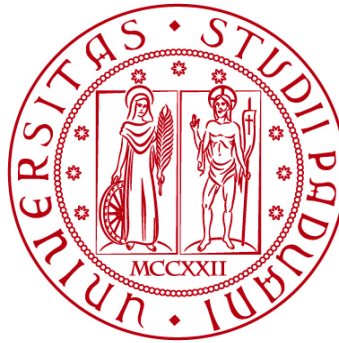


UNIVERSITÀ DEGLI STUDI DI PADOVA

DIPARTIMENTO DI BIOLOGIA

Corso di Laurea magistrale in Biologia Sanitaria



TESI DI LAUREA

IMPATTO DELL'ANALISI DELLE *COPY NUMBER VARIATIONS* (CNVs) NELLA DIAGNOSI MOLECOLARE DEI DISORDINI DEL NEUROSVILUPPO

**Relatore: Prof. Giovanni Vazza
Dipartimento di Biologia**

Laureanda: Eleonora Chiarotto

ANNO ACCADEMICO 2021/2022

Sommario

| | |
|---|-----------|
| Abstract | 2 |
| Introduzione | 3 |
| I disordini del neurosviluppo | 3 |
| Le <i>copy number variations</i> | 8 |
| Presentazione del progetto | 11 |
| Materiali e metodi..... | 13 |
| Sequenziamento dell'esoma e allineamento al genoma di riferimento | 13 |
| Analisi bioinformatica | 14 |
| Ricerca nei database e in letteratura | 17 |
| Validazione in laboratorio | 18 |
| <i>Real-time PCR</i> | 18 |
| <i>Array-CGH</i> | 22 |
| Risultati..... | 27 |
| Discussione..... | 31 |
| <i>Pazienti 1 e 8</i> | 33 |
| <i>Paziente 2</i> | 37 |
| <i>Paziente 3</i> | 39 |
| <i>Pazienti 4 e 6</i> | 42 |
| <i>Paziente 5</i> | 46 |
| <i>Paziente 7</i> | 47 |
| <i>Paziente 9</i> | 49 |
| <i>Paziente 10</i> | 50 |
| Bibliografia | 54 |
| Appendice | 57 |

Abstract

I disordini del neurosviluppo (NDDs) sono una classe di patologie ad esordio precoce che colpiscono la crescita e la funzionalità cerebrale e interessano più del 3% dei bambini su scala mondiale. Questi disordini presentano un'estrema eterogeneità sia clinica che genetica.

Nonostante l'impiego delle tecniche di sequenziamento di nuova generazione abbia portato alla possibilità di analizzare molti geni contemporaneamente, l'analisi delle sole varianti a singolo nucleotide (SNVs) dei geni implicati nei NDDs non porta alla diagnosi molecolare in tutti i pazienti.

Un'ulteriore categoria di varianti importanti nel contesto di questi disturbi è quella delle *copy number variations* (CNVs). Si tratta di sequenze di DNA ripetute, che variano nel numero di copie in individui diversi.

Lo scopo di questo progetto è stato quello di determinare l'impatto dell'analisi delle CNVs in pazienti con NDDs.

È stata condotta l'analisi delle CNVs su 100 esomi di pazienti affetti da NDDs, negativi alla precedente ricerca di SNVs mediante *whole exome sequencing* (WES). L'analisi bioinformatica ha portato all'individuazione di una variante CNV significativa in 10 dei pazienti analizzati, confermata successivamente mediante *real-time* PCR o *array-CGH*. Questo lavoro ha portato ad un aumento del 10% della *detection rate* nei pazienti analizzati, confermando l'importanza dell'analisi delle CNVs nei NDDs.

Introduzione

I disordini del neurosviluppo

Il DSM-5¹, ovvero l'ultima edizione del Manuale Diagnostico e Statistico dei disturbi mentali pubblicata dall'*American Psychiatric Association* nel 2013, include all'interno della classe dei disordini del neurosviluppo (NDDs²) quei disturbi che emergono precocemente nel periodo dello sviluppo, durante l'infanzia o l'adolescenza, e caratterizzati da deficit di gravità variabile che portano a difficoltà nel funzionamento personale, sociale, accademico o professionale.

Questi disordini comprendono:

- Disabilità intellettiva (disturbo dello sviluppo intellettivo): comporta deficit sia a livello della capacità cognitiva (attività quali ragionare, risolvere problemi, pianificare, giudicare, apprendere, pensare in modo astratto) che del funzionamento adattivo (attività della vita quotidiana quali comunicare, vivere in modo indipendente, vivere all'interno di un contesto sociale);
- Disturbi della comunicazione: determinano una difficoltà nella comprensione, produzione e uso del linguaggio verbale e non verbale, e/o nella comunicazione sociale;
- Disturbo dello spettro autistico: è definito da deficit nella comunicazione e nell'interazione sociale, e da comportamenti, interessi e attività ristretti e ripetitivi;
- Disturbo da deficit d'attenzione/iperattività: è caratterizzato da inattenzione (incapacità di concentrarsi) e/o iperattività (movimento eccessivo e fuori contesto) e impulsività (agire in modo frettoloso e senza pensare) eccessive per l'età o il livello di sviluppo;
- Disturbi specifici dell'apprendimento: comportano deficit specifici nella capacità di un individuo di percepire o elaborare in modo efficiente e accurato le informazioni. Si manifestano come difficoltà persistenti nell'ambito dell'apprendimento e dell'uso delle abilità scolastiche di lettura, scrittura e/o matematica;

¹ *Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition.*

² *Neurodevelopmental Disorders.*

- Disturbi del movimento: coinvolgono l'abilità motoria e comprendono il disturbo dello sviluppo della coordinazione e il disturbo da movimento stereotipato, oltre ai disturbi da tic.

In generale, i NDDs interessano più del 3% dei bambini nel mondo [1], e vengono diagnosticati più frequentemente nei maschi rispetto alle femmine [2].

Essi presentano un'estrema eterogeneità sia clinica che genetica [2]. L'eterogeneità genetica si manifesta quando mutazioni in geni diversi possono tradursi nello stesso fenotipo clinico; in molti casi questi geni appartengono alla stessa struttura cellulare o allo stesso *pathway* molecolare. L'eterogeneità clinica, invece, è presente quando mutazioni diverse nello stesso gene possono dare origine a due o più malattie differenti.

Nonostante quest'enorme variabilità genetica, la maggior parte delle varianti di rischio associate ai NDDs interessa geni che hanno un ruolo in un ristretto numero di *pathways* molecolari molto conservati [3]. Quest'ultimi risultano coinvolti nel mantenimento dell'equilibrio omeostatico durante lo sviluppo neurologico e comprendono in particolare tre processi biologici critici: la regolazione della sintesi proteica, la regolazione trascrizionale ed epigenetica e la segnalazione sinaptica [3].

La presenza di suddetti *pathways* molecolari condivisi rende ragione del fatto che frequentemente questi disturbi siano caratterizzati da comorbidità [2][3], per cui due o più di essi possono coesistere nello stesso individuo, e da *overlap* fenotipico [2], per cui alcune loro caratteristiche fenotipiche si sovrappongono.

Quanto detto implica che i NDDs (così come altri disturbi psichiatrici quali la schizofrenia e il disturbo bipolare), piuttosto che entità discrete, dovrebbero essere percepiti come parte di uno spettro, il cosiddetto *neurodevelopmental continuum* (**Figura 1**) [2]. Secondo la *neurodevelopmental gradient hypothesis*, gli stessi disordini sono classificati in accordo con la gravità del deficit dello sviluppo neurologico [2].

Gli attuali dati in letteratura indicano una natura multifattoriale o poligenica di questi disordini [3], sono cioè disordini causati da una combinazione di più eventi genetici, ognuno con un piccolo effetto sulla costituzione fenotipica, rispettivamente con o senza il contributo di fattori ambientali. Ne consegue che i NDDs con natura monogenica, cioè determinati da una semplice causa genetica,

da una o due varianti patogenetiche in un singolo gene, sono solo una piccola frazione dello spettro di queste malattie [2].

Si può pertanto dedurre che la correlazione genotipo-fenotipo sia difficile da interpretare. Nonostante ciò, una diagnosi molecolare precoce è essenziale per quanto riguarda la consulenza genetica e la gestione del paziente. Infatti, l'identificazione della possibile causa genetica alla base del fenotipo malattia può migliorare la valutazione del rischio di ricorrenza e dare la possibilità di prevedere future complicazioni mediche [3].

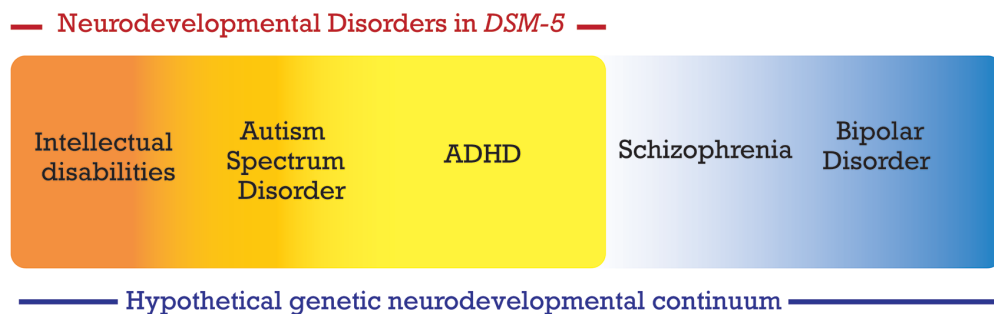


Figura 1. Spettro del neurodevelopmental continuum. Secondo questo modello i NDDs e altri disturbi psichiatrici rappresentano i diversi range di outcomes che si manifestano in seguito all'alterazione dello sviluppo e del funzionamento neurologico. Abbreviazione: ADHD, Attention-Deficit/Hyperactivity Disorder [2].

I concetti principali da tenere in considerazione per studiare l'effetto delle varianti genetiche sul fenotipo sono due:

1. Vulnerabilità genetica: consiste nella capacità di un gene di tollerare mutazioni potenzialmente deleterie. Minore è questa tolleranza, maggiore è il livello di vulnerabilità;
2. Carico mutazionale: rappresenta il peso genetico dato dall'insieme degli eventi mutazionali deleteri.

In generale, è presente una correlazione positiva tra l'aumento di ciascuno di questi due fattori e la complessità della manifestazione fenotipica (**Figura 2**) [3].

Alcuni geni associati ai NDDs causano un fenotipo patologico in seguito al meccanismo dell'aploinsufficienza. Questi geni sono estremamente sensibili al dosaggio [3], per cui anche una riduzione del 50% del prodotto proteico da loro codificato può dare origine ad un fenotipo patologico. Si tratta di prodotti genici che fanno parte di sistemi quantitativi di segnalazione, o che competono tra di loro

nel determinare il passaggio ad uno stadio di sviluppo, o sono coinvolti nel metabolismo. Questi geni sono geni altamente vulnerabili e, se colpiti da mutazioni deleterie, c'è una probabilità significativa che portino allo sviluppo di un fenotipo patologico causando forme monogeniche di NDDs. Varianti in questi geni altamente vulnerabili sono normalmente sotto una forte pressione selettiva negativa e sono categorizzabili come varianti rare associate ad un significativo rischio di sviluppare la malattia e ad un'elevata penetranza [3].

L'altra parte dello spettro della vulnerabilità comprende geni meno sensibili a mutazioni deleterie, per cui varianti in questi geni non sono sottoposte ad una pressione selettiva negativa e sono più frequentemente trasmesse nelle famiglie per generazioni. Singoli eventi mutazionali che colpiscono geni non vulnerabili non sono di per sé causa di malattia e ricadono quindi nella categoria di varianti comuni associate ad un basso rischio malattia [3].

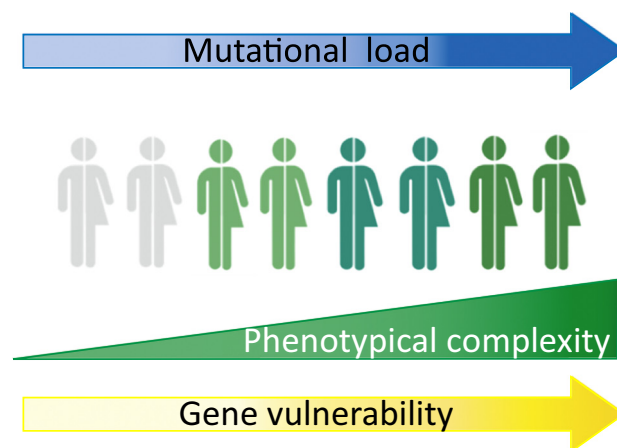


Figura 2. Vulnerabilità genetica e carico mutazionale. Sono concetti importanti per valutare l'effetto degli eventi mutazionali sul fenotipo: maggiore è ciascuno di questi due fattori, maggiore è la complessità dell'outcome fenotipico [3].

I NDDs si associano ad una ridotta fecondità [4], per cui la frequenza delle varianti rare associate ad un elevato rischio malattia per l'insorgenza di questa classe di patologie è funzione della pressione selettiva negativa a cui sono sottoposte e del tasso di insorgenza di mutazioni *de novo* [2].

Infatti, nella maggior parte dei casi (ad esclusione dei matrimoni consanguinei) questi disturbi sono dovuti a mutazioni che insorgono *de novo* nella prole di genitori non affetti, e si presentano come dominanti (**Figura 3A**) [3].

In alcuni casi, però, il probando eredita un allele mutato da ciascuno dei genitori, i quali risultano non affetti, per cui si osserva un'ereditarietà di tipo autosomico recessivo (**Figura 3B**) [3].

In altri casi più varianti genetiche possono essere responsabili dell'esordio del fenotipo patologico (**Figura 3 C, D, E**) [3].

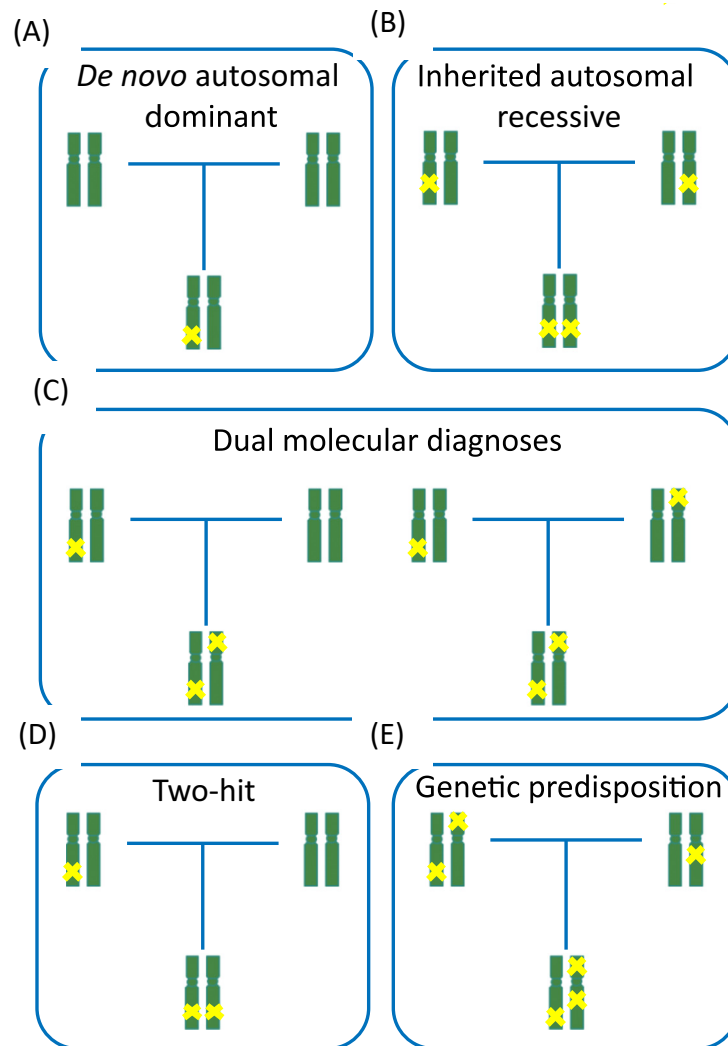


Figura 3. Meccanismi genetici responsabili dell'esordio dei NDDs. (A) La maggior parte delle cause genetiche dei NDDs è legata a mutazioni de novo; (B) in alcuni casi si osserva un'ereditarietà di tipo autosomico recessivo; (C) in altri casi ancora sono presenti mutazioni in diversi loci che possono essere insorte de novo o ereditate da uno o entrambi i genitori; (D) nel two-hit model una prima mutazione ereditata predispone il probando all'accumulo di una seconda mutazione somatica, la quale è poi responsabile della completa manifestazione fenotipica; (E) anche l'accumulo di varianti comuni ereditate può rendere il probando più vulnerabile all'esordio di NDDs [3].

I tipi di mutazione alla base dei NDDs possono essere molteplici: riarrangiamenti cromosomici, mutazioni puntiformi, piccole *indels* e *copy number variations* [3].

Le copy number variations

Le *copy number variations* (CNVs) sono una classe di sequenze che variano nel numero di copie. Si tratta di sequenze genomiche ripetute, il cui numero di ripetizioni è diverso tra individui diversi della stessa specie [5].

Si stima che le CNVs rappresentino tra il 4.8% e il 9.5% del genoma [6].

Nello specifico, questi riarrangiamenti strutturali del genoma possono presentarsi sotto forma di delezioni o di duplicazioni e, sebbene sia dibattuto in letteratura se debba o meno essere fatta una distinzione tra le piccole *indels* e le CNVs in base alle loro dimensioni, attualmente quest'ultime sono definite come varianti che spaziano da 50 bp a diverse Mb [5].

La loro formazione può dipendere sia da fattori genomici, ossia da diversi meccanismi associati alla ricombinazione, replicazione e riparazione del DNA, che da mutageni chimici e fisici (come le radiazioni) [5].

Per quanto riguarda i fattori ambientali, le radiazioni inducono la formazione di duplicazioni e delezioni in modo equivalente, impattando su loci in regioni casuali del genoma. Al contrario, i mutageni chimici inducono più frequentemente perdite del numero di copie rispetto ad una loro acquisizione, e hanno come *target* specifiche regioni genomiche [7].

Facendo invece riferimento ai meccanismi genomici alla base dell'origine delle CNVs, assume particolare rilevanza il fenomeno della ricombinazione omologa non allelica (NAHR³) [8]. Infatti, come affermato precedentemente, le CNVs sono dei riarrangiamenti strutturali del genoma e, come tali, possono originare da eventi di ricombinazione, quindi processi che prevedono rotture e unioni di molecole di DNA. Nello specifico, la NAHR si verifica quando la ricombinazione omologa, cioè lo scambio di materiale genetico tra due sequenze di DNA con un elevato grado di similarità, avviene tra sequenze non alleliche dei due cromosomi omologhi (intercromosomica), di due cromatidi fratelli (intracromosomica o intercromatidica) o di un singolo cromatidio (intracromatidica). Generalmente, essa avviene in presenza di sequenze di DNA ripetute, le cosiddette *Low-Copy Repeats* (LCRs) o duplicazioni segmentali, le quali risultano arricchite in prossimità dei punti di rottura. Si tratta di sequenze di DNA con una lunghezza stimata tra 10 e 300 kb e con un elevato grado di omologia (> 95%), presenti in due o più copie

³ *Non-Allelic Homologous Recombination*.

nel genoma. Grazie all'alto grado di similarità, copie non alleliche di LCRs possono allinearsi durante la meiosi, portando ad uno scambio ineguale di materiale genetico. Se le due copie paraloghe di LCRs sono presenti con orientamento diretto l'una rispetto all'altra, questo fenomeno di crossing-over ineguale può dare origine a reciproche delezioni e duplicazioni della sequenza unica tra le due sequenze LCRs se intercromosomica o intercromatidica, oppure esclusivamente a delezioni se intracromatidica. Di conseguenza, la frequenza delle delezioni generate tramite NAHR risulta essere più alta di quella delle duplicazioni, e la differenza tra queste due frequenze riflette a sua volta la frequenza della NAHR intracromatidica.

Siccome molte LCRs sono concentrate nel genoma umano a livello di *hot spots* di ricombinazione condivisi, queste regioni risultano più predisposte al fenomeno della NAHR e pertanto allo sviluppo di CNVs ricorrenti. Quest'ultime presentano la stessa dimensione e punti di rottura comuni arricchiti di LCRs, e sono presenti indipendentemente in diversi individui non imparentati tra loro. Al contrario, le CNVs non ricorrenti differiscono in dimensione e hanno diversi punti di rottura, ma possono anche condividere una più piccola regione di sovrapposizione (SRO⁴) tra pazienti differenti (**Figura 4**) [8].

Un meccanismo non replicativo che può portare alla formazione di semplici CNVs non ricorrenti è la giunzione non omologa delle estremità (NHEJ⁵) [8]. È un meccanismo di riparazione che interviene in caso di rottura a doppio filamento del DNA e che non richiede un'estesa omologia tra le sequenze delle due estremità libere per guidarne l'unione. Si riconoscono 4 step: identificazione del danno a livello del DNA, creazione di un ponte molecolare tra le due estremità libere, processamento delle estremità libere per renderle compatibili e complementari, ligazione finale delle due estremità. La terza fase può portare all'aggiunta o alla rimozione di nucleotidi alle estremità, lasciando una "firma" della NHEJ. È pertanto un meccanismo prone ad errori, che risulta essere poco preciso e accurato e può portare ad un cambiamento della sequenza originale di DNA quando le due estremità libere di DNA da unire non sono perfettamente compatibili.

Un meccanismo replicativo che può portare alla formazione di complesse CNVs non ricorrenti è il FoSTeS (*Fork Stalling and Template Switching*) [8], generalizzato e spiegato nei dettagli a livello molecolare con il modello della replicazione indotta

⁴ *Smallest Region of Overlap.*

⁵ *Non-Homologous End Joining.*

da rottura mediata da microomologia (MMBIR⁶). Questo modello prevede che, dopo lo stallo iniziale della forca di replicazione a causa di un danno al DNA, l'estremità 3' del filamento lento possa dissociarsi dal suo filamento stampo e appaiarsi su un altro filamento stampo, tramite microomologia, di un'altra forca di replicazione, continuando la sintesi del DNA. La dissociazione del filamento lento e l'invasione di altre forche di replicazione può verificarsi più volte prima della ripresa della replica sul modello originale. In questo modo, il prodotto finale contiene sequenze provenienti da diverse regioni genomiche, quindi si possono creare riarrangiamenti complessi, e, a seconda che il ritorno sul filamento stampo originale avvenga davanti o dietro la posizione della dissociazione iniziale, si può generare una delezione o una duplicazione rispettivamente.

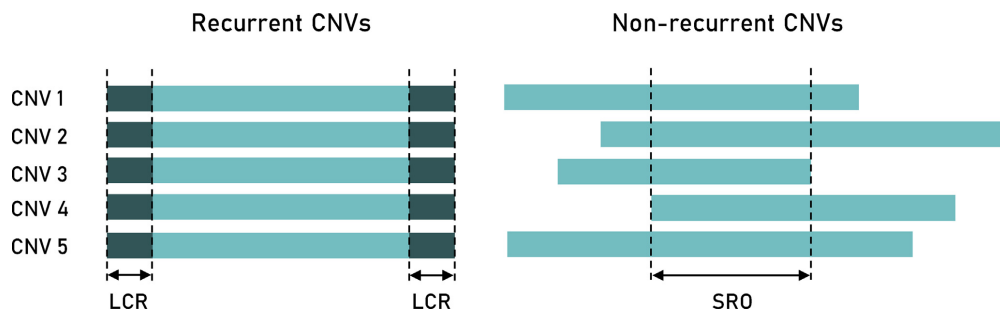


Figura 4. Confronto tra CNVs ricorrenti e non ricorrenti. Le CNVs ricorrenti hanno la stessa dimensione e punti di rottura comuni, e sono presenti in individui diversi. Le CNVs non ricorrenti differiscono nella dimensione e hanno diversi punti di rottura, ma possono condividere una piccola regione di sovrapposizione (SRO) [5].

Le CNVs possono insorgere *de novo* o essere ereditate. Se ereditate, non sono eventi ricorrenti, ma condividono gli stessi punti di rottura e perciò hanno un effetto simile sul fenotipo. Se *de novo*, sono eventi indipendenti che possono avere punti di rottura ricorrenti o non ricorrenti; anche se i punti di rottura non sono ricorrenti, queste varianti possono manifestare effetti sovrapposti in quanto colpiscono la stessa SRO [5].

Le CNVs contribuiscono alla variabilità genetica tra gli individui, a quel tipo di variabilità che coinvolge cambiamenti nel numero di copie di una sequenza di DNA. Non di meno, esse possono anche avere un ruolo importante nello sviluppo

⁶ *Microhomology-Mediated Break-Induced Replication.*

di malattie, compresi i NDDs, qualora vadano a coinvolgere geni sensibili al dosaggio.

Nello specifico, a seconda delle dimensioni del segmento genomico coinvolto, della sua posizione e del contesto genomico, nonché del numero di geni sensibili al dosaggio all'interno del segmento riarrangiato, insieme ad altri fattori di rischio, le CNVs si possono riscontrare in disordini con ereditarietà mendeliana, sindromi da geni contigui, disordini cromosomici, o altri tratti sporadici o complessi [5].

Presentazione del progetto

Nell'ambito dei NDDs, la possibilità di effettuare una precisa correlazione genotipo-fenotipo rappresenta generalmente una sfida.

Negli ultimi anni l'avanzamento nelle tecniche di sequenziamento di nuova generazione (NGS⁷) ha permesso l'analisi di molti geni contemporaneamente, ma valutando soltanto varianti a singolo nucleotide (SNVs⁸) o piccole inserzioni o delezioni. La sola valutazione di queste varianti non giustifica il totale ammontare dei casi osservati con questi disturbi e, pertanto, molti pazienti rimangono privi di diagnosi molecolare.

Si pensa che le CNVs potrebbero quindi essere uno dei fattori responsabili almeno di una parte dei casi di NDDs.

Nel laboratorio R&I Genetics (Padova) in cui è stato svolto questo progetto di tesi, la sola analisi delle SNVs tramite WES per i casi di NDDs ha portato ad un successo diagnostico del 37%, lasciando un 10% di casi inconclusivi e un 53% di casi negativi, ovvero senza diagnosi molecolare.

Il presente progetto si inserisce in questo contesto, con lo scopo di valutare l'impatto dell'analisi delle CNVs sulla diagnosi molecolare di pazienti affetti da NDDs. Per fare questo, sono stati analizzati gli esomi di 100 pazienti affetti da NDDs e risultati negativi alla precedente ricerca di SNVs tramite WES.

I 100 esomi sono stati inizialmente sottoposti ad analisi bioinformatica per individuare l'eventuale presenza di CNVs in geni correlati ai NDDs. Le CNVs così identificate sono state priorizzate tramite l'utilizzo dei database dedicati, della

⁷ *Next Generation Sequencing.*

⁸ *Single Nucleotide Variants.*

letteratura scientifica e delle linee guida ACMG⁹ per la classificazione delle varianti genetiche.

Per 10 campioni analizzati è stata trovata una variante CNV classificata come una variante di significatività elevata, che è stata poi validata in laboratorio, tramite *real-time* PCR o *array*-CGH a seconda della dimensione dell'evento mutazionale da confermare.

Infine, per i 5 probandi per cui erano disponibili anche i genitori è stata eseguita un'analisi di segregazione al fine di valutare se la variante in esame fosse ad insorgenza *de novo* o ereditata.

⁹ American College of Medical Genetics.

Materiali e metodi

Sequenziamento dell'esoma e allineamento al genoma di riferimento

Sono stati selezionati 273 campioni di pazienti affetti da NDDs, di cui alcuni con la disponibilità dei genitori, per un totale di 463 campioni nel progetto. I 273 probandi sono stati selezionati tra i campioni risultati negativi alla precedente analisi delle SNVs tramite WES nei geni correlati ai NDDs. Rispetto al totale di 273 probandi nel progetto, è stato scelto di portare avanti l'analisi per un gruppo di 100 campioni; in particolare, essendo già disponibili i dati per i campioni più recenti, sono stati selezionati i 100 campioni più vecchi.

Si è partiti da campioni di DNA genomico (gDNA) estratto da sangue. L'estrazione, condotta in maniera completamente automatizzata con QIAGEN EZ1 Advanced XL, si è basata sull'impiego di particelle magnetiche secondo i seguenti passaggi: i) lisi dei campioni; ii) legame degli acidi nucleici alle particelle magnetiche; iii) e infine lavaggio ed eluizione del DNA. L'estratto è stato poi quantificato spettrofotometricamente tramite NanoDrop.

Il gDNA è stato sottoposto a sequenziamento dell'esoma mediante tecniche di NGS. Infatti, nonostante le regioni codificanti rappresentino solo l'1-2% del genoma, l'85% delle cause delle malattie mendeliane risiede in regioni esoniche. Il flusso di lavoro in un esperimento di NGS prevede in generale tre fasi: i) costruzione di una libreria di sequenziamento; ii) sequenziamento vero e proprio; iii) e analisi bioinformatica dei dati ottenuti dal sequenziamento. Una libreria di sequenziamento contiene i frammenti derivati dalla frammentazione del gDNA e coniugati ad entrambe le estremità a delle sequenze definite adattatori. Trattandosi di sequenziamento dell'esoma, gli esoni devono essere isolati dalle regioni non codificanti. A tal fine, i frammenti dell'esoma coniugati agli adattatori vengono fatti ibridare con oligonucleotidi biotinilati e successivamente vengono catturati con delle biglie magnetiche ricoperte di streptavidina. A questo punto, i frammenti dell'esoma ibridati vengono amplificati tramite PCR e purificati, per poi essere sequenziati.

La costruzione delle librerie è stata effettuata mediante il kit Agilent SureSelect All Exon V6; il sequenziamento è stato poi effettuato utilizzando una tecnologia

Illumina. Quest'ultima prevede l'impiego di un supporto solido (*flow cell*), su cui sono ibridati dei *primers* complementari alle sequenze degli adattatori. La libreria di sequenziamento dell'esoma viene caricata sulla *flow cell*, in modo che ogni frammento si vada a legare ai *primers* complementari ai propri adattatori. Avviene quindi la reazione di amplificazione, attraverso cui vengono creati dei *cluster* monoclonali di ciascun frammento di ssDNA (*single strand DNA*), utilizzati per il successivo sequenziamento. Il sequenziamento con tecnologia Illumina si realizza per sintesi. I nucleotidi sono marcati con fluorofori diversi e vengono incorporati dalla polimerasi attraverso la complementarità naturale delle basi. Essi contengono anche dei terminatori reversibili che bloccano l'incorporazione della base successiva. Una volta letto il segnale fluorescente, che indica quale nucleotide è stato aggiunto, il terminatore viene tagliato per permettere l'incorporazione di un altro nucleotide al successivo ciclo di estensione. Attraverso questo processo vengono prodotte un insieme di sequenze di DNA definite *reads*. Infine, l'analisi bioinformatica consente l'allineamento delle *reads* prodotte e l'identificazione delle varianti rispetto al genoma di riferimento (GRCh37/hg19).

Analisi bioinformatica

In seguito al sequenziamento, sono stati prodotti i file *.bam* (*Binary Alignment Map*) e *.vcf* (*Variant Calling Format*), contenenti rispettivamente le sequenze delle *reads* allineate al genoma di riferimento (GRCh37/hg19), in formato binario, e le informazioni relative alle varianti trovate nel DNA del paziente in specifiche posizioni del genoma di riferimento.

I file *.bam* e *.vcf* relativi ai campioni nel progetto sono stati importati sul software VarSeq™ v2.x (Golden Helix, Inc., Bozeman, MT, www.goldenhelix.com) per l'analisi delle CNVs.

In particolare, VarSeq effettua la chiamata delle CNVs basandosi sul *coverage*. Il *coverage* rappresenta il numero medio di *reads* che mappano in una determinata regione; in questo caso si fa riferimento alle regioni esoniche. Per ognuna di queste regioni, l'algoritmo utilizzato dal programma per stabilire la presenza/assenza di CNVs misura la deviazione del *coverage* del campione in esame rispetto ad un set di campioni di riferimento definiti *reference samples*, oltre che rispetto alle altre regioni all'interno del campione stesso. I *reference samples* sono dei campioni

sequenziati in precedenza, appartenenti a diversi lotti di sequenziamento, e più simili possibile dal punto di vista del *coverage* al campione in analisi.

Gli eventi CNVs evidenziati tramite l'algoritmo sono stati filtrati con l'ausilio di tre filtri di base qualitativi e due filtri di ricorrenza:

1. *CNV State* (tipo di CNV): sono stati presi in considerazione gli eventi di duplicazione (*Duplicate*), delezione (*Deletion*) e delezione in eterozigosi (*Het Deletion*);
2. *Flags* (eventuale presenza di problemi statistici associati ad un determinato evento, per esempio un contenuto estremo in GC, una *Ratio* insufficiente, un basso *Z Score* (**Tabella 10** in Appendice)): sono stati presi in considerazione solo gli eventi privi di *flags* (*Missing*);
3. *p-value* (probabilità di rintracciare casualmente quello stesso evento in un pool di campioni generico): ≤ 0.01 ;
4. *#Samples* (numero di campioni nel progetto in cui si riscontra quella stessa identica CNV): ≤ 5 (1% nel progetto di quasi 500 campioni);
5. *#Match's type* (numero di eventi con lo stesso tipo di CNV che si sovrappongono nella regione in un database di 400 campioni popolato precedentemente con campioni esterni al progetto (**Tabella 11** in Appendice)): ≤ 4 (1%).

L'ultimo filtraggio è stato quello per pannello. Sono stati utilizzati due pannelli relativi ai NDDs, rispettivamente contenenti 1534 geni primari e 1231 geni candidati (per cui l'evidenza di associazione ai NDDs è meno diretta), ottenuti dai dati presenti in diversi database sulla base di una revisione aggiornata della letteratura.

Sono stati perciò selezionati nello specifico eventi CNVs in geni d'interesse per i NDDs, o comunque eventi molto grandi (> 500 kb), tipicamente eventi multi-sonici contigui, possibilmente implicati nel fenotipo malattia.

I campioni presenti nel progetto sono quelli risultati negativi nella precedente analisi delle SNVs tramite WES, per cui il lavoro svolto in questa tesi si è basato su campioni di DNA già estratti e sequenziati in precedenza.

Per ciascuno dei 100 campioni sottoposti all'analisi delle CNVs sono stati valutati i singoli eventi identificati dal programma e riportati con le seguenti informazioni (**Figura 5**):

- *Region*: numero del cromosoma e coordinate di inizio-fine dell'evento (sempre in riferimento alle regioni esoniche su cui è stato fatto il calcolo del *coverage*);
- *# Targets*: numero di esoni coinvolti nell'evento;
- *# Samples*: numero di campioni in cui si è riscontrato quello stesso identico evento all'interno del progetto (settato al massimo all'1%);
- *Span*: dimensione dell'evento in termini di paia di basi;
- *Gene Names*: gene/i coinvolto/i nell'evento;
- *Overlapping Exons (Clinically Relevant)*: esone/esoni coinvolti nell'evento (in riferimento al trascritto clinicamente rilevante se presenti più trascritti alternativi);
- CNV vera e propria, a sua volta comprendente il *CNV State* (*Duplicate*, *Deletion* o *Het Deletion*) e le due metriche associate all'evento, *Avg Z Score* e *Avg Ratio*.
 1. *Z Score*: è il numero di deviazioni standard dalla media dei *reference samples*;
 2. *Ratio*: rappresenta il *coverage* del campione diviso la media dei *reference samples*.

Oltre a questi valori medi, sono stati calcolati anche i valori delle metriche relative ad ogni singolo esone coinvolto nell'evento se quest'ultimo comprendeva più esoni.

Le due metriche sono caratterizzate da dei *range* di accettabilità:

- *Ratio*: per una regione *wild type* (wt) è circa 1, se è presente una duplicazione in eterozigosi è circa 1.5, se invece è presente una delezione è circa 0.5 in eterozigosi o 0 in omozigosi;
- *Z Score*: è 0 per una regione wt, in presenza di una duplicazione assume valori ideali da 3 in su, mentre in presenza di una delezione ha valori ideali da -3 in giù.

(Ci si riferisce a valori ideali perché si tratta di valori probabilistici che, come tali, vanno interpretati (**Tabella 10** in Appendice)).

Quando le due metriche sono concordanti, ossia assumono entrambe valori all'interno dei propri *range* di accettabilità, è più verosimile che quell'evento mutazionale a cui fanno riferimento sia effettivamente un evento reale.

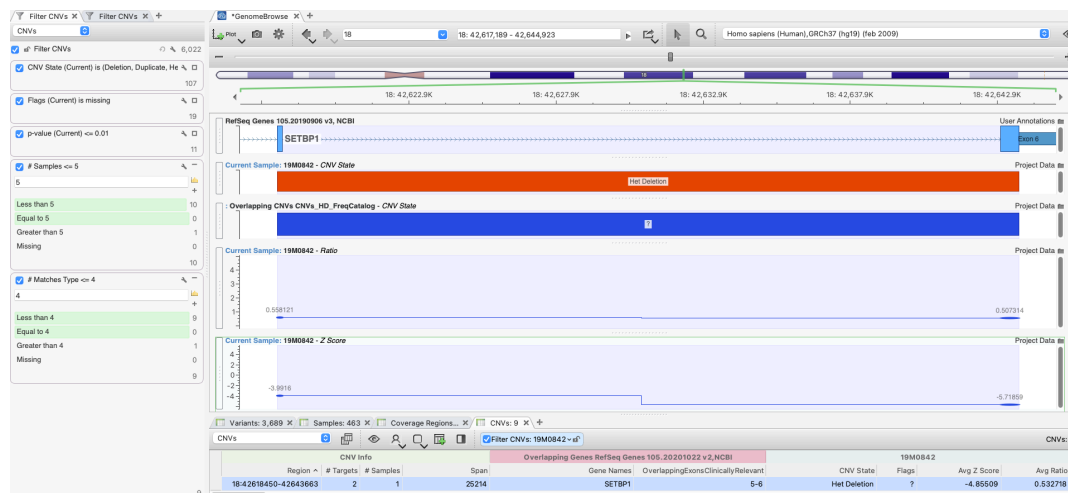


Figura 5. Schermata di VarSeq. A sinistra sono riportati i filtri utilizzati per tutti gli eventi CNVs; a destra è riportato nel dettaglio uno specifico evento CNV (riferito al campione 5), dove sono visibili tutte le informazioni relative all'evento stesso.

Ricerca nei database e in letteratura

L'analisi bioinformatica appena descritta ha portato all'identificazione della possibile presenza di CNVs correlate ai NDDs o comunque duplicazioni o delezioni molto grandi e quindi con un probabile effetto fenotipico nei campioni in esame. Tra tutte le varianti trovate, sono state selezionate quelle caratterizzate da metriche concordanti e sono state prioritizzate attraverso uno studio relativo alla possibile correlazione genotipo-fenotipo.

Sono stati consultati i database OMIM (<https://www.omim.org/>), DECIPHER (<https://www.deciphergenomics.org/>), DVG (<http://dgv.tcag.ca/dgv/app/home>), Orphanet (<https://www.orpha.net/consor/cgi-bin/index.php>), Genecards/Malacards (<https://www.genecards.org/>, <https://www.malacards.org/>), HGMD (<https://www.hgmd.cf.ac.uk/ac/index.php>), e la letteratura scientifica, al fine di confrontare i fenotipi patologici associati alle varianti CNVs identificate bioinformaticamente con i sintomi clinici riportati dal medico per ogni paziente.

Attraverso questa prioritizzazione, per 10 campioni è stata trovata una variante CNV con elevata significatività, quindi un evento CNV caratterizzato da una *Ratio* e uno *Z Score* concordanti e coinvolgente uno o più geni o una regione

cromosomica le cui varianti erano state precedentemente associate ad un fenotipo patologico compatibile col fenotipo clinico descritto per il paziente.

Validazione in laboratorio

Gli eventi CNVs significativi sono stati infine confermati sperimentalmente. In particolare, i 10 eventi sono stati distinti in due gruppi in base alle loro dimensioni:

1. Varianti piccole coinvolgenti due o più esoni di un singolo gene (3/10);
2. Varianti grandi comprendenti più geni (7/10).

Le prime sono state confermate tramite *real-time* PCR, mentre le seconde tramite *array*-CGH.

Real-time PCR

La *real-time* PCR (*Polymerase Chain Reaction*) è una tecnica di biologia molecolare che consente di amplificare uno specifico *target* genomico d'interesse e, allo stesso tempo, di seguire in tempo reale l'andamento dell'amplificazione stessa.

In maniera del tutto analoga alla PCR convenzionale, essa si compone di tre fasi ripetute in maniera ciclica: i) denaturazione della doppia elica; ii) *annealing* (appaiamento) dei *primers* ai filamenti di DNA denaturati; iii) ed estensione degli inneschi da parte della polimerasi.

Per seguire l'amplificazione in tempo reale è stato utilizzato il colorante fluorescente SYBR Green, un agente intercalante (ossia in grado di legare il DNA a doppio filamento insinuandosi nella struttura della doppia elica) che emette fluorescenza in modo proporzionale alla quantità di dsDNA (*double strand* DNA) presente. Il SYBR Green, tuttavia, è un intercalante aspecifico: questo implica che esso potrebbe legare anche eventuali prodotti aspecifici o impurità (ad esempio dimeri di *primers*) presenti in reazione, generando così dei segnali fluorescenti aggiuntivi che andrebbero ad inquinare quello relativo al *target* d'interesse. Per questo motivo, al termine dell'amplificazione si esegue la cosiddetta determinazione della curva di *melting*: la temperatura viene incrementata in maniera graduale e continua fino alla denaturazione del *target* amplificato e

conseguente perdita del segnale fluorescente. In base al profilo di riduzione del segnale fluorescente è possibile accertare la specificità dell'amplificazione ed escludere la presenza di prodotti inattesi/indesiderati.

Misurando la fluorescenza emessa in tempo reale è possibile tracciare una curva di amplificazione, nella quale si possono distinguere tre fasi: i) una prima fase esponenziale in cui i reagenti sono in eccesso e l'efficienza di amplificazione è del 100%, per cui il DNA *target* raddoppia ad ogni ciclo di amplificazione; ii) una seconda fase in cui il processo inizia a rallentare e passa attraverso una fase di amplificazione lineare; iii) e una fase di plateau che viene raggiunta quando i *primers* e/o i dNTPs si esauriscono.

La *real-time* PCR offre la possibilità di effettuare una quantificazione assoluta o relativa di campioni di DNA a concentrazione incognita. Nel presente progetto, la determinazione del numero di copie di una specifica regione genomica in un campione d'interesse è stata effettuata attraverso una metodica di quantificazione relativa nota come metodo $\Delta\Delta C_t$. Questo metodo consiste nel confronto tra il campione in esame e un campione di riferimento (*reference*) validato per l'assenza di CNVs nella regione indagata. Il parametro utile ad effettuare il confronto è il ciclo *threshold* (C_t), vale a dire il ciclo di amplificazione in cui la fluorescenza emessa dal prodotto supera il valore *threshold* di fluorescenza, che è generalmente collocato a metà della fase esponenziale. Il C_t è legato alla quantità di DNA iniziale secondo una correlazione inversa: più DNA è presente in partenza, meno cicli di amplificazione sono necessari per raggiungere il C_t . Sul campione in esame e il *reference* viene dunque amplificata sia la regione di cui si vuole valutare il numero di copie, sia una regione di un gene *housekeeping* (codificante per proteine fondamentali per la vita della cellula, il cui numero di copie difficilmente subisce alterazioni ed è pertanto verosimilmente amplificato alla stessa maniera in qualunque campione considerato). Il primo passaggio è il calcolo del ΔC_t : il C_t relativo all'amplificazione dell'*housekeeping* viene sottratto al C_t relativo all'amplificazione della regione indagata per la CNV. Questa normalizzazione ha lo scopo di escludere eventuali errori di caricamento o di quantificazione preliminare del DNA, ed è effettuata sia per il campione in esame che per il *reference*, ottenendo rispettivamente il ΔC_t campione ed il ΔC_t *reference*. Da questi si ottiene il $\Delta\Delta C_t$, calcolato come ΔC_t *campione* - ΔC_t *reference*. Infine, il rapporto relativo tra il numero di copie della regione indagata nel campione e quello nel *reference* risulta dall'operazione $2^{-\Delta\Delta C_t}$.

Gli esperimenti di *real-time* PCR sono stati condotti su piastra da 96 pozzetti Applied Biosystems™ MicroAmp® Fast 96-Well Reaction Plate (0.1mL), e ogni campione è stato analizzato in triplicato.

Il *reference* e il controllo negativo sono stati scelti tra quelli sottoposti di recente nel laboratorio all'analisi mediante *array*-CGH. In questo modo è stato possibile verificare che la regione d'interesse da amplificare non fosse né duplicata né deleta in entrambi i campioni. Inoltre, come controllo interno è stato utilizzato il gene *housekeeping* (HK) FBN1, un membro della famiglia proteica delle fibrilline. In particolare, per l'HK sono state impiegate due diverse coppie di *primers*, ottenendo due prodotti di dimensioni differenti.

Durante la fase preliminare sono state disegnate le coppie di *primers* per amplificare in modo specifico le regioni d'interesse. In particolare, sono stati disegnati i *primers* sia per gli esoni agli estremi degli eventi mutazionali identificati da VarSeq che per gli esoni fiancheggiati, in modo tale da valutare gli estremi reali degli eventi stessi.

Le coppie di *primers* sono state disegnate con Primer3Plus (<https://www.primer3plus.com/>), secondo i seguenti criteri:

- Localizzazione, cioè appaiamento rispetto alla sequenza da amplificare;
- Presenza di al massimo 3 basi identiche ripetute consecutivamente;
- Lunghezza ideale di 20 nucleotidi (nt);
- Temperatura di *melting* (T_m) intorno ai 60°C e non più di un grado di differenza tra le T_m dei due *primers* F (*Forward*) e R (*Reverse*);
- Composizione in GC intorno al 50% e non troppo differente tra i due *primers* F e R;
- Penalità complessiva per la coppia di *primers* (derivante da varie valutazioni e vari parametri) inferiore a 1;
- Presenza di eventuali strutture secondarie (dovute al *primer* che si ripiega su sé stesso o alla formazione di dimeri di *primers* F-F, R-R o F-R). In questo caso è importante che la T_m delle strutture secondarie, se presenti, sia molto più bassa di 60°C, temperatura a cui avviene l'*annealing*;
- T_m del prodotto non superiore a 95°C, temperatura a cui avviene la denaturazione;
- Lunghezza del prodotto di amplificazione tra 150-175 bp o tra 70-95 bp a seconda della lunghezza dell'HK impiegato, quello da 160 bp o da 80 bp rispettivamente.

Infine, è stato usato il *tool In-Silico* PCR presente su Human Genome Browser (<https://genome.ucsc.edu/>) per valutare la specificità delle coppie disegnate.

Le coppie di *primers* disegnate e utilizzate per i tre esperimenti condotti tramite la *real-time* PCR sono riportate nell'Appendice in **Tabella 12**.

Sono state poi preparate delle aliquote di ciascun *primer* secondo i seguenti volumi:

- 225 μL di acqua distillata (dH_2O);
- 25 μL di *primers* ricostituiti.

Una volta aliquotati i *primers*, i campioni di DNA del probando, dei genitori dove disponibili, del *reference* e del controllo negativo sono stati diluiti con dH_2O per ottenere un DNA di partenza concentrato 10 $\text{ng}/\mu\text{L}$.

Una volta preparate le diluizioni di DNA, sono state preparate le soluzioni per l'amplificazione di ciascun *target* (ogni esone da amplificare e l'HK), secondo i seguenti volumi riferiti ad un singolo pozzetto:

- 4.1 μL di dH_2O ;
- 0.2 μL di ciascun *primer*;
- 5 μL di mix preconstituita (2x AceQ[®] Universal SYBR Green qPCR Master Mix).

Infine, la piastra è stata caricata pipettando in ogni pozzetto 9,5 μL di soluzione per l'amplificazione e 0.5 μL di DNA diluito. Per ciascun *target* sono stati caricati anche 9.5 μL di bianco, costituito solo da dH_2O , *primers* e mix preconstituita (senza DNA), per valutare la presenza di un'eventuale contaminazione introdotta dall'esterno o insita nei reagenti stessi.

La reazione di amplificazione è avvenuta su strumento Applied Biosystems[™] StepOnePlus[™] Real-Time PCR System, con l'ausilio del software StepOne[™] Real-Time PCR System (v2.3). Il programma termico utilizzato è riportato nell'Appendice in **Tabella 13**.

Una volta conclusa la reazione, sono state analizzate le curve di amplificazione, sia dei bianchi che dei campioni in triplicato, e le curve di *melting* per valutare la specificità dell'amplificazione stessa.

I valori dei C_t sono stati esportati in Excel, dove è avvenuta l'analisi dei risultati mediante il metodo del $\Delta\Delta C_t$. Infine, i valori così ottenuti sono stati confrontati con i valori di riferimento di seguito riportati, al fine di valutare la reale presenza di un evento CNV:

- 1 se la regione è wt;

- 0.5 se c'è una delezione in eterozigosi (idealmente indeterminato se c'è una delezione in omozigosi, ma in realtà C_t molto alti);
- 1.5 se c'è una duplicazione in eterozigosi, poi si osserva un aumento di 0.5 in 0.5 per duplicazioni di più copie.

Array-CGH

L'*array*-CGH (*Comparative Genomic Hybridization*) è un metodo diffuso per verificare la presenza di varianti strutturali del DNA, come per l'appunto le CNVs. Si basa sul confronto tra il gDNA del paziente e un gDNA di riferimento. I due campioni sono marcati in modo differenziale con due diversi fluorofori e vengono fatti co-ibridare in ugual quantità su un *array*. In ogni spot dell'*array* è presente un *cluster* di sonde, rappresentate da molecole di DNA a singolo filamento con sequenza nota, identiche fra loro e corrispondenti ad una specifica regione. A seconda della quantità di molecole di DNA del paziente e del riferimento che competono per l'ibridazione con le sonde depositate su ciascuno spot dell'*array*, si ottiene una determinata fluorescenza:

- Se nel campione del paziente la regione in esame è wt, le due fluorescenze si equivalgono;
- Se nel probando è presente una duplicazione, più copie del suo DNA rispetto a quello di riferimento si vanno ad ibridare con le sonde di quella particolare regione, per cui prevale la fluorescenza del campione del paziente;
- Se nel probando è presente una delezione, meno copie del suo DNA rispetto a quello di riferimento si vanno ad ibridare con le sonde di quella specifica regione, per cui prevale la fluorescenza del *reference*.

Quindi, rapportando le due intensità di fluorescenza del probando e del *reference* si riesce a stabilire se nel primo è presente una variazione del numero di copie di una determinata regione rispetto al secondo.

Gli esperimenti di *array*-CGH sono stati condotti su *array* Agilent GenetiSure Cyto CGH Microarray Kit, 4 X 180 K, costituito da un vetrino con 4 *arrays* da 180.000 sonde. Essendo ogni vetrino dotato di 4 *arrays*, in ognuno di essi sono stati analizzati 8 campioni in contemporanea, 4 probandi e 4 *references* dello stesso sesso del probando corrispondente. I kit e i reagenti utilizzati sono elencati nell'Appendice in **Tabella 14**. Il protocollo prevede diverse fasi.

Fase preliminare: Valutazione quantitativa e qualitativa del gDNA del paziente.

La valutazione quantitativa è stata effettuata fluorometricamente con Invitrogen™ Qubit™ 4 Fluorometer e impiegando il Qubit™ dsDNA BR Assay Kit. Sono stati miscelati 190 µL di Working Solution con 10 µL di ciascuno dei due standard, rispettivamente concentrati 0 ng/µL e 100 ng/µL; e 199 µL di Working Solution con 1 µL di campione da quantificare. Le soluzioni sono state vortexate per alcuni secondi e tenute al buio per 3 minuti, per dare la possibilità al colorante fluorescente contenuto nella Working Solution di legarsi al DNA. È stata poi misurata la fluorescenza emessa dai due standard per creare la retta di taratura, con sulle ascisse la concentrazione e sulle ordinate l'intensità di fluorescenza. Infine, il valore d'intensità di fluorescenza misurato per il campione di gDNA del paziente è stato interpolato sulla retta di taratura, determinandone la concentrazione.

La valutazione qualitativa del gDNA del probando viene eseguita per valutarne la purezza (assenza di contaminanti quali proteine, carboidrati e tracce di solventi organici) e l'integrità (assenza di frammentazione).

La purezza è stata valutata tramite i rapporti tra le assorbanze calcolati al NanoDrop durante la prima quantificazione degli estratti: A260/280 per valutare l'eventuale contaminazione da proteine e A260/230 per valutare l'eventuale contaminazione da composti organici.

L'integrità è stata invece valutata mediante elettroforesi capillare su strumento Agilent 4159 TapeStation System, utilizzando i Genomic DNA Reagents e la Genomic DNA ScreenTape. Sono stati miscelati 10 µL di *buffer* di corsa con 1 µL di *Ladder* (marcatore di peso molecolare) o 1 µL di campione da analizzare. Le soluzioni sono state centrifugate a 2000 rpm per un minuto e sono state poi caricate sullo strumento insieme alla ScreenTape. Per l'analisi è stato utilizzato il software TapeStation Controller; mentre per la visualizzazione dei risultati è stato usato il software TapeStation Analysis. Per ogni campione è stato valutato il DIN (DNA Integrity Number) associato, ossia un indice da 1 a 10 che definisce l'integrità del DNA: più alto è questo valore, minore è il grado di frammentazione del DNA. Sono stati considerati accettabili DIN di almeno 6.

Una volta terminata la fase di valutazione preliminare, 500 ng di ciascun campione sono stati diluiti come segue:

$500/C_s + V_{H_2O} = 26 \mu L$, dove C_s è la concentrazione del campione calcolata nella fase precedente tramite il Qubit.

Fase 1: Marcatura fluorescente del gDNA con cianine.

Si compone di due sottofasi:

1. Frammentazione termica del gDNA, denaturazione dei frammenti e appaiamento dei Random Primers. Sono stati aggiunti 5 μL di Random Primers a ciascun campione e a ciascun *reference*. Le soluzioni sono state poste in termociclatore col programma termico “Marcatura, sottofase 1” descritto nell’Appendice in **Tabella 15**;
2. Marcatura differenziale del DNA dei probandi e dei *references*. Sono state preparate le mix di marcatura per ciascuna delle due cianine, con i seguenti volumi (riferiti al singolo vetrino):
 - 42 μL di 5X gDNA Buffer;
 - 21 μL di dNTPs;
 - 12.6 μL di dUTPs marcati con cianina 3 o 5;
 - 4.2 μL di Exo Klenow.

Sono stati aggiunti 19 μL di mix con cianina 5 nelle provette dei probandi e 19 μL di mix con cianina 3 nelle provette dei *references*. Le soluzioni sono state poste in termociclatore col programma termico “Marcatura, sottofase 2” descritto nell’Appendice in **Tabella 15**.

Fase 2: Purificazione del gDNA marcato.

I campioni di DNA marcati con le cianine sono stati purificati attraverso due lavaggi, con l’ausilio degli appositi filtri e colonnine di purificazione. È stato effettuato un primo lavaggio con 430 μL di TE 1X (pH 8.0), seguito da centrifugazione a 14000 g per 10 minuti; e un secondo lavaggio con 480 μL di TE 1X (pH 8.0), sempre seguito da centrifugazione a 14000 g per 10 minuti. I filtri sono stati poi capovolti in nuove provette e centrifugati a 1000 g per un minuto, in modo da raccogliere al loro interno l’eluato, ossia il gDNA marcato e purificato.

Il volume richiesto di eluato è di 21 μL , per cui le provette sono state poste col tappo aperto in concentratore a 45°C per 20 minuti. Al termine, il pellet è stato risospeso in 21 μL di TE 1X (pH 8.0).

Successivamente, con il NanoDrop è stata calcolata la concentrazione dei campioni purificati, utilizzando 1.5 μL di TE 1X (pH 8.0) come bianco, e la concentrazione molare delle cianine in pmol/ μL .

Con questi valori, per ciascun campione sono state calcolate:

- La resa:

$$\frac{C_S (\text{ng}/\mu\text{L}) \cdot V_S (\mu\text{L})}{1000}$$

Dove C_S è la concentrazione del campione e V_S è il volume del campione (21 μL);

- L'attività specifica delle cianine:

$$\frac{C_C (\text{pmol}/\mu\text{L})}{C_S (\text{ng}/\mu\text{L})} \cdot 1000$$

Dove C_C è la concentrazione della specifica cianina e C_S è la concentrazione del campione.

Con questi dati, i probandi e i *references* sono stati ordinati secondo valori decrescenti di resa e attività specifica. Ciascun probando è stato poi accoppiato col *reference* occupante la stessa posizione nell'ordine precedentemente determinato, dando la priorità all'attività specifica nel caso di discordanza tra i due parametri. Ciascun probando è stato quindi miscelato col corrispondente *reference*, formando le coppie così determinate.

Fase 3: Ibridazione del DNA marcato e purificato con le sonde dell'array.

Le miscele probando-*reference* sono state trattate con lo scopo di potenziare l'intensità del segnale, ridurre al massimo il rumore di fondo e ottimizzare la cinetica della successiva ibridazione su vetrino. A tal fine è stata preparata una mix di ibridazione con i seguenti volumi (riferiti al singolo vetrino):

- 21 μL di Cot-I DNA (1 $\text{mg}/\mu\text{L}$) (per prevenire l'ibridazione delle sequenze ripetute centromeriche e telomeriche alle sonde dell'array);
- 46.2 μL di Blocking Agent (per ridurre le ibridazioni aspecifiche), precedentemente ricostituito con 1.35 μL di dH_2O ;
- 231 μL di HI-RPM Hybridization Buffer (per prevenire le ibridazioni aspecifiche, minimizzare il segnale fluorescente di fondo sulla superficie dell'array, e concentrare il DNA e migliorare l'efficienza d'ibridazione complessiva).

Ad ogni miscela probando-*reference* sono stati aggiunti 71 μ L di mix di ibridazione. Le soluzioni sono state poste in termociclatore col programma termico "Ibridazione" descritto nell'Appendice in **Tabella 15**.

Al termine del programma, la camera d'ibridazione è stata assemblata aggiungendo 100 μ L di ciascuna coppia probando-*reference* sull'apposita *gasket slide* e appoggiando sopra il vetrino con gli *arrays*, mettendo le sonde a contatto con le miscele. La camera d'ibridazione è stata chiusa con l'apposito dispositivo metallico di chiusura ed è stata incubata a 67°C in rotazione bilanciata per 24 ore.

Fase 4: Lavaggi.

Al termine dell'incubazione, la camera d'ibridazione è stata disassemblata e il vetrino con gli *arrays* è stato trasferito in una vaschetta riempita con l'Agilent Oligo aCGH/ChIP-on-Chip Wash Buffer 1, in agitazione magnetica a 350 rpm per 5 minuti. Lo stesso vetrino è stato poi trasferito in una vaschetta riempita con l'Agilent Oligo aCGH/ChIP-on-Chip Wash Buffer 2 preriscaldato a 37°C, su piastra riscaldante e in agitazione magnetica per 1 minuto.

Infine, il vetrino è stato inserito nell'apposito supporto per la lettura da parte dello strumento Agilent SureScan Dx Microarray Scanner System.

Fase 5: Lettura del vetrino e analisi dei risultati.

Sono stati utilizzati i software: Agilent Microarray Scan Control per l'acquisizione delle immagini, Agilent Feature Extraction per la visualizzazione delle immagini acquisite e Agilent CytoGenomics per l'analisi dei risultati.

I valori di riferimento per la valutazione dei risultati sono riportati in **Tabella 1**.

Tabella 1. Valori di riferimento per l'analisi dei risultati dell'array-CGH.

| | <i>Average Sample CN</i> | <i>Reference CN</i> | <i>Ratio (S/R)</i> | <i>Ideal log₂ratio</i> |
|------------------------------------|--------------------------|---------------------|--------------------|-----------------------------------|
| <i>Homozygous Deletion</i> | 0 | 2 | 0 | $-\infty$ |
| <i>Heterozygous Deletion</i> | 1 | 2 | 0.5 | -1 |
| <i>Diploid</i> | 2 | 2 | 1 | 0 |
| <i>Duplication (1 copy gain)</i> | 3 | 2 | 1.5 | 0.58 |
| <i>Duplication (2 copies gain)</i> | 4 | 2 | 2 | 1 |

Risultati

L'analisi bioinformatica eseguita tramite il software VarSeq ha portato all'identificazione di molte varianti CNVs in geni correlati ai NDDs e di grandi eventi CNVs nei 100 pazienti analizzati. Tra questi eventi CNVs, quelli con metriche associate concordanti sono stati prioritizzati ricercando nei database dedicati e in letteratura i fenotipi patologici correlati al gene o ai geni coinvolti nella CNV. Successivamente è stata ricercata la presenza di tali varianti o di varianti simili in letteratura e nei database dedicati. Infine, per le varianti CNVs più interessanti si è fatto un confronto tra il fenotipo riportato in letteratura per quel gene/quei geni e il fenotipo clinico riportato dal medico per ogni paziente. Questo lavoro di prioritizzazione delle varianti CNVs individuate da VarSeq ha permesso di trovare in 10 dei pazienti analizzati (qui numerati per semplicità da 1 a 10) una CNV con elevata significatività.

In **Tabella 2** sono riportate le informazioni cliniche riferite dal medico per ciascuno dei 10 pazienti per cui è stata trovata una CNV significativa.

Tabella 2. Sesso, età e fenotipo clinico dei 10 pazienti per cui è stata identificata una CNV significativa.

| Paziente | Sesso | Età ¹⁰ | Fenotipo clinico |
|----------|-------|-------------------|--|
| 1 | F | 10 anni | Sospetta sindrome di Pitt-Hopkins o simile |
| 2 | M | 37 anni | Disabilità intellettiva sindromica |
| 3 | M | 12 anni | Disabilità intellettiva, disturbi specifici dell'apprendimento e ritardo delle acquisizioni |
| 4 | M | 38 anni | Encefalopatia congenita epilettogena con grave disordine dello sviluppo cognitivo e disturbo comportamentale |
| 5 | F | 20 anni | Deficit cognitivo cui si associano lievi note non specifiche |
| 6 | M | 30 anni | Ipoplasia cerebellare e deficit cognitivo severo |
| 7 | M | 29 anni | Sospetta sindrome di Bardet-Biedl, disabilità intellettiva e dismorfismi del viso |
| 8 | M | 7 anni | Ritardo globale dello sviluppo |
| 9 | M | 8 anni | Encefalopatia epilettica sindromica |
| 10 | F | 36 anni | Disabilità intellettiva, ipoacusia neurosensoriale e dismorfismi facciali |

¹⁰ Al momento del prelievo del campione biologico.

In **Tabella 3**, **Tabella 4** e **Tabella 5** sono riportate, per i 10 pazienti, le CNVs identificate da VarSeq, con le corrispondenti metriche associate e la relativa correlazione genotipo-fenotipo, rispettivamente in relazione alle duplicazioni, alle delezioni coinvolgenti parte di un singolo gene o al massimo due geni e alle delezioni coinvolgenti più geni di un determinato segmento genomico.

Tabella 3. Duplicazioni identificate da VarSeq, con metriche associate e relativa correlazione genotipo-fenotipo.

| Paziente | Segmento genomico | Dimensione | Avg Ratio | Avg Z Score | Correlazione genotipo-fenotipo |
|----------|--|------------|-----------|-------------|---|
| 1 | 22q11.21q11.22 (chr22:19,211,414-22,330,081dup) | 3.1 Mb | 1.31 | 3.35 | Sindrome da duplicazione 22q11.2, autosomica dominante (OMIM: 608363) |
| 2 | Xp22.33p22.11 (chrX:2,852,873-21,900,773dup) | 19 Mb | 1.52 | 3.79 | Duplicazione Xp22.33p22.12 in un paziente con disabilità intellettiva e dismorfismi facciali. [9] |
| 8 | 22q11.21 (chr22:18,893,886-21,414,817dup) | 2.5 Mb | 1.47 | 3.93 | Sindrome da duplicazione 22q11.2, autosomica dominante (OMIM: 608363) |

Tabella 4. Delezioni coinvolgenti parte di un singolo gene o al massimo due geni identificate da VarSeq, con metriche associate e relativa correlazione genotipo-fenotipo.

| Paziente | Gene | Esoni | Dimensione | Avg Ratio | Avg Z Score | Correlazione genotipo-fenotipo |
|----------|------------|------------|------------|-----------|-------------|--|
| 3 | SOX6 | 2-4 | 106.7 kb | 0.56 | - 3.01 | Varianti <i>de novo</i> nel gene SOX6 causano una sindrome del neurosviluppo associata con ADHD, craniosinostosi e osteocondroma. [10] |
| 5 | SETBP1 | 5-6 | 25 kb | 0.53 | -4.86 | Disordine associato all'aploinsufficienza di SETBP1. [11] |
| 9 | CHD2, RGMA | 31-29, 1-4 | 75 kb | 0.54 | -3.39 | Microdelezione 15q26.1 coinvolgente solo i geni CHD2 e RGMA in due adulti con disabilità intellettiva moderata, epilessia e obesità del tronco. [12] |

Tabella 5. Delezioni coinvolgenti più geni di un determinato segmento genomico identificate da VarSeq, con metriche associate e relativa correlazione genotipo-fenotipo.

| Paziente | Segmento genomico | Dimensione | Avg Ratio | Avg Z Score | Correlazione genotipo-fenotipo |
|----------|--|------------|-----------|-------------|---|
| 4 | 22q13.33 (chr22:50,297,486-51,169,740del) | 872.26 kb | 0.55 | -3.97 | Sindrome di Phelan-McDermid, autosomica dominante (OMIM: 606232) |
| 6 | 22q13.33 (chr22:50,297,486-51,169,740del) | 872.26 kb | 0.53 | -3.36 | Sindrome di Phelan-McDermid, autosomica dominante (OMIM: 606232) |
| 7 | 3p14.3p14.1 (chr3:56,696,413-63,898,903del) | 7.2 Mb | 0.53 | -3.66 | Una nuova microdelezione 3p14.2 in un paziente con disabilità intellettiva e deficit del linguaggio: <i>report</i> del caso e revisione della letteratura. [13] |
| 10 | 8p23.1 (chr8:8,998,304-10,782,340del) | 1.78 Mb | 0.51 | -4.28 | Correlazione genotipo-fenotipo in un nuovo caso di delezione 8p23.1 e revisione della letteratura. [14] |

Le dieci varianti sopra descritte sono state infine confermate sperimentalmente tramite *real-time* PCR o *array*-CGH, in dipendenza della loro dimensione. I risultati ottenuti sono presentati in **Tabella 6**, **Tabella 7** e **Tabella 8**, rispettivamente in relazione alle duplicazioni, alle delezioni coinvolgenti due o più esoni di un singolo gene validate tramite *real-time* PCR e alle delezioni coinvolgenti più geni di un determinato segmento genomico validate tramite *array*-CGH. Per i probandi per i quali erano disponibili i campioni di DNA dei genitori sono riportati anche i risultati dell'analisi di segregazione.

Tabella 6. Duplicazioni validate tramite array-CGH.

| Paziente | Segmento genomico | Dimensione | MeanLogRatio ¹¹ | Segregazione |
|----------|---|------------|----------------------------|--------------------------|
| 1 | 22q11.21q11.22 (19219899_23228483)x3 | 4 Mb | 0.50 | Genitori non disponibili |
| 2 | Xp22.33p22.11 (2701273_21962375)x4 | 19.3 Mb | 0.88 | Genitori non disponibili |
| 8 | 22q11.21 (18880598_21505417)x3 | 2.6 Mb | 0.46 | Ereditarietà paterna |

Tabella 7. Delezioni coinvolgenti due o più esoni di un singolo gene validate tramite real-time PCR.

| Paziente | Tipo di CNV | Gene | Esoni | MeanRatio | Segregazione |
|-----------------|--------------------------|--------|-------|-----------|--------------------------|
| 3 | Delezione in eterozigosi | SOX6 | 1-4 | 0.52 | Genitori non disponibili |
| 5 | Delezione in eterozigosi | SETBP1 | 5-6 | 0.33 | <i>De novo</i> |
| 9 ¹² | Delezione in eterozigosi | CHD2 | 31-39 | 0.47 | <i>De novo</i> |

Tabella 8. Delezioni coinvolgenti più geni di un determinato segmento genomico validate tramite array-CGH.

| Paziente | Segmento genomico | Dimensione | MeanLogRatio ¹⁰ | Segregazione |
|----------|--|------------|----------------------------|--------------------------|
| 4 | 22q13.32.q13.33 (49388594_51224252)x1 | 1.8 Mb | -0.75 | Genitori non disponibili |
| 6 | 22q13.33 (49531017_51178257)x1 | 1.65 Mb | -0.77 | Genitori non disponibili |
| 7 | 3p14.3p14.1 (56701139_63901725)x1 | 7.2 Mb | -0.79 | <i>De novo</i> |
| 10 | 8p23.1 (8930143_10777658)x1 | 1.85 Mb | -0.85 | <i>De novo</i> |

¹¹ I valori di riferimento sono riportati in **Tabella 1**.

¹² È stato validato solo il gene CHD2 in quanto è quello presente all'interno del pannello di geni associati al neurosviluppo.

Discussione

I NDDs sono una classe di patologie con esordio durante il periodo dello sviluppo che colpiscono la crescita e la funzionalità neurologica, comportando deficit che si traducono in difficoltà di funzionamento a livello cognitivo, comunicativo, comportamentale e psicomotorio [2][3].

Questi disturbi presentano un'ampia eterogeneità sia genetica che clinica che rende difficile l'interpretazione della correlazione genotipo-fenotipo. In questo senso, è importante considerare che la complessità del fenotipo malattia aumenta quando le mutazioni deleterie colpiscono geni sensibili al dosaggio (concetto di vulnerabilità genetica) e quando si sommano più eventi mutazionali deleteri (concetto di carico mutazionale) [3].

In questo contesto, la sola ricerca delle SNVs non rende ragione di tutti i casi osservati per questi disordini, includendo le CNVs come potenziale causa del fenotipo patologico. Le CNVs, essendo delle sequenze ripetute che variano nel numero di copie in individui diversi, sono responsabili di parte della variabilità genetica naturale che si riscontra tra gli individui, ma possono anche essere causa di malattia. Andando ad alterare il numero di copie di geni sensibili al dosaggio esse possono, per l'appunto, causare l'esordio di NDDs. In particolare, per interpretare la loro rilevanza clinica, si devono tenere in considerazione diversi aspetti: dimensione, tipo (delezione o duplicazione), contenuto genico, ereditarietà e frequenza [15][17].

Facendo riferimento al presente progetto svolto nel laboratorio R&I Genetics di Padova, l'analisi delle CNVs in 100 esomi di pazienti affetti da NDDs, e rimasti senza diagnosi in seguito alla precedente ricerca delle SNVs, ha portato all'identificazione e alla conferma della presenza di una CNV significativa in 10 dei probandi analizzati.

In particolare, di questi 10 casi positivi, tre portano una CNV che coinvolge due o più esoni di un singolo gene o al massimo due geni, mentre sette coinvolgono più geni contigui all'interno di un determinato segmento genomico. Sia le CNVs più grandi che quelle più piccole, che colpiscono un solo esone, sono associate a malattie umane [5]. In generale, gli eventi più grandi (> 500 kb) hanno più probabilità di risultare patogenetici rispetto agli eventi più piccoli. Infatti, i primi

generalmente comprendono più geni e quindi hanno una maggiore possibilità di colpire un gene sensibile al dosaggio e/o avere un effetto cumulativo, mentre i secondi devono colpire un gene critico o alterarne la funzionalità per avere un effetto significativo [15][16][17]. Questo è coerente col fatto che, anche in questo lavoro di tesi, più della metà degli eventi identificati come possibile causa genetica e confermati hanno una dimensione superiore a 500 kb. Inoltre, il fatto che i tre eventi più piccoli identificati siano tutti delle delezioni è coerente col fatto che le delezioni cominciano a manifestare un arricchimento per *range* di dimensioni minori rispetto alle duplicazioni [16]. Questo è in linea con un modello secondo cui le delezioni sono soggette ad una pressione selettiva maggiore, in quanto gli effetti fenotipici associati all'alpoinufficienza sono più gravi rispetto a quelli più lievi legati ad un aumento del dosaggio genico [15]. Di conseguenza, le delezioni hanno più probabilità di causare un fenotipo negativo rispetto alle duplicazioni [17] e tendono ad essere più penetranti [16]. Proprio per questo motivo può risultare più difficile determinare il significato funzionale di una duplicazione. Considerando poi che l'arricchimento più significativo a un unico *locus* è associato prevalentemente a *hot spots* di ricombinazione mediata da LCRs [16] e che il meccanismo di NAHR dà origine a delezioni in numero maggiore rispetto alle duplicazioni, è plausibile che solo tre degli eventi CNVs confermati siano duplicazioni.

Tutti i campioni per i quali è stato possibile effettuare un'analisi di segregazione, ad eccezione di uno, hanno mostrato un'insorgenza *de novo* per la variante identificata. Anche questo è un risultato in linea con i dati riportati in letteratura. Sebbene ci siano delle eccezioni, è più probabile che le CNVs *de novo* siano patogenetiche rispetto a quelle ereditate, e questo è vero in particolare per i disturbi più gravi [17]. Esiste anche una correlazione positiva tra la dimensione di una CNV e la sua probabilità di insorgere *de novo* [16]. Inoltre, nonostante esista una correlazione negativa tra il numero di eventi e la dimensione della CNV, per cui sono più le varianti piccole e comuni rispetto a quelle grandi e deleterie [5], sono emerse forti prove a sostegno di ruoli più importanti per eventi rari e *de novo* rispetto a CNVs più comuni [17]. Infatti, anche la frequenza è un parametro importante da considerare quando si valuta la potenziale patogenicità di una variante.

Le CNVs ereditate possono variare nel loro esito sia per quanto riguarda le caratteristiche fenotipiche che la loro gravità, e possono essere chiaramente patogenetiche anche se presenti in un genitore fenotipicamente normale [17]. In

questo contesto si fa riferimento a diverse sindromi ricorrenti, che risultano essere associate ad un'ampia gamma e gravità di fenotipi [17]. Diversi fattori sono stati proposti per spiegare quest'espressività variabile. Quest'ultima potrebbe essere dovuta al background genetico del singolo individuo, ad effetti epigenetici, a fenomeni come l'imprinting, o anche a fattori non genetici, quali i fattori ambientali [17].

Infine, anche il contenuto genico è un aspetto da tenere in considerazione: è più probabile che eventi CNVs che comprendono molti geni e/o geni già noti essere associati a malattie siano patogenetici rispetto a eventi che contengono pochi geni o geni con funzione incerta o sconosciuta [17].

Facendo riferimento al sesso biologico, solo 3/10 probandi che hanno ricevuto una diagnosi sono femmine. In generale, i maschi hanno da 2 a 4 volte un rischio maggiore rispetto alle femmine di sviluppare NDDs [18]. Nella maggior parte dei casi, ciò è dovuto a interazioni tra fattori genetici, ormonali e ambientali (come lo stress e l'aspetto socioculturale) [18].

Guardiamo ora nel dettaglio ognuno di questi 10 casi positivi.

Pazienti 1 e 8

Il paziente 1 è una femmina di 10 anni che presenta sintomi compatibili con la sindrome di Pitt-Hopkins o simile. La sindrome di Pitt-Hopkins (OMIM: 610954) è una malattia genetica autosomica dominante molto rara, che interessa in maniera equivalente maschi e femmine, e a cui si associano ritardo psicomotorio grave e precoce, anomalie respiratorie, dismorfismi facciali caratteristici (macrostomia, con denti distanziati, palato largo e poco profondo, labbra spesse, occhi infossati, naso con narici svasate e orecchie con elice spesso), e frequentemente disturbi gastrointestinali, microcefalia e crisi epilettiche. Il paziente in esame è clinicamente caratterizzato da: ritardo mentale medio/lieve, reflusso gastroesofageo, cefalea, schisi sacrale (S1), doppio distretto ureterale, dismorfismi (scafocefalia, microcefalia, macrostomia, ipertrofia gengivale, filtro lungo, *synophrys*, orecchie basse). L'analisi bioinformatica per la ricerca delle CNVs eseguita sull'esoma del paziente ha portato all'identificazione di una duplicazione in eterozigosi, di circa 3.1 Mb, localizzata a livello del locus genico 22q11.21q11.22 (chr22:19,211,414-22,330,081dup). Questa duplicazione comprende tra gli altri alcuni geni associati a patologie e presenti in OMIM: CDC45 (OMIM: 603465),

GP1BB (OMIM: 138720), TBX1 (OMIM: 602054), TXNRD2 (OMIM: 606448), COMT (OMIM: 116790), TANGO2 (OMIM: 616830), RTN4R (OMIM: 605566), SCARF2 (OMIM: 613619), PI4KA (OMIM: 600286), SERPIND1 (OMIM: 142360), SNAP29 (OMIM: 604202), LZTR1 (OMIM: 600574), GGT2 (OMIM: 137181), MAPK1 (OMIM: 176948). La successiva analisi mediante *array*-CGH ha confermato la presenza di una duplicazione in eterozigosi, di circa 4 Mb, a livello della banda citogenetica 22q11.21q11.22 (chr22:19,219,899-23,228,483dup). Quest'ultima comprende al suo interno tutti i geni malattia OMIM già identificati da VarSeq. Tuttavia, la dimensione reale dell'evento si estende circa 900 kb in più a livello distale rispetto a quella predetta da VarSeq. Non è stato possibile effettuare un'analisi di segregazione, in quanto non erano disponibili i DNA dei genitori del probando.

Il paziente 8 è un maschio di 7 anni con ritardo globale dello sviluppo. L'analisi bioinformatica per la ricerca delle CNVs eseguita sull'esoma del paziente ha portato all'identificazione di una duplicazione in eterozigosi, di circa 2.5 Mb, localizzata a livello del *locus* genico 22q11.21 (chr22:18,893,886-21,414,817dup). Questa duplicazione comprende tra gli altri i geni malattia presenti in OMIM: PRODH (OMIM: 606810), SLC25A1 (OMIM: 190315), CDC45 (OMIM: 603465), GP1BB (OMIM: 138720), TBX1 (OMIM: 602054), TXNRD2 (OMIM: 606448), COMT (OMIM: 116790), TANGO2 (OMIM: 616830), RTN4R (OMIM: 605566), SCARF2 (OMIM: 613619), PI4KA (OMIM: 600286), SERPIND1 (OMIM: 142360), SNAP29 (OMIM: 604202), LZTR1 (OMIM: 600574). La successiva analisi mediante *array*-CGH ha confermato la presenza di una duplicazione in eterozigosi, di circa 2.6 Mb, a livello della banda citogenetica 22q11.21 (chr22:18,880,598-21,505,417dup). Quest'ultima comprende al suo interno tutti i geni malattia OMIM già identificati da VarSeq. Tuttavia, la dimensione reale dell'evento si estende circa 100 kb in più a livello distale rispetto a quella predetta da VarSeq. L'analisi di segregazione eseguita sui genitori del probando ha rivelato la presenza della stessa duplicazione in eterozigosi anche nel padre.

Le duplicazioni dei probandi 1 e 8 interessano lo stesso *locus* genico. La variante portata dal probando 8 ha una dimensione minore e inizia a livello più prossimale rispetto a quella del probando 1. Inoltre, rispetto a quella del probando 1, la duplicazione del probando 8 coinvolge in più i geni PRODH e SLC25A1 a livello prossimale, mentre non include i geni GGT2 e MAPK1 a livello distale.

Duplicazioni di dimensioni variabili del *locus* genico 22q11.21 sono riportate in letteratura essere associate alla sindrome da duplicazione 22q11.2 autosomica dominante (OMIM: 608363). Si tratta di un raro disordine genomico

complementare, ma distinto dalla sindrome da delezione 22q11.2 [19], che include i disordini genomici della sindrome di DiGeorge (DGS) e della sindrome velo-cardio-facciale (VCFS) [20]. Questi riarrangiamenti sono verosimilmente il risultato di una NAHR durante la meiosi che utilizza come substrati per la ricombinazione le diverse LCRs presenti a livello del cromosoma 22 (LCR22s). Questo fenomeno di crossing-over ineguale può portare alla formazione di reciproche duplicazioni e delezioni in questa regione [20]. Il fenomeno di NAHR dà origine a CNVs ricorrenti [8]. In particolare, in questa regione si osservano frequenti riarrangiamenti, in quanto diverse LCRs possono essere usate alternativamente come substrato di ricombinazione [19] dando origine a delezioni e duplicazioni con diversi punti di rottura che variano in dimensione. Le duplicazioni variano in dimensione tra 1.5 Mb e 6 Mb [20] e includono la tipica duplicazione comune di 3 Mb e quella di 1.5 Mb interna alla prima [19]. In generale, sono riportati sia pazienti con varianti insorte *de novo* che ereditate da un genitore fenotipicamente normale o solo con anomalie lievi/minori [19][20]. Questa osservazione è in accordo con quanto riscontrato nel caso del probando 8, in cui la duplicazione è stata identificata anche nel padre non affetto. Questo trova riscontro in diversi riarrangiamenti ricorrenti, che possono essere associati ad un'ampia gamma e gravità di fenotipi in funzione di fattori genetici, epigenetici e/o non genetici/ambientali [17]. Lo studio di Ou et al. (2008) conferma il fenomeno dell'espressività variabile all'interno della stessa famiglia. Gli stessi autori affermano che virtualmente tutti gli individui con queste duplicazioni hanno un certo grado di ritardo dello sviluppo [19]. La sindrome da duplicazione 22q11.2 è caratterizzata da un fenotipo clinico estremamente variabile che spazia da normale a caratterizzato da disabilità intellettiva e difetti congeniti. Le caratteristiche fenotipiche più frequentemente osservate includono disabilità intellettiva/difficoltà di apprendimento, ritardo dello sviluppo psicomotorio, ritardo della crescita e ipotonia. Sono osservati anche disturbi comportamentali. I dismorfismi più comuni sono: ipertelorismo, naso largo e piatto, micrognazia, insufficienza velofaringea con o senza palatoschisi, orecchie displastiche, epicanto e *down-slanting* palpebrale [20]. Inoltre, i pazienti affetti dalla sindrome da duplicazione 22q11.2 hanno un aumentato rischio di sviluppare una varietà di complicazioni, incluse anomalie neurologiche, oftalmologiche, craniofacciali, otorinolaringoiatriche, cardiache, immunologiche, gastrointestinali, genitourinarie, muscoloscheletriche, endocrinologiche, cognitive e psichiatriche [21]. Il paziente 1 analizzato in questo studio presenta ritardo mentale medio/lieve

e ritardo psicomotorio (caratteristico della sindrome di Pitt-Hopkins), che sono due delle principali caratteristiche osservate nella sindrome da duplicazione 22q11.2. La bambina si presenta inoltre con un difetto genitourinario, microcefalia, cefalea e reflusso gastroesofageo che sono tutti sintomi che si possono riscontrare come complicazioni di questo disordine genomico [21]. Per il paziente 8 non è stata riportata una dettagliata descrizione fenotipica, tuttavia il ritardo globale dello sviluppo è una delle caratteristiche principali osservate nella sindrome da duplicazione 22q11.2, ed è infatti manifestata anche dall'altro probando dello studio con la mutazione nello stesso *locus* genico. Il disordine cromosomico in esame comprende tra gli altri il gene TBX1 (OMIM: 602054), considerato il principale gene malattia causativo della sindrome da delezione 22q11.2 [19][20]. Mutazioni patogenetiche in questo gene possono causare l'esordio delle sindromi DGV/VCFS autosomiche dominanti (OMIM: 188400; 192430). TBX1 è un fattore di trascrizione appartenente alla famiglia dei geni *T-box* coinvolto nella regolazione di processi dello sviluppo. La regolazione trascrizionale è uno dei processi critici maggiormente implicati nell'esordio di NDDs se alterato [3]. Nei modelli di topo è stato visto che sia una sua sovra- che sottoespressione può provocare il fenotipo DGV/VCFS; in più, sono state riportate mutazioni *gain of function* in questo gene in pazienti con un fenotipo sovrapponibile alle due sindromi citate [19]. Tali studi suggeriscono che un dosaggio genico fuori dal *range* di normalità per TBX1 possa colpire lo stesso *pathway* dello sviluppo nell'uomo e provocare sia la sindrome da delezione 22q11.2 che la sindrome da duplicazione 22q11.2 [19]. Il gene TBX1 è stato riscontrato duplicato in entrambi i pazienti di questo studio. Inoltre, l'aumentata espressione di questo gene, insieme ad altri fattori genetici, epigenetici o non genetici, è stata suggerita come una possibile spiegazione per l'ampia variabilità fenotipica osservata [19][20]. Un'ulteriore spiegazione è che il dosaggio genico alterato di TBX1 possa risultare in una sottoregolazione dei suoi geni *target* [20]. Le duplicazioni riscontrate nei pazienti 1 e 8 sono delle varianti di dimensioni elevate che includono al loro interno numerosi geni, tra cui diversi geni noti essere associati a malattie. Esse rappresentano delle varianti localizzate su un *hot spot* di ricombinazione. Inoltre, duplicazioni di dimensioni variabili del *locus* genico interessato da tali eventi sono riportate in letteratura essere associate a NDDs [19][20]. Pertanto, in accordo con quanto redatto da ACMG (2015) (**Tabella 16** in Appendice), in associazione alla consultazione della letteratura scientifica, le

duplicazioni riscontrate nei pazienti 1 e 8 sono da considerarsi patogenetiche e la loro presenza è compatibile con i fenotipi clinici indicati per i probandi.

Paziente 2

Il paziente 2 è un maschio di 37 anni affetto da disabilità intellettiva sindromica. È figlio di genitori non consanguinei e rappresenta un caso isolato in famiglia. Egli presenta turribrachicefalia, malformazione di Dandy-Walker (ipoplasia del verme cerebellare e cisti del IV ventricolo), epilessia, micrognazia, plica unica, riduzione IV-V metatarsale, clinodattilia del V dito, criptorchidismo, ipogenitalismo e difetto interventricolare. L'analisi bioinformatica per la ricerca delle CNVs eseguita sull'esoma del paziente ha portato all'identificazione di una duplicazione in eterozigosi, di circa 19 Mb, localizzata a livello del *locus* genico Xp22.33p22.11 (chrX:2,852,873-21,900,773dup). Questa duplicazione comprende tra gli altri i geni malattia presenti in OMIM: ARSE (OMIM: 300180), NLGN4X (OMIM: 300427), STS (OMIM: 300747), ANOS1 (OMIM: 300836), TBL1X (OMIM: 300196), GPR143 (OMIM: 300808), CLCN4 (OMIM: 302910), MID1 (OMIM: 300552), HCCS (OMIM: 300056), AMELX (OMIM: 300391), MSL3 (OMIM: 300609), FRMPD4 (OMIM: 300838), TLR7 (OMIM: 300365), TLR8 (OMIM: 300366), TRAPPC2 (OMIM: 300202), OFD1 (OMIM: 300170), GLRA2 (OMIM: 305990), FANCB (OMIM: 300515), PIGA (OMIM: 311770), AP1S2 (OMIM: 300629), NHS (OMIM: 300457), CDKL5 (OMIM: 300203), RS1 (OMIM: 300839), PHKA2 (OMIM: 300798), ADGRG2 (OMIM: 300572), PDHA1 (OMIM: 300502), SH3KBP1 (OMIM: 300374), RPS6KA3 (OMIM: 300075), CNKSR2 (OMIM: 300724), SMPX (OMIM: 300226), MBTPS2 (OMIM: 300294). La successiva analisi mediante *array*-CGH ha confermato la presenza di una duplicazione con guadagno di due copie rispetto al normale, di circa 19.3 Mb, a livello della banda citogenetica Xp22.33p22.11 (chrX:2,701,273-21,962,375dup). Quest'ultima comprende al suo interno tutti i geni malattia OMIM già identificati da VarSeq, più il gene malattia SMS (OMIM: 300105). La dimensione reale dell'evento è un po' più estesa rispetto a quella predetta da VarSeq. Tuttavia, sebbene VarSeq abbia identificato l'evento duplicativo, quest'ultimo è stato predetto dal programma come un guadagno di una sola copia e non di due copie in più rispetto al normale, come invece è stato rilevato tramite la validazione in laboratorio. Non è stato possibile effettuare un'analisi di segregazione, in quanto non erano disponibili i DNA dei genitori del probando.

La disabilità intellettiva X-linked è una condizione comune, clinicamente complessa e geneticamente eterogenea, che deriva da SNVs e/o CNVs in molti geni localizzati sul cromosoma X [9]. Lintas et al. (2016) riportano il caso di un uomo di 30 anni in cui è stata riscontrata una duplicazione *de novo*, di 19.98 Mb, a livello della banda citogenetica Xp22.33p22.12 (chrX:61,091-20,038,601dup). Il loro paziente è caratterizzato da disabilità intellettiva, ipotonia congenita e dismorfismi facciali. Si tratta di un quadro clinico compatibile con quello della disabilità intellettiva sindromica di cui è affetto il probando 2. In aggiunta, il paziente di Lintas et al. nel corso della sua storia clinica ha presentato criptorchidismo bilaterale, deformità scheletriche, lieve asimmetria dei ventricoli laterali e prollasso delle valvole mitrale e tricuspide. Anomalie genitali, scheletriche, cerebrali e cardiache sono manifestate anche dal probando 2. La duplicazione del paziente di Lintas et al. è più distale rispetto a quella identificata nel probando 2. Essa contiene tra gli altri i geni malattia presenti in OMIM: SHOX (OMIM: 312865), CSF2RA (OMIM: 306250), ARSE (OMIM: 300180), NLGN4X (OMIM: 300427), STS (OMIM: 300747), ANOS1 (OMIM: 300836), TBL1X (OMIM: 300196), GPR143 (OMIM: 300808), CLCN4 (OMIM: 302910), MID1 (OMIM: 300552), HCCS (OMIM: 300056), AMELX (OMIM: 300391), MSL3 (OMIM: 300609), FRMPD4 (OMIM: 300838), TLR7 (OMIM: 300365), TLR8 (OMIM: 300366), TRAPPC2 (OMIM: 300202), OFD1 (OMIM: 300170), GLRA2 (OMIM: 305990), FANCB (OMIM: 300515), PIGA (OMIM: 311770), AP1S2 (OMIM: 300629), NHS (OMIM: 300457), CDKL5 (OMIM: 300203), RS1 (OMIM: 300839), PHKA2 (OMIM: 300798), ADGRG2 (OMIM: 300572), PDHA1 (OMIM: 300502), SH3KBP1 (OMIM: 300374). Il paziente di Lintas et al., rispetto al probando 2, presenta in più i geni SHOX e CSF2RA a livello distale, però non include i geni RPS6KA3, CNKSR2, SMPX, MBTPS2 e SMS a livello prossimale. Entrambe le duplicazioni comprendono però geni malattia OMIM associati a NDDs:

- NLGN4X, associato ad autismo, disabilità intellettiva X-linked e sindrome di Asperger;
- CLCN4, associato a disabilità intellettiva X-linked;
- MSL3, associato a disabilità intellettiva sindromica X-linked;
- FRMPD4, associato a disabilità intellettiva X-linked;
- OFD1, associato alla sindrome di Joubert;
- GLRA2, associato a disabilità intellettiva sindromica X-linked;
- PIGA, associato a encefalopatia epilettica e dello sviluppo e a disordine del neurosviluppo con epilessia e emocromatosi;

- AP1S2, associato a disabilità intellettiva sindromica X-linked;
- CDKL5, associato a encefalopatia epilettica e dello sviluppo.

Altri geni importanti nel contesto dei NDDs e presenti solo nel paziente 2 di questo studio sono:

- RPS6KA3, associato alla sindrome di Coffin-Lowry e a disabilità intellettiva X-linked;
- CNKSR2, associato a disabilità intellettiva sindromica X-linked;
- SMS, associato a disabilità intellettiva sindromica X-linked.

Come affermato da Lintas et al., a causa dell'elevato numero di geni potenzialmente patogenetici, non si può correlare il fenotipo clinico ad un singolo gene, ma piuttosto esso potrebbe derivare dall'alterazione del dosaggio di più geni.

La duplicazione riscontrata nel paziente 2 è una variante di dimensioni elevate che include al suo interno numerosi geni, tra cui diversi geni noti essere associati a malattie, compresi i NDDs. Inoltre, in letteratura e nei database dedicati sono riportate duplicazioni coinvolgenti il braccio corto del cromosoma X, associate a NDDs e descritte in maschi affetti da disabilità intellettiva e varie anomalie congenite [9]. Pertanto, in accordo con quanto redatto da ACMG (2015) (**Tabella 16** in Appendice), in associazione alla consultazione della letteratura scientifica, la duplicazione riscontrata nel paziente 2 è da considerarsi potenzialmente patogenetica e la sua presenza è compatibile con il fenotipo clinico indicato per il probando.

Paziente 3

Il paziente 3 è un maschio di 12 anni affetto da disabilità intellettiva, disturbi specifici dell'apprendimento e ritardo delle acquisizioni. È nato a termine da parto eutocico dopo gravidanza decorsa con ripetute minacce d'aborto; i test di screening e le ecografie periodiche erano nella norma. È stata riferita ipossia perinatale e necessità di rianimazione neonatale. Il successivo periodo neonatale si è svolto in maniera regolare, con allattamento materno e suzione valida; tuttavia, lo svezzamento è risultato difficoltoso. Sono inoltre indicati i seguenti aspetti: presenza di ritmo sonno-veglia irregolare; accrescimento staturponderale regolare; predisposizione al contrarre infezioni respiratorie e intestinali; riferita allergia al frumento; dentizione in epoca; mai crisi comiziali,

neanche febbrili. Per quanto riguarda lo sviluppo psicomotorio: controllo del capo e stazione seduta acquisiti dopo l'anno e mezzo; cammino autonomo dai 18-20 mesi; prime parole dall'anno e mezzo, tuttavia il linguaggio è sempre risultato molto scarso e incomprensibile, la prima frase completa è stata prodotta intorno ai 4-5 anni. Continenza sfinterica raggiunta a 5 anni, con necessità di pannolino durante la notte per enuresi fino a 7 anni. Sono stati riferiti disturbi nella coordinazione e nell'equilibrio, assenza del riflesso alla caduta, soglia del dolore elevata. Il probando risulta quindi caratterizzato da una franca disprassia verbale e motoria. Per quanto riguarda l'ambito comportamentale, il bambino manifesta iperattività, irrequietezza e impulsività, oltre che una mancata percezione dei pericoli. Inoltre, la socializzazione con i pari risulta essere molto difficoltosa. Per queste ragioni, è stata fatta diagnosi di disturbo dello spettro autistico e disturbo comportamentale con ossessioni. Il probando è stato sottoposto ad un intervento di avulsione dentaria per carie, a seguito del quale ha manifestato un vasto ematoma del mento; allo studio coagulativo è stato evidenziato un deficit di attività del fattore di von Willebrand. All'esame obiettivo si presenta con una circonferenza cranica di 54 cm; non discromie; arti indenni, con modesta iperlassità degli arti superiori; palato indenne; non dismorfismi. L'analisi bioinformatica per la ricerca delle CNVs eseguita sull'esoma del paziente ha portato all'identificazione di una delezione in eterozigosi coinvolgente gli esoni 2-4 del gene SOX6 (OMIM: 607257), localizzato a livello della banda citogenetica 11p15.2. La successiva analisi mediante *real-time* PCR ha confermato la presenza di una delezione in eterozigosi coinvolgente gli esoni 1-4 del gene SOX6. È plausibile che VarSeq non abbia identificato anche l'esone 1 come incluso nella delezione, in quanto a volte può capitare che gli esoni a monte e a valle della sequenza abbiano dei problemi di qualità di sequenziamento. Non è stato possibile effettuare un'analisi di segregazione, in quanto non erano disponibili i DNA dei genitori del probando.

SOX6 appartiene alla famiglia dei geni SOX, i quali codificano fattori di trascrizione che controllano il destino e il differenziamento cellulare in molti processi dello sviluppo e nell'adulto. In particolare, SOX6 è coinvolto nei processi di neurogenesi e scheletogenesi. La regolazione trascrizionale è uno dei processi critici maggiormente implicati nell'esordio di NDDs se alterato [3]. Varianti in metà dei geni SOX sono state associate nell'uomo a disordini dello sviluppo, noti come SOXopatie, e varianti in diversi geni SOX sono state associate nello specifico a disabilità intellettiva e altre caratteristiche [10]. Tolchin et al. (2020) hanno

dimostrato come anche varianti associate ad un'aploinsufficienza del gene SOX6 siano causa di una SOXopatia del neurosviluppo. Il loro studio riporta i casi di 19 individui, provenienti da 17 famiglie non imparentate, con varianti in SOX6 e che condividono ritardo delle acquisizioni e/o disabilità intellettiva, più altre anomalie incostanti, inclusi dismorfismi lievi, craniosinostosi e osteocondroma. Si tratta di varianti private in ogni individuo affetto (a parte i parenti); tutte varianti in eterozigosi; 14 *de novo*, 4 ereditate da un padre affetto e 1 ereditata da un padre non affetto con mosaicismo del 22%. 8 dei 19 individui presentano una CNV e, nello specifico, sono tutte delezioni parziali del gene che non includono altri geni. Delezioni intrageniche sono predette inattivare l'allele mutato di SOX6 e pertanto i fenotipi patologici risultanti riflettono l'aploinsufficienza di questo gene. Un paziente nello studio di Tolchin et al. presenta inoltre la stessa delezione del probando 3 che coinvolge gli esoni 1-4 del gene in questione. Il loro paziente è un maschio di 10 anni affetto da disabilità intellettiva lieve, ADHD e ansia; nel suo caso la delezione risulta ereditata dal padre sano, che presenta un mosaicismo del 22%. In generale, tutti i soggetti condividono caratteristiche cliniche di una sindrome del neurosviluppo che include ritardo dello sviluppo e disabilità intellettiva da lieve a severa, associata in modo incostante con anomalie scheletriche, quali dismorfismi facciali lievi, craniosinostosi e osteocondroma. In conclusione, l'aploinsufficienza del gene SOX6 è associata alla sindrome di Tolchin-Le Caignec autosomica dominante (OMIM: 618971). Quest'ultima è caratterizzata da: dismorfismi facciali non specifici e variabili; difetti scheletrici variabili; ritardo dello sviluppo e disabilità intellettiva da lieve a severa. Per quanto riguarda il comportamento, si osservano caratteristiche autistiche, ADHD, comportamento aggressivo, disturbi del sonno, fragilità emotiva, ansia, irrequietezza, atteggiamento oppositivo, mancanza del senso del pericolo. L'esordio è precoce e il fenotipo variabile. La delezione riscontrata nel paziente 3 è una variante *loss of function* che comprende più esoni di un singolo gene. L'aploinsufficienza di quest'ultimo è associata ad una sindrome del neurosviluppo. Inoltre, la stessa variante riscontrata nel probando di questo studio è stata riportata in letteratura [10]. Pertanto, in accordo con quanto redatto da ACMG (2015) (**Tabella 16** in Appendice), in associazione alla consultazione della letteratura scientifica, la delezione riscontrata nel paziente 3 è da considerarsi patogenetica e la sua presenza è compatibile con il fenotipo clinico indicato per il probando.

Pazienti 4 e 6

Il paziente 4 è un maschio di 38 anni affetto da encefalopatia congenita epilettogena con grave insufficienza mentale, disturbo comportamentale, dismorfismi craniofacciali con macrocranio e steatosi epatica. Ha sviluppato la capacità di mantenere la posizione seduta a 9 mesi e di camminare a 18 mesi; lallazione a un anno, tuttavia non ha mai maturato un linguaggio verbale adeguato. Sono state escluse sindromi ormonali e/o dismetaboliche, tossiche e infettive, e la sindrome di Angelman. È caratterizzato da disabilità intellettiva ed autismo secondario. Dal 2003 ha iniziato a manifestare una sintomatologia caratterizzata da tremore, agitazione psicomotoria, alterazioni del ritmo sonno-veglia. Nel 2008 si è aggiunto un disordine del comportamento alimentare con alternanza di atteggiamenti bulimici e di rifiuto completo di cibo e liquidi. Presenta inoltre atteggiamenti auto-lesionistici. All'esame obiettivo: facies peculiare per peluria frontale e fissità dello sguardo con lagofalmo, severa congiuntivite con teleangectasie, bocca e orecchie grandi, parziale sindattilia del secondo e terzo dito dei piedi. L'analisi bioinformatica per la ricerca delle CNVs eseguita sull'esoma del paziente ha portato all'identificazione di una delezione in eterozigosi, di circa 872.26 kb, localizzata a livello del *locus* genico 22q13.33 (chr22:50,297,486-51,169,740del). Questa delezione comprende tra gli altri i geni malattia presenti in OMIM: ALG12 (OMIM: 607144), MLC1 (OMIM: 605908), MOV10L1 (OMIM: 605794), TUBGCP6 (OMIM: 610053), SBF1 (OMIM: 603560), SCO2 (OMIM: 604272), TYMP (OMIM: 131222), CHKB (OMIM: 612395), ARSA (OMIM: 607574), SHANK3 (OMIM: 606230). La successiva analisi mediante *array*-CGH ha confermato la presenza di una delezione in eterozigosi, di circa 1.8 Mb, a livello della banda citogenetica 22q13.32.q13.33 (chr22:49,388,594-51,224,252del). Quest'ultima comprende al suo interno tutti i geni malattia OMIM già identificati da VarSeq, più a livello distale il gene malattia ACR (OMIM: 102480). La dimensione reale dell'evento si estende circa 900 kb in più a livello prossimale rispetto a quella predetta da VarSeq. Non è stato possibile effettuare un'analisi di segregazione, in quanto non erano disponibili i DNA dei genitori del probando.

Il paziente 6 è un maschio di 30 anni con ipoplasia cerebellare e deficit cognitivo severo; atassia deambulatoria, tremore cinetico e posturale; e dismorfismi. La gravidanza e il parto sono decorsi nella norma. È stata riscontrata difficoltà di relazione già dai primi mesi di vita, nonché ritardo di acquisizione delle tappe dello

sviluppo psicomotorio (deambulazione a 18 mesi di vita, importante ritardo di acquisizione del linguaggio, costituito al momento del prelievo (2019) da parole e qualche frase minima). L'analisi del gene FMR1 è risultata nella norma. Il dosaggio di arisulfatasi A è stato eseguito nel 1991 ed è stato descritto come ridotto (non disponibile il referto). L'andatura è stata definita ciondolante e a base allargata. L'RMN cerebrale eseguita nel 1992 ha evidenziato ipoplasia del verme e degli emisferi cerebellari; nel tempo il quadro si è mantenuto essenzialmente stazionario. La valutazione neurologica nel 2016 ha indicato atassia deambulatoria, tremore cinetico e posturale. Nel 2017 è stato sottoposto a nefrectomia del rene destro, a seguito di malformazione dell'uretere e conseguente idronefrosi ingravescente. All'esame obiettivo statura, peso e circonferenza cranica risultano nei limiti di norma; sono tuttavia presenti alcune peculiarità morfologiche del volto (fronte sfuggente, lieve *synophrys*, strabismo, filtro poco modellato, ipoplasia malare, lieve micrognazia, ipomimia), collo tozzo e slargato, lieve sproporzione span/statura, lieve aracnodattilia e rigidità articolare, gomito valgo bilaterale. Non si riscontrano invece anomalie morfologiche significative a carico degli altri distretti. L'anamnesi familiare è negativa. L'analisi bioinformatica per la ricerca delle CNVs eseguita sull'esoma del paziente ha portato all'identificazione della stessa variante identificata nel probando 4, cioè una delezione in eterozigosi, di circa 872.26 kb, localizzata a livello del *locus* genico 22q13.33 (chr22:50,297,486-51,169,740del). Questa delezione comprende tra gli altri i geni malattia presenti in OMIM citati precedentemente. La successiva analisi mediante *array*-CGH ha confermato la presenza di una delezione in eterozigosi, di circa 1.65 Mb, a livello della banda citogenetica 22q13.33 (chr22: 49,531,017-51,178,257del). Quest'ultima comprende al suo interno tutti i geni malattia OMIM già identificati da VarSeq, più a livello distale il gene malattia ACR (OMIM: 102480). Anche in questo caso la dimensione reale dell'evento si estende circa 800 kb in più a livello prossimale rispetto a quello predetto da VarSeq. Non è stato possibile effettuare un'analisi di segregazione, in quanto non erano disponibili i DNA dei genitori del probando. Le delezioni dei probandi 4 e 6 interessano lo stesso *locus* genico e includono gli stessi geni malattia presenti in OMIM.

Delezioni di dimensioni variabili del *locus* genico 22q13.33 sono riportate in letteratura essere associate alla sindrome da delezione 22q13.3 autosomica dominante (OMIM: 606232), anche conosciuta come sindrome di Phelan-McDermid. Si tratta di una sindrome da geni continui risultante dalla delezione del

braccio lungo del cromosoma 22 distale [22]. Per la maggior parte si tratta di delezioni terminali che vanno da < 50 kb a > 9 Mb [23]. L'ampio range di dimensioni della delezione è accompagnato da un ampio spettro di variabilità fenotipica [22]. Le principali caratteristiche fenotipiche associate a questa sindrome sono riportate in **Tabella 9**.

Tabella 9. Principali caratteristiche fenotipiche associate alla sindrome di Phelan-McDermid [24].

| | |
|--|--|
| Central nervous system: behavioral and neurological manifestations | Global developmental delay; absence of expressive language (50%) or speech delay (100%); moderate-to-profound intellectual disability (about 87%); ASD (26%–84%); aggressive behavior (28%); impulsiveness (47%); biting (self or others) (46%); abnormal reflexes (48%); elevated pain threshold (77%); seizures (febrile and/or nonfebrile) (14%–41%); sleep disturbance (41%–46%); abnormal brain MRI (7%–75%) (arachnoid cyst, delayed myelination, generalized white matter atrophy, nonspecific white matter hyperintensities or gliosis, thinning or hypoplasia of the corpus callosum, ventriculomegaly, focal cortical atrophy, frontal lobe hypoplasia, enlarged cisterna magna, cerebellar vermis hypoplasia); neurologic, motor and cognitive regression in early childhood (18%); regression during adolescence within 1–3 years after the onset of psychiatric illness (66%–84%); cataplexy (53%); probable progressive neurodegeneration. |
| Central and peripheral musculoskeletal signs and symptoms | Generalized hypotonia (weak cry, poor head control, feeding difficulties, etc.) (29%–100%); delayed motor milestones; inability to walk; ataxia; muscle weakness; dynamic and static balance deficits; decreased upper limb strength; gross and fine motor coordination deficits; broad-based and unsteady gait; oral motor dysfunction; bladder dysfunction; scoliosis; hyperextensible joints (25%–61%); lax ligaments (65%). |
| Cardiovascular system | Heart defects, including tricuspid valve regurgitation, aortic regurgitation, patent ductus arteriosus, total anomalous venous return, atrial septal defect, and so on (3% to >25%). |
| Genitourinary system | Renal abnormalities or malformations, including vesicoureteral reflux, increased kidney size, dilated renal pelvis, renal cysts or polycystic kidney, hydronephrosis, horseshoe kidney, renal agenesis or dysplastic kidneys, duplicate kidney, and so on (17%–38%); urinary tract infections (8%). |
| Gastrointestinal system | Gastroesophageal reflux (42%–44%); constipation and/or diarrhea (38%–41%); cyclic vomiting, feeding difficulties and pica. |
| Endocrine-metabolic system | Hypothyroidism (3%–6%); precocious or delayed puberty (0%–12%); impaired hepatic function and autoimmune hepatitis and liver failure (rare). |
| Sensory organs (hearing and vision) | Cortical auditory and visual impairment (extensive use of peripheral vision, difficulty in processing cluttered images, problems with depth perception, difficulty discerning spoken words from background noise); hyperopia and myopia, blindness and optic nerve hypoplasia (rare). |
| Dysmorphic features | Bulbous nose (47%–80%); ears anomalies (27%–86%); long eyelashes (37%–93%); large fleshy hands (33%–68%); epicanthal folds (30%–73%); periorbital fullness (25%–60%); thin and dysplastic toenails (3%–78%); pointed chin (22%–62%); dolichocephaly (0%–86%); high arched palate (25%–47%); 2/3 toe syndactyly (9%–48%); ptosis (3%–57%); strabismus (26%); full cheeks (25%); sacral dimple (13%–37%); hypertelorism (13%–36%); deep-set eyes (6%–31%); macrocephaly (7%–31%); wide nasal bridge (16%); microcephaly (6%–14%). |
| Body growth | Short stature/delayed growth (0%–13%); tall stature/accelerated growth (3%–18%). |
| Miscellaneous | Lymphedema (22%–29%); dental problems (malocclusion, wide-spaced teeth, crowding); recurring upper respiratory tract infections (8%–53%); decreased skin perspiration with tendency to overheat (60%); skin rashes (39%); immune deficiency (12%). |

Abbreviations: ASD, autism spectrum disorder; MRI, magnetic resonance imaging.

I sintomi clinici presentati dai pazienti 4 e 6 trovano riscontro all'interno della suddetta tabella. Entrambi i probandi presentano grave disabilità intellettiva, disturbi comportamentali, dismorfismi e ritardo dello sviluppo psicomotorio. Il probando 4 presenta inoltre alterazioni del ritmo sonno-veglia, disordine alimentare ed epilessia. Il probando 6, invece, presenta anche ipoplasia cerebellare.

In generale, le ampie differenze cliniche che talvolta si osservano tra pazienti con delezioni identiche e il grado di gravità manifestato da alcuni pazienti con

mutazioni molto piccole, indicano un ruolo critico di fattori diversi dalla dimensione della delezione nel plasmare la sindrome a livello clinico [24]. Questi fattori includono ad esempio varianti in altri geni, fattori epigenetici e/o fattori ambientali [24].

Praticamente tutte le delezioni includono il gene SHANK3 (OMIM: 606230). Quest'ultimo è stato riscontrato deleto in entrambi i pazienti dello studio. SHANK3 codifica una proteina *scaffold* nella densità postsinaptica delle sinapsi eccitatorie che lega i recettori di membrana all'actina del citoscheletro [22]. La segnalazione sinaptica è uno dei processi critici maggiormente implicati nell'esordio di NDDs se alterato [3]. A causa della funzione importante che svolge, i suoi livelli devono essere strettamente controllati. La maggior parte dei difetti neurologici e comportamentali osservati nella sindrome da delezione 22q13.3 sono dovuti all'aploinsufficienza di questo gene [22]. Infatti, la sindrome di Phelan-McDermid, oltre che ad una delezione coinvolgente la banda citogenetica 22q13.3, può essere dovuta alla presenza di varianti patogenetiche a livello di questo gene [23]. Tuttavia, anche altri geni aploinsufficienti in questa regione possono svolgere un ruolo importante e modulare il fenotipo di ogni paziente [22]. Tra questi, il gene MAPK8IP2 (OMIM: 607755), coinvolto nella delezione di entrambi i pazienti in esame, potrebbe svolgere un ruolo nel determinare le caratteristiche neurologiche della maggior parte dei pazienti con questa sindrome, sebbene resti da vedere se questo gene mostri aploinsufficienza [22].

Nella maggior parte dei casi la delezione insorge *de novo*, ma può anche essere ereditata da un genitore che presenta un riarrangiamento bilanciato o un mosaicismo [23].

Le delezioni riscontrate nei pazienti 4 e 6 sono delle varianti di dimensioni elevate che includono al loro interno numerosi geni, tra cui diversi geni noti essere associati a malattie. Esse rappresentano delle varianti *loss of function*, localizzate su un *hot spot* di ricombinazione. Inoltre, delezioni di dimensioni variabili del *locus* genico interessato da tali eventi sono riportate in letteratura essere associate a NDDs [22]. Pertanto, in accordo con quanto redatto da ACMG (2015) (**Tabella 16** in Appendice), in associazione alla consultazione della letteratura scientifica, le delezioni riscontrate nei pazienti 4 e 6 sono da considerarsi patogenetiche e la loro presenza è compatibile con i fenotipi clinici indicati per i probandi.

Paziente 5

Il paziente 5 è una femmina di 20 anni con deficit cognitivo cui si associano lievi note non specifiche o significativamente orientanti. Il probando presenta brachidattilia del V dito bilaterale e del IV dito della mano sinistra; guance paffute e fronte stretta, ipertelorismo per telecanto, angoli della bocca rivolti verso il basso, rime palpebrali ristrette e micrognazia. La madre e la sorella sono ipotiroidee; la restante storia familiare appare negativa e non pertinente. A 18 anni, per verifica familiarità, è stato riscontrato ipotiroidismo subclinico non trattato. L'analisi bioinformatica per la ricerca delle CNVs eseguita sull'esoma del paziente ha portato all'identificazione di una delezione in eterozigosi coinvolgente gli esoni 5-6 del gene SETBP1 (OMIM: 611060), localizzato a livello della banda citogenetica 18q12.3. La successiva analisi mediante *real-time* PCR ha confermato la presenza di una delezione in eterozigosi coinvolgente gli esoni 5-6 del gene SETBP1. L'analisi di segregazione eseguita sui genitori del probando ha indicato un'insorgenza *de novo* per la variante in esame.

SETBP1 codifica una *SET-binding protein*, che regola i processi di trascrizione legandosi a diverse regioni promotoriali. La regolazione trascrizionale è uno dei processi critici maggiormente implicati nell'esordio di NDDs se alterato [3]. Anche se questa proteina è altamente espressa durante lo sviluppo cerebrale, le sue funzioni precise e i meccanismi molecolari nei *pathways* cerebrali e neuronali sono ancora ampiamente sconosciuti [11]. L'aploinsufficienza di questo gene si associa al *SETBP1 haploinsufficiency disorder* autosomico dominante, tipicamente causato da varianti patogenetiche *de novo*. Le caratteristiche cliniche più comunemente riscontrate sono lieve ritardo dello sviluppo motorio e ipotonia, disordini del linguaggio, disabilità intellettiva, ADHD, errori refrattivi e strabismo [11].

In conclusione, l'aploinsufficienza del gene SETBP1 è associata alla disabilità intellettiva 29 autosomica dominante (OMIM: 616078). Quest'ultima è caratterizzata da: dismorfismi facciali variabili; disabilità intellettiva da lieve a moderata; ritardo del linguaggio o assenza di linguaggio; ritardo dello sviluppo motorio; e ADHD.

La delezione riscontrata nel paziente 5 è una variante *loss of function de novo* che comprende più esoni di un singolo gene. L'aploinsufficienza di quest'ultimo è associata ad una sindrome del neurosviluppo. Inoltre, in letteratura e nei database dedicati sono riportate delezioni di dimensioni maggiori che includono l'intero gene o anche altri geni a livello della banda citogenetica 18q12.2q21.1, associate

in maniera variabile a ritardo dello sviluppo, ritardo del linguaggio espressivo, dimorfismi facciali e problemi comportamentali [25]. Pertanto, in accordo con quanto redatto da ACMG (2015) (**Tabella 16** in Appendice), in associazione alla consultazione della letteratura scientifica, la delezione riscontrata nel paziente 5 è da considerarsi potenzialmente patogenetica e la sua presenza è compatibile con il fenotipo clinico indicato per il probando.

Paziente 7

Il paziente 7 è un maschio di 29 anni con sospetta sindrome di Bardet-Biedl, disabilità intellettiva e dimorfismi del viso. La sindrome di Bardet-Biedl è una ciliopatia con coinvolgimento multisistemico. La malattia è caratterizzata dall'associazione tra obesità, retinite pigmentosa, polidattilia, problemi renali, ipogonadismo e difficoltà d'apprendimento, che di solito si presenta diversi anni dopo l'esordio della malattia. L'espressione clinica è variabile; è presente eterogeneità genetica; e nella maggior parte dei casi la malattia viene trasmessa come carattere autosomico recessivo. L'analisi bioinformatica per la ricerca delle CNVs eseguita sull'esoma del paziente ha portato all'identificazione di una delezione in eterozigosi, di circa 7.2 Mb, localizzata a livello del *locus* genico 3p14.3p14.1 (chr3:56,696,413-63,898,903del). Questa delezione comprende tra gli altri i geni malattia presenti in OMIM: IL17RD (OMIM: 606807), HESX1 (OMIM: 601802), APPL1 (OMIM: 604299), FLNB (OMIM: 603381), DNASE1L3 (OMIM: 602244), PDHB (OMIM: 179060), ACOX2 (OMIM: 601641), ATXN7 (OMIM: 607640). La successiva analisi mediante *array*-CGH ha confermato la presenza di una delezione in eterozigosi, di circa 7.2 Mb, a livello della banda citogenetica 3p14.3p14.1 (chr3:56,701,139-63,901,725del). Quest'ultima corrisponde per dimensione e coordinate alla delezione identificata da VarSeq, e pertanto comprende al suo interno tutti i geni malattia OMIM già identificati bioinformaticamente. L'analisi di segregazione eseguita sui genitori del probando ha indicato un'insorgenza *de novo* per la variante in esame.

Delezioni coinvolgenti la banda citogenetica 3p14, e ad insorgenza *de novo*, sono riportate in letteratura essere associate a NDDs [13][26]. Le delezioni interstiziali del braccio corto del cromosoma 3 sono rare. I riarrangiamenti variano in dimensione e si sovrappongono in un'ampia regione a partire da 3p12.1 fino a 3p14.3. Le sottoregioni coinvolte in questi riarrangiamenti contengono diversi geni

candidati responsabili di caratteristiche comuni quali disabilità intellettiva, ritardo dello sviluppo neurocognitivo e dismorfismi. Ci sono anche geni candidati responsabili di disordini più specifici, come disturbo dello spettro autistico, epilessia, malformazioni che colpiscono il cervello, gli arti e altri distretti, e altre anomalie (come l'ipotonia) [13]. Diversi autori hanno proposto una regione *core*, sulla base del fenotipo comune, a livello della banda citogenetica 3p14.2. Parmeggiani et al. (2018) riportano il caso di una femmina di 60 anni con disabilità intellettiva, severo ritardo del linguaggio e lievi dismorfismi, in cui è stata riscontrata una delezione *de novo* di 500 kb a livello della parte prossimale della banda citogenetica 3p14.2. Tale delezione coinvolge i geni OMIM: FEZF2, CADPS e PTPRG, che risultano deleti anche nel probando 7. FEZF2 (OMIM: 607414) codifica un fattore di trascrizione necessario per la specificazione dei neuroni corticospinali. CADPS (OMIM: 604667) codifica una specifica proteina neuroendocrina, citosolica e di membrana periferica, richiesta per l'esocitosi mediata dagli ioni calcio delle vescicole secretorie. PTPRG (OMIM: 176886) codifica una proteina membro della famiglia proteica delle tirosin-fosfatasi, molecole di segnale che regolano molti processi cellulari, inclusi la crescita e il differenziamento cellulare, il ciclo mitotico e la trasformazione oncogenica. Vista l'importante funzione svolta da questi tre geni nel contesto dello sviluppo del sistema nervoso centrale, Parmeggiani et al. propongono che il fenotipo della loro paziente, che comprende i sintomi caratteristici delle delezioni 3p14, possa essere correlato all'aploinsufficienza di questi tre geni, supportando l'ipotesi che una regione *core* 3p14.2 nelle delezioni 3p prossimali sia associata con NDDs. Schwaibold et al. (2013), oltre ai già citati geni FEZF2 e CADPS, prendono in considerazione anche il gene SYNPR, che codifica la proteina sinaptica sinaptoporina, e risulta deleto anche nel probando 7. Questi autori propongono che la delezione di uno o più di questi tre geni, in particolare CADPS, possa essere responsabile della disabilità intellettiva. Inoltre, anche loro propongono la regione 3p14.2 come potenziale *core* per le delezioni 3p prossimali associate a disabilità intellettiva, ritardo dello sviluppo e altre caratteristiche minori variabili. Questi studi suggeriscono che i sintomi clinici nei pazienti con una delezione 3p prossimale possano essere causati da una sindrome da geni contigui. Quest'ultima può spiegare le differenze fenotipiche tra i pazienti in dipendenza della dimensione della regione deleta [26].

La delezione riscontrata nel paziente 7 è una variante *loss of function de novo* di dimensioni elevate che include al suo interno numerosi geni, tra cui diversi geni

noti essere associati a malattie. Inoltre, in letteratura sono riportate delezioni 3p ristrette o espanse molto più della regione 3p14.2 che sembrano essere associate con un fenotipo di disabilità intellettiva, ritardo psicomotorio e lievi dismorfismi facciali [26]. Pertanto, in accordo con quanto redatto da ACMG (2015) (**Tabella 16** in Appendice), in associazione alla consultazione della letteratura scientifica, la delezione riscontrata nel paziente 7 è da considerarsi patogenetica e la sua presenza è compatibile con il fenotipo clinico legato al neurosviluppo indicato per il probando. I sintomi aggiuntivi legati alla sindrome di Bardet-Biedl sono probabilmente dovuti a una o più varianti patogenetiche in uno o più geni non inclusi nei pannelli del neurosviluppo.

Paziente 9

Il paziente 9 è un maschio di 8 anni affetto da encefalopatia epilettica sindromica. Il bambino presenta sindrome dismetabolica; ritardo mentale grave; epilessia, con esordio a 5 anni; obesità; fegato steatosico; iperglicemia e iperinsulinemia. I genitori del probando sono sani, non consanguinei e con anamnesi familiari negative. L'analisi bioinformatica per la ricerca delle CNVs eseguita sull'esoma del paziente ha portato all'identificazione di una delezione in eterozigosi coinvolgente gli esoni 31-39 del gene CHD2 (OMIM: 602119) e l'intero gene RGMA (OMIM: 607362), localizzati a livello della banda citogenetica 15q26.1. La successiva analisi mediante *real-time* PCR ha confermato la presenza di una delezione in eterozigosi coinvolgente gli esoni 31-39 del gene CHD2. La validazione in laboratorio è stata effettuata solo sul gene CHD2 in quanto è quello presente all'interno del pannello di geni associati al neurosviluppo (e anche del pannello di geni associati all'epilessia). L'analisi di segregazione eseguita sui genitori del probando ha indicato un'insorgenza *de novo* per la variante in esame.

CHD2 fa parte della famiglia delle proteine CHD, che modulano l'espressione genica agendo come rimodellatori della cromatina e influenzando l'acetilazione istonica. Il rimodellamento della cromatina è uno dei processi critici maggiormente implicati nell'esordio di NDDs se alterato [3]. Queste proteine sono coinvolte in processi come l'attivazione e la repressione genica, la ricombinazione e il riparo del DNA, la regolazione del ciclo cellulare, lo sviluppo e il differenziamento cellulare [27]. L'aploinsufficienza di questo gene è associata a encefalopatia epilettica e dello sviluppo 94 autosomica dominante (OMIM:

615369). Quest'ultima è caratterizzata da ritardo dello sviluppo psicomotorio, diversi tipi di crisi epilettiche, regressione psicomotoria e disabilità intellettiva; tutti i casi riportati sono dovuti a mutazioni *de novo*. Inoltre, Courage et al. (2014) riportano il caso di due pazienti con microdelezioni 15q26.1 coinvolgenti solo i geni CHD2 e RGMA. Entrambi i pazienti presentano un fenotipo distinto con disabilità intellettiva, epilessia, problemi comportamentali, obesità del tronco, scoliosi e dismorfismi facciali relativamente non specifici. Gli stessi autori suggeriscono che l'aploinsufficienza del gene CHD2 sia coinvolta nell'insorgenza della disabilità intellettiva e dell'epilessia, mentre l'aploinsufficienza del gene RGMA sia associata a obesità del tronco con insorgenza durante lo sviluppo puberale.

La delezione riscontrata nel paziente 9 è una variante *loss of function de novo* che comprende più esoni di due geni. L'aploinsufficienza di CHD2 è associata ad una sindrome del neurosviluppo. Inoltre, delezioni di diversa dimensione localizzate nel *locus* genico 15q26.1, comprendenti questo gene e ad insorgenza *de novo*, sono state riscontrate in più individui affetti da ritardo dello sviluppo globale, disabilità intellettiva, epilessia e disturbi del comportamento, con variabilità fenotipica tra di essi [27]. Pertanto, in accordo con quanto redatto da ACMG (2015) (**Tabella 16** in Appendice), in associazione alla consultazione della letteratura scientifica, la delezione riscontrata nel paziente 9 è da considerarsi potenzialmente patogenetica e la sua presenza è compatibile con il fenotipo clinico indicato per il probando.

Paziente 10

Il paziente 10 è una femmina di 36 anni con disabilità intellettiva, ipoacusia neurosensoriale e dismorfismi facciali. Quest'ultimi includono viso longitudinale, bocca ogivale, ugola grande, labbro superiore sottile e *down-slanting* palpebrale. L'analisi bioinformatica per la ricerca delle CNVs eseguita sull'esoma del paziente ha portato all'identificazione di una delezione in eterozigosi, di circa 1.78 Mb, localizzata a livello del *locus* genico 8p23.1 (chr8:8,998,304-10,782,340del). Questa delezione comprende tra gli altri il gene malattia presente in OMIM: RP1L1 (OMIM: 608581). La successiva analisi mediante *array*-CGH ha confermato la presenza di una delezione in eterozigosi, di circa 1.85 Mb, a livello della banda citogenetica 8p23.1 (chr8:8,930,143-10,777,658del). Quest'ultima è circa 70 kb più estesa a livello distale di quella identificata da VarSeq e comprende al suo

interno il gene malattia OMIM già identificato bioinformaticamente. L'analisi di segregazione eseguita sui genitori del probando ha indicato un'insorgenza *de novo* per la variante in esame.

Nello stesso *locus* genico sul cromosoma 8p23.1 sono riportati più riarrangiamenti cromosomici, con un numero elevato di punti di rottura (spesso considerati di significato incerto), identificati in alcuni pazienti con ritardo della crescita, disabilità intellettiva da moderata a lieve, ritardo psicomotorio e/o disordini dello spettro autistico con o senza dismorfismi [14].

Segno clinico caratteristico della sindrome da delezione 8p23.1 è la cardiopatia congenita, tipicamente nella forma di difetto del setto atrioventricolare, difetto del setto atriale e spesso stenosi polmonare [14]. Il gene GATA4 (OMIM: 600576) è deletato nella maggior parte dei pazienti con delezione 8p23.1 e anomalie cardiache. Questo gene codifica un fattore di trascrizione a dita di zinco e la sua aploinsufficienza è stata considerata un potenziale candidato per i difetti cardiaci [14]. Tuttavia, nel probando 10 non si riscontrano difetti cardiaci, caratteristici della sindrome da delezione 8p23.1, e non risulta nemmeno compreso nella delezione il gene GATA4.

La delezione riscontrata nel paziente 7 è una variante *loss of function de novo* di dimensioni elevate che include al suo interno numerosi geni, tra cui un solo gene noto essere associato a malattia. Inoltre, la microdelezione non è riportata in letteratura e nei database dedicati. Pertanto, in accordo con quanto redatto da ACMG (2015) (**Tabella 16** in Appendice), in associazione alla consultazione della letteratura scientifica, la delezione riscontrata nel paziente 10 è da considerarsi di significato incerto (VUS¹³). Tale evento è stato preso comunque in considerazione in quanto si tratta di un evento molto grande e ad insorgenza *de novo*. Dimensione ed ereditarietà sono due aspetti importanti nel valutare la rilevanza clinica di una variante CNV. Infatti, gli eventi più grandi (> 500 kb), rispetto a quelli più piccoli, hanno maggiore probabilità di risultare patogenetici; allo stesso modo, è più probabile che le varianti *de novo* siano patologiche rispetto a quelle ereditate [17]. Inoltre, è altrettanto vero che le delezioni hanno più probabilità di causare un fenotipo negativo [17], e tendono ad essere più penetranti [16], rispetto alle duplicazioni.

¹³ Variant of Uncertain Significance.

Tutti gli eventi identificati bioinformaticamente sono stati poi confermati essere reali attraverso una validazione in laboratorio tramite *real-time* PCR o *array*-CGH. In generale, in particolare per le CNVs più grandi coinvolgenti tanti geni, gli eventi reali sembrano essere più estesi di quelli identificati da VarSeq. Questo potrebbe essere dovuto al fatto che la chiamata delle CNVs effettuata dal programma si basa sull'associare ad ogni evento degli *scores* calcolati a livello probabilistico. Proprio per questo motivo è stato deciso di eseguire l'analisi di *real-time* PCR anche sugli esoni fiancheggiati a quelli identificati a livello bioinformatico come parte dell'evento mutazionale. Il fatto poi che alcuni valori di *Ratio* ottenuti in seguito alle analisi di laboratorio si discostino maggiormente dai valori ideali di riferimento è compatibile col fatto che si tratta di campioni di DNA estratti qualche anno fa, e infatti i valori che si discostano di più sono quelli associati ai campioni più vecchi.

In conclusione, il presente lavoro di tesi ha portato all'identificazione, in un totale di 100 probandi affetti da NDDs, di nove varianti patogenetiche o potenzialmente patogenetiche e di una variante classificata come VUS, anche se fortemente suggestiva e correlata al fenotipo clinico del probando che la porta. Pertanto, la ricerca delle CNVs in questa coorte di 100 pazienti affetti da NDDs ha portato ad una diagnosi molecolare in 10 probandi e quindi ad un aumento del 10% della *detection rate*. Tale percentuale di successo diagnostico è in linea con quella riscontrata in altri studi simili. Per esempio, Gao et al. (2019) hanno dimostrato come l'analisi di 54 casi di NDDs mediante Trio-WES accompagnato da *copy number variation sequencing* (CNVseq) abbia portato un significativo miglioramento del tasso di diagnosi rispetto al solo Trio-WES. Nella loro coorte di pazienti hanno infatti raggiunto un successo diagnostico del 57.4% e, nello specifico, un tasso di diagnosi del 44.4% dovuto al Trio-WES e del 14.8% dovuto al CNVseq. Il presente progetto fornisce quindi un'ulteriore prova a sostegno dell'importanza che rivestono le CNVs nel determinare una causa genetica nel contesto dei NDDs. C'è tuttavia da tener presente che il 90% di pazienti rimasti senza diagnosi rappresenta a sua volta un importante risultato che andrebbe discusso in maniera critica. Innanzitutto va considerato che sono state selezionate solo varianti ad alta significatività, quindi in questi 90 campioni potrebbe esserci una variante che è stata persa durante i vari filtri e/o durante il processo di prioritizzazione. Si tratta anche di campioni di DNA estratti qualche anno fa, pertanto la qualità del sequenziamento potrebbe aver influito sulle chiamate di CNVs fatte da VarSeq, che li compara con campioni di DNA sequenziati più

recentemente. Un'ulteriore spiegazione potrebbe essere che questi pazienti sono portatori di una variante in una regione non codificante del genoma o in un gene non ancora noto essere associato ai NDDs. Inoltre, va tenuto conto del fatto che, sebbene questo progetto si sia concentrato sulle cause monogeniche alla base dell'esordio dei NDDs, una parte di questi disturbi ha una natura poligenica o multifattoriale, quindi potrebbero esserci delle varianti in altre parti del genoma, come dei polimorfismi, o dei fattori non genetici che influenzano il fenotipo patologico dei pazienti con questi disordini.

Infine, le 10 varianti qui analizzate nel dettaglio sono state scelte, per l'appunto, mediante un processo di prioritizzazione, andando a considerare quelle migliori sia dal punto di vista degli *scores* bioinformatici che della possibile patogenicità. Pertanto, per portare avanti/approfondire questo lavoro si potrebbero prendere in considerazione anche altre varianti, magari con *scores* meno buoni, ma associate a fenotipi patologici. Un'ulteriore analisi che si potrebbe fare per migliorare ancora di più la percentuale di successo diagnostico nei pazienti con NDDs, sarebbe analizzare le CNVs in tutti quei pazienti con fenotipi recessivi, per cui con l'analisi delle SNVs è stata riscontrata una sola variante patogenetica e pertanto manca la seconda. Questa tipologia di pazienti non è stata presa in considerazione nel presente lavoro, ma sicuramente rappresenta un importante gruppo per il quale l'analisi delle CNVs potrebbe portare all'identificazione delle cause genetiche della patologia.

Bibliografia

- [1] Soden, S. E. et al. (2014). Effectiveness of exome and genome sequencing guided by acuity of illness for diagnosis of neurodevelopmental disorders. *Science translational medicine*, 6(265), 265ra168.
- [2] Morris-Rosendahl, D. J., & Crocq, M. A. (2020). Neurodevelopmental disorders-the history and future of a diagnostic concept. *Dialogues in clinical neuroscience*, 22(1), 65–72.
- [3] Parenti, I. et al. (2020). Neurodevelopmental Disorders: From Genetics to Functional Pathways. *Trends in neurosciences*, 43(8), 608–621.
- [4] Power, R. A. et al. (2013). Fecundity of patients with schizophrenia, autism, bipolar disorder, depression, anorexia nervosa, or substance abuse vs their unaffected siblings. *JAMA psychiatry*, 70(1), 22–30.
- [5] Pös, O. et al. (2021). DNA copy number variation: Main characteristics, evolutionary significance, and pathological aspects. *Biomedical journal*, 44(5), 548–559.
- [6] Zarrei, M. et al. (2015). A copy number variation map of the human genome. *Nature reviews. Genetics*, 16(3), 172–183.
- [7] Hovhannisyan, G. et al. (2019). DNA Copy Number Variations as Markers of Mutagenic Impact. *International journal of molecular sciences*, 20(19), 4723.
- [8] Gu, W. et al. (2008). Mechanisms for human genomic rearrangements. *PathoGenetics*, 1(1), 4.
- [9] Lintas, C. et al. (2016). Xp22.33p22.12 Duplication in a Patient with Intellectual Disability and Dysmorphic Facial Features. *Molecular syndromology*, 6(5), 236–241.
- [10] Tolchin, D. et al. (2020). De Novo SOX6 Variants Cause a Neurodevelopmental Syndrome Associated with ADHD, Craniosynostosis, and Osteochondromas. *American journal of human genetics*, 106(6), 830–845.
- [11] Morgan, A. et al. (2021). *SETBP1* Haploinsufficiency Disorder. In M. P. Adam (Eds.) et. al., *GeneReviews*®. University of Washington, Seattle.

- [12] Courage, C. et al. (2014). 15q26.1 microdeletion encompassing only CHD2 and RGMA in two adults with moderate intellectual disability, epilepsy and truncal obesity. *European journal of medical genetics*, 57(9), 520–523.
- [13] Parmeggiani, G. et al. (2018). A New 3p14.2 Microdeletion in a Patient with Intellectual Disability and Language Impairment: Case Report and Review of the Literature. *Molecular syndromology*, 9(4), 175–181.
- [14] Ballarati, L. et al. (2011). Genotype-phenotype correlations in a new case of 8p23.1 deletion and review of the literature. *European journal of medical genetics*, 54(1), 55–59.
- [15] Kaminsky, E. B. et al. (2011). An evidence-based approach to establish the functional and clinical significance of copy number variants in intellectual and developmental disabilities. *Genetics in medicine : official journal of the American College of Medical Genetics*, 13(9), 777–784.
- [16] Coe, B. P. et al. (2012). The genetic variability and commonality of neurodevelopmental disease. *American journal of medical genetics. Part C, Seminars in medical genetics*, 160C(2), 118–129.
- [17] Watson, C. T. et al. (2014). The genetics of microdeletion and microduplication syndromes: an update. *Annual review of genomics and human genetics*, 15, 215–244.
- [18] May, T. et al. (2019). Sex differences in neurodevelopmental disorders. *Current opinion in neurology*, 32(4), 622–626.
- [19] Ou, Z. et al. (2008). Microduplications of 22q11.2 are frequently inherited and are associated with variable phenotypes. *Genetics in medicine: official journal of the American College of Medical Genetics*, 10(4), 267–277.
- [20] Wentzel, C. et al. (2008). Clinical variability of the 22q11.2 duplication syndrome. *European journal of medical genetics*, 51(6), 501–510.
- [21] Bartik, L. E. et al. (2022). 22q11.2 duplications: Expanding the clinical presentation. *American journal of medical genetics. Part A*, 188(3), 779–787.
- [22] Phelan, K., & McDermid, H. E. (2012). The 22q13.3 Deletion Syndrome (Phelan-McDermid Syndrome). *Molecular syndromology*, 2(3-5), 186–201.

- [23] Phelan, K. et al. (2005). Phelan-McDermid Syndrome. In M. P. Adam (Eds.) et. al., *GeneReviews*[®]. University of Washington, Seattle.
- [24] Ricciardello, A. et al. (2021). Genotype-phenotype correlation in Phelan-McDermid syndrome: A comprehensive review of chromosome 22q13 deleted genes. *American journal of medical genetics. Part A*, 185(7), 2211–2233.
- [25] Filges, I. et al. (2011). Reduced expression by SETBP1 haploinsufficiency causes developmental and expressive language delay indicating a phenotype distinct from Schinzel-Giedion syndrome. *Journal of medical genetics*, 48(2), 117–122.
- [26] Schwaibold, E. M. et al. (2013). A 3p interstitial deletion in two monozygotic twin brothers and an 18-year-old man: further characterization and review. *American journal of medical genetics. Part A*, 161A(10), 2634–2640.
- [27] Chénier, S. et al. (2014). CHD2 haploinsufficiency is associated with developmental delay, intellectual disability, epilepsy and neurobehavioural problems. *Journal of neurodevelopmental disorders*, 6(1), 9.
- [28] Gao, C. et al. (2019). Diagnostic Yields of Trio-WES Accompanied by CNVseq for Rare Neurodevelopmental Disorders. *Frontiers in genetics*, 10, 485.

Appendice

Tabella 10. Flags "Contenuto estremo in GC", "Basso Z Score" e "Ratio insufficiente".

| | |
|--------------------------------|--|
| Contenuto estremo in GC | <25% o >75% |
| Basso Z Score | <p>Delezione in omozigosi: <i>Target Count</i>* ≤ 7 e <i>Z Score</i> > -0.65</p> <p>Delezione in eterozigosi:</p> <ul style="list-style-type: none"> • <i>Target Count</i>* = 1 e <i>Z Score</i> > -2.5 • <i>Target Count</i>* = 2 e <i>Z Score</i> > -1.4 • <i>Target Count</i>* < 7 e <i>Z Score</i> > -1.2 <p>Duplicazione:</p> <ul style="list-style-type: none"> • <i>Target Count</i>* = 1 e <i>Z Score</i> < 2.5 • <i>Target Count</i>* = 2 e <i>Z Score</i> < 2.0 • <i>Target Count</i>* < 7 e <i>Z Score</i> < 1.25 |
| Ratio insufficiente | <p>Delezione in eterozigosi e delezione in omozigosi:</p> <ul style="list-style-type: none"> • <i>Target Count</i>* ≤ 2 e <i>Ratio</i> > 0.6 • <i>Target Count</i>* < 8 e <i>Ratio</i> > 0.65 <p>Duplicazione:</p> <ul style="list-style-type: none"> • <i>Target Count</i>* = 1 e <i>Ratio</i> < 1.42 • <i>Target Count</i>* ≤ 3 e <i>Ratio</i> < 1.35 • <i>Target Count</i>* < 8 e <i>Ratio</i> < 1.28 |

**Target Count* = numero di esoni coinvolti nell'evento.

Tabella 11. Categorie di "Overlap Type".

| | |
|------------------------|---|
| Contains Region | La CNV* contiene interamente la regione** |
| Within Region | La CNV* è contenuta interamente dalla regione** |
| Identical | La CNV* è identica alla regione** |
| Partial Overlap | La CNV* si sovrappone parzialmente alla regione** |

*del campione d'interesse; **annotata nel database.

Tabella 12. Coppie di primers utilizzate per la real-time PCR per i pazienti 3, 5 e 9.

| Gene ed esone | Coppie di primers | Dimensione dell'amplificato |
|---------------------------------|---|------------------------------------|
| Gene SOX6 esone 1 | F: CCAGCACCCAGAAATCACTG R: GGGGACAACCTGCATGAGTTA | 170 bp |
| Gene SOX6 esone 2 | F: CTGTCCCAGTCAGCATCTTGT R: GAGGATGCAATGACCCAGGA | 164 bp |
| Gene SOX6 esone 3 | F: GAGCCCAACAGTGAAGTCCA R: GGAGTCGGACCGTGAGATA | 159 bp |
| Gene SOX6 esone 4 ¹⁴ | F: TTTCTCCAAGAAGTTCAGTGGT R: TCCTCCTGCATGGAAAACT | 82 bp |
| Gene SOX6 esone 5 | F: TCCAGTTCGTCAGTGCCTTT R: GGCAGCTCTCCACCATGATT | 175 bp |
| Gene SETBP1 esone 4 | F: ACGACCAAAGAAGCAGCCTT R: CGCTCATGGGTTTGCCTCT | 170 bp |
| Gene SETBP1 esone 5 | F: ATGACCAGGAAGAAGCCAGC R: AACAGGCCATTCTCAGTGGG | 173 bp |
| Gene SETBP1 esone 6 | F: CACCAAGATCCTGTCCACCA R: CGGCATTTGGTCTCTGGAGT | 169 bp |
| Gene CHD2 esone 30 | F: TGCTTAGATGGTCGTACCGT R: GTGCAGCTTTGACACGACAG | 169 bp |
| Gene CHD2 esone 31 | F: GACCCGAGCGGATTACTTGT R: GGGCCTCTCTTCTACCT | 155 bp |
| Gene CHD2 esone 39 | F: TACCACCAGCAGGATTTCCG R: GAGCGAGGATCTGAGACTGC | 161 bp |
| HK (introne 13) | F: TACGTGCAGGGTTACTGCTG R: TTCCGTGTCAAGCTCTGTTG | 160 bp |
| HK (esone 25) | F: ATTCCTCAGTACCCAGGCT R: GGAGTGCACCCTGCCTATTG | 80 bp |

¹⁴ Solo per l'esone 4 del gene SOX6 è stato utilizzato l'HK da 80 bp, in quanto il disegno della coppia di primers per amplificare in modo specifico quest'esone ha fatto emergere alcune problematiche (presenza di 4 basi consecutive identiche nel primer R, lunghezza di 22 bp del primer F e penalty complessiva superiore a 1), che risultavano amplificate con l'utilizzo dell'HK da 160 bp.

Tabella 13. Programma termico utilizzato per la real-time PCR.

| Stage | Step | Temperatura | Tempo |
|---------------------------------|---------------|--------------------|--------------|
| <i>Holding Stage</i> | <i>Step 1</i> | 95°C | 5 m |
| <i>Cycling Stage (40 cicli)</i> | <i>Step 1</i> | 95°C | 10 s |
| | <i>Step 2</i> | 60°C | 30 s |
| <i>Melt Curve Stage</i> | <i>Step 1</i> | 95°C | 15 s |
| | <i>Step 2</i> | 60°C | 1 m |
| | <i>Step 3</i> | 95°C | 15 s |

Tabella 14. Kit e reagenti utilizzati per l'array-CGH.

| Step | Kit/reagente | Contenuto |
|--------------------|---|--|
| Marcatura | SureTag Complete DNA Labeling Kit | Reagenti per la marcatura, colonnine per la purificazione, DNA maschile e femminile di riferimento |
| Ibridazione | Oligo aCGH/ChIP-on-chip Hybridization Kit | Reagenti per l'ibridazione |
| Ibridazione | Human Cot-I DNA | Reagente per bloccare l'ibridazione aspecifica |
| Lavaggi microarray | Oligo aCGH/ChIP-on-chip Wash Buffer Kit | Reagenti per i lavaggi del vetrino |

Tabella 15. Programmi termici utilizzati per l'array-CGH.

| Fase | Stage | Temperatura | Tempo |
|------------------------|----------------|--------------------|--------------|
| Marcatura, sottofase 1 | <i>Stage 1</i> | 98°C | 10 m |
| | <i>Stage 2</i> | 4°C | 5 m |
| Marcatura, sottofase 2 | <i>Stage 1</i> | 37°C | 2 h |
| | <i>Stage 2</i> | 65°C | 10 m |
| | <i>Stage 3</i> | 4°C | ∞ |
| Ibridazione | <i>Stage 1</i> | 98°C | 3 m |
| | <i>Stage 2</i> | 37°C | 30 m |

Tabella 16. Criteri di patogenicità adottati per l'interpretazione delle varianti genetiche.

| Classificazione | Criteri di patogenicità secondo ACMG | | | + Analisi della letteratura | |
|-----------------------------|--------------------------------------|--------------------------|----------------------|-----------------------------|--|
| Patogenetica | A | + almeno una tra B C D | - | Descritta in letteratura | Non descritta in letteratura = Potenzialmente patogenetica |
| | A | + almeno due tra F G H I | - | | |
| | A | + una tra F G H I | + una tra J K L | | |
| | A | + almeno due tra J K L | - | | |
| | Almeno 2 tra B C D E | - | - | | |
| | Una tra B C D E | + almeno 3 tra F G H I | - | | |
| | Una tra B C D E | + almeno 2 tra F G H I | + almeno 2 tra J K L | | |
| | Una tra B C D E | + una tra F G H I | + tutte le J K L | | |
| Potenzialmente patogenetica | A | + una tra F G H I | - | | |
| | Una tra B C D E | + almeno una tra F G H I | - | | |
| | Una tra B C D E | + almeno due tra J K L | - | | |
| | Almeno 3 tra F G H I | - | - | | |
| | Due tra F G H I | + almeno due tra J K L | - | | |
| | Una tra F G H I | + tutte le J K L | - | | |

Criteri di patogenicità:

- A: Varianti *loss of function*;
- B: Variante localizzata nello stesso codone e risultante nella stessa sostituzione amminoacidica nota essere patogenetica;
- C: Variante *de novo* confermata sui genitori del probando;
- D: Variante per la quale sono stati effettuati studi *in vitro* funzionali che dimostrano l'effetto dannoso sulla proteina;
- E: Varianti per le quali è nota una prevalenza, in termini di frequenza, negli individui affetti;
- F: Variante localizzata in un *hot spot* o in una zona funzionale proteica caratterizzata biochimicamente;
- G: Variante assente nel database di popolazione o presente con frequenza compatibile o minore all'incidenza del fenotipo associato;

- H: Per i fenotipi recessivi, variante localizzata *in trans* ad una seconda variante nello stesso gene;
- I: Variante localizzata in uno stesso codone nel quale è descritta una variante patogenetica;
- J: Variante che co-segrega con il fenotipo patologico;
- K: Variante missenso in un gene altamente conservato;
- L: Varianti per le quali l'analisi computazionale *in silico* indica con elevata probabilità un effetto danno sulla struttura o attività proteica.

Variante benigna e potenzialmente benigna

Variante per la quale sia soddisfatta una delle seguenti condizioni:

- Ha una frequenza allelica più alta della frequenza della patologia nella popolazione di riferimento;
- È riportata in letteratura come non associata al fenotipo patologico;
- È riscontrata *in cis* con una variante patogenetica (nel caso di fenotipi recessivi);
- È riscontrata in un familiare non affetto (nei casi di fenotipi dominanti in assenza di penetranza incompleta).

Variante di significato incerto

Variante per la quale siano soddisfatte una o più delle seguenti condizioni:

- Non è classificabile come patogenetica, potenzialmente patogenetica, potenzialmente benigna o benigna;
- È una variante riportata in un unico individuo affetto in mancanza di studi di segregazione ed informazioni cliniche soddisfacenti;
- È una variante per la quale siano riportate evidenze contrastanti;
- È una variante localizzata in una posizione limitrofa al sito di *splicing* canonico, qualora nei database dedicati siano descritte analoghe varianti.