UNIVERSITÀ DEGLI STUDI DI PALERMO

Dottorato di Ricerca in Genomica e Proteomica nella Ricerca Oncologica ed Endocrino-Metabolica – XXII ciclo

Settore Scientifico Disciplinare "Genetica" – BIO18

Searching for new genes in Autism Spectrum Disorders through high-resolution array-Comparative Genomic Hybridization

Authors

PhD Student: dr. Michele Falco

Dr. Marco Fichera (tutor)

Genetic Diagnosis Laboratory, IRCCS Oasi Maria SS, Troina (EN), Italy

Prof. Valentino Romano (co-tutor)

Dip. di Fisica e Tecnologie Relative, Università degli Studi di Palermo, Palermo, Italy

Anno Accademico 2007/08

Index

Introduction

Autism Spectrum Disorders (ASDs) are a spectrum of psychiatric conditions characterized by three major abnormal domains: limited or absent verbal and non verbal communication, a lack of reciprocal social communication or responsiveness, and repetitive, stereotyped and ritualistic patterns of interests and behaviors [1, 2]. The term "autism spectrum disorders" is used as shorthand to refer to any patient meeting these diagnostic criteria. But beyond this unifying definition, an extreme degree of clinical heterogeneity lies, ranging from profound to moderate impairments, always with functional disabilities. ASDs in fact include Autism itself (the prototypic pervasive developmental disorder, OMIM 209850); Asperger syndrome (see ASPG1; OMIM 608638); the pervasive developmental disorder, not otherwise specified (PDD-NOS); and Rett Syndrome (OMIM 312750). So, autism is not a single disease, rather a complex phenotype thought to be caused by different types of defects in common pathways, producing similar behavioral phenotypes. ASDs are usually diagnosed within three years, and have a prevalence of about 1/100, but closer to 1/300 for typical autism [3]. ASDs are more common in male than female patients, with a 4:1 ratio [4, 5]. Twin and family studies have conclusively established the strong genetic contribution to ASDs, with

concordance rates of 82–92% in monozygotic twins versus 1–10% in dizygotic twins.

Intellectual disabilities coexist in approximately two-thirds of individuals with ASDs, except for Asperger syndrome, in which mental retardation is conspicuously absent [6].

In approximately 10–25% of the affected individuals, autism is 'syndromic', occurring in children with a known genetic (such as fragile X syndrome, tuberous sclerosis, neurofibromatosis) or environmental (valproic syndrome, or brain herpes simplex infection) disorder [4, 6].Similarly to other child psychiatric disorders, two opposite hypotheses have been proposed to elucidate the genetics of ASDs. The early inability to detect mutations in single genes, that could explain ASDs as Mendelian inheritance diseases, prompted some geneticists to consider autism as a polygenic trait, and led to the ascendance of the "common variant common disease" hypothesis. This model supposes that any given common allele (defined as 5% or greater population frequency) is likely to be neither necessary nor sufficient to lead to the clinical phenotype, but in combination with other risk alleles would contribute to the emergence of the disease. Therefore, model-free linkage studies, such as affected sib-pair analyses, were performed to identify susceptibility genes. Although many genomic regions were detected, only a restricted number of loci were replicated in independent scans (e.g. 7q31 and 17q11); probably, an

explanation for the lack of definitive results is the much greater than anticipated degree of locus heterogeneity in ASDs. Afterward, the technological feasibility of genotyping one or a small number of common alleles in large cohorts of patients led to the publication of several studies, presenting interesting findings. These included the identification of EN2 [7], the oncogene MET [8], and CNTNAP2 [9] as autism-susceptibility genes. Finally, genome wide association studies (GWAS) with high-density Single Nucleotide Polymorphism (SNP) arrays, emerged as the gold standard for the identification of common small-effect alleles, have been applied to autism. Association studies provide major advantages compared with linkage studies: first, studies can include large samples of subjects because they are not restricted to multiplex families with two or more affected children. Second, the genomic regions associated with the trait are much narrower than in linkage studies, because linkage disequilibrium between relatively close genomic regions is less likely to affect the association. In addition, SNP arrays can be used to detect structural variants such as Copy Number Variants (CNVs), which may also contribute to the development of the disease [10], as we'll discuss later. Unfortunately, the three largest GWAS, performed on more than 1000 patients in each study, failed to detect the same genes [11-13]. The first one identified a significant association of ASDs to an intragenic region (5p14.1) between Cadherin 9 and Cadherin 10 genes. The second study found an association to a SNP

near Semaphorin 5A, and the third evidenced an association to the MACROD2 gene.

Despite the GWAS has failed to identify a significant proportion of the genetic risk for autism, the polygenic model cannot be excluded for several reasons. First, it is now well established that the clinical outcome of most "classical" monogenic disorders is actually modulated by additional genetic variations. Second, the experience in other polygenic traits, blood pressure and height for example, has shown that the effect of a single common variant on the phenotype can be very low (odds ratio <2) and can require a large sample size (>15 000 subjects) in order to be detected [14]; so far, the current GWAS scans in ASDs – the largest were performed on cohorts of less than 2000 subjects – may have missed the necessary statistical power to detect such common alleles with low effect. Finally, the polygenic heredity of autism could be explained by the presence of both common and rare variants: the number of deleterious mutations within the genome of one individual remains difficult to establish, but it is likely that a combination of multiple common and rare variants (such as small CNVs or coding and regulatory variations, not detect by GWAS) could affect specific biological pathways and therefore increase the risk for developing autism.

So, the converging results of both GWAS and linkage studies that have discovered only a small proportion of the genetic risk for common

disorders [14], the increasing relevance of rare variant findings in other neuropsychiatric area, the likely pathogenic involvement of CNVs and the identification of apparently monogenic forms of autism, each time affecting a limited number of patients (1–2% for the most replicated genes), have led some Authors to opt for the hypothesis of "rare variant common disease" (RVCD) to explain the etiology of ASDs. By this model, the "autisms" are a very heterogeneous group of neuropsychiatric phenotypes caused by rare variations (e.g., point mutations or CNVs) in different genes involved in common pathways (in this model, a single highly penetrant mutation would be sufficient to produce autism).

In recent years, enormous progress has been achieved in detecting genomic structural variants (i.e., CNVs) more or less rare [15], and at least nine studies have searched for such genomic imbalances in patients with ASDs [16-24]. Unfortunately, differences in the criteria used for patients' selection, genotyping methodologies and algorithms applied to detect deletions and duplications make the comparison of these results difficult. Nevertheless, it seems that there is a significant increase in rare inherited and *de novo* CNVs in the ASDs population compared with the general population. The overall rate of *de novo* CNVs in ASDs could range from 5–10% compared with 1% in the general population [20]. Not surprisingly, the frequency of rare and *de novo* CNVs increases for families with only one affected child (7–10%) compared with families with two or more

affected children (2–3%) [20]. The presence of dysmorphic features in the patient also increases the odds of detecting a rare or *de novo* CNV in up to 27.5% [17]. Based on current findings, the majority of the CNVs apparently affect only one copy of the gene (which can be either deleted or duplicated), suggesting that abnormal gene dosage or expression might play a key role in susceptibility to ASDs [25].

To date, despite the large efforts spent to identify common and rare alleles, the number of genes definitively accepted as high risk factors remains quite small. Nevertheless, lists of candidate genes are proliferating exponentially: for example, a list of 219 genes is available at AutDB, a public, web-based database for autism research (http://www.mindspec.org/autdb.html). Obviously, most of these genes remain only "candidates" because their association to autism was not confirmed by replication and/or validated by functional studies.

These candidate genes generally participate to three major biological processes, all related to the development and functioning of neuronal circuits.

(a) At the synapse, the following proteins have been associated to ASDs: adhesion molecules (cadherins, protocadherins, neuroligins and neurexins) involved in presynaptic assembly; scaffold proteins (SHANK and DLGAP2), which provide a link between postsynaptic membrane proteins and actin; receptors for glutamate

(GLUR) and GABA (GABAR), which play a major role in excitatory and inhibitory currents, respectively; FMRP, which regulates local translation of synaptic proteins, together with PTEN, NF1, TSC1 and TSC2, and c-MET; UBE3A, which participates in the targeting of proteins to the proteasome.

(b) In the neuron nucleus, the Methyl-CpG binding protein 2 and transcription factors like MEF2C regulate the expression of genes involved in neuronal circuit.

(c) At the node of Ranvier, contactins and CNTNAP2 organize the junction between the outgrowing axon and glia, together with cell adhesion and secreted proteins as NRCAM and SEMA5A.

Aim of the thesis

The aim of this study is to detect, through high-resolution array-CGH analysis, unreported Copy Number Variants in a cohort of 120 Italian autistic patients. After a validation of such results, through an *in silico* analysis of these variants, we expect to generate a list of new candidate genes in autism, which may be further investigated to better delineate their actual role in the etiology of ASDs.

Patients and methods

After obtaining informed written consent from parents or legal guardians, we collected blood samples from 120 ASD children (94 males, 26 female s). Individuals were included in this research project only if the following criteria were fulfilled:

- matching DSM-IV-TR criteria for PDD/ASD and scoring cut-off in Autism Diagnostic Interview-Revised [26] and/or in Autism Diagnostic Observation Schedule-Generic [27];
- being at least 3 years of age at the time of entering the research project.

The exclusion criteria were as follows:

- presence of profound mental retardation;
- history of serious head injury, encephalitis or tumors;
- presence of metabolic disease or known genetic syndrome.

All patients had a normal G-band karyotype, and were *FRAXA* (mutations or expansions) negative; subtelomeric rearrangements were also excluded.

DNA extractions were performed by salting out procedures. DNA quantity and quality were tested through NANODROP ND1000 spectrophotometer (Euroclone).

Autistic subjects were screened by array-CGH with the Agilent 60K, 180K and 244K arrays (median resolution 41.5 kb, 13 kb and 8.9 kb

respectively). Our array platforms were calibrated through a series of self-self and sex-mismatched hybridizations to experimentally establish the false positive and the false negative rates and the threshold to define copy number changes. A series of experiments using reference samples with DNA imbalances of variable but well-known sizes allowed us to check for the reliability and for the real resolution of our platform.

In addition, one male and one female (Coriell NA10851/ NA15510) DNAs, whose genomic CNV content has been well characterized by several studies, were used as male and female references, respectively. This strategy allowed us to discern between CNVs of reference DNA and those of sample DNA. Moreover, we were able to detect false negative results due to the presence of the same CNVs both in reference and sample.

In principle, 500 ng of patient and reference DNAs were double-digested with 5 units each of *RsaI* and *AluI* (Promega) for two hours at 37°C. Each digested sample was labeled by random priming (Agilent Genomic DNA Labeling Kit PLUS) for two hours at 37°C, using Cy5−dUTP for the patient DNA and Cy3−dUTP for the reference DNA. Labeled products were combined and column purified with a Microcon YM-30 filter (Millipore). After denaturation and pre-annealing with 50 µg of Cot-1 DNA (Invitrogen), DNAs were hybridized on a microarray at 65 °C with rotation for 40 hours. After two washing steps, the arrays were scanned on Agilent scanner and microarray images were analysed with Feature

Extraction software. Graphical representation of the results was obtained using the CGH-Analytics software.

The Agilent software pipeline has been developed together with a series of Quality Control (QC) parameters that give an important support to estimate if an experiment is sufficiently well-performed or need to be repeated. For each array analyzed the Feature Extraction software creates a file reporting information related to the ratio intensities, background intensities, spots distribution, out-/in-liers distributions and relative intensities and all the parameters relevant to the data analysis. Intra-experiment statistical features present in the software pipeline include: number and spatial distribution of outliers, net signal statistics, local background inliers, foreground surface fit, plot of background-corrected signals, spatial distribution of Up and Down regulated features and plot of LogRatio versus average Log Signals. Inter-experiment statistical features include: reproducibility statistics, microarray uniformity, reproducibility plot (spike-ins), spike-in LogRatio statistics and LogRatio plot for spike-ins.

The CNVs reported in our patients were compared with those reported in the most important database of genomic variants identified in normal controls, the "Toronto database" (http://projects.tcag.ca/variation/), and a database created in our laboratory (http://dbcnv.oasi.en.it/gvarianti/index.php), which collects CNVs data from the most important Italian Institutions studying MR and autistic

patients through array CGH platforms. Array-CGH findings of particular interest (i.e. non-polymorphic CNVs) were confirmed by alternative techniques, such as Multiplex Ligation-dependent Probe Amplification (MLPA) or Fluorescent In Situ Hybridization (FISH) assays, and tested on parents when possible.

We also referred to three international databases which collect information on genes and CNVs associated to autism (http://www.mindspec.org/autdb.html; http://wren.bcf.ku.edu/; http://projects.tcag.ca/autism_500k/). For the *in silico* study of the genes included in the reported CNVs, we referred to PubMed (http://www.ncbi.nlm.nih.gov/pubmed), OMIM (http://www.ncbi.nlm.nih.gov/omim/) and Entrez Gene (http://www.ncbi.nlm.nih.gov/gene/) databases.

Results and Discussion

Using the above mentioned Agilent protocol, 120 autistic patients were screened through high-resolution array-CGH assays. Compared to the reference DNA, each of them showed at least one CNV, but depending on the resolution of the array - which increased during the course of the project, as new array platforms were becoming available - this number could increase to ten and more. So, hundreds of CNVs were revealed and were to be interpreted. In order to assess the potential pathogenic role of these genomic imbalances, we defined an analytical procedure, consisting of several criteria that allowed us to prioritize polymorphic, "pathological" (P), "novel" (N) and "candidate" (C) CNVs and genes. The flowchart of this procedure is summarized in figure 1, and described in detail below.

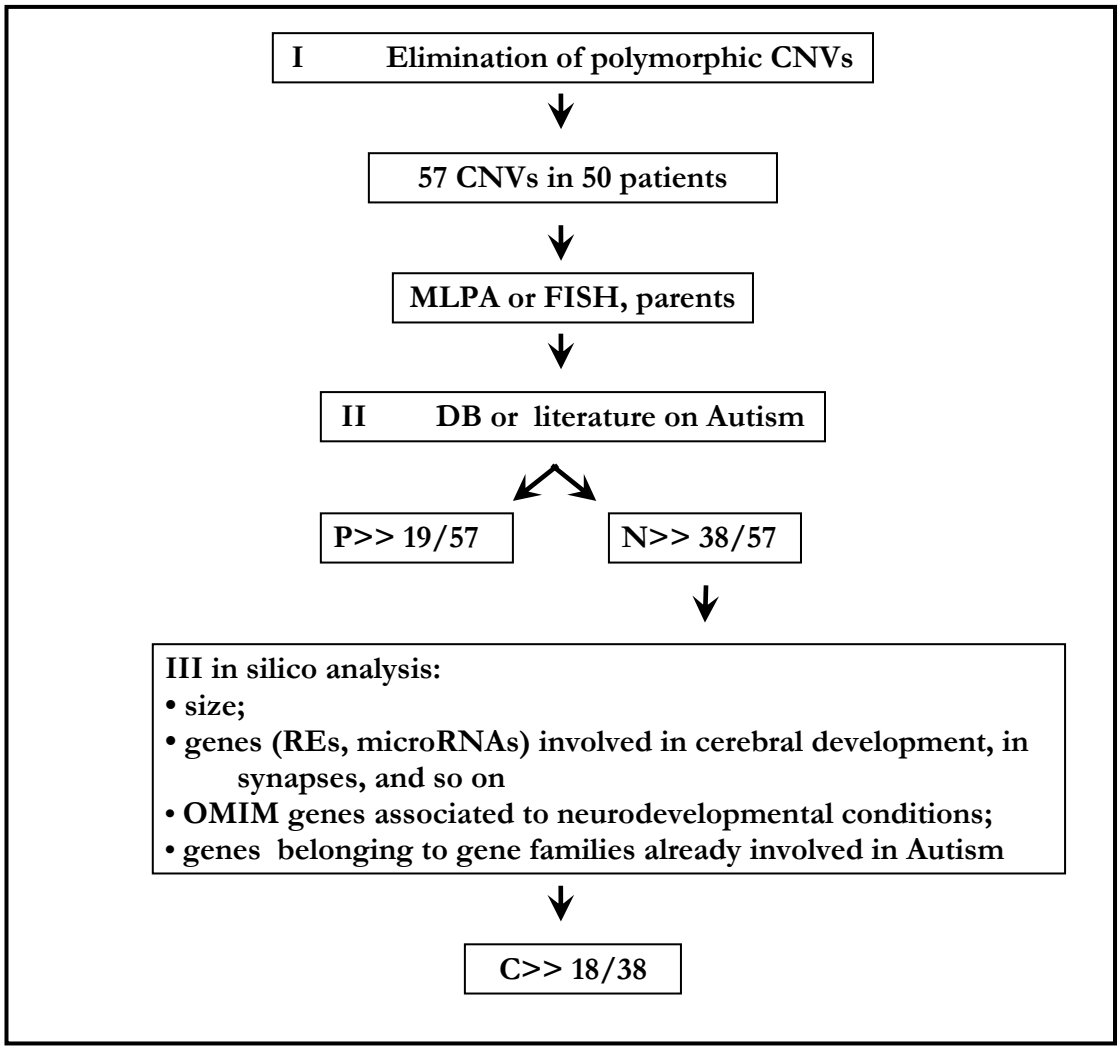

Figure 1. Procedure used to select the candidate CNVs. In roman numerals, the three selection steps are represented. P = pathological, N = novel (non polymorphic), C = candidate.

Criteria used to assess a pathogenic role of CNVs and their overlapping genes

The CNVs identified in the study were considered identical (recurrent) only if they were represented by the same probes and were of the same sign (loss or gain). As a first step, we searched for all the variants observed in the study within the Database of Genomic Variants, to eliminate from further investigations all the polymorphic variants: in so doing, we were able to select 57 different and non-polymorphic CNVs in 50 patients. 12 patients showed 2 variations each, 4 CNVs were recurrent in two patients each, one CNV presented identical in size, but in loss in one patient, and in gain in another one. The size of these variants was very variable, ranging from 11 kilobases to 3.5 megabases; 32 were deletions, 26 were duplications. These CNVs were confirmed by alternative methods, i.e. MLPA or FISH assays, to exclude false positives, and were also investigated in the parents (when possible), to establish if they were inherited or de novo events.

At this point, it was necessary to determine how to interpret these findings, and first understand which were to be considered pathogenic. So, the second selection step was the "replication": each of the 57 non-polymorphic CNVs was verified in the Autism and in the Italian databases described in the Methods, and in Pubmed, to search for the most recent data

literature; we considered a CNV pathological if it was already reported in autistic patients or (in one case, see below) if it contained one or more genes found mutated in autism. This analytical procedure enabled us to distinguish two groups of CNVs: 19 pathological, and 38 “novel” CNVs, never before associated to autism.

The data resulting from the application of this criterion and the related information are reported in Table 1, part one.

patient	gender	CNV type	heredity	loss/gain	chr	cytoband	size in Kb	genes	present in Autism DB or literature	evidence
03199	F	P	de novo	DEL	1	q42.2	18	DISC1	YES	inherited CNV
02951	M	P	F	DUP	2	p16.3	440	NRXN1	YES	mutations and CNV
03001	F	P	de novo	DEL	2	p25.2	1 Mb	SOX11	YES	de novo CNV
02875	M	P	F	DUP	3	p26.3	500	CNTN4 CNTN6	CNTN4 YES CNTN6 NO	CNV
02919	M	P	de novo	DEL	6	p21.32	128	SYNGAP1	YES	mutations and CNV
02331	**M**	**P**	**ND**	**DEL**	**6**	**q26**	**115**	**PARK2**	**YES**	**CNV**
02433	M	P	ND	DEL	7	q11.23	390	AUTS2	YES	mutations and CNV
03957	F	P	ND	DEL	7	q31	61	DOCK4	YES	CNV and SNP
03841	M	P	F	DEL	7	q35	35	CNTNAP2	YES	CNV and rare variants
03256	M	P	de novo	DEL	9	p24.3-24.2	3.5 Mb	DOCK8	YES	de novo CNV
03436	**F**	**P**	**ND**	**DEL**	**15**	**q11.2**	**425**	**NIPA1**	**YES**	**CNV**
03476	M	P	F	DEL	15	q11.2	443	NIPA1	YES	CNV
02104	M	P	de novo	DEL	22	q11.2	139	PRODH	YES	CNV
02985	F	P	de novo	DEL	22	q11.2	139	PRODH	YES	CNV
03788	**M**	**P**	**M**	**DEL**	**22**	**q11.21**	**116**	**DGCR6 PRODH**	**YES**	CNV
01579	M	P	M	DUP	22	q11.2-q11	2.5 Mb	PRODH TBX1	YES	CNV
01875	F	P	de novo	DEL	22	q13.33	25	SHANK3	YES	CNV and mutations
03233	F	P	de novo	DUP	X	p22.2	2.7 Mb	GRPR	YES	mutations
02997	**M**	**P**	**F**	**DUP**	**X**	**p22.33**	**18**	**ASMT**	**YES**	**inherited CNVs, SNP, mutations**
02051	M	P	M	DUP	X	p22.33	15	ASMT	YES	inherited CNVs, SNP, mutations
03820	**M**	**P**	**ND**	**DUP**	**X**	**p22.33**	**15**	**ASMT**	**YES**	**inherited CNVs, SNP, mutations**

Table 1 (part one). CNV type P= pathological; heredity F= from father, M= from mother, ND= not determined; in bold the patients with two novel CNVs reported; coloured lanes represent the recurrent findings.

Of these 19 CNVs, reported in 21 patients (two pairs of patients had a recurrent CNV each), 13 were deletions, 6 were duplications. However, it should be noted that some duplications, having their breakpoints within a gene, may disrupt its normal genomic structure, so one can expect a haploinsufficiency of that gene, rather than an overexpression.

8/21 patients had a *de novo* CNV, 8 subjects inherited their variant from an apparently healthy parent, and for 5 of them was not possible to investigate the parents.

Interestingly, 4 patients (02331, 03436, 03788, and 03820) had one P and one C CNVs each, one patient (02997) had one P and one N CNVs.

Two noteworthy findings:

(i) the gastrin-releasing peptide receptor has been found disrupted in an autistic girl due to a X:8 translocation [28], and then mutated in other patients, but never before it has been found involved in a CNV. It is an intriguing finding that our patient # 03233 (also a female) has a totally duplicated *GRPR* gene (see Table 1, part one) suggesting that both increased and decreased dosage of *GRPR* could be pathogenic in autistic females;

(ii) *CNTNAP2* (contactin associated protein-like 2) is a well established and strong autism gene [29]. We found a patient (# 03841 in Table 1, part one) bearing a deletion, inherited from the healthy father, encompassing only three probes, which lies

immediately proximal to the *CNTNAP2* gene (see figure 2): we suspect that regulatory sequences of this gene were lost with the deletion. Our findings on these two autism genes are unprecedented and deserve further investigation to elucidate the precise molecular mechanism altering their function.

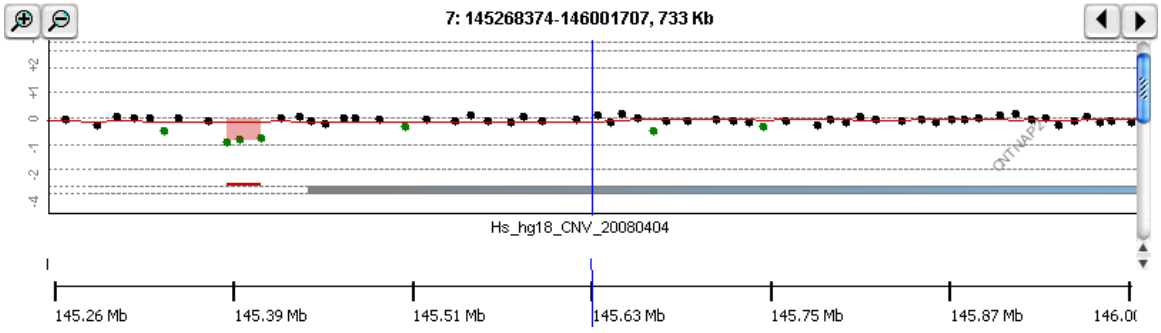

Figure 2. A 35 Kb deletion in 7q35 proximal to the CNTNAP2 mRNA (grey bar).

Criteria used to assess a candidate role of CNVs and their overlapping genes

38 CNVs were not previously described as polymorphic variants, and were never associated to autism. For the purpose of our study, they represented just the most intriguing findings, because these were the ones that could contain new genes of autism. To establish a priority order by which to submit these CNVs to further investigations, we applied the following criteria:

- CNV size: large deletions and duplications were tentatively considered as candidate pathogenic;
- positive *in silico* analyses: we performed on the gene content of the 38 CNVs an in silico study, using the databases described in methods, and we considered candidate a CNV if it contained genes expressed in the central nervous system and involved in brain development, neuronal functioning, or synaptic plasticity; moreover, OMIM genes, already associated to other neurological and/or neuropsychiatric conditions, especially those ones involving disturbances of neurodevelopment (pathological and candidate); finally, genes belonging to families whose other members have already been associated to autism. *In silico* analysis was also carried out to detect potential regulatory elements, microRNAs and

evolutionary conserved elements related to brain development and function. To this aim, we mainly used databases providing information on the expression pattern, on the sequence and functional similarities with other organisms, and on the existence of mouse knockout models. Reference databases are listed in the Methods' section.

De novo vs. inherited variant was not considered a selection criterion: in fact, though the *de novo* CNVs may represent highly penetrant candidate variations, the distinction between *de novo* and inherited CNVs was not taken into account. This choice was made to conform to the unified genetic theory for sporadic and inherited autism [30]. Indeed, according to this theory, causative CNVs may be inherited from an unaffected parent as well as *de novo* events.

Thanks to these criteria, we selected 18 CNVs as "candidate" for further investigations. The second part of Table 1 shows these CNVs, and their candidate genes. The patients listed in table are 20, having been two CNVs reported twice each.

Interestingly, 4 patients (03264, 01397, 01950, and 03799) had one C and one N CNVs each.

patient	gender	heredity	loss/gain	chr	cytoband	size in Kb	genes	Evidence for candidate
02806	M	F	DEL	1	q21.1	3.1 Mb	many	size
01995	F	F	DEL	2	q13	1.7 Mb	many	size
072	M	ND	DUP	17	q12	408	many	size
02031	M	F	DUP	13	p21.1-21.2	1.8 Mb	many, PCDH17	Size and GF
01397	**M**	**de novo**	**DEL**	**15**	**q24.1-24.2**	**2.5 Mb**	**many, SEMA7A**	**Size and GF**
02331	**M**	**ND**	**DUP**	**2**	**p12**	**325**	**CTNNA2**	**GF**
03264	**M**	**ND**	**DEL**	**11**	**q22.3**	**223**	**GRIA4**	**GF**
03820	**M**	**ND**	**DUP**	**4**	**q35.1**	**300**	**SLC25A4**	**GF**
03436	**F**	**ND**	**DUP**	**5**	**q35.3**	**56**	**GRM6**	**GF**
02993	M	M	DUP	10	q22.1	309	UNC5B SLC29A3 CDH23	GF
02993S	F	M	DUP	10	q22.1	309	UNC5B SLC29A3 CDH23	GF
03123	M	M	DEL	16	p12.1	79	HS3ST4	GF
01950	**M**	**ND**	**DUP**	**X**	**p22.33**	**11**	**ASMTL**	**GF**
0176	M	ND	DEL	X	q28	52	TMEM185A	GF
03799	**M**	**ND**	**DUP**	**X**	**q28**	**52**	**TMEM185A**	**GF**
02571	M	ND	DEL	7	p12.1	96	COBL	Neuronal morphogenesis
02572	M	ND	DEL	7	p12.1	113	COBL	Neuronal morphogenesis
02627	M	ND	DUP	6	q22.33	63	ARHGAP18	Schizophrenia
03153	F	F	DEL	7	q11.22	28	CALN1	Brain specific
03788	**M**	**M**	**DEL**	**15**	**q21.3**	**97**	**NEDD4**	**N**

Table 1 (part two). Heredity F= from father, M= from mother, ND= not determined; in bold the patients with two novel CNVs reported; coloured lanes represent the recurrent findings; 02993 and 02993S (in orange) are brother and sister; GF = gene belonging to Gene Families already associated to autism.

As one can see at the top of the table, five variants, containing many genes, were considered candidates because of their significant size. In particular, patients 02806 and 01995 both inherited the CNV from the father, and need further considerations. Two other patients, 02031 and 01397, had two CNVs that were considered "candidates" for both their size and because they contained respectively protocadherin 17 and semaphorin 7A, two genes belonging to gene families already involved in autism.

Eight other CNVs containing autism gene family members are listed in the middle of the table. The involved genes are:

- *CTNNA2* (catenin alpha 2), entirely duplicated in patient 02331;
- *GRIA4* (glutamate receptor, ionotrophic, AMPA 4), entirely deleted in patient 03264;
- the solute carrier *SLC25A4*, entirely duplicated in patient 03820;
- *GRM6* (glutamate receptor, metabotropic 6), entirely duplicated in patient 03436;
- *SLC29A3* and cadherin 23 (see figure 3);
- *HS3ST4* (heparan sulfate glucosamine 3-O-sulfotransferase 4), interrupted by a small deletion in patient 03123;
- *ASMTL* (acetylserotonin O-methyltransferase-like), an X linked gene interrupted by a very small duplication in the male patient 01950;

- *TMEM185A*, a transmembrane protein found interrupted in two identical CNVs (one deletion and one duplication) in two male patients, 0176 and 03799.

Figure 3. A 309 kb duplication in 10q22.1 in two patients, brother and sister, disrupting two genes (UNC5B and CDH23) and including the solute carrier SLC29A3.

Finally, a pathological effect due to an aberrant dosage could be supposed for the following genes, involved in microimbalances in the patients listed at the bottom of the table: *COBL*, *ARHGAP18*, *CALN1*, and *NEDD4*.

COBL gene was found deleted in two unrelated male patients (figure 4); it encodes for the Cordon-blue protein, a multifunctional brain-enriched vertebrate protein that uses three C-terminal WH2 domains to associate with actin and regulate actin dynamics [31]. To the best of my knowledge, literature lacks of *COBL* knock out model, and the protein was never associated to a disease.

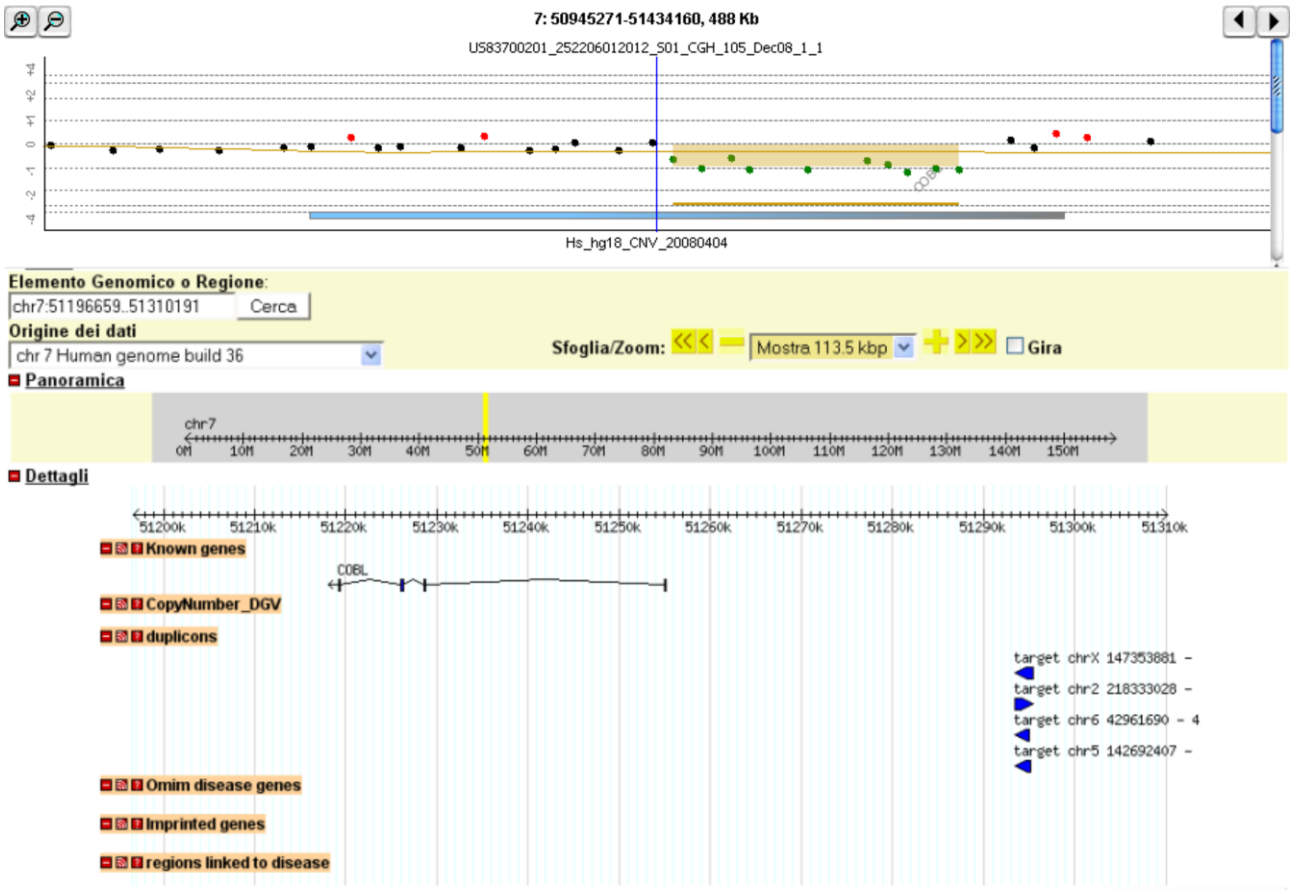

Figure 4. A recurrent deletion in 7p12.1 including the COBL gene.

ARHGAP18 (figure 5), totally duplicated in one patient, belongs to a family of Rho GTPase-activating proteins that modulate cell signalling [32] and was very recently associated to schizophrenia [33].

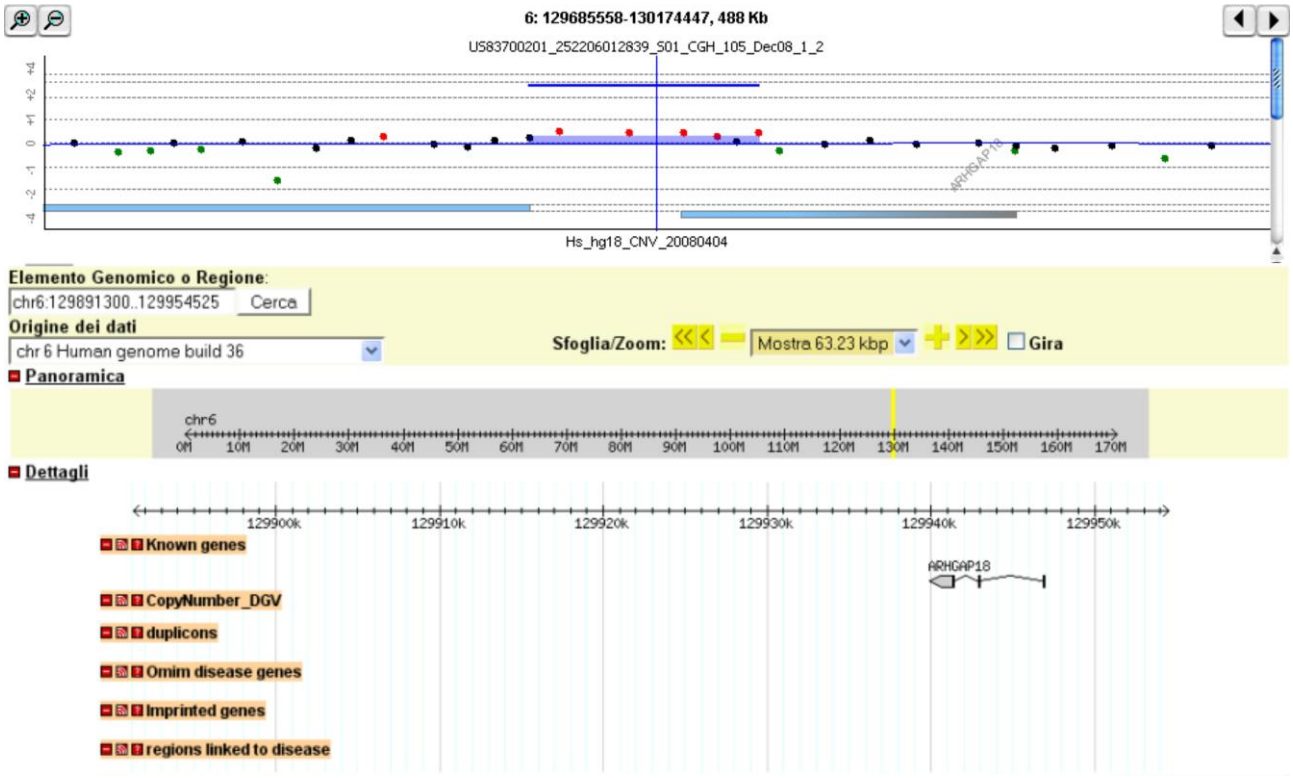

Figure 5. A 63 kb duplication in 6q22.33 including the ARHGAP18 gene.

Calneuron 1 (*CALN1*) gene is a brain specific member of the calmodulin superfamily [34] and never than before was linked to a neurodevelopmental disease. In one of our patients, it is interrupted by a 28 Kb deletion.

Finally, NEDD4 is a very interesting dosage-sensitive regulatory protein of the brain (OMIM 602278) which was interrupted by a microdeletion in one patient.

Future studies

To better delineate the possible role of the previously unreported CNVs in autism, we propose three parallel approaches, to be applied to the candidate variations described in the previous session.

First of all, to investigate if these CNVs represent rare polymorphisms not yet described, rather than autism susceptibility factors, we have planned to screening for these CNVs both a control group of about 1000 Caucasian healthy individuals, and a greater cohort of autistic patients, by a targeted MLPA approach. This will allow us to filter out all CNVs with frequencies not statistically different among patients and controls. Therefore, all CNVs with frequencies significantly different between the two cohorts will be regarded as potentially causative of ASDs.

Second, we propose to screen the array-CGH negative patients for a molecular analysis of the genes involved in the candidate variants: finding mutations in one of these genes could strongly confirm their pathological connotation.

Third, for those imbalanced genes that we know to be expressed in blood, we propose a gene expression analysis, to verify their non-physiological expression and its functional significance.

Acknowledgements

This work was supported by Telethon grant (GGP08226). We'd like to thank all the patients and their parents for participating in this study. The laboratory activities concerning this PhD thesis have been performed at the "Oasi Maria SS" Institute in Troina (Enna, Italy), in which the collaborating units were:

- Laboratorio di Diagnosi Genetica, and especially dr. Daniela Di Benedetto;
- Laboratorio di Genetica Molecolare (dr. Franco Calì, dr. Valeria Chiavetta, dr. Mirella Vinci);
- U.O. di Neurologia e Neurofisiopatologia Clinica e Strumentale (dr. Maurizio Elia);
- U.O. di Pediatria e Genetica Medica (dr. Corrado Romano);
- U.O. di Neurologia per il Ritardo Mentale (dr. Sebastiano Musumeci, dr. Stefania Giusto);
- U.O. di Psicologia (dr. Marinella Zingale, dr. Grazia Trubia).

References

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

2. Risch N, Spiker D, Lotspeich L, Nouri N, Hinds D, et al (1999) A genomic screen of autism: evidence for a multilocus etiology. Am J Hum Genet 65: 493-507.

3. Fernell E, Gillberg C (2010) Autism spectrum disorder diagnoses in Stockholm preschoolers. Res Dev Disabil 31: 680–685.

4. Freitag CM (2007) The genetics of autistic disorders and its clinical relevance: a review of the literature. Mol Psychiatry 12: 2–22.

5. Abrahams BS, Geschwind DH (2008) Advances in autism genetics: on the threshold of a new neurobiology. Nat Rev Genet 9: 341–355.

6. Gillberg C, Coleman M (2000) The Biology of the Autistic Syndromes, Oxford University Press.

7. Benayed R, Gharani N, Rossman I, Mancuso V, Lazar G (2005) Support for the homeobox transcription factor gene ENGRAILED 2 as an autism spectrum disorder susceptibility locus. Am J Hum Genet 77(5):851-68.

8. Campbell DB, Sutcliffe JS, Ebert PJ, Militerni R, Bravaccio C (2006) A genetic variant that disrupts MET transcription is associated with autism. Proc Natl Acad Sci U S A 103(45):16834-9.

9. Alarcón M, Abrahams BS, Stone JL, Duvall JA, Perederiy JV (2008) Linkage, association, and gene-expression analyses identify CNTNAP2 as an autism-susceptibility gene. Am J Hum Genet 82(1):150-9.

10. Cook EH Jr, Scherer SW (2008) Copy-number variations associated with neuropsychiatric conditions. Nature 455: 919–923.

11. Wang K, Zhang H, Ma D, Bucan M, Glessner JT, et al (2009) Common genetic variants on 5p14.1 associate with autism spectrum disorders. Nature 459: 528–533.

12. Weiss LA, Arking DE, Gene Discovery Project of Johns Hopkins & the Autism Consortium, Daly MJ, Chakravarti A, et al (2009) A genome-wide linkage and association scan reveals novel loci for autism. Nature 461: 802–808.

13. Anney, R. et al. (2010) A genomewide scan for common risk variants nominates phospholipase D and polysialyltransferase proteins for a role in autism (Submitted).

14. Manolio TA, Collins FS, Cox NJ, Goldstein DB, Hindorff LA, et al (2009) Finding the missing heritability of complex diseases. Nature 461(7265): 747-753.

15. Conrad DF, Pinto D, Redon R, Feuk L, Gokcumen O, et al (2010) Origins and functional impact of copy number variation in the human genome. Nature 464(7289): 704-12.

16. Autism Genome Project Consortium, Szatmari P, Paterson AD, Zwaigenbaum L, Roberts W et al (2007) Mapping autism risk loci using genetic linkage and chromosomal rearrangements. Nat Genet 39(3): 319-28.

17. Jacquemont ML, Sanlaville D, Redon R, Raoul O, Cormier-Daire V, et al (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): 843-9.

18. Glessner JT, Wang K, Cai G, Korvatska O, Kim CE, et al (2009) Autism genome-wide copy number variation reveals ubiquitin and neuronal genes. Nature 459(7246): 569-73.

19. Marshall CR, Noor A, Vincent JB, Lionel AC, Feuk L et al (2008) Structural variation of chromosomes in autism spectrum disorder. Am J Hum Genet 82(2): 477-88.

20. Sebat J, Lakshmi B, Malhotra D, Troge J, Lese-Martin C, et al (2007) Strong association of de novo copy number mutations with autism. Science 316(5823): 445-9.

21. Cho SC, Yim SH, Yoo HK, Kim MY, Jung GY, et al (2009) Copy number variations associated with idiopathic autism identified by whole-genome microarray-based comparative genomic hybridization. Psychiatr Genet 19(4): 177-85.

22. Christian SL, Brune CW, Sudi J, Kumar RA, Liu S, et al (2008) Novel submicroscopic chromosomal abnormalities detected in autism spectrum disorder. Biol Psychiatry 63(12): 20. Bucan M, Abrahams BS, Wang K, Glessner JT, Herman EI, et al (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.

23. Bucan M, Abrahams BS, Wang K, Glessner JT, Herman EI, et al (2009) Genome-wide analyses of exonic copy number variants in a family-based study point to novel autism susceptibility genes. PLoS Genet 2009 5(6):e1000536.

24. Pinto D, Pagnamenta AT, Klei L, Anney R, Merico D, et al (2010) Functional impact of global rare copy number variation in autism spectrum disorders. Nature 466(7304): 368-72.

25. Toro R, Konyukh M, Delorme R, Leblond C, Chaste P, at al (2010) Key role for gene dosage and synaptic homeostasis in autism spectrum disorders. Trends Genet 26(8): 363-72.

26. 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: 659-685.

27. Lord C, Risi S, Lambrecht L, Cook EH Jr, Leventhal BL, et al (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: 205-223.

28. Ishikawa-Brush Y, Powell JF, Bolton P, Miller AP, Francis F, et al (1997) Autism and multiple exostoses associated with an X;8 translocation occurring within the GRPR gene and 3' to the SDC2 gene. Hum Mol Genet 6(8): 1241-50.

29. Bakkaloglu B, O'Roak BJ, Louvi A, Gupta AR, Abelson JF, et al (2008) Molecular cytogenetic analysis and resequencing of contactin associated protein-like 2 in autism spectrum disorders. Am J Hum Genet 82(1): 165-73.

30. Zhao X, Leotta A, Kustanovich V, Lajonchere C, Geschwind DH, et al (2007) A unified genetic theory for sporadic and inherited autism. Proc Natl Acad Sci U S A 104(31):12831-6.

31. Renault L, Bugyi B, Carlier MF (2008) Spire and Cordon-bleu: multifunctional regulators of actin dynamics. Trends Cell Biol 18(10): 494-504.

32. Potkin SG, Macciardi F, Guffanti G, Fallon JH, Wang Q, et al (2010) Identifying gene regulatory networks in schizophrenia. Neuroimage 53(3): 839-47.

33. Potkin SG, Turner JA, Fallon JA, Lakatos A, Keator DB, et al (2009) Gene discovery through imaging genetics: identification of two novel genes associated with schizophrenia. Mol Psychiatry 14(4): 416-28.

34. Wu YQ, Lin X, Liu CM, Jamrich M, Shaffer LG (2001) Identification of a human brain-specific gene, calneuron 1, a new member of the calmodulin superfamily. Mol Genet Metab 72(4): 343-50.