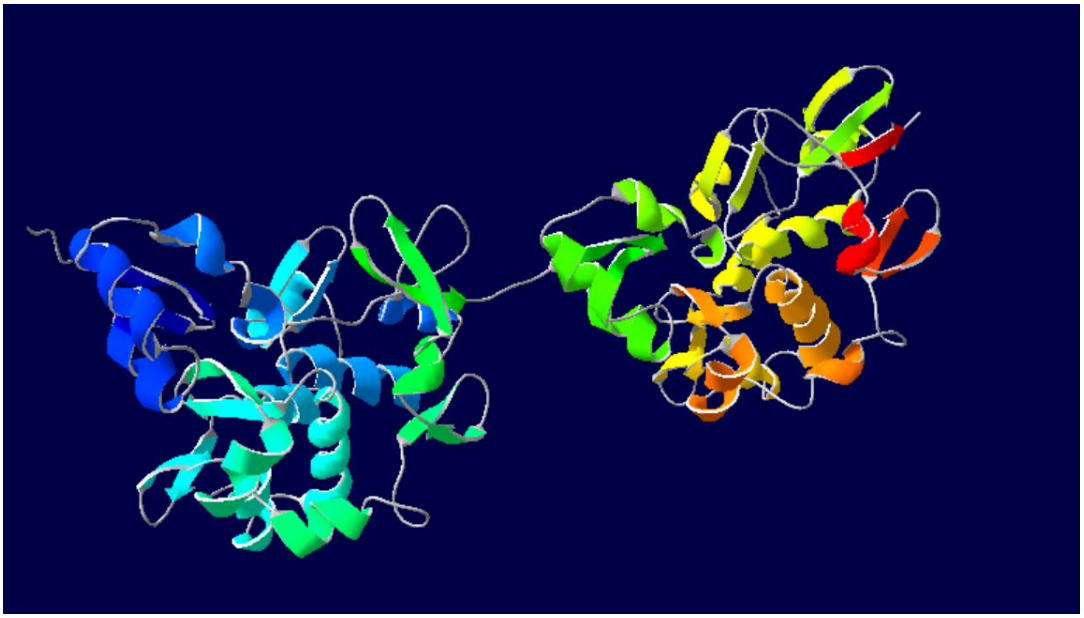


Figura 20. Immagine estrapolata dallo Swiss-PDB Viewer utilizzando la sequenza della desmoplachina presente nella banca dati PDB. In questa immagine si può osservare la forma a manubrio che adotta questa proteina e si possono apprezzare anche le strutture secondarie (α-elica e foglietti β). La colorazione è stata ottenuta mediante la funzione Color Secondary Structure

Succession attiva sul Ribbon, la sequenza amminoacidica viene colorata dal blu, che rappresentano i primi amminoacidi, al rosso, che sono gli ultimi.

Il dominio della testa globulare N-terminale della desmoplachina è composto da una serie di fasci α-elicoidali ed è richiesto sia per la localizzazione alla placca desmosomica che per l'interazione, mediante i sottodomini *NN*, *Z*, *Y*, *X*, *W* e *V*, con la regione N-terminale delle proteine armadillo, placofilina 1 e placoglobina, nonché per la desmocollina e la desmogleina[32].

La regione C-terminale della desmoplachina è composta da tre domini ripetuti delle plachine (PRD) chiamati *A*, *B* e *C* essenziali per l'ancoraggio dei filamenti intermedi ai dsmosomi[33,34] (Figura 21).

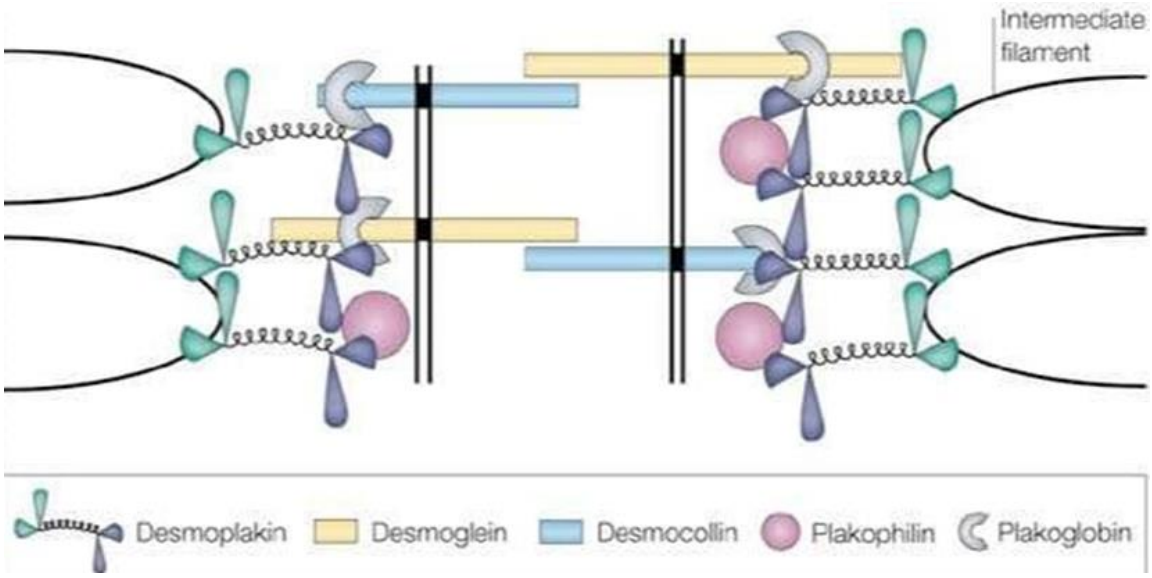


Figura 21. Raffigurazione che mostra le interazioni della desmoplachina nella sua porzione N-terminale, in viola, con la desmocollina e placoglobina e nella porzione C-terminale, in verde, con i filamenti intermedi[35].

All'estremità C-terminale vi è anche una regione ricca in glicina-serina-arginina. È stato dimostrato che il passaggio dallo stato fosforilato a quello non fosforilato della serina presente in questo dominio influenza l'interazione tra la desmoplachina e i filamenti intermedi[36].

Nella regione centrale della desmoplachina vi è un dominio coiled-coil, chiamato dominio *rod*, responsabile della omodimerizzazione[37] e quindi della forma a manubrio che assume la proteina (Figura 22).

Figura 22. Rappresentazione schematica della desmoplachina[35]. Nella regione N-terminale, in viola, si possono osservare i sottodomini necessari all'interazione con le proteine armadillo; a questa regione segue il dominio *rod* responsabile della omodimerizzazione che conferisce alla proteina la tipica struttura a manubrio ed infine i tre domini al C-terminale necessari per l'interazione con i filamenti intermedi.

La desmoplachina svolge numerose finzioni ed è essenziale per la normale adesione desmosomica. Principalmente interagisce con i filamenti intermedi di cheratina nelle cellule epiteliali, con i filamenti intermedi di desmina nei cardiomiociti e con i filamenti intermedi di vimentina nell'aracnoide e nelle cellule follicolari dendridiche. Fino a poco tempo fa si pensava che la desmoplachina interagisse esclusivamente con il citoscheletro del filamento intermedio. Tuttavia, recenti esperimenti hanno dimostrato che può anche essere coinvolta nella regolazione dell'organizzazione dei microtubuli; infatti, sembra che il reclutamento della proteina di ancoraggio dei microtubuli dipendente dalla desmoplachina sia essenziale per questo processo[29].

La desmoplachina esiste in tre isoforme:

- DSPI, con un peso molecolare di 332,0 kDa, è l'isoforma predominante espressa nel muscolo cardiaco;
- DSPII, con un peso molecolare di 260,0 kDa, mostra un dominio *rod* ridotto rispetto alla DSPI, ma comunque funzionale e viene espressa per lo più nell'epitelio stratificato e molto meno a livello cardiaco rispetto alla isoforma DSPI[38].
- DSPIa, recentemente individuata, è la isoforma minore. *L*a DSPIa non è stata descritta in precedenza, presumibilmente a causa di molteplici fattori principalmente derivanti dal basso livello di espressione rispetto al DSPI e al DSPII, ma anche dalla difficoltà di separare la proteina DSPII dal DSPIa usando tecniche standard di Western blotting[39].

La desmoplachina è espressa in diversi tessuti e cellule, principalmente nelle cellule epiteliali squamose e nel tessuto muscolare cardiaco (Figura 23).

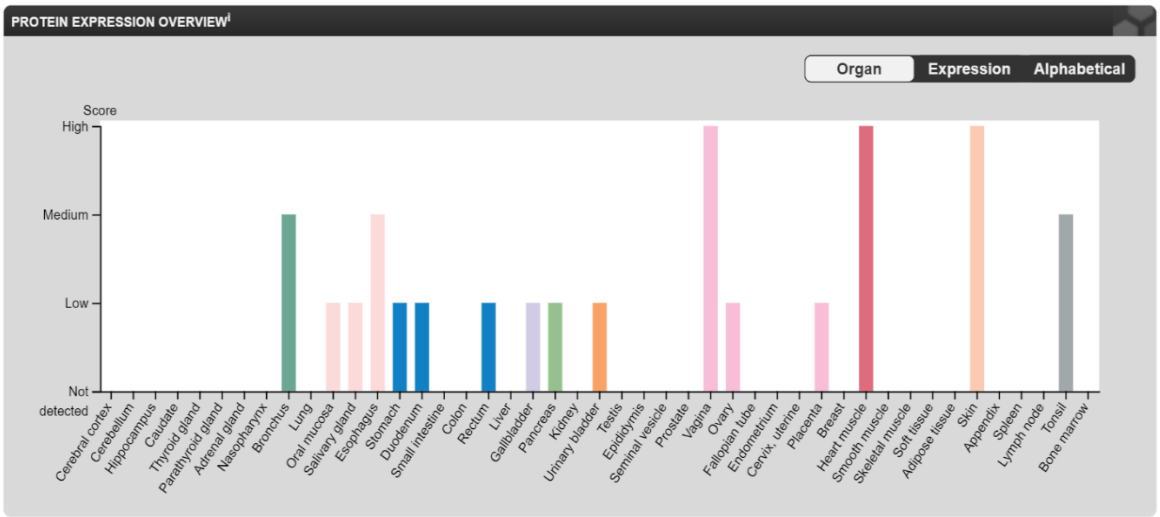


Figura 23. Immagine estrapolata dalla piattaforma The Human Protein Atlas, che mostra una panoramica dell'espressone della desmoplachina. Si vede come i livelli maggiori di espressione sono a livello dei tessuti in cui sono presenti cellule epiteliali squamose e il tessuto muscolare cardiaco.

Per quanto riguarda, invece, l'isoforma DSPIa sono stati condotti degli studi in cui è stato osservato che la sua espressione a livello tissutale era simile a quella delle altre due isoforme (DSPI e DSPII), ma sorprendentemente, a differenza delle altre due, l'isoforma DSPIa è stata rilevata anche a livello dell'aorta[39] (Figura 24).

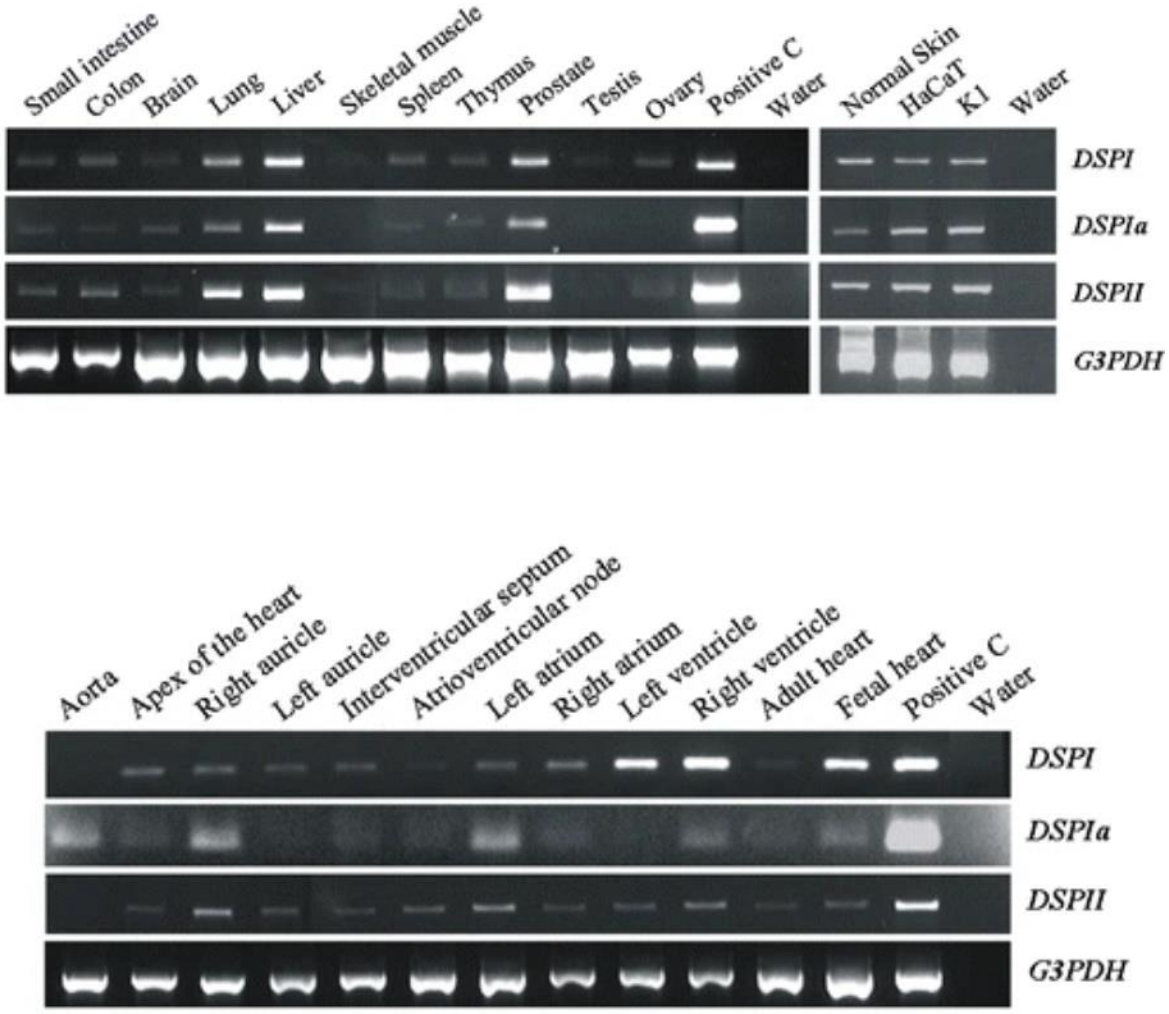


Figura 24. Amplificazione specifica delle tre isoforme DSP su un range di tessuti umani[40]. Si può osservare che la DSPIa è espressa per la maggior parte dei tessuti epiteliali, linee cellulari normali della pelle e dei cheratinociti con modelli di espressione simili a quelli di DSPI e di DSPII. L'isoforma DSPIa è espressa anche nel tessuto cardiovascolare, ma come si può vedere nella seconda figura è l'unica isoforma rilevata a livello dell'aorta. G3PDH è stato utilizzato come controllo negativo.

Capitolo 2

LE CARDIOMIOPTIE CORRELATE A VARIANTI DELLA DESMOPLACHINA

2.1 Varianti del gene e malattie correlate

Nel corso degli anni sono stati effettuati numerosi studi, in cui è emersa la correlazione di diverse mutazioni a carico del gene della desmoplachina con numerose patologie. Tali mutazioni si traducono in disturbi umani che variano da difetti della pelle e capelli senza coinvolgere il miocardio a una malattia letale della pelle fino alla cardiomiopatia aritmogena, un disturbo cardiaco che porta a morte improvvisa[41]. Le malattie studiate negli anni associate a mutazioni della demoplachina sono:

- *Cheratosi palmoplantare striata II*

  La prima mutazione eterozigote nel gene DSP venne descritta nel 1999 in una famiglia con una *forma striata di cheratoderma palmoplantare ereditario*, indicato poi come *tipo II*[42]. La mutazione consiste in una transizione C-T a livello dell'esone 4 del gene, che si traduce in un codone di stop prematuro nella regione N-terminale.

- *Sindrome di Carvajal*

  Nel 2000 venne descritta la prima mutazione umana recessiva del gene DSP, 7901delG, che si manifesta a livello fenotipico in una patologia che

prende il nome di *sindrome di Carvajal*. Questa sindrome si manifesta con cheratoderma striato generalizzato che colpisce prevalentemente l'epidermide palmoplantare, capelli lanosi e cardiomiopatia dilatata ventricolare sinistra[43]. In alcuni studi è stato osservato come un certo numero di pazienti con questo disturbo sindromico hanno subito un'insufficienza cardiaca durante la loro adolescenza, con conseguente morte precoce.

- *Cardiomiopatia dilatata con capelli lanosi e cheratoderma*

  Nel 2006 è stato descritto il caso di un paziente con cardiomiopatia dilatata aritmogena ereditaria (dovuta a mutazione recessiva) con coinvolgimento ventricolare sinistro e destro, cheratoderma palmoplantare epidermolitico e capelli lanosi[44]. Il paziente ha mostrato un grave fenotipo cardiaco con esordio precoce e rapida progressione all'insufficienza cardiaca all'età di quattro anni. In questo paziente è stata osservata una mutazione, R1267X, nell'esone 23 del gene della desmoplachina, che ha portato ad una proteina tronca dell'isoforma maggiore DSPI. Questa mutazione ha causato la perdita della maggior parte del dominio dell'asta e dell'area C-terminale, confermata dal Western blotting e dall'immunofluorescenza.

- *Cardiomiopatia dilatata con capelli lanosi, cheratoderma e agenesia dentale*

  Il primo caso è stato osservato nel 2006 in una paziente femmina, in cui è stata identificata un'inserzione di 30 bp nel gene DSP[45]. Nel 2016 in uno studio in cui sono stati analizzati i dati dell'esoma di una coorte di 496 pazienti con disturbi della cheratinizzazione, sono stati identificati 3 bambini non correlati con cardiomiopatia dilatata, capelli lanosi, eritrocheratoderma e agenesia dentale che sono eterozigoti per mutazioni missenso nel gene DSP: Q616P, H618P e L622P.

- *Fragilità cutanea – Sindrome dei capelli lanosi*

  In questa patologia è stata dimostrata l'etrezigosi per una combinazione di mutazione nonsenso e missenso in due proband studiati, da CYS809 a TER/ASN287 a LYS (C809X/N287K) e GLN664 a TER/ARG2366 a CYS (Q664X/R2366C). È stato, inoltre, dimostrato che portatori eterozigoti di queste mutazioni non mostrano anomalie fenotipiche anche se ci si aspetterebbe che tali mutazioni possano causare aploinsufficienza di desmoplachina[46].

- *Displasia ventricolare destra aritmogena familiare 8*

  In 1 delle 16 famiglie osservate nel nord Italia con displasia ventricolare destra aritmogena mappata al cromosoma 6p24 è stato osservato che i

membri colpiti presentavano eterozigosi per una mutazione missenso (S299R) a livello dell'esone 7 del gene DSP[47]. Quest mutazione è localizzata nel dominio N-terminale, coinvolto nel legame con la placoglobina e nella formazione del complesso con le caderine desmosomiali. L'assenza di difetti cutanei nei portatori eterozigoti della mutazione S299R può essere spiegata considerando che questa mutazione non influisce sul legame della desmoplachina con i filamenti intermedi, legame che viene meno in presenza di altre mutazioni che determinano un fenotipo caratterizzato da cheratoderma.

- *Epidermolisi acantolitica letale Bullosa*

  Nel 2005 uno studio ha evidenziato in un paziente (bambino morto all'età di 10 giorni) con grave fragilità della pelle e delle mucose, la correlazione di tale disordine con il troncamento genetico della coda della desmoplachina[48]. La fragilità della pelle era accompagnata da alopecia universale e perdita delle unghie.

2.2 Generalità sulle cardiomiopatie

Le cardiomiopatie sono un gruppo diversificato di disturbi cardiaci che spesso causano insufficienza cardiaca e morte, caratterizzate da un'alterazione anatomo-funzionale che riduce la capacità contrattile del cuore.

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) presentano 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"[49].

La classificazione dell'AHA del 2006 distingue le cardiomiopatie in:

- Primitive con interessamento prevalente o esclusivo del miocardio ventricolare senza manifestazioni sistemiche e si suddividono ulteriormente in genetiche, miste ed acquisite.
- Secondarie in cui l'interessamento miocardico è parte di un'ampia varietà di disordini nel contesto di una malattia sistemica come ischemia miocardia, infiammazione, infezione, aumentata pressione del miocardio e agenti tossici.

Un secondo criterio di classificazione tiene conto, invece, delle modifiche funzionali e strutturali a carico del cuore, dividendo pertanto la patologia in:

- Cardiomiopatia dilatativa, definita anche come "Malattia del Muscolo Cardiaco caratterizzata da dilatazione e ridotta contrattilità del ventricolo sinistro o di entrambi i ventricoli"[49], è una malattia genetica familiare che può manifestarsi in pazienti di tutte le età, ma l'esordio avviene nella maggior parte dei casi tra i 20 ed i 50 anni interessando prevalentemente il sesso maschile. Il danno alle cellule miocardiche è il fattore iniziale che porta alla morte cellulare e quindi ad una ridotta contrattilità del miocardio con conseguente attivazione di meccanismi di compenso che aiutano a mantenere una gittata cardiaca adeguata in un primo momento; con il progredire del danno miocardico questi meccanismi di compenso non saranno più sufficienti e si determinerà uno scompenso cardiaco. Clinicamente questa patologia si presenta in genere con scompenso cardiaco, per lo più sinistro, o con aritmie ventricolari o sopraventricolari. Un dolore toracico, per lo più da sforzo e talora con le caratteristiche di un'angina, rappresenta il sintomo principale d'esordio della patologia.
- Cardiomiopatia ipertrofica è una patologia primaria del miocardio che mostra carattere di ereditarietà autosomica dominante ed è caratterizzata

da ipertrofia del ventricolo, provocata dalla mutazione di geni codificanti proteine del sarcomero, incluse la Myosin Binding Protein C (MYPBC), la catena leggera della miosina regolatoria (MLC-1), la Troponina cardiaca T (cTnT), la Troponina cardiaca I (cTnI), la alfa-tropomiosina, la actina e la titina[50]. Questo ha portato a considerare la cardiomiopatia ipertrofica come una patologia del sarcomero. L'esistenza di un substrato genetico è un aspetto molto importante in quanto oltre a confermare o escludere la presenza della malattia in casi atipici, implica la possibilità di identificare i portatori delle mutazioni in fase preclinica attraverso uno screening familiare. Nella maggior parte dei casi l'ipertrofia del ventricolo sinistro è asimmetrica, interessa cioè in modo preponderante il setto interventricolare e in misura minore la parete libera del ventricolo sinistro, meno frequentemente ha una distribuzione uniforme nel ventricolo sinistro (ipertrofia concentrica), mentre in una piccola percentuale di pazienti l'ipertrofia è predominante a livello dell'apice del cuore. La cardiomiopatia ipertrofica 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. Il decorso clinico nella cardiomiopatia ipertrofica è estremamente variabile: alcuni pazienti restano asintomatici per tutta la vita, alcuni vanno incontro a severi sintomi di scompenso cardiaco, altri muoiono improvvisamente spesso in assenza di una precedente sintomatologia.

- Cardiomiopatia restrittiva è molto meno comune, non è sempre un disturbo cardiaco primario in quanto spesso associata con patologie sistemiche; anche se la causa è solitamente sconosciuta, può sorgere come conseguenza di disturbi sistemici o genetici. La cardiomiopatia restrittiva è caratterizzata da pareti ventricolari scarsamente distensibili, che si oppongono al riempimento diastolico; possono essere colpiti entrambi i ventricoli o uno solo, più frequentemente il sinistro[51]. I sintomi della cardiomiopatia restrittiva sono la dispnea da sforzo, l'ortopnea, dispnea parossistica notturna e l'edema periferico. L'astenia deriva da una gittata cardiaca fissa dovuta alla resistenza al riempimento ventricolare diastolico, sono frequenti le aritmie atriali e ventricolari e il blocco atrioventricolare; mentre l'angina e la sincope sono rare.

- Cardiomiopatia aritmogena è una malattia ereditaria del muscolo cardiaco caratterizzata da anomalie funzionali e strutturali del ventricolo destro. Le recenti acquisizioni escludono che questa sia una malattia esclusivamente del ventricolo destro; infatti, la parte ventricolare destra del cuore verrebbe colpita come la parte sinistra ma darebbe più precocemente manifestazione di sé per ragioni di tipo morfologico-strutturale e biofisico. Infatti, la parete muscolare del ventricolo destro è più sottile di quella del ventricolo sinistro, questo comporta che lo stress parietale destro sia maggiore del sinistro. Ad oggi si ritiene che la cardiomiopatia aritmogena sia una malattia dei desmosomi, questo spiegherebbe non solo le manifestazioni cardiache ma anche le manifestazioni extracardiache che troviamo nelle forme sindromiche. Dal punto di vista patologico, la malattia presenta una progressiva perdita di cardiomiociti che formano la parete libera del miocardio, con conseguente sostituzione adiposa o fibro-adiposa[51]. Questo cambiamento causa disomogeneità nel tessuto alterandone le caratteristiche di conduzione dell'impulso elettrico. Si creano in tal modo circuiti anatomici di rientro che danno origine ad aritmie ventricolari, le quali possono provocare fibrillazione ventricolare e morte improvvisa. La forma classica e più frequente si presenta in modo

autosomico dominante a penetranza incompleta, che porta alla presenza di soggetti portatori di mutazione e clinicamente affetti accanto ad individui portatori di mutazione ma senza diagnosi clinica di malattia, detti "portatori sani". Inoltre, il fenotipo, ovvero la manifestazione clinica della malattia, è variabile, ovvero può essere presente sia in forma grave che in forma moderata o lieve, sia in termini di alterazione anatomica che di instabilità elettrica.

2.3 Il ruolo della genetica nel quadro delle cardiomiopatie

L'eterogeneità clinica e genetica è una caratteristica cardine delle cardiomiopatie ereditarie e potenzialmente letali tra cui la cardiomiopatia ipertrofica, la cardiomiopatia dilatativa e la cardiomiopatia aritmogena del ventricolo desto, con più geni alla base dei loro meccanismi patogeni. I test genetici per molte di queste cardiomiopatie 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[51].

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 in una singola reazione. Questa tecnologia fornisce efficacemente una lista di ogni singola sostituzione nucleotidica e di ogni piccola inserzione/delezione (comune o rara, benigna o patogena) per ogni gene del genoma di un paziente, permettendo l'identificazione di nuove mutazioni correlate a patologie ereditarie.

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

- Il pannello multigenico include i geni appartenenti alla categoria di quelli riconosciuti fino ad oggi. Tuttavia, alcuni pannelli possono includere geni non appartenenti alla categoria di geni ad oggi noti; pertanto, 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 essere in grado di predire il genotipo. L'analisi della sequenza viene eseguita per prima e può essere seguita da un'analisi di delezione/duplicazione mirata al gene se non viene rilevata 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.

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, la loro validità clinica deve essere stabilita più ampiamente.

2.4 Varianti sul gene DSP correlate a cardiomiopatie

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.

2.4.1 Mutazioni: cenni generali

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.

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) e 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.

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.

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.

2.4.2 Varianti sul gene DSP

Le mutazioni nella codifica genica della desmoplachina sono state identificate per la prima volta in una forma autosomica recessiva di cardiomiopatia aritmogenica. Nel corso degli anni, sono stati effettuati numerosi studi di correlazione genotipo-fenotipo che hanno portato all'identificazione di numerose mutazioni nel gene DSP associate a cardiomiopatie, in particolare cardiomiopatia dilatativa e cardiomiopatia aritmogena. Allo stato attuale delle conoscenze l'individuazione delle mutazioni del gene DSP che risultano patogene è di fondamentale importanza per comprendere meglio le relazioni funzionali delle diverse parti della proteina e per tentare di stabilire correlazioni genotipo-fenotipo. Ad oggi sono state osservate circa 1859 varianti correlate a cardiomiopatie.

Capitolo 3

MATERIALI E METODI

3.1 Tecniche standard di studio delle varianti

Il primo approccio per lo studio delle varianti è rappresentato dall'utilizzo di databases e dalla consultazione della letteratura. Quando si utilizzano questi databases, i laboratori clinici devono valutare con quale frequenza essi vengono aggiornati e quali metodi sono stati utilizzati per l'inserimento dei dati; 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 dei dati, che potrebbe richiedere la lettura di pubblicazioni associate[53].

3.1.1 Il sequenziamento di 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 ossidile 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.

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[54] (Figura 25). In questo modo 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].

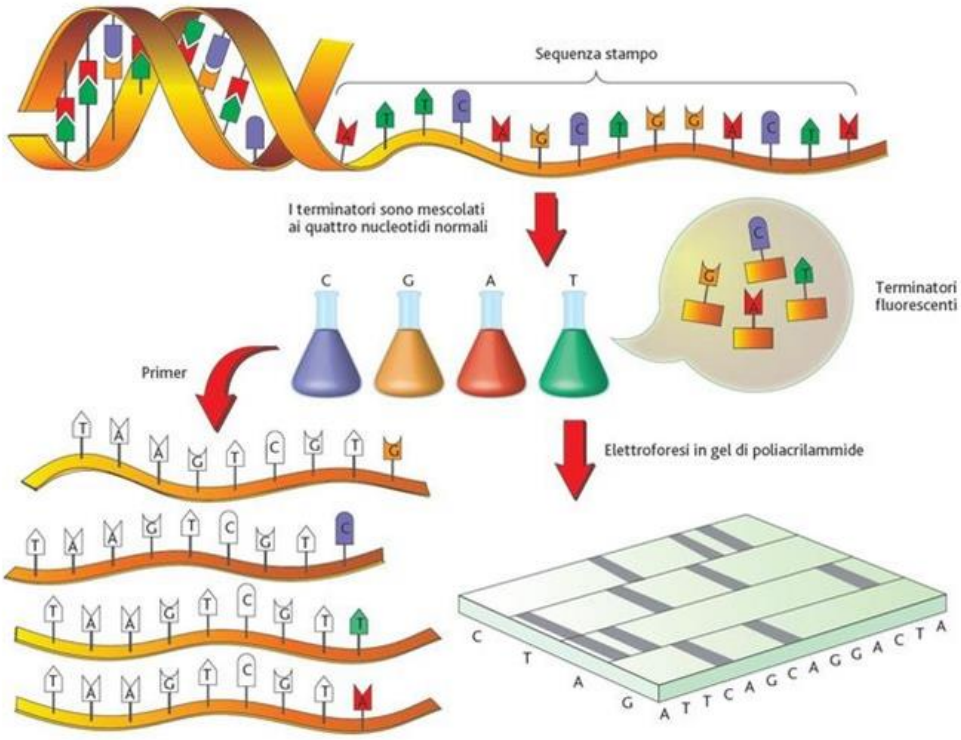


Figura 25. Immagine schematizzata degli step alla base del sequenziamento di Sanger[56].

3.1.2 Next Generation Sequencing

Negli ultimi anni, i ricercatori hanno sviluppato metodi e tecnologie sempre più innovative per determinare la sequenza nucleotidica da un campione biologico.

A partire dal metodo Sanger, introdotto nel 1977 che per diverso tempo ha rappresentato il "gold standard", le tecniche di sequenziamento sono state soggette a continua evoluzione.

Le tecniche di nuova generazione (*Next Generation Sequencing*), note anche come sequenziamento *high-throughput*, sono una serie di diverse tecnologie di sequenziamento moderne che consentono di ottenere le sequenze di DNA o RNA molto più rapidamente e con costi minori rispetto al sequenziamento di Sanger. La crescente potenza e la riduzione dei costi hanno portato ad un'enorme mole di applicazioni della tecnologia di sequenziamento di nuova generazione (NGS), che sta iniziando a diventare l'approccio standard da applicare in ambito clinico e di ricerca. A seguito della rapida evoluzione nelle tecnologie NGS, si è tentato di fare una classificazione di queste[57], differenziandole in:

- Tecniche di seconda generazione (2G), primi metodi di sequenziamento parallelo e massivo introdotti nel mercato. Esistono diversi metodi di sequenziamento di seconda generazione che differiscono per i processi biochimici alla base, ma che vengono accomunati da steps nella procedura. Il primo passo consiste nella preparazione di una libreria che solitamente viene ottenuta per frammentazione del DNA; a questi frammenti vengono aggiunti degli adattatori necessari ad immobilizzare i frammenti al supporto per gli step successivi di amplificazione e sequenziamento. Tra le tecniche 2G vi sono la tecnologia 454, che unisce la PCR in emulsione al pirosequenziamento[58]; la tecnologia

Illumina, in cui l'amplificazione clonale dei frammenti di DNA avviene secondo un procedimento a ponte, basato sulla cattura dei frammenti ripiegati ad arco che si ibridizzano ad un oligonucleotide adiacente; la tecnologia SOLiD, che sfrutta un meccanismo simile alla tecnologia 454; infine, la tecnologia Ion Torrent, che si basa sulla rilevazione diretta di ioni idrogeno che vengono rilasciati durante l'incorporazione delle singole basi da parte della DNA polimerasi.

- Tecniche di terza generazione (3G), note anche come metodiche a singola molecola di DNA, in quanto non richiedono una amplificazione del segmento da sequenziare, conferendo un enorme vantaggio nella loro applicazione. Le piattaforme attualmente in commercio sono la piattaforma Pacific Biosciences Single Molecule Real-Time (SMRTTM) e la piattaforma Helicos Biosciences true Single Molecule Sequencing (tSMS).
- Tecniche di quarta generazione (4G), sono le metodiche di sequenziamento più recenti, basate sull'utilizzo di una tecnologia a nanopori. In questo sistema la molecola di DNA viene fatta passare tra due camere riempite di una soluzione elettrolitica, tra cui è posta una membrana costituita da nanopori e all'interno di questa apertura il

passaggio della molecola causa una variazione del flusso di corrente ionica attraverso il poro.

I test eseguiti con piattaforme NGS generano una notevole quantità di informazioni. Infatti, a fronte di una riduzione dei tempi e dei costi di lavoro, risultano essere molto complesse le procedure di acquisizione e di elaborazione dei dati, soprattutto per la mole di informazioni che sono in grado di generare[59].

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.

3.2 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)[60]. Si possono pertanto 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, 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 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. Qualsiasi mutazione missenso non riportata precedentemente 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 patogenetiche.

  Questo gruppo contiene le mutazioni localizzate nella 5'UTR o 3'UTR, 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 (Figura 26), secondo le quali è importante andare a verificare diversi parametri che possono concorrere a spostare la diagnosi verso una forma benigna o patogenetica.

	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			

Figura 26. Tabella schematica dei criteri di cui tenere conto per la classificazione delle varianti secondo le linee guida dell'ACMG[60]. Per una determinata variante vengono selezionati i criteri in base alle prove osservate per la variante. I criteri vengono quindi combinati secondo regole di punteggio per scegliere una classificazione dal sistema. Le regole si applicano a tutti i dati disponibili su una variante, raccolti sul caso in esame o su dati pubblicati in letteratura. Abbreviazioni: BS, forte benigno; BP, supporto benigno; FH, storia familiare; LOF, perdita di funzione; MAF, frequenza allele minore; PM, moderato patogeno; PP, supporto patogeno; PS, forte patogeno; PVS, patogeno molto forte.

3.3 Materiali e Metodi

Sono stati sequenziati un gruppo di circa 180 casi provenienti dal reparto di cardiologia degli Ospedali Riuniti di Ancona, facenti parte di uno studio più ampio riguardo le malattie cardiache. I campioni sono campioni di sangue periferico raccolto in EDTA, da cui viene estratto il genoma e analizzato per un pannello di 101 geni, sui quali la presenza di mutazioni può determinare patologie cardiache.

L'analisi è stata effettuata mediante tecniche NGS e la conferma delle varianti è stata effettuata mediante sequenziamento Sanger.

3.3.1 Tecnica NGS

3.3.1.1 Estrazione

L'estrazione del DNA è stata effettuata mediante QIAamp DNA Micro Kit, che semplifica l'estrazione del DNA da piccoli campioni mediante la tecnica delle colonnine. La procedura con questo kit consta di quatto steps (Figura 27):

- Lisi – il campione viene lisato in forti condizioni di denaturazione ad elevate temperature e in presenza di proteinasi K e Buffer ATL
- Legame – il DNA nel lisato si lega alla membrana della colonna QIAamp MinElute, per favorire il legame viene aggiunto al lisato il Buffer AL.

- Lavaggio – la membrana viene lavata per allontanare eventuali contaminanti usando il Buffer AW1 e AW2.
- Eluizione – il DNA viene eluito dalla membrana utilizzando un piccolo volume di Buffer AE o acqua distillata.

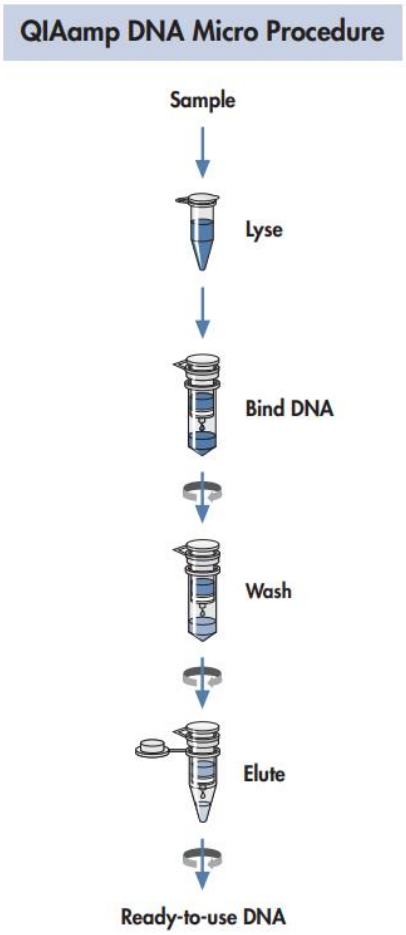


Figura 27. Passaggi schematizzati per l'estrazione del DNA

3.3.1.2 Quantificazione

La quantificazione viene effettuata utilizzando:

- Qubit® dsDNA HS Assay Kits con strumento Invitrogen Qubit – Qubit ThermoFisher Scientific.

  È necessario diluire il reagente utilizzando il tampone fornito, aggiungere il campione (qualsiasi volume compreso tra 1 µl e 20 µl è accettabile) e leggere la concentrazione usando il fluorometro Qubit. Il saggio presenta una performance ottimale a temperatura ambiente e fluttuazioni di temperatura possono influenzare l'accuratezza dell'analisi. Per consentire al saggio Qubit® di raggiungere una fluorescenza ottimale, incubare i preparati per i test del DNA per 2 minuti dopo aver miscelato il campione con la soluzione. Dopo questo periodo di incubazione, il segnale di fluorescenza è stabile per 3 ore a temperatura ambiente.

- Plexor® HY System – Preomega Corporation, saggio di PCR Real-Time, utilizzato con lo strumento per Real-Time Rotor Gene Q-Qiagen.

  Il kit contiene un controllo PCR interno (IPC) per testare i risultati falsi negativi che possono verificarsi in presenza di inibitori PCR e una funzione di curva di fusione per confermare che il prodotto corretto è stato amplificato. Il protocollo presuppone 2µl di templato per reazione:

  1. Scongelare Plexor® HY 2X Master Mix, Plexor® HY 20X Primer/IPC Mix and Water, Amplification Grade.

2. Vortexare Plexor® HY 2X Master Mix and Plexor® HY 20X Primer/IPC Mix per 10 secondi.
3. Determinare il numero di reazioni da impostare, incluse le reazioni di controllo senza templato. Aggiungi 1-3 reazioni a questo numero.
4. Preparare il mix di reazioni combinando acqua, Amplification Grade, Plexor® HY 2X Master Mix and Plexor® HY 20X Primer/IPC Mix.
5. Vortexare per 10 secondi per mescolare. Non centrifugare dopo il vortex.
6. Aggiungere 18µl di miscela di reazione nei pozzetti di una piastra che verrà utilizzata per la PCR.
7. Aggiungere 2µl di buffer TE^{-4} alle reazioni di controllo senza templato.
8. Aggiungere 2µl di campione di DNA nei pozzetti appropriati.
9. Sigillare le piastre e centrifugare brevemente.
10. Impostare e programmare lo strumento.
11. Posizionare la piastra PCR nello strumento e iniziare immediatamente il ciclo termico.

3.3.1.3 Preparazione delle librerie

Per la preparazione delle librerie viene utilizzato Ion AmpliSeq™ Library Kit 2.0, un kit progettato per la preparazione di librerie di ampliconi (Figura 28).

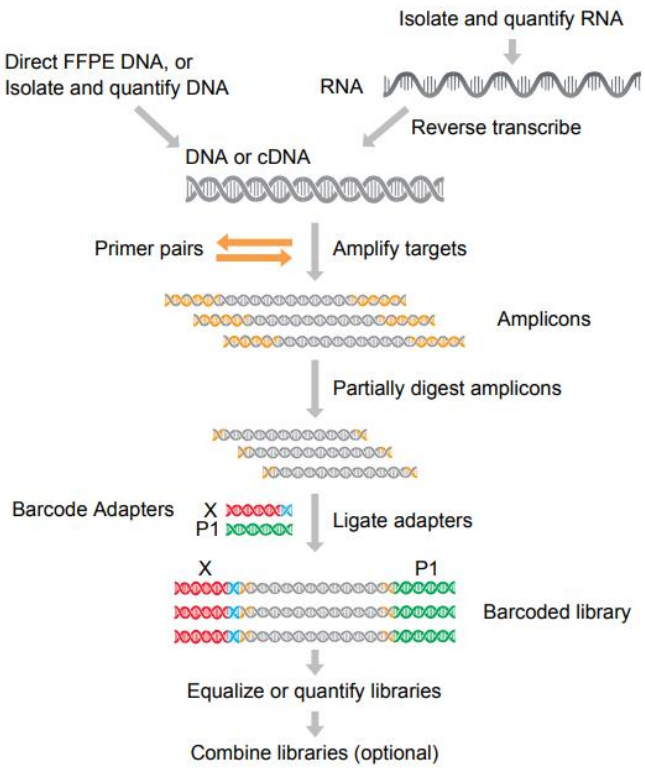


Figura 28. Schema della formazione delle librerie utilizzando Ion AmpliSeq™ Library Kit 2.0.

In questo studio è stata effettuata un'amplificazione con due pool. Il processo di preparazione delle librerie prevede:

1. Mescolare accuratamente Ion AmpliSeq™ HiFi Mix, DNA e Nuclease-free Water spipettando su e giù 5 volte, quindi trasferire il campione master mix in 2 pozzetti di una piastra PCR a 96 pozzetti.
2. Sigillare la piastra con una pellicola adesiva trasparente MicroAmp™. Per prevenire l'evaporazione durante l'amplificazione del target, utilizzare un applicatore per premere bene la pellicola attorno ogni pozzetto e al perimetro della piastra.
3. Posizionare un tampone MicroAmp™ Compression sulla piastra, quindi caricare la piastra nel termociclatore.
4. Procedere all'amplificazione del target.
5. Battere delicatamente la piastra su una superficie piana dura o centrifugare brevemente per raccogliere il contenuto in fondo ai pozzetti.
6. Rimuovere con cautela la guarnizione della piastra.
7. Per ogni campione, combinare le reazioni di amplificazione target da 10 µl. Il volume totale per ogni campione dovrebbe essere di 20 µl.
8. Procedere al passaggio di digestione degli ampliconi.
9. Aggiungere 2 µl di FuPa Reagent (tappo marrone) a ciascun campione amplificato. Il volume totale è ~22 µl.
10. Sigillare la piastra con una pellicola adesiva trasparente, vortexare accuratamente, quindi centrifugare per raccogliere eventuali goccioline.

11. Caricare la piastra nel termociclatore ed avviare la corsa.
12. Centrifugare brevemente la piastra per raccogliere il contenuto nel fondo dei pozzetti.
13. Rimuovere con cautela il sigillo della piastra, quindi aggiungere 45 µL di Agencourt™ AMPure™ XP Reagent a ciascuna libreria. Pipettare su e giù 5 volte per mescolare accuratamente la sospensione.
14. Incubare la miscela per 5 minuti a temperatura ambiente.
15. Posizionare la piastra in un supporto magnetico, quindi incubare per 2 minuti o finché la soluzione non si schiarisce. Rimuovere con attenzione, quindi scartare il surnatante senza prelevare anche il pellet.
16. Aggiungere 150 µl di etanolo al 70% appena preparato, quindi muovere la piastra da un lato all'altro nelle due posizioni del magnete per lavare le biglie. Rimuovere con attenzione, quindi scartare il surnatante senza prelevare il pellet e ripetere per un secondo lavaggio.
17. Assicurarsi che tutte le goccioline di etanolo vengano rimosse dai pozzetti. Mantenere la piastra all'interno del magnete, asciugare le biglie all'aria a temperatura ambiente per 5 minuti.
18. Procedere alla quantificazione delle librerie.

3.3.1.4 Quantificazione delle librerie

Per la quantificazione delle librerie è stato utilizzato Agilent High Sensitivity D1000 con lo strumento Agilent 2200 TapeStation nucleic acids system.

La procedura standard prevede l'allestimento di una provetta da 0.2 ml per ogni campione, avendo l'accortezza di aggiungere una provetta per il Genome DNA Ladder (una miscela di DNA con peso molecolare noto, che viene inserito nelle corse elettroforetiche per permettere all'operatore di risalire alle dimensioni dei frammenti processati. In ogni provetta vengono aggiunti 4 µl di Buffer e 2 µl di campione o Ladder. Dopo aver caricato la screentape ed i campioni nello strumento, si avvia la corsa (Figura 29).

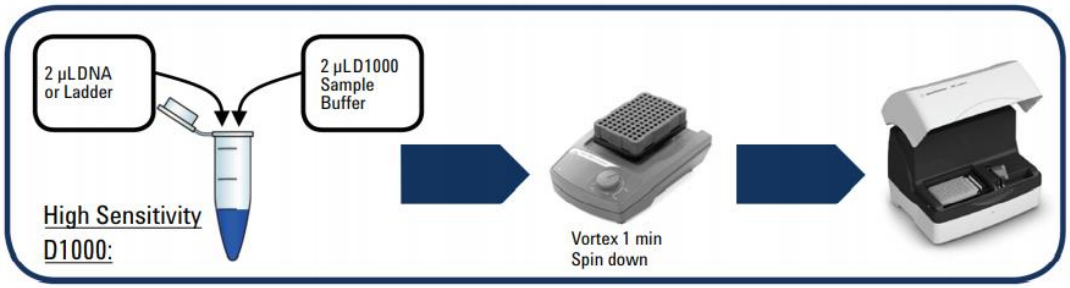


Figura 29. Step seguiti per la quantificazione delle librerie con lo strumento Agilent 2200 TapeStation nucleic acids system.

3.3.1.5 Sequenziamento

Per il sequenziamento viene utilizzato inizialmente Ion Chef System, utilizzato per la preparazione del campione al sequenziamento. Questo viene

combinato con Ion GeneStudio S5™ Systems, strumento utilizzato per il sequenziamento vero e proprio (Figura 30).

Figura 30. Strumenti utilizzati per il sequenziamento.

1. Caricare ogni chip che verrà utilizzato per i templati e sequenziamento nella centrifuga, quindi collegare a Chip Adapter per l'assemblaggio.
2. Caricare i gruppi adattatore/chip/bucket nella centrifuga di caricamento dei chip, assicurarsi del corretto bilanciamento della centrifuga e centrifugare.
3. Una volta effettuata la corsa rimuovere il chip e procedere al sequenziamento con Ion Gene Studio S5.

3.3.2 Conferma variante con Sanger

La conferma delle varianti è stata testata con il sequenziamento di Sanger, che prevede diversi steps.

3.3.2.1 Amplificazione

L'amplificazione è stata effettuata utilizzando il QIAGEN® Multiplex PCR Kit, allestendo una reaction mix che contiene tutti i componenti richiesti per una PCR, riportati nella tabella seguente.

Reaction mix	**Volume**
2x QIAGEN® Multiplex PCR Master Mix	5 µl
Primer Forward (0,2 µM)	1 µl
Primer Reverse (0,2 µM)	1 µl
DNA	3 µl
Volume totale	10 µl

Le condizioni di amplificazione sono riportate nella tabella seguente.

<table>
<tr><th colspan="5">PCR</th></tr>
<tr><td rowspan="3">95°C per
15 min</td><td colspan="3">30 cicli</td><td></td></tr>
<tr><td>Denaturazione</td><td>Annealing</td><td>Allungamento</td><td rowspan="2">60°C per
30 min</td></tr>
<tr><td>94°C per 30 min</td><td>65°C per 90 min</td><td>72°C per 60 min</td></tr>
</table>

3.3.2.2 Quantificazione dei prodotti di amplificazione

Tutti gli amplificati sono stati successivamente posti ad un *quality check*, per verificare che le sequenze di interesse siano state realmente amplificate. Per la quantificazione è stato utilizzato il sistema Agilent 2200 TapeStation nucleic acids system, descritto anche precedentemente per la quantificazione delle librerie nella procedura utilizzata per eseguire il sequenziamento.

3.3.2.3 Purificazione del DNA amplificato

Gli amplificati sono stati sottoposti ad una purificazione utilizzando il protocollo Exo-CIP PCR. Questo è un protocollo che viene utilizzato per purificare prodotti di PCR per il sequenziamento diretto. Si sfrutta l'azione di due enzimi per rimuovere i dNTP non incorporati e i primers in eccesso. Si basa sull'utilizzo di due enzimi: *exonuclease I* (Exo), che è un enzima che digerisce il DNA a singolo filamento come anche i primers e i prodotti di PCR incompleti; *Calf Intestinal alkaline Phosphatase* (CIP), che rimuove i dNTP che non sono stati incorporati nel DNA della PCR.

La purificazione è stata ottenuta incubando 1 µl di DNA con 9 µl di Exo-CIP. La quantità dei reagenti utilizzata è riportata nella tabella seguente.

Exo Mix		**CIP Mix**	
Buffer Exo	1,2 µl x 4 = 48 µl	Buffer CIP	1,25 µl x 4 = 5 µl
Exo I	0,3 µl x 4 = 1,2 µl	CIP	1,25 µl x 4 = 5 µl
H_2O	11 µl x 4 = 44 µl	H_2O	10 µl x 4 = 40 µl

3.3.2.4 Reazioni di sequenza

Il sequenziamento è stato eseguito con il BigDye® Terminator v1.1 Sequencing Buffer Kit della ThermoFisher. La soluzione per la reazione di sequenza è stata ottenuta unendo i vari componenti come riportato nella tabella successiva.

Reaction mix	**Volume**
BigDye Sequencing Buffer	2 µl
5x BigDye Terminator Sequencing Buffer	1 µl
Primer Forward o Reverse	1 µl
DNA	1 µl
H_2O	5 µl
Volume totale	10 µl

Sono stati effettuati 25 cicli di denaturazione a 96°C per 1 minuto e 96°C per 10 secondi, l'annealing a 50°C per 5 secondi e l'allungamento a 60°C per 4 minuti.

3.3.2.5 Rimozione dei ddNTPs non incorporati

Prima di sottoporre i prodotti del sequenziamento ad elettroforesi capillare, si è proceduto alla purificazione in piastra, rimuovendo i ddNTPs non incorporati. Per questo step è stato utilizzato il BigDye® X-Terminator Purification Kit, che prevede la preparazione di un premix per ogni campione composto da SAM Solution (45 µl) e X-Terminator Solution (10 µl). Quest'ultima è una resina contenente microbiglie che catturano i ddNTPs non incorporati, facendoli precipitare sul fondo della provetta e lasciando, quindi, in soluzione solo il DNA amplificato. Ciò avviene anche grazie al vortex di 30 minuti e alla successiva centrifugazione per 2 minuti a 3500 rpm.

3.3.2.6 Elettroforesi capillare

L'elettroforesi capillare è necessaria per visualizzare i prodotti del sequenziamento Sanger. La piastra contenente i prodotti di sequenziamento purificati è stata caricata direttamente sullo strumento Applied Biosystems® 3130 Genetic Analyzer.

Si è eseguita, infine, l'analisi dei dati usando una pipeline bioinformatica personalizzata e 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 benigna, una mutazione probabilmente o certamente patogenetica, oppure una mutazione di incerto significato, attenendosi alle linee guida del "The American College of Medical Genetics and Genomics" (ACMG) e grazie anche a tools bioinformatici (Varsome e InterVar), che forniscono i mezzi per interpretare le varianti riscontrate.

Capitolo 4

RISULTATI E DISCUSSIONE

4.1 Risultati

Dei campioni inviati dal reparto di cardiologia degli Ospedali Riuniti di Ancona nell'anno 2020, si riportano di seguito le due sequenze in cui è stata osservata una variante sul gene della desmoplachina. Entrambi i campioni sono stati trattati con la procedura descritta nel paragrafo dei materiali e metodi e i dati sono stati analizzati con software bioinformatici.

4.1.1 Paziente 1

Il primo caso riguarda un paziente di sesso femminile affetta da cardiomiopatia dilatativa (CMD) e capelli lanosi, si sospetta pertanto che sia affetta da Sindrome di Carvajal, nonostante non presenti interessamento dell'epidermide palmoplantare. Questa patologia in letteratura è descritta essere a trasmissione autosomica recessiva, perciò la mutazione del gene deve essere presente su entrambi gli alleli, in omozigosi.

L'indagine molecolare è stata condotta mediante un'analisi di sequenza di un pannello genico, che comprende 101 geni coinvolti in patologie cardiache ereditarie tra cui i geni coinvolti nella cardiomiopatia dilatativa.

Dall'analisi è risultata la variane DSP: NM_001319043: c.3739C>T: p.Arg1247* in omozigosi (Figura 31).

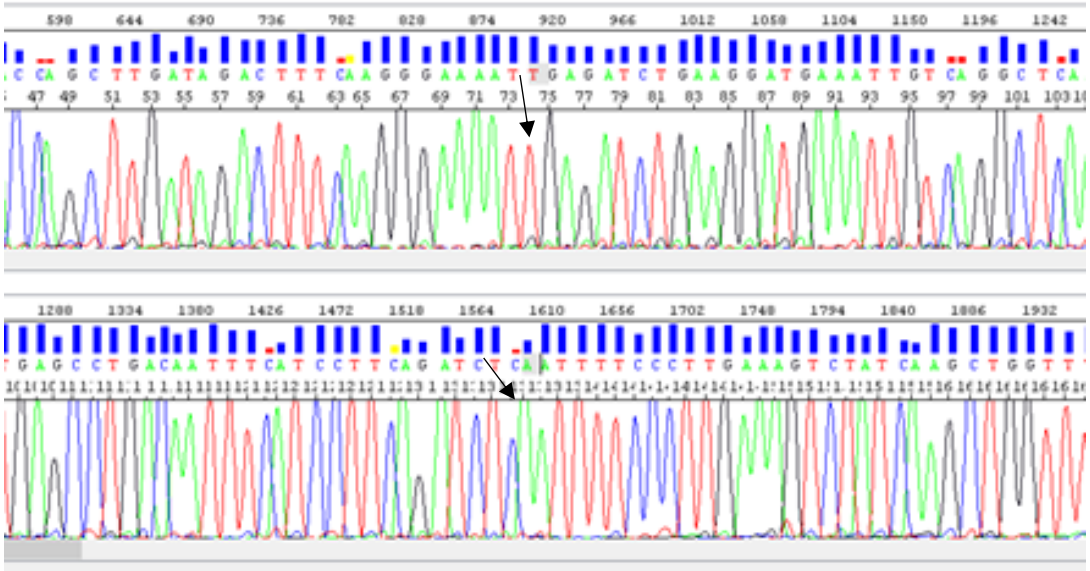


Figura 31. Nell'immagine è evidenziata in giallo la mutazione nel gene DSP, ritenuta responsabile della cardiomiopatia dilatativa nella Sindrome di Carvajal.

La mutazione è una sostituzione nonsenso del gene (C>T), che porta alla sostituzione di un amminoacido in un codone di stop. Questa mutazione porta alla sintesi di una proteina tronca che perciò non è in grado di svolgere la sua funzione. La variante osservata è presente sull'esone 23 del gene ed è una sostituzione che porta ad una mutazione nonsenso. È poco frequente nella popolazione, risultato che rispecchia le aspettative trattandosi di una variante che si traduce in una patologia rara e corrisponde ad un polimorfismo osservato che nel database SNP è designato con il codice rs1413961070 (Figura 32).

Figura 32. Immagine che raccoglie le varie informazioni riguardo la variante osservata, come il tipo di mutazione, l'esone in cui è presente la mutazione, la frequenza nella popolazione e studi in cui è stata osservata la medesima variante.

I risultati ottenuti sono poi stati interpretati con i tools bioinformatici Varsome (Figura 33) e InterVar (Figura 34) secondo le linee guida dell'ACMG. Da questa valutazione è emerso che la variante non è conosciuta in letteratura ed è classificabile come patogenetica per i seguenti criteri:

- PSV1: variante nonsenso presente nel gene DSP per il quale la perdita di funzione è un meccanismo patogenetico noto associato a cardiomiopatia

dilatativa, cardiomiopatia aritmogena del ventricolo destro e sindrome di Carvajal;

- PM2: variante rara nel database di frequenza GnomAD genomes (f=0.0032%) e assente in GnomAD exomes;
- PP3: cinque tools computazionali supportano la predizione di effetto deleterio della variante (BayesDel_addAF, DANN, EIGEN, FATHMM-MKL, MutationTaster) e nessuno supporta predizione di effetto benigno.

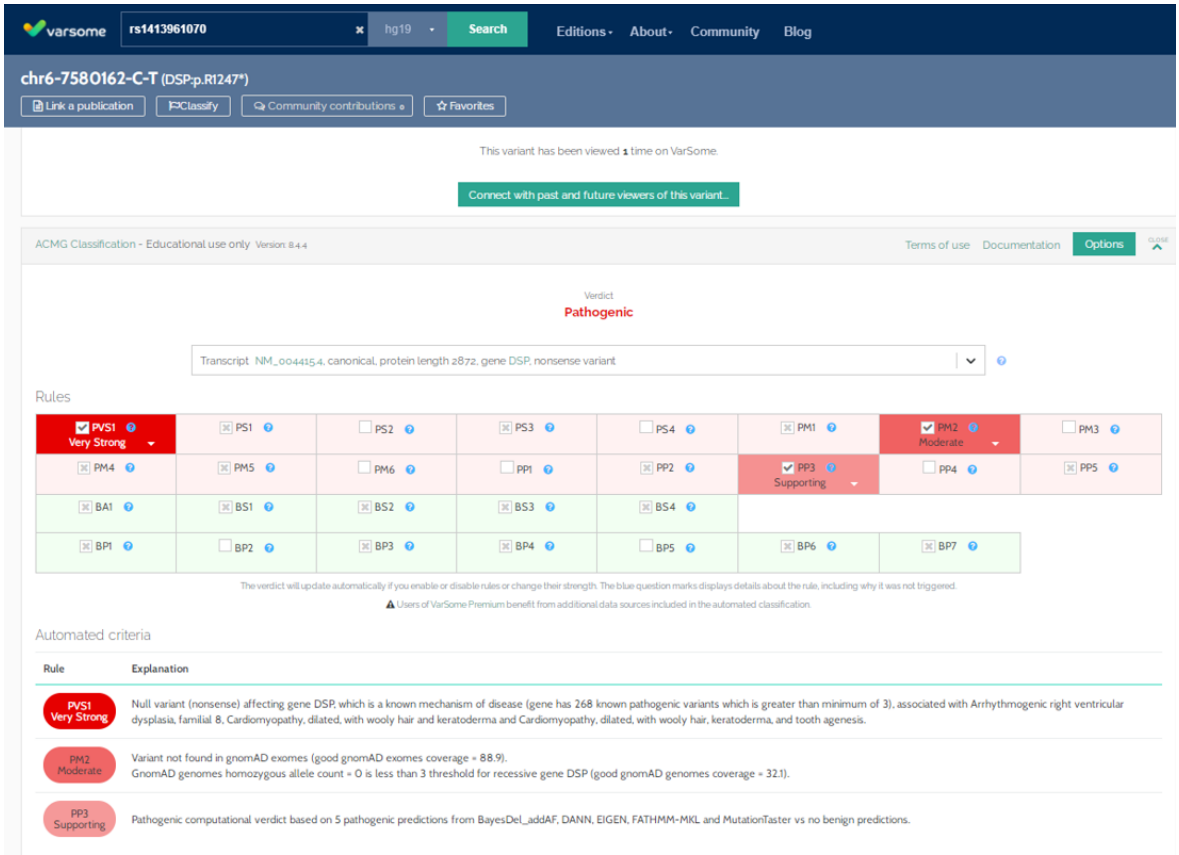


Figura 33. Analisi della patogenicità della variante mediante il tool Varsome. Vengono presi in considerazione anche altri tools computazionali come: DANN (Deleterious Annotation of genetic

variants using Neural Networks) è un tool informatico che ci permette di riconoscere varianti patogenetiche a partire da mutazioni genetiche, soprattutto non codificanti e si basa su un algoritmo detto “neural network”, i valori che otteniamo come risultato vanno da 0 a 1, con 1 che indica le varianti più pericolose. 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. 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.

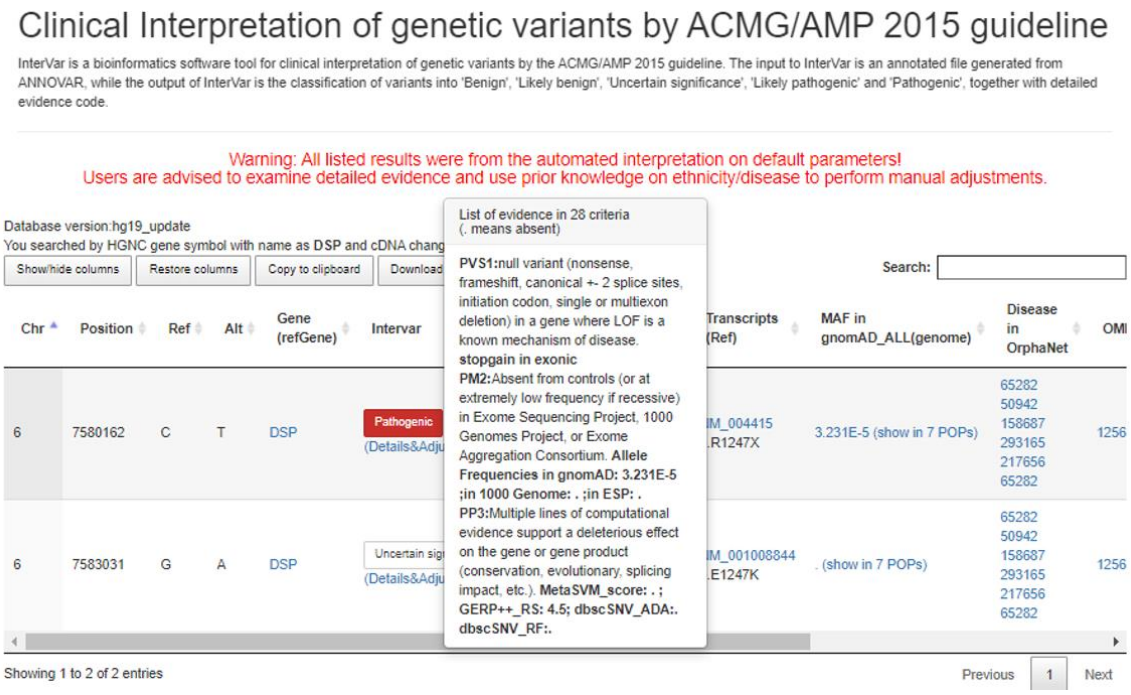
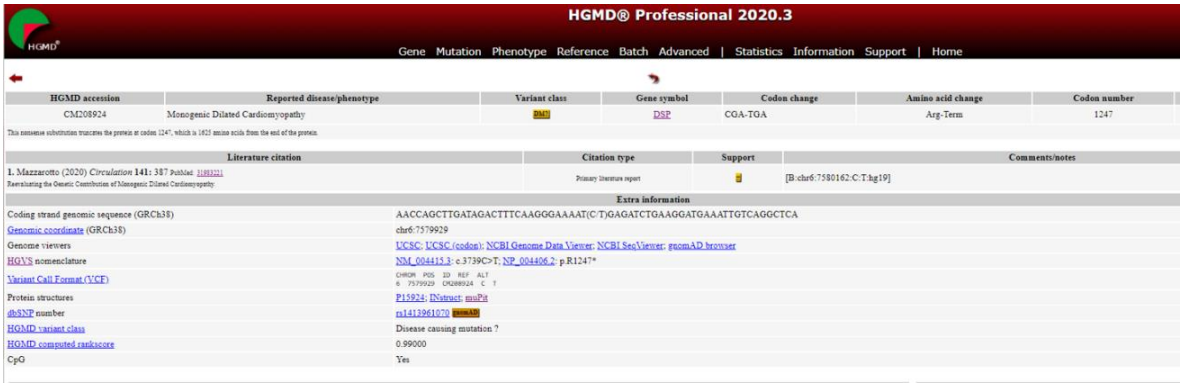


Figura 34. Immagine del tool InterVar per l’interpretazione clinica delle varianti genetiche secondo le linee guida ACMG.

Le caratteristiche riguardanti la variante osservata (Figura 35) sono riportate nel Human Gene Mutation Database (HGMD®), una raccolta completa di mutazioni patogenetiche associate a malattie ereditarie umane. HGMD, originariamente creato per la comprensione dei meccanismi mutazionali nei geni umani, ha ora ampliato il suo servizio includendo un'aggiornata e completa fonte bibliografica relativa ai vari tipi di mutazioni geniche associate a malattie ereditarie. HGMD annota sia mutazioni patogenetiche che polimorfismi funzionali/associati a malattie, riportati in letteratura, consentendone l'accesso a chi opera in istituzioni accademiche, cliniche o commerciali. Esso è il principale database in cui sono annotate mutazioni puntiformi in regioni codificanti (missenso e nonsenso), mutazioni in sequenze regolatorie e di splicing, micro-delezioni e micro-inserzioni, inserzioni/delezioni (indels), mutazioni da espansione di motivi ripetuti, grosse alterazioni geniche (delezioni, inserzioni e duplicazioni) e riarrangiamenti genici molto complessi.



HGMD® Professional 2020.3

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

HGMD accession	Reported disease/phenotype	Variant class	Gene symbol	Codon change	Amino acid change	Codon number
CM208924	Monogenic Dilated Cardiomyopathy	DM?	DSP	CGA-TGA	Arg-Term	1247

This nonsense substitution truncates the protein at codon 1247, which is 1625 amino acids from the end of the protein.

Literature citation	Citation type	Support	Comments/notes
1. Mazzarotto (2020) *Circulation* **141:** 387 PubMed: 31983221 Reevaluating the Genetic Contribution of Monogenic Dilated Cardiomyopathy.	Primary literature report		[B:chr6:7580162:C/T:hg19]

Extra information	
Coding strand genomic sequence (GRCh38)	AACCAGCTTGATAGACTTTCAAGGGAAAAT(C/T)GAGATCTGAAGGATGAAATTGTCAGGCTCA
Genomic coordinate (GRCh38)	chr6:7579929
Genome viewers	UCSC; UCSC (codon); NCBI Genome Data Viewer; NCBI SeqViewer; gnomAD browser
HGVS nomenclature	NM_004415.3: c.3739C>T; NP_004406.2: p.R1247*
Variant Call Format (VCF)	CHROM POS ID REF ALT 6 7579929 CM208924 C T
Protein structures	P15924; INstruct; mutPit
dbSNP number	rs1413961070 gnomAD
HGMD variant class	Disease causing mutation ?
HGMD computed rankscore	0.99000
CpG	Yes

Figura 35. Immagine del Database HGMD che riporta le caratteristiche riguardanti la variante osservata nel paziente a partire dalle caratteristiche generali come la posizione sul cromosoma, la sequenza genomica, la mutazione, la modifica al codone e dell'amminoacido dovuta alla mutazione, il codice di accesso al dbSNP e i riferimenti ad altri studi, qualora fossero presenti.

4.1.2 Paziente 2

Il secondo caso è relativo ad un paziente che presenta cardiomiopatia dilatativa (CDM). Il campione di sangue prelevato dal paziente è stato trattato seguendo i protocolli descritti nel paragrafo materiali e metodi. I risultati sono stati poi analizzati e interpretati facendo riferimento ai tools bioinformatici InterVar e Varsome.

L'indagine molecolare è stata condotta mediante un'analisi di sequenza di un pannello genico, che comprende 101 geni coinvolti in patologie cardiache ereditarie tra cui i geni coinvolti nella cardiomiopatia dilatativa.

Dall'analisi del sequenziamento è emerso che il soggetto presenta una la variante DSP: NM_001319034: c.3862A>C: p. (Lys1288Gln) in eterozigosi (Figura 36).

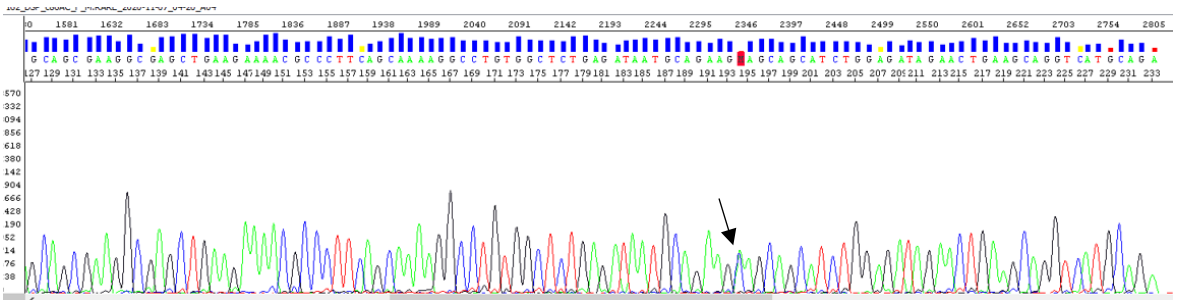


Figura 36. Analisi della sequenza in cui è possibile osservare la mutazione. È stato utilizzato come primer solo il forward poiché essendo una mutazione in eterozigosi il primer reverse non vedrebbe la variante.

La mutazione è una mutazione di tipo missenso (sostituzione di A in C). La variante osservata è presente nell'esone 23 del gene ed è una sostituzione che porta ad una variazione del codone che si traduce in una sostituzione dell'amminoacido presente in posizione 1288, da una lisina a una glutammina. Questa variante è stata osservata in numerosi altri studi ed è classificata nel database di polimorfismi sotto il codice rs138907450 (Figura 37).

Figura 37. Immagine che raccoglie le varie informazioni riguardo la variante osservata, come il tipo di mutazione, l'esone in cui è presente la mutazione, la frequenza nella popolazione e studi in cui è stata osservata la medesima variante.

I risultati ottenuti sono poi stati interpretati con i tools bioinformatici Varsome (Figura 38) e InterVar (Figura 39) secondo le linee guida dell'ACMG. Da questa valutazione è emerso che la variante osservata è nota in letteratura ed è classificabile come variante di incerto significato clinico (VUS) per i seguenti criteri:

- PM2: variante rara nei database di frequenza GnomAD genomes (f=0.000096) ed exomes (f=0.00027);
- PP5: fonti autorevoli registrano la variante come patogenetica (HGMD) o VUS (ClinVar, 10 submission)

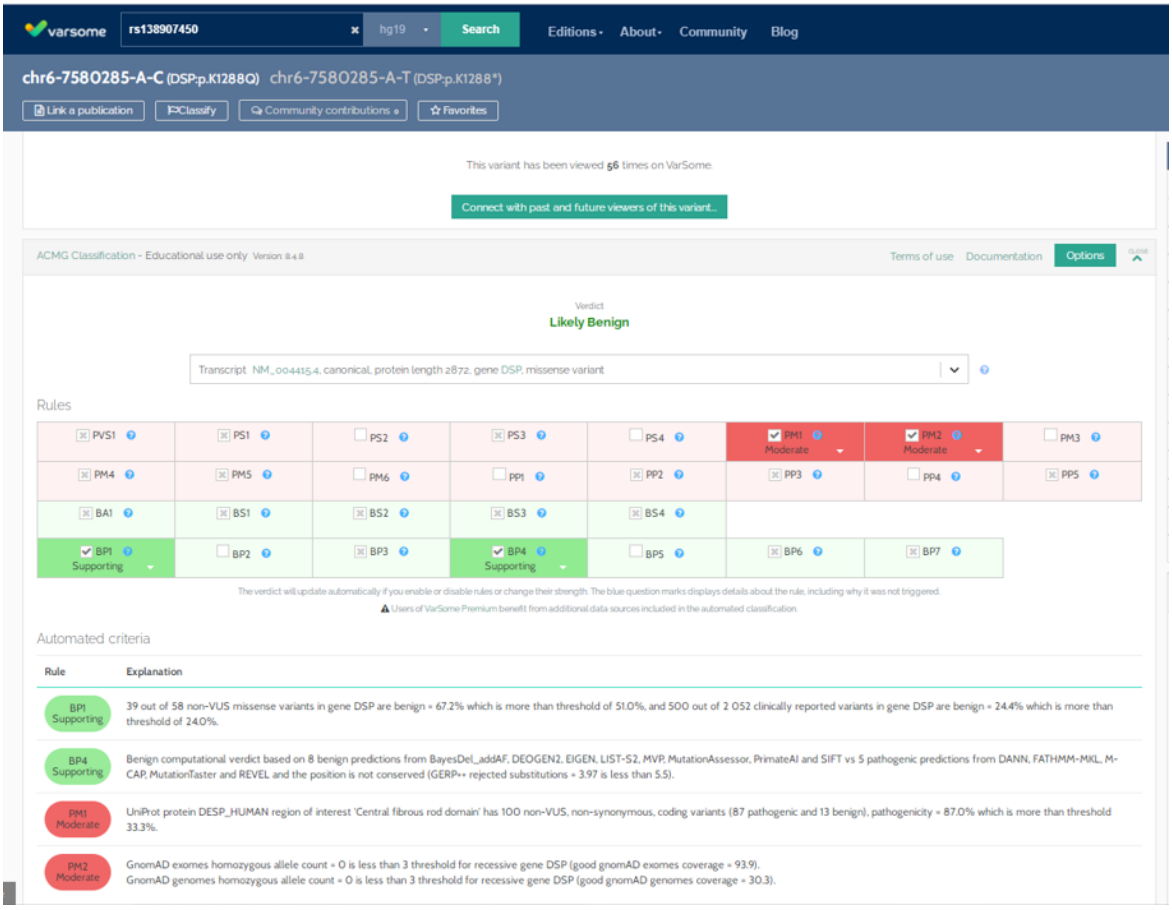


Figura 38. Immagine che mostra l'interpretazione della patogenicità della variante secondo il tool bioinformatico Varsome. Si può osservare che non è presente un'interpretazione univoca, ma ci sono alcuni tools bioinformatici che classificano tale mutazione come benigna (come i tools BayesDel_addAF, DEOGEN2, EIGEN, MVP, SIFT) e altri che classificano la variante come patogenetica (come DANN, FATHMM-MKL, MutationTaster, REVEL).

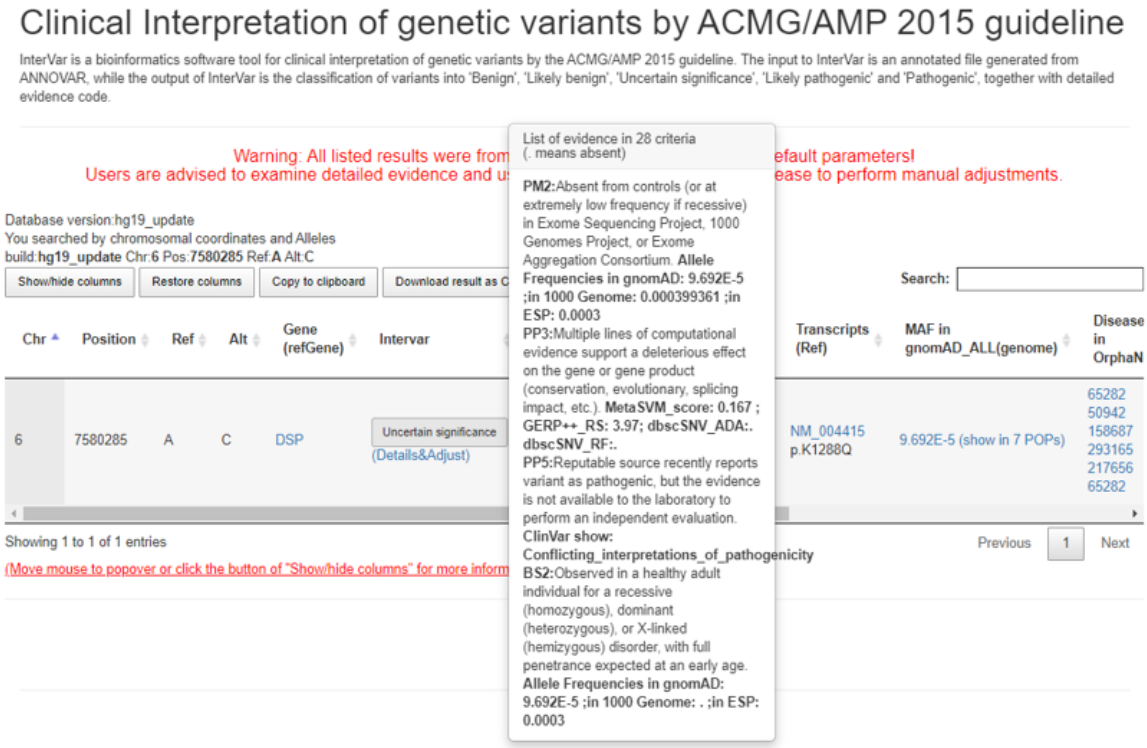


Figura 39. Immagine dell'interpretazione della variante secondo InterVar e i criteri che hanno portato alla sua classificazione come VUS, attenendosi alle linee guida dell'ACMG.

È stato consultato anche il database HGMD (Figura 40) che riporta tutte le caratteristiche della variante e riferimenti agli studi in cui è stata osservata questa variante.

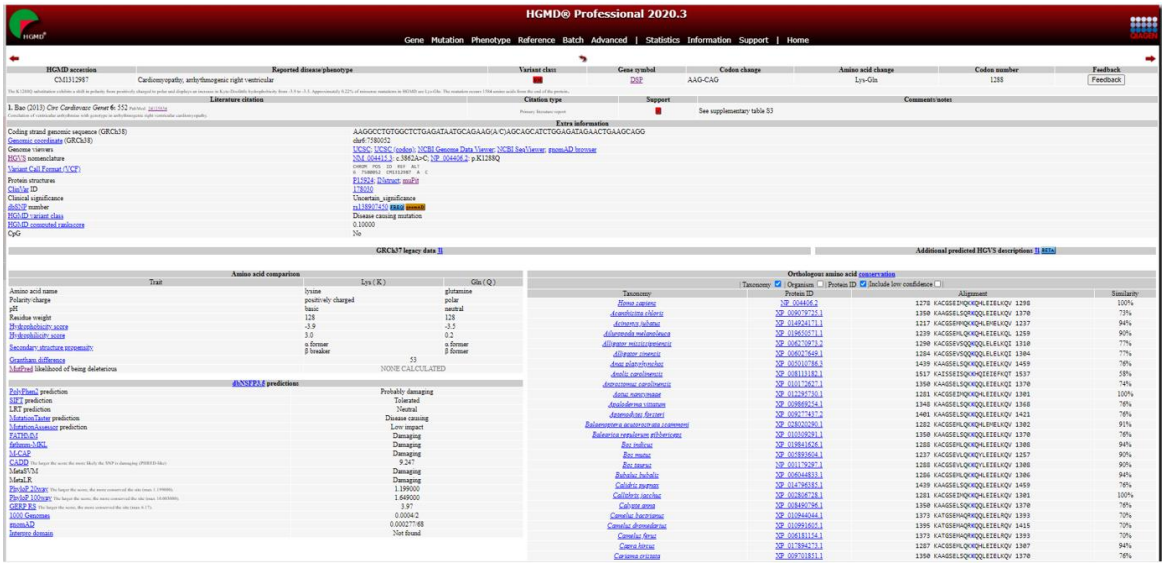


Figura 40. Immagine che descrive le informazioni riguardo la variante osservata riportata nel database HGMD.

È stato consultato anche CardioVAI (EnGenome), che interpreta le varianti non sinonime in 72 geni conosciuti per la loro associazione a disturbi cardiovascolari su base genetica. Questo tool segue le linee guida del ACMG e classifica una variante per ogni fenotipo associato, secondo la letteratura.

Per la variante osservata sono presenti 18 risultati che associano questa mutazione a diverse patologie, tra cui cardiomiopatia dilatativa, Sindrome di Carvajal, cardiomiopatia aritmogena del ventricolo destro, epidermolisi bullosa (Figura 41).

1	NC_000006.11:g.7583154A>C More Info ClinGen Allele Registry	DSP ENST00000379802: p.Lys1887Gln, c.5659A>C (missense_variant)	Cardiomyopathy C0878544	Uncertain significance 2.5	BP4, PM2, PP3
2	NC_000006.11:g.7583154A>C More Info ClinGen Allele Registry	DSP ENST00000379802: p.Lys1887Gln, c.5659A>C (missense_variant)	Cardiomyopathy, dilated, with woolly hair, keratoderma, and tooth agenesis CN221269 Autosomal dominant	Uncertain significance 2.5	BP4, PM2, PP3
3	NC_000006.11:g.7583154A>C More Info ClinGen Allele Registry	DSP ENST00000379802: p.Lys1887Gln, c.5659A>C (missense_variant)	CARDIOMYOPATHY, DILATED, WITH WOOLLY HAIR, KERATODERMA, AND TOOTH AGENESIS C4014393 Autosomal dominant	Uncertain significance 2.5	BP4, PM2, PP3
4	NC_000006.11:g.7583154A>C More Info ClinGen Allele Registry	DSP ENST00000379802: p.Lys1887Gln, c.5659A>C (missense_variant)	Epidermolysis bullosa, lethal acantholytic C1864826 Autosomal recessive	Uncertain significance 2.5	BP4, PM2, PP3
5	NC_000006.11:g.7583154A>C More Info ClinGen Allele Registry	DSP ENST00000379802: p.Lys1887Gln, c.5659A>C (missense_variant)	Cardiomyopathy dilated with woolly hair and keratoderma C1854063 Autosomal recessive	Uncertain significance 2.5	BP4, PM2, PP3
6	NC_000006.11:g.7583154A>C More Info ClinGen Allele Registry	DSP ENST00000379802: p.Lys1887Gln, c.5659A>C (missense_variant)	Arrhythmogenic right ventricular cardiomyopathy, type 8 C1843896 Autosomal dominant	Uncertain significance 2.5	BP4, PM2, PP3
7	NC_000006.11:g.7583154A>C More Info ClinGen Allele Registry	DSP ENST00000379802: p.Lys1887Gln, c.5659A>C (missense_variant)	Skin fragility woolly hair syndrome C1843292 Autosomal recessive	Uncertain significance 2.5	BP4, PM2, PP3
8	NC_000006.11:g.7583154A>C More Info ClinGen Allele Registry	DSP ENST00000379802: p.Lys1887Gln, c.5659A>C (missense_variant)	Keratosis palmoplantaris striata II C1852127	Uncertain significance 2.5	BP4, PM2, PP3
9	NC_000006.11:g.7583154A>C More Info ClinGen Allele Registry	DSP ENST00000379802: p.Lys1887Gln, c.5659A>C (missense_variant)	Arrhythmogenic right ventricular cardiomyopathy C0349788	Likely benign 2	BP4, BP1, PM2, PP3
10	NC_000006.11:g.7580285A>C More Info ClinGen Allele Registry	DSP ENST00000379802: p.Lys1288Gln, c.3862A>C (missense_variant)	Cardiomyopathy C0878544	Uncertain significance 0.5	BP6, PP3
11	NC_000006.11:g.7580285A>C More Info ClinGen Allele Registry	DSP ENST00000379802: p.Lys1288Gln, c.3862A>C (missense_variant)	Epidermolysis bullosa, lethal acantholytic C1864826 Autosomal recessive	Uncertain significance 0.5	BP6, PP3
12	NC_000006.11:g.7580285A>C More Info ClinGen Allele Registry	DSP ENST00000379802: p.Lys1288Gln, c.3862A>C (missense_variant)	Cardiomyopathy dilated with woolly hair and keratoderma C1854063 Autosomal recessive	Uncertain significance 0.5	BP6, PP3
13	NC_000006.11:g.7580285A>C More Info ClinGen Allele Registry	DSP ENST00000379802: p.Lys1288Gln, c.3862A>C (missense_variant)	Skin fragility woolly hair syndrome C1843292 Autosomal recessive	Uncertain significance 0.5	BP6, PP3
14	NC_000006.11:g.7580285A>C More Info ClinGen Allele Registry	DSP ENST00000379802: p.Lys1288Gln, c.3862A>C (missense_variant)	Keratosis palmoplantaris striata II C1852127	Uncertain significance 0.5	BP6, PP3
15	NC_000006.11:g.7580285A>C More Info ClinGen Allele Registry	DSP ENST00000379802: p.Lys1288Gln, c.3862A>C (missense_variant)	Cardiomyopathy, dilated, with woolly hair, keratoderma, and tooth agenesis CN221269 Autosomal dominant	Likely benign -0.5	BS1, BP6, PP3
16	NC_000006.11:g.7580285A>C More Info ClinGen Allele Registry	DSP ENST00000379802: p.Lys1288Gln, c.3862A>C (missense_variant)	CARDIOMYOPATHY, DILATED, WITH WOOLLY HAIR, KERATODERMA, AND TOOTH AGENESIS C4014393 Autosomal dominant	Likely benign -0.5	BS1, BP6, PP3
17	NC_000006.11:g.7580285A>C More Info ClinGen Allele Registry	DSP ENST00000379802: p.Lys1288Gln, c.3862A>C (missense_variant)	Arrhythmogenic right ventricular cardiomyopathy, type 8 C1843896 Autosomal dominant	Likely benign -0.5	BS1, BP6, PP3
18	NC_000006.11:g.7580285A>C More Info ClinGen Allele Registry	DSP ENST00000379802: p.Lys1288Gln, c.3862A>C (missense_variant)	Arrhythmogenic right ventricular cardiomyopathy C0349788	Likely benign -1	BP1, BS1, BP6, PP3

Figura 41. Immagine estrapolata da CardioVAI che raffigura le 18 associazioni della variante osservata nel paziente a diverse patologie, ad alcune attribuisce incerto significato, altre vengono, invece, classificate come potenzialmente benigne.

Ultima analisi è stata effettuata consultando ClinVar, un archivio pubblico consultabile gratuitamente sui rapporti tra una variante umana in esame e il fenotipo, quindi lo stato di salute del paziente.

Dall'analisi sono stati riscontrati 10 submissions che classificano la variante come variante di incerto significato e viene associata a diverse patologie cardiache e non (Figura 42).

Review status: ★★☆☆ criteria provided, multiple submitters, no conflicts
Submissions: 10 (Most recent: May 19, 2020)
Last evaluated: May 12, 2020
Accession: VCV000178030.10
Variation ID: 178030
Description: single nucleotide variant

Variant details
Conditions
Gene(s)

FEEDBACK

Aggregate interpretations per condition

Interpreted condition	Interpretation	Number of submissions	Review status	Last evaluated	Variation/condition record
not provided	Uncertain significance	3	criteria provided, multiple submitters, no conflicts	Apr 9, 2018	RCV000723628.2
not specified	Uncertain significance	1	criteria provided, single submitter	Apr 10, 2018	RCV000154713.4
Lethal acantholytic epidermolysis bullosa	Uncertain significance	1	criteria provided, single submitter	Jan 13, 2018	RCV000342008.2
Skin fragility-woolly hair-palmoplantar keratoderma syndrome	Uncertain significance	1	criteria provided, single submitter	Jan 13, 2018	RCV000403440.2
Arrhythmogenic right ventricular dysplasia 8	Uncertain significance	1	criteria provided, single submitter	Jan 13, 2018	RCV000404310.2
Arrhythmogenic right ventricular dysplasia 8 Dilated cardiomyopathy with woolly hair and keratoderma	Uncertain significance	1	criteria provided, single submitter	Dec 11, 2019	RCV000459782.5
Cardiovascular phenotype	Uncertain significance	1	criteria provided, single submitter	Aug 23, 2019	RCV000620854.1
Cardiomyopathy	Uncertain significance	1	criteria provided, single submitter	May 12, 2020	RCV000771831.2

Figura 42. Risultati ottenuti dalla consultazione dell'archivio ClinVar.

4.2 Discussione

4.2.1 Paziente 1

Una volta individuata una mutazione su un gene, per confermare una correlazione genotipo-fenotipo, è importante conoscere la storia familiare e analizzare anche le sequenze dei familiari per individuare la medesima variante, se presente, e la sua correlazione con eventuali manifestazioni cliniche presenti in questi.

In questo paziente, come descritto in precedenza, è stata osservata una mutazione sul gene DSP in omozigosi che causa una patologia a trasmissione autosomica recessiva, pertanto ci si aspetta che nei genitori sia presente la medesima variante in eterozigosi. Quindi, sono stati analizzati i genomi dei genitori, i campioni sono stati trattati con il procedimento descritto in materiali e metodi e dall'analisi è stata individuata la variante in eterozigosi come ci si aspettava (Figura 43, 44).

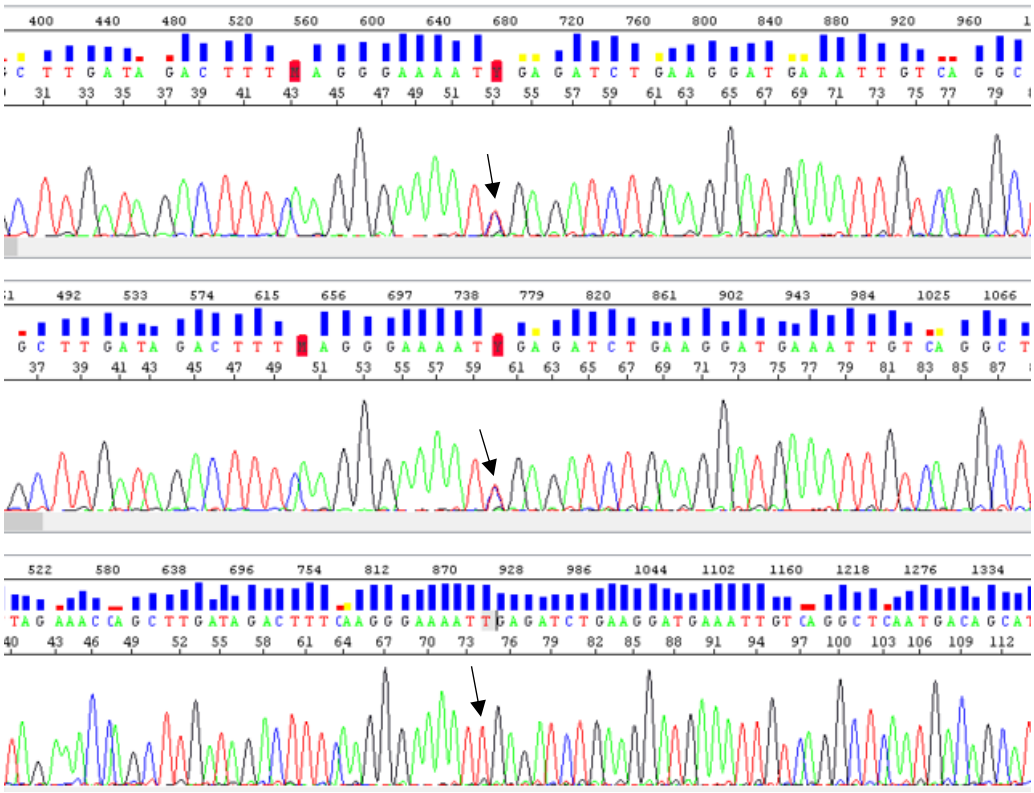


Figura 43. Sequenze dei genitori e del paziente ottenute con il primer forward. In alto è rappresentata la sequenza della madre, al centro quella del padre e in basso quella del paziente. Si può osservare come sia nella madre che nel padre la mutazione è presente in eterozigosi (CT), mentre nel paziente in omozigosi (TT).

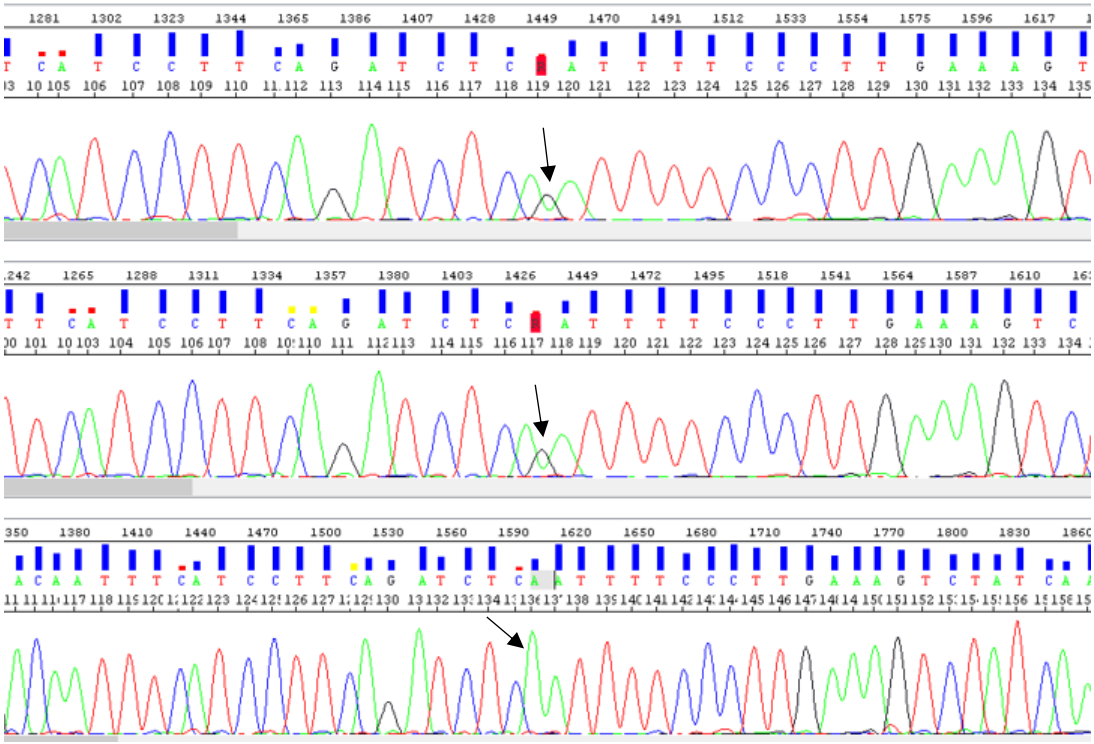


Figura 44. Sequenza ottenuta con il primer reverse. In alto è rappresentata la sequenza della madre, al centro quella del padre e in basso quella del paziente. Si può osservare come sia nella madre che nel padre la mutazione è presente in eterozigosi (AG), mentre nel paziente in omozigosi (AA).

Avendo ottenuto tali risultati sono state avviate varie analisi dei genitori e l'associazione con i segni clinici da loro manifestati. Se si dovesse riscontrare una forma patologica più o meno grave nei genitori, si potrebbe sostenere che la patologia viene trasmessa come autosomica dominante e non come autosomica recessiva.

Queste osservazioni hanno condotto all'analisi della sequenza genica del marito della paziente per rilevare un eventuale rischio riproduttivo, in quanto la paziente trasmetterà con certezza la mutazione alla prole.

Dall'analisi di sequenza del marito è risultato che quest'ultimo non è portatore della mutazione, ovvero è wild type (Figura 45,46).

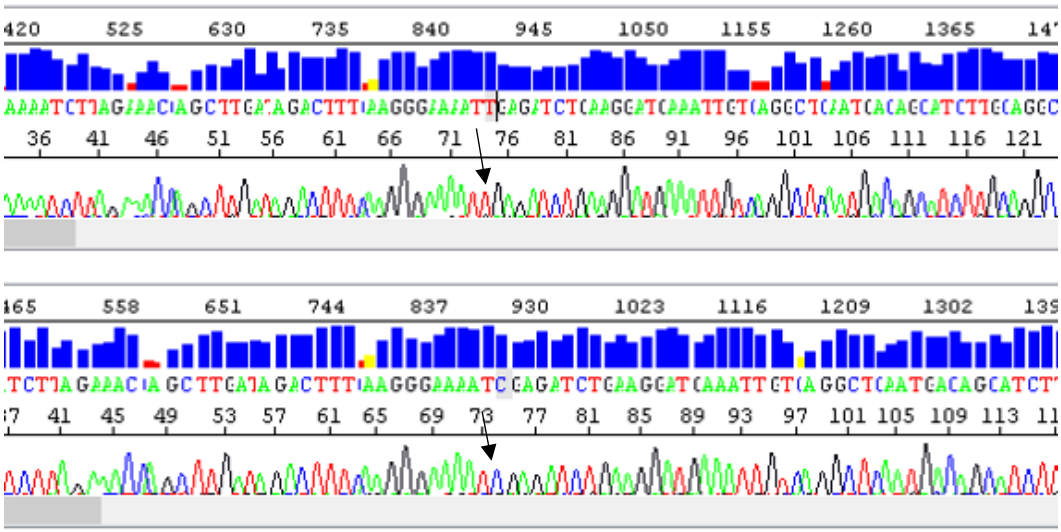


Figura 45. Sono riportate in figura le sequenze, ottenute con il primer forward, della paziente (in alto) e del marito (in basso). Si può osservare che nel marito la mutazione non è presente (CG).

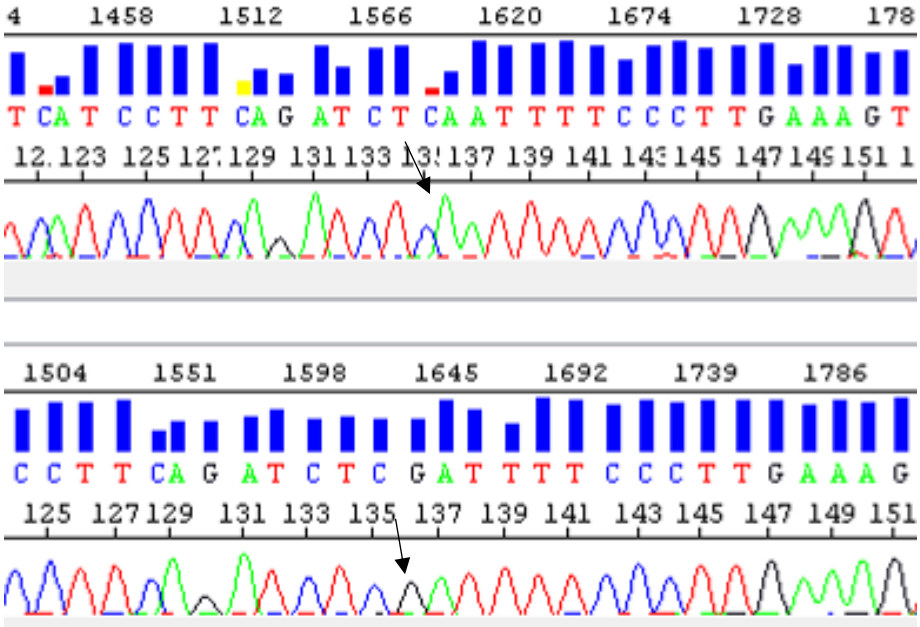


Figura 46. Sono riportate in figura le sequenze, ottenute con il primer reverse, della paziente (in alto) e del marito (in basso). Si può osservare che nel marito la mutazione non è presente (CG).

Da questi risultati si può affermare che i figli della paziente saranno tutti eterozigoti per questa variante, poiché il marito non è portatore (wt), mentre la paziente è omozigote (mut/mut).

Dallo studio dei genitori (mut/wt), quindi, si potrà affermare se i figli della paziente, essendo eterozigoti, presenteranno manifestazioni cliniche più o meno gravi e in che percentuale i figli saranno affetti (Figura 47).

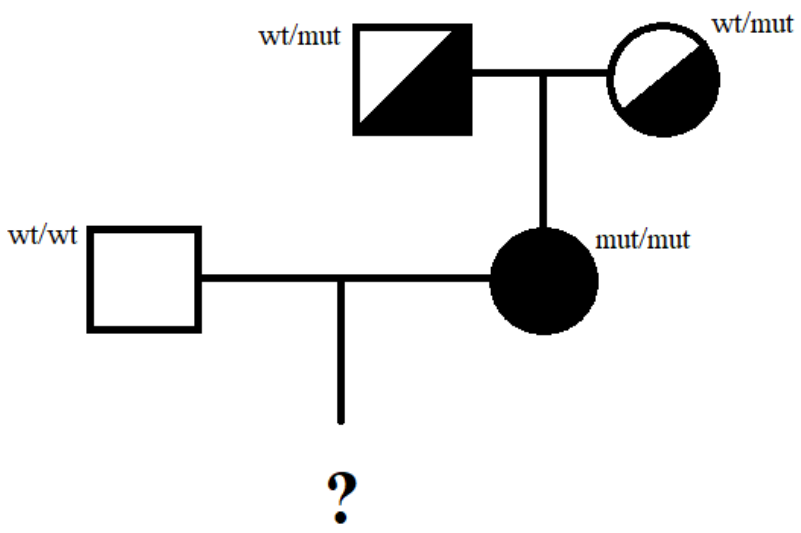


Figura 47. Albero delle mutazioni che schematizza la presenza della variante all'interno della famiglia presa in esame.

4.2.2 Paziente 2

Dall'analisi della sequenza genica di questo paziente è emersa una variante classificata come VUS. Quando emergono queste varianti sono necessari ulteriori studi che ne permettono di valutare la patogenicità, ovvero gli studi funzionali. Poiché questi tipi di studi richiedono numerose risorse e tempo, il primo step di approfondimento consiste nella ricerca della mutazione osservata a livello familiare. L'analisi nei familiari ha lo scopo di valutare quanto una determinata variante sia associata realmente ad una patologia piuttosto che essere una variante benigna. L'analisi a livello familiare viene effettuata con i cosiddetti studi di segregazione, ovvero quell'insieme di

metodiche che consentono la ricerca nei parenti più vicini le varianti individuate nel soggetto in esame. Questi studi hanno lo scopo di confermare l'ereditarietà della mutazione individuata in precedenza sulla base di manifestazioni cliniche.

Ritrovare una mutazione e un fenotipo ricorrenti in più di un individuo all'interno della stessa famiglia, includendo anche parenti non prossimi al paziente, supporta la teoria che la mutazione sia patogenetica.

Ulteriore studio importante per confermare la patogenicità di una mutazione è la caratterizzazione della funzione della variante e individuare la sua presenza in corrispondenza di un dominio funzionale che si traduce in un danno a livello fenotipico.

Infine, vi sono gli studi funzionali che hanno l'obiettivo di identificare e caratterizzare mutazioni e altri fattori regolatori che possono essere coinvolti nello sviluppo e progressione della malattia. Gli approcci tecnici che permettono di valutare il ruolo di un gene o un pattern di geni includono:

- Test *in silico*, che si basano sull'analisi dei dati, mediante tools bioinformatici online per la valutazione di una potenziale patogenicità della variante in esame, tenendo conto dei dati riscontrati e registrati precedentemente anche in altri centri.

- Test *in vitro*, che consistono nell'analisi su modello cellulare dell'effetto di una mutazione.
- Test *in vivo*, che consistono in esperimenti condotti su modello animale, in cui viene indotta la mutazione sul gene o in cui è assente l'intero gene (knock-out) e viene valutato l'effetto a livello fenotipico.

Sono stati effettuati degli studi[61] in cui sono stati creati topi knock-out per il gene della desmoplachina. È stato osservato che la delezione in omozigosi $DSP^{-/-}$ determina una letalità embrionale elevata, conseguenza di importanti anomalie cardiache. Invece, la delezione in eterozigosi $DSP^{+/-}$ si traduce in un eccesso di fibrosi e adipociti a livello del miocardio. Ciò determina anomalie strutturali sia a livello del ventricolo destro che sinistro, disfunzioni cardiache e aritmie ventricolari[62].

Pertanto, il test in vivo, in cui viene indotta la mutazione osservata in modelli animali, permette di confermare con maggior certezza una eventuale correlazione gene-malattia e di classificare con un certo grado di sicurezza una variante come patogena o benigna.

CONCLUSIONI

Le cardiomiopatie sono patologie con un importante rilevanza clinica dal momento in cui predispongono a morte cardiaca improvvisa e alterazioni della funzionalità cardiaca. Da un punto di vista epidemiologico le cardiomiopatie vengono riscontrate maggiormente in soggetti con un'età inferiore ai 30 anni. Tra le cardiomiopatie, la cardiomiopatia ipertrofica e la cardiomiopatia dilatativa presentano una maggior incidenza nei giovani (4-8% l'anno) e la cardiomiopatia aritmogena del ventricolo desto negli adulti (2-3% l'anno)[63,64].

Negli ultimi anni il rapido avanzamento delle conoscenze medico-scientifiche, l'accesso a tecniche avanzate, ma soprattutto lo sviluppo di tecniche di sequenziamento di nuova generazione del genoma, hanno permesso di ottenere importanti risultati in termini di diagnosi e gestione delle patologie. Ciò ha permesso di identificare sempre più spesso correlazioni tra varianti e manifestazioni cliniche associate al gene mutato.

Tuttavia, nonostante le moderne tecniche di diagnosi clinica ci sono casi in cui le mutazioni rilevate risultano essere misconosciute e si incorre nella rilevazione di mutazioni non identificate in precedenza (VUS, riscontrata anche nel nostro studio), la cui classificazione patogenetica dovrà essere confermata con studi di approfondimento. Per cui risulta di fondamentale importanza la raccolta di informazioni anamnestiche sia personali che familiari, a cui fanno

seguito test di segregazione per la ricerca della variante nei familiari e, infine, studi funzionali *in vitro* e *in vivo*.

L'eterogeneità delle mutazioni rende, pertanto, l'analisi genetica non sufficiente alla formulazione della diagnosi, che si basa ancora oggi principalmente su dati clinici, strumentali e sulla storia familiare dei pazienti.

La desmoplachina è conosciuta avere una funzione chiave nelle interazioni a livello dei desmosomi, i quali nel cuore sono diffusi nei dischi intercalari, ovvero giunzioni cellulari specializzate a livello dei siti di contatto tra i cardiomiociti, rivestendo perciò un ruolo chiave per lo sviluppo cardiaco.

Allo stato attuale delle conoscenze, l'individuazione delle mutazioni del gene DSP che risultano patogene è di fondamentale rilevanza per tentare di stabilire correlazioni gene-malattia, le quali potrebbero consentire ai cardiologi di formulare una prognosi precisa sulla base della mutazione patogena riscontrata nel paziente.

BIBLIOGRAFIA

1. Gerstein MB, Bruce C, Rozowsky JS, Zheng D, Du J, Korbel JO, Emanuelsson O, Zhang ZD, Weissman S, Snyder M, What is a gene, post-ENCODE? History and updated definition, in Genome Research, vol. 17, n. 6, 2007, pp. 669–681, DOI: 10.1101/gr.6339607

2. International Human Genome Sequencing Consortium, Finishing the euchromatic sequence of the human genome. in Nature, vol. 431, nº 7011, 2004, pp. 931–45.

3. Levy S, Sutton G, Ng PC, Feuk L, Halpern AL, Walenz BP, Axelrod N, Huang J, Kirkness EF, Denisov G, Lin Y, MacDonald JR, Pang AW, Shago M, Stockwell TB, Tsiamouri A, Bafna V, Bansal V, Kravitz SA, Busam DA, Beeson KY, McIntosh TC, Remington KA, Abril JF, Gill J, Borman J, Rogers YH, Frazier ME, Scherer SW, Strausberg RL, Venter JC. The diploid genome sequence of an individual human. Biol. 2007 Sep 4;5(10): e254.

4. NHS National Genetics and Genomics Education Centre.

5. Lewin, IL GENE VIII, Zanichelli editore S.p. A. Copyright © 2006.

6. Bruce Alberts, Alexander D. Johnson, Julian Lewis, David Morgan, Martin Raff, Keith Roberts e Peter Walter. Molecular Biology of the Cell.

7. Beverly, M. et al. “Label-free analysis of mRNA capping efficiency using RNase H probes and LC-MS.” Analytical and Bioanalytical Chemistry 408 (2016): 5021-5030.

8. Nick J. Proudfoot, Andre Furger, Michael J. Dye. Integrating mRNA Processing with Transcription. volume 108, issue 4, p501-512, february 22, 2002.

9. Patricia M. Watson and Dennis K. Watson. Alternative Splicing in Prostate and Breast Cancer. The Open Cancer Journal, 2010, 3, 62-76.

10. H. Curtis, N. S. Barnes, A. Schnek, G. Flores. Invito alla biologia. blu B – Biologia molecolare, genetica ed evoluzione 2. Zanichelli editore 2012.

11. Whittock, N. V., Ashton, G. H. S., Dopping-Hepenstal, P. J. C., Gratian, M. J., Keane, F. M., Eady, R. A. J., McGrath, J. A. Striate palmoplantar keratoderma resulting form desmoplakin haploinsufficiency. J. Invest. Derm. 113: 940-946, 1999.

12. Green KJ, Parry DA, Steinert PM, Virata ML, Wagner RM, Angst BD, Nilles LA. Structure of the human desmoplakins. Implications for function in the desmosomal plaque. *J Biol Chem.* 1990; 265:2603–2612.

13. Rita M. Cabral, Hong Wan, Clare L. Cole, Dominic J. Abrams, David P. Kelsell, Andrew P. South. "Identification and characterization of *DSPIa*, a novel isoform of human *desmoplakin*". Cell Tissue Res. 2010; 341(1): 121–129.

14. Kelly DF. "Fine structure of desmosome, hemidesmosomes, and an epidermal globular layer in developing newt epidermis." J Cell Biol 1966; 28: 51-72.

15. Sarah H. Vermij, Hugues Abriel, Toon A. B. van Veen, Refining the molecular organization of the cardiac intercalated disc, *Cardiovascular Research*, Volume 113, Issue 3, March 2017, Pages 259–275.

16. Miller K, Mattey D, Measures H, Hopkins C, Garrod D. "Localisation of the protein and glycoprotein components of bovine nasal epithelial desmosomes by immunoelectron microscopy". EMBO J 1987; 6, 885-889.

17. Huber O. "Structure and function of desmosomal proteins and their role in development and disease". Cell Mol Life Sci 2003; 60: 1872-1890.

18. Norman M, Simpson M, Mogensen J, Shaw A, Hughes S, Syrris P, SenChowdhry S, Rowland E, Crosby A, McKenna WJ: "Novel mutation in desmoplakin causes arrhythmogenic left ventricular cardiomyopathy". Circulation 2005, 112:636-642.

19. Heuser A, Plovie ER, Ellinor PT, et al. "Mutant desmocollin-2 causes arrhythmogenic right ventricular cardiomyopathy". Am J Hum Genet 2006; 79:1081– 8.

20. Syrris P, Ward D, Evans A, Asimaki A, Gandjbakhch E, Sen-Chowdhry S, McKenna WJ: "Arrhythmogenic right ventricular dysplasia/ cardiomyopathy associated with mutations in the desmosomal gene desmocollin-2". Am J Hum Genet 2006, 79:978-984.

21. Palka HL, Green KJ. "Roles of plakoglobin end domains in desmosome assembly". Journal of cell science 1997; 110:2359-2371.

22. Sen-Chowdhry, S Syrris P, McKenna WJ. "Genetics of Right ventricular cardiomyopathy". J Cardiovasc Electrophysiol 2005; 16: 927-935.

23. Gerull B, Heuser A, Wichter T, Paul M, Basson CT, McDermott DA, Lerman BB, Markowitz SM, Ellinor PT, MacRae CA, Peters S, Grossmann KS, Michely B, Sasse-Klaassen S, Birchmeier W, Dietz R, Breithardt G, Schulze-Bahr E, Thierfelder L: "Mutations in the

desmosomal protein plakophilin-2 are common in arrhythmogenic right ventricular cardiomyopathy". Nat Genet 2004, 36:1162-1164.

24. Kaplan SR, Gard JJ, Protonotarios N, Tsatsopoulou A, Spiliopoulou C, Anastasakis A, Squarcioni CP, McKenna WJ, Thiene G, Basso C, Brousse N, Fontaine G, Saffitz JE: Remodeling of myocyte gap junctions in arrhythmogenic right ventricular cardiomyopathy due to a deletion in plakoglobin (Naxos disease). Heart Rhythm 2004, 1:3-11.

25. Leung CL, Green KJ, Liem R.K.H. "Plakins: a family o versatile cytolinker proteins". TRENDS in Cell Biology 2002; 12:37-45.

26. Kowalczyk AP, Bornslaeger EA, Borgwardt JE, Palka HL, Dhaliwal AS, Corcoran CM, Denning MF, Green KJ. "The aminoterminal domain of desmoplakin binds to plakoglobin and clusters desmosomal cadherin-plakoglobin complexes". The Journal of cell Biology 1997; 139: 773-78.

27. D. Garrod, M. Chidgey. Desmosome structure, composition and function. BBA, 1778 (2008), pp. 572-587.

28. Beffagna G, Occhi G, Nava A, Vitiello L, Ditadi A, Basso C, Bauce B, Carraro G, Thiene G, Towbin JA, Danieli GA, Rampazzo A: "Regulatory mutations in transforming growth factor-beta3 gene cause

arrhythmogenic right ventricular cardiomyopathy type 1”. Cardiovasc Res 2005, 65:366-373.

29. Garrod D, Chidgey M (Mar 2008). "Desmosome structure, composition and function". *Biochimica et Biophysica Acta (BBA) - Biomembranes*. 1778 (3): 572–87.

30. David Garrod, "Desmosomes In Vivo", *Dermatology Research and Practice*, vol. 2010, Article ID 212439, 17 pages, 2010.

31. Green KJ, Parry DA, Steinert PM, Virata ML, Wagner RM, Angst BD, Nilles LA (Feb 1990). "Structure of the human desmoplakins. Implications for function in the desmosomal plaque". The Journal of Biological Chemistry. 265 (5): 2603–12.

32. Smith EA, Fuchs E (1998). "Defining the interactions between intermediate filaments and desmosomes". J. Cell Biol. 141 (5): 1229–41.

33. Stappenbeck TS, Bornslaeger EA, Corcoran CM, Luu HH, Virata ML, Green KJ (Nov 1993). "Functional analysis of desmoplakin domains: specification of the interaction with keratin versus vimentin intermediate filament networks". The Journal of Cell Biology. 123 (3): 691–705.

34. Choi HJ, Park-Snyder S, Pascoe LT, Green KJ, Weis WI (Aug 2002). "Structures of two intermediate filament-binding fragments of

desmoplakin reveal a unique repeat motif structure". Nature Structural Biology. 9 (8): 612–20.

35. Green, K., Gaudry, C. Are desmosomes more than tethers for intermediate filaments? *Nat Rev Mol Cell Biol* 1**,** 208–216 (2000).

36. Stappenbeck TS, Lamb JA, Corcoran CM, Green KJ (Nov 1994). "Phosphorylation of the desmoplakin COOH terminus negatively regulates its interaction with keratin intermediate filament networks". *The Journal of Biological Chemistry*. 269 (47): 29351–4.

37. T. Lechler, E. Fuchs, "Desmoplakin: an unexpected regulator of microtubule organization in the epidermis". J. Cell Biol. 176 (2007) 147–154.

38. Virata MLA, Wagner RM, Parry DAD, Green KJ. "Molecular structure of human desmoplakin I e II amino terminus". Proc Natl Acad Sci USA 1992;89 (2): 544-548.

39. Cabral, R.M., Wan, H., Cole, C.L. *et al.* "Identification and characterization of *DSPIa*, a novel isoform of human *desmoplakin*". *Cell Tissue Res* 341**,** 121–129 (2010).

40. Cabral, R.M., Wan, H., Cole, C.L. *et al.* Identification and characterization of *DSPIa*, a novel isoform of human *desmoplakin* . *Cell Tissue Res* 341**,** 121–129 (2010).

41. Jonkman, M.F., A.M.G. Pasmooij, S.G.M.A. Pasmans, M.P. van den Berg, H.J. Ter Horst, A. Timmer, and H.H. Pas. 2005. "Loss of desmoplakin tail causes lethal acantholytic epidermolysis bullosa". Am. J. Hum. Genet. 77:653–660.

42. Armstrong, D. K., McKenna, K. E., Purkis, P. E., Green, K. J., Eady, R. A. J., Leigh, I.M., Hughes, A. E. "Haploinsufficiency of desmoplakin causes a striate subtype of palmoplantar keratoderma". Hum. Molec. genet. 8: 143-148, 1999. Nota: Erratum: Hum. Molec. genet. 8: 943 solo, 1999.

43. Norgett, E. E., Hatsell, S. J., Carvajal-Huerta, L., Ruiz Cabezas, J.-C., Common, J., Purkis, P. E., Whittock, N., Leigh, I.M., Stevens, H. P., Kelsell, D. P. "La mutazione recessiva nel desmoplakin interrompe le interazioni del filamento desmoplakin-intermedio e causa cardiomiopatia dilatata, capelli lanosi e cheratoderma". Hum. Molec. genet. 9: 2761-2766, 2000.

44. Uzumcu, A., Norgett, E. E., Dindar, A., Uyguner, O., Nisli, K., Kayserili, H., Sahin, S. E., Dupont, E., Severs, N. J., Leigh, I. M., Yuksel-Apak, M., Kelsell, D. P., Wollnik, B. "Loss of desmoplakin isoform 1 causes early onset cardiomyopathy and heart failure in a Naxos-like syndrome". (Letter) J. Med. Genet. 43: e5, 2006.

45. Norgett, E. E., Lucke, T. W., Bowers, B., Munro, C. S., Leigh, I. M., Kelsell, D. P. "Early death from cardiomyopathy in a family with autosomal dominant striate palmoplantar keratoderma and wooly hair associated with a novel insertion mutation in desmoplakin". (Letter) J. Invest. Derm. 126: 1651-1654, 2006. Note: Erratum: J. Invest. Derm. 126: 2735 only, 2006.

46. Whittock, N. V., Wan, H., Morley, S. M., Garzon, M. C., Kristal, L., Hyde, P., McLean, W. H. I., Pulkkinen, L., Uitto, J., Christiano, A. M., Eady, R. A. J., McGrath, J. A. "Compound heterozygosity for non-sense and mis-sense mutations in desmoplakin underlies skin fragility/woolly hair syndrome". J. Invest. Derm. 118: 232-238, 2002.

47. Rampazzo, A., Nava, A., Malacrida, S., Beffagna, G., Bauce, B., Rossi, V., Zimbello, R., Simionati, B., Basso, C., Thiene, G., Towbin, J. A., Danieli, G. A. "Mutation in human desmoplakin domain binding to

plakoglobin causes a dominant form of arrhythmogenic right ventricular cardiomyopathy". Am. J. Hum. Genet. 71: 1200-1206, 2002.

48. Jonkman, M. F., Pasmooij, A. M. G., Pasmans, S. G. M. A., van den Berg, M. P., ter Horst, H. J., Timmer, A., Pas, H. H. "Loss of desmoplakin tail causes lethal acantholytic epidermolysis bullosa". Am. J. Hum. Genet. 77: 653-660, 2005.

49. Braunwald E. Heart Disease. A text book of cardiovascular medicine. VII ed. 2005. W.B. Saunders, New York.

50. Kubo T, Gimeno JR, Bahl A, Steffensen U, Steffensen M, Osman E, Thaman R, Mogensen J, Elliott PM, Doi Y, McKenna WJ. Prevalence, clinical significance, and genetic basis of hypertrophic cardiomyopathy with restrictive phenotype, J Am Coll Cardiol, 2007, vol. 49.

51. Thiene G, Nava A, Corrado D, Rossi L, Pennelli N Right ventricular cardiomyopathy and sudden death in young people. N Engl J Med 1988; 318: 129-133.

52. 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).

53. Chidgeavadze ZG, Beabealashvilli RS, Atrazhev AM, Kukhanova MK, Azhayev A V., Krayevsky AA. 2′, 3′-dideoxy-3′ amlnonudeo 5′ .triphosphates are the terminators of DNA synthesis catalyzed by DNA polymerases. *Nucleic Acids Res*. 1984;12(3):1671-1686.

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. Helmut Kae et al., 2011. Sanger method.

57. Che-Seng Ku, Roukos DH. From next-generation sequencing to nanopore sequencing technology:paving the way to personalized genomic medicine. Expert Rev.Med.Devices 10(1),1-6(2013).

58. Tawfik DS,Griffiths AD Man-made cell-like compartments for molecular evolution. Nat.Biotechnol 1998;16:652-6.

59. Rougemont J,Amzallag A, Iseli C Probabilistic base-calling of Solexa sequencing data. BMC Bioinformatics 2008; 9:431.

60. Richards S, Aziz N, Bale S, Bick D, Das S, Gastier-Foster J, Grody WW, Hegde M, Lyon E, Spector E, Voelkerding K, Rehm HL; ACMG Laboratory Quality Assurance Committee. 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. 2015 May;17(5):405-24. doi: 10.1038/gim.2015.30. Epub 2015 Mar 5.

61. Gallicano GI, Kouklis P, Bauer C, Yin M, Vasioukhin V, Degenstein L, Fuchs E Desmoplakin is required early in development for assembly of desmosomes and cytoskeletal linkage. J Cell biol 1998; 143: 2009-2022.

62. Gallicano GI, Bauer C, Fucus E Rescuing desmoplakin function in extra-embryonic ectoderm reveals the importance of this protein inembryonic heart, neuroepithelium, skin and vasculature. Developement 2001; 128: 929-941.

63. Balzanelli M.G., Catalano A.: La morte improvvisa. CIC Edizioni Internazionali, Roma, 2005.

64. Basso C., Calabrese F., Corrado D., Thiene G.: Postmortem diagnosis in sudden cardiac death victims: macroscopic, microscopic and molecular findings.Cardiovasc Res 2001; 50: 290-330. 5.