

ALMA MATER STUDIORUM
UNIVERSITA' DEGLI STUDI DI BOLOGNA

Ph.D. in Biologia Cellulare, Molecolare ed Industriale.
Progetto n. 1: Fisiologia e Biologia Cellulare.
XXIV Ciclo
Codice area scientifica Bio/18
Settore Concorsuale 05/I1

Analisi di “*Copy Number Variants*” ed
identificazione di nuovi geni candidati per
l’Autismo e Ritardo Mentale

candidato: **Minopoli Fiorella**

Coordinatore del progetto:

Prof.ssa Michela Rugolo

Supervisore:

Prof. ssa Elena Maestrini

Esame Finale-Aprile 2012, Bologna

ABSTRACT	5
1 Anomalie Cromosomiche	7
1.1 <i>Anomalie cromosomiche e malattie umane</i>	9
1.2 <i>Array-CGH e analisi delle anomalie cromosomiche</i>	10
1.2 <i>CNV: copy number variants</i>	12
1.3.1 <i>Formazione delle CNVs</i>	15
1.3.2 <i>Metodi per l'identificazione delle CNVs</i>	20
1.3.3 <i>Associazione tra le CNVs e le malattie</i>	23
2. AUTISMO	29
2.1 <i>Aspetti clinici</i>	30
2.2 <i>Anomalie del cervello</i>	32
2.3 <i>Anomalie biochimiche</i>	33
2.4 <i>Basi genetiche dell'autismo</i>	33
2.4.1 <i>Studi di linkage nell'autismo</i>	37
2.4.2 <i>Screening di mutazioni in geni candidati per l'autismo</i>	40
2.4.3 <i>Studi di associazione nell'autismo</i>	40
2.4.4 <i>Studi di associazione in geni candidati per l'autismo.</i>	43
2.4.5 <i>Studi di associazione genome-wide</i>	44
2.4.6 <i>Studi citogenetici</i>	46
2.5 <i>CNVs ed Autismo</i>	47
2.6 <i>Caratteri monogenici e varianti rare associati all'autismo</i>	52
2.7 <i>Architettura genetica dei disturbi dello spettro autistico</i>	56
3. il Ritardo Mentale	57
3.1 <i>Eziologia</i>	58
3.2 <i>Cause genetiche</i>	60
3.2.1 <i>Aberrazioni cromosomiche</i>	60
3.2.2 <i>CNV E MR</i>	61
3.2.3 <i>Forme di ID del cromosoma X</i>	64
3.2.4 <i>Forme autosomiche di MR</i>	66
3.2.5 <i>Alterata regolazione di geni soggetti ad imprinting</i>	68
3.3 <i>Funzioni biologiche e molecolari dei geni coinvolti nel ritardo mentale.</i>	69

3.3 Funzioni biologiche e molecolari dei geni coinvolti nel ritardo mentale.	69
3.3.2 Funzione e struttura sinaptica e ritardo mentale	72
4. Scopo della tesi	77
5. Materiali e Metodi	79
5.1 Campioni di DNA	79
5.2 Estrazione del DNA da sangue intero con kit Qiagen MAXI (per 5 ml di sangue)	79
5.3 Identificazione di CNVs tramite la tecnologia ILLUMINA	80
5.4 Validazione dei CNV tramite Real Time	83
5.5 Analisi di varianti strutturali mediante QMPSF della regione IMMP2L-DOCK4	88
5.6 Analisi di varianti strutturali mediante QMPSF della regione DOCK4	89
5.7 analisi semiquantitativa dell'espressione del trascritto di fusione IMMP2L-DOCK4 mediante RT-PCR.....	90
5.7 Screening di mutazioni dei geni CNTNAP5 e CADPS2.....	90
5.7a Disegno dei primers	90
5.7b Amplificazione tramite Polimerasi Chain Reaction (PCR)	90
5.7c Messa a punto del protocollo di PCR per l'amplificazione degli esoni del gene	92
5.7d Elettroforesi su gel di agarosio.....	92
5.7e Reazione di sequenziamento e analisi delle sequenze.....	92
5.8 Genotipizzazione di SNPs mediante digestione con enzimi di restrizione	94
5.9 Array Comparative Genomic Hybridization	94
5.9.1 Step 1. Digestione del DNA	95
5.9.2 Step 2. Marcatura del DNA	96
5.9.3 Step 3. Purificazione del DNA marcato	96
5.9.4 Step 4. Ibridazione	97
5.9.5 Step 5. Lavaggi post-ibridazione	97
5.9.6 Step 6 e 7. Scansione degli array e analisi dei risultati.....	97
6. RISULTATI:	102
Analisi dei geni IMMP2L, DOCK4 e CNTNAP5 in autismo e dislessia.....	102
6.1. Risultati Preliminari.....	102

6.1a: Analisi di associazione nei loci e AUTS1 ed AUTS5	102
6.1b: Analisi di Copy Number Variant (CNV) nei loci AUTS1 ed AUTS5 tramite il software QuantiSNP	104
6.2: Analisi di CNV nei geni IMMP2L e DOCK4 tramite QMPSF in individui con autismo	107
6.3: Caratterizzazione della delezione IMMP2L/DOCK4 nella famiglia 15-0084.....	111
6.3a. Definizione del break point della delezione IMMP2L/DOCK4.....	111
6.3b. Trascritto di fusione.....	113
6.3c. Longe range PCR su famiglia estesa della madre	115
6.4. Screening di DOCK4 su individui dislessici	116
6.5. Identificazione e caratterizzazione di una delezione in CNTNAP5	119
6.6. Screening di mutazione del gene CNTNAP5.....	122
7 RISULTATI: Analisi di CNV in ritardo mentale	127
7.1. CHERISH	127
7.2. Array CGH	129
7.3. Screening di mutazioni nel gene CADPS2 in individui affetti da ritardo mentale e/o autismo	134
7.4. Analisi dell'imprinting di CADPS2.....	139
8. DISCUSSIONE	143
8.1. Caratterizzazione di due CNVs identificate nella stessa famiglia	143
8.2. Copy Number Variants nel ritardo mentale.....	147
9.BIBLIOGRAFIA	153

ABSTRACT

Autism spectrum disorders (ASD) and intellectual disability (ID) are characterized by a complex and heterogeneous genetic etiology. Recent developments in genomic research have enabled the discovery of numerous copy number variants (CNVs) in the pathogenesis of these disorders, although their etiology remains unknown in the majority of cases. This work concerns the identification and characterization of specific CNVs in families with ASD and ID. I studied a microdeletion in 7q31 encompassing the two genes *DOCK4* and *IMMP2L*, transmitted from the mother (who has dyslexia) to two children with autism and to a daughter with dyslexia. In the same family we identified a second microdeletion in 2q14, that inactivates *CNTNAP5*, and is transmitted by the father (with ASD) to the two children with autism. We therefore hypothesized that *DOCK4* and *CNTNAP5* could be implicated in susceptibility to dyslexia and ASD, respectively. Screening of numerous affected individuals supported our hypothesis, leading to the identification of a new *DOCK4* microdeletion segregating with dyslexia, and 3 new missense variants in *CNTNAP5* in individuals with autism. Through array comparative genomic hybridization (aCGH) of individuals with ID, we also identified a 7q31.32 microdeletion involving the *CADPS2* gene in two brothers with ID and autistic features, probably inherited from the mother. Screening for mutations in this gene in individuals with autism or ID, has led to the identification of 3 maternally inherited nonsynonymous variants, absent in controls. Since *CADPS2* is located in a genomic region containing imprinted loci, we hypothesized that *CADPS2* itself could be subjected to imprinting, with maternal monoallelic expression. Expression analysis of *CADPS2* in blood cells supported this hypothesis, therefore suggesting *CADPS2* as a new susceptibility gene for ID and ASD, and as possible new imprinted gene .

I disturbi dello spettro autistico (DSA) ed il ritardo mentale (RM) sono caratterizzati da un'etiologia genetica complessa ed eterogenea. Grazie ai recenti sviluppi nella ricerca genomica, è stato possibile dimostrare il ruolo di numerose *copy number*

variants (CNVs) nella patogenesi di questi disturbi, anche se nella maggior parte dei casi l'eziologia rimane ancora sconosciuta.

Questo lavoro riguarda l'identificazione e la caratterizzazione di alcuni CNVs in famiglie con DSA e RM.

E' stata studiata una microdelezione in 7q31 che coinvolge i geni *IMMP2L* e *DOCK4*, trasmessa dalla madre con dislessia a due figli con autismo ed una figlia con dislessia. Nella stessa famiglia segrega una seconda microdelezione in 2q14 che inattiva il gene *CNTNAP5* ed è trasmessa dal padre (con tratti autistici) ai soli due figli con autismo. Abbiamo quindi ipotizzato che i geni *DOCK4* e *CNTNAP5* potessero essere implicati, rispettivamente, nella suscettibilità a dislessia e DSA. Lo screening di numerosi individui affetti ha supportato la nostra ipotesi, con l'identificazione di una nuova microdelezione di *DOCK4* che segrega con la dislessia, e 3 nuove varianti missenso in *CNTNAP5* in individui con autismo.

Dall'analisi genomica comparativa su array (aCGH) di individui con RM, è stata inoltre identificata una delezione nella regione 7q31.32, che coinvolge il gene *CADPS2*, in due fratelli con RM e tratti autistici, probabilmente ereditata dalla madre. Lo screening di mutazione di questo gene in individui con autismo o RM, ha portato all'identificazione di 3 varianti non sinonime, assenti nei controlli, ed ereditate per via materna. Poiché *CADPS2* risiede in una regione genomica che contiene loci soggetti ad imprinting, abbiamo ipotizzato che il gene *CADPS2* possa essere anch'esso caratterizzato da imprinting, con espressione monoallelica materna. Lo studio di espressione di *CADPS2* in cellule del sangue ha avvalorato questa ipotesi, implicando perciò *CADPS2* come un nuovo gene di suscettibilità per il RM e DSA.

1 Anomalie Cromosomiche

Da quando è stato accertato l'esatto numero dei cromosomi nelle cellule umane, nel 1956, sono stati sviluppati numerosi metodi per identificare le anomalie cromosomiche. Le anomalie cromosomiche sono alterazioni del numero o della struttura dei cromosomi che si verificano all'incirca nel 7,5% dei concepimenti. Nella maggior parte dei casi però esse sono incompatibili con la vita e portano ad un aborto spontaneo precoce, ne consegue che solo lo 0,6% dei nati vivi è affetto da malattie cromosomiche. Le anomalie numeriche dei cromosomi o *aneuploidie cromosomiche* vengono distinte in *monosomie*, quando è presente una sola copia del cromosoma anziché due, e *poliploidia* quando sono presenti più copie di uno stesso cromosoma, tra queste ultime le più diffuse sono le *trisomie* ossia la presenza di tre copie di uno stesso cromosoma.

I riarrangiamenti strutturali possono essere **bilanciati** o **sbilanciati**, se comportano o meno una perdita o un guadagno di materiale genetico, e possono essere distinti in sei categorie:

1. Le **delezioni** consistono nella perdita più o meno estesa di un frammento di cromosoma e vengono classificate come *terminali*, se avvengono nella parte finale del cromosoma (nei telomeri), e *interstiziali*, quando, in seguito alla rottura di un cromosoma in due punti, avviene la perdita del frammento intermedio e i punti di rottura vengono saldati tra di loro.
2. Le **uplicazioni** consistono nella presenza di due copie di uno stesso frammento nello stesso cromosoma; si dicono in *tandem* quando il segmento è ripetuto nello stesso orientamento, *inverse* quando la duplicazione (il frammento duplicato) ha una direzione opposta (al frammento originario).
3. Le **inversioni** consistono nella rotazione di 180° di un frammento. Esse non comportano perdite di materiale genetico ma possono determinare la perdita di funzionalità di geni importanti se i punti di rottura da cui hanno origine ricadono all'interno della struttura di un gene o nelle regioni regolatrici della trascrizione. Le inversioni possono essere di due tipi: *pericentriche*, se il segmento che subisce la rotazione include il centromero, e *paracentriche*, se non lo comprende. Quindi le inversioni producono un nuovo allineamento dei geni lungo l'asse di un cromosoma.

4. Le **traslocazioni** implicano un trasferimento di materiale tra due o più cromosomi diversi. Possono essere di due tipi: *reciproche* o *Robertsoniane*.

La *traslocazione reciproca* consiste in uno scambio bidirezionale di materiale genetico tra due cromosomi non omologhi e avviene sempre in seguito a rotture del DNA, in seguito alle quali si formano dei frammenti di cromosoma liberi che per errori nel processo di riparazione, sono saldati in maniera errata ad un altro cromosoma non omologo, dando così origine, a due cromosomi derivati. In questo caso si tratta di traslocazioni bilanciate.

Nella *traslocazione Robertsoniana*, invece, due cromosomi acrocentrici non omologhi si rompono a livello dei centromeri e i bracci lunghi si ritrovano attaccati a un unico centromero. I bracci corti si uniscono a loro volta a formare il prodotto reciproco, che generalmente contiene geni non essenziali e viene abitualmente perduto entro poche divisioni cellulari. Quindi i cromosomi che si originano da questo scambio sono uno dotato di uno o due centromeri ed uno acentrico (cioè senza centromero). Quest'ultimo è instabile ed è perso alla prima divisione cellulare.

5. I **cromosomi ad anello**, detti anche ring, sono originati dalla rottura di entrambe le braccia di un cromosoma, perdita delle regioni distali alle rotture e riunione delle due estremità in una struttura ad anello. I ring possono essere soprannumerari, ed in tal caso il portatore avrà 47 cromosomi e sarà trisomico per le regioni comprese nel ring, oppure possono aver sostituito un cromosoma normale, ed in tal caso il portatore avrà 46 cromosomi ma sarà parzialmente monosomico, per la perdita delle regioni distali alle rotture.
6. I **Markers o cromosomi marcatori**, sono anomalie cromosomiche particolari caratterizzate dalla presenza di piccole porzioni cromosomiche soprannumerarie di cui non si conosce l'origine, e cioè da quali cromosomi queste porzioni derivino.

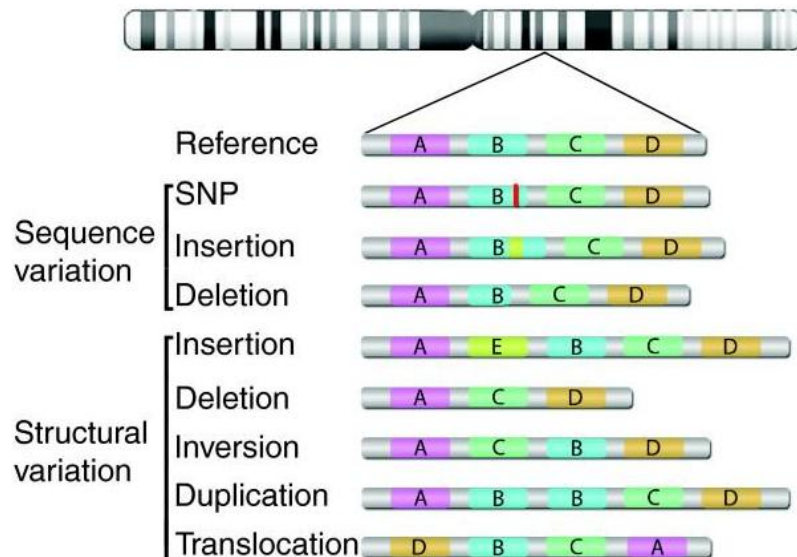


Figura 1: tipi di variazioni genetiche

1.1 Anomalie cromosomiche e malattie umane

La prima scoperta che ha messo in relazione gli sbilanciamenti genomici e le malattie umane risale al 1959, quando è stato dimostrato che la trisomia del cromosoma 21 era responsabile della sindrome di Down (Lejeune et al.1959).

Subito dopo, sono state descritte altre sindromi cliniche basate sull'identificazione delle stesse anomalie citogenetiche presenti in molti pazienti che mostravano lo stesso fenotipo clinico, come la trisomia del cromosoma 13 nella sindrome di Patau, e la trisomia del cromosoma 18 nella sindrome di Edwards.

L'inizio della citogenetica del cancro, nel 1960, è stata segnata dall'identificazione del cromosoma Philadelphia, causato da una traslocazione tra il cromosoma 9 ed il 22, e dalla sua associazione con la leucemia mieloide cronica.

In seguito all'invenzione delle tecniche di bandeggio del cromosoma, nel 1970, sono state identificate numerose aberrazioni cromosomiche strutturali associate a malattie umane.

Fu presto chiaro che la perdita o l'acquisizione di materiale genetico poteva essere associato a ritardo mentale e anomalie congenite multiple.

L'avvento dell'*array-CGH* descritta per la prima volta nel 1997, ha aperto la strada alla possibilità di una risoluzione più ampia nell'identificazione delle anomalie cromosomiche strutturali.

1.2 Array-CGH e analisi delle anomalie cromosomiche

La citogenetica tradizionale, pur utilissima nell'identificare un gran numero di anomalie cromosomiche, numeriche e strutturali, sia bilanciate che non bilanciate, è limitata nelle sue possibilità diagnostiche dal potere di risoluzione del microscopio, per cui la massima risoluzione che si può raggiungere con le tecniche di bandeggio dei cromosomi è di 3- 5 Mb. (Bruce R. K. 2000, 2001).

Lo sviluppo della tecnologia di ibridazione fluorescente in situ (***Fluorescence In Situ Hybridization -FISH***) ha sicuramente costituito un passo in avanti nello *screening* dei riarrangiamenti genomici ed ha portato alla nascita della citogenetica molecolare. Nell'analisi *FISH* le sonde di DNA, marcate con fluorofori, sono ibridate a preparazioni di cromosomi di cellule in interfase o metafase per determinare la presenza, la localizzazione e il numero di specifici segmenti genomici (Fig.1.2). Questa tecnologia, non solo permette di trovare piccole alterazioni genomiche, da 50 Kb a 100 Kb, ma permette anche la diretta visualizzazione di queste alterazioni nelle cellule, senza che sia necessario indurne la crescita in vitro.

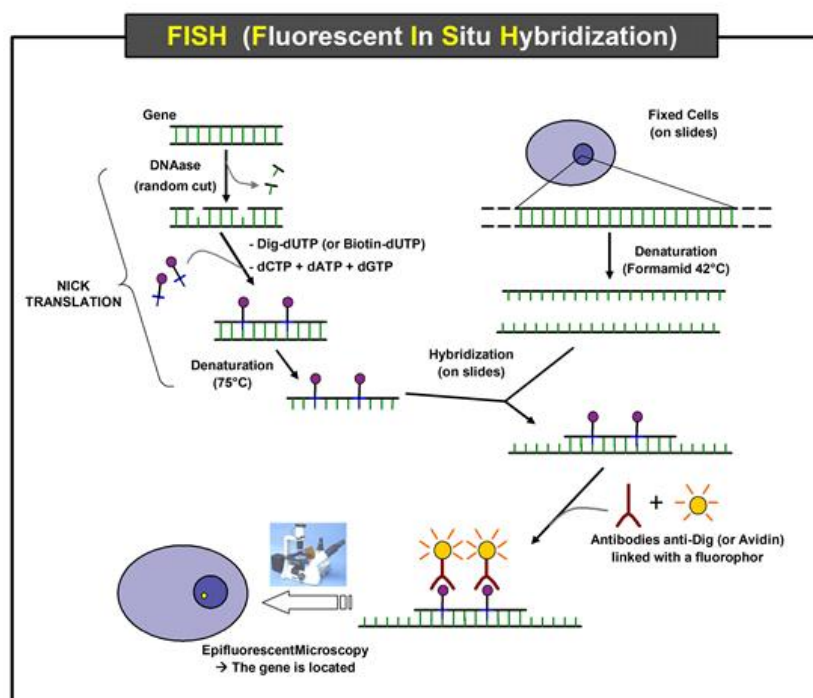


Figura 1.2: ibridazione fluorescente in situ

Queste caratteristiche hanno reso il test della *FISH* ideale nell'identificazione non solo delle sindromi da microdelezioni/microduplicazioni, ma anche per l'analisi di aneuploidie prenatali e per gli studi di genetica del cancro.

Sebbene la *FISH* permette la rilevazione di sbilanciamenti cromosomici con grande accuratezza, essa ha il grosso limite di rilevare solo particolari mutazioni a livello di precisi loci cromosomici, poiché può sondare solo sequenze specifiche che sono note o sospettate essere associate a sindromi note e non fornisce un'analisi dell'intero genoma. Questo limite è stato superato dall'“*Ibridazione Genomica Comparata*” (***Comparative Genomic Hybridization-CGH***) che permette di identificare la presenza delle eventuali anomalie cromosomiche a livello dell'intero genoma, senza sapere in anticipo cosa cercare.

La *CGH* è una tecnica che inizialmente è stata applicata allo studio di cellule tumorali ma che si è poi rivelata utile anche nella diagnosi delle aberrazioni cromosomiche costituzionali (Lapierre et al. 1998).

La metodica si basa su un'ibridazione in situ modificata, che sfrutta la competizione tra due campioni di DNA genomico marcati con fluorocromi diversi: un campione è quello del paziente, l'altro è costituito da un pool di DNA genomico di riferimento. Nella tecnica convenzionale i DNA marcati sono ibridati simultaneamente su preparati metafasici normali su una superficie di vetro (Fig.1.3).

Il vantaggio di tale tecnica è che in un'unica ibridazione si possono ottenere informazioni sulla dimensione e sulla localizzazione di tutti gli sbilanciamenti cromosomici, mentre il suo limite è il basso potere di risoluzione (3-10 Mb). (Kirchhoff, M. et al. 1999)

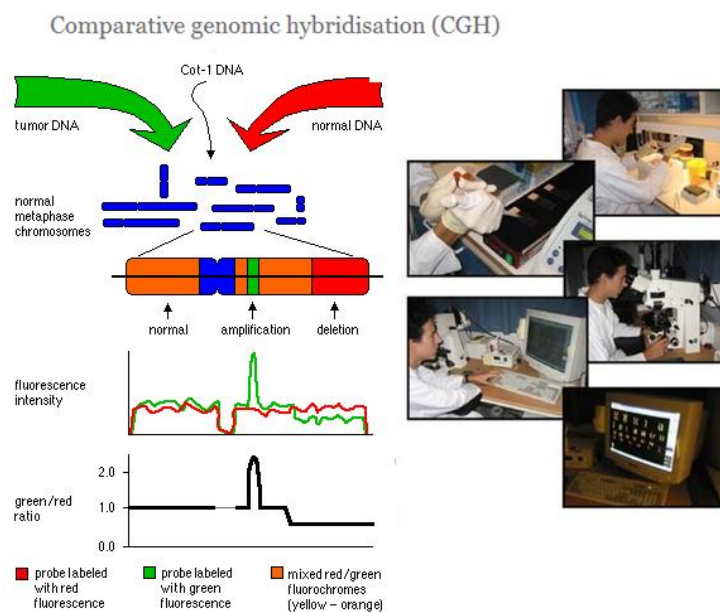


Figura 1.3: Ibridazione Genomica Comparata

L'avvento della tecnologia basata sui microarray, sviluppata nell'ultimo decennio, permette di esaminare tutto il genoma umano su un singolo chip con una risoluzione alta (fino ad un centinaio di coppie di basi), un processo anche noto come citogenetica basata sui microarray.

Nell'*array-CGH* i cromosomi in metafase sono stati sostituiti con sonde di DNA bersaglio roboticamente immobilizzati e distribuiti su un microscopico vetrino (chip) e questo ha aumentato significativamente la risoluzione e semplificato la procedura dell'analisi.

Le sonde possono essere pezzi di DNA genomico sotto forma di cromosomi batterici artificiali (BACs di 150-160 Kb) o cloni P1 (PAC) (dimensione di 75-200Kb), cloni inserto più piccoli come i cosmidi (dimensione 30-40 Kb) e fosmidi (dimensione 40-50 Kb), o oligonucleotidi (25-85 mers) e corrispondono a loci specifici di ogni singolo cromosoma, fino a comprendere tutto il genoma umano.

La risoluzione genomica delle differenti piattaforme *aCGH* è determinata dallo spazio e dalla lunghezza delle sonde di DNA. Questa risoluzione è grande almeno dieci volte quella della migliore analisi dei cromosomi in pro metafase, quindi rappresenta lo screening per delezioni o inserzioni più sensibile del genoma intero. Tuttavia l'*aCGH* non è in grado di rilevare anomalie cromosomiche bilanciate come traslocazioni ed inversioni. Perciò le tecniche citogenetiche convenzionali saranno sempre fondamentali e non potranno essere sostituite in toto, ma indubbiamente l'*array-CGH* apre le porte verso analisi sempre più accurate e precise a fronte di una maggiore semplicità e tempi di risposta molto più veloci.

1.2 CNV: copy number variants

Il Progetto Genoma Umano (*Human Genome Project*) ha fornito una dettagliata mappa fisica del genoma umano permettendo, così, lo sviluppo della tecnologia basata sugli array che ha rivoluzionato il campo della genetica, in quanto ha reso possibile ottenere in un singolo esperimento i genotipi di centinaia di migliaia di marcatori distribuiti in tutto il genoma. L'avvento delle moderne tecnologie di microarray ha anche permesso l'identificazione di piccole varianti strutturali dette *copy numbers variants* (CNVs).

Le CNVs possono essere delezioni o duplicazioni che portano ad un cambiamento del numero di copie di una specifica regione cromosomica. Le CNVs possono avere una dimensione che varia da una kilobase (kb) a diverse megabasi (Mb). Queste varianti strutturali quantitative, spesso, coinvolgono uno o più geni e comprendono, complessivamente, più nucleotidi per genoma rispetto al numero totale di polimorfismi a singolo nucleotide (SNPs) (Redon, R. *et al* 2006).

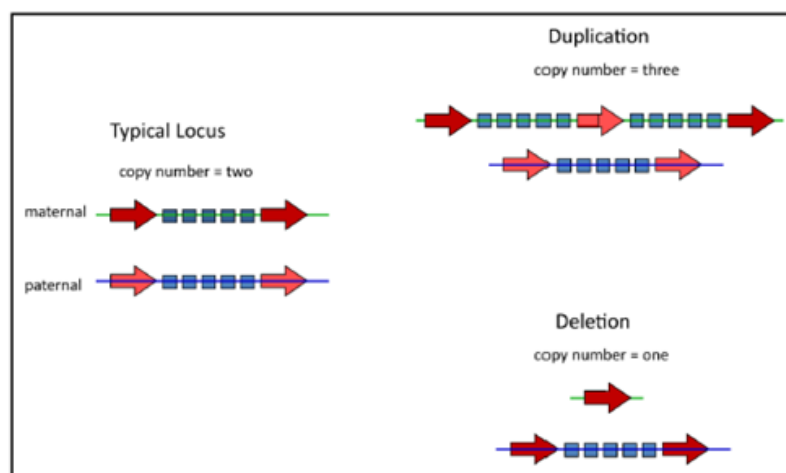


Figura 1.4: Copy Number Variants. In una situazione normale il numero di copie di un locus è due poiché una copia è ereditata dalla madre ed una dal padre. Alcuni loci possono subire duplicazioni o delezioni e possono variare in dimensione, coinvolgendo da nessuno a dozzine di geni, rappresentati dai rettangoli. Queste regioni del genoma variabili nel numero di copie sono spesso fiancheggiate da LCR o duplicazioni segmentali, rappresentate dalle frecce (Eric M. Morrow, Ph.D., M.D. 2010)

Quando una *CNV* è presente con una frequenza maggiore dell'1% si parla di *polimorfismo del numero di copie* (*Copy Number Polymorphism, CNP*).

Sebbene queste varianti strutturali in alcune regioni del genoma non hanno alcuna conseguenza fenotipica [Sharp et al. 2005; Sebat et al., 2004; Iafrate et al., 2004; Tuzun et al., 2005], altre influenzano il dosaggio di un gene, e di conseguenza possono causare malattie genetiche [Inoue et al., 2002].

Si stima che circa il 12% del genoma sia coperto da CNVs (Carter N.P. 2007; Reichenberg A, Weiser M, Rabinowitz J et al.2002; Redon et al. 2006) e più del 41% di tutte le CNVs identificate si sovrappongono a geni noti.

Inoltre, la presenza di un numero inaspettatamente grande di varianti strutturali, di dimensioni intermedie (1Kb-3Mb), anche nel genoma di individui sani ha sottolineato che le CNVs costituiscono una grande fonte di variabilità fenotipica nell'uomo e quindi molto probabilmente contribuiscono in modo preponderante alla diversità umana: si stima che due genomi scelti casualmente differiscano almeno dell'1% e che

la maggior parte di queste differenze siano dovute alle *CNVs* [Kallman FJ 2008; Conrad, 2006; Hinds, 2006; McCarroll 2006; Pinto D et al. 2007]. Quando presenti in emizigosi in individui sani, questi sbilanciamenti sono considerati “benigni” e cioè che non hanno un grande effetto fenotipico sullo sviluppo umano, ma molte *CNVs* sono state riconosciute anche come una delle cause più comuni delle malattie umane. Le *CNVs* possono essere distinte in *ereditarie o de novo* a seconda che esse siano o meno state trasmesse da almeno uno dei genitori (Mc Carroll, S.A et al. 2008).

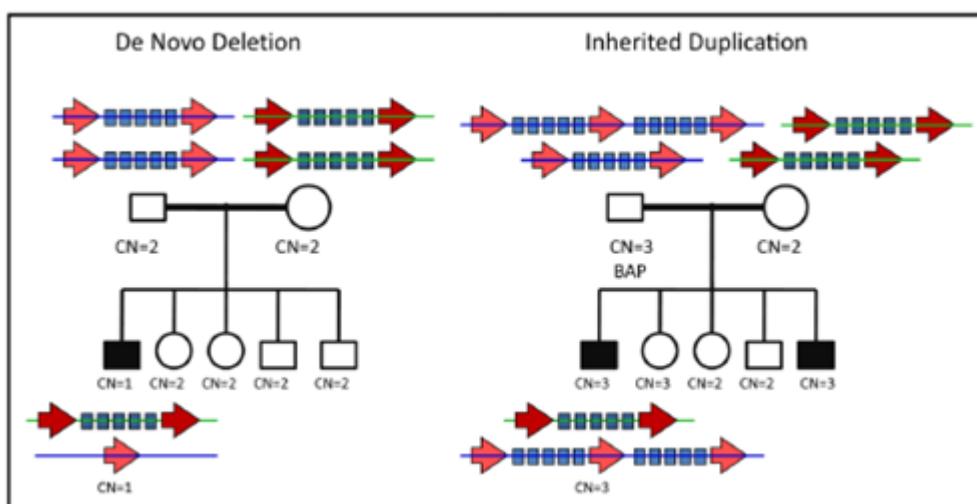


Figura 1.5: le Copy Number Variants possono essere ereditate o de novo, associate a malattie (altamente penetranti o con penetranza incompleta), benigne o di sconosciuta importanza clinica. Nota: il pedigree di sinistra rappresenta una CNV associata alla malattia altamente penetrante (non è presente nelle cellule della linea somatica dei genitori, ma compare spontaneamente nella linea germinale del padre). Il pedigree a destra rappresenta una CNV ereditata che può essere associata alla malattia con una penetranza incompleta. BAP= broader autism phenotype. (Eric M. Morrow, Ph.D., M.D. 2010)

Duplicazioni e delezioni di geni che agiscono nella risposta infiammatoria, nell’immunità, nella funzione olfattoria e nella proliferazione cellulare potrebbero essere state fissate e coinvolte nei fenomeni di differenziamento fenotipico adattativo degli uomini, topi e scimpanzè attraverso una selezione positiva (Nguyen, D.Q. et al. 2006; Young, J.M. et al. 2008). Questa diversità potrebbe probabilmente essere responsabile di una significativa parte della variabilità fenotipica (Carter N.P. 2007). Negli ultimi anni sono stati condotti numerosi studi al fine di capire il significato evolutivo di queste varianti, il meccanismo con cui si formano e la loro possibile associazione con malattie ereditarie ad eziologia ignota come l’autismo ed il ritardo mentale.

1.3.1 Formazione delle CNVs

Differenti meccanismi possono portare alla formazione delle CNVs. Le alterazioni cromosomiche strutturali sono il risultato di una o più rotture del DNA a cui fa seguito un ripristino erroneo della continuità dei filamenti nucleotidici coinvolti. Tali alterazioni possono derivare sia da un errore nel processo di riparazione che da un difetto del sistema di ricombinazione. I principali meccanismi che sono stati proposti per spiegare i riarrangiamenti cromosomici strutturali sono:

- la ricombinazione omologa non allelica (*Non Allelic Homologous Recombination* – NAHR),
- l'unione non omologa delle estremità (*Non Homologous End Joining* , NHEJ)
- il modello FoSTeS (*Fork Stalling and Template Switching*) (Gu et al. 2008).

1.3.1.a Non Allelic Homologous Recombination – NAHR

Spesso le CNVs sono fiancheggiate da grandi blocchi di sequenze ripetute di 200-400 kb che presentano un'elevata omologia di sequenza (>95%) chiamate LCRs (anche note come duplicazioni segmentali). Il 99% delle varianti strutturali si formano in seguito ad un evento di ricombinazione omologa non allelica (NAHR) durante le meiosi, causato dall'appaiamento scorretto delle sequenze ripetute che le fiancheggiano (Fig.1.6).

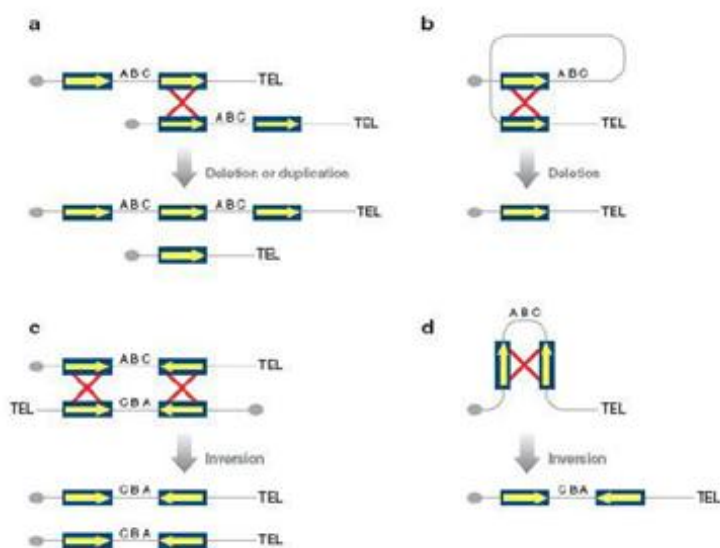


Figura 1.6: Esempi di ricombinazione omologa non allelica: le *Segmental duplications* fungono da substrato per la NAHR. (**a** e **b**) NAHR intercromosomica, intracromosomica, o intracromatidica tra sequenze ripetute orientate nello stesso senso causa delezioni e/o duplicazioni. (**c** e **d**) NAHR intercromosomica, intracromosomica, o intracromatidica tra sequenze ripetute orientate in senso opposto causa inversioni. (Le sequenze ripetute sono rappresentate dalle box blu con l'orientamento indicato dalla freccia gialla. L'evento di ricombinazione è rappresentato da una croce rossa).

La NAHR avviene esattamente con lo stesso meccanismo della ricombinazione omologa (appaiamento filamenti omologhi, formazione del chiasma, risoluzione del crossing-over) con l'eccezione che, nella NAHR, l'alta omologia di sequenza dei segmenti cromosomici che le fiancheggiano, fa sì che il meccanismo cellulare di controllo della ricombinazione non sia in grado di riconoscere l'appaiamento di segmenti non perfettamente identici (Shaffer & Lupski 2000).

L'appaiamento può avvenire fra LCRs su cromosomi omologhi (*intercromosomico*), su cromatidi fratelli (*intracromosomico*), o all'interno di un singolo cromatidio (*intracromatide*) (Stankiewicz & Lupski 2002).

In base al meccanismo di appaiamento, all'orientamento e alla complessità delle LCRs, la NAHR può risolversi in delezioni, duplicazioni, inversioni o altri riarrangiamenti più complessi che possono anche coinvolgere cromosomi non omologhi (Fig 1.7).

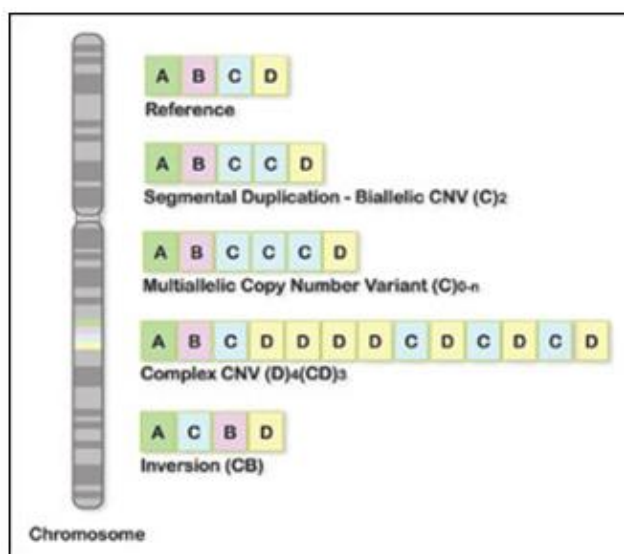


Figura 1.7: tipologie di riarrangiamenti strutturali che si possono ottenere in seguito ad un evento di NAHR.

Sia l'NAHR inter-cromatidica che inter-cromosomica portano ad una duplicazione e una delezione della sequenza interna alle LCRs omologhe non alleliche. Mentre l'NAHR intra-cromosomica, risulta solo in una delezione senza duplicazione. Questo suggerisce che il meccanismo di NAHR porta nell'insieme alla comparsa di un maggior numero di delezioni che di duplicazioni.

La probabilità che una CNV sia associata alle duplicazioni segmentali è maggiore per le duplicazioni intracromosomiche che per quelle intercromosomiche.

Per la ricombinazione omologa non allelica sembra sia necessaria la presenza di segmenti di DNA di lunghezza minima (300-500 bp) chiamati *minimal efficient processing segment* (MEPS) con un'identità di sequenza estremamente alta (>90%) (Reiter et al. 1998). Waldman e Rubnitz hanno dimostrato che l'introduzione di un mismatch anche di soli due nucleotidi riduce di circa 20 volte la frequenza di ricombinazione (Rubnitz & Subramani 1984; Waldman & Liskay 1988). La maggior parte delle LCRs identificate ai punti di rottura di riarrangiamenti cromosomici hanno dimensioni comprese tra 10 e 400Kb e hanno un'omologia di sequenza maggiore del 95% (Lupski 1998; Sharp et al. 2005)

Altri studi hanno osservato il verificarsi di NAHR anche in corrispondenza di sequenze ripetute e intersperse che, complessivamente, rappresentano circa il 45% del genoma. In particolare, sono stati osservati eventi di NAHR fra sequenze *Alu* (classe principale di *Short Interspersed Elements*, SINES) [Redon et al., 2006] e meno frequentemente fra sequenze L1 (classe principale di *Long Interspersed Elements*, LINES) (Burwinkel & Kilimann 1998; Kolomietz et al. 2002).

Tuttavia i tratti di omologia tra due sequenze ripetute intersperse hanno dimensioni molto minori rispetto a quelli presenti nelle LCRs ciò potrebbe spiegare una più bassa frequenza di eventi di ricombinazione mediati da tali sequenze (Gu et al. 2008).

La NAHR può avvenire anche durante la divisione cellulare mitotica come riarrangiamento genomico somatico nelle LCR fiancheggianti regioni che sono state identificate nelle linee cellulari del cancro (Darai-Ramqvist E et al.2008; Fridlyand J, Snijders AM, Ylstra B *et al* 2006).

Diversi studi hanno mostrato che gli eventi di ricombinazione non omologhi tra le LCRs avvengono a livello di siti preferenziali che costituiscono degli Hot Spot di ricombinazione [Lupski, 2004].

Oltre che fra LCRs, la NAHR può verificarsi in presenza di sequenze ripetute subtelomeriche.

Si è osservato che tali riarrangiamenti sono associati in modo significativo ad alcuni casi di ritardo mentale e/o a sindromi di malformazione [Flint et al.,1995; Knight et al.,1999].

Sono stati descritti almeno altri due meccanismi che possono generare delezioni o duplicazioni del DNA [Conrad *et al.*, 2010].

Infatti, meccanismi legati alla replicazione e la riparazione del DNA, come NHEJ (*non homologous end joining*, NHEJ) e MMBIR (*microhomology-mediated break-*

induced replication), possono determinare la formazione di riarrangiamenti più complessi.

1.3.1.b Non Homologous End Joining – NHEJ

Il meccanismo di NHEJ prevale durante la fase G1 del ciclo cellulare, quando non è ancora disponibile una seconda copia del genoma, necessaria per il processo di ricombinazione omologa.

Si basa sulla ligazione fra due segmenti di DNA in cui si è verificata una rottura del doppio filamento (*double strand break-DSB*), in assenza di una sequenza di riferimento (Lupski & Stankiewicz 2005), e può essere descritto in 4 passaggi: identificazione di una rottura del doppio filamento (DSB), formazione di un ponte molecolare che tenga unire le due estremità libere del DNA, processamento delle estremità danneggiate (Fig.1.8) Poiché l'unione delle due estremità avviene tramite l'aggiunta di nucleotidi ad entrambe le estremità in modo da renderle complementari, il meccanismo di NHEJ lascia un'"impronta genetica" fatta di corte sequenze inserite nel punto di rottura (Weterings & van Gent 2004).

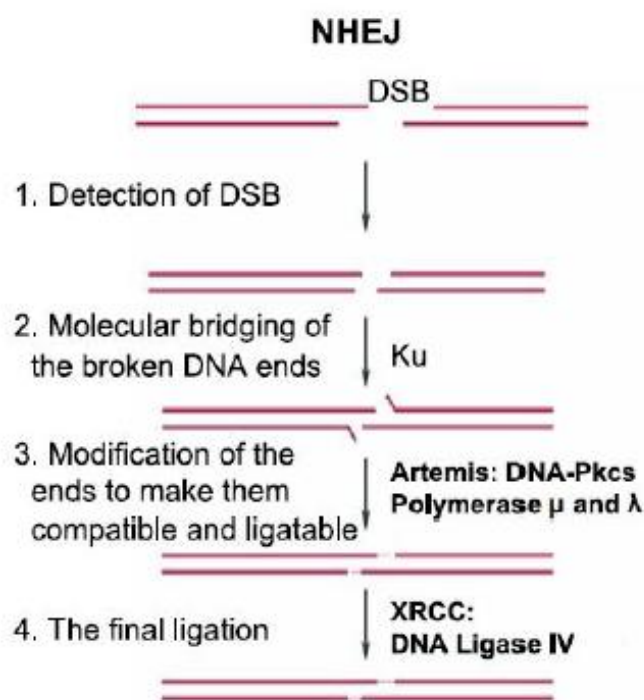


Figura 1.8: meccanismo del Non homologous end joining (NHEJ)

La riparazione è accurata se le estremità sono perfettamente compatibili. Una NHEJ inappropriata può essere fonte di aberrazioni cromosomiche. Perciò questo

meccanismo che non sfrutta l'omologia con il cromatidio fratello, è detto *error prone* cioè introduce mutazioni rispetto alla sequenza originaria.

1.3.1.c Fork Stalling and Template Switching FoSTeS

FoSTeS è un meccanismo basato sulla replicazione proposto inizialmente da Lee nel 2007 (Lee JA, Carvalho CM, Lupski JR 2007), ed in seguito rinominato replicazione indotta da rottura mediata da microomologia o MMBIR da Hastings et al. (Hastings PJ, Ira G, Lupski JR 2009).

Secondo questo modello, durante la replicazione, la forca di replicazione si ferma quando incontra un danno al DNA; il filamento lento si libera dal filamento stampo, si appaia, in virtù della presenza al 3' di regioni di microomologia ad un'altra forca di replicazione, e ricomincia la sintesi del DNA (Fig 1.9).

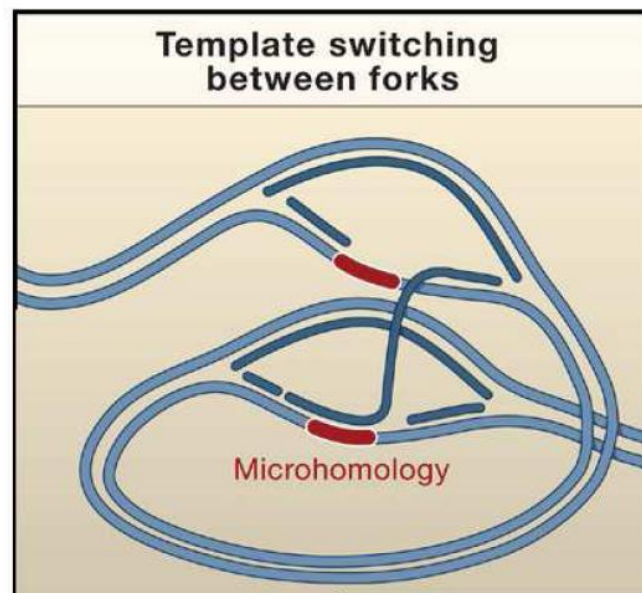


Figura 1.9: modello FoSTeS

L'invasione di una forca replicazione localizzata a valle risulterà in una delezione, mentre l'invasione di una forca localizzata a monte risulterà in una duplicazione. A seconda se è il filamento lento o il filamento guida ad essere invaso e copiato nella nuova forca, e a seconda della direzione di progressione della forca, il frammento incorporato erroneamente dalla nuova forca di replicazione sarà orientato in modo diretto o inverso rispetto alla sua posizione originale. Questo meccanismo di invasione/*annealing* e sintesi/estensione può avvenire numerose volte in serie,

riflettendo probabilmente la scarsa processività della DNA polimerasi coinvolta e causando così complessi riarrangiamenti (Lee et al. 2007).

FoSTeS/MMBIR è probabilmente la causa delle CNVs complesse, tipicamente duplicazioni interrotte da delezioni, attraverso cambiamenti (switches) dello stampo tra la forca di replicazione in associazione con le rotture del doppio filamento.

Queste CNVs sono generalmente marcate da regioni di microomia nei punti di rottura/giunzione che coinvolgono poche bp (2-5 bp)

E' probabile che molte varianti strutturali siano il risultato di una predisposizione al riarrangiamento cromosomico determinata dall'architettura stessa del genoma, che agisce come catalizzatore dell'instabilità cromosomica. Infatti, si pensa che alcuni brevi motivi di sequenza, tra cui le isole CpG, particolarmente presenti a livello dei promotori, determinino una struttura di DNA tale da favorire riarrangiamenti cromosomici.

1.3.2 Metodi per l'identificazione delle CNVs

Ci sono molte piattaforme tecnologiche disponibili per l'analisi delle variazioni del numero di copie di tutto il genoma con risoluzioni di 100-500 kb (rispetto alle 5-10 Mb del cariotipo).

Gli array attualmente disponibili per l'identificazione delle CNVs sono:

- **ROMA:** (*Representational Oligonucleotide Microarray Analysis*) utilizza array di oligonucleotidi, il DNA che deve essere ibridato all'array viene precedentemente trattato con enzimi di restrizione e ligato ad alcune sequenze adattatore. Questi frammenti vengono poi amplificati ottenendo un pool di segmenti rappresentativi dell'intero genoma ma con una complessità inferiore.

[Kennedy et al. 2003](Fig. 1.10)

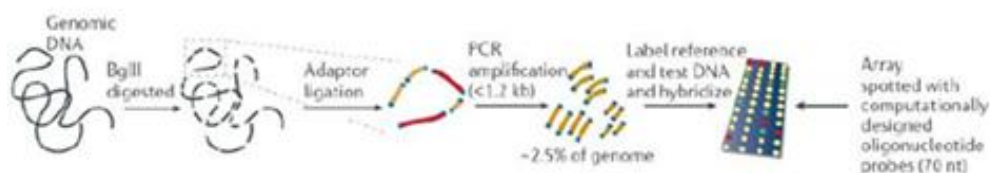


Figura 1.10: Representational oligonucleotide microarray analysis

- **Array di SNP:** Permettono l'identificazione di varianti strutturali con un alto grado di risoluzione. L'intensità di segnale dell'ibridazione di un campione di

DNA genomico di interesse viene confrontato con il valore medio di intensità di un campione di DNA controllo. Una deviazione d'intensità è indice di una variazione del numero di copie. Negli SNP arrays si paragona l'intensità del segnale dei campioni analizzati all'intensità del segnale di un insieme di sonde di DNA di riferimento, o a quella del resto della popolazione analizzata. Questa analisi genera due matrici, il *log2 ratio* e la *frequenza dell'allele B* (*B allele frequency (BAF)*). In una normale situazione diploide, il *log2 ratio* è circa zero. Un aumento del *log2 ratio* rappresenta un aumento del numero di copie e una diminuzione del *log2 ratio* indica una perdita del numero di copie. Il *BAF* dà informazioni sul genotipo e può essere usato per assegnare accuratamente il numero di copie nelle regioni diploidi del genoma da 0 a 4. L'intensità dei segnali di fluorescenza degli SNP arrays sono proporzionali alla quantità di DNA genomico. Quando più sonde consecutive hanno un'intensità del segnale aberrante (aumentata o diminuita), vuol dire che c'è una variazione del numero di copie (Fig. 1.11).

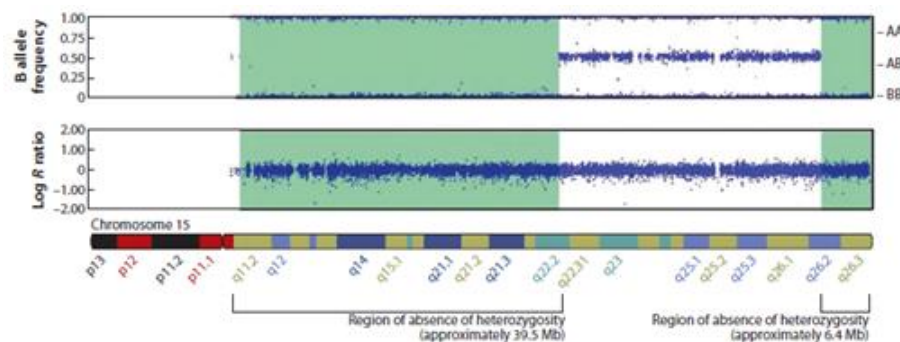


Figura 1.11: isodisomia ed eterodisomia in un cromosoma. Disomia uniparentale materna identificata in un paziente affetto dalla sindrome di Prader Willi mediante IlluminaHuman610Quad BeadChip. La *B-allele frequency* (pannello in alto) mostra due regioni di assenza di eterozigotità di circa 39.5 e 6.4 Mb interrotte da una regione di eterozigotità, indicando un evento di ricombinazione. Nel pannello in basso, il *log2 ratio* indica la presenza di due copie del cromosoma 15(Schaaf, Wiszniewska, L. Beaudet, 2011)

- **Array-CGH:** (*array-based Comparative Genome Hybridisation*) si basa sull'ibridazione competitiva di due campioni di DNA (caso e controllo) marcati diversamente ad una array con migliaia di frammenti genomici noti. [Solinas-Toldo et al. 1997; Pinkel et al. 1998] I primi studi di array CGH erano basati su piattaforme di cloni BAC o oligonucleotidi a bassa risoluzione e permettevano l'identificazione di CNVs tipicamente più grandi di 100 kb. Attualmente le piattaforme di array CGH utilizzate per il genoma intero

permettono l'identificazione di CNVs molto più piccole. (Conrad DF, Pinto D, Redon R et al 2009) Restano, comunque, molto utilizzati anche gli array CGH con cloni BACs perchè garantiscono un'ampia copertura del genoma e sono facilmente reperibili. [Iafrate et al. 2004; Tuzun et al. 2005]

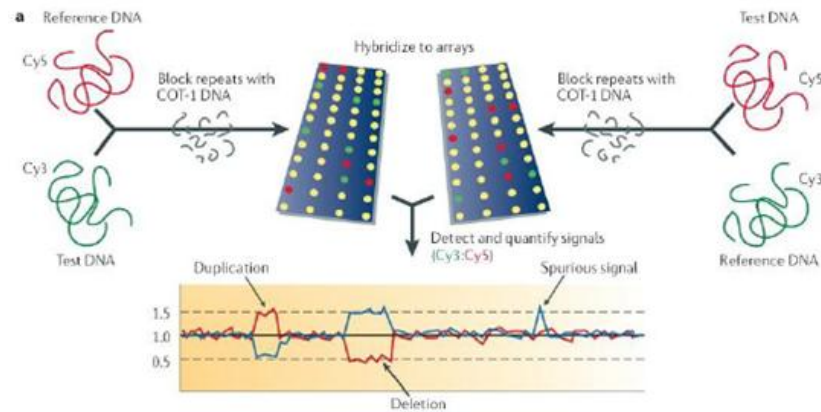


Figura 1.12: array-based Comparative Genome Hybridization.

Da numerosi studi pubblicati, la frequenza delle CNVs identificate come patogene (*pCNVs*) mediante microarray è del 15-20%, rispetto a quella del 3-5% ottenuta mediante l'analisi citogenetica classica nella stessa popolazione di pazienti. (Lupski, J. R. & Stankiewicz, P. 2005) In alcuni casi, il significato funzionale di una particolare CNV può essere sconosciuto e allora ci si riferisce ad essa come variante di incerta importanza clinica (*Variant Of Uncertain Significance, VOUS*). Quest'ultimo caso si verifica quando viene identificata una nuova CNV (ereditata o de novo) non descritta precedentemente. Gli studi delle famiglie possono aiutare nell'interpretazione clinica di questi casi, poichè la comparsa di una CNV de novo rafforza l'evidenza che essa sia patogena. Comunque, a causa dell'espressività variabile o della penetranza incompleta, l'importanza di alcune CNV può risultare incerta anche in seguito a studi sulle famiglie. Quindi, è estremamente utile eseguire analisi comparative di dati provenienti da popolazioni numerose di casi e controlli per associare definitivamente specifiche regioni cromosomiche a malattie umane.

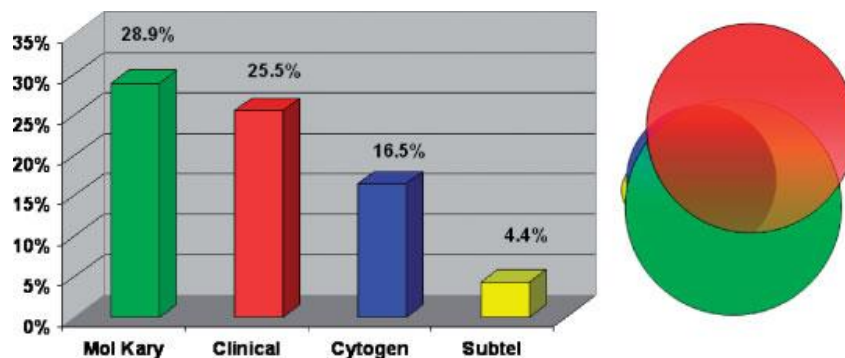


Figura 1.13: confronto tra le CNVs patogenetiche identificabili mediante i vari approcci diagnostici (a sinistra) e disegno schematico della loro sovrapposizione (a destra). (Rauch et al, 2006)

1.3.3 Associazione tra le CNVs e le malattie

Gli sbilanciamenti cromosomici possono essere la causa di diverse patologie quali, ad esempio, sindromi di malformazioni, ritardo mentale, autismo, epilessia e vari tipi di tumori.

Negli ultimi anni la tecnica *array-CGH* ha dimostrato che circa il 15% di soggetti con un fenotipo cromosomico (ritardo mentale/ dismorfismi/ anomalie congenite) e cariotipo normale sono portatori di uno sbilanciamento cromosomico criptico responsabile della loro condizione (Vissers et al. 2003; Shaw-Smith et al. 2004).

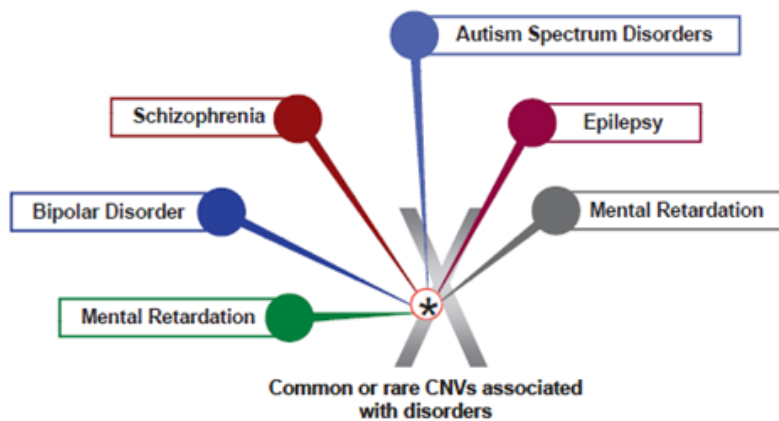


Figura 1.14: malattie dello sviluppo del sistema nervoso associate alle CNVs.

Studi recenti stanno mettendo in luce il loro possibile coinvolgimento come fattori di suscettibilità di malattie complesse ad eziologia ignota quali ad esempio la schizofrenia o l'autismo (Buckley et al. 2005). L'identificazione di un'aberrazione cromosomica in specifici pazienti si è dimostrata un modo efficace per identificare i

geni implicati e per far luce nella patogenesi di differenti condizioni genetiche (Visser, L.E. et al. 2004; Saito, H. et al. 2008)

Le CNVs hanno un tasso di mutazione de novo più alto rispetto alle mutazioni puntiformi (circa $1.7\% \times 10^{-6}$ per locus e per generazione delle CNVs in confronto a 1.8×10^{-8} delle varianti di sequenza) (van Ommen J.F. 2005; Lupski J.R.2007): contrariamente ai tassi di mutazione relativamente costanti degli SNPs in tutto il genoma, i tassi di mutazione delle CNVs possono ampiamente variare in differenti loci. Questo riflette probabilmente le differenze nei meccanismi di formazione delle CNVs. Il fatto che le CNVs coinvolgano almeno un decimo del genoma umano suggerisce che le varianti del numero di copie siano probabilmente importanti nell'eziologia delle malattie.

Le varianti strutturali possono funzionare da fattori di suscettibilità nelle malattie genetiche complesse, sia come causa diretta che come fattori di predisposizione alla malattia (Inoue et al., 2002). Si pensa che le CNVs associate alle malattie alterino il dosaggio genico e possano coinvolgere molteplici geni o regioni regolatorie. La delezione di un gene o di una sua parte può portare alla perdita della sua funzione, che può risultare in un effetto fenotipico dominante o recessivo in accordo alle funzioni cellulari del prodotto genico colpito. Le CNVs possono distruggere un elemento di regolazione, generare nuovi prodotti di fusione, o agire mediante un effetto di posizione, con varie possibili conseguenze positive o negative, come l'imprinting e l'espressione allelica differenziale (Feuk, L., Carson A.R., Shrerer S.W. 2006; Lupski J.R. et Stankiewicz P. 2005).

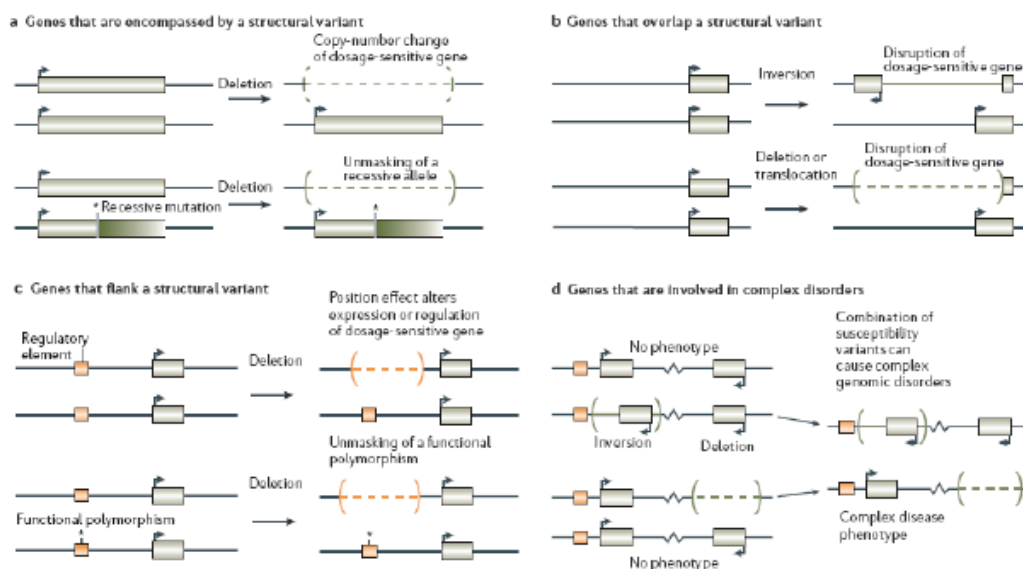


Figura 1.15: Effetti che una CNV può produrre a livello fenotipico

In generale le delezioni sono associate a fenotipi più gravi rispetto alle duplicazioni e le CNVs più grandi sono associate a caratteristiche cliniche più marcate rispetto a quelle più piccole. Alcune CNVs bilanciate che non portano a nessun effetto fenotipico nel portatore, ma possono creare instabilità genomica nelle future generazioni (Feuk,L.,Carson A.R.,Shrerer S.W. 2006). Infine, l'alterazione di specifici segmenti genomici può portare a particolari sindromi o fenotipi (LupskiJ.R. e Stankiewicz 2002). Le 'mutazioni' delle CNVs sono state trovate in tutti i tipi di caratteri mendeliani e in un numero crescente di disordini complessi (Buchanan, J.A. e Shrerer, S.W.2008), compresi quelli che si manifestano come malattie del sistema nervoso (Lee, J.A. e Lupski, J.R.2006).(Edwin H. Cook Jr1 & Stephen W. Scherer 2008, Feuk,L.,Carson A.R.,Shrerer S.W. 2006, Sebat J. Et al 2004)..

Tuttavia, determinare la patogenicità di un'alterazione non è sempre semplice, soprattutto considerando che molti riarrangiamenti ricorrenti sembrano essere caratterizzati da una notevole espressività variabile o penetranza incompleta, essendo appunto riscontrati in membri apparentemente sani della famiglia o con caratteristiche cliniche più lievi (Mefford et al., 2008; van Bon et al., 2009; Hannes et al., 2009).

Allo scopo di comprendere il significato funzionale e fenotipico di queste varianti, sono stati creati sia dei database che raggruppano tutte le CNVs comunemente trovate in individui "normali" di controllo, tra cui il più utilizzato è il "Database of Genomic Variant" (DGV) (Menten et al; J Med Genet 2006), sia database in cui vengono raccolte le informazioni relative al genotipo e al fenotipo dei pazienti che presenatno CNVs de novo o CNVs familiari ereditati da genitori fenotipicamente normali. Tra questi ultimi, quelli attualmente più utilizzati sono il database ECARUCA (<http://www.ecaruca.net>) e DECIPHER (<https://decipher.sanger.ac.uk/application/>).

Un esempio di quanto sia difficile interpretare il significato di alcune CNVs ci è fornito dalla sindrome da duplicazione 22q11 che rappresenta bene il nuovo paradosso incontrato nel cariotipo molecolare, ossia quanto diventano sfocate le relazioni causali tra un'anomalia cromosomica ed un fenotipo associato, per cui sbilanciamenti ereditati da genitori fenotipicamente normali possono contribuire allo sviluppo di una malattia tramite penetranza incompleta e/o espressività variabile, attraverso effetti epigenetici, o mediante la scoperta di una mutazione recessiva sull'allele deleto (Menten et al; J Med Genet 2006). Infatti, le duplicazioni del 22q11 possono portare a diversi fenotipi che vanno dalla normalità, a MR lieve o grave, e

possono conferire una suscettibilità all'insufficienza velo faringea caratteristica della sindrome di DiGeorge/VCFS (sindrome velo cardio facciale OMIM 193430) e anche ad altre caratteristiche distintive. (Menten et al; J Med Genet 2006).

Generalmente, l'interpretazione della causalità delle CNVs si basa su tre criteri: la loro dimensione, il pattern di ereditarietà (*de novo* vs ereditate), e la loro frequenza, cioè se la variazione è stata riportata precedentemente in una popolazione di controllo. Quindi, l'associazione delle CNVs alle malattie deve essere analizzata sistematicamente paragonando la frequenza di ciascuna CNV candidata nei pazienti e nei controlli.

Grazie all'utilizzo sempre più diffuso delle tecnologie basate sugli array sono emerse alcune CNVs ricorrenti che sembrano essere associate a malattie, poiché sono state identificate con una frequenza maggiore nei pazienti affetti rispetto alle popolazioni di controllo. Tali CNVs ricorrenti comprendono i segmenti genomici che coinvolgono le regioni dei cromosomi 1q21.1, 1q41-42, 2p15-q16.1, 3q29, 7q11.23, 9q22.3, 12q14, 14q112, 15q13.3, 15q24, 16p11.2, 16p11.2-12.2, 16p13.1, 17p11.2, 17q21.31, 19q13.11, 22q11.2, e Xq28 (Table 1). Tra le sindromi più comuni causate da CNV ci sono: la sindrome di Smith-Magenis dovuta a una delezione del 17q11.2(OMIM 182290), la sindrome di Williams causata da una delezione al 7q11.23 (OMIM 194050) e la sindrome di Prader-Willi (OMIM 176270) e la sindrome di Angelman (OMIM 105830), entrambe causate da delezioni di 15q11-13.

Table II. Summary of genomic alterations and main clinical features of novel microdeletion and microduplication syndromes			
Genomic aberration	Genomic location (NCBI Build 36.1)	Main phenotype	References
Del. 1q21.1	1.35 Mb Chr1: 145-146.35	Mild to moderate MR, DD, microcephaly, dysmorphic features, cardiac anomalies, and cataracts	31
Dup. 1q21.1	1.35 Mb Chr1: 145-146.35	Mild to moderate MR, DD, microcephaly, ASD	31
Del. 1q41-42	2.7 - 9 Mb Chr1: 219.49-220.66	Moderate to severe MR, coarse facies, deep-set eyes, broad nasal bridge, full lips, cleft palate, diaphragmatic hernia	32
Del. 2p15-16.1	0.5 - 5.7 Mb Chr2: 61.1-61.7	Moderate to severe MR, ASD, microcephaly, dysmorphic features, high nasal bridge, low-set ears, high palate, optic nerve hypoplasia	33
Dup. 3q29	1.61 - 1.76 Mb Chr3: 197.22-198.83	Mild to moderate MR and DD, microcephaly, round face, long arched eyebrows, broad nasal bridge, large eyes, ears and mouth, and obesity	34
Dup. 7q11.23	1.55 Mb	ASD, speech delay, and variable DD	35
Del. 8q21.3-22.1	4 Mb Chr8: 93.21-97.94	Normal to mild DD, microcephaly, tight shiny facial skin, blepharophimosis, contracture of large joints, camptodactyly, hypoplastic genitalia	36
Del. 9q22.3	6.5 Mb Chr9: 94.42-99.1	Severe MR, DD, trigonocephaly, dysmorphic facial features, overgrowth, ventriculomegaly, cerebral atrophy, advanced bone age, and hyperactivity	37
Del. 15q13.3	1.5 Mb Chr15: 28.06-30.79	Mild to moderate MR and DD, dysmorphic facial features, and seizures	38
Del. 15q24	1.7 - 3.9 Mb Chr15: 72.15-73.85	Mild to moderate DD, autistic trait, dysmorphic facial features, and seizures	39
Del. 16p11-12.1	0.6 - 8.7 Mb Chr16: 21.4-30.25	Mild to severe MR, DD, seizures, dysmorphic features, palate defect, aortic valve and other cardiac anomalies, and digital anomalies	40
Dup. 17p11.2	1.3 - 15.2 Mb Chr17: 16.47-18.03	Infantile hypotonia, failure to thrive, mental retardation, autistic features, sleep apnea, and structural cardiovascular anomalies	41
Del. 17q11.2-12	4.4 - 4.6 Mb Chr17: 26.28-31.03	Mild DD, coarse facial features, digital and skeletal anomalies	42
Del. 17q12	1.5 - 2.1 Mb Chr17: 31.83-33.35	Renal cystic dysplasia, renal hypoplasia, abnormal renal function, cryptorchidism, elevated hepatic enzymes, and MODY5	43
Del. 17q21.31	0.5 - 0.65 Mb Chr17: 41.03-41.52	Mild to moderate MR, severe hypotonia, ptosis, dysmorphic facial features, eye and ear anomalies	44
Del. 20q13.13-13.2	4 Mb Chr20: 49.76-50.84	DD, autistic tendencies, dysmorphic features, Duane anomaly, hearing loss, choanal atresia, renal and cardiac anomalies	45
Del. 21q22.12	0.7 Mb Chr21: 33.83-35.64	Growth restriction, dysmorphic features, DD, chronic thrombocytopenia, and predisposition to AML	46
Dup. 22q11.2	1 - 3 Mb Chr22: 17.2-22.1	Highly variable phenotype including MR, DD, learning difficulties, growth retardation, hypotonia, and dysmorphic facial features	12
Dup. Xq28	0.4 - 2.2 Mb ChrX: 152.5-152.9	Severe MR, progressive spasticity, proneness to infections, absent or limited speech	47

Tabella 1: comuni sindromi da microdelezione e micro duplicazione. (Marilyn M. Li and Hans C. Andersson, 2009)

Generalmente le microdelezioni e le microduplicazioni coinvolgono diversi geni, quindi possono determinare sindromi da geni contigui, oppure le caratteristiche fenotipiche associate possono essere determinate dall'aploinsufficienza di un solo gene sensibile al dosaggio. Per questo motivo nella maggior parte dei casi è difficile stabilire quale dei geni coinvolti sia responsabile di un determinato fenotipo clinico, e quali caratteristiche sono associate a grandi delezioni e quali sono dovute solo ad aploinsufficienza di un singolo gene.

Per esempio, per la sindrome di Williams e per la sindrome da delezione del 22q11, ancora non è stato pienamente risolto quale gene, localizzato nella regione deleta comune, sia la causa specifica del fenotipo caratteristico. In altri casi, invece, l'identificazione di un microscopico riarrangiamento cromosomico ha permesso di definire regioni critiche e di trovare il gene direttamente responsabile della patologia (Vissers et al., 2004; Stewart et al., 2004; Kleefstra et al., 2005; Van Esch et al., 2005). La conferma definitiva, in alcuni casi, è stata ottenuta attraverso lo screening della sequenza del gene e l'identificazione di mutazioni patogenetiche in soggetti affetti che non avevano la microdelezione. (Chelly et al., 2006; Raymond e Tarpey, 2006). Per esempio, l'identificazione del gene responsabile della sindrome da

delezione del 9q34(MIM 610253), è stata facilitata dall'analisi di una serie di pazienti, con delezioni di dimensioni sempre più piccole, che manifestavano un determinato fenotipo e all'identificazione di una traslocazione apparentemente bilanciata. Tuttavia, solo l'identificazione di una mutazione puntiforme nel gene EHMT1 (euchromatic histone methyl transferase 1), interno alla regione 9q34, ha permesso di attribuire a questo gene il caratteristico fenotipo di questa malattia.

2. AUTISMO

La parola “autismo” deriva dal greco “autòs”, che significa “se stesso”, ed è stata utilizzata per la prima volta nel 1908 da Eugene Bleuler (1857-1939), psichiatra svizzero, riferendosi ad una particolare forma di ritiro dal mondo, presente nei pazienti affetti da schizofrenia. L’autismo fu descritto per la prima volta attorno al 1940 da Leo Kanner e da Hans Asperger [Kanner et al., 1943; Asperger et al., 1949] che individuarono in alcuni bambini dei comportamenti particolari, non completamente riconducibili al ritardo mentale.

Sia Kanner che Asperger presupposero che tale malattia avesse origine neonatale; tuttavia, negli anni '50 tale ipotesi fu contestata, poiché si vide che i bambini affetti “sembravano intelligenti”, non presentavano caratteristiche dismorfiche o traumi alla nascita, e fu avanzata l’ipotesi che questo disturbo fosse da imputarsi ai genitori, considerati freddi, distanti e socialmente reticenti e si parlò della teoria delle “madri frigorifero” (Eisenberg, 1957; Bettelheim, 1967).

Solo a partire dagli anni '80 è divenuta chiara l’importanza di fattori genetici nella eziologia dell’autismo. Gli studi sui gemelli e sulle famiglie hanno messo in evidenza la complessa eziologia di questa malattia e, allo stesso tempo, la necessità di criteri diagnostici chiari e condivisi. Sono stati dunque elaborati degli algoritmi diagnostici per consentire una valutazione standardizzata dei pazienti affetti da autismo: Autism Diagnostic Interview-Revised (ADI-R) [Lord et al., 1994] e Autism Diagnostic Observation Schedule–Generic (ADOS-G) [Lord et al., 2000]. L’ADOS consiste in un’osservazione interattiva allo scopo di caratterizzare i comportamenti sociali e comunicativi dell’individuo, mentre l’ADI-R è un’intervista strutturata destinata ai genitori basata su domande relative ai comportamenti dell’individuo. I criteri diagnostici per l’autismo sono stati specificati dal Manuale Diagnostico e Statistico dei Disturbi Mentali dell’*American Psychiatric Association* (DSM-IV).

Secondo la definizione ufficiale, l’autismo (OMIM #209850) è una sindrome dello sviluppo del bambino, in particolare è un disturbo neuropsichiatrico, caratterizzato da un comportamento sociale atipico, da disturbi della comunicazione, sia verbale che non verbale, e da comportamenti ripetitivi, stereotipati, e interessi ristretti. La gravità e la sintomatologia dell’autismo variano molto da individuo a individuo e

generalmente i sintomi compaiono nei primi tre anni di vita e persistono per tutta l'età adulta (Volkmar et al., 2004).

L'autismo non rappresenta una categoria di disordini distinta, ma la forma più grave di un gruppo di disturbi classificati con il termine "disturbi pervasivi dello sviluppo" (*Pervasive Developmental Disorders*, PDD) o "disturbi dello spettro autistico" (*Autism Spectrum Disorders*, ASD), che oltre all'autismo propriamente detto, comprendono anche la sindrome di Asperger (AS) e il disturbo pervasivo dello sviluppo non altrimenti specificato (*non pervasive developmental disorder-not otherwise specified*, PDD-NOS). Meno comune, ma catalogata come ASDs, è il disturbo disintegrativo dell'infanzia (*childhood disintegrative disorder*, CDD)[Bailey et al. 2001]. Anche la sindrome di Rett, un disturbo genetico legato al cromosoma X caratterizzato da inadeguata crescita cerebrale, crisi epilettiche, perdita di un utilizzo finalizzato delle mani, era stata classificata come un PDD, ma ora è considerata una sindrome distinta. Infatti, è stato recentemente scoperto che, in circa 80% dei casi, la sindrome di Rett è causata da mutazioni nel gene che codifica per la proteina 2 di legame alle isole CpG metilate (methyl-CpG-binding protein-2, MECP2), che mappa sul cromosoma Xq28 [Amir et al.,1999].

Si stima che l'incidenza dell'autismo vero e proprio nella popolazione sia intorno a 15-20 persone su 10.000, con un rapporto maschi/femmine di 4:1. Quando si considerano invece tutti i disturbi del gruppo ASDs, l'incidenza sale a 116/10.000 [Baird et al., 2005].

Attualmente non esiste una cura per l'autismo, le varie terapie farmacologiche mirano a ridurre alcuni disturbi associati all'autismo: scarsa attenzione e concentrazione, fenomeni ossessivi e compulsivi, eccessiva ansietà, anomalie del sonno, depressione e autolesionismo. Sono stati messi a punto anche dei programmi psicopedagogici nei quali si lavora con il bambino per migliorarne le capacità cognitive e comportamentali.

2.1 Aspetti clinici

Il disturbo autistico compare tipicamente, entro i 3 anni di età, sebbene il riconoscimento diagnostico avvenga di solito verso i 4-5 anni. I sintomi clinici più comuni si manifestano generalmente intorno al 9°-10° mese dalla nascita e in particolare si osserva una diminuzione del pianto, dell'attività motoria in generale e dell'appetito.

In seguito si ha un ritardo nello sviluppo del linguaggio e i bambini sembrano non reagire ai suoni provenienti dall'ambiente circostante. In alcuni casi tuttavia si può osservare un normale sviluppo fino al 18° mese e talvolta anche fino al 24° mese di vita per poi assistere ad una regressione dello sviluppo sociale e comunicativo del bambino e alla comparsa dei sintomi caratteristici.

Le manifestazioni cliniche osservabili sono suddivisibili in tre distinte categorie:

1. Problemi nella comunicazione verbale e non verbale: il 50% dei soggetti autistici non è in grado di comunicare verbalmente; coloro che sono in grado di utilizzare il linguaggio si esprimono in molte occasioni in modo bizzarro; spesso ripetono parole, suoni o frasi sentite pronunciare (ecolalia). Molte volte è presente ritardo mentale; inoltre circa un terzo non sviluppa il senso del passato e possono riferirsi a sé stessi con il nome proprio piuttosto che con "io" o "me".

2. Difficoltà nell'interazione sociale: gli individui affetti da autismo mostrano un'apparente carenza di interesse, tendenza all'isolamento e alla chiusura sociale, apparente indifferenza emotiva agli stimoli o, al contrario, ipereccitabilità agli stessi. Hanno inoltre difficoltà ad instaurare un contatto visivo diretto (es. guardare negli occhi le persone), ad iniziare una conversazione o a rispondere alle domande. È frequente che bambini affetti da autismo vengano inizialmente sottoposti a controlli per verificare una sospetta sordità, dal momento che non mostrano apparenti reazioni (proprio come se avessero problemi uditivi) quando vengono chiamati per nome.

3. Presenza di comportamenti ripetitivi e stereotipati ed interessi ristretti: le attività motorie sono ristrette ad azioni ripetute in modo ossessivo o a sequenze di movimenti stereotipati (per es. torcersi o mordersi le mani). Talvolta la persona affetta da autismo tende ad isolarsi in una sorta di "mondo virtuale", in cui si sente di vivere a tutti gli effetti (dialogando talora con personaggi inventati), può esplodere in crisi di pianto e riso, diventare autolesionista, iperattiva ed aggressiva verso altri o verso oggetti. Al contrario, alcuni mostrano un'eccessiva passività e ipotonia che sembra renderli impermeabili a qualsiasi stimolo.

Tuttavia, è bene sottolineare che esiste una grande eterogeneità nella manifestazione clinica dell'autismo. Sono riportati casi rari di individui affetti da autismo che possiedono, per esempio, una straordinaria capacità di calcolo matematico, sensibilità musicale, eccezionale memoria audio-visiva o altri talenti in misura del tutto fuori dell'ordinario; altri mostrano una ridotta sensibilità al dolore, disturbi del sonno e problemi nel camminare e nello scrivere. Vi è un'estrema variabilità nel livello

cognitivo, che va dal ritardo mentale gravissimo a livelli cognitivi alti o superiori alla norma. Un bambino con ASD su quattro sviluppa crisi epilettiche, che di solito si manifestano nella prima infanzia o nell'adolescenza (Zappella 2010). L'epilessia, causata da un'attività elettrica cerebrale anormale, può provocare perdite di conoscenza momentanee, convulsioni, movimenti insoliti (Zappella 2010).

Il 25% circa dei bambini che soddisfano pienamente i requisiti clinici, all'età di 2-3 anni sono in grado successivamente di sviluppare il linguaggio e comunicare, mentre il restante 75% continua ad avere notevoli difficoltà e richiede un supporto costante e intenso da parte dei genitori, della scuola e dei servizi sociali (Volkmar and Pauls, 2003).

Vari disordini possono essere associati allo spettro dei disturbi autistici, i più comuni sono la sindrome dell'X fragile (MIM 300624), la sclerosi tuberosa (MIM 191100), la fenilchetonuria non trattata e la duplicazione 15q (AUTS4).

2.2 Anomalie del cervello

Studi post-mortem condotti su cervelli di individui affetti, e tecniche non invasive di neuroimaging suggeriscono che i disturbi ASDs potrebbero essere causati in parte da un alterato sviluppo del sistema nervoso. La struttura del sistema nervoso centrale più colpita sembra essere il cervelletto, in cui si riscontra un ridotto numero di cellule del Purkinje. Inoltre caratteristiche cerebrali anomale sono state riscontrate in quelle regioni cerebrali coinvolte nel comportamento sociale, quali l'amigdala e l'ippocampo. Numerosi studi epidemiologici hanno riportato la presenza di macrocefalia nella maggior parte degli individui affetti da autismo (Bailey et al. 1993; Woodhouse et al. 1996; Bolton et al. 2001). Le dimensioni del cervello nell'autismo sembrano essere normali alla nascita, ma dai 2-4 anni di età, il 90% dei pazienti ha un cervello con un volume più grande rispetto alla media normale. In particolare, questi bambini hanno maggiore quantità di materia grigia corticale cerebrale (10-12%), e più materia bianca cerebrale (18%) e cerebellare (39%) rispetto al normale (Courchesne et al. 2001; Courchesne 2002). Durante l'adolescenza queste differenze appaiono meno marcate. La crescita cerebrale e cerebellare rallenta con l'età nell'autismo, e spesso comporta una riduzione delle dimensioni delle strutture del cervello in soggetti autistici adolescenti e adulti, rispetto ai controlli. Il tasso di crescita inusuale del cervello nei primi anni di età potrebbe essere dovuta alla produzione anormale di sinapsi neuronali, alterazioni della morte cellulare e di eliminazione sinaptica

(processo di eliminazione neuronale che risulta in un'innervazione sinaptica aberrante) (Courchesne 2002).

2.3 Anomalie biochimiche

Elevati livelli di serotonina nel sangue, nelle urine, nel liquido cerebro-spinale e nelle piastrine, sono presenti in circa un terzo dei bambini affetti da autismo, rappresentando così i marcatori biochimici più consistenti nell'autismo (Pedersen et al. 1999). Un ruolo per il sistema della serotonina è stato dimostrato durante lo sviluppo del sistema nervoso centrale e nel comportamento sociale, nel sonno, nell'aggressività, nell'ansia e nella regolazione affettiva (Cook e Leventhal 1996). Il trattamento di soggetti autistici con inibitori del trasportatore della serotonina, ha dimostrato di migliorare alcuni sintomi autistici (Brodkin et al. 1997; Fatemi et al. 1998). Elevati livelli di neuropeptidi e fattori neurotrofici, come il fattore neurotrofico che deriva dal cervello (BDNF), neurotrofina 4 / 5 (NT4 / 5) o peptide vesointestinale (VIP) sono stati individuati in bambini autistici alla nascita, tuttavia, simili risultati sono stati ritrovati anche in bambini affetti da altri disturbi neurologici, e quindi sono considerati markers non specifici di sviluppo anomalo del cervello (Nelson et al. 2001).

2.4 Basi genetiche dell'autismo

Il ruolo centrale della genetica nell'eziologia dell'autismo divenne chiaro solo a partire dagli anni '70, quando furono condotti i primi studi epidemiologici su famiglie e su coppie di gemelli [Rutter e Folstein, 1977; Rutter, 2000; Folstein and Rosen-Sheidley, 2001; Lamb et al., 2002]. Da questi studi è emerso che i gemelli monozigoti (MZ) mostrano un tasso di concordanza per l'autismo del 40-60% e per l'ASD del 70-90%, mentre i gemelli dizigoti (DZ) presentano una concordanza molto inferiore: 0-20% (Folstein and Rutter 1977; Steffenburg et al. 1989; Bailey et al. 1995, Lauritsen et al. 2001). I gemelli MZ hanno lo stesso patrimonio genetico, identico al 100%, per cui devono avere le stesse alterazioni genetiche. Al contrario, i gemelli DZ si comportano, dal punto di vista genetico, come fratelli, condividendo solo la metà dei geni. La non totale concordanza dei dati per i gemelli MZ deriva dal fatto che nella genesi della malattia probabilmente intervengono anche altri fattori, ambientali o di altra natura, ancora sconosciuti; la notevole differenza fra i tassi di concordanza tra gemelli MZ e DZ è stata interpretata come indicativo di ereditarietà genetica.

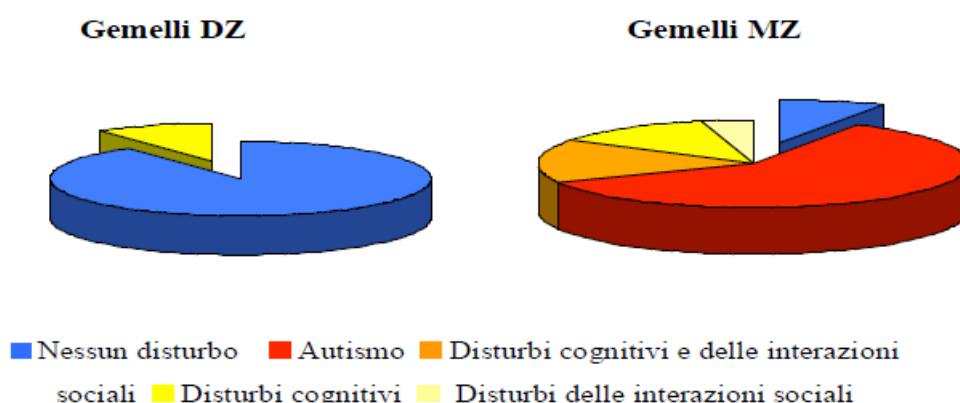


Figura 2.1: Concordanza nei gemelli monozygotici (MZ) e dizigotici (DZ) dello spettro dei disturbi connessi.

Inoltre, studi di ricorrenza familiare [Szatmari et al., 1998] stimano che la percentuale di rischio in fratelli di individui affetti da autismo è 30-40 volte maggiore rispetto al tasso di prevalenza nella popolazione generale. Rispetto al resto della popolazione, fratelli e parenti di individui autistici mostrano più frequentemente caratteristiche comportamentali e cognitive simili a quelle osservate nel probando, ma in forma più lieve (“*broader phenotype*”). Il *broader phenotype*, che può presentarsi come eccessiva timidezza, distacco e disinteresse per le interazioni sociali, ansia e interessi limitati, ma non è associato al ritardo mentale o a fenomeni di epilessia [Rutter, 2000], potrebbe essere dovuto al fatto che i parenti dell’individuo affetto presentano un numero inferiore di fattori di suscettibilità. Infatti, in accordo con il modello multifattoriale “a soglia”, l’autismo vero e proprio compare solo quando viene superata una determinata soglia di suscettibilità (Fig.2.2)

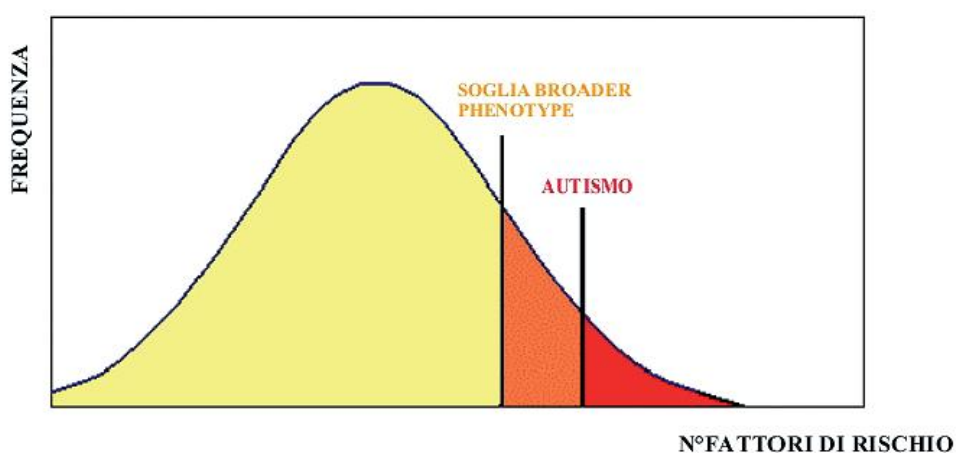


Figura 2.2: Modello multifattoriale a soglia: il raggiungimento della soglia del broader phenotype richiede un numero inferiore di fattori di rischio rispetto al raggiungimento della soglia del disturbo autistico.

Pur essendoci chiare evidenze scientifiche sulla base genetica dell'autismo, non sono ancora noti né il modello di ereditarietà genetica, né i fattori ambientali che conferiscono un rischio di insorgenza in individui geneticamente suscettibili.

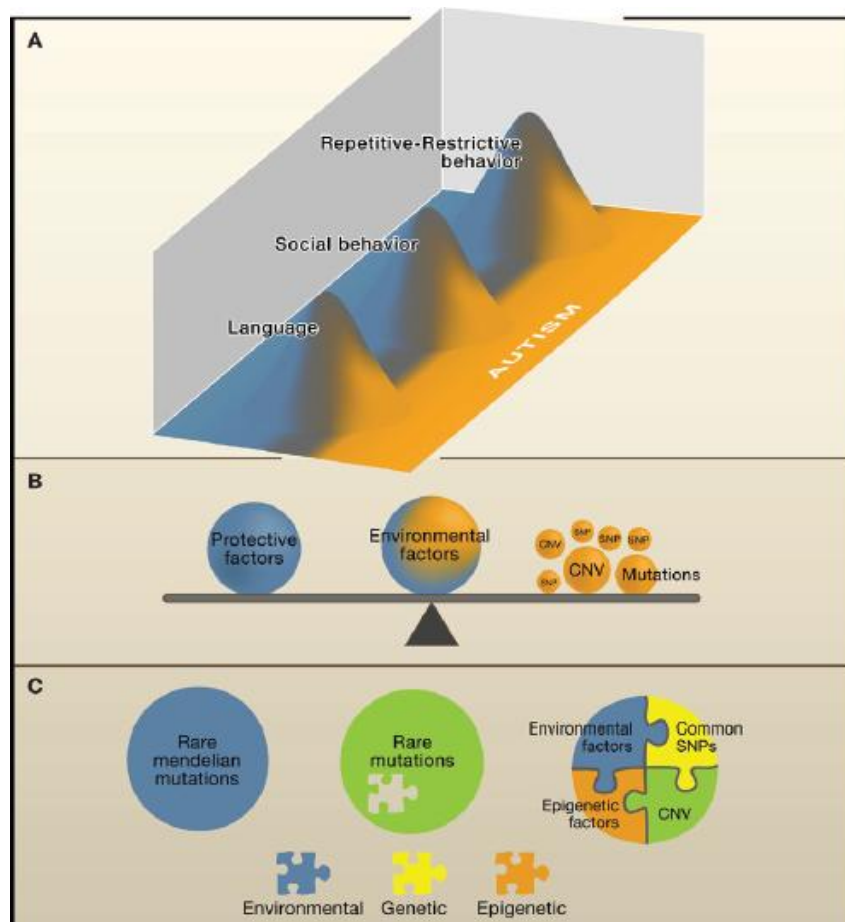


Figura 2.3: Variazione genetica, fattori ambientali ed eterogeneità genetica dell'autismo: in A sono indicate le tre principali categorie delle manifestazioni cliniche dell'ASD; affinché sia diagnosticato l'autismo è necessario che il paziente manifesti un difetto in tutte e tre le aree. (B) e (C) eterogeneità genetica dell'autismo. (Daniel H. Geschwind, 2008)

E' stato proposto un modello ad ereditarietà complessa secondo il quale si ipotizza siano coinvolti da un minimo di tre o quattro loci [Pickles et al.,1995] fino ad oltre 15 loci [Risch et al., 1999; Lamb et al., 2002], che interagiscono in modo epistatico (eredità poligenica o oligogenica).

La complessità clinica e l'eterogeneità del fenotipo autistico riflettono probabilmente la presenza di eterogeneità genetica (geni differenti o combinazioni diverse di geni possono essere coinvolti in famiglie diverse). Altri fattori come il sesso o i fattori ambientali possono dare un loro contributo ad una particolare espressione fenotipica. In un piccolo numero di casi, sono stati riportati anche fattori non genetici, quali danni prenatali, complicazioni ostetriche, esposizione intrauterina a droghe teratogene

ed infezioni virali durante o dopo la gravidanza, che possono incrementare il rischio per l'autismo.

Ricerche più recenti suggeriscono che anomalie citogenetiche submicroscopiche (CNVs) possono contribuire all'autismo in circa il 7%-20% dei casi e che una piccola percentuale di individui (5%-7% dei casi), che risultano inclusi nello spettro autistico dal punto di vista comportamentale, sono affetti da un altro disordine causato da mutazioni in singoli geni, come la sclerosi tuberosa, la sindrome dell'X-fragile, la sindrome di Down e la neurofibromatosi tipo 1 [Sykes & Lamb, 2007; Schaaf & Zoghbi, 2011]. Anche molte condizioni metaboliche sono state associate con il fenotipo ASD in circa il 5% dei casi, quali disordini mitocondriali, fenilchetonuria, deficienza di creatina e altri disordini legati alla biosintesi dello sterolo. Per cui il restante 70% è legato a cause genetiche di ASD non ancora identificate [Schaaf & Zoghbi, 2011] (Fig 2.4).

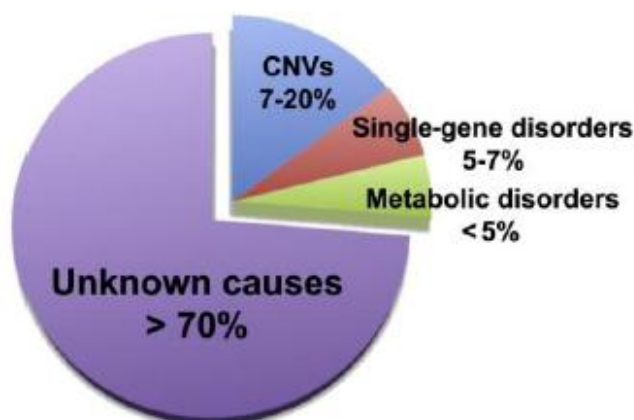


Figura 2.4: cause genetiche dei Disordini dello Spettro Autistico

Per lo studio di malattie complesse, come l'autismo, è necessario disporre di ampie casistiche (famiglie o individui affetti) sia per ottenere un sufficiente potere statistico, sia per avere un campione rappresentativo dell'eterogeneità della malattia a livello di fattori genetici ed ambientali [Lander et al.,1994]. La necessità di disporre di ampie casistiche per lo studio della genetica molecolare dell'autismo ha portato alla nascita di numerosi consorzi internazionali tra cui l'International Molecular Genetic Study of Autism Consortium (IMGSAC) e l'Autism Genome Project (AGP), di cui fa parte anche il nostro gruppo di ricerca. L'IMGSAC è un consorzio composto gruppi di ricercatori e clinici di diversi paesi (Gran Bretagna, Germania, Francia, Olanda, Danimarca, Italia, Grecia e U.S.A.), fondato nel 1994 con lo scopo di raccogliere un

numero elevato di famiglie con autismo ed ha collezionato circa 300 famiglie multiplex (con più di un individuo affetto da autismo). L'AGP (Autism Genome Project) è un consorzio internazionale, sorto nel 2003, allo scopo di mettere insieme le risorse e le competenze cliniche e scientifiche dei principali gruppi di ricerca che studiano l'autismo. Questo consorzio che coinvolge circa 50 differenti gruppi di ricerca in tutto il mondo, ha fino ad oggi collezionato 1500 famiglie multiplex e 3000 famiglie singleton (con un solo individuo affetto da autismo) provenienti dagli Stati Uniti, dal Canada e dall'Europa.

Gli approcci di genetica molecolare che sono stati utilizzati per identificare i geni coinvolti nell'autismo sono:

- Studi di linkage: trovano il legame fisico fra un marcatore e un gene analizzando l'ereditarietà all'interno di famiglie in cui si manifesta la patologia.
- Studi di associazione: mirano ad individuare un'associazione allelica tra marcatori e malattia a livello di popolazione.
- Analisi di anomalie cromosomiche: aiutano a definire intervalli cromosomici in cui è più probabile che sia localizzato un locus di suscettibilità.

2.4.1 Studi di linkage nell'autismo

Gli studi di linkage permettono di stabilire l'ordine dei geni sui cromosomi: il fenomeno del linkage si basa sulla ricombinazione (*crossing-over*) che avviene durante la meiosi tra i cromosomi omologhi e si deve alla stretta vicinanza di due geni o marcatori lungo un cromosoma. Praticamente si assume che se il marcatore ed il gene malattia sono molto vicini tra loro, vengono ereditati insieme con una frequenza maggiore di quella attesa per il solo effetto del caso. I marcatori genetici rappresentano caratteri mendeliani polimorfici, ossia con almeno due alleli alternativi, di cui l'allele più raro deve avere una frequenza nella popolazione significativamente alta (> 1%).

L'analisi di linkage è quindi una metodologia statistica e può essere:

- analisi di *linkage* classica o di tipo **parametrico** (nota anche come **model-based analysis**), basata sull'identificazione dei ricombinanti all'interno di famiglie. Questo tipo di linkage è usato nello studio di malattie mendeliane che si manifestano all'interno di famiglie a più generazioni contenenti diversi individui affetti, dai quali è possibile dedurre il modello di ereditarietà del locus malattia.

- *analisi di linkage di tipo non parametrico* (noto anche come *model-free analysis*) basata sulla condivisione allelica (*allele sharing*) tra due o più individui affetti all'interno di una famiglia. Si studiano alleli identici per discendenza (*Identity By Descent IBD*), cioè quegli alleli che sono identici negli individui affetti perché sono stati ereditati dallo stesso antenato comune. Quindi, se due fratelli sono entrambi affetti da una malattia hanno, con molta probabilità, ereditato dai loro genitori copie identiche della regione cromosomica contenente l'allele di suscettibilità alla malattia stessa. [Penrose, 1953; Suarez, 1978; Kruglyak, 1996].

Quando in uno studio di una malattia genetica non ci sono indicazioni sulle probabili regioni genomiche in cui mappa il gene o i geni coinvolti nella patologia, è necessario effettuare uno studio di *linkage* a livello genomico, *whole-genome scan*, utilizzando marcatori distribuiti il più uniformemente possibile lungo tutto il genoma. In generale un approccio di *whole-genome scan* conduce all'individuazione di regioni cromosomiche candidate, generalmente molto ampie, all'interno delle quali dovrebbero risiedere i loci responsabili della malattia.

Il primo genome scan sull'autismo è stato pubblicato nel 2001 dall'IMGSAC e ha portato all'identificazione di varie regioni cromosomiche in linkage con l'autismo. Le più significative sono sul cromosoma 7q21-q32 (AUTS1) sul cromosoma 2q24-q33(AUTS5) [IMGSAC,1998; IMGASC 2001].

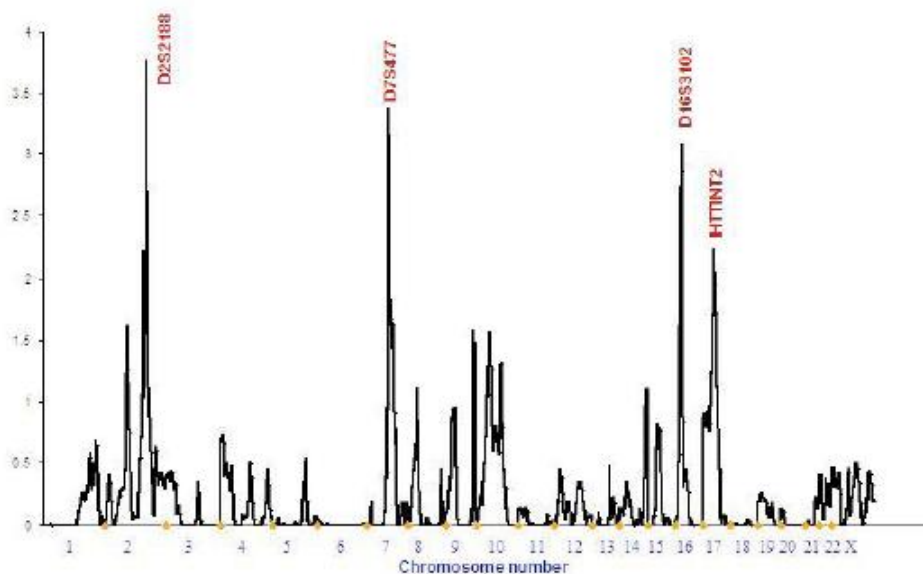


Figura 2.5: Risultati del *genome scan* effettuato dall'IMGSAC. In rosso sono indicati i marcatori in corrispondenza dei quali si ottiene il valore di LOD score più significativo, rispettivamente sui cromosomi 2q, 7q, 6p, 17q.

In seguito a numerose analisi di linkage estese all'intero genoma, sono state identificate regioni in linkage con l'autismo su quasi tutti i cromosomi. Tuttavia, solo una minoranza di loci, tra cui le regioni sul cromosoma 7q, 2q e 17q, sono state confermate da studi indipendenti.

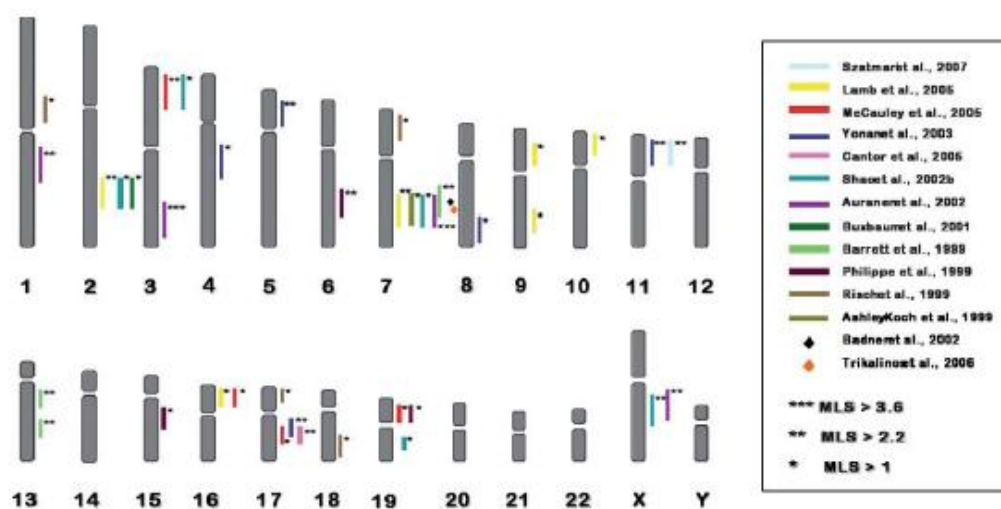


Figura 2.6: Risultati di altri genome scan indipendenti; le regioni sul cromosoma 2q, 7q e 17q risultano essere quelle maggiormente confermate.

La non replicazione dei risultati tra i vari studi di *linkage* indipendenti, potrebbe dipendere dalla presenza di eterogeneità genica e dall'esistenza di più geni che interagiscono in maniera epistatica. Al fine di aumentare l'omogeneità genetica e, di conseguenza, la probabilità di identificare loci di suscettibilità, sono stati perciò condotti studi di *linkage* utilizzando criteri di inclusione più stringenti oppure suddividendo le varie famiglie in sottogruppi, che presentano tratti caratteristici del fenotipo autistico (endofenocopie), senza, tuttavia ottenere risultati conclusivi, probabilmente a causa delle ridotte dimensioni dei sottogruppi analizzati.

Anche il genome scan condotto dall'AGP che ha analizzato all'incirca 1200 famiglie *multiplex* e che quindi rappresenta il più ampio studio di linkage esteso a tutto il genoma, ha riscontrato solo una minima sovrapposizione con le analisi precedenti ed ha portato all'individuazione di un risultato significativo solo sul cromosoma 11p12-p13. [AGP, 2007] Questo conferma ulteriormente la presenza di eterogeneità nel disordine autistico.

2.4.2 Screening di mutazioni in geni candidati per l'autismo

Gli studi di *linkage*, intrapresi negli ultimi dieci anni, hanno identificato diverse regioni cromosomiche che potrebbero essere implicate nella suscettibilità all'autismo. Tuttavia, gli intervalli cromosomici non sono stati sufficientemente circoscritti e per questo motivo molti gruppi di ricerca hanno iniziato ad effettuare *screening* di geni candidati e studi di associazione nelle regioni sub-cromosomiche definite dalle analisi di *linkage*. Un'analisi accurata dei geni candidati viene condotta attraverso il sequenziamento di tutto il gene in individui affetti e in individui sani. Tuttavia, a causa del problema rappresentato dall'elevato costo, il sequenziamento è solitamente ristretto alle regioni codificanti ed importanti dal punto di vista funzionale e/o regolatorio, in un numero limitato di individui. Negli ultimi dieci anni sono stati analizzati più di 100 geni candidati, scelti in base al loro profilo di espressione e al possibile ruolo nello sviluppo e funzionalità del SNC (*candidati funzionali*) e/o alla loro localizzazione all'interno di intervalli di *linkage* (*candidati posizionali*) alla ricerca di varianti implicate nell'eziologia del disturbo autistico. Tuttavia, per nessuno di questi geni è stato dimostrato un chiaro coinvolgimento nella patogenesi della malattia.

2.4.3 Studi di associazione nell'autismo

Per caratterizzare ulteriormente le regioni cromosomiche emerse dagli studi di *linkage* in genere si effettuano studi di associazione.

Gli studi di associazione mirano all'individuazione, all'interno di una popolazione, di varianti che si trovano più frequentemente associate alla patologia. Tale associazione può essere dovuta a diversi fattori quali:

- **effetto diretto** della variante genetica sul fenotipo malattia: ad esempio la variante potrebbe alterare la sequenza aminoacidica di una proteina funzionale coinvolta nel processo fisiopatologico, o alterare la regione regolatoria del gene determinando anomalie a livello di espressione o, infine, interferire con la regolazione dello *splicing* e la stabilità del messaggero
- **Presenza di *linkage disequilibrium* (LD)**: in questo caso la variante allelica mostra associazione con la malattia senza, però, essere direttamente implicata nella patogenesi in quanto i due loci sono vicini sul cromosoma e vengono ereditati insieme come unico blocco.

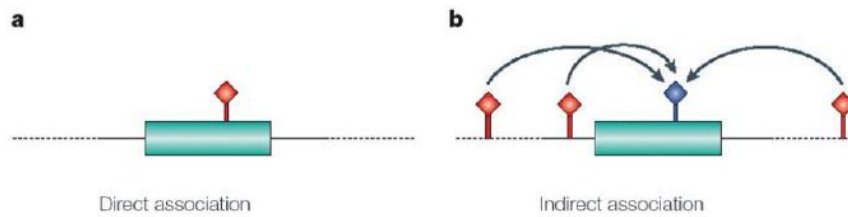


Figura 2.7: Esempi di associazione diretta e indiretta tra una variante allelica e un locus

- **Presenza di stratificazione nella popolazione:** la stratificazione è la presenza nella popolazione di 2 sottogruppi con frequenze alleliche del marcatore e del carattere diverse. Questo può generare un'associazione "spuria" tra il marcatore e il carattere dovuta solamente alla presenza di un errore di campionamento della popolazione studiata.

Le difficoltà riscontrate in uno studio di associazione, soprattutto per quanto concerne la riproducibilità dei risultati ottenuti, derivano dalle differenze tra le popolazioni in esame, dalla disomogeneità dei metodi utilizzati nella definizione del fenotipo e dall'eterogeneità genetica della patologia in esame.

Esistono due tipi di studi di associazione:

- lo **studio caso-controllo**, condotto a livello di popolazione, che confronta la frequenza di un determinato allele tra un campione di individui affetti (casi) e un campione di individui non affetti (controlli) tra loro non imparentati. I casi ed i controlli sono analizzati utilizzando marcatori polimorfici opportunamente scelti. C'è associazione tra un marcatore e la malattia se c'è una differenza statisticamente significativa tra le frequenze alleliche di questi marcatori nei casi e nei controlli. Il problema maggiore in questo tipo di studio è rappresentato dalla stratificazione della popolazione [Slager et al.,2001].
- **Il Disequilibrium Transmission Test (TDT)**, Questo test è condotto su famiglie composte da entrambi i genitori eterozigoti per il marcatore da analizzare e uno o più figli affetti e calcola quante volte i genitori eterozigoti per un certo allele trasmettono l'allele in questione al figlio affetto e quante volte non lo trasmettono [Spielman et al., 1993]. Se la trasmissione dell'allele avviene con una frequenza significativamente maggiore del 50% si è in presenza di associazione tra il marcatore e la malattia. Il TDT non risente del problema della stratificazione perchè le frequenze degli alleli non trasmessi dai genitori fungono da controllo

interno. Il suo limite principale è che ha un potere statistico minore rispetto agli studi caso-controllo [Spielman et al., 1993].

Negli studi di associazione e recentemente anche negli studi di linkage, i marcatori più utilizzati sono i polimorfismi di singolo nucleotide (SNPs) cioè varianti di singola base che si presentano nella popolazione con una frequenza superiore all'1%. Nel genoma umano sono stati osservati circa 10 milioni di SNPs che costituiscono il 90% della variabilità.

Il largo utilizzo degli SNPs è principalmente dovuto al grande contributo che è stato apportato dal progetto internazionale **HapMap**.

Questo progetto [The International HapMap Consortium 2003 e 2005] è sorto con lo scopo di determinare i comuni pattern di variabilità del genoma umano caratterizzando gli SNPs, le loro frequenze e il grado di linkage disequilibrium (LD) in campioni di DNA provenienti da quattro popolazioni diverse (Europei, Yoruba, Giapponesi e Cinesi Han). Durante la prima fase sono stati genotipizzati 1,1 milioni di SNPs (circa uno ogni 5Kb) in 269 individui non imparentati e nella seconda fase circa 5,8 milioni (uno SNPs ogni 1 Kb).

Dal progetto Hap Map è emerso che analizzando solo alcuni SNPs (tag SNPs) opportunamente scelti si può catturare tutta la variabilità genetica di quella regione. Questo implica che il genoma umano è costituito da regioni cromosomiche, dette “blocchi di aplotipi”, all'interno delle quali gli SNPs hanno un alto grado di associazione (figura 2.8).

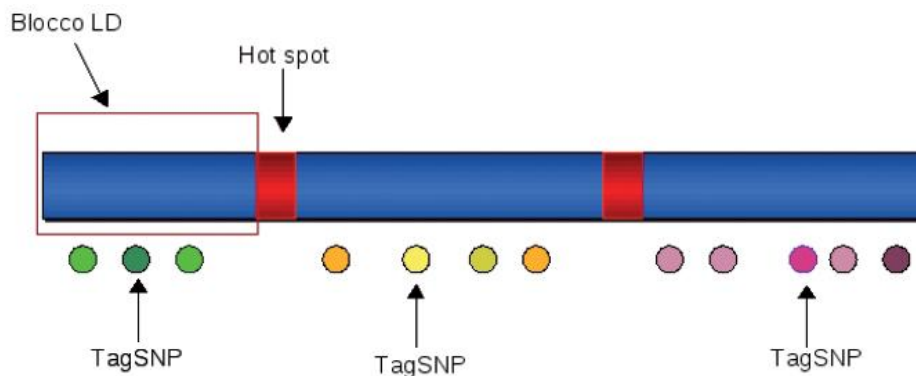


Figura 2.8: Struttura a blocchi di LD del genoma.

Un aplotipo rappresenta uno specifico insieme di alleli che si trovano sullo stesso cromosoma o nella medesima regione cromosomica. Questo significa che all'interno di un “blocco” la diversità di una popolazione umana è rappresentata solo da pochi aplotipi (Patil et al. 2001). I blocchi di aplotipi sono separati da regioni con un basso

livello di LD ed un'alta diversità, che corrispondono probabilmente a “hotspot” di ricombinazione.

Il linkage disequilibrium deriva dal fatto che alleli vicini, essendo associati in un unico aplotipo, vengono più facilmente coereditati (ereditati insieme). [Gabriel et al., 2002; Abecasis et al., 2001; Reich et al., 2001].

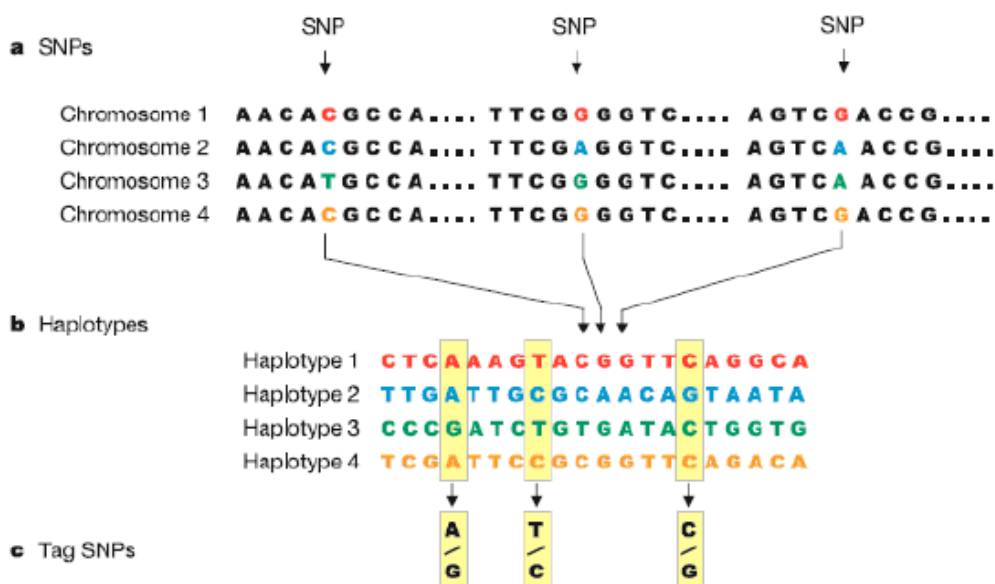


Figura 2.9: a) Sequenze genomiche caratterizzate dalla presenza di SNPs; b) Successioni di SNPs (aplotipi) che rappresentano una data sequenza genomica; c) Tag SNPs che rappresentano gli aplotipi di una data sequenza genomica.

2.4.4 Studi di associazione in geni candidati per l'autismo.

Le regioni di suscettibilità identificate mediante studi di *linkage* sono piuttosto ampie e contengono molti geni: ad esempio, i loci AUTS1 e AUTS5 si estendono per circa 40 Mb e comprendono approssimativamente 200 geni noti. Per analizzare i geni candidati più promettenti che mappano nelle regioni di *linkage*, vari gruppi di ricerca hanno intrapreso studi di associazione su singoli geni. Ad oggi, sono già stati analizzati oltre 200 geni, di cui più di 80 di essi hanno dato risultati positivi, anche se spesso questi non sono stati replicati da altri gruppi. Questo potrebbe essere spiegato in parte con l'eterogeneità del disturbo autistico, in parte con la variabilità tra questi studi in termini di potere di risoluzione, numero di campioni analizzati e densità di marcatori. Un esempio di studi di associazione con risultati controversi è rappresentato dal gene *SLC25A12*, che mappa nella regione di *linkage* 2q31.1. Esso codifica per un carrier aspartato/glutammato calcio-dipendente, localizzato nella

membrana interna del mitocondrio, ed è espresso nel cuore, nel muscolo e nel cervello [del Arco e Satustregui, 1998]. L'RNA messaggero e la proteina SLC25A12 sono presenti ad alti livelli nel sistema nervoso centrale del topo adulto. Risultati positivi di *linkage* e associazione sono stati ottenuti per due polimorfismi intronici (rs2056202 e rs2292813) in un campione di 411 famiglie multiplex [Ramos *et al.*, 2004]. Per entrambi gli SNPs, la rilevanza funzionale è poco chiara, dal momento che sono localizzati rispettivamente nel terzo e nel sedicesimo introne del gene. E' probabile quindi che la variante causale sia diversa da quelle identificate da Ramos e colleghi, ma in forte *linkage disequilibrium* con esse. L'associazione di questi due SNPs con l'autismo è stata successivamente confermata su un campione di famiglie Irlandesi [Segurado *et al.*, 2005], mentre altri due studi non hanno replicato questo risultato nelle loro collezioni di famiglie [Rabionet *et al.*, 2006, Blasi *et al.*, 2006]. La mancata replicazione potrebbe avere diverse interpretazioni. E' possibile che i risultati discordanti siano dovuti a differenze tra le popolazioni utilizzate, per esempio diverse caratteristiche fenotipiche non rilevate, in quanto sottili differenze nell'espressione della malattia nei vari sottogruppi potrebbero avere implicazioni genetiche importanti. Un'altra ipotesi è invece quella che queste varianti conferiscano un moderato rischio relativo, che non può essere rilevato da questi studi di dimensioni limitate.

Dato che molteplici studi di associazione condotti su geni candidati si sono dimostrati inconcludenti e non hanno permesso di evidenziare un chiaro coinvolgimento dei geni identificati nell'eziologia del disturbo autistico, e che comunque è molto difficile selezionare *a priori* un particolare gene candidato, principalmente perché un terzo di tutti i geni umani è espresso nel cervello o durante lo sviluppo cerebrale, si è sentita l'esigenza di effettuare studi di associazione su larga scala.

2.4.5 Studi di associazione genome-wide

Con l'avvento di nuovi metodi di genotipizzazione su larga scala, a costi accessibili, è stato possibile estendere gli studi di associazione a campioni molti più numerosi ed all'analisi dell'intero genoma. Nel 2009 è stato pubblicato il primo studio di associazione esteso all'intero genoma per l'autismo (GWAS). In particolare, è stata eseguita un'analisi TDT su 780 famiglie multiplex con individui affetti da autismo provenienti dall'*Autism Genetic Resource Exchange* (AGRE) e un'analisi caso-controllo comparando un campione di 1204 casi provenienti da vari stati americani

con 6491 controlli provenienti dal Children's Hospital of Philadelphia. [Wang et al., 2009] Tutti gli individui sono stati tipizzati utilizzando l'Illumina Human Hap 550 Bead Chip. Separatamente nei due campioni non è stata rivelata nessuna associazione statisticamente significativa, mentre l'analisi combinata dei due set di risultati (meta-analisi) ha permesso di identificare 6 SNPs fortemente associati all'autismo (p value $<1 \times 10^{-7}$) che mappano nella regione cromosomica 5p14.1. Analizzando questa regione si è visto che i vari SNPs identificati erano localizzati entro lo stesso blocco di linkage disequilibrium di circa 100 kd situato in una regione tra i geni CDH10 (caderina 10) e CDH9 (caderina 9). Le caderine rappresentano un grande gruppo di proteine transmembrana coinvolte nell'adesione cellulare e nella formazione delle connessioni neuronali durante lo sviluppo del cervello.

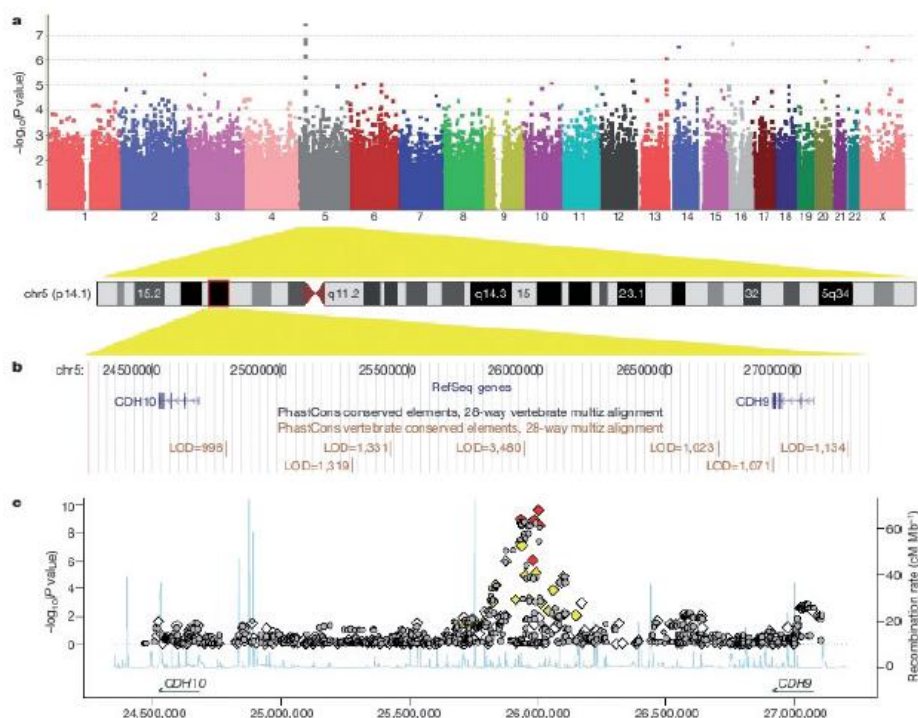


Figura 2.10: Risultati del primo studio di associazione caso-controllo e TDT. Combinando i due risultati si identificano alcuni SNPs (rosso) fortemente associati all'autismo. Mappano nella regione 5p14.1, si trovano nello stesso blocco di LD tra i geni CDH10 e CDH9. (Wang, Haitao Zhang, Deqiong Ma et al, 2009)

Un secondo studio GWAS per l'autismo è stato condotto su 1031 famiglie multiplex provenienti dalla collezione AGRE e da quella del National Institute for Mental Health Autism (NIMH). I soggetti sono stati genotipizzati con Affimetrix 500K e 5.0 [Weiss et al., 2009]. L'analisi iniziale non ha identificato nessuna associazione che raggiungesse soglia di significatività prefissata per il GWAS, ma replicando i risultati più significativi in un campione indipendente, è stata confermata un'associazione

significativa per lo SNP rs10513025, localizzato in 5p15. Anche in questo caso, lo SNP è stato trovato all'interno di un blocco di LD, circa 80 bp a monte del gene *SEMA5A*, un membro della famiglia delle semaforine assonali. *SEMA5A* sembra un buon gene candidato per l'autismo: la sua espressione risulta ridotta nel cervello di individui autistici, inoltre rappresenta il recettore della plexina B3, la cui via di segnalazione prevede la presenza della tirosina kinasi MET, anch'essa coinvolta nella patogenesi dell'autismo.

Un terzo studio di associazione *genome-wide* è stato recentemente condotto dall'AGP, che ha tipizzato 1 milione di SNPs in più di 1500 famiglie con ASD. Il risultato più significativo, che raggiunge la soglia di significatività prefissata ($P= 5 \times 10^{-8}$), è stato identificato in corrispondenza del gene *MACROD2*. Dato che lo SNP più associato in *MACROD2* sembra regolare l'espressione della fosfolipasi D2 (*PLD2*) e che ulteriori analisi, condotte suddividendo il campione in base a specifici fenotipi (IQ, linguaggio), hanno identificato segnali significativi di associazione per la fosfolipasi D5 (*PLD5*), si ipotizza che la famiglia delle fosfolipasi possa avere un ruolo importante nella suscettibilità all'autismo.

E' interessante notare come questo studio non abbia replicato i risultati ottenuti dai precedenti GWAS, suggerendo che per identificare varianti con rischi relativi molto modesti (OR intorno a 1.1-1.3) sia necessario analizzare campioni molto più numerosi. Questi studi dimostrano quindi che le varianti comuni, ad oggi identificate, che contribuiscono all'eziologia dell'autismo, rappresentano solo una piccola porzione dell'ereditabilità dei disturbi ASDs. Tuttavia, esse suggeriscono pathways molecolari che potrebbero avere un ruolo importante nella patogenesi dei disturbi ASDs

2.4.6 Studi citogenetici.

Un approccio potenzialmente utile nel mappaggio e nell'identificazione di geni malattia consiste nell'analisi di anomalie citogenetiche in pazienti autistici.

Studi citogenetici hanno indicato che una piccola percentuale di casi di autismo (6-7%) è associata ad una ampia gamma di anomalie cromosomiche, il cui significato nei confronti dei disturbi dello spettro autistico rimane poco chiaro. La maggior parte di tali anomalie è stata, infatti, riscontrata in casi sporadici e, in genere, si tratta di soggetti con forme sindromiche con fenotipo dismorfico e compromissione cognitiva. Dallo studio effettuato da Vorstman e collaboratori [Vorstman *et al.*, 2006], al fine di

identificare specifiche regioni cromosomiche associate al fenotipo autistico, è emerso che le principali regioni citogenetiche d'interesse (CROIs) non sono distribuite uniformemente nel genoma umano ma tendono ad organizzarsi in cluster, soprattutto in corrispondenza delle regioni cromosomiche 2q37, 5p14-p15, 7q, 15q11-q13 e Xp22.

Ad esempio, tra le anomalie citogenetiche più comuni vi sono alterazioni che riguardano la regione prossimale del cromosoma 15q (0.97%). Questa regione a causa della presenza di duplicazioni segmentali è altamente instabile cioè soggetta a frequenti riarrangiamenti tra cui delezioni, duplicazioni o triplicazioni di segmenti genomici. Le delezioni di regioni nell'intervallo 15q11-15q13 possono dare origine alle sindromi di Prader-Willi (PWS, MIM 176260) e di Angelman (AS, MIM 105830), due disordini dello sviluppo associati a ritardo mentale. In circa l'1-3% dei casi ASD si osservano duplicazioni nel cromosoma 15, principalmente di due tipi:

1. duplicazioni interstiziali della regione 15q11-q13; particolarmente in quei casi con ritardo mentale e attacchi epilettici [Cook, E.H., Jr., 2001; Repetto et al., 1998; Schroer et al., 1998; Lamb et al., 2002].
2. cromosoma 15 sopranumerario isodicentrico. (Rineer et al., 1998; Buoni et al., 2000; Wolpert et al., 2000),

Dal momento che i geni coinvolti nelle sindromi di PWS/AS sono soggetti ad *imprinting* e che tutte le duplicazioni osservate in questa regione nei casi ASD sono di origine materna, è possibile che al suo interno sia presente un locus di suscettibilità per l'autismo soggetto ad *imprinting*.

2.5 CNVs ed Autismo

Numerosi studi genome-wide, condotti al fine di testare l'associazione dei CNVs con malattie neuropsichiatriche, hanno rivelato che tali disordini sono per lo più causati da CNVs rari e altamente penetranti, che coinvolgono la perdita, l'acquisto o la distruzione di geni sensibili al dosaggio [Cook & Scherer, 2008]. Nello studio condotto dall'AGP e pubblicato nel 2007, analizzando circa 1500 famiglie multiplex con autismo idiopatico mediante l'Affimetrix Gene Chip 10K (10.000 SNPs), è stato identificato un numero inaspettatamente elevato di varianti strutturali submicroscopiche. In particolare sono state identificate 254 CNVs, di lunghezza media di 3.4 Mb, in 196 probandi provenienti da 173 famiglie e 10 CNVs *de novo*, tra

queste ultime, una delezione *de novo* sul cromosoma 2q16 presente in due fratelli affetti, che coinvolge il gene della neurexina1 [AGP, 2007] (Fig 2.11).

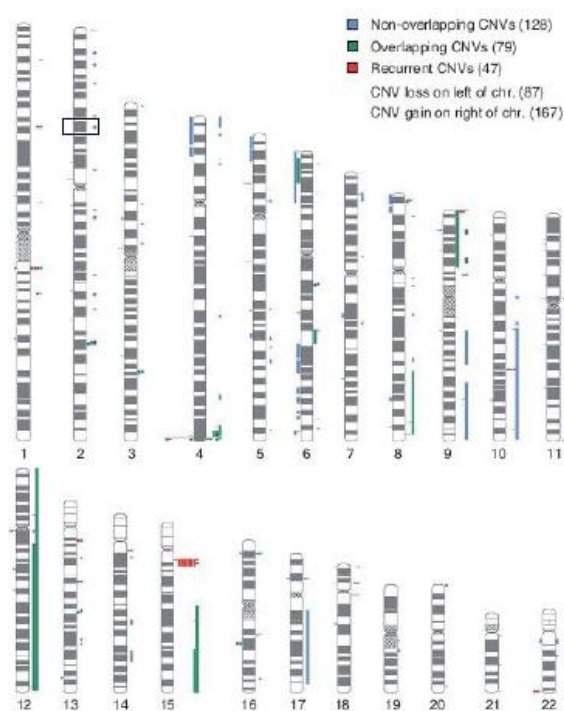


Figura 2.11: CNVs identificate nello studio pubblicato nel 2007 dall'AGP. Nel rettangolino è evidenziata la delezione *de novo* nel gene NRX1.

La neurexina è un buon candidato per l'ASD poiché svolge un ruolo importante nella sinaptogenesi e mutazioni rare in questo gene sembrano generare un aumento di rischio per lo sviluppo delle patologie dello spettro autistico [AGP, 2007].

Sempre nel 2007, è stato portato a termine un altro screening di CNVs mediante CGH array in 118 famiglie sigleton, 47 famiglie multiplex e 99 controlli. Nelle famiglie con ASD è stato identificato un maggior numero di CNV *de novo* rispetto ai controlli, differenza che diventa ancora più spiccata nelle famiglie sigleton rispetto alle multiplex (la percentuale di CNVs identificati è di circa 2-3% nelle multiplex, 7-10% nelle sigleton, 1% nei controlli). Ciò suggerisce che vi possano essere meccanismi diversi alla base dei casi sporadici e familiari [Sebat et al.,2007].

In molti casi individui non imparentati possiedono CNV, *de novo* o ereditati, allo stesso locus ed alcuni di questi coincidono con disordini genomici noti, mentre altri coinvolgono geni associati al ritardo mentale. Inoltre i CNVs che sembrano avere un ruolo nell'eziologia dell'autismo colpiscono soprattutto i geni dei complessi sinaptici neuronali, quali *SHANK3*, *NLGN4* e *NRXN1* [Cook & Scherer, 2008].

Un'ulteriore evidenza di associazione tra CNVs e autismo proviene da uno studio sull'intero genoma condotto nel 2008 mediante la piattaforma di microarray Affymetrix 5.0, allo scopo di trovare varianti genetiche associate all'autismo [Weiss *et al.*, 2008]. È stato tipizzato un campione di 751 famiglie multiplex con individui affetti da autismo ma senza anomalie cromosomiche provenienti dall'Autism Genetic Resource Exchange (AGRE)[Geschwind *et al.*, 2001]. Il risultato più significativo è stata l'identificazione di una regione nel cromosoma 16p11.2 che appare deleta o duplicata nell'1% delle famiglie multiplex analizzate. La regione contiene circa 25 geni tutti buoni candidati per l'autismo in quanto ampiamente espressi nel cervello e coinvolti nello sviluppo neuronale [Weiss *et al.*, 2008]. Successivi studi condotti sulla regione 16p11.2 hanno evidenziato che le varianti strutturali identificate sembrano non essere specifiche per l'autismo, ma sono presenti anche in casi di dimorfismo facciale e di ritardo mentale [Cook and Scherer, 2008]. Questi studi suggeriscono che perdite o acquisizioni di materiale al medesimo locus possano influenzare più di un carattere (effetto pleiotropico), accrescendo il rischio per una vasta gamma di patologie neuropsichiatriche. Ciò significa che questi loci hanno bassa penetranza ed espressività variabile, con fenotipi che variano da anomalie di media gravità fino al Ritardo Mentale (MR), all'autismo e alla schizofrenia [Stefansson *et al.*, 2008; Cook e Scherer, 2008]. Un altro esempio è dato dalle microdelezioni nella regione cromosomica in 15q13.3, che presentano un'estesa variabilità fenotipica, essendo presenti in individui normali, in casi di epilessia, di MR, di autismo o schizofrenia [Helbig *et al.*, 2009, Sharp *et al.*, 2008, Pagnamenta *et al.*, 2008, Miller *et al.*, 2008; Stefansson *et al.*, 2008]

Dall'analisi dei dati di SNP array utilizzati per il GWAS pubblicato dal gruppo di Wang (Wang *et al.*, 2009), sono emersi risultati significativi per delezioni o duplicazioni che sono individualmente molto rare, ma che tendono a coinvolgere geni implicati nell'adesione cellulare o appartenenti alla via dell'ubiquitina, meccanismo anch'esso coinvolto nella regolazione della connettività neuronale [Glessner *et al.*, 2009].

Per far luce sul contributo di varianti rare nei disturbi dello spettro autistico, nel 2010 l'AGP ha realizzato uno studio su larga scala, tipizzando 1275 casi ASDs e 1981 controlli con lo SNP array Illumina Infinium 1M [AGP, 2010]. Considerando solo gli individui di origine europea, almeno il 5,7% degli affetti risulta portatore di un CNV *de novo* e, a differenza di quanto osservato nei lavori precedenti, il tasso di CNVs *de*

novo risulta molto simile per famiglie simplex e multiplex (5,6% e 5,5%). Inoltre, allo scopo di valutare se geni e CNVs precedentemente associati ad ASD o a ritardo mentale fossero maggiormente presenti nei casi rispetto ai controlli, sono stati definiti tre gruppi di geni: geni implicati nell'ASD (*ASD implicated*), geni del ritardo mentale (*Intellectual disability*) e geni candidati ASD (*ASD candidates*). I risultati mostrati in figura 2.12 indicano che CNV rari (con frequenza minore dell'1%) che colpiscono i geni implicati in ASD e quelli del ritardo mentale sono maggiormente presenti nei casi rispetto ai controlli, sia quando questi due set di geni sono considerati singolarmente che in combinazione. Al contrario se si confronta la percentuale di CNV che interessano i geni candidati ASD la differenza tra casi e controlli risulta non significativa.

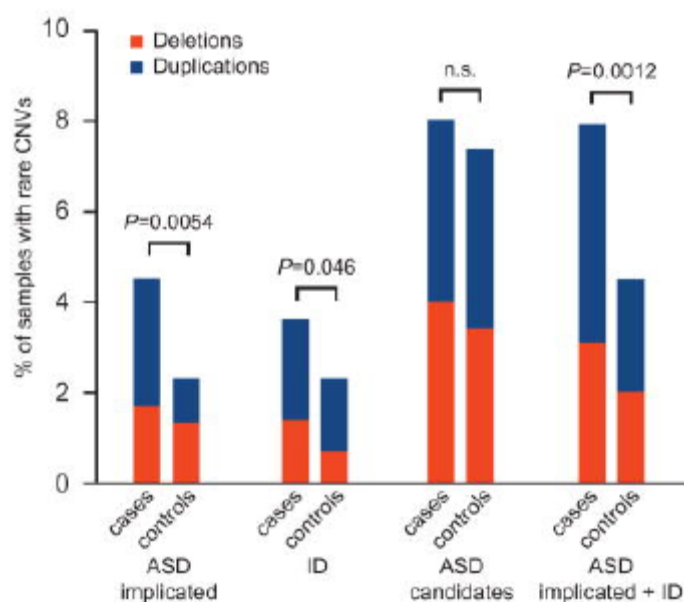


Figura 2.12: CNV identificati in geni noti ASD o implicati nel ritardo mentale

Infine, al fine di identificare set di geni colpiti da CNV che condividono una funzione comune o che operano nella stessa via del segnale (*pathway*), è stata costruita una mappa funzionale di geni coinvolti nei disturbi ASD: la maggior parte dei CNV identificati distruggono geni coinvolti nell'adesione cellulare, nella glicosilazione, nell'organizzazione del citoscheletro, nella via di segnalazione Ras/GTPasi e, soprattutto, nello sviluppo del sistema nervoso. (Fig.2.13)

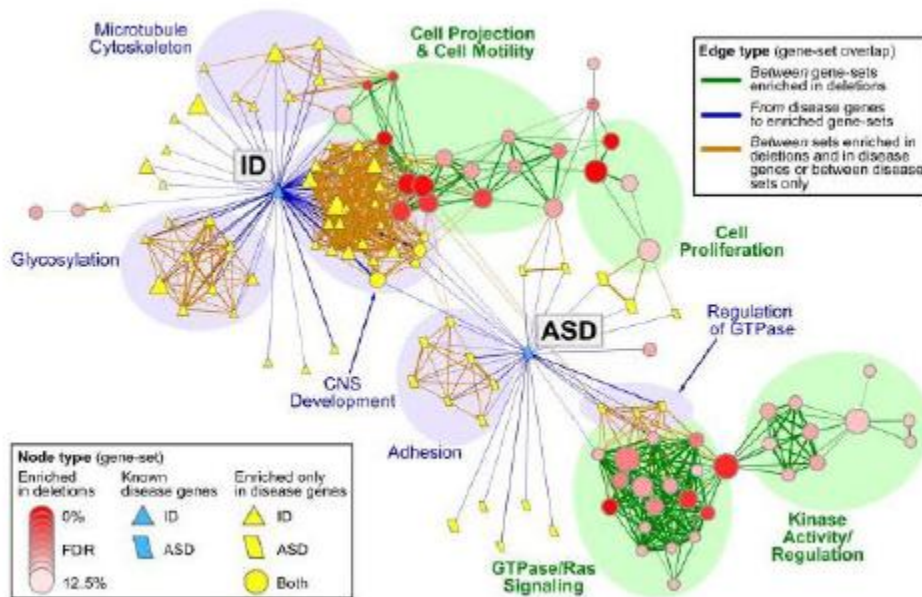


Figura 2.13: Mappa funzionale che mostra le connessioni tra le categorie geniche (*gene-sets*) maggiormente colpite da rari CNVs nei casi ASDs rispetto ai controlli. Ogni categoria è rappresentata da un nodo, le cui dimensioni sono direttamente proporzionali al numero di geni che ne fanno parte. I gruppi funzionali principali sono indicati dalle aree verdi (colpiti da delezioni) e blu (noti per essere in ASDs o in *Intellectual Disability*, ID).

Questi dati supportano quindi il coinvolgimento dei CNVs rari, sia sparsi in tutto il genoma che in specifici loci, nella suscettibilità all'ASD e sottolineano il ruolo di molecole, quali *SHANK3*, *NRXN1* e *NLGN3*, nella maturazione e funzionalità delle sinapsi glutammatergiche [Pinto et al., 2010].

Recentemente, usando differenti piattaforme di microarray sulla stessa coorte di pazienti, due studi indipendenti condotti da Sanders [Sanders et al., 2011] e Levy [Levy et al., 2011] hanno confermato il ruolo dei CNV *de novo* nell'eziologia dell'autismo idiopatico. L'analisi di un gran numero di famiglie singleton provenienti dalla collezione Simons Simplex (*Simons Simplex Collection*, SSC), ha permesso loro di confermare multipli loci ASD noti ma anche di identificarne dei nuovi. In particolare i risultati ottenuti dallo studio di Levy rilevano che un grande contributo alla suscettibilità dei disturbi dello spettro autistico è svolto da CNV "ultra rari", soprattutto duplicazioni ereditarie, presenti nei probandi, ma non nei loro fratelli non affetti. E' stata anche ipotizzata una maggiore resistenza al fenotipo autistico da parte di probandi di sesso femminile rispetto a quelli di sesso maschile: le femmine con ASD hanno una più alta frequenza di eventi *de novo* (11,7%) rispetto ai maschi (7,4%) ed, inoltre i CNV *de novo* colpiscono un maggior numero di geni. Sanders e collaboratori, invece, non hanno identificato un ruolo causale di CNV rari legati al

cromosoma X, dal momento che non hanno rilevato differenze significative tra probandi maschi e probandi di sesso femminile per quanto riguarda il tipo di CNV e il suo effetto fenotipico. Tuttavia, l'identificazione di un grande network biologico di geni colpiti da CNV rari *de novo* mostra in maniera convincente che sono richiesti più forti sconvolgimenti funzionali per scatenare il fenotipo autistico nelle femmine rispetto ai maschi. Inoltre, poichè tale network racchiude geni implicati in precedenza nel fenotipo autistico e principalmente coinvolti nello sviluppo sinaptico, targeting assonico e motilità neuronale, questo potrebbe dimostrare che l'analisi funzionale di varianti genetiche rare può essere un approccio valido per l'identificazione di pathway biologici alla base di fenotipi umani complessi [Gilman et al., 2011](Fig 2.14)

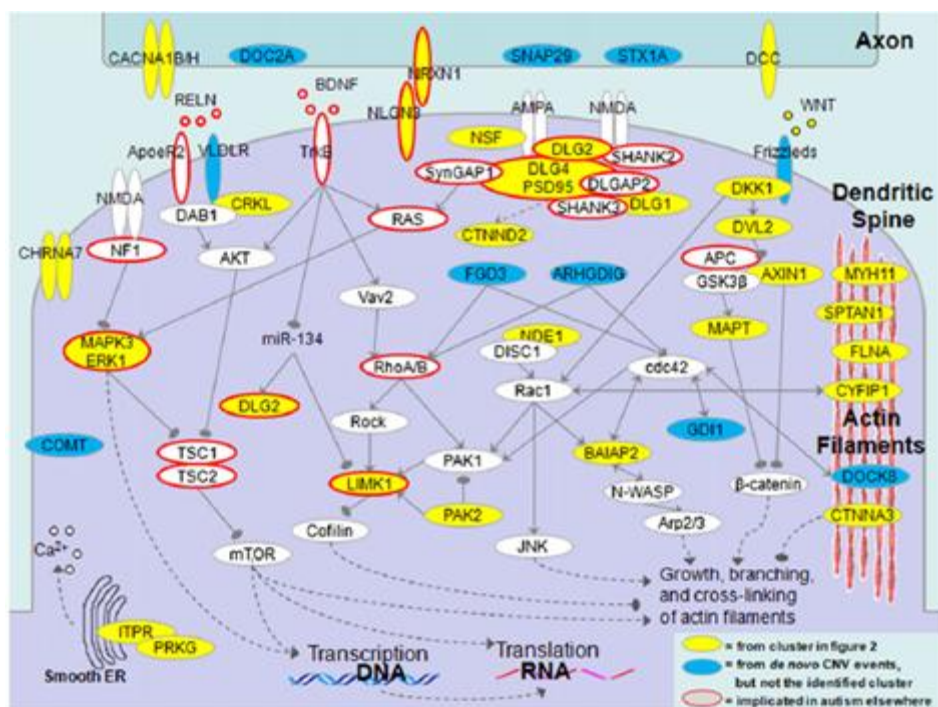


Figura 2.14: Geni associati con la morfogenesi delle spine dendritiche. Le proteine mostrate in figura giocano un ruolo cruciale nella formazione dei contatti fisici tra gli assoni e i dendriti, nell'organizzazione della densità post-sinaptica (PSD) e nei processi di segnalazione che controllano la morfologia delle spine dendritiche. Molti dei pathway convergono sulla regolazione della crescita e della ramificazione dei filamenti di actina, essenziali nel rimodellamento strutturale delle spine. Le proteine codificate da geni colpiti da CNV *de novo* associati al fenotipo autistico sono colorate in giallo, mentre altri geni colpiti da CNV *de novo* identificati nello studio condotto da Levy sono indicati in blu; i geni già implicati nell'autismo sono cerchiati invece in rosso. (Sarah R. Gilman, et al, 2011)

2.6 Caratteri monogenici e varianti rare associati all'autismo

In una minoranza di casi (<10%) l'autismo è associato a patologie causate da un singolo gene, tra cui la sindrome dell'X fragile (FRAXA), la sclerosi tuberosa (TSC),

la neurofibromatosi [Klauck, 2006] o a varie anomalie cromosomiche [Smalley et al. 1998; Dykens and Volkmar, 1997; Feinstein and Reiss, 1998], che colpiscono geni coinvolti nella funzione sinaptica, quali le neuroligine, le neurexine e il gene *SHANK3*.

La ***Sindrome dell'X-fragile*** è un disordine caratterizzato da ritardo mentale e con un'incidenza maggiore in individui di sesso maschile. E' causata dalla perdita di funzione della proteina FMRP1, la quale, legandosi ai poliribosomi localizzati a livello delle spine dendritiche, agisce da repressore trascrizionale di specifici mRNA target. La riduzione del prodotto proteico è associata all'ipermetilazione della tripletta CGG altamente ripetuta (>200 *repeats*), presente nella regione non tradotta al 5' del gene *FMR1* localizzato sul cromosoma Xq27 [Oostra et al., 2001; Sutcliff et al., 1992]. Il 2-5% di bambini con una diagnosi di autismo più lieve sono portatori di mutazioni associate a FRAXA [Fombonne, 2003; Bailey, 1993; Reddy, 2005; Wassink et al., 2001]. L'analisi di tale disordine permette la comprensione dei meccanismi molecolari alla base del ritardo mentale e dei sintomi sociali e di comunicazione, che si sovrappongono a quelli visti nei disordini autistici [Lord et al., 2000].

La ***Sclerosi Tuberosa*** è una malattia autosomica dominante caratterizzata dalla presenza di tubercoli nella regione corticale e cerebrale e da lesioni renali e attacchi epilettici durante l'infanzia [Bolton et al., 1997; Bolton et al., 2002; Gutierrez et al., 1998; Hunt et al., 1987; Jambaque et al., 2000]. Questo disordine è causato da mutazioni in due diversi geni, *TSC1* e *TSC2*, che mappano rispettivamente nelle regioni cromosomiche 9q34 e 16p13 [Roach et al., 2004]. Studi epidemiologici hanno mostrato che la prevalenza di Sclerosi Tuberosa in bambini autistici e di autismo in casi di Sclerosi Tuberosa è cento volte superiore a quella attesa [Harrison et al., 1997; Fombonne, 2003].

L'importanza di varianti rare nei disturbi dello spettro autistico è stata dimostrata dall'identificazione di alcune mutazioni rare in singoli geni. Le prime sono state individuate nei geni delle neuroligine *NLGN3* (*Neurologin 3*) e *NLGN4* (*Neurologin 4*), localizzati rispettivamente in Xp22.3 e Xq13

Le neuroligine (*NLGN*) rappresentano una famiglia di proteine transmembrana coinvolte nell'adesione cellulare post-sinaptica, sono localizzate sul versante citoplasmatico post-sinaptico delle sinapsi neuroni-neuroni; svolgono un ruolo importante nella sinaptogenesi e nel differenziamento pre-sinaptico di assoni

glutammatergici, poiché agiscono come ligandi sito specifici per le *beta-neurexine*, proteine essenziali per indurre la differenziazione dei neuroni cerebellari e ippocampali attraverso la loro interazione con le neuroligine, e sono coinvolte nei fenomeni di adesione cellulare e di formazione e rimodellamento delle sinapsi del Sistema Nervoso Centrale (Chih et al., 2004; Chubykin et al., 2005; Dean and Dresbach; 2006). Jamain e colleghi furono i primi a notare che due regioni del cromosoma X, Xp22.3 e Xq13, contenenti i geni *NLGN3* e *NLGN4* rispettivamente, erano associate al fenotipo autistico. Lo screening di mutazioni di questi 2 geni, in 36 coppie di fratelli affetti e 122 trios (famiglie composte da genitori e figlio affetto) con autismo o sindrome di Asperger, ha portato all'identificazione di una mutazione *frameshift* (insT 1186) nel gene *NLGN4*, che genera un codone di stop e causa la terminazione prematura della proteina. Questa variante, presente in due fratelli affetti (uno con la sindrome di Asperger, l'altro autistico), era invece assente nel fratello non affetto e nei 350 individui di controllo [Jamain et al., 2005]. In un'altra famiglia è stata individuata una sostituzione dissenso (C/T) nel gene *NLGN3*, in corrispondenza di un dominio funzionale estremamente conservato della proteina, trasmessa dalla madre ai due fratelli affetti ed assente nei 200 controlli. Questa mutazione causa la sostituzione dell'arginina 451, altamente conservata, con una cisteina (R451C) e influisce sulla localizzazione della proteina sulla superficie cellulare e sulle interazioni con altre proteine. La caratterizzazione di topi con questa mutazione in *Nlg3* [Tabuchi et al., 2007] suggerisce che uno squilibrio tra sinapsi eccitatorie ed inibitorie, a favore di quest'ultimo, possa avere effetti rilevanti nella determinazione della malattia. Inoltre, in una famiglia francese che presentava casi di autismo e di ritardo mentale, è stata trovata una delezione di 2bp nell'esone 5 del gene *NLGN4*, che porta alla formazione di un prematuro codone di stop. Tale mutazione segrega con ritardo mentale o con ASD [Laumonier et al., 2004].

Tuttavia, dal momento che questi risultati non sono stati replicati in altri 23 studi, si ritiene che tali geni abbiano un ruolo nella suscettibilità all'autismo solo raramente, in circa lo 0,4% dei casi. Questo potrebbe essere dovuto al fatto che tali geni raramente contengono mutazioni con un effetto così forte sulla suscettibilità all'autismo. Le neuroligine codificano per molecole di adesione cellulare localizzate nella membrana post-sinaptica e inducono il differenziamento pre-sinaptico di assoni glutammatergici [Graf et al., 2004]. Dal momento che le neuroligine legano le β -neurexine (*NRXN1*, *NRXN2*, *NRXN3*), è stato successivamente realizzato anche uno screening per

mutazioni nei tre geni delle neurexine, utilizzando 72 individui autistici [Feng *et al.* 2006]. L'identificazione di delezioni de novo e di varianti rare missenso nel gene *NRXN1* di individui autistici, supporta l'ipotesi del coinvolgimento delle neurexine (*NRXN*) nei casi ASD. Tale gene mappa nella regione 2q32 e codifica per una proteina di superficie altamente espressa nel cervello, che interagisce con le neuroligine. L'identificazione di rare varianti di sequenza anche in un altro gene appartenente alla famiglia delle neurexine, *CNTNAP2*, suggerisce che questa famiglia di proteine possa svolgere un ruolo importante nella suscettibilità all'autismo. Le neurexine rappresentano molecole di adesione cellulare pre-sinaptica coinvolte nel differenziamento post-sinaptico delle sinapsi glutammatergiche attraverso il legame con *NLGN3* e *NLGN4*. Dal momento che il glutammato è il principale neurotrasmettitore eccitatorio ed un fattore critico per lo sviluppo del cervello, disfunzioni a livello della neurotrasmissione glutammatergica potrebbero avere un ruolo importante nei disturbi dello spettro autistico.

Infine, mutazioni rare sono state individuate in *SHANK3* (SH3 and multiple Repeat Domains 3), un altro gene coinvolto nelle funzioni sinaptiche, che interagisce con le neuroligine. Questo gene mappa nella regione 22q13.3 e codifica per una proteina scaffold, abbondante nella densità post-sinaptica ed espressa preferenzialmente nella corteccia cerebrale e nel cervelletto [Bonaglia *et al.*, 2001]. Studi molto recenti condotti su individui con caratteristiche fenotipiche riconducibili a quelle dello spettro autistico, quali ritardo mentale e/o problemi nel linguaggio, hanno individuato in corrispondenza di *SHANK3* inserzioni o delezioni parziali e/o totali [Bonaglia *et al.*, 2001; Anderlid *et al.*, 2002; Wilson *et al.*, 2003; Durand *et al.*, 2007] con la conseguenza di un dosaggio anomalo del prodotto proteico corrispondente. Da questi dati emerge che lo sviluppo anomalo delle spine dendritiche sia coinvolto in vari disordini neuropsichiatrici associati al fenotipo ASD, che includono la sclerosi tuberosa, la sindrome di Rett, la sindrome di Angelman e la sindrome dell'X fragile. Inoltre, poiché l'insorgenza del fenotipo autistico si verifica durante il periodo critico di formazione e maturazione delle sinapsi, si ritiene che la disfunzione sinaptica potrebbe avere un ruolo nella patogenesi dell'autismo. Pertanto, varianti rare, che colpiscono i geni *SHANK3*, *NLGN3*, *NLGN4* e *NRX*, potrebbero causare alterazioni nello sviluppo delle sinapsi glutammatergiche nei casi ASD.

2.7 Architettura genetica dei disturbi dello spettro autistico

Per molto tempo, l'autismo è stato considerato una malattia multifattoriale, in cui molteplici varianti di rischio comuni, ciascuna di piccolo effetto, concorrono all'eziologia del disturbo. Tuttavia, molti degli studi di associazione condotti sui geni candidati hanno mostrato segnali di associazione al di sotto della soglia di significatività e spesso i risultati ottenuti non sono stati replicati da altri gruppi, dimostrando che le varianti comuni, ad oggi identificate, rappresentano solo una piccola porzione dei fattori genetici coinvolti nei disturbi ASDs. Sembra quindi che un ruolo preponderante nella suscettibilità al fenotipo autistico sia svolto da un insieme di varianti rare ed eterogenee, sia puntiformi che strutturali (CNVs), che singolarmente sono implicate in un piccolissimo numero di casi, ma che complessivamente potrebbero spiegare una buona parte dell'ereditabilità alla base della malattia. Tali varianti alterano molteplici geni con funzioni diverse, che tuttavia potrebbero agire a livello di pathways comuni (come, ad esempio, lo sviluppo e la funzione dei circuiti neuronali). Pertanto, attualmente i ricercatori mirano a sviluppare network di interazione proteica che potrebbero aiutare a comprendere meglio il ruolo all'interno del sistema nervoso centrale dei geni potenzialmente implicati nella patogenesi dell'autismo.

Tuttavia, anche se gli studi condotti negli ultimi anni hanno sottolineato il coinvolgimento di varianti rare e CNVs *de novo* nella suscettibilità ai disturbi dello spettro autistico [Zhao et al., 2007; Pinto et al., 2010] queste mutazioni, ad oggi, sono in grado di spiegare solo il 5-10% dei casi. Quindi, data l'elevata eterogeneità genetica e fenotipica dell'autismo, un ruolo causale delle mutazioni *de novo* non esclude completamente l'importanza anche della variazione comune nella suscettibilità al fenotipo ASD [Gillis and Rouleau, 2011].

3. il Ritardo Mentale

Il ritardo mentale (MR) o disabilità intellettiva (ID), insorge entro i 18 anni di età, ed è caratterizzato da uno sviluppo mentale incompleto, che induce significative limitazioni nelle funzioni cerebrali e nel comportamento adattativo, con una compromissione delle capacità pratiche, concettuali, cognitive, comunicative, motorie, affettive e sociali (American Psychiatric Association 1994).

La diagnosi del ritardo mentale è effettuata mediante test standardizzati, basati sull'età, che misurano il quoziente intellettivo, come, ad esempio, il Wechsler Intelligence Scale for Children (WISC) o for Adults (WAIS).

Sulla base del quoziente intellettivo (IQ) vengono distinte, tra i vari gradi di severità, due categorie principali di ritardo mentale: RM lieve con un QI compreso tra 50 e 70 e RM grave con un QI inferiore a 50.

Convenzionalmente, le forme ereditarie di ritardo mentale vengono suddivise in due classi principali: RM sindromico, caratterizzato dall'associazione di altre manifestazioni cliniche e RM non sindromico o non specifico, in cui il deficit cognitivo rappresenta l'unica manifestazione della patologia.

Sebbene questa distinzione sia molto utile per scopi clinici, recenti studi stanno indicando che il confine tra le forme di MR sindromiche e non sindromiche sta scomparendo, e alcune forme non sindromiche possono, successivamente, essere riconosciute come forme sindromiche (Frints SG et al. 2002; Chelly et al., 2006). Infatti, in diversi casi, distinte mutazioni sullo stesso gene possono determinare l'uno o l'altro tipo di ritardo mentale, a seconda del loro effetto sull'attività della proteina (es. perdita completa o parziale della funzione), del pattern di inattivazione dei geni X-linked nelle femmine affette o del background genetico (Vaillend et al., 2008).

Un esempio di gene coinvolto sia nelle forme sindromiche che non sindromiche di MR è il gene *RSK2* (*RPS6KA3*): la mutazione missenso R383W causa ritardo mentale X-linked (MRX), dovuto ad una parziale perdita di funzione dell'allele, che codifica una proteina con il 20% dell'attività della chinasi wild-type; al contrario la mutazione nulla di *RSK2* causa la syndrome di Coffin-Lowry, con un prominente coinvolgimento del tessuto connettivo e scheletrico. (Jennifer K. Inlow et Linda L. Restifo, 2003). Un altro esempio è dato dal gene *ATRX* (anche noto come *XH2* o *XNP*) che codifica per una proteina nucleare coinvolta nella struttura della cromatina

e nella regolazione della trascrizione. Infatti, la mutazione T1621M del gene causa MR non specifico lieve o moderato, mentre mutazioni missenso a 7 o 12 residui a monte, causano un fenotipo sindromico più grave, con difetti ematologici, scheletrici e genitali, suggerendo una maggiore distruzione della funzione di ATRX. (Jennifer K. Inlow et Linda L. Restifo, 2003)

Le relazioni genotipo-fenotipo sono ancora più, complesse per *MECP2* e *ARX*. Le femmine con mutazioni in *MECP2* mostrano una grande eterogeneità clinica, sia intra- che inter-familiare, che è probabilmente dovuta a diversi livelli di inattivazione dell'X e diversi siti di mutazione. Infatti, almeno 7 differenti mutazioni missenso in *MECP2*, sparse lungo tutta la proteina, causano MR non specifico; alcune di esse sono molto vicine ai siti delle mutazioni missenso che causano la sindrome di Rett. Per *ARX*, identiche mutazioni, che portano all'espansione del tratto di poli-alanine nell'omeodominio di questa proteina, possono causare in una famiglia MR non specifico e diverse sindromi neurologiche in altre famiglie. Questo suggerisce un possibile effetto del *background* genetico sui fenotipi *ARX*. (Jennifer K. Inlow et Linda L. Restifo, 2003)

Nei paesi sviluppati, MR rappresenta la causa più frequente di grave handicap nei bambini ed è una delle ragioni principali per cui si richiede una consulenza genetica, poiché le cause genetiche possono tener conto del 25–50% dei casi con MR grave.

Si stima che l'incidenza del ritardo mentale, nei paesi occidentali, sia generalmente dello 0.3–0.5% per il MR moderato o grave e dello 1–3% quando si considera anche il MR lieve, ma tende ad aumentare nelle nazioni in via di sviluppo, sia a causa di fattori ambientali, quali la malnutrizione, e le precarie condizioni di assistenza sanitaria, sia per l'azione di fattori eziologici a rilevanza genetica, come la consanguineità parentale (Chelly et al., 2006; Ropers HH, 2008).

Terapie comportamentali e cognitive possono aiutare i pazienti affetti da ritardo mentale a raggiungere il loro massimo potenziale (Bathae 2001; Butler *et al.* 2001), ma non sono curative e spesso mirano al trattamento delle manifestazioni cliniche più comuni che possono accompagnare MR come l'aggressività e i comportamenti autolesionisti (Long and Miltenberger 1998; Dosen and Day 2001).

3.1 Eziologia

Il ritardo mentale può essere causato da fattori genetici e/o ambientali, che agiscono sullo sviluppo e sul funzionamento del sistema nervoso centrale in età prenatale,

perinatale e postnatale; ma una causa precisa è stata trovata solo in circa il 50% dei casi con MR da moderato a grave, e in una percentuale ancora più bassa negli individui con MR lieve.

I fattori genetici includono aneuploidie cromosomiche, anomalie strutturali cromosomiche, disordini genomici e malattie monogeniche. Tali fattori incidono in circa il 30% dei ritardi moderati, mentre i fattori ambientali (malnutrizione durante la gravidanza, neurotossicità ambientale, nascita prematura, ischemia cerebrale perinatale, sindrome fetto-alcolica FAS, infezioni pre- o post-natali) giocano un ruolo importante nella patogenesi del ritardo mentale lieve.

I fattori prenatali (principalmente genetici, ma imputabili anche a problemi di crescita intrauterina, ad infezioni fetali, o all'assunzione di droga o alcol da parte della madre) sono i più frequenti e sono responsabili del 60-75% dei casi, mentre le cause perinatali (danni cerebrali da ipossia, infezioni) e postnatali (danni cerebrali di qualsiasi origine, specialmente da meningiti, encefaliti e traumi) sono meno comuni, anche se più suscettibili ad una prevenzione efficace (Aicardi, 1998).

Dal punto di vista eziologico, il 60-75% dei casi presenta MR grave, mentre solo il 38-55% ha MR lievi.

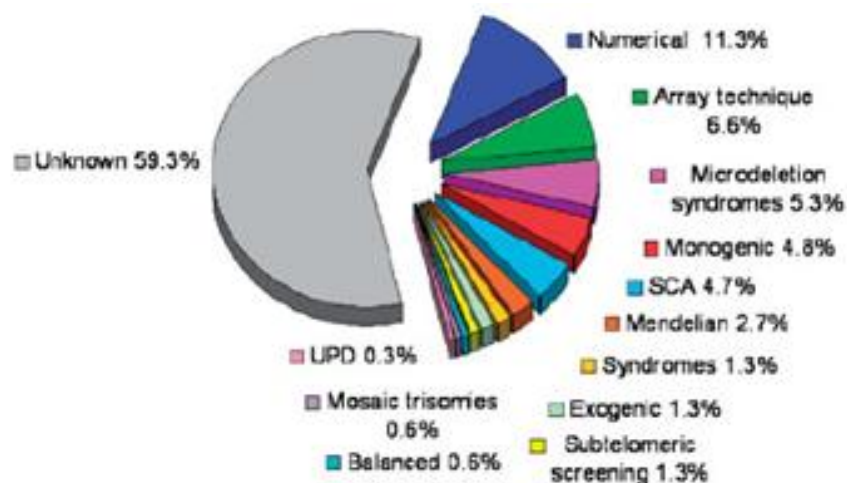


Figura 3.1: distribuzione di varie cause eziologiche raggruppate nelle categorie diagnostiche in gruppi di individui con ritardo nello sviluppo/ritardo mentale. Sca, anomalie cromosomiche segmentali. (Rauch et al, 2006)

3.2 Cause genetiche

Il ritardo mentale ha una penetranza ed un'espressività variabile. A causa della sua complessità genetica, ancora poco si sa sui fattori genetici di predisposizione al ritardo mentale lieve.

Si stima che le cause genetiche, responsabili del 25-50% dei casi di RM grave (McLaren and Bryson 1987), possono essere suddivise in :

- riarrangiamenti cromosomici, che determinano essenzialmente alterazioni deleterie del dosaggio genico e sono responsabili del 30% dei casi (considerando sia le anomalie cromosomiche citogeneticamente visibili che le anomalie strutturali submicroscopiche),
- alterata regolazione dell'imprinting di specifici geni o di regioni genomiche
- disfunzione di singoli geni importanti per lo sviluppo delle funzioni cognitive.

Oltre alle convenzionali aneuploidie cromosomiche, come la trisomia 21, che tiene conto di circa 1.2/1000 nati vivi, prove inoppugnabili stanno suggerendo che riarrangiamenti subtelomerici, nell'insieme, possono tener conto del 5-7% delle forme sindromiche di MR.

Per le cause monogeniche, sono stati trovati numerosi geni localizzati sul cromosoma X mutati in circa il 10% dei maschi affetti, ma un ruolo importante è svolto anche da geni autosomici, le cui alterazioni possono dare origine a casi isolati di forme dominanti o a forme recessive che solitamente si presentano come casi sporadici nelle piccole famiglie tipiche dei paesi industrializzati (Chelly, 2006; Ropers HH, 2008).

3.2.1 Aberrazioni cromosomiche

L'incidenza totale delle anomalie cromosomiche fra i nati vivi è di circa lo 0,3%, per quanto riguarda quelle che comportano un effetto fenotipico; a ciò va aggiunto uno 0,2% di portatori di traslocazioni bilanciate. Pertanto lo 0,5% dei nati vivi presenta un cariotipo anomalo.

L'effetto più comune di una anomalia cromosomica è il deficit cognitivo, ma frequentemente sono associate anche malformazioni cardiache e caratteristiche dismorfiche. In generale, uno sbilanciamento genico dovuto ad un'anomalia cromosomica provoca sempre un danno nello sviluppo. Se lo sbilanciamento è esteso,

questo può compromettere lo sviluppo embrionale (con conseguente aborto spontaneo) oppure determinare la nascita di bambini con gravi malformazioni, generalmente associate a ritardo mentale. Se esso è modesto, come nel caso di delezioni e duplicazioni di lieve entità, potranno non essere presenti malformazioni gravi e quadri dismorfici accentuati, mentre è probabile un ritardo psicomotorio, anche se limitato, anche se sono state identificate anche varianti prive di significato clinico (Raymond e Tarpey, 2006).

Nonostante il largo impiego di diagnosi prenatale soprattutto nelle donne di età avanzata, la sindrome di Down rimane la causa più comune di ritardo mentale (1:650 nati; è responsabile del 9.2% dei casi con ritardo mentale ed anomalie congenite multiple) e, complessivamente, tutte le alterazioni cromosomiche di numero e di struttura citogeneticamente visibili vengono riscontrate in un individuo affetto da deficit cognitivo grave su sette (Ropers HH, 2008).

Aberrazioni visibili citogeneticamente sono responsabili di circa il 15% di tutti i pazienti con MR grave, ma delezioni e duplicazioni microscopiche, non identificabili al microscopio, stanno emergendo come una causa altrettanto importante di MR (Ropers, 2010).

3.2.2 CNVE MR

Nei primi anni del 1990, sono state identificate piccole microdelezioni mediante l'utilizzo della tecnica di ibridazione fluorescente in situ (FISH); queste microdelezioni sono state, successivamente, associate a particolari sindromi: le sindromi di Wolf- Hirshhorn, Cri du Chat, Williams, Prader-Willi, Angelman, Rubinstein-Taybi, lissencefalia di Miller-Dieker, Smith-Magenis, Alagille, DiGeorge e la sindrome della delezione 22q11. Sebbene ciascuna sindrome ha caratteristiche distintive sono tutte associate a ritardo mentale di grado variabile. (F. Lucy Raymond et Patrick Tarpey, 2006) In particolare è stato ampiamente accertato che i riarrangiamenti telomerici, nell'insieme, siano responsabili del 5–7% di tutti i casi di MR e sono state riportate, in pazienti con ritardo mentale, delezioni nella maggior parte, ma non in tutte le estremità dei cromosomi.

Grazie allo sviluppo delle nuove tecnologie basate sui microarray, è aumentato il numero dei disordini genomici noti, ossia quelle patologie causate da variazioni del numero di copie di specifici segmenti genomici, dette copy number variants (CNVs).

Molte di queste CNV sono state trovate associate a specifici fenotipi, ed in pazienti con ritardo mentale sono stati identificati microdelezioni e duplicazioni interstiziali ricorrenti che coinvolgono numerosi intervalli genomici associati. Per alcuni di questi loci, sono state osservate delezioni e duplicazioni reciproche. (Mefford et al, 2012).

Table 1. Novel Recurrent Copy-Number Changes Associated with Intellectual Disability and Related Disorders.*

Chromosome Region	Coordinates in Mb†	Deletion or Duplication Associated with Disorder	Selected References
1q21.1	Chromosome 1: 145.0–146.35	Deletion: intellectual disability, schizophrenia, multiple congenital anomalies Duplication: intellectual disability, autism	Brunetti-Pierri et al., ³² Mefford et al., ³³ International Schizophrenia Consortium, ³⁴ Stefansson et al., ³⁵ Greenway et al., ³⁶ Haldeman-Englert and Jewett ³⁷
3q29	Chromosome 3: 197.4–198.9	Deletion: intellectual disability, schizophrenia Duplication: intellectual disability	Ballif et al., ³⁸ Lisi et al., ³⁹ Willatt et al. ⁴⁰
10q22-q23	Chromosome 10: 81.12–89.07	Deletion: intellectual disability	Balciuniene et al., ⁴¹ van Bon et al. ⁴²
15q11.2	Chromosome 15: 20.3–20.7	Deletion: intellectual disability, schizophrenia, epilepsy	Stefansson et al., ³⁵ de Kovel et al., ⁴³ Mefford et al., ⁴⁴ Burnside et al., ⁴⁵ Doornbos et al., ⁴⁶ Murthy et al., ⁴⁷ von der Lippe et al. ⁴⁸
15q13.3	Chromosome 15: 28.7–30.2	Deletion: intellectual disability, epilepsy, schizophrenia, autism	Stefansson et al., ³⁵ Helbig et al., ⁴⁹ Sharp et al., ⁵⁰ van Bon et al., ⁵¹ Ben-Shachar et al., ⁵² Pagnamenta et al., ⁵³ Miller et al. ⁵⁴
15q24	Chromosome 15: 72.2–73.8	Deletion: intellectual disability, autism	Andrieux et al., ⁵⁵ Sharp et al., ⁵⁶ Mefford et al., ⁵⁷ El-Hattab et al. ⁵⁸
16p11.2 (a)	Chromosome 16: 29.5–30.1	Deletion: intellectual disability, autism, obesity Duplication: schizophrenia	Weiss et al., ²⁹ Battaglia et al., ⁵⁹ Bijlsma et al., ⁶⁰ Hempel et al., ⁶¹ Shinawi et al., ⁶² Jacquemont et al., ⁶³ Walters et al., ⁶⁴ McCarthy et al. ⁶⁵
16p11.2 (b)	Chromosome 16: 28.7–29.0	Deletion: intellectual disability, obesity	Bachmann-Gagescu et al., ⁶⁶ Bochukova et al. ⁶⁷
16p12	Chromosome 16: 21.8–22.4	Deletion: intellectual disability	Girirajan et al. ⁶⁸
16p13.11	Chromosome 16: 15.4–16.4	Deletion: intellectual disability, epilepsy, autism, schizophrenia Duplication: intellectual disability, ADHD, autism	de Kovel et al., ⁴³ Mefford et al., ⁴⁴ Heinzen et al., ⁶⁹ Williams et al., ⁷⁰ Ullmann et al., ⁷¹ Kirov et al. ⁷²
17q12	Chromosome 17: 31.8–33.3	Deletion: intellectual disability, autism, schizophrenia	Moreno-De-Luca et al., ⁷³ Loirat et al. ⁷⁴
17q21.3	Chromosome 17: 41.0–41.7	Deletion: intellectual disability	Koolen et al., ²⁰ Sharp et al., ²³ Shaw-Smith et al., ²⁴ Koolen et al. ⁷⁵

In uno dei più grandi studi pubblicati fino ad oggi, condotto su 1010 bambini con ID idiopatico ed 2493 adulti sani, sono state identificate CNV patogene e ipoteticamente patogene in 5.4% dei pazienti. (Mefford et al.,2009). In questo studio, tra i cambiamenti considerati patogenetici è riportata una delezione nella regione cromosomica 16p11.2, con una frequenza dello 0.7%; questa stessa variante era stata precedentemente descritta nell'autismo, con una frequenza simile.(Weiss et al, 2008) Questi risultati supportano la teoria che alcune CNV abbiano un effetto pleiotropico, essendo implicate nella suscettibilità a diversi fenotipi come il ritardo mentale e l'autismo.

In base ai risultati di Mefford e collaboratori, negli individui con handicap intellettuali, la frequenza cumulativa delle delezioni e duplicazioni è circa 1.4%, che li rende comuni quanto la sindrome dell'X fragile. (Mefford et al.,2009)

Questi ed altri studi illustrano che le CNV sono la causa fondamentale di alcune forme specifiche di ritardo mentale e di altri disordini.

Molte nuove microdelezioni sono state identificate in pazienti che hanno un quadro clinico riconoscibile. Ad esempio la sindrome da microdelezione del 17q21.31, descritta simultaneamente da tre gruppi, è stata associata a MR, ipotonia, caratteristici tratti facciali, epilessia, difetti del cuore ed anomalie della pelle/urologiche ed ha una frequenza di 1 su 16000 persone. (Vissers et al, 2012) Questa delezione ha una dimensione di 500-650 kb, per cui non è identificabile mediante il cariotipo normale, e tutte le delezioni identificate nella regione 17q21.31 sono de novo, e non sono mai state trovate in individui di controllo sani.

Un altro esempio di sindrome da microdelezione ben riconoscibile clinicamente è la microdelezione 15q24. Infatti, mediante uno screening di circa 1200 individui con MR idiopatico, sono stati identificati quattro individui con delezioni submicroscopiche che si sovrappongono alla regione 15q24. Essi condividono varie caratteristiche cliniche, che comprendono MR, ritardo della crescita, microcefalia, anomalie digitali, ipospadia, perdita dei tessuti connettivi e caratteristici tratti facciali indicando che anche delezioni della regione 15q24 rappresentano una sindrome clinica. (Lisenka E L M Vissers,¹ Bert B A de Vries,^{1,2} Joris A Veltman¹;2012)

Al contrario alcune microdelezioni e duplicazioni ricorrenti sono state associate ad un'ampia varietà di fenotipi. Esempi di questo tipo sono stati riportati per le CNVs delle regioni 1q21.1, 15q13.3, e 16p13.11. Infatti, la delezione interna alla regione 1q21.1 è stata inizialmente identificata in individui affetti da MR non spiegato, autismo e altre anomalie congenite. Il fenotipo variava considerevolmente e comprendeva MR da lieve a moderato, microcefalia, anomalie cardiache e cataratta. Inoltre, sebbene siano stati identificati diversi portatori della delezione non affetti, l'alta frequenza delle delezioni 1q21.1 in persone affette da schizofrenia ha suggerito un ruolo di questa CNV nei disordini psichiatrici. Infine, la microduplicazione reciproca che coinvolge la regione 1q21.1 è stata associata all'autismo o a comportamenti autistici, e le caratteristiche fenotipiche comuni dei portatori delle duplicazioni 1q21.1 comprendevano MR da lieve a moderato, macrocefalia o microcefalia relativa, e lievi caratteristiche dismorfiche. Anche questa duplicazione è stata identificata in individui apparentemente normali. (Vissers et al, 2012)

3.2.3 Forme di ID del cromosoma X

Le forme di ritardo mentale legate all’X (XLID) sono facilmente identificabili grazie al loro caratteristico modello di ereditarietà, in cui il fenotipo recessivo si manifesta nei maschi emizigoti, ed hanno un’elevata prevalenza nella popolazione, mantenuta dalla presenza di femmine portatrici sane, la cui fitness riproduttiva è conservata (Raymond e Tarpey, 2006; Ropers HH, 2008; Inlow e Restifo, 2004)).

Le forme di ritardo mentale legato all’X, XLID, sono generalmente suddivise in forme sindromiche (S-XLID) e forme non-sindromiche (NS-XLID). (Ropers, 2010)

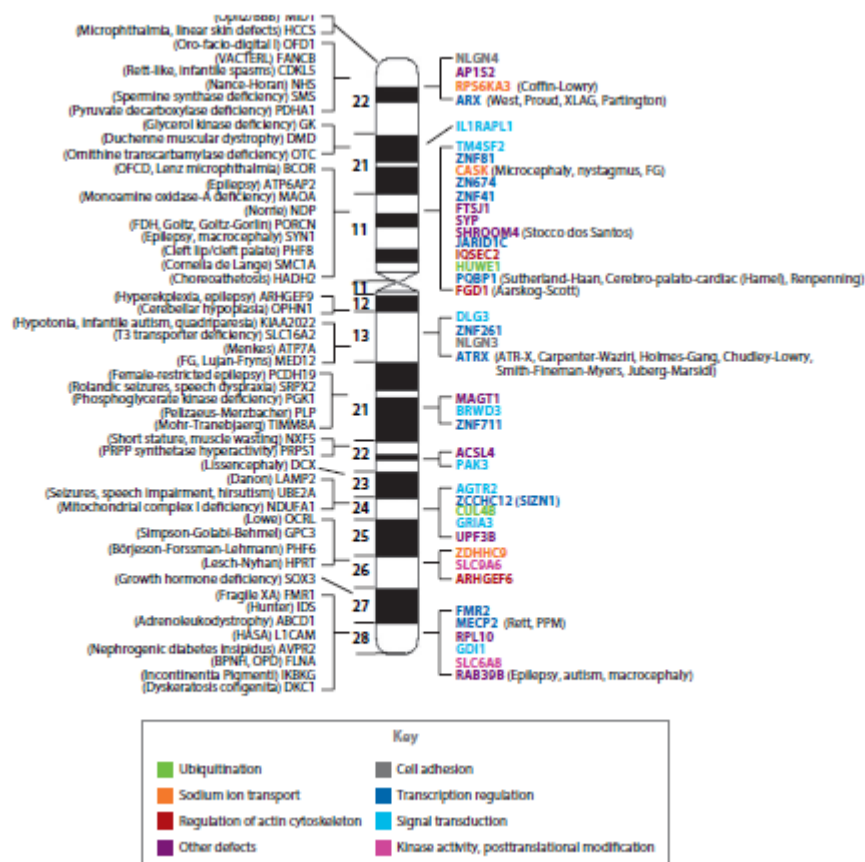


Figura 3.2: geni implicati nelle forme sindromiche e non sindromiche nel ritardo mentale X linked e la loro posizione sul cromosoma X umano. Le differenti classi funzionali sono colorate (Ropers, 2010).

Il cromosoma X umano contiene circa il 4% dei geni che codificano proteine, ma si pensa che difetti del cromosoma X siano responsabili dell’8-12% del MR osservato nei maschi. Nei maschi il ritardo mentale moderato o grave è circa 1.4 volte più frequente rispetto alle femmine e questa discrepanza non può essere ascritta solo ai difetti dei geni X-linked. Per spiegare questa differenza, sono state proposte varie

ipotesi che comprendono l'imprinting differenziale dei cromosomi X materni e paterni, l'inattivazione incompleta dell'X nelle femmine, che risulta in una più alta attività di alcuni geni del cromosoma X; effetti regolatori del cromosoma Y, sia indirettamente, attraverso meccanismi di controllo ormonale, che direttamente, tramite la regolazione della trascrizione mediata da i geni del cromosoma Y. (Ropers, 2010)

Infatti, a supporto dell'ipotesi che i geni del cromosoma Y agiscano come regolatori trascrizionali ci sono le recenti scoperte che un enzima chiave nel metabolismo della serotonina (la monoammina dell'ossidasi A, MAOA), con un importante ruolo nel comportamento, è direttamente sovra-regolato da SRY, il gene del cromosoma Y che controlla il differenziamento dei testicoli. Questi meccanismi, da soli o insieme, sono probabilmente responsabili della maggioranza reale o apparente dei maschi non solo nel ritardo mentale ma anche nell'autismo, dove i rapporti riportati maschio-femmina sono 4:1 (Ropers, 2010)

Ad oggi sono stati identificati più di 80 geni coinvolti nel ritardo mentale legato all'X, e di questi circa 30 geni sono stati trovati nelle forme "pure" non sindromiche dell'ID, anche se non c'è una netta separazione tra le due forme. (Ropers, 2010). Ad esempio mutazioni nel gene *OPHN1*, uno dei primi geni implicati in NS-XLID (Billuart P et al, 1998), sono poi stati trovati associati ad una sindrome che comprende atassia, epilessia, ed ipoplasia cerebrale. (Bergmann C et al,2003; des Portes V et al,2004).

Studi indipendenti hanno rivelato la presenza di CNV in circa il 10% di famiglie con XLID. (Ropers, 2010) (BautersM et al, 2008; Lugtenberg D et al.,2009)

Dopo la sindrome dell'X-fragile, responsabile del 25% delle famiglie con XLID, le mutazioni nel gene *ARX* si collocano al secondo posto, responsabili del 5% di forme sindromiche e non sindromiche di ID. Le mutazioni nei geni *CUL4B*, *JAD1C* e *SLC6A8* sono tutte relativamente frequenti, e ciascuna è responsabile del 2-3% delle famiglie, mentre i difetti in tutti gli altri geni X-linked noti sembrano essere significativamente meno comuni dell'1%.

Le mutazioni nei geni che regolano l'attività di piccole Rho e Ras GTPasi, sono tra le prime identificate in pazienti con le forme di XLID sia sindromiche che non sindromiche, si pensa che questi geni giochino un ruolo importante nello sviluppo dei neuroni e nella formazione e funzione delle sinapsi. (Billuart P et al,1998; KutscheK et al,2000) I geni implicati nell'XLID giocano un ruolo anche in altri processi cellulari tra cui l'adesione cellulare, la regolazione della trascrizione, la modificazione

post-traduzionali, la trasduzione del segnale, l'ubiquitinazione, la modificazione dell'actina del citoscheletro, e il trasporto del sodio.

3.2.4 Forme autosomiche di MR

Le forme autosomiche dominanti gravi di ritardo mentale (ADID) sono per lo più dovute a mutazioni *de novo*, poiché la fitness riproduttiva degli individui affetti da questa forma di ritardo mentale è bassa, tuttavia, la presenza di numerose CNV *de novo* in molti pazienti affetti da ritardo mentale idiopatico indica che le forme ADID non sono rare.

In un recente studio, è stato dimostrato che mutazioni puntiformi *de novo* possono contribuire sostanzialmente al ritardo mentale (Vissers et al, 2010). Il sequenziamento dell'esoma di dieci individui con ritardo mentale idiopatico e dei relativi genitori, ha portato all'identificazione di nove mutazioni *de novo* non sinonime, sei delle quali potenzialmente dannose, sulla base della funzione genica, della conservazione nell'evoluzione e del tipo di mutazione. Questo studio fornisce forti evidenze sperimentali a favore di un "paradigma *de novo*" per il ritardo mentale e suggerisce che mutazioni *de novo* potrebbero avere un ruolo prominente anche in casi sporadici di altri disordini neuropsichiatrici, come l'autismo e la schizofrenia.

Inoltre, vari disordini autosomici dominanti ben conosciuti, come la neurofibromatosi (con una incidenza di circa 1/4000), la sclerosi tuberosa (circa 1/6000), e la distrofia miotonica (circa 1/8000), sono spesso associati a ID di gravità variabile. (Nelson B, 2009) Per identificare nuovi geni autosomici dominanti, responsabili di ritardo mentale, sono stati raggruppati ed analizzati insieme individui affetti da ritardo mentale con distinte caratteristiche dismorfiche. Condizioni come la sindrome di Smith-Magenis e Rubinstein-Taybi sono state identificate in questo modo e sono stati identificati i geni che causano il fenotipo, mediante mappaggio sistematico delle delezioni e analisi di mutazioni puntiformi (F. Lucy Raymond et Patrick Tarpey, 2006)

Al contrario, si sa ancora molto poco sui difetti genetici responsabili ADID non sindromico.

Considerazioni funzionali e dati empirici ottenuti da esperimenti condotti su modelli di topo suggeriscono che la maggior parte dei difetti genetici sono ereditati come disordini recessivi.

La strategia migliore per spiegare i difetti molecolari responsabili delle forme autosomiche recessive di ritardo mentale (ARID) è il mappaggio dell'omozigosità in grandi famiglie di consanguinei, seguito dallo screening di geni candidati. Tali famiglie sono comuni nei paesi che si estendono dal Marocco all'India, mentre, nelle società Occidentali, le famiglie sono generalmente piccole, e la maggior parte dei pazienti con ARID sono casi isolati. Recentemente in uno studio condotto da Najmabadi e collaboratori (2011) su un campione di 136 grandi famiglie Iraniane di consanguinei con ARID (ritardo mentale non sindromico, autosomico recessivo), è stata dimostrata l'efficacia dell'applicazione combinata di due strategie: il mappaggio di regioni in omozigosi in grandi famiglie di consanguinei e il sequenziamento di tutte le regioni codificanti del genoma (il così detto "esoma"). Infatti, anziché sequenziare l'intero esoma delle famiglie di consanguinei, si sono focalizzati sugli esoni delle regioni in omozigosi, che sicuramente contengono la mutazione responsabile della malattia. Mediante questo approccio, oltre a 23 varianti note per essere coinvolte nel ritardo mentale o in altri disordini neurobiologici, sono stati identificati 50 nuovi geni candidati per il ritardo mentale e sono state trovate singole mutazioni causali in 78 su 136 famiglie analizzate. Alcuni di questi nuovi geni identificati da questo studio codificano per funzioni specifiche dei neuroni o del cervello, come il differenziamento delle cellule gliali o la trasmissione sinaptica, ma la maggior parte di essi hanno una funzione "housekeeping" essendo coinvolti soprattutto nelle funzioni istoniche, nella trascrizione e traduzione. Le proteine codificate da molti di questi geni interagiscono direttamente con i prodotti di geni notoriamente implicati nel ritardo mentale, e molti sono coinvolti in processi cellulari fondamentali come la traduzione e la trascrizione, il controllo del ciclo cellulare, il metabolismo energetico e la sintesi degli acidi grassi, che sembrano essere essenziali per il normale sviluppo e funzionamento del cervello. Solo in uno dei nuovi geni ARID identificati sono state trovate due mutazioni in due diversi nuclei familiari (*ZNF526*), confermando la natura estremamente eterogenea del ritardo mentale. Sebbene alcune delle varianti identificate potrebbero essere dei falsi "positivi" è incoraggiante che molti dei nuovi geni candidati interagiscono con geni coinvolti nel MR. (Hossein Najmabadi et al, 2011).

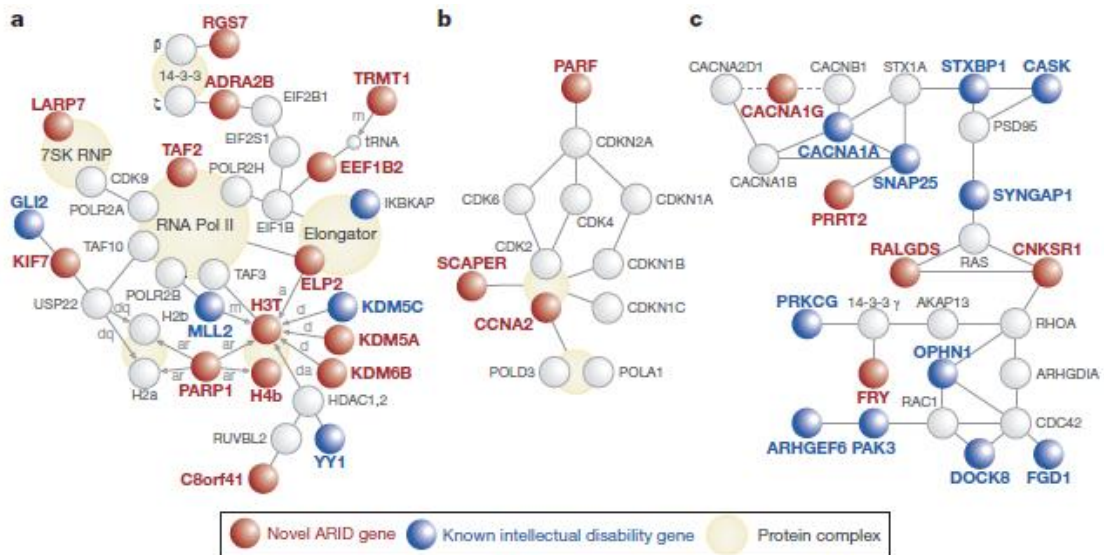


Figura 3.3: geni coinvolti nel ritardo mentale identificati nello studio di Najmabadi e collaboratori, 2011.

3.2.5 Alterata regolazione di geni soggetti ad imprinting

Il termine imprinting genomico descrive l'espressione preferenziale o esclusiva di un allele a seconda della sua origine parentale. L'espressione allele-specifica dei geni soggetti ad imprinting viene determinata da specifiche modificazioni epigenetiche, come la metilazione delle citosine e l'acetilazione o la metilazione degli istoni

Nelle cellule germinali, queste modificazioni epigenetiche sono eliminate, successivamente stabilite di nuovo in maniera genitore-specifica e mantenute dopo la fecondazione.

Ad oggi si ritiene che circa l'1% dei geni nel genoma siano sottoposti a questo tipo di regolazione, con 50 geni identificati e 150 predetti in base alle caratteristiche di sequenza (Amor e Halliday, 2008). Questi geni sono raggruppati in clusters e un chiaro effetto di imprinting è stato descritto per determinate regioni sui cromosomi 7, 11, 14 e 15 (Chelly et al., 2006).

L'alterata regolazione dei geni soggetti ad imprinting è stata osservata in numerose malattie umane, comprese le sindromi caratterizzate da disfunzione cerebrale e deficit cognitivo.

L'alterata espressione dei geni soggetti imprinting può essere dovuta a due tipi di meccanismi:

- di tipo “genetico” : delezioni/duplicazioni e mutazioni che coinvolgono geni sottoposti ad imprinting o i centri di regolazione dell’imprinting e la disomia uni parentale (UDP) , o
- di tipo “epigenetico” : alterazioni delle modificazioni post-traduzionali degli istoni o della metilazione del DNA (Amor e Halliday, 2008).

Tra le sindromi determinate da difetti di imprinting, le più conosciute sono quelle di Angelman e Prader-Willi, causate da geni diversi che si trovano nella stessa regione cromosomica (15q11.2q13), e dovute, spesso in seguito a delezioni, alla perdita dell’espressione dell’allele materno, nella sindrome di Angelman, e di quello paterno, nella sindrome di Prader-Willi (Chelly et al., 2006).

3.3 Funzioni biologiche e molecolari dei geni coinvolti nel ritardo mentale.

3.3 Funzioni biologiche e molecolari dei geni coinvolti nel ritardo mentale.

In molti casi, il ritardo mentale fa parte di una sindrome complessa che include anomalie cerebrali come microcefalia, lissencefalia (anomalia dello sviluppo della corteccia cerebrale), eterotopia dei neuroni (presenza dei neuroni in posizione anomala), agenesia (assenza di un organo), polimicrogyri (presenza di numerose circonvoluzioni, gyri, sulla superficie del cervello) e schizencefalia (rara malformazione della materia grigia del cervello). In questi casi il ritardo mentale è probabilmente un sintomo secondario, ed i geni coinvolti possono essere considerati come fattori richiesti per lo sviluppo normale del sistema nervoso centrale (SNC). Al contrario, nelle malattie in cui il ritardo mentale è associato ad una struttura ed architettura normale del cervello, ci sono, probabilmente, sottili anomalie funzionali, morfologiche, di connessione o di interazioni cellula-cellula di neuroni e/o cellule gliali. (Jamel Chelly et al, European journal of Human Genetics 2006).

Di conseguenza, possono essere distinti due principali gruppi di geni:

- geni coinvolti in disordini di ritardo mentale associati ad anomalie dello sviluppo cerebrale;
- geni coinvolti in disordini di ritardo mentale senza nessuna specifica anomalia cerebrale.

Alcuni esempi di geni noti appartenenti al primo gruppo sono: i geni che controllano l'identità cellulare dei precursori dei neuroni (*PTCH*, *SHH*, *SIX3*, *TDGF1*, *TGIF*, *ZIC2*) (Wallis e Muenke 2000), coinvolti nell'insorgenza della Oloprosencefalia (un fallimento a formare emisferi distinti nelle metà destra e sinistra del cervello) e mutazioni missenso dominanti nel gene *EMX2*, che codifica un fattore di trascrizione, responsabile della Schizencefalia ("cervello spaccato"). Difetti di migrazione dei neuroni causano anche Lissencefalia ("cervello liscio") dovuta a mutazioni nei geni *LIS1*, *DCX*, e *RELN*, *ARX* (alcuni alleli) e *FLNA* (Olson e Walsh 2002). (J. K. Inlow et L. Restifo, 2003)

La maggior parte dei geni noti, coinvolti in disordini di MR appartenenti al secondo gruppo, sono geni X-linked, sebbene stia progressivamente aumentando il numero dei geni e dei loci autosomici coinvolti nel ritardo mentale autosomico recessivo (NSMR), la maggior parte di essi può essere suddivisa in due sottogruppi:

- Geni coinvolti nel controllo trascrizione
- Geni implicati nel rimodellamento della cromatina

Una delle più studiate cascate di segnalazione della trascrizione coinvolte nei disordini cognitivi, è il meccanismo Ras-MAPK. È stato, infatti, dimostrato che le proteine codificate dai geni *NF1*, *RSK2* e *CBP*, sono coinvolte nei disordini di MR. (Chelly et al, 2006).

Le mutazioni di *NF1* sono responsabili della neurofibromatosi e sono state trovate in circa il 50% dei pazienti con ritardo mentale. (Chelly et al, 2006).

Esperimenti in topo, hanno dimostrato che l'interazione tra *NF1* e Ras è critica per l'apprendimento, facendo supporre che, nell'uomo, questa interazione sia coinvolta nelle patologie caratterizzate da deficit cognitivi. (Costa RM et al, 2001).

Il gene *RSK2* è responsabile della fosforilazione di CREB, che risulta danneggiata nei fibroblasti dei pazienti affetti dalla sindrome con MR di Coffin-Lowry (CLS), proprio a causa della distruzione di *RSK2* (Trivier E et al, 1996; De Cesare D et al, 1998, Harum KH et al, 2001).

CBP è il gene responsabile della sindrome di Rubinstein–Taybi (RTS) (Petrij F et al, 1995) ed ha un’attività intrinseca acetiltransferasica degli istoni, per cui svolge funzioni implicate nella regolazione genica e in eventi di rimodellamento della cromatina. (Swank MW et al,2001).

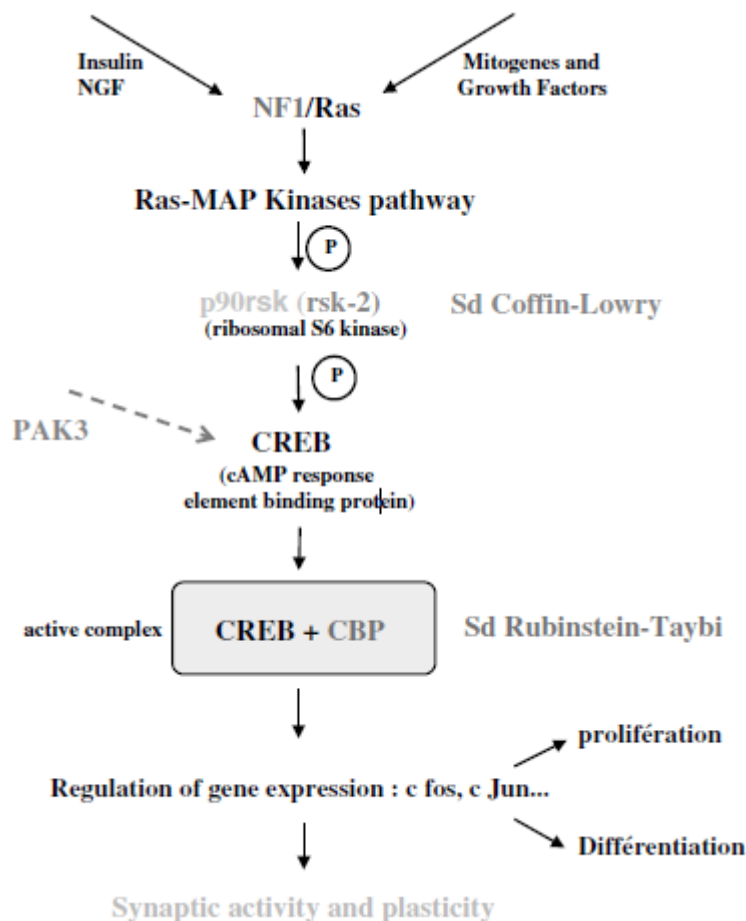


Figura 3.4: schema della cascata del segnale di trascrizione delle Ras/MAPK che mostra i geni e le proteine che sono note essere associate ai disordini di MR (NF1 e PAK3-coinvolte nel MR delle sindromi , Coffin –Lowry e Rubinstein – Taybi). (J. Chelly et al, 2006).

Un altro esempio di gene correlato al ritardo mentale, che codifica una proteina coinvolta nel rimodellamento della cromatina, è il gene *MECP2* (methyl-CpG-binding protein 2). Mutazioni in *MECP2* sono state trovate nella gran maggioranza dei pazienti con la sindrome di Rett (Kriaucionis S et al,2003; Caballero IM et al, 2005). Studi del gene *MECP2*, hanno evidenziato che, a seconda del tipo di mutazione e del modello di inattivazione del cromosoma X (XCI), si possono avere diversi fenotipi, non solo nelle femmine ma anche nei maschi. (Amir RE et al,2000; Bienvenu T et al,2000; Meloni I et al,2000; Couvert P et al,2001).

Inoltre, dati recenti hanno mostrato che le mutazioni nel gene *CDKL5*, (cyclin-dependent protein kinase-like 5, (CDKL5/STK9)), sono responsabili di malattie dello

sviluppo del sistema nervoso con caratteristiche cliniche simili a quelle della sindrome di Rett. (Tao J et al, 2004; Weaving LS et al, 2004).

Data la sovrapposizione clinica che risulta dalla mutazione dei geni *MECP2* e *CDKL5*, è stato ipotizzato il loro coinvolgimento in un comune meccanismo genetico e patogenetico ed ulteriori studi hanno mostrato che *Mecp2* e *Cdkl5* interagiscono e che la loro espressione spazio-temporale si sovrappone (Mari F et al, 2005).

Per comprendere i processi biologici e cellulari responsabili del MR è necessario identificare i geni bersaglio regolati dai geni coinvolti nelle vie di segnalazione della trascrizione (*NF1*, *RSK2* e *CDKL5*), dai fattori di rimodellamento della cromatina (*MECP2* e *CBP*) e dai fattori di trascrizione come *FMR2*, *SOX3* e *ARX*.

3.3.2 Funzione e struttura sinaptica e ritardo mentale

Recenti studi sul ritardo mentale suggeriscono che difetti nella sinaptogenesi, nell'attività e nella plasticità sinaptica sono processi cellulari cruciali che, probabilmente, sottendono il deficit cognitivo. In accordo a questa ipotesi emergente, la maggior parte delle proteine codificate da geni coinvolti nel MR sono localizzate nei compartimenti cellulari pre e/o post sinaptici.

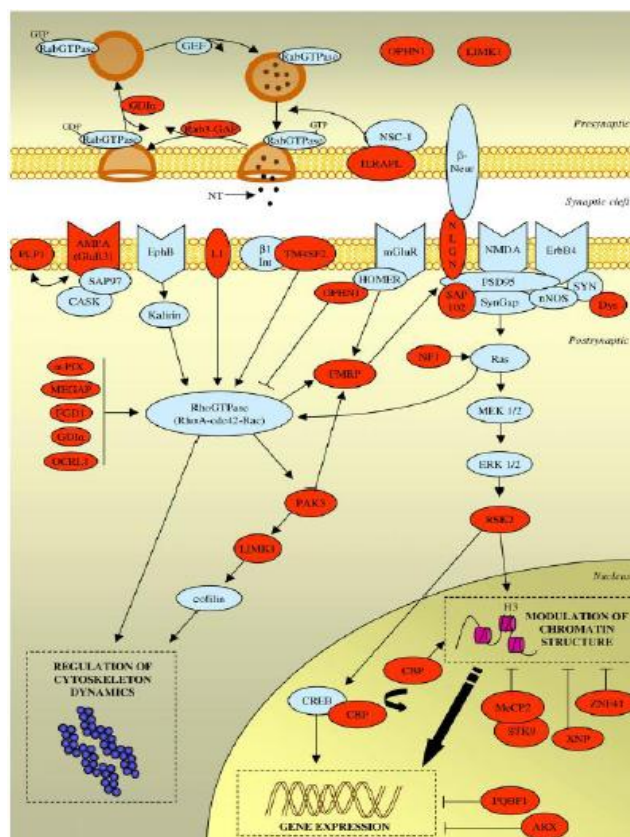


Figura 3.5: Geni associati a ritardo mentale e meccanismi di trasmissione e plasticità sinaptica. Rappresentazione schematica dei principali pathway di segnalazione che controllano il ciclo delle

vescicole sinaptiche, l'espressione genica e le dinamiche del citoscheletro. Le proteine responsabili di ritardo mentale sono colorate in rosso (Vaillend et al., 2008).

Ad eccezione dei fattori di trascrizione e delle proteine che rimodellano la cromatina, questa localizzazione cellulare è valida per la maggior parte delle proteine correlate al MR note, tra le quali FMRP, OPHN1, NLGN4, DLG3, Rab GDI, Neurotrypsina e probabilmente PAK3. (Chelly et al, 2006).

La sindrome dell'X fragile (FXS), è causata dal silenziamento trascrizionale del gene *FMR1*, mediato da un'espansione massiva di una ripetizione CGG nel 5'UTR.

Il prodotto del gene *FMR1*, FMRP, è una proteina che lega l'RNA ed è coinvolta nel trasporto dendritico degli mRNA. Nelle spine dendritiche FMRP è sintetizzata in risposta all'attivazione dei recettori metabotropici del glutammato (mGluR) e si associa ai poliribosomi (Weiler IJ et al, 1997). È stato dimostrato che anomalie, morfologiche e funzionali delle sinapsi nella corteccia cerebrale, nel cervello e nell'ippocampo contribuiscono al deficit cognitivo nei pazienti affetti dalla malattia dell'X fragile. (Bagni C et al, 2005; Chelly et al, 2006).

L'identificazione, in pazienti affetti da MR, di mutazioni che colpiscono geni implicati nel meccanismo di segnalazione delle Rho GTPase come *OPHN1*, *PAK3*, *ARHGEF6* e *FGD1*, ha ulteriormente sottolineato l'importanza della struttura, funzione e controllo delle sinapsi nella patofisiologia del MR. (J. Chelly et al, 2006).

OPHN1 interagisce con Homer, una proteina che lega i recettori del glutammato (GluR-1) e influenza la morfogenesi delle spine e la trasmissione sinaptica (Govek EE et al, 2004; J. Chelly et al, 2006).

PAK3 (chinasi attivata da p21) è un membro della famiglia delle proteine PAK attivate dalle 2 piccole GTPases, Rac e Cdc42 (Meng J et al, 2005; Boda B et al, 2004; Govek EE et al, 2004).

Il coinvolgimento di *PAK3* nella formazione e nella plasticità delle sinapsi, è stato chiaramente dimostrato: la sua inattivazione porta alla formazione di spine dendritiche anormali, una ridotta attività sinaptica spontanea e un difettoso potenziamento a lungo termine (LPT) (Boda B et al, 2004); topi knockout mancanti di *PAK3* e mostrano significative anomalie nella plasticità sinaptica, alterazioni nella morfologia, densità e lunghezza delle spine e deficit nella memoria e nell'apprendimento. (Meng J et al, 2005; J. Chelly et al, 2006).

È stato ipotizzato che anche un altro gruppo di geni coinvolti nel MR come *NLGN4*, *DLG3*, *IL1RAPL* e *GDI1*, possa essere implicato nella regolazione dell'attività sinaptica.

Ad esempio il gene *NLGN4* è stato trovato mutato in un ampio spettro di fenotipi, dal MR lieve alla sindrome di Asperger (Jamain S et al, 2003). La proteina NLGN4 è un membro della famiglia delle neuroligine, ed è particolarmente abbondante nelle membrane delle sinapsi eccitatorie ed inibitorie (Dean C et al, 2003; Scheiffele P et al, 2000 ; J. Chelly et al, 2006).

Le mutazioni che alterano la stechiometria interna dei complessi proteici delle sinapsi, o mutazioni dominanti negative che cambiano la struttura delle singole proteine, hanno probabilmente una conseguenza funzionale, non solo per le forme di MR X-linked ma anche per le forme autosomiche dominanti. Infatti, uno screening di mutazioni in un campione di 94 pazienti con ID idiopatico ha recentemente identificato tre mutazioni de novo in eterozigosi che troncano il gene *SYNGAP1*. (Hamdan FF et al, 2009).

SYNGAP1 è espresso nelle sinapsi e media la plasticità sinaptica, controllando l'inserimento dei recettori alfa-amino-3-idrossi-5-metil-4-isoazolo (AMPA) nella membrana sinaptica e regolando il citoscheletro dell'actina attraverso la fosforilazione della cofilina. (Hamdan FF et al, 2009).

Inoltre, in uno studio condotto su 95 pazienti con ID idiopatico sono state identificate due mutazioni in eterozigosi che inattivano il gene *STXBPI*, entrambe sono de novo ed assenti in controlli sani. *STXBPI* ha una funzione regolatoria essenziale nella fusione, mediata da SNARE, delle vescicole sinaptiche con la membrana presinaptica e nel rilascio dei neurotrasmettitori nello spazio sinaptico. (Hamdan FF et al, 2009).

Infine, varie mutazioni de novo che inattivano il gene *FOXG1* sono state trovate in una coorte di 310 pazienti con ID idiopatico. *FOXG1* codifica per un fattore di trascrizione, la presenza in eterozigosi di una mutazione che distrugge questo gene è stata associata a grave MR e anomalie dello sviluppo del cervello, e recentemente, diverse mutazioni di *FOXG1* sono state trovate in pazienti con varianti congenite della sindrome di Rett. (Shoichet SA et al, 2005)

Questi studi suggeriscono che, nelle popolazioni Occidentali, tra pazienti con MR idiopatico e sporadico sono sorprendentemente comuni mutazioni de novo nei geni espressi nel cervello. Sebbene le definizioni dei processi neurologici responsabili del MR sono ad uno stadio precoce, nell'insieme questi dati suggeriscono che la perdita

di regolazione dei meccanismi che regolano l'attività e la plasticità sinaptica, potrebbe essere una delle basi cellulari che contribuiscono al MR sia autosomico che X linked. (J. Chelly et al, 2006).

Kandel e collaboratori (2001), hanno costruito un parallelo tra la disfunzione sinaptica, responsabile del deficit cognitivo, e le variazioni biochimiche delle sinapsi, associate a cambiamenti di forme semplici di immagazzinamento della memoria a lungo e breve termine (Kandel ER et al, 2001). Nel loro 'modello' descrivono il processo di immagazzinamento della memoria e l'apprendimento come un 'dialogo tra i geni e le sinapsi': loro propongono che la memoria a breve termine risulta dagli immediati cambiamenti biochimici delle sinapsi (come l'attivazione di *CaMKII* e l'aumento nell'attività del recettore del glutammato AMPA), mentre l'immagazzinamento della memoria a lungo termine richiede generalmente la trascrizione e la traduzione di nuove proteine che incrementano la forza o il numero delle sinapsi attive.

A favore di questo modello, che combina effetti a breve e a lungo termine con le attività sinaptiche, c'è il coinvolgimento dei fattori di trascrizione e di rimodellamento della cromatina nel MR. (J. Chelly et al, 2006).

La disfunzione delle proteine codificate dai geni coinvolti nelle varie forme di MR potrebbe portare, tramite alterata regolazione di specifici meccanismi e processi cellulari, a difetti nella struttura e/o funzione sinaptica, e nella connessione dei neuroni, forse ostacolando la capacità del cervello di processare le informazioni. (J.Chelly et al, 2006).

La comprensione dei meccanismi patogenetici è un prerequisito fondamentale per lo sviluppo di medicinali. Infatti, studi in *Drosophila* hanno dimostrato che utilizzando antagonisti farmaceutici del mGluR si possono recuperare difetti di plasticità sinaptica, del corteggiamento e della crescita veloce del corpo in un modello della sindrome dell'X fragile: il trattamento delle mosche adulte con la pirimidina 2-metil-6-feniletinil (MPEP), l'antagonista dell'mGluR, può recuperare la memoria a breve termine e normali comportamenti di corteggiamento. (de Vrij FM et al, 2008; McBride SM et al, 2005). Lo scopo è ridurre l'eccessiva attivazione della traduzione di mRNA mediata dall'mGluR, per ridurre l'attività di una o varie proteine che sono sovra espresse a causa della perdita di regolazione mediata da FMRP, in modo tale da rafforzare le connessioni sinaptiche o restaurando l'attività di AMPA, o recuperando altre anomalie dei neurotrasmettitori nel sistema nervoso centrale.

La comprensione dei meccanismi patogenetici è un prerequisito fondamentale per lo sviluppo di medicinali, ma al contrario delle FXS, la funzione della maggioranza degli altri geni di ID è per lo più sconosciuta, eccetto per i disordini che sono dovuti a difetti metabolici.

4. Scopo della tesi

Le malattie complesse, come l'autismo ed il ritardo mentale, sono patologie la cui manifestazione fenotipica dipende dall'azione combinata di fattori ambientali e di fattori genetici.

Studi di associazione su larga scala condotti sull'autismo hanno suggerito che le cause molecolari responsabili di questa patologia siano da ricercarsi nella presenza di molteplici varianti genetiche rare, sia di sequenza che strutturali (CNVs), un modello che ricalca quello attualmente approvato per il ritardo mentale, in cui vari geni causano un fenotipo comportamentale simile e suggerisce l'esistenza di pathways comuni in questi due disturbi.

Lo studio delle anomalie cromosomiche strutturali, grazie allo sviluppo di tecniche di microarray, si è affermato negli ultimi anni come un potente mezzo per l'identificazione delle cause molecolari responsabili di fenotipi complessi come il ritardo mentale e l'autismo. Studi recenti mostrano che negli individui autistici è presente un numero di CNVs più elevato rispetto ai controlli, e i geni maggiormente colpiti sembrano essere geni coinvolti nell'adesione cellulare e nella trasmissione sinaptica, molti dei quali sono stati precedentemente implicati nel ritardo mentale.

In questo contesto si inserisce il mio progetto di tesi, volto all'identificazione e caratterizzazione di varianti strutturali ed allo screening di geni candidati in individui affetti da autismo e in individui affetti da ritardo mentale.

Nel primo periodo del mio dottorato mi sono occupata della caratterizzazione molecolare di due delezioni che segregano all'interno della stessa famiglia (15-0084) con individui affetti da autismo e dislessia, e localizzate, rispettivamente, sul cromosoma 7q e sul cromosoma 2q.

La delezione localizzata sul cromosoma 7q, coinvolge due geni contigui, *IMMP2L* e *DOCK4*, ed è trasmessa dalla madre (dislessica) a tutti e tre i figli (a due fratelli affetti da autismo e alla sorella dislessica). Ci siamo quindi proposti di identificare il breakpoint della delezione e di valutare gli effetti sull'espressione dei due geni coinvolti dalla delezione *IMMP2L* e *DOCK4*. Inoltre, visto il possibile coinvolgimento del gene *DOCK4* nella dislessia, grazie alla collaborazione di vari gruppi di ricerca abbiamo analizzato, 1940 individui affetti da dislessia e circa 4646 controlli, per valutare la frequenza di delezioni di questo gene in un campione esteso di casi.

Nella stessa famiglia è stata identificata una seconda microdelezione con possibile effetto patologico sul cromosoma 2q14.3, è trasmessa dal padre ad entrambi i figli affetti da autismo e distrugge il gene *CNTNAP5*, che rappresenta anche un buon candidato funzionale per l'autismo in quanto codifica per una proteina trans-membrana, appartenente alla superfamiglia delle neurexine, che mediano il contatto e le interazioni cellulari nel sistema nervoso, giocando un ruolo essenziale

nel suo corretto sviluppo e nel suo funzionamento. Quindi ho effettuato uno screening di mutazione delle regioni codificanti del gene *CNTNAP5*, al fine di individuare varianti rare di sequenza che potrebbero avere un ruolo eziologico nel disturbo autistico. Lo screening di mutazione è stato condotto su 143 casi provenienti da famiglie *multiplex* della collezione dell'IMGSAC.

Durante l'ultimo anno di dottorato, ho preso parte ad un progetto promosso da un consorzio di scienziati e clinici provenienti dall'Europa dell'est e dall'Asia centrale (il CHERISH), che ha l'obiettivo di collezionare un grande campione di pazienti affetti da ritardo mentale (ad oggi circa 1200) e identificare le cause molecolari di questa patologia.

In particolare io mi sono occupata di ricercare varianti strutturali in campione di 218 individui affetti da MR tramite array CGH (44K dell'Agilent), identificando 135 alterazioni in 127 pazienti.

La caratterizzazione delle anomalie strutturali identificate in questo campione e' ancora in corso e richiederà ulteriori ricerche. Io mi sono occupata in particolare della caratterizzazione di una microdelezione di circa 250 kb, localizzata sul cromosoma 7 a livello del locus *AUTS1*, precedentemente identificato come locus di suscettibilità per l'autismo, e quindi particolarmente interessante. La delezione è presente in eterozigosi in due individui affetti da MR ed autismo, appartenenti alla stesso nucleo familiare. Questa delezione distrugge il gene *CADPS2* che codifica per una proteina associata alle vescicole secretorie coinvolte nel rilascio delle neurotrofine (BDNF ed NT-3). Studi recenti hanno dimostrato che la perdita di *CADPS2* colpisce il rilascio sia costitutivo che regolato di BDNF, che è un fattore importante per la formazione delle sinapsi.

Pertanto, mi sono dedicata alla caratterizzazione della delezione, validando il risultato ottenuto da arrayCGH tramite Real Time PCR, definendo con precisione i confini, e effettuando uno screening di mutazione delle regioni codificanti del gene *CADPS2*, in un campione di 37 soggetti affetti da RM e in un campione di 94 individui affetti da autismo.

Infine e' stata valutata la possibile ipotesi che il gene *CADPS2* possa essere soggetto ad imprinting.

5. Materiali e Metodi

5.1 Campioni di DNA

I campioni di DNA genomico utilizzati in questo studio provengono dalla collezione di famiglie *multiplex* del Consorzio IMGSAC, di cui il nostro laboratorio fa parte e dalla collezione di famiglie del consorzio CHERISH. Il DNA genomico dei pazienti è stato estratto da sangue o da saliva secondo le condizioni descritte in Maestrini et al. [1999].

Sono state raccolte anche famiglie italiane grazie ad una collaborazione del nostro laboratorio di Bologna con il Dott. Agatino Battaglia dell'Istituto Stella Maris di Pisa.

Il DNA genomico degli individui appartenenti alle famiglie italiane è stato estratto utilizzando il kit QIAGEN MAXI. Come popolazione di controllo sono stati utilizzati i DNA di individui caucasici di sesso misto appartenenti alla collezione ECACC HRC (*European Collection of Cell Culture*).

Gli stock di DNA sono dapprima diluiti ad una concentrazione finale di 80ng/μl in TE 1X e conservati a -20°C. Per effettuare le reazioni di amplificazione tramite PCR, si opera una ulteriore diluizione dei campioni in TE 0.1X, ottenendo una concentrazione finale di 8 ng/μl. I campioni così diluiti possono essere conservati a 4°C per alcuni mesi.

5.2 Estrazione del DNA da sangue intero con kit Qiagen MAXI (per 5 ml di sangue)

Il sangue, fresco o scongelato, è versato con attenzione in un tubo da 50 ml contenente 500 μl di Proteasi Qiagen, e lisato con 6 ml di buffer di lisi (AL).

I campioni sono agitati e vortexati 3 volte per 5 secondi, quindi incubati per almeno 10' a 70°C. Dopo raffreddamento a temperatura ambiente si aggiungono 5 ml di etanolo assoluto, si vortexano i campioni tre volte per 5 secondi, e il lisato è versato nel tubo contenente la colonna Qiamp. Dopo aver centrifugato per 3' a 3000 rpm (1850 g), si butta via il filtrato e si aggiungono 5 ml di buffer di lavaggio (AW1) e si centrifuga per 2' a 4500 rpm. Senza buttare via il filtrato, si aggiungono 5 ml di buffer di lavaggio (AW2), si centrifuga 20' a 4500 rpm, si sposta la colonna in un tubo da 50 ml pulito e si aggiungono successivamente 600 μl di buffer di eluizione (AE). Si lascia avvenire l'eluizione per 10' a temperatura ambiente, quindi si centrifuga 5' a 4500 rpm e si aggiungono altri 600 μl di buffer di eluizione (AE), ripetendo il procedimento precedente una seconda volta. Il DNA eluito viene quindi messo in un tubo sterile da 1.5 ml, e quantificato allo spettrofotometro, con determinazione dell'assorbanza a 260 nm.

5.3 Identificazione di CNVs tramite la tecnologia ILLUMINA

Il sistema Illumina (BeadArray™) si basa su array di fibre ottiche. In una piastra sono disposti 96 arrays organizzati in una matrice 8x12. Ogni array è composto da circa 50.000 fibre ottiche, su ognuna delle quali è inciso un piccolo pozzetto che alloggia una bead di 3 micron, alla quale sono attaccate moltissimi probes con la stessa sequenza nucleotidica. Ogni array contiene più di 1500 diverse sonde oligonucleotidiche, ognuna delle quali è presente in circa 30 copie, rendendo quindi il risultato di una genotipizzazione più affidabile perché basato su delle repliche. Questa tecnologia permette quindi di genotipizzare 96 campioni per più di 1500 SNPs contemporaneamente.

Il saggio Illumina (GoldenGate™ genotyping assay) consente di discriminare le varianti di SNPs direttamente sul DNA genomico piuttosto che su prodotti di PCR, per questo motivo non possono essere introdotti errori causati dal processo di amplificazione. Il DNA genomico viene attivato coniugandolo ad un supporto solido (particelle sferiche paramagnetiche). Seguono un passaggio per l'ibridazione degli oligonucleotidi disegnati per il saggio di specifici SNPs di interesse e diversi lavaggi per eliminare gli oligonucleotidi che non si sono legati.

- Per ogni SNP da testare vengono disegnati due oligonucleotidi allele specifici (ASOs: Allele-Specific Oligonucleotides) e un oligonucleotide locus specifico (LSO: Locus Specific Oligonucleotides) circa 20bp a valle dell'oligonucleotide allele specifico. Ogni ASO termina al 3' con una base che si ibrida (in modo allele specifico) al DNA genomico in corrispondenza dello SNP, mentre al 5' contiene una regione complementare ad un primer universale per la PCR (P1 o P2 coniugati ad una cianina: Cy3 o Cy5 a seconda dell'allele). Il LSO, invece, è costituito da tre regioni: al 5' c'è una sequenza locus-specifica per lo SNP, nella porzione centrale una sequenza indirizzo complementare ad una delle 1500 sonde contenute in un array della piattaforma, e al 3' una sequenza complementare ad un primer universale (P3). Si utilizza quindi una polimerasi in grado di effettuare l'estensione di un oligonucleotide che si ibrida perfettamente al 3' con il target e priva di attività esonucleasica e di displacement di sequenze a valle. In questo modo viene riempito il gap fra ogni ASO e LSO generando un substrato allele specifico per la reazione di PCR. Si introducono i tre primer universali (P1, P2, P3) e si effettua un classico programma di amplificazione. I prodotti di PCR infine si ibridano alle sonde negli arrays tramite la sequenza indirizzo. Come precedentemente spiegato, ad ognuno dei 1500 SNP è assegnata una sequenza indirizzo contenuta nel LSO complementare ad una sola sonda presente nell'array.

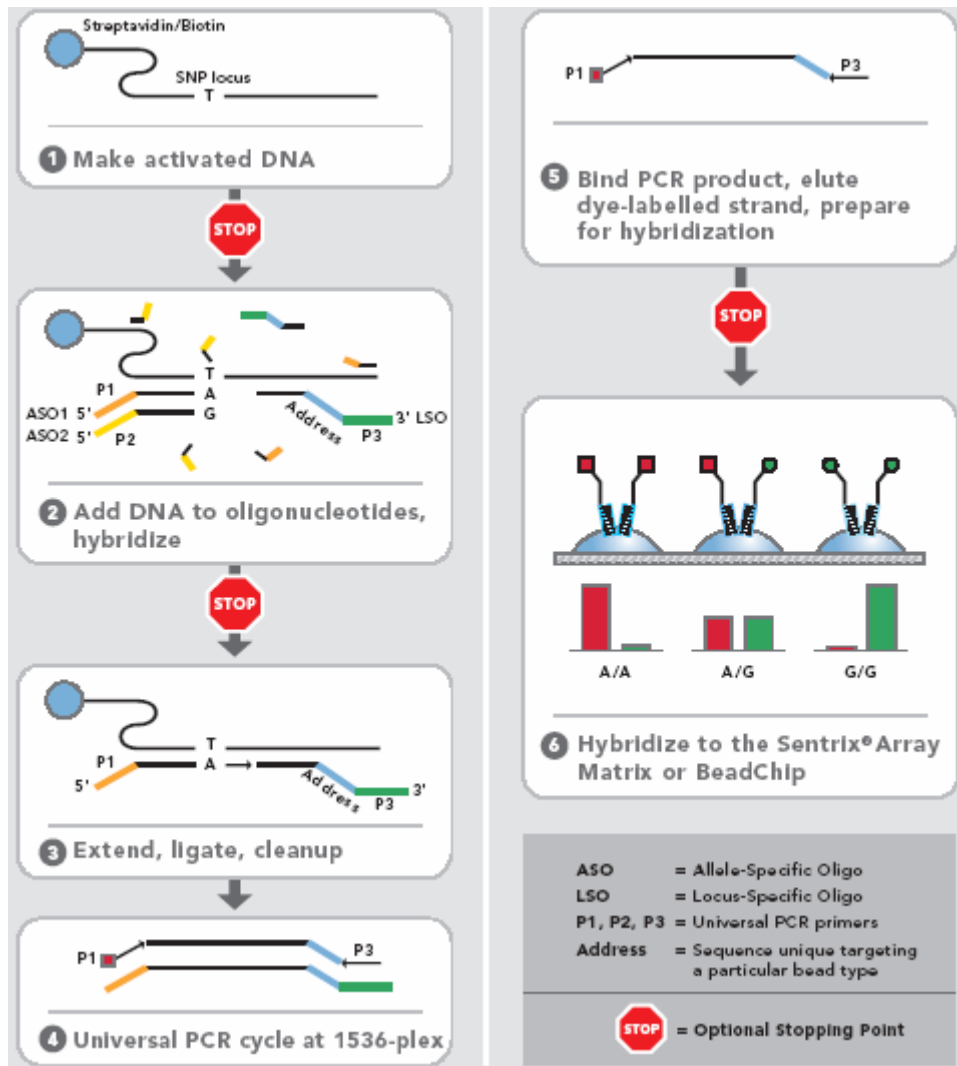


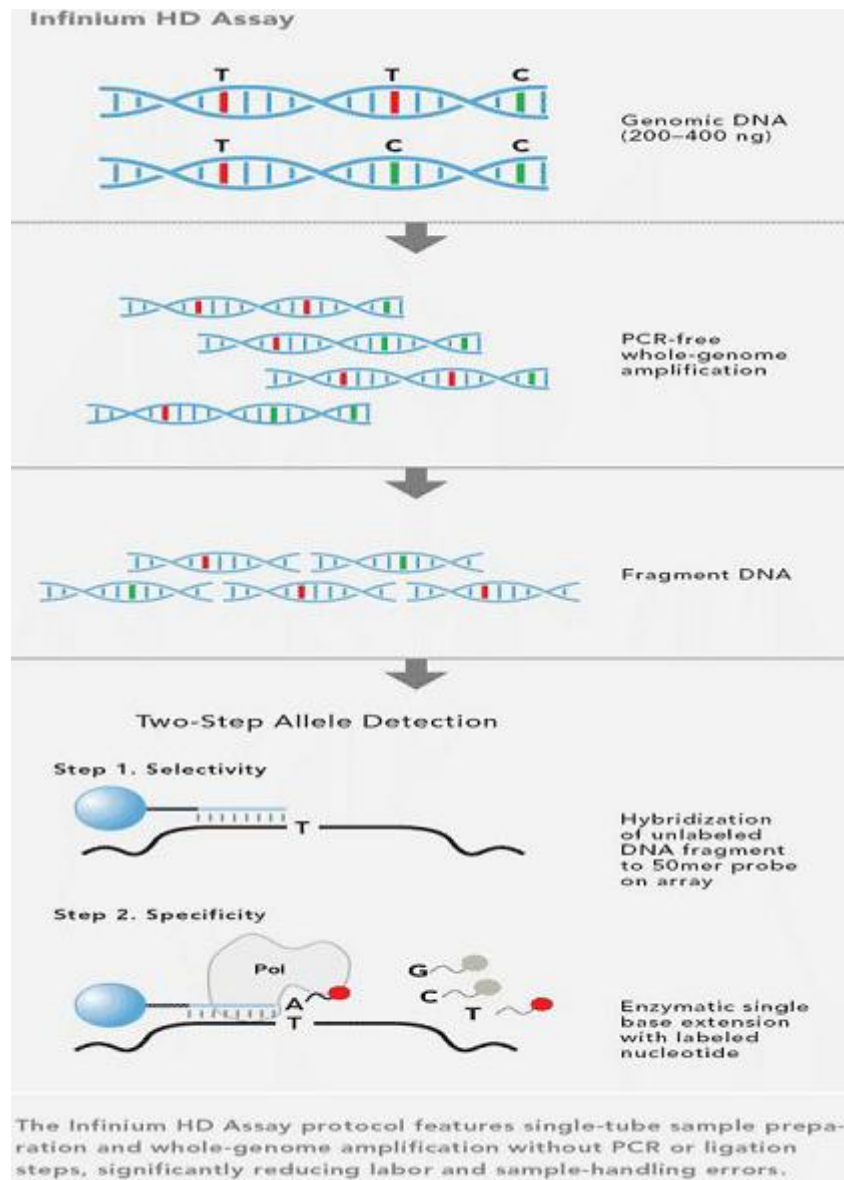
Figura 5.1: Schema del saggio Illumina (GoldenGate™ genotyping assay)

I rilevatori di fluorescenza utilizzati da Illumina scansionano infine la piastra contenente 96 arrays contemporaneamente a due diverse lunghezze d'onda adatte alla Cy3 e alla Cy5.

Nel caso di un campione omozigote per un dato SNP, il rilevatore osserva un solo tipo di fluorescenza (o quella dipendente dal primer P1 coniugato alla Cy3 o dal primer P2 coniugato alla Cy5), invece in presenza di un campione eterozigote verranno rilevate entrambe i tipi di fluorescenza. Quindi il rapporto dei segnali relativi alle due diverse fluorescenze identifica il genotipo come AA, AB, o BB.

L'array Illumina Infinium 1M (Infinium HD Human1M-Duo BeadChip) contiene più di 1 milione di sonde oligonucleotidiche lunghe 50 nucleotidi, ognuna delle quali è specifica per ogni SNP. Questa tecnologia permette quindi di genotipizzare due diversi campioni per più di un milione di SNPs simultaneamente. Inoltre consente di discriminare le varianti di SNPs direttamente sul DNA genomico piuttosto che su prodotti di PCR, per questo motivo non possono essere introdotti errori causati dal processo di amplificazione.

La detection delle varianti alleliche avviene in due fasi: inizialmente le sonde oligonucleotidiche legate ad ogni beads dell'array si ibridano in maniera selettiva ai frammenti di DNA genomico fermandosi una base prima dello SNP specifico; successivamente avviene l'estensione ad opera della polimerasi che incorpora un nucleotide differentemente marcato a seconda della variante allelica dello SNP.



Questo tipo di saggio utilizza un sistema di immagine che permette di rilevare sia l'emissione di fluorescenza dei due diversi fluorofori che l'intensità del segnale.

Nel caso di un campione omozigote per un dato SNP, il rilevatore osserva un solo tipo di fluorescenza (o quella dipendente dal fluoroforo che emette nel verde o quella dipendente dal fluoroforo che emette nel rosso), invece in presenza di un campione eterozigote verranno rilevate entrambe i tipi di fluorescenza. Quindi il rapporto dei segnali relativi alle due diverse fluorescenze identifica il genotipo come AA, AB, o BB (figura xxx).

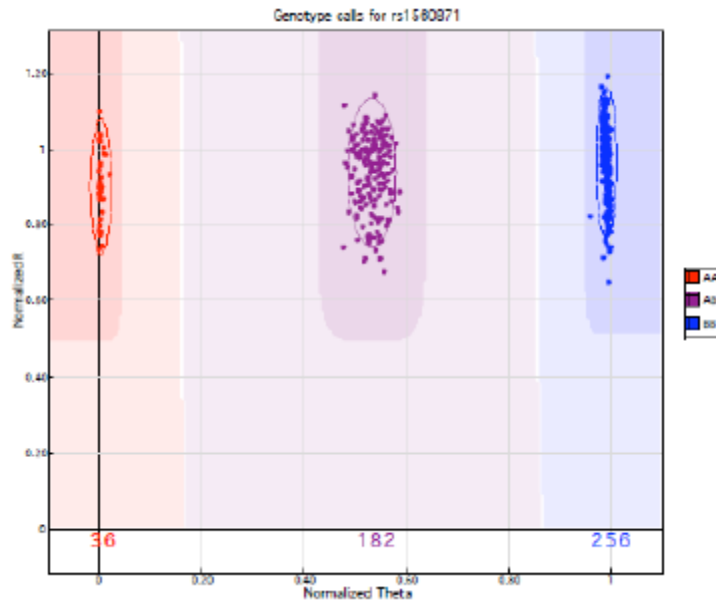


Figura 5.3: Rappresentazione grafica dei risultati forniti dal sistema Illumina per il *genotyping* di uno SNP (nell'esempio *rs1560871*). I tre *cluster* sono formati dai segnali di fluorescenza ottenuti per individui omozigoti AA (rosso), eterozigoti AB (viola), omozigoti BB (blu).

5.4 Validazione dei CNV tramite Real Time

La PCR Real Time è in grado di amplificare e quantificare il DNA in maniera simultanea. La quantificazione viene effettuata determinando un segnale di fluorescenza generato in seguito al legame di un reagente al prodotto amplificato nei primi stadi della reazione di PCR. Uno dei coloranti fluorescenti maggiormente utilizzati è il SYBR GREEN in grado di legarsi al solco minore del DNA. Ad ogni ciclo di amplificazione la fluorescenza emessa dal SYBR GREEN intercalato alla doppia elica è direttamente proporzionale al numero di copie prodotte dalla reazione, per cui, registrando la quantità di emissione fluorescente per ogni ciclo, è possibile monitorare la reazione di polimerizzazione durante la sua fase esponenziale, nella quale il primo incremento significativo di prodotti neo-sintetizzati è collegato alla concentrazione iniziale di stampo nel campione. Infatti, maggiore è il numero di copie iniziali dell'acido nucleico, prima si osserverà un incremento significativo della fluorescenza (figura 5.3). Tuttavia il legame del SYBR GREEN non è altamente specifico perché può legarsi ai dimeri di primers o anche ai prodotti non specifici di amplificazione.

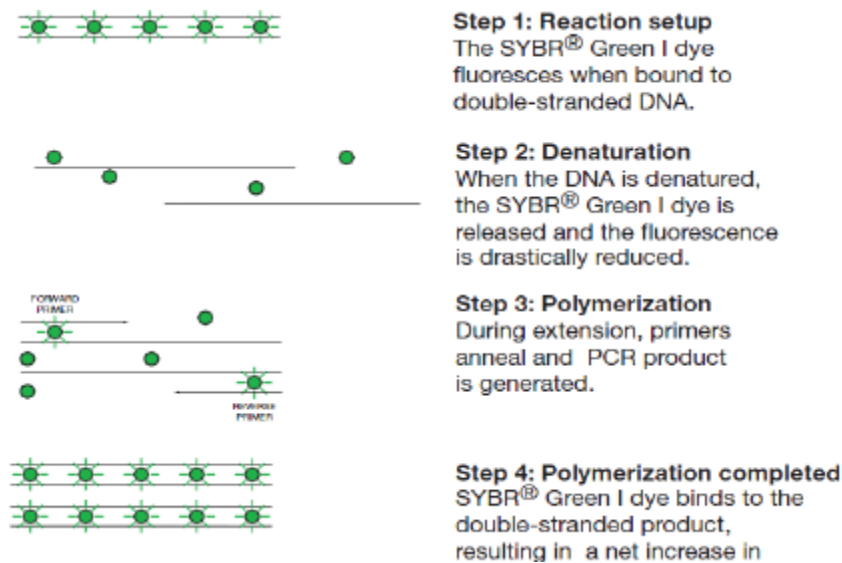


Figura 5.4: processo di reazione del colorante SYBR GREEN.

Un metodo di detection più specifico, ma tuttavia più costoso rispetto a quello basato sul SYBR GREEN, consiste nell'uso di sonde gene-specifiche (sonde TaqMan) che si appaiano nella zona compresa fra i due primers (forward e reverse) e contenenti un colorante fluorescente (*Reporter*), solitamente di colore verde, all'estremità 5' ed un colorante spegnitore (*Quencher*), di colore rosso, all'estremità 3'. Durante la fase di annealing della sonda al DNA stampo, grazie alla vicinanza del colorante ad alta energia in 5' e quello a bassa energia in 3', si verifica un trasferimento di energia fluorescente (FRET) dal primo al secondo annullando così l'emissione del segnale di fluorescenza; durante, invece, la fase di replicazione, la DNA-polimerasi sfrutta la propria attività esonucleasica 5'→3' per degradare la sonda correttamente appaiata; come conseguenza, l'allontanamento tra il reporter ed il quencher pone fine all'attività di assorbimento di quest'ultimo e fa in modo che il reporter inizi ad emettere fluorescenza. Quest'ultima incrementerà ad ogni ciclo in modo proporzionale al tasso di degradazione della sonda. L'accumularsi del prodotto amplificato viene rivelato monitorando quindi l'incremento di fluorescenza del reporter (figura xxx).

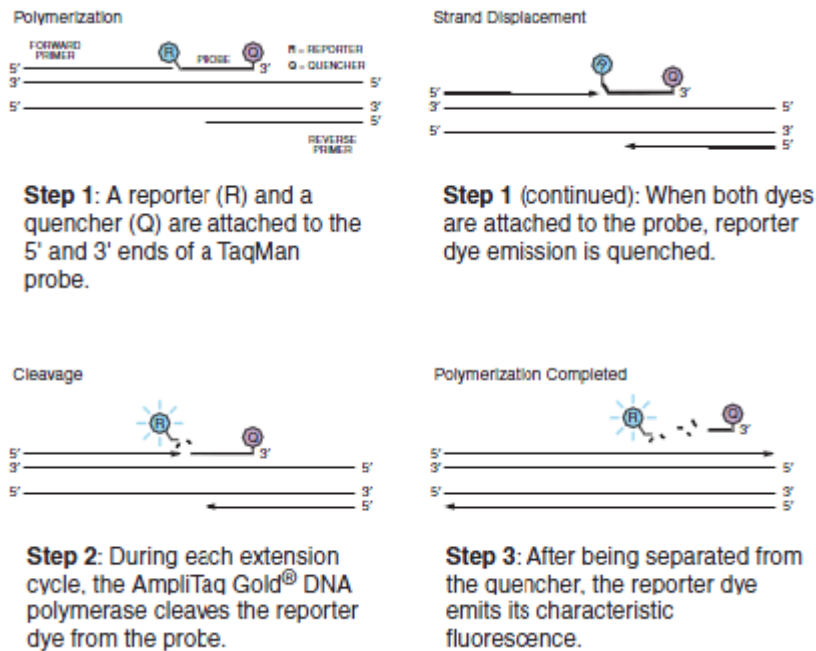


Figura 5.5: processo di reazione delle sonde TaqMan.

Dato l'utilizzo del SYBR GREEN come metodo di detection, prima di effettuare l'esperimento vero e proprio, è necessario allestire una curva standard di calibrazione. Si sfrutta il metodo della quantificazione assoluta che prevede l'utilizzo di un DNA genomico di controllo (reference), non contenente il CNV di interesse, a concentrazione nota, in triplicato, e diluizioni progressive (20 ng/μl; 10 ng/μl; 5 ng/μl; 2,5 ng/μl). L'analisi è stata eseguita sullo strumento 7500 Fast Real Time PCR System (Applied Biosystems). Ogni run consiste di una singola piastra da 96 pozzetti. La miscela di reazione contiene 1 μl di ciascun primer (forward e reverse) ad una concentrazione di 10 μM; 10 μl di SYBR Green PCR Master Mix (Applied Biosystems) contenente SYBR Green I Dye, AmpliTaq Gold DNA Polymerase, dNTPs e buffer; si arriva ad un volume finale di 15 μl con acqua e si aggiungono 5 μl di DNA. Per il disegno dei primers è stato utilizzato il programma Primer 3. Insieme ai primers specifici per i frammenti di interesse è stata utilizzata una coppia di primers che permette l'amplificazione di un gene di controllo (*FOXP2*).

Le condizioni di amplificazione per la costruzione della curva standard prevedono tre stadi:

- 1) *Enzyme Activation*: 20 sec a 95°C.
- 2) *Melting*: 3 sec a 95°C; 30 sec a 60°C.
- 3) *Annealing/Extension*: 15 sec a 95°C; 1 min a 60°C, per 40 cicli.

A questi 3 stadi viene aggiunto uno stadio di lenta dissociazione che permette di evidenziare l'eventuale presenza di dimeri di primers: consiste nell'uso di sonde gene-specifiche (sonde TaqMan) che si

appaiano nella zona compresa fra i due primers (forward e reverse) e contenenti un colorante fluorescente (Reporter), solitamente di colore verde, all'estremità 5' ed un colorante spegnitore (Quencher), di colore rosso, all'estremità 3'. Durante la fase di annealing della sonda al DNA stampo, grazie alla vicinanza del colorante ad alta energia in 5' e quello a bassa energia in 3', si verifica un trasferimento di energia fluorescente (FRET) dal primo al secondo annullando così l'emissione del segnale di fluorescenza; durante, invece, la fase di replicazione, la DNA-polimerasi sfrutta la propria attività esonucleasica 5' → 3' per degradare la sonda correttamente appaiata; come conseguenza, l'allontanamento tra il reporter ed il quencher pone fine all'attività di assorbimento di quest'ultimo e fa in modo che il reporter inizi ad emettere fluorescenza. Quest'ultima incrementerà ad ogni ciclo in modo proporzionale al tasso di degradazione della sonda. L'accumularsi del prodotto amplificato viene rivelato monitorando quindi l'incremento di fluorescenza del reporter (figura 5.4). Dato l'utilizzo del SYBR GREEN come metodo di detection, prima di effettuare l'esperimento vero e proprio, è necessario allestire una curva standard di calibrazione. Si sfrutta il metodo della quantificazione assoluta che prevede l'utilizzo di un DNA genomico di controllo (reference), non contenente il CNV di interesse, a concentrazione nota, in triplicato, e diluizioni progressive (20 ng/μl; 10 ng/μl; 5 ng/μl; 2,5 ng/μl). L'analisi è stata eseguita sullo strumento 7500 Fast Real Time PCR System (Applied Biosystems). Ogni run consiste di una singola piastra da 96 pozzetti. La miscela di reazione contiene 1 μl di ciascun primer (forward e reverse) ad una concentrazione di 10 μM; 10 μl di SYBR Green PCR Master Mix (Applied Biosystems) contenente SYBR Green I Dye, AmpliTaq Gold DNA Polymerase, dNTPs e buffer; si arriva ad un volume finale di 15 μl con acqua e si aggiungono 5 μl di DNA. Per il disegno dei primers è stato utilizzato il programma Primer 3. Insieme ai primers specifici per i frammenti di interesse è stata utilizzata una coppia di primers che permette l'amplificazione di un gene di controllo (FOXP2).

Le condizioni di amplificazione per la costruzione della curva standard prevedono tre stadi:

- 1) Enzyme Activation: 20 sec a 95°C.
- 2) Melting: 3 sec a 95°C; 30 sec a 60°C.
- 3) Annealing/Extension: 15 sec a 95°C; 1 min a 60°C, per 40 cicli.

A questi 3 stadi viene aggiunto una stadio di lenta dissociazione che permette di evidenziare l'eventuale presenza di dimeri di primers:

- 4) Dissociation: 15 sec a 95°C; 15 sec a 60°C.

Il risultato è in forma di un grafico con andamento sigmoidale che mostra il numero di cicli di PCR in funzione dell'incremento di fluorescenza (Rn), detto amplification plot (figura 5.5).

Una tipica curva di amplificazione contiene:

- Baseline: numero iniziale di cicli di PCR in cui c'è un piccolo cambiamento nel segnale di fluorescenza.
- Fase esponenziale: aumento significativo di fluorescenza in seguito alla reazione di amplificazione.
- Fase di plateau: fase in cui non si osserva più un incremento nei prodotti; corrisponde alla resa effettiva.
- Threshold: linea la cui intersezione con l'amplification plot definisce il Ct.

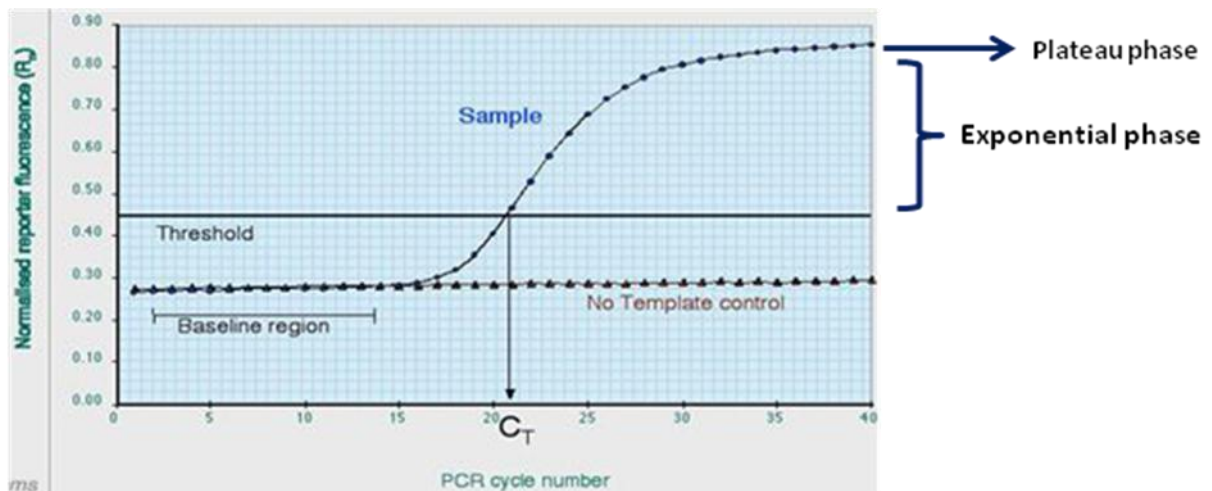


Figura 5.6: amplification plot.

Il ciclo della PCR (ciclo soglia, Ct), specifico per ogni campione, identifica il valore del ciclo di PCR in cui la curva in fase esponenziale interseca la linea soglia, ossia il ciclo in cui viene raggiunto il valore soglia di fluorescenza. Il numero dei cicli necessari perché un campione raggiunga il suo Ct è inversamente proporzionale al numero di copie target presenti inizialmente.

Il grafico che mostra i valori di Ct, misurati per campioni a concentrazione nota, in funzione del Log della concentrazione iniziale del target definisce la curva standard (figura 5.6). La pendenza della curva standard (slope) permette di determinare l'efficienza della reazione di amplificazione.

Lo strumento fornisce inoltre la curva di dissociazione (figura 5.7), un grafico che mostra il cambiamento di fluorescenza, dovuto all'interazione della sonda con il DNA a doppio filamento, in funzione della temperatura di melting (T_m), in corrispondenza della quale il 50% del DNA è a singolo filamento (figura 3). Durante la fase di denaturazione del DNA, quando viene raggiunta la temperatura di melting, si assiste ad un improvviso decremento della fluorescenza, dovuto alla dissociazione dei filamenti di DNA e al rilascio della sonda. In questo modo è possibile identificare i prodotti di amplificazione non specifici, come la formazione di dimeri di primer. Il prodotto specifico è mostrato con una temperatura di melting più alta rispetto al dimero di primer.

Per l'esperimento vero e proprio si prepara la stessa miscela di reazione usata per la curva standard, ma utilizzando il DNA dei campioni di interesse (in questo caso dell'individuo affetto da autismo,

dei genitori e di eventuali fratelli affetti e non) e il DNA di controllo a concentrazione nota (5 ng/μl). Allo stesso modo ogni frammento è amplificato in triplicato.

Il programma prevede due stadi e manca dello stadio di dissociazione:

- 1) Enzyme Activation: 20 sec a 95°C.
- 2) Melting: 3 sec a 95°C; 30 sec a 60°C, per 40 cicli.

Per ogni campione e per ogni sonda, si ottiene un diverso valore di Ct. Il DNA dei campioni di interesse è normalizzato sul DNA di controllo tramite il calcolo del Delta Ct (Ct_{reference} - Ct_{campione}). Altri due parametri calcolati sono il Ratio (Esonda(delta Ct) / EFOXP2 (delta Ct)), per normalizzare ogni sonda testata sul gene di controllo FOXP2, e il Copy number (2*Ratio) da cui è possibile valutare, rispetto al numero di copie di FOXP2 (valore di copy number intorno al 2), se un frammento del gene di interesse, amplificato con lo specifico primer, risulta essere deleto (valore di numero di copie compreso tra 0.7 e 1.3) o duplicato (valore di numero di copie compreso tra 2.7 e 3.3).

5.5 Analisi di varianti strutturali mediante OMPSF della regione IMMP2L-DOCK4

I primers (tabella X) sono stati disegnati con muPlex (<http://genomics14.bu.edu:8080/MuPlex/MuPlex.html>) e successivamente è stato utilizzato il software AutoDimer v.1 per testare l'assenza di prodotti aspecifici (<http://en.bio-soft.net/pcr/AutoDimer.html>).

La mix di primers presi da stock con una concentrazione di 100 ng/ul è così composta:

	IMMP2L_x2	IMMP2L_x3	DOCK4_X52(last)	LRRN3_x4	IMMP2L_x7	RNF20
F	3	3	4,5	3	3	3
R	3	3	4,5	3	3	3

Il volume finale è di 90ul (39ul primers + 51ul H₂O).

La mix per 1 campione è la seguente in un volume finale di 15 ul (10 ul pcr mix + 5 ul DNA ad una concentrazione di 10ng/ul)

- 1,5 ul di buffer 10X
- 1,2 ul di MgCl₂ (25mM)
- 0,9 ul di Mix primers
- 0,12 ul di dNTPs (25mM)
- 0,75 ul di DMSO (5%)
- 0,1 ul di Taq
- 5,43 ul di acqua

Per la reazione di amplificazione è utilizzato il seguente programma:

- 15 minuti a 95°

Per 20 cicli

- 30 secondi a 95°
- 30 secondi a 58°
- 30 secondi a 72°

5.6 Analisi di varianti strutturali mediante OMPSF della regione DOCK4

I primers (tabella X) sono stati disegnati con muPlex (<http://genomics14.bu.edu:8080/MuPlex/MuPlex.html>) e successivamente è stato utilizzato il software AutoDimer v.1 per testare l'assenza di prodotti aspecifici (<http://en.bio-soft.net/pcr/AutoDimer.html>).

La mix di primers presi da stock con una concentrazione di 100 ng/ul è così composta:

	DOCK4-X52	DOCK4-X31	DOCK4-X25	DOCK4-X15	DOCK4-X8	DOCK4-X1	RNF20
F	4,5 μ l	3 μ l	4,5 μ l	1,5 μ l	1,5 μ l	3 μ l	1,5 μ l
R	4,5 μ l	3 μ l	4,5 μ l	1,5 μ l	1,5 μ l	3 μ l	1,5 μ l

Il volume finale è di 90ul (39ul primers + 51ul H₂O).

La mix per 1 campione è la seguente in un volume finale di 15 ul (10 ul per mix + 5 ul DNA ad una concentrazione di 10ng/ul)

- 1,5 ul di buffer 10X
- 1,2 ul di MgCl₂ (25mM)
- 0,9 ul di Mix primers
- 0,12 ul di dNTPs (25mM)
- 0,75 ul di DMSO (5%)
- 0,1 ul di Taq
- 5,43 ul di acqua

Per la reazione di amplificazione è utilizzato il seguente programma:

- 15 minuti a 95°

Per 20 cicli

- 30 secondi a 95°
- 30 secondi a 60°
- 30 secondi a 72°

5.7 analisi semiquantitativa dell'espressione del trascritto di fusione *IMMP2L-DOCK4* mediante RT-PCR

Circa 500 μ l di sangue sono stati conservati in 1300 μ l di RNAlater (Ambion). In seguito è stato estratto l'RNA utilizzando il kit Ribo PureTM Blood Kit (Ambion) seguendo le indicazioni del fornitore. Il cDNA è stato sintetizzato da 500 ng di RNA totale utilizzando il kit Superscript III First Strand Synthesis SuperMix per qRT-PCR e primers random (Invitrogen).

La PCR semiquantitativa è stata condotta su una diluizione 1:20 di cDNA con un numero di cicli crescente (25-30-35-40).

I primer sono stati disegnati sull'esone 4 di *IMMP2L* e sull'esone 26 di *DOCK4*. Per permettere la quantificazione del prodotto fuso la RT-PCR semiquantitativa è stata ripetuta usando altre due coppie di primers una disegnata sull'esone 17-18 e l'altra sull'esone 47-48 di *DOCK4*. (Vedi tabella) Il prodotto di PCR è stato corso sul gel di agarosio al 2% con GelRedTM (Biotium) per permettere la visione del DNA.

5.7 Screening di mutazioni dei geni *CNTNAP5* e *CADPS2*

5.7a Disegno dei primers

Le sequenze e l'organizzazione genomica dei geni *IMMP2L*, *DOCK4*, *CNTNAP5* e *CADPS2* sono state ottenute dall'UCSC genome browser. I primers per le reazioni di amplificazione sono stati disegnati utilizzando il programma PRIMER 3 (http://www-genome.wi.edu/cgi-bin/primer/primer3_www.cgi o <http://frodo.wi.mit.edu/primer3>). Per lo screening di mutazioni, i primers sono stati disegnati in modo da coprire gli esoni e parte delle regioni introniche adiacenti, e almeno 300-400 bp del promotore. In particolare, l'esone 4 è stato suddiviso in tre frammenti.

5.7b Amplificazione tramite Polimerasi Chain Reaction (PCR)

Le reazioni di amplificazione sono effettuate in un volume finale di 15 μ l. Per ogni reazione di PCR è preparato un controllo negativo senza il DNA, per identificare eventuali contaminazioni. La composizione finale di una miscela di reazione standard è la seguente:

- 30 ng di DNA genomico in TE 0.1X
- 10X buffer per Amplitaq Gold DNA polimerasi (senza MgCl₂)

- 2,5 mM di MgCl₂
- 0.25 μM di primer Forward
- 0.25 μM di primer Reverse
- 0.2 mM di dNTPs
- 0.025 U/μl di AmpliTaq Gold DNA polimerasi (Applied Biosystems)
- H₂O a volume.

Per la PCR viene utilizzato il termociclatore a 96 posti della MJ Research e della Veriti (Applied Biosystems). In generale si effettua un primo step di 15 minuti a 95°C per l'attivazione dell'enzima AmpliTaq Gold, e per l'amplificazione si utilizza un protocollo definito touchdown. In questo tipo di protocollo lo step di annealing è dapprima impostato ad una temperatura più alta rispetto a quella calcolata sui primers utilizzati (T_{m1}), che viene diminuita di 0.5°C ad ogni successivo ciclo: in questo modo si conferisce una maggiore specificità iniziale alla reazione. Dopo 14 cicli si arriva alla temperatura di annealing (T_{m2}), che viene mantenuta costante per altri 25-30 cicli. Lo schema utilizzato per il protocollo di amplificazione touch-down è il seguente :

Lo schema utilizzato per il protocollo di amplificazione touch-down è il seguente:

- Pre-PCR step: 15 minuti a 95°C
- PCR 14 cicli:
 - 95°C 30 secondi o 1 minuto
 - T_{m1} 30 secondi con calo di 0.5°C ad ogni ciclo
 - 72°C 30 secondi o 1 minuto
- PCR 25-30 cicli:
 - 95°C 30 secondi o 1 minuto
 - T_{m2} 30 secondi o 1 minuto
 - 72°C 45 secondi o 1 minuto

L'amplificazione di regioni molto ricche in G-C, nelle quali le due coppie di basi sono tenute insieme da tre legami idrogeno, può risultare più difficoltosa; in tal caso la reazione è condotta con

l'aggiunta di 0.1 mM finale di 7-deaza-2'-deoxiguanosina-5'-trifosfato (Deaza GTP) e/o 5% di DMSO. Questi reagenti aiutano la reazione della polimerasi prevenendo la formazione di strutture secondarie del DNA e facilitandone la denaturazione.

5.7c Messa a punto del protocollo di PCR per l'amplificazione degli esoni del gene

La messa a punto di una reazione di PCR consiste nell'insieme di modifiche e prove che si possono effettuare per migliorare la resa o per eliminare eventuali prodotti aspecifici della reazione. La messa a punto, fatta in genere su un DNA controllo particolarmente puro e integro, prevede di:

- provare almeno 3 diverse condizioni di temperatura (nel nostro caso, sono stati sperimentati i Touch down 65°-60°C, 62°-57°C, 60°-55°C).

- variare la concentrazione di Mg²⁺, dalla quale dipende la specificità di attacco dei primers e la resa dell'amplificazione. Si testano in genere tre concentrazioni di Mg²⁺ (2 mM, 2.5 mM, 3 mM).

Le coppie di primers e le condizioni di amplificazione specifiche utilizzate in questo studio sono riportate nella tabella 2.

5.7d Elettroforesi su gel di agarosio.

Per la preparazione del gel si scioglie la quantità necessaria di agarosio in un volume di tampone TBE 1X che va dai 40 ml ai 400 ml, a seconda delle dimensioni e dello spessore desiderato del gel. Dopo essere stato portato ad ebollizione, il gel viene raffreddato a 60°C circa, successivamente vengono aggiunti 5 µl di Gel Red TM. Il gel viene versato su di un supporto orizzontale e lasciato polimerizzare per circa 30 minuti.

Si trasferisce poi il supporto in un apparato per elettroforesi orizzontale, contenente il tampone di corsa (TBE 1X). Per il caricamento si aggiunge ai campioni il loading buffer per DNA Orange G (Sigma). La corsa è effettuata a 5 volts/cm per il tempo necessario. Al termine della corsa elettroforetica il DNA viene visualizzato su un transilluminatore a luce UV e fotografato.

La percentuale di agarosio utilizzata varia dallo 0.8% al 2.5%, in funzione della grandezza dei frammenti da analizzare.

5.7e Reazione di sequenziamento e analisi delle sequenze

Per le reazioni di sequenza abbiamo utilizzato il kit ABI PRISM Big Dye Terminator Cycle Sequencing (V1.1 Applied Biosystems) ed i primers specifici di ciascun frammento.

I prodotti di amplificazione sono dapprima purificati usando l'ExoSAP-IT PCR Clean-up Kit, che contiene i due enzimi: exonuclease I (Exo), il quale rimuove i primers non utilizzati ed eventuali filamenti di DNA a singolo filamento, e Shrimp Alkaline Phosphatase (SAP), che rimuove i nucleotidi non incorporati. La purificazione viene realizzata aggiungendo 1 μ l di ExoSAP a 6 μ l di prodotto di amplificazione.

Per la reazione di purificazione è stato utilizzato il seguente programma:

- ✓ 37°C 15 minuti
- ✓ 80°C 15 minuti

Le reazioni di sequenza sono effettuate in un volume finale di 10 μ l.

La composizione finale di una miscela di reazione standard è la seguente:

- 5-20 ng di prodotto di amplificazione;
- 0.5 μ l di Big Dye Terminator Mix (contenente la DNA polimerasi e i dideossinucleotidi marcati con quattro differenti fluorocromi);
- 1.75 μ l di mix di buffer di reazione 5X (200 mM Tris HCl pH 8 /5 mM MgCl₂);
- 0.16 μ l di primer 10 μ M (forward o reverse)
- H₂O a volume

Per la reazione di sequenziamento è utilizzato il seguente programma:

Per 25 cicli :

- 96°C 1 minuto
- 96°C 10 secondi
- 50°C 5 secondi
- 60°C 4 secondi

I prodotti di sequenza sono quindi purificati attraverso una reazione di precipitazione selettiva; infatti, per effettuare una corretta lettura delle sequenze, è opportuno eliminare i dideossinucleotidi non incorporati. Ad ogni miscela di reazione (10 μ l) sono aggiunti:

- 2.5 volumi (55 μ l) di etanolo assoluto (100%);
- 1/10 di volume (2 μ l) di NaAcetato 3M pH 5.2;
- 10 μ l di H₂O.

Per comodità, viene preparata una "precipitation mix", contenente etanolo assoluto "freddo", Na-Acetato 3M e H₂O, nelle giuste proporzioni, in modo da poter aggiungere 67 μ l di mix a ciascun campione.

I campioni vengono lasciati per 15 minuti a temperatura ambiente, quindi centrifugati 30 minuti a 3000 rpm alla temperatura ottimale di 4°C. Il precipitato ottenuto è lavato con 70 µl di etanolo “freddo” al 70%. Si rimuove il supernatante capovolgendo i campioni su carta pulita e l’etanolo residuo è eliminato completamente centrifugandoli “a testa in giù” per 1 minuto a 50-60 g.

I prodotti di sequenza, dopo essere stati risospesi in 10 µl di formammide, vengono denaturati per 5” a 95°C, quindi posti in ghiaccio e corsi sul sequenziatore automatico ABI PRISM 3730 (Applied Biosystems).

Per l’analisi e la lettura dei file di sequenza è stato utilizzato il programma Sequencher (Gene Codes Corporation).

5.8 Genotipizzazione di SNPs mediante digestione con enzimi di restrizione

Per la scelta degli enzimi di restrizione è utilizzato il programma INSIZER, disponibile all’indirizzo <http://zeon.well.ox.ac.uk/git-bin/insizer>. Il programma permette di identificare gli enzimi di restrizione in commercio in grado di discriminare i due diversi genotipi dati da un particolare polimorfismo.

L’enzima di restrizione viene scelto sulla base delle seguenti caratteristiche:

- che abbia un unico sito di taglio all’interno del prodotto di PCR da analizzare;
- che produca frammenti di dimensioni sufficientemente diverse, in modo da permettere l’identificazione dei diversi genotipi in seguito a separazione elettroforetica.

Per l’enzima di digestione MnlI la mix preparata per 1 campione in un volume finale di 16 µl è la seguente:

- 1,6 µl di buffer NEB4
- 0,5 µl di enzima MnlI (5U/µl)
- 0,16 µl di BSA 100x
- 5,74 µl di acqua
- 8 µl di prodotto di PCR

L’incubazione è durata 2,5 ore a 37°C seguita da uno step di 20 minuti a 65°C

5.9 Array Comparative Genomic Hybridization

L’array-CGH è stato eseguito utilizzando microarray 44K (Agilent Human Genome CGH Microarray Kit 44A) con una risoluzione media di circa 75 Kb.

Il protocollo utilizzato per l’array-CGH è costituito da 7 step (figura 4).

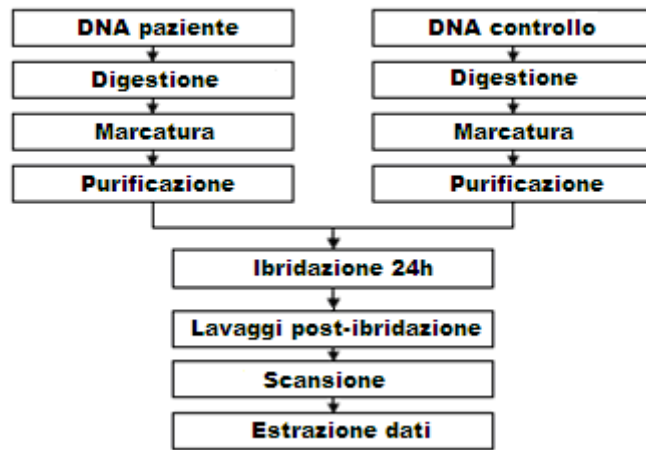


Figura 5.7: workflow array-CGH

5.9.1 Step 1. Digestione del DNA

500 ng di DNA del paziente e del controllo (dello stesso sesso) sono stati digeriti utilizzando la seguente mix di reazione:

Mix di reazione	Quantità x1 (µl)
Buffer 10X (Promega)	2.6
BSA (10 µg/µl) (Promega)	0.2
AluI (10U/µl)(Promega)	0.5
RsaI (10U/µl) (Promega)	0.5
TOTALE	3.8

Tab5.1: mix di digestione

I due enzimi di restrizione creano estremità piatte tagliando in corrispondenza delle sequenze 5'-AG↓CT-3' 3'-TC↓GA-5' (AluI) e 5'-GT↓AC-3' 3'-CA↓TG-5' (RsaI)

La reazione avviene a 37°C per 2 ore. Successivamente si inattivano le attività enzimatiche ponendo i campioni a 65°C per 20 min.

5.9.2 Step 2. Marcatura del DNA

Ciascun campione digerito viene mescolato con 5 μ l di random primers (Agilent Technologies) e incubato a 95°C per 5 min. Viene quindi preparata la seguente mix di marcatura per i pazienti (tabella 3) e per i controlli (tabella 4)

Mix di reazione	Quantità x1 (μ l)
Buffer 5X (Agilent)	10
dNTPs 10X (Agilent)	5
Cyanine 5-dUTP (1 mM)(Agilent)	3
Exo Klenow fragment	1
TOTALE	19 μ l

Tabella 5.2: mix di marcatura usata per i pazienti.

Mix di reazione	Quantità x1 (μ l)
Buffer 5X (Agilent)	10
dNTPs 10X (Agilent)	5
Cyanine 3-dUTP (1 mM)(Agilent)	3
Exo Klenow fragment	1
TOTALE	19 μ l

Tabella 5.3: mix di marcatura usata per i controlli.

La reazione di marcatura avviene per 2 ore a 37°C ed inattivata a 65°C per 10 min.

5.9.3 Step 3. Purificazione del DNA marcato

Dopo la marcatura, è necessario rimuovere dai campioni i nucleotidi marcati che non si sono incorporati nei filamenti di DNA. A tale fine i campioni vengono sottoposti a lavaggi con buffer TE 1X (Tris-HCl 10 mM, EDTA 1mM, pH 8,0) e a centrifugazioni attraverso colonnine contenenti membrane costituite da cellulosa (Amicon Ultra 30kDa, Millipore), che trattengono il DNA ad alto peso molecolare. Nello specifico, i passaggi previsti da questa procedura sono i seguenti: si diluisce ogni campione con 430 μ l di TE 1X e lo si trasferisce in una colonnina; si centrifuga a 13000 x g per 17 minuti e si rimuove l'eluato; si aggiungono 480 μ l di TE 1X e si centrifuga nuovamente a 13000 x g per 17 minuti; per eluire il DNA marcato, si capovolge la colonnina in una provetta pulita e si centrifuga a 1000 x g per 2

minuti. Il volume di DNA ottenuto dalla purificazione viene quindi portato a 80,5 µl con TE 1X. Il passo successivo è quello di verificare la concentrazione del DNA al NanoDrop e di calcolare l'attività specifica, cioè il livello di incorporazione dei fluorocromi attraverso la seguente formula:

Il DNA marcato e purificato viene controllato al nanodrop per determinarne la concentrazione finale (A260nm) e l'incorporazione del fluorocromo (A550 nm per Cy3 e A650 nm per Cy5).

Quindi viene calcolata l'attività specifica per ogni campione utilizzando la seguente formula:

$$[1] \quad \text{Attività specifica} = \text{pmol/} \mu\text{l di Cy3 o Cy5} / \mu\text{g/} \mu\text{l di DNA}$$

I valori ottimali da ottenere per quanto riguarda il recupero di DNA sono 5-7 µg, mentre relativamente all'attività specifica sono 25-40 pmol/ µg per la cianina-3 e 20-35 pmol/ µg per la cianina-5.

5.9.4 Step 4. Ibridazione

L'ibridazione è stata eseguita a partire da 5 µg di DNA marcato del paziente e del controllo che vengono mescolati. Successivamente sono stati aggiunti 5 µl di COT-1 DNA (1 mg/ml) (Invitrogen), 11 µl di Blocking Agent 10X (Agilent) e 55 µl di Hybridization Buffer 2x (Agilent). Il DNA Cot-1 (Invitrogen) blocca le sequenze altamente ripetute in modo da evitare che si ibridino in modo aspecifico agli oligonucleotidi

presenti sul vetrino.

I campioni così preparati vengono denaturati a 95°C per 3 min e poi incubati a 37°C per 30 min per permettere al COT-1 DNA di saturare le sequenze ripetute presenti nel genoma e ridurre così i segnali aspecifici.

L'ibridazione viene quindi effettuata su vetrini 4x44K che vengono incubati in rotazione (20 rpm) per 24 ore a 65°C.

5.9.5 Step 5. Lavaggi post-ibridazione

Questo passaggio consente di eliminare l'eccesso di DNA non ibridato. A questo scopo si utilizzano due diverse soluzioni di lavaggio: Oligo aCGH Wash Buffer 1, a temperatura ambiente per 5 minuti, e Oligo aCGH Wash Buffer 2, scaldata a 37°C, per 1 minuto.

5.9.6 Step 6 e 7. Scansione degli array e analisi dei risultati

La scansione degli array è stata eseguita con uno scanner Agilent (G2505C) che è in grado di rilevare i livelli di fluorescenza (Cy5 e Cy3) di ogni spot e fornisce l'immagine su cui poi si effettuerà l'elaborazione dei dati grazie all'utilizzo del software Feature Extraction V.10.1.1.1.

Esso elabora per ogni spot il logaritmo in base 2 del rapporto tra le intensità di fluorescenza emesse da Cy5 e Cy3 ($\log_2 \text{Cy5/Cy3}$) inoltre sottopone l'informazione ottenuta dalla scansione del vetrino

ad un processo di normalizzazione globale (sottrazione del background e correzione mediana) in grado di rimuovere le distorsioni sistematiche presenti.

Il rapporto delle intensità di fluorescenza tra i due fluorofori sono proporzionali al rapporto di quantità di materiale genomico tra campione e controllo. Se le intensità dei fluorofori sono uguali la regione analizzata non presenta alterazioni: avrò 2 copie di DNA da testare e 2 copie del DNA di controllo, per cui il rapporto

tra i 2 fluorocromi è pari a 1 ($\log_2 2/2 = 0$). Se invece il rapporto Cy5/Cy3 è alterato si è in presenza di amplificazioni o delezioni: in caso di trisomia il rapporto sarà pari a 1,5 e cioè $3/2$ ($\log_2 3/2 = 0,5$), viceversa in caso di monosomia il rapporto sarà pari a 0,5 e cioè $1/2$ ($\log_2 1/2 = -1$)

I dati elaborati da Feature Extraction vengono visualizzati grazie al software DNA Analytics V.4.0.76..

Per la valutazione dei dati si è fatto riferimento al Genome Assembly del 2006 (hg18, NCBI Build 36.1) e sono stati consultati il Database of Genomic Variants (<http://projects.tcag.ca/variation/>), il Database of Chromosomal Imbalance and Phenotype in Humans using Ensembl Resources (DECIPHER) (<http://decipher.sanger.ac.uk/>) e il sito UCSC Genome Browser (<http://genome.ucsc.edu/cgi-bin/hgGateway>)

FRAMMENTO	PRIMERS FORWARD	PRIMERS REVERSE
DOCK4_x1RT:	GGGACTGTGCATTGTTGTGA	GTATTTCTCGTGCTCCGTAGGT
DOCK4_x1RT_b:	GGAGGCCAAGAAGCAACTACAATTCC	TCACTGAAGCAAGCGAACACG
rt_dock4_ex35_F	AAATTGAGCGGGAAACATGG	AACATGCGACTTGGCTTACCT
rt_dock4_ex37_F	TGAACTGAACAAGGAGGAGATG	CCTTCCAATCTGAGCCAAA
rt_dock4_intr37_F	CCTCGTTCATTACCTACTGCT	CAAGCCACATTCCCTCAAAT
rt_dock4_ex38_F	ATGACGAGCTACTGGAATGGTC	AAGTTCTGGATGATGGTGAGG
rt_dock4_ex39_F	GGAAGATTGCAGAGCAGTATGA	TAATCACAAGGGGAGGCTATGG
rt_dock4_ex40_F	GCTTAGAATGTTTCTGTGCTT	TCCAACCTGAAGAAGCTCTGGT
CNTNAP5_rtà??Ex1	GATGGAGTGAAAGAGCGAGT	ATCCCAAATTCGAGTAGCC
CNTNAP5_rtà??Ex8	CCATCACATTTGTCAACTCCAG	ACAGAAGCAGACCATCCTTGTT
CNTNAP5_rtà??Ex10	ATCTACCGATTCCCAATGTTT	ACCTTGCTGAAGTAAATGAGG

Tabella 5.4: primers utilizzati nell'esperimento di real time di *DOCK4* e di *CNTNAP5* nelle famiglie 3043,2031, 15_0084

FRAMMENTO	PRIMERS FORWARD	PRIMERS REVERSE
CADPS2-x3	GAGCAAAGAGACAGTGTGAGC	CTGCACTTAGGGCCATTCTATT
CADPS2-x4	TTTGCTTAGTCTGTGGGACTG	CTGCTTGTTCATCTGCGTTATC
CADPS2-intr5	ATGCTCACCTCTGACAGCATAA	TGAGTGCCAACAATGGAGAA
CADPS2-x6	ACAATGGAAGTGAAGGAGAAA	GCATGTTTCTTTGGGAGGATTA
CADPS2-x7	CTCAAGGAGATTTACCACCAC	AACTCACCTTCCCAGTTCTTT
CADPS2-x26	TGTGCGATGGAAGAAATTGA	AGGATTCAGTTTCTGGGTTTGA
CADPS2-x28	TTCACAACAGAAACCAGGAATG	GCAAGGGAAGCTACTTACATCAA
CADPS2-x29	ATTTGCGTGTGGTTGACTGATA	TCCTTATAGCTTAGTGGGCAATTC

Tabella 5.5: primers utilizzati nell'esperimento di real time utilizzati per determinare i confini esatti della delezione nel gene *CADPS2*.

QMPSF <i>IMMP2L-DOCK4</i>			
Nome	Sequenza F (Marcato FAM)	Sequenza R	Condizioni di amplificazione
<i>IMMP2L</i> ex 2	GACAAGATTGCCCTTCGGTA	GACCCGATCCAAGAAAGTCA	60 + 20 (30"-30"-30"), 2 mM Mg 5uDNA Vf=15ul
<i>IMMP2L</i> ex 3	ATCCAGTGAGCCTTTTGA	TTTCCAGTGTTCAAAGCA	60 + 20 (30"-30"-30"), 2 mM Mg 5uDNA Vf=15ul
<i>IMMP2L</i> ex 6	CTTTATTGCTTTCTTTCTGTTTCTCA	GCAGTCATTCTCTTCTCTCTGT	60 + 20 (30"-30"-30"), 2 mM Mg 5uDNA Vf=15ul
<i>LRRN3</i> ex 4	CCCAAGCTGGAATCACTCAT	GGCTCCATGAATCGAATGTT	60 + 20 (30"-30"-30"), 2 mM Mg 5uDNA Vf=15ul
<i>DOCK4</i> ex 52	GCAGTCAAACAGCAAAGCA	AGGCAAGGTTTATCAGTGCAG	60 + 20 (30"-30"-30"), 2 mM Mg 5uDNA Vf=15ul
RNF20	GAGACAGCCGAATCACGAGT	CCCAAGGTGAGTCTTCCTGA	60 + 20 (30"-30"-30"), 2 mM Mg 5uDNA Vf=15ul

Tabella 5.6: primers utilizzati nell'analisi di PCR moltiplex per identificare la presenza di CNV all'interno dei geni *IMMP2L-DOCK4*

QMPSF <i>DOCK4</i>			
Nome	Sequenza F (Marcato FAM)	Sequenza R	Condizioni di amplificazione
<i>DOCK4</i> ex 1	GGCGTGTTTCGCTTAGTTCAG	CTTCTCCGGCTCACAACAAT	60 + 20 (30"-30"-30"), 2 mM Mg 5uDNA Vf=15ul
<i>DOCK4</i> ex 8	TGGCGTAGGACATCACAGTC	TCTCCCAGGTTGGAACACAT	60 + 20 (30"-30"-30"), 2 mM Mg 5uDNA Vf=15ul
<i>DOCK4</i> ex 15	AACAACAGTCCAGGTGGTC	CCTCCCCCTTCTGGCTAATC	60 + 20 (30"-30"-30"), 2 mM Mg 5uDNA Vf=15ul
<i>DOCK4</i> ex 25	GGATATTCTGCTGAGGACCA	TAACATCAAGCCACCTGCAA	60 + 20 (30"-30"-30"), 2 mM Mg 5uDNA Vf=15ul
<i>DOCK4</i> ex 31	TTTTATCCCTGCCCTGATTG	TCTGCTCCAGTCCATCATA	60 + 20 (30"-30"-30"), 2 mM Mg 5uDNA Vf=15ul
<i>DOCK4</i> ex 52	GCAGTCAAACAGCAAAGCA	AGGCAAGGTTTATCAGTGCAG	60 + 20 (30"-30"-30"), 2 mM Mg 5uDNA Vf=15ul
RNF20	GAGACAGCCGAATCACGAGT	CCCAAGGTGAGTCTTCCTGA	60 + 20 (30"-30"-30"), 2 mM Mg 5uDNA Vf=15ul

Tabella 5.7: primers utilizzati nell'analisi di PCR moltiplex per identificare la presenza di CNV all'interno del gene *DOCK4*

RT PCR semiquantitativa			
Nome	Sequenza F	Sequenza R	Condizioni di amplificazione
<i>IMMP2L_x4fusF</i>	ATCTCCTTCAAGAGCAATCACT		2,5 mM Mg, TD 65-58 + 30 cicli (30",30",30")
<i>DOCK4_x26fusR</i>		TCTCCTGTCCCTATTACGACAAA	2,5 mM Mg, TD 65-58 + 30 cicli (30",30",30")
<i>RT_DOCK4_ex17F</i>	CATGAAGGCCACAAAGGAGT		2,5 mM Mg, TD 65-58 + 30 cicli (30",30",30")
<i>RT_DOCK4_ex18R</i>		AGAGAGACAGCCAGTGATCTTG	2,5 mM Mg, TD 65-58 + 30 cicli (30",30",30")
<i>RT_DOCK4_ex47F</i>	AGCTGTCAACCGATATTCTTCC		2,5 mM Mg, TD 65-58 + 30 cicli (30",30",30")
<i>RT_DOCK4_ex48R</i>		AGCTTGAGGTAGATGGACTTGG	2,5 mM Mg, TD 65-58 + 30 cicli (30",30",30")

Tabella 5.8: primers utilizzati per vedere l'espressione semiquantitativa dei trascritti di *DOCK4* canonici e di fusione con *IMMP2L*

SCREENING MUTAZIONE CNTNAP5			
Nome	Sequenza F	Sequenza R	Condizioni di amplificazione
CNTNAP5 ex 1	CCCAGGCTGCTCTTATAGGC	CCAAAGCAGACACAATCGAG	2 mM Mg, 5% DMSO, TD 65-58 + 30 cicli (1',30'',1')
CNTNAP5 ex 2	TGGTCTCTCCCTCTGTCC	TTTCTGAAGACGGCCATT	2,5 mM Mg, TD 65-58 + 30 cicli (1',30'',1')
CNTNAP5 ex 3	CCATTTGGGGGTAGAGTCAT	GGCATGGGCCTTAAAAATTG	2,5 mM Mg, TD 65-58 + 30 cicli (1',30'',1')
CNTNAP5 ex 4	GCAGGATTTTCACCCAAGTC	CACCATCACAACCTACATTCCTACA	2,5 mM Mg, TD 65-58 + 30 cicli (1',30'',1')
CNTNAP5 ex 5	GTTTCCTGTTGAGACTAGACCTG	TGCTATACATGCCTTAAACATTTTC	2,5 mM Mg, TD 65-58 + 30 cicli (1',30'',1')
CNTNAP5 ex 6	GGCCTGTAGGGGAGACATCATT	TTCTCTGGCTGCACTCCTTT	2,5 mM Mg, TD 65-58 + 30 cicli (1',30'',1')
CNTNAP5 ex 7	CCATTTGGCTCAAGCATACA	AGACGCTGATGGGAECTCA	2,5 mM Mg, TD 65-58 + 30 cicli (1',30'',1')
CNTNAP5 ex 8	GTCAGTCCAGGTGGTTGTT	TGCTGCCTCTTCAGACTTGG	2,5 mM Mg, TD 65-58 + 30 cicli (1',30'',1')
CNTNAP5 ex 9	CTTGATCCAGCCGAGTGAG	TGACCAAACCAGAACCATCA	2,5 mM Mg, TD 65-58 + 30 cicli (1',30'',1')
CNTNAP5 ex 10	TGTGAATGAGAAAGCACAAA	GTGGCTAATGTCGGTGGATT	2,5 mM Mg, TD 65-58 + 30 cicli (1',30'',1')
CNTNAP5 ex 11	TGAGGCAGTTGGATATGAGACA	GCCCCATGCTAAATACATCC	2,5 mM Mg, TD 65-58 + 30 cicli (1',30'',1')
CNTNAP5 ex 12	GAAGGAGGGAAATGAATACACA	CTCCCCAAAGGAGACCAGTT	2,5 mM Mg, TD 65-58 + 30 cicli (1',30'',1')
CNTNAP5 ex 13	CCATGGGGAAACACAATTTTC	ACCAGACCCTCTTGTGTG	2,5 mM Mg, TD 65-58 + 30 cicli (1',30'',1')
CNTNAP5 ex 14	TCCAAGATATTTTCAGCTCAGTTTC	GGGGGAGCTCCAGTTTGTAT	2,5 mM Mg, TD 65-58 + 30 cicli (1',30'',1')
CNTNAP5 ex 15	GCAACTGGCCAAATCACTT	AAAACCTGTTTGTGTATCAGTTACA	2,5 mM Mg, TD 65-58 + 30 cicli (1',30'',1')
CNTNAP5 ex 16	GCAACTGGCCAAATCACTT	AAAACCTGTTTGTGTATCAGTTACA	2,5 mM Mg, TD 65-58 + 30 cicli (1',30'',1')
CNTNAP5 ex 17	GCTGCAGGATTTCAAAGA	TGACATTTTGATCATATCCAGA	2,5 mM Mg, TD 65-58 + 30 cicli (1',30'',1')
CNTNAP5 ex 18	AGACCTTCATGACAACAGAGAACT	AAACAGCGGGAGAGTAGATGA	2,5 mM Mg, TD 65-58 + 30 cicli (1',30'',1')
CNTNAP5 ex 19	GAAACGAGAAAGCAGCTAAGG	GGGACCAACTTGAGGTTGAA	2,5 mM Mg, TD 65-58 + 30 cicli (1',30'',1')
CNTNAP5 ex 20	AAGACCTGGGGCCCTAATAA	CAAGGCACATTGGCTGAAG	2,5 mM Mg, TD 65-58 + 30 cicli (1',30'',1')
CNTNAP5 ex 21	GGGGTCTGCTATCTGTATGC	CAAAGGCTCCAGACCTGAAT	2,5 mM Mg, TD 65-58 + 30 cicli (1',30'',1')
CNTNAP5 ex 22	GGAGGTGGGGAAAGGAGATA	AAAAACTGAGTCACATTAGCCAAC	2,5 mM Mg, TD 65-58 + 30 cicli (1',30'',1')
CNTNAP5 ex 23	ATTTGAGGGCACCTGGTGTGA	GAGTGCTGGATGGCCATAAT	2,5 mM Mg, TD 65-58 + 30 cicli (1',30'',1')
CNTNAP5 ex 24	GCTTCTCTCTACGCATTCC	CAGAACTGAACATGCAGGAGT	2,5 mM Mg, TD 65-58 + 30 cicli (1',30'',1')
Dock4_ex1_snp***	GGCGTGTTCGCTTAGTTTACAG	CTTCTCCGGCTCACAACAAT	2,5 mM Mg, TD 65-58 + 30 cicli (1',30'',1')

Tabella 5.9: primers utilizzati per lo screening di sequenza del gene *CNTNAP5*. L'ultima coppia di primers è stata utilizzata per verificare la presenza della variante nel primers utilizzato per la QMPSF

FRAMMENTO	PRIMERS FORWARD	PRIMERS REVERSE
CADPS2_X17SNP	GTTTTCCCTTTGGACGACCT	TAGCAGGAGATGCCTGGTTC
CADPS2_mRNAx1/x4	GAGCGGAGGATCCGCCTGGAG	ATCTGCGTTATCCAGCTGACA
CADPS2-express_x3a/x5a	GACTTGTGCAAACAGCCAAA	AAAATGCAGAGTTCTGTGAACG

Tabella 5.10: primers utilizzati per verificare che il gene *CADPS2* è soggetto ad imprinting.

SCREENING MUTAZIONE CADPS2			
Nome	Sequenza F	Sequenza R	Condizioni di amplificazione
CADPS2x1	GGCCAGTTTTCTTTTCGTTTC	AATACGGTGTCTGCGTTTTC	2,5 mM Mg,58°C + 40 cicli (1',30",1')
CADPS2x2	TTGAGATGGAGGTCTAACCATT	CAGCCTTCCTGTGACACCA	2,5 mM Mg,58°C + 40 cicli (1',30",1')
CADPS2x3	TGCTGAAAGCAAGTGGTGTT	AAAGGACATCTTCATAGGCTTG	2,5 mM Mg,58°C + 40 cicli (1',30",1')
CADPS2x4	GCCTTAGTCTGTGGGACTGC	TGTTAAAATTATGCTGGCTCTTTT	2 mM Mg,56°C + 40 cicli (1',30",1')
CADPS2x5	CAAAAATGCCAAAGCCTGAG	AAGTTACTGCAAAGGTCAACAGAA	2,5 mM Mg,58°C + 40 cicli (1',30",1')
CADPS2x6	CCTTATTACAGTTCATGCTGCTG	AGCATGTTTCTTTGGGAGGA	2,5 mM Mg,58°C + 40 cicli (1',30",1')
CADPS2x7	TTGCCTTCTATGGAGACACATC	TTCACATATGGCATACTGATCTACC	2mM Mg,58°C + 40 cicli (1',30",1')
CADPS2x8	TTGGATACATTTGGGTGTTGA	CTCCATGTGAAAGCCTGCTA	2 mM Mg,56°C + 40 cicli (1',30",1')
CADPS2x9	TGGTCTCAAAAATAAATACCTATGGA	TGTACTGCTTTTTATTTAAAGTGGCTTA	2mM Mg,58°C + 40 cicli (1',30",1')
CADPS2x10	TTTAATTGGGAAAAGAGTATTTTGAT	CATCGAGAGGGGTTTAAATTG	2mM Mg,58°C + 40 cicli (1',30",1')
CADPS2x11	AGCCATTCTTTTCCATGAGG	CTGAATACATTTGCCTCAATTTT	2,5 mM Mg,56°C + 40 cicli (1',30",1')
CADPS2x12	TAAAATCATGAGTAGAAATTGAGAAAA	GGGTTTGGCTAGAAAGTAGGG	2,5 mM Mg,56°C + 40 cicli (1',30",1')
CADPS2x13	GCTTGAGTCACGCATTTTGT	TCCCTCACTCTCACTACCCTTA	2,5 mM Mg,56°C + 40 cicli (1',30",1')
CADPS2x14	AAGGCCTAAAAGTGC GTTCA	GGTGATAATCGCAAAGCACA	2,5mM Mg, 5% DMSO,62°C + 40 cicli (1',30",1')
CADPS2x15	CAGTGAAC TTTTCTTGGATTTTT	GGGGGAAGTAAGGGTTGAGT	2,5 mM Mg,56°C + 40 cicli (1',30",1')
CADPS2x16	TGCCAAGTAATACAAGCAGGAA	ACGGGGTCTGTCTCCCTAGT	2mM Mg,56°C + 40 cicli (1',30",1')
CADPS2x17	TTTATTTTATGCAACATCTGTCTTTTT	CAGGAAAGGAAAGGACTTCTCA	2 mM Mg,58°C + 40 cicli (1',30",1')
CADPS2x18	TCAACATTCTAGTCTCATCATTACGTC	TATGGGCAGGCCACTTATTC	2 mM Mg,58°C + 40 cicli (1',30",1')
CADPS2x19	GCAATCACTAAAATGCTTGTC A	GGACTAATTGGCTTTAAAATGGA	2 mM Mg,58°C + 40 cicli (1',30",1')
CADPS2x20	TCATACCTGTTTGTGAAACTAAGACA	GGTGTGAGGGCACAGTAAAA	2 mM Mg,58°C + 40 cicli (1',30",1')
CADPS2x21	TGGCCCATATTTTCCAAATG	TGGCCATGTGTGAAGTCATAA	2 mM Mg,56°C + 40 cicli (1',30",1')
CADPS2x22	TGACCTTATGACTTCACACATGG	TGACAAATGCCATGCAGTCT	2 mM Mg,56°C + 40 cicli (1',30",1')
CADPS2x23	ACTGATGGCAATTGATGCAG	GGTGCATAATGAAACCTTTGC	2,5 mM Mg,56°C + 40 cicli (1',30",1')
CADPS2x24	GCATCTTGTAGTGTGATTTCACTTT	AAAGGGCATTTCCCATGC	2 mM Mg,56°C + 40 cicli (1',30",1')
CADPS2x25c/d	TTTTGTGTCCATGTCAACTAATGA	TTTTTCCTGCCCTAATCTTTT	2,5 mM Mg,56°C + 40 cicli (1',30",1')
CADPS2x26	TGAATTTTTGTGCGATGGAA	TGTGAAAGAAAGAGATAATCATGAGA	2,5 mM Mg,56°C + 40 cicli (1',30",1')
CADPS2x27	TTGAATGAGTTCTCTTCTGTGGTC	TAAGGGAATGGCACTGGTTT	2 mM Mg,56°C + 40 cicli (1',30",1')
CADPS2x28	TGAAGATGCTCACTTTTGGTTT	CTGGCCAGCTTCAACTTTTT	2,5 mM Mg,56°C + 40 cicli (1',30",1')
CADPS2x29	CCCCCTCATTATCGAATCTG	CAGTTTTCTTGTATGTCAGTATCTTGC	2,5 mM Mg,56°C + 40 cicli (1',30",1')
CADPS2x30	TCATTTTTGAAATGCCAACTG	GAGGATGCTCTTCTCCAAAAA	2,5 mM Mg,56°C + 40 cicli (1',30",1')

Tabella 5.10: primers utilizzati per lo screening di sequenza del gene *CADPS2*.

6. RISULTATI:

Analisi dei geni *IMMP2l*, *DOCK4* e *CNTNAP5* in autismo e dislessia.

6.1. Risultati Preliminari

6.1a: Analisi di associazione nei loci e *AUTS1* ed *AUTS5*

Il mio lavoro di dottorato parte dai risultati di uno studio di associazione e analisi di varianti strutturali condotto utilizzando SNP ad alta densità nei due loci di suscettibilità all'autismo, *AUTS1* ed *AUTS5*, precedentemente identificati dall'IMGSAC e localizzati, rispettivamente, sul cromosoma 7q e sul cromosoma 2q.

In particolare, utilizzando i dati resi disponibili dall'*International HapMap Project* per valutare il profilo di Linkage Disequilibrium (LD) nei due loci di interesse, sono stati selezionati in ciascun locus circa 3000 *tagging* SNPs, che erano in grado di catturare più dell'80% della variabilità totale presente in tutti i geni noti (173 geni sul cromosoma 2 e 270 geni sul cromosoma 7), nelle loro regioni regolatrici ed in regioni del genoma non codificanti ma molto conservate nell'evoluzione.

Inoltre sono stati analizzati 50 SNP, sparsi su tutto il genoma, per verificare l'eventuale presenza di stratificazione della popolazione, e 10 SNP sul cromosoma X, per controllare il sesso degli individui e per verificare la presenza di errori di genotyping.

Questo studio è stato condotto su un campione di 127 individui affetti e i relativi genitori, scelti tra le famiglie appartenenti al consorzio IMGSAC, che condividevano 1 o due alleli identici per discendenza, e su un campione di controllo, composto da 188 individui appartenenti alla collezione ECACC (European Collection of Cell Culture).

La genotipizzazione degli SNP è stata realizzata utilizzando la tecnologia Illumina (Golden Gate Assay), mentre per l'analisi dei genotipi è stato utilizzato il software BeadStudio (Illumina).

Tutti gli SNP (94%), che hanno superato il controllo della qualità, sono stati utilizzati per un'analisi di associazione sia caso-controllo che TDT, in modo da estrarre la massima informazione dal campione analizzato.

I risultati più importanti ottenuti dallo studio di associazione sono stati:

- Per il cromosoma 2 (tabella 6.1):
 - Analisi caso-controllo
 - *NOSTRIN* (rs7583629, $P = 3.2 \times 10^{-5}$; rs829957, $P = 9.0 \times 10^{-5}$; rs482435, $P = 1.4 \times 10^{-4}$)
 - *FAM130A2* (rs1020626, $P = 3.8 \times 10^{-4}$)
 - TDT test
 - *ZNF533* (rs11885327, $P = 8.0 \times 10^{-4}$; rs1964081, $P = 1.4 \times 10^{-3}$)
 - *UPP2* (rs6709528, $P = 8.0 \times 10^{-4}$)

SNP	Chr.	Posizione	Gene	Allele di rischio	TDT p-value	Case Control P-value
rs7583629	2	169381125	NOSTRIN	A	0,0027	3,22E-05
rs829957	2	169367080	NOSTRIN	T	0,0116	9,03E-05
rs482435	2	169384291	NOSTRIN	C	0,0084	1,39E-04
rs11885327	2	180276318	ZNF533	C	8,00E-04	0,0271
rs1964081	2	180299666	ZNF533	A	0,0014	
rs6709528	2	158678671	UPP2	T	8,00E-04	

Tabella 6.1: Riassunto degli SNPs sul cromosoma 2 che mostrano associazione. Sono riportati solo gli SNPs con un P-value minore di 0.005. Gli SNPs in grassetto hanno un P-value minore di 0.001

- Per il cromosoma 7 (tabella 6.2):
 - Analisi caso-controllo
 - *IMMP2L* (rs12537269, $P = 1.2 \times 10^{-4}$; rs1528039, $P = 6.3 \times 10^{-4}$)
 - *SMO* (rs6962740, $P = 3.4 \times 10^{-4}$)
 - TDT test
 - *Plexin A4* (rs4731863, $P = 1.2 \times 10^{-4}$)
 - *CUX1* (rs875659, $P = 2.0 \times 10^{-4}$)

SNP	Chr.	Posizione	Gene	Allele di rischio	TDT p-value	Case Control P-value
rs12537269	7	110184783	IMMP2L	A		1,20E-04
rs1528039	7	110230008	IMMP2L	C		6,28E-04
rs4731863	7	131674323	PLXNA4	T	1,00E-04	0,0321
rs875659	7	101696376	CUX1	C	2,00E-04	

Tabella 6.2: Riassunto degli SNPs sul cromosoma 7 che mostrano associazione. Sono riportati solo gli SNPs con un P-value minore di 0.005. Gli SNPs in grassetto hanno un P-value minore di 0.001

Sebbene i segnali più significativi identificati mediante i due approcci non coincidono, il confronto dei risultati ha permesso di focalizzare l'attenzione sui loci che erano supportati, anche se con forza differente, da entrambi i metodi di analisi. Quindi, i 56 SNPs (28 per ogni cromosoma), che avevano mostrato maggiore associazione, sono stati analizzati ulteriormente un campione di replicazione, composto da 390 casi (294 appartenenti all'IMGSAC e 204 casi di famiglie singleton dell'Olanda del Nord), 842 genitori e 188 individui di controllo. Il risultato più significativo è stato ottenuto con l'analisi TDT in corrispondenza del marcatore rs2217262 ($P = 9.2 \times 10^{-4}$) nel gene *DOCK4*. Questo è stato l'unico risultato che rimaneva significativo dopo aver applicato la correzione di Bonferroni per test multipli.

6.1b: Analisi di Copy Number Variant (CNV) nei loci AUTS1 ed AUTS5 tramite il software QuantiSNP

I dati dei genotipi degli SNP sono stati utilizzati anche per eseguire un'analisi di varianti strutturali all'interno dei loci AUTS1 ed AUTS5, mediante l'utilizzo del software QuantiSNP. QuantiSNP è un software per la predizione di varianti strutturali (delezioni o duplicazioni) con un alto potere di risoluzione. Per ogni CNV indentificato, il software associa la probabilità che il risultato sia o meno un falso positivo (fattore di Bayes).

I dati di predizione ottenuti da QuantiSNP, possono essere confermati, attraverso la visualizzazione grafica dei genotipi degli SNPs nella regione di interesse, tramite il software GenomeStudio (Illumina). In particolare i genotipi degli SNPs di tutti i campioni analizzati sono rappresentati all'interno di tre diversi cluster sotto forma di spots e indicati con colori diversi a seconda che l'individuo sia omozigote AA (rosso), eterozigote AB (viola), omozigote BB (blu) per un dato SNP. Il grafico mostra un segnale di intensità di fluorescenza, norm R (in ordinata) e la composizione allelica, norm tetha (in ascissa), normalizzate e relative alle posizioni dei clusters per un dato SNP (figura 6.1).

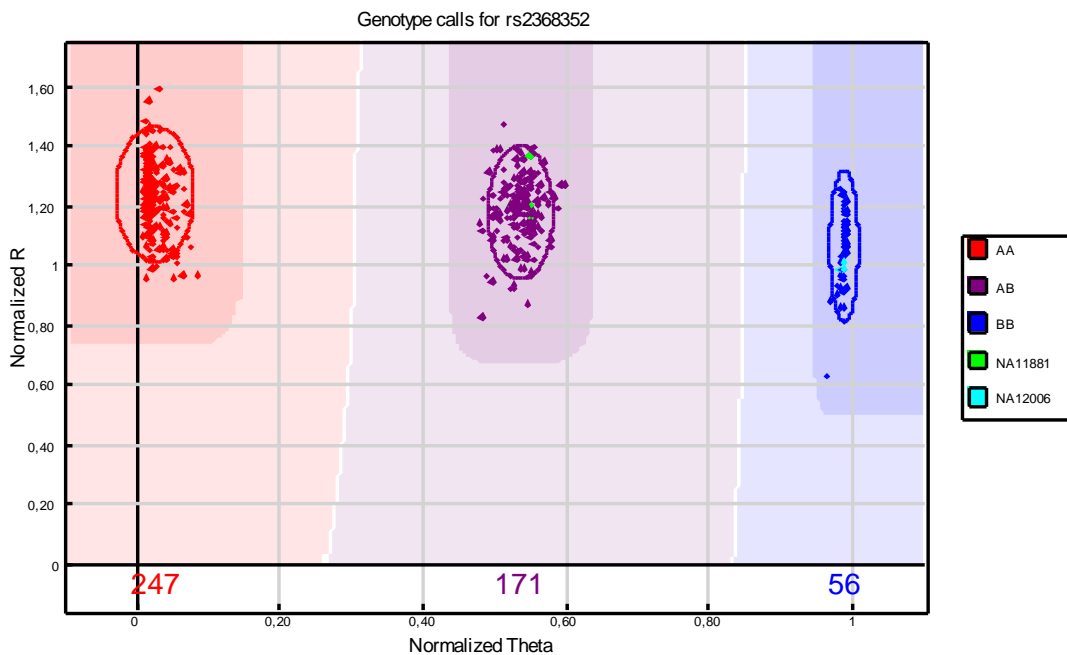


Figura 6.1: Rappresentazione grafica dei risultati forniti dal sistema Illumina per il *genotyping* di uno SNP (nell'esempio rs...). I tre *cluster* sono formati dai segnali di fluorescenza ottenuti per individui omozigoti AA (rosso), eterozigoti AB (viola), omozigoti BB (blu).

Tramite il software GenomeStudio è possibile calcolare vari parametri tra cui:

- La B-allele-frequency, che rappresenta la misura dell'allele minore B, è calcolata per ogni SNP, interpolando i valori di theta ottenuti per ogni cluster. In base al genotipo dello SNP (se è omozigote AA, eterozigote AB, oppure omozigote BB) può assumere valori di 0, 0.5, e 1.
- Log R ratio, che rappresenta il logaritmo del rapporto tra il valore di intensità di fluorescenza R osservato e quello atteso per un dato SNP. Pertanto, in un individuo normale, il rapporto tra le intensità di fluorescenza osservata e attesa è 1, ed il logaritmo è intorno allo 0 per tutti gli SNP.

In presenza di una delezione, la B-allele-frequency per gli SNPs interni alla regione deleva, assume solo valori di 0 o 1 (mancano gli eterozigoti), mentre il Log R ratio assume valori inferiori allo 0. In presenza di una duplicazione, invece, il log R ratio assume valori superiori allo 0, mentre la B-allele-frequency assume valori di 0 o 0.25 (corrispondente al genotipo AAB) o 0.75 (corrispondente al genotipo ABB) o 1 (Fig 6.2)

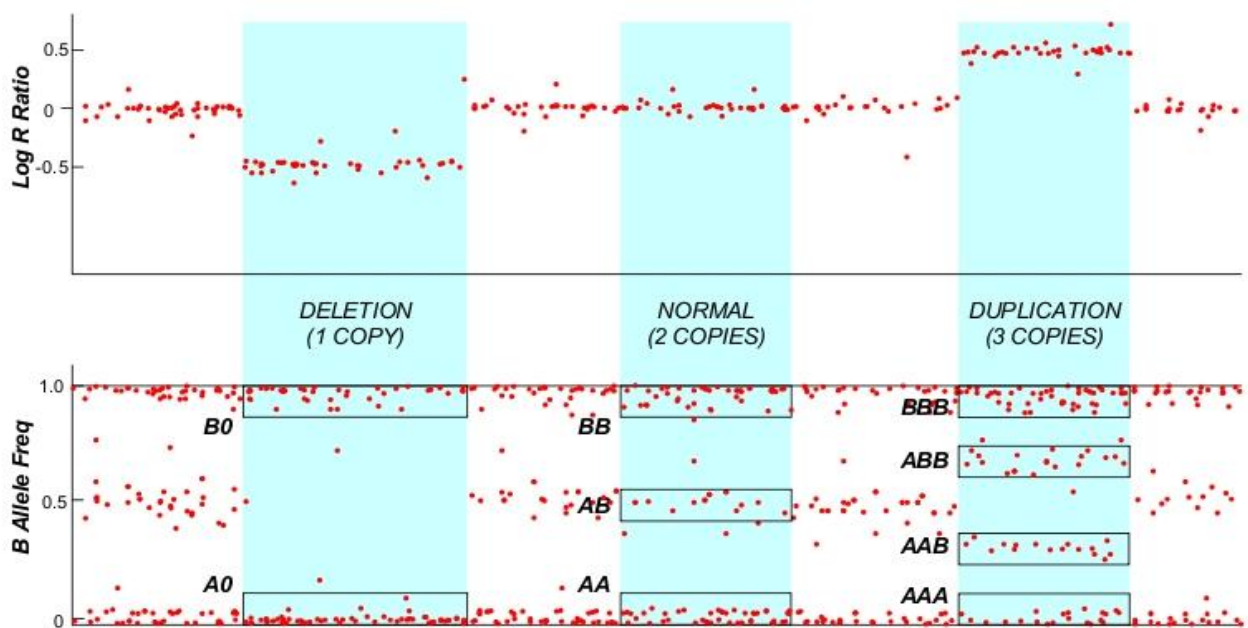


Figura 6.2: Identificazione dei *Copy Number Variants*: l'intensità del segnale di fluorescenza è proporzionale alla quantità di DNA genomico. Gli SNPs nelle zone delete o duplicate vengono mostrate come un aumento o una diminuzione del segnale rispetto alle regioni normali.

In questo modo sono state trovate altre 17 CNV nel cromosoma 7, e 6 CNV nel cromosoma 2, la maggioranza delle quali erano già descritte in DGV, suggerendo che esse non siano coinvolte nella suscettibilità all'autismo. Il risultato più interessante è stato una duplicazione, di circa 800 bp, sul cromosoma 7, identificata in una famiglia (13-3023) (Tabella XX). Questa duplicazione trasmessa dal padre al figlio affetto, coinvolge i 2 geni contigui, *IMMP2L* e *DOCK4* e ed è solo parzialmente sovrapposta a CNVs già descritte in DGV (60% nella regione corrispondente al gene *IMMP2L*).

Inoltre *IMMP2L* e *DOCK4* sono anche i due geni che hanno dato i migliori risultati negli studi di associazione.

Chr.	Locus	Start (Mb)	End (Mb)	Length (kb)	log BF	Copy number		Gene(s)	DGV (overlap)	
2	1	154,645	154,798	152,4	29,5	3	Mother	<i>GALNT13</i>	Yes (14%)	
	2	168,721	169,384	663,5	183,9	3	Control	<i>STK39, LASS6</i>	Yes (24%)	
	3	179,106	179,381	275,1	13,8	3	Mother	<i>TTN</i>	Yes (72%)	
		179,250	179,381	130,8	11,3	3	Mother	<i>TTN</i>	Yes (100%)	
	4	190,853	190,944	91,0	20,6	3	Father	<i>HIBCH, INPP1</i>	No	
5	191,224	191,538	313,9	10,4	1	Father	<i>NAB1, GLS</i>	Yes (2%)		
7		100,749	100,917	167,7	48,0	3	Proband	<i>EMID2 and RABL5</i>	Yes (100%)	
		100,749	100,896	147,1	13,8	3	Mother			
	1	100,749	100,917	167,7	45,3	3	Proband	<i>EMID2 and RABL5</i>		
		100,788	100,917	128,9	38,1	3	Mother			
		100,749	100,915	166,4	22,8	3	Proband	<i>EMID2 and RABL5</i>		
		100,749	100,917	167,7	30,9	3	Mother			
		100,788	100,917	128,9	50,4	3	Father	<i>EMID2</i>		
		100,788	100,915	127,6	12,7	3	Control	<i>EMID2</i>		
	2	110,198	110,271	73,0	40,0	1	Father	<i>IMMP2L (intronic)</i>		No
	3	110,639	111,454	815,6	119,3	3	Father	<i>IMMP2L and DOCK4</i>		Yes (60%)
110,708		111,450	742,6	56,0	3	Proband				
4	115,640	115,791	150,9	10,5	1	Control	<i>TES</i>	Yes (23%)		
5	124,276	125,280	1003,9	10,3	1	Control	<i>POT1</i>	Yes (47%)		
	124,521	124,856	334,6	12,8	1	Father	No RefSeq genes	Yes (85%)		
6	132,728	132,843	114,5	32,4	1	Father	<i>EXOC4</i>	No		
	132,728	132,829	100,7	11,5	4	Control	<i>EXOC4</i>			
7	133,784	133,793	8,7	14,7	1	Mother	<i>AKR1B1</i>	No		

Tabella6.3: CNVs identificate conQuantisNP. In giallo è sottolineata la duplicazione nei geni *IMMP2L-DOCK4*

6.2: Analisi di CNV nei geni *IMMP2L* e *DOCK4* tramite QMPSF in individui con autismo

È in questo contesto che si inserisce la prima parte del mio progetto di dottorato che è consistita nell'analisi di varianti strutturali nei geni *IMMP2L* e *DOCK4*. Al fine di testare se varianti strutturali nei geni *IMMP2L* e *DOCK4* segregano con il fenotipo autistico e mostrano una frequenza più alta nei casi rispetto ai controlli, abbiamo sviluppato un saggio di PCR multiplex con frammenti fluorescenti (QMPSF, *Quantitative Multiplex PCR of Short Fluorescent fragment*).

Il saggio di QMPSF consiste nell'amplificazione contemporanea di sonde fluorescenti specifiche per regioni target e per regioni di riferimento, in cui ogni sonda ha una lunghezza differente.

L'analisi viene condotta facendo un confronto dei rapporti tra l'area del picco corrispondente alla sonda specifica e l'area del picco corrispondente alla sonda di riferimento tra paziente e controllo. I dati vengono analizzati con il software GeneMapper (figura 6.3), che fornisce l'area in corrispondenza dei picchi delle varie sonde, e poi viene applicata la seguente formula:

$$\text{Numero di copie} = \frac{(A_{\text{probe_pz}}/A_{\text{probe_average}})}{(A_{\text{RNF_pz}}/A_{\text{RNF_average}})} \begin{cases} > 1.4 = \text{duplicazione} \\ < 0.6 = \text{delezione} \end{cases}$$

dove:

- A_{pz} è l'area del picco di ogni sonda locus specifica.
- A_{average} è l'area media di tutti i picchi di tutti i campioni corsi assieme per la medesima sonda.

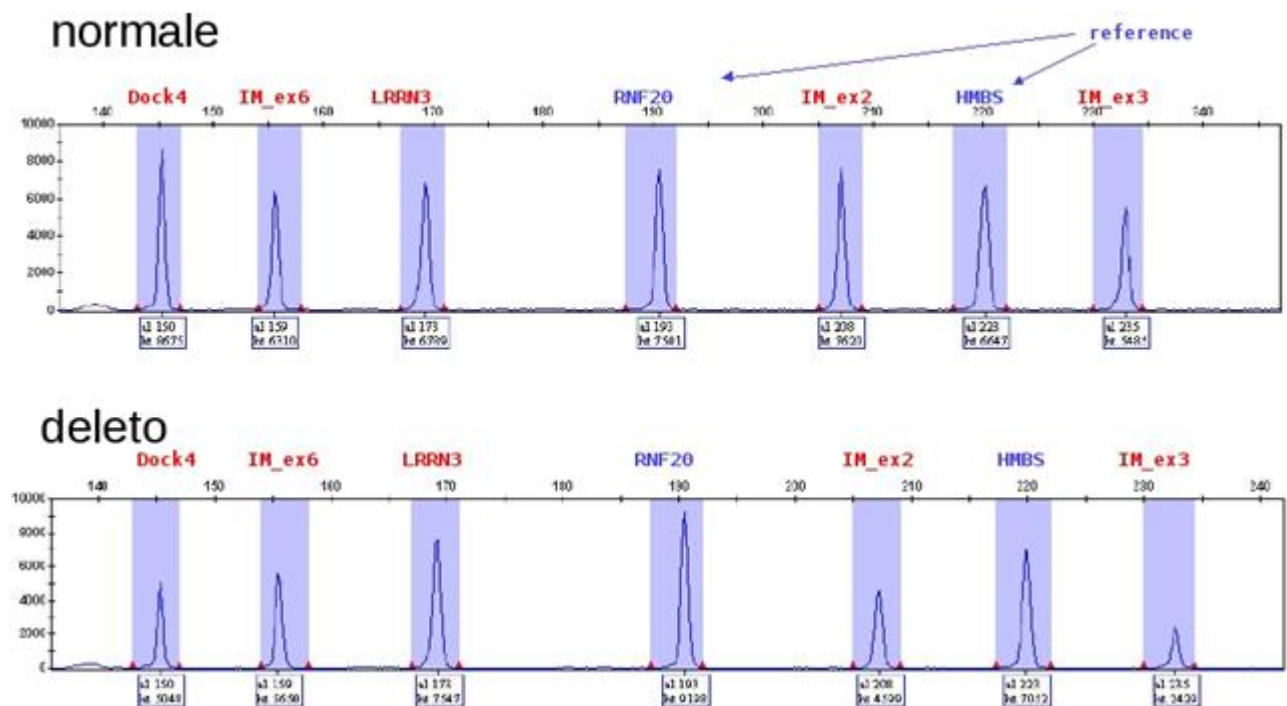


Figura 6.3: Analisi della QMPSF con Gene Mapper.

Per valori maggiori di 1.3 abbiamo un'indicazione di duplicazione nel locus dove mappa la sonda, per valori minori di 0.7 abbiamo indicazione di una delezione.

Il primo saggio di QMPSF che abbiamo effettuato comprende 6 sonde di cui 5 sono localizzate negli esoni 2, 3 e 6 di *IMMP2L*, l'esone 4 di *LRRN3* e l'ultimo esone di *DOCK4*, mentre la sesta sonda è quella di controllo che stata disegnata sul gene *RNF20* (fig 6.7).

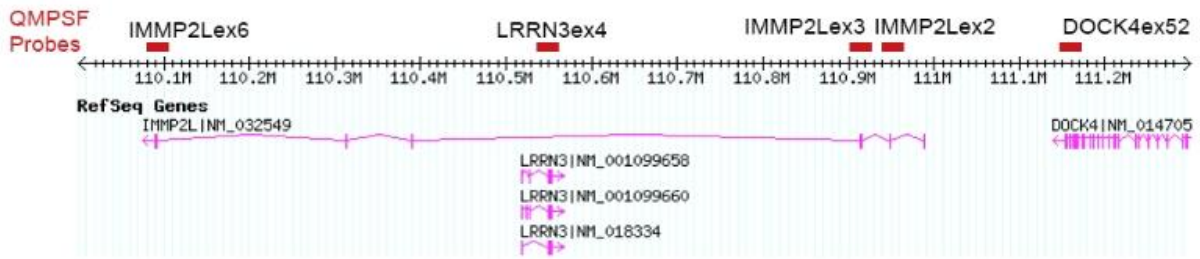


Figura 6.7: Localizzazione delle sonde utilizzate per il saggio di QMPSF

In questo modo abbiamo confermato la duplicazione nella famiglia 13-3023, identificata mediante QuantiSNP, ed abbiamo verificato che essa è trasmessa dal padre al figlio affetto, ma non è trasmessa agli altri due figli, di cui uno è anch'esso affetto da ASD. Tramite la QMPSF abbiamo anche analizzato 475 controlli UK, 285 famiglie multiplex dell'*IMGSAC* (con 487 individui affetti) e 180 famiglie trios italiane. Abbiamo così identificato altre sei delezioni di diversa lunghezza, che sono riportate in figura 6.8.

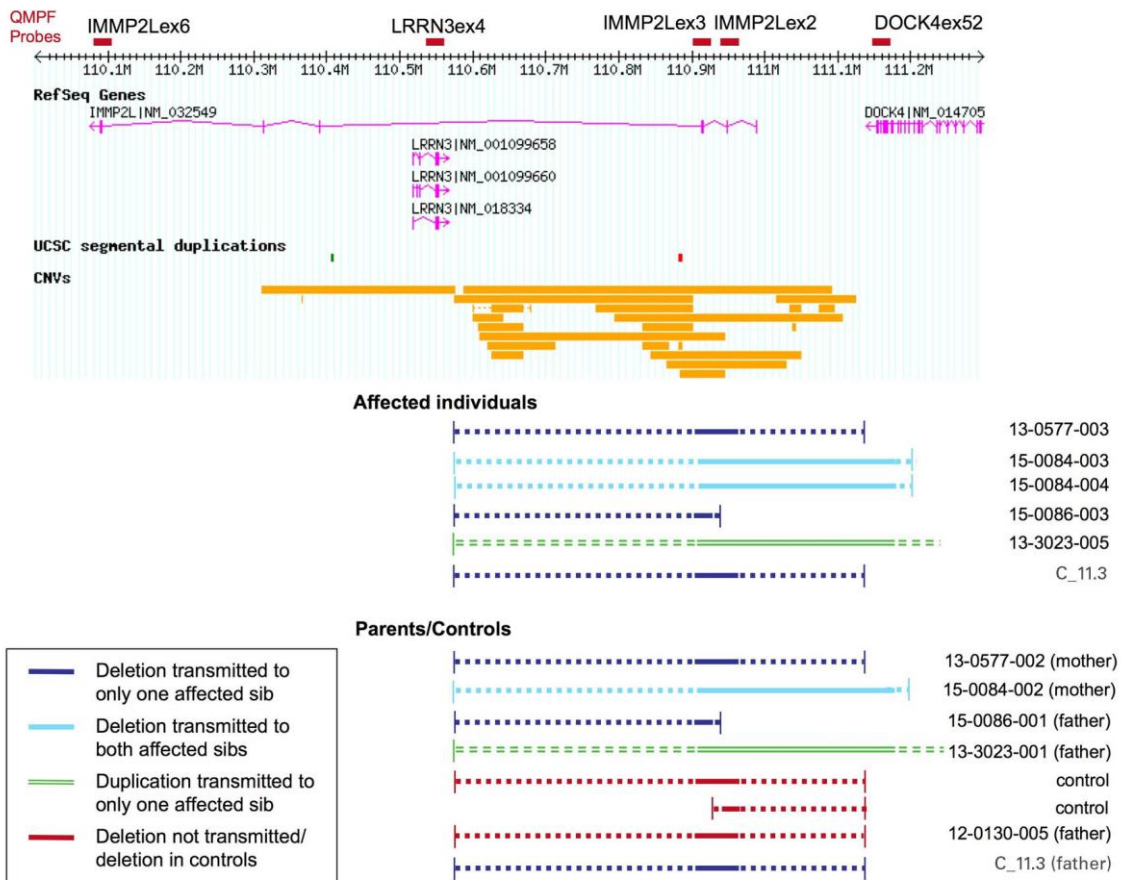


Figura 6.8: Copy Number Variants identificate nella regione IMMP2L-DOCK4. I CNVs presenti nel Database delle Genomic Variants (DGV) sono mostrati come barre arancioni. I trattini e le linee continue indicano la

lunghezza massima e minima delle delezioni o delle duplicazioni identificate negli individui affetti e nei controlli.

In particolare è stata trovata una delezione in una famiglia Olandese, la famiglia 15_0084, che distrugge il secondo e il terzo esone di *IMMP2L* e l'ultimo esone, l'esone 52, di *DOCK4*, ed è trasmessa dalla madre, ad entrambi i figli affetti, e alla figlia non affetta.

Tramite real time PCR, con sonde disegnate negli esoni 52, 37, 31, 14 e 7 di *DOCK4*, abbiamo visto che il punto di rottura distale della delezione è localizzato tra gli esoni 31 e 14 del gene *DOCK4*.

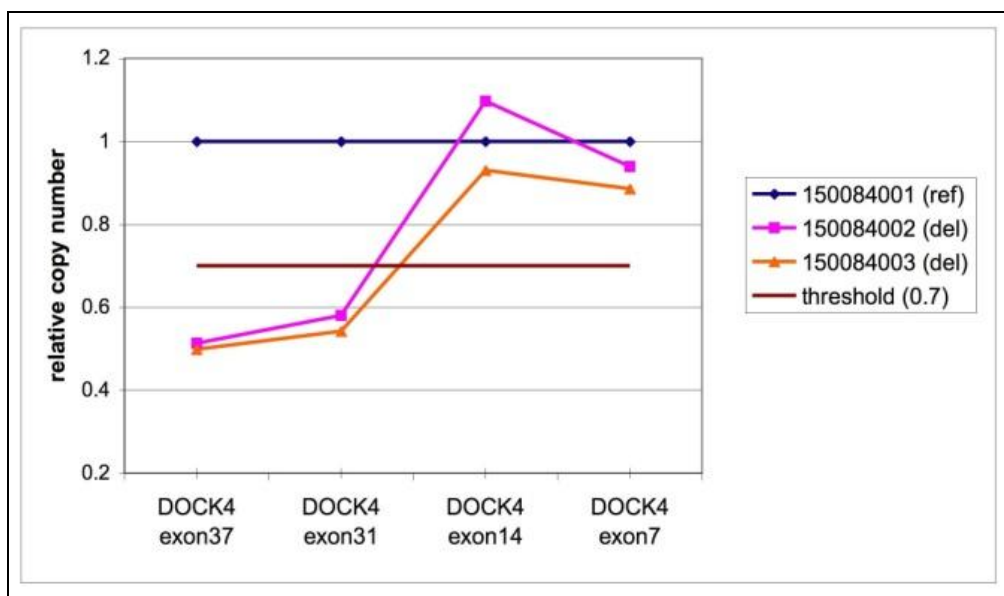


Figura 6.9: Numero di copie dei frammenti del gene *DOCK4* analizzati per Real Time PCR nella famiglia 150084. I frammenti amplificati con i primers specifici per l'esone 37 e l'esone 31 di *DOCK4* mostrano un relative copy number intorno allo 0.5 nella madre e nel probando, indice della presenza della delezione su uno dei due alleli del gene

Poiché, a differenza del gene *IMMP2L*, in DGV non sono descritte delezioni che coinvolgono gli esoni del gene *DOCK4*, abbiamo deciso di effettuare un saggio di QMPSF specifico per il gene *DOCK4* per accertare la frequenza di varianti strutturali in questo gene in altre 197 famiglie multiplex con autismo appartenenti al consorzio IMGSAC (320 individui affetti) e in un campione di controllo di 461 soggetti caucasici inglesi. Le sonde per la QMPSF sono state disegnate in modo da coprire l'intera regione codificante sugli esoni 1, 8, 15, 25, 31 e 52 di *DOCK4* (Fig. 6.10). Come controllo interno è stata utilizzata una sonda interna all'esone 7 del gene *RNF20* (gene sul cromosoma 9), utilizzata anche nel saggio precedente.

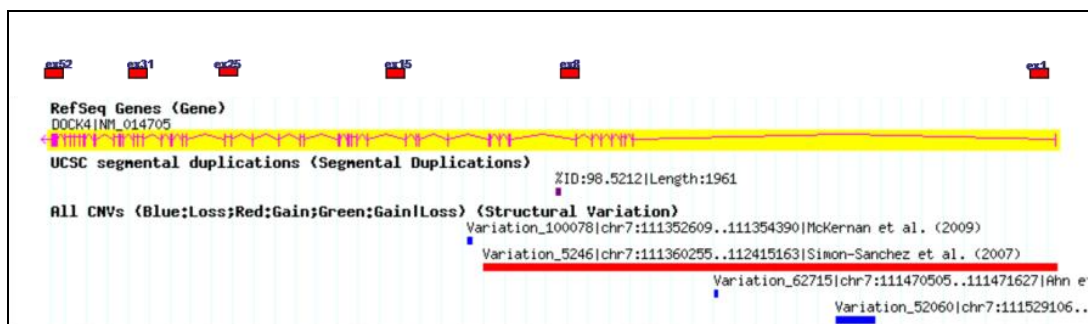


Figura 6.10: coppie di primers per la QMPSF disegnate sugli esoni 1, 8, 15, 25, 31 e 52 del gene *DOCK4*

Da questa analisi è emerso che la delezione della famiglia 15_0084 non coinvolge l'esone 25 di *DOCK4*. Analizzando i risultati non sono state identificate altre varianti strutturali in *DOCK4*, né nelle famiglie autistiche né nei controlli.

6.3: Caratterizzazione della delezione *IMMP2L/DOCK4* nella famiglia 15-0084

6.3a. Definizione del break point della delezione *IMMP2L/DOCK4*.

Per definire meglio le estremità di questa delezione, i campioni di DNA di tutti i membri della famiglia 15-0084 sono stati analizzati con l'Illumina 1M SNP Array, che comprende più di un milione di SNP e offre un'alta copertura globale del genoma.

L'analisi dei dati con il software QuantiSNP ha ulteriormente confermato la presenza della delezione, stabilendo che il punto di rottura prossimale della delezione risiedeva tra i marcatori rs37713 e rs37715, mentre il punto di rottura distale era situato tra i marcatori rs6966622 e rs0238664 (Fig 6.11). Questa risoluzione ci ha permesso di disegnare dei primers per una PCR long-range, in modo da amplificare un frammento di circa 3000 bp che contenesse i punti di rottura della delezione (Fig 6.12).

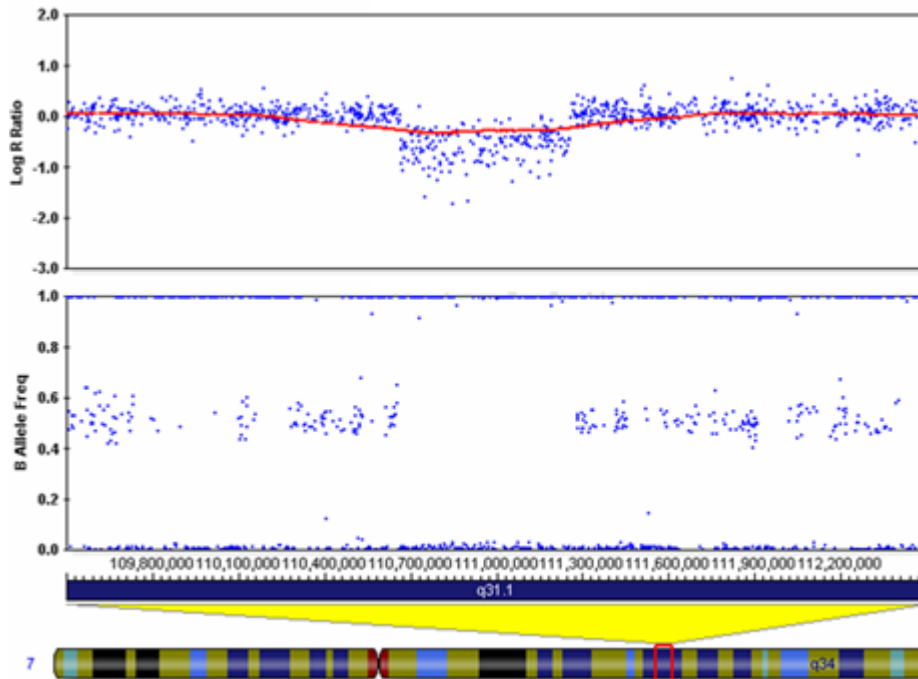


Figura 6.11: Plots del Log Ratio e della B Allele Frequency BeadStudio dei dati ottenuti dal 1M SNP array per il probando 15-0084-003.

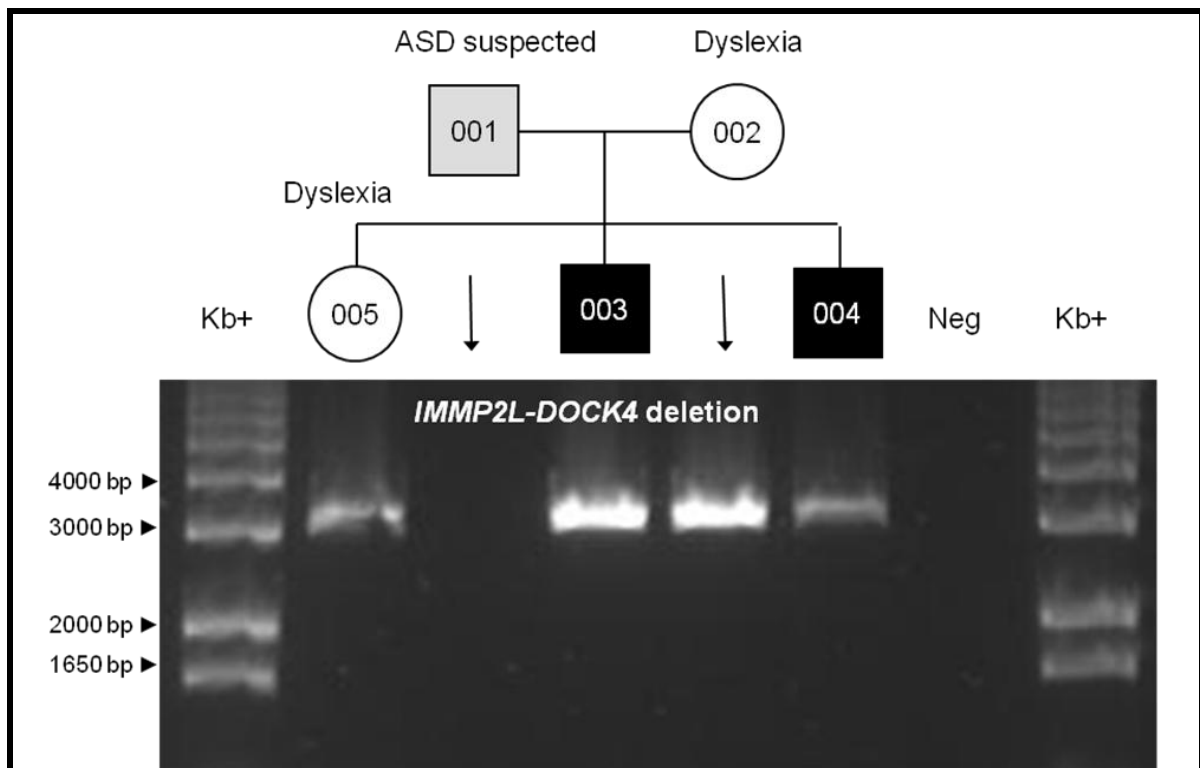
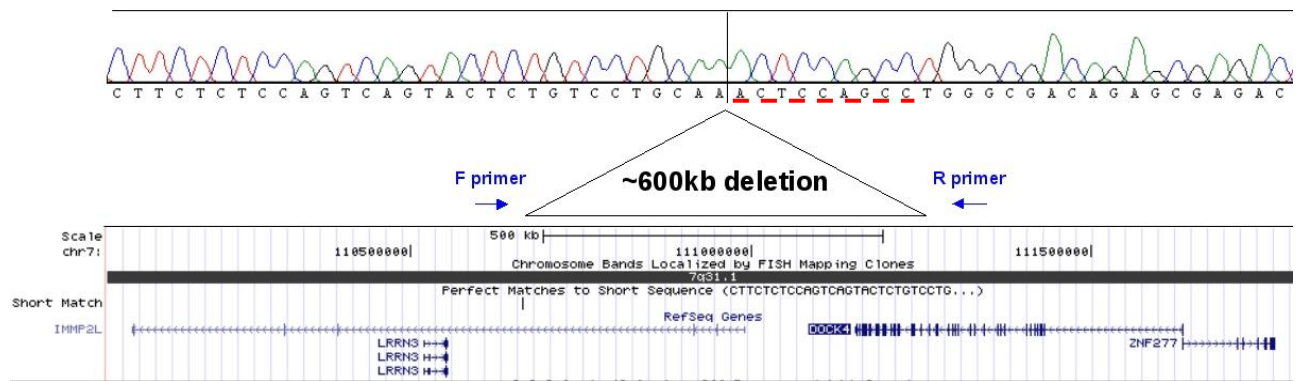


Figura 6.12: PCR long-range: Segregazione della delezione nei geni *IMMP2L-DOCK4* nella famiglia 15_0084. La banda (che indica la delezione) è presente nella madre e in tutti e tre i figli.

Il successivo sequenziamento di questo frammento ha evidenziato la presenza di un motivo ACTCYAGCC comune ad entrambi i punti di rottura e ha consentito di determinare che l'ampiezza della delezione è di circa 600 kb (594 kb), e comporta la perdita dall'esone 4 del gene *IMMP2L* fino all'esone 27 compreso del gene *DOCK4*.



ottenuta dalla long-range PCR. Con una linea tratteggiata rossa è mostrato il motivo ACTCYAGCC.

6.3b: Trascritto di fusione

Recenti studi hanno dimostrato che le delezioni che si estendono da un gene a quello vicino possono potenzialmente portare alla formazione di trascritti di fusione deleteri. Poiché *IMMP2L* e *DOCK4* sono trascritti nella stessa direzione, abbiamo ipotizzato che si potesse formare un trascritto di fusione *DOCK4-IMMP2L*.

Per confermare la presenza del trascritto fuso *DOCK4-IMMP2L* abbiamo quindi estratto l'RNA della madre 150084_002 che aveva la delezione *DOCK4-IMMP2*, abbiamo retrotrascritto l'RNA estratto, e abbiamo effettuato una reazione di RT-PCR, con un primer sull'esone 4 di *IMMP2L* e l'altro sull'esone 26 di *DOCK4*, ottenendo un frammento di circa 130 bp che è stato successivamente sequenziato (fig 6.14).

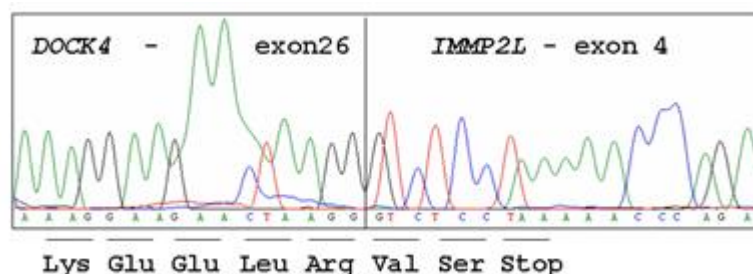


Figura 6.14: Sequenza dei prodotti di RT-PCR che provano l'esistenza di un trascritto fuso tra l'esone 26 del gene *DOCK4* e l'esone 4 di *IMMP2L* nella madre 150084_002.

Per ottenere una stima semiquantitativa del livello di espressione del trascritto di fusione, abbiamo ripetuto la reazione di RT-PCR, utilizzando un numero di cicli crescenti (25-30-35-40 cicli) e altre due coppie di primers, una disegnata sulla giunzione degli esoni 17 e18 e l'altra sulla giunzione tra gli esoni 47 e 48 di DOCK4. Dal confronto dell'intensità dei prodotti di amplificazione su gel di agarosio, si evince che il trascritto fuso è espresso ad un livello molto più basso rispetto al trascritto wild-type (figxx).

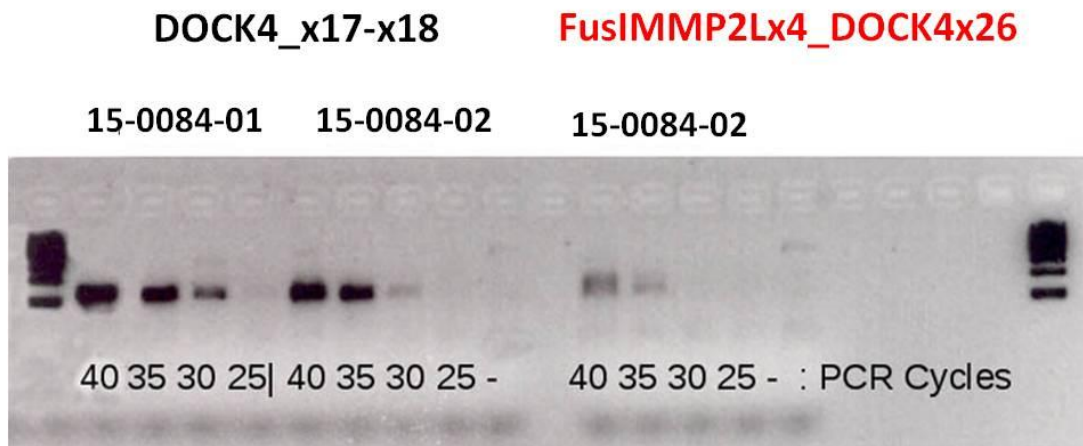


Figura 6.15: Stima semiquantitativa del livello di espressione del trascritto fuso IMMP2L-DOCK4. A 30 cicli di PCR si vede bene che nella madre l'espressione del trascritto di DOCK4 wild-type è inferiore a quella del padre con un normale numero di copie, mentre il trascritto di fusione IMMP2L e DOCK4 è veramente poco espresso, si inizia a vedere solo tra i 35 e i 45 cicli di PCR.

La sequenza proteica del prodotto di fusione corrisponde a quella di DOCK4 per i primi 945 amminoacidi, poi, in corrispondenza dell'esone 4 di *IMMP2L*, vengono inseriti gli amminoacidi Valina e Serina seguiti da un prematuro codone di stop. (fig 6.16)

```

MWIP EHEKYGVVIASFGVPYGLSLEIGDVQILEKCDGWYGFALKNPNIKGFIPSSYVHLKNA CVKKNKGQ
FEMVIPEDSVIEMSLDWGMWKQLYVNEGDLFHLWHIMNEILDQLVVGHLHDMKDVKHIALDWGNE
QLGLDLVPKEYAMVDPEDISIELYLMEHHKKDPVQASSHLFVQMKSLMCSNLGEELEVIFSLFDSKEN
PISEFFLLNGLPAPDKPEHCSLFDLGSSELKDIYIVHIIIIGMGAGEKKNACSVQYPFGCAVLSIADLLGE
KDDLILKVYMCNESEWYQIHENIHKLNAYNLGSNAGLAVSLQLLHGDIEQIEYSSVFSHGVSIKLGFNSI
IMP GEMNDLYIIEGEFEKGGKSVANVEVMFIVDSSGQLKDFISFGSGEPPASEYHSFVLYHNNSPWSELL
KLPIPVDKFGAHIFEFHCSKEKGEKKLFGFSFVPLMQEDGLPDGHELIVHKCEENNLQDYLLKLPFSKGIFL
GNNNQAMKAKESFCISFLCSKLQNGDMLDLLKWHDPKIGCLSKLKEIDGSEIVKFLQDLDFGILDENS
QKYGSKVFDLSLVHIINLLQDSKFHHFKPVMYDIESHFAGALAYDLIKVLKWYVDIEAEQEHIQEVLKAQ
EYIFKYIVQSLFSLAGGQNEEEFCCIQELLMSVFFLSQESKSGSALSQSQAVFLSSFPVAVYSELLKLFDVE
VANLVQDLGSLPILHVDDSLQAIKLQCIGKVESQLYNPDSYILLPVVLHHLHIHLQEQLDIMCAILSNVF
CLIKKNSSEKSVLEEIDVIVASLLDILLILEISPQPSSAMFQFQDVGEFVACLLSLLQMDHYQ QLLDSFNK
EELVSStop

```

Figura 6.16: Sequenza amminoacidica del prodotto di fusione *IMMP2L-DOCK4*. In nero è mostrata la sequenza di *DOCK4* con i suoi domini (SH3 evidenziato di verde e DHR-1 evidenziato di azzurro) mentre in rosso è mostrata la sequenza di *IMMP2L*. Il giallo è evidenziato il punto di rottura.

La presenza di uno stop prematuro fa ipotizzare che questo trascritto di fusione venga degradato mediante il meccanismo “nonsense mediated decay”. I risultati di RT-PCR semiquantitativa sono a supporto di questa ipotesi in quanto il trascritto di fusione è presente a livelli molto più bassi rispetto al trascritto wilde type (fig xxx).

L’espressione del trascritto normale di *DOCK4* è stata valutata nella madre con la delezione, mediante RT-PCR quantitativa, utilizzando le stesse due coppie di primers utilizzate in precedenza (una disegnata sulla giunzione degli esoni 17 e18 e l’altra sulla giunzione tra gli esoni 47 e 48 di *DOCK4*), ed è stata trovata dal 30% al 50% più bassa in confronto a quella del padre con un normale numero di copie (fig.6.17)

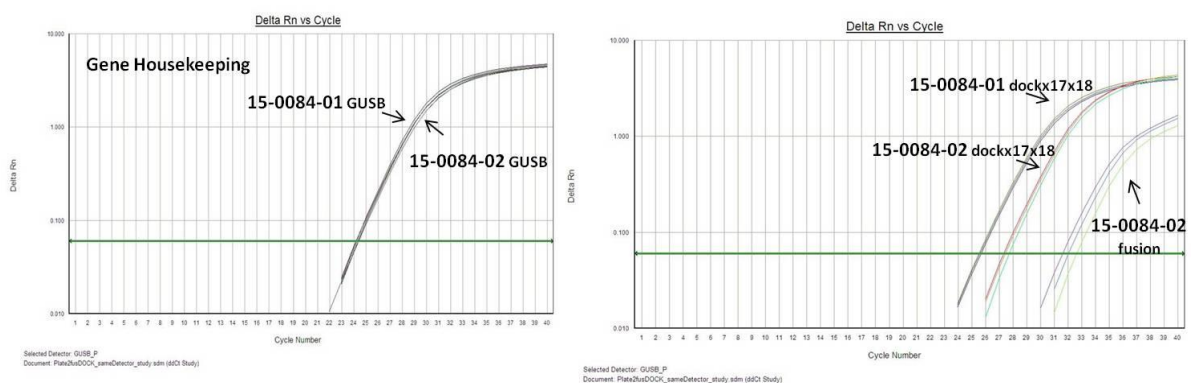


Figura 6.17: grafico dei risultati di real time PCR: a sinistra gene *GUSB*, il padre e la madre mostrano la stessa quantità di espressione, a destra si vede che il gene *DOCK4* nella madre (15-0084-02) è espresso in quantità inferiore rispetto a quello del padre e si vede anche che i livelli di espressione del trascritto di fusione sono molto più bassi di quelli di *DOCK4*.

6.3c. Longe range PCR su famiglia estesa della madre

I clinici che hanno in carico la famiglia 15-0084 hanno effettuato una più completa analisi fenotipica dei membri della famiglia 15-0084, da cui è emerso che sia la madre (10-0084-002), che la figlia (15-0084-005), entrambe portatrici della delezione in *IMMP2L-DOCK4*, e non affette da ASD, manifestano dislessia, ed hanno inoltre diagnosticato altri casi di dislessia in alcuni parenti da parte di madre. Abbiamo quindi deciso di testare la presenza della delezione in tutti gli individui appartenenti alla famiglia che fossero disposti a partecipare a questo studio.

Tramite long-range PCR, abbiamo verificato la presenza della delezione di circa 594 kb di *DOCK4-IMMP2L* in 5 individui, tre dei quali hanno una diagnosi formale di dislessia (fig.6.18).

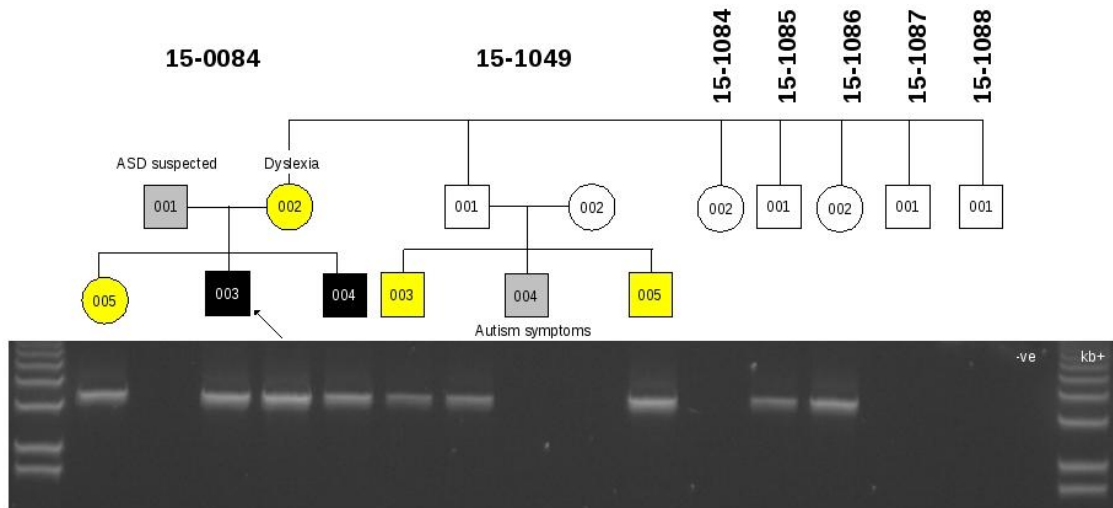


Figura 6.18: Famiglia 15-0084 estesa. La delezione *IMMP2L-DOCK4* è presente in 5 individui, di questi il 15-1049-001, 15-1049-003 e 15-1049-005 hanno una diagnosi formale di dislessia.

La presenza della delezione nella madre, nel fratello, nella sorella, nella zia e in due cugini del probando, ciascuno dei quali con difficoltà di lettura, ha quindi suggerito che questa delezione possa agire anche come un fattore di suscettibilità per la dislessia.

6.4. Screening di *DOCK4* su individui dislessici

Per meglio comprendere il potenziale ruolo di *DOCK4* nella suscettibilità alla dislessia, abbiamo deciso di testare la presenza di varianti strutturali nel gene *DOCK4* in altri individui con diagnosi di dislessia.

Tramite una collaborazione con il consorzio NeuroDys, abbiamo analizzato 606 individui con dislessia, non imparentati tra loro, che erano già stati genotipizzati con l’Infinium SNP array (Illumina) e 2091 controlli, genotipizzati mediante SNP array 550K (Illumina). Dall’analisi dei CNVs nella regione genomica di *DOCK4*, abbiamo identificato una delezione del gene *DOCK4* che rimuove la sequenza del gene dall’esone 38 al 52 presente in un solo individuo dislessico, mentre nessuna delezione è stata identificata nei controlli. Mediante QMPSF e PCR quantitativa abbiamo confermato la delezione identificata che è risultata essere trasmessa dal padre, che aveva avuto una diagnosi di dislessia durante l’infanzia, al figlio affetto, ma non all’altro figlio non affetto (Fig 6.19)

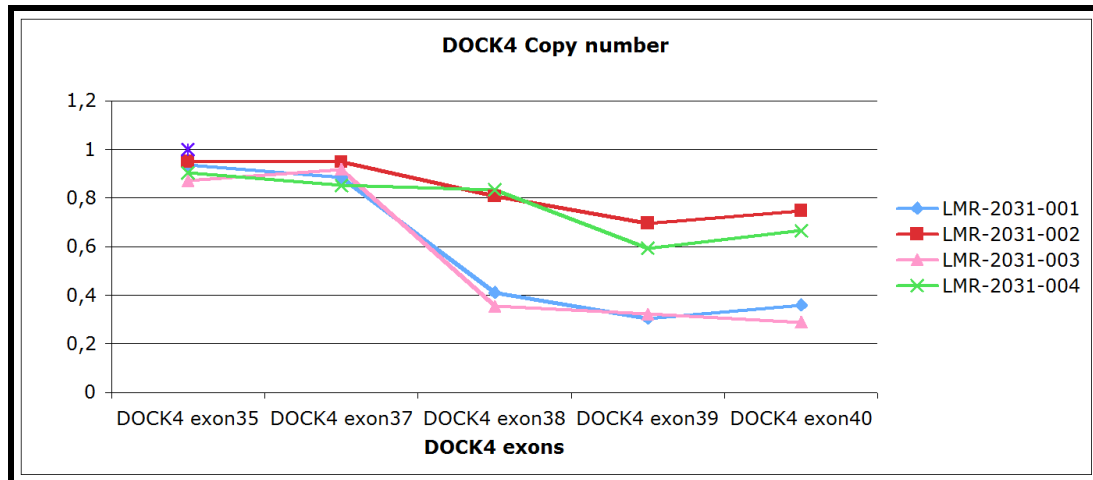


Figura 6.19: Numero di copie dei frammenti del gene *DOCK4* analizzati per Real Time PCR nella famiglia LMR-2031, appartenente al consorzio Neurodys. I frammenti amplificati con i primers specifici per l'esone 38, 39 e per l'esone 40 di *DOCK4* mostrano un relative copy number intorno allo 0.4 nel padre (LMR-2031-001) e nel probando (LMR-2031-003), indice della presenza della delezione su uno dei due alleli del gene.

L'identificazione di un'altra delezione del gene *DOCK4*, che segrega con il fenotipo dislessico, in un altro nucleo familiare, ha ulteriormente rafforzato la nostra ipotesi riguardante il coinvolgimento di delezioni di *DOCK4* nella suscettibilità alla dislessia.

Abbiamo quindi deciso di estendere lo screening di varianti strutturali nel gene *DOCK4*, tramite QMPSF, in un ulteriore campione di individui dislessici ed in altri individui di controllo.

A tal fine abbiamo analizzato:

- 794 casi tedeschi appartenenti al consorzio NEURODYS
- circa 350 casi di Oxford-Cardiff
- 15 campioni Francesi
- 733 controlli tedeschi appartenenti al consorzio NEURODYS

Da questo screening tramite QMPSF, inizialmente, avevamo identificato una delezione nell'esone 1 di *DOCK4* che sembrava essere trasmessa dalla madre al figlio affetto da dislessia in una famiglia tedesca (LW 3043), appartenente al consorzio NEURODYS. Abbiamo quindi messo appunto un esperimento di PCR quantitativa per confermare la nuova delezione identificata in *DOCK4*, utilizzando due sonde nell'esone 1 e una sonda in *FOXP2* (gene controllo). Purtroppo la Real Time non ha confermato la delezione (fig.xxx), per cui abbiamo ipotizzato che uno dei primers dell'esone 1 di *DOCK4*, utilizzati nell'esperimento di

QMPSF, potesse contenere una variante polimorfica. Per verificare quest'ultima ipotesi, abbiamo scritto un'altra coppia di primers, esterna a quella utilizzata nella QMPSF, in modo da amplificare e sequenziare il frammento ottenuto. Dall'analisi della sequenza così ottenuta, abbiamo effettivamente individuato la presenza di uno SNP, che non era descritto in dbSNP e che risulta essere molto raro nella popolazione, che cade proprio all'interno del primers forward utilizzato per amplificare l'esone 1 di *DOCK4* (Fig 6.20).

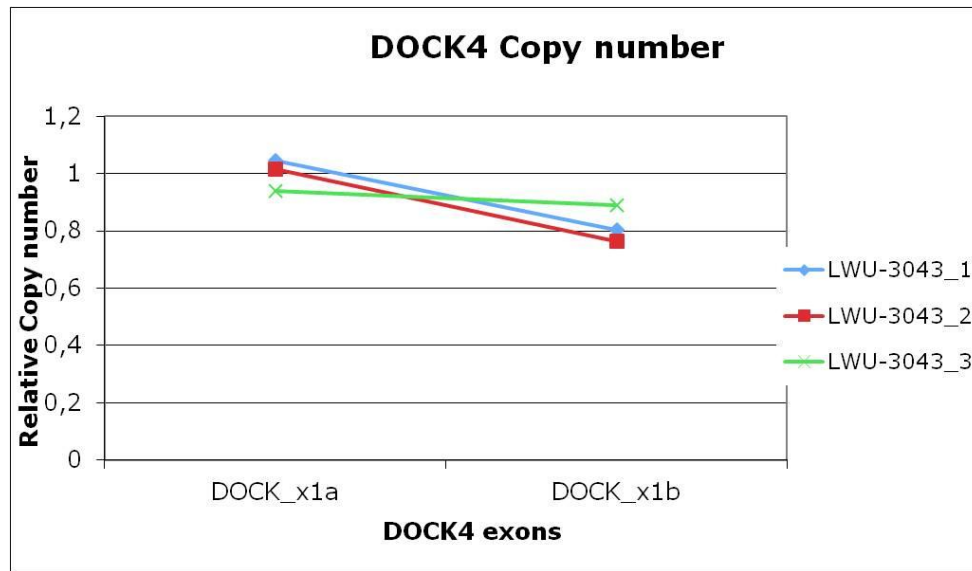


Figura 6.20: Numero di copie del frammento dell'esone 1 del gene *DOCK4* analizzato per Real Time PCR nella famiglia LWU-3043: tutti e tre i membri della famiglia hanno un copy number relative di 1 indicando che non c'è nessuna delezione.

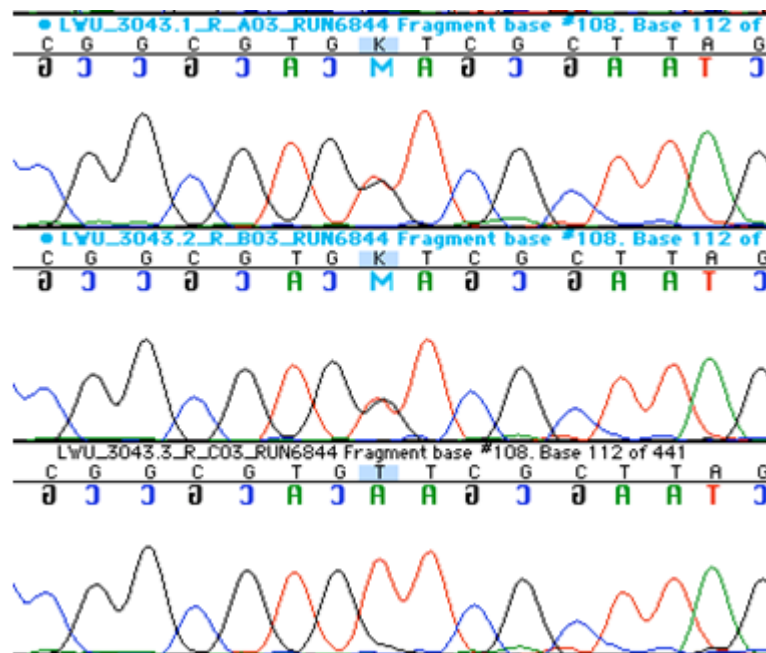


Figura 6.21: cromatogramma della sequenza dell'esone 1 di *DOCK4*, ottenuta mediante i nuovi primers esterni a quelli utilizzati nella analisi di QMPSF, nella famiglia LWU_3043. La madre (LWU_3043.2) trasmette la variante polimorfica(T → G) al figlio (LWU_3043.1).

Quindi, in definitiva, col saggio della QMPSF non abbiamo identificato nessuna altra CNV nel gene *DOCK4* nè tra i casi né tra i controlli.

Inoltre, grazie ad una collaborazione con il consorzio Genedys, abbiamo analizzato altri 175 casi di dislessia e 60 controlli, precedentemente genotipizzati con lo SNP array 660K (Illumina), senza identificare nessuna CNV in *DOCK4*.

Infine, nel contesto dello screening di tutto il genoma per l'analisi di CNV nell'autismo condotto dall'AGP (Pinto et al, 2010), sono stati analizzati 1287 controlli di origine europea SAGE con con l'array Illumina 1M, senza individuare nessuna ulteriore delezione in *DOCK4*. Lascia solo i SAGE, la fase 2 non e' neanche pubblicata.

N° CASI	CONSORZIO	METODICA ANALISI	CNV
606	Neurodys	Infinium SNP array Illumina	1 delezione in <i>DOCK4</i>
794	Neurodys	QMPSF <i>DOCK4</i>	
350	Oxford-Cardiff	QMPSF <i>DOCK4</i>	
15	Francesi	QMPSF <i>DOCK4</i>	
175	Genedys	660k SNP array	
1940			1 delezione in <i>DOCK4</i>
N° CNTRL	CONSORZIO	METODICA ANALISI	CNV
733	Neurodys	QMPSF <i>DOCK4</i>	
60	Genedys	660k SNP array	
2091	Neurodys	550 K SNP array Illumina	
475	ECACC	QMPSF <i>IMMP2L-DOCK4</i>	
1287	SAGE study	Illumina 1M SNP array	
4646			0 delezioni

Tabella 6.4: CNV identificate all'interno del gene *DOCK4* in 1940 individui affetti da dislessia e 4646 individui di controllo provenienti da vari consorzi: Neurodys, Genedys, ECACC, e SAGE (subjects from the Study on Addiction: Genetics and Environment (SAGE))

Quindi, se consideriamo insieme tutti i dati dei casi e dei controlli analizzati, abbiamo trovato 2 microdelezioni nel gene *DOCK4* su 1940 casi in confronto a 0 delezioni in 4646 controlli. Per determinare se questa differenza fosse o meno statisticamente significativa, abbiamo effettuato un test di Fisher, ottenendo un p-value di 0.08674.

6.5. Identificazione e caratterizzazione di una delezione in *CNTNAP5*

Dall'analisi di tutti i membri della famiglia 15-0084 con l'Illumina 1M SNP Array, sono emerse altre CNV, per la maggior parte già descritte in DGV, ma di cui una è risultata essere

particolarmente interessante. Infatti, è stata identificata un'altra delezione di circa 227 kb, in corrispondenza del gene *CNTNAP5*, sul cromosoma 2, che viene trasmessa dal padre ad entrambi i figli affetti da autismo.(tab.6.5, fig.6.22 e fig.6.23).

ID	Chr	Start (bp)	End (bp)	Length (bp)	Start	End	No. SNPs	Copy N	log BF	Gene(s)
1	2	124,836,663	125,063,827	227,164	rs11688892	rs4848944	84	1	117.45	<i>CNTNAP5</i>
2	7	110,654,771	111,256,808	602,037	rs214475	rs6966622	179	1	543.06	<i>DOCK4, IMMP2L</i>
3	7	110,666,487	111,256,808	590,321	rs37715	rs6966622	175	1	777.48	<i>DOCK4, IMMP2L</i>
3	2	124,836,663	125,063,827	227,164	rs11688892	rs4848944	84	1	313.38	<i>CNTNAP5</i>
4	7	110,254,457	111,266,360	1,011,903	rs10279573	rs7783121	310	1	169.92	<i>DOCK4, IMMP2L</i>
4	2	124,829,445	125,111,265	281,820	rs7578650	rs11900792	94	1	60.77	<i>CNTNAP5</i>
5	7	110,247,163	111,266,360	1,019,197	rs1569122	rs7783121	311	1	208.85	<i>DOCK4, IMMP2L</i>

Tabella 6.5: CNVs identificate nelle famiglia 15_0084 mediante QuantiSNP. Le posizioni fisiche mostrate sono basate sui dati dell'NCBI build 36

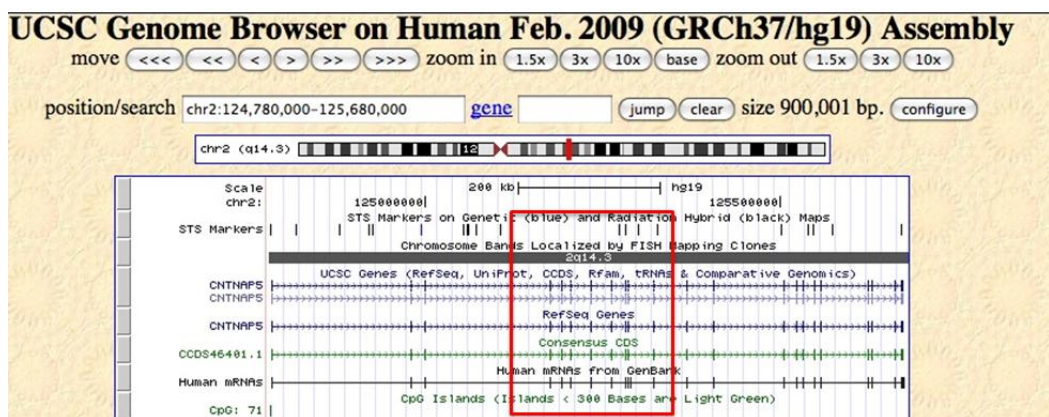


Figura 6.22: nel riquadro in rosso è indicata la regione deleta dallo SNP rs1168889 al SNP rs4848944 nel gene *CNTNAP5* identificata nella famiglia 14_0084 mediante analisi con l'Illumina 1M SNP Array

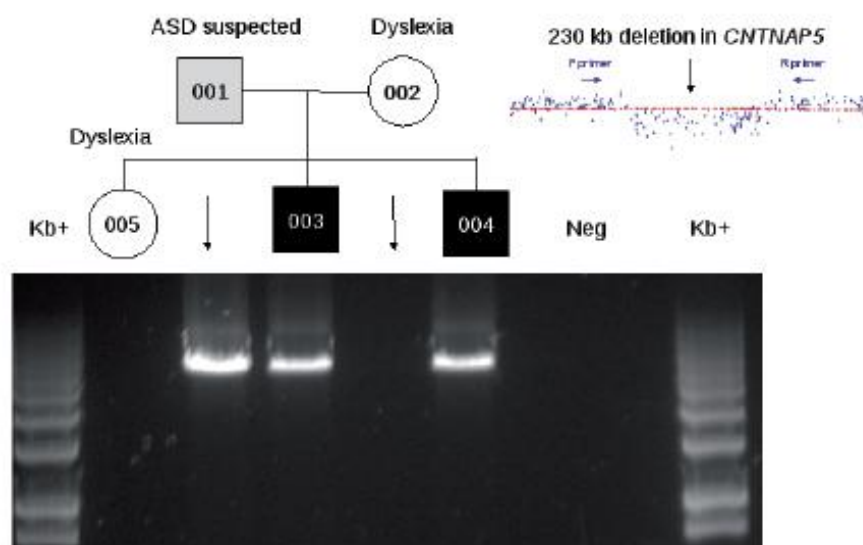


Figura 6.23: Segregazione della delezione nel gene *CNTNAP5* nella famiglia 15_0084

La presenza della delezione in *CNTNAP5* nella famiglia 15-0084, è stata confermata mediante Real Time PCR utilizzando delle sonde disegnate negli esoni 1, 8 e 10 di *CNTNAP5* e una in *FOXP2* come gene controllo. (figura 6.24)

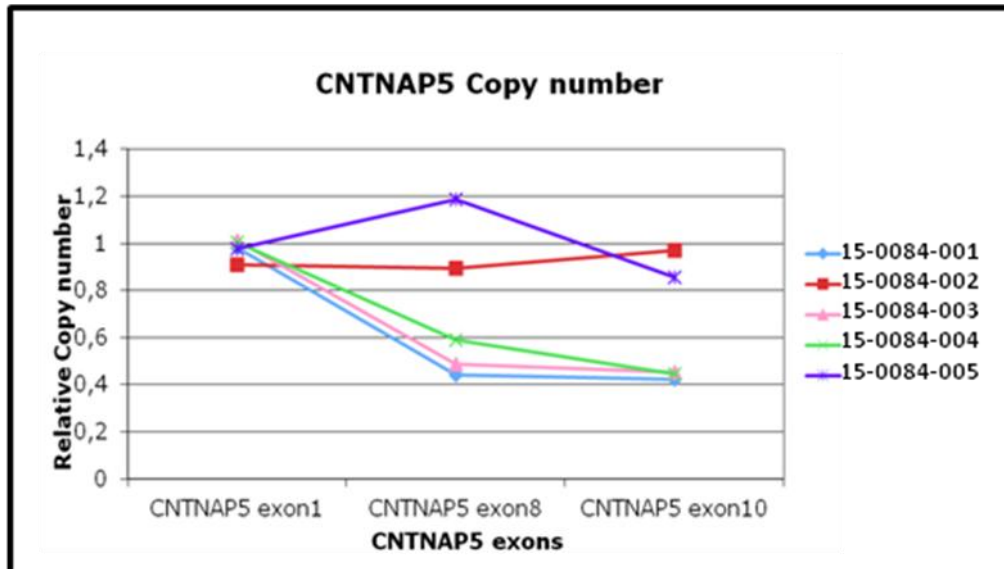


Figura 6.24: Numero di copie dei frammenti del gene *CNTNAP5* analizzati per Real Time PCR nella famiglia 15-0084. I frammenti amplificati con i primers specifici per l'esone 8 e per l'esone 10 di *CNTNAP5* mostrano un copy number intorno allo 0.4-0.6 nel padre (15-0084-001) e nei due probandi (15-0084-003 e 15-0084-004), indice della presenza della delezione su uno dei due alleli del gene.

Tramite long-range PCR, abbiamo determinato l'esatto punto di rottura, confermando che la delezione rimuove dal quarto all'undicesimo esone di *CNTNAP5* (fig 6.23). Tale delezione determinerebbe la formazione di una nuova proteina di 179 amminoacidi, di cui 125 corrisponderebbero al 5' della proteina *CNTNAP5* canonica, mentre 54 sarebbero nuovi residui seguiti da un codone di stop prematuro (fig xx). Poiché la delezione determina la perdita di più di 1000 amminoacidi (1178) dall'estremità 3' della proteina *CNTNAP5* canonica, è probabile che anche questo trascritto venga degradato, mediante il meccanismo di NMD (*nonsense mediated decay*). Tuttavia, poiché *CNTNAP5* non è espresso nel sangue, non è stato possibile confermare la presenza del nuovo trascritto tronco.

```

MDSLPRLTSLVLTLLFSGLWHLGLTATNYNCDDPLASLLSPMAFSSSSDLTGTHSPAQLNWRVGTGGWS
PADSQQWLQMDLGNRVEITAVATQGRYGSSDWVTSYSLMFSDTGRNWKQYKQEDSIWPSTSNPARCT
GTRGIQPASSTSTQMAAAHWDLSRCTAISLRTRSGHQCSTTIQSStop
  
```

Figura 6.25: Sequenza amminoacidica della delezione *CNTNAP5*. In nero è mostrata la sequenza di *CNTNAP5* fino all'esone 3 in giallo sono indicati i nuovi 54 amminoacidi che si formano in seguito alla delezione.

6.6. Screening di mutazione del gene *CNTNAP5*

Il gene *CNTNAP5* risulta essere un ottimo gene candidato per l'autismo per diverse ragioni. Innanzitutto, delezioni in *CNTNAP5* non sono descritte in DGV né sono state trovate nel campione di 2091 controlli analizzati mediante SNP Array 550K. Inoltre, questa delezione è trasmessa dal padre, che i clinici hanno riportato mostrare vari tratti autistici, anche se non ha una diagnosi formale di autismo, e nei due figli affetti da autismo, ma non nella sorella dislessica. Per di più, dall'anamnesi familiare è emerso che, anche altri familiari da parte del padre sembrano avere problemi neuropsichiatrici (figura xxx). Infine, il gene *CNTNAP5* codifica per una proteina transmembrana appartenente alla famiglia delle Neurexine, che media il contatto cellulare e le interazioni nel sistema nervoso, giocando un ruolo fondamentale nel suo sviluppo e nel suo funzionamento.

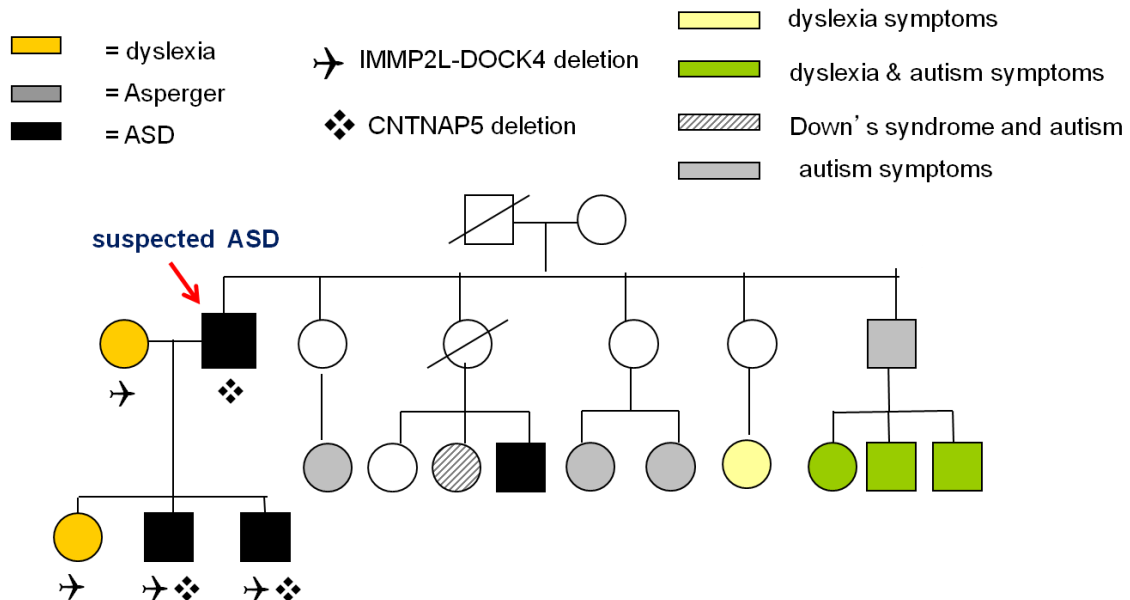


Figura 6.26: famiglia estesa del padre 15-0084-001. In grigio sono indicati gli individui della famiglia che mostrano sintomi autistici, in verde individui affetti da autismo e dislessia, in nero individui con ASD, in giallo quelli con la dislessia e nel simbolo a righe è indicata una sorella del padre con autismo e sindrome di Down.

A fine di verificare se il gene *CNTNAP5* potesse svolgere un ruolo nella suscettibilità all'ASD, abbiamo sequenziato tutti i 24 esoni di *CNTNAP5* in 143 individui affetti appartenenti alla collezione di famiglie multiplex dell'IMGSAC, che condividono 1 o 2 alleli identici per discendenza nella regione 2q14.3 dove è localizzato il gene *CNTNAP5*.

La sequenza ottenuta è stata analizzata comparandola con la sequenza di riferimento dell'mRNA di *CNTNAP5* (NM_130773).

Lo screening di mutazione ha portato all'identificazione di 3 mutazioni non sinonime in individui con ASD e non descritte in dbSNP (build 130).

Nell'esone 14, abbiamo evidenziato una transversione $C \Rightarrow G$ che provoca un cambio amminoacidico Prolina \Rightarrow Arginina, in posizione 714 (P714R) che ricade nel dominio fibrinogeno C-terminale. Il residuo di prolina è poco conservato durante l'evoluzione ma la sua perdita comporta alterazione nella struttura della proteina. Questa variante è trasmessa dalla madre ad entrambi i figli affetti ma non alla figlia sana (Fig. A)

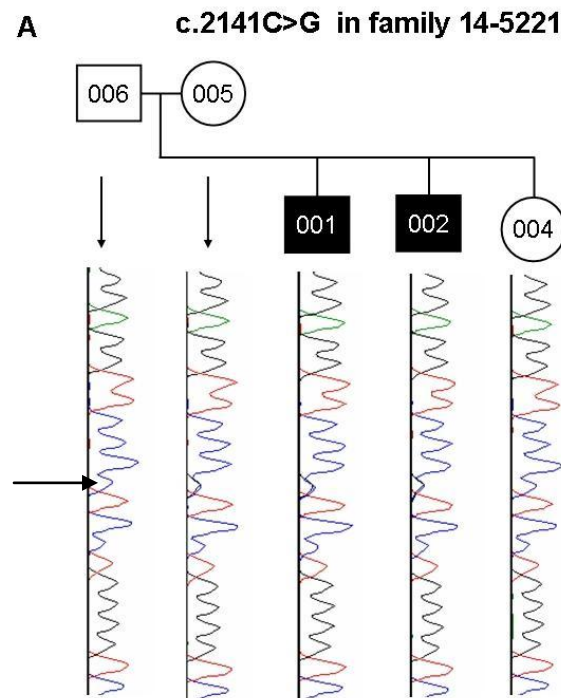


Figura A: Segregazione della mutazione nell'esone 14 di CNTNAP5 nella famiglia 14-5221. Lo SNP è indicato dalla freccia.

Per verificare se questa variante fosse presente anche in un campione di controllo, abbiamo analizzato 466 individui controllo, mediante PCR seguita da digestione con l'enzima di restrizione Mnl1. Mnl1 riconosce la sequenza CCTC e in assenza della variante G taglia il prodotto di PCR in un solo punto portando alla formazione di due frammenti (di 156 bp e di 206bp), invece la presenza della variante G porta alla formazione di un ulteriore sito di riconoscimento per l'enzima Mnl1, per cui, dopo la digestione si avranno 3 frammenti (151 bp, 205 bp e 211bp). La variante è stata trovata in un solo individuo.(Fig 6.27)

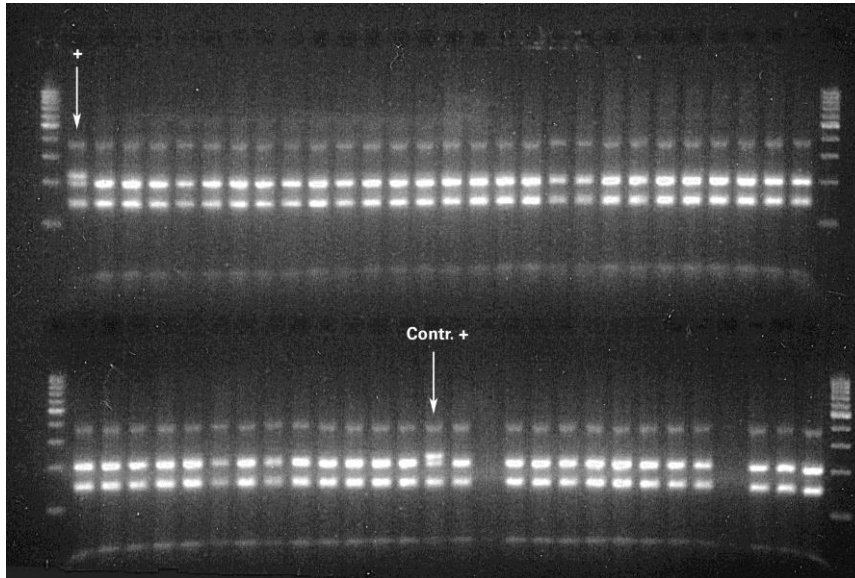


Figura 6.27: Gel di agarosio della digestione dei controlli con l'enzima *MnlI*. E' stato evidenziato l'individuo con lo SNPs (+) e il controllo risultato positivo (Contr. +) la banda in alto indica il frammento di PCR non digerito è di 374bp.

Nell'esone 18, abbiamo trovato una transizione $C \Rightarrow T$ che provoca un cambio amminoacidico Treonina \Rightarrow Metionina (T919M). Questa mutazione è trasmessa dalla madre ad entrambi i figli affetti ma non alle 2 figlie sane. (fig B)

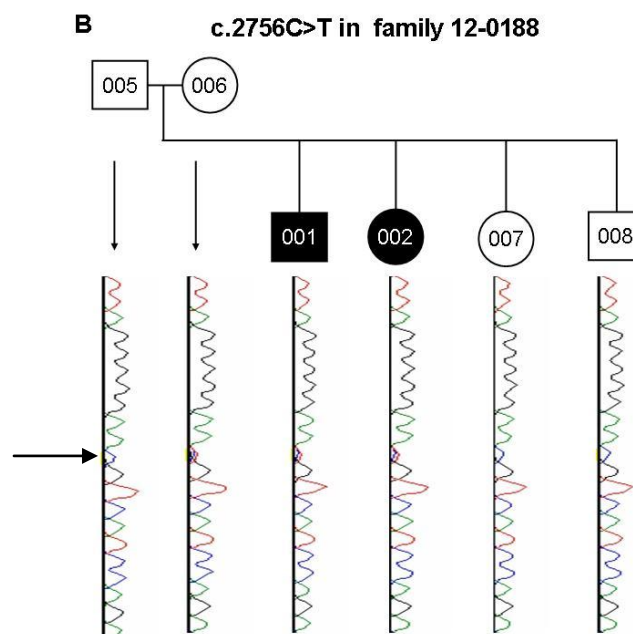


Figura B: Segregazione della mutazione nell'esone 18 di *CNTNAP5* nella famiglia 12-0188. Lo SNP è indicato dalla freccia.

Secondo le predizioni bioinformatiche, la mancanza della Treonina provoca un danno alla struttura della proteina.

Anche in questo caso abbiamo analizzato 611 controlli per la presenza della variante allelica T mediante PCR seguita da digestione con l'enzima di restrizione HpyCH4IV senza mai trovarla.

Nell'esone 22, abbiamo trovato una transizione $G \Rightarrow A$ che provoca un cambio amminoacidico Valina \Rightarrow Isoleucina (V1168I), all'interno del dominio 4 simile alla laminina G. Questa mutazione è trasmessa dalla madre ad entrambi i figli affetti ma non al fratello sano. (Fig C)

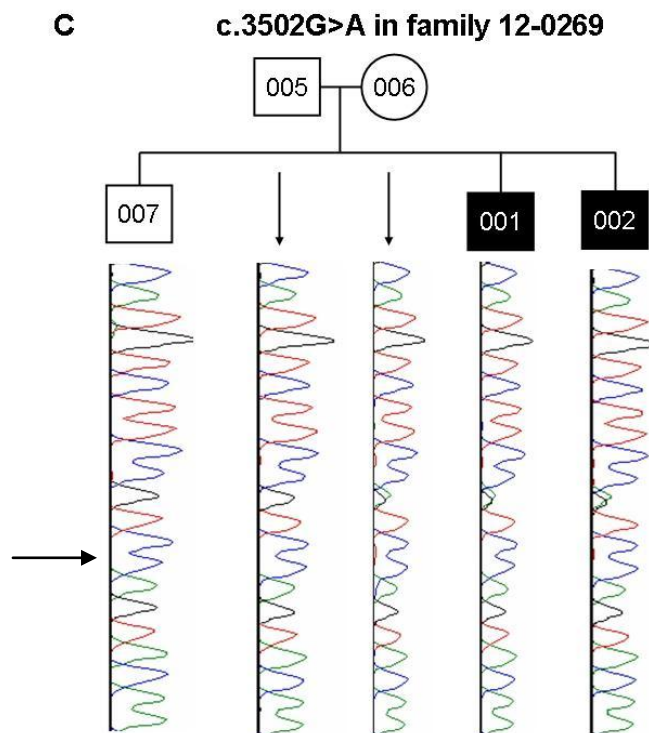


Figura C: Segregazione della mutazione nell'esone 22 di CNTNAP5 nella famiglia 12-0269. Lo SNP è indicato dalla freccia.

Sempre mediante digestione con l'enzima di restrizione HpyCH4IV sono stati analizzati 1232 controlli e l'alterazione non è mai stata trovata.

Sebbene entrambi gli amminoacidi siano alifatici, e secondo gli strumenti di predizione bioinformatici questa mutazione non sia dannosa, la Valina nella posizione 1168 è molto conservata tra le specie.

Le due varianti, che non sono state trovate negli individui di controllo (T919M e V1168I), sono state analizzate in altri 380 individui tedeschi con ASD. Mentre non è stato trovato nessun altro individuo con la variante T919M, la mutazione V1168I è stata trovata in un altro caso tedesco con ASD. Tuttavia, in questa seconda famiglia, la mutazione non è trasmessa anche all'altro fratello affetto. Dunque, in totale questa mutazione è stata trovata in 2/523

individui con ASD e in 0/616 individui di controllo. Tale differenza non risulta essere statisticamente significativa (test esatto di Fisher, p-value=0.211).

Tramite digestione con l'enzima di restrizione del prodotto di PCR dell'esone 22, abbiamo anche identificato un polimorfismo C/T che porta ad un cambiamento da Treonina a Metionina (T1195M). In totale, questo polimorfismo (rs34165507) è presente con una frequenza dell'allele minore del 3.1% nei casi con ASD (32 su 1046 cromosomi) in confronto a quella del 2.9% negli individui di controllo (36 su 1232 cromosomi). Tale differenza non risulta essere statisticamente significativa (test chi-quadrato, p-value=0.848).

Infine, lo screening di sequenza del gene *CNTNAP5* è stato effettuato anche in uno dei probandi della famiglia 15-0084, per determinare se oltre alla delezione di *CNTNAP5*, ereditata dal padre, fossero presenti altre varianti patologiche, ereditate dalla madre. Tuttavia, sono stati trovati solo un cambiamento C/T in una regione non conservata del 5'UTR ed un altro polimorfismo intronico comune (rs924802), nessuno dei quali può probabilmente colpire la funzione o l'espressione del gene.

7 RISULTATI: Analisi di CNV in ritardo mentale

7.1. CHERISH

Durante l'ultimo anno di dottorato, mi sono occupata dell'analisi di varianti strutturali di numero di copie (CNVs) in individui affetti da ritardo mentale (MR). A tale proposito, ho partecipato ad un progetto europeo, il Cherish, che unisce scienziati e clinici provenienti dall'Europa dell'est e dall'Asia centrale, con lo scopo di collezionare un grande campione di individui affetti da ritardo mentale (ad oggi circa 1200) e identificare le cause molecolari di questa patologia.

I paesi appartenenti al consorzio CHERISH sono: ITALIA, ARMENIA, POLONIA, CIPRO, RUSSIA, UCRAINA, REPUBBLICA CECA, LITUANIA, ESTONIA.

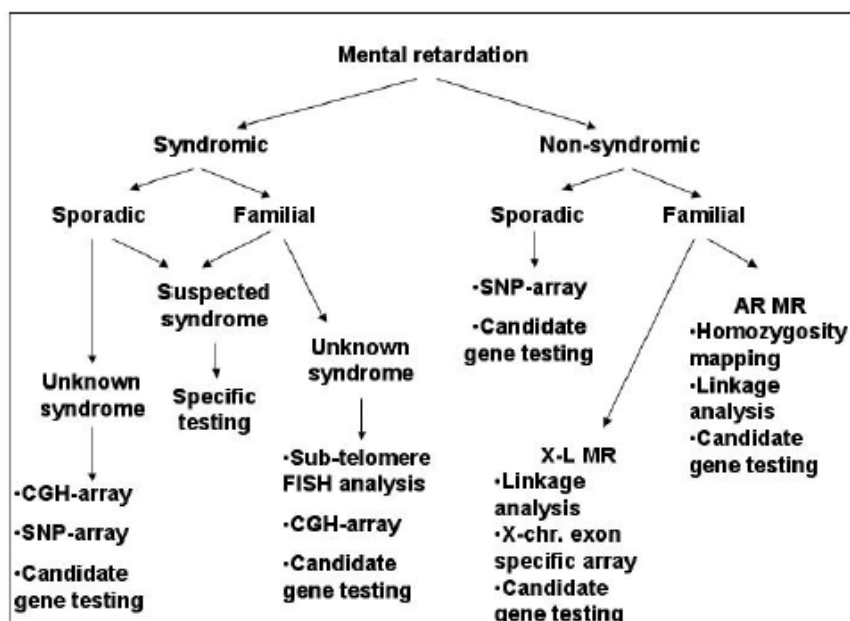
No.	Participant organisation name	Participant Name	Short name	Cities	Country
1	University of Bologna	Prof. Giovanni Romeo	UNIBO	Bologna	Italy
2	University of Tartu	Prof. Ants Kurg	UTARTU	Tartu	Estonia
3	Vilnius University Hospital Santariskiu Klinikos	Prof. Vaidutis Kucinskas	DHMG-VU	Vilnius	Lithuania
4	Charles University Prague	Prof. Zdenek Sedlacek	CUNI	Prague	Czech Republic
5	Poznan University of Medical Sciences	Prof. Anna Latos-Bielenska	PUMS	Poznan	Poland
6	National Academy of Sciences of Ukraine	Prof. Ludmila A. Livshits	NASU	Kiev	Ukraine
7	The Cyprus Institute of Neurology and Genetics	Dr. Philippos C. Patsalis	CING	Nicosia	Cyprus
8	Russian Academy of Medical Sciences, Tomsk Scientific Center	Prof. Igor Lebedev	IMG	Tomsk	Russia
9	Center of Medical Genetics and Primary Health Care	Prof. Tamara Sarkisjan	CMG	Yerevan	Armenia
10	Molecular Stamping s.r.l	Dr. Benjamin H. Larman	MS	Lodi	Italy

Gli obiettivi del progetto del Cherish sono :

- sviluppare un approccio standardizzato per la diagnosi del MR
- creare un grande database di pazienti con MR, sia sindromico che non sindromico, ben definito clinicamente.
- identificare i riarrangiamenti criptici del genoma attraverso l'analisi citogenetica molecolare

- sequenziare i geni del MR e analizzare l'epidemiologia molecolare del MR nelle popolazioni dell'Europa dell'Est
- sviluppare strumenti diagnostici per mutazioni ricorrenti o comuni
- identificare nuovi geni di MR nelle forme X-linked ed autosomiche recessive della malattia
- aumentare la conoscenza sulla possibile origine genetica del MR e le implicazioni per nuove strategie terapeutiche

In pratica, tutti i pazienti che risultano normali in seguito all'analisi classica del cariotipo e all'analisi FISH (nel caso in cui si ipotizzi che il ritardo mentale sia associato ad una sindrome sospetta), sono analizzati mediante MLPA (Multiplex Ligation-dependent Probe Amplification), MAPH (Multiplex Amplifiable Probe Hybridization) e FISH sub-telomerica, in modo da evidenziare l'eventuale presenza di riarrangiamenti nei telomeri. Quindi, i pazienti che non mostrano nessuna alterazione dei telomeri, sono analizzati mediante array-CGH (high-resolution whole-genome comparative genomic hybridization) e gli array di genotipizzazione di SNP (high-density whole-genome oligonucleotide single nucleotide polymorphisms (SNP)), in modo da identificare le alterazioni genomiche di più piccola dimensione, a seconda della densità degli SNP. Infine, tutte le varianti identificate sono confrontate a quelle depositate in DGV (database of genomic variant) in modo da eliminare tutte le varianti polimorfiche presenti nella popolazione generale. Le varianti nuove, sono inserite in un nuovo database creato dal Cherish per una consultazione rapida dei dati tra i vari partner del consorzio.



7.2. Array CGH

Nell'ambito del progetto CHERISH ho analizzato 218 individui affetti da MR tramite array CGH (60 Ucraini, 46 Russi, 6 Polacchi, 7 Lituani e 99 Italiani), usando la piattaforma 44K dell'Agilent, che in media consente di rilevare CNVs più grandi di 75-100 kb.

Questa analisi ha portato all'identificazione di 135 alterazioni strutturali in 127 pazienti, di cui:

- 14 sono sicuramente patogenetiche, in quanto associate a sindromi note o perché in letteratura sono state riportate in altri pazienti affetti da ritardo mentale (tabella 7.1)
- 25 sono probabilmente patogenetiche, perché non sono descritte in DGV e nella maggior parte dei casi coinvolgono un elevato numero di geni (tabella 7.2), ma non sono riportate associate al ritardo mentale né in DECIPHER né in letteratura, oppure perché sono state identificate in nuclei familiari con più di un individuo affetto e segregano con il ritardo mentale (tutti gli individui affetti hanno la stessa CNV)
- 23 sono di significato incerto: la maggior parte sono state ereditate da un genitore sano, alcune sono state riportate sia in DGV che in DECIPHER, altre, di cui alcune molto grandi, non sono state mai precedentemente descritte (tabella 7.3)
- 26 sono probabilmente benigne: per la quasi totale sovrapposizione a CNVs già descritte in DGV, oppure perché non contengono geni o contengono geni che non svolgono una funzione tale da poter essere associati al ritardo mentale (tabella 7.4)
- 46 alterazioni sicuramente benigne in quanto già ampiamente descritte in DGV

Abbiamo quindi determinato mediante array CGH l'origine parentale di 34 alterazioni strutturali, di cui 12 CNVs sono state selezionate tra quelle associate a sindromi note e 20 selezionate tra quelle mai descritte. Da quest'analisi, 14 CNVs sono risultate essere *de novo*, 11 di origine materna e 9 di origine paterna.

CNV patogenetiche							
BANDA CROMOSOMICA	POSIZIONE GENOMICA (Mb)	CNV	ORIGINE	SINDROME ASSOCIATA IN DECIPHER	DGV	GENI	DIMENSIONE (bp)
1q44	242,748-243,979	delezione	de novo	no	no	9	1,231,001
22q11.21	17,276-19,712	uplicazione		parte della sindrome da microuplicazione 22q11	no	84	2,436,001
22q11.21	19,084-19,712	uplicazione	paterna	parte della sindrome da microuplicazione 22q11	CNV	23	628,001
15q11.2-q13.1	21,250-26,199	uplicazione		ne sono descritte diverse su DECIPHER	no	112	4,991,001
Xp11.23-p11.22	47,937-52,858	uplicazione	familiare	no	no	141	4,921,001
16q22.3-q24.1	71,334-84,917	uplicazione	de novo	no	no	114	13,583,001
17pter-p13.3	0,05-2,361	delezione	de novo	regione della Miller-Dieker syndrome	no	54	2,311,001
8p23.1	7,290-12,285	delezione	de novo	più grande della sindrome da delezione 8p23.1	no	131	4,995,001
6q26-qter	162,314-170,734	uplicazione	de novo	no	no	64	8,420,001
8pter-p23.1	0,181-6,901	delezione	de novo	no	no	40	3,279,999
8p23.1-p21.3	12,627-20,512	uplicazione	de novo	no	no	46	7,885,001
2q35-qter**	218,312-242,690	uplicazione		no	no	305	24,378,001
10q26.3qter**	132,997-135,254	delezione		probabile derivativo	no	42	2,257,000
22q11.21	17,277-19,713	delezione		sindrome DiGeorge	no	84	2,436,001

Tabella 7.1: CNV patogenetiche. La posizioni genomiche mostrate sono basate sui dati dell'NCBI build36. Le CNVs con l'asterisco sono presenti nello stesso individuo .

CNV probabilmente patogenetiche							
BANDA CROMOSOMICA	POSIZIONE GENOMICA (Mb)	CNV	ORIGINE	SINDROMI ASSOCIATE IN DECIPHER	DGV	GENI	DIMENSIONE (bp)
16p11.2	29,581-30,105	delezione	de novo	locus di suscettibilità all'autismo	no	29	524
7q31.32	121,788-122,030	delezione	*materna	non descritta in DECIPHER	no	CADPS2	242,001
11p12-p11.2	42,873-43,733	uplicazione	paterna	non descritta in DECIPHER	no	10	860,001
15q13.2-q13.3	28,725-30,298	delezione	materna	parte della sindrome da microdelezione di 15q13.3	no	17	1,573,001
6q14.3	85,740-87,097	delezione		non descritta in DECIPHER	no	14	1,357,001
14q22.3-q23.1	54,493-58,868	delezione	de novo	non descritta in DECIPHER	no	49	1,573,000
9q31.3	111,673-111,860	delezione		non descritta in DECIPHER	no	4	187,001
16q23.3	82,387-82,593	uplicazione		non descritta in DECIPHER	no	5	206,001
8q22.3	104,453-104,841	uplicazione		non descritta in DECIPHER	no	5	388,001
11p14.2	27,033-27,181	uplicazione		descritta 1 dup in DECIPHER	no	BBOX1	148,001
6q26qter	161,079-170,734	delezione	de novo	2 delezioni simili riportate in DECIPHER	no	69	9,655,000
9q31.1q31.3	107,009-113,380	delezione	de novo	descritto un altro caso simile	no	45	6,371,000
1q21.1	145,031-146,201	uplicazione	materna	nota; possibile locus di suscettibilità per RM	no	29	1,170,001
3p25.3	9,391-9,474	delezione	de novo	non descritta in DECIPHER	no	3	83,001
21q22.3-qter	46,380-46,880	delezione	de novo	un pz in DECIPHER con altre alterazioni	no	12	500,001
15q11.2-q13.1	21,208-26,199	uplicazione	de novo	ne sono descritte diverse su DECIPHER	no	112	4,991,001
16p13-11p12.3	15,399-18,020	delezione	paterna	regione associata ad epilessia	no	19	2,621,001
16p11.2	29,581-30,106	delezione	paterna	locus di suscettibilità all'autismo	no	29	525,001
5q15-q22.1	92,766-110,875	delezione		non descritta in DECIPHER	no	93	18,109,001
2q32.3-q33.1	196,509-201,522	delezione		parte della sindrome da microdelezione 2q33.1	no	39	5,013,001
1p36.32-pter	0,749-3,552	delezione		parte della sindrome da microdelezione 1p36	no	95	2,803,001
2q37.1-qter	232,902-242,690	delezione		probabile derivativo/monosomia 2q37	no	152	9,788,001
3q27.3-qter	187,765-199,288	uplicazione		più grande della sindrome da microuplicazione 3q29	no	66	11,523,001
11p13	35,137-36,292	delezione		non descritta in DECIPHER	no	12	1,155,001
15q24.1q25.2	71,791-79,857	delezione		no	no	141	8,066,001

Tabella 7.2: CNV probabilmente patogenetiche. La posizioni genomiche sono basate sui dati dell'NCBI build 36. L'asterisco indica che l'origine parentale non è stata direttamente accertata ma dedotta poiché la presenza della CNV nei due figli affetti da MR e l'assenza della stessa CNV nel padre è indice che la CNV sia ereditata dalla madre

CNV di significato incerto							
BANDA CROMOSOMICA	POSIZIONE GENOMICA (Mb)	CNV	ORIGINE	DGV	DECIPHER	GENI	DIMENSIONE (bp)
Xp22.31	6,561-7,992	uplicazione		no	parziali	5	1,431,001
3p24.1	27,301-27,517	uplicazione	materna	no	no	2	217,001
9p24.2	3,817-4,577	uplicazione	paterna	no	1 dup	3	760,001
16q22.1	67,335-67,433	delezione	materna	no	no	CDH1	98,001
Xq13.3	74,410-74,566	uplicazione	materna	1	1 pz	4	156,001
1q21.1	144,343-144,458	uplicazione		si	si	3	115,001
13q12.12	22,464-23,788	uplicazione	paterna	1	1 pz	11	1,324,001
Xq28	153,059-153,158	delezione		1	1pz	6	99,001
3q23	143,649-143,760	uplicazione	materna	no	no	3	111,001
12q12	38,705-38,770	uplicazione	materna	no	no	SLC2A13	65,001
15q13.3	29,809-30,298	uplicazione	materna	si	diverse	3	489,001
Xp22.11	21,960-22,428	uplicazione		no	no	3	468,001
8q24.3	145,976-146,037	uplicazione	materna	no	no	6	61,001
11q22.3	103,555-104,643	uplicazione		1	no	11	1,088,001
Xp22.31	6,561-7,992	uplicazione	paterna*	no	no	5	1,431,001
Xp22.12-p22.11	21,797-21,911	uplicazione		no	no	2	114,001
probabile Xq28	152,975-153,062	delezione		no	no	2	87,001
16p12.2	21,382-21,859	delezione		si	1 pz	8	477,001
19q12	34,283-34,790	delezione	materna	no	no	4	507,001
Xp21.1	32,948-33,139	uplicazione		parziali	no	2	191,001
11q22.3	107,970-108,140	delezione		parziali	no	1	170,001
5q33.1	148,620-148,735	uplicazione		no	no	5	115,001
15q25.2	81,325-82,563	delezione	paterna	no	no	13	1,238,001

Tabella 7.3: CNV di significato incerto. La posizioni genomiche sono basate sui dati dell'NCBI build 36. L'asterisco sta ad indicare che anche il padre con quella duplicazione è affetto da lieve ritardo mentale.

CNV probabilmente benigne							
BANDA CROMOSOMICA	POSIZIONE GENOMICA (Mb)	CNV	ORIGINE	DECIPHER	DGV	GENI	DIMENSIONE (Bp)
9p24.3	0,601-0,883	duplicazione	*materna	no	parziale	4	115,001
17q23.2	56,464-56,682	delezione	paterna	no	no	2	282,001
Xq27.2	140,766-141,028	duplicazione	paterna	no	parziale	3	218,001
12q24.12	110,668-110,793	duplicazione	*paterna		no	3	125,001
14q32.33	104,918-105,354	delezione		no	parziali, LCR	21	436,001
19p13.3	0,798-1,209	delezione		no	parziali	21	411,001
15q11.2	20,336-20,636	duplicazione		no	si	5	300,001
3q25.32	159,795-159,806	duplicazione	materna	no	no	MLF1	11,001
2q12.3	106,738-107,638	duplicazione	paterna	no	no	3	900,001
12q24.12	110,668-110,793	duplicazione	paterna	no	no	3	125,001
18q22.1	62,405-64,017	delezione	*paterna	no	no	4	1,612,001
21q21.1	22,517-22,717	delezione	materna	no	no	0	200,001
15q11.2	20,335-20,636	delezione	paterna	no	si	0	301,001
Xp11.3	46,065-46,214	duplicazione	materna	no	no	2	149,001
6q27	169,938-169,970	duplicazione	paterna	no	no	2	32,001
20q11.21	29,352-29,692	duplicazione	paterna	no	no	14	340,001
13q12.11	19,104-19,329	duplicazione		no	2 duplicaz.	3	225,001
Xq22.2	103,107-103,155	duplicazione		no	2	4	48,001
1p36.22	10,204-10,363	duplicazione		no	si	KIF1B	159,001
Xq25	122,847-122,933	duplicazione		no	1	2	86,001
12p12.2p12.1	20,902-21,268	delezione		no	2	3	366,001
5q11.2	56,553-56,596	duplicazione	materna	no	no	GPBP1	43,001
22q12.3	33,086-33,729	duplicazione	paterna	no	no	1*	643,001
10q25.3	115,431-115,666	duplicazione	materna	no	no	4	235,001
2p22.3	32,696-33,102	duplicazione		no	si	3	406,001
3p26.3	0,653-0.902	duplicazione	paterna	no	no	0	249,001

Tabella 7.4: CNV probabilmente benigne. Le posizioni genomiche mostrate sono basate sui dati dell'NCBI build 36. L'asterisco indica che l'origine parentale non è stata direttamente accertata ma dedotta poiché la presenza della CNV in almeno due figli e l'assenza della stessa CNV in uno dei genitori è indice che la CNV sia stata ereditata dall'altro genitore non direttamente analizzato.

7.3. Screening di mutazioni nel gene *CADPS2* in individui affetti da ritardo mentale e/o autismo

Dai risultati dell'analisi di pazienti con ritardo mentale mediante array CGH è emersa una delezione presente in eterozigosi in due individui affetti da MR ed autismo, appartenenti ad una stessa famiglia (fratello e sorella) (fig 7.1). Questa delezione è lunga circa 250 kb ed è localizzata sul cromosoma 7q, in corrispondenza del locus AUTS1. Dato che questa delezione è presente in entrambi i figli affetti, mentre è assente nel padre, è probabile che essa sia stata ereditata dalla madre che aveva problemi cognitivi, ma di cui non abbiamo il DNA in quanto è deceduta in seguito ad un tumore (fig.7.2).

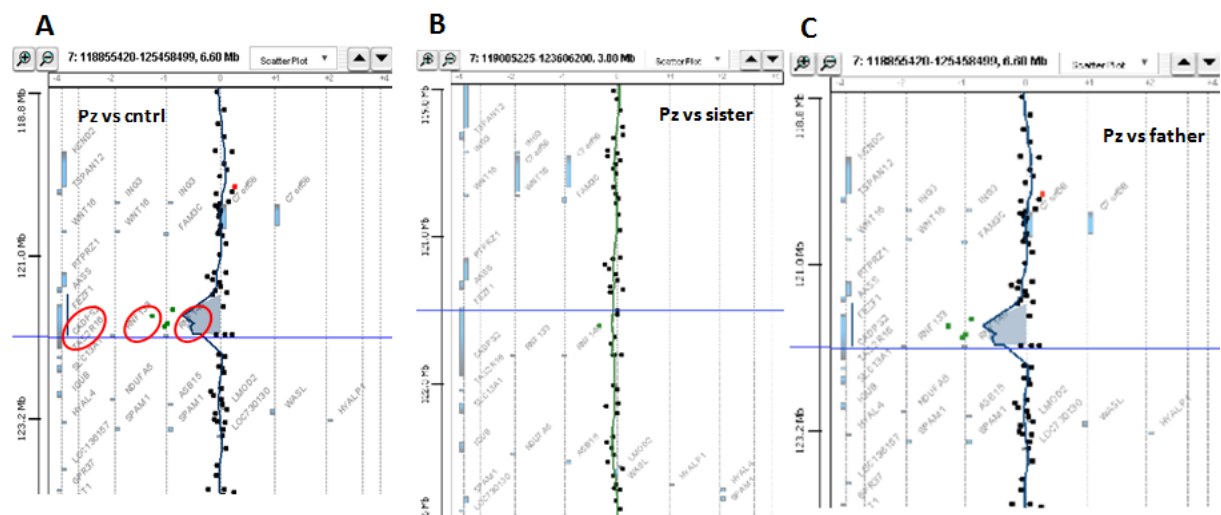


Figura 7.1: risultati array CGH. Tutti i pazienti sono stati analizzati mediante la metodica a loop (vedi materiali e metodi). A) identificazione della delezione che coinvolge il gene *CADPS2* in un paziente affetto da ritardo mentale e autismo B) Quando si analizza il paziente contro la sorella, entrambi affetti da ritardo mentale, la delezione non si osserva più, indicando che anche la sorella ha la stessa alterazione strutturale. C) quando il paziente è analizzato contro il padre si può notare la presenza della delezione nel paziente, indicando che il padre non è delemo in quella regione.

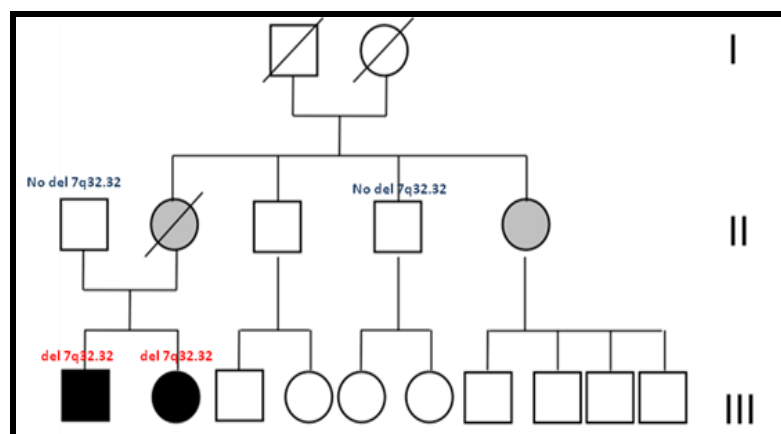


Figura 7.2: famiglia in cui è stata trovata la delezione del gene *CADPS2*

Questa delezione coinvolge il gene *CADPS2*, che rappresenta un buon candidato funzionale per l'autismo poiché è coinvolto nel rilascio di alcune neurotrofine (NT3 e BDNF) e codifica per una proteina che è localizzata nelle vescicole, a livello delle giunzioni presinaptiche.

Dai risultati di aCGH, la massima regione deleta va dall'esone 3 compreso fino alla fine del gene (esone 30), mentre la minima regione deleta, ossia quella sicuramente deleta, va dall'esone 7 all'esone 27 di *CADPS2* (fig.7.3)

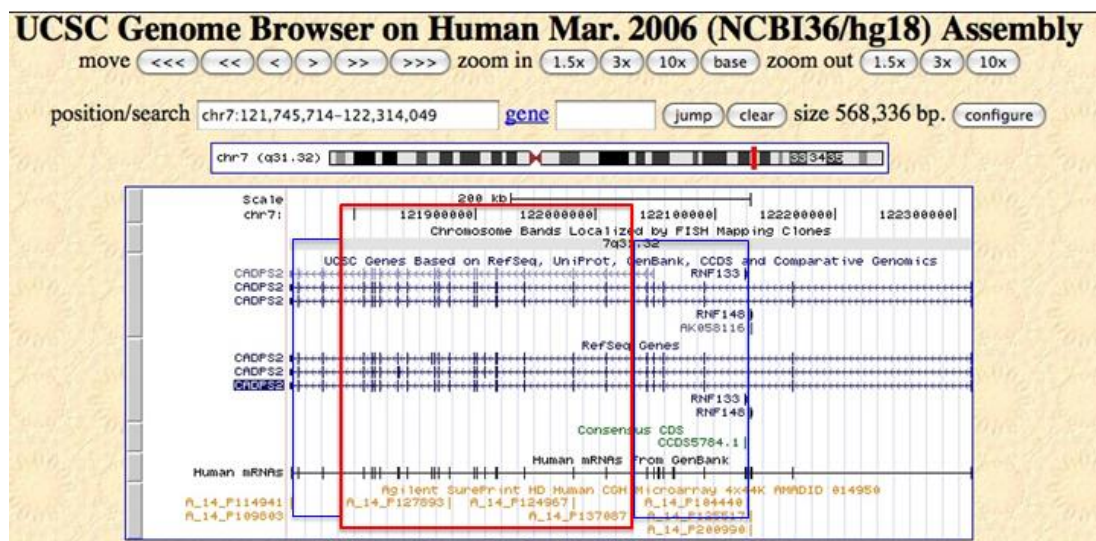


Figura 7.3: confini della delezione che coinvolge il gene *CADPS2*: nel rettangolo rosso è indicata la regione sicuramente deleta che comprende le sonde: da A_14_P109803 (chr7:121788199-121788258) a A_14_P137087 (chr7:122030022-122030081), nel rettangolo in blu sono indicate le regioni incerte, quelle sicuramente non delete sono la sonda A_14_P114941 (chr7:121747156-121747215) e la sonda A_14_P104440 (chr7:122126405-122126464),

Tramite Real time PCR, abbiamo determinato i confini esatti della delezione: al 5' il punto di rottura è tra l'esone 3 (NON deleta in real time) e l'esone 4 (deleta in real time). Al 3' invece il punto di rottura è tra l'esone 26 (deleta in real time) e l'esone 29 (NON deleta in real time) (Fig.7.4).

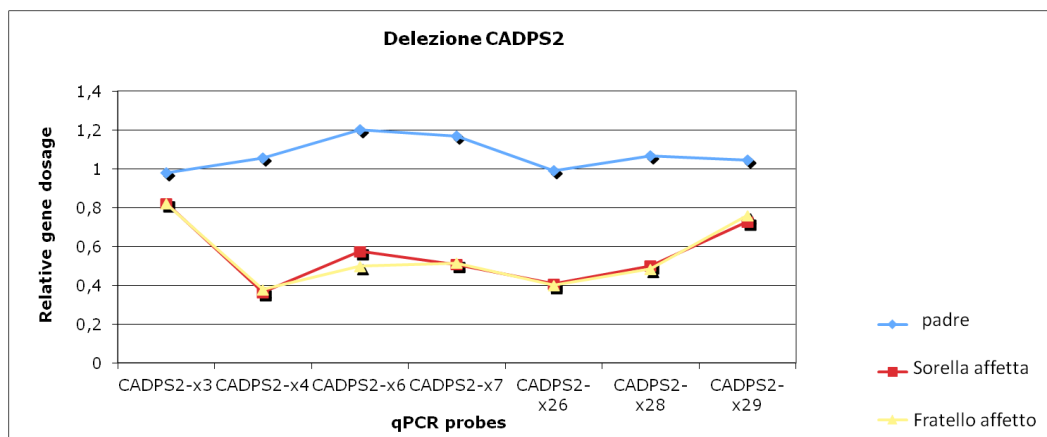


Figura 7.4: Numero di copie dei frammenti del gene *CADPS2* analizzati per Real Time PCR nella famiglia con due figli affetti da ritardo mentale ed autismo, appartenenti al consorzio CHERISH. I frammenti amplificati con i primers specifici per l'esone 4, 6, 7, 26, e per l'esone 28 di *CADPS2* mostrano un relative copy number intorno allo 0.4-0.6 nei due figli affetti (sorella e fratello), indice della presenza della delezione su uno dei due alleli del gene, mentre nel padre tutte le sonde utilizzate hanno un relative copy number intorno all'1, indicando che il padre ha due copie dell'allele e confermando i risultati dell'array CGH

Per verificare se *CADPS2* potesse avere un ruolo nella suscettibilità all'autismo e/o al ritardo mentale, abbiamo sequenziato l'intera regione codificante del gene (30 esoni) in 96 individui italiani affetti da autismo e 36 individui italiani affetti da MR.

Da questo screening di sequenza del gene *CADPS2*, abbiamo identificato due varianti esoniche comuni già descritte in dbSNP (rs2251761 nell'esone 3 e rs2074589 nell'esone 17,) e 6 varianti esoniche rare che sono state identificate in 6 individui differenti (tabella 7.5)

POSIZIONE NEL GENE	ALLELE MAGGIORE / MINORE	TIPO VARIANTE	POLYPHEN SCORE	ORIGINE	MAF NEI CASI DI ASD	MAF NEI CASI DI MR	MAF NEI CONTROLLI
Esone 1		Silent (A26A)	---	MATERNA	1/188	0/74	NA
Esone 3	A/G	Silent t(L252L) rs2251761	-	-	37/174	12/74	2/113**
Esone 6	C/G	Silent(A402A)	---	PATERNA	1/188	0/74	0/500
Esone 13	G/A	M630T	2.662	PATERNA	1/188	0/74	1/516
Esone 13	C/G	F645V	1.248	MATERNA	1/188	0/74	0/500
Esone 17	A/C	Silent (T819T) rs2074589			48/188	6/74	9/113**
Esone 25	G/A	D1113N		MATERNA	1/188	0/74	0/500
Esone 26	G/A	V1137M	1.162	MATERNA	0/188	1/74	0/500

Tabella 7.5: Varianti identificate nello screening del gene *CADPS2* in 94 pazienti affetti da autismo e 37 pazienti affetti da ritardo mentale. ** frequenza descritta in HapMap nella popolazione CEU.

Le varianti rare più interessanti identificate nel gene *CADPS2* sono:

- Nell'esone 13, due mutazioni missenso, la prima è una transizione $T \Rightarrow C$ che provoca un cambio amminoacidico Metionina \Rightarrow Treonina, in posizione 630 (M630T) della sequenza amminoacidica della proteina (fig.7.5); questa mutazione è trasmessa dal padre ed è stata trovata in un solo controllo (1/258 individui di controllo analizzati tramite analisi di sequenza).

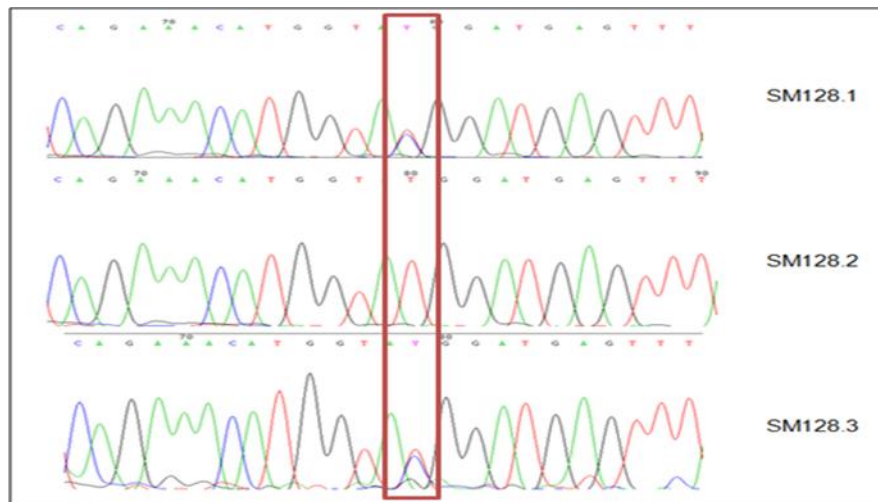


Figura 7.5: Cromatogramma della sequenza dell'esone 13 di *CADPS2* nella famiglia SM_128. Il padre (SM_128.1) trasmette la mutazione missenso (T → C) al figlio (SM_128.3).

- La seconda mutazione trovata nell'esone 13 di *CADPS2* è una transizione T ⇒ G che provoca un cambio amminoacidico Fenilalanina ⇒ Valina, in posizione 645 (F645V).

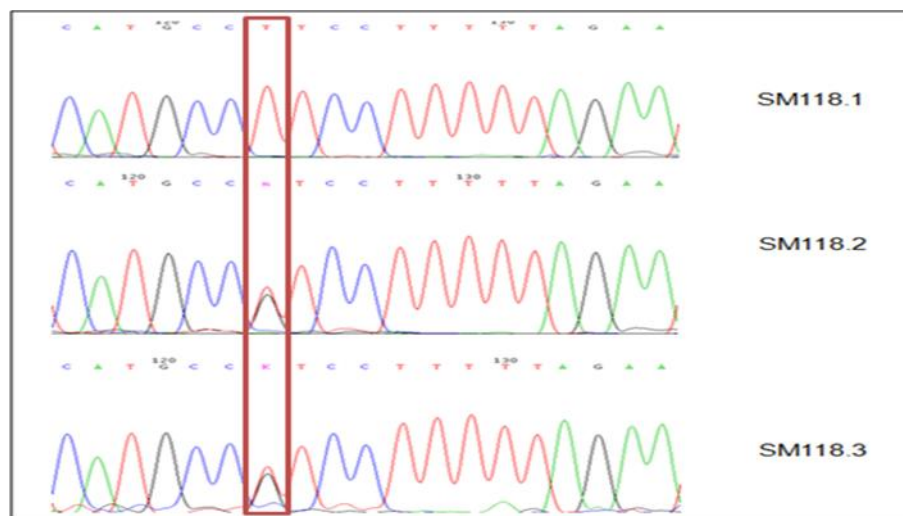


Figura 7.6: cromatogramma della sequenza dell'esone 13 di *CADPS2* nella famiglia SM_118. La madre (SM_118.2) trasmette la mutazione (T → G) al figlio (SM_118.3).

Questa mutazione F645V, ereditata dalla madre, non è stata trovata in circa 250 individui controllo.

- Nell'esone 25, una mutazione G⇒A, ereditata dalla madre, che porta ad un cambiamento D1113N nella sequenza amminoacidica della proteina. Questa variante, non è stata trovata in nessuno dei 250 individui di controllo analizzati.

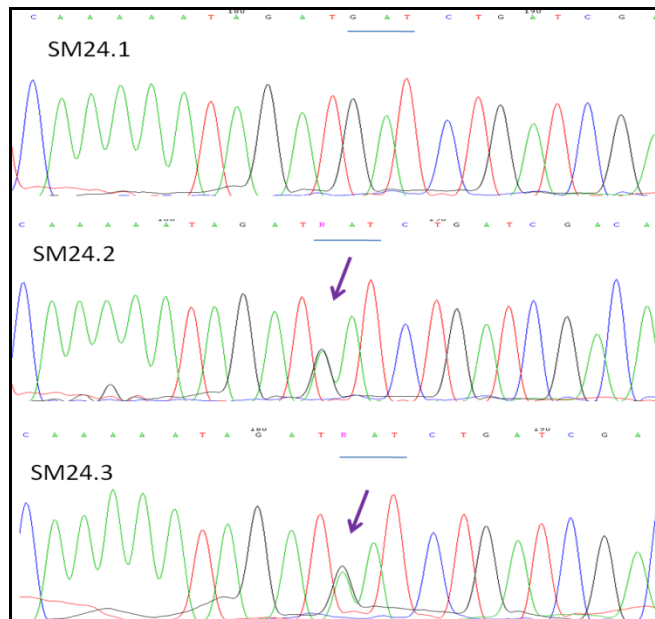


Figura 7.7: cromatogramma della sequenza dell'esone 25 di *CADPS2* nel quale si vede la variante A presente in eterozigosi nell'individuo SM_24.3 e nella madre SM_24.2.

- Nell'esone 26, una mutazione G⇒A in un paziente affetto da ritardo mentale. Questa mutazione porta ad un cambiamento V1137M nella proteina ed è ereditato dalla madre; da notare che questa variante non è stata trovata nei 250 controlli analizzati e non è presente nel fratello non affetto.

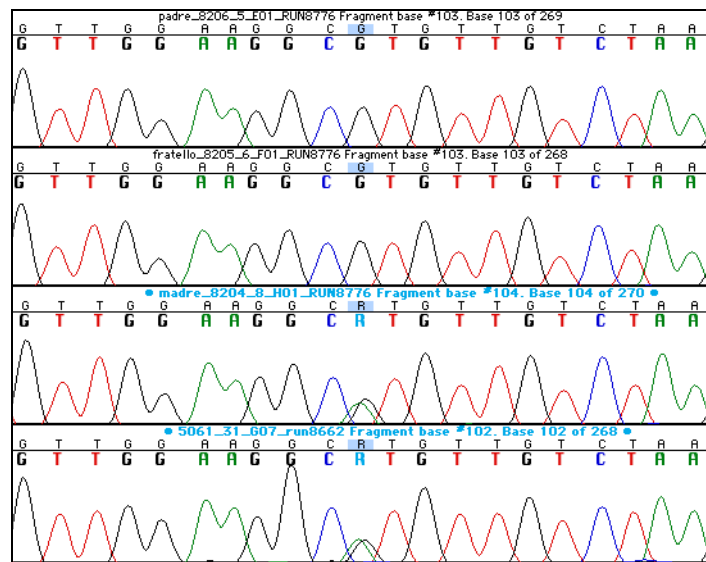


Figura 7.8: cromatogramma della sequenza dell'esone 26 di *CADPS2* nella famiglia con un figlio affetto da ritardo mentale. La madre (8204) trasmette la mutazione (T → G) al figlio (5061) affetto ma non al fratello non affetto .

Quindi, sebbene tutte le varianti trovate sono state ereditate da un genitore, nessuna di queste varianti è descritta in dbSNP o nel database 1000 genomi, e, ad eccezione di una sola variante, tutte sono assenti in 250 controlli italiani. Inoltre, tramite PolyPhen,

uno strumento di predizione bioinformatica abbiamo analizzato l'ipotetico effetto funzionale di queste varianti: solo una variante, la M630T nell'esone 13 di *CADPS2* è stata predetta avere un effetto "dannoso" sulla funzione della proteina da PolyPhen.

7.4. Analisi dell'imprinting di *CADPS2*

Poiché *CADPS2* è localizzato sul cromosoma 7, in una regione soggetta ad *imprinting*, e che la maggior parte delle varianti identificate, non presenti nei controlli, sono ereditate dalla madre, abbiamo deciso di verificare se *CADPS2* fosse soggetto ad imprinting.

A tal fine abbiamo sequenziato gli esoni 3 e 17 di una decina di individui di controllo per identificare quelli che erano eterozigoti almeno uno degli SNP rs2251761 e rs2074589. Successivamente abbiamo estratto e retrotrascritto l'RNA di 7 individui eterozigoti per uno di questi 2 SNPs per controllare se esprimevano entrambi gli alleli. L'analisi della sequenza ha evidenziato che nelle cellule del sangue *CADPS2* è espresso in modo monoallelico, in quanto tutti i controlli analizzati esprimevano un solo allele. (Fig. 7.9)

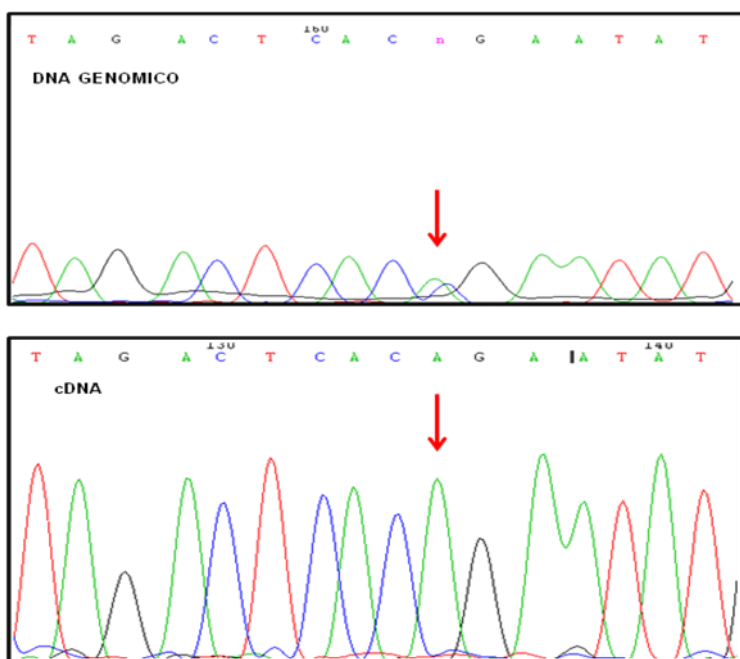


Fig 7.9: cromatogramma della sequenza dell'esone 17 di *CADPS2* del DNA genomico e del cDNA di un individuo di controllo eterozigote per lo SNP rs2074589, nella figura in basso si vede che esprime solo un allele. La posizione dello SNP è indicata dalla freccia rossa.

Al fine di capire se l'espressione monoallelica di *CADPS2* nelle cellule del sangue potesse dipendere dall'origine parentale, confermando l'ipotesi di imprinting, è stata analizzata la mia famiglia, in quanto io stessa sono risultata essere eterozigote per lo SNP rs2251761. E' stato analizzato il DNA di mia madre, mio padre, e mia nonna materna mediante PCR e sequenziamento dei prodotti di PCR, e come si può vedere in figura xxx, io, mia madre e mia nonna siamo eterozigoti per questo SNP, mentre mio padre è omozigote per la variante dello SNP più comune, la G. A questo punto sono andata a vedere quale allele esprimo in sangue e come si può vedere in figura xxx esprimo l'allele ereditato da mia madre.(fig.7.10 e 7.11)

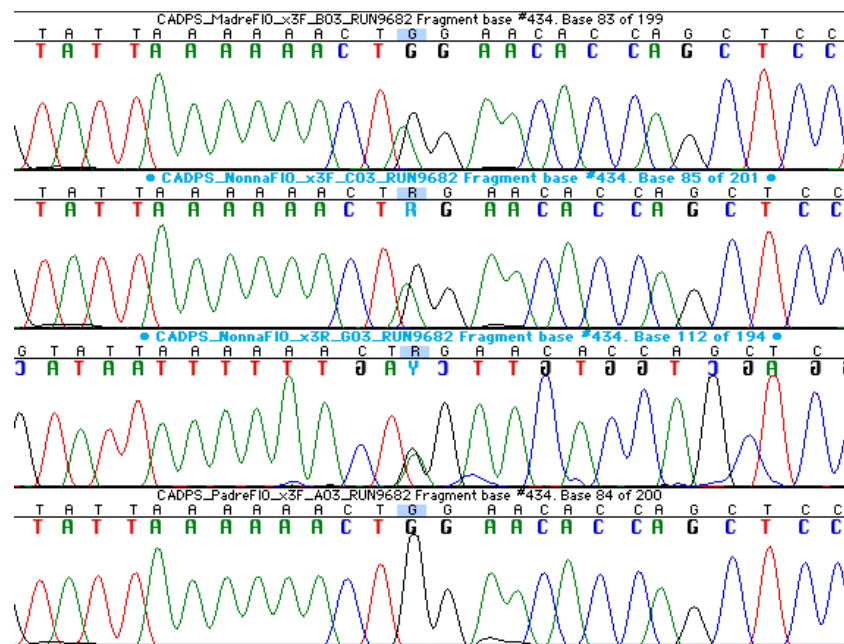


Figura 7.10: cromatogramma della sequenza dell'esone 3 di *CADPS2* nella mia famiglia, dove mia nonna, mia madre ed io siamo eterozigoti per lo SNP rs2251761 interno all'esone 3, mentre mio padre è omozigote per l'allele con la frequenza maggiore di questo SNP.

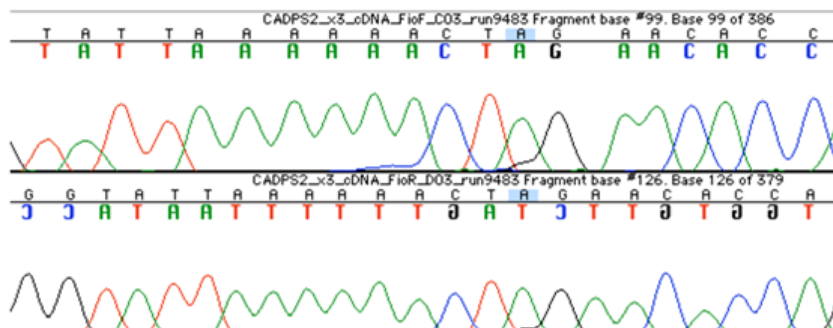


Figura 7.11: cromatogramma del mio cDNA dove si vede che esprimo solo la variante A ereditata da mia madre, indicando che almeno nelle cellule del sangue, *CADPS2* è soggetto ad imprinting ed è espresso solo l'allele materno.

Inoltre, abbiamo analizzato 6 famiglie con un individuo affetto da autismo, che non facevano parte del campione precedentemente analizzato per screening di sequenza, e che sono eterozigoti ed informative per almeno uno dei due SNPs esonici di *CADPS2*. Confrontando il DNA genomico ed il cDNA ottenuto da sangue abbiamo visto che l'unico allele espresso è quello ereditato dalla madre (fig.7.12), confermando ulteriormente i precedenti risultati osservati nella mia famiglia.

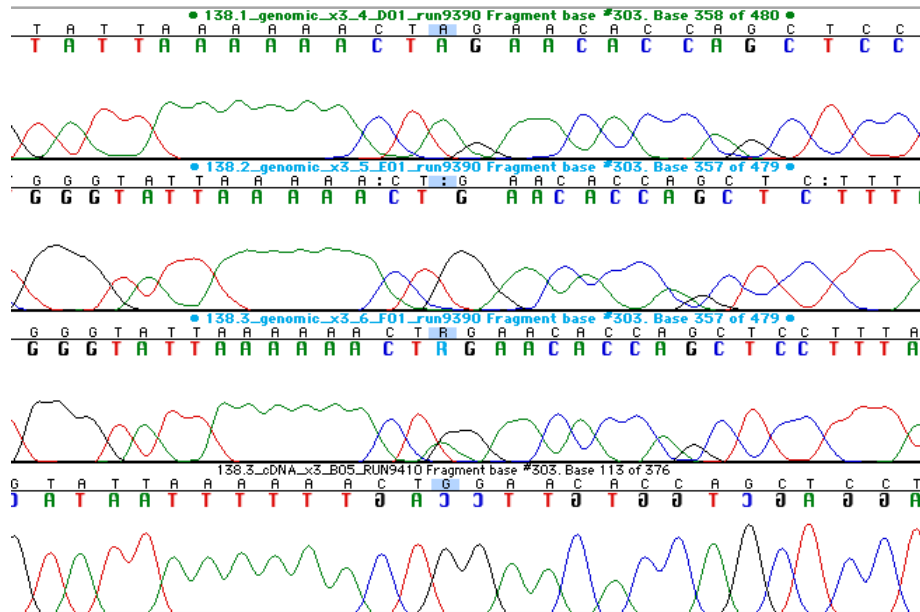


Figura 7.12: cromatogramma della sequenza dell'esone 3 di *CADPS2* nella famiglia SM_138, dove il padre è omozigote per la variante A dello SNP rs2251761, la madre è omozigote per la variante G, il figlio eterozigote, nell'ultimo cromatogramma in basso si vede che nel cDNA del figlio si trova solo la variante G ereditata dalla madre, indicando che almeno nelle cellule del sangue, *CADPS2* è soggetto ad imprinting ed è espresso solo l'allele materno.

E' stato infine possibile analizzare l'espressione allelica di *CADPS2* nella famiglia dell'individuo affetto da ritardo mentale con la variante V1137M nell'esone 26 ereditata dalla madre. Abbiamo potuto confermare che effettivamente il probando esprime solo l'allele materno (fig 7.15 e 7.16), quindi quello con la mutazione, mentre la madre pur essendo eterozigote per la mutazione, esprime solo l'allele wilde type.(fig. 7.13 e 7.14)

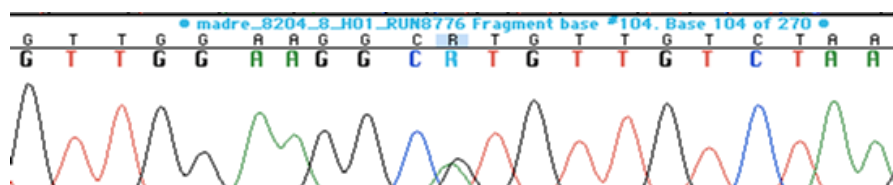


Figura 7.13: cromatogramma della sequenza dell'esone 26 di *CADPS2* del DNA genomico della madre dell'individuo affetto da ritardo mentale, dove si osserva che la madre è eterozigote per la mutazione G/A che nella proteina porta alla sostituzione amminoacidica Val1137Met .

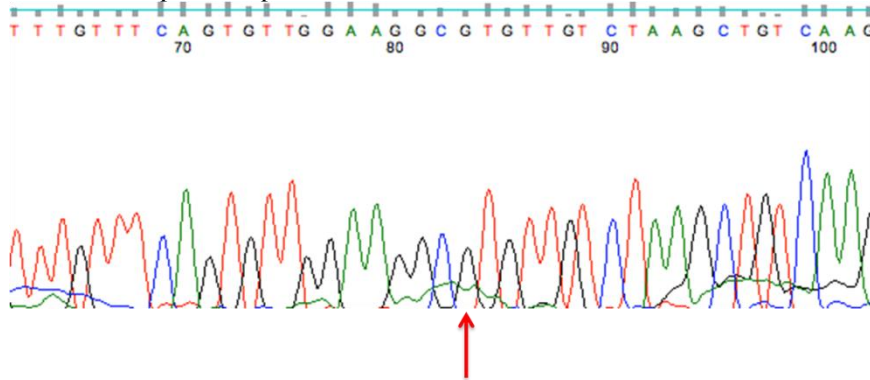


Figura 7.14: cromatogramma della sequenza dell'esone 26 di *CADPS2* del cDNA della madre dell'individuo affetto da ritardo mentale, entrambi eterozigoti per la mutazione G/A che nella proteina porta alla sostituzione amminoacidica Val1137Met , in figura si vede che la madre esprime solo l'allele wilde type (freccia rossa)

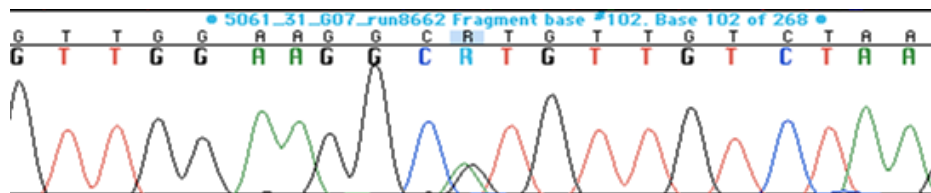


Figura 7.15: cromatogramma della sequenza dell'esone 26 di *CADPS2* del DNA genomico dell'individuo affetto da ritardo mentale, dove si osserva che è eterozigote per la mutazione G/A che nella proteina porta alla sostituzione amminoacidica Val1137Met .

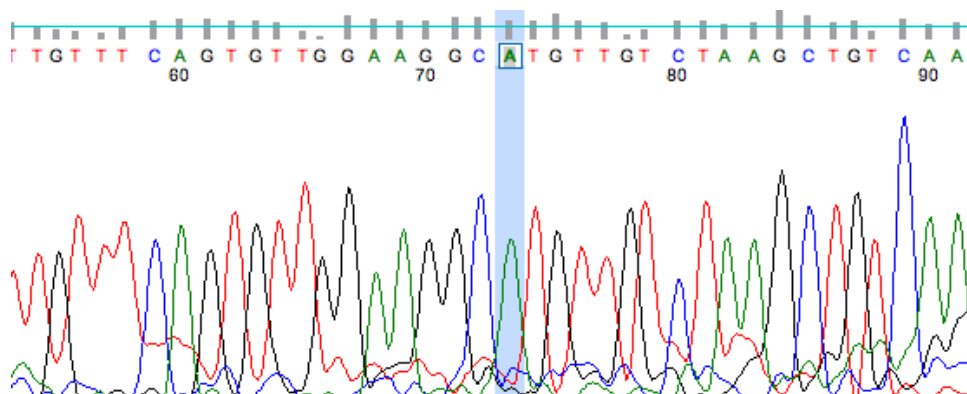


Figura 7.16: cromatogramma della sequenza dell'esone 26 di *CADPS2* del cDNA dell'individuo affetto da ritardo mentale eterozigote la mutazione G/A ereditata dalla madre, che nella proteina porta alla sostituzione amminoacidica Val1137Met , in figura si vede che esprime solo l'allele con la variante A (evidenziata)

8. DISCUSSIONE

8.1. Caratterizzazione di due CNVs identificate nella stessa famiglia

L'autismo ed il ritardo mentale sono malattie complesse la cui espressione fenotipica dipende dall'azione combinata di fattori ambientali e di fattori genetici. L'identificazione delle basi genetiche implicate in queste due patologie è complicata dalla presenza dell'eterogeneità genetica e fenotipica che le caratterizza.

Progressi significativi verso l'identificazione dei geni di suscettibilità a queste due patologie sono stati compiuti nell'ultimo decennio, grazie ad una sempre migliore caratterizzazione della variabilità presente nel genoma umano e allo sviluppo di tecniche basate su microarray. Ciò ha permesso, tramite studi condotti sull'intero genoma, di identificare numerose copy number variants (CNVs) in individui affetti da ritardo mentale ed autismo (~15-20%), alcune delle quali svolgono un ruolo importante nella suscettibilità a queste patologie. Per molto tempo l'autismo è stato considerato una malattia multifattoriale, in cui molteplici varianti di rischio comuni, ciascuna di piccolo effetto, concorrono all'eziologia del disturbo. Tuttavia, nonostante i numerosi studi di associazione condotti su ampie coorti di individui affetti da autismo, ad oggi sono descritti pochi loci con associazione significativa alla patologia e nessuno di questi è stato confermato in più studi. Questi dati indicano che, nella maggioranza dei casi, la varianti comuni hanno solo un debole effetto sul fenotipo autistico.

Per contro, recenti studi hanno avvalorato l'ipotesi che una significativa percentuale di casi di autismo e di ritardo mentale sia causata da singole varianti genetiche rare, sia CNVs che varianti di sequenza di piccola scala, ereditate o *de novo*, che, se considerate singolarmente, sono implicate in un piccolissimo numero di casi, ma che complessivamente potrebbero spiegare una buona parte dell'ereditabilità alla base della malattia ed indicare pathways preferenzialmente alterati in questa patologie (come, ad esempio, lo sviluppo e la funzione delle sinapsi) [Toro et al., 2010].

In questo contesto si inserisce il mio progetto di tesi, volto all'identificazione e caratterizzazione di varianti strutturali in individui affetti da autismo e in individui affetti da ritardo mentale.

Nell'ambito dell'autismo mi sono occupata della caratterizzazione molecolare di una delezione identificata in una famiglia con due individui affetti da autismo e che coinvolge due geni contigui: il gene *IMMP2L* e il gene *DOCK4*.

Il gene *IMMP2L* è localizzato sul cromosoma 7 e codifica per una proteina mitocondriale intermembrana. Riarrangiamenti cromosomici che distruggono questo gene sono stati identificati in individui affetti dalla sindrome di Tourette, una complessa malattia neuropsichiatrica con caratteristiche fenotipiche parzialmente sovrapponibili all'autismo. Tuttavia delezioni nel gene *IMMP2L* sono già descritte in DGV e nel nostro screening sono state trovate anche in individui di controllo.

Il gene *DOCK4* rappresenta un buon candidato funzionale per l'autismo, essendo ampiamente coinvolto nello sviluppo neuronale, codifica per un fattore di scambio nucleotidico della guanidina, GDP-GTP (GEF, Guanine nucleotide Exchange Factor) in grado di attivare Rap1 e Rac1, due piccole GTPasi. Le proteine Rap sono coinvolte nel controllo dell'adesione e delle giunzioni cellulari. La proteina *DOCK4* è attivata da RhoG, e regola la migrazione cellulare tramite l'attivazione di Rac, che è importante per la crescita e la formazione dei neuriti. Inoltre uno studio ha evidenziato la sua elevata espressione nel tessuto cerebrale, dove svolge un ruolo essenziale per la corretta formazione e morfologia dei dendriti. Infatti, esperimenti sui ratti hanno dimostrato che lo spegnimento (knockdown) del gene *DOCK4* risulta in una ridotta crescita e ramificazione dei dendriti [Ueda et al., 2008]. Inoltre, geni implicati nella via di segnalazione mediata dalle GTPase/Ras sono stati descritti tra i vari pathways coinvolti nell'ASD (Pinto et al, 2010). Analogamente, in un precedente studio, condotto dal laboratorio dove ho svolto la tesi, sono state trovate rare varianti non sinonime che segregano con l'ASD in cinque famiglie, nel gene *cAMP-GEFII* che codifica per un'altra GEF (Bacchelli et al, 2003). Infine nessuna delezione di *DOCK4* è presente in DGV e non abbiamo identificato nessuna CNV di *DOCK4* nei controlli analizzati.

La segregazione della delezione di *IMMP2L-DOCK4* all'interno della famiglia 15_0084 ci aveva indotto ad ipotizzare che l'aploinsufficienza del gene *DOCK4* potesse rappresentare un fattore di suscettibilità all'ASD, con una penetranza differente in relazione al sesso, in quanto questa delezione è trasmessa dalla madre dislessica, ai due figli maschi affetti da autismo e alla figlia affetta da dislessia.

Quindi abbiamo effettuato uno screening per la presenza di CNV di *DOCK4* in altre famiglie multiplex con autismo e in altri individui di controllo. Questo screening non

ha portato all'identificazione di nessun aggiuntiva CNV in *DOCK4*. Tuttavia, l'analisi di segregazione della delezione di *IMMP2L-DOCK4* in alcuni fratelli e nipoti della madre della famiglia 15-0084, ha evidenziato che i portatori della delezione erano affetti da dislessia, facendo quindi ipotizzare che *DOCK4* possa avere un ruolo nella suscettibilità alla dislessia. Grazie alla collaborazione con altri gruppi di ricerca (consorzio Neurodys, il consorzio Genedys, il gruppo di Oxford-Cardiff, il gruppo dei Francesi), abbiamo effettuato uno screening di CNVs in *DOCK4* in altri 1940 individui con diagnosi di dislessia e 4646 individui di controllo, che ci ha permesso di identificare un'altra delezione, che rimuove dall'esone 38 al 52 del gene *DOCK4*, e che segrega con il fenotipo dislessico. Quindi, in definitiva, abbiamo identificato su 1941 individui dislessici analizzati, 2 delezioni nel gene *DOCK4* che segregano con la dislessia in due differenti nuclei familiari, mentre nessuna delezione è stata individuata all'interno dei 4646 individui di controllo analizzati. Sebbene questa differenza non sia statisticamente significativa (0.09), la totale assenza di delezioni in un grande numero di controlli suggerisce che le delezioni nel gene *DOCK4* possano contribuire alla dislessia, anche se in un numero molto ridotto di casi. Tuttavia, ulteriori studi saranno necessari per comprendere il potenziale ruolo di *DOCK4* nella dislessia: ad esempio sarebbe interessante effettuare un'analisi di sequenza del gene *DOCK4* in un numeroso campione di individui affetti da dislessia al fine di verificare se, in aggiunta ai due casi di microdelezione, varianti rare di sequenza in *DOCK4* possano essere implicate nella suscettibilità a questa patologia.

In seguito, ci siamo chiesti se nella famiglia 15_0084 potesse segregare un'ulteriore alterazione genetica, in modo da spiegare la diversa espressione fenotipica nei vari membri della famiglia: grave autismo nei due fratelli e dislessia nella madre e nella sorella.

Mediante l'Illumina 1M SNP array, è stato analizzato l'intero genoma degli individui appartenenti alla famiglia 15-0084, al fine di identificare l'eventuale presenza di una seconda CNV, in un'altra regione del genoma, che potesse agire come modificatore genetico. Da quest'analisi abbiamo identificato una seconda rara microdelezione nella regione cromosomica 2q14.1 che distrugge il gene *CNTNAP5*. Questa seconda delezione segrega con il fenotipo autistico, in quanto è trasmessa dal padre, che presenta tratti autistici (anche se non ha una diagnosi formale di autismo), ad entrambi i figli con diagnosi di autismo classico e non alla sorella, che non presenta tratti autistici ma solo dislessia. L'identificazione di questa delezione è di particolare

interesse: *CNTNAP5* codifica, infatti, per una proteina trans-membrana, un membro della famiglia proteica delle Caspr, che appartengono alla superfamiglia delle neurexine, le quali mediano il contatto e le interazioni cellulari nel sistema nervoso, giocando un ruolo essenziale nel suo corretto sviluppo e nel suo funzionamento. Numerosi studi hanno evidenziato un ruolo importante delle neurexine e delle neuroligine, loro partner di interazione, nel fenotipo autistico (Jamain et al., 2003; Laumonnier et al., 2004; Jamain et al., 2005; Feng et al., 2006; AGP, 2007). Inoltre, è stato recentemente descritto un caso di un individuo affetto da ritardo mentale e problemi nel linguaggio, che presenta un complesso riarrangiamento cromosomico che coinvolge i cromosomi 1, 2 e 15 e che porta alla perdita completa di numerosi geni tra cui *CNTNAP5* [Ballarati et al., 2009]. Infine, un altro membro di questa famiglia, *CNTNAP2*, è stato recentemente implicato nella suscettibilità nell'autismo: alterazioni di questo gene sono state correlate alla presenza di ritardo del linguaggio e precisamente ad un ritardo dell'età in cui viene detta la prima parola o la prima frase di senso compiuto (D.A. Stephan, 2008; Alarcòn et al. 2008). Allo scopo di evidenziare altre eventuali mutazioni rare presenti in questo gene, abbiamo sequenziato tutta la regione codificante di *CNTNAP5* in 143 probandi. Questo screening ha identificato 3 mutazioni non sinonime, non descritte in dbSNP, che segregano con il fenotipo autistico in tutte e tre le famiglie, due delle quali sono completamente assenti nei controlli analizzati, suggerendo un possibile coinvolgimento di *CNTNAP5* nella suscettibilità all'autismo.

Per meglio capire il ruolo di *CNTNAP5* nell'autismo sarebbe utile estendere lo screening di questo gene su un campione molto più grande di individui affetti.

E' anche di particolare interesse l'identificazione di due diverse varianti strutturali probabilmente patogenetiche che segregano contemporaneamente nella stessa famiglia, le quali possono interagire tra loro nel determinare una diversa espressione fenotipica nella sfera cognitivo-comportamentale. Infatti, diversi recenti studi di CNV nell'autismo e in patologie neuropsichiatriche, hanno fatto emergere l'ipotesi di un meccanismo "two hit". Ad esempio Girirajan e collaboratori (2010), hanno dimostrato che in individui con la microdelezione 16p12.1 affetti da patologie cognitive o psichiatriche, la presenza di una seconda mutazione nella stessa regione genomica o in altri geni facenti parte dello stesso pathway funzionale, determina un fenotipo più grave rispetto alla presenza della sola delezione. Emerge così il possibile coinvolgimento di un ampio pathway sinaptico in cui mutazioni in vari geni coinvolti

nello sviluppo cerebrale (come *DOCK4* che svolge un ruolo importante nella corretta formazione dei dendriti e *CNTNAP5* che media il contatto e le interazioni nel sistema nervoso), agiscono in sinergia determinando la comparsa del fenotipo autistico. Nel nostro caso, in accordo con la “two hit hypothesis”, possiamo ipotizzare che solo la presenza di entrambe le delezioni nei geni *IMMP2L-DOCK4* e *CNTNAP5* porti allo sviluppo del fenotipo autistico, mentre la sola delezione in *DOCK4* possa portare all’insorgenza di dislessia.

8.2. Copy Number Variants nel ritardo mentale

Una delle maggiori difficoltà negli studi di CNV in patologie come l’autismo o il ritardo mentale e’ l’interpretazione del possibile effetto patologico delle CNV identificate. Questa difficoltà di interpretazione è esemplificata dai risultati del mio studio, volto all’analisi mediante array CGH di circa 200 pazienti affetti da ritardo mentale, alcuni dei quali oltre al ritardo mentale presentano anche autismo.

L’analisi dei cromosomi tramite microarray è stato recentemente proposto come test genetico prioritario da eseguire in tutti i casi di ritardo dello sviluppo/deficit intellettivo, disturbi dello spettro autistico o anomalie congenite multiple (Miller et al, 2010), permettendo di aumentare notevolmente la resa diagnostica rispetto all’analisi standard del cariotipo. Il coinvolgimento di numerose ed eterogenee varianti strutturali submicroscopiche nella patogenesi di questi disordini è stato infatti ampiamente documentato. Per questo motivo, nell’ambito del progetto europeo CHERISH, che ha come scopo quello di migliorare la diagnosi di ritardo mentale e approfondirne l’eziologia, abbiamo applicato questa tecnica per l’analisi di un’ampia casistica di 218 pazienti affetti da ritardo mentale, sia sindromico che isolato, e provenienti da diversi paesi europei.

Per l’analisi è stato utilizzato l’array 44k di Agilent, che ha una risoluzione relativamente bassa, in quanto permette l’identificazione delle CNVs con una dimensione media di 75-100 kb. Questo tipo di array è molto utilizzato nel contesto della diagnostica genetica, perché facilita l’interpretazione delle CNV identificate, in quanto le CNVs più piccole sono anche le più difficili da interpretare. Tuttavia, non si può escludere che anche piccole CNVs possano avere un ruolo nella

suscettibilità alla malattie e quindi lo screening mediante questo tipo di array potrebbe non essere esaustivo.

Al fine di effettuare una corretta classificazione delle nuove CNVs identificate, abbiamo seguito i criteri comunemente utilizzati, che principalmente si basano sul confronto con i dati di CNVs disponibili in database pubblici, sia relativi ad individui di controllo (presenti, ad esempio, in DGV), sia relativi a individui affetti da RM e autismo (che si possono trovare, ad esempio, in DECIPHER).

Per facilitare l'interpretazione delle CNV identificate sarebbe inoltre utile analizzare l'intero nucleo familiare, in quanto le CNV *de novo* hanno una maggiore probabilità di essere patogenetiche, anche se non si può escludere che anche quelle ereditate possano avere un ruolo nello sviluppo di questi fenotipi complessi. Infatti, in letteratura sono descritti numerosi casi di CNVs ereditate da genitori sani, che tuttavia presentano un ruolo nella patologia. Questa discrepanza viene generalmente giustificata dalla penetranza incompleta e dal differente background genetico. Ad esempio nel nostro studio la delezione 16p11.2, associata alla suscettibilità all'autismo in studi precedenti, è stata identificata in due individui affetti da MR, e in un caso risultava ereditata dal padre mentre nell'altro caso mostrava un'insorgenza *de-novo*.

Estendere l'analisi di una CNV all'intero nucleo familiare, è essenziale soprattutto per l'identificazione di nuove sindromi da microdelezioni/microduplicazioni. Questa analisi dei familiari dei casi risultati positivi non è stata sempre possibile a causa della mancanza di materiale genetico per alcuni genitori degli individui analizzati. Soltanto per 32 delle 135 CNV identificate è stato possibile determinare l'origine parentale.

Anche il confronto con i dati di letteratura può essere di aiuto ai fini di una corretta classificazione. Ad esempio la delezione 1q44, identificata in un paziente italiano di quattro anni che presentava un lieve ritardo psicomotorio e dimorfismi, è stata classificata come CNV patogenetica in quanto, sebbene non sia descritta in DECIPHER, si sovrappone alla regione comune delle delezioni descritte in letteratura in 5 pazienti con ritardo del linguaggio, crisi apoplettiche e ipoplasia del corpo calloso (Caliebe et al, 2010; Poot et al, 2007).

Infine, un altro elemento utile per classificare le CNV identificate è l'indagine del ruolo funzionale dei geni compresi nella regione alterata.

L'analisi dei fattori sopra descritti (segregazione familiare, descrizione in database e dati di letteratura, contenuto genico) all'interno di questo studio ha permesso di

definire la classificazione delle 135 CNVs identificate nei 218 pazienti analizzati. In particolare, 14 CNVs sono state definite come sicuramente patogenetiche perché descritte in letteratura e nel database DECIPHER; 25 sono state definite probabilmente patogenetiche perché sebbene non siano state precedentemente descritte, coinvolgono un elevato numero di geni oppure segregano con il ritardo mentale in nuclei familiari con più di un individuo affetto; 92 alterazioni sono risultate benigne in seguito alla valutazione di tutti i fattori descritti.

Tuttavia nel nostro studio, molte CNVs restano classificate come di significato incerto poiché ereditate da un genitore sano, oppure presenti sia in DGV che in DECIPHER o ancora perché, essendo duplicazioni e mai precedentemente descritte, è difficile darne un'interpretazione. Dopo l'analisi di tutte le alterazioni osservate, ben 23 rimangono non definibili. Ne è un esempio la duplicazione della regione Xq13, che sebbene sia descritta in un controllo in DGV e sia ereditata dalla madre, che non presenta il fenotipo, è anche descritta in un paziente affetto da ritardo mentale in DECIPHER e anche in questo caso viene ereditata da un genitore sano.

Per accertare il ruolo delle CNV più interessanti identificate in questo studio e stabilire una correlazione tra genotipo e fenotipo, ulteriori studi saranno effettuati da parte del nostro laboratorio o dagli altri laboratori appartenenti al consorzio CHERISH. In particolare, io mi sono occupata della caratterizzazione di una delezione presente in due fratelli, che oltre al RM mostravano anche dei tratti autistici. Questa delezione porta alla distruzione del gene *CADPS2* (*Calcium-dependent secretion activator 2*), localizzato sul cromosoma 7q, all'interno del locus di suscettibilità all'autismo. Le CADPS sono proteine coinvolte nei meccanismi esocitotici sia costitutivi che regolati, e partecipano alla fusione tra la membrana delle vescicole e la membrana plasmatica, in maniera dipendente dalla concentrazione del calcio per cui ad alte concentrazioni di calcio attivano la fusione delle membrane esocitotiche. CADPS2 è una proteina associata alle vescicole secretorie coinvolte nel rilascio delle neurofine (BDNF ed NT-3). Recenti studi hanno dimostrato che i topi knockout per *CADPS2* mostrano caratteristiche comportamentali simili ai pazienti autistici (diminuito comportamento esplorativo, ridotta interazione sociale, aumentata ansia per nuovi ambienti, iperattività, deficit nella regolazione dei ritmi circadiani, deficit nei comportamenti materni) e hanno un cervello con anomalie cellulari simili a quelle riscontrate in pazienti autistici. Questi autori hanno dimostrato che la perdita di

CADPS2 colpisce il rilascio sia costitutivo che regolato di BDNF, che è un fattore importante per la formazione delle sinapsi.(Sedakata et al, 2007- 2010).

Al fine di verificare se *CADPS2* potesse svolgere un ruolo nella suscettibilità all'ASD e al ritardo mentale, abbiamo sequenziato tutte le regioni codificanti in 37 individui affetti da ritardo mentale e 94 individui affetti da autismo. Il risultato più interessante di questa analisi è stata l'identificazione di 4 mutazioni missenso (3 in individui affetti da autismo ed una in un individuo con ritardo mentale) in 4 differenti nuclei familiari. Dato che 3 delle varianti identificate sono state ereditate dalla madre e che *CADPS2* è localizzato in una regione contenente geni generalmente soggetti ad *imprinting*, abbiamo verificato se anche *CADPS2* potesse essere soggetto ad *imprinting*. A tal fine, l'analisi dei genotipi di due SNP altamente polimorfici nell'esone 3 e nell'esone 17 di *CADPS2*, sia nel DNA genomico sia nell'RNA estratto dalle cellule del sangue, ci ha permesso di stabilire che nelle cellule del sangue *CADPS2* ha un'espressione monoallelica ed esprime solamente l'allele materno. Pertanto, i 3 individui portatori della variante missenso ereditata dalla madre, esprimono solamente l'allele mutato, mentre nei due fratelli con la delezione materna, identificata mediante array CGH, l'espressione di *CADPS2* è assente. Questo risultato è di particolare interesse se si considera che uno studio di linkage condotto dal consorzio IMGSAC ha mostrato l'esistenza di due loci sul cromosoma 7, tra loro adiacenti, che presentano evidenza di linkage significativa con il fenotipo autistico, in base all'origine parentale (Lamb et al, 2005), suggerendo la presenza, in queste due regioni, di geni soggetti ad *imprinting* coinvolti nella suscettibilità all'autismo. Il gene *CADPS2* si trova proprio all'interno della regione che mostra evidenza di linkage quando si considerano le trasmissioni materne.

Dal momento che sia l'autismo che il ritardo mentale sono patologie legate allo sviluppo e al funzionamento del sistema nervoso, sarebbe interessante poter testare se *CADPS2* sia soggetto ad *imprinting* anche a livello del SNC. A tale scopo, analizzeremo campioni di mRNA estratti da cervelletto, corteccia e amigdala, e il corrispondente DNA genomico di diversi individui di controllo.

Inoltre, dal momento che l'espressione allele-specifica dei geni soggetti ad *imprinting* è legata a specifiche modificazioni epigenetiche, come la metilazione delle citosine, analizzeremo se nella sequenza del DNA di *CADPS2* ci siano dei siti differenzialmente metilati.

Infine, per comprendere meglio il ruolo che varianti di sequenza rare in *CADPS2* possano avere nella suscettibilità all'ASD, estenderemo questo screening a tutti i probandi appartenenti alle famiglie del consorzio IMGSAC che condividono 2 alleli o solamente l'allele materno nella regione genomica corrispondente a *CADPS2*.

Quindi, nell'insieme, i risultati che ho presentato sottolineano l'importanza di variante genetica rare nella suscettibilità a fenotipi complessi come l'autismo ed il ritardo mentale. Tuttavia, le varianti rare ad oggi identificate sono in grado di spiegare solo in una piccola percentuale dei casi. Dato che, dai recenti studi genomici, è ormai chiaro che varianti comuni sembrano svolgere un ruolo minore nella suscettibilità a queste patologie, si ritiene che una grande porzione dell'ereditabilità che risulta ancora essere inspiegata (la cosiddetta "ereditabilità mancante", (Manolio et al, 2009)) sia da ricercarsi nelle varianti rare di sequenza.

Al fine di identificare tali varianti, la strategia vincente sarebbe quella di sequenziare l'intero genoma in un numero elevato di pazienti: questo approccio sta diventando sempre più fattibile grazie all'avvento e diffusione delle tecnologie di "sequenziamento di nuova generazione" (NGS), anche se i costi rappresentano ancora un limite per l'analisi di grandi coorti. Una strategia alternativa consiste nell'utilizzare la tecnologia NGS per analizzare tutte le regioni codificanti che compongono il genoma umano (il cosiddetto "esoma"). Questo approccio, rispetto al sequenziamento dell'intero genoma, riduce enormemente i costi e allo stesso tempo favorisce l'identificazione di rare varianti funzionali, senza la necessità di formulare ipotesi a priori. È stato già dimostrato che queste tecnologie sono un potente strumento di indagine per l'identificazione di geni implicati in malattie monogeniche (reviewed in Bamshad et al, 2011) e il recente lavoro di Najmabadi e collaboratori (2011), in cui hanno analizzato 136 famiglie consanguinee con RM non sindromico mediante una combinazione di mappaggio di omozigosigiosità e sequenziamento dell'esoma, ha sottolineato il potere delle tecnologie NGS nella comprensione delle basi genetiche delle malattie complesse. Dato che la maggiore difficoltà di questi studi consiste nel discriminare tra polimorfismi neutrali e mutazioni patogeniche, come emerge dallo studio di Najmabadi, l'applicazione di criteri specifici per la scelta degli individui da analizzare, quali la presenza di consanguineità, risulta essere molto utile.

Un altro criterio che stato utilizzato con successo è stato quello di restringere l'analisi alle sole varianti *de novo*, dato che una maggiore prevalenza di mutazioni *de novo*

potrebbe spiegare perchè malattie associate a fertilità molto ridotta, come disordini psichiatrici e dello sviluppo neurologico, rimangono frequenti nella popolazione umana (Vissers et al, 2010; B. J O'Roak et al,2011).

Infine, come si evince anche dal mio lavoro di tesi e dagli altri studi precedentemente discussi, per la comprensione dei meccanismi molecolari alla base di patologie complesse quali l'autismo ed il ritardo mentale è di cruciale importanza la collaborazione dei gruppi di ricerca e la creazione di consorzi internazionali, in cui ci sia una condivisione di risorse e conoscenze, al fine di collezionare migliaia di pazienti su cui poter effettuare analisi utilizzando le tecnologie all'avanguardia e di interpretare i risultati ottenuti.

9.BIBLIOGRAFIA

Abbasi-Moheb L, Garshasbi M, Kahrizi K, Jensen LR, Wrogemann K, et al. 2009. Two independent mutations in the *ZNF526* gene cause non syndromic autosomal recessive mental retardation. In *14th International Workshop on Fragile X and X-Linked Mental Retardation*, Abstr. no. 9, 17 pp. Bahia, Brazil

Abecasis, et al. Extent and distribution of linkage disequilibrium in three genomic regions. *Am. J. Hum. Genet.*, 2001; 68:191–197.

Abramson RK, Wright HH, Carpenter R, Brennan W, Lumpuy O, Cole E, Young SR. Elevated blood serotonin in autistic probands and their first-degree relatives. *J. Autism Dev. Disord.* 1989; 19: 397-407

AGP. Mapping autism risk loci using genetic linkage and chromosomal rearrangements. *Nat Genet.*, 2007; 39:319-28.

Aicardi J. (1998) The etiology of developmental delay. *Semin Pediatr Neurol*; 5(1):15-20.

Allen G, Courchesne, E. Differential effects of developmental cerebellar abnormality on cognitive and motor functions in the cerebellum: an fMRI study of autism. *American Journal of Psychiatry* 2003, 160:262-273.

Anney R, Klei L, Pinto D, Regan R, Bacchelli E, Bailey AJ, Baird G, Battaglia A, Berney T, Bolshakova N, Bölte S, Bolton PF, Bourgeron T, Brennan S, Brian J, Carson AR, Casallo G, Casey J, Chu SH, Duque F, Maestrini E, Mahoney W, Mantoulan C, Marshall CR, McConachie Ennis S, Hallmayer J et al. A genome-wide scan for common alleles affecting risk for autism. *Hum Mol Genet.* 2010.

Amor DJ, Halliday J. (2008) A review of known imprinting syndromes and their association with assisted reproduction technologies. *Hum Reprod*; 23(12):2826-34.

Ashley-Koch A, Wolpert CM, Menold MM, Zaem L, Basu S, Donnelly SL, Ravan SA, Powell CM, Qumsiyeh MB, Aylsworth AS, Vance JM, Gilbert JR, Wright HH, Abramson RK, DeLong GR, Cuccaro ML, Pericak-Vance MA. Genetic studies of autistic disorder and chromosome 7. *Genomics* 1999, 61:227-36

Autism Genome Project Consortium (AGP) Mapping autism risk loci using genetic linkage and chromosomal rearrangements. *Nat Genet.* 2007; 39(10):1285.

M. Alarcón, B. Abrahams, J. Stone, J. Duvall, J. Perederiy, J. Bomar, J. Sebat, M. Wigler, C. Martin, D. Ledbetter. Linkage, Association, and Gene-Expression Analyses Identify CNTNAP2 as an Autism-Susceptibility Gene. *The American Journal of Human Genetics*, 2008; 82:150-159.

Alkan C, Kidd JM, Marques-Bonet T: Personalized copy number and segmental duplication maps using next generation sequencing. *Nat Genet* 2009; 41: 1061-1067.

Amir RE, Van den Veyver IB, Schultz R et al: Influence of mutation type and X chromosome inactivation on Rett syndrome phenotypes. *Ann Neurol* 2000; 47: 670–679.

Amir R.E., Van den Veyver I.B., Wan M, Tran C.Q., Francke U., Zoghbi H.Y. Rett syndrome is caused by mutations in X-linked MECP2, encoding methyl- CpG-binding protein 2. *Nat. Genet.*, 1999; 23:185-188.

Asperger H, Die autistischen psychopathen im Kindesalter. *Archiv für Psychiatrie und Nervenkrankheiten*, 1949; 117:76–136.

Bacchelli E., Blasi F., et al. Screening of nine candidate genes for autism on chromosome 2q reveals rare nonsynonymous variants in the cAMP-GEFII gene. *Molecular Psychiatry*, 2003; 8: 916-924

Bailey, LeCouteur, Gottesman, Bolton, Simonoff, Yuzda, Rutter, Autism as a strongly genetic disorder: evidence from a British twin study. *Psychol Med*, 1995; 25: 63-77.

Bailey A, Luthert P, Bolton P, Le Couteur A, Rutter M, Harding B. Autism and megalencephaly. *Lancet* 1993, 341:1225-6

Bailey A, Luthert P, Dean A, Harding B, Janota I, Montgomery M, Rutter M, Lantos P. A clinicopathological study of autism. *Brain* 1998°, 121 (Pt 5):889-905

Bailey, Palferman, Heavey, LeCounter, Autism: the phenotype in relatives. *J Aut Dev Dis*, 1998; 28:439-445

Baron-Cohen S, Ring HA, Wheelwright S, Bullmore ET, Brammer MJ, Simmons A, Williams SC. Social intelligence in the normal and autistic brain: an fMRI study. *Eur J Neurosci* 1999, 11:1891-8

Berkel S, Marshall CR, Weiss B, Howe J, Roeth R, Moog U, Endris V, Roberts W, Szatmari P, Pinto D, Bonin M, Riess A, Engels H, Sprengel R, Scherer SW, Rappold GA. Mutations in the SHANK2 synaptic scaffolding gene in autism spectrum disorder and mental retardation. *Nat Genet.* 2010; 42(6):489-91

Bessou C, Giuglia JB, Franks CJ, Holden-Dye L, Ségalat L, Mutations in the *Caenorhabditis elegans* dystrophin-like gene *dys-1* lead to hyperactivity and suggest a link with cholinergic transmission, *Neurogenetics*, 1998. 2:61-72.

Bishop DV. Autism, Asperger's syndrome and semantic-pragmatic disorder: where are the boundaries? *Br J Disord Commun* 1989, 24:107-21

Bierut , L.J. et al. A genome-wide association study of alcohol dependence. *Proc Natl Acad Sci U S A*. 2010;107(11):5082-7.

Blasi F, Bacchelli E, Pesaresi G, Carone S, Bailey AJ, Maestrini E; International Molecular Genetic Study of Autism Consortium (IMGSAC). Absence of coding

- mutations in the X-linked genes neuroligin 3 and neuroligin 4 in individuals with autism from the IMGSAC collection. *Am J Med Genet B Neuropsychiatr Genet.* 2006; 141B(3):220-1
- Bagni C, Greenough WT: From mRNP trafficking to spine dysmorphogenesis: the roots of fragile X syndrome. *Nat RevNeurosci* 2005; 6: 376– 387.
- Bailey, LeCounteur, Gottesman, Bolton, Simonoff, Yuzda, Rutter. Autism as a strongly genetic disorder: evidence from a British twin study. *Psychol Med*, 1995; 25:63-77.
- Bailey A, Luthert P, Bolton P, Le Couteur A, Rutter M, Harding B. Autism and megalencephaly. *Lancet* 1993, 341:1225-6
- Bailey A, Luthert P, Dean A, Harding B, Janota I, Montgomery M, Rutter M, Lantos P. A clinicopathological study of autism. *Brain* 1998°, 121 (Pt 5):889-905
- Bailey A, Palferman S, Heavey L, Le Couteur A. Autism: the phenotype in relatives. *J Autism Dev Disord* 1998 b, 28:369-92
- Bailey A, Phillips W, Rutter M. Autism: towards an integration of clinical, genetic, neuropsychological, and neurobiological perspectives. *J Child Psychol Psychiatry* 1996, 37:89-126
- Bailey JA, Gu Z, Clark RA *et al*: Recent segmental duplications in the human genome. *Science* 2002; **297**: 1003-1007.
- Baird G, Charman T, Pickles A, Chandler S, Loucas T, Meldrum D, Carcani-Rathwell I, Serkana D, Simonoff E. Regression, developmental trajectory and associated problems in disorders in the autism spectrum: the SNAP study. *J Autism Dev Disord.* 2008; 38(10):1827-36.
- Ballarati L. *et al*. Cytogenetic, FISH and array-CGH characterization of a complex chromosomal rearrangement carried by a mentally and language impaired patient. *European Journal of Medical Genetics*, 2009; 52:218-223.
- Basel-Vanagaite L, Attia R, Yahav M, Ferland RJ, Anteki L, *et al*. 2006. The *CC2D1A*, a member of a new gene family with C2 domains, is involved in autosomal recessive non syndromic mental retardation. *J. Med. Genet.* 43:203–10
- Bamshad MJ, Ng SB, Bigham AW, Tabor HK, Emond MJ, Nickerson DA, Shendure J. Exome sequencing as a tool for Mendelian disease gene discovery. *Nat Rev Genet.* 2011 27;12(11):745-55.
- Bauters M, Van Esch H, Friez MJ, Boespflug-Tanguy O, Zenker M, *et al*. 2008. Non recurrent MECP2 duplications mediated by genomic architecture–driven DNA breaks and break-induced replication repair. *Genome Res.* 18:847–5810
- Bauters M, Weuts A, Vandewalle J, Nevelsteen J, Marynen P, *et al*. 2008. Detection and validation of copy number variation in X-linked mental retardation. *Cytogenet. Genome Res.* 123:44–53
- Bergmann C, Zerres K, Senderek J, Rudnik-Schoneborn S, Eggermann T, *et al*. 2003. *Oligophrenin 1 (OPHN1)* gene mutation causes syndromic X-linked mental

retardation with epilepsy, rostral ventricular enlargement and cerebellar hypoplasia. *Brain* 126:1537–4439.

Bettelheim B. (1967) *The Empty Fortress*. The Macmillan Company, New York.

Bienvenu T, Carrie A, de Roux N et al: MECP2 mutations account for most cases of typical forms of Rett syndrome. *Hum Mol Genet* 2000; 9: 1377– 1384.

Billuart P, Bienvenu T, Ronce N, des Portes V, Vinet MC, et al. 1998. Oligophrenin-1 encodes a rho GAP protein involved in X-linked mental retardation. *Nature* 392:923–2613

Bolton PF, Dennis NR, Browne CE, Thomas NS, Veltman MW, Thompson RJ et al. The phenotypic manifestations of interstitial duplications of proximal 15q with special reference to the autistic spectrum disorders. *Am J Med Genet* 2001; 105: 675–685.

Bolton PF, Park RJ, Higgins JN, Griffiths PD, Pickles A. Neuro-epileptic determinants of autism spectrum disorders in tuberous sclerosis complex. *Brain*. 2002 Jun;125(Pt 6):1247-55.

Buoni S, Sorrentino L, Farnetani MA, Pucci L, Fois A. The syndrome of inv dup (15): clinical, electroencephalographic, and imaging findings. *J Child Neurol*. 2000 Jun; 15(6):380-5.

Bonaglia MC, Giorda R, Mani E, Aceti G, Anderlid BM, Baroncini A, Pramparo T, Zuffardi O. Identification of a recurrent breakpoint within the SHANK3 gene in the 22q13.3 deletion syndrome. *J. Med. Genet*. 2006; 43: 822-828.

Caballero IM, Hendrich B: MeCP2 in neurons: closing in on the causes of Rett syndrome. *Hum Mol Genet* 2005; 14 (Spec No. 1):R19– R26.

Caliebe A, Kroes HY, van der Smagt JJ, Martin-Subero JI, Tönnies H, van 't Slot R, Nieveelstein RA, Muhle H, Stephani U, Alfke K, Stefanova I, Hellenbroich Y, Gillissen-Kaesbach G, Hochstenbach R, Siebert R, Poot M. (2010) Four patients with speech delay, seizures and variable corpus callosum thickness sharing a 0.440 Mb deletion in region 1q44 containing the HNRPU gene. *Eur J Med Genet.*;53(4):179-85.

Carlson, et al. Additional SNPs and linkage-disequilibrium analyses are necessary for wholegenome association studies in humans. *Nature Genet.*, 2003; 33: 518–520.

Chelly J, Khelifaoui M, Francis F, Chérif B, Bienvenu T. (2006) Genetics and pathophysiology of mental retardation. *Eur J Hum Genet*; 14(6):701-13.

Chen RZ, Akbarian S, Tudor M, Jaenisch R: Deficiency of methyl-CpG binding protein-2 in CNS neurons results in a Rett-like phenotype in mice. *Nat Genet* 2001; 27:327– 331.

Chih B, Afridi SK, Clark L, Scheiffele P. Disorder-associated mutations lead to functional inactivation of neuroligins. *Hum Mol Genet.* 2004 Jul 15;13(14):1471-7. Epub 2004 May 18.

Choi M, Scholl UI, Ji W, Liu T, Tikhonova IR, Zumbo P, Nayir A, Bakkaloğlu A, Ozen S, Sanjad S, Nelson-Williams C, Farhi A, Mane S, Lifton RP. Genetic diagnosis by whole exome capture and massively parallel DNA sequencing. *Proc Natl Acad Sci USA.* 2009; 106(45):19096-101..

Chubykin AA, Liu X, Comoletti D, Tsigelny I, Taylor P, Südhof TC. Dissection of synapse induction by neuroligins: effect of a neuroligin mutation associated with autism. *J Biol Chem.* 2005 Jun 10;280(23):22365-74. Epub 2005 Mar 29.

Colella S. et al., QuantiSNP: an Objective Bayes Hidden-Markov Model to detect and accurately map copy number variation using SNP genotyping data. *Nucleic Acids Research*, 2007, 35:2013-2025.

Conrad D. F., Andrews D., Carter N., Hurles M., Pritchard J., A high-resolution survey of deletion polymorphism in the human genome. *Nature Genetics*, 2006; 38: 75-81.

Conrad DF, Pinto D, Redon R *et al*: Origins and functional impact of copy number variation in the human genome. *Nature* 2009; Oct 7

Cook EH Jr, Scherer SW. Copy-number variations associated with neuropsychiatric conditions. *Nature*, 2008; 455:919-923.

Cook EH, *Jr. Genetics of autism.* Child Adolesc Psychiatr Clin N Am 2001, 10:333-50

Cook EH, Leventhal BL. The serotonin system in autism. *Curr Opin Pediatr.* 1996, 8:348-54

Costa RM, Yang T, Huynh DP et al: Learning deficits, but normal development and tumor predisposition, in mice lacking exon 23a of Nf1. *Nat Genet* 2001; 27: 399– 405.

Courchesne E, Karns CM, Davis HR, Ziccardi R, Carper RA, Tigue ZD, Chisum HJ, Moses P, Pierce K, Lord C, Lincoln AJ, Pizzo S, Schreibman L, Haas RH, Akshoomoff NA, Courchesne RY. Unusual brain growth patterns in early life in patients with autistic disorder: an MRI study. *Neurology* 2001, 57:245-54.

Courchesne E. Abnormal early brain development in autism. *Mol Psychiatry* 2002, 7 Suppl 2:S21-3.

Couvert P, Bienvenu T, Aquaviva C et al: MECP2 is highly mutated in X-linked mental retardation. *Hum Mol Genet* 2001; 10:941– 946.

D'Adamo MC, Imbrici P, Sponcichetti F, Pessia M. 1999. Mutations in the *KCNA1* gene associated with episodic ataxia type-1 syndrome impair heteromeric voltage-gated K⁺ channel function. *FASEB J.* 13:1335–4581

D'Adamo P, Welzl H, Papadimitriou S et al: Deletion of the mental retardation gene *Gdi1* impairs associative memory and alters social behavior in mice. *Hum Mol Genet* 2002; 11: 2567– 2580

Dean C, Scholl FG, Choih J et al: Neurexin mediates the assembly of presynaptic terminals. *Nat Neurosci* 2003; 6: 708– 716.

Dean C, Dresbach T. Neuroligins and neurexins: linking cell adhesion, synapse formation and cognitive function. *Trends Neurosci.* 2006 Jan;29(1):21-9. Epub 2005

De Cesare D, Jacquot S, Hanauer A, Sassone-Corsi P: Rsk-2 activity is necessary for epidermal growth factor-induced phosphorylation of CREB protein and transcription of c-fos gene. *Proc Natl Acad Sci USA* 1998; 95: 12202 –12207

del Arco A, Satrustegui J (Oct 1998). "Molecular cloning of Aralar, a new member of the mitochondrial carrier superfamily that binds calcium and is present in human muscle and brain". *J Biol Chem* **273** (36): 23327–34.

de Kovel CG, Trucks H, Helbig I, Mefford HC, Baker C, et al. 2010. Recurrent microdeletions at 15q11.2 and 16p13.11 predispose to idiopathic generalized epilepsies. *Brain* 133:23–3269

de Vries BB, Pfundt R, Leisink M, Koolen DA, Vissers LE, Janssen IM, Reijmersdal S, Nillesen WM, Huys EH, Leeuw N, Smeets D, Sistermans EA, Feuth T, van Ravenswaaij-Arts CM, van Kessel AG, Schoenmakers EF, Brunner HG, Veltman JA. (2005) Diagnostic genome profiling in mental retardation. *Am J Hum Genet*; 77(4):606-16.

de Vrij FM, Levenga J, van der Linde HC, Koekkoek SK, De Zeeuw CI, et al. 2008. Rescue of behavioral phenotype and neuronal protrusion morphology in *Fmr1* KO mice. *Neurobiol. Dis.* 31:127–32

Dykens E. and Volkmar F. R., Medical conditions associated with autism, in *Handbook of Autism and Pervasive Developmental Disorders*, 1997; (Volkmar C., ed.), Wiley, New York, pp. 388–410.

Donovan et al. Identification of loci associated with schizophrenia by genomewide association and follow-up. *Nature Genetics*, 2008-09; 9:1053-1055.

Durand CM, Betancur C, Boeckers TM, Bockmann J, Chaste P, Fauchereau F, Nygren G, Rastam M, Gillberg IC, Anckarsater H, Sponheim E, Goubran-Botros H, Delorme R, Chabane N, Mouren-Simeoni MC, de Mas P, Bieth E, Roge B, Heron D, Burglen L, Gillberg C, Leboyer M, Bourgeron T. Mutations in the gene encoding the synaptic scaffolding protein SHANK3 are associated with autism spectrum disorders. *Nature Genet.* 2007; 39: 25-27.

Feinstein C. and Reiss A. L., Autism: the point of view from fragile X studies. *J. Autism Dev. Disord.* 1998; 28: 393–405.

- Folstein S, Rutter M. Infantile autism: a genetic study of 21 twin pairs. *J Child Psychol Psychiatry*, 1977; 18:297–321.
- Folstein SE, Rosen-Sheidley B. Genetics of autism: complex aetiology for a heterogeneous disorder. *Nat Rev Genet*, 2001; 2:943–955.
- Fombonne E, Modern views of autism. *Can J Psychiatry*. 2003; 48: 503-5.
- Fombonne E, Epidemiology of Autistic Disorder and Other Pervasive Developmental Disorders. *J. Clin. Psychiatry*, 2005; 66:3–8.
- Fridlyand J, Snijders AM, Ylstra B *et al*: Breast tumor copy number aberration phenotypes and genomic instability. *BMC Cancer* 2006; 6: 96.
- Gabriel SB, Schaffner SF, Nguyen H, Moore JM, Roy J, Blumenstiel B, Higgins J, DeFelice M, Lochner A, Faggart M, Liu-Cordero SN, Rotimi C, Adeyemo A, Cooper R, Ward R, Lander ES, Daly MJ, Altshuler D. The structure of haplotype blocks in the human genome. *Science*, 2002; 296:2225-2229.
- Garshasbi M, Kahrizi K, Tzschach A, Behjati F, Falah M, et al. 2009. *ZC3H14* mutations cosegregate with nonsyndromic autosomal recessive mental retardation (NS-ARMR) in two Iranian families. In *14th International Workshop on Fragile X and X-Linked Mental Retardation*, Abstr. no. 62, 70 pp. Bahia, Brazil
- Garshasbi M, Kahrizi K, Tzschach A, Behjati F, Falah M, et al. 2009. *ZC3H14* mutations cosegregate with non-syndromic autosomal recessive mental retardation (NS-ARMR) in two Iranian families. Paper presented at 59th Annual Meeting of Am. Soc. Hum. Genet., Honolulu, Hawaii
- Geppert M, Goda Y, Stevens CF, Sudhof TC: The small GTP binding protein Rab3A regulates a late step in synaptic vesicle fusion. *Nature* 1997; 387: 810– 814
- Geschwind DH, Sowinski J, Lord C, Iversen P, Shestack J, Jones P, Ducat L, Spence SJ, AGRE Steering Committee. The autism genetic resource exchange: a resource for the study of autism and related neuropsychiatric conditions. *Am J Hum Genet*, 2001; 69:463-466.
- Gillis RF, Rouleau GA. The ongoing dissection of the genetic architecture of autistic spectrum disorder. *Mol Autism*. 2011;2(1):12.
- Gilman SR, Iossifov I, Levy D, Ronemus M, Wigler M, Vitkup D. Rare de novo variants associated with autism implicate a large functional network of genes involved in formation and function of synapses. *Neuron*. 2011;70(5):898-907.
- Giorda R, Bonaglia MC, Beri S, Fichera M, Novara F, Magini P, Urquhart J, Sharkey FH, Zucca C, Grasso R, Marelli S, Castiglia L, Di Benedetto D, Musumeci SA, Vitello GA, Failla P, Reitano S, Avola E, Bisulli F, Tinuper P, Mastrangelo M, Focchi I, Spaccini L, Torniero C, Fontana E, Lynch SA, Clayton-Smith J, Black G, Jonveaux P, Leheup B, Seri M, Romano C, dalla Bernardina B, Zuffardi O. Complex segmental duplications mediate a recurrent dup(X)(p11.22-p11.23) associated with

mental retardation, speech delay, and EEG anomalies in males and females.. *Am J Hum Genet.* 2009 ;85(3):394-400.

Girirajan S. *et al.* A recurrent 16p12.1 microdeletion supports a two-hit model for severe developmental delay. *Nat Genet* 2010; 42 (3): 203-209.

Glessner JT, Wang K, Cai G, Korvatska O, Kim CE, Wood S, Zhang H, Estes A, Brune CW, Bradfield JP, Imielinski M, Frackelton EC, Reichert J, Crawford EL, Munson J, Sleiman PM, *Nature*, 2009; 459:569-73.

Goldstein, Ahmadi, Weale, Wood, Genome scans and candidate gene approaches in the study of common diseases and variable drug responses. *Trends Genet.*, 2003; 19: 615–622.

Govek EE, Newey SE, Akerman CJ, Cross JR, Van der Veken L, VanAelst L: The X-linked mental retardation protein oligophrenin-1 is required for dendritic spine morphogenesis. *Nat Neurosci* 2004;7: 364– 372.

Graf ER, Zhang X, Jin SX, Linhoff MW, Craig AM. Neurexins induce differentiation of GABA and glutamate postsynaptic specializations via neuroligins. *Cell.* 2004 Dec 29;119(7):1013-26.

Gupton SL, Gertler FB. Filopodia: the fingers that do the walking. *Sci STKE.* 2007 (400):re5.

Gutierrez GC, Smalley SL, Tanguay PE. Autism in tuberous sclerosis complex. *J Autism Dev Disord.*1998; 28: 97-103.

Guy J, Hendrich B, Holmes M, Martin JE, Bird A: A mouse *Mecp2*-null mutation causes neurological symptoms that mimic Rett syndrome. *Nat Genet* 2001; 27: 322– 326.

Hamdan FF, Piton A, Gauthier J, Lortie A, Dubeau F, et al. 2009. De novo *STXBPI* mutations in mental retardation and nonsyndromic epilepsy. *Ann. Neurol.* 65:748–53

Hamdan FF, Gauthier J, Spiegelman D, Noreau A, Yang Y, et al. 2009. Mutations in *SYNGAP1* in autosomal nonsyndromic mental retardation. *N. Engl. J. Med.* 360:599–605

Hannes FD, Sharp AJ, Mefford HC, de Ravel T, Ruivenkamp CA, et al. 2009. Recurrent reciprocal deletions and duplications of 16p13.11: the deletion is a risk factor for MR/MCA while the duplication may be a rare benign variant. *J. Med. Genet.* 46:223–32124.

Harum KH, Alemi L, Johnston MV: Cognitive impairment in Coffin –Lowry syndrome correlates with reduced RSK2 activation. *Neurology* 2001; 56: 207– 214

Hastings PJ, Ira G, Lupski JR: A microhomology-mediated break-induced replication model for the origin of human copy number variation. *PLoS Genet* 2009; 5: e1000327.

Hedges DJ, Burges D, Powell E, Almonte C, Huang J, Young S, Boese B, Schmidt M, Pericak-Vance MA, Martin E, Zhang X, Harkins TT, Züchner S. Exome sequencing of a multigenerational human pedigree. *PLoS One*. 2009; 4(12):e8232.

Helbig I. et al. 15q13.3 microdeletions increase risk of idiopathic generalized epilepsy. *Nat Genet*, 2009; 2:160-162.

Higgins JJ, Pucilowska J, Lombardi RQ, Rooney JP. 2004. A mutation in a novel ATP-dependent Lonprotease gene in a kindred with mild mental retardation. *Neurology* 63:1927–31110

Hinds D.A., Kloek A. P., Jen M., Chen X., Frazer K.A., Common deletions and SNPs are in linkage disequilibrium in the human genome. *Nature Genetics*, February 2006; vol. 38, num. 1: 82-85.

Hotulainen P, Hoogenraad CC. Actin in dendritic spines: connecting dynamics to function. *J Cell Biol*. 2010;189(4):619-29.

Kallman FJ: The genetics of schizophrenia. New York: J. J. Augustin, 193863 Darai-Ramqvist E, Sandlund A, Müller S, Klein G, Imreh S, Kost-Alimova M: Segmental duplications and evolutionary plasticity at tumor chromosome break-prone regions. *Genome Res* 2008; 18: 370-379

Kandel ER: The molecular biology of memory storage: a dialogue between genes and synapses. *Science* 2001; 294: 1030– 1038

Kanner L, Autistic disturbance of affective contact. *Nervous Child*, 1943; 2: 217–250.

Kanner L (1955) Eisenberg, Early infantile Autism 1943-1955, in *American Journal of Orthopsychiatry*.

Kemper T.L. and Bauman P., Neuropathology of infantile autism. *J. Neuropathol. Exp. Neurol.* 1998; 57: 645-662.

Kennedy G. C., et al., Large-scale genotyping of complex DNA. *Nature Biotechnol.* 2003; 21: 1233–1237.

Klauck SM, Genetics of autism spectrum disorder, *Europ. Journ. of Hum. Genet.* 2006; 14: 714-720.

Koekkoek SK, Yamaguchi K, Milojkovic BA et al: Deletion of FMR1 in Purkinje cells enhances parallel fiber LTD, enlarges spines, and attenuates cerebellar eyelid conditioning in fragile X syndrome. *Neuron* 2005; 47: 339– 352.

Koolen DA, Pfundt R, de Leeuw N, Hehir-Kwa JY, Nillesen WM, Neefs I, Scheltinga I, Sijm A, Smeets D, Brunner HG, van Kessel AG, Veltman JA, de Vries BB. (2009) Genomic microarrays in mental retardation: a practical workflow for diagnostic applications. *Hum Mutat*; 30(3):283-92

- Kriaucionis S, Bird A: DNA methylation and Rett syndrome. *Hum Mol Genet* 2003; 12 (Spec No. 2): R221– R227.75
- Kruglyak, Daly, Reeve-Daly, Lander, Parametric and nonparametric linkage analysis: a unified multipoint approach. *Am J Hum Genet*, 1996; 58:1347– 1363.
- KutscheK, YntemaH, Brandt A, Jantke I, Nothwang HG, et al. 2000. Mutations in *ARHGEF6*, encoding a guanine nucleotide exchange factor for Rho GTPases, in patients with X-linked mental retardation. *Nat. Genet.* 26:247–50114
- Jamain S, Quach H, Betancur C, Rastam M, Colineaux C, Gillberg IC, Soderstrom H, Giros B, Leboyer M, Gillberg C, Bourgeron T; Paris Autism Research International Sibpair Study. Mutations of the X-linked genes encoding neuroligins NLGN3 and NLGN4 are associated with autism. *Nat Genet.* 2003; 34: 27-9.
- Jambaque I, Chiron C, Dumas C, Mumford J, Dulac O. Mental and behavioural outcome of infantile epilepsy treated by vigabatrin in tuberous sclerosis patients. *Epilepsy Res.* 2000; 38: 151-60.
- Jiang S, Seng S, Avraham HK, Fu Y, Avraham S. Process elongation of oligodendrocytes is promoted by the Kelch-related protein MRP2/KLHL1. *J Biol Chem.* 2007;282(16):12319-29.
- Iafate AJ, Feuk L, Rivera MN, Listewnik ML, Donahoe PK, Qi Y, Scherer SW, Lee C. Detection of large-scale variation in the human genome. *Nat Genet*, 2004; 36:949–951.
- IMGSAC, A full genome screen for autism with evidence for linkage to a region on chromosome 7q. *Hum Molec Genet*, 1998; 7:571–578.
- IMGSAC, A genomewide screen for autism: strong evidence linkage to chromosomes 2q, 7q and 16p. *Am J Hum Genet*, 2001; 69:570–581.
- Ingason A, Rujescu D, Cichon S, Sigurdsson E, Sigmundsson T, et al. 2009. Copy number variations of chromosome 16p13.1 region associated with schizophrenia. *Mol. Psychiatry* Sep. 29 doi:10.1038/mp.2009.101
- Inlow JK, Restifo LL. (2004) Molecular and comparative genetics of mental retardation. *Genetics*; 166(2):835-81.
- Inoue K, Lupski JR. Molecular mechanisms for genomic disorders. *Annu Rev Genomics Hum Genet.* 2002; 3:199-242. *J Autism Dev Disord.* 1989; 19: 349-51
- Ishizaki H, Miyoshi J, Kamiya H et al: Role of rab GDP dissociation inhibitor alpha in regulating plasticity of hippocampal neurotransmission. *Proc Natl Acad Sci USA* 2000; 97: 11587– 11592.
- L.A. Weiss, et al., Association between microdeletion and microduplication at 16p11.2 and autism, *New Engl. J. Med.*, 2008; 358:667-675.

Lamb JA, Parr JR, Bailey AJ, Monaco AP. Autism: in search of susceptibility genes. *Neuromolecular Med*, 2002; 2:11–28.

Lamb JA, Barnby G, Bonora E, Sykes N, Bacchelli E, Blasi F et al. Analysis of IMGSAC autism susceptibility loci: evidence for sex limited and parent of origin specific effects. *J Med Genet* 2005; 42: 132–137.

Lander ES, Schork NJ, Genetic dissection of complex traits. *Science*, 1994; 265: 2037-2048

Laumonier F, Bonnet-Brilhault F, Gomot M, Blanc R, David A, Moizard MP, Raynaud M, Ronce N, Lemonnier E, Calvas P, Laudier B, Chelly J, Fryns JP, Ropers HH, Hamel BC, Andres C, Barthelemy C, Moraine C, Briault S. X-linked mental retardation and autism are associated with a mutation in the NLGN4 gene, a member of the neuroligin family. *Am J Hum Genet*. 2004; 74: 552-7.

M. Lauritsen, H. Ewald The genetics of autism, j.1600-0447.2001.00086 *Acta Psychiatrica Scandinavica* Volume 103, Issue 6, pages 411–427, June 2001

Lee JA, Carvalho CM, Lupski JR: A DNA replication mechanism for generating non recurrent rearrangements associated with genomic disorders. *Cell* 2007; 131: 1235-124.

Levy D, Ronemus M, Yamrom B, Lee YH, Leotta A, Kendall J, Marks S, Lakshmi B, Pai D, Ye K, Buja A, Krieger A, Yoon S, Troge J, Rodgers L, Iossifov I, Wigler M. Rare de novo and transmitted copy-number variation in autistic spectrum disorders. *Neuron*. 2011;70(5):886-97.

Lord et al, The autism diagnostic observation schedule generic: a standard measure of social and communication deficits associated with the spectrum of autism. *J Autism Dev. Disord*. 2000; 30: 205-223.

Lord, Rutter, LeCouteur, Autism Diagnostic Interview-Revised: a revised version of a diagnostic interview for caregivers of individuals with possible pervasive developmental disorders. *J. Autism Dev. Disord*, 1994; 24:659-685

Lugtenberg D, Kleefstra T, Oudakker AR, Nillesen WM, Yntema HG, et al. 2009. Structural variation in Xq28: *MECP2* duplications in 1% of patients with unexplained XLMR and in 2% of male patients with severe encephalopathy. *Eur. J. Hum. Genet*. 17:444–5318.

Lupski JR. Hotspots of homologous recombination in the human genome: not all homologous sequences are equal. *Genome Biol*. 2004; 5(10):242.

Lupski JR, Stankiewicz P, Genomic Disorders: Molecular Mechanisms for Rearrangements and Conveyed Phenotypes. *Plos Genetics* , 2005; 1, Issue 6 e 49.

Lupski JR, Stankiewicz P, Genome architecture, rearrangements and genomic disorders. *Trends Genet.*, 2002;18:74–82.

Lupski JR, Shaw CJ, Implications of human genome architecture for rearrangement-based disorders: The genomic basis of disease. *Hum. Mol. Genet.*, 2004;13: 57–64.

- Maestrini E, Lai C, Marlow A, Matthews N, Wallace S, Bailey A, Cook EH, Weeks DE, Monaco AP, and IMGSAC. Serotonin transporter (5-HTT) and α -aminobutyric acid receptor subunit-3 (GABRB3) gene polymorphisms are not associated with autism in the IMGSAC families. *Am. J. Med. Genet.*, 1999;88:492-496.
- Maestrini E, Pagnamenta AT, Lamb JA, Bacchelli E, Sykes NH, Sousa I, Toma C, Barnby G, Butler H, Winchester L, Scerri TS, Minopoli F, Reichert J, Cai G, Buxbaum JD, Korvatska O, Schellenberg GD, Dawson G, de Bildt A, Minderaa RB, Mulder EJ, Morris AP, Bailey AJ, Monaco AP; IMGSAC. High-density SNP association study and copy number variation analysis of the AUTS1 and AUTS5 loci implicate the IMMP2L-DOCK4 gene region in autism susceptibility. *Mol Psychiatry*. 2010 ;15(9):954-68
- Manolio TA, Collins FS, Cox NJ, Goldstein DB, Hindorff LA, Hunter DJ, McCarthy MI, Ramos EM, Cardon LR, Chakravarti A, Cho JH, Guttmacher AE, Kong A, Kruglyak L, Mardis E, Rotimi CN, Slatkin M, Valle D, Whittemore AS, Boehnke M, Clark AG, Eichler EE, Gibson G, Haines JL, Mackay TF, McCarroll SA, Visscher PM. Finding the missing heritability of complex diseases. *Nature*. 2009 Oct 8;461(7265):747-53.
- Mari F, Azimonti S, Bertani I et al: CDKL5 belongs to the same molecular pathway of *MeCP2* and it is responsible for the early on set seizure variant of Rett syndrome. *Hum Mol Genet* 2005; 14:1935– 1946.
- Maria BL, Deidrick KM, Roach ES, Gutmann DH. Tuberous sclerosis complex: pathogenesis, diagnosis, strategies, therapies, and future research directions. *J Child Neurol*. 2004;19:632-42.
- Marshall CR, Noor A, Vincent JB, Lionel AC, Feuk L, Skaug J, Shago M, Moessner R, Pinto D, Ren Y, Thiruvahindrapduram B, Fiebig A, Schreiber S, Friedman J, Ketelaars CE, Vos YJ, Ficicioglu C, Kirkpatrick S, Nicolson R, Sloman L, Summers A, Gibbons CA, Teebi A, Chitayat D, Weksberg R, Thompson A, Vardy C, Crosbie V, Luscombe S, Baatjes R, Zwaigenbaum L, Roberts W, Fernandez B, Szatmari P, Scherer SW. Structural variation of chromosomes in autism spectrum disorder. *Am J Hum Genet*. 2008;82(2):477-88.
- McBride SM, Choi CH, Wang Y, Liebelt D, Braunstein E, et al. 2005. Pharmacological rescue of synaptic plasticity, courtship behavior, and mushroom body defects in a *Drosophila* model of fragile X syndrome. *Neuron* 45:753–6438.
- McCarroll S et al., Common deletion polymorphisms in the human genome. *Nature Genetics*, February 2006; vol. 38, num. 1: 86- 92.
- Mefford HC, Cooper GM, Zerr T, Smith JD, Baker C, et al. 2009. A method for rapid, targeted CNV genotyping identifies rare variants associated with neurocognitive disease. *Genome Res*. 19:1579–85165.
- Mefford HC, Batshaw ML, Hoffman EP. Genomics, intellectual disability, and autism. *N Engl J Med*. 2012 Feb 23;366(8):733-43.
- Meloni I, Bruttini M, Longo I et al: A mutation in the rett syndrome gene, *MECP2*, causes X-linked mental retardation and progressive spasticity in males. *Am J Hum Genet* 2000; 67:982– 985

Meng J, Meng Y, Hanna A, Janus C, Jia Z: Abnormal long-lasting synaptic plasticity and cognition in mice lacking the mental retardation gene Pak3. *J Neurosci* 2005; 25: 6641–6650

Menten B, Maas N, Thienpont B, Buysse K, Vandesompele J, Melotte C, de Ravel T, Van Vooren S, Balikova I, Backx L, Janssens S, De Paepe A, De Moor B, Moreau Y, Marynen P, Fryns JP, Mortier G, Devriendt K, Speleman F, Vermeesch JR. (2006) Emerging patterns of cryptic chromosomal imbalance in patients with idiopathic mental retardation and multiple congenital anomalies: a new series of 140 patients and review of published reports. *J Med Genet*; 43(8):625-33.

Miles JH, Hillman RE. Value of a clinical morphology examination in autism. *Am J Med Genet*. 2000; 91: 245-53.

Miller D.T. et al. Microdeletion duplication at 15q13.2q13.3 among individuals with features of autism and other neuropsychiatric disorders. *J Med Genet*, 2008.

Mir A, Kaufman L, Noor A, Motazacker MM, Jamil T, et al. 2009. Identification of mutations in TRAPPC9, which encodes the NIK- and IKK-beta-binding protein, in non syndromic autosomal recessive mental retardation. *Am. J. Hum. Genet.* 85:909–15104.

Mochida GH, Mahajnah M, Hill AD, Basel-Vanagaite L, Gleason D, et al. 2009. A truncating mutation of TRAPPC9 is associated with autosomal-recessive intellectual disability and postnatal microcephaly. *Am. J. Hum. Genet.* 85:897–902105

Molinari F, Rio M, Meskenaite V, Encha-Razavi F, Auge J, et al. 2002. Truncating neurotrypsin mutation in autosomal recessive non syndromic mental retardation. *Science* 298:1779–817.

Molinari F, Foulquier F, Tarpey PS, Morelle W, Boissel S, et al. 2008. Oligosaccharyl transferase-subunit mutations in non syndromic mental retardation. *Am. J. Hum. Genet.* 82:1150–57106.

Molinari F, Rio M, Meskenaite V, Encha-Razavi F, Auge J, et al. 2002. Truncating neurotrypsin mutation in autosomal recessive non syndromic mental retardation. *Science* 298:1779–81107.

Morton. Sequential tests for the detection of linkage. *Am J Hum Genet*, 1955;7:277–318.

Motazacker MM, Rost BR, Hucho T, Garshasbi M, Kahrizi K, et al. 2007. A defect in the *ionotropic glutamate receptor 6* gene (*GRIK2*) is associated with autosomal recessive mental retardation. *Am. J. Hum. Genet.* 81:792–981

Nakamoto M, Nalavadi V, Epstein MP, Narayanan U, Bassell GJ, Warren ST. 2007. Fragile X mental retardation protein deficiency leads to excessive mGluR5-dependent internalization of AMPA receptors. *Proc. Natl. Acad. Sci. USA* 104:15537–42112.
Najmabadi H, Hu H, Garshasbi M, Zemojtel T, Abedini SS, Chen W, Hosseini M, Behjati F, Haas S, Jamali P, Zecha A, Mohseni M, Püttmann L, Vahid LN, Jensen C,

Moheb LA, Bienek M, Larti F, Mueller I, Weissmann R, Darvish H, Wrogemann K, Hadavi V, Lipkowitz B, Esmaeeli-Nieh S, Wieczorek D, Kariminejad R, Firouzabadi SG, Cohen M, Fattahi Z, Rost I, Mojahedi F, Hertzberg C, Dehghan A, Rajab A, Banavandi MJ, Hoffer J, Falah M, Musante L, Kalscheuer V, Ullmann R, Kuss AW, Tzschach A, Kahrizi K, Ropers HH. Deep sequencing reveals 50 novel genes for recessive cognitive disorders. *Nature*. 2011 Sep 21;478(7367):57-63.

Narayanan U, Nalavadi V, Nakamoto M, Pallas DC, Ceman S, et al. 2007. FMRP phosphorylation reveals an immediate-early signaling pathway triggered by group I mGluR and mediated by PP2A. *J.Neurosci*. 27:14349–5746

Najmabadi H, Motazacker MM, Garshasbi M, Kahrizi K, Tzschach A, et al. 2007. Homozygosity mapping in consanguineous families reveals extreme heterogeneity of non syndromic autosomal recessive mental retardation and identifies 8 novel gene loci. *Hum. Genet*. 121:43–48103.

Nelson B. 2009. Mental retardation and intellectual disability. In *Vogul and Motulsky's Human Genetics*, ed. MR Speicher, SE Antonarakis, AG Motulsky. New York: Springer. 4th ed.

Ng SB, Buckingham KJ, Lee C, Bigham AW, Tabor HK, Dent KM, Huff CD, Shannon PT, Jabs EW, Nickerson DA, Shendure J, Bamshad MJ. (2010) Exome sequencing identifies the cause of a mendelian disorder. *Nat Genet*; 42(1):30-5.

Nuber UA, Kriaucionis S, Roloff TC et al: Up-regulation of glucocorticoid-regulated genes in a mouse model of Rett syndrome. *Hum Mol Genet* 2005; 14: 2247– 2256.

Oberle I, Rousseau F, HeitzD, Kretz C, DevysD, et al. 1991. Instability of a 550-base pair DNA segment and abnormal methylation in fragile X syndrome. *Science* 252:1097–102

O'Connell JR, Weeks DE. (1998) PedCheck: a program for identification of genotype incompatibilities in linkage analysis. *Am J Hum Genet*; 63(1):259-66.

Oostra BA, Chiurazzi P. The fragile X gene and its function. *Clin Genet*. 2001 Dec;60(6):399-408.

O'Roak BJ, Deriziotis P, Lee C, Vives L, Schwartz JJ, Girirajan S, Karakoc E, Mackenzie AP, Ng SB, Baker C, Rieder MJ, Nickerson DA, Bernier R, Fisher SE, Shendure J, Eichler EE. Exome sequencing in sporadic autism spectrum disorders identifies severe de novo mutations. *Nat Genet*. 2011 Jun;43(6):585-9.

Pagnamenta, A. et al. A 15q13.3 microdeletion segregating with autism. *Eur J Hum Genet*, 2008.

Pannekoek WJ, Kooistra MR, Zwartkruis FJ, Bos JL. Cell-cell junction formation: the role of Rap1 and Rap1 guanine nucleotide exchange factors. *Biochim Biophys Acta*. 2009;1788(4):790-6

Patil N, Berno AJ, Hinds DA, Barrett WA, Doshi JM, Hacker CR, Kautzer CR, Lee DH, Marjoribanks C, McDonough DP, Nguyen BT, Norris MC, Sheehan JB, Shen N, Stern D, Stokowski RP, Thomas DJ, Trulson MO, Vyas KR, Frazer KA, Fodor SP. Blocks of limited haplotype diversity revealed by high-resolution scanning of human chromosome 21. *Science*, 2001;294:1719–1723.

- Pedersen OS, Liu Y, Reichelt KL. Serotonin uptake stimulating peptide found in plasma of normal individuals and in some autistic urines. *J Pept Res* 1999, 53:641-6.
- Penrose, The general purpose sib-pair linkage test. *Ann Eugen*, 1953; 18:120–124.
- Petrij F, Giles RH, Dauwerse HG et al: Rubinstein-Taybi syndrome caused by mutations in the transcriptional co-activator CBP. *Nature* 1995; 376: 348–351
- Pinkel D. Et al., High resolution analysis of DNA copy number variation using comparative genomic hybridization to microarrays. *Nature Genet*, 1998; 20: 207–211.
- Pinto D, Pagnamenta A, Klei L, Anney R, Merico D, Regan R, Conroy J. et al. Functional impact of copy number variation in autism. *Nature* 2010 [Epub ahead of print].
- des Portes V, Boddaert N, Sacco S, Briault S, Maincent K, et al. 2004. Specific clinical and brain MRI features in mentally retarded patients with mutations in the *Oligophrenin-1* gene. *Am. J. Med. Genet. A* 124A:364–719.
- Piven J, Tsai GC, Nehme E, Coyle JT, Chase GA, Folstein SE. Platelet serotonin, a possible marker for familial autism. *J. Autism Dev. Disord.* 1991; 21: 51-59.
- Ponchel, F. et al. Real-time PCR based on SYBRGreen I fluorescence: An alternative to the TaqMan assay for a relative quantification of gene rearrangements, gene amplifications and micro gene deletions. *BMC Biotechnol* 2003; 3: 18.
- Poot M, Eleveld MJ, van 't Slot R, Ploos van Amstel HK, Hochstenbach R. 2010. Recurrent copy number changes in mentally retarded children harbor genes involved in cellular localization and the glutamate receptor complex. *Eur. J. Hum. Genet.* 18:39–4636.
- Rabionet R, McCauley JL, Jaworski JM, Ashley-Koch AE, Martin ER, Sutcliffe JS, Haines JL, DeLong GR, Abramson RK, Wright HH, Cuccaro ML, Gilbert JR, Pericak-Vance MA. Lack of association between autism and SLC25A12. *Am J Psychiatry.* 2006 May;163(5):929-31.
- Ramoz N, Reichert JG, Smith CJ, Silverman JM, Beshpalova IN, Davis KL, Buxbaum JD. Linkage and association of the mitochondrial aspartate/glutamate carrier SLC25A12 gene with autism *Am J Psychiatry.* 2004 Apr;161(4):662-9.
- Raymond FL, Tarpey P. (2006) The genetics of mental retardation. *Hum Mol Genet*; 15 Spec No 2:R110-6.
- Reddy KS. Cytogenetic abnormalities and fragile-X syndrome in Autism Spectrum Disorder. *BMC Med Genet.* 2005; 6: 3.
- Redon R, Ishikawa S, Fitch KR et al. Global variation in copy number in the human genome. *Nature*, 2006;444:444-454.

Reich, et al. Linkage disequilibrium in the human genome. *Nature*, 2001;411: 199–204.

Reichenberg A, Weiser M, Rabinowitz J *et al*: A population-based cohort study of pre morbid intellectual, language, and behavioral functioning in patients with schizophrenia, schizoaffective disorder, and non psychotic bipolar disorder. *Am J Psychiatry* 2002

Repetto GM, White LM, Bader PJ, Johnson D, Knoll JH. Interstitial duplications of chromosome region 15q11q13: clinical and molecular characterization *Am J Med Genet*. 1998 Sep 1;79(2):82-9.

Rineer S, Finucane B, Simon EW. Autistic symptoms among children and young adults with isodicentric chromosome 15. *Am J Med Genet*. 1998 Sep 7;81(5):428-33.

Risch N, Spiker D, Lotspeich L, Nouri N, Hinds D, Hallmayer J, Kalaydjieva L, et al. A genomic screen of autism: evidence for a multilocus etiology. *Am J Hum Genet* 1999, 65:493-507.

Ropers HH, Hamel BC. (2005) X-linked mental retardation. *Nat Rev Genet*; 6(1):46-57. Ropers HH. (2008) Genetics of intellectual disability. *Curr Opin Genet Dev*; 18(3):241-50.

Rutter M. Genetic Studies of Autism: From the 1970s into the Millennium. *Journal of Abnormal Child Psychology*. 2000; 28: 3-14.

Sanders SJ, et al. Multiple recurrent de novo CNVs, including duplications of the 7q11.23 Williams syndrome region, are strongly associated with autism. *Neuron*. 2011;70(5):863-85

Scheiffele P, Fan J, Choih J, Fetter R, Serafini T: Neuroligin expressed in non neuronal cells triggers presynaptic development in contacting axons. *Cell* 2000; 101: 657– 669

Schoumans J, Ruivenkamp C, Holmberg E, Kyllerman M, Anderlid BM, Nordenskjöld M. (2005) Detection of chromosomal imbalances in children with idiopathic mental retardation by array based comparative genomic hybridisation (array-CGH). *J Med Genet*; 42(9):699-705.

Sebat J. et al., Large-scale copy number polymorphism in the human genome. *Science*, 2004; 305:525–528

Sebat J. et al., Strong association of de novo copy number mutations with autism. *Science*, 2007; 316:445-449.

Segurado R, Hawi Z, Conroy J, Sheehan K, Lowe N, Kirley A, Shields D, Fitzgerald M, Gallagher L, Gill M. Preferential transmission of paternal alleles at risk genes in attention-deficit/hyperactivity disorder. *Am J Hum Genet*. 2005 Dec;77(6):958-65. Epub 2005 Oct 26.

- Seng S, Avraham HK, Jiang S, Venkatesh S, Avraham S., KLHL1/MRP2 mediates neurite outgrowth in a glycogen synthase kinase 3beta-dependent manner. *Mol Cell Biol.* 2006; 26(22):8371-84.
- Schaaf CP, Zoghbi HY. Solving the autism puzzle a few pieces at a time. *Neuron.* 2011 Jun 9; 70(5):806-8.
- Sharp, A. J. et al. A recurrent 15q13.3 microdeletion syndrome associated with mental retardation and seizures. *Nat Genet,* 2008; 40:322-328
- Sharp AJ, et al. Segmental duplications and copy number variation in the human genome. *Am. J. Hum. Genet.* 2005; 77: 78–88.
- Shinawi M, Cheung SW. (2008) The array CGH and its clinical applications. *Drug Discov Today;* 13(17-18):760-70. Epub 2008 Jul 17.
- Steffenburg S, Gillberg C, Hellgren L, Andersson L, Gillberg I, Jakobsson G et al. A twin study of autism in Denmark, Finland, Iceland, Norway, and Sweden. *J Child Psychol Psychiatry.* 1989; 30:405–416.
- Sykes NH, Lamb JA. Autism: the quest for the genes. *Expert Rev Mol Med.* 2007;9(24):1-15.
- Slager, Schaid, Case-control studies of genetic markers: power and sample size approximations for Armitage's test for trend. *Hum Hered* , 2001; 52: 149– 53.
- Slavotinek AM. (2008) Novel microdeletion syndromes detected by chromosome microarrays. *Hum Genet;* 124(1):1-17.
- Smalley SL, Autism and tuberous sclerosis. *J. Autism Dev. Disord.* 1998; 28: 407–414.
- Shoichet SA, Kunde SA, Viertel P, Schell-Apacik C, von VossH, et al. 2005. Haploinsufficiency of novel FOXG1B variants in a patient with severe mental retardation, brain malformations and microcephaly. *Hum. Genet.* 117:536–44
- Solinas-Toldo, Lampel S, Stilgenbauer S, Nickolenko J, Benner A, Döhner H, Cremer T, Lichter P., Matrix-based comparative genomic hybridization: biochips to screen for genomic imbalances. *Genes Chromosomes Cancer* 1997; 20: 399–407.
- Speicher MR., Monitoring chromosome rearrangements. *Adv Exp Med Biol.* 2005;570:19-41.
- Stewart DR, Huang A, Faravelli F, Anderlid BM, Medne L, Ciprero K, Kaur M, Rossi E, Tenconi R, Nordenskjöld M, Gripp KW, Nicholson L, Meschino WS, Capua E, Quarrell OW, Flint J, Irons M, Giampietro PF, Schowalter DB, Zaleski CA, Malacarne M, Zackai EH, Spinner NB, Krantz ID. (2004) Subtelomeric deletions of chromosome 9q: a novel microdeletion syndrome. *Am J Med Genet A;* 128A:340-51.
- Sadakata T, Furuichi T. Ca(2+)-dependent activator protein for secretion 2 and autistic-like phenotypes. *Neurosci Res.* 2010;67(3):197-202.

- Sadakata T, Furuichi T. Developmentally regulated Ca²⁺-dependent activator protein for secretion 2 (CAPS2) is involved in BDNF secretion and is associated with autism susceptibility. *Cerebellum*. 2009 ;8(3):312-22.
- Sadakata T, Kakegawa W, Mizoguchi A, Washida M, Katoh-Semba R, Shutoh F, Okamoto T, Nakashima H, Kimura K, Tanaka M, Sekine Y, Itohara S, Yuzaki M, Nagao S, Furuichi T. Impaired cerebellar development and function in mice lacking CAPS2, a protein involved in neurotrophin release. *J Neurosci*. 2007;27(10):2472-82.
- Sadakata T, Washida M, Morita N, Furuichi T. Tissue distribution of Ca²⁺-dependent activator protein for secretion family members CAPS1 and CAPS2 in mice. *J Histochem Cytochem*. 2007;55(3):301-11.
- Sadakata T, Washida M, Furuichi T. Alternative splicing variations in mouse CAPS2: differential expression and functional properties of splicing variants. *BMC Neurosci*. 2007;8:25.
- Sadakata T, Washida M, Iwayama Y, Shoji S, Sato Y, Ohkura T, Katoh-Semba R, Nakajima M, Sekine Y, Tanaka M, Nakamura K, Iwata Y, Tsuchiya KJ, Mori N, Detera-Wadleigh SD, Ichikawa H, Itohara S, Yoshikawa T, Furuichi T. Autistic-like phenotypes in Cadps2-knockout mice and aberrant CADPS2 splicing in autistic patients. *J Clin Invest*. 2007 ;117(4):931-43
- Schroer RJ, Phelan MC, Michaelis RC, Crawford EC, Skinner SA, Cuccaro M, Simensen RJ, Bishop J, Skinner C, Fender D, Stevenson RE. Autism and maternally derived aberrations of chromosome 15q. *Am J Med Genet*. 1998 Apr 1;76(4):327-36.
- Solinas-Toldo, et al., Matrix-based comparative genomic hybridization: biochips to screen for genomic imbalances. *Genes Chromosomes Cancer*, 1997;20: 399–407.
- Spielman, McGinnis, Ewens. Transmission test for linkage disequilibrium: the insulin gene region and insulin-dependent diabetes mellitus (IDDM). *Am J Hum Genet*, 1993; 52:506– 516.
- Stefansson H et al. Large recurrent microdeletions associated with schizophrenia. *Nature*, 2008; 455:232-236.
- Stefansson H, et al. Common variants conferring risk of schizophrenia. *Nature*, 2009; 460:744-747.
- Steffenburg S, Gillberg C, Hellgren L, Andersson L, Gillberg IC, Jakobsson G, Bohman M. A twin study of autism in Denmark, Finland, Iceland, Norway and Sweden. *J Child Psychol Psychiatry*. 1989 May;30(3):405-16
- Suarez, The affected sib pair IBD distribution for HLA-linked disease susceptibility genes. *Tissue Antigens*, 1978; 12:87– 93.
- Sutcliffe JS, Nelson DL, Zhang F, Pieretti M, Caskey CT, Saxe D, Warren ST. DNA methylation represses FMR-1 transcription in fragile X syndrome. *Hum Mol Genet*. 1992 Sep;1(6):397-400.
- Swank MW, Sweatt JD: Increased histone acetyltransferase and lysine acetyltransferase activity and biphasic activation of the ERK/RSK cascade in insular cortex during novel taste learning. *J Neurosci* 2001; 21: 3383–3391.
- Szatmari P, Jones MB, Zwaigenbaum L, MacLean JE. Genetics of autism: overview and new directions. *J Autism Dev Disord*. 1998; 28:351-368.

- Tabuchi K, Blundell J, Etherton MR, Hammer RE, Liu X, Powell CM, Südhof TC. *Science* A neuroligin-3 mutation implicated in autism increases inhibitory synaptic transmission in mice. 2007 Oct 5;318(5847):71-6. Epub 2007 Sep 6.
- Tao J, Van Esch H, Hagedorn-Greiwe M et al: Mutations in the X linked cyclin-dependent kinase-like 5 (CDKL5/STK9) gene are associated with severe neuro developmental retardation. *Am J Hum Genet* 2004; 75: 1149– 1154
- Tarpey P, Parnau J, Blow M et al: Mutations in the DLG3 gene cause non syndromic X-linked mental retardation. *Am J HumGenet* 2004; 75: 318– 324.
- Tarpey PS, Smith R, Pleasance E, Whibley A, et al: A systematic, large-scale resequencing screen of X-chromosome coding exons in mental retardation. *Nat Genet*; (2009) 41(5):535-43.
- Taylor SS, Radzio-Andzelm E, Hunter T. (1995) How do protein kinases discriminate between serine/threonine and tyrosine? Structural insights from the insulin receptor protein-tyrosine kinase. *FASEB J*; 9(13):1255-66.
- The International HapMap Consortium. The International HapMap Project. *Nature*, 2003; 426:789-796.
- The International HapMap Consortium. A haplotype map of the human genome. *Nature*, 2005; 437:1299-1320.
- Toro R, Konyukh M, Delorme R, Leblond C, Chaste P, Fauchereau F, Coleman M, Leboyer M, Gillberg C, Bourgeron T. Key role for gene dosage and synaptic homeostasis in autism spectrum disorders. *Trends Genet.* 2010;26(8):363-72.
- Trivier E, De Cesare D, Jacquot S et al: Mutations in the kinase Rsk-2 associated with Coffin –Lowry syndrome. *Nature* 1996; 384:567– 570
- Tuzun E, Sharp AJ, Bailey JA, Kaul R, Morrison VA, Pertz LM, Haugen E, Hayden H, Albertson D, Pinkel D, Olson MV, Eichler EE. Fine-scale structural variation of the human genome. *Nat Genet.*, 2005; 37: 660-661.
- Ueda S, Fujimoto S, Hiramoto K, et al. Dock4 regulates dendritic development in hippocampal neurons. *J. Neurosci. Res*; 2008. Vorstman JA, Staal WG, Hochstenbach PF, Franke L, van Daalen E, van Engeland H. Overview of cytogenetic regions of interest (CROIs) associated with the autism phenotype across the human genome. *Mol Psychiatry* 2006; 11: 18–28.
- Ullmann R, Turner G, Kirchhoff M, Chen W, Tonge B, et al. 2007. Array CGH identifies reciprocal 16p13.1 duplications and deletions that predispose to autism and/or mental retardation. *Hum. Mutat.* 28:674–8261
- Vaillend C, Poirier R, Laroche S. (2008) Genes, plasticity and mental retardation. *Behav Brain Res*; 192(1):88-105.

van Bon BW, Koolen DA, Borgatti R, Magee A, Garcia-Minaur S, Rooms L, Reardon W, Zollino M, Bonaglia MC, De Gregori M, Novara F, Grasso R, Ciccone R, van Duyvenvoorde HA, Aalbers AM, Guerrini R, Fazzi E, Nillesen WM, McCullough S, Kant SG, Marcelis CL, Pfundt R, de Leeuw N, Smeets D, Sistermans EA, Wit JM, Hamel BC, Brunner HG, Kooy F, Zuffardi O, de Vries BB. (2008) Clinical and molecular characteristics of 1qter microdeletion syndrome: delineating a critical region for corpus callosum agenesis/hypogenesis. *J Med Genet.*;45(6):346-54.

Van Esch H, Bauters M, Ignatius J, Jansen M, Raynaud M, Hollanders K, Lugtenberg D, Bienvenu T, Jensen LR, Gecz J, Moraine C, Marynen P, Fryns JP, Froyen G. (2005) Duplication of the MECP2 region is a frequent cause of severe mental retardation and progressive neurological symptoms in males. *Am J Hum Genet*; 77:442-53.

Verkerk AJ, Pieretti M, Sutcliffe JS, Fu YH, Kuhl DP, et al. 1991. Identification of a gene (*FMR-1*) containing a CGG repeat coincident with a breakpoint cluster region exhibiting length variation in fragile X syndrome. *Cell* 65:905-141

Vissers LE, van Ravenswaaij CM, Admiraal R, Hurst JA, de Vries BB, Janssen IM, van der Vliet WA, Huys EH, de Jong PJ, Hamel BC, Schoenmakers EF, Brunner HG, Veltman JA, Geurts van Kessel A. (2004) Mutations in a new member of the chromodomain gene family cause CHARGE syndrome. *Nat Genet*; 36:955-7.

Volkmar F, Stier DM, Cohen DJ. Age of recognition of pervasive developmental disorder. *American Journal of Psychiatry* 1985, 142:1450-1452.

Volkmar FR, Lord C, Bailey A, Schultz RT, Klin A (2004) Autism and pervasive developmental disorders. *J Child Psychol Psychiatry* 45:135-70.

Volkmar FR, Pauls D. Autism. *Lancet*. 2004 Jan 17;363(9404):250

Vorstman JA, Staal WG, Hochstenbach PF, Franke L, van Daalen E, van Engeland H. Overview of cytogenetic regions of interest (CROIs) associated with the autism phenotype across the human genome. *Mol Psychiatry* 2006; 11: 18-28.

Vorstman JAS, Staal WG, van Daalen E, van Engeland H, Hochstenbach PFR and Franke L. Identification of novel autism candidate regions through analysis of reported cytogenetic abnormalities associated with autism. *Molecular Psychiatry*, 2006;11:18-28.

Vorstman JAS, van Daalen E, Jalali GR, Schmidt ERE, Pasterkamp RJ, de Jonge M, Hennekam EAM, Janson E, Staal WG, van der Zwaag B, Burbach JPH, Kahn RS, Emanuel BS, van Engeland H, Ophoff RA. A double hit implicates DIAPH3 as an autism risk gene. *Mol Psychiatry* 2010.

Yoon S, Xuan Z, Makarov V, Ye K, Sebat J: Sensitive and accurate detection of copy number variants using read depth of coverage. *Genome Res* 2009; 19: 1586-1592

Wang K, Li M, Hadley D *et al*: Penn CNV: an integrated hidden Markov model designed for high-resolution copy number variation detection in whole-genome SNP genotyping data. *Genome Res* 2007; 17: 1665-1674

- Wang K, Zhang H, Ma D, Bucan M et al. Common genetic variants on 5p14.1 associate with autism spectrum disorders. *Nature*, 2009;459:528-533
- Wassink TH, Piven J, Patil SR. Chromosomal abnormalities in a clinic sample of individuals with autistic disorder. *Psychiatr Genet* 2001; 11: 57–63.
- Wassink TH, Piven J, Vieland VJ, Huang J, Swiderski RE, Pietila J et al. Evidence supporting WNT2 as an autism susceptibility gene. *Am J Med Genet* 2001; 105: 406–413.
- Weaving LS, Christodoulou J, Williamson SL et al: Mutations of *CDKL5* cause a severe neuro developmental disorder with infantile spasms and mental retardation. *Am J Hum Genet* 2004; 75:1079– 1093
- Weiss LA, Arking DE; Gene Discovery Project of Johns Hopkins & the Autism Consortium, Daly MJ, Chakravarti A. A genome-wide linkage and association scan reveals novel loci for autism. *Nature* 2009; 461: 802-8.
- Weiss LA, Shen Y, Korn JM, Arking DE, Miller DT, et al. 2008. Association between microdeletion and microduplication at 16p11.2 and autism. *N. Engl. J. Med.* 358:667–75158.
- Weiler IJ, Irwin SA, Klintsova AY, Spencer CM, Brazelton AD, et al. 1997. Fragile X mental retardation protein is translated near synapses in response to neurotransmitter activation. *Proc. Natl. Acad. Sci. USA*94:5395–40040
- Whitney ER, Kemper TL, Bauman ML, Rosene DL, Blatt GJ. Cerebellar Purkinje cells are reduced in a subpopulation of autistic brains: a stereological experiment using calbindin-D28k. *Cerebellum* 2008;7(3):406-16.
- Wilson HL, Wong ACC, Shaw SR, Tse, WY, Stapleton GA, Phelan MC, Hu S, Marshall J, McDermid HE. Molecular characterisation of the 22q13 deletion syndrome supports the role of haploinsufficiency of SHANK3/PROSAP2 in the major neurological symptoms. *J. Med. Genet.* 2003; 40: 575-584.
- Woodhouse W, Bailey A, Rutter M, Bolton P, Baird G, Le Couteur A. Head circumference in autism and other pervasive developmental disorders. *J Child Psychol Psychiatry.* 1996 Sep;37(6):665-71.
- Wong KK, deLeeuw RJ, Dosanjh NS et al: A Comprehensive Analysis of Common Copy-Number Variations in the Human Genome. *Am J Hum Genet* 2007; **80**: 91-104.
- Zappella M. Autistic regression with and without EEG abnormalities followed by favourable outcome. *Brain Dev.* 2010; 32(9):739-45.
- Zhao X, Leotta A, Lajonchere C, Geschwind DH, Law K, Law P, Qiu S, Lord C, Sebat J, Ye K, Wigler M. A unified genetic theory for sporadic and inherited autism. *Proc Natl Acad Sci U S A.* 2007;104(31):12831-6.

