



UNIVERSITÀ DEGLI STUDI DI TRIESTE

XXX CICLO DEL DOTTORATO DI RICERCA IN

NEUROSCIENZE E SCIENZE COGNITIVE

VALUTAZIONE DEL BACKGROUND GENETICO DI UNA COORTE DI INDIVIDUI DISLESSICI MEDIANTE L'UTILIZZO DI TECNOLOGIE AD ALTA PROCESSIVITÀ

Settore scientifico-disciplinare: M-PSI/01 PSICOLOGIA GENERALE

**DOTTORANDO
EMMANOUIL ATHANASAKIS**

**COORDINATORE
PROF. TIZIANO AGOSTINI**

**SUPERVISORE DI TESI
DOTTOR FLAVIO FALETRA**

**CO-SUPERVISORI DI TESI
PROF. WALTER GERBINO
PROF. PAOLO EDOMI**

ANNO ACCADEMICO 2016/2017

..... dedicata a mia moglie Paola

e ai miei figli Γκαρος και Τωνας

Abstract

Exploring the genetic background of a dyslexia cohort using high throughput technologies

Technological improvements and continued cost reduction have significantly contributed to the progress of identifying the genetic causes of complex traits. Here we report the results of a genetic screening on a dyslexia cohort combining targeted next generation sequencing and high density SNP array. The study cohort consists of 49 subjects with dyslexia and 52 subjects with dyslexia and other specific learning disabilities (dysorthographia, dysgraphia, dyscalculia). All samples were sequenced on Ion Torrent platform, targeting the coding regions and their exon-intron boundaries of 12 candidate dyslexia genes (*CMIP*, *CNTNAP2*, *CYP19A1*, *DCDC2*, *DIP2A*, *DYX1C1*, *GCFC2*, *KIAA0319*, *KIAA0319L*, *MRPL19*, *ROBO1*, *S100B*), with focus on novel and rare variants. A subset of 54 samples was further analyzed, genotyping over 1.7 M markers (Multi Ethnic Global Array design, Illumina), for copy number variation (CNV) discovery and characterization according to the literature. For this purpose, high confidence CNVs were obtained using the *cnvPartition* and the *PennCNV* calling algorithms. We report a total of 12 pathogenic predicted variants, among which two novel deleterious events (*DIP2A*:p.G1387* and *KIAA0319*:p.V774Afs*37) and a known rare splicing variant (*GCFC2*:c.266-2A>G). Moreover, several copy number variants were identified, overlapping some language related genes, but not any of the above sequenced genes. Finally, a sibling pair was found to harbor duplications in the chromosome band 16p13.11, a susceptibility region for several neurodevelopmental disorders. The present study enriches our knowledge about the genetic background in a dyslexia cohort. At the same time our findings emphasize the need for further research to attribute causative roles of these events for cohort phenotypes.

INDICE

1. INTRODUZIONE	1
1.1 I Disturbi di Apprendimento (<i>Learning Disabilities</i>)	1
1.2 Le <i>Raccomandazioni</i> sui Disturbi Specifici di Apprendimento	2
1.3 I Disturbi Specifici di Apprendimento	3
1.4 La Dislessia dello sviluppo	5
1.4.1 Dati Epidemiologici	5
1.4.2 Multifattorialità	5
1.4.3 DYX locus e geni candidati	7
1.4.4 Altri locus e geni candidati	11
1.4.5 Il ruolo delle <i>Copy Number Variation</i> (CNV)	13
1.5 Tecnologie di indagine genetica ad alta processività	14
1.5.1 <i>SNP-array</i>	14
1.5.2 <i>Targeted Next Generation Sequencing</i>	16
2. SCOPO DELLA TESI	18
3. MATERIALI E METODI	19
3.1 Coorte di studio	19
3.2 Bio-banca	20
3.3 <i>Targeted Next Generation Sequencing</i>	20
3.3.1 Pannello di geni candidati	21
3.3.2 Costruzione delle librerie	22
3.3.3 Sequenziamento NGS	26
3.4 Annotazione e prioritizzazione delle varianti	26
3.4.1 Annovar e Banche dati	27
3.4.2 Prioritizzazione delle varianti	28
3.5 Sequenziamento diretto (<i>Sanger sequencing</i>)	29
3.6 Analisi dei campioni mediante <i>SNP-array</i>	31
3.6.1 Genotipizzazione	33
3.6.2 Chiamata delle CNV	33
3.6.3 Selezione e caratterizzazione delle CNV ad “Alta Confidenza”	34

4. RISULTATI E DISCUSSIONE	36
4.1 <i>Targeted NGS</i>	36
4.1.1 Varianti predette con un effetto deleterio	36
4.1.2 Varianti predette altamente patogenetiche	38
4.2 Chiamata delle CNV	40
4.2.1 CNV contenenti geni associati (o riportati) con la lettura o linguaggio	43
4.2.2 CNV comuni tra fratelli	47
4.2.3 CNV non riportate nel DGV database	49
5. CONCLUSIONI	51
6. TABELLE SUPPLEMENTARI	53
Tabella S1	53
Tabella S2	56
Tabella S3	57
Tabella S4	60
7. FIGURE SUPPLEMENTARI	70
Figura S1	70
Figura S2	72
Figura S3	78
Figura S4	80
8. BIBLIOGRAFIA	83

1. INTRODUZIONE

1.1 I Disturbi di Apprendimento (*Learning Disabilities*)

I profili fondamentali che potrebbero sottostare a una difficoltà scolastica importante sono i seguenti: un disturbo specifico di apprendimento nel 4% dei casi, disturbi specifici collegati (4%), una condizione di handicap (1.2%), uno svantaggio socioculturale grave oppure difficoltà scolastiche in altre aree scolastiche rilevanti. L'importanza di considerare tutte queste categorie risiede nel fatto che non sempre i confini fra una categoria e l'altra sono evidenti, che due problematiche possono essere compresenti, senza la possibilità di stabilire in modo inequivocabile che l'una è la conseguenza dell'altra e soprattutto che le procedure diagnostiche e strategie di intervento possono essere simili indipendentemente dal fatto che le eziologie siano differenti (Cornoldi, 2007).

Nel 1990 Hammill ha presentato una definizione di *Learning Disabilities* (LD, Disturbo di Apprendimento) sostenendo che con questo termine ci si riferisce ad un gruppo eterogeneo di disturbi manifestati da significative difficoltà nell'acquisizione e nell'uso di abilità di ascolto, espressione orale, lettura, ragionamento e matematica, presumibilmente dovuti a disfunzioni del sistema nervoso centrale. Possono coesistere con la LD problemi nei comportamenti di autoregolazione, nella percezione sociale, pur non costituendo di per sé una LD. Inoltre, le LD possono verificarsi in concomitanza con altri fattori di disabilità o con influenze estrinseche (culturali, d'istruzione, ecc.) ma non sono il risultato di quelle condizioni o influenze (Hammill, 1990).

La categoria verrebbe a raccogliere una gamma diversificata di problematiche nello sviluppo cognitivo e nell'apprendimento scolastico, non imputabili primariamente a fattori di disabilità mentale grave e definibili in base al mancato raggiungimento di criteri attesi di apprendimento (per i quali esista un largo consenso) rispetto alle potenzialità generali del soggetto (Cornoldi, 1991).

1.2 Le Raccomandazioni sui Disturbi Specifici di Apprendimento

Dal gennaio 2007 in Italia sono disponibili le *Raccomandazioni* per la pratica clinica sui Disturbi Specifici dell'Apprendimento (DSA) elaborate con il metodo della *Consensus Conference*. Nelle Raccomandazioni si ribadisce che la principale caratteristica di definizione di questa "categoria nosografica" è quella della "specificità", con riferimento al fatto che il disturbo interessa uno specifico dominio di abilità in modo significativo ma circoscritto, lasciando intatto il funzionamento intellettivo generale. In questo senso, il principale criterio necessario per stabilire la diagnosi di DSA è quello della "discrepanza" tra abilità nel dominio specifico interessato (deficitaria in rapporto alle attese per l'età e/o la classe frequentata) e l'intelligenza generale (adeguata per l'età cronologica). Dal riconoscimento del criterio della discrepanza, come aspetto cardinale della definizione e della diagnosi di DSA, vengono fatte derivare alcune fondamentali implicazioni sul piano diagnostico, ovvero la necessità di usare test standardizzati, sia per misurare l'intelligenza generale, che l'abilità specifica e la necessità di escludere altre condizioni che potrebbero influenzare i risultati di questi test, come ad esempio menomazioni sensoriali e neurologiche gravi, disturbi significativi della sfera emotiva, situazioni ambientali di svantaggio socioculturale che possono interferire con un'adeguata istruzione. Anche se esistono alcune difformità, anche a livello internazionale, su come concettualizzare, operationalizzare, e applicare il criterio della discrepanza, la *Consensus Conference* ha rilevato un sostanziale accordo sul fatto che la compromissione dell'abilità specifica deve essere significativa (-2 deviazione standard (DS) dai valori normativi attesi per l'età o la classe frequentata qualora non coincida con l'età del bambino) e che il livello intellettivo deve essere nei limiti di norma, che operationalizzato significa un quoziente d'intelligenza non inferiore a -1DS (equivalente di solito a un valore di 85) rispetto ai valori medi attesi per l'età (Cornoldi, 2007).

Altre caratteristiche critiche dei DSA descritte nel documento della *Consensus* sono: il carattere "evolutivo" di questi disturbi; la diversa espressività del disturbo nelle diverse fasi evolutive dell'abilità in questione; la quasi costante associazione ad altri disturbi (comorbidità), fatto questo che determina la marcata eterogeneità dei profili funzionali e di espressività con cui i DSA si manifestano, e che comporta significative ricadute sul versante dell'indagine diagnostica; il carattere neurobiologico delle anomalie processuali che caratterizzano i DSA che interagiscono attivamente nella determinazione della comparsa del disturbo, con i fattori ambientali; il fatto che il disturbo specifico deve comportare un impatto significativo e negativo per l'andamento scolastico e/o per le attività della vita quotidiana. Viene inoltre riconosciuta la

possibile esistenza di un disturbo di apprendimento (non categorizzabile come specifico) in presenza di altre patologie o anomalie, sensoriali, neurologiche, cognitive e psicopatologiche, che normalmente costituiscono criteri di esclusione, quando l'entità del deficit settoriale è tale che non può essere spiegata solo sulla base di queste patologie. In tali casi, poiché lo stato attuale delle conoscenze non consente di distinguere in modo compiuto le relazioni eziopatogenetiche fra i disturbi compresenti, si esprime la raccomandazione ad estendere e ad approfondire la valutazione diagnostica su tutte le aree implicate (Cornoldi, 2007).

Tra i punti più significativi delle *Raccomandazioni*, è importante sottolineare i seguenti: la precisa definizione di dislessia, disortografia, disgrafia, discalculia; la specificazione del livello di discrepanza rispetto ai dati normativi da utilizzare nella valutazione psicometrica per soddisfare i criteri di inclusione (-2DS o sotto il 5° percentile) considerando sempre sia il parametro correttezza che quello di velocità; la raccomandazione a fare riferimento a strumenti con adeguate norme di riferimento e proprietà psicometriche, con un richiamo a considerare sempre l'errore standard di misura; l'indicazione di un QI inferiore a 85 come condizione di esclusione; l'invito a porre attenzione anche alle condizioni in cui non sono soddisfatti tutti i criteri di esclusione, per esempio presenza di problemi neurologici o sensoriali (disturbi non specifici), per capire se il problema di apprendimento sia realmente compatibile con questa condizione; l'assunzione di comorbidità (o comorbidity) come semplice co-occorrenza di disturbi senza sposare nessun modello eziopatogenetico comune; la raccomandazione di non fare diagnosi prima di due anni di regolare scolarizzazione per quanto riguarda dislessia e disortografia e di tre anni per quella di discalculia; la raccomandazione di concludere ogni valutazione con un referto scritto; l'indicazione di precise condizioni di rischio rilevabili alla fine della prima elementare da tenere sotto controllo anche con interventi di supporto all'apprendimento; l'indicazione precisa delle caratteristiche di un progetto riabilitativo e la specificazione delle differenze tra riabilitazione e abilitazione (Cornoldi, 2007).

1.3 I Disturbi Specifici di Apprendimento

I disturbi di apprendimento esaminati dalla *Consensus Conference* interessano la condizione clinica evolutiva di difficoltà di apprendimento della lettura, della scrittura e del calcolo che si manifesta con l'inizio della scolarizzazione. Sono pertanto escluse le patologie di apprendimento acquisite (*Consensus Conference*, Istituto Superiore di Sanità 2011).

I riferimenti internazionali utilizzati nella definizione e classificazione dei disturbi specifici dell'apprendimento (DSA) sono:

- ICD-10 (F81 Disturbi evolutivi specifici delle abilità scolastiche)
- DSM IV TR (315 Disturbi dell'apprendimento).

Si tratta di disturbi che coinvolgono uno specifico dominio di abilità, lasciando intatto il funzionamento intellettuale generale. Essi infatti interessano le competenze strumentali degli apprendimenti scolastici.

Sulla base del deficit funzionale vengono comunemente distinte le seguenti condizioni cliniche:

- Dislessia, cioè disturbo nella lettura (intesa come abilità di decodifica del testo)
- Disortografia, cioè disturbo nella scrittura (intesa come abilità di codifica fonografica e competenza ortografica)
- Disgrafia, cioè disturbo nella grafia (intesa come abilità grafo-motoria)
- Discalculia, cioè disturbo nelle abilità di numero e di calcolo (intese come capacità di comprendere e operare con i numeri).

La Dislessia dello sviluppo: Da un punto di vista clinico, la dislessia si manifesta attraverso una minore correttezza e rapidità della lettura a voce alta rispetto a quanto atteso per età anagrafica, classe frequentata, istruzione ricevuta. Risultano più o meno deficitarie la lettura di lettere, di parole e non-parole, di brani. In generale, l'aspetto evolutivo della dislessia può ricordare un semplice rallentamento del processo di sviluppo. Tale considerazione è utile per l'individuazione di eventuali segnali anticipatori, fin dalla scuola dell'infanzia.

La Disgrafia e la Disortografia: Il disturbo specifico di scrittura si definisce disgrafia o disortografia, a seconda che interessi rispettivamente la grafia o l'ortografia. La disgrafia fa riferimento al controllo degli aspetti grafici, formali, della scrittura manuale, ed è collegata al momento motorio-esecutivo della prestazione; La disgrafia si manifesta in una minore fluidità e qualità dell'aspetto grafico della scrittura. La disortografia riguarda invece l'utilizzo, in fase di scrittura, del codice linguistico in quanto tale. La disortografia è all'origine di una minore correttezza del testo scritto; entrambe, naturalmente, sono in rapporto all'età anagrafica dell'alunno. In particolare, la disortografia si può definire come un disordine di codifica del

testo scritto, che viene fatto risalire ad un deficit di funzionamento delle componenti centrali del processo di scrittura, responsabili della transcodifica del linguaggio orale nel linguaggio scritto.

La Discalculia: La discalculia riguarda l'abilità di calcolo, sia nella componente dell'organizzazione della cognizione numerica (intelligenza numerica basale), sia in quella delle procedure esecutive e del calcolo. Nel primo ambito, la discalculia interviene sugli elementi basali dell'abilità numerica: il riconoscimento immediato di piccole quantità, I meccanismi di quantificazione, la seriazione, la comparazione, le strategie di composizione e scomposizione di quantità, le strategie di calcolo a mente. Nell'ambito procedurale, invece, la discalculia rende difficoltose le procedure esecutive per lo più implicate nel calcolo scritto: la lettura e scrittura dei numeri, l'incolonnamento, il recupero dei fatti numerici e gli algoritmi del calcolo scritto vero e proprio.

1.4 La Dislessia dello sviluppo

1.4.1 Dati Epidemiologici

Anche se la dislessia è uno dei più comuni disturbi neurocomportamentali che interessano i bambini, la prevalenza è incerta e i dati disponibili sono scarsi e datati. Molti studi si sbilanciano sulle % di disturbi dell'apprendimento presenti in Italia, con valori che vanno dall'1.5% al 5%. La prevalenza nella regione Friuli Venezia Giulia in bambini di età scolare tra gli 8 e i 10 anni è di circa 3.1%-3.2% (Barbiero et al., 2012).

Johnson (1988) si sofferma sulle differenze sessuali, si ha nella maggioranza dei casi prova di una maggiore presenza di problemi di lettura nei maschi, effetto che è stato associato a fattori sia genetici, sia neurologici, sia ambientali. Johnson da un lato osserva come linguaggio e consapevolezza fonologica, due variabili cognitive fortemente legate al successo in lettura, siano generalmente superiori nelle femmine, dall'altro fa notare come le femmine tollerino meglio i problemi emergenti da difficoltà scolastiche.

1.4.2 Multifattorialità

Molti fattori influenzano la capacità di lettura. Da un lato, l'ambiente svolge un ruolo cruciale: l'esempio più drastico è che senza un'esposizione al testo, nessuna capacità di lettura può svilupparsi. Importanti fattori ambientali includono l'alfabetizzazione domestica, lo stato

socio-economico e l'istruzione genitoriale (Peterson e Pennington, 2015). I fattori genetici influenzano anche la diversificazione: le difficoltà di lettura sono familiari, e gli studi di ereditarietà hanno stimato che 0,40-0,80 della responsabilità alla dislessia è dovuto alla variazione genetica (De Fries et al., 1987). Le influenze genetiche non sono limitate alla diagnosi della dislessia, ma influenzano anche le variazioni normali nelle prestazioni di lettura (ereditarietà stimata da 0,57-0,67) (Harlaar et al., 2007; de Zeeuw et al., 2015) oltre a diverse abilità che supportano la capacità di lettura, inclusa la rapida denominazione (0,46) e la consapevolezza fonologica (0,61) (Petrill et al., 2006).

La componente genetica è complessa ed eterogenea dove molti geni possono essere responsabili delle difficoltà di lettura. Per quanto riguarda altri tratti complessi (Gottesman e Gould, 2003) si assume un modello di soglia multifattoriale; si pensa che la maggior parte delle varianti di rischio genetiche spieghino una frazione molto piccola (<1%) della varianza totale e sono comuni nella popolazione generale, cioè con una frequenza dell'allele minore (MAF) superiore al 5%. Le persone al di sopra della soglia diagnostica per la dislessia sono arricchite per quelle varianti che agiscono come fattori di rischio additivo. Tuttavia, le sole varianti comuni a basso effetto, potrebbero non essere in grado di spiegare completamente le stime di ereditarietà. Ci possono anche essere varianti che sono rare nella popolazione generale (MAF <1%) ma che hanno una penetranza sostanziale. Queste possono spiegare il modello di ereditarietà dominante Mendeliano della dislessia che è stato osservato in alcune famiglie. Quindi, sia le variazioni comuni che rare possono potenzialmente contribuire a tratti complessi come la lettura (Schork et al., 2009), e i fattori di rischio genetici per la responsabilità della dislessia sono eterogenei (Figura 1.1).

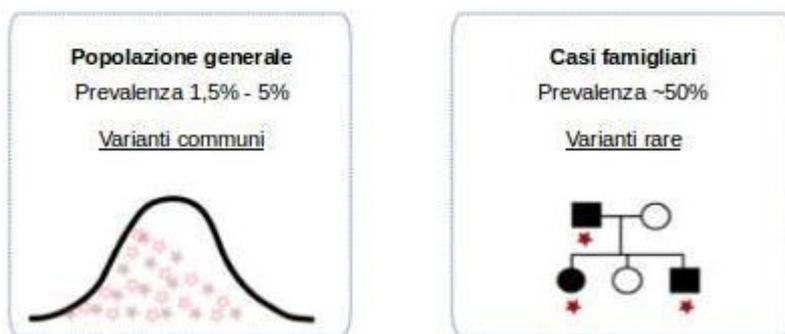


Figura 1.1: Rappresentazione schematica di due possibili tipi di contributo genetico alla base dell'insorgenza della dislessia. Lo schema a sinistra si riferisce a varianti comuni con un piccolo effetto additivo al fenotipo. Lo schema a destra, si tratta di casi famigliari dove poche rare o nuove varianti possono contribuire sostanzialmente al fenotipo.

1.4.3 DYX locus e geni candidati

DYX1 (15q21): Questo locus è stato il primo a essere associato alle disabilità di lettura, e questa associazione è stata confermata da diversi studi indipendenti (Grigorenko et al., 1997; Schulte-Korne et al., 1998; Chapman et al., 2004; Bates et al., 2007; Platko et al., 2008). In questa regione, studiando una famiglia finlandese, è stata identificata una traslocazione bilanciata che interrompeva il gene *DYX1C1* e che co-segregava con le disabilità di lettura (Nopola-Hemmi et al., 2000; Taipale et al., 2003). Più recentemente Anthoni e colleghi (2012) hanno effettuato esami più approfonditi su un'altra famiglia di dislessici, riportata sempre da Nopola-Hemmi, dove anche essa riportava una traslocazione simile alla precedente. Studi di associazione mirati hanno rilevato degli SNP presenti nel gene *CYP19A1*, associati a misure quantitative del linguaggio.

Il gene *DYX1C1* codifica per una proteina la quale prende parte nella migrazione neuronale, nell'assemblamento delle cilia e la loro motilità (Tammimies et al., 2013; Chandrasekar et al., 2013; Tarkar et al., 2013). Analogamente, il gene *CYP19A1* codifica per un enzima, aromatasi, che fa parte nella catena dei conversione degli androgeni in estrogeni. Comunque, anche questo enzima sembra prendere parte alla migrazione neuronale e nella crescita dei dendriti (Anthoni et al., 2012).

DYX2 (6p22.-p21.3): Le prime evidenze di un locus di suscettibilità alla dislessia nel cromosoma 6 sono arrivate nel 1994 da Cardon e colleghi indicando la regione 6p21.3, dove è localizzato anche la *Human Leukocyte Antigen complex* (HLA). In generale, la regione 6p21.3-23 è stata oggetto di tantissimi studi, usando diversi tratti relativi alla lettura, come anche diversi criteri di campionamento e di approcci di studio (Platko et al., 2008; Fisher et al., 2002; Kaplan et al., 2002; Fisher et al., 1999; Grigorenko et al., 1997). Lo stesso locus è stato confermato anche da uno studio su una coorte con SLI (SLI, *Specific Language Impairment*) da Rice e colleghi nel 2009.

Come per altri loci, anche qui ci sono degli studi contrastanti. Gli studi di Rice concernenti coorti con SLI, non hanno trovato evidenze di associazione del locus DYX2 (The SLI Consortium, 2002; Bartlett et al., 2002). Due altri studi, uno sulle famiglie tedesche, focalizzato sulla disabilità di spelling (Schulte-Korne et al., 1998), e uno studio di associazione delle abilità di lettura nella popolazione generale (Bates et al., 2007), non hanno potuto riportare nuove evidenze sul ruolo del locus DYX2 nella suscettibilità alla dislessia. Infine, analogamente anche il progetto Olandese DDP (*Dutch Dyslexia Project*) dove hanno studiato 108 famiglie composte

da almeno due membri affetti da dislessia, non ha trovato nessuna associazione tra questo locus e diversi tratti quantitativi in esame (de Kovel et al., 2008).

Tuttavia nei diversi studi che hanno sia confermato o smentito un ruolo di suscettibilità del locus *DYX2* alla dislessia, due geni candidati sono stati proposti: *KIAA0319* e *DCDC2*. Il primo gene, *KIAA0319*, era stato identificato mediante studi di associazione dove diversi SNPs situati nella regione del promotore presentavano una significativa associazione con la disabilità di lettura (Francks et al., 2004; Cope et al., 2005; Harold et al., 2006). SNPs presenti in questo gene sono stati associati anche con la SLI indicando un ruolo di questo gene più generale nelle capacità di apprendimento del linguaggio (Rice et al., 2009; Newbury et al., 2011).

Il secondo gene, *DCDC2*, era stato identificato mediante studi di associazione tra disabilità di lettura e due varianti genetiche: una piccola delezione (Meng et al., 2005; Marino et al., 2012) e un STR (*Short Tandem Repeat*) (Schumacher et al., 2006), entrambi situati nel medesimo gene. Altri marcatori sono stati proposti dopo la loro associazione con la dislessia e tratti relativi alla lettura (Meng et al., 2005; Schumacher et al., 2006; Wilcke et al., 2009; Newbury et al., 2011; Harold et al. 2006; Scerri et al., 2011).

Studi funzionali mediante knockdown dei geni *Kiaa0319* e *Dcdc2* hanno suggerito il coinvolgimento di entrambi nella migrazione neuronale durante lo sviluppo del sistema nervoso centrale (Velayos-Baeza et al., 2007; 2008; Peschansky et al., 2010; Adler et al., 2013; Meng et al., 2005; Wang et al., 2011). Inoltre, per il *Dcdc2* è stato proposto un suo ruolo nella regolazione della struttura ciliare, come anche della sua lunghezza e trasduzione del segnale (Massinen et al., 2011; Grati et al., 2015; Schueler et al., 2015).

DYX3 (2p12-p16): Nel 1999 Fagerhein e colleghi studiando una famiglia Norvegese dove la dislessia sembrava essere trasmessa secondo un modello dominante, hanno identificato e candidato la regione p12-p16 del cromosoma 2. Negli anni successivi nuovi studi hanno supportato il ruolo di questo locus (Fisher et al., 2002; Francks et al., 2002; Petryshen et al., 2002; Kaminen et al., 2003; de Kovel et al., 2008). Oltre a questa regione, anche la 2q23.3 è stata associata in diverse famiglie con dislessia (Raskind et al., 2005), come anche con la lettura di parole irregolari e di spelling nella popolazione generale (Bates et al., 2007)

I primi geni candidati erano i *SEMA4F*, *OTX1* e *TACR1* (Francks et al., 2002; Peyrard-Janvid et al., 2004). Studi successivi si sono focalizzati nella regione 2p12 proponendo come geni candidati i *FLJ13391*, *MRPL19* e *GCFC2 (C2ORF3)* (Anthoni et al., 2007). Studi di espressione genica su individui portatori eterozigoti dell'aplotipo di rischio una riduzione dell'espressione dei *MRPL19* e *GCFC2 (C2ORF3)*. Malgrado queste evidenze sperimentali,

Newbury e colleghi (2011) e Scerri e colleghi (2011) non hanno replicato l'associazione di varianti presenti in questi due geni con dei tratti del linguaggio e di lettura.

DYX4 (6q11.2-q12): La regione 6q11.2-q12 era stata associata con la dislessia e altri relativi tratti quantitativi nel 2001 da Petryshen e colleghi. Questa associazione non è stata replicata, tuttavia uno studio ha riportato una forte associazione della regione 6q15 allo spelling di parole irregolari (Bates et al., 2007)

DYX5 (3p12-q13): Nel 2001 Nopola-Hemmi e colleghi hanno descritto una famiglia finlandese di 4 generazioni che presentava difficoltà di lettura, ereditate secondo un modello dominante e con la regione 3p12-q13 candidata. Ulteriori supporti su questa regione sono arrivati secondo altri 3 studi da Fisher e colleghi (2002), Bates e colleghi (2007) e Platko e colleghi (2008).

I primi geni candidati sono stati i *5HT1F* e *DRD3* (Nopola-Hemmi et al., 2001), ma successivamente *ROBO1* è stato identificato da Hannula-Jouppi e colleghi (2005) attraverso un caso di dislessia che presentava una traslocazione *de novo* che coinvolgeva il locus 3p12. In seguito alla scoperta del *ROBO1*, Hannula-Jouppi e colleghi (2005) hanno identificato un aplotipo di rischio, nella famiglia di 4 generazioni studiata nel 2001, dove co-segregava con la dislessia in 19 dai 21 membri famigliari. Sequenziamento di *ROBO1* non ha identificato delle mutazioni nella regione codificante, tuttavia l'aplotipo di rischio attenuava l'espressione del gene nei linfociti isolati dai membri affetti (Hannula-Jouppi et al., 2005).

Successivamente sono stati riportati diversi polimorfismi associati alla memoria a breve termine ma non con i test della lettura (Bates et al., 2011).

Il gene *ROBO1* codifica per un recettore coinvolto nell'orientamento assonale ("axon guidance"), il quale indirizza gli assoni nel cervello ed è coinvolto in diversi processi del neurosviluppo, tra i quali migrazione cellulare, differenziamento e formazione delle sinapsi (Andrews et al., 2006 e 2008).

DYX6 (18p11.2): Due studi di *genome-wide linkage* su tratti quantitativi, in due gruppi di famiglie indipendenti, ha rilevato che i marcatori associati alle prestazioni di lettura di singole parole coincidevano ed erano localizzati nella regione 18p11.2 (Fisher et al., 2002). Uno studio successivo su famiglie tedesche non ha confermato i dati di Fisher e colleghi (Schumacher et al., 2006b) mentre altri due studi (Bates et al., 2007; Seshadri et al., 2007) hanno riportato delle associazioni con le misure di codifica fonologica e ortografica e con le prestazioni di lettura. In

questa regione sono stati identificati e candidati 4 geni da Scerri e colleghi (2010): *MC5R*, *DYM*, *NEDD4L* e *VAPA*.

DYX7 (11p15.5): Le prime evidenze di associazione della regione 11p15 sono state riportate da Fisher e DeFries (2002) attraverso uno studio di *genome-wide linkage*. Analisi successive mediante sequenziamento diretto del gene *DRD4* (regione 11p15.5), che codifica per il recettore D4 della dopamina ed è un gene candidato per l'ADHD, non hanno riportato nessuna associazione tra varianti comuni identificate e suscettibilità alla dislessia (Hsiung et al., 2004).

DYX8 (1p34-p36): Il locus DYX8, situato nel cromosoma 1, è stato uno dei primi a essere identificato. Era stata riportata una traslocazione che coinvolgeva la regione 1p22 a co-segregare con delle disabilità di scrittura e lettura (Rabin et al., 1993). Studi successivi sul cromosoma 1 hanno confermato una nuova regione di suscettibilità alla dislessia, sempre al braccio 1p (1p34-36), attraverso studi di associazione di fenotipi qualitativi e misure quantitative (Grigorenko et al., 2001). Ulteriori conferme, più specifiche, sono arrivate dal progetto Olandese DDP su studi di coppie di fratelli che ha messo in evidenza la regione 1p36 (de Kovel et al., 2008), come anche da studi successivi su famiglie con SLI (Rice et al., 2008).

A livello di geni candidati è stato proposto il gene *KIAA0319L*, nella regione 1p34, perché molto simile al *KIAA0319*, 61% di similarità della proteina codificata (Couto et al., 2008). Diminuzione dell'espressione genica del *Kiaa0319l* in cellule embrionali di ratto porta a un fenotipo simile a quello osservato con esperimenti simili mirati sul gene *Kiaa0319* (Platt et al., 2013).

DYX9 (Xq27.3-q28): Lo DDP studio ha analizzato mediante *genome-wide linkage* diverse famiglie di tre generazioni che riportavano multipli membri affetti da dislessia (de Kovel et al., 2004). In particolare in una famiglia con 15 su 25 membri con dislessia si è identificata la regione cromosomic Xq27.3. L'aplotipo di rischio era presente in 12 su 15 membri affetti; in specifico, tutti e 4 i maschi, su 12, avevano una dislessia severa, mentre le altre 8 femmine presentavano una grande variabilità fenotipica. L'aplotipo di rischio era presente anche in una femmina non affetta. Il modello proposto era quello dominante con mutazioni causative di ridotta penetranza. Analisi più approfondite della regione interessata ha portato al sequenziamento di 4 geni: *FMR1*, *CXORF1*, *DKFZp574M2010* e *KIAA1854*. Nessuna mutazione è stata identificata e i geni/varianti causative rimangono ancora sconosciute (de Kovel et al., 2004).

La stessa regione, Xq27.3, era stata candidata da due studi indipendenti di Bates e colleghi (2007) e Platko e colleghi (2008). Più recentemente, uno studio su famiglie francesi ha ulteriormente supportato la regione Xq27.3 con nuove prove. Tuttavia, nessuna variante che co-segregasse con la dislessia è stata identificata nei geni candidati nella regione (*FMRI*, *CXORF1*, *CXORF51*, *SLITRK2*, *FMR2*, *ASFMR1* e *FMR1NB*).

1.4.4 Altri locus e geni candidati

SLI1 (16q23.1-q24): Studi di associazione tra delle coorti di SLI e tratti relativi al linguaggio hanno permesso l'identificazione del locus SLI1 (The SLI Consortium, 2002; 2004; Monaco, 2007; Falcaro et al., 2008). Successivamente, Newbury e colleghi (2009) hanno identificato dei SNP associati a questa regione. Uno di essi era localizzato all'interno del gene *CMIP*. Esso codifica per una proteina che si ipotizza possa fungere da componente del citoscheletro e prende parte nella migrazione neuronale. Ulteriori evidenze di associazione di questo gene, alla lettura e spelling, arrivano da due studi indipendenti negli anni successivi (Newbury et al., 2011; Scerri et al., 2011) che indicano un coinvolgimento di questo gene in uno spettro più generale delle capacità di apprendimento del linguaggio e di lettura.

SLI4 (7q35-q6.1): Questo locus è stato analizzato mediante studi di associazione mirati sul gene *CNTNAP2* rilevando delle significative associazioni con 3 tratti del linguaggio, ripetizione delle non parole, espressione e ricezione del linguaggio (Vernes et al., 2008). Successivamente, nuovi SNP sono stati associati con la capacità di lettura (Newbury et al., 2011; Peter et al., 2011), come anche con ulteriori tratti relativi al linguaggio (Whitehouse et al., 2011), età delle prime parole (Alarcon et al., 2008), ed età delle prime frasi (Anney et al., 2012). Complessivamente, tutte queste associazioni indicano un coinvolgimento del gene *CNTNAP2* in molteplici funzioni cognitive e relative al linguaggio.

CNTNAP2 codifica per una proteina di adesione che si pensa abbia un importante ruolo nella migrazione neuronale, la secrezione e raggruppamento dei canali ionici nei nodi di Ranvier (Rodenas-Cuadrado et al., 2014).

DIP2A: Nel 2009 Poelmans e colleghi hanno identificato il gene *DIP2A* che era parzialmente deletato sulla regione del cromosoma 21q22.3 e che co-segrega con la dislessia. Il gene *DIP2A* codifica per la proteina 2 omologo A disco-interagente che è una proteina nucleare e altamente espressa nel cervello umano e può giocare un ruolo nel modellamento assonale nel sistema nervoso centrale.

Dalle analisi bioinformatiche si sa che *DIP2A* umana è ritenuta una molecola recettrice di tipo I con tre domini, contenente la proteina che lega DMAP associata alla DNA metiltransferasi I, CaiC e i domini enzimatici AMP-binding. *DIP2A* lega il complesso di trascrizione DMAP1-DNMT1-HDAC2 nei foci di replicazione del DNA (Rountree et al., 2000; Bai et al., 2006; Guan et al., 2009).

Collettivamente, il gene *DIP2A* è probabilmente implicato nella patogenesi molecolare della dislessia con la sua funzione nella protuberanza assonica, nella migrazione cellulare e nella plasticità sinaptica. Quest'ultima risulta cruciale durante i processi di apprendimento e memoria che sono deficitari nella dislessia (Swanson, 2011).

S100B: Come per *DIP2A*, lo stesso gruppo di ricerca ha identificato *S100B*, situato nella regione 21q22.3 (Poelmans et al., 2009). Successivamente, Poelmans e colleghi (2011) hanno proposto la partecipazione della proteina codificata da questo gene in un network molecolare coinvolto nella migrazione neuronale assieme ad altri geni candidati: *ROBO1*, *KIAA0319*, *KIAA0319L*, *S100B*, *DOCK4*, *FMRI*, *DIP2A*, *GTF2I*, *DYX1C1* and *DCDC2*.

Il gene *S100B* è principalmente e altamente espresso nel cervello, specialmente nell'ippocampo (Rothermundt et al., 2003). *S100B* codifica un peptide che lega il calcio che è principalmente prodotto dagli astrociti e che esercita degli effetti paracrini e autocrini sui neuroni e sulle cellule gliali. Inoltre, è noto che il legame di *S100B* al recettore per i prodotti finali della glicosilazione avanzata (RAGE) nei neuroni Ippocampali, attiva una *pathway* di segnalazione che risulta nella protuberanza assonica (Huttunen, 2000). La proteina è stata riportata per essere coinvolta nella modulazione della plasticità sinaptica (Nishiyama, 2002).

Nello studio di Matsson e colleghi (2015) è stata inoltre trovata una variante al 3'-UTR, ovvero rs9722, in associazione significativa con lo spelling. Le varianti di sequenza al 3'-UTR possono colpire la stabilità del trascritto, per esempio creando o distruggendo siti target per dei microRNAs (miRNAs) (Clop, 2006; Chin, 2008). Perciò sono stati cercati dei siti target per presunti miRNA che si sovrapponevano a rs9722 nelle predizioni TargetScan. Infatti, rs9722 era localizzato in siti target multipli predetti di miRNA.

1.4.5 Il ruolo delle *Copy Number Variation* (CNV)

Nei geni precedentemente descritti sono stati effettuati diversi studi di associazione tra tratti di lettura o del linguaggio e SNPs. Tuttavia, queste associazioni genetiche possono spiegare solo una piccola parte dell'ereditabilità dei tratti di lettura e del linguaggio (Peterson e Pennington, 2012; Newbury e Monaco, 2010; Graham e Fisher, 2015). Per tale motivo, una parte della mancata individuazione delle cause genetiche è stata attribuita alle CNV, che sono delle regioni larghe di più di 1 kb e che presentano uno stato non diploide.

Diverse CNV sono state individuate in disordini di neurosviluppo e psichiatrici, come schizofrenia, autismo e disabilità intellettuale (Grayton et al., 2012; Malhotra e Sebat, 2012) ma esistono anche diversi studi riguardo le CNV individuate nei soggetti con dislessia o SLI, come riportato sotto.

In 10 famiglie dislessiche indiane sono state identificate *de novo* CNV in diversi loci: *GABARAP* (17p13.1), *NEGR1* (1p31.1), *ACCN1* (17q11.21), *DCDC5* (11p14.1). Una CNV *de novo* è stata trovata anche nel locus SLI4 che coinvolgeva il gene candidato *CNTNAP2* (Veerappa et al., 2013a). Nelle stesse famiglie è stato identificato e proposto come gene candidato il *PCDH11X*, situato nel cromosoma X (Xq21.-q21.32) (Veerappa et al., 2013b). In una famiglia tedesca Poelmans e colleghi (2009) hanno identificato una delezione in eterozigosi nella regione 21q22.3 che co-segregava con la dislessia e conteneva i geni *PCNT*, *DIP2A*, *S100B* e *PRMT2*.

In uno dei più estesi studi finora effettuati, gli autori, hanno preso in esame 376 casi di dislessici e 337 controlli. Girirajan e colleghi (2011) hanno identificato diverse CNV, all'interno di uno dei loci di suscettibilità all'autismo (7q11.22), dove sovrappongono i geni *IMMP2L* e *AUTS2*.

Riguardo le difficoltà di apprendimento del linguaggio, Wisznieski e colleghi (2013) hanno identificato una delezione in eterozigosi che interrompeva il gene *TM4SF20* e che co-segregava con il ritardo del linguaggio in 15 famiglie asiatiche. In un'analisi di CNV di famiglie con SLI è stata identificata una piccola delezione esonica all'interno del gene *ZNF277*, situato nel locus 7q31.1 (Ceroni et al., 2014). In uno studio di *genome-wide CNV*, che hanno preso in considerazione 127 casi di SLI, hanno riportato delle *de novo* CNV come nuovi candidati, che interrompevano i geni *ACTR2* (2p14), *CSNK1A1* (5q33.1) e delle regioni coinvolte nella sindrome di duplicazione del 22q11.2 e 8p23.1 (Simpson et al., 2015). Recentemente, uno screening di CNV in una coorte longitudinale di bambini con difficoltà relative al linguaggio,

ha rivelato una *de novo* delezione nella regione 15q13.1-q13.3. Questa delezione era già stata riportata in casi di ritardo del neurosviluppo (Pettigrew et al., 2015).

Altre CNV sono state associate con problemi di lettura e del linguaggio nel contesto di comorbidità di altri disordini. Una delezione che interrompe il gene *DOCK4* e *IMMP2L* (7q31.3) co-segregava con difficoltà di lettura in una famiglia con due casi di autismo, come anche una delezione esonica a co-segregare con difficoltà di lettura in un'altra famiglia (Pagnamenta et al., 2010). Microdelezioni nella regione 16p11.2, di solito implicate nelle difficoltà cognitive, nel ritardo del sviluppo, problemi del linguaggio e autismo, sono state associate con dei casi aprassia verbale in due studi indipendenti (Raca et al., 2013; Fedorenko et al., 2016). Pazienti Prader-Willi/Angelman, che presentavano la duplicazione della regione 15q11.2, sono stati frequentemente riportati a presentare dei problemi di ritardo verbale e del linguaggio (Burnside et al., 2011). Analogamente, soggetti con la microdelezione del 2p15-p16.1 che tipicamente mostrano disabilità cognitive, linguistiche e psichiatriche, è stata identificata una microdelezione *de novo* che coinvolgeva il gene *BCL11A* dove il paziente mostrava aprassia, disartria e ritardo del linguaggio (Peter et al., 2014).

Recentemente, Stefansson e colleghi (2014) studiando l'effetto di diverse CNV, precedentemente associate alla schizofrenia o autismo, su diversi tratti cognitivi in una larga coorte di Islandesi ha riportato effettivamente un effetto di certe CNV sulle abilità cognitive. In particolare, la delezione della regione 16p11.2 e la duplicazione della regione 22q11.21 sono state associate a fluency delle lettere, mentre la delezione della regione 15q11.2 con dislessia e discalculia.

Infine, nel 2016 Gialluisi e colleghi hanno studiato gli effetti delle CNV sulle performance di lettura e del linguaggio. Queste analisi hanno riportato un'associazione con delle CNV che si sovrappongono con i geni *CNTN4*, *CTNNA3*, *CHRNA7* e *ZNF737*.

1.5 Tecnologie di indagine genetica ad alta processività

1.5.1 SNP-array

L'analisi dei polimorfismi a singolo nucleotide basata su *microarray* costituisce attualmente la tecnologia più avanzata di analisi citogenetica. Essa è impiegata per la diagnosi di anomalie cromosomiche quantitative, cioè di CNV in forma di delezioni, duplicazioni o amplificazioni. Inoltre, permette di identificare le disomie uniparentali (UPD, Uniparental Disomy), ovvero la presenza di due cromosomi omologhi provenienti da un solo genitore, come anche le *Loss of*

Heterozygosity (LOH), cioè regioni cromosomiche identiche determinate generalmente dalla consanguineità dei genitori.

La tecnica dello *SNP-array* si basa sulla microibridazione contemporanea di migliaia di specifici frammenti di DNA. Un *microarray* di DNA è costituito da un insieme di sonde di DNA, attaccate ad una superficie solida (array), che sfruttano l'ibridazione inversa (Etebari et al., 2015).

L'analisi dei dati ottenuti si basa su due misurazioni:

B allele frequency (BAF): fornisce la composizione allelica, ovvero, la proporzione del campione ibridizzato che porta l'allele B rispetto all'allele A in un determinato locus (Figura 1.2).

Generalmente in una situazione normale (CN=2), la frequenza dell'allele B assume 3 valori: BAF, un valore compreso tra 0 e 1 rappresenta la proporzione del contributo dell'allele B sul totale del contributo degli alleli e si calcola con $BAF: NB/(NA+NB)$, dove NA e NB sono rispettivamente il numero degli alleli A e B. Per un eterozigote AB il valore sarà dunque 0,5 ($1/1+1$), l'omozigote BB avrà come valore 1 ($2/2$), infine l'omozigote AA sarà 0 ($0/2$) (Priit, 2015).

Log R ratio (LRR) è il logaritmo del rapporto tra l'intensità del segnale di ibridizzazione osservata rispetto all'intensità attesa, per cui ogni deviazione dal valore 0 è indice di CNV (Figura 1.2). L'aumento del LRR rispetto al valore base risulta dall'incremento dell'intensità del segnale di una regione che presenta un ampliamento del numero di copie (duplicazione o amplificazione). La delezione è raffigurata come una riduzione del segnale nel plot LRR, a un valore medio di <0,45. Invece, la duplicazione del numero di copie (CN=3) è relativa a un $\text{Log R ratio} = \log_2 (\text{intensità osservata}/\text{intensità attesa}) = \log_2 (3/2) = 0,58$ (Priit, 2015).

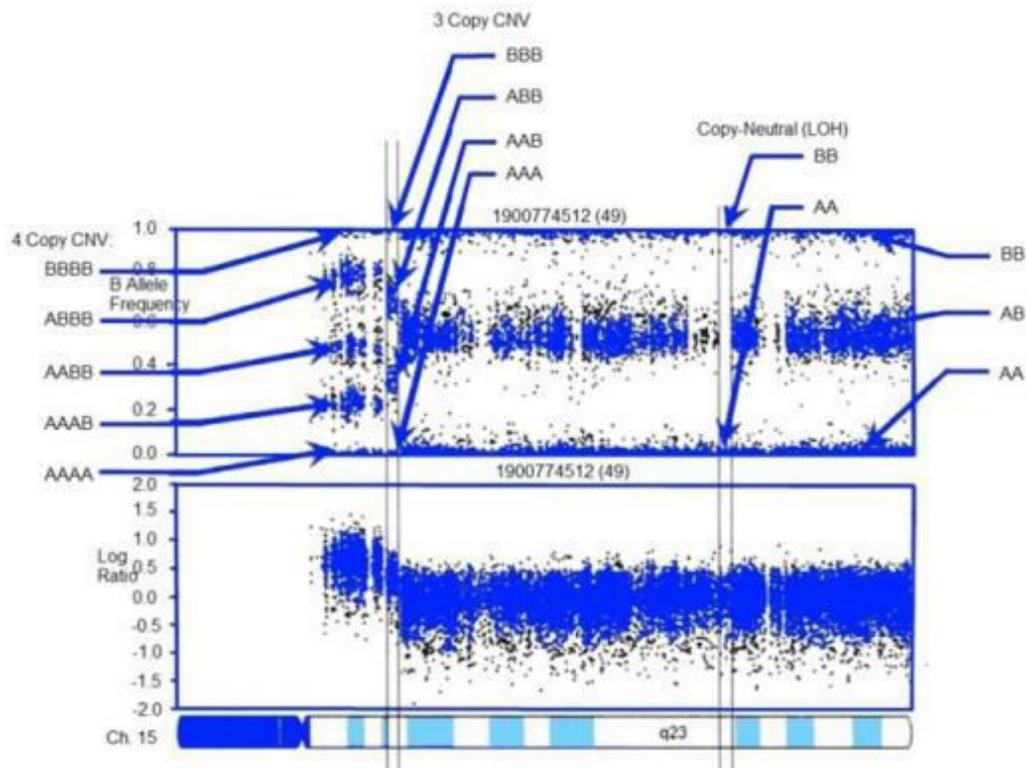


Figura 1.2: Esempio di valori LRR e BAF sul cromosoma 15 in relazione al numero di copie e in assenza di eterozigotà (LOH).

1.5.2 Targeted Next Generation Sequencing

Negli anni 2000 nasce la metodologia di sequenziamento di “seconda generazione”, *Next Generation Sequencing*, come successore del *Sanger Sequencing* (Figura 1.3). Caratteristica principale di questa metodica è l’alta processività, dove si generano parallelamente e in modo esponenziale miliardi di basi di DNA. L’NGS, oltre a permettere l’acquisizione di dati del genoma (Genome-Seq, Exome-Seq, Targeted-Seq), si è estesa anche in altri settori, come lo studio dell’intero trascrittoma e altro (RNA-Seq, miRNA-Seq, ecc.).

Nel presente studio si sono effettuati degli studi di targeted-NGS mediante l’utilizzo dell’*Ion Torrent PGM (Personal Genome Machine)* della Life Technologies (Thermo Fisher Scientific). Tale sistema sfrutta la tecnologia dei semiconduttori coniugata al principio biochimico, in base al quale, quando un singolo nucleotide è incorporato dall’enzima DNA polimerasi nella catena nascente di DNA, viene rilasciato uno ione idrogeno come sottoprodotto (Figura 1.3).

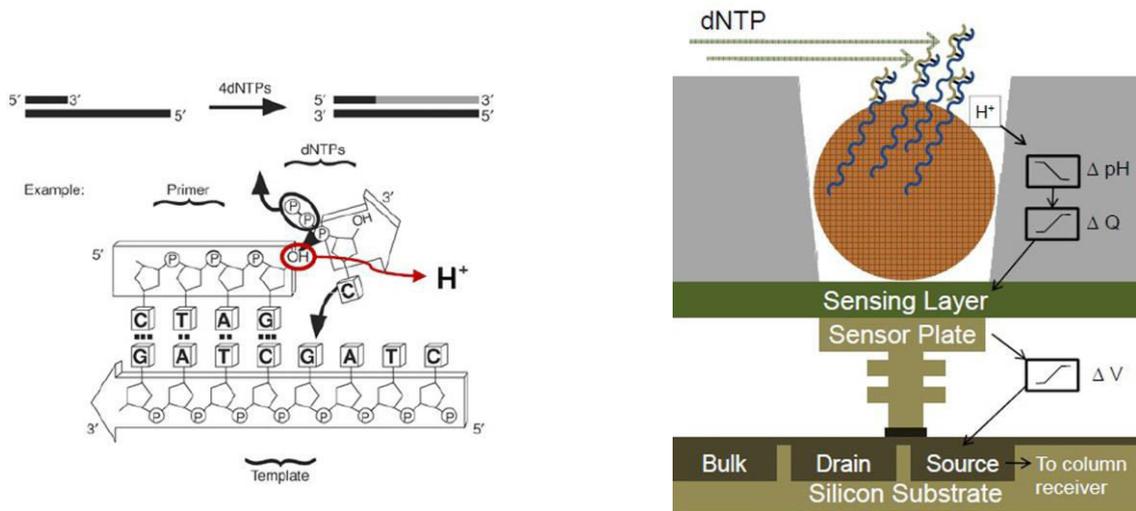


Figura 1.3: A sinistra viene raffigurata la reazione di incorporazione di un dNTP nella catena nascente del DNA, che comporta anche il rilascio di un protone. A destra, viene illustrata una cella di reazione dove il protone rilasciato dall'incorporazione del nucleotide provoca una variazione del pH dell'ambiente circostante. Tale variazione viene rilevata da dei sensori e il segnale viene interpretato nella base incorporata.

2. SCOPO DELLA TESI

La presente ricerca si focalizza nella messa a punto e nell'utilizzo di un protocollo di indagine genetica, basato su tecniche innovative come il *Next Generation Sequencing* e *SNP-array*, per esplorare la componente genetica della dislessia in una coorte di ragazzi dislessici.

In dettaglio gli obiettivi della presente tesi sono:

Obiettivo 1.

A: La selezione e il reclutamento di una serie di ragazzi con dislessia e con dislessia associata ad altri DSA, presenti nel territorio del Friuli Venezia Giulia, a partire da una lista di ragazzi diagnosticati con DSA presso l'ospedale pediatrico IRCCS materno infantile "Burlo Garofolo";

B: La creazione di una bio-banca di campioni biologici costituiti da saliva e DNA;

Obiettivo 2.

A: Lo sviluppo di uno metodo d'indagine genetico mediante tecnologia NGS. Questo obiettivo comprende la proposta di un pannello di geni associati alla dislessia e la sua valutazione tecnica;

B: L'identificazione di varianti a carico della coorte e la loro caratterizzazione; In questa analisi si è focalizzati sulle varianti nuove e rare come ad esse predette con un effetto deleterio o altamente patogenetico.

Obiettivo 3.

A: La valutazione tecnica dell'utilizzo di *SNP-array* ad alta densità (> 1.7 M marcatori) per l'identificazione delle CNV;

B: L'identificazione di CNV nella coorte di studio e la loro caratterizzazione.

3. MATERIALI E METODI

3.1 Coorte di studio

Uno dei centri di diagnosi dei DSA nella regione del Friuli Venezia Giulia è l'ospedale IRCCS Burlo Garofolo. In particolare, tutte le diagnosi vengono effettuate presso l'unità di Neuropsichiatria Infantile (NPI) secondo i criteri stabiliti dalla “*Consensus Conference (2007)*” e dalla “*International Classification of Diseases, ICD-10*” (Tabella 3.1).

Disturbo	Classificazione
Dislessia	F81.0
Disortografia	F81.1
Discalculia	F81.2
Disturbi misti delle abilità scolastiche	F81.3
Disgrafia	F81.8
Disturbi dello sviluppo delle abilità scolastiche, non specifici	F81.9

Tabella 3.1: Classificazione dei disturbi dello sviluppo delle abilità scolastiche secondo la “*International Classification of Diseases*”.

In collaborazione con l'unità di NPI e la struttura complessa di Genetica Medica del medesimo ospedale, previa approvazione del progetto dal “Comitato Etico Indipendente” con numero di protocollo 5/2015dd, è stata avviata la selezione e il reclutamento dei partecipanti.

Per il presente progetto sono stati consultati gli archivi del NPI e selezionate 94 famiglie. In totale, sono stati reclutati 101 soggetti, divisibili in due sotto-gruppi: 49 soggetti con sola dislessia e 52 soggetti con dislessia e altre comorbidità (Tabella S1, Capitolo 6 - Tabelle Supplementari). Il rapporto tra femmine e maschi è di 1:1,7, mentre la diagnosi dei DSA è avvenuta in una età degli partecipanti tra gli 8 e i 14 anni.

A fini legali e alla tutela della privacy, a ogni famiglia sono stati consegnati i seguenti moduli:

- L'informativa relativa al progetto di ricerca;
- L'informativa e il modulo di assenso, se il partecipante era un adolescente;
- L'informativa per il trattamento dei dati;
- Il consenso informato.

3.2 Bio-banca

Assieme all'avviamento del reclutamento dei soggetti è stata allestita una banca dati dedicata a questo progetto. Sono stati raccolti tutti i dati relativi alle diagnosi e necessari per le successive analisi e caratterizzazione dei campioni e a ogni soggetto partecipante è stato assegnato un codice (DDxxx, o DDxxxA se fratello).

Da ogni partecipante sono stati raccolti 2 mL di saliva utilizzando i collettori Norgen (Norgen Biotek Corporation, Ontario, Canada), dotati di soluzione di conservazione del campione. L'estrazione del DNA è stata effettuata utilizzando il sistema automatizzato QIASymphony con i kit QIASymphony DSP DNA Midi Kit (Qiagen, Hilden, Germany). La qualità e la quantità del DNA è stata valutata con lo spettrofotometro NanoDrop 1000 (Thermo Fisher Scientific, Waltham, MA, USA). Per l'analisi qualitativa sono stati considerati i rapporti di assorbanza: 260/280 per la purezza del campione dalle proteine; e 260/230 per la purezza del campione da polisaccaridi e fenoli.

3.3 Targeted Next Generation Sequencing

Tutti e 101 campioni sono stati sequenziati utilizzando la piattaforma *Ion Torrent PGM* della Life Technologies. La procedura consiste di 3 principali passaggi:

- Costruzione di un pannello di geni AmpliSeq;
- Preparazione delle librerie;
- Sequenziamento.

3.3.1 Pannello dei geni candidati

Basati sulla letteratura sono stati selezionati 12 geni candidati di suscettibilità alla dislessia. Tutti i geni selezionati erano stati replicati in almeno due studi indipendenti (vedi introduzione). In particolare, 7 dei geni sono localizzati all'interno degli *DYX1-9* locus (*CYP19A1*, *DCDC2*, *DYX1C1*, *GCFC2*, *KIAA0319*, *KIAA0319L*, *MRPL19*, *ROBO1*), mentre gli altri 5 (*CMIP*, *CNTNAP2*, *DIP2A*, *S100B*) sono dei nuovi geni candidati alla dislessia ma anche associati in generale con i disturbi del linguaggio (Tabella 3.2).

Locus	Cito-banda	Gene(i)
DYX1	15q21	<i>DYX1C1</i> , <i>CYP19A1</i>
DYX2	6p22.3-p21.3	<i>DCDC2</i> , <i>KIAA0319</i>
DYX3	2p12-p16	<i>MRPL19</i> , <i>GCFC2</i> (<i>C2ORF3</i>)
DYX5	3p12-q13	<i>ROBO1</i>
DYX8	1p36-p34	<i>KIAA0319L</i>
SLI1	16q23.1-q24	<i>CMIP</i>
SLI4	7q35-q36.1	<i>CNTNAP2</i>
-	21q22.3	<i>DIP2A</i> , <i>S100B</i>

Tabella 3.2: Lista dei geni analizzati mediante targeted-NGS. DYX: *Dyslexia susceptibility locus*; SLI: *Specific Language Impairment susceptibility locus*.

Utilizzando il programma *AmpliSeq Designer* (versione 4.4.2) sono stati disegnati due pool di *primers* (Pool-1, Pool-2) per il sequenziamento di tutti gli esoni codificanti e le giunzioni esone-introne, fino a 10 basi all'interno degli introni (Tabella 3.3).

Gene	Ampliconi	Basi totali (bp)	Basi coperte (bp)	Basi mancanti (bp)	Copertura totale
<i>CMIP</i>	33	2780	2748	32	99%
<i>CNTNAP2</i>	36	4476	4448	28	99%
<i>CYP19A1</i>	13	1692	1692	0	100%
<i>DCDC2</i>	14	1631	1631	0	100%
<i>DIP2A</i>	57	5605	5535	70	99%
<i>DYX1C1</i>	15	1562	1443	119	92%
<i>GCFC2</i>	28	2715	2715	0	100%
<i>KIAA0319</i>	31	3667	3667	0	100%

Gene	Ampliconi	Basi totali (bp)	Basi coperte (bp)	Basi mancanti (bp)	Copertura totale
<i>KIAA0319L</i>	27	3550	3550	0	100%
<i>MRPL19</i>	10	999	999	0	100%
<i>ROBO1</i>	47	5660	5660	0	100%
<i>S100B</i>	2	319	319	0	100%

Pool-1 = 179 ampliconi; Pool-2 = 172 ampliconi

Tabella 3.3: Riassunto delle caratteristiche del pannello di geni di tipo AmpliSeq.

3.3.2 Costruzione delle librerie

Per la costruzione delle due librerie, per ogni campione da sequenziare e per un totale di 101 campioni, viene utilizzato il protocollo “*Ion AmpliSeq DNA Library Preparation User Guide*”. Tutta la preparazione consiste di 5 fasi (Figura 3.1):

- Amplificazione delle regioni target;
- Digestione enzimatica degli ampliconi;
- Ligazione degli adattatori;
- Purificazione della libreria;
- Quantificazione della libreria mediante *Real-Time* PCR.

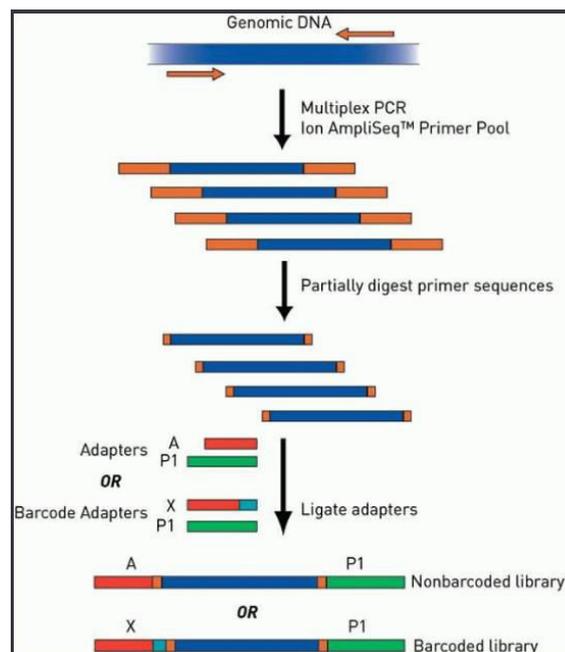


Figura 3.1: Rappresentazione schematica dei vari passaggi necessari per la costruzione delle librerie.

L'amplificazione delle regioni target avviene mediante due reazioni di *multiplex*-PCR, per ogni campione:

Allestimento delle reazioni di *multiplex*-PCR, per 1 DNA

2,0 µL *ReadyMix* 5X
5,0 µL Pool-1 di *primer* 2X
3,0 µL gDNA, 5.0 ng/µL (totale 15.0 ng)

10,0 µL TOTALE

2,0 µL *ReadyMix* 5X
5,0 µL Pool-2 di *primer* 2X
3,0 µL gDNA, 5.0 ng/µL (totale 15.0 ng)

10,0 µL TOTALE

Considerando il numero dei diversi ampliconi da generare da ogni reazione (Pool-1 = 179 ampliconi; Pool-2 = 172 ampliconi) si decide in base alle linee guida il numero di cicli della *multiplex*-PCR:

Ciclo termico

99°C x 2 min
[(99°C x 15 sec) (60°C x 5 min)] X 21 cicli
8°C x ∞

I cicli inizialmente indicati dalla guida sono 18, a questi abbiamo aggiunto 3 cicli in tal modo da aumentare la resa finale delle *multiplex*, in caso di difficoltà di amplificazione di qualche DNA, come in precedenza osservato.

Finita la *multiplex*-PCR, le due librerie vengono unite e poi aggiunti 2,0 µL di enzima FuPa:

Ciclo termico

50°C x 20 min
55°C x 20 min
60°C x 20 min
8°C ∞

Nel passaggio successivo vengono aggiunti l'adattatore universale e il *barcode* specifico per ogni libreria:

Allestimento della reazione di ligazione, per 1 libreria

4,0 μ L Switch solution
0,6 μ L P1, adattatore universale
1,0 μ L H₂O

5,6 μ L x libreria
 + 0,6 μ L *Barcode*
 + 2,0 μ L Ligasi

30,0 μ L TOTALE

Ciclo termico

22°C x 30 min
68°C x 5 min
72°C x 5 min
8°C x ∞

Alla fine della costruzione delle librerie esse vengono purificate secondo il protocollo qui sotto descritto:

- Aggiungere 45,0 μ L di AgentCourt AMPure a ciascun provetta dopo aver vortexato bene la bottiglietta.
- Settare la pipetta a 35,0 μ L e pipettare 10 volte.
- Incubare per 5 minuti a temperatura ambiente.
- *Mettere la provetta sul magnete per 4 minuti.*
- Eliminare il supernatante.
- Aggiungere 150,0 μ L di etanolo 70% fresco e attendere per 15 secondi.
- Eliminare il supernatante.
- Aggiungere 150,0 μ L di etanolo 70% fresco e attendere per 15 secondi.
- Eliminare il supernatante, ri-eliminare eventuali residui.
- *Togliere la provetta dal magnete e asciugare per 3 minuti.*
- Aggiungere 40,0 μ L di LowTE.
- Attendere 3 minuti, poi settare la pipetta a 25,0 μ L e pipettare 10 volte.
- Attendere 2 minuti.

- Vortexare 2 secondi.
- *Mettere la provetta sul magnete per 4 minuti.*
- Recuperare il supernatante (settare la pipetta a 35,0).

Una volta le librerie pronte, esse vengono quantificate mediante *Real-Time* PCR e utilizzando il kit “*Kappa Library Quantification Kit*”. In questo passaggio vengono preparati 4 “standard” per creare la Standard Curve: St.2 (1E7), St.3(1E6), St.4(1E5), St.5(1E4). Tutte le reazioni vengono effettuate in duplicato.

Allestimento della reazione di *Real-Time* PCR, per 1 campione e 1 replica

6,0 µL	<i>ReadyMix</i> 2X
2,0 µL	H ₂ O
2,0 µL	Libreria (diluita 1:500) o Standard o H ₂ O

10,0 µL	TOTALE

Ciclo termico

95°C x 5 min
 [(95°C x 30 sec) (60°C x 1 min)] X 32 cicli

In base alla quantità (pM) della libreria ottenuta, viene aggiudicata accettabile o meno per essere sequenziata.

500-100 pM	Ottimale
100-50 pM	Buona
50-30 pM	Accettabile
<30 pM	Non idonea

Prima del sequenziamento, le librerie selezionate vengono portate a 10-12pM e creati dei pool di più librerie.

3.3.3 Sequenziamento NGS

I pool di librerie vengono in seguito processate e sequenziate mediante una serie di strumenti della Life Technologies:

PCR in emulsione: questa reazione permette un'amplificazione clonale dei singoli ampliconi nelle *Ion Sphere Particle* (ISP). La reazione viene effettuata utilizzando lo strumento *Ion One Touch OT2* e il kit *Ion PGM Hi-Q OT2*. Questa emulsione viene preparata miscelando l'*Ion One Touch Reaction Oil* e l'*Amplification solution* e poi viene caricata su di una colonnina con alla base un filtro attraverso cui vengono generate microscopiche "gocce di reazione" che, nelle condizioni ideali, dovrebbero contenere una singola sfera e un singolo amplicone in modo da ottenere un'amplificazione monoclonale. Tali gocce attraversano una piastra di amplificazione dove avviene la PCR in emulsione. Le condizioni di reazione prevedono 60 cicli di PCR a due differenti gradienti di temperatura (95°C e 64°C).

Enrichment: al termine dell'amplificazione, tramite l'*Enrichment system* vengono selezionate solamente le sfere su cui ciascun amplicone si sarà legato e amplificato. Questa selezione avviene tramite l'utilizzo di biglie magnetiche di streptavidina che hanno affinità per le estremità biotinilate degli ampliconi. Una fase di denaturazione che avviene tramite un lavaggio con NaOH provvede a staccare le biglie di streptavidina dalle ISP amplificate.

Sequenziamento: è stato effettuato utilizzando lo strumento *Ion Torrent PGM (Personal Genome Machine, Life Technologies)*. Prima del caricamento del campione è prevista una fase di lavaggio del *Ion Torrent PGM* con una soluzione contenente clorito e successivamente una fase di inizializzazione dello strumento necessaria per la taratura del pH delle soluzioni utilizzate per il sequenziamento. Per questo studio sono stati utilizzati degli *Ion 316 Chip v2* costituiti da 6 milioni di pozzetti e in grado di produrre fino a 100 Mb.

3.4 Annotazione e prioritizzazione delle varianti

Durante la corsa tutti i dati vengono archiviati e analizzati sull'*Ion Torrent Browser* (Thermo Fisher Scientific). A fine della corsa le sequenze generate (reads) sono allineate al genoma umano (versione GRC37/hg19) e avviati due *plugin*: "*Coverage Analysis*" e "*Variant Caller*".

3.4.2 Prioritizzazione delle varianti

Per la prioritizzazione delle varianti sono stati effettuati due principali passaggi consecutivi:

- A) Selezione delle varianti con una MAF < 0.01 o non riportata;
- B1) Sotto selezione dal gruppo A delle varianti con un effetto deleterio (*stopgain*, *frameshift*, *indel*, *canonical splicing*);
- B2) Sotto-selezione dal gruppo A di varianti situate tra +/- 3-10 basi all'interno dell'esone o dell'introne.
- B3) Sotto selezione dal gruppo A delle varianti non-sinonime predette altamente patogenetiche da due dei tre predittori selezionati (vedi sotto).

Per filtrare e ottenere le varianti con una MAF < 0.01, è stato utilizzato il database gnomAD exome (Lek et al., 2016). Ove possibile, per le varianti che non erano presenti dati sulle frequenze è stato consultato il gnomAD genome database (Tabella 3.4).

Le varianti deleterie e quelle che erano localizzate nel *canonical splicing site* sono state direttamente candidate per essere validate con il metodo Sanger. Le varianti situate fino a 10 basi all'interno degli esoni o tra le +/- 3-10 basi all'interno degli introni, sono state classificate in base ai *tools* Spidex (Xiong et al., 2015) e scSNV (Jian et al., 2014). Le varianti che erano predette patogenetiche da entrambi i *tools* sono state poi verificate con il *tool* HSF (Desmet et al., 2009) (Tabella 3.4).

Per classificare le varianti non-sinonime secondo la loro patogenicità, basato sulla letteratura sono stati utilizzati i *tools* CADD (Kircher et al., 2014), M-CAP (Jagadeesh et al., 2016) e REVEL (Ioannidis et al., 2016). Il CADD è un classificatore di patogenicità molto “popolare” e usato da diversi autori. Il M-CAP presenta una percentuale di “*Misclassified known pathogenic variants*” pari al 5% (usando un cut-off di 0,025), rispetto al CADD che è stimata al 26% (usando un cut-off di 20). Il REVEL ha le migliori prestazioni complessive rispetto ai CADD, DANN (Quang et al., 2015) ed EIGEN (Ionita-Laza et al., 2016). Sono state considerate come varianti altamente patogenetiche quelle che superavano i cut-off di almeno due dei tre *tools* considerati (Tabella 3.4).

Minor Allele Frequency*	
gnomAD exome	<0.01
gnomAD genome	<0.01
Splicing alternativo	
Spidex	≤ -0.4 or ≥ 0.4
scSNV	≥ 0.6
Varianti non-sinonime	
CADD	≥ 20
REVEL	≥ 0.4
M-CAP	≥ 0.025

Tabella 3.4: Lista dei cut-off utilizzati durante la selezione delle varianti.

3.5 Sequenziamento diretto (*Sanger sequencing*)

Per confermare le varianti priorizzate dall'analisi dei dati NGS è stato utilizzato il metodo di sequenziamento diretto (Sanger), che è la tecnica più diffusa e affidabile per la conferma delle varianti identificate mediante NGS. Questa tecnica consiste dei seguenti 6 passaggi:

Disegno dei primers: una copia di *primer* per variante (varianti localizzate in regioni diverse) è stata disegnata utilizzando l'on-line *tool* Primer3 (<http://bioinfo.ut.ee/primer3-0.4.0/>) (Tabella 3.5). Le principali impostazioni riguardavano la lunghezza dell'amplicone (300-600 bp), la lunghezza del *primer* (20 - 25 basi), la temperatura di melting (58°C - 64°C) e la percentuale di GC del *primers* (40% - 60%).

Primer	Regione	Primer Forward	Primer Reverse
NGS_CMIP_01	chr16:81711917-81711917	atgaacggtgcctgtactcatgtg	atgaggcagtcgatgaggttg
NGS_CMIP_02	chr16:81736276-81736276	atcacccctccagtaaaccacct	ggaagagctggcgtgtaagta
NGS_CNTNAP2_01	chr7:147844744-147844744	ttgtaggacgtgacaggcttaga	atcaacagaggaagccatggt
NGS_CNTNAP2_02	chr7:147926780-147926780	tctaagagcaggaattgaggg	cctcaaaaacaaaccaatggccaagc
NGS_DIP2A_01	chr21:47983840-47983840	gcctctaaactcctgacctgac	atcctacataggaactgggtctggg
NGS_DIP2A_02	chr21:47980687-47980687	tggatggaagcatcagactgcc	tgtcctctcatgtcccaccca
NGS_DYX1C1_01	chr15:55731682-55731682	cttctggctcaagttatcatcc	tcaatccactggaggaagatggaa
NGS_GCFC2_01	chr2:75933779-75933779	gcctcctccctttctcattgca	gagcaattttggcactctgacaatcac
NGS_KIAA0319_01	chr6:24564533-24564540	cacatctgtcagcgtaaagacc	ggtaattcctggaaagtgcttgg

Primer	Regione	Primer Forward	Primer Reverse
NGS_KIAA0319_02	chr6:24596673-24596673	ttcagcatcatgtcccatagtc	gtgtgcccgaagcagtgagc
NGS_KIAA0319L_01	chr1:35944646-35944646	cctaaccacagacctgactgca	cacagtattcaagttatgcctccc
NGS_ROBO1_01	chr3:78700915-78700915	gtgtcacctgaggaccaagtca	tggataccgcagcacaataaacgc
NGS_ROBO1_02	chr3:78683121-78683121	gagatgtgaatcgctcatgaagag	agagacaagacacatgcctgagac
NGS_ROBO1_03	chr3:78717396-78717396	cattctctgctacggctgcccagc	ccagtgggaagcaacatggagtg
NGS_ROBO1_04	chr3:78987753-78987753	ctgaactataggcagcacatggttcc	gacaaagatgacctcgctcacacc

Tabella 3.5: *Primer* utilizzati per la conferma delle varianti rilevate mediante NGS.

PCR: ogni copia di *primers* è stata utilizzata per effettuare una *Polymerase Chain Reaction* (PCR) usando come template il DNA dei pazienti di interesse e secondo le istruzioni del kit *KAPA2G Fast ReadyMix PCR Kit* (Kapa Biosystems).

Allestimento delle reazioni di PCR, per 1 DNA

12,5 µL *ReadyMix 2X*

4,0 µL *primer Forward* (2 uM)

4,0 µL *primer Reverse* (2 uM)

4,5 µL H₂O sterile / gDNA, 25,0 ng/µL

25,0 µL TOTALE

Ciclo termico

96°C x 3 min

[(95°C x 15 sec) (T_a°C x 10 sec) (72°C x 15 sec)] X 30 cicli

4°C ∞

Purificazione PCR: in seguito alla PCR, 4,5 µl di prodotto di amplificazione è stato digerito mediante l'utilizzo di 1,5 µl della Endonucleasi I ExoStar (GE Healthcare). Questo passaggio mira i *primers* i quali vengono digeriti in tal modo da non influenzare i passaggi successivi. Le reazioni sono state incubate nel termociclatore con un profilo termico di 37°C per 20 minuti e 85°C per 15 minuti.

Marcatura: i prodotti di digestione sono sottoposti alla reazione di sequenziamento Sanger al fine di ottenere frammenti della regione target marcati e di lunghezza *random*. La reazione viene effettuata utilizzando 0,5 µl BigDye® (BigDye terminator v3.1 cycle sequencing kit,

Thermo Fisher Scientific), 2 µl di buffer 5X e 3 µl di H₂O sterile, 3 µl del prodotto di digestione, 2 µl del *primer* Forward (2 uM) e 2 µl del *primer* Reverse; ogni reazione di marcatura è sottoposta al seguente profilo termico:

Ciclo termico

96°C x 3 min

[(96°C x 30 sec) (Ta°C x 15 sec)] X 26 cicli

8°C x ∞

Purificazione del prodotto di sequenza: terminata la marcatura e prima della corsa dei campioni (vedi sotto) i prodotti marcati vanno purificati aggiungendo nella provetta 9 µl di resina e 41 µl di soluzione SAM del kit *BigDye X Terminator Purification kit* (Thermo Fisher Scientific). I campioni vanno agitati per almeno 30 minuti in tal modo da sequestrare/eliminare l'eccesso di BigDye, non incorporato nella precedente reazione.

Elettroforesi capillare e analisi dei dati: l'elettroforesi capillare è stata effettuata utilizzando il sequenziatore 3500 DX Genetic Analyzer (Thermo Fisher Scientific). I dati ottenuti sono stati visualizzati con il software Chromas.

3.6 Analisi dei campioni mediante *SNP-array*

Un totale di 55 campioni, provenienti da 45 famiglie e composti da 19 femmine e 31 maschi (Tabella S1, Capitolo 6 - Tabelle Supplementari), selezionati in base all'albero genealogico e la qualità del DNA, sono stati ulteriormente analizzati mediante l'utilizzo di *SNP-array* con il design MultiEthnic-Global Array (MEGA), (Illumina Inc., San Diego, CA). I marcatori in tutto sono più di 1.700.000 e la risoluzione mediana della spaziatura tra i marcatori è di 0,78 kb. Questo array permette un'analisi in parallelo di 8 campioni di DNA (Tabet et al., 2015).

Per lo svolgimento di questa analisi è stato utilizzato il DNA genomico dei campioni pre-selezionati a una concentrazione di 50 ng/µl (intervallo tra 30-70 ng/µl). Il protocollo utilizzato è costituito da 9 steps come riportati nella Figura 3.2.

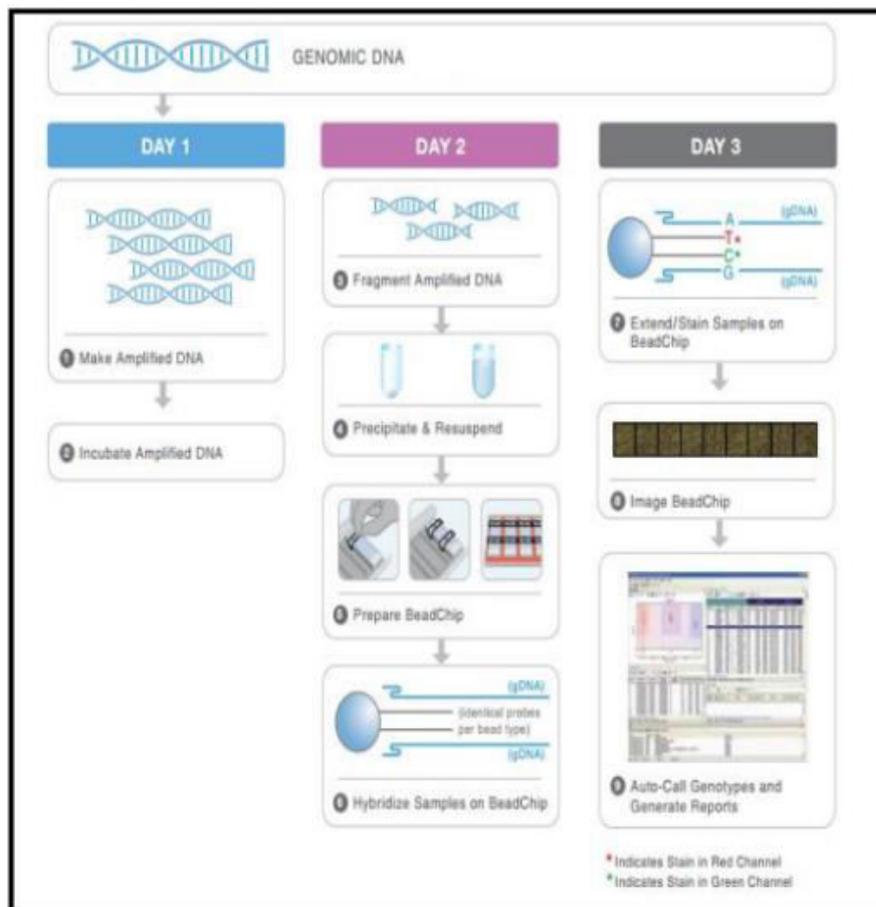


Figura 3.2: Rappresentazione schematica di un protocollo *SNP-array*.

La procedura inizia con un'amplificazione dei campioni (Step 1 e 2); i prodotti amplificati sono frammentati attraverso un processo enzimatico controllato che non richiede elettroforesi su gel (Step 3). Successivamente i DNA vengono fatti precipitare in alcool e poi risospesi (Step 4), i campioni risospesi vengono denaturati attraverso l'innalzamento della temperatura, adagiati sul array e incubati overnight nelle camere di ibridazione, affinché avvenga l'ibridazione del DNA alle sonde fissate sulle biglie (Step 5 e 6). Al termine dell'incubazione, vengono effettuati alcuni lavaggi per eliminare il DNA non fissato o non correttamente appaiato. La specificità è conferita attraverso estensione enzimatica delle basi, in cui sono incorporati dei nucleotidi marcati con dei fluorocromi (Step 7) e pertanto individuabili con la scansione dell'array tramite il sistema Illumina HiScan (Step 8). Finita la scansione dell'array inizia l'analisi dei dati ottenuti (Step 9).

3.6.1 Genotipizzazione

I dati generati dai chip sono stati processati con il software Illumina's GenomeStudio (versione 2011.1) seguendo le istruzioni del costruttore. Solo campioni con una *call rate* >98.5% sono stati inclusi nelle analisi successive.

Per valutare eventuali inconsistenze di parentela o parentele nascoste è stata effettuata un'analisi *identity-by-descent* (IBD). I dati di genotipizzazione di 55 campioni (includendo 5 copie di fratelli) sono stati processati con il *tool* Plink (versione 1.07) utilizzando i comandi: --maf, --hwe, --geno and --genome (Purcell et al., 2007). Tutte le coppie di fratelli (intervallo pi-hat 0.45-0.61) e i soggetti non imparentati (pi-hat cut off 0.05) sono stati confermati.

3.6.2 Chiamata delle CNV

I valori LRR e BAF sono stati esportati dal GenomeStudio per effettuare la chiamata delle CNV. Per ridurre i falsi positivi due algoritmi di chiamata delle CNV sono stati applicati: il *cnvPartition* (versione 3.2.0) di Illumina e il *PennCNV* (v1.0.4) come descritto da Wang e colleghi (2007). In entrambi le analisi prima è stato effettuato un test di controllo di ciascun campione. Sono stati inclusi dei campioni che rispettavano i seguenti criteri: $LRR_SD \leq 0.25$, $BAF_SD \leq 0.05$, $BAF_drift \leq 0.002$ and $|WF| \leq 0.05$ (Figura 3.3).

La chiamata delle CNV con *cnvPartition* è stata effettuata con la configurazione impostata dal costruttore: confidenza ≥ 35 , numero di marcatori ≥ 3 and lunghezza ≥ 1 kb. Per la chiamata delle CNV con il *PennCNV* è stata creata una pipeline che includeva due principali passaggi: prima, la chiamata delle CNV è stata effettuata utilizzando i seguenti parametri: confidenza ≥ 10 , numero di marcatori ≥ 5 e lunghezza delle CNV ≥ 2 kb; secondo, le CNV adiacenti sono state unite. Le CNV che si sovrappongono con dei *gaps* del genoma hg19 (>200 kb), le regioni di immunoglobuline, i centromeri e i telomeri (100 kb da ogni estremità di ogni cromosoma) sono state rimosse. Alla fine tutte le CNV di taglia inferiore alle 20 kb sono state rimosse da entrambi gli output (*cnvPartition* e *PennCNV*), mentre per le grandi CNV, oltre 10 mb, sono stati direttamente plotati i loro valori LRR e BAF (Figura 3.3). Le CNV rilevate nel cromosoma X, da entrambi gli algoritmi, sono state escluse da questo studio.

3.6.3 Selezione e caratterizzazione delle CNV ad “Alta Confidenza”

Per selezionare delle CNV ad “alta confidenza” prima sono stati eliminati i campioni che presentavano più di 100 CNV autosomiche con entrambi gli algoritmi dopo i vari passaggi di filtraggio. Secondo, sono state prese in considerazione solo le regioni che si intersecavano con una minima sovrapposizione di lunghezza di 0,5, tra le due liste finali delle CNV (cnvPartition e PennCNV, dopo i vari passaggi di filtraggio), sovrapposizione effettuata bi-direzionalmente (Figura 3.3).

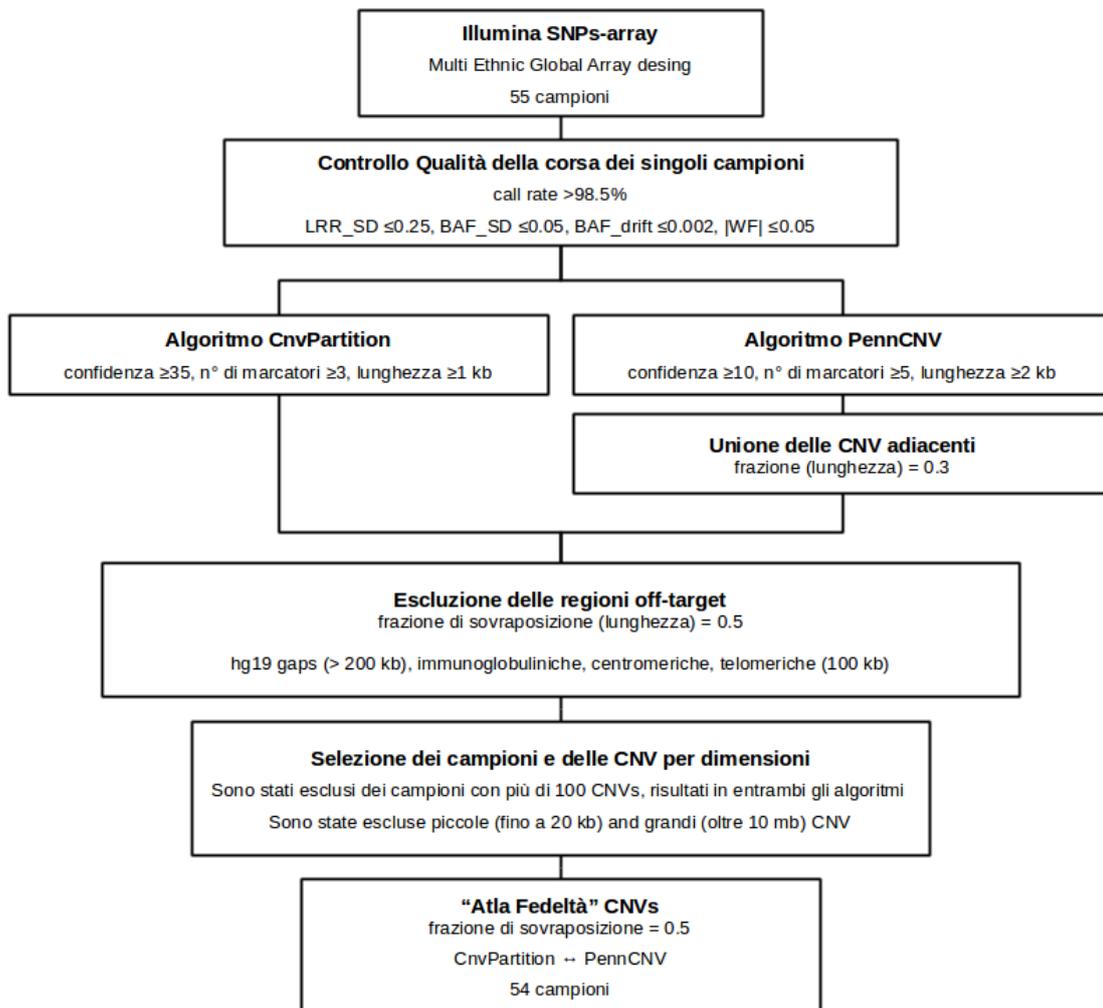


Figura 3.3 Rappresentazione schematica dei vari passaggi e filtri utilizzati a partire dalla corsa dei campioni fino alla chiamata delle CNV ad alta confidenza.

Per caratterizzare le CNV ad alta confidenza sono state filtrate con il *Database of Genomics Variants* (DGV, versione 2016-05-15) con una frazione di sovrapposizione di 0.5. In questo passaggio le CNV non riportate nel DGV sono state identificate e denominate come “nuove”.

Per esplorare il contenuto genico e possibili sovrapposizioni delle CNV con regioni regolatorie, le CNV sono state annotate con il hg19 RefSeq database, includendo 25 kb a monte e a valle delle CNV. Infine, si è calcolata la frequenza delle CNV nella popolazione generale, basandosi sul database DGV, definendo in questo studio le CNV come "rare" quelle con frequenza minore dell'1%.

4. RISULTATI E DISCUSSIONE

4.1 Targeted NGS

Il sequenziamento di 12 geni in 101 campioni ha generato di media 21 varianti per campione, con una profondità di copertura per corsa di oltre 74X e uniformità media di 93%. Iniziando da 71 varianti univoche (totale 122, Tabella S3, Capitolo 6 - Tabelle Supplementari) con una MAF <0.01 o no determinata, sono state selezionate 17 univoche e confermate mediante sequenziamento diretto 12 di esse (Tabella 4.1). Le varianti confermate sono distribuite in 6 geni (*CNTNAP2*, *DIP2A*, *DYX1C1*, *GCFC2*, *KIAA0319* e *ROBO1*) e in 11 soggetti.

Gruppo	Tipologia	Candidate (univoche)	Candidate (totali)	Confermate (univoche)	Confermate (totali)
B1	<i>Stopgain, Frameshift Indel, Canonical Splicing</i>	8	38	3	3
B2	<i>Splicing +/- 3-10, introniche / esoniche</i>	0	0	0	0
B3	Non-sinonime	9	9	9	9
Totale		17	47	12	12

Tabella 4.1: Riassunto delle varianti, rare o con MAF non determinata, candidate e confermate mediante sequenziamento diretto.

4.1.1 Varianti predette con un effetto deleterio

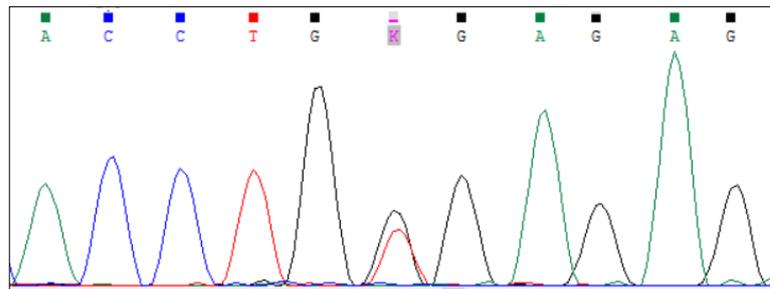
Due varianti confermate predicono la generazione di proteine tronche, mentre una terza variante di indurre uno splicing alternativo (Tabella 4.2, Figura 4.1). In dettaglio, il soggetto TS137 è portatore, eterozigote, di una variante stopgain (c.4159G>T, p.G1387*) identificata nel gene *DIP2A*. L'allele mutato produce una proteina mancante del secondo dominio "AMP-binding". Una variante frameshift dovuta alla delezione di 8 nucleotidi (c.2321_2328delTGGCTCTG, p.V774Afs*37) era stata confermata nel gene *KIAA0319* nel soggetto DD295. Questa variante porta alla perdita dei domini transmembrana e intracellulare. Infine, la variante rara rs138140952 (gnomExome-NFE, MAF 0.0006), situata nel canonical

splicing del gene *GCFC2* (c.266-2A>G), è stata identificata nel soggetto TS144. Questa variante è predetta da tutti e 3 i preditori di splicing (Spidex, scSNV e HSF) di indurre uno splicing alternativo.

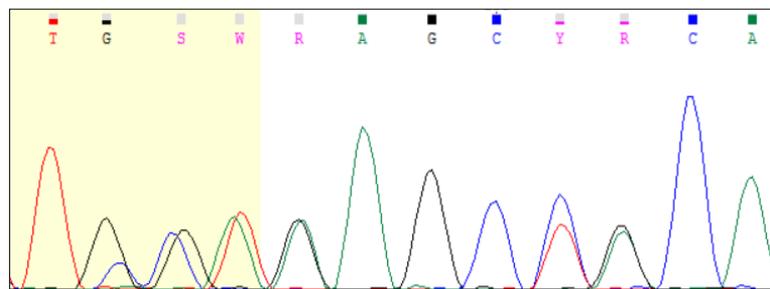
Tabella 4.2: Riassunto delle varianti predette con un effetto deleterio.

Varianti Nuove					
Gene	Chr	Inizio	Fine	Nomeclatura IUPAC	Soggetto
DIP2A	chr21	47983840	47983840	NM_015151:exon35:c.4159G>T:p.G1387*	TS137
KIAA0319	chr6	24564533	24564540	NM_014809:exon15:c.2321_2328del:p.V774fs*37	TS295
Varianti Rare					
Gene	Chr	Inizio	Fine	Nomeclatura IUPAC	Soggetto
GCFC2	chr2	75933779	75933779	NM_003203:exon2:c.266-2A>G	TS144

Figura 4.1: Elettroferogramma parziale delle nuove varianti DIP2A:c.4159G>T e KIAA0319:c.2321_2328del.



DIP2A (c.4159G>T): Sequenza visualizzata "ACCTG [K] GAGAG".



KIAA0319 (c.2321_2328del): sequenza visualizzata "TS [SWRAGCYR] CA".

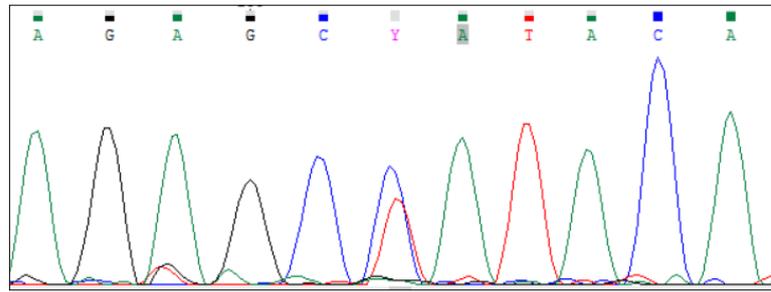
4.1.2 Varianti non-sinonime predette patogenetiche

I filtri applicati sulle varianti non-sinonime ha portato alla selezione di 9 varianti predette patogenetiche (Tabella 4.3, Figura 4.2). Tre varianti sono nuove, mentre le altre 5 sono state già riportate come SNPs nel dbSNPs, versione 147 (gnomExome-NFE, intervallo MAF 0.00001-0.00220). In dettaglio, sono stati identificati nel gene *CNTNAP2* una nuova variante e 2 SNPs, tutti localizzati nei domini “Laminin G-like”. Un punto interessante è che nel secondo SNP (c.2716C>T, p.R906C), l’arginina nella posizione 906 cambia in Istidina in un caso di autismo (Bakkaloglu et al., 2008). Nel gene *ROBO1* sono state validate 2 varianti nuove e un raro SNP. Le ultime 3 varianti sono dei rari SNPs identificati in 3 diversi geni: *DIP2A* (nel dominio “AMP-binding”), *DYX1C1* e *KIAA0319*.

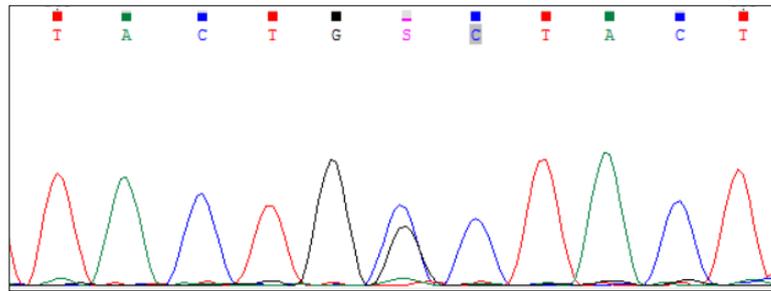
Tabella 4.3: Riassunto delle varianti non-sinonime predette con un effetto patogenetico.

Varianti Nuove					
Gene	Chr	Inizio	Fine	Nomeclatura IUPAC	Soggetto
<i>CNTNAP2</i>	chr7	147926780	147926780	NM_014141:exon20:c.3290C>T;p.P1097L	DD187
<i>ROBO1</i>	chr3	78987753	78987753	NM_002941:exon4:c.497C>G;p.A166G	DD144
<i>ROBO1</i>	chr3	78683121	78683121	NM_002941:exon24:c.3445T>C;p.S1149P	DD291
Varianti Rare					
Gene	Chr	Inizio	Fine	Nomeclatura IUPAC	Soggetto
<i>CNTNAP2</i>	chr7	147844744	147844744	NM_014141:exon17:c.2716C>T;p.R906C	DD309
<i>CNTNAP2</i>	chr7	147926851	147926851	NM_014141:exon20:c.3361G>A;p.E1121K	DD033
<i>DIP2A</i>	chr21	47980687	47980687	NM_015151:exon33:c.4022G>A;p.R1341H	DD276
<i>DYX1C1</i>	chr15	55731682	55731682	NM_130810:exon7:c.881A>T;p.K294M	DD020
<i>KIAA0319</i>	chr6	24596673	24596673	NM_014809:exon3:c.229T>C;p.C77R	DD317
<i>ROBO1</i>	chr3	78700915	78700915	NM_002941:exon19:c.2779G>A;p.G927R	DD004

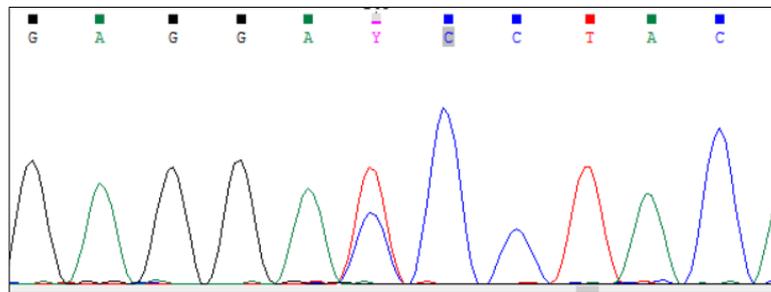
Figura 4.2: Elettroferogramma parziale delle nuove varianti CNTNAP2:c.3290C>T ROBO1:c.497C>G e ROBO1:c.3445T>C.



CNTNAP2 (c.3290C>T): sequenza visualizzata "AGAGC [Y] ATACA".



ROBO1 (c.497C>G): sequenza visualizzata "TACTG [S] CTACT".



ROBO1 (c.3445T>C): sequenza visualizzata "GAGGA [Y] CCTAC".

4.2 Chiamata delle CNV

Iniziando dai dati di genotipizzazione di 55 campioni, un campione è stato escluso dovuto all'eccesso (>100) di CNV autosomiche chiamate da entrambi gli algoritmi. Nella Tabella 4.4 sono riportati in dettaglio tutti i vari passaggi.

cnvPartition		PennCNV	
Campioni	55	55	Campioni
C 35, S 3, L 1 kb	4480	6515	C10, S 5, L 2kb
---	---	5276	Merge (0.3)
C 35, S 3, L >20 kb	1051	1302	C 10, S 5, L >20 kb
Off-target	-38	-85	Off-target
<i>Output</i>	1013	1217	<i>Output</i>
cnvPartition		PennCNV	
CNV >100	TS033, TS233	TS233, TS277	CNV >100
Campioni	54	54	Campioni
<i>Input</i>	830	1057	<i>Input</i>
Intersezione (cnvP→Penn), (0,5)	442	445	Intersezione (Penn→cnvP), (0,5)
<u>CNV ad “alta confidenza”</u>			
537			

Tabella 4.4: C = confidenza; S = marcatori; L = lunghezza. Off-target: hg19 gaps (>200 kb), le regioni di immunoglobuline, dei centromeri e dei telomeri (100 kb). cnvP = cnvPartition; Penn = PennCNV.

Il passaggio di intersezione tra le liste finali di CNV generate dal cnvPartition e dal PennCNV ha dato origine a 537 CNV ad alta confidenza, comprendenti 293 delezioni (Figura 4.3) e 244 duplicazioni (Figura 4.4), con una media di 7 CNV per campione. La lista completa delle CNV ad alta confidenza sono riportate nella Tabella S4, Capitolo 6 - Tabelle Supplementari, mentre la loro distribuzione lungo i 22 cromosomi autosomici si visualizza nella Figura S1, Capitolo 7 - Figure Supplementari.

Figura 4.3: Distribuzione delle CNV, delezioni, lungo il genoma. In rosso cnvPartition e in verde PennCNV. In blu le CNV ad alta confidenza risultate dall'intersezione delle CNV prodotte dai due algoritmi.

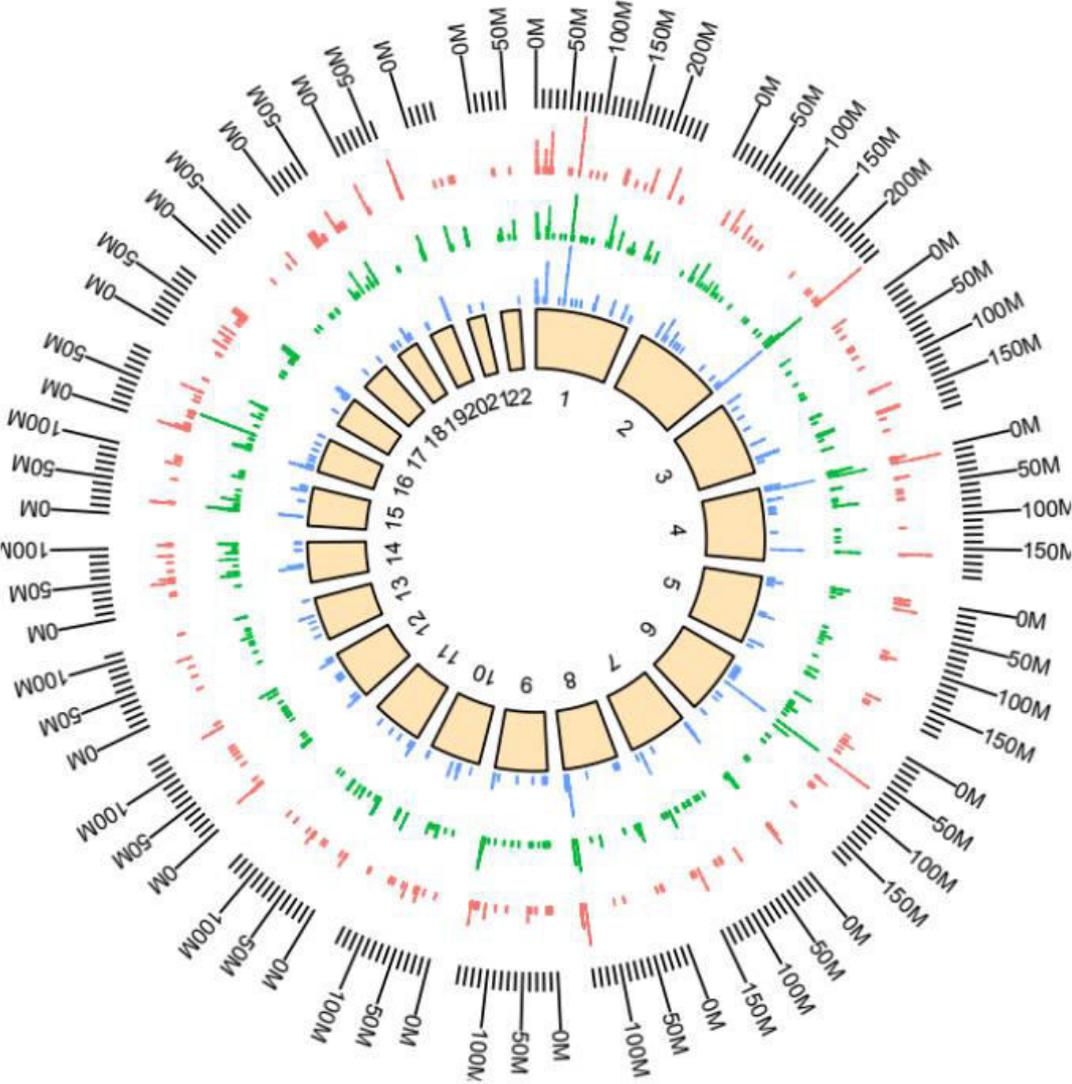
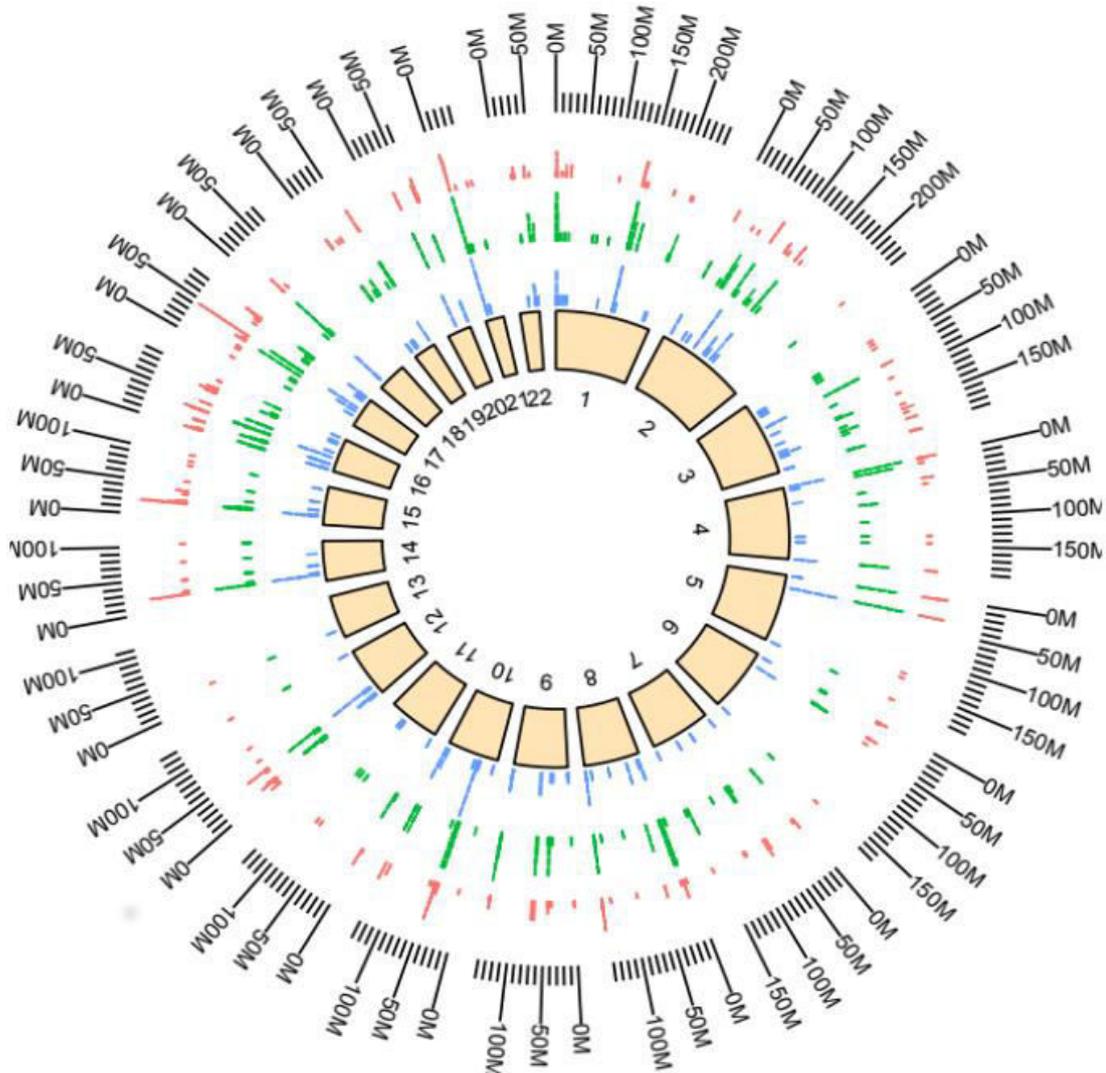


Figura 4.4: Distribuzione delle CNV, duplicazioni, lungo il genoma. In rosso cnvPartition e in verde PennCNV. In blu le CNV ad alta confidenza risultate dall'intersezione delle CNV prodotte dai due algoritmi.



La lunghezza delle CNV varia dalle ~14 kb fino a ~3.85 mb, con una lunghezza media di ~128 kb. Più del 70% delle CNV cade nell'intervallo di lunghezza di 14-100 kb (397 piccole CNVs), mentre 116 e 24 CNV appartengono rispettivamente ai gruppi di lunghezza di 100-500 kb (medie CNVs) and >500 kb (larghe CNVs).

4.2.1 CNV contenenti geni associati (o riportati) con la lettura o il linguaggio

L'annotazione delle CNV è stata effettuata mediante lo screening di singoli geni o geni contenuti in CNV, entrambi precedentemente associate (o riportate) con la lettura o con le disabilità del linguaggio (Carrion-Castillo et al., 2013; Eicher et al., 2014; Einarsdottir et al., 2017; Gialluisi et al., 2016; Graham and Fisher, 2015; Matsson et al., 2011; Sharma et al., 2017; Varshney et al., 2017; Veerappa et al., 2013a; Veerappa et al., 2014) (Tabella 4.5). I plot dei dati LRR e BAF per ciascuna CNV, partendo dai dati del PennCNV, sono riportati nella Figura S2, Capitolo 7 - Figure Supplementari.

Regione cromosomica	Lunghezza	Marcatori	CN	Gene(i)	Distanza	Soggetti
chr1:72767720-72805498	37778	16	0	NEGR1	19443	DD159
chr1:72767720-72806589	38869	17	0	NEGR1	19443	DD069
chr1:72767720-72806589	38869	17	0	NEGR1	19443	DD119A
chr1:72767720-72806589	38869	19	0	NEGR1	19443	DD119
chr1:72767720-72806589	38869	19	0	NEGR1	19443	DD168
chr1:72767720-72806589	38869	19	0	NEGR1	19443	DD203A
chr1:72767720-72806589	38869	19	0	NEGR1	19443	DD203
chr1:225391925-225459761	67836	30	3	DNAH14	0	DD069
chr6:168342746-168596251	253505	360	3	AFDN,FRMD1,HGC6.3, KIF25,KIF25-AS1	0	DD247
chr7:111148162-111288374	140212	43	1	IMMP2L	0	DD092
chr10:68076803-68115498	38695	24	1	CTNNA3	0	DD033
chr10:68076803-68115498	38695	24	1	CTNNA3	0	DD285
chr12:8090505-8115176	24671	11	3	SLC2A3	1613	DD226
chr12:8043600-8080159	36559	15	3	SLC2A14,SLC2A3	0	DD226
chr12:8008028-8115176	107148	53	1	SLC2A14,SLC2A3	0	DD122
chr12:8000978-8115176	114198	56	1	SLC2A14,SLC2A3	0	DD279
chr17:44167366-44434779	267413	152	3	ARL17A,ARL17B,KANSL1, KANSL1-AS1,LRRC37A	0	DD331
chr17:44171207-44368830	197623	146	3	KANSL1,KANSL1-AS1	0	DD208
chr17:44167366-44222424	55058	38	3	KANSL1	0	DD163
chr19:1456150-1484126	27976	108	3	APC2,C19orf25,PCSK4	0	DD189

Tabella 4.5: Lista delle CNV contenenti geni associati (o riportati) con la lettura o il linguaggio.

In dettaglio, una CNV identificata nella regione "chr6:168342746-168596251", in un unico paziente, era già stata descritta da Gialluisi e colleghi (2016). La CNV identificata in quel studio, "chr6:168336080-168597552", ha presentato l'associazione più significativa in uno studio di meta-analisi (Figura 4.5).

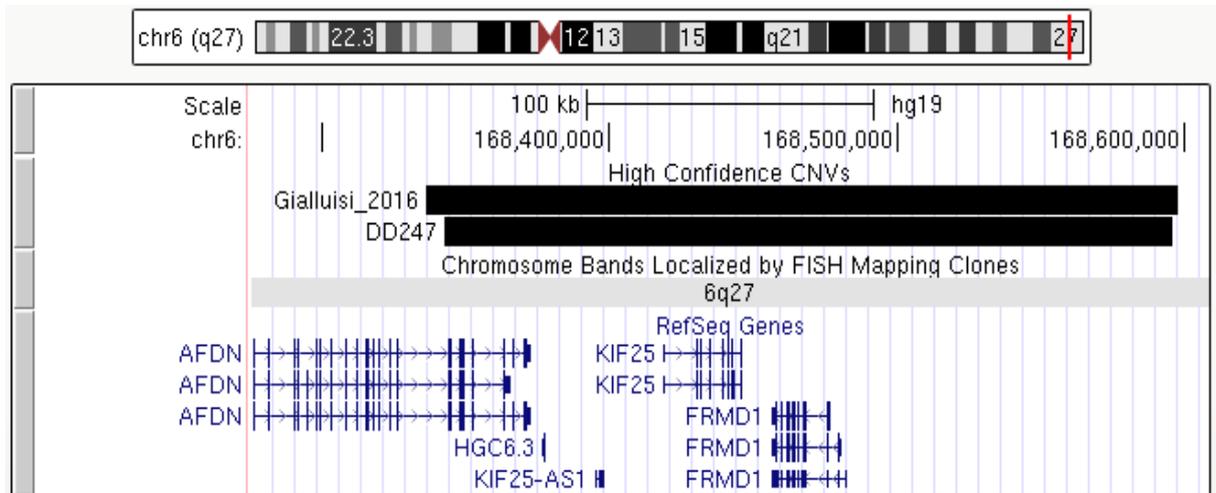


Figura 4.5: Visualizzazione delle CNV (color nero) su Genome Browser e la loro sovrapposizione sulle regioni geniche.

Una duplicazione di circa 267 kb è stata identificata nella regione 17q21.31 e si sovrappone con i geni *ARL17A*, *ARL17B*, *KANSL1*, *KANSL1-AS1* e *LRRC37A*. Altre due duplicazioni, che sovrappongono solo parte della lista di geni sopra riportati, sono state identificate in altri due soggetti non inparentati (Figura 4.6). CNV in queste regione erano già state riportate da Veerappa e colleghi (2014), descrivendo una delezione e 5 micro-duplicazioni e micro-delezioni in tre famiglie di dislessici.

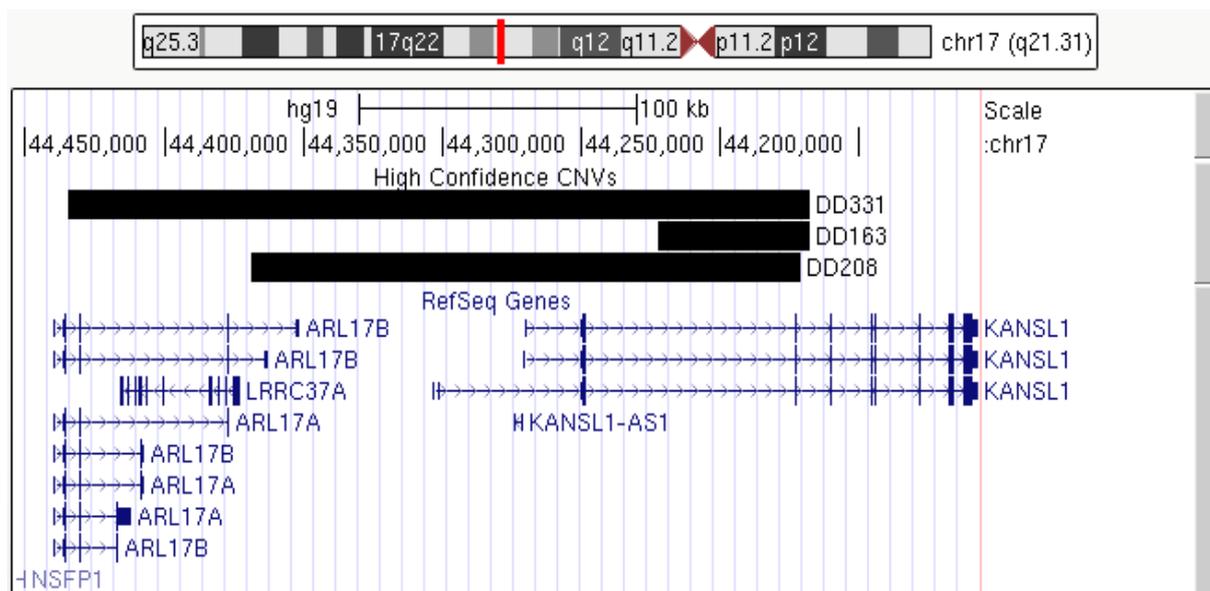


Figura 4.6: Visualizzazione delle CNV (color nero) su Genome Browser e la loro sovrapposizione sulle regioni geniche.

Una lista di CNV identificate in diversi soggetti e localizzate nelle regioni geniche già riportate da Gialluisi e colleghi (2016), sovrappongo vari geni o parte di essi o sono situate a monte o a valle di loro: *APC2*, *CTNNA3*, *C19ORF25*, *DNAH14*, *IMMP2L*, *NEGR1* e *PCSK4* (Figura 4.7).

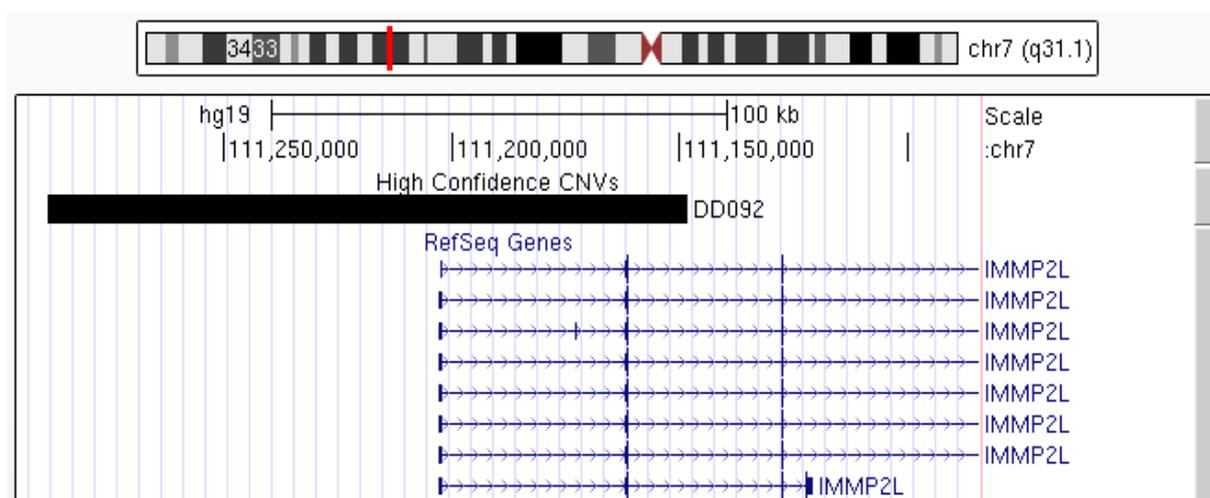


Figura 4.7: Visualizzazione delle CNV (color nero) su Genome Browser e la loro sovrapposizione sulle regioni geniche.

Alla fine, 4 CNV sono state riportate nella regione 12p12.31: due duplicazioni adiacenti presenti nello stesso soggetto, una a monte del gene *SLC2A3* e l'altra sovrapposta parzialmente con il medesimo gene; e due delezioni (~110 kb) in due soggetti non inparentanti, CNV che sovrappongono l'intero gene *SLC2A3*, una regione inter-genica a valle e parzialmente il gene *SLC2A14* (Figura 4.8).

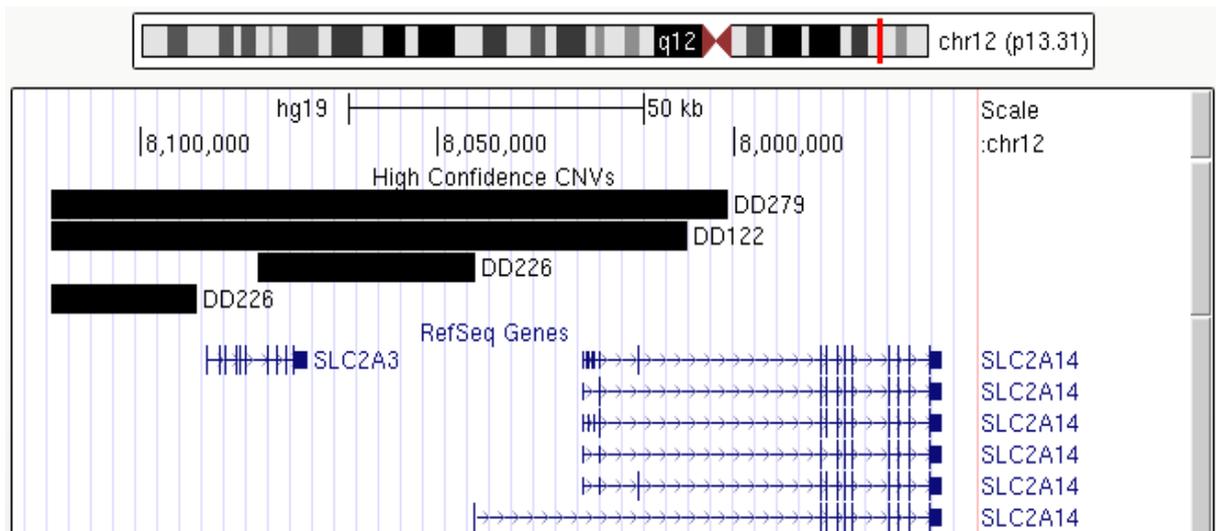


Figura 4.8: Visualizzazione delle CNV (color nero) su Genome Browser e la loro sovrapposizione sulle regioni geniche.

Tutte le CNV identificate sono frequenti (> 1%), in base al database DGV.

4.2.2 CNV comuni tra fratelli

Guardando per regioni comuni tra coppie di fratelli, diverse CNV sono state identificate con numero di copie 0, 1 e 3 (Tabella 4.6). Tra queste le più importanti sembrerebbero 3: la CNV = 0 a monte del gene *NEGR1* (già descritta nel sotto-capitolo 4.2.1); la CNV con numero di copie = 1 che si sovrappone al gene *TENM2* (descritta successivamente nel sotto-capitolo 4.2.3); e le CNV con numero di copie = 3 che si localizzano nella regione 16p13.11 dei fratelli DD119 e DD119A. I plot dei dati LRR e BAF per ciascuna CNV, partendo dai dati del PennCNV, sono riportati nella Figura S3, Capitolo 7 - Figure Supplementari.

Regione cromosomica	Lunghezza	Marcatori	CN	Gene(i)	Distanza	Soggetti
chr1:152561408-152585242	14	23834	0	LCE3C	0	DD004/DD004A
chr1:72767720-72806589	17	38869	0	NEGR1	19443	DD119/DD119A
chr3:162512480-162616738	61	104258	0	-	-	DD119/DD119A
chr4:10212896-10232936	23	20040	0	-	-	DD119/DD119A
chr16:15022065-15059248	13	37183	3	MIR6770-1,MIR6770-2, MIR6770-3,NPIPA1	0	DD119/DD119A
chr16:15839439-15960226	77	120787	3	FOPNL,MYH11	0	DD119/DD119A
chr16:16139714-16200716	50	61002	3	ABCC1	0	DD119/DD119A
chr16:16265935-16284192	48	18257	3	ABCC6	0	DD119/DD119A
chr4:34781075-34823780	9	42705	0	-	-	DD138/DD138A
chr12:11507321-11527351	18	20030	3	PRB1	0	DD138/DD138A
chr1:72767720-72806589	17	38869	0	NEGR1	19443	DD203/DD203A
chr2:52751394-52785233	7	33839	1	-	-	DD203/DD203A
chr4:10212896-10232936	23	20040	0	-	-	DD203/DD203A
chr14:41610224-41660172	16	49948	1	LOC644919		DD203/DD203A
chr13:72866419-72984051	70	117632	1	-	-	DD231/DD231A
chr4:161048044-161070928	21	22884	1	-	-	DD231/DD231A
chr5:166844833-166881177	30	36344	1	TENM2	0	DD231/DD231A

Tabella 4.6: CNV o regioni di CNV comuni tra coppie di fratelli.

Nella regione 16p13.11 sono localizzati una serie di geni (*NOMO3*, *ABCC6*, *ABCC1*, *FOPNL*, *MYH11*, *NDE1*, *MARF1*, *C16orf45* and *MPV17L*) che sono coinvolti in diverse CNV, sia delezioni che duplicazioni (Figura 4.9). Questi eventi di CNV sono stati riportati in diversi

disordini come: ritardo verbale e del linguaggio, deficit dell'attenzione e iperattività, autismo, disabilità intellettiva, schizofrenia ed epilessia (Nagamani et al., 2011; Ramalingam et al., 2011; Tropeano et al., 2013; Stefansson et al., 2014; Gialluisi et al., 2016).

Oltre a queste CNV condivise tra fratelli, loro presentano delle altre nella stessa regione non condivise. Parte di esse cadono sul gene *NDE1* (Reiner et al., 2016) e *NTAN1* (Hirai et al., 2006), due geni che si considerano dei candidati più forti di questa regione per i disordini di neurosviluppo.

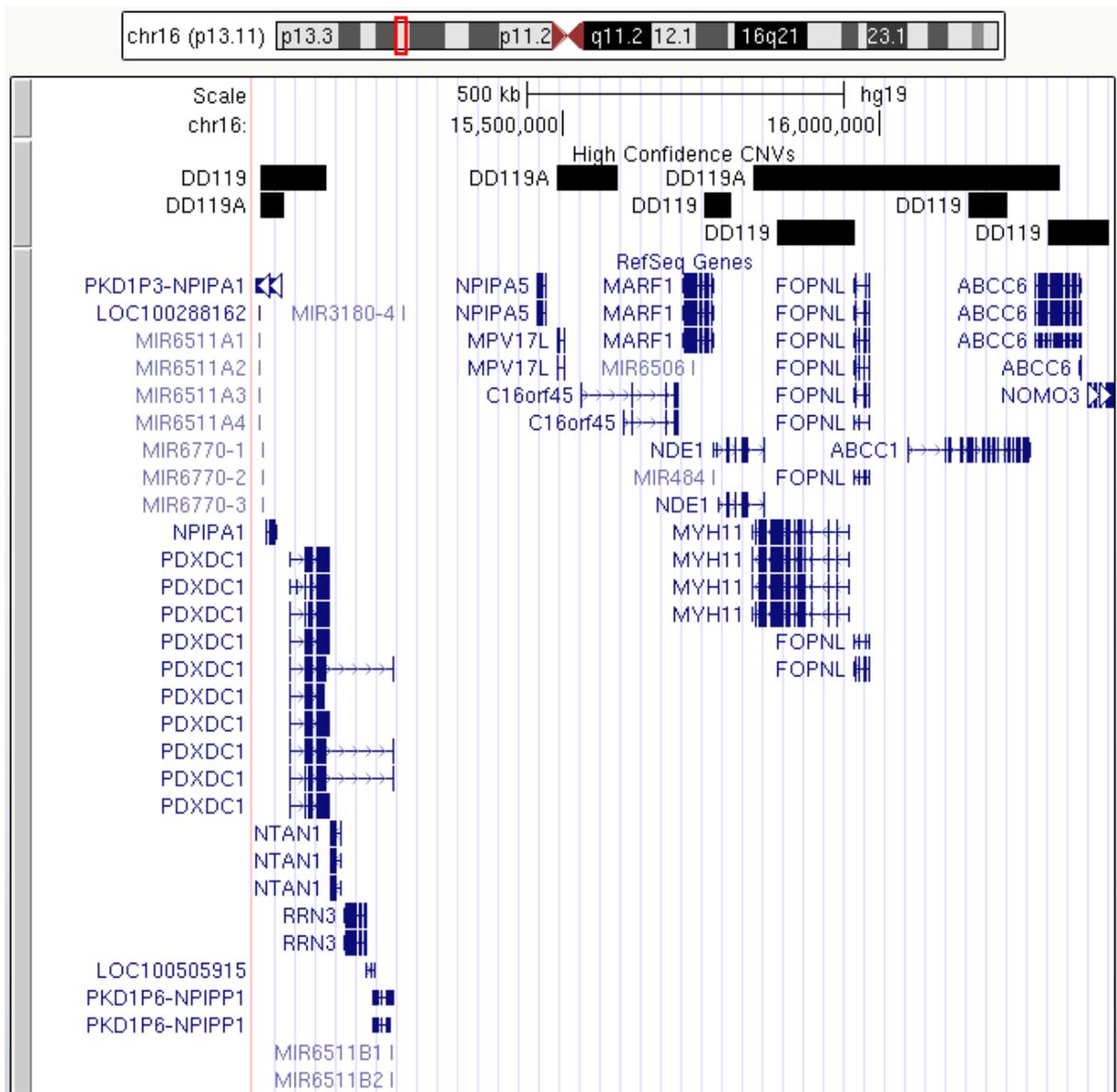


Figura 4.9: Visualizzazione delle CNV (color nero) su Genome Browser e la loro sovrapposizione sulle regioni geniche.

Tutte le CNV identificate sono frequenti (> 1%), in base al database DGV.

4.2.3 CNV non riportate nel DGV database

Per caratterizzare ulteriormente le CNV ad alta confidenza identificate nei 54 soggetti, esse sono state filtrate con il database DGV, con una frazione di sovrapposizione di lunghezza pari a 0,5 (Tabella 4.7). In totale sono state identificate 9 CNV non riportate nel DGV, composte da 6 CNV che si sovrappongono a un gene ciascuna, mentre le altre 3 non contengono nessun gene né si trovano vicino ad altri geni (25 kb da ciascun bordo delle CNV). I plot dei dati LRR e BAF per ciascuna CNV, partendo dai dati del PennCNV, sono riportati nella Figura S4, Capitolo 7 - Figure Supplementari.

Regione cromosomica	Lunghezza	Marcatori	CN	Gene(i)	Distanza	Soggetti
chr2:60438359-60485161	46803	12	1	-	-	DD223
chr2:178299726-178352746	53021	14	1	AGPS	0	DD277
chr2:214752273-214810651	58379	19	1	SPAG16	0	DD309
chr5:166844833-166881177	36345	30	1	TENM2	0	DD231
chr5:166844833-166881177	36345	30	1	TENM2	0	DD231A
chr6:25749957-25779736	32745	39	1	SLC17A4	0	DD010
chr6:40297288-40327096	32149	27	1	LINC00951	0	DD333
chr15:36274022-36359839	85818	44	3	-	-	DD069
chr21:9555540-9705758	150219	121	3	-	-	DD163

Tabella 4.7: Lista delle CNV non riportate nel database DGV.

Uno dei geni più interessanti è la *TENM2* dove la CNV presente in entrambi i fratelli si sovrappone al livello intronico dell'isoforma 1 (NM_001122679.1, isoforma non validata) (Figura 4.10). Questo gene codifica per la proteina teneurina-2, che appartiene a una larga famiglia di proteine della superficie cellulare di tipo II (Tucker et al., 2007). La *TENM2* è prevalentemente espressa nel cervello dove è coinvolta nella regolazione sinaptica, legando molto probabilmente la latrofilina che è un recettore neuronale (Mosca, 2015).

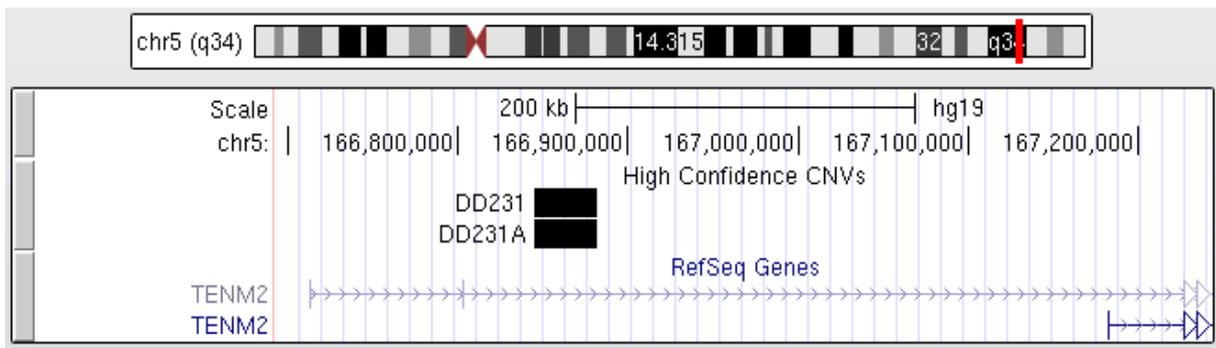


Figura 4.10: Visualizzazione delle CNV (color nero) su Genome Browser e la loro sovrapposizione sulle regioni geniche.

5. CONCLUSIONI

Diversi dei geni presi in considerazione in questo studio e analizzati mediante targeted-NGS, sono stati scoperti in famiglie dove essi erano localizzati nei punti di rottura di traslocazioni cromosomiche o in altri casi deleti. Il risultato di questi eventi solitamente porta alla ridotta trascrizione e quasi sempre alla riduzione dei livelli di proteina di interesse. Basandosi su questo sono state ricercate delle varianti deleterie (*stopgain, frameshift, canonical splicing*) che potessero replicare questo evento biologico. Tuttavia, nonostante siano state identificate 3 varianti facenti parte di questa categoria (Tabella 4.2), si devono attendere ulteriori analisi nei genitori o in altri fratelli sani, per poter fare un qualsiasi tipo di associazione tra genotipo e fenotipo.

Analogamente, sono state ricercate varianti rare o nuove con un potenziale effetto patogenetico, basandosi su un modello che prevede che diverse varianti, ognuna partecipante con un piccolo effetto additivo al fenotipo, siano all'origine dell'insorgenza della dislessia. In base a questo sono state identificate 9 varianti, 6 rare e 3 nuove, predette come patogenetiche da almeno due dei tre predittori usati (Tabella 4.3). Come detto precedentemente riguardo alle varianti identificate mediante NGS, anche in questo caso non si può fare attualmente nessuna associazione diretta tra genotipo e fenotipo fino alla costruzione di una coorte di controllo costituita da una popolazione generale di nazionalità italiana. Inoltre, vanno prese in considerazione in studi futuri anche le varianti frequenti (MAF >1%) per un eventuale effetto additivo di alcune di esse sul fenotipo.

In base alla letteratura l'utilizzo degli SNP-array (0.6 M - 1 M di marcatori) è un valido metodo per l'identificazione di CNV anche nel caso della dislessia. Fino ad oggi sono stati effettuati pochi studi che hanno riportato delle associazioni di CNV con i fenotipi studiati. L'analisi delle CNV sono state di aiuto più concreto nello studio di famiglie dislessiche dove hanno permesso l'identificazione di validi candidati nonché di regioni contenenti geni importanti per il neurosviluppo e altri aspetti biologici, soggette a variazione di numero di copie.

Nel nostro caso in *primis* abbiamo valutato l'utilizzo di *SNP-array* ad alta densità (>1.7 M di marcatori) in 55 campioni. Dai dati ottenuti è emerso che il numero delle piccole CNV (<20 kb) è altissimo, oltre il 75%, risultato che indicherebbe un alto numero di falsi positivi. A tal

proposito è stato messo appunto una pipeline che riduce drasticamente i falsi positivi (Tabella 4.4).

Per analizzare al meglio le CNV ottenute, denominate ad “alta confidenza” dovuto al fatto che sono state replicate da due tool indipendenti, ci siamo basati sulla letteratura. Come per l’interpretazione delle varianti anche in questo caso senza lo studio di famiglie intere (sia soggetti affetti che sani), come anche senza una coorte di controllo, non si possono effettuare delle associazioni genotipo-fenotipo. Da sottolineare che anche se in una seconda fase del progetto si creasse una coorte di controllo e/o fossero presi in considerazione tutti i membri di famiglie, solo degli appropriati studi funzionali sui geni candidati potrebbero portare a delle conclusioni più definitive.

In definitiva, nonostante i buoni risultati ottenuti, che hanno permesso di replicare alcuni dati noti e di identificarne di nuovi, per ottenere una effettiva significatività sarebbe necessario un ampliamento dello studio e una validazione funzionale delle varianti.

6. TABELLE SUPPLEMENTARI

Tabella S1: Fenotipi dei partecipanti allo studio e analisi effettuate.

ID	Sesso	Dislessia	Altri DSA	NGS	Chip-Aray
DD004	F	yes	no	yes	yes
DD004A	M	yes	no	yes	yes
DD008	F	yes	no	yes	no
DD010	F	yes	no	yes	yes
DD020	M	yes	Disgrafia	yes	no
DD033	F	yes	no	yes	yes
DD034	M	yes	Disgrafia + Discalculia	yes	no
DD037	F	yes	Disortografia	yes	no
DD038	M	yes	Disortografia	yes	yes
DD041	F	yes	Discalculia	yes	no
DD046	M	yes	Disortografia	yes	no
DD054	M	yes	no	yes	yes
DD068	M	yes	Disortografia	yes	no
DD069	M	yes	no	yes	yes
DD073	F	yes	no	yes	no
DD080	F	yes	Disortografia	yes	no
DD081	M	yes	Disgrafia + Disortografia	yes	no
DD084	M	yes	Disortografia	yes	no
DD088	F	yes	no	yes	no
DD092	M	yes	Disgrafia	yes	yes
DD102	F	yes	Discalculia	yes	no
DD105	M	yes	Disgrafia	yes	no
DD117	F	yes	no	yes	no
DD119	M	yes	Disgrafia + Disortografia	yes	yes
DD119A	F	yes	Disgrafia + Disortografia + Discalculia	yes	yes
DD120	F	yes	no	yes	no
DD122	F	yes	no	yes	yes
DD127	F	yes	no	yes	no
DD129	M	yes	no	yes	no
DD133	F	yes	no	yes	no
DD137	M	yes	Disortografia	yes	no
DD138	M	yes	no	yes	yes
DD138A	F	yes	Disgrafia + Disortografia + Discalculia	yes	yes
DD144	M	yes	Disortografia	yes	no

ID	Sesso	Dislessia	Altri DSA	NGS	Chip-Array
DD150	M	yes	Discalculia	yes	no
DD155	M	yes	Disortografia + Discalculia	yes	no
DD159	F	yes	no	yes	yes
DD163	M	yes	no	yes	yes
DD167	F	yes	no	yes	no
DD168	M	yes	no	yes	yes
DD172	M	yes	no	yes	no
DD180	M	yes	no	yes	yes
DD182	F	yes	no	yes	no
DD184	M	yes	no	yes	yes
DD187	M	yes	Disgrafia	yes	no
DD189	F	yes	Disortografia	yes	yes
DD195	M	yes	Disgrafia + Disortografia + Discalculia	yes	no
DD200	M	yes	no	yes	no
DD203	M	yes	no	yes	yes
DD203A	M	yes	no	yes	yes
DD206	M	yes	no	yes	no
DD207	F	yes	Disortografia + Discalculia	yes	no
DD208	M	yes	no	yes	yes
DD209	F	yes	no	yes	yes
DD211	M	yes	no	yes	yes
DD215	M	yes	no	yes	no
DD217	M	yes	Disgrafia + Disortografia	yes	no
DD218	M	yes	Disortografia	yes	yes
DD220	F	yes	no	yes	yes
DD223	F	yes	Disortografia	yes	yes
DD224	M	yes	no	yes	no
DD226	M	yes	Disortografia	yes	yes
DD227	F	yes	no	yes	no
DD228	M	yes	Disgrafia + Disortografia	yes	no
DD228A	M	yes	no	yes	no
DD231	M	yes	Disortografia + Discalculia	yes	yes
DD231A	F	yes	Disortografia	yes	yes
DD233	M	yes	Disortografia	yes	yes
DD235	M	yes	Disortografia	yes	yes
DD240	F	yes	Disortografia	yes	yes
DD243	F	yes	no	yes	no
DD243A	M	yes	Disortografia	yes	no
DD246	M	yes	Disortografia	yes	no
DD247	M	yes	Disortografia	yes	yes
DD249	M	yes	no	yes	no
DD250	F	yes	Disortografia + Discalculia	yes	no
DD263	F	yes	Discalculia	yes	no

ID	Sesso	Dislessia	Altri DSA	NGS	Chip-Aray
DD263	F	yes	Discalculia	yes	no
DD264	F	yes	no	yes	yes
DD265	M	yes	Disortografia	yes	yes
DD275	M	yes	Disgrafia + Disortografia	yes	no
DD276	M	yes	no	yes	yes
DD277	M	yes	no	yes	yes
DD279	M	yes	Disortografia	yes	yes
DD285	M	yes	no	yes	yes
DD291	M	yes	Disortografia	yes	yes
DD293	M	yes	Discalculia	yes	no
DD295	M	yes	Disortografia + Discalculia	yes	no
DD296	F	yes	Disortografia	yes	yes
DD298	M	yes	no	yes	yes
DD303	M	yes	Disortografia	yes	yes
DD306	M	yes	Disortografia	yes	yes
DD307	M	yes	no	yes	yes
DD309	M	yes	Disortografia	yes	yes
DD311	M	yes	no	yes	yes
DD316	M	yes	Disortografia	yes	yes
DD317	F	yes	Disortografia	yes	yes
DD319	M	yes	Disortografia	yes	yes
DD331	F	yes	no	yes	yes
DD332	M	yes	no	yes	yes
DD333	F	yes	no	yes	yes

Tabella S2: Lista dei database utilizzati per l'annotazione delle varianti con Annovar.

hg19_ALL.sites.2015_08.txt / hg19_ALL.sites.2015_08.txt.idx
hg19_avsnp147.txt / hg19_avsnp147.txt.idx
hg19_cadd13gt10.txt / hg19_cadd13gt10.txt.idx
hg19_caddindel.txt / hg19_caddindel.txt.idx
hg19_clinvar_20170130.txt / hg19_clinvar_20170130.txt.idx
hg19_cytoBand.txt
hg19_dbnsfp31a_interpro.txt / hg19_dbnsfp31a_interpro.txt.idx
hg19_dbnsfp33a.txt / hg19_dbnsfp33a.txt.idx
hg19_dbscsnv11.txt / hg19_dbscsnv11.txt.idx
hg19_ensGeneMrna.fa
hg19_ensGene.txt
hg19_esp6500siv2_all.txt / hg19_esp6500siv2_all.txt.idx
hg19_esp6500siv2_ea.txt / hg19_esp6500siv2_ea.txt.idx
hg19_EUR.sites.2015_08.txt / hg19_EUR.sites.2015_08.txt.idx
hg19_exac03.txt / hg19_exac03.txt.idx
hg19_GDI_full_10282015.txt
hg19_gnomad_exome.txt / hg19_gnomad_exome.txt.idx
hg19_gnomad_genome.txt / hg19_gnomad_genome.txt.idx
hg19_gwava.txt / hg19_gwava.txt.idx
hg19_intervar_20170202.txt / hg19_intervar_20170202.txt.idx
hg19_mcap.txt / hg19_mcap.txt.idx
hg19_refGeneMrna.fa
hg19_refGene.txt
hg19_refGeneVersion.txt
hg19_revel.txt / hg19_revel.txt.idx
hg19_RVIS_ExAC_4KW.txt
hg19_spidex.txt / hg19_spidex.txt.idx
hg19_wgEncodeGencodeBasicV19Mrna.fa

Tabella S3: Lista delle varianti rimaste dopo la selezione di loro con frequenza allelica <0.01 o non riportata.

Posizione genomica	Ref	Alt	Gene.refGene	ExonicFunc.refGene
chr1:35900572-35900572	G	A	KIAA0319L	synonymous SNV
chr1:35944645-35944646	GG	-	KIAA0319L	frameshift deletion
chr1:35944646-35944646	G	-	KIAA0319L	frameshift deletion
chr1:35944646-35944646	G	T	KIAA0319L	synonymous SNV
chr1:35944646-35944646	G	-	KIAA0319L	frameshift deletion
chr1:35944646-35944646	G	T	KIAA0319L	synonymous SNV
chr1:35944646-35944646	G	-	KIAA0319L	frameshift deletion
chr1:35944646-35944646	G	T	KIAA0319L	synonymous SNV
chr1:35944646-35944646	G	-	KIAA0319L	frameshift deletion
chr1:35944646-35944646	G	T	KIAA0319L	synonymous SNV
chr1:35944646-35944646	G	-	KIAA0319L	frameshift deletion
chr1:35944646-35944646	G	T	KIAA0319L	synonymous SNV
chr1:35944646-35944646	G	-	KIAA0319L	frameshift deletion
chr1:35944646-35944646	G	T	KIAA0319L	synonymous SNV
chr1:35944646-35944646	G	-	KIAA0319L	frameshift deletion
chr1:35944646-35944646	G	T	KIAA0319L	synonymous SNV
chr1:35944646-35944646	G	-	KIAA0319L	frameshift deletion
chr1:35944646-35944646	G	T	KIAA0319L	synonymous SNV
chr1:35944646-35944646	G	-	KIAA0319L	frameshift deletion
chr1:35944646-35944646	G	T	KIAA0319L	synonymous SNV
chr1:35944646-35944646	G	-	KIAA0319L	frameshift deletion
chr1:35944646-35944646	G	T	KIAA0319L	synonymous SNV
chr1:35944646-35944646	G	-	KIAA0319L	frameshift deletion
chr1:35944646-35944646	G	T	KIAA0319L	synonymous SNV
chr1:35944796-35944796	T	C	KIAA0319L	synonymous SNV
chr2:75873943-75873943	T	G	MRPL19	nonsynonymous SNV
chr2:75873955-75873955	G	A	MRPL19	nonsynonymous SNV
chr2:75874318-75874318	A	G	MRPL19	nonsynonymous SNV
chr2:75882199-75882199	A	G	MRPL19	nonsynonymous SNV
chr2:75893101-75893101	G	C	GCFC2	nonsynonymous SNV
chr2:75897256-75897256	C	T	GCFC2	splicing
chr2:75897256-75897256	C	T	GCFC2	splicing
chr2:75897256-75897256	C	T	GCFC2	splicing
chr2:75897347-75897347	G	A	GCFC2	synonymous SNV
chr2:75933779-75933779	T	C	GCFC2	splicing
chr3:78683121-78683121	A	G	ROBO1	nonsynonymous SNV
chr3:78700915-78700915	C	T	ROBO1	nonsynonymous SNV
chr3:78717094-78717094	C	T	ROBO1	synonymous SNV
chr3:78717396-78717396	-	T	ROBO1	frameshift insertion
chr3:78735014-78735014	G	A	ROBO1	synonymous SNV
chr3:78737843-78737843	G	A	ROBO1	synonymous SNV
chr3:78766544-78766544	C	T	ROBO1	synonymous SNV
chr3:78767040-78767040	A	G	ROBO1	splicing
chr3:78796054-78796054	T	C	ROBO1	splicing
chr3:78987753-78987753	G	C	ROBO1	nonsynonymous SNV
chr3:78988062-78988062	T	C	ROBO1	nonsynonymous SNV
chr6:24174967-24174967	G	A	DCDC2	synonymous SNV
chr6:24353858-24353858	A	-	DCDC2	splicing
chr6:24547456-24547456	C	T	KIAA0319	synonymous SNV
chr6:24559300-24559300	C	T	KIAA0319	nonsynonymous SNV
chr6:24564533-24564540	CAGAGCCA	-	KIAA0319	frameshift deletion
chr6:24566953-24566953	G	A	KIAA0319	nonsynonymous SNV
chr6:24566953-24566953	G	A	KIAA0319	nonsynonymous SNV
chr6:24569006-24569006	C	T	KIAA0319	splicing

Posizione genomica	Ref	Alt	Gene.refGene	ExonicFunc.refGene
chr6:24576736-24576736	T	C	KIAA0319	nonsynonymous SNV
chr6:24576771-24576771	A	T	KIAA0319	nonsynonymous SNV
chr6:24578421-24578421	G	A	KIAA0319	synonymous SNV
chr6:24580185-24580185	-	AA	KIAA0319	splicing
chr6:24580185-24580185	-	AA	KIAA0319	splicing
chr6:24580185-24580185	-	AA	KIAA0319	splicing
chr6:24580185-24580185	-	AA	KIAA0319	splicing
chr6:24580185-24580185	-	AA	KIAA0319	splicing
chr6:24588811-24588811	T	-	KIAA0319	splicing
chr6:24588811-24588811	T	A	KIAA0319	splicing
chr6:24588811-24588811	T	A	KIAA0319	splicing
chr6:24588960-24588960	G	A	KIAA0319	synonymous SNV
chr6:24588960-24588960	G	A	KIAA0319	synonymous SNV
chr6:24596461-24596461	A	C	KIAA0319	nonsynonymous SNV
chr6:24596526-24596526	G	T	KIAA0319	nonsynonymous SNV
chr6:24596673-24596673	A	G	KIAA0319	nonsynonymous SNV
chr6:24596675-24596675	C	A	KIAA0319	nonsynonymous SNV
chr7:146805369-146805369	C	T	CNTNAP2	synonymous SNV
chr7:146805369-146805369	C	T	CNTNAP2	synonymous SNV
chr7:146829564-146829564	C	T	CNTNAP2	synonymous SNV
chr7:146829564-146829564	C	T	CNTNAP2	synonymous SNV
chr7:146829564-146829564	C	T	CNTNAP2	synonymous SNV
chr7:146997363-146997363	C	T	CNTNAP2	synonymous SNV
chr7:147815334-147815334	T	C	CNTNAP2	synonymous SNV
chr7:147844744-147844744	C	T	CNTNAP2	nonsynonymous SNV
chr7:147914500-147914500	C	T	CNTNAP2	nonsynonymous SNV
chr7:147926780-147926780	C	T	CNTNAP2	nonsynonymous SNV
chr7:147926851-147926851	G	A	CNTNAP2	nonsynonymous SNV
chr7:148112662-148112662	A	G	CNTNAP2	nonsynonymous SNV
chr15:51503139-51503139	C	G	CYP19A1	nonsynonymous SNV
chr15:51510859-51510868	GAAAAAAAAA	AGAAAAAAAAAC	CYP19A1	splicing
chr15:55731682-55731682	T	A	DYX1C1	nonsynonymous SNV
chr16:81654484-81654485	GC	AA	CMIP	nonframeshift substitution
chr16:81711917-81711917	G	-	CMIP	frameshift deletion
chr16:81711917-81711917	G	-	CMIP	frameshift deletion
chr16:81711917-81711917	G	-	CMIP	frameshift deletion
chr16:81711917-81711917	G	-	CMIP	frameshift deletion
chr16:81711917-81711917	G	-	CMIP	frameshift deletion
chr16:81711917-81711917	G	-	CMIP	frameshift deletion
chr16:81711917-81711917	G	-	CMIP	frameshift deletion
chr16:81711917-81711917	G	-	CMIP	frameshift deletion
chr16:81711917-81711917	G	-	CMIP	frameshift deletion
chr16:81711917-81711917	G	-	CMIP	frameshift deletion
chr16:81712240-81712240	-	CCC	CMIP	splicing
chr16:81712241-81712241	-	C	CMIP	splicing
chr16:81712241-81712241	-	CC	CMIP	splicing
chr16:81712241-81712241	-	CC	CMIP	splicing
chr16:81726757-81726757	T	C	CMIP	splicing
chr16:81733292-81733292	G	A	CMIP	synonymous SNV
chr16:81736235-81736235	A	G	CMIP	synonymous SNV
chr16:81736276-81736276	-	C	CMIP	splicing
chr16:81736276-81736276	-	C	CMIP	splicing
chr16:81736276-81736276	-	C	CMIP	splicing
chr16:81736276-81736276	-	C	CMIP	splicing
chr16:81736276-81736276	-	C	CMIP	splicing
chr16:81736276-81736276	-	C	CMIP	splicing
chr16:81736276-81736276	-	C	CMIP	splicing
chr16:81736276-81736276	-	C	CMIP	splicing
chr16:81736276-81736276	-	C	CMIP	splicing

Posizione genomica	Ref	Alt	Gene.refGene	ExonicFunc.refGene
chr16:81736276-81736276	-	C	CMIP	splicing
chr16:81736276-81736276	-	C	CMIP	splicing
chr16:81736276-81736276	-	C	CMIP	splicing
chr16:81736276-81736276	-	C	CMIP	splicing
chr16:81736276-81736276	-	C	CMIP	splicing
chr16:81736276-81736276	-	C	CMIP	splicing
chr16:81736276-81736276	-	C	CMIP	splicing
chr16:81736276-81736276	-	C	CMIP	splicing
chr21:47931424-47931424	G	A	DIP2A	synonymous SNV
chr21:47971564-47971564	A	G	DIP2A	nonsynonymous SNV
chr21:47971664-47971664	G	A	DIP2A	splicing
chr21:47974558-47974558	G	A	DIP2A	synonymous SNV
chr21:47978183-47978183	C	T	DIP2A	synonymous SNV
chr21:47980687-47980687	G	A	DIP2A	nonsynonymous SNV
chr21:47983840-47983840	G	T	DIP2A	stopgain
chr21:47987475-47987475	C	T	DIP2A	synonymous SNV

Tabella S4: Lista delle CNV ad alta confidenza (*High Confidence CNV*).

Posizione genomica	Lunghezza	Copy Number	ID
chr1:102656247-102853046	196799	1	DD317
chr1:104113158-104159716	46558	3	DD209
chr1:1132785-1291507	158722	3	DD004A
chr1:115466098-115527488	61390	1	DD277
chr1:1163881-1295682	131801	3	DD189
chr1:1216667-1249296	32629	1	DD277
chr1:1216667-1269613	52946	3	DD279
chr1:1216667-1295682	79015	1	DD033
chr1:1221489-1250268	28779	1	DD264
chr1:1384091-1431702	47611	3	DD168
chr1:142539541-142659799	120258	3	DD211
chr1:142539541-143543821	1004280	3	DD163
chr1:142539541-144858260	2318719	3	DD208
chr1:142544440-143543821	999381	3	DD331
chr1:143180768-143543821	363053	3	DD211
chr1:148911143-149100302	189159	3	DD208
chr1:149109771-149420252	310481	1	DD316
chr1:152561408-152585242	23834	0	DD004A
chr1:152561408-152585242	23834	0	DD004
chr1:1634999-1686116	51117	1	DD264
chr1:17594362-17615542	21180	3	DD317
chr1:195829222-195874047	44825	1	DD265
chr1:196748331-196801989	53658	1	DD209
chr1:196748331-196801989	53658	1	DD240
chr1:196823300-196887457	64157	1	DD184
chr1:21030916-21051116	20200	1	DD033
chr1:225391925-225459761	67836	3	DD069
chr1:227149097-227172988	23891	1	DD277
chr1:228334590-228363176	28586	1	DD264
chr1:228335177-228363176	27999	1	DD277
chr1:228522483-228562438	39955	3	DD189
chr1:2374849-2492276	117427	3	DD189
chr1:2404307-2528133	123826	1	DD033
chr1:247267248-247294095	26847	1	DD277
chr1:248751809-248789061	37252	1	DD203
chr1:25601115-25629950	28835	1	DD184
chr1:25601115-25629950	28835	1	DD208
chr1:25601115-25629950	28835	1	DD223
chr1:25601115-25629950	28835	1	DD296
chr1:25601115-25633145	32030	1	DD010
chr1:25712227-25735307	23080	3	DD119
chr1:5113763-5174078	60315	1	DD209
chr1:534238-569399	35161	3	DD208
chr1:60005016-60046727	41711	1	DD259
chr1:72767720-72805498	37778	0	DD159
chr1:72767720-72806589	38869	0	DD069
chr1:72767720-72806589	38869	0	DD119A
chr1:72767720-72806589	38869	0	DD119
chr1:72767720-72806589	38869	0	DD168
chr1:72767720-72806589	38869	0	DD203A

Posizione genomica	Lunghezza	Copy Number	ID
chr1:72767720-72806589	38869	0	DD203
chr1:72974014-72999691	25677	1	DD069
chr1:894315-1026359	132044	1	DD033
chr1:894573-1021285	126712	3	DD189
chr1:896806-1026359	129553	1	DD264
chr1:91726109-91850744	124635	1	DD277
chr1:9322120-9401282	79162	3	DD054
chr2:109597818-109629073	31255	1	DD319
chr2:110852875-110885507	32632	3	DD122
chr2:110863055-110897388	34333	3	DD319
chr2:110929173-110982530	53357	3	DD122
chr2:117781868-117929711	147843	3	DD331
chr2:13202216-13264451	62235	3	DD038
chr2:13202216-13280599	78383	3	DD264
chr2:132769893-133046249	276356	3	DD208
chr2:133015666-133045919	30253	3	DD211
chr2:133017230-133046249	29019	3	DD331
chr2:142034759-142092565	57806	3	DD247
chr2:178299726-178352746	53020	1	DD277
chr2:214752273-214810651	58378	1	DD309
chr2:233247668-233271287	23619	1	DD119
chr2:239037635-239060552	22917	1	DD033
chr2:241827861-241874349	46488	1	DD264
chr2:242853383-242928850	75467	1	DD069
chr2:242857051-243017051	160000	1	DD038
chr2:242857051-243017051	160000	1	DD240
chr2:242857051-243017051	160000	1	DD333
chr2:242857051-243028504	171453	1	DD226
chr2:242861964-242961355	99391	1	DD218
chr2:242923648-242944732	21084	1	DD317
chr2:242975841-243017051	41210	1	DD218
chr2:34700291-34736476	36185	0	DD168
chr2:34700291-34736476	36185	0	DD203
chr2:35582468-35614344	31876	1	DD223
chr2:45309232-45329884	20652	3	DD180
chr2:52747188-52785233	38045	1	DD119
chr2:52747188-52785233	38045	1	DD203
chr2:52751394-52785233	33839	0	DD203A
chr2:57399161-57445471	46310	3	DD119A
chr2:60438359-60485161	46802	1	DD223
chr2:76543406-76582787	39381	1	DD189
chr2:76543406-76582787	39381	1	DD317
chr2:82203791-82315179	111388	3	DD218
chr2:82203791-82315179	111388	3	DD331
chr2:82205627-82315179	109552	3	DD285
chr2:82225509-82315179	89670	3	DD209
chr2:87758948-87974800	215852	1	DD211
chr2:97519158-97543701	24543	1	DD033
chr3:100351696-100437397	85701	3	DD309
chr3:126707680-126741478	33798	1	DD033
chr3:126707680-126751579	43899	1	DD277
chr3:128381086-128403989	22903	1	DD231A
chr3:128383446-128410285	26839	3	DD189
chr3:147613061-147639101	26040	3	DD231A
chr3:155481319-155508536	27217	3	DD209

Posizione genomica	Lunghezza	Copy Number	ID
chr3:155481319-155508536	27217	3	DD303
chr3:162512480-162616738	104258	0	DD119A
chr3:162512480-162626334	113854	0	DD119
chr3:162512480-162626334	113854	0	DD203
chr3:162558896-162619779	60883	4	DD138
chr3:165041586-165081353	39767	1	DD279
chr3:165041586-165081353	39767	1	DD285
chr3:165041586-165081353	39767	1	DD307
chr3:175887314-175913366	26052	1	DD240
chr3:194237025-194268752	31727	3	DD223
chr3:195327155-195448070	120915	3	DD211
chr3:26044946-26145346	100400	1	DD291
chr3:30720057-31265052	544995	3	DD279
chr3:37340833-37533255	192422	3	DD223
chr3:4056290-4167766	111476	1	DD264
chr3:4144542-4167766	23224	1	DD231
chr3:44922034-44971510	49476	3	DD184
chr3:52521391-52559251	37860	1	DD033
chr3:75444447-75756270	311823	3	DD208
chr3:75803258-75830131	26873	3	DD211
chr3:75803258-75913810	110552	3	DD208
chr3:75875795-75902919	27124	3	DD211
chr3:84852538-84872804	20266	1	DD235
chr4:10212896-10232936	20040	0	DD119A
chr4:10212896-10232936	20040	0	DD119
chr4:10212896-10232936	20040	0	DD138
chr4:10212896-10232936	20040	0	DD159
chr4:10212896-10232936	20040	0	DD203A
chr4:10212896-10232936	20040	0	DD203
chr4:117379910-117504213	124303	1	DD296
chr4:131975827-132351485	375658	3	DD279
chr4:142550-171534	28984	1	DD307
chr4:145004468-145039799	35331	3	DD259
chr4:161048044-161070928	22884	1	DD231A
chr4:161048044-161070928	22884	1	DD231
chr4:161048044-161070928	22884	1	DD276
chr4:161048044-161070928	22884	1	DD285
chr4:1764925-1808987	44062	1	DD277
chr4:1795768-1811798	16030	1	DD033
chr4:190684454-190720566	36112	3	DD231
chr4:190802326-190942728	140402	3	DD208
chr4:190802326-190942728	140402	3	DD331
chr4:34781075-34823780	42705	0	DD138A
chr4:34781075-34823780	42705	0	DD138
chr4:35374961-35410960	35999	1	DD317
chr4:44966982-45001235	34253	3	DD220
chr4:4972091-5435852	463761	1	DD122
chr4:55023892-55048152	24260	1	DD319
chr4:60879952-60918911	38959	1	DD184
chr4:69448344-69490045	41701	0	DD119
chr4:70130098-70160463	30365	1	DD331
chr4:8983974-9651706	667732	3	DD208
chr4:947024-985928	38904	3	DD189
chr4:951409-1019053	67644	1	DD033
chr4:951688-985320	33632	3	DD319

Posizione genomica	Lunghezza	Copy Number	ID
chr4:951830-985928	34098	3	DD004A
chr4:995586-1018360	22774	3	DD004A
chr5:1052730-1094325	41595	1	DD277
chr5:166844833-166881177	36344	1	DD231A
chr5:166844833-166881177	36344	1	DD231
chr5:17599616-17704897	105281	1	DD319
chr5:180114438-180214894	100456	3	DD291
chr5:180376719-180427092	50373	0	DD168
chr5:32108076-32151119	43043	3	DD285
chr5:32108076-32154527	46451	3	DD203
chr5:32109122-32151119	41997	3	DD054
chr5:32109122-32151119	41997	3	DD223
chr5:710047-739160	29113	3	DD119
chr5:757146-795706	38560	1	DD220
chr5:757407-777854	20447	3	DD069
chr5:801550-825329	23779	3	DD211
chr5:8708019-8747793	39774	1	DD010
chr5:8708019-8747793	39774	1	DD122
chr5:89999487-90021499	22012	1	DD277
chr5:97047867-97096042	48175	1	DD211
chr5:97048658-97096042	47384	1	DD306
chr5:97187118-97224746	37628	1	DD211
chr6:102328936-102349175	20239	1	DD264
chr6:103736629-103758776	22147	0	DD119
chr6:117510203-117561179	50976	1	DD276
chr6:162615186-162695645	80459	1	DD122
chr6:163010615-163040132	29517	1	DD276
chr6:168342746-168596251	253505	3	DD247
chr6:18821397-18843857	22460	1	DD122
chr6:25749957-25779736	29779	1	DD010
chr6:31359019-31451988	92969	3	DD163
chr6:31377362-31451988	74626	3	DD259
chr6:3156671-3211787	55116	3	DD119A
chr6:31963844-32007917	44073	1	DD223
chr6:40297288-40327096	29808	1	DD333
chr6:48787991-49093760	305769	1	DD279
chr6:67011900-67046761	34861	1	DD159
chr6:67012750-67040313	27563	1	DD122
chr6:67012750-67040313	27563	1	DD223
chr6:67012750-67040313	27563	1	DD291
chr6:67012750-67049039	36289	1	DD208
chr6:67012750-67049039	36289	1	DD209
chr7:100993851-101127455	133604	3	DD311
chr7:111148162-111288374	140212	1	DD092
chr7:141765516-141798374	32858	1	DD317
chr7:141766485-141786841	20356	1	DD276
chr7:141766485-141794295	27810	1	DD306
chr7:1482751-1525110	42359	1	DD264
chr7:1483937-1527021	43084	1	DD277
chr7:1484300-1529300	45000	3	DD004A
chr7:155241562-155264332	22770	1	DD033
chr7:158235670-158445063	209393	3	DD218
chr7:2533903-2583482	49579	1	DD033
chr7:4789141-4856978	67837	1	DD033
chr7:4820940-4856978	36038	1	DD277

Posizione genomica	Lunghezza	Copy Number	ID
chr7:4821215-4856880	35665	1	DD264
chr7:61756396-61843044	86648	3	DD218
chr7:62030718-62140786	110068	3	DD054
chr8:11869157-12524364	655207	3	DD208
chr8:130859185-131107094	247909	3	DD333
chr8:137688951-137855798	166847	1	DD218
chr8:137688951-137855798	166847	1	DD226
chr8:137688951-137855798	166847	1	DD264
chr8:137688951-137855798	166847	1	DD277
chr8:137688951-137855798	166847	1	DD317
chr8:144800915-144829407	28492	1	DD264
chr8:144808688-144893505	84817	3	DD004A
chr8:144872423-144893505	21082	1	DD033
chr8:144872423-144893505	21082	1	DD277
chr8:144987934-145060068	72134	1	DD264
chr8:144990455-145065435	74980	1	DD277
chr8:144990461-145065728	75267	3	DD189
chr8:144993667-145024872	31205	3	DD279
chr8:144996758-145057656	60898	3	DD054
chr8:145096006-145735117	639111	1	DD033
chr8:145096141-145138351	42210	3	DD004A
chr8:145096163-145120354	24191	1	DD264
chr8:145162727-145223291	60564	1	DD264
chr8:145615891-145806563	190672	1	DD264
chr8:145625536-145776355	150819	3	DD189
chr8:145736819-145773779	36960	1	DD033
chr8:15948235-16021468	73233	1	DD122
chr8:39232331-39386714	154383	3	DD138
chr8:39311568-39386714	75146	3	DD004A
chr8:5661662-5748525	86863	1	DD291
chr8:6837290-6861593	24303	1	DD138A
chr8:7154055-7925018	770963	3	DD211
chr8:7302949-7339918	36969	3	DD119
chr8:7770399-7797906	27507	3	DD119
chr8:8261624-8437136	175512	1	DD298
chr8:87198662-87293797	95135	3	DD311
chr8:89688172-89711868	23696	1	DD119
chr9:10445149-10467074	21925	1	DD004A
chr9:1057256-1393763	336507	3	DD038
chr9:11993394-12122223	128829	1	DD307
chr9:127237729-127269149	31420	1	DD033
chr9:13307629-13338731	31102	1	DD163
chr9:139219081-139276923	57842	3	DD004A
chr9:139228945-139257142	28197	1	DD264
chr9:139228945-139276923	47978	1	DD033
chr9:139378453-139444637	66184	3	DD004A
chr9:139379364-139426765	47401	3	DD189
chr9:139391682-139426765	35083	1	DD033
chr9:139564680-139592058	27378	3	DD004A
chr9:139566734-139591977	25243	3	DD189
chr9:139739872-139754515	14643	3	DD189
chr9:139835382-139935148	99766	3	DD189
chr9:140173379-140218591	45212	1	DD277
chr9:140320715-140397226	76511	1	DD277
chr9:140325554-140392689	67135	3	DD189

Posizione genomica	Lunghezza	Copy Number	ID
chr9:205964-407688	201724	3	DD038
chr9:38244221-38323006	78785	3	DD211
chr9:38646995-38774065	127070	1	DD285
chr9:414857-706654	291797	3	DD038
chr9:44462956-44892044	429088	3	DD208
chr9:4480200-4576851	96651	1	DD332
chr9:5162710-5252184	89474	1	DD004
chr9:66246366-69827686	3581320	3	DD163
chr9:66324996-70175679	3850683	3	DD208
chr9:712994-841285	128291	3	DD038
chr9:7733025-7772602	39577	1	DD010
chr9:79783751-79822026	38275	1	DD277
chr9:79932585-79973335	40750	1	DD277
chr9:843763-990085	146322	3	DD038
chr9:990868-1050981	60113	3	DD038
chr10:127575631-127616606	40975	3	DD163
chr10:127575631-127616606	40975	3	DD208
chr10:132912142-133043625	131483	1	DD069
chr10:135240371-135374888	134517	3	DD211
chr10:135240371-135379029	138658	1	DD285
chr10:135242873-135289883	47010	3	DD054
chr10:135242873-135379029	136156	3	DD223
chr10:16473984-16498058	24074	1	DD209
chr10:42355302-42400614	45312	3	DD208
chr10:42602655-42739181	136526	3	DD208
chr10:42794125-42925665	131540	3	DD168
chr10:43089965-43186721	96756	3	DD168
chr10:46952667-47606057	653390	3	DD138
chr10:46962104-46996515	34411	1	DD033
chr10:47000027-47090351	90324	1	DD033
chr10:47090175-47264810	174635	3	DD208
chr10:47131754-47756036	624282	1	DD033
chr10:47538589-47646751	108162	1	DD231A
chr10:47538589-47703946	165357	3	DD211
chr10:47538589-47706370	167781	3	DD331
chr10:47610225-47706370	96145	3	DD138
chr10:47619848-47667164	47316	3	DD054
chr10:47641862-47756036	114174	3	DD317
chr10:48314669-50533776	2219107	1	DD033
chr10:50535008-50819432	284424	1	DD033
chr10:50821191-51785598	964407	1	DD033
chr10:56448886-56469202	20316	3	DD298
chr10:68076803-68115498	38695	1	DD033
chr10:68076803-68115498	38695	1	DD285
chr10:6811851-6836760	24909	3	DD319
chr11:103006582-103091449	84867	1	DD277
chr11:1836476-1862331	25855	1	DD264
chr11:2408022-2429109	21087	1	DD264
chr11:2415292-2442314	27022	3	DD189
chr11:2424069-2445256	21187	1	DD033
chr11:40613584-40742299	128715	1	DD303
chr11:533357-563598	30241	1	DD033
chr11:66609971-66639202	29231	1	DD033
chr11:82593917-82702281	108364	3	DD235
chr11:89852034-90973763	1121729	3	DD279

Posizione genomica	Lunghezza	Copy Number	ID
chr11:9834123-10218423	384300	1	DD279
chr12:10571622-10592659	21037	1	DD033
chr12:11220670-11248858	28188	3	DD163
chr12:11224381-11244653	20272	3	DD168
chr12:11507049-11543432	36383	3	DD138A
chr12:11507321-11527351	20030	3	DD138
chr12:11512505-11536689	24184	1	DD306
chr12:123433185-123472168	38983	1	DD033
chr12:123444203-123471065	26862	1	DD277
chr12:131730411-131825321	94910	1	DD279
chr12:132379598-132426506	46908	1	DD277
chr12:133707090-133730705	23615	3	DD189
chr12:21006115-21408845	402730	1	DD163
chr12:27160662-27181391	20729	3	DD220
chr12:31296219-31386971	90752	3	DD306
chr12:31301888-31384444	82556	3	DD168
chr12:31305020-31406907	101887	3	DD138A
chr12:31310783-31406907	96124	3	DD247
chr12:32008660-32061168	52508	3	DD296
chr12:5711911-6081946	370035	3	DD296
chr12:73116514-73137647	21133	1	DD316
chr12:73949811-73997559	47748	1	DD264
chr12:8000978-8115176	114198	1	DD279
chr12:8008028-8115176	107148	1	DD122
chr12:8043600-8080159	36559	3	DD226
chr12:80632645-80731396	98751	1	DD184
chr12:8090505-8115176	24671	3	DD226
chr12:86950738-87007106	56368	1	DD209
chr12:9632783-9706571	73788	3	DD069
chr12:9633258-9729311	96053	3	DD138
chr13:114503858-114544722	40864	1	DD277
chr13:19401119-19715960	314841	1	DD279
chr13:24014194-24064269	50075	3	DD163
chr13:41430586-41468705	38119	1	DD235
chr13:57757118-57785520	28402	1	DD168
chr13:72866419-72984051	117632	1	DD231
chr13:72866419-72992226	125807	1	DD231A
chr14:103566560-103600074	33514	1	DD277
chr14:105931063-105968911	37848	1	DD277
chr14:19366709-20425051	1058342	3	DD208
chr14:19536544-19573344	36800	3	DD119
chr14:19607575-19636879	29304	3	DD159
chr14:19636879-20423059	786180	3	DD331
chr14:20209916-20337810	127894	3	DD119
chr14:20215941-20416499	200558	3	DD033
chr14:24921862-24987679	65817	3	DD277
chr14:41610224-41660172	49948	1	DD138
chr14:41610224-41660172	49948	1	DD203A
chr14:41610224-41670687	60463	1	DD203
chr14:42986382-43236053	249671	3	DD279
chr14:43829049-44247741	418692	3	DD092
chr14:44823308-44913238	89930	3	DD311
chr14:45183850-45224071	40221	1	DD189
chr14:48239789-48272674	32885	1	DD092
chr14:48239789-48272674	32885	1	DD235

Posizione genomica	Lunghezza	Copy Number	ID
chr14:48758500-48861803	103303	1	DD235
chr14:73997304-74019674	22370	3	DD004
chr14:86285503-86312977	27474	1	DD033
chr14:86459102-86484541	25439	1	DD122
chr15:100385477-100498724	113247	1	DD209
chr15:20603796-22179652	1575856	3	DD159
chr15:20797153-20833831	36678	3	DD069
chr15:21148522-22582149	1433627	4	DD208
chr15:21199990-21234252	34262	3	DD004A
chr15:22318872-22436665	117793	1	DD296
chr15:22384277-22420970	36693	1	DD317
chr15:22552748-22582149	29401	1	DD276
chr15:22560396-22582149	21753	1	DD296
chr15:22560396-22582149	21753	1	DD317
chr15:24634797-24665846	31049	3	DD231A
chr15:24694134-24722573	28439	3	DD168
chr15:36274022-36359839	85817	3	DD069
chr15:36360651-36470443	109792	3	DD069
chr15:36471670-36808006	336336	3	DD069
chr15:57643881-57731384	87503	3	DD316
chr15:57739523-57776671	37148	3	DD316
chr15:62197601-62361107	163506	1	DD277
chr15:87832806-87866035	33229	1	DD138
chr15:87832806-87866035	33229	1	DD331
chr15:91994436-92233988	239552	3	DD208
chr15:94152353-94242386	90033	1	DD004A
chr15:94780241-94828466	48225	1	DD316
chr16:10594180-10642072	47892	1	DD331
chr16:15022065-15059248	37183	3	DD119A
chr16:15022065-15128416	106351	3	DD119
chr16:15125441-15179788	54347	3	DD311
chr16:15489598-15586074	96476	3	DD119A
chr16:1569934-1592779	22845	1	DD033
chr16:15722934-15765403	42469	3	DD119
chr16:15800158-16284191	484033	3	DD119A
chr16:15839439-15960226	120787	3	DD119
chr16:16139714-16200716	61002	3	DD119
chr16:16265935-16362651	96716	3	DD119
chr16:1812884-1842305	29421	3	DD189
chr16:1813658-1842305	28647	3	DD054
chr16:2027115-2053042	25927	1	DD277
chr16:2027175-2087885	60710	3	DD189
chr16:2027994-2087885	59891	1	DD033
chr16:2121617-2296877	175260	3	DD189
chr16:2131754-2287867	156113	1	DD033
chr16:2138600-2161824	23224	1	DD264
chr16:22649699-22704258	54559	1	DD184
chr16:2281829-2345664	63835	1	DD264
chr16:2510648-2552119	41471	1	DD277
chr16:3169532-3200883	31351	1	DD264
chr16:32351177-33959975	1608798	3	DD163
chr16:32467367-33965582	1498215	3	DD211
chr16:32506957-32657013	150056	1	DD333
chr16:32522606-32604785	82179	3	DD298
chr16:32629757-32657013	27256	3	DD298

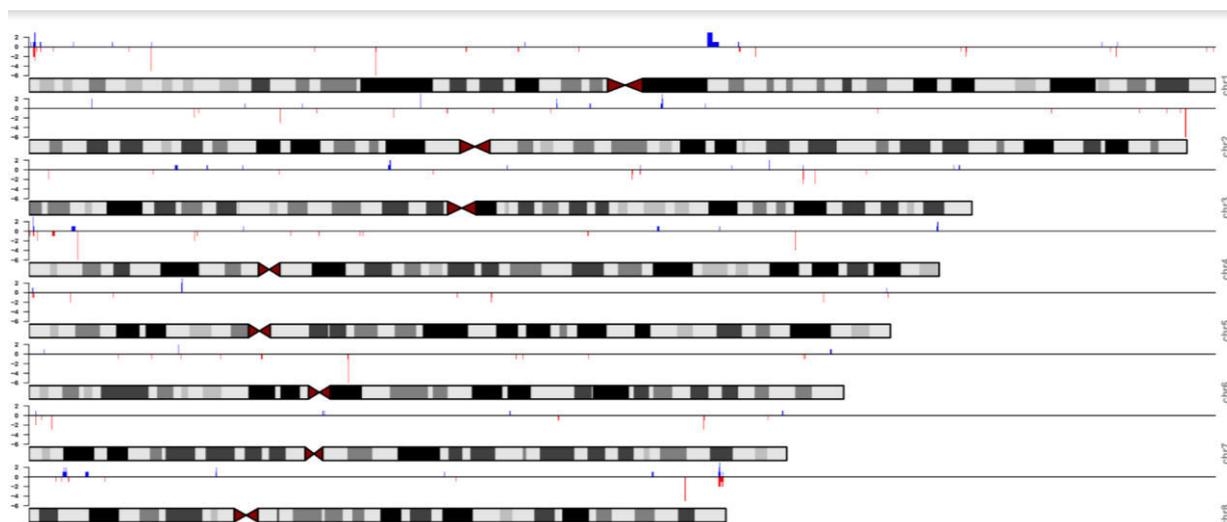
Posizione genomica	Lunghezza	Copy Number	ID
chr16:4410973-4432877	21904	1	DD264
chr16:46457807-46497305	39498	3	DD331
chr16:50321878-50343468	21590	1	DD033
chr16:588584-781623	193039	3	DD189
chr16:597527-767423	169896	3	DD004A
chr16:6503793-6604431	100638	1	DD231
chr16:67183307-67234866	51559	1	DD033
chr16:70011606-70196441	184835	3	DD211
chr16:74112313-74228281	115968	3	DD296
chr16:75545479-75572495	27016	3	DD291
chr16:814942-847075	32133	1	DD277
chr16:814942-864123	49181	1	DD264
chr16:814942-864123	49181	3	DD189
chr16:834714-858148	23434	1	DD033
chr16:88779739-88804443	24704	3	DD189
chr16:89161934-89220846	58912	1	DD231A
chr16:89760652-89790034	29382	1	DD264
chr17:21407795-21454676	46881	3	DD296
chr17:27893545-27949876	56331	1	DD033
chr17:34443811-34477867	34056	3	DD138A
chr17:34443811-34490448	46637	3	DD203
chr17:44167366-44222424	55058	3	DD163
chr17:44167366-44434779	267413	3	DD331
chr17:44171207-44368830	197623	3	DD208
chr17:66914191-66988666	74475	3	DD054
chr17:67136818-67190118	53300	1	DD054
chr17:73870612-73916185	45573	1	DD264
chr17:74448857-74470614	21757	1	DD264
chr17:74448857-74470614	21757	3	DD054
chr17:74448857-74473032	24175	1	DD033
chr17:76100570-76128899	28329	1	DD277
chr17:79409021-79430806	21785	1	DD264
chr17:79409573-79430201	20628	1	DD033
chr17:79409834-79430806	20972	3	DD189
chr17:79479956-79519221	39265	3	DD189
chr17:79857812-79920300	62488	3	DD004A
chr17:79961620-79987948	26328	3	DD054
chr17:79966961-80065351	98390	3	DD189
chr17:80006427-80051590	45163	3	DD004A
chr17:80159447-80197926	38479	1	DD033
chr17:80159447-80197926	38479	1	DD264
chr17:80176641-80198652	22011	3	DD189
chr18:14253170-14324745	71575	3	DD208
chr18:14460129-14524935	64806	3	DD208
chr18:15166155-15209181	43026	3	DD163
chr18:15166155-15209181	43026	3	DD208
chr18:15166155-15209181	43026	3	DD331
chr18:29429596-29477801	48205	1	DD277
chr18:76734664-76755461	20797	1	DD277
chr19:1219345-1255520	36175	3	DD189
chr19:1242460-1269171	26711	1	DD033
chr19:14130237-14153528	23291	3	DD296
chr19:1456150-1484126	27976	3	DD189
chr19:20596024-20717673	121649	1	DD203
chr19:20834714-20988708	153994	3	DD319

Posizione genomica	Lunghezza	Copy Number	ID
chr19:43372386-43420568	48182	1	DD279
chr19:43480854-43533562	52708	1	DD279
chr19:43703965-43744732	40767	1	DD317
chr19:43703965-43744732	40767	1	DD332
chr19:45641801-45683053	41252	1	DD033
chr19:53518590-53541112	22522	3	DD218
chr19:53518590-53545977	27387	3	DD208
chr19:53909822-53948212	38390	1	DD309
chr19:54726023-54746051	20028	3	DD138
chr19:55301229-55338667	37438	1	DD235
chr19:6957080-7013359	56279	1	DD240
chr20:1562873-1591985	29112	0	DD203
chr20:29419620-29653447	233827	3	DD331
chr20:29566283-29653447	87164	3	DD163
chr20:29569482-29653447	83965	3	DD208
chr20:3853190-5090075	1236885	1	DD189
chr20:60881267-60937927	56660	1	DD277
chr20:62160192-62202150	41958	1	DD277
chr20:62160896-62190235	29339	1	DD264
chr20:62160896-62223988	63092	1	DD033
chr20:62163984-62201702	37718	3	DD189
chr20:62175648-62200738	25090	3	DD279
chr20:62192969-62217074	24105	1	DD264
chr20:62312070-62340134	28064	1	DD264
chr20:62317159-62385479	68320	1	DD033
chr20:62319048-62374281	55233	3	DD054
chr20:62319118-62374281	55163	1	DD277
chr20:62319885-62374281	54396	3	DD189
chr20:62570844-62597666	26822	1	DD033
chr21:10848746-10892052	43306	3	DD208
chr21:14656752-14707306	50554	1	DD184
chr21:14758756-14778904	20148	1	DD184
chr21:20059375-20082121	22746	3	DD138
chr21:47530602-47581943	51341	1	DD277
chr21:9411773-10089889	678116	3	DD331
chr21:9411773-9523367	111594	3	DD332
chr21:9437719-10032416	594697	3	DD208
chr21:9437719-9487338	49619	3	DD163
chr21:9471017-9494520	23503	3	DD138
chr21:9555540-9705758	150218	3	DD163
chr21:9944047-10032416	88369	3	DD163
chr22:25684877-25739750	54873	3	DD033
chr22:25781394-25908383	126989	3	DD033
chr22:42908926-42945707	36781	3	DD038
chr22:42918711-42948995	30284	3	DD203
chr22:50613093-50660468	47375	3	DD004A
chr22:50615835-50660078	44243	1	DD033

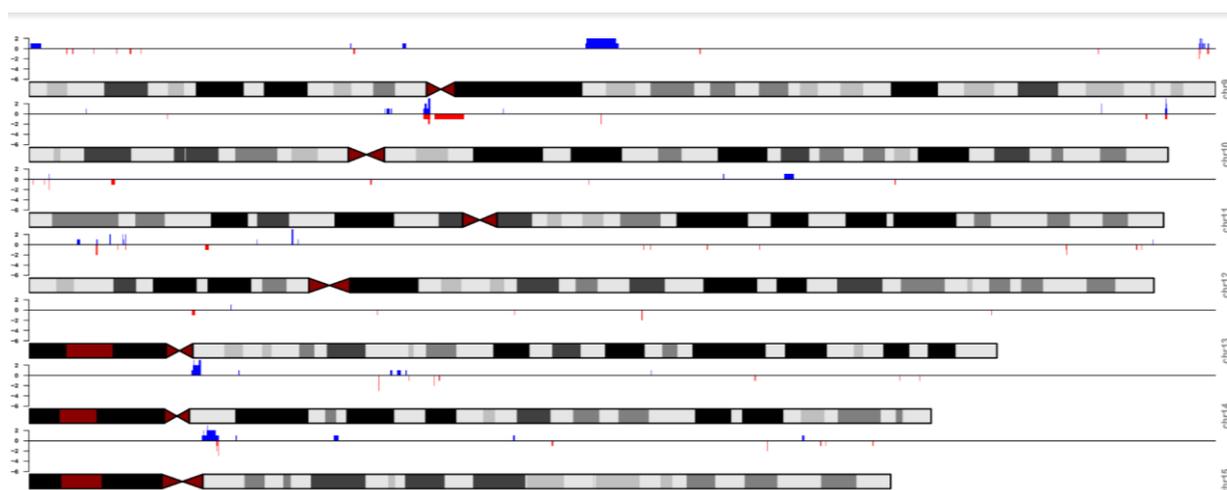
7. FIGURE SUPPLEMENTARI

Figura S1: Visualizzazione della localizzazione delle CNV ad alta confidenza sui singoli cromosomi (A, B, C). In blu le duplicazioni mentre in rosso le delezioni.

A) CNV localizzate sui cromosomi 1, 2, 3, 4, 5, 6, 7 e 8.



B) CNV localizzate sui cromosomi 9, 10, 11, 12, 13, 14 e 15.



C) CNV localizzate sui cromosomi 16, 17, 18, 19, 20, 21 e 22.

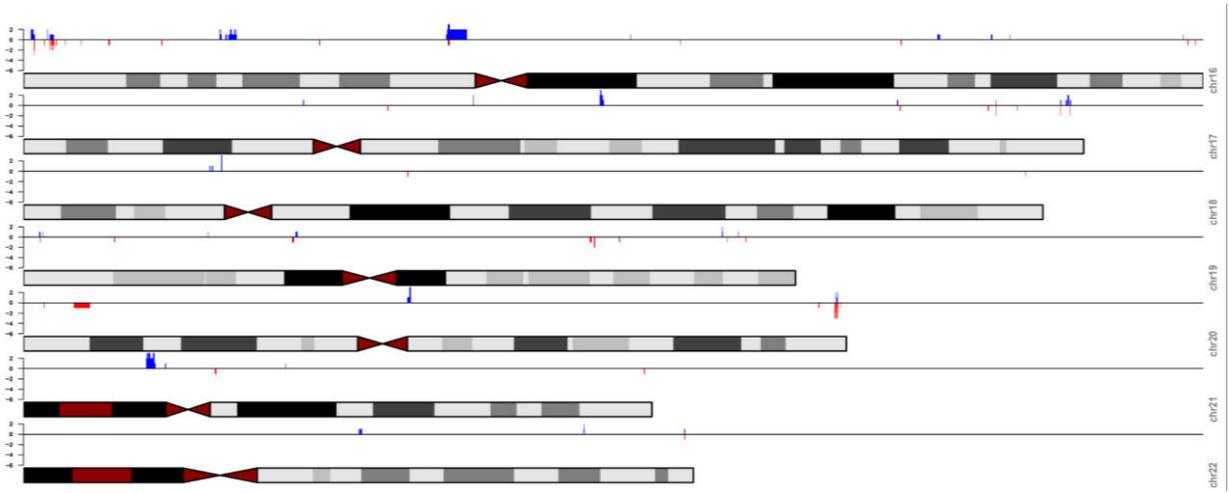
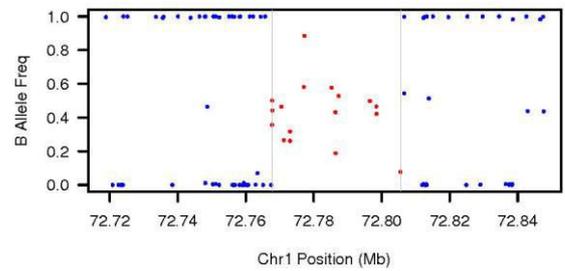
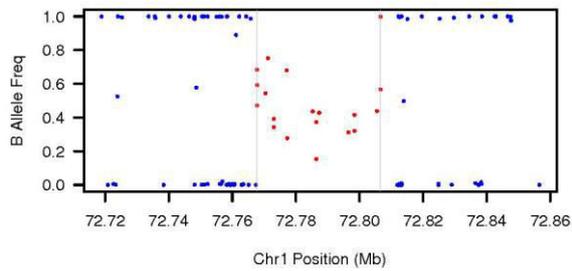
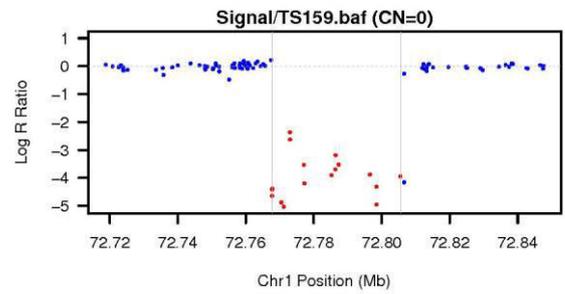
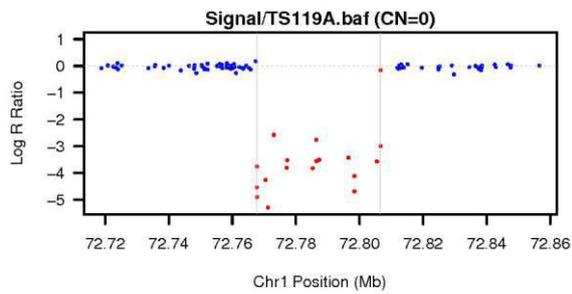
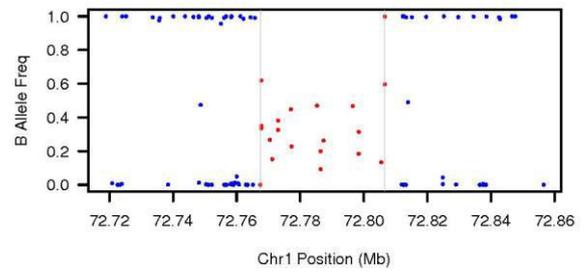
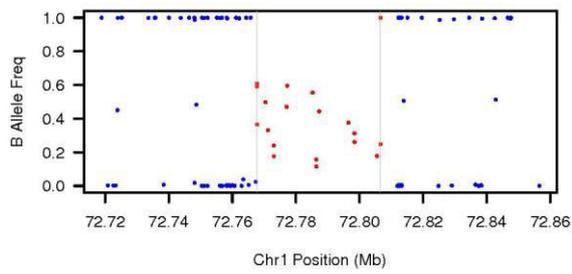
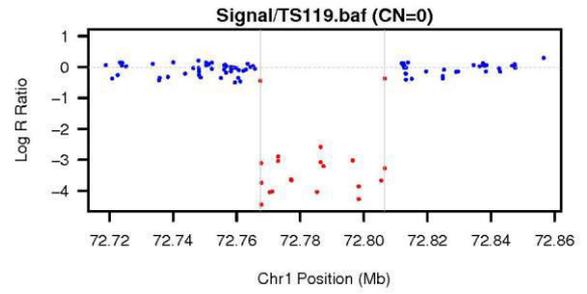
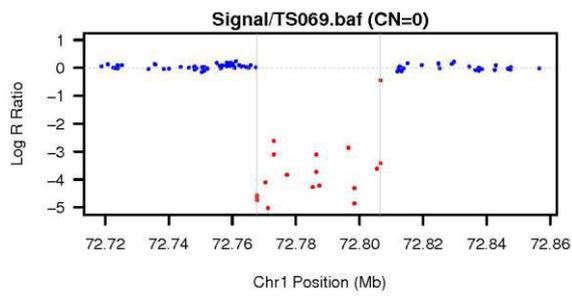
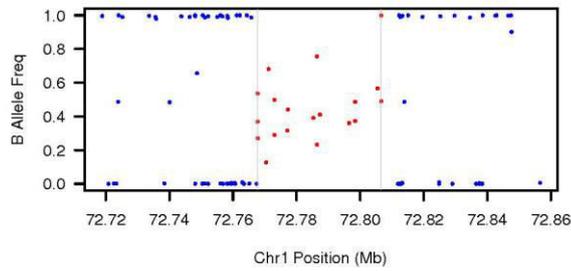
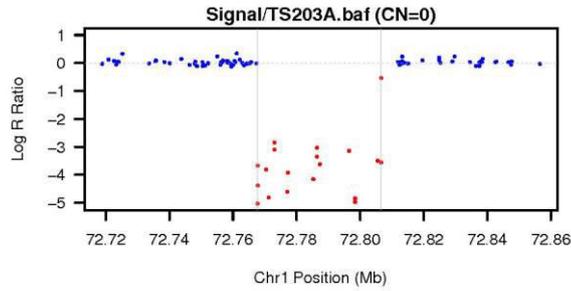
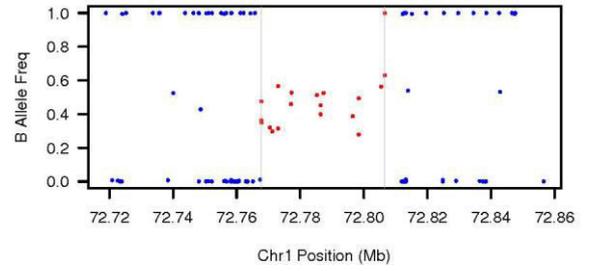
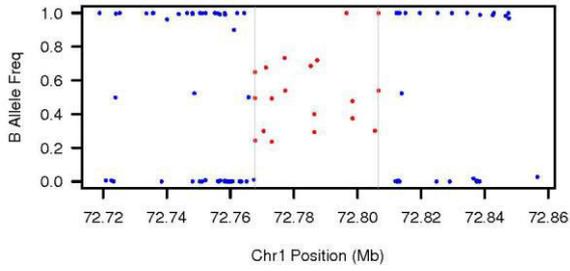
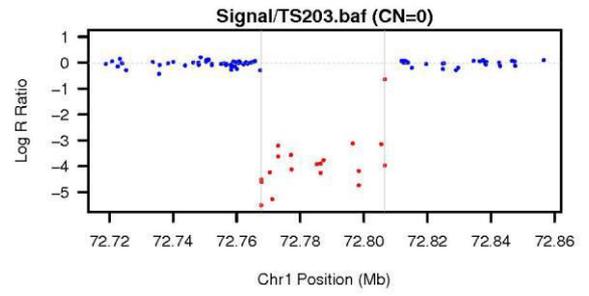
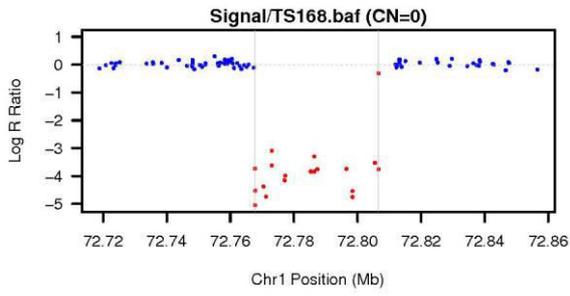


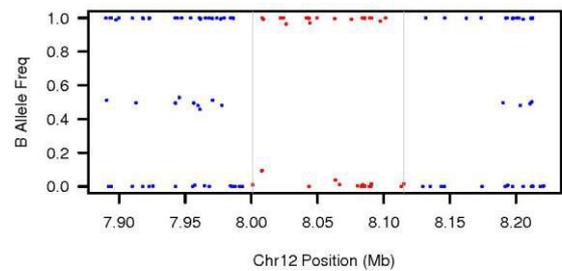
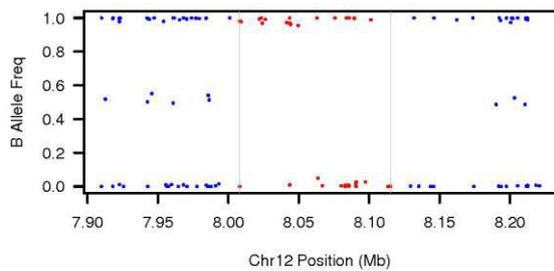
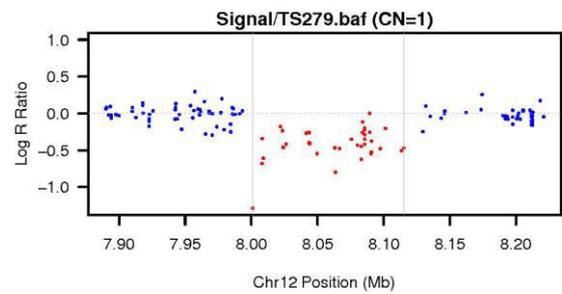
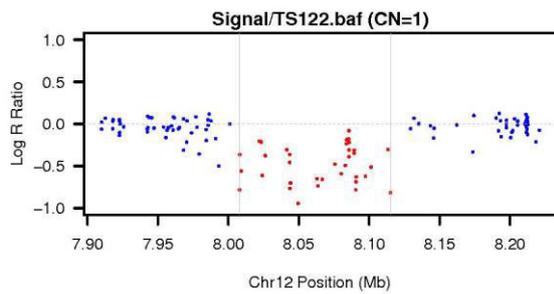
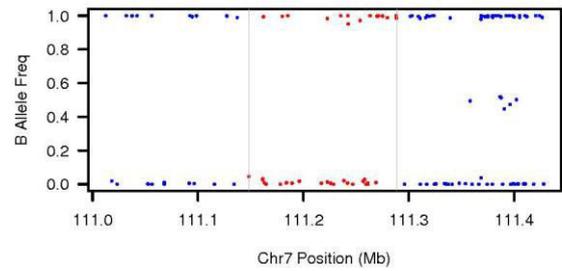
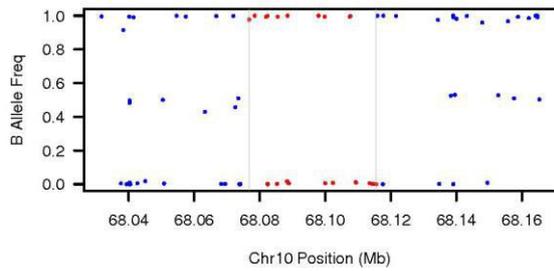
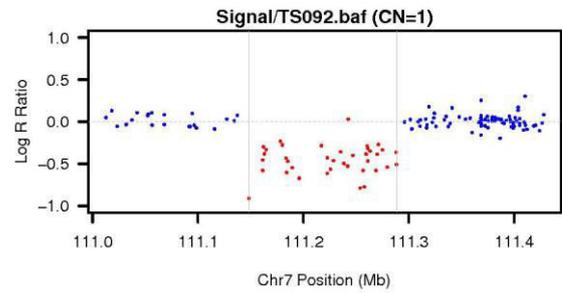
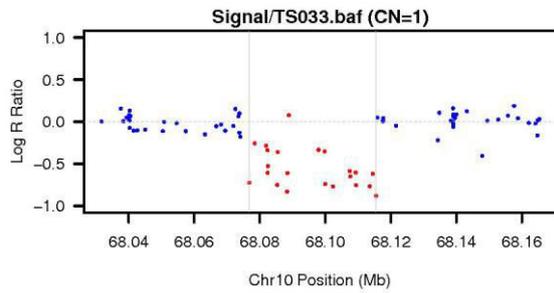
Figura S2: Plot dei dati LRR e BAF, partendo dai dati del PennCNV, delle CNV riportate nella Tabella 4.5, Sezione 4 - Risultati.

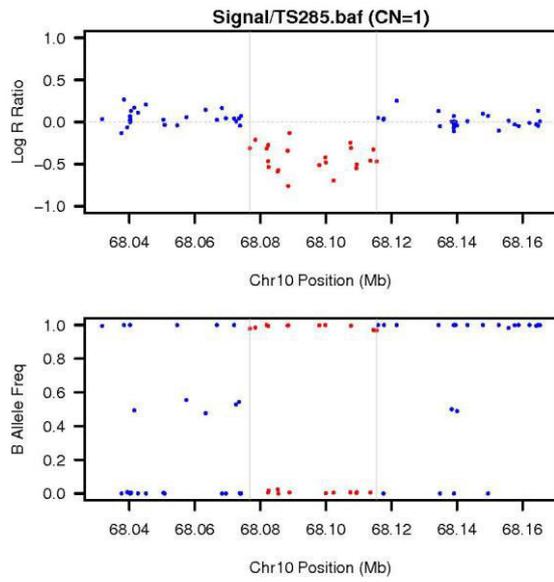
Copy Number 0



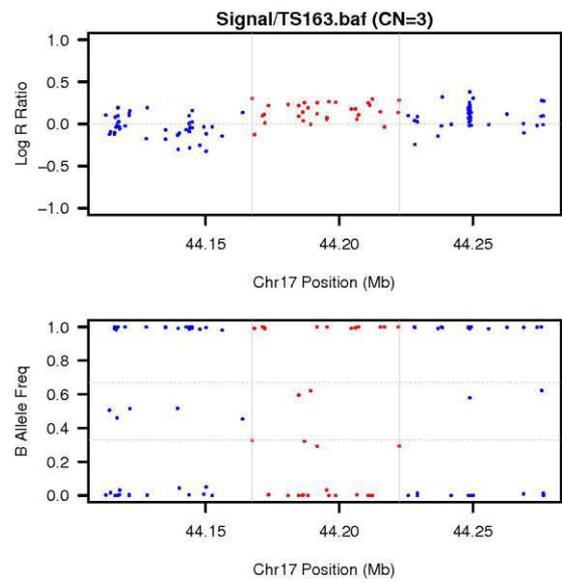
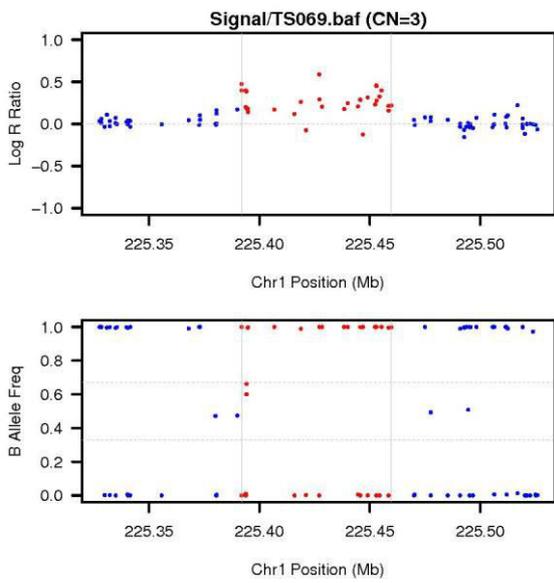


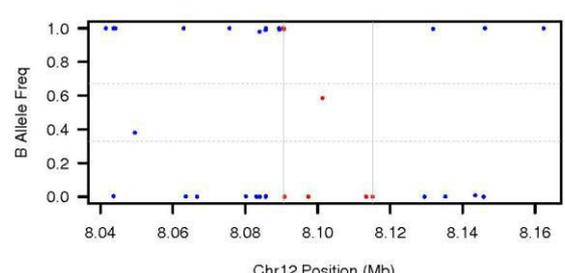
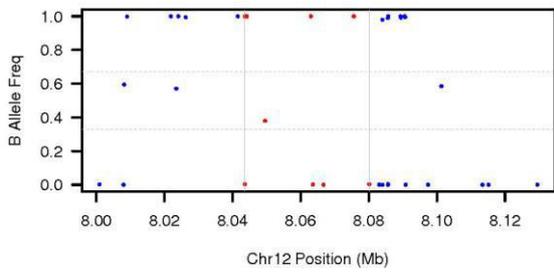
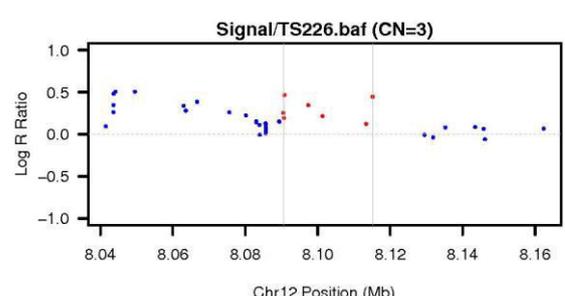
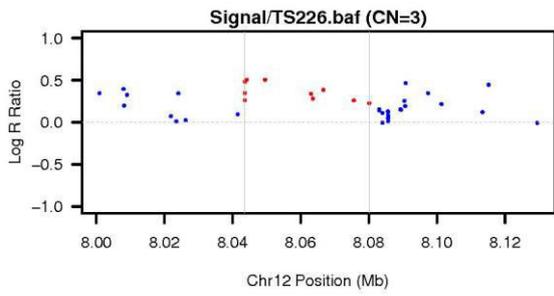
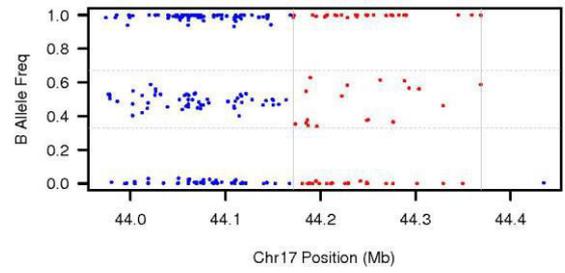
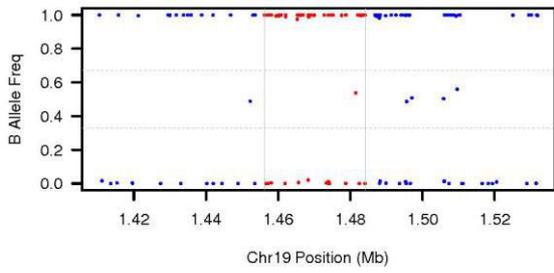
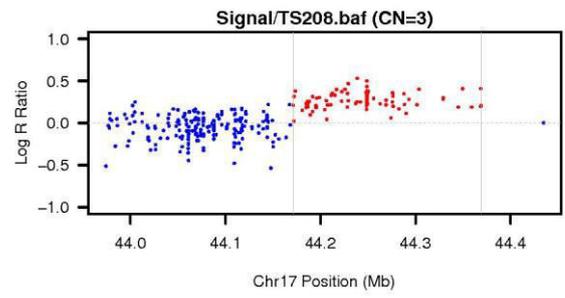
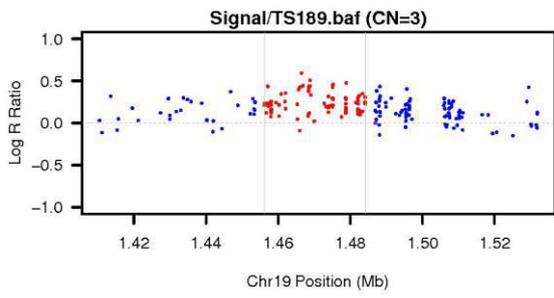
Copy Number 1





Copy Number 3





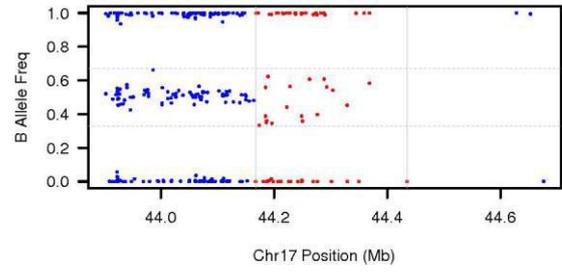
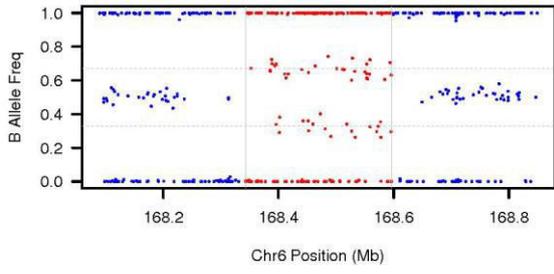
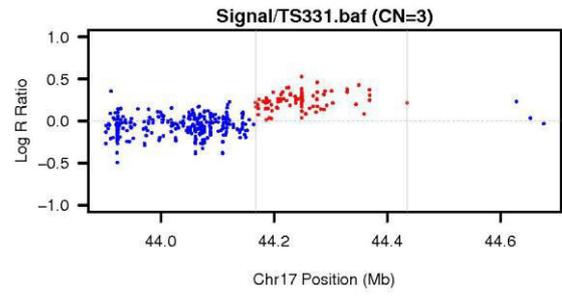
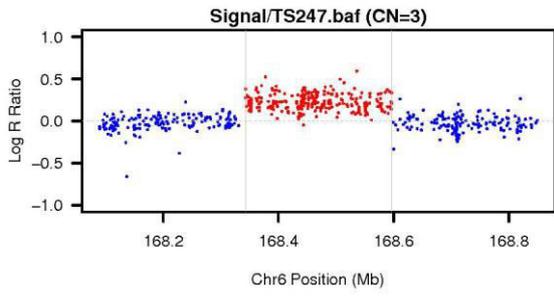
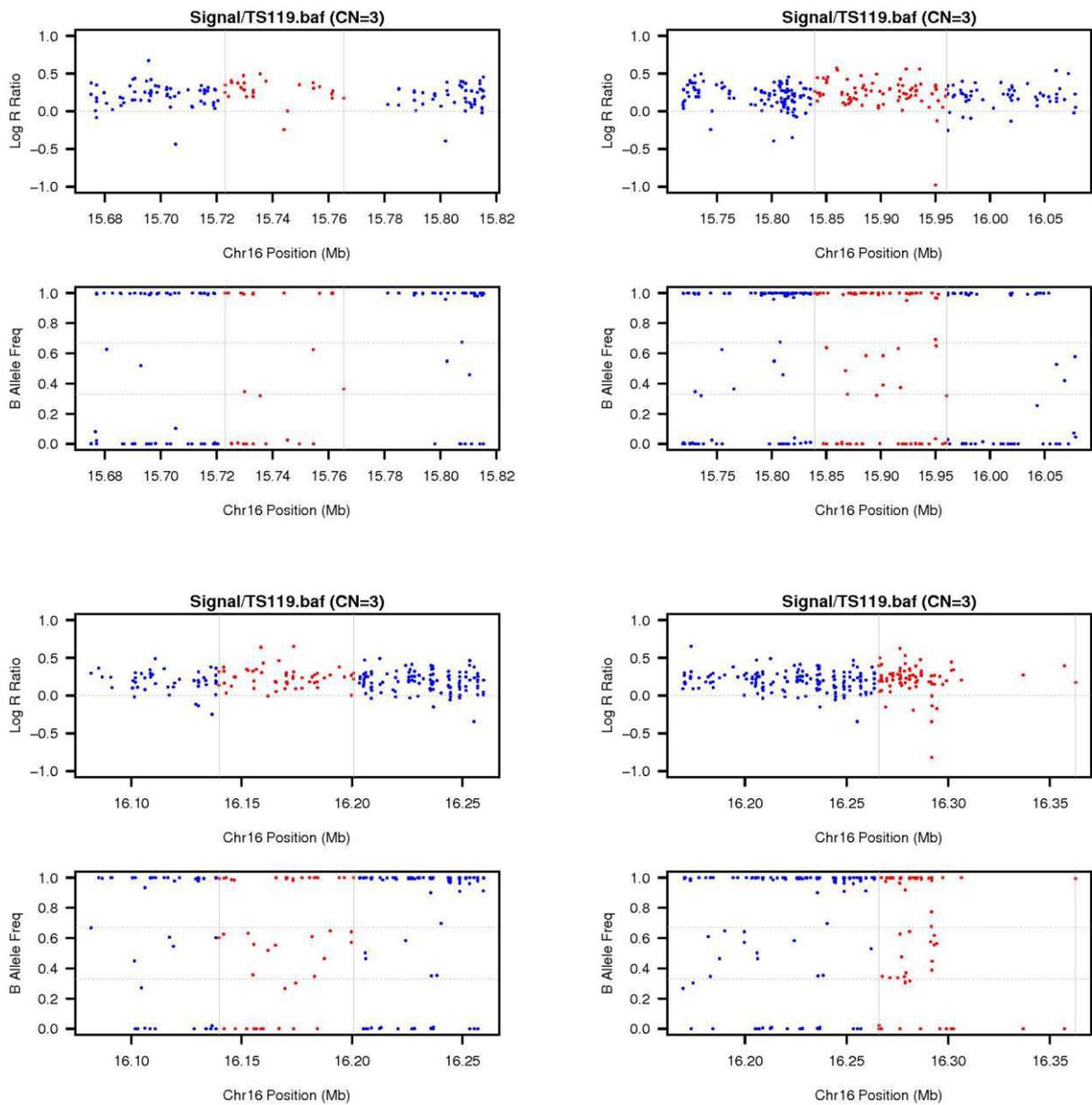


Figura S3: Plot dei dati LRR e BAF, partendo dai dati del PennCNV, delle CNV riportate nella Tabella 4.6, Sezione 4 - Risultati.

Copy Number 3



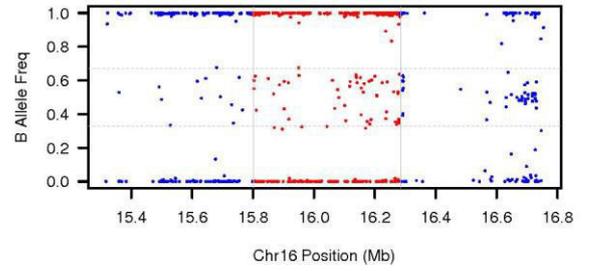
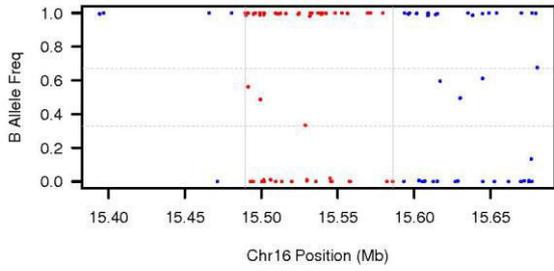
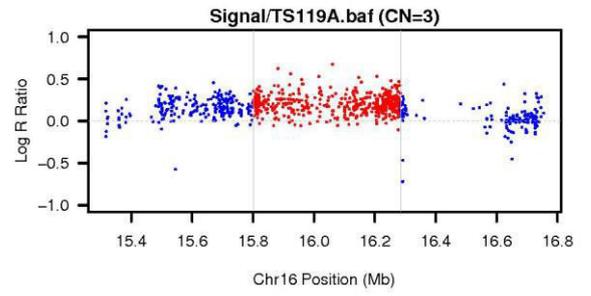
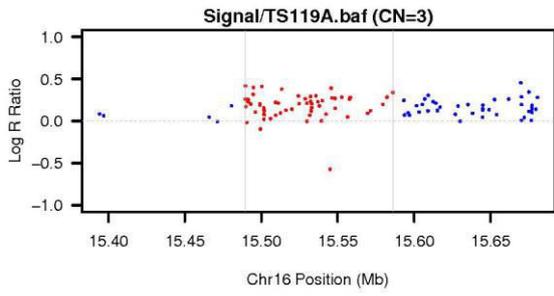
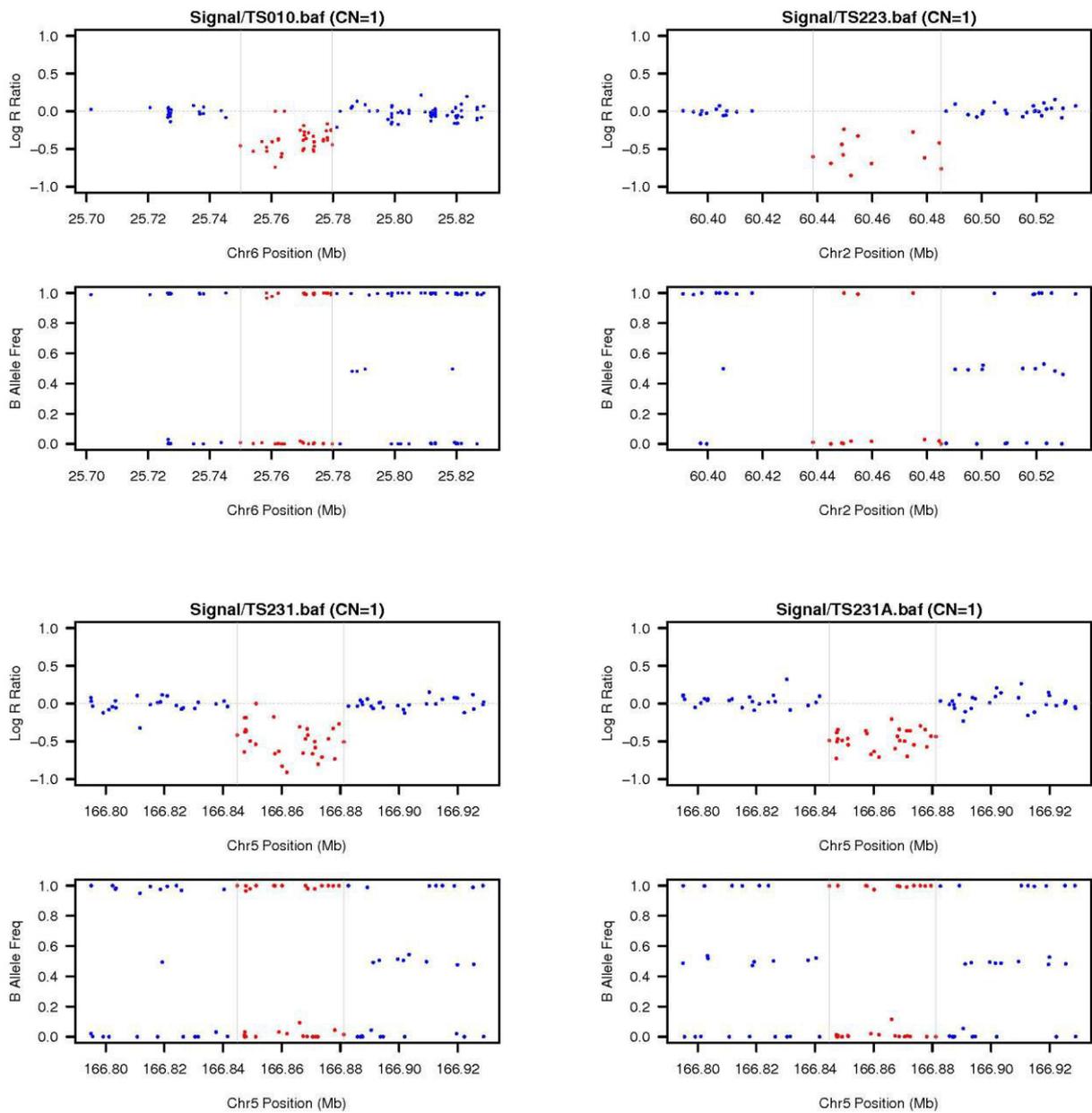
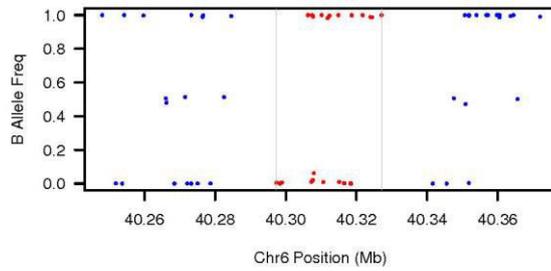
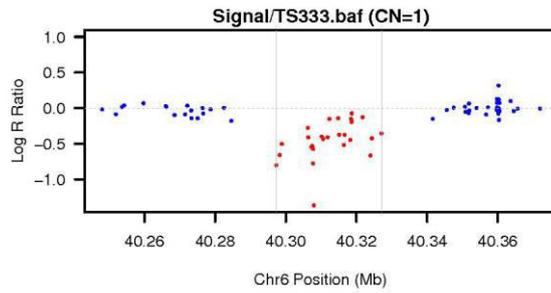
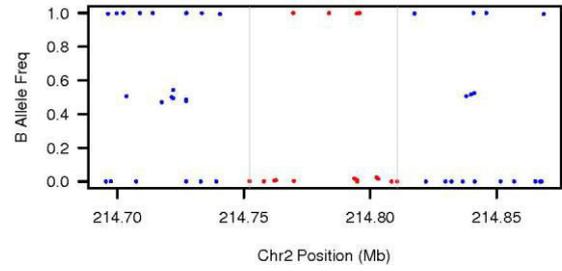
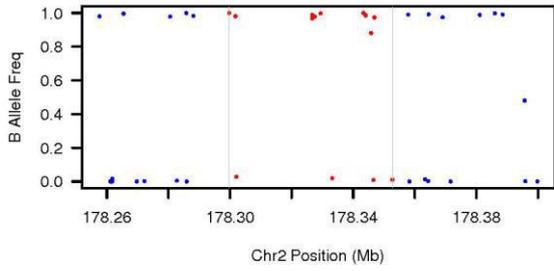
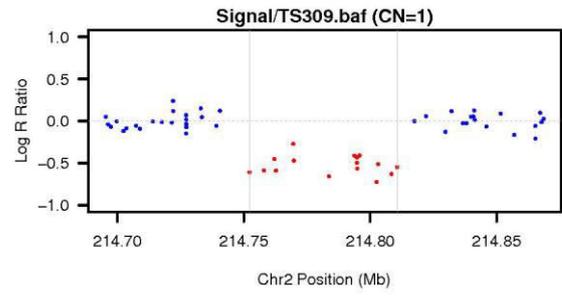
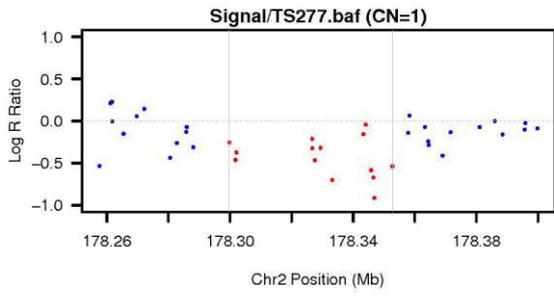


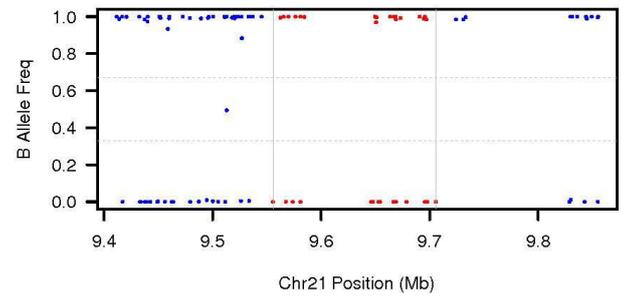
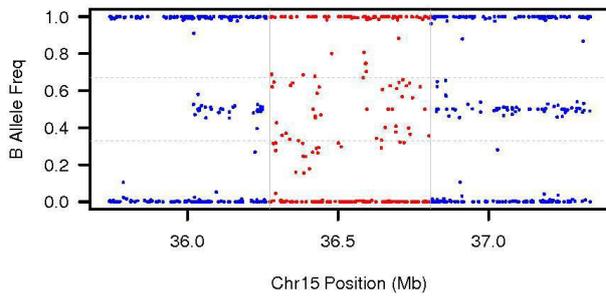
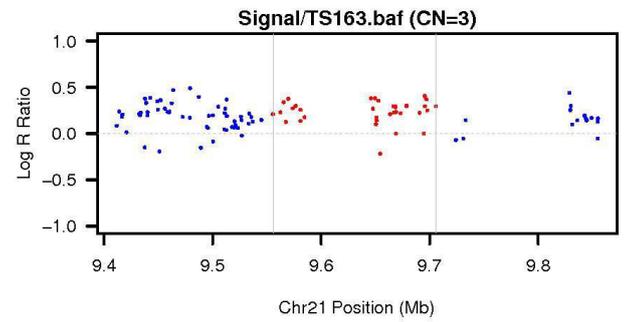
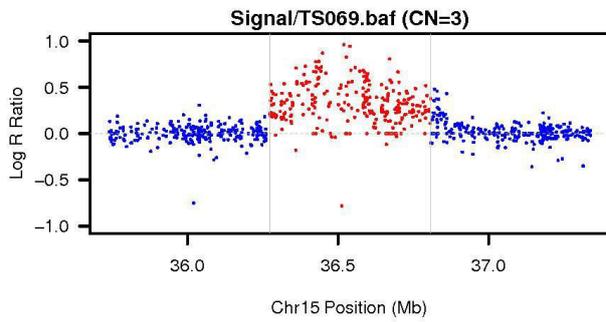
Figura S4: Plot dei dati LRR e BAF, partendo dai dati del PennCNV, delle CNV riportate nella Tabella 4.7, Sezione 4 - Risultati.

Copy Number 1





Copy Number 3



8. BIBLIOGRAFIA

- Adler WT, Platt MP, Mehlhorn AJ, et al. (2013) *Position of Neocortical Neurons Transfected at Different Gestational Ages with shRNA Targeted against Candidate Dyslexia Susceptibility Genes*. PLoS ONE, 8:e65179.
- Alarcon M, Abrahams B, Stone J, et al. (2008) *Linkage, association, and gene expression analyses identify CNTNAP2 as an autism-susceptibility gene*. Am J Hum Genet, 82:150-159.
- Andrews W, Barber M, Hernandez-Miranda LR, et al. (2008) *The role of Slit-Robo signaling in the generation, migration and morphological differentiation of cortical interneurons*. Dev Biol, 313:648-658.
- Andrews W, Liapi A, Plachez C, et al. (2006) *Robo1 regulates the development of major axon tracts and interneuron migration in the forebrain*. Development, 133:2243-2252.
- Anney R, Klei L, Pinto D, et al. (2012) *Individual common variants exert weak effects on the risk for autism spectrum disorders*. Hum Mol Genet, 21:4781-4792.
- Anthoni H, Sucheston L, Lewis B, et al. (2012) *The Aromatase Gene CYP19A1: Several Genetic and Functional Lines of Evidence Supporting a Role in Reading, Speech and Language*. Behav Genet, 42:509-527.
- Anthoni H, Zucchelli M, Matsson H, et al. (2007) *A locus on 2p12 containing the co-regulated MRPL19 and C2ORF3 genes is associated to dyslexia*. Hum Mol Genet 16: 667-677.
- Bai S, Ghoshal K, Jacob ST, et al. (2006) *Identification of T-cadherin as a novel target of DNA methyltransferase 3B and its role in the suppression of nerve growth factor-mediated neurite outgrowth in PC12 cells*. J Biol Chem, 281:13604-13611.
- Bakkaloglu B, O'Roak BJ, Louvi A, et al. (2008) *Molecular cytogenetic analysis and resequencing of contactin associated protein-like 2 in autism spectrum disorders*. Am J Hum Genet, 82:165-173.
- Barbiero C, Lonciari I, Montico M, et al. (2012) *The submerged dyslexia iceberg: how many school children are not diagnosed? Results from an Italian study*. PLoS One, 7:e48082.
- Bartlett CW, Flax JF, Logue MW, et al. (2002) *A Major Susceptibility Locus for Specific Language Impairment Is Located on 13q21*. Am J Hum Genet, 71:45-55.
- Bates TC, Luciano M, Castles A, et al. (2007) *Replication of reported linkages for dyslexia and spelling and suggestive evidence for novel regions on chromosomes 4 and 17*. Eur J Hum Genet, 15:194-203.

- Bates TC, Luciano M, Medland SE, et al. (2011) *Genetic variance in a component of the language acquisition device: ROBO1 polymorphisms associated with phonological buffer deficits*. Behav Genet, 41:50-57.
- Burnside RD, Pasion R, Mikhail FM, et al. 2011 *Microdeletion/microduplication of proximal 15q11.2 between BP1 and BP2: a susceptibility region for neurological dysfunction including developmental and language delay*. Hum Genet, 130:517-528.
- Cardon L, Smith S, Fulker D, et al. (1994) *Quantitative trait locus for reading disability on chromosome 6*. Science, 266:276-279.
- Carrion-Castillo A, Franke B, Fisher SE (2013) *Molecular genetics of dyslexia: an overview*. Dyslexia, 19:214-240.
- Ceroni F, Simpson NH, Francks C, et al. (2014) *Homozygous microdeletion of exon 5 in ZNF277 in a girl with specific language impairment*. Eur J Hum Genet, 22:1165-1171.
- Chandrasekar G, Vesterlund L, Hultenby K, et al. (2013) *The Zebrafish Orthologue of the Dyslexia Candidate Gene DYX1C1 Is Essential for Cilia Growth and Function*. PLoS ONE, 8:e63123.
- Chapman NH, Igo RP, Thomson JB, et al. (2004) *Linkage analyses of four regions previously implicated in dyslexia: Confirmation of a locus on chromosome 15q*. Am J Med Genet B Neuropsychiatr Genet, 131B:67-75.
- Chin LJ, Ratner E, Leng S, et al. (2008) *A SNP in a let-7 microRNA complementary site in the KRAS 3' untranslated region increases non-small cell lung cancer risk*. Cancer Res, 68:8535-8540.
- Clop A, Marcq F, Takeda H, et al. (2006) *A mutation creating a potential illegitimate microRNA target site in the myostatin gene affects muscularity in sheep*. Nat Genet, 38:813-818.
- Consensus Conference 3: *Disturbi specifici dell'apprendimento*, Roma, 6-7 Dicembre 2010.
- Cope N, Harold D, Hill G, et al. (2005) *Strong Evidence That KIAA0319 on Chromosome 6p Is a Susceptibility Gene for Developmental Dyslexia*. Am J Hum Genet, 76:581-591.
- Cornoldi C [1991], *Le parole della psicologia. Disturbi dell'apprendimento*, in <<Giornale Italiano di Psicologia>>, 18, pp.151-156.
- Cornoldi C [2007], *Difficoltà e disturbi dell'apprendimento*. Bologna, Il Mulino.
- Couto JM, Gomez L, Wigg K, et al. (2008) *The KIAA0319-like (KIAA0319L) gene on chromosome 1p34 as a candidate for reading disabilities*. J Neurogenet, 22:295-313.
- De Fries JC, Fulker DW, La Buda MC (1987) *Evidence for a genetic aetiology in reading disability of twins*. Nature, 329:537-539.
- de Kovel CG, Franke B, Hol FA, et al. (2008) *Confirmation of dyslexia susceptibility loci on chromosomes 1p and 2p, but not 6p in a Dutch sibpair collection*. Am J Med Genet B Neuropsychiatr Genet, 147:294-300.

de Kovel CG, Hol FA, Heister JG, et al. (2004) *Genome-wide scan identifies susceptibility locus for dyslexia on Xq27 in an extended Dutch family*. J Med Genet, 41:652-657.

de Zeeuw EL, van Beijsterveldt CE, Glasner TJ, et al. (2016) *Arithmetic, reading and writing performance has a strong genetic component: A study in primary school children*. Learn Individ Differ, 47:156-166.

Desmet FO, Hamroun D, Lalonde M, et al. (2009) *Human Splicing Finder: an online bioinformatics tool to predict splicing signals*. Nucleic Acids Res, 37:e67.

Eicher JD, Powers NR, Miller LL, et al. (2014) *Characterization of the DYX2 locus on chromosome 6p22 with reading disability, language impairment, and IQ*. Hum Genet, 133:869-881.

Einarsdottir E, Peyrard-Janvid M, Darki F, et al. (2017) *Identification of NCAN as a candidate gene for developmental dyslexia*. Sci Rep, 7:9294.

Etebari M, Navari M, Piccaluga PP (2015) *SNPs Array Karyotyping in Non-Hodgkin Lymphoma*. Microarrays (Basel), 4:551-569.

Falcaro M, Pickles A, Newbury DF, et al. (2008) *Genetic and phenotypic effects of phonological short-term memory and grammatical morphology in specific language impairment*. Genes Brain Behav, 7:393-402.

Fedorenko E, Morgan A, Murray E, et al. (2016) *A highly penetrant form of childhood apraxia of speech due to deletion of 16p11.2*. Eur J Hum Genet, 24:302-306.

Fisher SE and DeFries JC (2002) *Developmental dyslexia: genetic dissection of a complex cognitive trait*. Nat Rev Neurosci, 3:767-780.

Fisher SE, Francks C, Marlow AJ, et al. (2002) *Independent genome-wide scans identify a chromosome 18 quantitative-trait locus influencing dyslexia*. Nat Genet, 30:86-91.

Fisher S, Stein J, Monaco A (1999) *A genome-wide search strategy for identifying quantitative trait loci involved in reading and spelling disability (developmental dyslexia)*. Eur Child Adolesc Psychiatry, 8:S47-S51.

Francks C, Fisher SE, Olson RK, et al. (2002) *Fine mapping of the chromosome 2p12-16 dyslexia susceptibility locus: quantitative association analysis and positional candidate genes SEMA4F and OTX1*. Psychiatr Genet, 12:35-41.

Francks C, Paracchini S, Smith SD, et al. (2004) *A 77-Kilobase Region of Chromosome 6p22.2 Is Associated with Dyslexia in Families From the United Kingdom and From the United States*. Am J Hum Genet, 75:1046-1058.

Gialluisi A, Visconti A, Willcutt EG, et al. (2016) *Investigating the effects of copy number variants on reading and language performance*. J Neurodev Disord, 8:17.

Girirajan S, Brkanac Z, Coe BP, et al. (2011) *Relative burden of large CNVs on a range of neurodevelopmental phenotypes*. PLoS Genet, 7:e1002334.

- Gottesman II and Gould TD (2003) *The endophenotype concept in psychiatry: etymology and strategic intentions*. Am J Psychiatry, 160:636-645.
- Graham SA and Fisher SE (2015) *Understanding language from a genomic perspective*. Annu Rev Genet, 49:131-160.
- Grati MH, Chakchouk I, Ma Q, et al. (2015) *A missense mutation in DCDC2 causes human recessive deafness DFNB66, likely by interfering with sensory hair cell and supporting cell cilia length regulation*. Hum Molec Genet, 10.1093/hmg/ddv009.
- Grayton HM, Fernandes C, Rujescu D, et al. (2012) *Copy number variations in neurodevelopmental disorders*. Prog Neurobiol, 99:81-91.
- Grigorenko EL, Wood FB, Meyer MS, et al. (1997) *Susceptibility loci for distinct components of developmental dyslexia on chromosomes 6 and 15*. Am J Hum Genet, 60:27-39.
- Grigorenko EL, Wood FB, Meyer MS, et al. (2001) *Linkage studies suggest a possible locus for developmental dyslexia on chromosome 1p*. Am J Med Genet, 105:120-129.
- Guan J-S, Haggarty SJ, Giacometti E, et al. (2009) *HDAC2 negatively regulates memory formation and synaptic plasticity*. Nature, 459:55-60.
- Hammill DD (1990) *On defining learning disabilities: an emerging consensus*. J Learn Disabil, 23:74-84.
- Hannula-Jouppi K, Kaminen-Ahola N, Taipale M, et al. (2005) *The axon guidance receptor gene ROBO1 is a candidate gene for developmental dyslexia*. PLoS Genet, 1:e50.
- Harlaar N, Dale PS, Plomin R (2007) *From learning to read to reading to learn: substantial and stable genetic influence*. Child Dev, 78:116-131.
- Harold D, Paracchini S, Scerri T, et al. (2006) *Further evidence that the KIAA0319 gene confers susceptibility to developmental dyslexia*. Mol Psychiatry, 11:1085-1091.
- Hirai T, Taniura H, Goto Y, et al. (2006) *Stimulation of ubiquitin-proteasome pathway through the expression of amidohydrolase for N-terminal asparagine (Ntan1) in cultured rat hippocampal neurons exposed to static magnetism*. J Neurochem, 96:1519-1530.
- Hsiung GY, Kaplan BJ, Petryshen TL, et al. (2004) *A dyslexia susceptibility locus (DYX7) linked to dopamine D4 receptor (DRD4) region on chromosome 11p15.5*. Am J Med Genet B Neuropsychiatr Genet, 125B:112-119.
- Huttunen HJ, Kuja-Panula J, Sorci G, et al. (2000) *Coregulation of neurite outgrowth and cell survival by amphoterin and S100 proteins through receptor for advanced glycation end products (RAGE) activation*. J Biol Chem, 275:40096-40105.
- Ioannidis NM, Rothstein JH, Pejaver V, et al. (2016) *REVEL: An Ensemble Method for Predicting the Pathogenicity of Rare Missense Variants*. Am J Hum Genet, 99:877-885.
- Ionita-Laza I, McCallum K, Xu B, et al. (2016) *A spectral approach integrating functional genomic annotations for coding and noncoding variants*. Nat Genet, 48:214-220.

- Jagadeesh KA, Wenger AM, Berger MJ, et al. (2016) *M-CAP eliminates a majority of variants of uncertain significance in clinical exomes at high*. Nat Genet, 48:1581-1586.
- Jian X, Boerwinkle E, Liu X (2014) *In silico prediction of splice-altering single nucleotide variants in the human genome*. Nucleic Acids Res, 42:13534-13544.
- Johnson DJ [1988], *Review of research on specific reading, writing and mathematics disorders*, in JF Kavanagh e TJ Truss (a cura di), Learning, Disabilities, Parkton, Md., York Press, pp.79-163.
- Kaminen N, Hannula-Jouppi K, Kestila M, et al. (2003) *A genome scan for developmental dyslexia confirms linkage to chromosome 2p11 and suggests a new locus on 7q32*. J Med Genet, 40:340-345.
- Kaplan DE, Gayán J, Ahn J, et al. (2002) *Evidence for Linkage and Association with Reading Disability, on 6p21.3-22*. Am J Hum Genet, 70:1287-1298.
- Kircher M, Witten DM, Jain P, et al. *A general framework for estimating the relative pathogenicity of human genetic variants*. Nat Genet, 46:310-315.
- Lek M, Karczewski KJ, Minikel EV, et al. (2016) *Analysis of protein-coding genetic variation in 60,706 humans*. Nature, 536:285-291.
- Malhotra D and Sebat J (2012) *CNVs: harbingers of a rare variant revolution in psychiatric genetics*. Cell, 148:1223-1241.
- Marino C, Meng H, Mascheretti S, et al. (2012) *DCDC2 genetic variants and susceptibility to developmental dyslexia*. Psychiatr Genet, 22, 25-30.
- Massinen S, Hokkanen ME, Matsson H, et al. (2011) *Increased Expression of the Dyslexia Candidate Gene DCDC2 Affects Length and Signaling of Primary Cilia in Neurons*. PLoS ONE, 6:e20580.
- Matsson H, Huss M, Persson H, et al. (2015) *Polymorphisms in DCDC2 and S100B associate with developmental dyslexia*. Journal of Human Genetics, 60:399-401.
- Matsson H, Tammimies K, Zucchelli M, et al. (2011) *SNP variations in the 7q33 region containing DGKI are associated with dyslexia in the Finnish and German populations*. Behav Genet, 41:134-140.
- Meng H, Smith SD, Hager K, et al. (2005) *DCDC2 is associated with reading disability and modulates neuronal development in the brain*. Proc Natl Acad Sci USA, 102:17053-17058.
- Monaco AP (2007) *Multivariate linkage analysis of specific language impairment (SLI)*. Ann Hum Genet, 71:660-673.
- Mosca TJ (2015) *On the Teneurin track: a new synaptic organization molecule emerges*. Front Cell Neurosci, 9:204.
- Nagamani SC, Erez A, Bader P, et al. (2011) *Phenotypic manifestations of copy number variation in chromosome 16p13.11*. Eur J Hum Genet, 19:280-286.

- Newbury DF and Monaco AP (2010) *Genetic advances in the study of speech and language disorders*. *Neuron*, 68:309-320.
- Newbury DF, Paracchini S, Scerri TS, et al. (2011) *Investigation of dyslexia and SLI risk variants in reading- and language-impaired subjects*. *Behav Genet*, 41:90-104.
- Newbury D, Winchester L, Addis L, et al. (2009) *CMIP and ATP2C2 modulate phonological short-term memory in language impairment*. *Am J Hum Genet*, 85:264-272.
- Nishiyama H, Knopfel T, Endo S, et al. (2002) *Glial protein S100B modulates long-term neuronal synaptic plasticity*. *Proc Natl Acad Sci U S A*, 99:4037-4042.
- Nopola-Hemmi J, Myllyluoma B, Haltia T, et al. (2001) *A dominant gene for developmental dyslexia on chromosome 3*. *J Med Genet*, 38:658-664.
- Nopola-Hemmi J, Taipale M, Haltia T, et al. (2000) *Two translocations of chromosome 15q associated with dyslexia*. *J Med Genet*, 37:771-775.
- Pagnamenta AT, Bacchelli E, de Jonge MV, et al. (2010) *Characterization of a family with rare deletions in CNTNAP5 and DOCK4 suggests novel risk loci for autism and dyslexia*. *Biol Psychiatry*, 68:320-328.
- Peschansky VJ, Burbidge TJ, Volz AJ, et al. (2010) *The Effect of Variation in Expression of the Candidate Dyslexia Susceptibility Gene Homolog Kiaa0319 on Neuronal Migration and Dendritic Morphology in the Rat*. *Cereb Cortex*, 20:884-897.
- Peter B, Matsushita M, Oda K, et al. (2014) *De novo microdeletion of BCL11A is associated with severe speech sound disorder*. *Am J Med Genet A*, 164:2091-2096.
- Peter B, Raskind W, Matsushita M, et al. (2011) *Replication of CNTNAP2 association with nonword repetition and support for FOXP2 association with timed reading and motor activities in a dyslexia family sample*. *J Neurodevel Disord*, 3:39-49.
- Peterson RL and Pennington BF (2012) *Developmental dyslexia*. *Annu Rev Clin Psychol* 11:283-307.
- Peterson RL, Pennington BF. Developmental dyslexia*. *Lancet*, 379:1997-2007.
- Petrill SA, Thompson LA, Deater-Deckard K, et al. (2006) *Genetic and Environmental Effects of Serial Naming and Phonological Awareness on Early Reading Outcomes*. *J Educ Psychol*, 98:112-121.
- Petryshen TL, Kaplan BJ, Fu Liu M, et al. (2001) *Evidence for a susceptibility locus on chromosome 6q influencing phonological coding dyslexia*. *Am J Med Genet*, 105:507-517.
- Petryshen TL, Kaplan BJ, Hughes ML, et al. (2002) *Supportive evidence for the DYX3 dyslexia susceptibility gene in Canadian families*. *J Med Genet*, 39:125-126.
- Pettigrew KA, Reeves E, Leavett R, et al. (2015) *Copy number variation screen identifies a rare de novo deletion at chromosome 15q13.1-13.3 in a child with language impairment*. *PLoS ONE*, 10:e0134997.

- Peyrard-Janvid M, Anthoni H, Onkamo P, et al. (2004) *Fine mapping of the 2p11 dyslexia locus and exclusion of TACR1 as a candidate gene*. Hum Genet, 114:510-516.
- Platko JV, Wood FB, Pelsler I, et al. (2008) *Association of reading disability on chromosome 6p22 in the Afrikaner population*. Am J Med Genet B Neuropsychiatr Genet, 147B:1278-1287.
- Platko JV, Wood FB, Pelsler I, et al. (2008) *Association of reading disability on chromosome 6p22 in the Afrikaner population*. Am J Med Genet B Neuropsychiatr Genet, 147B:1278-1287.
- Platt MP, Adler WT, Mehlhorn AJ, et al. (2013) *Embryonic disruption of the candidate dyslexia susceptibility gene homolog Kiaa0319-like results in neuronal migration disorders*. Neuroscience, 248C:585-593.
- Poelmans G, Buitelaar JK, Pauls DL, et al. (2011) *A theoretical molecular network for dyslexia: Integrating available genetic findings*. Mol Psychiatry, 16:365-382.
- Poelmans G, Engelen JJM, Van Lent-Albrechts J, et al. (2009) *Identification of novel dyslexia candidate genes through the analysis of a chromosomal deletion*. Am J Med Genet B Neuropsychiatr Genet, 150B:140-147.
- Poelmans G, Engelen JJ, Van Lent-Albrechts J, et al. (2009) *Identification of novel dyslexia candidate genes through the analysis of a chromosomal deletion*. Am J Med Genet B Neuropsychiatr Genet, 150B:140-147.
- Priit P (2015) *Computational methods for DNA copy number detection*. The publication of this thesis is granted by the Institute of Molecular and Cell Biology, University of Tartu. ISBN 978-9949-32-927-4 (pdf). Room No 105, 23B Riia St., Tartu, on October 16th 2015, at 10.15.
- Purcell S, Neale B, Todd-Brown K, et al. (2007) *PLINK: a tool set for whole-genome association and population-based linkage analyses*. Am J Hum Genet, 81:559-575.
- Quang D, Chen Y, Xiaohui Xie (2015) *DANN: a deep learning approach for annotating the pathogenicity of genetic variants*. Bioinformatics, 31:761-763.
- Rabin M, Wen XL, Hepburn M, et al. (1993) *Suggestive linkage of developmental dyslexia to chromosome 1p34-p36*. Lancet, 342:178.
- Raca G, Baas BS, Kirmani S, et al. (2013) *Childhood apraxia of speech (CAS) in two patients with 16p11.2 microdeletion syndrome*. Eur J Hum Genet, 21:455-459.
- Ramalingam A, Zhou XG, Fiedler SD, et al. (2011) *16p13.11 duplication is a risk factor for a wide spectrum of neuropsychiatric disorders*. J Hum Genet, 56:541-544.
- Raskind WH, Igo RP, Chapman NH, et al. (2005) *A genome scan in multigenerational families with dyslexia: Identification of a novel locus on chromosome 2q that contributes to phonological decoding efficiency*. Mol Psychiatry, 10:699-711.
- Reiner O, Karzbrun E, Kshirsagar A, et al. (2016) *Regulation of neuronal migration, an emerging topic in autism spectrum disorders*. J Neurochem, 136:440-456.

- Rice ML, Smith SD, Gayan J (2009) *Convergent genetic linkage and associations to language, speech and reading measures in families of probands with Specific Language Impairment*. J Neurodev Disord, 1:264-282.
- Rodenas-Cuadrado P, Ho J, Vernes SC (2014) *Shining a light on CNTNAP2: complex functions to complex disorders*. Eur J Hum Genet, 22:171-178.
- Rothermundt M, Peters M, Prehn JH, et al. (2003) *S100B in brain damage and neurodegeneration*. Microsc Res Tech, 60:614-632.
- Rountree MR, Bachman KE, Baylin SB (2000) *DNMT1 binds HDAC2 and a new co-repressor, DMAP1, to form a complex at replication foci*. Nat Genet, 25:269-277.
- Scerri TS, Morris AP, Buckingham LL, et al. (2011) *DCDC2, KIAA0319 and CMIP Are Associated with Reading-Related Traits*. Biol Psychiatry, 70:237-245.
- Scerri TS, Paracchini S, Morris A, et al. (2010) *Identification of candidate genes for dyslexia susceptibility on chromosome 18*. PLoS ONE 5:e13712.
- Schork NJ, Murray SS, Frazer KA, et al. (2009) *Common vs. Rare allele hypotheses for complex diseases*. Curr Opin Genet Dev, 19:212-219.
- Schueler M, Braun Daniela A, Chandrasekar G, et al. (2015) *DCDC2 Mutations Cause a Renal-Hepatic Ciliopathy by Disrupting Wnt Signaling*. Am J Hum Genet, 96:81-92.
- Schulte-Körne G, Grimm T, Nöthen MM, et al. (1998) *Evidence for Linkage of Spelling Disability to Chromosome 15*. Am J Hum Genet, 63:279-282.
- Schumacher J, Anthoni H, Dahdouh F, et al. (2006a) *Strong Genetic Evidence of DCDC2 as a Susceptibility Gene for Dyslexia*. Am J Hum Genet, 78:52-62.
- Schumacher J, König IR, Plume E, et al. (2006b) *Linkage analyses of chromosomal region 18p11-q12 in dyslexia*. J Neural Transm, 113:417-423.
- Seshadri S, DeStefano AL, Au R, et al. (2007) *Genetic correlates of brain aging on MRI and cognitive test measures: a genome-wide association and linkage analysis in the Framingham Study*. BMC Med Genet, 8 Suppl 1:S15.
- Sharma P and Sagar R (2017) *Unfolding the genetic pathways of dyslexia in Asian population: A review*. Asian J Psychiatr, S1876-2018:30274-30275.
- Simpson NH, Ceroni F, Reader RH, et al. (2015) *Genome-wide analysis identifies a role for common copy number variants in specific language impairment*. Eur J Hum Genet, 23:1370-1377.
- Stefansson H, Meyer-Lindenberg A, Steinberg S, et al. (2014) *CNVs conferring risk of autism or schizophrenia affect cognition in controls*. Nature, 505:361-366.
- Swanson HL (2011) *Dynamic testing, working memory, and reading comprehension growth in children with reading disabilities*. J Learn Disabil, 44:358-371.

- Taipale M, Kaminen N, Nopola-Hemmi J, et al. (2003) *A candidate gene for developmental dyslexia encodes a nuclear tetratricopeptide repeat domain protein dynamically regulated in brain*. Proc Natl Acad Sci USA, 100:11553-11558.
- Tammimies K, Vitezic M, Matsson H, et al. (2013) *Molecular networks of DYX1C1 gene show connection to neuronal migration genes and cytoskeletal proteins*. Biol Psychiatry, 73:583-590.
- Tabet AN, Verloes A, Pilorge M, et al. (2015) *Complex nature of apparently balanced chromosomal rearrangements in patients with autism spectrum disorder*. Molecular Autism, 6:19.
- Tarkar A, Loges NT, Slagle CE, et al. (2013) *DYX1C1 is required for axonemal dynein assembly and ciliary motility*. Nat Genet, 45:995-1003.
- The SLI Consortium (2002) *A genome wide scan identifies two novel loci involved in specific language impairment*. Am J Hum Genet, 70:384-398.
- The SLI Consortium (2004) *Highly significant linkage to the SLI1 locus in an expanded sample of individuals affected by specific language impairment*. Am J Hum Genet, 74:1225-1238.
- Tropeano M, Ahn JW, Dobson RJ, et al. (2013) *Male-biased autosomal effect of 16p13.11 copy number variation in neurodevelopmental disorders*. PLoS One, 8:e61365.
- Tucker RP, Kenzelmann D, Trzebiatowska A, et al. (2007) *Teneurins: transmembrane proteins with fundamental roles in development*. Int J Biochem Cell Biol, 39:292-297.
- Varshney M and Nalvarte I (2017) *Genes, Gender, Environment, and Novel Functions of Estrogen Receptor Beta in the Susceptibility to Neurodevelopmental Disorders*. Brain Sci, 7(3).
- Veerappa AM, Saldanha M, Padakannaya P, et al. (2013a) *Family-based genome-wide copy number scan identifies five new genes of dyslexia involved in dendritic spinal plasticity*. J Hum Genet, 58:539-547.
- Veerappa AM, Saldanha M, Padakannaya P, et al. (2013b) *Genome-wide copy number scan identifies disruption of PCDH11X in developmental dyslexia*. Am J Med Genet B Neuropsychiatr Genet, 162B:889-897.
- Veerappa AM, Saldanha M, Padakannaya P, et al. (2014) *Family based genome-wide copy number scan identifies complex rearrangements at 17q21.31 in dyslexics*. Am J Med Genet B Neuropsychiatr Genet, 165B:572-580.
- Velayos-Baeza A, Toma C, da Roza S, et al. (2007) *Alternative splicing in the dyslexia-associated gene KIAA0319*. Mamm Genome, 18:627-634.
- Velayos-Baeza A, Toma C, Paracchini S et al. (2008) *The dyslexia-associated gene KIAA0319 encodes highly N- and O-glycosylated plasma membrane and secreted isoforms*. Hum Mol Genet, 17:859-871.
- Vernes SC, Newbury DF, Abrahams BS, et al. (2008) *A functional genetic link between distinct developmental language disorders*. N Engl J Med, 359:2337-2345.

Wang K, Li M, Hadley D, et al. (2007) *PennCNV: an integrated hidden Markov model designed for high-resolution copy number variation detection in whole-genome SNP genotyping data*. *Genome Res*, 17:1665-1674.

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

Wang Y, Yin X, Rosen G, et al. (2011) *Dcdc2 knockout mice display exacerbated developmental disruptions following knockdown of doublecortin*. *Neuroscience*, 190:398-408.

Whitehouse AJ, Bishop DV, Ang QW, et al. (2011) *CNTNAP2 variants affect early language development in the general population*. *Genes Brain Behav*, 10:451-456.

Wilcke A, Weissfuss J, Kirsten H, et al. (2009) *The role of gene DCDC2 in German dyslexics*. *Ann Dyslexia*, 59:1-11.

Wiszniewski W, Hunter JV, Hanchard NA, et al. (2013) *TM4SF20 ancestral deletion and susceptibility to a pediatric disorder of early language delay and cerebral white matter hyperintensities*. *Am J Hum Genet*, 293:197-210.

Xiong HY, Alipanahi B, Lee LJ, et al. (2015) *RNA splicing. The human splicing code reveals new insights into the genetic determinants of disease*. *Science*, 347:1254806.