Identification and characterization of disease-related copy number variations (CNVs) by high-dense SNP oligonucleotide microarrays

Núria Rivera Brugués

Núria Rivera Brugués

Identification and characterization of disease-related copy number variations (CNVs) by high-dense SNP oligonucleotide microarrays

DOCTORAL THESIS
Barcelona, 2011

PhD Program on Genetics
Departament de Genètica
Facultat de Biologia
Universitat de Barcelona
Biennium 2004-2006

PhD thesis presented by

Núria Rivera Brugués

entitled

Identification and characterization of disease-related copy number variations (CNVs) by high-dense SNP oligonucleotide microarrays

(Identificació i caracterització de variants en número de còpia (CNVs) relacionades amb malalties mitjançant microarrays de SNP oligonucleòtids d'alta densitat)

To be awarded a

PhD by Universitat de Barcelona

PhD research conducted at the Institute of Human Genetics (IHG), Helmholtz Center Munich (Neuherberg, Germany), supervised by Thomas Meitinger

Director:	Tutor:	Author:
Dr. Thomas Meitinger	Dr. Lluïsa Vilageliu Arqués	Núria Rivera Brugués

ACKNOWLEDGMENTS

The work presented in this thesis was performed in the Institute of Human Genetics at the Helmholtz Center Munich, Germany. The financial support was provided by a 4 years PhD-scholarship from the Helmholtz-DAAD funding program.
I would like to express my gratitude to all that have contributed to my work, helped and supported me during my time working on this thesis.

First of all, I would like to thank all the patients, their parents, and the clinicians for the participation and collaboration in my work. Without them this thesis could not have been done.

I especially want to thank my supervisor Tim M Strom for giving me the opportunity to perform my thesis work and be a part of his research team. I would like to extend my heartfelt gratitude to Professor Thomas Meitinger for accepting to be the director of my thesis. It was an honor to work under them and gain from their vast knowledge and expertise.
I am most grateful to Dr. Lluïsa Vilageliu Arqués for accepting to be the tutor of my thesis at the University of Barcelona. Many thanks for being very supportive and interested in my work.

I am also grateful to my colleagues of the Institute of Human Genetics for all their guidance in the lab and beyond and for creating a good and inspiring working environment. Especially I want to express my gratitude to the staff of technicians, researchers and clinicians connected to the Tim Strom's group for their contribution of excellent skills, enthusiasm and for creating a very nice working atmosphere.

I acknowledge the efforts of all the reviewers who have taken the time to read this thesis.

Sincere thanks goes to my parents Manel and Mª Angels, my sister Eva, brother-in-law Gori and my parents-in-law Fritz and Elisabeth, for taking interest in my work and for being supportive. I also want to thank the rest of my family in Spain. Finally, I want to express my sincere thanks to my boyfriend, Josef, for his endless support and love, and for always believing in me. This thesis is dedicated to you.

Table of contents

i. LIST OF ABBREVIATIONS

AAIDD	American Association in Intellectual and Developmental Disabilities
aCGH	Array CGH
ACTH	Adrenocorticotropic hormone
ADP	Adenosine diphosphate
AGH	Array genomic hybridization
AIDS	Acquired immune deficiency syndrome
APA	American Psychiatric Association
ASD	Autism spectrum disorder / Atrial septal defect
ATP	Adenosine triphosphate
BAC	Bacterial artificial chromosome
BAF	B allele frequency
bp	Base pair
BSA	Bovine Serum Albumin
°C	Celsius grad
CBS	Circular Binary Segementation algorithm
cDNA	Complementary DNA
CGH	Comparative genomic hybridisation based method
CHI	Congenital hyperinsulinism
CMT1A	Charcot Marie-Tooth type1A
CNV	Copy number variant
CODIS	Combined DNA Index System
CS	Cohen syndrome
Ct	Threshold cycle
DD	Developmental delay
Del	Deleted
DECIPHER	Database of Chromosomal Imbalance and Phenitype in Humans using Ensembl Resources
DGV	Database of Genomic Variants
DGS	DiGeore syndrome
DHPLC	Denaturing high-performance liquid chromatography
dH_2O	Distilled water
DNA	Deoxyribonucleic acid
DNase	Deoxyribonuclease
dNTPs	Deoxyribonucleotide phosphate
DSB	Double-strand breaks
DSM-IV-TR	Diagnostic and Statistical Manual of Mental Disorders, fourth edition, text revision

dsDNA	Double strand DNA
DS	Down syndrome
DVD	Developmental verbal dyspraxia
EDTA	Ethylenediamine tetraacetic acid
e.g.	*Exempli gratia* (latin: for example)
EMBL	European Molecular Biology Laboratory
EST	Expressed sequence tag
ESTR	Expanded simple tandem repeats
et al.	*Et alteres* (latin: and others)
EtOH	Ethanol
FISH	Fluorescence in situ hybridisation
FoSTeS	Fork Stalling and Template Switching
FOX	Forkhead-box
GADA	Genome Alteration Detection algorithm
GLAD	Gain and loss of DNA algorithm
GSIS	Glucose-stimulated insulin secretion
H_2O	Water
HI	Hyperinsulinemia of infancy
HIV	Human immunodeficiency virus
HGP	Human Genome Project
HOS	Holt-Oram syndrome
HMM	Hidden Markov Model
HNPP	Neuropathy with liability to pressure palsies
HPLC	High-pressure liquid chromatography
ICD-10	International Statistical Classification of Diseases and Related Health Problems, 10th edition
ID	Intellectual disability
i.e.	*id est* (latin: that is)
IQ	Intelligence quotient
INDEL	Short insertion or deletion polymorphisms
ins	Inserted
K^+_{ATP}	ATP-sensitive potassium channel
kb	Kilobase pair
KORA	Cooperative Health Research in the Augsburg Region
l	Liter
L1	Long interspersed element-1
LCR	Low copy repeat

LINE	Long interspersed nucleotide elements
log	Logarithm
LOH	Loss of heterozygosity
LTR	Long terminal repeat
MAD	Median absolute deviation
MAF	Minor allele frequency
MAPH	Multiplex amplifiable probe hybridisation
Mb	Megabase pair
mbar	Milibar
MER	Medium reiteration frequency interspersed repetitive element
µg	Microgram
mg	Miligram
$MgCl_2$	Magnesium Chloride
min	Minute
MIR	Mammalian interspersed repetitive element
ml	Mililiter
µl	Microliter
MLPA	Multiple ligation probe-dependent amplification
MM	Mismatch
MMEJ	Microhomology mediated end joining
mM	Millimolar
mm	Milimeter
µM	Micromolar
MR	Mental retardation
mRNA	Messenger RNA
MRNET	German Mental Retardation Network
NaCl	Sodium chloride
NAHR	Non allelic homologous recombination
NaOH	Sodium Hydroxide
NBTree	Naïve Bayesian Tree
NCBI	National Centre for Biotechnology Information
NGS	Next generation sequencing
ng	Nanogram
NH_4Cl	Ammonium chloride
NHEJ	Nonhomologous end-joining
NLS	Nuclear localization signal
nm	Nanometer
nM	Nanomolar
NTC	No-template control
oaCGH	Oligonucleotide-based array CGH
omb	Optomotor-blind
OMIM	Online Mendelian Inheritance in Man

PAC	P1-derived Artificial Chromosomes
PCR	Polymerase chain reaction
PET	Positron emission tomography
PM	Perfect match
Pmol	Picomol
qPCR	Quantitative PCR
RNA	Ribonucleic acid
Rnase	Ribonuclease
rpm	Revolutions per minute
RQ	Relative quantity
RT	Room temperature
SBE	Single base extension
SCIMM	SNP conditional mixture modeling
SD	Standard deviation / Segmental duplication
SNP	Single nucleotide polymorphism
SNR	Signal-to-noise ratio
SNV	Single-nucleotide variants
STR	Short tandem repeat
SV	Structural variants
SVA	SINE/VNTR/Alu element
SW	Smith-Waterman algorithm
SZ	Schizophrenia
TAAD	Thoracic aortic aneurysms and dissections
Taq	*Thermus aquaticus*
TBX	T-box transcription factor
TBE	T-box binding elements / Tris-borate-EDTA-electrophoresis buffer
Tris-HCl	Tris(hydroxymethyl)aminomethane Hydrogen Cloride
U	Unit
UCSC	University of California, Santa Cruz
UMS	Ulnar Mammary syndrome
UPD	Uniparental disomy
UV	Ultraviolate
V	Voltage
VCF	Velocardiofacial syndrome
VDCC	Voltage-dependent Ca2+ channels
VNTR	Variable number of tandem repeats
VSD	Ventricular septal defect

WAIS	Wechsler Intelligence Scales for Adult
WGA	Whole genome amplification
WHO	World Health Organization
WISC	Wechsler Intelligence Scales for Children
w/v	Weight/Volume
XCI	X-chromosome inactivation
xg	Hypergravity
YAC	Yeast Artificial Chromosomes

Symbols of nucleic acids

A	Adenine
C	Cytosine
G	Guanine
T	Thymine
U	Uracil

Symbols of amino acids

A (Ala)	Alanine
B (Asx)	Asparagine or Aspartic acid
C (Cys)	Cysteine
D (Asp)	Aspartic acid
E (Glu)	Glutamic acid
F (Phe)	Phenylalanine
G (Gly)	Glycine
H (His)	Histidine
I (Ile)	Isoleucine
K (Lys)	Lysine
L (Leu)	Leucine
M (Met)	Methionine
N (Asn)	Asparagine
P (Pro)	Proline
Q (Gln)	Glutamine
R (Arg)	Arginine
S (Ser)	Serine
T (Thr)	Threonine
V (Val)	Valine
W (Trp)	Tryptophan
Y (Tyr)	Tyrosine
Z (Glx)	Glutamine or Glutamic acid

ii. LIST OF TABLES

iii. LIST OF FIGURES

1. INTRODUCTION

1.1 THE HUMAN GENOME

The first draft of the human genome sequence was released in 2001, a decade ago, in parallel by the Human Genome Project (HGP), International Human Genome Sequencing Consortium (Lander et al 2001) and the company Celera Genomics (Venter et al 2001). The HGP declared the sequence completed in 2004 (Lander et al 2004) when over 99% of the genome sequence had been successfully elucidated. The draft, which was generated in 12 years, was still highly imperfect with 341 gaps spanning over 228 Mb across the genome. Few years after, as a result of the development of the massively parallel DNA sequencing technology, it was possible to decipher the first complete individual human genome in only few months (Levy et al 2007). In the next year, other human genomes were sequenced, including the one of the pioneer James Watson (Wheeler et al 2008, Bentley et al 2008, Wang et al 2008). These studies used the previously generated reference sequence as a guide on which the new genomes were built, enabling them to be assembled much more quickly and cheaply than the initial Human Genome Project. This approach is called re-sequencing. In 2008 the 1000 Genomes Project was launched with the goal of sequencing the genome of 1000 individuals from different ethnic groups. Nowadays (2011), and after the completion of the pilot phase (phase I) the project is in its production phase (phase II) with a target of sequencing upwards of 2000 individuals. However, given that the application of massively parallel sequencing methods now is becoming routine it is projected that the number of sequenced genomes generated worldwide will exceed 30,000 by the end of 2011 (Alkan et al 2011).

Before the first draft sequence was published, there was a lot of speculation on how many protein coding genes the human genome could contain and early guesses estimated a range from over 35,000 to 125,000 genes (Liang et al 2000). With the genome sequence completed, it came up as a surprise that the human genome contains approximately 23,000 genes only, similar to the number of genes in mice and twice the number of roundworms (e.g. Caenorhabditis elegans) and fruit flies. About 1.1% to 1.4% of the genome's sequence codes for proteins and these genes are distributed unevenly across the chromosomes (Lander et al 2004). As a characteristic feature the human genome has significantly more segmental duplications (SD) (nearly identical, repeated sections of DNA) than other mammalian genomes. These DNA segments may underlie the creation of new primate-specific genes and less than 7% of the protein families are vertebrate

specific (Lander et al 2004). In view of such trans-species similarities, understanding the differences in gene regulation will probably be more relevant than differences in gene identities per se. Therefore, deciphering gene expression patterns may provide clues to understand how the difference in complexity can be generated as well as how diseases are caused.

The reference human genome sequences provide the basis for the study of human genetics. However, a fundamental step in order to systematically investigate human variation is the construction of genome maps which describe types of DNA differences, allele frequencies and the correlation patterns between nearby variants. In general, the genome sequence lists the order of every DNA base in a genome whereas the genome map identifies the landmarks. An example of a variation map is the HapMap that catalogs genetic similarities and differences in human beings, thus permitting to find genes that affect health, disease, and individual responses to medications and environmental factors (International HapMap Project 2003). Currently major efforts are undertaken to integrate the new variations that are being identified, thus expanding public resource of genome variants in global populations (The International HapMap 3 Consortium 2011, The 1000 Genomes Project Consortium 2011).

1.1.1 Genetic variations

Four major genetic variations have been identified to account for the major polymorphic differences among individual human genomes; single nucleotide polymorphisms (SNPs), short insertion or deletion polymorphisms (INDELs), tandem repeats (STRs and VNTRs) and copy number variations (CNVs).

With more than 33 millions catalogued in the largest public database dbSNP 132, single nucleotide polymorphisms (SNPs) are the most abundant type of genetic variation. With the advent of the next generation sequencing (NGS) this number is expected to increase although it has been reported that 95% of the individual sequence variations are already represented in dbSNP (The 1000 Genomes Project Consortium 2011). The 1000 genomes project has reported a collection of 15 millions SNPs.
Initially, the notion of a SNP was defined as a single nucleotide variation at a specific location in the genome that is found in more than 1% of the population (Brookes 1999). However, in practice, this definition is not applied anymore, and sometimes short insertion or deletion polymorphisms (also called

INDELs) and SNP variations with less than 1% allele frequency are also referred to as SNPs. Recently, the 1000 genomes project data has allowed a comprehensive survey of rare SNPs with <0.1-0.5%. MAF (minor allele frequency). Typically, SNPs are biallelic, although very rarely tri- or tetraallelic forms can be found. SNPs are not evenly distributed across the genome, and SNP frequencies can vary by several hundred-fold between two regions. SNPs appear in the individual human genome once per 0.1-1 kb interval or on average 1 per 300 bp (Venter et al 2001). Therefore, due to the high density and the relatively easy genotyping assay, SNPs are the ideal genomic markers to perform linkage and association studies. SNPs can occur anywhere on a genome, but SNPs occur much less frequently in coding regions of the genome than in noncoding regions (Lander et al 2004). The phenotypic consequence of a SNP is related to the location where it occurs, as well as the nature of the variant. Complex disorders have been associated to common single nucleotide polymorphisms and Mendelian traits to rare missense, nonsense or splice site changes. However, most SNPs probably do not have an effect and do not affect the ability to survive or adapt.

Indel is a collective abbreviation to describe relative gain or loss of a segment of one or more nucleotides in a genomic sequence. It allows the designation of a difference between genomes in situations where the direction of sequence change cannot be inferred: for example, when a reference or ancestral sequence has not been defined. It has typically been used to denote relatively small-scale variants (particularly those smaller than 1 kb). Indels that are not multiples of 3 are particularly uncommon in coding regions but relatively common in non-coding regions. Indels can be used as genetic markers in natural populations, especially in phylogenetic studies. As for the case of SNPs, Indels can underly pathogenecity when the implied deletions or duplications are located in coding regions. The 1000 Genomes Project has reported a total of 1 million indels. Per individual 192-280 frameshifting indels were found.

Another type of polymorphism is the tandem repeat, including short tandem repeats (STRs) (sometimes referred to as microsatellites) and variable number of tandem repeats (VNTRs) (also known as minisatellites). STR and VNTR consist of genomic regions of variable length of sequence motifs repeating in tandem. A tandem repeat polymorphism is identified when there is a differerence in the number of repeats between individuals. The repeat unit of STR consists of 1-6 base pairs of DNA, whereas the repeat unit of VNTR is 10-100 base pairs long. They can be repeated 3 to 100 times. By identifying repeats of a

specific sequence at different locations in the genome it is possible to create a genetic profile of an individual. Therefore, STR/VNTRs are genetic markers widely employed in forensic studies and DNA paternity testing, by using techniques such as DNA fingerprinting. One example is the Combined DNA Index System (CODIS, www.fbi.gov/hq/lab/codis) developed by the Federal Bureau of Investigation (FBI) in the U.S. that uses 13 STR markers to create forensic DNA profiles for identification of individuals. There are more than 947,600 published tandem repeat polymorphisms across the human genome (Ames et al 2008) and typically, they are neutral. Tandem repeats with very short repeat blocks may be unstable and vary from one tissue to another within an individual and from one generation to another. Expanded simple tandem repeats (ESTR) have been associated to disease (Sutherland et al 1998).

Finally, there are structural variants (SV), including deletions, duplications, inversions and translocations of large DNA fragments. Inversions and translocations change the spatial organization of DNA but, per se, do not result in a net gain or loss of sequence. In 2004, two studies revealed that healthy human genomes harbor extensive multi-kilobase copy number variations (CNVs) (Sebat et al 2004, Iafrate et al 2004). CNVs are defined as DNA segments that present at variable copy number in comparison to a reference genome with the usual copy number of N=2 (Feuk et al 2006). Somewhat arbitrarily, the term CNV usually refers to DNA sequences larger than 1000 bp (1 kb). The repository for polymorphic CNVs, the Database of Genomic Variants (DGV), regards those CNVs in the 50-1000bp range to be a separate category of indel. This artificial division within a continuous range has historical reasons due to the size limits of the methods used to detect CNVs. However, the recent application of massively parallel sequencing methods has complemented microarray-based methods and therefore the operational spectrum of CNVa has widened to include much smaller events (for example, those >50 bp in length) (Alkan et al 2011).
So far, over 66,000 CNVs of approximately 16,000 loci have been identified in the apparently healthy population (Database of Genomics Variants). The 1000 human genomes project has identified 20.000 structural variants. In terms of total bases involved, SVs account for more differences among individuals than do SNPs. About 12% of the human genome is CNV (Perry et al 2008) with CNVs contributing between 0.12% and 7.3% of the genomic variability seen within humans (Redon et al 2006, Zogopoulos et al 2007). CNV altering dosage-sensitive genes have been associated both with Mendelian disorders and with complex diseases (see section 2.1.2).

1.1.2 Genomic rearrangements associated with disease and susceptibility to disorders

It has long been known that aneuploidies may be devastative conditions associated with high morbidity and mortality. Down syndrome was the first condition that was shown to be caused by an eneuploidy, i.e. trisomy of chromosome 21 (Lejeune et al 1959). As a consequence of the development of both chromosome banding techniques in the 70s (Seabright 1971) and fluorescent in situ hybridization (FISH) techniques in the 80s (Cremer et al 1988), microscopically visible and subvisible structural rearrangements were also found to be associated with recognizable syndromes and to be present in some individuals with intellectual disability (Jacobs et al 1978, Schmickel 1986).

Independently, standard molecular techniques such as Southern blotting and restriction-endonuclease assays have been used to identify subtle chromosomal imbalances. A submicroscopic deletion encompassing an alpha-globin gene was identified as a cause of alpha-thalassaemia by restriction-endonuclease assay (Higgs et al 1979). Several studies took use of Southern blot hybridization in order to search for pathogenic structural rearrangements of genomic candidate regions. Nathans et al (1986) reported that the red-green color blindness is caused by deletions and duplications that disrupt or eliminate the red and green photopigment genes. Few years later, two parallel studies identified the main cause of severe Haemophilia A, a recurrent invertion that disrupts the factor VIII gene (Naylor et al 1993, Lakich et al 1993). Also deletions encompassing this gene were reported (Bogdanova et al 2002).

The development of genome profiling technologies such as array-based platforms and next generation sequencing has facilitated the analysis of the human genome at a large scale, thus permitting the identification and characterization of CNVs underlying the pathogenesis of a broad range of conditions. Meanwhile, the list of CNVs related to Mendelian and complex diseases has become very long already. Some examples of clinical phenotypes caused by CNVs involving single or multiple genes are shown in table 1-1. Therefore, several databases cataloging chromosomal abnormalities and phenotype correlations have been established with DECIPHER (Database of Chromosomal Imbalance and Phenotype in Humans using Ensembl Resources) being one of the most prominent. DECIPHER is provided by the Wellcome Trust Sanger Institute. It links submicroscopic chromosomal imbalances to clinical phenotype descriptions. As of July 13, 2011, this database contained information on 10,091 individuals with 59 defined genetic syndromes.

Deletion and duplication rearrangements can cause a phenotype by several molecular mechanisms. CNVs may vary the copy number of a gene sensitive to a dosage effect. Gene dosage describes the number of copies of a gene in a cell, and the level of gene expression may be influenced by higher and lower gene dosages. The dosage-sensitive gene *PMP22* gene is a well known example with the associated genomic disorders Charcot Marie-Tooth type1A (CMT1A) caused by the duplication and neuropathy with liability to pressure palsies (HNPP) caused by the deletion of a 1.4-Mb genomic fragment at 17p12. Furthermore, deletions and duplications may lead to a dosage-sensitive effect by disrupting a gene coding sequence. Breakpoints of CNVs (or balanced rearrangements) located within a gene may cause a loss-of-function as in the cases mentioned above (Haemophilia A and red-color blindness). In addition, CNVs may unmask mutations or functional polymorphisms of the remaining allele and there cause recessive disorders. In this context, only the complete loss of the gene expression results in phenotypic consequences. As example, deletions exerting a negative effect by converging with other recessive point mutations have been identified in Sotos syndrome (Kurotaki et al 2003).

Besides gene expression effects caused by dosage change mechanisms should be taken into account changes in gene expression by positional effects, e.g., elimination of regulatory elements. These regulatory structures may be up to 1 Mb away from the gene under control, either upstream or downstream. The translocation detected in a patient with Peter's anomaly, a defect of the anterior chamber of the eye, is an example. This translocation disrupted the *HDAC9* gene at 7p21.11 but its reciprocal breakpoint mapped on chromosome 1 at a distance of about 500kb from *TGFB2*. Since tgfb2 null mice showed similar eye defects, the position effect on *TGFB2* was considered to be the underlying genetic cause rather than the disruption of *HDAC9* (David et al 2003).

Although some variants appear to have little effect being rather prevalent in the normal population they might contribute to a disease phenotype when combined with other genetic and environmental factors. Thus, deletions and duplications may function as predisposing or protective alleles. Since long it is known that the deletion of the alpha-globin gene (causing alpha-thalassaemia) protects against malaria (Flint et al 1986). Several regions containing genes of the immune system tend to be variable in copy number and some of their multiallelic copy number variants have been associated to disease. *FCGR3B* encodes a variant of the type III IgG receptor and occurs in variable number of copies. Low copy number has been associated with systemic autoimmune diseases such as systemic lupus erythematosus, polyangiitis and Wegener's granulomatosis (Fanciulli et al 2007). The unit of seven beta-defensin genes also has variable

copy number. High copy number has been associated with psoriasis (Hollox et al 2008) whereas low copy number predispose to Crohn's colitis (Fellermann et al 2006). Another example of gene with variable copy number is the chemokine gene *CCL3L1* which was shown to confer inter-population differences. Low *CCL3L1* copy number (below population average) is associated with enhanced HIV/AIDS susceptibility (Gonzalez et al 2005). However, diallelic copy number variants containing one or several genes may predispose to complex diseases as well. Investigations on complex neuropsychiatric disorders have provided multiple examples of such risk alleles (see section 1.3.1.2.4). In general, some rare large-scale CNVs have been found more frequently in the affected population group than in normal population, suggesting that they play a role in the pathogenesis. It has been hypothesized (Girirajan et al 2010) that a second risk allele (or more) pushes these rare CNVs toward a phenotypic manifestation. Recently, one study has been able to demonstrate that a particular CNV can predispose to different diseases depending on differently sensitized genomic backgrounds: Ermakova et al (2011) analyzed mice carrying a 0.8 Mb duplication or its reciprocal, a deletion on mouse chromosome 11, together with an ApoE $^{KO/+}$ allele. Five phenotypic conditions were tested (metabolic syndrome, hypersensitivity to immune dysfunction, cardiovascular function, behavioural traits, and cancer susceptibility) both in baseline state and in challenged state by using multiple sensitizers. The authors demonstrated that the genomic rearrangements alter the susceptibility to multiple disease phenotypes which are manifested when challenged by a second hit, either genetic or environmental. Recently, the recurrent 16p13.1 duplication, which is known to confer risk to a variety of neuropsychiatric disorders, has also been found to be associated with thoracic aortic aneurysms and dissections (TAAD) (Kuang et al 2011). TAAD patients with 16p13.1 duplications were more likely to harbor a second rare CNV ($p=0.012$) than patients without duplication. This suggests that 16p13.1 duplication predisposes to different complex disorders depending on the nature of a second rare CNV (and extensively, of the genetic background).

Table 1-1: **Examples of copy number variations (CNVs) and disorders that they cause**

Phenotype	OMIM	Locus	CNV
Mendelian (autosomal dominant)[a]			
Williams-Beuren syndrome	194050	7q11.23	del
7q11.23 duplication syndrome	609757	7q11.23	dup
Adult-onset leukodystrophy	169500	LMNB1	dup
Microduplication 22q11.2	608363	22q11.2	dup
DiGeorge syndrome	188400	22q11.2/TBX1	del
Miller-Dieker lissencephaly	247200	17p13.3/LIS1	del
Mental retardation	601545	17p13.3/LIS1	dup
Spinocerebellar ataxia type 20	608687	11q12	dup
Holt Oram syndrom	142900	12q24.21 /TBX5	del
Ulnar mammary syndrome	181450	12q24.21 /TBX3	del
HNPP	162500	17p12/PMP22	del
CMT1A	118220	17p12/PMP22	dup
Mendelian (autosomal recessive)			
Familial juvenile nephronophtisis	256100	2q13/NPHP1	del
Gaucher disease	230800	1q21/GBA	del
Pituitary dwarfism	262400	17q24/GH1	del
Spinal muscular atrophy	253300	5q13/SMN1	del
Beta-thalasemia	141900	11p15/beta-globin	del
Alpha-thalasemia	141760	16p13.3/HBA	del
Mendelian (X-linked)			
Hemophilia A	306700	F8	del/inv
Hunter syndrome	309900	IDS	del/inv
Ichthyosis	308100	STS	del
Mental retardation	300706	HUWE1	dup
Pelizaeus-Merzbacher disease	312080	PLP1	del/dup/tri
MR+SZ	300260	MECP2	dup
Red-green color blindness	303800	opsin genes	del
Complex traits			
Alzheimer disease	104300	APP	dup
Autism	612200	3q24	inherited homozygous del
	611913	16p11.2	del/dup
Crohn disease	266600	HBD-2	copy number loss
	612278	IRGM	del
HIV susceptibility	609423	CCL3L1	copy number loss
Mental retardation	612001	15q13.3	del
	610443	17q21.31	del
	300534	Xp11.22	dup
Pancreatitis	167800	PRSS1	tri
Psoriasis	177900	DEFB	copy number gain
Schizophrenia	612474	1q21.1	del
	181500	15q11.2	del
	612001	15q13.3	del
Parkinson disease	168600	SNCQ	dup/tri

Table modified from Zhang et al 2009
[a] also *de novo* sporadic

1.1.3 Mechanisms for CNV formation

Four major mechanisms, such as NAHR, NHEJ, FoSTeS and L1-mediated retrotransposition generate rearrangements in the human genome and probably account for the majority of CNVs (Korbel et al 2007, Kidd el al 2008).

1.1.3.1 Non-allelic homologous recombination (NAHR)

NAHR is caused by the alignment of and the subsequent crossover between two nonallelic (i.e, paralogous) DNA sequence repeats sharing high similarity to each other. Repeats on the same chromosome and with the same orientation mediate deletion and/or duplication, whereas inverted repeats mediate inversion of the genomic interval flanked by the repeat. NAHR between sequence repeats on different chromosomes can lead to chromosomal translocations.

Substrates for NAHR are LCRs (low copy repeats: also known as segmental duplications, SDs) of >10 kb in length with 95%-97% similarity. Different groups of LCRs are sometimes adjacent to each other, with some units in tandem and others in reverse orientation resulting in complex LCRs. Complex LCRs can themselves undergo CNV and specific structural variants may predispose susceptibility to chromosomal rearrangements. An illustrative example of this is the Williams-Beuren locus which is located in a highly dynamic region rich in segemental duplications. The hemizygous deletion of genes on the chromosome 7q11.23 arise from non-allelic homologous recombination (NAHR) between two specific blocks (B) of region-specific segmental duplications (SDs) that show 99.6% sequence identity over 105 kb (Bayés et al 2004). Heterozygosity for either a paracentric inversion (Osborne et al 2001) or a large CNV between proximal SDs (Cuscó et al 2008), appears to further predispose to chromosome misalignment in some progenitors of WBS patients. These findings indicate that large CNVs can act as susceptibility alleles for disease-associated genomic rearrangements in the progeny.

Retrotransposable L1 elements (such as Alus) or matching pseudogenes also can act as a substrate for NAHR. NAHR can occur in meiosis, leading to constitutional genomic rearrangements as well as in mitosis, resulting in mosaic populations of somatic cells. It is well appreciated that many cancers are related to somatic genomic rearrangements, some due to somatic NAHR (Darai-Ramqvist et al 2008).

The recognition that NAHR between two low-copy repeats is the cause of most of the known “microdeletion syndromes”, such as DiGeorge/Velocardiofacial-Syndrome, Williams-Beuren-Syndrome, Smith-Magenis-Syndrome or Hereditary Neuropathy with liability to Pressure Palsies (HNPP), led to

hypothesis that many *de novo* genomic disorders might have arisen from this mechanism. Consequently, predictions of regions prone to genomic instability were performed and tested with array-based techniques. This allowed the identification of an increasing list of novel genomic disorders, such as the microdeletion/microduplication syndromes 1q21.1, 15q13.3, 15q24, 16p13.11, 17q12 and 17q21.31 (Sharp et al 2006).

1.1.3.2 Nonhomologous end-joining (NHEJ)

NHEJ is utilized by human cells to repair DNA double-strand breaks (DSB) caused by endonuclease-induced breaks and damage by exogenous agents such as ionizing radiation or reactive oxygen species. NHEJ also plays a critical role in V(D)J recombination, the process by which B-cell and T-cell diversity is generated in the vertebrate immune system. Inappropriate NHEJ has been related to the formation of some translocations seen in leukemia patients (Nickoloff et al 2008). NHEJ is referred to as "non-homologous" because the break ends are directly ligated without the need for a homologous template, in contrast to homologous recombination which requires a homologous sequence to guide repair (Moore et al 1996). NHEJ typically utilizes short homologous DNA sequences called microhomologies to guide repair (or as mentioned do not need homology at all). These microhomologies are present in single-stranded overhangs on the end of double-strand breaks. NHEJ can rejoin DSB ends accurately or leading to small 1-4 bp deletions or insertions (Lieber 2008). Breakpoints of NHEJ-mediated rearrangements often fall within repetitive elements such as LTR, LINE, Alu, MIR and MER2. Of note, although NHEJ only requires microhomology it is assumed that specific genomic architecture may influence to this mechanism (Zhang et al 2009).

Microhomology-mediated end joining (MMEJ) is another pathway of DSB repair that either does not require homology or only needs very short microhomologies for repair. Two characteristic features that allow to distinguish MMEJ from NHEJ are; 1) MMEJ uses 5 to 25 bp long homologies to anneal two ends of DSB and 2) key proteins involved in NHEJ (Ku70/ Ku80) are not required for MMEJ (Hastings et al 2009). Nonhomologous end joining is a candidate recombination-based mechanism that may explain nonrecurrent rearrangements. Investigations of the structural variations in eight human genomes have shown that the predominant mechanism mediating the large structural variations of the human genome is indeed the non-allelic homologous recombination (Kidd et al 2008).

1.1.3.3 Fork Stalling and Template Switching (FoSTeS)

A recently proposed mechanism in the generation of complex CNVs is fork stalling and template switching (FoSTeS), a replication misstep (Lee et al 2007). FoSTeS may be considered a contributing mechanism for nonrecurrent rearrangements of all sizes and complexities. According to this model the active replication fork can stall and switch to another replication fork using complementary template microhomology to anneal and prime DNA replication. The involved forks may be adjacent on the same chromosome or in a close proximity in three-dimensional space. FoSTeS may cause the joining of different sequences from discrete genomic position that are far apart from each other. Upon annealing, the transferred strand primes its own template-driven extension at the transferred fork. Depending on the direction of the fork progression, the erroneously incorporated fragment can be in direct or inverted orientation. Furthermore, depending on whether the new fork is located upstream or downstream of the original fork, the template switching results in either a deletion or a duplication (Zhang et al 2009). This procedure of disengaging, invading, and synthesizing may occur multiple times in series, resulting in the observed complex rearrangemets. A recent CNV study showed sequence complexities and microhomologies in 22% CNV breakpoints, which were consistent with two or more FoSTeS events (Perry et al 2008). Similar replication based-models including “serial replication slippage” have been proposed for smaller complex rearrangements (Hastings et al 2009).

1.1.3.4 L1-mediated retrotransposition

Around 42% of the human genome is made up of retrotransposons. Of these, 17% are long interspersed elements of type 1 (L1) (Landers et al 2001, Zhang et al 2009).

L1 elements are the only active autonomous transposons in the human genome. Only those copies with the intact full length coding for a RNA-binding protein, the enzyme reverse transcriptase and for an endonuclease (e.g. RNase H) conserve the capability of retrotransposition. Therefore, L1 elements can amplify themselves (through an RNA intermediate) and integrate back into the genome. Retrotransposition can induce mutations by inserting near or within genes. However, in order to avoid deleterious effects on host or retrotransposon transpositions are possibly regulated both by retrotransposon- and host-encoded factors. It was found that specific LINE-1 retrotranspons in the human genome are actively transcribed and the associated LINE-1 RNAs are tightly bound to nucleosomes and are essential in the establishment of local chromatin environment (Chueh et al 2009).

One additional characteristic feature of L1s elements is that they can mobilize Alu elements, SVA elements and retrogenes, thus increasing the number of DNA segments that can integrate in another part of the genome. Therefore, L1-mediated retrotransposition may represent a contributing mechanism for the formation of insertions. This is in line with several studies that have attributed indels to retrotransposition. Kidd et al (2007) found that 15% of the structural variations events that they detected were mediated by retrotransposition and Xing et al (2009) found that about 10% of the indels >100 bp detected in the Venter's genome were associated with transposable DNA sequences, including L1, Alu and SVA.

1.2 DNA MICROARRAYS

Copy number profiling platforms based on a chip format have been developed over the last years enabling the detection of CNV of varying sizes, from a few kilobases in size to submicroscopic deletions or duplications that are up to megabases in size. These high-throughput techniques allow the detection of multiple CNVs in a single experiment by interrogating thousands or even millions of loci represented by DNA probes on a slide.

1.2.1 Array CGH

During the late 1990s, chromosome-based comparative genomic hybridization (CGH) was further developed into microarrays (Solinas-Toldo et al 1997, Pinkel et al 1998). Microarray-based comparative genomic hybridization (array CGH or aCGH) or array genomic hybridization (AGH)) is based on the same principle of competitive hybridization as chromosome-based CGH, but the normal targeted metaphase chromosome spreads are substituted by DNA sequences spotted onto a glass slide. Basically, two differentially fluorescence labeled DNAs (one from a patient and one from a reference control) are hybridized simultaneously onto an array and the relative number of copies are determined by comparing the capability of each sample to hybridize to the same target. The fluorescence ratio generated from control versus patient DNA for each probe represents the average copy-number ratio between the patient and the control DNA. A gain or loss of fluorescence signal intensity from the patient's DNA indicates a gain or loss of the patient's DNA copy numbers. Fluorescence dyes commonly used for DNA labeling include

Cy3, which has a fluorescence emission wavelength of 570 nm (corresponding to the green part of the light spectrum), and Cy5 with a fluorescence emission wavelength of 670 nm (corresponding to the red part of the light spectrum). Information of the genomic location of the targeted DNA fragments can be found in human genome browsers such as Ensembl or UCSC, and therefore the detected imbalances can be mapped and sized by plotting the deviated signal intensity ratio to each probe's position in the genome. Different types of DNA sequences have been used to produce microarrays, and they include large-insert clones (40–200 kb in size), small insert clones (1.5–4.5 kb), cDNA (0.5–2 kb), PCR products (100 bp–1.5 kb), and oligonucleotides (25–80 bp) (Carter 2007). Array CGH was initially developed using large-insert DNA from BAC (Bacterial Artificial Chromosomes), YAC (Yeast Artificial Chromosomes) or PAC (P1-derived Artificial Chromosomes) clones libraries. Initial aCGH studies used arrays with one BAC clone every 1 Mb across the human genome (comprising a total of 2400-3500 BACs clones) and demonstrated that CNVs occur both in normal population (Iafrate et al 2004) and in various disease populations (Vissers et al 2003, Shaw-Smith et al 2004, Schoumans et al 2005, Rosenberg et al 2006). Although the distance between probes used in 1 Mb BAC-based array permit to reach a resolution bellow the cytogenetic limit (i.e., below 5 Mb), the call of CNVs is limited to the genomic regions covered by the selected BAC clones. Therefore, the following efforts were made to construct arrays that truely covered the whole genome. Tiling-path BAC arrays, typically consisting of 432,000 overlapping BAC clones, permitted a 10-fold increase of resolution and resulted in a substantial increase of the number of relevant genomic imbalances detected (Ishkanian et al 2004, de Vries et al 2005). Yet, the maximal resolution that could be obtained with tiling BAC arrays was finite because of their large probe size (80-200kb). In order to overcome this problem and further increase the resolution, new CGH arrays were constructed using even larger numbers of features (spots) and shorter DNA sequences, such as cDNA clones (Pollack et al 1999), PCR-products (Dhami et al 2005) and oligonucleotides (Carvalho et al 2004). cDNA and PCR-product arrays have been abandoned because of the uneven distribution of genes and the high cost and laboriousness of probe generation, respectively. The problem of laboriousness of probe generation has been overcome by the use of automatically synthesized oligonucleotides. However, the method employed in array manufacturing still set an upper limit of 60,000 intact probes that could be spotted mechanically onto the microarray surface. Therefore, initial whole-genome aCGH oligonucleotide arrays could only reach a restricted resolution of one probe every 50 kb (Carter 2007). Several commercial microarray platforms were developed then which achieved higher probe density due to the synthesis of oligonucleotides directly onto the slide. This *in situ*

synthesis process printed oligonucleotide probes, base-by-base, from digital sequence files. Agilent Technologies, Inc (www.chem.agilent.com) uses ink-jet technology to apply the reagents necessary for oligonucleotide synthesis on a spot-by-spot basis. At present this technology can achieve 1 million of oligonucleotides (60mer) per array. NimbleGen Inc. (www.nimblegen.com) uses a programmable mirror array to synthesize an ultra-high density CGH array, containing 4.2 million target oliogonucleotides (50-75mer), directly on a glass surface using maskless photolithography. This oligonucleotide array achieves a median probe spacing of 284 bp across the genome and offers an overall resolution down to 1.4 kb. Agilent and NimbleGen chips are available having a smaller number of probes but multiple arrays on the chip, thus allowing for the analysis of multiple samples in parallel. Additionally, these platforms offer full customization of microarray content.

A disadvantage of oligonucleotide arrays, however, is the use of shorter DNA-sequences compared to the larger size of BAC DNA. This makes oligonucleotide arrays more vulnerable to unspecific hybridization, resulting in a higher level of background noise. Typically, the standard deviation (SD) of log2 ratios in an oligonucleotide array may reach 0.25, whereas the usual SD for a BAC clone–based array is in the range of 0.05 only (Carter 2007). Therefore, data from adjacent probes need to be averaged in order to make a confident CNV call. Most array CGH platforms require at least three (commonly five) consecutive oligonucleotides to reliably detect a single copy number variation, a fact that determines the overall resolution of the array. Moreover, when an array has a higher noise level more consecutive probes will be required to detect a CNV thus reducing the effective resolution of the array even further.

1.2.2 SNP genotyping arrays

High-throughput array technologies for SNP genotyping can also be used to identify CNVs. Typically these arrays contain short oligonucleotides (20–30 bases) and they are less well suited for identifying CNVs than microarrays with longer oligonucleotides (60-75 bases). However, intensity information of these arrays can be used to determine DNA copy numbers and SNPs in the same analysis. The dual-role of these arrays furthermore enables allelotyping, e.g., the detection of uniparental isodisomies (Raghavan et al 2005). Moreover SNP genotyping assays can be interrogated for both loss of heterozygosity (LOH) events. Therefore, these platforms do not only identify copy numbers of genomic segments by signal intensity but also are able to detect copy-neutral LOH events as well as uniparental

isodisomy, parental consanguinity, or non-paternity by allelotyping. As of July 2011, the available ready-made SNP arrays with the highest genome coverage are the Affymetrix array 6.0 (www.affymetrix.com) which contains 1.8 million genetic markers and the Illumina HumanOmni2.5-8 BeadChip (www.illumina.com) which contains 2.37 million genetic markers (to be superseded soon by an array with 5 million variants). These commercial vendors also provide assays for sample multiplexing and customization.

The construction of Affymetrix arrays is based on photolithographic synthesis on a silica substrate where light and light-sensitive masking agents are used to build the sequences nucleotide by nucleotide across the entire array. Illumina's technology is based on indexed beads which are covered with hundreds of thousands of copies of a specific oligonucleotide and that are randomly deposited into microwells of a substrate (fiber optic bundles or planar silica slides) with a uniform spacing of ~5.7 microns. Both the Affymetrix and the Illumina platforms are referred to as one-color technique because other than in arrayCGH only a single DNA sample is hybridized onto the array and therefore the intensity data provided for each probe or probe set indicate solely the relative level of hybridization of the tested sample. To quantify DNA abundance (or raw total copy number) the SNP array platforms essentially sum the fluorescence intensities from the two alleles investigated for a given SNP (averaging over the replicate probes and then summing). Typically, the raw intensity data (R value) are normalized and then compared with normalized values from another DNA sample (or a group of samples) that was (were) assayed in an independent hybridization(s), thus permitting to determine the relative number of DNA copies in a DNA sample (log R ratio value) at each locus.

The methods employed for genotyping SNPs can be divided into two main groups based on the two main principles used to discriminate SNP alleles; hybridization-based and enzyme-based techniques (reviewed by Kim et al 2007). The Affymetrix platform uses a hybridization reaction to discriminate SNP alleles whereas the Illumina platform uses an enzymatic reaction.

It is known that the strength of binding between two short complementary strands of DNA is altered by a single mismatched nucleotide. One nucleotide mismatch changes the denaturing temperature of the hybridized strands by 10 degrees (Wallace et al 1979). This change in stability allows sequence-specific separation of DNA molecules by using different denaturing conditions (temperature and salt concentration) in a washing solution. Several techniques for SNP genotyping use the hybridization principle, but to develop techniques with multiplexing levels has proven difficult since the hybridization and denaturing

conditions are sequence dependent (Southern et al 1999). However, Affymetrix developed an approach that enables the determination of alleles of over 906,600 SNPs in one experiment. The genotype of a single SNP is inferred from the fluorescent signal based on probe-to-target hybridization of up to 40 oligonucleotides. This redundant information is used to compensate for the sequence-specifc probe bias and thus to increase genotyping accuracy. Briefley, Affymetrix chips use short probe quartets, consisting of a perfect match (PM) and a mismatch (MM) 25-mer oligonucleotide for each allele (arbitrarily named allele A and allele B). Hence, a single SNP is interrogated by four different probes – PMA, PMB, MMA and MMB which represent the basic unit of quantifying allele-specific hybridization. Each SNP has multiple quartets querying different strands and shifts surrounding the polymorphic side (Xiao et al 2007). For its analysis, a DNA sample is digested with restriction enzymes, ligated to adaptors, and amplified using a universal primer that recognizes the adaptor sequences. The amplified DNA is subsequently fragmented, labeled, and hybridized to the oligonucleotide array under stringent conditions and followed by a washing step and fluorescence labeling. For CNV detection, the signal intensities of the match and mismatch probes are compared with values from another individual (or group of individuals) and the relative copy number per locus is determined.

The second and prevalent group of genotyping techniques uses an enzymatic discrimination step to detect SNP alleles (Kim et al 2007). These enzyme-based techniques seem to be more specific than the hybridization assays (Pastinen et al 1997). The employed enzymes are DNA polymerases and DNA ligases which are involved in DNA replication and repair and are highly sensitive to matched and mismatched nucleotides. For whole genome SNP genotyping the Infinium II assay developed by Illumina uses single base extension (SBE) (Steemers et al 2006). The protocol involves an isothermal amplification of the genomic DNA to be tested (non-PCR based), posterior fragmentation and hybridization onto a bead-array containing 50mer oligonucleotide primers designed so that their 3’ end hybridize to the nucleotide adjacent to the SNP. The DNA polymerase enzyme will extend the primer over the SNP site with differentially-labelled terminators. The Infinium assay produces two-color readouts (one color for each allele) for each SNP. Intensity values for each of the two-color channels, A and B, convey information about the allelic ratio at a genomic locus. Since samples that have identical genotypes at an assayed locus exhibit similar signal profiles (A and B values) and aggregate in one of the three possible clusters (AA, AB, BB), genotypes can be called by referencing assay signal intensities against known cluster data for a given locus. As for Affymetrix arrays, various signal-intensity data can be employed by CNV detection algorithms,

including the log R ratio (mentioned before) and the B allele frequency (BAF, calculated from the difference between the expected position of the cluster group and the actual value).
Initially, whole-genome genotyping technologies used probes that were not uniformly distributed across the genome (particularly sparse in regions of segmental duplication and other SNP deserts). To overcome this limitation, both Illumina and Affymetrix have designed microarrays that contain, apart of SNP oligonucleotides, additional non-polymorphic probes to assess chromosomal copy-number changes in genomic regions that are not well covered by SNPs. Last, similar problems with regard to the inherent background noise apply for SNP-based arrays. Therefore, as for the case of CGH-arrays the call of a CNV will require the minimum support of 3-5 consecutive SNPs.

In research work presented in this thesis, efficiency considerations led to the decision on which DNA microarray technology to be chosen. These considerations were based on the different capabilities and limitations of the different oligonuclelotide array types. The obvious advantage of SNP array platforms as compared to the aCGH technology is that they can detect CNVs and assay SNP genotypes simultaneously, thus allowing for a broader range of downstream analyses such as CNV, linkage and association-studies. Moreover, their ability to detect loss of heterozygosity (LOH) permits studies based on homozygosity mapping. Therefore, the versatility of SNP arrays played in their favor. At the onset of this thesis (September of 2006), Illumina as well Affymetrix offered whole genome SNP oligonucleotide microarrays with over 500,000 features. Pilot studies were performed in the institute in order to evaluate the Illumina HumanHap550 and Affymetrix Mapping 500K arrays from a viewpoint of platform performance. Both the signal-to-noise ratio and the SNP call rate seemed to be higher for the Illumina microarray data. Therefore, Illumina was the platform of choice for carrying out a set of different studies including the screening for CNVs in disease populations presented here.

1.2.3 CNV detection algorithms for oligonucleotide-based arrays

Data analysis includes the preprocessing of raw data (normalization) and the application of bioinformatic and statistic tools that allow the prediction of changes in genetic dosage. The process of normalization consists in adjusting the intensity raw data to a value of 1. Normalization implicitly assumes that the raw intensity values of the majority of probes correspond to an identical number of copies (two copies for a

diploid organism).

Many methods/algorithms have been proposed for detecting CNVs by array CGH or SNP arrays. The simplest method applies a cut-off value after smoothing the log R ratio. Using a moving average window of consecutive probes for smoothing, it is possible to detect duplication and deletion that surpass the specified threshold (Pollack et al 2002, Vermeesch et al 2005). Other smoothing algorithms have also been proposed including the quantile smoothing method (Eilers et al 2006) and the wavelet de-noising method (Hsu et al 2005). For a rigorous analysis of microarray data and determination of gains and losses more complex stadistical approaches are needed. At the moment a multitude of such computation intensive methods are available. They may be based on clustering methods (Wang et al 2005), maximum likelihood procedures (Jong et al 2004, Picard et al 2005) and Hidden Markov Models (HMMs) (Fridlyand et al 2004, Chen et al 2005, Wang et al 2007, Colella et al 2007, Korn et al 2008). Other programs widely implemented are based on segmentation algorithms such as the Circular Binary Segementation (CBS) (Olshen et al 2004), Gain and Loss of DNA algorithm (GLAD) (Hupé et al 2004), Smith-Waterman algorithm (SW) (Price et al 2005), Genome Alteration Detection algorithm (GADA) (Pique-Regi et al 2008), and SNP conditional mixture modelling (SCIMM) (Cooper et al 2008). It is relevant to note that all the analysis methods developed so far have their own pros and cons when modelling CNVs, and that some of them have been developed to suite specifically to a single platform. Thus, there is no well-formulated statistical method able to address the between-platform differences, assay methods, sensitivity and analytical complexity.

Several studies have compared the performance of methods or programs, mainly based on simulation data, arrayCGH data (Lai et al 2005, Willenbrock et al 2005) and Illumina SNP array data (Dellinger et al 2010). Lai et al. (2005) tested 11 programs using array CGH data and found that smoothing and clustering methods are effective when applied to simulation data but, as compared to other methods do not perform well enough in CNV detection when applied to experimental data. The best results were obtained with segmentation methods, especially with the Circular Binary Segementation algorithm. This method originally developed for arrayCGH analysis has been modified to allow SNP array analysis as well. The performance of CBS algorithm and of other six methods, including nonsegmentation methods such as QuantiSNP (Colella et al 2007) and PennCNV (Wang et al 2007), was evaluated in real and simulated SNP array data by Dellinger and colleagues (2010). These authours demonstrated that QuantiSNP outperformes the other methods both in real and simulated datasets. QuantiSNP was designed for Illumina data. It was shown, however, that it is the top method for Affymetrix data, too. These authors also reported that CBS

had problems in identifying CNVs that were detected by the majority of CNV detection methods, although in general it performed considerably well and it outperformed methods that were created for SNP arrays. Comparing different CNV calling algorithms it was frequently observed that the CNVs predicted by different algorithms differed in size and location. This has led several authors to recommend the use of two independent algorithms on a single dataset in order to increase confidence in the data and give clearer indications of the breakpoint events (Winchester et al 2009). However, the production of unequivocal CNV results by microarray techniques has proven difficult and ultimately it is desirable to validate CNVs with direct techniques such as FISH (fluorescent in situ hybridation), qPCR (quantitative polymerase chain reaction) or MLPA (multiple ligation probe-dependent amplification).

1.3 DISEASES INVESTIGATED WITH ILLUMINA SNP-ARRAYS

1.3.1 INTELLECTUAL DISABILITY

1.3.1.1 Definition, classification, prevalence and causes of intellectual disability

Mental retardation (MR) or intellectual disability (ID) (Schalock et al 2007), comprises a group of cognitive disorders in which brain development and/or function is impaired. As described by the American Association in Intellectual and Developmental Disabilities (AAIDD) and the American Psychiatric Association (APA), intellectual disability is defined as "a disability characterized by significant limitations both in intellectual functioning and in adaptive behavior, which covers many everyday social and practical skills. This disability originates before the age of 18." According to the World Health Organization (WHO) published international diagnostic system manual ICD-10 (International Statistical Classification of Diseases and Related Health Problems, 10th edition) intellectual disability is defined as "a condition of arrested or incomplete development of the mind, which is especially characterized by impairment of skills manifested during the developmental period, skills which contribute to the overall level of intelligence, i.e. cognitive, language, motor, and social abilities."

Intellectual functioning level is determined by a standardized age-dependent intelligence test (intelligence quotient or IQ), such as Wechsler Intelligence Scales for Children (WISC) or for adult (WAIS). Individuals with an IQ score below 70 (i.e., more than 2 SD bellow the median) and significant limitations in two or

more adaptive skill areas are considered to be intellectually disabled. However, intellectual disability varies in severity. The Diagnostic and Statistical Manual of Mental Disorders, fourth edition, text revision (DSM-IV-TR), classifies four different levels of intellectual disability: mild, moderate, severe, and profound. These categories are based on the person's level of functioning impairment, both on the intellectual area and on the adaptive skills. Table 1-2 shows the different degrees of intellectual disability classified by the IQ-score and the levels of functioning, according to the international diagnostic system manual ICD-10 (Blocks F70-F73).

Table 1-2: **Classification of intellectual disability (ID) based on the intelligence quotient (IQ)-score and capacity of functioning (according to the international diagnostic system manual ICD-10; Blocks F70-F73)**

Class	IQ-score	Levels of functioning
Mild	50-69 (in adults, mental age from 9 to under 12 years).	Learning difficulties in school. Many adults will be able to work and maintain good social relationships and contribute to society.
Moderate	35-49 (in adults, mental age from 6 to under 9 years).	Marked developmental delays in childhood but most can learn to develop some degree of independence in self-care and acquire adequate communication and academic skills. Adults will need varying degrees of support to live and work in the community.
Severe	20-34 (in adults, mental age from 3 to under 6 years).	Likely to result in continuous need of support.
Profound	<20 (in adults, mental age below 3 years).	Severe limitation in self-care, continence, communication and mobility

Syndromic intellectual disability refers to intellectual deficits associated with dysmorphic, metabolic, neuromuscular, or psychic manifestations. Non-syndromic intellectual disability refers to intellectual deficits that appear without other abnormalities. The prevalence of intellectual disability in children (syndromic and non-syndromic) is estimated to be 2-3% (Roeleveld et al 1997, Schevell et al 2003), although a review of several studies indicated that the range varies from 1% to 8% (Ropers 2010). Different rates can be expected depending on the population, methods of assessment, and criteria of assessment that are used. Interestingly, it has been observed that most of the variation in prevalence is associated with mild ID (IQ 50-69), whereas for moderate to severe mental retardation many studies agree on values between 0.3-0.5% (Roeleveld et al. 1997, Leonard et al 2002, Durkin 2002). One possible

explanation for this difference in variation is an ascertainment bias introduced when referring children with mild intellectual disability (IQ: 50-69). Mild ID could be overlooked specially for those cases without syndromic signs or symptoms.

Intellectual disability affects about 3% of the population, yet a diagnosis is obtained in only about a third of the cases (Johnson et al. 2006, Rauch et al 2006). In about 4 of 10 patients with moderate to severe intellectual disability (IQ<50) the etiology of mental retardation is unknown, whereas for the mild form (IQ 50-70) the cause cannot be found in 70-80% of the cases (Knight et al 1999).

Causes of intellectual disability can be due to biological or to environmental factors. Nongenetic causes include perinatal brain ischemia, fetal alcohol syndrome, pre- or postnatal infections, maternal exposure to certain types of disease or toxins, malnutrition and iodine deficiency. Genetic causes include chromosomal aberrations and single gene defects, which may arise *de novo* or through vertical transmission. About 40% of cases of intellectual disability are caused by hereditary factors (Rauch et al 2006). Chromosomal and other genetic disorders account for 30–40% of the cases with moderate to severe mental retardation, but only for 15% of the cases with mild intellectual disability (Knight et al 1999). In addition, epidemiologic studies on mentally handicapped individuals repeatedly have shown a sex bias, with a 30% excess of males over females (Zechner et al 2001, Mandel et al 2004). Most probably this excess is due to the fact that males have only one X-chromosome.

1.3.1.2 Genetic causes of intellectual disability

1.3.1.2.1 Chromosomal abnormalities

Chromosomal abnormalities are classified into two categories: numerical chromosome abnormalities and structural abnormalities.

Numerical chromosomal abnormalities (i.e. aneuploidy) are caused by an extra or a missing chromosome that results from nondisjunction or anaphase lag in meiosis and/or mitosis. Autosomal monosomies and trisomies usually are lethal in embryofetogenesis, with the exception of trisomies for chromosomes 13 (Patau syndrome), 18 (Edwards syndrome), and 21 (Down syndrome (DS)), which are observed in live born infants with an approximate frequency of 1:12000, 1:6000 and 1:700, respectively (Kuhn et al 1987). The fact that chromosome 13, 18 and 21 comprise a relative small number of genes may explain why only these polyploidies are viable. All the three types of trisomies result in mental impairment, but only trisomy

21 is compatible with long-term survival. Trisomy 21 (Down syndrome) is the most prevalent form of intellectual disability. In various studies on patients with ID the frequency of DS was as high as 12 % (Leonard et al 2002).

Numerical sex chromosome abnormalities are more common than numerical autosomal abnormalities, but not necessarily associated with mental retardation, possibly relating to the phenomenon that all except one X chromosome are subject to inactivation (Lyonization). Turner syndrome (females with only one X chromosome) and Klinefelter patients (XXY males) may be intellectually normal. However, with the number of additional X chromosomes the probability and severity of mental retardation increases. 70% of triple X syndrome patients, for instance, are mentally retarded.

Deletions, insertions, inversions and translocations may occur at any part of any chromosome and are produced by misrepair of chromosome breaks or by nonallelic homologous recombination either in meiosis or mitosis (see section 2.1.3 above). Translocations and inversion can remain without clinical consequences as long as they are balanced (without loss or gain of genetic material), do not interrupt a gene, and do not cause a pathogenic position shift of a chromosome segment (so-called position effect).

An example of balanced rearrangement is the Robertsonian translocation which results from the fusion of the long arms of two acrocentric chromosomes (13, 14, 15, 21 or 22) at their centromeres. As the short arm of these acrocentric chromosomes contains abundant ribosomal sequences only, the carrier of this Robertsonian translocation remains unaffected. However, carriers of a balanced rearrangement have an increased risk of infertility, miscarriages, and offspring with trisomy (due to the formation of unbalanced gametes) or uniparental disomy (resulting from trisomy rescue).

Unbalanced translocations, deletions and duplications are a major cause of intellectual disability. Only a minority is visible by conventional microscopic chromosome analysis. Examples of cytogenetically visible deletions are the Cri-du-chat syndrome (also known as chromosome 5p deletion syndrome or 5p monosomy**)** and the Wolf-Hirschhorn syndrome (also known as chromosome 4p deletion syndrome). The structural abnormalities not visible under a microscope are called “submicroscopic” or cryptic and in turn can be classified in interstitial or subtelomeric subtypes, depending on the position of the chromosome where they occur. Typically, submicroscopic rearrangements include copy number variants (CNVs), segmental duplications and submicroscopic inversions/translocations. Submicroscopic deletions or duplications that affect a contiguous series of genes represent the so-called microdeletion/microduplication

syndromes. These disorders were initially recognized in patients with recognizable phenotypic features and specific cognition profiles. Among them we can find Williams-Beuren syndrome, Smith-Magenis syndrome and DiGeorge/Velo-cardio-facial syndrome. Submicroscopic deletions or duplications also are found in Charcot-Marie-Tooth type 1, hereditary neuropathy with liability to pressure palsies (HNPP), Sotos syndrome, Neurofibromatosis type I, Prader-Willi syndrome and Angelman syndrome. In to all of these disorders the critical region is flanked by LCRs, thus being prone to NAHR, and consequently to the large CNV (Chen et al 1997). Such genomic disorders can be detected by region-specific FISH, MLPA or qPCR analysis or, more recently, by genome-wide microarray examination. As a result of the introduction of array-based analysis, novel microdeletion and microduplication syndromes have been described by identifying first overlapping genotypes in cohorts of patients with unspecific mental retardation and defining later the clinical phenotype. The 17q21.31 microdeletion syndrome was the first genomic disorder identified by this approach (Koolen et al 2006, Sharp et al 2006, Shaw-Smith et al 2006). Soon afterwards, 18 additional recurrent interstitial microdeletion/microduplication syndromes were identified. (Vissers et al 2009) Of these, the smallest recurrent CNVs, such as the deletions at 17q21.31 and 16p11.2, were at least 500 kb in size. Present array techniques can detect genomic imbalances as small as 500 bp. Their use has shown the existence of a large set of patients carrying nonrecurrent pathogenic imbalances (large and small) which seem to be private to each patient (Conrad et al 2010). Most of the rare, nonrecurrent CNVs probably are due to NAHR and FoSTeS (Ropers 2007).

Subtelomeric abnormalities constitute another major group of CNVs associated to intellectual disability. The telomeric regions are extremely gene-rich which explains why the relatively small deletions of subtelomeric sequences frequently cause intellectual disability such as observed in the Miller-Dieker syndrome (deletion of the 17p telomere) and the alpha-thalassemia/mental retardation syndrome (deletion of the 16p telomere). Molecular cytogenetic techniques have shown that causative submicroscopic rearrangements of the telomeric regions may be found in up to 5% of patients with intellectual disability (Flint et al 1995, Flint et al 2003, Knight et al 1999, de Vries et al 2003). While subtelomeric imbalances initially were searched for by multiplex FISH screening they can now be detected by the use of genome-wide microarrays.

Chromosomal abnormalities (both numerical and structural) may occur in mosaic state. Mosaicism is defined as the presence of two or more populations of cells in the same individual that differ genetically. Lethal autosomal aneuploidies may be viable in a mosaic state. A recent study of 2019 ID patients found

that 10% of the abnormalities diagnosed were mosaic aneuploidies (Conlin et al 2010). Microdeletions, microduplications, rings, and other types of structural rearrangement in a mosaic state may also lead to mental retardation (Friedman et al 2006, Gijsbers et al 2009, Koolen et al 2009). Mosaic states may not be detectable in blood cells thus evading detection by the usual diagnostic approach.

1.3.1.2.2 Monogenic disorders

Monogenic disorders follow the Mendelian patterns of dominant or recessive inheritance. Intellectual disability with Mendelian inheritance in humans have traditionally been identified and studied by searching for families with multiple affected members. The gene was localized by coinheritance of phenotype and genetic marker ("linkage analysis") and identified by positional cloning. In some cases the trait may be inherited as a X-chromosomal condition with males being more frequently affected than females. The first gene identified by linkage and positional cloning was the *FMR1* gene that causes Fragile X syndrome, the most common form of X-linked intellectual disability (XL-ID) (Oberle et al 1991). Before the identification of *FMR1,* the Fragile X syndrome was diagnosed cytogenetically by the detection of a visible fragile site at Xq27.3.

Large pedigrees are more common in X-linked conditions, since female carriers reproduce normally. Thus, linkage analysis has led to the discovery of over 60 X-linked genes of nonsyndromic intellectual disability while only 7 such genes have been identified on autosomes, primarily by homozygosity mapping in consanguineous families (Ropers 2006). Currently, point mutations and small insertions or deletions (indels) have been found in over 90 X-linked ID genes, which correspond to 10% of the coding genes of the X chromosome (Roper 2010).

Mental retardation occurs sporadically in most of the cases, which are not amenable to linkage mapping. Balanced reciprocal translocations that disrupt essential genes for brain development have been an important cytogenetic tool for gene discovery. An illustrative example is *TSPAN7* gene which was associated to nonsyndromic X-linked MR after being mapped to an X chromosome breakpoint in an individual with an X;autosome translocation (Zemni et al 2000). Another source of genes relevant to brain functioning and/or development are the contiguous gene microdeletions in which a primary candidate gene can be found. Several studies have shown that some of the known genomic disorders are also diagnosed in patients carrying point mutations in one of the genes contained in the common rearrangement instead of the

recurrent microdeletions. Few examples of these are *UBE3A* in Angelman syndrome (Kishino et al 1997), *RAI1* in Smith-Magenis syndrome (Slager et al 2003) and *CREBBP* in Rubinstein Taybi (Petrij et al 1995).
With the application of genomic microarrays it is now possible to identify genomic regions containing as few as one disease gene. In case of intellectual disability, for instance, the detection of cytogenetic microdeletions by microarrays technologies has led to the localization and identification of intellectual disability (ID) genes such as *PTCHD1, WDR13, FAAH2, GSPT2* (Whibley et al 2010).
Other methods also contribute to the detection of ID-causing genes. The use of mouse models with specific neurological phenotypes and the screening of genes that are in the same pathways or within protein complexes that include known genes are other approaches used to identify ID genes. For instance, as the glutamatergic pathway has been recognized to be a key player in ID, 197 candidate genes that code for members of known functional protein complexes were sequenced. Pathogenic *de novo* mutations were identified in four of them (Hamdan et al 2011).
Recently, technological advances in sequencing entire human coding regions (e.g., exome sequencing) have had a profound impact on ID gene identification. Tarpey et al (2009) sequenced all genes on the X chromosome in X-linked ID patients and identified nine new ID genes. One study that used exome sequencing to interrogate the whole genome identified *de novo* mutations in *MLL2* as cause of Kabuki syndrome, a well known syndromic condition whose molecular basis had been elusive for a long time (Ng et al 2010).
If the X chromosome is regarded as a model of the genome, one can predict that more than 1400 human autosomal genes may be involved in monogenic diseases (Raymond 2010). With the further incorporation of exome sequencing into the diagnostic laboratory, the number of genes responsible for ID should continue to increase.

1.3.1.2.3 Genomic imprinting and uniparental disomy (UPD)

Some patients with intellectual disability have been found to carry deleterious mutations (point mutations, indels and deletions/duplications) in genes which are directly or indirectly involved in imprinting modulation. Eighteen of the currently known ID genes play a role in modulation of chromatin structures, fact that permits to classify such disorders into the group namely imprinting syndromes (Kramer et al 2009).

Imprinting is an important mechanism to regulate gene expression. Epigenetic marks of the chromatin, such as CpG site methylation or post-translational modifications of histones (e.g. acetylation) may alter the chromatin structure and consequently the expression of a gene (or nearby genes) by activating or repressing, respectively. Currently, about 100 imprinted genes have been identified (Nakabayashi et al 2011). In total it is assumed that about 1% (200-300) of the genes in the human genome are imprinted. These genes are grouped in clusters and are under control of an imprinting center or imprinting control element (Franklin et al 2011). The discovery in 1999 of the molecular basis of Rett syndrome (i.e., defects of the gene encoding the methyl-CpG-binding protein 2, *MECP2*) was the first example of a relation between epigenetics with brain functioning (Amir et al 1999). Meanwhile, partly by the discovery of further imprinting syndromes, partly by the analysis of animal models, it is well known that neural mechanisms are fine-regulated by epigenetic modulation (reviewed in Kramer et al 2009 and Franklin et al 2011).

Genomic imprinting is a biological process by which a gene or a genomic domain is marked with information of its parental origin. Imprinted genes are preferentially expressed monoallelically from either the maternal or the paternal copy, resulting in unbalanced gene expression between homologous alleles.

Appropriate expression of imprinted genes is important for normal development. Alteration of the correct gene expression pattern of these clusters may lead to severe phenotypic consequences. These parent-of-origin effects of specific genes or groups of genes are most obvious in syndromic disorders caused by genomic imbalances. The best known example comprises two different disorders, Angelman syndrome and Prader-Willi syndrome, which are linked to the same imprinted autosomal domain on 15q11-q13. Microdeletion of the maternal allele gives rise to Angelman syndrome whereas microdeletion of the paternal allele cause Prader Willi syndrome. Another defect that can lead to Angelman or Prader Willi syndromes is uniparental disomy (UDP). In this situation, both copies of the chromosome 15 (or of a part of it) have the same parental origin. UPD can occur as a random event during the formation of egg or sperm cells or may happen in early fetal development as a result of so-called trisomic rescue. Other syndromes caused by the unique vulnerability of imprinted loci include Silver-Russell syndrome, Beckwith-Wiedemann syndrome and Turner syndrome. In fact, one strategy to identify imprinted genes is based on the delineation of UDP genotype/phenotype correlations. In turn, UDP can be identified with genetic markers by studying their parental origin. SNP oligonucleotide microarrays permit to both identify stretches of homozygosity (i.e. loss of heterogeneity, LOH) and to do parent-of-origin analyses, thus

allowing the identification of UDP. Finally, the use of epigenetic biomarkers can be used as a diagnostic tool. For example, determining the differential methylation of *HSA21* has been suggested as a method for noninvasive prenatal detection of *HSA21* trisomy, i.e., Down Syndrome (Tong et al 2010).

1.3.1.2.4 Complex neuropsychiatric and cognitive disorders

Polygenic or multifactorial disorders frequently are referred to as complex genetic disorders. Unlike monogenic traits, they are influenced by more than one gene underlying the susceptibility to the trait. Neuropsychiatric disorders, like autism spectrul disorder (ASD), schizophrenia (SZ) and bipolar disorder exhibit 50%-80% heritability - the fraction of variation in risk in a population that is attributable to inherited factors (Cook and Scherer 2008, Hirschhorn et al 2011) - but do not display an obvious pattern of Mendelian inheritance. It is also assumed that a multitude of genes that are neither necessary nor sufficient for the development of the phenotype contribute to intellectual disability, especially among those patients with mild and nonsyndromic forms (Ropers 2010).

Two hypotheses have been proposed in order to explain the genetic basis of complex neuropsychiatric disorders: 1) the polygenic model where a multitude of common alleles in the population may contribute to the phenotype with small to moderate effects, and 2) the oligogenic model where rare, highly penetrant variants underlie the susceptibility to the disease by means of large effects. Since 2006 genome-wide association studies (which identify common genetic variations associated to the disease) have identified several loci that contribute to risk of schizophrenia, bipolar disorder or autism spectrum disorder (Purcell et al 2009, Wang et al 2009, Weiss et al 2009) but none could be identified in case of intellectual disability (ID), probably due to the high degree of clinical and genetical heterogeneity. By contrast, several studies have shown that rare structural variants, deletions or duplications (and point mutations), can have substantial effects on all of these diseases. Case-control studies that used microarrays to screen for CNV in extended cohorts of unexplained ID, schizophrenia and autism spectrum disorder have shown that microdeletions and microduplications at the 1q21.1, 15q11.2, 15q13.3, 16p11.2, 16p13.11, and 22q11.2 loci are more frequent in cases than in controls (de Kobel et al 2010). These rare CNVs have also been found in relatives and in the control population, suggesting that the imbalances themselves are not sufficient to cause the phenotype (Cook and Scherer 2008). Therefore, they should be considered as risk alleles rather than as fully penetrant causative mutations. The fact that these CNVs can confer a pleiotropic susceptibility effect to a broad range of neuropsychiatric disorders pinpoints the existence of shared

biological pathways and the co-occurrence of other genetic, epigenetic or environmental factors which is the determinant of the peculiar symptomatology of each clinical condition.

It has been observed that these pathways mostly involve synapse formation and maintenance as well as neurotransmission with a special emphasis on glutamate and GABA (Guilmatre et al 2009). In addition, studies of structural genomic variations have revealed that a portion of these patients show two or more *de novo* CNVs (Marshall et al 2008, Xu et al 2008). These findings permitted to postulate a two-hit model; where a second genomic alteration, in the form of another large CNV, might serve as a second hit that would convert the recurrent microdeletion/microduplication from a risk factor to a determinant or modifier of the phenotype. Recently, investigations on children with severe developmental delay provided evidence in support of this model: Girirajan et al (2010) reported that the frequency of the 520kb deletion at 16p12.1 is four times higher in cases than in controls and a quarter of the affected individuals with 16p12.1 carried a second genomic alteration. They also found an excess of a second-site genomic events in other clinically variable microdeletion syndromes such as 15q13.3 deletion, 16p11.2 deletion, 22q11.2 duplication. Few or no double hits were observed in syndromes that almost always are caused by *de novo* and highly penetrant large CNVs such as William-Beuren syndrome and Smith-Magenis syndrome.

Lastly, CNV studies on neuropsychiatric disorders have shown that rare *de novo* structural variants sometimes are present in a single patient. Also inherited CNVs which seem to be private to each family have been reported (Cool et al 2008). It is therefore difficult to decipher which of these variations are causative, which are risk factors and which are only rare polymorphisms unrelated to any pathological phenotype. Since disease association of CNVs has to be tested systematically by comparing the frequency of each candidate CNV in patients and controls and given the very low frequency of these CNVs, the evaluation of their significance will require the study of large number of patients, i.e., by as meta-analytic approach. With more of such data at hand it will be interesting to see which molecular pathways are implicated in the development and functioning of the brain, which risk alleles or protective modifiers (genetic, epigenetic and environmental factors) can influence the expressivity of neuropsychiatric disease and in which combinatorial or hierarchical fashion of these factors takes in fact the phenotype across the threshold of liability.

1.3.1.3 Workflow and Diagnosis approach

Knowing the cause in a patient with intellectual disability is necessary for assessing the recurrence risk, the short and long term prognosis and the treatment options, but diagnosis still is a considerable challenge because of the broad spectrum of potential causes and the wide range of available tests. G-banded karyotyping has been the standard first-tier test for detection of genetic imbalances in patients with intellectual disability for the last 35 years. During the last decade, since the cost of array-based assays has dropped markedly and resolution limits and genome coverage have been improved, molecular karyotyping has become a routine diagnostic tool. A general observation is that the diagnostic rates of arrays are higher than the ones obtained with G-banding (Miller et al 2010). Consequently, this raised the debate about which clinical genetic test for this patient population should be adopted as a first instance.

In 2006, a study (Rauch et al 2006) evaluated a cohort of 1,170 ID patients with exhaustive clinical examination, conventional karyotyping, subtelomeric FISH, X-inactivation studies, molecular karyotyping and screening for metabolic disorders. The different types of analysis showed different detection rates but collectively yielded etiological diagnosis in 40% of the patients. The most common diagnoses were Down syndrome (9.2%), common microdeletion 22q11.2 (2.4%), William-Beuren syndrome (1.3%), Fragile-X syndrome (1.2%), Cohen syndrome (0.7%) and monosomy 1p36.3 (0,6%) (Rauch et al 2006). Based on detection rate and cost effective considerations the authors suggested examining patients without an etiological diagnosis with the following priority order; G-banded karyotyping, X-inactivation screening in mothers of boys, and molecular karyotyping. Solely when the last was not available it was recommended to use subtelomeric screening.

Beginning with 2008, however, several authors proposed CGH-array as the first choice after clinical evaluation (Shevell et al 2008). Gijsbers et al (2009) recommended using high-density SNP arrays instead of conventional karyotyping, due to the higher diagnostic yield and the ability to detect CNVs, mosaics, uniparenteral disomies and loss of heterozygosity. A review of 29 aCGH- and SNP-array-based studies, comprising 36,325 individuals with unexplained ID, pointed out that the balanced chromosome rearrangements such as reciprocal translocations and inversions, which are not detected by arrays, account for less than 1% of the diagnosed patients (Hochstenbach et al 2009). The most exhaustive comparison between microarray-based genomic copy number analysis and classical cytogenetics was conducted by Miller et al (2010), a retrospective study of 33 studies including 21,698 ID patients. After stating all the technical advantages and limitations of both techniques, they suggested that molecular karyotyping should

replace conventional chromosome analysis for patients with unknown ID, and reserving G-banded karyotyping only for patients with obvious chromosomal syndromes (e.g., Down syndrome), a family history of chromosomal rearrangements, or a history of multiple miscarriages. However, it is relevant to note that microarray-based genomic copy number analysis will not be widely implemented and accepted world-wide as the first-tier test until sufficient efforts of standardization are made. It is mandatory to unify guidelines for the expected clinical yield of ID, what in turn would permit to define the requirements of array-design, resolution and coverage, i.e., whole genome or locus specific. Another aspect that is still in a preliminary state is the creation of openly accessible centralized resources which include clinical interpretations of rare non-recurrent imbalances. This would facilitate the distinction of the pathogenic CNV among the thousands of variants and make possible the use of data by meta-analytic studies.

1.3.2 CONGENITAL HYPERINSULINISM

Hyperinsulinemia of infancy (HI; OMIM 601820) also known as persistent hyperinsulinemic hypoglycemia of infancy or congenital hyperinsulinism (CHI) is the most common cause of hypoglycemia during the newborn and infancy period (Stanley 1997). It is characterized by inappropiate oversecretion of insulin from pancreatic beta cells despite low blood glucose levels (Aynley-Green 1981). The primary aim of CHI treatment is to prevent irreversible brain damage in this group of patients by maintaining normoglycemia. The estimated incidence can vary from 1/40,000 live births in outbreed populations to 1/2500 in countries with high rates of consanguinity (Glaser et al 2000). CHI is a heterogeneous condition with respect to clinical presentation, histology and molecular basis. To date, 8 genes have been found to be mutated in CHI: *ABCC8, KCNJ11, GLUD1, GCK, HADH, HNF4A, SLC16A1* and *UCP2* (Thomas et al. 1996, Thomas 1995, Stanley et al. 1998, Glaser et al. 1998, Clayton et al 2001, Pearson et al. 2007, Otonkoski et al. 2007, González-Barroso et al 2008). The most common mechanism underlying CHI is dysfunction of the pancreatic β-cell ATP-sensitive potassium channel (K^+_{ATP} channel). This subtype of CHI thus are channelopathies. Recessive mutations in the SUR1 and Kir6.2 subunits of the K^+_{ATP} channel, which are encoded by *ABCC8* and *KCNJ11* respectively, account for 40-45% of all cases (Thomas et al 1995, Flanagan et al 2011). The remaining 5-10% of the CHI cases are caused by metabolopathies, Thus, gene mutations arise to enzyme or transcription factor deficiencies leading to an increase of beta cell

ATP/ADP ratio or accumulation of intermediary metabolites (Flanagan et al 2011). The various molecular mechanisms underlying CHI have been reviewed by Hussain (2007) and James et al (2010) and are presented here in a simplified version in Figure 1-2. Recessive CHI has been attribuited to, *ABCC8*, *KCNJ11* and *HADH* gene mutations, whereas autosomal dominant (or *de novo*) inheritance has been described for *GLUD1, GCK, HNF4A, SLC16A1* and *UCP2* genes. The clinical presentation of metabolopathies is very heterogeneous with respect to age of onset, severity of symptoms and degree of penetrance. In contrast, patients with recessive *ABCC8* or *KCNJ11* mutations tend to present a profound hypoglycaemia already in the neonatal period. However, also rare dominant-acting mutations in *ABCC8* or *KCNJ11* have been reported and they seem to be associated to milder phenotypes and better prognosis (Pinney et al 2008).

It is important to confirm the genetic subtype of CHI because the clinical management of the patients is very different. The treatment of CHI involves medical therapy and/or surgery. Diazoxide is the main treatment of CHI and works by binding to the SUR1 unit of the K^+_{ATP} channel and keeping the channel open, thus preventing insulin secretion. Therefore, as the target of this drug is the K^+_{ATP} channel, the majority of patients with mutations in *ABCC8* and *KCNJ11* do not respond to diazoxide treatment and require surgical removal of the pancreas (see Figure 1-1).

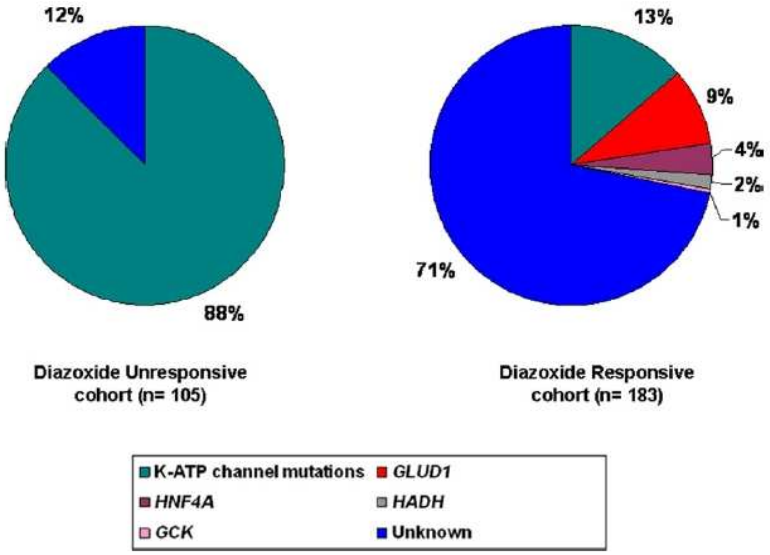


Figure 1-1: **Charts showing the genetic etiologies for 288 patients with CHI.** A genetic diagnosis was possible for 51 of 183 (28%) patients with diazoxide-responsive CHI and for 92 of 105 (88%) patients who did not respond to diazoxide therapy (Figure taken from Flanagan et al 2011)

CHI is not only genetically and clinically heterogeneous; there is also diversity in the pancreatic histopathology. Both diffuse and focal forms of CHI have been described (Rahier et al 1984). Focal CHI is characterized by a sporadic somatic islet-cell hyperplasia while diffuse CHI corresponds to a functional abnormality of insulin secretion in the whole pancreas. The typical diffuse form results from recessive mutations in the genes encoding the 2 subunits of the K^{+}_{ATP} channel. Instead, paternally inherited germ-line mutations of the *ABCC8* or *KCNJ11* genes, together with somatic haploinsufficiency of 11p15.5 in the maternal allele have been described to result in focal CHI (de Lonlay et al 1997). The somatically (only in the ß-cells) deleted 11p15.5 region contains the *ABCC8* and *KCNJ11* genes and the imprinted region with the *H19, P57KIP2, KVLQT1, HASH2* and *IGF*2 genes. Therefore, the unbalanced expression of these imprinted genes, including growth factor and tumor suppressor genes, and the hemizygosity of the paternal mutation result in a clonal proliferation of hyperfunctional pancreatic islets. Both focal and diffuse forms share similar clinical features, but require different surgical treatments. Focal lesions are effectively treated by limited pancreatic resection while diffuse lesions, that are caused by mutations in *ABCC8* and *KCNJ11* genes and are unresponsive to diaxozide, require subtotal pancreatectomy with a high risk of subsequent diabetes mellitus. To date the most frequently applied method to differenciate diffuse from focal CHI as well as to accurately locate the focal lesion is the fluoro-L-Dopa positron emission tomography (PET), a non-invasive technique (Ribeiro et al 2005, Otonkoski et al 2006).

In up to 50% of patients with CHI the genetic basis is still unknown. The low mutation detection rate may be explained by further locus heterogeneity or by missed mutations of known CHI genes. Intragenic and larger rearrangements, and mutations in regulatory or intronic regions may escape the usual analysis by direct sequencing and therefore a proportion of genetically unexplained cases might be caused by undetectable mutations of CHI genes, especially of *ABCC8* or *KCNJ11* that account for the largest fraction of genetically diagnosed CHI cases (see Figure 1-1). However, the fact that the vast majority of patients without genetic diagnosis respond well to diazoxide therapy (about 91% according to Flanagan et al 2011), suggests that there are further CHI genes implicated in ß-cell metabolism and/or encoding transcription factors. Hence, further investigations are necessary to understand the pathophysiology and molecular biology of the pancreatic ß-cells.

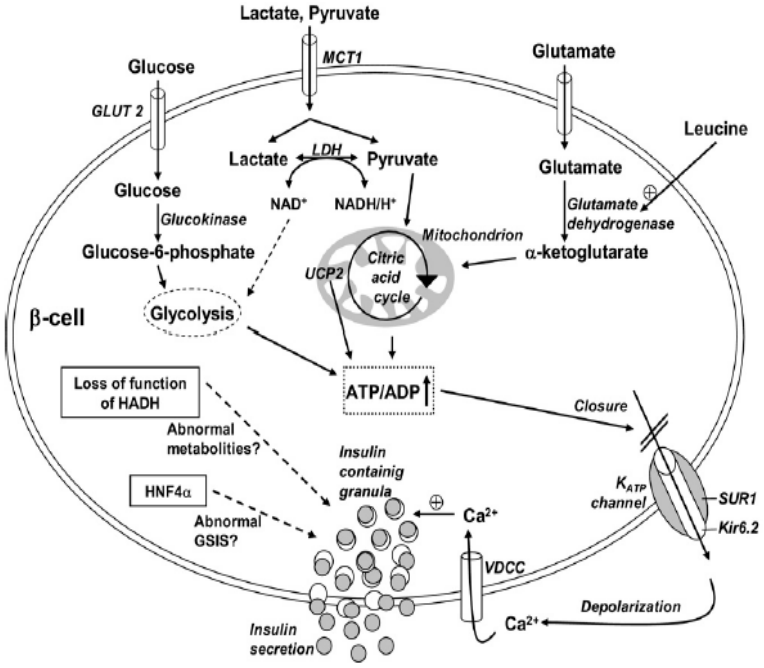


Figure 1-2: **Schematic representation of the regulation of insulin secretion in pancreatic beta cell and the known mechanisms implicated in CHI**. Following glucose uptake and metabolism, the ATP/ADP ratio increases and leads to closing of ATP-sensitive potassium (K^+_{ATP}) channels and depolarization of the cell membrane. This opens the voltage-dependent Ca2+ channels (VDCC), allows the influx of extracellular calcium, and finally the exocytosis of insulin. Defects in one of the 2 subunits of the K^+_{ATP} channel inhibits potassium outflow, leading to permanent depolarization of the cell membrane and unregulated insulin secretion. Other, less frequent mechanisms lead to an increased ATP/ADP ratio which triggers the closure of the K^+_{ATP} channel and the subsequent release of insulin. Some of these causes are: 1) Overactivity of *GCK* leading to increased glucose phosphorylation and lower threshold for glucose-stimulated insulin secretion (GSIS) 2) Overactivity of *GDH* and augmented leucine stimulation of glutamate oxidation which result in higher levels of alpha-ketoglutarate and consequent increase of the citric cycle, that is, ATP yield. 3) Abnormal expression of *MCT1* on the beta cells plasma membrane leading to oversensitivity to acute changes (by physical exercise) in the extracellular concentration of lactate and pyruvate, and consequent extra lactate and pyruvate uptake 4) Loss-of-function mutations of the mitochondrial uncoupling protein *UCP2* which normally inhibits the ATP yield from substrate oxidation. The molecular mechanisms of *HADH* (3-hydroxyacyl CoA dehydrogenase) hyperinsulinism and *HNF4a* hyperinsulinism are still unclear. (Marquard et al 2011)

1.3.3 TBX GENES IN HUMAN DISORDERS

The T-box genes constitute a family of transcription factors that have been implicated in the regulation of embryonic development of many organ systems. The first member was suspected in 1927 based on a mutant phenotype in mice. Mutations in the so-called T locus caused a short tail in heterozygotes and embryonic lethality in homozygotes (Dobrovolskaia-Zavadskaia 1927). However, the gene carrying the mutation, named *Brachyury* (short tail in Greek), was not cloned until 63 years later (Herrmann et al 1990). In 1992, Pflugfelder and colleagues discovered sequence homology between the mouse *T* gene and the Drosophila gene *omb*, which appeared to be a novel DNA binding domain (Kispert et al 1993). This finding made it possible that two years later a whole family of T-related genes was identified in the mouse genome (Bollag et al 1994). All the members shared a conserved protein motif with the *T* gene. This motif was therefore referred to as T-box or T-domain. Nowadays more than 50 genes belonging to the T-box family have been identified and mostly classified into 5 subfamilies (Wilson et al 2002). Orthologs and/or species-specific T-box genes have been found in the majority of animals investigated, such as *Caenorhabditis elegans*, Drosophila, sea urchin, ascidian, amphioxus, Xenopus, chick, newt, zebrafish, mouse, dog and human although some exceptions have been documented, such as the model plant *Arabidopsis thaliana* in which T-genes are absent (Papaioannou et al 1998, Wilson et al 2002). Phylogenetic studies have shown that the T-box family has an ancient origin, prior to the divergence of metazoans (Papaioannou et al 1998), and has arisen due to gene duplication and cluster dispersion (Agulnik et al 1996). This is evidenced by the existence of pairs of T-box genes showing tight linkage between them (Agulnik et al 1995). Two clusters of two T-box genes have been found in humans and, interestingly, their orthologs in mouse conserve identical synteny but a different chromosome location. In humans *TBX2* and *TBX4* are tightly linked on chromosome 17 and *TBX3* and *TBX5* are linked on chromosome 12, whereas in mouse these genes cluster in chomosomes 11 and 5, respectively (Agulnik et al 1996, Ruvinsky et al 1997). The initial expansion of the Tbx2/3/4/5 subfamily occurred at the beginning of vertebrate evolution (Papaioannou et al 1998) and as shown by Chapman et al (1996) the more recently duplicated members of this subfamily share high sequence similarity and common expression patterns. This led to some authors to suggest that they might have redundant functions as well (Papaioannou et al 1998). In general, the functional specificity of each T-box gene depends on the sequence-specific DNA-binding domain and on the transcriptional activator or repressor domain at C-terminal (Conlon et al 2001).

In addition, protein–protein interactions may influence the regulatory characteristics of T-domain proteins and also determine their target specificity (Showell et al 2004). Heterodimerization depends on the highly variable regions outside of the T-domain. Although embryological, biochemical and genetic studies have identified some of the molecular pathways in which T-box genes are involved, there is still little known about the regulation of the T-box genes, their transcriptional targets and their interaction partners.

However, in humans, the importance of T-box genes in the control of morphogenesis of a wide range of tissues and organs (skeletal or genital apparatus, cardiovascular or endocrine) has been well demonstrated since a significant number of inherited congenital disorders are caused by mutations in these genes. So far, the T-box genes implicated in human disorders are; *TBX5* in Holt-Oram syndrome (Basson et al 1997, Li et al 1997), *TBX3* in Ulnar-Mammary syndrome (Bamshad et al. 1997*), TBX1* in DiGeorge syndrome (Gong et al 2001), *TBX19* in isolated ACTH deficiency (Lamolet et al 2001), *TBX22* in X-linked cleft palate with ankyloglosia (Braybrook et al 2001), *TBX4* in Small Patella syndrome (Bongers et al. 2004), *TBX20* in congenital heart disease (Kirk et al 2007) and *TBX15* in Cousin syndrome (Lausch et al 2008). *TBX2* has been found to be amplified in a subset of cancers and its suppressor activity upon p19 ^{ARF} promoter suggests a role in carcinogenesis (Packham and Brook 2003). Moreover, duplications containing only the *T (brachyury*) gene have been found in four multiplex families with chordoma, thus pointing out another TBX gene to play a role in cancer susceptibility (Yang et al 2009). In addition, several studies have identified genome-wide significant association signals of SNPs near or within TBX genes; for example, SNP alleles of the *TBX4* locus are associated with developmental dysplasia of the hip (Wang et al 2010) and alleles close or within *TBX5* and *TBX3* are associated with both PR interval and blood pressure (Levy et al 2009, Pfeufer et al 2010, Holm et al 2010, Fox et al 2011, Smith et al 2011, Kato et al 2011). It is very likely that in the coming years more variants in T-box genes are found to be causative of or to confer susceptibility for other human diseases.

1.3.3.1 Holt-Oram syndrome

Holt-Oram syndrome (HOS, OMIM 142900) is an autosomal dominant disorder characterised by cardiac and skeletal congenital abnormailities (Holt and Oram 1960) and occurs with a frequency of 1 out of 100,000 live births (Elek et al 1991). HOS is completely penetrant but with a highly variable expression. Limb and heart malformations can vary from mild to severe even within families, and no correlation exists between the severity of the cardiac and skeletal abnormalities of a patient (Basson et al 1994, Newbury-

Ecob et al 1996). Clinically, there are 3 variations of HOS; affected individuals with only skeletal anomalies (27.4%), with only cardiac defects (3.9%), or with both (68.7%) (Lehner et al 2003).

The skeletal abnormalities affect the upper limbs including clinodactyly, limited supination, sloping shoulders and phocomelia. They affect the radial ray and are bilateral and asymmetrical (or even unilateral), frequently affecting the left side more severly than the right. The most prevalent findings are carpal bone anomalies which are present in almost all patients (Vaughan et al 2000). Cardiac defects include secundum atrial septal defect (ASD), ventricular septal defect (VSD), tetralogy of Fallot, hypoplastic left heart and cardiac conduction anomalies which may occur regardless of the presence or absence of structural cardiac defects. Heart malformations may require surgery in infancy in order to avoid sudden death due to heart block. Although cardiac anomalies are present in 72-75% of the HOS patients, they may also be entirely absent (Basson et al. 1994; Li et al 1997, Basson et al. 1999, Lehner et al 2003). Therefore, genetic background, environmental and stochastic modifiers are believed to influence greatly the severity of HOS manifestation (Basson et al 1999; Brassington et al 2003).

Holt-Oram syndrome is caused by mutations in the T-box transcription factor 5 (*TBX5*) gene. It consists of nine exons spanning over 54 kb of genomic DNA and comprises four alternative spliced isoforms. It regulates transcription of downstream genes important for cardiomyocyte differentiation and upper limb development such as the atrial natriuretic factor (*NPPA*), Sal-like protein 4 (*SALL4*) and fibroblast growth factor 10 (*FGF10*) by the binding to T-box binding elements (TBEs), often with interaction with the *NKX2-5* and *GATA4* transcription factor (Ghosh et al 2001, Hiroi et al 2001, Garg et al 2003, Koshiba-Takeuchi et al 2005).

From linkage studies, the gene was known to be located at 12q21.3-q22 (Bonnet et al 1994). The identification of a translocation at 12q2 further defined the interval containing the HOS locus (Terret et al 1994) and permitted to map the *TBX3* and *TBX5* genes within the critical region and to identify several mutations in *TBX5* in HOS patients (Li et al 1997, Basson et al 1997). Although *TBX5* mutations have been identified in familial cases of HOS, approximately 85% of affected individuals have HOS as the result of *de novo* mutations (Brassington et al 2003). There are more than 37 reported mutations (missense, nonsense, frameshift), and at least two recurrent mutations (Mori et al 2004, Garavelli et al 2008). Seventy percent of the *TBX5* mutations are null alleles which cause the HOS phenotypes by haploinsufficiency (Packham and Brook 2003, Postma et al 2008). Missense mutations are less frequent and lead to functional haploinsufficiency by diminishing *NKX2-5* and/or *GATA4* interaction or DNA binding (Cerbai et al 2008,

Boogerd et al 2010). Although the HOS-associated point mutations are distributed across all exons of *TBX5*, the majority are found within the T-box DNA binding domain (Mori et al 2004). Initial studies suggested that missense mutations at the 5' end of the T-box affected mainly the heart whereas those at the 3' end affected mainly the upper limb (Basson et al 1999, Yang et al 2000). However, mutational analyses performed in larger cohorts demonstrated that neither the kind of mutation nor its site seem to be predictive of the location of heart and limb defects or the severity of HOS manifestations (Brassington et al 2003, Heinritz et al 2005, Mcdermott et al 2005). Some studies showed that 22–42% of clinically diagnosed cases of HOS have a detectable mutation of *TBX5* (Cross et al 2000, Brassington et al 2003, Heinritz et al 2005, Borozdin et al 2006a), and up to 74% if more strict clinical criteria are applied (McDermott et al 2005). Several hypotheses have been proposed in order to explain the remaining proportion of HOS patients with unknown molecular basis (30-70%); 1) HOS is genetically heterogeneous 2) or/and the low mutation detection rate is due to the low sensitivity of conventional genetic testing (direct sequencing of coding exons of candidate genes). Genetic heterogeneity has been proposed in three families, but the lack of power to determine statistically significant linkage in one of the families and the presence of atypical clinical features (considered as exclusion criteria of HOS) in the others prevented the identification of a new HOS locus (Terret et al 1994, Bonnet et al 1995, Fryns et al 1996). In order to expand the spectrum of mutations associated with HOS several studies have looked for *TBX5* copy number changes in patients negative for *TBX5* point mutations. Akrami et al (2001) used multiplex amplificable probe hybridization (MAPH) for the screening of deletions/duplications whereas Borozdin et al (2006a, 2006b) employed quantitative real time PCR. These studies identified deletions of an exon(s) or the entire *TBX5* gene in about 2-5% of unselected HOS patients. Of note, all the mutations described in these studies have the same pathological mechanism in common, i.e. loss of protein function. Surprisingly, however, gain-of-function mutations have been also found to cause HOS. Postma et al (2008) reported a missense mutation G125R which lead to an increased transcriptional activation of *TBX5* target genes by enhancing the binding to their promoters. The affected family members with the G125R mutation showed mild skeletal anomalies and only few of them presented cardiac malformations. *TBX5* overexpression in transgenic mouse models produces the inhibition of ventricular chamber maturation (Liberatone et al 2000) whereas overexpression of *TBX5* in the developing chicken heart impairs cardiogenesis and inhibits cardiomyocyte proliferation (Kim et al 2000). In humans, a number of cases of 12q partial duplications, resulting in increased *TBX5* dosage, have been reported and shown to exhibit phenotypic features similar to Holt-Oram syndrome

including congenital heart defects and skeletal abnormalities (Harrod et al 1980, Melnyk et al 1981, Ieshima et al 1984, McCorquodale et al 1986, Dixon et al 1993, Cappellacci et al 2006). These data indicate therefore that balanced functional levels of *TBX5* are critical for the normal development of the heart and forelimbs, as both under- and overexpression of *TBX5* give rise to a HOS phenotype.

1.3.3.2 Ulnar-Mammary syndrome

The first case report of Ulnar-Mammary syndrome (UMS, OMIM 181450) was published by Pallister et al in 1976. UMS is an autosomal dominant disorder characterized by upper limb malformations on the ulnar side and hypoplasia of the breast and of other apocrine glands. Upper limb deformities may range from hypoplasia of the terminal phalanx of the fifth digit to complete absence of forearm and hand (Schinzel 1987). Duplication of posterior structures also may occur and in some individuals the ventral surface of posterior digits has been observed to be dorsalized (Bamshad et al 1997). Apocrine abnormalities include complete absence of breasts, axillary hair and perspiration but these clinical features may be minimal in other patients (Bamshad et al 1997). UMS patients also exhibit abnormal development of teeth and genitalia in both sexes as well as delayed onset of the puberty in males (Franceschini et al 1992). Additional symptoms in single cases may include pyloric- anal- and subglottic stenosis, obesity, short stature with or without growth hormone deficiency, cardiac malformations and pulmonary stenosis (Gonzalez et al 1976, Schinzel 1987, Bamshad et al 1996, Meneghini et al 2006). No association between the type or severity of limb involvement and apocrine or genital anomalies has been established (Bamshad et al 1996). The penetrance of the phenotype is complete while the expressivity highly variable, not only between families but also within them. Individuals with very few or only minor anomalies have had children with severe and disabling upper limb anomalies (Bamshad et al 1996). As in case of Holt-Oram syndrome (HOS, see above), this phenomenon is thought to be caused by epistatic modifiers (Bamshad et al 1999). Recently, Joss and colleagues (2011) noted that, regardless of the severity of the upper limb malformations, UMS patients seem to bear a typical facial appereance (broad nasal tip, broad jaw, prominent chin, and tongue frenulum), thus allowing to suspect the diagnosis of UMS even in patients with subtle phenotypes.

In 1995, linkage analysis in a pedigree with 33 affected members permitted to map the UMS locus on a 21-cM region at 12q23-q24.1 (Bamshad et al 1995), overlapping with the known HOS locus (Bonnet et al

1994). Both *TBX5* and *TBX3* genes were identified within this genomic region (Li et al 1997) and whereas mutations in *TBX5* were identified in Holt-Oram syndrome (Li et al 1997, Basson et al 1997) mutations in *TBX3* were found to cause Ulnar-Mammary syndrome (Bamshad et al 1997). *TBX3* contains 7 exons and spans 13.9 kb of genomic DNA. Three transcripts variants encode for different isoforms, but the full length of one of the variants is not completely determined yet (Bamshad et al 1999). *TBX3* encodes a transcription factor of the T-box family known to repress the cell cycle control gene *Cdkn2a* ($p14^{ARF}$) through the binding with a T-domain binding site which is specific for *TBX3* and *TBX2* (Carlson et al. 2001, He et al 1999). Crystallography studies showed that *TBX3* binds to its natural target sites as a monomer (Coll et al 2002). *TBX3* protein is expressed in forelimbs, hindlimbs, the epithelium of the mammary gland and the genital bud, but it has also been observed in a wide variety of tissues and organs that are not affected in UMS patients (Chapman et al 1996, Bamshad et al 1999, Davenport et al 2003). It is therefore believed that in these unaffected organs other genes from the T-box family may compensate for reduced levels of *TBX3* protein as a consequence of functional redundancy (Papaioannou et al 1998, Meneghini et al 2006).

To our knowledge a total of 62 patients from 15 families and 2 sporadic cases have been reported to date. Most of the molecular defects associated to UMS are nonsense or frameshift mutations leading to truncated proteins or missense mutations that result in impaired DNA binding (Bamshad et al 1997, Bamshad et al 1999, Sasaki et al 2002, Wollnick et al 2002, Joss et al 2011). As *TBX3* mutations are distributed across the whole gene (some located within the T-domain sequence and others located either upstream or downstream of the T-domain), several attempts for correlating different genotypes with specific phenotypes have been performed in the last years. Bamshad et al (1999) did not find any mutation-type or location to be associated with disease severity. In contrast, in a study that reviewed the mutation data in the literature and also reported some new cases Meneghini et al (2006) suggested that mutations affecting the T-domain of *TBX3* lead to more severe limb abnormalities. Actually, the number of cases published to date is too small, and additional studies that expand the mutational spectrum need to be done before any definitive conclusion about phenotype-genotype correlations can be reached. Since all the reported point mutations lead to a loss of function of the protein, haploinsuffiency for *TBX3* was postulated to be the pathogenetic mechanism underlying UMS (Bamshad et al 1997, Wollnik et al 2002). This was confirmed in 2006 when two studies identified genomic deletions encompassing the entire *TBX3* gene (Klopocki et al 2006, Borozdin et al 2006). A girl with severe ulnar limb malformations, dysmorphic facies, and mental retardation was shown to carry a 1.28Mb deletion containing only the *TBX3* gene (Klopocki et al 2006).

The authors suggested that the mental retardation of this patient might be caused by the deletion of non-coding sequences that control the expression of the nearby *THRAP2* gene which previously had been related to mental retardation (Muncke et al. 2003). Borozdin et al (2006) reported a 2.19 - 2.27Mb contiguous deletion encompassing the *TBX5*, *TBX3* and *RBM19* genes in patients with combined phenotype of HOS and UMS. The simultaneous loss of *TBX5* and *TBX3* would suggest very severe malformations of the forelimb but, in contrast, these patients exhibited mild limb involvement, only.

However, the variable degree of limb deformities observed in patients carrying *TBX3* deletions is in line with the broad range of phenotypic expressivity described for both UMS and HOS, and further supports the hypothesis that disease severity is highly dependent on the genetic background and the presence of epistatic modifiers

As mentioned in the previous section, several cases of 12q partial duplications have been published (Harrod et al 1980, Melnyk et al 1981, Ieshima et al 1984, McCorquodale et al 1986, Dixon et al 1993, Cappellacci et al 2006). Tipically, chromosomal alterations affecting the dosage of a large number of genes result in a mental retardation phenotype. Therefore the psychomotor delay presented by these patients may be attributable to the large 12q rearrangement which contains many genes. In addition to the mental retardation phenotype other clinical features such as cardiac malformations, limb anomalies, and spaced nipples were present in these patients. Such clinical features seemed to be associated specifically with the 12q duplication. Heart and limb defects are known to be caused by altered dosage of either *TBX5* or *TBX3* (Cernai et al 2008, Meneghini et al 2006) but since *TBX3* is known to be the major genetic determinant of breast development, the observed nipple alteration might be due to the overexpression of *TBX3*. The finding of patients with either cryptic *TBX3* duplications or gain-of-function *TBX3* point mutations would permit to elucidate whether the over-expression of *TBX3* (resulting in a over-repression of *Cdkn2a* ($p14^{ARF}$) and/or other potential *TBX3* target genes during embryogenesis) may impair the normal development of breast and limbs and, like *TBX3* haploinsufficiency, result in UMS.

2. AIMS OF THE INVESTIGATION

Due to a substantial increase in the power to detect copy number variations (CNVs) microarray analysis is rapidly replacing conventional chromosome analysis by karyotyping as the first step of a genomic examination.

The first goal of this study was to perform a systematic screen of a cohort of 219 sporadic novel patients with intellectual disability of unknown origin in order to evaluate and optimize the SNP-microarray technology for whole genome copy number analysis. For this purpose we used two different Illumina high density SNP microarrays; 1) the Human550-Quad arrays to study 109 DNA samples and 2) the updated Human610-Quadv1_B arrays to study a second set of 110 DNA samples.

The second and foremost aim of our application study was to use these arrays for reliable genome-wide detection of rare and/or *de novo* CNVs in patients of three different cohorts. In particular:

1) To perform SNP array analysis on DNA from 330 patients with unexplained intellectual disability (ID), most of them of unknown origin and an apparently balanced karyotype, to search for potentially pathogenic submicroscopic CNVs with a commercially available SNP array platform (Illumina).

2) To perform a genome-wide microarray scan for CNVs in 40 children referred for difuse congenital hyperinsulinism (CHI). As the conventional mutation screening in *ABCC8* and *KCNJ11* genes did not reveal any detectable abnormality, the patient DNAs were hybridized on high-dense SNP oligonucleotide arrays (Illumina).

3) To perform CNV genome-wide analysis in blood samples from an extensive german family with several members suffering from an atypical form of Holt-Oram syndrome. The index patient (ID: 31565), without detectable mutations in *TBX5* gene, was tested for the presence of cryptic chromosomal copy number variants using a high density SNP array (Illumina).

3. PATIENTS, MATERIALS AND METHODS

3.1 PATIENTS

Blood samples were sent to the Institute of Human Genetics of Helmholtz Research Center in Neuherberg from different genetics practices and infantile paediatrics departments of different national and international hospitals. Genomic DNA was isolated from peripheral cells and used in the following studies:

3.1.1 Patients with mental retardation

A cohort of 330 patients with intellectual disability and developmental delay (ID/DD) was referred to us, including sporadic cases with moderate to severe mental retardation (MR), mild MR and MR with epilepsy. The broad range of ID severity and the contribution of nonsyndromic and syndromic cases reflect the high heterogeneity degree of the ID condition. The majority of patients had been investigated for cytogenetic abnormalities (karyotype 500-550 G-band resolution). Most of the patients had shown normal G-banded chromosomes (banding level 500-550) and inconspicuous fragile X and subtelomeric FISH screenings. Three children with unexplained MR and an apparently *de novo* balanced reciprocal translocation detected by chromosomal karyotyping were also included in this study. We collected DNA material from 259 patients and their parents, whereas for the resting 71 patients samples of the parents were missing.

The patients were examined by experienced geneticist or pediatrician who sent us for each patient a clinical protocol where the information was collected and quantified for each of the possible manifestations and/or medical complications. Patients upon clinical suspicion for known causes of MR or with brain malformations were excluded from this study. We performed CNV genome-wide analysis on DNA from the patients to search for potentially pathogenic submicroscopic imbalances with a commercially available SNP array platform (Illumina).

Informed written consent was obtained from all the families and approval for the study was obtained by the ethical review boards of the participating institutions. The recruitment of patients was part of the German Mental Retardation Network (MRNET) study. The participating institutes and practices are listed below:

Dr. Stephanie Spranger and Dr. Bernd Kazmierczack, Practice of Human Genetics, Bremen
Dr. Michaela Nathrath, Depatment of Pediatrics, Munich Technical University, Munich
Dr. Dagmar Wahl, Practice of Human Genetics, Augsburg
Dr. Maja Hempel and Dr. Babett Heye Institute of Human Genetics, Kinderkrankenhaus, Munich
Dr. Florian Heinen, Dr. von Haunersches Kindersppital, Munich
Dr. Karl-Philipp Gloning and Dr. Cornelia Daumer-Haas, Prenatal Medicine, Munich
Dr. Steffen Leiz, Kinderklinic Dritter.Orden, Munich
Dr. Jürgen Kohlhase, Practice of Human Genetics, Freiburg
Dr. Nataliya Tyshchenko, Institute of Clinical Genetics, Dresden
Dr. Monika Cohen, KinderZentrum, Munich
Dr. Liesbeth Rooms, Center of Medical Genetics, Antwerpen

3.1.2 Patients with congenital hyperinsulinism

Molecular genetic analysis is performed in the institute to diagnose patients with diffuse and focal forms of congenital hyperinsulinism. The diffuse form is predominantly caused by autosomal recessive inheritance, with mutations in *ABCC8* and *KCNJ11* genes. A screen for mutations in the entire coding sequence of the mentioned genes is performed in a routine setting, allowing assigning a definite diagnosis in part of the patients. The underlying defect uses to remain unresolved in approximately 50% of the cases. This low diagnostic rate is most likely due to missed mutations in the known genes and further locus heterogeneity. Here, we built a cohort of 40 patients with diffuse congenital hyperinsulinism and inconspicuous mutation screening of *ABCC8* and *KCNJ11* genes. 23 of these patients had been screened for indels, cryptic splice sites and promotor mutations in the entire genomic region of *ABCC8* by deep sequencing on an Illumina GAII system. Furthermore, the exonic regions of the *ABCC8* gene had been checked for deletions by quantitative real-time PCR in all 40 patients. 8 cases missed a mutation in one allele and no mutations were found in 32 cases.

The aim of this study is to search for other CHI loci in these unresolved cases by genome-wide analysis of copy-number variations. Subjects were recruited from fifteen centers over Germany and ascertained by the following physicians;

Dr. Klaus Mohnike, Uni-Kinderklinik, Magdeburg

Dr. Oliver Blankenstein and Dr. Kühn, Charite Medical University, Berlin

Dr. Wolfgang Rabl, Klinikum rechts der Isar, Munich

Dr. Karl Schunck, Vivantes, Klinikim im Frierichshain, Berlin

Dr. Sabine Scholz, Kinderklinik, Hanau

Dr. Rolf- Dieter Stenger, Klinik für Kinder- und Jugendmedizin, Greifwald

Dr. Oliver Scheck, Klinik für Kinder- und Jugendmedizin, Erfurt

Dr. Sabine Heger, Kinderkrankenhaus auf der Blut, Hannover

Dr. Peter Müller and Dr. Johannes Weigel, Klinik für Kinder- und Jugendmedizin, Leipzig

Dr. Martin Lindner, Klinik für Kinder- und Jugendmedizin, Heidelberg

Dr. Jan Däbritz, für Kinder- und Jugendmedizin, Münster

Dr. Sibylle Linckersdorff, Josefinum Krankenhaus, Ausburg

Dr. Kai Rieckeheer, Klinikum Stuttgart Bürgerhospital M3, Stuttgart

Dr. Kaspers, Klinikum Leverkusen, Endocrinilogy and diabetology, Leverkusen

Dr. U Thier, Krankenhaus Dritter Orden, Munich

3.1.3 Patients with Holt-Oram-like syndrome

In addition to the MR and CHI cohort, blood samples from an extensive german family with several members suffering from an atypical form of Holt-Oram syndrome was sent to us for CNV genome-wide analysis. Affected individuals from the reported five generations pedigree presented hand anomalies, heart defect and additional mammary glands. The phenotype except the presence of accessory mammary glands resembled HOS syndrome.

No consanguinity was self-reported and the individuals were of Caucasian descent. Linkage analysis identified an interval on 12p24.21 with a LOD score of 3.1, encompassing the *TBX5* gene. *TBX5* mutation analysis by direct sequencing was reported to be negative.

All subjects provided written informed consent after being informed about the study. Blood samples were collected from 5 affected and 6 healthy individuals. The study was approved from the ethics committee of the referring institute:

Dr. Maja Hempel, Institute of Human Genetics, Kinderkrankenhaus, Munich

3.2 MATERIALS AND INSTRUMENTS

All chemicals used in this work had a degree of purity suitable *pro analyse*. Other chemicals, buffers, kits and additional materials are described under the corresponding methods.

3.2.1. Chemicals and kits

NaCl	Merk
Tris-HCl	Merk
Agarose	Biozym
Bromphenol blue	Merk
EDTA	Sigma
Ethanol	Merk
Isopropanol	Merk
dNTPs	Takara
1 kb DNA ladder	Gibco BRL
Mineral oil	Sigma
HPLC- H_2O	Merk
HhaI (20U)	New England Biolabs
HpaII (10U)	New England Biolabs
Hi-Di Formamide	Applied Biosystems
GeneScan-500LIZ Size Standard	Applied Biosystems
Ethidium bromide	Eurobio
Thermo-Start DNA Polymerase	Qiagen

3.2.2 Laboratory kits

Promega PCR Master Mix	Promega
GenomiPhi DNA Amplification kit	GE Healthcare
Infinium HD Assay	Illumina
Power SYBR® Green Master Mix	Applied Biosystems
Ampli Taq Gold PCR kit	Applied Biosystems
Big Dye Terminator v3.1 Cycle Sequencing Kit	Applied Biosystems
AmpFLSTR Identifiler PCR Amplification Kit	Applied Biosystems
iPLEX Gold Assay	Sequenom
LC Green PLUS	BioChem
NucleoFast 96 PCR clean up kit	Macherey-Nagel
QIAquick Gel Extraction Kit	Qiagen

3.2.3 Instruments

NanoDrop Spectrophotometer ND-1000 V.3.1.2	NanoDrop Technologies
ABI Prism 3730 DNA Sequencer	Applied Biosystems
DNA Engine Tetrad PTC-225	MJ Research
GeneAmp PCR system 9700	Applied Biosystems
MassARRAY system	Sequenom
LightScanner HR I 384 Instrument	Idaho Technologies
7900HT real-time PCR system	Applied Biosystems
Illumina Bead Array reader	Illumina
Hybridization Oven	Illumina
Vacuum manifold	Millipore
Microtube Centrifuge	Hettich
Plate Centrifuge	Sigma
Vortex REAX	Heidolf
Termo block	Eppendorf
Orbital shaker	

3.2.4 Computer programs

Sequence Detection System Version 2.3 software	Applied Biosystems
GeneMapper software	Applied Biosystems
Genome Studio software	Illumina
AssayDesign 3.1.2.2 software	Sequenom
SpectroTYPER 3.4 software	Sequenom
Hi-Res Melting software	Idaho Technologies
Call-IT software	Idaho Technologies
Staden Package	MCR Laboratory of Molecular Bioloy, Cambridge, UK

3.2.5 Databases and Internet-base Tools

Database of Genomic Variants	http://projects.tcag.ca/variaton
DECIPHER database	http://www.ac.uk/PostGenomics/decipher
Human BLAT Search	http://genome.ucsc.edu/cgi-bin/hgBlat
In Silico PCR	http://genome.ucsc.edu/cgi-bin/hgPcr?hgsid=162823421
Exon Primer	http://ihg.gsf.de/ihg/ExonPrimer.html
Primer 3	http://frodo.wi.mit.edu/
GeneBank	http://www.ncbi.nlm.nih.gov/Genbank
NCBI	http://www.ncbi.nlm.nih.gov
Repeat Masker at EMBL	http://www.woody.embl-heidelberg.de

OMIM	http://www.ncbi.nlm.nih.gov/OMIM
Pub Med	http://www.ncbi.nlm.nih.gov/sites/entrez
UCSC Genome Browser	http://genome.UCSC.edu/
Ensembl Genome Browser	http://www.ensembl.org/
dbSNP database	http://www.ncbi.nlm.nih.gov/SNP/
SNPper tool	http://snpper.chip.org/
Statistics (χ^2 test)	http://www.graphpad.com/quickcalcs/chisquared1.cfm

3.2.6 Tables of primers

Table 3-1: ***FOXP1* Idaho Primers**

ID	Forward primer (5' → 3')	Reverse Primer (5' → 3')	Size	°C
8797	CTTGGAAATCCTTGTATCAGG	GATCGCGATTAAGTGTAAAATTC	289	62
8798	GTCAGGACCCAGTCCCAAAG	TGCTCAACACAATCCACTCC	287	61
8799	GGCTGCGTGCTTCTGATTTC	GATTGCGAATGGAGTGATGG	257	61
8800	CAAAAGGGACAGTGGATTGG	AGCCACTAGATAGTCCTCTG	210	62
8801	CCATACTTTTAACCTTCCTTGGG	AAGAGACAACCCACCACCTC	283	62
8802/8825	GCTGTTAATTGATTTAATTATCC	TGTTAAAGGCAGTTTTGGACC	287	62
8803	CTCGTCTGGTGGAATGGTTC	TGAGCAAAGCATGAACGGTG	239	61
8804	GACCCGCTGCCTAGTTTATG	CTACAGAAATCTGGAATTTGAG	245	61
8805	TTTTGGTGTTAAGAGCCATGC	GGAAGTAGAAAAGGAATACTG	209	61
8806/8826	CTGGTTTTCTGAAGGTCTTC	GTCTACAAGAATTGCAGATTAC	302	61
8807	TTGTAGACCAGCCCCTTTTG	AAAGGTTTCAGCATGCTTGC	252	61
8808	GTGGAATGACGACTCTGCTG	GCACAACTGCATTTTATCAAC	241	61
8809	ACTGTTAGGCTGAATTTACTG	CATCTAGCAGCCAAAGCCTC	285	61
8810	AGCTAGGAACTCAGCAGTCG	TGAGTAGGGGAGACCTGTGC	186	61
8811	GTAGTTTCCTTGTCTTACATAC	TGTACCTAGGGAGACTAGGG	283	62
8812	TTTTCAGTGCACCTGAGCTG	AACAATTTCACTGCTAACTTTTG	296	62

Table 3-2: **X-Chromosome Inactivation Primers**

ID	Forward primer (5' → 3')	Reverse Primer (5' → 3')	Size	°C
8774	TCCAGAATCTGTTCCAGAGCGTGC	GCTGTGAAGGTTGCTGTTCCTCAT	287	58

Table 3-3: ***COH1* sequencing Primers**

ID	Forward primer (5' → 3')	Reverse Primer (5' → 3')	Size	°C
12448	AGGCAGAGGCCATACCTAGC	TCATTCAATTCCTTCTCTGGG	354	60
12449	GTAGAGTGGCAGCCGTGG	GCTTAGGAAACAATAGATTACCGTC	805	60
12450	GGCCTAACAAGGTATTTGATGC	GCCTGGCAACAGAGCAAG	557	60

12451	TCTCCTGTAGCTACCATAAACTGC	TCAAGTAATCCACCCGCC	331	60
12452	GCTAAATAATAACCACCTTTCTCATAC	TCAGTTGGCCCAAAGAAG	359	60
12453	TTCATTGCATATGGTCTGGC	CCTCTATCTACTTGAAGAGAGAAACAC	616	60
12454	TGTTTGGAGCATTTCTTTTATG	TCCCAATGGTAAAACGTAAGAATAG	356	60
12455	TGAGAGTCATAATGCTAAATTATTGG	AAGATTTCAAGTGGAGGAGCC	607	60
12456	CTGAATCTTAAAATTGGCCTTG	ATATGGTTAAACACTTCAGTAAGCAC	687	60
12457	TGGGCATCATAAAATCAAGG	CAATGTGTCCACCCAAACAC	277	60
12458	TTTTGGTCATCCTTTGTGTTG	TTGATTCTGAGGAAGGCTTTG	278	60
12459	TGTTTTAAGGCATTAAGCTACAAG	AAACGTAAAGTTAGCAAGACAACAC	393	60
12460	CAGGATTTGACCTTATAGTTTGGAG	ACACACTCAAGCCACTGTGC	383	60
12461	GCTTGCATAGAGGGAACTGC	TTATCTGATGATCCCAAAGCC	329	60
12462	TGCAGCTGTTGTAAAGCAAG	AAACTAGCATGGGCATAAGACAG	319	60
12463	CCTTTCTTCCCTTGCAACC	CGAGGAAATAAAACTGGATGG	460	60
12464	ATAATTAAGTTTAAATGCCTTGGTG	AAACGCAAGGCCAATTCTC	282	60
12465	GACATGTTCTGCATTCATTAGC	AGCACAACTGCAGAGTAAGCC	514	60
12466	GGTTTGCATTCCTCCACATC	ACGGAATTGTTTGAAAGGGG	400	60
12467	CCTTTGCTGCTTAAGAAATAGTG	CAGCATCCATGAAATTGTGG	298	60
12468	GAAACAATCTTGAAAATACGTTTGG	TTGTTACATGTGAGAAAAGTCTTACC	311	60
12469	TCTGGCGAAGATGTTAGGTG	TCAATGAGAAGACTGATTTGCTTATAC	418	60
12470	TTGTTTAGCACCGTGGTCTC	CCTGATCTTGATGGCCAAG	505	60
12471	AGTGGATTGAAACTGGAAGTTG	AACCATGATCTGGAAATCTGTG	333	60
12472	AGCCCCTTCACTGTAGCTTG	ATTAGCTGGGTGTCGTGGTG	781	60
12507	TCCAGCATGCATTTGTCAAT	AGTCTGGAGCCAAACTCTCAA	357	60
12473	TTTTAGTTCAGTGCAGTGTTGTG	CTCAAATTCTTGTAACCATTAGTCC	279	60
12474	AATACTCACTGAGGTCATTTTCTTC	CAAAGCACAACCAACTGCTC	611	60
12475	TTCCATTCCCTCAAAGATTTC	TGACCACTATATGTGTTTGCAGG	262	60
12476	TGCCCCAGTTTATTCATCAG	GGTTGGTTGGACATGCTAAAG	427	60
12477	CGATGACTGTAAAATATGCAAATAGG	CCTATGGGCTGAAACTGTGG	366	60
12478	ATTGTGTCAAGTGGCTGTGC	AAGAATATCAATCCAAAATGACTTATG	468	60
12479	AACATTGTTTATATGACACTTGGC	TTTAATCAATCTGGAAACTATGGG	833	60
12480	GCTTGCAAACAGTTATTTAACCAG	TTTTGCACGGAATGTCAAG	351	60
12481	AGCTTGTCAGAGAAGTAAAGGC	GAAATGTCCTCATCTCTAATGGC	618	60
12482	GCAAGACGACTCTGACAAAGG	GACAAATACCCTTTCTTACCACAC	430	60
12483	TTCAAGCAATAGAATTTCATCCTG	TATGATCAGCCCTTTCCCAC	933	60
12484	TCTGTACTTAGCCCTTTCTCTGC	AAGGAACAGGTTCTTAGTGAGGG	799	60
12485	ACTGCACCTGGCCAAACAC	AACCCCAGGATAACTTTGCC	518	60
12486	TTTTCACAAAATATTGGCTCATC	TGGTGCCCATCAAATTCTTC	539	60
12487	TGTAACAATCGCCACTGGG	TGCAGAAGAAATGAATCCCC	309	60
12488	TTTTCCAGCAATGAGACTTTC	TGCTCACATGCATGGTAGTAGTAG	373	60
12489	CCAAAGCAAAATTTCTTCTTGAC	ATTCCAGCTGTTTGATGCTG	1299	60
12490	GCTGGAATACTGCAACAAAGC	CATGCACACACCACTTATGG	1023	60
12491	GAGCATGGTGTGTTTGGAATC	CATAAACGCAACTTTACAATAAAATC	368	60
12492	TGGAGGGATATTTGGTTACCTTAG	AGACAGGTTAAGGGGATGGC	371	60
12493	TACTGTCCTGGCAGACAATTAG	CAGGGTTAGCTTTTCATTTGC	354	60

12494	TGCGTGATGGAGTGAGACTG	CCCTGCGACCATTGTATCTC	507	60
12495	TGTTCCAGAGGAGAGAGCATTC	GGCAACTGACAAAACCTCAG	680	60
12496	GGAGACATTTGAAGTCAAGCC	CCAAAAGCTTTAATACATTATGTGG	304	60
12497	CCAGACACAATGTACTGAAAGTTG	AAGCGTGTCATTTCTACCCG	1024	60
12498	ATGAATTGCAGGGAAAATGG	AAGGCCAAACAAAGCTAGAGG	384	60
12499	CTGAAACTACTGCCTTGGGG	GGGAGACCCTGCATCTCAG	303	60
12500	TGTTCCAGATGGAGCCTTTC	AGGGAACCCTTGGTATAAGC	321	60
12501	GCAGCATGAAACTGTTTTCC	GTACCAGTGACCAGCCTGC	1046	60
12502	TTTTGTTCTGGAACTCTCGTAAG	AGGCAAAGGTGGTGACTCAG	847	60
12503	CAGGGCCCTATATGTTCTTACC	AGTGAGCCATGACTGTTCCC	926	60
12504	AAACTTGGGGCTACAGCTTG	TCCTAATGACAAGCAGAGTATGG	913	60

Table 3-4: ***PTGER3*** **sequencing Primers**

ID	Forward primer (5' → 3')	Reverse Primer (5' → 3')	Size	°C
14695	CCCACCTCGGTTCTGTCC	CACGAAACCAGTGAGCAGC	506	60
14696	CTACACAGGCATGTGGGC	CACTGGACGGTGTACTGGC	549	60
14697	TCCTCGTTGTTCATCGCC	CTTAGGTTCCCGCTTACTCG	531	60
14698	CAAAGTACAGTGTAGAGGAGATGC	AGATAATGAGATAGGCTGCCC	315	60
14699	TATGCAGGTTTTGGTTTTGC	CGATTAGAACAGAGAGACGGG	229	60
14700	CATCTGTGGTTTAGCATCCTTC	GCAGAAAGTGAAGAGGTTTGG	167	60
14701	TTTGGTCAGTTTGGGAAGC	TTTCTCCACCATTTGAGGATAC	191	60
14702	AAGTGTCTCAGAATTACTTAGGATGG	GCTTGGAATTACAATGTTGACC	567	60
14703	TGCTGTAAGTGATTTACCATTCC	AACGATTTAGAGCTCCATGTTTC	491	60
14704	TGTTATCCAGAAGCTGTTATGTCC	TCTTTGACCTGGCACTTTCAC	414	60
14705	GGAGAAGAGGGGTGGAATG	GGGGCAGGGGTGGTATC	211	60
14706	CAAAGCCATGAGACTCTCCC	AAATATTGGAATTTCAGGGAACC	396	60
14707	TTTTGTTTCTGTGATGTGGCTC	GAGATAGGCTGACTTCCTGGG	282	60
14708	TCAGGGTGTCCCACACAAC	GCTTGGAATTACAATGTTGACC	598	60
14709	CCCCTACCAAAATCCCAG	AGCAGCAGTCTTGGCAATTC	659	60
14710	CCATGTGTTGATTGACTCGG	AAAGCCCCTTCTACTACATCACAC	478	60
14711	TCCTGGCTATTTCAGCTTCTATTC	TTTTGTCACGATTCCCTCC	799	60
14712	TTTCATGATAGCTTGCTTTTGC	GGGGACAGAAGAAAGGGAGG	702	60
14713	CCAACAACCAATTTATTGAAAACC	CTATCCCCGCTCCCCAATAG	684	60
14732	GGATAGGTTCCTTCACATTCTTG	GATGCTTGGTAGGCAGAACC	400	60
14733	TGATCAGAGATTGCTAAAGACTGC	GGCAGAGGGATGATGAGTTT	473	60
14734	GTAAACGCCGACCTCCG	TGGACAGGTACACGACGATG	441	60
14735	CTCTCCTCGTTGTTCATCGC	GAGGAAAGGACACTTCAGCG	602	60

Table 3-5: ***COH1*** **qPCR Primers (Patient 48737)**

ID	Forward primer (5' → 3')	Reverse Primer (5' → 3')	Size	°C
13254	TCATTGGTCAAATGTGCCTC	GTAAAACGCAAGGCCAATTC	125	60
13255	TGGAGTTTAGGCTCTGTTATTGG	CAGCTGCTGTTTGACTGTGG	130	60

15450	TCCAGTCTCTTCGGCCTTT	GCAGTGTACATAACAATAATGAAGCA	155	60
15457	AGGCAGGACTGACGTCTTTG	AGCATGTGGGGAGTTTCAGA	111	60
15458	GACCAAAAGATCTCAGATTGCT	AGCTTCAATCACTGCGGAAG	154	60
15480	TACTCAGAAGGCAGCCCAAA	GGCAGGGCCTGAACTCTAAT	169	60
15482	TGAAGAGGTGGCTCTTCTGG	TGACAGGTAGTAAATGGTCCAGC	122	60
15483	GATAGCAGTGCGGTGCATAG	GATTCAAAATTAATCCAGAAACAAAC	111	60
15484	AGATAGCTGGAATGCAACGG	TCAGTTGTTCAAGAAATGGCAC	120	60
15485	TGGTGACAATTTGAGTACGAAAG	GTCCACCCAAACACAGCC	167	60
15486	AGATTGGGTTTGTTGGTTGC	AAAGGATGGAGGAAACGGAC	113	60
15697	AGCTCACAGAAATGCAAGTTGA	ACCTCAACTGTAGTTCCTTCTTGTT	100	60
15698	CCTGCTCAGCATAAAGGTCA	GCTTTTTGTTGATGCATGGTT	167	60
15699	CGTGGAGGGACAGTTTGTTT	GCTCCCTCTGCACTCCAC	160	60
15700	GTGTTCAGGCGAAACTGTGA	TCAGTGAAGGGCTAGGGATG	163	60
15701	TCAGATACATCCATTCCAAAGAAA	TGAACACCTTTTGCAAAGCA	172	60

Table 3-6: ***COH1*** **qPCR (Patient 32140)**

ID	Forward primer (5' → 3')	Reverse Primer (5' → 3')	Size	°C
12332	CAAGCAAAACTACCAAAGACCC	TCAACAGTTGGCTGCATTTC	161	60
12333	CCACAAACCTGGACTTCTTCC	TTGGAACAAATGCAGTAAGCC	149	60
15417	TCCCTATTCCCTTTTATCCCC	AACCCATAGGTTCCACTAGGC	125	60
15418	GGTCTGGAACAAGATTCAGAAG	GCTGGTATCTGGAGGAGCTG	127	60
15419	TCCTCATCCCAGGGAGATTC	TCAGGATCTGGAGCAAAGAG	170	60
15420	TTGCAGAGGTTTGTATGGTCA	TTCATTATTCATTATTTACCTGCTTC	140	60
15421	TGCATAAAGTTATGTACAGATCCACAG	AGATCTGCCAAGGGGATTG	147	60
15422	TCCTGGATCAGAAATCGAAG	GCTGGATTTTGAGAATTCCTTTC	150	60
15423	GCGGCATCAAGAAAGGAGAG	CAATCCCTAACATACAAAGAGGAG	191	60
15424	TTTACTTTTGGTGAAAGAAATCTAGG	CAAGACAGACAGAAGGAAACCC	188	60
15425	GCTGATTAGAGATTATAAACAAACGAG	ACACCAGCATATTGCACAAC	271	60
15426	TGCCTAAAAGGATCTTCATGTG	TGTGCCATGTATTTCTTTGCC	273	60
15427	TGTTTACGTACAGTTTAATTTAGCCAC	TCCCTATTGATTCTCAAAACCAC	243	60
15428	TCAGGTCTGCTCCTTGCTTC	TCCCAGAGGCTGGTTAAATG	152	60
15429	GCTAGAAAGGCGGTATTAGGC	CAATCAATAAACCACCACGG	167	60
15430	ACATACAGCCAGCCAGGGTA	GAGGTTATGATTAGTGTAGTTGATGC	151	60
15431	TGCAGTGTGAAGAATGCTATG	TTTTGCTTATGTGCCTGTGC	126	60
15432	TTCATTCATTCCTCTGGGTACA	AAAAATTAAGGTTTTAGGAACAGAAGA	157	60
15433	CATGTAATACTCTGTGGGCTTTTT	GCCATAGTTCCCCATATTCAA	168	60
15434	TCAATTACTTTACAGTCATCTTACAGC	CCATAGGAGGGAAACTGGCTA	108	60
15435	GCCGAACTAAATTTATACTGATACTTT	CCTAATCTTTTATCTCAAGGTGCTG	150	60
15436	TGCCTCCCATTAATATAAACAGA	TGGTTATCACCTGAGACCATCT	156	60
15437	CCATTTCTTTGTTTATGTGGACAAT	AGCATTCCAAGGCACATAGG	115	60
15438	TCATTAAGAGTGCTGACCCTCTC	AAGTTTCAAAATTCAAAAGAAGGAA	101	60
15439	CATGATGGAGAGAAATGTATAATATGG	TCAAATTCCATAAAATGACCAATG	198	60
15440	CATCTTTTACCATGTACTTCCTTTGA	GGGAGGGAAAGCATTAGGAG	179	60

15441	GAAGAACTAAAATAGAGGCCAAAA	TGTGTGAGATATTAAATGCTTCCTG	195	60
15461	CACCATTTCCTCTGGGAATAA	CAGCTGAAATTTGCCCTTTT	136	60
15462	GGCCTTGAGTACACGCAAAT	TCAACTTGCCTGCTCTGTTG	142	60
15463	GAGATTTCAGGGAACCTTGG	AAATGCTATTATTTGCTTGTGTGG	108	60
15464	TTTCTATAGTAACCCTTCAAACAAAAA	ACAGTTTAGCTAGTTTTCCCCAGGA	100	60
15465	TCTCGTGCTGTGGTTTTCAG	AACTTCTCCGAGCAAAACGA	146	60
15466	TGATCCTTTTTCCATCTCCAA	ACATAAATTGGGCCACTGGA	131	60
15467	GCCACACTCACTGTTGCATT	AGACCACTGACCTCAACAAAAA	119	60
15468	TCTGAGGTTAGCTTACATTTCTTGTG	TGATTGTTTTTGAATGCAAATTGT	113	60
15469	GCAAGCCTCAGAGTAGACATCA	TGATGAGTTAATGGGTGCAG	142	60

Table 3-7: ***COH1*** **qPCR Primers (Patient 33147)**

ID	Forward primer (5' → 3')	Reverse Primer (5' → 3')	Size	°C
13251	TTGTCTCTCTCATTGTGGGAAG	CACCGGACTCAGCATCAAAG	159	60
13252	GCTAGGAATGTTTGGCTTGG	GGCAACCTGAAGGCCTATTC	183	60
13253	GCAAGGTGGCAGAATGTATTG	TCACCAGCAAAAGGAGTCAAC	161	60
13254	TCATTGGTCAAATGTGCCTC	GTAAAACGCAAGGCCAATTC	125	60
13255	TGGAGTTTAGGCTCTGTTATTGG	CAGCTGCTGTTTGACTGTGG	130	60
13256	CCCACCTAGATATACAAGCCCC	CCTATTTGTTATCCTTTTAACATCCC	153	60
15394	GGAGTTCTTAACAGCAGGGAAC	TTTCCCCATAGAGCAATTCAG	148	60
15395	TTGGCCAAAAGAAAGAGCAC	CCCTAAAGAAAACTTGGTTCTTG	124	60
15396	TTGAGCTCTTTTGTCATCTACTAGG	CAGCTACAAACTGCAGAGGC	129	60
15397	CATCATGTGGCTGGTGGTG	ATCTCCCTAACTGTTTTCTACTGC	217	60
15398	GCTTGCCTTGTTGTGTTTTG	AAGACAGCTATGGGCAAGAATATC	167	60
15399	TGGGCAAGAGGCTAAGAATC	CCCCAAACAGCATTATTTTC	163	60
15400	CATCATGTGGAACGATTTGC	CTAACCTGTCTTGGGAAGCG	191	60
15401	AATCCTCTGGGTAACCATTCC	GAATGGTTCAGGTTCTCTAGGC	164	60
15402	GGCACCATGTCTATCATCTGTG	TGTGCTTATTAAGGCGGACC	183	60
15403	TGAATGCAATCCACATGTCTG	CAAAACAGAACAGGGTAAGAAACC	126	60
15404	ATCTGAATCTGGTGGCATGG	CCAAAGTAGCTTATGAAACAGTAGAG	118	60
15405	AGAGCAGCCAAGATCCCTTT	ACCAACTGGCCCTTACAGGT	178	60
15406	CAGCTTTGTAGTGAGAGTTTGAATTT	TGGCCTTTACTTTTCATGCAC	187	60
15407	AAAGCCTCTTCTGAGCACATT	TGCAACATATAAACGATTGAGAGA	177	60
15408	GGGTAAATAAGACCCAGCACC	GCATACTATAGCACCTAGAAACCCC	134	60
15409	CCTTGTCTTCACTACTGTTGGAG	GCACAGGAAGAAGGAAGTGC	128	60
15410	TGTTTTCAGTGTTTATGGCACC	TGAAAAGGAAGCATACCACTCAC	117	60
15411	TGCAAGCCACCACTATTGAG	TGTGGGCTTTTAGAAGATGG	116	60
15412	TGTGTTCACCTTTCTTTCTGG	AAAATTCCTAAGTCTCCCAGGTG	123	60
15413	AAATTATACCCCACTCTTGCC	TCTGGCTAAGGTAACCAATCC	167	60
15414	TTTGTTTGAAAATTCATTGAACAC	GCAAACTGGACAAGATGCAAG	123	60
15415	CCCATTTTGCAATCATCTTG	GGGCCCAGAACCAGTATAAATC	165	60
15416	TCTTTTAGGTGCTGTACTTCTTTGC	TTGTTCGTTTCTGAGGCTGA	101	60

15450	TCCAGTCTCTTCGGCCTTT	GCAGTGTACATAACAATAATGAAGCA	155	60
15453	TTCTACTGTGACATAGGACTTTTGAA	CCTCAAAGGCTACTAAAATCAACC	150	60
15454	TGGGTTGATTTTAGTAGCCTTTG	AGCCAATACAGCCAAGGTTC	176	60
15455	TGATGGCTTTGTCAGTGTAAGAA	TCTTTCATCTGCATATTTTCCA	111	60
15456	TCTTCAGTTTTTATGTGTTTCAGTGTA	TTACTGCCATTATAAAACACATGG	188	60
15457	AGGCAGGACTGACGTCTTTG	AGCATGTGGGGAGTTTCAGA	111	60
15458	GACCAAAAGATCTCAGATTGCT	AGCTTCAATCACTGCGGAAG	154	60
15460	GAGAGGCATGTCTTCCAAGG	AATAAGGGGTACTTTAACAGAAAACA	144	60
15480	TACTCAGAAGGCAGCCCAAA	GGCAGGGCCTGAACTCTAAT	169	60
15482	TGAAGAGGTGGCTCTTCTGG	TGACAGGTAGTAAATGGTCCAGC	122	60
15483	GATAGCAGTGCGGTGCATAG	GATTCAAAATTAATCCAGAAACAAAC	111	60
15484	AGATAGCTGGAATGCAACGG	TCAGTTGTTCAAGAAATGGCAC	120	60
15485	TGGTGACAATTTGAGTACGAAAG	GTCCACCCAAACACAGCC	167	60
15486	AGATTGGGTTTGTTGGTTGC	AAAGGATGGAGGAAACGGAC	113	60
15697	AGCTCACAGAAATGCAAGTTGA	ACCTCAACTGTAGTTCCTTCTTGTT	100	60
15698	CCTGCTCAGCATAAAGGTCA	GCTTTTTGTTGATGCATGGTT	167	60

Table 3-8: ***PTGER3*** **qPCR Primers**

ID	Forward primer (5' → 3')	Reverse Primer (5' → 3')	Size	°C
14796	CTTTGCTGCAATGAGATGGA	GACAGCAAATCTCCCCCATA	152	60
14797	GCTGGGCTGACTTTAATTGC	CATCATTTCAAGGGGAGCTT	156	60
14798	TGCTTTTCTAGTGCGTGGAA	TGCTGGTTTCATTTCCCTTC	145	60
14799	CAAAGAAGCTAGACTTTAAACCCC	AAGCTCCCCTTGAAATGATG	156	60
14800	TGCAGACTTCAAGCATCAAAG	TGACTTTACCCTGGATGTGG	160	60
14801	AAACCACATCCAGGGTAAAGTC	TTCCAGTGTAACGCTTGGC	124	60
14802	GACGCCAAGCGTTACACTG	TCCTGTCCATCAAAAGTCTGTG	189	60
14803	TGAGGATATGAAAGTTACTAACGGC	AGTGCACAATTGGGAGTACG	128	60
14804	CAAAACCTGGAAGACTTGGC	CAGCAATGTTTTGTGTCATGG	114	60
14805	GCAGTCAAGCTGAGTGGGAG	AAGGGCTGCAATTTTGTATTC	195	60
14806	AACAAGCTGTAAAATGATCACCC	CAACTGAAAACCCATAATGCAG	159	60
15476	GTCCAGTTTCTCCCGCTGTA	CCAGCTGAAACATCAAACTCC	152	60
15477	ACCTGAGGGCTCTCCATAGG	AGGTGGGAGACAGTGGTGAG	162	60
15478	CCCATGAGCCCTATAGCATT	TACATTATACAAGTTGAATCCTTGACC	150	60

Table 3-9: ***TBX5-TBX3*** **qPCR Primers**

ID	Forward primer (5' → 3')	Reverse Primer (5' → 3')	Size	°C
12043	ACTGCCAGGGCTTGTTAGTG	CAACAAACCATCTCACCTTCC	154	60
12044	CGCAATTAAAGATTGAGAATAATCC	GTGGGGAGGAGAAAGTTGAG	124	60
12045	CCATGGACAGGCTACCCTAC	GAACATTCCCTCTCCCAGC	116	60
12046	TATGTCTCCTCTTCCCTGCC	GGGAGTTTCAAGATGTCATTGG	101	60
12047	CCCTTTCAAAAGAGTTGTCTGC	GAAGTGGAGCTGGAAAGTGG	123	60
12048	AGTCCATGAGGGTGTTTGATG	AGGTTCGATGTCCCTACAGTG	136	60

12112	GATGGGCATGAGAAATCACG	TTTCCTGTTCTTTCTACCGACTG	108	60
12113	TGAAGAACCCTACCATAACACTTTAC	AAGAGGCAAGACACAGTCACC	141	60
12114	CATTGCACTTGAAAGAAGCTG	AAATATGCCACATCCCCTTG	157	60
12115	CATCCAGAGTCTGTCCTCTGC	TGCCTTCATTCAGTAAACAGG	132	60
12116	ACCATCCACCGAGAATTGTG	CGTTTTATAGGCGAATGTTTCC	136	60
12117	CTTCAAATGCAATTCCTGCC	CTACCCCTCCCCAACATTTC	125	60
12118	TTGGGAGGCAGCGTCAG	GAGAGATCCGGTCATTCCTG	132	60
12119	GAAGTGAAACTGGACTTCAGTGAG	GGGGAATGTGTTAGCAGCTC	106	60
12194	CCCTTCCTCCTCAAGGTAGAC	TTTGGGATGTCCCAGTTTTC	150	60
12195	TTTAAGATGATGCCTGTGGG	GCTGCTGCTTCTTGTCTTTC	212	60
12196	TCAACACTGGAGGGACATCAG	TTCCTGAAACTGTGAGTGTTGG	105	60
12197	AAAGAAAAGGTGAGGTGCTG	ACTTTGGACCTTGGTGGC	83	60
12198	CTGAACGTCGGTTTCAGAAG	AAGACCTAGGGGCTGGAGAG	127	60
12199	CCGGTTCTCTCTCCAGCTC	CCGGCTTTTGTAGAACCGAG	108	60
12200	TTACCAATCGCCGAATTACC	AAGCAAGGATGGTCCAAGAC	115	60
12201	TACTTGGGGTGGAAGCAAAG	CTCCCATTTCCAAAGCTGAC	142	60
12223	GAAAGATAAATGCCTCCCACC	CTTGCGTGACTGCCATTG	102	60
12224	CTCTGCAAAACATCTCCGC	GGCTGGCTGATCTAAGTTGG	127	60
12225	CCTGAGTGCTGGAAAGATGG	TCTATCTTGGGGCTCACACTC	130	60
12226	GGGGATGAGCATGAAGAGAC	CAGTGGTCTCCATCAGGAGG	134	60
13553	TCTGCCCAGGTAGTCACCTC	GGGGTAAATCTTGAGTAACCTTCA	117	60
13554	AATCACATTCCACCCCTTGA	CACCCACAAAGACGACAATG	174	60
13555	TGGCCCCAAAACACTACATT	ATCGTGGTAAACCCAACAGC	109	60
13556	TGACGTTTAGCCATTCATTCTT	GCAGGTCTTTTAATCTAGGCAGA	150	60
13558	ATGTCTTGGACCATCCTTGC	TTTCTCCGTTCCCTTTCCAT	103	60
13560	GGGAAAATTACGAGGTGGTG	CCCCAGAGACTCATGAAACC	154	60
13561	CCTCCCCCAGACTGCTC	AGGGCAGGCAGGCAAA	98	60

Table 3-10: ***FOXP1* Sequenom Primers**

ID	Forward primer (5' → 3')	Reverse Primer (5' → 3')
FOXP1_1515CT	ACGTTGGATGTGGTTCACACGAATGTTTGC	ACGTTGGATGCAAGAATAAACAGGCCTGAG
FOXP1_975-14AG	ACGTTGGATGGCATGCTCACTGTTGAGATG	ACGTTGGATGCTTTTTGTTTCTCCATTTCAC
FOXP1_1188GA	ACGTTGGATGTCTGGTATCAAGTGTCACTC	ACGTTGGATGCGTTGGAGTATGAGGTAAGC
FOXP1_664+11AG	ACGTTGGATGCACCTCCACCCAAAGTCTTC	ACGTTGGATGCCTTCCCCTTCAACCTCTTG
FOXP1_181-29GA	ACGTTGGATGGGAATCCTTTTTTATGAAGAG	ACGTTGGATGAAGTGCCTGGAAGGAAAAAC
FOXP1_301AG	ACGTTGGATGTGGAGGATCTGCTGCATTTG	ACGTTGGATGACATTCTAGGTTCCCGTGTC
FOXP1_664+6CT	ACGTTGGATGCACCTCCACCCAAAGTCTTC	ACGTTGGATGCCTTCCCCTTCAACCTCTTG
FOXP1_768GA	ACGTTGGATGACAGGCAACAATCACAGCAG	ACGTTGGATGTAAGGAGGTCTTGGAAGGTG
FOXP1_1-5GA	ACGTTGGATGTCTCAGTCCCAGATTCTTGC	ACGTTGGATGCCCCATCTTGGAAATCCTTG
FOXP1_c181_30C_T	ACGTTGGATGGCACTTCCTGGAATCCTTTT	ACGTTGGATGAAGTGCCTGGAAGGAAAAAC
FOXP1_c180_149T_C	ACGTTGGATGTTCTGTCTGTCAGTGGCATC	ACGTTGGATGGTCTCTTAAGCCAACTGCTG
FOXP1_c1546A_G	ACGTTGGATGACTGCCCCTTTAACGTTTTC	ACGTTGGATGACTTGCTGCAGAATGCAGTG
FOXP1_c1168A_T	ACGTTGGATGTCCTTGCAGTTGAATCTGG	ACGTTGGATGGTATGAGGTAAGCTCTGTGG

FOXP1_c226_228dupCAG	ACGTTGGATGAGCTGGTTGTTTGTCATTCC	ACGTTGGATGCTCCTTCTTCAGCAGCAACA
FOXP1_1709AG	ACGTTGGATGAAACATGCAGAGCAGCCACG	ACGTTGGATGACGGCCGTTAATCTTACCTG
FOXP1_781TC	ACGTTGGATGTTGGATCTGACCACGACATG	ACGTTGGATGTGAGAGCTGTCCATTGGTAG
FOXP1_13TC	ACGTTGGATGCCCCATCTTGGAAATCCTTG	ACGTTGGATGCTGGATGGCTGAACCGTTAC
FOXP1_643CG	ACGTTGGATGCCAAGGCCTTCTGACAATTC	ACGTTGGATGACGCATACCTTGAGCAAGAG
FOXP1_1790AC	ACGTTGGATGGTTCAGCTCTTCCCGTATTG	ACGTTGGATGCTCTATACACTACCGCTTCC
FOXP1_A115_643CG	ACGTTGGATGGCCTTCTGTAAAATCAACGC	ACGTTGGATGCCTTCTGACAATTCAGCCCG
FOXP1_A50_1709AG	ACGTTGGATGAAACATGCAGAGCAGCCACG	ACGTTGGATGATTCGGGACGGCCGTTAATC
FOXP1_B146_664+6CT	ACGTTGGATGCTTCCCCTTCAACCTCTTGC	ACGTTGGATGCACCTCCACCCAAAGTCTTC
FOXP1_29949_768GA	ACGTTGGATGAGAAACCACAGGCAACAATC	ACGTTGGATGAAGGAGGTCTTGGAAGGTGC
FOXP1_A11_1546AG	ACGTTGGATGTACTTGCTGCAGAATGCAGT	ACGTTGGATGACTGCCCCTTTAACGTTTTC

ID	Unextended Primer_SEQ
FOXP1_1515CT	GCTTACTTCCGACGCAA
FOXP1_975-14AG	GATGTCTGCAACAATACA
FOXP1_1188GA	CTCTCTCCAAGTCCGCATC
FOXP1_664+11AG	AGCTATGCCTTCTGTAAAA
FOXP1_181-29GA	CCCTGAACCTGTGTAGTAAC
FOXP1_301AG	GATAACTTGAGGTGTCATCA
FOXP1_664+6CT	CTATGCCTTCTGTAAAATCAAC
FOXP1_768GA	ACAGCAGTTTGGATCTGACCAC
FOXP1_1-5GA	GTCCCAGATTCTTGCATCATGACT
FOXP1_c181_30C_T	CCTGAACCTGTGTAGTAA
FOXP1_c180_149T_C	GTCAGTGGCATCTCATAAAC
FOXP1_c1546A_G	ACACTTGTGAAGACTAAGAT
FOXP1_c1168A_T	GTTGAATCTGGTATCAAGTGTC
FOXP1_c226_228dupCAG	AATCCACTAACTTGCTGCTGCTG
FOXP1_1709AG	TACTGCACACCTCTCAA
FOXP1_781TC	CCACGACATGTGTCTCC
FOXP1_13TC	TGAGTCATGATGCAAGAA
FOXP1_643CG	GCCCGGGCAGCCTGCCCTT
FOXP1_1790AC	CCCGTATTGCGCTGGCTAAG
FOXP1_A115_643CG	GCAAGAGGTTGAAGGG
FOXP1_A50_1709AG	CTACTGCACACCTCTCAA
FOXP1_B146_664+6CT	CAACCTCTTGCTCAAGGTATG
FOXP1_29949_768GA	aCACAGCAGTTTGGATCTGACCA
FOXP1_A11_1546AG	CTTGCTGCAGAATGCAGTGCGTCA

Table 3-11: ***COH1*** **Sequenom Primers**

ID	Forward primer (5' → 3')	Reverse Primer (5' → 3')
NM_017890_3_c11827_11828insATG	ACGTTGGATGTGTTGCTCTGACAGTCTCTC	ACGTTGGATGTAAGGGTCTGGCAACTAATC
NM_017890_3_c5086C_T	ACGTTGGATGCTGGAGAAACTACTTTGGTG	ACGTTGGATGTTTGACCCCCGTTTTGACAG
NM_017890_3_c11505delA	ACGTTGGATGGGTATTTTACATGGAGCTGG	ACGTTGGATGCTGGTCAGCATGTAGATCAC
NM_017890_3_c3866C_G	ACGTTGGATGCCTTCATTTTGTAACAGGA	ACGTTGGATGAGCCCATCTGCTGACATTG

ID	Unextended Primer_SEQ
NM_017890_3_c11827_11828insATG	CTCTCTCGGACTCCATC
NM_017890_3_c5086C_T	AGGTGCTCCAGTTATTC
NM_017890_3_c11505delA	GGACTTTCTCAGCTTCCCAA
NM_017890_3_c3866C_G	ATTTTCAAAAACACTTACCTCA

Table 3-12: ***PTGER3* Sequenom Primers**

ID	Forward primer (5' → 3')	Reverse Primer (5' → 3')
NM_198718_c1185delC	ACGTTGGATGTAATGCTGCTCACGAGTACC	ACGTTGGATGAACTGCACATTGTGGGCAAG
ID	Unextended Primer_SEQ	
NM_198718_c1185delC	CTCTCTCGGACTCCATC	

Table 3-13: **Primers for amplification and sequencing of *FOXP1* deletion junction fragments**

Patient	ID	Forward primer (5' → 3')	Reverse Primer (5' → 3')	°C
1	B35_FOXP1F/B8_FOXP1R	ATGCTGAAGGTGGAATGGG	GGCCACATACGTGTTGTCAG	60
2	O06_for/Z02_rev	CGTTGCCAGCTCAAGGTTAT	TAAGTGTGTGCGAAGCCAAG	60
3	bp_FOXP1_3for/rev	GCACCTGACCCTCTAGCTCA	GGTTCAGCCACTGGTCTTTC	60

qPCR Primers for CNV Validation

In addition to the qPCR primers listed before, we designed 1313 CNV-specific primer sets for qPCR experiments. qPCR assays were used to validate CNV data obtained from microarray analysis, to map closer the deletion/duplication breaking points, and to determine the CNV inheritance. Primer sequences, melting temperature of the primers and annealing amplified product sizes are available upon request.

3.3 METHODS

3.3.1 DNA extraction from blood

DNA was extracted from EDTA-whole blood following a modified protocol from Miller *et al.* from peripheral blood leukocytes (Miller et al. 1988). 5-10 ml EDTA-Whole blood was mixed with 40-45 ml lysis buffer. Erythrocytes were lysed and separated from the leukocytes after centrifuging at 300 xg for 10 min at 10°C. 5 ml SE buffer, 25 µl pronase E (20 mg/ml) and 250 ml 20% SDS were added and the mixture was incubated overnight at 37 °C to allow nucleate cells to dissociate. Then 2-3 ml 5 M NaCl was added, gently vortexed and centrifuged at 1300 x g for 10 min at RT to achieve protein precipitation. The DNA-containing supernatant was transferred into a clean tube and two volumes of absolute ethanol were added for DNA precipitation. Precipitated DNA was washed with 70% ethanol, air dried and diluted in 100-1000 µl TE buffer.

Lysis buffer:	155 mM NH_4Cl, 10 mM $KHCO_3$, 0.1 mM EDTA up to 1 l dH_2O (pH 7.4)
SE buffer:	75 mM NaCl, 25 mM EDTA (pH 8)
TE buffer:	10 mM Tris-HCl (pH 8), 1 mM EDTA

3.3.2 Measurement of DNA concentration

The concentration of nucleic acids was determined with the NanoDrop Spectrophotometer ND-1000 V.3.1.2 by measuring the absorption of 1 µl of the samples at 260 nm. The quality of nucleic acids i.e. contamination with salt and proteins was controlled by measurements at 230, 280, and 320 nm (salt, proteins, and background respective absorption wavelength). The proper differential quotient E260 / E280 was around 1.8. Working aliquotes were prepared to the desired concentration for further downstream analysis.

3.3.3 Gel Electrophoresis of DNA

Analytical gels allow the separation and identification of nucleic acids on charge migration. In an electric field, DNA fragments migrate towards the anode due to their negatively charged phosphates. The migration

of DNA fragments is determined by size and conformation. The gels containing an appropriate agarose concentration were prepared depending on the size of the expected DNA fragments to be analyzed. To separate, identify, and/or purify DNA fragments ranging between 0.5 and 3 kb a concentration of 1-1.5% was used. For assessing integrity of the extracted genomic DNA a concentration of 0.6-0.8% was used. The DNA agarose (Biozym) was mixed with 1x TBE buffer in a glass vessel and boiled in a microwave. The gel was cooled to 55-60°C and 0.2 µg/ml ethidium bromide in TBE buffer was added. The agarose solution was poured into the gel tray to a thickness of approximately 6-7 mm. The DNA samples as well as the1 kb-ladder standard marker (New England Biolabs) were loaded into the gel wells. Electrophoresis was accomplished in a horizontal gel chamber at 1.5 V/cm for 30 min. The gel was viewed under a 300 nm UV light transilluminator and image capturing used a digital documentation system.

10x TBE-buffer:	840 mM Tris, 900 mM boric acid, 20 mM EDTA, dH_2O up to 1 l (pH 8)
Ethidium bromide:	10 mg/ml stock
Gel-loading buffer:	0.05% (w/v) orange G or bromophenol blue in 25% (w/v) Ficoll

3.3.4 Standard Polymerase Chain Reaction (PCR)

PCR is used to amplify exponentially a specific DNA fragment. The method relies on thermal cycling, consisting of cycles of repeated heating and cooling of the reaction for DNA melting and enzymatic replication of the DNA. A first step at 95°C allows the denaturation of the DNA, separating the two strands of the DNA double helix. A second step allows specific oligonucleotide primers to anneal to the targeted complementary DNA region. The temperature of the annealing step depends on the melting temperature of the primer, tipically being of 55-62°C. In a third step at 72°C the *Taq* DNA polymerase, a thermostable enzyme, assembles to the 3 prime of the primers and catalyses the synthesis of a new DNA molecule. This process is repeated 30-35 times, allowing the amplification of a specific DNA fragment across several orders of magnitude.

The polymerase chain reaction method (PCR) was carried out under the standard method described by Saiki (Saiki et al. 1988). Specific oligonucleotide primers for PCR were designed using the publicly

available Exon Primer (http://ihg.gsf.de/ihg/ExonPrimer.html/) and Primer3 (http://frodo.wi.mit.edu/) programs. Genomic sequences were gathered from the Santa Cruz Genome Browser (http://genome.UCSC.edu/). The use of Primer 3 required the previous masking of regions with low complexity structure or repetitive elements (www.woody.embl-heidelberg.de). To control for amplimer specificity a Blat search on all designed primers was perform using the program Human BLAT Search. An In-Silico PCR was performed to obtain the exact genomic position coordinates and to check further the PCR specificity. The two last electronic tools are integrated in the UCSC genome Browser. The primers were purchased from the company Metabion. Detailed descriptions of primer sequences, and annealing amplified product sizes are given in section 2.2.6.

The PCR reaction was carried out in a programmable thermal cycler (DNA Engine Tetrad PTC-225, MJ Research and GeneAmp PCR System 9700, Applied Bioystems) in 0.2 ml microcentrifuge tubes (Eppendorf) or in 96-well microtiter plates (ABgene). The standard reaction was prepared in a final volume of 25 µl using 50-100 ng genomic DNA as a starting amount of template. Negative controls were always used to control for contamination in the reaction mixture. The protocol of Promega PCR master mix was used as described below:

General reaction mixture for the PCR:

12.5 µl	Promega Master Mix 2X
1 µl	Forward primer (10 pmol/µl)
1 µl	Reverse primer (10 pmol/µl)
1 µl	Genomic DNA (50-100 ng / µl)
9.5 µl	dH_2O

(Promega master mix; 50 U / ml *Taq* DNA polymerase, 2x PCR buffer (pH 8.5), 400 mM of each dNTP and 3 mM $MgCl_2$)

General PCR conditions:

1st Initial Denaturation	94°C	2:00 min
2nd Denaturation	94°C	0:30 min
3 th Annealing	55-62°C	0:30 min
4th Elongation	72°C	0:30-2 min
5 th Additional Elongation	72°C	5:00 min

(The amplification stage comprises of 35 cycles from step 2 to 4. The annealing temperature of the primers depends on their nucleotidic composition and the elongation time depends on the size of the amplified DNA fragment)

3.3.5 Purification of PCR products

Before sequencing, PCR-amplified products that showed a single and distinct band in an agarose control gel were purified directly with the NucleoFast 96 ultrafiltration system, whereas PCR products that showed unspecific bands in an agarose control gel were purified from extraction of a new agarose gel using the QIAquick® Gel Extraction Kit.

3.3.5.1 Purification by ultrafiltration

PCR-amplified fragments were purified using the NucleoFast 96 PCR Clean-up ultrafiltration technology (Macherey-Nagel). During the procedure, the PCR samples are applied to an ultrafiltration membrane under vacuum to remove unincorporated nucleotides, enzymes and salt. The desired PCR products are retained on the membrane and can be recovered from the membrane after the addition of water. Briefly, 25-100 µl PCR products were loaded with a multichannel pipettor into the NucleoFast 96 plate. The plate was placed on a Millipore vacuum manifold and a vacuum of up to 600 mbar was applied for 10 minutes. Once the solution passed through the membrane the vacuum was released. The plate was taken out and immediately placed onto an absorbent pad to allow the resting liquid to drain off the outlets. After dispensing 25 µl DHPLC-H_2O (LiChrosolv, Merck) into the wells, the plate was positioned on an orbital shaker and incubated at 200 rpm for 10 minutes at room temperature. Then the sample volume was transferred by pipetting into a new PCR plate (Abgene). The plate was sealed with an adhesive film and stored at -4°C for further applications.

3.3.5.2 Purification by extraction of agarose gel electrophoresis

After separation of DNA on agarose gel, the DNA bands were excised with a scalpel blade and immediately purified using the QIAquick® Gel Extraction Kit. This method is designed to extract and purify DNA of 70 bp to 10 kb from standard agarose gels.

Three volumes of Buffer QG to 1 volume of gel (100 mg ~100 µl) were added to the tubes which contained the gel slices. Thereafter, the samples were incubated in the thermo-block at 50°C for 10 min at 1300 rpm

to allow the gel to dissolve (the samples were vortexed during the incubation every 2-3 min). Additionally, 1 gel volume of isopropanol was added to the samples when the DNA fragments were smaller than 500 bp or bigger than 4 kb. The samples were then applied to the QIAquick columns and centrifuged for 1 min at 12000 rpm. After, flow-through removal columns were placed in the same collection tubes. To remove all traces of agarose, 0.5 ml of Buffer QG was added to each column and centrifuged again for 1 min at 12000 rpm. To each column 0.75 ml of washing buffer PE was added. Samples were incubated at RT for 5 min and centrifuged for 1 min at 12000 rpm. The flow-through was discarded and the columns were dried by centrifugation at 12000 rpm for an additional minute. QIAquick columns were placed into clean 1.5 ml microcentrifuge tubes, incubated 5 min at RT with 25-30 µl of EB elution buffer and centrifuged for 1 min at 12000 rpm.

3.3.6 Automated DNA Sequencing

The automated DNA sequencing is based upon the dideoxynucleotide chain terminators method, developed by Sanger (Sanger et al. 1977). Sequencing was carried out in a 48 capillary Abi 3730 Genetic Analyzer (Applied Biosystems) after a termocycle-sequence reaction and using the BigDye®Ready Terminator v3.1 Sequencing Kit (Applied Biosystems). BigDye®Terminator Mix contains dNTPs, dye dideoxynucleotides terminators, $MgCl_2$, Tris-HCl, and modified *Taq* Polymerase.

The sequencing PCR reaction was prepared in a final volume of 5 µl containing; 1 µl of purified PCR product (200-300 ng), 1 µl primer (10 pmol/µl), 0.5 µl BigDye-Terminator mix v3.1, 1 µl buffer 5X and 1,5 µl dH_2O. Sequencing PCR conditions were as described bellow:

General sequencing PCR conditions

1st Initial Denaturation	95°C	1:00 min
2nd Denaturation	96°C	10:00 min
3 th Annealing	50°C	0:05 min
4th Elongation	60°C	4:00 min

(repetition of steps 2nd to 4th in 24 cycles)

After the sequencing reaction, the 5 µl volume sample was precipitated and redissolved into 25 µl DHPLC-H_2O. Briefly, 25 µl 100% ethanol was pipetted onto the PCR plate wells containing sequencing samples, mixed and incubated for 15 min at RT. The plate was centrifuged at 1300 rpm for 30 min at RT. The supernatant was discarded onto an absorvent pad and then 20 µl 70% ethanol was added, gently vortexed and centrifuged at 1300 rpm for 15 min at RT. After removing the supernant, the DNA pellet was air dried at RT and redissolved in 25-50 µl DHPLC-H_2O (LiChrosolv, Merck). The sequencing samples were transferred into a genetic analyzer plate (96 wells) and stored at 4ªC until sequencing. Sequencing data was analysed using the Staden Package.

3.3.7 X-Chomosome Inactivation

One male patient with one maternally inherited deletion on the X-chromosome might suffer from an inherited form of X-linked mental retardation. The inactivation pattern of the X-chromosome of the respective mother was analysed by the HUMARA assay, which detects methylation differences near a polymorphism in the X-linked androgen receptor (Allen et al. 1992). The digestion reaction was prepared in a final volume of 50 µl containing 1 µg genomic DNA (100ng/µl) and 10 Units methylation-sensitive restriction enzymes HhaI and HpaII (New England Biolabs).

Digestion reaction with HhaI		Digestion reaction with HpaII	
DNA(100ng/µl)	10 µl	DNA (100ng/µl)	10 µl
HhaI enzyme	1 µl	HpaII enzyme	2 µl
NEB4 Buffer (10X)	5 µl	NEB1 Buffer (10X)	5 µl
BSA (10X)	0.5 µl	dH_2O	33 µl
dH_2O	33.5 µl		

The mixture was incubated in an oven at 37ºC overnight. An 8 µg digestion aliquote was amplified in a termocycling reaction using a FAM labelled forward primer and the AmpliTaq Gold-PCR kit.

General reaction mixture for the AmpliTaq Gold-PCR:

AmpliTaq Gold Buffer (10x)	2.5 µl
dNTPs (2,5 mM)	1.25 µl
$MgCl_2$ (25 mM)	0.75 µl
Forward primer (pmol/µl)	1 µl
Reverse primer (pmol/µl)	1 µl
AmpliTaq Gold Polymerase	0.5 µl
Digested DNA	8 µl
dH_2O	10 µl

General PCR conditions:

1st Initial Denaturation	95°C	12:00 min
2nd Denaturation	95°C	0:20 min
3 th Annealing	58°C	0:20 min
4th Elongation	72°C	0:30 min
5 th Additional Elongation	72°C	10:00 min

(repetition of steps 2nd to 4th in 34 cycles)

For fragment analysis, 1 µl of the diluted PCR product (1:10 and 1:100) was mixed with 12 µl mix containing 100:1 Hi-Di Formamide / GeneScan-500LIZ Size Standard (Applied Biosystems) and dispensed into a genetic analyzer plate (96 wells). After a brief pulse centrifugation at 1200 rpm to collect at the bottom, the plate was placed into a preheated termocycler and incubated at 95ºC during 5 minutes to denaturate the DNA. Immediately after, the plate was placed on ice to fast cool down and stored at 4ºC. Fragments separation was performed in a ABI Prism 3730 DNA sequencer and analysis of the results used the GeneMapper Software (Applied Biosystems).

To asses for X chromosome inactivation we checked whether the mother showed skewed XCI pattern in the digested sample and whether or not the affected boy inherited the allele from the inactivated X-chromosome.

3.3.8 Paternity Test

We investigated the Probability of Paternity Exclusion using the AmpFLSTR Identifiler PCR Amplification Kit in one patient with two identified *de novo* CNVs.

AmpFLSTR Identifiler PCR Amplification Kit is a short tandem repeat (STR) multiplex assay that amplifies 15 tetranucleotide repeat loci (D8S1179, D21S11, D7S820, CSF1PO, D3S1358, TH01, D13S317, D16S539, D2S1338, D19S433, vWA, TPOX1, D18S51, D5S818, FGA microsatellites) and the Amelogenin gender determining marker in a single PCR amplification. The Identifier kit uses a five-dye fluorescent system for automated DNA fragment analysis. Reaction mixtures were prepared according to the protocols provided by the manufacturer (Applied Biosystems). Fragments separation was performed as described before, on an ABI 3730 sequencer and using the GeneMapper Software (Applied Biosystems).

General reaction mixture for the AmpFLSTR Identifiler PCR:

AmpFLSTR PCR reaction mix	9.2 µl
AmpFLSTR Identifier Primer Set	5 µl
AmpliTaq Gold Polymerase	0.8 µl
Genomic DNA(100ng/ µl)	1 µl
dH_2O	9 µl

(AmpFLSTR PCR reaction mix: $MgCl_2$, dNTPs, BSA, 0.05% sodium azide)

General conditions for the AmpFLSTR Identifiler PCR:

1st Initial Denaturation	95°C	11:00 min
2nd Denaturation	95°C	1:00 min
3 th Annealing	61°C	1:00 min
4th Elongation	72°C	1:00 min
5 th Additional Elongation	60°C	80:00 min

(repetition of steps 2nd to 4th in 34 cycles)

3.3.9 Illumina 550K and 610K arrays; protocol of hybridization

All patients were investigated using the Illumina high-throughput genotyping platform. The Infinium HD assay is based on direct hybridization of DNA samples onto SNP synthetic oligonucleotide microarrays. We studied 110 samples with HumanHap550 chips which offer genome-wide coverage of all human populations with up to 555,000 tagSNP markers down to 5.5kb median marker spacing. In addition to tag SNP content several auxiliary SNPs were included in the HumanHap550 chip to achieve the above mentioned even spacing across the genome. The rest of samples were studied with Human610-Quadv1_B chips, since the 550K array was transitioned to this new product. The 610K chip contains the same genetic variations of 550K plus 60K new markers targeting known CNV regions and SNP deserts (deCODE and DGV CNV content). Sequence-specific oligo probes of 50-mer and synthesized using Illumina's Oligator technology are coated to a 3-micron bead in the array. Each bead which contains hundreds of thousand copies of each oligonucleotide is assembled inside the optical fiber substrate micro-well on the array by electrostatic forces. Each array contains 30 copies per bead type which is in turn identified by a decoding system.

We first measured the DNA concentration with a NanoDrop spectrophotometer (ND-1000 V.3.1.2.) and assessed the sample integrity by means electrophoresis in a 0.8% agarose gel. After adjusting the DNA concentration to 50ng/µl we hybridized the DNA on the SNP-Chip. Briefly, 750 ng or 200 ng of total genomic DNA (for 550K or 610K chips respectively) is isothermally amplified in an overnight step. The amplified product is then fragmented and follow-up by alcohol precipitation. After resuspension, the DNA is applied to the Genotyping BeadChip and incubated overnight in the Illumina hybridization oven. The amplified and fragmented DNA samples anneal to locus-specific 50-mers during the hybridization step. After washing the non-hybridized and non-specifically hybridized DNA, an enzymatic base extension of the oligos, using the captured DNA as a template, is performed to confer allelic specificity. Afterwards, products are fluorescently stained. Scanning and image processing are performed with the Illumina BeadArray Reader and genotypes are called with the Illumina Genome Studio software.

Illumina microarray procedure (for 8 samples)

Day 1: genome amplification

Day 2: fragmentation, precipitation and array hybridization

Day 3: wash, staining, drying and scanning

Table 3-14: Reagents provided by the Illumina Infinium Assay Kit

Reagent	Description
MP1	neutralization solution
AMM	amplification master mix
FRG	fragmentation solution
PA1	precipitation buffer
RA1	resuspend, hybridization and wash buffer
PB1	hybridization buffer
PB2	humidifying buffer
XC1	stain solution 1
XC2	stain solution 2
XC3	stain solution 3
TEM	two-colour extension master mix
ATM	anti-stain two colour master mix
LTM	labeling two-colour master mix
XC4	stain and coating solution

3.3.9.1 Denaturation Step

A normalized concentration of 50 ng/µl of genomic DNA was used as a starting material for all reactions. Both reagents and DNA samples were well mixed by gentle vortexing and briefly centrifuged at 1250 rpm. 15 µl of DNA sample were dispensed to the MIDI plate into each designated well. To denaturate the DNA 15 µl 0.1N NaOH was added to each sample and mixed thoroughly by pippeting. The mixture was incubated for 10 minutes at RT. 270 µl of MP1 buffer was dispensed to neutralize the DNA and then 300 µl of AMM was added. This solution contains the enzymes necessary to amplify the DNA. The plate was sealed with a cap mat (ABgene) and inverted 10 times to mix contents. After a brief pulse centrifugation to 280 xg the plate was introduced into the Illumina hybridization oven and incubated for 20-24 hours at 37ºC.

After removing the plate from the oven, it followed a centrifugation to 50 xg for 1 minute. Then the samples were splitted into three additional wells, for a total of four wells per sample. Each well contained 150 µl of sample volume. The fragmentation buffer was thawed and tempered at RT. 50 µl FRG buffer was dispensed to each well. After sealing the plate with a cap mat the content was mixed by gentle vortexing at 1600 rpm for 1 minute. The plate was centrifuged at 50 xg for 1 minute and then incubated in a heat block for 1 hour at 37ºC to permit the enzymatic fragmentation of the DNA.

3.3.9.2 Precipitation Step

Before starting the precipitation step the heat block was preheated to 37ºC and the centrifuge set up to 4 ºC. The PA1 solution and the sample plate were pre warmed at RT, well mixed and briefly pulse centrifuged. 100 µl PA1 buffer were transferred to each well containing sample mixture. Then the plate was sealed, vortexed at 1600 rpm for 1 minute and centrifuged to 50 xg for 1 minute at RT. The plate was incubated on a heat block during 5 minutes at 37ºC and after centrifuging at 50 xg for 1 minute at RT, 300 µl 100% 2-propanol was dispensed to each well. The plate was sealed with a new, dry cap mat and it was inverted 10 times to mix contents. Afterwards, the sample plate was tempered for 30 minutes at 4ºC. Then it was centrifuged to 3000 xg for 30 minutes at 4ºC to precipitate DNA at the bottom of the well. Next, the supernatant was discarded by quickly inverting the plate and smacking it down onto an absorbent pad. The plate was tapped firmly on the pad several times over a period of 1 minute until all wells were completely devoid of liquid. The uncovered plate was left inverted on a tube rack for 1 hour at RT to air dry the DNA pellets.

3.3.9.3 Resuspension Step

Previous to resuspend the pellets, the oven was preheated to 48ºC and the heat sealer switched on 30 minutes before. The RA1 buffer was thawed to RT, vigorously mixed to re-dissolve the solution and dispensed into a modular reservoir. 44 µl RA1 was added to each well containing DNA pellet and the leftover reagent was reserved for the next day. The plate was heat-sealed with a foil seal and introduced into the Illumina hybridization oven. The incubation at 48ºC took 1 hour long. During this period the PB2 buffer for the hybridization step was thawed at RT and the heat block preheated at 95ªC. The plate was removed from the oven, vortexed at 1800 rpm for 1 minute and pulse centrifuged at 280 xg.

3.3.9.4 Hybridization Step

The rubber hybridization chamber gaskets were placed into the hybridization chambers and 200 µl PB2 buffer was dispensed into the humidifying buffer reservoir in each hybridization chamber. The chamber was sealed by closing and locking the chamber lid. The plate with the resuspended samples was incubated in a heat block at 95ºC for 20 minutes to denaturate the DNA. The plate was vortexed, centrifuged and allowed to cool to room temperature. Each sample that was splitted in 4 different wells was pooled back into the original well. A total volume of 150 µl was collected for each sample. Just before loading the DNA

samples, the bead chips were removed from their packages and placed onto the hybridization chamber insert. Using a precision pipettor, 83 µl DNA sample was dispensed in the appropriate section of the microarray inlet port. The hybridization chamber inserts containing the bead arrays were placed into the hybridization chamber. The lid was secured by closing down the clamps on the sides of the hybridization chamber. The hybridization chamber was placed into the preheated 48 ºC Illumina hybridization oven with the rocker adjusted to a standard speed of 5 and incubated for 16-24 hours.

The XC4 reagent that is used the next day during the staining step was prepared. 330 ml of EtOH 100% was added to the 310 ml XC4 bottle, vigorously mixed and the bottle was left on its side for 1 hour. Then it was mixed again for 15 seconds and left on the bench overnight.

3.3.9.5 Washing Step

The assay continued by washing the bead chips in WB1 and PB1 reagents to remove unhybridized and non-specifically hybridized DNA. For this, two wash dishes were prepared; one containing 200 ml WB1 and the other containing 200 ml PB1. The hybridization chamber was removed from the oven, opened and the microarrays were taken out. The coverseal from each array was pulled off and discarded. Bead chips were transferred to the slide rack and submerged in the WB1 wash dish. Using the wire handle, the rack with the bead arrays was moved up and down for 1 minute. Then the rack was placed immediately into the wash dish containing PB1, and moved up and down for 1 minute with gentle and slow agitation. To assemble the flow-through chamber, each bead chip was placed in a black frame into the multi-sample bead chip aligment fixture filled with 150 ml PB1. Afterwards a clear spacer was placed onto the top of each chip and an alignment bar was placed onto the alignment fixture. The clean glass back plates were placed on top of the spacer covering each array and secured with two metal clamps. Using scissors, the ends of the plastic spacers were trimmed from the ends of the flow-through chamber assembly. The arrays assembled into the flow-through chamber were immediately introduced in the chamber rack of the robot.

3.3.9.6 Single-base extension and Staining Step

The single-base extension, wash and staining steps were performed with the automated process. All reagents provided in the Illumina Infinium Kit, XC1, XC2, TEM, XC3, ATM and LTM, were thawed to RT, well mixed and centrifuged for 3 minutes at 3000 xg, checking that the tubes did not contain bubbles. PB1, RA1 and XC4 buffers were tempered to RT and mixed to dissolve any crystal that might be present. A fresh

95% formamide/1mM EDTA solution was prepared by mixing 15,25 ml 100% formamide, 30 µl EDTA 0.5 M and 720 µl dH_2O. The robot was prepared following the instructions provided in the automated protocol. The basic run parameter pane of the robot computer was set up with the corresponding number of arrays and the water circulator was switched on and set to 44 ºC. The chamber rack containing the bead chips, the quarter reservoir with 15 ml formamide/EDTA, the half reservoir with 10 ml RA1, the full reservoir with 50 ml XC3, and the XC1, XC2, TEM, ATM and LTM tubes into the robot tube rack were installed on the robot bed according to the bed map of the robot computer. After ensuring the correct position of the required items the process was started. Briefly, RA1 was used to wash away any remaining unhybridized and non-hybridized DNA sample. XC1 and XC2 were added to condition the chip surface. Then TEM reagent was dispensed to perform single-base extension of primers hybridized to DNA on the chip, thus incorporating labeled nucleotides into the extended primers. 95% formamide/1mM EDTA was added to remove the hybridized DNA. After neutralization using the XC3 reagent, the labelled extended primers underwent a multi-layer staining process with the resting reagents.

3.3.9.7 Coating and drying Step

When the robot finished, the flow-through chambers were removed from the chamber rack and placed horizontally on the bench. Two plastic wash dishes were prepared; one with 310 ml PB1 and the other with 310 ml XC4. The flow-through chamber was carefully disassembled and the bead array introduced into a staining rack. Using the rack handle, the rack was plunged in and out of the PB1 solution 10 times and left inside for 5 minutes. The rack was immediately removed and directly placed into the XC4 wash dish where it was moved up and down 10 times. The arrays were soaked on XC4 solution for an additional 5 minutes. Then the staining rack was placed on top of a tube rack. Carefully the arrays were taken and placed facing up on another tube rack. The tube rack with the arrays on top was placed in a vacuum desiccator to dry up by applying 508 mm Hg during 50-55 minutes. Once dry, the microarrays were stored in the dark until they were scanned.

3.3.9.8 Scanning Step

Bead chips were imaged using the Illumina Bead Array reader, a two channel 0.8 µm resolution confocal laser scanner. The decode data which comes with the microarray must first be loaded on the scanner computer. The chip was placed in the Bead Array reader after the barcode was scanned. The Illumina

BeadArrayTM Reader uses the laser to excite the fluorophore of the single-base extension product on the Chip and records high-resolution images of the light emitted. After scanning the bead array, the raw data was imported from the Illumina Genome Studio software.

3.3.10 Methods of normalization and CNV detection

Microarray data analysis was conducted in the Institute of Human Genetics at Helmholtz Research Center in Neuherberg by Dr. Tim Strom.

Illumina arrays use a one-color technique. In order to decide which dosis (copy number) is present in a single region of the genome, the normalized intensity values of a test chip have to be compared with one or multiple reference chips. However, due to quality differences of the DNAs or/and experimental variations, the signal intensities profiles between arrays showed to be different. Therefore, before CNV analysis the quality of the intensity profiles was checked and the arrays were subsequently grouped based on the raw intensity values. Genotype calls and raw intensity data were obtained from the Illumina Genome Studio software. Normalization and genotype-specific dosage calculation was performed according to Wagenstaller et al. 2007.

To account for experimental variations, we hierarchically clustered the euclidean distance matrix generated from the binary logarithm of the sum of median normalized intensity values of both alleles. The arrays were then ordered into groups with a similar intensity profile. Clusters with similar intensity profiles were analysed independently one to another in order to improve CNV calling.

The dosage values were calculated as follows. First the raw intensity values at each SNP locus was determined by separately calculating the mean of each allele. The raw intensity values were then median normalized. The log2 ratio of the intensity values were separately calculated for the three genotypes (AA, AB and BB) by dividing the normalized intensity values of the test array by the median values of all arrays with the same genotype for each SNP locus. One can show that the noise is lower when using genotype-specific intensity values as compared to using the sum of the intensity of both alleles for all genotypes. To make intensity values from male and female X-chromosomes comparable, the mean dosage of the male SNPs on the X-chromosome was adjusted to the mean dosage of the autosomal SNPs.

Two measures for assessment of data quality were used. First, the standard deviation (SD) and the median absolute deviation (MAD) of the $\log_2$ intensity ratios were calculated for all autosomal probes. The mean standard deviation (SD) was 0.15 for the 550K arrays and 0.19 for the 610K arrays. The mean median absolute deviation (MAD) was 0.11 and 0.12 for the 550K and 610K datasets, respectively. Second, a signal-to-noise ratio (SNR) was calculated in male DNA samples by subtracting the median log2 intensity ratio of the X-chromosomal SNPs from the median log2 ratio of the autosomal SNPs. This difference was then divided by half the sum of the MAD of the log2 intensity ratio of the autosomal and X-chromosomal SNPs. The obtained raw intensity profiles from 550K dataset had less variability in comparison to the ones from 610K. The mean SNR value was 5.44 for 550K arrays and 3.82 for 610K arrays.

$$\text{SNR} = \frac{\text{median (log2 (autosomal SNPs))} + \text{median (log2 (X-chromosomal SNPs))}}{0.5 * [\text{mad (log2(autosomal SNPs))} + \text{mad (log2(X-chromosomal SNPs)]}}$$

Segmentation was accomplished by applying two different algorithms. CNV analysis of the HumanHap550 arrays dataset was performed with an in-house developed Perl script (Wagenstaller et al. 2007) and the CNV analysis of the Human610-Quadv1_B chips-based data used the circular binary segmentation algorithms implemented in the R-package DNAcopy software (Olshen et al. 2004). In order to define a CNV, we used as a threshold or cutoff parameter the smoothing median of windows of more than five consecutive SNPs. CNVs with copy number values lower than 1.5 were suspicious for deletion and CNVs with copy number values of more than 2.5 were suspicious for duplication.

A Perl script and a relational database scheme were further implemented in order to automatically annotate the CNV outputs. In all datasets hg18 was used to map the segments to the human genome build. Conspicuous regions were compared with known CNVs as provided by the Database of Genomic Variants (http://projects.tcag.ca/variaton) and the DECIPHER database (http://www.ac.uk/PostGenomics/decipher).

3.3.11 Real Time Quantitative PCR

Real Time Quantitative PCR allows quantification of initial amounts of genomic DNA, cDNA, or RNA templates. Quantitative PCR is based on the detection of fluorescence of a reporter molecule that increases as the PCR product accumulates with each cycle of amplification. This method was used in the present study to genotype copy number variations and to map closer the deletion/duplication breaking points. Quantitative PCR (qPCR) is relatively easy to set up compared with other approaches such as MLPA and provides a high reliability when controlling for efficiency and specificity of the amplification. Therefore, qPCR was the method of choice for microarray data validation. A non-specific DNA binding dye such as Power SYBR® Green and a Taqman 9700HT instrument (Applied Biosystems) were employed. This dye emits relatively low fluorescence when being free in the solution, but its fluorescence increases more than thousand-fold when it binds into double-stranded DNA. The first cycle at which the amplified product generates sufficient fluorescent reading to overcome the background emission is called “Ct” or threshold cycle. The first significant increase in the amount of PCR product correlates therefore with the initial amount of target template. The main drawback of Real Time qPCR assays is the dramatic decrease in efficiency when amplifying products larger than 200 bp. To overcome this, the design of primers was limited to amplicons sized from 100 to 180 bp. One limitation confined to assays based on DNA-binding dye is the inherent non-specificity. A standard PCR and a subsequent agarose control gel were performed in order to check for unspecific amplification. Only primer pairs that amplified a single and distinct PCR product were used as detection templates. Monitoring of the PCR reaction and setting of baseline and threshold cycle values was accomplished automatically with the Sequence Detection System Version 2.3 Software (SDS 2.3, Applied Biosystems). To check further for non-specific product formation a dissociation curve analysis was performed using the SDS software.

Each assay included in duplicate a no-template control (NTC), 2 female and 2 male control DNAs, the patient DNA and the parent DNAs (when available) with a final concentration of 2,5 ng/µl in a total volume of 20 µl. 5 µl DNA template was transferred into a MicroAmp TM Optical 384-Well Reaction Plate (Applied Biosystems) and centrifuged to collect at the bottom of the wells. A master mix with the rest of the reagents was prepared in a 1.5 ml microcentrifuge tube (Eppendorf). After vortexing, 15 µl of the mixture was dispensed into each well containing sample using an electronic stepper pipettor (Rainin). The plate was sealed with a MicroAmp TM Optical Adhesive Film (Applied Biosystems), pulse centrifuged at

RT and stored at 4ºC until PCR amplification in a Taqman 9700HT instrument.

General reaction mixture for the Power SYBR Green PCR

10 µl	Power SYBR Green Master Mix
0.4 µl	Forward primer (10 pmol/µl)
0.4 µl	Reverse primer (10 pmol/µl)
5 µl	Genomic DNA (10 ng /µl)
4.2 µl	dH_2O

(SYBR Green Master Mix 2X: SYBR Green I Dye, AmpliTaq Gold DNA Polymerase, dNTPs, passive reference and optimized buffer components)

General Power SYBR Green PCR conditions

1st	50°C	2:00 min
2nd	95°C	10:00 min
(followed by 40 cycles)		
3 th	95°C	0:15 min
4th	58°C	1:00 min
(followed by 1 cycle; dissociation curve)		
5th	72°C	0:15 min
6 th	72°C	0:15 min
7 th	72°C	0:15 min

To control for differences in DNA concentration, reaction efficiency, and threshold cycles (C_t), the C_t values were normalized using the C_t value of a reference gene (*BNC1*) for each DNA sample. The relative quantification analysis based on the comparative C_t method was performed using an in-house developed Perl script.

Briefly, the comparative C_t method compares the Ct value of one target gene to the Ct value of the reference gene in the test sample. This value is then compared to the value obtained in the calibrator sample. It uses the following formula:

$$\frac{X_{test}}{X_{control}} = 2^{\Delta\Delta Ct} = 2^{(Ct,x - Ct,R)_{control} - (Ct,x - Ct,R)_{test}}$$

For the comparative Ct method to be valid, the efficiency of the target amplification and the efficiency of the reference amplification (endogenous control) must be approximately equal. Primer efficiency is defined as in equation:

$$\text{Efficiency} = 10^{\frac{-1}{slope}} - 1$$

where slope is the slope of the standard curve plot.

More than 1000 qPCR assays were used in the present study, thereby the run of a standard curve for both the endogeneous control primer set and the gene of interest primer set was unfeasible. To avoid to perform preliminary standard curve assays and since our study aimed to compare genomic dosage of two different loci (test and reference gene), identical amounts of starting material from the test sample and the calibrator sample were used in each assay. Additionally, information from two o more qPCR assays for each tested locus was obtained in order to validate a specific variant of copy number. The RQ values were expected to be 1 for normal genomic dosage (two copies), 1.5 for duplications (three copies) and 0.5 for deletions (one copy).

3.3.12 Whole genome amplification (WGA)

The whole genome amplification method was performed using the GenomiPhi DNA Amplification Kit (GE Healthcare). The kit enables to prepare sufficient DNA for genetic downstream analysis from biological materials of limited availability or just simplify simultaneous preparation of multiple samples.
We amplified the genomic DNA of 883 patients from the MRNET database in order to perform mutation analysis in the candidate gene *FOXP1*.
1 µl of genomic DNA was added to 9 µl sample buffer in a 0.2 ml microtiter plate (96 wells), well mixed by pipetting and pulse centrifugated. To denaturate the DNA, the plate was placed into a preheated thermocycler and incubated at 95°C for 3 minutes. Immediately after, the plate was placed on ice to fast

cool to 4°C. 9 µl mixture containing reaction buffer (with random hexamer primers) and 1 µl enzyme (bacteriophage Phi29 DNA polymerase) was dispensed into each sample. After mixing contents the plate was sealed with an adhesive film and incubated at 30°C for 1:30 hours. During the isothermal strand displacement reaction the double-stranded linear DNA template is exponentially amplified. DNA replication is extremely accurate because of the proofreading activity of Phi29 DNA polymerase.
An additional thermocycling step at 65°C for 10 minute was added to denaturate the enzyme. The concentration of the amplified DNA was roughly determined by running 1 µl aliquote of each sample onto a 0.8% agarose gel and comparing the band intensity respect to a control sample. The GenomiPhi-amplified DNA samples were used directly in downstream applications without purification.

3.3.13 Mutation screening with DNA melting curve analysis (Idaho platform)

The identification of a CNV affecting only a single gene allowed us to consider that particular gene as a candidate for the specific disease condition. We performed a mutation screening of *FOXP1* gene in a panel of 883 patients with ID using a high-throughput method such as Dye-binding/high-resolution DNA melting analysis. This method scans for sequence variations by using the melting curve of the PCR amplified fragment. A PCR product is amplified from a DNA fragment of the human diploid genome, thus generating DNA copies from both the maternal and paternal chromosome. If the paternal and maternal alleles show identical sequence, after applying a melting temperature, the PCR fragments will hybridize forming homoduplex hybrids. When a nucleic acid variant is present in one allele, in addition to the homoduplex, the DNA strands will hybridize forming a mismatched hybrid called heteroduplex. Therefore, mutations in PCR products can be detected by changes in the shape of the melting curve compared to a reference sample. A LightScanner HR I 384 instrument (Idaho Technology) was used to collect high-density information of fluorescence as a function of temperature in a mixture containing fluorescent double-strand DNA binding dye LCGreen PLUS (BioChem) and PCR product. LCGreen binds to dsDNA and has an optimum emission of fluorescence at 470-520 nm. Images of DNA melting (i.e. fluorescence) were captured by a CCD camera and results were evaluated with the Hi-Res Melting software and the Call-IT software (Idaho Technologies).
The assay was performed according to the recommendations of the manufacturer (Applied Biosystems). The primers were designed to flank all the coding exons and to generate amplicons of 300 bp of size. A

previous standard PCR and an analytical agarose gel were carried out for all 16 primer sets. A mother plate (96-well) was prepared with 300 µl of whole genome amplified DNA in a concentration of 2ng/µl. Every mother plate contained 94 patient samples and two negative controls. 5µl DNA sample from the mother plate was transferred in quadruplicates into a microtiter plates (384 wells); we amplified in duplicates two exons per plate. More than 12 daughter plates were prepared and left to dry up at RT during 4-5 days. Once dry, 5 µl PCR mixture containing LCGreen reagent and Thermo-Start DNA Polymerase were added into each well. After a pulse centrifugation, 8 µl mineral oil (Sigma) was dispensed into each tub to avoid evaporation of the sample during the termocycling reaction. After PCR amplification the plate was loaded into the LightScanner HR I 384 instrument for melting point analysis. Abnormal melting profiles were confirmed by direct sequencing of independent PCR products.

General reaction mixture for the LCGreen-Lightscanner PCR:

Thermo-Start High Performance Buffer (10x)	0.5 µl
dNTPs (2 mM)	0.5 µl
MgCl2 (25 mM)	0.5 µl
Forward Primer (0,4 µM)	0.2 µl
Reverse Primer (0,4 µM)	0.2 µl
LCGreen PLUS (1x)	0.25 µl
Thermo-Start DNA Polymerase (0,25 U)	0.05 µl
dH_2O	2.8 µl

General LCGreen-Lightscanner PCR conditions:

1st Initial Denaturation	95°C	15:00 min
2nd Denaturation	94°C	0:30 min
3 th Annealing	61-62°C	0:30 min
4th Elongation	72°C	0:30 min
5 th Final Denaturation	94°C	0:30 min
6th Hybridization	20°C	0:30 min

(repetition of steps 2nd to 4th in 40 cicles)

3.3.14 Variant Genotyping using Sequenom platform

Genotyping of selected variants was performed by Dr. Peter Lichtner in the Institute of Human Genetics at Helmholtz Research Center in Neuherberg using the iPLEX Gold-SNP Genotyping Assay (Sequenom genotyping platform).

This method relies on the MassARRAY system which permits to genotype multiple SNPs in a panel of hundreds of control samples by using multiplexed assays (36-plex). As a control group 676 unrelated subjects from the KORA-cohort (Cooperative Health Research in the Augsburg Region) was utilized. The assay consists of an initial locus-specific PCR reaction, followed by single base extension using mass-modified dideoxynucleotide terminators of an oligonucleotide primer which anneals immediately upstream of the polymorphic site of interest. The primer is extended, dependent upon the template sequence, resulting in an allele-specific difference in mass between extension products. Using MALDI-TOF mass spectrometry, the distinct mass of the extended primer identifies the SNP allele (Ross et al.1998).

The design of both PCR primers and MassEXTEND primers for multiplexed assays was performed using the AssayDesign 3.1.2.2 software. Genotype calling used the SpectroTYPER 3.4 software.

3.3.14.1 PCR amplification

The PCR conditions for the iPLEX assay have been optimized for amplification of multiplexed reactions.

Mixture for the PCR amplification:

PCR Buffer (10x)	0.625 µl
dNTPs (25 mM)	0.100 µl
MgCl2 (25 mM)	0.325 µl
Primers (500 nM each)	1.000 µl
Hotstar Taq (5 U/µl)	0.100 µl
Genomic DNA (5-10 ng/µl)	1.000 µl
dH_2O	1.850 µl

Conditions for the PCR amplification:

1st	Initial Denaturation	95°C	15:00 min
2nd	Denaturation	94°C	0:20 min

3 th Annealing	56°C	0:30 min
4th Elongation	72°C	1:00 min
5 th Additional Elongation	72°C	3:00 min

(repetition of steps 2nd to 4th in 44 cycles)

3.3.14.2 SAP treatment

Before the primer extension a SAP treatment was performed to dephosphorylate dNTPs that were not incorporated during the PCR reaction. Functional dNTPs can extend in the primer extension reaction causing contaminant peaks that greatly complicate data interpretation. It is essential that the PCR reaction and SAP are mixed and gently vortexed upon condition of the SAP.

Reaction mixture for the SAP treatment

SAP Buffer (10X)	0. 7 µl
SAP Enzyme (1 U/µl)	0.30 µl
PCR product	5.00 µl
dH_2O	7.00 µl

Conditions for SAP reaction:

1st	37 °C	40 min
2nd	85 °C	5 min
3rd	4 °C	10 min

3.3.14.3 iPLEX Reaction

The post- PCR primer extension reaction uses modified termination mix, DNA polymerase and cycling conditions. All reactions for the iPLEX assay are terminated after a single base extension.

Mixture for the iPLEX Reaction

iPLEX Buffer (10X)	0.200 µl
iPLEX Termination Mix	0.200 µl
Primer Mix (7 µM:14 µM)	0.804 µl
iPLEX Enzyme	0.041 µl

SAP mixture	7.000 µl
dH_2O	9.000 µl

iPLEX Reaction conditions:

1st	94 °C	30 sec
2nd	94 °C	5 sec
3rd	52 °C	5 sec
4th	80 °C	5 sec

(repetition of step 3rd to 4th in 5 cycles)

(repetition of step 2nd to 4th in 39 cycles)

5th	72 °C	3 min
6th	4°C	10 min

The 200-short-cycle program uses two cycling loops, one of five cycles that sits inside a loop of 40 cycles.

3.3.14.4 Clean Resin and MALDI-TOF analysis

The iPLEX reaction product is desalted to optimize mass espectometry analysis. The samples are diluted with 16 µl dH_2O and dispensed onto a 6 mg resine dimple plate. After centrifugation, the reaction product is transferred onto a 384-element SpectroCHIP bioarray and this one loaded into a MassARRAY system for mass spectrometry analysis. Genotype calling is performed with the SpectroTYPER 3.4 software

4. RESULTS

4.1 INTELLECTUAL DISABILITY

Intellectual disability (ID) is the most common problem in Clinical Genetics and represents a substantial social-economic burden of health care either in industrialized and developing countries. ID is characterized by subnormal cognitive functioning and deficits in adaptive behaviours with onset before the age of 18 years. ID has an estimated prevalence of 2-3% (Roeleveld et al. 1997). The spectrum of ID diseases results from a broad aetiology including genetic causes such as chromosomal aberrations (e.g., trisomies, unbalanced translocations), microdeletion/microduplication syndromes, and monogenic disorders. However, the disease-causing defect remains unknown in up to 50-80% of patients (Johnson et al. 2006, Rauch et al. 2006). Conventional cytogenetic analysis detects chromosomal abnormalities in about 10% of the ID patients, at least half of them being cases of trisomy 21. The resolution of microscopic cytogenetic techniques is limited to a level of 5-10 Mb. Analysis of subtelomeric regions with fluorescence in situ hybridization (FISH) or muliplex ligation-dependent probe amplification (MLPA) revealed submicroscopic rearrangements in an additional fraction of 5% of the affected children (Ravnan et al. 2006). To further increase the rate of detection of cryptic aneusomies, new methods based on array techniques have been developed over the recent years (Solinas-Toldo et al. 1997, Pinkel et al. 1998, Shaw-Smith et al. 2004, de Vries et al. 2005, Slater et al. 2005, Friedman et al. 2006). Array-based comparative genome hybridization (aCGH) and high-density single nucleotide polymorphism (SNP) microarrays not only provide higher resolution, but also permit the detection of submicroscopic gains and losses of genetic material (copy number variations, CNV) across the whole genome. Previous studies showed that these Chip-technologies raise the detection rate of chromosomal aberrations among ID patients to a level of up to 20% (Shaw-Smith et al. 2004, Ming et al. 2006, Miyake et al. 2006, Wagenstaller et al. 2007).

The study here reported used high-dense SNP oligonucleotide microarrays in order to screen for CNV in a cohort of patients with unexplained MR. The findings allowed the identification of several genes involved in ID, thus providing insights into the pathophysiology of syndromic and non-syndromic ID and into functional networks relevant for the development of the brain.

4.1.1 DATA ANALYSIS

4.1.1.1 Specificity and sensitivity; false positive detection rate

Genome-wide screening for CNVs in a clinical genetic setting at a high resolution is called molecular karyotyping and includes technologies such as array CGH or SNP oligonucleotide arrays. Several studies have demonstrated the potential value of genomic arrays in the diagnosis of unexplained intellectual disability disorders (Shaw-Smith et al. 2004, de Vries et al. 2005, Tyson et al. 2005, Schoumans et al. 2005, Ming et al. 2006, Miyake et al. 2006, Friedman et al. 2006, Hoyer et al. 2007, Wagenstaller et al. 2007, McMullan et al. 2009), and therefore it has been proposed that arrays are included on the genetic diagnostic work-up of these patients (Rauch et al. 2006, Vermeesch et al. 2007, Gijsbers et al. 2009, Miller et al. 2010). However, it is important to note that molecular karyotyping can be performed using different array platforms, different array types of increasing resolution, and different algorithms with varying thresholds. Thus the effective resolution in molecular karyotyping is not uniform. Moreover, the accuracy of a test in detecting chromosomal aberrations, that is, the ability to call true positive events and to avoid both false positives and false negatives is subject to experimental variability between different laboratories and even within a single laboratory. Thus, the challenge of an array experiment is to call all targets (thousands of loci) with high sensitivity and specificity levels. Specificity is defined as the probability of a test to be negative if the target is negative (no CNV detection if there is no CNV). Sensitivity is defined as the probability of a positive test result if the target is positive (CNV detection if there is a CNV). In other words, high specificity means low rate of false positives, high sensitivity means low rate of false negatives. Usually, an increase of sensitivity implies a decrease of specificity, and vice versa.

Here I report on the results in children with mental retardation that were examined using the Illumina high-throughput SNP-genotyping platform. A cohort of 327 sporadic novel patients with mental retardation of unknown origin and 3 patients with an apparently *de novo* balanced reciprocal translocation were screened systematically in order to identify and characterize causative copy number variations (CNVs). To set up the operating procedure and to optimise and standardise results, we started with 110 samples that were hybridised on Human550-Quad arrays and 109 samples that were hybridised on Human610-Quadv1_B arrays. In order to evaluate the resolution limit versus the specificity of each array type, the preliminary results of each dataset i.e. CNV calls, were validated by means of quantitative PCR. In other words, we

determined the rate of false positives for different limits of resolution in order to set up the threshold at which the detection of false positives is negligible. The resolution limit was varied by changing the number of consecutive SNPs that had to be aberrant (which in turn directly related to genomic size of the threshold). Once this cutoff parameter was established, the rest of the MR patients were analysed with Human610-Quadv1_B arrays and without follow-up validation experiments.

4.1.1.1.1 HumanHap550 arrays

A first set of 110 DNA samples were hybridized on Human550-Quad arrays. All analysis tools were implemented as R or Perl scripts. A total of 1,040 CNVs (749 losses, 291 gains) ranging in length from 1,604 to 17,510,855 base pairs (mean 170,517, median 32,959) were detected. For these CNVs we calculated the standard deviation (mean SD: 0.15) and the median absolute deviation (mean MAD: 0.11) of the log2 intensities ratios as well as the signal-to-noise ratio (mean SNR: 5.44). Next, we evaluated 293 out of 1,040 candidate regions with 2 or 3 quantitative PCRs each (638 qPCRs in total). This method was used in the present study to verify chromosomal copy number, to map closer the deletion/duplication breaking points, and to determine the CNV inheritance by including the parental DNA to each qPCR experiment.

The following criteria were used to select relevant CNVs; i) novel CNVs (i.e., not present in the Database of Genomic Variants DGV), ii) loss CNVs that were described in DGV as gains and confined genes, iii) and CNVs known to DGV that overlapped with genes associated with known brain disorders (OMIM database).

Consequently, 161 CNVs (55%) were determined to be true-positive findings and 132 CNVs (45%) emerged to be false-positive findings. However, our experimental design probably leads to the underestimation of the true positive rate due to the exclusion of known CNV polymorphisms in the selection of CNV to be evaluated by qPCR. Under the assumption that all detected known CNV polymorphisms are true positive, this rate would rise up to 87.3%. Consequently we would have detected an overall false positive rate of 12.7%. Interestingly, we observed that 113 (85,6%) of the false positives were detected in regions defined by <10 SNPs, 17 (12.9%) were indicated by 10-20 SNPs and only 2 (1.5%) by >20 SNPs. 132 out of the 161 true positive CNVs were determined to be inherited from one of the healthy parents. While no parental inheritance bias (p=0.2230) was detected (73 CNVs inherited from the mother, 59 CNVs inherited from the father, 2 CNVs detected in both parents, 13 CNVs without origin determination because of missing parental DNA), our data significantly (p= 0.0001) indicated that loss

CNVs (114 deletions) are more frequent than gains (47 duplications).

Table 4-1: Comprehensive results of a genome-wide screen of CNVs by Illumina SNP-arrays

	550K array/Custom R Scripts		**610K array/Dnacopy program**		**Total**	
Patients	110		109		219	
CNVs calls	1040	Mean length = 170,517 Median length = 32,959 Losses = 749 Gains = 291	3057	Mean length = 85,496 Median length = 24,352 Losses = 2407 Gains = 650	4097	Mean length = 108,900 Median length = 27,093 Losses = 3156 Gains = 941
Evaluated CNVs	293	qPCR = 638	170	qPCR = 362	463	qPCR = 1000
True-Positives	161 (55%)	Deletions = 114 Duplications = 47 p= 0.0001 Maternal = 73(-1) Paternal = 59(-1) p= 0.2230	69 (40.6%)	Deletions = 50 Duplications = 19 p = 0.0002 Maternal = 22(-2) Paternal = 28(-2) p = 0.3961	230 (49.8%)	Deletions = 164 Duplications = 66 P =0.0001 Maternal = 95 Paternal = 86 P=0.5035
False-Positives	132 (45%)	<10 SNPs = 113 (85.6%) 10-20 SNPs = 17 (12.9%) >20 SNPs = 2 (1.5%)	101 (59.4%)	<10 SNPs = 77 (76.2%) 10-20 SNPs = 19 (18.8%) >20 SNPs = 5 (5%)	233 (51.2%)	<10 SNPs = 190 (81.6%) 10-20 SNPs = 36 (15.4%) >20 SNPs = 7 (3%)
Overall False Positive Rate	12.7%.		3.3%		5.7%	
Diagnostic yield	15.5%		10.1%		12.8%	

4.1.1.1.2 Human610-Quadv1_B arrays

A second group of 109 mental retardation patients was investigated using the updated Illumina product, Human610-Quadv1_B array. We analysed the data quality in comparison to that of the 550k array (see table 4-1 and table 4-2). The variance of the log2 intensity ratio was larger (mean SD: 0.19 and mean MAD: 0.12) and the signal-to-noise ratio was reduced considerably (mean SNR: 3.82). This means that the inherent noise of the data was higher, possibly infering with the reliability of CNV calling. Illumina 610K arrays were shown to have a smaller SNR value not only in comparison to the former version 550K but also to the Affymetrix Mapping 100K and 500K Array Set. Our group previously had published the evaluation of the 100K Affymetrix platform for submicroscopic imbalances in a cohort of 67 MR patients

(Wagenstaller et al. 2007). For the Hind III array of that platform SNR was 4.16, and for XbaI array 5.14. Despite the higher SNP density and the higher call rate for the tag SNP set, the lower SNR of 610K arrays considerably impairs the reliability of CNV detection studies.

In order to deal with worse SNR values, we replaced the CNV analysis method by a more sophisticated algorithm which is implemented in the commercial dnacopy package (Olshen et al. 2004, Venkatraman et al. 2007). As before, candidate changes of number of copy were compared with a database of known polymorphisms (DGV) and possibly relevant regions were confirmed by qPCR. The applied selection criteria were slightly different. In order to consider a CNV as a candidate cause of MR: i) the CNVs had to match with OMIM genes, ii) it had to be a loss CNV that matched with a gain polymorphism embedding genes, iii) or it had to be an undescribed CNV lying in coding sequences. To simplify CNV testing, in a first approach, we excluded from the qPCR analysis the majority of novel CNV detected in introns, ESTs, and intergenic regions. Nevertheless, we bear in mind that they could be pathogenic through regulatory sequences or non-annotated genes.

We detected 3,057 CNVs, comprising 2,407 losses and 650 gains. The minimum and maximum length were 118 and 13,386,172 base pairs, respectively (mean 85,496, median 24,352). 170 out of 3,057 candidate regions were investigated with a total of 362 qPCR assays (see table 4-1). Of these, 69 (40.6%) were confirmed and 101 (59.4%) were false-positive. 77 (76.2%) of the false positive were detected in regions defined by <10 SNPs, 19 (18.8%) were indicated by 10-20 SNPs and 5 (5%) by >20 SNP. Under the assumption that all detected known CNV polymorphisms were true positives, the overall false positive rate would be 3.3%. 50 out of the 69 true positive CNVs were determined to be inherited from one of the healthy parents. 22 CNVs were maternally inherited, 28 CNVs were paternally inherited, and the origin of 9 CNVs could not be determined due to the lack of parental DNA. We detected 50 deletions and 19 duplications. Therefore, there was no parental inheritance bias ($p=0.3961$) and deletions were significantly more frequent than duplications ($p= 0.0002$).

Table 4-2: Statistical analysis of the resolution values of Illumina 550K and 610K arrays

Array Type; HumanHap550K		CNVs detected with custom R-scripts	CNVs confirmed by qPCR	< 10 SNPs bin (confirmed)	10-20 SNPs bin (confirmed)	> 20 SNPs bin (confirmed)
Length (bp):	Median	32,959	68,824	17,661	82,226	304,318
	Mean	170,517	637,927.46	23,514.63	93,296.13	1,560,451
	Minimum	1,604	2,249	2,249	16,383	37,016
	Maximum	17,510,855	13,295,012	125,103	238,988	13,295,012
	N	1,040	161	68	30	63
Log_2 intens. ratio:	Mean SD	0.15				
	Mean MAD	0.11				
Signal to noise ratio:	Mean	5.44				
Array Type: Human610-Quadv1_B		CNVs detected with Dnacopy	CNVs confirmed by qPCR	< 10 SNPs bin (confirmed)	10-20 SNPs bin (confirmed)	> 20 SNPs bin (confirmed)
Length (bp):	Median	24,352	80,023	14,280	41,684	539,048
	Mean	85,496	965,415.08	17,262.56	55,433.24	1,816,530
	Minimum	118	1,832	1,832	9,644	69,480
	Maximum	13,386,172	13,386,172	40,608	234,303	13,386,172
	N	3,057	69	16	17	36
Log_2 intens. ratio:	Mean SD	0.19				
	Mean MAD	0.12				
Signal to noise ratio:	Mean	3.82				

4.1.1.2 Distribution of the CNV length and of the SNPs bin length of 550K and 610K arrays

It has been estimated that typical individuals are hemizygous for 30-50 CNVs larger than 5 kb, totaling around 550-750 kb of euchromatic sequence across their genomes (Conrad et al. 2006).

In our study, the average of predicted CNVs per individual was 9 in the 550K dataset and 28 in the 610K dataset. In the first sample, 50% of the predicted CNVs were smaller than 32,959 bp, whereas in the second one they were smaller than 24,352 bp. However, when we considered solely the subgroup of CNVs confirmed by qPCR the observed median length was 68,824 bp and 80,023 bp respectively (see table 4-2 and table 4-3).

Table 4-3: Summary of genomic sizes of the CNVs obtained with Illumina SNP chips

Predicted CNVs from 550K arrays (N=1040)			**Predicted CNVs from 610K arrays (N=3057)**		
	count (snp)	length (bp)		count (snp)	length (bp)
mean	34	170,517	mean	20.54	85,496
median	8	32,959	median	8	24,352
min	5	1,604	min	5	118
max	3,545	17,510,855	max	3,280	13,386,172
Confirmed CNVs from 550K dataset (N=161)			**Confirmed CNVs from 610K dataset (N=69)**		
	count (snp)	length (bp)		count (snp)	length (bp)
mean	138.5	637,927.46	mean	206.5	965,415.08
median	14	68,824	median	26.5	80,023
min	5	2,249	min	5	1,832
max	3,545	13,295,012	max	3,280	13,386,172

In accordance to the literature (Conrad et al 2006), the length of the CNVs identified in our two samples followed an L-shaped distribution, both in the predicted and confirmed CNVs datasets. In general, there are many small CNVs and a few large ones. The larger ones seem to have a very low population frequency. Distributions are plotted as histograms and shown in the figure 4-1.

Additionally, there are 66% more predicted CNVs in the 610K sample, being mostly smaller. This difference of CNV prevalence between two cohorts of roughly 100 individuals is not only explained by the use of a higher probe density array and the use of an algorithm with higher SNP calling rate, but also by the detection of more false positives. The last is discussed in the next section (section 4.1.1.3).

On the contrary, when we compared the total number of confirmed CNVs, we observed that there is 42% more assessed CNVs in the 550K dataset than in the 610K dataset, also being mostly small (Figure 4-1). It is logical to detect more CNVs in the 550K dataset because we evaluated almost the double of selected CNV calls. However, we observed a higher proportion of small CNVs in the 550K dataset. The difference of proportion of small CNV calls between the two datasets can be explained most likely due to the better quality of the intensity signals of the 550K arrays. The larger SNR values of 550K arrays probably allowed working at higher resolution and with higher levels of sensitivity, thus identifying more small CNVs than when using 610K arrays.

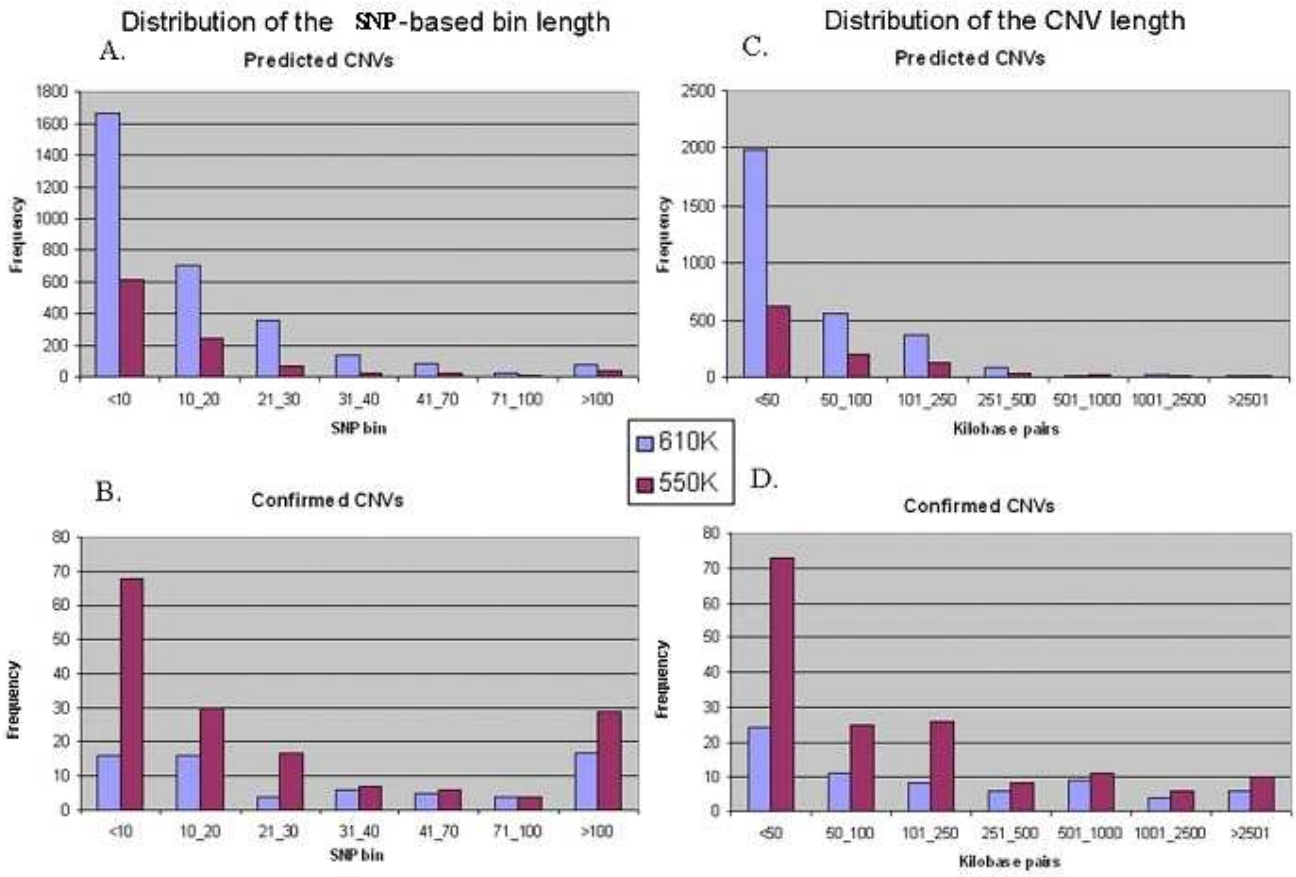


Figure 4-1: Frequency distribution histograms plotting the length values of 1040 and 3057 CNVs called with 550K and 610K arrays respectively. The purple bars represent the CNV data of the 550K arrays and the blue bars the CNV data of the 610K arrays. (A)(B) Length distribution of CNVs accounted into different SNP-based bins. (C)(D) Length distribution of CNVs accounted into different Kb-based clusters. (A)(C) Distribution of CNV calls. (B)(D) Distribution of confirmed CNVs.

4.1.1.3 Setting of the optimal resolution threshold

Due to the fact that the 550K arrays were transitioned to the Human610-Quadv1_B array and were not commercially available any longer, the remaining patients of the ID cohort here presented had to be analysed using the later version of the arrays. Hence, this section will put special emphasis on the data analysis and the setting of the achievable resolution of 610K arrays.

It was observed that the circular binary segmentation algorithms could pick up more and smaller copy number regions and *a priori* slightly improved the overall false positive rate (see table 4-1, table 4-2 and

table 4-3). First, the resolution of *de novo* genomic imbalances of Illumina 610K chips is as small as 1.8 kb, making the detection of causative intragenic rearrangements feasible. Second, under the assumption that all predicted known CNV polymorphisms are true positive, the false positive rate was observed to rise up to 3.3%. The lower false positive rate frequency of the 610K dataset could be explained due to the bias introduced when weighting the false positives to the total number of CNV calls. As mentioned before, 1,040 CNVs calls were detected in the 550K dataset versus 3,057 CNVs of the 610K dataset. Of these, only 293 CNVs and 170 CNVs respectively were evaluated with qPCR. Therefore, for the 610K dataset we are assuming that the unevaluated CNVs are true positives in a higher proportion, thus overestimating the true positive rate further in the 610K dataset.

To be more accurate when comparing the false positive rate between the two datasets, we weighted the false positives to the number of evaluated CNVs for each specific SNP bin (see table 4-4). We observed that the false positives rate detected in the 610K dataset was 82.8% in the <10 SNPs bin, 52.8% in the 10-20 SNPs bin and 12.2% in the >20 SNP bin. On the contrary, the false positive rates of the 550K dataset were 62.4%, 36.2% and 3.1% for each respective SNPs bin. Hence, we can conclude that due to the worse raw intensities profiles of the 610K chips and/or the lower specificity of the circular binary segmentation algorithms in all limits of resolution (i.e. SNP-based bin) the overall false positive rate of the 610K dataset is expected to be much higher than the one of 550K.

Both the in-house Perl script and the DNAcopy software call more deletions than duplications (see table 4-1), but the proportion of losses detected by the circular binary segmentation resulted to be slightly greater; the in-house software detected 61% more deletions than duplications, the DNAcopy detected 73% more deletions than duplications. Nonetheless, when we analyzed the percentage of false positive deletions respect the percentage of false positive duplications in both datasets we observed that, whereas the proportion of false positive deletions was much larger in all the SNP bins of the 610K dataset (91.5% more deletions than duplications in the <10 SNPs bin, 81.3% more deletions in the 10-20 SNPs bin and only false deletions in the >20 SNP bin), they were approximately the same in all 550K SNP bins (except for the >20 SNP bin in which only false positive duplications were detected) (see figure 4-2) . Therefore, the validation experiments demonstrated furthermore that the reliability of the circular binary segmentation algorithm is worse when calling deletions, as indicated for the higher rate of false positive deletions in all the SNP bins. Taken together these data suggest that the increased false postive rate observed in 610K dataset it is most likely a consequence of the combined performance of the microarrays (i.e. larger variance

of the log2 intensity ratio) and the CBS algorithm (which predicts more CNVs, calls more deletions and models losses with less specificity).

Table 4-4: True-positive and False-positive rates from the SNP-based bins of 550K and 610K datasets

	550K array/Custom R Scripts			**610K array/Dnacopy program**		
<10 SNPs bin	CNV calls	617		CNV calls	1668	
	CNVs tested	181	True-positive: 68 (37.6%)	CNVs tested	93	True-positive: 16 (17.2%)
			False.positive: 113 (62.4%)			False.positive: 77 (82.8%)
10-20 SNPs bin	CNV calls	241		CNV calls	704	
	CNVs tested	47	True.positive: 30 (63.8%)	CNVs tested	36	True-positive: 17 (47.2%)
			False.positive: 17 (36.2%)			False.positive: 19 (52.8%)
>20 SNPs bin	CNV calls	182		CNV calls	685	
	CNVs tested	65	True positive: 63 (96.9%)	CNVs tested	41	True-positive: 36 (87.8%)
			False.positive: 2 (3.1%)			False.positive: 5 (12.2%)

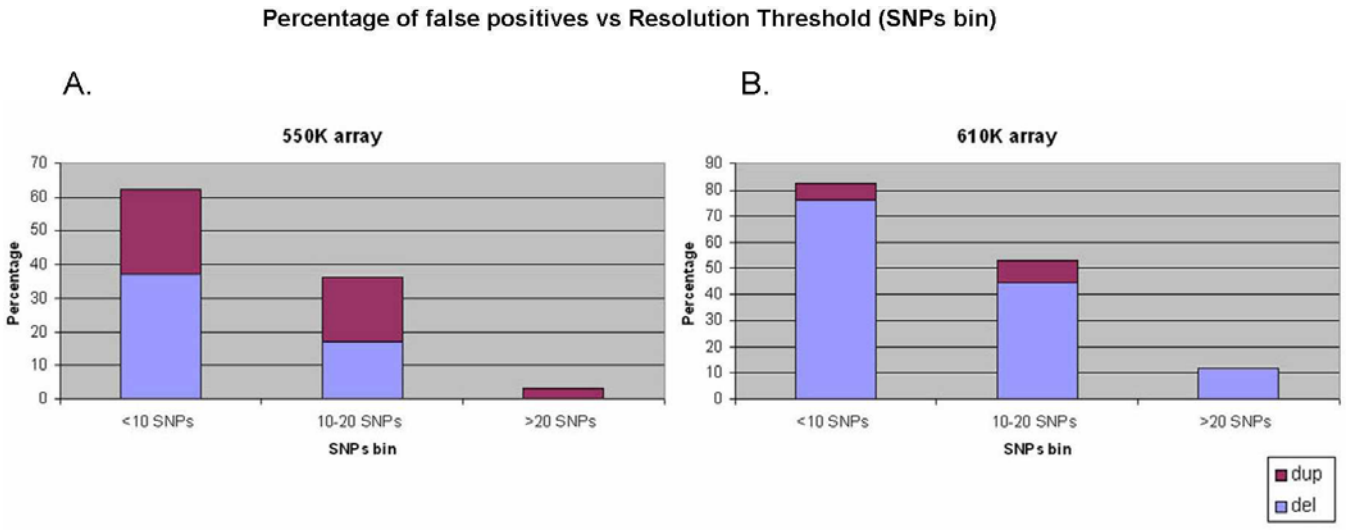


Figure 4-2: Diagram of bars displaying the false positive CNV results obtained at three different limits of resolution (specific bins indicated by <10 SNP, 10-20 SNPs and >20 SNPs). The purple section of the bars represents the fraction of false positive duplications, whereas the blue section of the bars indicates the number of false positive deletions. (A) Bar diagram plotting the false positive values of the 550K dataset. (b) Bar diagram plotting the false positive values of the 610K dataset.

In order to choose a level of resolution that balanced sensitivity and specificity, we used as a cuttoff parameter the smoothing median of windows of more than 30-40 consecutive SNPs, i.e. 200kb-500kb. The more or less stringent resolution threshold was applied depending on the quality of the signal intensity profile of each particular array.

Once the achievable resolution limit of the Illumina arrays was determined we performed a CNV genome-wide analysis in the remaining 111 MR patients without subsequent validation experiments. Genomic DNA of the patients were hybridised on Human610-Quadv1_B arrays and the CNV calling was performed with the DNAcopy software at the established resolution limit. Conspicuous regions were compared with known CNVs as provided by the Database of Genomic Variants. The parental DNAs of those patients with a conspicious chromosomal aberration were hybridized on 610K arrays in order to determine whether the specific imbalance emerged *de novo* or was inherited from one of the healthy parents.

4.1.1.4 Detection of chromosomal imbalances

4.1.1.4.1 *De novo* CNV

i) First set of 110 MR patients analyzed with 550K arrays

15 *de novo* CNVs (13 deletions and 2 duplications) which varied in size from 125 kb to 13.3 Mb, were identified. The only causative CNV without *de novo* confirmation is a known microdeletion syndrome (2q37 monosomy). Another six CNVs are also known genomic disorders, including the *FOXP* and *CFTR* deletion, 13q32 deletion (Holoprosencephaly), Prader-Willi syndrome, 22q11 duplication, 1q42-q44 deletion and 16p11.2-p12.2 deletion. Four CNVs overlap with deletions previously reported in the DECIPHER database. The remaining five *de novo* CNV were not described before.

Moreover, we identified a complex rearrangement in patient 32472. An apparently balanced Robertsonian translocation, 45,XY,t(14,22) was identified in this patient by standard cytogenetic analysis. However, the molecular karyotyping identified a complex *de novo* unbalanced translocation consisting of two different rearrangements in a mosaic form; a16.7 Mb duplication at 11q23.3-q25 and a 27.9 Mb deletion at the 18q22.3-q23 band.

The reported chromosomal imbalances are listed in table 17.

ii) Second set of 109 MR patients analised with 610K arrays

In addition to the *de novo* genomic imbalances described in the prior batch of samples, now we identified 13 *de novo* CNVs corresponding to 9 patients most likely to be causative of mental retardation. We detected 11 deletions and 2 duplications, which varied in size from 73 kb to 13.4 Mb. We identified three patients with known syndromes, 14q11.2 deletion, 16p11.2-p12.2 deletion and NF1 deletion type I, six patients with novel *de novo* CNVs and one patient with an unbalanced translocation (see Table 17).

We refined and narrowed-down further the breakages of a complex *de novo* unbalanced translocation. In this patient (28181) a *de novo* interstitial balanced translocation 46,XY,t(4;6)(q28.3q31;q23.1q22.2) was detected by traditional karyotyping. We recently identifited the breakages of a *de novo* unbalanced translocation in the mentioned patient, by using Affymetrix GeneChip Human Mapping 100K arrays (chromosomal bands 4q28.3-31.1 and 6q16.1-21, sized 3.9 and 14.3 Mb respectively) We decided to test either resolution of Human610-Quadv1_B chips and the analysis performance of DNAcopy program by screening CNVs again in this patient. We identified the same 4q28.3-q31 deletion but the loss of genetic material in chromosome 6 now corresponded to two different rearrangements, 13.4 Mb at 16q16.2-q2 and 1.6 Mb at 16q24.1.

Another patient, 38749, showed to have two deletions in the 16p11.2 deletion loci. The two deletions were flanking a region rich in segmental duplications, which might contribute to an artifactual signal intensity of this region. Therefore, further experiments such as qPCR and junction fragment amplification are needed to map accurately the breakpoints of this rearrangement and elucidate whether this result corresponds indeed to a complex rearrangement.

Another two *de novo* CNVs were found in another MR case, patient 39753; a 300 Kb deletion at 6p25.1 and a 349 Kb duplication at Xp21.3. In order to confirm the true origin of these CNVs, we performed the paternity test on the referred trio of samples (patient, father, and mother DNAs) with the AmpFLSTR Identifiler PCR Amplification Kit. The probability of paternity of the "father" was 0.9999992. Consequently, at least one of these two *de novo* rearrangement is expected to be the pathogenic imbalance responsible for the MR suffered for this child. The 349 Kb duplication at Xp21.3 identified in this patient includes exons 3-5 of *IL1RAPL1* (MIM 300231). *IL1RAPL1* is known to be involved in cognitive function (Gao et al 2008). Deletions encompassing (part of) this gene have been associated in multiple families with XLID and autism spectrum disorder (ASD) (Jin et al 2000, Zhang et al 2004, Tabolacci et al 2006, Piton et al 2008, Nawara et al 2008, Whibley et al 2010, Franek et al 2011). Here we report for the first time the

reciprocal intragenic *IL1RAPL1* duplication which is predicted to result, as for intragenic deletions, in loss-of-function, and to underlie the mutational mechanisms causal of disease.

Another patient (33361) showed to have a large complex rearrangement on chromosome 4, although its *de novo* status could not be determined. It consisted of a terminal 1.5 Mb deletion at 4p16.3 and a neighbouring 13.2 Mb duplication at the 4p16.3-p15.33 band.

iii) Third set of 111 MR patients analised with 610K arrays

In the last batch of MR samples 4 *de novo* CNVs were identified; 3 deletions and 1 duplication, with a length between 547 kb and 3.67 Mb.

Two deletions at the 22q11.22 band (one of 1,44Mb and the other of 740kb) were identified in the patient 38194, which presented an atypical form of Di-George Syndrome. Another patient was identified with a known chromosomal disorder, the 5p13 duplication syndrome. This patient, 38279, showed to have a 3,67 Mb duplication at the chromosomal band 5p13.1-p13.2. The remaining *de novo* CNV was not described before.

In addition, one patient (37210) with an apparently balanced translocation, 46,XX,t(2;10), was analysed in this group of samples. After SNP-chip analysis this structural rearrangement showed to be unbalanced. We detected a 7.808 Mb deletion at the 2q36.1-q37.1 band in a mosaic state. The *de novo* status of this rearrangement could not be confirmed due to the lack of parental samples.

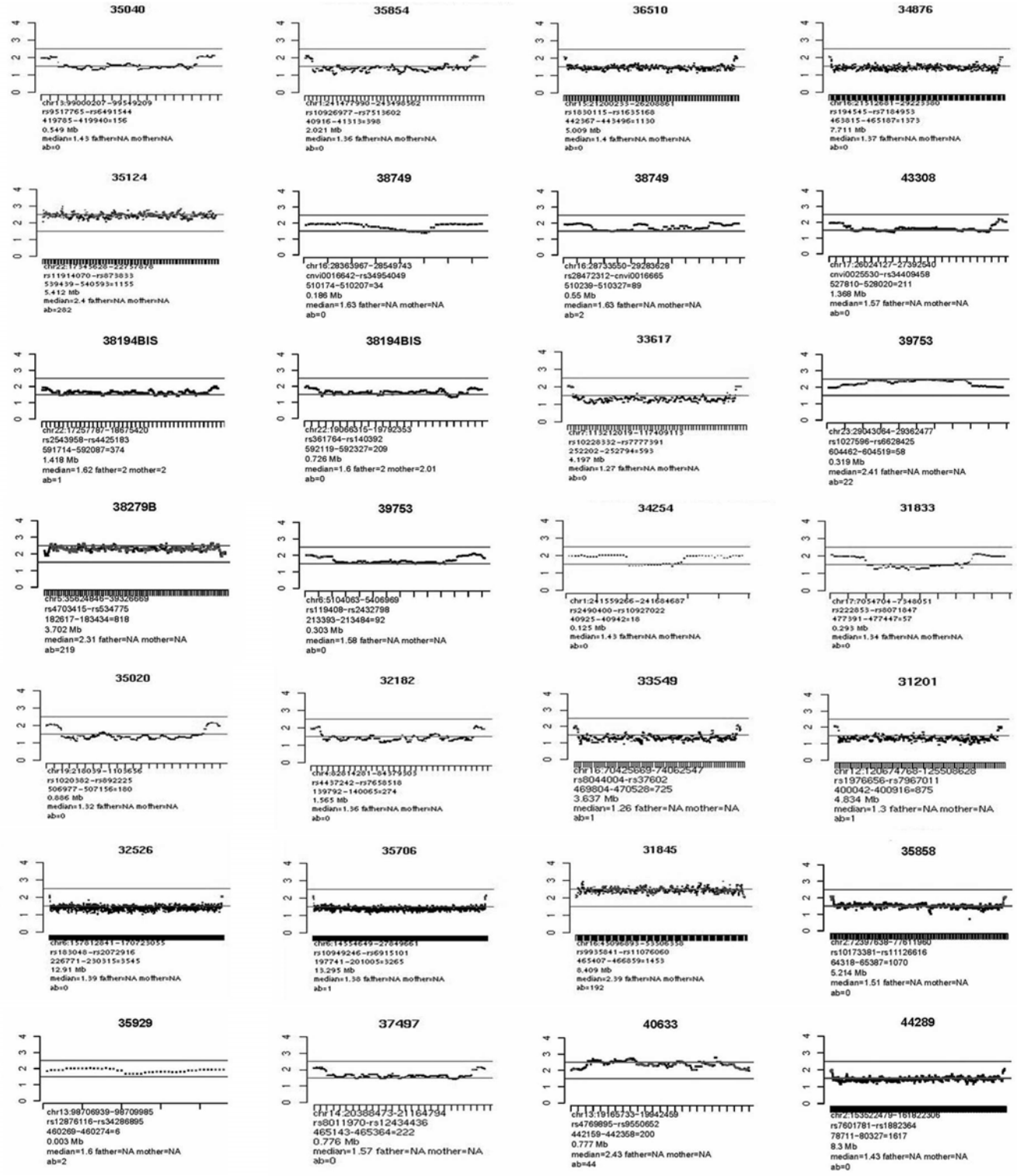
The whole cohort of MR could not be completely scrutinized, because of missing parental samples in some patients. CNV hits from these cases were mostly validated with qPCR, but their *de novo* appearance could not be tested. Therefore, 21 CNVs could not be defined or excluded as a polymorphism. 7 out of these CNVs without *de novo* confirmation were located at loci whose haploinsuficiency is known to cause MR. The following table lists the *de novo* CNVs and the potential disease-causative CNVs.

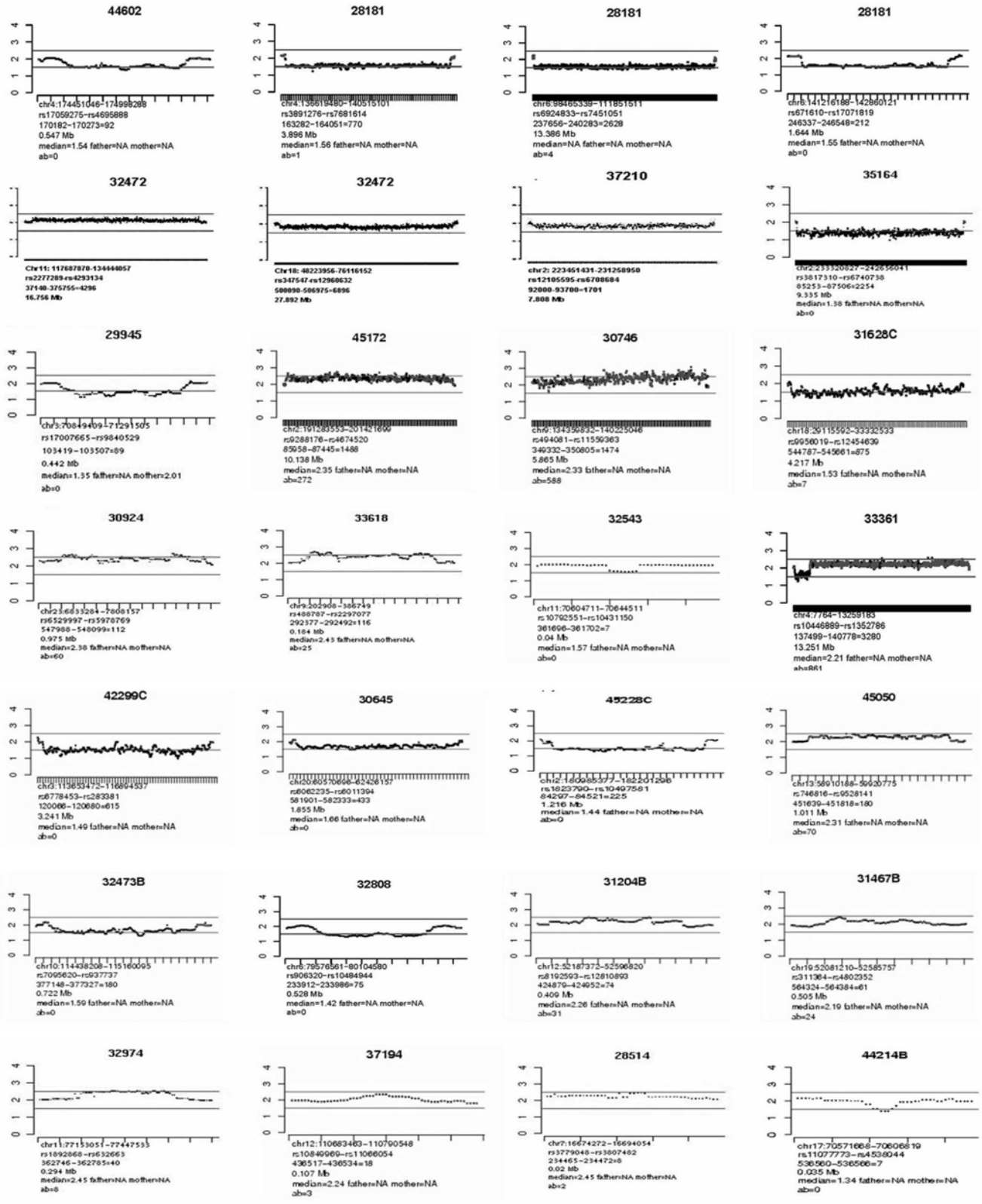
Table 4-5: Summary of the chromosomal imbalances identified with the Human550-Quad and Human610-Quad arrays (Illumina)

Patient ID	Chromosome band	Gain/ Loss	Position Start hg18	Position End hg18	Start rsSNP	End rsSNP	Number of SNPs	Length (Mb)	Array Type	Known Syndrome	Ensembl Genes	Parent of origin
35040	13q32.2	Loss	99000207	99549209	rs9517765	rs6491544	156	0.549	550K	Holoprosencephaly	22	de novo
35854	1q43-q44	Loss	241477990	243498562	rs10926977	rs7513602	398	2.021	550K	1q42-q44 del	52	de novo
36510	15q11.2-q13.1	Loss	21200233	26208861	rs1830115	rs1635168	1,130	5.009	550K	Prader-Willi syndrome	182	de novo
34876	16p11.2-p12.2	Loss	21512681	29223380	rs194545	rs7184953	1,373	7.711	550K	16p11.2-p12.2 del	163	de novo
35124	22q11.21-q11.23	Gain	17345628	22757878	rs11914070	rs873833	1,155	5.412	550K	22q11 dup	248	de novo
38749	16p11.2	Loss	28363967	28549743	cnvi0016642	rs34954049	34	0.185	610K	16p11.2 del	17	de novo
38749	16p11.2	Loss	28733550	29283628	rs28472312	cnvi0016665	89	0.550	610K	16p11.2 del	30	de novo
43308	17q11.2	Loss	26024127	27392540	cnvi0025530	rs34409458	211	1.368	610K	NF1 type I	34	de novo
38194	22q11.22	Loss	17257787	18675420	rs2543958	rs4425183	374	0.418	610K	Di-George syndrome	106	de novo
38194	22q11.22	Loss	19066315	19792353	rs361764	rs140392	209	0.726	610K	Di-George syndrome	56	de novo
33617	7q31.2-q31.31	Loss	113212019	117409113	rs10228332	rs7777391	593	4.197	550K	FOXP2/CFTR gene	36	de novo
39753	Xp21.3	Gain	29043064	29392222	rs1027596	rs7063024	60	0.349	610K	IL1RAPL1 gene	2	de novo
38279	5p13.1-p13.2	Gain	35624846	39304128	rs4703415	rs169048	815	3.679	610K	5p13 duplication	57	de novo
39753	6p25.1	Loss	5104063	5406969	rs119408	rs2432798	92	0.302	610K	no	3	de novo
34254	1q44	Loss	241559266	241684687	rs2490400	rs10927022	18	0.125	550K	no	3	de novo
31833	17p13.1	Loss	7054704	7348051	rs222853	rs8071847	57	0.293	550K	no	53	de novo
35020	19p13.3	Loss	218039	1103656	rs1020382	rs892225	180	0.885	550K	no	70	de novo
32182	4q21.22	Loss	82814281	84379303	rs4437242	rs7658518	274	1.565	550K	no	29	de novo
33549	16q22.2-q23.1	Loss	70425669	74062547	rs8044004	rs37602	725	3.637	550K	no	53	de novo
31201	12q24.31-q24.32	Loss	120674768	125508628	rs1976656	rs7967011	875	4.834	550K	no	109	de novo
32526	6q25.3-q27	Loss	157812841	170723055	rs183048	rs2072916	3,545	12.910	550K	no	265	de novo
35706	6p22.1-p23	Loss	14554649	27849661	rs10949246	rs6915101	3,265	13.295	550K	no	324	de novo
31845	16q11.2-q12.2	Gain	45096893	53506358	rs9935841	rs11076060	1,453	8.409	550K	no	80	de novo
35858	2p12-p13.3	Loss	72397638	77632757	rs10173381	rs12713918	1,074	5.235	610K	no	105	de novo
35929	13q32.3	Loss	98706939	98709985	rs12876116	rs34286895	6	0.003	610K	no	8	de novo
37497	14q11.2	Loss	20388473	21164794	rs8011970	rs12434436	222	0.776	610K	14q11.2 del	44	de novo
40633	13q12.11	Gain	19165733	19942459	rs4769895	rs9550652	200	0.776	610K	no	34	de novo
44289	2q23.3-q24.2	Loss	153522479	161822306	rs7601781	rs1882364	1,617	8.229	610K	no	69	de novo
44602	4q34.1	Loss	174451046	174998288	rs17059275	rs4695888	92	0.547	610K	no	10	de novo
28181	4q28.3-q31.1	Loss	136619480	140515101	rs3891276	rs7681614	770	3.895	610K	translocation	29	de novo
28181	6q16.2-q21	Loss	98465339	111851511	rs6924833	rs7451051	2,628	13.386	610K	translocation	272	de novo
28181	6q24.1	Loss	141216188	142860121	rs671610	rs17071819	212	1.643	610K	translocation	13	de novo
32472	11q23.3-q25	Gain	117697870	134444057	rs2277289	rs4293134	4,296	16.756	550K	translocation	mosaic-de novo	
32472	18q22.3-q23	Loss	48223956	76116152	rs347547	rs12960632	6,896	27.892	550K	translocation	mosaic-de novo	
37210	2q36.1-q37.1	Loss	223451431	231258950	rs12105595	rs6708684	1,701	7.808	610K	translocation	mosaic-ND	
35164	2q37.1-q37.3	Loss	233320827	242656041	rs3817310	rs6740738	2,254	9.335	550K	2q37 monosomy	194	ND
29945	3p14.1	Loss	70849409	71291505	rs17007665	rs9840529	89	0.442	550K	FOXP1 gene	1	ND
45172	2q32.2-q33.1	Gain	191283553	201421699	rs9288176	rs4674520	1,488	10.140	610K	2q32.2-q33.1 del	91	ND
30746	9q34.13-q34.3	Gain	134359832	140225046	rs494081	rs11559363	1,474	5.865	610K	9q subtel del	379	ND
31628	18q12.1-q12.2	Loss	29115592	33332533	rs9956019	rs12454639	875	4.217	610K	DTNA gene	67	ND
30924	Xp22.31	Gain	6833284	7808157	rs6529997	rs5978769	112	0.975	550K	STS region	7	ND
33618	9p24.3	Gain	202908	386749	rs488787	rs2297077	116	0.184	550K	DOCK8 gene	9	ND
32543	11q13.3	Loss	70604711	70644511	rs10792551	rs10431150	7	0.039	550K	Shank2 gene	1	NA
33361	4p16.3	Loss	7764	1504781	rs10446889	rs10029689	289	1.497	610K	no	61	ND
33361	4p16.3-p15.33	Gain	1504781	13259183	rs10029	rs1352786	2,991	13.754	610K	no	245	ND
42299	3q13.2-q13.31	Loss	113653472	116894537	rs6778453	rs283381	615	3.241	610K	no	52	ND
30645	20q13.33	Loss	60570696	62426157	rs6062235	rs6011394	433	1.855	610K	no	174	ND
45228	2q31.3	Loss	180985377	182201296	rs1823790	rs10497581	225	1.216	610K	no	12	ND
45050	13q21.1	Gain	58910188	59920775	rs746816	rs9528141	180	1.010	610K	no	17	ND
32473	10q25.2	Loss	114438208	115160095	rs7095620	rs937737	180	0.721	610K	no	17	ND
32808	6q14.1	Loss	79576561	80104580	rs906320	rs10484944	75	0.528	610K	no	3	ND
31204	12q13.13	Gain	52187372	52596820	rs8192593	rs12810893	74	0.409	610K	no	9	ND
31467	19q13.32	Gain	52081210	52585757	rs311364	rs4802352	61	0.504	610K	no	15	ND
32974	11q14.1	Gain	77153051	77447533	rs1892868	rs632663	40	0.294	550K	no	11	NA
37194	12q24.12	Gain	110683463	110790548	rs10849969	rs11066054	18	0.107	610K	no	4	NA
28514	7p21.1	Gain	16674272	16694054	rs3779048	rs3807482	8	0.019	550K	no	1	NA
44214	17q35.1	Loss	70571668	70606819	rs11077773	rs4538044	7	0.035	610K	no	2	ND

ND: de novo occurrence not determined (missing parental samples)
NA: CNV validation not available

Figure 4-3: Copy number values of 56 CNVs identified in MR patients; *de novo* and potential disease-causative CNVs (median smoothing with a window of 9 adjacent SNPs). CNVs with copy number values lower than 1.5 were suspicious for deletion and CNVs with copy number values of more than 2.5 were suspicious for duplication




44602
chr4:174451046-174998288
rs17059275-rs4695888
170182-170273=92
0.547 Mb
median=1.54 father=NA mother=NA
ab=0
28181
chr4:136619480-140515101
rs3891276-rs7681614
163282-164051=770
3.896 Mb
median=1.56 father=NA mother=NA
ab=1
28181
chr6:98465339-111851511
rs6924833-rs7451051
237656-240283=2628
13.386 Mb
median=NA father=NA mother=NA
ab=4
28181
chr6:141216188-142860121
rs671610-rs17071819
246337-246548=212
1.644 Mb
median=1.55 father=NA mother=NA
ab=0
32472
Chr11: 117687878-134444857
rs2277289-rs4293134
37140-375755=4296
16.756 Mb
32472
Chr18: 48223956-76116152
rs347547-rs12960632
500090-506975=6896
27.892 Mb
37210
chr2: 223451431-231258950
rs12105595-rs6708684
92000-93700=1701
7.808 Mb
35164
chr2:233320827-242856041
rs3817310-rs6740736
85253-87506=2254
9.535 Mb
median=1.38 father=NA mother=NA
ab=0
29945
chr3:70849409-71291505
rs17007665-rs9840529
103419-103507=89
0.442 Mb
median=1.35 father=NA mother=2.01
ab=0
45172
chr2:191283553-201421699
rs9288176-rs4674520
85958-87445=1488
10.138 Mb
median=2.35 father=NA mother=NA
ab=272
30746
chr9:134359832-140225046
rs494081-rs11559363
349332-350805=1474
5.865 Mb
median=2.33 father=NA mother=NA
ab=588
31628C
chr18:29115592-33332533
rs9956019-rs12454639
544787-545661=875
4.217 Mb
median=1.53 father=NA mother=NA
ab=7
30924
chr2:56033284-76808157
rs6529997-rs5978769
547988-548099=112
0.975 Mb
median=2.38 father=NA mother=NA
ab=60
33618
chr9:302908-388749
rs488787-rs2297077
292377-292492=116
0.184 Mb
median=2.43 father=NA mother=NA
ab=25
32543
chr11:70604711-70644511
rs10792551-rs10431150
361696-361702=7
0.04 Mb
median=1.57 father=NA mother=NA
ab=0
33361
chr4:7764-13259183
rs10446889-rs1352786
137499-140778=3280
13.251 Mb
median=2.21 father=NA mother=NA
ab=881
42299C
chr3:113653472-116894537
rs6778453-rs283081
120068-120680=615
3.241 Mb
median=1.49 father=NA mother=NA
ab=0
30645
chr20:60570696-62426157
rs6062235-rs6011394
581901-582333=433
1.855 Mb
median=1.66 father=NA mother=NA
ab=0
45228C
chr12:102065377-102201298
rs1823790-rs10497581
84297-84521=225
1.216 Mb
median=1.44 father=NA mother=NA
ab=0
45050
chr13:36910188-39020775
rs746816-rs9528141
451639-451818=180
1.011 Mb
median=2.31 father=NA mother=NA
ab=70
32473B
chr10:114438208-115160095
rs7095620-rs937737
377148-377327=180
0.722 Mb
median=1.59 father=NA mother=NA
ab=0
32808
chr6:79576561-80104580
rs906320-rs10484944
233912-233986=75
0.528 Mb
median=1.42 father=NA mother=NA
ab=0
31204B
chr12:52187373-52596820
rs8192593-rs12810893
424879-424952=74
0.409 Mb
median=2.26 father=NA mother=NA
ab=31
31467B
chr19:52081210-52585757
rs311364-rs4802352
564324-564384=61
0.505 Mb
median=2.19 father=NA mother=NA
ab=24
32974
chr11:77153051-77447535
rs1892868-rs632663
362746-362785=40
0.294 Mb
median=2.45 father=NA mother=NA
ab=8
37194
chr12:110683483-110790548
rs10849669-rs11066054
436517-436534=18
0.107 Mb
median=2.24 father=NA mother=NA
ab=3
28514
chr7:16674272-16694054
rs3779048-rs3807482
234465-234472=8
0.02 Mb
median=2.45 father=NA mother=NA
ab=2
44214B
chr17:70571668-70606819
rs11077773-rs4538044
536560-536566=7
0.035 Mb
median=1.34 father=NA mother=NA
ab=0

4.1.1.4.2 Rare inherited CNV

A *de novo* origin of an imbalance is generally thought to be an important argument suggesting the pathogenity of the imbalance. However given the high number of CNVs in the normal population, one has to consider that some of the *de novo* detected aberrations may represent a *de novo* occurrence of a non pathogenic CNV. On the contrary, the detection of imbalances which appear to be present in phenotypically normal parents of probands, permits to pinpoint these CNVs as benigns. However, it has been reported in the later years that some of those rare imbalances detected in individuals with MR, but shown to be inherited from phenotypically normal individuals, can be benign in some individuals but causal in others (Sharp et al. 2008, Girirajan et al. 2010, Mefford et al. 2009, Fernandez et al. 2009, Cook and Scherer 2008). The phenotypic expression of certain chromosomal imbalances can vary between patients possibly because of; 1) the unmasking of a recessive allele, 2) deletions or duplications of imprinted regions of the genome, 3) X-linked imbalances inherited from a normal mother 4) the existence of genetic or environmental modifiers. Therefore, inherited imbalances may turn out to be susceptibility loci contributing to pathogenity.

In addition to the *de novo* cases of MR, also patients suffering from an autosomal recessive form of MR were found in our cohort. We detected two partial deletions of the *COH1* gene at locus 8q22 which is mutated in Cohen syndrome (CS; OMIM 216550); a 124 kb deletion in patient 32140 and a 155 kb deletion in patient 33147. We sequenced the entire coding region and the exon/intron boundaries of *COH1* in DNA material of these two children. We identified a stop mutation and a frameshift mutation in patient 32140 and 33147 respectively. The identified null alleles lead to protein truncation and may rise to nonsense-mediated RNA decay. Therefore, two compound heterozygous mutations were identified in the *COH1* gene: c.5086C>T / c.3872-5024del [p.R1696X / p.G1291fsX42] and c.1207-2824del / c.11505delA [p.L403fsX11/ p.K3835fsX43], thus providing a distinctive Cohen Syndrome diagnose to two unrelated patients of our MR cohort. We assessed the cosegregation of each mutation in the respective patient families. The phenotypic features and the genetic analysis of these two patients are presented in a greater detail in the section 4.1.2.2.

One boy (ID 44399) exhibited a 40kb deletion at Xq13.1 which was proven to be inherited from his mother. Due to the excess of male patients with mental retardation, the importance of X-linked genes has been established and recently one study has shown a significant skewing of X-inactivation in 5.9% of

mothers of affected boys (Ropers et al. 2005, Rauch et al 2006). It is also known that structural aberrations on the X-chromosome can lead to non-random XCI in females (Willard HF 2000, Mumm et al 2001). Therefore we assumed that, in addition to the detectable cases of *de novo* imbalances maternally inherited XL-CNVs encompassing coding sequences might reveal novel genes responsible for X-linked forms of mental retardation. In this regard, we performed the XCI assay in this case, as described in the methods.

Our results uncovered a preferentially inactivated X chromosome in the mother and the transmission of this chromosome (shown by the inheritance of its related microsatellite allele) to the son. This finding supports the idea that the deleted gene content (*DGAT2L6* and *AWAT1*) of this CNV might have implications on XLMR. The proteins encoded by these genes belong to the diacylglycerol acyltransferase family. No polymorphisms affecting the exonic regions of these genes were described before. Recently, one X-chromsome CNV study based on families with two or more cases of ID in males and with a transmission pattern compatible with X-linked inheritance, detected a deletion encompassing the entire *AWAX* gene. However based on the no cosegregation of the mutation with the disease this gene was not considered as candidate for XLID causation (Whibley et al 2010). In our case, in addition to *AWAT1* also *DGAT2L6* is deleted in this CNV and this gene could be still considered a novel gene underlying XLID. Nevertheless, the absence of expression in brain and the functional information available for *DGAT2L6* (not compatible with neurological disease) precluded further mutational screenings in this gene.

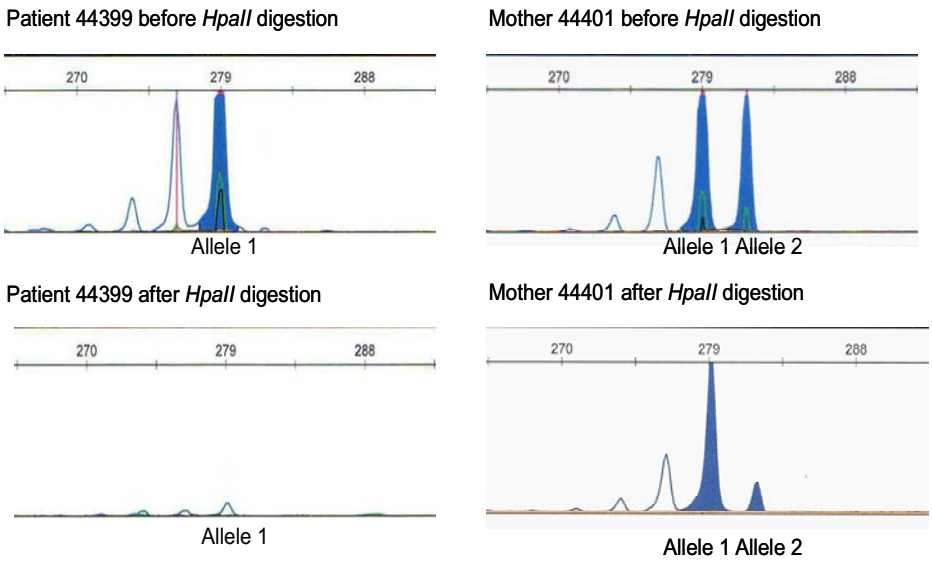


Figure 4-4: Representation of the chromosome X inactivation of the patient 44399 and his mother 44401. The figure shows the results before and after the digestion with the enzyme *HpaII*. The deletion carried by the patient and his mother is associated to the allele 1, and this allele is inactivated in the mother.

Additionally we found in our ID cohort some rare imbalances that were inherited from a phenotypically normal parent. Recent studies of structural genomic variations have revealed the existence of rare imbalances that are significantly enriched in cohorts of patients with unexplained intellectual disability (either *de novo* or inherited) compared with controls, suggesting that they might be neurocognitive disease risk factors. Therefore, these susceptibility loci collectively seem to account for a significant fraction of the genetic variance in common mental retardation. In the present study, we identified six large rare imbalances (at 16p13.11, two at 15q13.3, at 1q21.1, at 22q11.2 and at 12q14), previously associated with neuropsychiatric disorders.

A part of the 6 rearrangements mentioned above, 50 rare inhereted CNVs were identified in our cohort (see table 4-6). These rearrengements were found to contain coding sequences and were not known to the Database of Genomic Variants (DGV). We compared these rearrangements to an in-house database containing CNV data from 4104 control individuals. Controls consisted of 1146 individuals from popgen, 813 individuals of a population-based cohort (KORA study), 972 patients with cardiac ischemia (AGNES study), 482 patients with early-onset lung cancer (LUCY study), and 691 long-lived individuals (LLI study). 30 out of these 50 rare inherited CNVs were neither present in our control cohort.

The table bellow (table 4-6) lists the rare inherited chromosomal alterations identified in this study. The last column of the table shows the number of individuals of our control dataset in which the specific CNV could be detected.

Table 4-6: Summary of the rare inherited chromosomal imbalances identified with the Human550-Quad and Human610-Quad arrays (Illumina)

Patient ID	Chromosome band	Gain/ Loss	Position Start hg18	Position End hg18	Start rsSNP	End rsSNP	Number of SNPs	Length (Mb)	Array Type	Known Syndrome	Ensembl Genes	Parent of origin	ctrl
32140	8q22.2	Loss	100580459	100705562	rs10504991	rs2035618	7	0.124	550K	COH1 gene	6	maternal	-
33147	8q22.2	Loss	100250047	100405623	rs4442113	rs2291548	11	0.155	550K	COH1 gene	5	paternal	-
43488	16p13.11	Loss	15387380	16225138	rs251633	rs2856484	267	0.837	610K	16p13.11 locus	29	maternal	1
43033	15q13.2- q13.3	Loss	28723577	30302218	rs2046362	rs4779984	341	1.578	610K	15q13.3 locus	23	paternal	2
42261	15q13.3	Gain	29883814	30198807	rs905436	rs1392808	62	0.315	610K	15q13.3 locus	10	maternal	9
35701	1q21.1	Loss	144967972	146293282	rs10900384	rs2932454	227	1.325	550K	1q21.1 locus	51	paternal	-
33525	22q11.21	Gain	19876610	20244652	cnvi0017082	rs5754087	47	0.368	610K	22q11.2 locus	23	maternal	4
32794	12q14.2-14.3	Loss	62182448	64096441	rs7971738	rs1494505	273	1.914	550K	12q14 locus	26	paternal	-
43852	19q12-q13.11	Gain	32470668	38405768	rs28862711	rs8106493	1198	5.935	610K	no	56	maternal	-
43852	19qq13.42	Gain	60669721	60731863	rs4801655	rs645973	18	0.062	610K	no	4	maternal	3
36276	10p11.21-p11.22	Loss	33397558	35429849	rs2666260	rs12768019	481	2.032	550K	no	40	paternal	-
35723	15q21.3-q22.1	Gain	55790481	56699484	rs2733619	rs7167240	306	0.909	550K	no	16	maternal	-
44995	9p21.3	Gain	22373816	23221696	rs4452893	rs7021130	232	0.848	610K	no	2	paternal	1
35124	16p13.13	Loss	11300273	12115827	rs8057417	rs4781183	171	0.815	550K	no	23	maternal	-
31166	Xp22.2	Gain	15952583	16694026	rs5934299	rs4831057	90	0.741	610K	no	14	paternal	-
35397	2p11.2	Loss	85766473	86475381	rs17430897	rs6547681	149	0.708	550K	no	20	maternal	-
31468	3q25.2	Loss	155156188	155859208	rs16823342	rs12493149	102	0.703	550K	no	7	paternal	-
40153	5p15.1	Loss	15202009	15885573	rs2591978	rs13158449	119	0.683	610K	no	1	paternal	-
30921	2p25.3	Gain	1069320	1752354	rs4971432	rs2059412	195	0.683	610K	no	13	paternal	-
37684	4q13.3	Gain	72456449	73080742	rs16846308	rs11722798	153	0.624	610K	no	5	paternal	-
45385	5q32	Gain	146223189	146756043	rs7718587	rs2288803	146	0.533	610K	no	10	paternal	-
27734	9p24.1	Gain	7329425	7860017	rs1926381	rs10976636	247	0.530	550K	no	4	maternal	-
38173	21q21.1	Gain	17414349	18006458	rs2824238	rs2897192	130	0.502	610K	no	14	maternal	-
33525	1p21.1-p21.2	Loss	101620433	102075419	rs11164183	rs10493972	136	0.455	610K	no	6	paternal	2
45509	3p21.31	Gain	44914124	45284504	rs6768893	rs17077548	104	0.370	610K	no	10	maternal	-
37145	8q23.3	Gain	113036795	113351971	rs11775565	rs16883344	34	0.315	610K	no	5	paternal	-
35224	7q36.1	Gain	151431161	151670149	rs6464201	rs10279901	20	0.239	550K	no	4	paternal	1
35929	14q23.1	Gain	63050258	63284561	rs964955	rs8020821	20	0.234	610K	no	13	paternal	1
35040	11p14.2	Loss	26472778	26643141	rs11029583	rs10501063	72	0.170	550K	no	3	maternal	-
32526	12q23.1	Gain	98874747	99019557	rs7313123	rs17029986	24	0.145	550K	no	3	maternal	-
33053	5q33.3	Gain	159297964	159431165	rs11738480	rs4429888	20	0.133	550K	no	5	paternal	-
32369	2q37.3	Gain	238462548	238588384	rs302676	rs2121473	35	0.126	550K	no	2	maternal	1
36783	16p13.13	Loss	11283530	11404080	rs949429	rs4781083	31	0.120	610K	no	5	paternal	1
35170	15q21.3	Loss	53960438	54081109	cnvi0017954	rs16976674	34	0.120	610K	no	3	paternal	-
36580	10q23.33	Loss	96395319	96511239	rs12772169	rs17885857	27	0.115	610K	no	3	maternal	1
45118	2q24.3	Loss	166994290	167104384	rs3791253	rs1447218	22	0.110	610K	no	1	maternal	-
37194	12q24.12	Gain	110683463	110790548	rs10849969	rs11066054	18	0.107	610K	no	4	paternal	-
45385	12q24.11	Gain	108037232	108136783	rs34263	rs12831686	37	0.099	610K	no	4	maternal	1
34254	12q24.11	Gain	108049154	108133395	rs1642033	rs11613784	33	0.084	550K	no	3	paternal	1
35124	19p13.2	Gain	12269459	12356022	rs403386	rs8105975	12	0.086	550K	no	4	maternal	-
40271	1p21.3	Loss	98088873	98168469	rs4421623	rs4378243	18	0.079	610K	no	4	paternal	4
37373	12p13.31	Gain	9242915	9322669	rs2889718	rs6487735	31	0.079	610K	no	2	paternal	-
36183	10q26.11	Gain	121031723	121106202	rs10510056	rs17098766	20	0.074	550K	no	4	maternal	-
35740	8p21.3	Loss	19259291	19324722	rs11995233	rs11204039	27	0.065	550K	no	5	maternal	1
42742	19q13.43	Gain	61727824	61797304	rs11667052	rs7251367	24	0.069	610K	no	3	paternal	1
37394	2p16.2	Loss	54212021	54271149	rs10206954	rs843672	18	0.060	610K	no	2	maternal	1
34927	10q24.1	Loss	98112324	98165281	rs10736104	rs11188751	18	0.053	550K	no	3	maternal	-
40633	3q21.2	Loss	126025793	126067477	rs4678169	rs4234222	13	0.042	610K	no	1	paternal	-
32794	3q21.1	Loss	124129518	124170860	rs2276774	rs9816759	25	0.041	550K	no	2	maternal	-
44399	Xq13.1	Loss	69341389	69381997	rs5980905	rs5936869	6	0.040	610K	no	4	maternal	-
44171	3q13.13	Loss	110366760	110400929	rs1920028	rs1163397	7	0.034	610K	no	1	paternal	4
42940	17p11.2	Loss	19906179	19938969	rs16960660	rs7215803	10	0.033	610K	no	1	paternal	3
34632	6p22.1	Loss	26541308	26571639	rs1624440	rs13195509	8	0.030	550K	no	6	maternal	2
35580	16q22.1	Gain	65863014	65885954	rs7186310	rs3868143	7	0.023	550K	no	5	paternal	2
31160	11q13.4	Gain	73044399	73066760	rs3741147	rs3825003	7	0.022	550K	no	6	paternal	-
43036	2q33.3	Loss	208741802	208760548	rs7589531	rs12986551	8	0.018	610K	no	1	maternal	4
32546	11q12.3	Loss	61485639	61502845	rs195156	rs10897193	6	0.017	550K	no	5	maternal	1
40945	22q12.3	Loss	34917148	34930237	rs6000172	rs132739	7	0.013	610K	no	3	maternal	-

4.1.1.5 Diagnostic yield of MR cohort

We performed SNP array analysis on DNA from 330 patients with mental retardation, most of them of unknown origin and an apparently balanced karyotype, to search for potentially pathogenic submicroscopic CNVs with a commercially available SNP array platform (Illumina). We disposed parental samples from 78.5% of these patients (259 trios and 71 non-trios). To distinguish clinically relevant copy-number alterations from normal polymorphisms, phenotypically normal parents were tested using microarrays or targeted qPCRs. These parental samples provided information on the inheritance of the selected copy-number changes.

34 chromosomal alterations detected in 28 individual patients were not present neither in DNA of parental samples nor in DNA from of the 4104 ancestrally matched controls. 2 compound heterozygous mutations associated with Cohen syndrome, 1 unbalanced translocation and 1 known microdeletion syndrome, the last two without *de novo* confirmation, were also identified in our cohort. Interestingly, 2 of the 3 unbalanced translocations showed a low-level chromosomal mosaicism.
Some of the discovered variants occurred at well-established disease loci, including 12 microdeletion syndromes (Holoprosencephaly, 1q42-q44 deletion, Prader-Willi syndrome, 16p11.2-p12.2 deletion, 16p11.2 deletion, 14q11.2 deletion, Neurifibromatosis type I, Di-George syndrome, *FOXP2* gene deletion, 2q37 monosomy and Cohen syndrome (2)) and 1 microduplication syndrome (5p13 duplication). Additional 16 CNVs were observed to arise *de novo*, fact that indicates that these chromosomal alterations represent most likely clinically relevant CNV.
In summary, our final study allowed the identification of 13 known syndromes, 3 unbalanced translocation and 16 potentially pathogenic CNVs.
For other 21 imbalances, *de novo* occurrence could not be established as parental samples were not available. However, these alterations were never observed in any of the other parents tested neither in the dataset of 4104 control individuals. 7 of these CNVs without parent-of-origin analysis were located at loci whose haploinsuficiency is known to cause MR. We also detected 1 maternally inherited Xq13.1 deletion in a male patient which his mother showed to have a non random X inactivation pattern.
To conclude, the diagnostic yield of this study was of 9.7% (and up to 16.4%) and was found to be in accordance with previous published articles (de Vries et al. 2005, Friedman et al 2006, Vermeesch et al 2007, Hoyer et al 2007, Hochstenbach et al 2009). Furthermore, this study allowed the identification of 137

non-pathogenic variants not described before, including 95 novel polymorphic changes and 42 CNVs overlapping <50% with a DGV entry (data not shown).

4.1.2 SELECTED CASES REPORT

4.1.2.1 Heterozygous deletions of *FOXP1* gene

Here I present the identification of overlapping heterozygous deletions affecting the *FOXP1* (MIM# 605515) gene in three unrelated patients with MR and significant speech and language disorder (Horn and Kapeller et al 2010).

In a collaborative effort we performed a genome-wide microarray scan for CNVs in a German cohort of 1523 unrelated patients with unexplained mental retardation. The recruitment of patients was part of the German Mental Retardation Network (MRNET) study. CNV data were generated in different German institutions within the MRNET consortium using different array-based technologies. To pool and share array-based data a database was set up and made freely available to the entire consortium (www.german-mrnet.de).

These copy number studies revealed three heterozygous overlapping deletions at chromosome 3p14.1 affecting solely the forkhead box P1 (*FOXP1*) gene. Patient 1 was part of the cohort of 330 patients here reported, investigated with Infinium Human550-Quad and Human610-Quad arrays (Illumina). Patient 2 was part of a cohort of 188 patients investigated using whole genome oligonucleotide 244K arrays (Agilent Technologies, Santa Clara, CA). Patient 3 belonged to a cohort of 184 patients analyzed with genome-wide human SNP 6.0 arrays (Affymetrix, Santa Clara, CA). All three patients, two males and one female aged between 5.5 to 7 years, had moderate mental retardation and significant language and speech deficits.

4.1.2.1.1 Clinical Data

Detailed clinical investigation of the three patients revealed that they present moderate MR in combination with a general developmental delay (Table 4-7). In all patients, the non-verbal IQ score was assessed as ≤ 50 (3 SD below the mean). Speech and language development was estimated in the same range. All three patients started to speak at age of 3.5 years and used only combinations of two words at ages 5, 5.5 and 7 years, respectively. In all patients, the productive and receptive vocabulary came up to less than 100 words.

Expressive language was more affected than receptive abilities. Dysgrammatism and very poor speech articulation with difficulties producing consonants at the beginning of words was present in all patients. Two showed oromotoric problems including difficulties with lip protrusion. In infancy, patients had a tendency to keep their mouth open and patient 2 suffered from swallowing difficulties. In addition, all showed considerably retarded gross-motor development with unsupported walking as late as 24 to 36 months. Brain magnetic resonance imaging and electroencephalography did not reveal any abnormalities. None of the patients showed sensorineural hearing loss. Ophthalmologic testing disclosed moderate myopia only in patient 1. Upon physical examinations, normal growth parameters regarding height and occipitofrontal head circumference were documented for two patients but presence of obesity was striking in both. Consistent craniofacial anomalies seen in patients 2 and 3 included broad and prominent forehead and frontal hair upsweep (Figure 4-5).

Table 4-7: Clinical description of *FOXP1* deletion patients

	Patient 1	**Patient 2**	**Patient 3**
General features			
Age of last assessment	7 years	5.5 years	6 years
Sex	Male	Female	Male
Occipitofrontal head circumference (SD)	+1.2	+0.8	+1.3
Height (SD)	+1.4	+0.6	-0.4
Weight (SD)	+0.5	+2.7	+2.5
Facial gestalt	no dysmorphisms	Prominent Forehead, Frontal hair upsweep	Prominent Forehead, Frontal hair upsweep
Non-verbal performance IQ	<50	<50	50
Gross motor delay	+	+	+
Age at sitting	n.d.	12 months	8 months
Age at walking	24 months	36 months	24 months
Swallowing difficulties	n.d.	+	-
Oromotor problems (e.g. lip protusion, tongue elevation)	+	+	-
Speech and language development			
First vocalizing at age of	n.d.	4 months	12 months
First words at age of	3.5 years	3.5 years	3.5 years
Combined words at age of	7 years	5 years	5.5 years
Articulation problems	+	+	+
Poor grammar	+	+	+

n.d., not documented

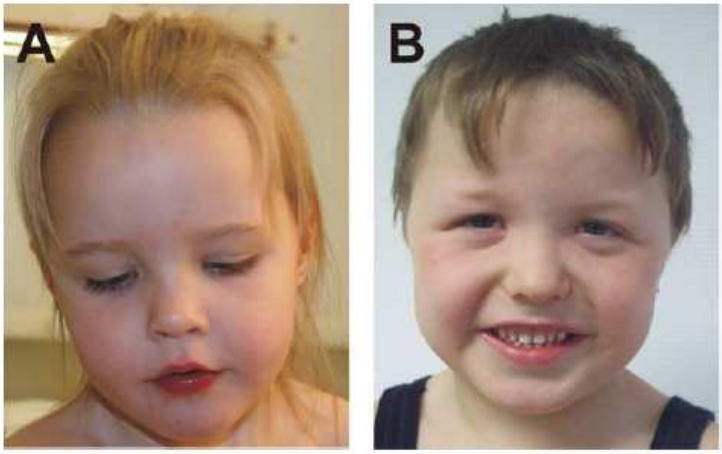


Figure 4-5: Facial Phenotype of patients with *FOXP1* deletion. Consistent facial features of patients showing broad and prominent forehead, frontal hair upsweep, and a short nose. (A) Patient 2 at the age of 4 years and 3 months. (B) Patient 3, at the age of 6 years, had in addition sparse lateral eye brows and down-slanting palpebral fissures.

4.1.2.1.2 Genetic Analysis

Microarray analysis showed that Patient 1 carries a 498 kb deletion affecting all but the first of the coding exons in the *FOXP1* gene (Wagenstaller et al 2007). This deletion, indicated by 89 SNPs, was confirmed by qPCR but its *de novo* origin could not be demonstrated since the paternal DNA was missing. However, it could be shown that the mother does not carry the deletion and that the deleted allele was of father origin. Sequencing of the coding region of the remaining *FOXP1* allele did not reveal any sequence variations compared to the reference sequence (NM_032682). We narrowed-down the deletion breakpoint by quantitative PCR and amplified and sequenced the junction fragment by PCR on genomic DNA of the patient (GenBank Accession No. EF504249). The deletion was sized 498,198 bp and located at position Chr3: 70,807,767-71,305,965 of the hg18, UCSC Human Genome March 2006. The distal breakpoint was found within a long interspersed nuclear element (LINE) repeat (L1/L1PA5) downstream to *FOXP1* and the proximal breakpoint was found in a MER1-type repeat (MER58A) within the intron 6. No recombination stimulating homologous sequences were found inside or around the breakpoint.

For the patients 2 and 3 we identified deletions of 659 kb and 1047 kb affecting the entire coding region of the *FOXP1* gene. The deletions were verified by fluorescence *in situ* hybridization (data not shown). DNA analysis of the unaffected parents by molecular karyotyping and FISH indicated that both patients carried *de novo* deletions. The parent-of-origin of the deletions was paternal in both cases. Sequencing of the

coding region of the remaining *FOXP1* allele did not reveal any sequence variations. Characterization of the breakpoints was achieved by amplification and sequencing of the respective junction fragments (GenBank Accession No HM124444, and GU980955). No repetitive sequences were present at the breakpoint loci in patient 2. However, the two breakpoints were located within a microhomology of three nucleotides (AGA) at the telomeric side at position 70,778,067-70,778,070 and the centromeric side at 71,437,354-71,437,357 resulting in a deletion size of 659,287 bp.

For patient 3, the telomeric breakpoint was located at position 70,341,246-70,341,247 and the centromeric breakpoint at position 71,388,173–71,388,174 resulting in a deletion size of 1,046,927 bp. All positions are given according to the hg18, UCSC Human Genome March 2006. The distal breakpoint of the deletion in patient 3 was located in a short interspersed nuclear element (SINE) while repetitive sequences were missing at the proximal breakpoint.

The absence of homologous regions at the breakpoints suggests that all three deletions were most likely generated by double-strand breaks and non-homologous end-joining.

.

We checked the *FOXP1* locus for CNVs in both the Database of Genomic Variants and an in-house database containing CNV data from 4104 ancestrally matched controls. No CNVs affecting the coding region were found in the DGV database, but a 1.3 Mb deletion at 3p14.1p13 affecting *FOXP1*, *EIF4E3* (MIM# 609896), *PROK2* (MIM# 607002) and *GPR27* (MIM# 605187) was detected in an individual of our control database. Application of Fisher's exact test on these data, not considering the controls contained in the DGV database, resulted in a nominal p-value of 0.06. Although *de novo* origin of the imbalances suggests a role of *FOXP1* in the clinical picture of these patients, it is necessary more population-genetic studies to elucidate whether or not incomplete penetrance applies in this chromosomal interval. However, three additional publications have confirmed that haploinsuficiency of *FOXP1* may cause severe language impairment and mental retardation (Pariani et al. 2009, Hamdan et al. 2010, Carr et al 2010).

Table 4-8: Summary of FOXP1 variants identified in controls and in patients with mental retardation

Variants			MR patients				Controls		
			Genotypes			Status of inheritance	Genotypes		
			11	12	22		11	12	22
Non-synonymous variants									
c.13T>C	p.Ser5Pro	Exon 6	822	1	0	maternal	674	2.	0
c.226-228dupCAG	p.Gln76dup	Exon 7	822	1	0	maternal	676	0	0
c.301A>G	p.Met101Val	Exon 8	822	1	0	n.a.	667	0.	0
c.643C>G	p.Pro215Ala	Exon 10	822	1	0	maternal	336	2	0
c.701T>C	p.Ser261Pro	Exon 11	822	1	0	n.a.	676	0	0
c.1168A>T	p.Thr390Ser	Exon 15	822	1	0	maternal	675	1	0
c.1709A>G	p.Asn570Ser	Exon 19	822	1	0	paternal	338	0	0
c.1790A>C	p.Asn597Thr	Exon 20	822	1	0	n.a.	676	0	0
Synonymous variants									
c.768G>A	p.Thr256Thr	Exon 11	822	1	0	n.a.	673	0	0
c.1188G>A	p.Ser396Ser	Exon 15	822	1	0	maternal	674	0	0
c.1790A>C	p.Asn505Asn	Exon 17	822	1	0	maternal	674	0	0
Non-coding variants							.		
c.1-5G>A		5'UTR	882	1	0	paternal	674	0	0
c.180+49T>C		Intron 6	80	30	1	not tested (dbSNP rs2037474)			
c.181-29G>A		Intron 6	875	7	1	1 x mat, 2 x pat, 1x mat + pat, 4 x n.a.	653	21	0
c.181-30C>T		Intron 6	822	1	0	na	676	0	0
c.664+11A>G		Intron 10	822	1	0	maternal	674	0	0
c.664+6C>T		Intron 10	822	1	0	n.a.	674	0	0
c.975-14A>G		Intron 12	822	1	0	maternal	673	1	0
c.1889+20A>C		Intron 20	95	15	1	not tested (dbSNP rs7638391)			
c.1890-15G>T		Intron 20	110	1	0	not tested (dbSNP rs7639736)			

In order to support further the idea that *FOXP1* deletions can cause mental retardation and active speech deficits we screened 883 patients from the MRNET database for point mutations in *FOXP1*. We identified eight different non-synonymous, three synonymous and nine non coding variants. All variants were genotyped in up to 676 unrelated healthy controls. Three of the identified coding variants, p.Ser5Pro, p.Pro215Ala and p.Thr390Ser were also present at similar frequencies in healthy matched controls, but none of the coding variants was found in the HapMap or the NCBI dbSNP database. Five of the non-synonymous variants, p.Ser5Pro, p.Gln76dup, p.Pro215Ala, p.Thr390Ser and p.Asn570Ser were transmitted through an apparently unaffected parent. For the resting three variants, p.Met101Val, p.Ser261Pro, p.Asn597Thr, we could not establish whether the origin is *de novo* due to the lack of parental DNA (see table 4-8).

Variants transmitted by an unaffected parent are usually classified as non-pathogenic. Still, presuming a multigenic threshold model for MR, disease relevance cannot be completely ruled out. To uncover the neurobiological significance of these variants, investigation of the functional effects of the putative mutations is needed.

De novo point mutations in FOXP1 have been reported in other studies to be causative for severe language impairment and mental retardation (Hamdan et al. 2010).

4.1.2.2 Cohen Syndrome Diagnosis

Here I present the identification of intragenic heterozygous deletions affecting the *COH1* (MIM#607817) gene in three unrelated patients with unexplained MR. Subsequent sequencing of the *COH1* gene revealed point mutations in the second allele in all three patients analyzed, thus providing Cohen Syndrome diagnosis (Rivera-Brugués et al. 2010).

We have analysed high density oligonucleotide array data from patients with unexplained mental retardation (n=1523) and normal controls (n=1612) for copy number variation (CNV) changes.

Recruitment of patients has been part of the German Mental Retardation Network (MRNET) study (http://www.german-mrnet.de/). As a control set, we used CNV data (n=1612) from the population based KORA study (Cooperative Health Research in the Augsburg Region).

In the present study we focus on three patients, in which CNVs in *COH1* gene were detected. Patient 1 and Patient 2 were part of the cohort of 330 patients previously reported, investigated with Infinium Human550-Quad and Human610-Quad arrays (Illumina). Patient 3 was part of a cohort recruited by the institution of Essen and was analysed with genome-wide human SNP 6.0 arrays (Affymetrix, Santa Clara, CA). All three patients, two males and one female aged between 18 months and 3 years, had severe mental retardation and developmental delay.

4.1.2.2.1 Clinical Data

The patient 1 was born after an uneventful pregnancy at term as the first child of unrelated and healthy Arabian parents. Birth weight, length and head circumference were not recorded. A delay in motor development became evident within his first year of life. Sitting started at the age of 18 months, walking at 24 months. A detailed examination at the age of 3 years showed a hypotonic boy with a height of 79cm (-

4.8SD), a weight of 9300g (-5.2SD) and a head circumference of 44cm (-4.9SD). A delay in speech development as well as in comprehension was obvious. There was mild craniofacial dysmorphism including horizontal eyebrows, a broad and downturned nasal tip, a broad columella, a short philtrum, a thin upper lip, and an everted lower lip (fig 4-6). The ophthalmologic examination revealed bilateral myopia, astigmatism and a slightly increased pigmentation of the retina. There was no neutropenia.

The patient 2 was born at term after an uneventful pregnancy as the second child of healthy unrelated German parents with a birth weight of 2460g (-2.4SD), a birth length of 46.5cm (-2.3SD), and a head circumference of 31.5cm (-2.9SD). Developmental milestones were delayed with a sitting age of 12 months and a crawling age of 17 months. Examination at the age of 18 months showed a hypotonic toddler with a weight of 10.1kg (-1.5SD), a height of 80cm (-1SD) and a severe microcephaly with an OFC of 42.5cm (-4.8SD). Speech development had not occurred but comprehension was nearly normal. Facial dysmorphism consisted of mild occipital flattening, horizontal eyebrows, almond shaped palpebral fissures, a broad nasal root, a round nasal tip, a thick columella, a short philtrum, an open appearance of the mouth with a prominent upper gingiva, and a large gap between the incisors (fig 4-6). Fundoscopy and complete blood count revealed no abnormalities.

The patient 3 is the second child of a healthy non consanguineous German/African couple. The premature birth occurred at 35 weeks of gestation with a birth weight of 2450g (-0.1SD), a birth length of 46cm (-0.4SD) and an OFC of 33.5cm (0.6SD). A heart defect (ASD II and pulmonary stenosis) and arrhythmia were diagnosed after birth. Developmental delay persisted until the 2 years of age and improved after surgical correction of the heart defect. The patient started to walk at the age of 24 months and at the age of 2 ¾ years she showed normal body measurement with a height of 95cm (0.4SD), a weight of 11.5kg (-1.4SD) and an OFC 50cm (0.6SD). She had a flat face with broad and flat nasal bridge and almond shaped eyes, a short philtrum with thin vermillion border and deep set ears with overfolded helices (fig 4-6). Fundoscopy and complete blood count revealed normal results.

Table 4-9: Clinical description of Cohen syndrome patients

	Patient 1	Patient 2	Patient 3
General features			
Age at assessment	3 years	18 months	2 ¾ years
Sex	Male	Male	Female
Ethnic origin	Arabian	German	German/African
Occipitofrontal head circumference[a] (SD)	-4.9	-2.9	+0.6
Height (SD)	-4.8	-2.3	-0.4
Weight (SD)	-5.2	-2.4	-0.1
Psychomotor retardation[a]	+	+	+
Delay in speech development	+	+	+
Obesity with slender extremities[a]	-	-	-
Cheerful disposition[a]	+	n.d.	-
Joint laxity[a]	-	-	-
Progressive myopia/retinopathy[a,b]	+	-	-
Characteristic Facial Dysmorphis[a,b]			
Thick hair and eyebrows	-	-	-
Downward slanting palpebral fissures	-	-	-
Almond shaped palpebral fissures	+	+	+
Prominent nose	+	+	+
Short philtrum	+	+	+
Hypotonic facial expression	+	+	-
Open mouth	+	+	-
Downturned corners of the mouth	-	+	-
Thick lower lip	+	-	-

[a] clinical criteria for Cohen syndrome provided by Kohlemainen et al. 2004; ; [b] clinical criteria for diagnosis of Cohen syndrome provided by Chandler et al. 2003; +, presence; -, absence; n.d., not documented

Figure 4-6: Craniofacial Phenotype of CS patients 1 to 3. Patient 1 with horizontal eyebrows, a broad and down turned nasal tip, a broad columella, a short philtrum, a small upper lip and everted lower lip; Patient 2 with horizontal eyebrows, almond shaped and downslanting palpebral fissures, a broad nasal root, a round nasal tip, thick columella, a short philtrum, open appearance of mouth with prominent upper gingiva; Patient 3 with round and flat face, bushy eyebrows with lateral flaring, broad nasal bridge, short philtrum and microtia with overfolded helices;

4.1.2.2.2 Genetic Analysis

Analysis of SNP oligonucleotide array and subsequent qPCR discovered CNVs in the *COH1* gene in three patients: a maternal inherited 67kb deletion encompassing exons 26 to 31 of the *COH1* gene (chr8: 100,573,090 …100,639,924; c.3871-5024del/p.G1291fsX42) in patient 1, a paternal inherited 193kb deletion encompassing exons 9 to 19 of the *COH1* gene (chr8: 100,216,034 …100,409,167; c.1207-2824del/p.L403fsX11) in patient 2 and a maternal inherited 315 kb deletion encompassing exon 4 of the *OSR2* gene and exons 1 to 17 of the *COH1* gene (chr8:100,015,029...100,347,846; c.1-2515del) in patient 3 (see figure 4-7). There were no such changes in the 1612 controls.

Sequencing of the *COH1* gene identified in patient 1 a paternal inherited missense mutation in exon 32 leading to a stop codon (c.5086C>T/p.R1696X), in patient 2 a maternal 1bp deletion in exon 60 leading to a stop codon (c.11505delA/p.K3835fsX43) and in patient 3 a heterozygous missense mutation in exon 25 (c.3866C>G/p.T1289S) and a heterozygous three base pairs insertion in exon 62 (c.11827_11828insATG/p.D3942_G3943insD), both inherited from the father (see table 4-10). These mutations were not present in 676 samples from a general population cohort (KORA) and not annotated as polymorphisms in HapMap or the NCBI dbSNP database.

Table 4-10: Summary of the mutations in the *COH1* gene

Patient ID	Nucleotide Exchange	Exons	Position UCSC hg18	Amino acid change/ Predicted consequence	Origin	Detection Method
Patient 1	c.5086C>T	32	chr8:100657123	p.R1696X	father	Sequencing
Patient 1	c. 3872-5024del	26-31	chr8:100573090-100639924	p.G1291fsX42	mother	SNP-Chip
Patient 2	c. 11505delA	60	chr8:100952228	p.K3835fsX43	mother	Sequencing
Patient 2	c. 1207-2824del	9-19	chr8:100216034-100409167	p.L403fsX11	father	SNP-Chip
Patient 3	c.3866C>G	25	chr8:100563202	p.T1289S	father	Sequencing
Patient 3	c.11827_11827insATG	62	chr8:100956824-100956826	p.D3942_G3943insD	father	Sequencing
Patient 3	c.1-2515del	1-17	chr8:100015029-100347846	-	mother	SNP-Chip

Gene model NM_17890.3/NP_060360 based on hg18, UCSC browser March 2006

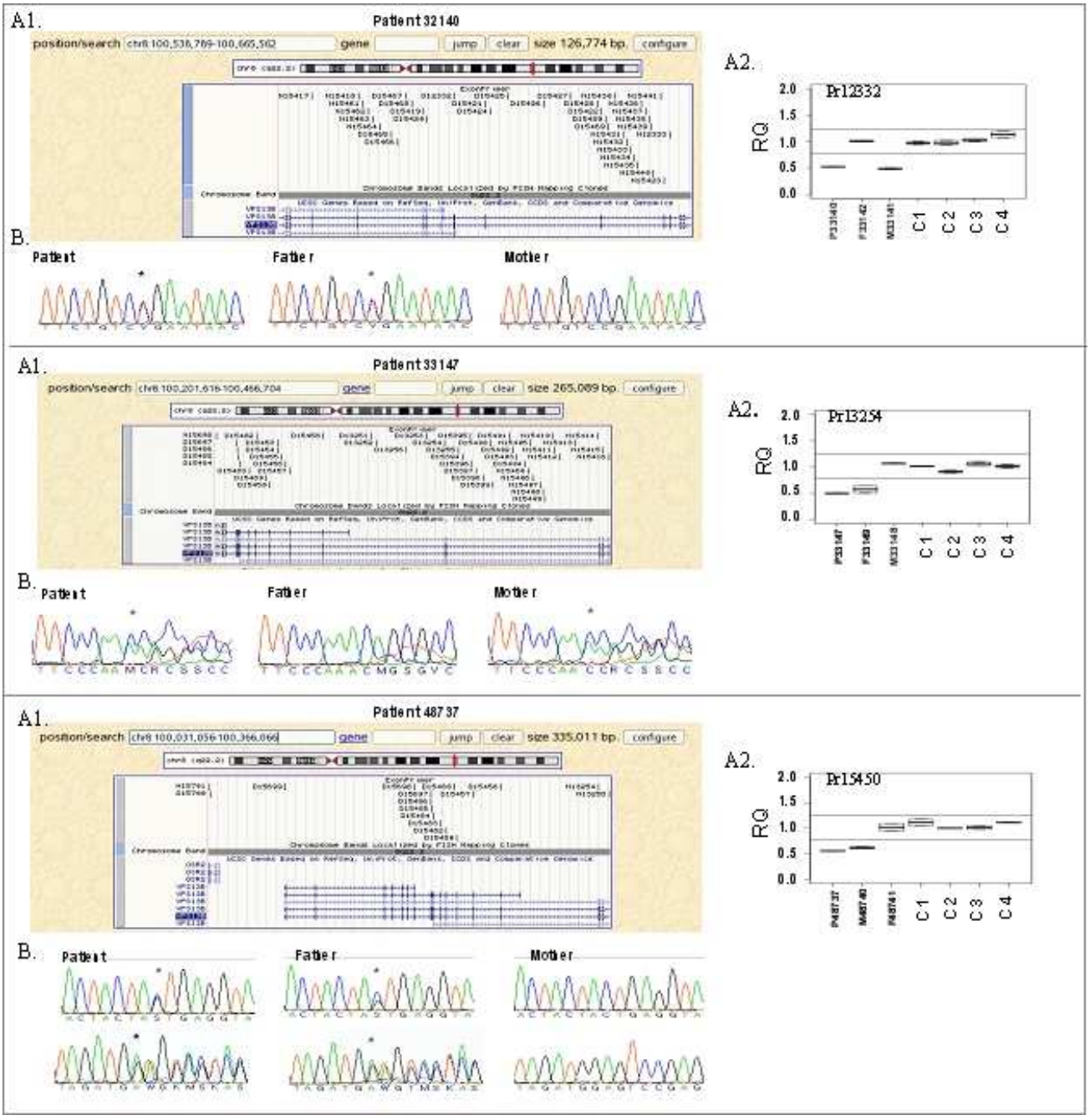


Figure 4-7: Molecular analysis of the *COH1* gene in three families with Cohen syndrome. (A) Intragenic heterozygous deletions encompassing 6, 11 and 17 exons of the *COH1* gene detected with SNP Chips. ***A1.*** Extend of the deletions and the positions of the primer pairs used for CNV validation shown within the UCSC genome browser. ***A2.*** Results of a representative qPCR. Shown are the RQ values of the patient (P), the father (F), the mother (M) and 4 controls (C1-C4). **(B)** Electropherograms of the *COH1* mutations; c.5086C>T, c. 11505delA, c.3866C>G and c.11827_11828insATG

4.2 CONGENITAL HYPERINSULINISM

Here I present the identification of three (up four) distinct chromosomal imbalances in five patients with Congenital Hyperinsulinism (CHI) to be most likely causative of the disease. The screening for point mutations in the candidate gene *PTGER3* did not reveal any pathogenic variant neither in the second allele of the patient in which a *de novo* deletion was detected nor in a cohort of 39 unrelated patients with unexplained CHI.

Congenital hyperinsulinism (CHI) is a heterogeneous condition with mutations in eight different genes described. The genetic basis of CHI involves defects in key genes which regulate insulin secretion from beta-cells. However, mutations in all these genes account for about 50% of the known causes of CHI. There are three major subtypes of CHI: diffuse, focal and atypical forms. The diffuse form is inherited mainly in an autosomal recessive fashion and the focal form is sporadic in inheritance.
Here we performed a genome-wide microarray scan for CNVs in 40 children referred for difuse congenital hyperinsulinism. As the conventional mutation screening in *ABCC8* and *KCNJ11* genes did not reveal any detectable abnormality, the patient DNAs were hybridized on Human610-Quad arrays (Illumina).

4.2.1 Genetic Analysis

We built a cohort of 40 patients with diffuse congenital hyperinsulinism and inconspicuous mutation screening of *ABCC8* and *KCNJ11* genes. 23 of these patients had been screened for indels, cryptic splice sites and promoter mutations in the entire genomic region of *ABCC8* by deep sequencing on an Illumina GAII system. Furthermore, the exonic regions of the *ABCC8* gene had been checked for deletions and duplications by quantitative real-time PCR in all 40 patients. 8 cases missed a mutation in one allele and no mutations were found in 32 cases. Testing by SNP arrays was applied in order to provide a more comprehensive screening of the patients for mutations across the genome and thus to identify new genes causative of CHI.
Analysis of SNP oligonucleotide arrays and subsequent qPCR validation revealed four chromosomal alterations in five patients with difuse CHI.

Two brothers (ID; 45743 and 45744) showed a similar *de novo* complex rearrangement, corresponding to two large deletions on the p-arm of chromosome 9 and one large duplication on the chromosome 13. The implication of two different chromosomes and the occurrence of the same type of rearrangement on the two affected brothers indicate that these patients carry an unbalanced translocation between chromosomes 9 and 13. This translocation was transmitted by one of the healthy parents carrying the reciprocal translocation t(9;13) balanced. To confirm this, standard chromosome analysis on both children and their parents is required. We observed that the breakages of the CNVs detected on chromosome 13 are almost identical between the two brothers, resulting in both cases in a 7,85 Mb duplication of 13q33.3-q34. However, the CNV losses from chromosome 9 differ slightly between the two kids. Patient 45743 presents a 6.3 Mb deletion of 9p24.3-p24.1 and a 4 Mb deletion of 9p24.1-p23, whereas patient 45744 shows a 4.3 Mb deletion at 9p24.3-p24.2 and a 7.8 Mb deletion at 9p24.1-p23. DNA of the patient 45744 was analyzed a second time with SNP-chips and in this analysis we only identified two large imbalances, a single 12,4 Mb deletion on chromosome 9 and a 7,9 Mb duplication on chromosome 13. The oligonucleotide chip data were used to initially pinpoint approximate breakpoint positions in the genome. This technology has a limitation of resolution at a base-pair level. Moreover, the inaccuracy of the calls depends to a large degree on the design of the array (i.e. SNP coverage), the algorithm used (i.e. different thresholds), and the experimental variability. Therefore, other techniques such as qPCR or MLPA and junction fragment amplification will be necessary to map accurately the breakpoints and determine the exact extension of these rearrangements.

Another patient (30370) with unexplained difuse congenital hyperinsulinism showed to have also a complex rearrangement. In this case, we detected two CNVs on chromosome X, a 2.7 Mb deletion at p22.33 and a 16.2 Mb duplication on q11.1-q22.1. Due to the lack of parental samples we could not investigate the *de novo* origin of these variants.

More than 300 Ensembl genes are annotated in the CNVs identified in the two brothers, and 426 genes in the altered regions of the latter patient. None of the genes embedded were found relevant to CHI, based on available data (expression in pancreas, function in metabolism, interaction with other disease related genes) from current public databases (UCSC Genome Browser, Ensembl, NCBI) and the literature.

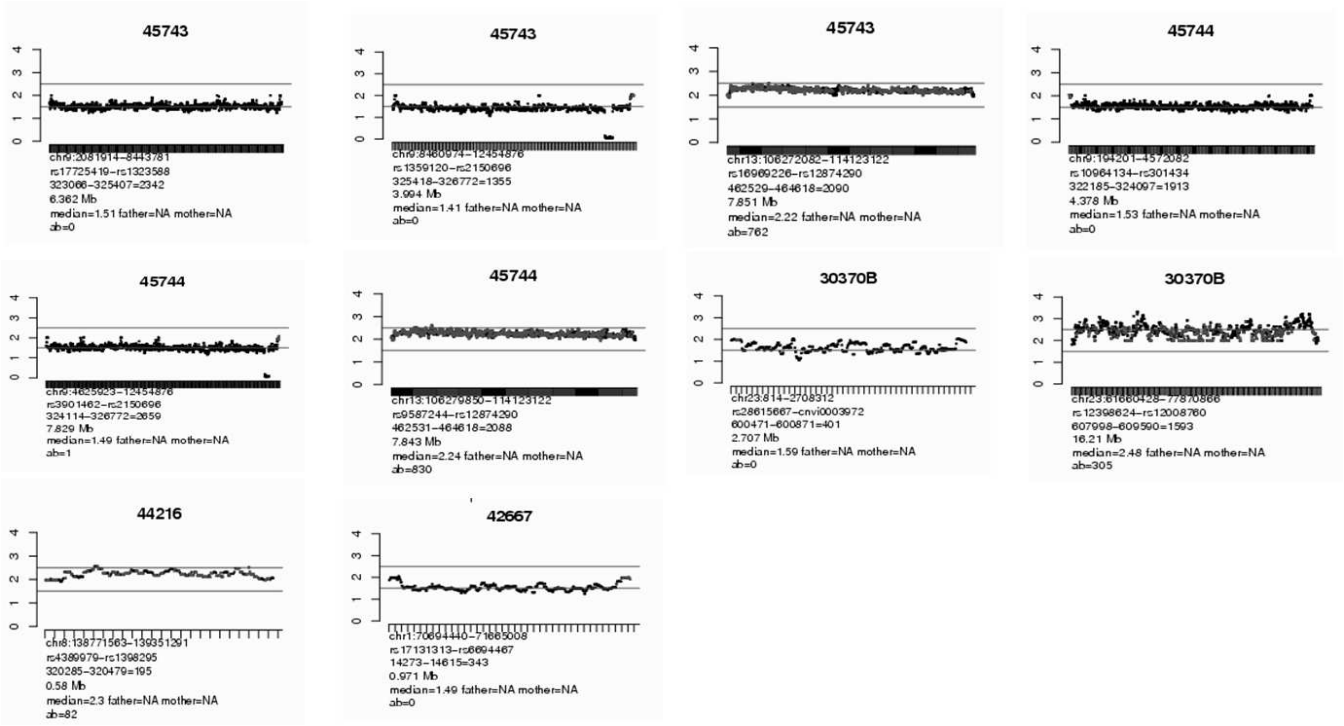
The next patient (44216) presented a duplication of 8q24.23 which extended 579kb and contained 195 SNPs. We could not determine its *de novo* occurrence because the DNA of the parents was not available. The distal breakpoint of the duplication was determined on intron 3 of *FAM135B* gene. Four Ensembl entries are annotated in the region of deletion corresponding to isoforms of *FAM135B*. This gene encodes a hypothetical protein, *LOC51059*, of unknown biological function. No polymorphic CNVs affecting exonic regions of this gene are described to date. However, other studies will be necessary to determine if the gene dosage of *FAM135B* or another gene of the genome is altered by this duplication, thus permitting to assure the pathogenicity of this gain and to identify *FAM135B* as a candidate gene for CHI.

The last rearrangement detected is a 970kb deletion on the small arm of chromosome 1 at the band p31.1 (physical position UCSC Genome Browser hg18; Chr1: 70,694,440-71,665,008). This CNV was detected on patient 42667 and by mean qPCR experiments it was demonstrated to arise *de novo*. This chromosomal copy number variation encompasses the *PTGER3* and *ZRANB2* genes and the last exon of the *NEGR1* gene. *ZRANB2* is a zinc finger protein required for alternative splicing of pre-mRNA. *NEGR1* encodes a neuronal growth regulator which may function as a trans-neural growth-promoting factor in regenerative axon sprouting in the mammalian brain. The third gene embedded in this region is *PTGER3* (prostaglandin E receptor 3) which encodes one of the receptors identified for prostaglandin E2 (PGE2). Prostaglandin E2 is an inhibitor of glucose-stimulated insulin secretion (GSIS) and it is known to act throught the receptor *PTGER3*. The implication of *PTGER3* on the mechanisms of regulation of the insulin has been previously highlighted in the literature. One publication reported the generation of a *PTGER3* knock-out mouse model (Ep3 -/-), which exhibited a phenotype characterized by elevated leptin and insulin levels and over 20% higher body weight compared with wildtype littermates (Sanchez-Alavez et al. 2007). Additionally, functional investigations of Meng et al. (2006) demonstrated that PGE2 can induce pancreatic beta cell dysfunction through the induction of *PTGER3* gene expression, thus altering the signaling pathway and the target transcription factors involved in the insulin secretion. Based on these informations, we hypothesized that the loss-of-function of *PTGER3* might be the cause of the hyperinsulinism suffered for the patient. Therefore, *PTGER3* was selected as candidate gene and our cohort was screened for point mutations by conventional Sanger sequencing.

Table 4-11: Summary of the chromosomal imbalances identified in CHI patients with Human610-Quad arrays (Illumina)

Patient ID	Chromosome band	Gain/ Loss	Position Start hg18	Position End hg18	Start rsSNP	End rsSNP	Number of SNPs	Length (Mb)	Ensemble Genes	Array Type	Known Syndrome	Parent of origin
45743	9p24.3-p24.1	Loss	20,811,914	8,443,781	rs17725419	rs1323588	2,342	6.362	163	610K	translocation-CHI	de novo
45743	9p24.1-p23	Loss	8,460,974	12,454,876	rs1359120	rs2150696	1,355	3.994	18	610K	translocation-CHI	de novo
45743	13q33.3-q34	Gain	106,272,082	114,123,122	rs16969226	rs12874290	2,090	7.851	137	610K	translocation-CHI	de novo
45744	9p24.3-p24.2	Loss	194,201	4,572,082	rs10964134	rs301434	1,913	4.378	74	610K	translocation-CHI	de novo
45744	9p24.1-p23	Loss	4,625,923	12,454,876	rs3901462	rs2150696	2,659	7.828	126	610K	translocation-CHI	de novo
45744	13q33.3-q34	Gain	106,279,850	114,123,123	rs9587244	rs12874290	2,088	7.843	137	610K	translocation-CHI	de novo
30370	Xp22.33	Loss	814	2,708,312	rs28615667	cnvi0003972	401	2.707	52	610K	CHI	ND
30370	Xq11.1-q22.1	Gain	61,660,428	77,870,866	rs12398624	rs12008760	1,593	16.210	374	610K	CHI	ND
44216	8q24.23	Gain	138,771,563	139,351,291	rs4389979	rs1398295	195	0.579	4	610K	CHI	ND
42667	1p31.1	Loss	70,694,440	71,665,008	rs17131313	rs6694467	343	0.970	31	610K	CHI	de novo

Figure 4-8: Copy number values of 10 potentially pathogenic CNVs identified in CHI patients using Illumina 610K arrays (median smoothing with a window of 9 adjacent SNPs). CNVs with copy number values lower than 1.5 were suspicious for deletion and CNVs with copy number values of more than 2.5 were suspicious for duplication



4.2.2 PTGER3 as candidate gene for CHI

We designed primers to amplify amplicons covering the exonic regions and the intron-exon boundaries of all 14 *PTGER3* transcripts. By direct sequencing we looked for point mutations in the remaining *PTGER3* allele of patient 42667 in order to uncover a putative recessive mutation. We also performed mutational analysis of *PTGER3* in the resting 39 patients of our CHI cohort.

Sequencing did not reveal any variant in patient 42667, but allowed us to identify a heterozygous 1bp deletion in exon 4 of *PTGER3* isoform 8 in patient 33531, which leads to a premature translation termination codon after twelve amino acids (c.1185delC / p.N395QfsX12; Gene model NM_198718.1/NP_942011.1 based on UCSC browser, hg18). The identified null allele predicts the protein truncation at the isoform-specific C-terminal domain and may also rise to nonsense-mediated RNA decay. The identified sequence variation was not annotated as a polymorphism in HapMap or the NCBI dbSNP database. However, segregation analysis showed that the variant was inherited from the apparently unaffected mother, thus pointing out its non-pathogenecity. We genotyped the selected variant in a control group of 676 unrelated subjects from the KORA-cohort by the MassARRAY system (Sequenom platform) and the variant was found in 18 individuals. In addition, we evaluated the specific expression of mRNA of the isoform EP3-8 on human pancreas. We performed RT-PCR analysis on human pancreas c-DNA (Marathon Ready cDNA, Clonthech) by using two nested-PCR amplification steps and the combination of three primer sets in order to amplify specific and unspecific *PTGER3* transcripts. We sequenced the RT-PCR amplified product and aligned the nucleotide sequence to BLAST. We confirmed that there is not expression of *PTGER3* isoform 8 (NM_198718.1) in pancreas tissue. Taken together, the genotyping and expression results demonstrated that the novel sequence variant identified in the *PTGER3* gene is not the molecular defect responsible of the hyperinsulinism suffered by the patient 33531. Instead, this variant corresponds to a rare polymorphism present in control population and it might affect solely to the carboxyl termini of the isoform EP3-8 which is not expressed in pancreas.

4.3 DETECTION OF A CONTIGUOUS DUPLICATION OF *TBX5* AND *TBX3* GENES

Here I report a contiguous heterozygous duplication of *TBX5* and *TBX3* resulting in a combined phenotype of Holt-Oram syndrome and supernumerary mammary glands. To our knowledge, this is the first description of *TBX5* gain-of-function caused by a genomic duplication in a family with mild skeletal symptoms and variable expression of cardiac defects. This is also the first time that a duplication including the whole *TBX3* gene is identified and associated with the development of accessory mammary glands.

4.3.1 Clinical Data

This study included a total of 11 individuals from an extended German family, which had five affected members with an initial diagnosis of Holt-Oram syndrome. All patients were previously investigated for the presence of mutations in the *TBX5* gene with negative results. The affected subjects presented anterior upper limb malformations with or without congenital heart defect or cardiac conduction anomalies, and some patients showed additional mammary glands.

The first member of this family possible affected by heart disease was a male born in 1852 (I/1). As we know by personal communication he had neurosis of the heart and was limited in his activities. He died in age of 84 years. One of his two sons was affected by neurosis of heart in a similar manner (II/1). He was exempt from military service. His daughter was healthy (III/2), but his son was affected by hand and heart diseases (III/1). His thumb joints were stiffened on both hands and he complained about heart pain. In age of 62 years he suffered from myocardial infarction. He died at 86 years. Three of his four children have hand abnormalities. One son has stiffed joints of left thumb (IV/1). Additionally, bicuspidal aortic valve was diagnosed, giving a systolic noise. One of his daughters suffers from immobilisation in thumbs on both hands (IV/2). The second daughter has stiffed thumbs joints as well (IV/3). In the next generation five members are affected. The two children of IV/1, twins, have immobile joints on both thumbs. An aortic stenosis was diagnosed in son (V/2), corrected by operation six weeks after birth. IV/3 has two unaffected sons and three of for children from IV/2 were affected. The son (V/3) has a limited extension of fifth finger on both hands. No more finger abnormalities were detected. After detailed investigation an atriovetricular septum defect (ASD II) was detected in age of seven years, followed by an operation. In addition accessory mammary glands were present on the anterior axiliar linie in left and right side of the thorax. His sister (V/4) has an extension deficit in the middle joint of thumb phalanx and a heart defect (ASDII) which was corrected by operation in the first year of her life. She presents accessory mammary gland. The next

daughter of IV/2 is unaffected, but the youngest daughter has complex heart defect (ASD II, stenosis of aorta) and severe contractures and deviation in fingers of both hands. Psychomotor development was normal in each affected individual.

Figure 4-9: Pedrigree of the HOS family

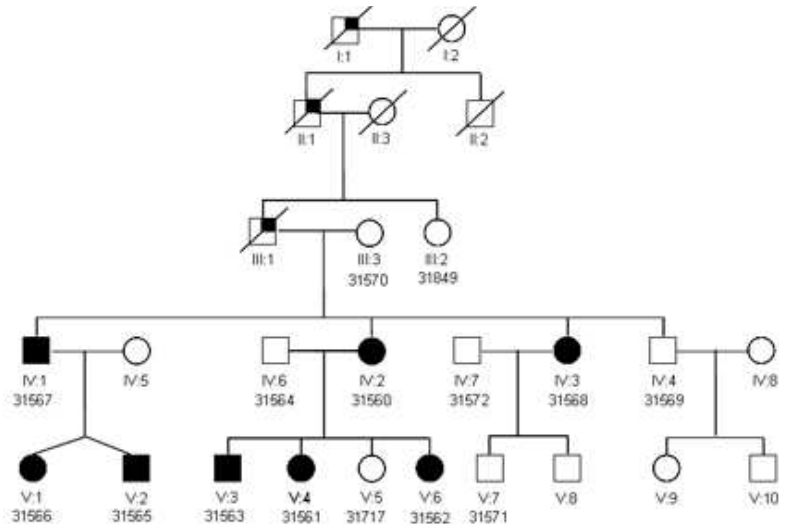


Table 4-12: Summary of clinical features

Patients	ID	Clinical data
I:1	-	cardioneurosis
I I:1	-	cardioneurosis
III:1	-	inflexible thumbs/ deficit in distal phalanx
IV:1	31567	inflexible left thumb/ bicuspidal aortic valve
IV:2	31560	inflexible thumbs/ accessory mammalians
IV:3	31568	inflexible thumbs/ flexion deficits in distal phalanx
V:1	31566	inflexible thumbs/ flexion deficits in distal phalanx
V:2	31565	inflexible thumbs/ clinodactyly/ coarctation aortae
V:3	31563	camptodactyly digit IV/ ASDII/ accessory mammary glands
V:4	31561	deficits in thumb abduction/ flexion deficits in distal phalanx digit VI/ ASDII/ accessory mammary gland
V:6	31562	ulnar deviation of fingers/ flexion deficit of distal phalanx/ deficit in thumbs abduction

Figure 4-10: Skeletal abnormalities and accessory mammalian gland in patients V:3, V:4 and V:6. (A) Family member V:3 (ID 31563) with camptodactyly digit V of left hand and accessory mammalian gland. (B) Family member V:4 (ID 31561) with immobile joints of thumbs, hypoplastic thenar on both hands and accessory mammalian gland. (C) Family member V:6 with ulnar deviation of fingers, limited mobility of thumbs and fingers, hypoplastic thenar and interphalangeal muscels

4.3.2 Genetic Analysis

The index patient (ID: 31565), without detectable mutations in *TBX5* gene, was tested for the presence of cryptic chromosomal copy number variants using a high density SNP array as described in section 2.3.9 and 2.3.10. The patient was found to carry a heterozygous duplication at the chromosomal band 12q24.21, spanning the whole coding region of the *TBX5* gene. This CNV was indicated by the increase of copy number values of 64 SNPs (median 2.37) and encompassed 210kb (chr12:113,275,838…113,486,329 based on UCSC browser, hg18) (See figure 4-11 and table 4-13). 9 additional CNVs were detected with this assay, but were either annotated as polymorphisms or did not contained OMIM disease genes.

CNV validation and breakpoint mapping was performed using quantitative Real Time PCR with SYBR Green as described in section 2.3.11. To confirm the duplication of *TBX5* gene all affected and unaffected individuals were examined by qPCR using amplicons positioned in exons 4 and 9 of the *TBX5* gene. 30 additional primer pairs were designed to fine map the proximal and distal breakpoints in the upstream and

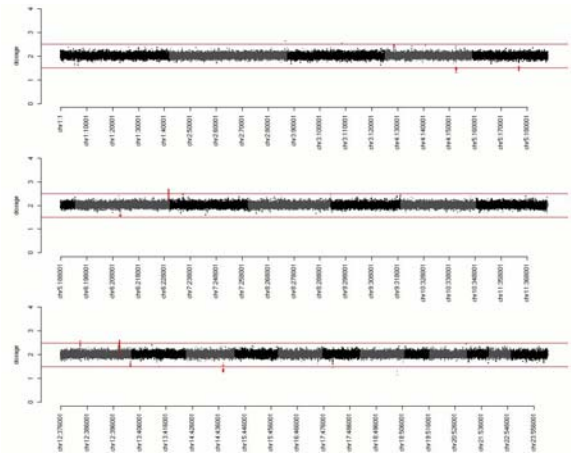
downstream neighboring region of *TBX5* gene. At the centromeric end of the duplication the first preserved amplicon was positioned at 113.46 Mb on the q-arm of chromosome 12, whereas at the telomeric end of the duplication the first preserved amplicon was positioned at 113.60 Mb (Figure 4-12A). Interestingly, the duplication affected not only the complete *TBX5* gene, but also the *TBX3* gene. *TBX3* is approximately 315 kb towards the telomere from the *TBX5* gene, both located at the 12q24.21 band. Therefore, after delineating the breakpoints with qPCR the maximal duplication size could be estimated as aproximately 345,6kb (chr12: 113262450…113608018, UCSC genome build 18), including both *TBX5* and TBX3 genes. Gene dosage assessment at specific genetic loci demonstrated the cosegregation of the duplication and the Holt-Oram syndrome/supernumerary mammary glands phenotype in this pedigree (see figure 4-12B).

Table 4-13: Summary of the chromosomal imbalance identified in the index HOS patient with Human610-Quad array (Illumina)

Patient ID	Chromosome band	Gain/ Loss	Position Start hg18	Position End hg18	Start rsSNP	End rsSNP	Number of SNPs	Length (Mb)	Ensemble Genes	Array Type	Known Syndrome	Parent of origin
31565	12q24.21	Gain	113,275,838	113,486,329	rs7316919	rs17678419	64	0.210	7	550K	TBX5	paternal

Figure 4-11: Genome-wide CNV analysis using Illumina Human550-Quad array (A) SNP-based array results using a 610K-features whole-genome microarray with one probe every 5kb, showing a duplication between the nucleotides 113,275,838 and 113,486,329 at 12q24.21 band (B) Copy number values of the 12q24.21 duplication (median smoothing with a window of 9 adjacent SNPs).

A.



B.

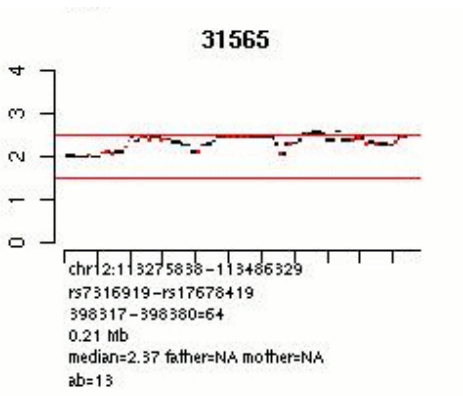
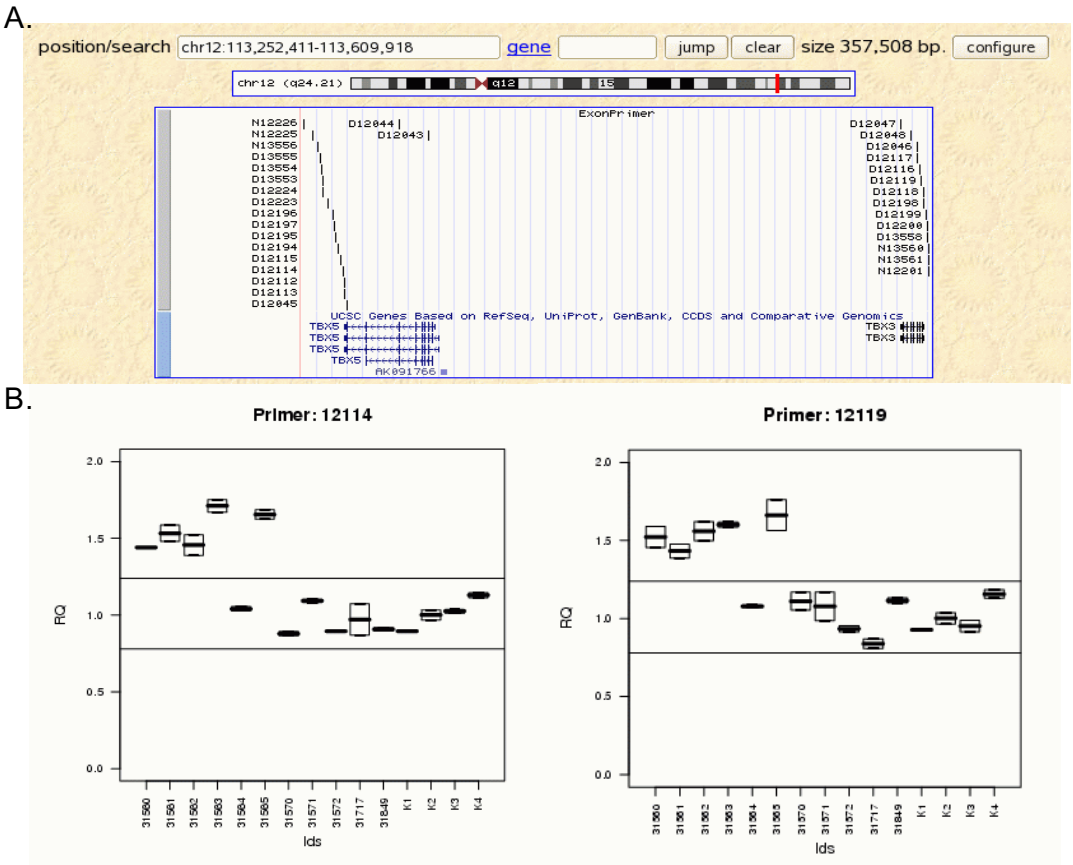


Figure 4-12: Molecular analysis of the *TBX5* and *TBX3* genes. (A) Extend of the duplication and the positions of the primer pairs used for CNV validation shown within the UCSC genome browser. N infront the qPCR primer ID indicates a normal copy number value whereas D indicates a deletion (B) Results of representative qPCR assays used for delineating the breakpoints. Of note, the duplication segragates with the Holt-Oram syndrome/additional mammary glands phenotype.



Here we identified a contiguous heterozygous duplication of *TBX5* and *TBX3* resulting in a combined phenotype of Holt-Oram syndrome and supernumerary mammary glands. The gain of function of *TBX5* causes mild skeletal deformations in the majority of affected patients and congenital heart disease in few of them. A first genotype-phenotype relationship for a *TBX3* gain-of-function mutation is also described. On the contrary to *TBX3* haploinsufficiency, which produces apoplasia or hypoplasia of the breast, the increased dosage of *TBX3* causes the development of supernumerary mammary glands.

Detailed clinical descriptions of each family member can be found in the previous section, 4.3.1. Table 4-12 contains a summary of the main clinical features observed in this family.

5- DISCUSSION

5.1 Methodological considerations

The first goal of this study was to perform a systematic screen of a cohort of 219 sporadic novel patients with intellectual disability of unknown origin in order to evaluate and optimize the SNP-microarray technology for whole genome copy number analysis. For this purpose we used two different Illumina high density SNP microarrays; 1) the HumanHap550 chips to study 109 DNA samples and 2) the updated Human610-Quadv1_B arrays to study a second set of 110 DNA samples.

In order to evaluate the specificity as a function of the resolution (i.e. sensitivity) we determined the rate of false positives for different limits of resolution. This allowed us to set up the threshold at which the detection of false positives is negligible. The low false-positive rate is especially important, as confirmation of aberrations with independent methods is laborious and time-consuming.

By validating the array data with quantitative PCR we observed that for 550K arrays, the false positives rate was 62.4% in the <10 supporting SNPs bin, 36.2% in the 10-20 SNPs bin and 3.1% in the >20 SNP bin, whereas for the 610K dataset the false positive rates were higher in all the SNP bin tested, 82.8% in the <10 supporting SNPs, 52.8% in the 10-20 SNPs bin and 12.2% in the >20 SNPs bin.

We observed therefore that CNVs supported by >20 SNPs were called with a high degree of specificity in both array types; 96.9% for 550K arrays and 87.8% for 610K arrays. Based on these data and applying a conservative security margin we determined as a cuttoff parameter the smoothing median of windows of more than 30-40 consecutive SNPs, i.e. 200kb-500kb. The more or less stringent resolution threshold was applied depending on the quality of the signal intensity profile of each particular array.

Once this window size was established, a third set of 111 DNA samples from patients with unknown ID were hybridised onto Human610-Quadv1_B arrays and only those CNVs indicated by at least 30-40 SNPs were considered for further analysis. In these cases, in order to distinguish clinically relevant copy-number alterations from normal polymorphisms, patients and their phenotypically normal parents were tested again using Human610-Quadv1_B microarrays. It is important to note that if we had initially used the threshold of 30 SNPs in the 219 patients examined in the evaluation step we would not have been able to detect 6 *de novo* and 28 inherited aberrations identified in our study.

The accuracy of a test in detecting chromosomal aberrations, that is, the ability to call true positive events and to avoid both false positives and false negatives is also subject to experimental variability. In order to

maximize accuracy we controlled the noise level. Standard deviation (SD) of log2 test-over-reference autosomal ratios was used as a measure of noise. Only samples with a SD <0.20 and a SNP call rate of >99.9% were included in the analysis. Based on these quality criteria the array analysis of 70 out of 330 (21.2%) patient DNA samples had to be repeated in order to perform an accurate screen of CNVs. We analysed the data quality of the 610K array in comparison to that of the 550K array (see table 4-1 and table 4-2). For 610K arrays the variance of the log2 intensity ratio was larger (mean SD: 0.19 and mean MAD: 0.12) and the signal-to-noise ratio was reduced considerably (mean SNR: 3.82). Here we demonstrated that the performance of platforms from generation to generation cannot be assumed to be constant.

In detail, in the evaluation cohort of 219 patients, microarray analysis indicated a total of 4,097 CNVs comprising 3,156 losses and 941 gains, ranging 118 to 17,510,855 bp in length (mean 108,900, median 267,093). 463 out of 4097 candidate CNVs were evaluated with 2 or 3 quantitative PCR experiments each. 230 out of 463 CNVs (49.8%) were confirmed and 233 (51.2%) emerged to be false-positive findings (see table 4-1). When the DNA samples from the parents were available we investigated whether CNVs originated *de novo*. For those CNVs that were inherited, no parental inheritance bias was detected (p=0.5035); 95 CNVs were maternal, 86 CNVs paternal and 3 CNVs were detected in both parents. Concerning the type of CNV, our data provided highly significant evidence (p=0.0001) that deletions (164) are more frequent than duplications (66). These frequencies do not seem to be in concordance with the idea that human genome tolerates gains of genetic material better than losses. It should be noted, however, that the array threshold algorithm favors detection of deletions as compared to duplications since the logarithm of the intensity ratio for a deletion (having a theoretical test-over-reference ratio of ½ with log2(½) = -1.0) is more distant from the mean (log2(1) = 0) than the logarithm of the intensity ratio for a duplication (log2(3/2) = 0.59). Therefore, there is a higher probability that duplications may be missed as the logarithms of their intensity ratios may be closer to the noise level of the microarray.

Both the in-house Perl script and the DNAcopy software called more deletions than duplications and with the circular binary segmentation (CBS) the fraction of losses was even larger. We also observed that the specificity of the CBS algorithm was worse when calling deletions, as indicated by the higher rate of false positive deletions in all the SNP bins (see section 4.1.1.3). Taken together these data suggest that the increased false positive rate observed in the 610K dataset most likely resulted both from the microarrays which had a larger variance of the log2 intensity ratio and from the CBS evaluation algorithm which predicts more CNVs but with less specificity, especially in case of deletions.

5.2 Characterization of genomic imbalances in patients with Intellectual Disabiliy (ID)

Since most cases of intellectual disability (ID; mental retardation) occur sporadically and cannot be classified phenotypically into existing categories of monocausal disorders, ID has long been considered as a multifactorial disorder due to the interaction of many genetic and environmental effects. However, the major causes of ID published to date are cytogenetic or submicroscopic chromosomal imbalances and mutations of single genes, although some cases have been presented recently (Girirajan et al 2010).where a multifactorial explanation may be applied.

In part, the idiopathic ID cases may be explained either by a high rate both of undetectable *de novo* submicroscopic deletions/duplications or *de novo* point mutations that give rise to sporadic forms of ID, or by a large number of unexplored genes responsible for autosomal recessive ID which in small outbred family settings could be mistaken as sporadic. In other words, these monocausal/Mendelian defects might have been elusive so far for several reasons; 1) the small sizes of causative CNVs below the resolution limits of available microarray technologies, 2) the lack of large and consanguineous families in industrialized countries, 3) the high degree of genetic heterogeneity, and 4) the reduced penetrance of some ID-causative variants. Moreover, the wide clinical spectrum of ID could overlap with the autism spectrum disorder or other complex neuropsychiatric conditions. Therefore, it can be hypothesized that a certain proportion of underlying disease causes is due to specific combinations of rare variants with large effects, thus representing a novel entity of ID.

Next-generation sequencing (NGS) of exomes and eventually whole genomes will probably reveal a number of additional loci. Several authors (Vissers et al 2010, Hamdan et al 2011) already provided evidence that *de novo* single-nucleotide variants (SNV) may contribute substantially to intellectual disability. These finding may explain one of the major paradox in evolutionary theory, namely the fact that ID diseases are prevalent among population despite severely reduced fecundity.

In addition, NGS of exomes has the potential to identify mutations in genes that cause autosomal-recessive forms of intellectual disability (Abou Jamra et al 2011, Rafiq et al 2011).

The ultimate resolution to screen the human genome for disease-causing CNVs is at the single nucleotide level. Using NGS technologies CNVs can now be detected at this level by identifying differences in read depth (coverage), i.e., the number of reads mapping to a specific genomic locus. However, CNV detection differs between different technologies.

For example, Zhang et al (2009) compared CGHarray-based CNV analysis with NGS. They found that many small CNVs (<50kb) remained unidentified with arrays whereas large-scale CNVs (especially duplications) were missed with NGS (due to limited read length). Thus, it is still not possible to detect all CNVs using one single technology. It is therefore expected that in the coming years NGS will be introduced in the field of clinical cytogenetics with the aim of complementing the DNA microarrays, not of substituting them.

During the last years genomic microarrays have been extensively used to study the genetic causes of ID. We have successfuly applied commercially available SNP-based oligonucleotide microarays to identify previously unknown submicroscopic rearrangements in patients with intellectual disability (ID). The platform of choice was purchased from Illumina and two different SNP-based array types were used, one with 550000 SNPs distributed across the whole genome and the other consisting of 610000 SNPs. We were able to detect apparently disease-causing or *de novo* cryptic chromosomal aberrations in 9.7% (and up to 16.4%) of unselected patients with intellectual disability of unknown aetiology. This is in accordance with the literature (de Vries et al. 2005, Friedman et al 2006, Vermeesch et al 2007, Hoyer et al 2007 Hochstenbach et al 2009) The failure to detect rearrangements in 298 out of 330 patients suggests either that they do not have a chromosomal rearrangement or that they carry a very small microdeletion or microduplication that was not identified with the 550K or 610K Illumina arrays. An even denser microarray or another technique with higher resolution for detection such as NGS technique might have revealed a larger number of such rearrangements.

The CNVs detected in this study can be divided into the following groups; 1) clearly pathogenic CNVs that overlap with known microdeletion/duplication syndromes, 2) *de novo* CNVs 3) potentially pathogenic CNVs (without *de novo* confirmation) 4) rare inherited CNVs and 5) common CNVs listed in the Database of Genomic Variants (DGV). Whereas common CNVs were excluded from the further analysis, those CNVs that were not described in the DGV were assessed with Ensembl and DECIPHER for gene content and similar cases, respectively

In total, we detected known syndromes in 13 of 330 ID patients, described unbalanced translocation in 3 cases, and identified rare *de novo* CNVs in 16 patients. Potentially causative CNVs (without *de novo* confirmation) were detected in 21 patients (in total, 16.4%). Additionally, rare inherited CNVs were observed in 56 patients with unexplained ID. A more detailed description is given in the following:

Seven of the discovered variants were considered to be pathogenic as they are known to be associated with

well-established microdeletion/duplication syndromes such as holoprosencephaly, Prader-Willi syndrome, neurifibromatosis type I, DiGeorge syndrome, 2q37 monosomy, and Cohen syndrome (2 cases). These syndromes were recognized afterwards by a clinical geneticist, which underlines the difficulty of establishing a diagnosis by clinical observation. Six patients showed CNVs that were recently identified in other studies; 16p11.2-p12.2 deletion, 16p11.2 deletion, 14q11.2 deletion, 1q42-q44 deletion, *FOXP2* gene deletion, 5p13 duplication. For these new syndromes an obvious phenotype has not been delineated yet, and more patients with the same abnormalities are needed to unravel the associated phenotype.

16 CNVs had a *de novo* origin, i.e., were detected in patients with unspecific mental retardation but were not detected in their normal parental samples. The interpretation of this group of CNVs is difficult. If a CNV is not only *de novo* in the patient but also rare, relatively large, and overlaps with genes, the probability of this CNV to have clinical relevance is high. This characterization complied with all 16 CNVs. Therefore, we concluded that they all are causative of the ID in the respective patients.

For 21 potentially pathogenic CNVs, the inheritance could not be determined. Interpretation of these CNVs is even more difficult. However, these alterations were never observed in any of the other parents neither in the dataset of 4104 control individuals. Moreover, 7 of these CNVs without parent-of-origin analysis were located at loci whose haploinsufficiency is known to cause ID. As in case of *de novo* CNVs, the effect of potentially pathogenic CNVs with indetermined inheritance can only be ascertained by identification of phenotypically concordant patients with the same abnormality. DECIPHER is a database that can be used for that puropose. For the 21 potentially pathogenic CNVs no complete overlapping cases were found in DECIPHER. Therefore, more array data on ID patients and healthy controls will be needed to determine the clinical relevance of these CNVs.

Additionally, we also wanted to assess the presence of rare or privately inherited CNVs, as an increasing number of genomic loci were recently reported with variable inheritance and penetrance (McMullan et al 2009). We found 56 rare imbalances in our ID cohort that were inherited from a phenotypically normal parent. Some of these imbalances may indeed be benign variations while others may cause disease (Sharp et al. 2008, Girirajan et al. 2010, Mefford et al. 2009, Fernandez et al. 2009, Cook and Scherer 2008), especially among those that are located on the X chromosome or are related to a parent-of-origin effect. We detected a maternally inherited Xq13.1 deletion in a male patient which encompassed the *DGAT2L6* and *AWAT1* genes. The X chromosome of the patient was shown to be non-randomly inactivated in the unaffected mother. However, based on the expression and functional information available for these genes

we concluded that the *AWAX* and *DGAT2L6* deletion was not relevant for the phenotype of the child (see section 4.1.1.4.2).

We also identified six large rare inherited aberrations, i.e., a 16p13.11 deletion, a 15q13.3 deletion, a 15q13.3 duplication, a 1q21.1 deletion, a 22q11.2 duplication and a 12q14 deletion, that were previously associated with neuropsychiatric disorders (Hannes et al 2009, Miller et al 2009, Brunetti-Pierri et al 2008, Ou et al 2008, Menten et al 2007). These findings demonstrated that there are limitations in considering CNVs as either benign when common and inherited, or causal when rare and *de novo*. Some of the rare imbalances detected in individuals with ID, but shown to be inherited from phenotypically normal individuals, can be benign in some individuals but pathogenic in others. Penetrance may be low or incomplete.

A part from the 6 known risk alleles we also identified 50 rare inherited CNVs which contained coding sequences and were not known to DGV. We compared these rearrangements to an in-house database containing CNV data of 4104 control individuals and found that 30 of the 50 privately inherited CNVs also were not present in our control cohort. Therefore, it is possible that these 30 CNVs are susceptibility loci for intellectual disability. They might comprise a dosage sensitive gene(s) affecting brain development with incomplete penetrance or they might uncover a recessive allele of an ID gene that is inherited from the other parent. Further investigations are needed to reach a better estimate on these possibilities.

In order to distinguish disease-associated CNVs from benign CNVs, Hehir-Kwa et al (2010) developed a computational method that uses a Naïve Bayesian Tree (NBTree) classifier and 13 diverse genomic features. This classification method was validated in a set of 1203 CNVs detected in 584 ID patients collected from routine diagnostics, achieving a high accuracy (94%) with a sensitivity of 88% and a specificity of 94%; 1085 of the 1154 of the common inherited CNVs were correctly classified, and 43 of 49 CNVs previously associated with ID were correctly classified as ID-associated.. They applied the classifier on two CNV datasets with unknown clinical significance; 1) CNVs unknown in the general population for which inheritance could not be established and 2) privately inherited CNV unknown in the general population. This method showed that 66% of the rare inherited CNVs contained genomic features similar to previously recognized ID-associated CNVs (such as high densities of functional elements including genes whose mouse orthologues, when knocked-out, lead to specific nervous system abnormalities), a proportion that was significant when it was compared to random segments of the genome. Furthermore, 87% of the rare CNVs with unknown inheritance were classified as ID-associated. These results support

the view that the interpretation of CNVs as potentially disease-causing should not be limited to rare *de novo* CNVs with fully penetrant dominant effect (Hehir-Kwa et al 2010).

5.3 Identification of FOXP1 deletions in three patients with moderate ID and severe speech and language impairment

Our large-scale study on chromosomal abnormalities in 1523 individuals with mental retardation permitted us to identify deletions encompassing solely the *FOXP1* gene (partly or entirely) in three unrelated patients (two males, one female) who had moderate intellectual disability, global developmental delay, and severe speech and language disorders. In the two patients in whom parent DNA samples were available, it was possible to demonstrate that the deletions arose *de novo*, thus indicating a causal role of *FOXP1* in the clinical finding. Three other patients with *de novo* deletions encompassing *FOXP1* have been recently described. Pariani et al (2009) identified a *de novo* 785 kb deletion of the 3p14.1p13 region including the genes *FOXP1*, *EIF4E3*, *PROK2, GPR27* in a patient with blepharophimosis (drooping eyelids) and arthrogryposis (contractures of hands and feet), developmental delay and speech and language disorders. Carr et al (2010) reported on a patient with Chiari I malformation, delayed gross motor skills, severe speech delay, and epileptiform discharge who had a 1 Mb deletion extending over 500kb beyond the 3' end of *FOXP1*. Lastly, Hamdan et al (2010) identified a *de novo* 390kb intragenic deletion in a female patient with non-syndromic intellectual disability (NSID), autistic features and severe language impairment. The *FOXP1* disruptions are likely to account for the similarities in these six patient phenotypes, namely for the deficits in motor development and the speech delay. It has been speculated that the blepharophimosis and arthrogryposis present in the case presented by Pariani et al and the Chiari I malformation and epileptiform discharge in the patient presented by Carr et al may be due to the deletion of adjacent genes (*EIF4E3*, *PROK2, GPR27)* and/or regulatory elements important for the expression of other genes located in the vicinity (Hamdan et al 2010). In two of our three patients we also found additional clinical features including obesity and mild craniofacial anomalies (prominent forehead and frontal hair upsweep). Since the deletion in our patients only comprised the *FOXP1* gene these features might be considered as specific for *FOXP1* deletion.

Noteworthingly, a similar 3p14.1p13 deletion affecting *FOXP1* was also observed in a control individual

who was not reported to have learning difficulties, fact that makes difficult the straightforward correlation between genetic variants and specific phenotypic features.

The human forkhead-box (FOX) gene family, to which *FOXP1* belongs, consists of over 40 members classified into 19 subfamilies (designated FOXA to FOXS), all of which encode 80-100 amino acids that form a DNA-binding winged helix or a forkhead domain (Newbury et al 2010). The FOXP subfamily comprises four genes (FOXP1-4) and is characterized by a DNA-binding dependent N-terminal transcriptional repression domain which encompasses both a zing finger and a leucine zipper motif (Pariani et al 2009). The proteins encoded by these genes act as transcriptional repressors by forming homo- or heterodimers (Li et al 2004) which play important roles in development, metabolism, immunology and cell cycle control (Carlsson and Mahlapuu 2002). In partiular, it has been suggested that *FOXP1* and *FOXP2* may regulate common processes, including the developing of foregut and brain (Shu et al 2007, Spiteri et al 2007). *FOXP1* is located on chromosome 3p14.1 where it covers 628kb. It consists of four isoforms, the most common being 677 amino acids in length. Heterozygous mutations in its closest relative, *FOXP2*, cause a rare monogenic form of developmental verbal dyspraxia (DVD) (i.e. problems in sequencing mouth movements underling speech) and language impairment (Lai et al 2001, MacDermot et al 2005, Feuk et al 2006, Vernes et al 2006). Therefore, it was hypothesized that mutations in *FOXP1* could likewise be associated with language impairment (Teramitsu et al 2004). In a subsequent screen of the *FOXP1* coding sequences in 49 patients with a clinical diagnosis of DVD and normal intelligence, Vernes et al (2009) could not identify any pathogenic variant. A nonsynonymous coding change (p.P215A) was a found in a patient but also in an unaffected control individual. This is in contrast with the clinical findings observed in patients with *de novo FOXP1* deletion (see above). Since these patients exhibit significant mental retardation and/or autistic features in addition to their speech and language disabilities, it may be possible that additional functions of *FOXP1* exist, presumably acting on a more global level in neuronal development as compared to *FOXP2.*

Characterization of the *FOXP1* deletions in our mental retardation cohort revealed that the affected allele had a paternal origin. This finding is in line with most of the mutations found in *FOXP2*, which are maternal uniparental disomies or deletions of the paternal allele (Marshall et al 2008). This suggests that both *FOXP1* and *FOXP2* may show differential parent-of-origin expression in human speech development. We analysed the breakpoints at a nucleotide level, observing that there were neither palindromic sequences nor similarities between the breakage regions. These interstitial deletions are therefore assumed to emerge

by a double-strand break (DSB) and non-homologous end joining (NHEJ).

To proof the causality of the *FOXP1* deletion following-up investigations were based in a screening of the entire coding region of *FOXP1* (exons and flanking intronic sequence) for nucleotide changes in a panel of 883 probands with mental retardation. Eight non-synonymous coding changes, three synonymous and nine non-coding variants were identified. Five of the non-synonymous variants were transmitted from a healthy parent and three of these five variants were also found in matched control individulals. As for the three remaining non-synonymous variants (p.M101V, p.S261P, p.N597T) no conclusive evidence for their pathogenicity was reached because parental DNA was not available.

A recent study on a male patient with NSID and autism (Hamdan et al 2010) identified a *de novo* nonsense mutation, c.1573C>T (p.R525X) of *FOXP1*. This alteration abolished the last 152 amino acids, including part of the forkhead DNA-binding region and a conserved nuclear localization signal (NLS). Functional studies demonstrated that the p.R525X mutation impaired the ability of FOXP1 to repress the SV40 promoter.

Other missense mutations have been reported in the literature which were likely to be non-pathogenic. Both in a patient and in a control individual Hamdan et al (2010) detected the same missense variant c.643C>A (p.P215A) that was reported previously by Vernes et al (2009). This variant was found to be transmitted from a healthy parent. In addition, the authors reported three additional heterozygous missense variants (p.V445M, p.T613N, p.N570S) that they found only in control individuals.

A recent trio-based exome sequencing study (O'Roak et al 2011) reported a *de novo FOXP1* mutation (p.Ala339SerfsX4) in a patient affected by severe autism. In addition to the *FOXP1* mutation, the proband also carried an inherited missense variant (p.His275Ala) of *CNTNAP2* which appears to be extremely rare or private. *CNTNAP2* is directly downregulated by FOXP2 and/or FOXP1 and has been associated independently with autism and specific language impairment (Bakkaloglu et al 2008, Vernes et al 2008, Arking et al 2008, Alarcón et al 2008, Jackman et al 2009, Poot et al 2010, O'Roak et al 2011). *CNTNAP2* encodes a neuroxin protein that is responsible for the localization of potassium channels in developing neurons and plays an important role in the facilitation of axonal-glial interactions (Poliak et al 2003, Zweier et al 2009). O'Roak et al (2011) hypothesized that *FOXP1* haploinsufficiency may result in overexpression of the CNTNAP2 protein, amplifying any deleterious effect of the mutation in the patient. It is generally thought that genetic mechanisms underlying speech and language disorders are multifactorial in nature, involving complex interactions between several genetic variants and environmental factors

(Newbury et al 2010). The function of genes may be modulated by a large number of stochastic variables that differ between individuals, yielding incomplete penetrance of a specific pathogenic variant. The finding of a missense mutation in *CNTNAP2* and a frame-shift mutation in *FOXP1* in the same patient provide a functional support for a multi-hit model for disease risk. Therefore, it is possible that not only the *de novo FOXP1* deletions detected in our study but also the non-synonymous single nucleotide variants that were transmitted from a normal parent contribute to disease generation in a multifactorial pathogenesis.

5.4 Cohen syndrome diagnosis in three patients with unknown mental retardation

Copy number analysis of high density oligonucleotide array data in a cohort of 1523 patients with unexplained mental retardation revealed intragenic heterozygous *COH1* deletions in three unrelated patients. These deletions were not detected in a cohort of 1612 control individuals. Subsequent sequencing of the *COH1* gene revealed point mutations in the second allele of all three patients thus providing a distinctive diagnosis of Cohen Syndrome (MIM # 216550). Our study demonstrated that Cohen syndrome, a rare autosomal recessive disorder, may be present in patients with unexplained mental retardation being only detectable by genome-wide screening methods such as high resolution oligonucleotide arrays but not by clinical examination (Rivera-Brugués et al 2011).

The phenotype of Cohen syndrome has been described to be fairly homogeneous in Finnish patients where the founder mutation c.3348_3349delCT is detected in about 75% of mutant alleles (Kolehmainen et al 2003) while in non-Finnish and especially in young Cohen patients a high genotypic and phenotypic variability occurs. Several clinical diagnostic criteria for Cohen syndrome have been introduced (Horn et al 2000, Chandler et al 2003, Mochida et al 2004, Kolehmainen et al 2004): Chandler et al (2003) proposed that besides significant learning disabilities two of the following criteria should be present for a Cohen syndrome diagnosis to be made: typical facial gestalt, pigmentary retinopathy, and neutropenia. Kohlemainen et al (2004) suggested Cohen syndrome in patients fulfilling at least six of the following criteria: developmental delay, microcephaly, typical facial gestalt, truncal obesity with slender extremities, overly sociable behavior, joint hypermobility, high myopia and/or retinal dystrophy, and neutropenia. El Chehadeh et al (2010) concluded that mutation analysis is not indicated in the absence of chorioretinal dystrophy or neutropenia.

Cohen syndrome was not suspected at first instance in the three patients here reported although all of them were examined by experienced pediatricians or clinical geneticists. This can be explained by the young age of the patients (16 months, 18 months, 2 ¾ years). The facial appearance of infants and young children with Cohen syndrome differs from that in adult patients, and myopia/ retinal pigmentary changes usually develop in the pre-school age (Chandler et al 2003). The common facial characteristics in our small series of young patients with Cohen syndrome included a hypotonic facial expression, almond shaped palpebral fissures, a prominent nose, and a short philtrum. All patients were affected by mental retardation and delay in motor and speech development. Pigmentary retinopathy and neutropenia were absent. The unusual phenotype in the patient 3 (see 4.1.2.2.1) may be due to the fact that in addition to exons 1 to 17 of the *COH1* gene the deletion affected the neighbouring exon 4 of the *ORS2* gene. Although the gene function of *ORS2* in humans is unknown, an influence on the phenotype cannot be ruled out.

The *COH1* gene that contains 62 exons distributed along 864315 bp of chromosome 8, is so far the only gene known to be associated with Cohen syndrome, and it is thought to play a key role in neuronal differentiation. More than 96 different mutations in *COH1* gene have been detected in association with Cohen syndrome. The majority of them are terminating mutations predicted to result in a functional null allele, whereas missense mutations and in frame deletions are less frequent. Recently, large deletions and duplications in *COH1* gene comprising multiple exons have been identified as a cause of Cohen syndrome (Balikova et al 2009, Parri et al 2010), although the frequency of copy number alterations in the *COH1* gene is still unknown. Parri et al. (2010) reported that *COH1* copy number variations account for 42% of *COH1* mutations. Balikova et al (2009) could increase their detection rate from 70% to 88% in typical Cohen patients when the introduced CNV analysis. In general, the detection rate depends on patient selection criteria, array resolution and workflow used in data analysis. It is likely that arrays analysis will further increase the detection rate of deletions and duplications.

The fact that we failed to detect CNVs in the *COH1* gene in 1612 controls from a general German population cohort indicates that CNVs in the *COH1* gene are rare. This is in contrast to the annotation of the *COH1* gene CNVs as benign polymorphisms in the UCSC genome browser and the database of genomic variants (DGV). Two studies reported the presence of CNVs in *COH1* at a frequency of up to 6% that affect the coding part of the gene. The reported frequency of the copy alterations would imply a disease frequency higher than 0.36% which is incompatible with the infrequent occurrence of Cohen

syndrome, suggesting that some of the reported CNVs may be smaller intronic imbalances or represent false positives (Balikova et al 2009).

The phenotype of Cohen syndrome defined by *COH1* mutations is fairly unspecific in particular in very young patients but in older children too. In addition, deletions in neighbouring genes may affect the phenotype of Cohen syndrome in the context of a contiguous gene syndrome.
Various diagnose criteria to identify Cohen syndrome patients have been established, including the ones described by Chnadler, Kolehmainen or El Cehadeh (see above). However the criteria mentioned above are not sensitive enough since applying stricly these criteria our patients would have been missed.
Nevertheless we can conclude that young patients with a hypotonic facial expression, almond shaped eyes, short philtrum, mental retardation and motor and speech delay are suspicious of Cohen syndrome. Microarrays have the potential to diagnose Cohen syndrome in very young patients and in patients with an atypical phenotype.

5.5 Identification of cryptic imbalances in patients with Congenital Hyperinsulism (CHI)

Diffuse congenital hyperinsulinsm is predominantly caused by autosomal recessive inheritance, with mutations in *ABCC8* and *KCNJ11* genes. A screen for point mutations in the entire coding sequence of these genes was performed using Sanger sequencing. Patients with negative results were collected to build a cohort of unexplained diffuse CHI cases which was further examined with SNP microarrays testing. The constructed cohort included 8 cases with mutation in one allele and 32 cases with no mutations in the mentioned genes. The typical low mutation detection rate on CHI may be explained by further locus heterogeneity or by missed mutations of known CHI genes.
Due to the fact that *ABCC8* account for the largest fraction of genetically diagnosed CHI cases, we looked for missed mutations in *ABCC8* gene by employing other molecular techniques such as quantitative PCR and deep sequencing on the Illumina GAII system. No deletions or duplications were detected in the exonic regions of the *ABCC8* gene by using quantitative PCR. No indels or single nucleotide changes were found in the entire genomic region of *ABCC8* by deep sequencing on an Illumina GAII system. An intriguing question is the frequent detection of a single, heterozygous mutation (8 out of 40 cases), by opposition to many recessively inherited CHI forms. This may suggest that the combination of different mechanisms

("synergistic heterozygosity") may be one cause of CHI. However, dominant-acting mutations in *ABCC8* or *KCNJ11* have also been reported (Pinney et al 2008). Therefore, functional investigations will be necessary to prove whether or not the heterozygote *ABCC8* and *KCNJ11* point mutations detected in these 8 patients have a pathogenic effect.

Other genes such as *GLUD1, GCK, HADH, HNF4A, SLC16A1* and *UCP2* have been found to be mutated in non-channelopathy CHI (Glaser et al. 1998, Clayton et al 2001, Pearson et al. 2007, Otonkoski et al. 2007, González-Barroso et al 2008). Mutations in these genes are detected only in 5 to 10% of the CHI cases (Flanagan et al 2011). To test for further locus heterogeneity we did not check for point mutations in these known candidate genes, but we screened the whole genome for deletions and duplications These genes then may potentially indicate metabolism-coupled regulation pathways of insulin expression. The discovery of novel genes and pathomechanisms currently is a major goal in CHI research and a focus of the present study. Indeed, CHI diversity itself calls for such studies since more than 50 % of the cases remain unexplained genetically (Thomas et al 1995, Flanagan et al 2011).

Chromosomal abnormalities detected by first SNP oligonucleotide arrays accounted for 20% of the studied cases. Four heterozygous chromosomal abnormalities were identified in five CHI patients. The detected imbalances were validated by a second independent test, such as quantitative PCR.

Unraveling the clinical relevance of potentially pathogenic CNV is a new challenge. Regions containing genes can be present in variable copy number without obvious clinical manifestations, which makes it very hard to determine whether a copy number has clinical significance. In this study we first excluded all polymorphic CNV by comparison with the database of genomic variants (DGV) and our in-house reference set. Second, for all CNVs containing coding genes annotated by Ensembl, the inheritance was determined by checking both parents (when available). For two of the four potentially pathogenic CNV we could establish that the rearrangement was arising *de novo, i.e.,* not transmitted from any of the unaffected parents. For the other two cases no parental DNA was available, thus precluding parent-of-origin analysis.

One of the imbalances without parent-of-origin analysis corresponded to a complex rearrangement on chromosome X; a 2.7 Mb deletion on p22.33 and a 16.2 Mb duplication on q11.1-q22.1. More than 420 genes were comprised in these CNVs making it likely that one of them was causal for the CHI although none of the genes was found to be a primary candidate gene. The second CNV in which parent-of-origin analysis could not be performed was a 579 kb duplication on 8q24.23. This CNV comprised the 3' end of the *FAM135B* gene which encodes a hypothetical protein, *LOC51059*, of unknown biological function.

Expression microarray data indicate that *FAM135B* is highly expressed in cerebellum and testes. However, the pathogenicity of this rearrangement could not be supported by large gene content as in the previous case. Moreover, it is worth to note that a direct tandem duplication would only alter the dosage of *FAM135B* if the gene were disrupted at the breakpoint. Therefore, the causality of this imbalance will necessitate additional molecular techniques such as next generation sequencing to demonstrate the disruption of *FAM135B* or another gene of the genome.

A particularly intriguing *de novo* rearrangement was observed in two affected brothers. These patients carried large *de novo* CNVs in chromosome 9 and 13. Whereas the 7.85 Mb duplication of 13q33.3-q34 was found to be identical in the two patients, the rearrangements of chromosome 9 appeared to be slightly different from each other. A 4 Mb deletion of 9p23-p24.1 and 6.3 Mb deletion of 9p24.1-24.3 were detected in one patient, while a 7.8 Mb deletion at 9p23-p24.1 and a 4.3 Mb deletion at 9p24.2-p24.3 were detected in the other patient. These brothers are thought to carry a *de novo* unbalanced translocation between chromosome 9 and 13 which is assumed to be the cause of the CHI of the patients. No obvious candidate genes were found within these specific chromosomal regions.

The last potentially pathogenic *de novo* aberration corresponded to a 970 kb deletion on chromosome 1p31.1. This chromosomal copy number variation encompassed the *PTGER3* and *ZRANB2* genes and the last exon of the *NEGR1* gene. Based on functional data the *PTGER3* was selected as a primary candidate gene and a mutational screening in our series of unexplained CHI patients was performed (see below). *PTGER3* (prostaglandin E receptor 3) encodes one of the receptors for prostaglandin E2 (PGE2). It has been long known that PGE2 impairs pancreatic beta cell function, and that the *PTGER3* agonist misoprostol mimicked the inhibitory action of PGE2 on glucose-stimulated insulin secretion (GSIS) (Tran et al 2002). Meng et al (2006) investigated the PGE2 signaling pathway in beta cell regulation. Their studies demonstrated that PGE2 specifically stimulated *PTGER3* gene expression and diminished cAMP generation. This was accompanied by the downregulation of Akt and Foxo phosphorylation resulting in the inhibition of GSIS. We hypothesized that the haploinsufficiency of *PTGER3* induces a 50% reduction of the stimulation by PGE2, thus diminishing the inhibition of GSIS and resulting in elevated insulin secretion. This hypothesis is in agreement with the phenotypic findings observed in the homozygous knock out model. Sanchez-Alaves et al (2007) reported that *PTGER3* knock-out mice (Ep3 -/-) exhibited increased frequency of feeding and developed an obese phenotype under a normal diet. Obesity in these

mice was characterized by elevated leptin and insulin levels and a more than 20% higher body weight compared to wildtype littermates.

EP3, the gene product of *PTGER3* (MIM#176806), is a member of the membrane-bound G protein-coupled prostanoid EP receptors family. This family includes the gene products of *PTGER1* (MIM# 176802), *PTGER2* (MIM# 176804), *PTGER3* (MIM#176806), *PTGER4* (MIM# 601586), and the variants of *PTGER3*. They all are defined on the basis of their pharmacologic profiles and signal transduction pathways (Coleman et al. 1994). PGE2 is the natural agonist of these seven-transmembrane-spanning receptors. *PTGER3* may have many biological functions, including those in digestion, nervous system activity, renal reabsorption, and uterine contraction. Northern blot analysis demonstrated its expression in pancreas as well as in small intestine, heart, gastric mucosa, mammary artery and pulmonary vessels. *PTGER3* gene spans more than 80 kb and is composed of 10 exons. Fourteen transcript variants generated by alternative mRNA splicing are described, encoding eleven distinct protein isoforms. Exon 1 and a portion of exon 2 encode the seven transmembrane domains and 10 amino acid residues of the cytoplasmic tail, which are common to all EP3 isoforms. The combinations of the other portion of exon 2 and 3-10 exons encode the EP3 isoform-specific C-terminal tail and confer specificity in the coupling to signal transduction pathways. *PTGER3* transcripts are expressed in a wide variety of human tissues, with each transcript showing a different tissue-specific distribution (Sugimoto et al. 2007, Kotani et al. 1997).

The mutation screening of *PTGER3* in our series of unresolved patients yielded a heterozygous 1bp deletion in exon 4 of *PTGER3* isoform 8 (p.N395QfsX12). This null allele predicts truncation at the isoform-specific C-terminal domain and may give rise to nonsense-mediated RNA decay. The identified sequence variation was not annotated as a polymorphism in HapMap or the NCBI dbSNP database. In order to discriminate between causative mutation and harmless polymorphism, this nucleotide change was assessed for segregation in the respective families and occurrence in the general population. Segregation analysis showed that the variant was inherited from the apparently unaffected mother and genotyping of the KORA population cohort revealed that 18 control individuals were also carriers of this nucleotide variant. Additionally, we evaluated the specific expression of mRNA of the EP3 isoforms by RT-PCR analysis observing that *PTGER3* isoform 8 (NM_198718.1) is not expressed in the human pancreas tissue. Taken together, the genotyping and expression results demonstrated that the novel sequence variant identified in the *PTGER3* gene is not causative of congenital hyperinsulinism. Instead, this variant corresponds to a rare

polymorphism that might affect solely the carboxyl termini of the isoform EP3-8 which is not expressed in pancreas.

Nonetheless, the haploinsufficiency of *PTGER3* may be regarded as a major contributor to the CHI phenotype of the patient in which the 970 kb deletion on 1p31.1 was detected. However, we can conclude that due to the negative results obtained in the mutational screening of our cohort *PTGER3* is not a frequent cause of CHI. However *PTGER3* may be still considered a promising candidate gene for CHI. Further screening in larger cohorts of individuals with similar phenotype to the index patient will permit to elucidate the real implication of this gene in the aetiology of CHI.

To our knowledge this is the first time that a SNP array testing is performed in a cohort of patients with diffuse congenital hyperinsulinism (CHI). Future investigations based on a mutational screening of the candidate genes *GLUD1, GCK, HADH, HNF4A, SLC16A1* and *UCP2* will be suitable to those patients of our cohort without point mutations in *ABCC8* or *KCNJ11* and without any conspicuous rearrangement detected by the reported CNV analysis.

5.6 Identification of a contiguous duplication of TBX5 and TBX3 in patients affected with Holt-Oram syndrome and supernumerary mammary glands

Mutations in *TBX5* gene cause Holt-Oram syndrome (HOS) (OMIM 142900), an autosomal dominant disorder characterized by upper limb and heart malformations (Li et al. 1997, Basson et al. 1997). Mutations in *TBX3* are known to cause Ulnar-Mammary syndrome (UMS) (OMIM 181450), also an autosomal dominant inherited disorder that affects posterior limb, apocrine gland, tooth, and genital development (Bamshad et al. 1996, Schinzel et al. 1987). Genetic studies by direct sequencing reported low mutation detection rates for these genes (Brassington et al. 2003; Cross et al. 2000; Heinritz et al. 2005). Most identified mutations are nonsense or frameshift mutations, leading to a functional null allele. Missense mutations are less frequent, but underly a similar pathological mechanism, i.e., loss of function, by diminishing transcription factor interaction or DNA binding (Basson et al. 1999; McDermott et al. 2005). To increase the diagnostic yield other detection methods such as multiplex amplificable probe hybridization (MAPH) and array CGH have been employed, permitting the identification of large genomic deletions causing haploinsufficiency of *TBX5* and *TBX3* gene, respectively (Akrami et al. 2001, Klopocki

et al. 2006). A contiguous hemizygous deletion of *TBX5, TBX3*, and *RBM19* resulting in a combined phenotype of Holt-Oram and Ulnar-mammary syndromes has been also identified by using qPCR (Borozdin et al. 2006). One missense mutation underlying a different pathological mechanism, i.e., a gain-of-function, has been found in individuals with atypical phenotypes of HOS. This finding suggested that other individuals with atypical or non-HOS phenotype might possibly have *TBX5* gain-of-function mutations. However, no duplications of the *TBX5* and *TBX3* genes have been identified to date.

We observed a contiguous heterozygous duplication containing the *TBX5* and *TBX3* genes that resulted in a combined phenotype of Holt-Oram syndrome and supernumerary mammary glands. The chromosomal rearrangement was first identified with a high density SNP oligonucleotide microarray. After delineating the breakpoints with qPCR the maximal duplication size could be estimated as approximately 345,6 kb. The index patient (31565) showed immobile joints on both thumbs and an aortic stenosis which was corrected six weeks after birth. Other clinical features identified in the affected members of this family included stiffened thumb joints, deficit in distal phalanx, bicuspidal aortic valve, atriovetricular septum defect, and myocardial infarction. Interestingly, besides the typical signs of HOS further accessory mammary glands were described in two affected members of this family.

Several cases with partial or terminal duplications of 12q, resulting in increased *TBX5* and *TBX3* dosage, have been reported (Harrod et al 1980, Melnyk et al 1981, Ieshima et al 1984, McCorquodale et al 1986, Dixon et al 1993, Cappellacci et al 2006). In a review of twelve cases with 12q duplications, Dixon et al. (1993) emphasized the similarity of the physical findings among the reported individuals, suggesting a clinically distinct syndrome. Clinical features that are common to the reviewed cases include the following: a flat nasal bridge, epicanthic folds, hypertelorism, down turned corners of the mouth, low set ears, excess nuchal skin, simian creases, widely spaced nipples, hypertonia, congenital heart disease, and skeletal deformities (ulnar deviation of the hands). By comparison with the HOS and UMS phenotypes it can be hypothesized that congenital heart disease and skeletal deformities present in 12q duplication cases result from the overexpression of *TBX5* (and/or possibly of *TBX3*). Additionally, it can be speculated that the breast alterations (i.e, widely spaced nipples) observed in most of the patients is the result of *TBX3* overexpression, since *TBX3* is one of the genetic determinants of the breast development. Of note, displaced nipples and areolae and limb malformations on the ulnar side are clinical features consistent with UMS (Borozdin et al 2006).

In the cryptic duplication of *TBX5* and *TBX3* the increase in gene dosis of *TBX5* may have caused mild skeletal deformations in most of the affected patients and congenital heart disease in some of them, a finding that is in a good agreement with Postma et al (2009). However, an effect of the increased *TBX3* gene dosis cannot be ruled out. The *TBX3* gene is expressed together with other T-box genes during cardiac development (Plageman and Yutzey 2005). Nevertheless, *Tbx3del/-* mice do not show heart defects (Davenport et al. 2003), and to date there is only a single report of a cardiac malformation in a patient with UMS (Meneghini et al. 2006). Therefore, the role of *TBX3* in inborn cardiac disease is still not clear.

A first genotype-phenotype relationship for a *TBX3* gain-of-function mutation is described in the present study. On the contrary to *TBX3* haploinsufficiency, which produces apoplasia or hypoplasia of the breast (Bamshad et al 1997), the increased dosage of *TBX3* causes the development of supernumerary mammary glands. Therefore, the overexpression of *TBX3* might lead to hyperplasia of the breast (which is in contrast with the hypoplasia of tissues caused by *TBX3* haploinsufficiency), most likely through an interference with apoptosis or differentiation during critical stages of embryogenesis (Bamshad et al 1997, Carlson et al 2002).

The penetrance of HOS and UMS syndromes is known to be complete, while the expressivity is highly variable, not only between families but also among sibs. This is well documented in our five-generation family where the same mutation leads to variable clinical symptoms; most of the patients showing mild hand abnormalities but a few of them presenting with breast and heart defects.

TBX5 gain-of-function mutations may result in a similar phenotypic spectrum as haploinsufficiency caused by loss-of-function mutations or deletions. Although this overlap is unusual it is not the first example: In the diGeorge/velocardiofacial syndrome (DGS/VCF) Zweier and colleagues observed that gain-of-function point mutations may lead to the same phenotype as 22q11.2 deletions (Zweier et al 2007).

Moreover, DGS/VCF syndrome is also associated with a highly variable phenotype despite the uniformity of the chromosomal deletion that causes the disease in most patients (80-90% share a heterozygous 3 Mb deletion). Interestingly, in the heterozygous mutant mouse model (Df1-/+), which model del22q11, the penetrance of cardiovascular defects is reduced similar to those seen in deleted patients while this is not the case for other del22q11-like findings (Taddei et al 2001). The authors reported that the varying genetic background of these mice had a major effect on penetrance of the cardiovascular defects by affecting embryonic recovery of the fourth pharyngeal arch artery growth abnormality. This effect could not be explained by allelic variation at the haploid locus, and instead it was thought to be caused by genetic

modifiers elsewhere in the genome (Taddei et al 2001).

Over the last decade several mouse models have been developed in order to dissect the genetic implication of *TBX5* or *TBX3* during development and disease. One of the first studies was conducted by Bruneau et al. 2001 which generated several mutant mouse models, such as tbx5 hetero- and homozygous mice on different genetic backgrounds. It was observed that Tbx5del/+ mice exhibited both cardiac and limb defects consistent with HOS features. Tbx5 haploinsufficiency also markedly decreased atrial natriuretic factor (Nppa) and connexin-40 (Cx40) transcription, implicating these as Tbx5 target genes and providing a mechanism by which 50% reduction of T-box transcription factors causes disease. Moreover, this study provided a potential explanation for Holt-Oram syndrome conduction system defects and suggested mechanisms for intrafamilial phenotypic variability (e.g., polymorphisms in Nkx2-5) (Bruneau et al. 2001). Tbx5 deficiency in homozygous mice (Tbx5 -/-) was seen to cause severe hypoplasia of posterior domains in the developing heart (Bruneau et al. 2001). Additionally, other studies demonstrated that homozygous inactivation of Tbx5 during forelimb development leads to failure of forelimb bud initiation, resulting in the complete absence of forelimb structures. This remarks the important role of Tbx5 in early limb development (Agarwal et al. 2003, Rallis et al. 2003).

Also, mutations in mouse Tbx3 have been generated to study the mechanisms by which *TBX3* haploinsufficiency causes breast and limb abnormalities in Ulnar-mammary syndrome. Whereas in heterozygous mice no limb or mammary gland defects were observed, homozygous Tbx3-/- mice showed the predominant features of human UMS with severe defects in forelimb and mammary gland development (Davenport et al 2003). Several possibilities could explain these species differencies; 1) differences in genetic background modifiers, 2) species-specific differences in dose sensitivity and 3) differences in the developmental role of Tbx3 between mouse and human. Additionally, the authors observed variable expressivity in mice with uniform genetic background as well as similar variability and range of expressivity in several mixed genetic backgrounds, indicating that phenotypic variability might be caused by stochastic events during development (Davenport et al 2003).

To further understand these conditions it would be helpful to develop other mouse mutants such as a mouse model heterozygous for TBX gene duplications. Specially, the Tbx3 dup/+ heterozygous mouse would be of great utility, since it would provide a means of investigation the involvement of Tbx3 overexpression in the development of additional mammary glands. This model could help to elucidate whether the

supernumerary mammary glands exhibited in humans is a unique observation or whether it is a consequence of a not yet demonstrated interaction between overexpressed *TBX5* and *TBX3* genes.

6- CONCLUSIONS

1) In this study we used and validated the Human550-Quad and Human610-Quadv1_B Illumina platforms. For validation, the arrays were applied to 110 and 109 patients with undiagnosed intellectual disability, respectively. By validating the array data with quantitative PCR we observed that CNVs supported by >20 SNPs were called with a high degree of specificity in both array types; 96.9% for 550K arrays and 87.8% for 610K arrays.

2) We were able to detect apparently disease-causing or *de novo* cryptic chromosomal aberrations in 9.7% of unselected patients with intellectual disability of unknown aetiology. In total, we detected known syndromes in 13 of 330 ID patients, described unbalanced translocation in 3 cases (2 of them being in a mosaic state), and identified rare *de novo* CNVs in 16 patients. Potentially causative CNVs (without *de novo* confirmation) were detected in 21 patients (in toal 16.4%). Additionally, rare or privately inherited CNVs were observed in 56 patients with unexplained ID.

3) We showed that SNP-based arrays allow the detection of intragenic deletions and duplications. The identification of a pathogenic CNV affecting only a single gene allowed us to consider that particular gene as a candidate for intellectual disability. This was the case for three unrelated patients with moderate intellectual disability, global developmental delay, and severe speech and language disorders in whom a *de novo* deletion encompassing solely the *FOXP1* gene was detected.

4) To prove further the causality of the *FOXP1* deletion, we screened the entire coding region of *FOXP1* (exons and flanking intronic sequences) for nucleotide changes in a panel of 883 probands with intellectual disability. Eight non-synonymous coding changes, three synonymous and nine non-coding variants were identified. No conclusive evidence for the pathogenicity of three non-synonymous variants (p.M101V, p.S261P, p.N597T) was reached because parental DNA was not available.

5) In addition to the *de novo* cases of ID, we also identified patients with an autosomal recessive form of ID in our cohorts. We detected three partial heterozygous deletions of the *COH1* gene at locus 8q22 which is mutated in Cohen syndrome.

6) By sequencing the entire coding region and the exon/intron boundaries of *COH1* we identified a stop mutation, a frameshift and two missense mutations in the remaining allele, respectively. Thus, three compound heterozygous mutations were detected in the *COH1* gene: c.5086C>T / c. 3872-5024del [p.R1696X / p.G1291fsX42], c. 1207-2824del / c. 11505delA [p.L403fsX11/ p.K3835fsX43] and c.1-2515del / (c.3866C>G/ c.11827_11828insATG) [- /(p.T1289S / p.D3942_G3943insD)], thus providing a distinctive Cohen Syndrome diagnosis to three unrelated patients of our ID cohort in whom the cause of ID was unknown before.

7) We studied the genetic basis of a rare human autosomal disorder, i.e., diffuse Congenital Hyperinsulinsm (CHI) in a cohort of 40 patients with inconspicuous mutation screening of the *ABCC8* and *KCNJ11* genes. Chromosomal abnormalities detected by SNP oligonucleotide arrays accounted for 20% of the studied cases. Four heterozygous chromosomal aberrations were identified in five CHI patients to be most likely causative of CHI.

8) The most interesting rearrangement found in the CHI cohort was a 970kb deletion at the chromosomal band 1p31.1 which was found to encompass the *PTGER3* and *ZRANB2* genes and the last exon of the *NEGR1* gene. We hypothesized that the haploinsufficiency of *PTGER3* gene induces a 50% reduction of the stimulation by PGE2, thus diminishing the inhibition of glucose-stimulated insulin secretion (GSIS) and resulting in elevated insulin secretion.

9) The screening for point mutations in the candidate gene *PTGER3* did not reveal any pathogenic variant neither in the second allele of the patient in which a *de novo* deletion had been detected nor in a cohort of 39 unrelated patients with unexplained CHI. Instead we identified a novel polymorphic variant which was also detected in 18 individuals of our control cohort.

10) CNV analysis in a family with both atypical Holt-Oram syndrome and additional mammary glands led to the detection of a heterozygous duplication at the chromosomal band 12q24.21. The maximal duplication size was estimated to be not longer than approximately 345.6 kb and it spanned the whole coding region of the *TBX5* and *TBX3* genes.

11) Locus-specific gene dosage assessment in the HOS pedigree demonstrated the cosegregation of the duplication and the Holt-Oram syndrome/supernumerary mammary glands phenotype, this being a strong indicator of the duplication's pathogenecity. Up to date, this is the first report of a heterozygous duplication encompassing both *TBX5* and *TBX3* genes, and consequently the first report of a combined phenotype of Holt-Oram syndrome and supernumerary mammary glands.

7- SUMMARY IN CATALAN

Introducció

Discapacitat Intel.lectual

La discapacitat intel.lectual (ID) és el problema més habitual en genètica clínica i representa una important càrrega socioeconòmica en l'àmbit de l'atenció sanitària, ja sigui en els països industrialitzats com en desenvolupament. ID es caracteritza per un funcionament cognitiu subnormal i un dèficit en les conductes d'adaptació que comença abans de l'edat de 18 anys. ID té una prevalença estimada de 2-3% (Roeleveld et al. 1997). L'espectre de malalties ID resulten d'una etiologia amplia, incloent causes genètiques com ara aberracions cromosòmiques (per exemple, trisomies, translocacions no balancejades), síndromes de microdeleció / microduplicació i malalties monogèniques. No obstant això, el defecte que causa la malaltia segueix sent desconegut en un 50-80% dels pacients (Johnson et al. 2006, Rauch et al. 2006). L'anàlisi citogenètica convencional detecta alteracions cromosòmiques en el 10% dels pacients ID, i almenys la meitat d'ells es deuen a casos de trisomia 21. La resolució de les tècniques de citogenètica microscòpica es limita a un nivell de 5-10 Mb. L'anàlisi de regions subtelomèriques utilitzant FISH (fluorescence in situ hybridization) o MLPA (muliplex ligation-dependent probe amplification) permet detectar reordenaments submicroscòpics en una fracció addicional del 5% dels nens afectats (Ravnan et al. 2006). Per augmentar encara més la taxa de detecció d'aneusomies críptiques, nous mètodes basats en tècniques de microarrays han estat desenvolupats en els darrers anys (Solina-Tendal et al. 1997, Pinkel et al. 1998, Shaw-Smith et al. 2004, de Vries et al. 2005, Slater et al. 2005, Friedman et al. 2006). La hibridació genòmica comparada basada en microarrays (aCGH) i els microarrays de SNP oligonucleòtids d'alta densitat no només proporcionen una major resolució, sinó que també permeten la detecció de guanys i pèrdues submicroscòpiques de material genètic (variants en nombre de còpia, CNVs) a través de tot el genoma. Estudis previs han demostrat que aquestes tecnologies basades en microarrays permeten elevar la taxa de detecció d'aberracions cromosòmiques entre els pacients amb ID a un nivell de fins a un 20% (Shaw-Smith et al. 2004, Ming et al. 2006, Miyake et al. 2006, Wagenstaller et al. 2007)

Aquí presento els resultats obtinguts en nens amb discapacitat intel.lectual que van ser examinats usant la plataforma Illumina d'alt rendiment de genotipació de SNPs. Una cohort de 327 pacients esporàdics amb discapacitat intel·lectual d'origen desconegut i 3 pacients amb una translocació *de novo* aparentment balancejada van ser estudiats de forma sistemàtica per tal d'identificar i caracteritzar les variants causals en nombre de còpia (CNVs).

Hiperinsulinisme congènit (CHI)

Hiperinsulinèmia de la infància (HI, OMIM 601.820), també conegut com hiperinsulinisme congènit (CHI) és la causa més comuna de hipoglicèmia durant el període neonatal i la infància (Stanley, 1997). Es caracteritza per una secreció excessiva inadequada d'insulina per les cèl lules beta pancreàtiques tot i els nivells baixos de glucosa a la sang (Aynley-Green, 1981). L'objectiu principal del tractament CHI és prevenir el dany cerebral irreversible en aquest grup de pacients mitjançant el manteniment de la normoglicèmia. La incidència estimada pot variar des d'1 / 40,000 nens nascuts vius a 1 / 2,500 en els països amb altes taxes de consanguinitat (Glaser et al 2000). CHI és una malaltia heterogènia pel que fa a la presentació clínica, la histologia i la base molecular. Fins ara, s'han trobat vuit gens mutats en CHI: *ABCC8, KCNJ11, GLUD1, GCK, HADH, HNF4A, SLC16A1* i *UCP2* (Thomas et al 1996, Thomas 1995, Stanley et al 1998, Glaser et al 1998, Clayton et al 2001, Pearson et al. 2007, Otonkoski et al. 2007, González-Barroso et al 2008)

El mecanisme més comú subjacent a CHI és la disfunció del canal de potassi sensible a ATP (K^+_{ATP} canal) de les cèl.lules β del pàncrees. Aquest subtipus de CHI correspon a les anomenades canalopaties. Les mutacions recessives en les subunitats SUR1 i Kir6.2 dels canals de K^+_{ATP}, que són codificades per *ABCC8* i *KCNJ11*, respectivament, representen el 40-45% dels casos (Thomas et al 1995, Flanagan et al 2011). El restant 5-10% dels casos de CHI són causats per les anomenades metabolopaties, Aquestes mutacions genètiques provoquen deficiències en factors de transcripció o enzims que condueixen a un augment de la ratio ATP/ADP en les cèl.lules beta o a l'acumulació de metabòlits intermediaris (Flanagan et al 2011).

El CHI d'herència recessiva s'atribueix a *ABCC8, KCNJ11* i a les mutacions del gen *HADH*, mentre que l'herència autosòmica dominant (o de novo) s'ha descrit pels gens *GLUD1, GCK, HNF4A, SLC16A1* i *UCP2*. La presentació clínica de les metabolopaties és molt heterogènia, pel que fa a l'edat d'inici, la gravetat dels símptomes i el grau de penetrància. En canvi, els pacients amb mutacions recessives en *ABCC8* o *KCNJ11* tendeixen a presentar una hipoglucèmia profunda ja en el període neonatal. No obstant això, també s'han identificat mutacions dominants rares en *ABCC8* o *KCNJ11* les quals semblen estar associades a fenotips més lleus i de millor pronòstic (Pinney et al 2008).

S'han descrit dues formes de CHI ; la difusa i la focal (Rahier et al 1984). El CHI focal es caracteritza per una hiperplàsia somàtica esporàdica de les cèl lules beta dels illots, mentre que la CHI difusa correspon a una alteració funcional de la secreció d'insulina en el pàncrees sencer. La forma difusa típica resulta de mutacions recessives en els gens que codifiquen les dues subunitats del canal de K^+_{ATP}. En canvi, la forma

focal resulta de mutacions dels gens *ABCC8* o *KCNJ11* en la línia germinal que són heretades del pare, juntament amb l'haploinsuficiencia somàtica de 11p15.5 en l'al lel matern (de Lonlay et al 1997).La deleció somàtica (només en les cèl lules ß) de la regió 11p15.5 conté els gens *ABCC8* i *KCNJ11* i la regió improntada que inclou els gens *H19, P57KIP2, KVLQT1, HASH2* i *IGF2*. Per tant, l'expressió desequilibrada d'aquests gens improntats, incloent factors de creixement i gens supressors de tumors, i la hemizigositat de la mutació paterna resulta en la proliferació clonal d'illots pancreàtics hiperfuncionals. Ambdues formes, la focal i la difusa comparteixen característiques clíniques similars, però requereixen diferents tractaments quirúrgics. Les lesions focals són tractades eficaçment amb resecció pancreàtica limitada, mentre que les lesions difuses requereixen, pancreatectomia subtotal amb un alt risc associat de diabetis mellitus. Fins ara, el mètode més freqüentment aplicat per diferenciar la forma difusa de la focal, així com per localitzar amb precisió la lesió focal és la fluoro-L-Dopa tomografia per emissió de positrons (PET), una tècnica no invasiva (Ribeiro et al 2005, Otonkoski et al 2006).

Els gens *TBX5* i *TBX3*

Les mutacions en el gen *TBX5* son la causa de la síndrome de Holt-Oram (HOS) (OMIM 142.900), un trastorn autosòmic dominant caracteritzat per alteracions en les extremitats superiors i malformacions del cor (Li et al. 1997, Basson et al. 1997). Les mutacions en *TBX3* se sap que causen la síndrome de Ulnar-mammary (UMS) (OMIM 181.450), un trastorn autosòmic dominant hereditari que afecta les extremitats posteriors, les glàndules apocrines, dents, i el desenvolupament genital (Bamshad et al. 1996, Schinzel et al. 1987). Els estudis genètics basats en seqüenciació directa reporten taxes baixes de detecció de mutacions en aquests gens (Brassington et al 2003, Cross et al 2000,. Heinritz et al 2005). La majoria de les mutacions identificades són mutacions sense sentit (nonsense) o de pauta de lectura (frameshift), donant lloc a un al.lel nul funcional. Les mutacions de canvi de sentit (missense) són menys freqüents, però subjacents a un mecanisme patològic similar, és a dir, de pèrdua de funció, degut a la disminució de la interacció a factors de transcripció o d'unió a l'ADN (Basson et al 1999;. McDermott et al 2005). Per augmentar el rendiment dels mètodes de detecció, s'han emprat mètodes com ara MAPH (multiplex amplificable probe hybridization) i CGH array, que permeteren la identificació de grans delecions genòmiques causants d'haploinsuficiència en els gens *TBX5* i *TBX3*, respectivament (Akrami et al. 2001, Klopocki et al . 2006). Mitjançant l'ús de qPCR també s'ha pogut identificar una deleció contigua hemizigota de *TBX5, TBX3* i *RBM19* la qual dona lloc a un fenotip combinat de les síndromes de Holt-

Oram i de Ulnar-mammary (Borozdin et al. 2006). A més, s'ha trobat una mutació missense subjacent a un mecanisme patològic diferent, és a dir, de guany de funció, en persones amb fenotips atípics de HOS. Aquesta troballa suggereix que altres individus amb fenotip atípic possiblement podrien tenir mutacions *TBX5* de guany de funció. No obstant això, fins ara no s'han trobat duplicacions genòmiques que incloguin els gens *TBX5* i *TBX3*.

Objectius

El primer objectiu d'aquest estudi va ser l'anàlisi sistemàtic d'una cohort de 219 pacients esporàdics amb discapacitat intel·lectual d'origen desconegut per tal d'avaluar i optimitzar la tecnologia de microarrays per a l'anàlisi de variants en nombre de còpia en tot el genoma. Per a això es van utilitzar dos microarrays de SNP oligonucleòtids d'alta densitat (Illumina): 1) els microarrays Human550-Quad per a l'estudi 109 mostres d'ADN i 2) els microarrays Human610-Quadv1_B per estudiar un segon conjunt de 110 mostres d'ADN.

El segon objectiu del nostre estudi va ser l'aplicació d'aquests microarrays per a la detecció de variants en nombre de còpia (CNVs) rares i/o *de novo* en pacients de tres cohorts diferents. Els objectius concrets van ser:

1) Realitzar l'anàlisi de SNP-microarrays en mostres d'ADN de 330 pacients amb discapacitat intel·lectual (ID) d'origen desconegut i amb un cariotip aparentment equilibrat en la majoria d'ells, per tal d'identificar CNVs submicroscòpiques potencialment patògenes mitjançant la plataforma de SNP-arrays de Illumina.

2) Realitzar una exploració de tot el genoma de CNVs en mostres de 40 nens referits per hiperinsulinisme congènit difús (CHI). Com que la detecció convencional de mutacions en els gens candidats *ABCC8* i *KCNJ11* no va revelar cap anomalia detectable, l'ADN dels pacients s'hibridà en microarrays de SNP-oligonucleòtids d'alta densitat (Illumina) per tal de detectar així alteracions cromosòmiques en el genoma i per tant identificar nous gens causants de CHI

3) Realitzar una anàlisi global del genoma de CNVs en mostres d'una extensa família alemanya amb diversos membres afectats d'una forma atípica de la síndrome de Holt-Oram. El cas índex (ID: 31565), sense mutacions detectables en el gen candidat *TBX5*, va ser testat per a la presència de variants cromosòmiques en nombre de còpies mitjançant microarrays de SNP-oligonucleòtids d'alta densitat (Illumina)..

Resultats i Discussió

Anàlisi de dades en una cohort de 330 patients amb ID

Per establir el procediment de funcionament i per optimitzar i estandarditzar els resultats, 110 mostres varen ser hibridades en microarrays Human550-Quad i 109 mostres varen ser hibridades en microarrays Human610-Quadv1_B. Per tal d'avaluar el límit de resolució respecte l'especificitat de cada tipus de microarray, els resultats preliminars de cada conjunt de dades, és a dir les CNVs, van ser validats mitjançant PCR quantitativa. En altres paraules, es va determinar la taxa de falsos positius per a diferents límits de resolució per tal d'establir el llindar en què la detecció de falsos positius és menyspreable. El límit de resolució va ser testat canviant el nombre de SNPs consecutius que havien de ser aberrants (que al seu torn està directament relacionat amb la mida genòmica del llindar). Quan aquest paràmetre de tall es va establir, la resta dels pacients amb ID (111 pacients) es van analitzar amb microarrays Human610-Quadv1_B sense seguiment d'experiments de validació.

Es va observar que pels microarrays 550K, la taxa de falsos positius és del 62,4% en el <10 SNPs bin, el 36,2% en el 10-20 SNPs bin i el 3,1% en el >20 SNPs bin. Per al conjunt de dades 610K les taxes de falsos positius van ser superiors en tots els SNPs bins testats: el 82,8% en el <10 SNPs bin, el 52,8% en el 10-20 SNPs bin i el 12,2% en el >20 SNPs bin. Observem, doncs, que CNVs amb el suport de >20 SNPs s'identifiquen amb un alt grau d'especificitat en els dos tipus de microarrays; 96,9% per als arrays 550K i 87,8% per als arrays 610K. Basant-nos en aquestes dades i aplicant un marge de seguretat conservador es va determinar com a paràmetre de tall la mitjana de les finestres de més de 30-40 SNPs consecutius, és a dir, 200kb-500kb. El llindar de resolució s'aplicà de manera més o menys rigorosa en funció de la qualitat del perfil del senyal d'intensitat de cada array en particular.

En concret, en el grup d'avaluació de 219 pacients, l'anàlisi dels microarrays indicà un total de 4097 variants en nombre de còpia (CNVs), les quals comprenen 3.156 delecions i 941 duplicacions, des de 118 a

17.510.855 parells de base de longitud. 463 de les 4.097 CNVs candidates varen ser avaluades amb 2 o 3 qPCRs cadascuna. 230 de les 463 CNVs avaluades (49,8%) van ser confirmades mentre que 233 CNVs (51,2%) van resultar ser falsos positius. Quan es disposà de mostres d'ADN dels pares es va investigar si la CNV s'originà *de novo*. Pel que fa a les CNVs heretades no es va detectar un biaix d'herència parental (p = 0,5035): 95 CNVs foren d'origen matern, 86 CNVs d'origen patern, i 3 CNVs es van detectar en ambdós pares. Pel que fa al tipus de la CNV, les nostres dades proporcionen una evidència molt significativa (p = 0,0001) que les delecions (164) són més freqüents que les duplicacions (66). Aquestes freqüències no semblen estar en concordança amb la idea que el genoma humà tolera guanys de material genètic millor que les pèrdues. Cal assenyalar, però, que el llindar de l'algoritme afavoreix la detecció de delecions en comparació amb les duplicacions ja que el logaritme del coeficient d'intensitats d'una deleció (que té un ràtio teòric de ½ amb log2 (½) = -1,0) està més distant de la mitjana (log2 (1) = 0) que el logaritme del coeficient d'intensitats d'una duplicació (log2 (3/2) = 0.59). Per tant, existeix una major probabilitat que les duplicacions puguin passar per alt ja que els logaritmes dels seus coeficients d'intensitat estan més a prop del nivell de soroll dels microarrays.

En resum, el nostre estudi ha permès la identificació de 13 síndromes conegudes, 3 translocacions desequilibrades (2 d'elles en estat de mosaic) i 16 CNVs *de novo* potencialment patògenes. Per unes altres 21 alteracions cromosòmiques, l'aparició *de novo* no es va poder establir ja que no disposàvem de mostres dels pares. No obstant això, aquestes CNVs no es van observar en cap dels altres pares testats ni en 4,104 individus control. A més, 7 d'aquestes alteracions sense un origen *de novo* provat es troben en loci on se sap que l'haploinsuficiència és causant de retard mental. A més, es varen detectar 56 CNVs rares heretades privadament. Entre elles cal destacar una deleció d'herència materna a la banda Xq13.1 en un pacient de sexe masculí. La seva mare va mostrar tenir un patró no aleatori d'inactivació del cromosoma X transmès. A més es van identificar sis reordenaments rars (en 16p13.11, dos a 15q13.3, en 1q21.1, en 22q11.2 i en 12q14), els qual han estat prèviament associats amb trastorns neuropsiquiàtrics. A part d'aquests 6 reordenaments, 50 CNVs rares heretades van ser identificades en la nostra cohort ID, les quals contenen seqüències codifiacants i no són conegudes per la Database of Genomic Variants (DGV). 30 d'aquestes 50 variants no van ser trobats tampoc en la nostra cohort de controls.

Per concloure, el rendiment diagnòstic d'aquest estudi va ser del 9,7% (i fins a un 16,4%) el qual es troba en concordança amb les dades publicades en altres articles (de Vries et al. 2005, Friedman et al. 2006, Vermeesch et al. 2007, Hoyer et al. 2007, Hochstenbach et al. 2009).

Delecions heterozigotes del gen *FOXP1*

Aquí presento la identificació de delecions heterozigotes que afecten el gen *FOXP1* (MIM # 605515) en tres pacients no relacionats amb ID i amb trastorn significatiu de la parla i del llenguatge (Horn i Kapeller et al. 2010)

En un esforç col·laboratiu de diferents centres es va realitzar una anàlisi global del genoma amb microarrays per tal d'identificar CNVs en una cohort alemanya de 1.523 pacients no relacionats amb retard mental. El reclutament de pacients formà part de l'estudi de la Xarxa Alemanya de Retard Mental (MRNET). Les dades de CNVs van ser generades per les diferents institucions alemanyes pertanyents al consorci MRNET fent ús de diferents tecnologies basades en microarrays. Per tal d'agrupar i compartir dades de microarrays es va crear una base de dades i es posà a la lliure disposició de tot el consorci (www.german-mrnet.de).

Aquests estudis de variants en nombre de còpia revelà tres delecions heterozigotes solapants en el cromosoma 3p14.1 que afecten únicament el gen forkhead box P1 (*FOXP1*). El pacient 1 forma part de la cohort de 330 pacients aquí reportats, i investigats amb Infinium Human550-Quad i Human610-Quad arrays (Illumina). El pacient 2 forma part d'una cohort de 188 pacients investigats utilitzant microarrays d'oligonucleòtids 244K (Agilent Technologies, Santa Clara, CA). El pacient 3 pertany a una cohort de 184 pacients analitzats amb SNP arrays 6.0 (Affymetrix, Santa Clara, CA). Els tres pacients, dos nens i una nena d'entre 5,5 a 7 anys, presenten un retard mental moderat i un dèficit significatiu del llenguatge i la parla.

La investigació clínica detallada dels tres pacients va revelar que presenten ID moderat en combinació amb un retard general del desenvolupament. En tots els pacients, el coeficient intel·lectual no verbal es va avaluar com ≤ 50 (3 SD sota de la mitjana). El desenvolupament de la parla i del llenguatge es va estimar en el mateix rang. Els tres pacients van començar a parlar a l'edat de 3,5 anys i utilitzaren només combinacions de dues paraules als 5, 5,5 i 7 anys, respectivament. En tots els pacients, el vocabulari productiu i receptiu va ser de menys de 100 paraules. El llenguatge expressiu es va veure més afectat que les habilitats receptives. En tots els pacients es trobà present una articulació de la parla molt pobre, i dificultats de producció de les consonants al principi de les paraules. Dos pacients van presentar problemes oromotòrics incloent dificultats de protrusió del llavi. En la infància, els pacients tenien una tendència a mantenir la boca oberta i el pacient 2 va patir dificultats per empassar. A més, tots van mostrar un considerable retard del desenvolupament motor per caminar sense suport a partir dels 24 a 36 mesos. La

ressonància magnètica cerebral i l'electroencefalograma no van revelar cap anormalitat. Cap dels pacients va mostrar una pèrdua auditiva neurosensorial. Proves oftalmològiques van revelar miopia moderada només en el pacient 1. En l'examen físic, els paràmetres de creixement en relació amb l'altura i el perímetre cefàlic occipitofrontal van ser documentats com a normals per dos pacients, tot i que la presència d'obesitat va ser notable en ambdós. Anomalies craniofacials consistents en els pacients 2 i 3 inclouen el front ample i prominent i cabell frontal elevat.

Les anàlisis de microarrays van mostrar que el pacient 1 presenta una deleció de 498 kb que afecta a tots els exons codificants, excepte el primer, del gen *FOXP1* (Wagenstaller et al 2007). Aquesta deleció, està indicada per 89 SNPs, i va ser confirmada per qPCR tot i que el seu origen *de novo* no es va poder demostrar ja que l'ADN del pare no estava disponible. No obstant això, es va poder demostrar que la mare no era portadora de la deleció i que l'al.lel delecionat era d'origen patern. La seqüenciació de la regió codificant de l'al.lel FOXP1 restant no va revelar cap variació en la seqüència en comparació amb la seqüència de referència (NM_032682). Varem acotar fins al punt de ruptura de la deleció mitjançant PCR quantitativa i varem amplificar i seqüenciar el fragment d'unió per PCR en l'ADN genòmic del pacient (GenBank No EF504249). La mida de la deleció és de 498.198 parells de base i es situa en la posició Chr3: 70.807.767-71.305.965 de hg18, UCSC Genoma Humà març de 2006. El punt de ruptura distal es troba dins d'una repetició LINE (long interspersed nuclear element) (L1/L1PA5) "downstream" de *FOXP1*, mentre que el punt de ruptura proximal es troba en una repetició tipus MER1 (MER58A) en l'intró 6. No es trobaren seqüències estimulants de recombinació homòloga dins o al voltant del punt de ruptura.

Per als pacients 2 i 3 es van identificar delecions de 659 kb i 1047 kb que afecten tota la regió codificant del gen *FOXP1*. Les delecions van ser verificades per hibridació in situ fluorescent (FISH). L'anàlisi de l'ADN dels pares mitjançant cariotipació molecular i FISH va indicar que tots dos pacients presenten delecions *de novo*. L'origen parental de les delecions és patern en ambdós casos. La seqüenciació de la regió codificant dels al.lels *FOXP1* restants no va revelar variacions en la seqüència. La caracterització dels punts de trencament es va aconseguir mitjançant l'amplificació i seqüenciació dels fragments d'unió respectius (GenBank No HM124444 i GU980955). No hi ha seqüències repetitives presents en els loci de punt de trencament en el pacient 2. No obstant això, els dos punts de ruptura es troben dins d'una microhomologia de tres nucleòtids (AGA) al costat telomèric en la posició 70.778.067-70.778.070 i al costat centromèric a 71.437.354-71.437.357 que resulta en una mida de la deleció de 659.287 parells de

bases.

Per al cas 3, el punt de ruptura telomèric es troba en la posició 70.341.246-70.341.247 i el punt de ruptura centromèric es troba a 71.388.173-71.388.174 resultant en una mida de la deleció de 1.046.927 parells de bases. En el pacient 3, el punt de ruptura distal de la deleció es troba en un SINE (short interspersed nuclear element), mentre que no hi ha seqüències repetitives en el punt de ruptura proximal.

L'absència de regions homòlogues en els punts de ruptura suggereix que les tres delecions van ser probablement generades per ruptures de doble cadena i una posterior fusió d'extrems no homòlegs, NHEJ.

A continuació varem comprovar el locus *FOXP1* per CNVs tant en la base de dades DGV (Database of Genomic Variants) com en una base de dades pròpia amb CNVs de 4.104 controls. No trobàrem CNVs que afectin la regió codificant a la base de dades DGV, però es va detectar una deleció de 1,3 Mb en 3p14.1p13 que abarca els gens *FOXP1, EIF4E3* (MIM # 609896), *PROK2* (MIM # 607002) i *GPR27* (MIM # 605187) en una persona de la nostra base de dades de controls. L'aplicació del test exacte de Fisher en aquestes dades, no tenint en compte els controls continguts en la base de dades DGV, va donar lloc a un p-valor nominal de 0,06. Encara que l'origen *de novo* de les delecions suggereix un paper de *FOXP1* en el quadre clínic d'aquests pacients, cal fer més estudis genètics poblacionals per dilucidar si aquest interval cromosòmic presenta o no penetrància incompleta. No obstant això, tres publicacions addicionals han confirmat que delecions *de novo* de *FOXP1* provoquen un deteriorament greu del llenguatge i retard mental (Pariani et al. 2009, Hamdan et al. 2010, Carr et al 2010).

Per tal de seguir donant suport a la idea que delecions de *FOXP1* causen retard mental i dèficit del llenguatge, es van revisar 883 pacients de la base de dades MRNET per mutacions puntuals en *FOXP1*. Es van identificar vuit variants no-sinònimes, tres sinònimes i nou variants no codificants. Totes les variants es genotiparen en un màxim de 676 controls sans no relacionats. Tres de les variants codificants identificades, p.Ser5Pro, p.Pro215Ala i p.Thr390Ser també es van trobar presents en freqüències similars en controls sans, però cap d'aquestes variants es trobà en la base de dades HapMap o en la base de dades NCBI dbSNP. Cinc de les variants no sinònimes, p.Ser5Pro, p.Gln76dup, p.Pro215Ala, p.Thr390Ser p.Asn570Ser es transmeteren a través d'un pare aparentment no afectat. Per a les tres variants restants, p.Met101Val, p.Ser261Pro, p.Asn597Thr, no es va poder establir si l'origen és *de novo*, a causa de la manca d'ADN parental. Les variants transmeses per un pare no afectat són generalment classificades com no patògenes. No obstant això, suposant que ID segueixi un model de llindar multigènic, no es pot descartar completament la pertinència d'aquestes variants a la malaltia. Per descobrir el significat neurobiològic

d'aquestes variants, és necessari que s'investiguin els efectes funcionals d'aquestes mutacions puntuals. No obstant això, altres estudis han reportat que mutacions puntuals *de novo* en *FOXP1* són la causa de greus trastorns del llenguatge i de retard mental (Hamdan et al. 2010).

Síndrome de Cohen

Aquí presento la identificació de delecions intragèniques heterozigotes que afecten el gen *COH1* (MIM # 607817) en tres pacients no relacionats amb discapacitat intel.lectual (ID). La seqüenciació posterior del gen *COH1* va revelar mutacions puntuals en el segon al.lel en els tres pacients analitzats, proporcionant així un diagnòstic diferencial de Síndrome de Cohen (Rivera-Brugués et al. 2010).

Hem analitzat les dades d'arrays de SNP oligonucleòtids d'alta densitat per tal d'identificar variants en nombre de còpies (CNVs) en pacients amb retard mental (n = 1523) i en controls normals (n = 1612). El reclutament de pacients ha estat part de l'estudi de la Xarxa Alemanya de Retard Mental (MRNET) (http://www.german-mrnet.de/). Com a conjunt de controls, hem utilitzat les dades de CNVs (n = 1612) de l'estudi poblacional KORA.

Els pacients 1 i 2 formen part de la cohort de 330 pacients aquí reportats, i investigats amb Infinium Human550-Quad i Human610-Quad arrays (Illumina). El pacient 3 forma part d'una cohort reclutada per la institució d'Essen i es va analitzar amb el SNP array 6.0 (Affymetrix, Santa Clara, CA). Els tres pacients, dos nens i una nena amb edats d'entre 18 mesos i 3 anys, presenten un sever retard mental i un retard del desenvolupament.

El pacient 1 va néixer a terme després d'un embaràs sense complicacions com el primer fill de pares àrabs no relacionats i sans. No es va registrar el pes en néixer, ni la longitud ni la circumferència del cap. En el seu primer any de vida es va fer evident un retard en el desenvolupament motor. Va començar a seure a l'edat de 18 mesos, i a caminar als 24 mesos. Un examen detallat a l'edat de 3 anys va mostrar un nen hipotònic amb una alçada de 79cm (-4.8SD), un pes de 9300g (-5.2SD) i una circumferència del cap de 44 cm (-4.9SD). Es va fer evident un retard en el desenvolupament de la parla, així com en la comprensió. El nen presentava dimorfisme craniofacial lleu, incloent les celles horitzontals, una punta del nas ampla, una columela àmplia, el solc nasolabial curt, un llavi superior prim, i un llavi inferior evertit. L'examen oftalmològic va revelar miopia bilateral, astigmatisme i un lleu augment de la pigmentació de la retina. No es va trobar neutropènia.

El pacient 2 va néixer a terme després d'un embaràs sense complicacions, com el segon fill de pares

alemanys sans no relacionats, amb un pes de 2460g (-2.4SD), una longitud de naixement de 46,5 cm (-2.3SD), i una circumferència del cap de 31,5 cm (-2.9SD). Les fites del desenvolupament es van endarrerir; amb una edat de seure de 12 mesos i una edat de gatejar de 17 mesos. L'examinació a l'edat de 18 mesos va mostrar un nen hipotònic amb un pes de 10.1kg (-1.5SD), una alçada de 80cm (-1SD) i una microcefàlia severa amb un OFC de 42.5cm (-4.8SD). El desenvolupament de la parla no s'havia produït, però la comprensió era pràcticament normal. El dimorfisme facial consistí en un lleu aplanament occipital, celles horitzontals, fissures palpebrals en forma d'ametlla, una arrel nasal àmplia, una punta nasal rodona, una columela gruixuda, un solc nasolabial curt, un aspecte obert de la boca amb una geniva superior prominent, i una gran bretxa entre els incisius. La funduscòpia i el comptatge sanguini complet no revelaren cap anormalitat.

La pacient 3 és la segona filla d'una parella sana no consanguínia alemany/africana. El naixement prematur es va produir a les 35 setmanes de gestació amb un pes en néixer de 2450g (-0.1SD), una longitud de 46cm (-0.4SD) i OFC de 33.5cm (0.6SD). Després del naixement se li va diagnosticar un defecte en el cor (ASD II i estenosi pulmonar) i arítmies. El retard en el desenvolupament es va mantenir fins als 2 anys d'edat el qual va millorar després de la correcció quirúrgica del defecte del cor. La pacient va començar a caminar a l'edat de 24 mesos i a l'edat de 2 ¾ anys va presentar mesures del cos normals, amb una alçada de 95cm (0.4SD), un pes de 11.5kg (-1.4SD) i OFC de 50cm (0,6 SD). Tenia una cara plana amb un pont nasal ampli i pla i els ulls ametllats, el solc nasolabial curt i orelles profundes amb les hèlixs replegades. La funduscòpia i el comptatge sanguini complet revelaren resultats normals.

L'anàlisi de microarray de SNP oligonucleòtids d'alta densitat i la posterior validació amb qPCR van permetre identificar CNVs en el gen *COH1* en els tres pacients: una deleció de 67kb heretada de la mare la qual abasta els exons 26-31 del gen *COH1* (chr8: 100.573.090...100.639.924; c.3871-5024del/p.G1291fsX42) en el pacient 1, una deleció de 193kb heretada del pare la qual abasta els exons 9-19 del gen *COH1* (chr8: 100.216.034 ... 100.409.167; c.1207-2824del/p.L403fsX11) en el pacients 2 i una deleció de 315 kb d'herència materna la qual abasta l'exó 4 del gen *OSR2* i els exons 1-17 del gen *COH1* (chr8: 100.015.029 ... 100.347.846; c.1-2515del) en el pacient 3. No es van detectar aquests canvis en els 1612 controls.

La seqüenciació dels gen *COH1* va permetre identificar en el pacient 1 una mutació paterna en l'exó 32 la qual provoca un codó de terminació (c.5086C> T/p.R1696X), en el patient 2 una deleció materna d'1pb en l'exó 60 que condueix a un codó de terminació (c.11505delA/p.K3835fsX43) i en el pacient 3 una mutació

de canvi de sentit heterozigota en l'exó 25 (c.3866C> G/p.T1289S) i una inserció heterozigota de tres parells de base en l'exó 62 (c.11827_11828insATG/p.D3942_G3943insD), ambdós canvis heretats del pare. Aquestes mutacions no es van trobar presents en 676 mostres procedents d'una cohort de població general (KORA) i no estaven anotats com a polimorfismes en HapMap o la base de dades NCBI dbSNP

El nostre estudi va demostrar que la síndrome de Cohen, un trastorn rar autosòmic recessiu, pot estar present en pacients amb retard mental i només ser detectable pels mètodes d'anàlisi global del genoma, com ara arrays d'oligonucleòtids d'alta resolució, i no per l'exàmen clínic.

Hiperinsulinisme congènit (CHI)

L'Hiperinsulinisme congènit (CHI) és una malaltia heterogènia amb mutacions en vuit gens diferents. La base genètica de CHI involucra defectes en els gens clau que regulen la secreció d'insulina de les cèl lules beta. No obstant això, les mutacions en tots aquests gens representen al voltant del 50% de les causes conegudes de CHI. Hi ha tres subtipus principals de CHI: formes difuses, focals i atípiques. La forma difusa s'hereta principalment en forma autosòmica recessiva i la forma focal presenta herència.esporàdica.

Aquí hem realitzat una anàlisi de tot el genoma amb microarrays per tal de detectar CNVs en 40 nens amb hiperinsulinisme congènit difús. Com que la detecció de mutacions convencionals en els gens *ABCC8* i *KCNJ11* no va revelar cap anomalia detectable, l'ADN dels pacients s'hibridà en Human610-Quad arrays (Illumina).

Hem construït una cohort de 40 pacients amb hiperinsulinisme congènit difús. 23 d'aquests pacients havien estat avaluats per indels, mutacions en llocs d'empalmament (splice-site) i mutacions en la regió promotora de *ABCC8* mitjançant la seqüenciació en un sistema Illumina GAII. D'altra banda, les regions exòniques del gen *ABCC8* varen ser estudiades per delecions i duplicacions amb PCR quantitativa en temps real en tots els 40 pacients. En 8 casos es va detectar una mutació en un al.lel, mentre que no es van trobar mutacions en els restants 32 casos. Es va aplicar l'anàlisi de SNP arrays per tal d'oferir una estudi més ampli dels pacients i detectar així alteracions cromosòmiques en el genoma i per tant identificar nous gens causants de CHI.

Mitjançant l'anàlisi d'arrays d'oligonucleòtids i la posterior validació de qPCR es van identificar quatre alteracions cromosòmiques en cinc pacients amb CHI difús.

Dos germans (ID: 45743 i 45744) van mostrar la mateixa alteració cromosòmica complexa *de novo,* que correspon a dos grans delecions en el braç p del cromosoma 9 i una duplicació en el cromosoma 13. La

implicació de dos cromosomes diferents i la ocurrència d'aquest tipus d'alteració en dos germans afectats indiquen que aquests pacients són portadors d'una translocació no equilibrada entre els cromosomes 9 i 13. Aquesta translocació va ser transmesa per un dels pares sans que porta la translocació recíproca t(9;13) equilibrada. Hem observat que els punts de ruptura de la CNV detectada en el cromosoma 13 són gairebé idèntics entre els dos germans, donant com a resultat en ambdós casos una duplicació de 7,85 Mb a 13q33.3-q34. En canvi, les delecions del cromosoma 9 difereixen lleugerament entre els dos nens. El pacient 45743 presenta una deleció de 6,3 Mb a 9p24.3-p24.1 i una deleció de 4 Mb a 9p24.1-p23, mentre que el pacient 45744 mostra una deleció de 4,3 Mb a 9p24.3-p24.2 i una deleció de 7,8 Mb a 9p24.1-p23. L'ADN del pacient 45744 s'ha analitzat per segona vegada amb SNP-arrays i en aquesta anàlisi només es van identificar dos grans desequilibris, una deleció de 12,4 Mb en el cromosoma 9 i una duplicació de 7,9 Mb en el cromosoma 13. Les dades dels arrays d'oligonucleòtids s'utilitzen per identificar inicialment posicions aproximades dels punts de ruptura en el genoma. No obstant, aquesta tecnologia té una limitació de resolució a nivell de parells de base. D'altra banda, la inexactitud de les CNVs depèn en gran mesura del disseny del microarray (és a dir, la cobertura de SNPs), l'algoritme utilitzat (és a dir, diferents llindars), i la variabilitat experimental. Per tant, altres tècniques com la PCR quantitativa, MLPA i l'amplificació de fragments d'unió seran necessàries per mapar amb precisió els punts de ruptura i determinar l'extensió exacta d'aquests reordenaments.

Un altre pacient (ID: 30370) amb hiperinsulinisme congènit difús va mostrar també una alteració complexa. En aquest cas, hem detectat dos CNVs en el cromosoma X, una deleció de 2,7 Mb a p22.33 i una duplicació de 16,2 Mb a q11.1-q22.1. A causa de la manca de mostres dels pares no es va poder investigar l'origen *de novo* d'aquestes variants.

Més uns 300 gens Ensembl són anotats en les CNVs identificades en els dos germans, i més de 426 gens en les regions alterades en l'últim pacient. Cap dels gens inclosos en aquestes CNVs es troben relacionats amb CHI, en base a les dades disponibles (expressió en el pàncrees, funció en el metabolisme, interacció amb gens relacionats amb la malaltia) de les actuals bases de dades públiques (UCSC Genome Browser, Ensembl, NCBI) i de la literatura.

El següent pacient (ID: 44216) presenta una duplicació de 579kb a 8q24.23 la qual conté 195 SNPs. No s'ha pogut determinar la seva ocurrència *de novo*, perquè l'ADN dels pares no estava disponible. El punt de ruptura distal de la duplicació es va determinar en l'intró 3 del gen *FAM135B*. Quatre entrades Ensembl són anotades a la regió de la corresponent CNV i corresponen a les isoformes de *FAM135B*. Aquest gen

codifica una proteïna hipotètica, LOC51059, la funció biològica de la qual és desconeguda. No s'ha descrit fins ara cap CNV polimòrfic que afecti les regions exòniques d'aquest gen. No obstant això, altres estudis seran necessaris per determinar si la dosi gènica de *FAM135B* o d'un altre gen del genoma és alterada per aquesta duplicació, la qual cosa permetria assegurar la patogenicitat d'aquest reordenament i identificar *FAM135B* com a gen candidat per a CHI.

L'última alteració cromosòmica detectada és una deleció de 970kb al braç petit del cromosoma 1 en la banda p31.1 (posició física hg18, UCSC Genoma Humà març de 2006; chr1: 70.694.440-71.665.008). Aquesta CNV es va detectar en el pacient 42667 i mitjançant experiments de qPCR es va demostrar que s'originà *de novo*. Aquesta variació en el nombre de còpies cromosòmiques abasta el gens *PTGER3* i *ZRANB2* i l'últim exó del gen *NEGR1*. El gen *PTGER3* (prostaglandin E receptor 3) codifica per un dels receptors identificats de la prostaglandina E2 (PGE2). La prostaglandina E2 és un inhibidor de la secreció d'insulina estimulada per glucosa (GSIS) i se sap que actua mitjançant el receptor *PTGER3*. La implicació de *PTGER3* sobre els mecanismes de regulació de la insulina ha estat destacada en la literatura. Una publicació va reportar la generació d'un model de ratolí knock-out *PTGER3* (EP3 -/-), el qual exhibeix un fenotip caracteritzat per nivells elevats de leptina i d'insulina i el pes de més d'un 20% en comparació amb els ratolins salvatges (Sánchez-Alavez et al. 2007). A més, les investigacions funcionals de Meng et al. (2006) van demostrar que la PGE2 pot induir la disfunció de les cèl lules beta del pàncrees mitjançant la inducció de l'expressió gènica de *PTGER3,* alterant així la via de senyalització i els factors de transcripció implicats específicament en la secreció d'insulina. D'acord amb aquestes informacions, aquí hipotetitzem que la pèrdua de funció de *PTGER3* podria ser la causa de l'hiperinsulinisme sofert pel pacient. Per tant, *PTGER3* va ser seleccionat com a gen candidat i la nostra cohort va ser escrutinada per mutacions puntuals.

Hem dissenyat encebadors per amplificar els amplicons que cobreixen les regions exòniques i els limits exó-intró dels 14 trànscrits de *PTGER3*.

Per seqüenciació directa varem buscar mutacions puntuals en l'al.lel *PTGER3* restant del pacient 42667 per tal de descobrir una mutació recessiva putativa. També es va realitzar l'anàlisi mutacional de *PTGER3* en els 39 pacients restants de la nostra cohort CHI. La seqüenciació no va revelar cap variant en el pacient 42667, però ens va permetre identificar en el pacient 33531 una deleció heterozigota de 1pb en l'exó 4 de la isoforma 8 de *PTGER3,* la qual condueix a un codó de terminació prematura de la traducció després de dotze aminoàcids (c.1185delC / p.N395QfsX12, model gènic NM_198718.1/NP_942011.1 basat en hg18,

UCSC Genoma Humà març de 2006). L'al.lel nul identificat prediu el truncament de la proteïna en el domini C-terminal de la isoforma específica. La variant de seqüència identificada, no va ser anotada com a polimorfisme en les bases de dades HapMap o NCBI dbSNP. No obstant això, l'anàlisi va mostrar que la variant és heretada de la mare la qual és aparentment no afecta, cosa que apunta a la falta de patogenicitat de la variant. Varem genotipar la variant seleccionada en un grup control de 676 individus no relacionats de la cohort KORA pel sistema MassARRAY (plataforma Sequenom) i varem trobar la variant en 18 individus control. A més, es va avaluar l'expressió específica d'ARNm de la isoforma EP3-8 en el pàncrees humà. Es va realitzar un anàlisi de RT-PCR en cDNA de pàncrees humà (Marathon Ready cDNA, Clonthech) mitjançant l'ús de dues passes imbricades d'amplificació PCR. La combinació de tres conjunts d'encebadors va permetre amplificar transcrits PTGER3 específics i inespecífics. Varem seqüenciar el producte RT-PCR amplificat i la seqüència de nucleòtids s'alineà a BLAST. Es va confirmar que no hi ha expressió de la isoforma PTGER3 8 (NM_198718.1) en els teixits del pàncrees. En conjunt, els resultats de genotipació i d'expressió demostren que la variant de seqüència identificada en el gen PTGER3 no és el defecte molecular responsable de l'hiperinsulinisme sofert pel pacient 33531. En canvi, aquesta variant correspon a un polimorfisme rar present en la població control i que podria afectar únicament a l'extrem carboxil de la isoforma EP3-8 la qual no s'expressa en el pàncrees.

Podem concloure que, a causa dels resultats negatius obtinguts en la detecció de mutacions en la nostra cohort CHI, *PTGER3* no és una causa freqüent de CHI. No obstant això *PTGER3* pot encara ser considerat com a un gen candidat prometedor per CHI. Estudis en cohorts més grans d'individus amb fenotip similar al del pacient índex permetrà dilucidar la implicació real d'aquest gen en l'etiologia de CHI

Identificació d'una duplicació contigua dels gens *TBX5* i *TBX3*

Aquí presento una duplicació contigua heterozigota dels gens *TBX5* i *TBX3* la qual produeix un fenotip combinat de la síndrome de Holt-Oram i glàndules mamàries supernumeràries. Al nostre entendre, aquesta és la primera descripció de guany de funció de *TBX5* causat per una duplicació genòmica en una família amb símptomes lleus de l'esquelet i defectes cardíacs d'expressió variable. Aquesta és també la primera vegada que una duplicació del gen *TBX3* és identificada i associada amb el desenvolupament de glàndules mamàries accessòries.

Aquest estudi va incloure un total de 11 persones pertanyents a una família alemanya, on cinc dels membres presentaven un diagnòstic inicial de síndrome de Holt-Oram.

Tots els pacients van ser investigats amb anterioritat per la presència de mutacions en el gen *TBX5* amb resultats negatius. Els subjectes afectes presenten malformacions del membre superior anterior, amb o sense cardiopatia congènita o anomalies de la conducció cardíaca, i alguns pacients van mostrar a més glàndules mamàries addicionals.

El primer membre d'aquesta família, possiblement afectat per una malaltia del cor, era un home nascut el 1852 (I / 1). Aquest individu havia patit neurosi del cor i es trobava limitat en les seves activitats. Va morir a l'edat de 84 anys. Un dels seus dos fills, es va veure afectat per la neurosi del cor d'una manera similar (II / 1). Aquest estava exempt del servei militar. La seva filla era sana (III / 2), però el seu fill es va veure afectat de la mà i per malalties del cor (III / 1). La seva articulació del polze es trobava limitada en ambdues mans i es queixava de dolor al cor. A l'edat de 62 anys va patir un infart agut de miocardi. Va morir als 86 anys. Tres dels seus quatre fills tenen anormalitats a la mà. Un dels fills presenta alteracions en les articulacions del polze esquerre (IV / 1). A més, se li va diagnosticar alteracions en la vàlvula aòrtica bicuspidal, donant lloc a un soroll sistòlic. Una de les seves filles pateix d'immobilització en els polzes de les dues mans (IV / 2). La segona filla té polzes i articulacions immòbils (IV / 3). En la propera generació cinc membres es veuen afectats. Els dos fills de IV / 1, bessons, tenen articulacions immòbils en els dos polzes. Es va diagnosticar una estenosi aòrtica en el fill (V / 2), corregida amb una operació a les sis setmanes després del naixement. IV / 3 té dos fills no afectes, mentre que tres dels nens de IV / 2 es veuen afectats. El fill (V / 3) té una extensió limitada del dit petit de les dues mans. Després d'una investigació detallada es va detectar un defecte del septe atrioventricular (ASD II) en l'edat de set anys, el qual va ser corregit amb una operació. A més el pacient presentava glàndules mamàries accessòries en la línia anterior axil·lar a la banda esquerra i dreta del tòrax. La seva germana (V / 4) té un dèficit d'extensió en l'articulació mitjana de la falange del polze i un defecte del cor (ASDII), que va ser corregit amb una operació en el primer any de la seva vida. També presenta glàndules mamàries accessòries. La següent filla de IV / 2 no es veu afectada, però la filla menor té cardiopatia complexa (ASD II, estenosi de l'aorta), contractures severes i una desviació dels dits d'ambdues mans. El desenvolupament psicomotor va ser normal en cada persona afectada.

El pacient índex (ID: 31565), sense mutacions detectables en el gen *TBX5*, va ser analitzat per a la presència de CNVs amb un microarray d'alta densitat d'oligonucleòtids. Es va identificar una duplicació heterozigota a la banda cromosòmica 12q24.21, que abasta tota la regió codificant del gen *TBX5*. Aquesta CNV s'indicà per l'augment dels valors de nombre de còpies de 64 SNP (mitjana 2,37) i que abasta 210kb

(chr12: 113.275.838 -113.486.329 ... basat en hg18, UCSC Genoma Humà març de 2006). 9 CNVs addicionals van ser detectades amb aquesta anàlisi, però foren anotades com a polimorfismes o no contenien gens OMIM. La validació de la CNV i la caracterització dels punts de ruptura es va realitzar mitjançant PCR quantitativa en temps real.

Per confirmar la duplicació del gen *TBX5* tots els individus afectes i no afectes van ser examinats per qPCR utilitzant amplicons posicionats en els exons 4 i 9 del gen *TBX5.* 30 parells d'encebadors addicionals van ser dissenyats per traçar bé els punts de ruptura proximal i distal de la regió veïna de TBX5. En el costat centromèric de la duplicació el primer amplicó conservat es va posicionar en 113,46 Mb al braç q del cromosoma 12, mentre que a l'extrem telomèric de la duplicació el primer amplicó conservat es va posicionar en 113,60 Mb. Curiosament, la duplicació afecta no només el gen *TBX5*, sinó també el gen *TBX3. TBX3* està aproximadament 315 kb cap als telòmers del gen *TBX5*, ambdós ubicats a la banda 12q24.21. Per tant, després de delinear els punts de ruptura amb qPCR la mida màxima de la duplicació es va estimar aproximadament en 345,6 kb (chr12: 113.262.450 ... 113.608.018, hg18, UCSC Genoma Humà març de 2006), i inclou tant el gen *TBX5* com el gen *TBX3*.

L'avaluació de dosis gènica en loci genètics específics va demostrar que la duplicació i el fenotip de la síndrome de Holt-Oram + glàndules mamàries addicionals cosegreguen en aquesta família.

Aquí hem identificat una duplicació contigua heterozigota contenint els gens *TBX5* i *TBX3* la qual dona lloc a un fenotip combinat de síndrome de Holt-Oram i glàndules mamàries addicionals. El guany de funció de *TBX5* causa símptomes lleus de deformacions esquelètiques en la majoria dels pacients afectats i una malaltia cardíaca congènita en només alguns d'ells. També es descriu una primera relació genotip-fenotip d'un guany de funció de *TBX3.* Al contrari a l'haploinsuficiència de *TBX3,* la qual produeix apoplàsia o hipoplàsia de la mama, l'augment de la dosi de TBX3 provoca el desenvolupament addicional de glàndules mamàries.

Conclusions

1) En aquest estudi hem utilitzat i validat les plataformes Human550-Quad i Human610-Quadv1_B de Illumina. Per a la validació, els microarrays van ser aplicats a 110 i 109 pacients amb discapacitat intel.lectual no diagnosticada, respectivament. Mitjançant la validació de les dades de microarrays amb PCR quantitativa observem que les CNVs amb el suport de >20 SNPs són identificades amb un alt grau d'especificitat en els dos tipus de microarrays; 96,9% per als arrays 550K i 87,8% per als arrays 610K.

2) Hem estat capaços de detectar aberracions cromosòmiques críptiques *de novo* o causants de malaltia en el 9,7% de pacients no seleccionats amb discapacitat intel.lectual d'etiologia desconeguda. En total, es van detectar síndromes conegudes en 13 dels 330 pacients ID, es van descriure translocacions desequilibrades en 3 casos (2 d'elles en estat de mosaic), i es van identificar CNVs rars *de novo* en 16 pacients. CNVs potencialment causals (sense confirmació *de novo*) van ser detectats en 21 pacients (en total 16,4%). A més, CNVs rares heretades privadament es van observar en 56 pacients amb ID desconegut.

3) Hem demostrat que els microarrays basats en SNPs permeten la detecció de delecions i duplicacions intragèniques. La identificació d'una CNV que afecta només a un sol gen permet que es consideri aquest gen en particular com a candidat per a la discapacitat intel.lectual. Aquest va ser el cas de tres pacients no relacionats amb discapacitat intel.lectual moderada, retard en el desenvolupament global, i trastorns greus del llenguatge en els quals es va detectar una deleció *de novo* que abasta únicament el gen *FOXP1*.

4) Per demostrar encara més la causalitat de les delecions de *FOXP1*, hem examinat tota la regió codificant de *FOXP1* (exons i seqüències intròniques flanquejants) per a canvis de nucleòtids en un panell de 883 pacients amb retard mental. Es van identificar vuit variants codificants no sinònimes, tres variants codificants sinònimes i nou variants no codificants. No hi ha proves concloents per a la patogenicitat de tres variants no sinònimes (p.M101V, p.S261P, p.N597T) ja que l'ADN dels pares no estava disponible.

5) A més dels casos *de novo* de ID, també es van identificar en la nostra cohort pacients amb una forma autosòmica recessiva de ID. Es van detectar tres delecions heterozigotes parcials del gen *COH1* en el locus 8q22 el qual es troba mutat en la síndrome de Cohen.

6) Al seqüenciar la regió codificant completa i els límits exó/intró de *COH1* es va identificar en l'al.lel restant una mutació stop, una mutació frameshift i dues mutacions missense, respectivament. Per tant, tres mutacions heterozigotes compostes van ser detectades en el gen *COH1*: c.5086C> T / c.3872-5024del [p.R1696X / p.G1291fsX42], c.1207-2824del / c.11505delA [p.L403fsX11 / p.K3835fsX43] i c.1-2515del / (c.3866C> G / c.11827_11828insATG) [- / (/ p.D3942_G3943insD)], proporcionant així un diagnòstic diferencial de la síndrome de Cohen en tres pacients no relacionats de la nostra cohort ID.

7) Hem estudiat la base genètica d'un trastorn rar autosòmic humà anomenat Hiperinsulinisme congènit (CHI) difús, en una cohort de 40 pacients sense mutacions en els gens candidats *ABCC8* i *KCNJ11*.
Les anomalies cromosòmiques detectades amb els microarrays d'oligonucleòtids representen el 20% dels casos estudiats. Quatre aberracions cromosòmiques heterozigotes van ser identificades en cinc pacients CHI com a la causa més probable de CHI.

8) El reordenament més interessant de la cohort CHI va ser la deleció de 970kb en la banda cromosòmica 1p31.1 que inclou els gens *PTGER3* i *ZRANB2* i l'últim exó del gen *NEGR1*. Aquí hipotetitzem que la haploinsuficiència del gen *PTGER3* indueix a una reducció del 50% de l'estimulació de la prostaglandina 2 (PGE2), disminuïnt així la inhibició de la secreció d'insulina estimulada per glucosa (GSIS) i per tant resultant en una secreció elevada d'insulina.

9) L'anàlisi de mutacions puntuals en el gen candidat *PTGER3* no va revelar cap variant patògena ni en el segon al.lel del pacient en el qual s'havia detectat la deleció *de novo* ni en una cohort de 39 pacients no relacionats amb CHI. En canvi es va identificar una nova variant polimòrfica, que també va ser detectada en 18 individus de la cohort de controls.

10) L'anàlisi de CNVs en una família amb síndrome de Holt-Oram i glàndules mamàries addicionals va conduir a la detecció d'una duplicació heterozigota de la banda cromosòmica 12q24.21. La grandària màxima de la duplicació es va estimar aproximadament en 345,6 kb i s'esten a tota la regió codificadora dels gens *TBX5* i *TBX3*.

11) L'avaluació de dosis gènica en locus-específic mitjançant qPCR va posar de manifest la cosegregació de la duplicació i el fenotip de la síndrome de Holt-Oram / glàndules mamàries supernumeràries en aquesta família. Això constitueix un fort indicador de la patogenicitat de la duplicació. Fins ara, aquest és el primer informe d'una duplicació heterozigota que abasta tant el gen *TBX5* com el gen *TBX3*, i per tant és el primer informe d'un fenotip combinat de síndrome de Holt-Oram i glàndules mamàries supernumeràries.

8. REFERENCES

Agarwal P, Wylie JN, Galceran J, Arkhitko O, Li C, et al. (2003) "Tbx5 is essential for forelimb bud initiation following patterning of the limb field in the mouse embryo" *Development* 130(3):623-33

Agulnik SI, Bollag RJ, Silver LM (1995) "Conservation of the T-box gene family from *Mus musculus* to *Caenorhabditis elegan*" *Genomics* 25:214-219

Agulnik SI, Garvey N, Hancock S, Ruvinsky I, Chapman DL, et al. (1996) "Evolution of mouse T-box genes by tandem duplication and cluster dispersion" Genetics 144:249-54

Akrami SM, Winter RM, Brook JD, Armour JA (2001) "Detection of a large TBX5 deletion in a family with Holt-Oram syndrome" *J Med Genet* 38(12): E44

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

Alkan C, Coe BP, Eichler EE (2011) "Genome structural variation discovery and genotyping" *Nat* Rev Genet 12(5):363-76

Allen RC, Zoghbi HY, Moseley AB, Rosenblatt HM, Belmont JW (1992) "Methylation of HpaII and HhaI sites near the polymorphic CAG repeat in the human androgen-receptor gene correlates with X chromosome inactivation." *Am J Hum Genet* 51(6): 1229-39

American Association on Intellectual and Developmental Disabilities (AAIDD) "Mental retardation: Definition, classification, and systems of supports (11th ed.)" 2010 Washington, DC.

American Psychiatric Association (APA). "Diagnostic and statistical manual of mental disorders, 4th edition, Text Revision (DSM-IV-TR)" 2000 Washington, DC

Ames D, Murphy N, Helentjaris T, Sun N, Chandler V (2008) "Comparative analyses of human single- and multilocus tandem repeats" *Genetics* 179(3):1693-704

Amir RE, van den Veyver I, Wan M, Tran CQ, Francke U, et al. (1999) "Rett syndrome is caused bymutations in X-linked MECP2, encoding methyl-CpG-binding protein 2." *Nat Genet* 2:185–8

Antonarakis SE, Waber PG, Kittur SD, Patel AS, Kazazian HHJr, et al, (1985) "Hemophilia A. Detection of molecular defects and of carriers by DNA analysis" *N Engl J Med* 313(14):842-8

Arking DE, Cutler DJ, Brune CW, Teslovich TM, West K, et al. (2008) "A common genetic variant in the neurexin superfamily member CNTNAP2 increases familial risk of autism" *Am J Hum Genet* 82(1):160-4

Aynsley-Green A (1981) “Nesidioblastosis of the pancreas in infancy” *Dev Med Child Neurol* 23:372-379

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

Balikova I, Lehesjoki AE, de Ravel TJ, Thientpont B, Chandler KE, et al. (2009) “Deletions in the VPS13B (COH1) gene as a cause of Cohen syndrome” *Hum Mutat* 30(9):e845-54

Bamshad M, Krakowiak PA, Watkins WS, Root S, Carey JC, et al. (1995) “A gene for ulnar-mammary syndrome maps to 12q23-q24.1” *Hum Mol Genet* 4:1973-1977

Bamshad M, Root S, Carey JC (1996) “Clinical analysis of large kindred with the Pallister ulnar-mammary syndrome” *Am J Med Genet* 65(4): 325-31

Bamshad M, Lin R C, Law DJ, Watkins WC, Krakowiak PA, et al. (1997) “Mutations in human TBX3 alter limb, apocrine and genital development in ulnar-mammary síndrome” *Nat Genet* 16(3):311-5

Bamshad, M, Le T, Watkins WS, Dixon ME, Kramer BE, et al. (1999) “The spectrum of mutations in TBX3: Genotype/Phenotype relationship in ulnar-mammary syndrome.” *Am J Hum Genet* 64(6):1550-62

Basson, CT, Cowley GS, Solomon SD, Weissman B, Poznanski AK, et al. (1994) “The clinical and genetic spectrum of the Holt-Oram syndrome (heart-hand syndrome)” *N Engl J Med* 330(13):885-91

Basson CT, Bachinsky DR, Lin RC, Levi T, Elkins JA, et al. (1997) “Mutations in human TBX5 cause limb and cardiac malformation in Holt-Oram syndrome.” *Nat Genet* 15:30-5

Basson CT, Huang T, Lin RC, Bachinsky DR, Weremowicz S, et al. (1999) “Different TBX5 interactions in heart and limb defined by Holt-Oram syndrome mutations.” *Proc Natl Acad Sci USA* 96(6): 2919-2924

Bayés M, Magano LF, Rivera N, Flores R, Pérez Jurado LA (2003) “Mutational mechanisms of Williams - Beuren syndrome deletions” *Am J Hum Genet* 73:131-151

Bentley DR, Balasubramanian S, Swerdlow HP, Smith GP, Milton J, et al. (2008) “Accurate whole human genome sequencing using reversible terminator chemistry.” *Nature* 456:53–9

Bogdanova N, Markoff A, Pollmann H, Nowak-Göttl U, Eisert R, et al (2002) “Prevalence of small rearrangements in the factor VIII gene F8C among patients with severe hemophilia A” *Hum Mutat* 20(3):236-7

Bollag RJ, Siegfried Z, Cebra-Thomas JA, Garvey N, Davison EM, et al. (1994) "An ancient family of embryonically expressed mouse genes sharing a conserved protein motif with the T locus." *Nat Genet* 7**:**383–389

Bongers EM, Duijf PH, van Beersum SE, Schoots J, Van Kampen A, et al. (2004) "Mutations in the human TBX4 gene cause small patella syndrome." *Am J Hum Genet* 74:1239-48

Bonnet D, Pelet A, Legeai-Mallet L, Sidi D, Mathieu M, et al. (1994) "A gene for Holt-Oram syndrome maps to the distal long arm of chromosome 12." *Nat Genet* 6(4):405-8

Boogerd CJ, Dooijes D, Ilgun A, Mathijssen IB, Hordijk R, et al. (2010) "Functional analysis of novel TBX5 T-box mutations associated with Holt-Oram syndrome." *Cardiovasc Res* 88(1):130-9

Borozdin W, Bravo-Ferrer Acosta AM, Bamshad MJ, Botzenhart EM, Froster UG, et al. (2006a) "Expanding the spectrum of TBX5 mutations in Holt-Oram syndrome: detection of two intragenic deletions by quantitative real time PCR, and report of eight novel point mutations" *Hum Mutat* 27(9):975-6

Borozdin W, Bravo-Ferrer Acosta AM, Seemanova E, Leipoldt M, et al. (2006b) "Contiguous hemizygous deletion of TBX5, TBX3, and RBM19 resulting in a combined phenotype of Holt-Oram and ulnar-mammary syndromes" *Am J Med Genet A* 140A(17):1880-6

Brassington AM, Sung SS, Toydemir RM, Le T, Roeder AD, et al. (2003) "Expressivity of Holt-Oram syndrome is not predicted by TBX5 genotype" *Am J Hum Genet* 73(1):74-85

Braybrook C, Doudney K, Marçano AC, Arnason A, Bjornsson A, (2001) "The T-box transcription factor gene TBX22 is mutated in X-linked cleft palate and ankyloglossia." *Nat Genet* 29(2):179-83

Brookes AJ (1999) "The essence of SNPs" *Gene* 234:177–86

Brunetti-Pierri N, Berg JS, Scaglia F, Belmont J, Bacino CA, et al. (2008) "Recurrent reciprocal 1q21.1 deletions and duplications associated with microcephaly or macrocephaly and developmental and behavioral abnormalities." *Nat Genet* 40;1466-71

Cappellacci S, Martinelli S, Rinaldi R, Martinelli E, Parisi P, et al. (2006) "De novo pure 12q22q24.33 duplication: first report of a case with mental retardation, ADHD, and Dandy-Walker malformation" *Am J Med Genet A* 140(11):1203-7

Carlson H, Ota S, Campbell CE, Hurlin PJ, et al. (2001) "A dominant repression domain in Tbx3 mediates transcriptional repression and cell immortalization: relevance to mutations in Tbx3 that cause ulnar-mammary syndrome." *Hum Mol Genet* 10(21):2403-13

Carlsson P, MAhlapuu (2002) "Forkhead transcription factos: key players in development and metabolism" *Dev Biol* 250:1-23

Carr CW, Moreno-De-Luca D, Parker C, Zimmerman HH, Ledbetter N, et al. (2010) "Chiari I malformation, delayed gross motor skills, severe speech delay, and epileptiform discharges in a child with FOXP1 haploinsufficiency" Eur J Hum Genet. 18(11):1216-20

Carter NP (2007) "Methods and strategies for analyzing copy number variation using DNA microarrays" *Nat Genet* 39(7 Suppl):16-21. Review

Carvalho B, Ouwerkerk E, Meijer GA, Ylstra B (2004) "High resolution microarray comparative genomic hybridisation analysis using spotted oligonucleotides." *J Clin Pathol* 57(6): 644-6

Cerbai E, Sartiani L (2008) "Holt-oram syndrome and atrial fibrillation: opening the (T)-box" *Circ Res* 102(11):1304-6

Chandler KE, Kidd A, Al-Gazali L Kolehmainen J, Lehesjoki AE, et al. (2003) "Diagnostic criteria, clinical characteristics, and natural history of Cohen syndrome." *J Med Genet* 40(4): 233-41

Chapman DL, Garvey N, Hancock S, Alexiou M, Agulnik SI, et al. (1996) "Expression of the T-box family genes, Tbx1-Tbx5, during early mouse development." *Dev Dynam* 206:379-390

Chen KS, Manian P, Koeuth T, Potocki L, Zhao Q, et al. (1997) "Homologous recombination of a flanking repeat gene cluster is a mechanism for a common contiguous gene deletion syndrome." Nat Genet 17(2):154-63

Chen QR, Bilke S, Khan J (2005) "High-resolution cDNA microarray-based comparative genomic hybridization analysis in neuroblastoma." *Cancer Lett* 2005 228(1-2):71-81

Chueh AC, Northrop EL, Brettingham-Moore KH, Choo KH, Wong LH (Jan 2009). "LINE retrotransposon RNA is an essential structural and functional epigenetic component of a core neocentromeric chromatin" *PLoS Genetics* 5 (1):e1000354

Clayton PT, Eaton S, Aynsley-Green A, Edginton M, Hussain K, et al. (2001) "Hyperinsulinism in short-chain L-3-hydroxyacyl-CoA dehydrogenase deficiency reveals the importance of beta-oxidation in insulin secretion" *J Clin Invest* 108(3):457-65

Colella S, Yau C, Taylor JM, Mirza G, Butler H, et al. (2007) "QuantiSNP: an Objective Bayes Hidden-Markov Model to detect and accurately map copy number variation using SNP genotyping data." *Nucleic Acids Res* 35:2013-2025

Coleman RA, Smith WL, Narumiya S (1994) "VIII. International union of pharmacology classification of prostanoid receptors: properties, distribution, and structure of the receptors and their subtypes." *Pharm Rev* 46: 205-229

Coll M, Seidman JG, Müller CW (2002) "Structure of the DNA-bound T-box domain of human TBX3, a transcription factor responsible for ulnar-mammary syndrome" *Structure* 10(3):343-56

Conlin KL, Thiel BD, Bonnemann CG, Medne L, Ernst L, et al. (2010) " Mechanisms of mosaicism, chimeris, and uniparental disomy identified by SNP array analsyis." *Hum Mol Genet* 19(7):1263-75

Conlon FL, Fairclough L, Price BM, Casey ES, Smith JC (2001) "Determinants of T box protein specificity" *Development*, 128:3749- 3758

Conrad DF, Andrews TD, Carter NP, Hurles ME, Pritchard JK (2006) A high-resolution survey of deletion polymorphism in the human genome. *Nat Genet* 38(1):75-81

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

Cook EH, Scherer SW (2008) "Copy-number variations associated with neuropsychiatric conditions." *Nature* 16:455(7215)

Cooper GM, Zerr T, Kidd JM, Eichler EE, Nickerson DA (2008) "Systematic assessment of copy number variant detection via genome-wide SNP genotyping" *Nat Genet* 40(10):1199-203

Cremer T, Lichter P, Borden J, Ward DC, Manuelidis L (1988) "Detection of chromosome aberrations in metaphase and interphase tumor cells by in situ hybridization using chromosome-specific library probes." *Hum Genet* 80(3): 235-46

Cross SJ, Ching YH, Li QY, Armstrong-Buisseret L, Spranger S, et al. (2000) "The mutation spectrum in Holt-Oram syndrome." *J Med Genet* 37(10):785-7

Cuscó I, Corominas R, Bayés M, Flores R, Rivera-Burgués N, et al. (2008) "Copy number variation at the 7q11.23 segmental duplications is a susceptibility factor for the Williams-Beuren syndrome deletion." *Genome Res* 18(5):683-94

Darai-Ramqvist E, Sandlund A, Müller S, Klein G, Imreh S, et al. (2008) "Segmental duplications and evolutionary plasticity at tumor chromosome break-prone regions" *Genome Res* 18**:**370-379

Davenport TG, Jerome-Majewska LA, Papaioannou VE (2003) "Mammary gland, limb and yolk sac defects in mice lacking Tbx3, the gene mutated in human ulnar mammary syndrome." *Development* 130(10):2263-73

David D, Cardoso J, Marques B, Marques R, Silva ED, et al. (2003) "Molecular characterization of a familial translocation implicates disruption of HDAC9 and possible position effect on TGFbeta2 in the pathogenesis of Peters' anomaly." *Genomics* 81(5):489-503

De Kovel CG, Trucks H, Helbig I, Mefford HC, Baker C, et al. (2010) "Recurrent microdeletions at 15q11.2 and 16p13.11 predispose to idiopathic generalized epilepsies". *Brain* 133(Pt 1):23-32

Dellinger AE, Saw SM, Goh LK, Seielstad M, Young TL, et al. (2010) "Comparative analyses of seven algorithms for copy number variant identification from single nucleotide polymorphism arrays." *Nucleic Acids Research* 38:e105

De Lonlay P, Fournet JC, Rahier J, Gross-Morand MS, Poggi-Travert F, et al. (1997) "Somatic deletion of the imprinted 11p15 region in sporadic persistent hyperinsulinemic hypoglycemia of infancy is specific of focal adenomatous hyperplasia and endorses partial pancreatectomy" *J Clin Invest* 100(4):802-7

De Vries BB, Pfundt R, Leisink M, Koolen DA, Vissers LE, et al. (2005) "Diagnostic genome profiling in mental retardation" *Am J Hum Genet* 77(4): 606-616

De Vries BB, Winter R, Schinzel A, van Ravenswaaij-Arts C (2003) "Telomeres: a diagnosis at the end of the chromosomes" *J Med Genet* 40:385-98

Dhami P, Coffey AJ, Abbs S, Vermeesch JR, Dumanski JP, et al. (2005) "Exon array CGH: detection of copy-number changes at the resolution of individual exons in the human genome." *Am J Hum Genet* 76:750–762

Dixon JW, Costa T, Teshima IE. (1993) "Mosaicism for duplication 12q (12q13!q24.2) in a dysmorphic male infant." *J Med Genet* 30:70–72

Dobrovolskaia-Zavadskaia N (1927) "Sur la mortification spontanée de la queue che la souris nouveau-née et sur l'existence d'un caractère (facteur) héréditaire ''non viable''. " *C R Seanc Soc Biol* 97**:**114–116

Durkin M (2002) "The epidemiology of developmental disabilities in low-income countries" Ment *Retard Dev Disabil Res Rev* 8:206-11

Eilers PH, de Menezes RX (2005) "Quantile smoothing of array CGH data" *Bioinformatics* 21:1146-1153

El Chehadeh S, Aral B, Gigot N, Thauvin-Robinet C, Donzel A, et al. (2010) "Search for the best indicators for the presence of a VPS13B gene mutation and confirmation of diagnostic criteria in a series of 34 patients genotyped for suspected Cohen syndrome" *J Med Genet* 47(8):549-53

Elek C, Vitez M, Czeizel E (1991) " Holt-Oram syndrome " *Orv Hetil* 132(2):73-74, 77-78

Engel E (2006) "Afascinationwith chromosome rescue in uniparental disomy: Mendelian recessive outlaws and imprinting copyrights infringements" *Eur J Hum Genet* 14:1158–69

Ermakova O, Piszczek L, Luciani L, Cavalli FM, Ferreira T, et al. (2011) "Sensitized phenotypic screening identifies gene dosage sensitive region on chromosome 11 that predispose to disease in mice" *EMBO Mol Med* 3(1):50-66

Fanciulli M, Norsworthy PJ, Petretto, et al. (2007) "FCGR3B copy number variation is associated with

susceptibility to systemic, but not organ-specific, autoimmunity" Nat Genet 39:721-23

Fellermann K, Stange DE, Schaeffeler E, et al. (2006) "A chromosome 8 gene-cluster polymorphism with low human beta-defensin 2 gene copy number predisposes to Crohn disease of the colon" Am J Hum Genet 79:439-48

Fernandez BA, Roberts W, Chung B, Weksberg R, Meyn S, et al. (2009) "Phenotypic spectrum associated with de novo and inherited deletions and duplications at 16p11.2 in individuals ascertained for diagnosis of Autism spectrum disorders." *J Med Genet* 47(3): 195-203

Feuk L, Carson AR, Scherer SW (2006) "Structural variation in the human genome" Nat Rev Genet 7:85-97

Feuk L, Kalervo A, Lipsanen-Nyman M, Skaug J, Nakabayashi K, et al. (2006) "Absence of a paternally inherited FOXP2 gene in developmental verbal dyspraxia"*Am J Hum Genet* 79(5):965-72.

Flanagan SE, Kapoor RR, Hussain K (2011) "Genetics of congenital hyperinsulinemic hypoglycemia" *Semin Pediatr Surg* 20(1):13-17

Flint J, Hill AV, Bowden DK, Oppenheimer SJ, Sill PR, et al. (1986) "High frequencies of alpha-thalassaemia are the result of natural selection by malaria" Nature 321: 744-50

Flint J, Wilkie AOM, Buckle VJ, Winter RB, Holland AJ, et al. (1995) "The detection of subtelomeric chromosomal rearrangements in idiopathic mental retardation." *Nature Genetics* 9: 132-139

Flint J, Knight S (2003) "The use of telomere probes to investigate submicroscopic rearrangements associated with mental retardation." *Curr Opin Genet Dev* 13(3):310-6

Fox ER, Young JH, Li Y, Dreisbach AW, Keating BJ, et al. (2011) "Association of genetic variation with systolic and diastolic blood pressure among African Americans: the Candidate Gene Association Resource study" *Hum Mol Genet* 20(11):2273-84

Franceschini P, Vardeu MP, Dalforno L, Signorile F, Franceschini D, et al. (1992) "Possible relationship between ulnar-mammary syndrome and split hand with aplasia of the ulna syndrome" *Am J Med Genet* 44(6):807-12

Franek KJ, Butler J, Johnson J, Simensen R, Friez MJ, et al. (2011) "Deletion of the immunoglobulin domain of IL1RAPL1 results in nonsyndromic X-linked intellectual disability associated with behavioral problems and mild dysmorphism." *Am J Med Genet* A155A(5):1109-14

Franklin TB, Mansuy IM (2011) "The involvement of epigenetic defects in mental retardation" Neurobiol Learn Mem 96(1):61-7

Fridlyand J, Snijders AM, Pinkel D, Albertson DG, Jain A (2004) "Hidden Markov models approach to the analysis of array CGH data." *J Multivariate Anal* 90:132-153

Friedman JM, Baross A, Delaney AD, et al. (2006) “Oligonucleotide microarray analysis of genomic imbalance in children with mental retardation” *Am J Hum Genet* 79(3): 500-13

Gao X, Xi G, Niu Y, Zhang S, Fu R, et al. (2008) ”A study on the correlation between IL1RAPL1 and human cognitive ability.” *Neurosci Lett* 438:163–167

Garavelli L, De Brasi D, Verri R, Guareschi E, Cariola F, et al. (2008) “Holt-Oram syndrome associated with anomalies of the feet” *Am J Med Genet A* 146A(9):1185-9

Garg V, Kathiriya IS, Barnes R, Schluterman MK, King IN, et al. (2003) “GATA4 mutations cause human congenital heart defects and reveal an interaction with TBX5” *Nature* 424(6947):443-7

Gijsbers AC, Lew JY, Bosch CA, Schuurs-Hoeijmakers JH, van Haeringen A, et al. (2009) “A new diagnostic workflow for patients with mental retardation and/or multiple congenital abnormalities: test arrays first.” *Eur J Hum Genet* 17(11): 1394-402

Girirajan S, Roselfeld J, Cooper GM, Antonacci F, Siswara P, et al. (2010) “A recurrent 16p12.1 microdeletion supports a two-hit model for severe developmental delay.” *Nat Genet* 42 (3); 203-9

Glaser B, Kesavan P, Heyman M, Davis E, Cuesta A, et al. (1998) “Familial hyperinsulinism caused by an activating glucokinase mutation” *N Engl J Med* 338(4):226-30

Glaser B, Thornton PS, Otonkoski T, Junien C (2000) “The genetics of neonatal hyperinsulinism” *Arch Dis Child* 82:79-86

Gong W, Gottlieb S, Collins J, Blescia A, Dietz H, et al. (2001) “Mutation analysis of TBX1 in non-deleted patients with features of DGS/VCFS orisolated cardiovascular defects” *J Med Genet* 38:E45

Gonzalez CH, Herrmann J, Opitz, JM (1976) “Studies of malformation syndromes of man XXXXIIB: mother and son affected with the ulnar-mammary syndrome type Pallister.” *Eur J Pediatr* 123(4):225-35

Gonzalez E, Kulkarni H, Bolivar H, et al. (2005) “The influence of CCL3L1 gene-containing segmental duplications on HIV-1/AIDS susceptibility” *Science* 307: 1434-40

González-Barroso MM, Giurgea I, Bouillaud F, et al. (2008) “Mutations in UCP2 in congenital hyperinsulinism reveal a role or regulation of insulin secretion” *PLoS ONE* 3:e3850

Ghosh TK, Packham EA, Bonser AJ, Robinson TE, Cross SJ, et al. (2001) “Characterization of the TBX5 binding site and analysis of mutations that cause Holt-Oram syndrome” *Hum Mol Genet* 10:1983-1994

Guilmatre A, Dubourg C, Mosca AL, Legallic S, Goldenberg A, et al. (2009) “Recurrent rearrangements in synaptic and neurodevelopmentlal genes and shared biologic pathways in schizophrenia, autism,

and mental retardation" *Arch Gen Psychiatry* 66(9): 947-956

Hamdan FF, Daoud H, Rochefort D, Piton A, Gauthier J, et al. (2010) "De novo mutations in FOXP1 in cases with intellectual disability, autism, and language impairment". *Am J Hum Genet* 87(5):671-8

Hamdan FF, Gauthier J, Araki Y, Lin DT, Yoshizawa Y, et al. (2011) "Excess of de novo deletierious mutations in genes associated with Glutamatergic systems in nonsyndromic intellectual disability" *Am J Hum Genet* 88: 306-16

Hannes FD, Sharp AJ, Mefford HC, de Ravel T, Ruivenkamp CA et al. (2009) "Recurrent reciprocal deletions and duplications of 16p13.11: the deletion is a risk factor for MR/MCA while the duplication may be a rare benign variant." *J Med Genet* 46(4):223-32

Harrod MJE, Byrne JB, Dev VG, Francke U (1980) "Duplication 12q mosaicism in two unrelated patients with a similar syndrome." *Am J Med Genet* 7:123–129

Hastings PJ, Lupski JR, Rosenberg SM, Ira G (2009) "Mechanisms of change in gene copy number" *Nat Rev Genet* 10(8): 551–564

He M, Wen L, Campbell CE, Wu JY, Rao Y (1999) "Transcription repression by Xenopus ET and its human ortholog TBX3, a gene involved in ulnar-mammary syndrome." *Proc Natl Acad Sci USA* 96(18):10212-7

Hehir-Kwa JY, Wieskamp N, Webber C, Pfundt R, Brunner HG, et al. (2010) "Accurate distinction of pathogenic from benign CNVs in mental retardation" *PLoS Comput Biol* 22;6(4): e1000752

Heinritz W, Shou L, Moschik A, Froster UG (2005) "The human TBX5 gene mutation database" *Hum Mutat* 26:397

Hennies HC, Rauch A, Seifert W, et al. (2004) "Allelic heterogeneity in the COH1 gene explains clinical variability in Cohen syndrome." *Am J Hum Genet* 75(1):138-45

Herrmann BG, Labiet S, Poustka A, King T, Lehrach H (1990) "Cloning of the T gene required in mesoderm formation in the mouse." *Nature* 343**:**617–622

Higgs DR, Pressley L, Old JM, Hunt DM, Clegg JB, et al. (1979) "Negro alpha-thalassemia is caused by deletion of a single alpha-globin gene" Lancet 2:272-76

Hinds DA, Kloek AP, Jen M, Chen X, Frazer KA (2006) "Common deletions and SNPs are in linkage disequilibrium in the human genome." *Nat Genet*. 38(1):82-5

Hiroi Y, Kudoh S, Monzen K, Ikeda Y, Yazaki Y, et al. (2001) "Tbx5 associates with Nkx2-5 and synergistically promotes cardiomyocyte differentiation" *Nat Genet* 28:276-280

Hirschhorn JN, Gajdos ZK (2011) "Genome-wide association studies; Results from the first few years and

potential implications for clinical medicine" Annu Rev Med 62:11-24

Hochstenbach R, van Binsbergen E, Engelen J, Nieuwint A, Polstra A, et al. (2009) "Array analysis and karyotyping: Workflow consequences based on a retrospective study of 36,325 patients with idiopathic developmental delay in the Netherlands." *Eur J Med Genet* 52(4):161-9

Hollox EJ, Huffmeier U, Zeeuwen PL, et al. (2008) "Psoriasis is associated with increased beta-defensin genomic copy number" Nat Genet 40:23-25

Holm H, Gudbjartsson DF, Arnar DO, Thorleifsson G, Thorgeirsson G, et al. (2010) "Several common variants modulate heart rate, PR interval and QRS duration." *Nat Genet* 42(2):117-22

Holt M and Oram S (1960) "Familial heart disease with skeletal malformations" *Br Heart J* 22:236-42.

Horn D, Kapeller J, Rivera-Brugués N, Moog U, Lorenz-Depiereux B, et al. (2010) "Identification of FOXP1 deletions in three unrelated patients with mental retardation and significant speech and language deficits." *Hum Mutat* 31(11):E1851-60

Horn D, Krebsova A, Kunze J*, et al.* Homozygosity mapping in a family with microcephaly, mental retardation, and short stature to a Cohen syndrome region on 8q21.3 8q22.1: redefining a clinical entity. *Am J Med Genet* 2000;92(4):285-92

Hoyer J, Dreweke A, Becker C*,* et al. (2007) "Molecular karyotyping in patients with mental retardation using 100K single-nucleotide polymorphism arrays." *J Med Genet* 44(10): 629-36

Hsu L, Self SG, Grove D, Randolph T, Wang K, et al. (2005) "Denoising array-based comparative genomic hybridization data using wavelets." *Biostatistics* 6:211-226

Hupé P, Stransky N, Thiery JP, Radvanyi F, Barillot E (2004) "Analysis of array CGH data: from signal ratio to gain and loss of DNA regions." *Bioinformatics* 20, 3413–22

Hussain K (2007) "Insights in congenital hyperinsulinism" *Endocr Dev* 11:106-21. Review

Iafrate AJ, Feuk L, Rivera MN, Listewnik ML, Donahoe PK, et al. (2004) "Detection of large-scale variation in the human genome." *Nat Genet* 36(9):949-51

Ieshima A, Yorita T, Ohta S, Kuroki Y (1984) "A female infant with pure duplication 12q24.2 to qter." *Jpn J HumGenet* 29:391–397

International HapMap Project (2003) "The International HapMap Project" *Nature* 426(6968):789-96

Ishkanian AS, Malloff CA, Watson SK, DeLeeuw RJ, Chi B, et al. (2004) "A tiling resolution DNA microarray with complete coverage of the human genome." *Nat Genet* 36:299–303

Jackman C, Horn ND, Molleston JP, Sokol DK (2009) “Gene associated with seizures, autism, and hepatomegaly in an Amish girl” *Pediatr Neurol* 40(4):310-3.

Jacobs PA, Matsuura JS, Mayer M, Newlands IM (1978) “A cytogenetic survey of an institution for the mentally retarded; I. Chromosome abnormalities” *Clin Genet* 13:37-60

James C, Kapoor RR, Ismail D, Hussain K (2009) “The genetic basis of congenital hyperinsulinism” *J Med Genet* 46(5):289-99. Review

Jin H, Gardner RJ, Viswesvaraiah R, Muntoni F, Roberts RG (2000) “Two novel members of the interleukin-1 receptor gene family, one deleted in Xp22.1-Xp21.3 mental retardation.” *Eur J Hum Genet* 8(2):87-94

Johnson CP, Walker JWO, Palomo-Gonzalez SA, Curry CJ (2006) ”Mental retardation: diagnosis, management, and family support.” *Curr Probl Pediatr Adolesc Health Care* 36:126-165

Jong K, Marchiori E, Meijer G, Vaart AV, Ylstra B (2004) “Breakpoint identification and smoothing of array comparative genomic hybridization data.” *Bioinformatics* 20(18):3636-37

Joss S, Kini U, Fisher R, Mundlos S, Prescott K, et al. (2011) “The face of Ulnar Mammary syndrome?” *Eur J Med Genet* 54(3):301-5

Kato N, Takeuchi F, Tabara Y, Kelly TN, Go MJ, et al. (2011) “Meta-analysis of genome-wide association studies identifies common variants associated with blood pressure variation in east Asians” *Nat Genet* 43(6):531-8

Kidd JM, Cooper GM, Donahue WF, Hayden HS, Sampas N, et al. (2008) “Mapping and sequencing of structural variation from eight human genomes” *Nature* 453(7191):56-64

Kim MS, Hatcher CJ, Wong B, Goldstein MM, Mikawa T, et al. (2000) “Tbx5 transcription factor: a cellular arrest signal during vertebrate cardiogenesis.” *Circulation* 102:100-110

Kim S, Misra A (2007) “SNP Genotyping: Technologies and Biomedical Applications” *Annu Rev Biomed Eng* 9:289-320

Kirk EP, Sunde M, Costa MW, Rankin SA, Wolstein O, et al. (2007) “Mutations in cardiac T-box factor gene TBX20 are associated with diverse cardiac pathologies, including defects of septation and valvulogenesis and cardiomyopathy” *Am J Hum Genet* 81(2):280-91

Kishino T, Lalande M, Wagstaff J (1997) “UBE3A/E6-AP mutations cause Angelman syndrome“ *Nat Genet* 15(1):70-73

Kispert A, Herrmann BG (1993) “The Brachyury gene encodes a novel DNA binding protein.“ *EMBO J* 12(8):3211-20

Klopocki E, Neumann LM, Tönnies H, Ropers HH, Mundlos S, **et al.** (2006) "Ulnar-mammary syndrome with dysmorphic facies and mental retardation caused by a novel 1.28 Mb deletion encompassing the TBX3 gene" *Eur J Hum Genet* 14(12):1274-9

Knight SJ, Regan R, Nicod A, Horsley SW, Kearney L, et al. (1999) "Subtle chromosomal rearrangements in children with unexplained mental retardation.." *Lancet* 354(9191): 1676-81

Kolehmainen J, Wilkinson R, Lehesjoki AE, Chandler K, Kivitie-Kallio S, et al. (2004) "Delineation of Cohen syndrome following a large-scale genotype-phenotype screen" *Am J Hum Genet* 75(1):122-7

Koolen DA, Vissers LE, Pfundt R, de Leeuw N, Knight SJ, et al. (2006) "A new chromosome 17q21.31 microdeletion syndrome associated with a common inversion polymorphism." *Nat Genet* 38(9):999-100

Koolen DA, Pfundt R, de Leeuw N, Hehir-Kwa JY, Nillesen WM, **et al.** (2009) "Genomic microarrays in mental retardation: a practical workflow for diagnostic applications." *Hum Mutat* 30(3):283-92

Korbel JO, Urban AE, Affourtit JP, Godwin B, Grubert F, **et al.** (2007) "Paired-end mapping reveals extensive structural variation in the human genome."*Science* 318(5849):420-6

Korn JM, Kuruvilla FG, McCarroll SA, Wysoker A, Nemesh J, et al. (2008) "Integrated genotype calling and association analysis of SNPs, common copy number polymorphisms and rare CNVs." *Nat Genet* 40:1253-1260

Koshiba-Takeuchi K, Takeuchi JK, Arruda EP, Kathiriya IS, Mo R, et al. (2006) "Cooperative and antagonistic interactions between Sall4 and Tbx5 pattern the mouse limb and heart" *Nat Genet* 38(2):175-83

Kotani M, Tanaka I, Ogawa Y, Usui T, Tamura N, et al. (1997) "Structural organization of the human prostaglandin EP3 receptor subtype gene (PTGER3)." *Genomics* 40(3): 425-34

Kuang SQ, Guo DC, Prakash SK, McDonald ML, Johnson RJ, et al. (2011) "Recurrent chromosome 16p13.1 duplications are a risk factor for aortic dissections" PLoS Genet 7(6):e1002118

Kuhn EM, Sarto GE, Bates BJ, Therman E (1987) "Gene-rich chromosome regions and autosomal trisomy. A case of chromosome 3 trisomy mosaicism." *Hum Genet* 77(3):214-20

Kurotaki N, Harada N, Shimokawa O, Miyake N, Kawame H, et al**.** (2003) " Fifty microdeletions among 112 cases of Sotos syndrome: low copy repeats possibly mediate the common deletion." *Hum Mutat* 22(5):378-87

Kramer JM, van Bokhoven H (2009) "Genetic and epigenetic defects in mental retardation" Int J Biochem Cell Biol 41(1):96-107

Lai CS, Fisher SE, Hurst JA, Vargha-Khadem F, Monaco AP (2001) "A forkhead-domain gene is mutated in a severe speech and language disorder" *Nature*. 2001 Oct 4;413(6855):519-23

Lai WR, Johnson MD, Kucherlapati, Park J (2005) "Comparative analysis of algorithms for identifying amplifications and deletions in array CGH data" *Bioinformatics* 21(19):3763-70

Lakich D, Kazazian HH Jr, Antonarakis SE, Gitschier J (1993) "Inversions disrupting the factor VIII gene are a common cause of severe haemophilia A." Nat Genet 5(3): 236-41

Lamolet B, Pulichino AM, Lamonerie T, Gauthier Y, Brue T, et al. (2001) "A pituitary cell-restricted T box factor, Tpit, activates POMC transcription in cooperation with Pitx homeoproteins." *Cell* 104(6):849-59

Lander ES Linton LM, Birren B; International Human Genome Sequencing Consortium (2001) "Initial sequencing and analysis of the human genome." *Nature* 409(6822): 860-921

Lander ES et al; International Human Genome Sequencing Consortium (2004) "Finishing the euchromatic sequence of the human genome" *Nature* 431(7011): 931-45

Lausch E, Hermanns P, Farin HF, Alanay Y, Unger S, et al. (2008) "TBX15 mutations cause craniofacial dysmorphism, hypoplasia of scapula and pelvis, and short stature in Cousin syndrome" *Am J Hum Genet*. 83(5): 649–655

Lee JA, Carvalho CM, Lupski JR (2007). "A DNA replication mechanism for generating nonrecurrent rearrangements associated with genomic disorders" *Cell* 131 (7): 1235–47

Lehner R, Goharkhay N, Tringler B, Fasching C, Hengstschläger M (2003) "Pedigree analysis and descriptive investigation of three classic phenotypes associated with Holt-Oram síndrome" J Reprod Med. 48(3):153-9. Review

Lejeune J, Gautier M, Turpin R (1959) "Etude des chromosomes comatiques de neuf enfants mongoliens" *C.R. Acad Sci* 248:1721-22

Leonard H, Wen X (2002) "The epidemiology of mental retardation; challenges and opportunities in the new millenium" *Ment Retard Dev Disabil Res Rev* 8:117-34

Levy D, Ehret GB, Rice K, Verwoert GC, Launer LJ, et al. **(**2009) "Genome-wide association study of blood pressure and Hipertensión**"** *Nat Genet* 41(6):677-87

Levy S, Sutton G, Ng PC, Feuk L, Halpern AL, et al. (2007) "The diploid genome sequence of an individual human." *PLoS Biol* 5:e254

Li QY, Newbury-Ecob RA, Terrett JA, Wilson DI, Curtis AR, et al. (1997) "Holt-Oram syndrome is caused by mutations in TBX5, a member of the Brachyury (T) gene family" *Nat Genet* 15(1): 21-9

Li S, Weidenfeld J, Morrisey EE. (2004) "Transcriptional and DNA binding activity of the Foxp1/2/4 family is modulated by heterotypic and homotypic protein interactions" *Mol Cell Biol* 24(2):809-22

Liberatore CM, Searcy-Schrick RD, Yutzey KE (2000) "Ventricular expression of tbx5 inhibits normal heart chamber development" *Dev Biol* 223(1):169-80

Lieber MR (2008) "The mechanism of human nonhomologous DNA end joining" *J Biol Chem* 283 (1):1-5

Liang F, Holt I, Pertea G, Karamycheva S, Salzberg SL, et al. (2000) "Gene index analysis of the human genome estimates approximately 120,000 genes" *Nat Genet* 25(2): 239-40

Lupski JR, de Oca-Luna RM, Slaugenhaupt S, Pentao L, Guzzetta V, et al. (1991) "DNA duplication associated with Charcot-Marie-Tooth disease type 1A" Cell 66:219-32

MacDermot KD, Bonora E, Sykes N, Coupe AM, Lai CS, et al. (2005) "Identification of FOXP2 truncation as a novel cause of developmental speech and language deficits" *Am J Hum Genet* 76(6):1074-80

Mandel JL, Chelly (2004) J "Monogenic X-linked mental retardation: Is it as frequent as currently estimated? The paradox of the ARX (Aristaless X) mutations" *Eur J Hum Genet* 12:689–693

Marquard J, Palladino AA, Stanley CA, Mayatepek E, Meissner T (2011) "Rare forms of congenital hyperinsulinism" *Semin Pediatr Surg* 20(1):38-44. Review

Marshall, CR*, et al.* (2008) "Structural variation of chromosomes in autism spectrum disorder." *Am J Hum Genet* 82:477–488

McCarroll SA, Hadnott TN, Perry GH, Sabeti PC, Zody MC, et al. (2006) "Common deletion polymorphisms in the human genome" *Nat Gene*t 38(1):86-92

McCorquodale MM, Rolf J, Ruppert ES, Kurczynski TW, Kolacki P (1986) "Duplication (12q) syndrome in female cousins, resulting from maternal (11;12)(p15.5;q24.2) translocations. " *Am J Med Genet* 24:613-622

McDermott DA, Bressan MC, He J, Lee JS, Aftimos S, et al. (2005) "TBX5 genetic testing validates strict clinical criteria for Holt-Oram syndromes." *Pediatr Res* 58:981–986

McMullan DJ, Bonin M, Hehir-Kwa JY, de Vries BBA, Dufke A, et al. (2009) "Molecular karyotyping of patients with unexplained mental retardation by SNP arrays: A multicenter study." *Hum Mut* 30(7): 1082-92

Mefford HC, Cooper GM. Zerr T, Smith JD, Baker C, et al. (2009) A method for rapid, targeted CNV grnotyping identifies rare variants associated with neurocognitive disease. *Genome Res* 19(9);1579-85

Melnyk AR, Weiss L, Van Dyke DL, Jarvi P (1981) "Malformation syndrome of duplication 12q24.1-qter" *Am J Med Genet* 10:357-365

Meneghini V, Odent S, Platonova N, Egeo A, Merlo GR (2006) "Novel TBX3 mutation data in families with Ulnar-Mammary syndrome indicate a genotype-phenotype relationship: mutations that do not disrupt the T-domain are associated with less severe limb defects." *Eur J Med Genet* 49(2):151-8

Meng ZX, Sun JX, Ling JJ, Lv JH, Zhu DY, et al. (2006) "Prostaglandin E2 regulates Foxo activity via the Akt pathway: implications for pancreatic islet beta cell dysfunction." *Diabetologia* 49:2959-2968

Menten B, Buysse K, Zahir F, Hellemans J, Hamilton SJ, et al (2007) "Osteopoikilosis, short stature and mental retardation as key features of a new microdeletion syndrome on 12q14" *J Med Genet* 44:264-8

Miller DT, Adam MP, Aradhya S, Biesecker LG, Brothman AR, et al. (2010) "Consensus Statement: Chromosomal Microarray is a first-tier clinical diagnostic test for individuals with developmental disabilities or congenital anomalies." *Am J Hum Genet* 86: 749-764

Miller DT, Shen Y, Weiss LA, Korn J, Anselm I, et al. (2009) "Microdeletion/duplication at 15q13.2q13.3 among individuals with features of autism and other neuropsychiatric disorders." *J Med Genet* 46(4):242-8

Miller SA, Dykes DD, Polesky HF (1988). "A simple salting out procedure for extracting DNA from human nucleated cells." *Nucleic Acids Res* 16(3): 1215

Ming JE, Geiger E, James AC, Ciprero KL, Nimmakayalu M, et al. (2006) "Rapid detection of submicroscopic chromosomal rearrangements in children with multiple congenital anomalies using high density oligonucleotide arrays" *Hum Mut* 27(5):467-473

Miyake N, Shimokawa O, Harada N, Sosonkina N, Okubo A, et al. (2006) "BAC array CGH reveals genomic aberrations in idiopathic mental retardation." *Am J Med Genet* 140A:205-211

Mochida GH, Rajab A, Eyaid W*, et al.* (2004) "Broader geographical spectrum of Cohen syndrome due to COH1 mutations" *J Med Genet* 41(6):e87

Moore JK, Haber JE. (1996) "Cell cycle and genetic requirements of two pathways of nonhomologous end-joining repair of double-strand breaks in Saccharomyces cerevisiae" *Mol Cell Biol* 16(5):2164-73

Mori AD, Bruneau BG (2004) "TBX5 mutations and congenital heart disease: Holt-Oram syndrome revealed." *Curr Opin Cardiol* 19(3):211-5

Mumm S, Herrera L, Waeltz PW, Scardovi A, Nagaraja R, et al. (2001) "X/autosomal translocations in the Xq critical region associated with premature ovarian failure fall within and outside genes" *Genomics* 76 (1-3): 30-6

Muncke N, Jung C, Rudiger H, Ulmer H, Roeth R et al. (2003) "Missense mutations and gene interruption in PROSIT240, a novel TRAP240-like gene, in patients with congenital heart defect (transposition of the great arteries)" *Circulation* 108(23):2843-50

Nakabayashi K, Trujillo AM, Tayama C, Camprubi C, Yoshida W, et al. (2011) "Methylation screening of reciprocal genome-wide UPDs identifies novel human-specific imprinted genes" *Hum Mol Genet* 20(16):3188-97.

Nathans J, Piantanida TP, Hedí RL, Shows TB, Hogness DS (1986) "Molecular genetics of inherited variation in human color vision" *Science* 232: 203-10

Nawara M, Klapecki J, Borg K, Jurek M, Moreno S, et al. (2008) "Novel mutation of IL1RAPL1 gene in a nonspecific X-linked mental retardation (MRX) family." *Am J Med Genet Part A* 146A:3167–3172

Naylor J, Brinke A, Hassock S, Green PM, Giannelli F (1993) "Characteristic mRNA abnormality found in half the patients with severe haemophilia A is due to large DNA inversions." *Hum Mol Genet* 2(11):1773-8

Newbury-Ecob RA, Leanage R, Raeburn JA, Young ID (1996) "Holt-Oram syndrome: a clinical genetic study." *J Med Genet* 33(4):300-7

Newbury DF, Monaco AP.(2010) "Genetic advances in the study of speech and language disorders" *Neuron* 68(2):309-20. Review.

Ng SB, Bigham AW, Buckingham KJ, Hannibal MC, McMillin MJ, et al. (2010) "Exome sequencing identifies MLL2 mutations as a cause of Kabuki syndrome." *Nat Genet* 42: 790–793

Nickoloff JA, De Haro LP, Wray J, Hromas R (2008) "Mechanisms of leukemia translocations." *Curr Opin Hematol* 15:338–4

Oberle I, Rousseau F, Heitz D, Kretz C, Devys D, et al. (1991) "Instability of a 550-base pair DNA segment and abnormal methylation in fragile X syndrome." *Science* 252:1097–1102

O'Roak BJ, Deriziotis P, Lee C, Vives L, Schwartz JJ, et l. (2011) "Exome sequencing in sporadic autism spectrum disorders identifies severe de novo mutations" *Nat Genet* 43(6):585-9

Olshen AB, Venkatraman ES, Lucito R, Wigler M (2004) "Circular binary segmentation for the analysis of array-based DNA copy number data" *Biostatistics* 5(4):557-572

Osborne LR, Li M, Pober B, Chitayat D, Bodurtha J, et al. (2001) "A 15 million-base pair inversion polymorphism in families with Williams-Beuren syndrome." *Nature Genet* 29:321-325

Otonkoski T, Näntö-Salonen K, Seppänen M, Veijola R, Huopio H, et al. (2006) "Noninvasive diagnosis of focal hyperinsulinism of infancy with [18F]-DOPA positron emission tomography" *Diabetes*

55(1):13-8

Otonkoski T, Jiao H, Kaminen-Ahola N, Tapia-Paez I, Ullah MS, et al. (2007) "Physical exercise-induced hypoglycemia caused by failed silencing of monocarboxylate transporter 1 in pancreatic beta cells" *Am J Hum Genet*. 81(3):467-74

Ou Z, Berg JS, Yonath H, Enciso VB, Miller DT, et al.(2008) "Microduplications of 22q11.2 are frequently inherited and are associated with variable phenotypes" *Genet Med* 10:267-77

Packham EA, Brook JD (2003) "T-box genes in human disorders" *Hum Mol Genet* 12 Spec No 1:R37-44. Review

Pallister PD, Herrman, J. and Opitz JM (1976). Studies of malformation syndromes in man XXXXII: a pleiotropic dominant mutation affecting skeletal, sexual and apocrine-mammary development. *Birth Defects Orig Artic Ser* 12(5):247-54

Papaioannou VE, Silver LM (1998) "The T-box gene family" *Bioessays* 20(1): 9-19

Pariani MJ, Spencer A, Graham JM Jr, Rimoin DL (2009) "A 785kb deletion of 3p14.1p13, including the FOXP1 gene, associated with speech delay, contractures, hypertonia and blepharophimosis." *Eur J Med Genet*. 52(2-3):123-7

Parri V, Katzaki E, Uliana V, Scionti F, Tita R, et al. (2010) "High frequency of COH1 intragenic deletions and duplications detected by MLPA in patients with Cohen syndrome" *Eur J Hum Genet* 18(10):1133-40

Pastinen T, Kurg A, Metspalu A, Peltonen L, Syvänen AC (1997) "Minisequencing: a specific tool for DNA analysis and diagnostics on oligonucleotide arrays." *Genome Res* 7(6):606-14

Pearson ER, Boj SF, Steele AM, Barrett T, Stals K, et al. (2007) "Macrosomia and hyperinsulinaemic hypoglycaemia in patients with heterozygous mutations in the HNF4A gene" *PLoS Med* 4(4):e118

Perry GH, Ben-Dor A, Tsalenko A, Sampas N, Rodriguez-Revenga L, et al. (2008) "The fine scale and complex architecture of human copy-number variation." *Am J Hum Genet* 82(3):685-95

Petrij F, Giles RH, Dauwerse HG, Saris JJ, Hennekam RC, et al. (1995) "Rubinstein-Taybi syndrome caused by mutations in the transcriptional co-activator CBP" *Nature*. 376(6538):348-51

Pfeufer A, van Noord C, Marciante KD, Arking DE, Larson MG, et al. (2010) "Genome-wide association study of PR interval" *Nat Genet* 42(2):153-9

Pflugfelder GO, Roth H, Poeck B (1992) "A homology domain shared between Drosophila optomotor-blindand mouse Brachyuryis involved in DNA binding." *Biochem Biophys Res Commun* 186**:**918–925

Picard F, Robin S, Lavielle M, Vaisse C, Daudin JJ (2005) “A statistical approach for array CGH data analysis.” *BMC Bioinformatics* 6(1):27

Pinkel D, Segraves R, Sudar D et al. (1998) “Hight reslution analysis of DNA copy number variation using comparative genomic hybridization to microarrays” *Nat Genet* 20:207-211

Pinney SE, MacMullen C, Becker S, Lin YW, Hanna C, et al. (2008) “Clinical characteristics and biochemical mechanisms of congenital hyperinsulinism associated with dominant KATP channel mutations” *J Clin Invest* 118(8):2877-86

Pique-Regi R, Monso-Varona J, Ortega A et al (2008) “Sparse representation and Bayesian detection of genme copy number alterations from microarray data” *Bioinformatics* 24(3):309-18

Piton A, Michaud JL, Peng H, Aradhya S, Gauthier J, et al. (2008) “Mutations in the calciumrelated gene IL1RAPL1 are associated with autism.” *Hum Mol Genet* 17:3965–3974

Plageman TF, Yutzey KE (2005) “T-box genes and heart development: putting the "T" in heart.” *Dev Dyn* 232(1):11-20.

Poliak S, Salomon D, Elhanany H, Sabanay H, Kiernan B, et al. (2003) “Juxtaparanodal clustering of Shaker-like K+ channels in myelinated axons depends on Caspr2 and TAG-1” *J Cell Biol* 162(6):1149-60

Pollack JR, Perou CM, Alizadeh AA, Eisen MB, Pergamenschikov A, et al. (1999) “Genome-wide analysis of DNA copy-number changes using cDNA microarrays.” *Nat Genet* 23(1):41–46

Pollack JR, Sørlie T, Perou CM, Rees CA, Jeffrey SS, et al. (2002) “Microarray analysis reveals a major direct role of DNA copy number alteration in the transcriptional program of human breast tumors.” *Proc Natl Acad Sci USA* 99:12963-12968

Poot M, Beyer V, Schwaab I, Damatova N, Van't Slot R, et al. (2010) “Disruption of CNTNAP2 and additional structural genome changes in a boy with speech delay and autism spectrum disorder” *Neurogenetics* 11(1):81-9

Price TS, Regan R, Mott R, Hedman A, Honey B, et al. (2005) “SW-ARRAY: a dynamic programming solution for the identification of copy-number changes in genomic DNA using array comparative genome hybridization data.” *Nucleic Acids Res* 33(11):3455-64

Purcell SM, Wray NR, Stone JL, Visscher PM, O’Donovan MC, et al. (2009) “Common polygenic variation contributes to risk of schizophrenia and bipolar disorder” *Nature* 460(7256):748-52

Raghavan M, Lillington DM, Skoulakis S, Debernardi S, Chaplin T, et al. (2005) “Genome-wide single nucleotide polymorphism analysis reveals frequent partial uniparental disomy due to somatic recombination in acute myeloid leukemias.” *Cancer Res* 65:375–378

Rahier J, Falt K, Muntefering H, Becker K, Gepts W, et al. (1984) "The basic structural lesion of persistent neonatal hypoglycaemia with hyperinsulinism: deficiency of pancreatic D cells or hyperactivity of B cells? *Diabetologia* 26:282-289

Redon R, Ishikawa S, Fitch KR, Feuk L, Perry G, et al. (2006) "Global variation in copy number in the human genome." *Nature* 444:444–54

Rallis C, Bruneau BG, Del Buono J, Seidman CE, Seidman JG, et al. (2003) "Tbx5 is required for forelimb bud formation and continued outgrowth" *Development* 130(12):2741-51

Rauch A, Hoyer J, Guth S, Zweier C, Kraus C, et al. (2006) "Diagnostic yield of various genetic approaches in patients with unexplained developmental delay or mental retardation" *Am J Med Genet* 140(19): 2063-74

Ravnan JB, Tepperberg JH, Papenhausen P et al. (2006) "Subtelomere FISH analysis of 11688 cases: an evaluation of the frequency and pattern of subtelomere rearrangement in individuals with developmental disabilities." *J Med Genet* 43:478-489

Raymond FL (2010) "Monogenic Causes of Mental Retardation" *Genetics of Mental Retardation. Monogr Hum Genet. Basel*, Karger, 18:89–100

Ribeiro MJ, De Lonlay P, Delzescaux T, Boddaert N, Jaubert F, et al. (2005) "Characterization of hyperinsulinism in infancy assessed with PET and 18F-fluoro-L-DOPA." *J Nucl Med* 46(4):560-6

Rivera-Brugués N, Albrecht B, Wieczorek D, Schmidt H, Keller T, et al. (2011) "Cohen syndrome diagnosis using whole genome arrays." *J Med Genet* 48(2):136-40

Roeleveld N, Zielhuis GA, Gabreels F (1997) "The prevalence of mental retardation: a critical review of recent literature." *Dev Med Child Neurol* 39(2):125-32

Rosenberg C, Knijnenburg J, Bakker E, Vianna-Morgante AM, Sloos W, et al. (2006) "Array-CGH detection of micro rearrangements in mentally retarded individuals: clinical significance of imbalances present both in affected children and normal parents." *J Med Genet* 43(2): 180-6

Ross P, Hall L, Smirnov I, Haff L (1998) "High-level multiplex genotyping by MALDI-TOF mass spectrometry." *Nat Biotechnol* 16(13): 1347-51

Ropers HH, Hamel BC (2005) "X-linked mental retardation" *Nat Rev Genet* 6(1):46-57

Ropers HH (2006) "X-linked mental retardation: many genes for a complex disorder." *Curr Opin Genet Dev* 16: 260–269

Ropers HH (2007) "New perspectives for the elucidation of genetic disorders" *Am J Hum Genet* 81:199-207

Ropers HH (2010) “Genetics of Early Onset Cognitive Impairment” *Annu Rev Genomics Hum Genet* 11:161-87

Ruvinsky I, Silver LM (1997) “Newly identified paralogous groups on mouse chromosomes 5 and 11 reveal the age of aT-box cluster duplication.” *Genomics* 40**:**262-266

Saiki RK, Gelfand DH, Stoffel S. Scharf SJ, Higuchi R, et al. (1988) "Primer-directed enzymatic amplification of DNA with a thermostable DNA polymerase" *Science* 239:487-491

Sanchez-Alavez M, Klein I, Brownell SE, Tabarean IV, Davis CN, et al. (2007) “Night eating and obesity in the EP3R-deficient mouse” *Proc Nat Acad Sci* 104: 3009-3014

Sanger F, Nicklen S, Coulson AR (1977). "DNA sequencing with chain-terminating inhibitors" *Proc Natl Acad Sci U S A* 74(12): 5463-7

Sasaki G, Ogata T, Ishii T, Hasegawa T, Sato S, et al. (2002) “Novel mutation of TBX3 in a Japanese family with ulnar-mammary syndrome: implication for impaired sex development.” *Am J Med Genet* 110(4):365-9

Schalock RL, Luckasson RA, Shogren KA, Borthwick-Duffy V et al. (2007) “ The renaming of mental retardation; understanding the change to the term intellectual disability” *Intellect Dev Disabil* 45;116-24

Shaw-Smith C, Redon R, Rickman L, Rio M, Willatt L, et al. (2004) “Microarray based comparative genomic hybridisation (array-CGH) detects submicroscopic chromosomal deletions and duplications in patients with learning disability/mental retardation and dysmorphic features.” *J Med Genet* 2004; 41(4): 241-8

Shu W, Yang H, Zhang L, Lu MM, Morrisey EE (2001) “Characterization of a new subfamily of winged-helix/forkhead (Fox) genes that are expressed in the lung and act as transcriptional repressors” *J Biol Chem* 276(29):27488-97

Schinzel A (1987) “Ulnar-mammary syndrome.” *J Med Genet* 24(12):778-81

Schinzel A, Illig R, Prader A (1987) “The ulnar-mammary syndrome: an autosomal dominant pleiotropic gene.” *Clin Genet* 32(3):160-8

Schoumans J, Ruivenkamp C, Holmberg E, Kyllerman M, Anderlid BM, et al. (2005) “Detection of chromosomal imbalances in children with idiopathic mental retardation by array based comparative genomic hybridisation (array-CGH).” *J Med Genet* 2005; 42(9): 699-705

Seabright M (1971) “A rapid banding technique for human chromosomes.” *Lancet* 2(7731): 971-2

Seifert W, Holder-Espinasse M, Spranger S, et al. (2006) ”Mutational spectrum of COH1 and clinical heterogeneity in Cohen syndrome” *J Med Genet* 43(5):e22.

Seifert W, Holder-Espinasse M, Kuhnisch J, et al. (2009) *“*Expanded mutational spectrum in Cohen syndrome, tissue expression, and transcript variants of COH1.” *Hum Mutat* 30(2):E404-20.

She X, Jiang Z, Clark RA, Liu G, Cheng Z, (2004) “Shotgun sequence assembly and recent segmental duplications within the human genome” *Nature* 431(7011):927-30

Sharp AJ, Hansen S, Selzer RR, Cheng Z, Regan R, (2006) "Discovery of previously unidentified genomic disorders from the duplication architecture of the human genome." *Nat Genet* 38(9): 1038-42

Schevell M, Ashwal S, Donley D, Flint J, Gingold M, et al. (2003) “ Practice parameter: evaluation of the child with global developmental delay: Report of Quality Standards Subcommittee of the American Academy of Neurology and The Practice Committee of the Child Neurology Society.” *Neurology* 60, 367-380

Schmickel RD (1986) “Contiguous gene syndromes: a component of recognizable syndromes.” *J Pediatr* 109:231–41

Schinzel A (1987) “Ulnar-mammary syndrome” *J Med Genet* 24(12): 778-81

Schoumans J, Ruivenkamp C, Holmberg E, Kyllerman M, Anderlid BM, et al. (2005) “Detection of chromosomal imbalances in children with idiopathic mental retardation by array based comparative genomic hybridization (array-CGH).” *J Med Genet* 42:699-705

Sebat J, Lakshmi B, Troge J, Alexander J, Young J, et al. (2004) “Large-scale copy number polymorphism in the human genome.” *Science* 23;305(5683):525-8

Sharp AJ, Locke DP, McGrath SD, Cheng Z, Bailey JA, et al. (2005) “Segmental duplications and copy-number variation in the human genome.” Am J Hum Genet 77(1):78-88

Sharp AJ, Mefford HC, Li K, Baker C, Skinner C, et al. (2008) “A recurrent 15q13.3 microdeletion syndrome associated with mental retardation and seizures.” *Nat Genet* 40(3):322-8

Shaw-Smith C, Redon R, Rickman L, Rio M, Fiegler H, et al. (2004) “Microarray based comparative genomic hybridization (array-CGH) detects submicroscopic chromosomal deletions and duplications in patients with learning disability/mental retardation and dysmorphic features.” *J Med Genet* 41:241-248

Shaw-Smith C, Pittman AM, Willatt L, Martin H, Rickman L (2006)“Microdeletion encompassing MAPT at chromosome 17q21.3 is associated with developmental delay and learning disability.” *Nat Genet*. 38(9):1032-7

Shevell MI, Bejjani BA, Srour M, Rorem EA, Hall N, et al. (2008) “Array comparative genomic

hybridization in global developmental delay.“ *Am J Med Genet B Neuropsychiatr Genet* 147B(7):1101-8

Showell C, Binder O, Conlon FL (2004) “T-box genes in early embryogenesis” *Dev Dyn* 229(1):201-18

Slager RE, Newton TL, Vlangos CN, Finucane B, Elsea SH (2003) “Mutations in RAI1 associated with Smith-Magenis syndrome.” *Nat Genet* 33(4):466-8

Slater HR, Bailey DK, Ren H, Cao M, Bell K, et al. (2005) “High-resolution identification of chromosomal abnormalities using oligonucleotide arrays containing 116,204 SNPs.” *Am J Hum Genet* 77: 709-726

Smith JG, Magnani JW, Palmer C, Meng YA, Soliman EZ, et al. (2011) “Genome-Wide Association Studies of the PR Interval in African Americans.” *PLOS Genetics* 7(2):e1001304

Southern E, Mir K, Shchepinov M (1999) “Molecular interactions on microarrays“ *Nat Genet* 21(1 Suppl):5-9

Spiteri E, Konopka G, Coppola G, Bomar J, Oldham M, et al. (2007) “Identification of the transcriptional targets of FOXP2, a gene linked to speech and language, in developing human brain” *Am J Hum Genet*. 81(6):1144-57

Stanley CA (1997) “Hyperinsulinism in infants and children” *Pediatr Clin North Am* 44:363-74

Stanley CA, Lieu YK, Hsu BY, Burlina AB, Greenberg CR, et al. (1998) “Hyperinsulinism and hyperammonemia in infants with regulatory mutations of the glutamate dehydrogenase gene” *N Engl J Med* 338(19):1352-7

Steemers FJ, Chang W, Lee G, Barker DL et al (2006) “Whole-genome genotyping with the single-base extension assay.” *Nat Methods* 3(1):31-3

Solinas-Toldo S. Lampel S, Stilgenbauer S, Nickolenko J, Benner A, et al. (1997) “Matrix-based comparative genomic hybridization: biochips to screen for genomic imbalances.” *Genes Chromosomes Cancer*, 20:399-407

Sugimoto, Y., Narumiya, S (2007) “Prostaglandin E receptors.” *J Biol Chem* 282 (16): 11613-7

Sutherland GR, Baker E, Richards RI (1998) “Fragile sites still breaking” *Trends Genet* 14:501-516

Tabolacci E, Pomponi MG, Pietrobono R, Terracciano A, Chiurazzi P, et al. (2006) “A truncating mutation in the IL1RAPL1 gene is responsible for X-linked mental retardation in the MRX21 family.” *Am J Med Genet Part A* 140A:482–487

Taddei I, Morishima M, Huynh T, Lindsay EA.(2001) “Genetic factors are major determinants of phenotypic variability in a mouse model of the DiGeorge/del22q11 syndromes” *Proc Natl Acad Sci U S A* 98(20):11428-31

Teramitsu I, Kudo LC, London SE, Geschwind DH, White SA. (2004) "Parallel FoxP1 and FoxP2 expression in songbird and human brain predicts functional interaction" *J Neurosci.* 24(13):3152-63

Tarpey PS, Smith R, Pleasance E, Whibley A, Edkins S, et al. (2009) "A systematic, large scale resequencing screen of X-chromosome coding exons in mental retardation." *Nat Genet* 41: 535–543

Terrett JA, Newbury-Ecob R, Cross G.S, Fenton I, Raeburn JA, et al. (1994) "Holt-Oram syndrome is a genetically heterogeneous disease with one locus mapping to human chromosome 12q" *Nature Genet* 6: 401-404

The 1000 Genomes Project Consortium (2011) "A map of human genome variation from population scale Sequencing" *Nature* 467(7319): 1061–1073

The International HapMap 3 Consortium (2010) "Integrating common and rare genetic variation in diverse human populations" *Nature* 467(7311):52-8.

Thomas P, Ye Y, Lightner E (1996) "Mutation of the pancreatic islet inward rectifier Kir6.2 also leads to familial persistent hyperinsulinemic hypoglycemia on infancy" *Hum Mol Genet* 5:1809-12

Thomas PM, Cote GJ, Wohllk N, Haddad B, Mathew PM, et al. (1995) "Mutations in the sulfonylurea receptor gene in familial persistent hyperinsulinemic hypoglycemia of infancy" *Science* 268:426-9

Tong YK, Jin S, Chiu RW, Ding C, Chan KC, et al. (2010) "Noninvasive prenatal detection of trisomy 21 by an epigenetic-genetic chromosome-dosage approach." *Clinical Chemistry*, 56, 90–98

Tran PO, Gleason CE, Robertson RP (2002) " Inhibition of interleukin-1beta-induced COX2 and EP3 gene expression by sodium salicylate enhances pancreatic islet beta-cell function." *Diabetes* 51:1772-1778

Tuzun E, Sharp AJ, Bailey JA, Kaul R, Morrison VA, et al. (2005) Fine-scale structural variation of the human genome. *Nat Genet*. 37(7):727-32

Tyson C, Hardvard C, Locker R, Friedman JM, Langlois S, et al. (2005) "Submicroscopic deletions and duplications in individuals with intellectual disability detected by array-CGH." *Am J Med Genet* 139A:173-185

Vaughan CJ, Basson CT (2000) "Molecular Determinants of Atrial and Ventricular Septal Defects and Patent Ductus Arteriosus"*Am J Med Genet* 97(4):304-9. Review

Vermeesch JR, Melotte C, Froyen G, Van Vooren S, Dutta B, et al. (2005) "Molecular karyotyping: array CGH quality criteria for constitutional genetic diagnosis." *J Histochem Cytochem* 53:413–422

Vermeesch JR, Fiegler H, de Leeuw N, Szuhai K, Schoumans J, et al. (2007) "Guidelines for molecular karyotyping in constitutional genetic diagnosis." *Eur J Hum Genet* 15(11):1105-14

Vernes SC, MacDermot KD, Monaco AP, Fisher SE (2009) "Assessing the impact of *FOXP1* mutations on developmental verbal dyspraxia" *Eur J Hum Genet* 17(10):1354-8

Vernes SC, Newbury DF, Abrahams BS, Winchester L, Nicod J, (2008) "A functional genetic link between distinct developmental language disorders" *N Engl J Med* 359(22):2337-45

Vernes SC, Nicod J, Elahi FM, Coventry JA, Kenny N, et al. (2006) "Functional genetic analysis of mutations implicated in a human speech and language disorder" *Mol Genet* 15(21):3154-67

Venkatraman E, Olshen A (2007) "A faster circular binary segmentation algorithm for the analysis of aCGH data." *Bioinformatics* 23, 657–663

Venter JC et al (2001) "The sequence of the human genome." *Science* 291(5507):1304-51

Vissers LE, de Vries BB, Osoegawa K, Janssen IM, Feuth T, et al. (2003) "Array-based comparative genomic hybridization for the genomewide detection of submicroscopic chromosomal abnormalities." *Am J Hum Genet* 73(6): 1261-70

Vissers LE, de Vries BB, Veltman JA (2009) "Genomic microarrays in mental retardation: from CNV to gene, from research to diagnosis" *J Med Genet* 47(5):289-97

Vissers L, de Ligt J, Gilissen C, Janssen I, Steehouwer M, et al. (2010) "A de novo paradigm for mental retardation" *Nat Genet* 42(12):1109-12

Wagenstaller J, Spranger S, Lorenz-Depiereux B, Kazmierczak B, Nathrath M, et al. (2007) "Copy-number variations measured by single-nucleotide-polymorphism oligonucleotide arrays in patients with mental retardation" *Am J Hum Genet* 81(4):768-79

Wallace RB, Shaffer J, Murphy RF, Bonner J, Hirose T, et al. (1979) "Hybridization of synthetic oligodeoxyribonucleotides to phi chi 174 DNA: the effect of single base pair mismatch." *Nucleic Acids Res* 6(11):3543-57

Wang P, Kim Y, Pollack J, Narasimhan B, Tibshirani R (2005) "A method for calling gains and losses in array CGH data." *Biostatistics* 6:45-58

Wang K, ZhangH, Ma D, Bucan M, Glessner JT, et al. (2009) "Common genetic variants on 5p14.1 associate with autism spectrum disorders." *Nature* 459:528–33

Wang K, Shi D, Zhu P, Dai J, Zhu L, et al. (2010) "Association of a single nucleotide polymorphism in Tbx4 with developmental dysplasia of the hip: a case-control study." *Osteoarthritis Cartilage* 18(12):1592-5

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

Wang J, Wang W, Li R, Li Y, Tian G, et al. (2008) "The diploid genome sequence of an Asian individual." *Nature* 456 (7218):60–5

Weiss LA, Arking DE, Daly MJ, et al. (2009) "A genome-wide linkage and association scan reveals novel loci for autism." *Nature* 461:802–8

Wheeler DA, Srinivasan M, Egholm M, Shen Y, Chen L, et al. (2008) "The complete genome of an individual by massively parallel DNA sequencing." *Nature* 452:872–6

Whibley AC, Plagnol V, Tarpey PS, Abidi F, Fullston T, et al. (2020) "Fine-scale survey of X Chromosome copy number variants and indels underlying intellectual disability" *Am J Med Genet* 87:173-188

Willard HF (2000) "The sex chromosomes and X-chromosome inactivation. In: The Metabolic and Molecular Bases of Inherited Diseas" *New York, USA:McGraw-Hill* 1191-1221

Willenbrock H, Fridlyand J (2005) "A comparison study: applying segmentation to array CGH data for downstream analyses." *Bioinformatics* 21(22):4084-91

Wilson V, Conlon FL (2002) "The T-box family" *Genome Biol* 3(6): REVIEWS3008.

Winchester L, Yau C, Ragoussis J (2009) "Comparing CNV detection methods for SNP arrays." *Brief Funct Genomic Proteomic* 8(5):353-66

World Heath Organization "The ICD-10 Classification of Mental and Behavioral Disorders" International Statistical Classification of Diseases and Related Health Problems (10th ed.) 2007 Geneva

Xiao Y, Segal M, Yang Y, Yeh R (2007) "A multi-array multi-SNP genotyping algorithm for Affymetrix SNP microarrays" *Bioinformatics* 23(12):1459-1467

Xing J, Zhang Y, Han K, Salem AH, Sen SK, et al. (2009) "Mobile elements create structural variation: analysis of a complete human genome" *Genome Res* 19(9):1516-26

Xu B, Roos JL, Levy S, van Rensburg EJ, Gogos JA, et al. (2008) "Strong association of *de novo* copy number mutations with sporadic schizophrenia." *Nature Genet* 40, 880–885

Yang J, Hu D, Xia J, Yang Y, Ying B, et al. (2000) "Three novel TBX5 mutations in Chinese patients with Holt-Oram syndrome." *Am J Med Genet* 92:237–240

Yang XR, Ng D, Alcorta DA, Liebsch NJ, Sheridan E, et al. (2009) " Copy number variation in human health, disease, and evolution" *Annu Rev Genomics Human Genet* 10:451-81

Zhang F, Gu W, Hurles ME, Lupski JR (2009) "Copy Number Variation in Human Health, Disease, and Evolution" *Annu Rev Genomics Hum Genet* 10:451–81

Zhang YH, Huang BL, Niakan KK, McCabe LL, MacCabe ER, et al. (2004) "IL1RAPL1 is associated with mental retardation in patients with complex glycerol kinase deficiency who have deletions extending telomeric of DAX1" *Hum Mutat* 24(3):273

Zechner U, Wilda M, Kehrer-Sawatzki H, Vogel W, Fundele R, et al. (2001) "A high density of X-linked genes for general cognitive ability: a run-away process shaping human evolution? " *Trends Genet* 17(12): 697-701

Zemni R, Bienvenu T, Vinet MC, Sefiani A, Carrie A, et al. (2000) "A new gene involved in X-linked mental retardation identified by analysis of an X;2 balanced translocation." *Nat Genet* 24:167–170

Zogopoulos G, Ha KC, Naqib F, Moore S, Kim H, Montpetit A, et al. (2007) "Germ-line DNA copy number variation frequencies in a large North American population." *Hum Genet* 122:345–53

Zweier C, de Jong EK, Zweier M, Orrico A, Ousager LB, Collins AL, Bijlsma EK, Oortveld MA, Ekici AB, Reis A, Schenck A, Rauch A. (2009) "CNTNAP2 and NRXN1 are mutated in autosomal-recessive Pitt-Hopkins-like mental retardation and determine the level of a common synaptic protein in Drosophila" *Am J Hum Genet* 85(5):655-66

Zweier C, Sticht H, Aydin-Yaylagul I, Campbell CE, Rauch A (2007) "Human TBX1 missense mutations cause gain of function resulting in the same phenotype as 22q11.2 deletions" *Am J Hum Genet* 80:510-5

APPENDIX

HUMAN MUTATION Mutation in Brief 31: E1851-E1860 (2010) Online

MUTATION IN BRIEF

HUMAN MUTATION

Identification of FOXP1 Deletions in Three Unrelated Patients with Mental Retardation and Significant Speech and Language Deficits

Denise Horn[1,§], Johannes Kapeller[2,§], Núria Rivera-Brugués[3], Ute Moog[4], Bettina Lorenz-Depiereux[3], Sebastian Eck[3], Maja Hempel[5], Janine Wagenstaller[3], Alex Gawthrope[6], Anthony P. Monaco[6], Michael Bonin[7], Olaf Riess[7], Eva Wohlleber[8], Thomas Illig[9], Connie R. Bezzina[10], Andre Franke[11], Stephanie Spranger[12], Pablo Villavicencio-Lorini[1], Wenke Seifert[1,13,14], Jochen Rosenfeld[15], Eva Klopocki[1], Gudrun A. Rappold[2,§,$], and Tim M. Strom[3,5,§,$]

[1] Institute of Medical Genetics, Charité, University Medicine of Berlin, Berlin, Germany; [2] Department of Molecular Human Genetics, Ruprecht-Karls-University, Heidelberg, Germany; [3] Institute of Human Genetics, Helmholtz Zentrum München, German Research Center for Environmental Health, Neuherberg, Germany; [4] Institute of Human Genetics, Ruprecht-Karls-University, Heidelberg, Germany; [5] Institute of Human Genetics, Technische Universität München, Munich, Germany; [6] Wellcome Trust Centre for Human Genetics, University of Oxford, Oxford, UK; [7] Department of Medical Genetics, Institute of Human Genetics, University of Tübingen,Tübingen, Germany; [8] Institute of Human Genetics, Rheinische Friedrich-Wilhelms-University, Bonn, Germany;[9] Institute of Epidemiology, Helmholtz Zentrum München, German Research Center for Environmental Health, Neuherberg, Germany; [10] Heart Failure Research Center, Department of Experimental Cardiology, Academic Medical Center, University of Amsterdam, Amsterdam, The Netherlands; [11] Institute for Clinical Molecular Biology, Christian-Albrechts-University zu Kiel, Kiel, Germany;[12] Praxis für Humangenetik, Bremen, Germany;[13] Cologne Center for Genomics, Universität zu Köln, Cologne, Germany; [14] Faculty of Biology, Chemistry, and Pharmacy, Free University of Berlin, Berlin, Germany; [15] Department of Audiology and Phoniatrics, Charité, University Medicine of Berlin, Berlin, Germany

[$]Correspondence to Dr. Tim M. Strom, Institute of Human Genetics, Helmholtz Zentrum München, Ingolstädter Landstr. 1, 85764 Neuherberg, Germany. Phone: +49 89 3187 3296; E-mail: TimStrom@helmholtz-muenchen.de and Dr. Gudrun A. Rappold, Deptartment of Human Molecular Genetics, Institute of Human Genetics, Im Neuenheimer Feld 366, 69120 Heidelberg, Germany. Phone: +49 6221 565059; E-mail: gudrun.rappold@med.uni-heidelberg.de.

[§]These authors contributed equally to this work.

Communicated by Mark H. Paalman

ABSTRACT: Mental retardation affects 2-3% of the population and shows a high heritability. Neurodevelopmental disorders that include pronounced impairment in language and speech skills occur less frequently. For most cases, the molecular basis of mental retardation with or without speech and language disorder is unknown due to the heterogeneity of underlying genetic factors. We have used molecular karyotyping on 1523 patients with mental retardation to detect copy number variations (CNVs) including deletions or duplications. These studies revealed three heterozygous overlapping deletions solely affecting the forkhead box P1 (FOXP1) gene. All three patients had moderate mental retardation and significant language and speech deficits. Since our results are consistent with a de novo occurrence of these deletions, we considered them as causal

Received 25 May 2010; accepted revised manuscript 31 August 2010.

DOI:10.1002/humu.21362

although we detected a single large deletion including FOXP1 and additional genes in 4104 ancestrally matched controls. These findings are of interest with regard to the structural and functional relationship between FOXP1 and FOXP2. Mutations in FOXP2 have been previously related to monogenic cases of developmental verbal dyspraxia. Both FOXP1 and FOXP2 are expressed in songbird and human brain regions that are important for the developmental processes that culminate in speech and language.

KEY WORDS: FOXP1, mental retardation, copy number variations, language and speech deficits

INTRODUCTION

Mental retardation (MR) is defined as a significant impairment of cognitive and adaptive functions with an onset before the age of 18 years (Ropers, 2008). Based on the assessment of patients' intelligence quotients (IQs), MR is usually classified into mild (IQ 50-70), moderate (IQ > 35) and severe (IQ > 20) forms. It has been shown that the causes of the disorder include environmental and genetic factors (Inlow and Restifo, 2004), still for most cases, the pathological basis remains unexplained. The high degree of heritability of MR is highlighted by the estimation that up to 50% of severe cases are caused by genetic abnormalities (Inlow and Restifo, 2004). Due to the heterogeneity of the underlying genetic factors, the identification of MR candidate genes still remains difficult to date. A promising approach to detect small chromosomal copy number variants (CNVs) in the genome is the use of molecular karyotyping techniques (Marshall, et al., 2008). CNVs have been shown as the underlying cause or susceptibility factors for a variety of autism spectrum disorders and conditions associated with MR (Berkel, et al., 2010; Marshall, et al., 2008; Sebat, et al., 2007; Zweier, et al., 2010). Array-based comparative genomic hybridization has also enabled the detection of interstitial submicroscopic copy number alterations in about 10 % of patients with MR (de Vries, et al., 2005). Here we present the identification of overlapping heterozygous deletions affecting the *FOXP1* (MIM# 605515) gene in three unrelated patients with MR and significant speech and language disorder.

MATERIALS AND METHODS

Oligonucleotide arrays

CNV data were generated in different institutions within the MRNET consortium (Supp. Table S1). Patient 1 was part of a cohort of 387 patients investigated with Infinium Human550-Quad and Human610-Quad arrays (Illumina). Intensity data were normalized as described previously (Wagenstaller, et al., 2007). Segmentation was performed with circular binary segmentation as implemented in the R-package 'DNAcopy'. Patient 2 was part of a cohort of 188 patients investigated using whole genome oligonucleotide 244K arrays (Agilent Technologies, Santa Clara, CA). Image data were analyzed using Feature Extraction 9.5.3.1 and CGH Analytics 3.4.40 software (Agilent Technologies, Santa Clara, CA) with the following analysis settings: aberration algorithm ADM-2; threshold: 6.0; window size: 0.2 Mb; filter: 5probes, log2ratio = 0.29. Patient 3 belonged to a cohort of 184 patients analyzed with genome-wide human SNP 6.0 arrays (Affymetrix, Santa Clara, CA). Analysis of data was performed using the Genotyping Console Software 3.0 (Affymetrix). For the detection of genomic deletions and duplications, automated analysis by Segment Reporting Tools was used. Regions showing at least 5-10 aberrant neighboring SNPs / markers and having a size of at least 100 kb were classified as being significant. Additional CNV data of 764 MR patients from other institutions not listed here were obtained through the database of the MRNET (www.german-mrnet.de).

Control populations

Controls consisted of 1146 individuals from popgen, 813 individuals of a population-based cohort (KORA study), 972 patients with cardiac ischemia (AGNES study), 482 patients with early-onset lung cancer (LUCY study), and 691 long-lived individuals (LLI study).

Breakpoint identification

PCR reactions on genomic DNA level were performed to amplify the junction fragments that spanning the telomeric and centromeric breakpoints. Fragments were directly sequenced with the respective forward and reverse primer: B35_FOXP1F/B8_FOXP1R (patient 1): 5'-atgctgaaggtggaatggg-3', 5'-ggccacatacgtgttgtcag-3'; O06_for/Z02_rev (patient 2): 5'-cgttgccagctcaaggttat-3', 5'-taagtgtgtgcgaagccaag-3'; bp_FOXP1_3for/rev (patient 3): 5'-gcacctgaccctctagctca-3', 5'-ggttcagccactggtctttc-3'.

Fluorescence *in situ* hybridization (FISH)

Preparation of chromosome metaphases of patient 3 and his parents and FISH were performed according to standard protocols using BAC DNA probes RP11-215K24, RP11-154H23, CTD-3121O8 (Invitrogen, Darmstadt, Germany) and RP11-788D09 (BACPAC Resource Center, Oakland, USA).

Mutation screening

FOXP1 (NM_032682.4) exons were PCR amplified using intronic primers and investigated with 2 different methods: 197 DNAs were investigated by direct sequencing, 772 DNAs were analyzed using dye-binding/high-resolution DNA melting point analysis (LightScanner HR I 384, Idaho Technology). Genotyping of the controls were performed on a MALDI-TOF mass-spectrometer (Sequenom MassArray system) using the homogeneous mass-extension (hME) process for producing primer extension products. Primers were designed with ExonPrimer (http://ihg.helmholtz-muenchen.de/ihg/ExonPrimer.html) and are available on request. Nucleotide numbering reflects cDNA numbering with +1 corresponding to the A of the ATG translation initiation codon.

SLIC study

The SLIC probands were selected from samples ascertained by the Newcomen Centre at Guys Hospital and by the Manchester Language Study (The SLI Consortium, 2002) and all had severe language impairments with language skills more than 1.5 SD below the normative mean for their chronological age in combination with full IQ scores at least 0.7 SD below that expected for their age. IQ scores are derived from the Wechsler Intelligence Scales for Children (WISC-III). The 46 (16 males, 30 females; average age: 11 years and 3 months) individuals would not meet strict SLI diagnostic criteria as individuals with IQ problems are usually excluded. Nonetheless all probands had severe language impairments which require special schooling arrangements and continued support.

RESULTS

FOXP1 deletions

In a collaborative effort we performed a genome-wide microarray scan for CNVs in a German cohort of 1523 unrelated patients with unexplained mental retardation. Standard diagnostic tests like chromosomal karyotyping, fragile X testing and subtelomeric screenings were performed in most cases to rule out known causes of MR. The recruitment of patients was part of the German Mental Retardation Network (MRNET) study. Approval for the study was obtained by the ethical review boards of the participating institutions and informed written consent was obtained from all participants.

Copy number analysis revealed overlapping deletions at chromosome 3p14.1 affecting solely the *FOXP1* gene in three unrelated cases, two males and one female aged between 5.5 to 7 years. Deletion sizes of 498 kb, 659 kb and 1047 kb included all but the first of the coding exons in patient 1 (Wagenstaller, et al., 2007) and the entire coding region in the other two patients 2 and 3 (Figure 1A). The deletions were verified by fluorescence *in situ* hybridization (FISH; Figure 1B) or quantitative real-time PCR (data not shown). DNA analysis of the unaffected parents by molecular karyotyping and FISH indicated that patients 2 and 3 carried *de novo* deletions, while it could only be shown that the mother of patient 1 does not carry the deletion since the father was not available for investigation. We found that the deleted alleles were of paternal origin in all three patients. Sequencing of the coding exons of the remaining *FOXP1* allele did not reveal any sequence variation compared to the annotated reference sequence (NM_032682). Characterization of the breakpoints was achieved by amplification and sequencing of the respective junction fragments (GenBank EF504249, HM124444, and GU980955; Figure 1C).

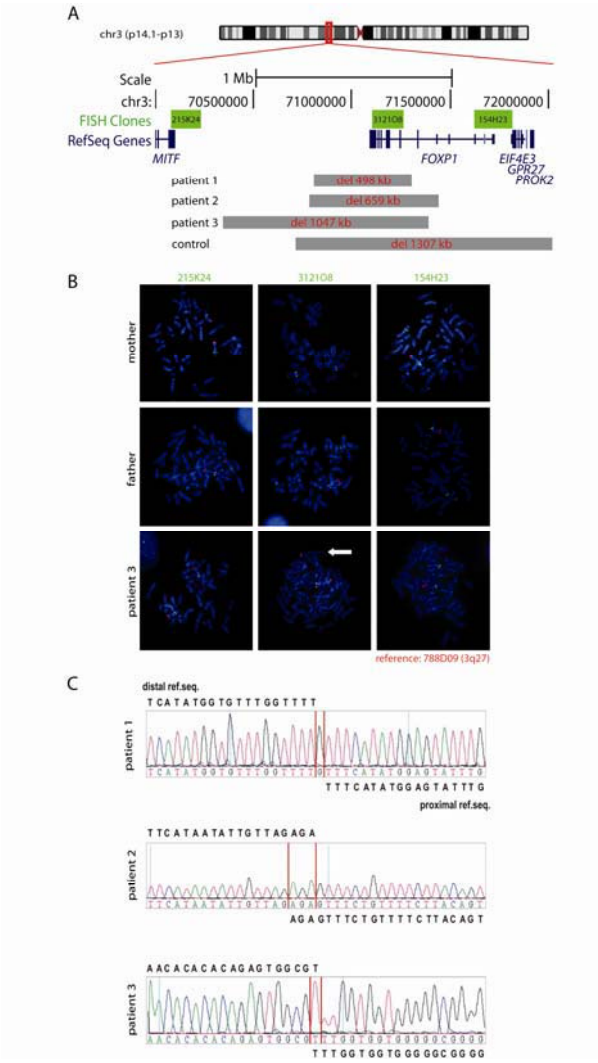


Figure 1. *FOXP1* deletions in patients with MR.
(A) Schematic presentation of the position and size of the deleted regions (grey bars) at chromosome 3p14.1 in three patients with MR and in a control individual. Only the *FOXP1* gene (NM_032682, minus strand) is affected. The coding exons are indicated by larger vertical lines compared to the non-coding exons.
(B) Results of fluorescence *in situ* hybridization (FISH) shown for patient 3 and his parents. The positions of the BAC clones RP11-215K24 and RP11-154H23 flanking the deletion, as well as CTD-3121O8 located within the deletion (green signals) are given in Figure 1A. BAC clone RP11-778D09 was used as a reference probe located on chromosome 3q27 (red signals). The white arrow indicates the missing signal for clone CTD-3121O8 on one of the patient's chromosomes 3 while the signal is present in both chromosomes of the parents.
(C) Breakpoint characterization by sequencing of junction fragments amplified by PCR on genomic DNA of the patients.
In patient 1, the distal breakpoint was located at position 70,807,767, the proximal breakpoint at position 71,305,965 with a deletion size of 498,198 bp. In patient 2, the two breakpoints were located within a microhomology of three nucleotides (AGA) at the telomeric side at position 70,778,067-70,778,070 and the centromeric side at 71,437,354-71,437,357 resulting in a deletion size of 659,287 bp. For patient 3, the telomeric breakpoint is located at position 70,341,246-70,341,247 and the centromeric breakpoint at position 71,388,173–71,388,174 with a deletion size of 1,046,927 bp. All positions are given according to hg18, UCSC Human Genome March 2006.

For the deletion in patient 1, the distal breakpoint is located within a long interspersed nuclear element (LINE) repeat (L1/L1PA5) while the proximal breakpoint lies in a MER1-type repeat (MER58A). No repetitive sequences are present at the breakpoint loci in patient 2, however, microhomoloy of 3 bases (AGA) is observed at the breakpoints. The distal breakpoint of the deletion in patient 3 is located in a short interspersed nuclear element (SINE) while repetitive sequences are missing at the proximal breakpoint. The absence of homologous regions at the breakpoints suggests that the deletions were likely generated by double-strand breaks and non-homologous end-joining.

Control populations

To assess the significance of these findings, we checked the *FOXP1* region for CNVs in the *Database of Genomic Variants* (DGV) (Iafrate, et al., 2004) and in 4104 ancestrally matched controls which have been investigated with different types of high-density SNP arrays containing at least 500,000 SNPs (Supp. Table S1). No CNVs affecting the coding region were found in the DGV database. Copy number analysis in the 4104 control individuals revealed the presence of a single large 3p14.1p13 deletion of approximately 1.3 Mb affecting *FOXP1*, *EIF4E3* (MIM# 609896), *PROK2* (MIM# 607002) and *GPR27* (MIM# 605187) in an individual of the LLI study, with no indications of a comparable MR phenotype. Detailed clinical data on intellectual abilities were lacking and

further investigations not possible due to the design of the study. In particular, we could not investigate the presence of somatic mosaicism. Application of Fisher's exact test on these data, not considering the controls contained in the DGV database, resulted in a nominal p-value of 0.06. Thus, although the absence of deletions in the unaffected parents is consistent with a role of *FOXP1* in the clinical findings of the patients incomplete penetrance of deletions in this chromosomal interval cannot be ruled out.

Speech and language development

Detailed clinical investigation of the three patients revealed that they present moderate MR in combination with a general developmental delay (Table 1). In all patients, the non-verbal IQ score was assessed as ≤ 50 (3 SD below the mean). Speech and language development was estimated in the same range. All three patients started to speak at age of 3.5 years and used only combinations of two words at ages 5, 5.5 and 7 years, respectively. In all patients, the productive and receptive vocabulary came up to less than 100 words. Expressive language was more affected than receptive abilities. Dysgrammatism and very poor speech articulation with difficulties producing consonants at the beginning of words was present in all patients. Two showed oromotoric problems including difficulties with lip protrusion. In infancy, patients had a tendency to keep their mouth open and patient 2 suffered from swallowing difficulties. In addition, all showed considerably retarded gross-motor development with unsupported walking as late as 24 to 36 months. Brain magnetic resonance imaging and electroencephalography did not reveal any abnormalities. None of the patients showed sensorineural hearing loss. Ophthalmologic testing disclosed moderate myopia only in patient 1. Upon physical examinations, normal growth parameters regarding height and occipitofrontal head circumference were documented for two patients but presence of obesity was striking in both. Consistent craniofacial anomalies seen in patients 2 and 3 included broad and prominent forehead and frontal hair upsweep (Figure 2).

Table 1. Clinical description of *FOXP1* deletion patients

	Patient 1	Patient 2	Patient 3
General features			
Age of last assessment	7 years	5.5 years	6 years
Sex	Male	Female	Male
Occipitofrontal head circumference (SD)	+1.2	+0.8	+1.3
Height (SD)	+1.4	+0.6	-0.4
Weight (SD)	+0.5	+2.7	+2.5
Facial gestalt	no dysmorphisms	Prominent forehead, frontal hair upsweep	Prominent forehead, frontal hair upsweep
Non-verbal performance IQ	< 50	< 50	50
Gross motor delay	+	+	+
Age at sitting	n.d.	12 months	8 months
Age at walking	24 months	36 months	24 months
Swallowing difficulties	n.d.	+	-
Oromotor problems (e.g. lip protrusion, tongue elevation)	+	+	-
Speech and language development			
First vocalizing at age of	n.d.	4 months	12 months
First words at age of	3.5 years	3.5 years	3.5 years
Combined words at age of	7 years	5 years	5.5 years
Articulation problems	+	+	+
Poor grammar	+	+	+

n.d., not documented

FOXP1 point mutations

In order to detect possible *FOXP1* point mutations, we sequenced the coding exons in 883 patients with MR from the MRNET database and discovered eight different non-synonymous, three synonymous and nine non-

coding variants (Table 2). All variants were genotyped in up to 676 unrelated healthy controls. Three of the identified coding variants, p.Ser5Pro, p.Pro215Ala and p.Thr390Ser were also present at similar frequencies in healthy matched controls, but none of the coding variants was found in the HapMap or the NCBI dbSNP database. Five of the non-synonymous variants were transmitted through an apparently unaffected parent; for the other variants, heredity could not be established as either one or both parents were not available (Table 2). Variants transmitted by an unaffected parent are usually classified as non-pathogenic. Still, presuming a multigenic threshold model for MR, disease relevance cannot be completely ruled out. To uncover the neurobiological significance of these variants, investigation of the functional effects of the putative mutations is needed.

Since the core phenotype of the three patients carrying *FOXP1* deletions consisted of MR with significant speech and language deficits, we additionally sequenced *FOXP1* coding exons in DNA of 40 patients with a tentative diagnosis of Angelman syndrome (without microcephaly and negative for *SNRPN* imprinting). We also sequenced the DNA of 46 probands from the Specific Language Impairment Consortium (SLIC) collection (The SLI Consortium, 2002). For both groups, no sequence variants in the coding region of *FOXP1* were detected.

Table 2. Summary of *FOXP1* variants identified in controls and in patients with mental retardation

Variants			MR patients				Controls		
			Genotypes			Status of inheritance	Genotypes		
			11	**12**	**22**		**11**	**12**	**22**
Non-synonymous variants									
c.13T>C	p.Ser5Pro	Exon 6	882	1	0	maternal	674	2	0
c.226_228dupCAG	p.Gln76dup	Exon 7	882	1	0	maternal	676	0	0
c.301A>G	p.Met101Val	Exon 8	882	1	0	n.a.	667	0	0
c.643C>G†	p.Pro215Ala	Exon 10	882	1	0	maternal	336	2	0
c.781T>C	p.Ser261Pro	Exon 11	882	1	0	n.a.	676	0	0
c.1168A>T	p.Thr390Ser	Exon 15	882	1	0	maternal	675	1	0
c.1709A>G	p.Asn570Ser	Exon 19	882	1	0	paternal	338	0	0
c.1790A>C	p.Asn597Thr	Exon 20	882	1	0	n.a.	676	0	0
Synonymous variants									
c.768G>A	p.Thr256Thr	Exon 11	882	1	0	n.a.	673	0	0
c.1188G>A	p.Ser396Ser	Exon 15	882	1	0	maternal	674	0	0
c.1515C>T	p.Asn505Asn	Exon 17	882	1	0	maternal	674	0	0
Non-coding variants									
c.1-5G>A		5'UTR	882	1	0	paternal	674	0	0
c.180+49T>C		Intron 6	80	30	1	not tested (dbSNP rs2037474)			
c.181-29G>A†		Intron 6	875	7	1	1 x mat, 2 x pat, 1 x mat + pat, 4 x n.a.	653	21	0
c.181-30C>T		Intron 6	882	1	0	n.a.	676	0	0
c.664+11A>G		Intron 10	882	1	0	maternal	674	0	0
c.664+6C>T		Intron 10	882	1	0	n.a.	674	0	0
c.975-14A>G		Intron 12	882	1	0	maternal	673	1	0
c.1889+20A>C†		Intron 20	95	15	1	not tested (dbSNP rs7638391)			
c.1890-15G>T†		Intron 20	110	1	0	not tested (dbSNP rs7639736)			

Listed is the number of individuals carrying the respective homozygous or heterozygous genotypes (11, 12 or 22; GenBank NM_032682.4). Patients and matched healthy controls were of German origin. †also found by Vernes *et al*, 2009. MR, mental retardation; n.a., parents not available; mat, maternal; pat, paternal. Nucleotide numbering reflects cDNA numbering with +1 corresponding to the A of the ATG translation initiation codon.

DISCUSSION

This is the first report of 3p14.1 deletions that solely affect the *FOXP1* gene in three unrelated patients with moderate mental retardation in combination with a significant impairment of speech and language abilities. We investigated a large control population and the apparent finding of a large 3p14.1p13 deletion in this cohort affecting *FOXP1* and further genes illustrates the difficulties to assess the significance of rare mutational events in extremely heterogeneous diseases and may point to incomplete penetrance. The additional findings of obesity and mild craniofacial anomalies (prominent forehead and frontal hair upsweep) in two patients might be part of the

clinical spectrum associated with *FOXP1* deletions. A *de novo* deletion of this chromosomal interval on 3p14.1p13 affecting part of *FOXP1* and three additional genes (*EIF4E3*, *PROK2* and *GPR27*) has been previously described in a boy with multiple abnormalities including speech and developmental delay (Pariani, et al., 2009). Compared to the clinical findings in our patients, there is a clear overlap with regard to the speech and developmental delay suggesting that haploinsufficiency of FOXP1 is indeed causative for this phenotype. As our patients did not show any of the other clinical signs like contractures, hypertonia and blepharophimosis, present in the boy carrying the larger 3p14.1p13 deletion (Pariani, et al., 2009), these might rather be caused by the haploinsufficiency of the additionally affected genes.

FOXP1 belongs to a functionally diverse family of forkhead box (FOX) transcription factors that are all characterized by a highly conserved FOX domain. *FOX* genes have been shown to play important roles in diverse cellular functions including metabolic and developmental processes (Carlsson and Mahlapuu, 2002). While the function of FOXP1 during blood cell, lung and heart formation has already been addressed (Hu, et al., 2006; Shi, et al., 2004; Shu, et al., 2001; Wang, et al., 2004), the role of FOXP1 in neuronal processes, in particular brain development, is still unclear. In a mouse mutant model, Foxp1 has recently been shown to be an important accessory factor in Hox transcriptional output, thus regulating motor neuron diversification and connectivity to target muscles (Dasen, et al., 2008; Rousso, et al., 2008). These findings are of particular interest with regard to the gross-motor and oromotor deficits seen in our patients with *FOXP1* deletions.

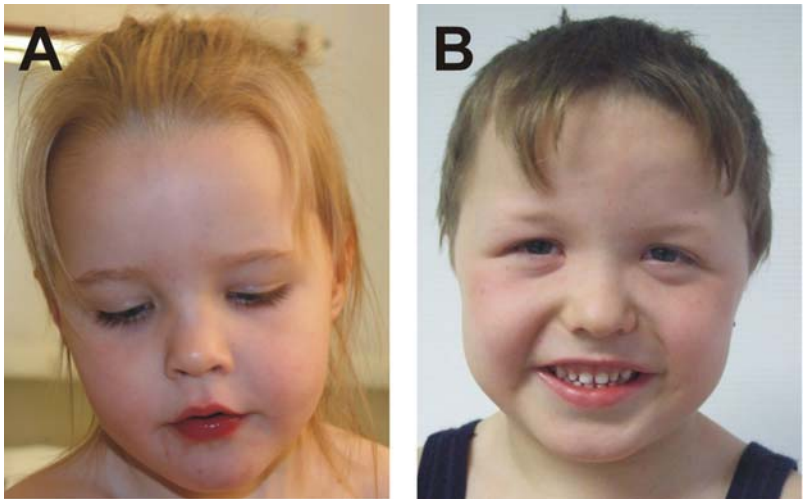


Figure 2. Facial Phenotype of patients 2 and 3. Consistent facial features of patients showing broad and prominent forehead, frontal hair upsweep, and a short nose. (A) Patient 2, at the age of 4 years and 3 months. (B) Patient 3, at the age of 6 years, had in addition sparse lateral eye brows and down-slanting palpebral fissures.

Currently, four *FOX* genes (*FOXC1* [MIM# 601090], *FOXC2* [MIM# 602402], *FOXP2* [MIM# 605317] and *FOXP3* [MIM# 300292]) are listed in the Online Mendelian Inheritance in Man (OMIM) database that have been shown to be causative for human diseases (Hannenhalli and Kaestner, 2009). With regard to structure and expression profiles, the most interesting gene related to *FOXP1* is *FOXP2*. Rare mutations of *FOXP2* have been described in individuals having expressive and receptive language and speech deficits generally described as developmental verbal dyspraxia (Hurst, et al., 1990; Lai, et al., 2001; MacDermot, et al., 2005; Marshall, et al., 2008). It is striking, that patients with *FOXP1* deletions show a comparable speech and language deficit. Most of the mutations affecting *FOXP2* are deletions and maternal uniparental disomies. In almost all cases, the affected allele is of paternal origin suggesting differential parent-of-origin expression of *FOXP2* in human speech development (Marshall, et al., 2008). Of note, there is some parallelism as the three *FOXP1* deletions we describe are also of paternal origin.

A functional relationship between Foxp1 and Foxp2 has been previously demonstrated in mouse, as they form homo- and heteromers necessary for efficient DNA binding (Marshall, et al., 2008). *FoxP1* and *FoxP2* also show overlapping expression patterns within brains of zebra finches and fetal human brains, particularly in subcortical regions that play important roles in sensorimotor integration and coordinated movements important for vocalization and speech respectively (Teramitsu, et al., 2004). From these findings it was speculated that FOXP1 might play an important role in the development of brain circuits that coordinate fine sequential motor control required for articulation.

Although the mean non-verbal IQ of the members of the large family carrying *FOXP2* mutation was lower than of the unaffected members, non-verbal intellectual impairment could not be considered characteristic of the phenotype associated with this *FOXP2* mutation (Lai, et al., 2001). This is in contrast to the clinical findings in our patients with *FOXP1* deletions who, in addition to their speech and language disabilities, exhibit significantly reduced non-verbal IQ scores. This may indicate that additional distinct functions of FOXP1 exist, presumably acting on a more global level in neuronal development compared to FOXP2. This would be in line with results from a previous study on 49 patients with developmental verbal dyspraxia and normal intelligence, where no causative *FOXP1* mutations could be detected (Vernes, et al., 2009).

In conclusion, we report a previously unknown cause of MR associated with significant speech and language disorder defined by *FOXP1* deletions. We propose that haploinsufficiency of FOXP1 leads to abnormal development of neural structures that coordinate general cognitive and psychomotoric as well as verbal abilities.

ACKNOWLEDGMENTS

We thank all patients and their families for participating in this study. We thank Dianne Newbury, Kirsty Wing and Richard Holt from the Wellcome Trust Centre in Oxford for their support, Erich Wichmann (Institute of Epidemiology, Helmholtz Zentrum München) for providing population control microarray data, and Michael Wittig for providing CNV data for the popgen and LLI studies. G.R. is a member of *CellNetworks* – Cluster of Excellence (EXC81).

This work was supported by a grant from the German Ministry for Education and Research (01GS08168). The research was conducted within the MRNET consortium. The sequencing of SLIC samples was funded by a Wellcome Trust programme grant (grant no. 076566) using core facilities funded by a Wellcome Trust core award grant (grant no. 075491). The popgen and LLI study received infrastructure support from the popgen biobank and the cluster of excellence *Inflammation at Interfaces*.

REFERENCES

Berkel S, Marshall CR, Weiss B, Howe J, Roeth R, Moog U, Endris V, Roberts W, Szatmari P, Pinto D and others. 2010. Mutations in the SHANK2 synaptic scaffolding gene in autism spectrum disorder and mental retardation. Nat Genet 42:489-491.

Carlsson P, Mahlapuu M. 2002. Forkhead transcription factors: key players in development and metabolism. Dev Biol 250:1-23.

Dasen JS, De Camilli A, Wang B, Tucker PW, Jessell TM. 2008. Hox repertoires for motor neuron diversity and connectivity gated by a single accessory factor, FoxP1. Cell 134:304-316.

de Vries BB, Pfundt R, Leisink M, Koolen DA, Vissers LE, Janssen IM, Reijmersdal S, Nillesen WM, Huys EH, Leeuw N and others. 2005. Diagnostic genome profiling in mental retardation. Am J Hum Genet 77:606-616.

Feuk L, Carson AR, Scherer SW. 2006a. Structural variation in the human genome. Nat Rev Genet 7:85-97.

Feuk L, Kalervo A, Lipsanen-Nyman M, Skaug J, Nakabayashi K, Finucane B, Hartung D, Innes M, Kerem B, Nowaczyk MJ and others. 2006b. Absence of a paternally inherited FOXP2 gene in developmental verbal dyspraxia. Am J Hum Genet 79:965-972.

Hannenhalli S, Kaestner KH. 2009. The evolution of Fox genes and their role in development and disease. Nat Rev Genet 10:233-240.

Hollox EJ, Huffmeier U, Zeeuwen PL, Palla R, Lascorz J, Rodijk-Olthuis D, van de Kerkhof PC, Traupe H, de Jongh G, den Heijer M and others. 2008. Psoriasis is associated with increased beta-defensin genomic copy number. Nat Genet 40:23-25.

Hu H, Wang B, Borde M, Nardone J, Maika S, Allred L, Tucker PW, Rao A. 2006. Foxp1 is an essential transcriptional regulator of B cell development. Nat Immunol 7:819-826.

Hurst JA, Baraitser M, Auger E, Graham F, Norell S. 1990. An extended family with a dominantly inherited speech disorder. Dev Med Child Neurol 32:352-355.

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

Inlow JK, Restifo LL. 2004. Molecular and comparative genetics of mental retardation. Genetics 166:835-881.

Lai CS, Fisher SE, Hurst JA, Vargha-Khadem F, Monaco AP. 2001. A forkhead-domain gene is mutated in a severe speech and language disorder. Nature 413:519-523.

Li S, Weidenfeld J, Morrisey EE. 2004. Transcriptional and DNA binding activity of the Foxp1/2/4 family is modulated by heterotypic and homotypic protein interactions. Mol Cell Biol 24:809-822.

Lupski JR, de Oca-Luna RM, Slaugenhaupt S, Pentao L, Guzzetta V, Trask BJ, Saucedo-Cardenas O, Barker DF, Killian JM, Garcia CA and others. 1991. DNA duplication associated with Charcot-Marie-Tooth disease type 1A. Cell 66:219-232.

MacDermot KD, Bonora E, Sykes N, Coupe AM, Lai CS, Vernes SC, Vargha-Khadem F, McKenzie F, Smith RL, Monaco AP and others. 2005. Identification of FOXP2 truncation as a novel cause of developmental speech and language deficits. Am J Hum Genet 76:1074-1080.

Marshall CR, Noor A, Vincent JB, Lionel AC, Feuk L, Skaug J, Shago M, Moessner R, Pinto D, Ren Y and others. 2008. Structural variation of chromosomes in autism spectrum disorder. Am J Hum Genet 82:477-488.

Pariani MJ, Spencer A, Graham JM, Jr., Rimoin DL. 2009. A 785kb deletion of 3p14.1p13, including the FOXP1 gene, associated with speech delay, contractures, hypertonia and blepharophimosis. Eur J Med Genet 52:123-127.

Ropers HH. 2008. Genetics of intellectual disability. Curr Opin Genet Dev 18:241-250.

Rousso DL, Gaber ZB, Wellik D, Morrisey EE, Novitch BG. 2008. Coordinated actions of the forkhead protein Foxp1 and Hox proteins in the columnar organization of spinal motor neurons. Neuron 59:226-240.

Sebat J, Lakshmi B, Malhotra D, Troge J, Lese-Martin C, Walsh T, Yamrom B, Yoon S, Krasnitz A, Kendall J and others. 2007. Strong association of de novo copy number mutations with autism. Science 316:445-449.

Shu W, Yang H, Zhang L, Lu MM, Morrisey EE. 2001. Characterization of a new subfamily of winged-helix/forkhead (Fox) genes that are expressed in the lung and act as transcriptional repressors. J Biol Chem 276:27488-27497.

Teramitsu I, Kudo LC, London SE, Geschwind DH, White SA. 2004. Parallel FoxP1 and FoxP2 expression in songbird and human brain predicts functional interaction. J Neurosci 24:3152-3163.

The SLI Consortium. 2002. A genomewide scan identifies two novel loci involved in specific language impairment. Am J Hum Genet 70:384-398.

Vernes SC, MacDermot KD, Monaco AP, Fisher SE. 2009. Assessing the impact of FOXP1 mutations on developmental verbal dyspraxia. Eur J Hum Genet 17:1354-1358.

Wagenstaller J, Spranger S, Lorenz-Depiereux B, Kazmierczak B, Nathrath M, Wahl D, Heye B, Glaser D, Liebscher V, Meitinger T and others. 2007. Copy-number variations measured by single-nucleotide-polymorphism oligonucleotide arrays in patients with mental retardation. Am J Hum Genet 81:768-779.

Wang B, Weidenfeld J, Lu MM, Maika S, Kuziel WA, Morrisey EE, Tucker PW. 2004. Foxp1 regulates cardiac outflow tract, endocardial cushion morphogenesis and myocyte proliferation and maturation. Development 131:4477-4487.

Zweier M, Gregor A, Zweier C, Engels H, Sticht H, Wohlleber E, Bijlsma EK, Holder SE, Zenker M, Rossier E and others. 2010. Mutations in MEF2C from the 5q14.3q15 microdeletion syndrome region are a frequent cause of severe mental retardation and diminish MECP2 and CDKL5 expression. Hum Mutat 31:722-733.

Supp. Table S1. Setup of molecular karyotyping analyses in patients with MR and control individuals

Cohort	Arrays	Data analysis; thresholds / settings	Individuals
Patients MRNET			
Munich	Infinium Human550-Quad and Human610-Quad (Illumina)	Data were normalized according to Wagenstaller et al. (2007). Segmentation was performed with circular binary segmentation as implemented in the R-package 'DNAcopy'.	387
Berlin	Whole-genome oligonucleotide 244K array (Agilent Technologies, Santa Clara, CA)	Feature Extraction 9.5.3.1 and CGH Analytics 3.4.40; aberration algorithm ADM-2, threshold: 6.0; window size: 0.2 Mb; filter: 5 probes; log2ratio = 0.29	188
Heidelberg	Genome-wide human SNP Array 6.0 (Affymetrix, Santa Clara, CA)	Segment Reporting Tools of the Genotyping Console 3.0 software; only CNVs > 100 kb with at least 5-10 aberrant SNPs	184
Other*			764
Sum			**1523**
Control individuals			
popgen study	Genome-wide human SNP Array 6.0 (Affymetrix, Santa Clara, CA)	CNVineta (www.ikmb.uni-kiel.de/cnvineta)	1146
KORA study	Infinium Human550-Quad (Illumina)	Data were normalized according to Wagenstaller et al. (2007). Segmentation was performed with circular binary segmentation as implemented in the R-package 'DNAcopy'.	813
AGNES study	Infinium Human6100-Quad (Illumina)		972
LUCY study	Infinium Human550-Quad (Illumina)		482
LLI study	Genome-wide human SNP Array 6.0 (Affymetrix, Santa Clara, CA)	CNVineta (www.ikmb.uni-kiel.de/cnvineta)	691
Sum			**4104**

Given are the setups of the molecular karyotyping performed in MR patients by the authors' institutions. *Additional CNV data of MR patients from other institutions (not listed here) were obtained through the database of the MRNET (www.german-mrnet.de); popgen study (www.popgen.de); KORA study (www.helmholtz-muenchen.de/kora); MR, mental retardation.

Cohen syndrome diagnosis using whole genome arrays

Nuria Rivera-Brugués,[1] Beate Albrecht,[2] Dagmar Wieczorek,[2] Heinrich Schmidt,[3] Thomas Keller,[4] Ina Göhring,[5] Arif B Ekici,[5] Andreas Tzschach,[6] Masoud Garshasbi,[6] Kathlen Franke,[7] Norman Klopp,[8] H-Erich Wichmann,[8] Thomas Meitinger,[1,9] Tim M Strom,[1,9] Maja Hempel[1,9]

[1]Institute of Human Genetics, Helmholtz Zentrum München, Neuherberg, Germany
[2]Institut für Humangenetik, Universitätsklinikum Essen, Essen, Germany
[3]Department of Pediatrics, Ludwig-Maximilians-Universität, Munich, Germany
[4]Department of Pediatrics, Josefinum Augsburg, Augsburg, Germany
[5]Institute of Human Genetics, Friedrich-Alexander-University Erlangen-Nuremberg, Erlangen, Germany
[6]Max Planck Institute for Molecular Genetics, Department Human Molecular Genetics, Berlin, Germany
[7]Private Clinic B.Prager & A. Junge, Dresden, Germany
[8]Institute of Epidemiology, Helmholtz Zentrum München, Neuherberg, Germany
[9]Institute of Human Genetics, Technische Universität München, Munich, Germany

Correspondence to
Dr Maja Hempel, Institute of Human Genetics, Technische Universität München, Trogerstraße 32, D-81675 Munich, Germany; hempel@humangenetik.med.tu-muenchen.de

Received 21 June 2010
Revised 12 August 2010
Accepted 17 August 2010
Published Online First 4 October 2010

ABSTRACT

Background Cohen syndrome is a rare autosomal recessive disorder with a complex phenotype including psychomotor retardation, microcephaly, obesity with slender extremities, joint laxity, progressive chorioretinal dystrophy/myopia, intermittent isolated neutropenia, a cheerful disposition, and characteristic facial features. The *COH1* gene, which contains 62 exons, is so far the only gene known to be associated with Cohen syndrome. Point mutations, deletions and duplications have been described in this gene. Oligonucleotide arrays have reached a resolution which allows the detection of intragenic deletions and duplications, especially in large genes such as *COH1*.

Method and results High density oligonucleotide array data from patients with unexplained mental retardation (n=1523) and normal controls (n=1612) were analysed for copy number variation (CNV) changes. Intragenic heterozygous deletions in the *COH1* gene were detected in three patients but no such changes were detected in the controls. Subsequent sequencing of the *COH1* gene revealed point mutations in the second allele in all three patients analysed.

Conclusion Genome-wide CNV screening with high density arrays provides a tool to detect intragenic deletions in the *COH1* gene. This report presents an example of how microarrays can be used to identify autosomal recessive syndromes and to extend the phenotypic and mutational spectrum of recessive disorders.

INTRODUCTION

The phenotype of Cohen syndrome (MIM 216550), a rare autosomal recessive disorder, has been described to be fairly homogeneous in Finnish patients where the founder mutation c.3348_3349delCT is detected in about 75% of mutant alleles.[1] But in non-Finnish and especially in young Cohen patients, a high genotypic and phenotypic variability occurs. Several clinical diagnostic criteria for Cohen syndrome have been introduced.[2–5] Chandler *et al* proposed that, next to significant learning disabilities, two of the following criteria should be present for Cohen syndrome diagnosis: facial gestalt, pigmentary retinopathy, and neutropenia.[3] Kohlemainen *et al* suggested Cohen syndrome in patients fulfilling at least six of the following criteria: developmental delay, microcephaly, typical facial gestalt, truncal obesity with slender extremities, overly sociable behaviour, joint hypermobility, high myopia and/or retinal dystrophy, and neutropenia.[5] El Chehadeh *et al* concluded that mutation analysis is not indicated in the absence of chorioretinal dystrophy or neutropenia.[6]

The *COH1* gene (VPS13B, MIM 607817), so far the only gene known to be associated with Cohen syndrome, is one of the largest known genes in the human genome and comprises 62 exons distributed along 864315 bp of chromosome 8. More than 96 different mutations in *COH1* gene have been detected in association with Cohen syndrome. The majority of them are terminating mutations predicted to result in a functional null allele.[5 7–9] Recently large intragenic deletions and duplications in the *COH1* gene have been identified as a cause of Cohen syndrome.[10 11] Molecular diagnosis of syndromes has been improved by the introduction of oligonucleotide arrays which have reached a resolution facilitating the detection of intragenic copy number variations (CNVs).

ARRAY DATASETS AND METHODS

Array datasets

Recruitment of patients (n=1523) had been part of the German Mental Retardation Network (MRNET) study (http://www.german-mrnet.de/). In the present study we focus on three patients, in which CNVs in the *COH1* gene were detected. As a control set, CNV data (n=1612) had been generated from the population based KORA study (Cooperative Health Research in the Augsburg Region).

CNV analysis

Genome-wide screening for CNVs was performed using the Infinium Human550 Genotyping BeadChip (Illumina, SanDiego, California, USA) in patients 1 and 2 and the Genome-Wide Human SNP Array 6.0 (Affymetrix, Santa Clara, California, USA) in patient 3. The controls have been investigated with Illumina Infinium Human550-Quad or Human610-Quad arrays.

The genomic DNA of the patients and controls was isolated from peripheral blood lymphocytes according to standard procedures and processed following the manufacturer's instructions. Arrays were scanned with the Illumina BeadArray Reader and with the Affymetrix Scanner 3000 7G. Genotypes were called with the Illumina GenomeStudio Software or using the Affymetrix Genotyping Console Software (version 3.0.2), respectively.

Data analysis of the Illumina arrays was performed according to Wagenstaller *et al*.[12] CNV profiling of the Affymetrix array data was accomplished by using the Segment reporting Tools of the Genotyping Console Software. To determine a deletion we used as a cut-off the smoothing median of five or more adjacent single nucleotide polymorphisms (SNPs) with copy number values ≤1.5 or with log2 intensity ratios ≤−1 for Illumina and Affymetrix arrays, respectively. CNVs with copy number values ≥2.5 or with log2 intensity ratios ≥1 were suspicious for duplication. All CNVs were checked for gene content and overlap with known genetic variants as provided by the genome browser of the University of California Santa Cruz (UCSC) (http://genome.ucsc.edu, hg 18) and the Database of Genomic Variants (DGV, http://projects.tcag.ca/variation). CNVs not annotated as structural polymorphisms and containing RefSeq genes were genotyped by quantitative PCR (qPCR). Monitoring of the PCR reaction and setting of baseline and threshold cycle values were accomplished automatically with the Sequence Detection System Version 2.3 Software (SDS 2.3, Applied Biosystems, Darmstadt, Germany). The relative quantification analysis based on the comparative C_t method was performed using an in-house developed Perl script.

Sequencing

In the patients with a partial heterozygous deletion of the *COH1* gene, direct sequencing of the entire coding region and the exon/intron boundaries of *COH1* was carried out using BigDye Ready Terminator Sequencing Kit and an 48 capillary Abi 3730 Genetic Analyzer (Applied Biosystems) in accordance with standard procedures. All identified variants were genotyped in 676 individuals of a population based cohort (KORA-cohort) via the MassARRAY system (Sequenom genotyping platform) and the iPLEX Gold chemistry. The assay design used the AssayDesign 3.1.2.2 software with default parameters. Genotype calling was performed by the SpectroTYPER 3.4 software.

Nomenclature

Gene model NM_17890.3/NP_060360 based on UCSC browser was used to describe the detected *COH1* gene variants (http://genome.ucsc.edu hg18).

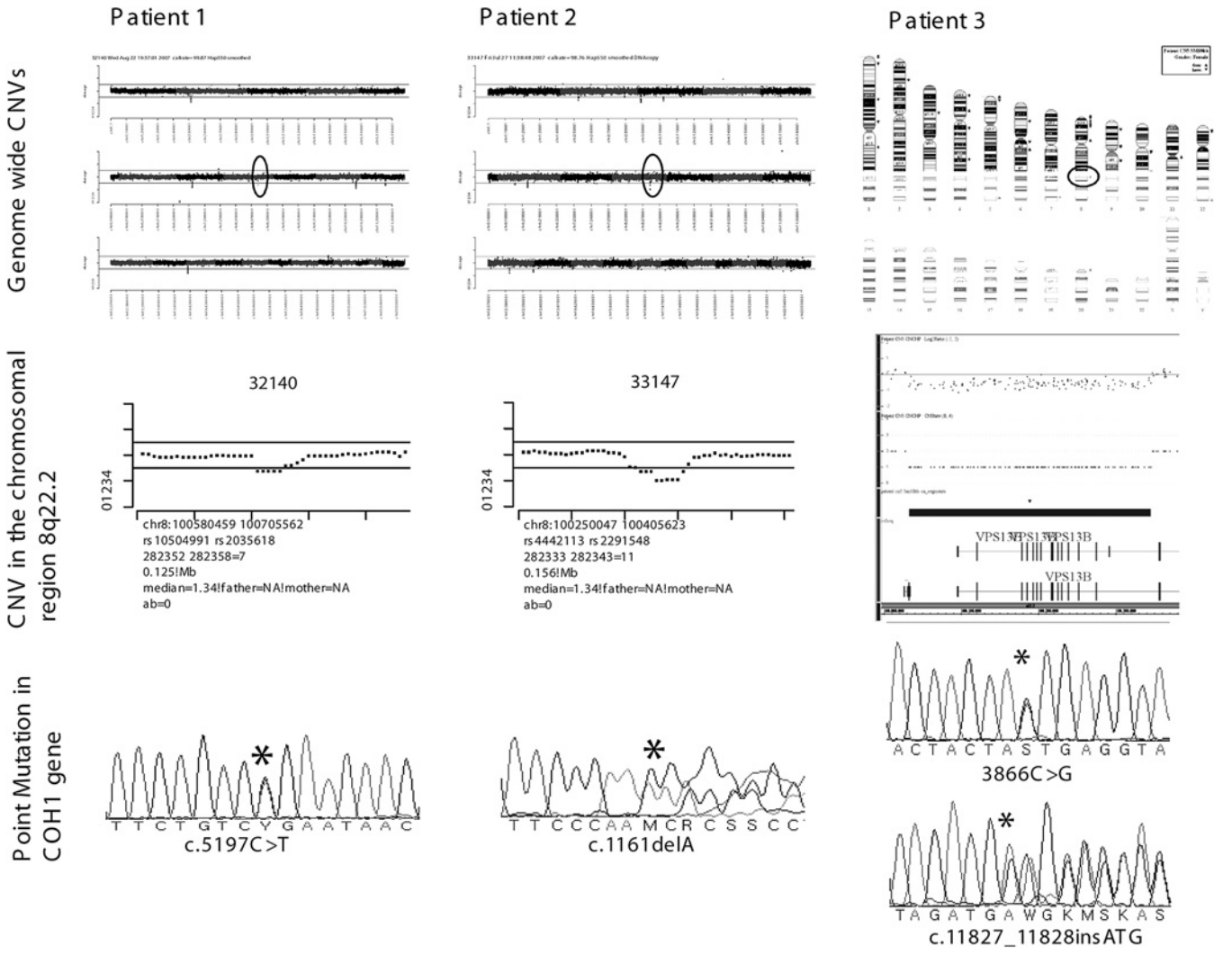


Figure 1 Array results and electropherograms of the *COH1* mutations in patients 1 to 3. Patient 1 is affected by a 125 kb deletion encompassing exons 26 to 31 of the *COH1* gene (chr.8: 100 563 167 …100 642 020) and a missense mutation c.5197C→T of the second allele. In Patient 2 a 156 kb deletion encompassing exons 16 to 19 of the *COH1* gene (chr.8:100 250 047…100 405 623) and a one base pair deletion c.1161delA in the second allele was detected. In Patient 3 a 315 kb deletion encompassing exons 1 to 17 of the *COH1* gene (chr8:100 015 029…100 347 846) and missense mutation 3866C→G and three base pair insertion c.11827_11828insATG in the second allele were found. CNV, copy number variation.

Figure 2 Craniofacial phenotype of patients 1 to 3. Patient 1 at the age of 3 years with horizontal eyebrows, a broad and down turned nasal tip, a broad columella, a short philtrum, a small upper lip and everted lower lip. Patient 2 at the age of 18 months with horizontal eyebrows, almond shaped and downslanting palpebral fissures, a broad nasal root, a round nasal tip, thick columella, a short philtrum, and open appearance of mouth with prominent upper gingiva. Patient 3 at the age of $2^{8}/_{12}$ years with round and flat face, bushy eyebrows with lateral flaring, broad nasal bridge, short philtrum and microtia with overfolded helices.

RESULTS

Molecular findings

Analysis of SNP oligonucleotide array and subsequent qPCR discovered CNVs in the *COH1* gene in three patients: a maternal 67 kb deletion encompassing exons 26 to 31 of the *COH1* gene (chr8: 100 573 090…100 639 924; c.3871-5024del/p.G1291fsX42) in patient 1, a paternal 193 kb deletion encompassing exons 9 to 19 of the *COH1* gene (chr8: 100 216 034…100 409 167; c.1207-2824del/p.L403fsX11) in patient 2, and a maternal 315 kb deletion encompassing exon 4 of the *OSR2* gene and exons 1 to 17 of the *COH1* gene (chr8:100 015 029…100 347 846; c.1-2515del) in patient 3 (figure 1). There were no such changes in the 1612 controls.

Sequencing of the *COH1* gene identified in patient 1 a paternal missense mutation in exon 32 leading to a stop codon (c.5086C→T/p.R1696X), in patient 2 a maternal 1 bp deletion in exon 60 leading to a stop codon (c.11505delA/p.K3835fsX43), and in patient 3 a heterozygous missense mutation in exon 25 (c.3866C→G/p.T1289S) and a heterozygous three base pairs insertion in exon 62 (c.11827_11828insATG/p.D3942_G3943insD), both inherited from the father (figure 1). These mutations were not present in 676 samples from a general population cohort (KORA) and not annotated as polymorphisms in dbSNP (NCBI, http://www.ncbi.nlm.nih.gov/SNP).

Clinical data

Patient 1

The boy was born after an uneventful pregnancy at term as the first child of unrelated and healthy Arabian parents. Birth weight, length and head circumference were not recorded. A delay in motor development became evident within his first year of life. Sitting started at the age of 18 months, walking at 24 months. A detailed examination at the age of 3 years showed a hypotonic boy with a height of 79 cm (−4.8 SD), a weight of 9300 g (−5.2 SD), and a head circumference of 44 cm (−4.9 SD). A delay in speech development as well as in comprehension was obvious. There was mild craniofacial dysmorphism including horizontal eyebrows, a broad and downturned nasal tip, a broad columella, a short philtrum, a thin upper lip, and an everted lower lip (figure 2). The ophthalmologic examination revealed bilateral myopia, astigmatism and a slightly increased pigmentation of the retina. There was no neutropenia.

Patient 2

This boy was born at term after an uneventful pregnancy as the second child of healthy unrelated German parents, with a birth weight of 2460 g (−2.4 SD), a birth length of 46.5 cm (−2.3 SD), and a head circumference of 31.5 cm (−2.9 SD). Developmental milestones were delayed with a sitting age of 12 months and a crawling age of 17 months. Examination at the age of

18 months showed a hypotonic toddler with a weight of 10.1 kg (−1.5 SD), a height of 80 cm (−1 SD), and a severe microcephaly with an occipitofrontal circumference (OFC) of 42.5 cm (−4.8 SD). Speech development had not occurred but comprehension was nearly normal. Facial dysmorphism consisted of mild occipital flattening, horizontal eyebrows, almond shaped palpebral fissures, a broad nasal root, a round nasal tip, a thick columella, a short philtrum, an open appearance of the mouth with a prominent upper gingiva, and a large gap between the incisors (figure 2). Funduscopy and complete blood count revealed no abnormalities.

Patient 3

The patient is the second child of a healthy non-consanguineous German/African couple. The premature birth occurred at 35 weeks of gestation with a birth weight of 2450 g (−0.1 SD), a birth length of 46 cm (−0.4 SD), and an OFC of 33.5 cm (0.6 SD). A heart defect (atrioseptal defect (ASD) II and pulmonary stenosis) and arrhythmia were diagnosed after birth. Developmental delay persisted until 2 years of age and improved after surgical correction of the heart defect. The patient started to walk at the age of 24 months and at the age of $2^3/_4$ years she showed normal body measurement with a height of 95 cm (0.4 SD), a weight of 11.5 kg (−1.4 SD), and an OFC 50 cm (0.6 SD). She had a flat face with broad and flat nasal bridge and almond shaped eyes, a short philtrum with thin vermillion border, and deep set ears with overfolded helices (figure 2). Funduscopy and complete blood count revealed normal results.

DISCUSSION

Our report demonstrates that the diagnosis of Cohen syndrome can be reached in patients with unexplained mental retardation by applying high resolution oligonucleotide arrays. Although multiple exon deletions in the *COH1* gene have been reported in single patients,[5 7 8 13] the distribution of deletions and duplications has only been recognised recently.[10 11] The frequency of copy number alterations in the *COH1* gene is unknown. Parri *et al* disclosed that *COH1* CNVs account for 42% of *COH1* mutations.[11] Balikova *et al* reported an increase in the detection rate of 18% (88% instead of 70%) in typical Cohen patients.[10] Arrays targeted at individual exons will further increase the detection rate. We have recently identified—by molecular analysis—a homozygous 66 kb deletion comprising exons 32 and 33 of *COH1* which had escaped detection by current array comparative genomic hybridisation (CGH) analysis. The fact that we failed to detect CNVs in the *COH1* gene in 1612 controls from a general German population cohort indicates that CNVs in the *COH1* gene are rare. This is in contrast to the annotation of the *COH1* gene CNVs as benign polymorphisms in the UCSC genome browser and the DGV.

To our knowledge this is the first report on patients with Cohen syndrome diagnosed by molecular whole genome analyses but not by clinical examination. Cohen syndrome was not suspected at first instance in the patients, although all of them were examined by experienced paediatricians or clinical geneticists. This can be explained by the young age of patients (16 months, 18 months, $2^3/_4$ years). The facial appearance of the infant and young children with Cohen syndrome differs from adult patients and myopia/retinal pigmentary changes usually develop in the pre-school age.[3] The common facial characteristics in our small series of young patients with Cohen syndrome were a hypotonic facial expression, almond shaped palpebral fissures, a prominent nose, and a short philtrum. All patients were affected by mental retardation and delay in motor and speech development. Pigmentary retinopathy and neutropenia were absent in all patients. The unusual phenotype in patient 3 may due to the fact that in addition to exons 1 to 17 of the *COH1* gene, the deletion affected the neighbouring exon 4 of the *ORS2* gene. Although the gene function of *ORS2* in humans is unknown, an influence of the phenotype cannot be ruled out.

In conclusion, the phenotype of Cohen syndrome defined by *COH1* mutations is fairly unspecific, particularly in very young patients but in older children too. In addition, deletions in neighbouring genes may affect the phenotype of Cohen syndrome in the context of a contiguous gene syndrome. Nevertheless, young patients with a hypotonic facial expression, almond shaped eyes, short philtrum, mental retardation, and motor and speech delay are suspicious for Cohen syndrome. Microarrays have the potential to diagnose Cohen syndrome in very young patients and in patients with an atypical phenotype.

Acknowledgements We thank all patients and their families for participating in this study. We thank Peter Lichtner, Institute of Human Genetics, Helmholtz Zentrum München for genotyping the *COH* variants in controls and Jürgen Kohlhase, Center for Human Genetics, Freiburg for *CHD7* mutation analysis in patient 3. We also thank Monika Hartig, Ilona Dugdale and Lawrence Haw for critical revision of the manuscript and editorial assistance. This work was supported by a grant from the German Ministry for Education and Research (NGFNplus/http://www.ngfn.de/englisch/15.htm, project reference numbers 01GS08160, 01GS08161, 01GS08163 and 01GS08167). The research was conducted within the MRNET consortium (http://www.german-mrnet.de/).

Funding NGFN Geschäftsstelle; c/o Deutsches Krebsforschungszentrum - DKFZIm Neuenheimer Feld 580, V025; 69120 Heidelberg. Other funders: BMBF, Germany.

Competing interests None.

Patient consent Obtained.

Ethics approval Approval for the study had been obtained by the ethical review boards of the participating institutions.

Provenance and peer review Not commissioned; externally peer reviewed.

REFERENCES

1. **Kolehmainen J,** Black GC, Saarinen A, Chandler K, Clayton-Smith J, Träskelin AL, Perveen R, Kivitie-Kallio S, Norio R, Warburg M, Fryns JP, de la Chapelle A, Lehesjoki AE. Cohen syndrome is caused by mutations in a novel gene, COH1, encoding a transmembrane protein with a presumed role in vesicle-mediated sorting and intracellular protein transport. *Am J Hum Genet* 2003;**72**:1359–69.
2. **Horn D,** Krebsova A, Kunze J, Reis A. Homozygosity mapping in a family with microcephaly, mental retardation, and short stature to a Cohen syndrome region on 8q21.3-8q22.1: redefining a clinical entity. *Am J Med Genet* 2000;**92**:285–92.
3. **Chandler KE,** Kidd A, Al-Gazali L, Kolehmainen J, Lehesjoki AE, Black GC, Clayton-Smith J. Diagnostic criteria, clinical characteristics, and natural history of Cohen syndrome. *J Med Genet* 2003;**40**:233–41.
4. **Mochida GH,** Rajab A, Eyaid W, Lu A, Al-Nouri D, Kosaki K, Noruzinia M, Sarda P, Ishihara J, Bodell A, Apse K, Walsh CA. Broader geographical spectrum of Cohen syndrome due to COH1 mutations. *J Med Genet* 2004;**41**:e87.
5. **Kolehmainen J,** Wilkinson R, Lehesjoki AE, Chandler K, Kivitie-Kallio S, Clayton-Smith J, Träskelin AL, Waris L, Saarinen A, Khan J, Gross-Tsur V, Traboulsi EI, Warburg M, Fryns JP, Norio R, Black GC, Manson FD. Delineation of Cohen syndrome following a large-scale genotype-phenotype screen. *Am J Hum Genet* 2004;**75**:122–7.
6. **El Chehadeh S,** Aral B, Gigot N, Thauvin-Robinet C, Donzel A, Delrue MA, Lacombe D, David A, Burglen L, Philip N, Moncla A, Cormier-Daire V, Rio M, Edery P, Verloes A, Bonneau D, Afenjar A, Jacquette A, Heron D, Sarda P, Pinson L, Doray B, Vigneron J, Leheup B, Frances-Guidet AM, Dienne G, Holder M, Masurel-Paulet A, Huet F, Teyssier JR, Faivre L. Search for the best indicators for the presence of a VPS13B gene mutation and confirmation of diagnostic criteria in a series of 34 patients genotyped for suspected Cohen syndrome. *J Med Genet* 2010;**47**:549–53.
7. **Seifert W,** Holder-Espinasse M, Spranger S, Hoeltzenbein M, Rossier E, Dollfus H, Lacombe D, Verloes A, Chrzanowska KH, Maegawa GH, Chitayat D, Kotzot D, Huhle D, Meinecke P, Albrecht B, Mathijssen I, Leheup B, Raile K, Hennies HC, Horn D. Mutational spectrum of COH1 and clinical heterogeneity in Cohen syndrome. *J Med Genet* 2006;**43**:e22.

8. **Seifert W,** Holder-Espinasse M, Kuhnisch J, Kahrizi K, Tzschach A, Garshasbi M, Najmabadi H, Walter Kuss A, Kress W, Laureys G, Loeys B, Brilstra E, Mancini GM, Dollfus H, Dahan K, Apse K, Hennies HC, Horn D. Expanded mutational spectrum in Cohen syndrome, tissue expression, and transcript variants of COH1. *Hum Mutat* 2009;**30**:E404–20.
9. **Hennies HC,** Rauch A, Seifert W, Schumi C, Moser E, Al-Taji E, Tariverdian G, Chrzanowska KH, Krajewska-Walasek M, Rajab A, Giugliani R, Neumann TE, Eckl KM, Karbasiyan M, Reis A, Horn D. Allelic heterogeneity in the COH1 gene explains clinical variability in Cohen syndrome. *Am J Hum Genet* 2004;**75**:138–45.
10. **Balikova I,** Lehesjoki AE, de Ravel TJ, Thienpont B, Chandler KE, Clayton-Smith J, Träskelin AL, Fryns JP, Vermeesch JR. Deletions in the VPS13B (COH1) gene as a cause of Cohen syndrome. *Hum Mutat* 2009;**30**:E845–54.
11. **Parri V,** Katzaki E, Uliana V, Scionti F, Tita R, Artuso R, Longo I, Boschloo R, Vijzelaar R, Selicorni A, Brancati F, Dallapiccola B, Zelante L, Hamel CP, Sarda P, Lalani SR, Grasso R, Buoni S, Hayek J, Servais L, de Vries BB, Georgoudi N, Nakou S, Petersen MB, Mari F, Renieri A, Ariani F. High frequency of COH1 intragenic deletions and duplications detected by MLPA in patients with Cohen syndrome. *Eur J Hum Genet* 2010;**10**:1133–40.
12. **Wagenstaller J,** Spranger S, Lorenz-Depiereux B, Kazmierczak B, Nathrath M, Wahl D, Heye B, Glaser D, Liebscher V, Meitinger T, Strom TM. Copy-number variations measured by single-nucleotide-polymorphism oligonucleotide arrays in patients with mental retardation. *Am J Hum Genet* 2007;**81**:768–79.
13. **Taban M,** Memoracion-Peralta DS, Wang H, Al-Gazali LI, Traboulsi EI. Cohen syndrome: report of nine cases and review of the literature, with emphasis on ophthalmic features. *J AAPOS* 2007;**11**:431–7.

Poetry

The battle of replication fork

Small soldiers march swiftly, the battle draws near
Helicase from the cannons split legions, no fear
Commanders ride forth with bellowing calls
Sending signals to primers, engage, pair them all!
Deoxy stands back, a path must be made
Those RNA boys bear the brunt of the trade
At last help's arrived; they've come through the trees
3-primed and ready, polymerase threes
RNA clears the field, a job nicely done
Their stations soon filled by polymerase one
"Who leads?" cries the general, he asks for a name
It's 3-prime not 5, the phosphates to blame
Fragmented and lagging are the men from Japan
Send ligase post haste, it's part of the plan
The fighting drones on, with no end in sight
Gyrase eases tensions, gets men through the night
At last the sun rises, dust settles, all clear
Polymerase checks that the win is sincere
Not a moment for rest, they're worked to the ground
For over those hills, more ori abound

Victor Laurion

Correspondence to Victor Laurion, Bryn Mawr College, 900 Montgomery Avenue, Bryn Mawr, PA 19010, USA; victor.laurion@gmail.com

Competing interests None declared.

Provenance and peer review Not commissioned; not externally peer reviewed.

Received 20 October 2010
Accepted 16 November 2010
Published Online First 22 December 2010

J Med Genet 2011;**48**:140. doi:10.1136/jmg.2010.086553

RESEARCH ARTICLE

AMERICAN JOURNAL OF medical genetics PART A

Microdeletion Syndrome 16p11.2-p12.2: Clinical and Molecular Characterization

Maja Hempel,[1,2]* Nuria Rivera Brugués,[1,2] Janine Wagenstaller,[1,2] Gaby Lederer,[1] Andrea Weitensteiner,[3] Heide Seidel,[1,4] Thomas Meitinger,[1,2] and Tim M. Strom[1,2]

[1]Institute of Human Genetics, Technische Universität München, Munich, Germany

[2]Institute of Human Genetics, Helmholtz Zentrum München, German Research Center for Environmental Health, Neuherberg, Germany

[3]Department of Pediatrics, Technische Universität München, Munich, Germany

[4]Institute of Human Genetics, Ludwig-Maximilians-Universität München, Munich, Germany

Received 5 March 2009; Accepted 13 July 2009

The pericentromeric region on 16p appears to be susceptible to chromosomal rearrangements and several patients with rearrangements in this region have been described. We report on a further patient with a microdeletion 16p11.2-p12.2 in the context of described patients with a deletion in the pericentromeric region of 16p. Minor facial anomalies, feeding difficulties, significant delay in speech development, and recurrent ear infections are common symptoms of the microdeletion syndrome 16p11.2-p12.2. All reported patients so far share a common distal breakpoint at 16p12.2 but vary in the proximal breakpoint at 16p11.2. The microdeletion 16p11.2-p12.2 should be distinguished from the ~500 kb microdeletion in 16p11.2 which seems to be associated with autism but not with facial manifestations, feeding difficulties, or developmental delay. © 2009 Wiley-Liss, Inc.

Key words: microdeletion; 16p11.2-p12.2; facial manifestation; feeding problems; speech delay; ear infection; SNP oligonucleotide array

How to Cite this Article:
Hempel M, Rivera Brugués N, Wagenstaller J, Lederer G, Weitensteiner A, Seidel H, Meitinger T, Strom TM. 2009. Microdeletion syndrome 16p11.2-p12.2: Clinical and molecular characterization.
Am J Med Genet Part A 149A:2106–2112.

INTRODUCTION

Since the implementation of array analysis, numerous new microdeletion syndromes have been described [Potocki et al., 2000; Koolen et al., 2006; Slavotinek, 2008]. Their identification and clinical description is most relevant for the clinical approach, counseling, and the management of patients with developmental disabilities. One of these newly discovered syndromic entities is the recurring microdeletion 16p11.2-p12.2.

An interstitial deletion 16p11.2 was first described by Hernando et al. [2002] in a male newborn. This deletion was visible on chromosome analysis and was confirmed by CGH. The boy showed growth retardation, congenital malformations, and facial dysmorphisms and died at the age of 5 months. Ballif et al. [2007] reported on four patients with microdeletions related to the pericentric region 16p11.2-p12.2. Despite different lengths of the microdeletions, the authors defined common characteristics in these patients, such as distinctive facial features, orofacial clefting, heart defects, frequent ear infections, short stature, feeding difficulties, hypotonia, and development delay. A further patient with a microdeletion 16p11.2-p12.2 has been recently described by Battaglia et al. [2009]. This girl showed the clinical symptoms similar to the patients reported by Ballif et al. [2007]. In this article, we discuss the findings and the development of another patient with a microdeletion 16p11.2-p12.2 in relation to the described patients.

CLINICAL REPORT

The boy we report on was born with a low birth weight (2,210 g) and a low birth length (49 cm) after 38 weeks of an uneventful pregnancy as the third child of healthy parents. The family history was unremarkable. A glandular hypospadias and hydrocele were observed soon after birth. The postnatal period was complicated by hypotonia and poor sucking. Retardation in motor development was recognized at the age of 6 months. Crawling started at 18 months, and walking at 23 months. Gross motor abilities

Grant sponsor: NGFN; Grant Number: 01GS08163.
*Correspondence to:
Maja Hempel, M.D., Institute of Human Genetics, Technische Universität München, Trogerstraße 32, D-81675 Munich, Germany.
E-mail: hempel@humangenetik.med.tu-muenchen.de
Published online 12 August 2009 in Wiley InterScience (www.interscience.wiley.com)
DOI 10.1002/ajmg.a.33042

improved over the years, but deficits in the fine motor skills were present. Hypotonia as well as a poor facial expression and an open mouth appearance with salivation were mentioned in all neurological investigations.

The most profound delay was noted in speech development. No phonation was observed at the age of 14 months. At 3 years of age there was still no speech development. The boy imitated noises but was not able to say a word.

Initial genetic examination occurred at the age of 3 years and showed subtle dysmorphic features: a round and flat face, a high and broad forehead, deep set eyes, sparse medial eyebrows, a small mouth, overfolded helix, and clinodactyly on both hands (Fig. 1). Further investigations showed mild mixed hearing loss, which could not completely explain the severe speech delay. Consequently hearing aids were applied. Although the speech development did not essentially improve, better comprehension occurred and facilitated communication between the boy and his parents. In the following years he had recurring ear infections. Repeated operations were necessary to treat the recurring cholesteatomas and retracted tympanic membrane. Feeding problems continued during the whole infancy until present. Poor sucking and problems in chewing and swallowing did not improve significantly over time.

FIG. 1. Facial appearance at the age of 3 years showed a round and flat face, a high and broad forehead, deep set eyes, sparse medial eyebrows, a small mouth, and overfolded helix. [Color figure can be viewed in the online issue, which is available at www.interscience.wiley.com.]

Re-examination at the age of 13 years revealed mild dysmorphic features: a long flat face with frontal upsweep of the hair, deep set eyes, a long nose with bulbous nasal tip, a short and smooth philtrum, a small mouth, and clinodactyly (Fig. 2). Height, body weight, and head circumference were within the normal range. Deficits in the fine motor skills such as writing and drawing were substantial. Testing showed moderate mental retardation. He developed a few self-help skills and was able to perform processes demanding three to four successive steps. Expressive language was severely impaired; he spoke only a few words and interacted with his parents using sounds and his own sign language. The boy also showed an introverted behavior. But in comparison to patients with autism he demonstrated multiple nonverbal behaviors. In addition

FIG. 2. Facial appearance at the age of 13 years with mild dysmorphic features: a long flat face with frontal upsweep of the hair, deep set eyes, a long nose with bulbous nasal tip, a short and smooth philtrum and a small mouth. [Color figure can be viewed in the online issue, which is available at www.interscience.wiley.com.]

he developed peer relationships, he sought enjoyment, and he initiated communication. He did not show restricted repetitive and stereotyped patterns of behavior. Using the DSM-IV-TR Diagnostic Criteria for Autistic Disorders, autism could be excluded in this patient. Further investigations including EEG, ECG, and vision testing were normal. Ultrasound of abdomen showed kidneys of small size (left volume of 41 ml, right volume of 43 ml).

METHODS

Chromosome Analysis and FISH

Chromosome analysis by GTG banding was performed in the patient and both his parents in 1996. Furthermore fluorescence in situ hybridization (FISH) using telomeric probes for 16p (RT100, 16p13.3, Rubinstein-Taybi-region) and 16q (Oncor P5432, 16q24) as well as chromosome 16 painting (WCP 16) was performed to refine the chromosome aberration on chromosome 16.

SNP Array

For further delineation of the chromosomal rearrangement on 16p, SNP oligonucleotide array analysis was performed using the Sentrix® HumanHap 550 Genotyping BeadChip (Illumina®). The array contains ~550,000 markers, majority of which are tag SNPs selected from the International HapMap Project. Additional SNPs ensure an even spacing across the genome. On average, there is one SNP every 6 kb. Genomic DNA of the patient and his parents was isolated from peripheral blood lymphocytes by salting-out procedure and processed according to the manufacturer's instructions (for detailed description see Infinium II Assay HC Two-Sample, Manual, Experienced User Card, Illumina). The Chips were scanned with the Illumina BeadArray™ Reader. Results were called with the Genotyping Module of BeadStudio® using the default call threshold of 0.15. Data analysis was performed as previously

described by Wagenstaller et al. [2007]. CNVs with intensity values lower than 1.5 expanding more than five consecutive SNPs were suspicious for deletion, CNVs with intensity values of more than 2.5 and expanding more than five consecutive SNPs for duplication. For all CNVs suspected to be involved in deletions or duplications, the following information was recorded: start and end SNP, start and end position on the chromosome, total length of the chromosomal aberration, number of included genes using the Ensembl (http://www.ensembl.org), and the University of California Santa Cruz (UCSC) genome browser (http://genome.ucsc.edu, hg 18) (compare Table I). Overlap with known genetic variants was checked by using the Database of Genomic Variants (DGV, http://projects.tcag.ca/variation).

If a detected CNV could not be assigned to a known genetic variant, 1–3 amplicons were designed for validation by quantitative real-time PCR (qPCR). qPCR was performed on a 7900HT real-time PCR system (Applied Biosystems, Darmstadt, Germany) as described in detail by Wagenstaller et al. [2007].

Segmental duplications located in the chromosome region 16p11.2 and 16p12.2 were identified using the UCSC human genome Browser (assembly March 2006, http://genome.ucsc.edu, hg18) and the Human Genome Segmental Duplication Database (http://projects.tcag.ca/humandup).

RESULTS

Chromosome analysis by GTG banding at a level of 450 bands demonstrated an aberration on chromosome 16p with the karyotype: 46,XY,?del(16)(p) (Fig. 3). Both parents had a normal karyotype. Rearrangements of telomeric regions of chromosome 16 could be excluded by normal FISH using telomeric probes 16p13.3 and 16q24. Chromosome 16 painting showed a complete painting of derivative 16 but no material from chromosome 16 on other chromosomes.

SNP oligonucleotide array analysis confirmed an aberration in the pericentric region on the short arm of chromosome 16. We defined a deletion involving the chromosome region 16p11.2–16p12.2, which encompassed 7.7 Mb and contained 1,373 SNPs (CNV no. 10 in Table I). The deletion was confirmed by quantitative PCR. Since qPCR results in the parents were normal, the deletion occurred de novo. The distal breakpoint of the deletion has been determined 21.5 Mb from the 16p telomere in 16p12.2, the proximal breakpoint 29.2 Mb from the 16p telomere in 16p11.2. Both breakpoints were flanked by multiple clusters of segmental duplications.

One hundred sixty-three Ensembl genes are annotated in the deleted region, 70 genes are listed in OMIM and 11 genes are classified as OMIM disease genes (*OTOA*, *SCNN1G*, *SCNN1B*, *COG7*, *PALB2*, *IL4R*, *IL21R*, *CLN3*, *TUFM*, *ATP2A1*, *CD19*). According to the DGV database there is a negligible overlap with known CNPs (0.4%).

Eleven further CNVs were detected in the patient. Nine of these have been described in the Database of Genomic Variance as known CNVs without pathological relevance. One of the two remaining CNVs does not include known genes; the other CNV was inherited from the mother (see Table I).

TABLE I. CNV of the Present Patient Detected by SNP Oligonucleotide Microarray Analysis

CNV no.	CNV chromosome position	SNP.start	SNP.end	Length of CNV (bp)	Count of SNP in CNV	Median of SNP copies in CNV	Ensembl genes	Confirmed by qPCR	Inherited	Overlap with known CNV (%)
1	2:41092148–41099005	rs12617846	rs2373974	6,857	9	1.21	0	NA	NA	100
2	2:50817046–50850456	rs4971692	rs3892750	33,410	16	1.32	0	Yes	Maternal	100
3	3:8808878–8828715	rs237852	rs1386976	19,837	7	2.51	0	No	NA	100
4	4:463935–492911	rs17721028	rs17164861	28,976	7	2.58	3	Yes	Maternal	0
5	5:154891408–154903174	rs17117886	rs1296135	11,766	5	2.56	0	NA	NA	0
6	6:79029920–79088461	rs818249	rs818280	58,541	24	0.00	0	NA	NA	100
8	11:122977757–123005969	rs12291891	rs2027767	28,212	10	2.44	3	NA	NA	100
9	11:55127597–55201444	rs2456022	rs295635	73,847	10	1.40	4	NA	NA	100
10	**16:21512681–29223380**	**rs194545**	**rs7184953**	**7,710,699**	**1373**	**1.37**	**163**	**Yes**	**De novo**	**0.4**
11	18:6539372–6536292	rs12958013	rs7230140	3.554	6	1.52	1	Yes	Maternal	100
12	19:2042200–20473895	rs10408291	rs2021399	51,695	9	1.35	1	NA	NA	100
13	23:112788193–113114824	rs6643336	rs7058109	326,631	28	0.02	5	Yes	Maternal	80

CNV no. 10 indicates the microdeletion 16p11.2-p12.2 (bold row); NA, not analyzed.

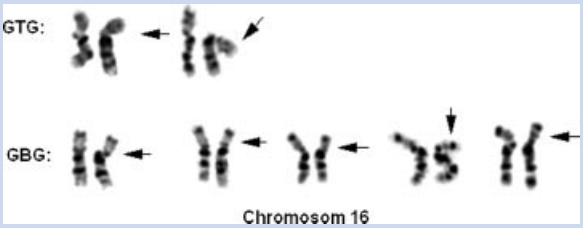


FIG. 3. Partial G-banded karyotypes of the homologous chromosomes 16 of the patient. The aberration on 16p is indicated by an arrow.

DISCUSSION

In 1996 we had detected an aberration on the short arm of chromosome 16 in a patient with mental retardation, severe speech delay, and recurring ear infections. But the chromosome aberration could not be refined at this time. This became possible by the implementation of array analyses. Using a SNP oligonucleotide array we were able to define the aberration as a microdeletion 16p11.2-p12.2 of 7.7 Mb.

Ballif et al. [2007] detected a microdeletion 16p11.2-p12.2 in four patients with using CGH array including a minimum of three BAC clones at the most proximal end of the pericentromeric region of each chromosome. Microdeletions extended to ~7.1, ~7.9, and ~8.7 Mb. Battaglia et al. [2009] reported a further patient with microdeletion 16p11.2-p12.2 identified by array-CHG analysis. This deletion spanned ~8.2 Mb. In all so far characterized patients with microdeletion 16p11.2-p12.2 the distal breakpoint was delineated ~21.4 Mb from the 16p telomere in chromosomal region 16p12.2. Although the distal breakpoint in the patient reported here was estimated to be ~21.5 Mb from the 16p telomere, we assume that the microdeletion in our patient shares this common distal breakpoint. The proximal breakpoint in the present patient was delineated to ~29.3 Mb from the 16p telomere close to the proximal breakpoint of Patients 3 and 4 reported by Ballif et al. [2007]. In the remaining patients with microdeletion 16p11.2-o12.2 the proximal breakpoint was ~28.5 Mb [Patient 2, Ballif et al., 2007], ~29.6 Mb [Battaglia et al., 2009], and ~30.1 Mb [Patient 1, Ballif et al., 2007] from the 16p telomere. Apparently patients with microdeletion 16p11.2-p12.2 share a common distal breakpoint at 21.4 Mb from 16p telomere in the chromosomal region 16p12.2 but differ in the proximal breakpoint in the chromosomal region 16p11.2.

Both breakpoints of the microdeletion in the reported patient are flanked by numerous segmental duplications located in different duplication clusters (DC). The 215 kb DC 3558 is located close to the distal breakpoint, as well as the DC 3560 extending 136 kb. Both DC contain segmental duplications which share 98% sequence identity with segmental duplications located in DC 3573 flanking the proximal breakpoint. The accumulation of DC including numerous segmental duplications in 16p11.2 and 16p12.2 seems to make these chromosome regions susceptible to rearrangements mediated by nonallelic homologous recombination events [Ballif et al., 2007; Battaglia et al., 2009].

Despite different microdeletion sizes, Ballif et al. [2007] pointed out that all patients had distinctive facial features, including flat faces, downslanting palpebral fissures, low set and malformed ears, and eye anomalies. Frequent malformations described in these patients were orofacial clefting, heart defects, and minor hand and foot anomalies. All patients had frequent ear infections, feeding difficulties, hypotonia, and developmental delay.

The facial features, such as flat face, deep set eyes and low set, and posterior rotated ears were also present in the girl reported by Battaglia et al. [2009]. In addition she had frequent ear infections, feeding difficulties, development delay, and muscular hypotonia. The latter involved the face and the mouth with consequent swallowing difficulties.

The present patient shared most of these facial anomalies. Like all described patients he was affected by frequent ear infections which made PE tube replacement necessary. Severe language impairment seems to be another common characteristic in the 16p11.2-p12.2 microdeletion syndrome, which was significant in the reported patient and could not be explained by recurring ear infections. He suffered also from feeding difficulties, another common symptom in this microdeletion syndrome (Table II).

Hernando et al. [2002] described a neonate with a deletion 16p11.2, detected by chromosome analysis and further delineated by CGH analysis. The newborn had severe intrauterine growth retardation and multiple malformations (coloboma, unilateral chorioretinitis, tetralogy of Fallot, pulmonal atresia, unilateral renal agenesis, articular limitation, cubital deviation of hands, and hemivertebrae at L1 level). Facial dysmorphisms included a flat face, blepharophimosis, a short nose with hypoplastic alae nasi, low-set malformed ears, glossoptosis with hypoplastic palate, and micro-retrognathia. The boy developed cardial decompensation and died at the age of 5 months. Contrary to the patient described here and the patients described by Ballif et al. [2007] and Battaglia et al. [2009], this boy was more severely affected. This discrepancy may be due to a larger deletion in 16p than defined by Hernando et al. [2002].

Association between a microdeletion at 16p11.2 and autism was suggested first by Weiss et al. [2008]. Performing a genome-wide association study, they were able to detect a 593 kb de novo deletion in 16p11.2 (29.5–30.1 Mb from the 16p telomere) in about 1% of the families with autism as well as in patients with psychiatric disorders. This deletion was also found in 1.5% of patients with developmental or speech delay in a clinical setting.

Additional patients with autism and a microdeletion 16p11.2 were described by Kumar et al. [2008]. In total 712 autism propositi and 837 control subjects were analyzed by array-CGH using a 19k whole genome tiling path bacterial artificial chromosome array. A ~500 kb microdeletion 16p11.2 (29.6–30.2 Mb from 16p telomere) was found in four patients with autism (0.6%), but none in the control subjects.

A similar microdeletion 16p11.2 was identified in monozygotic twins who presented with craniosynostosis, aortic valve abnormalities, seizures, and mild mental retardation as reported by Ghebranious et al. [2007]. These symptoms have never been described in other patients with a microdeletion 16p11.2. So it has not been verified if craniosynostosis and heart defects are related to this microdeletion.

TABLE II. Clinical Findings in Patients With Microdeletion 16p11.2-p12.2

	Ballif et al. [2007]				Battaglia et al. [2009]	Present patient
	Case 1	Case 2	Case 3	Case 4		
Size of deletion (Mb)	8.7	7.1	7.8	7.8	8.2	7.7
Age of diagnosis	13 years 8 months	13 years 9 months	3 years 1 month	2 years 9 months	6 years	13 years 3 months
Facial dysmorphism						
Flat face	+	+	+	−	+	+
High forehead/frontal bossing	+	−	+	+	−	+
Downslanting palpebral fissures	+	+	+	+	−	−
Epicantal folds	+	−	+	+	−	−
Deep set eyes	+	−	−	+	+	+
Ears						
Low-set, malformated	+	−	+	+	+	+
Recurrent infections	**+**	**+**	**+**	**+**	**+**	**+**
Hearing loss	+	−	−	−	−	+
Hands						
Single palmar creases	+	+	+	−	+	−
Absent flexion creases	+	+	−	−	−	−
Cardiovascular Abnormalities	+	−	−	+	−	−
Stature						
Height $\leq$3.P.	+	+	−	−	−	−
Weight $<$3.P.	+	−	−	+	−	−
Gastrointestinal problems						
Feeding difficulties	**+**	**+**	**+**	**+**	**+**	**+**
GE reflux	+	+	+	+	−	−
Neurologic problems						
Hypotonia	+	−	+	+	+	+
Development						
Delay in motor development	**+**	**+**	**+**	**+**	**+**	**+**
Intellectual disability	**+**	**+**	**+**	**+**	**+**	**+**
Severe speech delay	**+**	**+**	**+**	**+**	**+**	**+**
Behavior						
Autism	−	−	−	−	−	−
Other problems than autism	−	+	−	+	+	+

Bold indicates that these clinical findings were present in all reported patients with microdeletion 16p11.2-p12.2.

Although behavioral problems seem to be a frequent symptom in patients with microdeletion 16p11.2-p12.2, an autism spectrum disorder has not been observed. Behavioral problems included anxiety [Patient 2, Ballif et al., 2007], irritability [Patient 4, Ballif et al., 2007], hyperactivity, short attention ability [Battaglia et al., 2009], and introverted behavior (present patient). The absence of autism in patients with the recurrent microdeletion 16p11.2-p12.2 could be explained by the fact, that this microdeletion approaches but did not overlap the microdeletions described by Weiss et al. [2008] and Kumar et al. [2008] except Patient 1 reported by Ballif et al. [2007] (Fig. 4).

Eleven disease causing genes are annotated in the 7.7 Mb microdeletion region on 16p11.2-p12.2 (*OTOA*, *SCNN1G*, *SCNN1B*, *COG7*, *PALB2*, *IL4R*, *IL21R*, *CLN3*, *TUFM*, *ATP2A1*, *CD19*). Currently only a few of these genes can be linked to the phenotype. Otoancorin (*OTOA*, OMIM #607038) encodes the protein otoancorin, specifically located at the interface between the apical surface of the sensory epithelia and their overlying acellular gels, and also entirely specific for the inner ear. Homozygous mutations in the *OTOA* gene were found in patients with mild to moderate prelingual sensoneural recessive deafness in one Palestinian family [Zwaenepoel et al., 2002]. It could be speculated that loss of heterozygosity in the *OTOA* gene is responsible for the hearing loss in the patient described here and in the Patient 1 described by Ballif et al. [2007]. This hypothesis is supported by the fact that a duplication in 16p11.2-p12.1, including the *OTOA* gene, is associated with hyperacusis as described by Ballif et al. in the Patient 5 [2007]. The opposite phenotypes indicate that duplication of the *OTOA* gene may lead to a gain of function and hyperacusis and a deletion of the *OTOA* gene may lead to a loss of function and hearing impairment.

CD19 is a cell surface molecule expressed exclusively by B lymphocytes and follicular dendritic cells of the hematopoetic system. Mutations in both alleles of the *CD19* gene are associated with hypogammaglobulinemia and increased susceptibility to infections, which is due to the defective response of mature B cells to

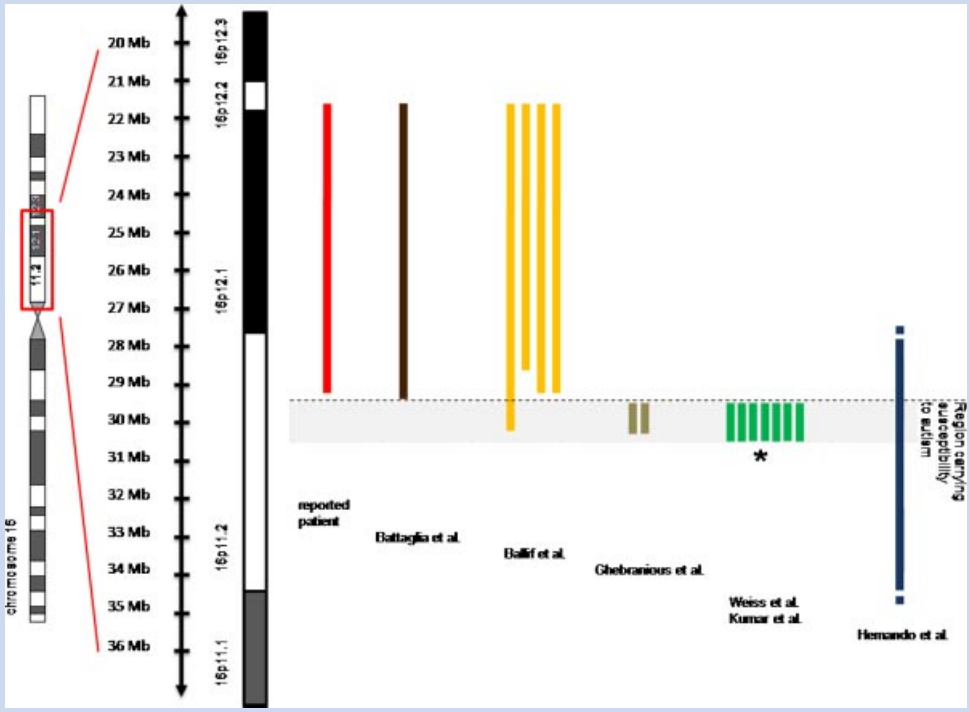


FIG. 4. Reported microdeletions in chromosome region 16p11.2-p12.2 illustrating the overlap with microdeletion in the present patient, *indicated patients with autism. [Color figure can be viewed in the online issue, which is available at www.interscience.wiley.com.]

antigenic stimulation [van Zelm et al., 2006]. It seems to be possible that the heterozygous deletion of the CD19 gene is one factor leading to recurrent ear infections.

In conclusion the patient described here confirms the observations in five recently reported patients affected by a microdeletion in 16p11.2-p12.2. Dysmorphic features, feeding difficulties, significant delay in speech development, and recurrent ear infections seem to be common symptoms of this microdeletion syndrome. All reported patients share a common distal breakpoint at 16p12.2 but vary in the proximal breakpoint at 16p11.2. The recurrent microdeletion 16p11.2-p12.2 should be distinguished from the microdeletion 16p11.2 which is associated with autism [Kumar et al., 2008; Weiss et al., 2008]. The accumulation of segmental duplications in the chromosomal region 16p11.2 and 16p12.2 makes these regions susceptible to chromosomal rearrangements. It can be expected that further patients with a microdeletion 16p11.2-p12.2 will be observed. Further research is needed for a comprehensive characterization of a genotype–phenotype correlation.

ACKNOWLEDGMENTS

We thank the patient and his parents for permission to publicize this detailed case report. We also thank Ilona Dugdale for her editorial assistance. This work was partly funded by a grant from the NGFN (grant 01GS08163).

REFERENCES

Ballif BC, Hornor SA, Jenkins E, Madan-Khetarpal S, Surti U, Jackson KE, Asamoah A, Brock PL, Gowans GC, Conway RL, Graham JM Jr, Medne L, Zackai EH, Shaikh TH, Geoghegan J, Selzer RR, Eis PS, Bejjani BA, Shaffer LG. 2007. Discovery of a previously unrecognized microdeletion syndrome of 16p11.2-p12.2. Nat Genet 39:1071–1073.

Battaglia A, Novelli A, Bernardini L, Igliozzi R, Parrini B. 2009. Further characterization of the new microdeletion syndrome of 16p11.2-p12.2. Am J Med Genet Part A 149A:1200–1204.

Ghebranious N, Giampietro PF, Wesbrook FP, Rezkalla SH. 2007. A novel microdeletion at 16p11.2 harbors candidate genes for aortic valve development, seizure disorder, and mild mental retardation. Am J Med Genet Part A 143A:1462–1471.

Hernando C, Plaja A, Rigola MA, Perez MM, Vendrell T, Egocue J, Fuster C. 2002. Comparative genomic hybridisation shows a partial de novo deletion 16p11.2 in a neonate with multiple congenital malformations. J Med Genet 39:E24.

Koolen DA, Vissers LE, Pfundt R, de Leeuw N, Knight SJ, Regan R, Kooy RF, Reyniers E, Romano C, Fichera M, Schinzel A, Baumer A, Anderlid BM, Schoumans J, Knoers NV, van Kessel AG, Sistermans EA, Veltman JA, Brunner HG, de Vries BB. 2006. A new chromosome 17q21.31 microdeletion syndrome associated with a common inversion polymorphism. Nat Genet 38:999–1001.

Kumar RA, KaraMohamed S, Sudi J, Conrad DF, Brune C, Badner JA, Gilliam TC, Nowak NJ, Cook EH Jr, Dobyns WB, Christian SL. 2008. Recurrent 16p11.2 microdeletions in autism. Hum Mol Genet 17:628–638.

Potocki L, Chen KS, Park SS, Osterholm DE, Withers MA, Kimonis V, Summers AM, Meschino WS, Anyane-Yeboa K, Kashork CD, Shaffer LG, Lupski JR. 2000. Molecular mechanism for duplication 17p11.2—The homologous recombination reciprocal of the Smith-Magenis microdeletion. Nat Genet 24:84–87.

Slavotinek AM. 2008. Novel microdeletion syndromes detected by chromosome microarrays. Hum Genet 124:1–17.

van Zelm MC, Reisli I, van der Burg M, Castano D, van Noesel CJ, van Tol MJ, Woellner C, Grimbacher B, Patino PJ, van Dongen JJ, Franco JL. 2006. An antibody-deficiency syndrome due to mutations in the CD19 gene. N Engl J Med 354:1901–1912.

Wagenstaller J, Spranger S, Lorenz-Depiereux B, Kazmierczak B, Nathrath M, Wahl D, Heye B, Glaser D, Liebscher V, Meitinger T, Strom TM. 2007. Copy-number variations measured by single-nucleotide-polymorphism oligonucleotide arrays in patients with mental retardation. Am J Hum Genet 81:768–779.

Weiss LA, Shen Y, Korn JM, Arking DE, Miller DT, Fossdal R, Saemundsen E, Stefansson H, Ferreira MA, Green T, Platt OS, Ruderfer DM, Walsh CA, Altshuler D, Chakravarti A, Tanzi RE, Stefansson K, Santangelo SL, Gusella JF, Sklar P, Wu BL, Daly MJ. 2008. Association between microdeletion and microduplication at 16p11.2 and autism. N Engl J Med 358:667–675.

Zwaenepoel I, Mustapha M, Leibovici M, Verpy E, Goodyear R, Liu XZ, Nouaille S, Nance WE, Kanaan M, Avraham KB, Tekaia F, Loiselet J, Lathrop M, Richardson G, Petit C. 2002. Otoancorin, an inner ear protein restricted to the interface between the apical surface of sensory epithelia and their overlying acellular gels, is defective in autosomal recessive deafness DFNB22. Proc Natl Acad Sci USA 99:6240–6245.

American Journal of Medical Genetics Part A 146A:2960–2965 (2008)

Research Letter

Polyneuropathy, Scoliosis, Tall Stature, and Oligodontia Represent Novel Features of the Interstitial 6p Deletion Phenotype

Birgit Zirn,[1]* Maja Hempel,[2] Andreas Hahn,[3] Bernd Neubauer,[3] Janine Wagenstaller,[4] Núria Rivera-Bruguès,[4] Tim Matthias Strom,[4] and Angelika Köhler[1]

[1]Institute of Human Genetics, University of Giessen, Giessen, Germany
[2]Institute of Human Genetics, Technische Universität München, Munich, Germany
[3]Department of Neuropediatrics, University of Giessen, Giessen, Germany
[4]HelmholtzZentrum München, Neuherberg, Germany

Received 20 May 2008; Accepted 13 July 2008

How to cite this article: Zirn B, Hempel M, Hahn A, Neubauer B, Wagenstaller J, Rivera-Bruguès N, Strom TM, Köhler A. 2008. Polyneuropathy, scoliosis, tall stature, and oligodontia represent novel features of the interstitial 6p deletion phenotype. Am J Med Genet Part A 146A:2960–2965.

To the Editor:

Deletions of the short arm of chromosome 6 are rare events and have first been defined as syndrome by Palmer et al. [1991]. Subsequently, two distinct 6p deletion syndromes have been delineated according to the localization of the deletion [Davies et al., 1999a]. These include terminal (6p24-pter) and more proximal, interstitial deletions (6p22–6p24). Terminal 6p deletions result in eye anomalies, deafness and developmental delay. Ocular manifestations (corneal opacity, iridogoniodysgenesis, Axenfeld-Rieger anomaly, glaucoma) are caused by haploinsufficiency of the *FOXC1* (*FKHL7*, *IRID1*, *FREAC3*) gene located in 6p25 [Nishimura et al., 1998] and/or the TFAP2A (AP-2alpha) gene in 6p24 [Nottoli et al., 1998], whereas deafness and developmental delay have not yet been assigned to specific genes. Interstitial 6p deletions (mostly within 6p22–p24) are associated with heart, kidney and brain defects, craniofacial anomalies (abnormal skull shape, structural eye and ear anomalies, short neck), and mental retardation [Davies et al., 1999a]. There are several candidate genes within 6p including *ATXN1*, *BMP6*, and *EDN1* [Davies et al., 1999a; Mirza et al., 2004] that might cause one or several of the developmental anomalies found. However, functional relevance has not been proven for any of them.

Distinction between these two 6p deletion syndromes became possible by the refinement of methodology for delineation of deletions, in particular fluorescence in situ hybridization (FISH) techniques. Several 6p deletions, which were originally classified as terminal deletions, were shown to be interstitial and vice versa after FISH analysis (e.g., patient 6 in Mirza et al. [2004]; patient JW in Davies et al. [1999a]). To date, only eight patients with FISH confirmed interstitial 6p deletions have been reported [Davies et al., 1996, 1999a,b; Mirza et al., 2004]. Here, we describe a large interstitial deletion in a patient with yet unrecognized clinical features of the interstitial 6p deletion syndrome. The deletion was delineated by SNP oligonucleotide array analysis and represents the most proximally located deletion reported to date.

The girl is the third of four children of healthy nonconsanguineous parents. There is no family history of malformations or developmental delay. Pregnancy was complicated by bleeding and imminent abortion in the 10th gestational week. The girl was born at term without complications (birth weight: 3,220 g, 25–50th centile; length: 52 cm, 75th centile; head circumference: 34 cm, 25–50th centile). Several facial anomalies were noted shortly after birth and became more prominent over time: brachycephaly, deep-set eyes with laterally downslanting palpebral fissures, ptosis, prominent supraorbital ridges, deep-set and posteriorly rotated ears with overfolded helices, broad nasal bridge, large

Birgit Zirn and Maja Hempel contributed equally to this work.

*Correspondence to: Birgit Zirn, M.D., Ph.D., Institute of Human Genetics, University of Giessen, Schlangenzahl 14, D-35392 Giessen, Germany. E-mail: birgit.zirn@humangenetik.med.uni-giessen.de

Published online 16 October 2008 in Wiley InterScience (www.interscience.wiley.com)
DOI 10.1002/ajmg.a.32536

mouth with open-mouth appearance and a large tongue. Moreover, thoracic asymmetry and scoliosis were recognized. Echocardiography showed a ventricular septum defect, which closed spontaneously during the first months of life. Ultrasound of the kidneys revealed a mild non-progressive unilateral obstruction of the ureteropelvic junction. At age four years, the girl's mental and motor development

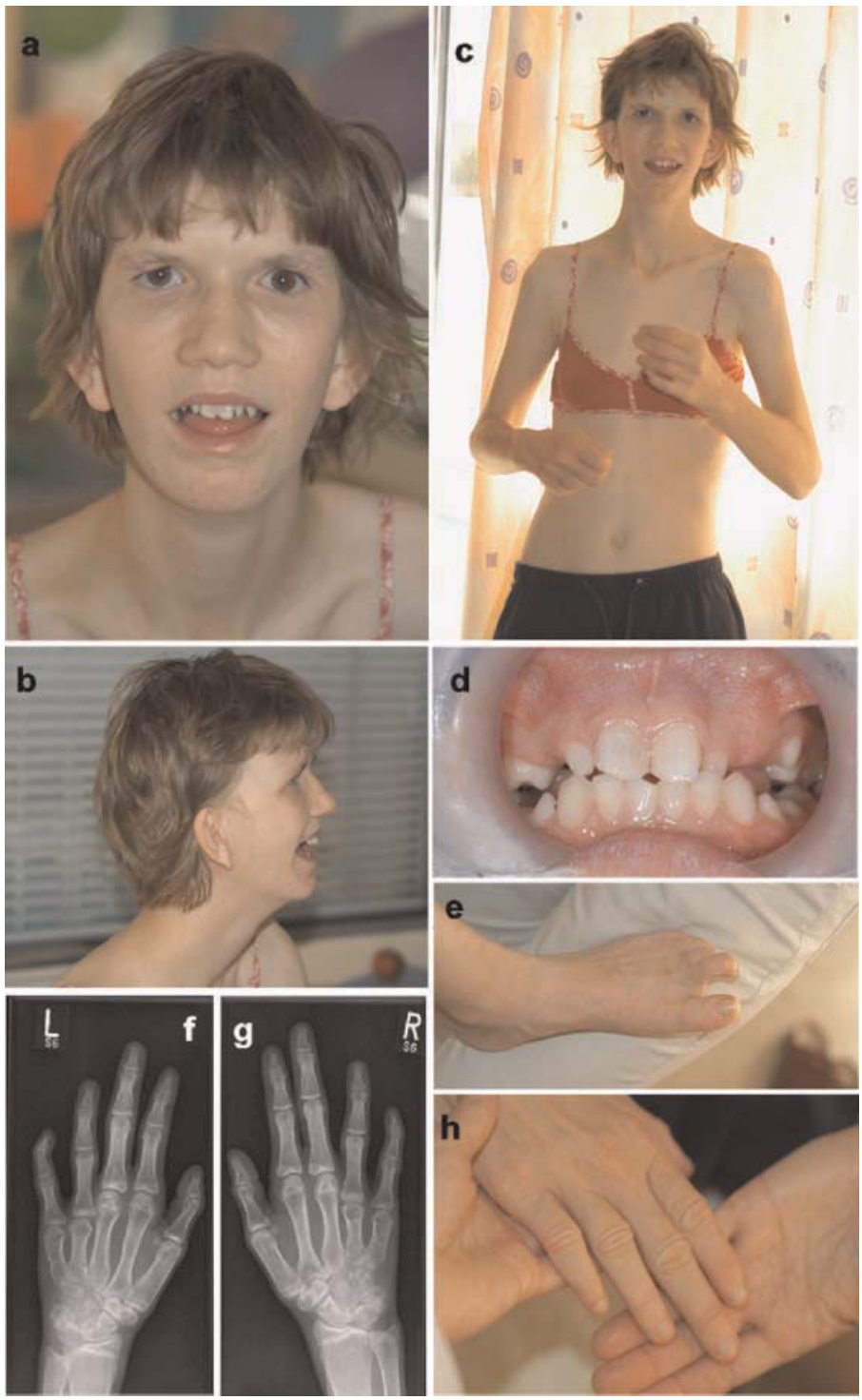


FIG. 1. Phenotype of the girl at age 15 years. **a**,**b**: Facial appearance: note brachycephaly, deep-set eyes with downslanting palpebral fissures, ptosis and prominent supraorbital ridges, broad nose, large mouth with open-mouth appearance, large and deep-set ears with overfolded helices. **c**: Thoracic asymmetry, contractures of elbows, thin and tall aspect. **d**: Mishaped teeth and oligodontia. **e**: Claw toes. **f**–**h**: X-rays and photographs of hands, note shortened phalanges IV and V, and wrinkled skin. [Color figure can be viewed in the online issue, which is available at www.interscience.wiley.com.]

was delayed, and her growth accelerated remarkably. While her height was above the 97th centile, weight and head circumference remained between the 25th and 50th centiles.

At the age of 15 years, motor functions declined and she became increasingly immobile owing to severe pain in legs and back. At that time, her height was 1.80 m (3 cm above the 97th centile, +3SD) and her bone age was 15 years. Skeletal anomalies included: scoliosis (Cobb angle 20°), thoracic asymmetry with flat left thorax, long neck, prominent trochanter bones of the hip, elbow contractures, short 4th and 5th phalanges of both hands, large feet with claw toes II-V and increased distance between digits I and II. Finger and toe nails were spoon-shaped. In addition to the facial features described above, she had a high arched palate with oligodontia (eight permanent teeth missing), a low frontal hairline and an inverse hair growth pattern in the neck. Her skin was generally dry with rough and wrinkled lesions at her hands (Fig. 1). Both parents and all siblings are within the normal range for height.

Neurological examination showed absent Achilles tendon reflexes. The girl's gait was broad-based and her movements were clumsy, but no unequivocal atactic symptoms were detectable. She spoke single words, followed simple instructions and was hyperactive. Median nerve sensory evoked potentials were normal, while those elicited from the tibial nerve were slightly delayed. Motor and sensory nerve conduction of both median nerves was slightly reduced. No sensory nerve action potentials could be obtained with surface electrodes from both sural nerves. Distal latencies of both peroneal and the right tibial nerves were prolonged. Motor nerve conduction velocities of all three nerves were moderately reduced. The compound motor action potentials were within the normal ranges. Taken together, these findings correspond to a mild motor and sensory polyneuropathy with preponderance of the legs. Brain imaging (MRI) was normal. The parents gave informed consent to the publication of photographs and clinical data of their daughter.

Chromosome analysis was performed at an average of 550 bands per haploid genome on G-banded metaphases, which had been prepared from cultured peripheral lymphocytes following standard protocols. A small deletion of the distal short arm of one chromosome 6 was found (Fig. 2), and the patient's karyotype was determined as follows: 46,XX,del(6)(p23p?25). The deletion was shown to be interstitial by FISH using a subtelomeric probe of chromosome 6p (Abbott). Karyotypes of the parents were normal.

FIG. 2. Partial G-banded karyotype of chromosomes 6 of the patient. The deleted chromosome is right.

To refine the 6p deletion region, SNP oligonucleotide array analysis was performed using Sentrix® HumanHap 550 Genotyping BeadChips (Illumina®). Genomic DNA of the patient and her parents was isolated from peripheral blood lymphocytes by salting-out procedure and processed according to the manufacturer's instructions (for detailed description of the method see Infinium II Assay HC Two-Sample, Manual, Experienced User Card, Illumina). The Chips were scanned with the Illumina BeadArray™ Reader, which uses a laser to excite the fluorophore of the single-base extension product on the Chip and records high-resolution images of the light emitted.

TABLE I. Results of Microarray Analysis: CNV No. 4 Represents the Exact Delineation of the Cytogenetically Detected 6p Deletion

CNV No.	Chr	Start position	End position	Length of CNV bp	Median of SNP copies in CNV	Confirmed by qPCR	De novo CNV	Ensembl genes	Overlap with known CNV (%)
1	2	81376544	81408438	31,894	1.14	NA	NA	5	100
2	5	9955340	9977172	21,832	1.20	NA	NA	13	100
3	5	97075236	97102304	27,068	1.23	NA	NA	11	96
4	**6**	**14554649**	**27849661**	**13,295012**	**1.38**	**Yes**	**De novo**	**326**	**1**
5	8	5583294	5591903	8,609	1.46	NA	NA	6	100
6	11	81181640	81194909	13,269	1.44	NA	NA	9	100
7	14	61030245	61037956	7,711	1.49	No	NA	6	0
8	16	34373576	34464860	91,284	2.58	NA	NA	7	100

NA, not applicable.

Results were called with Genotyping Module of BeadStudio using the default call threshold of 0.15. Data analysis and validation of CNVs by quantitative real time PCR (qPCR) was performed as previously described by Wagenstaller et al. [2007].

The deletion in the short arm of chromosome 6 was confirmed by microarray analysis (Table I; CNV No. 4). However, the deletion was shown to extend proximally instead of distally from 6p23 as assumed by analysis of G-banded chromosomes. The deletion extended 13.3Mb and contained 3,265 SNPs. The proximal breakpoint of the deletion was determined on region 6p22.1 (chr6:14554649), the distal breakpoint on 6p23 (chr6:27849661). The proximal breakpoint of 6p22.1 was flanked by a segmental duplication (UCSC segmental duplication). Three hundred twenty six Ensembl genes are annotated in the region of deletion, 68 of these genes are listed as OMIM genes and 8 genes are classified as OMIM disease genes (*DTNBP1*, *ATXN1*, *NHLRC1*, *TPMT*, *DEK*, *ALDH5A1*, *KIAA0319*, *HFE*). In addition, seven other CNVs were detected. Six of these (CNV No. 1, 2, 3, 5, 6, 8) were noted in the Database of Genomic Variants and represent known

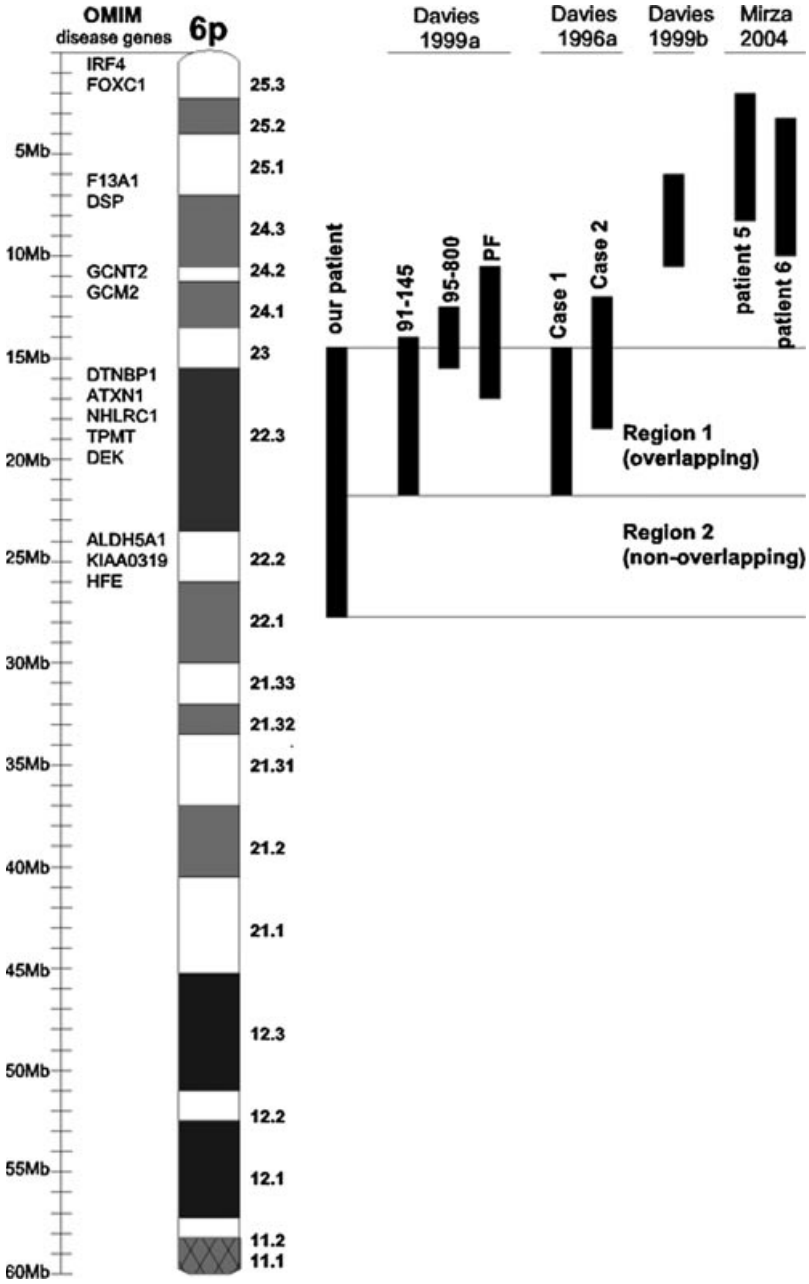


Fig. 3. Extent of 6p deletions in previously published patients and in our patient.

variants without pathological relevance. One CNV (No. 7) could not be confirmed by quantitative real time PCR.

The interstitial 6p deletion of 13.3 Mb (6p22.1–6p23) reported here represents the most proximal deletion described to date and is the first 6p deletion that was analyzed by array technology. It partially overlaps with five previously published interstitial 6p deletions (patients 91–145, 95–800, PF in Davies et al. [1999a]; Patients 1 and 2 in Davies et al. [1996]), which had been characterized by FISH analysis (Fig. 3; overlapping region 1). Developmental delay, craniofacial anomalies as well as heart and urogenitary tract abnomalities are clinical symptoms in our patient that match well with the interstitial 6p deletion phenotype as it was initially described by Davies et al. [1999a]. In contrast, some of the girls' symptoms such as polyneuropathy, tall stature, scoliosis, and oligodontia have not been previously reported (Table II and Fig. 1).

We posit that the additional clinical features of our patient are caused by deletion of genes in the more proximal part of 6p (6p22.2–22.1 or non-overlapping region 2 in Fig. 3). This region contains 54 OMIM genes of which three are classified as OMIM disease genes (*ALDH5A1*, *KIAA0319*, *HFE*). However, there is no obvious functional link of these three genes to the neurological and skeletal anomalies and oligodontia in our patient, since the expression of these genes is low or not present in nerve and bone tissue (http://www.ncbi.nlm.nih.gov/UniGene/ESTProfileViewer). It is possible that the deleted region alters the expression of more distant genes.

Previous publications discussed a genotype–phenotype correlation in patients with more proximally located deletions of chromosome 6p. Mirza et al. [2004] proposed that especially haploinsufficiency of *ATXN1* (ataxin 1), which is also deleted in our patient, contributes to the developmental

TABLE II. Summary of clinical features in our patient and patients with interstitial 6p deletions reported in the literature

	Our patient	Davies et al. [1999a]			Davies et al. [1996]		Davies et al. [1999b]	Mirza et al. [2004]	
Patient		91–145	95–800	PF	1	2		5	6
Sex	F	F	M	M	M	F	M	F	F
Follow-up	16 y	3 y	20 y	4.5 y	15 y	1 y	2 y	0.5 y	1 y
Delivery, at term	+	+	+	+	+	+	W 36	+	W 35
Birth weight, normal	+	+	− (a)	+	+	+	+	+	NR
Craniofacial dysmorphisms	+	+	+	+	+	+	+	+	+
Brachy- or dolichocephaly	+	−	+	+	−	−	−	+	−
Retrognathia	+	−	−	−	−	+	−	−	−
Short neck	− (b)	+	−	+	−	+	−	+	−
Bilateral cleft lip	−	−	−	−	−	−	−	−	+
Structural eye defects (e.g. downslanting palpebral fissures, hypertelorism)	+	+	−	+	+	−	+	−	+
Strabism	+	−	−	+	−	−	−	−	−
Nystagmus	−	−	−	+	−	−	−	−	−
Visual impairment	−	−	+	−	+	−	+	−	−
Structural ear defects	+	−	−	−	−	+	+	+	+
Hearing loss	−	+ (c)	−	−	−	−	−	−	−
Abnormal teeth	+	−	−	−	+	−	−	−	−
Mishaped teeth	+	−	−	−	+	−	−	−	−
Oligodontia	+	−	−	−	−	−	−	−	−
Heart defect	+ (d)	−	−	−	−	+ (e)	+ (d,f)	−	+ (e,f)
Kidney defect/abnormality	+	−	+	−	−	+	−	−	+ (g)
Hand/feet abnormalities	+	+	−	+	+	+	+	+	+
Short phalanges	+	−	−	−	−	−	−	−	−
Nail abnormalities	+	−	−	+	+	+	−	−	+
Accelerated growth/macrosomia	+	−	−	−	−	−	−	−	−
Sceletal anomalies	+	−	+	−	−	−	+	+	+
Scoliosis	+	−	−	−	−	−	−	−	−
Stiffness/contractures of joints	+	−	−	−	−	−	−	−	−
Torticollis	+	−	−	−	−	−	−	−	−
Pectus excavatum/carinatum	−	−	+	−	−	−	+	+	+
Skin abnormalities	+	−	+	−	−	−	−	−	+
Sclerodermia	+	−	−	−	−	−	−	−	−
Eczema	+	−	−	−	−	−	−	−	+
Demyelinating polyneuropathy	+	−	−	−	−	−	−	−	−
Developmental delay	+	+	+	+	+	+	− (h)	+	− (h)
Abnormal brain imaging	−	+ (i)	−	+ (j)	−	+ (k)	−	−	−

F, female; M, male; NR, not reported; w, week; (a) 2,370 g, 0.4th centile, (b) long neck, (c) probably moderate, difficult to assess, (d) VSD, (e) PDA, (f) ASD, (g) mild pelviectasis, (h) no developmental delay at 1 year, (i) cerebral atrophy, increased extraaxial fluid, (j) gliosis in the white matter with increased perivascular spaces, subarachnoidal cysts in the temporal lobes, (k) non-progressive hydrocephalus, no intracranial hypertension.

delay in individuals with 6p interstitial deletions. Interestingly, the expansion of an unstable CAG repeat within *ATXN1* may cause demyelinating polyneuropathy in young adults with spinocerebellar ataxia type 1 [Ragno et al., 2005].

Mirza et al. [2004] assume that the size of 6p deletions does not correlate with the severity of the phenotype. Some reported patients with interstitial 6p deletions are young and may develop additional clinical features later in life. For this reason and because the number of reported patients is still small, it remains difficult to derive an exact genotype–phenotype correlation in interstitial 6p deletions. However, the expanding use of array technologies will not only detect additional patients with interstitial 6p deletions but will also help to further delineate the phenotype. This may enable a more specific prognosis in affected patients and in prenatally diagnosed cases.

REFERENCES

Davies AF, Olavesen MG, Stephens RJ, Davidson R, Delneste D, Van Regemorter N, Vamos E, Flinter F, Abusaad I, Ragoussis J. 1996. A detailed investigation of two cases exhibiting characteristics of the 6p deletion syndrome. Hum Genet 98:454–459.

Davies AF, Mirza G, Sekhon G, Turnpenny P, Leroy F, Speleman F, Law C, van Regemorter N, Vamos E, Flinter F, Ragoussis J. 1999a. Delineation of two distinct 6p deletion syndromes. Hum Genet 104:64–72.

Davies AF, Mirza G, Flinter F, Ragoussis J. 1999b. An interstitial deletion of 6p24-p25 proximal to the FKHL7 locus and including AP-2alpha that affects anterior eye chamber development. J Med Genet 36:708–710.

Mirza G, Williams RR, Mohammed S, Clark R, Newbury-Ecob R, Baldinger S, Flinter F, Ragoussis J. 2004. Refined genotype-phenotype correlations in cases of chromosome 6p deletion syndromes. Eur J Hum Genet 12:718–728.

Nishimura DY, Swiderski RE, Alward WL, Searby CC, Patil SR, Bennet SR, Kanis AB, Gastier JM, Stone EM, Sheffield VC. 1998. The forkhead transcription factor gene FKHL7 is responsible for glaucoma phenotypes which map to 6p25. Nat Genet 19:140–147.

Nottoli T, Hagopian-Donaldson S, Zhang J, Perkins A, Williams T. 1998. AP-2-null cells disrupt morphogenesis of the eye, face, and limbs in chimeric mice. Proc Natl Acad Sci USA 95:13714–13719.

Palmer CG, Bader P, Slovak ML, Comings DE, Pettenati MJ. 1991. Partial Deletion of chromosome 6p: Delineation of the syndrome. Am J Med Genet 39:155–160.

Ragno M, Perretti AC, Castaldo I, Scarcella M, Acciarri S, Manganelli F, Santoro L. 2005. Multimodal electrophysiologic follow-up study in 3 mutated but presymptomatic members of a spinocerebellar ataxia type 1 (SCA1) family. Neurol Sci 26: 67–71.

Wagenstaller J, Spranger S, Lorenz-Depiereux B, Kazmierczak B, Nathrath M, Wahl D, Heye B, Glaser D, Liebscher V, Meitinger T, Strom TM. 2007. Copy-number variations measured by single-nucleotide-polymorphism oligonucleotide arrays in patients with mental retardation. Am J Hum Genet 81:768–779.

European Journal of Medical Genetics 54 (2011) 225–230

Contents lists available at ScienceDirect

European Journal of Medical Genetics

journal homepage: http://www.elsevier.com/locate/ejmg

Original article

3.7 Mb tandem microduplication in chromosome 5p13.1-p13.2 associated with developmental delay, macrocephaly, obesity, and lymphedema. Further characterization of the dup(5p13) syndrome

Konrad Oexle [a,*], Maja Hempel [a], Anna Jauch [c], Thomas Meitinger [a,b], Núria Rivera-Brugués [b], Sabine Stengel-Rutkowski [d], Tim Strom [b]

[a] Institute of Human Genetics, Technische Universität München, Munich, Germany
[b] Institute of Human Genetics, Helmholtz Zentrum München, Neuherberg, Germany
[c] Institute of Human Genetics, University Hospital Heidelberg, Heidelberg, Germany
[d] Institute of Human Genetics, Ludwig-Maximilians Universität München, München, Germany

ARTICLE INFO

Article history:
Received 13 September 2010
Accepted 24 December 2010
Available online 4 January 2011

Keywords:
Mental retardation
Autism
Low posterior hairline
Obesity
Lymphedema
Dup(5p) syndrome

ABSTRACT

In a male patient with developmental delay, autistic behaviour, obesity, lymphedema, hypertension, macrocephaly, and facial features of chromosome 5p duplication (trisomy 5p) a 3.7 Mb *de novo* tandem microduplication of 5p13.1-13.2 (rs4703415-rs261752, i.e., chr5:35.62–39.36 Mb) was identified. This observation contributes to the characterization and dissection of the 5p13 duplication syndrome. The possible role of increased *NIPBL* gene dosage is discussed.

1. Introduction

Trisomy of the short arm of chromosome 5 was first reported by Lejeune et al. [1] in a family with 5p translocation that also provided one of the first three patients with "cri du chat"-syndrome (i.e., distal monosomy 5p). Trisomy 5p causes a variable syndrome of macrocephalic brain dysfunction, congenital heart defect, respiratory compromise or infection, club feet, and craniofacial features including dolichocephaly, short neck, dysplastic ears, epicanthus, upslanting palpebral fissures, hypertelorism, broad nasal root, short nose, high-arched palate, and microre trognathia [2]. In patients with terminal duplication 5p due to unbalanced translocation, malformations were found to be more frequent and severe if the breakpoint was proximal of 5p14 [3]. Examination of rare patients with intrachromosomal duplications or with a marker chromosome containing a segment of 5p indicated that band 5p13 may be crucial for the development of the phenotype [2,4–6]. There is a paucity of cases reported with pure duplications of the proximal region of 5p [6]. Here we describe a patient with chromosome 5p13.1–13.2 tandem microduplication that is helpful in the further dissection of 5p13 duplication syndrome.

1.1. Case report

The propositus was the single child of non-consanguineous German parents aged 31 yrs (mother) and 34 yrs (father). Parental heights were 170 cm and 162 cm, respectively. Fetal movements were weak. At birth (42 nd week of pregnancy), length (53 cm), weight (3.4 kg), and head size (OFC, 35 cm) were in the upper normal range. Muscular hypotonia, weak cry, dysplastic ears, and talipes varus were noticed. Macrocephaly was observed at the age of 4 months, cranial ultrasound revealed mildly enlarged ventricular spaces. At the age of 1.5 yrs diastasis recti was diagnosed. Mental and motor development were delayed (walking at 2.5 yrs, first words at 4 yrs). Muscular hypotonia persisted. (Craniofacial features are shown in Fig. 1). At the age of 5.6 yrs, wide intercanthal distance, epicanthal folds, slightly prominent bulbi, diffuse lining of the lateral eyebrows, broad and prominent nasal root, short and broad nose with anteverted nostrils, shallow philtrum, high-arched

* Corresponding author. Institute of Human Genetics, Technische Universität München, Trogerstr. 32, 81675 München, Germany. Tel.: +49 89 4140 6386; fax: +49 89 4140 6382.
E-mail address: oexle@humangenetik.med.tum.de (K. Oexle).

1769-7212/$ – see front matter © 2010 Elsevier Masson SAS. All rights reserved.
doi:10.1016/j.ejmg.2010.12.012

Fig. 1. Propositus at the age of 1.4 yrs with dolichocephalic macrocephaly, low-set ears, short palpebral fissures, broad nasal root, short neck, and augmented nuchal tissue.

palate, everted lower lip, and small, low set, posteriorly rotated ears with broad transversal helix and pointed tragus were observed (Suppl.Fig. 1). Hearing was normal. Soft tissue of the neck was augmented. The posture was kyphotic with a protruding abdomen. Fingers had a tapered form. One testis was located in the inguinal canal. There was an attention deficit and perception skills were not adequate for the patient's age.

During adolescence the propositus became overweight and behaviour was described as autistic. He finished public school, but needed special support to achieve some level of professional training. He had a tendency to echolalia and stereotypic responses. Self-care was insufficient and functioning in daily life remained dependent on parental help. Obesity further increased to a BMI of 40 at the age of 22 yrs. Body height reached 175 cm with mild scoliosis, OFC was 62 cm (2 cm above the 97th centile, [7]). High blood pressure necessitated antihypertensive medication. Uric acid was elevated while renal, hepatic and lipid parameters were normal. Swelling of feet and lower legs was diagnosed as lymphedema with positive Stemmer's sign. Echocardiography and Doppler sonography of abdominal and leg veins gave normal results. Fig. 2 displays the craniofacial features of the propositus at the age of ~10 and 23 yrs: Macrocephalus with mild dolichocephaly, short neck, low posterior hairline, posteriorly rotated, mildly dysplastic ears, short palpebral fissures, epicanthus (right eye), broad and prominent nasal root and bridge, short nose with anteverted nostrils, long and shallow philtrum, high-arched palate (not shown), thick and everted lower lip, and small chin.

Fig. 2. Propositus at the age of ~10 and 23 yrs. See main text for a detailed description.

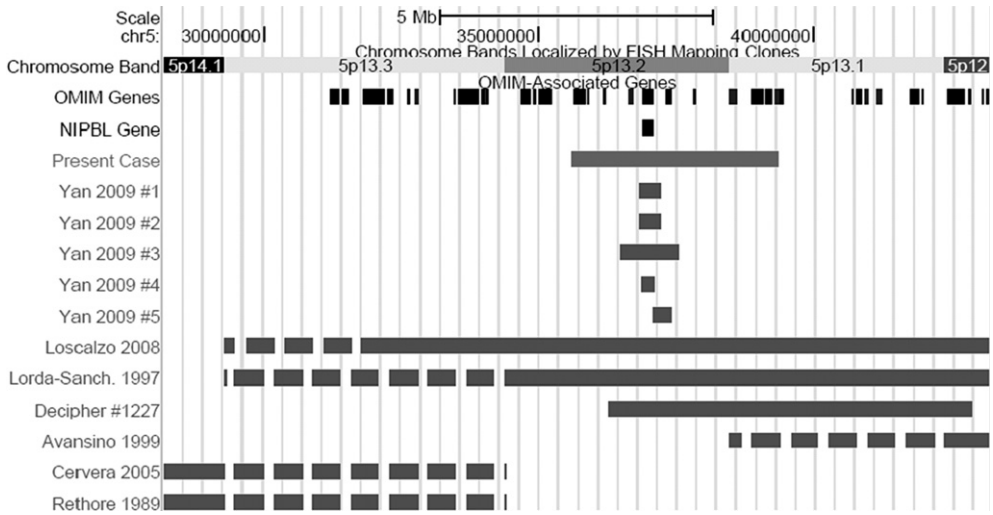


Fig. 3. Origin of the duplicated material in the present case as compared to previously reported cases with small duplications involving chromosomal band 5p13. Dashed lines indicate the approximate ends of aberrations in cases that were delimited by cytogentic methods. The position of the NIPBL gene is also indicated. Phenotypic features are presented in Table 1.

2. Materials and methods

Conventional cytogenetics (500 bands) and FISH analyses were done according to standard procedures. For FISH analysis of 5p13.2 three BAC clones (CTD-2267H19, CTD-2007F2, and RP11-7M4) of this chromosomal region were labelled and hybridized as previously described [8]. Copy number variation (CNV) was detected using Infinium® Human 610-Quad BeadChips (Illumina, San Diego, CA, USA) and Illumina GenomeStudio software. Data were analyzed as described in Wagenstaller et al. [9]. The resolution limit was in the range of 0.01–0.1 Mb but varied due to inhomogeneous SNP representation on the chip. SNP positions were indicated according to UCSC genome browser (http://genome.ucsc.edu), build hg18. All CNVs were checked for gene content (www.ensembl.org/Homo_sapiens) and overlap with known genetic variants as listed in the Database of Genomic Variants (DGV, [10], http://projects.tcag.ca/variation/).

3. Results

CNV analysis (Suppl.Fig. 2) revealed a *de novo* 3.7 Mb microduplication at chromosome 5p13.1–13.2 (rs4703415-rs261752, i.e., chr5:35,624,846–39,364,263). FISH analysis (Suppl.Fig. 3) using three BAC clones from 5p13.2 showed that this was an intrachromosomal tandem microduplication, indicated by hyperintense or split signals on one chromosome 5p in metaphase spreads and by two of three signals being in close proximity to each other in interphase nuclei.

BeadChip analysis revealed 11 other gene-containing CNVs in the patient. However, these CNVs were also present in the parents of the patient. Moreover, they overlapped entirely with benign polymorphisms described previously in the Database of Genomic Variants. Conventional cytogenetics and FISH analyses of 4pter, 4qter, 21qter, 7q11.23, 22q11.2, and 22q13 gave normal results.

4. Discussion

Developmental delay, muscular hypotonia, macrocephaly, and craniofacial appearance of the propositus fitted the diagnosis of 5p duplication syndrome. Previous studies of patients with intrachromosomal duplications of 5p proposed a critical region to reside between 5p12 and 5p13.3 [2,4,6,11; also see Fig. 3 and Table 1]. Our finding further delimits this region to 5p13.1–13.2. The region contains the *NIPBL* (nipped B-like) gene at 5p13.2. Haploinsufficiency of *NIPBL* causes Cornelia de Lange syndrome (CdLS). Recently, Yan et al. [12] reported on four patients with 0.25–1.1 Mb duplications that comprised only the *NIPBL* gene or only the *NIPBL* gene and its close neighbours. Similar to the present case these patients had mental retardation, behavioural abnormalities, and at least some of the craniofacial features of dup(5p) syndrome. Considering that the dup(5p) phenotype is variable [13], the increased *NIPBL* gene dosage may therefore be regarded as a major contributor to the phenotype. Indeed, an OMIM entry (#613174) on "5p13 duplication syndrome" has been based solely on the four patients of Yan et al. [12].

Some observations qualify but do not exclude the notion that increased dosage of the *NIPBL* gene at 5p13.2 is the dominant influence: i) Yan et al. [12] also reported a patient with a 0.33 Mb duplication that overlapped only with the distal end of the *NIPBL* gene and did not necessarily increase its dosage. This patient had mental retardation and some, albeit unspecific craniofacial features. ii) In the two patients described by Rethoré et al. [14] and Cervera et al. [15] intrachromosomal duplications appeared to have their proximal breakpoint at 5p13.3, i.e., thus not including the *NIPBL* gene at 5p13.2. iii) The chromosomal material contained in the marker chromosome (5p10-5p13.1) of the patient described by Avansino et al. [5] was located proximal of 5p13.2. iv) Some signs of dup(5p13) syndrome have been observed in patients with familial translocations with breakpoints at or distal of 5p13 [13] who have siblings with "cri du chat"-syndrome but not CdLS, indicating that the translocated material did not include the *NIPBL* gene.

It is thus likely that besides *NIPBL* other genetic factors influence the dup(5p) phenotype and that even the dup(5p13) phenotype may not be attributable to the effect of this gene only. The duplication described here comprises 22 RefSeq genes. Some of them might possibly contribute to the phenotype, such as the "glial cell line-derived neutrophic factor" gene *GDNF* (OMIM +600837), for instance, that may be of interest in relation to the macrocephaly.

Individual features present or absent in the propositus deserve to be discussed in comparison with other dup(5p13) patients as soon as a larger number of them is collected. For now, we focus on the following: i) Neither the propositus nor any of the patients in

Table 1
Phenotypic features of the present and of previous cases with small duplications at 5p13 (also see Fig. 3).

	Present case	Yan et al., 2009 #1	Yan et al., 2009 #2	Yan et al., 2009 #3	Yan et al., 2009 #4	Yan et al., 2009 #5	Loscalzo et al., 2008	Lorda-S. et al., 1997	Decipher #1227	Avansino et al., 1999	Cervera et al., 2005	Rethoré et al., 1989
Reference		[12]	[12]	[12]	[12]	[12]	[6]	[2]	Decipher	[5]	[15]	[14]
Gender	M	F	F	F	M	F	F	M	M	F	F	M
Age [yrs]	23	18	6	30	0.75	5	5.5	0.5	3	0.1	5	1
Height	P25–P50	Normal	Normal	Normal	<P5	<P3	P10	25%		<P3	P25	P25
Weight	>P97	>P95	P90	Normal	<P5	P50	P80	P50		<P3	<P3	P10–25
Mental retard/ Develop.delay	+	+	+	+	+	+	+	+	+	+	+	+
Behavioural abnormalities	+	+	+	+		−						
Muscular hypotonia	+	−	+		+	−	+	+		+	+	+
Seizures		−	Staring spells		−	+	+	+			+	
Ventriculo-megaly	(+)						(+)	+		(+)	−	
Other brain dysmorphies					Callosal agenesis			Callosal hypogen.			Frontal atrophy	
Macrocephaly	+	+					+	(+, P75–90)		+	(+, P50)	+
Other skull features (1)	d, pf	pf, bn	fb	fb, bn	cs	pf	d, bn	d, fb, bn, cs, lfo	pf	d, pf, bn, lfo	d, lfo	pf, bn
	(1) d, dolichocephaly; pf, prominent forehead; fb, frontal bossing; bn, bitemporal narrowing; cs, craniosynostosis; lfo, large fontanelle											
Low set, dysplastic ears	+	+	+		+		+	+			+	+
Ocular region (2)	spf, epc	spf, upf	Astigma-tism, strabism	epc, hypo	Hyper, proptosis	spf	spf, upf	epc, apparent hyper		upf, epc, apparent hyper	upf, epc	spf, epc, coloboma
	(2) spf, short palpebral fissures; upf, upslanting palpebral fissures; epc, epicanthal folds; hypo, hypotelorism; hype, hypertelorism											
Nose, midface (3)	sns, bnr, avn	bnr, bulbous tip	bnr			Paranasal groove	fnb, anv	sns, mfh, fnb	fnb	sns, mfh, fnb	bnr, mfh, avn	sns, fnb
	(3) sns, short nose; mfh, midface hypoplasia; bnr, broad nasal root; fnb, flat nasal bridge; avn, anteverted nostrils											
Mouth (4)	swp, hap, everted lower lip	stp, hap, dtc	stp	stp, hap	hap	swp, missing incisors		fll, open mouth	stp, dtc, fll	hap	stp, hap	stp, swp, everted lower lip
	(4) stp, short philtrum; swp, shallow philtrum; hap, high-arched or narrow palate; dtc, down-turned corners of the mouth; fll, full lips											
Chin	Micrognathia	Prognathia		Micrognathia				Microre-trognathia				Microre-trognathia
Neck (5)	stn, lph			lph								stn
	(5) stn, short neck; lph, low posterior hairline											
Hands (6)	Tapered fingers	spc, lfi, ulnar dev. of thumb	lfi			hsc, short 5th finger		lfi		hsc, lfi	spc, 4th& 5th finger dysplasia	lfi
	(6) spc, single palmar crease; hsc, hockey stick crease; lfi, long fingers											
Lower limbs	Talipes varus	Large feet	Genu valgum					Short hallux, club feet	Narrow feet	Short hallux	Dorsiflex. hallux, crowded toes	Crowed toes
Heart defect	−						ASD		PDA	VSD	−	
Respir. difficulties	−							+		+		
Abdom. muscle hypoplasia/ hernia	+							+		+		+

Cryptorchidism	+				+		
Others	Lymphedema, high blood pressure	Scant hair, eyebrows absent	Preauricular pits	Laryngomalacia, gastro-esophag. reflux	Laryngomalacia, bell-shaped thorax	Pectus excavat., auricular tags	IVF, polyhydramn., intestinal malrotat., preauricular pit

Yan et al. [12] with small duplications comprising the *NIPBL* gene (see above) were reported to have a heart defect. In contrast, of 13 previously reported dup(5p) and dup(5p13) patients who had echocardiography, 10 showed heart defects [6,11]. Increased *NIPBL* gene dosage may thus not be relevant for impaired heart development in dup(5p), even though haploinsufficiency of this gene is associated with heart defects in CdLS. ii) Respiratory problems that are frequent in dup(5p) patients (50–60%, [5]) also were not reported in our patient and the patients of Yan et al. [12]. iii) Obesity developing in childhood or adolescence as in our patient may be a feature of patients with interstitial duplications at 5p13. Patient 1 of Yan et al. [12] was obese at an age of 18 yrs, the weight of their patient 2 was at the 90th centile at the age of 6 yrs. The patient with dup(5)(p11-p13.3) described by Loscalzo et al. [6] had a BMI above the 97th centile (105 cm, 22 kg) at an age 5.5 yrs. However, of the obesity susceptibility loci that are listed in OMIM or were identified in recent genome-wide association studies [16–19], none maps to 5p13. The lymphedema observed in the present patient may or may not result from his obesity. An association of hereditary lymphedema to 5p13 has not been described previously (cf OMIM #153100).

In conclusion, using high resolution array technology patients with microduplications of chromosome 5p13 can be identified that are helpful in elucidating the genotype–phenotype relation of these aberrations. Increased gene dosage of *NIPBL* appears to be a major, but not the only, player in that relation.

A tabular description of this case (TUM255925) is represented in the Decipher database.

Acknowledgment

We are indebted to our patient for his permission to present his case to the scientific and medical community. The research was supported by a grant from the German Ministry for Education and Research (NGFNplus/www.ngfn.de/englisch/15.htm, project number 01GS08163) and participated in the MRNET consortium (http://www.german-mrnet.de/). The BAC clones from 5p13.2 were kindly provided by Joris Vermeesch, Center for Human Genetics, Leuven, Belgium.

Appendix. Supplementary material

Supplementary material associated with this paper can be found, in the online version, at doi:10.1016/j.ejmg.2010.12.012.

References

[1] J. Lejeune, J. Lafourcade, R. Berger, R. Turpin, Ségrégation familiale d'une translocation 5~13 déterminant une monosomie et une trisomie partielles du bras court du chromosome 5: Maladie du "Cri du chat" et sa "réciproque", C. R. Hebd. Seances. Acad. Sci. 258 (1964) 5767–5770.

[2] I. Lorda-Sánchez, M. Urioste, A. Villa, M.C. Carrascosa, M.S. Vázquez, A. Martínez, M.L. Martínez-Frías, Proximal partial 5p trisomy resulting from a maternal (19;5) insertion, Am. J. Med. Genet. 68 (1997) 476–480.

[3] B. Zabel, W. Baumann, J. Gehler, G. Conrad, Partial trisomy for short and long arm of chromosome no. 5. Two cases of two possible syndromes, J. Med. Genet. 15 (1978) 143–147.

[4] N.L. Chia, L.R. Bousfield, B.H. Johnson, A case report of a de novo tandem duplication (5p)(p14→pter), Clin. Genet. 31 (1987) 65–69.

[5] J.R. Avansino, T.R. Dennis, P. Spallone, A.D. Stock, M.L. Levin, Proximal 5p trisomy resulting from a marker chromosome implicates band 5p13 in 5p trisomy syndrome, Am. J. Med. Genet. 87 (1999) 6–11.

[6] M.L. Loscalzo, T.A. Becker, M. Sutcliffe, A patient with an interstitial duplication of chromosome 5p11-p13.3 further confirming a critical region for 5p duplication syndrome, Eur. J. Med. Genet. 51 (2008) 54–60.

[7] K.M. Bushby, T. Cole, J.N. Matthews, J.A. Goodship, Centiles for adult head circumference, Arch. Dis. Child. 67 (1992) 1286–1287.

[8] L. Blackx, R. Thoelen, H. Van Esch, J.R. Vermeesch, Direct fluorescent labelling of clones by DOP PCR, Mol. Cytogenet. (2008) 1–4.

[9] J. Wagenstaller, S. Spranger, B. Lorenz-Depiereux, B. Kazmierczak, M. Nathrath, D. Wahl, B. Heye, D. Glaser, V. Liebscher, T. Meitinger, T.M. Strom, Copy-number variations measured by single-nucleotide-polymorphism oligonucleotide arrays in patients with mental retardation, Am. J. Hum. Genet. 81 (2007) 768–779.
[10] A.J. Iafrate, L. Feuk, M.N. Rivera, M.L. Listewnik, P.K. Donahoe, Y. Qi, S.W. Scherer, C. Lee, Detection of large-scale variation in the human genome, Nat. Genet. 36 (2004) 949–951.
[11] A. Vera-Carbonell, J.A. Bafalliu, E. Guillén-Navarro, A. Escalona, M.J. Ballesta-Martínez, C. Fuster, A. Fernández, I. López-Expósito, Characterization of a de novo complex chromosomal rearrangement in a patient with cri-du-chat and trisomy 5p syndromes, Am. J. Med. Genet. A 149 (2009) 2513–2521.
[12] J. Yan, F. Zhang, E. Brundage, A. Scheuerle, B. Lanpher, R.P. Erickson, Z. Powis, H.B. Robinson, P.L. Trapane, D. Stachiw-Hietpas, K.M. Keppler-Noreuil, S.R. Lalani, T. Sahoo, A.C. Chinault, A. Patel, S.W. Cheung, J.R. Lupski, Genomic duplication resulting in increased copy number of genes encoding the sister chromatid cohesion complex conveys clinical consequences distinct from Cornelia de Lange, J. Med. Genet. 46 (2009) 626–634.
[13] A. Schinzel, Catalogue of unbalanced chromosome aberrations in man, second ed. Walter de Gruyter, Berlin, 2001, 242–246.
[14] M.O. Rethoré, M.C. Blois, M. Peeters, P. Popowski, C. Pangalos, J. Lejeune, Pure partial trisomy of the short arm of chromosome 5, Hum. Genet. 82 (1989) 296–298.
[15] M. Cervera, S. Sánchez, B. Molina, M.A. Alcántara, V. Del Castillo, A. Carnevale, A. González-del Angel, Trisomy of the short arm of chromosome 5 due to a de novo inversion and duplication (5)(p15.3 p13.3), Am. J. Med. Genet. A 136 (2005) 381–385.
[16] T.M. Teslovich, K. Musunuru, A.V. Smith, A.C. Edmondson, I.M. Stylianou, M. Koseki, J.P. Pirruccello, S. Ripatti, D.I. Chasman, C.J. Willer, C.T. Johansen, S.W. Fouchier, A. Isaacs, G.M. Peloso, M. Barbalic, S.L. Ricketts, et al., Biological, clinical and population relevance of 95 loci for blood lipids, Nature 466 (2010) 707–713.
[17] I.M. Heid, A.U. Jackson, J.C. Randall, T.W. Winkler, L. Qi, V. Steinthorsdottir, G. Thorleifsson, M.C. Zillikens, E.K. Speliotes, R. Mägi, T. Workalemahu, C.C. White, N. Bouatia-Naji, T.B. Harris, S.I. Berndt, E. Ingelsson, et al., Meta-analysis identifies 13 new loci associated with waist–hip ratio and reveals sexual dimorphism in the genetic basis of fat distribution, Nat. Genet. 42 (2010) 949–960.
[18] E.K. Speliotes, C.J. Willer, S.I. Berndt, K.L. Monda, G. Thorleifsson, A.U. Jackson, H.L. Allen, C.M. Lindgren, J. Luan, R. Mägi, J.C. Randall, S. Vedantam, T.W. Winkler, L. Qi, T. Workalemahu, I.M. Heid, et al., Association analyses of 249,796 individuals reveal 18 new loci associated with body mass index, Nat Genet 42 (2010) 937–948.
[19] A. Scherag, C. Dina, A. Hinney, V. Vatin, S. Scherag, C.I. Vogel, T.D. Müller, H. Grallert, H.E. Wichmann, B. Balkau, B. Heude, M.R. Jarvelin, A.L. Hartikainen, C. Levy-Marchal, J. Weill, J. Delplanque, et al., Two new loci for body-weight regulation identified in a joint analysis of genome-wide association studies for early-onset extreme obesity in French and German study groups, PLoS Genet. 6 (2010) e1000916.