



ALMA MATER STUDIORUM  
UNIVERSITA' DI BOLOGNA

**Alma Mater Studiorum - Università di Bologna**

---

*TESI DI DOTTORATO IN BIOCHIMICA*

Settore Scientifico Disciplinare MED/03 (GENETICA MEDICA)

*Analisi di varianti strutturali e di sequenza in geni  
candidati per l'autismo sul cromosoma 2q*

Tesi di Dottorato di:  
Dott.ssa **Simona Carone**

Relatore:  
Chiar.mo **Prof. Giovanni Romeo**

Correlatore:  
Chiar.ma **Prof.ssa Elena Maestrini**

Coordinatore:  
Chiar.mo **Prof. Giorgio Lenaz**

*XIX ciclo*

---

Anno Accademico 2005-2006



*Analisi di varianti strutturali e di sequenza in geni  
candidati per l'autismo sul cromosoma 2q*



## ***Indice***

### ***1. Le Malattie Complesse***

1.1. Introduzione	1
1.2. Come determinare le componenti genetiche di una malattia complessa	3
1.3. Lo studio delle malattie complesse: metodi usati per il mappaggio di geni di suscettibilità	4
1.3a. L'analisi di linkage	5
1.3b. Gli studi di associazione	9
1.4. HapMap: una risorsa importante per gli studi di associazione delle malattie complesse	12
1.5. Varianti strutturali nel genoma	14
1.5a. Caratteristiche generali	14
1.5b. Instabilità genomica e variabilità genetica	16
1.5c. Formazione di CNVs tramite NAHR	17
1.5d. CNVs e fenotipo	19
1.5e. Metodi per l'identificazione di CNVs	20

### ***2. Autismo***

2.1. Autismo: caratteri generali	24
2.2. Fenotipo clinico e criteri diagnostici	24
2.3. Trattamenti terapeutici	26
2.4. Autismo: dalla sua prima descrizione al riconoscimento dell'importanza dei fattori genetici	27
2.5. Disordini monogenici associati all'autismo	31
2.6. Identificazione dei geni di suscettibilità dell'autismo: metodi usati per il mappaggio	33
2.6a. Studi di linkage per l'autismo estesi all'intero genoma	33
2.6b. Anomalie cromosomiche associate all'autismo	38
2.7. Geni candidati per l'autismo	41

<b>3. Il cromosoma 2: screening di geni candidati per l'autismo</b>	44
3.1. Geni di suscettibilità sul cromosoma 2	45
<b>4. Il cromosoma 2: studio di associazione su larga scala e individuazione di CNVs</b>	50
4.1 Genotyping su larga scala di SNPs tramite il sistema Illumina	50
4.2 Tecnologia Illumina: tecnologia array-based per il genotyping di SNPs su larga scala	51
<b>5. Scopo della ricerca</b>	54
<b>6. Materiali e metodi</b>	56
6.1. Campioni di DNA	56
6.2. Estrazione del DNA da sangue intero	56
6.3. Disegno dei primers	57
6.4. Amplificazione tramite Polimerasi Chain Reaction (PCR)	57
6.5. Elettroforesi su gel di agarosio	63
6.6. Reazione di sequenziamento ed analisi delle sequenze	63
6.7. Genotipizzazione SNPs mediante digestione con enzimi di restrizione	65
6.8. Metodo della Ligase Detection Reaction (LDR)	65
6.9. Analisi di associazione: caso controllo e TDT	69
<b>7. Risultati</b>	70
7.1. Screening di mutazioni nei geni candidati	70
7.2. Studi di associazione allelica nei geni candidati	77
7.2a. INNP1 e NCKAP1	77
7.2b. KCNJ3, NR4A2, NAB1, UBE2E3	80
7.3. Individuazione di delezioni submicroscopiche nella regione 2q24-q32.2 in famiglie con autismo	81
7.4. Mappaggio della delezione nel gene UPP2	89
7.5. Analisi della delezione in UPP2 in famiglie con autismo	91

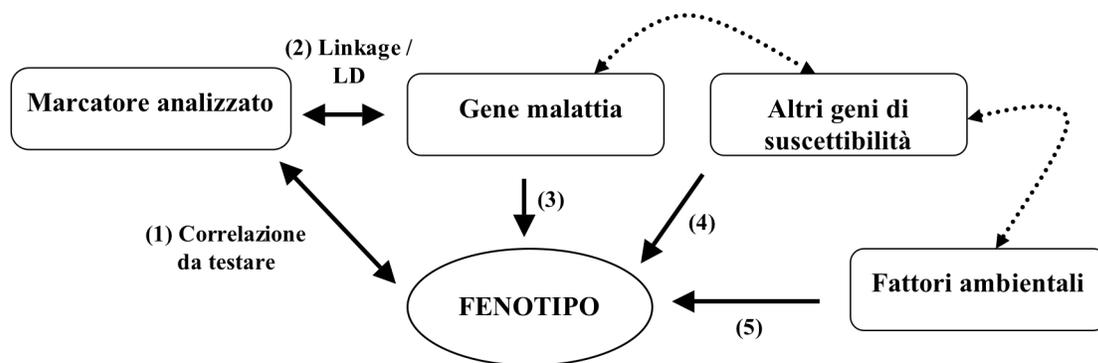
7.6. Screening di mutazioni del gene UPP2	94
7.7. Studi di associazione allelica nel gene UPP2	96
<b>8. Discussione</b>	97
<b>9. Bibliografia</b>	103



# 1. Le Malattie Complesse

## 1.1. Introduzione

Le malattie complesse rappresentano un insieme di patologie per le quali fattori genetici e ambientali concorrono alla manifestazione fenotipica e, a differenza delle malattie monogeniche, non mostrano un classico modello di ereditarietà mendeliana. Il fenotipo delle malattie monogeniche può essere complesso, ma in ogni caso dipende dall'azione di un unico gene, mentre in una malattia complessa, le varie componenti fenotipiche osservabili dipendono dalle interazioni di più geni che interagiscono tra loro in maniera epistatica (Figura 1). La difficoltà nello studio di queste patologie complesse consiste proprio nel correlare le varie componenti fenotipiche ad uno specifico fattore genetico e/o ambientale [Elston, 1995]



**Figura 1:** Interazioni fra fattori genetici e/o ambientali nel caso di una malattia genetica complessa: (1) In uno studio di mappaggio genico si ricerca una correlazione tra fenotipo e marcatore analizzato. Il suo successo dipende (2) sia dalla correlazione tra marcatore e gene malattia (*linkage*), sia dalla correlazione tra il gene malattia ed il fenotipo (3). Nelle malattie mendeliane questa correlazione tra il gene malattia ed il fenotipo (3) è forte ed il successo dipende unicamente dalla vicinanza tra il marcatore analizzato e il gene malattia. La presenza di altri fattori che predispongono alla malattia come per es. altri geni di suscettibilità (4), fattori ambientali (5) riduce la correlazione tra gene malattia e fenotipo (3).

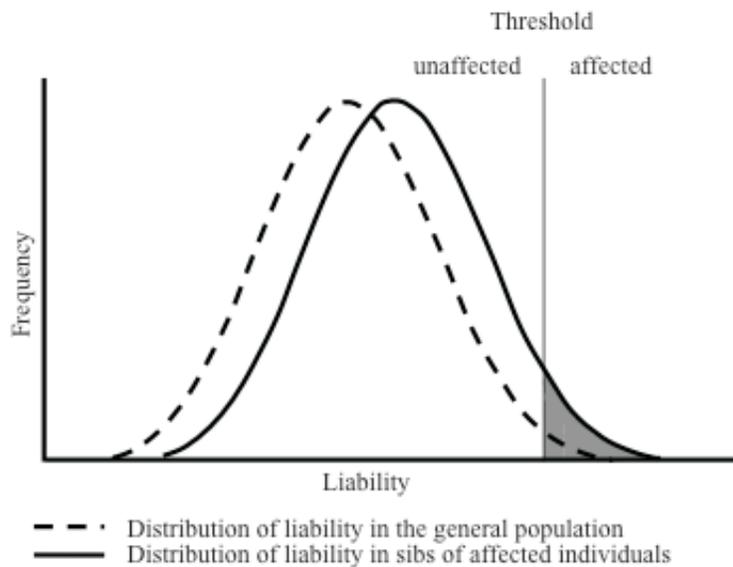
Generalmente ogni gene coinvolto in una patologia complessa ne è responsabile solo in parte e per questo si dice che è un “gene di suscettibilità”, il cui effetto sulla manifestazione fenotipica può essere additivo o meno rispetto a quello di altri geni coinvolti nella malattia stessa. Per lo studio di malattie complesse questo comporta la necessità di disporre di ampie casistiche

(famiglie o individui affetti) sia per ottenere un sufficiente potere statistico, sia per avere un campione rappresentativo dell'eterogeneità della malattia a livello di fattori genici e ambientali [Lander 1994]. Più una patologia è complessa, infatti, meno la componente genetica risulta influente sul piano eziologico, e di conseguenza il rischio relativo di sviluppare la malattia, attribuibile ai soli fattori genetici, si rivela modesto e inversamente proporzionale al numero di pazienti necessari per lo studio. Quindi malattie complesse, caratterizzate da un rischio genetico relativo, difficilmente quantificabile, richiedono una dimensione del campione a volte difficile da raggiungere sul piano pratico.

Per tutti questi motivi, lo studio delle malattie complesse risulta più arduo rispetto a quello delle malattie monogeniche e non sempre è possibile adattare le metodologie di analisi genetica di malattie monofattoriali alle malattie multifattoriali se non integrando molteplici approcci [Barnette et al., 2005].

## 1.2. Come determinare le componenti genetiche di una malattia complessa

Molte malattie complesse sono causate dalla combinazione di un numero variabile di geni di suscettibilità e di fattori ambientali. Solo gli individui che presentano un insieme di fattori di predisposizione superiore ad un livello definito “livello soglia” manifestano la malattia (Figura 2). In presenza di una forte determinante genetica, i parenti di individui affetti mostrano una maggiore probabilità rispetto alla popolazione in generale di sviluppare la malattia stessa (aumento del rischio relativo,  $\lambda_s$ ) [Farrer and Cupples, 1998].



**Figura 2:** Modello multifattoriale a soglia: I fattori che predispongono all’insorgere della malattia sono distribuiti normalmente nella popolazione. Gli individui la cui predisposizione supera questa soglia manifestano la malattia. I fratelli di individui affetti mostrano un rischio maggiore rispetto alla media della popolazione e una maggiore proporzione di questi ultimi presenta una predisposizione che supera la soglia.

La determinazione di una componente genetica è un passaggio fondamentale negli studi di mappaggio genetico in quanto il suo successo dipende dalla correlazione tra genotipo e fenotipo. Il primo passo consiste nella determinazione della ricorrenza familiare ( $\lambda_r$ ). Quando una malattia ha una base genetica, i parenti di primo grado di probandi affetti hanno un rischio maggiore rispetto a parenti più lontani, o rispetto alla popolazione in generale, di sviluppare la malattia. Quindi, il grado di aggregazione (*clustering*) familiare può essere misurato come rischio relativo di ricorrenza familiare  $\lambda_r$ , cioè come la proporzione del rischio dei parenti dei probandi affetti

rispetto alla prevalenza della malattia nella popolazione ( $\lambda_r = k_r/k$ , dove  $\lambda_r$  rappresenta il rischio relativo di ricorrenza familiare,  $k_r$  rappresenta il rischio dei parenti dei probandi affetti, e  $k$  rappresenta la prevalenza nella popolazione in generale [Risch, 1990]. Il rischio di ricorrenza familiare maggiormente utilizzato è il rischio nei fratelli,  $\lambda_s$ , che rappresenta il rischio di ricorrenza nei fratelli di probandi rispetto al rischio nella popolazione in generale.

Poiché la ricorrenza familiare di un tratto non è un mezzo sufficiente per provare la base genetica di una malattia, in quanto individui tra loro imparentati condividono gli stessi geni, ma spesso anche gli stessi fattori ambientali, sono stati sviluppati metodi più sofisticati. Per quelle patologie dove una distinzione tra fattori genetici e fattori ambientali risultava difficile sono stati condotti studi su gemelli e su individui adottati. Un classico studio su gemelli paragona la concordanza tra gemelli monozigotici (MZ) e gemelli dizigotici (DZ) per un carattere che indica la relativa importanza di effetti genetici e di effetti ambientali. La differenza di concordanza tra gemelli MZ e DZ è indice di una forte componente genetica, mentre una uguale concordanza suggerisce una forte componente ambientale [Allen et al., 1967].

L'analisi di segregazione è utilizzata per determinare se nell'eziologia di un carattere fenotipico è coinvolta una componente genetica preponderante.

### ***1.3 Lo studio delle malattie complesse: metodi usati per il mappaggio di geni di suscettibilità***

Per intraprendere lo studio genetico di una malattia complessa, occorre considerare anzitutto alcuni aspetti fondamentali:

1. La conoscenza del processo fisiopatologico:
  - Se è noto, oppure esistono già ipotesi sul ruolo nella patologia di specifici geni, allora è applicabile l'approccio del *gene candidato* utilizzando marcatori nelle vicinanze del gene candidato.
  - Se non è noto, è possibile utilizzare marcatori distribuiti su tutto il genoma e applicare un approccio *genome scan*.
2. La strategia metodologica ottimale in base alle informazioni genetiche disponibili:
  - *Studi di linkage* che mirano a trovare il legame fisico fra un marcatore e un gene analizzando l'ereditarietà all'interno di famiglie in cui si manifesta la patologia.
  - *Studi di associazione* che mirano ad individuare un'associazione allelica tra marcatori e malattia a livello di popolazione.

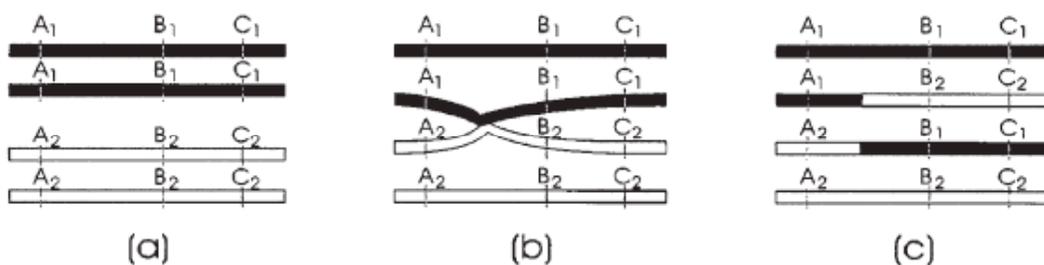
Non avendo chiari indizi sulle responsabilità genetiche in una malattia complessa, una prima fase per l'identificazione di geni di suscettibilità è l'individuazione di specifiche regioni sub-cromosomiche potenzialmente correlate alla malattia.

In assenza di anomalie cromosomiche o modelli animali della malattia, che potrebbero indicare la regione di interesse, questa localizzazione è effettuata mediante analisi di *linkage* seguite da analisi di associazione a più alta risoluzione.

### 1.3a. L'analisi di *linkage*

L'analisi di *linkage* classica, di tipo *parametrico* (nota anche come *model-based analysis*), rappresenta una metodologia statistica classica dell'analisi genetica e si basa sull'identificazione di marcatori che segregano insieme alla malattia all'interno di famiglie. Il fenomeno del *linkage* si deve alla stretta vicinanza di due geni o marcatori lungo un cromosoma e si verifica qualora la cosegregazione tra un marcatore (per lo più microsatellite) ed un gene che influenza il fenotipo di interesse si manifesta con una frequenza maggiore di quella attesa per il puro effetto del caso.

La base biologica è il concetto di ricombinazione o *crossing-over*, grazie al quale durante la prima fase meiotica avviene lo scambio fisico di materiale genetico tra i cromosomi omologhi paterno e materno (Figura 3).



**Figura 3:** Ricombinazione tra cromosomi omologhi durante la I fase meiotica. In (a) i due cromosomi si accoppiano e ciascun cromosoma (materno e paterno) si divide in due cromatidi identici. Sono indicati gli alleli appartenenti ai tre loci (A, B e C). In (b) si osserva un *crossing-over* tra i loci A e B. In (c) sono raffigurati i quattro gameti risultanti; in due è avvenuta una ricombinazione tra i loci A e B e tra i loci A e C; nessuna ricombinazione è avvenuta tra B e C.

Se tra due loci avviene un *crossing-over*, allora gli alleli che il figlio riceve dal genitore si dicono “ricombinanti” non essendo più identici a quelli trasmessi dal genitore stesso.

In pratica ciò che si osserva è il grado di indipendenza con cui avviene la trasmissione di un marcatore rispetto ad un altro, ciò riflette il grado di ricombinazione fra tali marcatori. Il parametro che esprime la frazione di ricombinazione è  $\theta$ , intesa come la percentuale dei gameti che ci si attende siano ricombinanti. Secondo questo parametro, se due marcatori sono indipendenti (non in *linkage*), la frazione di ricombinazione  $\theta$  è massima ( $\theta = 0,5$ ), altrimenti la frazione di ricombinazione diminuisce progressivamente con l'aumentare dello stato di *linkage* dei due marcatori in esame. L'obiettivo dell'analisi di *linkage* consiste appunto nello stabilire se la frequenza di ricombinazione (misura della distanza genetica tra due loci), è significativamente  $< 50\%$  ( $\theta < 1/2$ ).

Solitamente, in un'analisi di *linkage* di tipo parametrico, si usano alberi genealogici (*pedigrees*) a più generazioni contenenti diversi individui affetti, dai quali è possibile dedurre il modello di ereditarietà del locus malattia.

Un *genomewide screen* consiste nella tipizzazione del genoma (*genotyping*) attraverso 300-400 marcatori altamente polimorfici (microsatelliti), distribuiti in maniera uniforme ad una distanza di circa 10 cM l'uno dall'altro, in un elevato numero di famiglie contenenti coppie di fratelli affetti (*sib-pairs*) o altri tipi di parenti affetti (*relative pairs*). I dati di *linkage*, ottenuti in seguito all'analisi di tutti i marcatori, sono utilizzati per effettuare un conteggio diretto dei ricombinanti per ottenere il valore di  $\theta$ . In alternativa si applica il *metodo del LOD Score*.

Il LOD Score è il logaritmo in base 10 del rapporto tra la probabilità che due loci siano associati e quella che non lo siano (Figura 4).

$$\text{LOD score } Z(\theta) = \Sigma \log_{10} = \frac{\text{L (family se } \theta=X)}{\text{L (family se } \theta=1/2)}$$

**Figura 4:** Formula di LODscore in funzione di  $\theta$  (frazione di ricombinazione). L: probabilità che i dati osservati all'interno di una famiglia si verificano per un dato valore assunto di  $\theta$ .

La soglia di significatività di *linkage* è fissata per LODscore  $\geq 3$  [Morton, 1955]. Poiché il valore di LODscore è il logaritmo in base dieci della probabilità in favore di *linkage*, un valore di LODscore pari a 3 corrisponde alla probabilità 1000:1 in favore di *linkage*. Se due loci si trovano lontani tra loro, allora i LOD score saranno per lo più negativi. Si assume che la frazione di

ricombinazione sia pari al valore di  $\theta$  che massimizza la funzione  $Z(\theta)$ , il cui valore massimo è detto massimo LODscore (MLS). In questo modo, in seguito ad un'analisi di *linkage* si generano dei profili di MLS a più punti (*multipoint* MLS) lungo i 23 cromosomi. I picchi lungo questi profili, rappresentati graficamente, indicano la posizione di possibili loci di suscettibilità.

I metodi di *linkage* parametrici sono generalmente applicati ai tratti mendeliani e necessitano della preventiva conoscenza del tipo di corrispondenza fra genotipo e fenotipo (modello di ereditarietà) che spiega la trasmissione della malattia nelle famiglie.

Nel caso di malattie complesse, il modello di ereditarietà non è conosciuto e, tuttavia, qualora sia possibile individuare un sottogruppo di famiglie in cui segrega la malattia, influenzate da un *majour locus* con segregazione di tipo mendeliano, si possono usare metodi parametrici. Tuttavia, il rischio di questo tipo di approccio è includere nell'analisi di *linkage* falsi positivi o falsi negativi.

La migliore strategia per localizzare geni coinvolti nella suscettibilità alle malattie complesse (*common diseases*) è rappresentata dal *linkage non parametrico* (noto anche come *model-free analysis*) che si basa sulla condivisione allelica (*allele sharing*) tra due o più individui affetti all'interno della stessa famiglia. Questo tipo di analisi, non necessitando della specificazione di un modello di ereditarietà, sopperisce al limite imposto dal metodo LODscore [Weeks & Lathrop, 1995; Lander & Schork, 1994; Kruglyak et al., 1996]. Si studiano alleli identici per discendenza (*Identity By Descent* IBD), cioè quegli alleli che sono identici negli individui affetti perchè sono stati ereditati dallo stesso antenato comune. La struttura familiare più semplice alla quale si applicano i metodi di condivisione allelica è costituita da coppie di fratelli affetti (*Affected Sib-Pair method*, ASP) (Figura 5) in cui si contano il numero di alleli trasmessi dai genitori e condivisi dai due probandi [Penrose, 1953; Suarez, 1978]. Questo metodo può essere esteso anche a coppie di parenti di tipo diverso (*Affected Relative-Pair method*).

Quindi, se due fratelli sono entrambi affetti da una malattia hanno, con tutta probabilità, ereditato dai loro genitori copie identiche della regione cromosomica contenente l'allele di suscettibilità alla malattia stessa. Con il metodo "model-free" ciò che si osserva è se i probandi affetti (coppie di fratelli affetti) condividono marcatori allelici IBD con una frequenza maggiore rispetto a quella prevista dalla segregazione casuale. Infatti, si suppone che, in quanto entrambi malati, avranno in comune il segmento di cromosoma che contiene il locus malattia. Due fratelli condividono in media per ciascun locus zero, uno o due alleli IBD rispettivamente con una probabilità del 25, 50 e 25%. In un campione di ASPs, in assenza di *linkage* tra un marcatore e il locus malattia, ci si

attendenderà dunque una simile distribuzione, mentre in caso di *linkage* si osserverà un aumento di ASPs che condividono uno o due alleli IBD (ed una diminuzione delle coppie che condividono zero alleli IBD) (Tabella 1).

<b>N°alleli condivisi</b>	<b>2</b>	<b>1</b>	<b>2</b>
Caso di non linkage	25%	50%	25%
<b>Linkage</b>	<b>&gt;25%</b>		<b>&lt;25%</b>

**Tabella 1:** Frequenze con cui due fratelli affetti condividono gli alleli di un locus in caso di linkage oppure in caso di assenza di linkage.

L'evidenza di *linkage* può essere stimata:

- I. comparando il numero osservato ed atteso di coppie che condividono zero, uno o due alleli identici per stato (IBS) mediante un test del  $\chi^2$  con due gradi di libertà (o test della media) [Ott, 1999].
- II. tramite il calcolo statistico del massimo LODscore (MLS) che valuta l'incremento di condivisione allelica tra parenti affetti. Per le malattie monogeniche il valore soglia significativo per il *linkage* è rappresentato da un  $MLS=3.0$ , mentre per le malattie complesse è stato proposto di utilizzare un di  $MLS \geq 3.6$ , valore che corrisponde ad una percentuale del 5% di falsi positivi per un intero *genome scan*. Tuttavia, poichè in uno studio di media grandezza (100-200 coppie ASPs) e in presenza di loci ad effetto moderato questo valore soglia non è sempre raggiungibile, si usa riportare tutti i punti del genoma che indicano presenza di *linkage*, anche se al di sotto della soglia di  $MLS \geq 3.6$ , accettando la possibilità che alcuni di questi siano "falsi positivi". Ulteriori evidenze sperimentali sono, quindi, necessarie per distinguere tra "veri" e "falsi positivi". La replicazione indipendente di un valore positivo di *linkage* è generalmente accettata come evidenza di un reale risultato positivo anche se, solitamente, è piuttosto difficile da ottenere in quanto il numero di famiglie necessarie per replicare un risultato di *linkage* è notevolmente maggiore rispetto al numero necessario nello studio originario [Suarez et al., 1994].

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

### ***1.3b. Gli studi di associazione***

Gli studi di *linkage model-free* hanno un potere relativamente basso di individuare geni con effetto moderato, quindi, quando si inizia lo studio di una malattia complessa, di cui non si conosce l'eziologia e i geni coinvolti, l'approccio generalmente seguito prevede, inizialmente, un'analisi dell'intero genoma utilizzando metodi di *linkage* non-parametrici e, successivamente, un'analisi sempre più fine delle regioni di *linkage* individuate utilizzando ulteriori marcatori interni alle regioni stesse.

A tal fine, gli studi di associazione rappresentano un approccio particolarmente utile in quanto mirano all'individuazione, all'interno di una popolazione, di varianti che si trovano più frequentemente associate alla patologia. Tale associazione può essere dovuta a diversi fattori quali: un effetto diretto della variante genica sul fenotipo malattia, il *linkage disequilibrium* (LD), o associazioni spurie dovute alla stratificazione della popolazione. Una variante allelica può avere un effetto diretto sul fenotipo in quanto può alterare la sequenza aminoacidica di una proteina funzionale coinvolta nel processo fisiopatologico, o alterare la regione regolatoria del gene determinando anomalie a livello di espressione o, infine, interferire con la regolazione dello *splicing* e la stabilità del messaggero. Al contrario, una variante allelica può mostrare associazione con la malattia senza, però, essere direttamente implicata nella patogenesi. Quest'ultima situazione si verifica in presenza di LD, quindi si ha una associazione non casuale tra la variante allelica analizzata e la vera variante eziologica in quanto i due loci sono vicini sul cromosoma e vengono ereditati insieme come unico blocco.

Esistono due tipi di studi di associazione:

- lo **studio caso-controllo**, condotto a livello di popolazione, che confronta la frequenza di un determinato allele tra un campione di individui affetti (casi) e un campione di individui non affetti (controlli) tra loro non imparentati.
- il ***Disequilibrium Transmission Test*** (TDT), condotto su base familiare, che permette di osservare se un allele di un marcatore è trasmesso da un genitore eterozigote al figlio affetto con una frequenza significativamente più alta rispetto a quella con la quale non è trasmesso (Figura 5).

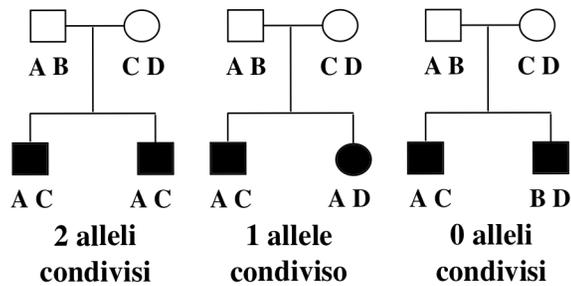
Gli studi di associazione del tipo **caso-controllo** sono quelli condotti a livello di popolazione. Gli individui, affetti e non, sono selezionati da una stessa popolazione e analizzati utilizzando marcatori polimorfici opportunamente scelti. Se le frequenze alleliche di tali marcatori, osservate nel campione di individui affetti e nel campione di individui sani, differiscono in modo significativo è stata trovata associazione. Il problema maggiore in questo tipo di studio è rappresentato dalla stratificazione della popolazione e, quindi, dalla presenza di sottopopolazioni geneticamente differenti, con diverse distribuzioni alleliche [Slager et al., 2001].

Per ovviare questo problema è utilizzato un test di associazione basato sulla trasmissione di alleli all'interno di nuclei familiari: il ***Transmission Disequilibrium Test*** (TDT). Questo test è condotto su famiglie composte da entrambi i genitori informativi (eterozigoti) per il marcatore da analizzare e uno o più figli affetti e calcola quante volte i genitori eterozigoti per un certo allele trasmettono l'allele in questione al figlio affetto e quante volte non lo trasmettono [Spielman et al., 1993]. Se la trasmissione dell'allele avviene con una frequenza significativamente  $> 50\%$  si è in presenza di associazione allelica e *linkage* tra marcatore e malattia. Si osserva associazione se l'allele è direttamente implicato nell'eziologia della patologia o se è in LD con il locus malattia, motivo per cui questo tipo di approccio viene utilizzato, di solito, o per analizzare polimorfismi presenti all'interno di geni candidati o per mappare in maniera più fine una regione di interesse individuata da precedenti studi di *linkage*

## Linkage

Si basa sulla cosegregazione della malattia con marcatori molecolari in famiglie con più individui affetti (multiplex)

### Affected Sib-Pair Method



Marcatore <u>non in linkage</u> con il gene di suscettibilità	<b>25%</b>	<b>50%</b>	<b>25%</b>
Le coppie di fratelli affetti condividono gli alleli secondo la probabilità casuale			

Marcatore <u>in linkage</u> con il gene di suscettibilità	<b>40%</b>	<b>55%</b>	<b>5%</b>
Le coppie di fratelli affetti condividono gli alleli più frequentemente rispetto alla probabilità casuale			

## Associazione

In una popolazione la frequenza di un allele è significativamente più alta negli affetti rispetto ai controlli

### Transmission Disequilibrium Test

Gli alleli dei genitori sono usati come controlli interni per evitare falsi risultati di associazione dovuti alla stratificazione della popolazione

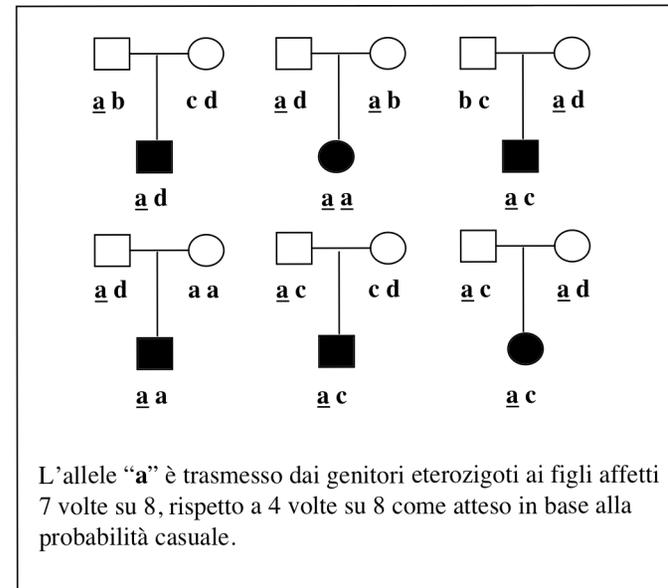


Figura 5: Linkage e Associazione: strumenti statistici per mappare malattie genetiche.

#### ***1.4. HapMap: una risorsa importante per gli studi di associazione delle malattie complesse***

Il progetto internazionale *HapMap* ha ampliato le prospettive degli studi di genetica umana in quanto ha portato alla formazione di *database* delle varianti comuni del genoma umano. Grazie alle informazioni contenute in questi *database* è oggi possibile intraprendere degli studi di associazione molto più efficienti ed estesi all'intero genoma, un approccio praticamente irrealizzabile fino a pochi anni fa. Il progetto HapMap è nato nel 2002 con lo scopo di mappare la struttura di *linkage disequilibrium* lungo tutto il genoma umano. Nella prima fase del progetto è stato effettuato il *genotyping* di più di un milione di SNPs (*Single Nucleotide Polymorphisms*), uno ogni circa 5 Kb di DNA, in 269 individui facenti parte di quattro diverse aree geografiche. Di questi individui fanno parte i componenti di 30 famiglie composte da madre, padre e un figlio della popolazione Yoruba della penisola Ibadan in Nigeria (YRI), 30 analoghe famiglie del progetto CEPH in Utah (CEU), 45 individui non imparentati della popolazione cinese Han di Beijing (CHB), altri 45 individui non imparentati giapponesi della zona di Tokyo (JPT). Nella seconda fase del progetto sono stati analizzati altri quattro milioni di SNPs nelle stesse popolazioni, più ulteriori SNPs in popolazioni diverse [The International HapMap Consortium: 2003, 2005].

Secondo alcune stime, nella popolazione mondiale, si trovano circa dieci milioni di SNPs per i quali entrambi gli alleli sono presenti con una frequenza superiore all'1% . Tali SNPs (*common SNPs*) costituiscono il 90% della variabilità nella popolazione mondiale, mentre il restante 10% è costituito da un insieme diversificato di altre varianti rare [Kruglyak et al., 2001; Reich, 2003].

La formazione di una variante si ha in seguito ad un singolo evento mutazionale che si manifesta con una frequenza molto bassa (nell'ordine di  $10^{-8}$  per generazione) rispetto al numero di generazioni che separano due individui dal loro più prossimo antenato comune (nell'ordine di  $10^4$  generazioni). Per questo motivo, ogni nuova variante è inizialmente associata agli altri alleli che si trovano nelle vicinanze del locus cromosomico in cui si è verificata la mutazione. Lo specifico insieme di alleli che si trovano sullo stesso cromosoma o nella medesima regione sub-cromosomiale è detto *aplotipo*, quindi l'insorgere di una mutazione porta alla formazione di un nuovo *aplotipo*, così come avviene la formazione di un *aplotipo* mosaico nel caso di ricombinazione fra due *aplotipi* diversi presenti sul cromosoma di origine materna e paterna [Pääbo, 2003].

Il *linkage disequilibrium* (LD) deriva dal fatto che alleli vicini vengono più facilmente co-ereditati in blocchi di *aplotipi* risultando, quindi, associati all'interno di una popolazione. Poiché la probabilità di ricombinazione fra due SNPs aumenta con l'aumentare della distanza fisica fra i



malattie complesse: la conoscenza del pattern di LD e dei *tag* SNPs del genoma umano ha determinato la velocizzazione e l'ottimizzazione dei costi di *genotyping* dei campioni genomici oggetto di studio.

## ***1.5. Varianti strutturali nel genoma***

### ***1.5a. Caratteristiche generali***

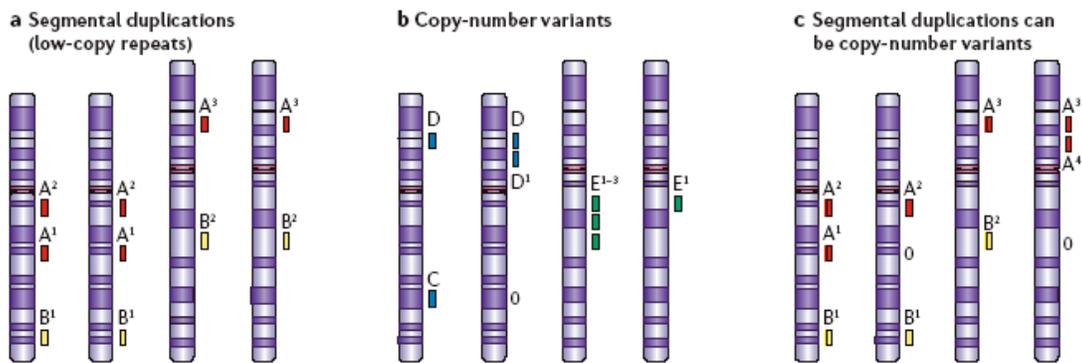
Prima dell'avvento delle innovazioni tecnologiche che hanno consentito un'analisi molecolare dettagliata e su larga scala del genoma, le differenze osservabili a livello genomico riguardavano rari casi di variazione del numero e della struttura dei cromosomi, spesso associati a gravi patologie (aneuploidie, riarrangiamenti, eteromorfismi, siti fragili, ecc.) osservabili tramite l'utilizzo del microscopio [Warburton, 1991; Jacobs et al., 1992]. Questo tipo di varianti coinvolgono segmenti di DNA di lunghezza pari o superiore a 3Mb, abbastanza grandi da poter essere individuati a livello microscopico con osservazioni citogenetiche. In seguito, con il diffondersi di tecniche di sequenziamento sempre più efficienti, sono state individuate varianti strutturali di dimensioni molto minori ( $\leq$  1Kb) e molto più frequenti nella popolazione. Tali varianti comprendono SNPs (polimorfismi a singola basa), piccole ripetizioni in tandem (*Short Tandem Repeats*, STRs o microsatelliti), e piccole inserzioni, delezioni, duplicazioni, inversioni. Molte di queste varianti sono state caratterizzate e catalogate in una banca dati, chiamata dbSNP (<http://www.ncbi.nlm.nih.gov/projects/SNP/>). Gli SNPs sono il tipo di variante più frequente (mediamente una ogni 100-300 bp) e hanno destato negli ultimi anni un grande interesse in quanto considerate la principale fonte di variabilità genetica e fenotipica [Kruglyak et al., 2001]. Solo molto recentemente, in seguito al diffondersi di tecnologie specifiche per gli studi a livello genomico, è stato possibile individuare la presenza di un numero inaspettatamente grande di varianti strutturali di dimensioni intermedie (comprese fra 1Kb e 3Mb) complessivamente denominate *Copy-Number Variants* (CNVs) (Tabella 2) e (Figura7).

<p><b>Copy-number variant (CNV):</b>          Segmento di DNA di lunghezza superiore ad 1Kb presente in un numero variabile di copie rispetto ad un genoma di riferimento. Diverse classi di CNVs comprendono inserzioni, delezioni, duplicazioni. Delle CNVs fanno parte anche le <i>large-scale copy-number variants</i> che riguardano segmenti di DNA superiori a 50Kb.</p>
<p><b>Copy-number polymorphism (CNP):</b>          Un CNV presente nella popolazione con una frequenza maggiore dell'1%.</p>
<p><b>Segmental duplication o low-copy repeat (LCR):</b>          Segmento di DNA di lunghezza superiore ad 1Kb presente in due o più copie per genoma aploide tale per cui le copie hanno un'identità di sequenza del 90%. Le LCRs possono anche essere CNVs (fig.x).</p>
<p><b>Inversioni:</b>          Un segmento di DNA con orientamento invertito rispetto al resto del cromosoma.</p>
<p><b>Traslocazioni:</b>          Un cambiamento di posizione all'interno del genoma di un segmento di DNA senza perdita o guadagno di materiale genetico. Le traslocazioni possono essere <i>intra-</i> o <i>inter-</i>cromosomiche.</p>
<p><b>Segmental uniparental disomy:</b>          Si ha quando una porzione di cromosoma deriva da una coppia di cromosomi omologhi ereditata da un singolo genitore.</p>

**Tabella 2:** Definizioni relative alle diverse tipologie di varianti strutturali.

Le varianti strutturali costituiscono una fonte di variabilità di milioni di nucleotidi all'interno di ogni genoma e quindi molto probabilmente contribuiscono in modo preponderante alla diversità umana, così come fin dall'inizio era stato ipotizzato per gli SNPs [Sharp et al. 2005]. Inoltre, questo tipo di varianti può comprendere milioni di basi (nucleotidi) contenenti interi geni o intere regioni regolatorie [Sharp et al., 2005; Sebat et al., 2004; Iafrate et al., 2004; Tuzun et al., 2005; Stefansson et al., 2005]. Sebbene queste varianti strutturali in alcune regioni del genoma non hanno alcuna conseguenza fenotipica [Sharp et al. 2005; Sebat et al., 2004; Iafrate et al., 2004; Tuzun et al., 2005], altre influenzano il dosaggio di un gene, e di conseguenza, con la concomitanza di ulteriori fattori genetici ed ambientali, possono causare malattie genetiche [Inoue et al., 2002].

Negli ultimi anni sono stati condotti numerosi studi allo scopo di chiarire il significato evolutivo di questo tipo di varianti, il meccanismo in seguito al quale si formano, in che modo contribuiscono alla variabilità genetica e la loro possibile associazione con malattie ereditarie e complesse ad eziologia ignota quali, ad esempio, l'autismo [Buckley et al., 2005].



**Figura 7:** a) Segmental duplication: copie di un segmento genomico con omologia minima del 90% intersperse nel genoma aploide; b) Copy-number variants: quando un segmento genomico può essere presente un numero variabile di volte rispetto ad un genoma di riferimento; c) Esempio in cui alcune segmental duplication possono essere copy-number variant.

### ***1.5b. Instabilità genomica e variabilità genetica***

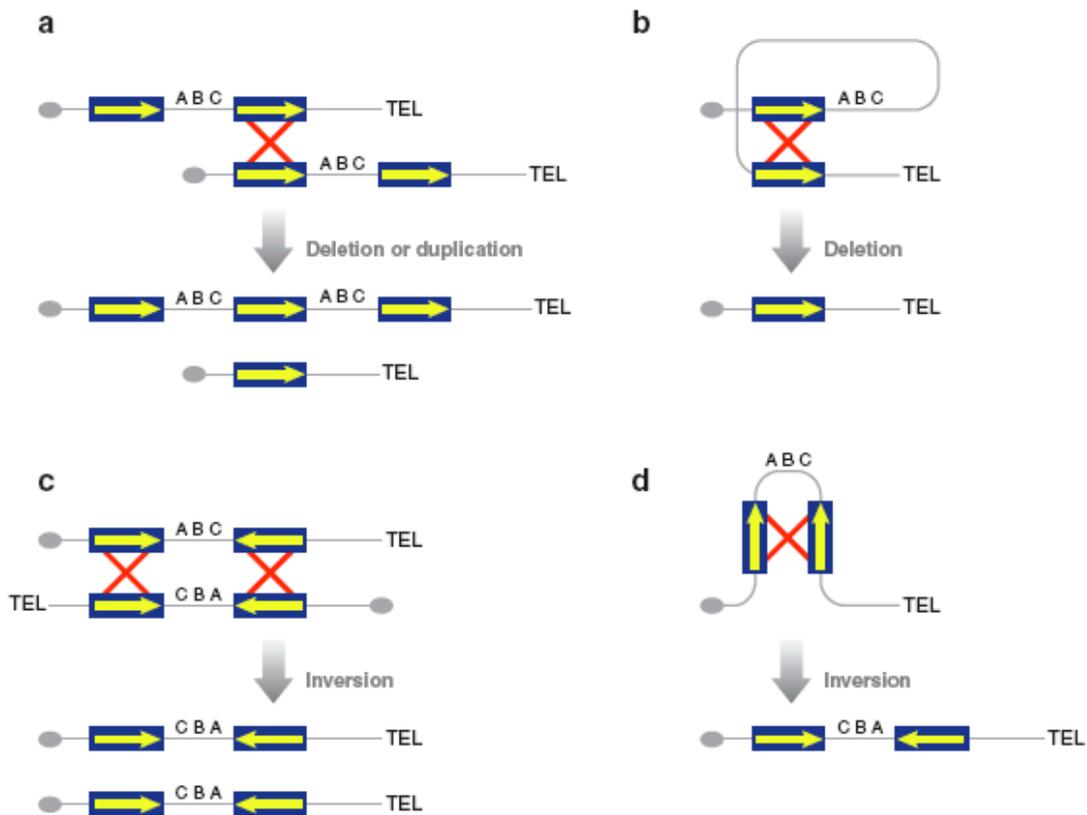
Diversi studi hanno evidenziato la presenza nel genoma di alcuni segmenti particolarmente suscettibili a riarrangiamenti cromosomici che causano un alto grado di instabilità genomica a livello di specifiche regioni cromosomiche. Si ipotizza che ciò sia uno dei meccanismi responsabili dell'introduzione di variabilità genetica all'interno della specie. I meccanismi che generano tali riarrangiamenti e le conseguenze fenotipiche che determinano, sono stati oggetto di numerosi studi negli ultimi anni. E' noto che riarrangiamenti cromosomici rari, piuttosto che mutazioni puntiformi nella sequenza di DNA, sono la causa dell'insorgenza di alcune malattie. Si indicano infatti con il termine di *malattie genomiche* tutti quei disturbi che sono dovuti a riarrangiamenti cromosomici per i quali si verifica, ad esempio, un dosaggio anomalo o la scorretta regolazione di uno o più geni [Lupski, 1998]. Analogamente, si ipotizza che la presenza nel genoma di varianti strutturali comuni quali le CNVs, sia responsabile di parte della comune

variabilità fenotipica degli individui. Per molte malattie genomiche (CMT1A, *Charcot-Marie-Tooth1A*; HNPP, *Hereditary Neuropathy with liability to pressure palsies*; SMS, *Smith-Magenis syndrome*), così come per le varianti strutturali comuni, si è osservato che le alterazioni a livello genomico che ne sono la causa possono essere ereditate dai genitori oppure insorgere *de novo* sporadicamente.

### **1.5c. Formazione di CNVs tramite NAHR**

Spesso le CNVs sono delimitate da punti di rottura che mappano in corrispondenza di specifiche *segmental duplications* (o low copy repeats, LCRs) fiancheggianti la delezione o la duplicazione in questione. Si ipotizza che, nel 99% dei casi, tali varianti strutturali siano il risultato di un evento di ricombinazione omologa non allelica (*non-allelic homologous recombination*, NAHR) causato dall'appaiamento scorretto tra sequenze ripetute che le fianleggiano. La NAHR avviene esattamente con lo stesso meccanismo della ricombinazione omologa (appaiamento filamenti omologhi, formazione del chiasma, risoluzione del crossing-over) con l'eccezione che, nella NAHR, l'alta omologia di sequenza dei segmenti cromosomici che costituiscono ad esempio le LCRs, fa sì che il meccanismo cellulare di controllo della ricombinazione non sia in grado di riconoscere l'appaiamento di segmenti non perfettamente identici [Lupski e Shaffer 2000]. L'appaiamento può avvenire fra LCRs appartenenti a cromosomi omologhi (*intercromosomico*), fra cromatidi fratelli (*intracromosomico*), all'interno di un singolo cromatidio (*intracromatide*) [Lupski and Stankiewicz, 2002]. In base al meccanismo di appaiamento, dell'orientamento e della complessità delle LCRs, la NAHR può risolversi in delezioni, duplicazioni, inversioni o altri riarrangiamenti più complessi che possono anche coinvolgere cromosomi non omologhi (Figura 8).

Solo in pochi casi sono stati osservati riarrangiamenti originatisi da eventi di ricombinazione non omologa (*Non-Homologous End-Joining*, NHEJ), un tipo di ricombinazione che non richiede un grado minimo di omologia di sequenza e che si basa sulla ligazione fra due segmenti di DNA in cui si è verificata una rottura del doppio filamento [Lupski e Stankiewicz, 2005].



**Figura 8:** Esempi di ricombinazione omologa non allelica : Le *Segmental duplications* fungono da substrato per la NAHR. (**a e b**) NAHR intercromosomale, intracromosomale, o intracromatidica tra sequenze ripetute orientate nello stesso senso causa delezioni e/o duplicazioni. (**c e d**) NAHR intercromosomale, intracromosomale, o intracromatidica tra sequenze ripetute orientate in senso opposto causa inversioni. Le sequenze ripetute sono rappresentate dalle box blu con l'orientamento indicato dalla freccia gialla. L'evento di ricombinazione è rappresentato da una croce rossa

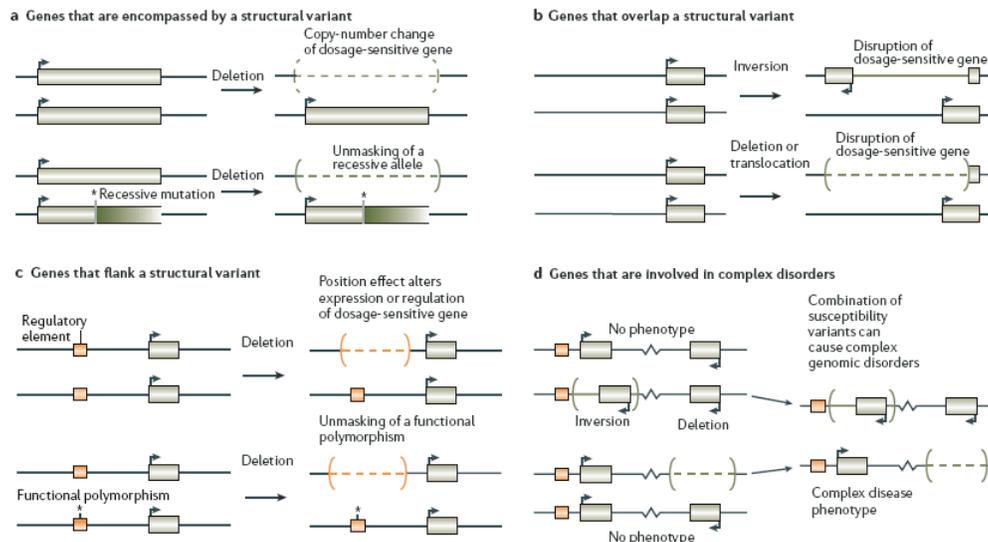
Oltre che fra LCRs, la NAHR può verificarsi in presenza di sequenze ripetute sub-telomeriche. Si è osservato che tali riarrangiamenti sono associati in modo significativo ad alcuni casi di ritardo mentale e/o a sindromi di malformazione [Flint et al., 1995; Knight et al., 1999]. Altri studi hanno invece osservato il verificarsi della NAHR in corrispondenza di sequenze ripetute e intersperse che, complessivamente, rappresentano circa il 30% del genoma. In particolare, sono stati osservati eventi di NAHR fra sequenze *Alu* (classe principale di *Short Interspersed Elements*, SINEs) e meno frequentemente fra sequenze L1 (classe principale di *Long Interspersed Elements*, LINEs) [Kolomietz et al., 2002; Burwinkel et al., 1998].

E' probabile che molte varianti strutturali non rappresentino eventi rari, ma bensì il risultato di una predisposizione al riarrangiamento cromosomico determinata dall'architettura stessa del genoma che agisce come catalizzatore dell'instabilità cromosomica.

#### ***1.5d. CNVs e fenotipo***

Gli studi effettuati sulle varianti strutturali nel genoma hanno mostrato che individui sani differiscono per il numero di copie di ampi e numerosi segmenti cromosomici [Conrad, 2006; Hinds, 2006; McCarroll, 2006]. Ad oggi non è chiaro in che misura tali riarrangiamenti cromosomici possano influire su tratti fenotipici complessi (ad esempio tratti comportamentali), essere causa di suscettibilità per una patologia o essere semplici varianti polimorfiche. Attualmente è tecnicamente impossibile vagliare in tutti gli individui l'associazione di CNVs a tutti i fenotipi possibili. Inoltre lo *screening* dovrebbe essere condotto su individui di tutte le età, poichè è stato osservato che alcune varianti strutturali, come il riarrangiamento genomico che causa HNPP (*Hereditary Neuropathy with liability to Pressure Palsies*), hanno una penetranza età-dipendente [Lupski and Chance, 2005].

In generale, le varianti strutturali possono essere causa di variabilità o dell'insorgere di una patologia, secondo molteplici meccanismi (Figura 9). Ad esempio, una CNV può influenzare direttamente il dosaggio di un gene oppure, indirettamente, alterarne l'espressione genica per effetto di posizione in quanto si trova in una regione regolatoria. Oppure una variante strutturale, pur essendo ininfluenza dal punto di vista dell'espressione genica, potrebbe predisporre il locus in cui risiede ad ulteriori riarrangiamenti cromosomici che invece causano un particolare fenotipo. Infatti, nei portatori di inversioni polimorfiche, la minore frequenza di ricombinazione causata dall'orientamento opposto dei segmenti cromosomici coinvolti, può aumentare la probabilità di un appaiamento non allelico fra CNVs o LCRs aumentando il rischio dell'insorgere di delezioni *de novo* o altri riarrangiamenti cromosomici a livello meiotico [Sankaranarayanan and Wassom, 2005].



**Figura 9:** Casistica di effetti che la presenza di una CNV può produrre a livello fenotipico.

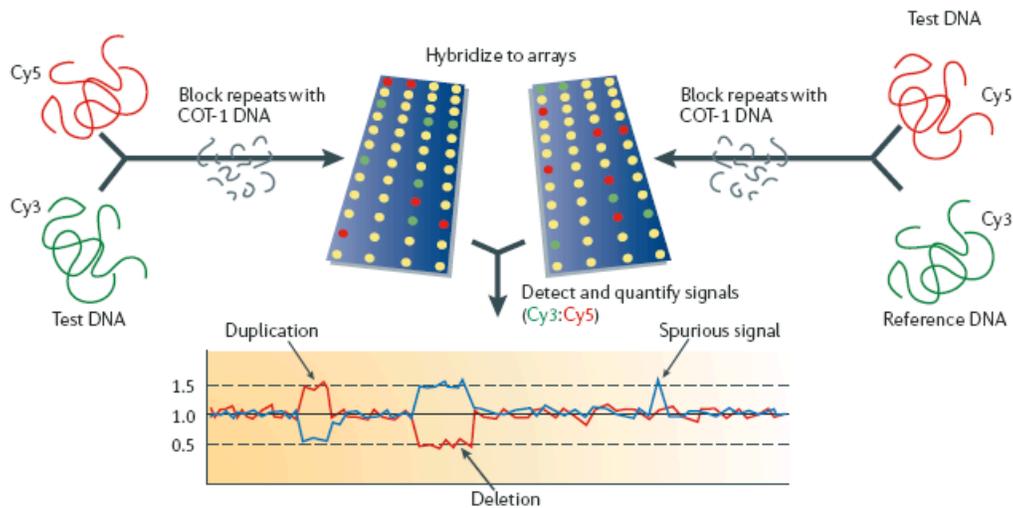
Quindi, le varianti strutturali possono o agire in modo diretto nel causare o predisporre ad una particolare patologia, oppure possono fungere da alleli di suscettibilità nel caso di malattie complesse. Sebbene alcune varianti sembrano essere benigne e frequenti in certe popolazioni, in combinazione con altri fattori genetici e/o ambientali (compresi SNPs, CNVs o inversioni) potrebbero contribuire al fenotipo malattia.

### 1.5e. Metodi per l'identificazione di CNVs

Lo sviluppo di nuove strategie sperimentali e computazionali ha permesso di analizzare ad una risoluzione sempre maggiore le varianti strutturali umane [Speicher et al., 2005].

L'identificazione di nuove varianti strutturali segue principalmente due approcci: uno basato sull'uso di array specificamente progettati, l'altro su metodi di PCR quantitativa.

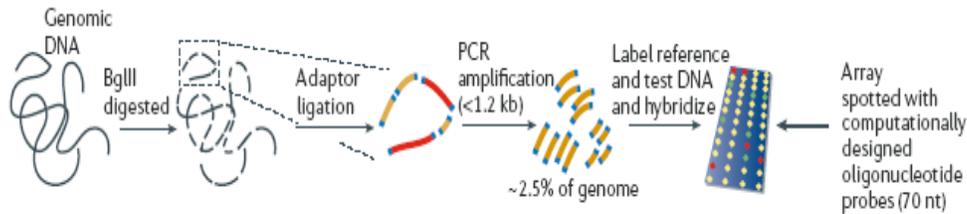
**Approcci basati sull'utilizzo di array:** l'approccio attualmente più robusto è quello basato su esperimenti di genomica comparativa con array-CGH (*array-based Comparative Genome Hybridization*) [Solinas-Toldo et al., 1997; Pinkel et al., 1998]. Questo approccio sfrutta l'ibridazione competitiva di segmenti di un genoma di interesse e di un genoma di confronto diversamente marcati ad un *array* di target genomici (BACs, cDNAs, prodotti di PCR, oligonucleotidi) (Figura 10).



**Figura 10:** Schema di un esperimento con array-CGH.

In particolare, gli array-CGH con BACs [Iafate et al., 2004; Tuzun et al., 2005] sono molto diffusi perché provvedono ad un'ampia copertura del genoma (1 clone BAC ogni 1 Mb), perché sono disponibili mappe di BACs altamente affidabili, perché i cloni BAC sono facilmente reperibili e perché il loro limite di risoluzione è abbastanza alto, (parte da 50Kb fino a raggiungere, teoricamente, poche Kb) [Feuk et al., 2006].

Gli array-CGH con oligonucleotidi di 60-100 bp sono in grado di migliorare tale limite di risoluzione e vengono in genere utilizzati con un metodo detto ROMA (*Representational Oligonucleotide Microarray Analysis*) [Lucito et al., 2003; Sebat et al., 2004]. Il metodo ROMA consente di aumentare il rapporto “segnale/rumore” riducendo la complessità del DNA utilizzato per l'analisi [Kennedy et al., 2003]. Infatti, il DNA che deve essere ibridato all'array viene precedentemente trattato con enzimi di restrizione e poi legato ad alcune sequenze “adattatore”. Tali frammenti di sintesi vengono quindi amplificati ottenendo un pool di segmenti di lunghezza paragonabile che sono rappresentativi dell'intero genoma ma hanno una complessità inferiore (Figura 11).



**Figura 11:** Metodo ROMA (*Representational Oligonucleotide Microarray Analysis*), una variante degli esperimenti con array-CGH.

Un altro metodo ultimamente molto utilizzato sfrutta gli **array per il genotyping di SNPs** allo scopo di rilevare anche CNVs [Slater et al. 2005]. In questo caso le intensità dei segnali di ibridazione di un campione di DNA genomico di interesse sono confrontati con il valore medio di un DNA di controllo. Una deviazione dell'intensità prodotta dal campione rispetto a quella prodotta dal controllo è indicativa di una variazione del numero di copie e può corrispondere ad esempio ad una delezione o presenza di disomia uniparentale (che può essere considerata anche una particolare forma di variante strutturale) [Lindblad-toh et al., 2000; Mei et al., 2000; Altug-Teber et al., 2005; Raghavan et al., 2005; Bruce et al., 2005].

Un approccio simile è stato utilizzato nel corso di questa tesi per l'individuazione di CNVs che potessero essere eventualmente associate all'autismo.(Cap.4, § 4.1)

**Approcci basati su PCR quantitativa:** per identificare CNVs in specifiche regioni genomiche è preferibile un approccio che si basa su PCR. In particolare il metodo della *real-time quantitative PCR* (qPCR) è molto affidabile per l'individuazione di singole delezioni o duplicazioni, ma si adatta più difficilmente ad esperimenti che mirano alla rilevazione di diverse CNVs contemporaneamente in diverse regioni genomiche [Ponchel et al. 2003]. A tale proposito sono stati messi a punto metodi alternativi quali la QMPSF (*Quantitative Multiplex PCR of Short Fluorescent Fragments*) [Charbonnier et al. 2000], la MAPH (*Multiplex Amplifiable Probe Hybridization*) [Armour et al. 2000] e la MLPA (*Multiplex Ligation-dependent Probe Amplification*) [Schouten et al 2002] che sono in grado di rilevare in un solo esperimento la presenza di diverse CNVs fino a 40 regioni diverse.

Ogni metodo per l'identificazione di varianti strutturali nel genoma risulta ottimale per alcune tipologie di varianti strutturali. Quindi, uno studio completo di questo tipo di variabilità a livello genomico richiederà, in futuro, un approccio integrato di più metodi oltre che la validazione con

metodi indipendenti (es. FISH) delle CNVs individuate. Di attuale interesse è la definizione di “genoma di riferimento”, in quanto su di esso si basano tutti i metodi per l’individuazione di CNVs. L’assenza di un genoma di riferimento standard causa infatti problemi nel confronto dei risultati ottenuti in progetti diversi e ostacola la costruzione di un banca dati universale delle varianti strutturali del genoma.

## **2. Autismo**

### **2.1. Autismo: caratteri generali**

Il termine **autismo** è stato reso pubblico per la prima volta intorno alla metà del secolo scorso da Leo Kanner e Hans Asperger [Kanner, 1943; Asperger, 1944] che individuarono nei bambini uno specifico stato psicopatologico differente dal ritardo mentale [Kanner, 1943]. Da allora, e specialmente negli ultimi 20 anni, il quadro della sintomatologia del disturbo autistico si è delineato con precisione sempre maggiore.

L'autismo non rappresenta una categoria di disordini distinta, ma il prototipo di uno spettro più ampio di problemi nell'interazione sociale, comportamentale e della comunicazione.

Tali disordini sono descritti nel *Diagnostic and Statistical Manual of Mental Disorders (DSM-IV)* e nell' *International Classification of Mental and Behavioral Disorders (ICD-10)* sotto il termine generico di “disturbi pervasivi dello sviluppo” (*Pervasive Developmental Disorders, PDD*) o “disturbi dello spettro autistico” (*Autism Spectrum Disorders, ASD*), un gruppo di disturbi che oltre all'autismo propriamente detto, comprendono anche la sindrome di Asperger (AS), la sindrome di Rett (RTT) il disturbo pervasivo dello sviluppo non altrimenti specificato (*non pervasive developmental disorder-not otherwise specified, PDD-NOS*), il disturbo disintegrativo dell'infanzia (*childhood disintegrative disorder, CDD*) [Bailey et al. 2001].

### **2.2. Fenotipo clinico e criteri diagnostici**

L'autismo si manifesta con sintomi variabili da un individuo all'altro e secondo molteplici livelli di gravità anche fra membri della stessa famiglia. Generalmente i sintomi della malattia si riscontrano entro i primi tre o quattro anni di età e persistono per tutta la vita dell'individuo.

Le manifestazioni cliniche che si possono osservare negli individui con disturbo autistico sono suddivisibili in tre distinte categorie:

- **Problemi nella comunicazione verbale e non verbale:**

La persona utilizza il linguaggio in modo bizzarro o appare del tutto muta; spesso ripete parole, suoni o frasi che sente pronunciare (ecolalia). Anche se le capacità imitative sono integre, queste persone in molti casi hanno notevoli difficoltà ad impiegare i nuovi apprendimenti in modo costruttivo in situazioni diverse. Molte volte è presente ritardo mentale. Mostrano un ritardo nel

linguaggio parlato e possono riferirsi a sé stessi con il nome proprio piuttosto che con “io” o “me”.

- ***Difficoltà nell'interazione sociale:***

Apparente carenza di interesse e di reciprocità con gli altri; tendenza all'isolamento e alla chiusura; apparente indifferenza emotiva agli stimoli o ipereccitabilità agli stessi; difficoltà ad instaurare un contatto visivo (es. guardare negli occhi le persone ), ad iniziare una conversazione o a rispettarne i turni, difficoltà a rispondere alle domande e a partecipare alla vita o ai giochi di gruppo. Appaiono inconsapevoli dei sentimenti altrui nei propri confronti e dell'impatto negativo del proprio comportamento sugli altri. Non è infrequente che i bambini affetti da autismo siano inizialmente diagnosticati come sordi perché non mostrano alcuna reazione quando sono chiamati per nome.

- ***Presenza di comportamenti ripetitivi e stereotipati ed interessi ristretti:***

Di solito un limitato repertorio di comportamenti viene ripetuto in modo ossessivo; si possono osservare sequenze di movimenti stereotipati (es. torcersi o mordersi le mani, sventolarle in aria, dondolarsi, ecc.) detti appunto stereotipie. Queste persone possono manifestare eccessivo interesse per gli oggetti o parti di essi, in particolare se hanno forme tondeggianti o possono ruotare (biglie, trottole, eliche, ecc.). Alcuni parlano con voce cantilenante di un numero ristretto di argomenti preferiti. Si riscontra una resistenza al cambiamento che per alcuni può assumere le caratteristiche di un vero e proprio terrore fobico. La persona può esplodere in crisi di pianto e riso. Può diventare autolesionista, iperattiva ed aggressiva verso altri o verso oggetti. Al contrario, alcuni mostrano un'eccessiva passività e ipotonia che sembra renderli impermeabili a qualsiasi stimolo.

***Altri comportamenti osservati comunemente in soggetti autistici:***

Le persone autistiche spesso hanno risposte anomale ai suoni, al tatto o ad altri stimoli sensoriali. Molti mostrano una ridotta sensibilità al dolore, ma possono anche essere straordinariamente sensibili ad altre sensazioni. Questa sensibilità alterata può contribuire ai sintomi comportamentali, come la resistenza ad essere abbracciati. Presentano inoltre: disturbi del sonno, preferenze limitate per i cibi e problemi nel camminare e nello scrivere.

Secondo stime più recenti, l'autismo si presenta nella popolazione con una frequenza quattro volte superiore negli individui di sesso maschile rispetto agli individui di sesso femminile (male : female = 4 : 1), valore che tende ad aumentare se si considerano forme più lievi.

La percentuale di casi di autismo nella popolazione generale sembra essere aumentata nel corso degli ultimi vent'anni, con una stima di circa 10 individui su 10.000 per l'autismo in senso stretto fino ad un valore di 60 individui su 10.000 per le forme di ASD più lievi [Fombonne, 2003]. Si ritiene che questo apparente aumento di casi sia dovuto principalmente ad una maggiore conoscenza degli ASDs da parte degli operatori ed al miglioramento dei metodi di diagnosi. Per la diagnosi, gli strumenti più comunemente utilizzati a livello internazionale sono l'**ADI-R** (*Autism-Diagnostic Interview-Revised*) [Lord et al., 1994] e l'**ADOS** (*Autism Diagnostic Observation Schedule*) [Lord et al., 2000] che, insieme, permettono una diagnosi corretta e non ambigua, fondamentale non solo dal punto di vista clinico ma anche dal punto di vista dell'analisi genetica per un campionamento degli individui che sia il più omogeneo possibile.

### **2.3. Trattamenti terapeutici**

Attualmente non esiste una cura per l'autismo. Le terapie o gli interventi vengono scelti in base ai sintomi specifici di ciascun individuo e comprendono un programma psicopedagogico, volto a migliorarne le capacità cognitive e comportamentali, e un trattamento farmacologico. Sebbene questi interventi non curino l'autismo, spesso portano a un miglioramento sostanziale.

**Trattamenti educativi e comportamentali** : si lavora con il bambino per aiutarlo a sviluppare capacità sociali e di linguaggio. Poiché i bambini imparano tanto più efficacemente tanto più sono piccoli, questo tipo di terapia dovrebbe iniziare il più presto possibile. Recentemente è stato evidenziato che un intervento precoce ha buone possibilità di influenzare positivamente lo sviluppo cerebrale.

**Trattamenti farmacologici**: poiché l'eziologia del disturbo autistico e delle disfunzioni ad esso connesse non sono attualmente note, ogni terapia farmacologica è da considerarsi esclusivamente sintomatica. Le terapie più diffuse mirano a ridurre alcuni disturbi associati all'autismo: scarsa attenzione e concentrazione, fenomeni ossessivi e compulsivi, eccessiva ansietà, anomalie del sonno, depressione e autolesionismo. In particolare, i farmaci che regolano la via serotoninergica, come gli inibitori dei trasportatori o del *re-uptake* selettivo della serotonina (SSRIs), riducono i comportamenti ripetitivi ed aggressivi [McDougle et al. 1996; Gordon 1993; Sanchez et al. 1996; Remington 2001]. Altri farmaci inibiscono la trasmissione dell'impulso nei neuroni dopaminergici, riducendo alcune sintomatologie comportamentali [Anderson et al. 1989; Naruse et al. 1982]. Altri ancora agiscono sul sistema adrenergico e noradrenergico riducendo

l'iperattività [Gordon 1992; Jaselskis 1992; Lichter 1996]. Di particolare interesse per lo scopo di questa tesi è l'assunzione di farmaci a base di uridina per mitigare i disturbi del sonno [Page et al 1997] (Cap. 3, § 3.1).

#### ***2.4. Autismo: dalla sua prima descrizione al riconoscimento dell'importanza dei fattori genetici***

***Fattori ambientali nell'autismo:*** Come già detto, il termine ***autismo*** è stato reso pubblico per la prima volta intorno alla metà del secolo scorso da *Leo Kanner* (1943) il quale descrisse 11 bambini (8 maschi, 3 femmine) la cui condizione differiva dal ritardo mentale. Kanner notò che la caratteristica che accomunava questi bambini era l'“inabilità innata per i contatti interpersonali” e denominò questo disturbo come “autismo infantile” [Kanner, 1943].

Un anno dopo in Germania *Hans Asperger* descrisse casi simili che chiamò “psicopatia autistica” [Asperger, 1944].

Negli anni '60-'70 alcuni studiosi, quali Bergmann, Escalona, Rimlaand e Pronovost, sostennero l'ipotesi di una base psicologica nell'autismo

Secondo questi ultimi la causa della sindrome autistica era da ricercare nel rapporto patologico del bambino con i genitori e in particolar modo con le madri, definite “madri frigorifero”, ed in particolare sulla rottura troppo precoce del rapporto di dipendenza madre-bambino o su una non risposta della madre al bambino.

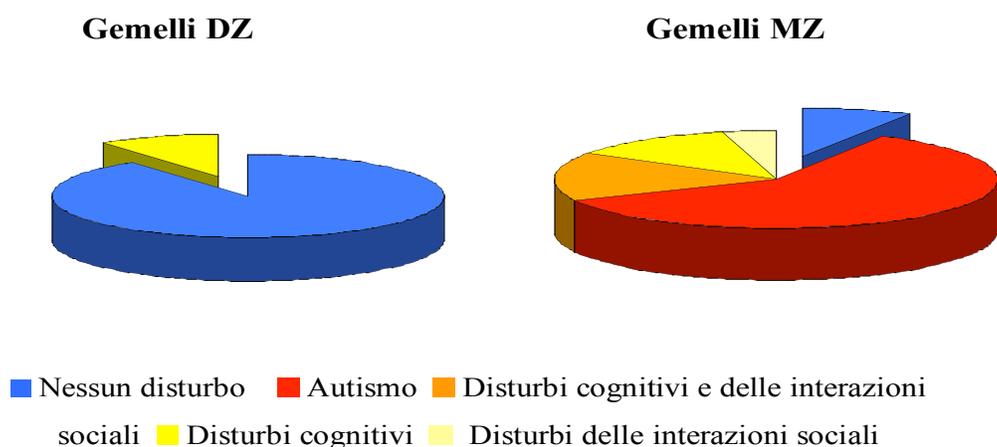
Successivamente, tuttavia l'infondatezza delle teorie psicologiche è stata chiaramente dimostrata. Nella ricerca di fattori ambientali che potessero dare un contributo all'eziologia del disturbo sono stati presi in esame gli alimenti nella dieta di quella minoranza di bambini autistici che mostravano problemi gastrointestinali [Page, 2000]. È stata anche avanzata l'ipotesi di un'infezione virale da Rosolia, *Hemophilus influenzae*, citomegalovirus durante o dopo la gravidanza, o l'utilizzo di vaccini contenenti mercurio (come il vaccino trivalente *measles-mumps-rubella*, MMR) [Wakefield et al., 1998] o thimerosal. Tuttavia nessuna di queste associazioni è risultata significativa [Peltola et al., 1998; Taylor et al., 1999; Medical Research Council, 2001].

***Studi neuroanatomici e neurochimici:*** Nel tentativo di individuare specifiche anomalie neuroanatomiche associate all'autismo sono stati condotti studi post-mortem di cervelli di individui affetti [Baily et al., 1998]. I vari studi hanno evidenziato anomalie dello sviluppo corticale, un aumento della densità neuronale a livello del sistema limbico, anomalie nello

sviluppo del tronco encefalico ed un ridotto numero delle cellule del Purkinje nel cervelletto [Baily et al., 1998; Kemper e Bauman, 1998]. Tuttavia nessuna di queste anomalie è stata riscontrata in un numero significativo di individui. La scoperta di tipo neuroanatomico più interessante è stata l'osservazione che circa il 20% degli individui autistici, provenienti da sette studi diversi, mostrava macrocefalia [Miles et al., 2000]. Questa associazione si è dimostrata così chiara in alcune popolazioni da suggerire l'utilizzo della circonferenza cranica come uno degli strumenti di *screening* per l'autismo [Bolton et al., 2001].

Altri studi hanno evidenziato il coinvolgimento di vari neurotrasmettitori; le evidenze più significative riguardano la presenza di elevati livelli di serotonina (5-HT) nel sangue, nelle urine, nel liquido cerebro-spinale e nelle piastrine di individui autistici [Schain & Freedman, 1961; Gillberg & Coleman et al., 1992; Cook&Levental, 1996]. Questi dati hanno acceso l'interesse per il sistema serotoninergico, dato il ruolo della serotonina nello sviluppo del sistema nervoso centrale (CNS) e il suo coinvolgimento nei meccanismi alla base del comportamento sociale ed affettivo, del sonno, dell'aggressività e dell'ansia: tutti aspetti compatibili con alcune caratteristiche fenotipiche dell'autismo. Il gene che codifica per il trasportatore della serotonina (*5-HTT*) è stato oggetto di studio come possibile gene candidato per l'autismo, ma i risultati ottenuti sono contrastanti e poco chiari [Cook et al., 1997; Klauck et al., 1997; Maestrini et al., 1999; Zhong et al., 1999; Persico et al., 2000; Cook et al., 2001].

***I fattori genetici nell'autismo:*** Alcuni decenni dopo la prima descrizione dell'autismo, a partire dagli anni '70, sono stati condotti i primi studi epidemiologici su famiglie e su coppie di gemelli che riconobbero il ruolo centrale della genetica nell'eziologia dell'autismo [Rutter e Folstein, 1977; Rutter, 2000; Folstein and Rosen-Sheidley, 2001; Lamb et al., 2002]. Nello studio più ampio condotto finora [Bailey et al. 1995] il tasso di concordanza per l'autismo in coppie di gemelli monozigotici (MZ) risulta notevolmente alto (circa 60%), rispetto al valore riscontrato nelle coppie di gemelli dizigotici dello stesso sesso (DZ) (circa 3%). Se si considerano anche disturbi sociali e di comunicazione meno gravi, ma qualitativamente simili a quelli dell'autismo (il cosiddetto "*broader phenotype*"), il tasso di concordanza nei gemelli MZ aumenta fino a 92% e nei DZ raggiunge il 10% (Figura 12).



**Figura 12:** Concordanza nei gemelli monozigotici (MZ) e diziotici (DZ) dello spettro dei disturbi connessi all'autismo.

Inoltre, studi di ricorrenza familiare [Szatmari et al., 1998] stimano che la percentuale di rischio in fratelli di individui affetti sia attorno al 3-4%, quindi con un rischio di ricorrenza ( $\lambda_s$ ) per i fratelli di bambini autistici di 30-40 volte maggiore rispetto al tasso di prevalenza nella popolazione generale.

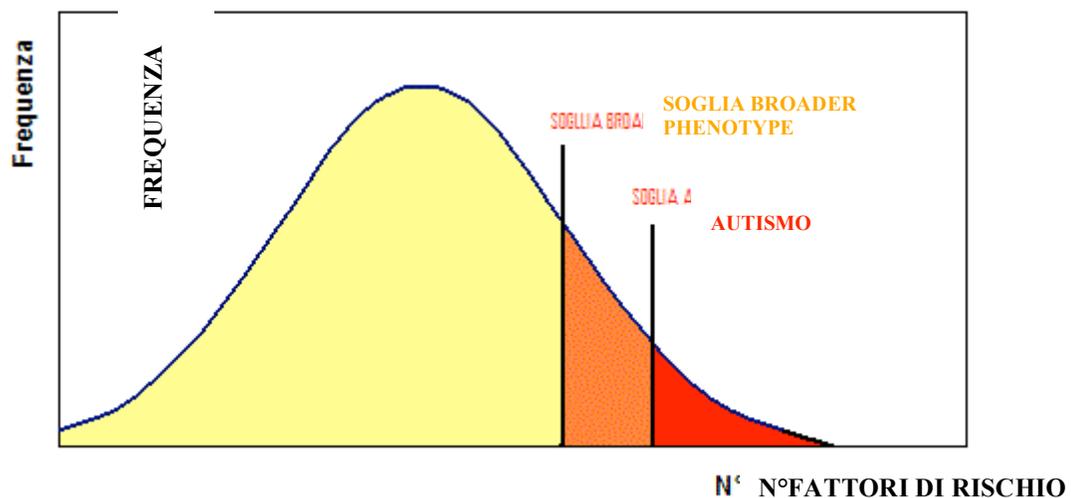
Gli studi su famiglie e su coppie di gemelli hanno evidenziato che individui che fanno parte di una famiglia in cui c'è un soggetto autistico, presentano più frequentemente caratteristiche del *broader phenotype*. Il *broader phenotype* si differenzia dall'autismo in quanto comprende lievi anomalie del comportamento quali eccessiva timidezza, distacco e disinteresse per le interazioni sociali, ansia e interessi limitati, ma non è associato con ritardo mentale o fenomeni di epilessia [Rutter, 2000]. Attualmente non è chiaro se il *broader phenotype* derivi dall'aver ereditato un numero inferiore di geni di suscettibilità, un pattern diverso di tali geni, o entrambe queste condizioni. Probabilmente, all'interno di una stessa famiglia, rispetto ad un individuo che presenta *broader phenotype*, l'autismo propriamente detto si manifesta quando un individuo eredita un numero maggiore di fattori di rischio (o vi è sottoposto a causa di particolari fattori ambientali).

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

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

La complessità clinica e l'eterogeneità del fenotipo autistico riflettono probabilmente la presenza di eterogeneità genetica (geni differenti o combinazioni diverse di geni possono essere coinvolti in famiglie diverse). Altri fattori come il sesso o i fattori ambientali possono dare un loro contributo ad una particolare espressione fenotipica. Inoltre meccanismi epigenetici come difetti nella metilazione del DNA o anomalie nell'imprinting sono stati proposti come possibili fattori coinvolti nell'eziologia del disturbo autistico [Abdolmaleky et al., 2004; Lamb et al., 2005].

Probabilmente varianti funzionali nei geni di suscettibilità hanno un effetto debole o moderato nel processo patologico: ciascun locus, da solo, contribuisce ad aumentare il rischio di sviluppare la malattia, ma non è in grado di determinare il completo fenotipo clinico (Figura13).



**Figura 13:** Modello multifattoriale a soglia: il raggiungimento della soglia del *broader phenotype* richiede un numero inferiore di fattori di rischio rispetto al raggiungimento della soglia del disturbo autistico.

## **2.5. Disordini monogenici associati all'autismo**

La presenza di eterogeneità genetica è anche indicata dal fatto che in una minoranza di casi (<10%) l'autismo è associato a patologie causate da un singolo gene, tra cui la sindrome dell'X fragile (FRAXA), la sclerosi tuberosa (TSC), la neurofibromatosi [Klauck, 2006] o a varie anomalie cromosomiche [Smalley et al. 1998; Dykens and Volkmar, 1997; Feinstein and Reiss, 1998].

**TSC:** La Sclerosi Tuberosa (Omim # 191100) è un disordine autosomico dominante neurocutaneo caratterizzato dalla presenza di tubercoli nella regione corticale e cerebrale, lesioni renali, attacchi epilettici durante l'infanzia, resistenza a trattamenti antiepilettici [Bolton et al., 1997; Bolton et al., 2002; Gutierrez et al., 1998; Hunt et al., 1987; Jambaque et al., 2000]. Questo disordine è causato da mutazioni in due diversi geni: *TSC1* (Omim # 605284) e *TSC2* (Omim # 191092), che mappano rispettivamente nelle regioni cromosomiche 9q34 e 16p13 [Roach et al., 2004]. Studi epidemiologici hanno mostrato che la prevalenza di Sclerosi Tuberosa in bambini autistici e di autismo in casi di Sclerosi Tuberosa è cento volte superiore a quella attesa [Harrison et al., 1997; Fombonne, 2003].

**FRAXA:** La Sindrome dell'X-fragile (Omim # 300624) è un disordine caratterizzato da ritardo mentale con un'incidenza di 1 : 4000 negli individui di sesso maschile e 1 : 8000 negli individui di sesso femminile [Lombroso, 2003]. Questo disordine è causato dalla ipermetilazione di una tripletta (CGG) altamente ripetuta (>200 *repeats*) presente nella regione non tradotta al 5' del gene *FMRI* (Omim # 309550) localizzato sul cromosoma Xq27; con conseguente riduzione del prodotto proteico [Oostra et al., 2001; Sutcliffe et al., 1992]. Il 2-5% di bambini con una diagnosi di autismo più lieve sono portatori di mutazioni associate a FRAXA [Fombonne, 2003; Bailey, 1993; Reddy, 2005; Wassink et al., 2001].

**Neuroligine:** Le neuroligine (*NLGN*) rappresentano una famiglia di proteine transmembrana coinvolte nell'adesione cellulare post-sinaptica con un ruolo importante nella sinaptogenesi e nel rimodellamento delle sinapsi [Zoghbi, 2003]. Uno studio interessante ha identificato la presenza di una sostituzione missenso nel gene *NLGN3* (Omim # 300336), in corrispondenza di un dominio funzionale estremamente conservato della proteina, e una mutazione *frameshift* nel gene *NLGN4* (Omim # 300427), che genera un codone di stop e causa la terminazione prematura della proteina [Jamain et al., 2003]. In uno studio più recente condotto su una famiglia estesa, nel gene *NLGN4* è stata identificata una mutazione *frameshift* che segrega con ritardo mentale o con ASD [Laumonnier et al., 2004]. Lo *screening* di mutazioni di questi due geni, condotto su ulteriori individui affetti, ha comunque prodotto risultati contrastanti [Gauthier et al., 2005; Vincent et al.,

2004; Yan et al., 2005]. L'ipotesi più plausibile, comunque, è che difetti funzionali di questi geni possano influire sullo sviluppo cognitivo e condurre ad un fenotipo autistico solo in una piccola percentuale di casi.

***NRXNI***: Il gene neurexina 1 (*NRXNI*) (Omim # 600565) mappa nella regione 2q32 e codifica per una proteina di superficie altamente espressa nel cervello. Le neurexine rappresentano una famiglia di proteine coinvolte nella formazione e nel mantenimento delle sinapsi attraverso il legame con le neuroligine che, come già detto, se mutate possono avere un ruolo nella manifestazione del fenotipo autistico. Secondo Graf e collab. [2004], le neurexine inducono il differenziamento post-sinaptico nel contatto tra i dendriti, mentre le neuroligine inducono il differenziamento pre-sinaptico negli assoni glutamatergici [Graf et al., 2004], a dimostrazione del fatto che il legame “neurexine-neuroligine” è di fondamentale importanza nella sinaptogenesi glutamatergica. In un recentissimo studio condotto dall'*Autism Genome Project Consortium* [AGP, 2007], attraverso la genotipizzazione su larga scala di SNPs, in un campione di 1168 famiglie *multiplex*, è stata analizzata la correlazione tra le *Copy Number Variants* (CNVs) individuate e il fenotipo autistico. Quest'analisi ha permesso di individuare una delezione submicroscopica, in eterozigosi, in una coppia di fratelli affetti in corrispondenza di esoni codificanti del gene *NRXNI*, delezione non presente in un campione di individui di controllo. Questa osservazione, accoppiata alla scoperta di varianti rare missenso in individui autistici da parte Feng e collab [2006] e alle conoscenze sulle interazioni tra neurexine e neuroligine è molto interessante in quanto mette nuovamente in risalto il ruolo del sistema glutamatergico nella manifestazione del fenotipo autistico.

***SHANK3***: Il gene *SHANK3* (*SH3 and Multiple Ankyrin Repeat Domains 3*) (Omim # 606230) mappa nella regione 22q13.3 e codifica per una proteina espressa preferenzialmente nella corteccia cerebrale e nel cervelletto [Bonaglia et al., 2001]. Studi molto recenti condotti su individui con caratteristiche fenotipiche riconducibili a quelle dello spettro autistico, quali ritardo mentale e/o problemi nel linguaggio, hanno individuato in corrispondenza di *SHANK3* inserzioni o delezioni parziali e/o totali [Bonaglia et al., 2001; Anderlid et al., 2002; Wilson et al., 2003; Durand et al., 2007] con la conseguenza di un dosaggio anomalo del prodotto proteico corrispondente.

I geni specifici che contribuiscono alla patogenesi nella restante maggioranza di forme idiopatiche di autismo rimangono ad oggi sconosciuti.

## **2.6. Identificazione dei geni di suscettibilità dell'autismo: metodi usati per il mappaggio**

Benchè l'autismo e i disordini dello spettro autistico manifestino un elevato grado di ereditabilità, identificare i geni di suscettibilità per questo tipo di disordini si è rivelato alquanto difficoltoso. Le difficoltà maggiori sono rappresentate dal modello di ereditarietà che, per l'autismo, come già detto, è un modello complesso, con un probabile coinvolgimento di più geni che interagiscono fra loro (da 3 a 10 loci stimati) in maniera epistatica e presenza di eterogeneità.

Quindi, non essendoci chiari indizi sulle responsabilità genetiche dell'autismo, una prima fase per l'identificazione dei geni di suscettibilità è l'individuazione di specifiche regioni sub-cromosomiche potenzialmente correlate alla malattia.

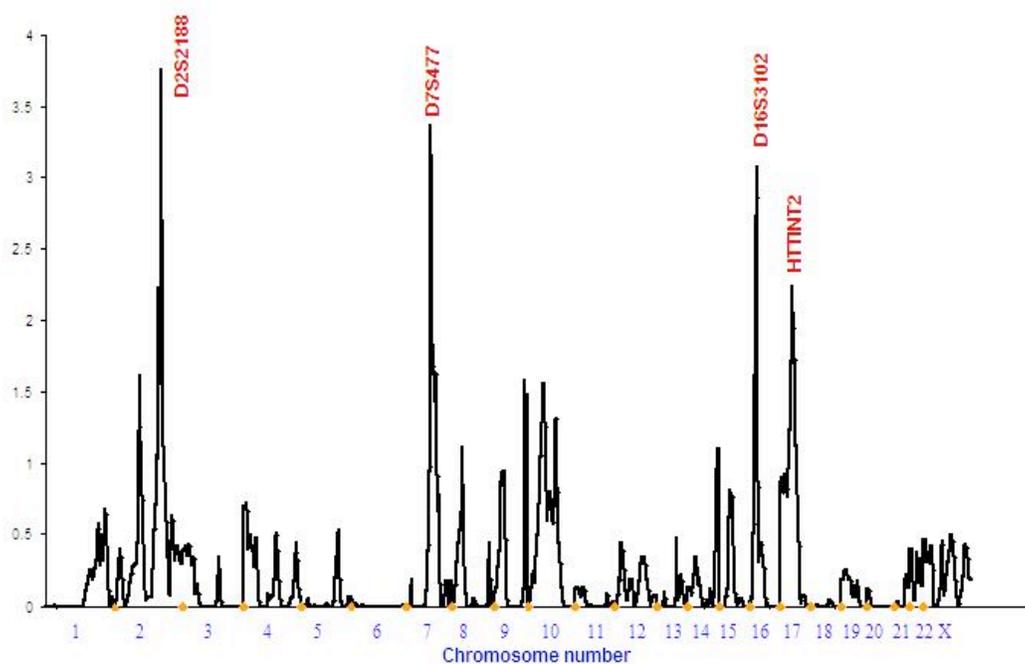
In assenza di anomalie cromosomiche o modelli animali della malattia, che potrebbero indicare la regione di interesse, questa localizzazione è effettuata mediante analisi di *linkage* seguite da analisi di associazione a più alta risoluzione.

### **2.6a. Studi di linkage per l'autismo estesi all'intero genoma**

Essendo ampiamente sconosciute le basi neurobiologiche dell'autismo, negli ultimi anni diversi gruppi di ricerca hanno effettuato studi di *linkage* non parametrici estesi all'intero genoma con lo scopo di identificare quelle regioni che contengono potenziali loci di suscettibilità (*affected-relative pair method*, descritto nel Cap 1) [Ashley-Koch et al., 1999; Auranem et al. 2002; Barret et al., 1999; Barlett et al., 2005; Buxbaum et al., 2001; Cantor et al., 2005; Coon et al., 2005; IMSCAC, 1998; IMGSAAC, 2001; Lamb et al., 2005; Lauritsen et al., 2006; Liu et al., 2001; McCauley et al., 2005; Philippe et al., 1999; Risch et al. 1999; Shao et al., 2002; Vincent et al., 2005; Yonan et al., 2003; Ylisaukko et al., 2004].

Il primo *genome scan* sull'autismo è stato portato a termine, nel 1998, dall' ***International Molecular Genetic Study of Autism Consortium*** (IMGSAAC), un consorzio composto da ricercatori e clinici di diversi paesi (Gran Bretagna, Germania, Francia, Olanda, Danimarca, Grecia e U.S.A.) e di cui fa parte anche il nostro gruppo di ricerca. La necessità di costruire un consorzio internazionale per questo tipo di studi deriva dalla relativa rarità di famiglie con più soggetti affetti e dalla necessità di analizzare un grande numero di famiglie diagnosticate con criteri omogenei. In un campione di 83 coppie di fratelli affetti (*affected sibling pairs*, ASPs), sono stati analizzati 354 marcatori ed è stato calcolato il MLS *multipoint* lungo i 23 cromosomi. Sono state individuate 12 regioni, ma la regione con maggiore evidenza di *linkage* si è rivelata la regione 7q con un MLS=2,53 [IMGSAAC, 1998]. Nel 2001, dallo stesso consorzio, il *genome scan* è stato esteso ad un campione di 152 ASPs, utilizzando 119 marcatori, all'interno delle regioni di

*linkage* individuate nel primo *genome scan*. La regione con maggiore evidenza di *linkage* è risultata essere localizzata nella porzione cromosomica **2q** (MLS=3.74) in corrispondenza del marcatore D2S2188, mentre se si considerano le coppie di fratelli affetti che soddisfano i criteri diagnostici più stringenti per l'autismo il valore di MLS sale a 4.80 [IMGSAC, 2001]; tale valore rappresenta un risultato estremamente significativo ed il dato più importante finora ottenuto nei differenti *genome scans* effettuati per l'autismo. Altre regioni in cui è stato osservato un valore significativo di *linkage* sono la regione 7q (MLS=3.20), 16p (MLS=2.93), 16p (MLS=2.93) e 17q (MLS=2.34) [IMGSAC, 2001] (Figura 14).



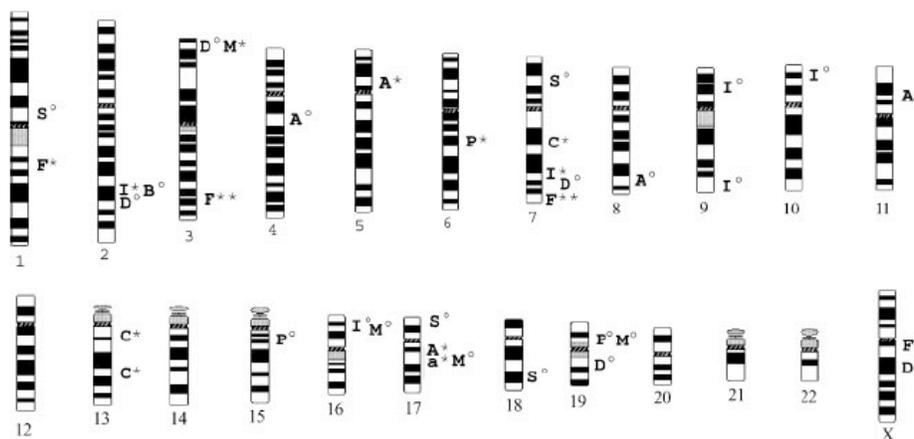
**Figura 14:** Risultati del primo *genome scan* effettuato dall'IMGSAC. In rosso sono indicati i marcatori in corrispondenza dei quali si ottiene il valore di LODscore più significativo rispettivamente sui cromosomi 2q, 7q, 6p, 17q.

Nel 2005 il consorzio ha esteso l'analisi a 219 ASPs analizzando 396 marcatori in corrispondenza dei loci precedentemente individuati sui cromosomi 2, 7, 9, 10, 16, 17 e includendo nell'analisi, sulla base dei risultati riportati in letteratura, anche i cromosomi 3, 13 e 15 [Lamb et al., 2005]. Valori di  $MLS > 2$  sono stati trovati per i cromosomi **2q** (MLS=2.54) in corrispondenza del marcatore D2S2314, posto nelle vicinanze del marcatore D2S2218 individuato nel primo *genome*

scan, 7q (MLS=2.31) in corrispondenza del marcatore D7S530 e 9p (MLS=2.12) in corrispondenza del marcatore D9S161.

I dati di *linkage* ottenuti dall'IMGSAC sono stati, in parte, replicati da altri 8 *genome scans*, condotti da altrettanti gruppi, utilizzando campioni di famiglie indipendenti, da cui sono emersi valori positivi (MLS>1) per regioni a livello del cromosoma 6q [Philippe et al., 1999], 7q e 13q [Barrett et al., 1999], 1p [Rich et al., 1999], 2q [Buxbaum et al., 2001], 5q, 7q, 19q e X [Liu et al., 2001], 2q, 3p, 7q, 19q, Xq [Shao et al., 2002], 1p, 3q, 7q, 17q, Xq [Auranen et al., 2003], e a livello dei cromosomi 17q e 5p [Yonan et al., 2003] a dimostrazione dell'eterogeneità di questa patologia.

La figura qui sotto (Figura 15) mostra un quadro generale di tutte le regioni cromosomiche che presentano valori significativi di *linkage* individuate da dieci *genome-wide scans* effettuati da diversi gruppi di ricerca nel corso degli ultimi dieci anni.



**Figure 15:** Regioni cromosomiche che presentano valori significativi di linkage individuate da 10 genome-wide scans effettuati per la ricerca di loci di suscettibilità per ASD.

I risultati di ogni studio sono indicati da una lettera seguita da \*\* se  $MLS > 3.6$ , da \* se  $MLS > 2.2$ , e da ° se  $MLS > 1$ . **A:** AGRE [Yonan et al., 2003]; **(a)** AGRE2 [Cantor et al., 2005]; **(B)** [Buxbaum et al., 2001], **(C)** CLSA [Barrett et al., 1999]; **(D)** Duke [Shao et al., 2002b]; **(F)** Finland [Auranen et al., 2002]; **(I)** IMGSAC [Lamb et al., 2005]; **(M)** [McCauley et al., 2005]; **(P)** PARIS [Philippe et al., 1999]; **(S)** Stanford [Risch et al., 1999].

Le regioni sui cromosomi **2q**, 7q, 3p, 17q, 19q sono quelle per le quali sono stati trovati risultati parzialmente sovrapponibili, confermati da almeno due diversi studi di *linkage* condotti da gruppi indipendenti.

La sovrapposizione dei risultati è incoraggiante e rappresenta un primo passo verso l'identificazione di loci di suscettibilità all'autismo. Tuttavia, l'interpretazione dei dati di *linkage* provenienti da diversi *genome scans* comporta alcune difficoltà insite in questo tipo di approccio applicato alle malattie complesse. Infatti, è tuttora poco chiaro quale valore di *linkage* sia da considerarsi significativo in uno studio esteso all'intero genoma, in quanto vengono effettuati molteplici test statistici e non si conosce il numero dei reali loci di suscettibilità.

Come già detto, pur utilizzando strumenti diagnostici internazionalmente standardizzati, la variabilità che si osserva tra i risultati ottenuti dai diversi studi indipendenti può dipendere da diversi fattori quali:

- criteri di inclusione adottati
- differenze nei metodi analitici
- differenze nella scelta dei marcatori tipizzati
- grandezza del campione

Per tutti questi motivi, anche condurre una meta-analisi dei dati, messi attualmente a disposizione dai vari gruppi di ricerca, risulta particolarmente complicato [Lamb et al., 2002; Klauck, 2006; Trikalinos et al, 2006].

Dato che l'autismo segue un modello di ereditarietà complesso in cui sono coinvolti molti loci di suscettibilità, ciascuno con un effetto moderato, la non replicazione di un risultato in uno studio indipendente non esclude completamente la possibilità che questo sia un risultato reale, ma suggerisce semplicemente la presenza di eterogeneità genica e l'esistenza di più geni che interagiscono in maniera epistatica.

Per limitare il problema rappresentato dall'eterogeneità genetica sono stati condotti studi di *linkage* o utilizzando criteri di inclusione più stringenti o suddividendo le varie famiglie in sottogruppi che presentano tratti caratteristici del fenotipo autistico (endofenocopie) quali problemi nel linguaggio, difficoltà nella lettura, comportamenti stereotipati e ripetitivi, sesso. Per esempio, stratificando le famiglie sulla base di tratti relativi al linguaggio si è osservato un aumento del potere di *linkage* in corrispondenza del cromosoma 7q [Folstein e Mankoski. 2000; Lai et al. 2001; Bradford et al. 2001] e del cromosoma **2q** [Buxbaum et al., 2001; Shao et al., 2002], mentre utilizzando come endofenocopie famiglie con individui con comportamenti

ossessivi-compulsivi è stato individuato un ulteriore locus di suscettibilità sul cromosoma 1q [Buxbaum et al., 2004; Silverman et al., 2002]. Inoltre, l'osservazione di una prevalenza maggiore di casi di autismo in individui di sesso maschile rispetto ad individui di sesso femminile ha suggerito il probabile coinvolgimento di fattori sesso-specifici e quindi la possibilità di stratificare le famiglie in base al sesso del probando [Cantor et al., 2005; Stone et al., 2004; IMGSAC, 2001]. Anche in questo caso, la non riproducibilità dei risultati porta alla conclusione di uno scarso coinvolgimento del cromosoma X nella maggioranza dei casi di autismo.

Benché la stratificazione delle famiglie può essere un buon sistema per aumentare l'omogeneità genetica e, di conseguenza, la probabilità di identificare loci di suscettibilità, il problema fondamentale negli studi condotti fino ad ora rimane probabilmente la dimensione del campione analizzato che non è sufficiente grande per identificare varianti con effetto moderato e per discriminare i veri dai falsi positivi [Bacchelli et al., 2006].

Per ovviare questo problema è stato creato un Consorzio Internazionale, *l'Autism Genome Consortium* (AGP), di cui il nostro gruppo di ricerca fa parte, che coinvolge 50 differenti gruppi di ricerca in tutto il mondo. Il consorzio ha raccolto un totale di 1168 famiglie *multiplex* provenienti da Stati Uniti, Canada ed Europa, creando un campione di dimensione tre volte superiore a quello utilizzato da qualsiasi altro studio genetico condotto fino ad oggi sull'autismo. È stata condotta un'analisi di *linkage* estesa all'intero genoma analizzando più di 10.000 SNPs e utilizzando la piattaforma Affimetrix. Per ridurre l'eterogeneità presente nel campione, le famiglie sono state suddivise sulla base di tre diversi livelli diagnostici: famiglie con un fenotipo autistico stringente, famiglie con un fenotipo più lieve, tutte le famiglie. Il valore di *linkage* statisticamente più significativo è stato trovato in corrispondenza della regione 11p12-p13 se si considerano tutte le famiglie, e in corrispondenza delle regioni 2q e 7q se si considerano solo le famiglie di origine europea [AGP, 2007],

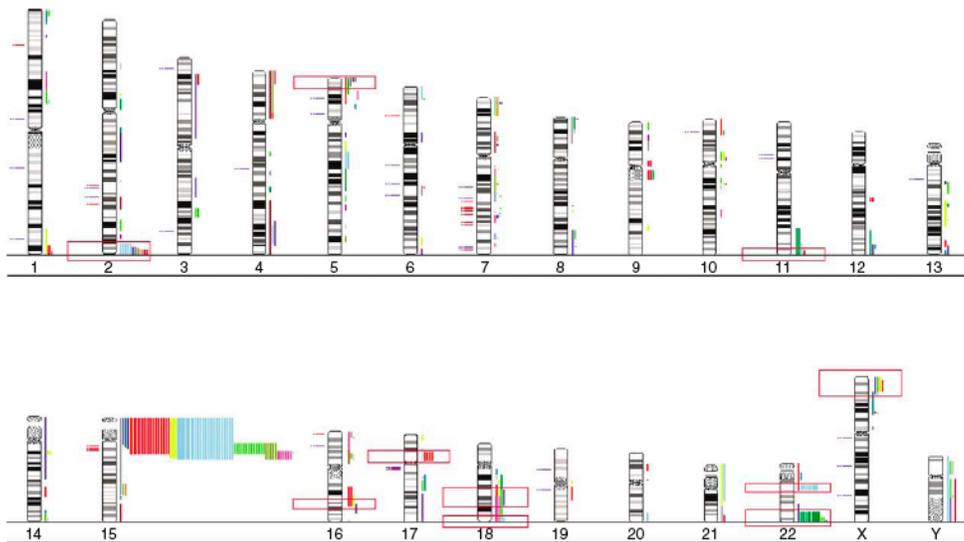
Tenendo conto della generale difficoltà nella replicazione dei risultati di *linkage* nello studio delle malattie complesse, è quindi molto interessante ed incoraggiante notare come le regioni sul cromosoma 7q e cromosoma 2q, inizialmente identificate dall'IMGSAC, siano state successivamente confermate da ulteriori studi indipendenti.

### 2.6b. Anomalie cromosomiche associate all'autismo

L'analisi di anomalie cromosomiche (delezioni, inserzioni, traslocazioni o inversioni) rappresenta un approccio alternativo e complementare che si è dimostrato molto utile nel mappaggio e nell'identificazione di geni malattia in quanto può aiutare o ad individuare direttamente il gene malattia, qualora il punto di rottura cada in corrispondenza del gene stesso, o a definire intervalli cromosomici critici per la localizzazione di un locus di suscettibilità.

Numerosi studi condotti su soggetti autistici hanno indicato che una percentuale del 3-5% di casi è associata ad un ampio spettro di anomalie cromosomiche [Freitag, 2006; Folstein and Rosen-Sheidley, 2001; Reddy, 2005; Chakrabarti&Forbonne, 2001; Wassink et al., 2001; Vorstman et al., 2006] rafforzando in questo modo l'idea che l'influenza genetica sia importante nello sviluppo dell'autismo.

Come si può osservare dalla figura sotto (Figura 16), anomalie cromosomiche associate al disturbo autistico, o a tratti riconducibili al fenotipo autistico, sono state osservate nella maggior parte dei cromosomi umani [Vorstman et al., 2006].



**Figura 16:** Anomali cromosomiche: panoramica di tutte le regioni citogenetiche di interesse (*Cytogenetic Region of Interest*, CROI) associate al fenotipo autistico sparse lungo tutto il genoma. A sinistra di ogni cromosoma la doppia linea blu indica le regioni dove è stato osservato linkage (LOD>2.0), mentre la linea rossa singola indica le regioni in cui è stata trovata associazione ( $P<0.05$ ). Sul lato destro dei cromosomi le CROI sono rappresentate da barre colorate. Barre con lo stesso colore affiancate indicano la stessa CROI individuata in più di un case-report, mentre le barre più spesse indicano la stessa CROI individuata in più di due casi nello stesso case-report. I quadrati rossi indicano nuove potenziali regioni riportate in più di quattro case-report in corrispondenza dello stesso locus che non era stato individuato da una precedente analisi di linkage o associazione.

Le anomalie cromosomiche determinano cambiamenti genetici funzionali che si possono verificare in vario modo:

- “*effetto dosaggio*”: come conseguenza della variazione del numero di copie di un gene in seguito, per esempio a delezioni o duplicazioni
- “*effetto di posizione*”: come conseguenza di un riarrangiamento cromosomico uno o più geni possono risultare separati da sequenze regolatrici
- i punti di rottura direttamente associati ad un riarrangiamento cromosomico possono distruggere il gene
- delezioni in uno dei due alleli possono “smascherare” una mutazione puntiforme localizzata nella regione corrispondente sul cromosoma omologo non deletato
- [Vorstman et al., 2006]

La sovrapposizione di regioni che presentano riarrangiamenti cromosomici con regioni individuate dai vari studi di *linkage* da una parte rappresenta una conferma di *linkage*, e dall'altra permette di localizzare in maniera più precisa la regione cromosomica di interesse.

**cromosoma 15:** l'anomalia cromosomica più frequentemente associata con una frequenza dell'1-3% ad alcune forme del disturbo autistico è rappresentata da:

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

Questa regione è tra le più complesse in termini di instabilità genomica e di *imprinting* [Reik et al., 1998]. Mutazioni che coinvolgono geni soggetti ad *imprinting* in questa regione danno origine alle sindromi di Prader-Willi (PWS, MIM 176260) e di Angelman (AS, MIM 105830) che presentano, in un sottoinsieme di casi, tratti fenotipici simili all'autismo. Tutte le duplicazioni osservate associate all'autismo in questa regione sono di origine materna, quindi è possibile che esista un locus di suscettibilità per l'autismo soggetto ad *imprinting*.

**cromosoma 22:** l'anomalia del cromosoma 22 più frequentemente associata ad alcune forme di autismo è rappresentata da delezioni delle regioni q11.1 e q13. Delezioni a carico della regione 22q13 sono associate ad una sindrome che, dal punto di vista clinico, si manifesta con comportamenti pervasivi e problemi nello sviluppo del linguaggio; mentre la delezione a carico

della regione 22q11.1 è associata alla sindrome velo-cardio-facciale con caratteristiche fisiche simili a quelle riscontrate in individui autistici.

La mancanza di una forte evidenza di *linkage* per queste due regioni potrebbe essere spiegata dalla presenza di eterogeneità genetica e/o dal fatto che questi riarrangiamenti cromosomici siano associati ad uno specifico sottogruppo fenotipico, ed è quindi probabile che geni presenti all'interno di tali regioni sui cromosomi 15 e 22 non conferiscano suscettibilità all'autismo in individui citogeneticamente normali [Vorstman et al., 2006].

**cromosoma X:** le anomalie cromosomiche a livello dei cromosomi sessuali, e in particolare del cromosoma X, sono di particolare interesse in riferimento alla maggiore incidenza del disturbo autistico nei soggetti di sesso maschile e nelle femmine affette dalla sindrome di Turner (45, X) [Skuse et al., 1997].

**cromosoma 7:** numerose anomalie cromosomiche sono anche state riscontrate a livello del cromosoma 7, alcune delle quali presentano il punto di rottura in corrispondenza delle regioni 7q21-q22 e 7q31-q32 e sono di particolare interesse in quanto rientrano nell'intervallo individuato dagli studi di *linkage* e di associazione [Yu et al., 2002; Vincent et al., 2000; Waburton et al., 2000; Ashley-Koch et al., 1999]. Data l'importanza di questo cromosoma, è stato ideato un *database* delle anomalie del cromosoma 7 associate all'autismo, mantenuto continuamente aggiornato dal TCAG (The Centre for Applied Genomics).

**cromosoma 2:** sono state osservate diverse delezioni della porzione distale del braccio lungo del cromosoma 2 (2q37) in bambini autistici o che presentavano tratti autistici associati ad altre anomalie fisiche [Ghaziuddin et al., 1999; Smith et al., 2001]. Nella maggior parte dei casi pubblicati l'ampiezza della delezione non è stata determinata a livello molecolare, anche se è probabile che delezioni più piccole siano associate solo al comportamento autistico e al ritardo nel linguaggio [Casas et al., 2002; Smith et al., 2001], mentre delezioni più estese sono associate ad spettro di deficits più ampio che comprende ritardo mentale e anomalie fisiche.

Oltre ai casi sopra elencati, risultati interessanti sono stati trovati anche in corrispondenza di altre regioni cromosomiche: 5p14-p15, 11q25, 16q22.3, 17p11.2, 18q21.1-q23.

Quale sia il significato funzionale di queste anomalie cromosomiche in relazione ai disturbi dello spettro autistico rimane ancora da chiarire. La maggior parte di queste anomalie si verifica *de novo* e viene riscontrata in casi sporadici ed in genere in soggetti con diagnosi di autismo "atipico" e con problemi nell'apprendimento. Inoltre, la loro marcata variabilità non ha finora permesso di identificare una specifica regione genomica in cui ricercare i geni di suscettibilità.

Tuttavia, la raccolta e caratterizzazione di ulteriori anomalie citogenetiche associate all'autismo, specialmente se si considerano quelle localizzate nelle regioni già evidenziate dai diversi studi di *linkage*, potrebbe essere di grande aiuto nell'identificazione di geni che predispongono alla malattia.

### **2.7. Geni candidati per l'autismo**

Gli studi mediante *linkage*, intrapresi negli ultimi dieci anni, hanno identificato diverse regioni cromosomiche di interesse per la suscettibilità all'autismo. Tuttavia, questo tipo di analisi non ha circoscritto abbastanza gli intervalli cromosomici e per questo motivo molti gruppi di ricerca hanno iniziato ad effettuare *screening* di geni candidati e studi di associazione nelle regioni sub-cromosomiche definite dalle analisi di *linkage*.

La disponibilità della sequenza completa del genoma umano ed il suo costante aggiornamento ha facilitato l'efficienza nello *screening* di geni candidati che mappano in determinate regioni cromosomiche. Un'indagine della letteratura mostra l'enorme numero (circa un centinaio) di geni candidati analizzati e la quantità di lavori sull'argomento pubblicati negli ultimi dieci anni.

Di solito, i geni candidati sono essere scelti in base al loro profilo di espressione e al possibile ruolo nello sviluppo e funzionalità del SNC (candidati funzionali), alla loro localizzazione all'interno di intervalli individuati tramite un'analisi di *linkage* o alla presenza di riarrangiamenti cromosomici associati alla patologia (candidati posizionali).

Un'analisi accurata dei geni candidati viene condotta attraverso il sequenziamento di tutto il gene in individui affetti e in individui sani. Questo metodo viene adottato nel caso di patologie con eterogeneità allelica, cioè quando alleli rari che causano la patologia sono presenti all'interno dello stesso gene. Tuttavia, a causa del problema rappresentato dall'elevato costo, il sequenziamento è solitamente ristretto alle regioni codificanti ed importanti dal punto di vista funzionale e/o regolatorio, in un numero limitato di individui. Una strategia per aumentare la probabilità di identificare varianti del DNA che possano contribuire all'eziologia del disturbo autistico è quella di analizzare individui autistici non imparentati, selezionati tra le famiglie che contribuiscono maggiormente al risultato di *linkage* in una specifica regione cromosomica.

Di solito, gli studi di associazione si basano su un approccio "diretto" in base al quale una probabile variante casuale (come una variante missenso) viene analizzata per individuare la correlazione con la malattia.

Nel caso di malattie complesse non è immediato riuscire ad identificare varianti funzionali in geni candidati anche perchè queste varianti possono essere localizzate all'interno di regioni introniche o lontane da elementi regolatori. Per questo motivo, la maggior parte degli studi di associazione partono da un approccio "indiretto" basato sul *linkage disequilibrium* (LD), dove per LD si intende un'associazione non-casuale di alleli vicini. Quindi, un'associazione di tipo indiretto viene individuata solo se l'allele di un marcatore è in LD con la variante di suscettibilità, anche se il marcatore stesso non ha alcun effetto funzionale. Il vantaggio del mappaggio tramite LD (*LD mapping*) è che permette di rilevare effetti genetici più deboli rispetto al *linkage*, mentre il principale problema è che richiede l'utilizzo di un numero molto elevato di marcatori.

Negli ultimi dieci anni sono stati analizzati più di 100 geni candidati (posizionali e funzionali) alla ricerca di varianti implicate nell'eziologia del disturbo autistico. Tuttavia, per nessuno di questi geni è stato dimostrato un chiaro coinvolgimento nella patogenesi della malattia. Alcuni esempi sono descritti a seguito:

**RELN:** La Relina (*RELN*) (Omim # 600514) è un gene che mappa nella regione 7q22, una regione di *linkage* identificata da diversi *genome scan*. Questo gene codifica per una proteina coinvolta nella migrazione neuronale nel corso dello sviluppo cerebrale. Inoltre, si è visto che il topo *reeler*, un mutante naturale portatore di una larga delezione di questo gene, mostra alterazioni nella posizione dei neuroni nella corteccia celebrale e nell'ippocampo, le stesse regioni cerebrali in cui sono state riscontrate anomalie in individui autistici [D'Arcangelo et al., 1995; Bailey et al., 1998]. In uno studio iniziale [Persico et al., 2001] era stata riportata l'esistenza di una trasmissione preferenziale dell'allele più lungo di una sequenza ripetuta (tripletta) presente nella regione non tradotta al 5' del gene, ma la replicazione di tale risultato in campioni indipendenti ha fornito risultati contrastanti o, comunque, nessuna evidente associazione tra *RELN* e fenotipo autistico. [Krebs et al., 2002; Zhang et al., 2002; Devlin et al., 2004; Li et al., 2004; Skaaret al., 2005]. Successivamente, uno *screening* di mutazioni di questo gene [Bonora et al., 2003] ha portato all'individuazione di 4 varianti missenso non osservate in un campione di controllo, ma queste varianti sono troppo rare per poter essere considerate responsabili del valore di *linkage* sul cromosoma 7q22.

**SLC6A4:** Questo gene codifica per il trasportatore della serotonina (*SLC6A4/5-HTT*) (Omim # 182138) ed ha suscitato notevole interesse in seguito alle osservazioni sugli elevati livelli di serotonina nel sangue e nelle urine di individui autistici e alla risposta positiva a sintomi caratteristici del fenotipo autistico (comportamenti aggressivi e ripetitivi) dopo trattamento con

inibitori del trasportatore stesso [Anderson et al.,1987; Hollander et al., 2005]. Questo gene mappa nella regione cromosomica 17q11-12, precedentemente identificata tra tre distinti studi di *linkage* [Risch et al, 1999; Yonan et al., 2003; Cantor et al., 2005; McCauley et al, 2005]. Tuttavia, studi di associazione di un polimorfismo, in particolare una delezione, presente nella regione promotrice di questo gene hanno mostrato risultati non significativi [Maestriniet al., 1999; Zhong et al., 1999; Betancuret al., 2002; Persico et al., 2002;Wu et al.,2005] o comunque contrastanti [Cook et al., 1997; Kim et al., 2002; Conroy et al., 2004; Devlin et al., 2005;McCauley et al., 2005; Klauck et al., 1997; Tordjman et al.,2001; Yirmiya et al., 2001], evidenziando la presenza di eterogeneità tra gli studi condotti dai diversi gruppi di ricerca.

**MECP2:** Il gene *MECP2 (methyl-CpG-binding protein-2)* (Omim # 300496) mappa nella regione cromosomica Xq28 e codifica per una proteina responsabile di un altro disordine neurologico, la Sindrome di Rett [Amir et a., 1999] classificato tra i disordini pervasivi dello sviluppo. *Screening* di mutazioni di questo gene in individui autistici non ha identificato alcuna variante funzionale coinvolta nell'eziologia del disturbo autistico [Vourc'h et al., 2001].

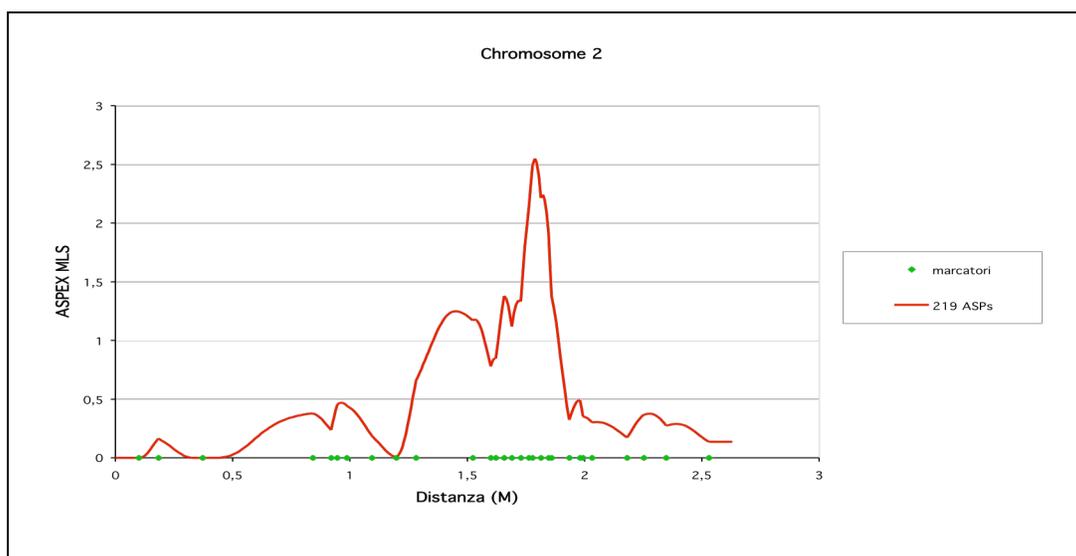
**MET :** Il gene *MET* (Omim # 164860) mappa nella regione 7q31 e codifica per un recettore tirosin chinasi. Nonostante il suo ruolo nella formazione di metastasi in diversi tipi di cancro, *MET* è stato considerato un buon candidato per l'autismo in quanto contribuisce allo sviluppo della corteccia cerebrale e del cervelletto, regioni altamente compromesse negli individui autistici. Un recentissimo studio condotto su 204 famiglie *multiplex*, e successivamente replicato su altre 539 famiglie, ha infatti rilevato un'associazione significativa tra un polimorfismo funzionale nella regione promotrice del gene *MET* ed il fenotipo autistico [Campbell et al, 2006].

**FOXP2:** Come già detto, i valori più alti di *linkage* sul cromosoma 7 sono stati trovati in quelle famiglie con individui con problemi nel linguaggio. *FOXP2 (Forkhead Box P2)* (Omim # 605317) è un gene che mappa sul cromosoma 7q ed è stato considerato un buon candidato in quanto mutazioni di questo gene erano state osservate in una famiglia con individui con problemi nel linguaggio [Fisher et al. 1998; Lai et al., 2000]. Tuttavia, con l'eccezione di due studi dai risultati discutibili [Gong et al. 2004; Li et al., 2005] non è stata mai trovata associazione tra il fenotipo autistico e le varianti individuate in *FOXP2* [Gauthier et al., 2003; Marui et al., 2005; Wassink et al., 2002; Newbury et al., 2002].

Questi sono solo alcuni esempi che mettono in evidenza la difficile riproducibilità dei risultati ottenuti dai diversi studi condotti sui geni candidati e le scarse evidenze sul chiaro coinvolgimento di nessuno di questi geni nell'eziologia del disturbo autistico.

### 3. Il cromosoma 2: screening di geni candidati per l'autismo

I risultati degli *screening* genomici condotti dal consorzio IMGSAC [IMGSAC, 1998; IMGSAC, 2001; Lamb et al., 2005] hanno individuato un locus di suscettibilità per l'autismo nella regione 2q21-33 (Cap.2, § 2.6a). In Figura 16 è riportato il profilo di MLS sul cromosoma 2q relativo ai 219 ASPs analizzati nell'ultimo *genoma scan* condotto dall'IMGSAC [Lamb et al., 2005].



**Figura 16:** Risultati di *linkage* sul cromosoma 2: In figura è riportato il profilo di MLS relativo ai 219 ASPs [Lamb et al., 2005].

I risultati positivi riportati per la stessa regione cromosomica da altri tre studi indipendenti di *genome scan* [Philippe et al., 1999; Buxbaum et al., 2001; Shao et al., 2002] sono a supporto dell'evidenza di *linkage* in questa regione.

Visti i promettenti presupposti, il gruppo di ricerca ha intrapreso la caratterizzazione del locus di suscettibilità per l'autismo sul cromosoma 2q. Tuttavia, identificare i geni legati all'autismo richiede un lavoro impegnativo perché la regione di *linkage* sul cromosoma 2q è molto ampia (~ 40 cM) e contiene molti geni che potrebbero essere responsabili del valore di *linkage* trovato. Inoltre, un'altra difficoltà che si incontra negli studi di *linkage* per malattie multifattoriali è che la

localizzazione dei risultati per loci ad effetto moderato non è precisa e può essere distante diversi centimorgan dal locus reale.

Per questi motivi, per caratterizzare ulteriormente la regione del cromosoma 2q21-q33, abbiamo deciso di seguire due approcci:

1. *Screening* sistematico di geni che mappano nella regione di maggior *linkage* (tra i marcatori D2S2370 e D2S364) e che rappresentano dei buoni candidati funzionali, con particolare attenzione ai geni localizzati all'interno di regioni che presentano anomalie cromosomiche.
2. *Genotyping* di SNPs ad alta densità all'interno della regione cromosomica 2q, allo scopo di effettuare studi di *LD mapping* e identificare, per via indiretta, le varianti di rischio per l'autismo (Cap.4 §4.1).

### **3.1. Geni di suscettibilità sul cromosoma 2**

All'inizio del mio lavoro di dottorato il gruppo di ricerca del quale faccio parte aveva già effettuato lo screening di 12 geni candidati funzionali: T-box Brain 1 (*TBR1*), Glutamic Acid Decarboxylase 1 (*GDAI*), Distal-less 1 (*DLX1*), Distal-less 2 (*DLX2*), cAMP Guanine nucleotide Exchange Factor II (*cAMP-GEFII*), Chimerin 1 (*CHN1*), Activating Transcription Factor 2 (*ATF2*), (Homeo box D1) HOXD1, Neurogenic Differentiation 1 (*NEUROD1*), Frizzled-related protein 1 (*FRZB1*), Engulfment Adaptor Protein; GULP (*CED-6/GULP*), Hypothetical protein, selected from TESS library (*KIAA 1604*)

Nei geni *cAMP-GEFII* e *TBR1* sono state identificate nuove varianti codificanti non riscontrate in un vasto gruppo di controlli. Tuttavia, la bassa frequenza di tali varianti non offre una chiara spiegazione del ruolo di questi geni nella suscettibilità all'autismo [Bacchelli, Blasi et al., 2003].

Una parte del lavoro svolto durante il dottorato ha, quindi, riguardato l'analisi di otto geni candidati scelti, in corrispondenza della regione con il valore più alto di *linkage* sul cromosoma 2, in base al loro ruolo nello sviluppo e funzionalità del sistema nervoso centrale e/o alla loro localizzazione in corrispondenza di anomalie cromosomiche rilevate in individui affetti. (Tabella 3)

GENE		FUNZIONE
<b>REPRIMO</b>	TP53 dependant G2 arrest mediator	regolazione del ciclo cellulare
<b>KCNJ3</b>	potassium inwardly-rectifying channel J3	eccitazione delle cellule nervose
<b>NR4A2</b>	nuclear receptor subfamily 4, group A, member 2	regolazione del rilascio dell'ormone corticotropico
<b>UPP2</b>	Uridine phosphorylase 2	biosintesi dei nucleotidi
<b>UBE2E3</b>	Ubiquitin-conjugating enzyme E2E	ubiquitinazione
<b>NCKAP1</b>	NCK Associated Protein	apoptosi
<b>INPP-1</b>	Inositol Polyphosphate 1- Phosphatase	metabolismo dei fosfoinositidi
<b>NAB1</b>	NGFI-A binding protein 1	regolazione trascrizionale

**Tabella 3:** gli 8 geni analizzati che mappano nella regione di *linkage* 2q21-q33

Gli 8 geni analizzati sono: NGFI-A binding protein (*NAB1*), ubiquitin-conjugating enzyme E2E3 (*UBE2E3*), enzyme inositol polyphosphate-1-phosphatase (*INPP1*), NCK-associated protein 1 (*NCKAP1*), Uridine phosphorylase 2 (*UPP2*), potassium inwardly-rectifying channel J3 (*KCNJ3*), Nuclear Receptor Subfamily 4, Group A, Member 2 (*NR4A2*), e TP53 dependant G2 arrest mediator (*REPRIMO*).

*NAB1* fa parte di una famiglia altamente conservata di proteine che regolano negativamente l'attività trascrizionale di *NGFI-A*, una famiglia di fattori di trascrizione del tipo “*zinc finger*” che, in risposta a stimoli extracellulari, svolgono un ruolo importante nella regolazione della crescita cellulare, nel differenziamento neuronale, nell'attività sinaptica, nel potenziamento a lungo termine e nell'apoptosi [Swirnoff et al., 1998]. Recenti studi hanno proposto, per *NAB1*, un ruolo nella mielinizzazione nel sistema nervoso periferico e un coinvolgimento nello sviluppo del cervello [Mechta-Grigoriou et al., 2000].

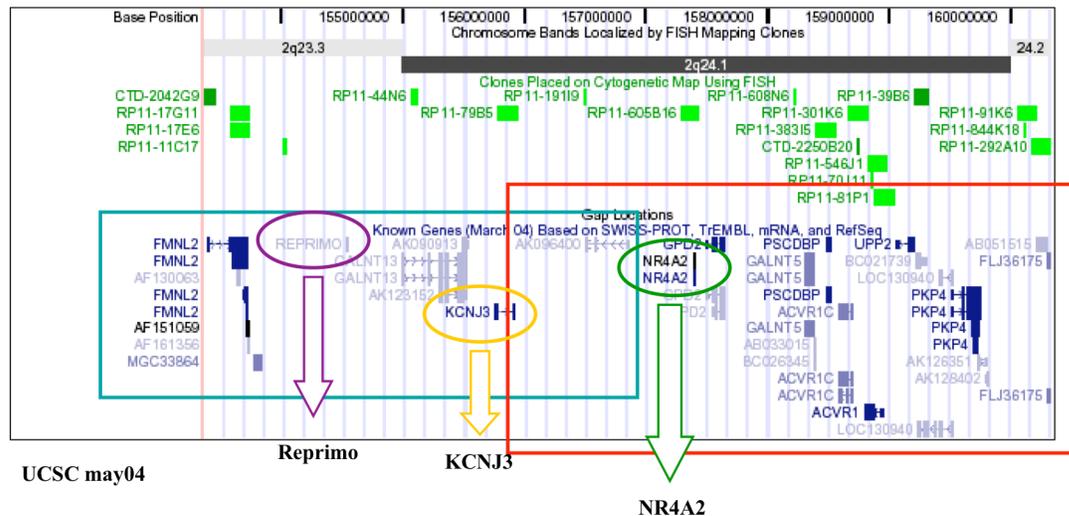
*UBE2E3* è uno degli enzimi coinvolti nell'ubiquitinazione, un importante meccanismo cellulare di modificazione post traduzionale implicato in varie funzioni cellulari quali l'apoptosi, la progressione del ciclo cellulare, la trascrizione e l'endocitosi [Hershko et al., 1998]. Questo meccanismo prevede l'intervento di tre classi di enzimi: E1 (*Ub-activating enzyme*), E2 (*Ub-conjugating enzyme*), E3 (*Ub-ligase*). Il gene *UBE2E3* fa parte della classe E2 e la sua funzione è quella di catalizzare il legame covalente dell'ubiquitina ad altre proteine [Matuschewski et al., 1996]. Poichè studi recenti hanno dimostrato che mutazioni nel gene *UBE3a*, appartenente alla

classe E3 delle ligasi, sono implicate nella Sindrome di Angelman, un disordine neurologico con caratteristiche fenotipiche riconducibili a quelle dello spettro autistico [Matsuura et al., 1997; Malzac et al., 1998; Fang et al., 1999; Rapakko et al., 2004], si è pensato che il gene *UBE2E3*, appartenente allo stesso gruppo di enzimi coinvolti nell'ubiquitinazione, potesse essere un buon candidato funzionale da studiare nel nostro campione di pazienti autistici.

**INPPI** codifica per l'*inositolo polifosfato fosfatasi 1*, un enzima coinvolto nei processi di trasmissione del segnale del fosfatidilinositolo; questo enzima è importante per numerosi processi connessi alla funzionalità del sistema nervoso centrale quali: sopravvivenza e differenziamento neuronale, trasduzione dei segnali di fattori di crescita, neurotrasmettitori e recettori accoppiati a proteine G. Questo enzima è espresso, con livelli rilevanti di mRNA, nei testicoli, nelle ghiandole surrenali, nell'ipofisi e in diverse aree cerebrali [York et al., 1993].

**NCKAPI** codifica per una proteina espressa in tutti i tessuti umani, con livelli di espressione più alti nel cuore, nei tessuti muscolari e nel cervello (ippocampo e amigdala), in particolare nel cervello adulto rispetto al cervello fetale [Suzuki e coll., 2000]. Alcuni ricercatori hanno anche dimostrato che la soppressione di *NCKAPI* induce l'apoptosi delle cellule neuronali, probabilmente a causa della perdita dei continui segnali di sopravvivenza dipendenti da *NCKAPI* [Suzuki e coll., 2000].

I geni *KCNJ3*, *NR4A2* e *REPRIMO* sono stati scelti come candidati funzionali preferenziali in quanto, nel corso del periodo del mio dottorato, sono giunti alla nostra attenzione due pazienti autistici, non imparentati tra loro, entrambi con una delezione in corrispondenza della regione cromosomica 2q24.1. Il mappaggio più fine di questa regione, effettuato tramite FISH, è stato condotto ad Oxford ed ha messo in evidenza che queste due delezioni presentano una regione di sovrapposizione di circa 1 Mb comprendente il solo gene, *KCNJ3*. Come si può vedere dalla figura sotto (Figura 17), nel paziente con la delezione più piccola, di ~3,4Mb (153,084,416 - 156,481,525), *KCNJ3* è completamente deleta, mentre nell'altro paziente con la delezione maggiore, di ~4,5Mb (155,495,564 - 159,951,720), *KCNJ3* è deleta solo parzialmente. Benchè la delezione non corrisponde alla regione subtelomerica, 2q37 comunemente deleta negli individui autistici [Lukusa et al., 2004], è sembrato molto interessante studiarla in quanto, non solo corrisponde ad una regione in cui diversi gruppi di ricerca hanno trovato un significativo valore di *linkage* [Philippe et al., 1999; IMGSA, 2001; Buxbaum et al., 2001; Shao et al., 2001], ma anche perchè sembra essere coinvolta nello sviluppo del linguaggio [Barlett et al., 2004] e dell'IQ [Posthuma et al., 2005].



**Figura 17:** Regione di sovrapposizione delle due delezioni. Il riquadro rosso indica la delezione più grande (~4,5Mb) che distrugge il gene *KCNJ3* solo parzialmente, mentre il riquadro blu indica la delezione più piccola (~3,4Mb) che include tutto il gene *KCNJ3*.

*KCNJ3* codifica per una proteina integrale di membrana, un canale del potassio del tipo “*inward-rectifier*”, abbondantemente espressa nel cervello [Stoffel et al., 1994; Schoots et al., 1997]. Questi canali controllano l’eccitabilità delle cellule neuronali favorendo l’entrata del potassio all’interno della cellula attraverso un meccanismo non dipendente dal voltaggio. Inoltre, sono attivati direttamente dal legame con proteine G e svolgono un ruolo nel mantenimento del potenziale di membrana a riposo, nella generazione di potenziali d’azione prolungati e nella modulazione dell’eccitabilità cellulare. Recentissimi studi hanno analizzato il loro ruolo in forme idiopatiche di epilessia [Chioza et al., 2002].

*NR4A2* codifica per un membro di una superfamiglia di recettori ormonali (steroidi-tiroidei) nucleari essenziali nel promuovere il differenziamento di cellule nervose che producono dopamina, neurotrasmettitore necessario per il controllo del movimento e degli stati emozionali. Mutazioni in questo gene sono state associate a disordini correlati a disfunzioni del sistema dopaminergico quali la schizofrenia [Chen et al., 2001; Buervenich et al., 2000], la sindrome maniaco-depressiva e forme familiari di Parkinson [Le et al., 2003; Xu et al., Hering et al., 2004; Buervenich et al., 2000].

**REPRIMO** (*RPRM*) è un gene che codifica per una proteina coinvolta nell'arresto del ciclo cellulare mediato da p53 in corrispondenza della fase G2 [Ohki et al., 2000].

Il gene *UPP2* è stato scelto come ottimo candidato in quanto, in seguito all'esperimento di *genotyping* di SNPs ad alta densità all'interno della regione cromosomica 2q (Cap.7, §7.3) è stata individuata una CNV, in particolare una delezione, che comprende l'esone 6 e l'esone 7 di questo gene. Si è deciso quindi di analizzare il gene *UPP2* allo scopo di evidenziare altre eventuali varianti presenti sia negli individui autistici portatori della delezione, sia in altri soggetti autistici, scelti fra le famiglie che contribuiscono maggiormente al picco di *linkage* in corrispondenza di questo gene sul cromosoma 2.

**UPP2** codifica per una proteina simile all'uridina fosforilasi (*UPase*) e alla timidina fosforilasi (TPase), denominata *uridina fosforilasi 2* (UPase-2) che catalizza, in maniera reversibile, la fosforolisi dell'uridina in uracile [Cao et al., 2005]. L'uridina fosforilasi è un enzima fondamentale per la biosintesi dei nucleotidi e per la regolazione omeostatica dell'uridina a livello plasmatico [Leer et al. 1977; Vita et al.1986]. Inoltre, l'attività catalitica delle pirimidine fosforilasi interviene nel meccanismo di recupero delle basi pirimidiniche per la biosintesi dei nucleotidi con un meccanismo energeticamente meno costoso rispetto alla sintesi *de novo* [Pizzorno et al. 2002]. Di particolare interesse, per il nostro gruppo di ricerca, è il coinvolgimento dell'uridina nella regolazione del sonno e il suo utilizzo per il trattamento dell'epilessia. Infatti, alcuni casi di autismo sono associati a disturbi del sonno e a fenomeni più o meno gravi di epilessia. In particolare, in alcuni casi di autismo associato ad epilessia, la somministrazione orale di uridina determina miglioramenti nella capacità linguistica, nel comportamento ed una diminuzione della frequenza di crisi epilettiche [Page et al. 1997, Cao et al., 2005].

## **4. Il cromosoma 2: studio di associazione su larga scala e individuazione di CNVs**

### **4.1 Genotyping su larga scala di SNPs tramite il sistema Illumina**

Sulla base di tutti questi presupposti e sull'osservazione che lo *screening* di mutazione nei migliori geni candidati non ha portato all'identificazione di nessuna variante chiaramente implicata nell'eziologia dell'autismo, il gruppo di ricerca con cui ho svolto il lavoro di dottorato ha intrapreso un approccio più sistematico. Tale approccio consiste nel *genotyping* di SNPs ad alta densità all'interno della regione cromosomica 2q, allo scopo di effettuare studi di *LD mapping* e identificare per via indiretta le varianti di rischio per l'autismo. Questo approccio è stato reso possibile dalla recente disponibilità di risorse pubbliche quali il *International Haplotype Mapping Consortium* (HapMap), che sta caratterizzando i pattern più comuni di varianti di sequenza presenti nel genoma umano, e dalle nuove tecnologie che permettono una tipizzazione rapida ed efficiente di SNPs su larga scala.

Sfruttando la grande quantità di dati relativi all'analisi di SNPs è possibile mappare a livello genomico, su larga scala e a basso costo, la nuova fonte di variabilità rappresentata dalle CNVs. Dopo un'estesa analisi statistica per valutare la strategia migliore in termini di numero di campioni di DNA e di SNPs da analizzare, data l'ampiezza dei picchi di *linkage* (~ 40 Mb), per ottimizzare la selezione degli SNPs, è stata effettuata un'approfondita analisi bioinformatica utilizzando i dati disponibili in HapMap per valutare il profilo di LD e la struttura a blocchi di aplotipi nella regione di interesse

Abbiamo selezionato 1500 *haplotype tagging* SNPs (htSNPs) relativi ai blocchi in corrispondenza di tutti i geni noti e delle loro regioni regolatorie, riducendo così il carico di *genotyping* necessario, pur caratterizzando la maggior parte della variabilità genetica nell'80% di tutte le regioni geniche. Questi SNPs sono quindi stati tipizzati in un campione di 576 individui costituito da 126 famiglie con autismo (individui affetti con i loro genitori scelti fra le famiglie *multiplex* che contribuivano maggiormente al picco di *linkage* sul cromosoma 2q) e 192 individui di controllo. Il sistema utilizzato per la tipizzazione degli SNPs è stato la tecnologia Illumina (*GoldenGate® assay*).

Da questo esperimento sono stati ottenuti circa 850.000 genotipi, i quali sono stati utilizzati per effettuare un'analisi di associazione con il fenotipo autistico, utilizzando sia approcci di tipo caso-controllo, sia il test di disequilibrio della trasmissione (TDT). Ciò ha permesso di individuare alcuni aplotipi nell'ambito di due diversi geni, con una significativa evidenza di associazione con l'autismo. I risultati più significativi sono stati trovati in corrispondenza di due

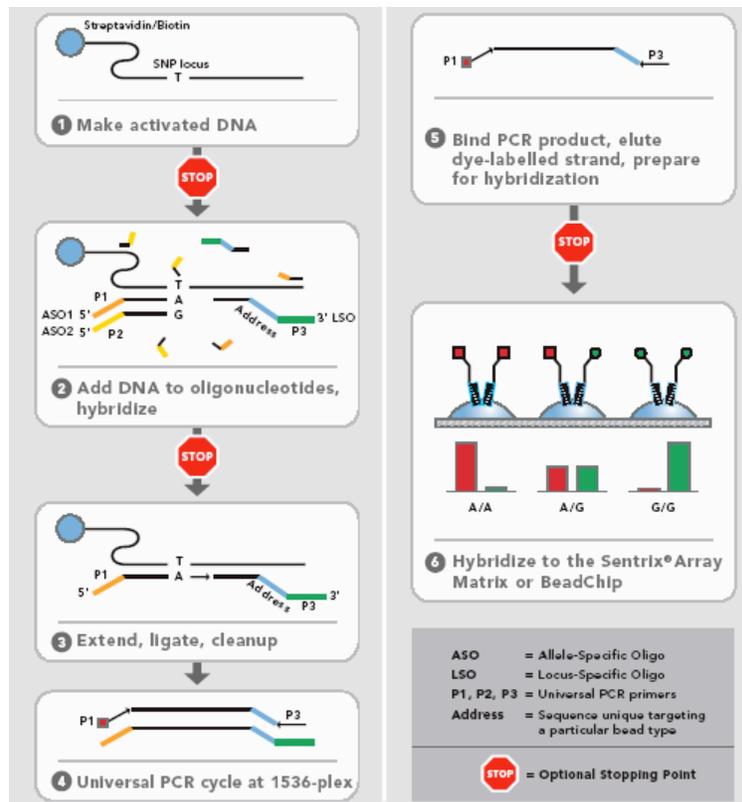
geni: *NOSTRIN* (*Nitric Oxide Synthase Trafficker*) e *ZNF533*. E' attualmente in corso una replicazione dell'esperimento in un campione di individui indipendente.

Tuttavia, nell'ambito del suddetto esperimento Illumina, l'obiettivo del mio lavoro ha riguardato un diverso aspetto dell'analisi dei genotipi prodotti: l'individuazione di *Copy-Number Variants* (CNVs) o varianti strutturali (cap7, §7.3). Cio' si basa sull'osservazione che la presenza di una delezione in un individuo fa si che il risultato di *genotyping* di uno SNP emizigotico (in quanto ricade all'interno della delezione) venga interpretato erroneamente come omozigotico. Sulla base di tale osservazione, se lo SNP in questione è polimorfico e i genitori sono informativi per tale SNP, una delezione trasmessa da un genitore al figlio si evidenzia in risultati di *genotyping* dello SNP che, apparentemente, violano le leggi di Mendel.

#### **4.2 Tecnologia Illumina: tecnologia array-based per il genotyping di SNPs su larga scala**

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

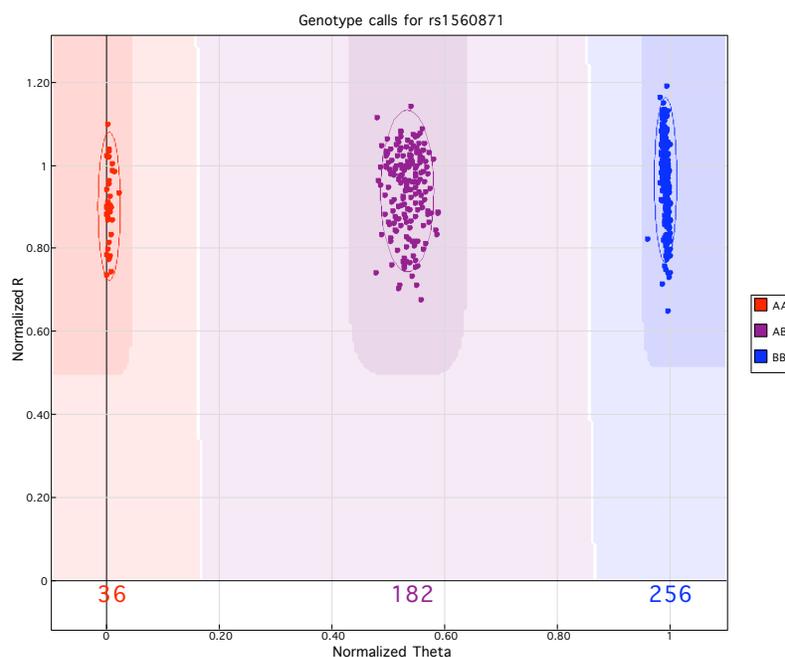
Il saggio Illumina (*GoldenGate<sup>TM</sup> genotyping assay*) (Figura 18) consente di discriminare le varianti di SNPs direttamente sul DNA genomico piuttosto che su prodotti di PCR, per questo motivo non possono essere introdotti errori causati dal processo di amplificazione. Il DNA genomico viene attivato coniugandolo ad un supporto solido (particelle sferiche paramagnetiche). Seguono un passaggio per l'ibridazione degli oligonucleotidi disegnati per il saggio di specifici SNPs di interesse e diversi lavaggi per eliminare gli oligonucleotidi che non si sono legati.



**Figura 18:** Schema del saggio Illumina (*GoldenGate<sup>TM</sup> genotyping assay*).

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

negli *arrays* tramite la sequenza indirizzo. Come precedentemente spiegato, ad ognuno dei 1500 SNPs è assegnata una sequenza indirizzo contenuta nel LSO complementare ad una sola sonda presente nell'*array*. I rilevatori di fluorescenza utilizzati da Illumina scansionano infine la piastra contenente 96 *arrays* contemporaneamente a due diverse lunghezze d'onda adatte alla Cy3 e alla Cy5. Nel caso di un campione omozigote per un dato SNP, il rilevatore osserva un solo tipo di fluorescenza (o quella dipendente dal primer P1 coniugato alla Cy3 o dal primer P2 coniugato alla Cy5), invece in presenza di un campione eterozigote verranno rilevate entrambi i tipi di fluorescenza. Quindi il rapporto dei segnali relativi alle due diverse fluorescenze identifica il genotipo come AA, AB, o BB (Figura 19).



**Figura 19:** Rappresentazione grafica dei risultati forniti dal sistema Illumina per il *genotyping* di uno SNP (nell'esempio *rs1560871*). I tre *cluster* sono formati dai segnali di fluorescenza ottenuti per individui omozigoti AA (rosso), eterozigoti AB (viola), omozigoti BB (blu).

## 5. Scopo della ricerca

L'autismo è un grave disturbo neuropsichiatrico che si manifesta durante l'infanzia e i cui sintomi persistono per tutta la vita dell'individuo. È una malattia eterogenea sia da un punto di vista clinico che eziologico. Gli studi epidemiologici compiuti negli ultimi vent'anni hanno indicato la presenza di una forte componente genetica alla base di questo grave disturbo. Il modello di ereditarietà è tuttavia complesso, ed è probabilmente influenzato dall'azione concertata di diversi geni.

Studi di *linkage* estesi all'intero genoma hanno evidenziato regioni, su diversi cromosomi, in cui potrebbero essere presenti geni di suscettibilità per il disturbo autistico. In particolare, in base ai risultati ottenuti dall'*International Molecular Genetic Study of Autism Consortium* (IMGSAC) e da altri gruppi di ricerca in studi di *linkage*, è stato identificato un potenziale locus di suscettibilità all'autismo nella regione cromosomica 2q21-q33. Lo scopo del presente lavoro è la caratterizzazione di questo locus attraverso due approcci complementari.

In primo luogo è stata intrapresa un'analisi dei migliori geni candidati nella regione di maggior *linkage* compresa tra i marcatori D2S2370 e D2S364. Sono stati presi in esame otto geni candidati che in base al loro ruolo nello sviluppo e funzionalità del sistema nervoso centrale e/o alla loro localizzazione in corrispondenza di anomalie cromosomiche rilevate in individui affetti, rappresentano ottimi candidati per il disturbo autistico. Tutte le regioni codificanti ed importanti da un punto di vista funzionale e/o regolatorio sono state analizzate in modo sistematico per la presenza di varianti di sequenza. Tutte le varianti con un possibile significato funzionale identificate in ogni gene sono state ulteriormente analizzate nell'intera collezione di famiglie *multiplex* dell'IMGSAC, paragonando la frequenza ottenuta negli individui affetti con quella ottenuta negli individui sani. Le altre varianti più informative identificate da questa analisi e altri polimorfismi del DNA a singola base (*Single Nucleotide Polymorphisms*, SNPs), scelti in banca dati, localizzati all'interno o in prossimità dei geni candidati sono stati utilizzati per effettuare studi di associazione allelica.

Negli ultimi anni diversi studi hanno evidenziato che un grande contributo alla variabilità genetica ed in particolare alla suscettibilità ad alcune malattie, è dato da polimorfismi che consistono nella variazione del numero di determinati segmenti genomici. Questi ultimi interessano intervalli compresi fra 1Kb e 1-2Mb (inserzioni, delezioni, duplicazioni) e sono denominati *Copy-Number Variants* (CNVs). Recenti studi hanno verificato la possibilità di

rilevare, in modo efficiente, la presenza di CNVs usufruendo della grandissima quantità di dati, attualmente già disponibile, relativa all'analisi su larga scala di SNPs.

Il secondo scopo di questo lavoro consiste appunto nell'individuazione di CNVs a partire da dati di *genotyping* di SNPs ottenuti nel nostro laboratorio tramite la tecnologia Illumina (Cap.4, § 4.2) e lo studio dell'eventuale correlazione delle CNVs così individuate con la manifestazione del disturbo autistico. Dall'analisi dei dati forniti dall'esperimento Illumina, abbiamo individuato 7 SNPs che presentavano errori di ereditarietà compatibili con la presenza di una delezione (Cap.7, §7.3). Ci siamo quindi proposti di:

- Verificare tramite sequenziamento i genotipi corrispondenti agli SNPs che presentano errori mendeliani.
- Qualora venga confermata la presenza di un errore mendeliano, determinare se ciò sia attribuibile alla presenza di una delezione e caratterizzarla. Verificare una possibile correlazione fra le CNVs individuate e il disturbo autistico tramite lo studio della loro frequenza e segregazione in un ampio campione di famiglie con autismo, oltre che in una popolazione di controllo.
- Qualora le CNVs individuate siano localizzate all'interno di regioni codificanti, effettuare uno *screening* di mutazione del gene corrispondente per rilevare eventuali altre varianti associabili al disturbo autistico.

## **6. Materiali e metodi**

### **6.1. Campioni di DNA**

I campioni di DNA genomico utilizzati in questo studio provengono dalla collezione di famiglie *multiplex* del Consorzio IMGSAC, di cui il nostro laboratorio fa parte. I criteri applicati per la diagnosi dell'autismo e per la raccolta delle famiglie, e le metodologie utilizzate per la preparazione dei campioni, sono stati precedentemente descritti [IMGSAC, 1998; 2001]. Il DNA genomico dei pazienti è stato estratto da sangue o da saliva secondo le condizioni descritte in Maestrini et al. [1999].

Nel corso di questi tre anni sono state inoltre raccolte 98 famiglie *singleton* (con un solo figlio affetto da autismo) inglesi del consorzio IMGSAC e 69 famiglie *singleton* italiane, queste ultime grazie ad una collaborazione del nostro laboratorio di Bologna con il Dott. Agatino Battaglia dell'Istituto Stella Maris di Pisa.

Il DNA genomico degli individui appartenenti alle famiglie *singleton* italiane è stato estratto utilizzando il kit QIAGEN MAXI (§ 6.2)

Come popolazione di controllo sono stati utilizzati i DNA di 196 individui caucasici di sesso misto appartenenti alla collezione ECACC (SIGMA).

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

### **6.2. Estrazione del DNA da sangue intero con kit Qiagen MAXI (per 5 ml di sangue)**

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

I campioni sono agitati e vortexati 3 volte per 5 secondi, quindi incubati per almeno 10' a 70°C. Dopo raffreddamento a temperatura ambiente si aggiungono 5 ml di etanolo assoluto, si vortexano i campioni tre volte per 5 secondi, e il lisato è versato nel tubo contenente la colonna Qiam. Dopo aver centrifugato per 3' a 3000 rpm (1850 g), si butta via il filtrato e si aggiungono 5 ml di buffer di lavaggio (AW1) e si centrifuga per 20' a 4500 rpm. Senza buttare via il filtrato, si aggiungono 5 ml di buffer di lavaggio (AW2), si centrifuga 20' a 4500 rpm, si sposta la colonna in un tubo da 50 ml pulito e si aggiungono successivamente 600 μl di buffer di eluizione (AE). Si lascia avvenire l'eluizione per 10' a temperatura ambiente, quindi si centrifuga 5' a 4500 rpm e si aggiungono altri 600 μl di buffer di eluizione (AE),

ripetendo il procedimento precedente una seconda volta. Il DNA eluito viene quindi messo in un tubo sterile da 1.5 ml, e quantificato allo spettrofotometro, con determinazione dell'assorbanza a 260 nm. L'integrità del DNA genomico è verificata tramite elettroforesi su gel di agarosio.

### **6.3. Disegno dei primers**

I primers per le reazioni di amplificazione sono stati disegnati utilizzando il programma PRIMER 3 ([http://www-genome.wi.edu/cgi-bin/primer/primer3\\_www.cgi](http://www-genome.wi.edu/cgi-bin/primer/primer3_www.cgi)). Per lo *screening* di mutazioni, i primers sono stati disegnati in modo da coprire gli esoni e parte delle regioni introniche adiacenti, e almeno 300-400 bp del promotore.

### **6.4. Amplificazione tramite Polimerasi Chain Reaction (PCR)**

Le reazioni di amplificazione sono effettuate in un volume finale variabile da 10 a 50  $\mu$ l. Per ogni reazione tramite PCR è preparato un controllo negativo senza il DNA, per identificare eventuali contaminazioni. La composizione finale di una miscela di reazione standard è la seguente:

40 ng di DNA genomico in TE 0.1X (24 ng nelle reazioni in  $V_f = 10 \mu$ l, ; 40 ng nelle reazioni in  $V_f = 15 \mu$ l)

1X buffer per AmpliTaq Gold DNA polimerasi (senza MgCl<sub>2</sub>)

2-3 mM di MgCl<sub>2</sub>

0.3  $\mu$ M di primer Forward

0.3  $\mu$ M di primer Reverse

0.2 mM di dNTPs

0.02 U/ $\mu$ l di AmpliTaq Gold DNA polimerasi (Applied Biosystems)

H<sub>2</sub>O a volume

Per la PCR vengono utilizzati apparecchi a 96 posti della MJ Research e dell'Applied Biosystems. In generale si effettua un primo step di 15 minuti a 95°C per l'attivazione dell'enzima AmpliTaq Gold, e per l'amplificazione si utilizza un protocollo definito *touch-down*: in questo tipo di protocollo lo step di *annealing* è dapprima impostato ad una temperatura più alta rispetto a quella calcolata sui primers utilizzati ( $T_m$ ), che viene diminuita di 0.5°C ad ogni successivo ciclo: in questo modo si conferisce una maggiore

specificità iniziale alla reazione. Dopo 14 cicli si arriva alla temperatura di *annealing* ( $T_m2$ ), che viene mantenuta costante per altri 25-30 cicli. Lo schema utilizzato per il protocollo di amplificazione *touch-down* è il seguente :

Pre-PCR step: 15' a 95°C

PCR 14 cicli:

95°C	30 secondi o 1 minuto
$T_m$ 1	30 secondi con calo di 0.5°C ad ogni ciclo
72°C	30 secondi o 1 minuto

PCR 25-30 cicli:

95°C	30 secondi o 1 minuto
$T_m$ 2	30 secondi o 1 minuto
72°C	30 secondi o 1 minuto

L'amplificazione di regioni molto ricche in G-C, nelle quali le due coppie di basi sono tenute insieme da tre legami idrogeno, può risultare più difficoltosa; in tal caso la reazione è condotta con l'aggiunta di 0.1 mM finale di 7-deaza-2'-deoxiguanosina-5'-trifosfato (Deaza GTP) e/o 5% di DMSO. Questi reagenti aiutano la reazione della polimerasi prevenendo la formazione di strutture secondarie del DNA e facilitandone la denaturazione.

**N.B.:** Per la reazione di amplificazione di frammenti molto lunghi (es. amplificazione allele delete per l'individuazione del punto di rottura nel gene *UPP2*) è utilizzata la DNA polimerasi Taq Herculase, nelle condizioni indicate dal fornitore. Il protocollo di amplificazione prevede una fase a 72°C più lunga (4 minuti).

Al termine della reazione l'amplificazione è verificata tramite elettroforesi su gel di agarosio. Tutte le coppie di primers e le condizioni di amplificazione utilizzate in questo studio sono riportate in Tabella 4 (a-h)

**Tabella 4a:** set di primers e condizioni utilizzate per le PCR per i geni *NAB1*, *UBE2E3*, *NCKAP1*

Nome frammento	primer Forward	primer Reverse	condizioni di PCR
<b><i>NAB1</i></b>			
promotore	GCCTCCTAGAGGATGTTGGA	GGGGAAGGGGAAGGTGGTG	
ex1	AGAGCGGAATCTCCCAATGT	CTCAGCCGGGAAAGTGTCT	
ex1/Promot**	promF	ex1R	2.5mM Mg TD: 62-55+30 (1'-1'-1') + DEAZ.AdNTPs(2mM)+DMSO(5%Vfin)
ex2	cttcgtcagggcagctctttt	tctctcaaccattcaagetta	2.5mM Mg TD: 62-55+30 (30"-30"-30")
ex3	tttttgaaacaatgattaccg	ctctctctcccactctgaa	2.5mM Mg TD: 62-55+30 (30"-30"-30")
ex4_1	cttcacagcagctcagcagc	GGATATTCCCAGCCATGTTG	
ex4_2	GCAAGCCCTTCATGTTAGA	CACACTCAGCCACAGAGAGC	
ex4_3	CAGGGGAAGTCAGATGTGGT	tgatcacattttggcacaact	
ex4_1/ex4_2*	ex4_1F	X4_2R	2.5mM Mg TD: 65-58+30 (1'-30"-1')
ex4_3*	ex4_2F	X4_3R	2.5mM Mg TD: 65-58+30 (1'-30"-1')
ex5	tttttaagtgaaccttttgcatt	tcaatagatgatgttccaactct	3mM Mg TD: 65-58+30 (1'-30"-30")
ex6	gctgagatgccctttggaata	tgaaagccatttttcaactagttt	2.5mM Mg TD: 62-55+30 (30"-30"-30")
ex7	tcagctagctgagataaagatttc	gaaataaaaagctcagcctcaga	2.5mM Mg TD: 62-55+30 (30"-30"-30")
ex8	ttgctgagctcgaatttgg	cagcactctctcttcttcttc	2.5mM Mg TD: 62-55+30 (30"-30"-30")
ex9	cctagagcttctcagcaca	tttttgatggggagagaaat	3mM Mg TD: 65-58+30 (30"-30"-30")
ex10	gagtttggggctttgtttta	TTGCAACACTCTTGATTTTCTC	2.5mM Mg TD: 62-55+30 (30"-30"-30")
*per amplificare ex4_3 abbiamo usato il seguente incrocio: ex4_2F/ex4_3R			
**per amplificare ex1 abbiamo usato PromF/ex1R			
<b><i>UBE2E3</i></b>			
ex1*	gtgagggaggctgtgag	aaacaagcccaagaagtg	1.5mM Mg TD: 62-55+29 (1'-30"-1') + doppio dei primers + DEAZ.AdNTPs(2mM) + DMSO(5%Vfin)
ex1_alt	cgcatggatggaaccag	cgcgagctcagcaagttc	2mM Mg TD: 65-58+29 (1'-30"-1')
ex2	gaagctttttcaacaagcag	ctgctatttgagtaaaaccaac	2mM Mg TD: 65-58+29 (1'-30"-1') + DEAZ.AdNTPs(2mM) + DMSO(5%Vfin)
ex3	gagccagctcgtcaatcaac	ttgtctctctctgctcac	2mM Mg TD: 65-58+29 (1'-30"-1')
ex4	ttaacactccctggttttagg	tcaatccatgaaagcaatg	2mM Mg TD: 65-58+29 (1'-30"-1') + doppio dei primers
ex5	agctccagtcacaataatagatc	agggcatttgataattctc	1.5mM Mg TD: 67-60+29 (30"-30"-30")
ex6a	gaaagggactttcctaactgg	GGCTCCAAGGATAAAAGCAG	2.5mM Mg TD: 65-58+29 (30"-30"-30")
ex6b	AAGGCATCTTCTCACTGTGC	CTGAGTACAAAGGGTAAAGG	2.5mM Mg TD: 65-58+29 (30"-30"-30")
ex1*= non si riesce a sequenziare.			
i primers per amplificare l'esone1 si chiamano ex1F_prova e ex1R_prova			
<b><i>NCKAP1</i></b>			
ex 1	gccacactagtgagcag	cttcccttgccttcaac	2mM Mg TD=60-53+32 (1'-30"-1') + DEAZ.AdNTPs(2mM) + DMSO(5%Vfin)
ex 2	tgacagattgctttttgattt	gctactataagcactgtgtgacca	2.5mM Mg TD=62-55+29 (30"-30"-30")
ex 3	aaaagctgtgtctccatagc	tttacaactgacacagcaatca	2.5mM Mg TD=62-55+29 (30"-30"-30")
ex 4	cttccagaactgaagttaacaca	agctaaagaaagcaagctggc	2.5mM Mg TD=62-55+29 (30"-30"-30")
ex 5	gcttctatgaagattttgtgtg	GGCGTGGGTATCTCTGTCA	2.5mM Mg TD=62-55+29 (30"-30"-30")
ex5-Falt per seq	gaaagaaattagcattttagaacagg	GGCGTGGGTATCTCTGTCA	2.5mM Mg TD=62-55+29 (30"-30"-30")
ex 6	CA'TGGAGCAAGgtagatttgc	atcccttatecaaccatce	2.5mM Mg TD=62-55+29 (30"-30"-30")
ex 7	agcaattatagctcaatgtaaa	agaacaacaagcaacagca	2.5mM Mg TD=62-55+29 (30"-30"-30")
ex 8	agggaggaagtgtgctcaat	gcacaatgctaagcagagag	2.5mM Mg TD=62-55+29 (30"-30"-30")
ex 9	tcaaaagctacactgccaacaa	ccctttatccccaaaactica	2.5mM Mg TD=62-55+29 (30"-30"-30")
ex 10	tgccagtttctgaattgat	ttcaacaagaatgatgtttaaga	2.5mM Mg TD=55-50+29 (30"-30"-30")
ex 11	ttggaagtgcaccataatgct	ctttttatgtagctcctgactca	2.5mM Mg TD=62-55+29 (30"-30"-30")
ex 12	cactcttgccaatgcttttc	tcagacataaccctattgta	2.5mM Mg TD=62-55+29 (30"-30"-30")
ex 13	gaaataagattaaggcagcagaa	ggcactgatatttcaaatgct	2.5mM Mg TD=62-55+29 (30"-30"-30")
ex 14	ccgtccactaatgatagactgc	tgaaatctttgctcagagaa	2.5mM Mg TD=62-55+29 (30"-30"-30")
ex 15	aatgtgcttatagattgtgaaaca	tcaaaaactcatattggttaagt	2.5mM Mg TD=62-55+29 (30"-30"-30")
ex 16	gtgatgcttttccccctgta	tgcttaagactgtgtaacacaaa	2.5mM Mg TD=62-55+29 (30"-30"-30")
ex 17	tgctaccagttatcattca	ttggttgctcaggtttga	2.5mM Mg TD=62-55+29 (30"-30"-30")
ex 18	ccrttagcctaattttccatgc	aaacaacaactggctaaagca	2.5mM Mg TD=62-55+29 (30"-30"-30")
ex 19	tatttgcacacagcctgga	aaaatgcttggaaatgaaga	2.5mM Mg TD=62-55+29 (30"-30"-30")
ex 20	gcattatctccaaataggtgat	aaactaatccagtrctactaatgct	2.5mM Mg TD=62-55+29 (30"-30"-30")
ex 21	aacagttaggtgaggtttat	ctgcatataatttaagtgaactg	2.5mM Mg TD=62-55+29 (30"-30"-30")
ex 22	caacttaaaatfatgacggaatg	tcaaaagtctaaagaatcaagac	2.5mM Mg TD=62-55+29 (30"-30"-30")
ex 23	tttgctatttctctctcttca	tttctgtctctatccctga	2.5mM Mg TD=62-55+29 (30"-30"-30")
ex 24	ttccaagctctcttatctca	caaaatgtgacccaagctaa	2.5mM Mg TD=60-53+29 (30"-30"-30")
ex 25	ttttccaagttaagaagaatcc	cattgtgaactgaactgaa	2.5mM Mg TD=62-55+35 (30"-30"-30")
ex 26	aaatttaggggttgcatt	caagcagctcaactcaaccactca	2.5mM Mg TD=62-55+29 (30"-30"-30")
ex 27	tccgaattctctggtcttc	aaagcagcattgaaatattgg	2.5mM Mg TD=62-55+29 (30"-30"-30")
ex 28	tcaacttctaactacttgaagaaa	cagactaccctctctccac	2.5mM Mg TD=62-55+29 (1'-30"-1')
ex 29	gccctcttatgacagattg	attcttcaattttggtgtgga	2.5mM Mg TD=62-55+29 (30"-30"-30")
ex 30	tggttggagttgaacatacct	agatgcaagaagaagattctca	2.5mM Mg TD=62-55+29 (30"-30"-30")
ex 31a	tgccctactggttgcataaaa	TGATCAGAAAATACAACCGTATGA	2.5mM Mg TD=62-55+29 (30"-30"-30")
ex 31b	TGTGTGCTTGGTITTTACCTG	TTTGGGAAATGATCTGCTTCT	2.5mM Mg TD=62-55+29 (30"-30"-30")

**Tabella 4b:** set di primers e condizioni utilizzate per le PCR per i geni *KCNJ3*, *RPRM*, *NR4A2*

Nome frammento	primer Forward	primer Reverse	condizioni di PCR
<b>KCNJ3</b>			
promotore	caggaaaaagtgggaacgag	GGCCCTCCTGATTATCCTG	2,5mM Mg + DMSO+ deazadNTPs TD: 65-58+30 (1'-30"-1')
ex1a	tctggtcCAGGCAGGTAAAG	TGGACGATGTGGTCACTACC	2,5mM Mg TD: 65-58+30 (1'-30"-1') + DEAZAdNTPs(2mM) + DMSO(5%Vfin)
ex1b	CCCCACCTCCTTATTGGTG	AGAGGAAGGCAGAAAGGGAAG	2,5mM Mg TD: 65-58+30 (1'-30"-1')
ex1c	GCTACCTCTCGGACCTCTTC	aaaggaagtttacctgtgcta	2,5mM Mg TD: 65-58+30 (1'-30"-1')
ex2	attactgttggctgccttg	tgagtcatttgcceaatc	2,5mM Mg TD: 65-58+30 (1'-30"-1')
ex3a	caaacgctttacatgtgc	CCCGGAACCTGAACTTATTCG	2,5mM Mg TD: 65-58+30 (1'-30"-1')
ex3b	TCTGTGGAATGCTTAGATGGAC	CCCTCCGTAACCTGTGCTTTC	2,5mM Mg TD: 65-58+30 (1'-30"-1')
<b>RPRM</b>			
promotore	ttctccacctccaagag	CGCATTATGTACAGGCTACGC	2,5mM Mg TD: 66-59+30 (1'-30"-1') + DMSO(5%Vfin)
ex1a	CGTTCCTGCGATGAATCC	GGCATGAGGACTTTCAGAGG	2,5mM Mg TD: 66-59+30 (1'-30"-1') + DMSO(5%Vfin)
ex1a-Falt per seq	GGCCCTAGGCAACCAGAC		
ex1b	CTACTGACCCGCCCTCTG	GGGAGCAGGAGAAGAGTGG	2mM Mg TD: 65-58+30 (1'-30"-1')
<b>NR4A2</b>			
promotore	aggtgggtgggaaggttag	CCtttattatgtgcttgc	2mM Mg TD: 65-58+30 (1'-30"-1')
ex1	cagctcgaaatataccaagc	tgccccacaagtggaaag	2mM Mg TD: 65-58+30 (1'-30"-1')
ex2	aacaggaagcctggaag	ttacaggtttgcttgc	2,5mM Mg TD: 65-58+30 (1'-30"-1')
ex3a	cggaagaaagtttcagttg	TTCTGGTGGAAAGTTGTGGAG	2,5mM Mg TD: 65-58+30 (1'-30"-1')
ex3b	CGGGTCGGTTTACTACAAGC	ctactggccaagcagag	2mM Mg TD: 65-58+30 (1'-30"-1')
ex4	tttggggaggaatcgtagac	cgtaatctggagtatgagc	2mM Mg TD: 65-58+30 (1'-30"-1')
ex5	tgaagctttacaagcactgg	ttacccccgttgaatctgag	2mM Mg TD: 65-58+30 (1'-30"-1')
ex6	attaattgccccgaacatc	ttccattccaaggttcc	2mM Mg TD: 65-58+30 (1'-30"-1')
ex7	aacctttcaagctccactg	tgcaaaaaccaagagaatc	2mM Mg TD: 65-58+30 (1'-30"-1')
ex8a	ccttggttttccaactagc	AAGGGGGCTAGGAGGGTTAC	2mM Mg TD: 67-60+30 (1'-30"-1')
ex8b	GCAAGTCACATGGGCAGAG	CCACCCACGCAACATTTAG	2mM Mg TD: 67-60+30 (1'-30"-1')
<b>UPP2</b>			
ex1	tcgtatatgggttttctgcat	tggttgatgtttccacatttc	2mM Mg TD: 65-58+30 (1'-30"-1')
ex2	ctctcaagatttgggttttg	tatgaaaaggactaacgc	2mM Mg TD: 65-58+30 (1'-30"-1')
ex3	tcaaaagggcctatctct	aatggcctctacagttaagca	2mM Mg TD: 65-58+30 (1'-30"-1')
ex4	ccctaaggaaggtttatgttca	catttccacctcaacaagaca	2mM Mg TD: 65-58+30 (1'-30"-1')
ex5	catgaaacataaacatgacatc	ttcttttccctccagett	2mM Mg TD: 65-58+30 (1'-30"-1')
ex6	gggtcccaattaggaggact	catcctattcagagcccagc	2mM Mg TD: 65-58+30 (1'-30"-1')
ex7_bis	tctaaacccatcacattgtaaa	gcaactttgtttattcaaccc	2mM Mg TD: 65-58+30 (1'-30"-1')
ex8	ggagacagctgtagagttta	tattacctggctgaagcc	2mM Mg TD: 65-58+30 (1'-30"-1')
ex9	tggtgttttggctaagct	TTTGAGTTCATATGCAGCA	2mM Mg TD: 65-58+30 (1'-30"-1')

**Tabella 4c:** set di primers e condizioni utilizzate per le PCR per il gene *INPPI*. [ID: numero identificativo dello SNP in banca dati].

Nome frammento	SNP ID	Primer FORWARD	Primer REVERSE	CONDIZIONI di PCR
INPP-X2	rs4656 (G456T)	cagagccacctgacatcac	atggagaagccacagatgc	2.5mM Mg,TD 62-55 + 30 cicli (1'-30"-1')
INPP-X4	rs11544940 (A606G) + rs10931450 (G651A)	ATGGCTTATCGTGTGGCTTT	CCCAAAATGTCTGTGGGAAG	2.5mM Mg,TD 64-57 + 32 cicli (1',30",1')
INPP-X6A	rs7592352 (A985G)	tttctcgtatcattaagtgtgaaag	GCAGCACAAAGAGTCCCATT	2.5mM Mg,TD 62-55 + 32 cicli (1'-30"-1')
INPP-X6B	rs1882891 (C1276A)	GGGCTGGTTATAAGAGCCTATG	GGACAAAGCCATCCTCAAAA	2.5mM Mg,TD 62-55 + 32 cicli (1'-30"-1')
Inpp_5UTR	rs3791809	AGCAGAAAGTTTGCCCACTGT	CAAGCAAGCGGTGTTTAGT	2.5mM Mg,TD 62-55 + 30 cicli (1'-30"-1')
rs1108939	rs1108939	TGGATTCTAGAAAACAGGGAACA	TACGACCTCAGCTTTGCTTG	2.5mM Mg,TD 64-57 + 30 cicli (30"-30"-40")
Inpp_intr5	rs4853696	GGTGGAAACAAGCAGTTCAGC	AGAAAGGCCCAATAGCACT	2.5mM Mg,TD 62-55 + 30 cicli (1'-30"-1'40")
Inpp_3UTR	rs6747000 + rs2582755	TGCATACAGATCCAGGAAGC	TCCCTTCGAGAGGAAACAAA	2.5mM Mg,TD 64-57 + 32 cicli (1'-30"-1'40")

**Tabella 4d:** Enzimi utilizzati per l'analisi di restrizione di 3 SNPs nel gene *INPPI*

Variante	Enzima	Frammento (bp)	Frammenti ottenuti dal taglio (bp)
INPP-X2 (G->T) T	<i>Hinf I</i>	388	112+276
INPP-X6A (A->G) G	<i>Ava II</i>	405	126+279
rs1108939 (G->T) G	<i>Hae III</i>	245	101+144

**Tabella 4e:** Set di sonde utilizzate negli esperimenti di LDR per il gene *INPPI*. Gli SNP indicati con uno stesso colore nella colonna "SNP ID", e aventi la medesima temperatura di ligazione, sono stati analizzati in contemporanea

SNP ID	Sonde	Lunghezza prodotto ligato (bp)	Temp. ligazione
RS_11544940	FAM-TTACCTTGAGGTTGTGTTCAACAGAGGAG NED-TTACCTTGAGGTTGTGTTCAACAGAGGAA PHOSPHO-GAAAACAGCAGAGCTTCTTAGCAAAGTCC	57 57	60.5 60.5
RS_10931450	FAM-TTCTTAGCAAAGTCTCAATGGTAACAAGGTA NED-CITAGCAAAGTCTCAATGGTAACAAGGTG PHOSPHO-GCATCTGAAGCATTAGCCAGGGTTG	57 55	61.5 61.5
RS_1882891	FAM-TCCTGTGCTGCTCATGCCATACTGC NED-ACCTGTGCTGCTCATGCCATACTGA PHOSPHO-GGGCCATGGGTGGGGGAATA	45 47	61.5 61.5
RS_2582755	FAM-aGGGGAGGCCAAGACGGGC NED-TGGGGAGGCCAAGACGGGT PHOSPHO-GGATCACGAGGTCAAGAAATCAGACC	46 46	61 61
RS_6747000	FAM-CATGAGCCACTGCTCTGGTCAATAC NED-CATGAGCCACTGCTCTGGTCAATAT PHOSPHO-TTATTGAAATGTTTCAAGCAGCACTAACAG	57 57	61 61
RS_3791809	FAM-GATAATAGTATAAGGATGAAAGTAATCCACTGAACTAC NED-GATAATAGTATAAGGATGAAAGTAATCCACTGAACTAT PHOSPHO-TCCTGATGGTGAATAGCTAAGTGTATCTCA	69 69	61 61
RS_4853696	FAM-taCAAGCTACTTACTTCTCTAAGCCTGTTTCC NED-tGCAAGCTACTTACTTCTCTAAGCCTGTTTCT PHOSPHO-TTATCTGTAAATATGTAACAACATAGACAGTAAGTTAAGa	73 73	61 61

**Tabella 4f:** set di primers e condizioni utilizzate per le PCR per il gene *INPP1*

Nome frammento	primer Forward	primer Reverse	condizioni di PCR
<b>INNP1</b>			
rs 3791809	AGCAGAAAGTTTGCCCACTGT	CAAGCAAGCGGTGGTTTAGT	2.5mM Mg,TD 62-55 + 30 cicli (1',30",1')
INPP-PX1	gaatcategctccact	ccccagatccaaggaggag	2.5mM Mg +DMSO,TD 68-61 + 32 cicli (1',30",1')
INPP-X2	cagaggcacctgacatcac	atggagaagccacagatgc	2.5mM Mg,TD 62-55 + 30 cicli (1',30",1')
INPP-X2 seq	tctgaactgctttctttcaca	tttgacgttaccttggctatga	2.5mM Mg,TD 67-60 + 30 cicli (1',30",1')
INPP-X3	ggaatttcagtctgattgctc	tggcttccaacatttagtca	2.5mM Mg,TD 62-55 + 30 cicli (1',30",1')
INPP-X4	tctctgttctttccactgaaac	tttattttccaacaccatec	2.5mM Mg,TD 64-57 + 32 cicli (1',30",1')
INPP-X4 bis	ATGGCTTATCGTGTGGCTTT	CCCCAAATGTCCTGTGGAAG	2.5mM Mg,TD 64-57 + 32 cicli (1',30",1')
INPP-X5	agagaagtcatatggaacagagtctt	caacaagcaagctgaggtc	2.5mM Mg,TD 62-55 + 32 cicli (1',30",1')
rs1108939	TGGATTCTAGAAACAGGGAACA	TACGACCTCAGCTTTGCTTG	2.5mM Mg,TD 64-57 + 30 cicli (30",30",40")
Inpp_intr5	GGTGGAAACAAGCAGTTCAGC	AGA AAGGCCCAATAGCACT	2.5mM Mg,TD 62-55 + 30 cicli (1',30",1'40")
INPP-X6A	tttctctgctataagttgaaag	GCAGCACAAAGAGTCCCATT	2.5mM Mg,TD 62-55 + 32 cicli (1',30",1')
INPP-X6B	GGGCTGGTTATAAGAGCCTATG	GGACAAAGCCATCCTCAAAA	2.5mM Mg,TD 62-55 + 32 cicli (1',30",1')
Inpp_3UTR	TGCATACAGATCCAGGAAGC	CAGTCTCGCTTTGTCACCAG	2.5mM Mg,TD 62-55 + 30 cicli (1',30",1'40")
Inpp_3UTR Rev alt		TCCCTTCGAGAGGAAACAAA	2.5mM Mg,TD 64-57 + 32 cicli (1',30",1'40")

**Tabella 4g:** set di primers e condizioni utilizzate per le PCR per il gene *UPP2*

Nome frammento	primer Forward	primer Reverse	condizioni di PCR
<b>UPP2</b>			
ex1	tcgtatattgggtttttcgc	tggtgatgtttccacatttc	2mM Mg TD: 65-58+30 (1'-30"-1')
ex2	ctctcacaagatttgggttttg	tatgaaaaaggaactaacgc	2mM Mg TD: 65-58+30 (1'-30"-1')
ex3	tcaaaggggctctctctct	aatggcctctacagtttaagca	2mM Mg TD: 65-58+30 (1'-30"-1')
ex4	ccctaaaggaagtttatgttca	catttccacctcaacaagaca	2mM Mg TD: 65-58+30 (1'-30"-1')
ex5	catggaacataaacatgacatc	ttcttttccctccagctt	2mM Mg TD: 65-58+30 (1'-30"-1')
ex6	ggcgcaattagggaggaact	catcctattcagagccagc	2mM Mg TD: 65-58+30 (1'-30"-1')
ex7_bis	tctaaccctcatcattgtaaa	gcaactctttgtattcaaaccc	2mM Mg TD: 65-58+30 (1'-30"-1')
ex8	gggagcagctgtagagtta	tattacctggctgtagagc	2mM Mg TD: 65-58+30 (1'-30"-1')
ex9	tggtgtgtttggctaagct	TTTGGAGTTTCATATGCAGCA	2mM Mg TD: 65-58+30 (1'-30"-1')

**Tabella 4h:** set di primers e condizioni utilizzate per delimitare la delezione nel il gene *UPP2*. [ID: numero identificativo dello SNP in banca dati].

SNP ID*	FORWARD	REVERSE	PCR condition
rs6437134	TGAAAGAGAGCACAAAGTCCA	CCTTTCTGGTTCCTATTGC	2mM Mg TD: 63-56+30 (1'-30"-1')
rs6734022	AGGCTTGGCGGAGCTATAATC	GAATCCACAGAAGGGGAAGCA	2mM Mg TD: 63-56+30 (1'-30"-1')
rs10190321	TCTCGAGGGTCACTCACTGAT	AATGTCGATGCGAGTGTCTG	2mM Mg TD: 63-56+30 (1'-30"-1')
rs6709528	tcctgtgatggacctgtc	ctccaaccctcagtgagt	2mM Mg TD: 63-56+30 (1'-30"-1')
rs13418976	CAGGGAGATCCCACATTACAC	GCTAATTTGCATCATCAACCTG	2mM Mg TD: 63-56+30 (1'-30"-1')
rs6437135+ rs6437136+ rs6437137	AGGAGAAGAAAAAGAGGTCCAG	TGCTCTTTAAAAATTATGATTGATGG	2mM Mg TD: 63-56+30 (1'-30"-1')
rs6437138	CACAGGCTTGCCTGACAATA	CCTGCAATTGGTCTCCTTGT	2mM Mg TD: 63-56+30 (1'-30"-1')
rs1468523	GACCAGCCCTGTTGACTTC	CCCTTGGGAACAATTATGG	2mM Mg TD: 63-56+30 (1'-30"-1')

### **6.5. Elettroforesi su gel di agarosio**

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

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

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

### **6.6. Reazione di sequenziamento ed analisi delle sequenze**

Per le reazioni di sequenza è utilizzato il kit ABI PRISM Big Dye Terminator Cycle Sequencing (V1.1 o V2.1 Perkin Elmer) ed i primers specifici di ciascun frammento. I prodotti di amplificazione sono dapprima purificati usando il sistema di filtrazione Multiscreen PCR della Millipore, seguendo le indicazioni del fornitore.

Le reazioni sono effettuate in un volume finale di 10 µl. La composizione finale di una miscela di reazione standard è la seguente:

5-20 ng di prodotto di amplificazione

1 µl di Big Dye Terminator Mix (contenente la DNA polimerasi e i dideossinucleotidi marcati con quattro differenti fluorocromi)

1,5 µl di mix di buffer di reazione (200 mM Tris HCl pH 8 / 5 mM MgCl<sub>2</sub>)

0.32 µl primer 10 µM ( reverse o forward)

H<sub>2</sub>O a volume

Per la reazione di sequenziamento è utilizzato il seguente programma:

Per 30 cicli :

96°C            10 secondi

50°C            5 secondi

60°C            4 minuti

I prodotti di sequenza sono quindi purificati attraverso una reazione di precipitazione selettiva; infatti, per effettuare una corretta lettura delle sequenze, è opportuno eliminare i dideossinucleotidi non incorporati.

Ad ogni miscela di reazione (10 µl) sono aggiunti:

2.5 volumi di etanolo assoluto (100%)

1/10 di volume di NaAcetato 3M pH 5.2

10 µl di H<sub>2</sub>O

Per comodità, viene preparata una “precipitation mix”, contenente etanolo assoluto “freddo”, NaAcetato 3M e H<sub>2</sub>O nelle giuste proporzioni, in modo da poter aggiungere 67 µl di mix a ciascun campione.

I campioni vengono lasciati per 15 minuti preferibilmente a -20°C, quindi centrifugati 30 minuti a 3000 rpm alla temperatura ottimale di 4°C. Il precipitato ottenuto è lavato con 150 µl di etanolo “freddo” al 70%. Si rimuove il supernatante capovolgendo i campioni su carta pulita e l’etanolo residuo è eliminato completamente centrifugandoli “a testa in giù” per 1 minuto a 50-60 g.

Le reazioni di sequenza sono state corse su un sequenziatore ABI PRISM 3730 (Applied Biosystems).

Per l’analisi e la lettura dei file di sequenza sono utilizzati i programmi *Genescan* e *Sequence Navigator* (Perkin Elmer) e *Sequence Analysis* (Perkin Elmer).

### **6.7. Genotipizzazione SNPs mediante digestione con enzimi di restrizione**

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

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

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

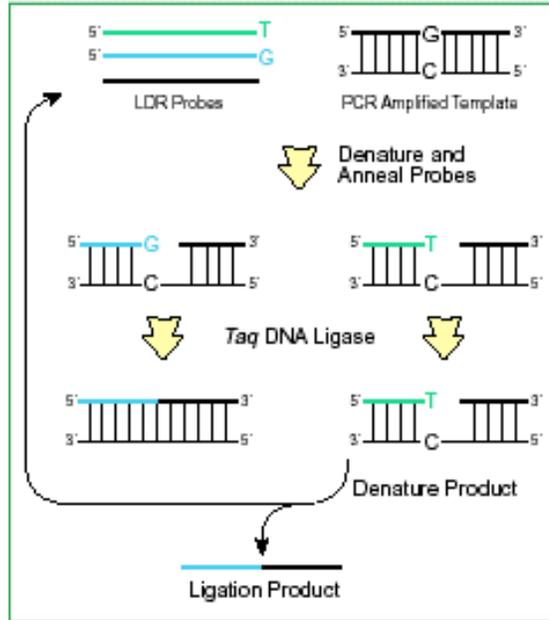
### **6.8. Metodo della Ligase Detection Reaction (LDR)**

La tecnica dell'LDR è un metodo rapido e sensibile per tipizzare polimorfismi del DNA (SNPs, microinserzioni, microdelezioni) con alta efficienza in un numero elevato di individui. Questa tecnica, sviluppata da Francys Barany [Barany et al.,1991], sfrutta la capacità della DNA ligasi di saldare preferenzialmente oligonucleotidi adiacenti, ibridizzati ad una molecola di DNA bersaglio, solo se c'è perfetta complementarità nel punto dove è presente l'interruzione a singolo filamento tra i due oligonucleotidi.

Per tipizzare SNPs biallelici, si disegnano 3 sonde, una comune e due allele-specifiche (Figura 19):

- **Sonda comune:** si lega al prodotto di PCR che funziona da stampo, in posizione subito a valle del nucleotide polimorfico.
- **Prima sonda allele-specifica:** è disegnata in modo da avere all'estremità 3' il nucleotide corrispondente all'allele *wild-type* dello SNP da tipizzare.
- **Seconda sonda allele-specifica:** ha al 3' il nucleotide corrispondente all'allele variante.

Se si denatura e rinatura il campione, in presenza delle 3 sonde descritte, la due sonde allele specifiche competono per l'appaiamento al template, in posizione adiacente alla sonda comune, che è fosforilata al 5'. Pertanto si genera una regione a doppio filamento, che possiede una rottura dovuta alla mancanza di un legame fosfodiesterico in corrispondenza del nucleotide che deve essere testato. Solamente la sonda allele-specifica che è perfettamente complementare al filamento stampo, sarà legata alla sonda comune per opera della DNA ligasi.



**Figura 19:** Esempio della tecnica dell'LDR: tipizzazione di uno SNP G/T. La sonda allelica che corrisponde all'allele *wild-type* è quella blu e ha una G al 3'. La sonda comune è quella nera. In questo esempio, l'individuo da tipizzare è omozigote per l'allele G *wild-type*. Quindi solamente la sonda allelica *wild-type* è legata alla sonda comune

Utilizzando una DNA ligasi termostabile, è possibile effettuare più cicli di denaturazione-rinaturazione-ligazione, ottenendo un aumento lineare del prodotto finale: infatti tale enzima rimane stabile alle temperature necessarie per la denaturazione del template, e catalizza efficientemente la reazione di ligazione a temperature compatibili con l'appaiamento controllato e specifico tra sonde e template.

Le sonde allele-specifiche possono essere disegnate in modo da avere una lunghezza diversa, cosicché i prodotti di ligazione contenenti l'allele *wild-type* o l'allele variante possano essere separati sulla base della dimensione del frammento.

Oppure, le due sonde allele-specifiche possono essere marcate con fluorocromi diversi, in modo che i prodotti di ligazione possano essere separati in base al colore.

Uno degli aspetti più interessanti dell'LDR è la possibilità di tipizzare contemporaneamente diversi SNPs (*multiplexed LDR*), utilizzando in una stessa reazione differenti set di sonde (una sonda comune e due sonde allele-specifiche, per ciascuno SNP da testare).

Per fare questo, i set devono essere progettati in modo che i prodotti finali di ligazione, corrispondenti ai diversi SNPs in analisi, possano essere discriminati tra loro in base alla dimensione (almeno quattro basi di differenza).

Nelle reazioni di LDR *multiplex* è necessario avere alcune accortezze:

- Le sonde LDR utilizzate per le diverse reazioni non si devono appaiare tra loro
- Le sonde LDR devono avere tutte temperature di appaiamento simili tra loro
- Ciascun prodotto di ligazione deve poter essere discriminato sulla base della lunghezza e del fluorocromo utilizzato.

In questo studio abbiamo utilizzato sonde allele-specifiche marcate con fluorocromi differenti al 5' (rispettivamente 6-FAM e NED).

Tutte le sonde sono state disegnate utilizzando il programma “*LDR Tm calculator*” (<http://www.well.ox.ac.uk/hill/Protocols.shtml>). Questo programma consente di calcolare tutte le temperature di *melting* ( $T_m$ ) di frammenti di lunghezza crescente disegnati sulla base della sequenza importata come template (in genere i frammenti hanno una lunghezza che non è mai minore di 20 basi e non supera mai le 40 basi).

Il set di sonde necessarie per una reazione di LDR si scelgono in modo tale che abbiano una temperatura di *melting* molto simile (non più di un grado di differenza). Poiché a volte la sostituzione nucleotidica da analizzare causa una notevole variazione di  $T_m$  della sonda allele-specifica rispetto all'altra sonda a parità di lunghezza, è necessario includere una o più basi complementari al 5' della sonda con temperatura più sfavorevole, e aggiungere uno o più nucleotidi aspecifici, che non si appaiano sul template, all'altra sonda per mantenere una lunghezza identica tra le due sonde.

Per una reazione *multiplex*, la scelta deve anche tener conto della dimensione di tutti i prodotti finali, che devono essere distinti tra loro oltre ad avere temperature di *melting* comparabili.

Il protocollo prevede l'amplificazione tramite PCR dei frammenti contenenti gli SNPs da analizzare. La reazione di PCR è effettuata in un volume finale di 10  $\mu$ l nelle condizioni precedentemente descritte.

Per effettuare reazioni *multiplex*, i prodotti di amplificazione dei diversi frammenti sono mescolati insieme, si crea così un pool dei prodotti di PCR da testare.

Questo pool è trattato con 1  $\mu$ l di Proteinasi K (soluzione 1mg/ml in 10 mM EDTA pH 8) per inattivare l'enzima Taq polimerasi, prima della reazione di ligazione.

Per la reazione di inattivazione è utilizzato il seguente programma:

37°C	15 minuti
55°C	10 minuti
90°C	10 minuti (per inattivare la Proteinasi K).

La reazione di ligazione sul pool trattato con Proteinasi K è effettuata in un volume finale di 7 µl, come segue:

2.5 µl pool dei prodotti di PCR specifici  
0.017 µM sonde allele-specifiche  
0.05 µM sonde comuni fosforilate  
0.7 µl buffer 10X per TaqDNA Ligasi  
0.25 U di TaqDNA Ligasi (Ampligase, CAMBIO)(5u/µl) per ciascun “trio” di sonde  
H<sub>2</sub>O a volume

Lo schema utilizzato per il protocollo di ligazione è il seguente :

95°C	1 minuto
Per 15 cicli:	
95°C	15 secondi
T <sub>m</sub> specifica	4 minuti

Infine, 1 µl di ciascun prodotto di ligazione viene risospeso in 9 µl di una miscela formamide-size standard (8.8 µl formamide + 0.2 µl Liz120) e corso sul sequenziatore ABI PRISM 3730 (Applied Biosystems). L'analisi dei genotipi viene fatta utilizzando il software GENEMAPPER<sup>TM</sup> (Applied Biosystems).

### **6.9. Analisi di associazione: caso controllo e TDT**

Per l'analisi di associazione caso-controllo è stato utilizzato il software cocophase (v2.40) del pacchetto UNPHASED (Dudbridge F (2003) Pedigree disequilibrium tests for multilocus haplotypes. Genet Epidemiol 25:115-221). Si considerano significative le differenze alleliche per le quali il p-value è  $<0.05$

Il software utilizzato per l'analisi di associazione tramite TDT è il sib\_tdt del pacchetto ASPEX (<ftp://lahmed.stanford.edu/pub/aspex>). La versione di sib\_tdt calcola le probabilità empiriche della statistica del chi<sup>2</sup>, permutando gli alleli dei genitori, mentre mantiene fisso l'IBD status dei fratelli affetti all'interno di una famiglia. In questo modo, il programma sib\_tdt dà una misura accurata dell'associazione indipendentemente dal linkage, permettendo così di utilizzare per l'analisi più fratelli affetti all'interno di famiglie nucleari.

## 7. Risultati

La regione di *linkage* per l'autismo sul cromosoma 2q è stata analizzata in dettaglio con l'obiettivo di identificare possibili alleli di rischio per il disturbo autistico, utilizzando due approcci complementari: ricerca diretta di mutazioni in geni candidati e studi di associazione. I dati relativi allo studio di associazione su larga scala condotto mediante la tecnologia Illumina sono stati utilizzati nel presente lavoro per completare la caratterizzazione dei geni candidati in esame e per l'identificazione di possibili varianti strutturali submicroscopiche nella regione cromosomica di interesse.

### 7.1. Risultati: screening di mutazioni nei geni candidati

È stata effettuata una ricerca di mutazioni in otto geni candidati, localizzati nell'intervallo di maggiore *linkage* per l'autismo sul cromosoma 2q, tra i marcatori D2S2370 e D2S364. I geni analizzati sono: *NCKAP1*, *INPP1*, *NAB1*, *UBE2E3*, la cui funzione biologica può essere correlata al disturbo autistico; *KCNJ3*, *NR4A2*, *REPRIMO*, localizzati all'interno di una o entrambe le delezioni mappate in due individui affetti da autismo nella regione 2q24 (Cap.3, §paragr. 3.1); *UPP2*, in cui è stata identificata una variante strutturale submicroscopica e che pertanto verrà descritto in un paragrafo a parte (cap 7 § 7.3-7.7).

I geni sono stati analizzati in un campione di 48 individui affetti da autismo non imparentati, selezionati tra le famiglie *multiplex* dell'IMGSAC che condividono due alleli identici per discendenza (IBD) nella regione di interesse. Questo criterio di scelta, cioè l'inclusione nell'analisi di quelle famiglie che maggiormente contribuiscono al valore positivo di *linkage* nella regione cromosomica in esame, rende più probabile l'identificazione di varianti di suscettibilità per l'autismo, se presenti. La struttura dei geni e loro sequenza completa era disponibile in banca dati (UCSC Genome Browser, <http://genome.ucsc.edu/cgi-bin/hgGateway>). Tutte le regioni codificanti ed importanti dal punto di vista funzionale (inclusendo le regioni introniche critiche per lo *splicing* e le regioni di regolazione dell'espressione genica note) sono state analizzate in modo sistematico alla ricerca di varianti nella sequenza del DNA. La ricerca di varianti di sequenza è stata effettuata mediante sequenziamento diretto (Cap. 6 § 6.6 ). Le posizioni dei geni nel genoma (UCSC Genome Browser, Maggio 2004 Assembly), la grandezza in basi dell'mRNA corrispondente, il numero di esoni, ed il numero di bp di DNA genomico che sono state analizzate per ciascun gene sono riportati in Tabella 5.

Nell'analisi sono state individuate un totale di 47 varianti di sequenza, così distribuite: 14 in *NCKAP1*, 6 in *INPP1*, 3 in *NAB1*, 1 in *UBE2E3*, 7 in *KCNJ3*, 11 in *NR4A2* e 5 in *REPRIMO*.(Tabella 6a-c)

<b>Gene</b>	<b>Accession n</b>	<b>banda</b>	<b>posizione sul genoma§</b>	<b>mRNA</b>	<b>n esoni</b>	<b>proteina</b>	<b>bp analizzate</b>
<b>REPRIMO</b>	NM_01984	2q23.3	154159360 - 154160830	1496 basi	1	109 aa	1558 bp
<b>UPP2</b>	NM_173355	2q24.1	158677199 - 158817985	2236 basi	9	317 aa	3455 bp
<b>NR4A2</b>	NM_173173	2q24.1	157006459 - 157014708	3269 basi	8	535 aa	4984 bp
<b>KCNJ3</b>	NM_002239	2q24.1	155380601 - 155538522	2890 basi	3	501 aa	3071 bp
<b>UBE2E3</b>	NM_006357	2q31.3	181670848 - 181753656	1559 basi	6	207 aa	3363 bp
<b>NCKAP1</b>	NM_013436	2q32.1	183615112 - 183728735	4619 basi	31	1128 aa	9953 bp
<b>NAB1</b>	NM_005966	2q32.2	191339354 - 191382997	4499 basi	10	487 aa	4866 bp
<b>INPP1</b>	NM_002194	2q32.2	191033941 - 191061813	1682 basi	6	399 aa	2602 bp

§ UCSC *Genome browser*, May 2004 *Assembly*

**Tabella 5:** per ognuno degli 8 geni studiati sono riportati l'Accession number, la banda cromosomica, le posizioni nel genoma, la grandezza in basi dell'mRNA corrispondente, il numero di esoni e il numero di paia di basi (bp) analizzate.

Di tutte le varianti identificate dallo *screening* di mutazioni, solo una determina un cambio di aminoacido nella proteina corrispondente, ed è la transizione A985G nell'esone 6 del gene *INPPI*. Questa variante, che causa la sostituzione della treonina in posizione 228 con alanina (Thr228Ala), era già descritta in banca dati (dbSNP, <http://www.ncbi.nlm.nih.gov/SNP/>), ma non ne era riportata la frequenza allelica. Per determinare se vi fosse una differenza nella frequenza di tale variante tra casi e controlli è stata effettuata un'analisi su 77 individui non affetti. Solamente 1 su 77 individui di controllo presenta la variante in eterozigosi (1.3 %), contro i 5 casi su 48 rilevati nei pazienti (10.4 %) (dati non mostrati).

Le altre varianti individuate nello screening dei 7 geni candidati sono in regioni introniche, in regioni non tradotte, oppure determinano sostituzioni sinonime nella relativa sequenza proteica. Inoltre, nessuna di queste varianti risulta interferire con le sequenze consenso per lo *splicing*. Tuttavia, varianti a livello degli esoni possono determinare uno *splicing* anomalo se alterano particolari sequenze consenso esoniche che servono per il legame di specifici fattori di regolazione dello *splicing*; tali sequenze sono denominate *exonic splicing enhancer*, ESE (<http://exon.cshl.org/ESE/>). Il programma ESE Finder non ha previsto alcun potenziale effetto dannoso per lo *splicing* di nessuna delle varianti esoniche silenti analizzate.

**Tabella 6:** risultati dello *screening* di mutazioni in 7 geni candidati per l'autismo. In rosso sono indicate le sostituzioni non sinonime. Sono evidenziati con lo stesso colore gli SNPs in completo LD fra loro nei geni *INPPI* e *NCKAPI*.

Gene (mRNA AC)	Posizione nel gene	posizione nel cDNA	Tipo di mutazione	Allele maggiore/ minore	dbSNP	Nr eterozigoti in 48 famiglie IMGSAC
<b>NCKAPI</b> (NM_013436)	esone 1	259	5' UTR	+/ins (GGC)		1
	introne 4	641 - 15		+/ins T		9
	introne 10	1275 + 22		G/A		9
	introne 10	1276 - 4		A/G	rs1400130	12
	esone 11	1333	silente (Ala 354)	T/C		2
	introne 18	2153 - 50		C/G		9
	introne 20	2425 - 32		G/A		3
	introne 21	2644 - 28		C/T		1
	introne 22	2775 - 120		A/G		1
	esone 26	3052	silente (Val 927)	C/G		12
	introne 27	3225 - 26		T/C		9
	introne 30	3541 + 46		G/A		1
	esone 31	3686	3' UTR	G/A		1
esone 31	4016	3' UTR	T/C	rs4666601	4	
<b>NAB1</b> (NM_005966)	esone 3	519	5'UTR	C/A	rs2293765	11
	introne 6	1577-23		A/C		1
	introne 9	1947+62		G/C		1
<b>UBE2E3</b> (NM_006357)	esone 2	459	silente (R25)	A/T		1

**Tabella 6a:** *NCKAPI*, *NAB1*, *UBE2E3*

Gene (mRNA AC)	Posizione nel gene	posizione nel cDNA	Tipo di mutazione	Allele maggiore/ minore	dbSNP	Nr eterozigoti in 48 famiglie IMGSAC
<b>REPRIMO</b> (NM_01984)	prom	1-79		G/T		1
	prom	1-11		C/A		1
	prom	1-7		C/G		1
	esone 1	748	3'-UTR	A/G		1
	esone 1		3'-UTR	G/C	rs1063728	19
<b>NR4A2</b> (NM_173173)	prom	1-434		ins(A)		15
	prom	1-198		C/T		1
	esone 1	1-125		G/A		1
	esone 1	1-94		G/A		2
	esone 1	1-91		G/A		1
	esone 5	1420	silente (L368)	C/T		1
	introne 5	1475+13		C/T		1
	introne 6	1679+18		ins(G)		14
	introne 7	1858 + 33		C/T	rs7562656	3
	esone 8	2445	3'UTR	T/G	rs12803	27
esone 8	2532	3'UTR	ins/del(CA)	rs3832066	1	
<b>KCNJ3</b> (NM_002239)	prom	1-590		C/T		1
	esone 1	1-3		G/A	rs3111034	2
	esone 1	108	5'UTR	ins/del (C)	rs5835535	15
	esone 1	273	silente (P26)	C/T		1
	esone 1	786	silente (S197)	C/T	rs3111033	2
	esone 3	1233	silente (H346)	T/C	rs17642086	21
	esone 3	2854	silente (D498)	T/C		1

**Tabella 6b:** *REPRIMO*, *NR4A2*, *KCNJ3*

Gene (mRNA AC)	Posizione nel gene	posizione nel cDNA	Tipo di mutazione	Allele maggiore/ minore	dbSNP	Nr eterozigoti in 48 famiglie IMGSAC
<b>INPP1</b> (NM_002194)	introne 1	241-13		T/G	rs2067402	16
	esone 2	456	silente (Thr 51)	G/T	rs4656	22
	esone 4	606	silente (Glu 101)	A/G	rs11544940	27
	esone 4	651	silente (Val 116)	G/A	rs10931450	16
	<b>esone 6</b>	<b>985</b>	<b>missenso (Thr 228 Ala)</b>	<b>A/G</b>	rs7592352	<b>5</b>
	esone 6	1276	silente (Arg 325)	C/A	rs1882891	11

**Tabella 6c:** *INPP1*

## **7.2. Risultati: Studi di associazione allelica nei geni candidati**

### **7.2a. INPPI e NCKAPI**

Le varianti più frequenti identificate durante lo *screening* di mutazioni dei geni *INPPI* e *NCKAPI* sono state utilizzate per effettuare studi di associazione allelica con l'autismo in 152 famiglie *multiplex* dell'IMGSAC mediante analisi di restrizione (*Restriction Length Fragment Polymorphism*, RLFP), oppure con la tecnica della *Ligation Detection Reaction* (LDR) (Cap. 6, §6.8). Dall'analisi dei genotipi degli individui studiati è stato possibile identificare in ciascun gene gli SNPs che erano in completo *Linkage Disequilibrium* (LD) fra loro, presenti sempre contemporaneamente sullo stesso aplotipo, e di conseguenza analizzarne solamente uno per ciascun aplotipo (Tabella 6a-c). Nel gene *NCKAPI* sono stati tipizzati gli SNPs C/G nell'introne 18 e C/G nell'esone 26. Nel gene *INPPI* il grado di LD riscontrato è basso, di conseguenza sono stati tipizzati gli SNPs *rs4656*, *rs11544940*, *rs10931450*, *rs7592352* (C1276A), *rs1882891* (A985G), quest'ultimo corrispondente alla sostituzione non sinonima Thr228Ala, che ha mostrato evidenza di associazione con l'autismo nel campione di 48 individui utilizzati per lo *screening*.

I genotipi ottenuti sono stati analizzati con il *Transmission Disequilibrium Test* (TDT), utilizzando il programma ASPEX sib\_tdt (Cap.6, §6.9). Nel gene *NCKAPI* non è stata evidenziata associazione con nessuno degli SNPs analizzati. Nel gene *INPPI*, l'analisi dello SNP C1276A, corrispondente alla sostituzione sinonima Arg325 nell'esone 6, ha fornito un risultato di TDT significativo ( $p\text{-value}=0.02$ ), soprattutto se si considerano separatamente le trasmissioni materne ( $p\text{-value}=0.0022$ )(Tabella 7). La variante A985G (Thr228Ala), anch'essa nell'esone 6 di *INPPI*, mostra un trend di trasmissione preferenziale dell'allele raro G, senza tuttavia raggiungere la significatività statistica ( $p\text{-value}=0.07$ ).

Le due varianti nel gene *INPPI* sono state ulteriormente analizzate in un campione allargato comprendente un totale di 256 famiglie *multiplex* dell'IMGSAC e 173 individui di controllo, per effettuare sia un test TDT che caso-controllo (Tabelle 7 e 8) Come si può osservare dai risultati del TDT, l'aggiunta di ulteriori famiglie ha aumentato il livello di significatività per la trasmissione preferenziale dell'allele C1276 ( $p\text{-value}=0.006$ )(Tabella 7). Poichè l'evidenza di associazione è con l'allele più frequente, è improbabile che questa rappresenti una variante eziologica, ma potrebbe essere in LD con una vicina variante di suscettibilità. La tipizzazione di 5 SNPs aggiuntivi per catturare tutta la variabilità presente nel gene *INPPI* e l'analisi degli aplotipi

effettuata per valutare gli effetti di differenti combinazioni alleliche di SNPs fiancheggiati, non ha tuttavia fornito risultati significativi (dato non mostrato).

Per la variante A985G, il test caso-controllo ha evidenziato una frequenza significativamente maggiore dell'allele raro 985G negli individui affetti da autismo rispetto al campione di controllo (test  $\chi^2$ ,  $p$ -value= 0,0007)(Tabella 8), come osservato nei dati preliminari ottenuti su 48 pazienti; tuttavia, la stessa variante non mostra un trend di trasmissione preferenziale nel campione più ampio nel test TDT.

Poiché la replicazione dei risultati in campioni indipendenti è un requisito essenziale per distinguere associazioni reali da associazioni spurie, le due varianti A985G e C1276A sono state analizzate tramite TDT in un addizionale campione di 127 famiglie *singleton* (IMGSAC e italiane). In nessun caso è stato osservato un  $p$ -value significativo.

**Tabella 7:** Risultati di TDT per le varianti A985G e C1276A nell'esone 6 di *INPP1*. Sono indicate separatamente le trasmissioni paterne e materne.

<b>INPP exon6 SNPs: 152 famiglie multiplex</b>											
	Allele	%	Trasmissione paterna			Trasmissione materna			Combinata		
			TR	NT	Chi2	TR	NT	Chi2	TR	NT	Chi2
<b>INPP_X6A</b> missenso (Thr228Ala)	G	97.9	2	10	5.33	5	5	0	7	15	2.91
	A	2.1	10	2	5.33	5	5	0	15	7	2.91
			p-val = 0.22			p-val = 1			<b>p-val = 0.07</b>		
<b>INPP_X6B</b> silente (Arg325)	A	15.5	24	29	0.47	15	36	8.65	48	74	5.54
	C	84.5	29	24	0.47	36	15	8.65	74	48	5.54
			p-val = 0.57			<b>p-val = 0.0022</b>			<b>p-val = 0.02</b>		

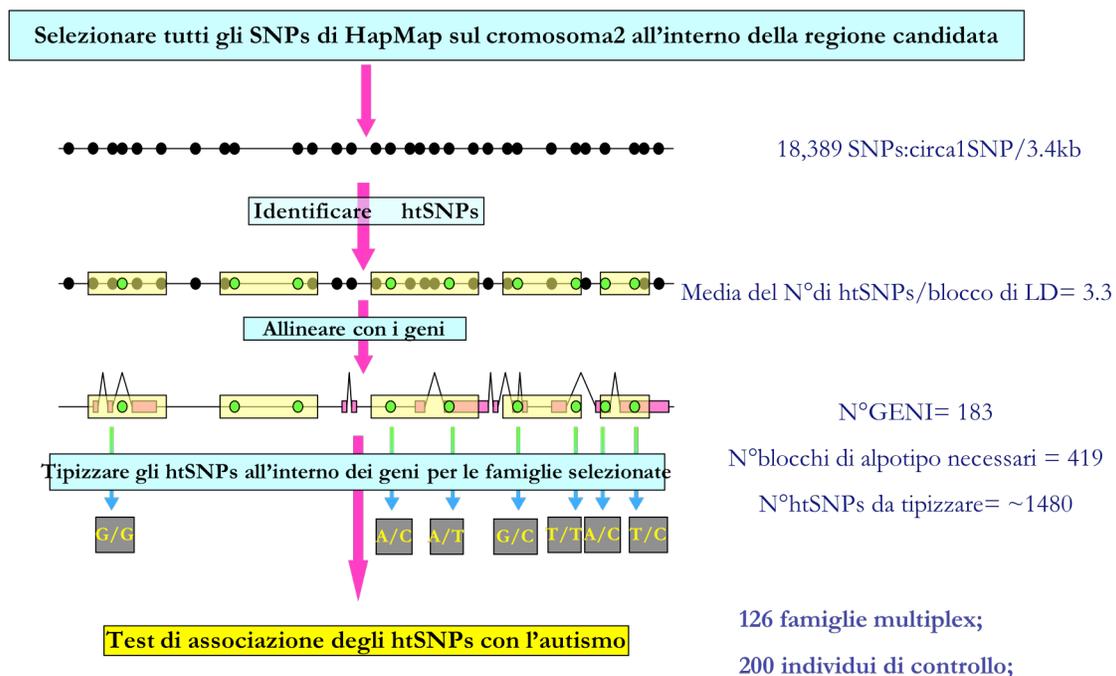
<b>INPP exon6 SNPs: 256 famiglie multiplex</b>											
	Allele	%	Trasmissione paterna			Trasmissione materna			Combinata		
			TR	NT	Chi2	TR	NT	Chi2	TR	NT	Chi2
<b>INPP_X6A</b> missenso (Thr228Ala)	G	97,2	8	17	3,24	5	9	1,14	15	28	3.93
	A	2,8	17	8	3,24	9	5	1,14	28	15	3.93
			p= 0.16						p= 0.14		
<b>INPP_X6B</b> silente (Arg325)	A	14,8	28	39	1,81	36	65	8,49	76	116	8,42
	C	85,2	39	28	1,81	65	36	8,49	116	76	8,42
			p= 0.22			p= 0.0042			<b>p-val = 0.0060</b>		

**Tabella 8:** Test caso-controllo per le varianti A985G e C1276A.  
a Numero di individui analizzati

	<b>Cases</b>				<b>Controls</b>				<b>Chi2 (p)</b>
	<b>Genotype %</b>			<b>N<sup>a</sup></b>	<b>Genotype %</b>			<b>N<sup>a</sup></b>	
	<b>11</b>	<b>12</b>	<b>22</b>		<b>11</b>	<b>12</b>	<b>22</b>		
<b>INPP1_X6A</b>	93	7	0.47	259	99	0.01	0	173	<b>0,0007</b>
<b>INPP1_X6B</b>	0.83	0.17	0	259	0.87	0.13	0.01	173	0.23

### 7.2b. *KCNJ3*, *NR4A2*, *NAB1*, *UBE2E3*

Anche per i geni *KCNJ3*, *NR4A2*, *NAB1*, *UBE2E3* è stato condotto uno studio di associazione per mettere in evidenza la eventuale correlazione tra varianti del DNA ed il fenotipo autistico. I geni sono stati inclusi nell'esperimento Illumina (Cap 4, § 4.2) e i genotipi ottenuti sono stati utilizzati tramite test TDT e caso-controllo. Lo schema sperimentale seguito per l'esperimento Illumina è descritto in Figura 20.



**Figura 20:** schema sperimentale seguito per l'esperimento Illumina

Brevemente, utilizzando i dati messi a disposizione da Hap Map, sono stati definiti i blocchi di elevato LD nella regione di *linkage* sul cromosoma 2q e sono stati selezionati i tagSNPs. I blocchi sono stati allineati con i geni di interesse, includendo anche 1000 bp a monte di ciascun gene, al fine di coprire le regioni prossimali di regolazione trascrizionale. Il numero di blocchi presenti e di SNPs analizzati per i geni *KCNJ3*, *NR4A2*, *NAB1*, *UBE2E3* sono in Tabella 9. Il gene *REPRIMO* cade all'interno di una regione a basso LD, e non è stato possibile includerlo in questo tipo di analisi.

GENE	Nr blocchi	Nr TagSNPs
UBE2E3	2	9
KCNJ3	4	16
NAB1	3	10
NR4A2	1	4

**Tabella 9:** numero dei blocchi e di SNPs analizzati per i geni *KCNJ3*, *NR4A2*, *NAB1*, *UBE2E3*

Gli SNPs sono stati analizzati in un campione di 126 famiglie trios (padre-madre-figlio affetto), selezionate tra le famiglie *multiplex* dell'IMGSAC che maggiormente contribuiscono al *linkage* nella regione 2q, e 192 individui di controllo. In nessun caso è stato ottenuto un *p-value* significativo, sia nel TDT che nel test caso-controllo.

### **7.3. Individuazione di delezioni submicroscopiche nella regione 2q24-q32.2 in famiglie con autismo**

Delezioni submicroscopiche possono essere evidenziate dalla presenza di errori di segregazione degli alleli di marcatori all'interno di famiglie. Partendo da questo presupposto, i circa 850.000 genotipi ottenuti nelle 126 famiglie analizzate nell'esperimento Illumina sono stati controllati per identificare l'eventuale presenza di errori di ereditarietà. E' da sottolineare che la causa più frequente di errori di ereditarietà è rappresentata da errori di *genotyping*, che è tra l'altro importante individuare e correggere prima di procedere all'analisi dei dati.

Nel nostro campione sono stati individuati 34 SNPs che presentano genotipi non compatibili con le leggi di segregazione mendeliana, 7 dei quali sono risultati compatibili con la presenza di delezioni e sono stati ulteriormente analizzati in questo studio (Tabella 10). La presenza di una delezione comporta che il genotipo per un dato SNP, localizzato all'interno di essa, sia emizigotico, e quindi nel saggio Illumina si legherà solo la sonda marcata specifica per l'unico allele presente, causando l'attribuzione di un erroneo genotipo omozigotico agli individui portatori della delezione. Inoltre, il segnale di fluorescenza risulterà meno intenso, in quanto prodotto dall'unica sonda che si è riuscita a legare, e esterno al raggruppamento del genotipo corrispondente nella rappresentazione grafica dei dati. Un esempio è riportato in Figura 21.

Pattern compatibili con la presenza di una delezione possono essere dovuti anche a errori di *genotyping* determinati da SNPs non noti che si trovano nelle vicinanze dello SNP in questione, o da alleli non descritti dello SNP stesso. In questo caso l'errore è dovuto al fatto che la sonda non riesce a legarsi in modo efficiente al template in presenza di basi varianti, causando perdita di alleli e un pattern di ereditarietà anomalo. La causa dei pattern di trasmissione anomali ottenuti per i 7 SNPs di interesse nel corso dell'esperimento Illumina è stata verificata analizzandoli tramite amplificazione per PCR e sequenziamento nelle famiglie in cui hanno mostrato segregazione errata.

I genotipi determinati per sequenza sono riportati in Tabella 10.

SNPs	Genotipi da Illumina	Genotipi da sequenza	Genotipi varianti addizionali
<b>rs2113480</b>	83.1: BB	83.1: BB	(Del 4bp a + 1)
	83.2: AA	83.2: <b>AA</b>	+/+
	83.4: BB	83.4: <b>AB</b>	+/-
	135.1: BB	135.1: <b>AB</b>	+/-
	135.2: AA	135.2: AA	+/+
	135.3: AA	135.3: <b>AA</b>	+/-
<b>rs6433368</b>	34.1: AA	34.1: <b>AA</b>	(Del 3bp a + 1)
	34.2: AB	34.2: <b>AB</b>	+/-
	34.3: BB	34.3: <b>AB</b>	+/+
	175.1: AA	175.1: <b>AA</b>	+/-
	175.2: BB	175.2: <b>BB</b>	+/+
	175.3: BB	175.3: <b>AB</b>	+/-
<b>rs2287618</b>	268.1: AA	268.1: <b>AB</b>	(SNP a + 2bp; SNP a + 16 bp)
	268.2: AB	268.2: <b>AB</b>	C/T; G/A
	268.3: BB	268.3: <b>BB</b>	C/C; G/G
	257.1: AA	257.1: AA	C/T; G/A
	257.2: BB	257.2: <b>BB</b>	C/C; G/G
	257.4: AA	257.4: <b>AB</b>	C/T; G/A
<b>rs1267070</b>	89.1: AB	89.1: AB	(SNP a -1 bp)
	89.2: BB	89.2: <b>BB</b>	C/C
	89.3: AA	89.3: <b>AB</b>	A/C
<b>rs2592790</b>	35.1: AB	35.1: AB	A/C
	35.2: AA	35.2: <b>AC</b>	G/C
	35.3: BB	35.3: <b>BC</b>	A/G
<b>rs6720043</b>	38.1: AA	38.1: AA	(SNP a + 2 bp)
	38.2: BB	38.2: <b>AB</b>	A/A
	38.3: AA	38.3: <b>AA</b>	A/G
<b>rs7585982</b>	194.1: AB	194.1: AB	-
	194.2: AA	194.2: AA	-
	194.4: BB	194.4: <b>BB</b>	-

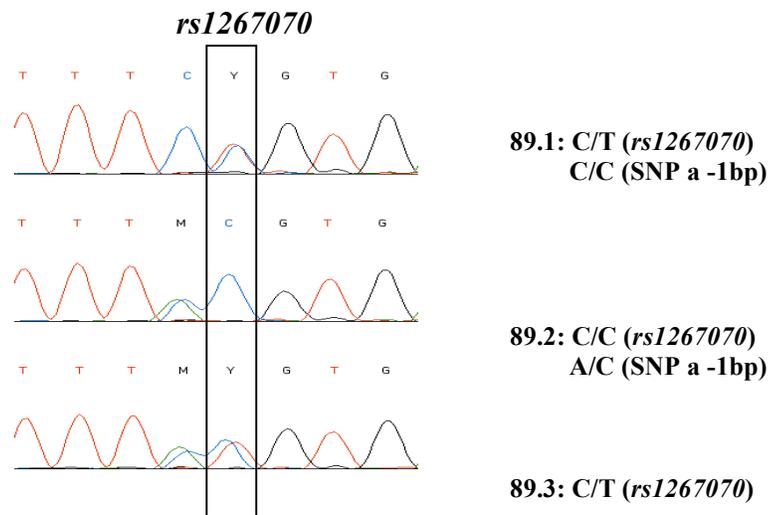
**Tabella 10:** Confronto genotipi degli SNPs ottenuti con esperimento Illumina e tramite sequenza. Gli alleli degli SNPs ai quali la sonda dell'esperimento Illumina non riesce ad ibridizzarsi sono evidenziati in grassetto

**SNP *rs2113480*:** dall'analisi delle sequenze dei membri delle famiglie 83 e 135 è stato stabilito che gli individui 83.2 (madre) e 135.1 (padre) sono portatori in eterozigosi di una delezione di 4 bp, posta 1bp a valle dello SNP *rs2113480*, che trasmettono ai figli 83.4 e 135.3. La presenza della delezione è la causa dell'errore di tipizzazione dello SNP tramite il saggio Illumina.

**SNP *rs6433368*:** la sequenza del frammento contenente lo SNP *rs6433368* nelle famiglie 34 e 175 rivela un pattern di ereditarietà mendeliano, in quanto gli individui 34.3 e 175.3 sono eterozigoti, e non omozigoti come risultato dal saggio Illumina. Gli individui 34.1 e 175.1 sono portatori in eterozigosi di una delezione di 3 bp posta 1bp a valle dello SNP *rs6433368*, che trasmettono ai figli 34.4 e 175.3. La presenza della delezione è la causa dell'errore di tipizzazione dello SNP tramite il saggio Illumina.

**SNP *rs2287618*:** in seguito alla lettura delle sequenze per lo SNP *rs2287618* nelle famiglie 268 e 257, si osserva un pattern di ereditarietà mendeliano. Gli individui 268.1 e 257.4 sono eterozigoti e non omozigoti come risultato dal saggio Illumina. Gli individui 268.1 e 257.2 sono risultati eterozigoti per due SNPs rari, 2bp e 16 bp a valle dello SNP *rs2287618* e non precedentemente descritti, che trasmettono ai figli 268.3 e 257.4. Tali varianti, se presenti, sono la causa dell'apparente perdita di un allele nel saggio Illumina.

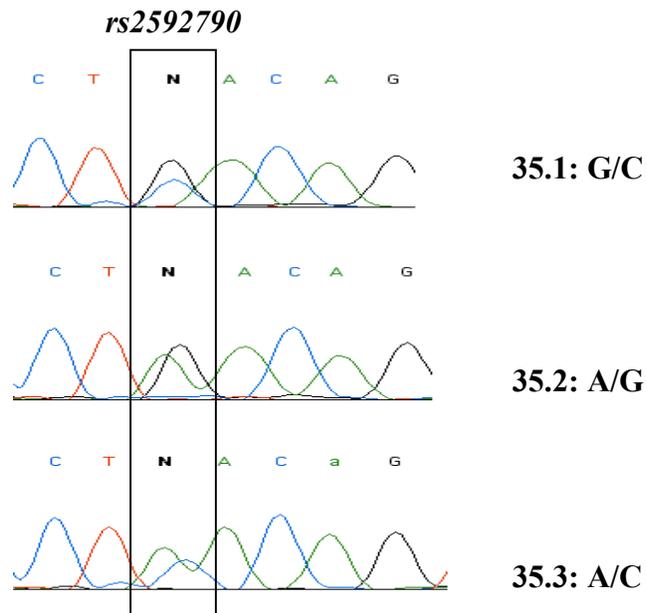
**SNP *rs1267070*:** in seguito alla lettura delle sequenze per lo SNP *rs1267070* nella famiglia 89, si osserva un pattern di ereditarietà mendeliano. L'individuo 89.3 è eterozigote e non omozigote come risultato dal saggio Illumina. Infatti, l'individuo 89.2 presenta uno SNP 1bp a monte dello SNP *rs1267070* e non precedentemente descritto (Figura 22), che trasmette al figlio 89.3. Questa variante è la causa dell'apparente perdita di un allele nel saggio Illumina.



**Figura 22:** Elettroferogramma della sequenza intorno allo SNP *rs1267070*

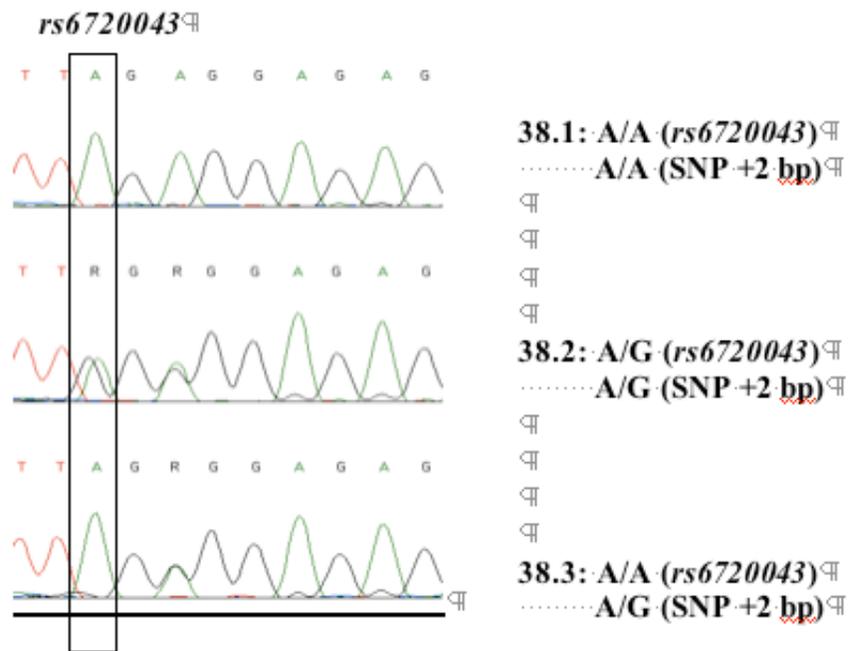
Gli individui 89.2 e 89.3 risultano essere eterozigoti per uno SNP non noto in posizione -1bp rispetto allo SNP *rs1267070*

**SNP *rs2592790*:** in seguito alla lettura delle sequenze per lo SNP *rs2592790* nella famiglia 35, si osserva un pattern di ereditarietà mendeliano. L'individuo 35.2 è eterozigote e non omozigote come risultato dal saggio Illumina. L'individuo 35.2 presenta un allele alternativo non noto (allele A), che trasmette al figlio 35.3 (Figura 23). In presenza di tale variante nucleotidica la sonda non è in grado di legarsi al template con conseguente perdita di un allele.



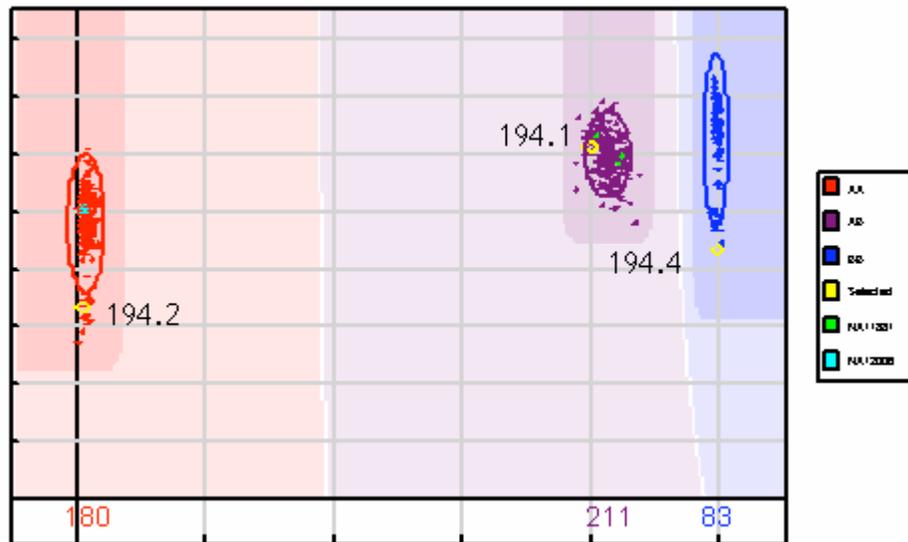
**Figura 23:** Elettroferogramma delle sequenze intorno allo SNP *rs2592790* Nel riquadro è evidenziato lo SNP *rs2592790* triallelico (A/C/G).

**SNP *rs6720043*:** in seguito alla lettura delle sequenze per lo SNP *rs6720043* nella famiglia 38, si osserva un pattern di ereditarietà mendeliano. L'individuo 38.2 è eterozigote e non omozigote come risultato dal saggio Illumina. L'individuo 38.2 presenta uno SNP 2bp a valle dello SNP *rs6720043* non precedentemente descritto (Figura 24), che trasmette al figlio 38.3. In presenza di tale variante nucleotidica la sonda non è in grado di legarsi al template con conseguente perdita di un allele.



**Figura 24:** Elettroferogramma delle sequenze intorno allo SNP *rs6720043*. Gli individui 38.2 e 38.3 risultano essere eterozigoti per uno SNP non noto in posizione +2bp rispetto allo SNP *rs6720043*.

**SNP *rs7585982*:** in seguito alla lettura delle sequenze per lo SNP *rs7585982* nella famiglia 194, sono stati confermati i genotipi prodotti dall'esperienza Illumina, che mostrano una segregazione non mendeliana compatibile con la presenza di una delezione in questa regione. Inoltre, gli individui 194.2 e 194.4 hanno prodotto nel saggio Illumina per lo SNP *rs7585982* delle fluorescenze di intensità inferiore rispetto alla media, e nella rappresentazione grafica dei risultati figurano al di sotto dei *cluster* (Figura 25).



**Figura 25:** Immagine dei cluster prodotti dal software Illumina relativi allo SNP *rs7585982*. I cerchietti gialli evidenziano i dati di fluorescenza relativi agli individui della famiglia 194.

I dati ottenuti fanno quindi ipotizzare che l'individuo 194.2 sia portatore in eterozigosi di una delezione che comprende lo SNP *rs7585982* ed è trasmessa al figlio 194.4. La delezione comprende i primers utilizzati per amplificare e sequenziare il frammento contenente lo SNP *rs7585982*, localizzato nell'introne 6 del gene *UPP2*.

#### 7.4. Mappaggio della delezione nel gene UPP2

Per caratterizzare la delezione che segrega nella famiglia 194 all'interno del gene *UPP2* è stato utilizzato il seguente approccio sperimentale. I dati di *genotyping* ottenuti nell'esperimento Illumina per gli SNPs a monte e a valle dello SNP deleto *rs7585982* hanno permesso di definire la lunghezza massima della delezione (Tabella 11). Questa è compresa fra gli SNPs *rs6709528* e *rs1468523*, che risultano informativi e non deleti per la famiglia 194 e distano circa 17,4 kb. Per delimitare ulteriormente la delezione, sono stati tipizzati nella famiglia 194 altri 8 SNPs, localizzati nell'intervallo in esame (fra gli SNP *rs6709528* e *rs1468523*) e che presentano un grado di correlazione elevato ( $r^2 > 0.8$ ) con lo SNP deleto; questo per aumentare la probabilità che gli SNPs scelti siano informativi nella famiglia. I dati ottenuti sono riassunti in Tabella 11 e mostrano che la delezione interessa un minimo di 5,9 Kb (dallo SNP *rs6437134* allo SNP *rs10190321*) fino ad un massimo di 9,3 Kb (dallo SNP *rs13418976* allo SNP *rs6437135*).

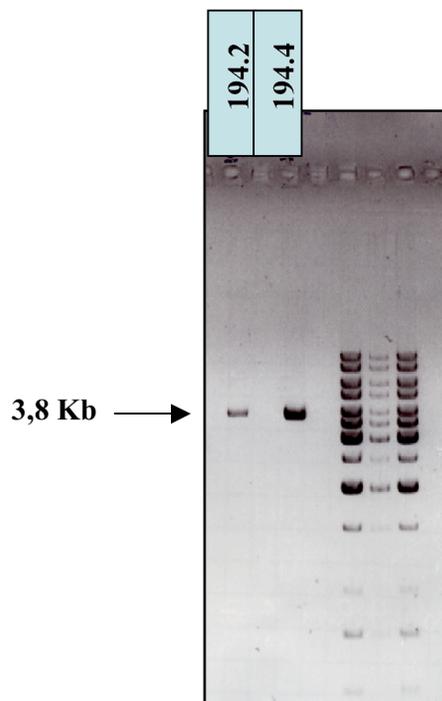
SNP	POSIZIONE UCSC May2004	GENOTIPI						METODO DI GENOTYPING
		194.1		194.2		194.4		
rs6709528	158795933	T	C	T	C	C	C	ILLUMINA
rs13418976	158796745	A	G	A	G	G	G	Sequenza
rs6437134	158798817	C	T	C	-	T	-	Sequenza
<b>rs7585982</b>	<b>158801104</b>	<b>A</b>	<b>G</b>	<b>A</b>	<b>-</b>	<b>G</b>	<b>-</b>	<b>ILLUMINA</b>
rs6734022	158802654	C	T	C	-	T	-	Sequenza
rs10190321	158804751	G	A	G	-	A	-	Sequenza
rs6437135	158806085	T	A	T	A	A	A	Sequenza
rs6437136	158806085	G	A	G	A	A	A	Sequenza
rs6437137	158806085	C	T	C	T	T	T	Sequenza
rs6437138	158810238	C	T	C	C	T	C	Sequenza
rs1468523	158813341	A	T	A	T	T	T	ILLUMINA

9,3 Kb

5,9 Kb

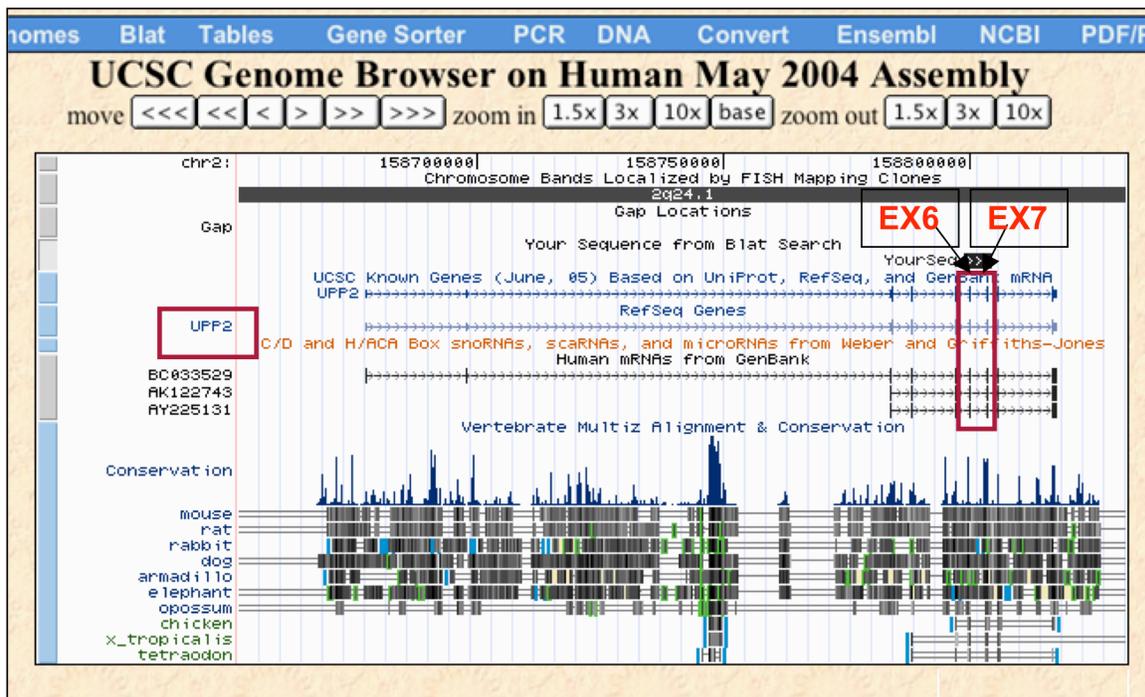
**Tabella 11:** Elenco degli SNP analizzati, relativa posizione cromosomica e genotipo nella famiglia 194. Il cerchio rosso evidenzia gli SNPs compresi nella delezione

I punti di rottura della delezione sono stati individuati mediante esperimenti di *long-extension PCR*. Per la reazione di amplificazione sono stati utilizzati il primer *forward* precedentemente utilizzato per l'amplificazione dello SNP *rs13418976* e quello *reverse* utilizzato per lo SNP *rs6437135* (entrambi gli SNPs mappano al di fuori della delezione). L'allele non delecto risulta troppo lungo per essere amplificato, mentre quello delecto risulta di circa 3,8Kb (Figura. 26).



**Figura 26:** Corsa elettroforetica su gel del prodotto di long extension PCR per la delezione in UPP2. E' visibile una banda di circa 3.8 kb in corrispondenza dei campioni 194.2 e 194.4 (in cui si è amplificato solo l'allele delecto). Nel campione 194.1 non è visibile nessuna banda, in quanto gli alleli non delecti sono troppo grandi per potere essere amplificati con questo protocollo.

In base all'analisi di sequenza, la delezione è risultata avere un'estensione di 5896 bp, con i punti di rottura in corrispondenza rispettivamente delle basi 158798874 e 158804770 del cromosoma 2 (UCSC *Genome Browser*, May2004 *Assembly*, NCBI Build 35). La porzione delecta è all'interno del gene *UPP2* (2q24.1) e comprende interamente due esoni codificanti (esone 6 ed esone 7) (Figura 27). A livello del messaggero questa delezione determina un *frameshift* e l'introduzione di uno *stop codon*, con probabile perdita di funzione del prodotto proteico.



**Figura 27:** Struttura del gene *UPP2* (sito web di UCSC *Genome Browser*, May 2004). La delezione comprende l'esone 6 e 7 del gene *UPP2* (rettangolo rosso).

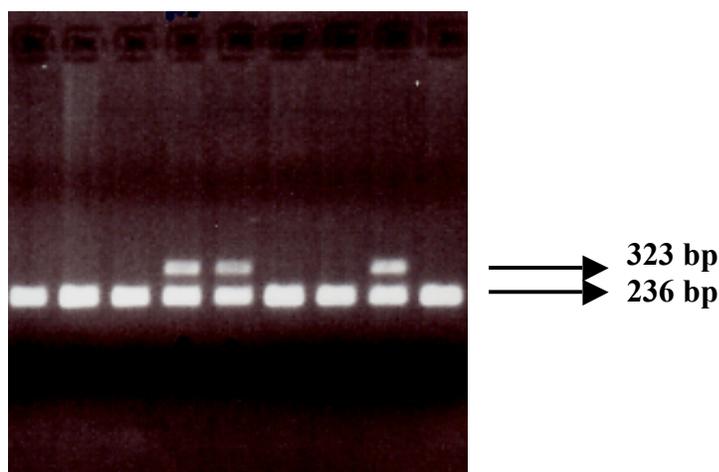
### 7.5. Analisi della delezione in *UPP2* in famiglie con autismo

La delezione che segrega nella famiglia 194 rappresenta una *Copy Number Variant* (CNV). Il passo successivo è stato analizzare l'eventuale associazione di questa variante con il fenotipo autistico. A tale scopo è stata determinata la frequenza della delezione in 256 famiglie multiplex con autismo della collezione IMGSAC, e in 188 individui di controllo ECACC (rapporto maschio:femmina di 4:1, come per il campione con autismo).

Lo *screening* della delezione è stato effettuato mediante la co-amplificazione di due frammenti. Nella miscela per la reazione di amplificazione sono state infatti utilizzate due differenti coppie di primers: una coppia amplifica un frammento interno alla delezione (236 bp negli alleli non deleti), l'altra coppia amplifica un frammento comprendente le regioni esterne alla delezione (323 bp negli alleli deleti). Le dimensioni dei frammenti sono tali per cui è possibile discriminarli mediante corsa elettroforetica del prodotto di amplificazione.

Il pattern osservabile è il seguente (Figura.28):

- Omozigoti WT: si amplifica solo il frammento interno alla delezione (quello che comprende le regioni esterne alla delezione è troppo lungo per essere amplificato);
- Eterozigoti per la delezione: si amplificano entrambi i frammenti;
- Omozigoti per la delezione: si amplifica solo il frammento che comprende le regioni esterne alla delezione (quello interno alla delezione non può amplificarsi).



**Figura 28:** Test PCR per la delezione individuata in *UPP2*: esempio di elettroforesi su gel dei prodotti di co-amplificazione. Sono visibili 2 bande in corrispondenza dei campioni che hanno la delezione in eterozigosi. I campioni in cui si amplifica solo una banda di 236 sono omozigoti non deleti.

I risultati dello screening per la CNV individuata nel gene *UPP2* mostrano che la delezione è in generale poco frequente e non presenta differenze significative tra il campione di individui con autismo e il gruppo di controllo (Tabella 12). Inoltre dall'analisi della segregazione all'interno delle famiglie è risultato che la delezione non segrega con il fenotipo autistico, in quanto non è preferenzialmente trasmessa dai genitori a uno o entrambi i figli affetti (Tabella 13), ed è talvolta trasmessa anche ai figli non affetti (dato non mostrato).

**Tabella 12:** Test caso-controllo per la delezione individuata in *UPP2* su 126 individui affetti da autismo e 188 individui di controllo.

<b>Allele</b>	<b>N° allele nei casi</b>	<b>Frequenza dell'allele nei casi</b>	<b>N° allele nei controlli</b>	<b>Frequenza dell'allele nei controlli</b>	<b>p value</b>
1	238	0,9835	364	0.9681	
2 (deleto)	4	0,01653	12	0.03191	0,2258

**Tabella 13:** Analisi della frequenza e della segregazione della delezione individuata in *UPP2* sull'intero campione di famiglie con autismo analizzate.

<b>N° famiglie multiplex analizzate</b>	<b>N° genitori con un allele deleto</b>	<b>Freq allele deleto nei genitori</b>	<b>N° trasmissioni allele deleto a un solo figlio affetto (%)</b>	<b>N° trasmissioni allele deleto a tutti e 2 i figli affetti (%)</b>
265	30	0,028	8 (26.7%)	5 (16.7%)

### 7.6. Screening di mutazioni del gene *UPP2*

La CNV individuata nel gene *UPP2* non mostra associazione significativa con l'autismo in base alle analisi effettuate. Tuttavia, non è da escludere che la presenza di altre varianti all'interno dello stesso gene in individui portatori della delezione possa contribuire al fenotipo autistico. Il gene *UPP2* si trova nella regione di *linkage* per l'autismo sul cromosoma 2 e rappresenta un interessante candidato in base a recenti evidenze sperimentali, come precedentemente descritto (Cap. 3, § 3.1).

Lo *screening* di mutazioni del gene *UPP2* è stato condotto in 47 individui con disturbo autistico, di cui 12 portatori della delezione (emizigoti per gli esoni 6 e 7) e 35 scelti fra le famiglie che contribuiscono maggiormente al picco di *linkage* in corrispondenza di questo gene. I risultati sono mostrati in Tabella 14. Non è stata identificata alcuna variante del DNA con un possibile effetto funzionale a livello delle regioni codificanti e di regolazione del gene *UPP2*.

**Tabella 14:** *screening* di mutazioni del gene *UPP2* in 47 soggetti con disturbo autistico

Gene (mRNA AC)	Posizione nel gene	posizione nel cDNA	Tipo di mutazione	Allele maggiore/ minore	dbSNP	Nr eterozigoti in 48 famiglie IMGSAC
<b>UPP2</b> (NM_173355)	intron 1	81+32	-	G/C	rs13414112	1
	intron 2	168-174	-	C/T	rs6739665	1
	intron 2	168-1	-	insA	rs11368509	1
	exon 3	181	5'UTR	T/C		1
	exon 4	337	Silente T49	C/A		1
	intron 7	855+48	-	G/C	rs10804389	12
	exon 9	1147	3'UTR	C/T		1
	exon 9	1268	3'UTR	C/G		1

### 7.7. Studi di associazione allelica nel gene UPP2

Per completare l'analisi del gene *UPP2* in relazione al disturbo autistico, i dati prodotti nell'esperimento Illumina sono stati utilizzati per effettuare studi di associazione. Il gene presenta una struttura con 5 principali blocchi di LD la cui variabilità è catturata da 18 tagSNPs. Come si può vedere in Tabella 15, valori positivi di associazione sono stati ottenuti per alcuni SNPs localizzati nel blocco 151, sia mediante TDT che mediante test caso-controllo. È interessante notare che 2 degli SNPs per i quali è stata trovata associazione con TDT ricadono all'interno della CNV individuata in questo studio. L'analisi degli SNPs *rs7585982* e *rs2074955* è stata ripetuta eliminando le famiglie in cui è stata individuata la CNV mediante il test PCR (Tabella 16). Il *p-value* rimane inalterato per lo SNP *rs2074955*, mentre aumenta per lo SNP *rs7585982* (da 0.006 a 0.012), pur rimanendo significativo.

**Tabella 15:** test TDT e caso-controllo per i18 tagSNPs analizzati: in grassetto sono indicati i 2 SNPs che cadono nella CNV individuata in questo studio.

BLOCCo di LD	nome del marcatore	TDTPHASE p	COCAPHASE P
147	rs717378	0,694851	0,455857
147	rs6745721	0,5523	0,452
148	rs6437126	0,8185	0,0998
148	rs1860164	0,3129	0,1705
148	rs4664920	0,4051	0,2211
149	rs2098044	0,7531	0,4382
149	rs6759860	0,1949	0,5139
150	rs1860165	0,409	0,4547
151	rs13020547	0,7531	0,6017
151	rs13383619	<b>0,0001897</b>	0,1487
151	rs6437129	0,2635	0,2835
151	rs6437133	<b>0,00426</b>	0,06959
151	rs6709528	<b>0,0007733</b>	0,07937
151	<b>rs7585982</b>	<b>0,01193</b>	0,1605
151	<b>rs2074955</b>	<b>0,03325</b>	0,8472
151	rs1468523	<b>0,0182</b>	0,1179
151	rs2074954	1	0,04229
151	rs1357184	<b>0,01142</b>	0,08431

**Tabella 16:** analisi dei 2 SNPs *rs7585982* e *rs2074955* effettuata includendo o escludendo le famiglie in cui è stata individuata la CNV individuata in questo studio.

<b>rs7585982</b>	<b>Allele</b>	<b>T</b>	<b>freq-T</b>	<b>NT</b>	<b>freq-NT</b>	<b>p</b>
	1	42	0.3717	71	0.6283	0.00608
	2	71	0.6283	42	0.3717	
	1	41	0.3796	67	0.6204	0.01193
	2	67	0.6204	41	0.3796	

<b>rs2074955</b>	<b>Allele</b>	<b>T</b>	<b>freq-T</b>	<b>NT</b>	<b>freq-NT</b>	<b>p</b>
	1	15	0.3409	29	0.6591	0.03325
	2	29	0.6591	15	0.3409	
	1	15	0.3409	29	0.6591	0.03325
	2	29	0.6591	15	0.3409	

## 8. *Discussione*

Gli studi epidemiologici condotti negli ultimi due decenni hanno evidenziato un fondamentale contributo dei fattori genetici nell'eziologia dell'autismo. Tali studi hanno inoltre indicato che la predisposizione genetica si estende ad un gruppo più ampio di disturbi sociali e di comunicazione, con caratteristiche simili all'autismo, ma presenti in forma più lieve (“*broader phenotype*”). Il modello di ereditarietà è tuttavia complesso: previsioni statistiche hanno suggerito il coinvolgimento di molteplici geni di suscettibilità ad azione epistatica con probabile presenza di eterogeneità.

Negli ultimi anni diversi gruppi di ricerca hanno effettuato analisi di *linkage* estese all'intero genoma e hanno individuato alcune regioni cromosomiche in cui è probabilmente localizzato un locus di suscettibilità per l'autismo. In particolare, l'IMGSAC (*International Molecular Genetic Study of Autism Consortium*) ha identificato una regione sul cromosoma 2q21-q33, con un MLS di 3.74 in corrispondenza del marcatore D2S2188 [IMGSAC, 2001], regione successivamente confermata in più studi indipendenti [Philippe et al., 1999; Buxbaum et al., 2001; Shao et al., 2002]. Questo è un risultato particolarmente incoraggiante, data la generale difficoltà nel replicare le evidenze di *linkage* nello studio delle malattie complesse. Il locus sul cromosoma 2 copre una regione genomica piuttosto ampia (~ 40 cM) in cui sono localizzati centinaia di geni noti, problema anche questo comune agli studi di *linkage* per le malattie complesse, e che complica il successivo passo di identificazione del gene o dei geni responsabili della suscettibilità conferita da un particolare locus. Inoltre, ancora oggi si sa poco sulla natura delle varianti genetiche alla base della predisposizione a patologie multifattoriali.

Per identificare varianti di rischio per l'autismo in corrispondenza della regione di *linkage* 2q è stato deciso di seguire nel presente studio due approcci complementari. Sono stati scelti i migliori geni candidati in base al loro ruolo nello sviluppo e funzionalità del sistema nervoso centrale e/o alla localizzazione all'interno di due delezioni mappate nella regione 2q24 in due individui affetti da autismo giunti alla nostra attenzione durante questi tre anni di dottorato. L'analisi dei geni è stata quindi condotta sia mediante ricerca diretta di mutazioni nelle regioni codificanti e di regolazione dell'espressione genica, sia mediante un approccio indiretto basato su studi di associazione allelica (*LD mapping*).

Grazie alle scoperte scientifiche inerenti la struttura a blocchi del genoma umano e la possibilità di catturare una gran parte della variabilità presente in una determinata regione utilizzando un numero ridotto di marcatori (htSNPs), alle recenti innovazioni tecnologiche che consentono di

effettuare analisi ad alta efficienza con costi e tempi più contenuti rispetto al passato, e alla disponibilità di risorse pubbliche basate su progetti importanti e innovativi quali l'*International HapMap Consortium*, è oggi possibile condurre studi di associazione più sistematici ed esaustivi. In particolare, in collaborazione con il gruppo di ricerca del Prof. Anthony Monaco presso il Wellcome Trust Center for Human Genetic di Oxford, e' stato condotto uno studio di associazione su larga scala utilizzando una tecnologia innovativa ed efficiente, quale il *Golden Gate Assay*<sup>TM</sup> di Illumina, I risultati ottenuti dall'esperimento Illumina sono stati utilizzati nel presente lavoro per affiancare l'analisi di mutazioni, e gli studi di associazione preliminari da noi condotti nei geni candidati sul locus 2q.

Tra tutti i geni candidati esaminati nella prima fase del presente lavoro, solo nel gene *INPPI* sono state identificate varianti di sequenza con evidenza di associazione al fenotipo autistico. Lo *screening* di mutazioni ha rilevato una maggiore frequenza di una rara sostituzione missenso nel campione di individui autistici analizzati, rispetto ad un campione di controllo. Il risultato non è stato confermato su un campione indipendente di individui affetti, inoltre l'iniziale associazione rilevata sul campione totale di famiglie *multiplex* dell'IMGSAC non è sufficiente a giustificare il valore di *linkage* sul cromosoma 2. Anche il polimorfismo C1276A nell'esone 6 del gene *INPPI* ha fornito un'evidenza di associazione con l'allele più frequente, non replicata nel campione di famiglie indipendenti. Analogamente, un altro studio ha riportato associazione tra due SNPs nel gene *INPPI* e l'autismo [Serajee et al., 2003]; gli stessi SNPs non sono però risultati significativi nel nostro campione. Infine, l'analisi del gene *INPPI*, nell'esperimento Illumina, non ha evidenziato la presenza di alleli o aplotipi di rischio, suggerendo che varianti nel gene *INPPI* non siano coinvolte nella suscettibilità al disturbo autistico.

Le scarse evidenze neurobiologiche disponibili, la mancata concordanza tra i valori di *linkage* ottenuti nei diversi *genome-scan* e i risultati contraddittori dei geni candidati e degli studi di associazione ottenuti da diversi gruppi di ricerca indipendenti suggeriscono che ,probabilmente, molti loci di suscettibilità contribuiscono alla predisposizione genetica al fenotipo autistico. Inoltre, evidenze scientifiche suggeriscono che il 7% dei casi di individui affetti da autismo è determinato dalla presenza di anomalie cromosomiche visibili citogeneticamente (The Autism Chromosome Rearrangement Database, <http://projects.tcag.ca/autism/>). Sulla base di queste osservazioni, nella seconda parte del lavoro i dati dell'esperimento Illumina sono stati utilizzati per l'identificazione e caratterizzazione di eventuali varianti strutturali submicroscopiche, o *Copy Number Variants* (CNVs) presenti nel campione di famiglie con autismo. Le CNVs sono caratterizzate dalla variazione del numero di copie (delezioni, duplicazioni) di segmenti genomici

le cui dimensioni sono comprese fra 1Kb e 1-2Mb circa [Buckley et al. 2005]. A causa delle loro dimensioni “intermedie” le CNVs non sono identificabili ne’ con le classiche tecniche di citogenetica (che permettono di rilevare varianti cromosomiche di dimensioni  $\geq 3$ Mb), ne’ tramite sequenziamento o tecniche molecolari standard (che generalmente permettono di rilevare solamente varianti del DNA di piccole dimensioni). Recentissimi studi, condotti sull’intero genoma, hanno rilevato l’elevata frequenza di queste CNVs nella popolazione generale, evidenziando come tali varianti strutturali submicroscopiche possano, secondo molteplici meccanismi, dare un contributo non solo alla normale variabilità fenotipica, ma anche alla suscettibilità ad alcune malattie, sia mendeliane che complesse, quali per esempio il ritardo mentale non sindromico [Buckley et al. 2005]. Ad esempio, una CNV può influenzare direttamente il dosaggio di un gene oppure, indirettamente, alterarne l’espressione genica per effetto di posizione in quanto si trova in una regione regolatoria. Inoltre, una variante strutturale, pur essendo ininfluenza dal punto di vista dell’espressione genica, potrebbe predisporre il locus in cui risiede ad ulteriori riarrangiamenti cromosomici che invece causano un particolare fenotipo. Per quanto riguarda lo studio dell’eziologia delle malattie complesse, il mappaggio e la caratterizzazione di questa nuova tipologia di varianti genetiche, sia nella popolazione normale che in individui affetti da patologie, potrebbe rivelarsi una nuova fonte di informazione per l’individuazione di varianti di suscettibilità.

Recenti studi hanno evidenziato la possibilità di rilevare in modo efficiente le CNVs del tipo “delezioni” a partire dai dati ottenuti dall’analisi su larga scala di SNPs in accordo con quanto proposto nel presente lavoro.

Dei 34 SNPs che fornivano dati anomali di ereditarietà nell’esperimento Illumina, solamente 7 erano compatibili con delle delezioni e pertanto sono stati ulteriormente analizzati. L’analisi di sequenza nelle famiglie in cui sono stati osservati errori di eredità mendeliana ha evidenziato per 6 dei 7 SNPs studiati un pattern di ereditarietà mendeliano. E’ stato verificato che gli errori di ereditarietà precedentemente osservati erano causati da varianti non note in prossimità o in corrispondenza dello SNPs stesso. Di fatto, la presenza di tali varianti impediva l’ibridazione al template delle sonde appositamente progettate sulla base della sequenza nota che comprende lo SNP di interesse. Per lo SNP *rs7585982* la sequenza ha confermato la possibile presenza di una delezione, che pertanto è stata in seguito caratterizzata, portando all’identificazione di una CNV di circa 5,9 Kb che comprende interamente due esoni codificanti (esone 6 ed esone 7) del gene *UPP2*. A livello del messaggero questa delezione determina un *frameshift* e l’introduzione di uno *stop codon*, con probabile perdita di funzione del prodotto proteico. Il gene *UPP2* codifica per

una uridina fosforilasi necessaria per la conversione reversibile dell'uridina in uracile, pertanto la traduzione di questo allele alternativo di *UPP2*, deleto in due esoni, potrebbe avere effetti sull'attività e sull'efficienza del *pathway* metabolico di questo enzima, fondamentale per l'omeostasi a livello plasmatico dell'uridina. L'uridina è coinvolta in molteplici funzioni biologiche fra le quali, di particolare interesse per il nostro gruppo di ricerca, è il coinvolgimento dell'uridina nella regolazione del sonno e il suo utilizzo per il trattamento dell'epilessia. Infatti, il 33% dei casi di autismo sono associati ad epilessia ed inoltre è stata riportata una maggiore frequenza di disturbi del sonno in individui affetti da autismo. In alcuni casi di autismo associato ad epilessia, la somministrazione orale di uridina determina miglioramenti della capacità linguistica, del comportamento e della diminuzione delle frequenza di crisi epilettiche epilettiche [Page et al. 1997, Cao et al., 2005].

L'analisi di associazione mediante test caso-controllo e TDT non ha evidenziato associazione tra la CNV individuata nel gene *UPP2* e il disturbo autistico. Questa delezione rappresenta probabilmente un polimorfismo presente all'interno della popolazione europea, oppure potrebbe essere implicata nella determinazione di un fenotipo diverso dal disturbo autistico. Lo *screening* di mutazione del gene *UPP2* non ha evidenziato altre varianti codificanti con un possibile effetto sulla funzionalità proteica. I dati di associazione, invece, hanno mostrato valori di *p-value* significativi per polimorfismi localizzati in un singolo blocco di LD in corrispondenza della regione al 3' del gene. Questo blocco comprende gli esoni 6 e 7 e attualmente è in corso un esperimento per replicare i risultati ottenuti. Tuttavia, è da notare che i valori di associazione riscontrati non rimangono significativi dopo la correzione effettuata per tener conto dell'elevato numero di test condotti per analizzare i 1500 SNPs tipizzati nel corso dell'esperimento Illumina.

Da un punto di vista metodologico, è interessante notare come l'eliminazione delle famiglie in cui è stata riscontrata la presenza della delezione nella fase di analisi dei dati, ha abbassato l'evidenza di associazione ottenuta per uno dei due SNPs localizzati all'interno della delezione stessa. Questo sottolinea l'importanza delle analisi svolte anche nell'ottica della correzione di tutti i possibili errori presenti nei dati prima di procedere alle elaborazioni statistiche.

Il metodo utilizzato in questo lavoro per l'individuazione di CNVs si basa esclusivamente sull'analisi degli errori mendeliani dei dati di *genotyping* prodotti da esperimenti con tecnologia Illumina. Questo non è sufficiente per ottenere un'efficienza di rilevazione delle CNVs paragonabile a quella delle tecnologie più specifiche, come ad esempio *array-CGH* di ultima generazione, appositamente progettate per l'individuazione a livello genomico di questo tipo di varianti strutturali. Tuttavia, il *software* di ultima generazione sviluppato dall'Illumina permette

di analizzare l'intensità delle fluorescenze prodotte da un esperimento di *genotyping*, consentendo di rilevare in modo più efficiente la presenza di CNVs, sia per quanto riguarda delezioni che duplicazioni. Sarà quindi interessante rianalizzare i dati ottenuti utilizzando il nuovo *software*.

L'individuazione di questa CNV nella regione sul cromosoma 2, in corrispondenza del gene *UPP2*, risulta di notevole importanza anche alla luce di un progetto più grande, *l'Autism Genome Project (AGP) Consortium*, che ha visto coinvolti cinquanta diversi gruppi di ricerca nel mondo e al quale il mio gruppo di ricerca, diretto dalla Prof.ssa Elena Mestrini, ha preso parte come IMGSAC.

Su un campione totale di 1168 famiglie *multiplex* con autismo provenienti da Canada, Stati Uniti ed Europa è stata condotta la più ampia analisi di *linkage* mai effettuata prima. Quest'analisi, condotta utilizzando la tecnologia Affimetrix (Affimetrix GeneChip®Mapping 10K 2.0 SNParray), ha permesso di identificare la presenza di un inaspettato numero di CNVs precedentemente mai individuate con le classiche tecniche citogenetiche. In particolare, in una coppia di fratelli affetti è stata trovata una delezione *de novo* in corrispondenza del gene della Neurexina (*NRXN1*) (Cap.2, §2.5), osservazione che, da sola, potrebbe essere poco significativa dato l'elevato numero di CNVs presenti nel genoma umano. Tuttavia, si sa che questo gene codifica per una proteina che interagisce con le Neuroligine, proteine implicate nel differenziamento pre-sinaptico negli assoni glutamatergici [Scheiffele et al., 2000] e precedentemente trovate mutate in alcuni pazienti affetti da ASD [Jamain et al., 2003]. È probabile che il legame "neurexine-neuroligine" possa avviare una risposta molecolare pre-sinaptica che porta alla riorganizzazione di tutto il compartimento pre-sinaptico [Graf et al. 2004; Prange et al. 2004] e che quindi varianti in questi geni, o in altri geni con i quali interagiscono, possano determinare alterazioni del *pathway* del glutamato e, di conseguenza, avere un ruolo cruciale nella suscettibilità al disturbo autistico.

Capire il ruolo di queste varianti strutturali nella suscettibilità al disturbo autistico rappresenta una chiave di svolta fondamentale nella comprensione della natura delle disfunzioni cerebrali negli individui autistici e nell'interpretazione delle relazioni esistenti tra le varianti di rischio e differenti aspetti del comportamento autistico. Inoltre, queste varianti possono rappresentare dei bersagli importanti per lo sviluppo di approcci terapeutici e di prevenzione, per il miglioramento dei metodi diagnostici e per l'identificazione di eventuali fattori di rischio dovuti all'ambiente permettendo in questo modo di condurre ulteriori studi di tipo genetico-epidemiologico.



## 9. Bibliografia

Abdolmaleky HM, Smith CL, Faraone SV, Shafa R, Stone W, Glatt SJ, Tsuang MT. Methylomics in psychiatry: Modulation of gene-environment interactions may be through DNA methylation. *Am J Med Genet B Neuropsychiatr Genet.* 2004; 127: 51-9.

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

Allen G, Genetics and the prevention of mental subnormality. *Proc Annu Meet Am Psychopathol Assoc.* 1967; 56:202-21.

Altug-Teber O. et al. A rapid microarray based whole genome analysis for detection of uniparental disomy. *Hum. Mutat.* 2005; 26: 153–159

American Psychiatric Association. Diagnostic and statistical manual of Mental Disorders, DSM-IV-TR. Washington DC, 2000.

American Psychiatric Association. Diagnostic and statistical manual of Mental Disorders. American Psychiatric Association: Washington, DC, 1994.

Anderlid BM, Schoumans J, Anneren G, Tapia-Paez I, Dumanski J, Blennow E, Nordenskjold M. FISH-mapping of a 100-kb terminal 22q13 deletion. *Hum. Genet.* 2002; 110: 439-443,.

Anderson GM, Minderaa RB, Cho SC, Volkmar FR, Cohen DJ. The issue of hyperserotonemia and platelet serotonin exposure: a preliminary study.

Anderson LT, Campbell M, Adams P, Small AM, Perry R, Shell J. The effects of haloperidol on discrimination learning and behavioral symptoms in autistic children. *J Autism Dev Disord.* 1989; 19: 227-39.

Armour J. A., Sismani C., Patsalis P. C., Cross G., Measurement of locus copy number by hybridisation with amplifiable probes. *Nucleic Acids Res.* 2000; 28: 605–609.

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

Asperger H. Die ‘Autistischen Psychopathen’ im Kindesalter. *Arch Psychiat Nerven* 1944; 117: 73–136.

Asperger, Die autistischen psychopathen im Kindesalter. *Archiv fur Psychiatrie und Nervenkrankheiten*, 1944; 117: 76–136.

August GJ, Stewart MA, Tsai L. The incidence of cognitive disabilities in the siblings of autistic children. *Br J Psychiatry* 1981; 138: 416–422.

Auranen M, Varilo T, Alen R, Vanhala R, Ayers K, Kempas E, Ylisaukko-Oja T, Peltonen L, Jarvela I. Evidence for allelic association on chromosome 3q25-27 in families with autism spectrum disorders originating from a subisolate of Finland. *Mol Psychiatry.* 2003;8: 879-84.

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

Bacolla A., Jaworski A., Larson J.E., Jakupciak J.P., Chuzhanova N., et al., Breakpoints of gross deletions coincide with non-B DNA conformations. *Proc Natl Acad Sci U S A* , 2004; 101: 14162–14167.

Bailey A, Phillips W, Rutter M. Autism: towards an integration of clinical, genetic, neuropsychological, and neurobiological perspectives. *J Child Psychol Psychiatry*. 1996; 37:89-126.

Bailey AJ. The biology of autism. *Psychol Med* 1993; 23: 7–11.

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

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

Bailey, Rutter, Philips, Autism: the phenotype in relatives. *J Aut Dev Dis*, 1996; 28:439-445.

Barnette T, Gourraud PA, Cambon-Thomsen A, Strategies in analysis of the genetic component of multifactorial diseases; biostatistical aspects. *Transpl Immunol*. 2005; 14(3-4): 255-66.

Baron-Cohen S, Knickmeyer RC, Belmonte MK. Sex differences in the brain: implications for explaining autism. *Science* 2005; 310: 819–823.

Barrett S, Beck JC, Bernier R, Bisson E, Braun TA, Casavant TL et al. An autosomal genomic screen for autism. *Am J Med Genet* 1999; 88: 609–615.

Benzi G., Villa R.F., Dossena M., Vercesi L., Gorini A., Pastoris O., Cerebral endogenous substrate utilization during the recovery period after profound hypoglycemia. *J. Neurosci. Res*. 1984; 11: 437– 450.

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

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

Bruce SK et al. Global analysis of uniparental disomy using high-density genotyping arrays. *J. Med. Genet*. 2005; 42: 847–851

Buckley P. G., Mantripragada K. K., Piotrowski A., Diaz de Stahl T., Dumanski J.P., Copy-number polymorphisms: mining the tip of an iceberg. *Trends in Genetics* , 2005; vol.21, N°6 .

Buoni et al., The syndrome of inv dup (15): clinical, electroencephalographic, and imaging findings. *J. Child. Neurol*. 2000; 15: 380-385.

Burwinkel B., Kilimann M.W., Unequal homologous recombination between LINE-1 elements as a mutational mechanism in human genetic disease. *J. Mol. Biol*. 1998; 277: 513–517.

Buxbaum JD, Silverman J, Keddache M, Smith CJ, Hollander E, Ramoz N et al. Linkage analysis for autism in a subset families with obsessive-compulsive behaviors: evidence for an autism susceptibility gene on chromosome 1 and further support for susceptibility genes on chromosome 6 and 19. *Mol Psychiatry* 2004; 9: 144–150. 150–159.

Buxbaum JD, Silverman JM, Smith CJ, Kilifarski M, Reichert J, Hollander E et al. Evidence for a susceptibility gene for autism on chromosome 2 and for genetic heterogeneity. *Am J Hum Genet* 2001; 68: 1514–1520.

Campbell DB, Sutcliffe JS, Ebert PJ, Militeri R, Bravaccio C, Trillo S, Elia M, Schneider C, Melmed R, Sacco R, Persico AM, Levitt P. A genetic variant that disrupts MET transcription is associated with autism. *Proc Natl Acad Sci U S A*. 2006; 103: 16834-9.

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

Casas et al., Four new patients with 2q terminal deletion syndrome: hemidiaphragmatic hernia as a novel findings and patterns of malformation among 47 reported cases. *Am. J. Hum. Genet.*, 2002; 71(4): A569.

Chakrabarti S., Fombonne E. Pervasive developmental disorders in preschool children, *Jama* 2001; 285:3093-9.

Chance PF, Lupski JR. Inherited neuropathies: Charcot-Marie-Tooth disease and related disorders. *Baillieres Clin Neurol*. 1994; 3(2):373-85.

Charbonnier F. et al., Detection of exon deletions and duplications of the mismatch repair genes in hereditary nonpolyposis colorectal cancer families using multiplex polymerase chain reaction of short fluorescent fragments. *Cancer Res*. 2000; 60: 2760–2763.

Connolly G.P., Simmonds H.A., Duley J.A., Pyrimidines and CNS Regulation. *Trends Pharmacol. Sci.* , 1996; 17: 106– 107.

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

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

Cook et al., Trasmision disequilibrium mapping in the serotonin transporter gene (SLC6A4) region in autistic disorder. Presentato dall' American Society of Human Genetics, San Diego, CA. 2001

Dawson E, Abecasis GR, Bumpstead S, Chen Y, Hunt S, Beare DM, Pabial J, Dibling T, Tinsley E, Kirby S, Carter D, Papaspyridonos M, Livingstone S, Ganske R, Lohmussaar E, Zernant J, Tonisson N, Remm M, Magi R, Puurand T, Vilo J, Kurg A, Rice K, Deloukas P, Mott R, Metspalu A, Bentley DR, Cardon LR, Dunham I. A first-generation linkage disequilibrium map of human chromosome 22. *Nature*, 2002; 418: 544–548.

Delong GR, Dwyer JT. Correlation of family history with specific autistic subgroups: Aspergers and bipolar affective disease. *J Autism Dev Disord* 1988; 18: 593–600.

Diagnostic and Statistical Cores of the UW Autism International Molecular Genetic Study of Autism Consortium. A genomewide screen for autism: strong evidence for linkage to chromosomes 2q, 7q, and 16p. *Am J Hum Genet* 2001; 69: 570–581.

Dudbridge F, Pedigree disequilibrium tests for multilocus haplotypes. *Genet Epidemiol* 2003; 25:115-221.

Durand CM, Betancur C, Boeckers TM, Bockmann J, Chaste P, Fauchereau F, Nygren G, Rastam M, Gillberg IC, Anckarsater H, Sponheim E, Goubran-Botros H, and 11 others.

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

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

Elston, The genetic dissection of multifactorial traits. *Clin Exp Allergy* , 1995; 25:103– 6.

Fan et al. Highly parallel SNP genotyping. *Cold Spring Harbor Symposia on Quantitative Biology*, 2003; vol. LXVIII: 69-78.

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

Feng, J. et al. High frequency of neurexin 1b signal peptide structural variants in patients with autism. *Neurosci.* 2006; 409, 10–13.

Feuk L., Carson A. R., and Scherer S. W., Structural variation in the human genome. *Nature Genetics*, February 2006; 7: 85-97.

Flint J., Wilkie M.A.O., Bucke V.J., Winter R.M., Holland A.J., McDermid H.E., The detection of sub-telomeric chromosomal rearrangements in idiopathic mental retardation. *Nat. Genet.*, 1995; 9: 132–140.

Folstein S, Rutter M. Infantile autism: a genetic study of 21 twin pairs. *J Child Psychol Psychiatry* 1977; 18: 297–321.

Folstein SE, Rosen-Sheidley B. Genetics of autism: complex aetiology for a heterogeneous disorder. *Nat Rev Genet* 2001; 2:943–55.

Folstein SE, Santangelo SL, Gilman SE, Piven J, Landa R, Lainhart J et al. Predictors of cognitive test patterns in autism families. *J Child Psychol Psychiatry* 1999; 40: 1117–1128.

Fombonne E, Modern views of autism. *Can J Psychiatry.* 2003; 48: 503-5.

Fombonne E. Epidemiological surveys of autism and other pervasive developmental disorders: an update. *J Autism Dev Disord* 2003; 33: 365–382.

Gabriel SB, Schaffner SF, Nguyen H, Moore JM, Roy J, Blumenstiel B, Higgins J, DeFelice M, Lochner A, Faggart M, Liu-Cordero SN, Rotimi C, Adeyemo A, Cooper R, Ward R, Lander ES, Daly MJ, Altshuler D. The structure of haplotype blocks in the human genome. *Science* 2002; 296: 2225-9.

Gauthier J, Bonnel A, St-Onge J, Karemera L, Laurent S, Mottron L, Fombonne E, Joober R, Rouleau GA. NLGN3/NLGN4 gene mutations are not responsible for autism in the Quebec population. *Am J Med Genet B Neuropsychiatr Genet.* 2005; 132: 74-5.

Geiger A., Yamasaki S., Cytidine and uridine requirement of the brain. *J. Neurochem.*, 1956; 1: 93– 100.

Ghaziuddin et al., Deletion of chromosome 2q37 and autism: a distinct subtype. *J. Autism Dev. Disord.*, 1999; 29: 259-263.

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

Gonzalez E., Kulkarni H., Bolivar H., Mangano A., Sanchez R., et al., The influence of CCL3L1 gene-containing segmental duplications on HIV-1/AIDS susceptibility. *Science*, 2005; 307: 1434–1440.

Gordon CT, Rapoport JL, Hamburger SD, State RC, Mannheim GB. Differential response of seven subjects with autistic disorder to clomipramine and desipramine. *Am J Psychiatry.* 1992; 149: 363-6.

Gordon CT, State RC, Nelson JE, Hamburger SD, Rapoport JL. A double-blind comparison of clomipramine, desipramine, and placebo in the treatment of autistic disorder. *Arch Gen Psychiatry.* 1993; 50: 441-7.

Graf ER, Zhang X, Jin SX, Linhoff MW & Craig AM. Neurexins induced differentiation of GABA and glutamate postsynaptic specializations via neuroligins. *Cell.* 2004; 119: 1013–1026.

Graf ER, Zhang X, Jin SX, Linhoff MW, Craig AM. Neurexins induce differentiation of GABA and glutamneurologins. *Cell* 2004; 119:1013-26.

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

Hinds D.A., Kloek A. P., Jen M., Chen X., Frazer K.A., Common deletions and SNPs are in linkage disequilibrium in the human genome. *Nature Genetics*, February 200; vol. 38, num. 1: 82-85.

Hyer, Julier, Buckley, Trucco, Rotter, Spielman, et al. High-resolution linkage mapping for susceptibility genes in human polygenic disease: insulin-dependent diabetes mellitus and chromosome 11q. *Am J Hum Genet* , 1991; 48(2):243–57.

Iafraite AJ, Feuk L, Rivera MN, Listewnik ML, Donahoe PK, Qi Y, Scherer SW, Lee C. Detection of large-scale variation in the human genome. *Nat Genet* 2004; 36:949–951.

IMGSAC (International Molecular Genetic Study of Autism Consortium), A full genome screen for autism with evidence for linkage to a region on chromosome 7q. *Hum. Mol. Genet.*, 1998; 7: 571–578.

IMGSAC (International Molecular Genetic Study of Autism Consortium), A genomewide screen for autism: strong evidence for linkage to chromosomes 2q, 7q, and 16p. *Am J Hum Genet.* 2001; 69: 570-81.

Inoue K, Lupski JR. Molecular mechanisms for genomic disorders. *Annu Rev Genomics Hum Genet.* 2002; 3:199-242. *J Autism Dev Disord.* 1989; 19: 349-51.

Jacobs PA, Browne C, Gregson N, Joyce C, White H, Estimates of the frequency of chromosome abnormalities detectable in unselected newborns using moderate levels of banding. *J. Med. Genet.* 1992; 29: 103–108.

Jamain S, Quach H, Betancur C, Rastam M, Colineaux C, Gillberg IC, Soderstrom H, Giros B, Leboyer M, Gillberg C, Bourgeron T; Paris Autism Research International Sibpair Study. Mutations of the X-linked genes encoding neuroligins NLGN3 and NLGN4 are associated with autism. *Nat Genet.* 2003; 34: 27-9.

Jambaque I, Chiron C, Dumas C, Mumford J, Dulac O. Mental and behavioural outcome of infantile epilepsy treated by vigabatrin in tuberous sclerosis patients. *Epilepsy Res.* 2000; 38: 151-60.

Jaselskis CA, Cook EH Jr, Fletcher KE, Leventhal BL. Clonidine treatment of hyperactive and impulsive children with autistic disorder. *J Clin Psychopharmacol.* 1992; 12: 322-7.

Johansson M., Identification of a novel human Uridine Phosphorylase. *Biochemical and Biophysical Research Communications*, 2003; 307: 41–46.

Johnson, et al. Haplotype tagging for the identification of common disease genes. *Nature Genet.*, 2001; 29: 233–237.

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

Karkishchenko N.N., Makliakov I.S., Stradomskii B.V., Pyrimidine derivatives: their psychotropic properties and the molecular mechanisms of their central action. *Farmakol. Toksikol.*, 1990; 53: 67–72.

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

Kimura T., Kuze J., Watanabe K., Kondo S., Ho I.K., Yamamoto I., N3-phenacyluridine, a novel hypnotic compound, interacts with the benzodiazepine receptor. *Eur. J. Pharmacol.* , 1996; 311: 265–269.

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

Klauck SM, Munstermann E, Bieber-Martig B, Ruhl D, Lisch S, Schmotzer G, Poustka A, Poustka F. Molecular genetic analysis of the FMR-1 gene in a large collection of autistic patients. *Hum Genet.* 1997; 100: 224-9.

Knight AJ, Regan R, Nicod A, Horsley SW, Kearny L, Homfray T, Winter RM, Bolton P, Flint J, Subtle chromosomal rearrangements in children with unexplained mental retardation. *Lancet* 1999; 354: 1676–1681.

Kolomietz E, Meyn MS, Pandita A, Squire JA, The role of Alu repeat clusters as mediators of recurrent chromosomal aberrations in tumors. *Genes, Chromosomes Cancer* 2002; 35: 97–112.

Kruglyak, Daly, Reeve-Daly, Lander, Parametric and nonparametric linkage analysis: a unified multipoint approach. *Am J Hum Genet* , 1996; 58(6):1347– 63.

- Kruglyak, Nickerson, Variation is the spice of life. *Nature Genet.*, 2001; 27: 234–236.
- Lamb JA, Barnby G, Bonora E, Sykes N, Bacchelli E, Blasi F et al. Analysis of IMGSAC autism susceptibility loci: evidence for sex limited and parent of origin specific effects. *J Med Genet* 2005; 42: 132–137.
- Lamb JA, Parr JR, Bailey AJ, Monaco AP. Autism: in search of susceptibility genes. *Neuromolecular Med* 2002; 2:11–28.5.
- Lander ES, Schork NJ, Genetic dissection of complex traits. *Science*. 1994; 265: 2037-48.
- Laumonier F, Bonnet-Brilhault F, Gomot M, Blanc R, David A, Moizard MP, Raynaud M, Ronce N, Lemonnier E, Calvas P, Laudier B, Chelly J, Fryns JP, Ropers HH, Hamel BC, Andres C, Barthelemy C, Moraine C, Briault S. X-linked mental retardation and autism are associated with a mutation in the NLGN4 gene, a member of the neuroligin family. *Am J Hum Genet*. 2004; 74: 552-7.
- Le Couteur A, Rutter M, Lord C, Rios P, Robertson S, Holdgrafer M et al. Autism diagnostic interview: a standardized investigatorbased instrument. *J Autism Dev Disord* 1989; 19: 363–387.
- Leer JC, Hammer-Jespersen K, Schwartz M, *Eur. J. Biochem.* , 1977; 75: 217-224.
- Leyva A., Van Groeningen C.J., Kraal I., Peters G.J., Lankelma J., Pinedo H.M., Phase I and pharmacokinetic studies of high-dose uridine intended for rescue from 5-fluorouracil toxicity. *Cancer Res.* , 1984; 44: 5928– 5933.
- Lindblad-Toh K, Winchester E, Daly MJ, Wang DG, Hirschhorn JN, Laviolette JP, Ardlie K, Reich DE, Robinson E, Sklar P, Shah N, Thomas D, Fan JB, Gingeras T, Warrington J, Patil N, Hudson TJ, Lander ES. Large-scale discovery and genotyping of single-nucleotide polymorphisms in the mouse. *Nat Genet*. 2000; 24: 381-6.
- Liskay R.M., Letsou A., Stachelek J.L., Homology requirement for efficient gene conversion between duplicated chromosomal sequences in mammalian cells. *Genetics* 1987; 115: 161–167.
- Liu JJ, Nyholt DR, Magnussen P, Parano E, Pavone P, Geschwind D et al. A genomewide screen for autism susceptibility loci. *Am J Hum Genet* 2001; 69: 327–340.
- Lord et al. The autism diagnostic observation schedule generic: a standard measure of social and communication deficits associated with the spectrum of autism. *J Autism Dev. Disord* .2000; 30: 205-223.
- Lord, Rutter, LeCouteur, Autism Diagnostic Interview-Revised: a revised version of a diagnostic interview for caregivers of individuals with possible pervasive developmental disorders. *J. Autism Dev. Disord*. 1994; 24: 659-685.
- Lord, Schopler, Revicki, Sex difference in autism. *Journal of Autism and developmental disorders*, 1982; 12: 317-330.
- Lucito R., et al., Representational oligonucleotide microarray analysis: a high-resolution method to detect genome copy number variation. *Genome Res*. 2003; 13: 2291–2305.
- Lupski JR, Genomic disorders: Structural features of the genome can lead to DNA rearrangements and human disease traits. *Trends Genet*. 1998; 14: 417–422.

Lupski JR, Chance PF, Garc. Inherited primary peripheral neuropathies. Molecular genetics and clinical implications of CMT1A and HNPP. *JAMA* 1993; 270: 2326-30.

Lupski JR, Hotspots of homologous recombination in the human genome: Not all homologous sequences are equal. *Genome Biol.* 2004; 5: 242.

Lupski JR, Shaffer LG, Molecular mechanisms for constitutional chromosomal rearrangements in humans. *Ann. Rev. Genet.* 2000; 34: 297–329.

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

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

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

Maestrini E, Lai C, Marlow A, Matthews N, Wallace S, Bailey A, Cook EH, Weeks DE, Monaco AP, and IMGSAC. Serotonin transporter (5-HTT) and  $\gamma$ -aminobutyric acid receptor subunit  $\beta$ -3 (GABRB3) gene polymorphisms are not associated with autism in the IMGSAC families. *Am. J. Med. Genet.*, 1999; 88: 492-496.

Maestrini E, Paul A, and Bailey AP. Identifying Autism Susceptibility Genes. *Neuron* 2000; 28, 19–24.

Maria BL, Deidrick KM, Roach ES, Gutmann DH. Tuberous sclerosis complex: pathogenesis, diagnosis, strategies, therapies, and future research directions. *J Child Neurol.* 2004;19:632-42.

McCarroll S et al., Common deletion polymorphisms in the human genome. *Nature Genetics*, February 2006; vol. 38, num. 1: 86- 92.

McCauley JL, Li C, Jiang L, Olson LM, Crockett G, Gainer K et al. Genome-wide and Ordered-Subset linkage analyses provide support for autism loci on 17q and 19p with evidence of phenotypic and interlocus genetic correlates. *BMC Med Genet* 2005; 6: 1.

McDougle CJ, Epperson CN, Price LH. Obsessive-compulsive symptoms with-neuroleptics. *J Am Acad Child Adolesc Psychiatry.* 1996; 35: 837-8.

Mei R, et al. Genome-wide detection of allelic imbalance using human SNPs and high-density DNA arrays. *Genome Res.* 2000; 10, 1126–1137.

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

Morton, Sequential tests for the detection of linkage. *Am J Hum Genet* , 1955;7(3):277 –318.

Olson JM, Witte JS, Elston RC. Tutorial in biostatistics. Genetic mapping of complex traits. *Stat Med* 1999; 18: 2961-81.

Ott J, Bhat A, Linkage analysis in heterogeneous and complex traits. *Eur Child Adolesc Psychiatry.* 1999; 3: 43-6.

- Paabo S, The mosaic that is our genome. *Nature*. 2003 Jan 23;421(6921):409-12.
- Page T., Yu A., Fontanesi J., Nyhan W.L., Developmental disorder associated with increased cellular nucleotidase activity. *Proc. Natl. Acad. Sci.* , 1997; 94: 11601– 11606.
- Patil N, Berno AJ, Hinds DA, Barrett WA, Doshi JM, Hacker CR, Kautzer CR, Lee DH, Marjoribanks C, McDonough DP, Nguyen BT, Norris MC, Sheehan JB, Shen N, Stern D, Stokowski RP, Thomas DJ, Trulson MO, Vyas KR, Frazer KA, Fodor SP. Blocks of limited haplotype diversity revealed by high-resolution scanning of human chromosome 21. *Science* 2001; 294: 1719–1723.
- Peltola H, Patja A, Leinikki P, Valle M, Davidkin I, Paunio M. No evidence for measles, mumps, and rubella vaccine-associated inflammatory bowel disease or autism in a 14-year prospective study. *Lancet*. 1998; 351: 1327-8.
- Penrose, The general purpose sib-pair linkage test. *Ann Eugen* , 1953; 18(2):120–4.
- Persico AM, D'Agruma L, Malorano N, Totaro A, Militerni R, Bravaccio C et al. Reelin gene alleles and haplotypes as a factor predisposing to autistic disorder. *Mol Psychiatry* 2001; 6: 150–159.
- Persico AM, Militerni R, Bravaccio C, Schneider C, Melmed R, Conciatori M, Damiani V, Baldi A, Keller F. Lack of association between serotonin transporter gene promoter variants and autistic disorder in two ethnically distinct samples. *Am J Med Genet*. 2000;96:123-7.
- Philippe A, Martinez M, Guilloudbataille M, Gillberg C, Rastam M, Sponheim E et al. Genome-wide scan for autism susceptibility genes. *Hum Mol Genet* 1999; 8: 805–812.
- Pickles A, Bolton P, Macdonald H, Bailey A, Le Couteur A, Sim CH, Rutter M. Latent-class analysis of recurrence risks for complex phenotypes with selection and measurement error: a twin and family history study of autism. *Am J Hum Genet*. 1995; 57: 717-26.
- Pinkel D. Et al., High resolution analysis of DNA copy number variation using comparative genomic hybridization to microarrays. *Nature Genet*. 1998; 20: 207–211.
- Pizzorno G., Deliang C., Leffert J.J., Russell R.L., Zhang D., Handschumacher R.E., Homeostatic control of uridine and the role of uridine phosphorylase: a biological and clinical update. *Biochim. Biophys. Acta* 1587, 2002; 133–144.
- Ponchel, F. et al. Real-time PCR based on SYBRGreen I fluorescence: An alternative to the TaqMan assay for a relative quantification of gene rearrangements, gene amplifications and micro gene deletions. *BMC Biotechnol* 2003; 3: 18.
- Prange O, Wong TP, Gerrow K, Wang YT, El-Husseini A. A balance between excitatory and inhibitory synapses is controlled by PSD-95 and neuroligin. *Proc Natl Acad Sci* 2004; 101:13915-20.
- Raghavan M. et al. Genome-wide single nucleotide polymorphism analysis reveals frequent partial uniparental disomy due to somatic recombination in acute myeloid leukemias. *Cancer Res*. 2005; 65: 375–378.
- Reddy KS. Cytogenetic abnormalities and fragile-X syndrome in Autism Spectrum Disorder. *BMC Med Genet*. 2005; 6: 3.
- Reich DE, Lander ES. On the allelic spectrum of human disease. *Trends genet*. 2001; 17: 502-10.

- Reich, et al. Linkage disequilibrium in the human genome. *Nature*, 2001; 411: 199–204.
- Reich, Gabriel, Altshuler, Quality and completeness of SNP databases. *Nature Genet.*, 2003; 33: 457–458.
- Remington G, Sloman L, Konstantareas M, Parker K, Gow R. Clomipramine versus haloperidol in the treatment of autistic disorder: a double-blind, placebo-controlled, crossover study. *J Clin Psychopharmacol.* 2001; 21: 440-4.
- Rineer et al., Autistic symptoms among children and young adults with isodicentric chromosome 15. *Am. J. Med. Genet.* 1998; 81: 428-433.
- Risch N, Linkage strategies for genetically complex traits. I. Multilocus models. *Am J Hum Genet.* 1990; 46(2):229-41.
- Risch N, Linkage strategies for genetically complex traits. II. The power of affected relative pairs. *Am J Hum Genet.* 1990; 46: 222-8.
- Risch N, Spiker D, Lotspeich L, Nouri N, Hinds D, Hallmayer J et al. A genomic screen of autism: evidence for a multilocus etiology. *Am J Hum Genet* 1999; 65: 493–507.
- Ronquist G., Niklasson F., Uridine, xanthine, and urate contents in human seminal plasma. *Arch. Androl.* , 1984; 13: 63– 70.
- Ronquist G., Stegmayr B., Niklasson F., Sperm motility and interactions among seminal uridine, xanthine, urate, and ATPase in fertile and infertile men. *Arch. Androl.* , 1985; 15: 21– 27.
- Rutter M. Concepts of autism: a review of research. *J Child Psychol Psychiatry* 1968; 9: 1–25.
- Rutter M. Genetic Studies of Autism: From the 1970s into the Millennium. *Journal of Abnormal Child Psychology.* 2000; 28: 3-14.
- Sankaranarayanan K, Wassom JS, Ionizing radiation and genetic risks XIV. Potential research directions in the post-genome era based on knowledge of repair of radiation-induced DNA double-strand breaks in mammalian somatic cells and the origin of deletions associated with human genomic disorders. *Mutat Res.* 2005; 578: 3.
- Schain, Freedman DX. Studies on 5-hydroxyindole metabolism in autistic and other mentally retarded children. *J Pediatr.* 1961; 58: 315-20.
- Schouten J. P. et al. Relative quantification of 40 nucleic acid sequences by multiplex ligation-dependent probe amplification. *Nucleic Acids Res.* 2002; 30: 57.
- Sebat J. et al., Large-scale copy number polymorphism in the human genome. *Science*, 2004; 305: 525–528.
- Seifert R., Schultz G., Involvement of pyrimidinoceptors in the regulation of cell functions by uridine and by uracil nucleotides. *Trends Pharmacol. Sci.* , 1989; 10: 365–369.

- Shao YJ, Raiford KL, Wolpert CM, Cope HA, Ravan SA, AshleyKoch AA et al. Phenotypic homogeneity provides increased support for linkage on chromosome 2 in autistic disorder. *Am J Hum Genet* 2002; 70: 1058–1061.
- Shao YJ, Wolpert CM, Raiford KL, Menold MM, Donnelly SL, Ravan SA et al. Genomic screen and follow-up analysis for autistic disorder. *Am J Med Genet* 2002; 114: 99–105.
- Sharp AJ, Cheng Z, Eichler EE, Structural Variation of the Human Genome. *Annu. Rev: Human Genet.* 2006; 7:407-442.
- Sharp AJ, et al. Segmental duplications and copy number variation in the human genome. *Am. J. Hum. Genet.* 2005; 77: 78–88.
- Shen et al. High-throughput SNP genotyping on universal bead arrays. *Mutation Research*, 2005; 573: 70-82.
- Skuse et al., Evidence from syndrome of an imprinted X-linked locus affecting cognitive function. *Nature*, 1997; 387: 705-708.
- Slager, Schaid, Case-control studies of genetic markers: power and sample size approximations for Armitage's test for trend. *Hum Hered* , 2001; 52: 149– 53.
- Slater SR et al. High-resolution identification of chromosomal abnormalities using oligonucleotide arrays containing 116,204 SNPs. *Am. J. Hum. Genet.* 2005; 77: 709–726.
- Smalley SL, Asarnow RF, Spence A. Autism and genetics. *Arch Gen Psychiatry.* 1988; 45: 953–961.
- Smalley SL, Autism and tuberous sclerosis. *J. Autism Dev. Disord.* 1998; 28: 407–414.
- Smith et al., Molecular genetic delineation of 2q37.3 deletion in autism and osteodystrophy: report of a case and of new markers for deletion screening by PCR. *Cytogenet. Cell. Genet.*, 2001; 94: 15-22.
- Solinas-Toldo, et al., Matrix-based comparative genomic hybridization: biochips to screen for genomic imbalances. *Genes Chromosomes Cancer* 1997; 20: 399–407.
- Spielman, McGinnis, Ewens, Transmission test for linkage disequilibrium: the insulin gene region and insulin-dependent diabetes mellitus (IDDM). *Am J Hum Genet*, 1993; 52: 506– 16.
- Steffenburg S, Gillberg C, Hellgren L, Andersson L, Gillberg I, Jakobsson G et al. A twin study of autism in Denmark, Finland, Iceland, Norway, and Sweden. *J Child Psychol Psychiatry.* 1989; 30: 405–416.
- Stone JL, Merriman B, Cantor RM, Yonan AL, Gilliam TC, Geschwind DH et al. Evidence for sex-specific risk alleles in autism spectrum disorder. *Am J Hum Genet* 2004; 75: 1117–1123.
- Suarez BK, Hampe CL, Linkage and association. *Am J Hum Genet.* 1994; 54: 554-9.
- Suarez, The affected sib pair IBD distribution for HLA-linked disease susceptibility genes. *Tissue Antigens*, 1978; 12: 87– 93.
- Szatmari P, Jones MB, Zwaigenbaum L, MacLean JE. Genetics of autism: overview and new directions. *J Autism Dev Disord.* 1998; 28: 351-68.

Taylor DC, Neville BG, Cross JH. Autistic spectrum disorders in childhood epilepsy surgery candidates. *Eur Child Adolesc Psychiatry*. 1999; 8:189-92.

The International HapMap Consortium, A haplotype map of the human genome. *Nature*, 2005; 437: 1299-1320.

The International HapMap Consortium, The International HapMap Project. *Nature*, 2003; 426: 789-796.

Traut T.W., Physiological concentrations of purines and pyrimidines. *Mol. Cell. Biochem.*, 1994; 140: 1-22.

Trikalinos TA, Karvouni A, Zintzaras E, Ylisaukkoja T, Peltonen L, Jarvela I et al. A heterogeneity-based genome search metaanalysis for autism-spectrum disorders. *Mol Psychiatry* 2006; 11: 29–36.

Tuzun E, Sharp AJ, Bailey JA, Kaul R, Morrison VA, Pertz LM, Haugen E, Hayden H, Albertson D, Pinkel D, Olson MV, Eichler EE, Fine-scale structural variation of the human genome. *Nat Genet*. 2005; 37: 660-1.

Vincent JB, Kolozsvari D, Roberts WS, Bolton PF, Gurling HM, Scherer SW.

Vincent JB, Kolozsvari D, Roberts WS, Bolton PF, Gurling HM, Scherer SW. Mutation screening of X-chromosomal neuroligin genes: no mutations in 196 autism probands. *Am J Med Genet B Neuropsychiatr Genet*. 2004;129: 82-4.

Vita A, Amici A, Cacciamani T, Lanciotti M & Magni G, *Int.. J. Biochem*. 1986; 18: 431-436.

Vorstman JA, Staal WG, Hochstenbach PF, Franke L, van Daalen E, van Engeland H. Overview of cytogenetic regions of interest (CROIs) associated with the autism phenotype across the human genome. *Mol Psychiatry* 2006; 11: 18–28.

Vorstman JAS, Staal WG, van Daalen E, van Engeland H, Hochstenbach PFR and Franke L. Identification of novel autism candidate regions through analysis of reported cytogenetic abnormalities associated with autism. *Molecular Psychiatry*. 2006; 11: 18–28.

Warburton D., De novo balanced chromosome rearrangements and extra marker chromosomes identified at prenatal diagnosis: clinical significance and distribution of breakpoints. *Am. J. Hum. Genet*, 1991; 49: 995–1013.

Wassink TH, Brzustowicz LM, Bartlett CW, Szatmari P. The search for autism disease genes. *Mental Retard Dev Disabil Res Rev* 2004; 10: 272–283.

Wassink TH, Piven J, Patil SR. Chromosomal abnormalities in a clinic sample of individuals with autistic disorder. *Psychiatr Genet* 2001; 11: 57–63.

Wassink TH, Piven J, Vieland VJ, Huang J, Swiderski RE, Pietila J et al. Evidence supporting WNT2 as an autism susceptibility gene. *Am J Med Genet* 2001; 105: 406–413.

Weeks DE, Lathrop GM, Polygenic disease: methods for mapping complex disease traits. *Trends Genet*. 1995; 11: 513-9.

Wilson HL, Wong ACC, Shaw SR, Tse WY, Stapleton GA, Phelan MC, Hu S, Marshall J, McDermid HE. Molecular characterisation of the 22q13 deletion syndrome supports the role of haploinsufficiency of SHANK3/PROSAP2 in the major neurological symptoms. *J. Med. Genet.* 2003; 40: 575-584.

Wolff DJ, Clifton K, Karr C, Charles J. Pilot assessment of the subtelomeric regions of children with autism: detection of a 2q deletion. *Genet Med* 2002; 4: 10–14.

Wolpert et al., Three probands with autistic disorder and iodicentric chromosome 15. *Am. J. Med. Genet.* 2000; 96: 365-372.

World Health Organisation. The ICD-10 Classification of Mental and Behavioural Disorders. Clinical Descriptions and Diagnostic Guidelines. World Health Organisation: Geneva, 1992.

Yan J, Oliveira G, Coutinho A, Yang C, Feng J, Katz C, Sram J, Bockholt A, Jones IR, Craddock N, Cook EH Jr, Vicente A, Sommer SS Analysis of the neuroligin 3 and 4 genes in autism and other neuropsychiatric patients. *Mol Psychiatry.* 2005; 10: 329-32.

Yonan AL, Alarcon M, Cheng R, Magnusson PKE, Spence SJ, Palmer AA et al. A genomewide screen of 345 families for autism susceptibility loci. *Am J Hum Genet* 2003; 73: 886–897.

Zhao Q., Marolewski A., Rusche J., Holmes G. Effects of uridine in models of epileptogenesis and seizures. *Epilepsy Research* , 2006.

Zoghbi HY. Postnatal neurodevelopmental disorders: meeting at the synapse. *Science.* 2003; 302: 826-30.