

UNIVERSITÀ DEGLI STUDI DI NAPOLI FEDERICO II



DOTTORATO DI RICERCA IN

MEDICINA CLINICA E SPERIMENTALE

CURRICULUM IN SCIENZE PEDIATRICHE TRASLAZIONALI

XXXII Ciclo

(Anni 2017-2020)

Coordinatore: Prof. Francesco Beguinot

TESI DI DOTTORATO

TITOLO

**Sequenziamento dell'esoma in pazienti pediatrici con malattia senza
diagnosi**

TUTOR/RELATORE

Chiar.mo

Prof. Nicola Brunetti Pierri

Handwritten signature of Prof. Nicola Brunetti Pierri.

CANDIDATA

Dott.ssa Gerarda Cappuccio

Sommario

Introduzione	pagina 4
<i>Sequenziamento del DNA di ultima generazione: metodiche</i>	pagina 5
<i>Il sequenziamento dell'esoma, Whole Exome Sequencing (WES)</i>	pagina 8
<i>Applicazioni cliniche della NGS</i>	pagina 11
<i>Interpretazione delle varianti esoniche</i>	pagina 15
<i>Valutazioni cliniche pre- e post- analisi genomica e problemi etici</i>	pagina 17
<i>Programmi per malattie "senza diagnosi"</i>	pagina 20
Obiettivi dello studio	pagina 23
Pazienti e Metodi	pagina 24
<i>Criteri di inclusione ed esclusione dello studio</i>	pagina 24
<i>WES ed analisi bioinformatica</i>	pagina 25
<i>Relazione causa-effetto tra la variante identificata da WES e fenotipo</i>	pagina 27
Risultati	pagina 31
Caso 1- Nuova scoperta	pagina 34
Caso 2- Conclusivo con presentazione atipica	pagina 39
Caso 3- Gene candidato	pagina 44
Discussione	pagina 47
Conclusioni	pagina 52
Referenze	pagina 53

Introduzione

Le malattie genetiche incidono pesantemente sulla salute pubblica. Prese singolarmente, la maggior parte di loro è rara (incidenza pari a 5:10,000) o molto rara (definita in UK come patologia con incidenza 1:50,000 e negli USA come patologia che colpisce meno di 2,000 persone al mondo) [1]. Come gruppo, invece, esse hanno un'incidenza complessiva di 40-50:100,000 [2]. L'osservatorio delle malattie rare stima che in Italia ci siano 20 casi di malattie rare ogni 10,000 abitanti e ogni anno circa 19,000 nuovi casi. Il 20% dei casi riguarda pazienti in età pediatrica. Esse rappresentano, inoltre, una delle principali cause di morbidità e mortalità pediatrica: si stima che circa il 2.5-3% dei neonati presenti malformazioni congenite eventualmente associate a disabilità cognitiva su base genetica e che il 71% dei ricoveri ospedalieri pediatrici nel mondo occidentale riguardi pazienti con emergenti problematiche associate a sindromi genetiche [3, 4]. Il catalogo delle malattie genetiche OMIM (Online Mendelian Inheritance in Man) colleziona ad oggi 8,600 fenotipi e 6,200 di questi sono associati ad uno specifico difetto genetico [5]. Dal 2014 infatti OMIM ha incluso approssimativamente 300 nuovi fenotipi per anno e il numero totale dei geni associati a malattia continua ad aumentare [5].

Il raggiungimento di una specifica diagnosi genetica in pazienti affetti da una sospetta malattia genetica è un obiettivo di primaria importanza per le conseguenze sul paziente, la sua famiglia ed il sistema sanitario. La diagnosi di malattia genetica in bambini con anomalie congenite multiple e ritardo dello sviluppo sulla base della consulenza dismorfologica si raggiunge soltanto nel 22.5%-47% dei casi [2]. Negli ultimi anni, la disponibilità di metodiche diagnostiche come l'array-CGH (*array comparative genomic hybridization*), che interrogano l'intero genoma umano, ha permesso di migliorare notevolmente le capacità diagnostiche in pazienti con dismorfismi, anomalie congenite e disturbi neurocognitivi. Per questo motivo, l'array-CGH è diventato test di primo livello nella valutazione di questi pazienti ed ha permesso di definire che i riarrangiamenti cromosomici (o *copy numbers variants - CNVs*), sono una causa frequente causa di ritardo dello sviluppo, disabilità

cognitiva, disturbi comportamentali associata o meno ad anomalie congenite. Oltre all'array-CGH altri strumenti diagnostici come test di metilazione per specifiche regioni cromosomiche o l'analisi mirata di specifici geni sono disponibili quando un sospetto specifico di malattia è formulato. I pazienti impiegano in media 5 anni per raggiungere una diagnosi definitiva, accumulando così una grande quantità di dati medici e test diagnostici spesso invasivi, complessi e costosi [6]. Nonostante la disponibilità di questi strumenti diagnostici una percentuale ancora troppo elevata di pazienti nei quali viene sospettata una causa genetica rimane senza una diagnosi [7].

Le famiglie con bambini affetti da patologie gravi e sconosciute (pazienti “senza diagnosi”) rappresentano una sfida drammatica. Queste famiglie vivono un intenso stress emozionale a causa delle incertezze sulla prognosi e sul rischio di ricorrenza della malattia in future gravidanze. L'assenza di diagnosi inoltre preclude l'uso di terapie qualora queste siano disponibili. Pertanto la diagnosi genetica rimane un obiettivo fondamentale e necessario per aiutare la comprensione dei meccanismi alla base di queste patologie ma, anche e soprattutto, per garantire una più mirata ed efficace presa in carico e una migliore qualità di vita per i pazienti e le loro famiglie. Il rapido progresso di nuove tecnologie per il sequenziamento genico sta iniziando a dare un grosso contributo al raggiungimento di questi obiettivi.

Sequenziamento del DNA di ultima generazione: metodiche

Il sequenziamento del DNA basato sulla sequenziamento di nuova generazione (*Next Generation Sequencing*-NGS) include metodiche e tecnologie che differiscono in maniera significativa dal sequenziamento convenzionale di Sanger. Quest'ultimo è basato sul sequenziamento diretto di porzioni di genoma, singolarmente amplificate mediante *polymerase chain reaction* (PCR) e consente di sequenziare fino a 800 bp circa per singola reazione, mentre le tecniche NGS consentono di sequenziare parallelamente più regioni del genoma, da ~400 a 3000 Mb in una singola corsa dello

strumento. Allo stato attuale è possibile sequenziare 25 gigabasi al giorno (l'equivalente di circa 8 genomi umani) [8] a costi molto ridotti [9]. Le tecnologie di NGS richiedono tutte nella loro fase iniziale la generazione di una libreria genomica che può includere l'intero genoma del soggetto studiato o parte di esso (*targeted sequencing*) (**Figura 1**) [8]. Il genoma viene dapprima frammentato in molecole di circa 35-400 bp, che sono successivamente amplificate clonalmente in parallelo fino a generare circa 50 copie identiche per singola molecola (**Figura 1**). Successivamente, viene effettuato il sequenziamento vero e proprio per il quale possono essere utilizzati due differenti approcci: sequenziamento “per sintesi”, basato sul metodo di Sanger oppure sulla reazione di pirosequenza e il sequenziamento “per ligazione” [10, 11].

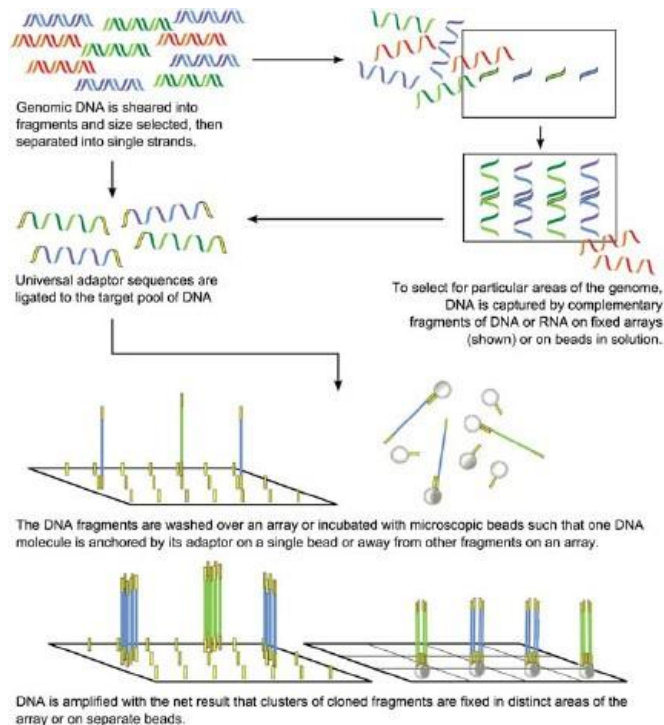


Figura 1. Formazione della libreria genomica. Il DNA è frammentato e le sequenze di interesse vengono legate a delle sequenze adattatrici. I frammenti di DNA così composti vengono legati ad un array o a delle biglie ed amplificati (Raffan et al.).

In particolar modo il sistema Roche/454 è basato su una combinazione di PCR in emulsione e piro-sequenziamento. Le sequenze target vengono legate a delle biglie e confinate in delle gocce di emulsione, in cui avviene l'amplificazione per PCR. Le sequenze ottenute vengono poi collocate su un vetrino, su cui viene letta la luminescenza generata dal rilascio di pirofosfato. Il sequenziamento Illumina/Solexa utilizza, anch'esso, una reazione di sintesi ma la sua particolarità è l'uso di nucleotidi modificati (marcati ognuno con un diverso fluorocromo). Questo metodo permette di sequenziare in maniera più precisa sequenze ripetute di uno stesso nucleotide. La tecnologia SOLID è l'unica basata non su reazioni di sintesi ma di ligasi e il sequenziamento avviene tramite l'uso di sonde di ligasi (*ligation probes*) della lunghezza di 8 bp, che includono un sito di legame (prima base), un sito di clivaggio (quinta base) e quattro differenti coloranti fluorescenti legati all'ultima base. Oltre a queste tecnologie già consolidate sono in avanzata fase di sviluppo nuove tecnologie che permettono il sequenziamento di singola molecola di DNA. Queste nuove tecnologie non richiedono la fase iniziale di amplificazione per PCR e si basano sul sequenziamento diretto del DNA. Questa metodica riduce le possibilità di errori legati all'amplificazione e richiede una quantità inferiore di materiale genomico di partenza [11].

Con la NGS l'acquisizione dei dati di sequenza avviene contestualmente per tutte le reazioni, con un notevole risparmio di tempo rispetto alla tecnologia precedente. L'enorme mole di dati generati, opportunamente processata e analizzata per via bioinformatica, viene poi confrontata con un genoma di riferimento [12]. L'elevata processività riduce molto tempi e costi, rendendo la tecnica sostenibile anche per un uso clinico più vasto. Rispetto al sequenziamento tradizionale di Sanger infatti, le tecniche di NGS (200 volte più veloci) sequenziano dai 25 milioni (454 sequencer) ai 18,000 miliardi di basi (Illumina HiSeq X Ten) per corsa con un costo di sequenziamento/bp stimato di 5 milioni di volte inferiore rispetto a quello della metodica Sanger [13]. Si calcola, per esempio, che

con questa tecnica il progetto “Genoma Umano” sarebbe durato solo 30 ore e avrebbe avuto un costo molto più basso (**Tabella 1**).

Sanger		NGS
3h	Tempo necessario per una corsa	10h-6 gg
Max 96	Sequenze analizzate per corsa	Milioni
80 Kb	Totale basi lette per corsa	500 Mb – 50 Gb
70 milioni \$	Costo per sequenziamento di un intero genoma umano	5000 \$

Tabella 1. Confronto tra sequenziamento Sanger e NGS.

Le sequenze ottenute per sequenziamento di Sanger rimangono comunque più accurate di quelle ottenute per NGS. In particolare, la NGS può avere una bassa copertura in porzioni genomiche che presentano un *coverage* minore di 20X (una stessa sequenza viene letta meno di 20 volte) e in regioni ricche in C e G, le cui varianti, specialmente se in eterozigosi, spesso sono misconosciute. Pertanto, nonostante la diffusione delle nuove tecnologie NGS, la tecnica di sequenziamento di Sanger continua ad avere una sua utilità e importanza nell’ambito clinico. Il metodo Sanger è, infatti, caratterizzato dalla facilità di allestimento e rimane un ottimo strumento per la ricerca di mutazioni causative in singolo gene [8]. Il maggior *coverage* e la maggiore sicurezza ne fanno inoltre uno strumento indispensabile per la conferma delle mutazioni riscontrate con la tecnica NGS. Infatti, nell’ambito del progetto oggetto di questa tesi, ogni variante ritenuta causativa è stata validata con metodica Sanger [12].

Il sequenziamento dell’esoma, Whole Exome Sequencing (WES)

Lo sviluppo di metodiche di sequenziamento del DNA di ultima generazione [*Next Generation Sequencing* (NGS)], permette il sequenziamento parallelo di un’immensa quantità di materiale genico, con la sostituzione dell’approccio gene per gene, e si avvale di apparecchiature e tecnologie miniaturizzate ed estremamente avanzate, consentendo il sequenziamento simultaneo di un numero elevatissimo di geni fino all’intero esoma [*whole exome sequencing* (WES), sequenziamento

dell'intero esoma], ossia di tutti i 20,000 geni umani. L'NGS può essere applicato anche all'intero genoma [*whole genome sequencing* (WGS), sequenziamento dell'intero genoma]. Rispetto al sequenziamento dell'intero genoma, il sequenziamento dell'esoma, ossia delle porzioni codificanti del genoma (i.e. esoni), pari a circa l'1-1.5% del genoma, è un approccio decisamente più semplice e meno costoso, perché porta all'identificazione di un numero di varianti decisamente inferiori (**Tabella 2**).

Target	Basi del DNA presenti nel target	Copertura mediana richiesta	Basi del DNA da sequenziare	Varianti attese	Costo minimo della sola NGS**
WGS (genoma)	3.100.000.000	30x	> 120 Gb [*]	3.000.000	2.500 €
WES (esoma)	50.000.000	100x	10 Gb	30.000	900 €
Pannello esteso	1.500.000	200x	1 Gb	1.000	400 €
Pannello focalizzato	50.000	300x	0.05 Gb	30	300 €

* Gb = 1 miliardo di nucleotidi sequenziati

** I costi sono il minimo sul mercato e si riferiscono ai materiali per il solo sequenziamento senza considerare i costi accessori delle validazioni con metodica Sanger, l'analisi bioinformatica dei dati e lo studio di altri parenti

Tabella 2. Dimensione e costi degli studi NGS in base al target (Nigro 2015).

Poiché la maggioranza delle malattie mendeliane è dovuta a varianti nelle sequenze codificanti o nei siti di *splicing* il sequenziamento dell'esoma ha la potenzialità di identificare la maggioranza delle varianti che causano malattia. Il sequenziamento dell'esoma rappresenta un avanzamento importante nello studio delle malattie mendeliane [14], soprattutto per quelle nelle quali vi è un limitato numero di soggetti affetti per famiglia [15] e nelle quali vi è una significativa eterogeneità genetica e fenotipica [14, 16]: in questi casi infatti le tecniche tradizionali di *linkage* si sono rivelate in gran parte inefficaci, soprattutto per la difficoltà di reclutare per lo studio un numero adeguato di pazienti e di pedigree sufficientemente estesi.

Il sequenziamento dell'esoma si basa sull'applicazione del sequenziamento NGS a regioni d'interesse del genoma 'catturate' e separate dal resto del genoma mediante frammentazione e successiva amplificazione di queste regioni. Se si volesse utilizzare l'approccio tradizionale di

sequenziamento di Sanger per raggiungere lo stesso obiettivo ottenuto mediante NGS, sarebbe necessario disegnare *primers* per PCR per i circa 180,000 esoni del genoma umano. Quest'approccio sarebbe inevitabilmente molto più costoso ed estremamente impegnativo in termini di tempo e lavoro. Le metodiche di 'cattura' che hanno trovato maggiore applicazione nel sequenziamento dell'esoma sono: l'ibridazione in fase solida [17–19] e l'ibridazione in fase liquida [20]. Per quanto riguarda la prima, la sua applicazione al sequenziamento dell'esoma è possibile grazie alla realizzazione di chip o array, ossia matrici solide sulle quali vengono adese sonde di DNA complementari alle sequenze esoniche. L'applicazione sul supporto del DNA genomico opportunamente frammentato e successivi cicli di ibridazione e lavaggio consentono di arricchire il campione nel contenuto delle porzioni codificanti del genoma. L'ibridazione in fase liquida ricalca la precedente, con la differenza che le sonde non sono attaccate a un supporto solido, ma sono complessate con biotina. Dopo il passaggio di ibridazione, le sonde (con il DNA complementare legato) vengono legate da molecole di streptavidina coniugata a biglie magnetiche e, dopo opportuni lavaggi, il DNA d'interesse viene eluito. Per questo metodo vengono utilizzate sia sonde a RNA che a DNA. La copertura effettiva delle regioni d'interesse, ottenuta con le sonde selezionate, è un elemento cruciale per il successo dell'intero processo. Purtroppo la cattura dell'esoma non è mai completa (in media è pari a circa l'85%) e questo può determinare il mancato rilevamento di mutazioni patogeniche, se queste cadono in esoni non catturati. Bisogna però sottolineare che, per quanto diretta, questa ricerca della/e mutazione/i causativa/e non è semplice; basti pensare che il sequenziamento dell'esoma di un individuo sano riscontra generalmente da 20,000 a 25,000 varianti di cui circa 10,000 sono non-sinonime e quindi potenzialmente causa di patologia genetica. Questi numeri spiegano bene la grossa quantità di lavoro che è richiesta per l'interpretazione dei dati di sequenziamento genomico.

Mutazioni in regioni non-codificanti però sfuggono al sequenziamento dell'esoma, mentre possono comunque essere identificate mediante sequenziamento dell'intero genoma. Sia il

sequenziamento dell'esoma che di tutto il genoma, tuttavia, non permettono di identificare microdelezioni e microduplicazioni che possono essere escluse mediante altre metodiche (*i.e.*, array CGH) (**Figura 2 e Tabella 2**).

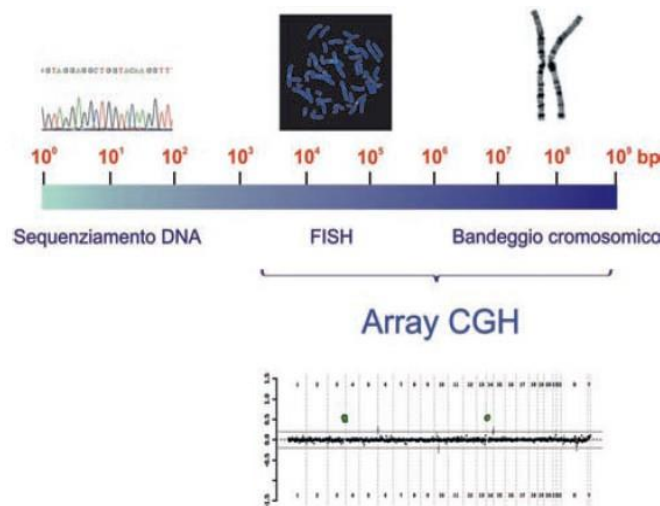


Figura 2. Potere di risoluzione delle indagini per lo studio del genoma umano dal cariotipo al sequenziamento di Sanger (Brunetti Pierri).

Applicazioni cliniche della NGS

La NGS ha un notevole valore nello studio di condizioni pediatriche, soprattutto genetiche e metaboliche, che spesso sono caratterizzate da grande eterogeneità clinica e molecolare [21]. Le tecnologie di NGS hanno conosciuto un rapido sviluppo nell'ultimo decennio che ha portato ad una loro ampia diffusione e soprattutto ad una riduzione dei tempi necessari per l'analisi e dei costi. Nella pratica clinica tre sono stati i principali usi della tecnologia NGS: “*targeted sequencing*” (TS), WES e WGS. Ciascuna di queste applicazioni presenta delle caratteristiche che sono riassunte nella **Tabella 2** [22].

Queste metodiche hanno trovato importanti applicazioni in molti campi della medicina ed hanno avuto un grosso successo nell'identificazione dei difetti genetici di moltissime malattie. Il

sequenziamento di pannelli di geni, noto anche come TS, consiste nel sequenziamento mirato ad un numero più o meno ampio di geni. Il TS è principalmente utilizzato per analizzare gruppi di geni responsabili di uno specifico fenotipo, come ad esempio l'epilessia, le malattie metaboliche ereditarie o gruppi di esse, la disabilità intellettiva. L'uso di pannelli di pochi geni non ha sostanziali differenze con la diagnostica tradizionale nell'interpretazione dei risultati e non pone significative problematiche etiche a contrario del WES. Ad esempio la tecnica NGS, con approccio TS, è utilizzata per la diagnosi delle iperfenilalaninemia tramite l'analisi dei geni *PAH*, *PTS*, *QDPR*, *GCHI* e *PCBD1* [23], collagenopatie o osteogenesi imperfecta [24], per condizioni geneticamente molto eterogenee come le ciliopatie [25], malattie neuromuscolari [26] ed anemia di Fanconi [27]. Esistono inoltre pannelli di geni codificanti proteine coinvolte nelle stesse vie metaboliche, come il pannello Lysoplex che include 891 geni coinvolti in malattie da accumulo lisosomiale, autofagia e *pathway* endocitico [28], pannelli di geni responsabili di malattie mitocondriali, malattie perossisomiali o difetti congeniti di glicosilazione.

Un approccio più estensivo del TS è il “*clinical exome sequencing*” o esoma clinico noto anche come “*Mendelioma*” [29, 30]. Con questo pannello il sequenziamento è mirato ad oggi fino a 5,228 geni responsabili di malattie mendeliane [31]. Il vantaggio di TS è quello di permettere di ottenere una diagnosi senza dover ricorrere a indagini diagnostiche invasive (*e.g.*, biopsia muscolare nel sospetto di malattie mitocondriali), costose (esami neuroradiologici) o difficilmente accessibili (dosaggi enzimatici). Uno dei vantaggi del TS rispetto al WES è la maggiore facilità di interpretazione dei risultati in quanto questa analisi risulta in un numero inferiore di varianti da interpretare e non richiede sempre l'analisi contestuale del DNA di entrambi i genitori con ovvie ripercussioni sui costi e sulla rapidità dell'esame. Il vantaggio di costo dei TS rispetto a WES e WGS, si è progressivamente assottigliato nel corso del tempo ed è atteso che nei prossimi anni si ridurrà ulteriormente. Nel caso in cui vi sia un sospetto preciso di sindrome genetica dovuta a più geni, il TS è comunque la metodica

da preferire, almeno allo stato attuale. La sensibilità dei vari approcci è funzione del grado di copertura (*coverage*) delle sequenze indagate, ovvero del numero di “letture” (*reads*) dei segmenti di DNA genomico esaminati. Avendo un *coverage* elevato, l’analisi TS ha una sensibilità maggiore rispetto alla WES e WGS.

L’esoma rappresenta solo una piccola porzione del genoma umano pari a circa 1-1.5% o 30 Mb, per un totale di circa 180,000 esoni. Circa 85% delle mutazioni causa delle patologie mendeliane interessa le sequenze esoniche [8]. La copertura effettiva delle regioni d’interesse, ottenuta con sonde selezionate, è un elemento cruciale per il successo dell’intero processo. In generale, per il WES sarebbe auspicabile avere un *coverage* medio di almeno di 100X perché questo grado di copertura permette di identificare in maniera attendibile mutazioni in eterozigosi che sono soggette a fenomeni di amplificazione preferenziale di uno dei due alleli. La maggior parte degli studi hanno riportato che la risoluzione diagnostica del WES è variabile dal 25% al 50% [32–35]. Il tasso di successo diagnostico varia molto in base ai criteri di selezione dei pazienti, infatti nel caso in cui vengano analizzati pazienti con maggior omogeneità fenotipica, i tassi diagnostici della WES sono risultati molto più alti: ad esempio 48.8 % per patologie neuromuscolari [36] e 85% per immunodeficienze primarie [37]. Altri studi hanno evidenziato la necessità di un *coverage* di almeno 20X per una accurata genotipizzazione e hanno stimato una sensibilità e una specificità del 100% e un tasso di riproducibilità intra- ed inter-operatore del 99% (dato calcolato su un gruppo ristretto di 18 pazienti) validando così ulteriormente questa tecnica nella pratica clinica [38].

Il WES offre la possibilità di individuare nuovi geni malattia e dal 2009 ha portato alla scoperta di almeno 200 nuove malattie genetiche all’anno ed è atteso che questo numero continui a crescere nei prossimi anni [39–42]. L’importanza di questa tecnica risiede non solo nella possibilità di individuare nuovi geni malattia o nuove correlazioni tra nuovi quadri fenotipici e geni malattia già conosciuti, ma anche nella possibilità, una volta individuata la causa, di modificare, in alcuni casi

selezionati, la terapia e la prognosi dei pazienti. Ripercussioni terapeutiche del WES si ottengono soprattutto nel caso di patologie metaboliche [43]. In uno studio su un gruppo di disordini neuro-metabolici, nel 44% dei casi il risultato del WES ha portato a cambiamenti della terapia in atto con evidenti miglioramenti prognostici [44]. Va considerato che la mancata identificazione di un difetto genetico con il WES non implica che nessuna variante causativa è presente nel DNA del paziente ed è possibile che varianti patogeniche siano presenti in geni con minore copertura o in sequenze regolatorie non studiate con il WES. Ampie delezioni e duplicazioni, specie in eterozigosi, sono difficilmente identificabili mediante NGS. Pertanto è consigliabile far precedere od associare tecniche quali l'array-CGH. Inoltre lunghe sequenze ripetute, come quelle osservabili nella sindrome dell'X fragile o in altre malattie da espansione di triplette, sono ugualmente non valutabili. Infine, delezioni in sequenze ripetute come quelle responsabili dell'atrofia muscolare spinale (SMA) non sono valutabili mediante WES.

Il WGS ha avuto finora una limitata applicazione in ambito clinico e diagnostico, a causa dei costi elevati (circa 2,500-3,000 euro per genoma sequenziato), della difficoltà interpretativa e della difficile gestione dell'enorme quantità di dati che esso genera. Rispetto al WES, il WGS identifica un numero 150 volte maggiore di varianti, localizzate in sequenze introniche e regolatorie dell'espressione genica (promotori ed *enhancers*) (**Tabella 3**) [45].

	Sequenziamento intero genoma	Sequenziamento esoma
Numero di varianti	3 milioni	20.000
Varianti esoniche	+	+
Varianti di splicing	+	+
Varianti introniche e promotori	+	-
Microdelezioni e microduplicazioni	-	-

Tabella 3. Principali differenze nelle varianti individuate tramite WES e WGS.

La vasta quantità di dati prodotta dalla metodica NGS inoltre comporta la necessità di avere una sufficiente quantità di computer *storage*. Infine i dati ottenuti pongono serie difficoltà di interpretazione per l'immensa quantità di dati e problematiche di natura etica per l'identificazione di varianti non ricercate con possibili conseguenze sulla salute [*incidental findings* - IF]. La probabilità di ottenere IF e Varianti di Significato Incerto (VUS) dipende dall'ampiezza della porzione di genoma interrogata. In generale, quanto più è ampia la sequenza analizzata, tanto maggiore è il rischio di riscontrare IF e VUS. Infine quanto più ampia la porzione di genoma analizzata, tanto maggiore è il volume di dati prodotti che richiede adeguate e costose piattaforme informatiche per l'archiviazione.

Interpretazione delle varianti esoniche

Il WES porta all'identificazione di migliaia di variazioni in ciascun paziente analizzato ed è necessario l'uso di opportuni filtri per isolare un ristretto numero di variazioni, tra le quali è presumibilmente contenuta la variazione patogena. L'interpretazione delle varianti presenti nell'esoma o genoma di un individuo richiede prima di tutto un accurato filtraggio. Una volta escluse le varianti che non determinano cambiamenti nella sequenza amminoacidica, il primo filtro è basato sulla frequenza con cui ciascuna variante è presente nella popolazione di controlli sani: tutte le varianti presenti nella popolazione con un'alta frequenza vengono considerate benigne. Per quest'operazione di filtro, vengono utilizzati i database che contengono tutte le varianti individuate in una popolazione controllo (ossia una popolazione di soggetti senza patologie conclamate) [46]. Il più esteso di questi database è il *Genome Aggregation Database* (gnomAD), che include dati di esoma di 125,748 individui e di genomi da 71,702 individui [47] (<https://gnomad.broadinstitute.org/>). Altri database sono HapMap e dbSNP [8]. Se le varianti riscontrate nel soggetto in studio sono presenti in questi database, la probabilità che siano patogene è bassa, considerata la rarità delle malattie mendeliane. Anche se la frequenza di popolazione riveste un ruolo fondamentale, non è sufficiente come unico

filtro perché in media riduce il numero di varianti candidate a circa duecento [48]. Un ulteriore filtro è applicato usando strumenti come Polyphen2 [49], SIFT [50] and PhyloP [51] per la predizione della patogenicità delle varianti in base al grado di conservazione della sequenza nella scala evolutiva e le possibili conseguenze sulla struttura secondaria e terziaria della proteina e di proteine della stessa famiglia. Questi algoritmi si basano sull'assunzione che i residui amminoacidici maggiormente conservati nell'evoluzione sono verosimilmente i più importanti dal punto di vista funzionale e mutazioni che interessano questi residui sono più probabilmente patogeniche. Sfortunatamente, queste metodiche non hanno alta specificità e sensibilità [48] e recentemente sono stati sviluppati altri metodi tra cui il *Combined Annotation-Depended Depletion (CADD)* che è un ulteriore strumento per dare un punteggio di patogenicità a possibili varianti del singolo nucleotide o di piccole inserzioni e delezioni (indel), calcolato su multiple variabili [46, 52]

Un'altra importante operazione di filtro consiste nell'escludere le variazioni sinonime, avendo queste un potenziale patogenetico molto più basso rispetto a variazioni *missense*, *nonsense*, ad inserzioni o delezioni *frameshift* o a mutazioni che alterano lo *splicing*. Infine per filtrare ulteriormente i risultati ottenuti dal sequenziamento dell'esoma è utile considerare la modalità di trasmissione della patologia. Infatti qualora si tratti di una malattia dominante, l'analisi si può limitare alle variazioni presenti in eterozigosi. Per le patologie autosomico recessive invece vanno considerate non solo le variazioni presenti in omozigosi ma anche quelle in eterozigosi composta, qualora due o più di queste colpiscano lo stesso gene. La segregazione delle varianti nella famiglia è uno strumento prezioso per filtrare le varianti candidate. L'analisi contestuale dei genitori del probando (*trio*) o di *quartet* con un altro figlio (affetto o non affetto) è un altro importante approccio che consente di filtrare varianti potenzialmente patogene da quelle benigne [29]. Nel caso di varianti in geni non precedentemente associate a malattie sono necessari ulteriori approfondimenti per provarne la patogenicità come un saggio funzionale che può fornire una forte evidenza sulla patogenicità ma richiede tempo, risorse e

lo sviluppo di un progetto di ricerca. Un'altra prova molto importante è l'identificazione di altri pazienti con mutazioni nello stesso gene ed un fenotipo simile. Questa seconda strategia è resa possibile da iniziative di *data sharing* per malattie rare come DECIPHER [53], Café Variome [54], GeneMatcher [55], RD-connect [56] e PhenomeCentral [57]. “Matchmaker Exchange” è un'iniziativa recente che ha lo scopo di integrare le informazioni di tutti questi *databases* [58]. Il successo di questa strategia è legato alla qualità ed accuratezza della descrizione del fenotipo. Per questa ragione sono stati sviluppati diversi programmi che permettono la condivisione di dati clinici. Lo *human phenotype ontology* (HPO) che contiene 250,000 annotazioni fenotipiche è il programma più usato [59].

Una volta completato il processo di filtro, le varianti individuate possono essere classificate in tre gruppi principali [22]:

1. Varianti causative: sono quelle varianti alle quali è possibile attribuire con certezza un ruolo patogenetico. Si tratta di varianti già associate a malattie.
2. Varianti responsabili di fenotipi non collegati al quesito clinico o anche detti IF [60]. Il riscontro di queste varianti non è inatteso perché è ben noto che ogni individuo ha una discreta probabilità di essere portatore di un numero non trascurabile di varianti genetiche associate a potenziali conseguenze cliniche (ognuno di noi è portatore di circa 3-5 alleli associati a patologie autosomico recessive) [61].
3. Varianti con effetti funzionali e clinici non definiti (VUS).

L'American College of Medical Genetics and Genomics (ACMG) nel 2013 ha fornito delle linee guida per la refertazione delle varianti riscontrate con NGS. Essi raccomandano l'uso di una terminologia standard per le varianti: “patogenica”, “probabilmente patogenica”, “di incerto significato”, “probabilmente benigna”, “benigna” [60].

Valutazioni cliniche pre- e post- analisi genomica e problemi etici

Rispetto alle metodiche tradizionali, le metodiche di analisi genomica individuano un numero di varianti, anche non correlate con il quesito clinico e che possono interessare la salute futura di chi ha effettuato l'analisi. Prima di effettuare l'analisi genomica deve essere fornita al paziente o ai genitori un'adeguata informazione per permettere la comprensione degli obiettivi dell'esame, della sua tempistica e dei possibili risultati. In questo contesto vanno descritte alcune importanti implicazioni dell'esame relativo agli IF. Le IF possono essere classificate in **(Tabella 4)** [62]:

- 1) IF clinicamente utili;
- 2) IF clinicamente valide ma non utili;
- 3) IF clinicamente non utili e non validi.

Il consultando dovrà poi essere supportato nella decisione autonoma di ricevere o meno informazioni sulle prime due classi. Il terzo gruppo non dovrebbe essere oggetto né di comunicazione né di refertazione. Il riscontro di IF può rivelare condizioni nuove e inaspettate per il paziente e la sua famiglia.

La consulenza post-test pertanto richiede la comunicazione e spiegazione dei risultati con particolare attenzione alle VUS e alle varianti associate a possibile rischio di malattie con esordio tardivo. La comunicazione dei risultati è tutt'ora oggetto di dibattito [63–66]. L'ACMG raccomanda di estendere l'analisi NGS a un set di 57 geni responsabili di 24 patologie per le quali si ravvisa un'utilità clinica certa, in quanto associate alla disponibilità di terapie e mezzi di prevenzione efficaci [78]. Questa posizione è stata oggetto di dibattito e non è condivisa da tutti gli esperti. Inoltre, alcune informazioni derivate dalle analisi WES o WGS riguardano patologie per le quali, non esistendo trattamenti validi (*e.g.*, condizioni neurodegenerative), la loro individuazione è di dubbia utilità, se non addirittura potenzialmente dannosa. Al contrario informazioni di interesse farmacogenetico, che riguardano varianti polimorfiche predittive dell'efficacia terapeutica o di effetti avversi di specifici farmaci, potrebbero essere utili per i pazienti.

Utilità clinica	Si	No	No
Validità clinica	Si	Si	No
Comunicazione al paziente	Si (se dato consenso)	No (<i>si su consenso specifico</i> dopo adeguata consulenza)	No
Esempi di geni e condizioni cliniche	<ul style="list-style-type: none"> -<i>BRCA1/BRCA1</i> (carcinoma mammario/ovarico ereditario) -<i>MSH2/MLH1/MSH6</i> (sindrome di Lynch) -<i>KCNQ1, KCNH2, SCN5A</i> (sindrome del QT lungo) -<i>OTC</i> (deficit di ornitina transcarbamilasi) -Condizione di eterozigosi per patologie autosomiche recessive relativamente frequenti -<i>rs4149056</i> genotipo a bassa attività (CC) → rischio elevato di miopatia indotta da statine (dose inferiore o farmaco alternativo) -<i>CYP2D6 poor metabolizers</i> (*4/*4) → evitare uso di codeina 	<ul style="list-style-type: none"> -<i>APOE</i> (malattia di Alzheimer) -Otosclerosi -Deficit alfa1-antitripsina 	<ul style="list-style-type: none"> -VUS in regioni non comprese nei target specifici del test -Varianti polimorfiche di suscettibilità associate a rischi di patologia <2X rispetto alla popolazione generale

Tabella 4. *Classificazione degli IF secondo utilità e validità clinica (Documento commissione SIGU, NGS).*

Per quanto riguarda la gestione dei dati, controversie di tipo etico vengono sollevate sulla conservazione, rianalisi e condivisione dei dati. Altri punti di criticità sotto il profilo etico sono il tempo di conservazione di queste informazioni e l'accessibilità dei dati. Molti sono i quesiti irrisolti: “I pazienti dovrebbe avere un accesso libero e diretto a questi dati? I minori dovrebbero poter accedere ai dati generati anni prima, una volta raggiunta la maggiore età? Bisognerebbe ricontattare i pazienti prima di cercare nuove diagnosi o nuove opzioni terapeutiche non disponibili in precedenza?” [67].

Programmi per malattie “senza diagnosi”

Nelle ultime decadi si è raggiunta una sempre maggiore consapevolezza delle difficoltà nel raggiungimento di una diagnosi definitiva in pazienti con fenotipo complesso. Per questo motivo, nel 2008 l'Office of Rare Disease Research del National Institute of Health (NIH) degli USA ha stanziato consistenti fondi per l'approfondimento delle patologie genetiche complesse e nel maggio 2008 è stato lanciato il *Undiagnosed Disease Program* (UDP) [68] che ha come obiettivi primari il raggiungimento della diagnosi per i pazienti senza diagnosi e la scoperta di nuove sindromi e dei sottostanti meccanismi molecolari. Nel corso degli anni, il progetto, nato come programma intramurale del NIH, si è espanso sul territorio nazionale, evolvendosi in UDN (*undiagnosed disease network*) (**Figura 3**). Nel corso dei primi 7 anni di attività il progetto UDP ha ricevuto 2,954 richieste e ne ha valutata 863 per una media di 120-130 pazienti/anno [68, 69]. I pazienti ammessi al programma sono stati sottoposti ad una valutazione clinica completa presso centri specialistici per una dettagliata e standardizzata analisi del fenotipo utilizzando i termini HPO [70].

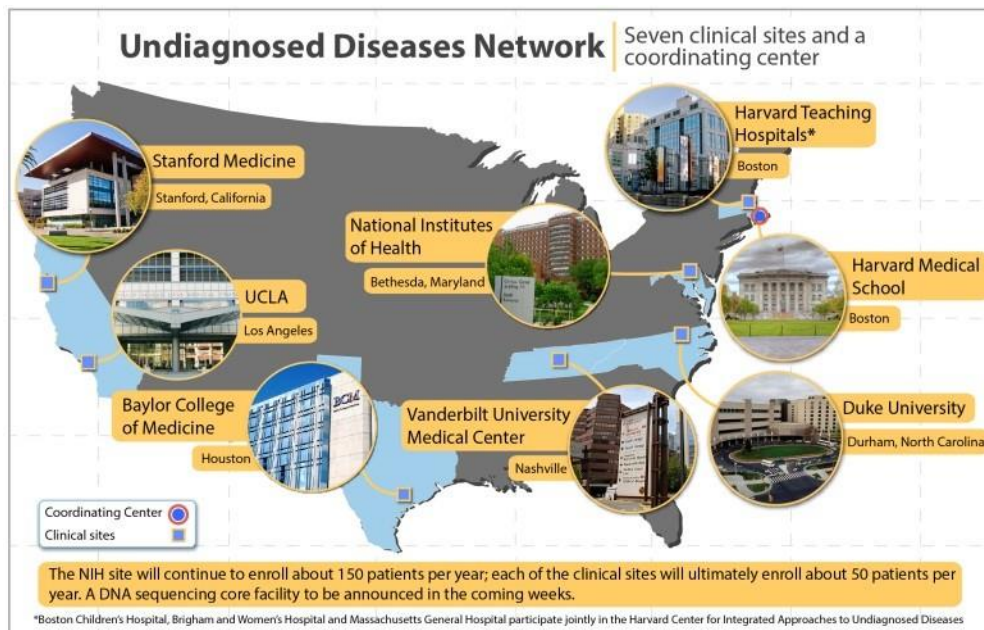


Figura 3. Undiagnosed Diseases Network (Healthline news).

Sulla base del modello americano sono stati sviluppati molti altri programmi nazionali per le malattie senza diagnosi come in Canada, Australia, Giappone, Italia e altri. Successivamente è stato anche stabilito una rete internazionale che coinvolge questi vari programmi nazionali con gli obiettivi primari di:

- migliorare il livello delle diagnosi del paziente con malattie non diagnosticate;
- facilitare la ricerca per queste patologie grazie alla raccolta standardizzata e condivisione dei dati;
- creare una comunità di ricerca integrata e collaborativa tra i vari centri clinici [71].

Nel 2015 in Italia è cominciato il programma "Malattie senza Diagnosi" della Fondazione Telethon (TUDP) (<http://www.telethon.it/cosa-facciamo/malattie-senza-diagnosi>) che ha l'obiettivo di studiare pazienti in età pediatrica (età <18 anni) con fenotipo particolarmente complesso e una lunga storia di studi metabolici e genetici non conclusivi da un punto di vista diagnostico. Il

programma coinvolge il TIGEM, dove vengono effettuate le analisi NGS e molteplici centri nazionali con elevata esperienza nella genetica clinica e nella pediatria (**Figura 4**). Altri centri e altri medici al di fuori di questi due centri hanno la possibilità di candidare i propri pazienti per il programma o riferirli per una valutazione presso questi centri.

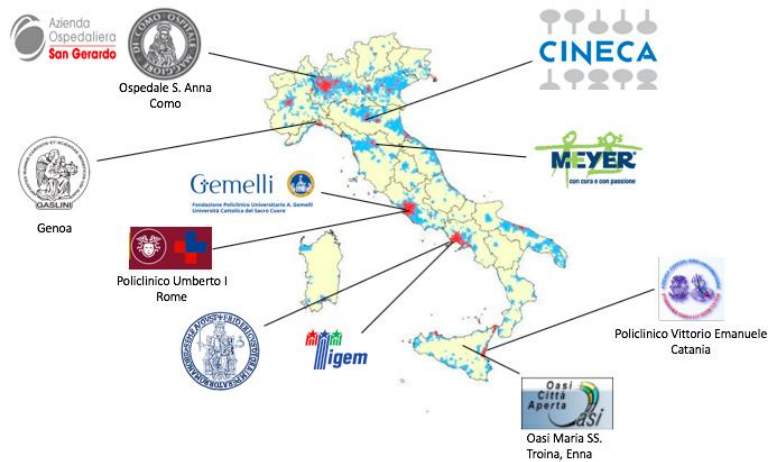


Figura 4. Centri di riferimento progetto Telethon “Malattie senza diagnosi”.

L’attività di questa tesi è stata svolta nell’ambito di uno dei centri clinici partecipanti a questo progetto (Azienda Ospedaliera Federico II di Napoli).

Obiettivi dello studio

Lo studio ha l'obiettivo principale è quello di applicare la tecnologia WES a pazienti “senza diagnosi” selezionati attraverso una caratterizzazione clinica e diagnostica ampia e standardizzata.

I centri clinici partecipanti allo studio, ben distribuiti geograficamente sul territorio italiano nelle aree di più elevata densità di popolazione (**Figura 4**), selezionano i pazienti da sottoporre a WES in base alla severità ed interessamento multi-sistemico del quadro clinico non inquadrabile in nessuna sindrome genetica nota, età di esordio e storia familiare positiva. Priorità verrà data ai pazienti ampiamente studiati nei quali siano state escluse anomalie cromosomiche o di geni responsabili di condizioni geniche note. I pazienti sono riferiti per WES anche da altri centri italiani, pediatri o medici di famiglia.

Il programma TUDP ha i seguenti tre obiettivi principali:

- L'individuazione di criteri standardizzati e condivisi per la scelta dei pazienti candidati da sottoporre a WES. I pazienti sono ordinati in base a criteri pre-stabiliti quali: persistente mancanza di una diagnosi eziologica, severità e complessità della condizione clinica, età di esordio e storia familiare. Maggiore priorità è data ai casi risultati negativi ad estensiva analisi per la presenza di anomalie cromosomiche o difetti genici in grado di spiegare il quadro in esame (“pazienti con malattia senza diagnosi”). Con questa metodica sono stati selezionati, nel corso di tre anni, circa 500 casi con maggiore probabilità di rappresentare condizioni genetiche ancora sconosciute. I pazienti che partecipano allo studio, ma non sono selezionati per il WES, possono, comunque beneficiare di una rivalutazione clinica da parte di un gruppo di esperti, che possono suggerire una diagnosi fino a quel momento non considerata.

- Identificare la causa genetica della condizione selezionata usando il WES. Per questa analisi vengono utilizzati i campioni del *trio* (padre-madre-proband) che permette una selezione ed

identificazione più efficiente delle varianti candidate grazie allo studio della segregazione delle varianti nella famiglia. In casi selezionati, l'analisi è ampliata anche ad uno o più familiari del caso indice.

- La condivisione dei risultati con altri progetti internazionali simili, quali *NIH Undiagnosed Disease Program*, *Finding of Rare Disease Gene* (FORGE) in Canada e altri per la ricerca di altri pazienti con mutazioni nello stesso gene e più approfondita caratterizzazione clinica del fenotipo causato da mutazioni in un determinato gene.

Pazienti e Metodi

Criteri di inclusione ed esclusione dello studio

Sono stati candidati per lo studio tutti i soggetti da 0 a 18 anni affetti da sindromi complesse che presentavano anomalie congenite maggiori e minori, alterazioni della crescita, disabilità intellettuale di eziologia sconosciuta nonostante un esteso work-up diagnostico-strumentale (compreso l'array-CGH). Sono stati esclusi i soggetti in cui è sospettata una specifica sindrome genetica, in cui le manifestazioni cliniche siano ascrivibili a fattori ambientali (teratogeni e infezioni perinatali) o con difetti isolati (*i.e.*, retinopatia, sordità, disabilità intellettiva). Sono anche stati esclusi soggetti per i quali non è stato possibile ottenere i campioni del trio familiare.

Per i pazienti riferiti da altri centri o medici è stata effettuata una valutazione in due step: nella prima fase il medico del paziente ha compilato un modulo *online* (web-form) che includeva domande dettagliate sulla storia familiare e medica del paziente, crescita, sviluppo, dismorfismi e anomalie d'organo, esami strumentali ed esami clinici e genetici già effettuati. I pazienti inseriti nella web-form sono stati divisi in tre macro-gruppi:

1. inadeguata documentazione clinica e conseguente inleggibilità per lo studio;

2. sospetto di una specifica diagnosi che richiede ulteriori approfondimenti. Il paziente è rinviato al medico di riferimento con i necessari suggerimenti;
3. il caso è eleggibile.

I pazienti considerati eleggibili sono stati valutati in uno dei centri affiliati al progetto in base della maggiore vicinanza geografica. I dati del paziente comprendenti la storia familiare e l'albero genealogico, l'anamnesi gestazionale, perinatale e neonatale, le caratteristiche cliniche inclusive degli elementi dismorfici accompagnate da foto del paziente, dati auxologici ed altri elementi clinici, indagini strumentali (RMN-encefalo, eco-cardio, eco-addome, etc) e precedenti risultati di test diagnostici (array-CGH, test genetici) sono stati raccolti su una piattaforma telematica dedicata e standardizzata: "PhenomeCentral" (<https://www.phenomecentral.org/>). Tale piattaforma permette sia di standardizzare la procedura di raccolta dei dati clinici dei pazienti sia di condividere i dati con i centri clinici coinvolti e con altri centri clinici coinvolti in simili iniziative per l'identificazione di altri casi con fenotipo simili e/o difetti negli stessi geni eventualmente identificati dal WES. I pazienti inseriti nella piattaforma sono stati discussi con gli altri centri partner del progetto durante delle riunioni a cadenza mensile "Clinical Plenary Meeting". Le riunioni avevano lo scopo di discutere la completezza dell'analisi clinica, di proporre eventuali ipotesi diagnostiche, e di decidere dell'inclusione del caso nel progetto nel caso in cui il quadro clinico non è inquadrabile in una sindrome genetica nota. I pazienti che hanno superato anche quest'ultima fase di selezione sono stati poi richiamati per il prelievo di sangue, che è stato poi inviato al laboratorio per eseguire il WES. L'analisi WES è stata condotta su almeno tre soggetti (probando, madre e padre).

WES ed analisi bioinformatica

Il DNA genomico è stato estratto secondo tecniche standard da sangue periferico dal probando e dai genitori (*trio*) ed eventualmente dai fratelli non affetti (*quartet*, *quintet*). La scelta di analizzare *trio* e non *singleton* (analisi inclusive solo del probando) è stata consigliata da precedenti studi che hanno osservato che la strategia basata sull'analisi unica del probando è meno efficace rispetto a quella del *trio* e comporta costi maggiori in quanto richiede un numero significativamente più alto di analisi di conferma mediante tecnologia tradizionale Sanger [29].

La NGS è stata praticata usando tecniche ad alta copertura (HC-WES) mediante kit Illumina HiSeq1000. HC-WES si basa su più recenti tecnologie e fornisce un numero di *reads* con una copertura di 100x su circa l'80-90% dei targets e circa 30x sul 95-99% dei targets. Il WES è stato eseguito usando "the SureSelect target capture kit" (Agilent). I risultati ottenuti sono stati valutati usando una sequenza di operazioni pre-stabilita: 1. controllo di qualità delle *reads* al fine di selezionare per l'allineamento con il genoma le *reads* che al contrario passano il controllo di qualità (FastQ format) [72]; 2. allineamento tramite Burrows-Wheeler Aligner; 3. rimozione delle *reads* presenti in duplicato; 4. controllo di qualità per scartare le *reads* di bassa qualità.

Le varianti sono state raggruppate in quattro gruppi: *de novo*, omozigosi/emizigosi, eterozigote composto ed eterozigote singolo. Questa quarta variante (variante ereditata da genitore sano) non viene considerata in prima battuta, escludendo così inizialmente meccanismi di penetranza incompleta.

L'interpretazione bioinformatica delle varianti identificate si basa su prioritizzazione sulla base della frequenza in database di popolazioni controllo, sul modello di ereditarietà della patologia e sulla segregazione nei genitori, sulla patogenicità della variante, sulla compatibilità fenotipica rispetto alla verosimile funzione del gene e della letteratura disponibile.

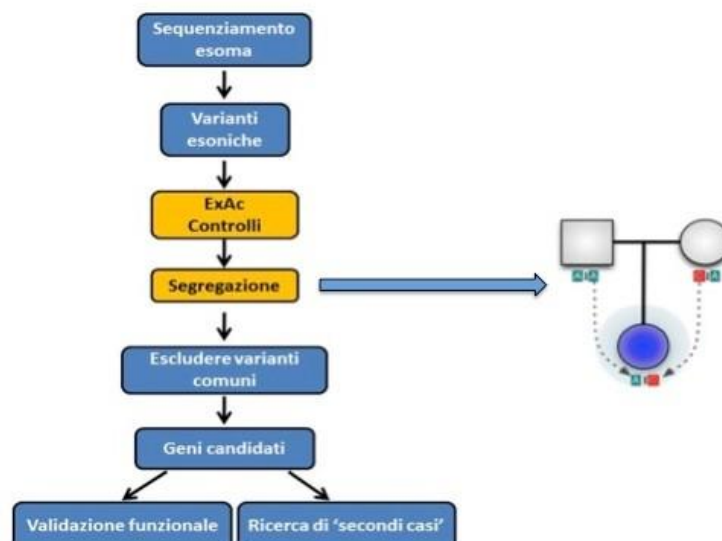


Figura 5. Percorso di analisi delle varianti individuate con metodica WES nel progetto TUDP.

Solo le varianti rare (frequenza allelica minore dell'1% nei nostri controlli WES o minore del 0.5 % in GnomAD) sono state considerate come possibili candidate (**Figura 5**). GnomAD permette di identificare la frequenza di una data variante in varie popolazioni raggruppate sulla base della provenienza geografica (Europei non finlandesi, Finlandesi, Africani, Asiatici dell'Est, Asiatici Sud, Latino, Altri). La presenza di una variante in omozigosi in questo database verosimilmente annulla la patogenicità della stessa variante. Altra funzione disponibile consultando il database GnomAD è la probabilità di intolleranza alle mutazioni con perdita di funzione o mutazioni *loss-of-function* (*probability of being loss-of-function intolerant*, pLI), che viene calcolata sulla base del numero atteso di mutazioni *loss of function* verso il numero realmente osservato. Quanto più la probabilità è vicina a 1 tanto più il gene appare essere aploinsufficiente quindi intollerante alle *loss-of-function* dato, che suggerisce che un dato gene potrebbe essere un gene malattia. Le varianti sono state poi valutate sulla base delle loro predette conseguenze funzionali. Varianti

nonsense o *frameshift* suggeriscono un effetto più severo sulla funzione della proteina rispetto a mutazioni missenso e pertanto sono più probabilmente patogeniche. Per varianti di incerto significato si è valutato l'effetto della variante sulla funzione della proteina attraverso vari gli strumenti bioinformatici di predizione discussi precedentemente (SIFT, Polyphen-2, CADD). Nella valutazione del potenziale ruolo patogenico delle varianti identificate ha un ruolo importante anche lo studio delle informazioni sulla funzione del gene, qualora queste siano disponibili. Ad esempio un gene implicato nella sinaptogenesi è un ottimo candidato se il fenotipo del paziente comprende la disabilità intellettiva o disturbo dello spettro autistico. Le varianti geniche selezionate sono state confermate mediante sequenziamento di Sanger.

Relazione causa-effetto tra la variante identificata da WES e fenotipo

La dimostrazione della patogenicità delle varianti geniche identificate si è basata sull'analisi di segregazione nella famiglia, sulla frequenza della variante nella popolazione, su saggi funzionali e sull'identificazioni di mutazioni nello stesso gene in pazienti con fenotipi simili [73]. L'esistenza di modelli murini *knock-out* per il gene che ricapitolano il fenotipo umano è stata un'altra informazione importante per la patogenicità. In caso di riscontro di mutazioni in geni implicati in *pathway* metabolici alterazioni dei livelli dei metaboliti è stato importante. Per validare la relazione di causalità tra il gene mutato e la malattia occorre soprattutto dimostrare un coinvolgimento dello stesso gene in almeno un secondo paziente con un quadro clinico sovrapponibile. Per facilitare l'identificazione di 'secondi casi' sono stati utilizzati diversi database: GeneMatcher, e PhenomeCentral. Genematcher è uno strumento web creato nel 2013, accessibile gratuitamente, sviluppato con l'obiettivo di individuare altri soggetti con fenotipi rari causati da varianti nello stesso gene malattia candidato. Inoltre PhenomeCentral è stato progettato

per consentire l'interazione tra clinici e ricercatori con l'obiettivo di connettere i geni ai fenotipi mendeliani e aumentare la conoscenza su malattie rare. I dati inseriti in GeneMatcher e in PhenomeCentral possono essere salvati in forma anonima allegando il contatto del medico che li inserisce [74].

I risultati dell'analisi bioinformatica sono stati suddivisi in 4 macro-gruppi: risultati conclusivi, potenzialmente conclusivi, inconclusivi e nuova scoperta (**Figura 6**):

1a. Risultati conclusivi: un risultato è stato conclusivo nel caso di identificazione di varianti bialleliche, *de novo* o legate al sesso in geni associati a una ereditarietà e fenotipo compatibile (match fenotipico). In particolare per disordini autosomici recessivi, questo risultato è stato ottenuto quando le varianti considerate si trovavano *in trans*, coinvolgevano un gene già precedentemente noto come causa malattia e sono già state precedentemente indagate per valutarne la patogenicità. In questo caso il clinico ha valutato la correlazione tra la presentazione clinica del paziente e quella riportata in letteratura. Per disordini con trasmissione autosomica dominante: questo risultato è stato ottenuto quando la singola variante identificata è molto rara (frequenza in GnomAD <0.001%), era *de novo* nel paziente o co-segregava con il fenotipo in altri parenti affetti della famiglia. La variante interessa un gene già noto essere causa di patologia e la sua patogenicità è già stata validata. In questo caso la conferma tramite sequenziamento Sanger è stata sufficiente a porre diagnosi definitiva.

1b. Risultati conclusivi- fenotipo atipico. Questo caso riguarda patologie recessive, dominanti o legate al sesso in pazienti con caratteristiche cliniche non associate al gene-malattia identificato. Il riconoscimento di nuove caratteristiche cliniche associate a una malattia genetica già nota permette l'espansione del fenotipo della malattia che è indispensabile sia per il riconoscimento di ulteriori pazienti, che per un giudizio diagnostico e prognostico completo.

2. Nuova scoperta- Disordini allelici. Ricadono in questo gruppo pazienti con: a) mutazioni in geni non ancora implicati in patologie umane e b) mutazioni in geni già noti per essere associati a patologie genetiche ben distinte sul piano fenotipico e patogenetico. L'identificazione di nuove malattie genetiche si basa sul riconoscimento di almeno due pazienti con fenotipo sovrapponibile e mutazioni nello stesso gene. Studi *in silico*, *in vitro* e *in vivo* sono necessari per ulteriore conferma della patogenicità. Nuove malattie alleliche vengono definite quando almeno due individui con mutazioni nella stesso dominio del gene presentano fenotipo sovrapponibile. Studi funzionali *in silico*, *in vitro* e *in vivo* sono necessari anche per l'identificazione di nuove malattie alleliche.

3. Potenzialmente conclusivi-geni candidati. Se la variante non è stata descritta la sua patogenicità deve essere validata tramite strumenti bioinformatici (Polyphen, SIFT e CADD) e/o esperimenti funzionali. In questa circostanza è stata effettuata attraverso database internazionali quali Genematcher e PhenomeCentral una ricerca di pazienti con mutazioni nel gene candidato per paragonarne i fenotipi. Quando le varianti potenzialmente patogeniche sono state riportate in una sola famiglia, il gene è descritto come candidato. Anche in questo caso studi *in silico*, *in vitro* ed *in vivo* sono necessari per la conferma della relazione causa-effetto.

4. Non conclusivo; nessuna variante candidata o troppe varianti potenzialmente causative vengono identificate. Questi casi potrebbero usufruire di periodiche rivalutazioni dell'analisi WES alla luce di una migliore definizione delle varianti polimorfiche o potrebbero essere candidati per studi più approfonditi quali il WGS, per l'individuazione di varianti causative nelle zone introniche, nei siti di splicing o nelle regioni regolatrici del DNA. Non sono ancora disponibili linee guida per la gestione di questi casi [75].

In caso di conferma della patogenicità della variante si è proceduto alla comunicazione dei risultati ai genitori dei pazienti da parte del medico che ha effettuato l'arruolamento dei pazienti nello studio.

Lo studio è stato approvato dal Comitato Etico della Facoltà di Medicina e Chirurgia dell'Università Federico II di Napoli (data di approvazione: Aprile 2016). Ogni paziente o tutore legale ha fornito consenso informato per la partecipazione allo studio e per la pubblicazione eventuale dei risultati, dei dati del paziente e delle foto. In quest'ottica, tutti i pazienti (o i loro tutori legali), se eleggibili per il programma TUDP, sono stati riconvocati dai centri coordinatori per una sessione informativa. In quest'occasione gli esperti hanno esposto e rivalutato in maniera critica tutti gli aspetti del programma (obiettivi, risultati attesi e possibili effetti avversi) e un opuscolo informativo è stato consegnato ai diretti interessati. Per i bambini e i pazienti minorenni, capaci di intendere, sono disponibili opuscoli semplificati e specifici per età. Al termine di questo incontro, dopo aver valutato la reale comprensione da parte del paziente dei diritti e doveri in quanto partecipanti dello studio, si è richiesta la volontà di partecipazione al progetto.

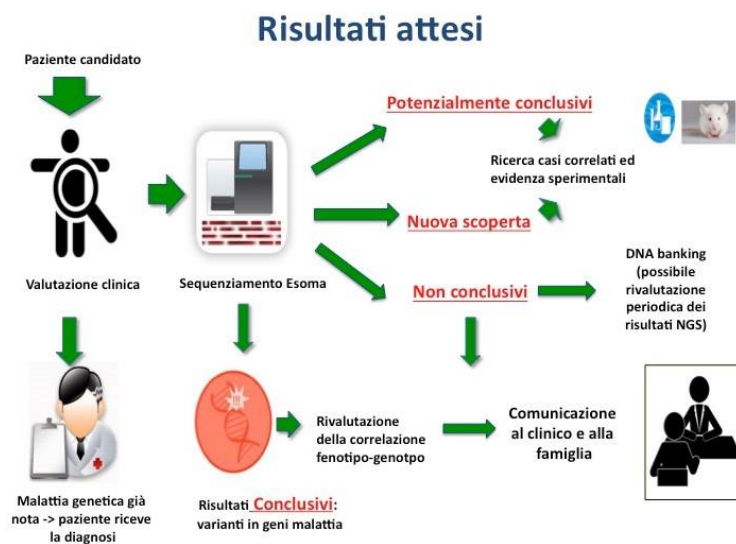


Figura 6. Classificazione dei risultati possibili dell'analisi delle varianti.

Risultati

Nello studio che è tutt'ora in corso sono stati arruolati presso il Centro di Napoli, Dipartimento ad Attività Integrata di Pediatria dell'Azienda Ospedaliera Universitaria Federico II di Napoli, ad oggi 163 pazienti. Tutti i pazienti erano già stati precedentemente sottoposti a estese valutazioni cliniche e diagnostiche che comprendevano tutte le metodiche di approfondimento clinico e genetico (*e.g.*, cariotipo, array CGH, sequenziamento Sanger di geni specifici) ed esami metabolici, senza raggiungere una diagnosi definitiva. Per tale ragione, questi pazienti venivano definiti come pazienti con “malattia senza diagnosi”. Dei 163 pazienti candidati, 152 hanno passato la prima fase di selezione e sono stati ritenuti idonei per l'analisi WES. Al momento della stesura di questa tesi, sono disponibili le analisi dei dati degli esomi per 116 famiglie. I casi con un risultato conclusivo sono stati 23, quelli con un risultato conclusivo con una espansione del fenotipo 20, mentre nuove condizioni genetiche sono state identificate in 7 famiglie (**Tabella 5, Figura 7**). In totale i casi conclusi sono stati 50 ed il tasso diagnostico in questo studio risulta essere al momento pari al 43% (50/116). I pazienti con un risultato potenzialmente conclusivo con identificazione di geni candidati sono 32 (**Tabella 5, Figura 7**). Per alcuni di questi geni sono in corso studi funzionali e ricerca di secondi casi mediante degli appositi strumenti di *matching*. Sia per i casi conclusivi che per quelli potenzialmente conclusivi le mutazioni/varianti *de novo* sono quelle più frequentemente riconosciute (25/50 e 15/32 rispettivamente per i casi conclusi e quelli potenzialmente conclusivi). I pazienti con un risultato non-conclusivo sono risultati 33 (**Figura 7**). I casi non-conclusivi sono rappresentati da quei pazienti per i quali l'analisi non ha individuato alcuna variante potenzialmente patogena o per i quali, nonostante la presenza di uno o più geni candidati, non è stato possibile delineare un quadro chiaro e definitivo di associazione genotipo-fenotipo. Questi pazienti potranno essere candidati

per ulteriori analisi, quali il sequenziamento WGS, o essere rianalizzati in futuro sulla base dei nuovi progressi scientifici e tecnologici.

Geni implicati in casi risolti			Geni candidati in pazienti potenzialmente conclusi		
Trasmissione <i>De novo</i>	Trasmissione Biallelica	Trasmissione legata al sesso	Trasmissione e <i>De novo</i>	Trasmissione e Biallelica	Trasmissione legata al sesso
<i>GRIN2B</i>	<i>RARS2</i>	<i>DDX3X</i> (<i>n=2</i>)	<i>SMARCD3</i>	<i>TEP1</i>	<i>FLNA</i>
<i>ASXL3</i>	<i>DST</i>	<i>APLS2</i>	<i>SIRT6</i>	<i>ADAT3</i>	<i>FRMPD4</i>
<i>SMAD4</i>	<i>WDR81</i>	<i>SLC16A2</i>	<i>RAB10</i>	<i>PHKB</i>	<i>IQSEC2</i>
<i>POLR2A</i>	<i>WDR62</i>	<i>PDHA1</i>	<i>HNRNPUL1</i>	<i>TEFM</i>	<i>MED14</i>
<i>KIF1A</i>	<i>KARS</i>	<i>RPL10 (n=3)</i>	<i>HOXA13</i>	<i>PDCD6IP</i>	<i>POLA1</i>
<i>CTNNB1</i>	<i>NDUEA2</i>	<i>TFE3</i>	<i>GNE</i>	<i>TRIM21</i>	<i>KDM6A</i>
<i>PURA</i>	<i>PRRT2/CYC</i>	<i>UBE2A</i>	<i>SMC3</i>	<i>ATG13</i>	
<i>NOTCH3</i>	<i>SMPD4</i>	<i>MSL3</i>	<i>SPTBN5</i>	<i>MACF1</i>	
<i>DHX37</i>	<i>RAB3GAPI</i> (<i>n=2</i>)	<i>CREBBP</i>	<i>POLR2B</i>	<i>FLAD1</i>	
<i>SMARCA2</i>	<i>B4GALT5</i>		<i>DYNC1LI1</i>	<i>RYR1</i>	
<i>SMARCA2</i>	<i>NMNAT1</i>		<i>SMARCC1</i>	<i>NUP210</i>	
<i>GLI2,</i>	<i>CENPE</i>		<i>ANKRD11</i>	<i>LRP2</i>	
<i>GRIN1</i>	<i>SLC17A5</i>		<i>EP400</i>		

<i>ATP6V1B2</i>			<i>LMNB2</i>		
<i>IRF2BPL</i>			<i>FOXP2</i>		
<i>AHDC1</i>					
<i>ADNP</i>					
<i>SMARCA4</i>					
<i>GABRB2</i>					
<i>GNB1</i>					
<i>CHD2</i>					
<i>HNRNPU</i>					
<i>CDK8</i>					
<i>EHMT1</i>					
<i>EXT1</i>					

Tabella 5. Lista dei geni malattia e geni candidati identificati nei pazienti reclutati nel Programma Malattie Senza Diagnosi Telethon. I geni associati a fenotipi atipici nei pazienti reclutati sono sottolineati. In grassetto indichiamo sia i nuovi geni-malattia sia le nuove malattie alleliche a geni già noti.

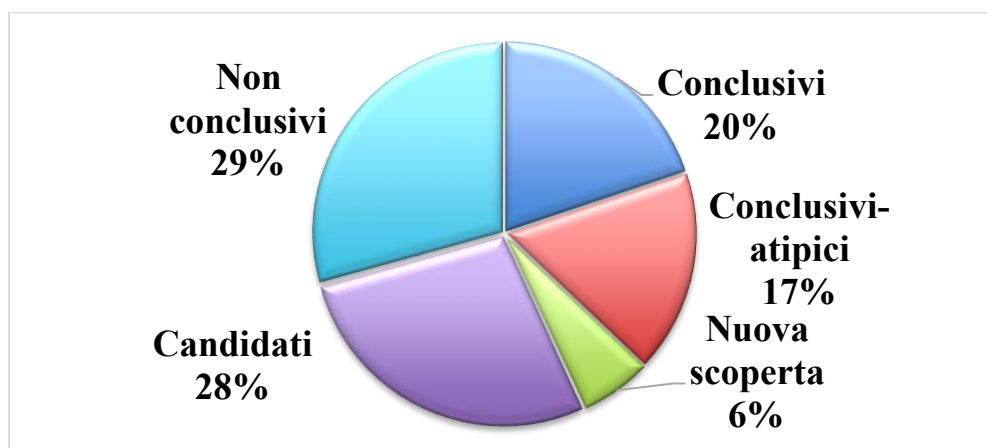


Figura 7. Distribuzione dei risultati del WED nei pazienti arruolati nel progetto TUDP .

Nelle sezioni successive presenteremo 3 casi esemplificativi in cui a seguito dell'odissea diagnostica è stata raggiunta una diagnosi genetica. In un caso è stato identificato il difetto in un gene recentemente associato a patologia (nuova scoperta), in altro caso è stato identificato un difetto genetico molto raro con una presentazione precoce ed atipica (caso risolto con presentazione atipica) ed in un terzo caso presentiamo un paziente con un risultato potenzialmente conclusivo con un forte gene candidato.

Caso 1- Nuova scoperta

Descrizione clinica e risultati del WES

Il paziente è un bambino di 7 anni è stato seguito da molti anni per ritardo dello sviluppo e microcefalia e non ha mai mostrato regressione psicomotoria.

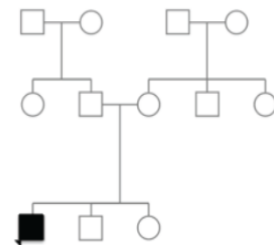
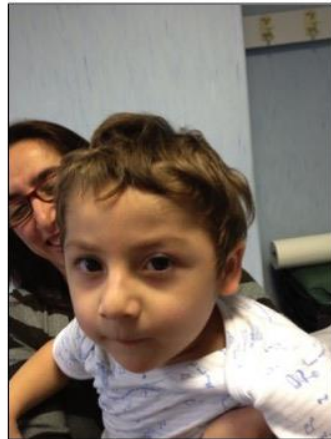


Figura 8. *Caratteristiche cliniche ed albero genealogico del paziente.*

Il bambino è nato da una coppia di genitori sani e non consanguinei con una storia familiare non contributiva (**Figura 8**). E' nato dopo 40 settimane di gravidanza normocondotta ed i fenomeni perinatali sono risultati nella norma con l'eccezione del riscontro nei primi giorni di vita di lievi

tremori ed ipotonia assiale. Il peso alla nascita era di 3,440 gr (25°-50° pc). Nel primo anno di vita il bambino ha presentato ricorrenti episodi di infezioni del tratto respiratorio, durante i quali era letargico, scarsamente responsivo ed ipotonico. Questi episodi di torpore non si sono più presentati successivamente ma all'epoca del loro esordio avevano fatto sospettare una malattia metabolica ereditaria esclusa poi dagli esami di screening metabolico. All'età di 4 anni ha sviluppato episodi critici tonico-clonico generalizzati. L'EEG ha mostrato modiche anomalie irritative in sede temporo-occipitale. Ha sofferto di stitichezza cronica con frequenti episodi di dolore addominale. Il bambino ha raggiunto le tappe principali dello sviluppo psicomotorio con ritardo: a 4 anni circa manteneva la posizione seduta ma non quella eretta ed eseguiva comandi semplici solo saltuariamente. Il linguaggio non era ancora sviluppato a 4 anni e a 6 anni riusciva a pronunciare solo due paroline.

All'esame obiettivo veniva notata la presenza di microcefalia [all'età di 4 anni di età l'OFC risultava di 47.5 cm (<-2SD)]. Non presentava dismorfismi facciali significativi (**Figura 8**). La RMN cerebrale ha mostrato dilatazione dei ventricoli, ipoplasia del verme e degli emisferi cerebrali e iperintensità dei gangli basali T2 (**Figura 9**).

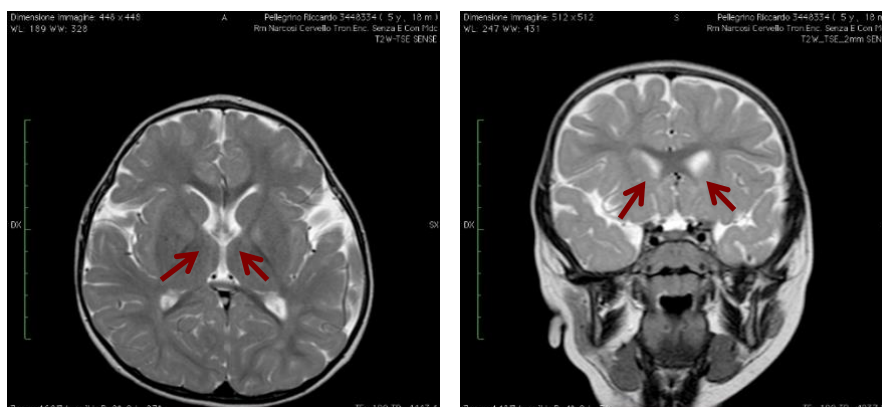


Figura 9. RMN encefalo del paziente all'età di 24 mesi: le iperintensità peri-ventricolari dei nuclei della base sono indicate dalle frecce.

L'ABR è risultato nella norma. La consulenza oculistica non ha riscontrato anomalie di rilievo

ma ha evidenziato solo una lieve esotropia dell'occhio destro. Le ecografie cardiaca ed addominale non hanno mostrato alterazioni.

Per approfondimento eziologico il piccolo ha effettuato: cariotipo (46, XY), analisi molecolare per *FMRI* e test di metilazione per sindrome di Angelman e sequenziamento per *UBE3A* che sono risultati tutti nella norma. Ha poi effettuato un ulteriore approfondimento cromosomico con array-CGH che ha rivelato una microdelezione 14q32.33 e una microdelezione 8p23.1 trasmessa dalla madre. La microdelezione 14q32.33 non contiene geni noti, è stata riportata in soggetti normali e pertanto è stata interpretata come variante cromosomica benigna. Entrambe le microdelezioni sono state pertanto considerate benigne.

Considerata la negatività di tutti gli esami eziologici eseguiti, il bambino è stato inserito nel programma “malattie senza diagnosi” e ha effettuato l'analisi WES che ha individuato un totale di 47,684 varianti, di cui 43,565 erano varianti a singolo nucleotide e 4,119 erano indel (inserzioni o delezioni). Di queste solo 9,315 risultavano avere ruolo deleterio per la funzionalità della proteina codificata (in base a strumenti di predizione quali Polyphen, SIFT e CADD) e solo 491 risultavano avere sia un significato funzionale che una frequenza < 0.001% su database GnomAD. Di queste ultime, 8 varianti erano *de novo*, 5 erano legate all'X e 478 erano trasmesse con modalità autosomica recessiva (omozigosi, variante in eterozigosi ereditata da un solo genitore ma compatibile con ereditarietà autosomica recessiva ed eterozigosi composta). Da un primo studio delle informazioni note sui singoli geni, sono risultate interessanti varianti in eterozigosi composta presenti in 9 geni. Tra questi geni abbiamo inizialmente considerato il gene calsintenina 2 (*CLSTN2*) che non era stato in precedenza associato a patologia e sembrava essere un ottimo gene candidato in quanto coinvolto nello sviluppo di sinapsi inibitorie e nella modulazione della trasmissione GABAergica [79]. Sebbene una di queste varianti fosse a bassa frequenza in

GnomAD, la seconda variante aveva un'elevata frequenza nella popolazione rendendo così improbabile un ruolo causativo di questo gene nel determinare la malattia del nostro paziente.

La variante *de novo* nel gene *POLR2A* aveva caratteristiche che la rendevano più probabilmente causativa. Questa variante determina una mutazione missenso p.Y1109H (c.3325T>C), assente in database di controllo e le analisi di predizione ne predicevano patogenicità (punteggio CADD pari a 28). Il gene *POLR2A* codifica per la proteina RPB1 che è la subunità più grande del complessi dell'RNA polimesari 2. Questo complesso è responsabile per la trascrizione di tutti i geni condificanti, di alcuni long and short -non-coding RNA.

Queste evidenze a favore del ruolo causativo della variante di *POLR2A* nella patologia del paziente hanno trovato un'ulteriore evidenza dall'individuazione di altri pazienti con fenotipo simile attraverso il portale web 'GeneMatcher'. Tale strumento ci ha permesso di collaborare con un altro gruppo olandese che ha raccolto 14 ulteriori pazienti con mutazioni nel gene *POLR2A*. Come il nostro paziente, i casi dello studio olandese, presentano un ritardo dello sviluppo psicomotorio di diversa gravità ed ipotonia in epoca infantile. Tutte le mutazioni dei pazienti ricadevano in domini funzionali della proteina altamente conservati nelle altre specie (**Figura 10**)

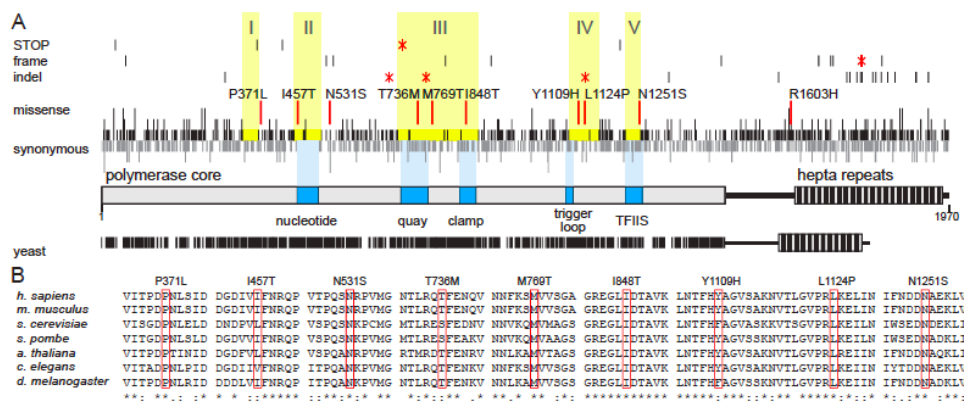


Figura 10. A) Rappresentazione grafica del gene *POLR2A*. Gli elementi funzionali della proteina sono evidenziati in blue. La posizione delle mutazioni di stop, frameshifts, indels, missenso e sinonime presenti in GnomAD sono rappresentate da barre verticali. La varianti dei

pazienti sono indicate da barre rosse e specificamente indicate. Regioni “prive di mutazioni missenso” sono evidenziate in giallo e indicate da numeri romani. Tutte le mutazioni missenso identificate nei pazienti ricadevano in regioni “prive di mutazioni missenso”. B) Viene rappresentata la conservazione delle mutazioni nelle diverse specie. Tutte le mutazioni missenso dei pazienti inclusi nello studio ricadevano in regioni altamente conservate in diverse specie.

Per ogni paziente è stato calcolato uno score di severità clinica che è stato messo in relazione con la “severità” della mutazione identificata mediante il punteggio CADD. La severità clinica aveva una correlazione inversa con la severità della mutazione dettata dal punteggio CADD. Si evinceva che le mutazioni con perdita di funzione con più alto CADD erano meglio tollerate causando un fenotipo più lieve (**Figure 11**). Le mutazioni missenso per le quali si attende solo un minimo danno strutturale, determinerebbero invece un più grave malfunzionamento della proteina RPB1 inducendo un effetto dominante negativo. Le mutazioni in *POLR2A* dei 15 pazienti sono state validate anche mediante studi in lieviti e saggi di funzionalità in cellule HeLa [80].

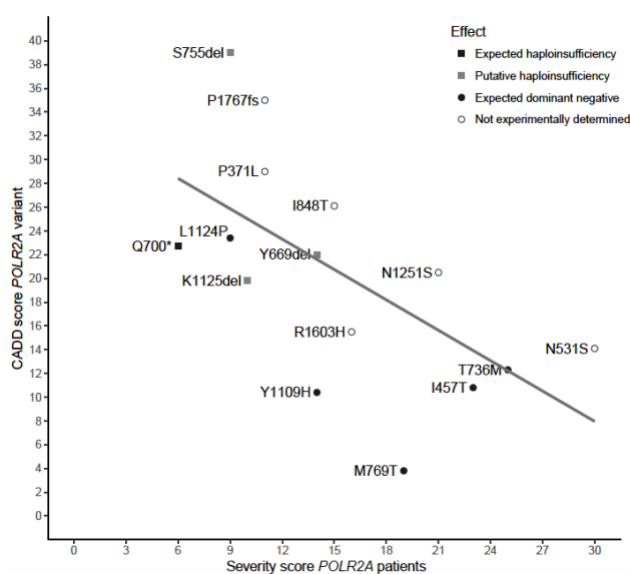


Figura 11. Correlazione inversa tra punteggi CADD e lo score clinico calcolato per ogni paziente. Le mutazioni con potenziale effetto dominante negative sono mutazioni missenso che sono associate ad uno score clinico maggiore.

Discussione

Il gene *POLR2A* codifica per una subunità della RNA polimerasi II (RPB1), l'enzima responsabile della sintesi del RNA messaggero nelle cellule eucariote. La RNA polimerasi II coordina virtualmente ogni fase della trascrizione, dalle fasi iniziali fino allo *splicing* e alle altre modifiche post-trascrizionali. Mutazioni di questo gene non sono state finora mai associate a malattie nell'uomo. La perdita completa di funzione di questa polimerasi è considerata incompatibile con la vita. La letale tossina amanitina è un potente inibitore della RNA polimerasi II.

L'identificazione di mutazioni in 15 individui con caratteristiche cliniche sovrapponibili principalmente caratterizzate da un disordine neuro-comportamentale con ipotonia infantile, la presenza di varianti in domini della proteina altamente conservati, e di studi funzionali che depongono per la patogenicità delle stesse hanno permesso di identificare un nuovo disordine mendeliano "*POLR2A*-related disorder", oggetto di una recente pubblicazione (Haijes HA et al., 2019) [80]. Questo paziente illustra la potenzialità del WES nell'identificare nuovi geni-malattia e l'importanza della condivisione delle informazioni con altri centri impegnati in programmi simili al nostro.

Caso 2- Conclusivo con presentazione atipica

Descrizione clinica e risultati del WES

La paziente è una bambina di 2 anni venuta alla nostra osservazione per la prima volta per la presenza di un difetto cardiaco congenito e note dismorfiche (**Figura 12**). La bambina è secondogenita di genitori sani e non consanguinei ed è nata da taglio cesareo dopo 38 settimane di gravidanza complicata dal riscontro mediante ecografie prenatali di tetralogia di Fallot e di ritardo di crescita intrauterino (IUGR).

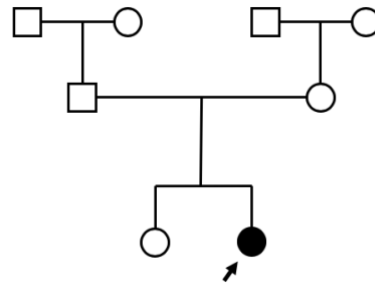


Figura 12. Albero genealogico e caratteristiche facciali della paziente descritta.

L'amniocentesi praticata per il riscontro di queste anomalie ha mostrato un cariotipo fetale normale. Alla nascita il peso era di 2300 gr (3° pc). Subito dopo la nascita, la tetralogia di Fallot è stata confermata e la piccola è stata trasferita nel reparto di cardiocirurgia, dove è stata sottoposta a due interventi di cateterismo cardiaco a 4 e 20 giorni di vita. Nei primi mesi la piccola si è alimentata con difficoltà ed ha presentato frequenti episodi di rigurgito. L'ecografia addominale praticata in questo periodo è risultata nella norma mentre l'ecografia cerebrale trans-fontanellare ha mostrato una moderata iperecogenicità periventricolare. A 5 mesi di vita ha praticato un terzo intervento di cateterismo per valvuloplastica e a 8 mesi circa si è sottoposta ad ulteriore intervento chirurgico di correzione della cardiopatia congenita.



Figura 13. Corectopia dell'iride della paziente.

All'esame clinico veniva evidenziata corectopia (o *ectopia pupillae*) ed esotropia dell'occhio sinistro mentre il cristallino ed il fondo oculare sono risultati nella norma (**Figura 13**). A 11 mesi è stata ricoverata in seguito ad un episodio di ascite ed effusione pleurica. A 2 anni circa ha presentato un grave episodio di pericardite che ha richiesto un nuovo ricovero. La paziente ha presentato sin dai primi mesi di vita uno scarso accrescimento staturico-ponderale e microcefalia. All'esame clinico venivano notati dismorfismi facciali quali ipertelorismo, fessure palpebrali strette, ponte nasale piatto, palato ogivale, mani piccole e brachidattilia. Tuttavia questi dismorfismi non erano suggestivi di una specifica sindrome genetica. Lo sviluppo psicomotorio è stato leggermente ritardato: a 16 mesi la bambina era in grado di pronunciare 3-4 paroline, raccogliere gli oggetti, sorridere e riconoscere i familiari ma non riusciva ancora a stare in piedi da sola e camminare. Nel sospetto di una sindrome genetica è stato richiesto l'array-CGH che ha evidenziato le microdelezioni 3q29, 4q13.3, 7q34, 15q14 e la microduplicazione 7q11.21. Tutte queste varianti essendo presenti in database di soggetti controllo, venivano interpretate come non-patogeniche. La paziente è stata quindi arruolata nel programma "Malattie senza Diagnosi" e l'analisi WES effettuata sul trio ha rilevato un totale di 56,437 varianti nella probanda, di cui 50,828 erano varianti di un singolo nucleotide e 5,609 erano indel. Di queste solo 10,505 risultavano avere ruolo deleterio per la funzionalità della proteina codificata (in base a strumenti di predizione quali Polyphen, SIFT e CADD) e solo 502 risultavano avere sia un significato funzionale che una frequenza < 0.001% su database ExAC. Di queste ultime, 14 varianti erano *de novo*, 2 erano legate all'X e 486 avevano una trasmissione di tipo autosomico recessiva (omozigosi, variante in eterozigosi ereditata da un solo genitore ma compatibile con ereditarietà autosomica recessiva ed eterozigosi composta). Tra le varianti identificate, molto significativa è

sembrata essere la variante *de novo* (c.1498A>G) del gene *SMAD4*. Questa variante risulta in una mutazione missenso (p.Ile500Val) con un punteggio CADD pari a 23 che è stata già precedentemente descritta in pazienti con sindrome di Myhre. La bambina in effetti presentava varie caratteristiche associate alla sindrome quali le anomalie cardiache, la brachidattilia e condivideva con i casi descritti in letteratura i dismorfismi facciali (**Figura 14**).

Discussione

La sindrome di Myhre è una malattia descritta per la prima volta circa 30 anni fa in due fratelli che presentavano simili dimorfismi facciali, bassa statura, ipertrofia muscolare, limitazione articolare, sordità e disabilità intellettiva. Nel 2009 è stato individuato il gene *SMAD4* come responsabile della malattia [81]. Questa sindrome è caratterizzata dalla presenza di ritardo di crescita, rigidità articolare, dismorfismi facciali (fessure palpebrali strette, ipoplasia mascellare, bocca piccola con labbro superiore sottile e prognatismo), anomalie scheletriche (platispondilia, ipoplasia delle ali dell'ileo e brachidattilia), schisi del palato, malformazioni cardiache, ipoacusia e disabilità intellettiva.

La presentazione clinica della sindrome di Myhre ha un certo grado di sovrapposizione con le caratteristiche cliniche di altre sindromi quali la displasia gelefisica e acromelica, la sindrome di Weil Marchesani e la LAPS (laryngotracheal stenosis, progressive, with short stature and arthropathy syndrome) [81–83]. Fino al 2016 sono stati descritti 83 casi in letteratura dei quali nessuno ha ricevuto la diagnosi prima dei due anni [84]. La nostra paziente è ad oggi la più piccola paziente mai descritta con sindrome di Myhre. Il ritardo nella diagnosi è principalmente dovuto all'assenza di molti tratti caratteristici di questo disordine nei primi anni di vita; le uniche caratteristiche, che sembrano essere precoci sono, infatti, la presenza della brachidattilia e delle anomalie cardiache mentre altre caratteristiche quali i problemi articolari, l'ipertrofia muscolare e

i dismorfismi diventano evidenti solo più avanti negli anni.

La nostra paziente presentava un'anomalia cardiaca, la tetralogia di Fallot, non ancora riportata nella sindrome di Myhre. I difetti cardiaci congeniti finora descritti comprendono difetto del setto ventricolare e atriale, pervietà del dotto arterioso, stenosi della valvola aortica e coartazione aortica [85]. Un'altra caratteristica peculiare della nostra paziente è la presenza di corectopia che non è mai stata riportata in pazienti con la sindrome di Myhre ma è stata già descritta in pazienti con mutazioni di *ADAMTSL4* che codifica per un gene che come *SMAD4* è coinvolto nel *pathway* di TGF- β [86].



Figura 14. Confronto tra le caratteristiche facciali della nostra paziente e quelle di altri pazienti con sindrome di Myhre (Garavelli et al.; Michot et al.). Tra i pazienti si notano dismorfismi facciali comuni quali ipertelorismo, fessure palpebrali strette, ponte nasale piatto e punta del naso allargata.

Le complicanze cardiache (pericardite e cardiomiopatia restrittiva) e otorinolaringoiatriche (stenosi tracheali) mettono a rischio la vita di questi pazienti [84]. Queste manifestazioni, spesso anche precoci, sottolineano ulteriormente l'importanza di una diagnosi tempestiva per individuare precocemente queste complicazioni potenzialmente letali.

SMAD4 codifica per una proteina coinvolta nella regolazione dei *pathway* di BMP e TGF- β , che hanno un ruolo fondamentale nella formazione della matrice extracellulare (ECM). Fino ad oggi sono state descritte tre varianti missenso in eterozigosi a carico del codone che codifica per Ile500

come nella nostra paziente [87] o più raramente a carico del codone Arg496 [81]. Questi amminoacidi sono localizzato nel dominio MH2 (Mad homology 2) della proteina situato vicino al sito di mono-ubiquitinazione Lys519. Nelle cellule dei pazienti è stata descritta anche un'alterazione della deposizione dell'ECM anche se minore rispetto alle patologie con mutazioni di *FBNI*, quali la sindrome di Marfan [88, 89]. Modelli cellulari hanno evidenziato l'efficacia della terapia farmacologica con Losartan nel migliorare la deposizione dell'ECM [88, 89].

Nel caso della nostra paziente, il raggiungimento di una diagnosi molecolare ha aperto una prospettiva terapeutica, visto che il Losartan un farmaco già in uso ed approvato per l'età pediatrica. La probanda ha iniziato terapia con il Losartan ed è in corso uno studio per la valutazione dei risultati del trial.

Caso 3: Gene candidato

Descrizione clinica e risultati del WES

Il paziente è un bambino di 8 anni e 8 mesi inviato alla valutazione genetica per ritardo dello sviluppo scarso accrescimento staturale-ponderale, microcefalia pre-natale, crisi epilettiche e letargia in trattamento con Levetiracetam, deficit visivo di origine centrale, disturbo della deglutizione e storia di neutropenia (**Figura 15**). La RMN-encefalo individuava agenesia del corpo calloso, della commissura ippocampale, e l'EEG anomalie epilettogene focali. Il probando ha effettuato differenti test diagnostici di approfondimento: test di metilazione 15q11-13 nel sospetto della sindrome di Prader-Willi, analisi molecolare del gene *SBDS* nel sospetto della sindrome di Shwachman-Diamond, array-CGH, amminoacidemia, acilcarnitine plasmatiche, acidi organici urinari, mucopolisaccaridi urinari che sono risultati tutti negativi.

L'analisi del sequenziamento dell'esoma del trio, con copertura a 20x >90% per tutti i membri, ha identificato 1,162 varianti filtrate secondo database di popolazione. Nessuna variante era

identificabile in omozigosi, mentre circa 30 erano *de novo* di cui 7 con punteggio CADD superiore a 25, 6 geni con varianti in eterozigosi composta e 3 trasmesse con modalità X-linked.



Figura 15. Albero genealogico del paziente e caratteristiche facciali. Il probando presenta una ipotonia assiale che rende difficoltoso il mantenimento della posizione seduta con appoggio.

Tra le mutazioni *de novo* veniva identificata una mutazione missenso del gene *RAB10*, p.(T23I) (NM_016131, c.68C>T), assente in database di controlli (GnomAD e database interno) e predetta patogenica secondo Sift, Polyphen e CADD (28.5). Il residuo amminoacidico modificato è altamente conservato nelle altre specie.

RAB10 appartiene alla famiglia delle proteine RAB, la più ampia famiglia di piccole proteine ad attività GTPasica che controllano diversi passaggi del traffico intracellulare. Il traffico intracellulare è fondamentale per lo sviluppo assonale e del trasporto assonale anterogrado e retrogrado e la morfologia neuronale. *RAB10* media anche il trasporto agli assoni terminali attraverso le interazioni con le Kinesine [90].

Similmente alle altre GTPasi l'attività delle proteine Rab è strettamente controllata dal ricircolo tra una forma attiva legata al GTP e una inattiva legata al GDP [91]. Il passaggio tra i due diversi stati, attivo ed inattivo è mediato da proteine regolatorie: *guanine nucleotide exchange factors* (GEFs) che mediano lo scambio del GDP in GTP e le *GTPase-activating proteins* (GAPs) che stimolano l'attività GTPasica intrinseca delle proteine Rab (**Figura 18**) [92]. In *RAB10* il codone T23 fa parte del sito che lega il GTP, pertanto una mutazione in questo sito potrebbe

causare l'alterazione del legame del GDP e GTP risultando in una forma inattiva. In particolare il mutante inattivo con la mutazione p.T23N, che è localizzata nello stesso codone della variante presentata dal nostro paziente, causa un blocco in *RAB10* nella forma inattiva (*GDP-locked*) con delle note conseguenze funzionali studiate *in vitro* mediante un effetto dominante negativo [92].

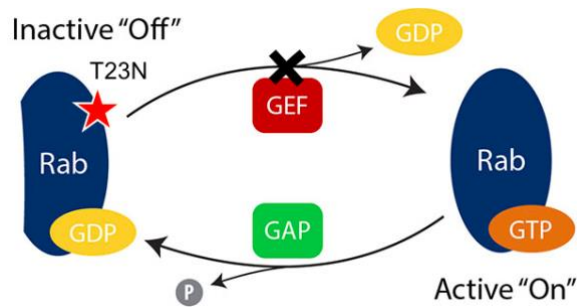


Figura 16. L'attivazione delle proteine Rab è mediata da due proteine, GEF (Guanine nucleotide exchange factor) e GAP (GTPase activating protein).

A livello intracellulare il mutante T23N altera drammaticamente la morfologia del reticolo endoplasmico, con un aumento delle regioni cisternali e riduce significativamente l'estensione la fusione dei tubuli del reticolo endoplasmico [93]. Inoltre il mutante T23N causa una alterazione morfologica del reticolo endoplasmico che coinvolge alcuni siti di sintesi dei fosfolipidi, in particolari gli enzimi fosfatidilinositolo sintasi e CEPT1 [93]. Il mutante *RAB10* GDP-locked causa una inibizione della formazione e della funzione degli enzimi di sintesi dei fosfolipidi [93]. *RAB10*-T23N inibisce l'arborizzazione assonale e dendritica [94–96] (**Figura 17**).

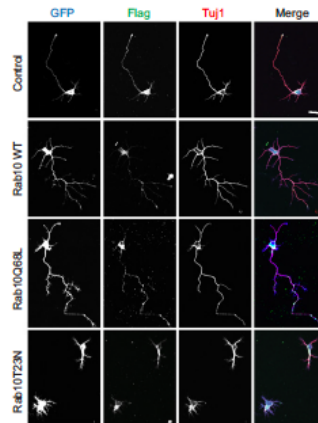


Figura 17. Neuroni ippocampali sono transfettati con GFP, o co-transfettati con Flag-per i diversi costrutti di RAB10 e poi colorati con anticorpi specifici. I neuroni con RAB10 GDP-locked da mutazione T23N presentano un significativo difetto di arborizzazione dendritica neuronale rispetto ai controlli e alla forma RAB10 GTP-locked (Q86L) (Wang et al., 2011).

Difetti in altre proteine della stessa famiglia come *RAB11* e *RAB39* causano forme sindromiche di disabilità cognitiva. La mutazione del probando è stata depositata in *Genematcher* e sono stati presi contatti diretti con diversi centri nazionali ed internazionali, ma ad oggi non sono stati identificati altri pazienti con quadro clinico sovrapponibile e mutazione p.T23I in *RAB10*.

Discussione

Le malattie “rare” vengono definite come malattie con un’incidenza inferiore a 5/10,000 nella popolazione (Orphan Drug Regulation 141/2000). Molte di queste malattie sono croniche e gravemente invalidanti e sono causa di mortalità precoce. La mancanza di politiche sanitarie specifiche e la scarsità di conoscenze a riguardo si sono tradotte negli anni in gravi ritardi nella diagnosi e in difficoltà di accesso alle cure. Nell’ambito di queste malattie, ancora molte sono rarissime e restano misconosciute e non diagnosticate. Solo circa 7,000 dei 22,000 geni umani ha una funzione nota e solo 6,200 sono stati associati ad un disordine mendeliano ad oggi. Sebbene mutazioni in molti geni sono probabilmente incompatibili con la vita vi è il forte sospetto che molti disordini genetici debbano essere ancora riconosciuti. Studi recenti hanno dimostrato che meno

della metà dei pazienti con sindromi genetiche hanno ricevuto una diagnosi usando approcci genetici standard e che il 33% dei pazienti con malattie rare impiega da 1-5 anni per l'identificazione della causa genetica e il 15% dei pazienti più di 5 anni (dati NIH Office of Rare Disease Research). La percentuale di diagnosi sembra invece aumentare in maniera drammatica con le nuove metodologie di indagine genetica [97]. Il WES, infatti, si sta diffondendo come strumento importante per la definizione diagnostica e l'individuazione delle cause in moltissime di queste malattie. Il programma Telethon per le “malattie senza diagnosi” che è l'argomento di questa tesi è stato il primo esempio di applicazione sistematica in Italia del WES in pazienti senza diagnosi. In molti casi, compresi quelli illustrati in questo lavoro di tesi, quest'unico test ha consentito di mettere fine alla lunga odissea diagnostica, risultando persino economicamente vantaggioso se confrontato con i costi complessivi delle indagini effettuate prima del WES [98].

La probabilità di successo diagnostico del WES nei pazienti senza diagnosi si stima dal 25% al 50% [32–35]. In uno studio il WES eseguito su 3,040 casi (analizzati per lo più *trios*) ha permesso di identificare una diagnosi definitiva di malattia genetica nel 28.8% dei casi mentre un risultato possibile o probabile è stato riscontrato nel 51.8% dei casi; un gene candidato era suggerito nel 7.6% dei casi e nel 11.8% dei casi l'analisi risultava negativa [41]. Il tasso di successo diagnostico incrementa quando il numero di membri familiari analizzati aumenta: nei *singleton* (analisi del probando) è del 23.6% mentre per i *trios* è del 31%. Il tasso di successo diagnostico è significativamente elevato quando viene applicata un'attenta selezione dei pazienti e raggiunge il 68% in pazienti con presentazioni fortemente suggestive di malattie neurometaboliche [44]. Questo lavoro di tesi svolto su 116 famiglie, ha permesso di identificare il difetto molecolare nel 43% e mentre nel 28% delle famiglie arruolate è stato identificato un gene candidato. La nostra esperienza sembra confermare il principio che una accurata selezione pre-analitica dei pazienti

amenti l'efficacia diagnostica del WES. La conoscenza approfondita del fenotipo clinico, infatti, permette di facilitare ed accelerare molto l'analisi dell'ampio numero di varianti riscontrate al WES. Rispetto ad altre simili iniziative in ambito internazionale, la peculiarità del progetto pilota all'interno del quale è stata svolta questa attività di tesi è che esso è rivolto solo alla popolazione pediatrica e permette a numerose figure (medici, pediatri, neonatologi, genetisti, neuropsichiatri infantili) su tutto il territorio nazionale di poter richiedere l'accesso di pazienti al programma, inserendo le informazioni dei pazienti su una piattaforma web dedicata. I centri coinvolti hanno condiviso sempre ogni passaggio del percorso ed ogni eventuale risultato raggiunto, con i medici che lo hanno attivato.

Si stima che l'esoma di ciascun individuo contenga all'incirca 10,000 varianti non sinonime ed è portatore di 50-100 mutazioni in eterozigosi potenzialmente correlate a patologie recessive [99]. Usando filtri appositi secondo un ordine ben preciso (esclusione di varianti polimorfiche, segregazione familiare) il numero di varianti prese in considerazione viene drasticamente ridotto a 5-40 varianti per paziente [41]. Questi numeri ci fanno ben comprendere come proprio l'analisi dei dati rappresenti la parte più critica di questa indagine. Nonostante questo e nonostante i numerosi ostacoli che il percorso analitico pone ai medici/ricercatori, identificare la causa di queste malattie rarissime rimane un obiettivo di primaria importanza, perchè consente di dare una spiegazione ai genitori delle problematiche del bambino, di arrestare l'odissea diagnostica nella quale spesso questi pazienti si trovano e di offrire una consulenza genetica precisa con la possibilità di eseguire una diagnosi prenatale ed una precisa definizione del rischio di ricorrenza. Inoltre l'individuazione della causa di una malattia apre alla possibilità di sviluppare approcci terapeutici mirati, oltre a consentire una migliore comprensione dei *pathways* biologici coinvolti nella malattia. Poter finalmente attribuire un nome alla patologia che si combatte da anni e potersi

confrontare con l'esperienza di altri pazienti con la stessa malattia ha un importante risvolto psicologico per i genitori dei bambini affetti. Dal punto di vista clinico invece il raggiungimento di una diagnosi precisa permette di sviluppare un approccio assistenziale mirato. In alcuni casi selezionati la diagnosi eziologica permette di instaurare una terapia o di candidare i pazienti per trial clinici sperimentali in corso. Quest'opportunità è stata presa in considerazione per esempio per la terapia con Losartan della bambina descritta nel caso 2. Nonostante l'uso del WES, vi sono ancora pazienti che rimangono senza diagnosi, nella nostra casistica circa il 29%. In questi casi il difetto potrebbe essere a carico di regioni regolatorie non codificanti (introni, splicing), non analizzate dal WES o in regioni codificante ma non ben studiate. Inoltre in casi di mutazioni presenti in stato di mosaicismo il difetto potrebbe non essere identificato.

Studi di WGS probabilmente permetteranno di risolvere nel futuro almeno una porzione di questi casi. Tuttavia data l'enorme quantità di dati generati dal WGS e la loro difficile interpretazione questa analisi non è ancora pronta per applicazioni diffuse in pazienti senza diagnosi. WGS esamina il 90% del genoma, ma la sua utilità diagnostica non differisce sostanzialmente da quella del WES; si stima infatti che la *detection rate* del WGS è superiore a quella del WES solo di circa il 3% [100]. Diversi studi hanno valutato l'utilità diagnostica del trascrittoma, come uno strumento complementare al WES/WGS. Il trascrittoma permette infatti di chiarire varianti incerte, anche sinonime, di splicing o introniche profonde, evidenziate dal WES/WGS, risultando in un incremento del rate diagnostico fino al 35% [101, 102]. Inoltre nuove metodologie permettono lo studio del WES *read depth*, predicendo la presenza di CNV, potrebbe estendere l'utilità clinica dell'esoma [103].

Oltre agli indubbi vantaggi per i pazienti e le loro famiglie, lo studio di malattie genetiche molto rare permette anche di poter acquisire importanti conoscenze sulla fisiologia e biologia, fornendo

importanti spunti per la ricerca di base. La scoperta di nuove associazioni tra geni e malattia permette per esempio la scoperta di nuovi *pathways* e la rivalutazione delle implicazioni mediche di *pathways* già noti. Nel 1657 William Harvey scriveva: *‘Nature is nowhere accustomed more openly to display her secret mysteries than in cases where she shows tracings of her workings apart from the beaten paths; nor is there any better way to advance the proper practice of medicine than to give our minds to the discovery of the usual law of nature, by careful investigation of cases of rarer forms of disease.’* Analogamente lo studio delle basi molecolari di malattie genetiche rare che possono interessare anche pochissimi pazienti può portare alla scoperta di *pathways* molecolari importanti che possono avere implicazioni molto più vaste di quelle attese. Per esempio lo studio della rara ipercolesterolemia familiare dovuta a mutazioni in omozigosi del gene per il recettore delle LDL ha permesso di comprendere la patofisiologia dell’ipercolesterolemia, una delle malattie croniche più diffuse nel mondo occidentale.

In conclusione, i programmi di “malattie senza diagnosi” hanno grossa valenza su diversi livelli: per il paziente e la famiglia, per le organizzazioni istituzionali e per l’avanzamento delle conoscenze scientifiche. In assenza di una diagnosi, la speranza di un potenziale trattamento è nulla, l'utilizzo coscienzioso delle strutture e risorse sanitarie impossibile e la medicina personalizzata una realtà ancora troppo lontana.

La possibilità di un sequenziamento rapido ed economicamente sostenibile del genoma modificherà in maniera sostanziale la medicina moderna nel prossimo futuro. Queste tecnologie diventeranno sempre più validi strumenti di applicazione clinica permettendo diagnosi più accurate e lo sviluppo di nuove terapie. L’uso clinico del WES in pazienti con sospette malattie metaboliche ereditarie, inoltre, ha grosse potenzialità perché per varie di queste malattie è disponibile una terapia e con queste metodiche è possibile intervenire modificando la storia

naturale della malattia. Sulla base di queste considerazioni è stato anche proposto uno screening neonatale di tipo genomico da accompagnare allo screening neonatale espanso basato sul dosaggio di metaboliti. Per alcune di queste malattie è in fase avanzata di sperimentazione la terapia genica che potrebbe offrire una cura definitiva.

Un ulteriore campo di applicazione potrebbe essere, inoltre, quello della diagnostica prenatale. Negli ultimi anni si sta, infatti, diffondendo la diagnosi prenatale non-invasiva basata sull'analisi del DNA fetale libero nella circolazione materna. L'utilizzo di metodiche NGS potrebbe consentire, quindi, in un futuro non molto lontano lo screening allargato sull'intero genoma fetale [104].

Conclusioni

La NGS ha un incredibile potenziale per la risoluzione di casi non diagnosticati di condizioni genetiche. Questo comporta notevoli vantaggi sia sul piano clinico, aumentando considerevolmente le potenzialità dei test genetici di raggiungere la diagnosi causale di malattia sia sul piano della ricerca, rendendo più facile l'individuazione di nuovi geni e *pathway* metabolici. I risultati preliminari presentati in questo lavoro di tesi sono molto incoraggianti poiché hanno fornito risultati definitivi a 50 famiglie in 7 di queste abbiamo identificato un nuovo gene-malattia. I risultati finora ottenuti mostrano che la NGS ha potenzialità importanti per la diagnosi e nell'immediato futuro potrebbe modificare in maniera sostanziale l'algoritmo diagnostico riducendo o evitando molti test molecolari routinariamente richiesti e procedure invasive, quali biopsie. Un'attenta valutazione clinica rimane però una priorità in quanto essa è indispensabile per un corretta interpretazione dei risultati di sequenziamento.

Referenze

- 1 Beck M. Rare and Ultra Rare Diseases; *J Dev Drugs* 2012;**01**. doi:10.4172/2329-6631.1000e107
- 2 Richter T, Nestler-Parr S, Babela R, Khan ZM, Tesoro T, Molsen E, Hughes DA, International Society for Pharmacoeconomics and Outcomes Research Rare Disease Special Interest Group. Rare Disease Terminology and Definitions-A Systematic Global Review: Report of the ISPOR Rare Disease Special Interest Group. *Value Health* 2015;**18**:906–14.
- 3 Salvador-Carulla L, Reed GM, Vaez-Azizi LM, Cooper S-A, Martinez-Leal R, Bertelli M, Adnams C, Cooray S, Deb S, Akoury-Dirani L, Girimaji SC, Katz G, Kwok H, Luckasson R, Simeonsson R, Walsh C, Munir K, Saxena S. Intellectual developmental disorders: towards a new name, definition and framework for ‘mental retardation/intellectual disability’ in ICD-11. *World Psychiatry* 2011;**10**:175–80.
- 4 Douzgou S, Chervinsky E, Gyftodimou Y, Kitsiou-Tzeli S, Shalev S, Kanavakis E, Donnai D, Clayton-Smith J. Dysmorphology services: a snapshot of current practices and a vision for the future. *Clin Genet* 2016;**89**:27–33.
- 5 Amberger JS, Bocchini CA, Scott AF, Hamosh A. OMIM.org: leveraging knowledge across phenotype-gene relationships. *Nucleic Acids Res* 2019;**47**:D1038–43.
- 6 Tiff CJ, Adams DR. The National Institutes of Health undiagnosed diseases program. *Curr Opin Pediatr* 2014;**26**:626–33.
- 7 Ahn JW, Bint S, Bergbaum A, Mann K, Hall RP, Ogilvie CM. Array CGH as a first line diagnostic test in place of karyotyping for postnatal referrals - results from four years’ clinical application for over 8,700 patients. *Mol Cytogenet* 2013;**6**:16.
- 8 Raffan E, Semple RK. Next generation sequencing--implications for clinical practice. *Br Med*

- Bull* 2011;**99**:53–71.
- 9 Nigro V. La next generation sequencing è entrata nella pratica pediatrica. *Prospett Pediatr* 2015;**45**:137–42.
- 10 Pareek CS, Smoczynski R, Tretyn A. Sequencing technologies and genome sequencing. *J Appl Genet* 2011;**52**:413–35.
- 11 Xuan J, Yu Y, Qing T, Guo L, Shi L. Next-generation sequencing in the clinic: promises and challenges. *Cancer Lett* 2013;**340**:284–95.
- 12 Metzker ML. Sequencing technologies - the next generation. *Nat Rev Genet* 2010;**11**:31–46.
- 13 Chin ELH, da Silva C, Hegde M. Assessment of clinical analytical sensitivity and specificity of next-generation sequencing for detection of simple and complex mutations. *BMC Genet* 2013;**14**:6.
- 14 Ng SB, Bigham AW, Buckingham KJ, Hannibal MC, McMillin MJ, Gildersleeve HI, Beck AE, Tabor HK, Cooper GM, Mefford HC, Lee C, Turner EH, Smith JD, Rieder MJ, Yoshiura K-I, Matsumoto N, Ohta T, Niikawa N, Nickerson DA, Bamshad MJ, Shendure J. Exome sequencing identifies MLL2 mutations as a cause of Kabuki syndrome. *Nat Genet* 2010;**42**:790–3.
- 15 Hoischen A, van Bon BWM, Gilissen C, Arts P, van Lier B, Steehouwer M, de Vries P, de Reuver R, Wieskamp N, Mortier G, Devriendt K, Amorim MZ, Revencu N, Kidd A, Barbosa M, Turner A, Smith J, Oley C, Henderson A, Hayes IM, Thompson EM, Brunner HG, de Vries BBA, Veltman JA. De novo mutations of SETBP1 cause Schinzel-Giedion syndrome. *Nat Genet* 2010;**42**:483–5.
- 16 Paulussen ADC, Stegmann APA, Blok MJ, Tserpelis D, Posma-Velter C, Detisch Y,

- Smeets EEJGL, Wagemans A, Schrandt JJP, van den Boogaard M-JH, van der Smagt J, van Haeringen A, Stolte-Dijkstra I, Kerstjens-Frederikse WS, Mancini GM, Wessels MW, Hennekam RCM, Vreeburg M, Geraedts J, de Ravel T, Fryns J-P, Smeets HJ, Devriendt K, Schrandt-Stumpel CTRM. MLL2 mutation spectrum in 45 patients with Kabuki syndrome. *Hum Mutat* 2011;**32**:E2018-25.
- 17 Choi M, Scholl UI, Ji W, Liu T, Tikhonova IR, Zumbo P, Nayir A, Bakkaloğlu A, Ozen S, Sanjad S, Nelson-Williams C, Farhi A, Mane S, Lifton RP. Genetic diagnosis by whole exome capture and massively parallel DNA sequencing. *Proc Natl Acad Sci U S A* 2009;**106**:19096–101.
- 18 Ng SB, Turner EH, Robertson PD, Flygare SD, Bigham AW, Lee C, Shaffer T, Wong M, Bhattacharjee A, Eichler EE, Bamshad M, Nickerson DA, Shendure J. Targeted capture and massively parallel sequencing of 12 human exomes. *Nature* 2009;**461**:272–6.
- 19 Hodges E, Xuan Z, Balija V, Kramer M, Molla MN, Smith SW, Middle CM, Rodesch MJ, Albert TJ, Hannon GJ, McCombie WR. Genome-wide in situ exon capture for selective resequencing. *Nat Genet* 2007;**39**:1522–7.
- 20 Bainbridge MN, Wang M, Burgess DL, Kovar C, Rodesch MJ, D’Ascenzo M, Kitzman J, Wu Y-Q, Newsham I, Richmond TA, Jeddloh JA, Muzny D, Albert TJ, Gibbs RA. Whole exome capture in solution with 3 Gbp of data. *Genome Biol* 2010;**11**:R62.
- 21 von Bubnoff A. Next-generation sequencing: the race is on. *Cell* 2008;**132**:721–3.
- 22 SIGU. Il Sequenziamento del DNA di nuova generazione indicazioni per l’impiego clinico. 2016.
- 23 Cao Y, Qu Y, Song F, Zhang T, Bai J, Jin Y, Wang H. Fast clinical molecular diagnosis of hyperphenylalaninemia using next-generation sequencing-based on a custom AmpliSeq™

- panel and Ion Torrent PGM sequencing. *Mol Genet Metab* 2014;**113**:261–6.
- 24 Acke FR, Malfait F, Vanakker OM, Steyaert W, De Leeneer K, Mortier G, Dhooge I, De Paepe A, De Leenheer EMR, Coucke PJ. Novel pathogenic COL11A1/COL11A2 variants in Stickler syndrome detected by targeted NGS and exome sequencing. *Mol Genet Metab* 2014;**113**:230–5.
- 25 Halbritter J, Diaz K, Chaki M, Porath JD, TARRIER B, Fu C, Innis JL, Allen SJ, Lyons RH, Stefanidis CJ, Omran H, Soliman NA, Otto EA. High-throughput mutation analysis in patients with a nephronophthisis-associated ciliopathy applying multiplexed barcoded array-based PCR amplification and next-generation sequencing. *J Med Genet* 2012;**49**:756–67.
- 26 Aparisi MJ, Aller E, Fuster-García C, García-García G, Rodrigo R, Vázquez-Manrique RP, Blanco-Kelly F, Ayuso C, Roux A-F, Jaijo T, Millán JM. Targeted next generation sequencing for molecular diagnosis of Usher syndrome. *Orphanet J Rare Dis* 2014;**9**:168.
- 27 De Rocco D, Bottega R, Cappelli E, Cavani S, Criscuolo M, Nicchia E, Corsolini F, Greco C, Borriello A, Svahn J, Pillon M, Mecucci C, Casazza G, Verzegnassi F, Cugno C, Locasciulli A, Farruggia P, Longoni D, Ramenghi U, Barberi W, Tucci F, Perrotta S, Grammatico P, Hanenberg H, Della Ragione F, Dufour C, Savoia A, Bone Marrow Failure Study Group of the Italian Association of Pediatric Onco-Hematology. Molecular analysis of Fanconi anemia: the experience of the Bone Marrow Failure Study Group of the Italian Association of Pediatric Onco-Hematology. *Haematologica* 2014;**99**:1022–31.
- 28 Di Fruscio G, Schulz A, De Cegli R, Savarese M, Mutarelli M, Parenti G, Banfi S, Braulke T, Nigro V, Ballabio A. Lysoplex: An efficient toolkit to detect DNA sequence variations in the autophagy-lysosomal pathway. *Autophagy* 2015;**11**:928–38.
- 29 Lee H, Deignan JL, Dorrani N, Strom SP, Kantarci S, Quintero-Rivera F, Das K, Toy T,

- Harry B, Yourshaw M, Fox M, Fogel BL, Martinez-Agosto JA, Wong DA, Chang VY, Shieh PB, Palmer CGS, Dipple KM, Grody WW, Vilain E, Nelson SF. Clinical exome sequencing for genetic identification of rare Mendelian disorders. *JAMA* 2014;**312**:1880–7.
- 30 Al-Dewik N, Mohd H, Al-Mureikhi M, Ali R, Al-Mesaifri F, Mahmoud L, Shahbeck N, El-Akouri K, Almulla M, Al Sulaiman R, Musa S, Al-Marri AA, Richard G, Juusola J, Solomon BD, Alkuraya FS, Ben-Omran T. Clinical exome sequencing in 509 Middle Eastern families with suspected Mendelian diseases: The Qatari experience. *Am J Med Genet Part A* 2019;**179**:927–35.
- 31 Saudi Mendeliome Group. Comprehensive gene panels provide advantages over clinical exome sequencing for Mendelian diseases. *Genome Biol* 2015;**16**:134.
- 32 Ankala A, da Silva C, Gualandi F, Ferlini A, Bean LJH, Collins C, Tanner AK, Hegde MR. A comprehensive genomic approach for neuromuscular diseases gives a high diagnostic yield. *Ann Neurol* 2015;**77**:206–14.
- 33 Chong JX, Buckingham KJ, Jhangiani SN, Boehm C, Sobreira N, Smith JD, Harrell TM, McMillin MJ, Wiszniewski W, Gambin T, Coban Akdemir ZH, Doheny K, Scott AF, Avramopoulos D, Chakravarti A, Hoover-Fong J, Mathews D, Witmer PD, Ling H, Hetrick K, Watkins L, Patterson KE, Reinier F, Blue E, Muzny D, Kircher M, Bilguvar K, López-Giráldez F, Sutton VR, Tabor HK, Leal SM, Gunel M, Mane S, Gibbs RA, Boerwinkle E, Hamosh A, Shendure J, Lupski JR, Lifton RP, Valle D, Nickerson DA, Bamshad MJ. The Genetic Basis of Mendelian Phenotypes: Discoveries, Challenges, and Opportunities. *Am J Hum Genet* 2015;**97**:199–215.
- 34 Taylor JC, Martin HC, Lise S, Broxholme J, Cazier J-B, Rimmer A, Kanapin A, Lunter G, Fiddy S, Allan C, Aricescu AR, Attar M, Babbs C, Becq J, Beeson D, Bento C, Bignell P,

- Blair E, Buckle VJ, Bull K, Cais O, Cario H, Chapel H, Copley RR, Cornall R, Craft J, Dahan K, Davenport EE, Dendrou C, Devuyst O, Fenwick AL, Flint J, Fugger L, Gilbert RD, Goriely A, Green A, Greger IH, Grocock R, Gruszczyk A V, Hastings R, Hatton E, Higgs D, Hill A, Holmes C, Howard M, Hughes L, Humburg P, Johnson D, Karpe F, Kingsbury Z, Kini U, Knight JC, Krohn J, Lambie S, Langman C, Lonie L, Luck J, McCarthy D, McGowan SJ, McMullin MF, Miller KA, Murray L, Németh AH, Nesbit MA, Nutt D, Ormondroyd E, Oturai AB, Pagnamenta A, Patel SY, Percy M, Petousi N, Piazza P, Piret SE, Polanco-Echeverry G, Popitsch N, Powrie F, Pugh C, Quek L, Robbins PA, Robson K, Russo A, Sahgal N, van Schouwenburg PA, Schuh A, Silverman E, Simmons A, Sørensen PS, Sweeney E, Taylor J, Thakker R V, Tomlinson I, Trebes A, Twigg SRF, Uhlig HH, Vyas P, Vyse T, Wall SA, Watkins H, Whyte MP, Witty L, Wright B, Yau C, Buck D, Humphray S, Ratcliffe PJ, Bell JI, Wilkie AOM, Bentley D, Donnelly P, McVean G. Factors influencing success of clinical genome sequencing across a broad spectrum of disorders. *Nat Genet* 2015;**47**:717–26.
- 35 Yang Y, Muzny DM, Xia F, Niu Z, Person R, Ding Y, Ward P, Braxton A, Wang M, Buhay C, Veeraraghavan N, Hawes A, Chiang T, Leduc M, Beuten J, Zhang J, He W, Scull J, Willis A, Landsverk M, Craigen WJ, Bekheirnia MR, Stray-Pedersen A, Liu P, Wen S, Alcaraz W, Cui H, Walkiewicz M, Reid J, Bainbridge M, Patel A, Boerwinkle E, Beaudet AL, Lupski JR, Plon SE, Gibbs RA, Eng CM. Molecular Findings Among Patients Referred for Clinical Whole-Exome Sequencing. *JAMA* 2014;**312**:1870.
- 36 Chae JH, Vasta V, Cho A, Lim BC, Zhang Q, Eun SH, Hahn SH. Utility of next generation sequencing in genetic diagnosis of early onset neuromuscular disorders. *J Med Genet* 2015;**52**:208–16.
- 37 Nijman IJ, van Montfrans JM, Hoogstraat M, Boes ML, van de Corput L, Renner ED,

- van Zon P, van Lieshout S, Elferink MG, van der Burg M, Vermont CL, van der Zwaag B, Janson E, Cuppen E, Ploos van Amstel JK, van Gijn ME. Targeted next-generation sequencing: a novel diagnostic tool for primary immunodeficiencies. *J Allergy Clin Immunol* 2014;**133**:529–34.
- 38 Yohe S, Hauge A, Bunjer K, Kemmer T, Bower M, Schomaker M, Onsongo G, Wilson J, Erdmann J, Zhou Y, Deshpande A, Spears MD, Beckman K, Silverstein KAT, Thyagarajan B. Clinical validation of targeted next-generation sequencing for inherited disorders. *Arch Pathol Lab Med* 2015;**139**:204–10.
- 39 Boycott KM, Rath A, Chong JX, Hartley T, Alkuraya FS, Baynam G, Brookes AJ, Brudno M, Carracedo A, den Dunnen JT, Dyke SOM, Estivill X, Goldblatt J, Gonthier C, Groft SC, Gut I, Hamosh A, Hieter P, Höhn S, Hurles ME, Kaufmann P, Knoppers BM, Krischer JP, Macek M, Matthijs G, Olry A, Parker S, Paschall J, Philippakis AA, Rehm HL, Robinson PN, Sham P-C, Stefanov R, Taruscio D, Unni D, Vanstone MR, Zhang F, Brunner H, Bamshad MJ, Lochmüller H. International Cooperation to Enable the Diagnosis of All Rare Genetic Diseases. *Am J Hum Genet* 2017;**100**:695–705.
- 40 Bamshad MJ, Ng SB, Bigham AW, Tabor HK, Emond MJ, Nickerson DA, Shendure J. Exome sequencing as a tool for Mendelian disease gene discovery. *Nat Rev Genet* 2011;**12**:745–55.
- 41 Retterer K, Jussola J, Cho MT, Vitazka P, Millan F, Gibellini F, Vertino-Bell A, Smaoui N, Neidich J, Monaghan KG, McKnight D, Bai R, Suchy S, Friedman B, Tahiliani J, Pineda-Alvarez D, Richard G, Brandt T, Haverfield E, Chung WK, Bale S. Clinical application of whole-exome sequencing across clinical indications. *Genet Med* 2016;**18**:696–704.
- 42 Zhang X. Exome sequencing greatly expedites the progressive research of Mendelian

- diseases. *Front Med* 2014;**8**:42–57.
- 43 Parenti. Come la ricerca sta affrontando le problematiche irrisolte nella diagnosi e nella terapia delle malattie metaboliche ereditarie. prospettive in pediatria. 2016;**25**–37.
- 44 Tarailo-Graovac M, Shyr C, Ross CJ, Horvath GA, Salvarinova R, Ye XC, Zhang L-H, Bhavsar AP, Lee JJY, Drögemöller BI, Abdelsayed M, Alfadhel M, Armstrong L, Baumgartner MR, Burda P, Connolly MB, Cameron J, Demos M, Dewan T, Dionne J, Evans AM, Friedman JM, Garber I, Lewis S, Ling J, Mandal R, Mattman A, McKinnon M, Michoulas A, Metzger D, Ogunbayo OA, Rakic B, Rozmus J, Ruben P, Sayson B, Santra S, Schultz KR, Selby K, Shekel P, Sirrs S, Skrypnik C, Superti-Furga A, Turvey SE, Van Allen MI, Wishart D, Wu J, Wu J, Zafeiriou D, Kluijtmans L, Wevers RA, Eydoux P, Lehman AM, Vallance H, Stockler-Ipsiroglu S, Sinclair G, Wasserman WW, van Karnebeek CD. Exome Sequencing and the Management of Neurometabolic Disorders. *N Engl J Med* 2016;**374**:2246–55.
- 45 Dewey FE, Grove ME, Pan C, Goldstein BA, Bernstein JA, Chaib H, Merker JD, Goldfeder RL, Enns GM, David SP, Pakdaman N, Ormond KE, Caleshu C, Kingham K, Klein TE, Whirl-Carrillo M, Sakamoto K, Wheeler MT, Butte AJ, Ford JM, Boxer L, Ioannidis JPA, Yeung AC, Altman RB, Assimes TL, Snyder M, Ashley EA, Quertermous T. Clinical interpretation and implications of whole-genome sequencing. *JAMA* 2014;**311**:1035–45.
- 46 Lelieveld SH, Veltman JA, Gilissen C. Novel bioinformatic developments for exome sequencing. *Hum Genet* 2016;**135**:603–14.
- 47 Clark MM, Stark Z, Farnaes L, Tan TY, White SM, Dimmock D, Kingsmore SF. Meta-analysis of the diagnostic and clinical utility of genome and exome sequencing and chromosomal microarray in children with suspected genetic diseases. *npj Genomic Med* 2018;**3**:16.

- 48 Gilissen C, Hoischen A, Brunner HG, Veltman JA. Unlocking Mendelian disease using exome sequencing. *Genome Biol* 2011;**12**:228.
- 49 Adzhubei IA, Schmidt S, Peshkin L, Ramensky VE, Gerasimova A, Bork P, Kondrashov AS, Sunyaev SR. A method and server for predicting damaging missense mutations. *Nat Methods* 2010;**7**:248–9.
- 50 Kumar S, Suleski MP, Markov GJ, Lawrence S, Marco A, Filipinski AJ. Positional conservation and amino acids shape the correct diagnosis and population frequencies of benign and damaging personal amino acid mutations. *Genome Res* 2009;**19**:1562–9.
- 51 Pollard KS, Hubisz MJ, Rosenbloom KR, Siepel A. Detection of nonneutral substitution rates on mammalian phylogenies. *Genome Res* 2010;**20**:110–21.
- 52 van der Velde KJ, Kuiper J, Thompson BA, Plazzer J-P, van Valkenhoef G, de Haan M, Jongbloed JDH, Wijmenga C, de Koning TJ, Abbott KM, Sinke R, Spurdle AB, Macrae F, Genuardi M, Sijmons RH, Swertz MA, InSiGHT Group. Evaluation of CADD Scores in Curated Mismatch Repair Gene Variants Yields a Model for Clinical Validation and Prioritization. *Hum Mutat* 2015;**36**:712–9.
- 53 Bragin E, Chatzimichali EA, Wright CF, Hurles ME, Firth H V., Bevan AP, Swaminathan GJ. DECIPHER: database for the interpretation of phenotype-linked plausibly pathogenic sequence and copy-number variation. *Nucleic Acids Res* 2014;**42**:D993–1000.
- 54 Lancaster O, Beck T, Atlán D, Swertz M, Thangavelu D, Veal C, Dalglish R, Brookes AJ. Cafe Variome: general-purpose software for making genotype-phenotype data discoverable in restricted or open access contexts. *Hum Mutat* 2015;**36**:957–64.
- 55 Sobreira N, Schiettecatte F, Valle D, Hamosh A. GeneMatcher: a matching tool for connecting investigators with an interest in the same gene. *Hum Mutat* 2015;**36**:928–30.

- 56 Thompson R, Johnston L, Taruscio D, Monaco L, Bérout C, Gut IG, Hansson MG, 't Hoen P-BA, Patrinos GP, Dawkins H, Ensini M, Zatloukal K, Koubi D, Heslop E, Paschall JE, Posada M, Robinson PN, Bushby K, Lochmüller H. RD-Connect: an integrated platform connecting databases, registries, biobanks and clinical bioinformatics for rare disease research. *J Gen Intern Med* 2014;**29 Suppl 3**:S780-7.
- 57 Buske OJ, Girdea M, Dumitriu S, Gallinger B, Hartley T, Trang H, Misyura A, Friedman T, Beaulieu C, Bone WP, Links AE, Washington NL, Haendel MA, Robinson PN, Boerkoel CF, Adams D, Gahl WA, Boycott KM, Brudno M. PhenomeCentral: a portal for phenotypic and genotypic matchmaking of patients with rare genetic diseases. *Hum Mutat* 2015;**36**:931–40.
- 58 Philippakis AA, Azzariti DR, Beltran S, Brookes AJ, Brownstein CA, Brudno M, Brunner HG, Buske OJ, Carey K, Doll C, Dumitriu S, Dyke SOM, den Dunnen JT, Firth H V, Gibbs RA, Girdea M, Gonzalez M, Haendel MA, Hamosh A, Holm IA, Huang L, Hurles ME, Hutton B, Krier JB, Misyura A, Mungall CJ, Paschall J, Paten B, Robinson PN, Schiettecatte F, Sobreira NL, Swaminathan GJ, Taschner PE, Terry SF, Washington NL, Züchner S, Boycott KM, Rehm HL. The Matchmaker Exchange: a platform for rare disease gene discovery. *Hum Mutat* 2015;**36**:915–21.
- 59 Köhler S, Doelken SC, Mungall CJ, Bauer S, Firth H V, Bailleul-Forestier I, Black GCM, Brown DL, Brudno M, Campbell J, FitzPatrick DR, Eppig JT, Jackson AP, Freson K, Girdea M, Helbig I, Hurst JA, Jähn J, Jackson LG, Kelly AM, Ledbetter DH, Mansour S, Martin CL, Moss C, Mumford A, Ouwehand WH, Park S-M, Riggs ER, Scott RH, Sisodiya S, Van Vooren S, Wapner RJ, Wilkie AOM, Wright CF, Vulto-van Silfhout AT, de Leeuw N, de Vries BBA, Washington NL, Smith CL, Westerfield M, Schofield P, Ruef BJ, Gkoutos G V, Haendel M,

- Smedley D, Lewis SE, Robinson PN. The Human Phenotype Ontology project: linking molecular biology and disease through phenotype data. *Nucleic Acids Res* 2014;**42**:D966-74.
- 60 Richards S, Aziz N, Bale S, Bick D, Das S, Gastier-Foster J, Grody WW, Hegde M, Lyon E, Spector E, Voelkerding K, Rehm HL, ACMG Laboratory Quality Assurance Committee. Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. *Genet Med* 2015;**17**:405–24.
- 61 Facio FM, Sapp JC, Linn A, Biesecker LG. Approaches to informed consent for hypothesis-testing and hypothesis-generating clinical genomics research. *BMC Med Genomics* 2012;**5**:45.
- 62 Green RC, Berg JS, Berry GT, Biesecker LG, Dimmock DP, Evans JP, Grody WW, Hegde MR, Kalia S, Korf BR, Krantz I, McGuire AL, Miller DT, Murray MF, Nussbaum RL, Plon SE, Rehm HL, Jacob HJ. Exploring concordance and discordance for return of incidental findings from clinical sequencing. *Genet Med* 2012;**14**:405–10.
- 63 Ayuso C, Millán JM, Mancheño M, Dal-Ré R. Informed consent for whole-genome sequencing studies in the clinical setting. Proposed recommendations on essential content and process. *Eur J Hum Genet* 2013;**21**:1054–9.
- 64 Townsend A, Adam S, Birch PH, Lohn Z, Rousseau F, Friedman JM. ‘I want to know what’s in Pandora’s Box’: comparing stakeholder perspectives on incidental findings in clinical whole genomic sequencing. *Am J Med Genet A* 2012;**158A**:2519–25.
- 65 Davey S. Next generation sequencing: considering the ethics. *Int J Immunogenet* 2014;**41**:457–62.
- 66 Krabbenborg L, Vissers LELM, Schieving J, Kleefstra T, Kamsteeg EJ, Veltman JA,

- Willemsen MA, Van der Burg S. Understanding the Psychosocial Effects of WES Test Results on Parents of Children with Rare Diseases. *J Genet Couns* 2016;**25**:1207–14.
- 67 Pinxten W, Howard HC. Ethical issues raised by whole genome sequencing. *Best Pract Res Clin Gastroenterol* 2014;**28**:269–79.
- 68 Gahl WA, Markello TC, Toro C, Fajardo KF, Sincan M, Gill F, Carlson-Donohoe H, Gropman A, Pierson TM, Golas G, Wolfe L, Groden C, Godfrey R, Nehrebecky M, Wahl C, Landis DMD, Yang S, Madeo A, Mullikin JC, Boerkoel CF, Tiftt CJ, Adams D. The National Institutes of Health Undiagnosed Diseases Program: insights into rare diseases. *Genet Med* 2012;**14**:51–9.
- 69 Gahl WA, Mulvihill JJ, Toro C, Markello TC, Wise AL, Ramoni RB, Adams DR, Tiftt CJ, UDN. The NIH Undiagnosed Diseases Program and Network: Applications to modern medicine. *Mol Genet Metab* 2016;**117**:393–400.
- 70 Brownstein CA, Holm IA, Ramoni R, Goldstein DB. Data Sharing in the Undiagnosed Diseases Network. *Hum Mutat* 2015;**36**:985–8.
- 71 Taruscio D, Groft SC, Cederroth H, Melegh B, Lasko P, Kosaki K, Baynam G, McCray A, Gahl WA. Undiagnosed Diseases Network International (UDNI): White paper for global actions to meet patient needs. *Mol Genet Metab* 2015;**116**:223–5.
- 72 Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* 2009;**25**:1754–60.
- 73 MacArthur DG, Manolio TA, Dimmock DP, Rehm HL, Shendure J, Abecasis GR, Adams DR, Altman RB, Antonarakis SE, Ashley EA, Barrett JC, Biesecker LG, Conrad DF, Cooper GM, Cox NJ, Daly MJ, Gerstein MB, Goldstein DB, Hirschhorn JN, Leal SM, Pennacchio LA, Stamatoyannopoulos JA, Sunyaev SR, Valle D, Voight BF, Winckler W,

- Gunter C. Guidelines for investigating causality of sequence variants in human disease. *Nature* 2014;**508**:469–76.
- 74 Sobreira N, Schiettecatte F, Valle D, Hamosh A. GeneMatcher: A Matching Tool for Connecting Investigators with an Interest in the Same Gene. *Hum Mutat* 2015;**36**:928–30.
- 75 Biesecker LG, Green RC. Diagnostic clinical genome and exome sequencing. *N Engl J Med* 2014;**371**:1170.
- 76 Adams SA, Coppinger J, Saitta SC, Stroud T, Kandamurugu M, Fan Z, Ballif BC, Shaffer LG, Bejjani BA. Impact of genotype-first diagnosis: the detection of microdeletion and microduplication syndromes with cancer predisposition by aCGH. *Genet Med* 2009;**11**:314–22.
- 77 Pichert G, Mohammed SN, Ahn JW, Ogilvie CM, Izatt L. Unexpected findings in cancer predisposition genes detected by array comparative genomic hybridisation: what are the issues? *J Med Genet* 2011;**48**:535–9.
- 78 Green RC, Berg JS, Grody WW, Kalia SS, Korf BR, Martin CL, McGuire AL, Nussbaum RL, O’Daniel JM, Ormond KE, Rehm HL, Watson MS, Williams MS, Biesecker LG, American College of Medical Genetics and Genomics. ACMG recommendations for reporting of incidental findings in clinical exome and genome sequencing. *Genet Med* 2013;**15**:565–74.
- 79 Lipina T V, Prasad T, Yokomaku D, Luo L, Connor SA, Kawabe H, Wang YT, Brose N, Roder JC, Craig AM. Cognitive Deficits in Calsyntenin-2-deficient Mice Associated with Reduced GABAergic Transmission. *Neuropsychopharmacology* 2016;**41**:802–10.
- 80 Haijes HA, Koster MJE, Rehmann H, Li D, Hakonarson H, Cappuccio G, Hancarova M, Lehalle D, Reardon W, Schaefer GB, Lehman A, van de Laar IMBH, Tesselaar CD, Turner C, Goldenberg A, Patrier S, Thevenon J, Pinelli M, Brunetti-Pierri N, Prchalová D, Havlovicová

- M, Vlckova M, Sedláček Z, Lopez E, Ragoussis V, Pagnamenta AT, Kini U, Vos HR, van Es RM, van Schaik RFMA, van Essen TAJ, Kibaek M, Taylor JC, Sullivan J, Shashi V, Petrovski S, Fagerberg C, Martin DM, van Gassen KLI, Pfundt R, Falk MJ, McCormick EM, Timmers HTM, van Hasselt PM. De Novo Heterozygous POLR2A Variants Cause a Neurodevelopmental Syndrome with Profound Infantile-Onset Hypotonia. *Am J Hum Genet* 2019;**105**:283–301.
- 81 Caputo V, Cianetti L, Niceta M, Carta C, Ciolfi A, Bocchinfuso G, Carrani E, Dentici ML, Biamino E, Belligni E, Garavelli L, Boccone L, Melis D, Andria G, Gelb BD, Stella L, Silengo M, Dallapiccola B, Tartaglia M. A Restricted Spectrum of Mutations in the SMAD4 Tumor-Suppressor Gene Underlies Myhre Syndrome. *Am J Hum Genet* 2012;**90**:161–9.
- 82 Michot C, Le Goff C, Mahaut C, Afenjar A, Brooks AS, Campeau PM, Destree A, Di Rocco M, Donnai D, Hennekam R, Heron D, Jacquemont S, Kannu P, Lin AE, Manouvrier-Hanu S, Mansour S, Marlin S, McGowan R, Murphy H, Raas-Rothschild A, Rio M, Simon M, Stolte-Dijkstra I, Stone JR, Sznajder Y, Tolmie J, Touraine R, van den Ende J, Van der Aa N, van Essen T, Verloes A, Munnich A, Cormier-Daire V. Myhre and LAPS syndromes: clinical and molecular review of 32 patients. *Eur J Hum Genet* 2014;**22**:1272–7.
- 83 Lindor NM, Gunawardena SR, Thibodeau SN. Mutations of SMAD4 account for both LAPS and Myhre syndromes. *Am J Med Genet Part A* 2012;**158A**:1520–1.
- 84 Garavelli L, Maini I, Baccilieri F, Ivanovski I, Pollazzon M, Rosato S, Iughetti L, Unger S, Superti-Furga A, Tartaglia M. Natural history and life-threatening complications in Myhre syndrome and review of the literature. *Eur J Pediatr* 2016;**175**:1307–15.
- 85 Lin AE, Michot C, Cormier-Daire V, L’Ecuyer TJ, Matherne GP, Barnes BH, Humberson JB, Edmondson AC, Zackai E, O’Connor MJ, Kaplan JD, Ebeid MR, Krier J,

- Krieg E, Ghoshhajra B, Lindsay ME. Gain-of-function mutations in SMAD4 cause a distinctive repertoire of cardiovascular phenotypes in patients with Myhre syndrome. *Am J Med Genet A* 2016;**170**:2617–31.
- 86 Christensen AE, Fiskerstrand T, Knappskog PM, Boman H, Rødahl E. A novel ADAMTSL4 mutation in autosomal recessive ectopia lentis et pupillae. *Invest Ophthalmol Vis Sci* 2010;**51**:6369–73.
- 87 Le Goff C, Mahaut C, Abhyankar A, Le Goff W, Serre V, Afenjar A, Destrée A, di Rocco M, Héron D, Jacquemont S, Marlin S, Simon M, Tolmie J, Verloes A, Casanova J-L, Munnich A, Cormier-Daire V. Mutations at a single codon in Mad homology 2 domain of SMAD4 cause Myhre syndrome. *Nat Genet* 2011;**44**:85–8.
- 88 Piccolo P, Mithbaokar P, Sabatino V, Tolmie J, Melis D, Schiaffino MC, Filocamo M, Andria G, Brunetti-Pierri N. SMAD4 mutations causing Myhre syndrome result in disorganization of extracellular matrix improved by losartan. *Eur J Hum Genet* 2014;**22**:988–94.
- 89 Piccolo P, Sabatino V, Mithbaokar P, Polishchuk E, Hicks J, Polishchuk R, Bacino CA, Brunetti-Pierri N. Skin fibroblasts of patients with geleophysic dysplasia due to FBN1 mutations have lysosomal inclusions and losartan improves their microfibril deposition defect. *Mol Genet genomic Med* 2019;**7**:e844.
- 90 Guadagno NA, Progida C. Rab GTPases: Switching to Human Diseases. *Cells* 2019;**8**. doi:10.3390/cells8080909
- 91 Müller MP, Goody RS. Molecular control of Rab activity by GEFs, GAPs and GDI. *Small GTPases* 2018;**9**:5–21.
- 92 Stenmark H, Parton RG, Steele-Mortimer O, Lütcke A, Gruenberg J, Zerial M. Inhibition

- of rab5 GTPase activity stimulates membrane fusion in endocytosis. *EMBO J* 1994;**13**:1287–96.
- 93 English AR, Voeltz GK. Rab10 GTPase regulates ER dynamics and morphology. *Nat Cell Biol* 2013;**15**:169–78.
- 94 Mignogna ML, D’Adamo P. Critical importance of RAB proteins for synaptic function. *Small GTPases* 2018;**9**:145–57.
- 95 Wang T, Liu Y, Xu X-H, Deng C-Y, Wu K-Y, Zhu J, Fu X-Q, He M, Luo Z-G. Lgl1 Activation of Rab10 Promotes Axonal Membrane Trafficking Underlying Neuronal Polarization. *Dev Cell* 2011;**21**:431–44.
- 96 Taylor CA, Yan J, Howell AS, Dong X, Shen K. RAB-10 Regulates Dendritic Branching by Balancing Dendritic Transport. *PLOS Genet* 2015;**11**:e1005695.
- 97 Thiffault I, Lantos J. The Challenge of Analyzing the Results of Next-Generation Sequencing in Children. *Pediatrics* 2016;**137 Suppl**:S3-7.
- 98 Grody WW, Thompson BH, Hudgins L. Whole-exome/genome sequencing and genomics. *Pediatrics* 2013;**132**:S211-5.
- 99 Rabbani B, Tekin M, Mahdiah N. The promise of whole-exome sequencing in medical genetics. *J Hum Genet* 2014;**59**:5–15.
- 100 Gonorazky HD, Naumenko S, Ramani AK, Nelakuditi V, Mashouri P, Wang P, Kao D, Ohri K, Viththiyapaskaran S, Tarnopolsky MA, Mathews KD, Moore SA, Osorio AN, Villanova D, Kemaladewi DU, Cohn RD, Brudno M, Dowling JJ. Expanding the Boundaries of RNA Sequencing as a Diagnostic Tool for Rare Mendelian Disease. *Am J Hum Genet* 2019;**104**:1007.
- 101 Cummings BB, Marshall JL, Tukiainen T, Lek M, Donkervoort S, Foley AR, Bolduc V,

Waddell LB, Sandaradura SA, O'Grady GL, Estrella E, Reddy HM, Zhao F, Weisburd B, Karczewski KJ, O'Donnell-Luria AH, Birnbaum D, Sarkozy A, Hu Y, Gonorazky H, Claeys K, Joshi H, Bournazos A, Oates EC, Ghaoui R, Davis MR, Laing NG, Topf A, Genotype-Tissue Expression Consortium, Kang PB, Beggs AH, North KN, Straub V, Dowling JJ, Muntoni F, Clarke NF, Cooper ST, Bönnemann CG, MacArthur DG. Improving genetic diagnosis in Mendelian disease with transcriptome sequencing. *Sci Transl Med* 2017;**9**.

doi:10.1126/scitranslmed.aal5209

102 Wrighton KH. Genetic testing: The diagnostic power of RNA-seq. *Nat Rev Genet* 2017;**18**:392–3.

103 Marchuk DS, Crooks K, Strande N, Kaiser-Rogers K, Milko L V, Brandt A, Arreola A, Tilley CR, Bizon C, Vora NL, Wilhelmsen KC, Evans JP, Berg JS. Increasing the diagnostic yield of exome sequencing by copy number variant analysis. *PLoS One* 2018;**13**:e0209185.

104 Scala I, Parenti G, Andria G. Universal screening for inherited metabolic diseases in the neonate (and the fetus). *J Matern Fetal Neonatal Med* 2012;**25**:4–6.