University of Siena

Ph.D in Medical Genetics

Interpretation of molecular imbalances detected by array-CGH analysis

Filomena Tiziana Papa

Supervisor: Prof. Alessandra Renieri

Academic Year 2009-2010
Ph.D dissertation board

Prof. Alessandra Renieri
Professor of Medical Genetics, University of Siena, Italy

Prof. Mario De Marchi
Professor of Medical Genetics, University of Torino, Italy

Prof. Francesca Gensini
Professor of Medical Genetics, University of Firenze, Italy
4.2 CNV in Autism and RETT syndrome.................................................................40
  4.2.1 Genetic susceptibility factors in ASD patients........................................41
  4.2.2 Genetic modifier factor ............................................................................56

4.3 Human Copy Number Variations Database ....................................................69

5. DISCUSSION, CONCLUSIONS and FUTURE PERSPECTIVES ....................74
  5.1 DISCUSSION........................................................................................................75
  5.2 CONCLUSIONS AND FUTURE PERSPECTIVES........................................83

6. REFERENCES..........................................................................................................84

Curriculum Vitae......................................................................................................90
...with love to my family
Acknowledgements

I would like to thank Prof. Alessandra Renieri for believing in me and giving me the opportunity to do my abroad experience. Her teaching and trust helped me believe in my possibilities and follow my goals.

Thanks to …

… Chiara who introduced me to genetic and cared me at the beginning of my no-student-world-work.

… Rossella, pioneer of array technique and Eleni, Mafalda and Viria, we enjoyed working together!

… all my colleagues and friends of Medical Genetics staff, in particular Mirella, Francesca A., Francesca M. for their patience and Marzia, Annabella, Laura and Cinzia my favorite medical doctors.

… Rosangela who after a “stormy” start is became my inseparable family.

… Katia who always cheered me up.

… Giulio who always supported me.

… my new Lexington friends who in these last months helped me not feeling lonely.

A special thank to my “roommate” Maria Antonietta, I miss our conversation in the night!

The experience of life changes your own personality and sometime carries you far from your hearth and home, but you know you can always find friends who wait for you, listen to you and enjoy with you in every part of world we are. Grazia, Francesca, Nicoletta, Monica … thanks, I am so glad to have friends like you!
1. INTRODUCTION
1. INTRODUCTION

1.1 Historical perspective

Since karyotyping became a routine technique in clinical genetics, mental impairment, with or without other abnormalities, has often been found to be associated with chromosome rearrangements. However, in the majority of the patients, the G banded karyotype is normal, and in about half of them no obvious cause for the impairment is found. More recently, cryptic chromosome rearrangements have been reported in patients with an apparently normal karyotype and an unexplained abnormal phenotype. The best characterized rearrangements are the recurrent microdeletion syndromes, such as the Miller-Dieker lissencephaly and DiGeorge syndrome. In addition, subtelomeric imbalances of variable sizes cause intellectual disability in 5-7% of these cytogenetically "normal" cases. In the last few years, microarray-based comparative genomic hybridization (aCGH) analysis has became available, and appears to be a robust tool for detecting genomic imbalances in patients, with a much higher resolution than permitted by cytogenetic analyses based on chromosome banding. [1]

Conventional G-banded karyotyping remains the gold standard, but it is time-consuming (about 10-14 days is required to obtain the result), labor-intensive and the limitations to detect imbalances larger than 5-10 Mb. To overcome these limitations, rapid fluorescent in situ hybridization (FISH) and multiplex ligation-dependent probe amplification (MLPA) and micro-array platforms have been developed and introduced to this field to detect such cryptic genomic aberrations [2] (Fig. 1).

In the 1990s the introduction of molecular cytogenetic techniques into the clinical laboratory setting represented a major advance in the ability to detect known syndromes and identify chromosomal rearrangements of unknown
origin. FISH, which is the annealing of fluorescently labelled locus-specific probes to their complimentary sequences in the genome, allowed the detection of specific microdeletion syndromes. [3] There are currently a number of commercially available FISH probes for the most common disorders and this method is still predominantly used when the clinical phenotype is suggestive of a particular disorder. Several other FISH-based methods, including spectral karyotyping (SKY), multicolour FISH (m-FISH), and comparative genomic hybridization (CGH) have proven extremely useful in the identification of unknown chromosomal material. SKY and m-FISH rely mainly on the principal of differentially labelling each chromosome using a unique combination of fluorochromes and are especially beneficial for identifying the origin and content of supernumerary marker chromosomes (SMCs) and complex chromosome rearrangements (CCRs) that involve more than two chromosomes. CGH was originally introduced for the cytogenetic analysis of solid tumors, which can be difficult to culture and involves the differential labeling of DNA from a test sample and a reference sample. The fluorescently labelled reactions are combined and hybridized to metaphase spreads from chromosomally normal individuals. Gains and losses of the genome in the test sample relative to the control sample are represented as ratios that are quantified from digital image analysis. This method allows the investigation of the whole genome and is very useful for determining the origin of unknown genetic material, such as SMCs and other unbalanced rearrangements. However, CGH does not detect balanced rearrangements, the resolution is on the order of 5-10 Mb and consequently many genomic disorders cannot be detected. [4]
Fig. 1  A. G banded karyotype. B. Fluorescence in situ hybridization (FISH) of metaphase human chromosomes. C. aCGH result.
The need to screen the whole genome at a resolution that surpassed the existing technologies led to the implementation of microarray based CGH. The principle is very similar to that employed for traditional CGH, where two differentially labelled specimens are co-hybridized in the presence of Cot1 DNA. However, instead of metaphase spreads, the hybridization targets are DNA substrates immobilized on a glass slide. [5] [6] [7] Subsequently, the arrays are scanned and the resultant data are analyzed by software that computes the log 2 ratios for a variety of copy number differences between a patient and reference sample (Fig. 2).

![Fig. 2](image)

Schematic representation of an array-CGH experiment. a) Test and reference DNA are differentially labelled, co-precipitated and hybridised to an array. b) and c) After wash procedures, the slides are analysed through a scanner and fluorescence intensities of each probe are determined. d) After imaging processing and data normalization, the log2 ratios of the probes are plotted as a function of chromosomal position. Probes with a value of zero represent equal fluorescence intensity ratio between sample and reference. Each dot represents a single probe spotted on the array. In this representation, copy number loss shift the ratio to the left and copy number gains shift the ratio to the right.
Consequently, aCGH is an entirely molecular technique with a cytogenetic application and represents a hybrid method that requires the expertise of both specialties. The current limitations of the technology include the inability to detect balanced chromosome rearrangements and the equivocal nature of copy number alterations of unknown significance that may be identified. Nevertheless, it is being used routinely in the clinical setting with a normal chromosome result in cases of intellectual disability and/or multiple congenital anomalies (ID/MCA); as a result the diagnostic yield in this patient group has increased considerably. [4]

1.2 Array – CGH Methodologies

Two major types of array targets are currently being utilized. Initially, bacterial artificial chromosomes (BACs) were the array target of choice. [6] However, more recently, oligonucleotide arrays have been adopted due to the increased genome coverage they afford. The design of both array types was made possible by the availability of the complete map and sequence of the human genome. The BAC arrays may contain DNA isolated from large insert clones that range in size from 150–200 Kb, spotted directly onto the array or may employ the spotting of PCR products amplified from the BAC clones. [8] These arrays are generally very sensitive and results can be directly validated with FISH using the BAC DNA as a probe. However, production of BAC DNA is labor-intensive and the resolution is limited to 50-100 Kb, even on a whole genome tiling path array. [9] Oligonucleotide arrays offer a flexible format with the potential for very high resolution and customization. Several different platforms are available for oligonucleotide arrays, some of which were adapted from genome wide SNP-based oligonucleotide markers and others that were created from a library of virtual probes that span the genome, and consequently can be constructed to have extremely high resolution. [10] Both BAC and
Oligonucleotide arrays have been used successfully to detect copy number changes in patients with ID/MCA and autism. A number of different array design approaches have been taken for diagnostic purposes. A targeted array is one that contains specific regions of the genome, such as the subtelomeres and those responsible for known microdeletion/microduplication syndromes, but does not have probes that span the whole genome. [11] [12] [13] This type of array was initially used for clinical applications in postnatal specimens but has also been implemented for prenatal specimens with an abnormal ultrasound finding or for general screening purposes. [14] [15] [16] A whole genome or tiling path array offers full genome coverage with a resolution that is dependent on the spacing of the probes. For clinical testing the resolution generally involves a spacing of 50 Kb to 1 Mb between adjacent probes on the array often with additional coverage at the subtelomeric regions. [17] [18] The enhanced coverage of whole genome arrays identifies an additional 5% of abnormalities when compared to a targeted array. [19] [20] For research purposes, very high density oligonucleotide whole genome arrays and region specific custom arrays have been instrumental in defining new syndromes, detecting target gene deletions and characterizing breakpoint regions. [21] [22] [23]
1.3 Applications of array-CGH in clinical genetics

aCGH is a highly effective technique that is entering routine clinical use much faster than other microarray technologies. Indeed, compared to, for example, expression microarrays, aCGH enjoys several technical advantages: (1) genomic DNA samples are less prone to degradation than mRNA samples, (2) genomic DNA samples show much less variation between biological replicates than mRNA samples, and (3) interpretation of chromosomal imbalances is much easier than that of expression fingerprints. These advantages explain why aCGH for the diagnosis of constitutional anomalies is progressing faster towards the clinic than expression microarrays for the prediction of clinical outcome (e.g., in cancer), [24] for which a few applications are now entering clinical practice.

aCGH have revolutionized cytogenetic diagnostics and, in turn, the clinical management of patients with developmental delays and multiple congenital anomalies [25] [26] [27] aCGH has given the clinician a greater appreciation of variability in the clinical presentation of many well-described conditions, [28] [29] has identified numerous previously unrecognized chromosomal syndromes with relatively mild phenotypes [30] [31] [32] [33] [34] [35] and has refined critical regions for established genetic defects. [36] aCGH allowed to appreciate the prevalence of mosaicism in individuals with developmental delay and ascertain the unexpected frequency of copy number variants (CNVs) across the genome. [37] Furthermore, the application of aCGH has created a paradigm shift in genetics that has moved the description and discovery of genetic conditions from the "phenotype-first" approach, in which patients exhibiting similar clinical features are identified prior to the discovery of an underlying etiology, to a "genotype-first" approach, in which a collection of individuals with similar copy-number imbalances can be examined for common clinical features. [38] Moreover we performed aCGH to detect recurrent genomic alterations in tumor tissue, underlying genetic susceptibility factors in a consistent number of
autism spectrum disorders (ASD) patients, to identify new genetic determinants in known syndrome.

1.3.1 Application in known and unknown chromosomal syndrome

"Recognizable syndromes" are recognizable because they exhibit, to the trained clinician, a constellation of signs and symptoms that arouse sufficient suspicion to cause the clinician to order a test that will confirm the clinical diagnosis.

The advent of molecular tools such as aCGH, allows us to define the rearrangements in a more detailed and comprehensive manner. A report has highlighted the usefulness of aCGH to characterize the Angelman syndrome/Prader-Willi syndrome (AS/PWS) region. [15] This age-old pattern of medical practice creates a loop that includes the patient, the clinician, and the laboratory and, in doing so, reinforces these recognizable features, cements them to the syndrome, and makes the clinician more confident in his/her diagnostic skills. With the application of aCGH to individuals with nonspecific developmental delay (DD) and/or intellectual disability (ID), with or without dysmorphic features (DF), it is now clear that many recognizable microdeletions and microduplication syndromes have a much wider spectrum of clinical presentation than was previously appreciated. [39] [25] A more complete understanding of the full clinical spectrum of these disorders will be achieved as the use of aCGH in the clinic becomes more prevalent and as correlations of these clinical findings with the genomic lesions are made. Existing website resources such as DECIPHER (see 1.5). [40] Many recognizable microdeletion syndromes are caused by nonallelic homologous recombination (NAHR) mediated by flanking low-copy repeat (LCR) sequences. [41] Interchromosomal and interchromatid NAHR between LCRs in direct orientation result in reciprocal duplication and deletion, whereas intrachromatid NAHR only creates deletion.
This mechanism predicts that the prevalence of the reciprocal duplication product in the population should be 1:2 to the frequency of deletion. However, duplications have not been observed until fairly recently, likely because, in general, individuals with duplications tend to have a milder phenotype than those with the complementary deletions [43] [44] [45] [46] [47] and this milder phenotype may not lead to clinical investigation. [48] [49] The introduction of aCGH in clinical practice has virtually eliminated all the technical impediments of traditional cytogenetics and FISH and allowed the detection of such conditions with relative-but not complete-independence from the clinician's diagnostic judgment. Therefore, recent reviews of cohorts of patients ascertained with aCGH showed that the frequency of these duplications is much higher than heretofore appreciated. As aCGH becomes the primary method of testing individuals with even mild DD/ID, the frequency of microduplications at the common microdeletion syndrome loci will likely increase. [37] [50]

Molecular karyotyping is being increasingly applied to delineate novel disease causing microaberrations and related syndromes in patients with ID of unknown aetiology. The use of aCGH has allowed the discovery of many new syndromes in the last few years and has revitalized clinical cytogenetics. [51] [52] [53] [54]

1.3.2 Application to identify new genetic defects in known syndrome

Until recently, the cause of CHARGE syndrome, a sporadic malformation, was unknown. Vissers et coll. localized the gene for CHARGE syndrome by identifying and characterizing microdeletions by aCGH. CHARGE syndrome is a pleiotropic disorder comprising of coloboma, heart defects, choanal atresia, retarded growth and development, genital hypoplasia, ear anomalies and deafness. They tested 18 patients with CHARGE syndrome on a 1 Mb resolution genome-wide BAC array. One de novo microdeletion of 4.8 Mb was identified on
8q12. Another CHARGE patient originally reported with a balanced chromosome 8 translocation revealed a complex microdeletion partially overlapping with the one encountered in our index patient. No microdeletions were identified in 17 additional CHARGE patients tested on a tiling resolution chromosome 8 BAC array. Sequence analysis of nine genes located within the minimal region of deletion overlap revealed causative mutations in CHD7, a novel member of the chromodomain helicase DNA-binding gene family, in the majority of CHARGE patients without microdeletions. From these results, they concluded that CHARGE syndrome is caused by haploinsufficiency of the CHD7 gene, either by a microdeletion encompassing the CHD7 gene or by single base changes within this gene. This study showed that aCGH can indeed serve as an effective new approach to localize disease-causing genes. [55]

1.3.3 Application to detect the prevalence of mosaicism

Even though the effect of mosaicism on embryonic development and pregnancy outcome is not entirely clear, mosaic chromosomal imbalances have been shown to affect the development of in vitro-generated preimplantation embryos. [56] However, the detection of mosaicism in only 5% of aneuploid spontaneous miscarriages between 6-20 weeks gestation [57] and in only 12% of viable pregnancies screened by chorionic villus sampling (CVS) [58][59] indicates that the incidence of mosaicism decreases through the first and second trimesters of pregnancy and is even rarer in live births. This dramatic reduction in mosaicism from the early stages of embryonic development through the late stages of clinically established pregnancies suggests that there is significant selection against mosaicism. Nevertheless, detecting low-level mosaicism for clinically significant chromosome abnormalities remained a pressing diagnostic challenge for conventional cytogenetic testing until the advent of aCGH. The first systematic study of mosaicism in a large cohort identified mosaicism in as
little as 3% of the cells on the basis of metaphase counts. [60] aCGH has revealed a surprising prevalence of mosaicism in unselected populations of individuals who are referred for evaluation owing to DD, ID, and DF; the frequency of mosaicism among such individuals may be as high as 8-10%. [25] [60] Perhaps most significant is that aCGH does not require the stimulation of cell cultures by phytohemagglutinin, as classically done in clinical cytogenetics, which may distort the percentage of mosaic cells and inhibit the detection of some mosaic abnormalities by chromosome analysis. [61]

1.3.4 Application to ascertain the frequency of copy number variants across the genome.

Improvements in microarray resolution and hybridization robustness have resulted in the widespread implementation of genomic microarray technologies in medical research and diagnostics. This technology is most effective in detecting genomic deletions and duplications larger than 1 Kb, known as CNVs. Genomic microarrays are commonly used to identify rare, but highly penetrant, and commonly single CNVs in patients suffering from neurological disorders such as autism, [62] [63] schizophrenia [64] [65] and intellectual disability. [50] [66] [67] However CNVs have also been recently recognized as a common form of genomic structural variation: high resolution microarrays and sequencing approaches are able to identify 600-900 CNVs in a single individual. [68] [69] [70] [71] [72] Current clinical interpretation therefore needs to contrast the frequencies of a CNV in affected versus unaffected individuals, as well as determining the inheritance of CNVs via parental analysis. [73] [74] The size distributions of CNVs detected is dependent on the technology used. Increases in microarray resolution are revealing both a much higher rate of rare CNVs than previously thought [75] and an increasing number of genomic loci being reported that show variable inheritance and penetrance. There are limitations in
considering CNVs as either benign when common and inherited, or causal when rare and de novo. At present up to 5% of the human genome has been shown to vary in large scale copy number in numerous healthy controls [71] [76] and novel CNVs continue to be identified. In particular, CNV regions are enriched in repetitive sequences of near identical DNA known as segmental duplications [77] and are less prone to recombination. [78] The CNVs reported to date are documented in the Database of Genomic Variants (DGVs) (http://projects.tcag.ca/variation). During our analysis we found some rearrangements not reported in that database. We built, in collaboration with other Italian laboratory, an Human Copy Number Variations database (See 1.5).

1.4 CNV in Autism and RETT syndrome

The application of high-resolution genome analysis in research and clinical laboratories will uncover the genomic basis of many such disorders and will allow for better correlation of the many known CNVs with specific phenotypes. Although this promises to be a very challenging exercise, [79] [80] much work has already been initiated in cancer, neurological and neuropsychological conditions, infectious diseases, and others, suggesting that the clinical utility and applicability of such investigations cannot be too distant. Recent publications document the contribution of CNV to genetic diversity in humans and human disease. CNV has been shown to contribute to phenotype in model organisms and to important production and disease traits in domesticated livestock species. [81] A significant fraction of CNVs are likely to have functional consequences, due to gene dosage alteration, disruption of genes, positional effects or the uncovering of deleterious alleles. [82]

Autistic disorder is the prototypic form of a group of conditions, the autism spectrum disorders, which share common characteristics including
impairments in reciprocal social interaction and communication, and the presence of restricted and repetitive interests and behaviors. Individuals with an ASD vary greatly in language ability and cognitive development. For example, cognitive development ranges from above average to intellectual disability. [83]

Autism was first described by Leo Kanner in 1943 as a childhood developmental disorder characterized by the "inability to relate themselves in the ordinary way to other people" and by "insistence on sameness". The diagnostic criteria include qualitative impairment of reciprocal social interaction and communicative development and restricted interests and repetitive behaviors. The prevalence of autism is 1:500 with a 4:1 male to female ratio. In addition to autism, ASDs include Asperger syndrome, pervasive developmental disorder not otherwise specified, and a few specific conditions such as Rett syndrome. [84]

Twin and family studies indicate a predominantly genetic basis for ASD susceptibility and provide support for considering these disorders as a clinical spectrum. Concordance of monozygotic twins for autistic disorder is ~ 60%, but rises to ~ 90% when less severe cognitive and social deficits are considered. [85] Additionally, subclinical autistic traits are sometimes observed in first-degree relatives [86] and, to a lesser extent, in the general population. [87] The architecture of ASD genetic risk is complex. [88] Some 5-15% of individuals with an ASD have an identifiable genetic etiology corresponding to known rare single-gene disorders (e.g., fragile X syndrome, tuberous sclerosis 1 (TSC1), tuberous sclerosis 2 (TSC2), methyl CpG binding protein 2 (MECP2) in Rett syndrome and phosphatase and tensin homolog (PTEN)) and chromosomal rearrangements (e.g., maternal duplication of 15q11q13). [89][90]

Rett syndrome RTT is an X-linked dominant neurodegenerative disorder that predominantly affects women. Individuals with RTT show progressive
deficits beginning at 6 to 18 months that include atrophy of verbal and skilled motor abilities, social withdrawal, hand wringing, respiratory difficulties, seizures, and autism spectrum behavior. RTT symptoms are variable and depend on the pattern of X-chromosome inactivation of the mutant allele, 121, 125-127 the nature of the MECP2 mutation, and epigenetic factors. [91] The syndrome has been the object of extensive investigations, revealing a wide spectrum of clinical phenotypes including the classic form, the early onset seizure variant, the Zappella variant (Z-RTT), the congenital variant, the 'forme fruste' variant, and the late regression variant. [92] The majority of cases of RTT are the result of de novo mutations in MECP2 gene that can arise in either parent, but more frequently occur in the paternal gamete. Mutations in the this gene account for half of girls with classic RTT. This mutation is prenatally lethal in males. A few surviving RTT male cases have been reported, but these were generally mosaic for MECP2 mutations. [93] MeCP2 inhibits transcription by binding DNA methylated at CpG dinucleotides and translation via direct interaction with RNA. Mutations in another gene, CDKL5 which encodes cyclin-dependent kinase-like 5 and FOXG1 gene, transcription factors which may play a role in the development of the brain and telencephalon, have also been implicated in a Rett like syndrome phenotype.

1.5 DECIPHER and Human Copy Number Variations database

As we already mentioned before, many patients suffering from developmental disorders harbor submicroscopic deletions or duplications that, by affecting the copy number of dosage-sensitive genes or disrupting normal gene expression, lead to disease. However, many aberrations are novel or extremely rare, making clinical interpretation problematic and genotype-phenotype correlations uncertain. Identification of patients sharing a genomic rearrangement and having phenotypic features in common leads to greater
certainty in the pathogenic nature of the rearrangement and enables new syndromes to be defined. To facilitate the analysis of these rare events, an interactive web-based database has been developed called DECIPHER (Database of Chromosomal Imbalance and Phenotype in Humans using Ensembl Resources; http://www.sanger.ac.uk PostGenomics/decipher1) (Wellcome Trust Sanger Inst. 2007). It may facilitate widespread appreciation of such phenotypic variability, which incorporates a suite of tools designed to aid the interpretation of submicroscopic chromosomal imbalance, inversions, and translocations. DECIPHER catalogs common copy-number changes in normal populations and thus, by exclusion, enables changes that are novel and potentially pathogenic to be identified. DECIPHER enhances genetic counseling by retrieving relevant information from a variety of bioinformatics resources. Known and predicted genes within an aberration are listed in the DECIPHER patient report, and genes of recognized clinical importance are highlighted and prioritized. DECIPHER enables clinical scientists worldwide to maintain records of phenotype and chromosome rearrangement for their patients and, with informed consent, share this information with the wider clinical research community through display in the genome browser Ensembl. By sharing cases worldwide, clusters of rare cases having phenotype and structural rearrangement in common can be identified, leading to the delineation of new syndromes and furthering understanding of gene function. [40]

The possible differences in frequency of particular CNVs in different populations and the variety of different CNVs observed in a given study may be significantly limited by the number and ethnic origin of individuals examined. From a collaboration with some Italian laboratories, an interactive web-based database, who collect rearrangements, de novo and inherited, not described in DGVs, has been developed called Human Copy Number Variations Database (HCNVs) (http://dbcnv.oasi.en.it/gvarianti/index.php).
2. RATIONALE, AIM and OUTLINE
2. RATIONALE, AIM and OUTLINE OF THE STUDY

The objectives of this study were: (a) to discover emerging syndromes in a cohort of patients with ID and MCA, (b) to detect the underlying genetic susceptibility factors in a consistent number of ASD patients, and (c) to explore the hypothesis that CNVs leading to over/underexpression of genes may modulate the phenotype by comparing Rett patient with different mutation. It was hypothesized that accurate characterization of the underlying molecular defect could help predict clinical outcomes, and therefore aid with early, appropriate interventions. Moreover all inherited or de novo rearrangements, not reported in DGVs, have been collected in HCNVs database.

By the use of nucleotide array with 44,000 and 99,000 probes and an average resolution of about 100-130 Kb (44K, Agilent) and 50-65 Kb (105K, Agilent), respectively, we have analyzed 696 patients with mild to severe intellectual disability associated with facial dysmorphisms and/or congenital anomalies. We found 426 (61%) negative while in 88 patients (13%) the analysis is still ongoing. We identified an inherited or de novo rearrangement in 165 cases (24%) while in 17 cases (2%) we detected novel de novo deletions not reported in the literature.

We selected 95 unrelated patients with ASD. In addition to the well known recurrent rearrangements involving the 15q11q13, 16p13 and 22q13 regions, recurrent microdeletions and microduplications at 16p11.2 have been recently identified and are shown to confer susceptibility to ASDs in up to 1% of autistic patients. In fifteen patients (16%) CNVs were considered disease associated, while in 20 patients (21%) CNVs were classified as uncertain cofactor.
We compare two discordant pairs of Rett sisters and four additional discordant pairs of unrelated Rett girls matched by mutation type to identify one major common modifier gene/region.

All patients described in this thesis have been inserted in the DECIPHER database (http://www.sanger.ac.uk/PostGenomic/decipher/) and HCNVs database (http://dbcnv.oasi.en.it/gvarianti/index.php) permitting us to create collaborations and extend the clinical phenotype associated with new microdeletion/microduplication syndromes.
3. MATERIALS and METHODS
3. MATERIALS & METHODS

3.1 Patients collection

Patients with ID and MCA enrolled in this study have been selected among those attending the Medical Genetics Unit of the University Hospital of Siena. All they were evaluated in genetic counseling and a clinically recognizable condition was excluded a diagnosis of a recognizable syndrome, and all patients.

3.2 Array-based CGH

3.2.1 Samples preparation

Genomic DNA of normal controls was obtained from Promega. Genomic DNA was extracted from peripheral blood using a QIAamp DNA Blood Maxi kit according to the manufacturer protocol (Qiagen, www.qiagen.com). The OD260/280 method on a photometer was employed to determine the appropriate DNA concentration. [94] Patient and control DNA samples were sonicated to produce a homogeneous smear DNA extending from approximately 600 bp to 2 Kb. DNA samples were then purified using the DNA Clean and Concentrator kit (Zymo Research, Orange, CA). Ten micrograms of genomic DNA both from the patient and from the control were sonicated. Test and reference DNA samples were subsequently purified using dedicated columns (DNA Clean and Concentrator, Zymo research, CA92867-4619, USA) and the appropriate DNA concentrations were determined by a DyNA Quant™ 200 Fluorometer (GE Healthcare).

3.2.2 Human oligonucleotides array

Array based CGH analysis was performed using commercially available oligonucleotide microarrays containing about 43,000 60-mer probes with an
estimated average resolution of about 100-130 Kb (Human Genome CGH Microarray 44B Kit, Agilent Technologies) and microarrays containing 99,000 60-mer probes with an estimate average resolution of 50-65 Kb (Human Genome CGH Microarray 105A Kit, Agilent Technologies). Physical positions of the probes correspond to the UCSC genome browser - NCBI build 36, March 2006. (http://genome.ucsc.edu). DNA labelling was executed essentially according to the Agilent protocol (Oligonucleotide Array-Based CGH for Genomic DNA Analysis 2.0v) using the Bioprime DNA labelling system (Invitrogen). Genomic DNA (2 μg) was mixed with 20 μl of 2.5X Random primer solution (Invitrogen) and MilliQ water to a total volume of 41 μl. The mix was denaturated at 95°C for 7 minutes and then incubated in ice/water for 5 minutes. Each sample was added with 5 μl of 10X dUTP nucleotide mix (1.2 mM dATP, dGTP, dCTP, 0.6 mM dTTP in 10 mM Tris pH 8 and 1 mM EDTA), 2.5 μl of Cy5-dUTP (test sample) or 2.5 μl of Cy3-dUTP (reference sample) and with 1.5 μl of Exo-Klenow (40 U/μl, Invitrogen). Labeled samples were subsequently purified using CyScribe GFX Purification kit (Amersham Biosciences) according to the manufacturer protocol. Test and reference DNA were pooled and mixed with 50 μg of Human Cot I DNA (Invitrogen), 50 μl of Blocking buffer (Agilent Technologies) and 250 μl of Hybridization buffer (Agilent Technologies). Before hybridization to the array the mix was denatured at 95°C for 7 minutes and then pre-associated at 37°C for 30 minutes. Probes were applied to the slide using an Agilent microarray hybridization station. Hybridization was carried out for 24/40 hrs at 65°C in a rotating oven (20 rpm). The array was disassembled and washed according to the manufacturer protocol with wash buffers supplied with the Agilent kit. The slides were dried and scanned using an Agilent G2565BA DNA microarray scanner. Image analysis was performed using the CGH Analytics software v. 3.4.40 with default settings. The software automatically determines the fluorescence intensities of the spots for both fluorochromes performing
background subtraction and data normalization, and compiles the data into a spreadsheet that links the fluorescent signal of every oligo on the array to the oligo name, its position on the array and its position in the genome. The linear order of the oligos is reconstituted in the ratio plots consistent with an ideogram. The ratio plot is arbitrarily assigned such that gains and losses in DNA copy number at a particular locus are observed as a deviation of the ratio plot from a modal value of 1.0.

3.3 Real-time quantitative PCR

Some aCGH data were confirmed by Real-time Quantitative PCR experiments. To design adequate probes in different regions of the human genome, we used an TaqMan Gene Expression Assays by design which provides pre-designed primers-probe set for real-time PCR experiments (Applied Biosystems, https://products.appliedbiosystems.com) PCR was carried out using an ABI prism 7000 (Applied Biosystems) in a 96-well optical plate with a final reaction volume of 50 μl. A total of 100 ng (10 μl) was dispensed in each of the four sample wells for quadruplicate reactions. Thermal cycling conditions included a prerun of 2 min at 50°C and 10 min at 95°C. Cycle conditions were 40 cycles at 95°C for 15 sec and 60°C for 1 min according to the TaqMan Universal PCR Protocol (ABI). The TaqMan Universal PCR Master Mix and Microamp reaction tubes were supplied by Applied Biosystems. The starting copy number of the unknown samples was determined using the comparative Ct method as previously described. [95]
3.4 Multiplex Ligation-dependent Probe Amplification (MLPA)

MLPA analysis was performed according to the provider's protocol with a specifically designed set of probes for testing critical regions in DiGeorge syndrome (SALSA P023 kit; MRC-Holland, Amsterdam, Netherlands: http://www.mrc-holland.com), 1p-deletion syndrome, Williams syndrome, Smith-Magenis syndrome, Miller-Dieker syndrome, DiGeorge syndrome, Prader-Willi syndrome, Alagille syndrome, Saethre-Chotzen syndrome, Sotos syndrome: (SALSA P064B MR1 kit) and subtelomere regions (SALSA P036D subtelomeric primer kit). The ligation products were amplified by PCR using the common primer set with the 6-FAM label distributed by the supplier. Briefly, 100 ng of genomic DNA was diluted with TE buffer to 5 μl, denatured at 98°C for 5 minutes and hybridized with SALSA Probe-mix at 60°C overnight. Ligase-65 mix was then added and ligation was performed at 54°C for 15 minutes. The ligase was successively inactivated by heat, 98°C for 5 minutes. PCR reaction was performed in a 50 μl volume. Primers, dNTP and polymerase were added and amplification was carried out for 35 cycles (30 seconds at 95°C, 30 seconds at 60°C and 60 seconds at 72°C). Amplification products were identified and quantified by capillary electrophoresis on an ABI 310 genetic analyzer, using GENESCAN software (version 3.7) all from Applied Biosystems (Foster City, CA, USA). The peak areas of the PCR products were determined by GENOTYPER software (Applied Biosystems). A spreadsheet was developed in MicrosoftTM Excel in order to process the sample data efficiently. Data were normalized by dividing each probe's peak area by the average peak area of the sample. This normalized peak pattern was divided by the average normalized peak pattern of all the samples in the same experiment. [96]
4. RESULTS
4.1 Novel deletions/duplications
A 3 Mb Deletion in 14q12 Causes Severe Mental Retardation, Mild Facial Dysmorphisms and Rett-like Features.

Filomena Tiziana Papa, Maria Antonietta Mencarelli, Rossella Caselli, Eleni Katzaki, Katia Sampieri, Ilaria Meloni, Francesca Ariani, Ilaria Longo, Angela Maggio, Paolo Balestri, Salvatore Grosso, Maria Angela Farnetani, Rosario Berardi, Francesca Mari, and Alessandra Renieri.

Clinical Report

A 3 Mb Deletion in 14q12 Causes Severe Mental Retardation, Mild Facial Dysmorphisms and Rett-like Features

Filomena Tiziana Papa,1 Maria Antonietta Mencarelli,1 Rossella Caselli,1 Eleni Katzaki,1 Katia Sampieri,1 Ilaria Meloni,1 Francesca Ariani,1 Ilaria Longo,1 Angela Maggio,2 Paolo Balestri,2 Salvatore Grosso,2 Maria Angela Farnetani,2 Rosario Berardi,2 Francesca Mali,1 and Alessandra Renieri1,2

1Medical Genetics Molecular Biology Department, University of Siena, Siena, Italy
2Department of Pediatrics, University of Siena, Italy

Received 31 October 2007; Accepted 7 April 2008

The present report describes a 7-year-old girl with a de novo 3 Mb interstitial deletion of chromosome 14q12, identified by oligo array CGH. The region is gene poor and contains only five genes of two of them, FOXP1B and PRKCD being deleted also in a previously reported case with a very similar phenotype. Both patients present prominent metopic suture, epicanthic folds, bulbous nasal tip, tense upper lip, eversion lower lip and large ears and a clinical course like Rett syndrome, including normal perinatal period, postnatal microcephaly, seizures, and severe mental retardation. FOXP1B (forkhead box G1B) is a very intriguing candidate gene since it is known to promote neuronal progenitor proliferation and to suppress premature neurogenesis and its disruption is reported in a patient with postnatal microcephaly, corpus callosum agenesis, seizures, and severe mental retardation. © 2008 Wiley-Liss, Inc.

Key words: chromosome 14; array-CGH; 14q deletion; Rett-like features; FOXP1B gene


INTRODUCTION

Interstitial deletions of the long arm of chromosome 14 are a relatively rare finding. Reported deletions have ranged from the loss of multiple bands to smaller deletions involving a single band (Kaminszaran et al., 2001; Petruk et al., 2003). The clinical phenotype varies, but some features commonly seen include psychomotor retardation, hypotonia, microcephaly and craniofacial anomalies.

Bisgaard et al. (2006) have recently described a female patient with an interstitial deletion of 14q12 of 2.65–3.5 Mb, identified by metaphase CGH. The girl presented psychomotor retardation, epilepsy, microcephaly and unusual facial features (Bisgaard et al., 2006).

Here we report on the clinical description and the molecular characterization of a patient with a very similar 3 Mb deletion of the long arm of chromosome 14 involving band q12, identified by oligo array-CGH. The clinical phenotype is represented by severe mental retardation, epilepsy, microcephaly, Rett-like features and mild facial dysmorphisms.

To our knowledge this is the second case reported of a patient with a deletion smaller than 3.5 Mb involving the band 14q12. The small number of genes included in the deleted region (only five) may allow to identify candidate genes responsible for the clinical features.

CLINICAL REPORT

The patient is a 7-year-old girl, only child of healthy, nonconsanguineous parents. At the time of delivery, the mother and the father were 30 years old. No teratogen exposure during pregnancy was
reported. In the third pregnancy trimester, ultrasound
ectography showed intrauterine growth retardation.
The girl was born at 39 weeks of gestation by cesarean. Her birth weight was 2730 g (25th cent),
length was 48.5 cm (25th cent), OFC was 33 cm (10–25th cent), and the Apgar score was 8 at 1' and 9 at 5'.
Psychomotor development was severely impaired since early infancy. Head control was acquired at the
age of 4 months. No gain of motor milestones or
social contact was achieved until now. An almost
normal development during the first 6 months of life
followed by a period of regression is referred by the
mother. At that time, the child experienced an
epileptic episode characterized by staring gaze
followed by clonic jerking. At the age of 3 years
and 10 months, seizures with similar electroclinical
pattern re-occurred. EEG recording showed paroxysmal
activity in the tempo-centro-occipital area,
more prominent in the tempo-centro-central
region bilaterally. Results of a brain MRI at age 1 year
and a recent new evaluation were normal. An ophthalmologic examination showed hypermetropia.
Physical examination at 7 years showed: height
102 cm (<<3rd cent), weight 15.4 kg (<<3rd cent), OFC
46.5 cm (<<3rd cent), prominent metopic suture,
bluish epicanthal folds, intermittent exotropia,
fullness of orbital region, depressed nasal bridge,
bulbous nasal tip, large ears, thick lips with everted
lower lip, prognathism (Fig. 1). She showed
generalized hypotonia, small eyelids, gauche-oesophageal
reflux, bruxism, sclerocoria and cold extremities.
Constant stereotypic movements of hands and
tongue were also evident. Hand movements are
asymmetrical and more often not in the midline.
They are rocking/rhythmic movements, hand
waving, movements of scratching, pinching and patting
with fingers. The patient presents also a double
collarbone crease on the left and a double ankle crease on the
right.
Results of the karyotype (320 bands), analysis of
telomeric rearrangements, sequence analysis of the
CDKL5 gene and analysis of 15q11q13 region were
normal. DHPLC analysis of the MBD2 gene showed a
nonpathogenic variant (c. 1168_1173del) inherited
from the father.

Array-CGH Analysis

Array-based CGH analysis was performed as
previously reported [Resovitz et al., 2007]. We used
commercially available oligonucleotide microarrays
containing about 99,000 50-mer probes (Human
Genome CGH Microarray 105A Kit, Agilent
Technologies, Santa Clara, California). The average
resolution is about 25 kb.

Quantitative PCR Analysis

Real-time quantitative PCR was performed to con
firm array-CGH data. We used a pre-designed set of
primers and probe specific for the FOXG1E gene
(14q12 locus) provided by the Array-by-Design
service (Applied Biosystems, Foster City, CA). Primers
and probe were designed in the gene, FOXG1E, for
ward primer: 5'-GTCAGCCCCACCGCGCT-3',
FOXG1E reverse primer: 5'-GGCAAAACGCAGTCAT
TATTTAGATAACA-3', FOXG1E_TaqMan probe: 5'-
TTGACTTCAACGCTATATCT-3' reverse PCR was
carried out as previously described [Resovitz et al.,
2007]. The starting copy number of the unknown
samples was determined using the comparative Ct
method, as reported by Livak [1997].

Array-CGH analysis identified a chromosome
14 interstitial deletion of about 3.5 Mb (45,XX,del(14)(q12q12)) (Fig. 2). Both breakpoints are mapped in
14q12, with the last present oligonucleotide located
in 27.47 MB and the first deleted in 27.55 MB. The
deletion breakpoint is located between 20.48 MB, last
oligonucleotide deleted, and 30.47 MB, first oligo
nucleotide present (according to UCSC Genome
Browser, http://genome.ucsc.edu, on Human May
2004 Assembly). Only five genes are included in the
region: FOXG1E, FRK17, SCPD1, CC2H and
STRG5.
Array-CGH analysis performed in both healthy parents showed a normal result indicating that the rearrangement is de novo in nature (results not shown). To confirm array data, real-time quantitative PCR experiments were performed in the patient and her parents. The deletion was confirmed in the patient, while parents showed a normal result (Fig. 2b).

Array-CGH analysis was performed on DNA from patient described by Bipaard et al. (2006), for a better definition of deletion extension and breakpoints. The deletion is about 3.12 Mb, with both breakpoints mapped in 14q12. The last present oligonucleotide is located in 26.50 Mb and the first deleted in 26.68 Mb. The distal breakpoint is located between 30.03 Mb, last oligonucleotide deleted, and 30.09 Mb, first oligonucleotide present (according to UCSC Genome Browser, http://genome.ucsc.edu, on Human May 2004 Assembly). Only two genes are included in the region: FOXG1B and FOXC1.

**DISCUSSION**

Given the clinical history and the neurological features of the patient, we first hypothesized a Rett spectrum disorder. The severe neurological impairment, dyspraxia, stereotypes, gastroesophageal reflux, constipation, scoliosis and cold extremities together with a normal perinatal period were quite suggestive of Rett syndrome. However, we were unable to demonstrate a mutation in one of the two main genes of Rett syndrome, *MECP2* and *CDKL5*.

Considering the presence of distinctive facial features, which are not present in Rett syndrome, we decided to perform the array-CGH analysis in the patient, and we found a pathogenic de novo deletion in 14q12. In this region no copy number variation is reported in normal individuals and it contains a very low amount of genes. In a 3 Mb region only five known genes are included. None of them is reported...
in the databases of imprinted genes (http://www.
genomeimprint.com and http://ucg.otago.ac.nz).

Although LCRs are not reported near the breakpoints (BioRTDatabase: http://projects.tcag.ca/bioxrt/), the deletion identified in our patient is very similar to another recently reported by Bisgaard et al. [2006] (Table 1). The described facial phenotype is very similar between the two patients: both present postnatal microcephaly, prominent metopic suture, epicanthic folds, bulbous nasal tip, tented upper lip, evverted lower lip and large ears. Moreover, the two patients both have mental retardation and seizures. Two genes are included in the common deleted region: FOXG1B and PRKDI.

The PRKDI gene encodes for a cyslosine serinethreonine kinase that binds to the trans-Golgi network and regulates the fission of transport carriers specifically destined to the cell surface [Johannes et al., 1994]. To date a clinical phenotype has not been associated with mutations or deletions of this gene. FOXG1B (forkhead box G1B) gene is of a very interesting factor. This transcription factor is potentially a strong candidate gene for determining forebrain size in vertebrates due to its role in the development of the telencephalon, where it promotes progenitor proliferation and suppresses premature neurogenesis [Bredenkamp et al., 2007].

Shioch et al. [2005] reported a patient with an inversion in chromosome 14 disrupting at the breakpoints the FOXG1B gene. The patient exhibited severe cognitive disability associated with postnatal microcephaly, seizures, agenesis of the corpus callosum and frontal and parietal myelination defects. The patient was very severely affected being in the status of tetraplegia at the age of 7 years. Similar features are described also in two patients harboring larger deletions that involve FOXG1B, previously reported by Kaminosaram et al. [2001]. It is plausible that neurological features result at least in part from FOXG1B haploinsufficiency.

On the telomeric end of the deletion, outside the deleted region shared with the patient reported by Bisgaard there are three additional genes: SCP1D, COCH and STRN3, SCIDH (Ssc1 family domain containing 1) may be involved in vesicular transport between the endoplasmic reticulum and the Golgi and may interact functionally with Ras protein. To date a clinical phenotype has not been associated with mutations or deletions of this gene. COCH encodes for cochlin, the major non-collagen component of extracellular matrix of the inner ear. Dominant negative COCH mutations cause dominant progressive hearing loss and vestibular dysfunction (DFNA9) [Makishima et al.,

<table>
<thead>
<tr>
<th>TABLE I: Comparison of the Clinical Features of the Present Case with the one Described by Bisgaard et al. [2006]</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Present case</strong></td>
</tr>
<tr>
<td>Deletion breakpoints</td>
</tr>
<tr>
<td>Age</td>
</tr>
<tr>
<td>Birth weight</td>
</tr>
<tr>
<td>Birth length</td>
</tr>
<tr>
<td>Birth OFC</td>
</tr>
<tr>
<td>Prominent growth retardation</td>
</tr>
<tr>
<td>Psychomotor retardation/mental retardation</td>
</tr>
<tr>
<td>Seizures</td>
</tr>
<tr>
<td>Brain MRI</td>
</tr>
<tr>
<td>Microcephaly</td>
</tr>
<tr>
<td>Prominent metopic suture</td>
</tr>
<tr>
<td>Tarsus</td>
</tr>
<tr>
<td>Epicanthal folds</td>
</tr>
<tr>
<td>Nasal bridge</td>
</tr>
<tr>
<td>Nasal tip</td>
</tr>
<tr>
<td>Lips</td>
</tr>
<tr>
<td>Prognathism</td>
</tr>
<tr>
<td>Hypertonia</td>
</tr>
<tr>
<td>Scoliosis</td>
</tr>
<tr>
<td>Gastroesophageal reflux</td>
</tr>
<tr>
<td>Epilepsy</td>
</tr>
<tr>
<td>Seizures</td>
</tr>
<tr>
<td>Cold extremities</td>
</tr>
<tr>
<td>Postural abnormality</td>
</tr>
<tr>
<td>Stereotypic movements</td>
</tr>
</tbody>
</table>

N.A.: not available.

*Oligonucleotide positions according to UCSC Genome Browser: http://genome.ucsc.edu; on Human May 2004 Assembly, numbers refer to last oligonucleotide present, last oligonucleotide deleted, last oligonucleotide deleted and last oligonucleotide present, respectively.
2005]. Our patient presents a normal auditory function and this is in accordance with the evidence that COCH haploinsufficiency does not cause DFNA9 [Makishima et al., 2005].

The distal breakpoint of the deletion interrupts the STRN3 gene. STRN3 or SG2NA is a supposedly nuclear protein, which shares with striatin identical protein–protein interaction domains and the same overall domain structure [Gastets et al., 2001]. Proteins of the striatin family are principally expressed in neurons with a somato-dendritic localization and high concentration in dendritic spines. Recent investigations have demonstrated the involvement of proteins of the striatin family in numerous developmental and nondevelopmental signaling pathways as well as in vesicular trafficking (Gaillard et al., 2005). This data could suggest a possible involvement also of STRN3 gene in the neurological phenotype of our patient.

ACKNOWLEDGMENTS
This work was supported by Telethon grant GGP06170 and EuroRETT E-RARE Network to A.R. and grant from University of Siena (PAR 2005) to F.M. We gratefully acknowledge Anne Marie Bisgaard, Maria Kirchhoff and Monica Cohen for their collaboration in providing the DNA sample of their patient that has permitted a better definition of the deletion breakpoints.

REFERENCES
4.2 CNV in Autism and RETT syndrome
4.2.1 Genetic susceptibility factors in ASD patients
Practical hints for management of array-CGH output in autism spectrum disorders: susceptibility regions versus uncertain cofactors.

Filomena Tiziana Papa, Roberto Canitano, Maria Antonietta Mencarelli, Mucciolo Mafalda, Disciglio Vittoria, Annabella Marozza, Radice Lucia, Cinzia Castagnini, Laura Dosa, Marzia Pollazzon, Joussef Hayek, Alessandra Renieri, Francesca Mari.

Submitted to Am J Med Genet Part B
Practical hints for management of array-CGH output in autism spectrum disorders: susceptibility regions versus uncertain cofactors.

Filomena T. Papa¹, Roberto Canitano², Maria Antonietta Mencarelli¹, Mafalda Mucciolo¹, Vittoria Disciglio¹, Annabella Marozza¹, Lucia Radice², Cinzia Castagnini¹, Laura Dosa¹, Marzia Pollazzon¹, Joussef Hayek², Alessandra Renieri¹, Francesca Mari¹.

¹ Medical Genetics-Molecular Biology Department-University of Siena-Policlinico S. Maria alle Scotte-53100 Siena-Italy
² Child Neuropsychiatry-University Hospital of Siena-53100 Siena-Italy

BACKGROUND: Comparative genomic hybridization (CGH) has provided insights into the underlying genetic causes of autism spectrum disorders (ASD) by associating CNVs at 16p11.2, 16p13.1 and 15q13.3 with susceptibility to ASD in up to 1% of patients.

METHODS: Using 44K oligo array-CGH we analyzed 95 unrelated subjects with ASD, thoroughly investigated by both child neuropsychiatrists and clinical geneticists. In most cases parents were available for the analysis. All patients except 3 with Asperger syndrome, had mental retardation, 24% of them had congenital anomalies and 13% had epilepsy.

RESULTS: In fifteen patients (16%) CNV(s) were considered disease associated. These include two de novo deletions (del16p11.2; delXq12), five known susceptibility regions and eight new potential candidate regions (dup2q11.2, del20p12.1, del7q31.1, dup2p16.3, dup4p16.1, dup5q21.1, del9p13.1p13.2, del17q25.3). In 20 patients (21%), including three patients with two rearrangements each, one inherited by the mother and one by the father, CNVs were classified “as uncertain cofactors”. Overall, CNVs were more common among patients with congenital anomalies and less common in patients with epilepsy.

CONCLUSIONS: Our study confirmed that array-CGH analysis is able to detect the underlying genetic cause in about 16% of ASD patients, strongly indicating that it has become an essential diagnostic tool for ASD patient assessment. Moreover, this study underlines the difficulty in interpreting array-CGH data in an even larger number of patients (21%), and provides some practical recommendations for the management of such data and for the communication of findings to the family.

Keywords: array-CGH; CNVs; autism spectrum disorders; susceptibility region; uncertain cofactors

INTRODUCTION

Autism Spectrum Disorders (ASDs) present complex and heterogeneous aetiology with a strong evidence of genetic involvement. In addition to Autism, ASDs include Asperger syndrome and pervasive developmental disorder not otherwise specified (PDD-NOS). The overall estimated prevalence of ASD is 1:160, while the prevalence of Autism is considered 1:500 with a 4:1 male to female ratio.[1]
The complexity of the disorder, resulting from the interaction of several genes and environmental factors, makes the identification of contributory genes extremely difficult. Recently, comparative genomic hybridization (CGH) technology has been used in research studies and in clinical assessment of ASDs in order to detect copy number variants (CNVs) throughout the whole genome.

The detection rate in different studies is variable from 9% to 44% and it is highly influenced by the resolution of the applied array platform, the clinical selection of patients (family history, associated anomalies) and the criteria used to define the CNVs as pathogenic.

One of the first applications of array-CGH analysis in ASDs was published in 2006. Array-CGH with 1 Mb resolution allowed the identification of 8 clinically relevant CNVs among a cohort of 29 patients with syndromic ASD (~28%).[2] Subsequently, a larger series of 427 unrelated ASD cases has been investigated by single nucleotide polymorphism (SNP) microarray. This analysis identified 277 CNVs in 44% of the patients. These were classified as pathogenic in 14% of cases (~7% de novo and ~7% overlapping/recurrent in 2 or more unrelated samples).[3] In the same year, Christian and colleagues, using BAC array, studied nearly 400 patients and showed an overall CNVs rate of 11.6%.[4] More recently, Qiao Y et al investigated 100 subjects with syndromic ASD using BAC array-CGH and identified 9 pathogenic CNVs (9%) and three rearrangements of unknown significance.[5] In the latter study, the pathogenic role of CNVs was determined on the basis of these criteria: CNVs not reported in the database of genomic variants or in the control population; X-linked CNVs present in males and maternally inherited; and CNVs already reported in ASD.

Overall, array-CGH analysis increased the diagnostic yield in ASDs and allowed the identification of new genetic determinants. In addition to the well known recurrent rearrangements involving the 15q11q13, 16p13 and 22q13 regions, recurrent microdeletions and microduplications at 16p11.2 have been recently identified and are shown to confer susceptibility to ASDs in up to 1% of autistic patients.[3]

Beside these emerging data, the clinical interpretation of several newly discovered private CNVs is still uncertain. In the above reported literature, only a fraction of CNVs ranging from 9% to 27% have been classified as pathogenic. Our analysis focused on the identification of new pathogenic loci and on the practical management of those CNVs with uncertain clinical significance.

METHODS
Study cohort.
A cohort of 95 unrelated patients with ASD with unknown aetiology was selected for this study. Eighty were males and 15 were females (M:F; 5.3:1). The age range was 2 years and 8 months and 22 years and 2 months with a mean of 10 years and 10 months, and median of 10 years and 4 months. This cohort included 3 patients with Asperger syndrome, 48 with Autism and 44 with PDD-NOS. All patients belong to “Cell Lines and DNA Bank of Rett syndrome, X linked mental retardation and other genetic diseases” - Telethon Genetic Biobank Network, and were: Asperger = AU237, MR1297, MR1346; Autism = MR841, MR593, MR755, AU277, MR1081, AU243, MR886, MR1187, AU212, MR681, MR925, PW66, MR1203, AU204, AU303, MR1460, AU274, MR322, AU180, AU295, AU193, MR1475, MR868, MR815, MR940, MR1505, MR798, AU265, AU197, AU229, AU237, MR1181, MR746, MR460, MR1297, MR850, AU186, AU271, AU256, MR1269, MR708, RET196, MR1681, MR1346, MR1537; PDD-NOS = AU135, AU94, AU142, MR790, AU227, MR518, MR148, MR1748, AU183t, XXM1, MR521, MR1005, MR889, MR1708, MR1022,
The ASD diagnosis for all subjects was made by the same child neuropsychiatrist (RC) based on standardized Diagnostic and Statistical Manual 4th Edition (DSM-IV) criteria using Autism Diagnostic Interview-Revised (ADI-R) and/or Autism Diagnostic Observation Schedule Generic (ADOS-G) standards. All patients underwent a comprehensive evaluation by a clinical geneticist (FM or MAM) who excluded a diagnosis of a recognizable syndrome, and all patients, except those with microcephaly, tested negative for FMR1 gene expansion.

**Array-CGH analysis:** Genomic DNA of the patients was isolated from an EDTA peripheral blood sample using the QIAamp DNA Blood Kit according to the manufacturer’s protocol (Qiagen, www.qiagen.com). Genomic DNA of normal male and female controls was obtained from Promega. Ten micrograms of genomic DNA from the patient (test sample) and the control (reference sample) were sonicated. Test and reference DNA samples were subsequently purified using dedicated columns (DNA Clean and Concentrator, Zymo Research) and the appropriate DNA concentrations were determined by a DyNA Quant™ 200 Fluorometer (GE Healthcare).

Array-CGH analysis was performed using commercially available oligonucleotide microarrays containing about 43,000 60-mer probes (Human Genome CGH Microarray 44B Kit, Agilent Technologies, Santa Clara, California) as previously reported.[6] The median functional resolution was 75-100 Kb. Copy number variations (CNVs) were considered significant if they were defined by three or more oligonucleotides, and were not present in the Database of Genomic Variants. Confirmation of results was performed by a second independent experiment. Segregation analysis of each rearrangement was performed in 22/35 (63%) using the same technique.

**Statistical analysis.** Differences in proportions were calculated using Fisher Exact Test with p<0.05 set for significance.

**RESULTS**

**CNV discovery.**
Among the 95 patients, 35 patients exhibited at least one rearrangement (Tables 1 and 2). In these 35 patients, 49 specific rearrangements were identified: 22 deletions and 27 duplications, 45 in autosomes and 4 in the X-chromosome. The size of rearrangements ranged from 2Kb to 1.8Mb. The mean size of rearrangements was 326 Kb. Among the 49 rearrangements two were de novo, 13 were inherited by the mother and 13 by the father, while for 21 inheritance was unknown. Twenty-five patients had only one rearrangement and 10 had two or more rearrangements up to four in one case. Interestingly, three patients had two rearrangements, one inherited by the father and the second by the mother, suggesting digenic inheritance.

In fifteen patients (16%) CNV(s) were considered disease associated. These include two de novo deletions (del16p11.2; delXq12), five known susceptibility regions and eight new potential candidate regions (dup2q11.2, del20p12.1, del7q31.1, dup2p16.3, dup4p16.1, dup5q21.1, del9p13.1p13.2, del17q25.3). In 20 patients (21%) CNVs were classified as uncertain cofactors.
Table 1 Susceptibility regions (Class I)

<table>
<thead>
<tr>
<th>Sample ID</th>
<th>Gender</th>
<th>Chromosome</th>
<th>Rearrangement</th>
<th>Inheritance</th>
<th>Proximal breakpoint</th>
<th>Distal breakpoint</th>
<th>Size (Kb)</th>
<th>Comments</th>
</tr>
</thead>
<tbody>
<tr>
<td>MR518</td>
<td>M</td>
<td>16p11.2</td>
<td>Loss</td>
<td>De novo</td>
<td>29 581 455</td>
<td>30 027 213</td>
<td>450</td>
<td>Susceptibility to ASD.[3]</td>
</tr>
<tr>
<td>AU142</td>
<td>M</td>
<td>Xq12</td>
<td>Loss</td>
<td>De novo</td>
<td>65 732 215</td>
<td>65 811 681</td>
<td>80</td>
<td>One gene: EDI2R, a susceptibility gene for androgenic alopecia.[7] The distal breakpoint is located 1.36 Mb from the XLID gene, OPHN1.</td>
</tr>
<tr>
<td>MR1537</td>
<td>M</td>
<td>16p13.11</td>
<td>Gain</td>
<td>maternal</td>
<td>14 852 061</td>
<td>15 867 585</td>
<td>1.1</td>
<td>Susceptibility to ASD.[8]</td>
</tr>
<tr>
<td>AU221</td>
<td>M</td>
<td>15q13.3</td>
<td>Gain</td>
<td>unknown</td>
<td>29 809 025</td>
<td>30 298 096</td>
<td>500</td>
<td>Susceptibility to ASD.[9]</td>
</tr>
<tr>
<td>XXM1</td>
<td>M</td>
<td>17q12</td>
<td>Gain</td>
<td>paternal</td>
<td>31 925 650</td>
<td>33 726 698</td>
<td>1.8</td>
<td>Susceptibility to cognitive impairment and behavioral abnormalities.[10]</td>
</tr>
<tr>
<td>MR1475</td>
<td>M</td>
<td>11p12</td>
<td>Loss</td>
<td>maternal</td>
<td>40 239 489</td>
<td>41 342 038</td>
<td>1.1</td>
<td>Recurrent and overlapping locus in ASD.[3]</td>
</tr>
<tr>
<td>MR746</td>
<td>M</td>
<td>15q11.2</td>
<td>Loss</td>
<td>unknown</td>
<td>20 335 887</td>
<td>20 636 537</td>
<td>300</td>
<td>Susceptibility to developmental delay, behavioral problems and dysmorphisms.[11]</td>
</tr>
<tr>
<td>MR1346</td>
<td>M</td>
<td>2q11.2</td>
<td>Gain</td>
<td>maternal</td>
<td>99 991 724</td>
<td>100 088 972</td>
<td>97</td>
<td>One gene: AFF3 associated with limb and brain abnormalities when deleted.[12]</td>
</tr>
<tr>
<td>MR681</td>
<td>M</td>
<td>20p12.1</td>
<td>Loss</td>
<td>maternal</td>
<td>14 772 372</td>
<td>15 216 002</td>
<td>444</td>
<td>One gene: C20orf133, disrupted in a patient with Niikawa-Kuroki syndrome.[13]</td>
</tr>
<tr>
<td>MR460 (*)</td>
<td>M</td>
<td>2p16.3</td>
<td>Gain</td>
<td>paternal</td>
<td>47 866 371</td>
<td>47 938 504</td>
<td>72</td>
<td>Two genes including MSH6, mutated in Lynch syndrome. The proximal breakpoint is located 3Mb far from the neurexin I gene which has been associated with schizophrenia.[14]</td>
</tr>
<tr>
<td>MR1392 (*)</td>
<td>M</td>
<td>7q31.1</td>
<td>Loss</td>
<td>unknown</td>
<td>110 917 835</td>
<td>110 989 204</td>
<td>70</td>
<td>One gene: IMMP2L, a candidate gene for ASD.[15]</td>
</tr>
<tr>
<td>AU224</td>
<td>M</td>
<td>4p16.1</td>
<td>Gain</td>
<td>paternal</td>
<td>9 686 502</td>
<td>9 751 028</td>
<td>65</td>
<td>One gene: WDR1, upregulated in the dorsolateral prefrontal cortex in schizophrenia patients.[16]</td>
</tr>
<tr>
<td>AU197 (*)</td>
<td>M</td>
<td>5q21.1</td>
<td>Gain</td>
<td>paternal</td>
<td>100 219 716</td>
<td>100 401 892</td>
<td>182</td>
<td>Only one gene: STRESS44, involved in polysialylation of the neural cell adhesion molecule, a critical step in neural development.[17]</td>
</tr>
<tr>
<td>AU183 (*)</td>
<td>M</td>
<td>9p13.1p13.2</td>
<td>Loss</td>
<td>unknown</td>
<td>37 847 269</td>
<td>38 040 719</td>
<td>193</td>
<td>Three genes, including SHB, playing a role in neuronal cells[18]; and WDR22 (DCAP10) interacting with the XLID gene, CUL4B. [19]</td>
</tr>
<tr>
<td>AU271 (*)</td>
<td>M</td>
<td>17q25.3</td>
<td>Loss</td>
<td>unknown</td>
<td>78 536 419</td>
<td>78 653 660</td>
<td>117</td>
<td>Two genes, including METRNL, which codes a glial cell differentiation regulator-like protein. [20]</td>
</tr>
</tbody>
</table>

(*) These patients have additional CNV of unknown significance, see table 2.
<table>
<thead>
<tr>
<th>Sample ID</th>
<th>Gender</th>
<th>Chromosome</th>
<th>Rearrang</th>
<th>Inheritance</th>
<th>Proximal breakpoint</th>
<th>Distal breakpoint</th>
<th>Size</th>
<th>Gene content</th>
</tr>
</thead>
<tbody>
<tr>
<td>PW66</td>
<td>M</td>
<td>Xp22.31</td>
<td>Gain</td>
<td>maternal</td>
<td>7 565 292</td>
<td>8 226 181</td>
<td>660 Kb</td>
<td>This region includes 3 genes (VCX, FNL1A and VCX2) and is located in distal region involved in Xp22.31 microduplication.</td>
</tr>
<tr>
<td>MR940</td>
<td>M</td>
<td>Xq22.3</td>
<td>Gain</td>
<td>maternal</td>
<td>106 238 368</td>
<td>106 515 657</td>
<td>277 Kb</td>
<td>2 genes</td>
</tr>
<tr>
<td>MR1681</td>
<td>M</td>
<td>Xq26.2</td>
<td>Gain</td>
<td>unknown</td>
<td>130 398 834</td>
<td>130 788 239</td>
<td>390 Kb</td>
<td>2 genes</td>
</tr>
<tr>
<td>AS100</td>
<td>M</td>
<td>1p22.1</td>
<td>Loss</td>
<td>maternal</td>
<td>87 179 292</td>
<td>87 273 767</td>
<td>94 Kb</td>
<td>No genes</td>
</tr>
<tr>
<td>MR952</td>
<td>M</td>
<td>4q32.2</td>
<td>Gain</td>
<td>paternal</td>
<td>162 409 396</td>
<td>162 900 008</td>
<td>493 Kb</td>
<td>1 gene</td>
</tr>
<tr>
<td>RE216</td>
<td>F</td>
<td>9p21.3</td>
<td>Loss</td>
<td>unknown</td>
<td>21 451 614</td>
<td>21 557 336</td>
<td>105 Kb</td>
<td>2 genes</td>
</tr>
<tr>
<td>AS100</td>
<td>M</td>
<td>1q23.1</td>
<td>Gain</td>
<td>paternal</td>
<td>111 421 787</td>
<td>111 470 536</td>
<td>50 Kb</td>
<td>1 gene</td>
</tr>
</tbody>
</table>

Likely digenic inheritance (Class IIb)

<table>
<thead>
<tr>
<th>Sample ID</th>
<th>Gender</th>
<th>Chromosome</th>
<th>Rearrang</th>
<th>Inheritance</th>
<th>Proximal breakpoint</th>
<th>Distal breakpoint</th>
<th>Size</th>
<th>Gene content</th>
</tr>
</thead>
<tbody>
<tr>
<td>AS100</td>
<td>M</td>
<td>1q23.1</td>
<td>Gain</td>
<td>unknown</td>
<td>130 398 834</td>
<td>130 788 239</td>
<td>390 Kb</td>
<td>2 genes</td>
</tr>
</tbody>
</table>

Likely cofactors for additional signs (Class IIc)

<table>
<thead>
<tr>
<th>Sample ID</th>
<th>Gender</th>
<th>Chromosome</th>
<th>Rearrang</th>
<th>Inheritance</th>
<th>Proximal breakpoint</th>
<th>Distal breakpoint</th>
<th>Size</th>
<th>Gene content</th>
</tr>
</thead>
<tbody>
<tr>
<td>PW66</td>
<td>M</td>
<td>Xq26.2</td>
<td>Gain</td>
<td>unknown</td>
<td>130 398 834</td>
<td>130 788 239</td>
<td>390 Kb</td>
<td>2 genes</td>
</tr>
</tbody>
</table>

Unknown significance (Class IId)

<table>
<thead>
<tr>
<th>Sample ID</th>
<th>Gender</th>
<th>Chromosome</th>
<th>Rearrang</th>
<th>Inheritance</th>
<th>Proximal breakpoint</th>
<th>Distal breakpoint</th>
<th>Size</th>
<th>Gene content</th>
</tr>
</thead>
<tbody>
<tr>
<td>MR853</td>
<td>M</td>
<td>1q43</td>
<td>Gain</td>
<td>paternal</td>
<td>234 698 142</td>
<td>234 814 728</td>
<td>116 Kb</td>
<td>2 genes including EDARADD gene associated with anhidrotic ectodermal dysplasia when mutated both in homozygous and heterozygous state.</td>
</tr>
<tr>
<td>MR1505</td>
<td>F</td>
<td>1p21.2</td>
<td>Loss</td>
<td>unknown</td>
<td>101 246 819</td>
<td>101 277 111</td>
<td>29 Kb</td>
<td>1 gene</td>
</tr>
<tr>
<td>MR1748</td>
<td>F</td>
<td>2p15</td>
<td>Gain</td>
<td>unknown</td>
<td>63 459 048</td>
<td>63 670 473</td>
<td>210 Kb</td>
<td>2 genes</td>
</tr>
<tr>
<td>MR886</td>
<td>M</td>
<td>4p16.3p16.2</td>
<td>Gain</td>
<td>unknown</td>
<td>2 902 096</td>
<td>3 114 307</td>
<td>212 Kb</td>
<td>5 genes including HTT gene responsible for Huntington disease in heterozygous state (gain of function).</td>
</tr>
<tr>
<td>AU293</td>
<td>M</td>
<td>5p13.3</td>
<td>Loss</td>
<td>unknown</td>
<td>37 552 360</td>
<td>38 292 964</td>
<td>740 Kb</td>
<td>2 genes including GDNF gene associated with susceptibility to Hirschsprung disease.</td>
</tr>
<tr>
<td>MR1501</td>
<td>M</td>
<td>1q21.1</td>
<td>Gain</td>
<td>unknown</td>
<td>59 944 705</td>
<td>60 678 299</td>
<td>733 Kb</td>
<td>4 genes</td>
</tr>
<tr>
<td>MR384</td>
<td>M</td>
<td>14q21.1</td>
<td>Loss</td>
<td>unknown</td>
<td>40 088 478</td>
<td>40 380 622</td>
<td>292 Kb</td>
<td>No genes</td>
</tr>
<tr>
<td>MR798</td>
<td>M</td>
<td>15q15</td>
<td>Loss</td>
<td>unknown</td>
<td>41 676 219</td>
<td>41 830 276</td>
<td>154 Kb</td>
<td>6 genes including STTC gene associated with deafness when mutated in homozygous state.</td>
</tr>
<tr>
<td>MR1501</td>
<td>M</td>
<td>17q12</td>
<td>Loss</td>
<td>unknown</td>
<td>30 711 469</td>
<td>30 762 521</td>
<td>51 Kb</td>
<td>2 genes</td>
</tr>
<tr>
<td>AU277</td>
<td>M</td>
<td>10q26.3</td>
<td>Gain</td>
<td>unknown</td>
<td>135 202 362</td>
<td>135 222 424</td>
<td>20 Kb</td>
<td>2 genes</td>
</tr>
<tr>
<td>MR1460</td>
<td>M</td>
<td>7q34</td>
<td>Gain</td>
<td>unknown</td>
<td>142 469 982</td>
<td>142 591 458</td>
<td>121 Kb</td>
<td>3 genes</td>
</tr>
<tr>
<td>MR1392</td>
<td>M</td>
<td>2q11.1</td>
<td>Gain</td>
<td>unknown</td>
<td>130 731 967</td>
<td>130 861 076</td>
<td>129 Kb</td>
<td>3 genes</td>
</tr>
<tr>
<td>MR1022</td>
<td>M</td>
<td>10q26.3</td>
<td>Loss</td>
<td>unknown</td>
<td>135 104 029</td>
<td>135 222 424</td>
<td>118 Kb</td>
<td>2 genes</td>
</tr>
<tr>
<td>AU271</td>
<td>M</td>
<td>17q21.31</td>
<td>Loss</td>
<td>unknown</td>
<td>41 544 224</td>
<td>41 706 870</td>
<td>163 Kb</td>
<td>Only one gene KIAA1267. This gene extends into commonly variable region around 41.6Mb and in part its sequence overlaps the distal region involved in 17q21.31 microdeletion syndrome. Note: no OMIM gene.</td>
</tr>
<tr>
<td>PW66</td>
<td>M</td>
<td>Xp22.31</td>
<td>Gain</td>
<td>maternal</td>
<td>955 992</td>
<td>1 131 225</td>
<td>175 Kb</td>
<td>2 genes</td>
</tr>
<tr>
<td>AU193</td>
<td>F</td>
<td>3q13.2</td>
<td>Loss</td>
<td>unknown</td>
<td>113 666 633</td>
<td>113 913 199</td>
<td>245 Kb</td>
<td>4 genes</td>
</tr>
</tbody>
</table>

(*) These patients have one additional pathogenic CNV, see table 1.
(**) These patients have one additional pathogenic CNV, see table 2 – Class IIa.
CNVs affecting susceptibility regions

Six patients showed a rearrangement in known susceptibility regions for ASD: 11p12, 15q11.2, 15q13.3, 16p11.2, 16p13.11, and 17q12 (Table 1).[3] [11] [9] [8] [10] The 16p11.2 rearrangement was proven to be de novo, while the remaining rearrangements were inherited or with unknown origin. The case with 17q12 CNV was previously reported in detail.[21]

While the rearrangements in 16p11.2, 16p13.1 and 17q12 fall within the range of a previously reported susceptibility region [3] [8] [10], the rearrangements in 15q13.3, 15q11.2 and 11p12 call for specific description. [3] [22] [23] [24] [25] [8] [26] [27] [11] Szafranski et al. recently investigated the microduplication in 15q13.3 of about 500kb that includes the CHRNA7 gene and overlaps the class 3 microduplication. They demonstrated that the mutation may exert a predisposition to neurodevelopmental and neuropsychiatric phenotypes in association with other genetic modifiers.[9] In our case, the de novo or inherited origin of the rearrangement could not be assessed: the CHRNA7 duplication was excluded in mother’s DNA, while father was not available for the analysis. The family history was intensely investigated for the presence of cognitive impairment or neuropsychiatric disorder, which were both ruled out.

The microdeletion 15q11.2 was recently described by Doornbos et al, who suggested a possible predisposing effect to developmental delay, behavioural problems and dysmorphic features, such as hypertelorism, palate and ear anomalies.[11] The rearrangement, partially overlapping PWS/AS critical region, is of about 300 kb and it includes four genes. Three of them seem to be important for neurological development. The NIPA1 gene is highly expressed in neuronal tissue and encodes a magnesium membrane transporter. Gain of function mutations in this gene have been associated with autosomal dominant spastic paraplegia 6.[28] The CYFIP1 gene is a protein that interacts with FMRP (fragile X mental retardation protein) and GTPase RAC1, a protein involved in the development and maintenance of neuronal structures.[29] A third gene, TUBGCP5, which is highly expressed in the subthalamic nuclei of the brain, could be involved in the behavioral disorders, such as ADHD and obsessive-compulsive behavior.[30]

Another patient had a deletion in 11p12, a recurrent rearrangement in ASD patients in the series described by Marshall et al.[3] The deletion in our patient is of about 1 Mb and is larger than the deletion reported by Marshall et al. It disrupts only one gene, NGL1, which is highly expressed in the striatum and in cerebral cortex, and is a specific binding partner for NTNG1, a member of the netrin family of axon guidance molecules. Functional experiments using mouse thalamic neurons and chick embryos suggested that the interaction of the two genes is crucial for the growth of thalamocortical axons.[31]

Identification of novel susceptibility regions

Nine new susceptibility regions were identified on the basis of either the gene content or de novo inheritance (Table 1). One case had a de novo deletion of about 100 Kb in Xq12, including a single gene: EDA2R. This gene, also known as XEDAR, is the X-linked ectodysplasin receptor, a type III transmembrane protein with an extracellular N terminus and a cellular cytoplasmic C terminus.[32]

The remaining eight CNVs either included genes that, according to their expression and function or according to literature data, were good candidates for the neurological phenotype or fall at 1-3Mb distance from such genes (Table 1). The 7q31.1 deletion includes the IMMP2L gene, a candidate gene for ASD by association studies.[15] The 9p13.1p13.2
deletion contains the *SHB* gene, which plays a role in neuronal cells, and the *WDR32* gene (also known as DCAF10), which interacts with *CUL4B*, an X-linked mental retardation gene. In one male patient an inherited 20p12.1 deletion of nearly 450 Kb interrupting the C20orf133 has been identified. In 2007 a *de novo* deletion of the same gene was identified in a patient with a clinical diagnosis of Niikawa-Kuroki syndrome. Our patient does not show clinical features of this syndrome and the analysis of different cohorts of patients with Niikawa-Kuroki syndrome did not reveal any C20orf133 mutation. Nevertheless, this gene remains a good candidate gene for the neurocognitive phenotype.

The duplication in 2p16.3 included two genes: *MSH6* and *FBXO11* (also known as *VIT1*). MSH6 is a component of the post-replicative DNA mismatch repair system and defects in *MSH6* are the cause of Lynch syndrome. *FBXO11* is known to be down-regulated in melanocytes from patients with vitiligo, a skin disorder that results in depigmentation, while polymorphisms in this gene are associated with chronic otitis media with effusion and recurrent otitis media. The duplication of *FBXO11* may be responsible for the presence in the patient of hyperchromic areas localized in the abdomen. About 3Mb from the proximal breakpoint is the neurexin 1 gene (*NRXN1*), which has been implicated in a variety of conditions including autism, schizophrenia, and nicotine dependence.

The small duplication on 4p16.1 included only the *WDR1* gene. This gene is included in a 200 kb region linked by haplotype analysis to bipolar affective disorder. *WDR1* encodes a brain expressed protein that provides protein–protein interactions thought to induce the disassembly of actin filaments. This function offers a plausible mechanism for neuronal dysfunction. The 5q21.1 duplication included the *ST8SIA4* gene that promotes neurite outgrowth and sprouting and has a role in development and neuronal plasticity. The deletion in 17q25.3 included only two genes: *B3GNTL1* and *METRN*. *B3GNTL1* codes for a glycosyltransferase, a subgroup of enzymes involved in a wide range of functions in all living organisms. *METRN* is a small gene, with a guanine cytosine (GC)-rich 5′ untranslated region resembling the features of genes involved in epigenetic control. The gene product is a protein akin to meteoin, a secreted protein expressed in undifferentiated neural progenitors and in the astroglial lineage. In the embryonal stages, meteoin is involved in the development of the nervous system, contributing to glial cell differentiation and axonal network formation. The small duplication on 4p16.1 included only the *WDR1* gene. This gene is included in a 200 kb region linked by haplotype analysis to bipolar affective disorder. *WDR1* encodes a brain expressed protein that provides protein–protein interactions thought to induce the disassembly of actin filaments. This function offers a plausible mechanism for neuronal dysfunction. The 5q21.1 duplication included the *ST8SIA4* gene that promotes neurite outgrowth and sprouting and has a role in development and neuronal plasticity. The deletion in 17q25.3 included only two genes: *B3GNTL1* and *METRN*. *B3GNTL1* codes for a glycosyltransferase, a subgroup of enzymes involved in a wide range of functions in all living organisms. *METRN* is a small gene, with a guanine cytosine (GC)-rich 5′ untranslated region resembling the features of genes involved in epigenetic control. The gene product is a protein akin to meteoin, a secreted protein expressed in undifferentiated neural progenitors and in the astroglial lineage. In the embryonal stages, meteoin is involved in the development of the nervous system, contributing to glial cell differentiation and axonal network formation. The 2q11.2 duplication includes a single gene, *AFF3* (also known as *LAF4*), which is interrupted. This gene, which belongs to a family of four genes that includes *FMR2*, encodes a nuclear transcriptional activator that is preferentially expressed in lymphoid tissue, and that may function in lymphoid development and oncogenesis. Haploinsufficiency for *AFF3* has been recently associated with brain, limb and urogenital anomalies.

**Uncertain cofactors**

A total of 34 CNVs of unknown significance were identified (Table 2). Twenty-eight were present in 20 patients, who had only this kind of rearrangements. Instead, four patients had the remaining 6 rearrangements of unknown significance in association with a pathogenic CNV (Table 1). All these CNVs were neither reported in the database of genomic variant nor previously identified in other controls/patients. They ranged in size from 2Kb to 940Kb and they had a gene content from 0/1 to 7 genes. The mean content of genes was 2.5.

**Correlation of clinical features with CNVs**

Clinical data of all patients is provided in Table 3. Between the cohort of patients with pathogenic CNVs and the cohort without CNVs, no statistically
relevant differences were identified in gender, weight, head circumference and peculiar facial features. A statistically significant correlation was found between epilepsy and absence of CNVs, especially considering cases without CNVs versus cases with CNVs - all classes (p=0.002). A nearly statistically significant association was identified between congenital anomalies and the presence of CNVs - class I only (p=0.09). CNVs - all classes are over-represented in patients with Autism in respect to PDD-NOS, both considering the solely group of patients with Autism (p=0.06) and that with Autism plus Asperger (p=0.07). In addition CNVs - all classes are under-represented (p=0.06) among patients with height >97°cnt. 

### Table 3 Comparison of clinical features of patients with and without CNVs.

<table>
<thead>
<tr>
<th>Phenotypes</th>
<th>All cases</th>
<th>Cases without CNVs</th>
<th>Cases with CNVs - all classes</th>
<th>Cases with CNVs - class I only</th>
<th>P value (a)</th>
<th>P value (b)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Total cases</strong></td>
<td>95</td>
<td>60</td>
<td>35</td>
<td>15</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Gender</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>80</td>
<td>50</td>
<td>30</td>
<td>15</td>
<td>0.22</td>
<td>0.09</td>
</tr>
<tr>
<td>Female</td>
<td>15</td>
<td>10</td>
<td>5</td>
<td>0</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>ASD</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Autism</td>
<td>48</td>
<td>27</td>
<td>21</td>
<td>7</td>
<td>0.06</td>
<td>0.22</td>
</tr>
<tr>
<td>PDD-NOS</td>
<td>44</td>
<td>31</td>
<td>13</td>
<td>7</td>
<td>0.07</td>
<td>0.21</td>
</tr>
<tr>
<td>Asperger</td>
<td>3</td>
<td>2</td>
<td>1</td>
<td>0</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td><strong>ID</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Yes (IQ&lt;70)</td>
<td>92</td>
<td>58</td>
<td>34</td>
<td>15</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>No (IQ≥70)</td>
<td>3</td>
<td>2</td>
<td>1</td>
<td>0</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Weight &lt;3rd centile</strong></td>
<td>4</td>
<td>3</td>
<td>1</td>
<td>0</td>
<td>0.38</td>
<td>0.51</td>
</tr>
<tr>
<td><strong>Height &lt;3rd centile</strong></td>
<td>4</td>
<td>3</td>
<td>1</td>
<td>0</td>
<td>0.38</td>
<td>0.51</td>
</tr>
<tr>
<td><strong>Weight &gt;97th centile</strong></td>
<td>17</td>
<td>11</td>
<td>6</td>
<td>3</td>
<td>0.22</td>
<td>0.28</td>
</tr>
<tr>
<td><strong>Height &gt;97th centile</strong></td>
<td>13</td>
<td>11</td>
<td>2</td>
<td>3</td>
<td>0.06</td>
<td>0.28</td>
</tr>
<tr>
<td>Microcephaly</td>
<td>5</td>
<td>2</td>
<td>3</td>
<td>0</td>
<td>0.20</td>
<td>0.64</td>
</tr>
<tr>
<td>Macrocephaly</td>
<td>20</td>
<td>14</td>
<td>6</td>
<td>3</td>
<td>0.16</td>
<td>0.26</td>
</tr>
<tr>
<td>Epilepsy</td>
<td>12</td>
<td>11</td>
<td>1</td>
<td>0</td>
<td>0.02</td>
<td>0.07</td>
</tr>
<tr>
<td>Peculiar facial features</td>
<td>59</td>
<td>38</td>
<td>21</td>
<td>9</td>
<td>0.16</td>
<td>0.22</td>
</tr>
<tr>
<td>Congenital anomalies</td>
<td>23</td>
<td>13</td>
<td>10</td>
<td>6</td>
<td>0.14</td>
<td>0.09</td>
</tr>
</tbody>
</table>

Note a, b: number of patients.
Note d: P value comparing cases without CNVs versus cases with CNVs (all classes).
Note e: P value comparing cases without CNVs versus cases with CNVs (class I).
DISCUSSION

We report the investigation of CNVs in a cohort of 95 patients with ASD. We could confidently assign pathogenicity in 15 cases, which corresponds to about 16% of the cohort. This is higher that the 9% and 11.6% reported by Qiao Y et al 2009 and Christian et al 2008, respectively, owing to the higher resolution of oligo-array platform in respect to BAC platform. [5] [4] On the basis of this, the beneficial use of higher resolution array is clear. Our detection rate, considering only pathogenic CNVs, is similar to that reported in Marshall 2008, 14%, who used SNP array. [3] We prioritized CNV pathogenicity using the following criteria. Firstly, we considered the de novo inheritance of CNV (class Ia of table 1). Although it is possible to detect a de novo event of a benign variant, in the clinical practice we suggest that all de novo events be considered as pathogenic, keeping in mind that using this as the only criterion, a percentage of false positive cases will be included. As a second criterion, we used the literature data on known susceptibility CNVs (class Ib). We suggest clearly stating in genetic counseling reports, that these CNVs, although in most cases inherited from healthy parents, are a cofactor of the disease. As a third criterion we used literature data on association studies and knowledge on gene function/expression (class Ic). In this case, we propose that it should be indicated in the medical report that the region is a “likely” cofactor in the disease. Overall, in our opinion CNVs of class Ia, Ib and Ic should be considered as pathogenic.

As an example of class Ic, we discuss more extensively the case of the 7q31 deletion. The involvement of the 7q31 deletion is supported by three pieces of evidence. Firstly, this region includes the IMMP2L gene, designated autism susceptibility locus 1 (AUTS1). Secondly, a disruption of IMMP2L has been reported in a patient with Tourette syndrome. Lastly, IMMP2L intron 3 contains a gene that is highly expressed in fetal brain (LRRN3).[15]

In addition to the 15 cases with pathogenic CNVs, we identified CNVs of unknown significance (class II) in 20 cases, which correspond to about 21% of the cohort. This group, which is even larger than class I group in our cohort, represents a challenge in genetic counseling. Previous studies reported this category at a variable rate from 3% in Qiao et al to 30% in Marshall et al.[35] [3] In these cases a deep investigation of gene content together with 1Mb of the surrounding region is essential. We suggest that these CNVs should be considered to be “uncertain cofactors” and to clearly state in genetic counseling reports that “an involvement of this CNV in the disease is not completely ruled out”.

CNVs in class II can be prioritized using the following combination of criteria. Firstly, we considered those rearrangements on the X chromosome identified in a male patient and maternally inherited, since they could potentially imply an X-linked recessive defect (class IIa of table 2). Secondly, we considered the concurrent inheritance of two rearrangements, one from the mother and one from the father, potentially to imply digenic inheritance (class IIb).

As a third criterion, we examined literature data pertaining to genes that were potentially implicated in secondary, but not necessarily psychiatric neurological, phenotypic signs. These data implied that rearrangements in these genes were, at least partially, correlated with the phenotype (class IIc). One example of this category is the maternally inherited 3q29 microduplication that included the RPL35A gene, whose loss of function mutations in heterozygous state cause Diamond-Blackfan Anemia. In this condition macrocytic anemia, short stature, microcephaly and mental retardation are present. Interestingly, our patient (AU271) has mental retardation associated with microcytic anemia, tall stature, and
macrocephaly, these latter characteristics being present also in the mother. Therefore, one can be hypothesized that the RLP35A gene duplication is responsible for the reciprocal phenotype of Diamond-Blackfan Anemia present both in the patient and in the mother, while the susceptibility to autism of the patient was well explained by the concurrent 11p12 microdeletion (table 1-class Ib). All the remaining cases that do not fit the three previous categories are classified as class IId.

Overall, in order to obtain a correct interpretation of array CGH output in Autistic Spectrum Disorders we suggest not only accurate literature revision and gene content analysis but also accurate parents investigation both from molecular and clinical point of view. This process makes array-CGH analysis among one of the more complex of genetic investigation. In explaining class Ila output during the genetic counseling a recurrence risk up to 50% of male sibling have to be considered. In class IIb the consultant has to explain that the recurrence risk may rise up to the 25% independently to the gender. In class IIc specific clinical signs may be allocated to specific rearrangements potentially leading the accurate prediction of fetus phenotype. In class IId output it could be clearly stated that the recurrence risk is not predictable.

Overall, in this cohort CNVs were identified in 35 patients out of 95, corresponding to a detection rate of 37%. Duplications and deletions had the same frequency even considering only pathogenic CNVs (class I). These data are in contrast with the CNVs identified in syndromic ID in which deletions are over-represented.[36] In syndromic ID the rearrangement is usually de novo, and it is considered the major determinant of the disease. The reciprocal duplication is usually associated with a milder phenotype that may escape CNVs investigation. On the contrary, ASDs are considered multifactorial diseases in which the CNV may represent one among several etiological factors. Therefore, deletions and duplications are expected to occur at the same rate.

Our most important finding coming from the correlation of clinical features with CNVs was that positive output of array-CGH analysis is more expected in Autism and Asperger than in PDD-NOS pinpoint that CNV may represent a stronger cofactor in the first two disorders. While the association with CNVs and congenital anomalies was expected, the inverse relationship with epilepsy is a quite unexpected finding. This finding deserves to be confirmed in other larger cohort of patients.

In conclusion, our study confirmed that array-CGH analysis is able to detect the underlying genetic susceptibility factors in a consistent number of ASD patients, strongly indicating that it has became an essential diagnostic tool for assessing ASD patients. Moreover, this study has provided additional data on the genetic causes of ASD, and has proposed to take into account the possible interaction between genes located in different chromosomal regions when attempting to identify new disease genes. Furthermore, our findings underscore a challenge in genetic counselling in interpreting array-CGH data in a consistent number of patients, and we propose some practical recommendations on the management of such data and on the communication of these findings to the family.

Acknowledgements
This work was supported by Ricerca Finalizzata 2007 Ministero della Salute to A.R. and by PAR 2006 University of Siena to FM. We acknowledge “Cell Lines and DNA Bank of Rett syndrome, X linked mental retardation and other genetic diseases” (Medical Genetics-Siena)-Telethon Genetic Biobank Network (Project No. GTB07001C)

The Corresponding Author has the right to grant on behalf of all authors and does grant on behalf of all authors, an exclusive licence (or non-exclusive for government employees) on a worldwide basis to the BMJ Publishing Group Ltd and its Licensees to permit this article (if accepted) to be published in Journal of Medical Genetics and any other BMJPL products to exploit all subsidiary rights, as set out in our licence.
References


13. Maas NM, Van de Putte T, Melotte C, Francis A, Schrander-Stumpel CT, Sanlaville...


4.2.2 Genetic modifier factor
Investigation of modifier genes within copy number variations in
Rett syndrome.


Submitted to Molecular Autism
Investigation of modifier genes within copy number variations in Rett syndrome.

Rosangela Artuso1*, Filomena T Papa1*, Mafalda Mucciolo1, Dag H Yasui2, Keith W Dunaway2, Vittoria Disciglio1, Maria Antonietta Mencarelli1, Marzia Pollazzon1, Francesca Ariani1, Francesca Mari1, Janine M LaSalle2, Alessandra Renieri1§

1 Medical Genetics, University of Siena Italy
2 Medical Microbiology and Immunology, Genome Center, School of Medicine, University of California, Davis, CA
*These authors contributed equally to this work
§Corresponding author

Abstract

Background Mutations in the X-liked MECP2 gene are responsible for two different phenotypes in females, classical and Zappella variant Rett syndrome. In classical Rett, girls show somatic hypoevolutism, are unable to speak, possess severely impaired motor function and have acute mental retardation. Zappella variant patients recover the ability to speak in either simple or complex often ecolalic phrases, regain purposeful hand function, have moderate mental retardation and generally show normal head circumference, weight and height.

Methods We explored the hypothesis that Copy Number Variants (CNVs) leading to over/underexpression of genes may modulate the phenotype by comparing by array-CGH two discordant pairs of Rett sisters and four additional discordant pairs of unrelated Rett girls matched by mutation type. By ChIP-chip analysis we also identify hypothetical MeCP2 targets included in the identified CNVs.

Results Our study failed to identify one major common modifier gene/region, suggesting genetic modifiers may be complex and variable between cases. However, we demonstrated that several common CNVs are enriched for MeCP2 bound promoters and could influence the severity of the phenotype in complex ways.

On 1p36.13 CROCC, an hypothetical MeCP2 target, results duplicated in a Zappella variant and deleted in a classic Rett patient. It encodes for a major structural component of the ciliary rootlet and ciliary motility is required for correct brain development and function. On 1q31.3 CFHR1 and CFHR3, deleted in a Zappella variant and duplicated in two classic Rett patients, are involved in the regulation of complement whose proteins are important for CNS synapse elimination. The duplication on 10q11.22, present in two Zappella variant patients, includes GPRIN2, involved in the regulation of neurite outgrowth and PPYRI, a key regulator of energy homeostasis and directly implicated in the regulation of food intake.

Conclusions Finding genetic modifiers of the Rett phenotype by CNVs analyses is by nature complex because MeCP2 pathways are influenced by both genetic and epigenetic variables. However this study highlights important hints to further investigate by expression and/or statistical analysis in a higher number of patients.
**Background**

Rett syndrome (RTT, OMIM#312750) is an X-linked neurodevelopmental disorder predominantly affecting females. In the classic form, after a period of normal development (6-18 months), patients show growth retardation and regression of speech and purposeful hand movements, with appearance of stereotyped hand movements, microcephaly, autism, seizures, and somatic hypoevolutism [1] [2]. RTT syndrome has been the object of extensive investigations, revealing a wide spectrum of clinical phenotypes including: the Zappella variant (Z-RTT), the early onset seizure variant, the congenital variant, the ‘forme fruste’, and the late regression variant [3]. Z-RTT, firstly described by M. Zappella in 1992, represents the most common RTT variant and it is characterized by the recovery of the ability to speak in single words or third person phrases and by an improvement of purposeful hand movements [4] [5]. Z-RTT patients also show milder intellectual disabilities (up to IQ of 50) and often normal head circumference, weight and height respect to classic RTT [5].

De novo mutations in the MECP2 gene (Xq28) account for the majority of girls with classic RTT (95%) and for about half of cases with Z-RTT. Only a few familial cases have been reported. Some cases have been explained by skewing of X-inactivation towards the wild type allele in an asymptomatic carrier [6] [7]. In others cases, germline mosaicism has been a possible explanation [8] [9] [10].

RTT symptoms can vary depending on the pattern of X-chromosome inactivation, the type of MECP2 mutation, and environmental factors [5] [11] [12]. However, all these factors are not sufficient to explain RTT clinical variability suggesting that other elements may be involved. In fact there are cases of RTT sisters with identical MECP2 mutation, balanced X-inctivation, similar environmental stimulation and discordant phenotype (one classic and one Z-RTT sister) [6] [8].

Copy Number Variations (CNVs) are segments of DNA ranging from kilobases to multiple megabases in length that are present at a variable number of copies compared with the reference genome sequence. It has been demonstrated that CNVs are associated with quantitative gene expression levels that in some cases are likely to have causative, functional effects [13]. CNVs have been reported to be associated with human diseases such as neurological and autoimmune disorders and cancer [14] [15] [16]. CNVs, to a larger extent than Single Nuclotide Polimorphisms (SNPs), represent an important source of variability in both phenotypically normal subjects and individuals with diseases [17] [18]. It is therefore reasonable to hypothesize that CNVs can modulate the phenotypic expression of RTT syndrome.

In order to test this hypothesis, we analyzed by array-CGH two couples of RTT sisters and four additional couples of unrelated RTT girls matched by mutation type showing discordant phenotype (classic and Z-RTT). By ChIP-chip analysis we also identify hypothetical MeCP2 targets included in the identified CNVs.

**Methods**

**Patients**

From the Italian RTT database and biobank, we recruited two rare familial cases with two RTT sisters with discordant phenotype: one classic (#897 and #138) and one Z-RTT (#896 and #139) (www.biobank.unisi.it) [19]. These cases were screened by both DHPLC and MLPA techniques to identify MECP2 mutations. The first couple bears a MECP2 large deletion involving exon 3 and exon 4, while the second one has a late truncating mutation: c.1157del32. Clinical descriptions of these patients have been...
reported in previous manuscripts [6] [8]. Furthermore, we selected four more pairs (#565/601, #185/119, #421/109, #402/368) of unrelated RTT patients showing discordant phenotype (classic and Z-RTT) coupled by MECP2 mutation type (c.1163del26, p.R306C, c.1159del44, p.R133C). All cases contained in the bank have been clinically evaluated by the Medical Genetics Unit of Siena. Patients were classified in classic and RTT variant according to the international criteria [20].

**Genomic DNA isolation**

Genomic DNA of the patients was isolated from an EDTA peripheral blood sample using the QIAamp DNA Blood Kit according to the manufacturer’s protocol (Qiagen, www.qiagen.com). Genomic DNA of normal male and female controls was obtained from Promega. Ten micrograms of genomic DNA from the patient (test sample) and the control (reference sample) were sonicated. Test and reference DNA samples were subsequently purified using dedicated columns (DNA Clean and Concentrator, Zymo Research) and the appropriate DNA concentrations were determined by a DyNA Quant™ 200 Fluorometer (GE Healthcare).

**Array Comparative Genomic Hybridization**

Array CGH analysis was performed using commercially available oligonucleotide microarrays containing about 99,000 60-mer probes with an estimated average resolution of about 65 Kb. Physical positions of the probes correspond to the UCSC genome browser - NCBI build 36/hg18, March 2006 (http://genome.ucsc.edu).

DNA labeling was performed according to the Agilent Genomic DNA Labeling Kit Plus protocol (Oligonucleotide Array-Based CGH for Genomic DNA Analysis 2.0v). Genomic DNA (3.5 μg) from patients with classical Rett syndrome was mixed with Cy5-dNTP while genomic DNA from patient with Zappella variant was mixed with Cy3-dNTP, as previously reported [21]. The array was disassembled and washed according to the manufacturer protocol with wash buffers supplied with the Agilent 105A kit. The slides was dried and scanned using an Agilent G2565BA DNA microarray scanner.

**Array-CGH image and data analysis**

Image analysis was performed using the CGH Analytics software v. 5.0.14 with aberration algorithms Z-Score and ADM-1 provided by the CGH-Analytics software. The software automatically determines the fluorescence intensities of the spots for both fluorochromes performing background subtraction and data normalization, and compiles the data into a spreadsheet that links the fluorescent signal of every oligo on the array to the oligo name, its position on the array and its position in the genome. The linear order of the oligos is reconstituted in the ratio plots consistent with an ideogram. The ratio plot is arbitrarily assigned such that gains and losses in DNA copy number at a particular locus are observed as a deviation of the ratio plot from a modal value of 1.0.

**Identification of MeCP2 bound promoters**

ChIP-chip analysis was performed as described previously [22] using a modified protocol from Öberley et al. [23]. Briefly, chromatin from three separate SH-SY5Y neuroblastoma cultures differentiated by 48h treatment with phorbol 12-myristate 13-acetate (PMA) was crosslinked by addition of formaldehyde to 1% by volume. Crosslinked cells were lysed by incubation in SDS containing buffer and dounce homogenization and then sonicated to fragment chromatin. MeCP2 bound chromatin was isolated by immunoprecipitation with a custom anti-MeCP2 antibody. As a control, chromatin was immunoprecipitated in parallel with anti-RNA polymerase II antibody (Covance, Berkeley, CA). DNA was isolated from immunoprecipitated
chromatin by proteinase K digestion and phenol/chloroform extraction prior to linear amplification by ligation mediated PCR (LM-PCR).

Amplicons from ChIP samples were subsequently labeled with Cy5 according to Nimblegen protocol (Roche Nimblegen, Madison, WI) and hybridized along with Cy3 labeled sonicated total SH-SY5Y genomic DNA to a commercial genome wide promoter microarray produced by Nimblegen according to the manufacturer’s instructions. In this 1.5 kb promoter array, tiled oligonucleotide probes extend 1.3 kb upstream and 0.2 kb downstream of the transcriptional start sites of 24,275 human transcripts. Arrays were washed, dried and scanned using a GenePix 4000B scanner (Molecular Devices, Sunnyvale, CA) to obtain raw fluorescent intensity data. NimbleScan software (Nimblegen) was used to assemble log2 signal intensity for each tiled oligonucleotide for further analysis. Statistical analysis of promoter ChIP-chip data indicated that 2600-4300 promoters were bound by MeCP2 with 1524 promoters common to two replicate hybridizations. Promoters were ranked according to MeCP2 binding levels for the two arrays.

Analyses on phenotypically discordant RTT pairs resulted in 28 CNV’s that gave a list of 55 genes which could potentially modify RTT phenotype. A Perl script was used to search for these 55 genes within two experimentally replicated promoter hit files that ranked MeCP2 binding based on ChIP-chip analysis. Aliases of genes were acquired from genome.ucsc.edu and genelist.org.

**Results**

Results of CNVs analysis in phenotypically discordant RTT pairs are shown in Tables 1 and 2. Overall, we indentified 29 CNVs, 28 of them being known polymorphic regions and one being an apparently private rearrangement duplicated in only one Z-RTT patient. This private CNV localized to 3q13.12, extended for about 99 Kb and contained two coding genes: GUCA1C expressed in retina and MORC1 expressed in testis.

Some polymorphic CNVs were found in more than one pair of RTT pairs. Among the 28 CNVs, we considered 14 of them as “unlikely modifier” since results were apparently not associated with phenotypic severity (Table 2). These include regions containing olfactory receptors and class II HLA molecules that are not expected to directly correlate with the phenotypic variability related to classic/Z-RTT phenotype.

To determine if the CNVs found in phenotypically discordant RTT pairs contained possible MeCP2 target genes, we compared promoter rankings of MeCP2 binding using promoter-wide ChIP-chip analysis [22]. The ranking from total number of genes from 1 to 24,134 is shown for two replicate MeCP2-ChIP microarrays (MeCP2 B and MeCP2 C promoter hits rank, Tables 1 and 2). Genes with promoters in the top 10% of MeCP2 promoter hits for at least one replicate are shaded in blue. Among CNVs classified as “likely modifiers”, ChIP-chip analysis identified potential MeCP2 target genes within the 1p36.13 (CROCC gene whose duplication was found in the Z-RTT # 896 and deletion in the classic form #402) and the 2p25.2 (TSSC1 gene whose deletion was found in the Z-RTT #896) region. Among CNVs classified as “unlikely modifiers”, ChIP-chip analysis identified potential MeCP2 target genes on 14q11 (OR4Q3 and OR4Q1, deleted in a classic patient #138 and duplicated in another classic patient #421) and on 16p11.2 (NFATC2IP and SPNS1, duplicated in both a classic #897 and a Z-RTT patient #368).

**Discussion**

In order to test the hypothesis that CNVs may modulate RTT phenotype, we analyzed by array-CGH two couples of RTT sisters and four additional couples of unrelated RTT girls matched by MECP2 mutation type showing discordant phenotype: classic and Z-RTT. Our study
failed to identify one major common modifier gene/region, suggesting genetic modifiers may be complex and variable between cases (Tables 1 and 2). In total we found 29 CNVs that were divided into two groups: “likely modifiers” and “unlikely modifiers” (Tables 1 and 2).

Among the first group, the rearrangement on 1p36.13 includes CROCC (ciliary rootlet coiled-coil), that represents an interesting potential modifier gene. This gene is duplicated in the Z-RTT patient # 896 and deleted in the classical patient # 402, suggesting that change in its expression may modulate RTT outcome. Moreover, according to ChIP-chip analysis, CROCC could be a potential MeCP2 target gene (Table 1). CROCC encodes for a major structural component (Rootletin) of the ciliary rootlet, a cytoskeletal-like structure in ciliated cells which originates from the basal body at the proximal end of a cilium and extends proximally toward the cell nucleus. In nonciliated cells, a miniature ciliary rootlet is located at the centrosome and does not project a fibrous network into the cytoplasm. Rootletin is expressed in retina, brain, trachea and kidney [24]. Cilia generate specialized structures that perform critical functions of several broad types: sensation, development, fluid movement, sperm motility, and cell signaling. Their functional significance in tissues is reflected in the severity and diversity of pathologies caused by defects in cilia. These include anosmia, retinitis pigmentosa and retinal degeneration, polycystic kidney disease, diabetes, neural tube defects and neural patterning defects, chronic sinusitus and bronchiectasis, obesity, heterotaxias, polydactyly, and infertility [25] [26] [27]. Defects in cilia are therefore underlying causes of several diseases with pleiotropic symptoms [28]. Several pleiotropic disorders (Bardet-Biedl syndrome, Alstrom syndrome, Meckel-Gruber syndrome and Joubert syndrome) caused by disruption of the function of cilia present mental retardation or other cognitive defects as part of their phenotypic spectrum [29]. The presence of cilia in different types of neurons supports the notion that dysfunction in specific neuronal populations might explain, at least in part, such defects. If MeCP2 acts as a positive regulator of CROCC, it can be hypothesized that higher protein level due to three copies of such gene may counteract the MeCP2 functional reduction, while lower protein level due to single gene copy may worsen the phenotype.

The CFHR gene family members (CFHR1 and CFHR3) located on 1q31.3 are duplicated in classic girls ( #185 and #402) and deleted in Z-RTT (#368), suggesting that the phenotype may benefit of reduced expression of these proteins involved in complement regulation [30]. The complement system is a tightly controlled part of the host innate immune defense. Imbalances in this control contribute to tissue injury and can result in autoimmune diseases. In particular, CFHR1 and CFHR3 was previously associated with hemolytic uremic syndrome (HUS) and age related macular degeneration (AMD) [31] [32] [33]. It is well known that the immune system participate in the development and functioning of the CNS and an immune etiology for RTT and autism has been recently hypothesized [34]. Interestingly, complement proteins have been demonstrated to be fundamental for CNS synapse elimination [35]. Morphological studies in postmortem brain samples from RTT individuals described a characteristic neuropathology which included decreased dendritic arborization, fewer dendritic spines, and increased packing density [36]. It is therefore possible that CFHR genes could be involved in the modelling of synaptic connections and that they could influence RTT severity.

The duplication on 10q11.22, present in two Z-RTT patients (#139 and #109), includes two interesting candidate modifier genes: GPRIN2 and PPYR1. The GPRIN2 gene, highly expressed in the cerebellum, interacts with activated
members of the Gi subfamily of G protein α subunits and acts together with GPRIN1 to regulate neurite outgrowth [37]. PPYR1, also named as neuropeptide Y receptor or pancreatic polypeptide 1, is a key regulator of energy homeostasis and directly involved in the regulation of food intake. Previous studies have reinforced the potential influence of PPYR1 on body weight in humans [38]. Moreover, it has been demonstrated that PPYR1 knockout mice display lower body weight and reduced white adipose tissue [39]. Thus a higher level of PPYR1 expression due to duplication may correlate with the higher body weight characterizing Z-RTT patients respect to classic RTT [5]. In contrast, a recent study demonstrated that 10q11.22 gain is associated with lower body mass index value in the Chinese population. However this CNV is bigger respect to the one reported here and includes two additional genes [40].

The 2p25.2 region, deleted in the Z-RTT patient #896, contains a potential MeCP2 target gene, TSSC1 (Table 1). Very little is known about this gene. It shares sequence homology with RbAp48 and CAF1, suggesting that it could play a role in gene silencing [41].

The 1q42.12 region, duplicated in one Z-RTT patient (#896), includes ENAH. This gene was identified as a mammalian homolog of Drosophila Ena and initially named Mena (Mammalian enabled) [42]. It localizes to cell-substrate adhesion sites and sites of dynamic actin assembly and disassembly. It is a member of the Ena/VASP family that also includes VASP and EVL in vertebrates. Work carried out in Drosophila, C. elegans and mice showed that these proteins participate in axonal outgrowth, dendritic morphology, synapse formation and also function downstream of attractive and repulsive axon guidance pathways [43]. Previous evidence shows that knocking out the three murine ENA/VASP proteins results in a blockade of axon fiber tract formation in the cortex in vivo, and that failure in neurite initiation is the underlying cause [44] [45]. ENAH therefore represents an interesting candidate gene as modifier in RTT. Further investigations are necessary in order to test whether the duplication of ENAH gene in Z-RTT #896 effectively corresponds to increase mRNA levels in brain and whether this mechanism is confined to one couple of discordant girls or is a common mechanism in Z-RTT possibly throughout SNP modulation.

**Conclusions**

One limitation of our study is that the number of patients is too low to perform a statistically significant analysis of CNVs in classic and Z-RTT and this is principally due to the difficulty in recruiting Z-RTT cases. Another limitation is that mRNA expression analysis has not been performed. However, this analysis would not be conclusive because it could be executed only in blood or other accessible human tissues. Regions/genes found rearranged could rather represent interesting hints for further studies in animal models or in new cellular models such as human induced pluripotent stem (iPS).

**Competing interests**

The authors declare that they have no competing interests.

**Acknowledgments**

This work was supported by Ricerca Finalizzata 2007 Ministero della Salute to A.R. and by PAR 2006 University of Siena to FM. We acknowledge “Cell Lines and DNA Bank of Rett syndrome, X linked mental retardation and other genetic diseases” (Medical Genetics-Siena) - Telethon Genetic Biobank Network (Project No. GTB07001C).
## Table 1

Likely modifier regions. Genes with promoters in the top 10% of MeCP2 promoter hits for at least one replicate are shaded in blue.
### Table 2 Unlikely modifier regions. Genes with promoters in the top 10% of MeCP2 promoter hits for at least one replicate are shaded in blue.

<table>
<thead>
<tr>
<th>Polymorphic CNVs</th>
<th>Breakpoints (Bp)</th>
<th>Gene content</th>
<th>MeCP2_B promoter hits rank</th>
<th>MeCP2_C promoter hits rank</th>
<th>$%$ C/ 89 Z</th>
<th>139 Z</th>
<th>565 C/ 601 Z</th>
<th>185 C/ 119 Z</th>
<th>421 C/ 109 Z</th>
<th>402 C/ 368 Z</th>
</tr>
</thead>
<tbody>
<tr>
<td>1q44 (58kb)</td>
<td>246,794,522-246,832,126</td>
<td>OR2T34 22567 12444</td>
<td><strong>Del C</strong></td>
<td>Del C</td>
<td>Del Z</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>OR2T10 22348 8566</td>
<td>Del Z</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>*IGKV2D - -</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>*IGKV1D - -</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>*IGKV3D - -</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2p11 (494kb)</td>
<td>89,401,838-89,895,566</td>
<td>-</td>
<td>-</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>IGKV2D - -</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>IGKV1D - -</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>IGKV3D - -</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3q26 (134kb)</td>
<td>163,997,228-164,101,776</td>
<td>-</td>
<td>-</td>
<td>Del C</td>
<td>Del Z</td>
<td>Del C/ Del Z</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3q29 (36kb)</td>
<td>196,905,767-196,942,158</td>
<td>MUC20 - -</td>
<td></td>
<td>Dup C</td>
<td>Dup Z</td>
<td>Del C</td>
<td>Del C</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4q13.2 (108kb)</td>
<td>69,037,735-69,165,814</td>
<td>UGT2B17 21719 18336</td>
<td>Del C</td>
<td>Del Z</td>
<td>Del C</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6p21.32 (65kb)</td>
<td>29,939,288-30004,636</td>
<td>HCG4P6 - -</td>
<td>Del C</td>
<td>Del Z</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6p21.33 (77kb)</td>
<td>32,595,402-32,672,983</td>
<td>HLA-DRB5 - -</td>
<td></td>
<td>Dup C</td>
<td><strong>Amp C</strong></td>
<td>Amp C/ Amp Z</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8p11.23 (143kb)</td>
<td>39,336,595-39,499,752</td>
<td>ADAM5P 14070 2916</td>
<td>Del C</td>
<td>Del C</td>
<td>Amp Z</td>
<td>Amp C/ Amp Z</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10q11.22 (84kb)</td>
<td>47,017,598-47,161,232</td>
<td>- - -</td>
<td>Del C</td>
<td>Del Z</td>
<td>Del Z</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>14q11 (860kb)</td>
<td>18,624,383-19,484,013</td>
<td>- - -</td>
<td>Del C</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>OR11H13P - -</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>OR4Q3 23539</td>
<td>100</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>OR4M1 24054</td>
<td>4686</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>OR4N2 23383</td>
<td>7030</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>OR4K2 21957</td>
<td>3567</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>OR4K5 21814</td>
<td>7944</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>15q11.2 (727)</td>
<td>18,810,004-19,537,035</td>
<td>- -</td>
<td>Del C</td>
<td>Del C</td>
<td>Del Z</td>
<td>Del C</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>16p11.2 (220kb)</td>
<td>28,732,295-28,952,218</td>
<td>-</td>
<td>-</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>ATXN2L - -</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>TUFM 7097</td>
<td>3848</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>SH2B1 12675</td>
<td>12680</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>ATP2A1 15182</td>
<td>23366</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>RABEP2 10298</td>
<td>16794</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>CD19 16469</td>
<td>16504</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>17q21.31 (163kb)</td>
<td>41,544,234-41,706,870</td>
<td>KIAA1267 - -</td>
<td></td>
<td>Amp C/ Dup Z</td>
<td>Amp Z</td>
<td>Del Z</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>22q13.2 (43kb)</td>
<td>41,237,731-41,287,066</td>
<td>SERHL1 199</td>
<td>20317</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*16 isoforms
** smaller rearrangement in these patients

*Table 2* Unlikely modifier regions. Genes with promoters in the top 10% of MeCP2 promoter hits for at least one replicate are shaded in blue.
References


4.3 Human Copy Number Variations Database
Fig. 3  Homepage of Human Copy Number Variations Database
<table>
<thead>
<tr>
<th>Molecular karyotype</th>
<th>Inheritance</th>
<th>Proximal breakpoints</th>
<th>Distal breakpoints</th>
<th>Disease genes</th>
</tr>
</thead>
<tbody>
<tr>
<td>dup(1)(p22.3)(p22.3)</td>
<td>M</td>
<td>87 179 292</td>
<td>87 273 826</td>
<td></td>
</tr>
<tr>
<td>dup(1)(p12)(p12)</td>
<td>P</td>
<td>118 294 990</td>
<td>118 303 853</td>
<td></td>
</tr>
<tr>
<td>dup(1)(q43)(q43)</td>
<td>P</td>
<td>234 698 142</td>
<td>234 814 787</td>
<td>EDARADD</td>
</tr>
<tr>
<td>dup(2)(p16.3)(p16.3)</td>
<td>P</td>
<td>47 866 371</td>
<td>47 938 504</td>
<td>MSH6</td>
</tr>
<tr>
<td>dup(2)(q13q13)</td>
<td>P, M</td>
<td>110 199 004</td>
<td>110 337 631</td>
<td>NPHP1*</td>
</tr>
<tr>
<td>dup(2)(q14)(q14)</td>
<td>P</td>
<td>119 843 354</td>
<td>120 283 921</td>
<td></td>
</tr>
<tr>
<td>del(2)(q32q32.2)</td>
<td>M</td>
<td>189 573 876</td>
<td>189 633 678</td>
<td>COL3A1, COL5A2</td>
</tr>
<tr>
<td>del(2)(q36)(q36.3)</td>
<td>P</td>
<td>229 638 936</td>
<td>230 533 004</td>
<td></td>
</tr>
<tr>
<td>dup(3)(p22.1)(p22.1)</td>
<td>M</td>
<td>41 729 441</td>
<td>41 940 304</td>
<td></td>
</tr>
<tr>
<td>dup(4)(p16.2)(p16.3)</td>
<td>M</td>
<td>2 902 096</td>
<td>3 114 307</td>
<td>HTT</td>
</tr>
<tr>
<td>del(4)(q13)(q13.3)</td>
<td>P</td>
<td>74 105 031</td>
<td>74 245 908</td>
<td></td>
</tr>
<tr>
<td>dup(4)(q13.3)(q13.3)</td>
<td>P</td>
<td>75 458 705</td>
<td>76 190 527</td>
<td></td>
</tr>
<tr>
<td>dup(4)(q32)(q32.2)</td>
<td>P</td>
<td>162 406 396</td>
<td>162 900 008</td>
<td></td>
</tr>
<tr>
<td>del(4)(q34.1)(q34.1)</td>
<td>M</td>
<td>174 912 401</td>
<td>175 910 342</td>
<td>HPGD</td>
</tr>
<tr>
<td>del(4)(q35)(q35.2)</td>
<td>P</td>
<td>186 466 884</td>
<td>187 446 875</td>
<td>TLR3, CYP4V2, KLKB1, F11</td>
</tr>
<tr>
<td>del(4)(q35.2)(q35.2)</td>
<td>P</td>
<td>189 846 235</td>
<td>191 133 668</td>
<td></td>
</tr>
<tr>
<td>dup(5)(p13.2)(p13.2)</td>
<td>M</td>
<td>37 327 849</td>
<td>37 552 419</td>
<td></td>
</tr>
<tr>
<td>dup(5)(q21.1)(q21.1)</td>
<td>M</td>
<td>100 219 716</td>
<td>100 401 892</td>
<td></td>
</tr>
<tr>
<td>del(6)(p24.3)(p24.3)</td>
<td>P</td>
<td>9 953 753</td>
<td>10 085 727</td>
<td></td>
</tr>
<tr>
<td>del(7)(q31.3)(q31.3)</td>
<td>M</td>
<td>124 394 469</td>
<td>124 482 617</td>
<td></td>
</tr>
<tr>
<td>dup(8)(q12.1)(q12.1)</td>
<td>P</td>
<td>56 590 647</td>
<td>57 085 155</td>
<td></td>
</tr>
<tr>
<td>del(9)(p24.1)(p24.1)</td>
<td>M</td>
<td>6 318 511</td>
<td>6 321 081</td>
<td></td>
</tr>
<tr>
<td>dup(9)(p23)(q23)</td>
<td>M</td>
<td>11 447 340</td>
<td>11 687 635</td>
<td></td>
</tr>
<tr>
<td>dup(9)(q22.2)(q22.2)</td>
<td>M</td>
<td>97 470 882</td>
<td>98 104 136</td>
<td>HSD17B3</td>
</tr>
<tr>
<td>dup(10)(p21.3)(p21.3)</td>
<td>M</td>
<td>69 661 546</td>
<td>70 076 154</td>
<td></td>
</tr>
<tr>
<td>dup(11)(p11.2)(p11.2)</td>
<td>M</td>
<td>44 108 216</td>
<td>44 184 944</td>
<td>EXT2**</td>
</tr>
<tr>
<td>del(11)(q22.3)(q22.3)</td>
<td>M</td>
<td>107 761 839</td>
<td>107 807 671</td>
<td></td>
</tr>
<tr>
<td>dup(12)(p12.1)(p12.1)</td>
<td>S</td>
<td>22 369 323</td>
<td>22 586 257</td>
<td></td>
</tr>
<tr>
<td>dup(12)(p11.2)(p11.23)</td>
<td>M</td>
<td>27 270 178</td>
<td>27 659 659</td>
<td></td>
</tr>
<tr>
<td>dup(12)(p11.22)(p11.22)</td>
<td>M</td>
<td>29 489 556</td>
<td>29 844 564</td>
<td></td>
</tr>
<tr>
<td>dup(12)(q21.3)(q21.31)</td>
<td>M</td>
<td>81 274 846</td>
<td>81 814 232</td>
<td></td>
</tr>
<tr>
<td>dup(12)(q21.3)(q21.32)</td>
<td>M</td>
<td>84 474 053</td>
<td>86 008 465</td>
<td></td>
</tr>
<tr>
<td>dup(12)(q24.1)(q24.1)</td>
<td>M</td>
<td>108 061 580</td>
<td>108 123 789</td>
<td></td>
</tr>
<tr>
<td>dup(13)(q13.2)(q13.3)</td>
<td>P</td>
<td>34 517 499</td>
<td>35 022 728</td>
<td></td>
</tr>
<tr>
<td>dup(13)(q21.3)(q21.3)</td>
<td>M</td>
<td>72 153 664</td>
<td>72 228 260</td>
<td></td>
</tr>
</tbody>
</table>
Table 3a. Autosomal inherited private rearrangements found in 330 patients. The base pair positions are according to UCSC Genome Browser, http://genome.ucsc.edu, on Human May 2006 Assembly.

<table>
<thead>
<tr>
<th>Molecular karyotype</th>
<th>Patient sex</th>
<th>Inheritance</th>
<th>Proximal breakpoints</th>
<th>Distal breakpoints</th>
</tr>
</thead>
<tbody>
<tr>
<td>dup(X)(p22.31p22.31){0.8 Mb}</td>
<td>F</td>
<td>P</td>
<td>8 458 107</td>
<td>9 042 705</td>
</tr>
<tr>
<td>dup(X)(p11.4p11.4){0.2 Mb}</td>
<td>F</td>
<td>P</td>
<td>38 376 483</td>
<td>38 432 836</td>
</tr>
<tr>
<td>dup(X)(q22.3q22.3){0.43 Mb}</td>
<td>M</td>
<td>M</td>
<td>106 238 368</td>
<td>106 515 657</td>
</tr>
<tr>
<td>del(X)(q25q25){3.8 Mb}</td>
<td>M</td>
<td>M</td>
<td>124 267 011</td>
<td>127 627 617</td>
</tr>
<tr>
<td>dup(X)(q26.2q26.2){0.6 Mb}</td>
<td>F</td>
<td>P</td>
<td>130 526 525</td>
<td>130 788 298</td>
</tr>
<tr>
<td>dup(X)(q27.2q27.2){0.21 Mb}</td>
<td>M</td>
<td>M</td>
<td>140 617 007</td>
<td>140 824 390</td>
</tr>
<tr>
<td>dup(X)(q28q28){0.4 Mb}</td>
<td>F</td>
<td>P</td>
<td>148 880 115</td>
<td>149 137 402</td>
</tr>
</tbody>
</table>

Table 3b. Inherited private rearrangements on the X chromosome found in 330 patients.
<table>
<thead>
<tr>
<th>Molecular karyotype</th>
<th>Disease genes</th>
<th>number</th>
<th>Frequency %</th>
</tr>
</thead>
<tbody>
<tr>
<td>dup(2)(q13q13){1.33 Mb}</td>
<td>NPHP1</td>
<td>3</td>
<td>2.14%</td>
</tr>
<tr>
<td>del(2)(q32.2q32.2){0.13 Mb}</td>
<td>COL3A1, COL5A2</td>
<td>1</td>
<td>0.71%</td>
</tr>
<tr>
<td>dup(3)(p22.1p22.1){0.5 Mb}</td>
<td></td>
<td>1</td>
<td>0.71%</td>
</tr>
<tr>
<td>dup(4)(p16.1p16.1){0.06 Mb}</td>
<td></td>
<td>2</td>
<td>1.42%</td>
</tr>
<tr>
<td>dup(9)(p23p23){0.60 Mb}</td>
<td></td>
<td>1</td>
<td>0.71%</td>
</tr>
<tr>
<td>dup(13)(q13.2q13.3){0.8 Mb}</td>
<td></td>
<td>2</td>
<td>1.42%</td>
</tr>
</tbody>
</table>

**Table 4a.** Autosomal private rearrangements found in 140 controls.

<table>
<thead>
<tr>
<th>Molecular karyotype</th>
<th>Disease genes</th>
<th>number</th>
<th>Frequency %</th>
</tr>
</thead>
<tbody>
<tr>
<td>dup(X)(p22.31p22.31){0.8 Mb}</td>
<td>KAL1</td>
<td>1</td>
<td>0.71%</td>
</tr>
</tbody>
</table>

**Table 4b.** Inherited private rearrangements on the X chromosome found in 140 controls.
5. DISCUSSION, CONCLUSIONS and FUTURE PERSPECTIVES
5.1 DISCUSSION

Microarray-based comparative genomic hybridization has revolutionized clinical cytogenetic, as it provides a relatively quick method to scan the genome for gains and losses of chromosomal material with significantly higher resolution and greater clinical yield than was previously possible. Numerous different aCGH platforms have emerged and have been used successfully in the diagnostic setting. In the past few years, these new methodologies have led to the identification of novel genomic disorders in patients with developmental delay/intellectual disability and/or multiple congenital anomalies as well as to the discovery that each individual carries inherited copy number variations whose contributions to genetic variation, and complex disease are not well understood. Although aCGH is currently being used as an adjunct test for to standard karyotype analysis, it may to become the genetic test of choice, especially in cases of idiopathic ID/MCA. [4]

As such y, in this work this innovative technique was employed to study a group of MCA/ID patients. In all cases where the aCGH analysis indicated that the rearrangements arisen “de novo”, we performed standard karyotype analysis of both parents in order to exclude the presence of a balanced rearrangement favouring the unbalanced rearrangement in the child. In all cases but two the parents’ karyotype resulted normal and during genetic counselling a low recurrence risk was given to the family.

Among 696 patients, we found 426 (61%) negative while in 88 patients (13%) the analysis is still ongoing. We identified an inherited or de novo rearrangement in 165 cases (24%) while in 17 cases (2%) we detected novel de novo deletions not reported in the literature (Fig. 4).
Among 165 positive cases 10% were novel de novo rearrangements, 15% were rearrangements of known syndromes, 40% were non polymorphic rearrangements inherited from mother, 35% were non polymorphic rearrangements inherited from father.

The strategy for the identification of the candidate gene responsible for the clinical features is to analyze the genic content of rearrangements. In the patient with a de novo 3 Mb interstitial deletion of chromosome 14q12 this approach allowed us the identification of a new gene: FOXG1B. Later FOXG1B was demonstrated to be a gene responsible for the most severe form of RTT, the congenital variant. [97] The patient showed dysmorphic features and a Rett-like clinical course, including normal perinatal period, postnatal microcephaly, seizures, and severe intellectual disability. The deleted region was gene poor, indeed it contained only five genes. Among them, FOXG1 turned out to be a very interesting gene because it encodes a brain-specific transcriptional repressor.
A first analysis of this gene with a combination of both DHPLC and real-time quantitative PCR in a cohort of 53 MECP2/CDKL5 mutation-negative RTT patients, allowed the characterization of a different de novo FOXG1 truncating mutation in two RTT congenital variant patients. Soon after, FOXG1 mutations have been identified in several patients, independently classified as congenital Rett variants from France, Spain, Latvia and Canada. [98]

The search of additional patients having overlapping 14q12 deletions in the existing databases, (ECARUCA and DECIPHER) which collect cases with chromosomal aberrations, allowed the identification, in the DECIPHER database, of two new patients and the further characterization of the 14q12 microdeletion syndrome. [99]

Nevertheless, when attempting to carry out a genotype-phenotype correlation after the identification of a certain submicroscopic anomaly, the classic approach of disease mapping based on the sole genic content may not be sufficient.

Autism spectrum disorders are a group of common neurodevelopmental conditions characterised by impairments in communication, social interaction, and behaviour. At least one of every 166 children is likely to be affected by this spectrum of disorders, with a male: female (M:F) ratio of 4:1, and it is the most heritable of all complex neuropsychiatric conditions. Since ASDs are considered complex genetic disorders, resulting from the interaction of several genes and environmental factors, the lumping together of all cases of ASD, with no subgrouping based on phenotypic characteristics, makes the identification of contributory genes extremely difficult. The fact that autism is known to be associated with several distinct medical/genetic disorders further highlights its genetic heterogeneity. A variety of different approaches to identify genes for ASDs have been undertaken, including cytogenetic assessment for chromosome
abnormalities, genome scan linkage studies, and association studies. The overlap of results among different study methods is limited, largely attributed to the significant clinical and genetic heterogeneity of ASDs based on variable behavioural indices, variations among study populations, and limitation of methods for detecting ASD susceptibility genes of mild to moderate effect. [100]

We selected 95 unrelated patients with ASD. In addition to the well known recurrent rearrangements involving the 15q11q13, 16p13 and 22q13 regions, recurrent microdeletions and microduplications at 16p11.2 have been recently identified and have been shown to confer susceptibility to ASDs in up to 1% of autistic patients. In fifteen patients (16%) CNVs were considered as disease associated, while in 20 patients (21%) CNVs were classified as uncertain cofactor. Our findings have important clinical and research implications. These results emphasise that aCGH analysis is able to detect genetic cause in ASD patients, stimulating the need for further investigation of CNVs in ASD. These findings underscore a challenge in genetic counselling in interpreting aCGH data and give practical advices about the management of such data and the communication process to the family.

The presence of a CNVs in a coding region usually correlates with changes in the abundance of corresponding transcripts. Absence or excess of the protein product of a dosage sensitive gene may influence cell differentiation or migration and tissue formation early during development. Excess of a protein produced as a result of gene amplification may also lead to protein misfolding in an age-dependent manner and overload intracellular pathways of protein transport, proteosome degradation, or recycling commonly seen in late onset neurodegenerative disorders. Intracellular aggregation of proteins and inclusions
may affect protein-protein interactions, enhance free radical formation, cause mitochondrial damage and dysfunction, and initiate apoptosis.

In addition, genomic rearrangements may also be associated with molecular mechanisms other than affecting transcript levels to influence gene dosage and expression. Such complex mechanisms include gene interruption, gene fusion, unmasking a recessive allele or silenced gene, and interruption of regulatory gene-gene and chromosomal interactions. The investigations of the roles of CNVs during evolution and their biologic significance in health and diseases are rapidly progressing. Even before the completion of the Human Genome Project, the pathogenic significance of gene dosage was realized in several disorders of the central and peripheral nervous system. The highlighted examples demonstrate how gene dosage effects may influence cell function and development of common disorders often characterized by heterogeneous genetic etiology. [101] To test the hypothesis that CNVs can modulate the phenotypic expression of RTT syndrome we analyzed by aCGH two couples of RTT sisters and four additional couples of unrelated RTT girls matched by mutation type showing discordant phenotype (classic and Z-RTT). Results of CNVs analysis of phenotypically discordant RTT pairs are shown in Tables 1 and 2. Overall, we indentified 28 CNVs, 27 of them being known polymorphic regions and one being an apparently private rearrangement duplicated in only one Z-TT patient. This private CNV localized to 3q13.12, extended for about 99 Kb and contained three genes expressed in retina (GUCA1C), testis (MORC1) and one putative uncharacterized protein C3orf66.

Some polymorphic CNVs were found in more than one pair of phenotypically discordant RTT pairs. Among the 29 CNVs we considered 14 of them as “unlikely modifier regions” since results were apparently not associated with phenotypic severity (Table 2). For instance, the CNV was observed in both a classic and a Z-RTT member of different pairs. These include regions containing
olfactory receptors and class II HLA molecules that are not expected to directly correlate with the phenotypic variability related to classic/Z-RTT phenotype. The remaining 15 CNVs were considered potential modifiers (Table 1). For 3 of them results were consistent in two couples. Analyses on phenotypically discordant RTT pairs resulted in 28 CNV's that gave a list of 55 genes which could potentially modify RTT phenotype. A Perl script was used to search for these 55 genes within two experimentally replicated promoter hit files that ranked MeCP2 binding based on ChIP-chip analysis. Aliases of genes were acquired from genome.ucsc.edu and genelist.org. However, twenty genes were not found on the promoter hit list primarily because most these genes were discovered after the array was designed. The ranking from total number of genes from 1 to 24,134 is shown for two replicate MeCP2-ChIP microarrays (MeCP2 B and MeCP2 C promoter hits rank, Tables 1 and 2). Genes with promoters in the top 10% of MeCP2 promoter hits for at least one replicate are shaded in blue.

In conclusion, our study failed to identify one major common modifier gene/region suggesting that genetic modifiers may be complex and different case by case. However, we identified several genes within loci that may have a role in modulating the diseases in single cases. Environmental and epigenetic differences may further modify RTT syndrome phenotype and this study suggests that investigating genetic modifiers of the RTT phenotype by CNV analyses is by nature complex because MeCP2 pathways and gene targets are influenced by both genetic and epigenetic variables.

Microarray-based comparative genomic hybridisation enables the accurate detection of submicroscopic CNVs and has been increasingly used to investigate patients with intellectual disability or congenital abnormalities in some clinical settings. In aCGH analysis the CNVs have been identified using the aberration
algorithms Z-Score and ADM-1 provided by the CGH-Analytics software. The presence/absence of CNVs depends both on the experimental procedure and the aberration algorithm used to this purpose. The intrinsic variability of these procedures does inevitably lead to false negative and false positive calls. In order to reduce the false positive results, we disregarded all the calls containing less than three contiguous probes for experiments of excellent quality. We adopted some criteria to discern between benign and pathogenic CNVs. Firstly, we considered benign, and thus disregarded, all the CNVs reported in the last release of DGVs. [102] Since this catalogue contains many benign CNVs detected by BAC arrays, which in general are overestimated in size, [68] we compared our CNVs only with those reported in the database detected by oligo/SNP arrays. Then, we considered as clinically relevant all the CNVs arisen de novo, the imbalances known to be pathogenic and the CNVs encompassing genes whose up/downregulation is known to cause a disease.

Until August 2009 we selected of a cohort of 330 patients with MCA/ID in which we identified 61 CNVs, ranging from 0.06 to 2.9 Mb, that were not described in the DGVs (Tab. 3a and 3b). Among these 14 of which are regions that include disease genes, 44 are gains and 17 are losses. In all cases the same rearrangement was inherited from a healthy parent. Fifty nine of the rearrangements were private while 2 were found in more patients. In particular a del17q12 was found in 3 patients and a del9p24.1 was found in 2 patients. We collected 140 subjects from the first year of Medicine and Surgery faculty at University of Siena and from IRCCS OASI Maria SS. Troina and Biologia Generale e Genetica Medica, University of Pavia to detect the frequency of selected CNVs in normal population. In the 140 control subjects analyzed we identified only 7 out of 61 rearrangements previously described (11.4%) (Tab. 4a and 4b). In particular a dup2q13 of 1.33 Mb was identified in 2.14% of control subjects; a dup4p16.1 of 0.06 Mb and a dup13q13 of 0.8 Mb were identified in
1.4% of control subjects; a del2q32.2 of 0.3 Mb, a dup3p22.1 of 0.5 Mb and a dup9p23 of 0.6 Mb were identified in 0.7% of control subjects. Moreover, none of the 2 rearrangements identified in more than one patient (del17q12q12 and del9p24.1) was detected until known in the normal population. Each of the 7 rearrangements found in the control population has the same extension of those previously detected in the 330 patients. By analyzing 140 control subjects we found that 13.6% of the duplications and 6.2% of the deletions detected were recurrent in the control population. The dup2q13 seems to be the most frequent rearrangement detected so far. Three of the 7 recurrent rearrangements are regions that include known disease genes. One of these 3 rearrangements is included in region known as a susceptibility locus in autism (dupXp22.31).

The project of “Genomic Structural Variation Studies in Mentally Retarded and Normal Individuals in Italy” was born from collaboration with other two genetic units to characterize Italian CNVs not reported in DGVs. Currently, after 2 years, the total of Italian genetic units that decided to share aCGH data increased to 19. (Fig.3) In HCNV database about 14,000 CNVs are collected, so far.

The HCNV database resulted to be an helpful tool. In fact increasing the CNVs detected by aCGH, the clinician will have more data to explain not polymorphic but inherited or de novo rearrangements overlapping with other patients of database. Our knowledge is still limited, you can not accurately predict the clinical syndrome based on genetic backgrounds, you can only predict a possible variation.
5.2 CONCLUSIONS AND FUTURE PERSPECTIVES

aCGH can detect cytogenetic anomalies at a resolution unachievable by conventional techniques. Imbalances as small as the size of the used clones will identify new syndromes and will elucidate the molecular basis of clinically recognized syndromes. Given the great achievements in the past few years, aCGH will change the diagnostic approach to many congenital and acquired genetic diseases such as intellectual disability, birth defects, and cancer. However, diagnostic guidelines should be established though in order to enable the reliable and accurate detection of chromosome imbalances by aCGH in the cytogenetic laboratory. [103]

Our study allowed the characterization of several chromosomal imbalances in patients with complex phenotype, confirming the power of the aCGH method to clarify the molecular basis of these difficult cases. Through the employment of this innovative approach, several families finally received a definitive diagnosis and a correct recurrence risk.

For the future we plan to continue the consultation of the literature to find new disease candidate genes and new emerging low penetrance syndromes. We plan to reconsider the rearrangements found in our cohort and to provide to our patients a correct clinical follow up. We will continue to re-analyze our cohort paying attention to the CNVs regions and taking into consideration the possibility of modifier genes.
6. REFERENCES
REFERENCES


64. Walsh, T., et al., Rare structural variants disrupt multiple genes in neurodevelopmental pathways in schizophrenia. Science, 2008. 320(5875): p. 539-43.
Curriculum Vitae

Personal data
Name: Filomena Tiziana
Surname: Papa
Date of birth: Feb 02, 1977
Place of birth: Foggia
Nationality: Italian

EDUCATION AND TRAINING

1996: School-leaving certificate
Scientific High School: “Guglielmo Marconi” of Foggia.

2005: Degree in Biological Sciences
in the field of Molecular Biology, achieved on November 15 with marks 106/110. Degree thesis in Molecular biology about "proteomic analysis of oocytes of Danio Rerio”
Supervisor: Prof. Luca Bini

2005: Professional Qualification for Biologist

2006: Winner of four year PhD in “Oncology and Genetics” - October 1th, 2006

2006-2010: Scholarship for a 4-year Doctoral in Oncology and Genetics
– Medical Genetics Course (coordinator prof. A. Renieri), University of Siena.

From April 2010: International Exchange Scholar – Department of Neurlogy, University of Kentucky, Lexington-USA

TEACHING RESPONSIBILITIES

Component of examen committee for:
Integrated course in Biology 1st year Degree course in Medical School from 2007-2009
Integrated course in Genetics 2nd year Degree course in Midwifery from 2007-2009

Lessons

Tutorial teaching activities in Medical Genetics:
Genetics 2nd year Degree course in Medical School from 2007-2009
A 2.6 Mb deletion of 6q24.3-25.1 in a patient with growth failure, cardiac septal defect, thin upperlip and asymmetric dysmorphic ears. 

Clinical and molecular characterization of Italian patients affected by Cohen syndrome.

Expanding the phenotype of 22q11 deletion syndrome: the MURCS association.
Clin Dysmorphol 2008, Jan;17(1):13-17


A 3 Mb deletion in 14q12 causes severe mental retardation, mild facial dysmorphisms and Rett-like features.

Private inherited microdeletion/microduplications: implications in clinical practice.

Array comparative genomic hybridization in retinoma and retinoblastoma tissues.

14q12 Microdeletion syndrome and congenital variant of Rett syndrome.

A 9.3 Mb microdeletion of 3q27.3q29 associated with psychomotor and growth delay, tricuspid valve dysplasia and bifid thumb.
Is HSD17B1 a new sex reversal gene in human?
Katzaki E, Papa FT, Mucciolo M, Uliana V, Renieri A.

3.2 Mb microdeletion in chromosome 7 bands q22.2-q22.3 associated with overgrowth and delayed bone age.

Syndromic mental retardation with thrombocytopenia due to 21q22.11q22.12 deletion: Report of three patients.

Association between primary open-angle glaucoma (POAG) and WDR36 sequence variance in Italian families affected by POAG.
Br J Ophthalmol. 2010 Sep 2. [Epub ahead of print]

Practical hints for management of array-CGH output in autism spectrum disorders: susceptibility regions versus uncertain cofactors.

BOOK CHAPTER
Collaboration for a book chapter:
Mental Retardation. Chapter 7: X-Linked Mental Retardation: a Diagnostic, Clinical and Molecular update.
Caselli R, Papa FT, Ariani F, Meloni I, Renieri A. Mariani Foundation Paediatric Neurology Editor. 2007