



Università degli Studi di Cagliari

## **DOTTORATO DI RICERCA**

**Terapia Pediatrica e Farmacologia dello Sviluppo**

Ciclo XXV

### **TITOLO TESI**

**Nuovi approcci molecolari per lo studio di malattie monogeniche rare: utilizzo dell'exome sequencing per la ricerca di geni malattia.**

Settore/i scientifico disciplinari di afferenza

BIO/11

Presentata da: Dott.ssa Carla Cossu

Coordinatore Dottorato Prof. Paolo Moi

Tutor/Relatore Prof.ssa Maria Cristina Rosatelli

Esame finale anno accademico 2011 – 2012

# INDICE

<b>SOMMARIO</b>	<b>pag.3</b>
<b>INTRODUZIONE</b>	<b>pag.5</b>
<b>LA SINDROME DI JEUNE</b>	<b>pag.35</b>
<b>SCOPO DEL PROGETTO</b>	<b>pag.41</b>
<b>PAZIENTI E METODI</b>	<b>pag.43</b>
<b>RISULTATI</b>	<b>pag.51</b>
<b>DISCUSSIONE</b>	<b>pag.58</b>
<b>CONCLUSIONI</b>	<b>pag.67</b>
<b>APPENDICE</b>	<b>pag.70</b>
<b>BIBLIOGRAFIA</b>	<b>pag.82</b>

## SOMMARIO

Le malattie “rare” rappresentano un insieme di patologie in cui il numero di affetti, secondo quanto definito dall’Unione Europea, è inferiore a 5/10000 nella popolazione generale (Orphan Drug Regulation 141/2000).

Tali patologie, rare per definizione, costituiscono in realtà un problema socio-sanitario di primaria importanza.

Molte di queste malattie sono infatti spesso croniche o gravemente invalidanti, e sono, in molti casi, causa di mortalità precoce. La mancanza di politiche sanitarie specifiche e la scarsità di conoscenze si traducono in sono tradotte negli anni in ritardi alla diagnosi e in difficoltà di accesso alle cure, che hanno come conseguenza ulteriori danni fisici, psicologici e intellettuali e trattamenti spesso inadeguati, benché certe malattie rare siano compatibili con una vita normale se diagnosticate in tempo e trattate in modo appropriato (Commissione delle Comunità Europee; Bruxelles, 11.11.2008 COM(2008))

Sebbene il gruppo della malattie rare comprenda forme tumorali, malattie autoimmuni, malformazioni congenite e patologie di origine infettiva o tossica, la maggior parte di esse è rappresentata da patologie di origine genetica.

Identificare le basi genetiche di tali malattie è di primaria importanza perché, da un lato, consente di comprendere i pathways biologici alterati e di attuare approcci terapeutici mirati e, dall’altro, di effettuare una diagnosi molecolare diretta e di predisporre misure di prevenzione tramite il calcolo del rischio di ricorrenza e la Consulenza Genetica.

Molti geni responsabili di malattie mendeliane sono stati individuati e caratterizzati utilizzando strategie come il clonaggio posizionale e il sequenziamento con il metodo di Sanger. Tuttavia, fattori come la ridotta disponibilità di soggetti affetti e di famiglie da studiare, l’eterogeneità di locus e la ridotta fitness riproduttiva, spesso limitano il potere di questi approcci convenzionali.

I progressi tecnologici raggiunti nel campo delle strategie di sequenziamento degli acidi nucleici (“Next Generation Sequencing”) sono oggi in grado di fornire un metodo alternativo per la caratterizzazione di alleli-malattia attraverso lo studio dell’informazione contenuta all’interno del genoma umano. In particolare, il

sequenziamento della sola porzione codificante del genoma (esoma) consente di ridurre i costi e i tempi del sequenziamento senza che le probabilità di successo risultino ridotte in maniera significativa. Le patologie ad ereditarietà mendeliana sono infatti determinate da varianti alleliche che nell'85% dei casi si ritrovano all'interno di sequenze codificanti per proteine o nei siti canonici di splicing.

Lo scopo di questo progetto è stato lo studio dell'eziopatogenesi di malattie mendeliane rare per le quali il gene causativo non è noto mediante l'analisi dell'esoma di individui affetti. In particolare, l'approccio dell'exome sequencing è stato utilizzato per ricercare uno o più geni candidati per una rara osteocondrodisplasia, la Sindrome di Jeune (JATD, Asphyxiating Thoracic Dysplasia, MIM 208500), negli affetti di due famiglie di origine sarda non correlate, in studio presso il nostro Dipartimento.

# INTRODUZIONE

## **Il sequenziamento del genoma umano per la caratterizzazione di alleli malattia**

Le strategie di identificazione di geni-malattia utilizzate nei decenni passati con lo scopo di definire le basi eziopatogenetiche di malattie ad ereditarietà mendeliana hanno consentito nel complesso di chiarire i meccanismi patogenetici di circa il 50% dei fenotipi con base genetica (fig.1). Nel caso di malattie a base biochimica nota, come l'anemia falciforme e la fenilchetonuria, è stato possibile associare il fenotipo patologico al gene responsabile attraverso la ricerca e la caratterizzazione del prodotto proteico assente (genetica funzionale). Con la strategia dei "geni candidati" le basi eziopatogenetiche sono state definite mediante la ricerca, in pazienti affetti, di mutazioni patogenetiche in geni con funzione nota. Per la maggior parte delle malattie mendeliane, per le quali non si avevano dati sui pathway biologici alterati, strategie di mappatura genica come l'analisi di linkage e il mappaggio per omozigosità hanno permesso di identificare il gene-malattia tramite l'impiego di marcatori molecolari polimorfici genome-wide che consentono di locus malattia all'interno di una regione cromosomica definita mediante l'analisi di eventi di ricombinazione meiotica in famiglie con individui affetti.

La limitata disponibilità di affetti da analizzare, la penetranza variabile, la ridotta fitness riproduttiva, l'eterogeneità genetica e fenotipica rappresentano i fattori che più spesso hanno ridotto il successo delle tradizionali metodiche impiegate per la ricerca di geni-malattia.

I progressi attualmente raggiunti nel campo delle tecnologie di sequenziamento degli acidi nucleici, complessivamente note con il termine "Next-Generation Sequencing" (NGS) o "Massively Parallel Sequencing" (MPS), consentono di poter identificare e analizzare ogni tipo di variazione genetica presente nel genoma umano, ad un elevato livello di risoluzione e con throughput di dati estremamente significativo, fornendo un metodo alternativo valido per l'individuazione di alleli-malattia.

Negli ultimi anni, molti geni responsabili di malattie genetiche sono stati identificati grazie all'analisi di varianti alleliche rilevate in progetti di sequenziamento massivo parallelo. In alcuni studi, le nuove metodiche hanno semplicemente sostituito l'elettroforesi capillare basata sulla biochimica di Sanger per il sequenziamento di intervalli di linkage. In tal senso, il sequenziamento MPS è stato applicato, per esempio, per identificare il gene della tetraspanina 12 (TSPAN12, MIM 613138) in un individuo affetto da una grave retinopatia, la FEVR (Familial Exudative Vitreo-retinopathy, MIM 613310), per la quale era stato caratterizzato un intervallo di linkage esteso 40Mb contenente circa 300 geni e il sequenziamento con il metodo Sanger, effettuato solo per alcuni di essi, non aveva permesso di identificare varianti causative per la malattia (Nikopoulos et al., 2010).

In altri progetti di sequenziamento, applicati all'intero genoma (WGS, Whole Genome Sequencing) o alla sola porzione codificante (Exome Sequencing), è stato possibile identificare mutazioni causative di malattie genetiche mediante l'analisi diretta delle varianti rilevate, prescindendo dai dati di linkage (fig.2).

I costi ancora elevati e la mole di dati risultante dal sequenziamento dell'intero genoma, che richiede complesse analisi di tipo bioinformatico a cui si associa la difficoltà nell'attribuire un significato funzionale a molte delle varianti rilevate, fanno sì che il WGS trovi attualmente la sua maggiore applicabilità nel campo della ricerca e, sebbene in futuro i costi relativi al sequenziamento genomico siano progressivamente destinati a ridursi, il trasferimento di tale metodica nella pratica clinica appare ancora difficile.

Un'alternativa valida al WGS è rappresentata dalla possibilità di selezionare e sequenziare solo la porzione del genoma corrispondente alle regioni codificanti (circa il 2% dell'intero genoma) attraverso l'exome sequencing. In generale il sequenziamento dell'esoma consente di ridurre i costi e i tempi di sequenziamento senza che le probabilità di successo siano significativamente ridotte. Il limite principale dello studio dell'esoma è, chiaramente, la mancata determinazione di varianti presenti nella parte non codificante, come varianti strutturali e riarrangiamenti cromosomici.

L'utilizzo di tale approccio è però giustificato dall'evidenza che gli studi di positional cloning hanno spesso portato all'identificazione di varianti responsabili di malattie

genetiche all'interno di sequenze codificanti e che molti alleli responsabili di disordini di tipo Mendeliano alterano in modo deleterio tali sequenze. Molte varianti causative rare che si trovano all'interno di sequenze codificanti sono sostituzioni missenso o non senso o piccole inserzioni e delezioni per le quali si può predire un impatto grave o deleterio a livello funzionale. Inoltre, sebbene gli studi genome-wide abbiano portato all'individuazione di più di 250 varianti comuni che contribuiscono al fenotipo patologico di un vasto range di disordini complessi, parte del contributo genetico rimane non definito, suggerendo che varianti rare con effetto importante a livello fenotipico possano spiegare la parte della variabilità mancante nelle patologie complesse.

Anche nei casi in cui si possano utilizzare le metodiche convenzionali, sequenziare l'esoma rappresenta comunque un modo per semplificare la diagnosi molecolare nel caso di malattie rare e/o di elevata eterogeneità genetica.

I risultati del primo progetto di exome sequencing applicato all'identificazione del gene responsabile di una malattia genetica rara, la Sindrome di Miller (MIM 263750), vengono pubblicati dal gruppo di Ng nel 2010 (Ng et al., 2010). Dal 2009, anno in cui è stata dimostrata l'efficacia dell'exome sequencing nel rilevamento di alleli-malattia (Ng et al., 2009), al 2012 più di 100 geni sono stati associati a patologie genetiche rare grazie all'exome sequencing (Rabbani et al., 2012), per un totale di più di 10000 esomi sequenziati.

Numero di voci inserite in OMIM (all'8 marzo 2013)					
Descrizione	Autosomica	X-linked	Y-linked	Mitocondriale	Totale
Gene e fenotipo, combinati	124	4	0	2	130
Fenotipo con basi molecolari note	3430	272	4	28	3734
Fenotipo o locus con basi molecolari non note	1625	133	5	0	1763
Altri fenotipo con sospette basi mendeliane	1757	124	2	0	1883

**Fig. 1** Dati estratti dal sito dell'O2MIM (On Line Mendelian Inheritance In Man, <http://omim.org/>).

<b>Disorder</b>	<b>Inheritance</b>	<b>Gene identified</b>	<b>Scope</b>
Congenital chloride diarrhea	Recessive	<i>SLC26A3</i>	Exome
Miller syndrome	Recessive	<i>DHODH</i>	Exome
Charcot-Marie-Tooth neuropathy	Recessive	<i>SH3TC2</i>	Genome
Metachondromatosis	Dominant	<i>PTPN11</i>	Genome
Schinzel-Giedion syndrome	Dominant	<i>SETBP1</i>	Exome
Nonsyndromic hearing loss	Recessive	<i>GPSM2</i>	Exome
Perrault syndrome	Recessive	<i>HSD17B4</i>	Exome
Hyperphosphatasia mental retardation syndrome	Recessive	<i>PIGV</i>	Exome
Sensenbrenner syndrome	Recessive	<i>WDR35</i>	Exome
Cerebral cortical malformations	Recessive	<i>WDR62</i>	Exome
Kaposi sarcoma	Recessive	<i>STIM1</i>	Exome
Spinocerebellar ataxia	Dominant	<i>TGM6</i>	Exome
Combined hypolipidemia	Recessive	<i>ANGPTL3</i>	Exome
Complex I deficiency	Recessive	<i>ACAD9</i>	Exome
Autoimmune lymphoproliferative syndrome	Recessive	<i>FADD</i>	Exome
Amyotrophic lateral sclerosis	Dominant	<i>VCP</i>	Exome
Nonsyndromic mental retardation	Dominant	Various	Exome
Kabuki syndrome	Dominant	<i>MLL2</i>	Exome
Inflammatory bowel disease	Dominant	<i>XIAP</i>	Exome
Nonsyndromic mental retardation	Recessive	<i>TECR</i>	Exome
Retinitis pigmentosa	Recessive	<i>DHDDS</i>	Exome
Osteogenesis imperfecta	Recessive	<i>SERPINF1</i>	Exome
Dilated cardiomyopathy	Dominant	<i>BAG3</i>	Exome
Hajdu-Cheney syndrome	Dominant	<i>NOTCH2</i>	Exome
Hajdu-Cheney syndrome	Dominant	<i>NOTCH2</i>	Exome
Skeletal dysplasia	Recessive	<i>POP1</i>	Exome
Amelogenesis	Recessive	<i>FAM20A</i>	Exome
Chondrodysplasia and abnormal joint development	Recessive	<i>IMPAD1</i>	Exome
Progeroid syndrome	Recessive	<i>BANF1</i>	Exome
Infantile mitochondrial cardiomyopathy	Recessive	<i>AARS2</i>	Exome
Sensory neuropathy with dementia and hearing loss	Dominant	<i>DNMT1</i>	Exome
Autism	Dominant	Various	Exome

**Fig.2.** Progetti di sequenziamento NGS applicati alla caratterizzazione di alleli-malattia (Gilissen et al., 2011)



## **Strategie per l'identificazione di nuovi geni malattia attraverso il sequenziamento NGS**

L'identificazione di geni-malattia per patologie genetiche rare per le quali non sono note mutazioni in geni conosciuti richiedono strategie di filtraggio che consentano di pervenire all'identificazione di varianti presumibilmente patogenetiche selezionandole da un background comprendente variazioni non patogenetiche (polimorfismi) e errori di sequenziamento.

Mentre con il sequenziamento genomico umano si rilevano mediamente più di 3 milioni di variazioni di singoli nucleotidi (Bentley DR et al., 2008), con l'exome sequencing si identificano da 20000 a 40000 varianti di singoli nucleotidi (SNV) (Ng et al., 2009) con differenze popolazione-specifiche (fig.3).

I criteri di filtraggio delle varianti nel caso di patologie mono o oligogeniche sono generalmente basati sulle seguenti considerazioni: 1) mutazioni in uno o pochi geni sono generalmente sufficienti a causare la malattia, 2) tali mutazioni dovrebbero essere rare e a elevata penetranza, 3) per il loro impatto sono probabilmente localizzate nella parte genica codificante (mutazioni missenso, nonsense, frameshift o mutazioni di splicing).

Altri fattori che si considerano nelle strategie di filtraggio sono la modalità di ereditarietà del fenotipo, la possibilità di mutazioni *de novo* e la potenziale eterogeneità genetica per un locus.

Tipo di variante	N medio di varianti (+ds) Afro-Americani	N medio di varianti (+ds) Euro-Americani
<b>Nuove varianti</b>		
Missenso	303 (±32)	192 (±21)
Non senso	5 (±2)	5 (±2)
Sinonime	209 (±26)	109 (±16)
Splicing	2 (±1)	2 (±1)
Totale	520 (±53)	307 (±33)
<b>Varianti note</b>		
Missenso	10828 (±342)	9319 (±233)
Non senso	98 (±8)	89 (±6)
Sinonime	12567 (±416)	10536 (±280)
Splicing	36 (±4)	32 (±3)
Totale	23529 (±751)	19976 (±505)
<b>Varianti totali</b>		
Missenso	11131 (±364)	9511 (±244)
Non senso	103 (±8)	93 (±6)
Sinonime	12776 (±434)	10645 (±286)
Splicing	38 (±5)	34 (±4)
Totale	24049 (±791)	20283 (±532)

**Fig.3** Numero medio di varianti identificate in progetti di sequenziamento esomico negli Afro-Americani e negli Euro-Americani (Ng et al., 2009).

#### *Filtraggio contro set di polimorfismi*

Dal momento che le varianti identificate con il sequenziamento NGS sono per il 95% annotate nei database, esse possono essere filtrate contro set di polimorfismi utilizzando database pubblici come dbSNP (<http://www.ncbi.nlm.nih.gov/snp>) o 1000 Genomi (<http://www.1000genomes.org/>), che raccoglie le varianti genomiche individuate con progetti di sequenziamento NGS che abbiano frequenza pari ad almeno l'1% nella popolazione.

Il vantaggio principale di questo metodo di filtraggio è dato dal fatto che solo il 2% delle variazioni di singoli nucleotidi (SNV, single nucleotide variants) di un individuo sono “nuove”, cioè non annotate nei database di riferimento per i polimorfismi, per cui il

sequenziamento di un numero anche modesto di affetti e l'applicazione del filtraggio può ridurre considerevolmente il numero di geni candidati risultando sufficientemente "potente" per l'identificazione di varianti responsabili di malattie mendeliane rare.

Applicare un filtraggio di questo tipo implica assumere che all'interno del filtering set non siano inclusi alleli di individui che presentano il fenotipo oggetto di studio incrementando il rischio di escludere dall'analisi alleli patogenetici. Il database dbSNP, per esempio, è, in una certa misura, "contaminato" da mutazioni patogenetiche. Inoltre, dal momento che sono sempre più numerosi i risultati dei progetti di genome ed exome sequencing, il filtraggio di alleli basato sulla loro Minor Allele Frequency (MAF) aumenta la probabilità di escludere dall'analisi alleli che segregano nella popolazione con una bassa ma apprezzabile frequenza, soprattutto nel caso di patologie ad ereditarietà recessiva (Bamshad et al., 2011).

Alternativamente al filtraggio con i database pubblici si possono confrontare le varianti rilevate con varianti di controllo della popolazione cui appartengono i campioni sequenziati, riducendo in tal modo il rischio di escludere potenziali varianti patogenetiche.

Gli alleli candidati possono essere ulteriormente stratificati sulla base della predetta patogenicità, esaminando quindi solo le varianti che determinano mutazioni non senso e missenso, le mutazioni frameshift, e le varianti che alterano i siti canonici di splicing.

Programmi come SIFT ([sift.jcvi.org/](http://sift.jcvi.org/)) e PolyPhen2 ([genetics.bwh.harvard.edu/pph2/](http://genetics.bwh.harvard.edu/pph2/)) saggiano la potenziale patogenicità di una variante a seconda del tipo di sostituzione aminoacidica cui danno luogo.

Infine, si possono selezionare varianti presumibilmente patogenetiche considerando che, spesso, quelle localizzate all'interno di regioni conservate nella scala evolutiva sono deleterie. Le analisi comparative di sequenze ortologhe possono essere effettuate mediante i programmi phastCONS (<http://compgen.bscb.cornell.edu/phast/>) e GERP (Davydov et al., 2010).

#### *Filtraggio per modello di ereditarietà*

Il tipo di ereditarietà di una patologia genetica influenza sia il disegno sperimentale e, quindi, la scelta del numero di casi da sequenziare, sia l'approccio analitico conseguente.

Il filtraggio è più efficace nel caso di patologie recessive, per cui anche il numero richiesto di casi da sequenziare è minore; si stima infatti che, in un individuo, siano presenti nel genoma un numero di geni con due alleli mutati inferiore di 50 volte al numero di geni con un solo allele variante (Bamshad et al., Nature Reviews Genetics 2011).

*Il sequenziamento di individui correlati.* Per i fenotipi Mendeliani, l'utilizzo di informazioni derivate dall'albero genealogico può indirizzare verso la scelta del soggetto più informativo da sequenziare. Se non sono disponibili dati di analisi di linkage, e avendo un albero genealogico esteso, la scelta più appropriata sarebbe quella di sequenziare gli individui con lo stesso fenotipo più distanti nell'albero genealogico in quanto, per gli alleli molto rari, la probabilità di identità per discendenza è molto elevata anche nel caso di individui non strettamente correlati.

Il sequenziamento del trio genitori-probanda è l'approccio più appropriato nel caso in cui si ricerchino mutazioni *de novo*, in patologie con casi sporadici, quando si sospetta una trasmissione con modalità dominante o ci si aspetta eterogeneità di locus.

Lo studio dell'esoma in individui correlati può semplificare l'identificazione di nuove varianti o di geni candidati anche nel caso di patologie molto eterogenee dal punto di vista clinico e genetico, come le atassie spinocerebellari autosomiche dominanti (Spinocerebellar ataxia, SCA). Nel 2010, il gruppo di Wang ha identificato in una famiglia Cinese con quattro individui affetti da SCA-35 (MIM 613908) una mutazione missenso precedentemente non descritta nel gene TGM6. Il risultato dell'exome sequencing è stato ulteriormente confermato da analisi di linkage, che ha evidenziato la presenza della mutazione in una regione condivisa dagli affetti (Wang et al., 2010).

Integrare i dati di sequenza con quelli del mappaggio per omozigotà può semplificare la ricerca del gene-malattia in famiglie consanguinee con patologie ad ereditarietà recessiva. Il gruppo di Bolze, per esempio, ha identificato le varianti causative responsabili di una sindrome caratterizzata da infezioni ricorrenti, encefalopatia, disfunzione epatica e malformazioni cardiovascolari (MIM 613759) individuando due regioni di omozigotà nei cromosomi 11 e 18 in tre affetti correlati e al contempo sequenziando l'esoma di uno dei pazienti. Tali analisi hanno permesso di restringere il

numero di varianti da più di 23000 a 67 nel cromosoma 11 e a 14 nel cromosoma 18. Il filtraggio con i database di SNP ha portato all'identificazione di una sola mutazione non sinonima in un gene (FADD) localizzato nella regione condivisa del cromosoma 11 (Bolze et al., 2010).

*Il sequenziamento di individui non correlati.* Il sequenziamento di nuova generazione ha permesso di superare uno dei maggiori ostacoli che si incontrano nello studio di malattie mendeliane estremamente rare, ovvero la scarsa disponibilità di affetti.

Nel 2009, Ng et al. hanno individuato mutazioni nel gene MYH3 in quattro casi non correlati affetti dalla Sindrome di Freeman-Sheldon (FSS, MIM 193700) (Ng et al, 2009).

Lo stesso gruppo ha identificato le varianti responsabili della Sindrome di Kabuki (MIM 147920), una patologia rara geneticamente e fenotipicamente eterogenea, attraverso lo studio dell'esoma di dieci individui non correlati. L'analisi ha portato all'individuazione di mutazioni nel gene MLL2 in sette pazienti. Il sequenziamento Sanger ha poi consentito di individuare mutazioni nel gene in due dei dieci pazienti che erano stati sottoposti ad exome sequencing e in altri 26/43 casi studiati. Il lavoro di Ng evidenzia l'importanza della definizione accurata del fenotipo clinico del campione di pazienti il cui genoma debba essere analizzato con metodiche di sequenziamento NGS.

## **Il sequenziamento NGS per la diagnosi molecolare di mutazioni in geni noti**

Oltre che come strategia di identificazione di nuovi geni-malattia, il sequenziamento NGS viene oggi sfruttato in ambito clinico per la diagnosi di malattie a base genetica nota.

Uno dei primi progetti di sequenziamento genomico per la delucidazione delle basi di una malattia genetica è stato pubblicato nel 2010 nel *The New England Journal Of Medicine* e rappresenta, di fatto, la prima applicazione del sequenziamento WGS per la diagnosi di una malattia genetica. Nel lavoro il gruppo di Lupski, infatti, riporta i risultati del sequenziamento genomico per la caratterizzazione degli alleli responsabili della Sindrome di Charcot-Marie-Tooth (CMT, MIM 118200, 118220). La Sindrome di Charcot-Marie-Tooth è una neuropatia ereditaria per la quale sono noti almeno 39 loci di suscettibilità che possono segregare con modalità autosomica dominante, autosomica recessiva o X-linked. Il sequenziamento del genoma è stato condotto nel probando di una famiglia in cui la malattia segregava con modalità autosomica recessiva e in cui il resequencing di alcuni dei geni candidati non aveva portato all'individuazione di varianti patogenetiche. L'analisi delle varianti rilevate con il sequenziamento WGS ha comportato l'elaborazione di una quantità di dati di sequenza pari a 89.6 Gb e il rilevamento di 3420306 SNPs, di cui più di due milioni localizzati nelle regioni extrageniche e circa un milione all'interno di regioni geniche. L'applicazione di filtri consecutivi ha permesso di identificare nel gene SH3TC2 una mutazione non senso già associata alla patologia e una nuova mutazione missenso, presenti sia nel probando che negli altri membri della famiglia affetti (Lupski et al., 2010).

Il sequenziamento genomico del probando CMT non ha determinato l'individuazione di nuovi geni-malattia ma ha permesso di effettuare la diagnosi molecolare, negli affetti della famiglia in esame, di una malattia altamente eterogenea dal punto di vista genetico quale è la sindrome di Charcot-Marie-Tooth.

Recentemente, protocolli di sequenziamento massivo parallelo sono stati sviluppati con l'obiettivo di poter effettuare la diagnosi prenatale non invasiva di patologie genetiche attraverso la quantificazione del DNA fetale libero presente nel plasma materno. Grazie

al sequenziamento NGS effettuato su DNA estratto da plasma si ottengono infatti milioni di sequenze che derivano sia dal sequenziamento del DNA materno che da quello della frazione fetale. Quantificando gli alleli feto-specifici e gli alleli condivisi da madre e feto è possibile dedurre la concentrazione della parte fetale e verificare l'eventuale presenza di mutazioni per la diagnosi molecolare di malattie monogeniche o di sbilanciamenti cromosomici nel caso di aneuploidie (Papasavva et al., 2013; Liang et al., 2013).

### **Il sequenziamento NGS per la diagnosi clinica**

In alcuni casi il sequenziamento NGS ha permesso di effettuare o rivalutare la diagnosi clinica di patologie genetiche. Choi et al., per esempio, hanno identificato mutazioni nel gene SLC26A3, associato alla CLD (Congenital Chloride Diarrhea, MIM 214700) in un paziente in cui la prima diagnosi differenziale era di sindrome di Bartter (MIM 607364), consentendo di rivalutare e modificare la diagnosi clinica iniziale (Choi et al., 2009).

Il gruppo di Worthey ha invece riportato un caso in cui il rilevamento di mutazioni nel gene XIAP ha consentito di poter effettuare in un paziente diagnosi di malattia infiammatoria intestinale (MIM 266600) (Worthey et al., 2011).

## **La tecnologia di sequenziamento “Next Generation”**

Le piattaforme che vengono impiegate attualmente nei progetti di sequenziamento massivo parallelo si caratterizzano per una elevata processività, accuratezza dei dati di sequenza (oltre il 99% per alcune piattaforme) e costi sostanzialmente ridotti rispetto ai sequenziatori ad elettroforesi capillare basati sulla biochimica di Sanger (Lin et al., 2012). Le piattaforme di sequenziamento NGS comprendono il 454 della 454 (acquisito in seguito dalla ditta Roche), introdotto nel mercato nel 2005, il Genome Analyzer Solexa, acquisito nel 2007 da Illumina, il SOLID (Life Technologies), il Polonator (Dover/Harvard) e l' HeliScope (Helicos).

Accanto a queste piattaforme, che restituiscono un output di dati che va da 0.7 Gb nel caso del 454 GsFlex (Roche) a 600 Gb nel caso dell'HiSeq2000 (Illumina) e che, per questo, vengono comunemente impiegate nei progetti di sequenziamento genomico ed esomico, sono attualmente presenti sul mercato sequenziatori come il Personal Genome Machiner (PGM, Life Technologies) e il MiSeq (Illumina), che offrono un throughput di dati minore, che va da 10 a 100 MB nel caso del PGM e da 120 a 1.5 GB nel caso del MiSeq. Per le loro performance, il PGM e il MiSeq vengono utilizzati principalmente in ambito clinico-diagnostico nel sequenziamento di pannelli di geni o di ampliconi per la diagnosi molecolare di malattie genetiche.

Sebbene si differenzino per il tipo di biochimica alla base e per i metodi di acquisizione e di elaborazione dei dati, le piattaforme NGS sono accomunate da un work-flow operativo simile che prevede tre fasi principali:

1. Preparazione del template
2. Sequenziamento e imaging.
3. Analisi bioinformatica dei dati.

### **1. Preparazione del template**

La preparazione del template avviene mediante la costruzione di una libreria di DNA a partire da quantità variabili di DNA genomico (3-20µg). La libreria è il risultato della frammentazione del DNA in molecole di lunghezza appropriata al sequenziamento e del



loro fissaggio su supporti solidi. L'immobilizzazione delle molecole di template spazialmente separate permette a migliaia di reazioni di sequenziamento di avvenire in maniera simultanea.

La criticità maggiore di questa fase è rappresentata dalla qualità del materiale genomico di partenza, che deve essere privo di contaminanti (RNA, proteine, fenolo o altri solventi utilizzati per l'estrazione) e non degradato. Un DNA degradato o non puro può infatti compromettere la fase di preparazione della libreria e le successive applicazioni.

Nel caso in cui si vogliono analizzare l'esoma (fig.4) o specifiche porzioni genomiche, la libreria deve essere allestita a partire dalle regioni di interesse che si intendono sequenziare.

L'arricchimento di specifiche regioni genomiche, estese da centinaia di Kb fino a diverse Mb nel caso dell'intero esoma, viene effettuato mediante ibridizzazione in fase liquida o solida utilizzando probes di RNA (Agilent) o DNA (Roche NimbleGen).

Per l'esoma, il targeting medio consentito dai kit è pari a 30Mb che comprendono le sequenze esoniche, i siti di splicing e le regioni UTR. I kit di ultima generazione, come il TruSeq Exome Enrichment Kit della Illumina, permettono di estendere il targeting fino a 62Mb coprendo più di 200000 esoni e 20794 geni.

*Amplificazione clonale della libreria.* Gli strumenti NGS hanno sistemi di analisi dei dati di immagine che non sono in grado di rilevare emissioni di fluorescenza singole, per cui la pre-amplificazione dei template, tramite PCR, permette di generare un segnale di fluorescenza sufficientemente intenso da poter essere rilevato dallo strumento.

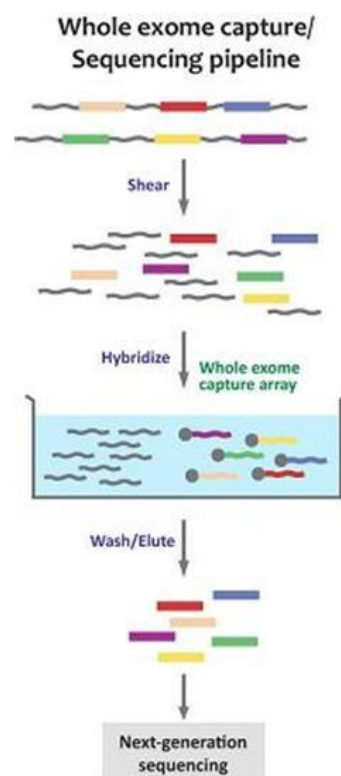
L'amplificazione della libreria può avvenire mediante PCR in emulsione (*emPCR*) o amplificazione in fase solida. Entrambe le metodiche consentono di ottenere milioni di copie identiche (cloni) delle molecole che compongono la libreria.

L'*emPCR* richiede che, alle estremità delle molecole della libreria, vengano legati oligonucleotidi universali a sequenza nota, chiamati "adattatori", che funzionano come primers consentendo la reazione di amplificazione. Dopo denaturazione, i filamenti single-stranded della libreria vengono "catturati" sulla superficie di biglie, in condizioni chimiche tali da favorire il legame di una singola molecola per biglia. La PCR ha luogo in

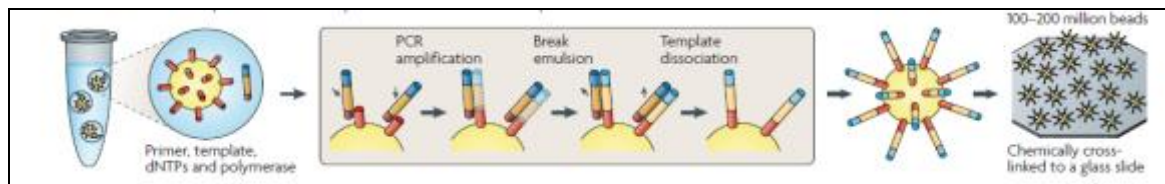
un'emulsione acqua-olio, contenente il complesso biglia-DNA, che agisce come un microreattore all'interno del quale si verifica la reazione di amplificazione.

Dopo la rottura dell'emulsione, i clusters vengono legati chimicamente sulla superficie di un supporto di vetro (Life/APG; Polonator) o depositati nei pozzetti di una PicoTiterPlate (Roche/454) su cui avverrà la reazione di sequenziamento (fig.5).

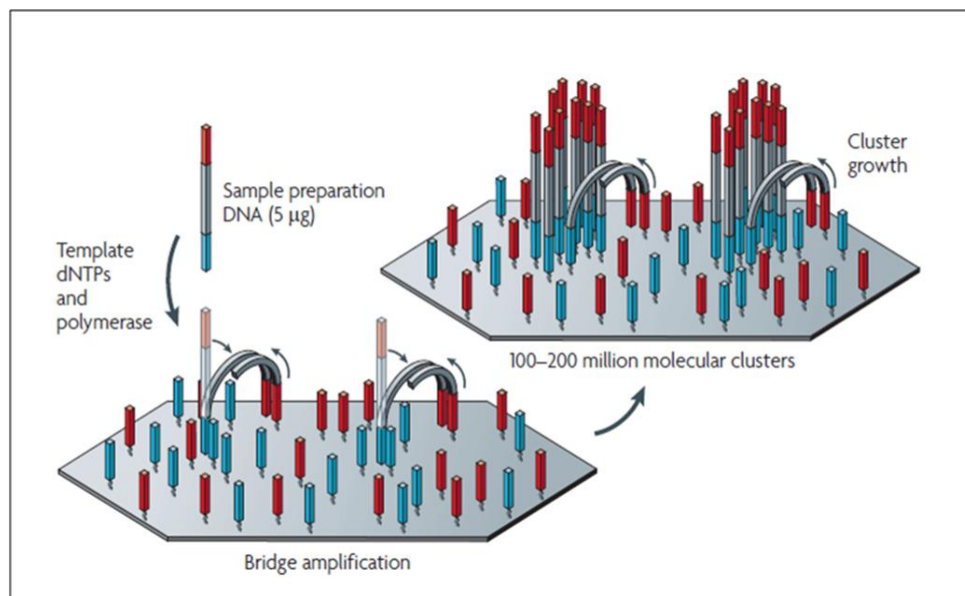
L'*amplificazione in fase solida* prevede dapprima il legame dei primer e l'estensione della molecola di template single-stranded sulla superficie di un vetrino. Successivamente, le molecole di template immobilizzate sul vetrino vengono amplificate con un meccanismo a ponte (*bridge PCR*) generando 100-200 milioni di clusters fisicamente separati aventi un'estremità libera su cui viene ibridizzato il primer specifico per la reazione di sequenziamento (fig.6)



**Fig.4** Workflow del processo di cattura dell'esoma (Gerald Goh, Murim Choi)



**Fig.5 PCR in emulsione.** Nella emPCR una miscela acqua-olio incapsula le biglie legate alle molecole di DNA. L'amplificazione consente di ottenere milioni di copie clonali del template di DNA legato alla biglia. Il prodotto della emPCR può essere depositato sulla superficie di picotiter plates o di supporti di vetro. (Metzker ML, 2009)



**Fig.6 Amplificazione in fase solida.** Nella PCR in fase solida le molecole template di DNA single-stranded vengono immobilizzate tramite adattatori su un supporto solido. L'amplificazione avviene con un meccanismo "a ponte" per ibridizzazione con i primers delle molecole adiacenti, portando alla formazione di milioni di clusters. (Metzker ML, 2009)

## 2. Sequenziamento e imaging

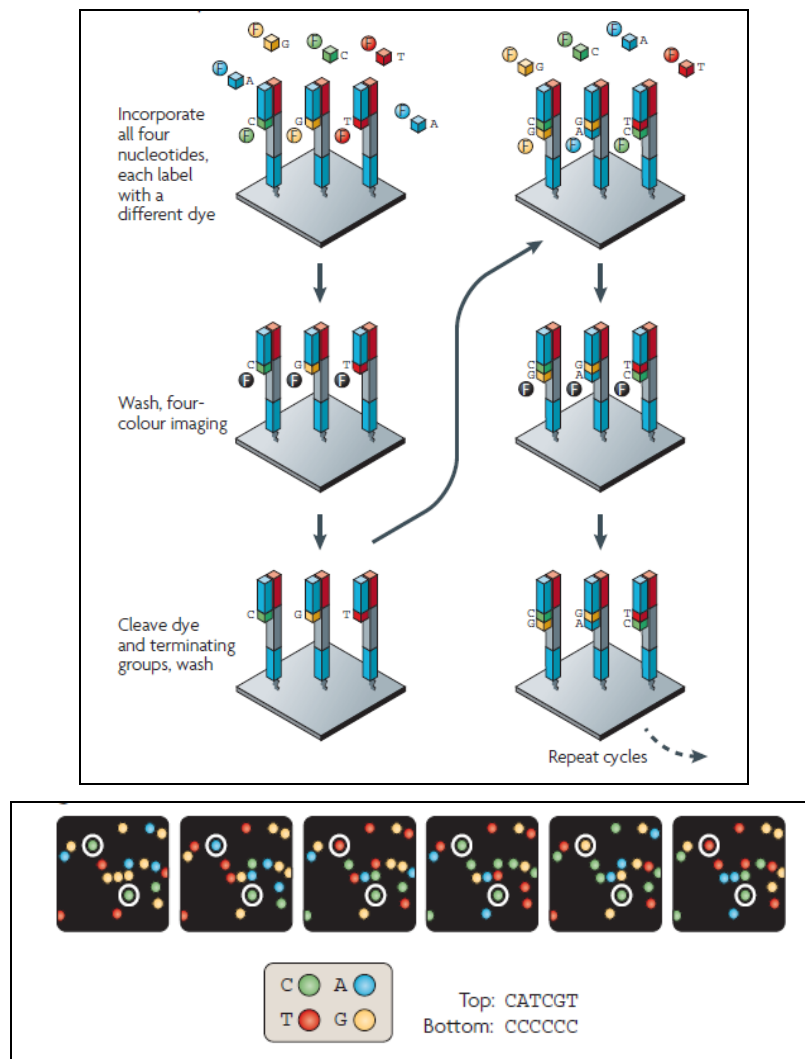
*Cyclic reversible termination (CRT).* Il sequenziamento ciclico con terminatori reversibili (CRT) si compie in un numero variabile di cicli ciascuno dei quali si compone di (I) una fase di incorporamento di nucleotidi terminatori, modificati e marcati con fluorocromi, ad opera di una DNA polimerasi, (II) sintesi di una nuova catena complementare al

templato, (III) acquisizione del segnale di fluorescenza tramite imaging e rimozione (*cleavage*) dei terminatori (fig.7).

Nel sequenziamento CRT i terminatori reversibili possono essere bloccati oppure non bloccati in posizione 3'.

Le piattaforme Illumina associano la metodica di amplificazione clonale in fase solida al metodo di sequenziamento con terminatori reversibili bloccati reversibilmente in 3' da fluorocromi specifici per ciascun nucleotide. L'inserimento di una base da parte della DNA polimerasi causa l'arresto della sintesi durante il quale il segnale di fluorescenza base-specifico viene acquisito in contemporanea per i quattro fluorocromi ("*four-colour imaging*"). Il clivaggio successivo rimuove i fluorocromi terminatori e ripristina il gruppo ossidrilico in 3' permettendo alla polimerasi di proseguire nella sintesi della catena in nuovo ciclo di sequenziamento. Il principale limite di questo approccio è dato dagli errori di incorporazione dei nucleotidi modificati in 3', che porta alla determinazione di false variazioni di singoli nucleotidi, con un'elevata percentuale di errore che si verifica quando la base incorporata è una G. I nucleotidi bloccati richiedono, inoltre, l'utilizzo di una DNA polimerasi mutagenizzata che sia in grado di incorporare in modo efficace le basi modificate.

I terminatori reversibili non bloccati in 3' vengono incorporati in modo più efficace e consentono l'utilizzo di una DNA polimerasi non mutagenizzata. La differenza principale tra i due tipi di terminatori è che l'arresto della sintesi della catena si verifica dopo l'incorporamento di ogni singolo nucleotide. La LaserGen Inc. e la Helicos Biosciences hanno sviluppato terminatori di questa tipologia rappresentati dai "Lightning terminators" e dai "Virtual terminators".



Sequenziamento CRT (Ciclyc Reversible Termination)

**Fig.7 Sequenziamento CRT.** Nel sequenziamento CRT una DNA polimerasi mutagenizzata incorpora in ogni ciclo terminatori reversibili bloccati reversibilmente in 3'. L'arresto della sintesi permette di rilevare in contemporanea la fluorescenza relativa ai quattro dinucleotidi per TIRF (Total Internal Reflection Fluorescence). Il clivaggio dei terminatori dà avvio al secondo ciclo di sequenziamento. (Metzker, 2009)

*Sequencing by ligation (SBL).* Il sequenziamento SBL prevede l'utilizzo di una DNA ligasi e cosiddette "one base" o "two bases" probes (1-probes e 1,2-probes).

Le 1-probes sono sequenze oligonucleotidiche ottameriche in cui la prima base è associata con un particolare terminatore, per esempio la A al verde, mentre le altre sono

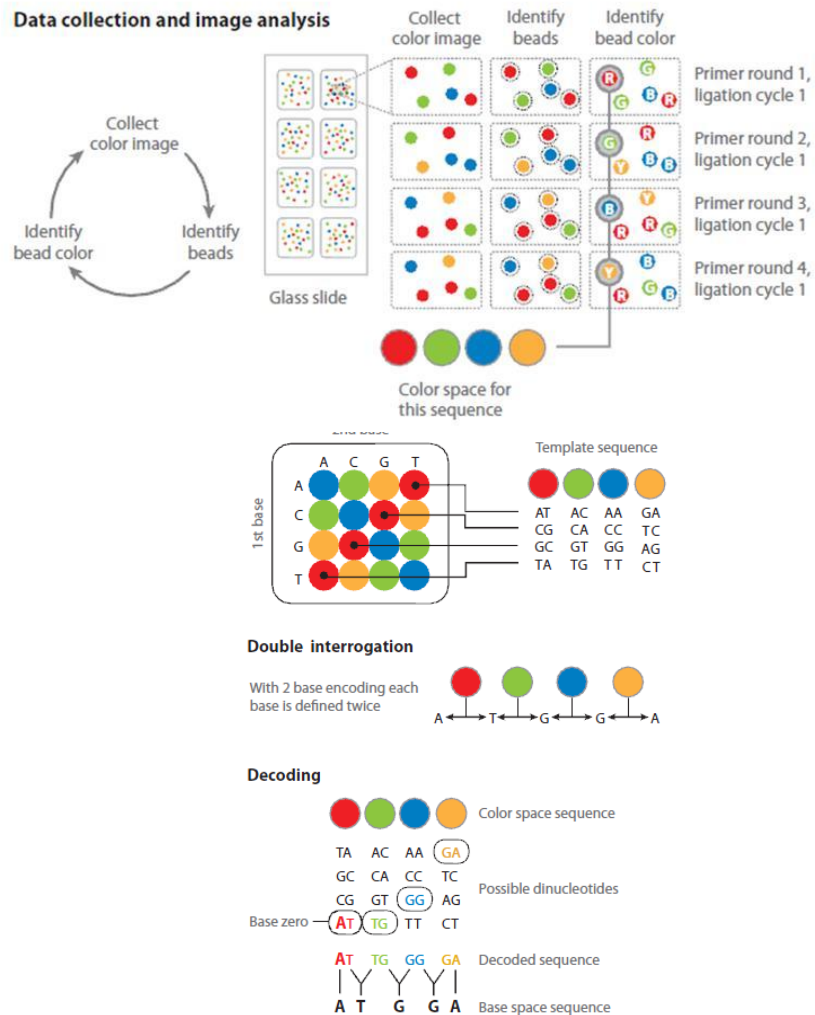
quattro possibili basi degenerate o universali. La prima base è quella che viene “interrogata” durante la lettura della sequenza.

Le 1,2-probes sono sequenze oligonucleotidiche in cui due basi vengono interrogate e sono associate con un particolare terminatore, per esempio AA, CC, GG e TT sono associate al terminatore blu. Anche in questo caso, le altre basi della probe sono degenerate o universali. Come mostrato in figura, una probe fluorescente si ibrida con la sequenza complementare adiacente al primer a cui viene legata da una DNA ligasi. Il legame della probe al primer può avvenire sia dal gruppo fosfato in 5' sia dal gruppo ossidrilico in 3'. Il four-colour imaging consente di determinare l'identità della probe ligata.

La piattaforma SOLiD (Support Oligonucleotide Ligation Detection, Life Technologies) utilizza le 1,2-probes. Nel SOLiD ogni ciclo è costituito da (I) fase di ibridizzazione delle 1,2-probes, (II) ligazione, (III) imaging e clivaggio delle probes (fig.8).

Alla fine di un ciclo, il primo primer viene rimosso e sostituito da un primer  $n-1$ , che resetta le basi interrogate precedentemente e i corrispondenti colori slittando di una posizione sulla sinistra. Si effettuano secondo questo schema cinque cicli, in ciascuno dei quali vengono performati dieci cicli di ligazione. Alla fine, i colori rilevati durante i cinque cicli vengono riordinati in una sequenza lineare costituendo il cosiddetto “color space”. Utilizzando le 1,2-probes, infatti, il segnale di fluorescenza che si ottiene con l'imaging è associato a quattro sequenze dinucleotidiche con una base in 3' e una base in 5'. Il color space corrisponde alla sequenza di dinucleotidi che, sovrapponendosi, vanno a codificare simultaneamente quattro sequenze nucleotidiche. L'allineamento con la sequenza di riferimento consente di tradurre il colour space in una specifica sequenza nucleotidica. Come per l'Illumina, l'errore più frequente sono le sostituzioni e una minore rappresentazione delle regioni AT-rich e GC-rich.



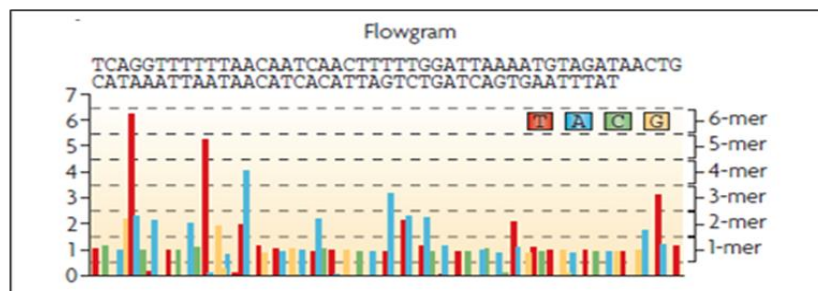
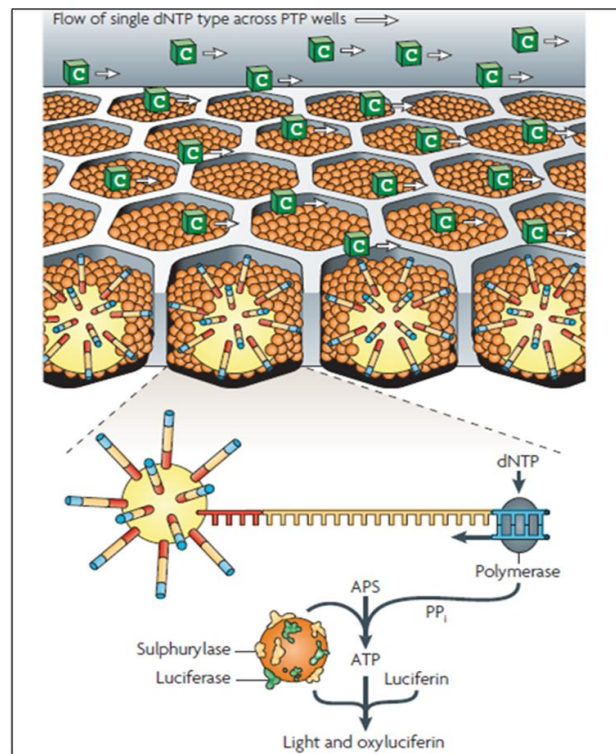


**Fig.8 Sequencing by ligation (SBL).** Nel SOLiD (Life Technologies), i frammenti di DNA sono amplificati sulla superficie di biglie magnetiche e depositati su un supporto solido. Il sequenziamento è mediato dall'enzima ligasi e comincia con l'annealing di un primer a partire da sequenze adattatrici presenti in ogni frammento. Al termine di ogni ligazione si ha il rilevamento della fluorescenza e la sostituzione del primer che dà avvio ad un altro ciclo. Poiché ogni gruppo di fluorescenza identifica una combinazione di due basi, la sequenza della reads può essere dedotta allineando le singole reads alla sequenza di riferimento (Mardis, 2008).



*Pirosequenziamento.* Il pirosequenziamento è un metodo di sequenziamento non elettroforetico, basato sulla rilevazione di bioluminescenza che si verifica in seguito al rilascio di pirofosfato inorganico e alla sua conversione in luce visibile attraverso una serie di reazioni enzimatiche (fig.9).

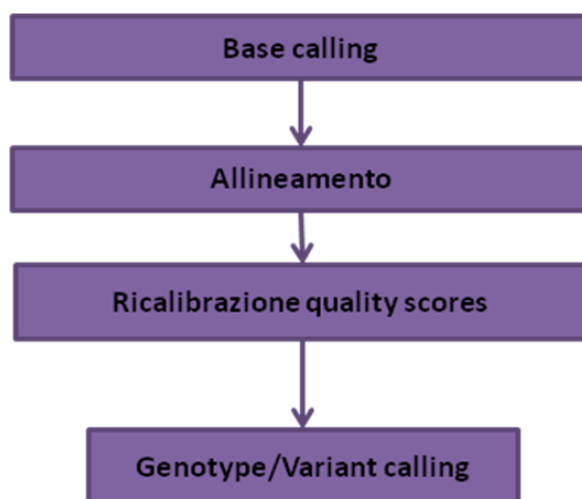
Il template di DNA viene preparato mediante emPCR e gli ampliconi clonali sono catturati sulla superficie di biglie di 28µm. Le biglie associate alle molecole di template vengono incubate con una Bst polimerasi (*Bacillus stearothermophilus*) e depositate nei pozzetti di una picotiter plate costituita da 1-2 milioni di pozzetti (PTP). Ulteriori biglie contenenti gli enzimi necessari al pirosequenziamento (luciferasi e ATP sulfurilase) vengono depositate nei pozzetti in modo da circondare le biglie legate ai templati. Durante ciascuno dei centinaia di cicli di sequenziamento, i dNTPs vengono aggiunti singolarmente secondo un ordine predeterminato. Ogni volta che si ha l'incorporamento di un dNTP in un template, si verifica il rilascio di una molecola di pirofosfato e, grazie all'ATP sulfurilasi e alla luciferasi, si genera un segnale luminescente che viene captato in corrispondenza di ogni singolo pozzetto da una "charge-coupled device camera (CCD)". L'errore più frequente legato al pirosequenziamento è dato da falsi positivi di inserzione/delezione di basi, frequenti in caso di stretch contigui di basi ripetute (omopolimeri).



**Fig.9 Pirosequenziamento.** Nel 454 (Roche) gli ampliconi clonali legati alle biglie magnetiche sono depositati su una picotiter plate ed incubati con una *Bst* polimerasi. Ogni volta che si ha l'incorporamento di un nucleotide, si verifica il rilascio di pirofosfato che viene convertito, in una serie di reazioni enzimatiche mediate dall'ATP sulfirase e dalla luciferasi, in segnale luminescente captato in corrispondenza di ogni pozzetto della plate da una CCD. La luce generata dalla cascata enzimatica è registrata come una serie di picchi che costituiscono il flowgramma (Metzker, 2009).

### 3. Analisi bioinformatica

Dal punto di vista dell'analisi dei dati, la tecnologia NGS prevede un insieme di processi computazionali, schematizzati in figura 10.



**Fig.10** Fasi di analisi dei dati originati da sequenziamento NGS

#### *Base calling*

Per poter essere analizzate, le sequenze nucleotidiche devono essere “estratte” dai dati di immagine generati dalle piattaforme di sequenziamento.

Gli algoritmi di base-calling consentono di convertire l'informazione di immagine in informazione di sequenza. Poiché ciascuna piattaforma è prona ad un errore specifico, anche gli algoritmi si differenziano e sono specifici per la definizione delle basi nelle differenti piattaforme.

Per esempio, nel caso del pirosequenziamento, gli algoritmi di base calling sono formulati per poter determinare in maniera esatta il numero di nucleotidi nelle regioni omopolimeriche. Infatti, l'intensità di emissione di fluorescenza non correla, oltre un numero soglia di nucleotidi uguali incorporati, con il numero di basi effettivo, causando falsi positivi di variazioni indels.

Nel processo di base-calling vengono inoltre corretti artefatti come il “crosstalk” e il “phasing”. Il “crosstalk” è causato dalle sovrapposizioni di emissioni di fluorescenza di nucleotidi diversi. Il “phasing” è dovuto alla dispersione e alla diffusione del segnale tra i cicli, che si verifica quando le molecole di un cluster non incorporano un nucleotide

oppure quando la sintesi non si arresta dopo l'incorporamento. Questo fa sì che se per esempio una guanina non è inserita nel ciclo 2, si rilevi il segnale aspecifico di quella base nei cicli 1 e 3.

Con il base-calling a ciascuna base viene assegnato un valore sulla qualità della definizione della base stessa, chiamato "Phred quality score" (Q). Un valore di Q20 corrisponde ad un errore nell'assegnazione della base pari a 1/100, mentre Q30 indica una probabilità di errore nella chiamata della base pari a 1/1000. Alcune piattaforme assegnano quality scores che sono specifici dello strumento ma che comunque possono essere convertiti nel quality score standard.

### *Allineamento*

Le short reads sono generate a partire da frammenti di DNA relativamente lunghi (da 200 bp a 8000 bp) che vengono sequenziati per un certo numero di basi: da 35 bp a 150 bp nel caso delle piattaforme Illumina, in media 400 bp nel caso della piattaforma 454 Roche, a partire da una o da entrambe le estremità dei frammenti (single-end o paired-end reads). Obiettivo dell'allineamento è determinare l'esatta localizzazione delle reads su una sequenza di riferimento.

Attualmente sono disponibili un insieme di softwares specificatamente disegnati o modificati per allineare reads derivate dal sequenziamento NGS. In genere, tali strumenti bioinformatici sono predisposti per l'allineamento delle short reads di una specie sul genoma di riferimento relativo alla stessa specie. Gli algoritmi sono infatti formulati in modo tale da eseguire il mappaggio considerando fattori quali la variabilità genetica di una popolazione (ci si aspetta di trovare un certo numero di mismatches), l'errore nel sequenziamento, la lunghezza e la quantità di reads ottenute.

È possibile allineare le reads di una specie con il genoma di riferimento di un'altra nel caso in cui si voglia, per esempio, valutare la distanza genetica fra specie differenti.

In generale, l'allineamento è più difficile per le regioni che risultano molto diverse rispetto alla sequenza di riferimento. Questa difficoltà può essere in parte superata producendo reads di lunghezza maggiore e "paired-end", che derivano dal

sequenziamento sia in direzione 5'-3' (forward) che in direzione 3'-5' (reverse) della molecola di DNA templato.

Per il corretto assemblaggio della sequenza un fattore estremamente importante è il "coverage". Il coverage rappresenta infatti il numero di volte in cui una sequenza viene allineata in corrispondenza della sequenza di riferimento. Un coverage medio di 30X per un gene indica quindi che ogni nucleotide del gene è rappresentato mediamente 30 volte in short reads diverse e sovrapponibili.

Le reads troppo corte non possono essere interpretate perché possono essere allineate e corrispondere a sequenze presenti in più parti del genoma, mentre le reads che possono essere assemblate e corrispondono ad una sequenza unica nel riferimento rappresentano le "reads mappabili".

I programmi bioinformatici più frequentemente utilizzati per il processo di allineamento sono impostati su pochi algoritmi, tra cui quelli basati sulla trasformazione *Burrows Wheeler* (BWT) (Burrows and Wheeler, 1994; Li and Durbin, 2009) e sulle implementazioni "hash table-based" (Flicek and Birney, 2009).

Sia gli algoritmi hash-based che quelli basati su Burrows-Wheeler possono essere applicati per l'allineamento delle reads delle piattaforme Life Technologies, Illumina e Roche.

I programmi di allineamento basati su BWT, come BWA (Li and Durbin, 2010), risultano veloci ed efficaci nell'allineamento delle short reads, ma sono complessivamente meno sensibili rispetto agli algoritmi hash-based. I programmi MAQ, SOAP, SHRiMP, Novoalign e Stampy, che utilizzano l'approccio hash-based, sono invece tra gli aligners che producono i risultati di allineamento più accurati.

#### *Calibrazione dei quality scores*

Il Phred quality score derivato dagli algoritmi di allineamento spesso non riflette in modo accurato l'errore reale dovuto al base calling, per cui è necessario "ricalibrarlo".

In generale, il quality score viene ricalibrato tenendo conto del valore del raw quality score, della posizione relativa della base all'interno della read e del contesto dinucleotidico in cui è inserita la base.

Il programma SOAPsnp, per esempio, effettua la ricalibrazione confrontando le sequenze ottenute con la sequenza di riferimento in corrispondenza di siti che non presentano varianti note. Il software GATK invece basa la ricalibrazione del quality score su fattori quali il ciclo di sequenziamento e il contesto dinucleotidico in cui è inserita la base (Nielsen R et al., 2011).

La ricalibrazione dei quality scores è importante per la determinazione precisa delle varianti e dei genotipi (Variant e Genotype calling).

### *Variant/Genotype calling*

Una volta allineate le reads al genoma di riferimento, il processo che consente di convertire base calls e quality scores in un insieme di genotipi si articola in due fasi che prendono il nome di “variant calling” e “genotype calling”.

Il processo di variant calling determina in quali posizioni sono presenti polimorfismi o una base differisce rispetto alla sequenza di riferimento.

Il genotype calling è invece il processo mediante cui si definisce il genotipo in una determinata posizione e, generalmente, si effettua solo per i siti che differiscono rispetto alla sequenza di riferimento.

La chiamata delle varianti e la definizione dei genotipi sono processi che risentono degli errori che possono derivare dall’allineamento e dalla chiamata delle basi.

Il grado di incertezza associato ai due processi può essere ridotto sequenziando le regioni target con un coverage sufficientemente elevato, superiore a 20X. Infatti, l’errore più frequente in cui si incorre quando il coverage è molto basso (<5X) è quello di ottenere la sequenza di uno solo dei due cromosomi di un individuo diploide.

Soprattutto nei primi progetti di sequenziamento NGS, variant e genotype calling venivano eseguiti impostando dei valori di cutoff, solitamente rappresentato da un valore di Phred quality score pari a 20 (Q=20). Secondo un cutoff di questo tipo, l’individuo viene considerato eterozigote se la proporzione fra l’allele reference e il non reference varia fra il 20% e l’80%; in caso contrario, il genotipo sarà considerato omozigote (Nielsen et al., 2011). Una determinazione dei genotipi di questo tipo è

efficace se il coverage è sufficientemente alto (superiore a 20X), per cui la probabilità che un individuo sia eterozigote al di fuori del range 20-80% è molto bassa.

Il principale limite di questo metodo è che nel caso di coverage basso aumenta la probabilità di sottostimare i genotipi eterozigoti. Inoltre, il filtraggio basato sul cutoff dei quality scores determina la perdita di informazioni sulla qualità delle singole reads e non dà una stima del grado di incertezza associato al genotipo calcolato.

Per poter associare una misura del grado di incertezza si può determinare per ciascun genotipo una probabilità *a posteriori* a partire dai valori di quality scores (Li H et al., 2008; Li R et al., 2009). La probabilità a posteriori viene derivata dalla probabilità *a priori* mediante la formula di Bayes. Considerando quindi un genotipo a priori,  $p(G)$ , si calcola la probabilità a posteriori del genotipo  $G$  mediante la formula  $p(X|G)$ , dove  $X$  rappresenta tutti i dati della reads per un individuo in una posizione specifica.

Il genotipo con la probabilità a posteriori più elevata viene assegnato come genotipo definitivo, mentre il rapporto fra la prima e la seconda probabilità a posteriori più elevata viene usato per stimare il livello di confidenza del dato.

Nel calcolo delle probabilità a posteriori si considerano generalmente anche gli errori "correlati", ovvero gli errori che non si verificano in maniera indipendente per cui una base erroneamente chiamata in una posizione incrementa la possibilità di errore anche nei siti vicini.

La probabilità *a priori* viene calcolata in modo diverso a seconda che si consideri un singolo campione o più campioni sequenziati nello stesso esperimento.

Nel caso di un campione singolo, la probabilità a priori può essere calcolata assegnando a ciascun genotipo possibile lo stesso livello di probabilità di verificarsi oppure basandosi sui dati esterni disponibili, che possono essere quelli relativi alla sequenza di riferimento o i dati sui polimorfismi riportati nei database pubblici, come nel caso di SOAP2 (Short Oligonucleotide Analysis Package, Li et al., 2009),

Nel caso di analisi parallela di più campioni, la probabilità *a priori* può essere assegnata considerando le frequenze alleliche o genotipiche derivate da un campione sufficientemente ampio.

Se sono note le frequenze alleliche, le probabilità genotipiche possono essere derivate applicando la legge dell'equilibrio di Hardy-Weinberg (HWE) oppure sulla base della frequenza dei diversi genotipi nell'intero campione; sapendo per esempio che le probabilità dei genotipi AT e AA sono equivalenti, nel genotype calling di un singolo campione si potrà scegliere in modo indifferente se considerare nella probabilità a priori l'uno o l'altro genotipo. Se invece la frequenza dell'allele A è particolarmente bassa, si sceglierà nel genotype calling di assegnare il genotipo AT.

In alcuni programmi, possono essere integrate nell'analisi le informazioni sul linkage disequilibrium, utili nel caso in cui ci siano delle ambiguità nell'interpretazione di una sequenza in un campione di una popolazione in cui è nota la presenza di un aplotipo specifico.

La figura 11 estratta da Nature Reviews Genetics (Nielsen et al., 2011), riporta una lista dei programmi più utilizzati per i processi di calling.

Software	Available from	Calling method	Prerequisites	Comments
SOAP2	<a href="http://soap.genomics.org.cn/index.html">http://soap.genomics.org.cn/index.html</a>	Single-sample	High-quality variant database (for example, dbSNP)	Package for NGS data analysis, which includes a single individual genotype caller (SOAPsnp)
realSFS	<a href="http://128.32.118.212/thorfinn/realSFS/">http://128.32.118.212/thorfinn/realSFS/</a>	Single-sample	Aligned reads	Software for SNP and genotype calling using single individuals and allele frequencies. Site frequency spectrum (SFS) estimation
Samtools	<a href="http://samtools.sourceforge.net/">http://samtools.sourceforge.net/</a>	Multi-sample	Aligned reads	Package for manipulation of NGS alignments, which includes a computation of genotype likelihoods (samtools) and SNP and genotype calling (bcftools)
GATK	<a href="http://www.broadinstitute.org/gsa/wiki/index.php/The_Genome_Analysis_Toolkit">http://www.broadinstitute.org/gsa/wiki/index.php/The_Genome_Analysis_Toolkit</a>	Multi-sample	Aligned reads	Package for aligned NGS data analysis, which includes a SNP and genotype caller (Unified Genotyper), SNP filtering (Variant Filtration) and SNP quality recalibration (Variant Recalibrator)
Beagle	<a href="http://faculty.washington.edu/browning/beagle/beagle.html">http://faculty.washington.edu/browning/beagle/beagle.html</a>	Multi-sample LD	Candidate SNPs, genotype likelihoods	Software for imputation, phasing and association that includes a mode for genotype calling
IMPUTE2	<a href="http://mathgen.stats.ox.ac.uk/impute/impute_v2.html">http://mathgen.stats.ox.ac.uk/impute/impute_v2.html</a>	Multi-sample LD	Candidate SNPs, genotype likelihoods	Software for imputation and phasing, including a mode for genotype calling. Requires fine-scale linkage map
QCall	<a href="ftp://ftp.sanger.ac.uk/pub/rd/QCALL">ftp://ftp.sanger.ac.uk/pub/rd/QCALL</a>	Multi-sample LD	'Feasible' genealogies at a dense set of loci, genotype likelihoods	Software for SNP and genotype calling, including a method for generating candidate SNPs without LD information (NLDA) and a method for incorporating LD information (LDA). The 'feasible' genealogies can be generated using Margarita ( <a href="http://www.sanger.ac.uk/resources/software/margarita">http://www.sanger.ac.uk/resources/software/margarita</a> )
MaCH	<a href="http://genome.sph.umich.edu/wiki/Thunder">http://genome.sph.umich.edu/wiki/Thunder</a>	Multi-sample LD	Genotype likelihoods	Software for SNP and genotype calling, including a method (GPT_Freq) for generating candidate SNPs without LD information and a method (thunder_glf_freq) for incorporating LD information

**Fig.11** Programmi utilizzati per il processo di definizione delle basi nel sequenziamento NGS (Nielsen et al., 2009).



Anche se, in genere, il calcolo delle probabilità a posteriori per ciascun sito consente di ottenere dei dati sufficientemente precisi, a cui è associata anche una misura dell'errore relativo, talvolta si può decidere di filtrare i dati grezzi affinché il processo di calling risulti più accurato, per esempio nel caso di bias di strand per cui un filamento viene sequenziato più efficacemente dell'altro o per differenze nei valori di quality scores associati all'allele più frequente e a quello meno frequente.

Nel progetto 1000Genomi, ad esempio, sono stati eliminati tutti i dati di sequenza che presentavano grosse discrepanze con i genotipi rilevati nel progetto HapMap.

### **Limitazioni tecniche ed analitiche dell'exome sequencing**

Negli ultimi anni è stato spesso possibile caratterizzare nuovi geni-malattia grazie al sequenziamento dell'esoma e all'applicazione di varie strategie di filtraggio.

Più difficoltoso risulta effettuare una stima degli esiti negativi che, generalmente, possono derivare o da errori di tipo tecnico e/o da errori di tipo analitico.

L'errore tecnico è riconducibile alla mancata copertura di regioni genomiche più o meno estese, o perché non sono presenti in una determinata regione geni noti oppure perché la cattura fallisce in corrispondenza di certe sequenze per cui solo parte della sequenza di un gene è compresa nel targeting. Falsi negativi vengono determinati, inoltre, se il coverage è esiguo in certe regioni, per cattura inefficiente o per un basso grado di sequenziamento.

I falsi positivi possono essere generati da errori nell'allineamento delle reads ed essere rilevati in corrispondenza di regioni genomiche polimorfiche e geni di dimensioni elevate (Zhi et al., 2012). L'identificazione di nuove varianti richiede pertanto che esse vengano sempre validate mediante il risequenziamento diretto con il metodo di Sanger.

Altro limite tecnico dell'exome sequencing è che non consente la cattura e la conseguente analisi di elementi funzionali non codificanti, come tutte le sequenze regolatorie, e il rilevamento di variazioni strutturali come copy number variations (CNV), traslocazioni e inversioni (Goh and Choi, 2012).

L'errore di tipo analitico può derivare dalle limitazioni dei metodi di filtraggio dei dati descritti in precedenza.

Da una prospettiva analitica, la presenza di falsi negativi, la potenziale presenza di alleli patogenetici nel gruppo di controllo e la ridotta penetranza possono risultare in un rapporto segnale/rumore di fondo che è, talvolta, indistinguibile dall'eterogeneità genetica (Bamshad et al., 2011).

# LA SINDROME DI JEUNE

## Caratteristiche cliniche

La Sindrome di Jeune (JATD, Asphyxiating Thoracic Dystrophy, MIM 208500) è una osteocondrodiplosia rara, con incidenza stimata intorno a 1-5/500000 nati vivi, che si trasmette con modalità di ereditarietà autosomica recessiva.

Il fenotipo è caratterizzato da gravi malformazioni scheletriche (fig.12) comprendenti coste orizzontali corte e deformi che causano il restringimento e la caratteristica morfologia “a campana” della gabbia toracica (Langer, 1968; Oberklaid et al., 1977; Cortina et al., 1979), struttura irregolare delle epifisi e fusione tra epifisi e metafisi delle falangi distali e medie, metafisi delle ossa lunghe più corte della norma, clavicole anormali (a “manubrio di bicicletta”), irregolarità delle ossa pelviche con riduzione delle ossa iliache e forma a tridente dell’acetabolo, con o senza polidattilia di mani e/o piedi.

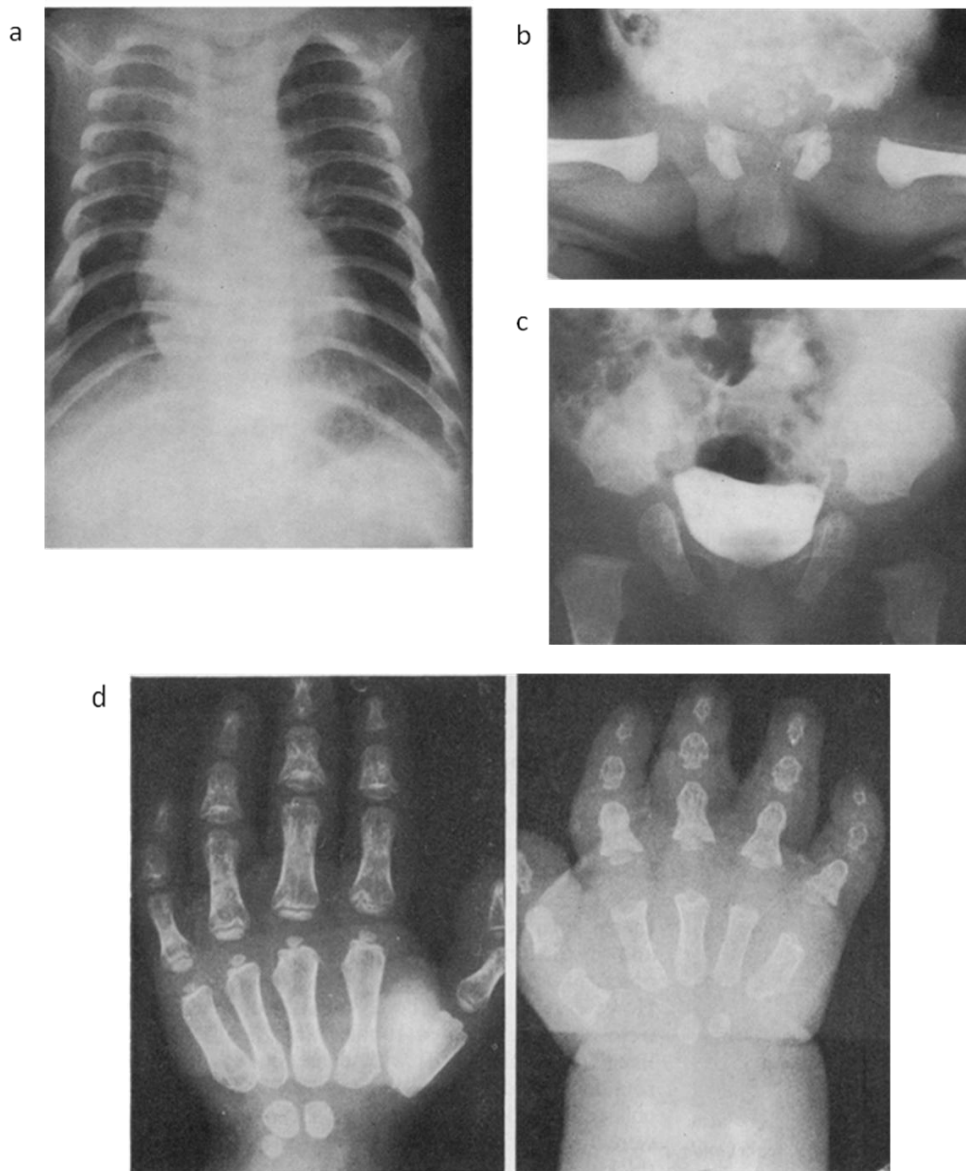
Talvolta si associano alle malformazioni di tipo scheletrico anomalie pancreatiche che includono fibrosi e cisti, iperbilirubinemia diretta con ittero prolungato, degenerazione retinica. Solo raramente sono state osservate alterazioni nello sviluppo cognitivo.

Il fenotipo della Sindrome di Jeune è variabile e può essere più o meno severo, letale o in alcuni casi latente (Tüysüz et al., 2013). La maggior parte dei pazienti è affetta dalla forma più grave, in cui il torace ristretto impedisce una corretta respirazione e predispone a gravi infezioni. L’asfissia rappresenta infatti causa di morte nel periodo perinatale per il 60% degli affetti (Oberklaid F, 1977). Approssimativamente solo 1/5 dei bambini affetti da JATD supera il periodo neonatale sviluppando patologia renale con cistogenesi e fibrosi periglomerulare che sfocia in insufficienza renale cronica. Anche il fegato può risultare fortemente compromesso con cirrosi biliare che causa morbidità precoce.

La Sindrome di Jeune è fenotipicamente correlata al gruppo delle Short Rib Polydactyly Syndromes (SRPs), un gruppo eterogeneo di sindromi malformative scheletriche, caratterizzate da espressività variabile e spesso letali in periodo perinatale.

Le SRPs sono le più frequenti osteocondrodiplosie ad eredità autosomica-recessiva.

Si distinguono sei condizioni autosomiche recessive di cui quattro letali: la SRP I o sindrome di Saldino-Noonan (MIM 263530), la SRP II o Sindrome di Majewski (MIM 263520), la SRP III o Sindrome di Verma-Naumoff (MIM 263510), e la SRP IV o Sindrome di Beemer-Langer (MIM 269860). Sebbene geneticamente distinte, le SRPs condividono caratteristiche fenotipiche e radiografiche con la JATD e la Sindrome di Ellis-Van Creveld (EVC, MIM 225500). In particolare, la Sindrome di Jeune è correlata alla SRP di tipo III: le due patologie sono accomunate dalle stesse caratteristiche radiografiche, compresa la polidattilia, tuttavia la Sindrome di Verma-Naumoff presenta un quadro clinico più grave, caratterizzato da insorgenza e letalità precoce in epoca prenatale e da malformazioni a carico di apparato gastrointestinale, urogenitale, cerebrale e cardiaco. La sindrome di Jeune è inoltre correlata sia fenotipicamente che geneticamente con la sindrome di Sensenbrenner (Cranioectodermal dysplasia; MIM 218330) (Bredrup C, 2011) e con la sindrome di Mainzer-Saldino (Conorenal syndrome; MIM 266920) (Perrault I, 2012).



**Fig.12** *Caratteristiche radiologiche della Sindrome di Jeune. Si possono notare in figura le alterazioni a carico delle coste (a), le anomalie alla pelvi (b,c) e le modificazioni alle falangi (d). Immagini tratte da Oberklaid F et al. (Archives of Disease in Childhood, 1977).*

## Genetica della Sindrome di Jeune

Le basi molecolari della Sindrome di Jeune sono state solo di recente parzialmente chiarite.

Nel 1995 l'individuazione di una delezione cromosomica in un paziente di 5 anni affetto da una osteocondrodisplasia, con caratteristiche radiografiche sovrapponibili a quelle della Sindrome di Jeune, suggerì che il locus malattia fosse localizzato nel cromosoma 12 (Nagai et al., 1995) mentre, nel 2003, in seguito ad un'analisi genome-wide effettuata su cinque famiglie consanguinee, si evidenziò una regione in linkage con la patologia nel cromosoma 15q13; l'analisi estesa ad altre cinque famiglie consanguinee consentì di definire una regione di omozigotità, condivisa da due probandi, contenente due geni, GREMLIN e FORMIN. La potenziale implicazione di GREMLIN e FORMIN nell'eziopatogenesi della Sindrome di Jeune venne però esclusa poiché, in seguito ad analisi molecolare, non si identificarono varianti patogenetiche nei due geni (Morgan et al., 2003).

Negli ultimi sei anni, sono state associate alla Sindrome di Jeune mutazioni in geni codificanti per proteine espresse a livello delle strutture ciliari e, nello specifico, implicate nei meccanismi di trasporto intraflagellare necessari per la genesi, l'omeostasi e la sopravvivenza di tali strutture, portando, di fatto, a considerare la Sindrome di Jeune parte della famiglia di patologie genetiche note come "ciliopatie".

Sebbene finora siano stati associati alla Sindrome di Jeune quattro geni candidati, molti casi rimangono non definiti, evidenziando una certa eterogeneità genetica per la patologia e suggerendo che altri geni ciliari possano essere coinvolti nell'insorgenza della malattia.

Il primo gene ciliare candidato per la Sindrome di Jeune, IFT80, venne individuato nel 2007 grazie ad uno studio genome-wide effettuato su 12 famiglie consanguinee e su 27 non consanguinee originarie del Pakistan e della Turchia (Beales et al., 2007). Il gene IFT80 mappa sul cromosoma 3, è costituito da 17 esoni e codifica per una proteina di 777 residui aminoacidici componente del complesso di trasporto intraflagellare di tipo B,

coinvolto nel trasporto anterogrado e quindi essenziale per la sintesi e il mantenimento sia delle cilia sensoriali che mobili.

Due lavori pubblicati nel 2009 sull'American Journal of Human Genetics individuarono come altro gene candidato per le SRPs il gene DYNC2H1; mutazioni nel gene sono state associate sia alla Sindrome di Jeune che alla SRPIII (Dagoneau et al., 2009; Merrill et al., 2009). Di recente, il gruppo di Schmidts, tramite uno studio di exome sequencing, ha evidenziato mutazioni in DYNC2H1 nel 41% degli affetti JATD con fenotipo caratterizzato esclusivamente da alterazioni di tipo scheletrico (Schmidts et al., 2013). Il gene DYNC2H1 mappa sul cromosoma 11q21-22 e codifica per la catena pesante della dineina, una proteina ciliare di 4307 aminoacidi componente del complesso citoplasmatico della dineina. Le dineine citoplasmatiche hanno un ruolo essenziale nei meccanismi di trasporto retrogrado che avvengono a livello del cilium. In particolare, DYNC2H1, in associazione con le catene leggera intermedia, intermedia e leggera, è componente del complesso della dineina citoplasmatica-2 (DYNC2) ed è direttamente coinvolta nella traslocazione del complesso lungo i microtubuli mediante il suo motor domain.

Effettuando uno screening sistematico di tutti i geni potenzialmente coinvolti nella biogenesi e nella funzionalità ciliari, nel 2011 Davis e colleghi hanno rilevato due mutazioni nel gene TTC21B presenti in una famiglia JATD e assenti in 796 cromosomi di controllo. Il gene TTC21B, che mappa nel cromosoma 2q24.3, codifica per THM1 (tetratricopeptide repeat-containing hedgehog modulator-1), una proteina di 1317 aminoacidi contenente diversi domini TPR (tetratricopeptide domains) che sembrano importanti per la funzionalità ciliare e che sono presenti in numerose proteine coinvolte in ciliopatie dell'uomo, come BBS4 e BBS8 che, mutati, causano il fenotipo Bardet-Biedl (MIM 209900). TTC21B è una proteina assonemale necessaria per il trasporto retrogrado intraflagellare. Mutazioni nell'ortologo murino sono responsabili del fenotipo *alien* che, nel topo, è letale e provoca anomalie anatomiche sovrapponibili con quelle riscontrate in diverse ciliopatie umane. Oltre ad essere causalmente associati alla Sindrome di Jeune, gli alleli mutati di TTC21B possono interagire in *trans* con alleli

responsabili di altre ciliopatie determinando un certa variabilità nella severità del fenotipo (Davis E. et al, 2011).

Sempre nel 2011, analizzando l'esoma di un campione di pazienti diagnosticati per differenti ciliopatie, il gruppo di Bredrup ha rilevato una mutazione missenso in omozigosi nel gene WDR19 in una famiglia JATD originaria della Germania e due mutazioni in eterozigosi composta in una famiglia norvegese affetta da sindrome di Sensebrenner, il cui fenotipo clinico è in parte sovrapponibile a quello Jeune (Bredrup et al., 2011). Il gruppo di Bredrup ha inoltre evidenziato come, nelle cilia dei fibroblasti, la mancata espressione di IFT144, codificata da WDR19 e componente del trasporto intraflagellare di tipo A, sia in grado di determinare una marcata alterazione della morfologia ciliare.



## SCOPO DEL PROGETTO

In questo progetto di ricerca la strategia dell'exome sequencing è stata utilizzata per identificare uno o più geni candidati per la Sindrome di Jeune (JATD, Asphyxiating Thoracic Dystrophy, MIM 208500) in tre individui affetti appartenenti a due famiglie di origine sarda non correlate tra loro in studio presso il nostro Dipartimento.

Negli anni passati questi pazienti sono stati sottoposti ad un insieme di indagini molecolari con l'obiettivo di chiarire le basi eziopatogenetiche della malattia. Le regioni 12p11-p12 e 15q13, individuate in passato come regioni candidate in linkage con la patologia (Nagai et al., 1995; Morgan et al., 2005), sono state analizzate nelle nostre famiglie attraverso l'impiego di marcatori molecolari. Nella regione 12p11-p12, individuata da Nagai come sede di una delezione cromosomica in un bambino con fenotipo Jeune, non stati evidenziati riarrangiamenti cromosomici. L'analisi della regione 15q13 ha invece rivelato nei fratelli affetti di una famiglia la condivisione allelica in una regione di 0.5Mb. Nella regione genomica individuata non era presente però alcun gene che per la sua funzione potesse candidarsi ad essere il gene causativo della sindrome.

Per quanto riguarda i geni ciliari che sono stati associati alla patologia negli ultimi anni, sono state ricercate le mutazioni descritte in letteratura per il gene IFT80 (Beales et al., 2007) e per il gene DYNC2H1 (Dagoneau et al., 2009; Merrill et al., 2009). Tali mutazioni non sono state però rilevate nei nostri pazienti.

Considerata l'eterogeneità genetica della patologia e la possibilità che, oltre ai geni già descritti, possano essere implicati nel fenotipo Jeune geni ancora non identificati, l'analisi dell'esoma si proponeva come una strategia appropriata per la ricerca di alleli patogenetici. L'approccio del sequenziamento dell'intero nucleo familiare, rappresentato da due fratelli affetti e dai rispettivi genitori nel caso di una famiglia e dal trio probando-genitori nel caso dell'altra, è stato scelto in relazione alla modalità di ereditarietà autosomica-recessiva della malattia. L'analisi dell'esoma degli affetti di due famiglie non correlate ma originarie di una popolazione geneticamente isolata come

quella sarda avrebbe inoltre agevolato l'individuazione di alleli mutati in un unico gene presumibilmente patogenetico per la malattia in entrambe le famiglie.

La comprensione dei meccanismi patogenetici alla base della Sindrome di Jeune ha come fine ultimo poter predisporre interventi di tipo terapeutico e adeguate misure di prevenzione della malattia attraverso l'identificazione dei portatori sani, delle coppie a rischio e la diagnosi prenatale.

## PAZIENTI E METODI

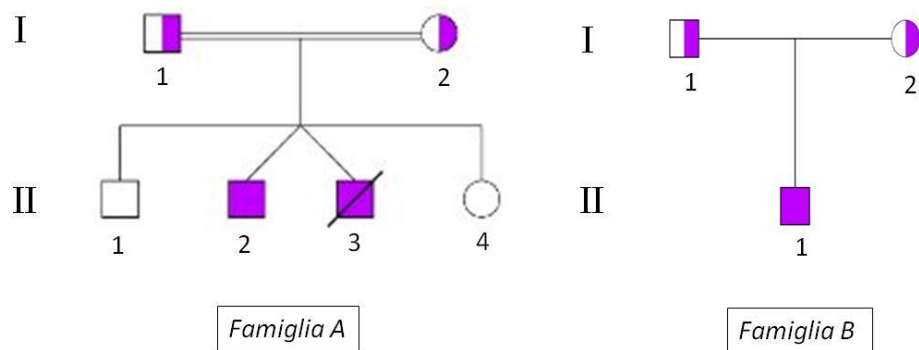
### Soggetti coinvolti nello studio

Sono state reclutate per questo studio due famiglie non correlate di origine sarda i cui alberi genealogici sono riportati nella figura 18.

Alla famiglia A appartengono due gemelli dizigoti (II2; II3) affetti da Sindrome di Jeune (JATD) nati da genitori con una lontana consanguineità (coefficiente di inbreeding  $F=1/128$ ). I soggetti, nati pretermine nel 2008, hanno presentato sin dalla nascita un fenotipo grave, caratterizzato da malformazioni scheletriche comprendenti una grave costrizione della gabbia toracica, polidattilia postassiale delle mani, riduzione delle epifisi e delle metafisi delle ossa lunghe, associate a retinopatia e policistosi renale. Uno dei due affetti (II3) è deceduto dopo sei mesi di vita per una grave insufficienza respiratoria.

Il probando JATD della famiglia B (II1), nato a termine nel 2009, presenta un fenotipo grave caratterizzato da alterazioni esclusivamente di tipo scheletrico che comprendono osteopenia diffusa, svasamento metafisario omerale, glenoideo e costale, ipoplasia e calcificazioni dei nuclei epifisari, scoliosi dorsale con platispondilia e aspetto dismorfico di numerose vertebre. Il torace ristretto predispone il paziente a infezioni ricorrenti respiratorie con episodi di insufficienza respiratoria che hanno richiesto il ricovero e la ventilazione meccanica.

Lo studio dell'esoma è stato eseguito in entrambe le famiglie nei probandi e nei genitori previo consenso informato per l'accettazione delle indagini molecolari (Fig.18).



**Fig.18 Alberi genealogici delle due famiglie JATD analizzate.** Famiglia A: è stato sottoposto a sequenziamento e analisi dell'esoma il DNA dei gemelli dizigoti affetti (II2; II3) e dei genitori (I1; I2). Famiglia B: è stato sottoposto a sequenziamento e analisi dell'esoma il DNA del trio costituito da affetto (II1) e genitori (I1; I2).

## Analisi dell'esoma

Il workflow di analisi si è articolato nelle seguenti fasi:

1. Preparazione dei campioni e cattura dell'esoma
2. Sequenziamento con la piattaforma Illumina
3. Analisi bioinformatica dei dati

### 1. Preparazione dei campioni e cattura dell'esoma

Il DNA genomico è stato estratto da 5mL di sangue periferico secondo la procedura "salting out" e sospeso in buffer Tris-EDTA (Tris 10mM+EDTA 1mM) pH 8 (Miller, S.A., Dykes, D.D. & Polensky, H.F. (1998) A simple salting out procedure for extracting DNA from human nucleated cells. *Nucleic Acids Research*, 16, 1215).

Le librerie di DNA sono state preparate presso la ditta Centrillion Biosciences Inc (Palo Alto, CA, USA) da 10ug di DNA genomico, secondo il protocollo previsto dal kit SureSelect Human All Exon V4 (Agilent Technologies) per la cattura dell'esoma.

Il targeting dell'esoma ha permesso di ottenere una copertura media di 51Mb per ciascun campione. L'amplificazione delle librerie è stata effettuata secondo quanto previsto dal workflow specifico per il sequenziamento con piattaforma HiSeq2000 (Illumina).

## 2. Sequenziamento con piattaforma HiSeq2000 (Illumina)

Le librerie di ciascun campione sono state sequenziate paired-end con la piattaforma HiSeq2000 dalla ditta Centrillion Biosciences Inc (Palo Alto, CA, USA).

Il coverage medio è risultato pari a 30X e la lunghezza media delle reads è stata di 75bp.

Il numero di reads e di basi sequenziate per ciascun campione sono riassunte in tabella 1.

Campione	Reads (M)	Basi (G)
I 1	42.48	6.37
I 2	38.67	5.8
II 2	52.71	7.91
II 3	48.67	7.34

*Famiglia A*

Campione	Reads (M)	Basi (G)
I 1	47.65	7.15
I 2	46.18	6.93
II 1	44.22	6.63

*Famiglia B*

**Tabella 1.** Numero medio di reads e basi sequenziate per ciascun campione nelle famiglie A e B.

### 3. *Analisi bioinformatica*

I programmi utilizzati per l'analisi dei dati sono implementati nella piattaforma di web analisi Galaxy (<http://galaxyproject.org/>) disponibile presso la Sezione di Biostatistica e Bioinformatica del nostro Dipartimento.

**FASTQC** (<http://www.bioinformatics.babraham.ac.uk/projects/fastqc/>)

Il programma esegue un'analisi qualitativa dei dati grezzi di sequenziamento. L'analisi è stata condotta sui file in formato fastq (Cock et al., 2010; Deorowicz et al., 2011) prodotti per ciascun campione. Il tool produce un output in formato HTML che contiene i risultati relativi a quality scores, contenuto in GC, distribuzione di lunghezza, livello di duplicazione delle reads, sequenze over-rappresentate.

**BWA** (<http://bio-bwa.sourceforge.net/>)

Il programma di allineamento BWA permette di allineare sequenze relativamente corte ("short reads") con una sequenza di riferimento estesa come quella del genoma umano.

La sequenza di riferimento utilizzata per il processo di allineamento è la versione ucsc\_GRCh37 (hg19) del genoma umano. L'output ottenuto è in formato SAM (Li et al., 2009). Per l'analisi sono stati utilizzati i parametri di default del software (<http://bio-bwa.sourceforge.net/>).

**SAMTOOLS** (<http://samtools.sourceforge.net>)

Il toolkit SAMTOOLS, che implementa una serie di utilities per le analisi post-allineamento (indicizzazione, chiamata delle varianti, viewer di allineamento), è stato utilizzato per convertire il file output di BWA, in formato SAM, nel formato binario BAM, che contiene tutte le informazioni del SAM in forma compressa.

**PICARD** (<http://picard.sourceforge.net>)

Il programma Picard utilizza linee di comando basate su Java che permettono di manipolare i file di tipo SAM. L'applicazione "mark duplicate reads" del programma PICARD è stata utilizzata per contrassegnare e rimuovere le reads duplicate (duplicati ottici e artefatti di PCR) presenti nel file BAM .

**GATK** (<http://www.broadinstitute.org/gatk/>)

Il tool Unified Genotyper incluso nel programma GATK ha permesso di generare i file in formato VCF (Variant Caller Format, v4.1; Danecek et al., 2011) a partire dai file BAM di

input dei campioni sequenziati. Sono state incluse nel file vcf solo le varianti che hanno superato i filtri GATK standard per qualità, coverage, lunghezza degli omopolimeri contigui (HRun>5.0, QD<2.0, MQ <40.0, FS > 60.0, HaplotypeScore >13.0, MQRankSum<12.5, ReadPosRankSum<-8.0). Per ogni individuo sono stati generati due file vcf che riportano le informazioni relative alle variazioni di singoli nucleotidi (SNPs) e quelle relative alle inserzioni/delezioni (In/Dels).

### **Filtraggi applicati per l'individuazione di varianti candidate**

Al fine di individuare una lista di geni candidati per la patologia, le varianti sono state analizzate e filtrate con wAnnovar (<http://wannovar.usc.edu/>), il web-server del software di analisi Annovar (<http://www.openbioinformatics.org/annovar/>).

wAnnovar annota le conseguenze funzionali di ciascuna variante riportata nel file vcf di input (fig.19).

Le varianti sono state filtrate impostando i seguenti parametri di analisi:

- Coverage  $\geq$  20X
- Quality score  $\geq$  30
- Minor Allele Frequency (MAF) in 1000Genomi ([www.1000genomes.org](http://www.1000genomes.org)) e NHLBI-ESP 6500 (NHLBI Exome Sequencing Project [evs.gs.washington.edu/EVS/](http://evs.gs.washington.edu/EVS/))  $\leq$  0.01
- Modalità di ereditarietà autosomica recessiva
- Sostituzioni missenso, nonsense, frameshift, nei siti di splicing non riportate in dbSNP135 o associate in dbSNP135 a fenotipi clinici.

Type	Column	Description
Gene annotation	Variation function	Exonic, intronic, intergenic, UTR, etc
	Gene	Impacted gene or neighbouring gene (with distance)
	Exonic variation function	Non-synonymous, synonymous, stopgain, etc
Variation databases	AChange	mRNA and amino acid change for coding variants
	ESP5400_ALL	Allele frequency in 5400 NHLBI-ESP exomes
	1000G_ALL	Allele frequency in 1000 Genomes Project (currently, version 2012 Feb)
Functional prediction	dbSNP	dbSNP identifier (currently, version 135)
	AVSIFT	Base-level SIFT scores
	LJB_SIFT	1-SIFT scores and predictions (D: damaging, T: tolerated)
	LJB_PolyPhen2	PolyPhen 2 scores and predictions (D: probably damaging; P: possibly damaging; B: benign)
	LJB_LRT	LRT scores and predictions (D: deleterious; N: neutral; U: unknown)
	LJB_MutationTaster	MutationTaster scores and predictions (A: disease_causing_automatic; D: disease_causing; N: polymorphism; P: polymorphism_automatic)
	LJB_PhyloP	PhyloP conservation scores and predictions (C: conserved, N: non-conserved)
Region annotation	GERP++	GERP++ scores for exonic variants
	Conserved	Region-level phastCons LOD scores
	SegDup	Located in segmental duplication region and the sequence identity score

**Fig.19** Annotazioni funzionali del software wANNOVAR (Chang et al., 2012; Wang et al., 2010)

### Predizione della patogenicità delle varianti

La patogenicità delle varianti è stata valutata con il software Polyphen-2 (Polymorphism Phenotyping v2).

Polyphen è un programma che predice il possibile impatto funzionale di una sostituzione aminoacidica missenso sulla struttura e la funzione di una proteina umana.

Il software usa informazioni basate sull'allineamento con otto sequenze distinte e sulla struttura della proteina, confrontando le proprietà specifiche dell'allele wild-type con quelle dell'allele mutato. L'importanza funzionale di un allele viene predetta mediante un classificatore bayesiano (Adzhubei et al., 2010).

### Annotazioni funzionali geni candidati

I geni candidati sono stati valutati dal punto di vista funzionale con il software gratuito DAVID Bioinformatic Database v6.7 (The Database for Annotation, Visualization and Integrated Discovery; <http://david.abcc.ncifcrf.gov/tools.jsp>).



Il software DAVID consiste di un insieme di tools analitici che permettono di estrarre da una lista di geni/proteine informazioni inerenti il loro specifico significato biologico. Il programma effettua quindi, per ciascun gene individuato, una classificazione di tipo funzionale (Huang et al., 2009).

DAVID consente inoltre di visualizzare i geni nelle mappe di pathway BioCarta ([www.biocarta.com/genes/index.asp](http://www.biocarta.com/genes/index.asp)) e KEGG ([www.genome.jp/kegg/pathway.html](http://www.genome.jp/kegg/pathway.html)), di verificare la presenza di eventuali proteine interagenti e connessioni gene-malattia.

### **Prioritizzazione geni candidati**

I geni individuati sono stati prioritizzati con il software ToppGene utilizzando l'applicazione Candidate Gene Prioritization (Chen et al., 2009).

Questa applicazione effettua la prioritizzazione dei geni inseriti in una lista "test" basandosi sulla similarità funzionale con una lista di geni "training".

La prioritizzazione impostata sulle annotazioni funzionali computa la similarità tra due geni in base alle annotazioni semantiche gene-specifiche, tramite il calcolo del coefficiente di correlazione di Pearson. Gli scores di similarità per ciascuna proprietà calcolata sono combinati in uno score totale (overall score) tramite meta-analisi statistica; ad ogni annotazione viene associato un P-value derivato dalla probabilità che, all'interno del genoma, ci siano altri geni con analoghe annotazioni funzionali.

Nell'analisi da noi condotta, la training list includeva i geni già associati in passato alla Sindrome di Jeune (IFT80, DYNC2H1, TTC21B, WDR19), mentre la test list era costituita dai geni risultanti dai filtraggi dei dati di sequenziamento esomico. Per il calcolo degli scores di similarità sono stati considerati la funzione molecolare, eventuali implicazioni in fenotipi umani e murini, domini proteici, pathways funzionali associati, interazioni proteina-proteina, associazioni con la letteratura PubMed.

### **Risequenziamento varianti candidate**

Le regioni geniche contenenti le varianti candidate rilevate con il sequenziamento dell'esoma sono state amplificate e risequenziate con la chimica BigDye Terminator (Life Technologies) nel sequenziatore ad elettroforesi capillare 3130XL Genetic Analyzer (Life Technologies).

I primers e le condizioni di amplificazione utilizzati per risequenziare l'esone 25 del gene DYNC2H1 sono riportati nel lavoro di Dagoneau et al., 2009. I primers utilizzati per l'amplificazione dell'esone 5 del gene MAP1S sono stati disegnati con il software Primer3web version 4.0.0 (<http://primer3.wi.mit.edu/>, Steve Rozen and Helen J. Skaletsky, 2000).

## RISULTATI

### Sequenziamento dell'esoma nella famiglia A

Il sequenziamento dell'esoma, effettuato per i gemelli affetti e per i genitori, ha consentito di individuare più di 150000 variazioni di singoli nucleotidi e più di 10000 variazioni di inserzione/delezione per campione (Tabella 2).

#### Famiglia A

Campione	SNPs totali	In/Dels totali
I 1	156275	10413
I 2	152498	10228
II 2	153280	11019
II 3	200429	11028

#### Famiglia B

Campione	SNPs totali	In/Dels totali
I 1	176768	12078
I 2	160908	10957
II 1	146394	9924

**Tabella 2.** Numero di varianti totali rilevate nel sequenziamento dell'esoma della famiglia A e della famiglia B.

Tra queste, sono state considerate solo le varianti che avessero superato i principali filtri di tipo qualitativo ( $Q \geq 30$ ) e quelle per cui potessero essere analizzate almeno venti reads di sequenza distinte e sovrapponibili (coverage  $\geq 20$ ).

Con l'applicazione di filtri discreti e consecutivi, sono state selezionate le varianti rare, non riportate nei database pubblici (dbSNP135, 1000 Genomi e NHLBI-ESP 6500 esomi) o riportate nei medesimi con una frequenza dell'allele minore  $\leq 0.01$ .

In seguito all'applicazione dei modelli di filtraggio per modalità di ereditarietà e per tipologia di variante (missenso, nonsense, siti di splicing), sono stati selezionati 30 geni nel probando II2 e 35 nel probando II3, aventi almeno due alleli potenzialmente deleteri. Di questi, 11 geni presentavano varianti di singoli nucleotidi (SNV) comuni ad entrambi gli affetti (tabella 2). Gli stessi filtraggi, applicati per l'analisi delle In/Del, hanno consentito di ottenere una lista finale di 7 geni aventi almeno due alleli mutati in comune nei due affetti (tabella 3).

L'analisi bioinformatica dei dati di sequenziamento esomico ha determinato l'individuazione di una variante nel gene DYNC2H1 (MIM 603297), c.3694G>A, presente in omozigosi negli affetti e in eterozigosi nei genitori, portatori obbligati della mutazione. La variante è una mutazione missenso che causa la sostituzione di un residuo di acido aspartico con un residuo di asparagina (p.D1232N) nella catena pesante della dineina citoplasmatica 2. Il sequenziamento con la metodica di Sanger ha confermato la presenza della variazione in eterozigosi nei genitori e in omozigosi nei probandi, mentre i fratelli non affetti sono risultati normali per la variante ricercata.

Il coinvolgimento nell'eziopatogenesi della Sindrome di Jeune degli altri geni individuati con l'analisi bioinformatica (tabelle 3 e 4) è stato escluso per mancanza di evidenze scientifiche sul loro ruolo in pattern relativi ai meccanismi di trasporto e/o segnalazione ciliare.

Gene	Numero alleli deleteri II 2	Numero alleli deleteri II 3
AGAP3	6	6
DOCK1	2	2
DYNC2H1	2	2
GYG2	2	2
KCNC3	2	2
MAFIP	2	2
OR8U1	2	2
SPATC1	3	2
TAS2R31	3	11
TAS2R46	5	2
ZNF717	14	11

**Tabella 3.** Lista dei geni con variazioni di singoli nucleotidi (SNV) in comune ai due probandi della famiglia A (II 2; II 3), identificati con l'exome sequencing in seguito all'applicazione dei filtri per tipologia di variante (nonsense, missenso, siti di splicing canonici), frequenza globale dell'allele minore ( $MAF \leq 0.01$ ) e modello di ereditarietà autosomica recessiva.

Gene	Numero alleli deleteri II2	Numero alleli deleteri II3
ALOX5AP	2	2
EI24	2	2
HLA	2	2
NPEPPS	2	2
SLC38A3	2	2
SON	2	2
TPM2	2	2

**Tabella 4.** Lista dei geni con variazioni di inserzione/delezione (In/Del) in comune ai due probandi della famiglia A (II2; II3), identificati con l'exome sequencing in seguito all'applicazione dei filtri per tipologia di variante (frameshift), frequenza globale dell'allele minore ( $MAF \leq 0.01$ ) e modello di ereditarietà autosomica recessiva.

### Sequenziamento dell'esoma nella famiglia B

L'analisi dell'esoma dei componenti della famiglia B ci ha permesso di identificare in media 161000 sostituzioni di singoli nucleotidi (SNV) e 10986 variazioni di inserzione/delezione (In/Del) (Tabella 2).

Le varianti sono state analizzate utilizzando gli stessi tipi di filtraggio applicati per la famiglia A, selezionando quindi le variazioni di singoli nucleotidi (SNV) che danno origine a sostituzioni aminoacidiche missenso, non senso o che ricadono nei siti canonici di splicing e le varianti In/Dels che causano mutazioni frameshift (tabelle 5 e 6).

Nonostante le caratteristiche di isolato genetico della Sardegna e la rarità della patologia facessero presupporre che il gene *DYNC2H1* fosse mutato anche nel probando della famiglia B, l'exome sequencing non ha portato all'individuazione di varianti patogenetiche all'interno di questo gene.

Nella lista finale di geni risultante dall'analisi bioinformatica è presente solo un gene, MAP1S (MIM 607573), che, per la funzione e per il pattern di espressione, poteva essere considerato un valido candidato nell'eziopatogenesi della malattia.

Utilizzando il software ToppGene, il gene MAP1S è stato prioritizzato nella rank1 ottenendo, nel confronto per similarità funzionale con i geni della training list, l'overall score più elevato rispetto agli altri geni della test list.

MAP1S codifica per una proteina implicata nell'ancoraggio del centro di organizzazione dei microtubuli ai centrosomi. L'overespressione dell'ortologo MAP8 determina, nel modello murino, tossicità cellulare presumibilmente riconducibile ad un'alterazione nell'organizzazione della struttura o della dinamica dei microtubuli che interferisce con i movimenti delle proteine di trasporto. In particolare, l'accumulo di MAP8 determina l'intrappolamento delle dineine in aggregati insolubili e altera la funzionalità dei meccanismi di trasporto assonale (Ding J et al., 2006).

L'eventuale implicazione di MAP1S è stata esclusa in quanto il sequenziamento esomico ha evidenziato la presenza nel probando di due mutazioni in eterozigosi che non sono state confermate tramite risequenziamento diretto.

Gene	Numero alleli deleteri II 1	Gene	Numero alleli deleteri II 1
ALK	2	MYH3	2
AQP7	2	NOTCH1	2
BRD8	2	OBSCN	2
CCNK	2	PCDHA9	2
CDR71	2	POLR3B	2
CPSF3	2	SAFB2	2
DSPP	2	SBF1	2
FAM186A	2	SELPLG	2
KCNC3	2	SPEN	2
LDB3	2	TCHH	2
MAFIP	2	TSHZ1	2
MAP1S	2	USP6	2
MBOAT7	2		

**Tabella 5.** Lista dei geni con variazioni di singoli nucleotidi (SNV) identificati con l'exome sequencing nel probando della famiglia B (II1) in seguito all'applicazione dei filtraggi per tipologia di variante (nonsense, missenso, siti di splicing canonici), frequenza globale dell'allele minore ( $MAF \leq 0.01$ ) e modello di ereditarietà autosomica recessiva.



Gene	Numero alleli deleteri II 1
EI24	2
ACOT4	2
ALOX5AP	2
PLCL2	4
PRIM2	3
ZNF492	2
C8ORF38	2
KIAA0040	2
SLC38A3	2
ZNF720	2
TPM2	2

**Tabella 6.** Lista dei geni con variazioni di inserzione/delezione (In/Dels) identificati con l'exome sequencing nel probando della famiglia B (II 1) in seguito all'applicazione dei filtri per tipologia di variante (frameshift), frequenza globale dell'allele minore ( $MAF \leq 0.01$ ) e modello di ereditarietà autosomica recessiva.

## DISCUSSIONE

L'analisi dell'esoma nella famiglia A ha portato all'individuazione di una mutazione (p.D1232N) nel gene DYNC2H1, presente in omozigosi nei gemelli affetti.

Studi precedenti avevano già identificato DYNC2H1 come gene candidato per la Sindrome di Jeune e per la SRP di tipo III (Short rib-polydactyly Syndrome, type III) o Sindrome di Verma-Naumoff (MIM 263510) (Dagoneau N, 2009; Merrill AE, 2009). Mutazioni nel gene DYNC2H1 sono state associate anche alla SRP di tipo II (Short rib-polydactyly Syndrome, type II) o Sindrome di Majewski (MIM 263520) (El Hokayem J, 2012).

Un recentissimo articolo pubblicato sul Journal of Medical Genetics riconosce DYNC2H1 come primo gene candidato per la Sindrome di Jeune nella popolazione Nord Europea (Schmidts M, 2013).

Il gene DYNC2H1 mappa nel cromosoma 11, è costituito da 90 esoni e codifica per una proteina di 4314 aminoacidi che rappresenta la subunità più grande della dineina-2.

Nei vertebrati sono espresse circa quindici forme di dineine. La maggior parte di esse costituisce la classe delle dineine "assonemali", responsabili dei movimenti di cilia e flagelli. Solo due forme costituiscono invece la classe delle dineine citoplasmatiche, implicate nei movimenti di trasporto lungo i microtubuli all'interno delle strutture ciliari.

Una di queste, descritta per la prima volta da Paschal nel 1987 come proteina associata ai microtubuli 1C (MAP1C), viene anche chiamata dineina citoplasmatica 1 ed è largamente espressa nelle cellule, dove ha molteplici ruoli che includono il trasporto di organelli di membrana, l'orientamento del fuso mitotico, la migrazione nucleare e cellulare.

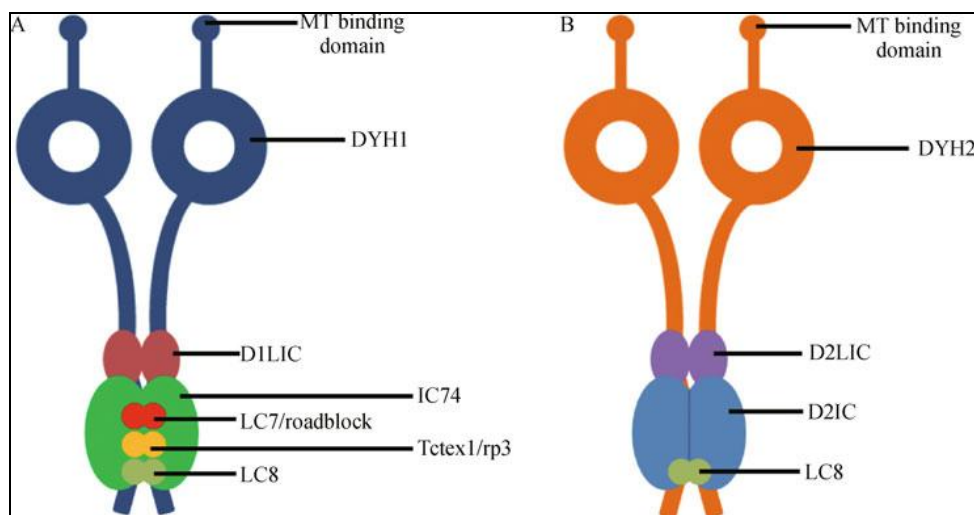
La dineina-1b, o dineina-2, ha invece una più ridotta espressione e funzioni limitate. Nei tessuti è espressa quasi esclusivamente nelle cellule ciliate sia mobili che immobili.

Tutte le dineine sono olocomplessi enzimatici che si formano in seguito alla dimerizzazione della catena pesante (HC, heavy chain) e alla sua associazione con subunità accessorie rappresentate dalle catene leggere, intermedie e leggere-intermedie (LCs, light chains, ICs, intermediate chains, LICs, light-intermediate chains),

alcune delle quali comuni sia alle dineine citoplasmatiche che a quelle assonemali, altre specifiche per le diverse isoforme.

La dineina-1a è un omodimero formato da due catene pesanti identiche (DYH1) a cui si associano come proteine accessorie la catena intermedia (IC74) e la catena leggera intermedia (D1LIC) (King et al, 2002; Pfister et al., 2006). La dineina 1 contiene inoltre tre catene leggere (LC7/roadblock, LC8, Tctex1/rp3) (Rajagopalan et al, 2012) (Fig.20A)

Nel caso della dineina-2, sebbene i primi studi di sedimentazione e immunoprecipitazione avessero consentito di identificare, dapprima in cellule di mammifero e poi in *Chlamydomonas* e nel ratto, solo la catena leggera intermedia (D2LIC) (Grissom et al, 2002; Perrone et al., 2003; Hou et al., 2004; Mikami et al.,2002) sembra ormai definitivamente stabilita la presenza di una catena intermedia, identificata in estratti di *Chlamydomonas* e chiamata FAP133 o D2IC, e di una catena leggera, LC8 (Rompolas et al., 2007) (Fig.20B)



**Fig.20** *Struttura dei complessi citoplasmatici della dineina-1 (A) e della dineina-2 (B).* Entrambi i complessi sono il risultato dell'assemblaggio del dimero di catene pesanti (DYH1; DYH2) con catene leggere intermedie (D1LIC; D2LIC), catene intermedie (IC74; D2IC) e catene leggere (LC7/roadblock, LC8, Tctex1/rp3; LC8) (Rajagopalan V, 2012; Cytoplasmic dynein-2: from molecules to human diseases).

### *Struttura della catena pesante della dineina-2*

La catena pesante della dineina-2 è costituita da un insieme di domini funzionali, schematizzati in figura (Fig.21).

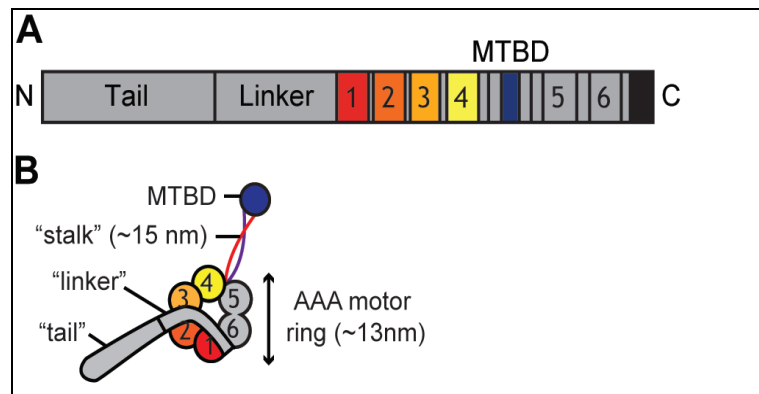
Il dominio "tail" occupa la parte N-terminale della proteina corrispondente a circa 1300 aminoacidi. Questo dominio contiene i siti necessari per la dimerizzazione della catena pesante, per l'interazione con le subunità accessorie del complesso e il sito di legame per il cargo da trasportare sulla superficie dei microtubuli.

Il tail domain può essere scomposto in due subdomini, di cui uno, quello maggiormente N-terminale, è chiamato DHC\_N1, l'altro, DHC\_N2, si colloca in una posizione intermedia tra il DHC\_N1 e il primo dei domini ATPasici che costituiscono il "motor" domain.

Il tail domain è connesso, tramite un dominio linker di recente identificazione, al motor domain, che è il dominio maggiormente conservato in tutte le catene pesanti delle dineine, dalle alghe ai vertebrati e dalle forme citoplasmatiche a quelle assomali. Questo dominio occupa larga parte della catena pesante, corrisponde infatti a circa 350 kDa della parte C-terminale ed è costituito da più di 3000 residui aminoacidici.

Studi di microscopia elettronica hanno evidenziato la presenza, nel motor domain, di sei domini AAA+ ATPasici connessi tra loro in un'unica catena polipeptidica e organizzati in una struttura ad anello esamerico. I domini ATPasici sono responsabili del legame e dell'idrolisi dell'ATP e/o dell'ADP. Mentre dal legame dell'ATP dipende il movimento lungo i microtubuli, in quanto dall'idrolisi dell'ATP si genera la forza motrice, il legame dell'ADP sarebbe necessario, in alcune isoforme, per regolare la velocità del complesso lungo i microtubuli.

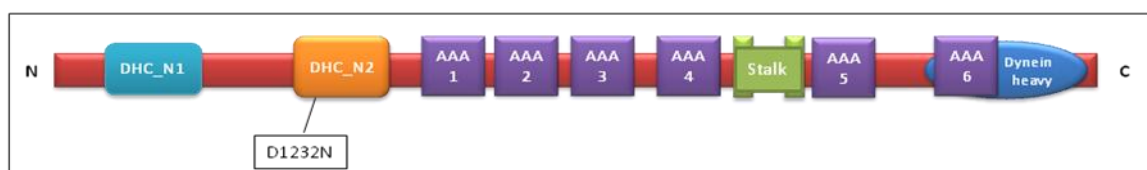
Fra il quarto e il quinto modulo AAA è localizzato il dominio "stalk", un dominio coiled-coil antiparallelo al cui apice è presente un piccolo dominio globulare di legame ai microtubuli ATP-dipendente. La funzione del dominio stalk non è stata ancora definitivamente chiarita, ma sembrerebbe che esso sia in grado, attraverso modificazioni dinamiche nelle interazioni elica-elica, di trasmettere cambiamenti di conformazione lungo il complesso enzimatico durante l'attività di movimento.



**Fig.21 A: Domini funzionali della catena pesante della dineina-2. B: Rappresentazione tridimensionale dei domini funzionali (Reck-Peterson SL, Handbook of Dynein).**

#### *Ruolo eziopatogenetico della variante c.3694G>A*

La variante individuata con l'exome sequencing nella famiglia A è una mutazione missenso che causa la sostituzione di un residuo di acido aspartico con un residuo di asparagina (p. D1232N) in corrispondenza del dominio DHC\_N2 del tail domain della catena pesante della dineina-2 (fig.22).



**Fig.22 Domini funzionali della catena pesante della dineina-2 e localizzazione della mutazione individuata.**

Il ruolo patogenetico della mutazione individuata dovrà essere chiarito mediante studi funzionali.

In generale, mutazioni a carico della catena pesante della dineina determinano difetti nei meccanismi di trasporto intraflagellare retrogrado per perdita della funzionalità di DYNC2H1 e accumulo delle proteine di trasporto anterogrado all'apice del cilium.

Data la localizzazione a livello del tail domain, è plausibile che la sostituzione dell'acido aspartico (D) in posizione 1232 con l'asparagina (N) causi un'alterazione strutturale e quindi la riduzione o la totale perdita della capacità di interazione della catena pesante con le catene accessorie del complesso o l'abolizione del processo di dimerizzazione delle catene pesanti.

Recenti lavori sembrano evidenziare per il tail domain anche una funzione di regolazione dell'attività del motor domain.

Mutazioni nel tail domain della dineina citoplasmatica causano infatti nel modello murino *Loa* (Leggs at odd angles) degenerazione neuronale per difetti nel meccanismo di trasporto retrogrado assonale (Ori-McKenney KM, 2010). Il modello murino *Loa* è portatore di una mutazione (p.F580Y) che ricade nella porzione N-terminale della catena pesante dell'orologio murino *Dync2h1*. Il meccanismo molecolare mediante cui questa mutazione causa una disfunzione della dineina citoplasmatica deve essere ancora chiarito ma lo studio pubblicato nel 2010 da McKenney et al. ipotizza che, nonostante la localizzazione della mutazione *Loa* a livello di tail domain, ci possa essere una diretta comunicazione tra tail e motor domain che si riflette in una riduzione generale della stabilità della proteina nel mutante *Loa*. Nel topo *Loa* la mutazione p.F580Y causerebbe un difetto di coordinazione tra i due domini modificando la loro posizione relativa e la loro capacità di interagire e, in definitiva, un'alterazione nell'affinità tra dineina e microtubuli.

Sebbene il meccanismo specifico mediante cui la catena pesante della dineina è in grado di determinare l'insorgenza del fenotipo caratteristico della Sindrome di Jeune debba essere ancora in parte chiarito, sembra che mutazioni nel gene *DYNC2H1* risultino principalmente in malformazioni scheletriche dovute ad un anomalo sviluppo delle strutture ossee durante l'embriogenesi, come evidenziato recentemente dal lavoro di Schmidts e colleghi con il quale, in seguito ad exome sequencing, sono state rilevate 34 mutazioni nel gene in 29/71 pazienti JATD di origine Nord-Europea il cui fenotipo clinico era caratterizzato da riduzione della gabbia toracica e brachidattilia, senza polidattilia e alterazioni di tipo extra-scheletrico (retinopatia o compromissione renale) (Schmidt, 2013).

Il fenotipo clinico dei probandi della famiglia A è caratterizzato, invece, non solo da una severa riduzione della gabbia toracica associata ad altre anomalie di origine scheletrica, inclusa la polidattilia delle mani, ma anche da alterazioni a carico della retina e policistosi renale, suggerendo la possibilità che alleli modificatori o fattori epigenetici possano avere un ruolo nella determinazione della severità del fenotipo nei pazienti.

Dal momento che, per la formazione delle cilia primarie, sono necessari numerosi geni, è possibile quindi che l'espressività variabile correlata alle mutazioni nel gene *DYNC2H1* possa essere dovuta all'eterozigotità per mutazioni in altri geni ciliari (Ocbina R, 2011).

Per poter investigare sulla presenza di alleli che possano in qualche modo influenzare la severità del fenotipo sono state ricercate, nei nostri pazienti, varianti in geni noti per avere effetti modificatori sul fenotipo delle ciliopatie, tra cui *TTC21B*, *RPGRIP1L*, *AHI1* e *CEP290*.

Come evidenziato dal lavoro del 2011 di Davis et al., la presenza di due alleli mutati nel gene *TTC21B* è in grado di causare il fenotipo Jeune. Mutazioni in eterozigosi sono state però rilevate associate in *trans* con altre mutazioni in geni responsabili di ciliopatie, in particolare la mutazione p.P209L, in eterozigosi nel gene *TTC21B*, sarebbe responsabile dell'insorgenza di cisti renali in un paziente Bardet-Biedl, in cui la cistogenesi è un fenomeno con prevalenza inferiore al 24%. (Davis et al., 2011).

I pazienti affetti esaminati nel nostro studio presentano cinque varianti nel gene *TTC21B*, di cui due danno origine a sostituzioni aminoacidiche sinonime, mentre le altre causano sostituzioni non sinonime. Queste varianti sono annotate nei database come polimorfismi non patogenetici.

È stato osservato che mutazioni in *RPGRIP1L* (MIM 610937), responsabili della Sindrome di Meckel (MIM 611561) e della Sindrome di Joubert (MIM 611560), sono in grado di modificare la penetranza della degenerazione retinica in ciliopatie causate da mutazioni in altri loci, mentre la presenza di un solo allele variante nel gene *AHI1* (MIM 608894) può modificare il grado di disfunzione a carico della retina in individui con nefronoftisi (*NPHP1*, MIM 256100) (Davis et al., 2011; Khanna H, 2009; Louie CM, 2010).

Nessuno dei due pazienti JATD presenta varianti nel gene RPGRIP1L, mentre solo uno dei due ha una variante in eterozigosi nel gene AHI1 che dà luogo ad una sostituzione aminoacidica sinonima (p.S1005S), riportata in dbSNP con MAF = 0.0133 (rs41287054).

Sono stati ricercati alleli varianti e potenzialmente patogenetici anche nel gene CEP290, che codifica per una proteina centrosomica e le cui mutazioni sono associate ad un range di ciliopatie, accomunate da degenerazione retinica, che comprendono la sindrome di Joubert di tipo 5 (BTS5, MIM:610188), la sindrome di Senior-Loken di tipo 6 (SLSN6, MIM:610189), la sindrome di Meckel di tipo 4 (MKS4, MIM:611134) e la sindrome di Bardet-Biedl di tipo 14 (BBS14, MIM:209900). Nei nostri pazienti JATD è presente solo un polimorfismo (rs2468255), in comune ai due affetti, che dà luogo ad una sostituzione di tipo sinonimo (p.S756S).

Analizzando tutti gli alleli che determinano sostituzioni non sinonime non annotati nei database di polimorfismi, è emerso che entrambi gli affetti presentano due varianti in eterozigosi, di cui una nel gene PTCH1 (p.E1280K) e l'altra nel gene BBS4 (p.C264F), codificanti entrambi per proteine espresse a livello ciliare e connesse con le funzioni di trasporto regolate dai complessi della dineina.

Il gene PTCH1 codifica per il recettore hedgehog Patched1 che, legandosi alla molecola Shh, determina l'attivazione della cascata di segnalazione Sonic hedgehog che risulta nell'attivazione della trascrizione di geni target regolati da Gli. È stata evidenziata una connessione diretta tra DYNC2H1 e il pathway di segnalazione Sonic hedgehog; in particolare, gli studi effettuati con doppi mutanti murini *Dync2h1* dimostrano che il complesso della dineina è indispensabile per una corretta segnalazione a valle di Ptch1 e a monte dei fattori di trascrizione Gli (Obkina R., 2011). La polidattilia osservata in molte ciliopatie sarebbe inoltre proprio da ricondurre a difetti nei sistemi di segnalazione hedgehog che, durante l'embriogenesi, regolano la morfogenesi e i meccanismi di differenziazione e crescita di tessuti e organi (Cardenas-Rodriguez M, Badano JL, 2009).

Il gene BBS4 codifica per una delle proteine appartenenti alla famiglia delle proteine BBS (BBS1-8). Molte di queste proteine sono localizzate nei corpi basali, nell'assonema ciliare e nelle regioni cellulari pericentriolari. Le proteine BBS sono coinvolte nel traffico intracellulare ciliare e mutazioni nei geni BBS sono associate al fenotipo Bardet-Biedl.



BBS4 è espresso nei centrioli satellite dei centrosomi e nei corpi basali delle ciglia primarie, dove funziona come proteina adattatrice della subunità p150 del complesso di trasporto della dineina per reclutare, a livello dei satelliti, PCM1 (pericentriolar material protein 1), una proteina che si ritiene essere coinvolta nel trasporto dei componenti necessari per la replicazione del centrosoma e nell'organizzazione e nell'ancoraggio dei microtubuli che derivano dal centro di organizzazione dei microtubuli (MTOC).

La deplezione di BBS4 nelle cellule di mammifero risulta in difetti strutturali e funzionali del centrosoma (Kim et al., 2004), mentre il silenziamento di BBS4 induce la mislocalizzazione di PCM1 e il concomitante disancoraggio dei microtubuli del centrosoma, arresto della divisione e morte cellulare per apoptosi.

Il sequenziamento esomico non ha, ad oggi, permesso di chiarire l'eziopatogenesi della Sindrome di Jeune nella famiglia B.

Nonostante la rarità della patologia e le caratteristiche di isolato genetico della Sardegna non abbiamo rilevato nella famiglia B varianti potenzialmente patogenetiche nel gene DYNC2H1. La presenza della variante c.3694G>A nel gene DYNC2H1 è stata esclusa nel probando sia con il sequenziamento esomico che con re-sequencing dell'esone 25.

Il paziente presenta, nel gene DYNC2H1, sei varianti riportate in dbSNP135. Una di queste, rs189806840 (c.3682C>A), localizzata nell'esone 25, dà luogo ad una sostituzione aminoacidica missenso (p.L1228I), ha una MAF globale < 0.01 e non è stata rilevata dal progetto 1000Genomi nella popolazione CEU. La variante, presente in eterozigosi, è stata identificata recentemente in un individuo affetto che presentava altre due mutazioni nel gene ma si ritiene debba essere esclusa come causa di malattia in quanto polimorfismo annotato come non patogenico (Schmidts M, 2013).

Sempre nell'esone 25, nel probando della famiglia B è stata identificata una variazione (c.3660T>C) non annotata nei database pubblici. La variante è presente in eterozigosi e causa una sostituzione sinonima (p.P1220P).

È stata evidenziata per il gene DYNC2H1 una incompleta copertura esonica (72/90 esoni), per cui il coinvolgimento di questo gene nella patologia non può essere escluso in maniera definitiva, anche considerato che le mutazioni sinora descritte sono localizzate lungo tutta la sequenza del gene.

La lista finale di geni risultante dall'analisi bioinformatica conteneva solo un gene, MAP1S, che poteva essere considerato un valido candidato per la malattia in quanto codifica per una proteina implicata nell'ancoraggio del centro di organizzazione dei microtubuli ai centrosomi. Il re-sequencing dell'esone 5 del gene MAP1S non ha però consentito di confermare nel probando la presenza delle varianti rilevate con l'exome sequencing.

## CONCLUSIONI

La Sindrome di Jeune (JATD) appartiene alla famiglia di disordini scheletrici noti come “ciliopatie”, associati a disfunzioni delle cilia primarie. Le ciliopatie sono causate da mutazioni in geni che codificano per proteine essenziali per la genesi e il mantenimento delle strutture ciliari. La maggior parte di queste patologie sembra avere un tipo di ereditarietà oligogenica, e studi recenti evidenziano che mutazioni in geni differenti possono avere un effetto modificatore sul fenotipo. Complessivamente, sono noti almeno 50 loci associati a 15 ciliopatie (Davis E, Katsanis N, 2012).

Per la sola Sindrome di Jeune sono stati finora individuati quattro geni candidati ma, considerando che il funzionamento delle cilia dipende da complesse interazioni tra proteine coinvolte nella formazione delle strutture, nei meccanismi di trasporto e nei pattern di segnalazione importanti durante l’embriogenesi, potrebbero essere coinvolti nell’eziopatogenesi della malattia geni ancora non identificati.

Negli anni passati, la diagnosi molecolare di JATD è stata ostacolata dalla eterogeneità genetica della patologia che, insieme al numero ridotto degli affetti, limita più fortemente l’identificazione di geni responsabili di malattie rare.

Attualmente, i progressi raggiunti nel campo delle metodiche di sequenziamento del DNA consentono di superare le maggiori criticità legate all’individuazione di varianti rare attraverso il sequenziamento massivo parallelo (MPS) del genoma intero o della sola porzione codificante.

Questo progetto di ricerca si prefiggeva di individuare, attraverso il sequenziamento dell’esoma, uno o più geni candidati responsabili della Sindrome di Jeune nelle famiglie in studio presso il nostro Dipartimento.

Il sequenziamento ha permesso di chiarire le basi eziopatogenetiche della patologia in una delle due famiglie, sostituendosi alle tradizionali metodiche di indagine genetica come l’analisi di linkage e il sequenziamento diretto dei geni candidati con la metodica di Sanger. Mediante l’analisi bioinformatica dei dati e l’applicazione di filtri discreti e consecutivi delle varianti è stata rilevata una nuova sostituzione (c.3694G>A, p.D1232N)

nel gene *DYNC2H1* che codifica per la catena pesante della dineina citoplasmatica 2, confermandolo come principale gene responsabile delle anomalie scheletriche associate al fenotipo Jeune.

La caratterizzazione dei difetti molecolari alla base della Sindrome di Jeune consentirà di poter sviluppare un protocollo diagnostico-terapeutico e di mettere in atto misure di prevenzione attraverso l'identificazione dei portatori sani, delle coppie a rischio e la diagnosi prenatale.

### **Obiettivi futuri**

Il ruolo patogenetico della variante c.3694G>A verrà approfondito mediante studi funzionali. In particolare, esperimenti di immunoprecipitazione in cellule di mammifero transfettate con i costrutti della catena pesante mutagenizzata e delle subunità accessorie (catena leggera intermedia e catena leggera) consentiranno di stabilire se la sostituzione D1232N alteri la funzionalità della proteina determinando una riduzione o una mancata interazione tra la catena pesante e le altre catene del complesso.

La preparazione e la trasfezione di due costrutti della catena pesante con epitopi differenti, riconoscibili da due anticorpi distinti, consentirà inoltre di verificare se la mutazione causa alterazioni nel processo di omodimerizzazione, fondamentale per la formazione e l'attività di proteina di trasporto ciliare dell'olocomplexo enzimatico della dineina.

Ulteriori esperimenti saranno condotti per definire se e in che modo una mutazione che ricade nel DHC\_2 del tail domain sia in grado di compromettere i processi motori regolati dal motor domain, come nel caso del modello murino *Loa*. Per testare l'effetto della variante p.D1232N sull'attività meccanochimica della dineina, sarà allestito un saggio di motilità in cellule di mammifero. Misurando l'attività ATPasica della dineina sarà possibile stabilire eventuali variazioni, tra dineina mutante e dineina wild-type, nei valori della costante di Michaelis per i microtubuli ( $K_{mt}$ ), che rappresenta il grado di affinità della dineina per i microtubuli durante l'idrolisi dell'ATP.

Studi funzionali in vivo e in vitro permetteranno inoltre di poter valutare l'effetto delle varianti rilevate in geni candidati modificatori.

Per quanto riguarda il probando della famiglia B non sono state rilevate nel gene DYNC2H1 varianti presumibilmente patogenetiche. Poiché la cattura dell'esoma non ha permesso la copertura di tutti i 90 esoni del gene, gli esoni non analizzati verranno risequenziati con la tecnologia di sequenziamento di nuova generazione "Ion Torrent"(Life Technologies). Data la rarità della malattia e le caratteristiche di isolato genetico della Sardegna, è infatti plausibile che anche il probando della famiglia B presenti mutazioni nel gene DYNC2H1 che causano alterazioni nella funzionalità della dineina-2.

## Appendice

### Ciliopatie

Il termine “ciliopatie” designa un insieme di patologie genetiche rare con base eziologica comune rappresentata da alterazioni a carico di proteine espresse a livello di particolari strutture cellulari, le cilia. Sebbene le singole ciliopatie possano essere considerate malattie genetiche rare, considerate complessivamente la prevalenza è pari a 1/2000.

Mentre negli invertebrati l'espressione delle cilia è confinata ai neuroni sensoriali specializzati nella ricezione e trasduzione di stimoli di tipo chimico/ambientale, nei vertebrati le cilia hanno distribuzione diffusa in tutti gli organi e tessuti (Evans et al., 2006). Proprio per la loro espressione pressochè ubiquitaria e per la molteplicità dei meccanismi biologici che, almeno in parte, vengono regolati dalle cilia, i difetti a carico di queste strutture sono alla base di un vasto range di fenotipi patologici che possono causare disfunzioni limitate a singoli organi, come nel caso della sindrome del rene policistico (Polycystic Kidney Disease, PKD) o specifiche sindromi pleiotropiche con fenotipo e manifestazioni cliniche, comprendenti cisti renali, degenerazione retinica, polidattilia e alterazioni più o meno severe a carico dell'apparato scheletrico, sovrapponibili (fig.23)

Le manifestazioni cliniche e la severità dei fenotipi sono quindi correlate al tipo di organo compromesso dalle alterazioni ciliari, al tipo di cilia (cilia mobili o primarie) e al ruolo specifico delle cilia in un determinato organo/tessuto.

Le anomalie a carico di proteine implicate nella formazione e nella funzione delle cilia mobili sono state associate a immotilità degli spermatozoi e delle cilia del tratto respiratorio già nel 1976, quando Afzelius ricollegò la mancanza dei movimenti ciliari ad alterazioni nei bracci delle dineine assonemali nei pazienti affetti da Discinesia Ciliare Primaria (PCD, primary ciliary dyskinesia), nota anche come Sindrome di Kartagener (MIM 244400) (Afzelius BA, 1976). Studi successivi hanno poi confermato il ruolo delle mutazioni a carico delle catene pesanti e intermedie delle dineine assonemali (DNAH5,

DNAH11 e DNAI1) nell'eziopatogenesi della PCD (Olbrich et al., 2002; Knowles et al., 2012; Zariwala et al., 2006; Zariwala et al., 2007).

Mentre le disfunzioni che determinano l'immotilità delle cilia sono state da tempo correlate a differenti manifestazioni cliniche, il ruolo delle cilia primarie nella patogenesi di numerose malattie è emerso solo recentemente, grazie soprattutto a studi effettuati su modelli murini. Gli studi sul topo *orpk* (Oak Ridge Polycystic Kidney) modello della sindrome autosomica recessiva del rene policistico (ARKPD, OMIM 263200), hanno per esempio consentito di chiarire il ruolo delle cilia primarie nella patogenesi delle patologie caratterizzate da cistogenesi renale (CDKs, cystic diseases of the kidney).

Si stima che almeno 100 disordini rari possano essere determinati, almeno parzialmente, da difetti strutturali/funzionali a carico delle cilia primarie e, ad oggi, sono stati identificati circa 50 loci associati a 15 ciliopatie (Katsanis N, Davis E, 2012).

Ciliopathy		Ciliopathy Phenotype									
		Retinitis pigmentosa	Renal cystic disease	Situs inversus	Mental retardation/ developmental delay	Hypoplasia of corpus callosum	Dandy-Walker malformation	Posterior encephalocele	Hepatic disease	Polydactyly	Craniofacial abnormality
Leber Congenital Amaurosis	LCA	○			○						
Nephronophthisis	NPHP		○								
Senior-Loken Syndrome	SLS	○	○								
Joubert Syndrome	JBTS	○	○	○	○						
Bardet-Biedl Syndrome [119]	BBS	○	○	○	○	○		○	○	○	○
Meckel-Gruber Syndrome	MKS		○	○	○	○	○	○	○	○	
Oro-facial-digital Syndrome	OFD		○		○	○					
Cranioectodermal Dysplasia	CED	○	○		○	○				○	
Short Rib-Polydactyly	SRP	○	○						○	○	○
Jeune Asphyxiating Thoracic Dystrophy	JATD	○	○						○	○	○

**Fig.23** Sovrapposizione fenotipica in dieci ciliopatie. (Katsanis N et al.,2012).

### Struttura e classificazione delle cilia

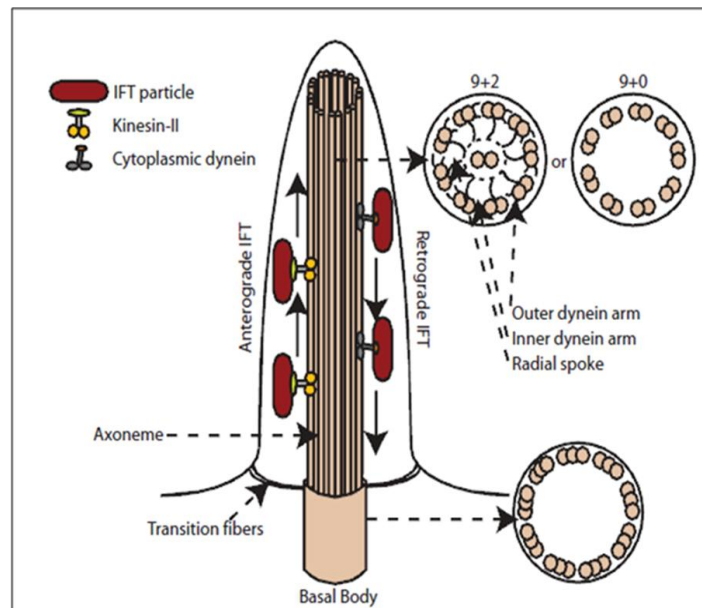
Le cilia sono strutture microtubulari che si originano da centrioli modificati chiamati corpi-basali. Pur esprimendosi in maniera preferenziale negli epitelii, le strutture ciliari hanno distribuzione pressoché ubiquitaria con poche eccezioni: sono presenti infatti nelle cellule endoteliali, nei fibroblasti, nei neuroni e nei condrociti, non sono espresse invece nelle cellule della linea mieloide e linfoide.

Dal punto di vista strutturale le cilia si differenziano in cilia *mobili* e in cilia *primarie*.

Le cilia *mobili* sono sottili protrusioni della membrana plasmatica che si estendono per circa 20  $\mu\text{m}$  dalla superficie cellulare. Le cilia mobili hanno una tipica struttura microtubulare "9+2", nella quale nove doppiette di microtubuli periferici sono disposti intorno ad un nucleo centrale costituito da un'altra doppietta di microtubuli. La struttura, circondata da un prolungamento della membrana plasmatica, viene chiamata "assonema" (fig.24). Alla base del cilium è situato il corpo basale, una struttura cilindrica che deriva dal centriolo madre e che collega il cilium al citoplasma, consentendogli di estendersi al di fuori della cellula. Le cilia mobili tendono ad essere concentrate sulla superficie apicale della cellula e a muoversi in maniera sincrona determinando il flusso dei liquidi fisiologici come il muco nell'epitelio respiratorio o il fluido cerebrospinale nei ventricoli cerebrali e nel midollo spinale; sono inoltre responsabili del movimento dello sperma nelle tube di Falloppio. Affinché il movimento delle cilia avvenga in maniera sincrona e coordinata, proteine accessorie, come le dineine assonemali, regolano la forza e la direzione dello spostamento spaziale del cilium.

Le cilia *primarie* (fig.24) si differenziano da quelle mobili per la mancanza della doppietta di microtubuli centrale, hanno infatti una struttura "9+0" e non sono generalmente dotate di movimento, ad eccezione di quelle che ricoprono il nodo dell'embrione nei vertebrati. Le cilia primarie hanno un ruolo fondamentale nei meccanismi di trasduzione del segnale in diversi pathways di segnalazione, nella regolazione della disposizione degli organi nell'asse destro/sinistro (LR, left right symmetry) durante l'embriogenesi, nella regolazione del ciclo cellulare.





**Fig.24 Struttura base del cilium.** Rappresentazione schematica del cilium e cross-sezione del corpo basale costituito da triplette di microtubuli e struttura dell'assonema "9+0" e "9+2"; sono riportate in figura i bracci delle dineine necessari per la produzione della forza necessaria per il movimento. Lungo la doppietta di microtubuli esterna i motori molecolari trasportano le particelle IFT (Cardenas-Rodriguez and Badano, 2009).

### Il trasporto intraflagellare

Poiché all'interno delle ciglia non avviene sintesi proteica, le cellule hanno sviluppato un meccanismo specializzato chiamato "trasporto intraflagellare" (IFT, Intraflagellar Transport) che consente il trasferimento di nutrienti e altre molecole essenziali dai diversi compartimenti cellulari lungo l'asse ciliare (Pedersen and Rosenbaum, 2008).

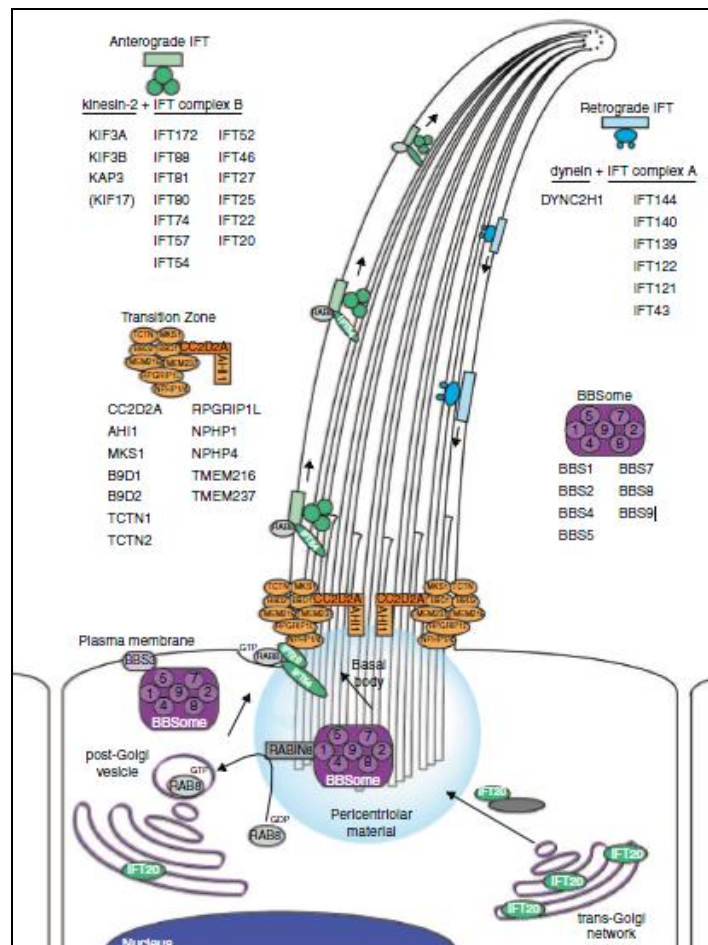
Il trasporto intraflagellare, descritto per la prima volta nell'alga flagellata unicellulare verde *Chlamydomonas reinhardtii* (Kozminski et al., 1993), richiede l'azione coordinata di motori molecolari necessari per il movimento delle particelle IFT dal corpo basale alla parte apicale del cilium, e viceversa, con due meccanismi distinti: il trasporto anterogrado e il trasporto retrogrado (fig.26).

Il movimento anterogrado è coordinato dal complesso intraflagellare di tipo B, costituito da almeno 14 proteine IFT che si associano alla Kinesina II, un complesso proteico eterotrimerico formato da due subunità motrici, KIF3A e KIF3B nei vertebrati, e da una

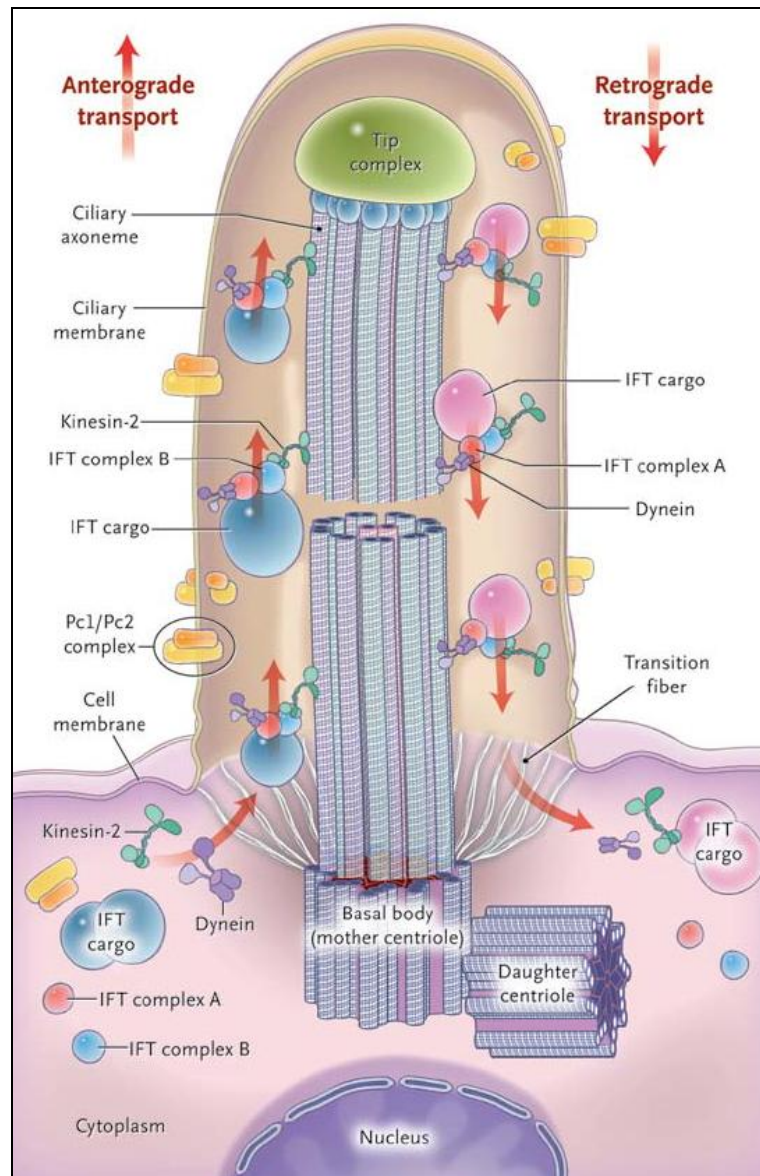
subunità non motrice chiamata KAP. Possono essere reclutate per il movimento anterogrado anche altre chinesine accessorie (Huber C, Cormier-Daire V, 2012) (fig.25-26).

Il motore responsabile del trasporto retrogrado è il complesso intraflagellare di tipo A, costituito da sei proteine IFT in associazione con il motore della dineina.

Il movimento IFT è cruciale per la funzionalità delle cilia e anomalie a carico delle strutture che coordinano il trasporto intraflagellare sono associate a differenti fenotipi patologici nell'uomo (Cardenas-Rodriguez and Badano, 2009). Un tipo cellulare particolarmente dipendente dal corretto funzionamento delle proteine coinvolte nell'IFT è il fotorecettore. Il segmento esterno del fotorecettore deriva infatti dalla membrana plasmatica di un cilium primario modificato che lo connette con il segmento interno. In generale, i difetti nei meccanismi di trasporto intraflagellare alterano la funzionalità del fotorecettore in misura proporzionale alla compromissione dell'IFT causando degenerazione retinica, una caratteristica conseguenza fenotipica della disfunzione ciliare (Insinna and Besharse, 2008). Mutazioni nelle proteine IFT88 o IFT57, che causano la totale abolizione o la riduzione del trasporto IFT, risultano nella perdita o in una ridotta estensione del segmento esterno del cilium (Pazour et al., 2002; Krock and Perkins, 2008), mentre la deplezione di Kif3a, in seguito alla quale si verifica l'accumulo di opsina e arrestina, proteine normalmente trasportate nel segmento esterno, causano la morte del fotorecettore (Marszalek et al., 2000).



**Fig.25** Proteine coinvolte nei meccanismi di trasporto intraflagellare del cilium. Sono schematizzate in figura le proteine espresse a livello ciliare divise a seconda della funzione e della loro localizzazione relativa nelle diverse strutture del cilium (Davis & Katsanis, *Current Opinion in Genetics & Development* 2012).



**Fig.26 Struttura del cilium e trasporto intraflagellare.** Il cilium consiste di un assonema rivestito da una membrana plasmatica specializzata e assemblato a partire da un centriolo madre (corpo basale). Lungo la struttura dell'assonema si verificano due tipi di trasporto intraflagellare (anterogrado e retrogrado) regolati da specifiche proteine cargo e accessorie che mediano i meccanismi di trasporto dal corpo basale all'apice del cilium e viceversa. (Hildebrandt F et al., 2011)

### *Cilia e Organogenesi*

La mancata regolazione dei meccanismi alla base della corretta disposizione degli organi nell'asse sinistro/destro (LR) durante l'embriogenesi è causa del cosiddetto "situs inversus", fenotipo comune a molte ciliopatie ed evidenziato già nel 1976 da Afzelius nei pazienti affetti da PCD, che, oltre ad infezioni respiratorie e sterilità per immotilità degli spermatozoi, presentavano una disposizione degli organi speculare rispetto alla norma (Afzelius, 1976).

L'importanza delle cilia nella regolazione di tali meccanismi è stata in seguito confermata dagli esperimenti di Nonaka. Mediante esperimenti sul modello murino *Kif3b* venne infatti dimostrato che le cilia mobili del nodo embrionale sono responsabili di un movimento direzionale di fluido extra embrionico verso sinistra che rappresenta il primo segnale per la disposizione degli organi nell'asse LR (Nonaka et al., 1998). Mutazioni nella dineina *Ird* nel mutante murino *iv* (inversus viscerum) confermarono l'importanza delle cilia del nodo embrionale evidenziando, anche in questo caso, come le alterazioni nei flussi embrionali diretti dalle cilia determinassero anomalie nella disposizione degli organi interni (Okada et al., 1999; Supp et al., 1999).

Solo nel 2003, McGrath propose che, oltre alle cilia mobili, ricoprissero un ruolo fondamentale nell'embriogenesi anche le cilia primarie sensoriali, non mobili, disposte intorno alle cilia mobili della parte centrale del nodo embrionale. Secondo McGrath, le cilia sensoriali si piegherebbero in risposta ai flussi verso sinistra regolati dalle cilia mobili attivando un pathway di segnalazione mediato dal rilascio di calcio (McGrath et al., 2003).

Conferma del ruolo delle cilia sensoriali nei pattern di regolazione LR è data dall'evidenza che i modelli murini mutanti *Pkd2* sviluppano situs inversus in presenza di cilia mobili del nodo morfologicamente normali e che la segnalazione al calcio mediata dalle cilia sensoriali, e attivata in risposta al piegamento delle cilia durante il flusso di urina, dipende dalla policistina-2, proteina frequentemente mutata nella sindrome del rene policistico (PKD) (Pennekamp et al., 2002).

### *Cilia e pathways di segnalazione cellulare*

Sulla membrana del cilio primario sono espressi recettori necessari alla cellula per rilevare stimoli fisici/meccanici e legare ormoni, chemiochine e fattori di crescita, oppure per legare morfogeni tramite cui avviene la modulazione dei pathways di segnalazione. Il cilio si comporta quindi come una sorta di “antenna” in grado di captare una grande varietà di segnali extracellulari inerenti l’osmolarità, la termosensibilità, la meccanosensibilità, la risposta agli ormoni, e nel tradurre tali segnali in un insieme di risposte cellulari come la proliferazione, la polarità, la crescita e la differenziazione cellulare.

Negli ultimi anni, diversi meccanismi di segnalazione, come quella Wnt (canonica e non canonica), HH (Hedgehog), FGF (Fibroblast Growth Factor pathway), Notch, mTor, PDGF e Hippo, sono stati associati al cilium primario. Tra questi, le cilia e i componenti del citoscheletro sembrano avere un ruolo particolarmente importante nei pathways di segnalazione Wnt e Hedgehog.

Sia la segnalazione Wnt canonica che quella non canonica sono mediate da molecole della famiglia di glicoproteine Wnt che legano specifici recettori di membrana (May-Simera and Kelley, 2012). I due meccanismi di segnalazione si differenziano per gli effettori a valle. Il pathway Wnt canonico risulta infatti nella modulazione della  $\beta$ -actina citoplasmatica e nell’attivazione della trascrizione di geni target. Il pathway Wnt non canonico è invece coinvolto nei meccanismi di regolazione della polarità cellulare durante la morfogenesi ed è necessario per i riarrangiamenti polari della cellula conosciuti come “estensione convergente”. L’implicazione delle proteine ciliari nella regolazione dei meccanismi Wnt è stata evidenziata sia per il pathway canonico che per quello non canonico (fig.27).

Altri pathways in cui l’integrità delle strutture ciliari è essenziale per i meccanismi di segnalazione sono quelli mediati dalle molecole “hedgehog”. Nei mammiferi, il Sonic hedgehog (Shh), l’Indian hedgehog (Ihh) e il Desert hedgehog (Dhh) costituiscono la famiglia di proteine segnale hedgehog (Hh) che, legandosi ai recettori Patched (Ptc),

sono in grado di attivare specifiche cascate di trasduzione del segnale (Chen et al., 2009; Humke et al., 2010; Lum and Beachy, 2004).

L'Indian hedgehog agisce infatti come molecola segnale nell'ossificazione dell'endocondrio regolando la fase di maturazione dei condrociti, mentre il Sonic hedgehog e i suoi effettori a valle sono necessari nelle fasi precoci dei pattern di formazione dello scheletro. Il fenotipo dei topi knockout per *Ihh* o per *Pthrp* (parathyroid-hormone related protein), un ormone i cui livelli sono regolati da *Ihh* e *Gli3*, si caratterizza per un anomalo sviluppo osseo che comprende riduzione delle coste e della gabbia toracica.

Sebbene il meccanismo di segnalazione hedgehog non sia stato ancora definito in modo preciso, sembra che l'esatta localizzazione subcellulare delle differenti molecole segnale del pathway Hh dipenda dai meccanismi di trasporto intraflagellare del cilium, che sarebbe pertanto indispensabile perché la segnalazione avvenga in maniera corretta.

La connessione tra cilia e Sonic hedgehog è stata evidenziata per la prima volta nei mutanti murini per il gene *Ift* "wim" (*Ift172*) e "flexo" (*Ift88*) (Huangfu et al, 2003) e, in seguito, nel topo *alien*, nel quale mutazioni nel gene *Ift139* determinano difetti nelle cilia primarie e iperattivazione del *Shh*, confermando l'importanza dell'integrità delle strutture ciliari e il ruolo delle proteine di trasporto intraflagellare nella segnalazione Hh. Difetti nei pathways di segnalazione Hh sono generalmente associati alle anomalie scheletriche di patologie come le SRPs, la Sindrome di Ellis-Van Creveld, la Sindrome di Jeune, la Sindrome di Sensenbrenner e la disostosi acrofaciale di Weyers e a fenotipi come la polidattilia post-assiale, anomalie dei genitali esterni e difetti craniofaciali.

Canonical Wnt	Cilia or Wnt association
Wnt proteins	
Inv (NPHP2)	Ciliary localization and physical interaction with Dvl
Dvl	Docking and polarization of basal bodies, basal body localization, targeted for degradation by Inv
Cilia-associated proteins	
Bbs-associated	Hyperactive Wnt response in knockout cell lines
Kif3a <sup>2</sup>	Up regulation of cellular Wnt response in knockout cell lines and mutant mice
Ift88 <sup>2</sup>	Up regulation of cellular Wnt response in mutant mice
Ift40	Increase in expression of canonical Wnt pathway genes in kidney of mutant mice
Ift20	Increase in nuclear beta-catenin and expression of Wnt target genes in kidney of mutant mice
Ofd1	Up regulation of cellular Wnt response in mutant mice
Chibby	Binds beta-catenin preventing nuclear entry negatively regulating Wnt signaling
Seahorse	Binds to Dvl, constrains Wnt signaling in zebrafish
SREBP1c	Over expression disrupts ciliogenesis and increases canonical Wnt signaling in Xenopus.
Ahi1/Jbn	Abrogated Wnt signaling in kidney and cerebellum of mutant mice, facilitates beta-catenin entry into nucleus
<b>Non-canonical Wnt</b>	
PCP effectors	
Inturned	Highly expressed in ciliated tissue, required for ciliogenesis. Actin assembly, Rho localization, docking of basal bodies
Fuzzy	Required for axoneme elongation, predicted role in vesicular trafficking
Fritz	Expressed in ciliated tissue, required for ciliogenesis, mutations identified in human ciliopathies
Core PCP proteins	
Dubroya	Regulates ciliogenesis, Apical Actin assembly
Frizzled	Defective ciliogenesis at zebrafish Kupffers Vesicle
Dvl	Regulates ciliogenesis, Actin assembly, Rho localization, docking of basal bodies, associated with human ciliopathy proteins TMEM216 and TMEM67
Celsr2/Celsr3	Regulates ciliogenesis in multiciliated ependymal cells via basal body docking at apical plasma membrane
Prickle	Regulates cilium length in zebrafish
Vangl2 <sup>2</sup>	Localizes to some cilia, xenopus basal body localization and ciliogenesis, zebrafish conflicting data

**Fig.27** *Pathways di segnalazione Wnt canonici e non canonici associati a proteine ciliari (May-Simera, Kelley, 2012)*

### *Cilia e ciclo cellulare.*

Recenti studi indicano un ruolo fondamentale delle cilia anche nel controllo della divisione cellulare. Le cilia sono strutture post-mitotiche presenti durante la fase G0/G1 e all'inizio della fase S del ciclo cellulare. Affinché la cellula possa dividersi è pertanto necessario che il cilio primario venga disassemblato rendendo disponibile il centriolo. L'assonema ciliare deriva dal corpo basale, il centriolo madre del centrosoma che dirige l'assemblaggio del fuso bipolare durante la mitosi. Durante la divisione cellulare la proteina del centrosoma Aurora A (AurA), interagendo con HEF1, guida il riassorbimento del cilium attraverso la regolazione dell'istone deacetilasi HDAC6 e la deacetilazione della tubulina assonemale (Hildebrandt F et al., 2011).

La progressione del ciclo cellulare è regolata non solo dalla disponibilità dei centrioli ma anche dall'inibizione della ciliogenesi ad opera delle proteine del centrosoma. Per



esempio, CP110, una proteina che controlla la duplicazione del centrosoma e la citogenesi, inibisce la ciliogenesi attraverso l'interazione con Cep270 e CEP290, una proteina che risulta mutata in diverse ciliopatie.

Altre proteine necessarie per i meccanismi di coordinamento tra ciclo cellulare e ciliogenesi sono le chinasi NIMA-related, o Neks; mutazioni in Nek1 e Nek8 causano nei modelli murini policistici del rene, tratto caratteristico di molte ciliopatie (Upadhyaya et al., 2000; Liu et al., 2002).

In vivo, la disfunzione ciliare non sembra avere ripercussioni sulla progressione del ciclo cellulare associate a fenotipi oncologici, suggerendo che i checkpoints del ciclo cellulare non siano compromessi in modo deleterio dalle anomalie ciliari.

## BIBLIOGRAFIA

- Afzelius BA. A human syndrome caused by immotile cilia. *Science*, 1976 Jul 23;193(4250):317-9.
- Adzhubei IA et al. A method and server for predicting damaging missense mutations. *Nature methods*, VOL.7 NO.4 APRIL 2010.
- Bamshad MJ et al. Exome sequencing as a tool for Mendelian disease gene discovery. *Nature Reviews Genetics* Volume 12, November 2011.
- Beales PL et al. IFT80, which encodes a conserved intraflagellar transport protein, is mutated in Jeune asphyxiating thoracic dystrophy. *Nat Genet*. 2007 Jun;39(6):727-9.
- Bentley DR et al. Accurate whole human genome sequencing using reversible terminator chemistry. *Nature*. 2008 Nov 6;456(7218):53-9
- Bolze A et al. Whole-exome-sequencing-based discovery of human FADD deficiency. *Am J Hum Genet*. 2010 Dec 10;87(6):873-81
- Botstein D, Risch N. Discovering genotypes underlying human phenotypes: past successes for mendelian disease, future approaches for complex disease. *Nature Genetics supplement*, volume 33, 2003
- Bredrup C et al. Ciliopathies with Skeletal Anomalies and Renal Insufficiency due to Mutations in the IFT-A Gene WDR19. *Am J Hum Genet*. 2011 Nov 11;89(5):634-43
- Burrows M, Wheeler DJ. A Block-sorting Lossless Data Compression Algorithm Digital Systems Research Center Research Report 124, 1994
- Cardenas-Rodriguez M and Badano JI. Ciliary biology: understanding the cellular and genetic basis of human ciliopathies. *Am J Med Genet C Semin Med Genet*. 2009 Nov 15;151C(4):263-80.

Chang X, Wang K. wANNOVAR: annotating genetic variants for personal genomes via the web. *J Med Genet*. 2012 Jun 20

Chen J et al. ToppGene Suite for gene list enrichment analysis and candidate gene prioritization. *Nucleic Acids Research* 2009 Jul;37(Web Server issue):W305-11

Choi M et al. Genetic diagnosis by whole exome capture and massively parallel DNA sequencing. *Proc Natl Acad Sci U S A*. 2009 Nov 10;106(45):19096-101

Cock JA et al. The Sanger FASTQ file format for sequences with quality scores, and the Solexa/Illumina FASTQ variants. *Nucleic Acids Research*, 2010, Vol. 38, No. 6 1767–1771

Cortina H, Beltran J, Olague R, Ceres L, Alonso A, Lanuza A. 1979. The wide spectrum of the asphyxiating thoracic dysplasia. *Pediatr Radiol* 8:93–99.

Dagoneau N et al. DYNC2H1 mutations cause asphyxiating thoracic dystrophy and short rib-polydactyly syndrome, type III. *Am J Hum Genet*. 2009 May;84(5):706-11

Danecek P et al. The variant call format and VCFtools. *Bioinformatics applications note*. Vol.27 no.15 2011, 2156-2158.

Davis EE et al. TTC21B contributes both causal and modifying alleles across the ciliopathy spectrum. *Nat. Genet*. 2011 March ; 43(3): 189–196.

Deorowicz S, Grabowski S. Compression of DNA sequence reads in FASTQ format. *Bioinformatics Applications Note* Vol. 27 no. 6 2011, pages 860–862

Ding J et al. Gene targeting of GAN in mouse causes a toxic accumulation of microtubule-associated protein 8 and impaired retrograde axonal transport. *Hum Mol Genet*. 2006 May 1;15(9):1451-63

El Hokayem J et al. NEK1 and DYNC2H1 are both involved in short rib polydactyly Majewski type but not in Beemer Langer cases. *J Med Genet* 2012;49:227e233.

Ferragina P, Manzini G, 2000. Opportunistic Data Structures with Applications Proceedings of the 41st Annual Symposium on Foundations of Computer. *Science* p.390, 2000

Flicek P, Birney E. Sense from sequence reads: methods for alignment and assembly. *Nat Methods*. 2009 Nov;6(11 Suppl):S6-S12.

Frazer KA et al. Human genetic variation and its contribution to complex traits. *Nature Reviews Genetics* volume 10, April 2009

Gilissen C et al. Disease gene identification strategies for exome sequencing. *European Journal of Human Genetics* (2012) 20, 490–497

Gilissen C et al. Unlocking Mendelian disease using exome sequencing. *Genome Biology* 2011, 12:228

Goh G, Choi M. Application of whole exome sequencing to identify disease-causing variants in inherited human diseases. *Genomics Inform*. 2012 Dec;10(4):214-9.

Grissom PM et al. Identification of a novel light intermediate chain (D2LIC) for mammalian cytoplasmic dynein 2. *Mol Biol Cell*. 2002 Mar;13(3):817-29.

Hildebrandt F. et al. Ciliopathies. *N Engl J Med* 2011;364:1533-43.

Hou Y et al. A Dynein Light Intermediate Chain, D1bLIC, Is Required for Retrograde Intraflagellar Transport. *Mol Biol Cell*. 2004 October; 15(10): 4382–4394.

Huang DW et al. Systematic and integrative analysis of large gene lists using DAVID Bioinformatics Resources *Nature Protoc*. 2009;4(1):44-57

Huang DW et al. Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. *Nucleic Acids Res*. 2009;37(1):1-13

Huangfu D et al. Hedgehog signalling in the mouse requires intraflagellar transport proteins. *Nature*. 2003 Nov 6;426(6962):83-7.

Huber C, Cormier-Daire V. Ciliary Disorder of the Skeleton. *American Journal of Medical Genetics Part C (Seminars in Medical Genetics)* 160C:165–174 (2012)

Humke EW et al. The output of Hedgehog signaling is controlled by the dynamic association between Suppressor of Fused and the Gli proteins. *Genes Dev.* 2010 Apr 1;24(7):670-82.

Insinna C, Besharse JC. Intraflagellar transport and the sensory outer segment of vertebrate photoreceptors. *Dev Dyn.* 2008 Aug;237(8):1982-92.

Katsanis N et al. BBS4 is a Minor Contributor to Bardet-Biedl Syndrome and May Also Participate in Triallelic Inheritance. *Am J Hum Genet.* 2002 Jul;71(1):22-9

Katsanis N, Davis E. The ciliopathies: a transitional model into systems biology of human genetic disease. *Current Opinion in Genetics & Development* 2012, 22:290–303.

Khanna H et al. A common allele in RPGRIP1L is a modifier of retinal degeneration in ciliopathies. *Nat Genet.* 2009 Jun;41(6):739-45.

Kim JC et al. The Bardet-Biedl protein BBS4 targets cargo to the pericentriolar region and is required for microtubule anchoring and cell cycle progression. *Nature Genetics* 36, 462 - 470 (2004)

King SH et al. Subunit organization in cytoplasmic dynein subcomplexes. *Protein Sci.* 11 (2002), 1239–1250.

Knowles MR et al. Mutations of DNAH11 in patients with primary ciliary dyskinesia with normal ciliary ultrastructure. *Thorax.* 2012 May;67(5):433-41.

Kozminski et al. A motility in the eukaryotic flagellum unrelated to flagellar beating. *Proc. Natl. Acad. Sci. USA* Vol. 90, pp. 5519-5523, June 1993

Krock BL, Perkins BD. The intraflagellar transport protein IFT57 is required for cilia maintenance and regulates IFT-particle-kinesin-II dissociation in vertebrate photoreceptors. *J Cell Sci.* 2008 Jun 1;121(Pt 11):1907-15.

Langer LO Jr. 1968. Thoracic-pelvic-phalangeal dystrophy: Asphyxiating thoracic dystrophy of the newborn, infantile thoracic dystrophy. *Radiology* 91:447–456.

Langlois S et al. Current status in non-invasive prenatal detection of down syndrome, trisomy 18, and trisomy 13 using cell-free DNA in maternal plasma. *J Obstet Gynaecol Can.* 2013

Liang D et al., 2013. Non-invasive prenatal testing of fetal whole chromosome aneuploidy by massively parallel sequencing. *Prenat Diagn.* 2013 May;33(5):409-15.

Li H et al. The Sequence Alignment/Map format and SAMtools. *Bioinformatics Applications Note* Vol. 25 no. 16 2009, pages 2078–2079

Li H et al. Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics Original Paper* Vol. 25 no. 14 2009, pages 1754–1760

Li H et al. Mapping short DNA sequencing reads and calling variants using mapping quality scores. *Genome Res.* 2008 November; 18(11): 1851–1858.

Li R et al. De novo assembly of human genomes with massively parallel short read sequencing. *Genome Res.* 2010 20: 265-272

Li R. SOAP2: an improved ultrafast tool for short read alignment. *Bioinformatics.* 2009

Lin Liu et al. Comparison of Next-Generation Sequencing Systems. *Journal of Biomedicine and Biotechnology*, Volume 2012

Liu S et al. A defect in a novel Nek-family kinase causes cystic kidney disease in the mouse and in zebrafish. *Development.* 2002 Dec;129(24):5839-46

Louie CM et al. AHI1 is required for photoreceptor outer segment development and is a modifier for retinal degeneration in nephronophthisis. *Nat Genet.* 2010 Feb;42(2):175-80.

Lupski JR et al. Whole-Genome Sequencing in a Patient with Charcot–Marie–Tooth Neuropathy. *N Engl J Med* 2010;362:1181-91.

Mardis ER. Next-Generation DNA Sequencing Methods. *Annu. Rev. Genomics Hum. Genet.* 2008.9:387–402

Marszalek JR et al. Genetic evidence for selective transport of opsin and arrestin by kinesin-II in mammalian photoreceptors. *Cell.* 2000 Jul 21;102(2):175-87

May-Simera HL, Kelley MW. Cilia, Wnt signaling, and the cytoskeleton. *Cilia* 2012, 1:7

McGrath J et al. Two Populations of Node Monocilia Initiate Left-Right Asymmetry in the Mouse. *Cell*, Vol. 114, 61–73, July 11, 2003

Merrill AE et al. Ciliary abnormalities due to defects in the retrograde transport protein DYNC2H1 in short-rib polydactyly syndrome. *Am J Hum Genet.* 2009 Apr;84(4):542-9.

Metzker ML. Sequencing technologies-the next generation. *Nature Reviews Genetics* volume 11, January 2010

Mikami A et al. Molecular structure of cytoplasmic dynein 2 and its distribution in neuronal and ciliated cells. *J Cell Sci.* 2002 Dec 15;115(Pt 24):4801-8.

Miller S.A. et al. (1998) A simple salting out procedure for extracting DNA from human nucleated cells. *Nucleic Acids Research*, 16, 1215

Morgan NV et al. A locus for asphyxiating thoracic dystrophy, ATD, maps to chromosome 15q13. *J Med Genet* 2003;40:431-435

Nagai T et al. Del(12)(p11.21p12.2) associated with an asphyxiating thoracic dystrophy or chondroectodermal dysplasia-like syndrome. *Am J Med Genet* 1995 Jan 2;55(1):16-8

Ng SB et al. Exome sequencing identifies the cause of a Mendelian disorder. *Nat Genet.* 2010 January; 42(1): 30–35. doi:10.1038/ng.499.

Ng SB et al. Massively parallel sequencing and rare disease. *Human Molecular Genetics*, 2010, Vol. 19, Review Issue 2.

Ng SB et al. Targeted capture and massively parallel sequencing of 12 human exomes. *Nature*. 2009 Sep 10;461(7261):272-6.

Nielsen R et al. Genotype and SNP calling from next-generation sequencing data. *Nature Reviews Genetics*, volume 12 June 2011

Nikopoulos K et al. Next-Generation Sequencing of a 40 Mb Linkage Interval Reveals TSPAN12 Mutations in Patients with Familial Exudative Vitreoretinopathy. *Am J Hum Genet*. 2010 February 12; 86(2): 240–247.

Nonaka S et al. Randomization of left-right asymmetry due to loss of nodal cilia generating leftward flow of extraembryonic fluid in mice lacking KIF3B motor protein. *Cell*. 1998 Dec 11;95(6):829-37

Oberklaid F et al. Asphyxiating thoracic dysplasia, Clinical, radiological, and pathological information on 10 patients. *Archives of Disease in Childhood*, 1977, 52, 758-765

Ocbina PJR et al. Complex interactions between genes controlling trafficking in primary cilia. *Nat. Genet.* 43, 547–553

Okada Y et al. Abnormal nodal flow precedes situs inversus in *iv* and *inv* mice. *Mol Cell*. 1999 Oct;4(4):459-68.

Olbrich H et al. Mutations in DNAH5 cause primary ciliary dyskinesia and randomization of left-right asymmetry. *Nat Genet.* 2002 Feb;30(2):143-4.

Ori-McKenney KM et al. A cytoplasmic dynein tail mutation impairs motor processivity. *Nat Cell Biol.* 2010 Dec;12(12):1228-34.

Papasavva T et al. Next generation sequencing of SNPs for non-invasive prenatal diagnosis: challenges and feasibility as illustrated by an application to  $\beta$ -thalassaemia. *Eur J Hum Genet*. 2013 Apr 10

Pazour GJ, Rosenbaum JL. Intraflagellar transport and cilia-dependent diseases. *Trends Cell Biol.* 2002 Dec;12(12):551-5.



Pedersen LB, Rosenbaum JL. Intraflagellar transport (IFT) role in ciliary assembly, resorption and signalling. *Curr Top Dev Biol.* 2008;85:23-61.

Perrault I et al. Mainzer-Saldino syndrome is a ciliopathy caused by IFT140 mutations. *Am J Hum Genet.* 2012 May 4;90(5):864-70.

Perrone CA et al. A novel dynein light intermediate chain colocalizes with the retrograde motor for intraflagellar transport at sites of axoneme assembly in chlamydomonas and Mammalian cells. *Mol Biol Cell.* 2003 May;14(5):2041-56.

Pfister KK et al. Genetic Analysis of the Cytoplasmic Dynein Subunit Families. *PLoS Genet.* 2006 Jan;2(1):e1.

Rajagopalan V et al. Cytoplasmic dynein-2: from molecules to human diseases. *Frontiers in Biology* February 2013, Volume 8, Issue 1, pp 119-126

Reck-Peterson SL et al. Motile properties of cytoplasmic dynein. In *"Handbook of dynein."* Pan Stanford Publishing. Editors: Keiko Hirose and Linda Amos.

Rompolas P et al. Chlamydomonas FAP133 is a dynein intermediate chain associated with the retrograde intraflagellar transport motor. *J Cell Sci.* 2007 Oct 15;120(Pt 20):3653-65.

Rozen S Skaletsky HJ. Primer3 on the WWW for general users and for biologist programmers. *Methods Mol Biol.* 2000;132:365-86.

Schmidts M et al. Exome sequencing identifies DYNC2H1 mutations as a common cause of asphyxiating thoracic dystrophy (Jeune syndrome) without major polydactyly, renal or retinal involvement. *J Med Genet.* 2013 May;50(5):309-23.

Supp DM et al. Mutation of an axonemal dynein affects left-right asymmetry in inversus viscerum mice. *Nature.* 1997 Oct 30;389(6654):963-6.

Tüysüz B et al. Clinical Variability of Asphyxiating Thoracic Dystrophy (Jeune) Syndrome: Evaluation and Classification of 13 Patients. *Am J Med Genet Part A* 149A:1727–1733, 2009

Wang K, Li M, Hakonarson H. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Research*, 38:e164, 2010

Wang JL et al. TGM6 identified as a novel causative gene of spinocerebellar ataxias using exome sequencing. *Brain*. 2010 Dec;133(Pt 12):3510-8.

Worthey EA et al. Making a definitive diagnosis: successful clinical application of whole exome sequencing in a child with intractable inflammatory bowel disease. *Genet. Med.* 2011 Mar;13(3):255-62.

Zariwala MA et al. Genetic Defects in Ciliary Structure and Function. *Annu. Rev. Physiol.* 2007. 69:423–50

Zariwala MA et al. Mutations of DNAI1 in primary ciliary dyskinesia: evidence of founder effect in a common mutation. *Am J Respir Crit Care Med.* 2006 Oct 15;174(8):858-66.

Zhang J et al. The impact of next-generation sequencing on genomics. *J Genet Genomics.* 2011 March 20; 38(3): 95–109.