

**Alma Mater Studiorum – Università di Bologna**

**DOTTORATO DI RICERCA IN**

**Biologia Cellulare e Molecolare**

**Ciclo XXVII**

Settore Concorsuale di afferenza: **05/I1**

Settore Scientifico disciplinare: **BIO/18**

**Analysis of Copy Number Variants  
identifies new candidate genes  
for Autism Spectrum Disorder and Intellectual Disability**

Presentata da: **Silvia Lomartire**

Coordinatore Dottorato  
**Prof. Davide Zannoni**

Relatore  
**Prof.ssa Elena Maestrini**

Co-relatore  
**Dott.ssa Elena Bacchelli**

**Esame finale-Aprile 2015, Bologna**

## TABLE OF CONTENTS

<b>ABSTRACT</b> .....	<b>1</b>
<b>CHAPTER 1:</b> .....	<b>2</b>
<b>COPY NUMBER VARIANTS</b> .....	<b>2</b>
1.1 Copy Number Variants .....	2
1.2 Mutational mechanisms .....	3
1.2.1 Non-Allelic Homologous Recombination (NAHR). .....	4
1.2.2 Non-Homologous End Joining (NHEJ) .....	6
1.2.3 Fork Stalling and Template Switching.....	7
1.2.4 Retrotransposition .....	9
1.3 CNV detection methods .....	10
1.3.1 CGH-array and ROMA .....	11
1.3.2 SNPs arrays .....	13
1.3.3 PCR-based approaches.....	14
1.3.4 Next Generation Sequencing-based approaches .....	15
1.4 CNVs and neurodevelopmental disorders.....	17
1.5 Recurrent CNVs on chromosome 15q11-13.....	21
<b>CHAPTER 2:</b> .....	<b>25</b>
<b>AUTISM SPECTRUM DISORDER</b> .....	<b>25</b>
2.1 Autism Spectrum Disorder.....	25
2.2 Clinical phenotype and diagnostic instruments .....	26
2.3 The genetic basis of ASD.....	28
2.4 Molecular genetic studies of ASD susceptibility .....	30
2.4.1 Linkage studies in ASD .....	30
2.3.2 Association studies in ASD.....	31
2.5 CNVs and ASD.....	32
2.6 Next Generation Sequencing Technologies (NGS) and ASD .....	35
<b>CHAPTER 3:</b> .....	<b>39</b>
<b>ASD CANDIDATE GENES</b> .....	<b>39</b>
3.1 ASD candidate genes .....	39
3.2 ASD-related syndromes .....	40
3.3 “Synaptic” genes: neuroligins, neuroligins and SHANK family .....	41
3.4 Candidate genes object of this thesis .....	44
3.4.1 The $\alpha$ -catenins .....	44
3.4.2 The <i>CADPS</i> family.....	46
3.4.3 The <i>CHRNA7</i> gene.....	48

3.4.4 The kelch-like (KLHL) gene family .....	51
<b>CHAPTER 4:.....</b>	<b>53</b>
<b>INTELLECTUAL DISABILITY .....</b>	<b>53</b>
4.1 Intellectual disability.....	53
4.2 Causes of ID.....	54
4.2.1 Chromosomal aberrations and ID .....	54
4.2.2 X-linked intellectual disability (XLID).....	55
4.2.3 Autosomal forms of ID .....	57
4.3 Co-morbidity of ID with ASD .....	58
4.3.1 Phenotypic overlap between ID and ASD .....	59
4.3.2 Genetic overlap between ID and ASD.....	59
<b>CHAPTER 5:.....</b>	<b>62</b>
<b>MATERIALS AND METHODS .....</b>	<b>62</b>
5.1 ASD Samples .....	62
5.1.1 Italian ASD Cohort .....	62
5.1.2. AGP Cohort.....	62
5.1.3 IMGSAC Cohort.....	63
5.2 Control samples.....	64
5.2.1 AGP control cohorts.....	64
5.2.2 Italian controls.....	65
5.2.3 European controls .....	65
5.3 Identification of CNVs by Illumina Human 1M-Duo BeadChip array .....	65
5.4 Validation of CNVs by Real-Time PCR.....	68
5.5 Mutation screening.....	70
5.5.1 Primers design.....	70
5.5.2 Polymerase Chain Reaction (PCR) assay .....	70
5.5.3 PCR purification .....	72
5.5.4 Sanger Sequencing reaction .....	73
5.5.5 Ethanol-EDTA Precipitation of Sequencing Reactions .....	73
5.6 Prediction tools .....	73
5.6 RNA extraction and cDNA synthesis .....	74
5.7 Analysis of the candidate gene <i>CTNNA3</i> .....	74
5.7.1 Characterisation and segregation analysis of <i>CTNNA3</i> deletions.....	74
5.7.2 <i>CTNNA3</i> and <i>LRRTM3</i> exon sequencing.....	75
5.7.3 <i>CTNNA3</i> expression.....	75
5.7.4 Western blot analysis .....	75
5.8 Analysis of the candidate gene <i>CHRNA7</i> .....	76

5.8.1 Validation of the <i>CHRNA7</i> duplication .....	76
5.8.2 <i>CHRNA7</i> sequencing.....	76
5.9 Analysis of the microdeletion on chromosome 2q31.1.....	77
5.9.1 Validation of the microdeletion in the discovery pedigree .....	77
5.9.2 <i>KLHL23</i> and <i>PHOSHO2-KLHL23</i> expression analyses .....	77
5.9.3 <i>KLHL23</i> and <i>PHOSHO2-KLHL23</i> mutation screening .....	78
5.9.4 PCR-Restriction Fragment Length Polymorphism (RFLP) analysis.....	78
5.10 Analysis of the <i>CADPS2</i> gene .....	79
5.10.1 Microsatellite analysis.....	79
5.10.2 <i>CADPS2</i> expression analysis .....	80
5.10.3 MALDI-TOF MS methylation analysis .....	80
5.10.4 PCR and sequencing of <i>CADPS2</i> intron 1 .....	84
5.10.5 Colony screening.....	84
<b>CHAPTER 6:.....</b>	<b>94</b>
<b>AIM OF THE STUDY AND PRELIMINARY RESULTS.....</b>	<b>94</b>
6.1 Aim of the study.....	94
6.2 Preliminary results .....	95
6.2.1 Analysis of three CNVs identified in ASD Italian families.....	95
6.2.2 Analysis of the <i>CADPS2</i> gene in subjects with ID and/or ASD.....	96
<b>CHAPTER 7:.....</b>	<b>101</b>
<b>RESULTS .....</b>	<b>101</b>
Characterization of three CNVs identified in ASD Italian families .....	101
7.1 Analysis of the <i>CTNNA3</i> gene .....	101
7.1.1 Fine-mapping and segregation analysis of <i>CTNNA3</i> deletion in family 3456.....	101
7.1.2 <i>CTNNA3</i> exonic deletion frequency in ASD cases and controls .....	105
7.1.3 Segregation analysis of <i>CTNNA3</i> exonic deletions and mutation screening of <i>CTNNA3</i> and <i>LRRTM3</i> in four ASD families.....	107
7.1.4 <i>CTNNA3</i> expression analysis .....	109
7.2 Analysis of rare variants in the <i>CHRNA7</i> gene.....	111
7.2.1 Validation of the <i>CHRNA7</i> microduplication in family 3474.....	111
7.2.2 CNV analysis and mutation screening in 135 ASD subjects.....	113
7.2.2a CNV analysis.....	113
7.2.2b <i>CHRNA7</i> mutation screening .....	113
7.3 Analysis of a rare microdeletion on chromosome 2q31.1 .....	118
7.3.1 Fine mapping of a rare microdeletion on chromosome 2q31.1 .....	118
7.3.2 CNV search in the 2q31.1 locus.....	120
7.3.3 <i>KLHL23</i> and <i>PHOSHO2-KLHL23</i> expression analyses .....	120
7.3.4 <i>KLHL23</i> and <i>PHOSHO2-KLHL23</i> mutation screening .....	121

<b>CHAPTER 8:</b> .....	<b>124</b>
<b>RESULTS</b> .....	<b>124</b>
Analysis of the <i>CADPS2</i> gene in subjects with ID and/or ASD.....	124
8.1 Maternal inheritance of the <i>CADPS2</i> deletion.....	124
8.2 <i>CADPS2</i> allelic expression analysis in brain tissues .....	124
8.3 Quantitative methylation analysis of <i>CADPS2</i> CpG regions.....	125
<b>CHAPTER 9:</b> .....	<b>130</b>
<b>DISCUSSION</b> .....	<b>130</b>
9.1 Analysis of rare CNVs implicated in ASD and ID susceptibility.....	130
9.2 Analysis of a compound heterozygous deletion in <i>CTNNA3</i> .....	130
9.3 Analysis of a microduplication in the <i>CHRNA7</i> gene.....	132
9.4 Analysis of a rare microdeletion on chromosome 2q31.1 .....	135
9.5 Analysis of <i>CADPS2</i> in individuals with ASD and ID.....	136
CONCLUSIONS.....	137
<b>References</b> .....	<b>139</b>

**ABSTRACT**

Autism spectrum disorder (ASD) and Intellectual Disability (ID) are complex neuropsychiatric disorders characterized by extensive clinical and genetic heterogeneity and with overlapping risk factors.

The aim of my project was to further investigate the role of Copy Numbers Variants (CNVs), identified through genome-wide studies performed by the Autism Geome Project (AGP) and the CHERISH consortium in large cohorts of ASD and ID cases, respectively.

Specifically, I focused on four rare genic CNVs, selected on the basis of their impact on interesting ASD/ID candidate genes: a) a compound heterozygous deletion involving *CTNNA3*, predicted to cause the lack of functional protein; b) a 15q13.3 duplication containing *CHRNA7*; c) a 2q31.1 microdeletion encompassing *KLHL23*, *SSB* and *METTL5*; d) Lastly, I investigated the putative imprinting regulation of the *CADPS2* gene, disrupted by a maternal deletion in two siblings with ASD and ID.

This study provides further evidence for the role of *CTNNA3*, *CHRNA7*, *KLHL23* and *CADPS2* as ASD and/or ID susceptibility genes, and highlights that rare genetic variation contributes to disease risk in different ways: some rare mutations, such as those impacting *CTNNA3*, act in a recessive mode of inheritance, while other CNVs, such as those occurring in the 15q13.3 region, are implicated in multiple developmental and/or neurological disorders possibly interacting with other susceptibility variants elsewhere in the genome. On the other hand, the discovery of a tissue-specific monoallelic expression for the *CADPS2* gene, implicates the involvement of epigenetic regulatory mechanisms as risk factors conferring susceptibility to ASD/ID.

## CHAPTER 1: COPY NUMBER VARIANTS

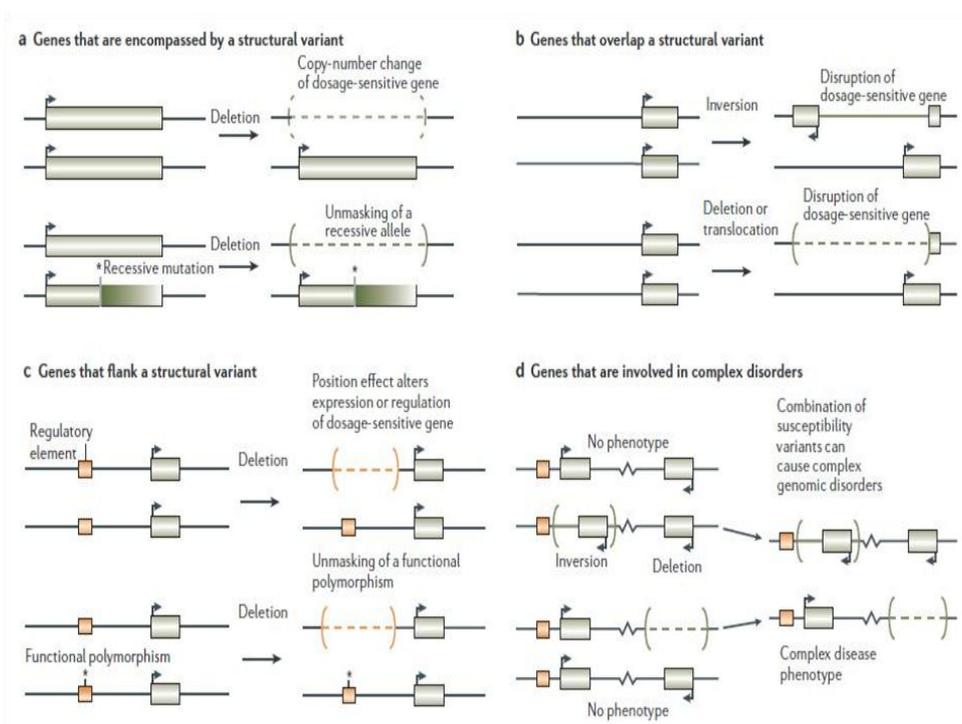
### **1.1 Copy Number Variants**

Before the availability of sequencing technology, the first differences observed in the human genetic composition were mainly rare changes in the quantity and structure of chromosomes, defined as microscopic structural variants (~ 3Mb or more in size). Subsequently, with the advent of DNA sequencing, smaller and more abundant alterations were observed: individual base changes called Single-Nucleotide Polymorphisms (SNPs) are the most numerous variants in the genome, with estimates of at least 10 million SNPs within the human population (Kruglyak & Nickerson, 2001); various repetitive elements that involve relatively short DNA sequences (for example, micro- and mini-satellites), and small (usually <1 kb) insertions, deletions, inversions and duplications. More recently, new strategies and tools for analysis of the human genome have revealed DNA variation (ranging from ~1 kb to 3 Mb in size), involving segments that are smaller than those identified using a microscope, but larger than those detectable by conventional sequence analysis. These submicroscopic structural variants, defined as Copy number variants (CNVs), are DNA segments longer than 1kb, with a variable copy number compared to a reference genome (Feuk, Carson, & Scherer, 2006). CNVs have gained considerable interest as a source of genetic diversity likely to play a role in functional variation, as they are widespread in normal human genomes (Iafate et al., 2004; Sebat et al., 2004). Recently two studies (Conrad et al., 2010; Mills et al., 2011) have shown that there are > 1000 CNVs in the genome, accounting for million base pairs of genomic difference. Therefore, although the number of SNPs in the genome exceed the number of CNVs, their relative contribution to genetic heterogeneity is similar if we consider the variation in terms of nucleotides implicated.

A CNV can be a deletion, insertion, duplication or an inversion and it can be inherited or may arise *de novo* on a paternally or maternally inherited chromosome. Moreover, these types of variants can encompass millions of bases of DNA, containing entire genes and their regulatory regions (Sebat et al., 2004; Sharp et al., 2005; Tuzun et al., 2005).

Although some structural variants represent benign CNVs, others can predispose to or directly cause a disease or, in combination with other genetic and environmental factors, they might function as susceptibility alleles in complex genetic disorders.

Genes might be influenced in several ways, depending on their proximity to CNVs (**figure 1.1**) and their phenotypic effects usually depend on whether dosage-sensitive genes or regulatory regions are affected. In the simplest cases deletions, inversions or translocations can affect gene dosage directly or, in some cases, deletions can “unmask” a recessive pathogenic allele in the heterozygous state (**figure 1.1a**); CNVs can also lead to the reduced expression (**figure 1.1b**), indirectly alter gene expression through position effects (**figure 1.1c**). or result in fusion or abnormal gene products with a new function (Holt et al., 2012). Moreover two CNVs in combination could predispose to a complex disorder (**figure 1.1d**).



**Figure 1.1** (Feuk et al., 2006). **Influence of structural variants on phenotype.** **a)** Dosage-sensitive genes that are encompassed by a structural variant can cause disease through a deletion event (upper panel); dosage-insensitive genes can also cause disease if a deletion of the gene unmask a recessive mutation on the homologous chromosome (lower panel). **b)** Genes that overlap structural variants can be disrupted directly by inversion (upper panel), translocation or deletion (lower panel). **c)** Structural variants that are located at a distance from dosage-sensitive genes can affect expression through position effects (upper panel). Alternatively, deletion of a functional element could unmask a functional polymorphism within an effector which could have consequences for gene function (lower panel). **d)** Structural variants can function as susceptibility alleles, where a combination of several genetic factors are required to produce the phenotype.

## 1.2 Mutational mechanisms

Different mechanisms can lead to the formation of CNVs.

Structural variants can arise as a result of one or more DNA breaks, which are followed by a wrong restoration of the continuity of filaments involved. These alterations may result either from an error

in the repair process or from a defect of the recombination system. The four major mechanisms of structural mutations are: non-allelic homologous recombination (NAHR); non-homologous end joining (NHEJ), fork stalling and template switching (FoSTeS) and retrotransposition. Recent findings from the 1000 Genomes Project (Mills et al., 2011) have shown that about 70.8% of the deletions are caused by NHEJ; 89.6% of small insertions are attributed to retrotransposition activity, while most tandem duplications are likely formed by FoSTeS.

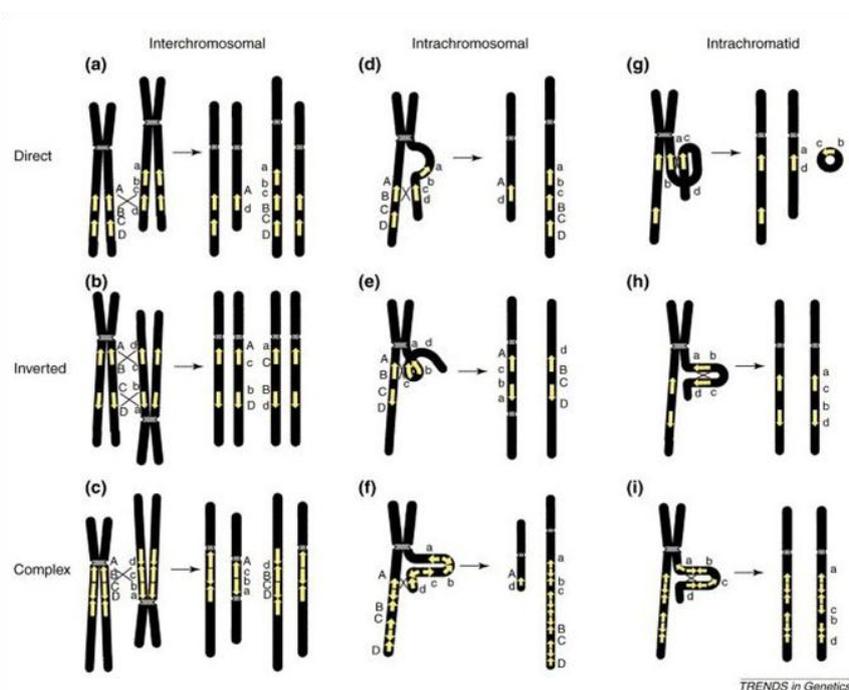
### 1.2.1 Non-Allelic Homologous Recombination (NAHR).

Low copy repeats (LCRs) or segmental duplications are DNA segments >1 kb in size that occur in two or more copies per haploid genome, with the different copies sharing >90% sequence identity.

The size, orientation and relative arrangement of LCRs affect the genome architecture such that they result in genomic regions that are unstable and prone to subsequent Non-Allelic Homologous Recombination (NAHR). NAHR occurs by a chromosome/chromatid misalignment that enables non-allelic LCRs to pair as substrates for homologous recombination (Stankiewicz & Lupski, 2002). NAHR can involve genomic rearrangements between paralogues on homologous chromosomes (*inter-chromosomal*), on sister chromatids (*intra-chromosomal*) and on the same chromatid (*intra-chromatid*) (Gu, Zhang, & Lupski, 2008).

The result of unequal crossing over between flanking LCRs is the formation of homologous chromosomes, carrying a reciprocal deletion and a duplication of intervening sequence, which segregate from each other at the next cell division, leading to a change in copy number in both daughter cells. These two reciprocal products (tandem duplication and deletion) are produced when LCRs are directly oriented on homologous chromosomes (**Figure 1.2a**) or on sister chromatids (**Figure 1.2d**). Alternatively, unequal crossing-over between direct LCRs on the same chromatid leads to the loop excision and deletion (**Figure 1.2g**). Instead, mispairing between inverted repeats results in an inversion when crossover involves both repeats (**Figure 1.2b and 1.2h**). Some LCRs have complex structures with sequences among the LCRs oriented in a direct manner whereas others are inverted and they can lead to both inversion or deletion and duplication (**Figure 1.2c, 1.2f, 1.2i**).

NAHR can occur both in meiosis (Turner et al., 2008) and, at lower frequency, in mitotically dividing cells. The majority of NAHR events occurs between LCRs which have a sequence identity greater than ~97%, a size that ranges from about 10 to ~400 kb and are located at a distance less than ~10 Mb from each other (Lupski, 1998; Sharp et al., 2005).

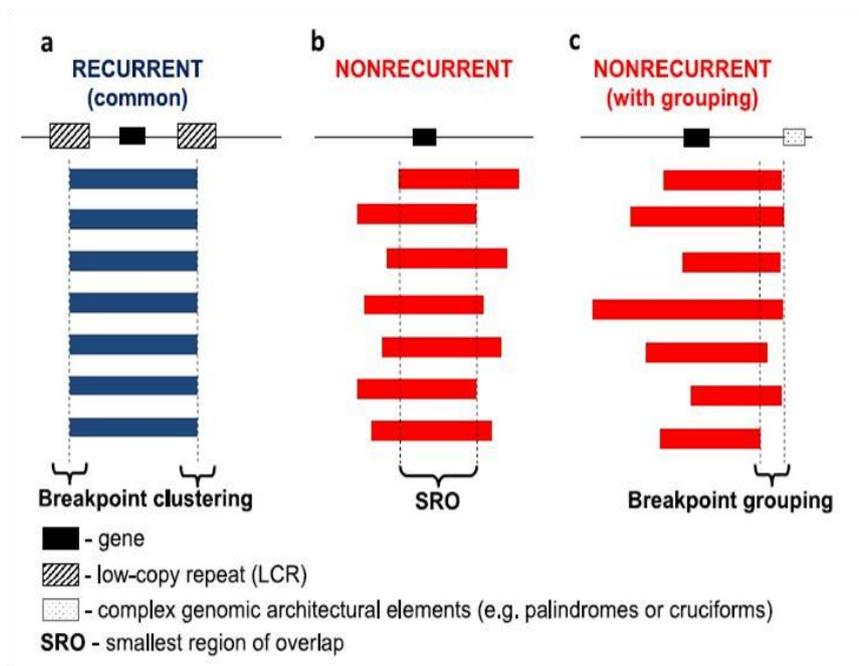


**Figure 1.2** (Stankiewicz & Lupski, 2002). **Schematic representation of LCRs-NAHR based mechanisms for genomic rearrangements.** LCRs are depicted as yellow arrows and their orientation is indicated. Interchromosomal misalignment leads to deletion/duplication (a) and inversion (b). Intrachromosomal mispairing of direct repeats results also in deletion/duplication (d). Intrachromatid misalignment of directed repeats (g) can result in deletion and an acentric fragment. Intrachromatid loop of inverted repeats results in inversion (h). Unequal exchange between complex LCRs can be responsible for deletion/duplication or inversion (c; f; i).

NAHR has been shown to require segments of a minimal length (300-500 bp) sharing extremely high similarity or identity between the LCRs, called *minimal efficient processing segments* (MEPS) (>90%). Two studies (Rubnitz & Subramani, 1984; Waldman & Liskay, 1988) showed that introduction of only two single-nucleotide mismatches leads to a 20-fold reduction of the recombination frequency. Other studies have found that NAHR can also be mediated by highly repetitive sequences, such as *Alu sequences* (which are the main class of *Short Interspersed Elements*, SINES) (Redon et al., 2006) or by L1 elements (which are the main class of *Long Interspersed Elements*, LINEs), but also by retrotransposons or satellite DNAs. Moreover, different studies have provided evidence that the majority of strand exchanges or crossovers are restricted to a specific region (positional recombination hotspots) located within the repeat and are responsible for some genomic disorders (Lupski, 2004; Reiter et al., 1996).

Most of the rearrangements resulting from NAHR are “recurrent” rearrangements (**figure 1.3a**), as they recur in multiple individuals, share a common size and have fixed breakpoints, which are clustered inside the LCRs. Instead, the “non-recurrent” rearrangements (**figure 1.3b**) present different sizes and distinct breakpoints for each event. However all of the non-recurrent rearrangements share a common genomic region of overlap, indicated as the smallest region of

overlap (SRO). Some of non-recurrent rearrangements have grouping of breakpoints, as one of their breakpoints localized in one small genomic region (**figure 1.3c**) (Gu et al., 2008).

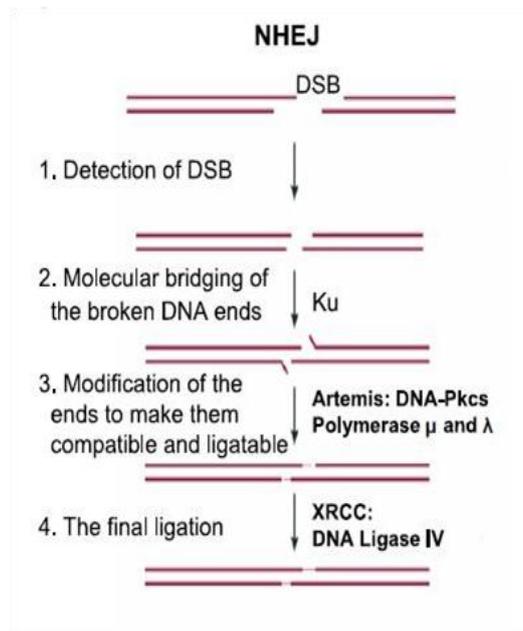


**Figure 1.3** (Gu et al., 2008) **Recurrent and non-recurrent rearrangements.** The black line indicates the genomic region hit by CNVs, blue and red bars indicate the rearrangements with their breakpoints. **a.** Recurrent rearrangements have a common size and clustered breakpoints; both breakpoints map within the same LCRs (hatching rectangles). **b.** The non-recurrent rearrangements have different length and different breakpoints, but might share a common region of overlap (SRO). In this example, the SRO encompasses one gene, indicated by the black rectangle. **c.** Some non-recurrent rearrangements show one of their breakpoints localized in one small genomic region: the grouping of one breakpoint may occur in proximity of an architectural element important to the rearrangements mechanism.

### 1.2.2 Non-Homologous End Joining (NHEJ)

Another mechanism that can lead to CNVs formation is Non-Homologous End Joining (NHEJ), one of the main pathways for repairing double-stranded DNA breaks (DSBs).

NHEJ proceeds in four steps (**figure 1.4**): detection of DSB; molecular bridging of both broken DNA ends; modification of the ends to make them compatible and ligatable; and the final ligation step (Weterings & van Gent, 2004)

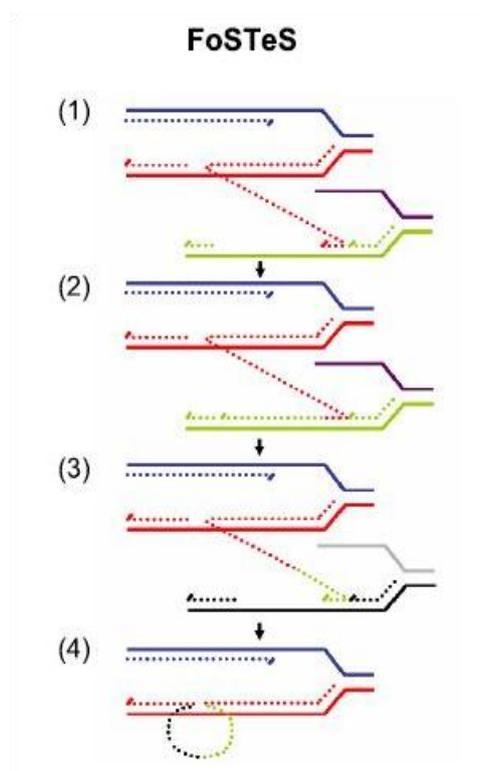


**Figure 1.4** (Gu et al., 2008). **The NHEJ mechanism consists of four steps.** The two thick lines represents the double stranded DNA with double strand breaks (DSB). When DSBs occur, they are detected (1) and the molecular machinery of NHEJ mediates the molecular bridging of the broken DNA ends (2). The DNA ends are modified to be compatible (3) and, finally, they are ligated (4) to restore the structural integrity.

The ends of DNA double-strand breaks are repaired through many rounds of enzymatic activity and, since linkage of the two ends requires some addition or deletion of bases in order to make them complementary, NHEJ leaves “information scar” at the rejoining site. This mechanism is guided entirely by the information contained within or near the DNA lesion for repair, which makes it “error prone”, since it introduces mutations compared to reference sequence (Lieber, 2008). In contrast to NAHR, NHEJ does not require LCRs, MEPS or sequence homology to mediate the recombination. However, it can frequently occur within repetitive elements, such as long terminal repeats (LTRs), SINEs, LINEs, and mammalian interspersed repeats (MIRs).

### 1.2.3 Fork Stalling and Template Switching

Fork Stalling and Template Switching (FoSteS) is a replication-based genomic rearrangement mechanism induced by errors during DNA replication process: during DNA replication, the DNA replication fork stalls when it encounters a nick in a template strand, the lagging strand releases the original template and anneals, by virtue of micro-homology at the 3’ end, to another replication fork nearby; then restarts the DNA synthesis (Lee, Carvalho, & Lupski, 2007) (**figure 1.5**).



**Figure 1.5** (Gu et al., 2008). A **schematic representation of the Fork Stalling and Template Switching (FoSteS) mechanism**. Solid lines indicate the template DNA, dotted lines instead the newly synthesised strands. The original replication fork is represented by the red and blue lines. After the stalling, the lagging strand (red, dotted line) invades a second fork (indicated in purple and green) via microhomology (1). The 3' end of the lagging strand allows the DNA extension in the second fork (green dotted line) (2). Serial replication fork disengaging and invasion could occur several times (3) before resumption of replication on the original template (4).

A further generalization of the FoSteS mechanism is known as the Microhomology-Mediated Break-Induced Replication model (MMBIR) (Hastings, Ira, & Lupski, 2009). It is used to repair single double-strand ends when the 3' end of a collapse fork is able to anneal to any single-stranded DNA stretch, available in physical proximity, with which it shares micro-homology. Whether the template switch occurs in front of the position of the original collapse, a deletion will be created, whereas switching to a fork located upstream will result in a duplication. Moreover, whether the switch occurs in direct or opposite orientation determines if the erroneously incorporated fragment will be in direct or inverted orientation with respect to its original position.

Genomic rearrangements generated by this mutational mechanism vary greatly in size and complexity (Hastings, Lupski, Rosenberg, & Ira, 2009). When FoSteS/MMBIR is mediated by large inverted repeats (>300 kb apart) and coupled with NHEJ, it can cause also the formation of complex rearrangements with duplication-triplication/inversion-duplication structures (Carvalho et al., 2011).

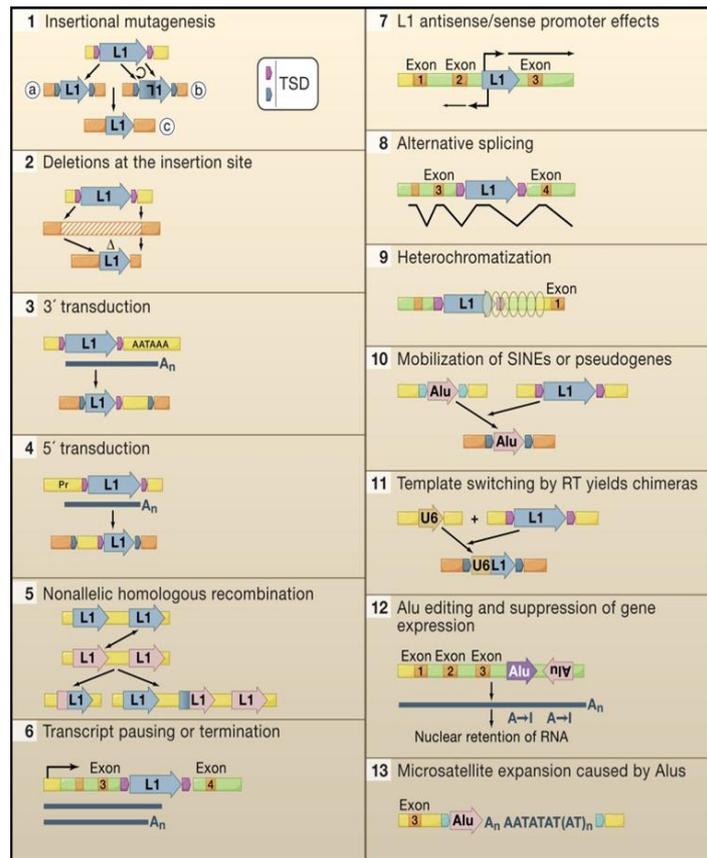
### 1.2.4 Retrotransposition

Retrotransposons, mainly LINEs (Long Interspersed Nuclear Elements), SINEs (Short Interspersed Nuclear Elements), and endogenous retroviruses, make up about 40% of the mammalian genome.

LINE1 or L1 are the only currently active class of retrotransposons in humans, moving their own and SINE sequences into new genomic locations.

Although ~500,000 copies are present in the genome, only about 80-100 are active full-length elements (6 Kb) and are able to transpose their own sequences or non-autonomous elements to new genomic locations by a “Target Primed Reverse Transcription” (TPRT) mechanism (Goodier & Kazazian, 2008). TPRT results in the insertion of a new, often 5'-truncated, L1 copy at a new genomic location that generally is flanked by “target site duplications” (TSDs) (**figure 1.6.1**). Retrotransposition occasionally can generate target site deletions (**figure 1.6.2**). Moreover recombination between retrotransposons may causes deletions, duplications, or rearrangements of gene sequence (**fig 1.6.5**). However, the greatest impact of retrotransposon insertions is on the expression of nearby genes: such as pausing in transcriptional elongation and premature transcript termination (**figure 1.6.6**), producing new transcription start sites (**fig 1.6.7**), leading to new exons within genes (**figure 1.6.8**) or altering the chromatin state (**figure 1.6.9**), producing chimeric insertions in the genome (**figure 1.6.11**) and suppressing gene expression (**fig 1.6.12**).

Both germline and somatic L1 activity contribute significantly to structural variation in human genomes (Lupski, 2010).



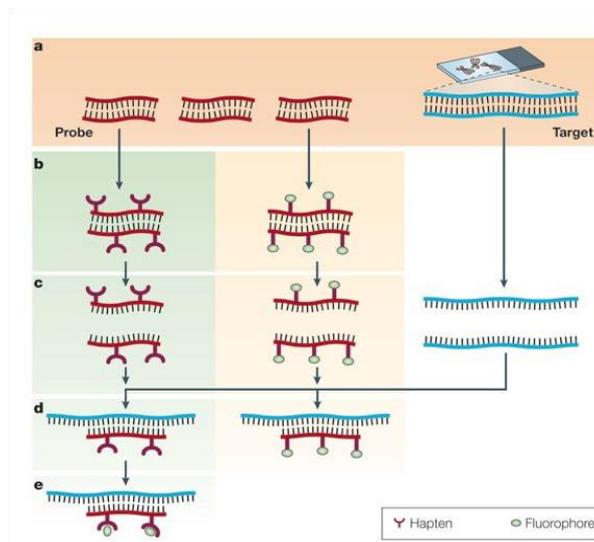
**Figure 1.6** (Goodier & Kazazian, 2008). **Examples of genomic changes caused by retrotransposons.** (1) Insertion of the L1 element to a new location. In the example a, the insertion causes the formation of TSDs. (2) The insertion of the L1 determines a deletion at the insertion site. (3 and 4) Regions flanking the retrotransposon in the original location (at 5' or 3') may be carried along with the L1 element during the retrotransposition. (5) Mismatching and crossing over between LINE or SINE elements via NAHR, leading to deletions and duplications. (6) The retrotransposon sequence can cause a pausing in the transcriptional elongation, and poly(A) signals within an L1 can lead to premature termination of transcription. (7) The antisense promoter in the L1 5' UTR can produce new transcription start sites for genes upstream of the L1 on the opposite strand. (8) Splice sites within L1s residing in introns can lead to new exons within genes, in a process called “exonization”. (9) L1s can alter the chromatin state, thereby altering gene expression. (10) L1 reverse transcriptase can mobilize Alu, leading to further genome expansion. (11) Template switching of L1 reverse transcriptase from L1 RNA to other sequences, such as U6 RNA, can produce chimeric insertions in the genome. (12) Editing of inverted Alu can suppress gene expression by nuclear retention of the mRNA. (13) Alu elements can promote formation and expansion of microsatellites.

### 1.3 CNV detection methods

Chromosome banding techniques, developed in the early 1970s, enabled the identification of each of the 23 pairs of human chromosomes and the specific bands involved in rearrangements.

However, since the resolution obtained using banding methods is 2-5 Mb, molecular cytogenetic methods, such as *fluorescence in situ hybridization* (FISH) (**figure 1.7**) (Speicher & Carter, 2005), have been required to visualize the submicroscopic rearrangements, which are too subtle to be detected by routine analysis. FISH analysis uses fluorescent DNA probes, that hybridize to either metaphase or interphase chromosomes in order to detect and localize specific DNA sequences.

This technology allows the identification of small genomic alterations, from 50 Kb to 100 Kb, and the diagnosis of both micro-deletion/micro-duplications syndromes and cancer. However, the main limitation of FISH is represented by the detection of specific chromosomal loci clearly associated with known syndromes, not providing thus an analysis of the entire genome.



**Figure 1.7** (Speicher & Carter 2005). **Fluorescence *in situ* hybridization (FISH)**. a) FISH uses a DNA probe and a target sequence. b) the DNA probe is labelled before hybridization; there are two main labelling strategies: for indirect labelling (left panel) probes are labelled with modified nucleotides that contain a HAPTEN, whereas direct labelling (right panel) uses nucleotides that have been directly modified to contain a fluorophore. c) The labelled probe and the target DNA are denatured d) Then they anneal to complementary DNA sequences. e) If the probe has been labelled indirectly, an extra step is required for visualization of the non-fluorescent haptent that uses an enzymatic or immunological detection system.

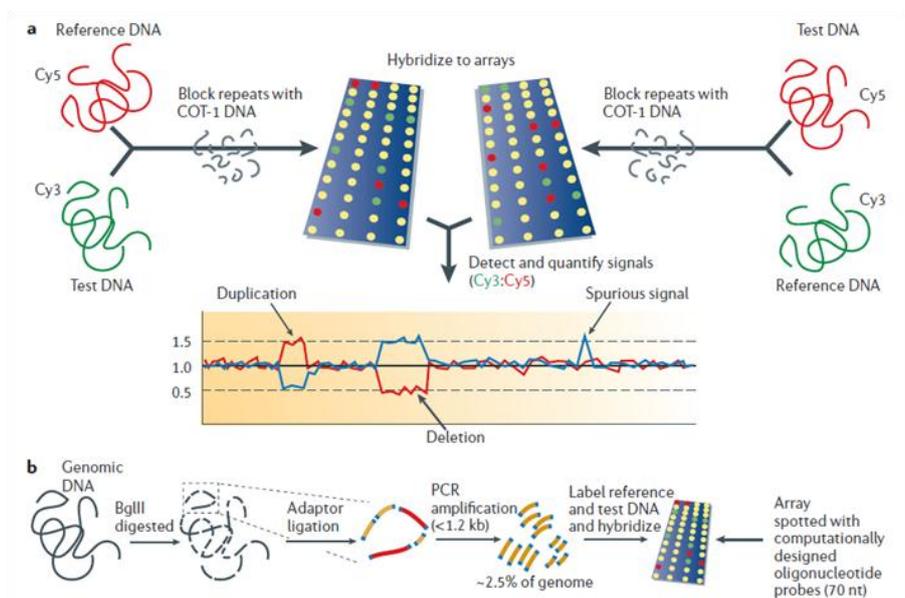
The advent of new experimental methods, which are able to assay the genome in either a genome-wide (such as array-CGH and SNPs arrays) or a targeted manner (such as Real-time qPCR) with varying degrees of resolution (ranging from kb to base pair resolution), has led to dramatically increase the number of identified CNVs. Moreover, the development of Next Generation Sequencing (NGS) technologies has allowed sequence-based approaches for mapping CNVs at fine scale.

### 1.3.1 CGH-array and ROMA

Array Comparative Genomic Hybridization (CGH-array) is the most robust array-based method for carrying out genome-wide scans to find novel CNVs (Pinkel et al., 1998). In array-CGH two DNA sample, a reference DNA and a DNA of interest (test DNA), labelled with different fluorescent dyes (e.g. Cy5 and Cy3), compete to hybridize to arrays that are spotted with DNA fragments to cover the whole genome. After hybridization, the fluorescence ratio between the two dyes is determined, revealing copy-number differences between the two genomes (**figure 1.8a**). Typically, array-CGH

is carried out using a ‘dye-swap’ method, in which the initial labelling of the reference and test DNA samples is reversed for a second hybridization (indicated by the left and right sides of the figure 1.8a). This allows the detection of spurious signals, which are not common to both hybridizations. The array can be spotted with one of several DNA sources, including genomic clones (for example, BACs), PCR fragments or oligonucleotides. The use of BACs gives the advantage of extensive coverage of the genome, reliable mapping data and ready access to clones. Instead the use of CGH with long oligonucleotides (60-100 bp) can improve the detection resolution (from 50 kb to, theoretically, a few kb) compared to BACs.

A variant of array-CGH is Representational Oligonucleotide Microarray Analysis (ROMA) (**figure 1.8b**) (LaFramboise, 2009; Lucito et al., 2003) in which, in order to reduce the complexity of the DNA that will be hybridized to the array, the reference and test DNA samples are digested with a restriction enzyme and then ligated to adaptors, which results in the PCR-based amplification of fragments in a specific size-range. It is estimated that around 200,000 fragments of DNA are amplified (approximately 2.5% of the human genome), leading thus to a reduction in background noise.

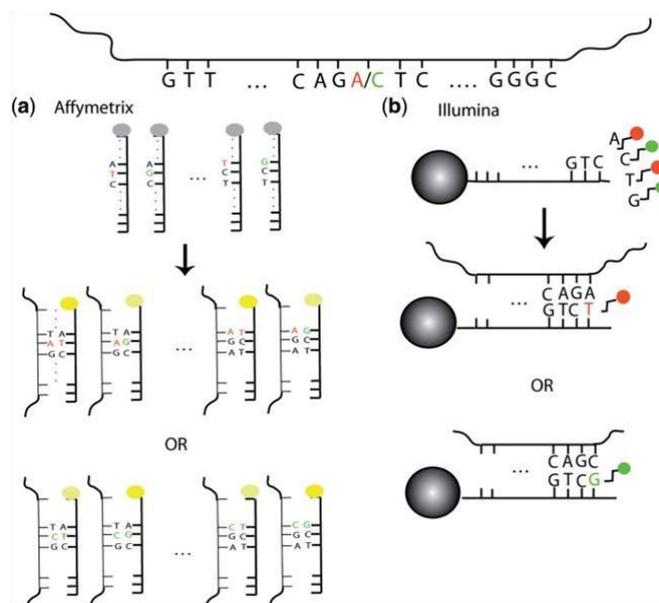


**Figure 1.8** (Feuk et al., 2006). **a) Array-based comparative genome hybridization (array-CGH) and b) Representational Oligonucleotide Microarray Analysis (ROMA).** **a)** Reference and test DNA samples are differentially labeled with fluorescent tags (Cy5 and Cy3, respectively), and are then hybridized to genomic arrays after repetitive-element binding is blocked using COT-1 DNA (which is mainly composed of repetitive sequences). After hybridization, the fluorescence ratio (Cy3: Cy5) is determined, which reveals copy-number differences between the two DNA samples. An example output for a dye-swap experiment is shown at the bottom: the red line represents the original hybridization, whereas the blue line represents the reciprocal, or dye-swapped, hybridization. **b)** Before hybridization the reference and test DNA samples are digested with a restriction enzyme that has uniformly distributed cleavage sites (BglII in this example). Adaptors (with PCR primer sites) are then ligated to each fragment, which are amplified by PCR. Only DNA of less than 1.2 kb (yellow) is amplified, while fragments that are greater than this size (red) are lost, therefore reducing the complexity of the DNA that will be hybridized to the array. An oligonucleotide array is used, which is spotted with computationally designed 70-nt probes.

### 1.3.2 SNPs arrays

Others array-based approaches are SNPs arrays which, compared to CGH-arrays, have the added advantage of providing information about both copy number and genotype, leading to considerable progress in genome-wide detection of various types of DNA sequence-level human variation.

Two commercial SNPs arrays, which work using different chemistries but sharing several aspects, are the Affymetrix and Illumina SNPs arrays (**figure 1.9**) (LaFramboise, 2009). Both protocols are based on the hybridization of fragmented single-stranded DNA to arrays containing hundreds of thousands of oligonucleotidic sequences. Each probe is designed to bind to a target DNA subsequence and it is represented multiple times within an array. The number of SNPs represented on the array through these probes is proportional to the resolution of the array. After the hybridization, a detection system measures the fluorescent signal associated with each probe. The signal intensity depends upon the amount of target DNA in the sample, as well as the affinity between target and probe.

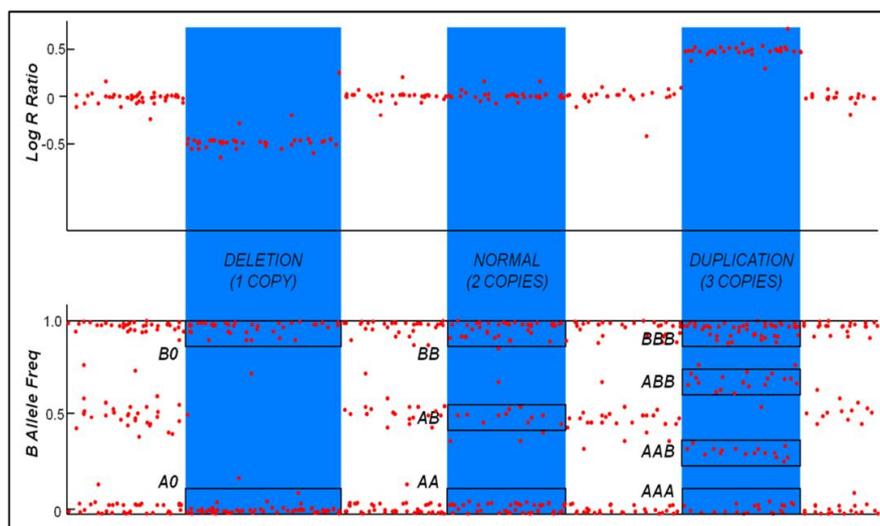


**Figure 1.9** (LaFramboise, 2009). **Overview of SNP array technologies.** A genomic region containing the SNP A/C is shown at the top. **(a) Affymetrix** assay: each probe (25-nt long) targets either allele A or allele B of each SNP interrogated. The location of the SNP locus varies from probe to probe. The DNA of interest binds to the complementary probes on the array, regardless of the allele it carries. However, it does so more efficiently when it is complementary to all 25 bases rather than mismatching the SNP site (indicated by a less bright yellow signal). **(b) Illumina** BeadArray: there are 50-mer probes consisting of a sequence complementary to sequence adjacent to the SNP site. A single-base extension (T or G) with labelled nucleotides results in a appropriated-colour signal (red or green, respectively).

In contrast with array-CGH, in SNP arrays the signal intensities are not compared with those of a reference sample, but they are directly proportional to DNA quantity: an aberrant, increased or decreased, signal intensity means a copy number variation.

The intensity measures are converted into genotype inferences (AA, AB or BB) by a computational analysis of the raw signal data (**figure 1.10**). The inferences regard the presence or absence of each of the two alleles of the SNP, which are labeled as A and B. Each individual usually inherits one copy of each SNP position from each parent, therefore the genotype at a SNP site is typically either AA (homozygous for allele A), AB (heterozygote) or BB (homozygous for allele B).

The computational algorithms that have been developed to convert the signal intensities into genotypes use two measures, the Log R Ratio (LRR) and the B Allele Frequency (BAF). LRR is a normalized measure of the total signal intensity at each SNP. In autosomic regions without CNVs (copy number = 2), LRR is ~0. LRR lower than zero may indicate a deletion, LRR>0 a duplication. BAF represents the relative ratio of the fluorescent signals between two probes/alleles (B/A) at each SNP. BAF values range from 0 to 1: BAF close to 1 indicate that the marker is homozygous for allele B, while BAF close to 0 indicate that the marker is homozygous for allele A. Values close to 0.5 indicate a heterozygous genotype AB. Duplicated regions are characterized by intermediate BAF values (between 0.5 and 1 and between 0.5 and 0), correspondent to the genotypes ABB and AAB.



**Figure 1.10.** The figure shows examples of LRR and BAF plots for a deleted region (1 copy, genotype B0 or A0), for a region with 2 normal copies (three possible genotypes for each SNP: AA, AB, BB) and a duplicated region (3 copies, four possible genotypes for each SNP: AAA, AAB, ABB, BBB).

### 1.3.3 PCR-based approaches

The PCR-based assays are useful for screening targeted regions of the genome.

The best established of these approaches is real-time quantitative PCR (qPCR), that allows to confirm the presence of a predicted deletion or duplication. In a qPCR assay a DNA-binding dye (such as SYBR Green) binds to DNA, causing fluorescence of the dye. An increase in DNA product

during PCR leads to an increase in fluorescence intensity and is measured at each cycle, thus allowing DNA concentrations to be quantified. The comparison between a target region and a region with known copy number determines whether there is a gain or loss. Alternative methods to qPCR for the simultaneous interrogation of multiple regions include Multiplex Ligation-Dependent Probe Amplification (MLPA) (Schouten et al., 2002) and Quantitative Multiplex PCR of Short Fluorescent Fragments (QMPSF) (Charbonnier et al., 2000).

Long range PCR can instead be used in order to accurately characterize the breakpoints of a CNV. For the deletions, primers spanning the CNV breakpoints generate a PCR product shorter than the expected for the reference genomic sequence and, if resolution is sufficient, this allows the fine mapping of the boundaries. For copy number gains instead, this approach is complicated by the two possible orientations of the duplicated region.

### **1.3.4 Next Generation Sequencing-based approaches**

Despite progress in detection of human variation, the nucleotide resolution architecture of most structural variants remains unknown. The detailed characterization of structural variation requires the knowledge of precise sequences. The recent introduction of *Next Generation Sequencing* technologies (NGS) has enabled the mapping of CNVs at fine scale.

Massively parallel sequencing platforms, such as the Roche 454 System and the Illumina HiSeq System, are able to perform sequencing of millions of small DNA fragments in parallel, providing an unprecedented increase in DNA sequencing throughput. Indeed these technologies produce high-quality short “reads” (from 25 to 500 bp in length), which are substantially shorter than the ones that can be obtained by the capillary-based sequencing technology (500 pb-700 bp in length). Moreover the total number of base pairs sequenced in a given run is orders of magnitude higher.

These newer approaches use various strategies that rely on a combination of template preparation, sequencing and imaging, genome alignment and assembly methods. The unique combination of specific protocols distinguishes one technology from another and determines the type of data produced from each platform (Metzker, 2010).

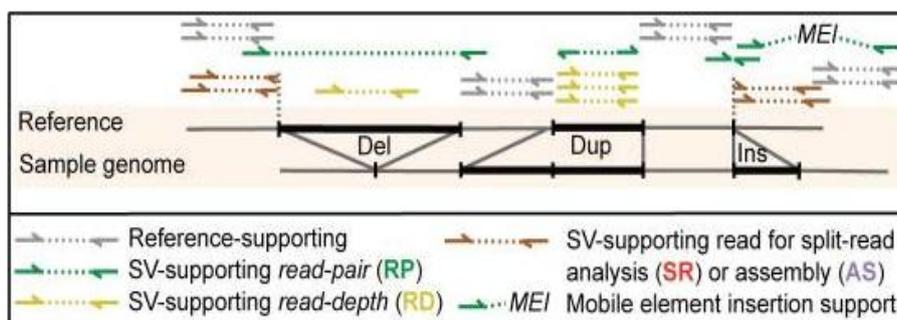
Moreover, the ability to sequence the whole genome of many related organisms has allowed to perform large-scale comparative and evolutionary studies and to understand how genetic differences affect health and disease. Several projects, aimed at sequencing more individuals, have arisen. In particular the 1000 Genomes Project is the first project to sequence the genomes of a large number of people, combining the data from 2500 samples in order to provide a comprehensive resource on human genetic variation. Based on whole genome DNA sequencing data from 185

human genomes, the Structural Variation Analysis Group of the 1000GP has recently generated a comprehensive map of human genetic variation, integrating evidence from complementary sequence-based approaches with extensive experimental validations (Mills et al., 2011).

One of the main advantages of the sequence-based approaches is the base-pair resolution of the CNVs detected, which also enables the identification of the breakpoint position. Analysis of DNA region motifs surrounding the breakpoints allows to hypothesize the formation mechanism.

These sequence-based approaches include (**figure 1.11**):

- *Paired-end mapping* (or read pair ‘RP’ analysis): it is based on sequencing and analysis of abnormally mapping pairs of clone ends (Kidd et al., 2008; Korbelt et al., 2007; Tuzun et al., 2005) or high-throughput sequencing fragments. When the region spanned by the paired-ends in the sample genome is shorter than the correspondent region in the reference genome, this can indicate a deletion; when it is longer instead, this might indicate a simple insertion.
- *Read-depth* (‘RD’) *analysis*: evaluates the read depth-of-coverage, which is measure by counting the number of reads mapping to a certain genomic window. An increase or a decrease of the normalized read count in a certain region may indicate a gain or a loss, respectively (Yoon, Xuan, Makarov, Ye, & Sebat, 2009).
- *Split-read* (‘SR’) *analysis*: maps the boundaries (breakpoints) of structural variants by sequence alignment (K. Ye, Schulz, Long, Apweiler, & Ning, 2009). When a read spans across the breakpoint of a deletion, this sequence does not map to a single position on the reference genome, but it will be split into two fragments that map separately, indicating the position of the breakpoints.
- *Sequence assembly* (‘AS’), which enables the fine-scale discovery of CNVs, including novel sequence insertions (Hajirasouliha et al., 2010; Simpson et al., 2009), by using algorithms to assemble together the reads that do not map to any region of the genome.



**Figure 1.11** (Mills et al., 2011). **Schematic representation of the different sequence-based approaches to detect CNVs.** The arrows indicate the reads. Different sequence-based CNV-detection approaches are represented by different coloured reads.

In addition to whole-genome sequencing, the NGS platforms can be used to target specific regions of interest, such as all the protein-coding regions of the human genome. Since the exons constitute only approximately 1% of the human genome, but harbor 85% of the mutations with large effects on disease-related traits, sequencing complete coding regions (i.e., “whole exome sequencing”) has the potential to play a major role in disease gene discovery and also in clinical use for establishing a genetic diagnosis (Choi et al., 2009; Risch & Merikangas, 1996).

### **1.4 CNVs and neurodevelopmental disorders**

Neurodevelopmental disorders are characterized by impairment of growth and development of the brain often associated with cognitive, neurological, or psychiatric dysfunction, such as intellectual disability (ID), developmental delay (DD), autism spectrum disorders (ASDs), schizophrenia (SCZ), and bipolar disorder (BD) (Coe, Girirajan, & Eichler, 2012). These different neurological conditions show a high heritability, but they have been proven to have a complex genetic architecture, in which multiple loci contribute to the overall risk.

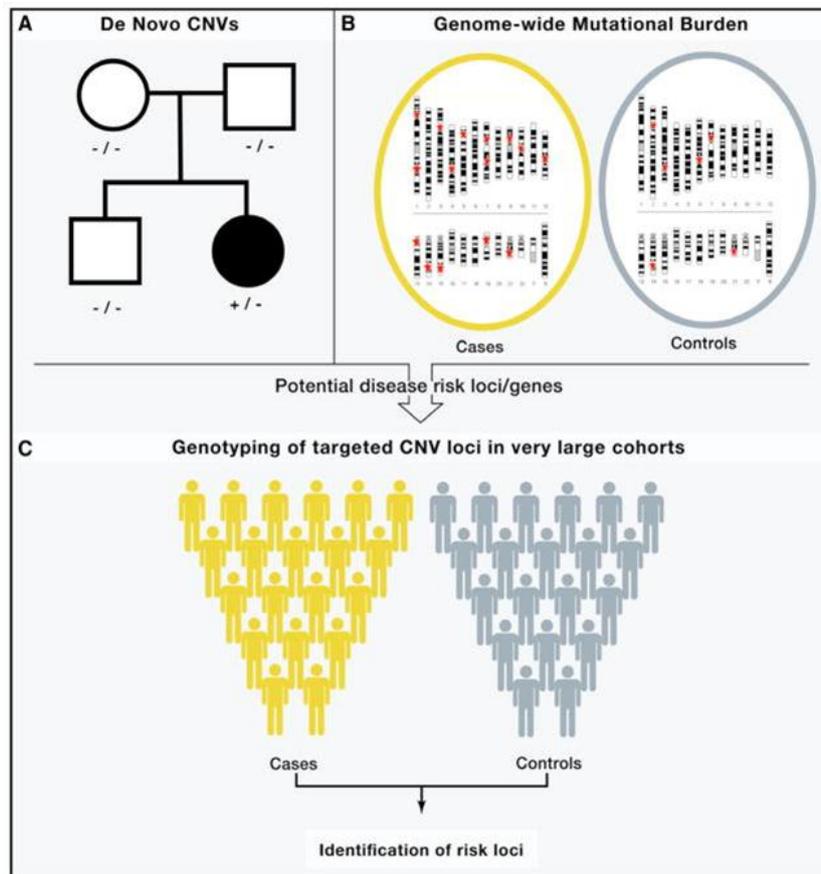
In the past two decades, the mapping of genes underlying these diseases has been focused on two alternative hypotheses: the “common disease-common variants” model (Risch & Merikangas, 1996) and the “common disease-rare variants” model.

Studies attempting to test the first hypothesis have found that common variants confer only a small or moderate level of risk, suggesting that the aetiology of these common diseases might be explained instead by the “common disease-rare variants” model, in which a number of different causes (SNPs or CNVs), each of them with low frequency in the population and typically highly penetrant, could collectively account for a large proportion of attributable risk (McClellan & King, 2010). These two hypotheses have been subsequently integrated in a new multifactorial model, in which common disorders could be result of a heterogeneous set of numerous rare and common variants, with different impact on the phenotype and collectively implicating a large number of different genes.

Detection of CNVs has become an important field of genetic studies of complex disorders, as CNVs can make a substantial contribution to the genetic mechanisms underlying disease susceptibility and, in particular, rarer CNVs have been indicated as a potential source of missing heritability (Manolio et al., 2009).

Within CNVs research, three study designs have been widely used (**figure 1.12**).

- 1) Family-based approach: this approach, which examines CNVs at individuals level, enables the identification of *de novo* CNVs, the determination of their frequency and the association of these mutations with the disorder;
- 2) Case-control analysis of CNV burden: this approach analyses CNVs at population levels and examines whether the cases show a greater CNV genome-wide burden (i.e. the number of CNVs carried by an individual) compared to controls. The comparison of the collective frequency of rare variants (with frequency less than 1% in the general population) between cases and controls allows to investigate the contribution of rare CNVs to the disease.
- 3) Association of Target regions or Genes: this approach analyses the association of specific CNV loci with the disease phenotype in large case control cohorts.



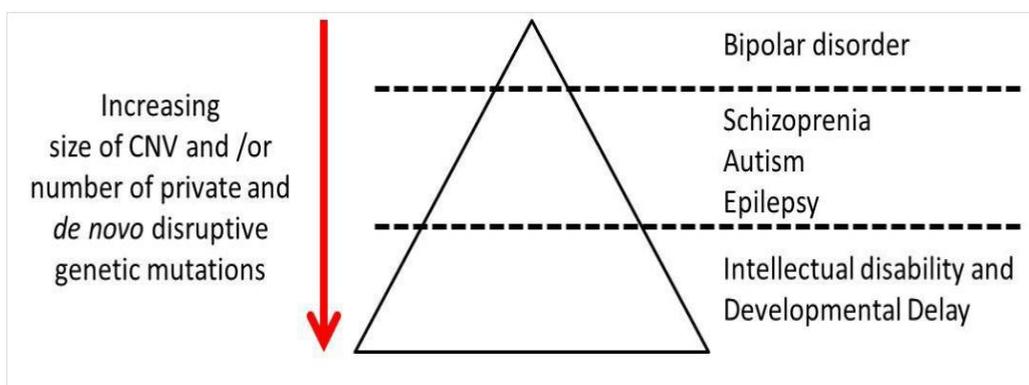
**Figure 1.12** (Molhotra 2012). Genetic approaches for CNV discovery. (A) Family-based studies of de novo mutation and (B) case control studies of genome-wide CNV burden, with CNV positions denoted by a red star (C) followed by single marker tests for association in large cohorts.

Several interesting themes have begun to emerge from CNV studies in psychiatric and neurodevelopmental diseases.

Studies of large cohorts have revealed several highly penetrant loci associated with these disorders and a significant increase in large rare/*de novo* CNV burden in affected individuals compared to unaffected siblings and controls. In particular, in autism and schizophrenia, burden analyses have shown a general enrichment of potentially pathogenic duplications in a larger size range (>500 kb), and of deletions in smaller size range (30-500 kb) (Pinto et al., 2010).

CNVs larger than 500 kb usually contain multiple genes and can be found in a small proportion of the general population (~8%), suggesting that these large rearrangements are under purifying selection and they are likely to be potentially pathogenic. Small CNVs instead are more frequent in the general population and this complicates the identification of pathogenic CNVs; however small CNVs can also represent important risk factors and the investigation of their role needs future studies with sufficient power in sample size and resolution. Moreover, it has been hypothesized that *de novo* CNVs might contribute to risk of sporadic forms of these disorders, since *de novo* mutations seem to have a stronger and more robust effect size compared to inherited CNVs. In particular, family-based studies have showed that *de novo* CNVs represent a contributing factor in 5-10% of ASD patients (D. Levy et al., 2011; Pinto et al., 2010; Sanders et al., 2011), 5-10% of schizophrenia cases (Kirov et al., 2012; Malhotra et al., 2011) and 4.3% of individuals affected by bipolar disorder (Malhotra et al., 2011).

The common theme amongst the CNV studies is the fact that the size and the rate of rare/*de novo* CNVs seem to correlate with the severity of the phenotype. The highest burden of large rare/*de novo* CNVs has been observed in cases with intellectual disability (ID) and dysmorphic features (Girirajan et al., 2012), the lowest burden in bipolar cases, in between the extremities are schizophrenia and autism. This trend seems to support a model where neurodevelopmental disorders, based on their severity and co-morbidities of ID, are considered part of a continuum (Figure 1.13).



**Figure 1.13** (Coe et al., 2012). **An oligogenic model for neurodevelopmental disorders.** In this model, a higher number of large, rare, *de novo* CNVs and, more in general, disruptive genetic mutations correlates with an increase in the severity of the clinical phenotype.

Another interesting observation is that a CNV locus can be recurrently identified in association with a variety of multiple neurological phenotypes (*pleiotropic effect*). In some cases, a CNV is necessary and sufficient to result in a specific phenotype: these CNVs are associated with known syndromes (“syndromic CNVs”) and most often they occur in patients with moderate-to-severe ID. Examples are the 17q21.31 microdeletion syndrome and the Williams syndrome deletion on chromosome 7q11.23. By contrast, there are CNVs which are much more variable in their outcome. These loci exhibit a *variable expressivity*, which means that individuals carrying the same CNV show either a qualitative or quantitative phenotypic variation, and a *reduced penetrance*, as these CNVs can be identified also in asymptomatic carriers.

Therefore, these recurrent CNVs often show an incomplete segregation within multiplex families and are likely to be inherited from a parent, who may present any of the phenotypic manifestations associated with the CNV or have normal phenotype. Some of these CNV hotspots are 16p11.2, 15q13.3, 1q21.2, 3q29, 17q21.31 and 7q11.23 loci (Coe et al., 2012).

Moreover, losses and gains occurring at the same locus can lead to drastically different phenotypes or, instead, to surprisingly overlapping phenotypes. An example of different outcomes associated with CNVs at the same locus is given by the 16p11.2 region. Both the deletions and duplications in this locus have been observed in multiple conditions with significant enrichment compared to healthy controls. However, while deletions are associated with more severe phenotypes, including cases with autism, ID, macrocephaly, dysmorphic features, and obesity (Walters et al., 2010), the reciprocal duplications are seen in a wider range of conditions, including clinically underweight cases, schizophrenia and microcephaly. Similarly, duplications of 1q21.1 locus were identified in patients with ASD, whereas both deletions and duplications were associated with mental retardation. The 15q13.3 microdeletion is also detected across multiple phenotypes. In particular deletions in this locus have been identified in individuals with idiopathic generalized epilepsy (Helbig et al., 2009), ID (Sharp et al., 2008), autism and schizophrenia. Instead the reciprocal duplication shows a more variably expression across neurological phenotypes, such as ID, bipolar disorder and autism, and its role in pathogenicity remains unclear (Szafranski et al., 2010).

These examples suggest that differences between clinical phenotypes caused by reciprocal CNVs (duplication/deletion) can be attributed in part to the variable dosage sensitivity of the genes encompassed by the CNV: for some genes, an increased or a reduced dosage can determine opposite consequences (“mirror” phenotype); for other genes, the dosage imbalance alters certain cellular functions, irrespectively of the type of copy-number change. Moreover, the variable expressivity observed among the cases carrying the same CNV might be determined by secondary

“hits”, including additional pathogenic CNVs or damaging sequence mutations and by differences in the genetic background and epigenetic regulation. However, it is worth noting that this complex variability in phenotypic expression is not limited to the hotspots loci but might be true also for other genomic regions.

Another interesting theme that emerged from CNVs analyses is that CNVs implicated in neurodevelopmental disorders affect genes that converge on common pathways, suggesting that there are some key genes important for several aspects of brain development and that, if mutated, can contribute to a range of neurological conditions, depending on the genetic background.

Therefore, genic CNVs can be an important source for the discovery of the genes contributing to complex disorders and of pathways implicated in complex disorders.

Interestingly, whole-genome or exome studies are very fruitful to identify particular candidate genes. Moreover, network analyses have allowed to estimate the amount of genes involved in neurodevelopmental disorders, confirming heterogeneity underlying these complex diseases. In the next years, the gene discovery effort is expected to be facilitated by the increasing availability of high-throughput CNV and sequencing data. The integration of structural and sequence information will allow the capture of a larger fraction of the rare disease-causing variants and to explain a larger proportion of the risk.

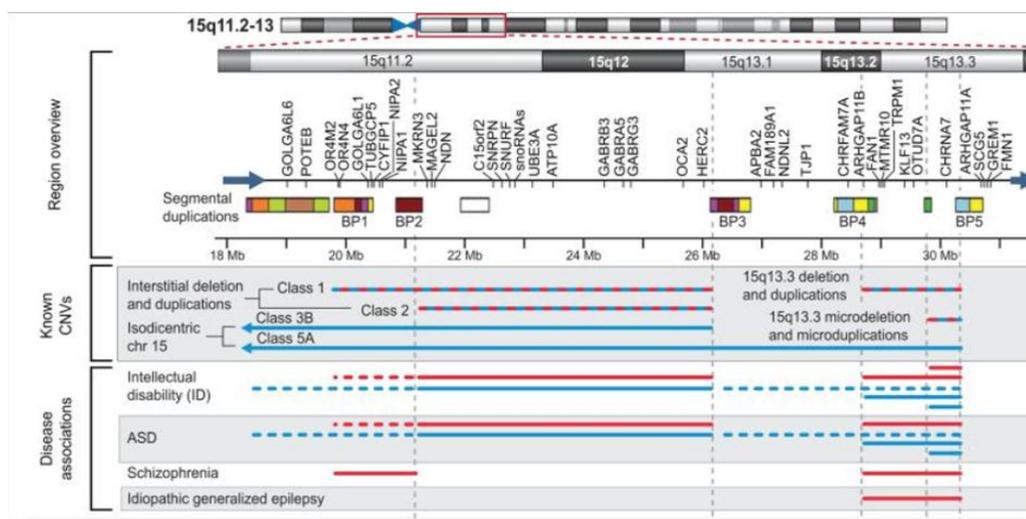
### **1.5 Recurrent CNVs on chromosome 15q11-13.**

The proximal region of the long arm of chromosome 15 (15q11-q13) is a well-known hotspot for CNVs, due to the presence of complex patterns of highly homologous LCRs, that make this locus one of the most unstable regions in the human (Makoff & Flomen, 2007; Toth-Fejel et al., 1995). CNVs in this region include deletions, duplications, translocations, inversions and, also, supernumerary inv-dup(15) chromosomes (Jauch, Robson, & Smith, 1995; Schinzel et al., 1994). Deletions and duplications are likely to be caused by NAHR between the LCRs. The breakpoints of these CNVs are localized within clusters of LCRs, that have been designated BP1-BP5 (**figure 1.14**).

The structural and sequence features of the 15q11-13 region can lead to a number of different rearrangements:

1. BP1-BP3 deletions or BP2-BP3 deletions can result in Prader-Willi Syndrome (PWS), if the deletion is inherited from the father, or Angelman Syndrome (AS), if the deletion is inherited from the mother (Amos-Landgraf et al., 1999).

2. Duplications instead can be associated with learning disabilities, seizures and autism. The 15q11-q13 maternal duplication is the most frequent cytogenetic cause of autism, which occurs in 1-3% of individuals with ASD (Veenstra-VanderWeele & Cook, 2004). This duplication includes imprinted genes, therefore the effects of mutations in this region depend on the origin (maternal or paternal) of the chromosome in which they occur.
3. BP1-BP2 microdeletions and microduplications have been proposed as risk factors for a range of neurological problems, in particular, language delay and developmental delay (Doornbos et al., 2009), ID, ASD (Sanders et al., 2011; van der Zwaag et al., 2010), schizophrenia (Kirov et al., 2012; Stefansson et al., 2008) and epilepsy (de Kovel et al., 2010), but they have been observed also in controls .
4. recurrent large BP4-BP5 deletions and duplications have been found across multiple conditions and also in asymptomatic carriers. However, BP4 and BP5 have a complex organization and can lead also to smaller microdeletions and microduplications (350-680 kb), which have also been identified in a similar range of neuropsychiatric phenotypes (Szafranski et al., 2010).



**Figure 1.14** (Sanders et al., 2011) The genomic architecture of the BP1-BP5 region on chromosome 15q. The position of each BP is indicated. Class 1 indicates BP1-BP3 deletions, class 2 instead indicates the BP2-BP3 deletions.

The typical rearrangements occurring between BP4 and BP5 are CNVs with a size of ~1.6 Mb and encompass six RefSeq genes (*MTMR15*, *MTMR10*, *TRPM1*, *KLF13*, *OTUD7A* and *CHRNA7*) and the microRNA gene *hsa-mir211*. In particular, the BP4-BP5 deletion has been reported in cases of mental retardation with seizures (Sharp et al., 2008), autism (Miller et al., 2009; Pagnamenta et al., 2009), schizophrenia (Stefansson et al., 2008), bipolar disorder, epilepsy (de Kovel et al., 2010) and language delay (Ben-Shachar et al., 2009).

COPY NUMBER VARIANTS

A recent study (Moreno-De-Luca et al., 2013) performed on a group of 31,516 cases (including patients with developmental delay, ID, ASD or multiple congenital abnormalities) has found statistical support for a pathological role for several CNV loci, including the BP4-BP5 region (Table 1.1 and Table 1.2). The BP4-BP5 deletion was detected in 88 cases, whereas the BP4-BP5 duplication was identified in 34 cases and 5 controls. Both CNV types were reported to have a statistical significant increase in cases compared to controls. However, while the deletion appears to have a complete penetrance, the reciprocal microduplication has a less certain clinical significance in comparison with the deletions. The clinical uncertainty of the 15q13.3 duplications could be due to the fact that a larger sample size is necessary to detect a low penetrant effect.

**Table 1.** Deleterious recurrent deletions in clinical cohorts

Deletion region	Syndrome	Coordinates (Mb)	Cases (31 516)	Frequency	Controls (13 696)	OR	P
<i>Complete penetrance</i>							
22q11.2	DiGeorge/Velo-cardio-facial	chr22:17.4–18.67	189	1 in 167	0	∞	$2.2 \times 10^{-16}$
15q13.2-q13.3 (BP4-BP5)		chr15:28.92–30.27	88	1 in 358	0	∞	$2.53 \times 10^{-14}$
7q11.23	Williams–Beuren	chr7:72.38–73.78	76	1 in 415	0	∞	$1.49 \times 10^{-12}$
15q11.2-q13 (BP2-BP3)	Angelman/Prader–Willi	chr15:22.37–26.1	57	1 in 553	0	∞	$1.79 \times 10^{-9}$
17q21.31		chr17:41.06–41.54	45	1 in 700	0	∞	$1.95 \times 10^{-7}$
17p11.2	Smith–Magenis	chr17:16.65–20.42	32	1 in 985	0	∞	$1.79 \times 10^{-5}$
22q11.2 (distal)		chr22:20.24–21.98	26	1 in 1 212	0	∞	$1.32 \times 10^{-4}$
8p23.1		chr8:8.13–11.93	17	1 in 1 854	0	∞	$2.88 \times 10^{-3}$
5q35	Sotos	chr5:175.65–176.99	16	1 in 1 970	0	∞	$4.81 \times 10^{-3}$
3q29		chr3:197.23–198.84	15	1 in 2 101	0	∞	$8.40 \times 10^{-3}$
10q23		chr10:81.95–88.79	14	1 in 2 251	0	∞	$8.20 \times 10^{-3}$
17q11.2	Neurofibromatosis type 1	chr17:26.19–27.24	13	1 in 2 424	0	∞	0.01
<i>Incomplete penetrance</i>							
16p11.2		chr16:29.56–30.11	131	1 in 241	7	8.16	$2.25 \times 10^{-13}$
1q21.1		chr1:145.04–145.86	102	1 in 309	4	11.11	$7.18 \times 10^{-12}$
16p12.1		chr16:21.85–22.37	54	1 in 584	5	4.7	$9.17 \times 10^{-5}$
16p13.11		chr16:15.41–16.2	40	1 in 788	4	4.35	$1.46 \times 10^{-3}$
17q12	Renal cysts and diabetes	chr17:31.89–33.28	32	1 in 985	2	6.96	$1.09 \times 10^{-3}$
1q21(TAR)	Thrombocytopenia-absent radius	chr1:144–144.34	30	1 in 1 051	3	4.35	$6.94 \times 10^{-3}$
16p11.2 (distal)		chr16:28.68–29.02	23	1 in 1 370	2	5	0.01
Total deletions			1 000	1 in 32			

Abbreviations: BP, breakpoint; OR, odds ratio; Mb, megabase. All coordinates are given in hg18. CNVs are ordered according to frequency in clinical collections.

**Table 1.1** (Moreno-De-Luca et al., 2013). Deleterious recurrent deletions in clinical cohorts.

**Table 2.** Deleterious recurrent duplications in clinical cohorts

Duplication region	Syndrome	Coordinates (Mb)	Cases (31 516)	Frequency	Controls (13 696)	OR	P
<i>Complete penetrance</i>							
15q11.2-q13 (BP2-BP3)		chr15:22.37–26.1	62	1 in 508	0	∞	$2.38 \times 10^{-10}$
7q11.23		chr7:72.38–73.78	32	1 in 985	0	∞	$1.79 \times 10^{-5}$
17p11.2	Potocki-Lupski	chr17:16.65–20.42	24	1 in 1 313	0	∞	0.00
8p23.1		chr8:8.13–11.93	13	1 in 2 424	0	∞	0.01
22q11.2 (distal)		chr22:20.24–21.98	11	1 in 2 865	0	∞	0.04
<i>Incomplete penetrance</i>							
22q11.2		chr22:17.4–18.67	82	1 in 384	5	7.14	$2.69 \times 10^{-8}$
16p11.2		chr16:29.56–30.11	67	1 in 470	4	7.29	$4.19 \times 10^{-7}$
1q21.1		chr1:145.04–145.86	54	1 in 584	4	5.87	$2.17 \times 10^{-5}$
17q12		chr17:31.89–33.28	39	1 in 808	6	2.83	0.01
15q13.2-q13.3 (BP4-BP5)		chr15:28.92–30.27	34	1 in 927	5	2.96	0.01
16p11.2 (distal)		chr16:28.68–29.02	25	1 in 1 261	3	3.62	0.02
Total duplications			443	1 in 71			

Abbreviations: BP, breakpoint; OR, odds ratio; Mb, megabase. All coordinates are given in hg18. CNVs are ordered according to frequency in clinical collections.

**Table 1.2** (Moreno-De-Luca et al., 2013). Deleterious recurrent duplications in clinical cohorts

In conclusion, BP1-BP2 and BP4-BP5 CNVs can be observed in a wide spectrum of clinical phenotypes of variable severity and some of them, such as BP1-BP2 microdeletions and BP4-BP5 microduplications, are present also in healthy individuals, suggesting that they have variable expressivity and incomplete penetrance and are probably not sufficient to cause pathological phenotypes (Leblond et al., 2012; van Bon et al., 2009). However, since they involve interesting candidate genes and have a higher frequency in clinical cohorts, they might contribute to the susceptibility to certain neuropsychiatric disorders in specific genetic backgrounds, where secondary alterations could have an additive or epistatic effect.

## CHAPTER 2:

### AUTISM SPECTRUM DISORDER

#### 2.1 Autism Spectrum Disorder

The word “autism” was first used in 1943 by psychiatrist Leo Kanner (Kanner, 1968) who described 11 children, mostly boys, with a combination of severe social and variable language dysfunction and the presence of repetitive restrictive behaviors. Kanner’s reuse of *autism* led to decades of confused terminology like *infantile schizophrenia*, and it was suggested that autism may be related to a “genuine lack of maternal warmth”. The “refrigerator mother” theory was then rejected and, starting in the late 1960s, autism was established as a distinct syndrome, distinguishing it from intellectual disability and schizophrenia and from other developmental disorders, and demonstrating the benefits of involving parents in active programs of therapy (Fombonne, 2003).

A major change in perspective came with the pioneering twin studies of Rutter and Folstein that demonstrated a genetic susceptibility to the disorder (Folstein & Rutter, 1977a, 1977b). Over the past two decades, the concept of autism has extended from the strict diagnosis of autistic disorder to include a group of related lifelong neurodevelopmental conditions, called Autism Spectrum Disorders (ASDs) (**figure 1.1**). According to the fourth edition of the Diagnostic and Statistical Manual of Mental disorders (DSM-IV) ([APA], 1994), the ASDs are characterized by delays or abnormal functioning in one or more of the following three domains: social interaction; social communication and restricted and repetitive behaviours and interests, and they include four disorders:

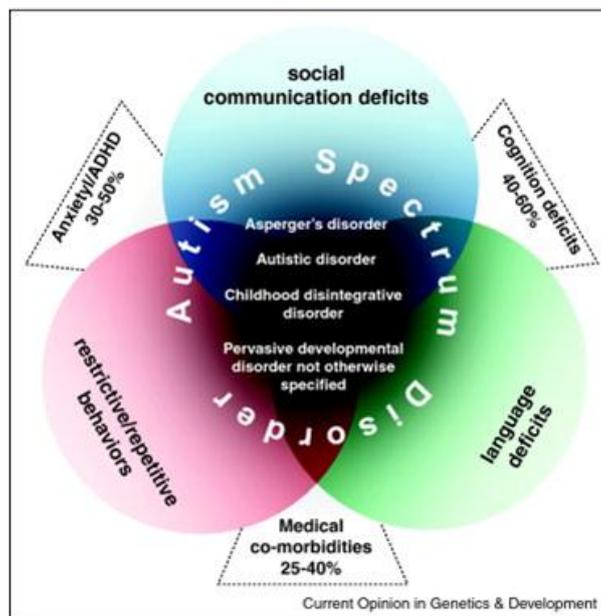
- 1) Autism, which presents deficits in communication, social interactions and repetitive behaviours;
- 2) Asperger’s disorder, which is characterized by the absence of clinically significant delay in language and cognitive development;
- 3) Childhood Disintegrative Disorder (CDD), which typically occurs later than autism and involves a more dramatic loss of skills (regression);
- 4) Pervasive Development Disorder-not otherwise specified (PDD-NOS), which presents sub-threshold symptoms and/or later onset.

These neurological conditions differ in the severity and in the pattern of the core symptoms, developmental course, and cognitive and language abilities. Therefore, given the absence of a clear

discrimination, both etiological and clinical, between these categories, in the last version of manual, the DSM-V ([APA], 2013), the category of Asperger syndrome was removed and the diagnostic criteria for autism were modified under the new heading of Autism Spectrum Disorder (ASD). The word “spectrum” refers to the decomposition of light through a prism, in which multiple dimensions are present, indicating the heterogeneity underlying these disorders.

Therefore, individuals with ASD are best represented as a single diagnostic category because they show similar types of symptoms and are better differentiated by clinical presentations (i.e., severity, verbal ability) and by associated features (i.e., known genetic disorders, epilepsy and intellectual disability). An individual with an ASD diagnosis will be described in terms of severity of social communication symptoms, severity of fixated or restricted behaviors or interests and associated features.

ASD has an estimated prevalence of ~60/10,000 individuals (Elsabbagh et al., 2012) and a male to female gender bias, with a ratio of ~4:1 (E. Williams, Thomas, Sidebotham, & Emond, 2008).



**Figure 2.1** (Devlin & Scherer, 2012). Autism phenotypic classification. Autism is the prototypic form of a group of conditions, collectively called autism spectrum disorders (ASD), which share defects in three core domains: social communication; language development and restricted and repetitive behaviours and interests).

## **2.2 Clinical phenotype and diagnostic instruments**

Autism is a developmental neuropsychiatric syndrome characterized by onset prior to 3 years of age and by a triad of behavioral signs and symptoms including the following domains:

- 1) social interaction: people with autism have social impairments which become apparent early in childhood. They show lack of interest, tendency to isolation, less attention to social

stimuli and smile. They also have difficulties to establish a direct visual contact, to start a conversation and to answer questions or to their own name.

- 2) communication: about 50% of autistic children not develop any verbal language or show a marked delay in the development of spoken language (Hus, Pickles, Cook, Risi, & Lord, 2007). Impairments in communication include babbling, unusual gestures which are not integrated with words, vocal patterns that are not synchronized with the caregiver, deficits in joint attention and imaginative play. They often repeat others words, expressions or sounds (echolalia) and they may refer to themselves with the own name.
- 3) restricted interests and repetitive behavior: autistic individuals display repetitive movements, such as hand flapping, head rolling, or body rocking (stereotypy); compulsive behaviors, like arranging objects in stacks or lines; ritualistic behaviors which involve a invariable pattern of daily activities, such as an unchanging menu or a dressing ritual; restricted behaviors or movements which cause self-injury, such as eye-poking, skin-picking, hand-biting and head-banging.

However, it should be noted that autism shows a high heterogeneity of symptomatology. Rare autistic cases with an extraordinary capacity for mathematical calculation, musical sensitivity, exceptional audio-visual memory or other unusual abilities, are reported. Autism presents also a high variation in cognitive skills, ranging from profound mental retardation to cognitive levels high or higher than normal. Moreover one ASD child out of four develops seizures, which usually occur in early childhood or adolescence (Zappella, 2010), while others may show motor retardation, dysmorphisms, gastrointestinal problems and other co-morbidities.

Two diagnostic instruments commonly used in autism diagnosis are the Autism Diagnostic Interview-Revised (ADI-R) (Lord, Rutter, & Le Couteur, 1994) and the Autism Diagnostic Observation Schedule–Generic (ADOS-G) (Lord et al., 2000). The ADI-R is a standardized interview conducted with caregivers of autistic individuals. The questionnaire explores the areas of communication, social interactions, restricted and repetitive behaviours and developmental history. An algorithm was generated to make a standard diagnosis on the basis of the ADI-R scores obtained in each area and it includes only the items that more closely depicted the phenotypic abnormalities described in the Diagnostic and Statistical Manual of mental disorders-4th edition (DSM-IV) ([APA], 1994). The algorithm specifies a cutoff score for each of three core domains of autism, therefore, only the individuals who meet all the cutoffs, meet diagnostic criteria for autism, the most severe form of ASD. The ADOS-G is an interactive test that aims to assess social interactions,

communication, play and spontaneous behaviours in a standardized context. ADOS-G is an implementation of ADOS (Lord et al., 1989), which was proposed as a complementary instrument to ADI. As for ADI-R, subsets of items in each module of ADOS-G were selected to generate diagnostic algorithms. Classification is made on the basis of exceeding cutoffs in social behaviour, communication and social-communication totals.

### **2.3 The genetic basis of ASD**

Twin and family studies have indicated a strong genetic basis for ASD susceptibility.

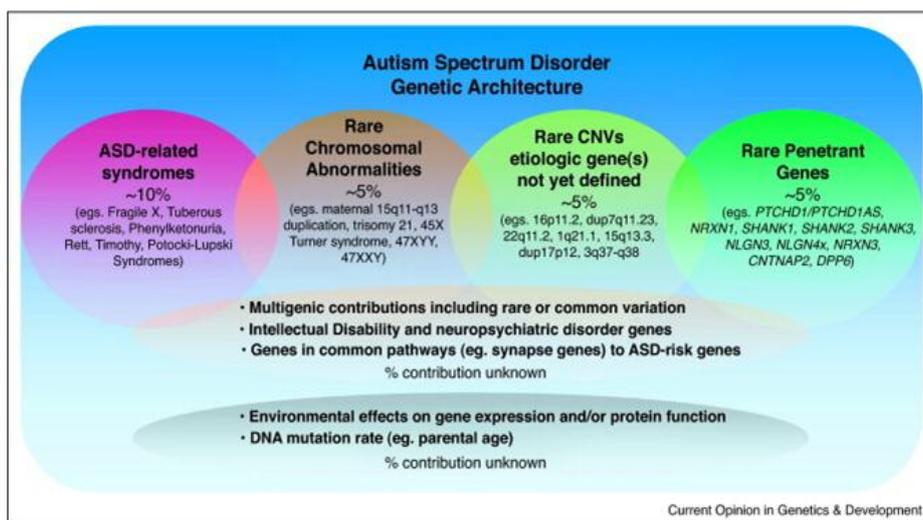
The first twin studies observed concordance rates of 82-92% in monozygotic twins (MZ) compared with 1-10% in dizygotic twins (DZ) (Bailey et al., 1995; Folstein & Rutter, 1977b). A higher disease co-occurrence in MZ twins than DZ twins supports a genetic etiology, because MZ share 100% of their genetic material and dizygotic twins (DZ) share 50% (similar to non-twin siblings) and both share the in utero environment with their twin. However, more recent studies have indicated that the concordance in dizygotic twins might be higher (>20%), suggesting that shared environment could play a larger role than had been previously estimated (Hallmayer et al., 2011). Concordance rates between MZ twins do not take into account the genetic factors that may differ in co-twins, such as epigenetic factors, X-inactivation and mutations *de novo* arisen after the separation of the embryos. Therefore, these genetic and epigenetic differences can produce phenotypic differences between MZ which are not due to environmental factors, but that can result in a decrease of estimated heritability.

A further support for a strong genetic background in ASD comes from family-based studies, which have found that first-degree relatives of autistic probands have a markedly increased risk for autism compared with the population, as they show behavioral or cognitive features which are similar to ASD proband (such as social or language dysfunction), but in lesser forms. These mild forms of impairments, usually affecting only one of the core domains, are classified as “broader phenotypes” (Losh et al., 2009). The recurrence risk in families with one child with ASD was initially estimated at 5%, compared with 1% in the general population, but recent studies investigating early signs of ASD in siblings of individuals with ASD revealed an even higher recurrence rate (up to 20%) (Hallmayer et al., 2011; Ozonoff et al., 2011). These sub-threshold autistic traits are highly heritable and show a continuous distribution in the population, suggesting that different features of autism represent a quantitative continuum of function that could be inherited in distinct patterns.

However, despite high heritability of ASD, its genetic architecture is complex, as several different genes are involved in the disorder (Betancur, 2011) and the exact underlying causes are identifiable only in a minority of patients.

The first genetic causes of ASD (**Figure 2.2**) were identified in known single-gene disorders, accounting for approximately 10% of individuals with ASD, with fragile X syndrome being the most common (1%-3% of cases), followed by *PTEN* macrocephaly syndrome, tuberous sclerosis and Rett syndrome (each accounting for approximately 1% of children diagnosed with autism) (Miles, 2011). In addition to this syndromic forms of autism, rare cytogenetically visible chromosome rearrangements may account in about 5% of individuals with ASD. Among these cytogenetic abnormalities, the maternal 15q11-q13 duplication is the most common, detected in 1-3% of cases (Baker, Piven, Schwartz, & Patil, 1994).

On the other hand, recent findings have shown that some rare, highly penetrant mutations in several ASD candidate genes, including *NLGN3*, *NLGN4* (Jamain et al., 2003; Laumonnier et al., 2004), *SHANK1* (Sato et al., 2012), *SHANK2* (Berkel et al., 2010), *SHANK3* (Durand et al., 2007), *NRXN1* (Ching et al., 2010), *NRXN3* (Vaags et al., 2012), *DPYD* (Carter et al., 2011), *DPP6* (Marshall et al., 2008) and *CNTNAP2* (Bakkaloglu et al., 2008) could collectively account for a large proportion of risk (5% of ASD cases). Moreover, recent high-throughput screenings have begun to uncover a large number of individually rare submicroscopic structural variants (Copy Number Variants, CNVs), both *de novo* and inherited, potentially contributing to the ASD susceptibility in about 5% of individuals with ASD (Devlin & Scherer, 2012). (For a detailed discussion of Copy Number Variants see Chapter 1).



**Figure 2.2** (Devlin and Scherer, 2012). **Genetic architecture in ASD.** Rare genetic risk factors and their estimated contribution to ASD are represented in four groupings. Additional rare and common genetic variants may modify clinical presentation or operate under a threshold model. Genetic contributions to ASD can also arise from direct or indirect effects on genes and proteins by environmental influences.

## **2.4 Molecular genetic studies of ASD susceptibility**

In order to identify the genetic risk factors conferring susceptibility to ASD, linkage and association studies have been performed. Both approaches are based on the genotyping and analysis of genetic markers, which are polymorphic variants with a known position in the genome (e.g. SNPs or microsatellites).

### **2.4.1 Linkage studies in ASD**

Linkage studies are based on the principle that polymorphic genetic markers located in proximity of disease-causing variants co-segregate with the affection status, across generations: the linkage between these markers and the disease-causing variants is detected when they are transmitted together to the offspring more often than expected under independent inheritance. Their probability of being separated by a crossing over during meiosis is proportional to their distance on the chromosome. These studies involve related individuals and can include large pedigrees (with extended families and/or multiple generations) or a large number of small nuclear families (consisting of a father, a mother, and their children).

Linkage studies usually identify large regions of susceptibility (in the order of Mb), encompassing a large number of genes, and their success is affected by genetic and phenotypic heterogeneity of the disorder.

Several whole-genome linkage studies have been performed for ASD and, although many loci have been implicated, replicated regions between samples are rare, reflecting the extensive heterogeneity underlying the disorder and the likely small effect size attributable to single genes. The first regions linked to autism were identified by the International Molecular Genetic Study of Autism Consortium ((IMGSAC), 2001; Maestrini et al., 2010) on chromosomes 7q (designated AUTS1, 7q21-q32, OMIM 209850) and 2q (designated AUTS5, 2q24-q33, OMIM 606053) and these have been also the most consistently replicated loci (Badner & Gershon, 2002; Buxbaum et al., 2001; Schellenberg et al., 2006; Shao et al., 2002; Trikalinos et al., 2006).

A possible strategy to increase the chances of identifying contributory risk genes in a context of high heterogeneity is to analyse a large sample of multiplex families. However, although linkage studies with higher resolution and larger cohorts of ASD families (Szatmari et al., 2007; Weiss, Arking, Daly, Chakravarti, & Consortium, 2009) have been performed, they showed that the linkage regions described above failed to reach genome-wide significance. In the study realized by the Autism Genome Project (Szatmari et al., 2007) using Affymetrix 10K SNP arrays and 1,168 families with at least two affected individuals, suggestive linkage was obtained only for a region on

chromosome 11p12-p13. The samples were also stratified in categories, but, even in these subsets, the loci on 7q and 2q reached only suggestive evidence of linkage in the individuals of European ancestry, confirming the genetic heterogeneity underlying the ASD and hindering the identification of genetic cause.

### 2.3.2 Association studies in ASD

Association studies aim to find an allelic association between specific genetic variants and the disease phenotype in a population, assuming that the marker itself or a variant close to it confers susceptibility to the disease. When an allelic variant is associated with a trait, this result can be interpreted as:

- a) a direct association, if the variant has a causal role in the phenotype susceptibility;
- b) an indirect association, if the variant is in Linkage Disequilibrium with the causal variant (LD is a phenomenon arising from alleles at linked loci, that tend to co-segregate more often than expected by chance and forming “haplotypes blocks”);
- c) a false positive, that may be due to chance or to problems such as population stratification or inappropriate statistical methods.

Generally, association studies require a higher density of markers than linkage studies, but they identify candidate chromosome regions with a better resolution.

Results of association studies are usually expressed as P values or  $-\log_{10}(P)$ : very low P values provide strong evidence for association. The significance threshold depends on the number of markers: as the number of markers increases, the required number of tests increases and the significance threshold becomes more stringent.

International projects, such as the HapMap Project (<http://hapmap.ncbi.nlm.nih.gov/>) and the 1000 Genome Project (<http://www.1000genomes.org/>) have been a useful resource for association studies, providing a deep characterization on human genetic variation in different populations and determining its Linkage Disequilibrium structure. Moreover, the advent of microarrays and the availability of sufficiently large samples have allowed to perform genome-wide association studies (GWAS).

Several GWAS in large cohorts of autistic individuals have been carried out (Anney et al., 2012; Anney et al., 2010; Wang et al., 2009; Weiss et al., 2009). These studies identified genome-wide significant association signals in some genomic regions, suggesting putative molecular pathways with a crucial role in ASD pathogenesis. For example, an altered functional connectivity in frontal lobe circuits has been associated with common genetic SNP variation in *CNTNAP2* (Scott-Van

Zeeland et al., 2010), a putative ASD risk gene identified through the study of rare variants. However, none of these studies has replicated any of the previous linkage or association findings, supporting the hypothesis that only a few common variants have a substantial impact on risk, whereas risk alleles with low minor allele frequency or rare variants might be relevant in the aetiology of ASD.

## **2.5 CNVs and ASD**

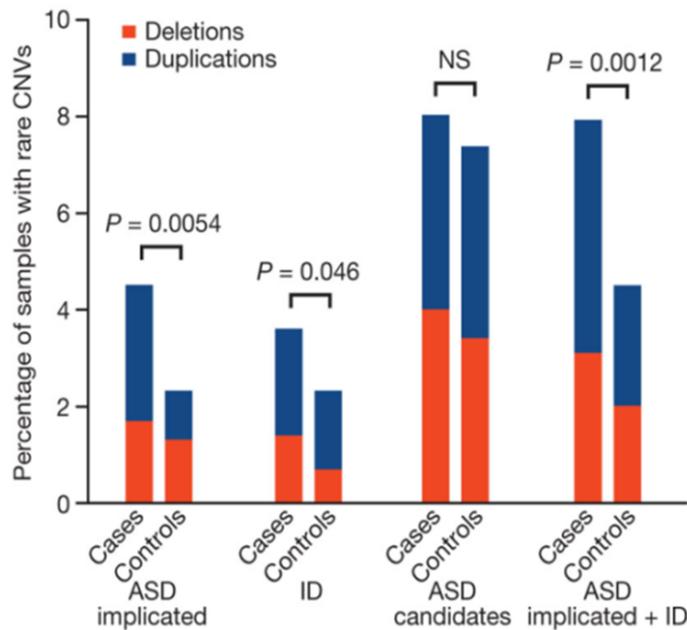
Screening for CNVs, performed using different microarray platforms (see Chapter 1), has proven to be a rapid method to analyse the genomes of large cohorts of ASD cases and controls and to quantify the role of rare CNVs in ASD susceptibility.

Early genome-wide studies performed on ASD families, revealed that CNVs are present in about 6-10% of individuals with ASD, compared with 1-3% in the general population (Christian et al., 2008; Jacquemont et al., 2006; Marshall et al., 2008; Sebat et al., 2007). Notably, these studies found that the proportion of *de novo* CNVs is three-fold to five-fold higher in patients with ASD than in unaffected individuals. Moreover, *de novo* CNVs are larger and affect more genes in individuals with ASD compared with their unaffected siblings and controls. However, since unaffected siblings display a frequency of *de novo* variants similar to controls, this enrichment does not reflect a higher risk of parents to produce these *de novo* variants.

Another interesting finding is that simplex and multiplex families differ in the proportion of *de novo* CNVs.

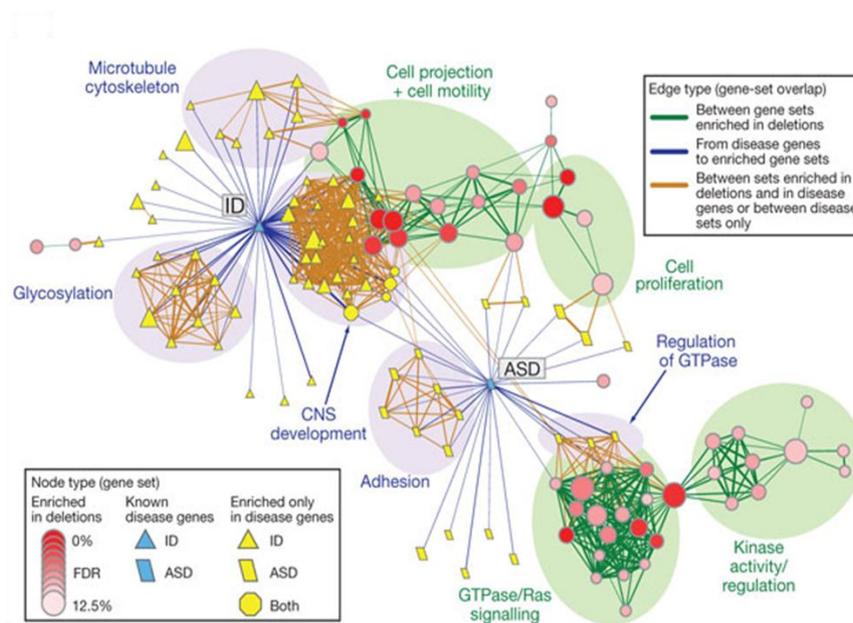
Other studies (Marshall et al., 2008; Pinto et al., 2010) confirmed a greater abundance of large *de novo* CNVs in simplex versus multiplex ASD families and highlighted some of the CNVs recognized as risk loci and their frequency of occurrence (all individually less than 1%) in ASD cases. Further studies, however, failed to replicate the genome-wide differences initially found in CNV frequency between ASD patients and controls (Bucan et al., 2009; Glessner et al., 2009; Pinto et al., 2010).

When comparing 996 ASD individuals of European ancestry to 1,287 matched controls, genotyped on the Illumina Infinium 1M-single SNP-microarray, Pinto et al. (2010) did not find a significant difference between cases and controls. Instead, when focusing on gene-containing segments, a significant increase (1.19-fold increase, empirical  $P=0.012$ ) in the number of genes intersected by rare CNVs was discovered in cases; interestingly this difference was more statistically significant (1.69 fold,  $P= 3.4\times 10^{-4}$ ) restricting the analysis to loci previously implicated in either ASD and/or intellectual disability (**Figure 2.3**).



**Figure 2.3** (Pinto et al., 2010). **CNV burden in known ASD and/or ID genes.** Proportion of samples with rare CNVs overlapping genes known to be associated in ASD with or without ID or ID only. A higher proportion of cases with rare CNVs overlapping “ASD implicated” genes was observed compared to controls (4.3% versus 2.3%, Fisher exact test  $P= 5.4 \times 10^{-3}$ ). Combining the “ASD implicated” and “ID” gene-sets result that 7.6% of cases has rare CNVs preferentially affecting ASD/ID genes compared to 4.5% in controls (Fisher exact test  $P= 1.2 \times 10^{-3}$ , Fig. 2a).

A total of 226 *de novo* and inherited CNVs not observed in controls and affecting single genes were found, and among these CNVs, some of them implicated many novel ASD genes, such as *SHANK2*, *SYNGAP1* and *DLGAP2*. Moreover in order to identify groups of genes affected by CNVs that share a common function or operate in the same pathway (gene-sets), a functional enrichment mapping approach was used. Both gene-sets known to be involved in ASDs and new candidate ASD pathways were discovered: for example, gene-sets involved in cell and neuronal development and function (including projection, motility, and proliferation) previously reported in ASD-associated phenotypes, and new groups such as GTPase/Ras signaling, known to be involved in regulating dendrite and spine plasticity and associated with ID (**figure 2.4**).



**Figure 2.4** (Pinto et al., 2010). **A functional map of ASD.** The gene-sets of known ASD/ID genes are represented by nodes and node size is proportional to the total number of genes in each set. The color indicates the class of gene-set: enriched in deletions, red; known ASD and/or ID genes, blue; enriched only in disease genes, yellow. Edge color represents the overlap between gene-sets enriched in deletions (green), from disease genes to enriched sets (blue), and between sets enriched in deletions and in disease genes or between disease gene-sets only (orange). The major functional groups are highlighted by filled circles (enriched in deletions, green; enriched in ASD/ID, blue).

These findings suggest that CNV location and its functional relevance may play a more important role instead of mean CNV number and size. Moreover these CNV studies provide strong support for the involvement of multiple rare genic CNVs, both genome-wide and at specific loci, in ASD and highlight the role of molecules such as *NRXN1*, *NLGN3/4X* and *SHANK3*, in maturation and function of glutamatergic synapses. Other network-based functional analyses of rare CNVs have confirmed the involvement of these loci in axon targeting, neuron motility and synapse development (Gilman et al., 2011).

Two independent recent studies (D. Levy et al., 2011; Sanders et al., 2011), performed on the same large ASD cohorts of simplex cases, have supported the role of *de-novo* CNVs in idiopathic autism, reporting a significant increase in the burden of rare *de novo* CNVs in ASD cases (5.8%-7.9%) compared to unaffected siblings (1.7%-1.9%). Moreover, they confirmed several known ASD loci, but identified also many new candidate regions, such as the 16p13.2 and *CDH13* loci.

Therefore, the occurrence of multiple different *de novo* CNVs identified in the probands, but not in their unaffected siblings, indicates that autism is mostly caused by rare mutations with most *de novo* events being unique to each proband.

More specifically, Levi et al found evidence for a major contribution of inherited “ultrarare” duplications in cases than in unaffected siblings and showed that relative to males, females have

greater resistance to autism, as they have a higher frequency of *de novo* events than males (11.7% versus 7.4%,  $p$  value = 0.16) and these genomic imbalances are larger and affect more genes.

In summary, the studies conducted to date on ASD revealed that a large number of individually rare (with a frequency less than 1%) CNVs and encompassing several different genes, are involved in the etiology of ASD, further showing the extent of genetic heterogeneity in this disorder.

Some CNVs have a large impact on ASD expression, cause more severe ASD symptoms, and be more prevalent among sporadic ASD cases, whereas other CNVs have moderate or mild effects and constitute risk factors, which likely require other genetic (other CNVs, mutations) or non-genetic factors to predispose to ASD or to other disorders. Some of these CNVs show a variable phenotypic expression and are observed in non-ASD relatives and in some controls.

Indeed up to 40% of family-specific CNVs are inherited from an apparently non ASD parent, who may display some autistic traits, but without satisfying criteria for autistic disorder and this is consistent with incomplete penetrance.

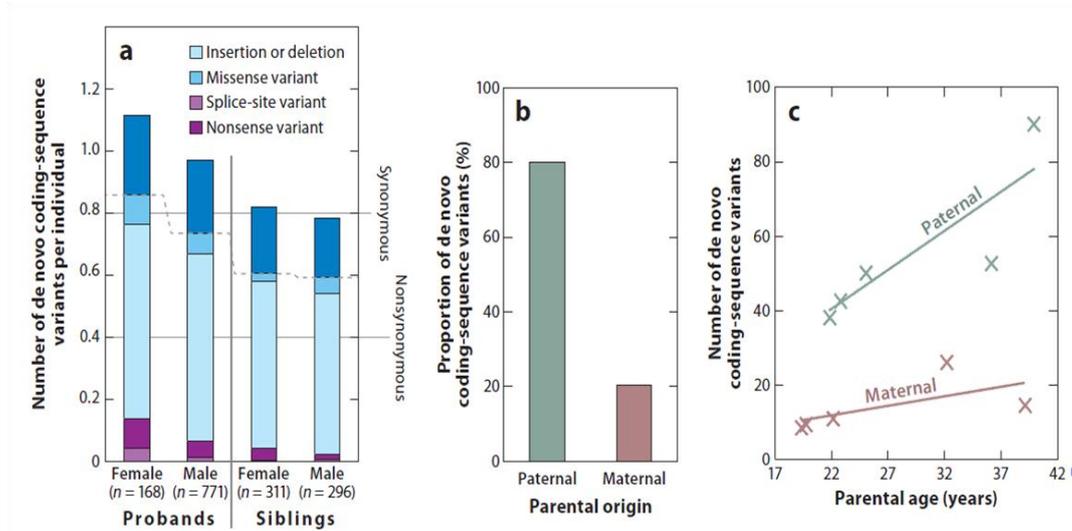
### **2.6 Next Generation Sequencing Technologies (NGS) and ASD**

The recent introduction of *Next Generation Sequencing* (NGS) technologies such as whole-exome and whole-genome sequencing, had an enormous impact on gene discovery, accelerating the characterization of genomic variation and allowing the mapping of both CNVs and single nucleotide variants.

In particular, whole exome sequencing has proven to be a powerful tool for discovering risk-conferring variation, such as *de-novo* mutations, which are sufficiently rare and may thus provide evidence for a causal link to ASD.

The first large-scale WES (Kong et al., 2012; Neale et al., 2012; O'Roak et al., 2012; Sanders et al., 2012) have been carried out assessing more than 1000 ASD subjects, in order to estimate the contribution of *de novo* coding sequence mutations, especially *de novo* loss-of-function (LoF) variations (mutations that are predicted to disrupt gene function), which are more deleterious than inherited variants. These studies showed that the average rate of *de-novo* LoF (including missense, splicing, frameshift and stop-gain variants) was higher in ASD cases than unaffected siblings, and this difference was statistically significant when the analysis was restricted to genes expressed in brain. Specifically, females with ASD tend to have more *de-novo* coding sequence variants than affected males (**figure 2.5a**). Moreover the proportion of these protein-altering point mutations was three time more likely to come from paternal chromosome than the maternal one (**figure 2.5b**) and

the rate of these variants increased with paternal age (**figure 2.5c**) (Kong et al., 2012), suggesting that children of older father have a higher autism risk.



**Figure 2.5** (Huguet, Ey and Bourgeron 2013). The burden of de novo coding-sequence variants in ASD. (a) Distribution of de novo coding-sequence variants in individuals with ASD and controls after stratification by sex. Data are from Neale et al.2012, O’Roak et al. 2012, Sanders et al. 2012; Iossinov et al, 2012. (b) Parental origin of de novo coding-sequence variants identified in individuals with ASD. Data are from References Kong et al., 2012 and O’Roak et al. 2012. (c) Relationship between parental age and the number of de novo coding-sequence variants in the child. Data are from Reference Kong et al., 2012.

However, only a small number of these variants are causative, and those that confer risk to ASD are distributed in many autism-related genes and show an incomplete penetrance. Moreover, many of the disrupted genes are involved in important gene networks, including synaptic plasticity or catenin/chromatin remodelling, and several *de novo* variants impact genes previously implicated in other neurodevelopmental disorders and in intellectual disability. Therefore, these findings further support the involvement of synaptic pathways in ASD susceptibility.

Another interesting result originating from WES is that ASD risk was increased when two extremely rare ( $\leq 5\%$ ) damaging variants, such as nonsense and essential splice site, affect both copies of a protein coding gene, suggesting a role of recessive non-synonymous changes in ASD susceptibility. Similarly, rare hemizygous mutations on the X chromosome, which truncate the protein in males, were enriched in ASD males compared to controls (Lim et al., 2013). These results provide thus evidence that homozygous or compound heterozygous LoF variants play an important role in ASD pathogenesis.

A further evidence supporting a role of biallelic mutations comes from WES applied to consanguineous and/or multiplex families with ASD (Yu et al., 2013), which are extremely useful to identify inherited mutations responsible for rare heritable conditions. These approach allowed the

identification of many hypomorphic mutations that partially impair gene function and could result in atypical milder forms of some diseases, associated with a ASD phenotype. Interestingly these rare variants have been detected in genes not directly implicated in synaptic activity, but involved in neurometabolic disorders, such as *AMT* and *PEX7*, suggesting that unexpected pathways may be involved in ASD aetiology. Alternatively, these new candidate genes may be implicated in synaptic pathways that have yet to be characterized: for example *AMT* encodes for an enzyme essential for glycine degradation and it is known that glycine is a crucial inhibitory neurotransmitter (Baer, Waldvogel, Faull, & Rees, 2009), whereas *PEX7* encodes for a receptor required for import of proteins into the peroxisomes, which are abundant in dendrites (J. Kou et al., 2011)

Therefore, since ASD is an etiologically and phenotypically heterogeneous disorder that involved hundreds of genomic loci, identifying highly penetrant variants in very large sample size is likely the better approach to understand the biological pathways linked to ASD.

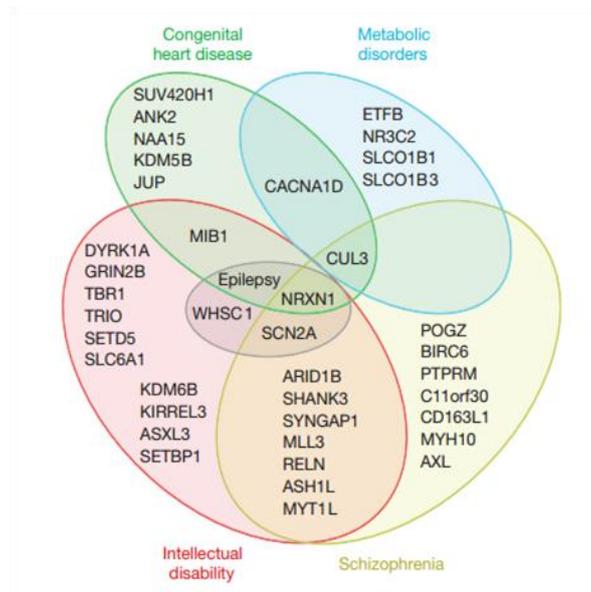
To this end two recent WES have been performed (De Rubeis et al., 2014).

By sequencing more than 2500 ASD simplex families Iossifov et al (Iossifov et al., 2014) estimated that 13% of *de novo* missense mutations and 43% of *de novo* likely gene-disrupting (LGD) mutations (including non sense, frameshift and splice site) contribute to 12% and 9% of ASD diagnoses, respectively. Moreover *de novo* variants occur three time as often in the paternal background and mutation rates are increased with age of either parent, confirming thus previous findings (Kong et al., 2012; O'Roak et al., 2012).

When comparing probands and unaffected siblings, probands were found to carry a significant enrichment for *de novo* LGD in certain functional categories, including FMRP targets, chromatin modifiers and genes expressed in embryonic development. Specifically, affected individuals with higher IQs had a greater incidence of LGD mutations than unaffected siblings, but a lower incidence than affected females or males with lower IQs. Moreover, in comparison with males with higher IQ, females and males with lower IQ showed an enrichment of these variants in the FMRP-associated genes, chromatin modifiers, embryonically expressed genes and in published targets for intellectual disability and schizophrenia.

Additional gene-sets implicated in ASD susceptibility have been discovered by the largest ASD WES study conducted so far (De Rubeis et al., 2014) in more than 3800 autistic cases and in about 10.000 ancestry-matched or parental controls. By analysing the distribution of relative risks across genes for three particular classes of sequence changes, including *de novo* LoF, *de novo* probably damaging missense (Mis3) variants and transmitted LoF, 22 autosomal genes at a false discovery rate (FDR) < 0.05, plus a set of 107 autosomal genes strongly enriched for those likely to affect risk

(FDR < 0.30), were discovered. These 107 genes include 21 candidate genes for intellectual disability, 3 for epilepsy, 17 for schizophrenia, 9 for congenital heart disease and 6 for metabolic disorders (**figure 2.6**).



**Figure 2.6** (De Rubeis et al., 2014). Venn diagram showing the overlap in disease involvement for the ASD genes.

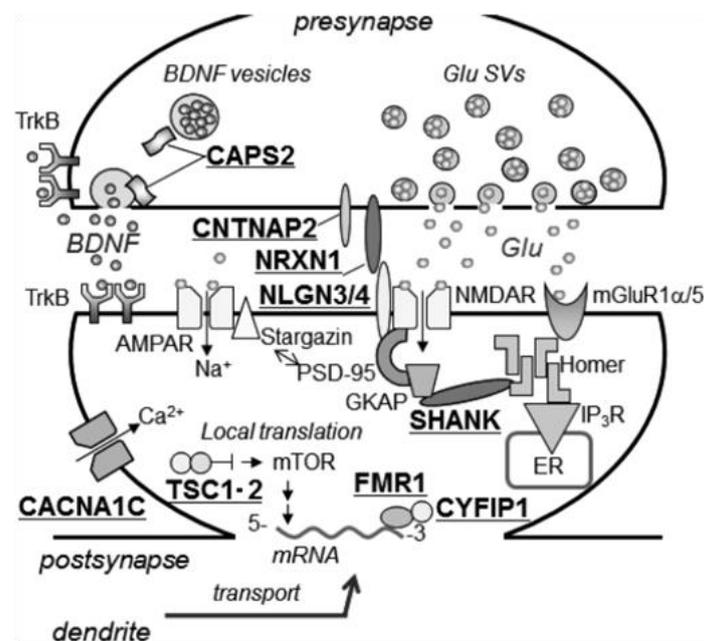
As previously observed, male and female ASD subjects showed a difference in the effect on risk for a class of variants: females have greatly reduced rates of ASD relative to males (a ‘female protective effect’) and a higher liability threshold, requiring a larger genetic burden before being diagnosed. Moreover, three critical pathways were found to be damaged by risk variation: chromatin remodelling, which involves histone-modifying enzymes and chromatin remodellers that recognize specific histone post-translational modifications and orchestrate their effects on chromatin; transcription and splicing, which includes the RBFOX splicing factors; synaptic function, which involves multiple classes and components of synaptic networks, from receptors and ion channels to scaffolding proteins. These findings suggest that alterations of the chromatin dynamics and transcription may affect the synaptic function, that is essential for neural physiology.

## CHAPTER 3: ASD CANDIDATE GENES

### 3.1 ASD candidate genes

Many ASD loci and genes have been identified and are just beginning to be connected in functional networks. Moreover, several databases, such as Autism KB (<http://autismkb.cbi.pku.edu.cn>) and SFARI Gene (<https://gene.sfari.org>), are providing functional annotation of genes associated with ASD.

The genes implicated in ASD susceptibility are numerous and involved in multiple cellular functions. Most of them encode proteins involved in the development and function of neural circuits (**figure 3.1**), including neuronal differentiation, migration and circuit formation (*CDH10*, *CDH9*, *SEMA5A*, *RELN*, *PTEN*), regulation of synaptic adhesion (*NRXN1*, *CNTNAP2*, *NLGN3*, and *NLGN4X*), synaptic transmission (*CADPS2*, *CACNA1C*, *GABRB3* and *SHANK3*), and transcription and translation (*TSC1*, *TSC2*, *FMR1*, and *MECP2*) (Bill & Geschwind, 2009; O'Roak et al., 2012; Pinto et al., 2010).



**Figure 3.1** (Sadakata, Shinoda, Sato, et al., 2013). Many ASD-associated genes and gene candidates are involved in synapse function and structure. ASD-associated genes and gene candidates (indicated by underlining) are involved in synaptic connections, synaptic transmission, synaptic plasticity, activity-dependent gene expression, and local translation in synapses.

### **3.2 ASD-related syndromes**

The first genetic causes of ASD were identified in monogenic disorders which occur in approximately 10% of all ASD cases and include known genetic syndromes, such as tuberous sclerosis (associated with *TSC1* and *TSC2*), neurofibromatosis (associated with *NFI*), Rett syndrome (associated with *MECP2*), and Cowden syndrome (associated with *PTEN*).

Fragile X syndrome (FXS) is the second most common cause of mental impairment after trisomy 21 (Rousseau, Rouillard, Morel, Khandjian, & Morgan, 1995), accounting for 2% of ASD cases. FXS is caused by the unstable expansion of a CGG repeat (>200 repeats) at 5' of the *FMRI* gene, located in Xq27.3, producing abnormal methylation, *FMRI* transcription silencing and decreased FMRP protein levels in the brain (Bagni & Greenough, 2005). Fragile X syndrome in the ASD cases is caused by premutation repeat expansions (55-200 CGG repeats) which can result in RNA toxicity and can also lead to aging effects including tremor, ataxia and cognitive decline, called "Fragile X-associated tremor ataxia syndrome". ASD in FXS is mainly characterized by deficits in social interaction and communication, although impairments in theory-of-mind and pragmatic language distinguish FXS with and without ASD (Kaufmann et al., 2004; Losh, Martin, Klusek, Hogan-Brown, & Sideris, 2012; McDuffie et al., 2010). Recent studies have demonstrated that individuals with ASD and FXS have a lower IQ and greater deficits in receptive and expressive language compared to patients with FXS alone (McDuffie et al., 2010).

Tuberous sclerosis (TS) is an autosomal dominant disease with high penetrance, caused by inactivating mutations, including nonsense, missense, insertion and deletion mutations, in either of two genes, *TSC1* or *TSC2*, located in 9q34 and 16p13.3, respectively. It is characterized by tumor-like lesions in multiple organs but also by other clinical manifestations such as epilepsy, learning difficulties and behavioral problems. In particular, subject with *TSC2* mutations are significantly more likely to display greater severity compared to those with *TSC1* mutations, including autistic disorder, infantile spasms and a lower intelligent quotient (Crino, Nathanson, & Henske, 2006; Gutierrez, Smalley, & Tanguay, 1998).

While *FMRI* mutations may also contribute to the etiology of non-syndromic ASD, particularly in women (Chaste et al., 2012), an enrichment of rare functional variants in the *TSC1* and *TSC2* genes was not reported in a sample of 300 ASD trios, thus excluding that mutations in the *TSC1/2* genes are rare causes of non-syndromic autism (Bahl et al., 2013).

Rett syndrome is caused by *de novo* loss-of-function mutations in the *MECP2* gene, which encodes a member of the methyl-CpG-binding domain family of proteins. *MeCP2* binds to methylated CpG dinucleotides and recruits histone deacetylase 1 (*HDAC1*) and other proteins involved in chromatin

repression at specific gene promoters. Loss of MeCP2 results in a delay of neuronal maturation and synaptogenesis, suggesting that this gene play an important role in correct brain function and development. Rett syndrome occurs in approximately 70% of affected females, while it is generally lethal in males (Amir et al., 1999; Chahrouh & Zoghbi, 2007).

Depending on specific mutation affecting *MECP2*, genetic background of the affected individual and most specifically on X-inactivation pattern which tends to be highly altered in presence of mutations affecting X-linked genes, the *de novo* mutations in *MECP2* can result in mild mental retardation or relatively asymptomatic phenotypes. Moreover *MECP2* variants, most of them are *de novo*, but some are inherited from mothers with borderline cognitive functioning, have also been identified in non-syndromic autistic girls (Young et al., 2008).

Mutations in the *PTEN* gene (Phosphatase and Tensin homolog), encoding a tumor suppressor involved in cell-cycle arrest in G1 and apoptosis, are associated with a broad spectrum of disorders, including Cowden syndrome. Moreover genetic syndromes linked to *PTEN* haploinsufficiency are often associated with autism or mental retardation (Goffin, Hoefsloot, Bosgoed, Swillen, & Fryns, 2001). In particular autistic individuals with mutations in *PTEN* are characterized by severe to extreme macrocephaly and the most of them have a high incidence of *de novo* mutations and an increased risk of developing cancers during adulthood (Lintas & Persico, 2009).

These ASD-related syndromes are characterized by an aberrant mRNA translation which lead to increased synthesis of synaptic proteins, since the proteins encoded by *FMRI*, *TSC1/2*, *NF1* and *PTEN* normally inhibit translation directly at synapses (FMRP) or through the PI3K-mTOR signaling pathway (*TSC1*, *TSC2*, *NF1*, and *PTEN*). Thus, these findings indicate that abnormally increased levels of plasticity-related proteins may affect synaptic connectivity, producing cognitive impairment (Kelleher & Bear, 2008). Moreover, it is evident that also these syndromic forms of autism show clinical phenotypes highly heterogeneous even in the presence of the same well characterized mutation, supporting the hypothesis that differences in genetic background and epigenetic influences underlie ASD.

### **3.3 “Synaptic” genes: neuroligins, neurexins and SHANK family**

Other monogenic causes of ASD, each accounting for less than 1% of the general ASD population, are due to rare highly penetrant mutations in some synaptic genes, such as those encoding neuroligins, neurexins, and SHANKs, which are crucial proteins for synapse formation, maturation and stabilization.

Neurexins (NRXN) and neuroligins (NLGN) are cell adhesion molecules present in excitatory and inhibitory synapses, and they are required for correct neuron network function. At the extracellular level, postsynaptic neuroligins interact with presynaptic  $\alpha$  or  $\beta$  neurexins modulating the synaptic development and function (Fabrichny et al., 2007; Scheiffele, Fan, Choih, Fetter, & Serafini, 2000), whereas at the intracellular level, neuroligins associate with postsynaptic scaffolding proteins, such as SHANK3 (Gerrow et al., 2006). In addition to their role in interaction with neuroligins, neurexins are important mediators for neurotransmitter release by linking calcium ( $\text{Ca}^{2+}$ ) channels to synaptic vesicle exocytosis (Missler et al., 2003).

The involvement of neuroligins in ASD was first reported by Jamain and colleagues, who identified deleterious mutations in the *NLGN3* (Xq13) and *NLGN4* (Xp22.3) genes in siblings with autism or Asperger Syndrome (Jamain et al., 2003). Moreover a 2-bp deletion in fifth exon of *NLGN4* was found in a large family including members affected by nonspecific X-linked mental retardation, with or without autism or pervasive developmental disorder in affected male patients (Laumonnier et al., 2004). This deletion leads to a premature stop codon in the middle of the sequence and results in a protein that is predicted to be truncated by ~50% of its normal sequence and to lose sequences important for the dimerization of neuroligins that are required for cell-cell interaction through binding to  $\beta$ -neurexins.

These results suggest that *NLGN4* is not only involved in autism, as previously described, but also in mental retardation, indicating that some types of autistic disorder and mental retardation may have common genetic origins and highlighting the phenotypic heterogeneity of this X-linked condition. However these and others studies have demonstrated that the frequency of NLGN gene mutations among idiopathic ASD patients is low (Lintas & Persico, 2009).

Several studies have reported rare sequence variants or CNVs affecting the *NRXN1* gene. Similarly to neuroligins, individuals with mutations in *NRXN1* show vastly different clinical phenotypes, ranging from ASD, ID (Zahir et al., 2008) and schizophrenia (Kirov et al., 2009; Rujescu et al., 2009) to specific language disorder and muscle hypotonia (Ching et al., 2010).

Rare mutations have also been identified in the three members of the *SHANK* gene family, *SHANK1* (Sato et al., 2012), *SHANK2* (Leblond et al., 2012) and *SHANK3* (Durand et al., 2007).

The *SHANK3* gene maps in the critical region of the 22q13 deletion syndrome, also known as Phelan-McDermid Syndrome, which is characterized by hypotonia, ID, severely speech delay or absence of speech, mild dysmorphic and autistic traits (Phelan & McDermid, 2012).

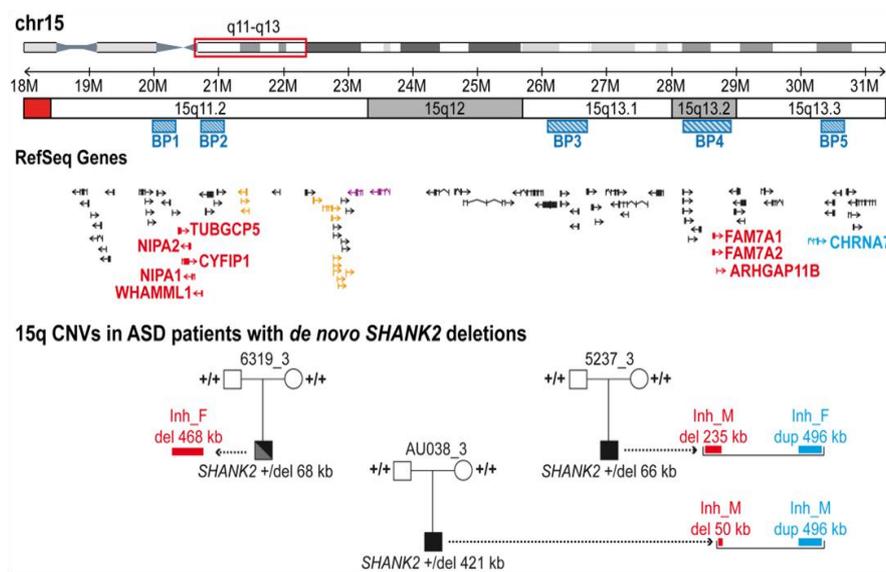
The phenotypic features observed in the Phelan–McDermid Syndrome are likely due to different genes involved in the deletion, but *SHANK3* haploinsufficiency seems to be the most cause of the

autistic traits reported in the ASD patients. Indeed several studies have reported *de novo* mutations (such as microdeletions or loss-of function sequence variants) encompassing only the *SHANK3* locus in ASD individuals. These ASD subjects are characterized by severe language impairment, often accompanied by ID and neurodevelopmental delay (Lintas & Persico, 2009). Moreover, two *de novo* mutations, never found in ASD patients, have been found in individuals with schizophrenia (Gauthier et al., 2010). Notably, autistic carriers of inherited 22q13 deletions involving *SHANK3* have siblings with partial 22q13 trisomy diagnosed with Asperger's syndrome (Durand et al., 2007; Moessner et al., 2007).

Deletions and rare sequence variants, both inherited and *de novo*, have also recently been reported in the *SHANK2* gene in subjects with ASD and ID (Berkel et al., 2010). However, while *de novo* mutations have a clear pathogenic effect, the role of inherited variants is difficult to ascertain, since they are frequently inherited from an apparently healthy parent and can also be found in unaffected siblings of probands with autism and also in controls.

However, a significant enrichment ( $P=0.004$ ) of variants affecting conserved amino acids of *SHANK2* were observed in subject with ASD (3.4%) compared with controls (1.5%) and, notably, several variants identified in patients were associated with a reduced synaptic density at dendrites compared to the variants only detected in controls ( $P = 0.0013$ ) (Leblond et al., 2012).

Interestingly, three patients with *de novo* *SHANK2* deletions also carried inherited 15q11-q13 CNVs previously associated with neuropsychiatric disorders. Specifically, two patients carried a duplication involving the nicotinic receptor *CHRNA7*, while the other ASD individual carried a deletion encompassing the synaptic translation repressor *CYFIP1* (figure 3.1).



**Figure 3.1** (Leblond et al., 2012). Inherited 15q11-q13 CNVs identified in three ASD patients carrier of a *de novo* *SHANK2* deletion. Genes altered by the CNVs are indicated in blue (duplications) or red (deletions). BP, breakpoint; Inh\_M, inherited by mother; Inh\_F, inherited by father.

These results strongly support an oligogenic, “multiple hit” model and suggest that alterations in the *SHANK2* gene might contribute to ASD risk interacting with other variants involving genes which play a role in the same pathway.

Another study (Sato et al., 2012) has indeed indicated that *SHANK1* deletions are associated with ASD with higher functioning in males. Interestingly, a stop mutation of the *PCDHGA11* gene, a member of the protocadherin gamma gene cluster, which have an important role in establishing connections in the brain, was found to segregate precisely with *SHANK1* deletion, suggesting that, similarly to individuals with *SHANK2* mutations, additional hits increase the risk of ASD in individuals with *SHANK1* mutations.

### **3.4 Candidate genes object of this thesis**

#### **3.4.1 The $\alpha$ -catenins**

Other synaptic cell-adhesion proteins implicated in ASD susceptibility are the cadherins (CDH), the protocadherin (PCDH), the contactins (CNTN1-CNTN6) and the contactin-associated proteins (CNTNAP1–CNTNAP5) (Betancur, Sakurai, & Buxbaum, 2009).

Specifically, the cadherins regulate the synapse morphogenesis and plasticity within the mammalian central nervous system through the binding to their cytosolic partners, the catenins.

The catenins are subdivided into three separate groups: two  $\beta$ -catenin-like proteins ( $\beta$ -catenin and plakoglobin), three  $\alpha$ -catenins and four p120 catenin-related proteins (Arikkath & Reichardt, 2008).

The  $\alpha$ -catenins anchor the cadherin/catenin complex to the actin cytoskeleton by binding to  $\beta$ -catenins or cadherins on one side, and form a direct or indirect (through  $\alpha$ -actinin) interaction with actin filaments on the other side. The cadherin-catenin complex has been proposed to be crucial in mediating adhesion between presynaptic and postsynaptic membranes.

There are three  $\alpha$ -catenin genes, which differ in pattern of expression: *CTNNA1* (alpha E-catenin) is ubiquitously expressed but mainly in epithelial tissues (Nagafuchi, Takeichi, & Tsukita, 1991), *CTNNA2* (alpha N-catenin) shows a neural specific expression pattern (Hirano, Kimoto, Shimoyama, Hirohashi, & Takeichi, 1992) and *CTNNA3* (alpha T-catenin) is expressed primarily in the heart and testis but at lower levels in the brain (Janssens et al., 2001).

However, the sequence conservation between the three  $\alpha$ -catenins is particularly elevated in some functional domains, such as binding sites for  $\beta$ -catenin,  $\alpha$ -actinin and actin. In addition to playing a structural role in cell-cell adhesion,  $\alpha$ -catenins are also involved in signal transduction, including

the Wnt signaling, previously implicated in Alzheimer disease (Sehgal, Gumbiner, & Reichardt, 1997).

$\alpha$ E-catenin is a well-known invasion suppressor, as defects in its expression have been observed in several invasive cell lines. *In vitro* experiments have reported that  $\alpha$ E-catenin function can be substituted by  $\alpha$ N-catenin and  $\alpha$ T-catenin. However, since  $\alpha$ N-catenin and  $\alpha$ T-catenin show a restricted expression pattern, they likely have other specific functions.  $\alpha$ N-catenins are localized to adherens junctions bordering active zones in developing and mature synapses throughout the brain and it has been suggested that they contribute to the maintenance of positional information during forebrain development (Park, Falls, Finger, Longo-Guess, & Ackerman, 2002). Moreover  $\alpha$ N-catenin is a key regulator for the stability of synaptic contacts, since it was been shown that in the absence of  $\alpha$ N-catenin, spine heads were abnormally motile, actively protruding filopodia from their synaptic contact sites.

Alternative splicing generate different  $\alpha$ -catenin protein isoforms. A C-terminally truncated isoform of  $\alpha$ E-catenin (Vanpoucke, Nollet, Tejpar, Cassiman, & van Roy, 2002) or an isoform of  $\alpha$ N-catenin with in-frame insertion in the C-terminal region (Uchida et al., 1994) have been identified. In addition, transcripts of *CTNNA2* and *CTNNA3* that contain an alternative 5' exon instead of the conventional initial exon encode isoforms with truncated N-termini (Goossens et al., 2007; Mexal et al., 2008). Moreover a testis-specific N-terminally truncated *CTNNA3* isoform that does not bind  $\beta$ -catenin has been detected in mouse (Goossens et al., 2007).

The role of  $\alpha$ T-catenin has been primarily investigated in the heart, due to its high expression in cardiac tissue and co-localization with plakophilin 2 (Goossens et al., 2007) and in testis but it is expressed also in brain (Busby et al., 2004). Even if its neuronal function remains largely unexplored, several evidence have indicated that *CTNNA3* is a very promising ASD candidate gene. Functional studies performed on mice have shown that overexpression of *Ctnna3*, similarly to *Ctnna1* and *Ctnna2*, causes an increase in spine and synapse density (Abe, Chisaka, Van Roy, & Takeichi, 2004), suggesting that all  $\alpha$ -catenin isoforms share the same spine-stabilizing activity. Moreover, like other  $\alpha$ -catenins, *Ctnna3* participates in the canonical Wnt signalling pathway (Busby et al., 2004), which plays an important role in brain development and synaptic function. Specifically, CNV and association studies investigating several genes involved in this pathway have provided evidence that Wnt signalling might be affected at least in a subset of individuals with ASD (Kalkman, 2012). In addition some studies have reported common single nucleotide polymorphism (SNP) association (Wang et al., 2009; Weiss et al., 2009) and the occurrence of rare CNVs intersecting *CTNNA3* in ASD cases (Girirajan et al., 2012; D. Levy et al., 2011; O'Roak et al., 2012;

Prasad et al., 2012). Another study have reported a *de novo* exonic deletion in this gene associated to ASD and in a recent analysis of exon-disrupting CNVs affecting known autism candidate genes, *CTNNA3* was found borderline enriched in the autism cohort as opposed to controls (22 out of 2,588 autism cases versus 12 out of 2,670 controls,  $P = 0.050$ ) (Girirajan et al., 2012).

Another evidence that support the role of  $\alpha$ T-catenin in ASD susceptibility is its genomic localization. Indeed, *CTNNA3* is located in a common fragile site (FRA10D) and it was suggested that some fragile sites, which are considered hot spots for genomic instability, may be associated to neuropsychiatric diseases, including autism (Savelyeva, Sagulenko, Schmitt, & Schwab, 2006; Smith, Zhu, McAvoy, & Kuhn, 2006).

Moreover, like the others  $\alpha$ -catenins, *CTNNA3* contains the antisense oriented leucine rich repeat transmembrane 3 gene (*LRRTM3*) within its seventh intron (Laurén, Airaksinen, Saarma, & Timmusk, 2003). The *LRRTM* genes encode transmembrane proteins that regulate presynaptic differentiation (Linhoff et al., 2009). The location of the *LRRTM3* genes within  $\alpha$ -catenin family members suggests that  $\alpha$ -catenin genes and the respective LRRTMs share common transcriptional regulation mechanisms.

*LRRTM3* is expressed predominantly in the brain, notably in the hippocampus, and, similar to *CTNNA3*, is thought to mediate cell adhesion (Laurén et al., 2003; Majercak et al., 2006). Interestingly, a significant association of SNPs in *LRRTM3* with ASD was reported, supporting that the *CTNNA3/LRRTM3* genomic region confers susceptibility to ASD.

### 3.4.2 The *CADPS* family

The *CADPS* ( $\text{Ca}^{2+}$ -dependent activator protein for secretion) family is a secretory-related protein family involved in the regulation of secretory granule exocytosis (Berwin, Floor, & Martin, 1998), including monoamines and neuropeptides, as well as in the Golgi trafficking (Sadakata & Furuichi, 2010; Sadakata, Shinoda, Oka, Sekine, & Furuichi, 2013; Shinoda et al., 2011).

In mammals this family includes two members, *CADPS1* and *CADPS2*. These two isoforms are structurally and functionally highly related and show striking difference in their spatial and temporal expression patterns: *CADPS1* is expressed almost exclusively in neuroendocrine cells and in brain and at lower levels before birth, but it increases postnatally, whereas *CADPS2* expression is restricted to certain brain regions, but also in lung, liver and testis and its expression is constant between embryonic and postnatal stage (Speidel et al., 2003).

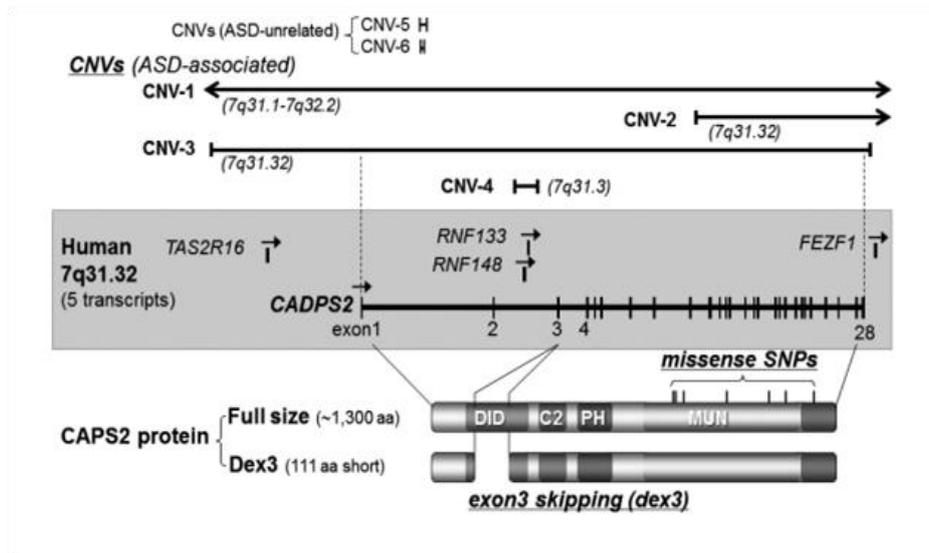
Several studies (Sadakata & Furuichi, 2009; Sadakata et al., 2004; Shinoda et al., 2011) have found that *CADPS2* promotes secretion of brain-derived neurotrophic factor (BDNF), which plays a

crucial role in neuronal survival and differentiation and in synaptic development and plasticity (Greenberg, Xu, Lu, & Hempstead, 2009) as well as in development of hippocampal GABAergic interneuron networks (Hong, McCord, & Greenberg, 2008). *CADPS2*-KO mice show decreased BDNF amounts and significant deficits in hippocampal GABAergic systems at multiple levels, ranging from inhibitory synaptic architectures to related behavioral traits, including decreased social interaction, increased anxiety-like behaviors and defective circadian rhythms (Sadakata, Kakegawa, et al., 2007), suggesting thus that BDNF is involved in psychiatric disorders, such as schizophrenia and depression, and developmental disorders, such as autism.

The human *CADPS2* gene is located within 7q31.32. Five transcripts (*CAPS2/CADPS2*, *TAS2R16*, *RNF148*, *RNF133*, and *FEZF1*) have been annotated in this region, although *CADPS2* is the only gene with an exon-intron structure and characterized to be expressed as a functional protein (**figure 3.2**).

Comparison between *CADPS2* and the mouse homolog, which is located on chromosome 6, showed several differences on exonic structure, suggesting that alternative splicing may occur for this gene. In mouse brain, the major form of *CAPS2* protein (full) consists of approximately 1300 amino acid residues (1275-1355 aa). An aberrant alternative splicing of *CADPS2* mRNA caused by exon 3 skipping (*CAPS2-dex3*), predicting a deletion of 111 aa residues in the human *CADPS2* protein, was found abnormally increased in some patients with autism compared to healthy controls (Sadakata, Washida, et al., 2007). Interestingly, in the *Caps2-dex3* mice models the exon 3-skipped *CADPS2* protein localizes at cell soma and it is not transported to presynaptic terminal, suggesting an increased BDNF release from cell soma and a decreased BDNF release from the axon terminal. The altered local BDNF secretion patterns thereby affect the correct development of synapses and neural circuits. Furthermore, decreased social interaction, increased anxiety, impairment of maternal nurturing behavior and circadian rhythm abnormalities were observed, indicating that increased *CADPS2-dex3* expression leads to increased risk for developing ASD-like behaviors.

At least 13 CNVs around the 7q31.3 region and at least 12 single nucleotide variations, including missense, silence, and insertion mutations in the *CADPS2* gene have so far been reported in ASD patients (Cisternas, Vincent, Scherer, & Ray, 2003). Some of the CNVs have been identified in multiplex ASD families and one CNV (CNV-3 in Figure 2) causes a deletion of 0.75 Mb, including four genes (*CAPS2*, *TAS2R16*, *RNF148*, and *RNF133*) (AGP, 2007). Moreover in a recent study on exon-disrupting CNVs in 253 autism candidate genes a paternally inherited duplication of 0.43 Mb within the *CAPS2* gene was identified (Girirajan et al., 2013). These data suggest that *CADPS2* could be an excellent candidate for neurologic development abnormalities associated with ASD.



**Figure 3.2** (Sadakata, Shinoda, Sato, et al., 2013) In the figure are represented six CNVs and seven non-synonymous single-nucleotide polymorphisms (SNPs) affecting the human *CADPS2* gene. Among the six CNVs four CNVs (CNV 1-4) have been identified in ASD patients, whereas CNV-5 (Bentley et al., 2008) and CNV-6 (S. Levy et al., 2007), which delete very short sequences of *CADPS2* intron 1, had no evidence for a connection with ASD. CNV-2 causes a 0.94 Mb deletion which extend from exon 9 to the end of the *CADPS2* gene (Christian et al., 2008). CNV-1 deletes 15 Mb, affecting about 50 genes including also *CADPS2*. CNV-4 removes *CADPS2* intron 2 and the *RNF148* and *RNF133* gene (Bucan et al., 2009).

### 3.4.3 The *CHRNA7* gene

As discussed in the first chapter, the PB4-PB5 recurrent microdeletion contains six genes (*MTMR15*, *MTMR10*, *TRPM1*, *KLF13*, *OTUD7A*, and *CHRNA7*) and an miRNA gene (*hsa-mir-211*). Strong evidence supporting *CHRNA7* as responsible for the majority of neurodevelopmental phenotypes resulting from PB4-BP5 deletion comes from the identification of individuals carrying smaller deletions which encompass the entire *CHRNA7* gene and the first exon of *OTUD7A* (Shinawi et al., 2009) or even smaller deletions, including only *CHRNA7* (Hoppman-Chaney, Wain, Seger, Superneau, & Hodge, 2013; Masurel-Paulet et al., 2010; Mikhail et al., 2011) and conveying most or all of the phenotypic abnormalities associated with the larger 15q13.3 recurrent deletions.

*CHRNA7* encodes the  $\alpha 7$  subunit of the neuronal nicotinic acetylcholine receptor, which is the only subunit able to form a homopentameric chloride channel receptor, and is highly expressed in the brain. Receptors containing  $\alpha 7$  are localized both pre- and post-synaptically and regulate the release of both the inhibitory neurotransmitter GABA and excitatory neurotransmitter glutamate in the hippocampal formation (Albuquerque, Pereira, Alkondon, & Rogers, 2009). Moreover alpha-7 nicotinic acetylcholine receptor mediated signaling causes an influx of  $\text{Ca}^{2+}$  into the cell (Vijayaraghavan, Pugh, Zhang, Rathouz, & Berg, 1992).

Mobilization of intracellular  $\text{Ca}^{2+}$  plays a critical role in synaptic plasticity and immediate early gene expression associated with learning and memory (Benfenati, 2007). Another evidence

supporting *CHRNA7* involvement in neuropsychiatric disorders comes from a recent study which has reported a significantly reduced *CHRNA7* expression in the frontal cortex of individuals with Rett syndrome or with typical ASD (Yasui et al., 2011). In addition *CHRNA7* has been implicated as a candidate gene for schizophrenia and its expression was reported to be reduced in several brain regions in schizophrenic subjects compared with control subjects. In particular several polymorphisms were identified in the *CHRNA7* core promoter in schizophrenic individuals and functional analysis of these promoter variants indicated that *CHRNA7* transcription was reduced (S. Leonard et al., 2002).

Moreover it is known that the binding of the methyl CpG binding protein 2 (MeCP2), encoded by *MECP2*, influences the chromatin loop organization of the much larger 15q11.2-13.3 region that includes the Prader-Willi/Angelmansyndrome region, and is required for optimal expression of AS/PWS region genes implicated in the ASD phenotype.

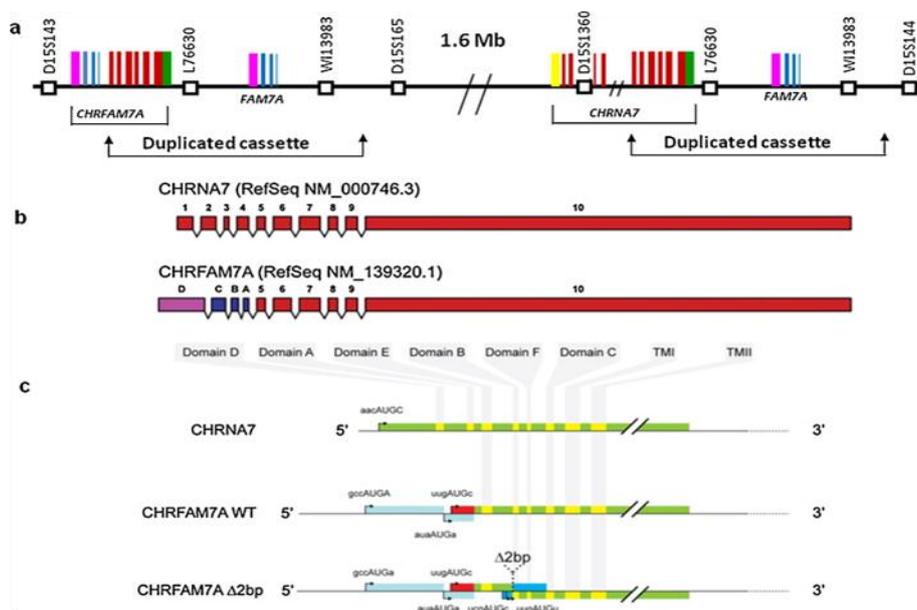
Therefore these discoveries suggest that *CHRNA7* transcription is modulated by these regulatory elements and its reduced expression may be involved in ASD-like phenotypes (Yasui et al., 2011).

#### Function and structural organization of the *CHRNA7* gene

The  $\alpha 7$  receptor subunit gene, *CHRNA7*, is a member of a large gene family of neuronal receptors which are expressed in mammalian brain as pentameric, ligand-gated ion channels (Vijayaraghavan et al., 1992). Two principal classes of these receptors are present in the nervous system: receptors with  $\alpha 2$ - $\alpha 6$  subunits bind nicotine with high affinity and require an association with  $\beta$  subunits for functional expression (Deneris et al., 1989; Goldman et al., 1987), whereas receptors with  $\alpha 7$ - $\alpha 9$  subunits, including *CHRNA7*, bind nicotine with low affinity but they have a high affinity for the antagonist,  $\alpha$ -bungarotoxin, and can function as homomeric ion channels in vitro.

The  $\alpha 7$  receptors are located both pre- and postsynaptically: presynaptically, they are involved in neurotransmitter release, including release of GABA, glutamate, and dopamine from specific terminals (Jones & Wonnacott, 2004; Schilström et al., 2000), while postsynaptically they are localized in or near the postsynaptic density (PSD), where the calcium flux increases phosphorylation and affects gene expression.

*CHRNA7* contains ten exons and it is partially duplicated (**figure 3.3a,b**). Indeed, the exons 5-10, along with a large cassette of DNA (about 300 kb), are duplicated and map approximately 1.6 Mb proximal to the full-length *CHRNA7* gene (Gault et al., 1998).



**Figure 3.3** (Araud et al., 2011). Structure of the *CHRNA7/CHRFAM7A* gene cluster on chromosome 15q13.3. a) Map of the partial duplication of *CHRNA7* on 15q13.3. Exons 5-10 of *CHRNA7* were duplicated in a duplcon of ~300kb, mapping centromeric by 1.6 Mb. The duplcon interrupted a partial duplication of a second gene, *ULK4*. *CHRNA7* exons, red; *ULK4* exons, blue; exon D, pink. b) Schematic representation of the exon organisation of the transcripts coding for *CHRNA7*, *CHRFAM7A*, *CHRFAM7A* based on the RefSeq NM\_000746.3 and NM\_139320.1. c) Putative translation products from *CHRNA7*, *CHRFAM7A*, and *CHRFAM7AΔ2bp* mRNAs. Amino acid sequence of  $\alpha 7$  is represented in green and yellow (for the different domains). Alternative amino acids from *CHRFAM7A* are indicated in red and alternative amino acids from *CHRFAM7AΔ2bp* in blue.

*CHRFAM7A* is a chimeric gene derived by the fusion of *CHRNA7* (exons 5-10) to one of many copies of a novel gene called *FAM7A*, that encodes exons A-E (Gault et al., 1998; Riley, Williamson, Collier, Wilkie, & Makoff, 2002). Exons C-A are the result of a partial duplication of a putative kinase-like gene (*ULK4*) on chromosome 3p22.1, while exon D is of unknown provenance. *CHRNA7* DNA sequence in *CHRFAM7A* is 99.9% conserved. Since *CHRFAM7A* has not been found in closely related primates (Locke et al., 2003) and in rodents, it is probably that the duplication has a recent origin. The *CHRFAM7A* gene has a variable copy number: approximately 30% of individuals have only one copy of *CHRFAM7A*, while rare individuals (about 5%) are missing both copies (Gault et al., 1998; Sinkus et al., 2009). Mutation screening of *CHRFAM7A* has revealed a 2bp deletion in exon 6. This polymorphism has been associated with schizophrenia and with a gene inversion (Flomen et al., 2008). The wild type allele is in a head-to-head orientation with respect to the full-length *CHRNA7* gene, but *CHRFAM7AΔ2bp* is oriented in the same direction (**figure 3.4c**).

A recent study has revealed that expression of *CHRFAM7A* alone generates a protein product but it is not functional, whereas co-expression of  $\alpha 7$  subunits and the *CHRFAM7A* gene results in

decreased ACh stimulated current. However, this co-expression does not reduce the ligand binding and does not alter *CHRNA7* transcription, suggesting thus that *CHRFAM7A* acts as a dominant negative modulator of *CHRNA7* function and is critical for receptor regulation in humans (Araud et al., 2011).

#### **3.4.4 The kelch-like (KLHL) gene family**

The KLHL (Kelch-like) gene family includes proteins that constitute a subgroup at the intersection between the BTB/POZ domain and Kelch domain superfamilies. In the human genome there are 42 KLHL family members, dispersed on different chromosomes and with a number of exons ranging from a single coding exon to 15 coding exons.

However the number of KLHL genes is conserved between mammalian species and all generally contain a BTB/POZ domain, a BACK domain, and five to six Kelch domains. The BTB (Broad Complex, tramtrack and bric à brac)/POZ (poxviruses and zinc finger) domain has been found primarily in zinc finger proteins (Zollman, Godt, Privé, Couderc, & Laski, 1994). It is present at the amino terminus and it is involved in the protein-protein interaction interface (Bardwell & Treisman, 1994) and in both dimer and heterodimer formation in vitro (Albagli, Dhordain, Deweindt, Lecocq, & Leprince, 1995). BTB-containing proteins have different functions such as control of cytoskeletal organization (M. I. Kang, Kobayashi, Wakabayashi, Kim, & Yamamoto, 2004), ion channel gating (Minor et al., 2000), transcription suppression (Melnick et al., 2000) and protein targeting for ubiquitination (Furukawa, He, Borchers, & Xiong, 2003; Xu et al., 2003).

The central BACK domain does not have a known function yet, but it is likely to be of functional significance since mutations in this region have been shown to cause human disease (Bomont et al., 2000; Liang, Avraham, Jiang, & Avraham, 2004). The Kelch repeats, at the C-terminus, form a tertiary structure of  $\beta$ -propellers, which has roles in extracellular communication/interaction, cell morphology, gene expression and actin binding.

KLHL proteins are known to be involved in several cellular processes and are responsible for different human diseases, from cancer to Mendelian diseases.

Mutations in *KLHL7* have been found in patients with autosomal dominant retinitis pigmentosa (Friedman et al., 2009); a missense mutation in the *KLHL9* gene has been associated with development of distal myopathy (Cirak et al., 2010), while mutations in *KLHL16* are linked to human giant axonal neuropathy (Bomont et al., 2000). Other KLHL members are involved in cancer. Examples are *KLHL20*, implicated in prostate cancer progression (Yuan et al., 2011) and *KLHL37* associated with brain tumors (Liang et al., 2004).

### The *KLHL* genes in neuropsychiatric disorders

Despite the specific roles for each *KLHL* family member have not been completely elucidated, as they are involved in different biological processes and disease, some *KLHL* proteins have an established role in neurodevelopment (Jiang, Seng, Avraham, Fu, & Avraham, 2007; S. K. Williams et al., 2005).

An example is *KLHL2*/Mayven, an actin-binding protein constitutively expressed in developing and mature oligodendrocytes and neurons, where it binds directly to F-actin through its Kelch repeats and plays an important role in the organization of the actin cytoskeleton (Jiang et al., 2005; Jiang et al., 2007). Another actin-binding protein is *KLHL1*/MRP2, mainly expressed in specific brain regions, including the cerebellum, the area most affected by spinocerebellar ataxia type 8 (SCA8), an inherited disease of the central nervous system caused by a CTG expansion mutation in the natural antisense RNA of *KLHL1*. MRP2 plays an important role in neurite outgrowth and colocalizes with the cytoskeletal protein talin at the neuronal growth cones and with actin in differentiated primary rat hippocampal neurons (Seng, Avraham, Jiang, Venkatesh, & Avraham, 2006). MRP2 is also involved in the process elongation of oligodendrocytes (OLGs) and its interaction with actin is significantly increased in differentiated OLGs (Jiang et al., 2007).

Additional evidences have supported the involvement of some *KLHL* genes in complex neuropsychiatric disorders.

A partial deletion of the *KLHL15* gene was identified in a patient with severe ID, epilepsy and anomalies of cortical development in a recent array CGH study (Mignon-Ravix et al., 2014). Moreover, proteins containing the kelch domain have been also detected as one of the functional categories enriched among *de novo* genic CNVs in schizophrenia (Malhotra et al., 2011). In addition a deletion involving *KLHL17* has been identified in a subject with early infantile epileptic encephalopathy (Paciorkowski et al., 2011).

Therefore, the occurrence of CNVs in some *KLHL* gene family members in multiple unrelated patients with the same neuropsychiatric disorder or co-morbid conditions might indicate that these genes represent new and interesting candidate for complex disorders.

## CHAPTER 4: INTELLECTUAL DISABILITY

### **4.1 Intellectual disability**

Intellectual disability (ID), also referred to as mental retardation (MR) is a common neurodevelopmental disorder with onset before age 18 years and characterized by social, cognitive, and adaptive skill deficits.

ID diagnosis is based on use of standardized age-dependent tests that measure the intelligence quotient (IQ), such as the Wechsler Intelligence Scale for Children (WISC) or Wechsler Adult Intelligence Scales Intelligence Scale for Adults (WAIS).

Depending on the IQ, several degrees of severity are distinguished: mild, moderate, severe, profound and unable to classify. However ID is commonly divided into 2 principal categories: mild ID (IQ 50-70) and severe ID (IQ<50) (Ropers & Hamel, 2005).

Mild mental retardation is the most common type of intellectual impairment accounting for about 85% of cases (Deb & Prasad, 1994). Despite a causative diagnosis of moderate to severe mental retardation is established in up to 65% of cases, the underlying cause of mild ID mental remains unknown in up to 80% of patients.

In addition to categorization by severity/IQ level, ID can also be divided into syndromic intellectual disability (S-ID), which is characterized by one or multiple clinical features or co-morbidities in addition to ID, and non-syndromic intellectual disability (NS-ID), in which intellectual disability is the sole clinical feature. However the distinction between S-ID and NS-ID is unclear, since the diagnosis is often difficult because of the cognitive impairment of the patients. Moreover recent studies have indicated that, in some cases, different mutations in the same gene can result in S-ID or NS-ID, depending on their effect on the protein product (complete loss or partial loss of the protein), on the X-inactivation pattern in females and on genetic background (Vaillend, Poirier, & Laroche, 2008).

The prevalence of ID is between 1% and 3% (H. Leonard & Wen, 2002) and tends to be higher in developing countries or in areas of lower socioeconomic status, likely due to a variety of non-genetic factors such as malnutrition, cultural deprivation and poor health care (Durkin, 2002), and to etiologic factors, such as parental consanguinity (Ropers, 2008), since inbreeding is associated with reduced cognitive performance (Bashi, 1977).

Mild ID tends to be more prevalent in males. However, the ratio of males to females decreases as IQ decreases (McLaren & Bryson, 1987) and some studies have revealed that severe ID may be more prevalent among females (Katusic et al., 1996).

### **4.2 Causes of ID**

ID can be caused by environmental and/or genetic factors.

Environmental exposure to certain teratogens, viruses or radiation can cause ID, but 25-50% of cases are due to genetic factors. However for up to 60% of cases, there is no an identifiable cause (Rauch et al., 2006).

The genetic causes underlying mental retardation include chromosomal abnormalities, autosomal trisomies compatible with human viability, aneuploidies of the X-chromosome, and monogenic diseases. These genetic factors contribute to about 30% of moderate ID cases, while the environmental factors, including malnutrition during pregnancy, preterm birth, cerebral ischemia and infections pre and postnatal, play an important role in mild ID pathogenesis. In particular, prenatal factors (principally genetic, but also due to threatened abortion and maternal diabetes), are the most frequent and are responsible of about 60-75% of cases, while perinatal causes (low birth weight and intra-uterine growth retardation), and postnatal factors (for example cerebral injuries), are less common (Aicardi, 1998).

Given its genetic complexity, still little is known about the genetic factors contributing to mild ID. Some forms of mild ID segregate as Mendelian traits, while others are multifactorial. Moreover, many single gene causes of NS-ID may also cause not only S-ID, but also autism or other neurodevelopmental phenotypes, suggesting thus that other genetic modifiers or environmental factors may be involved in disease etiology.

Cytogenetically visible chromosome aberrations account for approximately 15% of patients with severe ID (H. Leonard & Wen, 2002) and recently, also small duplications and deletions, have been identified as risk factors for both mild and severe ID. Moreover about 10% of ID found in males are caused by X-linked gene defects (Ropers & Hamel, 2005), much less than previously thought (Herbst & Miller, 1980), indicating that other factors are responsible for prevalence of cognitive impairment in males than females. This has left ample room for autosomal gene defects that appear as both dominant or recessive forms of ID in sporadic cases.

#### **4.2.1 Chromosomal aberrations and ID**

With a prevalence of 1/750 to 1/800, Down syndrome (DS, or trisomy 21) is the most common known cause of mental retardation (Collins, Muggli, Riley, Palma, & Halliday, 2008).

Other numerical and structural aberrations that can be detected by light microscopy are far less common, but taken together, cytogenetically visible chromosomal aberrations are found in one out of seven individuals with severe cognitive impairment (H. Leonard & Wen, 2002).

Moreover, submicroscopic rearrangements, including deletions as well as balanced translocations and other chromosomal aberrations have been implicated in the etiology of NS-ID. These microdeletions include a number of known syndromes, such as Prader-Willi and Angelman syndromes, Smith-Magenis, Miller-Dieker, and DiGeorge syndromes.

After the advent of high-resolution array CGH the identification of submicroscopic subtelomeric rearrangements has become more robust and several pathogenic CNVs, both rare and *de novo*, as well as rare inherited mutations, were detected in 10-15% of individuals with idiopathic ID and normal karyotype (Koolen et al., 2009).

Interestingly, mapping the breakpoints of these CNVs was a successful method in determining autosomal dominant as well as X-linked causes of ID.

### **4.2.2 X-linked intellectual disability (XLID)**

X-linked forms of ID (XLID) are common causes of moderate to severe intellectual disability in males and are easily identifiable because of their characteristic inheritance pattern.

XLID is usually subdivided into syndromic (S-XLID) and non-syndromic forms (NS-XLID) and generally NS-XLID are more common than syndromic ones (Fishburn, Turner, Daniel, & Brookwell, 1983).

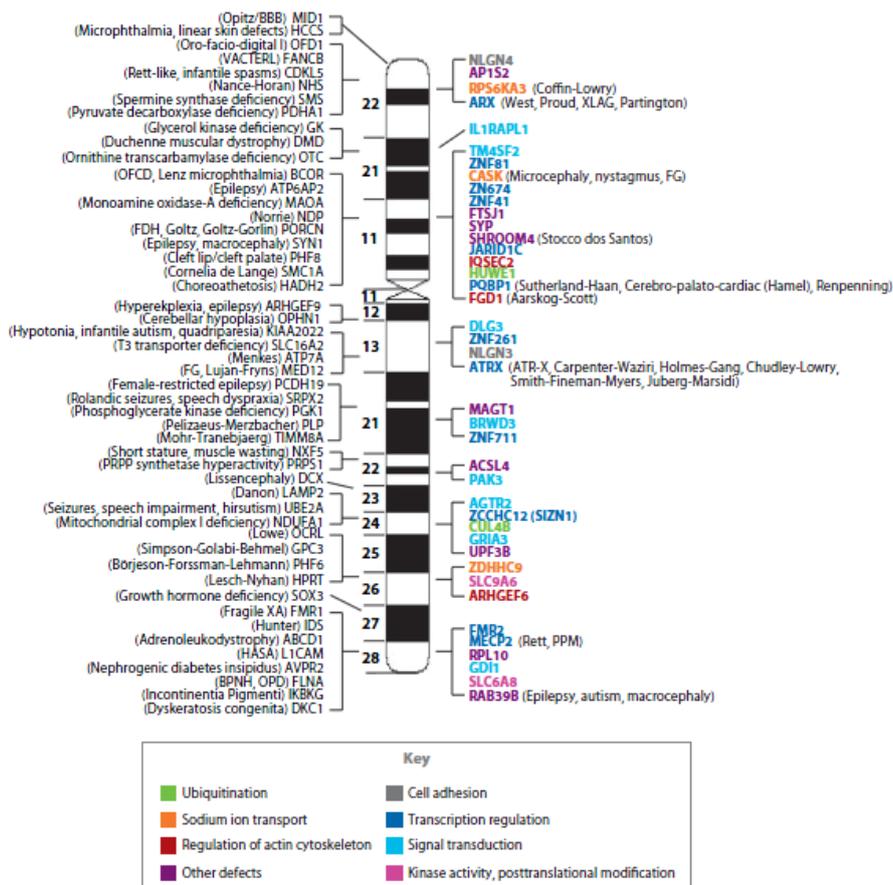
However, some of the genes involved in NS-XLID have also been implicated in syndromic XLID, depending on the mutation, or may even vary within families, possibly modulated by additional factors, indicating thus that there is no clear separation between S-XLID and NS-XLID.

For example, defects of *OPHNI*, one of the first genes implicated in NS-XLID (Billuart et al., 1998), were later found to be associated with a syndrome including ataxia, epilepsy, and cerebellar hypoplasia. Additionally, several genes that classically cause ID syndromes may also cause NS-XLID, such as the *MEPC2* gene, which causes Rett syndrome.

The human X-chromosome carries only about 4% of the protein-coding genes in the human genome, but the X-linked gene defects are responsible for about 10-12% of the ID found in males, which means that there must be other factors to explain why cognitive impairment is far more common in males than females. These factors include differential imprinting of maternal and paternal X-chromosomes, regulatory effects of the Y-chromosome or incomplete X-inactivation in females (Nguyen & Distèche, 2006; Skuse, 2007).

The most common form of XLID, accounting in 25% of ID cases (Fishburn et al., 1983), is the Fragile X syndrome, while ARX mutations rank second giving rise to non-syndromic XLID and to a variety of syndromic forms in more than 5% of the families (Gécz, Cloosterman, & Partington, 2006). Mutations in the *CULAB*, *JARID1C* and *SLC6A8* genes are all relatively frequent, each accounting for 2-3% of the families, whereas defects of all other known XLID genes seem to be significantly less common (about 1%).

The first mutations identified in patients with syndromic and nonsyndromic forms of XLID impact genes which play an important role in neuronal development and synapse formation and function, such as genes regulating the activity of small Rho and Ras GTPases (Kutsche et al., 2000). Other genes implicated in XLID are involved in many other important cellular processes, such as cell adhesion, post-translational modification, signal transduction, transcription regulation, actin cytoskeleton modification and ubiquitination (figure 4.1).



**Figure 4.1** (Ropers, 2010). Genes implicated in syndromic and/or nonsyndromic forms of XLID, and their position on the human X-chromosome. Gene defects that may give rise to non-syndromic XLID are shown on the right, and colors relate to different functional classes (gray, cell adhesion; dark blue, transcription regulation; pink, kinase activity, post-translational modification; light blue, signal transduction; green, ubiquitination; red, regulation of actin cytoskeleton; orange, sodium ion transport; and purple, other defects).

### 4.2.3 Autosomal forms of ID

Autosomal forms of ID (AID) include both autosomal dominant forms (ADID) and autosomal recessive forms (ARID).

Well-known autosomal dominant disorders often associated with ID of varying severity include neurofibromatosis, tuberous sclerosis, and myotonic dystrophy. Other ADID, associated with a more severe phenotype, are mainly caused by *de novo* mutations and they are rarely familial because in general, affected individuals will not reproduce.

The first method used to identify autosomal dominant disease causing genes was the characterization of chromosomal aberrations by breakpoint analysis, which allowed the discovery of several candidate genes for non-syndromic autosomal dominant ID (NS-ADID), such as *DOCK8*, *MBD5*, *CDH15* and *KIRREL3* (Bhalla et al., 2008; Griggs, Ladd, Saul, DuPont, & Srivastava, 2008). Further screening of some of these genes in cohorts of ID patients led to the identification of several sequence mutations.

Others genes that were all identified as autosomal dominant causes of NS-ID using candidate gene sequencing are *SYNGAP1*, *STXBPI*, and *SHANK2* (Hamdan et al., 2009).

In particular, the *SYNGAP1* and *SHANK2* (Berkel et al., 2010; Pinto et al., 2010) genes have also been associated with autism, suggesting thus that ASD and ID show a genetic overlap.

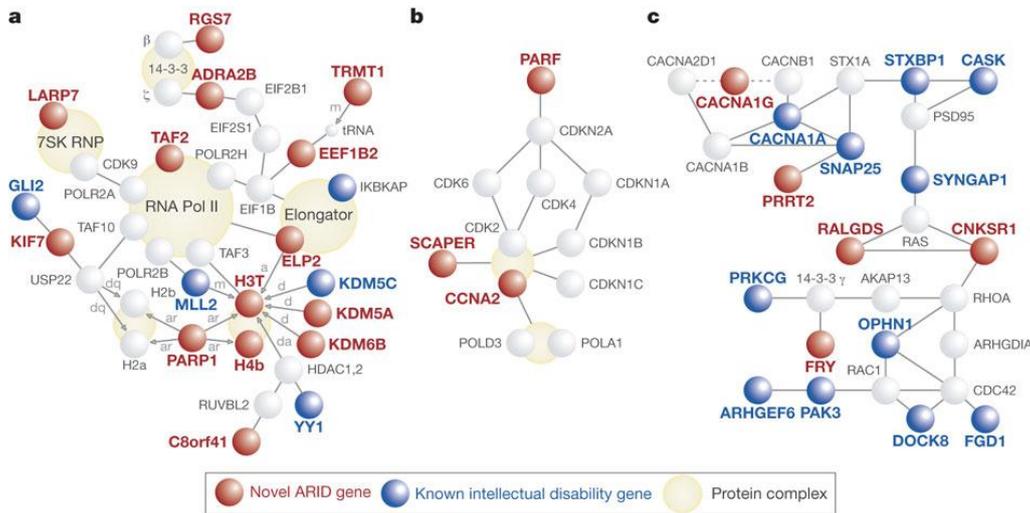
In general, all these studies have provide evidence that *de novo* mutations in autosomal genes are common among sporadic patients with idiopathic ID and that some mutations, mainly truncating mutations, copy number changes or dominant-negative mutations affecting genes involved in the synapses of the central nervous system, have functional consequences not only for X-linked, but also for autosomal dominant forms of ID.

The best strategy to elucidate the molecular defects underlying the autosomal recessive forms of ID is microarray technology combined with homozygosity mapping using large consanguineous families, followed by mutation screening of candidate genes (Houwen et al., 1994). Consanguineous families are common in countries belonging to a consanguinity belt, such as Morocco and India, while in the industrialized countries large families and parental consanguinity are rare.

The first studies ruled out the presence of common gene defects that account for more than a few percent of the patients and revealed that ARID is highly heterogeneous.

The largest study published so far for ARID has been performed by Najmabadi et al., (Najmabadi et al., 2011) using homozygosity mapping, targeted exon enrichment and next-generation sequencing in 136 consanguineous families from Iran and elsewhere.

Instead of sequencing entire exomes in consanguineous families, they have focused on the exons from homozygous linkage intervals known to carry the genetic defect. Additional mutations in 23 genes previously implicated in intellectual disability or related neurological disorders, as well as single, probably disease-causing variants in 50 novel candidate genes, have been identified (figure 4.2).



**Figure 4.2** (Najmabadi et al., 2011). Known and novel intellectual disability genes form protein and regulatory networks. Connecting edges indicate protein–protein interactions. Arrows define direction of post-translational protein modifications: a, acetylation; ar, ADP-ribosylation; d, demethylation; da, deacetylation; dq, deubiquitination; m, methylation. Dotted lines indicate modulation of gene function. **a)** Transcriptional/translational network. **b)** Cell-cycle-related network. **c)** Ras/Rho/PSD95 network.

Relatively few of the novel defects identified involve synapse- or neuron-specific genes and the majority among them are ubiquitously expressed genes with a role in several cellular functions, such as DNA transcription, translation, protein degradation and mRNA splicing. However, some proteins encoded by these genes interact directly with products of known intellectual disability genes and therefore they seem to be pivotal for normal brain development and function.

This study provides interesting implications for the diagnosis and prevention of intellectual disability, and perhaps also for autism, schizophrenia and epilepsy, which often co-exist in intellectual disability patients and are frequently associated with mutations in the same genes.

### 4.3 Co-morbidity of ID with ASD

In addition to social, cognitive, and adaptive skill deficits, intellectual disability is often accompanied by different co-morbidities, ranging from challenging behaviors to anxiety, depression, and schizophrenia (Cherry, Matson, & Paclawskyj, 1997). The disorders with the greatest overlap with ID are those in the autism spectrum.

Several studies have revealed that, besides the heterogeneity of ID and ASDs, they share overlapping risk factors (Betancur, 2011) and both exist together in the majority of patients.

About 70% of patients with ASD have ID (Fombonne, 2003), while the remaining 30%, in addition to cognitive dysfunction, have also speech or behavior disability (Mefford, Batshaw, & Hoffman, 2012). Conversely, at least 10% of individuals with ID have ASDs, with some ID conditions exhibiting a much higher level of co-morbidity.

### **4.3.1 Phenotypic overlap between ID and ASD**

ID and ASD have multiple overlapping phenotypic domains. Indeed, the three major phenotypes that characterize autism (language abnormalities, social deficits and stereotypies) can often be seen to varying degrees in ID individuals. Both ASD and ID may present with developmental delay, abnormal language and social difficulties and are often associated with medical conditions such as epilepsy or autoimmune disorders (Depienne et al., 2007; Schieve et al., 2012).

Deficits in verbal and non-verbal communication as well as repetitive and restrictive activities are more common in persons with autism and ID versus autism alone (Deb & Prasad, 1994). Individuals with severe and profound ID show often severe language deficits and stereotypies that tend to become more evident than subjects with mild ID. Moreover with increasing of IQ the autistic subjects present less improvements in social and daily living skills. In addition, patients with severe ID and autism as a comorbid disorder have much greater deficits in social and adaptive behavior in comparison with subjects whose comorbid disorder with ID was psychosis (Matson, Bielecki, Mayville, & Matson, 2003). In addition, as IQ decreases the rates of challenging behaviors became higher, self-injury in particular. Symptoms of psychopathology, such as mania, anxiety, psychosis, mood disorder and schizophrenia are present in persons with ID and ASD. Specifically these symptoms are evident in severe autism with the fewest symptoms in the ID only group. Compared to persons with ID alone, patients with autism and ID show higher rates of inattention, hyperactivity, and impulse behaviors. Other common comorbid disorders are depression and bipolar disorder (Bradley & Bolton, 2006). What is evident is that this ID and ASD group is distinctly different from persons with ID or with normal IQ and ASD and, at present, the clinicians are attempting to distinguish between these groups, by developing scales that are specific to this aim

### **4.3.2 Genetic overlap between ID and ASD**

Although ID and ASD constitute two distinct disorders, they are characterized by some genetic overlaps which are difficult to quantify genetically because of the heterogeneity of both conditions

and the contribution of rare genetic variants to both diseases. Therefore understanding of the molecular mechanisms underlying ID and ASD remain limited, since most deleterious variants in the ASD and/or ID-associated genes have a very low prevalence and many rare and inherited mutations display often incomplete penetrance. However, several CNVs as well as single gene mutations appear to contribute significantly to the etiology of both ID and autism.

It has recently been reported that 10-20% of ASD cases have an identified genetic anomaly and, notably, these genetic anomalies are also found in individuals with ID. Moreover, many patients affected by an ID syndrome, such as fragile X Syndrome, Down syndrome, tuberous sclerosis (TSC), but also Angelman syndrome and Rett syndrome, have a concomitant autism/PDD diagnosis (Fombonne, 2003).

Several CNVs have been identified in patients with ASD and ID, such as the terminal deletion of the long arm of chromosome 2 (Galasso et al., 2008), the 1.5 Mb duplication on chromosome 16p-13.1 (Ullmann et al., 2007), the deletion at the 7q 11.23 (Depienne et al., 2007) and many others.

A mutation in *NLGN4*, a gene known to be associated to autism (Marshall et al., 2008), has been identified in a family containing subjects with NS-ID, with or without ASD (Laumonnier et al., 2004). Others genes that have been implicated both in autism and NS-ID are *PTCHD1* (Noor et al., 2010), *SHANK3*, *ILIRAPL1* (Piton et al., 2008) and *JARID1C* (Adegbola, Gao, Sommer, & Browning, 2008)

Polymorphisms in *GRIK2*, one of the genes implicated in non-syndromic autosomal recessive intellectual disability (NS-ARID), has been associated with an increased susceptibility to autism and it has been found in linkage disequilibrium in an autistic population (Jamain et al., 2002). Others independent studies performed on different autistic populations have revealed similar results (Kim, Kim, Park, Cho, & Yoo, 2007). However, while in autism mutations in *GRIK2* act in a dominant pattern of inheritance, in NS-ID they show a recessive inheritance, suggesting that polymorphisms in this gene may contribute to the overall ASD susceptibility.

These findings indicate that ID and ASD may be on the same continuum, as opposed to being different clinical entities, sharing common etiologies and showing many genetic similarities. Furthermore these observations confirm the involvement of similar cellular and molecular processes underlying these two disorders. Indeed, using a combined network and systems biology approach to predict candidate genes for ASD and ID, Kou et al (Y. Kou, Betancur, Xu, Buxbaum, & Ma'ayan, 2012) found that both conditions share common pathways and have similar clusters of genes (**Figure 4.3**). This approach is very useful to understand the cellular and molecular mechanisms



## CHAPTER 5:

### MATERIALS AND METHODS

#### **5.1 ASD Samples**

##### **5.1.1 Italian ASD Cohort**

The Italian ASD cohort is formed by 228 Italian simplex families recruited at the Stella Maris Clinical Research Institute for Child and Adolescent Neuropsychiatry (Calambrone, Pisa, Italy). ASD diagnosis was based on the the Autism Diagnostic Interview-Revised (ADIR) (Lord et al., 1994) and the Autism Diagnostic Observation Schedule (ADOS) (Lord et al., 2000) and a clinical evaluation was undertaken in order to exclude known syndromes associated with autism. Standard karyotyping, fragile-X testing, EEG, and array-based comparative genomic hybridization (aCGH) were obtained for all probands.

DNA was extracted from blood with the QIAGEN DNA Blood extraction kit.

80 Italian ASD families were included in the Autism Genome Project (AGP) genome-wide study (combined sample of stage 1 and stage 2 families) (Pinto et al., 2014; Pinto et al., 2010) (see paragraph 5.1.2).

The three Italian ASD individuals carrying the CNVs reported in this thesis (3456\_3; 3474\_3 and 3423\_3) belong to Italian ASD cohort included in the AGP genome-wide study.

135 ASD subjects belonging to Italian ASD cohort were analysed in the *CHRNA7* mutation screening.

##### **5.1.2. AGP Cohort**

The AGP collection, including a total of 2,845 ASD families (Pinto et al., 2014), comprises simplex families with only one ASD case and multiplex families with at least two ASD individuals. The affection status was determined using the diagnostic tools Autism Diagnostic Interview-Revised (ADIR) (Lord et al., 1994) and the Autism Diagnostic Observation Schedule (ADOS) (Lord et al., 2000). Three categories of affected individuals were established (strict, broad and spectrum ASD), based on proband diagnostic measures:

- the strict class included affected individuals who met criteria for autism on both ADI-R and ADOS instruments;
- the broad class included individuals who met full autism criteria on one diagnostic instrument and ASD criteria on the other one;

- the spectrum class included all individuals who were classified as ASD on both the ADI-R and ADOS or who were not evaluated on one of the instruments but were diagnosed with autism on the other instrument.

Given the international and multi-site nature of the project and the range of chronological and mental ages of the probands, a range of cognitive tests were administered, and standard scores were combined across tests to provide consolidated IQ estimates.

DNA, extracted from blood, buccal-swabs or cell-lines, has been genotyped with Illumina Human 1M-single and Illumina Human 1M-Duo BeadChip arrays (Pinto et al., 2014; Pinto et al., 2010). All CNVs reported in ASD samples are high-confidence CNVs predicted by intersecting CNV calls from at least two algorithms between iPattern, PennCNV and QuantiSNP. This strategy ensures maximum specificity because each of these algorithms employs unique strategies for CNV calling, allowing their strengths to be leveraged. Previous analysis showed that validation rates were approximately 95% for CNVs identified using this method (Pinto et al., 2010).

CNV data on a total of 2,147 European ASD families from the AGP study were used to examine the frequency of *CTNNA3* exonic deletions in autism.

### 5.1.3 IMGSAC Cohort

The International Molecular Genetic Study of Autism Consortium (IMGSAC) ((IMGSAC), 2001) collected multiplex ASD families from different countries (UK, Netherlands, France, USA, Germany, Denmark, Greece). The individuals had predominantly a Caucasian origin (>90%). Clinical diagnosis of ASD was made using the ADI-R, ADOS or ADOS-Generic (ADOS-G).

The multiplex families had fragile X testing and were karyotyped whenever possible. The following items were used to exclude individuals so as to keep the sample set more homogeneous:

- Any medical condition likely to be etiological (e.g. tuberous sclerosis, fragile X, focal epilepsy, infantile spasms, single gene disorders involving the central nervous system).
- Any neurological disorder involving pathology above the brain stem, other than uncomplicated non-focal epilepsy.
- Contemporaneous evidence, or unequivocal retrospective evidence, of probable neonatal brain damage. Clinically significant visual or auditory impairment after correction.
- Rearing in adoptive or foster homes.
- Institutional rearing during the first 4 years when there is any possibility that this led to an autistic-like picture.

- Any circumstances that might possibly account for the picture of autism (e.g. very severe nutritional or psychological deprivation).
- Birth in a place making it difficult to obtain satisfactory obstetric data (this would ordinarily exclude those born in a developing country).
- Autism secondary to some other psychiatric disorder (e.g., schizophrenia), but not psychiatric co-morbidity.
- Observational data that cast doubt on the diagnosis.
- *In vitro* fertilisation as a means of conception.
- Cases arising from consanguineous marriage.

A number of features were not used as exclusion criteria, these are: epilepsy (unless focal); psychiatric co-morbidity; head circumference over 97%; mental illness in a parent and belonging to an ethnic minority.

Genomic DNA was extracted from blood using Nucleon® kit (IMGSAC, 1998). In a minority of cases in which a blood sample could not be obtained, DNA was extracted from buccal swabs (IMGSAC, 1998). In addition, when possible, lymphoblastoid cell lines (LCLs) have been generated from peripheral blood leukocytes.

96 unrelated ASD probands selected from the IMGSAC multiplex families were included in the *KLHL23* and *PHOSPHO2-KLHL23* mutation screening.

## **5.2 Control samples**

### **5.2.1 AGP control cohorts**

CNV data from 6,639 European controls belonging to the AGP control cohorts were used to examine the frequency of *CTNNA3* exonic deletions in control populations.

Control cohort microarray data include 1,287 unrelated European control subjects from the Study of Addiction: Genetics and Environment cohort (SAGE) (Bierut et al., 2010) genotyped with Illumina Human 1 M-single BeadChip arrays, 1,123 Northern Europeans from the German PopGen project (POPGEN) (Krawczak et al., 2006) genotyped on the Affymetrix 6.0 SNP array (Affymetrix, Santa Clara, CA, USA), 1,234 individuals of European descent from the Ottawa River Valley (OHI) (Stewart et al., 2009) genotyped on the Affymetrix 6.0 SNP array, 1,320 European control subjects routinely seen at primary care and well-child clinic practices within the Children's Hospital of Philadelphia (CHOP) Health Care (Shaikh et al., 2009) genotyped with Illumina 550 K BeadChip, 435 unrelated European control subjects from the Ontario Colorectal Cancer Case-Control study (OC) (Figueiredo et al., 2011) genotyped with the Illumina 1 M single array and 1,240 European

controls from the NHGR-CIDR Visceral Adiposity Study (Fox et al., 2012) genotyped on Illumina 1 M-duo BeadChip arrays. For all these control samples (except for the CHOP samples, for which the CNV data are available at <http://cnv.chop.edu>), the heterozygous state of exonic *CTNNA3* deletions has been determined by inspecting the genotypes and/or plotting B allele freq and log R ratios for each region.

Statistical comparison of *CTNNA3* exonic deletion frequencies between ASD cases and controls was performed using Fisher's exact test.

### 5.2.2 Italian controls

The Italian control sample consists of 174 unrelated individuals selected from University of Bologna with no psychiatric disorders. This sample was used to test the presence of the -86/-241 variants identified in the *CHRNA7* mutation screening.

### 5.2.3 European controls

462 Caucasian controls from the European Collection of Cell Culture (ECACC) (<http://www.hpacultures.org.uk/products/dna/hrcdna/hrcdna.jsp>), were instead used to test by PCR-RFLP the presence of the non-synonymous variant p.Met65Val detected in the *KLHL23* and *PHOSPHO2-KLHL23* mutation screening.

All data from either affected patients or their parents and controls, including informed consent, were handled in accordance with the local ethical committee's approved protocols and in compliance with the Helsinki declaration.

## **5.3 Identification of CNVs by Illumina Human 1M-Duo BeadChip array**

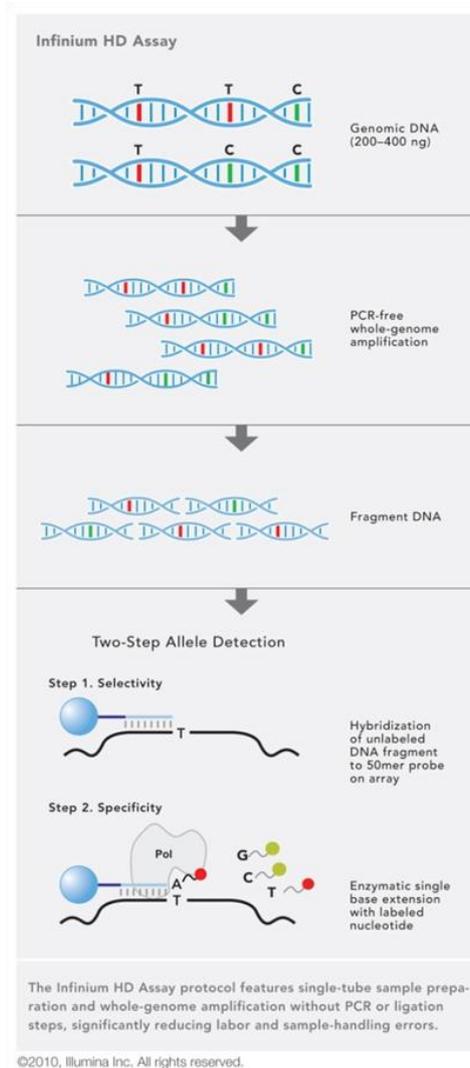
The Illumina Human 1M-duo BeadChip array generates genotype calls for more than 1 million markers using the Infinium HD technology (**Figure 5.1**) and it is able to genotype two samples per Chip.

This assay is based on a two-step detection process:

- a) fragments of the DNA of interest selectively hybridize to specific probes (50-mer oligonucleotides), designed to be complementary to the loci of interest, but stopping one base before the interrogated SNP;
- b) an enzymatic single-base extension incorporates a nucleotide labelled with a fluorescent dye, complementary to the base present at the SNP site.

## MATERIALS AND METHODS

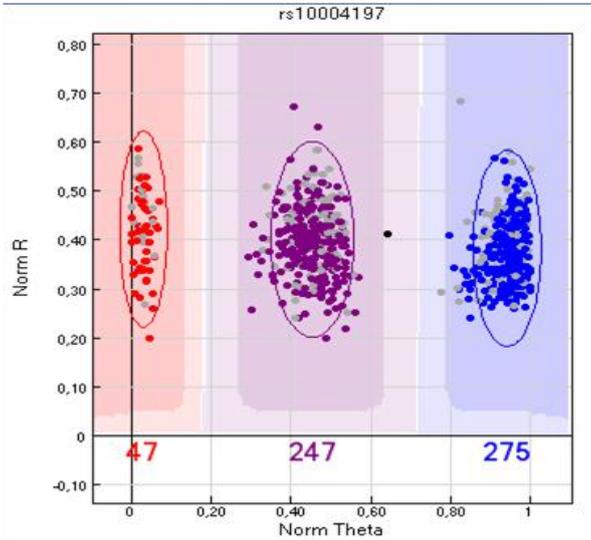
For each SNP, Cy3 and Cy5 fluorescence signals specify the two alleles (referred to as allele A and allele B). Dual-colour staining of the labelled nucleotides is followed by an image scanning performed by Illumina's iScan imaging system, which detects both colour and signal intensity. Homozygous genotypes are specified by red/red or green/green signals, heterozygous genotypes are indicated by red/green (yellow) signals.



**Figure 5.1.** Schematic representation of the Illumina Infinium HD Assay protocol.

The data generated by a SNP genotyping array can be analyzed and visualized with GenomeStudio Data Analysis software (Illumina). This program can convert the raw image scan into quantitative values and calculate the signal intensities for alleles A and B at each SNP, indicated as X and Y, respectively. The allele specific intensities are normalized using a proprietary algorithm of the Illumina GenomeStudio software: this procedure adjusts for background and makes markers more comparable to each other. Normalized allelic intensities are used to calculate the total fluorescent

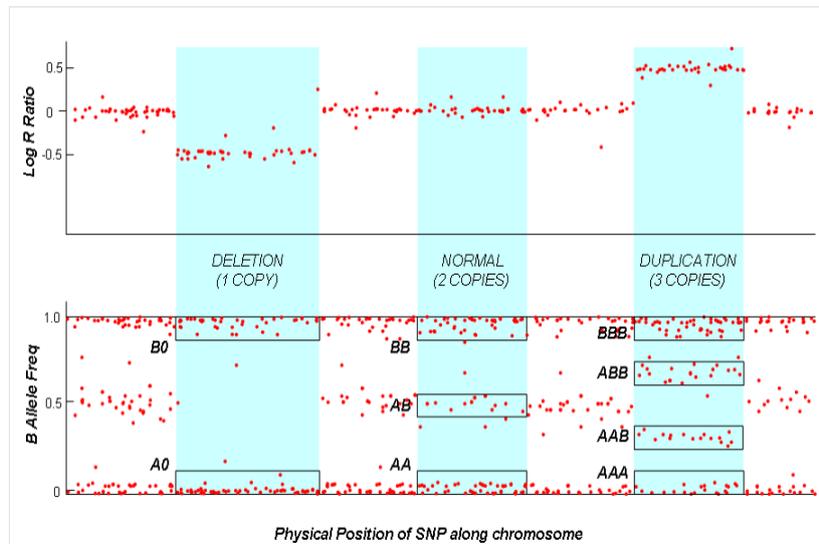
intensity signal ( $R$ ) and the allelic intensity ratio ( $\theta$ ).  $R$  is calculated as a combined SNP intensity:  $R_{\text{observed}} = X+Y$ , while  $\theta$  is calculated as  $\arctan(Y/X)/(\pi/2)$ .  $R$  and  $\theta$  values are calibrated to canonical genotype clusters generated from a large panel of normal samples, used to determine the  $R$  and theta values expected for each genotype (AA, AB and BB) (**Figure 5.2**).



**Figure 5.2** The graphical display of results in GenomeStudio is a Genoplot with data points color coded for the call (red=AA; purple=AB; blue=BB). Genotype are called for each sample (dots) by their signal intensity (Norm-R; y-axis) and Allele Frequency (Norm Theta; x-axis) relative canonical cluster position for a given SNP marker.

$R$  and  $\theta$  are then converted to two important measures: Log R Ratio (LRR) and B Allele Frequency (BAF). LRR, which represents a normalized measure of the total signal intensity at each SNP, derives from the comparison of the  $R_{\text{observed}}$  with the  $R$  obtained from a reference sample population ( $R_{\text{expected}}$ ) and it is calculated as  $\log_2(R_{\text{observed}}/R_{\text{expected}})$ . In autosomic regions without CNVs (copy number = 2), LRR is  $\sim 0$ : LRR lower than zero may indicate a deletion, LRR higher than zero a duplication. BAF, which derives from the normalized  $\theta$ , represents the proportion contributed by allele B to the total copy number. BAF represents an estimate of  $N_B/(N_A+N_B)$ , where  $N_A$  and  $N_B$  are the number of A and B alleles, respectively, therefore its value range from 0 to 1. BAF close to 1 indicate that all alleles for that marker are B alleles (e.g. BB, BBB or B/-), *viceversa* BAF close to 0 indicate that all alleles for that SNP are A alleles (e.g. AA, AAA or A/-), values close to 0.5 indicate a heterozygous genotype AB.

These two transformed parameters, LRR and BAF, are plotted along each chromosome for all SNPs on the array and can be then visually inspected (**Figure 5.3**). The exported values of LRR and BAF for each SNP in each individual can be used for the identification of changes in copy number by QuantiSNP and PennCNV.



**Figure 5.3.** Examples of LRR and BAF plots for a deleted region (1 copy, genotype B/- or A/-), region with 2 normal copies (three possible genotypes for each SNP: AA, AB, BB) and duplicated region (3 copies, four possible genotypes for each SNP: AAA, AAB, ABB, BBB).

The validation and the resolution of the extension of the three CNVs identified in probands 3456\_3 (*CTNNA3* deletion); 3474\_3 (*CHRNA7* duplication) and 3423\_3 (2q31.1 deletion), were carried out using the SNP data from the Illumina 1M-duo array generated as part of the large genome-wide scan for CNVs performed by the AGP (Pinto et al., 2014). The presence of the CNVs was also verified by visual inspection of the log R ratios and B allele frequencies in GenomeStudio.

### **5.4 Validation of CNVs by Real-Time PCR**

CNVs were validated by real-time quantitative PCR (qPCR) using Fast SYBR Green Master Mix (Life Technologies). Briefly, SYBR Green is a non-specific dye that emits fluorescence when it intercalates with double stranded DNA. In a qPCR experiment, the fluorescence is measured after each extension step and this allows the monitoring of the increasing amount of DNA produced during the PCR reaction. The analysis of qPCR data is based on the threshold Cycle (Ct), which represents the cycle at which the fluorescence passes the threshold level, within the exponential phase, and it is a relative measure of the amount of target in the reaction.

All the primers used for the CNV validation were designed with Primer3 (<http://bioinfo.ut.ee/primer3-0.4.0/primer3/>). Oligonucleotide primers were designed taking into account an optimal primer length of 20-22 nucleotides, a GC content of 40–60% and an optimal PCR product size of 90-140 bp. In order to evaluate the PCR efficiency, a standard curve was set up for each primer pair, using three replicates of a control DNA and 4 template concentrations (four DNA template amounts deriving from a 1:2 serial dilution). The primers used for the CNV

validation were selected to have a PCR efficiency in the range of 90%-110%. We made sure the concentration of the DNA we would be using occurred within the serial dilution range.

For the relative quantification, the DNA template was used with a concentration of 5 ng/μl.

Each reaction was set up in triplicate using the following conditions:

qPCR Reaction MIX	Amount (μl)
DNA template	5
SYBR Green MasterMix (2X)	10
Primer F (10 μM)	1
Primer R (10 μM)	1
H <sub>2</sub> O	3
Final volume	20

The qPCR experiments were run on the 7500 fast Real-Time PCR system (Life Technologies) according to the following qPCR program, which includes three main stages:

- 1) Polymerase activation: incubation at 95°C for 20”;
- 2) Anneal/extend: 40 cycles of incubation at 95°C for 3” , at 60°C for 30”;
- 3) Melting curve: incubation at 95°C for 15” , followed by an incubation at 60°C for 1’ , with an increase of 0.5°C each cycle.

The melting curve analysis, also called dissociation curve, consists of a step in which the temperature is gradually increased while the fluorescence is constantly monitored. When the temperature is high enough, the double strands of DNA fragments are denatured and the SYBR Green dye dissociates from the ds-DNA, causing a decrease in fluorescence. The melting temperature (T<sub>m</sub>) depends on primer features (such as sequence complementarity and G-C composition) and reaction conditions. Analysis of melting curves allows the detection of primer-dimers or other non-specific PCR products that could reduce the PCR efficiency and give spurious fluorescence signals. Primer-dimers usually have lower T<sub>m</sub> compared to the PCR products, since they have a smaller size.

For each sample, the qPCR data were compared against a control gene (*FOXP2*) (**Table 5.16**) and a control subject predicted to have two normal copies of the tested region. The number of copies of each amplified fragment was calculated using the 2<sup>-ΔΔCt</sup> method (Livak & Schmittgen, 2001):

$$\Delta C_t \text{ sampleDNA} = C_t \text{ fragment of interest} - C_t \text{ FOXP2}$$

$$\Delta C_t \text{ control DNA} = C_t \text{ fragment of interest} - C_t \text{ FOXP2}$$

$$\Delta \Delta C_t = \Delta C_t \text{ sample DNA} - \Delta C_t \text{ control DNA}$$

$$\text{ratio} = 2^{-\Delta \Delta C_t}$$

$$\text{Copy number} = 2 * \text{ratio}$$

## **5.5 Mutation screening**

### **5.5.1 Primers design**

DNA sequences and genomic organization of the genes analyzed are from the UCSC genome Browser (hg 19). All the primers were designed using Primer3 (<http://bioinfo.ut.ee/primer3-0.4.0/primer3/>). All coding exons, intron-exon boundaries and the 5-UTR were sequenced by Sanger method.

### **5.5.2 Polymerase Chain Reaction (PCR) assay**

The PCR assays were performed using the kit provided with AmpliTaq Gold DNA Polymerase (Life Technologies). AmpliTaq Gold is a *Hot Start* polymerase, that is inactive at room temperature and is activated during the initial denaturing step at 95°C.

The PCR conditions were optimized for each primer pair using four different Mg<sup>2+</sup> concentrations (1,5 mM; 2 mM; 2,5 mM or 3 mM). The reactions were set up in a final volume of 15 or 20 µl using 30 or 50 ng of template DNA according to the following PCR reaction setup:

<b>Component</b>	<b>20 µl rxn</b>	<b>15 µl rxn</b>	<b>Final conc.</b>
PCR Buffer (10X)	2	1,5	1X
MgCl <sub>2</sub> (25 mM)	variable	variable	1,5mM; 2mM; 2,5mM or 3mM
dNTPs mix (25 mM)	0,16	0,12	0,2 mM each
Forward primer (10 µM)	0,5	0,4	0,25 mM
Reverse primer (10 µM)	0,5	0,4	0,25 mM
Taq Gold Polymerase (5 U/µl)	0,06	0,05	0,015 U/ µl
H <sub>2</sub> O	variable	variable	variable
DNA template	variable	variable	30-50 ng

The amplification of some fragments containing a high GC content was performed adding to the PCR mix the dimethyl sulfoxide (DMSO) and/or 7-deaza-2'-deoxiguanosina-5'-trifosfato (Deaza GTP), which improve the denaturation of the template DNA, help the reaction and overcome the problem of non-specific annealing. The GC pair has a higher number of hydrogen bonds compared to the AT pair, therefore, GC-rich stretches are more stable and require a higher melting temperature.

The PCR conditions were optimized for each primer pair used in the screening. For some of them, we used a traditional PCR program, with one annealing temperature (Ta):

## MATERIALS AND METHODS

<b>program</b>	
Initial denaturation	95°C for 15 minutes
Amplification (35 cycles)	95°C for 30 seconds
	Ta <sub>1</sub> °C for 30 seconds
	72°C for seconds/minutes

For other fragments, in order to increase the specificity and the yield of the reaction, we used a Touch-Down PCR program (TD) (Korbie & Mattick, 2008). In a TD program, the initial Ta (1) is higher to ensure a specific annealing of the primers to the template, then it is progressively decreased until it reaches a second, lower Ta (2), which is maintained constant for remaining amplification cycles.

<b>TD program</b>	
Initial denaturation	95°C for 15 minutes
Touch-Down (Ta decreases 0,5°C at each cycle)	95°C for 30 seconds or 1 minute
	Ta <sub>1</sub> °C for 30 seconds
	72°C for 30-45 seconds or 1 minute
Amplification (30 cycles)	95°C for 30 seconds or 1 minute
	Ta <sub>2</sub> °C for 30-45 seconds or 1 minute
	72°C for 1 minute

Alternatively, some fragments were amplified using KAPA HiFi HotStart PCR (Resnova), which contains a novel HotStart DNA Polymerase, engineered for fast and versatile high-fidelity PCR and it is able to amplify fragments up to 15 kb from genomic DNA. The manufacturer's protocol is the following:

PCR mix	25 µl rxn	Final conc.
5X KAPA HiFi Fidelity Buffer (contains 2.0 mM Mg <sup>2+</sup> at 1X)	5	1X
KAPA dNTP Mix (10 mM each dNTP)	0,75	0,3 mM each
Forward primer (10 µM)	0,75	0,3 µM each
Reverse primer (10 µM)	0,75	0,3 µM each
DMSO (100%) (for amplicons with a GC content >70%)	1,25	5%
KAPA HiFi HotStart DNA Polymerase (1 U/µl)	0,5	0,5 U/25 µl rxn
PCR grade water	11	-
Template DNA (10ng/µl)	5	50 ng

The PCR reaction was performed on the Veriti thermocycler (*Life Technologies*) using the following “touch-down” protocol:

<b>KAPA HiFi HotStart cycling protocol</b>	
Initial denaturation	95°C for 5 minutes
Touch-Down ( $T_a$ decreases 0,5°C at each cycle)	98°C for 20 seconds
	$T_{a1}$ °C for 15 seconds
	72°C for 15-60 seconds/kb
Amplification (25 cycles)	98°C for 20 seconds
	$T_{a2}$ °C for 15 seconds
	72°C for 15-60 seconds/kb
Final extension	72°C for 1-5 minutes

The PCR products were visualized with a UV transilluminator after electrophoresis on a 0,8%, 1%, 2% or 2,5% agarose gel (the percentage of agarose varies depending on the PCR fragment) and GelRed staining (*Biotium*). The 100 bp and 1kb DNA ladders (*NEB*) were loaded to check the PCR fragment size.

### 5.5.3 PCR purification

Prior to sequencing, the PCR products were purified with Exonuclease I (ExoI) and Shrimp Alkaline Phosphatase (SAP) using the ExoSAP-IT PCR Clean-up Kit (GE Healthcare). The enzyme ExoI removes single-strand DNA (primers or intermediate products), the enzyme SAP removes the unconsumed dNTPs remaining in the PCR mixture. The PCR cleanup was performed adding the Exo-SAP reaction mix to the PCR products, followed by a two-step incubation:

<b>EXOSAP mix</b>	<b>Amount (µl)</b>	<b>Program</b>	
SAP enzyme	0,5	Treatment	37°C x 15 minutes
ExoI enzyme	0,5	Enzymatic inactivation	80°C x 15 minutes
PCR product	6		
Final volume	10		

Alternatively, PCR products were purified using a Millipore Multi Screen PCR plate (Millipore) according to the following procedure:

1. at the conclusion of the PCR reaction, pipet 70 µl H<sub>2</sub>O to the reaction products and transfer the final volume into the 96 well Millipore Multiscreen PCR plate;
2. place the Multiscreen plate on top of the Multiscreen Vacuum and apply vacuum until no liquid remains in the wells;
3. to resuspend PCR products, add 25-30 ml of H<sub>2</sub>O to each well. Cover plate with cover provided and shake vigorously on a plate shaker for 10 min.
5. remove as much as possible of the resuspended products from each well and place in a new, 96 well plate for storage.

### 5.5.4 Sanger Sequencing reaction

The Sanger Sequencing reaction was performed using the BigDye Terminator kit v1.1 (Life Technologies) according to the following manufacturer's protocol:

Sequencing mix	Amount (µl)	Program	
Big Dye terminator buffer (5X)	1,75	Initial Denaturation	96°C for 1 minute
Big Dye terminator	0,5	Amplification (29 cycles)	96°C for 10 seconds
Primer (10 µM)	0,16		50°C for 5 seconds
H <sub>2</sub> O	variable		60°C for 4 minutes
PCR product	variable		
<b>Final volume</b>	10		

### 5.5.5 Ethanol-EDTA Precipitation of Sequencing Reactions

In order to remove unincorporated dNTPs and primers, the products of sequence reactions were precipitated using the following procedure:

1. A "precipitation mix" was prepared and added to each sample (volume of the sequence reaction: 10 µl). This solution was composed by 10 µl of H<sub>2</sub>O, 2.5 volumes of 100% EtOH (55 µl), 1/10 of the volume of sodium acetate (NaAc) 3M pH 5.2 (2 µl).
2. The samples were spun in a centrifuge for 30 minutes at 3000 rpm (4°C).
3. The supernatant was discarded.
4. The pellet was rinsed with 70 µl of 70% EtOH.
5. The samples were spun in a centrifuge for 10 minutes at 3000 rpm (4°C).
6. The supernatant was discarded.
7. The microplates were spun upside down for 1 minute at 300 rpm, to remove traces of EtOH or NaAc.
8. The pellet was dried at room temperature and stored away from direct light.

The purified sequencing products were then resuspended in 15 µl of Injection Solution (DNA Sequencing Reaction Cleanup kit, *Millipore*) by pipetting up and down several times and/or using a microplate shaker. The sequencing products were run on the ABI PRISM 3730 DNA analyser (Life Technologies). The sequencing products were analysed and compared to the reference sequence using the software Sequencher 5.0 (*Gene Code Corporation*).

## 5.6 Prediction tools

We used two online bioinformatic tools to predict the possible impact of non-synonymous coding SNPs on the structure and function of the human proteins: PolyPhen-2 (**P**oly**M**orphism

**Phenotyping v2**, <http://genetics.bwh.harvard.edu/pph2/>) (Adzhubei et al., 2010) and SIFT (Sorting Tolerant From Intolerant, <http://sift.jcvi.org/>) (Kumar, Henikoff, & Ng, 2009). SIFT is a sequence-based algorithm that uses sequence homology to predict the effect of amino-acid replacements, assuming that important positions in a protein sequence have been conserved throughout evolution. Polyphen-2 instead is a prediction algorithm, that incorporates sequence conservation information and protein structure annotations to predict the impact of the non-synonymous change. For SIFT, the score ranges from 0 (damaging) to 1 (neutral); for Polyphen-2, the score ranges from 0 (neutral) to 1 (damaging).

### **5.6 RNA extraction and cDNA synthesis**

Total RNA from human frozen brain tissues (30-40 mg) was extracted using the Qiagen Total RNA kit (Qiagen), according to the manufacturer's protocol.

Total RNA from whole blood was instead extracted using the RiboPure™-Blood Kit (Life Technologies) according to the manufacturer's protocol. 300-500 µl of blood can be processed immediately, or they can be stored in RNAlater Solution (a solution for the stabilization of RNA in whole blood) for a few days at ambient temperature or for longer periods at -20°C prior to RNA extraction.

RNA quality was assessed on a 1% agarose gel and its concentration and purity was determined by UV Spectrophotometry.

Reverse transcriptase PCR (RT-PCR) was performed using the Superscript III First Strand Synthesis SuperMix (Life Technologies) using approximately 1 µg of RNA as template.

### **5.7 Analysis of the candidate gene *CTNNA3***

#### **5.7.1 Characterisation and segregation analysis of *CTNNA3* deletions**

The maternal and the paternal exon 11 *CTNNA3* microdeletions segregating in family 3456 were amplified via several PCR assays using the AmpliTaq Gold polymerase (Life Technologies) (**Table 5.1**). The paternal 301-bp and the maternal 949-bp deletion spanning amplicons were purified using Exosap (GE Healthcare, Little Chalfont, UK) and then sequenced using BigDye Terminator kit v1.1 (Life Technologies), as described before, to determine the exact boundaries of the deletions.

An additional primer pair that amplifies exon 11 of *CTNNA3* (not deleted allele) was subsequently used to confirm that the identified deletions in family 3456 were in the heterozygous or homozygous status (**Table 5.1**).

The experimental validation of *CTNNA3* exonic deletions in four other ASD families was carried out by qPCR using Fast SYBR Green as reported in section 5.4. Each assay was conducted with at least three sets of primers corresponding to the region of interest (**Table 5.2**) and another mapping to the control region on *FOXP2* gene (**Table 5.16**). The number of copies of each amplified fragment was calculated using the  $2^{-\Delta\Delta C_t}$  method. The parents and additional affected or unaffected siblings were also tested for inheritance and segregation of CNVs, respectively.

Numbering for *CTNNA3* exons is based on the Reference Sequence (RefSeq) NM\_013266.

### 5.7.2 *CTNNA3* and *LRRTM3* exon sequencing

All coding exons, intron-exon boundaries and the 5'-UTR of the long isoforms of *CTNNA3* (NM\_001127384; NM\_013266) and the nested gene *LRRTM3* (NM\_178011.4) have been amplified by PCR in all members of four families carrying *CTNNA3* exonic deletions .

Primer sequences and PCR conditions used for amplification (20 amplicons for *CTNNA3* and 5 amplicons for *LRRTM3*) are listed in **tables 5.3** and **5.4**. PCR products were purified using Exosap (GE Healthcare) and then sequenced using BigDye v1.1 (Life Technologies).

### 5.7.3 *CTNNA3* expression

Human brain samples from the frontal cerebral cortex and cerebellum were obtained from deep frozen (-80°C) slices of two informative heterozygous adult control subjects for a *CTNNA3* exon 13 coding SNP (rs4548513; p.Ser596Asn).

Total RNA from human frozen brain tissues (30-40 mg) was extracted using the Qiagen Total RNA kit (Qiagen) and RT-PCR was performed using the Superscript III First Strand Synthesis SuperMix (Life Technologies) as described in section 5.6. Two microliters of complementary DNA (cDNA) was used for testing *CTNNA3* expression in the human frontal cerebral cortex and cerebellum using a forward primer designed in exon 10 and a reverse primer mapping in exon 15, in order to amplify a cDNA fragment including the coding SNP rs4548513 (pSer596Asn). PCR products were purified and sequenced as described before using primers mapping in exon 12 and exon 14. (**Table 5.5**)

### 5.7.4 Western blot analysis

The hippocampus and cortex were dissected from three mice (C57Bl/6 N) at each time point (from P0 to P90) pooled together. Samples were lysed with lysis buffer containing 1% SDS and boiled. Forty micrograms of total proteins were loaded onto a 4%-12% polyacrylamide gel (Life Technologies) and then transferred to a nitrocellulose membrane (Whatman). Filters were

hybridized with antibodies against N-catenin (Santa Cruz Biotechnology, Inc., Dallas, TX, USA) or T-catenin [30] or GAPDH (Millipore, Billerica, MA, USA) as housekeeping control and then revealed by using HRP-conjugated specific secondary antibodies (BioRad, Hercules, CA, USA) and ECL (GE Healthcare). ImageJ was used to quantify bands. Experiments were done according to the animal protocols approved by the Institutional Animal Care and Use Committee San Raffaele (IACUC) (San Raffaele, Milan, Italy) and were approved by the National Ministry of Health, IACUC ID 470. All experiments were carried out in accordance with the guidelines established by the European Community Council Directive of 24 November 1986 on the use of animals in research (86/609/EEC). All efforts were made to minimize animal suffering and to use only the number of animals necessary to produce reliable results.

## **5.8 Analysis of the candidate gene *CHRNA7***

### **5.8.1 Validation of the *CHRNA7* duplication**

All members of the discovery family 3474 were analyzed for the *CHRNA7* microduplication by qPCR as described in section 5.4. A primer pair mapping in exon 3, specific for *CHRNA7* was used to confirm the predicted duplication (**Table 5.6**). 67 additional ASD probands were tested by qPCR using the same probe on exon 3.

### **5.8.2 *CHRNA7* sequencing**

All 10 exons and intron-exon boundaries of *CHRNA7* were sequenced by Sanger method in 135 ASD individuals.

The first four exons, which are specific for *CHRNA7*, were amplified using exon specific primers corresponding to flanking intronic sequences using the AmpliTaq Gold polymerase (Life Technologies), purified with Exosap (GE Healthcare) and then subjected to Sanger sequencing.

The amplification of the fragments containing exon 1 and exon 2 was particularly problematic, since these regions are located in a CpG island extending to intron 2, and therefore they were amplified adding to the PCR mix 7-deaza-2'-deoxyguanosine-5'-triphosphate (Deaza GTP) and/or DMSO (5%). The primers and the PCR conditions used for amplification of the first four exons are listed in **table 5.7**.

To selectively amplify the *CHRNA7* exons 5-10, which are duplicated in the *CHRFAM7A* gene, two long range PCRs (LR-PCRs) were performed to amplify: a) a segment encompassing exon 5 to exon 8 (x5-x8 LR-PCR) using a primer F outside the duplication (mapping to *CHRNA7* specific intron 4) and a primer R in the duplicated region (mapping to intron 8); b) a segment encompassing

exon 9 to exon 10 (x9-x10 LR-PCR) using primer F in the duplicated region (mapping to intron 8) and a primer R mapping in a *CHRNA7* specific region located at 3'UTR. LR-PCR was performed using KAPA HiFi HotStart PCR (Resnova) according to the manufacturer's protocol, as described in section 5.5.2. The primers and the PCR conditions used for amplification of the two LR-PCRs are listed in **table 5.8**.

All LR-PCR products were purified using a Millipore Multi Screen PCR 96 well (section 5.5.3) and then they were directly sequenced by Sanger method using specific primers for each exon (**Table 5.10**). In particular to sequence the *CHRNA7* exon 10, we carried out an additional PCR assay with the AmpliTaq Gold polymerase using 3 µl of the purified product from x9-x10 LR-PCR and primers specific for exon 10 (**Table 5.9**).

The *CHRNA7* intron 5 and intron 9 were amplified using primers flanking these two regions, purified with Exosap and then sequenced with the same primers (this protocol leads to co-amplification of both *CHRNA7* and *CHRFAM7A*) or they were amplified using 3 µl of the purified product from the x5-x8 and x9-x10 LR-PCRs (**Table 5.7**).

## **5.9 Analysis of the microdeletion on chromosome 2q31.1**

### **5.9.1 Validation of the microdeletion in the discovery pedigree**

The presence of the deletion was validated in blood-derived genomic DNA from family 3423 by qPCR using Fast SYBR-green as described in section 5.4. The copy number was normalized against the father 3423\_1.

Eight primer pairs were used (**Table 5.11**): two qPCR fragments, mapping to *KLHL23* exon 3 and *METTL5* intron 7, are located outside the deletion, whereas six probes, mapping to intron 3 and exon 4 of *KLHL23*, to intron 9 and intron 11 of *SSB*, and to intron 6 and exon 6 of the *METTL5* gene, are inside the deletion.

### **5.9.2 *KLHL23* and *PHOSHO2-KLHL23* expression analyses**

Total RNA from all member of family 3423 was extracted from peripheral blood using the RiboPure™-Blood Kit (Life Technologies) and cDNA was synthesized using the Superscript III First Strand Synthesis SuperMix (Life Technologies). *KLHL23* expression in family 3423 was tested by quantitative RT-PCR with Sybr Green using 1 µl of cDNA. Two qPCR fragments were used: one spanning exons 2-3 (outside of the microdeletion) and another one spanning exons 3-4 (exon 4 maps inside the microdeletion) (**Table 5.12**). All the data were normalized using the housekeeping gene *GUSB* as reference gene (**Table 5.16**). Expression levels were also normalized

against a control individual. The  $2^{-\Delta\Delta Ct}$  method was applied to estimate the difference in the gene expression between samples:

$$\Delta Ct_{\text{sample cDNA}} = Ct_{\text{fragment of interest}} - Ct_{GUSB}$$

$$\Delta Ct_{\text{control cDNA}} = Ct_{\text{fragment of interest}} - Ct_{GUSB}$$

$$\Delta\Delta Ct = \Delta Ct_{\text{sample cDNA}} - \Delta Ct_{\text{control cDNA}}$$

$$\text{Expression level} = 2^{-\Delta\Delta Ct}$$

*KLHL23* and *PHOSPHO2-KLHL23* expression was analysed in a multiple tissue cDNA panel (Human MTC Panel I, Clontech) by PCR with the TaqGold polymerase using 1µg of cDNA from each tissue and two primer pairs: one primer pair specific for *KLHL23* and another primer pair specific for the *PHOSPHO2-KLHL23* gene. PCR primers and condition are reported in **Table 5.13**.

### 5.9.3 *KLHL23* and *PHOSHO2-KLHL23* mutation screening

The entire open reading frames of *KLHL23* (NM\_144711.5) and *PHOSPHO2-KLHL23* (NM\_001199290.1) were screened in the 3423\_3 individual and in 96 additional unrelated ASD probands using primers and reaction conditions reported in **tables 5.14** and **5.15**. Probands were selected from the *International Molecular Genetic Study of Autism Consortium* (IMGSAC) multiplex families (see paragraph 5.1.3).

Bioinformatic analysis of novel non synonymous variants was carried out using Poly-Phen2 (<http://genetics.bwh.harvard.edu/pph2>) and SIFT (<http://sift.jcvi.org/>) (section 5.6).

### 5.9.4 PCR-Restriction Fragment Length Polymorphism (RFLP) analysis

PCR-Restriction Fragment Length Polymorphism (RFLP)-based analysis was used for SNPs genotyping. The first step in a PCR-RFLP analysis is the amplification of a fragment containing the variation. Since the presence or absence of the restriction enzyme recognition site results in the formation of restriction fragments of different sizes, allele identification can be done by electrophoretic resolvment of the fragments.

We used the bioinformatic program INSIZER (<http://zeon.well.ox.ac.uk/git-bin/insizer>, not available anymore) to find a restriction endonuclease that could specifically discriminate between the two allelic variants (A>G) located in exon 2 of *KLHL23*.

The PCR fragment containing exon 2 was amplified with the same primer pair used for *KLHL23* mutation screening (**Table 5.14**). Subsequently, the digestion reaction was assembled using the enzyme BtsI (NEB, *New England Biolabs*) and the following conditions:

Digestion mix	Amount (µl)	Program	
NEB buffer 10X	1,5	Digestion	55°C for 3 hours
BtsI enzyme (2 U/µl)	0,2	Heat inactivation	80°C for 20 minutes
BSA (10X)	1,5		
H <sub>2</sub> O	8,8		
PCR product	3		
<b>Final volume</b>	<b>15</b>		

The restriction fragments were separated by electrophoresis on a 2,5% agarose gel (0,7 g of agarose and 0,8 g of Nusieve).

## **5.10 Analysis of the *CADPS2* gene**

### **5.10.1 Microsatellite analysis**

Two microsatellite markers mapping to *CADPS2* intron 1 (*CADPS2\_21xAC*, genomic position: chr7:122,414,324-122,414,365) and intron 2 (*CADPS2\_26xAT*, genomic position: chr7:122,348,575- 122,348,626) were genotyped from blood-derived DNA of father, proband, and affected sister according to the following PCR conditions:

Component	10 µl rxn	Final conc.
PCR Buffer (10X)	1	1X
MgCl <sub>2</sub> (25 mM)	1	2,5mM
dNTPs mix (2 mM)	1	0,2 mM each
Forward primer FAM-labeled (10 µM)	0,25	0,5 µM
Reverse primer FAM-labeled (10 µM)	0,25	0,5 µM
Taq Polymerase (5 U/µl)	0,03	0,015 U
H <sub>2</sub> O	3,47	Up to
DNA template	3	30 ng

The following touch-down program was used:

TD program	
Initial denaturation	95°C for 15 minutes
Touch-Down (T <sub>a</sub> decreases 0,5°C at each cycle)	95°C for 30 seconds
	62°C for 30 seconds
	72°C for 30 seconds
	72°C for 30 seconds
Amplification (30 cycles)	95°C for 30 seconds
	57°C for 30 seconds
	72°C for 7 minutes

Samples were diluted 1/10 and 1 µl of dilution was run onto the automated 3730 ABI sequencing machine with LIZ(500) size marker (Life Technologies, Foster City, CA, USA). Genotype call was performed with GeneMapper v3.7.

### 5.10.2 CADPS2 expression analysis

Human brain samples from frontal cerebral cortex, amygdala, entorhinal cortex and cerebellum were obtained from deep-frozen (-80°C) slices of 3 adults showing either no significant histopathological changes or neurodegenerative histopathological lesions of variable severity. Total RNA from these brain tissues (30-40 mg) was extracted using with the QIAGEN Total RNA kit (QIAGEN) and cDNA was synthesized using the Superscript III First Strand Synthesis SuperMix (Life Technologies) according to the manufacturer's protocol.

PCR was performed using a forward primer designed in *CADPS2* exon 3 and a reverse primer mapping in exon 5 (**Table 5.17**), in order to amplify a cDNA fragment including the SNP rs2251761. PCR products were purified and sequenced as described before using the same PCR primers.

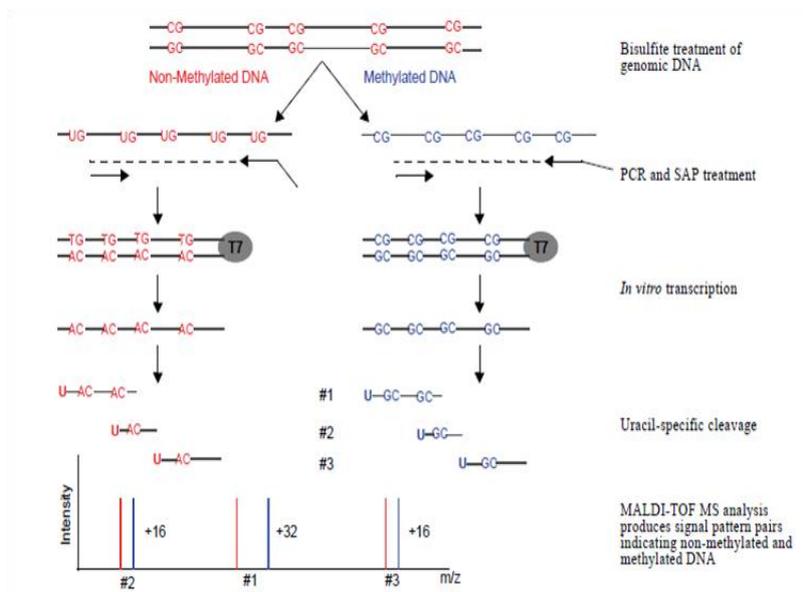
### 5.10.3 MALDI-TOF MS methylation analysis

DNA methylation of *CADPS2* gene was determined with the MassARRAY platform Sequenom, which uses the EpiTYPER assay, a tool for the detection and quantitative analysis of DNA methylation using base-specific cleavage and MALDI-TOF MS (Matrix-Assisted Laser Desorption/Ionization Time-of-Flight Mass Spectrometry). This technology consists of four main steps (**Figure 5.4**):

- 1) The first step is the bisulfite treatment of genomic DNA, where nonmethylated cytosine is converted to uracil, while methylated cytosine remains unchanged.
- 2) The second step is PCR amplification with a T7-promotor tagged primers and the neutralization of unincorporated dNTPs using shrimp alkaline phosphatase (SAP).
- 3) The third step is the generation of a single strand RNA molecule by in vitro RNA transcription followed by base-specific cleavage using RNase A. In this step the reverse strand of both methylated and not methylated regions is cleaved at every U to produce fragments containing at least one CpG site each.
- 4) The different cleavage products have a mass difference of 16 dalton for each CpGs between the methylated and un-methylated regions and they can be analyzed by mass spectrometry. In analyzing the mass spectrum, the relative amount of methylation can be calculated by comparing the

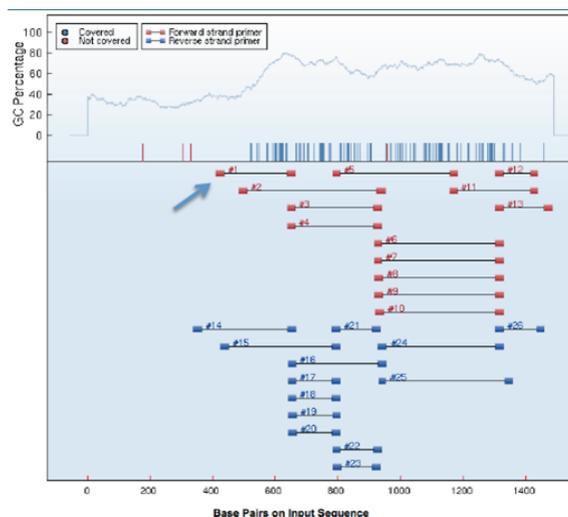
## MATERIALS AND METHODS

difference in signal intensity between mass signals derived from methylated and nonmethylated template DNA.



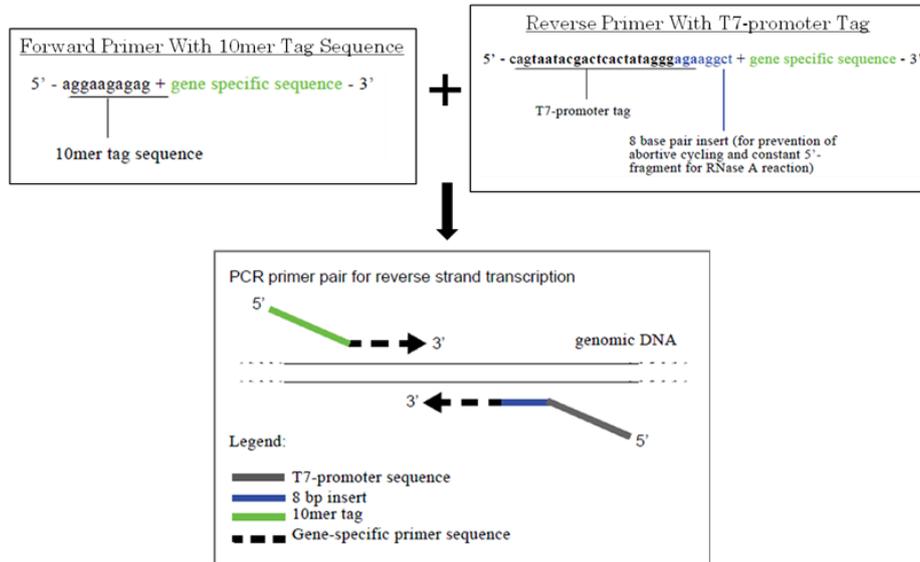
**Figure 5.4** Protocol of the MALDI-TOF MS methylation analysis

The EpiDesigner BETA software was used to design primers for PCR amplification of bisulfite treated DNA. This tool allows to design PCR amplicons which must cover the majority of CpG dense areas in close proximity to or overlapping with the annotated transcription start. The **figure 5.5** shows an overview of a genomic region submitted to software for predicting the CpG islands: the blue bars are the CpG covered by MS analysis, while the red bars are the CpG cannot be analyzed; all possible primer combinations to cover the area of interest are also represented.



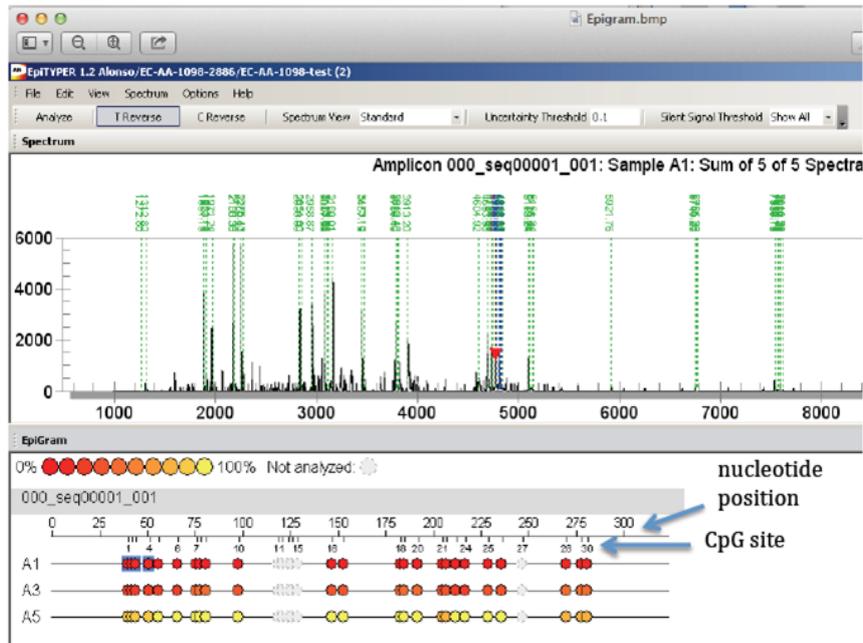
**Figure 5.5** Overview of the genomic area submitted to the EpiDesigner BETA software, with a pictorial view of the GC content and all possible primer combinations to cover the area of interest.

PCR amplification is carried out using a forward primer with a 10mer-tag sequence and a reverse primer with a T7 promoter-tag that is incorporated into the amplification product for in vitro transcription (**Figure 5.6**).



**Figure 5.6** Primers used for the T7-promoter-tagged PCR amplification of bisulfite-converted DNA

The software also displays the mass spectrum for all analyzed amplicons and identifies the selected CpG site. The software matches the reference spectrum (green lines) with the mass of the fragments generated by each amplicon (black lines) once they are run in the MassArray. The output is a graphical representation (epigram) where a series of individual CpGs are represented by circles on the same line and they are arranged in ascending numerical order from left to right. The numbers indicate the base pairs relative to the amplified PCR product and the position of the CpGs. Per amplicon, the color within the circle denotes the level of methylation found at this particular site in the selected sample. The color ranges from red (0% methylation) to yellow (100% methylation) (**Figure 5.7**).



**Figure 5.7** Outputs of the EpyTYPER BETA software. At the top panel is shown the spectrum pane for a specific amplicon; the reference spectrum is indicated by green lines, while the spectrum of mass of the fragment of interest is indicated by black lines. At the bottom panel is shown the epigram tab pane: this pane provides graphical representations of the CpG sites within the selected amplicon.

One microgram of genomic DNA was submitted to bisulfite conversion, using the EZ DNA Methylation Kit (Zymo Research, CA, USA), according to the manufacturer’s protocol, except for the conversion step consisting of 21 cycles at 95°C for 30 s and 50°C for 15 min. Universal unmethylated and methylated DNAs (Millipore, MA, USA) were used as internal controls. In addition, a 477-bp amplicon for the IGF2 region was included in the assay for quality assessment of bisulfate treated DNA. A region length spanning 1.9 kb of promoter and 1.8 kb of intron 1 of *CADPS2* gene (corresponding to genomic coordinates chr7: 122,524,253-122,528,311, hg19) was submitted to EpiDesigner BETA software (Sequenom) for predicting the CpG islands. Four amplicons including 92 CpGs were selected; 63 of these CpGs were analyzable through this methodology (**Table 5.18**).

25 µl H<sub>2</sub>O MilliQ and 4 µg resin were added to the cleaved fragments. 20 µl of the analytes was nanodispensed onto 384-element silicon chips preloaded with matrix. Mass spectra were collected by using a matrix-assisted laser desorption/ionization time-of-flight (MALDI-TOF) MS, and spectra’s methylation ratios were generated by EpiTYPER software V.1.2.22. Samples were run in duplicate/triplicate when possible.

#### 5.10.4 PCR and sequencing of *CADPS2* intron 1

The *CADPS2* intron 1 fragment containing SNP rs981321 (g.122,525,329 G>A) was amplified with primers flanking this region using the AmpliTaq Gold DNA Polymerase (Life Technologies) with 30 ng of genomic DNA in a final volume of 15  $\mu$ l. The PCR products were purified using a Millipore Multi Screen PCR 96 well (section 5.5.3) and then sequenced as described in section 5.5.4. The PCR primers and conditions are reported in **table 5.19**.

#### 5.10.5 Colony screening

To analyze the methylation status of four control individuals heterozygous for the SNP rs981321, we performed colony analysis of an intron 1 fragment containing both the SNP rs981321 and the three CpG sites (CpG\_14, CpG\_15, CpG\_16).

##### 1) Producing PCR products.

One microgram of genomic DNA was submitted to bisulfite conversion, using the EZ DNA Methylation Kit (Zymo Research, CA, USA), according to the manufacturer's protocol.

Six microliters of bisulfite-treated DNA was used for PCR amplification using the KAPA2G ReadyMix (KAPA Biosystems) in a final volume of 50  $\mu$ l. KAPA2G ReadyMix allows the amplification of DNA fragments with high GC- or AT content and it contains a HotStart DNA Polymerase (1 U per 25  $\mu$ l reaction) in a proprietary reaction buffer containing dNTPs (0.2 mM of each dNTP at 1X), MgCl<sub>2</sub> (2 mM at 1X) and stabilizers. The PCR protocol is the following:

Component	50 $\mu$ l rxn	Final conc.
KAPA 2G (2X)	25	1X
MgCl <sub>2</sub> (25 mM)	1	0,5mM
DMSO 100%	2,5	5%
Forward primer (10 $\mu$ M)	2,5	0,5 $\mu$ M
Reverse primer (10 $\mu$ M)	2,5	0,5 $\mu$ M
H <sub>2</sub> O	10,5	
DNA template	6	

The PCR primers and PCR program are reported in **table 5.20**.

All 50  $\mu$ l of the PCR products were purified using the GenElute Gel Extraction Kit (Sigma) according to the manufacturer's protocol.

##### 2) Cloning into pCR@2.1 vector.

The PCR products were then cloned into pCR@2.1 vector using the Original TA cloning kit (Life Technologies). This vector contains the lacZ gene, encoding  $\beta$ -galactosidase, and it allows for blue-white screening. The following ligation reaction was used:

## MATERIALS AND METHODS

---

Component	Amount ( $\mu$ l)
10X Ligation Buffer	1
pCR®2.1 vector (25 ng/ $\mu$ l)	2
T4 DNA Ligase (4.0 Weiss units)	1
H <sub>2</sub> O	7
PCR product	10

The ligation reaction was incubated at 14°C overnight.

### 3) Transforming Competent Cells.

After the ligation of insert into pCR®2.1, the construct was transformed into competent DH5 $\alpha$  *E. coli* strains according to this protocol:

1. Thaw on ice one 50  $\mu$ l vial of frozen Competent Cells for each transformation.
2. Pipette 2  $\mu$ l of each ligation reaction directly into the vial of competent cells and mix
3. Incubate the vials on ice for 30 minutes.
4. Heat shock the cells for 30 seconds at 42°C without shaking.
5. Immediately transfer the vials to ice for 5 minutes.
6. Add 300  $\mu$ l of room temperature LB medium.
7. Shake the vials horizontally at 37°C for 1 hour
8. Spread the entire volume on LB agar plates containing ampicillin (10mg/ml).
9. Incubate plates overnight at 37°C.

White colonies were grown and the plasmid DNA was purified and sequenced with universal M13 forward and reverse primers (**Table 5.21**) using the BigDye v1.1 kit (Life Technologies). Sequences were run onto the ABI 3730 automated sequencing machine, and electropherograms were analyzed with Sequencher 5.0 (*Gene Code Corporation*).

**Table 5.1: CTNNA3 deletion in family 3456 (validation by PCR) (\* primers used for sequencing)**

Primer name	Sequence (5'-3')	Primer size (bp)	Product size (pb)	Fragment position (hg19)	PCR with TaqGold	
					[Mg <sup>2+</sup> ]	PCR program
CTNNA3_father_del_5F	TCTCCGAAGTCCCAAAGAA	20	301 bp	chr10:68227982-68327931	2,5 mM	95°C-60°C-72°C (30",30",1'30") for 35 cycles
CTNNA3_father_del_5Rseq*	GTCAAGCATCACTTCTTGGCTA	22				
CTNNA3_mother_del_1F	TCGCTTTCAGAGGTTGGAGT	20	949 bp	chr10:68272399-68333091	2,5 mM	95°C-60°C-72°C (30",30",6') for 35 cycles
CTNNA3_mother_del_1Rseq*	GGATTGTGCTTCTGGCATT	20				
CTNNA3_delx11_rt_F	ACCAGATACAGCAAGGAAGTCA	22	91 bp	chr10:68280373-68280463	2,5 mM	TD 65-58°C (30",30",30") for 30 cycles
CTNNA3_delx11_rt_R	GTACAAGCGTACATGGGAGAAT	22				

**Table 5.2: CTNNA3 deletions in four ASD families (validation by qPCR)**

Primer name	Sequence (5'-3')	Primer size (bp)	Product size (pb)	Fragment position (hg19)	Name in the figure 7.5
CTNNA3_del3093_1F	TCTGTGCCTTTGTCCTATCACT	22	107	chr10:68373163-68373269	Intron 10_1
CTNNA3_del3093_1R	GGTGGTCCGTAAGTATTTGTTG	22			
CTNNA3_del3093_2F	AGTGCCCTCACACATTTCCA	20	100	chr10:68372197-68372297	Intron 10_2
CTNNA3_del3093_2R	GGGCTACAACCTTACCTTTCCA	21			
CTNNA3_del3093_3F	TTTCATTGCTTCTGGCCTCT	20	105	chr10: 68352567-68352671	Intron 10_3
CTNNA3_del3093_3R	CAAAGTGTCTGCACAGGGTAA	21			
CTNNA3_del3311_1F	CTTTCAGTTGACTCCTTGTGCT	22	114	chr10:68065197-68065310	Intron 12
CTNNA3_del3311_1R	GACTCTGCATTTGGCTCATT	20			
CTNNA3_del3311_2F	CCTCCTCTTGTGTATCCCATTT	22	116	chr10:68031618-68031733	Intron13
CTNNA3_del3311_2R	TCACTAACTAGCGCGGCTTT	20			
CTNNA3_del3476_1F	GGCTTAGTTTATGGCTCTCCAA	22	111	chr10:68547946-68548056	Intron7_1
CTNNA3_del3476_1R	TTAATGACCACCACCCTCCT	20			
CTNNA3_del3476_2F	TTTGGTATGCTTCTGCTGCT	20	107	chr10:68537601-68537707	Intron7_2
CTNNA3_del3476_2R	ACTCCTGTTGTGTTGGCTGT	20			
CTNNA3_delx11_rt_F	ACCAGATACAGCAAGGAAGTCA	22	91	chr10:68280373-68280463	Exon 11
CTNNA3_delx11_rt_R	GTACAAGCGTACATGGGAGAAT	22			

**Table 5.3: CTNNA3 mutation screening** (\* primers used for sequencing)

Primer name	Sequence (5'-3')	Primer size (bp)	Product size (pb)	Fragment position (hg19)	PCR with TaqGold		
					[Mg <sup>2+</sup> ]	[DMSO]	PCR program
CTNNA3_x1F*	TGTTTGTAGGTCTCTAGGCCAAAG	23	394	chr10:69455688-69456081	2 mM		TD 67-60°C (1',30",1') for 30 cycles
CTNNA3_x1R	CACAGGAATTCAATGCAGAA	20					
CTNNA3_x1alt_F*	GCTCTGGGCTATGCAGAAAAG	20	476	chr10:69408315-69408791	2,5 mM		TD 67-60°C (1',30",1') for 30 cycles
CTNNA3_x1alt_R	CAGGGTGGCTACAATTCCTT	20					
CTNNA3_x1bF*	TACCCTTCACTGGCATGTTG	20	340	chr10:69425192-69425531	2 mM		TD 62-57°C (30",30",30") for 30 cycles
CTNNA3_x1bR	CCTTTCATTCCCCCAAAGT	20					
CTNNA3_x2F	TTTTGTAGCCTGGTCTCTGA	21	467	chr10:69406956-69407422	2 mM	5%	TD 67-60°C (1',30",1') for 30 cycles
CTNNA3_x2R*	AAAGAGAATGTGGGGGAAACT	21					
CTNNA3_x3F	TCCCAGGACTGTGTTCTCCT	20	467	chr10:69366850-69367317	2 mM,		TD 67-60°C (1',30",1') for 30 cycles
CTNNA3_x3R*	GGGGACCGAGCAACTAAAA	19					
CTNNA3_x4F*	TCAGCAAATGCAGAAAGTTGG	20	542	chr10:69299171-69299712	2 mM		TD 67-60°C (1',30",1') for 30 cycles
CTNNA3_x4R	GCCAAGATGCACTAGGATCG	20					
CTNNA3_x5F	GAGGCTCATTATCGACTCTTCG	22	350	chr10:69281462-69281811	2 mM		TD 67-60°C (1',30",1') for 30 cycles
CTNNA3_x5R*	GATCAGCACAGCAACCTG	20					
CTNNA3_x6F*	TGTAGAAATAATAGATGCCCTTGA	24	567	chr10:68979217-68979783	2,5 mM		TD 62-55°C (1',30",1') for 30 cycles
CTNNA3_x6R	TTTGGAAAGACTCAAGAAGGTGA	22					
CTNNA3_x7F*	AGCCGAGGCAGTAAAACACA	20	499	chr10:68939912-68940410	2,5 mM		TD 67-60°C (1',30",1') for 30 cycles
CTNNA3_x7R	TTCCCTGTAATAACCAAGCAGA	23					
CTNNA3_x8F*	GCTCTGCATTGTCACACTG	21	435	chr10:68535107-68535541	2 mM		TD 67-60°C (1',30",1') for 30 cycles
CTNNA3_x8R	TGAAAGGAACAAAACAAGAAACA	23					
CTNNA3_x9F	TTTCTTTTTGCCTGGATATTTTTG	23	513	chr10:68525851-68526363	3 mM;	5%	TD 67-60°C (1',30",1') for 30 cycles
CTNNA3_x9R*	TGGAGTGATCTATAACAATCGGACA	24					
CTNNA3_x10F*	CATTGAGATTCTTGATGAATGTGT	24	585	chr10:68381063-68381647	3 mM ;	5%	TD 67-60°C (1',30",1') for 30 cycles
CTNNA3_x10R	GGACAATTCTGTCTTCCATGC	21					
CTNNA3_x11F*	TGTTTTTCTATTGTGCGATGATG	23	582	chr10:68280297-68280878	2,5 mM	5%,	TD 67-60°C (1',30",1') for 30 cycles
CTNNA3_x11R	ACCATGCCTGTCCCAGTATT	20					
CTNNA3_x12F*	AGTCTTCCATTTCCAATGTGC	21	471	chr10:68138763- 68139233	2,5 mM		TD 67-60°C (1',30",1') for 30 cycles
CTNNA3_x12R	CGCATAACCAAAGTGTGCAA	20					
CTNNA3_x13F	CAGAAGCCTGGGTTTAGAA	20	467	chr10:68040148- 68040614	3 mM		TD 67-60°C (1',30",1') for 30 cycles
CTNNA3_x13R*	AGCATTCTTGACAGTCATCC	20					
CTNNA3_x14F	TAAGGTACTGCCGCGTTGTT	20	305	chr10:67862781-67863085	3 mM		TD 67-60°C (1',30",1') for 30 cycles
CTNNA3_x14R*	TTTCTCCATGTGCTTACAGAA	21					
CTNNA3_x15F*	TGGATTTAAGTATGTTGGACACC	23	546	chr10:67828960-67829505	2,5 mM		TD 67-60°C (1',30",1') for 30 cycles
CTNNA3_x15R	CCACCTGATTTTTGGCACTT	20					
CTNNA3_x16F	GATGGATGGACGCATAGATT	20	459	chr10:67748199-67748657	3 mM		TD 67-60°C (1',30",1') for 30 cycles
CTNNA3_x16R*	TTTGATGCTGAAGAGCTGATACT	23					
CTNNA3_x17F*	GAAGCTACTTGGTCCTGTATGAA	23	350	chr10:67726290-67726639	2,5 mM		TD 67-60°C (1',30",1') for 30 cycles
CTNNA3_x17R	TGGTCATGTAACAAGGGTGT	21					
CTNNA3_x18F	AACAAAACCTCCTGGAATAATAGC	24	479	chr10:67679946-67680424	2,5 mM		TD 67-60°C (1',30",1') for 30 cycles
CTNNA3_x18R*	TTGATTTAGCGCCCAATATTTTA	23					

**Table 5.4: *LRRTM3* mutation screening** (\* primers used for sequencing)

Primer name	Sequence (5'-3')	Primer size (bp)	Product size (pb)	Fragment position (hg19)	PCR with TaqGold	
					[Mg <sup>2+</sup> ]	PCR program
LRRTM3_x1F*	GTGCTCATCACGGGAAGCTG	19	587	chr10:68685905-68686491	2,5 mM	TD 65-60°C (30",30", 1') for 30 cycles
LRRTM3_x1R	GGAAACATGCCACTAAGCAAA	21				
LRRTM3_x2Fa	TTTGCTTAGTGGCATGTTTCC	21	845	chr10:68686471-68687315	2,5 mM	TD 65-60°C (30",30", 1') for 30 cycles
LRRTM3_x2Ra	TGCTCCAGGTGAAGTTCTTTG	21				
LRRTM3_x2Fb	GAACCATCCCTGTGCGAATA	20	645	chr10:68687183-68687827	2,5 mM	TD 65-60°C (30",30", 1') for 30 cycles
LRRTM3_x2Rb	GGAGCTTGGGCTTAAACGTC	20				
LRRTM3_x2Fc	ATATGGGAATGCAGCAGAAA	20	690	chr10:68687611-68688300	2,5 mM	TD 65-60°C (30",30", 1') for 30 cycles
LRRTM3_x2Rc	CATCGCGTTCCCTTGATAGT	20				
LRRTM3_x3F*	GCGGGATAATATAGGGACACC	21	473	chr10:68857201-68857673	2,5 mM	TD 65-60°C (30",30", 1') for 30 cycles
LRRTM3_x3R	TTTGAACAGGGGATTTTGTG	20				
LRRTM3_x2RaSeq*	TGGAACATAAGAATCAGCTCTTTGA	24				
LRRTM3_x2RbSeq*	TGAAAGCTTCGATCTCATTGC	21				
LRRTM3_x2FcSeq*	CCGAGCACATCTCTTTCCAT	20				

**Table 5.5: *CTNNA3* expression** (\* primers used for sequencing)

Primer name	Sequence (5'-3')	Primer size (bp)	Product size (pb)	Fragment position (hg19)	PCR conditions (Kapa HiFi)	
					[Mg <sup>2+</sup> ]	PCR program
CTNNA3_express_x10F	CCAATCATTGGAAACCTTGTG	22	731	chr10:68381457- 68857673	2,5 mM	TD 65-60°C (20",15", 45") for 35 cycles
CTNNA3_express_x15R	CTCAATCTCAGCATCCAGCTTA	22				
CTNNA3_express_x12F	GTTACGAGCCAGGGGCTTAC*	20				
CTNNA3_express_x14R	CAAGGTCAGAAACATCCTCCA*	21				

**Table 5.6: *CHRNA7* duplication (validation by qPCR)**

Primer name	Sequence (5'-3')	Primer size (bp)	Product size (pb)	Fragment position (hg19)
CHRNA7_x3F	GGCTGCAAATGGTAAGTTAAGAG	23	111	chr15:32393540-32393650
CHRNA7_x3R	AACAGGACCTCTCAGAAGCAAG	22		

**Table 5.7: *CHRNA7* mutation screening exons 1-4 (\* primers used for sequencing)**

Primer name	Sequence (5'-3')	Primer size (bp)	Product size (pb)	Fragment position (hg19)	PCR with TaqGold			
					[Mg <sup>2+</sup> ]	[Deaza GTP]	[DMSO]	PCR program
CHRNA7_x1F	AGTACCTCCCGCTCACACCT	20	480	chr15:32322529-32323009	1,5 mM	0.1mM	5%	TD 65-60°C (30",30", 30") for 30 cycles
CHRNA7_x1R*	GTGCAGCCCAGACAAGCA	18						
CHRNA7_x2F	CTCCGGGACTCAACATGC	18	578	chr15: 32322784-32323008	1,5 mM	0.1mM	5%	TD 65-60°C (30",30", 30") for 30 cycles
CHRNA7_x2R*	AGCTTGGGGCCAACTAGAG	19						
CHRNA7_x3F	CCACACACAACAACGCTCTC	20	252	chr15:32393397-32393646	2,5 mM			TD 65-60°C (30",30", 30") for 30 cycles
CHRNA7_x3R*	GGACCTCTCAGAAGCAAGCA	20						
CHRNA7_x4F	TGGAATTCTCTTTGGTTTTGC	21	388	chr15:32403893-32404280	2,5 mM			TD 65-60°C (30",30", 30") for 30 cycles
CHRNA7_x4R*	GCTGGCTTACAGGGACAGAG	20						
CHRNA7_intron5F	GACTTGACCATAACATGACTTTCC	24	252	chr15:32448127-32448507	2,5 mM			TD 65-60°C (30",30", 30") for 30 cycles
CHRNA7_intron5R*	CAGACAATCCCCCTCCATTA	20						
CHRNA7_intron9F	AGTGCCGCCTGCTTGTA	18	298	chr15:32458604-32458902	2,5 mM			TD 65-60°C (30",30", 30") for 30 cycles
CHRNA7_intron9R*	TTCCCTGAAATTATCCAGATCCT	23						

**Table 5.8: *CHRNA7* long range PCR**

	Primer name	Sequence (5'-3')	Primer size (bp)	Product size (pb)	Fragment position (hg19)	PCR with KAPA HiFi	
						[DMSO]	PCR program
x5-x8 LR-PCR	CHRNA7_intron4F	CACCTGCAGTTCAGTCATTCAA	22	6752	chr15:32445306-32452057	5%	TD 65-60°C (20",15",7) for 35 cycles
	CHRNA7_intron8R	AAAGTCAAACCTCAAAGCTGAA	22				
x9-x10 LR-PCR	CHRNA7_intron8F	AGTGCATGGAAGTGCAATGA	20	7475	chr15:32455200-32462675		TD 65-60°C (20",15",7) for 35 cycles
	CHRNA7_3'-UTR	CACTTCTACTTGTCTAAAGACACTG	27				

**Table 5.9: *CHRNA7* mutation screening exon 10**

Primer name	Sequence (5'-3')	Primer size (bp)	Product size (pb)	Fragment position (hg19)	PCR with TaqGold	
					[Mg <sup>2+</sup> ]	PCR program
CHRNA7_x10F	AATGAAGCAGGGCTTGTATTG	21	728	chr15:32460047-32460774	1,5 mM	TD 65-60°C (30",30", 30") for 30 cycles
CHRNA7_x10R	AGGGAACACTGGAGTTGTGG	20				

**Table 5.10: *CHRNA7* sequencing** (Primer used for sequencing specific exons from the x5-x8 LR-PCR and the x9-x10 LR-PCR products)

Primer name	Sequence (5'-3')	Primer size (bp)	Primer position (hg19)
CHRNA7_x5F	CAAGGTCTTTGCTGCTCCAT	20	chr15:32446012-32446031
CHRNA7_x5R	TCTGGGGGTAGAAAGCACAC	20	chr15:32446341-32446360
CHRNA7_x6F	GGCTGAAGGAACTGCTGTGT	20	chr15:32449730-32449749
CHRNA7_x6R	TGTTTCTCCTCCTGTAGCTCTCA	23	chr15:32450107-32450129
CHRNA7_x7F	ATGGCTCCTTCTCTCCTCCT	20	chr15:32450583-32450602
CHRNA7_x7R	AATCCCCAGGAACCCTGAT	19	chr15: 32446333-32446360
CHRNA7_x8F	GGAGCCCTCGTTAGACAGAA	20	chr15:32451695-32451714
CHRNA7_x8R	TCTAGTTTCATCTGCTGGGAAAT	23	chr15:32451918-32451940
CHRNA7_x9F	GGTGTGCCTGTCCTGTGAC	19	chr15: 32455356-32455376
CHRNA7_x9R	AATTGGCCAGGTGTGGTG	18	chr15: 32455802-32455819
CHRNA7_x10F	AATGAAGCAGGGCTTGATTG	21	chr15:32460047-32460067
CHRNA7_x10R	AGGGAACACTGGAGTTGTGG	20	chr15:32460755-32460774

**Table 5.11: *KLHL23* deletion in the discovery pedigree (validation by qPCR)**

Primer name	Sequence (5'-3')	Primer size	Product size (pb)	Fragment position (hg19)	Name in the figure 8.14
KLHL23del_x3F	ACGTAGGTGTGGGAAATGCTAC	22	117	chr2:170,597,889-170,598,005	1
KLHL23del_x3R	CGGAATTGTAGCTCTGAACTTTG	23			
KLHL23del_intron3F	AGGACAAGCTGAAGGATAATGG	22	104	chr2:170,605,701-170,605,804	2
KLHL23del_intron3R	ATTCAGGCACACAATACTGCAA	22			
KLHL23del_x4F	AAGGGAACGTATCTTCAGAGCA	22	108	chr2:170606123-170606230	3
KLHL23del_x4R	CACACAAACACACCCATGAGAC	22			
SSBdel_intron9F	AAAGACATGGAAGGTTTGCAG	21	92	chr2:170,667,262-170,667,353	4
SSBdel_intron9R	GAAAACAAAAGCCATACCCTTG	22			
SSBdel_intron11F	CGATTTACCTCGGTTATGCTGT	22	113	chr2:170,667,962-170,668,074	5
SSBdel_intron11R	ACTTCCTTGACCAATGAGAAA	22			
METTL5del_intron7F	GCTATTGACGATGGGCCTTT	20	91	chr2:170,668,636-170,668,726	6
METTL5del_intron7R	TTCTAGTGTTACCAACTTTCCTGTG	25			
METTL5del_intron6F	AAGTGTTAAGTGCTGGTAAGCTGAA	25	112	chr2:170,669,106-170,669,217	7
METTL5del_intron6R	GGCCATAATAACAGAAATGCCAAG	24			
METTL5del_exon6F	TCAGGCTAAGTGTTCACTCTGC	22	105	chr2:170,671,819-170,671,923	8
METTL5del_exon6R	CATGCAGGTGCTGAGAAGATAA	22			

**Table 5.12: *KLHL23* expression in family 3423 by qPCR**

Primers name	Sequence (5'-3')	Primer size (bp)	Product size (pb)	Fragment position (hg19)
KLHL23_expr_x2F	GCAGAAGAGGCTGAGTTCTATGA	23	123	chr2:170,592,671-170,597,950
KLHL23_expr_x3R	GCCACCAATGACGTAGATAACA	22		
KLHL23_expr_x3F	AACGAATGGAGCCTCATCAC	20	125	chr2:170,598,011- 170,606,019
KLHL23_expr_x4R	TG TTCAGGGTCATAGCATTCTG	22		

**Table 5.13: *KLHL23* and *PHOSPHO2-KLHL23* expression in a multiple tissue cDNA panel**

	Primers name	Sequence (5'-3')	Primer size (bp)	Product size (pb)	Fragment position (hg19)	PCR with TaqGold	
						[Mg <sup>2+</sup> ]	PCR program
PCR fragment specific for <i>KLHL23</i>	KLHL23_expr_x1F	CCTCTTCCAAAGATGGTCAGA	21	156	chr2:170,590,503-170,591,583	2,5 mM	TD 63-58°C (30",30", 30") for 30 cycles
	PHOSPHO2_expr_x4R	ACTGGATGTGTTGAATCCTTGA	22				
PCR fragment specific for <i>PHOSPHO2-KLHL23</i>	PHOSPHO2_expr_x3F	CCAGAAGACTCTGTTCCCTGTA	22	200	chr2:170,553,903-170,591,583	2,5 mM	TD 63-58°C (30",30", 30") for 30 cycles
	PHOSPHO2_expr_x4R	ACTGGATGTGTTGAATCCTTGA	22				

**Table 5.14: *KLHL23* mutation screening** (\* primers used for sequencing; <sup>a</sup> primers used also for mutation screening of *PHOSPHO2-KLHL23* exons 4,5,6; <sup>b</sup> primers used for exon 2 PCR-RFLP)

Fragment name	Sequence (5'-3')	Primer size (bp)	Product size (pb)	Fragment position (hg19)	PCR with TaqGold	
					[Mg <sup>2+</sup> ]	PCR program
KLHL23_x1F*	GTCTCCTGGGGAAGCAGTC	19	474	chr2:170590262-170590735	2,5 mM	TD 65-60°C (30",30", 45") for 30 cycles
KLHL23_x1R	CATTGTCTGCGCTCCTCTC	19				
KLHL23_x2aF* <sup>a,b</sup>	TCCGATGATAGTCAAGTTATTTAGC	25	565	chr2:170591436-170592000	2,5 mM	TD 65-60°C (30",30", 45") for 30 cycles
KLHL23_x2aR <sup>a,b</sup>	CGAGATTCCTTCTCTAGTTCTGG	23				
KLHL23_x2bF* <sup>a</sup>	TTCAGTAAAGAAGGCTTGTGAGC	23	552	chr2:170591884-170592435	2,5 mM	TD 65-60°C (30",30", 45") for 30 cycles
KLHL23_x2bR <sup>a</sup>	TCTGCTCCCTGAATCCAAAC	20				
KLHL23_x2cF* <sup>a</sup>	GCCTGCTCACCGAAAATAAG	20	580	chr2:170592258-170592837	2,5 mM	TD 65-60°C (30",30", 45") for 30 cycles
KLHL23_x2cR <sup>a</sup>	TCCTAGATAGCATCCCAAAGTGA	23				
KLHL23_x3F* <sup>a</sup>	CAGTACCTTGCAATTTACCATCA	23	465	chr2:170597779-170598243	2,5 mM	TD 65-60°C (30",30", 45") for 30 cycles
KLHL23_x3R <sup>a</sup>	TGTAGATGTGACCCAATCAGAA	22				
KLHL23_x4F* <sup>a</sup>	AGCCTCAGCGTTGCAGTATT	20	584	chr2:170605773-170606356	2,5 mM	TD 65-60°C (30",30", 45") for 30 cycles
KLHL23_x4R <sup>a,b</sup>	TCCCCTATCATGTGCCAGAC	20				

**Table 5.15: PHOSPHO2-KLHL23 mutation screening (\* primers used for sequencing)**

Fragment name	Sequence (5'-3')	Primer size (bp)	Product size (pb)	Fragment position (hg19)	PCR with TaqGold	
					[Mg <sup>2+</sup> ]	PCR program
PHOSPHO2-KLHL23_x1F*	CGTACCAGCATCTCTGACGA	20	423	chr2:170550809-170551232	2,5 mM	TD 65-60°C (30",30", 45") for 30 cycles
PHOSPHO2-KLHL23_x1R	GTAGGCATCGGACCCAATC	19				
PHOSPHO2-KLHL23_x2F*	AACAAGTTGGCTCTGACGTG	20	260	chr2:170550809-170551866	2,5 mM	TD 65-60°C (30",30", 45") for 30 cycles
PHOSPHO2-KLHL23_x2R	CCCCAGTTCTTGACACTTT	20				
PHOSPHO2-KLHL23_x3F*	CCATTTCTGACCTCGTGAT	20	394	chr2:170553745-170554139	2,5 mM	TD 65-60°C (30",30", 45") for 30 cycles
PHOSPHO2-KLHL23_x3R	GTTCCCCTGCTGATCTGAAA	20				

**Table 5.16 Control genes (Real time PCR)**

Fragment name	Sequence (5'-3')	Primer size (bp)	Product size (pb)	Fragment position (hg19)
FOXP2_F	TGCTAGAGGAGTGGGACAAGTA	22	139	chr7:114121559-114121698
FOXP2_R	GAAGCAGGACTCTAAGTGCAGA	22		
GUSB_8F	CACCTAGAATCTGCTGGCTACT	22	93	chr7:65435290-65439310
GUSB_9R	AGAGTTGCTCACAAAGGTCACA	22		

**Table 5.17: CADPS2 expression analysis SNP rs2251761**

Fragment name	Sequence (5'-3')	Primer size (bp)	Product size (pb)	Fragment position (hg19)
CADPS2_express_x3F	GACTTGTGCAAACAGCCAAA	20	393	chr7: 122261603-122303431
CADPS2_express_x5R	AAAATGCAGAGTTCTGTGAACG	22		

**Table 5.18: primer sequences for CADPS2 MS analysis and CpG content details for each amplicons.**

Amplicon position from CDS	Target length (bp)	Primer Forward	Primer reverse	Strand	CpGs contained	CpGs analyzable
-1421; -927 (a) <sup>a</sup>	495	GAAAGTGGTTTGAAAAAGTTAAAATTG	ACAAAATAATACTAATTCAACTCAAACA	Forward	12	10
-911; -486 (b)	426	TTTGTTTTGAGTTTGTAGGATTAGAA	CCTAAAACCCTAACACACAATTTACA	Forward	22	13
-515; -112 (c)	404	GAAAGGAAAATTGGTTAGGGT	CCTTACAAACTATATACTAAAACCTCCAAA	Reverse	41	31
+550; +973 (d)	424	GGGTTGGTAGGAGTTGTTAGTTTATT	CAAACCTACCCAAACCAAAAATTAC	Forward	17	9

<sup>a</sup> (a) to (d) refers to positions indicated in Figure 9

**Figure 5.19: *CADPS2* intron 1 (primers used for PCR amplification on genomic DNA) (\* primer used for sequencing)**

Fragment name	Sequence (5'-3')	Primer size (bp)	Product size (pb)	Fragment position (hg19)	PCR with TaqGold	
					[Mg <sup>2+</sup> ]	PCR program
rs981321_F*	GGCAGGCCTGATTAATGAAA	20	186	chr7: 122525243-122525428	2,5 mM	TD 63-58°C (30",30", 30") for 30 cycles
rs981321_R	CCCTGGGGTCTTGTTTTGA	19				

**Figure 5.20: *CADPS2* intron 1 (primers used for PCR amplification on bisulfite treated DNA) (\* primers used for sequencing)**

Fragment name	Sequence (5'-3')	Primer size (bp)	Product size (pb)	Fragment position (hg19)	PCR with KAPA2G	
					PCR program	
CADPS2_int1bisF+*	GGGTTTGAATGAGTGTGTAG	21	366	chr7: 122525285-122525650	95°C 1'; 95°C 15"; 56°C 15"; 72°C 20" (for 47 cycles)	
CADPS2_int1bisG-*	AAAAGGTAAGAAGTTTGTTTTAGG	24				

**Figure 5.21: *CADPS2* colony screening**

Fragment name	Sequence (5'-3')	Primer size (bp)
M13_F	GTAAAACGACGGCCAGT	17
M13_R	CAGGAAACAGCTATGAC	17

## CHAPTER 6:

### AIM OF THE STUDY AND PRELIMINARY RESULTS

#### **6.1 Aim of the study**

Intellectual disability (ID) and autism spectrum disorders (ASDs) are complex neuropsychiatric conditions with a clinical and genetic heterogeneity.

Despite high heritability estimates and the identification of many monogenic and chromosomal causes of both these two disorders, their genetic basis remains poorly explained. Genome-wide association studies (GWAS) have so far provided only tenuous evidence for individual common variants that affect risk of complex neurological traits, shifting the attention to rare variation. Indeed several recent studies have revealed that a number of different causes (SNPs or CNVs), each of them with low frequency in the population and typically highly penetrant, could collectively account for a large proportion of attributable (McClellan & King, 2010).

Considerable progress in detection of risk genes underlying ASD and ID has been made after the advent of array-based approaches (array-CGH and SNP arrays), which have allowed the discovery of a large number of pathogenic Copy Number Variants, both *de novo* and inherited, and the genomic analysis of large cohorts of affected subjects and controls. One of the most interesting findings emerging from these CNV screenings is that ASD cases carry a higher number of rare genic CNVs in comparison with healthy controls, and that the affected genes converge on common pathways including cell-adhesion and synaptic function (Pinto et al., 2010). Furthermore, these findings have indicated that, besides the high heterogeneity, ID and ASD share overlapping risk factors (Betancur, 2011) and they can co-occur in many patients.

My PhD project has started from the discovery of CNVs impacting ASD/ID candidate genes, which were identified in two genome wide CNVs studies (SNP array and CGH-array) performed by two International Projects, the Autism Genome Project (AGP) and the CHERISH.

The AGP is an ongoing international project which gathers more than 50 research groups from different countries in North America and Europe, and is aimed at searching autism-susceptibility genes.

The CHERISH includes clinicians and research groups from Central Asia and Eastern Europe in order to collect a large sample of ID patients and to understand the molecular causes underlying this complex disease.

Specifically, I analyzed three rare genic CNVs detected by the AGP study in Italian ASD individuals (see Table 6.1) (Pinto et al., 2014) and an intragenic deletion in *CADPS2* in individuals with either ASD/ID identified by the CHERISH project (see paragraph 6.2.2)

The aim of this study was to validate and characterize these rare CNVs, in order to understand their contribution and their role in ASD/ID susceptibility.

## **6.2 Preliminary results**

### **6.2.1 Analysis of three CNVs identified in ASD Italian families**

In order to investigate the role of *de novo* and inherited CNVs in the genetic risk for ASD, the International Consortium Autism Genome Project (AGP) has recently performed the largest genome-wide scan for CNVs reported to date on over 2,845 ASD families and 4,768 control subjects (Pinto et al., 2014).

The samples were genotyped on the Illumina Human 1M-single and Illumina Human 1M-Duo BeadChip arrays and three CNV prediction algorithms, namely, QuantiSNP (Colella et al., 2007), PennCNV (Wang et al., 2007) and iPattern (Pinto et al., 2011), were used to obtain high-confidence CNV call and to minimize the number of potential false discoveries. Specifically, CNVs were identified by using QuantiSNP and iPattern, while PennCNV was used to confirm inheritance status of the resulting CNV calls in trios families (parents and affected proband). Array data of each family were exploited by manual inspection of Log R ratio and B allele frequency in Illumina's GenomeStudio software, as described in Materials and Methods.

80 out of 2,845 ASD families included in the AGP study belong to the Italian ASD cohort.

Among the CNVs identified in the Italian cohort, three interesting rare CNVs impact promising ASD candidate genes (**Table 6.1**).

**Table 6.1** CNVs identified in the Italian ASD Cohort in the AGP study

AGP ID	CNV	Coordinates (hg 19)	Length (bp)	Ref Seq genes	Inheritance
<b>3456_3</b>	Homozygous deletion	chr10:68228166-68332668	104,503	<i>CTNNA3</i>	Maternal/paternal
<b>3476_3</b>	Heterozygous duplication	chr15: 32,005,348-32,515,973	510, 626	<i>CHRNA7, OTUD7A</i>	Paternal
<b>3423_3</b>	Heterozygous deletion	chr2:170,603,578-170,666,813	63,236	<i>PHOSPHO2-KLHL23, KLHL23, SSB</i>	Maternal

The first CNV is a compound heterozygous deletion involving the *CTNNA3* gene, encoding  $\alpha$ T-catenin (see paragraph 3.4.1).  $\alpha$ T-catenin is a member of the  $\alpha$ -catenin family and it has a crucial

role in cell adhesion, one of the major pathways implicated in ASD (Bucan et al., 2009). The deletion, identified in a male proband (3456\_3) with ASD and borderline cognitive impairment, is inherited from both parents, each heterozygote for a deletion of slightly different length, and causes the homozygous loss of a coding exon (exon11) in the affected individual. His younger sister has normal cognition, with normal social and communication skills, while both parents do not show any evident cognitive or behavioural impairment.

The second CNV is a small duplication on chromosome 15q13.3, spanning a region of approximately 500 kb, including exon 1 of the longer isoform of *OTUD7A* and the entire *CHRNA7* gene. The duplication was detected in a male proband with ASD and epilepsy (3476\_3). His father, who transmitted the microdeletion, as well as his mother are both apparently normal. However, the paternal grandfather's sister was affected by ID, while the paternal grandfather's brother had speech delay. The *CHRNA7* gene, encoding for the alpha7 subunit of the neuronal nicotinic acetylcholine receptor (see paragraph 3.4.3), is considered the likely culprit gene in mediating neurological phenotypes in cases with 15q13.3 recurrent microdeletions. Even if the reciprocal microduplication has a less certain clinical significance in comparison with the deletions (see paragraph 1.5), *de novo* and inherited duplications involving this genomic region are associated with a wide spectrum of neuropsychiatric disorders, including ASD.

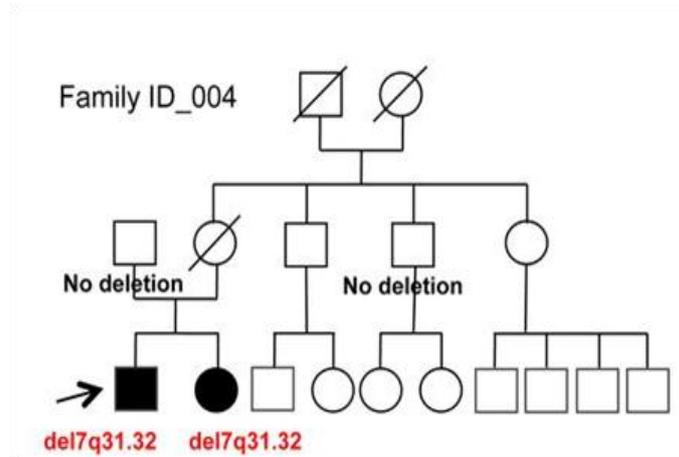
The third CNV is a rare maternal inherited microdeletion of about 63 kb encompassing three neighbouring genes on the chromosome 2q31.1, *KLHL23*, *PHOSPHO2-KLHL23* and *SSB*. This deletion, mapping in a region previously found to be in linkage with ASD, the AUTS5 locus ((IMGSAC), 2001; Maestrini et al., 2010), was identified in a male subject with ASD and borderline cognitive impairments (3423\_3).

### **6.2.2 Analysis of the *CADPS2* gene in subjects with ID and/or ASD**

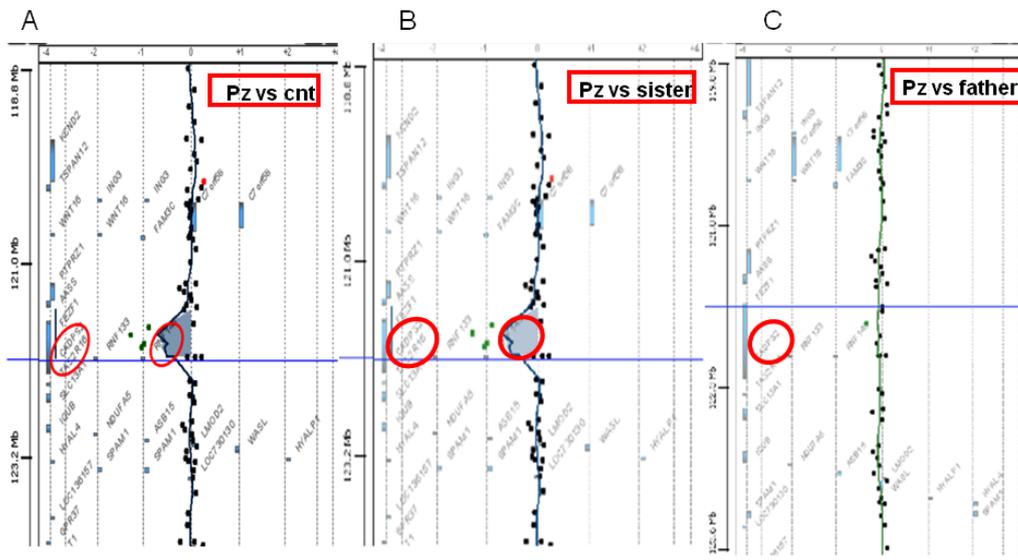
#### *Identification and validation of a rare deletion in the *CADPS2* gene*

During an array-CGH analysis with the Agilent platform 44 K, performed on ID patients recruited in the framework of the CHERISH project, an intragenic deletion of ~285 kb in the *CADPS2* gene was identified in a pair of siblings (male and female) with behavioral problems, borderline ID, and epilepsy (**Figures 6.1 and 6.2**). The deletion is likely to be inherited from the deceased mother, since the father did not carry it.

*CADPS2* is an excellent candidate gene for neurologic development abnormalities (see paragraph 3.4.2) and maps to the “autism susceptibility locus 1” on chromosome 7q31-q33 (Lamb et al., 2005).

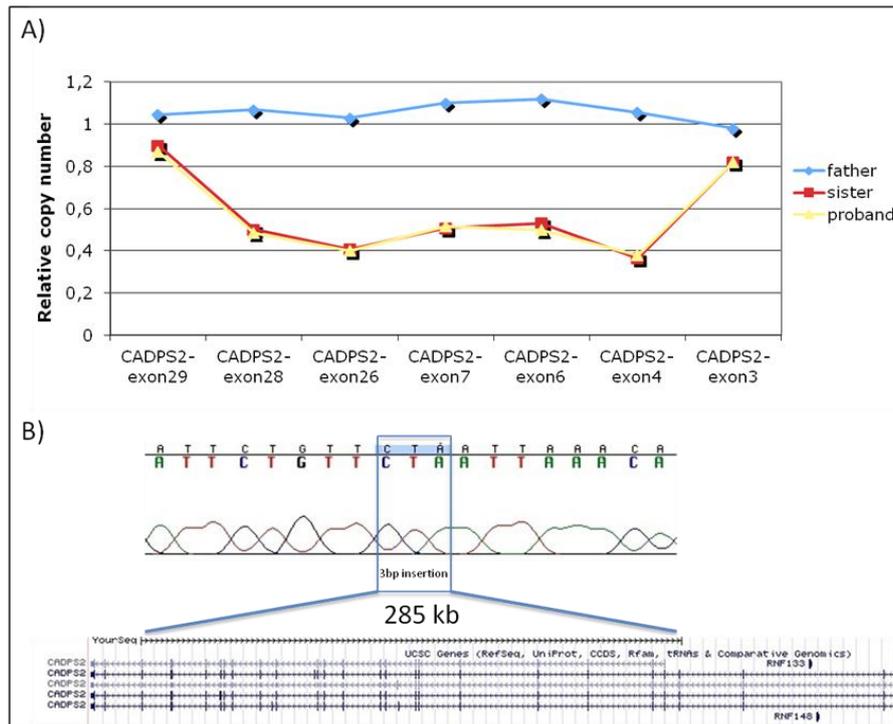


**Figure 6.1** (Bonora et al., 2014). Family tree of the ID sibs carrying the novel intragenic deletion.



**Figure 6.2 Results of CGH array in family ID\_004.** A) Identification of the *CADPS2* deletion in a subject with ID and autism. B) When the analysis was performed on patient vs sister CGH cannot detect it because it is present in both affected siblings; this suggest that also the affected sister has the same alteration. C) When the patient vs father were compared the deletion is seen only in the patient suggesting that the father does not carry the deletion.

The deletion was validated in father and in the two affected siblings by quantitative PCR and the deletion boundaries were mapped between intron 3 and intron 28 of *CADPS2*. Moreover, the deletion breakpoints were defined by long-range PCR between bp 121,984,852-122,270,267 of chromosome 7 (hg19) (**Figure 6.3**).



**Figure 6.3** (Bonora et al., 2014) A) Fine mapping of *CADPS2* deletion by real-time qPCR using different probes across the region. All data were normalized using as reference gene *FOXP2*. B) *CADPS2* deletion breakpoint mapping, showing the insertion of 3 base pairs at the breakpoint (upper panel) and the corresponding location on chromosome 7q (UCSC Genome Browser).

### *CADPS2* mutation screening in ASD/ID patients

To test whether rare variants in *CADPS2* could contribute to ASD and/or ID risk, all 30 exons and exon-intron boundaries of the *CADPS2* gene (NM\_017954.10) were screened in 36 Italian patients with ID and in 187 probands with ASD, of which 94 from Italy and 93 from the International Molecular Genetic Study of Autism Consortium (IMGSAC) collection (IMGSAC, 2001). Two synonymous and five missense heterozygous rare variants were identified (**Table 6.2**). The most interesting variant is a novel missense variant p.Asp113Ans, never observed in databases and in Italian controls. This variant is predicted to be damaging by PolyPhen-2 (PolyPhen score=0.936), and it has a functional effect since it disrupts *CADPS2* binding to the dopamine receptor type 2, one of its few known interactors (Bonora et al., 2014).

## AIM OF THE STUDY AND PRELIMINARY RESULTS

**Table 6.2** (Bonora et al., 2014). **Coding variants in *CADPS2* identified either in ASD or in ID patients.** The p.Asp1113Asn missense variant is indicated in red.

Position on chr7 (hg19)	Type of change (N_060424.9) NM_017954.10	PolyPhen-2 score (HumDiv)	SIFT prediction (cutoff=0.05)	Parental origin	Het <sup>a</sup> in ASD/ID (N=223)	Het <sup>a</sup> in Italian controls (N=250)	Het <sup>a</sup> in EVS (European-American)	P value <sup>b</sup> (Fisher's <i>t</i> -test)
g.122,5 <sup>36</sup> ,314G>A none <sup>c</sup>	<b>p.Ala26=</b>	na	na	Maternal	<b>1/223 (ASD)</b>	<b>0</b>	<b>0</b>	<b>0,0467</b>
g.122,255,252G>C none <sup>c</sup>	<b>p.Ala402=</b>	na	na	Paternal <sup>d</sup>	<b>1/223 (ASD)</b>	<b>0</b>	<b>0</b>	<b>0,0467</b>
<b>g.122,114,544A&gt;G rs199713510<sup>c</sup></b>	p.Met630Thr	0,917 (possibly damaging)	0 (damaging)	Paternal <sup>d</sup>	1/223 (ASD)	2/250	10/4113	0,477
<b>g.122,114,500A&gt;C rs201536376<sup>c</sup></b>	p.Phe645Val	0,001 (benign)	0,531 (tolerated)	Maternal	1/223 (ASD)	0	9/4128	0,392
<b>g.122,027,130C&gt;T rs76528953<sup>c</sup></b>	p.Asp1088Asn	1 (probably damaging)	0,001 (damaging)	Paternal	2/223 <sup>e</sup> (ASD)	0	28/4145	0,429
<b>g.122,019,472C&gt;T none<sup>c</sup></b>	<b>p.Asp1113Asn</b>	<b>0,936 (probably damaging)</b>	<b>0,039 (tolerated)</b>	<b>Maternal</b>	<b>1/223 (ASD)</b>	<b>0</b>	<b>0</b>	<b>0,0467</b>
<b>g.122,001,046C&gt;T rs200984050<sup>c</sup></b>	p.val1137Met	0,997 (probably damaging)	0,003 (damaging)	Maternal	1/223 (ID)	0	6/4122	0,294

<sup>a</sup> Het = number of heterozygous individuals.

<sup>b</sup> Fisher's exact test calculated considering the number of heterozygous individuals in the total control group [Italian and European-American from EVS (<http://evs.gs.washington.edu/EVS/> accession March 2013)].

<sup>c</sup> dbSNP entry reported for the corresponding variant; none = not present in dbSNP.

<sup>d</sup> The two paternal changes were present in the same ASD individual.

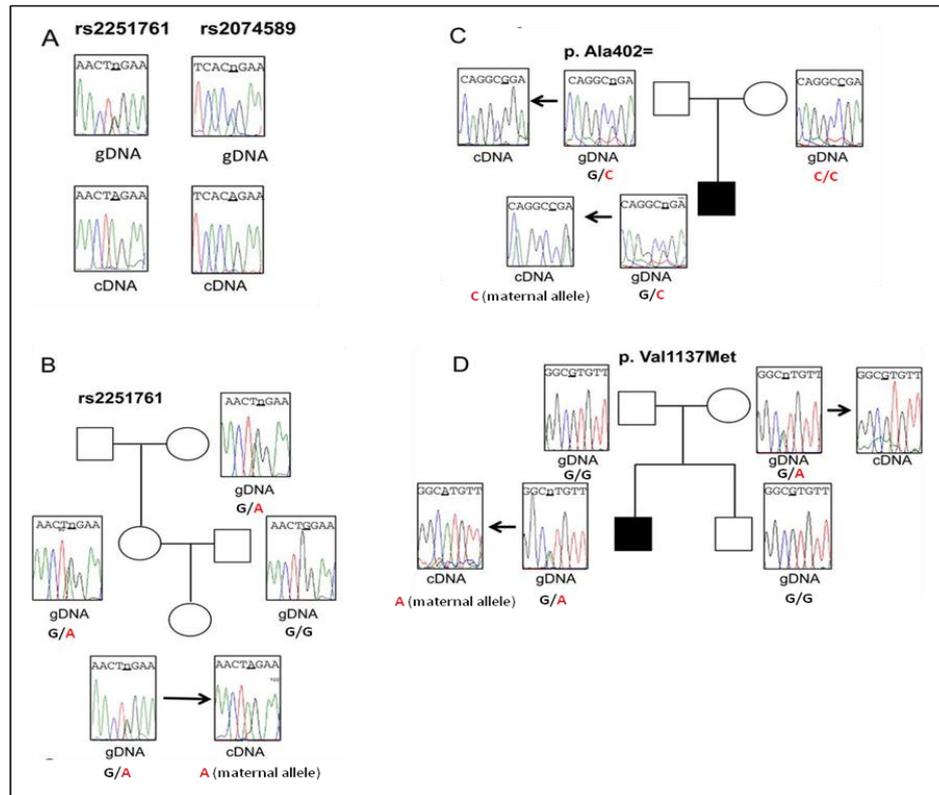
<sup>e</sup> The two individuals from multiplex ASD families inherited the change from the father; however in both cases, it did not co-segregate with the phenotype  
SNVs not found in dbSNP and EVS are shown in bold.

### *CADPS2 allelic expression analysis in blood*

*CADPS2* maps on a region of chromosome 7 containing a cluster of imprinted genes (Schneider et al., 2012); moreover, the *CADPS2* deletion and the majority of rare missense variants were of maternal origin. Thus *CADPS2* was tested for a possible parent-of-origin regulation.

In order to test this hypothesis the gene expression was analyzed in blood of seven informative heterozygous controls for two *CADPS2* coding SNPs: rs2251761 (exon 3) and rs2074589 (exon 17) (**figure 6.4A**). The results showed that only one allele is expressed in blood cDNA, suggesting that *CADPS2* shows a monoallelic expression in blood.

Moreover, in order to establish if the expressed allele was of maternal origin, three different families with heterozygous individuals for three SNPs (rs2251761; p.Ala402=; p.Val1137Met) were analyzed (**Figure 6.4B,C**). Since in each family the expressed allele in blood cDNA was the maternal one, these findings suggest that *CADPS2* is maternally expressed in blood.



**Figure 6.4** (Bonora et al., 2014). *CADPS2* allelic expression in blood. **A**) Allelic expression in blood cDNA of SNP rs2251761 (A/G alleles) and rs2074589 (A/C alleles) in two heterozygous individuals for the two SNPs. Upper panel: sequence electropherograms from genomic DNA (gDNA) showing the heterozygous state, lower panel: sequence electropherograms from blood cDNA; the SNP position is shown by the underscore marking. **B**) Allelic expression from blood cDNA of SNP rs2251761 in an informative control family: the expressed A allele in the offspring inherits it from the mother. **C**) Allelic expression of the variant p.402Ala=: the change is inherited from the heterozygous father, who expresses only the variant allele in blood, whereas the affected child expresses only the maternal allele in blood. **D**) Allelic expression of the variant p.Val1137Met: the maternally inherited variant is the only one expressed in blood cDNA of the affected son, whereas the healthy mother expresses the wild-type allele. The healthy brother does not inherit the change.

**CHAPTER 7:****RESULTS****Characterization of three CNVs identified in ASD Italian families**

Given the discovery in our Italian cohort of three interesting CNVs involving promising ASD candidate genes (see preliminary results, paragraph 6.2.1), we investigated their pathogenic role in ASD susceptibility performing expression analyses and assessing the contribution of rare sequence variants in these genes by mutation screening.

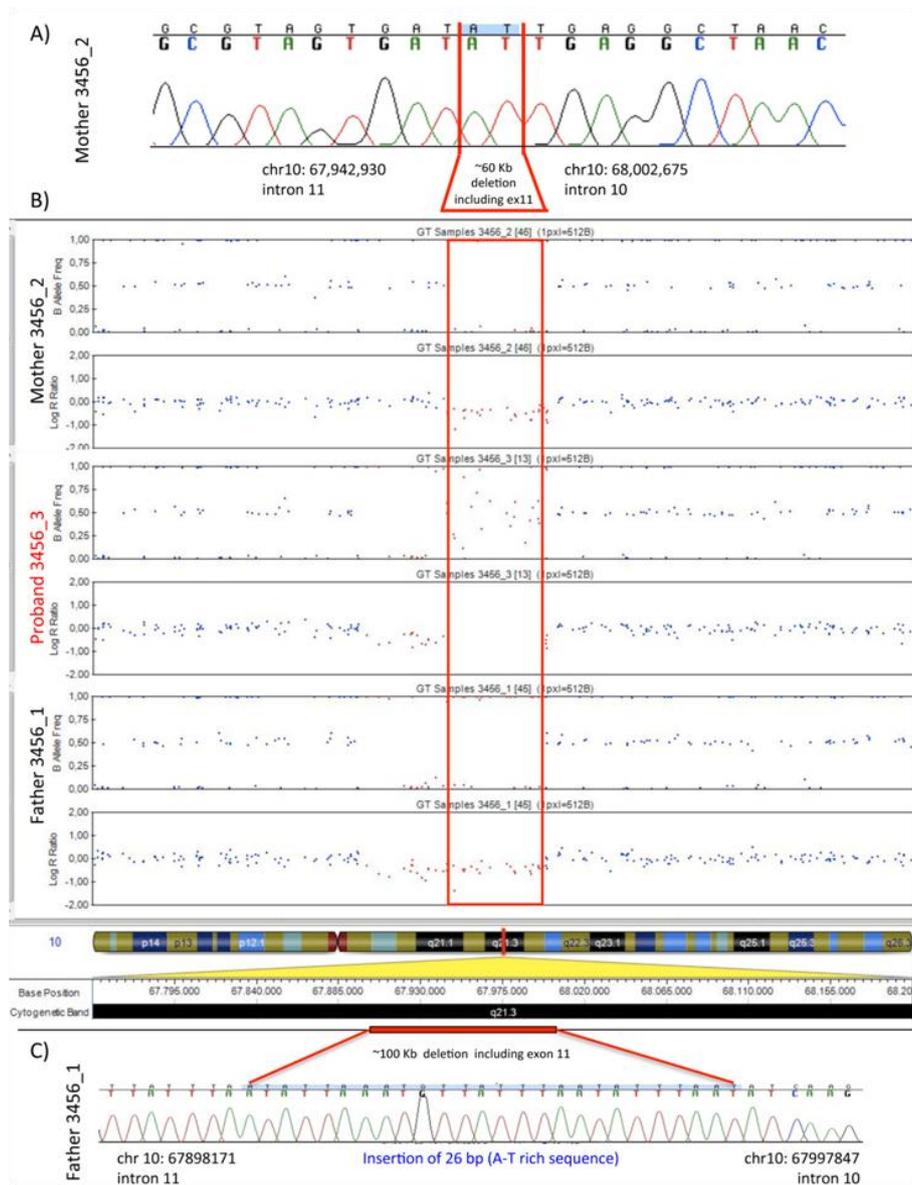
**7.1 Analysis of the *CTNNA3* gene****7.1.1 Fine-mapping and segregation analysis of *CTNNA3* deletion in family 3456**

Manual inspection of the log R ratios and B allele frequencies of family 3456 in GenomeStudio demonstrated that one copy of the microdeletion was transmitted to the proband (3456\_3) from each parent, who were both heterozygotes. Indeed the mother (3456\_2) and the father (3456\_1) show a decrease in the log R ratio and values of B-allele frequency around 0 and 1 (there is a lack of heterozygous variants), while in the proband the region is deleted on both alleles (**Figure 7.1B**).

Using several PCR assays with primers designed to flank the predicted breakpoints, we defined and sequenced the breakpoints of the two slightly different microdeletions carried by the parents of family 3456: the maternal microdeletion is of about 60 Kb and it encompasses chr10:67,942,931-68,002,674 (**Figure 7.1A**), while the paternal one has a length of about 100 Kb and it spans chr10:67,898,172-67,997,846 (National Center for Biotechnology Information build 36 coordinates); both the deletions remove the exon 11 of the *CTNNA3* (**Figure 7.1C**).

Sequencing of the breakpoints showed the presence of 2 and 5 bp microhomology at the junctions of the maternal and paternal deletions, respectively, as well as the insertion of a 26-bp A-T rich sequence at the paternal deletion junction (**figures 7.1A,C and 7.2**). These observations, together with the non-recurrent nature of the deletions and the absence of flanking low copy repeats (LCR), suggest that these deletions are likely to be generated through a microhomology-mediated repair mechanism.

## RESULTS



**Figure 7.1** (Bacchelli et al., 2014). **Detection and breakpoint definition of *CTNNA3* deletions in the discovery pedigree.** (A, C) Sequence electropherograms with breakpoint-spanning sequences in the mother and in the father, respectively (NCBI build 36 coordinates). In (A) the two base pairs (AT), common to both ends, are highlighted in blue. (B) GenomeStudio plots of the log *R* ratio and B allele frequency data from the 1 M-Duo SNP array for family 3456. The deletions result in a decrease in the log *R* ratio and a lack of heterozygous variants (the expected allelic ratio for heterozygous variants is 0.5). The deleted SNPs are depicted in red, and the red rectangle indicates the region deleted on both alleles in the proband.

## RESULTS

```
Alignment Maternal Deletion
Prox CCTATTTATTTTCTCACTGCTTTTCCTTCTGGAATTTCTAAAACAGGAAT
Del  CCTATTTATTTTCTCACTGCTTTTCCTTCTGGAATTTCTAAAACAGGAAT
Dist CAATGCTTTACCATCATCTGCCAAGACGTGTGTGCAAATATCAGTCTTG

Prox AGACTATTTGTTTGCATAGTGATATCCCATATGTCATGTAACCTTTTTTC
Del  AGACTATTTGTTTGCATAGTGATATGAGGCTAACCTGCCATCTGCTTAT
Dist CAAACAGACAATATTTGCACTACATGAGGCTAACCTGCCATCTGCTTAT

Prox ATTTTATCTTTTTCTCTTTTTGCCTAATTCGCTTATTTCAAAGATC
Del  CAGATGTTCCGGAGTCAAATTGTACAGTTTTCTTGTGTTTCCCCTAATG
Dist CAGATGTTCCGGAGTCAAATTGTACAGTTTTCTTGTGTTTCCCCTAATG

Alignment Paternal Deletion
Prox GTGTAGGAACAGTACATACAACAAGGCATCTTTAAAGGACAGAGTATTT
Del  GTGTAGGAACAGTACATACAACAAGGCATCTTTAAAGGACAGAGTATTT
Dist TATACTTCACATTTACAGTAATTACTATGGCTTTACTTTCAATTTGCTTCT

Prox CCTGAAC TACATATTAATGTTATTTAGACAGATTTTTTAAAAGTTTGATA
Del  CCTGAAC TACATATTAATGTTATTTAATATAAATGTTATTTAATATTTA
Dist CTCAGTTTTCTTCATCTTGCAACAAGACA-----

Prox TGTACATCAATATAAAATTTAAAAGGGAGTCATGGGTGGGGAAATGGGA
Del  ATATCAAGAAAAAGCAGGTAGTAGGTAGCTTCTAACAGAAGAACAGACAA
Dist --ATCAAGAAAAAGCAGGTAGTAGGTAGCTTCTAACAGAAGAACAGACAA

Prox AAATGCTGGTCAAGGTGTACAAAAGTTTCATTTAGGTAGCA
Del  AAGAGAGCATAAAGAAGATATTAGCCAAGAAGTGATGCTT
Dist AAGAGAGCATAAAGAAGATATTAGCCAAGAAGTGATGCTT
```

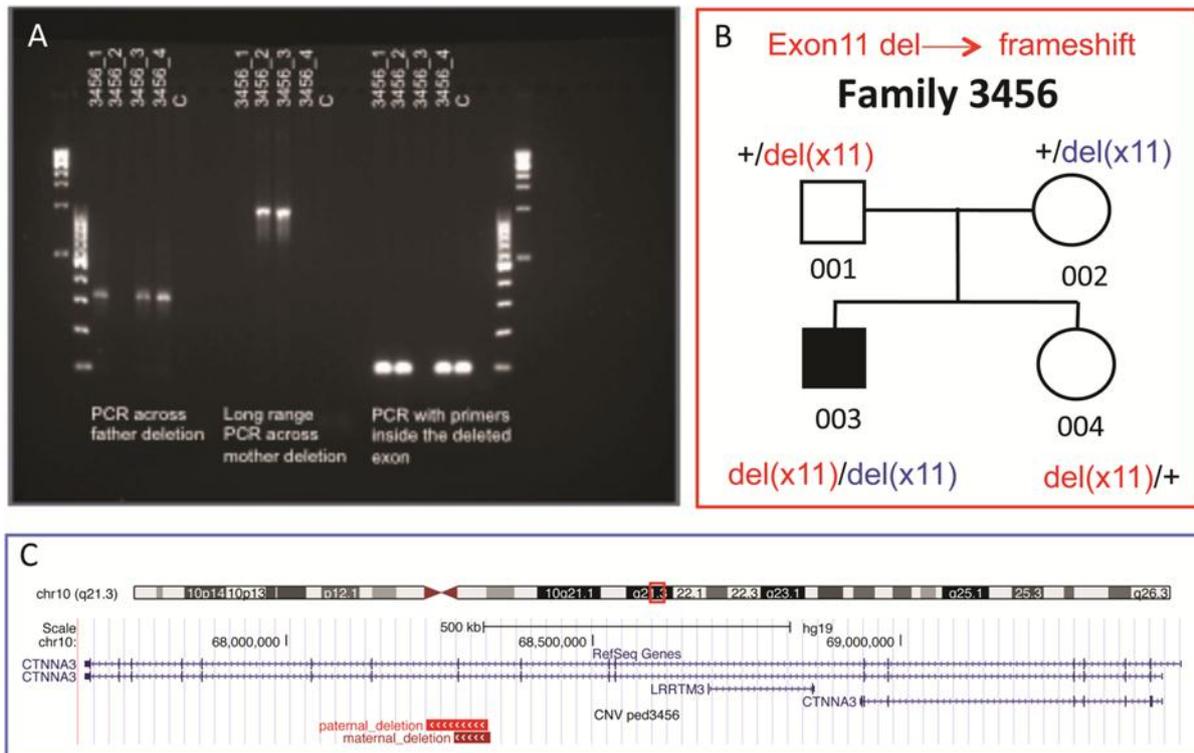
**Figure 7.2** (Bacchelli et al., 2014). Junction fragments of maternal and paternal deletions in family 3456. DNA sequences, obtained from direct sequencing of the junction fragments, were aligned to the normal wild-type proximal and distal sequences. The presence of bases with perfect microhomology to the normal proximal and distal wild-type sequences is shown in red.

The same PCR assays were also used to analyse the segregation of the two microdeletions in the discovery pedigree. Three PCR assays were performed: two PCR assays with primers designed across the paternal and maternal microdeletion and one PCR with primers within the deleted exon 11. Visualization of the PCR products on agarose gel showed that the proband inherits both the microdeletions, one heterozygous deletion from the father and one from the mother, resulting in the homozygous loss of exon 11; instead the unaffected sister inherits only the paternal microdeletion (**Figure 7.3A,B**).

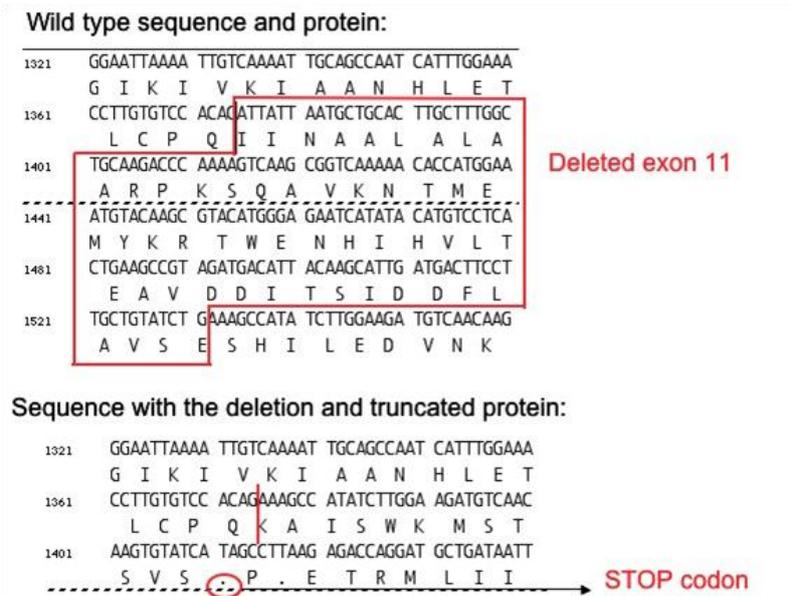
Exon 11 is present in both the two full-length *CTNNA3* isoforms (NM\_013266 and NM\_001127384), which are two transcript variants that differ only for the first 5' non-coding exon and encoding the same protein of 895 aminoacids (**Figure 7.3C**).

The absence of exon 11 causes a frameshift in the ORF and the introduction of 12 novel amino acids followed by a premature stop codon (**Figure 7.4**), resulting in a truncated protein with a length of 470 aminoacids. Given the likelihood of nonsense-mediated mRNA decay, this homozygous microdeletion would thus be predicted to result in a complete lack of functional protein in the affected individual.

## RESULTS



**Figure 7.3** (Bacchelli et al., 2014). Further characterisation of *CTNNA3* deletions in the pedigree 3456. **(A)** From the left, results of PCR across father's and mother's deletion breakpoints, respectively: only the allele with the deletion can be amplified and visualized as a band of 301 bp in the father (3456\_1), in the proband (4356\_3) and the unaffected sister (3456\_4), and as a band of 949 bp in the mother (3456\_2) and in the proband. On the right, the amplification with primers mapping in the deleted exon of *CTNNA3* indicates the presence of at least one allele without the deletion. At both extremities of the gel, 1-kb Plus and 100-bp DNA ladder were loaded. **(B)** The segregation pattern for these two deletions involving *CTNNA3* in the discovery pedigree. Autism is indicated in *black filling*. **(C)** Schematic from the UCSC genome browser. The figure shows the position of the two *CTNNA3* deletions in the pedigree 3456. The region shown corresponds to approximately 1.8 Mb on 10q21.3 (NCBI build 37 coordinates).



**Figure 7.4** At the top are shown the wild type sequence and protein of the *CTNNA3* gene; the red box indicates the exon 11 that is deleted on both alleles in the proband. At the bottom are shown the sequence without exon 11 and truncated protein; the premature stop codon is indicated as a red circle.

## RESULTS

### 7.1.2 *CTNNA3* exonic deletion frequency in ASD cases and controls

To compare the frequency of *CTNNA3* exonic deletions between ASD cases and controls, we used the existing CNV data on a total of 2,147 European ASD families from the AGP genome-wide study (combined sample of stage 1 and stage 2 families) (Pinto et al., 2014) and CNV data on 6,639 European controls (Bierut et al., 2010; Figueiredo et al., 2011; Fox et al., 2012; Krawczak et al., 2006; Shaikh et al., 2009; Stewart et al., 2009).

We identified a total of 14 additional heterozygous exonic deletions in ASD probands (allelic frequency = 0.37%) (Table 7.1; Table 7.2a), thus showing a modest deviation from the Hardy-Weinberg equilibrium (exact test  $P = 0.028$ ). Instead, in the control population, we found 43 *CTNNA3* exonic deletions (allelic frequency 0.32%) (Table 7.1; Table 7.2b), indicating a comparable frequency between ASD cases and controls ( $P = 0.62$ ).

**Table 7.1** (Bacchelli et al., 2014). **Frequency of *CTNNA3* exonic deletions in ASD cases and controls.**

	Number of subjects	Exonic deletions	Exonic deletion frequency (%)	$P$ value <sup>a</sup>	Frameshift deletions	Frameshift deletion frequency (%)	$P$ value <sup>a</sup>
<b>ASD cases</b>	2,147	16 <sup>b</sup>	0,37	0,62	6 <sup>b</sup>	0,14	0,56
<b>Controls</b>	6,639	43	0,32		14	0,11	

<sup>a</sup> Fisher's exact test; <sup>b</sup> Including the two exon 11 deletions in family 3456

**Table 7.2a *CTNNA3* exonic deletions in ASD cases**

Sample ID	Inheritance	Array_platform	Genomic coordinates (hg18) <sup>a</sup>	Length (bp)	Deleted exons <sup>b</sup>	Effect on CDS
5237_3	paternal	illumina 1Mv1 single	chr10:67628183-67752490	124308	x13	frameshift
3311_003	maternal	illumina 1Mv1 single	chr10:67688367-67759307	70941	x13	frameshift
3456_3	<b>maternal/paternal</b>	<b>illumina 1Mv3 duo</b>	<b>chr10:67898172-68002674<sup>c</sup></b>	<b>104502</b>	<b>x11</b>	<b>frameshift</b>
4196_1	mother	illumina 1Mv1 single	chr10:67912678-67976820	64143	x11	frameshift
4527_1	mother	illumina 1Mv1 single	chr10:67947835-67968052	20218	x11	frameshift
4228_1	N/A	illumina 1Mv1 single	chr10:67957765-68226404	268640	x10	in frame
6372_3	maternal (father N/A)	illumina 1Mv3 duo	chr10:67979174-68114740	135567	x10	in frame
14299_4200	paternal	illumina 1Mv1 single	chr10:67987089-68056520	69432	x10	in frame
3093_004	maternal	illumina 1Mv1 single	chr10:67987089-68067310	80222	x10	in frame
5013_3	paternal	illumina 1Mv1 single	chr10:68023745-68071037	47293	x10	in frame
3169_004	maternal	illumina 1Mv1 single	chr10:68029140-68183933	154794	x10	in frame
4211_1	father	illumina 1Mv1 single	chr10:68029140-68183933	154794	x10	in frame
4291_1	N/A	illumina 1Mv1 single	chr10:68052141-68210655	158515	x8-x9	in frame
5065_3	maternal	illumina 1Mv1 single	chr10:68138586-68227559	88974	x8-x9	in frame
3476_3	paternal	illumina 1Mv3 duo	chr10:68154851-68247375	92525	x8-x9	in frame

<sup>a</sup> The size of CNV shown is as detected by microarrays; <sup>b</sup> Exons numbering according to RefSeq NM\_013266; <sup>c</sup> Exact breakpoint coordinates; <sup>d</sup> The deletion includes also exon1 of *LRRTM3*; <sup>e</sup> The deletion includes also exon1 and exon2 of *LRRTM3*.

## RESULTS

**Table 7.2b CTNNA3 exonic deletions in controls**

Sample ID	Control cohort	Array_platform	Genomic coordinates (hg18)	Length (bp)	Deleted exons	Effect on CDS
B601040_1007870358	SAGE	illumina 1Mv1 single	chr10:67628183-67752490	124308	x13	frameshift
B782997_1007853703	SAGE	illumina 1Mv1 single	chr10:67628183-67754983	126801	x13	frameshift
B974175_1007875270	SAGE	illumina 1Mv1 single	chr10:67748487-67889985	141499	x12	in frame
CONSPC2_f_179516	POPGEN	Affy6	chr10:67754797-67871675	116879	x12	in frame
B355026_0067942568	SAGE	illumina 1Mv1 single	chr10:67820257-68046869	226613	x11	frameshift
B291548_1007841762	SAGE	illumina 1Mv1 single	chr10:67834450-68127819	293370	x11 & x10	frameshift
CONSPC_m_183371	POPGEN	Affy6	chr10:67911535-67975943	64409	x11	frameshift
HABC_900384_900384	HABC	Illumina 1Mv3 duo	chr10:67912678-67968052	55375	x11	frameshift
HABC_900268_900268	HABC	Illumina 1Mv3 duo	chr10:67920659-68059454	138796	x11 & x10	frameshift
CONT1635	OHI	Affy6	chr10:67937332-67971629	34298	x11	frameshift
CONT-1870	OHI	Affy6	chr10:67938271-68044090	105820	x11	frameshift
110036018737	OC	illumina 1Mv1 single	chr10:67938287-68059454	121168	x11 & x10	frameshift
Caucasian	CHOP	Illumina 550K	chr10:67944050-68069165	125116	x10-x11	frameshift
Caucasian	CHOP	Illumina 550K	chr10:67944050-68077867	133818	x10-x11	frameshift
Caucasian	CHOP	Illumina 550K	chr10:67944050-68088852	144803	x10-x11	frameshift
CONT1792	OHI	Affy6	chr10:67955244-68119974	164731	x10	in frame
Caucasian	CHOP	Illumina 550K	chr10:67960600-68052141	91542	x10	in frame
CONT2294	OHI	Affy6	chr10:67981580-68139445	157866	x10	in frame
B431282_1007873513	SAGE	illumina 1Mv1 single	chr10:67987089-68088852	101764	x10	in frame
Caucasian	CHOP	Illumina 550K	chr10:67997021-68053979	56959	x10	in frame
HABC_900966_900966	HABC	Illumina 1Mv3 duo	chr10:67997021-68091312	94292	x10	in frame
CONT-2085	OHI	Affy6	chr10:68009168-68066669	57502	x10	in frame
HABC_901209_901209	HABC	Illumina 1Mv3 duo	chr10:68013385-68156906	143522	x10	in frame
CONSPC_f_186937	POPGEN	Affy6	chr10:68020146-68070424	50279	x10	in frame
B779950_1007875276	SAGE	illumina 1Mv1 single	chr10:68023745-68071037	47293	x10	in frame
B941932_1007873623	SAGE	illumina 1Mv1 single	chr10:68023745-68071037	47293	x10	in frame
HABC_900941_900941	HABC	Illumina 1Mv3 duo	chr10:68023745-68127819	104075	x10	in frame
NCA07346	OHI	Affy6	chr10:68024710-68070424	45715	x10	in frame
CONT1593	OHI	Affy6	chr10:68024710-68075715	51006	x10	in frame
Caucasian	CHOP	Illumina 550K	chr10:68034046-68088852	54807	x10	in frame
CONBSP_m_213905	POPGEN	Affy6	chr10:68037680-68109008	71329	x10	in frame
Caucasian	CHOP	Illumina 550K	chr10:68040186-68088852	48667	x10	in frame
CONSPC2_m_195755	POPGEN	Affy6	chr10:68050416-68077020	26605	x10	in frame
B764705_1007853643	SAGE	illumina 1Mv1 single	chr10:68091312-68222395	131084	x9 & x8	in frame
HABC_902495_902495	HABC	Illumina 1Mv3 duo	chr10:68113705-68200470	86766	x9	in frame
Caucasian	CHOP	Illumina 550K	chr10:68127614-68216254	88641	x8-x9	in frame
Caucasian	CHOP	Illumina 550K	chr10:68160982-68305039	144058	x8-x9	in frame
Caucasian	CHOP	Illumina 550K	chr10:68180377-68268934	88558	x8-x9	in frame
Caucasian	CHOP	Illumina 550K	chr10:68190451-68252582	62132	x8-x9	in frame
NCA07559	OHI	Affy6	chr10:68197491-68520899	323409	x8 <sup>d</sup>	in frame
CONT-1970	OHI	Affy6	chr10:68211267-68626505	415239	x7 <sup>e</sup>	frameshift
CONBSP_f_186933	POPGEN	Affy6	chr10:68622864-68770434	147571	x6	in frame

Parental information was available for 13 out of 15 ASD families, and in none of these cases, *CTNNA3* exonic microdeletions were *de novo*. The observed exonic deletions are different in position, size, sequence junctions and genomic content, removing one or two exons, spanning from exon 6 to exon 13 of *CTNNA3* (**Table 7.2**). This is in accordance with *CTNNA3* being located in a common fragile site (Smith et al., 2006), a region characterised by increased genomic instability (Dillon, Pierce, Ng, & Wang, 2013).

Since deletions that result in a frameshift cause unambiguous loss-of-function alleles, we focused our attention to this class of deletions. Among the observed exonic *CTNNA3* deletions in cases and controls, only deletions that remove exon 7, exons 10-11, exon 11 and exon 13 (NM\_013266) are predicted to induce frameshifts, but their frequency is not significantly different between cases and controls (**Table 7.1**,  $P = 0.56$ ).

### **7.1.3 Segregation analysis of *CTNNA3* exonic deletions and mutation screening of *CTNNA3* and *LRRTM3* in four ASD families**

To investigate if *CTNNA3* exonic deletions segregate with the ASD phenotype we performed a quantitative real time PCR in three ASD multiplex families (3093, 3169 and 3311) and one singleton family (3476) with heterozygous *CTNNA3* deletions. All deletions segregate with the ASD phenotype except for the deletion in exon 13 that is transmitted from the mother to only two out of three affected children (**figure 7.5**).

Moreover, in order to test the hypothesis that the *CTNNA3* deletions could act by unmasking rare variants in the not deleted allele, we sequenced the entire coding sequence and intron-exon boundaries of *CTNNA3* and of the nested gene *LRRTM3* (leucine-rich repeat transmembrane neuronal 3) in all members of these four ASD families.

We did not detect any novel exonic variants in *CTNNA3*, while a previously undescribed missense change (p.Arg275Ser; G>A) was identified in the *LRRTM3* exon 2 in family 3476 (**figure 7.6**); however, this missense variant was transmitted from the unaffected father, who also carries the *CTNNA3* deletion, to the affected child. This result is thus not compatible with a two-hit model, since both the *CTNNA3* deletion and the *LRRTM3* missense variant are present in the unaffected father.

RESULTS

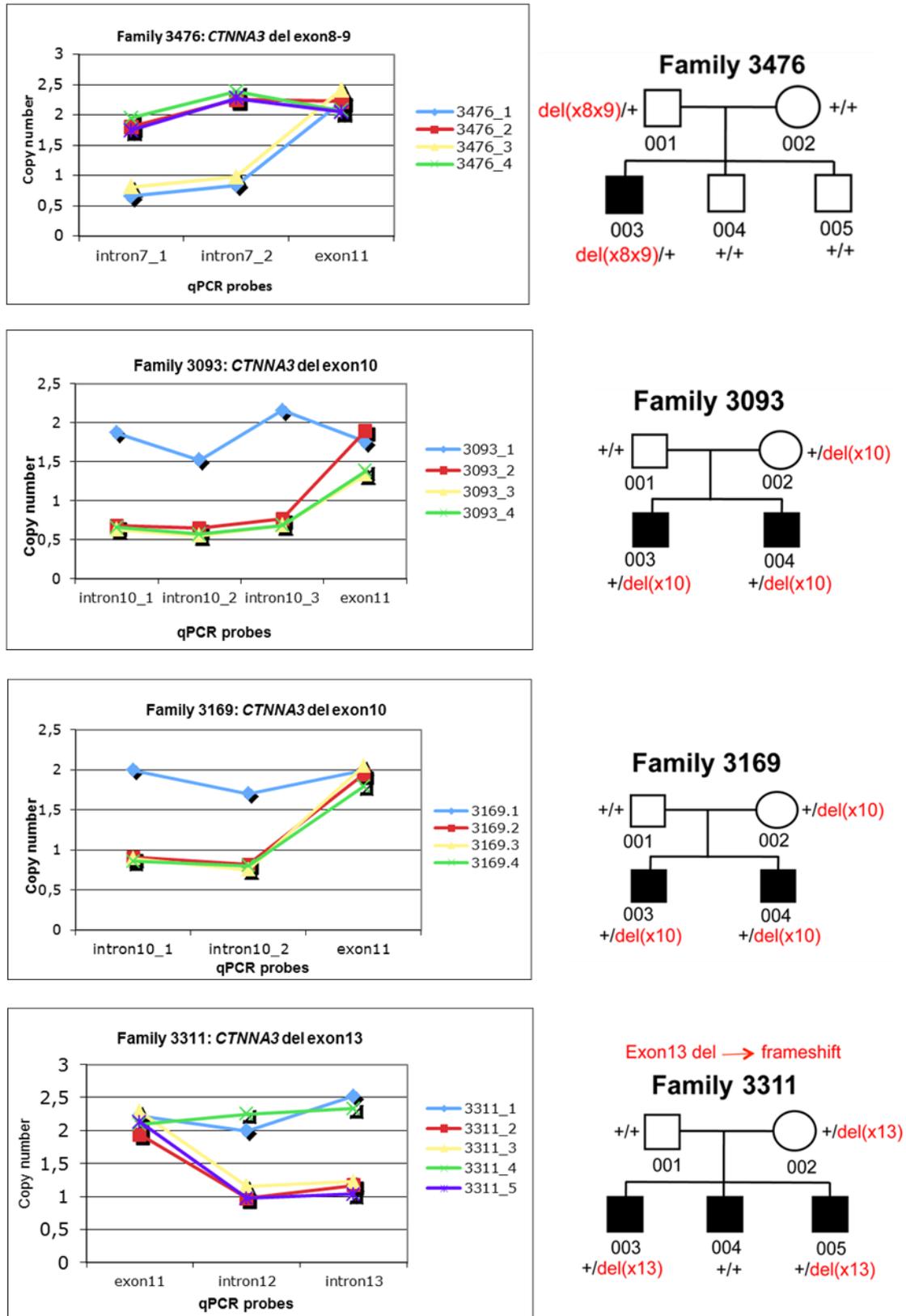
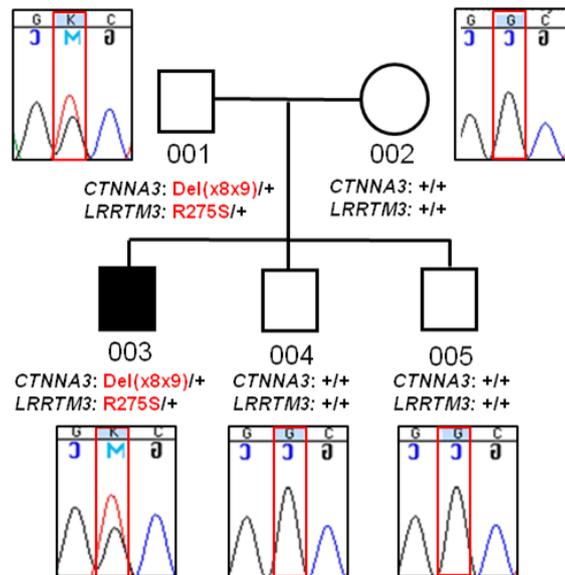


Figure 7.5 Pedigree of four ASD families carrying CTNNA3 exonic deletions. Black filling indicates ASD diagnosis.



**Figure 7.6:** Pedigree of family 3476 carrying the *CTNNA3* deletion of exons 8 and 9 and the missense change p.Arg275Ser in the *LRRTM3* gene.

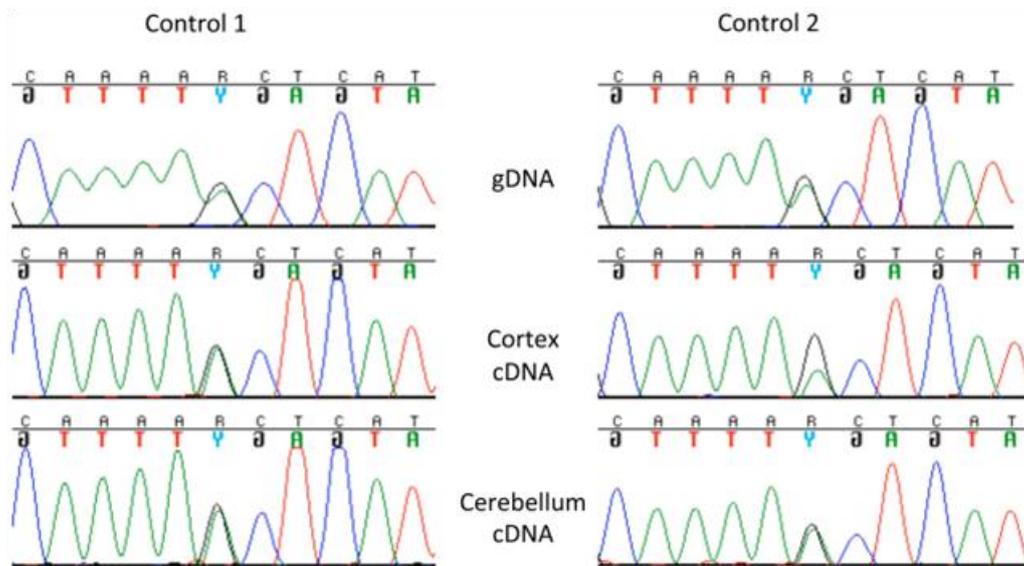
#### 7.1.4 *CTNNA3* expression analysis

In order to test the functional effect of the heterozygous and homozygous loss of exon 11 in family 3456, we analysed the *CTNNA3* expression in blood RNA or Epstein Barr virus (EBV)-transformed cell lines from family 3456 by RT-PCR. However, since *CTNNA3* is expressed in blood at very low levels, we were unable to obtain a reliable amplification of the *CTNNA3* transcript.

It has been previously reported that *CTNNA3* is subject to genomic imprinting, with preferential monoallelic expression of the maternal allele in placental tissue (Van Dijk et al., 2004). Therefore, to investigate whether *CTNNA3* is also monoallelically expressed in brain, we analyzed the allelic expression in the cerebellum and cerebral cortex of two informative heterozygous controls for a coding SNP (rs4548513, p.Ser596Asn) located in exon 13.

RT-PCR analysis showed high expression levels in these two brain tissues and Sanger sequencing of the PCR products revealed a biallelic gene expression, thus excluding the hypothesis of a monoallelic expression of *CTNNA3* in brain (**figure 7.7**).

## RESULTS



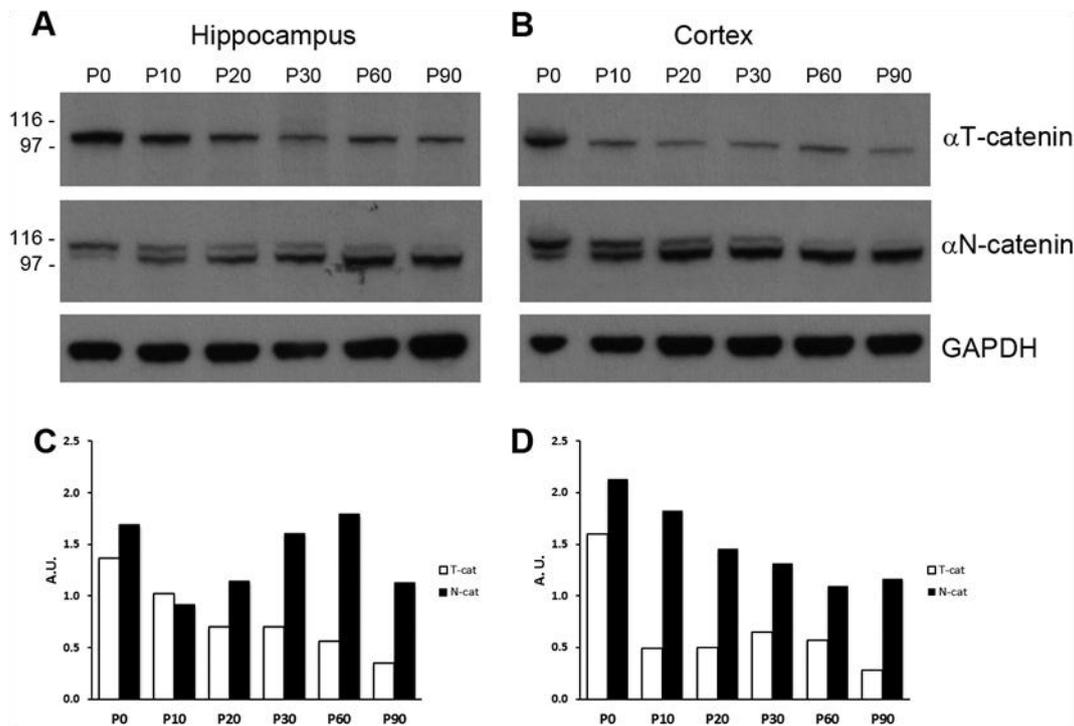
**Figure 7.7.** (Bacchelli et al., 2014). *CTNNA3* biallelic expression in the cerebellum and cerebral cortex. Sanger sequencing of rs4548513 (p.Ser596Asn) from genomic DNA (gDNA) and brain cDNA (cortex and cerebellum) of two adult controls showing the heterozygosity of the SNP and demonstrating biallelic expression of *CTNNA3* in both brain areas.

Since the mouse *Ctnna3* cDNA encodes a protein showing 95% identity to human *CTNNA3* and the genomic structures of the mouse *Ctnna3* and human *CTNNA3* genes are completely conserved, a Western blot analysis was performed to get a quantitative characterisation of *Ctnna3* expression during early brain mouse development compared with the neural specific *Ctnna2*.

Protein extracts of mouse cortex and hippocampus at different developmental stages (from P0 to P90) were probed with anti-N-catenin antibody that recognizes specifically the C-terminus of *Ctnna2* as a doublet band and a rabbit polyclonal anti- $\alpha$ T-catenin antibody (#952), which recognizes a specific peptide corresponding to the C-terminus of *Ctnna3*. As shown in **Figure 7.8**, not only *Ctnna2* is highly expressed at all brain developmental stages analysed, but also *Ctnna3* is present in both brain areas analysed. Noticeably, *Ctnna3* showed a higher expression in the hippocampus and cortex at P0, suggesting a specific neuronal role in very early developmental stages.

The Western blot experiment was performed at San Raffaele Scientific Institute of Milan in collaboration with Dr. M. Giannandrea and Dr. P. D'Adamo.

## RESULTS



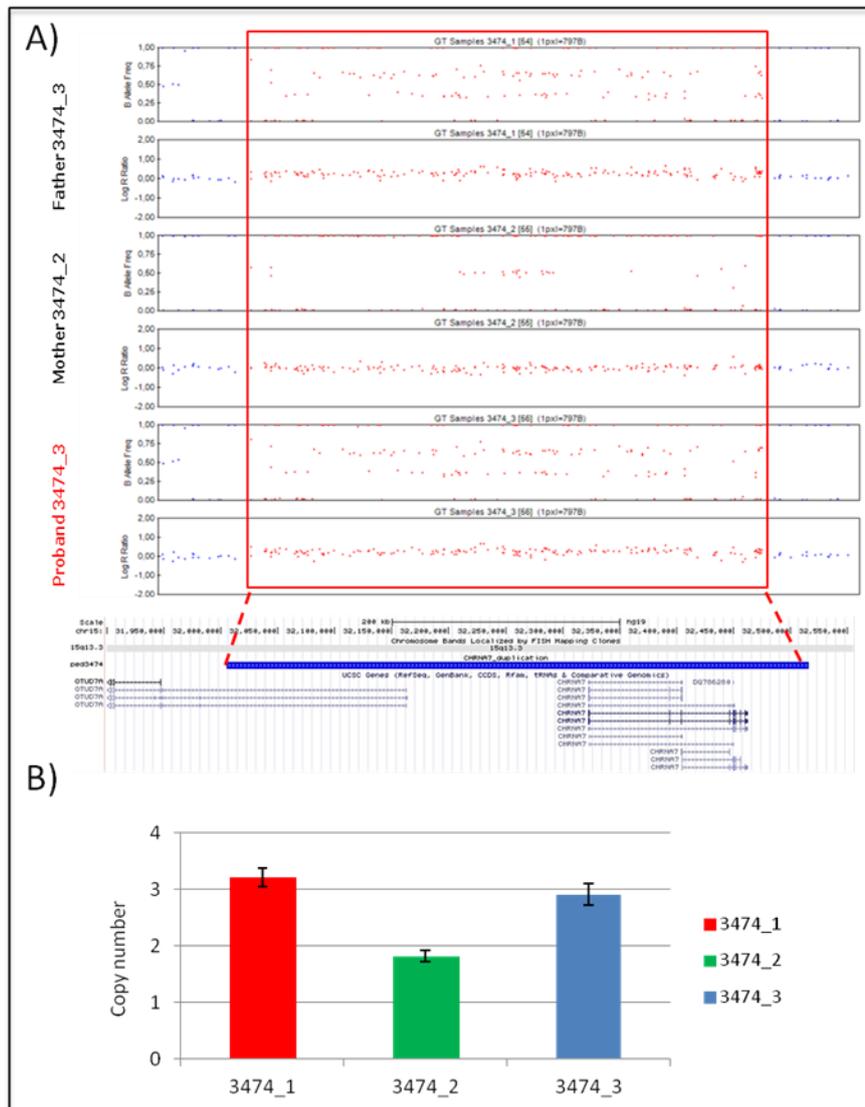
**Figure 7.8** (Bacchelli et al., 2014). **Expression analysis of *cttna3* in mouse brain.** Western blot analysis of  $\alpha$ T-catenin and  $\alpha$ N-catenin in mouse hippocampus (A) and cortex (B) at different developmental stages (from P0 to P90). (C, D) Histograms showing the band intensity normalised by *GAPDH* as internal loading control.

## 7.2 Analysis of rare variants in the *CHRNA7* gene

### 7.2.1 Validation of the *CHRNA7* microduplication in family 3474

The presence of the microduplication identified in the proband 3474\_3 by SNP array (see preliminary results) was verified by visual inspection of the 1M Illumina SNP data, by plotting intensities (logR ratio) and allelic ratios (B allele frequency) of all members of the family 3474 (Figure 7.9A). This CNV, including exon 1 of the longer isoform of *OTUD7A* and the entire *CHRNA7* gene (chr15: 32,005,348-32,515,973 NCBI build 37 coordinates), is transmitted from the father (3474\_1) to the affected son. The duplication was also validated by quantitative PCR in all family members. As shown in Figure 7.9B, the father and the proband have 3 copies of *CHRNA7*, while the mother (3474\_2) has a normal number of copies.

## RESULTS



**Figure 7.9** (Bacchelli et al., 2015) **A)** GenomeStudio screenshot showing B-allele frequency and log R ratio for family 3474. The duplication results in an increase to the Log R ratio and deviation in the allelic ratio for heterozygote variants away from the expected 0.5. The SNPs included in the duplication are boxed. **B)** qPCR results for the *CHRNA7* duplication in family 3474.

CNV analysis of family 3474 led also to the identification of a ~55 kb paternal deletion of *ARHGAP11B*, which is known to accompany the majority of *CHRNA7* microduplications (Szafranski et al., 2010) and a rare maternal gain of unknown significance on chromosome 5q13.2 (chr5:68,594,539-68,638,941). This duplication involves the first seven exons of *CCDC125* (Coiled-coil domain-containing protein 125), a gene with a putative role in controlling the cell motility of immune systems (Araya et al., 2009).

## RESULTS

**Table 7.3** Other CNVs identified in proband 3474 by CNV analysis

CNV	Gene	Position (hg19)	Length	Inheritance
<b>del15q13.2</b>	<i>ARHGAP11B</i>	chr15:30,913,207-30,968,006	54,800	paternal
<b>dup5q13.2</b>	<i>CCDC125</i>	chr5:68,594,539-68,638,941	44,402	maternal

### 7.2.2 CNV analysis and mutation screening in 135 ASD subjects

In order to investigate if rare variants in *CHRNA7* could play an important role in ASD, we performed CNV analysis and sequence mutation screening of the coding sequence of the *CHRNA7* gene in a sample of 135 Italian ASD probands. The main clinical features of the 135 Italian ASD cases included in the study are shown in **Table 8.4**.

	Total	Sex		Epilepsy	Intellectual disability					
		Male	Female		severe (20<IQ<34)	moderate (35<IQ<49)	Mild (50<IQ<69)	BCI (70<IQ<85)	normal (IQ>85)	Unknown
<b>Autism</b>	98	76	22	2	0	11	27	20	17	23
<b>Atypical autism</b>	34	28	6	0	2	3	11	3	7	8
<b>Asperger syndrome</b>	3	2	1	0	0	0	0	1	1	1
<b>Total</b>	135	106	29	2	2	14	38	24	25	32

BCI, borderline cognitive impairments

#### 7.2.2a CNV analysis

68 out of 135 Italian ASD probands have been already included in the CNV screening performed by the AGP. Instead, 67 individuals that were not previously analysed by SNP array, were tested by qPCR to identify CNVs in *CHRNA7* using a probe mapping in exon 3, specific for *CHRNA7*. The proband of family 3474 was included as positive control. No additional deletion or duplication was identified in this sample.

#### 7.2.2b *CHRNA7* mutation screening

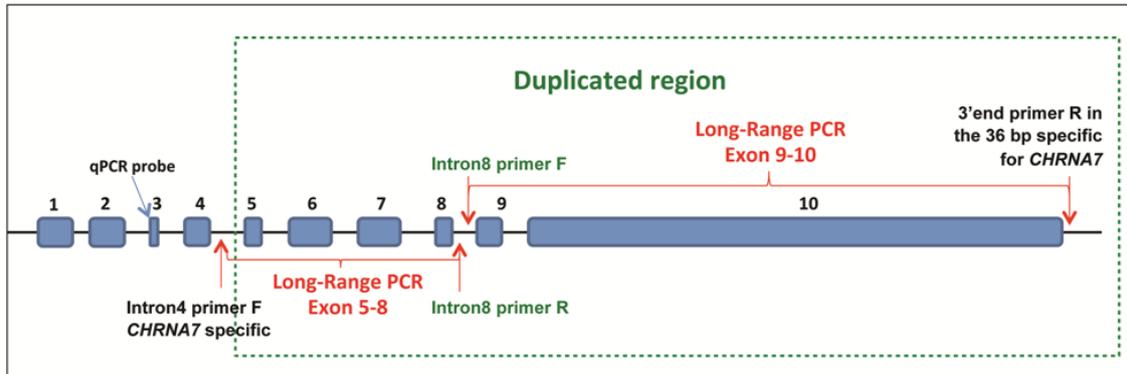
All 10 exons and intron-exon boundaries of *CHRNA7* were sequenced by Sanger method in all 135 ASD individuals .

The first four exons, which are specific for *CHRNA7*, were amplified using exon specific primers corresponding to flanking intronic sequences and then subjected to Sanger sequencing.

The genomic sequence comprising *CHRNA7* exons 5-10 is duplicated and nearly identical (>99%) in the *CHRFAM7A* gene, complicating the mutation screening (Gault et al., 1998) (chapter 3, paragraph 3.4.3). Therefore, to selectively amplify *CHRNA7* exons 5-10, we performed two different long range PCRs (LR-PCRs): x5-x8 LR-PCR encompasses exon 5 to exon 8 and amplifies

## RESULTS

a segment of 6752 bp; x9-x10 LR-PCR encompasses exon 9 to exon 10 and amplifies a segment of 7500 bp. For the x5-x8 LR-PCR we used a primer F outside the duplication (mapping to *CHRNA7* specific intron 4) and a primer R in the duplicated region (mapping to intron 8), while for the x9-x10 amplicon we used a primer F in the duplicated region (mapping to intron 8) and a primer R mapping in a *CHRNA7* specific region located at 3'UTR which contains a 36 bp insertion (**Figure 7.10 and 7.11C**). These two LR-PCR products were sequenced with primers specific for each exon.



**Figure 7.10** (Bacchelli et al., 2015). Genomic structure of the human *CHRNA7* gene, with the position of the primers used for the x5-x8 and x9-x10 LR-PCRs and the qPCR probe. The dashed rectangle indicates the region duplicated in *CHRFAM7A*.

By aligning the genomic sequences of *CHRNA7* and *CHRFAM7A* as reported in human genome reference sequence (GRCh37, hg19) we noted that, in addition to 36 bp at the 3' end of *CHRNA7* (chr15:32,462,638-32,462,673) (**Figure 7.11C**), these two genes differ for the presence/absence of some bp located in intron 5 and intron 9. As shown in **Figure 7.11A,B**, intron 5 and intron 9 of *CHRNA7* contain the 4 bp GTCT (chr15:32,448,389-32,448,392) and the 5 bp AAGAT (chr15:32,458,725-32,458,729), respectively, which are absent in the duplicate gene *CHRFAM7A*.

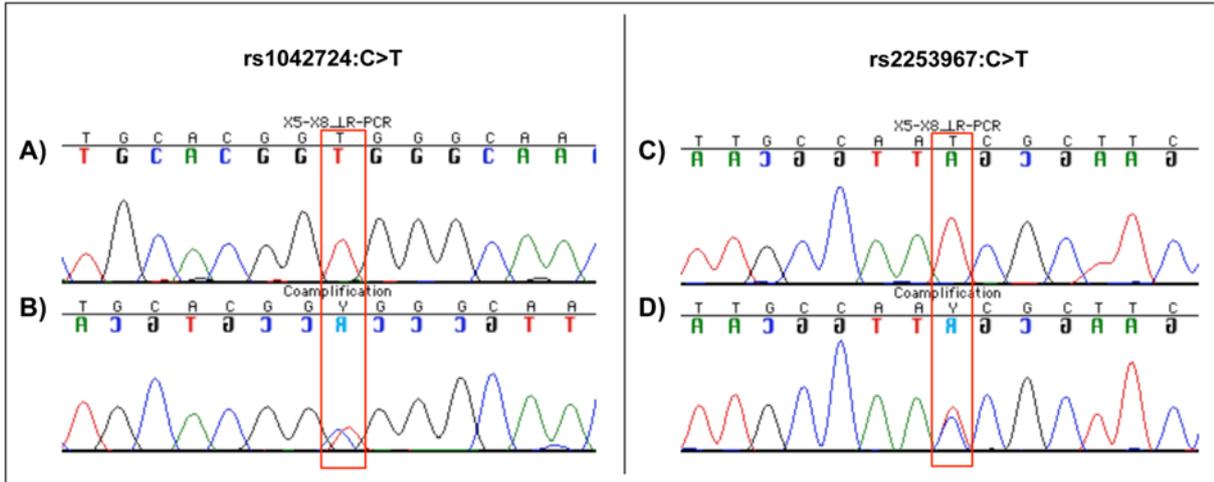
<b>A) Intron 5</b>	
<i>CHRNA7</i>	32,448,362 TCTGCTTCTGTCTCTAGAAATTCTCTGTCTGTCTCGGGCCATTATTTTCGTAGAAATG 32,448,419
<i>CHRFAM7A</i>	30,666,781 TCTGCTTCTGTCTCTAGAAATTCTCT:::GTCTCGGGCCATTATTTTCGTAGAAATG 30,666,728
<b>B) Intron 9</b>	
<i>CHRNA7</i>	32,458,696 CAGACCTGTTTGTTCAGTTATCTTTGATAAGATGTTATAAAAAGATGACCAAGAATG 32,458,753
<i>CHRFAM7A</i>	30,656,454 CAGACCTGTTTGTTCAGTTATCTTTGAT:::GTTATAAAAAGATGACCAAGAATG 30,656,402
<b>C) 3' end</b>	
<i>CHRNA7</i>	32,462,625 CGGTTTTTTTTTTTTTTTCTGTTACAGTGTCTTTAGAAACAAGTAGAAGTGTTTTGAT 32,462,682
<i>CHRFAM7A</i>	30,652,528 CGGTTTTTTTTTTT:::C:::TGTTTGAT 30,652,506

**Figure 7.11** (Bacchelli et al., 2015) Sequence differences in the duplicated regions of *CHRNA7* and *CHRFAM7A*, according to NCBI human build 37 (hg19). The genomic position indicate the first and to the last nucleotide of each line. The exact coordinates of the 3 small insertions present in *CHRNA7* and absent in *CHRFAM7A* are: **A)** chr15: 32,448,389-32,448,392 for the 4 bp in intron 5; **B)** chr15: 32,458,725-32,458,729 for the 5 bp in intron 9; **C)** chr15: 32,462,638-32,462,673 for the 36 bp at the 3' end of *CHRNA7*.



## RESULTS

insertion in intron 9 as reported GRCh37 is actually present in both *CHRNA7* and *CHRFAM7A*, at least in most individuals.



**Figure**

**re 7.13** (Bacchelli et al., 2015). Genotypes of rs1042724 and rs2253967 in two different probands, obtained by sequencing the x8-x9 LR-PCR (A, C) versus the coamplification product (B, D).

By sequencing the entire open reading frame of *CHRNA7* we identified four putatively functional rare (<1% minor allele frequency) variants: a non-synonymous variant in exon 10 and three variants located in the proximal promoter region (S. Leonard et al., 2002) (**Table 7.5a**).

**Table 7.5** (Bacchelli et al., 2015). Rare variants identified in the *CHRNA7* gene by mutation screening.

Region (bp distance from ATG)	hg19 position (SNP ID)	Protein NP_000737.1 (cDNA)	135 ASD cases Genotype Count (MAF)	Controls (1000 Genomes)		
				98 TSI Genotype Count (MAF)	85 CEU Genotype Count (MAF)	379 EUR* Genotype Count (MAF)
<b>a) Rare putative functional variants</b>						
Promoter (-241)	g.32322557A>G (rs188889623)	(c.1-129A>G)	AA=132;AG=3 (0,011)	AA=96; AG=2 (0,010)	AA=85 (0)	AA=374; AG=5 (0,006)
Promoter (-191)	g.32322607G>A	(c.1-79G>A)	GG=134;GA=1 (0,004)	-	-	-
Promoter (-182)	g.32322616 delG	(c.1-70_1-69delG)	GG=134;G/delG=1 (0,004)	-	-	-
Exon 10	g.32460504 G>A (rs199504752)	p.E452K (c.1466G>A)	GG=134;GA=1 (0,004)	-	-	-
<b>b) 5'-UTR variant found in the same individual carrying the -241 variant</b>						
5'-UTR(-86)	g.32322712C>T (rs149637464)	(c.27C>T)	CC=112;CT=23 (0,085)	CC=85;CT=13 (0,066)	CC=75; CT=10 (0,059)	CC=336; CT=43 (0,057)

## RESULTS

---

The non-synonymous variant in exon 10 causes an amino acid substitution from Glutamate to Lysine (p.E452K, rs199504752) and it has been identified in only one ASD proband who inherited it from the unaffected mother. Since this variant is located in *CHRNA7* exon10, which is duplicated in the *CHRFAM7A* gene, it was not possible to estimate the frequency of this non-synonymous variant from public databases.

Therefore we have sequenced *CHRNA7* exon 10 by LR-PCR in a control sample consisting of 125 Italian unaffected individuals and we detected the p.E452K variant in three unrelated individuals. However, since the frequency of this non-synonymous variant is not different between our ASD sample and the control sample (1/135 in ASD vs. 3/125 in controls, two sided Fisher exact test P-value.0.35) and in silico analysis using PolyPhen-2 (<http://genetics.bwh.harvard.edu/pph2/>) and SIFT (<http://sift.jcvi.org/>) does not predict any deleterious effect on the protein, it is unlikely that this variant contributes to ASD risk.

The three variants identified in the proximal promoter region are located at -241 bp (rs188889623, c.1-129A>G), at -191 bp (c.1-79G>A), and at -182 bp from ATG (c.1-70\_1-69delG) (**Table 7.5a**). One out of the three individual with the -241 bp variant (proband 3377\_3) also carries another more frequent variant in the 5' UTR region (g.32322712C>T, rs149637464, -86 bp from ATG) (**Table 7.5b**). Interestingly, these two variants have been previously reported to have a functional effect on *CHRNA7* gene transcription, being strongly associated with a significant decrease of promoter activity ( $P<0,0001$ ) (S. Leonard et al., 2002). Although these two variants have been individually found in the sample of 174 Italian unaffected individuals and are separately reported in the 1000 Genome project (**Table 7.5**), none of the 174 Italian controls or the 379 EUR patients carry both variants.

Segregation analysis of the -86/-241 variants in the Italian family showed that the -86 bp variant was inherited from the mother, while the -241 bp variant was inherited from the father (**Figure 7.13**), indicating thus that these two variants are not on the same chromosome and suggesting that *CHRNA7* expression might be significantly decreased in proband 3377\_3. Unfortunately, we were not able to test this hypothesis, since *CHRNA7* mRNA is not detectable in blood.

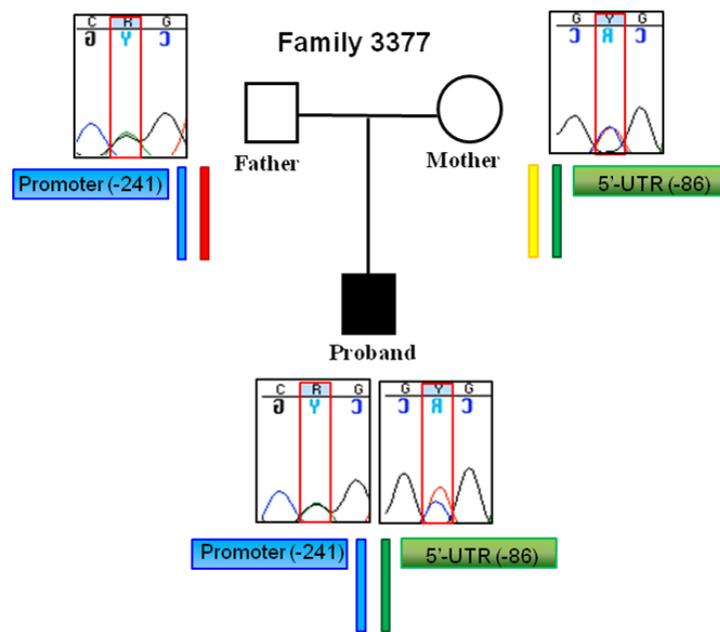


Figure 7.13. Segregation analysis of the -86/-241 variants in family 3377

### 7.3 Analysis of a rare microdeletion on chromosome 2q31.1

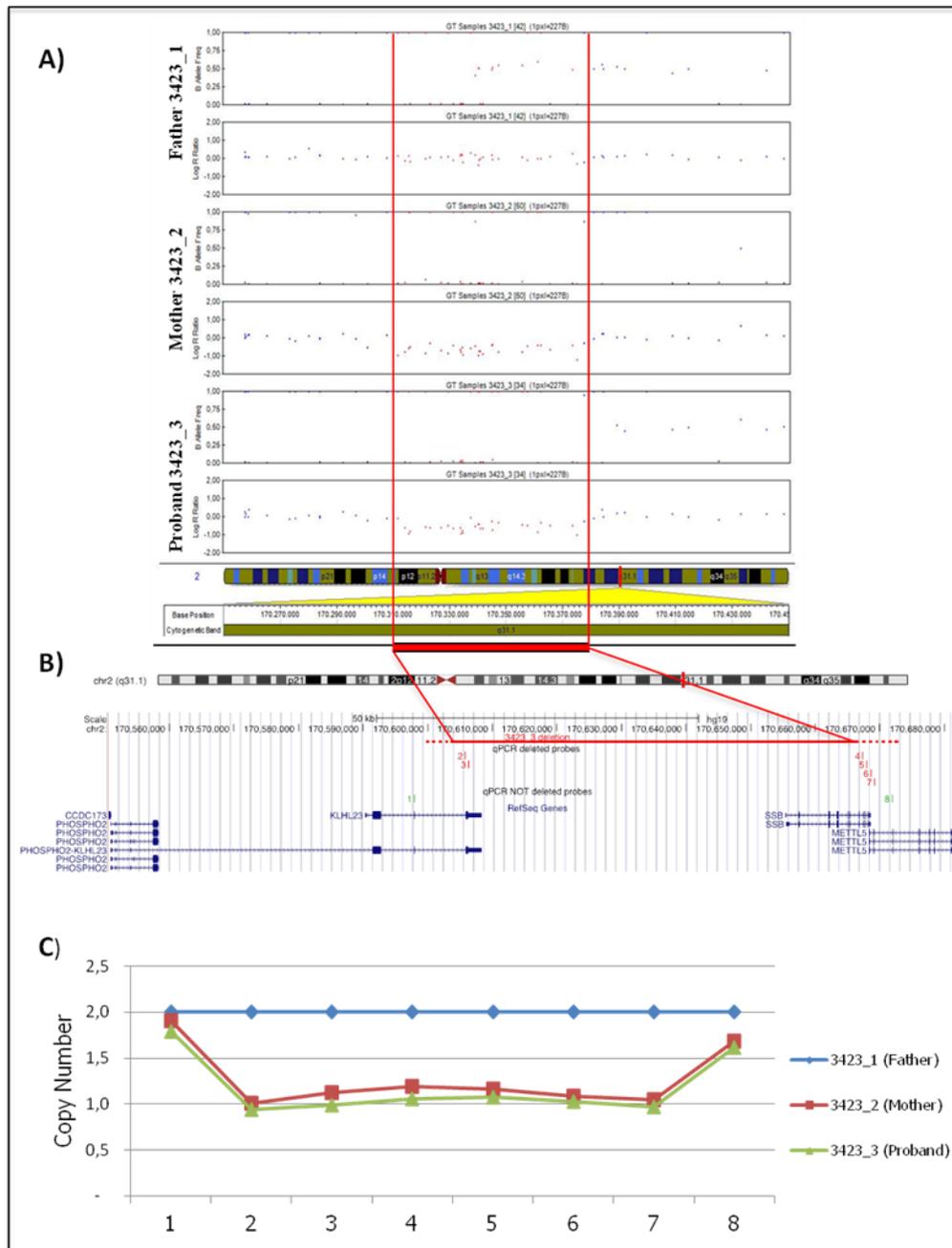
#### 7.3.1 Fine mapping of a rare microdeletion on chromosome 2q31.1

The presence of the microdeletion encompassing three neighbouring genes on the chromosome 2q31.1, *KLHL23*, *PHOSPHO2-KLHL23*, *SSB*, was checked by manual inspection of the log R ratios and B allele frequencies of family 3423 in Genome Studio: the mother (3423\_2) and the affected son (3423\_3) show a decrease in the Log R ratio and values of B-allele frequency around 0 and 1, while the father (3423\_1) shows values of log R ratio around zero and values of B-allele frequency around zero, 0.5 and 1, as he does not carry the deletion (**Figure 7.14A**).

The minimum and the maximum deleted regions were determined to be from rs6745705 to rs3754913 (chr2:170603578-170666813, GRCh37) and from rs6725295 to rs10190853 (chr2:170599690-170672843, GRCh37), respectively, in the mother and the affected subject (**Figure 7.14B**).

However, at the distal end, this resolution was not sufficient to determine if the deletion affect also the *METTL5* gene. Therefore, in order to validate and fine map the deletion further, we performed a quantitative PCR in the family 3423 using eight different probes, encompassing *KLHL23*, *SSB* and *METTL5*. The results showed that in the mother and in the affected son the last exon of *KLHL23*, the entire *SSB* gene and the last two exons of the *METTL5* gene are deleted, resulting in one copy of this genomic region (**Figure 7.14C**).

## RESULTS



**Figure 7.14: Fine-mapping of the deletion in family 3423.** **A)** GenomeStudio screenshot showing B-allele frequency and log R ratio for family 3423. SNP data are from the Illumina Infinium 1M-Single SNP array. SNPs within the deletion are boxed and highlighted in red. **B)** Schematic representation from the UCSC genome browser (GRCh37/hg19) showing the *PHOSPHO2-KLHL23*, *KLHL23*, *SSB* and *METTL5* loci with respect to chromosome 2. The red bar represents the minimum deleted region identified by Pinto et al. (2014) in proband 3423\_3 (chr2:170603578-170666813), whereas the maximum deleted region (chr2:170599690-170672843), which extends also to the *METTL5* gene, is indicated as a dotted line. The probes used for qPCR validation are depicted in red (deleted probes) and in green (not-deleted probes). **C)** qPCR results for the exonic deletion in family 3423. The eight qPCR fragments map to exon 3, intron 3 and exon 4 of *KLHL23*, to intron 9 and intron 11 of *SSB*, and to intron 6, exon 6 and intron 7 of the *METTL5* gene (probes 1-8, respectively). The number of copies of each amplified fragment was calculated using the  $2^{-\Delta\Delta C_t}$  method. The qPCR data were compared against a control gene (*FOXP2*) and the copy number has been normalized against the father 3423\_1.

### 7.3.2 CNV search in the 2q31.1 locus

In order to investigate the frequency of CNVs overlapping the 2q31.1 locus, we searched CNVs data published in the SFARI AutDB database (Basu, Kollu, & Banerjee-Basu, 2009), an evolving database containing different kinds of genetic data about autism. While no CNVs affecting *SSB* and/or *METTL5* are present, an intragenic deletion in *PHOSPHO2-KLHL23*, a read-through transcript between the neighbouring *PHOSPHO2* and *KLHL23* genes which encodes the same *KLHL23* protein, has been reported (Prasad et al., 2012).

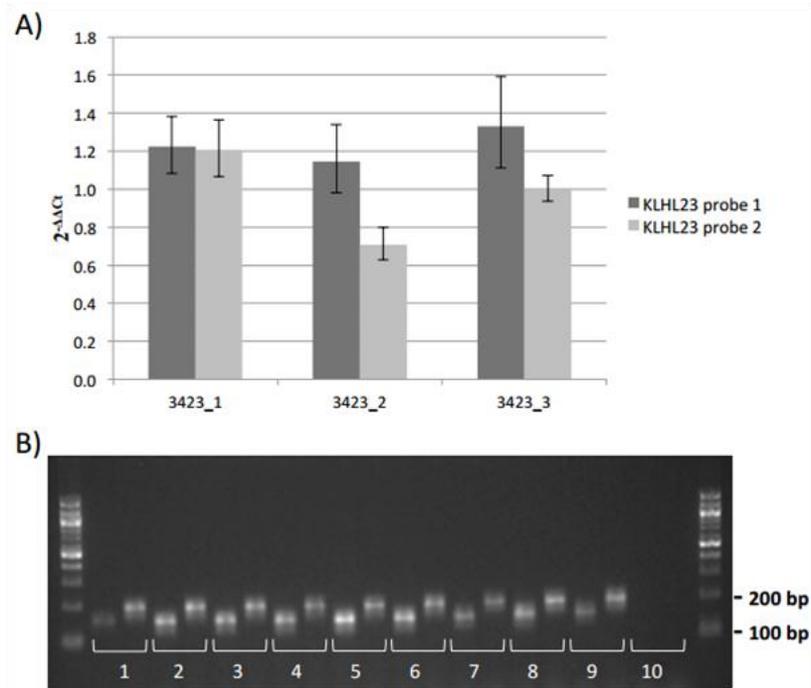
Therefore, given the presence of another CNV impacting only *KLHL23* and considering the biological function of the KLHL (Kelch-like) family members and their involvement in neuropsychiatric disorders (see paragraph 3.4.4), we focused our attention on *KLHL23* performing expression analyses and mutation screening.

### 7.3.3 *KLHL23* and *PHOSHO2-KLHL23* expression analyses

To investigate if *KLHL23* haploinsufficiency may influence the autism phenotype, we tested the *KLHL23* expression in family 3423 by quantitative RT-PCR. Expression analysis revealed decreased levels of the wild type *KLHL23* transcript, assessed using a probe extending across exon 3 and the deleted exon 4 (probe 2), in the mother and the proband, while no difference in expression was shown by the other probe (probe 1) mapping across the two non-deleted exons 2 and 3 (**Figure 7.15A**).

Moreover, in order to get a global overview of the expression of *KLHL23* and the *PHOSPHO2-KLHL23* fusion gene, we tested their expression across multiple tissues using a commercially available cDNA panel (Human MTC Panel I, Clontech). Both transcripts are ubiquitously expressed in all the tested tissues (**Figure 7.15B**), thus not supporting the hypothesis of a tissue-specific expression of one of them.

## RESULTS



**Figure 7.15 Expression analyses.** **A)** Two qPCR fragments were tested to analyse the *KLHL23* transcript levels: one spanning exons 2-3 (probe1, outside of the microdeletion) and another one spanning exons 3-4 (probe 2, exon 4 maps inside the microdeletion). All the data were normalized using the housekeeping gene *GUSB* as reference gene. Expression levels were also normalized against a control individual. The  $2^{-\Delta\Delta C_t}$  method was applied to estimate the difference in the gene expression between samples. **B)** *KLHL23* and *PHOSPHO2-KLHL23* expression in a multiple tissue cDNA panel. The different tissues are indicated by numbers: 1) Lung; 2) Heart; 3) Kidney; 4) Placenta; 5) Pancreas; 6) Liver; 7) Skeletal muscle; 8) Brain; 9) Human control; 10) Negative control. The *KLHL23*-specific amplicon is visualized as a band of 156 bp, while amplicon specific for the *PHOSPHO2-KLHL23* gives a band of 200 bp.

### 7.3.4 *KLHL23* and *PHOSHO2-KLHL23* mutation screening

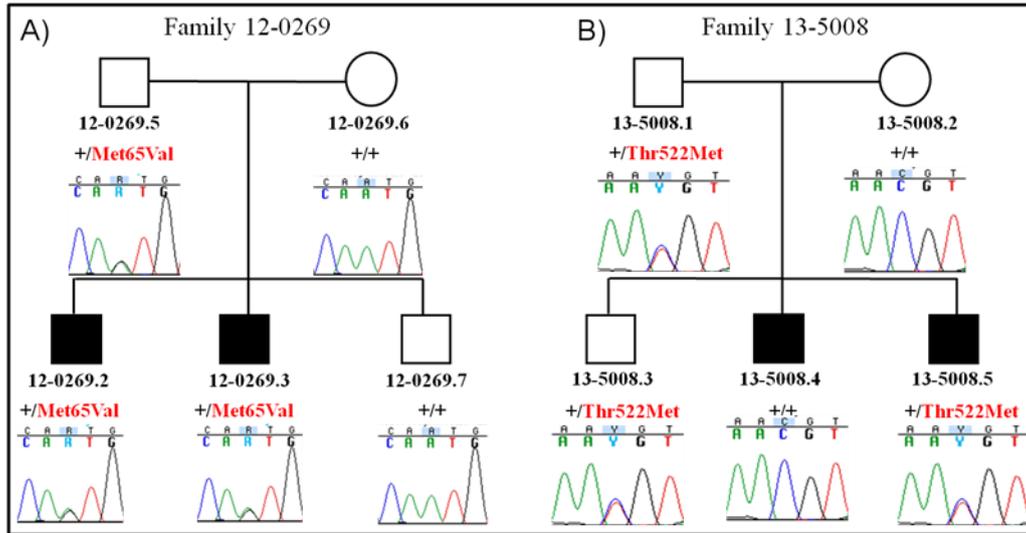
We sequenced the entire coding region of *KLHL23* (NM\_144711.5) and *PHOSPHO2-KLHL23* (NM\_001199290.1) in proband 3423\_3, in order to test the hypothesis that the deletion could act by unmasking rare variants in the not-deleted allele, and in 96 additional unrelated ASD probands from the IMGSAC multiplex families, in order to identify rare sequence variants which may have an etiological role in ASD.

No rare point mutations were found in proband 3423\_3, while two new non-synonymous rare variants, p.Met65Val (g.chr2:170,591,717 A>G, GRCh37) and p.Thr522Met (g.chr2:170,606,130 C>T, GRCh37), were detected in exon 2 and exon 4 of *KLHL23*, respectively, in two ASD unrelated individuals, 12-0269.2 and 13-5008.5 (**Table 8.6**). The other variants were identified in the non coding region of *PHOSPHO2-KLHL23* and they were reported in dbSNP.

Bioinformatic analysis predicted that both the missense changes are likely to have a damaging effect (PolyPhen2 score: 0.999 for p.Met65Val and 0.796 for p.Thr522Met; SIFT score: 0 for both p.Met65Val and p.Thr522Met).

## RESULTS

Segregation analysis in these two multiplex families showed that the p.Met65Val variant is transmitted from the father (12-0269.5) to both two affected children (12-0269.2; 12-0269.3), but not to the non-affected one (12-0269.7) (**Figure 7.16A**). Instead, the p.Thr522Met variant is inherited from the father (13-5008.1) by one affected (13-5008.5) and non-affected child (13-5008.3), but not by the other affected son (13-5008.4) (**Figure 7.16B**)



**Figure 7.16** Electropherograms showing the segregation analysis in the two multiplex families (family 12 02269 and family 13 5008) of the two missense variants, p.Met65Val and p.Thr522Met, respectively.

**Table 7.6** Variants identified in *PHOSPO2-KLHL23* and *KLHL23* mutation screening. The two missense changes are indicated in red.

Region (bp distance from ATG)	hg 19 position (SNP ID)	97 ASD cases Genotype Count	Type of change (NP_653312)	PolyPhen-2 score (HumDiv)	SIFT prediction (cutoff=0.05)
-328 (NM_001199290.1)	g.170551011C>G (rs115888469)	CC=96; CG=1			
-290 (NM_001199290.1)	g.170551049 A>G (rs76102705)	AA=85; AG=10; GG=2			
-207 (NM_001199290.1)	g.170551132 G>A (rs77728018)	GG=86; GA=9; AA=2			
-191 (NM_001199290.1)	g.170551740 T>G (rs542956085)	TT=95; TG=2			
Exon 2 (NM_144711.5)	g.170,591,717 A>G none	AA=96; AG=1	p.Met65Val	0.999 (probably damaging)	0 (damaging)
Exon 4 (NM_144711.5)	g.170,606,130 C>T none	CC=96; CT=1	p.Thr522Val	0.796 (possibly damaging)	0 (damaging)

Since the p.Met65Val segregates with the ASD phenotype in the family, we genotyped 462 Caucasian controls from the European Collection of Cell Culture (ECACC) by endonuclease restriction analysis using the BtsI enzyme, that is able to cut a PCR fragment containing the minor

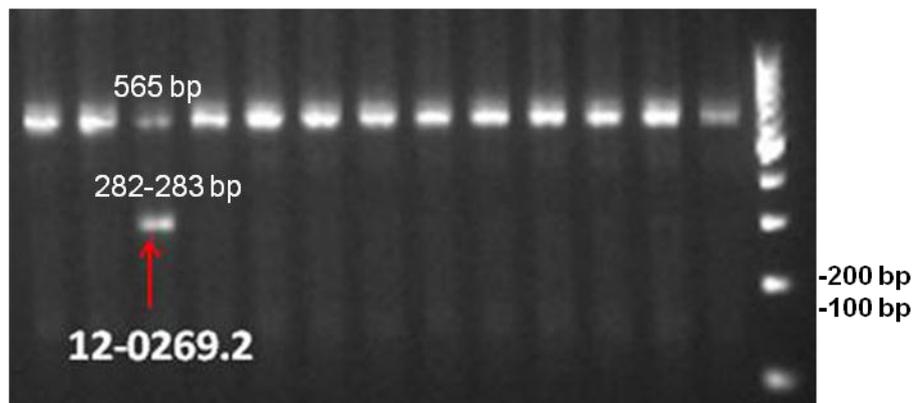
## RESULTS

---

allele G of the p.Met65Val variant. In our case, the PCR fragment used to amplify the exon 4 and the surrounding intronic regions had a size of 565 bp and a unique target site for BtsI (**figure 7.17**). In the presence of the minor allele G, the PCR fragment was cut in three fragments of 565 bp, 282 bp and 283 bp, while in the presence of the wild type site the restriction site was not recognized by the enzyme BtsI and the PCR fragment was not cut.

Visualization of the restriction products on a 2% agarose gel allowed the discrimination of homozygous individuals of the reference allele (A/A), who presented one band of 565 bp and the and heterozygotes individuals (A/G), who displayed two bands: the 565 bp fragment derived from the reference allele A and one band of about 280 bp (282 bp and 283 bp fragments) from the minor allele G (**Figure 7.18**).

By PCR-RFLP the p.Met65Val variant was not detected in the 924 Caucasian control chromosomes.



**Figure 7.18:** Example of genotyping of the p.Met65Val variant using the restriction endonuclease BtsI in Caucasian controls. The red arrow indicates the heterozygous individual (12-0269.2) carrying the variant p.Met65Val. The ladder 100 bp was loaded in the last lane.

## CHAPTER 8:

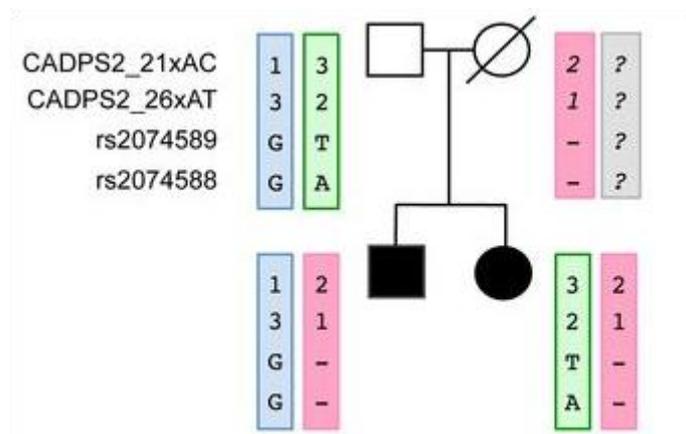
## RESULTS

**Analysis of the *CADPS2* gene in subjects with ID and/or ASD**

Given the discovery of a novel intragenic deletion and of a new functional SNV (p. Asp1113Asn), both of maternal origin, encompassing the *CADPS2* gene in individuals with either ASD/ID, and considering that this gene shows a maternal monoallelic expression in blood (see preliminary results, chapter 6), we have tested whether *CADPS2* might be subjected to a parent of origin regulation.

**8.1 Maternal inheritance of the *CADPS2* deletion**

In order to confirm the maternal inheritance of the *CADPS2* intragenic deletion, we performed a microsatellite markers and single nucleotide polymorphisms (SNPs) analysis on blood derived DNA of father, proband and affected sister. The haplotype analysis showed that both the affected siblings inherit two different chromosome by the father, while they share the same maternal chromosome carrying the deletion (**Figure 8.1**).



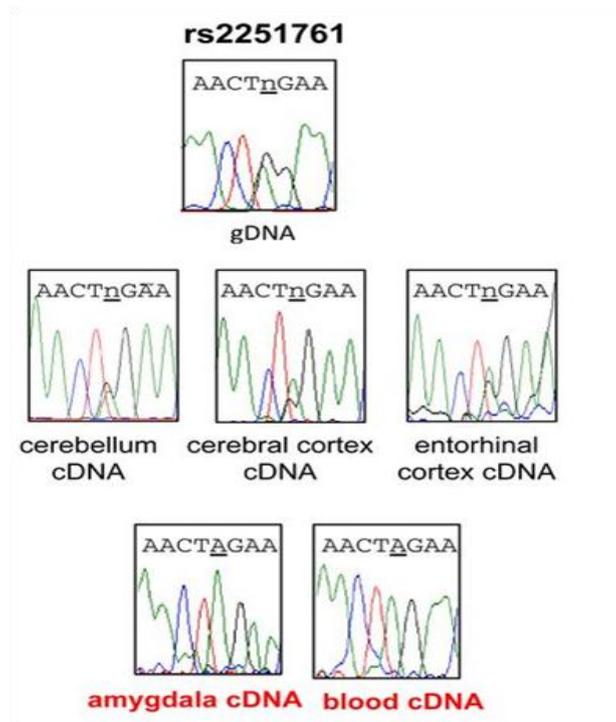
**Figure 8.1** (Bonora et al., 2014). Microsatellite and SNP analysis showing the maternal inheritance of *CADPS2* intragenic deletion. The two microsatellite markers map to *CADPS2* intron 1 and intron 2 (CADPS2\_21xAC; CADPS2\_26xAT), while the two SNPs (rs2074589; rs20745889) map to *CADPS2* exon 17 and intron 21.

**8.2 *CADPS2* allelic expression analysis in brain tissues**

In order to test if *CADPS2* is also monoallelically expressed in brain, we analysed the gene expression in four different human brain regions (amygdala, cerebellum, cerebral cortex and enthorinal cortex) of three informative heterozygous controls for SNP rs2251761. Sequencing of the

## RESULTS

PCR products showed that *CADPS2* is always monoallelically expressed in the amygdala, whereas it is biallelic in the other areas (**Figure 8.2**). *CADPS2* expression analysis in blood cDNA of the same individuals confirmed the monoallelic expression and indicated that the same allele was expressed in both amygdala and blood, suggesting that *CADPS2* might be subjected to tissue-specific monoallelic expression and putative imprinting.

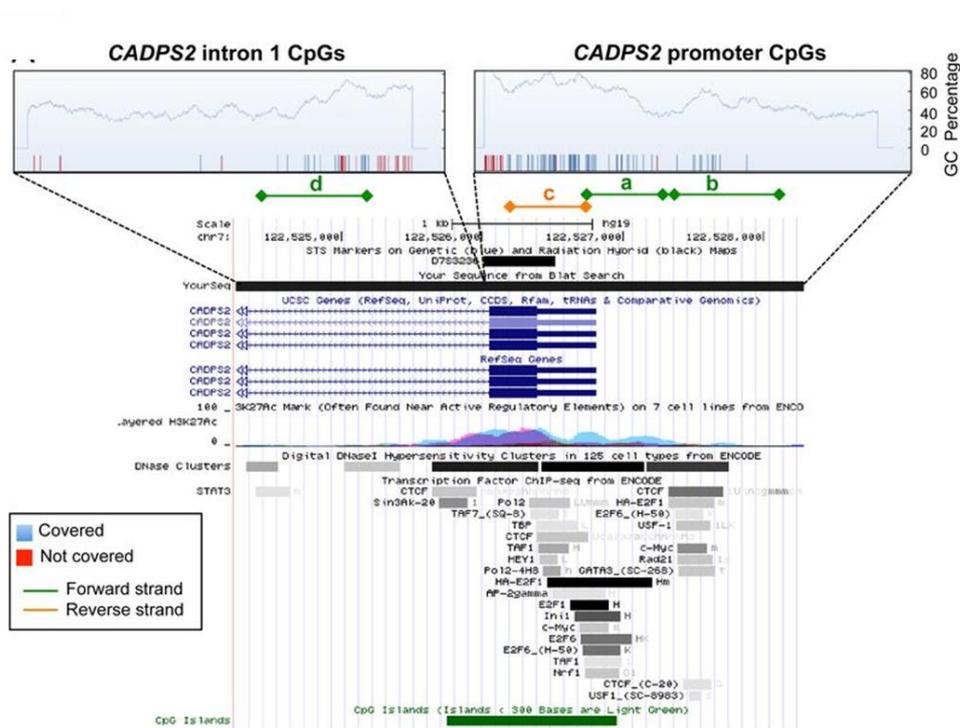


**Figure 8.2** (Bonora et al., 2014). *CADPS2* allelic expression in different human tissues. Electropherograms showing *CADPS2* allelic expression in different brain areas of one control individual heterozygous at SNP rs2251761. In the lower panel, it is shown that the allele expressed by amygdala and blood cDNA is the same. gDNA = genomic DNA.

### 8.3 Quantitative methylation analysis of *CADPS2* CpG regions

In order to analyze the methylation pattern of the *CADPS2* gene we performed a quantitative epigenetic analysis of *CADPS2* CpG regions located in the promoter and first intron, using the Sequenom's mass array system. It uses the EpiTYPER assay for the detection and quantitative analysis of DNA methylation using base-specific cleavage and Mass Spectrometry (paragraph 5.10.3). The EpiDesigner BETA software was used to design primers for PCR amplification of bisulfite-treated DNA. This tool allows to design amplicons which must cover the majority of the CpG sites. Four amplicons (namely a,b,c,d) including 92 CpGs were selected; 63 of these CpGs were suitable for analysis by gene-specific amplification using *in vitro* transcription coupled with mass spectrometry (MS) (**Figure 8.3**).

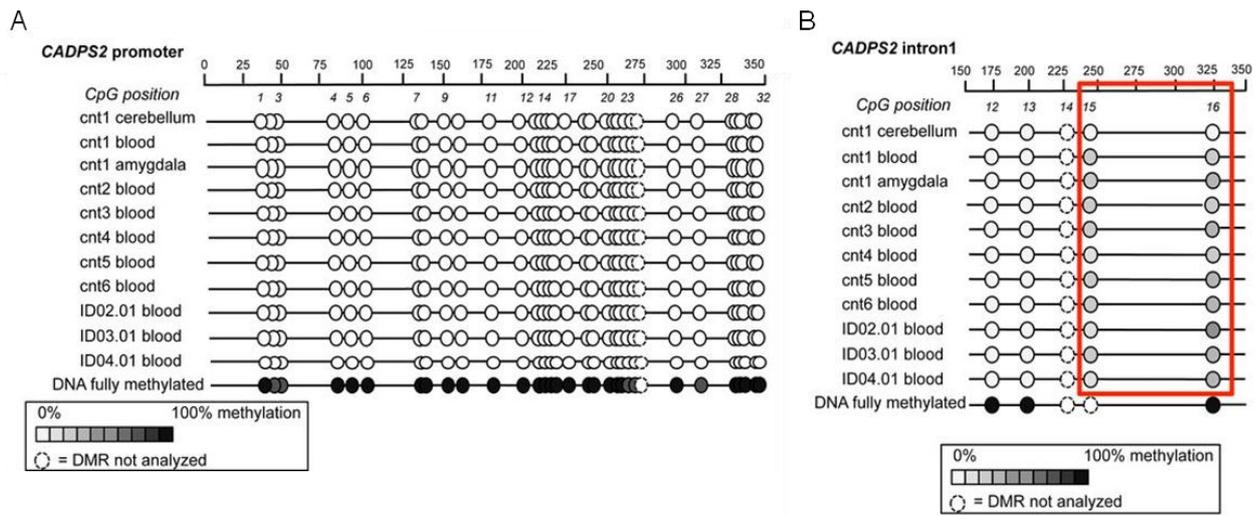
## RESULTS



**Figure 8.3** (Bonora et al., 2014). **Methylation analysis of *CADPS2* promoter and intron 1 CpG regions.** Upper panel: Epidesigner beta output of CpG prediction in the promoter and intron 1 of *CADPS2* genomic region; blue bars = CpG covered by MS analysis, red bars = CpG that cannot be analyzed by MS analysis, (a,b,d) sequences covered in the MS analysis using a forward primer in the *in vitro* transcription, (c) sequence covered by MS analysis using a reverse primer in the *in vitro* transcription; lower panel: transcription factors, CpG island prediction, and regulatory sites in the corresponding genomic region as reported in UCSC Genome Browser (hg19).

Epityper beta output is a graphical representation where a series of individual CpGs are represented by circles on the same line and the color within the circle denotes the level of methylation from 0% to 100%. The numbers indicate the base pairs relative to the amplified PCR product and the position of the CpGs. Analysis was performed on bisulfite-treated DNA from blood, amygdala and cerebellum of 9 different individuals (3 ID and 6 controls individuals). The results showed that the promoter region was unmethylated across all the three regions analyzed (**Figure 8.4A**), whereas in the first intron of *CADPS2* we identified two differentially methylated sites, (CpG<sub>15</sub>, chr7:122,525,608) and (CpG<sub>16</sub>, chr7:122,525,525). For these two sites, cerebellum DNA showed a complete unmethylated status, whereas in blood and amygdala a consistent hemimethylated pattern was observed (**Figure 8.4B**).

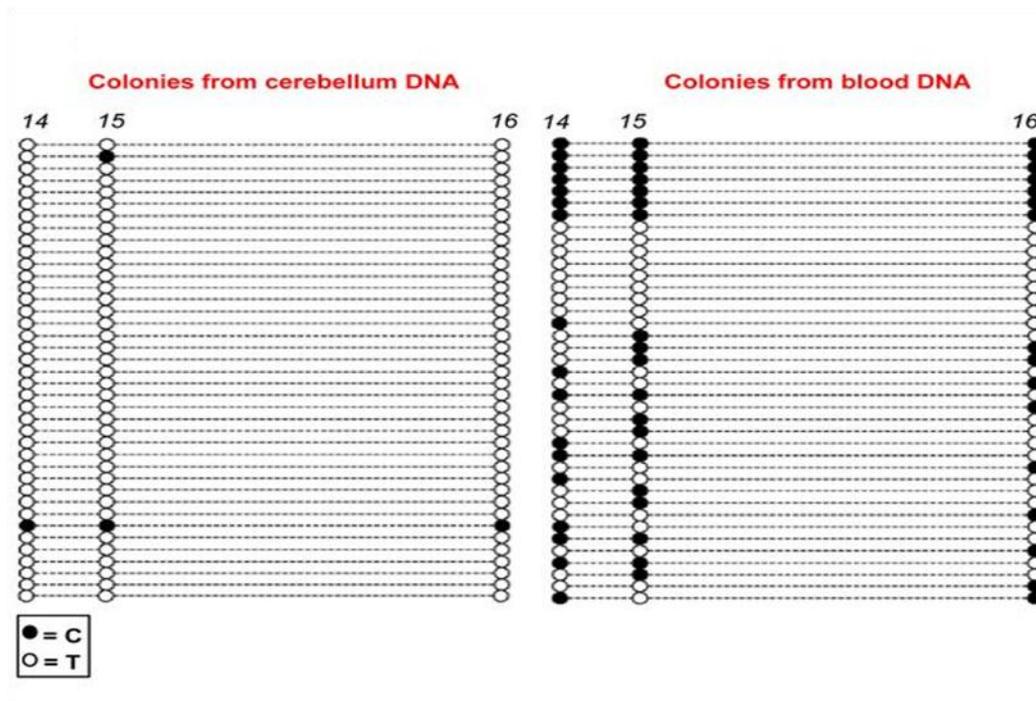
## RESULTS



**Figure 8.4** (Bonora et al., 2014). **Methylation analysis of *CADPS2* promoter and intron 1 CpG regions.** A,B) Epityper beta output of *CADPS2* promoter and intron 1 quantitative methylation analysis via MS in cerebellum, amygdala, and blood genomic DNA bisulfite-treated from different individuals; the percentage of differential methylation is indicated by the different shades of gray as shown in the corresponding box; numbers indicate the base pairs relative to the amplified PCR product (upper lane) and the position of the CpG (lower lane); the differential methylation pattern is appreciable for CpG<sub>15</sub> and CpG<sub>16</sub> in intron 1.

These data were corroborated by cloning the PCR products of intron 1 obtained from bisulfite-treated blood and cerebellum DNA of the same individual (cnt1) and colony screening for the presence of either C (methylated) or T (unmethylated) alleles at the CpGs. This approach led us to analyze also CpG<sub>14</sub> (position chr7:122,525,624) that was not possible to study with MS.

Out of 39 colonies from blood, we found 43.58% methylation at position chr7:122,525,624 (CpG<sub>14</sub>), 48.72% methylation at position chr7:122,525,608 (CpG<sub>15</sub>), and 41.03% methylation allele at position chr7:122,525,525 (CpG<sub>16</sub>), confirming the hemimethylated pattern of the blood. Instead, the percentage of colonies derived from the PCR products of bisulfite-treated cerebellum DNA carrying the methylated allele was 2.56% at bp122,525,624 (CpG<sub>14</sub>), 5.13% at bp122,525,608 (CpG<sub>15</sub>), and 2.56% at bp122,525,525 (CpG<sub>16</sub>), confirming the extensive unmethylation observed by MS analysis (**Figure 8.5**).

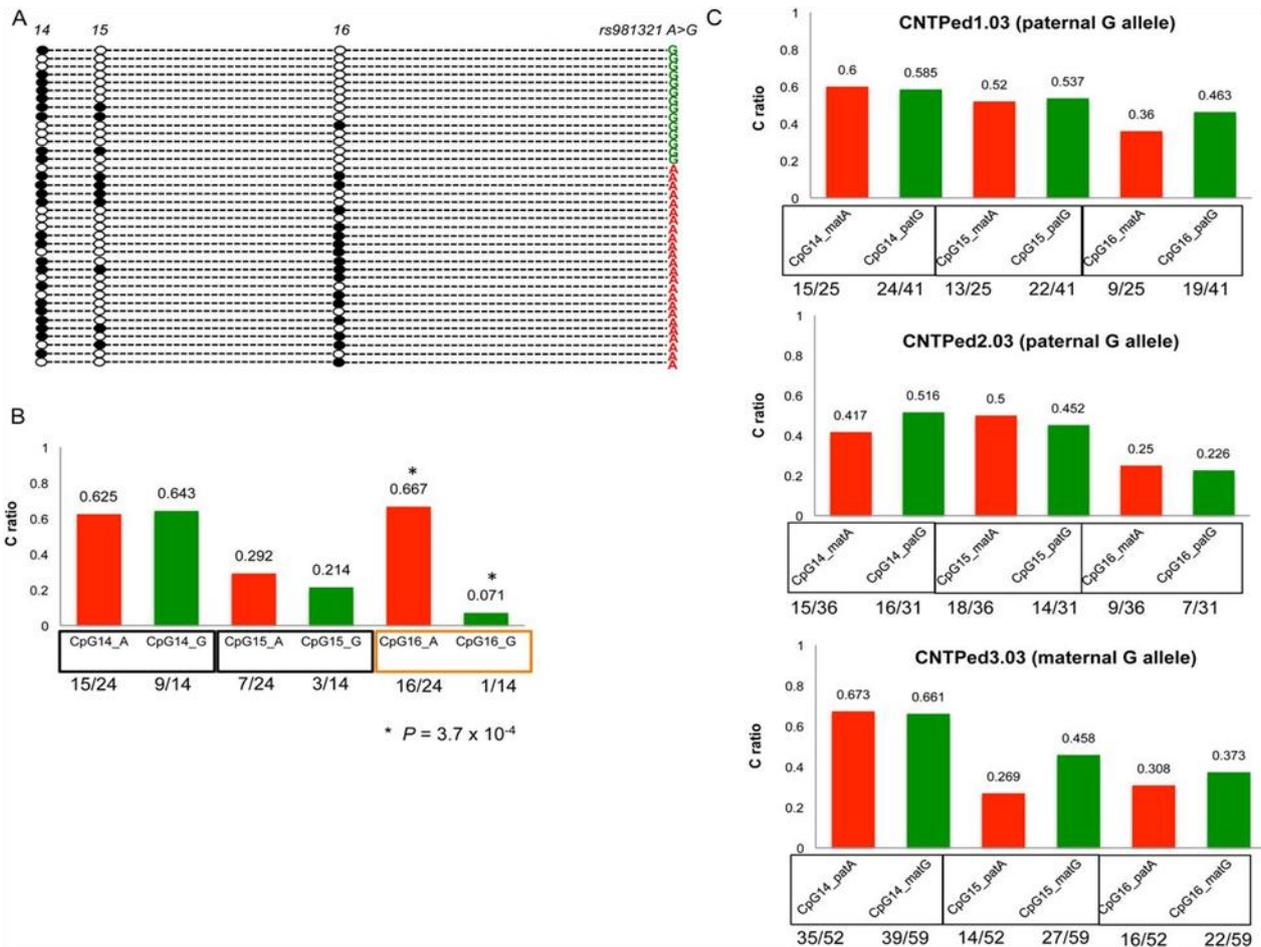


**Figure 8.5** (Bonora et al., 2014). **Methylation analysis of *CADPS2* promoter and intron 1 CpG regions.** Colonies carrying either the T (unmethylated, white circle) or the C (methylated, black circle) alleles at CpG\_14, CpG\_15, and CpG\_16 of *CADPS2* from the colonies obtained by cloning the intron 1 PCR products from bisulfite- treated DNA of cerebellum and blood of individual cnt1.

In order to detect whether these sites (CpG\_14, CpG\_15, CpG\_16) show a parent-of-origin specific differential methylation, we analyzed the methylation status of four control individuals heterozygous for an adjacent SNP (rs981321, g.122,525,329 G>A), by performing colony analysis of an intron 1 fragment containing both the SNP rs981321 and the three CpG sites.

In one control individual, we could detect a statistically significant preferential methylation of the A allele of the SNP rs981321 at CpG\_16, while CpG15 and CpG14 did not show a differential methylation pattern for the two alleles; however, it was not possible to test the parental origin of the two alleles in this individual (**Figure 8.6A,B**). Instead, for the three other heterozygous controls, the parental origin of the alleles was known, but we did not detect a difference in the methylation level at any of the three CpGs (**Figure 8.6C**).

## RESULTS



**Figure 8.6** (Bonora et al., 2014). **Colony screening for parent-of-origin methylation analysis.** **A)** Colonies carrying either the T (unmethylated, white circle) or the C (methylated, black circle) alleles at CpG\_14, CpG\_15, and CpG\_16 and either G (green) or A (red) alleles at SNP rs981321 (A/G in *CADPS2* intron 1 PCR product; sequencing analysis performed on the colonies carrying the PCR product for intron 1 obtained from bisulfite-treated blood DNA of a heterozygous individual for SNP rs981321). **B)** Histogram showing the ratio of colonies with the C methylated allele, for each CpG, for each rs981321 allele. Lower line: number of colonies carrying the C methylated allele out of the total number of colonies with the same allele at rs981321. Significant  $P$ -values are marked with star (Fisher's exact test). **C)** Parental origin of the methylated and unmethylated alleles at CpG\_14, CpG\_15, and CpG\_16 for three individuals for whom parental origin of alleles at rs981321 was known as shown in the figure.

## CHAPTER 9: DISCUSSION

### 9.1 Analysis of rare CNVs implicated in ASD and ID susceptibility

Autism spectrum disorder (ASD) and intellectual disability (ID) are complex neuropsychiatric disorders with an extensive clinic and genetic heterogeneity and a prevalence of 0.7-1.1% (Elsabbagh et al., 2012) and of 1-3% (van Bokhoven, 2011), respectively, in the general population.

Despite high heritability estimates for both ASD and ID, relatively few common risk variants have been convincingly associated with these diseases (Marshall et al., 2008), drawing attention to the contribution of rare genetic variation to susceptibility of complex traits.

Indeed, recent studies have highlighted the involvement of rare (<1% frequency) inherited or *de novo* Copy Number Variants (CNVs) and single nucleotide variants (SNVs) in 5-10% of idiopathic ASD cases and in 10-15% of individuals with idiopathic ID (Devlin & Scherer, 2012). Despite the heterogeneous genetic landscape for both ASD and ID, the genes impacted by rare variants seem to converge in common pathways affecting neuronal and synaptic homeostasis and many susceptibility genes involved in ASD are also involved in intellectual disability, indicating that ID and ASD may be on the same continuum, sharing common aetiologies and showing many genetic similarities (Y. Kou et al., 2012).

My PhD project focused on the further characterization of rare genic CNVs previously identified in individuals with ASD or ID by two International Consortia, the Autism Genome Project (AGP) and the CHERISH project.

### 9.2 Analysis of a compound heterozygous deletion in *CTNNA3*

The compound heterozygous deletion involving the *CTNNA3* gene appeared to be an interesting finding to follow-up as it causes the homozygous loss of a coding exon, predicted to result in a complete lack of *CTNNA3* functional protein.

The role of rare recessive inherited variants in ASD risk has been highlighted in recent studies which applied whole-exome sequencing and homozygosity mapping to consanguineous and/or multiplex families with ASD (Yu et al., 2013) and in a case-control study (Lim et al., 2013), which estimated a 3% contribution to ASD risk from recessive mutations. The proposed involvement of recessive mutations in ASD susceptibility is also in accordance with the high heritability of ASD

(Constantino & Todd, 2005) and with the observation that the majority of parents of ASD individuals are unaffected.

*CTNNA3* represents an interesting candidate gene for ASD based on its biological function. Indeed it encodes  $\alpha$ T-catenin, a member of the  $\alpha$ -catenin family, and it has a crucial role in cell adhesion, one of the major pathways implicated in ASD (Betancur et al., 2009).

Several genetic findings have provided evidence for a potential role of *CTNNA3* in ASD susceptibility. Some studies have reported common single nucleotide polymorphism (SNP) association (Wang et al., 2009; Weiss et al., 2009) and the occurrence of rare CNVs intersecting *CTNNA3* in ASD cases (D. Levy et al., 2011; O'Roak et al., 2012; Prasad et al., 2012). Moreover, in a recent analysis of exon-disrupting CNVs affecting known ASD candidate genes, *CTNNA3* was found borderline enriched in ASD cases in comparison with controls (Girirajan et al., 2012).

Even if we did not observed a statistically significant difference for *CTNNA3* exonic deletions between ASD cases and controls, no homozygous or compound heterozygous exonic deletions were found in a sample of over 6,600 controls, suggesting that *CTNNA3* is haplosufficient and that only a recessive loss of function may play a role in ASD susceptibility. This hypothesis is consistent with the segregation of *CTNNA3* deletions in the discovery family, as the unaffected sister inherited the exonic deletion in the heterozygous form.

Mutation screening of coding region of *CTNNA3* and the nested gene *LRRTM3* in four ASD families carrying exonic *CTNNA3* deletions did not reveal any sequence variants in the non-deleted allele, but we cannot exclude the presence of point mutations in non coding regions disrupting gene regulation or splicing.

Linkage and association studies have indicated that the *CTNNA3/LRRTM3* locus may play a role in susceptibility to late-onset Alzheimer's disease and plasma amyloid  $\beta$  levels (Ertekin-Taner et al., 2003; Martin et al., 2005) but the role of *CTNNA3* in brain remains largely unexplored. Western blot characterisation of mouse *Ctnna3* expression, have suggested that *Ctnna3* has a specific neuronal function in very early developmental stages, as it displays higher expression in the hippocampus and cortex at P0. Interestingly other functional studies have revealed that, similarly to other  $\alpha$ -catenins, overexpression of *Ctnna3* in dendrites causes an increase in spine and synapse density, suggesting that all  $\alpha$ -catenin isoforms share the same spine-stabilizing activity (Abe et al., 2004). The role of  $\alpha$ T-catenin has been primarily investigated in the heart, due to its high expression in cardiac tissue and co-localization with plakophilin 2 (Goossens et al., 2007). An experiment performed on a *CTNNA3* loss of function murine model has revealed that the loss of  $\alpha$ T-catenin alters the adhering junctions in the heart and leads to dilated cardiomyopathy and to increased risk

of ventricular arrhythmia following acute ischemia (Li et al., 2012). Notably, the affected proband 3456\_3 with the compound heterozygous *CTNNA3* deletion had a normal EKG and heart ultrasound, suggesting that *CTNNA3* loss of function does not lead to cardiac dysfunction in human. Another feature that makes *CTNNA3* particularly interesting is the presence of the nested gene *LRRTM3*, a member of leucine-rich repeat transmembrane neuronal (LRRTM) gene family, which are transmembrane proteins highly expressed in brain and involved in synaptic organization during synapse development (Linhoff et al., 2009). *LRRTM3* displays expression predominantly in the hippocampus and, similar to *CTNNA3*, plays a role in cell adhesion (Laurén et al., 2003). Since all three  $\alpha$ -catenins harbour a *LRRTM* gene within their largest intron, it is possible that  $\alpha$ -catenin genes and the respective *LRRTMs* share common transcriptional regulation mechanisms. Therefore, it would be interesting to investigate if deletions in *CTNNA3* could cause dysregulation of *LRRTM3* expression.

In conclusion, our findings point to involvement of  $\alpha$ -catenins in ASD pathogenesis, given their crucial role in both the development and maintenance of the nervous system. Specifically, the recessive inheritance pattern seen in the discovery pedigree, together with the absence of exonic homozygous microdeletions in *CTNNA3* in over 6,600 control subjects and the high expression of *Ctnna3* in developing mouse brain, led us to hypothesize that the identified compound heterozygous exonic deletion in *CTNNA3* contributes to ASD phenotype in family 3456.

In order to understand the potential role for recessive *CTNNA3* defects in other neuropsychiatric disorders, alone or co-occurring with ASD, we are currently sequencing *CTNNA3* coding region in two subjects, one with early infantile epileptic encephalopathy, developmental delay and ASD and the other one with language delay, learning disability associated to dysmorphic features, both of them carry two different *CTNNA3* exonic deletions.

Further studies in larger cohort with point or structural mutations affecting  $\alpha$ -catenin's function will be very useful to elucidate their contribution to ASD and other neuropsychiatric disorders pathogenesis.

### **9.3 Analysis of a microduplication in the *CHRNA7* gene**

The identification of a small 15q13.3 duplication, which involves the entire *CHRNA7* gene, encoding the alpha7 subunit of the neuronal nicotinic acetylcholine receptor, was of particular interest as *CHRNA7* is considered the likely culprit gene in mediating neurological phenotypes in patients with 15q13.3 recurrent CNVs.

The 15q13.3 microdeletion is highly but not always fully penetrant, and it is significantly enriched in cases of intellectual disability, autism, epilepsy, schizophrenia, and bipolar disorder (Pagnamenta et al., 2009; Sharp et al., 2008; Stefansson et al., 2008; van Bon et al., 2009). Instead, the significance of the 15q13.3 reciprocal microduplications has been more challenging to interpret, given that they have been detected across the same spectrum of neuropsychiatric disorders of the microdeletions, but with high variability in expressivity and reduced penetrance compared with deletions.

In order to capture the entire spectrum of genetic variation in *CHRNA7* contributing to ASD risk, it is essential to integrate both CNV and sequence data. However, given the existence of a fusion gene, *CHRFAM7A*, which includes a nearly identical partial duplication of *CHRNA7*, sequence variation in *CHRNA7* remains largely unexplored. Hence, attempts to sequence coding exons must distinguish between *CHRNA7* and *CHRFAM7A*, making next-generation sequencing approaches unreliable for this purpose.

No additional duplications or deletions were detected in *CHRNA7* by our CNV analysis, while mutation screening led to the identification of three rare variants located in the proximal promoter region. Notably, one of these promoter mutations (-241 bp from ATG) was found in one ASD individual (3377\_3) who also carries a more frequent 5'-UTR variant (-86 bp from ATG) on the other chromosome. Both variants have been previously associated with *CHRNA7* decreased transcription in vitro (S. Leonard & Freedman, 2006; S. Leonard et al., 2002) suggesting that the presence of the compound variants (-241 bp and -86 bp from ATG) in proband 3377\_3 might lead to a marked reduction of *CHRNA7* expression. Unfortunately, it was not possible to test the functional effect of the -86/-241 variants variant in this ASD subject, as *CHRNA7* expression is too low in blood to be assessed by RT-PCR.

Interestingly, the subject with the *CHRNA7* duplication developed complex partial seizures with secondary generalization, at age of 13 years. However, while epilepsy has been strongly associated with microdeletions of *CHRNA7* (Helbig et al., 2009), seizures are not reported as a common features of patients carrying the microduplication (Szafranski et al., 2010). However, it is not possible to exclude that some participants of pediatric cohorts may develop epilepsy at a later time.

To better understand the complex genotype-phenotype correlations of the reciprocal microdeletions and microduplications, a detailed clinical characterization of the individuals carrying the CNV would be very useful, especially if followed up over time. The co-occurrence of ASD and epilepsy in the proband with the *CHRNA7* duplication, may suggest that the microduplication involving *CHRNA7* could have the same role of the deletion in ASD/epilepsy susceptibility although with

lower penetrance. In accordance with this hypothesis, *CHRNA7* transcript levels were recently found reduced in neuronal cells (Meguro-Horike et al., 2011) or brain samples with maternal 15q duplication, in contrast to what is expected according to the gene copy number.

The observation that deletions and duplications at the same locus may yield similar phenotypes is quite common and it could be explained by the sensitivity of certain cellular functions to dosage imbalance, as described for the 1q21.1 region (Harvard et al., 2011).

The phenotypic variability at locus 15q13.3 may also be controlled by second-site CNVs, in line with the recently proposed “two-hit model” for severe developmental delay (Girirajan et al., 2010). Support to this model comes from a recent study where five out of 11 patients with small microduplications involving *CHRNA7* and showing a variety of neuropsychiatric disorders, carried at least one additional different CNV of potential clinical relevance (Szafranski et al., 2010).

In addition, two ASD individuals carrying a small *CHRNA7* duplication and a *de novo* *SHANK2* deletion on distinct parental chromosomes, were recently observed, suggesting the presence of epistasis between these two loci (Leblond et al., 2012). Another report described a boy with severe ID, language impairment, and behavioral anomalies, carrying a *de novo* balanced translocation disrupting the *SHANK2* gene as well as an inherited duplication of *CHRNA7* (Chilian et al., 2013).

In our study, no other clearly pathogenic CNV has been identified in proband 3474\_3, and no *CHRNA7* sequence variants have been detected, even if we cannot exclude the presence of a sequence variant elsewhere in the genome that could act within the same pathway to increase the risk of ASD.

In conclusion, our results do not support the hypothesis that rare sequence variants in *CHRNA7* significantly contribute to ASD susceptibility in our ASD cohort, which is characterized by a low degree of medical comorbidities, e.g., the frequency of epilepsy in our sample is only 1.5%, while it is commonly reported to occur in 5% to 46% of individuals with ASD (Viscidi et al., 2013).

Therefore, analysis of much larger cohorts of individuals is thus warranted to elucidate the role of rare *CHRNA7* sequence variants in ASD risk and to discriminate if *CHRNA7* might be mainly implicated in ASD cases associated to other clinical features that would be consistent with the significant enrichment of 15q13.3 CNVs observed in individuals with comorbid phenotypes.

Moreover, it would be interesting to extend the analysis of the *CHRNA7* promoter to the distal upstream regulatory region, which contains SNP rs3087454 (-1831bp), reported to be associated with schizophrenia (Stephens et al., 2009).

We are also currently investigating the role of the chimeric *CHRFAM7A* gene product (*dupa7*) in the ASD phenotype. Recent data have suggested that *dupa7* acts as a dominant negative regulator of

*CHRNA7* function, as it assembles with  $\alpha 7$  subunits and causes a decrease of acetylcholine-stimulated current. Furthermore, the *CHRFAM7A $\Delta$ 2bp* gene product (dup $\alpha\Delta 2$ ), containing a 2bp deletion in exon 6, is a more potent inhibitor in comparison with the wild-type dup $\alpha 7$  (Araud et al., 2011) and interestingly, this 2 bp deletion was significantly associated with schizophrenia (Sinkus et al., 2009). Therefore, we plan to assess the *CHRFAM7A* copy number and the presence of the 2 bp deletion in order to evaluate *CHRNA7* function in human diseases.

#### **9.4 Analysis of a rare microdeletion on chromosome 2q31.1**

The rare chromosome 2q31.1 microdeletion was identified in a subject with ASD and borderline cognitive impairments and it encompasses three genes: the last exon of *KLHL23*, the entire *SSB* gene and the last two exons of *METTL5*.

*SSB* (Sjogren syndrome antigen B) encodes La protein, that is involved in different aspects of RNA metabolism, including binding and protecting poly(U) termini of nascent RNA polymerase III transcripts from exonuclease digestion (Teplova et al., 2006). Autoantibodies reacting with this protein are found in the sera of patients with Sjogren syndrome and systemic lupus erythematosus.

The *METTL5* gene (methyltransferase like 5) encodes a protein of unknown function, probably belonging to the methyltransferase superfamily. The methyltransferases catalyze the transfer of a methyl group to diverse substrates, including nucleic acids (DNA and RNA), proteins and lipids, and they essentially influence multiple cellular regulatory mechanisms by modifying their targets.

*KLHL23* is a member of the kelch-like (KLHL) gene family, whose specific role has not yet been elucidated. The KLHL family is conserved throughout evolution and consists of proteins with a variety of functions, including actin binding, cytoskeletal organization and ubiquitination (Adams, Kelso, & Cooley, 2000). In general, KLHL proteins have similar structural motifs as they contain one BTB/POZ domain, one BACK domain, and five to six Kelch motifs.

It is not straightforward to distinguish which of the genes in the deleted region, alone or in combination, contribute to the disorder. To this aim we searched the available databases for the presence of additional deletions overlapping this locus, which might help to restrict the region of interest.

In the Simons Foundation Autism Research Initiative (SFARI) AutDB database (Basu et al., 2009), we identified an intragenic deletion in *PHOSPHO2-KLHL23* (Prasad et al., 2012), while there are no CNVs affecting *SSB* and/or *METTL5*, thus suggesting a causative association between disruption of *KLHL23* and ASD.

The identification of two rare damaging missense variants (p.Met65Val and p.Thr522Met) in two unrelated ASD probands, affecting two conserved domains (the BTB domain and the last Kelch-repeat, respectively), provides another line of evidence for a role of *KLHL23* in ASD susceptibility. In conclusion, we propose that *KLHL23* might constitute a new ASD gene, even if confirmation in large ASD cohorts is needed to implicate it with certainty; further studies are also warranted to understand its biological function and its role in the developing brain.

### **9.5 Analysis of *CADPS2* in individuals with ASD and ID**

My role on this topic was to further investigate the hypothesis that *CADPS2* might be subjected to imprinting.

Previous results have demonstrated that *CADPS2* is monoallelically expressed in blood. My first task was to investigate *CADPS2* expression in different brain regions. This analysis revealed that the gene is monoallelically expressed in blood and amygdala, and the expressed allele is the maternal one.

Subsequently, I carried out a quantitative epigenetic analysis of *CADPS2* CpGs regions. This led to the identification of a cluster of differentially methylated CpGs sites in the first intron of the gene in bisulfate treated DNA from the blood and amygdala, while a differential methylation in this region was not observed in the cerebellum, in agreement with the biallelic expression of the gene in this brain area.

The regulation of *CADPS2* expression via monoallelic expression in the amygdala, which plays a key role in regulating social interactions, supports the importance of a fine modulation of *CADPS2* for human behavior. Although the function of *CADPS2* in amygdala has not yet been elucidated, recent findings have shown that human *CADPS2* expression in amygdala is lower in the prenatal period and starts to increase in late fetal stage until mid-childhood (H. J. Kang et al., 2011). Moreover, given that amygdala is known to be involved in attention, perception and in memory formation, it is not surprising that amygdala dysfunction has been reported in various psychiatric conditions such as schizophrenia, autism and anxiety (LeDoux, 2007).

In order to detect whether the CpG sites identified in *CADPS2* first intron show a parent-of-origin-specific differential methylation, I analysed the methylation status of four control individuals heterozygous for a SNP near to these sites. The analysis did not reveal a reproducible parent-of-origin methylation profile for *CADPS2*, suggesting that other genomic regions in this gene or other regulatory mechanisms may be related to the presence of a monoallelic pattern of expression.

Therefore, further gene expression and methylation analyses of *CADPS2* genomic region in diverse brain areas and in larger samples are warranted in order to elucidate the regulation pattern of this gene and to identify additional maternal variants, which may contribute to the ID/ASD phenotype.

### CONCLUSIONS

Taken together these results confirm the striking complexity of pathophysiology of ASD and Intellectual Disability and the role of rare genetic variants, both CNVs and point mutations, in the etiology of these complex disorders.

Specifically, our findings further support the observation that rare genetic variation contributes to disease risk in different ways: some rare mutations, such as those impacting the *CTNNA3* gene, act in a recessive mode of inheritance, while other chromosomal rearrangements, such as those occurring in the 15q13.3 region, are recurrent CNVs, implicated in multiple developmental and/or neurological disorders, which could contribute to overall risk, possibly interacting with other susceptibility variants elsewhere in the genome. On the other hand, the discovery of a tissue-specific monoallelic expression for the *CADPS2* gene, implicates the involvement of epigenetic regulatory mechanisms as risk factors conferring susceptibility to ASD/ID.

Considerable progresses in the genetics of ASD and ID have been made with the recent introduction of Next-Generation Sequencing (NGS) technologies, which have allowed to accelerate the identification of rare risk-conferring variation.

Whole exome sequencing (WES) studies in large cohorts of ASD/ID patients, provided further evidence for a contribution of rare inherited and *de novo* variants to heritability of these complex traits. More recently, whole-genome sequencing (WGS) are becoming a powerful method to assess genetic variation, as they allow the detection of all classes and size of mutations. Although the largest WGS study performed to date in 170 ASD individuals (Yuen et al., 2015) had highlighted the genetic heterogeneity underlying ASD, both between and within the multiplex families analyzed, the data emerged from this analysis represent an important first step in a much large initiative to sequence the genome of about one thousand of other ASD families.

Although many rare variants are predicted to disrupt the gene function, only few mutations are expected to be causative, as most of them have an unknown clinical significance and therefore it is essential to distinguish disease-causing variants from false-positive findings. Therefore the greatest challenge in the next few years will be to develop bioinformatic methods, analytical strategies and tools to interpret the significance of the huge amount of genomic data stemming from NGS studies.

## DISCUSSION

---

Multidisciplinary approaches will also be crucial to enhance our knowledge on the biology of neurodevelopmental disorders, including deep phenotypic characterization of cases, the development of animal models, pluripotent stem cells (iPSCs) derived from patients and epigenetic studies.

Additionally, since neurodevelopmental disorders, such as ASD and ID, share many risk factors and they often co-exist in the same individual, the analysis of large cohorts of individuals with different co-morbidities might be very useful to better understand the mechanisms by which specific molecular pathways result in common vs. different developmental outcomes. Understanding this is essential to identify targets for novel aetiology-based treatments.

---

## References

- APA, A. P. A., 2013, Diagnostic and statistical manual of mental disorders (5th ed.), Arlington, VA: American Psychiatric Publishing.
- (IMGSAC), I. M. G. S. A. C. (2001). A genomewide screen for autism: strong evidence for linkage to chromosomes 2q, 7q, and 16p. *Am J Hum Genet*, 69(3), 570-581. doi: 10.1086/323264
- Abe, K., Chisaka, O., Van Roy, F., & Takeichi, M. (2004). Stability of dendritic spines and synaptic contacts is controlled by alpha N-catenin. *Nat Neurosci*, 7(4), 357-363. doi: 10.1038/nn1212
- Adams, J., Kelso, R., & Cooley, L. (2000). The kelch repeat superfamily of proteins: propellers of cell function. *Trends Cell Biol*, 10(1), 17-24.
- Adegbola, A., Gao, H., Sommer, S., & Browning, M. (2008). A novel mutation in JARID1C/SMCX in a patient with autism spectrum disorder (ASD). *Am J Med Genet A*, 146A(4), 505-511. doi: 10.1002/ajmg.a.32142
- Adzhubei, I. A., Schmidt, S., Peshkin, L., Ramensky, V. E., Gerasimova, A., Bork, P., . . . Sunyaev, S. R. (2010). A method and server for predicting damaging missense mutations. *Nat Methods*, 7(4), 248-249. doi: 10.1038/nmeth0410-248
- Aicardi, J. (1998). The etiology of developmental delay. *Semin Pediatr Neurol*, 5(1), 15-20.
- Albagli, O., Dhordain, P., Deweindt, C., Lecocq, G., & Leprince, D. (1995). The BTB/POZ domain: a new protein-protein interaction motif common to DNA- and actin-binding proteins. *Cell Growth Differ*, 6(9), 1193-1198.
- Albuquerque, E. X., Pereira, E. F., Alkondon, M., & Rogers, S. W. (2009). Mammalian nicotinic acetylcholine receptors: from structure to function. *Physiol Rev*, 89(1), 73-120. doi: 10.1152/physrev.00015.2008
- Amir, R. E., Van den Veyver, I. B., Wan, M., Tran, C. Q., Francke, U., & Zoghbi, H. Y. (1999). Rett syndrome is caused by mutations in X-linked MECP2, encoding methyl-CpG-binding protein 2. *Nat Genet*, 23(2), 185-188. doi: 10.1038/13810
- Amos-Landgraf, J. M., Ji, Y., Gottlieb, W., Depinet, T., Wandstrat, A. E., Cassidy, S. B., . . . Nicholls, R. D. (1999). Chromosome breakage in the Prader-Willi and Angelman syndromes involves recombination between large, transcribed repeats at proximal and distal breakpoints. *Am J Hum Genet*, 65(2), 370-386. doi: 10.1086/302510
- Anney, R., Klei, L., Pinto, D., Almeida, J., Bacchelli, E., Baird, G., . . . Devlin, B. (2012). Individual common variants exert weak effects on the risk for autism spectrum disorders. *Hum Mol Genet*, 21(21), 4781-4792. doi: 10.1093/hmg/ddc301
- Anney, R., Klei, L., Pinto, D., Regan, R., Conroy, J., Magalhaes, T. R., . . . Hallmayer, J. (2010). A genome-wide scan for common alleles affecting risk for autism. *Hum Mol Genet*, 19(20), 4072-4082. doi: 10.1093/hmg/ddq307
- Araud, T., Graw, S., Berger, R., Lee, M., Neveu, E., Bertrand, D., & Leonard, S. (2011). The chimeric gene CHRFAM7A, a partial duplication of the CHRNA7 gene, is a dominant negative regulator of  $\alpha 7$ \*nAChR function. *Biochem Pharmacol*, 82(8), 904-914. doi: 10.1016/j.bcp.2011.06.018
- Araya, N., Arimura, H., Kawahara, K., Yagishita, N., Ishida, J., Fujii, R., . . . Nakajima, T. (2009). Role of Kenae/CCDC125 in cell motility through the deregulation of RhoGTPase. *Int J Mol Med*, 24(5), 605-611.
- Arikkath, J., & Reichardt, L. F. (2008). Cadherins and catenins at synapses: roles in synaptogenesis and synaptic plasticity. *Trends Neurosci*, 31(9), 487-494. doi: 10.1016/j.tins.2008.07.001
- Bacchelli, E., Battaglia, A., Cameli, C., Lomartire, S., Tancredi, R., Thomson, S., . . . Maestrini, E. (2015). Analysis of CHRNA7 rare variants in autism spectrum disorder susceptibility. *Am J Med Genet A*. doi: 10.1002/ajmg.a.36847

- Bacchelli, E., Ceroni, F., Pinto, D., Lomartire, S., Giannandrea, M., D'Adamo, P., . . . Maestrini, E. (2014). A CTNNA3 compound heterozygous deletion implicates a role for  $\alpha$ T-catenin in susceptibility to autism spectrum disorder. *J Neurodev Disord*, *6*(1), 17. doi: 10.1186/1866-1955-6-17
- Badner, J. A., & Gershon, E. S. (2002). Regional meta-analysis of published data supports linkage of autism with markers on chromosome 7. *Mol Psychiatry*, *7*(1), 56-66. doi: 10.1038/sj/mp/4000922
- Baer, K., Waldvogel, H. J., Faull, R. L., & Rees, M. I. (2009). Localization of glycine receptors in the human forebrain, brainstem, and cervical spinal cord: an immunohistochemical review. *Front Mol Neurosci*, *2*, 25. doi: 10.3389/neuro.02.025.2009
- Bagni, C., & Greenough, W. T. (2005). From mRNP trafficking to spine dysmorphogenesis: the roots of fragile X syndrome. *Nat Rev Neurosci*, *6*(5), 376-387. doi: 10.1038/nrn1667
- Bahl, S., Chiang, C., Beauchamp, R. L., Neale, B. M., Daly, M. J., Gusella, J. F., . . . Ramesh, V. (2013). Lack of association of rare functional variants in TSC1/TSC2 genes with autism spectrum disorder. *Mol Autism*, *4*(1), 5. doi: 10.1186/2040-2392-4-5
- Bailey, A., Le Couteur, A., Gottesman, I., Bolton, P., Simonoff, E., Yuzda, E., & Rutter, M. (1995). Autism as a strongly genetic disorder: evidence from a British twin study. *Psychol Med*, *25*(1), 63-77.
- Baker, P., Piven, J., Schwartz, S., & Patil, S. (1994). Brief report: duplication of chromosome 15q11-13 in two individuals with autistic disorder. *J Autism Dev Disord*, *24*(4), 529-535.
- Bakkaloglu, B., O'Roak, B. J., Louvi, A., Gupta, A. R., Abelson, J. F., Morgan, T. M., . . . State, M. W. (2008). Molecular cytogenetic analysis and resequencing of contactin associated protein-like 2 in autism spectrum disorders. *Am J Hum Genet*, *82*(1), 165-173. doi: 10.1016/j.ajhg.2007.09.017
- Bardwell, V. J., & Treisman, R. (1994). The POZ domain: a conserved protein-protein interaction motif. *Genes Dev*, *8*(14), 1664-1677.
- Bashi, J. (1977). Effects of inbreeding on cognitive performance. *Nature*, *266*(5601), 440-442.
- Basu, S. N., Kollu, R., & Banerjee-Basu, S. (2009). AutDB: a gene reference resource for autism research. *Nucleic Acids Res*, *37*(Database issue), D832-836. doi: 10.1093/nar/gkn835
- Ben-Shachar, S., Lanpher, B., German, J. R., Qasaymeh, M., Potocki, L., Nagamani, S. C., . . . Sahoo, T. (2009). Microdeletion 15q13.3: a locus with incomplete penetrance for autism, mental retardation, and psychiatric disorders. *J Med Genet*, *46*(6), 382-388. doi: 10.1136/jmg.2008.064378
- Benfenati, F. (2007). Synaptic plasticity and the neurobiology of learning and memory. *Acta Biomed*, *78 Suppl 1*, 58-66.
- Bentley, D. R., Balasubramanian, S., Swerdlow, H. P., Smith, G. P., Milton, J., Brown, C. G., . . . Smith, A. J. (2008). Accurate whole human genome sequencing using reversible terminator chemistry. *Nature*, *456*(7218), 53-59. doi: 10.1038/nature07517
- Berkel, S., Marshall, C. R., Weiss, B., Howe, J., Roeth, R., Moog, U., . . . Rappold, G. A. (2010). Mutations in the SHANK2 synaptic scaffolding gene in autism spectrum disorder and mental retardation. *Nat Genet*, *42*(6), 489-491. doi: 10.1038/ng.589
- Berwin, B., Floor, E., & Martin, T. F. (1998). CAPS (mammalian UNC-31) protein localizes to membranes involved in dense-core vesicle exocytosis. *Neuron*, *21*(1), 137-145.
- Betancur, C. (2011). Etiological heterogeneity in autism spectrum disorders: more than 100 genetic and genomic disorders and still counting. *Brain Res*, *1380*, 42-77. doi: 10.1016/j.brainres.2010.11.078
- Betancur, C., Sakurai, T., & Buxbaum, J. D. (2009). The emerging role of synaptic cell-adhesion pathways in the pathogenesis of autism spectrum disorders. *Trends Neurosci*, *32*(7), 402-412. doi: 10.1016/j.tins.2009.04.003

- Bhalla, K., Luo, Y., Buchan, T., Beachem, M. A., Guzauskas, G. F., Ladd, S., . . . Srivastava, A. K. (2008). Alterations in CDH15 and KIRREL3 in patients with mild to severe intellectual disability. *Am J Hum Genet*, *83*(6), 703-713. doi: 10.1016/j.ajhg.2008.10.020
- Bierut, L. J., Agrawal, A., Bucholz, K. K., Doheny, K. F., Laurie, C., Pugh, E., . . . Gene, E. v. A. S. C. (2010). A genome-wide association study of alcohol dependence. *Proc Natl Acad Sci U S A*, *107*(11), 5082-5087. doi: 10.1073/pnas.0911109107
- Bill, B. R., & Geschwind, D. H. (2009). Genetic advances in autism: heterogeneity and convergence on shared pathways. *Curr Opin Genet Dev*, *19*(3), 271-278. doi: 10.1016/j.gde.2009.04.004
- Billuart, P., Bienvenu, T., Ronce, N., des Portes, V., Vinet, M. C., Zemni, R., . . . Chelly, J. (1998). Oligophrenin-1 encodes a rhoGAP protein involved in X-linked mental retardation. *Nature*, *392*(6679), 923-926. doi: 10.1038/31940
- Bomont, P., Cavalier, L., Blondeau, F., Ben Hamida, C., Belal, S., Tazir, M., . . . Koenig, M. (2000). The gene encoding gigaxonin, a new member of the cytoskeletal BTB/kelch repeat family, is mutated in giant axonal neuropathy. *Nat Genet*, *26*(3), 370-374. doi: 10.1038/81701
- Bonora, E., Graziano, C., Minopoli, F., Bacchelli, E., Magini, P., Diquigiovanni, C., . . . IMGSAC. (2014). Maternally inherited genetic variants of CADPS2 are present in autism spectrum disorders and intellectual disability patients. *EMBO Mol Med*, *6*(6), 795-809. doi: 10.1002/emmm.201303235
- Bradley, E., & Bolton, P. (2006). Episodic psychiatric disorders in teenagers with learning disabilities with and without autism. *Br J Psychiatry*, *189*, 361-366. doi: 10.1192/bjp.bp.105.018127
- Bucan, M., Abrahams, B. S., Wang, K., Glessner, J. T., Herman, E. I., Sonnenblick, L. I., . . . Hakonarson, H. (2009). Genome-wide analyses of exonic copy number variants in a family-based study point to novel autism susceptibility genes. *PLoS Genet*, *5*(6), e1000536. doi: 10.1371/journal.pgen.1000536
- Busby, V., Goossens, S., Nowotny, P., Hamilton, G., Smemo, S., Harold, D., . . . Lovestone, S. (2004). Alpha-T-catenin is expressed in human brain and interacts with the Wnt signaling pathway but is not responsible for linkage to chromosome 10 in Alzheimer's disease. *Neuromolecular Med*, *5*(2), 133-146. doi: 10.1385/NMM:5:2:133
- Buxbaum, J. D., Silverman, J. M., Smith, C. J., Kilifarski, M., Reichert, J., Hollander, E., . . . Davis, K. L. (2001). Evidence for a susceptibility gene for autism on chromosome 2 and for genetic heterogeneity. *Am J Hum Genet*, *68*(6), 1514-1520. doi: 10.1086/320588
- Carter, M. T., Nikkel, S. M., Fernandez, B. A., Marshall, C. R., Noor, A., Lionel, A. C., . . . Scherer, S. W. (2011). Hemizygous deletions on chromosome 1p21.3 involving the DPYD gene in individuals with autism spectrum disorder. *Clin Genet*, *80*(5), 435-443. doi: 10.1111/j.1399-0004.2010.01578.x
- Carvalho, C. M., Ramocki, M. B., Pehlivan, D., Franco, L. M., Gonzaga-Jauregui, C., Fang, P., . . . Lupski, J. R. (2011). Inverted genomic segments and complex triplication rearrangements are mediated by inverted repeats in the human genome. *Nat Genet*, *43*(11), 1074-1081. doi: 10.1038/ng.944
- Chahrouh, M., & Zoghbi, H. Y. (2007). The story of Rett syndrome: from clinic to neurobiology. *Neuron*, *56*(3), 422-437. doi: 10.1016/j.neuron.2007.10.001
- Charbonnier, F., Raux, G., Wang, Q., Drouot, N., Cordier, F., Limacher, J. M., . . . Frebourg, T. (2000). 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*, *60*(11), 2760-2763.

- Chaste, P., Betancur, C., Gérard-Blanluet, M., Bargiacchi, A., Kuzbari, S., Drunat, S., . . . Delorme, R. (2012). High-functioning autism spectrum disorder and fragile X syndrome: report of two affected sisters. *Mol Autism*, *3*(1), 5. doi: 10.1186/2040-2392-3-5
- Cherry, K. E., Matson, J. L., & Paclawskyj, T. R. (1997). Psychopathology in older adults with severe and profound mental retardation. *Am J Ment Retard*, *101*(5), 445-458.
- Chilian, B., Abdollahpour, H., Bierhals, T., Haltrich, I., Fekete, G., Nagel, I., . . . Kutsche, K. (2013). Dysfunction of SHANK2 and CHRNA7 in a patient with intellectual disability and language impairment supports genetic epistasis of the two loci. *Clin Genet*, *84*(6), 560-565. doi: 10.1111/cge.12105
- Ching, M. S., Shen, Y., Tan, W. H., Jeste, S. S., Morrow, E. M., Chen, X., . . . Group, C. s. H. B. G. P. S. (2010). Deletions of NRXN1 (neurexin-1) predispose to a wide spectrum of developmental disorders. *Am J Med Genet B Neuropsychiatr Genet*, *153B*(4), 937-947. doi: 10.1002/ajmg.b.31063
- Choi, M., Scholl, U. I., Ji, W., Liu, T., Tikhonova, I. R., Zumbo, P., . . . Lifton, R. P. (2009). Genetic diagnosis by whole exome capture and massively parallel DNA sequencing. *Proc Natl Acad Sci U S A*, *106*(45), 19096-19101. doi: 10.1073/pnas.0910672106
- Christian, S. L., Brune, C. W., Sudi, J., Kumar, R. A., Liu, S., Karamohamed, S., . . . Cook, E. H. (2008). Novel submicroscopic chromosomal abnormalities detected in autism spectrum disorder. *Biol Psychiatry*, *63*(12), 1111-1117. doi: 10.1016/j.biopsych.2008.01.009
- Cirak, S., von Deimling, F., Sachdev, S., Errington, W. J., Herrmann, R., Bönnemann, C., . . . Voit, T. (2010). Kelch-like homologue 9 mutation is associated with an early onset autosomal dominant distal myopathy. *Brain*, *133*(Pt 7), 2123-2135. doi: 10.1093/brain/awq108
- Cisternas, F. A., Vincent, J. B., Scherer, S. W., & Ray, P. N. (2003). Cloning and characterization of human CADPS and CADPS2, new members of the Ca<sup>2+</sup>-dependent activator for secretion protein family. *Genomics*, *81*(3), 279-291.
- Coe, B. P., Girirajan, S., & Eichler, E. E. (2012). The genetic variability and commonality of neurodevelopmental disease. *Am J Med Genet C Semin Med Genet*, *160C*(2), 118-129. doi: 10.1002/ajmg.c.31327
- Colella, S., Yau, C., Taylor, J. M., Mirza, G., Butler, H., Clouston, P., . . . Ragoussis, J. (2007). QuantiSNP: an Objective Bayes Hidden-Markov Model to detect and accurately map copy number variation using SNP genotyping data. *Nucleic Acids Res*, *35*(6), 2013-2025. doi: 10.1093/nar/gkm076
- Collins, V. R., Muggli, E. E., Riley, M., Palma, S., & Halliday, J. L. (2008). Is Down syndrome a disappearing birth defect? *J Pediatr*, *152*(1), 20-24, 24.e21. doi: 10.1016/j.jpeds.2007.07.045
- Conrad, D. F., Pinto, D., Redon, R., Feuk, L., Gokcumen, O., Zhang, Y., . . . Consortium, W. T. C. C. (2010). Origins and functional impact of copy number variation in the human genome. *Nature*, *464*(7289), 704-712. doi: 10.1038/nature08516
- Constantino, J. N., & Todd, R. D. (2005). Intergenerational transmission of subthreshold autistic traits in the general population. *Biol Psychiatry*, *57*(6), 655-660. doi: 10.1016/j.biopsych.2004.12.014
- Crino, P. B., Nathanson, K. L., & Henske, E. P. (2006). The tuberous sclerosis complex. *N Engl J Med*, *355*(13), 1345-1356. doi: 10.1056/NEJMra055323
- de Kovel, C. G., Trucks, H., Helbig, I., Mefford, H. C., Baker, C., Leu, C., . . . Sander, T. (2010). Recurrent microdeletions at 15q11.2 and 16p13.11 predispose to idiopathic generalized epilepsies. *Brain*, *133*(Pt 1), 23-32. doi: 10.1093/brain/awp262
- De Rubeis, S., He, X., Goldberg, A. P., Poultney, C. S., Samocha, K., Cicek, A. E., . . . Consortium, U. K. (2014). Synaptic, transcriptional and chromatin genes disrupted in autism. *Nature*, *515*(7526), 209-215. doi: 10.1038/nature13772

- Deb, S., & Prasad, K. B. (1994). The prevalence of autistic disorder among children with a learning disability. *Br J Psychiatry*, *165*(3), 395-399.
- Deneris, E. S., Boulter, J., Connolly, J., Wada, E., Wada, K., Goldman, D., . . . Heinemann, S. (1989). Genes encoding neuronal nicotinic acetylcholine receptors. *Clin Chem*, *35*(5), 731-737.
- Depienne, C., Heron, D., Betancur, C., Benyahia, B., Trouillard, O., Bouteiller, D., . . . Brice, A. (2007). Autism, language delay and mental retardation in a patient with 7q11 duplication. *J Med Genet*, *44*(7), 452-458. doi: 10.1136/jmg.2006.047092
- Devlin, B., & Scherer, S. W. (2012). Genetic architecture in autism spectrum disorder. *Curr Opin Genet Dev*, *22*(3), 229-237. doi: 10.1016/j.gde.2012.03.002
- Dillon, L. W., Pierce, L. C., Ng, M. C., & Wang, Y. H. (2013). Role of DNA secondary structures in fragile site breakage along human chromosome 10. *Hum Mol Genet*, *22*(7), 1443-1456. doi: 10.1093/hmg/dds561
- Doornbos, M., Sikkema-Raddatz, B., Ruijvenkamp, C. A., Dijkhuizen, T., Bijlsma, E. K., Gijsbers, A. C., . . . van Ravenswaaij-Arts, C. M. (2009). Nine patients with a microdeletion 15q11.2 between breakpoints 1 and 2 of the Prader-Willi critical region, possibly associated with behavioural disturbances. *Eur J Med Genet*, *52*(2-3), 108-115. doi: 10.1016/j.ejmg.2009.03.010
- Durand, C. M., Betancur, C., Boeckers, T. M., Bockmann, J., Chaste, P., Fauchereau, F., . . . Bourgeron, T. (2007). Mutations in the gene encoding the synaptic scaffolding protein SHANK3 are associated with autism spectrum disorders. *Nat Genet*, *39*(1), 25-27. doi: 10.1038/ng1933
- Durkin, M. (2002). The epidemiology of developmental disabilities in low-income countries. *Ment Retard Dev Disabil Res Rev*, *8*(3), 206-211. doi: 10.1002/mrdd.10039
- Elsabbagh, M., Divan, G., Koh, Y. J., Kim, Y. S., Kauchali, S., Marcín, C., . . . Fombonne, E. (2012). Global prevalence of autism and other pervasive developmental disorders. *Autism Res*, *5*(3), 160-179. doi: 10.1002/aur.239
- Ertekin-Taner, N., Ronald, J., Asahara, H., Younkin, L., Hella, M., Jain, S., . . . Hutton, M. (2003). Fine mapping of the alpha-T catenin gene to a quantitative trait locus on chromosome 10 in late-onset Alzheimer's disease pedigrees. *Hum Mol Genet*, *12*(23), 3133-3143. doi: 10.1093/hmg/ddg343
- Fabrichny, I. P., Leone, P., Sulzenbacher, G., Comoletti, D., Miller, M. T., Taylor, P., . . . Marchot, P. (2007). Structural analysis of the synaptic protein neuroligin and its beta-neurexin complex: determinants for folding and cell adhesion. *Neuron*, *56*(6), 979-991. doi: 10.1016/j.neuron.2007.11.013
- Feuk, L., Carson, A. R., & Scherer, S. W. (2006). Structural variation in the human genome. *Nat Rev Genet*, *7*(2), 85-97. doi: 10.1038/nrg1767
- Figueiredo, J. C., Lewinger, J. P., Song, C., Campbell, P. T., Conti, D. V., Edlund, C. K., . . . Casey, G. (2011). Genotype-environment interactions in microsatellite stable/microsatellite instability-low colorectal cancer: results from a genome-wide association study. *Cancer Epidemiol Biomarkers Prev*, *20*(5), 758-766. doi: 10.1158/1055-9965.EPI-10-0675
- Fishburn, J., Turner, G., Daniel, A., & Brookwell, R. (1983). The diagnosis and frequency of X-linked conditions in a cohort of moderately retarded males with affected brothers. *Am J Med Genet*, *14*(4), 713-724. doi: 10.1002/ajmg.1320140413
- Flomen, R. H., Davies, A. F., Di Forti, M., La Cascia, C., Mackie-Ogilvie, C., Murray, R., & Makoff, A. J. (2008). The copy number variant involving part of the alpha7 nicotinic receptor gene contains a polymorphic inversion. *Eur J Hum Genet*, *16*(11), 1364-1371. doi: 10.1038/ejhg.2008.112

- Folstein, S., & Rutter, M. (1977a). Genetic influences and infantile autism. *Nature*, 265(5596), 726-728.
- Folstein, S., & Rutter, M. (1977b). Infantile autism: a genetic study of 21 twin pairs. *J Child Psychol Psychiatry*, 18(4), 297-321.
- Fombonne, E. (2003). Modern views of autism. *Can J Psychiatry*, 48(8), 503-505.
- Fox, C. S., Liu, Y., White, C. C., Feitosa, M., Smith, A. V., Heard-Costa, N., . . . Consortium, G. (2012). Genome-wide association for abdominal subcutaneous and visceral adipose reveals a novel locus for visceral fat in women. *PLoS Genet*, 8(5), e1002695. doi: 10.1371/journal.pgen.1002695
- Friedman, J. S., Ray, J. W., Waseem, N., Johnson, K., Brooks, M. J., Hugosson, T., . . . Swaroop, A. (2009). Mutations in a BTB-Kelch protein, KLHL7, cause autosomal-dominant retinitis pigmentosa. *Am J Hum Genet*, 84(6), 792-800. doi: 10.1016/j.ajhg.2009.05.007
- Furukawa, M., He, Y. J., Borchers, C., & Xiong, Y. (2003). Targeting of protein ubiquitination by BTB-Cullin 3-Roc1 ubiquitin ligases. *Nat Cell Biol*, 5(11), 1001-1007. doi: 10.1038/ncb1056
- Galasso, C., Lo-Castro, A., Lalli, C., Nardone, A. M., Gullotta, F., & Curatolo, P. (2008). Deletion 2q37: an identifiable clinical syndrome with mental retardation and autism. *J Child Neurol*, 23(7), 802-806. doi: 10.1177/0883073808314150
- Gault, J., Robinson, M., Berger, R., Drebing, C., Logel, J., Hopkins, J., . . . Leonard, S. (1998). Genomic organization and partial duplication of the human alpha7 neuronal nicotinic acetylcholine receptor gene (CHRNA7). *Genomics*, 52(2), 173-185. doi: 10.1006/geno.1998.5363
- Gauthier, J., Champagne, N., Lafrenière, R. G., Xiong, L., Spiegelman, D., Brustein, E., . . . Team, S. D. (2010). De novo mutations in the gene encoding the synaptic scaffolding protein SHANK3 in patients ascertained for schizophrenia. *Proc Natl Acad Sci U S A*, 107(17), 7863-7868. doi: 10.1073/pnas.0906232107
- Gerrow, K., Romorini, S., Nabi, S. M., Colicos, M. A., Sala, C., & El-Husseini, A. (2006). A preformed complex of postsynaptic proteins is involved in excitatory synapse development. *Neuron*, 49(4), 547-562. doi: 10.1016/j.neuron.2006.01.015
- Gilman, S. R., Iossifov, I., Levy, D., Ronemus, M., Wigler, M., & Vitkup, D. (2011). Rare de novo variants associated with autism implicate a large functional network of genes involved in formation and function of synapses. *Neuron*, 70(5), 898-907. doi: 10.1016/j.neuron.2011.05.021
- Girirajan, S., Dennis, M. Y., Baker, C., Malig, M., Coe, B. P., Campbell, C. D., . . . Eichler, E. E. (2013). Refinement and discovery of new hotspots of copy-number variation associated with autism spectrum disorder. *Am J Hum Genet*, 92(2), 221-237. doi: 10.1016/j.ajhg.2012.12.016
- Girirajan, S., Rosenfeld, J. A., Coe, B. P., Parikh, S., Friedman, N., Goldstein, A., . . . Eichler, E. E. (2012). Phenotypic heterogeneity of genomic disorders and rare copy-number variants. *N Engl J Med*, 367(14), 1321-1331. doi: 10.1056/NEJMoa1200395
- Girirajan, S., Rosenfeld, J. A., Cooper, G. M., Antonacci, F., Siswara, P., Itsara, A., . . . Eichler, E. E. (2010). A recurrent 16p12.1 microdeletion supports a two-hit model for severe developmental delay. *Nat Genet*, 42(3), 203-209. doi: 10.1038/ng.534
- Glessner, J. T., Wang, K., Cai, G., Korvatska, O., Kim, C. E., Wood, S., . . . Hakonarson, H. (2009). Autism genome-wide copy number variation reveals ubiquitin and neuronal genes. *Nature*, 459(7246), 569-573. doi: 10.1038/nature07953
- Goffin, A., Hoefsloot, L. H., Bosgoed, E., Swillen, A., & Fryns, J. P. (2001). PTEN mutation in a family with Cowden syndrome and autism. *Am J Med Genet*, 105(6), 521-524.

- Goldman, D., Deneris, E., Luyten, W., Kochhar, A., Patrick, J., & Heinemann, S. (1987). Members of a nicotinic acetylcholine receptor gene family are expressed in different regions of the mammalian central nervous system. *Cell*, *48*(6), 965-973.
- Goodier, J. L., & Kazazian, H. H. (2008). Retrotransposons revisited: the restraint and rehabilitation of parasites. *Cell*, *135*(1), 23-35. doi: 10.1016/j.cell.2008.09.022
- Goossens, S., Janssens, B., Bonn e, S., De Rycke, R., Braet, F., van Hengel, J., & van Roy, F. (2007). A unique and specific interaction between alphaT-catenin and plakophilin-2 in the area composita, the mixed-type junctional structure of cardiac intercalated discs. *J Cell Sci*, *120*(Pt 12), 2126-2136. doi: 10.1242/jcs.004713
- Greenberg, M. E., Xu, B., Lu, B., & Hempstead, B. L. (2009). New insights in the biology of BDNF synthesis and release: implications in CNS function. *J Neurosci*, *29*(41), 12764-12767. doi: 10.1523/JNEUROSCI.3566-09.2009
- Griggs, B. L., Ladd, S., Saul, R. A., DuPont, B. R., & Srivastava, A. K. (2008). Deducator of cytokinesis 8 is disrupted in two patients with mental retardation and developmental disabilities. *Genomics*, *91*(2), 195-202. doi: 10.1016/j.ygeno.2007.10.011
- Gu, W., Zhang, F., & Lupski, J. R. (2008). Mechanisms for human genomic rearrangements. *Pathogenetics*, *1*(1), 4. doi: 10.1186/1755-8417-1-4
- Gutierrez, G. C., Smalley, S. L., & Tanguay, P. E. (1998). Autism in tuberous sclerosis complex. *J Autism Dev Disord*, *28*(2), 97-103.
- G ecz, J., Cloosterman, D., & Partington, M. (2006). ARX: a gene for all seasons. *Curr Opin Genet Dev*, *16*(3), 308-316. doi: 10.1016/j.gde.2006.04.003
- Hajirasouliha, I., Hormozdiari, F., Alkan, C., Kidd, J. M., Birol, I., Eichler, E. E., & Sahinalp, S. C. (2010). Detection and characterization of novel sequence insertions using paired-end next-generation sequencing. *Bioinformatics*, *26*(10), 1277-1283. doi: 10.1093/bioinformatics/btq152
- Hallmayer, J., Cleveland, S., Torres, A., Phillips, J., Cohen, B., Torigoe, T., . . . Risch, N. (2011). Genetic heritability and shared environmental factors among twin pairs with autism. *Arch Gen Psychiatry*, *68*(11), 1095-1102. doi: 10.1001/archgenpsychiatry.2011.76
- Hamdan, F. F., Piton, A., Gauthier, J., Lortie, A., Dubeau, F., Dobrzyniecka, S., . . . Michaud, J. L. (2009). De novo STXBP1 mutations in mental retardation and nonsyndromic epilepsy. *Ann Neurol*, *65*(6), 748-753. doi: 10.1002/ana.21625
- Harvard, C., Strong, E., Mercier, E., Colnaghi, R., Alcantara, D., Chow, E., . . . Rajcan-Separovic, E. (2011). Understanding the impact of 1q21.1 copy number variant. *Orphanet J Rare Dis*, *6*, 54. doi: 10.1186/1750-1172-6-54
- Hastings, P. J., Ira, G., & Lupski, J. R. (2009). A microhomology-mediated break-induced replication model for the origin of human copy number variation. *PLoS Genet*, *5*(1), e1000327. doi: 10.1371/journal.pgen.1000327
- Hastings, P. J., Lupski, J. R., Rosenberg, S. M., & Ira, G. (2009). Mechanisms of change in gene copy number. *Nat Rev Genet*, *10*(8), 551-564. doi: 10.1038/nrg2593
- Helbig, I., Mefford, H. C., Sharp, A. J., Guipponi, M., Fichera, M., Franke, A., . . . Sander, T. (2009). 15q13.3 microdeletions increase risk of idiopathic generalized epilepsy. *Nat Genet*, *41*(2), 160-162. doi: 10.1038/ng.292
- Herbst, D. S., & Miller, J. R. (1980). Nonspecific X-linked mental retardation II: the frequency in British Columbia. *Am J Med Genet*, *7*(4), 461-469. doi: 10.1002/ajmg.1320070407
- Hirano, S., Kimoto, N., Shimoyama, Y., Hirohashi, S., & Takeichi, M. (1992). Identification of a neural alpha-catenin as a key regulator of cadherin function and multicellular organization. *Cell*, *70*(2), 293-301.

- Holt, R., Sykes, N. H., Conceição, I. C., Cazier, J. B., Anney, R. J., Oliveira, G., . . . Pagnamenta, A. T. (2012). CNVs leading to fusion transcripts in individuals with autism spectrum disorder. *Eur J Hum Genet*, *20*(11), 1141-1147. doi: 10.1038/ejhg.2012.73
- Hong, E. J., McCord, A. E., & Greenberg, M. E. (2008). A biological function for the neuronal activity-dependent component of Bdnf transcription in the development of cortical inhibition. *Neuron*, *60*(4), 610-624. doi: 10.1016/j.neuron.2008.09.024
- Hoppman-Chaney, N., Wain, K., Seger, P. R., Superneau, D. W., & Hodge, J. C. (2013). Identification of single gene deletions at 15q13.3: further evidence that CHRNA7 causes the 15q13.3 microdeletion syndrome phenotype. *Clin Genet*, *83*(4), 345-351. doi: 10.1111/j.1399-0004.2012.01925.x
- Houwen, R. H., Baharloo, S., Blankenship, K., Raeymaekers, P., Juyn, J., Sandkuijl, L. A., & Freimer, N. B. (1994). Genome screening by searching for shared segments: mapping a gene for benign recurrent intrahepatic cholestasis. *Nat Genet*, *8*(4), 380-386. doi: 10.1038/ng1294-380
- Hus, V., Pickles, A., Cook, E. H., Risi, S., & Lord, C. (2007). Using the autism diagnostic interview--revised to increase phenotypic homogeneity in genetic studies of autism. *Biol Psychiatry*, *61*(4), 438-448. doi: 10.1016/j.biopsych.2006.08.044
- Iafrate, A. J., Feuk, L., Rivera, M. N., Listewnik, M. L., Donahoe, P. K., Qi, Y., . . . Lee, C. (2004). Detection of large-scale variation in the human genome. *Nat Genet*, *36*(9), 949-951. doi: 10.1038/ng1416
- Iossifov, I., O'Roak, B. J., Sanders, S. J., Ronemus, M., Krumm, N., Levy, D., . . . Wigler, M. (2014). The contribution of de novo coding mutations to autism spectrum disorder. *Nature*, *515*(7526), 216-221. doi: 10.1038/nature13908
- Jacquemont, M. L., Sanlaville, D., Redon, R., Raoul, O., Cormier-Daire, V., Lyonnet, S., . . . Philippe, A. (2006). Array-based comparative genomic hybridisation identifies high frequency of cryptic chromosomal rearrangements in patients with syndromic autism spectrum disorders. *J Med Genet*, *43*(11), 843-849. doi: 10.1136/jmg.2006.043166
- Jamain, S., Betancur, C., Quach, H., Philippe, A., Fellous, M., Giros, B., . . . Study, P. A. R. I. S. P. (2002). Linkage and association of the glutamate receptor 6 gene with autism. *Mol Psychiatry*, *7*(3), 302-310. doi: 10.1038/sj.mp.4000979
- Jamain, S., Quach, H., Betancur, C., Rastam, M., Colineaux, C., Gillberg, I. C., . . . Bourgeron, T. (2003). Mutations of the X-linked genes encoding neuroligins NLGN3 and NLGN4 are associated with autism. *Nat Genet*, *34*(1), 27-29. doi: 10.1038/ng1136
- Janssens, B., Goossens, S., Staes, K., Gilbert, B., van Hengel, J., Colpaert, C., . . . van Roy, F. (2001). alphaT-catenin: a novel tissue-specific beta-catenin-binding protein mediating strong cell-cell adhesion. *J Cell Sci*, *114*(Pt 17), 3177-3188.
- Jauch, A., Robson, L., & Smith, A. (1995). Investigations with fluorescence in situ hybridization (FISH) demonstrate loss of the telomeres on the reciprocal chromosome in three unbalanced translocations involving chromosome 15 in the Prader-Willi and Angelman syndromes. *Hum Genet*, *96*(3), 345-349.
- Jiang, S., Avraham, H. K., Park, S. Y., Kim, T. A., Bu, X., Seng, S., & Avraham, S. (2005). Process elongation of oligodendrocytes is promoted by the Kelch-related actin-binding protein Mayven. *J Neurochem*, *92*(5), 1191-1203. doi: 10.1111/j.1471-4159.2004.02946.x
- Jiang, S., Seng, S., Avraham, H. K., Fu, Y., & Avraham, S. (2007). Process elongation of oligodendrocytes is promoted by the Kelch-related protein MRP2/KLHL1. *J Biol Chem*, *282*(16), 12319-12329. doi: 10.1074/jbc.M701019200
- Jones, I. W., & Wonnacott, S. (2004). Precise localization of alpha7 nicotinic acetylcholine receptors on glutamatergic axon terminals in the rat ventral tegmental area. *J Neurosci*, *24*(50), 11244-11252. doi: 10.1523/JNEUROSCI.3009-04.2004

- Kalkman, H. O. (2012). A review of the evidence for the canonical Wnt pathway in autism spectrum disorders. *Mol Autism*, 3(1), 10. doi: 10.1186/2040-2392-3-10
- Kang, H. J., Kawasawa, Y. I., Cheng, F., Zhu, Y., Xu, X., Li, M., . . . Sestan, N. (2011). Spatio-temporal transcriptome of the human brain. *Nature*, 478(7370), 483-489. doi: 10.1038/nature10523
- Kang, M. I., Kobayashi, A., Wakabayashi, N., Kim, S. G., & Yamamoto, M. (2004). Scaffolding of Keap1 to the actin cytoskeleton controls the function of Nrf2 as key regulator of cytoprotective phase 2 genes. *Proc Natl Acad Sci U S A*, 101(7), 2046-2051. doi: 10.1073/pnas.0308347100
- Kanner, L. (1968). Autistic disturbances of affective contact. *Acta Paedopsychiatr*, 35(4), 100-136.
- Katusic, S. K., Colligan, R. C., Beard, C. M., O'Fallon, W. M., Bergstralh, E. J., Jacobsen, S. J., & Kurland, L. T. (1996). Mental retardation in a birth cohort, 1976-1980, Rochester, Minnesota. *Am J Ment Retard*, 100(4), 335-344.
- Kaufmann, W. E., Cortell, R., Kau, A. S., Bukelis, I., Tierney, E., Gray, R. M., . . . Stanard, P. (2004). Autism spectrum disorder in fragile X syndrome: communication, social interaction, and specific behaviors. *Am J Med Genet A*, 129A(3), 225-234. doi: 10.1002/ajmg.a.30229
- Kelleher, R. J., & Bear, M. F. (2008). The autistic neuron: troubled translation? *Cell*, 135(3), 401-406. doi: 10.1016/j.cell.2008.10.017
- Kidd, J. M., Cooper, G. M., Donahue, W. F., Hayden, H. S., Sampas, N., Graves, T., . . . Eichler, E. E. (2008). Mapping and sequencing of structural variation from eight human genomes. *Nature*, 453(7191), 56-64. doi: 10.1038/nature06862
- Kim, S. A., Kim, J. H., Park, M., Cho, I. H., & Yoo, H. J. (2007). Family-based association study between GRIK2 polymorphisms and autism spectrum disorders in the Korean trios. *Neurosci Res*, 58(3), 332-335. doi: 10.1016/j.neures.2007.03.002
- Kirov, G., Grozeva, D., Norton, N., Ivanov, D., Mantripragada, K. K., Holmans, P., . . . Consortium, W. T. C. C. (2009). Support for the involvement of large copy number variants in the pathogenesis of schizophrenia. *Hum Mol Genet*, 18(8), 1497-1503. doi: 10.1093/hmg/ddp043
- Kirov, G., Pocklington, A. J., Holmans, P., Ivanov, D., Ikeda, M., Ruderfer, D., . . . Owen, M. J. (2012). De novo CNV analysis implicates specific abnormalities of postsynaptic signalling complexes in the pathogenesis of schizophrenia. *Mol Psychiatry*, 17(2), 142-153. doi: 10.1038/mp.2011.154
- Kong, A., Frigge, M. L., Masson, G., Besenbacher, S., Sulem, P., Magnusson, G., . . . Stefansson, K. (2012). Rate of de novo mutations and the importance of father's age to disease risk. *Nature*, 488(7412), 471-475. doi: 10.1038/nature11396
- Koolen, D. A., Pfundt, R., de Leeuw, N., Hahir-Kwa, J. Y., Nillesen, W. M., Neefs, I., . . . de Vries, B. B. (2009). Genomic microarrays in mental retardation: a practical workflow for diagnostic applications. *Hum Mutat*, 30(3), 283-292. doi: 10.1002/humu.20883
- Korbel, J. O., Urban, A. E., Affourtit, J. P., Godwin, B., Grubert, F., Simons, J. F., . . . Snyder, M. (2007). Paired-end mapping reveals extensive structural variation in the human genome. *Science*, 318(5849), 420-426. doi: 10.1126/science.1149504
- Korbie, D. J., & Mattick, J. S. (2008). Touchdown PCR for increased specificity and sensitivity in PCR amplification. *Nat Protoc*, 3(9), 1452-1456. doi: 10.1038/nprot.2008.133
- Kou, J., Kovacs, G. G., Höftberger, R., Kulik, W., Brodde, A., Forss-Petter, S., . . . Berger, J. (2011). Peroxisomal alterations in Alzheimer's disease. *Acta Neuropathol*, 122(3), 271-283. doi: 10.1007/s00401-011-0836-9
- Kou, Y., Betancur, C., Xu, H., Buxbaum, J. D., & Ma'ayan, A. (2012). Network- and attribute-based classifiers can prioritize genes and pathways for autism spectrum disorders and

- intellectual disability. *Am J Med Genet C Semin Med Genet*, 160C(2), 130-142. doi: 10.1002/ajmg.c.31330
- Krawczak, M., Nikolaus, S., von Eberstein, H., Croucher, P. J., El Mokhtari, N. E., & Schreiber, S. (2006). PopGen: population-based recruitment of patients and controls for the analysis of complex genotype-phenotype relationships. *Community Genet*, 9(1), 55-61. doi: 10.1159/000090694
- Kruglyak, L., & Nickerson, D. A. (2001). Variation is the spice of life. *Nat Genet*, 27(3), 234-236. doi: 10.1038/85776
- Kumar, P., Henikoff, S., & Ng, P. C. (2009). Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm. *Nat Protoc*, 4(7), 1073-1081. doi: 10.1038/nprot.2009.86
- Kutsche, K., Yntema, H., Brandt, A., Jantke, I., Nothwang, H. G., Orth, U., . . . Gal, A. (2000). Mutations in ARHGEF6, encoding a guanine nucleotide exchange factor for Rho GTPases, in patients with X-linked mental retardation. *Nat Genet*, 26(2), 247-250. doi: 10.1038/80002
- LaFramboise, T. (2009). Single nucleotide polymorphism arrays: a decade of biological, computational and technological advances. *Nucleic Acids Res*, 37(13), 4181-4193. doi: 10.1093/nar/gkp552
- Lamb, J. A., Barnby, G., Bonora, E., Sykes, N., Bacchelli, E., Blasi, F., . . . (IMGSAC), I. M. G. S. o. A. C. (2005). Analysis of IMGSAC autism susceptibility loci: evidence for sex limited and parent of origin specific effects. *J Med Genet*, 42(2), 132-137. doi: 10.1136/jmg.2004.025668
- Laumonnier, F., Bonnet-Brilhault, F., Gomot, M., Blanc, R., David, A., Moizard, M. P., . . . Briault, S. (2004). 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*, 74(3), 552-557. doi: 10.1086/382137
- Laurén, J., Airaksinen, M. S., Saarma, M., & Timmusk, T. (2003). A novel gene family encoding leucine-rich repeat transmembrane proteins differentially expressed in the nervous system. *Genomics*, 81(4), 411-421.
- Leblond, C. S., Heinrich, J., Delorme, R., Proepper, C., Betancur, C., Huguet, G., . . . Bourgeron, T. (2012). Genetic and functional analyses of SHANK2 mutations suggest a multiple hit model of autism spectrum disorders. *PLoS Genet*, 8(2), e1002521. doi: 10.1371/journal.pgen.1002521
- LeDoux, J. (2007). The amygdala. *Curr Biol*, 17(20), R868-874. doi: 10.1016/j.cub.2007.08.005
- Lee, J. A., Carvalho, C. M., & Lupski, J. R. (2007). A DNA replication mechanism for generating nonrecurrent rearrangements associated with genomic disorders. *Cell*, 131(7), 1235-1247. doi: 10.1016/j.cell.2007.11.037
- Leonard, H., & Wen, X. (2002). The epidemiology of mental retardation: challenges and opportunities in the new millennium. *Ment Retard Dev Disabil Res Rev*, 8(3), 117-134. doi: 10.1002/mrdd.10031
- Leonard, S., & Freedman, R. (2006). Genetics of chromosome 15q13-q14 in schizophrenia. *Biol Psychiatry*, 60(2), 115-122. doi: 10.1016/j.biopsych.2006.03.054
- Leonard, S., Gault, J., Hopkins, J., Logel, J., Vianzon, R., Short, M., . . . Freedman, R. (2002). Association of promoter variants in the alpha7 nicotinic acetylcholine receptor subunit gene with an inhibitory deficit found in schizophrenia. *Arch Gen Psychiatry*, 59(12), 1085-1096.
- Levy, D., Ronemus, M., Yamrom, B., Lee, Y. H., Leotta, A., Kendall, J., . . . Wigler, M. (2011). Rare de novo and transmitted copy-number variation in autistic spectrum disorders. *Neuron*, 70(5), 886-897. doi: 10.1016/j.neuron.2011.05.015

- Levy, S., Sutton, G., Ng, P. C., Feuk, L., Halpern, A. L., Walenz, B. P., . . . Venter, J. C. (2007). The diploid genome sequence of an individual human. *PLoS Biol*, *5*(10), e254. doi: 10.1371/journal.pbio.0050254
- Li, J., Goossens, S., van Hengel, J., Gao, E., Cheng, L., Tyberghein, K., . . . Radice, G. L. (2012). Loss of  $\alpha$ T-catenin alters the hybrid adhering junctions in the heart and leads to dilated cardiomyopathy and ventricular arrhythmia following acute ischemia. *J Cell Sci*, *125*(Pt 4), 1058-1067. doi: 10.1242/jcs.098640
- Liang, X. Q., Avraham, H. K., Jiang, S., & Avraham, S. (2004). Genetic alterations of the NRP/B gene are associated with human brain tumors. *Oncogene*, *23*(35), 5890-5900. doi: 10.1038/sj.onc.1207776
- Lieber, M. R. (2008). The mechanism of human nonhomologous DNA end joining. *J Biol Chem*, *283*(1), 1-5. doi: 10.1074/jbc.R700039200
- Lim, E. T., Raychaudhuri, S., Sanders, S. J., Stevens, C., Sabo, A., MacArthur, D. G., . . . Daly, M. J. (2013). Rare complete knockouts in humans: population distribution and significant role in autism spectrum disorders. *Neuron*, *77*(2), 235-242. doi: 10.1016/j.neuron.2012.12.029
- Linhoff, M. W., Laurén, J., Cassidy, R. M., Dobie, F. A., Takahashi, H., Nygaard, H. B., . . . Craig, A. M. (2009). An unbiased expression screen for synaptogenic proteins identifies the LRRTM protein family as synaptic organizers. *Neuron*, *61*(5), 734-749. doi: 10.1016/j.neuron.2009.01.017
- Lintas, C., & Persico, A. M. (2009). Autistic phenotypes and genetic testing: state-of-the-art for the clinical geneticist. *J Med Genet*, *46*(1), 1-8. doi: 10.1136/jmg.2008.060871
- Livak, K. J., & Schmittgen, T. D. (2001). Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) Method. *Methods*, *25*(4), 402-408. doi: 10.1006/meth.2001.1262
- Locke, D. P., Archidiacono, N., Miscio, D., Cardone, M. F., Deschamps, S., Roe, B., . . . Eichler, E. E. (2003). Refinement of a chimpanzee pericentric inversion breakpoint to a segmental duplication cluster. *Genome Biol*, *4*(8), R50. doi: 10.1186/gb-2003-4-8-r50
- Lord, C., Risi, S., Lambrecht, L., Cook, E. H., Leventhal, B. L., DiLavore, P. C., . . . Rutter, M. (2000). The autism diagnostic observation schedule-generic: a standard measure of social and communication deficits associated with the spectrum of autism. *J Autism Dev Disord*, *30*(3), 205-223.
- Lord, C., Rutter, M., Goode, S., Heemsbergen, J., Jordan, H., Mawhood, L., & Schopler, E. (1989). Autism diagnostic observation schedule: a standardized observation of communicative and social behavior. *J Autism Dev Disord*, *19*(2), 185-212.
- Lord, C., Rutter, M., & Le Couteur, A. (1994). Autism Diagnostic Interview-Revised: a revised version of a diagnostic interview for caregivers of individuals with possible pervasive developmental disorders. *J Autism Dev Disord*, *24*(5), 659-685.
- Losh, M., Adolphs, R., Poe, M. D., Couture, S., Penn, D., Baranek, G. T., & Piven, J. (2009). Neuropsychological profile of autism and the broad autism phenotype. *Arch Gen Psychiatry*, *66*(5), 518-526. doi: 10.1001/archgenpsychiatry.2009.34
- Losh, M., Martin, G. E., Klusek, J., Hogan-Brown, A. L., & Sideris, J. (2012). Social communication and theory of mind in boys with autism and fragile x syndrome. *Front Psychol*, *3*, 266. doi: 10.3389/fpsyg.2012.00266
- Lucito, R., Healy, J., Alexander, J., Reiner, A., Esposito, D., Chi, M., . . . Wigler, M. (2003). Representational oligonucleotide microarray analysis: a high-resolution method to detect genome copy number variation. *Genome Res*, *13*(10), 2291-2305. doi: 10.1101/gr.1349003
- Lupski, J. R. (1998). Genomic disorders: structural features of the genome can lead to DNA rearrangements and human disease traits. *Trends Genet*, *14*(10), 417-422.

- Lupski, J. R. (2004). Hotspots of homologous recombination in the human genome: not all homologous sequences are equal. *Genome Biol*, 5(10), 242. doi: 10.1186/gb-2004-5-10-242
- Lupski, J. R. (2010). Retrotransposition and structural variation in the human genome. *Cell*, 141(7), 1110-1112. doi: 10.1016/j.cell.2010.06.014
- Maestrini, E., Pagnamenta, A. T., Lamb, J. A., Bacchelli, E., Sykes, N. H., Sousa, I., . . . IMGSAC. (2010). High-density SNP association study and copy number variation analysis of the AUTS1 and AUTS5 loci implicate the IMMP2L-DOCK4 gene region in autism susceptibility. *Mol Psychiatry*, 15(9), 954-968. doi: 10.1038/mp.2009.34
- Majercak, J., Ray, W. J., Espeseth, A., Simon, A., Shi, X. P., Wolffe, C., . . . Stone, D. J. (2006). LRRTM3 promotes processing of amyloid-precursor protein by BACE1 and is a positional candidate gene for late-onset Alzheimer's disease. *Proc Natl Acad Sci U S A*, 103(47), 17967-17972. doi: 10.1073/pnas.0605461103
- Makoff, A. J., & Flomen, R. H. (2007). Detailed analysis of 15q11-q14 sequence corrects errors and gaps in the public access sequence to fully reveal large segmental duplications at breakpoints for Prader-Willi, Angelman, and inv dup(15) syndromes. *Genome Biol*, 8(6), R114. doi: 10.1186/gb-2007-8-6-r114
- Malhotra, D., McCarthy, S., Michaelson, J. J., Vacic, V., Burdick, K. E., Yoon, S., . . . Sebat, J. (2011). High frequencies of de novo CNVs in bipolar disorder and schizophrenia. *Neuron*, 72(6), 951-963. doi: 10.1016/j.neuron.2011.11.007
- Manolio, T. A., Collins, F. S., Cox, N. J., Goldstein, D. B., Hindorff, L. A., Hunter, D. J., . . . Visscher, P. M. (2009). Finding the missing heritability of complex diseases. *Nature*, 461(7265), 747-753. doi: 10.1038/nature08494
- Marshall, C. R., Noor, A., Vincent, J. B., Lionel, A. C., Feuk, L., Skaug, J., . . . Scherer, S. W. (2008). Structural variation of chromosomes in autism spectrum disorder. *Am J Hum Genet*, 82(2), 477-488. doi: 10.1016/j.ajhg.2007.12.009
- Martin, E. R., Bronson, P. G., Li, Y. J., Wall, N., Chung, R. H., Schmechel, D. E., . . . Pericak-Vance, M. A. (2005). Interaction between the alpha-T catenin gene (VR22) and APOE in Alzheimer's disease. *J Med Genet*, 42(10), 787-792. doi: 10.1136/jmg.2004.029553
- Masurel-Paulet, A., Andrieux, J., Callier, P., Cuisset, J. M., Le Caignec, C., Holder, M., . . . Faivre, L. (2010). Delineation of 15q13.3 microdeletions. *Clin Genet*, 78(2), 149-161. doi: 10.1111/j.1399-0004.2010.01374.x
- Matson, J. L., Bielecki, J., Mayville, S. B., & Matson, M. L. (2003). Psychopharmacology research for individuals with mental retardation: methodological issues and suggestions. *Res Dev Disabil*, 24(3), 149-157.
- McClellan, J., & King, M. C. (2010). Genetic heterogeneity in human disease. *Cell*, 141(2), 210-217. doi: 10.1016/j.cell.2010.03.032
- McDuffie, A., Abbeduto, L., Lewis, P., Kover, S., Kim, J. S., Weber, A., & Brown, W. T. (2010). Autism spectrum disorder in children and adolescents with fragile X syndrome: within-syndrome differences and age-related changes. *Am J Intellect Dev Disabil*, 115(4), 307-326. doi: 10.1352/1944-7558-115.4.307
- McLaren, J., & Bryson, S. E. (1987). Review of recent epidemiological studies of mental retardation: prevalence, associated disorders, and etiology. *Am J Ment Retard*, 92(3), 243-254.
- McMullan, D. J., Bonin, M., Hehir-Kwa, J. Y., de Vries, B. B., Dufke, A., Rattenberry, E., . . . Veltman, J. A. (2009). Molecular karyotyping of patients with unexplained mental retardation by SNP arrays: a multicenter study. *Hum Mutat*, 30(7), 1082-1092. doi: 10.1002/humu.21015
- Mefford, H. C., Batshaw, M. L., & Hoffman, E. P. (2012). Genomics, intellectual disability, and autism. *N Engl J Med*, 366(8), 733-743. doi: 10.1056/NEJMra1114194

- Meguro-Horike, M., Yasui, D. H., Powell, W., Schroeder, D. I., Oshimura, M., Lasalle, J. M., & Horike, S. (2011). Neuron-specific impairment of inter-chromosomal pairing and transcription in a novel model of human 15q-duplication syndrome. *Hum Mol Genet*, *20*(19), 3798-3810. doi: 10.1093/hmg/ddr298
- Melnick, A., Ahmad, K. F., Arai, S., Polinger, A., Ball, H., Borden, K. L., . . . Licht, J. D. (2000). In-depth mutational analysis of the promyelocytic leukemia zinc finger BTB/POZ domain reveals motifs and residues required for biological and transcriptional functions. *Mol Cell Biol*, *20*(17), 6550-6567.
- Metzker, M. L. (2010). Sequencing technologies - the next generation. *Nat Rev Genet*, *11*(1), 31-46. doi: 10.1038/nrg2626
- Mexal, S., Berger, R., Pearce, L., Barton, A., Logel, J., Adams, C. E., . . . Leonard, S. (2008). Regulation of a novel alphaN-catenin splice variant in schizophrenic smokers. *Am J Med Genet B Neuropsychiatr Genet*, *147B*(6), 759-768. doi: 10.1002/ajmg.b.30679
- Mignon-Ravix, C., Cacciagli, P., Choucair, N., Popovici, C., Missirian, C., Milh, M., . . . Villard, L. (2014). Intragenic rearrangements in X-linked intellectual deficiency: results of a-CGH in a series of 54 patients and identification of TRPC5 and KLHL15 as potential XLID genes. *Am J Med Genet A*, *164A*(8), 1991-1997. doi: 10.1002/ajmg.a.36602
- Mikhail, F. M., Lose, E. J., Robin, N. H., Descartes, M. D., Rutledge, K. D., Rutledge, S. L., . . . Carroll, A. J. (2011). Clinically relevant single gene or intragenic deletions encompassing critical neurodevelopmental genes in patients with developmental delay, mental retardation, and/or autism spectrum disorders. *Am J Med Genet A*, *155A*(10), 2386-2396. doi: 10.1002/ajmg.a.34177
- Miles, J. H. (2011). Autism spectrum disorders--a genetics review. *Genet Med*, *13*(4), 278-294. doi: 10.1097/GIM.0b013e3181ff67ba
- Miller, D. T., Shen, Y., Weiss, L. A., Korn, J., Anselm, I., Bridgemohan, C., . . . Wu, B. L. (2009). Microdeletion/duplication at 15q13.2q13.3 among individuals with features of autism and other neuropsychiatric disorders. *J Med Genet*, *46*(4), 242-248. doi: 10.1136/jmg.2008.059907
- Mills, R. E., Walter, K., Stewart, C., Handsaker, R. E., Chen, K., Alkan, C., . . . Project, G. (2011). Mapping copy number variation by population-scale genome sequencing. *Nature*, *470*(7332), 59-65. doi: 10.1038/nature09708
- Minor, D. L., Lin, Y. F., Mobley, B. C., Avelar, A., Jan, Y. N., Jan, L. Y., & Berger, J. M. (2000). The polar T1 interface is linked to conformational changes that open the voltage-gated potassium channel. *Cell*, *102*(5), 657-670.
- Missler, M., Zhang, W., Rohlmann, A., Kattenstroth, G., Hammer, R. E., Gottmann, K., & Südhof, T. C. (2003). Alpha-neurexins couple Ca<sup>2+</sup> channels to synaptic vesicle exocytosis. *Nature*, *423*(6943), 939-948. doi: 10.1038/nature01755
- Moessner, R., Marshall, C. R., Sutcliffe, J. S., Skaug, J., Pinto, D., Vincent, J., . . . Scherer, S. W. (2007). Contribution of SHANK3 mutations to autism spectrum disorder. *Am J Hum Genet*, *81*(6), 1289-1297. doi: 10.1086/522590
- Moreno-De-Luca, D., Sanders, S. J., Willsey, A. J., Mulle, J. G., Lowe, J. K., Geschwind, D. H., . . . Ledbetter, D. H. (2013). Using large clinical data sets to infer pathogenicity for rare copy number variants in autism cohorts. *Mol Psychiatry*, *18*(10), 1090-1095. doi: 10.1038/mp.2012.138
- Nagafuchi, A., Takeichi, M., & Tsukita, S. (1991). The 102 kd cadherin-associated protein: similarity to vinculin and posttranscriptional regulation of expression. *Cell*, *65*(5), 849-857.
- Najmabadi, H., Hu, H., Garshasbi, M., Zemojtel, T., Abedini, S. S., Chen, W., . . . Ropers, H. H. (2011). Deep sequencing reveals 50 novel genes for recessive cognitive disorders. *Nature*, *478*(7367), 57-63. doi: 10.1038/nature10423

- Neale, B. M., Kou, Y., Liu, L., Ma'ayan, A., Samocha, K. E., Sabo, A., . . . Daly, M. J. (2012). Patterns and rates of exonic de novo mutations in autism spectrum disorders. *Nature*, *485*(7397), 242-245. doi: 10.1038/nature11011
- Nguyen, D. K., & Disteche, C. M. (2006). Dosage compensation of the active X chromosome in mammals. *Nat Genet*, *38*(1), 47-53. doi: 10.1038/ng1705
- Noor, A., Whibley, A., Marshall, C. R., Gianakopoulos, P. J., Piton, A., Carson, A. R., . . . Consortium, A. G. P. (2010). Disruption at the PTCHD1 Locus on Xp22.11 in Autism spectrum disorder and intellectual disability. *Sci Transl Med*, *2*(49), 49ra68. doi: 10.1126/scitranslmed.3001267
- O'Roak, B. J., Vives, L., Girirajan, S., Karakoc, E., Krumm, N., Coe, B. P., . . . Eichler, E. E. (2012). Sporadic autism exomes reveal a highly interconnected protein network of de novo mutations. *Nature*, *485*(7397), 246-250. doi: 10.1038/nature10989
- Ozonoff, S., Young, G. S., Carter, A., Messinger, D., Yirmiya, N., Zwaigenbaum, L., . . . Stone, W. L. (2011). Recurrence risk for autism spectrum disorders: a Baby Siblings Research Consortium study. *Pediatrics*, *128*(3), e488-495. doi: 10.1542/peds.2010-2825
- Paciorkowski, A. R., Thio, L. L., Rosenfeld, J. A., GajECKa, M., Gurnett, C. A., Kulkarni, S., . . . DobyNS, W. B. (2011). Copy number variants and infantile spasms: evidence for abnormalities in ventral forebrain development and pathways of synaptic function. *Eur J Hum Genet*, *19*(12), 1238-1245. doi: 10.1038/ejhg.2011.121
- Pagnamenta, A. T., Wing, K., Sadighi Akha, E., Knight, S. J., Bölte, S., SchMötzer, G., . . . Consortium, I. M. G. S. o. A. (2009). A 15q13.3 microdeletion segregating with autism. *Eur J Hum Genet*, *17*(5), 687-692. doi: 10.1038/ejhg.2008.228
- Park, C., Falls, W., Finger, J. H., Longo-Guess, C. M., & Ackerman, S. L. (2002). Deletion in *Catna2*, encoding alpha N-catenin, causes cerebellar and hippocampal lamination defects and impaired startle modulation. *Nat Genet*, *31*(3), 279-284. doi: 10.1038/ng908
- Phelan, K., & McDermid, H. (2012). The 22q13.3 Deletion Syndrome (Phelan-McDermid Syndrome). *Mol Syndromol*, *2*(3-5), 186-201. doi: 10.1159/000334260
- Pinkel, D., SeGraves, R., Sudar, D., Clark, S., Poole, I., Kowbel, D., . . . Albertson, D. G. (1998). High resolution analysis of DNA copy number variation using comparative genomic hybridization to microarrays. *Nat Genet*, *20*(2), 207-211. doi: 10.1038/2524
- Pinto, D., Darvishi, K., Shi, X., Rajan, D., Rigler, D., Fitzgerald, T., . . . Feuk, L. (2011). Comprehensive assessment of array-based platforms and calling algorithms for detection of copy number variants. *Nat Biotechnol*, *29*(6), 512-520. doi: 10.1038/nbt.1852
- Pinto, D., Delaby, E., Merico, D., Barbosa, M., Merikangas, A., Klei, L., . . . Scherer, S. W. (2014). Convergence of genes and cellular pathways dysregulated in autism spectrum disorders. *Am J Hum Genet*, *94*(5), 677-694. doi: 10.1016/j.ajhg.2014.03.018
- Pinto, D., Pagnamenta, A. T., Klei, L., Anney, R., Merico, D., Regan, R., . . . Betancur, C. (2010). Functional impact of global rare copy number variation in autism spectrum disorders. *Nature*, *466*(7304), 368-372. doi: 10.1038/nature09146
- Piton, A., Michaud, J. L., Peng, H., Aradhya, S., Gauthier, J., Mottron, L., . . . team, S. D. (2008). Mutations in the calcium-related gene *IL1RAPL1* are associated with autism. *Hum Mol Genet*, *17*(24), 3965-3974. doi: 10.1093/hmg/ddn300
- Prasad, A., Merico, D., Thiruvahindrapuram, B., Wei, J., Lionel, A. C., Sato, D., . . . Scherer, S. W. (2012). A discovery resource of rare copy number variations in individuals with autism spectrum disorder. *G3 (Bethesda)*, *2*(12), 1665-1685. doi: 10.1534/g3.112.004689
- Rauch, A., Hoyer, J., Guth, S., Zweier, C., Kraus, C., Becker, C., . . . Trautmann, U. (2006). Diagnostic yield of various genetic approaches in patients with unexplained developmental delay or mental retardation. *Am J Med Genet A*, *140*(19), 2063-2074. doi: 10.1002/ajmg.a.31416

- Redon, R., Ishikawa, S., Fitch, K. R., Feuk, L., Perry, G. H., Andrews, T. D., . . . Hurles, M. E. (2006). Global variation in copy number in the human genome. *Nature*, *444*(7118), 444-454. doi: 10.1038/nature05329
- Reiter, L. T., Murakami, T., Koeuth, T., Pentao, L., Muzny, D. M., Gibbs, R. A., & Lupski, J. R. (1996). A recombination hotspot responsible for two inherited peripheral neuropathies is located near a mariner transposon-like element. *Nat Genet*, *12*(3), 288-297. doi: 10.1038/ng0396-288
- Riley, B., Williamson, M., Collier, D., Wilkie, H., & Makoff, A. (2002). A 3-Mb map of a large Segmental duplication overlapping the alpha7-nicotinic acetylcholine receptor gene (CHRNA7) at human 15q13-q14. *Genomics*, *79*(2), 197-209. doi: 10.1006/geno.2002.6694
- Risch, N., & Merikangas, K. (1996). The future of genetic studies of complex human diseases. *Science*, *273*(5281), 1516-1517.
- Ropers, H. H. (2008). Genetics of intellectual disability. *Curr Opin Genet Dev*, *18*(3), 241-250. doi: 10.1016/j.gde.2008.07.008
- Ropers, H. H. (2010). Genetics of early onset cognitive impairment. *Annu Rev Genomics Hum Genet*, *11*, 161-187. doi: 10.1146/annurev-genom-082509-141640
- Ropers, H. H., & Hamel, B. C. (2005). X-linked mental retardation. *Nat Rev Genet*, *6*(1), 46-57. doi: 10.1038/nrg1501
- Rousseau, F., Rouillard, P., Morel, M. L., Khandjian, E. W., & Morgan, K. (1995). Prevalence of carriers of premutation-size alleles of the FMRI gene--and implications for the population genetics of the fragile X syndrome. *Am J Hum Genet*, *57*(5), 1006-1018.
- Rubnitz, J., & Subramani, S. (1984). The minimum amount of homology required for homologous recombination in mammalian cells. *Mol Cell Biol*, *4*(11), 2253-2258.
- Rujescu, D., Ingason, A., Cichon, S., Pietilainen, O. P., Barnes, M. R., Toulopoulou, T., . . . Collier, D. A. (2009). Disruption of the neurexin 1 gene is associated with schizophrenia. *Hum Mol Genet*, *18*(5), 988-996. doi: 10.1093/hmg/ddn351
- Sadakata, T., & Furuichi, T. (2009). Developmentally regulated Ca<sup>2+</sup>-dependent activator protein for secretion 2 (CAPS2) is involved in BDNF secretion and is associated with autism susceptibility. *Cerebellum*, *8*(3), 312-322. doi: 10.1007/s12311-009-0097-5
- Sadakata, T., & Furuichi, T. (2010). Ca(2+)-dependent activator protein for secretion 2 and autistic-like phenotypes. *Neurosci Res*, *67*(3), 197-202. doi: 10.1016/j.neures.2010.03.006
- Sadakata, T., Kakegawa, W., Mizoguchi, A., Washida, M., Katoh-Semba, R., Shutoh, F., . . . Furuichi, T. (2007). Impaired cerebellar development and function in mice lacking CAPS2, a protein involved in neurotrophin release. *J Neurosci*, *27*(10), 2472-2482. doi: 10.1523/JNEUROSCI.2279-06.2007
- Sadakata, T., Mizoguchi, A., Sato, Y., Katoh-Semba, R., Fukuda, M., Mikoshiba, K., & Furuichi, T. (2004). The secretory granule-associated protein CAPS2 regulates neurotrophin release and cell survival. *J Neurosci*, *24*(1), 43-52. doi: 10.1523/JNEUROSCI.2528-03.2004
- Sadakata, T., Shinoda, Y., Oka, M., Sekine, Y., & Furuichi, T. (2013). Autistic-like behavioral phenotypes in a mouse model with copy number variation of the CAPS2/CADPS2 gene. *FEBS Lett*, *587*(1), 54-59. doi: 10.1016/j.febslet.2012.10.047
- Sadakata, T., Shinoda, Y., Sato, A., Iguchi, H., Ishii, C., Matsuo, M., . . . Furuichi, T. (2013). Mouse models of mutations and variations in autism spectrum disorder-associated genes: mice expressing Caps2/Cadps2 copy number and alternative splicing variants. *Int J Environ Res Public Health*, *10*(12), 6335-6353. doi: 10.3390/ijerph10126335
- Sadakata, T., Washida, M., Iwayama, Y., Shoji, S., Sato, Y., Ohkura, T., . . . Furuichi, T. (2007). Autistic-like phenotypes in Cadps2-knockout mice and aberrant CADPS2 splicing in autistic patients. *J Clin Invest*, *117*(4), 931-943. doi: 10.1172/JCI29031

- Sanders, S. J., Ercan-Sencicek, A. G., Hus, V., Luo, R., Murtha, M. T., Moreno-De-Luca, D., . . . State, M. W. (2011). Multiple recurrent de novo CNVs, including duplications of the 7q11.23 Williams syndrome region, are strongly associated with autism. *Neuron*, *70*(5), 863-885. doi: 10.1016/j.neuron.2011.05.002
- Sanders, S. J., Murtha, M. T., Gupta, A. R., Murdoch, J. D., Raubeson, M. J., Willsey, A. J., . . . State, M. W. (2012). De novo mutations revealed by whole-exome sequencing are strongly associated with autism. *Nature*, *485*(7397), 237-241. doi: 10.1038/nature10945
- Sato, D., Lionel, A. C., Leblond, C. S., Prasad, A., Pinto, D., Walker, S., . . . Scherer, S. W. (2012). SHANK1 Deletions in Males with Autism Spectrum Disorder. *Am J Hum Genet*, *90*(5), 879-887. doi: 10.1016/j.ajhg.2012.03.017
- Savelyeva, L., Sagulenko, E., Schmitt, J. G., & Schwab, M. (2006). Low-frequency common fragile sites: link to neuropsychiatric disorders? *Cancer Lett*, *232*(1), 58-69. doi: 10.1016/j.canlet.2005.08.033
- Scheiffele, P., Fan, J., Choih, J., Fetter, R., & Serafini, T. (2000). Neuroligin expressed in nonneuronal cells triggers presynaptic development in contacting axons. *Cell*, *101*(6), 657-669.
- Schellenberg, G. D., Dawson, G., Sung, Y. J., Estes, A., Munson, J., Rosenthal, E., . . . Wijsman, E. M. (2006). Evidence for multiple loci from a genome scan of autism kindreds. *Mol Psychiatry*, *11*(11), 1049-1060, 1979. doi: 10.1038/sj.mp.4001874
- Schieve, L. A., Gonzalez, V., Boulet, S. L., Visser, S. N., Rice, C. E., Van Naarden Braun, K., & Boyle, C. A. (2012). Concurrent medical conditions and health care use and needs among children with learning and behavioral developmental disabilities, National Health Interview Survey, 2006-2010. *Res Dev Disabil*, *33*(2), 467-476. doi: 10.1016/j.ridd.2011.10.008
- Schilström, B., Fagerquist, M. V., Zhang, X., Hertel, P., Panagis, G., Nomikos, G. G., & Svensson, T. H. (2000). Putative role of presynaptic alpha7\* nicotinic receptors in nicotine stimulated increases of extracellular levels of glutamate and aspartate in the ventral tegmental area. *Synapse*, *38*(4), 375-383. doi: 10.1002/1098-2396(20001215)38:4<375::AID-SYN2>3.0.CO;2-Y
- Schinzel, A. A., Brecevic, L., Bernasconi, F., Binkert, F., Berthet, F., Wuilloud, A., & Robinson, W. P. (1994). Intrachromosomal triplication of 15q11-q13. *J Med Genet*, *31*(10), 798-803.
- Schneider, E., Mayer, S., El Hajj, N., Jensen, L. R., Kuss, A. W., Zischler, H., . . . Haaf, T. (2012). Methylation and expression analyses of the 7q autism susceptibility locus genes MEST, COPG2, and TSGA14 in human and anthropoid primate cortices. *Cytogenet Genome Res*, *136*(4), 278-287. doi: 10.1159/000337298
- Schouten, J. P., McElgunn, C. J., Waaijer, R., Zwiijnenburg, D., Diepvens, F., & Pals, G. (2002). Relative quantification of 40 nucleic acid sequences by multiplex ligation-dependent probe amplification. *Nucleic Acids Res*, *30*(12), e57.
- Scott-Van Zeeland, A. A., Abrahams, B. S., Alvarez-Retuerto, A. I., Sonnenblick, L. I., Rudie, J. D., Ghahremani, D., . . . Bookheimer, S. Y. (2010). Altered functional connectivity in frontal lobe circuits is associated with variation in the autism risk gene CNTNAP2. *Sci Transl Med*, *2*(56), 56ra80. doi: 10.1126/scitranslmed.3001344
- Sebat, J., Lakshmi, B., Malhotra, D., Troge, J., Lese-Martin, C., Walsh, T., . . . Wigler, M. (2007). Strong association of de novo copy number mutations with autism. *Science*, *316*(5823), 445-449. doi: 10.1126/science.1138659
- Sebat, J., Lakshmi, B., Troge, J., Alexander, J., Young, J., Lundin, P., . . . Wigler, M. (2004). Large-scale copy number polymorphism in the human genome. *Science*, *305*(5683), 525-528. doi: 10.1126/science.1098918

- Sehgal, R. N., Gumbiner, B. M., & Reichardt, L. F. (1997). Antagonism of cell adhesion by an alpha-catenin mutant, and of the Wnt-signaling pathway by alpha-catenin in *Xenopus* embryos. *J Cell Biol*, *139*(4), 1033-1046.
- Seng, S., Avraham, H. K., Jiang, S., Venkatesh, S., & Avraham, S. (2006). KLHL1/MRP2 mediates neurite outgrowth in a glycogen synthase kinase 3beta-dependent manner. *Mol Cell Biol*, *26*(22), 8371-8384. doi: 10.1128/MCB.02167-05
- Shaikh, T. H., Gai, X., Perin, J. C., Glessner, J. T., Xie, H., Murphy, K., . . . Hakonarson, H. (2009). High-resolution mapping and analysis of copy number variations in the human genome: a data resource for clinical and research applications. *Genome Res*, *19*(9), 1682-1690. doi: 10.1101/gr.083501.108
- Shao, Y., Raiford, K. L., Wolpert, C. M., Cope, H. A., Ravan, S. A., Ashley-Koch, A. A., . . . Pericak-Vance, M. A. (2002). Phenotypic homogeneity provides increased support for linkage on chromosome 2 in autistic disorder. *Am J Hum Genet*, *70*(4), 1058-1061. doi: 10.1086/339765
- Sharp, A. J., Locke, D. P., McGrath, S. D., Cheng, Z., Bailey, J. A., Vallente, R. U., . . . Eichler, E. E. (2005). Segmental duplications and copy-number variation in the human genome. *Am J Hum Genet*, *77*(1), 78-88. doi: 10.1086/431652
- Sharp, A. J., Mefford, H. C., Li, K., Baker, C., Skinner, C., Stevenson, R. E., . . . Eichler, E. E. (2008). A recurrent 15q13.3 microdeletion syndrome associated with mental retardation and seizures. *Nat Genet*, *40*(3), 322-328. doi: 10.1038/ng.93
- Shinawi, M., Schaaf, C. P., Bhatt, S. S., Xia, Z., Patel, A., Cheung, S. W., . . . Stankiewicz, P. (2009). A small recurrent deletion within 15q13.3 is associated with a range of neurodevelopmental phenotypes. *Nat Genet*, *41*(12), 1269-1271. doi: 10.1038/ng.481
- Shinoda, Y., Sadakata, T., Nakao, K., Katoh-Semba, R., Kinameri, E., Furuya, A., . . . Furuichi, T. (2011). Calcium-dependent activator protein for secretion 2 (CAPS2) promotes BDNF secretion and is critical for the development of GABAergic interneuron network. *Proc Natl Acad Sci U S A*, *108*(1), 373-378. doi: 10.1073/pnas.1012220108
- Simpson, J. T., Wong, K., Jackman, S. D., Schein, J. E., Jones, S. J., & Birol, I. (2009). ABySS: a parallel assembler for short read sequence data. *Genome Res*, *19*(6), 1117-1123. doi: 10.1101/gr.089532.108
- Sinkus, M. L., Lee, M. J., Gault, J., Logel, J., Short, M., Freedman, R., . . . Leonard, S. (2009). A 2-base pair deletion polymorphism in the partial duplication of the alpha7 nicotinic acetylcholine gene (CHRFAM7A) on chromosome 15q14 is associated with schizophrenia. *Brain Res*, *1291*, 1-11. doi: 10.1016/j.brainres.2009.07.041
- Skuse, D. H. (2007). Rethinking the nature of genetic vulnerability to autistic spectrum disorders. *Trends Genet*, *23*(8), 387-395. doi: 10.1016/j.tig.2007.06.003
- Smith, D. I., Zhu, Y., McAvoy, S., & Kuhn, R. (2006). Common fragile sites, extremely large genes, neural development and cancer. *Cancer Lett*, *232*(1), 48-57. doi: 10.1016/j.canlet.2005.06.049
- Speicher, M. R., & Carter, N. P. (2005). The new cytogenetics: blurring the boundaries with molecular biology. *Nat Rev Genet*, *6*(10), 782-792. doi: 10.1038/nrg1692
- Speidel, D., Varoqueaux, F., Enk, C., Nojiri, M., Grishanin, R. N., Martin, T. F., . . . Reim, K. (2003). A family of Ca<sup>2+</sup>-dependent activator proteins for secretion: comparative analysis of structure, expression, localization, and function. *J Biol Chem*, *278*(52), 52802-52809. doi: 10.1074/jbc.M304727200
- Stankiewicz, P., & Lupski, J. R. (2002). Genome architecture, rearrangements and genomic disorders. *Trends Genet*, *18*(2), 74-82.

- Stefansson, H., Rujescu, D., Cichon, S., Pietiläinen, O. P., Ingason, A., Steinberg, S., . . . GROUP. (2008). Large recurrent microdeletions associated with schizophrenia. *Nature*, *455*(7210), 232-236. doi: 10.1038/nature07229
- Stephens, S. H., Logel, J., Barton, A., Franks, A., Schultz, J., Short, M., . . . Leonard, S. (2009). Association of the 5'-upstream regulatory region of the alpha7 nicotinic acetylcholine receptor subunit gene (CHRNA7) with schizophrenia. *Schizophr Res*, *109*(1-3), 102-112. doi: 10.1016/j.schres.2008.12.017
- Stewart, A. F., Dandona, S., Chen, L., Assogba, O., Belanger, M., Ewart, G., . . . Roberts, R. (2009). Kinesin family member 6 variant Trp719Arg does not associate with angiographically defined coronary artery disease in the Ottawa Heart Genomics Study. *J Am Coll Cardiol*, *53*(16), 1471-1472. doi: 10.1016/j.jacc.2008.12.051
- Szafranski, P., Schaaf, C. P., Person, R. E., Gibson, I. B., Xia, Z., Mahadevan, S., . . . Stankiewicz, P. (2010). Structures and molecular mechanisms for common 15q13.3 microduplications involving CHRNA7: benign or pathological? *Hum Mutat*, *31*(7), 840-850. doi: 10.1002/humu.21284
- Szatmari, P., Paterson, A. D., Zwaigenbaum, L., Roberts, W., Brian, J., Liu, X. Q., . . . Meyer, K. J. (2007). Mapping autism risk loci using genetic linkage and chromosomal rearrangements. *Nat Genet*, *39*(3), 319-328. doi: 10.1038/ng1985
- Teplova, M., Yuan, Y. R., Phan, A. T., Malinina, L., Ilin, S., Teplov, A., & Patel, D. J. (2006). Structural basis for recognition and sequestration of UUU(OH) 3' termini of nascent RNA polymerase III transcripts by La, a rheumatic disease autoantigen. *Mol Cell*, *21*(1), 75-85. doi: 10.1016/j.molcel.2005.10.027
- Toth-Fejel, S., Magenis, R. E., Leff, S., Brown, M. G., Comegys, B., Lawce, H., . . . Olson, S. (1995). Prenatal diagnosis of chromosome 15 abnormalities in the Prader-Willi/Angelman syndrome region by traditional and molecular cytogenetics. *Am J Med Genet*, *55*(4), 444-452. doi: 10.1002/ajmg.1320550411
- Trikalinos, T. A., Karvouni, A., Zintzaras, E., Ylisaukko-oja, T., Peltonen, L., Järvelä, I., & Ioannidis, J. P. (2006). A heterogeneity-based genome search meta-analysis for autism-spectrum disorders. *Mol Psychiatry*, *11*(1), 29-36. doi: 10.1038/sj.mp.4001750
- Turner, D. J., Miretti, M., Rajan, D., Fiegler, H., Carter, N. P., Blayney, M. L., . . . Hurles, M. E. (2008). Germline rates of de novo meiotic deletions and duplications causing several genomic disorders. *Nat Genet*, *40*(1), 90-95. doi: 10.1038/ng.2007.40
- Tuzun, E., Sharp, A. J., Bailey, J. A., Kaul, R., Morrison, V. A., Pertz, L. M., . . . Eichler, E. E. (2005). Fine-scale structural variation of the human genome. *Nat Genet*, *37*(7), 727-732. doi: 10.1038/ng1562
- Uchida, N., Shimamura, K., Miyatani, S., Copeland, N. G., Gilbert, D. J., Jenkins, N. A., & Takeichi, M. (1994). Mouse alpha N-catenin: two isoforms, specific expression in the nervous system, and chromosomal localization of the gene. *Dev Biol*, *163*(1), 75-85. doi: 10.1006/dbio.1994.1124
- Ullmann, R., Turner, G., Kirchhoff, M., Chen, W., Tonge, B., Rosenberg, C., . . . Ropers, H. H. (2007). Array CGH identifies reciprocal 16p13.1 duplications and deletions that predispose to autism and/or mental retardation. *Hum Mutat*, *28*(7), 674-682. doi: 10.1002/humu.20546
- Vaags, A. K., Lionel, A. C., Sato, D., Goodenberger, M., Stein, Q. P., Curran, S., . . . Scherer, S. W. (2012). Rare deletions at the neurexin 3 locus in autism spectrum disorder. *Am J Hum Genet*, *90*(1), 133-141. doi: 10.1016/j.ajhg.2011.11.025
- Vaillend, C., Poirier, R., & Laroche, S. (2008). Genes, plasticity and mental retardation. *Behav Brain Res*, *192*(1), 88-105. doi: 10.1016/j.bbr.2008.01.009
- van Bokhoven, H. (2011). Genetic and epigenetic networks in intellectual disabilities. *Annu Rev Genet*, *45*, 81-104. doi: 10.1146/annurev-genet-110410-132512

- van Bon, B. W., Mefford, H. C., Menten, B., Koolen, D. A., Sharp, A. J., Nillesen, W. M., . . . de Vries, B. B. (2009). Further delineation of the 15q13 microdeletion and duplication syndromes: a clinical spectrum varying from non-pathogenic to a severe outcome. *J Med Genet*, *46*(8), 511-523. doi: 10.1136/jmg.2008.063412
- van der Zwaag, B., Staal, W. G., Hochstenbach, R., Poot, M., Spierenburg, H. A., de Jonge, M. V., . . . Burbach, J. P. (2010). A co-segregating microduplication of chromosome 15q11.2 pinpoints two risk genes for autism spectrum disorder. *Am J Med Genet B Neuropsychiatr Genet*, *153B*(4), 960-966. doi: 10.1002/ajmg.b.31055
- Vanpoucke, G., Nollet, F., Tejpar, S., Cassiman, J. J., & van Roy, F. (2002). The human alphaE-catenin gene CTNNA1: mutational analysis and rare occurrence of a truncated splice variant. *Biochim Biophys Acta*, *1574*(3), 262-268.
- Veenstra-VanderWeele, J., & Cook, E. H. (2004). Molecular genetics of autism spectrum disorder. *Mol Psychiatry*, *9*(9), 819-832. doi: 10.1038/sj.mp.4001505
- Vijayaraghavan, S., Pugh, P. C., Zhang, Z. W., Rathouz, M. M., & Berg, D. K. (1992). Nicotinic receptors that bind alpha-bungarotoxin on neurons raise intracellular free Ca<sup>2+</sup>. *Neuron*, *8*(2), 353-362.
- Viscidi, E. W., Triche, E. W., Pescosolido, M. F., McLean, R. L., Joseph, R. M., Spence, S. J., & Morrow, E. M. (2013). Clinical characteristics of children with autism spectrum disorder and co-occurring epilepsy. *PLoS One*, *8*(7), e67797. doi: 10.1371/journal.pone.0067797
- Waldman, A. S., & Liskay, R. M. (1988). Dependence of intrachromosomal recombination in mammalian cells on uninterrupted homology. *Mol Cell Biol*, *8*(12), 5350-5357.
- Walters, R. G., Jacquemont, S., Valsesia, A., de Smith, A. J., Martinet, D., Andersson, J., . . . Beckmann, J. S. (2010). A new highly penetrant form of obesity due to deletions on chromosome 16p11.2. *Nature*, *463*(7281), 671-675. doi: 10.1038/nature08727
- Wang, K., Li, M., Hadley, D., Liu, R., Glessner, J., Grant, S. F., . . . Bucan, M. (2007). PennCNV: an integrated hidden Markov model designed for high-resolution copy number variation detection in whole-genome SNP genotyping data. *Genome Res*, *17*(11), 1665-1674. doi: 10.1101/gr.6861907
- Wang, K., Zhang, H., Ma, D., Bucan, M., Glessner, J. T., Abrahams, B. S., . . . Hakonarson, H. (2009). Common genetic variants on 5p14.1 associate with autism spectrum disorders. *Nature*, *459*(7246), 528-533. doi: 10.1038/nature07999
- Weiss, L. A., Arking, D. E., Daly, M. J., Chakravarti, A., & Consortium, G. D. P. o. J. H. t. A. (2009). A genome-wide linkage and association scan reveals novel loci for autism. *Nature*, *461*(7265), 802-808. doi: 10.1038/nature08490
- Weterings, E., & van Gent, D. C. (2004). The mechanism of non-homologous end-joining: a synopsis of synapsis. *DNA Repair (Amst)*, *3*(11), 1425-1435. doi: 10.1016/j.dnarep.2004.06.003
- Williams, E., Thomas, K., Sidebotham, H., & Emond, A. (2008). Prevalence and characteristics of autistic spectrum disorders in the ALSPAC cohort. *Dev Med Child Neurol*, *50*(9), 672-677. doi: 10.1111/j.1469-8749.2008.03042.x
- Williams, S. K., Spence, H. J., Rodgers, R. R., Ozanne, B. W., Fitzgerald, U., & Barnett, S. C. (2005). Role of Mayven, a kelch-related protein in oligodendrocyte process formation. *J Neurosci Res*, *81*(5), 622-631. doi: 10.1002/jnr.20588
- Xu, L., Wei, Y., Reboul, J., Vaglio, P., Shin, T. H., Vidal, M., . . . Harper, J. W. (2003). BTB proteins are substrate-specific adaptors in an SCF-like modular ubiquitin ligase containing CUL-3. *Nature*, *425*(6955), 316-321. doi: 10.1038/nature01985
- Yasui, D. H., Scoles, H. A., Horike, S., Meguro-Horike, M., Dunaway, K. W., Schroeder, D. I., & Lasalle, J. M. (2011). 15q11.2-13.3 chromatin analysis reveals epigenetic regulation of

- CHRNA7 with deficiencies in Rett and autism brain. *Hum Mol Genet*, 20(22), 4311-4323. doi: 10.1093/hmg/ddr357
- Ye, H., Liu, J., & Wu, J. Y. (2010). Cell adhesion molecules and their involvement in autism spectrum disorder. *Neurosignals*, 18(2), 62-71. doi: 10.1159/000322543
- Ye, K., Schulz, M. H., Long, Q., Apweiler, R., & Ning, Z. (2009). Pindel: a pattern growth approach to detect break points of large deletions and medium sized insertions from paired-end short reads. *Bioinformatics*, 25(21), 2865-2871. doi: 10.1093/bioinformatics/btp394
- Yoon, S., Xuan, Z., Makarov, V., Ye, K., & Sebat, J. (2009). Sensitive and accurate detection of copy number variants using read depth of coverage. *Genome Res*, 19(9), 1586-1592. doi: 10.1101/gr.092981.109
- Young, D. J., Bebbington, A., Anderson, A., Ravine, D., Ellaway, C., Kulkarni, A., . . . Leonard, H. (2008). The diagnosis of autism in a female: could it be Rett syndrome? *Eur J Pediatr*, 167(6), 661-669. doi: 10.1007/s00431-007-0569-x
- Yu, T. W., Chahrour, M. H., Coulter, M. E., Jiralerspong, S., Okamura-Ikeda, K., Ataman, B., . . . Walsh, C. A. (2013). Using whole-exome sequencing to identify inherited causes of autism. *Neuron*, 77(2), 259-273. doi: 10.1016/j.neuron.2012.11.002
- Yuan, W. C., Lee, Y. R., Huang, S. F., Lin, Y. M., Chen, T. Y., Chung, H. C., . . . Chen, R. H. (2011). A Cullin3-KLHL20 Ubiquitin ligase-dependent pathway targets PML to potentiate HIF-1 signaling and prostate cancer progression. *Cancer Cell*, 20(2), 214-228. doi: 10.1016/j.ccr.2011.07.008
- Yuen, R. K., Thiruvahindrapuram, B., Merico, D., Walker, S., Tammimies, K., Hoang, N., . . . Scherer, S. W. (2015). Whole-genome sequencing of quartet families with autism spectrum disorder. *Nat Med*, 21(2), 185-191. doi: 10.1038/nm.3792
- Zahir, F. R., Baross, A., Delaney, A. D., Eydoux, P., Fernandes, N. D., Pugh, T., . . . Friedman, J. M. (2008). A patient with vertebral, cognitive and behavioural abnormalities and a de novo deletion of NRXN1alpha *J Med Genet* (Vol. 45, pp. 239-243). England.
- Zappella, M. (2010). Autistic regression with and without EEG abnormalities followed by favourable outcome. *Brain Dev*, 32(9), 739-745. doi: 10.1016/j.braindev.2010.05.004
- Zollman, S., Godt, D., Privé, G. G., Couderc, J. L., & Laski, F. A. (1994). The BTB domain, found primarily in zinc finger proteins, defines an evolutionarily conserved family that includes several developmentally regulated genes in *Drosophila*. *Proc Natl Acad Sci U S A*, 91(22), 10717-10721.