Universidade de Aveiro Departamento de Química
Ano 2011

Fátima Daniela Teixeira Lopes

Alterações de dosagem no genoma de doentes com atraso mental

Genomic imbalances in patients with intellectual disability

Universidade de Aveiro Departamento de Química
Ano 2011

Fátima Daniela Teixeira Lopes

Alterações de dosagem no genoma de doentes com atraso mental

Genomic imbalances in patients with intellectual disability

Dissertação apresentada à Universidade de Aveiro para cumprimento dos requisitos necessários à obtenção do grau de Mestre em Biotecnologia, realizada sob a orientação científica da Doutora Patrícia Espinheira de Sá Maciel, Professora Auxiliar da Escola de Ciências da Saúde da Universidade do Minho, e da Doutora Odete Cruz e Silva, Professora Auxiliar do Departamento de Biologia da Universidade de Aveiro.

Apoio financeiro da FCT
(PIC/IC/83026/2007) no âmbito do III
Quadro Comunitário de Apoio.

“Sooner or later we will find the cause, and it will be in exactly the same way as we have found genes for other disorders - by lucky coincidence. (…)There isn't anything unique about not having discovered it yet."

John Marius Opitz, 1997

Por este motivo, este pequeno contributo é dedicado aos pais das crianças envolvidas neste estudo.

o júri

presidente

Professor Doutor Jorge Manuel Alexandre Saraiva
Investigador Auxiliar do Departamento de Química da Universidade de Aveiro

Professora Doutora Filipa Abreu Gomes de Carvalho
Professora Associada com Agregação da Faculdade de Medicina da Universidade do Porto

Professora Doutora Patrícia Espinheira de Sá Maciel
Professora Auxiliar da Escola de Ciências da Saúde da Universidade do Minho

Professora Doutora Odete Cruz e Silva
Professora Auxiliar do Departamento de Biologia da Universidade de Aveiro.

agradecimentos

Em primeiro lugar gostaria de agradecer aos meus pais por estarem sempre ao meu lado e me apoiarem em tudo o que sempre precisei.

Ao Albino pelo apoio incondicional, pela paciência, pelo carinho, pela amizade e por acreditar em mim mesmo quando nem eu acredito.

À Patrícia pela oportunidade de fazer parte da equipa em primeiro lugar, pelo voto de confiança na realização deste trabalho, pela oportunidade de realizar o mestrado em conjunto com as minhas funções no ICVS e pela sua orientação. Sinceramente muito obrigada!

Ao doutor Jorge Saraiva e ao doutor João Coutinho por me terem recebido como aluna do mestrado em Biotecnologia.

A todos os elementos do ICVS, em particular aos NeRDs.

Aos meus colegas de laboratório em especial: à Sara, à A. Carvalho, à Anabela, à Ana João, à A. Castro, à Sofia M., à Sofia L., à Sandra e à Susana por toda a ajuda quando precisei, por todos os conselhos acertados que me deram.

To my colleagues at MAF in VU Medical Center, in particular to doctor Bauke Ylstra for receiving me in his laboratory.

A todos os elementos do projecto dos aCGH e do XLMR, assim como à FCT pelo financiamento.

palavras-chave Atraso do desenvolvimento; Atraso mental; aCGH; Alterações do número de cópias

resumo

O atraso mental (AM) é uma doença que apresenta desafios para a vida tanto nas famílias como na sociedade. É caracterizada por um funcionamento sub-óptimo do sistema nervoso central que manifesta limitações quer ao nível da capacidade intelectual quer do comportamento adaptativo e tem uma incidência de aproximadamente 2-3% de nados vivos em todo o mundo. A determinação da causa da doença é importante para o correcto encaminhamento clínico, aconselhamento genético e estratégias de coping a desenvolver pela família.

A técnica de hibridação genómica comparativa (aCGH), também designada de cariótipo molecular (CM), permite medir variações no número de cópias entre o genoma do doente e de um controlo. A totalidade do genoma de um indivíduo é representado num cariótipo virtual de alta resolução, permitindo assim a detecção de alterações submicroscópicas indetectáveis por métodos de cariotipaem standard ou de alta resolução.

Em Portugal, as recomendações para os testes genéticos a realizar num doente com AM indicam o cariótipo tradicional de bandas G como primeira abordagem e, sempre que o resultado seja negativo, a pesquisa relativa às patologias unigénicas mais comuns (Síndrome de X frágil, por exemplo), rearranjos subteloméricos e sondas de FISH específicas. No entanto, estas tecnologias não são aplicáveis para análise de todo o genoma no contexto de diagnóstico de rotina, tanto pela sua baixa resolução (caso do cariítpo convencional) como pela elevada laboriosidade e/ou custos monetários (caso do FISH).

A introdução de aCGH pretende contribuir para a classificação etiológica de grande parte dos doentes com AM, assim como, concluir acerca da utilidade desta tecnologia para o diagnóstico de AM idiopático no contexto clínico.

keywords Development delay; Intellectual disability; aCGH; Copy number variations

abstract

Development delay/Intellectual disability (DD/ID) is a serious and life-long condition which represents a challenge for families and public health services. It is characterized by suboptimal functioning of the central nervous system resulting in limitations both in intellectual functioning and in adaptive behavior, and it is observed in approximately 2-3% of the population worldwide. Establishing the disease etiology is important for clinical management, genetic counseling and coping strategies of the families.

Array-based comparative genomic hybridization technique (aCGH), also called molecular karyotyping (MC), allows us to directly measure genomic copy number variations between the patient and a control DNA. The whole genome of an individual is represented in a high-resolution “virtual karyotype”, allowing the detection of submicroscopic alterations, undetectable by standard or high-resolution karyotyping techniques.

In Portugal, the guidelines for testing patients with DD/ID indicate the G-banding karyotyping in first place and, whenever the result is normal, testing for the most common single gene disorders (Fragile X, for instance), for subtelomeric rearrangements and with specific FISH probes. However, the latter technologies are not suitable for whole genome scans in routine diagnosis, both because of the lower resolution levels (the case of conventional karyotyping), the extensive time consumption and high costs (the FISH case).

The introduction of aCGH should contribute to the etiological classification of a large proportion of the DD/ID patients as well as to conclude about the utility of using these technologies for diagnosis of idiopathic DD/ID in the clinical context.

Publicações

Comunicações Orais

Mafalda Barbosa, Maria João Sá, Cristina Dias, João Silva, Ana Maria Fortuna, **Fátima Lopes**, Patrícia Maciel, Hanna Verdin, Elfride de Baere, Eduardo Silva, Gabriela Soares "*FOXL2 mutations: from 1 base pair to 10 Megabase*". 14ª Reunião da Sociedade Portuguesa de Genética Humana. Coimbra; 18-20 Novembro de 2010; Comunicação oral.

João Silva, **Fátima Lopes**, Gabriela Soares, Mafalda Barbosa, Miguel Rocha, Gisela Barros, Catarina Gomes, Teresa Temudo, Bauke Ylstra, Janneke Weiss, Erik Sistermans, Patrícia Maciel. *"aCGH analysis of Portuguese patients with idiopathic mental retardation".* 14ª Reunião da Sociedade Portuguesa de Genética Humana. Coimbra; 18-20 Novembro de 2010; Comunicação oral.

Apresentações em poster

Carlos Bessa, **Fátima Lopes**, Gabriela Soares, João Silva, Mafalda Barbosa, Miguel Rocha, Margarida Reis Lima, Ana Maria Fortuna, Frederico Duque, Guiomar Oliveira, Carla Marques, Jorge Pinto-Basto, Gisela Barros, Catarina Gomes, Teresa Temudo, Bauke Ylstra, Janneke Weiss, Erik Sistermans, Patrícia Maciel. "*New microdeletion/microduplications detected by aCGH in patients with intelectual disability: seacrh for the key genes*" Conferência Internacional Jerome Lejeune "Genetic intellectual disabilities: progress toward targeted treatments?" Paris 24-26 Março de 2011; Apresentação em poster

Fátima Lopes, Henedina Antunes, Gisela Barros, Bauke Ylstra, Janneke Weiss, Erik Sistermans, Patrícia Maciel. "*Small 3q29 microduplication covering the genes KIAA0226 and MIR922 in a girl with severe obesity, intelectual disability and microcephaly"*. 14ª Reunião da Sociedade Portuguesa de Genética Humana. Coimbra; 18-20 Novembro de 2010; Apresentação em poster.

Patrícia Maciel, Anabela Fernandes, Sofia Moura, **Fátima Lopes**, Sandra Pereira, Bruno Costa, Bauke Ylstra, Janneke Weiss, Teresa Temudo, Teresa Lourenço, José Pedro Vieira. *"Variant Rett syndrome in a girl with a pericentric X-chromosome inversion leading to overexpression of the MECP2 gene in the absence of gene duplication*". 14ª Reunião da Sociedade Portuguesa de Genética Humana. Coimbra; 18-20 Novembro de 2010; Apresentação em poster.

Fátima Lopes, Gabriela Soares, João Silva, Mafalda Barbosa, Miguel Rocha, Gisela Barros, Catarina Gomes, Teresa Temudo, Bauke Ylstra, Janneke Weiss, Erik Sistermans, Patrícia Maciel. "*aCGH analysis of Portuguese patients with idiopathic mental retardation*". Conferência Jaques Monod "Mental retardation: from genes to synapses, functions and dysfunctions", Roscoff, França. Outubro de 2010, Apresentação em poster.

Contents

List of figures

Chapter 1

Chapter 2

List of tables

Chapter 1

Chapter 2

Abbreviations

aCGH – array comparative genomic hybridization
Alu - Short interspersed nuclear elements
AM – Atraso mental
AMPA – Amino-3-hydroxyl-5-methyl-4-isoxazolepropionic acid
AMPAr - Amino-3-hydroxyl-5-methyl-4-isoxazolepropionic acid receptor
ASD – Autism spectrum disorders
bp – Base pair
BAC – Bacterial artificial chromosome
CA – Chronological age
CAMs – cell adhesion molecules
cDNA – Complementary deoxyribonucleic acid
CHARGE - Coloboma of the eye, Heart defects, Atresia of the nasal choanae, Retardation of growth and/or development, Genital and/or urinary abnormalities, and Ear abnormalities and deafness syndrome
CM – Cariótipo molecular
CNV – Copy number variation
CD – Cortical dysgenesis
Cy3 – Cyanine dye 3
Cy5 - Cyanine dye 5
COFS - Cerebro-Oculo-Facio-Skeletal syndrome
DD – Developmental disabilities
DNA – Deoxyribonucleic acid
DQ – Development quotient
DA – Developmental age
DECIPHER - DatabasE of Chromosomal Imbalances and Phenotype in Humans using Ensembl Resources
DGV - Toronto Database of Genomic Variants
ECM - Extracellular matrix
Ephs - ephrins
FISH – Fluorescent in situ hybridization
FOC – Frontal-occipital circumference
FoSTeS - Fork stalling and template switching
FXS – Fragile X syndrome
GABA – gamma-aminobutyric acid
$GABA_A$ – gamma-aminobutyric acid A
ICF - Immunodeficiency, centromeric instability and facial dysmorphisms syndrome
ID – Intellectual disability
ISCA – International Standards for Cytogenomic Arrays Consortium

IQ – Intelligence quotient
LCH – lissencephaly with cerebellar hypoplasia
LCR – Low copy repeat
LINE - Long interspersed nuclear elements
LOH - Loss of heterozygosity
LTD – Long-term depression
LTP – Long-term potentioation
LTR - Long terminal repeats
MA – Mental age
Mb – Megabase
MC – Molecular karyotiping
MCA – Multiple congenital abnormalities
MCPH – Microcephaly
miRNAs – Micro ribonucleic acids
MLPA – Multiplex ligation-dependent probe amplification
MMBIR - Microhomology-mediated break-induced replication
M-FISH – Multiplex fluorescent in situ hybridization
mRNA – Messenger ribonucleic acid
NAHR - Nonalelic-homologous recombination
NGS – Next generation sequencing
NHEJ - Nonhomologous end-joining
NLGNs – neuroligins
NMDA – N-Methyl-D-aspartic acid
NRXNs – neurexins
MPS – Massive parallel sequencing
PSD – Post synaptic density
PCR – Polymerase chain reaction
qPCR – Quantitative polymerase chain reaction
RNA – Ribonucleic acid
RTT – Rett syndrome
SNARE – soluble NSF attachment receptor
SNP – Single nucleotide polymorphism
SKY-FISH - Spectral karyotyping fluorescent in situ hybridization
Kb – Kilobases
XCI – X chromosome inactivation
XLAG - X-linked lissencephaly with absent corpus callosum and ambiguous genitalia
Wnts – Wingless-type signaling proteins
UPS - Ubiquitin proteasome system

General introduction and thesis overview

Genomic imbalances in the genome can occur due to a wide variety of alterations in DNA sequences. Alterations in the dosage of a certain genomic portion are currently accepted as an appreciable cause of a wide spectrum of diseases. This type of genetic alterations can adopt many forms of presentation, ranging from deletions/duplications large enough to be observed by regular karyotyping, insertions, multisite rearrangements or microdeletions or microduplications only detectable with high resolution techniques [1].

Many times, and besides the incredible technical advances in the field, the real functional significance of some alterations is difficult to evaluate. Many times, inside a copy number variable region only a small quantity of alterations turns to be causative of the disease. Furthermore, copy number variations can occur at "gene desert" regions of the genome and still indirectly influence the expression of other genes [2].

Several efforts have been taken in order to fully, or at least better understand, the real impact of this type of genetic variation in human health and disease. Projects such as *The Copy Number Variation Project* (http://www.sanger.ac.uk/humgen/cnv/) intend to help in the clarification of how frequent genomic imbalances really are, their impact in gene expression, mechanism of origin and its contribution for evolution, variability, genetic diseases and complex trait disorders.

Despite the amount of doubts still existing around this type of genetic alteration, its association with diseases such as cancer and developmental disorders is currently accepted. We can say that, so far, all kinds of genomic imbalances were found to be associated with intellectual disability and that at least one type of these alterations were found in each of the human chromosomes [3].

In this dissertation, the author will review the physiological mechanisms thought to be associated with intellectual disability with focus on genomic imbalances as the causative mechanism (chapter 1), present the experimental work and results regarding the presence of this type of alterations in a group of Portuguese patients with idiopathic intellectual disability and a brief discussion of the work (chapter 2).

The results respecting the FISH and MLPA analysis were performed in other laboratory. All the data presented in this work regarding the validation of findings by quantitative real time PCR was performed by a colleague and, consequently, will be not described in detail in the present thesis.

Chapter 1 Introduction

Intellectual dysfunction: pathogenic mechanisms, genetic causes and diagnostic tools

Developmental disabilities

The term “disability” refers, with respect to an individual, to a physical and/or mental impairment that substantially limits one or more major life activities [4]. Developmental disabilities (DD) are a large group of chronic disorders that begin during any step of the development process (including conception, birth and growth). These diseases stem from fundamental deficits in the developing brain due to several factors (genetic, prenatal, perinatal, metabolic and environmental). Prevention and an early diagnosis, as well as intervention, are extremely important to mitigate the personal and socioeconomical impact of these disorders [4, 5].

According to the International Classification of Functioning, Disability and Health of World Health Organization, a person is considered to have a development disability if he/she has impairment in three key components: (I) body functioning and activity (problems in body structure and functioning as a significant deviation or loss); (II) activity (limitations an individual may have in executing activities); (III) participation (restrictions an individual may have in dealing with a life situation). When these tree main categories are impaired the term disability is used [6].

When a child is proceeding normally, the development process is considered normal in four domains: (I) motor development (fine and gross motor domains); (II) speech and language (receptive and expressive domains); (III) social/emotional (combination of other domains such as fine-motor adaptative abilities, communication, cognitive), (IV) cognitive development (visual perception, problem-solving skills, intellectual abilities). Although not all children achieve the predicted milestones at the same time, the manner of progression is predictable in a typical healthy child. For example, the generalized reflexes evolve to complex motor patterns for later voluntary motor development to be achieved [7-9].

The rate of development may measured by the development quotient (DQ) [5]. This quotient measures the progression in a given development domain and it is calculate as follows: DQ = developmental age (DA) / chronological age (CA) X 100. Whenever this value is less than 70 a significant delay is considered to occur in the child. The intelligence quotient (IQ) is a measure of cognitive abilities and it is calculated as follows: IQ = mental age (MA) / chronological age (CA) X 100. A child is considered to present an atypical development whenever a delay, deviation, dissociation or regression is present [5].

Intellectual disability

Intellectual disability (ID) is one of the most frequent and disabling neurological impairments in school-age children [8].

In order to define ID in a child 3 main criteria are necessary: I) significant sub-average general intellectual functioning; II) criterium I needs to be accompanied by limitations in adaptative behavior

in at last 2 of the following skills: communication, self-care, home living, social/interpersonal skills, use of community resources, self-direction, functional academic skills, work, leisure, health and safety; III) the onset of the symptoms should be before 18 years [8,13].

In a more summarized way, general ID can be diagnosed in children below 18 years, with compromised general intellectual functioning (defined by IQ below 70) and impairment in adaptative behavior (how effectively individuals cope with common life demands). Although the observations regarding the adaptative functioning of the child are less objective (since they rely on information provided by family, teachers and medical history) they are extremely useful in assessing the children and recommended by ICD-10 Classification of Mental and Behavior Disorders (World Health Organization) and the DSM-IV Diagnostic Classification (USA) [6].

According to ICD-10, ID is defined as a "condition of arrested or incomplete development of the mind, which is specially characterized by impairment of skills manifested during the developmental period, skills which contribute to the overall level of intelligence, i.e. cognition, language, motor and social abilities". ID can be seen alone or together with other mental and/or physical abnormalities [6].

In the general population IQ in normally distributed around a mean of 100. It is accepted that when an individual presents an IQ lower than 70, this person is classified as intellectually handicapped. The levels of ID are divided in mild ID (IQ between 50 and 69), moderate (IQ of 35-49), severe (IQ of 20-34) and profound (IQ lower than 20) (8,15). The most frequent form of ID are the mild and moderate forms that have a frequency of 2-3% in the population, while 0,5%-1% presents severe to profound ID [8, 7].

The different degrees of ID are stipulated by conventionally standardized tests that can be used together with scales for assessing social adaptation to the surrounding environment. Depending on the severity of ID and of the social rehabilitation of the patient since early life, the ID level of a given individual can change over time, although always in a very poor and unextended way [14, 7].

In table 1.1 are described the expected outcomes for patients with ID according to their IQ level.

Table 1.1 - Classification of ID level according to the IQ range, percentage of individuals affected, correspondent mental age in adults and supported needed in daily living activities. The predicted adult outcome for the individuals in each group is also described (adapted from [5-7]).

Severity Level	% Individuals with ID	IQ Range	Equivalent MA (as adults)	Support needed in daily living activities	Predicted adult outcome
Mild	85%	50-69	9-12 years	Intermittent; Support on "as needed" basis, episodic or short term. Likely to results in learning difficult in school.	Many adults will be able to work and maintain good social relationships and contribute to society. Reads at 4th-5th grade level; able to perform simple multiplications / divisions; writes simple letters, lists; completes job applications; basic independent job skills (arrive on time, stay at task, interact with co-workers); uses public transportation; may qualify for recipes.
Moderate	10%	35-49	6-9 years	Limited. Consistent over time but time limited. Likely to result in marked DD in childhood but most can learn to develop some levels 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. Sight-word reading; copies information (address); matches written numbers to number of items; recognizes time on clock; communicates; some independence in self-care; housekeeping with supervision or cue cards; meal preparation; can follow picture recipe cards; job skills learned with much repetition; uses public transportation with some supervision.
Severe	4%	20-34	3-6 years	Extensive. Regular, consistent, lifetime support. Regular support in at least one aspect such as school, work or home.	Likely to result in continuous need of support and supervision; may communicate wants and needs sometimes with augmentation communication techniques.
Profound	1%	<20	<3 years	Pervasive. High intensity across all environments, lifetime and potentially life-sustaining. Results in severe limitations in self-care, continence, communication and mobility.	Limitations of self-care; continence; communication and mobility; may need complete custodial or nursing care.

The severity of ID can be assessed based on a combination of IQ level, adaptative functioning and intensity of support needed [5]. When the level of ID cannot be reliably assessed but there is a high level of confidence based on the clinical observations, the diagnosis of ID can be made without specifying the severity of intellectual impairment [6]. The differences in the ID prevalence found between studies reflect the variation in the definition used, method of assessment and the different characteristics of the studied populations. Nevertheless, based on the normal distribution of IQ levels in the general population and using two standard deviations below the mean as a cutoff, it is predictable that around 2,5% of the population has ID. Among these 2,5%, around 85% of the individuals are reported to have mild ID, 10% to have moderate ID, 4% to have severe and 1% to have profound ID [6, 7]. ID is more frequent in more males than in females. The ratio of males to females has been substantiated by numerous studies that arrive to the same conclusion: an approximately 30% excess of males being affected by ID [16, 17, 18].

Causes of intellectual disability

The likelihood of identifying a specific etiology increases with the severity of the ID [6].

The most common form of ID (mild) is mainly associated with environmental risk factors and a specific etiology can be identified in less than half of the cases. In contrast, in severe ID, more than 2/3 of the affected individuals are found to have a biological or neurological etiology [7]. The most common conditions identified in children with severe ID include genetic imbalances, single gene disorders, congenital brain malformations, congenital infections, inborn errors of metabolism and birth injury [9, 7].

According to etiology and severity of ID, children present distinct symptoms and signs. Children who have severe ID generally present an earlier symptomatology while children with mild ID are only diagnosed at school age. Among the most common signs that parents report to the clinicians are the following: the child is late to sit, crawl or walk; is late to learn to talk; has difficulties in speech; finds it hard to remember things; has trouble understanding social rules; has trouble seeing the results of his/her actions; has trouble solving problems [8, 7].

Once the ID is suspected by the parents and clinicians, additional clinical history should be obtained and a complete physical examination should be performed. By definition, a diagnosis of ID requires evaluation by qualified examiners using standardized psychometric instruments. The nature of the tests should take into account the mental age of the child and should be culturally sensitive. A workup should also include complete audiologic and vision evaluation in all children [8, 7, 19].

Parents are frequently faced with the need to understand and establish an etiology for their child's condition. In about 75% of individuals with severe ID a biological cause can be found. In many cases disease-specific features may indicate which tests to order and, if more children are planned, genetic counseling and pre-natal diagnosis can be offered to the parents [15, 17 - 18, 20, 21]. Adding the

fact that associated complications can often be expected, etiologic information may suggest that the specific case may be treatable, or allow the development of prevention strategies [5]. Contribution to the research, early intervention for behavioral problems and help in log-term life planning are also factors that contribute to the decision of searching for an etiology of the disease [7, 8].

Some of the known causes of ID include fetal alcohol syndrome, brain malformations, iron deficiency, genetic syndromes and several single gene mutations (such as fragile X syndrome and tuberous sclerosis) [8, 22].

When the clinical characteristics don´t allow the suspicion of a syndrome, a workup of genetic testing, neuroimagin analysis and metabolic screenings is needed. The yield of these tests in dentifying a cause of disease varies depending on the presence or absence of associated symptoms and signs. In developed countries, newborn screening programs generally identify metabolic errors early after birth and the yield of metabolic testing in a child with ID is lower than 1%. Neuroimaging approaches to detect abnormalities in brain have yields of 33% to 63% although they may or may not be helpful in establishing the cause of ID (many times the findings are seen as a consequence rather than a cause) [6]. In the genetic testing field, the outcomes are strictly related with the extent of the investigation, in which cost is an important limitation [19] .

The importance of a diagnosis

In the Portuguese national public health system, when a child is born a detailed clinical evaluation is performed by the obstetrician and then by a pediatrician. Whenever the child is inserted in a functional and caring family or environment, development surveillance is performed by the parents or other caretakers, teachers and health care professionals (usually the family doctors or pediatricians). The observation and summary of the developmental progress of a child has the main goal of identification of an individual who may have a developmental problem [20]. When a suspicion is raised, the child must be evaluated using more structured tests performed by a credited health-care professional. The goal is to identify a specific developmental disorder and its etiology if possible. If the evaluation performed by the clinician yields suspicion for a developmental disorder, a formal and advanced battery of tests is necessary (including neuroimaging, electroencephalography and specific laboratory tests for genetic or metabolic alterations). Consultations with appropriated medical specialties should also be recommended if necessary [10,7].

All children with significant developmental delay or regression should also be referred to a more detailed evaluation by a clinical psychologist who may propose an intervention plan [21].

Pathophysiology

ID is a very heterogeneous group of diseases that can result from genetic and non-genetic alterations (maternal intoxication, prematurity, ischemia, infectious diseases). When the intellectual limitation is the only manifestation of the disease ID is categorized as non-syndromic. On the other hand, when it occurs as one phenotypic feature of a more global clinical presentation, it is called syndromic [22].

The disruption of prenatal and early postnatal brain development can be associated with ID in humans [23]. During development, a strict spatiotemporal control of neurogenesis and cell migration occurs, and neurons form intricate axonal and denditric connections resulting from both intrinsic genetic characteristics and functional cell to cell interactions. Small disruptions in any of these processes during development can lead to cognitive dysfunction in children [24, 23].

Cerebral human cortex development: a matter of fine-tuning regulations

The mammalian cortex is a remarkably complex organ formed by several different types of neurons, oligodendrocytes and glial cells, forming intricate connections with other regions of the nervous system [25].

In mammals, cerebral cortex development is not only a quantitative process. Cortical expansion and development requires a series of tightly regulated procedures at the genetic, molecular, cellular and anatomic levels [26].

In the late eighties, two main theoretical hypotheses regarding the human cortical development were proposed: the protocortex hypothesis and the protomap hypothesis. The protocortex theory stated that at the onset of cortical development, all the cortical neurons had the same potential and that their development pattern resulted from external specific stimulations. On the other hand, the protomap theory stated that neurons, at their birth at the ventricular zone (a pseudostratified neuroepithelium that surrounds the lateral ventricules), already had the genetic instructions required to correctly migrate, connect and establish their final place at the developed cortex. This model also states that the correct road of each neuron is find through their recognition of secreted molecular patterns that acts like “maps” for the guidance of the neuron (protomap) [25, 26]. Currently, the protocortex and protomap hypotheses of cortical development are not seen anymore as alternative and mutually exclusive, instead they are increasingly being viewed as complementary aspects of a single mechanism [27].

Many times disorders of the cerebral cortex, cortical dysgenesis (CD), are associated with intellectual disability, autism, schizophrenia, epilepsy and focal or widespread neurological deficits [28]. Currently it is known that external injuries (chemical or physical) during embryonic development can affect the development by deregulating molecular patterns at crucial time points.

Cortical malformations may arise in isolation, together with other brain malformations or occur as a component of a generalized malformation syndrome. The cerebral cortex is the outer layer of the brain and is a complex laminated structure composed of six organized layers of neurons, nerve fibers and supporting cells. During development, the brain transforms from a smooth hollow tube to a complex organ with a surface full of folds (gyri) and fissures (sulci) [26, 27]. The neurons present in the cortical layer are not born there, but instead are generated from neuronal precursors in the germinal zone. As a consequence, they have to migrate along specialized radial-glial cells to their final destination. It is currently believe that this complex process is genetically regulated (protomap theory) and that even the smallest deregulation may result in significant cortical malformations and account for the ID phenotype in many patients [30].

It is currently accepted that the classification of cortical developmental abnormalities should take into account pathological, genetic and neuroimaging features during the three main stages of cortical development: neuronal proliferation, growth and differentiation (neurogenesis), neuronal migration and cortical maturation and folding (organization). Whenever damage occurs in one of these 3 steps of cortical development, a malformation will occur [31] .

Malformations associated with abnormal neurogenesis

During embryonic development, the first formed neurons arise from two different daughter cells: one daughter cell gives rise to a neuron that will migrate to the cortex, while the other daughter cell continues to proliferate as a stem cell or dies [24]. Nowadays it is known that, although in the cerebral hemispheres the majority of neurogenesis occurs in the first half of gestation, neurogenesis also occurs in the olfactory bulb, sub-ventricular zone and hippocampus in adults [24, 27].

Excess or defects in germinal epithelial proliferation can lead to macro- or microcephaly. This classification enrolls a heterogeneous group of disorders that can be de novo or familial, often associated with an increase incidence of cognitive impairments [24, 32].

Regarding the malformations associated with alterations in neuronal proliferation, they can be divided in microcephaly, megalencephaly, hemimegalencephaly and cortical dysplasia.

Microcephaly (MCPH) is characterized by a reduced frontal-occipital head circumference (FOC), more than 3 standard deviations below the mean of age and sex-matched controls [25]. MCPH can be classified as primary or secondary (acquired). Patients with primary MCPH usually display small but architecturally normal brains or with mildly simplified gyral patterns, and have mild ID. This for or MCPH is distinct from holoprosencephaly and lissencephaly, two forms of microcephaly associated with cortical malformations [33, 34]. In secondary MCPH the cause in presumed to be traumatic, in primary MCPH the cause is genetic and heterogeneous [33], [25]. Mutations in several genes have been identified in primary MCPH, all of them associated with cell division and cell cycle regulation [34]. In figure 1.1 is represented a scheme for the different types of microcephaly.

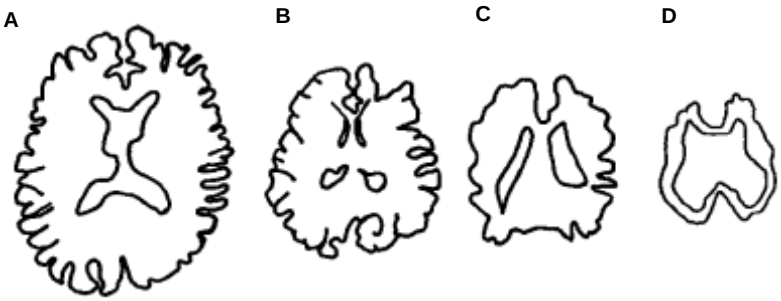


Figure 1.1 – Schematic representation of microcephalic brains. (A) Normal brain; **(B)** microcephalic brain with conserved architecture; **(C)** microcephalic brain with a simplified gyral pattern; **(D)** microlissencephalic brain. Adapted from [34].

Megalencephaly refers to an abnormal large head (macrocephaly) due to the enlargement of the brain parenchyma. This condition is also related with disturbances in cell proliferation. Non-syndromic macrocephaly refers to a condition in which the brain enlargement is the predominant abnormality and not associated with any other physical alterations. Minor craniofacial changes can be present but they are a consequence of the enlarged cranial cavity [36]. The classification of syndromic megalencephaly is used when a large brain size is associated with other significant physical and behavior abnormalities. Syndromic megalencephaly should also be distinguished from other genetic syndromes in which an enlarged brain size is an occasional but not consistent finding (autism, Fragile X Syndrome and Sotos syndrome, for instance) [31, 36].

Hemimegalencephaly is the classification attributed to the enlargement of only one hemisphere. Histopathologicaly presents similar features to the ones of cortical dysplasia, with the difference that is located in one side of the brain specifically. Refractory epilepsy, hemiparesis and ID are common clinical findings among hemimegalencephaly patients and not megalencephaly [31]. Hemimegalencephaly has been observed in patients with tuberous sclerosis complex, Neurofibromatosis type I and linear epidermal nevus syndrome [34, 36].

Cortical dysplasia is characterized by the disturbance of the cortical lamination and the term is applied to focal areas of cortical malformations in which neuronal orientation is disrupted [34]. In this type of malformations dysmorphic neurons are seen; these may be undifferentiated balloon-shaped cells or giant cells, both resulting from abnormal regulation of cell growth [34]. Cortical dysplasia is recognized as the major cause of intractable epilepsy in childhood and the second most common etiology in adult epilepsy patients recommended for surgery [31].

Malformations associated with abnormal neuronal migration

After neurogenesis, postmitotic neurons are organized in columns and migrate away from the ventricular zone in a radial pattern [24]. Neuronal migration consists in the movement of those

postmitotic neurons from the ventricular zone to their final destination in the cerebral cortex [37]. The strict regulation of the migration pattern and timing is essential for the correct development and functioning of the intellectual functioning in an individual [30].

Neuronal migration disorders are a heterogeneous group of disorders of the nervous system development and are considered one of the most common causes of neurologic, development and epileptic seizures in childhood [28, 37]. This category of disorders includes lissencephaly and heterotopia.

Lissencephaly (which literally means “smooth brain”) refers to the occurrence of a smoother brain surface without gyri or sulci and it is associated with severe ID and refractory epilepsy. In some cases it is possible to observe, in the same patient, brain region without gyri (agyria) and brain regions with broad and large gyri (pachygyria) [31, 34]. At least five distinct types of lissencephaly exist: type I lissencephaly (also known as classic), type II lissencephaly (also called Cobblestone lissencephaly), lissencephaly X-linked with agenesis of the corpus callosum (XLAG), lissencephaly with cerebellar hypoplasia (LCH) and microlissencephaly [34]. A schematic representation of a lissencephalic brain is presented in figure 1.2.

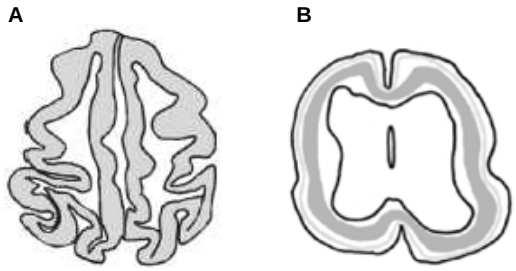


Figure 1.2 – Schematic representation of a lissencephalic brain (B) and normal brain (A) (adapted from [38]).

Classic lissencephaly causes a combination of agyria and pachygyria and presents a poorly structured cortex with only four immature layers instead of the six fully organized layers present in the well-developed brain. The cortex is extremely thick as well as the underlying white matter that, sometimes, can contain single neurons or small heterotopias [36, 32].

Cobblestone lissencephaly (or type II lissencephaly) is a complex brain malformation characterized by global disorganization of cerebral organogenesis [38]. In this form of lissencephaly the surface of the cortex is not necessarily smooth but, instead, frequently shows a broad smooth paving that looks like cobblestones. As in type I lissencephaly, the cortex layer is thickened with no recognizable layers. However, in this type of lissencephaly the ratio of cortex to white matter in approximately 1:1 (similar to the one present in the normal adult brain) and not 4:1 as in lissencephaly type I [32].

X-linked lissencephaly with agenesis of corpus callosum (XLAG) includes thickened cortex with gyral malformations that are more severe in the posterior than anterior brain regions together with agenesis of corpus callosum and ambiguous genitalia [38].

Lissencephaly with cerebellar hypoplasia (LCH) is the term adopted to classify the type of lissencephaly that is not included in either classic or cobblestone type. It is characterized by cerebellar underdevelopment ranging from vermian hypoplasia to total aplasia [37, 38].

Microlissencephaly is characterized by the presence of a severe microcephaly in the patients and it is caused by an abnormal proliferation together with a defect in neuronal migration [38]. A summary of the genes known to be associated with each type of lissencephaly as well as the syndromes commonly presenting each type is present in table 1.2.

Table 1.2 - Classification of lissencephalies, syndromes and genes associated (adapted from [38]).

Type of Lyssencephaly	Associated syndrome	Genes involved
Type I Lissencephaly	• Isolated lissencephaly sequence • Miller-Dieker syndrome	• *LIS1, DCX, TUBA1A* • *LIS1* and *YHAWAE* deletion
Cobblestone Lissencephaly	• Walker-Warburg syndrome • Muscle-Eye-Brain disease • Fukuyama congenital muscular dystrophy	• *POMT1, POMT2. FKTN*, *FKRP, LARGE* • *POMGnT1* • *FKTN*
Lissencephaly X-Linked with agenesis of corpus callosum (XLAG)	• Lissencephaly	• *ARX*
Lissencephaly with cerebellar hypoplasia (LCH)	• Lissencephaly	• *RELN, VLDLR*
Microlissencephaly	• Norman-Roberts syndrome • Barth syndrome • Primordial osteodysplastic dwarfism and microcephaly (MOPD type 1)	• Not described

Heterotopia is a neuronal migration disorder characterized by the collection of grey matter in inappropriate places. There are three main groups of heterotopia: periventricular (the most frequent form), subcortical and leptomeningeal [30, 32].

Subcortical band heterotopia (or double cortex) (figure 1.3) refers to the occurrence of a second cortex layer. In this pathology bands of gray matter are interposed with bands of white matter between the cortex and the lateral ventricles [28, 30]. In other forms of heterotopias (that are not band heterotopia), neuronal migration is only affected in subsets of neurons along the cortex. These alterations may be found in patients with ID varying from mild to severe [31].

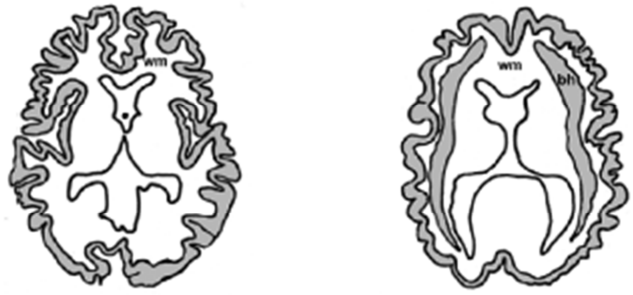


Figure 1.3 – Schematic representation of subcortical laminar heterotopia. A normal brain **(A)** and a brain with subcortical heterotopia **(B)** where it is possible to see a second layer of grey matter (adapted from [37]).

Malformations associated with abnormal cortical organization

When neurons finally reach their destination in the cortex, it is necessary that they both send axons and receive dendrites and synapses with other neurons [25]. Maturation of the cortex includes not only the establishment of intercellular connections but also the acquisition of a population of glial cells in order to unpack the closely arranged immature neurons [34]. Cortical neurons also maturate in a specific phenotype thought the production of appropriate neurotransmitters and receptors. Similar to the previous stages of cortical development, in this phase there is a balanced molecular regulation where cells must be able to receive afferent connections, target efferent pathways and develop specialized synaptic connections. Therefore, any disturbance in synaptogenesis, either by a delaying in the onset or by preventing the formation of the necessary number of synapses, can have a dramatic effect in the future cognitive function of an individual [24], [34].

Polymicrogyria is a cortical malformation characterized by an irregular brain surface with an excessive number of small and partly fused gyri separated by shallow sulci. Two types of polymicrogyria can be identified histopatologically: simplified four-layered form (layer of intracortical laminar necrosis with alterations of late migration and postmigration disruption of cortical organization) and an unlayered form (the molecular layer is continuous and does not follow the borders of convolutions. Polymicrogyria can also be focal or diffused, bilateral or unilateral [36]. In this type of malformation, convolutions and the neurons below have radial distribution while laminar organization is absent) [28], [29], [30]. Because of its heterogeneity polymicrogyria has unknown incidence. Almost all the children with polymicrogyria have a high risk of developing epilepsy. Among other characteristics are global development, motor and language delay, and mild to moderate ID [36].

Schizencephaly is one of the most common cortical malformations (prevalence of 1.5/100 000 newborns) and occurs when a deep vascular insult reaches the white matter and sometimes even reaches the ventricle, leading to the formation of a fissure [31], [34]. The cleft spans the cerebral hemisphere from the pial surface to the lateral ventricle and lined by cortical gray matter. Cleft

organization varies widely but the perisylvian region is more frequently involved [37]. In patients, the extent of the malformation correlates with the severity of the clinical manifestation either in the cognition and motor areas. When the fissure is open, patients have mild to moderate ID and hemiparesis; patients with bilateral fissures present severe ID and severe motor abnormalities such as quadriparesis. Several types of seizures have been reported in patients with schizencephaly, such as tonic-clonic, partial motor and sensorial seizures, being usually resistant to medical therapy [34, 37].

Synaptogenesis

The points of contact that allow exchange of information between neurons are called synapses. The synapse is the most fundamental mean of transmission in the nervous system and can be divided in two types: electrical or chemical. Electrical synapses are less common than chemical synapses in the brain and occur by means of current flow through gap junctions that connect the cytoplasm of pre- and postsynaptic cells. This type of transmission is an extremely rapid process [39]. The chemical synapse represents a specialized functional and morphological structure where a presynaptic neuron communicates with a postsynaptic one. For most neurons, in a chemical synapse a set of small molecules (called neurotransmitters) mediate information transfer between neurons. The presynaptic neuron is responsible for the controlled release of neurotransmitters that bind and activate specific receptors proteins located in the pre and postsynaptic membranes, triggering the opening of ion channels and activation of signaling cascades [40]. The process that determines how, when, were and with who a synapse will be established is called synaptogenesis.
During brain development, when neurons arrive at the cortical layer they start to extend apical and basilar dendrites followed by oblique branches of the apical dendrite. Synapses appear in the human cerebral cortex in the third month of gestation development. The number of synapses increases during the following months together with the elaboration of the dendritic spines and the specificity of synaptic connections appears to be determined by molecular clues [24]. The number of cortical synapses increases until eight months of life in the striatum and until 24 months of life in the cortex. The specificity of the synaptic connections is determined by molecular clues and the existing molecules in the neuronal surface and surrounding area determine the linking of axons to specific neurons by chemoaffinity [24]. Moreover, several factors secreted by neurons and glial cells and cell-adhesion molecules (CAMs) are involved in the stimulation of axon growth and guidance. Synaptogenesis requires the activation of genes coding for synaptic proteins as well as elements involved in vesicle formation, accumulation and trafficking of proteins complexes involved in both pre and postsynaptic levels [20]. We can even compare the process of forming a new synapse with the guidance patterns seen in the neuronal migration process from the ventricular zone to the cortex: it involves coordinated cell morphologic and structures changes, intracellular signaling cascades and protein complexes remodeling; all these processes don´t occur in an isolated form

and there are also cell intrinsic determining factors that contribute to the pre-establishment of long before the arrival of the axon to the synaptic site, as well as to the prohibit formation of inappropriate synapses [40]. Additionally, there is a period that follows the synaptic overgrowth where a massive pruning of synaptic connections occurs [41]. Even though it is amazing how the destruction of such a massive amount of synapses doesn´t have a negative effect in brain function, this mechanism is indeed essential for the correct establishment of functional and fully developed memory mechanisms through the elimination of weaker synapses [42].

Axon guidance and synapse formation

The correct target of axons to their final destination is an essential process of the correct neuronal network formation. Axon guidance requires the action of receptor proteins located in the surface of the growth cone and proteins located in the cell surface of cells found in the migration pathway of the target field [40]. The creation of a connection between the ligand and receptors molecules of the growth cone and target field molecules, usually targets the activation of signaling cascades that induce complex cell morphological changes in the axon growth cone, making it move forward (attraction) or away from (repulsion) the guidance molecular clue, either by contact (contact attraction and repulsion) or by chemical signaling (chemical attraction and repulsion) [43]. The morphological changes required for the migration of the axon, as well as for the formation and plasticity of dendritic spines, are mediated by the correct assembly and disassembly of the F-actin cytoskeleton [44]. In axon growth, the rapid extension and retraction of F-actin cytoskeleton is mediated by the search for local molecular clues. However, protrusions arrive from the dendritic shafts and “evaluate” the surrounding space in order to search for presynaptic axons “willing” to establish a synaptic [40]. An illustrative picture scheme of the processes described above is presented in figure 1.4.

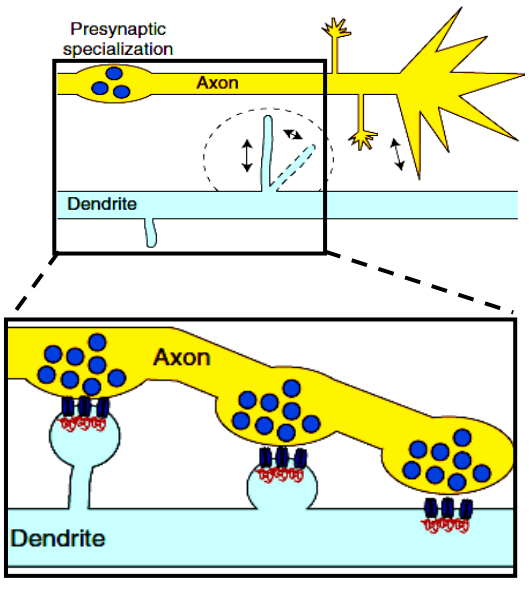


Figure 1.4 - Schematic representation of an initial stage of synaptic formation. In yellow is represented an axon and in blue a nearby dendrite. These two cells interact via cell-to-cell contacts through collateral axon branches and dendritic filopodial extensions. In some regions of the axon may happen that there are pre-established synaptic specialization regions that mark the location of a future synapse. Whenever a stable contact is established proteins, vesicles and ion channels both at presynaptic and postsynaptic sites are recruited to the synaptic site in order to mature and definitely establish a functional synapse (adapted from [40]).

After the establishment of synapses there is a period of synaptic rearrangements that has a fundamental role in the future normal neural functioning. During this period, weaker connections are eliminated while the stronger ones are reinforced trough competition for trophic substances. A series of synaptic rearrangements occur as a response to neuronal activity. The most ilustractive case is the developing of the visual cortex. In this area of the cortex, cells from layer IV are arranged in ocular dominant columns. If the normal pattern of neural activity is interrupted during this crucial period in the first few months of life by closing one eye, the inputs into the mature cortex are permanently altered, resulting in abnormal binocular vision and decreased acuity for the rest of life. However, if the eye is closed after this critical period, no damage is created in the adult visual cortex activity [24]. These findings support the idea that the development of connections at neuronal level accommodates inputs from a combination of spontaneously generated and experience-dependent events [20]. All these structural rearrangements that occur during cortex maturation seem to be related with the balance between excitatory and inhibitory neurotransmission as well as with activity-dependent forms of synaptic plasticity such as long-term depression (LTD) and long-term potentiation (LTP). Mechanisms involved in the transformation of translating activities into structural relevant changes are also considered to play a role in the establishment of synapses during development. One of the proposed mechanisms is the local release of neurotrophins such as BDNF (Brain Derived Neurorophic Factor) that may have a role in controlling neuronal survival and differentiation [20].

Besides the general proposed mechanisms of synaptogenesis, it is currently known that different synapse types have different formation and plasticity mechanisms. The most extensively studied cases are the formation of the cholinergic neuromuscular junction synapses and the glutamatergic axo-dendritic synapse [40].

In the basis of learning and memory is the information storage in the brain by strengthening or weakening already existing synapses by creating or eliminating dendritic spines, which leads to the formation of elimination of synaptic sites [44].

Axon growth and synapse stabilization – the role of the cytoskeleton

Axon guidance and synaptic maturation involve partially similar processes and shared molecular elements. During axon elongation, actin filaments move towards areas of attraction, signaled by molecules that not only regulate the straight movement of the axon but also condition the turning of the axons. On the other hand, synapse formation involves the transformation of an undifferentiated region of the presynaptic plasma membrane into a highly specialized complex [40].

Dendritic filopodia are dynamic protrusions that are thought to play an active role in synaptogenesis and serve as precursors to spine synapses. These structures cover the dendritic space, increase the probability that this dendritic region will interact with an axon and increase the number of

synapses formed. Indeed, filopodial motility is an essential aspect of the complex process of synapse creation and facilitates proper synapse density [45].

Once a synaptic site is determined, a complex chain of molecular partners must be recruited in order to form a stable and efficient synapse. At the presynaptic level, synaptic vesicles, transmembrane receptors, ion channels, vesicle docking and mitochondrial proteins are allocated, while at the postsynaptic region, signaling proteins, ion channels, scaffolding proteins and translational machinery are recruited [40].

BDNF/ TrkB

BDNF and its receptor, tropomycin-related kinase B (*TrkB*), are important molecules for axon guidance, synapse formation and plasticity[46]. At the early stages of synaptogenesis, these two molecules play an important role in the determination of dendritic spine density and morphology in a variety of neuron types, as well as in the determination of the synapse number. This pathway is thoughy to regulate synapse number in early development by stimulating filopodial motility. At more developed and mature synapses, *BDNF/ TrkB* signaling pathway function is more restricted to the regulation of synapse morphology and plasticity [40].

Ephrins and Ephs

Ephrins are cell-surface expressed molecules recognized by Eph receptors. Together, these two types of molecules play a crucial role in guidance of pathfinding axons [47]. The ephrin receptor EphB2 (*EphB2*) is required not only for filopodial motility and increased frequency of axon-dendritic contacts, but also for stabilization of synaptic contacts trough cell to cell adhesion mechanisms. *EphB2* has an extracellular domain function that may function as a contact stabilizing protein. In fact, the extracellular domain of *EphB2* recruits NMDA receptors and is implicate in dendritic formation and increased number of synapses. Also, *Ephrins* and *Ephs* seem to be related with presynaptic maturation and development as well. In summary, *Ephrins* and *Ephs* are currently associated with synapse formation either by regulation of cell to cell adhesions, activation of cytoskeleton, recruitment of proteins and signaling activation [47].

Wnts

The WNT signaling members are a family of highly conserved developmental control genes that play important roles in embryonic development as well as in neuronal development and plasticity in the nervous system. At least two Wnt proteins have been associated with functions at synaptic levels - Wnt-7a and Wnt-5a. Wnt-7a stimulates presynaptic protein clustering and facilitates

neurotransmitter release while wnt-5a modulates glutamatergic synaptic transmission trough a postsynaptic mechanism [40, 48].

Semaphorins, Neuropilins and Plexins

Semaphorins are a large family of highly conserved proteins that play a crucial role in axon guidance during synaptogenesis. Neuropilins and plexins are transmembrane proteins that, for being semaphorin receptors, are also involved in neuronal development [49].
So far, there are 3 main semaphorin classes involved in synaptic development: class 3 semaphorins (which promote axonal repulsion or attraction), class 4 semaphorins (regulators of synaptogenesis) and class 5 semaphorins (which regulate synaptic density). Animal models lacking semaphorins, neuropilins and plexins have altered patterns of spine density and sizes. These proteins are thought to re related neurodevelopmental disorders, such as intellectual disability [40, 49].

NOGO Receptor

NOGO molecules constitute a group of myelin-associated inhibitors that don´t belong to any of the known families of axon guidance molecules. NOGO ligands or receptors were originally described for their function in blocking axon outgrowth after spinal cord injury but were recently associated with hyppocampal synaptic plasticity [50].
In a mouse model, it was described that nogo-66 receptor 1 (*NgR1*) interacts with fibroblast growth factor 1 and 2 (*FGF1* and *FGF2*, respectively) and that, whenever deregulated, *NgR1* inhibits *FGF2*-elicited axonal branching. Loss of NgR1 results in altered spine morphology along apical dendrites in hippocampus [50, 40].

Synapse structure and function

Chemical synapses regulate the electrical communication within neurons and pass information from presynaptic axon terminals to postsynaptic dendritic regions. Most excitatory synapses in the brain are formed at tiny dendritic protrusions called dendritic spines [44].
The molecular architecture of chemical synapses consists of pre-synaptic axon terminals harboring synaptic vesicles and a post-synaptic region (on dendrites) containing neurotransmitter receptors. The presynaptic and postsynaptic sites are separated by the synaptic cleft (10 to 25 nm) and a variety of cell adhesion molecules hold them together at the proper distance [28, 11].
On the presynaptic region neurotransmitters, such as glutamate or ɣ - aminobutyric acid (GABA), are produced and stored in synaptic vesicles at the presynaptic terminals. When a nerve impulse travels through the axon of the presynaptic neuron, it cannot cross the synaptic cleft in order to communicate with the next neuron [52]. A regulated exocytosis pathway is necessary for synaptic

vesicles to make contact with the plasma membrane, where vesicle fusion and exocytosis of the neurotransmitter occur. The vesicles of this pathway undergo exocytosis only in response to a particular stimulus, such as an increase in calcium ion concentration [29, 30]. The docking and fusion of the vesicles of the membrane is controlled by the soluble N–ethylmalei–mide–sensitive factor attachment protein receptor (SNARE) complex. Many other presynaptic proteins play a role not only in synaptic vesicle fusion but also in other steps of synaptic vesicle trafficking such as targeting, docking and priming [53].

Information about the mechanisms of synaptic vesicle docking and fusion are the key to understand synaptic transmission itself and also to discover the transmission modifications that may play a role in synaptic plasticity, learning and memory [54].

In the postsynaptic structures, the dendritic spines, is located the postsynaptic machinery. This includes a series of neurotransmitter receptors (glutamate receptors), postsynaptic density (PSD), actin cytoskeleton and a wide variety of membrane-bound organelles (mitochondria, endoplasmatic reticulum and endosomes). There are two types of postsynaptic receptors that recognize neurotransmitters: (I) ionotropic receptors that are ligand-gated ion channels (AMPA-type [amino-3-hydroxyl-5-methyl-4-isoxazolepropionic acid] and (II) NMDA-type [N-methyl-D-aspartate], that lead to excitatory synapse transmission. An inhibitory synapse is present when an interaction between GABA and $GABA_A$ receptors allows an influx of negatively charged chloride ions. Precise mechanisms must exist to maintain the balance between the excitatory and inhibitory synapses (E/I balance). The presence of the postsynaptic density (PSD) is characteristic of excitatory synapses. Glutamate receptors and PSD proteins play an important role in excitatory synaptic plasticity [52].

Numerous regulatory pathways have been found to participate in an ever-growing signaling pathway network in dendritic spines.

Dendritic spine morphology and neuronal plasticity

Dendritic spines are small protrusions that receive input from a single excitatory presynaptic terminal. Spine density varies with the type of neuron where it is located, but in the majority of neurons they occur at a density of 1-10 spines per micrometer of dendritic length [44].

Dendrites of neurons are covered by small protrusions known as dendritic spines. These structures are the main site of synaptic input in neurons and neurotransmitter receptors (localized to the surface of dendritic spines in the PSD). The classical structure of a spine consists of a head that is connected to the dendritic shaft by a narrow neck and can be divided in three different basic compartments: (I) a delta-shaped base at the junction with the dendritic shaft, (II) a constricted neck and (III) a bulbous head that contacts with the axon. In addition to this general morphology, there are several categories of spines based on their shape and size: thin (filodopia-like protrusions), stubby (short spines without a well-defined neck) and mushroom shaped (spines with a large bulbous head) [44, 45]. One interesting finding about dendritic spines is that they are not static

in time, but change their morphology continuously even in adulthood. The shape of the spines is strongly related with maturity of the synaptic connection and it is currently though that spines don´t present the same shape throughout their entire life, but instead, there is a continuum change of shape between the categories related above (11,35,38). The size of the spine head is related with the size of the PSD; for instance, mushroom shaped spines have larger and more complex PSDs with a higher density of glutamate receptors. Genes that encode factors involved in spine structure and organization have been found to be altered in several brain diseases [52]. Adjacent to the PSD is a region specialized in the endocytosis of postsynaptic receptors. The main function of this region is to capture and recycle the synaptic pool of mobile AMPA-type glutamate receptors required for synaptic potentiation [44].

Independently of the factor(s) that contributed to the disease, abnormalities in dendritic spine structures and morphology are often associated with ID [56].

Cytoskeleton organization in dendritic spines

Since the structure and dynamics of dendritic spines are determined by the underlying cytoskeleton, mutations in genes involved in the organization of these structures could be the cause of the ID phenotype in patients. The actin cytoskeleton, which controls neuronal connectivity and synaptic function, is crucial for morphological differentiation, cell polarity, neurite outgrowth, dendritic morphology, synapse formation, synapse plasticity and protein transport [26,29,31]. The structure of dendritic spines achieved during brain development is essential for the function of mature synapses as well as for the formation of new ones. When an excitatory transmission occurs, the actin network mediates structural changes in dendritic spines [53].

It is currently assumed that there is a strong relation between spine morphology, synapse strength and the actin cytoskeleton. The volume enlargement seen in dendritic spines during long LTP and the decrease during LTD is an activity-dependent spine structural plasticity caused by reorganization of the actin cytoskeleton. Additionally, an ever increasing number of ID-associated genes have been described many of which are components or somehow related with actin remodeling and cytoskeleton organization [57]. The function of actin filaments in developed synapses is to stabilize postsynaptic proteins and modulate spine head structure in response to signaling. A network of long and short branching actin filaments is present in the spine neck, while a series of short branched filaments are localized on the spine head just underneath the PSD [44].

Actin filaments are polar structures that grow more rapidly in one end (barbed end) than in the other (pointed end). In the migrating cell, the barbed end pushes the cell membrane and inducing changes in the shape of the cell that look like filament extensions. In the spine head, a branched actin filament network and capping proteins are present in great amount. In the spine neck and base, there is an actin network composed by the combination of branched and intersecting linear filaments [57].

The first cytoskeleton genes to be associated with ID were members of the Rho GTPase pathway, frequently associated with the cytoskeletal regulation. Rho GTPases are guanine-nucleotide-binding proteins that act as recycling molecules, switching between an active GTP-bound and an inactive GDP-bound state. The proteins belonging to this cascade are best known for their effects on the actin cytoskeleton and microtubule organization [58]. Rho GTPases include Ras homologous member A (RhoA), Ras-related C3 botulinum toxin substrate 1 (Rac1)) and cell division cycle 42 (Cdc42). Studies in several animal models have revealed the implication of these Rho GTPases in dendritic and spine growth and morphogenesis. RhoA, Rac1 and Cdc42 regulate microtubule growth and stabilization, actin polymerization and assembly, myosin contractibility and local protein translation through interaction with FMRP protein [20].

Table 1.3 lists the most widely studied genes from the Rho GTPase pathway associated with ID or ID-related syndromes.

Table 1.3 - Rho GTPase pathway associated genes involved in ID or ID-related phenotypes (adapted from [58, 20].

Gene	Locus	Protein	Function	Associated phenotype	Reference
OPHN1	Xq12	Oligophrenin-1	OPHN1 protein negatively regulates RhoA, Rac1 and Cdcc42	ID	[58]
PAK3	Xq22	PAK3	PAK3 protein mediates Cdc42 and Rac1 effects on the cytoskeleton	ID	
ARHGEF6	Xq26	ARHGEF6	ARHGEF6 protein interact with PAK kinases and is essential for spine morphogenesis	ID	
MEGAP	3p25	MEGAP	MEGAP is a GTPase activating protein that acts towards both Rac1 and Cdc42	3p deletion syndrome	
LIMK1	7q11	LIMK1	LIMK1 is a Ser/Thr kinase acting downstream of Rac1/Cdc42	Williams syndrome	

Presynaptic vesicle cycling

The electric signal that is travelling in a neuron is not able to cross the synaptic cleft by itself. As a consequence, the signal translation is carried by neurotransmitters that are transported by synaptic vesicles. These structures make contact and merge with the presynaptic plasma membrane, where the fusion and exocytosis occurs [52]. Even though a large catalog of proteins known to be involved in vesicle synthesis, targeting, mobilization, fusion and recycling already exists, there is still a large lack of information regarding the exact mechanisms by which synaptic vesicles are regulated. In the centerpiece of these hypothesis is the SNARE-mediated membrane fusion and the proteins associated with remodeling and assembly of the SNARE complex [53].

The essence of neurotransmission relies on the fusion of the vesicle and plasma membranes. The neuronal SNARE proteins VAMP2 (vesicle associated membrane protein 2), syntaxin 1 and SNAP-25 (synaptosome-associated protein of 25 kDa) are essential for membrane fusion [53]. Polymorphisms in SNAP25 gene have been associated with attention deficit and hyperactivity disorder (Feng et al, 2005). *STXBP1* (syntaxin-binding protein 1) is the human homologous of rat n-Sec1 gene. This gene is a neuronal specific syntaxin-binding protein that plays a role in in the regulation of synaptic vesicle docking and fusion. STXBP1 is also a regulatory protein of VAMP2, syntaxin 1 and SNAP-25 proteins. Mutations in *STXBP1* gene were found in patients with autosomal dominant ID and epilepsy [59].

The trafficking cycle of synaptic vesicles is regulated by Rab GTPase activity. Three major proteins involved in this process have been associated with ID: the *GDI1* (GDP dissociation inhibitor-1 gene), that regulates the GDP-GTP exchange reaction of members of the rab family, the IL1RAPL (interleukin 1 receptor accessory protein-like 1) that interacts with neuronal calcium sensor-1 (NCS1) through its specific C-terminal domain and plays a role in the regulation of calcium-dependent exocytosis and *RAB3GAP1* (RAB3 GTPase-activating protein, catalytic subunit) that specifically converts active RAB3-GTP to the inactive form RAB3-GDP and regulate glutamate release [20, 60, 61].

More recently, mutations in genes encoding components of the synaptic vesicle or elements necessary for its formation were found in ID or ID associated syndromes. In 2004, mutations in synapsin 1 gene (*SYN1*) gene, a neuronal phosphoprotein associated with the membranes of small synaptic vesicles, were found in members of families with epilepsy and variable learning disabilities and behavioural disorders[62]. *SLC6A8* (solute carrier family 6, neurotransmitter and creatine transporter, member 8) gene has also been associated with X-linked ID with high levels of creatine in urine. Creatine transport has important roles in the maintenance of energy supply to the cell and resonance spectroscopic imaging studies in ID patients showed absence of creatine in the whole brain [63, 64]. *SAP102* (synapse associated protein 102), also known as *DLG3*, belongs to the subfamily of the membrane-associated guanylate kinase (MAGUK) protein family and is an important component of specialized cell junctions. Mutations in in the human *DLG3* gene were found in families with moderate to severe X-linked ID [65].

Organization of postsynaptic protein complexes and cell adhesion molecules

Cell adhesion molecules (CAM) such as neurexins (NRXNs) and neuroligins (NLGNs) are involved in the formation of functional presynaptic regions specialized in vesicle fusion to the plasma membrane and correct release of the neurotransmitters to the synaptic cleft. NRXNs are presynaptic receptors while NLGNs are the ligands of NRXNs located in the postsynaptic side [52, 20]. Both these molecules are thought to be crucial for synapse maturation but not for their initial development. Mutations in *NLGN4* (neuroligin 4) gene were found in patients with ID and/or ASD. NLGN4 is involved in formation of active regions at presynaptic terminals through interactions with

its presynaptic receptor β-neurexin [20]. More recently, mutations in neurexin 1 and 2 were described in patients with ASD and schizophrenia [66].

During postsynaptic differentiation a series of scaffolding proteins of the membrane-associated guanylate kinase (MAGUK) family are required. Among these, we can refer *PSD95* (postsynaptic density 95, also known as *DLG4*) and SAP102, that are required for recruitment of NMDAR. *SAP97* (synapse-associated protein 97, also known as *DLG1*) is required for the recruitment of AMPAr. MAGUKs play a role in the determination of the number of glutamate receptors at the synapse as well as the synaptic trafficking of these receptors during the morphological changes that are related with synapse plasticity [20]. One of the genes from this group of proteins involved in ID is the AMPAr subunit GluR3 (glutamate receptor 3) that was found to be mutated in X-linked ID [67].

Shank proteins are multidomain scaffold proteins of the postsynaptic density, connecting neurotransmitter receptors and other membrane proteins with signaling proteins and the actin cytoskeleton. By virtue of their protein interactions, Shank proteins assemble signaling platforms for G-protein-mediated signaling and the control of calcium homeostasis in dendritic spines. In addition, they participate in morphological changes, leading to maturation of dendritic spines and synapse formation. The importance of the Shank scaffolding function is demonstrated by genetically determined forms of ID, which may be caused by haploinsufficiency for the *SHANK3* (SH3 and multiple ankyrin repeat domains 3) gene and *SHANK2* (SH3 and multiple ankyrin repeat domains 2) in which copy number variations were recently described to be associated with ID and ASD [20, 68, 69].

Transcription regulation

Several ID disorders result from mutations in genes encoding regulators of signal transduction pathways, transcription factors and cofactors involved in chromatin remodeling, gene expression and protein maturation [20]. The synthesis of functional proteins is a multistage process also determined by the balance between transcriptional activators/repressors and the regulation of RNA splicing, export and degradation [11, 48]. Altered regulation of chromatin structure and transcription mechanisms may result in deregulation of gene expression and abnormal expression of genes in specific cell-types and developmental stages, resulting in defective brain development and/or functioning [20].

Several genes known to be involved in transcription regulation were already described as causative of ID or ID related syndromes. Examples of those genes are *NF1* (neurofibromatosis type 1) gene, responsible for the ID phenotype in neurofibromatosis type 1 patients, the *RSK2* (ribosomal S6 kinase 2) gene, responsible for Coffin-Lowry syndrome and *CBP* (CREB-binding protein), involved in Rubinstein-Taybi syndrome. These three genes MAPK/ERK cascade displaying functions in the chromatin remodeling and transcription processes [20]. Genes encoding for proteins that are located in the cell nucleus and participate in large chromatin remodeling complexes that control gene transcription are also associated with ID, including *MECP2* (methyl-CpG-binding protein 2) gene,

causative of about 80% of the Rett Syndrome cases, *ZNF41* (zinc finger protein 41), *ATRX* (helicase 2), *CDKL5* (cycling dependent kinase 5), *ARX* (Aristaless-related homebox), *PQBP1* (polyglutamine-binding protein 1) and *FMR2* (fragile X mental retardation 2) are examples of genes that, once disrupted, may account of ID phenotype [20]. In table 1.4 there is a summary of some transcription regulator genes known to be involved ID or ID associated phenotypes.

Table 1.4- Transcription regulators genes known to be associated with ID (adapted from [71]).

	Gene	Locus	Protein	Function	Associated phenotype
Chromatin remodelling	*ATRX*	Xq13	ATRX	ATPase/helicase	Alpha-thalassemia/mental retardation syndrome
	CHD7	8q12.1	CHD7	ATP-dependent chromatin remodeller	CHARGE syndrome
	ERCC6	10q11	Rad26	ATPase/helicase	Cerebrooculofacioskeletal (COFS) syndrome
DNA methylation	*CDKL5*	Xp22	CDKL5	Dnmt1 phosphorylase, MeCP2-binding protein	West syndrome; Rett syndrome variant
	DNMT3B	20q11.2	DNMT3B	DNA methyltransferase	Immunodeficiency, centromeric instability and facial dysmorphisms (ICF) syndrome
	MECP2	Xq28	MeCP2	Methyl DNA-binding protein	Rett syndrome
Histone modification	*BCOR*	Xp11.4	BCOR	Histone modifier binding protein	Oculofaciocardiodental (OFCD) syndrome
	CREBBP	16p13.3	CREBBP	Histone acethyltransferase	Rubinstein-Taybi syndrome
	EHMT1	9q34.3	EHMT1	Histone methyltransferase	Kleefstra syndrome
	EP300	22q13	P300	Histone acethyltransferase	Rubinstein-Taybi syndrome
	HUWE1	Xp11.2	HUWE1	Histone ubiquitin ligase	Non-syndromic X-linked ID
	MED12	Xq13	TRAP230	Histone methyltransferase binding protein, histone phosphorylation activator	Optiz-Kaveggia syndrome; Lujan-Fryns syndrome
	NSD1	5q35	NSD1	Histone methyltransferase	Sotos syndrome; weaver syndrome; Beckwith-Wiedemann syndrome
	PHF8	Xp11.2	PHF8	Histone demetyhylase	Siderius X-linked ID syndrome
	RPS6KA3	Xp22.2-p22.1	RPS6KA3	Kinase-histone phosphorylation	Coffin-Lowry syndrome; X-linked ID
	JARID1C	Xp11.22-p11.21	JARID1C	Histone demetyhylase	X-linked ID
DNA /Chromatin binding	*BRWD3*	Xq13	BRWD3	Chromatine binding protein (putative)	Non-syndromic X-linked ID with macrocephaly
	PHF6	Xq26.3	PHF6	Chromatine binding protein (putative)	Borjeson-Forssman-Lehmann syndrome
	ZNF41	Xp22.1	ZNF41	DNA-binding protein	Non-syndromic X-linked ID
	ZNF81	Xp221-p11	ZNF81	DNA-binding protein	Non-syndromic X-linked ID
	ZNF674	Xp11	ZNF674	DNA-binding protein	Non-syndromic X-linked ID
	ZNF711	Xq21.1-q21.1	ZNF711	DNA-binding protein	Non-syndromic X-linked ID

mRNA regulation

Mutations in some RNA binding proteins are also known to account for some forms of ID. The most widely known case is FMRP (produced by *FMR1* gene) and that is causative of fragile X syndrome. FMRP is an RNA binding protein that associates with many mRNAs, some of which encode proteins important for neuronal development and plasticity. FMRP controls activity-dependent dendritic mRNA localization and translational efficiency of dendritic mRNAs in response to stimulation of mGluRs [49,48]. It is currently known that neurons are highly specialized cells where some specific aspects of RNA metabolism play critical roles for their function specially during development, when trafficking of mRNAs to growth cones (axonal and dendritic) regulates neuronal growth. After synapse formation, mRNAs continue to be transported to dendrites and axons where they are locally translated at the synapse. This process has extreme importance for synaptic plasticity (the biological correlate of memory/learning) and its deregulation can have an important impact on cognition [70].

miRNAs

miRNAs are small non-coding RNA molecules (~21 bp) encoded in the genome that regulate gene expression by binding to the 3'UTR of specific target mRNAs. This regulatory process may occur via translational repression and/or mRNA destabilization/degradation [73, 74].

miRNAs are known to be involved in several developmental processes, such as development timing, apoptosis, myogenesis and glucose homeostasis and currently there are 1424 miRNA sequences listed for the human genome in the online database miRBase (http://www.mirbase.org/cgi-bin/mirna_summary.pl?org=hsa) [75]. Many of these miRNAs are expressed in the brain and seem to be essential for neuronal cell development and to interfere in processes such as neurite outgrowth, synaptic development and neuronal plasticity [74, 75].

Although a number of miRNAs have been shown to be important for neuronal development very little is known about the mechanisms by which these molecules regulate synapse formation and maturation. One of the most well studied cases is the miR132 that has been shown to influence the Rac1-PAK-mediated spinogenesis and influence dendritic growth in hippocampal neurons [73]. Additionally, miR132 was also shown to regulate expression of *MECP2*, the gene responsible for Rett syndrome[73, 76]. Indeed, miR132 transgenic mice displaied decreased Mecp2 protein expression and showed significant deficits in novel object recognition [73]. The authors propose that miR132 is a regulatory miRNA in neurons and raise the hypothesis that miR132 may contribute not only for the cognitive defect in Rett syndrome, but also for a larger set of ID-related disorders.

In order to further explore the actual function of these small molecules in brain development it would be interesting to characterize the function of the miRNAs known to be specifically expressed in central nervous system.

Ubiquitin-proteasome system (UPS)

The programmed, fast and specific degradation of proteins in the cell is essential for survival and plays an important role in virtually all the cell events, and thus also in the function of neurons and synapses. The process of targeting proteins for degradation is accomplished by covalent attachment of ubiquitin to the protein substract. These ubiquin tags act as a label for proteins that must be degraded because they are no longer useful for the cell [77]. The specificity and timing of ubiquitin attachment to particular proteins is determined by ubiquitin ligase proteins (E3 ubiquitin ligases, for instance) is extremely important for the fine tuning of protein levels in the cell. Cullins are highly conserved among species and serve as catalytic core components of E3 enzymes. Recently, mutations in cullin proteins have been described in patients with ID [22, 46, 47]. Examples of genes involved in this regulatory pathway that are associated with ID are *CUL4B* (cullin 4B), *CUL7* (cullin 7) and *HUWE1* (HECT, UBA, and WWE domains-containing protein 1) [79, 80].

Additionally to his role in substrate tagging for degradation, the UPS is also associated with neurotransmitter receptors, protein kinases, synaptic proteins and transcription factors regulation, all considered important molecules for synaptic plasticity [81].

Synaptic and dendritic spine malformation in intellectual disability: two examples

Abnormalities in dendritic spines in ID patients were first described in 1970 by Huttenlocher [82]. That report showed that dendrites of cortical neuron of ID children had abnormal long, thin spines instead of short and thick spines. Nowadays it is generally accepted that abnormalities in dendritic structure and/or dendritic spine morphology give rise to altered neuronal activity and contribute to the cognitive disabilities of ID.

Changes in the size of the spines as well as the number of those have been associated with learning. For example, it has been demonstrated that spatial learning is associated with an increase in the number of spine densities in hippocampus while the reduction of those same spines reflect deficiencies in spatial memory in aged mice [56]. ID is often accompanied by an increase in spine length, indicating a relationship between the length of the dendritic spines and the efficacy of transmission [56].

Fragile X syndrome (FXS)

FXS is the most common form of inherited ID affecting approximately 1:4000 males and 1:8000 females. This syndrome is characterized by intellectual disability, development delay, hyperactivity, hypersensitive to stimuli, mood instability and autism. The FXS is caused by the expansion of the trinucleotide repeat CGC at *FMR1* gene in the X chromosome, leading to the absence of FMRP protein. Although its function is still not completely characterized, FMRP seems to act as an RNA

binding protein and some mRNAs responsible for the synthesis of several proteins with variable functions were found to bind to FMRP [32, 33, 34] .

It is thought that FMRP plays an important role in synaptic plasticity through regulation of mRNA transport and translational inhibition of local protein synthesis ate synapses [85]. In neurons, FMRP is localized at the base of dendritic spines in association with the polyribosome. It is currently thought that FMRP is a translation repressor of microtubule associated protein 1B (MAP1B). MAP1B is essential for stabilization of microtubules during elongation of dendrites and neuritis and it is involved in the dynamics of morphological structures. It is hypothesized that the lack of FMRP results in altered microtubule dynamics that may contribute to the altered spine morphology seen in FXS patients. Studies in patients and animal models of FXS have identified increased spine density and an excess of abnormal long, thin and immature spines which are indicative of alterations in synapse development and/or function in this patients [86].

Rett syndrome (RTT)

RTT is a neurodevelopmental disease also associated with abnormalities in brain size, branching and synaptic morphology, neurotransmitter receptors and genes expression. Taken together, these alterations strongly suggest a disorder of brain growth, especially the growth and development of gray matter and axonodendritic connections [87]. Post-mortem studies revealed that the neuronal cell packing and synaptic density lower in RTT cerebral cortex when compared to controls [88]. At the mRNA level, the same type of studies revealed that synaptic markers such as *MAP-2*, synapsin II and synaptotagmins 1 and 5 presented decreased levels of expression. Mutations in the *MECP2* gene can account for the disruption of synaptic formation, maturation and refinement [44, 45].

The increasing number of genes identified over the last years as associated with ID suggest that this phenotype can emerge as the final common pathway of many different types of abnormal cellular processes, and that there isn´t a single main general mechanism responsible for the disease. These findings went against the first hopes of ID researchers who thought that the identification of the genes responsible for the disease would lead them into the mechanisms of pathology in a very straightforward way. On the other hand, it is consistent with the fact that intellectual processing in humans is both sophisticated and complex. The purpose of understanding the patophysiology of this disease is to provide a better clinical service to the patients and their families [7].

The emerging theory of the tetrapartite synapse: non-neuronal brain plasticity

It has been said in this work that chemical synapse are units of signal transfer the brain that are composed by a bipartite mechanism: presynaptic cells and postsynaptic cells. However, over the

last years new findings have proposed glia and the extracellular matrix (ECM) as important participants in the synaptic machinery, giving rise to the hypothesis of a tetrapartite synapse [90, 91, 92]. For long it has been thought that glia must be doing something more than just filling the spaces between neurons. In spite of not being able to fire action potentials, glial cells are highly excitable cells that communicate with neurons and other glia through chemical signal [93]. The major types of glia that in the central nervous system communicates with synapses are the astrocytes, oligodendrocites (myelination glia) and microglia. Besides the glia, there is also the ECM that accounts for about 20% of the total volume of the mature brain [91]. ECM molecules are synthesized by neurons, glial and non-neuronal cells and secreted into the extracellular space and is manly composed by collagen, proteoglycans, glycoproteins and integrins [90].

Astrocytes are the most abundant cell type in the brain and are able to “listen” and respond to synaptic activity. During development, the formation of mature neuronal circuits requires the selective elimination of inappropriate synaptic connections as well as the stabilization of other ones. It is currently believed that glial cells may have a role in these processes as well [93].

In the last decade, closer attention has been given to the many forms of structural and synaptic changes that rely on interactions between the ECM and astroglia, and it is becoming increasingly clear that glia plays a crucial and dynamic role in the synapse development and influence functions that for years have been strictly associated with neuronal control [90, 93]. Consequently, further attention must be paid to the function of these two factors (ECM and glia) in the search for the genes associated with ID.

Types of mutations underlying ID

Chromosomal rearrangements and structural variations in the genome

Chromosomal aberrations can be numeric or structural. Numeric aberration can be due to the loss or gain of an entire chromosome (leading to the monosomy or trisomy) or of the whole chromosome complement (given rise to triploidy and tetraploidy). In a general perspective, the absence of a chromosome is far more drastic than its excess. These consequences are not the same for autosomal and sexual chromosomes since the absence of an entire autosomal chromosome is not compatible with life while the same alteration in a sexual chromosome may result in a liveborn female (45, X – Turner syndrome) [94]. This type of alteration is usually found in cancer cells but can also lead to a variety of human genetic disorders, such as Down syndrome, Cri du Chat syndrome, Klinefelter's syndrome, etc. [95]. Chromosomal aberrations that large can be detected by low-resolution cytogenetic methods.

Structural variation is a term used to describe all types of genomic rearrangements, including deletions, duplications, insertions, inversions, translocations, loss of heterozygosity (LOH) and more complex alterations. The most common form of structural variation in the genome are copy number variations [96]. Structural abnormalities are a consequence of double-strand breaks and inappropriate DNA repair. These types of alterations can be balanced or unbalanced, but both require a break in the DNA phosphodiester backbone. In figure 1.5 is represented a scheme of the most common structural variations.

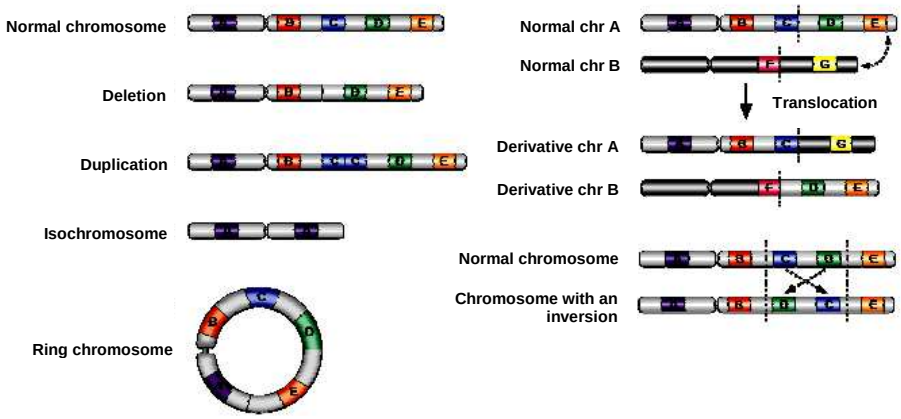


Figure 1.5 – Schematic representation of the most common structural variations occurring in the genome (adapted from [94]).

During many years genomic DNA copy-number gains and losses were assumed to be few and to have limited impact on the total content of human genetic variations [97]. With the development and improvement of genome-wide analysis tools to study the genome, it was shown that CNVs are relatively frequent, spread throughout the genome and represent a very significant source of

genetic variation in human populations. Until now, thousands of heritable copy number variations have been identifying regions potential variability corresponding to approximadetly 13% of the human genome [98].

Using aCGH methodologies it is possible to identify unbalanced alterations within the genome. However, alterations where there isn´t an alteration in the amount of DNA, such as balanced inversions that only alter the orientation of a portion of a genomic region, are not detected by aCGH. As a consequence, and despite the great improvements in mapping deletions and duplications, the mapping of inversions has lagged behind [96]. Thus the real number of common inversion and regions prone to develop this kind of alteration is currently not clear. With the availability of themost recent high-throughput sequencing techniques, this type of variation is possible to detect and, consequently, our knowledge about them has increased greatly [1].

Copy number variations

The definition of a CNV says that it is a DNA segment larger than 1Kb that present variable copy number when compared to another reference genome. If these alterations are so common, can happen along the entire genome and reach a considerable size, it is likely that CNVs can influence the transcriptional and/or translational levels of overlapping and nearby genes [56, 57]. In fact, whenever the deletion or duplication harbors one or more dosage-sensitive genes or occurs simultaneously with point mutations in the other allele, an abnormal phenotype can result from the alteration [1].

Copy number variations are often sporadic and caused by *de novo* rearrangements. It is accepted that these types of variations can occur with/at a 1000 to 10.000 higher frequency than point mutations [1,100]. Although CNVs occur at similar frequencies between populations, several significant differences in the frequency of some CNVs were found between different populations in the world, revealing that variations in the genomic architecture can not only account for disease, but also for ethnic differences and selective evolution [101].

The distribution of CNVs in the genome is thought to be non-random and several “hot spots” in the genome have been identified, such as in centromere and telomere regions). The higher frequency of CNVs in these regions is perhaps due to their repetitive nature. Also, the CNV occurrence rate does not correlate with the chromosome size. Some chromosomes have relatively high proportion of CNVs, while others have lower proportions. For instance, about 20% of the chromosome 18 sequence is covered by CNVs, while for chromosomes 16, 17, 19 and 22 more than 40% of their sequence is covered by CNVs [102]. There are two main theories to explain this non-random distribution of CNVs throughout the genome: (I) the mutational hypothesis that defends that most CNVs are phenotypically neutral but, because they are flanked by repetitive sequences that predispose to its occurrence, are more frequent in that regions and (II) the selection hypothesis that

states that negative and positive selection determine the bias occurrence of CNVs in certain regions [101].

Despite the extensive studies performed in the last decade, the real amount, position, size, gene content and distribution in population of common CNVs remain unclear. The wider application of high-resolution aCGH technologies and next generation sequencing will reveal will contribute to the description of many more different CNVs between individuals and populations, as the majority of CNVs will appear at smaller sizes [1].

CNVs described in presumelly healthy individuals have been cataloged at DGV (Toronto Database of Genomic Variants, site), while clinically relevant CNVs can be found in DECIPHER (DatabasE of Chromosomal Imbalances and Phenotype in Humans using Ensembl Resources, site). However, Decipher database is not always clear since it describes all the CNVs found for the patient, not always has the phenotypic description of the case and parental analysis for the alterations.

CNVs occur in about 13% of the human genome and can be categorized into two major groups based on the breakpoint pattern analysis: (I) recurrent rearrangements, occurring in multiple unrelated individuals with clustering of breakpoints and sharing a rearrangements with a minimal overlapping region and (II) non-recurrent rearrangements, with variable breakpoints [1, 100, 101].

It is estimated that about 45% of the human genome is composed by repetitive sequences including transposon-derived repeats, processed pseudogenes, simple sequence repeats and blocks of tandem repetitive sequences. These types of variants are many times called common repeats [103]. In addition to these elements, low copy repeats (LCRs), also called segmental duplications, constitute another kind of repeated sequences that account for about 5% of the genome. LCRs are DNA fragments that range in size from 1Kb to 400 Kb, and share a high level of sequence identity (higher than 90%). These alterations are present in at least two copies of the genome and many times have complex structures.

Both LCRs and common repeats cause genomic instability and have been implicated in the generation of constitutional genomic rearrangements [100, 103].

Mechanisms underlying copy number variation origin

So far, four mechanisms of origin have been proposed to underlie these alterations: nonalelic-homologous recombination (NAHR), nonhomologous end-joining (NHEJ), L1 retrotransposition and fork stalling and template switching (FoSTeS). Nonalelic-homologous recombination and nonhomologous end-joining account for the majority of CNVs [3, 6, 58, 59, 60, 61, 62]. It is important to say that many structural variations display features of origin that can be interpreted as NAHR, NHEJ, FoSTeS and retrotransposition, and many times, there isn´t a single possible mechanism associated with the alteration [100].

Nonalelic-homologous recombination (NAHR)

NAHR is caused by the alignment and crossover of two nonallelic DNA sequence repeats sharing high similarity to each other [100]. When two LCRs have sequence identity higher than 97% and are located at a distance less than 10 Mb from each other, they can lead to misalignment of chromatids and mediate an alteration in the genome. The formation of a stable reciprocal translocation mediated by interchromosomal NAHR between LCRs is dependent upon the orientation of the LCRs and the chromosome arms involved. When LCRs are directly oriented they give raise to deletions or duplications of the region between them. If the LCRs are in different directions they will result in an inversion of the region [1, 108].

Recurrent interstitial rearrangements are often flanked by LCRs, indicating that NAHR is a major causative mechanism for these structural variations.

In addition to LCRs, smaller but equally highly similar DNA sequences can lead to NAHR. These sequences can be retrotransposable L1 elements, LINE elements (long interspersed nuclear elements), *Alu* elements (a class of SINEs, short interspersed nuclear elements), LTRs (long terminal repeats) or matching pseudogenes [102, 94, 100]. Such shorter sequence subtracts mediate NAHR of smaller DNA regions and that many times are out of reach of aCGH platforms of lower resolutions. With the advent of massive paralel sequencing techniques, the amount of NAHR-related alterations mediated by these types of sequences that are detected will certainly increase [108].

Nonhomologous end-joining (NHEJ)

While in recurrent rearrangements the most common originating method proposed is NAHR, regarding nonrecurrent CNVs the most common method causative method is thought to be nonhomologous end joining (NHEJ). NHEJ is a natural DNA repair mechanism responsible for correction of DNA double strands breaks caused by ionizing radiation or reactive oxygen species [1]. Unlike NAHR, NHEJ does not require LCRs or any other type of highly similar sequences to occur and often leaves a “DNA scar” resulting from the addition of extra nucleotides at the DNA junction. Breakpoints of NHEJ-mediated rearrangements are often located within repetitive elements such as *Alu*, LINE or LTR sequences. Additionally, sequence motifs like TTTAAA, known to be able to cause double strand DNA breaking and curving of the DNA, are present in the proximity of many of these breakpoints. These data suggest that, even without any necessity of LCRs for occurring, NHEJ may still be stimulated by them [1, 100, 109].

Fork stalling and template switching (FoSTeS)

FoSTeS is a replication error based mechanism that has been recently implicated in the origin of genomic-disorder-associated nonrecurrent rearrangements of complex structure (for instance, deletions or duplications interrupted by normal copy number regions and also triplicated segments) [1]. According to this mechanism, during DNA replication the replication fork can be stalled and the polymerase can switch to other replication fork and restart DNA replication in the new fork. This exchange occurs via the microhomology between the switched template site and the original fork [100], [106]. The new template sequence doesn´t have to be necessarily adjacent to the original replication fork, but needs to at least be located in three-dimensional physical distance [100].

Depending on the direction of the fork progression and whether the lagging or leading strand of the new fork is used as template, the wrongly incorporated fragment can be in the same direction or inverted when related with its original position. Moreover, if the new fork is located downstream of the original fork, the alteration will give rise to a deletion, while if the new fork is located upstream of the original one, the new fragment will be duplicated [1, 100, 106].

This mechanism can generate not only complex and relatively small rearrangements (sometimes even single exons), but also large genomic alterations. Sometimes, analysis of the breakpoints reveals not only one but more FoTSeS events as originators of the rearrangement [106]. One example is the genomic rearrangements find in *MECP2* locus, that for many years were not explained by NHAR or NHEJ, is now though to be caused by FoTSeS [1].

L1 Retrotransposition

Transposons are mobile genetic elements that can affect gene expression in virtually all higher eukaryotes. Transposons are elements that can be transcribed into RNA, reverse transcribed into cDNA, and then reintegrated as cDNAs into the genome at a new location. In the human genome, long interspersed element-1 (L1) elements cover around 17% of the genome [100]. L1s are also capable of mobilizing Alu and SINE elements as well as retrogenes within the genome. A high amount of inversions and deletions in the genome are attributed to retrotransposition [100].

A schematic representation of these 4 mechanisms of origin is described in figure 1.6.

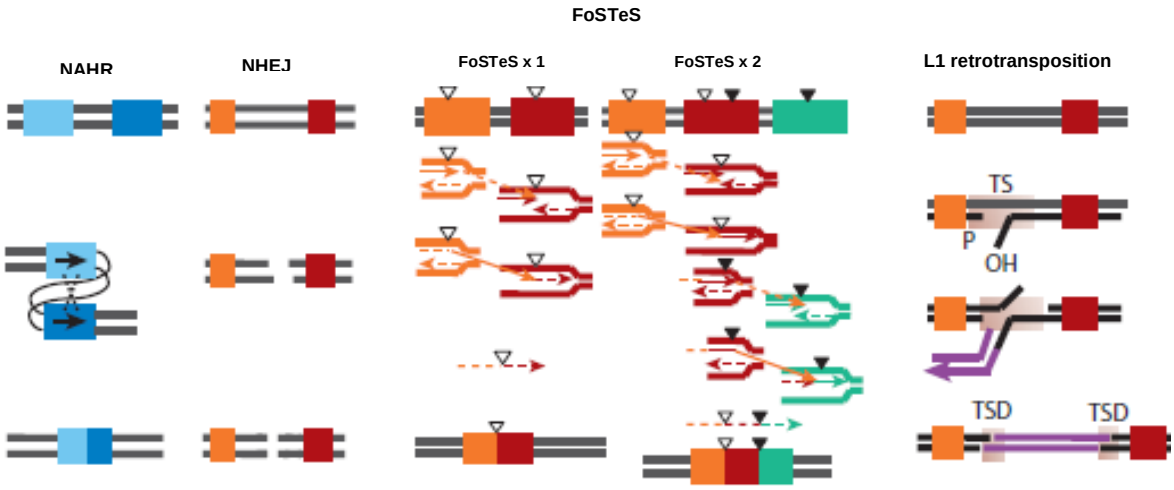


Figure1.6 – Schematic representation of the four most common molecular mechanisms of origin by which CNVs can occur (adapted from [100])

Microhomology-mediated break-induced replication (MMBIR)

Microhomology-mediated break-induced replication is a general model, based on observations not only in humans but also in other animal models, proposed to explain nonrecurrent CNVs and complex genomic rearrangements in all life forms [1]. The MMBIR model follows the same basic principles of the FoSTeS model: the rearrangement is initiated by a single-end double-strand DNA break resulting from a collapsed replication fork [94]. The single-strand 3' tails from the broken replication fork will anneal with a microhomologous single-stranded DNA nearby, where it forms a new replication fork. Similar to what happens in the FoSTeS mechanism, the replication in the new fork is of low processivity and the extended end will dissociate and invade different templates [94, 106]. The schematic representation of the MMBRI mechanism is briefly summarized in figure 1.7.

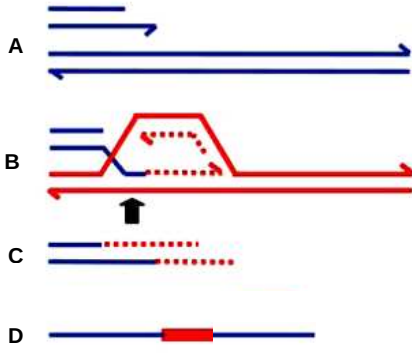


Figure 1.7 – Schematic representation of MMBIR mechanism. (A) Shows the broken arm of a collapsed replication fork that forms a microhomology junction (arrow) leading to the formation of a new low-processivity fork **(B).** The extended end dissociates **(C)** originating a duplicated region in the genome **(D).** Adapted from [94]

It has been proposed that both FoSTeS and MMBIR mechanisms may be a major cause of nonrecurrent complex structural variations in the human genome, as well as in the formation of LCRs that, consequently, will predispose to NAHR [1, 100].

Copy number variations in health and disease

Before the last decade, it was thought that healthy individuals were diploid for each chromosome inheriting one copy from the mother and other from the father and that any alteration of this pattern could result in an altered phenotype. Also, before the beginning of the sequencing of the human genome, it was thought that two healthy individuals should share about 99,9% of the genome, being the 0,1% of due to single nucleotide polymorphisms [102]. This perception of the human genome has completely changed in the last decade, mainly because of the technical improvements for the detection of CNVs and other structural variants.

Copy number variations are frequent and widespread through all the genome. Genomic rearrangements may cause Mendelian diseases, complex traits with complex inheritance patterns (such as behavioral disorders) or just represent benign polymorphisms with relatively high frequency in general population [1]. During the years many diseases have been associated with DNA rearrangements that result in gain, loss or disruption of genes for which dosage is critical. The impairment in gene expression may not only be due to deletions or duplications covering the gene (and consequently less or more quantity of the gene, respectively), but also alterations in gene regulatory regions. Whatever the mechanism, these disorders are often characterized by overlapping genomic features and clinical symptoms. Genomic imbalance disorders are often sporadic and caused by de novo alterations [110].

It is known that CNVs can affect gene expression and play an important role in the etiology of common diseases. About 15% of the genes present in OMIM database are within CNV regions and, many times, these dosage alterations turn to be relevant for molecular-environmental interactions [101, 111].

There are several molecular mechanisms that can lead to clinically significant alterations. The most obvious and widely described mechanism is the coverage of an entire gene that is sensitive to dosage, such as duplications or deletions. Another important aspect is when the alteration doesn´t cover the entire region and there is a breakpoint occurring inside the gene. These breakpoints can be due to deletions, duplication, translocations or inversions (figure 8). This type of alteration may lead to loss-of-function of the gene or gain-of-function (through the creation of new transcripts by gene fusion or exon shuffling). Structural alterations can also lead to deregulation of a gene (either by down regulation or over expression) by affecting regions outside the gene (such as the promoter region or other important regulatory elements). Many times alteration in regions that don´t cover any gene are found and present themselves as almost impossible to interpret because of the lack of information regarding the sequence [94].

Not only CNVs in dosage-sensitive genes can account for the phenotype. A deletion in one allele can unmask a recessive mutation in the other allele or even a heterozygous polymorphism with functional significance while in homozygosity. Finally, it is also hypothesized that a CNV in one allele can interfere with the interaction mechanism between them (transvection) [94]. In figure 1.8 there is a schematic representation of the mechanisms described above.

.

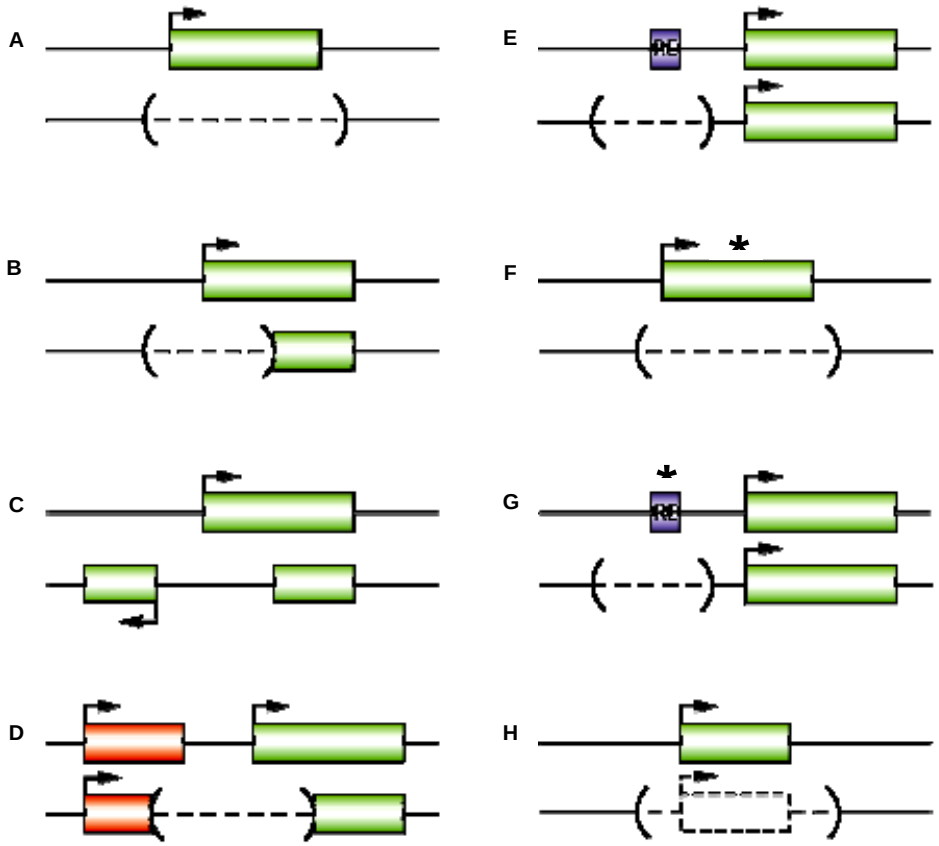


Figure 1.8 - Molecular mechanisms by which chromosomal rearrangements can influence phenotypes. The rearrangement can encompass a dosage sensitive gene that causes disease (**A**); disrupt a dosage sensitive gene through deletion (**B**), duplication, translocation or inversion (**C**); create a fusion gene (**D**); exert a position effect by affecting a regulatory element (**E**); unmask a recessive allele (**F**) or functional polymorphism (**G**) on the homologous chromosome; interrupt effects of transvection (**H**) where the deletion of a gene affects communication between alleles. Genes are depicted as green or red rectangles, regulatory elements (purple boxes) as purple rectangles and an asterisk (*) indicates a recessive point nucleotide variation. Adapted from [94].

Genetic testing of children with ID

A child with ID is deserving of the best diagnostic evaluation available in order to improve the health and well being of that child and his/her family. Clinicians report that the family of a child with global DD and/or ID often experiences the feeling of loss of control and guilt. A specific diagnosis is one that can be used by the family in obtaining information about prognosis, recurrence risks and available therapy [6].

The preliminary clarification of the type of DD presented by the child is an important step because it is going to lead the path of the genetic investigation followed. The diagnostic approach that should be undertaken for a child with ID includes the following steps: collection of the clinical history (including prenatal and birth histories), family history and physical and neurologic examination. These two last parameters are of extreme importance since the extended examination for minor anomalies, neurologic and behavioral signs may suggest a specific recognizable syndrome. Only after the clinical genetic evaluation and determination of the diagnosis path to be followed, laboratory tests, imaging and other consultants should be required [4]. During the years, several studies have estimated 4 to 40% of all cases of ID to have a genetic etiology, with variable rates of success in this estimate likely to be due to patient selection criteria and available laboratory technologies [112].

The reason most commonly cited by parents for pursuing diagnostic testing is “to know”. The establishment of a medical cause for their child´s condition and the end of the uncertain of “not knowing” is one of the most important aspects for families affected by the disease. Several additional reasons and expected outcomes for establishing a diagnosis were: to understand prognosis and future needs, to improve response by medical and educational services, to obtain early access to educational and social services, to test other offspring with questionable symptoms and make future reproductive decisions based on appropriately estimated recurrence rate [21].

The final clinical benefits of achieving a specific genetic diagnosis from the clinical point of view are, among others, the following: end the diagnostic odyssey, alleviate parents´ fears about other causes, pass on patients to an appropriated specialist, determine possible prognosis, guide optimal management and surveillance of associated co-morbidities, advise of the risk of recurrence in future offspring, increase knowledge about precise genotype-phenotype correlations; provide potential insight into disease mechanisms and eventual development of therapeutic interventions [56].

Standard cytogenetic analysis

The purpose of cytogenetic testing is to look for chromosomal rearrangements that could explain the disease in the child. It is now known that the most commonly found chromosomal rearrangements are copy number variations. These types of alterations consist in deletions or duplications of large segments of DNA (above 1Kb) and have been described the in context of

diseases and also in healthy controls. Less common, but likely to account for some causes of DD in children are apparently balanced translocations (exchanges of DNA segments of the same size between chromosomes) and inversions (inversion of a DNA segment within a chromosome) [37, 38].

Cytogenetic studies are recommended in the evaluation of all children with ID whose etiology is not clinically evident [114]. The currently guidelines adopted in Portugal for genetic testing of these patients suggest the search of genomic imbalances or rearrangements by regular G-banded karyotyping in first place, followed, whenever the result is normal, by analysis of subtelomeric rearrangements, use of specific FISH probes according to clinical presentation and by the testing of the most common single gene disorders (for instance, Fragile X syndrome) [115].

Conventional cytogenetic techniques were developed in the early 1970's and have allowed the identification of each of the human chromosomes. This kind of analysis uses light microscopy to examine stained metaphase or prometaphase chromosomes to produce a distinct banding pattern for each chromosome [116]. This approach requires dividing cells (usually peripheral blood leucocytes, bone marrow cells or fibroblasts) at a stage when chromosomes are highly condensed allowing a resolution (detectable size of chromosomal abnormalities) of a series of 400-550 light/dark bands in each genome [116, 117, 118, 119, 120]. In spite of its important role as a medical diagnostic tool, subtle deletions and duplications can´t be detected by regular chromosome banding [121]. To overcome this obstacle, high-resolution chromosome analysis, developed in 1976 by J.J. Yunis, allows a resolution up to 2000 bands per haploid genome [122]. This approach is only used when a suspicion is raised about a specific chromosomal region, because it requires extended prometaphase chromosomes (that are often overlapped and entangled after the cell preparation), which represent increased difficulty in the analysis [121].

Other stains and techniques can be used to improve information and resolution. The most commonly used technique to investigate target regions is fluorescent in situ hybridization (FISH). In the patients with ID/DD, when karyotype analysis is normal, FISH can be used to detect submicroscopic alterations. This type of technique requires that the differential diagnosis must be first decided so that the specific chromosomal location for analysis could be targeted by the FISH probe [57, 58]. FISH uses fluorescence-labeled DNA probes to detect the presence, number and location of genomic regions that are too small to be visualized by more conventional techniques [121]. This technique can confirm or exclude a clinical suspicion of know and clinically established microdeletions and microduplications syndromes [124]. When a differential diagnosis and/or target analysis are not possible, a screening of the subtelomeric regions is recommended [125]. The use of subtelomere FISH panels during the years has revealed that, in the absence of specific clinical features suggestive of a syndrome, patients with ID should also be screened for this group of genomic imbalances. The developments in FISH techniques generated the second revolution in cytogenetic analysis [126]. First, the chromosome painting allowed the observation of complex rearrangements and second, the Human Genome Project allowed the generation of single locus probes that facilitated the delineation of critical deleted/duplicated regions associated with specific disease phenotypes [61, 62]. This development was the first step for the shift from the traditional

"phenotype-first" approach, to a "genotype-first" approach. This last approach allows the characterization of a common genetic alteration in first place, followed by a more detailed clinical characterization as more patients with the same alteration are discovered [62, 63].

During the years, the development of labeling and detection methods that allowed the visualization of the all chromosomes in different colors provided the first prospect of a "molecular" karyotype. These techniques, named multiplex-FISH (M-FISH) and spectral karyotyping (SKY-FISH), combined the global screening power with the accuracy of the molecular analysis [126]. Besides its incredible ability to discriminate the components of highly complex rearrangements, obtaining metaphase cells is not always easy and increases the cost of the analysis. The development of FISH techniques that don´t require dividing cells (interphase FISH) enabled the analysis of samples which otherwise would be impossible to analyze [61, 64].

Rearrangements of the subtelomeric regions of chromosomes are a significant cause of both idiopathic and familial ID [42, 43, 44]. The telomere protects the chromosome from end-to-end fusion and degradation. However, due to the sequence similarity that exists between non-homologous chromosomes in these areas, rearrangements in these regions are relatively frequent. Because such rearrangements are too small to be detected by conventional karyotyping, a series of different protocols have been described to be useful for the detection of subtelomeric alterations (multiprobe FISH, microsatellite marker analysis, multiplex ligation-dependent probe amplification, quantitative PCR, multicolor FISH, among others), MLPA being the most commonly used technique for this kind of analysis [43, 45, 46] .

Although it has been broadly used for more than 30 years, normal karyotyping has a low resolution power (it can miss alterations within the 3Mb-10Mb size range) and it is also subject to inter-observer variations, leading to variable detection rates between operators and laboratories. Importantly, it leaves the majority of cases without identification of the genetic abnormality associated with the child´s condition [5, 133].

aCGH based cytogenetic analysis

Microarray technology was developed more than a decade ago and has now become a routine tool in genetic research. This methodology was first used in the clinical genetics field for the analysis of CNVs in all human telomeres in patients with ID (48,20). Since then, aCGH has been adopted by hundreds of genetic laboratories although the vast majority of the platforms now target the entire genome and not only the telomeres [135].

aCGH formats and technical principles

In an aCGH experiment, DNA from a test sample (e.g. patient with ID) and a reference DNA are differentially labeled with a fluorescent dye and hybridized in a glass slide containing the

representation of the genome (probes) (figure 1.9). These arrays are generally produced on glass slides, silicon or plastic substrates and can have from a hundred to many thousands probes. The construction of the array slide involves the physical immobilization of the previously selected and synthetically generated probes onto the solid support in specific sites [71, 137].

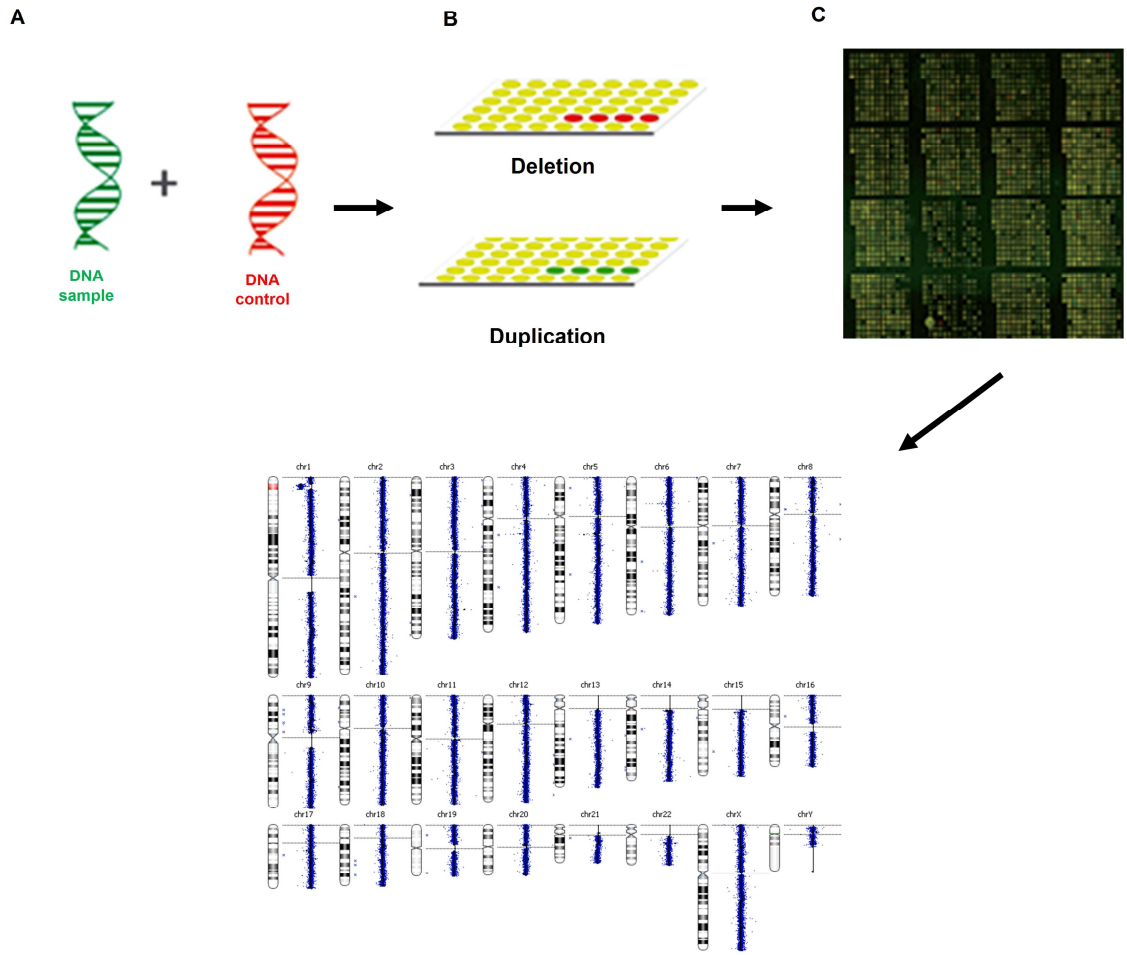


Figure 1.9 - aCGH methodology schematic representation. Schematic representation of the DNA staining with two different fluorescent dyes (Cy3 and Cy5) (**A**), hybridization representation (**B**), scan image of the slide after hybridization (**C**) and Nexus software (BioDiscovery, Inc) virtual schematic karyotype representation of all human chromosomes (**D**). Notice the deletion in the short arm of chromosome 1. Adapted from [307, 136].

In this methodology the genomic DNA from a test and reference samples are labeled with different fluorophores and hybridized with genomic regions (probes) located on the slide. The intensity of the colors from both test and reference DNAs is compared and measured into an intensity value (figure 1.9) [136].

The vast majority of aCGH data available today in public databases and published articles originates from studies using Bacterial Artificial Clone-based CGH (BAC arrays). BAC probes vary from 150 to 200 Kb and require high amounts of DNA for hybridization, leading to the need of PCR

based amplifications before this step. Over the years, oligonucleotide-based arrays (oligo arrays) have been considered to have large advantages over the BAC arrays. These platforms are characterized by single-stranded oligonucleotides (25 to 85 bp in length, depending on the platform) attached to the array slide, allowing CNV detection and analysis with much higher resolution [3].

The development of the aCGH platforms has lead to an important increase in the resolution of CNV detection. Although BAC arrays present the advantage of clearly identifying copy number changes in a sensitive and very precise way, the spatial resolution of the array is determined by the sensitivity, the number, the chromosomal distribution and the length of the probes in the array. Facing this fact, platforms with shorter probes will allow more probes in the same space, leading to higher resolution. In the early years of this technology, the "proof of principle" required before aCGH could be adopted to diagnostic uses was still missing and there was a lack of international databases and population studies regarding pathogenic and non-pathogenic CNVs, which lead to use of BAC arrays as a more safe and robust strategy. As the years went by, the increase in knowledge concerning the significance of CNVs lead to the need of platforms with higher resolution power and, as a consequence, the oligo arrays started to be used in a widespread and routinely manner [50, 49, 51, 52].

The successful implementation and use of aCGH technologies requires the use of many statistical methods and techniques for synthesis and spotting of the probes in the right place, detection of hybridized samples and informatic analysis of the data. The informatic analysis usually requires the detection of the ratio of the signal generated at each probe location in the test and control sample [52, 53, 54]. The slide is scanned to obtain the complete hybridization pattern of the entire array (figure 9-C). During the years, a series of bioinformatic tools have been developed to transform the complex data and artifacts into useful information [55, 62].

CNVs and diagnosis of ID

The increased offer in commercially available oligo microarrays allowed the relatively easy implementation of the technique in clinical practice. The increasing genome coverage and resolution and the high levels of production according to standardized and controlled conditions seems to overcome all the major technical problems that were present years ago. However, the clinical relevance of many of the CNVs found is still presenting challenges and doubts. The problem is the existence of CNVs in about 12% of the human genome and the fact that often those are present both in healthy individuals and in subjects with the most varied pathological conditions [106].

The main strategy for distinguish pathogenic CNVs from benign variants is the study of unaffected relatives (mostly parents) of a patient, as well as large normal control cohorts. However, often both parents are not available for testing, are not completely asymptomatic (which difficult the CNV

interpretation and association with the disease) and many times the clinical information regarding the parents is also not available and/or assessed. Additionally, many times the currently used databases don´t provide enough information to clarify whether the CNV is a very rare alteration that, alone or together with other theoretically benign CNVs, can account for the phenotype [63, 37]. In spite of our current knowledge about the clinical implications of some of the alterations found, it is unlikely that the CNVs most frequently found in population are directly related with ID. However, we should be careful when comparing CNVs from ID patients with the ones described in public databases only by the presence-absence criteria. Instead, attention to the amount of overlap and number of individuals described should also be paid [134].

There are public databases created with the purpose of help clarifying these questions, such as the Database of Genomic Variants (http://projects.tcag.ca/cgi-bin/variation/gbrowse/hg18/). However, it is important to realize that these databases were created with research purposes and, as a consequence, most of the CNVs are identified by high-resolution microarrays or high throughput sequencing techniques. Also, only a small portion of the CNVs described have been validated leading to doubts in the real percentage of false positive CNVs. Also, when the CNVs reported were discovered using low resolution platforms (BAC arrays), it is difficult to determine the size of the genomic region affected and, therefore, the real size of the relevant variation may be much smaller than previously estimated [48, 56].

Another limitation of the aCGH technique is the incapacity to detect balanced translocations and inversions, that occur in 0,75% of all ID patients [145].

Thus, G-banded karyotyping is still important for detection of balanced rearrangements and low-level mosaicism. These two types of alterations are not yet uniformly detected by aCGH since it depends of the existence of pathogenic balanced translocations and percentage of mosaic material present in the patient. A study performed by Miller DT *et al* in 2010 concluded that (I) truly balanced rearrangements account only for a small proportion (<1%) of clinically significant genomic events in patients and (II) many apparently balanced rearrangements detected by regular karyotyping show submicroscopic imbalances at breakpoints when analyzed by aCGH [114].

Microdeletions/microduplications syndromes in ID

The development in aCGH technique allows genome-wide studies in patients with ID. As a consequence, the discovery of new microdeletion and microduplication syndromes has substantially increased. When searching for the genetic basis of clinical phenotypes different types of approaches can be used, such as: (I) collection of a series of patients with similar clinical phenotypes and identifications of the genomic imbalance that account for phenotype (phenotype first), (II) the identification of additional patients with previously described clinical features and the proposal of the genetic etiology of the disease, (III) delineation of a minimal critical region that can account for the phenotype as well as further identification of more patients (genotype first) and (IV)

identification of genomic imbalances by large-scale screening techniques common to several patients and highlight of the common phenotypical features among them (reverse phenotyping) [40, 65]. Any of these approaches have proven to be successful considering the amount of established phenotypes nowadays. Table 1.5 summarizes the microdeletion and microduplication syndromes associated with ID reported so far in the Decipher database (http://decipher.sanger.ac.uk/).

Table 1.5 - Recurrent interstitial microdeletion/duplications associated with ID. List of syndromes associated with ID first recognized by molecular karyotyping. Adapted from [146], [147].

Syndrome Name	Other Alias	Minimal Alteration Size	Supposed Origin Mechanism	Clinical Features
1q21.1 microdeletion	-	1,35 Mb	NAHR	Highly variable: mild to moderate ID, microcephaly, multiple congenital abnormalities. It's clear the incomplete penetrance of the alteration since it was find in patients and in unaffected parents.
1q21.1 microduplication	-	1,35 Mb	NAHR	Mild to moderate ID, mild facial dysmorphic features, microcephaly/macrocephaly, autism and autistic features.
1q41q42 microdeletion	-	1,2 Mb	Unknown	ID, seizures, dysmorphic features, cleft palate and diaphragmatic hernia.
2p15p16.1 microdeletion	-	3,9 Mb	Unknown	Moderate to severe ID, feeding problems, microcephaly, optic nerve hypoplasia, hydronephrosis, dysmorphic features, telecanthus, ptosis and downslanting, short palpebral fissures, high nasal bridge, long straight eyelashes.
3q29 microdeletion	-	1,6 Mb	NAHR	Mild to moderate ID, mildly dysmorphic features (including long narrow face, short philtrum and high nasal bridge).
3q29 microduplication	-	1,6 Mb	LCR	Very variable with few common features. Mild to moderate ID, Microcephaly and obesity. The clinical significance of this finding remains somewhat uncertain.
7q11.23 microdeletion	Williams-Beuren Syndrome	1,55 Mb	NAHR	Mild to moderate ID or learning disabilities. Facial features include periorbital fullness, bulbous nasal tip, long philtrum, wide mouth, full lips, full cheeks and small widely spaced teeth. Anxiety, attention deficit hyperactivity disorder, overfriendliness and congenital heart defects.
7q11.23 microduplication	-	1.5 Mb	NAHR	Speech and language delay, mild growth retardation and subtle facial dysmorphism (high and broad nose, short philtrum). Intellectual strengths and weaknesses are in direct contrast to those typically seen in Williams-Beuren syndrome. The patient has a normal cardiovascular system and autistic features.
8p23.1 microdeletion	-	3,4 Mb	NAHR	Congenital heart disease, congenital diaphragmatic hernia, developmental delay and a characteristic behavior profile with hyperactivity and impulsiveness.
8p23.1 microduplication	-	3,75 Mb	NAHR	Prominent forehead, arched eyebrows, development delay (not present in all individuals) and variable congenital abnormalities.
9q22.3 microdeletion	-	6,5 Mb	?	ID, hyperactivity, overgrowth, trigonocephaly, macrocephaly and facial dysmorphisms.

9q34.3 microdeletion	-	0,7 Mb	?	Severe development delay and learning disabilities, hypotonia, microcephaly, congenital heart disease, seizures and facial dysmorphisms.
11p11.2 microdeletion	Potocki-Shaffer syndrome	2,1 Mb	?	Development delay, ID, multiple exostoses, parietal foramina, enlarged anterior fontanel, minor craniofacial abnormalities, ophthalmologic anomalies and genital anomalies in males.
11p13 microdeletion	WAGR syndrome	1 Mb	?	WAGR: Wilms tumor, Aniridia, Genitourinary abnormalities, growth retardation and ID are the cardinal features of the syndrome. In females, malformations of the uterus and/or vagina may also be present.
12q14 microdeletion	-	3,4 Mb	NHEJ	Mild ID, failure to thrive, proportionate short stature and osteopoikilosis.
14q11.2 microdeletion	-	0,4 Mb	?	ID, widely spaced eyes, short nose with flat nasal bridge, long philtrum, Cupid´s bow of the upper lip, full lower lip and auricular abnormalities.
15q13.3 microdeletion	-	1,5 Mb	NHAR	Developmental delay with mild to moderated ID, seizures, subtle facial dysmorphism, hypotonia, short stature and cardiac defects.
15q24 microdeletion	-	1,7 Mb	NAHR	Mild ID, growth retardation, microcephaly, digital abnormalities, genital abnormalities, hypospadias and loose connective tissue. High frontal hairline, broad medial eyebrows, downslanted palpebral fissures and long philtrum.
16p11.2 microdeletion/duplication	-	0,6 Mb	NAHR	Associated with autism, schizophrenia and ID.
16p11.2-p12.2 microdeletion	-	7,1 Mb	NAHR	ID and development delay and subtly dysmorphic facial features including flat faces, downslanting palpebral fissures, low-set ears and eye anomalies.
16p13.1 microdeletion	-	1,7 Mb	LCRs	ID, microcephaly, epilepsy, short stature. A wide range of phenotypic variability.
16p13.1 microduplication	-	1,7 Mb	LCRs	Associated with autism (significance uncertain); Neurocognitive disorders susceptibility locus.
17p11.2 microdeletion	Smith-Magenis Syndrome	3,7 Mb	NAHR	Mild to severe ID, short stature, obesity, small hands and feet, hypotonia, developmental delay and behavior disturbance (sometimes food-searching behavior).
17p11.2 microduplication	Potocki-Lupski syndrome	3,7 Mb	LCR	Mild ID to borderline normal cognitive function. Attention deficit disorder, hyperactivity and autistic features. Most have short stature and cardiovascular anomalies.
17q21.31 microdeletion	-	0,5 Mb	NAHR	Very variable phenotype. Neonatal hypotonia, poor feeding in infancy, oromotor dyspraxia, moderated ID and friendly/amiable behavior. Epilepsy, heart defects (ASD, VSD) and kidney/urological anomalies.

19q13.11 microdeletion	-	0,7 Mb	?	ID, pre and postnatal growth retardation, primary microcephaly, hypospoadias, ectodermal dysplasia, dysplastic nails and dry skin.
22q11.2 microduplication	-	3,7 Mb	NAHR	Highly variable. ID and facial dysmorphisms (widely spaced eyes, downslanting palpebral fissures, velopharyngeal insufficiency, conotruncal hearth disease.
22q11.2 distal microdeletion	-	1,4 Mb	NAHR	ID, prenatal and postnatal growth delay, mild skeletal abnormalities, arched eyebrows, deep-set eyes, smooth philtrum and thin upper lip.
Xq28 microduplication		0,4 Mb	?	ID, severe hypotonia, progressive lower limb spasticity, absent or very limited speech.

Cytogenetic diagnostic: past, present and future

Conventional cytogenetic testing has been available for almost 40 years and has the advantage of being a widely accepted and used technique. In contrast, molecular karyotyping is a relatively new and more diverse technique concerning platforms used, genome coverage and data interpretation. A review study performed by Miller D *et al* in 2010 concluded that molecular kayotyping detected pathogenic genomic imbalances in 10% more patients than G-banded karyotyping alone. However, it is important to say that this review was performed in 33 published studies (total of 21 698 subjects) of which only 10 of the publications were performed in oligo-based arrays platforms. Additionally, these types of platforms often revealed higher positives yields compared with the BAC-based studies also covered by the review [114].

G-banded karyotyping techniques can sometimes detect genomic imbalances as small as 3 Mb depending on the genomic region involved and/or assay conditions. However, it can easily miss alterations in the 5-10 Mb range the analysis of which is based on “subjective” assessment by technicians, leading to variability in detection rates among laboratories [114].

Currently many clinical geneticists and pediatricians are increasingly requesting molecular karyotyping studies to clarify the cause of the disease in children with DD/ID [133]. In many countries MC is already used as the first-tier genetic test instead of G-banded karyotype and current international recommendations of the International Standard Cytogenetic Array (ISCA) Consortium (https://www.iscaconsortium.org/) and the American College of Medical Genetics (http://www.acmg.net//AM/Template.cfm?Section=Home3) propose this approach in first place as well (5,66). Despite being a relatively expensive technique, it may cost less than a regular karyotype analysis plus a customized FISH, such as subtelomeric FISH and has increased yield. Following these same guidelines, the circumstances in which traditional cytogenetic methods are indicated instead of MC is when the patient has a recognizable chromosomal syndrome such as trisomy 21, trisomy 13, Turner syndrome or Klinefelter syndrome [114].

In order to replace conventional karyotyping MC has to provide high resolution whole-genome coverage. This higher resolution across the entire genome also raises concerns about the interpretation of CNVs of unknown significance leading to hesitation about adoption of MC as a first-tier test. However this issue can be adequately addressed through (I) choosing a level of resolution that balances sensitivity and specificity, (II) increased data sharing trough national and international databases and (III) parental studies to determine whether CNVs are de novo or inherited [5, 7, 66].

During the last years enormous technical progress has been achieved not only the array-technology field but also in DNA sequencing technologies. These two technologies are leading to new strategies in the analysis of monogenic diseases allowing the identification of causative mutations in disorders which weren´t previously possible to identify by linkage analysis and positional cloning. Over the past two decades the automated Sanger sequencing was used as the

dominant method in the sequencing industry and led to a number of accomplishments including the completion of the human genome sequence. Despite many technical improvements during the years, the limitations of automated Sanger sequencing showed a need for a new and improved technologies for sequencing large numbers of human genomes [150].

Next generation sequencing (NGS) or massive parallel sequencing (MPS) has some fundamental differences compared to conventional Sanger sequencing. Sanger sequencing always followed the same technical principle (sequencing was directed to a candidate target regions), but NGS uses a number of different technical approaches and is not directed, except if there is a previous enrichment step. NGS is based on sequencing clonally amplified single molecules of genomic DNA and needs between 10-50 reads of the same base to reliable identify heterozygous sequence variants since every read shows only one of the possible two alleles in the sequence. Other advantages of NGS are the possibility of dosage quantification, simultaneously with sequence determining, and the fact that is much cheaper than Sanger sequencing on a per basis manner (although still quite expensive on a genome scale). Another important aspect contributing to the NGS limitations nowadays is the huge amount of data generated and the even more dramatic limitations in accurately distinguish the disease-associated alterations from the benign and evolution-related genome aspects. Besides the limitations faced today by this technology, we can´t overlook the fact that many obstacles now attributed to next-generation sequencing were experienced in much of the same way by Sanger sequencing and MC at their early stages [68, 69].

In the years ahead of us, we will certainly see NGS acquiring the “first-tier test” position leaving behind MC in the same way that we watch in the present regular karyotyping being left behind by MC. Nevertheless, we cannot forget that with these new powerful diagnostic instruments come large amounts of genomic data and with this come increased responsibility in the data interpretation, achieved conclusion and communication to the patient, also with new ethical issues.

Aims of the work

The main goal of this thesis was to study 130 consecutive patients with idiopathic ID using an aCGH platform (Agilent 180K), in order to contribute for the etiologic classification of these cases. The laboratorial work of this thesis was divided in 3 parts: (I) establishment of a DNA bank from blood samples for the patients the patients, both parents and, whenever significant, other relatives; (II) study genomic imbalance for the whole genome in patients with idiopathic ID by aCGH technique; (III) to determine the contribution of the CNVs found in the patients to their phenotype.

The specific aims of this work were:

1. Establishment of a DNA bank for trios of families for which the child have unexplained ID. The establishment of these DNA bank will allow the possibility for future study of the negative cases by different techniques.
2. Establishment of a neurocognitive profile for all the patients involved in the study for further correlations of genetic and clinical features.
3. Discovery of CNVs that allow the etiologic classification of the disease in the patients.
4. Discovery of previously unreported CNVs that can account for the phenotype of ID.
5. Pinpoint the genes, within the CNVs found, that may be important for nervous system development or synaptic functioning.
6. Conclude about the utility of using these technologies for diagnosis of idiopathic ID in the clinical context.

Chapter 2

Results and Discussion

Copy number variations in Portuguese patients with idiopathic intellectual disability

Foreground for the work

The increasing number of publications and guidelines referring aCGH as the first line diagnostic tool while testing children with ID gives rise to the need for implementation of such technique as a diagnostic tool in Portugal. The performance of a pioneer study using aCGH to analyze patients with previously unknown ID etiology, has advantages such as (I) the determination of detection rate in previously negative cases and (II) the epidemiologic study of the Portuguese ID patients for CNVs type and incidence.

The determination of the CNVs responsible for this disorder in patients either with syndromic and non-syndromic ID is important for the clinical correspondence in a future diagnostic perspective. The discovery of previously reported as well as new CNVs in a set of well characterized patients is also relevant for a research point of view as enabling the possibility for the discovery of new genes associated with the clinical features of the patient and in the elucidation of the mechanisms involved in neuronal development, maturation and functioning of the nervous system.

Since the real extent of CNVs occurrence in the human genome is not yet completely characterized and this kind of variations usually occur *de novo* and spread along the genome, another challenging aspect in the molecular study of a group of Portuguese patients is the lack of knowledge regarding the genomic landscape of the population *per se*. This aspect may give rise to problems in distinguish from pathogenic CNVs from rare and population specific benign variants. Additionally, the determination of the alteration effect at molecular and functional level can also become a challenge. In order to accurately distinguish a pathogenic CNV from a benign alteration it is necessary (I) search for the alteration in described control populations, (II) to study the parents and any other relevant sibling for the presence of the alteration and (III) review the literature for the genes involved in the alteration and try to hypothesize about their relevance in the molecular pathways involved in ID.

Material and methods

Patients

We have collected detailed clinical information and DNA samples from 260 patients with idiopathic ID. So far we have performed aCGH and analyzed data for a total of 130 patients (number of patients described in this study).

The vast majority of the patients were observed by a clinical geneticist, a neuropediatrician and a psychologist.

All the 130 patients have detailed clinical information regarding the clinical genetic field, 59 were observed by a neuropediatrician and 128 were evaluated by a psychologist. Their clinical features and cognitive profile have been registered in detail, but in an anonymous manner, in an online secure database in parallel to the genetic study, allowing collaborators from different regions of Portugal to share and discuss the information (http://www.neurodevgen.org/). Clinical and genetic information is kept anonymous and confidential in the same database. A DNA bank from blood samples for patients and families was also established.

All the patients included in the study tested negative for the routinely performed laboratory tests such as, documented high-resolution G-banding karyotype, flourescence in situ hybridization (FISH) studies when specific syndrome suspected and fragile X testing when clinically indicated (FRAXA and FRAXE). The eligibility criteria for the patients included in the study are listed in more detail in table 2.1.

All the patients involved in the study were selected by clinical geneticists from Centro de Genética Médica Jacinto Magalhães do INSA (Porto), and neuropediatricians from Hospital de Geral de Stº António and Hospital Pediátrico de Coimbra.

Table 2.1 – Eligibility criteria for the studied patients. All the patients included in the study needed to fulfill the criteria described below.

Eligibility criteria for the studied patients	
Inclusion criteria	
1.	Documented developmental delay/ ID on basis of an individually administered IQ test equal/below 70 for individuals with ages 3 or higher;
2.	Dysfunction/impairment in more than 2 areas of: communication, self-care, home living, social/interpersonal skills, use of community resources, self direction, functional academic skills, work, leisure and safety.
3.	Documented developmental delay on basis of clinical evaluation by a paediatrician for individuals with ages between 1 and 3 years of age. Individuals included in this group will be individually administered IQ test when they reach age of three.
4.	Unknown aetiology/unexplained developmental delay/ID.
5.	Onset of the developmental delay/ID during childhood.
6.	Previous normal investigation including: baseline general investigation including biochemistry workup (renal, liver, thyroid function tests) and full blood count; documented high-resolution karyotype (>650 bands); fragile X testing when clinically indicated (FRAXA and FRAXE); ATRX (HgH inclusion bodies in red cells); metabolic screen when clinical findings or laboratory anormalities suggesting a metabolic disorder; flourescence in situ hybridization (FISH) studies when specific syndrome suspected (e.g. VeloCadioFacial, Williams-Beuren, Smith-Magenis...); pregnancy TORCH serologies if available or Guthrie's
7.	DNA of the proband and both parents available.
Exclusion criteria	
1.	Not meeting any of the inclusion criteria;
2.	Presence of a recognizable pattern pattern of anomalies (known syndrome or association);
3.	Establishment of the aetiology, or specific diagnosis during the ongoing study.

Within the 130 patients studied in this work, 56 patients have non-syndromic ID while 74 were syndromic. We have colected a total of 126 trios; for 4 patients, only one of the parents was collected. A total of 40 females and 90 males were included in this analysis. Only 53 patients characterized didn´t present familial history of ID, congenital abnormalities and/or repetitive abortions. 77 of the cases had familial history of ID, CA or epilepsy. Of those 77 cases, 80% presented familial history of ID, 19.5% familial history of congenital abnormalities and 13% familial history of epilepsy.

Neurologic alterations were present in were present in 48 of the cases (epilepsy in 33% of the cases, ataxia in 16%, tremor in 4.2%, hypotony in 46% and diminished sensibility to pain in 2%), Ophtalmologic impairments were found in 47 cases while otorrinolaringologic alterations were present in 19 cases. Electroencephalogram analysis revealed alterations for 13 cases while nuclear magnetic resonance and/or computed axial tomography revealed alterations for 44 cases.

Facial dysmorphisms were present in 74 cases and a total of 37 patient presented head circumference alteration (26 cases with microcephaly and 11 cases with macrocephaly).

Behavior abnormalities were present in 70 of the patients (autism or autistic features in 12 cases, aggressiveness in 25 cases and hyperactivity in 52 cases. Sleep disturbances were aslo present in 6 cases.

Skeleton alterations were found in 19 patients, skin alterations in 36 cases and stereotypies in 33 patients.

DNA extraction

Genomic DNA was extracted from peripheral blood using the Citogene® DNA isolation kit (Citomed, Portugal) according to the manufacturer´s protocol (http://www.citomed.com/index.php?option=com_content&view=article&id=46&Itemid=63).

Oligonucleotide aCGH

The aCGH analysis was performed on a human genome CGH Agilent 180K custom array designed by the *Low Lands Consortium* (*LLC*, Professor Klass Kok) in order to be used in the analysis of children with ID/DD (AMADID:023363; Agilent, Santa Clara, CA). The array contained about 180.000 in situ synthesized 60-mer oligonucleotide probes with a mean resolution of approximately one probe every 17Kb (the resolution can reach 11Kb in the RefSeq genes rich regions) designed specially for the study of neurodevelopment and/or congenital abnormalities associated disorders (probes enrichment in regions known to be associated with DD/ID spectrum disorders). In the lower resolution areas, the array is garranteed to detect at least CNVs above 75Kb. DNA labeling was performed using the ENZO Labeling Kit for Oligo Arrays (Enzo Life Sciences, Inc.). With this kit random hexamers are annealed to a denatured template DNA, and Cy3 or Cy5 dUTPs are then incorporated into new DNA strands by exo-Klenow fragment. A total of 500 ng of DNA was labeled with Cy3 (test DNA) and Cy5 (reference DNA) using Klenow fragment at 37ºC during 4 hours (http://www.enzolifesciences.com/ENZ-42671/cgh-labeling-kit-for-oligo-arrays/). The removal of uncoupled nucleotides as carried out in a MinElute PCR purification kit (Qiagen) according to the manufacturer´s instructions (http://www.qiagen.com/Products/DnaCleanup/GelPcrSiCleanupSystems/MinEluteReactionCleanupKit.aspx?r=1644).

As reference DNA was used Kreatech´s MegaPoll Reference DNA (Kreatech Diagnostics), a homogeneous DNA pool from male or female human genomic DNA which has been isolated from 100 different anonymous healthy individuals. The ethnic background of the DNA poll is 85% of Caucasian and Hispanic and 15% of Afro-American. This product was specifically developed as a reference DNA for aCGH use.

Arrays were then hybridized using the Agilent SurePint G3 Human CGH Microarray Kit. Briefly, samples and controls were hybridized in the presence of Cot-1 DNA and blocking agents for 24 hours, 65ºC and 20rpm. After the hybridization period, the slides were washed and scanned with Agilent Microrray Scanner. Data was extracted with the Agilent Feature Extraction (FE) Software v10.5 using default settings for CGH hybridizations.

HUMARA X-chromosome inactivation (XCI) assay

The X-inactivation assay was carried out in the HUMARA (human androgen receptor) locus, which is a highly polymorphic CAG repeat with a very near Hhal enzyme restriction site, according to the protocol designed by [153]. Enzymatic digestion was carried out in 100 ng of genomic DNA at 37ºC for 16 hours. The restriction enzyme hydrolyzed only the unmethylated alleles.

PCR reaction was carried out for both digested and non-digested DNA in a final volume of 25 µl reaction mixtures (1x enzyme buffer, 37.5 mM of $MgCl^2$, 5 mM of dNTPs, 2.5 µM of each primer and 1U of Taq DNA polymerase (Invitrogen)). The thermal cycling profile (Eppendorf) consisted of an initial denaturation step for 10 minutes at 96ºC, followed by 40 cycles of denaturing at 96ºC for 45 seconds, annealing at 67ºC for 30 seconds and extension at 72ºC for 45 seconds, followed by a final extension at 72ºC for 15 minutes. PCR primers and cycling conditions are described in detail in appendix 1 (table S1).

After PCR gel visualization the products were separated on an automated DNA-sequencer (ABI-3130 Genetic Analyzer, Applied Biosystems, Life technologies) serving the allele peak height as the semiquantitative measure of the amount of PCR product amplified from each allele. The samples were defined as having a deviated XCI patter if they presented a corrected allele ratio out of the 0.33 – 3 range.

Data analysis

Image analysis were performed using the across-array methodology described previously [140]. CGH data was analyzed using Nexus Copy Number 5.0 software with FASST Segmentation algorithm and a minimum of three probes in a region required to be considered a copy number alteration. The FASST Segmentation algorithm is a novel Hidden Markov Model (HMM) based approach that, unlike other HMM methods, does not aim to estimate the copy number state at each probe but uses many states to cover possibilities, such as mosaic events, and then make calls based on a second level threshold. The stipulated minimal thresholds for calling a copy number gain were 0.2 (Copy number Gain) and 0.6 (High Copy Gain) and for calling a copy number loss were -0.2 (Copy number Loss) and -1 (Homozygous Copy Loss).

The arrays design, database consultation and comparative analysis was performed using genome build 36.1/HG18.

Interpretation of CNVs

Interpretation of CNVs found was carried out based on the workflow proposed by [154]. For each patient the total number of CNVs was listed according to the position in the chromosome and

classified according to the workflow represented in figure 2.1. The list of all the CNVs found for the patients with pathogenic or very likely pathogenic alterations can be seen in the appendix 2 of the thesis.

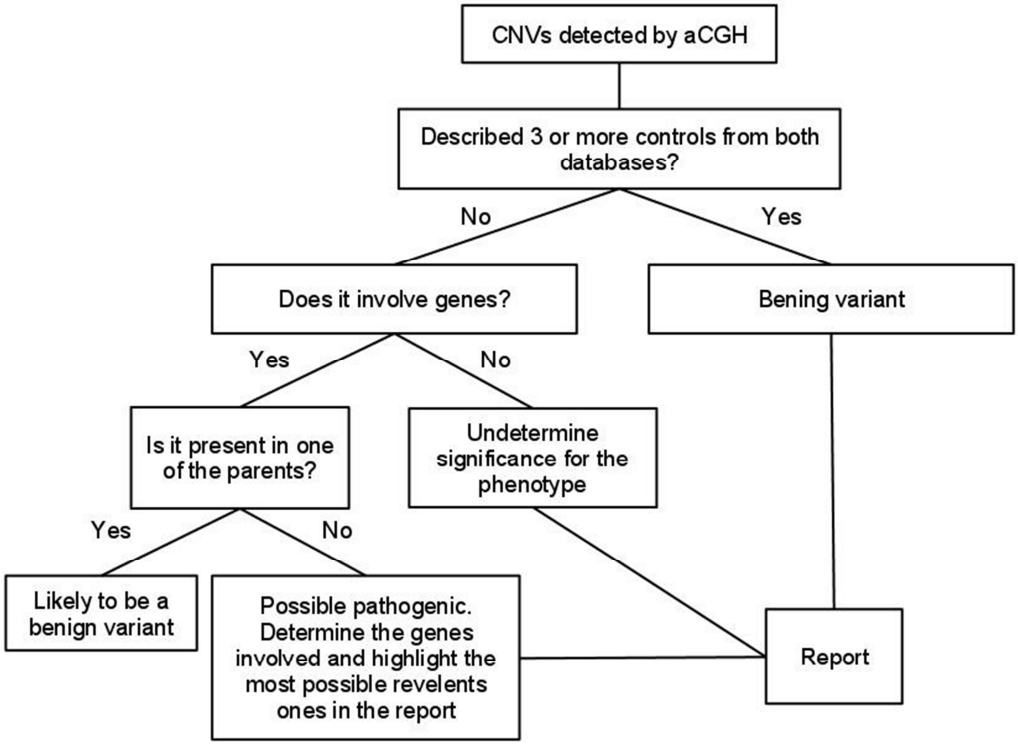


Figure 2.1 – Schematic representation of the workflow followed for the interpretation of the CNVs found in each patient.

Gene Prioritization

The prioritization of the genes discussed for each patient was based on the data regarding its function described in OMIM and PubMed databases. An informatics analysis was also performed for each patient using the TS-CoExp Disease Gene Prediction web tool (http://87.253.99.109/ts-coexp/index.php?site_section=NULL). TS-CoExp basis its predictions on conserved co-expression data at either multi-tissue, tissue-specific or condition specific levels. This tool leveraged with data from the OMIM/PubMed databases regarding described/predicted protein function. For all the patients the search was performed using the "Brain" network and a MinMiner value fo 0, 4. The list of reference syndromes (training set) was obtained by the search in OMIM database for entrances with "mental retardation" or "Intellectual disability" expressions in title, text and synopsis and the results were limited to the entries with phenotype description and known basis. The search

retrieved a list of 727 OMIM entrances (appendix 1, table S2) for the previous criteria. For the discussion were selected the 3 genes with the best p-values.

Web resources

The following web resources were used for the interpretation of CNVs significance in the patients:

- Database of Genomic Variants (http://projects.tcag.ca/variation/), used for search of CNVs described in healthy controls;
- UCSC Genome Bioinformatics (http://genome.ucsc.edu/), used for upload and visualization of "house-detected" CNVs databases;
- DECIPHER (https://decipher.sanger.ac.uk/application/), used for search of clinically significant CNVs;
- Ensembl (http://www.ensembl.org/index.html), used for gene search;
- OMIM- Online Mendelian Inheritance in Man (http://www.ncbi.nlm.nih.gov/omim) used for gene and clinical associated data search;
- PUBMED (http://www.ncbi.nlm.nih.gov/pubmed/), for bibliographic information search;
- WHO - International Statistical Classification of Diseases and Related Health Problems (http://apps.who.int/classifications/apps/icd/icd10online/), for ID classification.
- TS-CoExp - Disease Gene Prediction web tool (http://87.253.99.109/ts-coexp/index.php?site_section=NULL)

Results

We have analyzed the aCGH data from 130 patients and discovered non polymorphic genomic imbalances in 53 patients. For each analyzed genome an average of 25 CNVs with more than 3 consecutive altered probes was found. The lists of all the CNVs found for each patient are present in appendix 1 (table S3). Of the 53 patients, 27 patients had copy number variations of unknown significance and, consequently, further studies need to be performed to clarify the pathogenicity of those alterations. A summary of these alterations is described below:

- 3 cases presented already described microdeletion /microduplication syndromes:
    - *MECP2* gene duplication syndrome (2 patients);
    - 8p23.1 microdeletion syndrome (1 patient);
- In 20 patients we found alterations not described previously and that, considering the size and the implicated genes, are very likely to be pathogenic:
    - 1p36.21p36.23 deletion;
    - 1q43-q44 deletion;
    - 2p13.1-13.3 deletion;
    - 2q11.2-q12.2 deletion;
    - 2q11.2 duplication;
    - 3q22.1-q23 deletion;
    - 5q35.3 dup + 1q21.3-q22 dup + 19p13.2 dup;
    - 5q35.3 duplication;
    - 6q25.3 deletion;
    - 7q11.23 deletion (2 patients);
    - 7q33 deletion;
    - 9q34.2-q34.3 duplication + 14q32.31-q32.33 duplication;
    - 9q33.2-q33.3 duplication;
    - 10q26.3 deletion;
    - 12q24.21-24.2 deletion;
    - 12q24.21 duplication;
    - 20q13.12-q13.13 deletion;
    - Xq24 duplication;
- In 3 patients genomic imbalances concerning susceptibility loci were found:
    - 1q21.1 duplication;
    - 16p13.11 duplication (2 patients);
- For 27 other patients, several smaller but likely significant alterations were found. For those, it is necessary the search in the parents in order to clarify the relevance for the phenotype.

The summary of the alterations as well as the total number of patients for each alteration is described in table 2.2.

Table 2.2 - Summary of the alteration found in the patients.

Classification	Number of patients	Patient code	Description	Size (Mb)	Number of genes involved	Known syndrome	Parents analysis		Total number of patients
							Done	Result	
Known syndromes	2	#1 and #2	Xq28 Dup	2.8; 0.5	78; 16	*MECP2* gene duplication	Yes	*De novo*; Carrier mother	3
	1	#3	8p23.1 Del	5.5	70	8p23.1 microdeletion	Yes	*De novo*	
Novel, very likely pathogenic CNVs	1	#4	1p36.21p36.23 Del	6.7	86	-	Yes	*De novo*	20
	1	#5	1q43-q44 *(AKT3)* Del	0.18	2	-	Yes	De novo	
	1	#6	2p13.1-13.3 Del	4	62	-	Yes	*De novo*	
	1	#7	2q11.2-q12.2 deletion	4.5	24	-	No	-	
	1	#8	2q11.2 Dup	1.5	24	-	No	-	
	1	#9	3q22.1-q23 (*FOXL2*) Del	10.2	62	BPES + ID	Yes	*De novo*	
	1	#10	5q35.3 dup + 1q21.3-q22 dup + 19p13.2 dup	0.12 + 0.2 + 0.28	6 + 6 + 6	Inside the Sotos syndrome region	Yes	*De novo*	
	1	#11	5q35.3 Dup	0.9	12	-	No	-	
	1	#12	6q21 dup + 9p22.2 dup +16q24.3 dup + Xq26.3 dup	0.3 + 0.05 + 0.1	3 + 2 + 2	-	Yes	De novo and inherited	
	1	#13	6q25.3 Del	2.6	7	-	Yes	*De novo*	
	2	#14 and #15	7q11.23 Del	0.4; 0.3	11; 12	Inside the WBS region	No	-	
	1	#16	7q33 Del	2	15	-	No	-	
	1	#17	9q34.2-q34.3 Dup + 14q32.31-q32.33 Dup	5.5 + 3	135 + 46	-	Yes	9q34.2-q34.3 dup (*de novo*); 14q32.31-q32.33 dup (*de novo*)	
	1	#18	9q33.2-q33.3 Dup	3.6	52	-	No	-	
	1	#19	10q26.3 Del (EBF3)	0.65	3	-	No	-	
	1	#20	12q24.21-24.2 Del	1	8	-	No	-	
	1	#21	12q24.21 Dup	0.3	2	-	No	-	
	1	#22	20q13.12-q13.13 Del	5.5	88	-	Yes	De novo	
	1	#23	Xq24 Dup	0.3	4	-	Yes	Carrier mother	
Susceptibility loci	1	#24	1q21.1 Del	2.5	43	1q21.1 Del	Yes	Carrier father	3
	2	#25 and #26	16p13.11 Dup	0.8; 0.9	9; 11	16p13.11 Dup	Yes	Carrier mother	
CNVs of unknown significance	27	-	Several	Variable	Variable	-	No	-	27
								TOTAL	**53**

Patients with copy number changes associated with known syndromes

1. *MECP2* duplication syndrome

Patient #1

This patient is a 15 years old boy with profound ID (IQ= 17), familial history of ID (second grade cousin from the mother side) (figure 2.2C), squint, gastroesophageal reflux disease, recurrent respiratory infections, cortical atrophy, autistic features, plagiocephaly (right size of the skull bigger than the left), dysmorphisms, high arched palate, thin upper lip, sunken eyes, kyphosis and ataxia.

aCGH revealed a 2.8 Mb *de novo* duplication at chromosome region Xq28 (chrX:152,001,572-154,881,207) containing 78 genes, of which *MECP2*, *RAB39B*, *GDI1* and *SLC6A8* are good candidates for explaining the phenotype in the patient. Figure 2.2A presents the facial appearance of patient 1 while figure 2.2B shows the alteration detected by aCGH.

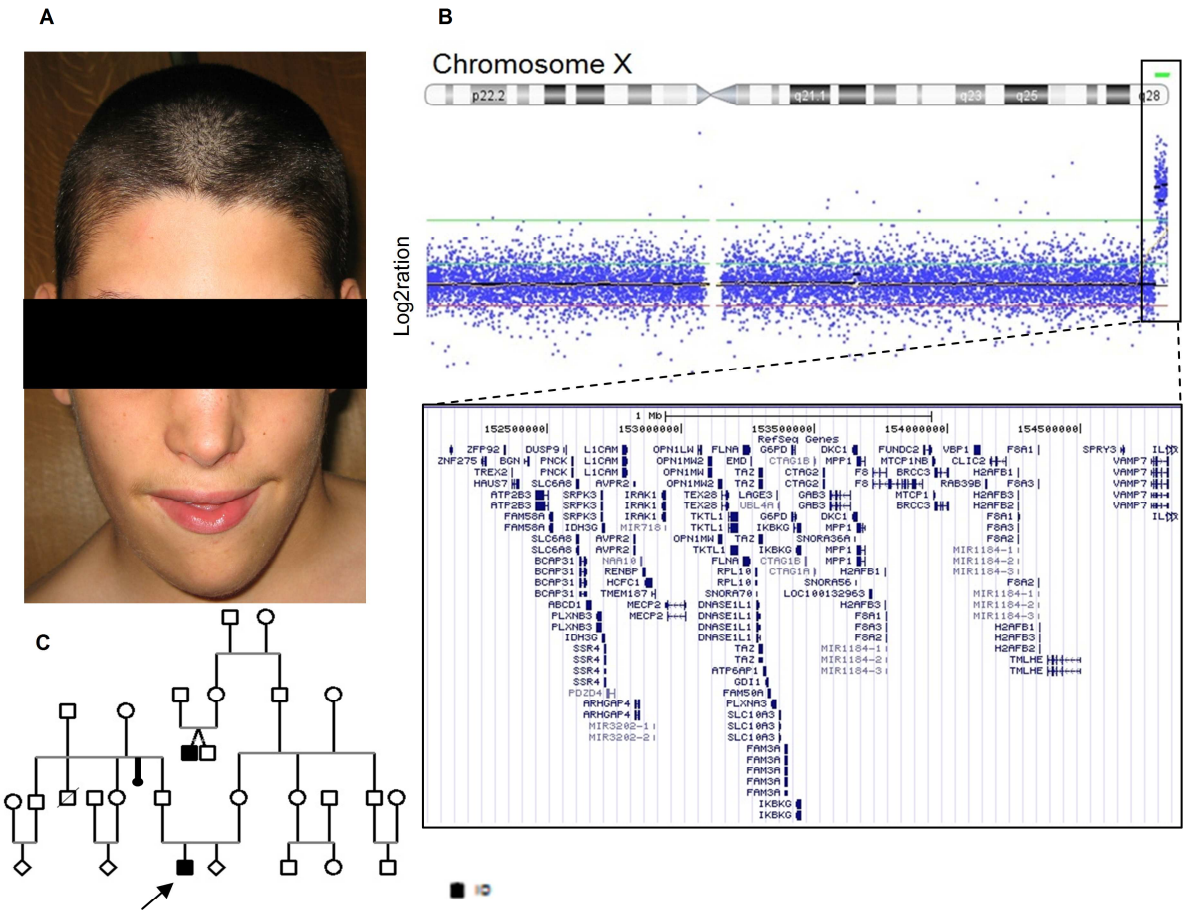


Figure 2.2 – Facial appearance, schematic representation of the genomic alteration and pedigree representation for patient 1. (A) The facial appearance of the patient reveals its plagiocephaly (right size of the skull bigger than the left) thin upper lip and sunken eyes. **(B)** The genomic alteration is represented by the green line above the Xq28 cytoband where it is possible to observe the elevated mean ratio of the probes contained in that region. The UCSC genes present in the genomic region altered in the patient in also represented. **(C)** The patient is represented by the arrow. A second grade cousin of the patient also displays ID, however this individual belong to the fathers side of the family being unlikely that bot phenotypes are related with the same alteration.

The *MECP2* (methyl-CpG-binding protein 2) gene encodes for a homonymous protein (MeCP2) with capacity to both activate and repress transcription. MeCP2 is required for maturation of neurons and is correct development. Mutations in the *MECP2* gene are associated with Rett Syndrome, a severe neurodevelopment disorder that affects 1:10000 girls worldwide. Mutations in *MECP2* were also found in boys with severe infantile encephalopathy, ID and development delay [155].

The *RAB39B* (Ras-associated protein RAB39B) gene encodes a small GTPase involved in the regulation of vesicular trafficking between membrane compartments with high expression levels neuronal precursors and neurons of the brain, mainly in hippocampus [156]. Mutations in *RAB39B* gene were identified in patients with X-linked ID in two different families [156]. Further studies in a mouse model reveal that absence of RAB39B protein resulted in decreased number of growth cones in neuritis and, consequently, decreased number of neuronal branches. This reduced number of neuronal terminals is thought to be associated with deficient synapse formation and instability. Until this date, there is no CNVs described involving *RAB39B* gene in PUBMED or OMIM databases.

The *SLC6A8* (solute carrier family 6; neurotransmitter transporter, creatine; member 8) gene encodes a creatine transporter that was found to be mutated in patients with ID. SLC6A8 protein has a crucial role in the temporal and spatial maintenance of the energy supply to muscle and brain [157]. The lack of SLC6A8 leads to a very strong deficiency of creatine in the central nervous system and patients affected by mutations in this gene can present severe neurodevelopment delay, ID, impairment in speech acquisition and also epilepsy [158].

The *GDI1* (GDP dissociation inhibitor 1) gene encodes a protein necessary for the GDP-GTP exchange reaction of members of the Rab family. Inactive Rab GTPases are maintained in the cytosol by binding to GDP-dissociation inhibitor. GDI1 is highly expressed in the brain and has been considered the most likely candidate for the explanation of ID phenotype in patients carrying duplications at Xq28 region not covering the *MECP2* gene. In fact, GDI1 duplicated copies perfectly correlated with ID severity in patients with this type of alterations [159].

Another 7 genes involved in the duplication encode for proteins that act or may act in different molecular pathways known to be involved in ID etiology. These genes are listed in table 2.3.

Table 2.3 – Summary of other genes covered by the duplication that may contribute for the patient's phenotype.

Gene	Locus	Protein	Function	Phenotype
HAUS7	Xq28	HAUS7	Is a subunit of a complex that is a microtubule-binding complex involved in microtubule generation within the mitotic spindle and is vital to mitotic spindle assembly	Not described
FAM58A	Xq28	FAM58A	Contains a cyclin box-fold domain, a protein-binding domain found in cyclins with a role in cell cycle and transcription control.	STAR syndrome
PLXNB3	Xq28	PLXNB3	Is a plexin with high affinity for semaphorin-5a	Possibly associated with X-linked adrenoleukodystrophy together with ALD gene
L1CAM	Xq28	L1CAM	Is a cellular adhesion molecule found primarily in the nervous system of several species and may be more aptly called a neural recognition molecule.	X-linked hydrocephalus; MASA syndrome;X-linked spastic paraplegia
ARHGAP4	Xq28	ARHGAP4	Is a Rho GTPase-activating protein 4	Possible association with nephrogenic diabetes insipidus
FLNA	Xq28	Filamin A	Actin-binding protein that regulates reorganization of the actin cytoskeleton by interacting with integrins, transmembrane receptor complexes, and second messengers.	X-linked dominant periventricular heterotopia;
PLXNA3	Xq28	PLXNA3	Is a plexin with high affinity for semaphorin-3f	Not described

FLNA (Filamin A) gene encodes a protein named filamin A that regulates reorganization of the actin cytoskeleton by interacting with integrins, transmembrane receptor complexes, and second messengers. Filamins are responsible for the creation of crosslinks between actin filaments generating orthogonal networks in the cytoplasm and participating in the anchoring of membrane proteins to the actin cytoskeleton. Remodeling of the cytoskeleton is a central process involved in the modulation of cell shape during migration [160]. Mutations in *FLNA* gene have been identified in patients with periventricular heterotopia and otopalatodigital syndrome spectrum disorders. Otopalatodigital syndrome is characterized by malformations that arrive from the incorrect development of craniofacial and long bones [161]. Although there are no duplications described affecting the entire gene in otopalatodigital syndrome patients, the duplication of this gene in our patient may account for the facial dysmorphisms and unequal skull structure of the patient.

Even though the *MECP2* duplication syndrome is a severe neurodevelopment disorder affecting males, patient 1 seems to have a more severe phenotype when compared with the reported features in the literature. This aspect may be due the due the fact that the alteration in patient 1

encompasses three additional ID-related genes. The predicted outcome of duplications covering entire genes is not well characterized. Even if the isolated extra amount of genomic DNA for these genes resulted in a benign alteration, patient 1 has at least three more genes related with ID (besides *MECP2*) that are duplicated at the same time. We believe that the *MECP2* duplication syndrome phenotype in the patient is due to *MECP2* alteration, while the others genes are contributing for the severity of the phenotype and malformations found in the patient.

TS-CoExp analysis for patient 1 retrieved *RAB39B* (RAB39B, member RAS oncogene family), *ATP6AP1* (ATPase, H+ transporting, lysosomal accessory protein 1) and *L1CAM* (L1 cell adhesion molecule) genes as the 3 more likely to cause ID, among those covered by the duplication. *ATP6AP1* gene encodes for a vacuolar H(+)-ATPase protein expressed in microglia [162]. Mutations in a member of the same family - *ATP6AP2* (ATPase, H+ transporting, lysosomal accessory protein 2) were described in patients with X-linked ID and epilepsy [163]. The other 2 genes were described above.

A summary of the alteration, genes involved and genes pinpointed by the TS-CoExp tool is represented in table 2.4.

Table 2.4 – Summary of the alteration in patient 1 and genes involved. In bold are represented the most likely significant gene for the phenotype based on their function.

Patient Number	Event	Number of Genes Involved	Number of Probes Involved	Genes	TS-CoExp
#1	Xq28 (152,001,572-154,881,207) X 2	73	260	*ABCD1, ARHGAP4, ATP2B3, ATP6AP1, AVPR2, BCAP31, BGN, BRCC3, CLIC2, CTAG1A, CTAG1B, CTAG2, DKC1, DNASE1L1, DUSP9, EMD, F8, F8A1, F8A2, F8A3, FAM3A, FAM50A, FAM58A, FLNA, FUNDC2, G6PD, GAB3, GDI1, H2AFB1, H2AFB2, H2AFB3, HAUS7, HCFC1, IDH3G, IKBKG, IRAK1, L1CAM, LAGE3, LOC100132963, MAGEA1,* ***MECP2****, MIR718, MPP1, MTCP1, MTCP1NB, NAA10, NCRNA00204, NCRNA00204B, OPN1LW, OPN1MW, OPN1MW2, PDZD4, PLXNA3, PLXNB3, PNCK,* ***RAB39B****, RENBP, RPL10, SLC10A3,* ***SLC6A8****, SNORA36A, SNORA56, SNORA70, SRPK3, SSR4, TAZ, TEX28, TKTL1, TMEM187, TMLHE, TREX2, UBL4A, VBP1, ZFP92, ZNF275, IL9R, SPRY3, VAMP7*	*RAB39B, ATP6AP1, L1CAM*

Patient #2

This patient is a 5 years old boy with mild ID (IQ= 62) with familial history of ID (brothers) (figure 2.3C), congenital abnormalities, moderate and difuse brain atrophy, brachycephaly and light plagiocephaly, round face, reduced vision, sepsis and constipation during the neonatal period.
aCGH revealed a 0.5 Mb duplication at chromosome region Xq28 (chrX:152,783,739-153,255,487) containing 16 genes, of which the *MECP2* gene is a good candidate for explaining the phenotype in the patient. The duplication was inherited from the mother and it is very likely that the two affected brothers of the patient carried the same alteration. Figure 2.3A presents the facial appearance of patient 2 while figure 2.3B shows the alteration detected by aCGH.

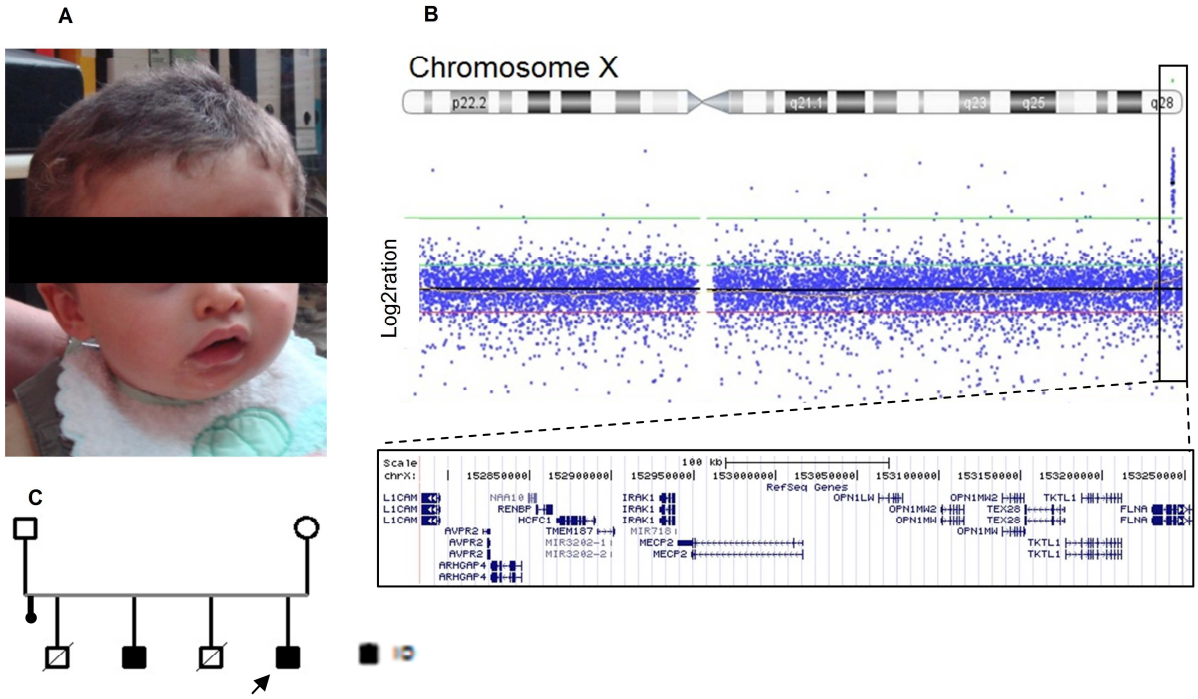


Figure 2.3 - **Facial appearance, schematic representation of the genomic alteration and pedigree representation for patient 2. (A)** The facial appearance of the patient reveals its braquicephaly, plagiocephaly and round face. **(B)** The genomic alteration is represented by the green line above the Xq28 cytoband where it is possible to observe the elevated mean ratio of the probes contained in that region. The UCSC genes present in the genomic region altered in the patient in also represented. **(C)** The patient is represented by the arrow. The patient´s mother had 1 miscarriage and 4 live-born sons, two of which passed away. The patient has a brother alive that also displaysID.

Patient 1 and 2 share duplication of *MECP2* gene. The predominant clinical features of males with *MECP2* duplication syndrome are DD/ID, hypotonia, absent or severely delayed and history of recurrent infections. Less common but also present are genital or digital abnormalities; seizures and autism are often present. Many times, the reported cases of *MECP2* duplications are familial cases where the mothers are asymptomatic carriers and often display skewed X chromosome inactivation. This is probably the case present for patient 2. The fact that the mother carries the duplication indicates that the siblings (also with ID) probably inherited the duplicated X chromosome from the mother as well. It is our knowledge that one of the siblings passed away at 1

month of age after a premature labour. It would be interesting to study the other brother for the same alteration.

Similar to patient 1, patient 2 also displays plagiocephaly. The fact that the *FLNA* gene is also duplicated in this patient lead us to believe that it may also account for the cranial abnormalities in patient 2.

We also believe that the differences between the ID severity in both cases is due to the size of the duplication that, in patient 2, doesn´t cover any other previously described ID associated gene.

The *MECP2* duplication syndrome is 100% penetrant in males, leading to severe ID, infantile hypotonia and speech impairment. Table 2.5 describes the main features presented by patient 1 and 2 and compared with the clinical accepted phenotype for the *MECP2* duplication syndrome patients.

Table 2.5 – Comparison between patient 1 and 2.

***MECP2* duplication patients**	**Patient 1**	**Patient 2**
Severe ID	Present (Profound ID)	Present (Mild ID)
Poor or absent speech	Absent	Poor
Hypotonia	No	No
Progressive spasticity	Not determined	Not determined
Recurrent respiratory infections	Present	Not determined
Seizures	No	No
Severe impairment in mobility	Yes (ataxia)	No
Autistic features	Present	No
Gastrointestinal dysfunction	Present	Present

TS-CoExp analysis for patient 2 retrieved *L1CAM* , *MECP2* and *FLNA* genes as the 3 more likely to cause ID, among those covered by the duplication. These 3 genes were described above.

A summary of the alteration, genes involved and genes pinpointed by the TS-CoExp tool is represented on table 2.6.

Table 2.6 - Summary of the alteration in patient 2 and genes involved. In bold are represented the most likely significant gene for the phenotype based on their function.

Patient Number	Event	Number of Genes Involved	Number of Probes Involved	Genes	TS-CoExp
#2	Xq28 (152,783,739-153,255,487) X 2	16	57	*ARHGAP4, AVPR2, FLNA, HCFC1, IRAK1,* ***L1CAM****,* ***MECP2****, MIR718, NAA10, OPN1LW, OPN1MW, OPN1MW2, RENBP, TEX28, TKTL1, TMEM187*	*L1CAM, MECP2, FLNA*

The XCI study for the mother of the patient 2 retrieved no conclusive results since both alleles for HUMARA studied locus have the same repeat number, making it impossible to determine the XCI pattern for this locus.

2. 8p23.1 deletion syndrome

Patient #3

This patient is a 13 years old girl with moderate ID (IQ= 48) with microcephaly (head circumference 4.4 standard deviations below the normal), lower weight, arched eyebrows, high arched palate, pulmonary stenosis, ventricular septal defect, hyperactivity and attention deficit. aCGH revealed a 5.5 Mb *de novo* deletion at chromosome region 8p23.1 (chr8:7,026,686-12,529,929) containing 70 genes. Figure 2.4A presents the facial appearance of patient 3 while figure 2.4B shows the alteration detected by aCGH.

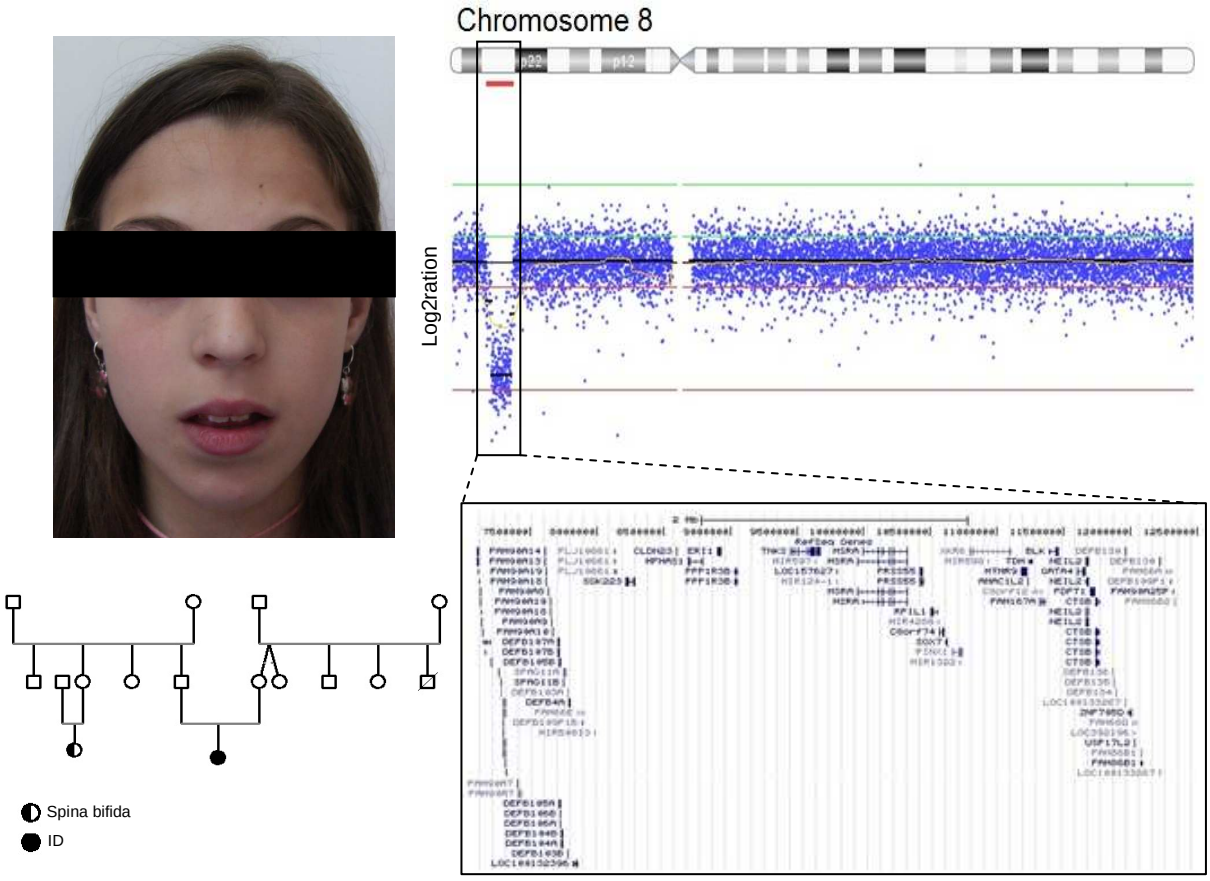


Figure 2.4 - Facial appearance, schematic representation of the genomic alteration and pedigree representation for patient 3. (A) The facial appearance of the patient reveals its microcephaly and arched eyebrows. **(B)** The genomic alteration is represented by the red line bellow the 8p23.1 cytoband where it is possible to observe the diminished mean ratio of the probes contained in that region. The UCSC genes present in the genomic region altered in the patient in also represented. **(C)** The patient is represented by the arrow.

The 8p23.1 deletion syndrome is a clinically recognized syndrome characterized by congenital heart abnormalities (atrial and/or ventricular septal defects and general heart abnormalities), congenital diaphragmatic hernia, hyperactivity, intellectual disability and development delay[164]. In

general terms, the congenital heart defects presented by the patients are attributed to the *GATA4* (GATA-binding protein 4) gene haploinsufficiency. GATA4 is a zinc finger transcription factor that belongs to a group of transcription factors that recognize a consensus sequence known as "GATA" motif present in the promoter region of many genes. GATA-binding proteins control gene expression and cell differentiation in a variety of tissues [165]. Due to the large variability of heart malformations found among patients carrying 8p23.1 deletions, it was hypothesized that maybe not only the *GATA4* gene plays a crucial role in heart formation. Additionally to this gene, *SOX7* (SRY-box 7) has been associated with the cardiac phenotype seen in the patients. *SOX7* encodes a protein containing a high mobility group (HMG) characteristic of proteins that are transcription factors with critical roles in the regulation of diverse developmental processes [166]. It seems that patients carrying deletions involving both *GATA4* and *SOX7* have more severe cardiac defects [167]. A recent publication by Ballarati L and colleagues, report the case of a boy with an alteration very similar to the one present in our patient [168]. As in the reported case, our patient has a 5 Mb deletion flanked by two low copy repeats created by the olfactory receptors gene clusters REPD and REPP [169, 170]. Figure 2.5 represents the alignment of the alteration found in patient 3 with the ones reviewed in Ballarati *et al*, 2011.

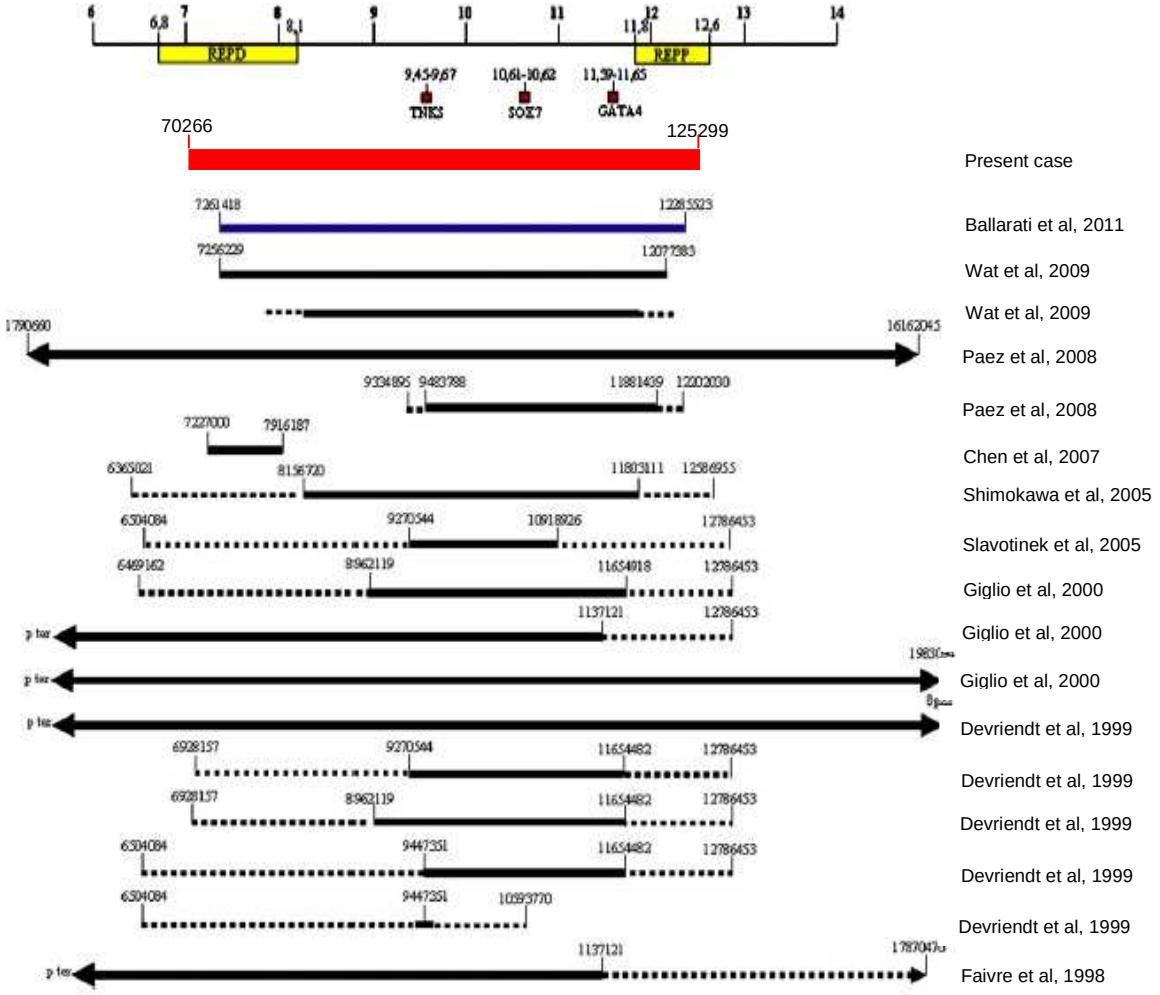


Figure 2.5 – Schematic comparison of the alteration present in patient 3 with the ones described in the literature. Patient 3 (in red) carries a 5.5 Mb deletion highly similar to the one described by [168, 167].

The 8p23.1 deletion has an associated clinical phenotype of growth impairment, ID, facial dysmorphisms, cardiac abnormalities, microcephaly and psychiatric troubles. Our patient has a 5 Mb deletion and shows the main clinical features of the syndrome. As the deletion is located inside the two LCR clusters that are thought to be responsible for the origin of this imbalance, we believe that NAHR was the originating mechanisms for the deletion in the patient. As was said before, our patient has a very similar alteration to the one recently described by Ballarati *et al*, 2011. The clinical comparison is made in table 2.7.

Table 2.7 – Clinical comparison between the patient described by Ballarati *et al*, 2001 [168] and patient 3.

8p23.1 deletion syndrome	Ballarati *et al*, 2011 [168]	Present case
Cardiac malformations	Yes (ventricular septal defect)	Yes (ventricular septal defect)
Development delay	Yes	Yes
ID	Yes (mild ID)	Yes (moderate ID)
Diaphragmatic hernia	No	No
Hypospadias	No	Not applicable
Hyperactivity/attention deficit/impulsiveness	Yes (Attention deficit disorder)	Yes (Attention deficit disorder)
Microcephaly	Yes	Yes
Dysmorphic facial features	Yes (sloping forehead, high nasal bridge, deep-set eyes)	Yes (arched eyebrows, high arched palate)

Another interesting gene present in the deleted region is *TNKS* (TRF1-interacting, ankyrin-related ADP-ribose polymerase), which was recently proposed as responsible for Cornelia de Lange syndrome[1] [171]. This gene encodes a protein (tankyrase) that acts as an acceptor for adenosine diphosphate (ADP) ribosylation. Although *TNKS* function have not been assessed in patients with 8p231 deletion and Cornelia de Lange syndrome yet, the phenotype of the patients is very different and most likely not related with *TNKS* haploinsufficiency.

TS-CoExp analysis for patient 3 retrieved *CTSB* (cathepsin B), *FDFT1* (farnesyl-diphosphate farnesyltransferase 1) and *MTMR9* (myotubularin related protein 9) genes as the 3 more likely to cause ID, among those covered by the deletion.
The *CTSB* gene, also called amyloid precursor protein secretase (APP), suffers a cleavage of the amyloid beta peptide sequence during constitutive processing of its precursor, leading to the deposition of Aβ. It was suggested that a deregulation in this mechanism could be in related with

[1] Cornelia de Lange syndrome is characterized by typical facial dysmorphisms, hirsutism, limb abnormalities, internal organ abnormalities (including diaphragmatic hernia), impaired pre and postnatal growth, microcephaly and ID.

Alzheimer disease[2] [172]. CTSB was found to be overexpressed in tumors of the lung, prostate, colon, breast and stomach [173]. According to TS-CoExp tool, the *CSTB* gene is co-expressed in the brain with *PRPS1* (phosphoribosyl pyrophosphate synthetase 1) gene. *PRPS1* catalyzes the phosphoribosylation of ribose 5-phosphate to 5-phosphoribosyl-1-pyrophosphate, necessary for purine and pyrimidine biosynthesis pathways [174]. Mutations in this gene were described in patients with neurodevelopment abnormalities, deafness and hyperuricemia[3]. However, these phenotypes were found in patients carrying mutations causing gain of function in the protein leading to phosphoribosyl pyrophosphate synthetase 1 superactivity [175].

Farnesyldiphosphate farnesyltransferase (FDFT1) catalyzes the conversion of trans-farnesyldiphosphate to squalene during cholesterol biosynthesis pathway. Patterson et al, 1995 studied the breakpoints of an 8p23.1 inversion and verified that it affected the 5-prime coding region of the *FDFT1* gene. This alteration was found in patients with Recombinant chromosome 8 syndrome[4], an autosomal dominant disorder characterized by growth retardation, multiple congenital malformations (such as cardiac and genitourinary defects), postnatal microcephaly, facial dysmorphisms and DD/ID [176]. These findings suggest *FDFT1* gene as a good candidate for the ID phenotype displayed by the patient.

MTMR9 is a myotubularin-related protein that contains a double-helical motif though to be associated with a function in cell proliferation [177]. This characteristic may indicates MTMR9 as a good candidate, together with the previously described gene), for the severe microcephaly of the patient.

A summary of the alteration, genes involved and genes pinpointed by the TS-CoExp tool is represented in table 2.8.

[2] Alzheimer disease is a pathology associated with memory loss and other intellectual abilities. Symptoms usually develop slowly and get worse over time, interfering with daily life activities. Alzheimer's disease accounts for 50 to 80 percent of dementia cases.

[3] Hyperuricemia refers to the abnormally high levels of uric acid in the blood.

[4] Recombinant chromosome 8 syndrome is caused by a recombinant chromosome 8 characterized by duplication of 8q22.1-qter and deletion of 8pter-p23.1.

Table 2.8 - Summary of the alteration in patient 3 and genes involved. In bold are represented the most likely significant gene for the phenotype based on their function.

Patient Number	Event	Number of Genes Involved	Number of Probes Involved	Genes	TS-CoExp
#3	8p23.1(7,026,686-12,529,929)X1	70	275	*AMAC1L2, BLK, C8orf12, C8orf74, CLDN23, CTSB, DEFB103A, DEFB103B, DEFB104A, DEFB104B, DEFB105A, DEFB105B, DEFB106A, DEFB106B, DEFB107A, DEFB107B, DEFB109P1, DEFB109P1B, DEFB130, DEFB134, DEFB135, DEFB136, DEFB4, ERI1, FAM167A, FAM66A, FAM66B, FAM66D, FAM66E, FAM86B1, FAM86B2, FAM90A10, FAM90A13, FAM90A14, FAM90A18, FAM90A19, FAM90A20, FAM90A5, FAM90A7, FAM90A8, FAM90A9, FDFT1, FLJ10661,* ***GATA4****, LOC157627, LOC349196, LOC392196, MFHAS1, MIR124-1, MIR1322, MIR548I3, MIR597, MIR598, MSRA, MTMR9, NEIL2, PINX1, PPP1R3B, PRAGMIN, RP1L1,* ***SOX7****, SPAG11A, SPAG11B, T-SP1, TDH,* ***TNKS****, USP17L2, XKR6, ZNF705D, ZNF705G*	*CTSB, MTMR9, FDFT1*

Patients with novel, very likely pathogenic CNVs

3. 1p36.21-p36.23 deletion

Patient #4

This patient is a 23 years old man with moderate ID (IQ= 49) and with familial history of a second grade cousin with ID (figure 2.6C). The patient presented failure to thrive, microcephaly, broad nasal bridge, hypoplastic nares, microretrognathia, kyphosis, hypertelorism and telecanthus.
aCGH revealed a 6.7 Mb *de novo* deletion at chromosome region 1p36.21-p36.23 (chr1:8,516,261-15,269,259) containing 86 genes. Figure 2.6A presents the facial appearance of patient 4 while figure 2.6B shows the alteration detected by aCGH.

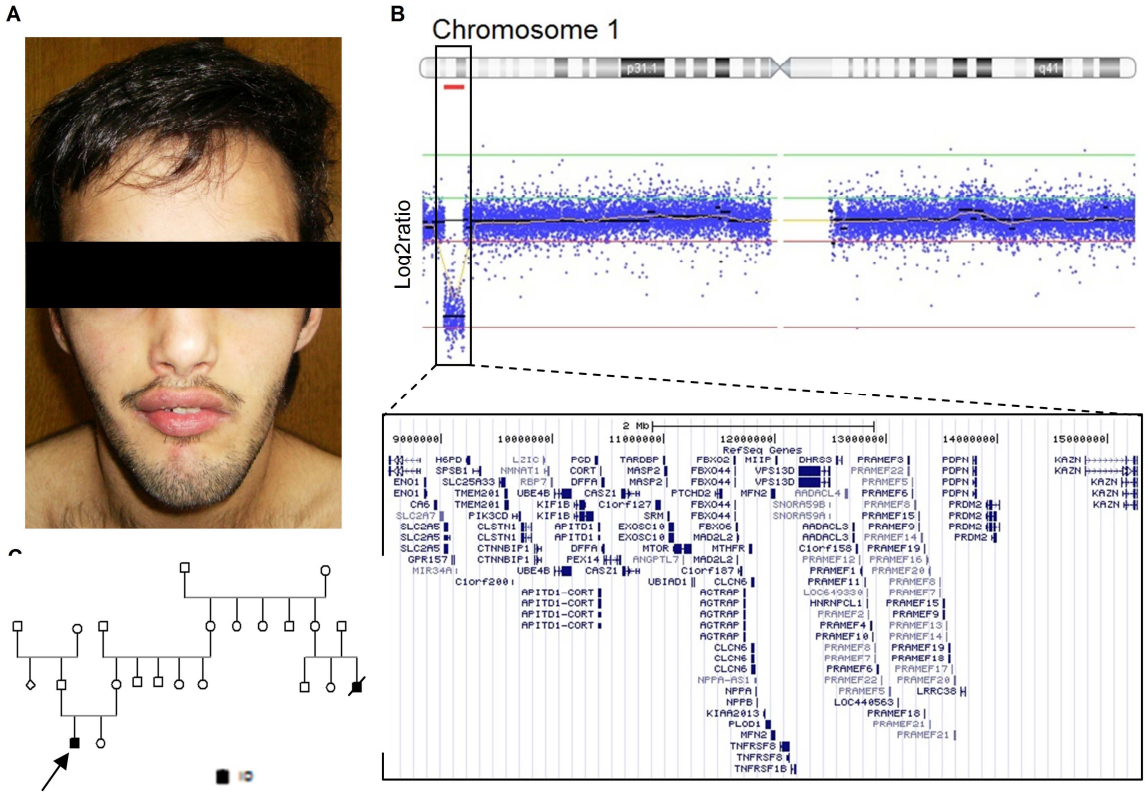


Figure 2.6 - Facial appearance, schematic representation of the genomic alteration and pedigree representation for patient 4. (A) The facial appearance of the patient reveals its microcephaly, broad nasal bridge, hypoplastic nares, microretrognathia, hypertelorism and telecanthus. **(B)** The genomic alteration is represented by the red line bellow the 1p36.21-p36.23 cytoband where it is possible to observe the diminished mean ratio of the probes contained in that region. The UCSC genes present in the genomic region altered in the patient in also represented. **(C)** The patient is represented by the arrow. The patient has a deceased second grade maternal cousin that also had ID.

The PubMed search for alterations described in this region retrieved no results. However, there are alterations described in Decipher database in patients with ID/DD that partially overlap with the one found in our patient. The case described in Dechiper that has the most similar alteration (#251601)

with the one find in patient 4 has a deletion of 6.15 Mb and a clinical presentation of ID, microcephaly, camptodactyly, seizures, short stature, eczema and facial dysmorphisms (down-turned corners of the mouth, lateral hypoplasia of eyebrows, palpebral fissures slant up and pits of ear helix). When compared to our patient, we can see that they share features such as ID, microcephaly and facial dysmorphisms (even though they are not exactly the same).

Regarding the genes altered in both cases, there are 4 genes referred in Decipher as likely pathogenic: *ANGPTL7* (angiopoietin-like 7), *CASZ1* (castor zinc-finger protein 1), *MAD2L2* (mitotic arrest-deficient 2, S. Cerevisiae, homolog-like 2) and *RERE* (RE repeats-encoding gene).

The *ANGPTL7* gene encodes the angiopoietin-like 7 protein that is a molecular of the extra-cellular matrix and belongs to the Angiopoietin-like proteins family. These proteins contain an amino-terminal coiled-coil domain and a carboxyl-terminal fibrinogen-like domain and are known as angiogenic factors. Several members of ANGPTLs also regulate lipid metabolism independently of angiogenic effects, but most of their functions during vertebrate development are not demonstrated, besides the expression of some members of the family in the central nervous system [178].

The *CASZ1 g*ene encodes a zinc finger transcription factor and is a neural fate-determination gene [179].

MAD2L2 encodes proteins that play a role in the mitotic spindle checkpoint [180].

Rere act as transcriptional corepressor by binding to histone deacetylase 1 and 2. *RERE* gene mutation leads to the formation of symmetrical somites in mouse embryos, similar to embryos deprived of retinoic acid. Furthermore, Rere also controls retinoic acid signalling, which is required to maintain somite symmetry [181]. Based on their functional roles, these four genes can be good candidates for explaining the ID and morphologic alterations presented by the patient.

Unexpectedly, TS-CoExp analysis for patient 4 retrieved *PRAMEF2* (PRAME family member 2), *PRAMEF1* (PRAME family member 1) and *PRAMEF13* (PRAME family member 13) genes as the 3 more likely to cause ID, among those covered by the deletion.

These three genes belong to the PRAME gene family, for which the expression limited in normal tissue (such as the testis) and increased in a variety of cancers (such as melanomas) [182]. The most reasonable explanation for the detection of these 3 genes as good candidates for causing ID by the Ts-CoExp web tool is their expression in neurologic tumors.

A summary of the alteration, genes involved and genes pinpointed by the TS-CoExp tool is represented in table 2.9.

Table 2.9 - Summary of the alteration in patient 4 and genes involved. In bold are represented the most likely significant gene for the phenotype based on their function.

Patient Number	Event	Number of Genes Involved	Number of Probes Involved	Genes	TS-CoExp
#4	1p36.21-p36.23(chr1:8,516,261-15,269,259)X1	86	364	*AADACL3, AADACL4, AGTRAP, ANGPTL7, APITD1, C1orf127, C1orf158, C1orf187, C1orf200, CA6,* ***CASZ1****, CLCN6, CLSTN1, CORT, CTNNBIP1, DFFA, DHRS3, ENO1, EXOSC10, FBXO2, FBXO44, FBXO6, GPR157, H6PD, HNRNPCL1, KIAA1026, KIAA2013, KIF1B, LOC440563, LOC649330, LZIC,* ***MAD2L2****, MASP2, MFN2, MIIP, MIR34A, MTHFR, MTOR, NMNAT1, NPPA, NPPB, PDPN, PEX14, PGD, PIK3CD, PLOD1, PRAMEF1, PRAMEF10, PRAMEF11, PRAMEF12, PRAMEF13, PRAMEF14, PRAMEF15, PRAMEF16, PRAMEF17, PRAMEF18, PRAMEF19, PRAMEF2, PRAMEF20, PRAMEF21, PRAMEF22, PRAMEF3, PRAMEF4, PRAMEF5, PRAMEF6, PRAMEF7, PRAMEF8, PRAMEF9, PRDM2, PTCHD2, RBP7,* ***RERE****, SLC25A33, SLC2A5, SLC2A7, SNORA59A, SNORA59B, SPSB1, SRM, TARDBP, TMEM201, TNFRSF1B, TNFRSF8, UBE4B, UBIAD1, VPS13D*	*PRAMEF2, PRAMEF13, PRAMEF1*

4. 1q43-q44 deletion

Patient #5

This patient is a 14 years old girl with mild ID (IQ=63), microcephaly and development delay. The patient doesn´t display any significant dysmorphic features.

aCGH revealed a *de novo* 0.18 Mb deletion at chromosome region 1q43-q44 (chr1:241,618,630-241,805,298) containing 2 genes. Figure 2.7A presents the facial appearance of patient 5 while figure 2.7B shows the alteration detected by aCGH.

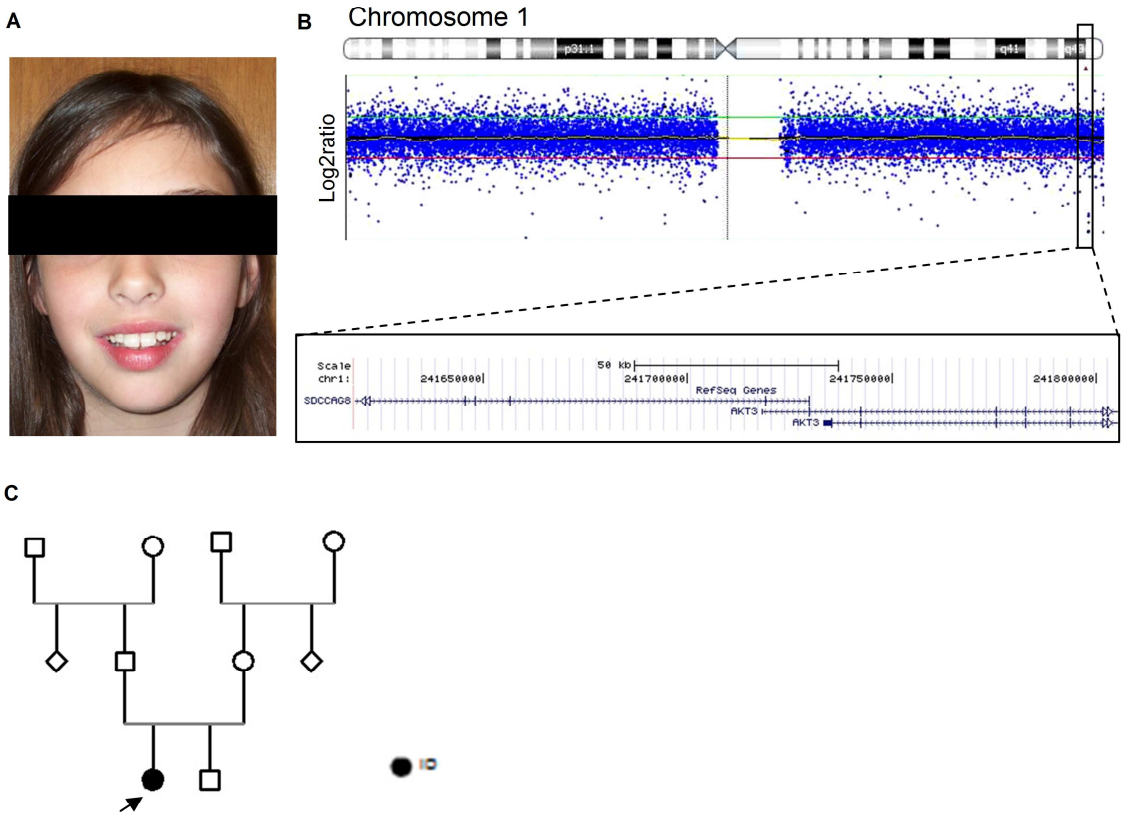


Figure 2.7 - Facial appearance, schematic representation of the genomic alteration and pedigree representation for patient 5. (A) The facial appearance of the patient reveals her lack of significant facial dysmorphisms. **(B)** The genomic alteration is represented by the red line bellow the 1q43-q44 cytoband where it is possible to observe the diminished mean ratio of the probes contained in that region. The UCSC genes present in the genomic region altered in the patient in also represented. **(C)** The patient is represented by the arrow.

If located in tandem, the alteration present in this patient might disrupt two genes: *SDCCAG8* (serological defined colon cancer antigen 8) and *AKT3* (V-AKT murine thymoma viral oncogene homolog 3). The *SDCCAG8* gene encodes for a protein thought to be a stable centrosomal component with a structural role in the centrosomal architecture or the microtubule-organizing activities of the centrosome matrix. Recently, mutations in *SDCCAG8* gene were described in patients with nephronophthisis-related ciliopathies [183]. Even though some features of these patients

include ID, the alteration that is causing disease is recessive [183]. On the other hand, AKT3 is highly expressed in the brain and is implicated in numerous biological processes including cell proliferation [184]. Deletions and translocation breakpoints have been described for region 1q44 in patients with microcephaly and agenesis of the corpus callosum. It is important to refer that not all the patients described in the paper that had deletion mapping for the region presented agenesis of corpus callosum. Some of them presented microcephaly without anomalies in corpus callosum, a phenotype that is partially overlapping with that of our patient. The authors propose the *AKT3* gene as an excellent candidate for developmental human microcephaly and agenesis of the corpus callosum [185].

Another important aspect to highlight is that there are patients described in Decipher with similar genomic imbalances (Decipher# 252432 and #252434) for which no phenotypic description is provided. Decipher# 252432 carries a 0.12 Mb deletion inherited from a normal parent covering a very similar alteration to the one present in patient 18. Decipher#252434 carries a 0.35 Mb *de novo* deletion that affects one gene more than the one present in our patient. In spite of the presence of a similar case eventually not associated with disease in Decipher, we cannot exclude the fact that this alteration arises *de novo* in our patient and is phenotypically correlated with the literature.

TS-CoExp analysis for patient 5 retrieved only *AKT3* gene, as the one more likely to cause ID, among those covered by the deletion (table 2.10).

Table 2.10 - Summary of the alteration in patient 5 and genes involved. In bold are represented the most likely significant gene for the phenotype based on their function.

Patient Number	Event	Number of Genes Involved	Number of Probes Involved	Genes	TS-CoExp
#5	1q43-q44(chr1:241,618,630-241,805,298)X1	2	11	***AKT3***, *SDCCAG8*	*AKT3*

5. 2p13.3-13.1 deletion

Patient #6

This patient is a 9 years old boy without psychometric criteria for classification as ID (IQ= 73) but with severe delayed language acquisition that lead clinicians to classify him as intellectually handicapped. The patient also presents dysmorphisms (triangular face, brachycephaly, small ears with hypoplastic helix), thin lips, hypertrophic gums, pointed chin and short neck. Additionally, he also presents stereotypies, aggressiveness and hyperactivity and attention deficit.

aCGH revealed a 4 Mb *de novo* deletion at chromosome region 2p13.3-p13.1 (chr2:70,748,414-74,840,026) containing 62 genes. Figure 2.8A presents the facial appearance of patient 6 while figure 2.8B shows the alteration detected by aCGH.

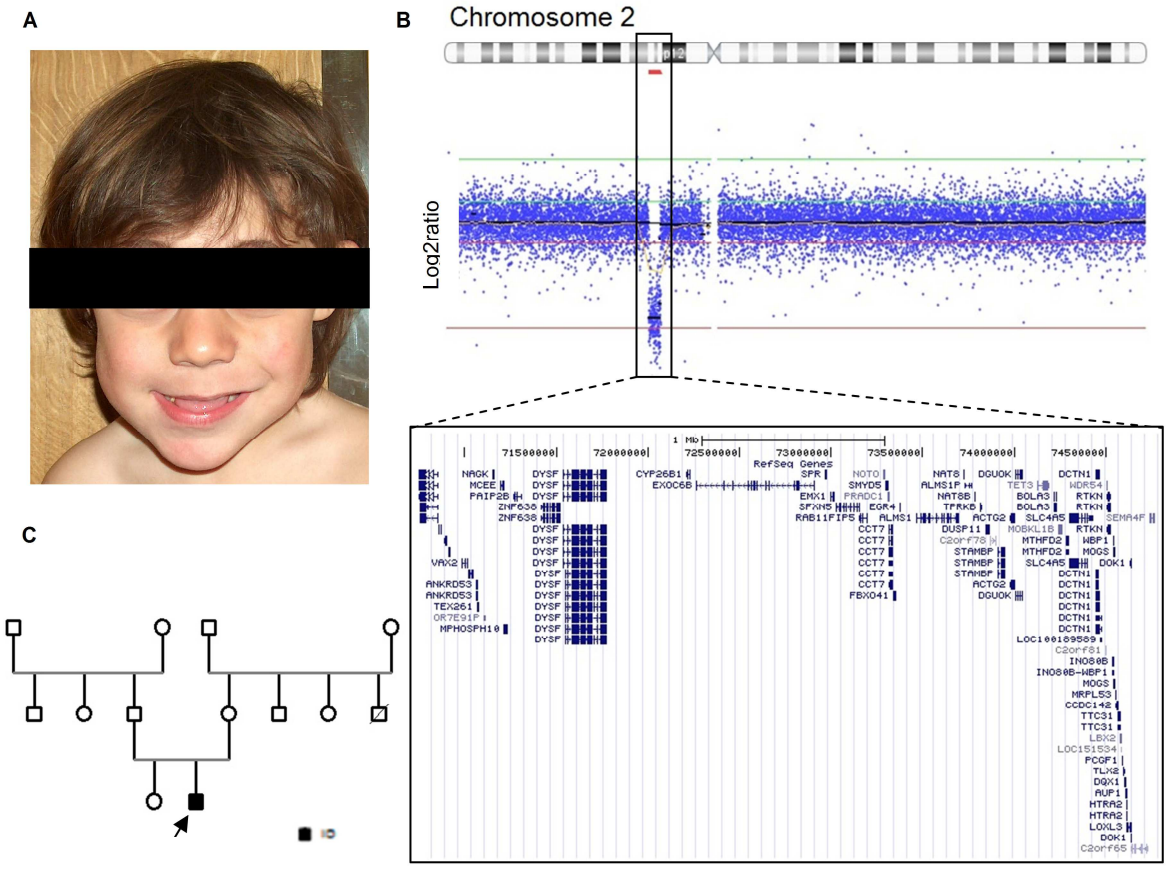


Figure 2.8 - Facial appearance, schematic representation of the genomic alteration and pedigree representation for patient 6. (A) The facial appearance of the patient reveals its triangular face, brachycephaly, small ears with hypoplastic helix, thin lips, hypertrophic gums, pointed chin and short neck. **(B)** The genomic alteration is represented by the red line bellow the 2p13.3-p13.1 cytoband where it is possible to observe the diminished mean ratio of the probes contained in that region. The UCSC genes present in the genomic region altered in the patient in also represented. **(C)** The patient is represented by the arrow. The patient has a deceased maternal that died at 3 months old.

Two our knowledge no deletion similar to the one present in patient 9 has been described so far. This alteration covers almost the total region of cytobands 2p13.1, 2p13.2 and 2p13.3, affecting 62

genes. In the Decipher database there are only 2 patients with alterations partially overlapping with the one present in patient 6. Decipher patient #248972 carries a de novo 12 Mb duplication involving cytobands 2p13.3-p12, while decipher patient #249203 carries a de novo 0.9 Mb deletion on 2p13.2 cytoband. The schematic comparison between the 3 cases is represented in figure 2.9.

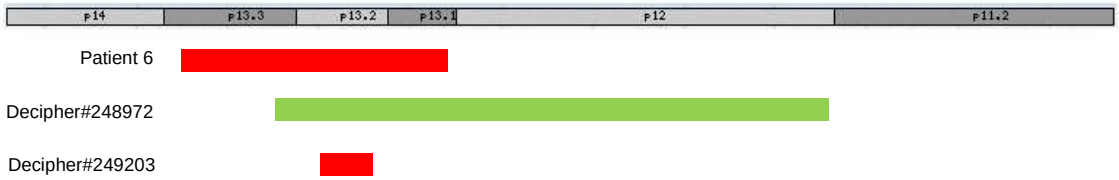


Figure 2.9 - Schematic comparison of the alteration present in patient 6 with the ones described in decipher database. Patient 6 carries a 5.5 Mb deletion (red), Decipher#248972 carries a 12 Mb duplication (green) and Decipher#249203 carries a 0.9 Mb deletion (red).

Although the comparison of the 3 cases is very limited due to the differences in the imbalance sizes, the ID feature is common to all the cases.

In patient 6, one of the genes deleted is the *SEMA4F* (semaphorin 4) gene. Even though no direct association between *SEMA4F* deficiency and ID phenotype has been made, this gene encodes a very important protein for axon guidance (as reviewed in the introduction), making it a good candidate to cause ID [186]. Only the Decipher case #248972 has this gene altered as well.

TS-CoExp analysis for patient 6 retrieved *ADD2* (adducin 2, beta), *SEMA4F* (semaphorin 4F) and *DYSF* (dysferlin, limb girdle muscular dystrophy 2B (autosomal recessive) genes as the 3 more likely to cause ID, among those covered by the deletion.

Add2 is a heterodimeric calmodulin-binding protein of the cell-membrane skeleton that is thought to play a role in assembly of the spectrin-actin lattice that underlies the plasma membrane [187]. A study performed by Gillian *et al*, 1999, demonstrates the important role of Add2 for red blood cells membrane stability [188].

Dysferlin is a transmembrane protein that has is implicated in calcium-dependent membrane fusion events and has an important role in muscle fiber repair, being mutated in patients with myopathies [189].

A summary of the alteration, genes involved and genes pinpointed by the TS-CoExp tool is represented in table 2.11.

Table 2.11 - Summary of the alteration in patient 6 and genes involved. In bold are represented the most likely significant gene for the phenotype based on their function.

Patient Number	Event	Number of Genes Involved	Number of Probes Involved	Genes	TS-CoExp
#6	2p13.3-p13.1(70,748,414- 74,840,026)X1	62	265	***ACTG2***, *ADD2, ALMS1, ALMS1P, ANKRD53, ATP6V1B1, BOLA3, C2orf7, C2orf78, C2orf81, CCT7, CD207, CLEC4F, CYP26B1, DCTN1, DGUOK, DUSP11, DYSF, EGR4, EMX1, EXOC6B, FBXO41, FIGLA, INO80B, LOC100189589, MCEE, MOBKL1B, MPHOSPH10, MTHFD2, NAGK, NAT8, NAT8B, NOTO, OR7E91P, PAIP2B, RAB11FIP5, RTKN, SFXN5, SLC4A5, SMYD5, SPR, STAMBP,* ***TET3***, *TEX261, TPRKB, VAX2, WBP1, WDR54, ZNF638, AUP1, C2orf65, CCDC142, DOK1, DQX1, HTRA2, LBX2, LOC151534, LOXL3, PCGF1,* ***SEMA4F***, *TLX2, TTC31*	*ADD2, SEMA4F, DYSF*

6. 2q11.2-q12.2 deletion

Patient #7

This patient is a 12 years old girl for which no psychometric evaluation could be performed yet. She has ventricular enlargement, dysmorphic features (low set ears, bulbous nose, arched eyebrows, open mouth appearance, dysplasic teeth), hirsutism, aggressiveness and hyperactivity. Her mother is also suspected to have mild ID (picture 2.10C), however this impression has not been confirmed by psychometric assessment.

aCGH revealed a 4.5 Mb deletion at chromosome region 2q11.2-q12.2 (chr2:101,122,697-105,631,450) containing 24 genes. Figure 2.10A presents the facial appearance of patient 7 while figure 2.10B shows the alteration detected by aCGH.

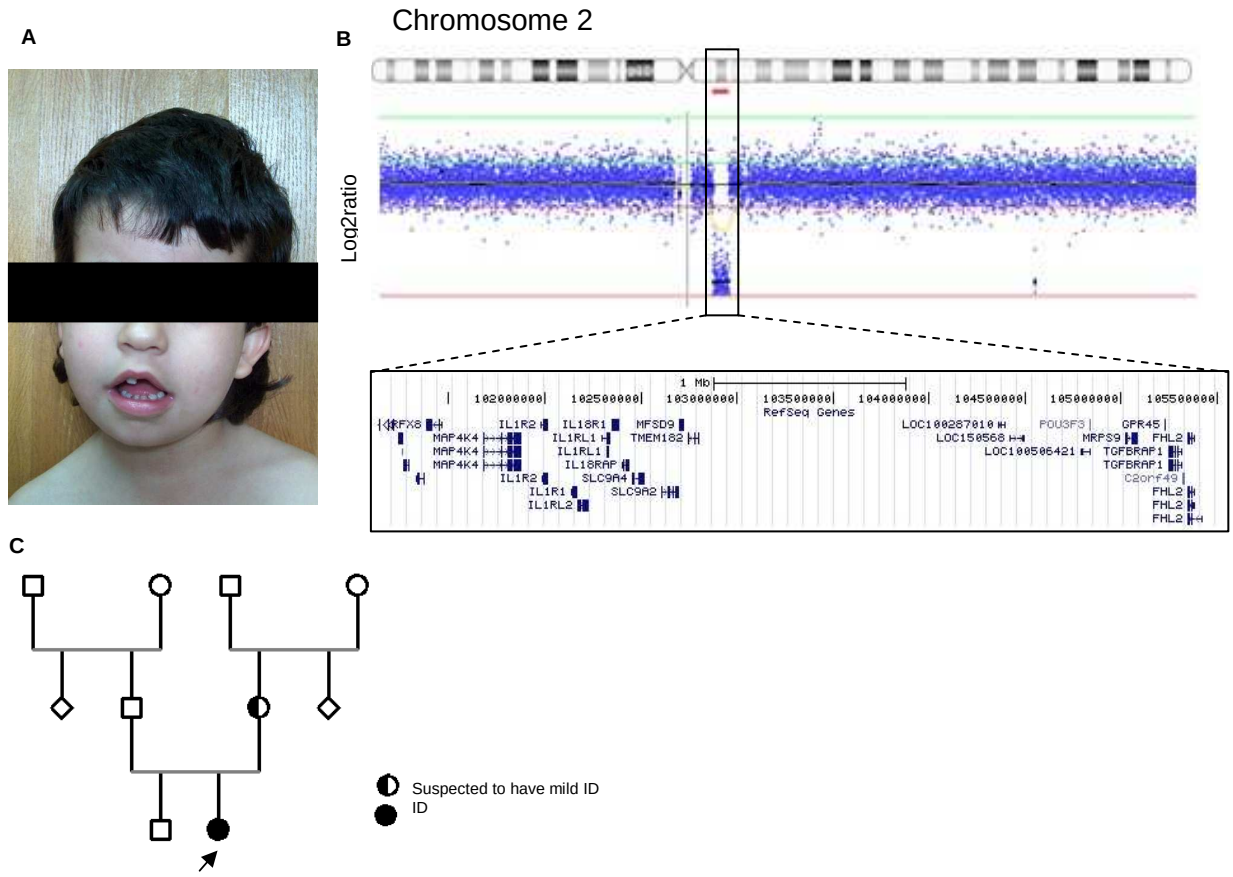


Figure 2.10 - Facial appearance, schematic representation of the genomic alteration and pedigree representation for patient 7. (A) The facial appearance of the patient shows her low set ears, bulbous nose, arched eyebrows, and open mouth appearance. **(B)** The genomic alteration is represented by the red line bellow the 2q11.2-q12.2 cytoband where it is possible to observe the diminished mean ratio of the probes contained in that region. The UCSC genes present in the genomic region altered in the patient in also represented. **(C)** The patient is represented by the arrow; her mother is suspected to have mild ID.

The Decipher patient with the most similar alteration (Decipher #250511) has a 6.6 Mb deletion that partially overlaps with the one found in patient 7. Decipher #250511 displays some overlapping characteristics with our patient such as, ID, open mouth appearance, irregular teeth and facial dysmorphisms.

The fact that the mother of the patient is suspected to have mild ID leads us to hypothesize if the same or a similar alteration is also present in the mother. In order to clarify this aspect aCGH in the mother will be also performed.

The OMIM search for the genes covered by the deletion revealed, among others, three interesting candidates: *GPR45* (G protein-coupled receptor 45), *POU3F3* (POU domain, class 3, transcription factor 3) and *GREG2* (cellular repressor of E1A-stimulated genes).

GPR45 encodes a protein belonging to the G protein-coupled receptors that are integral membrane proteins and mediate signals to the interior of the cell. GPR45 transcript was detected in basal forebrain, frontal cortex and caudate [190].

POU3F3 is a member of the class III POU family of transcription factors that are expressed in the central nervous system and thought to control multiple aspects of the neurogenic program [191].

CREG2 is a member of the cellular repressor of E1A-stimulated genes, and encodes a secreted glycoprotein expressed specifically in the brain [192].

TS-CoExp analysis for patient 7 retrieved *MRPS9* (mitochondrial ribosomal protein S9), *IL1RL1* (interleukin 1 receptor-like 1) and *CREG2* (cellular repressor of E1A-stimulated genes) genes as the 3 more likely to cause ID, among those covered by the deletion.

MRPS9 is a component of the small subunit of the mitochondrial ribosome that is encoded by the nuclear genome [193]. IL1RL1 is a receptor for interleukin 1 and an association between atopic dermatitis and single nucleotide polymorphisms in IL1RL1 promoter [194]. Although referred by TS-CoExp tool as good candidate genes for ID in the patients it is likely that these two genes are false positive detections of the software and not related with the phenotype in the patient.

A summary of the alteration, genes involved and genes pinpointed by the TS-CoExp tool is represented in table 2.12.

Table 2.12 - Summary of the alteration in patient 7 and genes involved. In bold are represented the most likely significant gene for the phenotype based on their function.

Patient Number	Event	Number of Genes Involved	Number of Probes Involved	Genes	TS-CoExp
#7	2q11.2-q12.2 (101,122,697-105,631,450)X1	24	196	*C2orf29, C2orf49,* ***CREG2****, FHL2,* ***GPR45****, IL18R1, IL18RAP, IL1R1, IL1R2, IL1RL1, IL1RL2, LOC150568, MAP4K4, MFSD9, MRPS9,* ***POU3F3****, RFX8, RNF149, SLC9A2, SLC9A4, SNORD89, TBC1D8, TGFBRAP1, TMEM182*	*CREG2, IL1R1, MRPS9*

7. 2q11.2 duplication

Patient #8

This patient is a 9 years old boy with mild ID (IQ=61) and skin pigment alterations (one *café-au-lait* spot). The patient presented no significant facial dysmorphisms but has abnormal skin texture (thin skin). No photograph or genealogic tree is available yet for this patient.

aCGH revealed a 1.5 Mb duplication at chromosome region 2q11.2 (chr2:96,098,910-97,594,697) containing 24 genes. Figure 2.11 shows the alteration detected by aCGH.

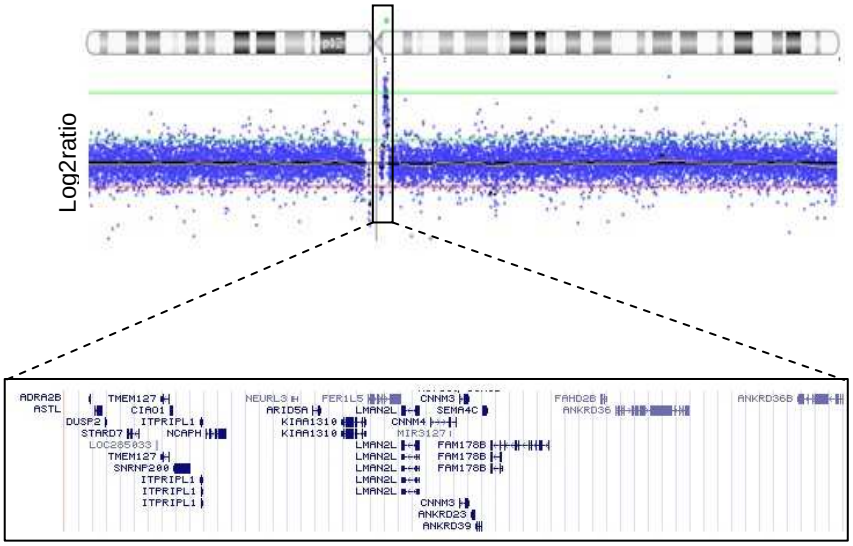


Figure 2.11 - Facial appearance, schematic representation of the genomic alteration and pedigree representation for patient 8. (A) The facial appearance of the patient shows his lack of significant dysmorphisms. **(B)** The genomic alteration is represented by the green line above the 2q11.2 cytoband where it is possible to observe the increased mean ratio of the probes contained in that region. The UCSC genes present in the genomic region altered in the patient in also represented. **(C)** The patient is represented by the arrow.

A similar alteration was reported by Rudd *et al* in 2009. The patient carries a 1.5 duplication partially overlapping with the one from patient 8. The reported patient had moderate ID, short stature, macrocephaly and was hypotonic.

In Decipher there is also a patient described (Decipher #254924) with a smaller (0.9 Mb) duplication affecting partially the 2q11.2 cytoband. The alteration in Decipher #254924 was inherited from a parent with similar phenotype as the child. Additionally to this duplication, the patient also carries a de novo Xp22.31 duplication covering 5 genes. Smaller duplications in controls from the DGV databases are described but only covering the *FAHD2B* (fumarylacetoacetate hydrolase domain containing 2B), *ANKRD36* (ankyrin repeat domain 36) and *ANKRD36B* (ankyrin repeat domain 36 B) genes. No description of the phenotype of the patient Decipher#254924 is present in Decipher database.

Regarding the genes present in the duplicated region of patient 8 we can find *SEMA4C* (semaphorin 4C) which encodes a semaphorin. Semaphorins are known to be transmembrane and secreted proteins that appear to function during growth cone guidance and network establishment in the developing brain (as reviewed in the first chapter).
ADRA2B (alpha 2B adrenergic receptor) gene encodes an adrenergic receptor with high expression in liver and kidney. Polymorphisms in *ADRA2B* gene are thought to be associated with metabolic disorder by altering the autonomic nervous systems normal functions [195].

TS-CoExp analysis for patient 8 retrieved *CNNM3* (cyclin M3), *TMEM127* (transmembrane protein 127) and *LMAN2L* (lectin, mannose-binding 2 like protein) genes as the 3 more likely to cause ID, among those covered by the deletion.
CNNM3 encodes an ancient conserved domain protein that is evolutionarily conserved in diverse species and is ubiquitously expressed in developing and adult tissues [196]. No association with disease was established so far for this gene.
TMEM127 encodes a transmembrane protein that localizes in the Golgi and lysosomes, suggesting that it may be involved in protein trafficking between these structures [197]. Quin *et al*, 2010 suggested that TMEM127 limits mTORC1 activation and that it occupies the same intracellular domain as active mTOR. These authors also concluded that mutations in TMEM127 conferred susceptibility for pheochromocytoma[5] [197].
The LMAN2L protein highly expressed in kidney and skeletal muscle; it is also expressed in brain. This protein is thought to participate in the exportation of glycoproteins from the endoplasmic reticulum [198].

A summary of the alteration, genes involved and genes pinpointed by the TS-CoExp tool is represented in table 2.13.

Table 2.13 - Summary of the alteration in patient 8 and genes involved. In bold are represented the most likely significant gene for the phenotype based on their function.

Patient Number	Event	Number of Genes Involved	Number of Probes Involved	Genes	TS-CoExp
#8	2q11.2 (96,098,910-97,594,697)X3	24	88	*ADRA2B, ANKRD23,* ***ANKRD36****,* ***ANKRD36B****, ANKRD39, ARID5A, ASTL, CIAO1, CNNM3, CNNM4, DUSP2,* ***FAHD2B****, FAM178B, FER1L5, ITPRIPL1, KIAA1310, LMAN2L, LOC285033, NCAPH, NEURL3,* ***SEMA4C****, SNRNP200, STARD7, TMEM127*	*CNNM3, TMEM127, LMAN2L,*

[5] Pheochromocytomas are neuroendocrine tumors of the medulla and adrenal glands of neural crest origin that are frequently hereditary.

8. 3q22.1-q23 deletion (*FOXL2*)

Patient #9

This patient is a 9 years old boy with mild ID (IQ= 54), Von Willebrand disease[6], cortical atrophy, interventricular communication, low weight, hypoplastic genitalia, hypospadias, sandal gap, brachicephaly, arched highbrows, micrognathia, small mouth and thin lips. This patient has also a peculiar eye dysmorphology characterized by microphtalmia, epicanthus and ptosis. Additionally, the patient also displays a hyperactivity disorder.

aCGH revealed a 10.2 Mb *de novo* deletion at chromosome region 3q22.1-q23 (chr3:132,898,329-143,101,242) containing 65 genes. Figure 2.12A presents the facial appearance of patient 9 while figure 2.12B shows the alteration detected by aCGH.

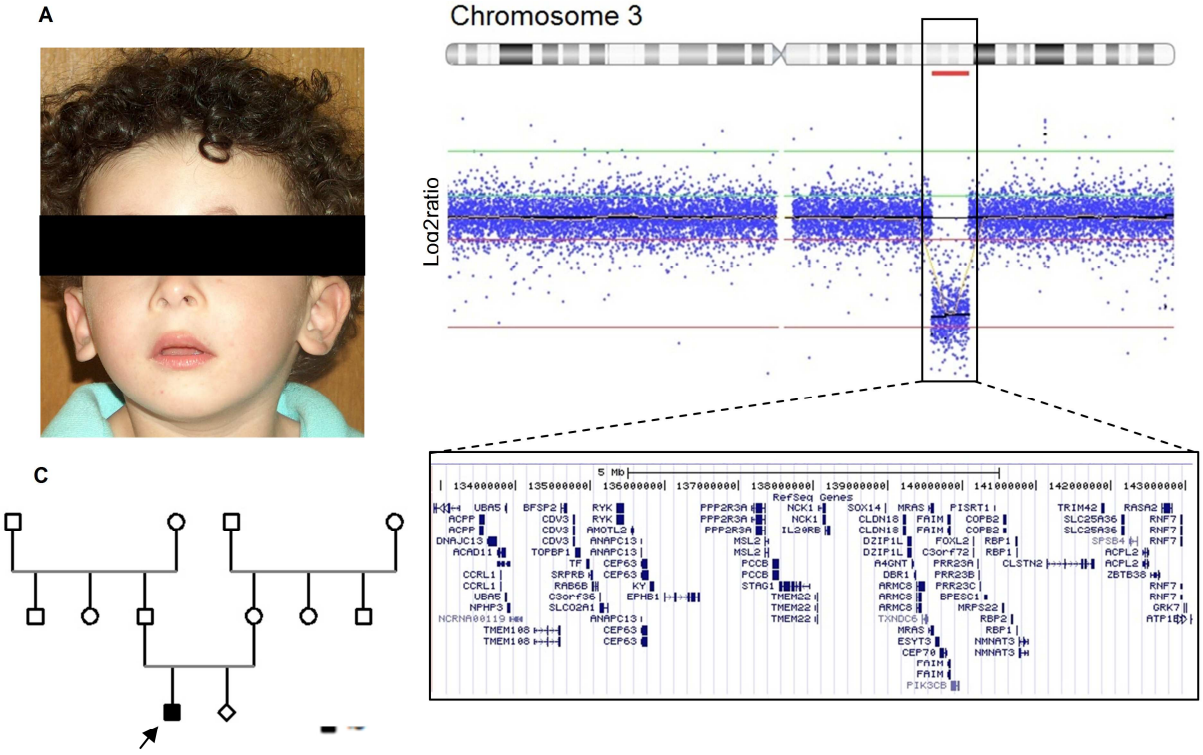


Figure 2.12 - Facial appearance, schematic representation of the genomic alteration and pedigree representation for patient 9. (A) The facial appearance of the patient shows his brachicephaly, arched highbrows, micrognathia, small mouth and thin lips. **(B)** The genomic alteration is represented by the red line bellow the3q22.1-q23 cytoband where it is possible to observe the diminished mean ratio of the probes contained in that region. The UCSC genes present in the genomic region altered in the patient in also represented. **(C)** The patient is represented by the arrow and no familial history of ID is described in the family.

In 2008, Croft *et al* described a 15 years old female with a 3q22.1-q23 deletion, blepharophimosis, ptosis and epicanthus inversus (BPES), microcephaly and ID [199]. The alteration covered several

[6] Von Willebrand disease is a bleeding disorder caused by a defect or deficiency of a blood clotting protein, called von Willebrand Factor.

genes, among them the *FOXL2* (helix/forkhead transcription factor) gene, for which haploinsufficiency has been confirmed as causative of the BPES phenotype [200].

Patient 9 displays similar phenotypic characteristics with the ones described in the literature: BPES, short stature, facial dysmorphisms, ID/DD. We believe that the BPES phenotype in the patient is due to the haploinsufficiency of *FOXL2* gene. Since point mutations *FOXL2* have been described in BPES patients without any intellectual limitation or development abnormality, it is currently stipulated that *FOXL2* haploinsufficiency is only responsible for the eye phenotype in the patients carrying large genomic rearrangements. The causative gene for the ID/DD phenotype was never fully investigated in those cases.

EPHB1 (ephrin receptor EphB1), *RAB6B* (RAB6B, member RAS oncogene family) and *UBA5* (ubiquitin-activating enzyme E1 domain-containing 1) are genes covered by the deletion and that constitute good candidates for the ID phenotype in the patient.

EPHB1 protein is an ephrin receptor highly expressed in brain regions were the facial and inner ear efferent neurons are localized [201, 202]. Rab6b belongs to the subfamily of small GTPases Rab6 and is predominantly expressed in brain. The actual function of this protein still remains unclear. However, Wanschers BF in 2007 proposed that Rab6b interacts with dynein/dynactin complex playing a role in the retrograde transport of cargo in neurons [203]. UBA5 is an E1-like activating protein that activates ubiquitin-fold modifier 1 (Umf1) and SUMO2 [204, 205]. Alterations in this pathway may interfere with the correct degradation of other proteins (substrates), leading to a variety of diseases (such as ID).

TS-CoExp analysis for patient 9 retrieved *COPB2* (coatomer protein complex, subunit beta 2, beta prime), *MRPS22* (mitochondrial ribosomal protein S22) and *RAB6B* (RAB6B, member RAS oncogene family) genes as the 3 more likely to cause ID, among those covered by the deletion.

COPB2 encodes a protein that belongs to the Golgi coatomer protein complex, a set of proteins that participate in protein transport between the endoplasmic reticulum and Gogi complex. No association with COPB2 and the disease has been made so far [206].

The *MRPS22* gene was identified in the BPES syndrome critical region. Mutations in this gene were recently associated with Cornelia de Lange-like phenotype, brain abnormalities and hypertrophic cardiomyopathy revealing the importance of mitochondrial synthesized proteins [207].

RAB6B gene encodes a small GTPase that is predominantly expressed in the brain and thought to regulate protein transport in neuronal cells by interaction with the dynein/dynactin complex [208].

A summary of the alteration, genes involved and genes pinpointed by the TS-CoExp tool is represented in table 2.14.

Table 2.14 - Summary of the alteration in patient 9 and genes involved. In bold are represented the most likely significant gene for the phenotype based on their function.

Patient Number	Event	Number of Genes Involved	Number of Probes Involved	Genes	TS-CoExp
#9	3q22.1-q23(132,898,329-143,101,242)X1	65	622	*CPNE4, ACAD11, ACPP, AMOTL2, ANAPC13, BFSP2, C3orf36, CCRL1, CDV3, CEP63, CPNE4, DNAJC13,* ***EPHB1****, KY, NCRNA00119, NPHP3,* ***RAB6B****, RYK, SLCO2A1, SRPRB, TF, TMEM108, TOPBP1,* ***UBA5****, A4GNT, ACPL2, ARMC8, ATP1B3, BPESC1, C3orf72, CEP70, CLDN18, CLSTN2, COPB2, DBR1, DZIP1L, ESYT3, FAIM,* ***FOXL2****, GRK7, IL20RB, MRAS, MRPS22, MSL2, NCK1, NMNAT3, PCCB, PIK3CB, PISRT1, PPP2R3A, PRR23A, PRR23B, PRR23C, RASA2, RBP1, RBP2, RNF7, SLC25A36, SOX14, SPSB4, STAG1, TMEM22, TRIM42, TXNDC6, ZBTB38*	*RAB6B, COPB2, MRPS22*

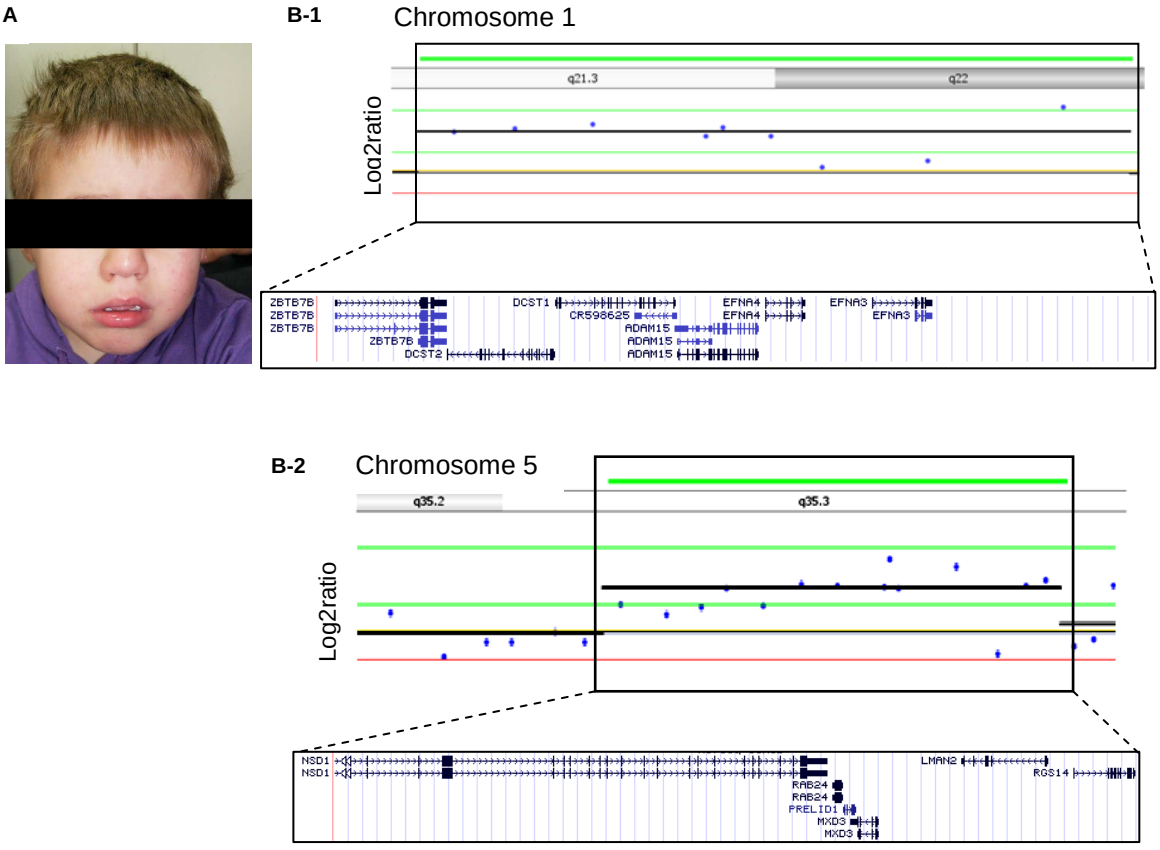
9. 5q35.3 duplication

Patient #10

This patient is a 6 years old boy with moderate ID (IQ=48), with familial history of ID (paternal grandmother suspected to have mild ID although no psychometric assessment was performed) who at birth presented weight, stature and head circumference above 98th centile. The patient also has liquor space enlargement, epilepsy (controlled with medication) and hypermetropia. He also presents dysmorphic face (epicanthus, bulbous nose, and harsh hair), generalized hypopigmentation of the skin and hair, weight above the 98th centile, aggressiveness, anxious behavior and hyperactivity and attention deficit disorder. Additionally he displays motor stereotypies (shaking hands when excited).

aCGH revealed a 1q21.3-q22 *de novo* duplication (chr1:153,239,357-153,361,162), containing 6 genes, a *de novo* 0.2 Mb duplication at chromosome region 5q35.3 (chr5:176,544,003-176,734,904), containing 6 genes and a 19p13.2 *de novo* duplication (chr19:9,468,575-9,748,852), containing 6 genes.

Figure 2.13A presents the facial appearance of patient 10 while figure 2.13B 1-3 shows the alteration detected by aCGH.



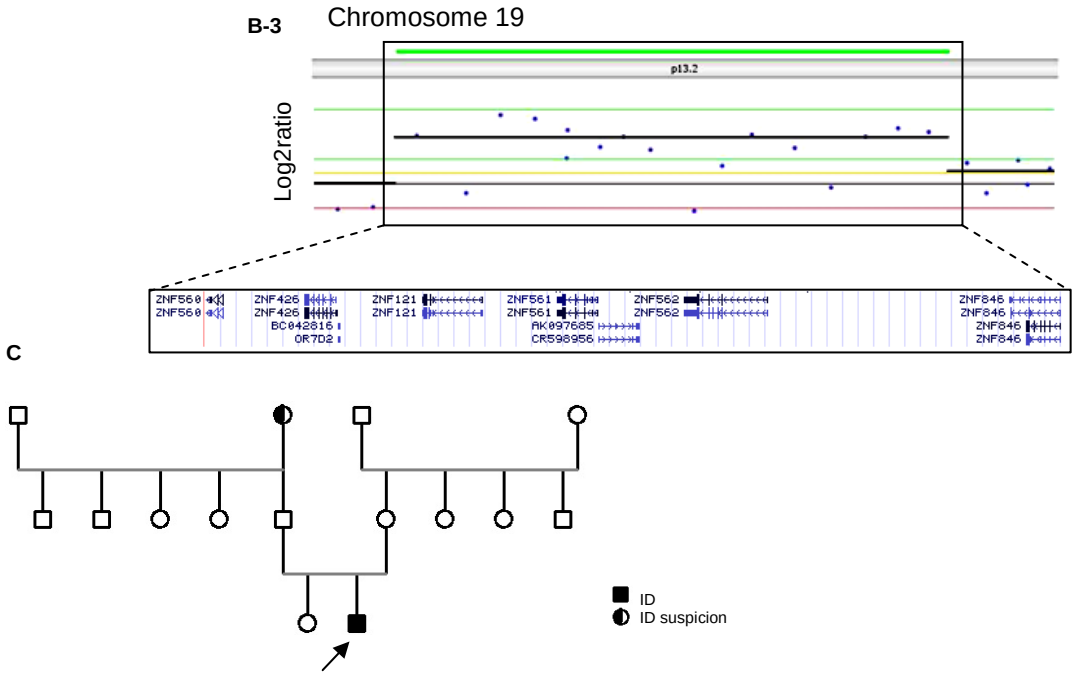


Figure 2.13 - Facial appearance, schematic representation of the genomic alteration and pedigree representation for patient 10. (A) The facial appearance of the patient shows his dysmorphic face (epicanthus, bulbous nose, and bitter hair) and generalized hypopigmentation of the skin and hair. **(B)** The genomic alteration is represented by the green line above the 1q21.3-22 cytoband **(B-1),** above the 5q35.3 cytoband **(B-2)** and above the 19p13.2 cytoband **(B-3)** where it is possible to observe the increased mean ratio of the probes contained in those regions. The UCSC genes present in the genomic region altered in the patient in also represented. **(C)** The patient is represented by the arrow. His mother is suspected to have a milder form of ID, although not formally accessed.

The 1q21.3-q22 duplication covers, among others, the *ADAM15* (disintegrin and metalloproteinase domain 15), *EFNA3* (ephrin A3) and *EFNA4* (ephrin A4) genes.

ADAM15 encodes a transmembrane glycoprotein that is important for cell adhesion and proteolytic shedding of cell surface receptors [209]. Few functional studies have been performed so far. In mouse, Adam15 is highly expressed in vascular cells and in hippocampus and cerebellum. In the same mouse model, the null mice showed no major development defects and reduced neovascularization[210].

EphA3 protein is the ligand for EphA4 and, in mice, is localized in dendritic spines of pyramidal neurons in the adult hippocampus. EphrA4 and EphnA3 communication in neuroglial cells regulates the structure of synaptic connections (activation of EphA4 by ephrin-A3 was found to induce spinal retraction and reduce spine density) [211].

The 5q35.3 duplication, if located in tandem, may disrupt the *NSD1* (nuclear receptor-binding Su-var, enhancer of zeste and trithorax domain protein 1) gene. NSD1protein is a coregulator protein of the steroid receptor superfamily [212]. Haploinsufficiency of NSD1 protein is the major cause of Sotos syndrome, also called cerebral gigantism, which is a disorder characterized by excessively

rapid growth, acromegalic features, and a nonprogressive cerebral disorder with ID. High-arched palate and prominent jaw are also common features among the patients. At birth, length is usually between the 90th and 97th centiles in all the patients and bone age is also advanced for the majority of the cases [213]. Nagai et al, in 2003 reviewed the phenotypes for two groups of patients with Sotos syndrome: patients carrying point mutations in NSD1 and patients with a deletion involving the whole *NSD1* gene. The results suggested that the clinical features of Sotos syndrome patients can de divided in two major categories: (I) overgrowth, advanced maturation in infancy to early childhood, ID, hypotonia, hyperreflexia and minor anomalies were present in patients with point mutations in the gene, while (II) major anomalies in the central nervous system (such as agenesis or hypoplasia of the corpus callosum), cardiovascular system defects and urinary system abnormalities were exhibited by patients with deletions [214].
Additionally to NSD1, *PRELID1* (preli domain-containing protein 1) and *RAB24* (RAs-associated protein 24) genes are also covered by the duplication and may be contributing for the phenotype of the patient. *PRELID1* encodes a protein (PRELI) that in chicken is expressed in late embryogenesis and that is thought to contribute for hemopoiesis. In human, PRELI is known to be a mitochondrial regulator of human primary T-helper cell apoptosis and Th-2 cell differentiation [215]. RAB24 is a small GTPase of the Rab subfamily of Ras-related proteins that regulate intracellular protein trafficking, highly expressed in the brain [216].

The 19p13.2 duplication covers 6 genes, for which only ZNF121 (zinc finger protein 121) has entrance in OMIM. ZNF121 was isolated from a brain cDNA library and localizes to a region in chromosome 19 where other zinc finger sequences and multiple cosmid clones containing zinc fingers had previously been localized [217].

TS-CoExp analysis for patient 10 retrieved *ADAM15*, *ZNF562* (zinc finger protein 562) and *ZNF561* (zinc finger protein 561) genes as the 3 more likely to cause ID, among those covered by the alterations.
The *ADAM15* gene was described before (see page 88) and the *ZNF562* and *ZNF561* genes don´t have an entrance in OMIM and PubMed databases.

A summary of the alteration, genes involved and genes pinpointed by the TS-CoExp tool is represented in table 2.15.

Table 2.15 - Summary of the alteration in patient 10 and genes involved. In bold are represented the most likely significant gene for the phenotype based on their function.

Patient Number	Event	Number of Genes Involved	Number of Probes Involved	Genes	TS-CoExp
#10	1q21.3-q22 (153,239,357-153,361,162)X3	6	9	***ADAM15***, *DCST1*, *DCST2*, ***EFNA3***, ***EFNA4***, *ZBTB7B*	*ADAM15, ZNF562, ZNF561*
	5q35.3 (176,544,003-176,734,904)X3	6	14	*LMAN2*, *MXD3*, ***NSD1***, ***PRELID1***, *RAB24*, *RGS14*	
	19p13.2 (9,468,575-9,748,852) X3	6	17	***ZNF121***, *ZNF426*, *ZNF560*, *ZNF561*, *ZNF562*, *ZNF846*	

5q35.3 duplication

Patient #11

This patient is a 9 years old boy with mild ID (IQ=58) and without familial history of ID. The patient also has myopia, squint, and right ptosis, inguinal herniorrhaphy, liquor space enlargement. He also has a Sotos-like facial appearance, hypoplasic genitals and is hypotonic.

aCGH revealed a 0.9 Mb duplication at chromosome region 5q35.3 (chr5:179,345,860-180,237,221) containing 12 genes. Figure 2.14A presents the facial appearance of patient 11 while figure 2.14B shows the alteration detected by aCGH.

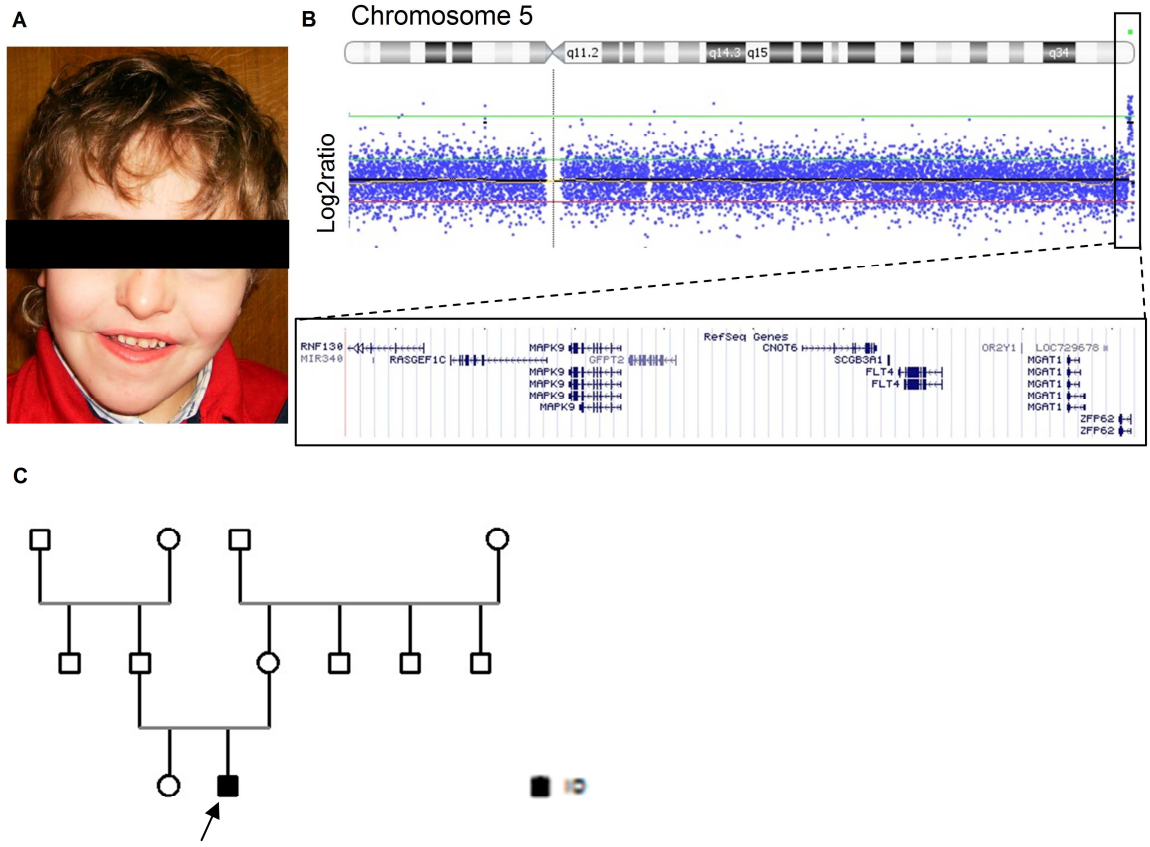


Figure 2.14 - Facial appearance, schematic representation of the genomic alteration and pedigree representation for patient 11. (A) The facial appearance of the patient shows his Sotos-like facial appearance, squint and right ptosis. **(B)** The genomic alteration is represented by the green line above 5q35.3 cytoband where it is possible to observe the increased mean ratio of the probes contained in that region. The UCSC genes present in the genomic region altered in the patient in also represented. **(C)** The patient is represented by the arrow.

This patient presents a Sotos-like facial appearance even though the alteration is not inside the Sotos syndrome critical region. There are no patients with similar alterations described in Decipher and a PubMed search only retrieved one case of 19p13.2 microduplication in a three-generation family segregating with a Sotos-like disorder [218].

Among the affected genes, *CNOT6* (CCR4-not transcription complex, subunit 6) and *MAPK9* (mitogen-activated protein kinase 9) genes may constitute good candidates for the ID phenotype of the patient.

CNOT6 is a general transcriptional regulatory complex involved in several aspects of mRNA metabolism, including transcription initiation and elongation and mRNA degradation, with high expression levels in brain [219].

MAPK9 is a mitogen-activated protein kinase that plays an important role in during brain development through its essential role in regulation of region-specific apoptosis during brain development [220].

Both genes described above are good candidates for the phenotype displayed by the patient.

TS-CoExp analysis for patient 11 retrieved *RNF130* (ring finger protein 130) and *MAPK9* genes as the 2 more likely to cause ID, among those covered by the duplication.

RNF130 encodes a protein that has a protease-associated, a transmembrane and a RING domains. RNF130 protein is the human homolog of Drosophila´s g1, a zinc-finger protein involved in apoptosis in embryonic development [221].

A summary of the alteration, genes involved and genes pinpointed by the TS-CoExp tool is represented in table 2.16.

Table 2.16 - Summary of the alteration in patient 11 and genes involved. In bold are represented the most likely significant gene for the phenotype based on their function.

Patient Number	Event	Number of Genes Involved	Number of Probes Involved	Genes	TS-CoExp
#11	5q35.3(179,345,860-180,237,221)X3	12	57	*CNOT6, FLT4, GFPT2, LOC729678, MAPK9, MGAT1, MIR340, OR2Y1, RASGEF1C, RNF130, SCGB3A1, ZFP62*	*RNF130, MAPK9*

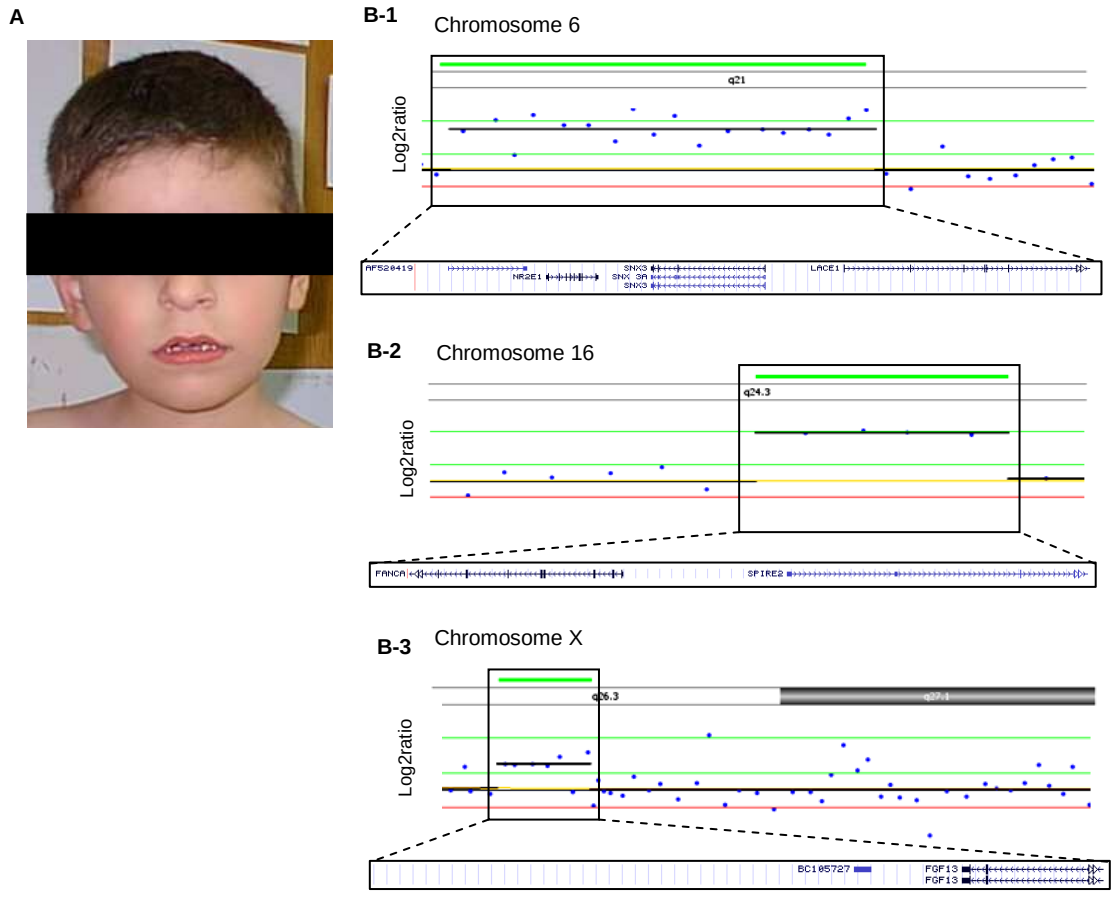
10. 6q21 duplication, 16q24.3 duplication (*FANCA* gene) and Xq26.3 duplication

Patient #12

This patient is a 13 years old boy with severe ID (IQ= 27) with familial history of epilepsy (father), speech delay and learning difficulties (maternal cousins) and congenital anomalies (2nd grade paternal cousin born with meromelia [absence of both feet]). The mother was suspected of toxoplasmosis infection during 2nd-3rd trimester. The patient has thin corpus callosum, hyperintensive signal in the substantia nigra, ventricular asymmetry, repetitive ear infections, squint and constipation. The patient also presents stature and weight above the 98th centile, facial dysmorphisms (thin and low set eyebrows, bulbous, pending and right deviated nose, cupid bow upper lip, wrong dental implantation, and anteriorly located anus. Additionally, the patient also presents recurrent kneecap lesion and increased activity.

aCGH revealed a 0.3 Mb maternal duplication at chromosome region 6q21 (chr6:108,537,896-108,829,534) containing 3 genes, a 0.05 Mb maternal duplication at chromosome 16q24.3 (chr16:88,395,085-88,444,115) containing 2 genes and a 0.1 Mb de novo duplication at chromosome Xq26.3 (chrX:137,447,004-137,563,727) containing 2 genes.

Figure 2.15A presents the facial appearance of patient 12 while figure 2.15B 1-3 shows the alterations detected by aCGH.



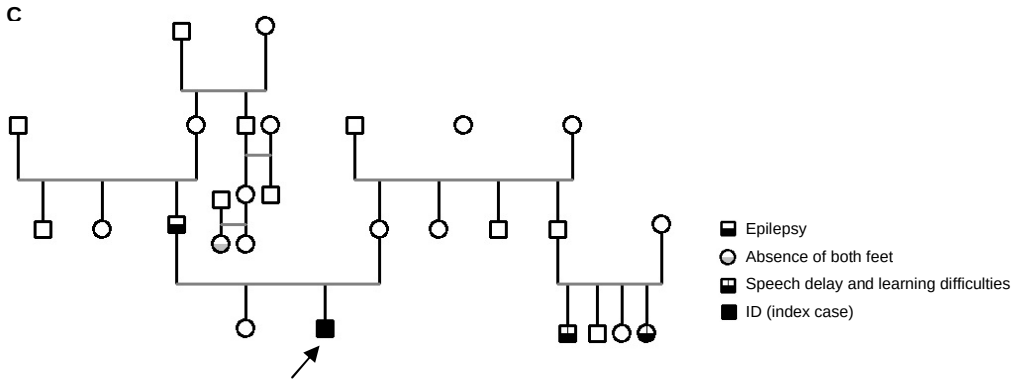


Figure 2.15 - Facial appearance, schematic representation of the genomic alteration and pedigree representation for patient 12. (A) The facial appearance of the patient shows his thin and low set eyebrows, bulbous, pending and right deviated nose and cupid bow upper lip. **(B)** The genomic alteration is represented by the green line above 6q21 cytoband **(B-1)**, 16q24.3 cytoband **(B-2)** and Xq26.3 cytoband **(B-3)** where it is possible to observe the increased mean ratio of the probes contained in those regions. The UCSC genes present in the genomic region altered in the patient in also represented for each case. **(C)** The patient is represented by the arrow; he has two cousins with learning difficulties and a 3rd grade cousin with meromelia. His father also displays epilepsy.

The 6q21 duplication was inherited from the mother and covers the *NR2E1* (nuclear receptor subfamily 2, group E, member 1) gene. The *NR2E1* drosophila homologous gene (*TLX*) is highly expressed in brain and is required for its development [222]. Shi *et al* in 2004 demonstrated that TLX was able to maintain adult neural stem cells in an undifferentiated state thought its recruitment of a set of histone deacetylase that target gene sot repress their expression [223]. *TLX* null mice are viable at birth but display reduced size of the olfactory, infrarhinal and entorhinal cortex, amygdala and dentate gyrus structures in the brain. Also, males tended to be more aggressive and females lacked maternal instincts [224]. 6q16-q22 has also been described as a locus for occipital polymicrogyria [225].

The 16q24.3 duplication, if located in tandem, can disrupt the *FANCA* (Fanconi anemia[7] complementation group A) gene. Although patient 12 doesn't have Fanconi anemia, this disorder is recessive and some of the patients also show ID and congenital malformations. The *FANCA* gene has a transcript that is exclusively expressed in brain where it is expressed in a stage-specific manner during embryogenesis [226]. Another important aspect of this finding is the fact that both mother and son may be carriers for a recessive mutation for Faconi anemia (is the gene is disrupted). In order to clarify this aspect and provide better counseling to the parents, FISH analysis need to be performed in order to determine if the duplication is in tandem or not.

[7] Fanconi anemia is an autosomal recessive disorder affecting all bone marrow elements and associated with cardiac, renal, and limb malformations as well as dermal pigmentary changes.

The Xq26.3 duplication occurs *de novo* in the patient and covers the *FGF13* (fibroblast growth factor 13) gene. FGF13 is expressed in the developing and adult central nervous system (mainly in the cerebellum and cortex) and has a mitogenic and cell survival function [227].

TS-CoExp analysis for patient 12 retrieved *SNX3* (sorting nexin 3), *FGF13* (fibroblast growth factor 13) and *SPIRE2* (spire, drosophila homolog of, 2) genes as the 3 more likely to cause ID, among those covered by the alterations.

SNX3 encodes a protein member of the sorting nexin family of molecules that acts in intracellular trafficking of proteins to various organelles [228]. A translocation between chromosome 6 and 13 that disrupted the *SNX3* gene was described in a patient with microcephaly, microphtalmia, ectrodactyly and prognathism [229].

SPIRE2 encodes a protein that belongs to the Spir protein family, a class of WH2 domain. These proteins initiate actin polymerization and, in a mice model, Spire-2 was mainly expressed in developing and adult neuronal cells [230].

A summary of the alteration, genes involved and genes pinpointed by the TS-CoExp tool is represented in table 2.17.

Table 2.17 - Summary of the alteration in patient 12 and genes involved. In bold are represented the most likely significant gene for the phenotype based on their function.

Patient Number	Event	Number of Genes Involved	Number of Probes Involved	Genes	TS-CoExp
#12	6q21 (108,537,896-108,829,534)X3	3	18	*LACE1*, ***NR2E1***, *SNX3*	*SNX3, FGF13, SPIRE2*
	16q24.3 (88,395,085-88,444,115)X3	2	4	***FANCA***, *SPIRE2*	
	Xq26.3 (137,447,004-137,563,727)X3	2	7	***FGF13***, *LOC158696*	

11. **6q25.3 deletion**

Patient #13

This patient is a 7 years old boy with mild ID (IQ= 56) with familial history of ID/DD (cousin and second - grade cousins), who suffered sepsis after birth, has an enlarged liver and spleen (without criteria for organomegaly), hirsutism, dysmorphic face (bulbous nose, large mouth, thin upper lip), squint, short stature, severe delay in language acquisition, hyperactivity, aggressiveness and stereotypies (shaking hands when excited).

aCGH revealed a 2.6 Mb *de novo* deletion at chromosome region 6q25.3 (chr6:156,054,446-158,724,482) containing 7 genes. Figure 2.16A presents the facial appearance of patient 13 while figure 2.16B shows the alteration detected by aCGH

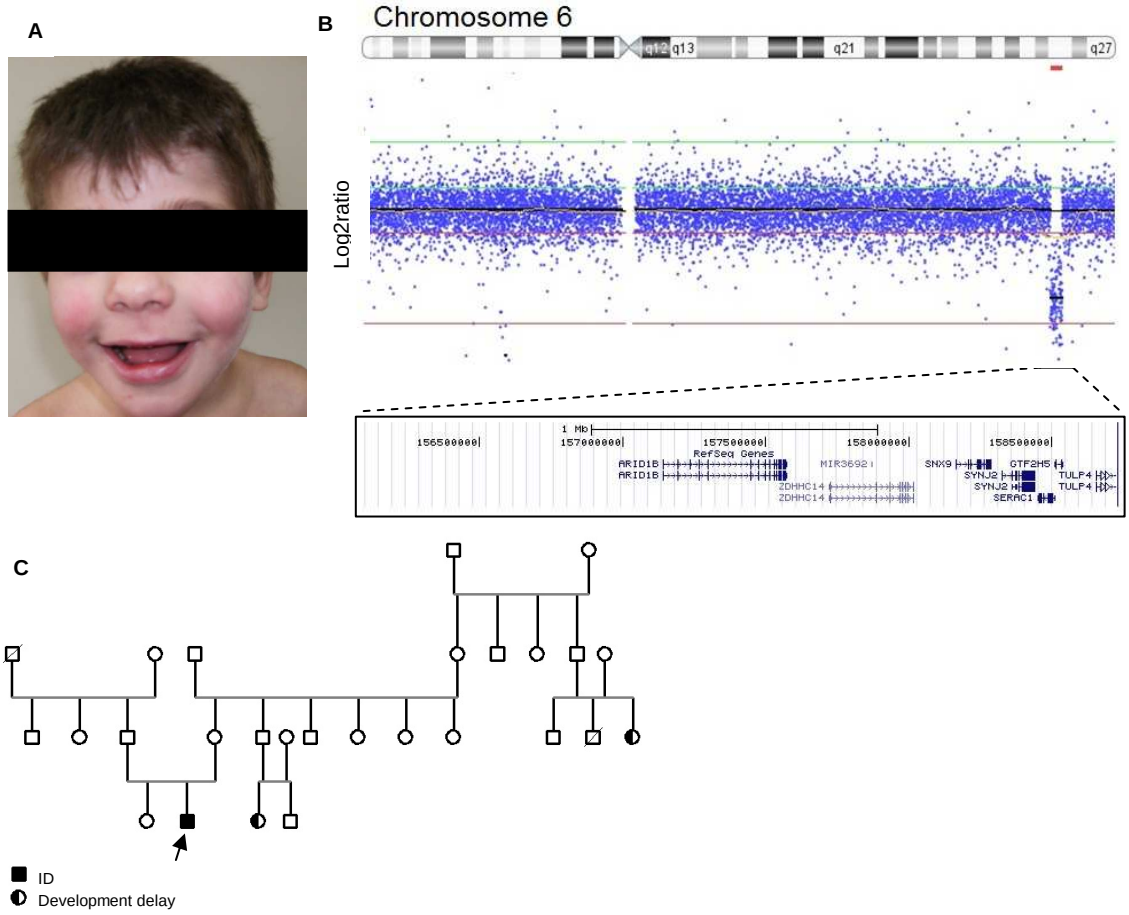


Figure 2.16 - Facial appearance, schematic representation of the genomic alteration and pedigree representation for patient 13. (A) The facial appearance of the patient shows his bulbous nose, large mouth, thin upper lip and squint. **(B)** The genomic alteration is represented by the red line bellow 6q25.3 cytoband where it is possible to observe the diminished mean ratio of the probes contained in that region. The UCSC genes present in the genomic region altered in the patient are also represented. **(C)** The patient is represented by the arrow; he has a cousin and a second-grade cousin with development delay.

Patient 13 aCGH revealed a 2.6 Mb deletion encompassing a critical region associated with sacro/anorectal malformations [231]. A publication by Nagamani S. *et al*, 2009 described partially overlapping deletions (2q25.2-q25.3) associated with DD, hearing loss, microcephaly and dysmorphic features [232]. Search in the Decipher database revealed the existence of a patient (Decipher#248472) with a deletion very similar to the one of our patient. Similar to our patient, Decipher#248472 presented ID and facial dysmorphic features. However, Decipher#248472 had an atrial septum defect while our patient doesn´t seem to have cardiac problems.

The genes pinpointed as crucial for the phenotype in these patients that are also deleted in patient 13 are *SYNJ2* (synaptojanin 2), *SERAC1* (serine active site containing 1), *GTF2H5* (general transcriptor factor IIH, polypeptide 5) and *TULP4* (tubby-like protein 4).

SYNJ2 encodes a protein shown to be involved in the regulation of invadopodia and lamellipodia in a glioblastoma cell line [233]. *SERAC1* mutations were shown to cause sterility in mice. Mutations in GTF2H5 subunits cause trichothiodystrophy syndrome[8]. We found that, although not described before, this patient also shows brittle hair. This feature may be due to the deficiency of GTF2H5 protein [234].

TULP4 is a putative transcriptor factor of unknown function so far. However, in Drosophila it was shown that the TULP protein is mostly expressed in the embryonic central and peripheral nervous systems [235].

TS-CoExp analysis for patient 13 retrieved *GTF2H5* (general transcriptor factor IIH, polypeptide 5), *SYNJ2* (synaptojanin 2) and *SNX9* (shorting nexin 9) genes as those with the 3 more likely to cause ID, among those covered by the deletion.

SNX9 contains a phospholipid-binding module called the phox homology and plays a role in endocytosis through physically binding F-actin nucleation to plasma membrane during endocytosis [236].

A summary of the alteration, genes involved and genes pinpointed by then TS-CoExp tool is represented on table 2.18.

[8] Trichothiodystrophy syndrome is caused by mutations in 2 subunits of GTF2H5 gene. Ppatients with trichothiodystrophy have brittle hair and nails (because of reduced content of cysteine-rich matrix proteins), ichthyotic skin, and physical and mental retardation. Approximately half of the patients display photosensitivity, correlated with a nucleotide excision repair defect.

Table 2.18 - Summary of the alteration in patient 13 and genes involved. In bold are represented the most likely significant gene for the phenotype based on their function.

Patient Number	Event	Number of Genes Involved	Number of Probes Involved	Genes	TS-CoExp
#13	6q25.3(156,054,446 -158,724,482)X1	7	140	*ARID1B*, ***GTF2H5***, ***SERAC1***, *SNX9*, ***SYNJ2***, ***TULP4***, *ZDHHC14*	*GTF2H5, SYNJ2, SNX9*

12. 7q11.23 deletion (Williams-Beuren syndrome region)

Patient #14

This patient is a 13 years old boy with severe ID (IQ=21) and possible familial history of ID (the mother is suspected to have mild ID, although no psychometric assessment was performed). The patient also displays cerebral atrophy, microcephaly, recurrent respiratory infections, epilepsy, plagiocephaly, hypoplastic teeth, short neck, hirsutism, genital abnormalities (cryptorchid) and facial dysmorphisms. Additionally the patient also presents motor stereotypies (hand clapping, hand mouthing and shifting the weight from one leg to the other), as well increased activity, aggressiveness and sleep disturbances.

aCGH revealed a 0.4 Mb deletion at chromosome region 7q11.23 (chr7:72,379,797-72,783,852) containing 11 genes. Figure 2.17A presents the facial appearance of patient 14 while figure 2.17B shows the alteration detected by aCGH.

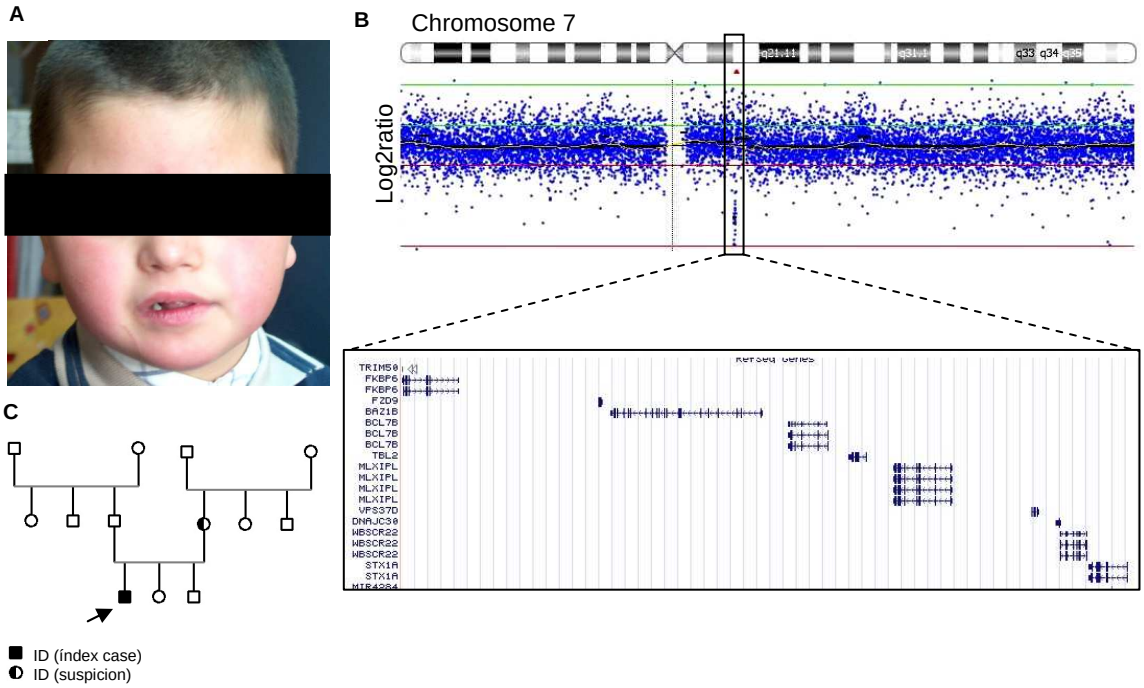


Figure 2.17 - Facial appearance, schematic representation of the genomic alteration and pedigree representation for patient 14. (A) The facial appearance of the patient shows his facial dysmorphisms, plagiocephaly and short neck. **(B)** The genomic alteration is represented by the red line bellow 7q11.23 cytoband where it is possible to observe the diminished mean ratio of the probes contained in that region. The UCSC genes present in the genomic region altered in the patient in also represented. **(C)** The patient is represented by the arrow; his mother is suspected to have ID although not psychometric assessment was performed.

This alteration is located inside the Williams-Beuren syndrome (WBS) critical region, a microdeletion syndrome characterized by mild to moderate ID, weakness in visuospatial construction with relatively high performance in verbal short term memory and language. The

patient also presents characteristic facial dysmorphisms (bulbous nose, wide mouth, full lips and cheeks, small widely spaced teeth and periorbital fullness), anxious behavior, attention deficit and hyperactive disorder, overfriendliness and congenital heart disease (usually supravalvular aortic stenosis). This feature is often associated to the haploinsufficiency of the *ELN* (elastin) gene [237].

Search in Decipher revealed no patients with deletion of a similar size to the one found in patient 14.

Two interesting genes affected by the deletion are *STX1A* (syntaxin 1A) and *MLXIPL* (MLX-interacting protein-like). *STX1A* encodes a protein that acts in the vesicle fusion process and plays a central role in neurotransmitter release through multiple protein-protein interactions [238]. Syntaxin 1a is thought to give rise to some behavioral phenotypes of Williams-Beuren syndrome such as hyperactivity and attention deficit disorder [239].

MLXIPL encodes a protein expressed in multiples tissues, including regions of the brain and the intestinal tract. This protein belongs to the family of helix-loop-helix leucine-zipper transcription factors. It currently supposed that MLXIPL is a key element of a transcription factor network and plays a role in cell proliferation and/or differentiation. It was suggested that haploinsufficiency of this gene may contribute to some features of Williams-Beuren syndrome, however the precise correlation between MLXIPL and a specific feature was never described [240].

The elastin gene is not covered by the deletion present in our patient, which is concordant with the fact that he has no apparent cardiac abnormalities.

It is intriguing that, although patient 14 has a smaller deletion than is common in WBS patients, he seems to display, except for the cardiac problems, a more severe ID phenotype than WBS classical patients.

TS-CoExp analysis for patient 14 retrieved *DNAJC30* (DnaJ (Hsp40) homolog, subfamily C, member 30), *STX1A* (syntaxin 1A) and *BAZ1B* (bromodomain adjacent to zinc finger domain, 1B) genes as the 3 more likely to cause ID, among those covered by the deletion.

DNAJC30 has no relevant information regarding its function in OMIM and PubMed databases.

The BAZ1B protein contains a bromodomain motif which is characteristic of proteins involved in chromatin-dependent transcription regulation. *BAZ1B* is thought to be the Williams syndrome transcription factor associated gene [241].

A summary of the alteration, genes involved and genes pinpointed by the TS-CoExp tool is represented in table 2.19.

Table 2.19 - Summary of the alteration in patient 14 and genes involved. In bold are represented the most likely significant gene for the phenotype based on their function.

Patient Number	Event	Number of Genes Involved	Number of Probes Involved	Genes	TS-CoExp
#14	7q11.23 (72,379,797-72,783,852)X1	11	28	*BAZ1B, BCL7B, DNAJC30, FKBP6, FZD9,* ***MLXIPL****,* ***STX1A****, TBL2, TRIM50, VPS37D, WBSCR22*	*DNAJC30, STX1A, BAZ1B*

Patient #15

This patient is a 10 years old boy with mild ID (IQ=52) and familial history of ID (cousin from the maternal side of the family), who has microcephaly (head circumference bellow 3rd centile), recurrent respiratory infections, hyperactivity and no significant dysmorphic features (the only facial observation for the patient is his hypotelorism).

aCGH revealed a 0.3 Mb deletion at chromosome region 7q11.23 (chr7:72,052,516-72,359,842) containing 12 genes. Figure 2.18A presents the facial appearance of patient 15 while figure 2.18B shows the alteration detected by aCGH.

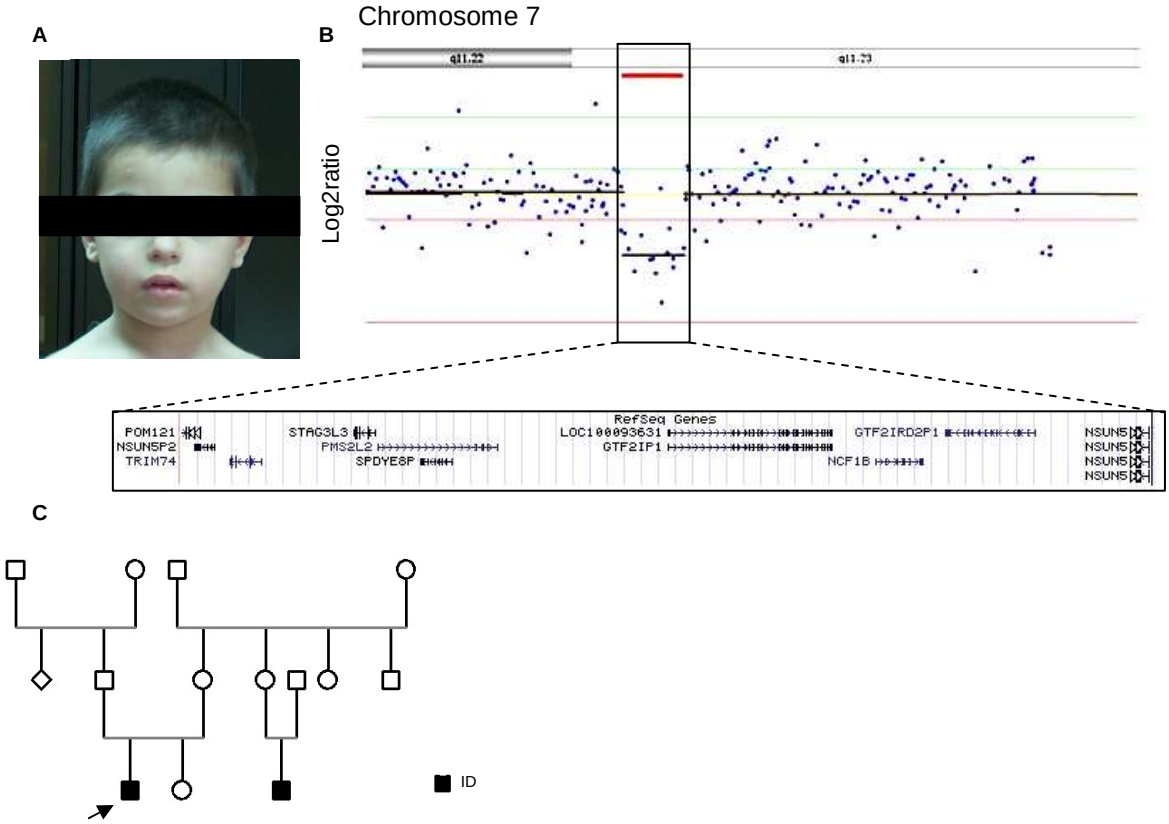


Figure 2.18 - Facial appearance, schematic representation of the genomic alteration and pedigree representation for patient 15. (A) The facial appearance of the patient shows his lack of significant facial dysmorphisms. **(B)** The genomic alteration is represented by the red line bellow 7q11.23 cytoband where it is possible to observe the diminished mean ratio of the probes contained in that region. The UCSC genes present in the genomic region altered in the patient in also represented. **(C)** The patient is represented by the arrow; he has a maternal cousin with ID.

The deletion present in patient 15 doesn´t overlap with the one described for patient 14, although it is still inside the WBS critical region. The only feature shared by patient 15 with WBS classic patients and patient 14 is the hyperactivity disorder. Considering only patient 14 and 15 they share microcephaly and hyperactivity.

The most interesting genes deleted in this patient are *GTF2IP1* (general transcription factor II-I, pseudogene) and *GTF2IRD2P1* (2I repeat domain containing 2 pseudogene) genes.
GTF2IP1 is the centromeric copy of GTF2I (general transcription factor II-I) and seems to be partially truncated expressed pseudogene with no protein product. These two genes (GTF2I, located in a more telomeric region, and *GTF2IP1,* located closer to the centromere) seems to have a role in the predisposition for WBS deletion [242]. Only *GTF2IP1* is deleted in patient 15 and only the haploinsufficiency of GTF2I have been attributed to ID feature of WBS [243].

GTF2IRD2P1 is also a pseudogene that has been implicated in the visual-spatial construction defects seem in WBS patients. A study performed by Dai L et al in 2009 in WBS patients with atypical rare small deletions retrieved the conclusion that GTF2IRD2P1 deletion could be associated with WBS facial appearance and visual-spatial construction [244]

TS-CoExp analysis for patient 15 retrieved *POM121* (POM121 membrane glycoprotein, rat) gene as the most likely to cause ID, among those covered by the deletion.
POM121 encodes an integral membrane component of the nuclear pore complex [245]. Other than integrating WBS deletions, no association with disease has been made so far for *POM121* gene.

A summary of the alteration, genes involved and genes pinpointed by the TS-CoExp tool is represented in table 2.20.

Table 2.20 - Summary of the alteration in patient 14 and genes involved. In bold are represented the most likely significant gene for the phenotype based on their function.

Patient Number	Event	Number of Genes Involved	Number of Probes Involved	Genes	TS-CoExp
#15	7q11.23 (chr7:72,052,516-72,359,842)X1	12	14	***GTF2IP1***, ***GTF2IRD2P1***, *LOC100093631, NCF1B, NSUN5, NSUN5C, NSUN5P2, PMS2L2, POM121, SPDYE8P, STAG3L3, TRIM74*	*POM121*

In table 2.21 is possible to observe a brief comparison between WBS classic patients, patient 14 and patient 15.

Table 2.21 – Phenotypic comparison between WBS core features, patient 14 and patient 15.

WBS feature	Patient 14	Patient 15
Mild to moderated ID	No (severe ID)	Yes (mild ID)
Facial dysmorphisms	Present (some of them)	Not present
Congenital heart disease	No	No
Anxious behavior	No	No
Attention deficit and hyperactive disorder	Yes	Yes
Overfriendliness	No (aggressiveness)	No

13. 7q33 deletion

Patient #16

This patient is a 19 years old boy with mild ID (IQ= 53), familial history of ID (mother and sister affected; a second-grade cousin with ID, dysmorphic features and epilepsy), seizures, reduced vision, hypotony, dysmorphisms (sunken eyes, high and thin palate, thin upper lip, bulbous nose, prognathism, open mouth), aggressiveness and hyperactivity.

aCGH revealed a 2 Mb deletion at chromosome region 7q33 (chr7:132,827,191-134,903,411) containing 15 genes. Figure 2.19A presents the facial appearance of patient 16 while figure 2.19B shows the alteration detected by aCGH.

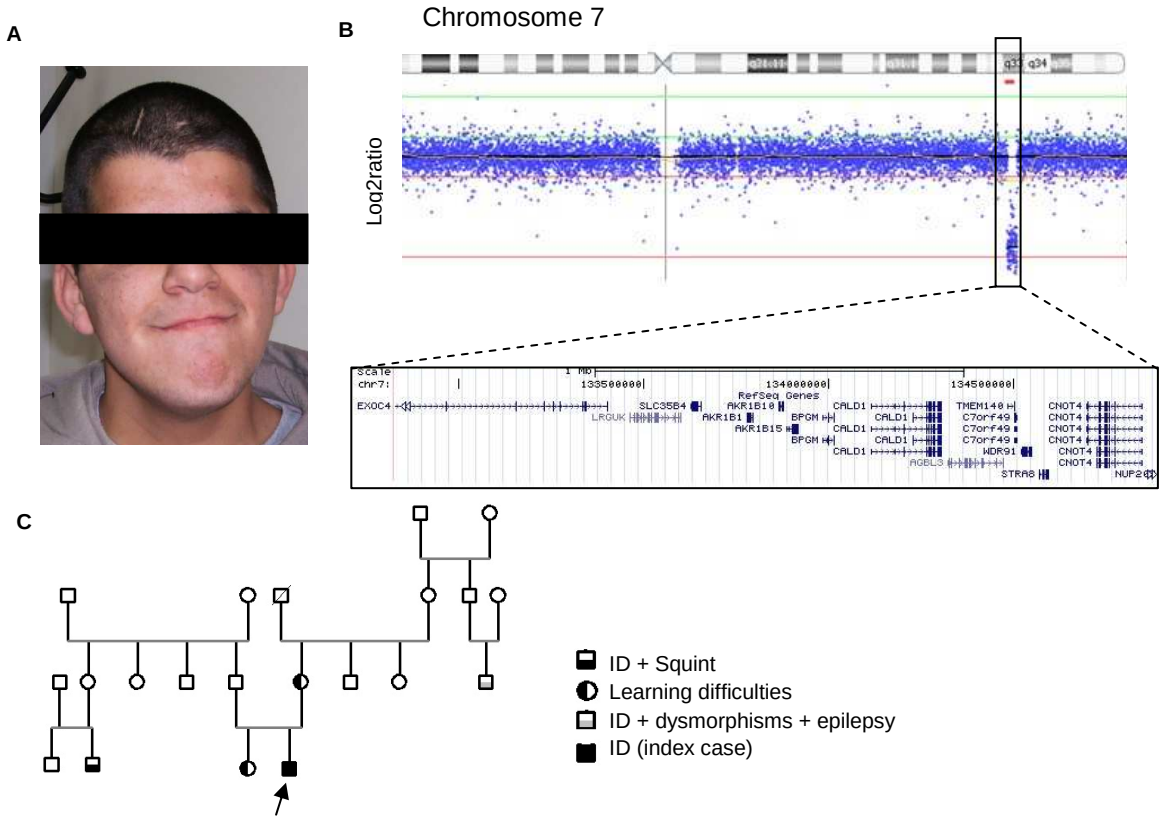


Figure 2.19 - Facial appearance, schematic representation of the genomic alteration and pedigree representation for patient 16. (A) The facial appearance of the patient shows his sunken eyes, thin upper lip, bulbous nose and prognathism. **(B)** The genomic alteration is represented by the red line bellow 7q33 cytoband where it is possible to observe the diminished mean ratio of the probes contained in that region. The UCSC genes present in the genomic region altered in the patient in also represented. **(C)** The patient is represented by the arrow; he has the sister and the mother with learning difficulties and a second-grade cousin with ID, dysmorphic features and epilepsy.

A slightly larger deletion (3 Mb) is described in one patient from Decipher (Decipher #256036) in a patient with ID and behavioral problems (emotional instability) among other features, two characteristics are also present in patient 16.

One of the genes covered by the deletion is the *CNOT4* (CCR4-NOT transcription factor complex, subunit 4) gene. The CNOT4 protein has both positive and negative roles in transcriptional regulation and a positive role in transcriptional elongation. In yeast, the human homolog (Not4) functioned as an E3 ubiquitin ligase and controlled the level of Jhd2 (the yest ortholog of *JARID1C*) [246]. Mutations in *JARID1C* (lysine-specific demethylase 5C) were reported in patients with X-linked ID, reveling that the correct expression of this protein is probably essential for correct neuronal function [247].

TS-CoExp analysis for patient 16 retrieved *CALD1* (caldesmon 1), *CNOT4* (CCR4-NOT transcription complex, subunit 4) and *EXOC4* (exocyst complex component 4) genes as the 3 more likely to cause ID, among those covered by the deletion.

CALD1 encodes for the caldesmon protein which is found in smoothe muscle and nonmuscle cells (in this last case, the low molecular weight caldesmon variant). Caldesmon is an actin-linked regulatory protein that binds and stabilizes actin filaments, and regulates actin-myosin interaction [248]. Since caldesmon has numerous functions in cell motility (such as migration, invasion, and proliferation), executed through the reorganization of the actin cytoskeleton, this protein is likely to have a functional contribution for ID pathogenesis [249].

EXOC4 is part of a multiprotein complex involved in vesicle trafficking at neural synapses and vesicle targeting in polarized epithelial cells [250].

A summary of the alteration, genes involved and genes pinpointed by the TS-CoExp tool is represented in table 2.22.

Table 2.22 - Summary of the alteration in patient 16 and genes involved. In bold are represented the most likely significant gene for the phenotype based on their function.

Patient Number	Event	Number of Genes Involved	Number of Probes Involved	Genes	Ts-CoExp
#16	7q33(132,827,191 -134,903,411)X1	15	131	*AGBL3, AKR1B1, AKR1B10, AKR1B15, BPGM, C7orf49,* ***CALD1****,* ***CNOT4****,* ***EXOC4****, LRGUK, NUP205, SLC35B4, STRA8, TMEM140, WDR91*	*EXOC4, CNOT4*

14. 9q34.13-q34.3 duplication and 14q32.31-q32.33 duplication

Patient #17

This patient is a 10 years old boy with mild ID (IQ= 57), elph-like face, low-set and prominent and large ears, submucous cleft palate, short philtrum, hypochromic macules and *café-au-lait* spots, stereotypies, obsessive behaviors aggressiveness, hyperactivity and attention deficit.

aCGH revealed a 5.5 Mb *de novo* duplication at chromosome region 9q34.13-q34.3 (chr9:134,757,732-140,273,252) containing 135 genes and a 3 Mb *de novo* (figure 2.20B) duplication at chromosome region 14q32.31-q32.33 (chr14:102,028,863- 105,602,384) containing 46 genes (figure 2.21B). Figure 2.20A presents the facial appearance of patient 17 while figure 2.20B shows the 9q34.13-q34.3 alteration detected by aCGH.

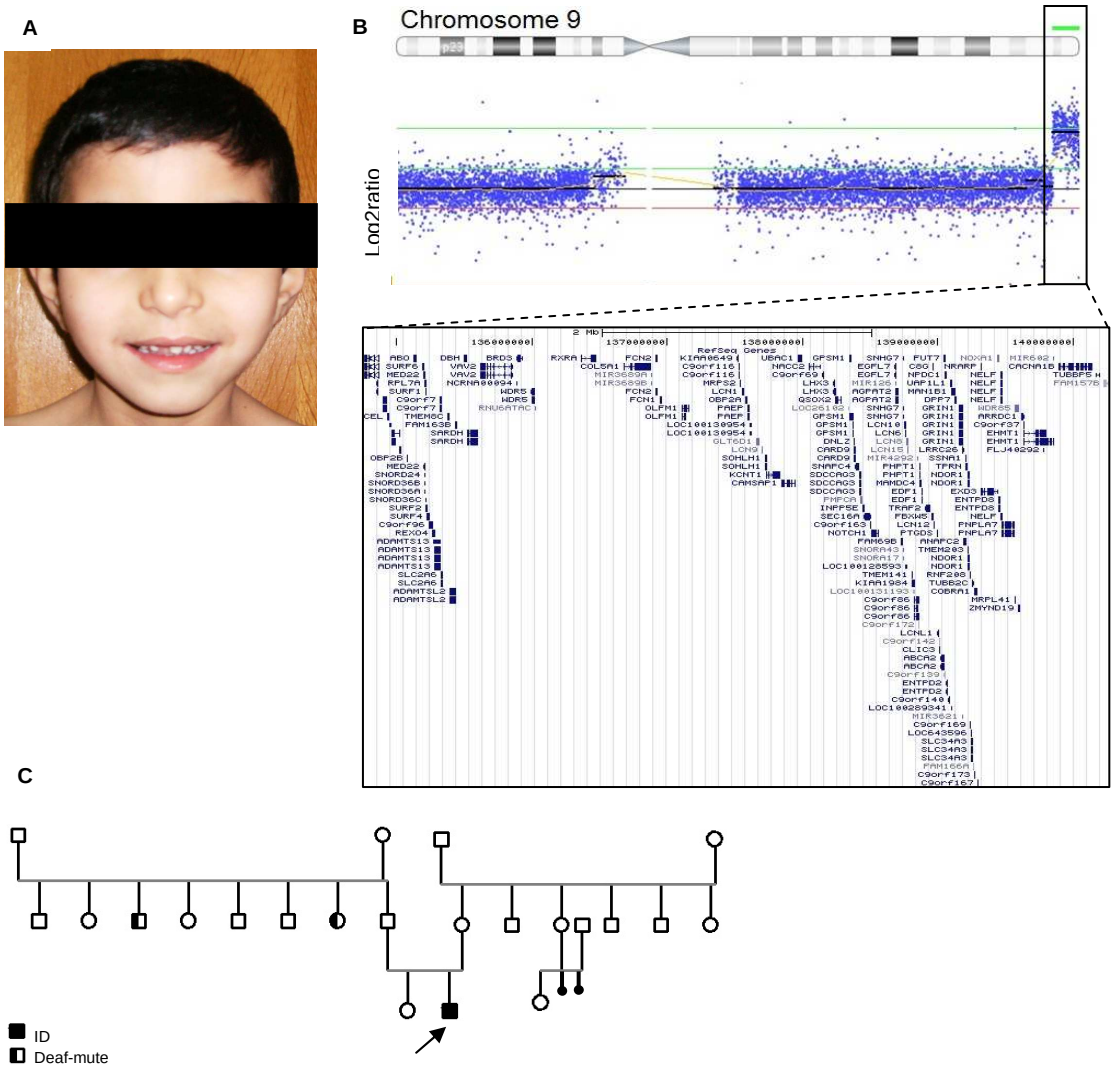


Figure 2.20 - Facial appearance, schematic representation of the genomic alteration and pedigree representation for patient 17. (A) The facial appearance of the patient shows his elph-like face, low-set and prominent and large ears. **(B)** The genomic alteration is represented by the green line above 9q34.13-q34.3 cytoband where it is possible to observe the increased mean ratio of the probes contained in that region. The UCSC genes present in the genomic region altered in the patient in also represented. **(C)** The patient is represented by the arrow; he has a deaf-mute uncle and aunt.

The Decipher patient with the most similar alteration (Decipher#254131) has a 6.3 Mb duplication (larger than the one from patient 17) and has hypotonia, ID, alterations in skin pigmentation and speech delay. When compared to our patient, we can see that they overlap in ID and pigment alteration but not in hypotonia (which is not present in our patient) [251].

Kleefstra syndrome (also known as 9q34.3 microdeletion syndrome) is characterized by ID, childhood hypotonia and distinctive facial features. It is currently believed that the *EHMT1* gene is the key gene, or at least of the most relevant, for the cognitive phenotype presented by the patients.

The duplicated region present in our patient also covers the *EHMT1* (euchromatic histone methyl transferase 1) gene. *EHMT1* haploinsufficiency is thought to be the core molecular change explaining the cognitive phenotype of this entity. Studies in a Drosophila EHMT mutant model revealed that its expression is required for memory and cognitive function. Moreover, the re-expression of EHMT in adulthood could rescue the cognitive defects caused by its lack during development [252].

TS-CoExp analysis for 9q34.13-q34.3 duplication retrieved *RXRA* (retinoid X receptor, alfa), *GRIN1* (glutamate receptor, ionotropic, N-methyl D-aspartate 1) and *UAP1L1* (UDP-N-acteylglucosamine pyrophosphorylase 1-like 1) genes as the 3 more likely to cause ID, among those covered by the duplication.

RXRA is a nuclear receptor protein, expressed in the brain, member of the steroid and thyroid hormone receptor superfamily of transcriptional regulators. A RXRA knockout mice model displayed embryonic heart failure by low levels of ventricular contractions [253].

The *GRIN1* gene encodes a glutamate receptor. Polymorphisms in *GRIN1* were associated with susceptibility for schizophrenia, suggesting that this gene may indeed be a good candidate gene for the disorder [254].

For *UAP1L1* gene no entrance is found in OMIM database and PubMed only retrieves one paper regarding breast cancer.

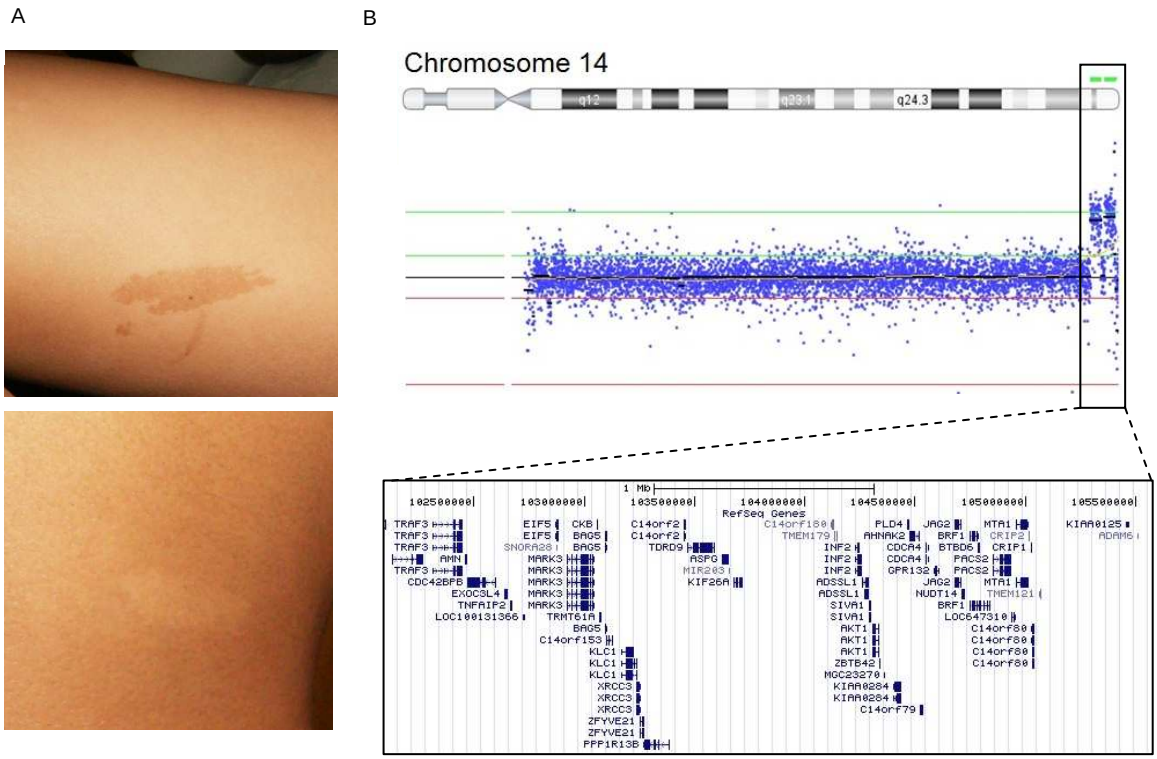


Figure 2.21 – Skin pigmentation alterations and schematic representation of the genomic alteration. The patient presents and café-au-lait **(A)** and hypochromic **(B)** spots on the skin. **(C)** The genomic alteration is represented by the green line above 14q32.31-q32.33 cytoband where it is possible to observe the increased mean ratio of the probes contained in that region. The UCSC genes present in the genomic region altered in the patient in also represented.

The 14q32.31-q32.33 duplication (figure 29) covers the 46 genes, of which MARK3 (MAP/microtubule affinity-regulating kinase 3) and KLC1 (kinesin light chain 1) could be candidate genes for the cognitive phenotype in the patient.

MARK3 belongs to the MARK proteins family known to be involved in the phosphorylation of microtubule associated proteins (such as TAU) [255]. A recent study from 2010 performed by Chia CY and colleagues reported that MARK3 is a gene with very stable expression, making it a good reference gene [256].

KLC1 encodes a kinesin that acts as a tubulin molecular precursor for transport of organelles within cells and in the movement of chromosomes along microtubules during cell division. Moreover, in neurons the axonal transport of APP is mediated by its binding to kinesin light chain subunit of kinesin-1 [257].

TS-CoExp analysis for 14q32.31-q32.33 duplication retrieved *EIF5* (eukaryotic translation initiation factor 5) and *MRPL41* (mitochondrial ribosomal protein L41) genes as those with the 2 more likely to cause ID, among those covered by the duplication.

EIF5 function as a GTPase-activating protein that plays a role in the establishment of the EIF2-GTP-tRNAi(Met) ternary complex of the ribosome during mRNA translation [258].

MRPL41 is one of the protein components of mitochondrial ribosomes encoded by the nuclear genome of the cell [259].

A summary of the alteration, genes involved and genes pinpointed by the TS-CoExp tool is represented in table 2.23.

Table 2.23 - Summary of the alteration in patient 17 and genes involved. In bold are represented the most likely significant gene for the phenotype based on their function.

Patient Number	Event	Number of Genes Involved	Number of Probes Involved	Genes	TS-CoExp
#17	9q34.13-q34.3(134,757,732-140,273,252)X3	135	371	*ABCA2, ABO, ADAMTS13, ADAMTSL2, AGPAT2, ANAPC2, ARRDC1, BRD3, C8G, C9orf116, C9orf139, C9orf140, C9orf142, C9orf163, C9orf167, C9orf169, C9orf172, C9orf173, C9orf37, C9orf69, C9orf7, C9orf75, C9orf86, C9orf96, CACNA1B, CAMSAP1, CARD9, CEL, CELP, CLIC3, COBRA1, COL5A1, DBH, DNLZ, DPP7, EDF1, EGFL7,* ***EHMT1****, ENTPD2, ENTPD8, EXD3, FAM157B, FAM163B, FAM166A, FAM69B, FBXW5, FCN1, FCN2, FLJ40292, FUT7, GBGT1, GFI1B, GLT6D1, GPSM1, GRIN1, GTF3C5, INPP5E, KCNT1, KIAA0649, KIAA1984, LCN1, LCN10, LCN12, LCN15, LCN6, LCN8, LCN9, LCNL1, LHX3, LOC100131193, LOC100289341, LOC26102, LRRC26, MAMDC4, MAN1B1, MED22, MIR126, MIR602, MRPL41, MRPS2, NACC2, NCRNA00094, NDOR1, NELF, NOTCH1, NOXA1, NPDC1, NRARP, OBP2A, OBP2B, OLFM1, PAEP, PHPT1, PMPCA, PNPLA7, PTGDS, QSOX2, RALGDS, REXO4, RNF208, RNU6ATAC, RPL7A, RXRA, SARDH, SDCCAG3, SEC16A, SLC2A6, SLC34A3, SNAPC4, SNHG7, SNORA17, SNORA43, SNORD24, SNORD36A, SNORD36B, SNORD36C, SOHLH1, SSNA1, SURF1, SURF2, SURF4, SURF6, TMEM141, TMEM203, TMEM8C, TRAF2, TSC1, TUBB2C, TUBBP5, UAP1L1, UBAC1, VAV2, WDR5, WDR85, ZMYND19*	*RXRA, GRIN1, UAP1L1*
	14q32.31-q32.33(102,028,863-105,602,384)X3	46	211	*AMN, ANKRD9,* ***ASPG****,* ***BAG5****, C14orf153, C14orf2, C14orf73, CDC42BPB, CKB, EIF5,* ***KLC1****,* ***MARK3****, PPP1R13B, RCOR1, SNORA28, TDRD9, TECPR2, TNFAIP2, TRAF3, TRMT61A, XRCC3, ZFYVE21, ADAM6, ADSSL1, AHNAK2,* ***AKT1****, BRF1, BTBD6, C14orf79, C14orf80, CDCA4, CRIP1, CRIP2, GPR132, INF2, JAG2, KIAA0125, KIAA0284, MGC23270, MTA1, NUDT14, PACS2, PLD4, SIVA1, TMEM121, ZBTB42*	*EIF5,, MRPL41*

When notified for the participation in the study, subtelomeric studies were still ongoing for patient 17. Together with the aCGH result, the subtelomeric studies for the patient retrieved an altered result. MLPA studies reveal a duplication of the gene EHMT1 (chromosome 9) and gene MAT1 (chromosome 14). These results were confirmed further confirmed by FISH.

In figure 2.22 is possible to see the FISH studies for patient 17.

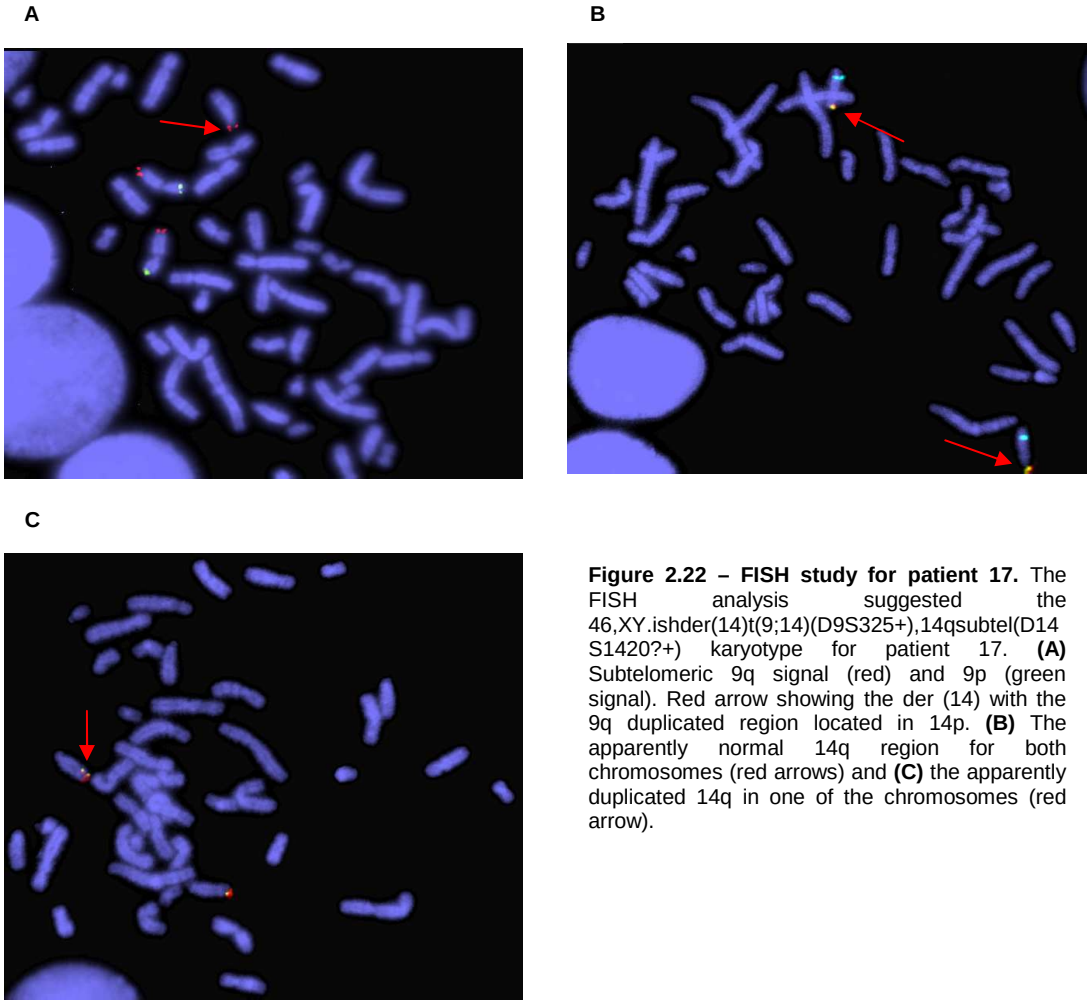


Figure 2.22 – FISH study for patient 17. The FISH analysis suggested the 46,XY.ishder(14)t(9;14)(D9S325+),14qsubtel(D14S1420?+) karyotype for patient 17. **(A)** Subtelomeric 9q signal (red) and 9p (green signal). Red arrow showing the der (14) with the 9q duplicated region located in 14p. **(B)** The apparently normal 14q region for both chromosomes (red arrows) and **(C)** the apparently duplicated 14q in one of the chromosomes (red arrow).

The MLPA and FISH analysis suggested a the 46,XY.ish der(14)t(9;14)(D9S325+),14qsubtel(D14S1420?+).mlpa9qsubtel(P036,P070)x3,14qsubtel(P036,P070)x3 karyotype for the patient. Patient 17 has a derivative chromosome that results from a *de novo* translocation for which the 9q duplicated region is located in 14p. It was not possible to determine in which chromosome 14 is located the 14q duplication (in the derivative or in the normal one).

Because it is located in tandem, the 14q32.31-q32.33 duplication may lead to the disruption of the genes involved in the breakpoints. However, since both duplications are very large and cover very interesting genes the most likely scenario is that both duplications encompass dosage sensitive genes and together account for the phenotype in the children**.**

15. **9q33.2-q33.3 duplication**

Patient #18

This patient is a 12 years old girl with mild ID (IQ=53) and familial history of ID (figure 25C). During the perinatal period the patient manifested epilepsy, hypothyroidism and cardiac insufficiency (heart murmur). Currently she has a three years delay in bone maturation, short stature, relative macrocephaly, short neck and dysmorphic face (high nasal bridge, high palate, micro and retrognathia, epicantus and long face and expressionless face) and reduced activity.

aCGH revealed a 3.6 Mb duplication at chromosome region 9q33.2-q33.3 (chr9:122,564,885-126,227,440) containing 52 genes. Figure 2.23A presents the facial appearance of patient 18 while figure 2.23B shows the alteration detected by aCGH.

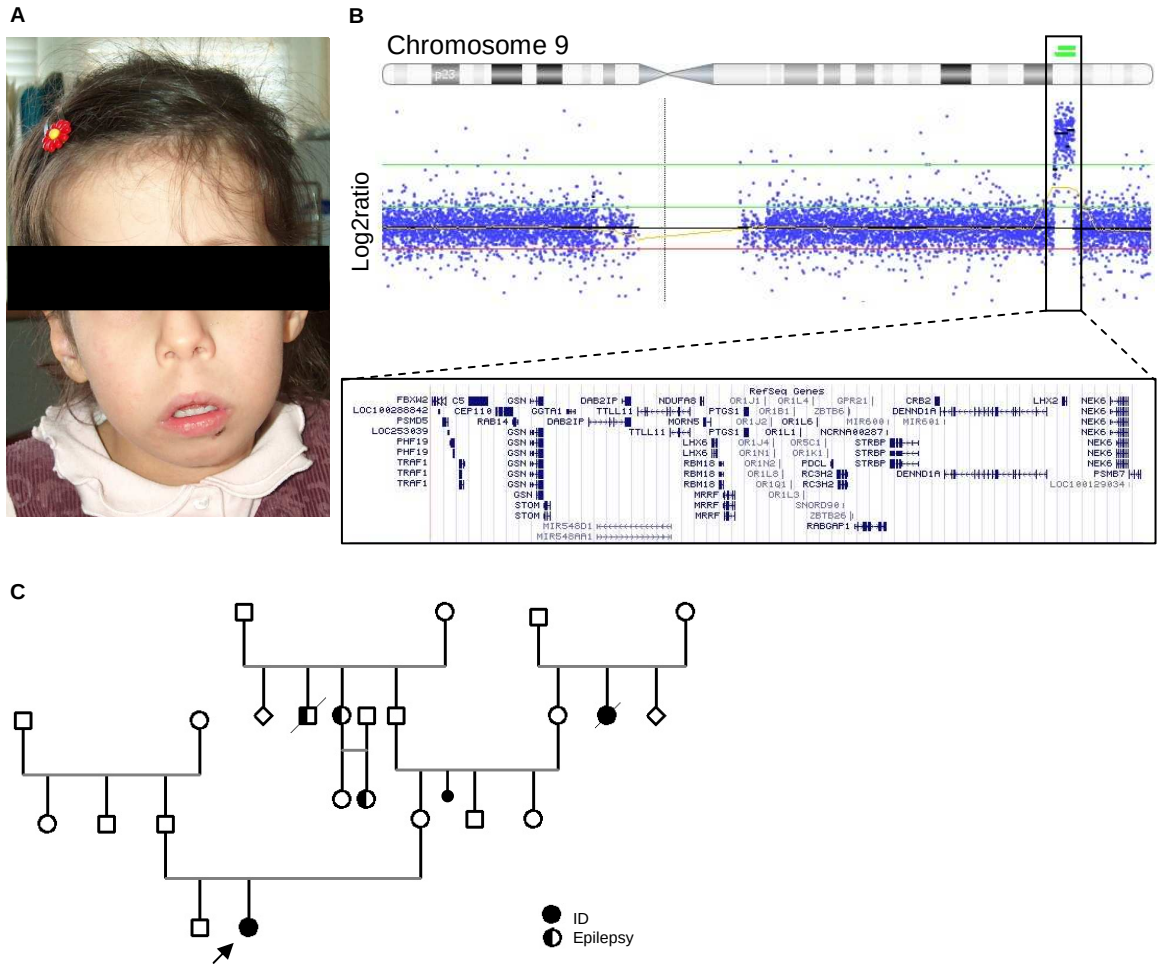


Figure 2.23 - Facial appearance, schematic representation of the genomic alteration and pedigree representation for patient 18. (A) The facial appearance of the patient shows her relative macrocephaly, short neck, high nasal bridge, micro and retrognathia, epicantus and long face and expressionless face. **(B)** The genomic alteration is represented by the green line above 9q33.2-q33.3 cytoband where it is possible to observe the increased mean ratio of the probes contained in that region. The UCSC genes present in the genomic region altered in the patient in also represented. **(C)** The patient is represented by the arrow. An aunt of the mother´s patients had ID (maternal side). Uncles and direct cousins of the mother's patient have epilepsy.

The most similar alteration described in Decipher is a 2.65 Mb deletion (Decipher# 249332) partially overlapping with the duplication of patient 18, however, the only feature shared by both patients seems to be ID.
Among the genes affected, *LHX2* (LIM homebox gene 2), *LHX6* (LIM homebox gene 6) and *STRBP* (Spermatid perinuclear RNA-binding protein) seem to play important functions in brain.

LHX2 encodes a protein expressed both in developing and adult rodent central nervous system. LHX2 acts as a classic selector gene and an essential intrinsic determinant of cortical identity. Indeed, Lhx2 -/- mice had defects in the development of the eyes, forebrain, and erythrocytes and died a few days before birth due to severe anemia [260].
LHX6 is expressed in GABAergic neurons from the medial ganglionic eminence, an early neuronal population that migrates to the developing cerebral cortex [261, 262].
STRBP encodes a protein mainly expressed in testis but with lower levels of expression in other tissues (such as brain). STRBP is both a RNA-binding protein and a microtubule-associated protein and, although it is mainly studied in sperm cells, its function in the assembly of the microtubules in other tissues cannot be excluded. In fact, a mouse model lacking STRBP expression displayed severe defects in brain, spinal cord and dorsal root ganglia [263].

TS-CoExp analysis for patient 18 retrieved *LHX2* (LIM homebox gene 2), *LHX6* (LIM homebox gene 6) and *STOM* (stomatin) genes as the 3 more likely to cause ID, among those covered by the duplication.
STOM, also known as *EPB72*, encodes the band integral membrane protein, a major protein of the erythrocyte membrane. The *STOM* -/- mice, despite the complete absence of protein, presented no hemolytic anemia and red blood cells were morphologically normal. Thus, their experiments suggested that 7.2b deficiency plays no direct role in the etiology of stomatocytosis and excluded any role of this protein as a mediator of cation transport in red blood cells [264].

A summary of the alteration, genes involved and genes pinpointed by the TS-CoExp tool is represented in table 2.24.

Table 2.24 - Summary of the alteration in patient 18 and genes involved. In bold are represented the most likely significant gene for the phenotype based on their function.

Patient Number	Event	Number of Genes Involved	Number of Probes Involved	Genes	Ts-CoExp
#18	9q33.2-q33.3 (122,564,885-126,227,440)X3	52	232	*C5, FBXW2, LOC253039, LOC402377, PHF19, PSMD5, TRAF1, C9orf45, CEP110, CRB2, DAB2IP, DENND1A, GGTA1, GPR21, GSN, LHX2, LHX6, LOC100129034, MIR548D1, MIR600, MIR601, MORN5, MRRF, NDUFA8, NEK6, OR1B1, OR1J1, OR1J2, OR1J4, OR1K1, OR1L1, OR1L3, OR1L4, OR1L6, OR1L8, OR1N1, OR1N2, OR1Q1, OR5C1, PDCL, PSMB7, PTGS1, RAB14, RABGAP1, RBM18, RC3H2, SNORD90, STOM, STRBP, TTLL11, ZBTB26, ZBTB6*	*LHX2, LHX6 STOM*

16. 10q26.3 deletion

Patient #19

This patient is a 5 years old girl with mild ID (IQ=52), short stature, squint, short neck, mild hypotony, lack of sensibility to peripheral pain, axial hypertonia, altered EEG, conductive hypoacusia, reduced activity and dysmorphic face (triangular face - looking like Trisomy 21, although similar to the parents' - and arched eyebrows). The patient also presents spherocytosis[9], as do several members of her family (maternal side).

aCGH revealed a *de novo* (determined by qPCR analysis) 0.65 Mb deletion at chromosome region 10q26.3 (chr10:131,264,691-131,920,458) containing 3 genes Figure 2.24A presents the facial appearance of patient 19 while figure 2.24B shows the alteration detected by aCGH.

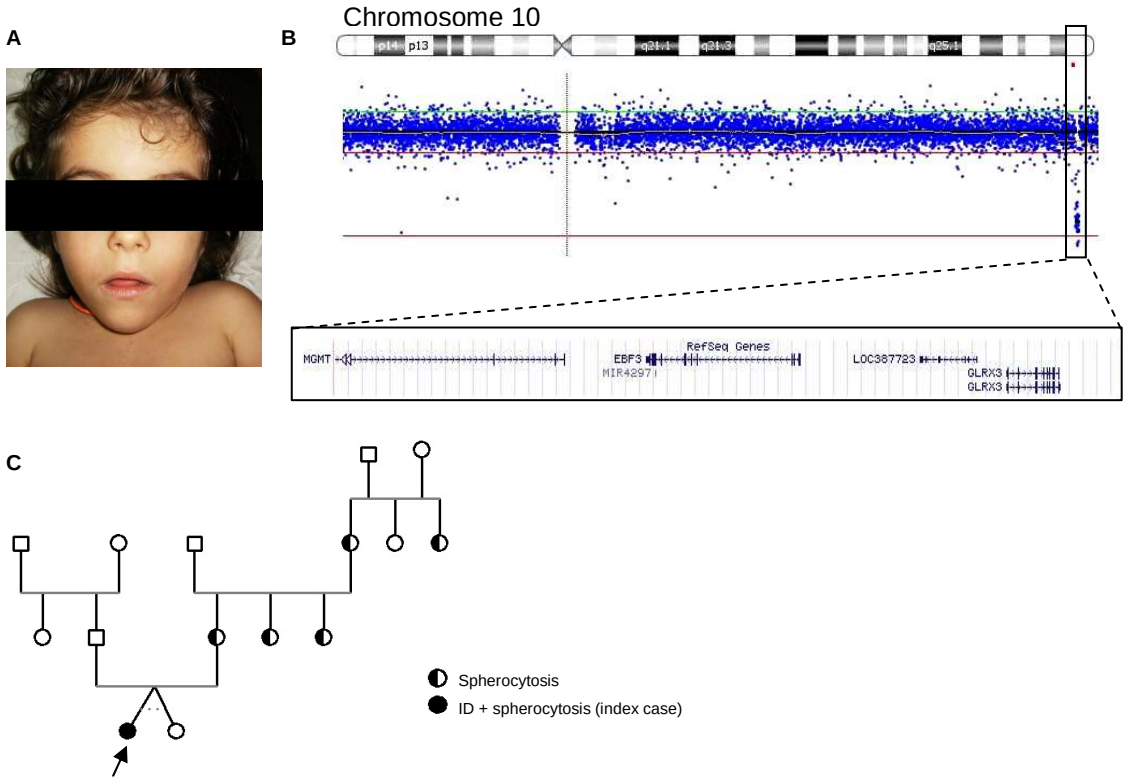


Figure 2.24 - Facial appearance, schematic representation of the genomic alteration and pedigree representation for patient 19. (A) The facial appearance of the patient shows her squint, short neck, triangular and arched eyebrows. **(B)** The genomic alteration is represented by the red line bellow 10q26.3 cytoband where it is possible to observe the decreased mean ratio of the probes contained in that region. The UCSC genes present in the genomic region altered in the patient in also represented. **(C)** The patient is represented by the arrow. Together with DD/ID she also presents spherocytosis, as do many other members from the maternal side of the family.

[9] Spherocytosis is an auto-hemolytic anemia characterized by the production of abnormal red blood cells. In spherocytosis patients these cell are sphere-shaped instead of bi-concave disk shaped. This disorder is caused by alterations in one or more proteins of the red blood cells cytoskeleton.

A patient with a smaller deletion causing the disruption of *EBF3* (early B-cell factor 3) and *MGMT* (methylguanine-DNA methyl transferase) genes is described in the Decipher database (Decipher#257229). The phenotype and classification of the alteration are not described in Decipher.

The *EBF3* gene is a transcription factor expressed in adult brain that regulates the transcription of genes involved in neurogenesis [265]. *EBF3* gene expression is regulated by ARX, a gene known to be involved in syndromic and non-syndromic forms of ID [266].

Besides the *EBF3* gene, the deletion present in the patient also covers the *GLRX3* (glutaredoxin 3) and *MGMT* (methylguanine-DNA methyltranferase) genes.

GLRX3 is a protein kinase c-interacting cousin of thioredoxin (PICOT) that inhibits pressure overload, inducing cardiac hypertrophy in rats by disrupting the Mlp-calcineurin interaction and thereby negatively regulating calcineurin-Nfat signaling [267]. Interestingly, calcineurin-Nfat signaling regulates genes necessary for myelination and is essential for neuregulin and ErbB signaling, neural crest diversification, and differentiation of Schwann cells [268].

MGMT is a repair protein involved in the correction of the DNA replicative error O(6)-alkylguanine. In glioblastoma patients the methylation status of MGMT predicts sensitivity to temozolomide, an alkylating agent used to treat glioblastoma patients [269].

TS-CoExp analysis for patient 19 retrieved *GLRX3* (glutaredoxin 3) gene as the one more likely to cause ID, among those covered by the deletion.

A summary of the alteration, genes involved and genes pinpointed by the TS-CoExp tool is represented in table 2.25.

Table 2.25 - Summary of the alteration in patient 19 and genes involved. In bold are represented the most likely significant gene for the phenotype based on their function.

Patient Number	Event	Number of Genes Involved	Number of Probes Involved	Genes	Ts-CoExp
#19	10q26.3 (131,264,691-131,920,458)X1	3	40	*EBF3, GLRX3, MGMT*	*GLRX3*

17. 12q24.21-q24.22 deletion

Patient #20

This patient is a 3 years old girl with moderate ID (IQ= 49) who has a second-grade cousin with learning difficulties. She also presents cardiac defects (intraventricular communication), abnormal skull shape, facial dysmorphisms (low nasal bridge, linear eyebrows, enophtalmia and entropion), photophobia, hypoplastic nipples, and anteriorly localized anus. The patient also displays reduced interaction with the external environment (only interacts when stimulated).

aCGH revealed a *de novo* (determined by qPCR analysis) 1 Mb deletion at chromosome region 12q24.21-q24.22 (chr12:113,989,883- 115,926,066), containing 8 genes. Figure 2.25A presents the facial appearance of patient 20 while figure 2.25B shows the alteration detected by aCGH.

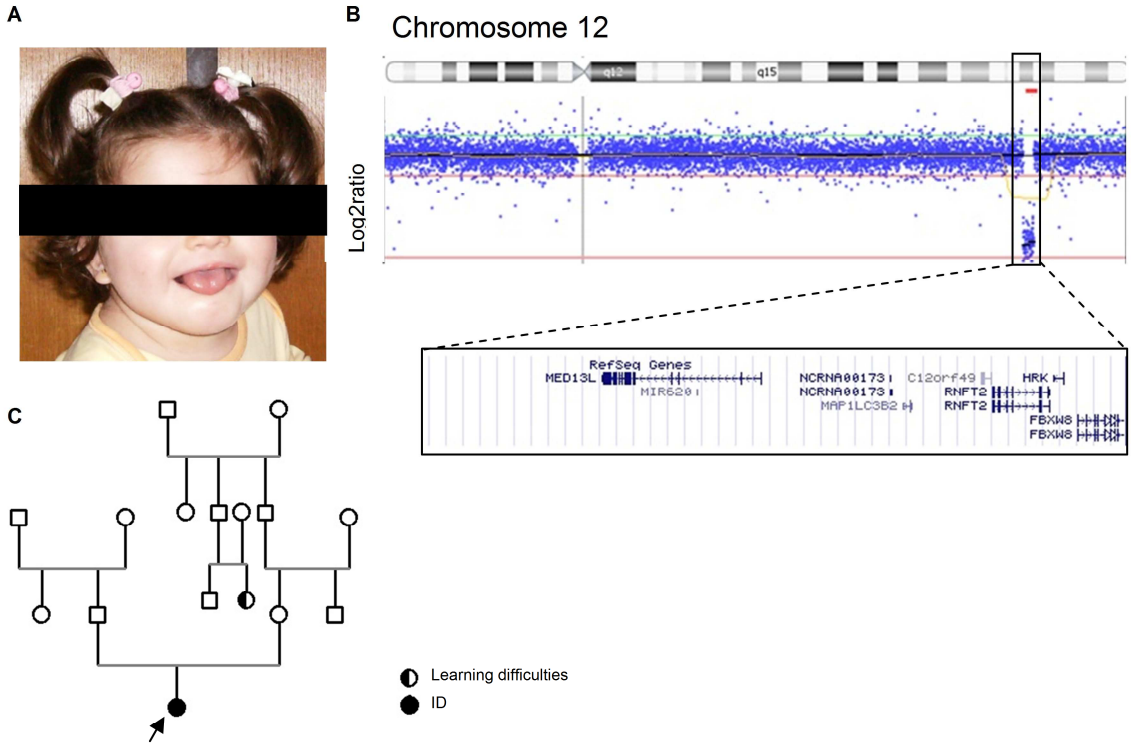


Figure 2.25 - Facial appearance, schematic representation of the genomic alteration and pedigree representation for patient 20. (A) The facial appearance of the patient shows her abnormal skull shape, low nasal bridge and linear eyebrows. **(B)** The genomic alteration is represented by the red line bellow 12q24.21-q24.22 cytoband where it is possible to observe the descreased mean ratio of the probes contained in that region. The UCSC genes present in the genomic region altered in the patient in also represented. **(C)** The patient is represented by the arrow; she has a second-grade cousin with learning difficulties.

In Decipher, only one larger deletion (1.2 Mb) encompassing only the *MED13L* (mediator complex subunit 13-like) gene is described in one patient (Decipher #4135). A related alteration was made

in 2006 by Ruiter *et al*, who described a 2.3 Mb duplication partially overlapping with the one from patient 20.

Similar to Decipher #4135, our patient has ID, anal malformations and facial dysmorphisms. Additionally, our patient also has cardiac defects. The fact that these two patients share the *MED13L* absence may indicate its importance for the disease. *MED13L* encodes the THRAP2 (thyroid hormone receptor-associated protein 2) protein, that belongs to a family of proteins with important functions in embryonic development. More precisely, it is a subunit of a large mediator complex that functions with DNA-binding transcription factors and RNA polymerase II for gene activation. Point mutations in the *MED13L* gene have been described in patients with cardiac malformations, namely isolated dextro-looped transposition of the great arteries, and one of these patients also presented ID [270].

Regarding the case described by Ruiter *et al*, patient 20 shared the ID phenotype, dysmorphic facial appearance and hypoplastic nipples. Although the alteration described by the authors was a duplication while the one present in our case is a deletion, the number of overlapping genes is very consistent. Besides *MED13L*, other interesting genes are located within the deleted region, for instance: *FBXW8* (F-box and WD40 domain protein 8) and *HRK* (karakiri).

FBXW8 encodes for a protein that is part of a complex that act as protein-ubiquitin ligases and binds to cullin 7. CUL7 protein was also found to interact with that complex and have a crucial role in the stabilization of the complex [271].

HRK encodes an apoptosis regulator protein, which is thought to activate cell death, and is expressed also in brian [272].

TS-CoExp analysis for patient 20 retrieved *MED13L* (mediator complex subunit 13-like), *HRK* (harakiri, BCL2 interacting protein, contains only BH3 domain) and *RNFT2* (ring finger protein, transmembrane 2) genes the 3 more likely to cause ID, among those covered by the deletion.

RNFT2 gene has currently no entrance in OMIM and PubMed databases.

A summary of the alteration, genes involved and genes pinpointed by the TS-CoExp tool is represented in table 2.26.

Table 2.26 - Summary of the alteration in patient 20 and genes involved. In bold are represented the most likely significant gene for the phenotype based on their function.

Patient Number	Event	Number of Genes Involved	Number of Probes Involved	Genes	Ts-CoExp
#20	12q24.21-q24.22(113,989,883-115,926,066)X1	8	114	*C12orf49,* ***FBXW8****,* ***HRK****, MAP1LC3B2,* ***MED13L****, MIR620, NCRNA00173, RNFT2*	*MED13L HRK, RNFT2*

18. 12q24.21 duplication

Patient #21

This patient is a 12 years old boy with moderate ID (IQ=43). The patient´s mother has epilepsy but suspended medication before pregnancy, however she suffered from pre-eclampsia during pregnancy. The patient has hyperactivity and is aggressive, which complicates the detailed clinical observation. The patient has only minor facial dysmorphisms (facial appearance similar to the mother) such as large nasal bridge, large mouth and arched eyebrows). The patient has a severe speech delay, two small hyperchromic spots on the skin, uncoordinated walking, camptodactily of the inferior members and artrogriposis of the hands.

aCGH revealed a 0.3 Mb *de novo* duplication (determined by qPCR) at chromosome region 12q24.21 (chr12:114,893,119-115,188,686) containing 2 genes. Figure 2.26A presents the facial appearance of patient 21 while figure 2.26B shows the alteration detected by aCGH.

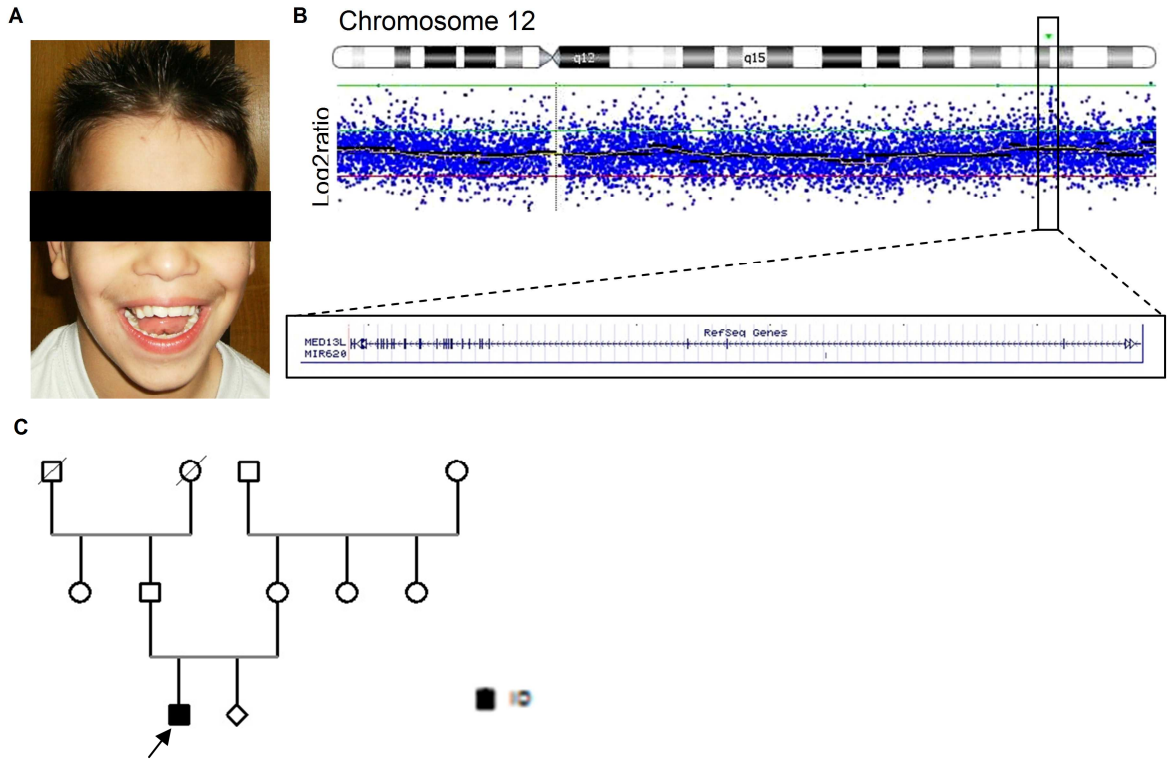


Figure 2.26 - Facial appearance, schematic representation of the genomic alteration and pedigree representation for patient 21. (A) The facial appearance of the patient shows his large nasal bridge, large mouth and arched eyebrows. **(B)** The genomic alteration is represented by the green line above 12q24.21 cytoband where it is possible to observe the increased mean ratio of the probes contained in that region. The UCSC genes present in the genomic region altered in the patient in also represented. **(C)** The patient is represented by the arrow.

Patient 21 has a smaller duplication in the same region where patient 20 has a deletion. This duplication encompasses the *MED13L* gene and, if located in tandem, may cause its disruption. The only shared clinical feature between patient 20 and21 seems to be the moderate level of ID.
In Decipher there is one patient (Decipher# 255109) with a 0.06 Mb *de novo* duplication that can disrupt the *MED13L* gene as well. Decipher# 255109 is described as having ID, auricular tags and macrostomia. Patient 21 also have ID and an increased mouth size, however, for the last one it is not clear if it a dysmorphisms or a just a familial similarity (mother).

TS-CoExp analysis for patient 21 retrieved *MED13L* (mediator complex subunit 13-like) gene as the one more likely to cause ID, among those covered by the duplication.

A summary of the alteration, genes involved and genes pinpointed by the TS-CoExp tool is represented in table 2.27.

Table 2.27 - Summary of the alteration in patient 21 and genes involved. In bold are represented the most likely significant gene for the phenotype based on their function.

Patient Number	Event	Number of Genes Involved	Number of Probes Involved	Genes	Ts-CoExp
#21	12q24.21(114,893,119-115,188,686)X3	2	19	***MED13L***, *MIR620*	*MED13L*

Figure 2.27 displays a schematic comparison between the alteration present in patients 20 and 21, and the two referred Decipher cases.

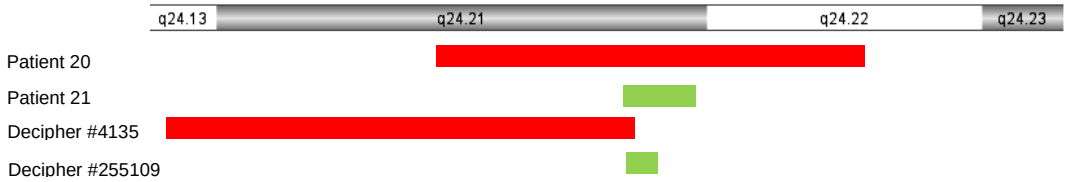


Figure 2.27 – Schematic comparison of the alterations present in patients 20 and 21 with the ones described in Decipher database. Patient 20 carries a 1 Mb deletion, patient 21 carries a 0,3 Mb duplication (both covering the *MED13L* gene), decipher #4135 carries a 1,23 Mb deletion and decipher #255109 carriers a 0,06 Mb duplication. Bothe Decipher cases are thought to also encompass the *MED13L* gene.

19. 20q13.12-q13.13 deletion

Patient #22

This patient is a 10 years old girl with mild ID (IQ= 56), microcephaly, sloping forehead, astigmatism, hypermetropia, retrognatia, “screwdriver teeth”, hypoplasic genitals, hyperactivity and speech delay.

aCGH revealed a 5.5 Mb *de novo* deletion at chromosome region 20q13.12-q13.13 (chr20:42,717,234-48,284,251) containing 88 genes. Figure 2.28A presents the facial appearance of patient 22 while figure 2.28B shows the alteration detected by aCGH.

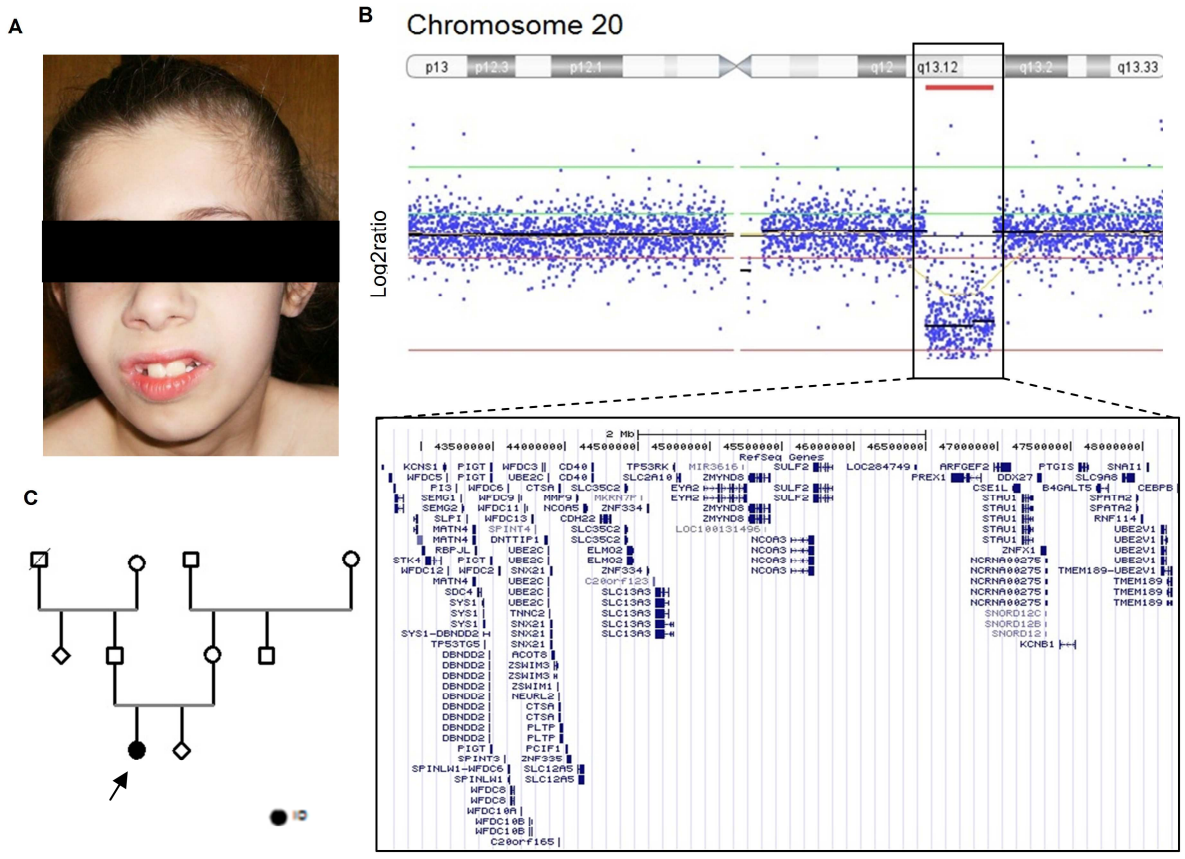


Figure 2.28 - Facial appearance, schematic representation of the genomic alteration and pedigree representation for patient 22. (A) The facial appearance of the patient reveals her sloping forehead, retrognatia and “screwdriver teeth”. **(B)** The genomic alteration is represented by the red line bellow the 20q13.12-q13.13 cytoband where it is possible to observe the decreased mean ratio of the probes contained in that region. The UCSC genes present in the genomic region altered in the patient in also represented. **(C)** The patient is represented by the arrow.

During the literature research, the case with the most similar genomic imbalance to our patient was Decipher patient #309. This patient has a 3.3 Mb deletion covering almost the majority of cytoband 20q13.13 and a small portion of 20q13.12. Decipher #309 is a male with no described ID but with

speech delay, microcephaly, hypoplasic genitals (small penis), feeding problems and facial dysmorphisms. When compared with patient 22, we can see that they share the microcephaly, hypoplasic genitals and speech delay. One of the genes referred as likely haploinsufficient within the affected region is the *CSE1L* (chromosome segregation 1-like), that encodes a protein important for apoptosis and cell proliferation. Recently duplications in the 20q13.13 region were described, which are associated with isolated growth hormone deficiency, a condition where infants have, among others features, micropenis and growth failure [273].

SPATA2 (spermatogenesis-associated protein 2) encodes a highly hydrophilic protein expressed, among other tissues, in brain. In *zebrafish*, the Spata2 orthologue was shown to be expressed and strongly localized at central nervous system at early development stages. Thus, although this protein was first described as involved in the spermatogenic process, further functional studies need to be performed in order to clarify its importance for brain development [274].

UBE2C (ubiquitin-conjugating enzyme E2C) encodes a protein that is essential for the transition of the anaphase-promoting complex into a state of permission for cyclin and accumulation during G1 phase of mitosis [275, 276].

TS-CoExp analysis retrieved *ARFGEF2* (ADP-ribosylation factor guanine nucleotide exchange factor 2), *RNF114* (ring finger protein 114) and *CEBPB* (CCAAT/enhancer binding protein (C/EBP), beta) genes as the 3 more likely to cause ID, among those covered by the deletion.

The ARFGEF2/BIG2 protein is required for vesicle and membrane trafficking from the TGN and inhibition of *ARFGEF2* has been shown to decrease neural expansion. Mutations in the *ARFGEF2* gene are the cause of autosomal recessive periventricular heterotopia with microcephaly [277].

RNF114 belongs to a family of RING domain E3 ubiquitin ligases characterized by 3 zinc fingers and an ubiquitin-interacting motif. RNF114 has high expression levels in testis and much lower expression in the other tissues [278].

CEBPB has a DNA-binding activity in the S phase of mitosis. Studies in mouse revealed that CEBPB was expressed in cortical progenitor cells and could induce expression of a reporter gene containing the minimal promoter of alpha-tubulin (a neuron-specific gene) [279].

A summary of the alteration, genes involved and genes pinpointed by the TS-CoExp tool is represented in table 2.28.

Table 2.28 - Summary of the alteration in patient 22 and genes involved. In bold are represented the most likely significant gene for the phenotype based on their function.

Patient Number	Event	Number of Genes Involved	Number of Probes Involved	Genes	TS-CoExp
#22	20q13.12-q13.13 (42,717,234-48,284,251)X1	88	363	*ACOT8, **ARFGEF2**, B4GALT5, C20orf123, C20orf165, C20orf199, CD40, CDH22, CEBPB, **CSE1L**, CTSA, DBNDD2, DDX27, DNTTIP1, ELMO2, EYA2, KCNB1, KCNK15, KCNS1, LOC100131496, LOC100240726, LOC284749, MATN4, MIR1259, MMP9, NCOA3, NCOA5, NEURL2, PABPC1L, PCIF1, PI3, PIGT, PLTP, PREX1, PTGIS, RBPJL, RIMS4, RNF114, SDC4, SEMG1, SEMG2, SLC12A5, SLC13A3, SLC2A10, SLC35C2, SLC9A8, SLPI, SNAI1, SNORD12, SNORD12B, SNORD12C, **SNX21, SPATA2**, SPINLW1, SPINT3, SPINT4, STAU1, STK4, SULF2, SYS1, SYS1-DBNDD2, TMEM189, TMEM189-UBE2V1, TNNC2, TOMM34, TP53RK, TP53TG5, **UBE2C, UBE2V1**, WFDC10A, WFDC10B, WFDC11, WFDC12, WFDC13, WFDC2, WFDC3, WFDC5, WFDC6, WFDC8, WFDC9, WISP2, YWHAB, ZMYND8, ZNF334, ZNF335, ZNFX1, ZSWIM1, ZSWIM3*	*ARFGEF2, RNF114, CEBPB*

20. Xq24 duplication (*CUL4B* gene duplication)

Patient #23

This patient is a 8 years old boy without psychometric criteria for classification as ID (IQ= 80), but who has familial history of ID (two brothers and cousins with ID), cardiac arrhythmia[10] (apparently benign) generalized obesity, testicular asymmetry (atrophic left testicle), stereotypies (shaking hands when excited) and hyperactivity.

aCGH revealed a 0.3 Mb duplication inherited from the mother at chromosome region Xq24 (chrX:119,476,634-119,789,009) containing 4 genes. Figure 2.29A presents the facial appearance of patient 23 while figure 2.29B shows the alteration detected by aCGH.

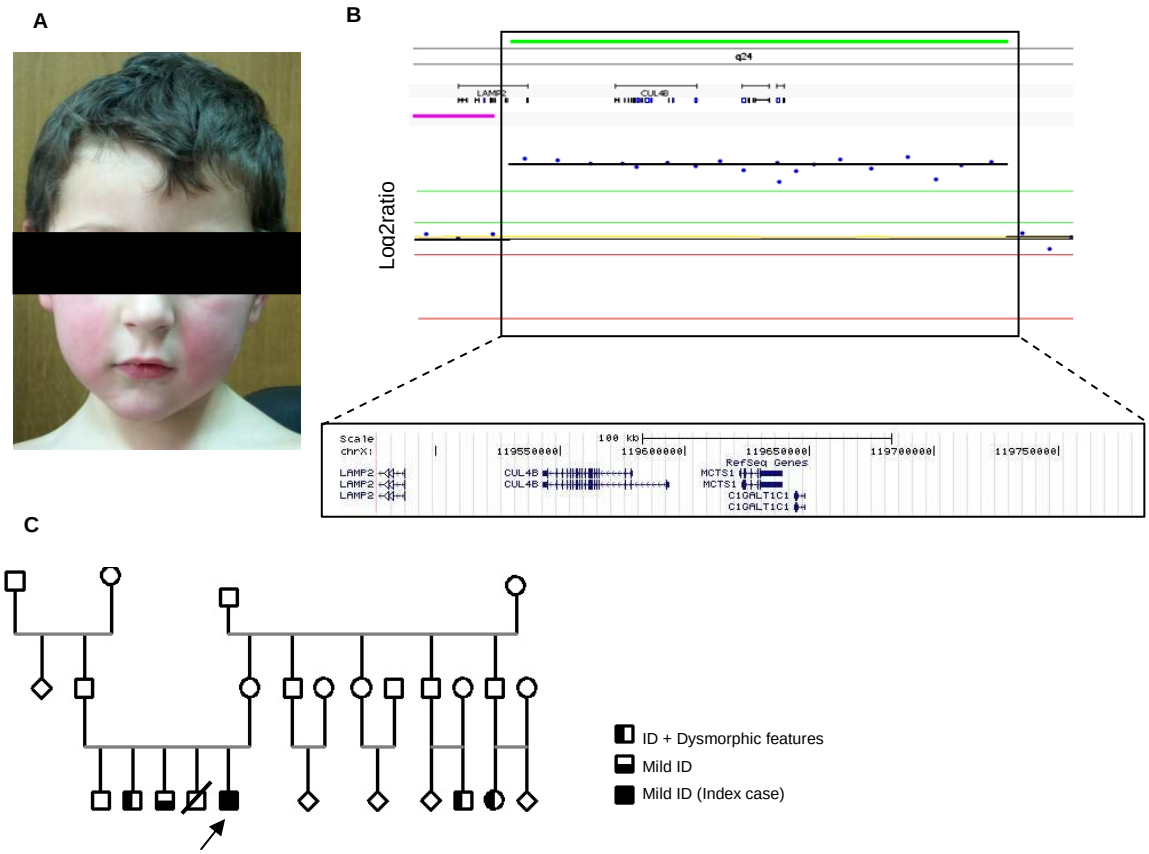


Figure 2.29 - Facial appearance, schematic representation of the genomic alteration and pedigree representation for patient 23. (A) The facial appearance of the patient. **(B)** The genomic alteration is represented by the green line above the Xq24 cytoband where it is possible to observe the increased mean ratio of the probes contained in that region. The UCSC genes present in the genomic region altered in the patient in also represented. **(C)** The patient is represented by the arrow. The patient has a brother with mild ID and other one with ID and dysmorphic features. The patient also has 2 cousins (both sons of the mothers' brother) with ID and dysmorphisms.

[10] Cardiac arrhythmia is a designation used to describe a large and heterogeneous group of conditions in which there is abnormal electrical activity of the heart (the heart beat is irregular, to fast or to slow).

The duplication found the patient covers, among others, the *CUL4B* gene. Point mutations and a deletion encompassing the entire *CUL4B* gene were described in patients with X-linked ID [280]. Until now, duplications affecting *CUL4B* have not been described.

CUL4B encodes for a ubiquitously expressed cullin protein that is a component of the ubiquitin ligase complex. Mutations in *CUL4B* gene were found in patients with X-linked ID, generalized obesity, delayed puberty, hypogonadism, relative macrocephaly, moderate short stature and aggressiveness.

CUL4B qPCR quantification analysis was performed for the patient, one of the brothers (for the other brother it was not possible to collect DNA sample) and the mother. The qPCR analysis revealed the existence of 2 *CUL4B* copies in the index case as well as in one of the affected brothers and three copies in the mother, leading us to believe that the alteration segregates with the disease. The brother for which was not possible to collect DNA samples is now a working adult who is thought to have a very mild cognitive impairment. The duplication present in the patient also covers the *C1GALT1C1* (C1GALT1 - core 1 beta-3-galactosyltransferase-specific molecular chaperone 1), *LAMP2* (lysosome-associated membrane protein 2) and *MCTS1* (malignant T-cell amplified sequence 1).

C1GALT1C1 is a molecular chaperone that contains beta-1,3-galactosyltransferase activity and specifically associates with C1GALT1. Mutations in C1GALT1 have been described in patients with Tn syndrome[11] [281].

The LAMP2 protein is thought to protect the lysosomal membrane from proteolytic enzymes as well as to display a receptor function for proteins to be imported into lysosomes [282]. Patients with Danon disease[12] were found to carry mutations in *LAMP2*. Mice models for this disorder revealed that LAMP2 has a crucial role in autophagy and that Danon disease is caused by the accumulation if autophagic material in the striated myocytes [283]. Since *LAMP2* gene may be disrupted with the duplication, a more detailed cardiac assessment should be carried out in the patient.

MCTS1 gene was found to be amplified in a T-cell lymphoma cell line and, consequently, was named after that [284, 285]. The protein is currently thought to be involved in cell cycle progression and is ubiquitously expressed in tissues [285].

[11] Tn syndrome is a rare autoimmune disease in which subpopulations of blood cells of all lineages carry an incompletely glycosylated membrane glycoprotein (called the Tn antigen). This syndrome is caused by mutation is the *C1GALT1* coding gene. This pathology is caused by recessive mutations in somatic cells.

[12] Danon disease, also known as X-linked vacuolar cardiomyopathy and myopathy, is an X-linked dominant disorder predominantly affecting cardiac muscle and that in some cases can be seen together with skeletal muscle involvement and ID. The disease is caused by mutations in LAMP2 gene which leads to the intracytoplasmic autophagic vacuoles with sarcolemmal features.

TS-CoExp analysis retrieved *CUL4B* and *LAMP2* genes as the 2 more likely to cause ID, among those covered by the duplication.

A summary of the alteration, genes involved and genes pinpointed by the TS-CoExp tool is represented in table 2.29.

Table 2.29 - Summary of the alteration in patient 23 and genes involved. In bold are represented the most likely significant gene for the phenotype based on their function.

Patient Number	Event	Number of Genes Involved	Number of Probes Involved	Genes	Ts-CoExp
#23	Xq24(119,476,634-119,789,009)X3	4	19	*C1GALT1C1,* ***CUL4B****, LAMP2, MCTS1*	***CUL4B, LAMP2***

XCI for the patient´s mother revealed a skewed XCI with a normalized ratio of 4.5 (about 85% of XCI deviation). This value is consistent with the fact that the mother is not suspected to have any form of ID and that she has 3 affected sons. This data also lead us to believe that similar to what happens in *CUL4B* point mutations and deletions, its duplication is also causing the disease. The most likely scenario is that *CUL4B* gene is a dosage-sensitive gene essential for intellectual functioning.

Patients with copy number changes associated with Susceptibility loci

1. 1q21.1 recurrent microduplication

Patient #24

This patient is a 5 years old boy with moderate ID (IQ= 44), familiar history of psychiatric disorders (brother with Autism/Asperger and a second grade cousin with ID), thin corpus calosum, autism, hyperactivity and attention deficit. The patient doesn´t display any significant dysmorphisms or malformation.

aCGH revealed a 2.5 Mb duplication inherited from the father at chromosome region 1q21.1 (chr1:144,577,524-147,095,314) containing 43 genes. Figure 2.30A presents the facial appearance of patient 24 while figure 2.30B shows the alteration detected by aCGH.

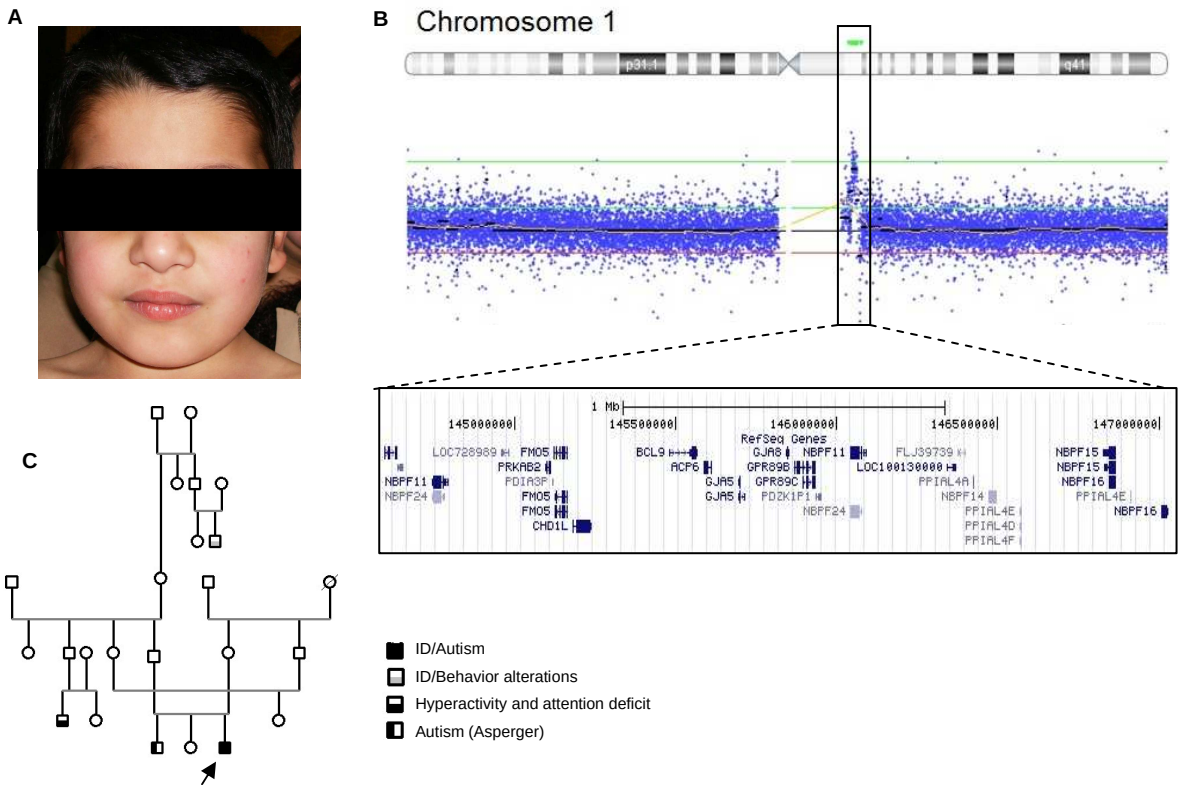


Figure 2.30 - Facial appearance, schematic representation of the genomic alteration and pedigree representation for patient 24. (A) The facial appearance of the patient reveals lack of significant dysmorphisms. **(B)** The genomic alteration is represented by the green line above the 1q21.1 cytoband where it is possible to observe the increased mean ratio of the probes contained in that region. The UCSC genes present in the genomic region altered in the patient in also represented. **(C)** The patient is represented by the arrow. The patient has a brother with Asperger's syndrome and a cousin with hyperactivity and attention deficit. A second grade cousin from the father side of the family also displays ID and behavioral alterations.

The 1q21.1 recurrent microduplication is considered a possible susceptibility locus for neurodevelopment disorders. The main features of patients carrying duplications in this region are autism or autism spectrum disorders, development delay and ID. Less frequent, but also present in some cases reported, are macrocephaly (around 50% of the cases) and dysmorphic features [286]. A peculiar characteristic of this region and other susceptibility loci, is that the alterations present in the patients are also commonly found in the phenotypically normal parents, implying the incomplete penetrance of this genomic imbalance [286, 287].

Our patient displays phenotypic features concordant with those described in the literature [286]. The father is a carrier for this alteration and closer analyses of his IQ level and psychiatric profile should be performed in order to determine is he is a truly non affected individual or an underdiagnosed case of a less severe psychiatric trait. The fact that the patient also has a familial history of autism (brother) and ID (second grade cousin from the father side) indicates that the father could also have transmitted the alteration to the brother and even that the alteration could also be present in more members of the family.

The phenotypically diversity and incomplete penetrance of this alteration raises complex counseling questions for the families.

The duplication present in the patient covers, among others, three interesting genes based on their described function: *BCL9* (B-cell CLL/lymphoma 9), *GJA5* (GAP junction protein, alpha-5) and *GJA8* (GAP junction protein, alpha-8).

BCL9 protein is described as playing a critical role in the regulation of Wnt signaling pathway in the developing embryonic body axis [288]. Genome wide association studies in schizophrenic patients have described variants in *BCL9* gene as conferring risk for the disease [289].

GJA5 encodes for a conexin protein that belongs to a wide family of protein that form intercellular channels (so called GAP junctions) that participate in the exchange of ions and small molecules between adjacent molecules [290]. Mice models lacking *GJA5* showed cardiac malformations, indicating that its reduced expression may have a role in cardiac tissue correct development [291].

GJA8 encodes also a conexin protein, for which mutations were described in patients with visual impairments caused by cataract [292].

TS-CoExp analysis retrieved *PRKAB2* (protein kinase, AMP-activated, beta 2 non-catalytic subunit) and *CHD1L* (chromodomain helicase DNA binding protein 1-like) genes as the 2 more likely to cause ID, among those covered by the duplication.

PRKAB2 encodes an AMP-activated protein kinase subunit that acts a metabolic stress-sensing protein kinase and is highly expressed in skeletal muscle [293]. No association between *PRKAB2* gene and disease has been established so far.

CHD1L interacts with poly(ADP-ribose) (PAR) and contributes to the cromatin relaxation following DNA damage [294]. CHD1L overexpression has been found in hepatocellular carcinoma [295].

A summary of the alteration, genes involved and genes pinpointed by the TS-CoExp tool is represented in table 2.30.

Table 2.30 - Summary of the alteration in patient 24 and genes involved. In bold are represented the most likely significant gene for the phenotype based on their function.

Patient Number	Event	Number of Genes Involved	Number of Probes Involved	Genes	Ts-CoExp
#24	1q21.1 (144,577,524-147,095,314)X3	43	93	*ACP6, AF131738, AF380582, AK023548, AK023809, AK091688, AK309393, BC036212, BC041003, BC110832,* ***BCL9****, CHD1L, CR610404, CR617196, DKFZp434H1228, DKFZp451B1418, DQ576969, FAM108A3, FMO5,* ***GJA5, GJA8****, GPR89A, GPR89B, GPR89C, KIAA1693, LOC200030, LOC728989, NBPF1, NBPF10, NBPF11, NBPF12, NBPF14, NBPF15, NBPF16, NBPF20, NBPF24, PDIA3P, PDZK1P1, PPIAL4D, PPIAL4E, PPIAL4F, PRKAB2, hPACPL1*	*PRKAB2, CHD1L*

2. 16p13.11 recurrent microduplication

Patient #25

This patient is a 5 years old boy without psychometric criteria for classification as ID (IQ= 88) but with verbal IQ scoring 58 (speech delay), who has familiar history of ID (cousin with ID and sister with learning difficulties) and was born with an interventricular communication corrected by surgery. The patient has small stature, squint, dysmorphic features (micrognathia, retrognathia, bulbous nose, thin upper lip), short neck, hirsutism and hyperactivity.

aCGH revealed a 0.8 Mb duplication inherited from the mother at chromosome region 16p13.11 (chr16:15,391,681-16,215,845) containing 9 genes. Figure 2.31A presents the facial appearance of patient 25 while figure 2.31B shows the alteration detected by aCGH.

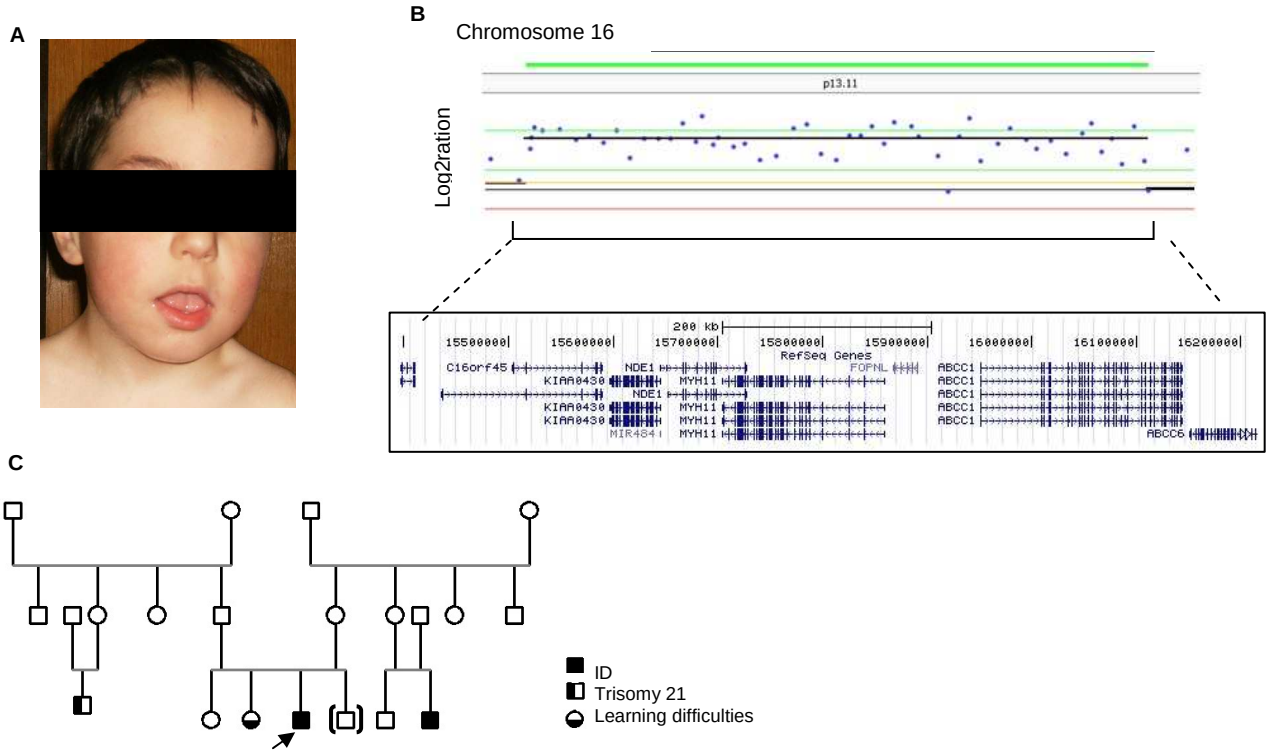


Figure 2.31 - Facial appearance, schematic representation of the genomic alteration and pedigree representation for patient 25. (A) The facial appearance of the patient reveals squint, dysmorphic features (micrognathia, retrognathia, bulbous nose, thin upper lip) and a short neck. **(B)** The genomic alteration is represented by the green line above the 16p13.11 cytoband where it is possible to observe the increased mean ratio of the probes contained in that region. The UCSC genes present in the genomic region altered in the patient in also represented. **(C)** The patient is represented by the arrow. The patient has a sister with learning difficulties in school and a cousin with ID (mother´s side of the family). A cousin from the father side of the family has Down syndrome.

Despite the fact that alteration in the 16p13.11 region still raise conflicts regarding its pathogenicity or only susceptibility for neurocognitive disorders, deletions in this region have been found in patients with ID and/or multiple congenital abnormalities as well as in phenotypically healthy

parents. It is currently widely accepted that deletion is a risk factor for ID and congenital malformations while the real significance of the duplications remains uncertain [296].
In the patients reported by Hannes FD and colleagues, brain abnormalities (namely microcephaly) and seizures were both main features in the patients. Additionally, small stature and some unspecified dysmorphic features were also present. Patient 5 presents no IQ criteria for classification as intellectually handicapped, was born with a heart congenital malformation, has short stature and facial dysmorphisms. The fat that the mother carries the same alteration and that the patient has a cousin (from the mother side) with ID may indicate that this alteration is also present in the mother´s sister and son. The absence of ID in the patient but the presence of malformations and concordant characteristics with previously described cases, together with the fact that the mother is a carrier and there are more cases of ID in the maternal side of the family reveal once again the incomplete penetrance find in this kind of alterations.

A very interesting gene present covered by the duplication in this patient is the *NED1* (nudE nuclear distribution gene E homolog 1, A. nidulans) gene. In the work published by Hannes FD *et al*, 2008 the authors propose the *NDE1* as a good candidate for the abnormal brain size detected in patients with ID and/or multiple congenital anomalies. NED1 protein is highly expressed in the brain and *Nde1*-null mice have microcephaly. However, the screening for mutations in *NED1* searching for sequence alterations in the remaining allele revealed no mutations [296]. However, our patient doesn´t have an altered brain size.

TS-CoExp analysis retrieved *ABCC1* (ATP-binding cassette, sub-family C (CFTR/MRP), member 1), *NDE1* (nudE nuclear distribution gene E homolog 1, *A. nidulans*) and *KIAA0430* (KIAA0430) genes as the 3 more likely to cause ID, among those covered by the duplication.
ABCC1, also called multidrug resistance-associated protein (MRP), is a plasma membrane drug-efflux pump. The blocking of ABCC1 function increases the cells susceptibility for suffering toxic effects. One example is the case of the increase in unconjugated bilirubin and its association with neonatal encephalopathy [297]. The *KIAA0430* gene has no entrance in OMIM and very little is known about its function.
A summary of the alteration, genes involved and genes pinpointed by the TS-CoExp tool is represented in table 2.31.

Table 2.31 - Summary of the alteration in patient 25 and genes involved. In bold are represented the most likely significant genes for the phenotype based on their function.

Patient Number	Event	Number of Genes Involved	Number of Probes Involved	Genes	TS-CoExp
#25	16p13.11(15,391,681 -16,215,845)X3	9	51	*ABCC1, ABCC6, C16orf45, C16orf63, KIAA0430, MIR484, MPV17L, MYH11,* ***NDE1***	*ABCC1, NDE1, KIAA0430*

Patient #26

This patient is a 4 years old boy with moderate ID (IQ= 49); the most significant fact in the family history is the epilepsy presented by the patient´s aunt (mother´s sister). The mother had chickenpox at 36 weeks of pregnancy; the patient was born with normal somatometry and facial palsy. The patient has weight and occipital-frontal circumference close to 95th centile, is hypotonic, has abnormal positioning of the arms (pronation), autism and dysmorphisms, such as brachycephaly, round face, small nose and short neck.

aCGH revealed a 0.9 Mb duplication inherited from the mother at chromosome region 16p13.11 (chr16:15,384,327-16,351,469), containing 11 genes. Figure 2.32A presents the facial appearance of patient 26 while figure 2.32B shows the alteration detected by aCGH.

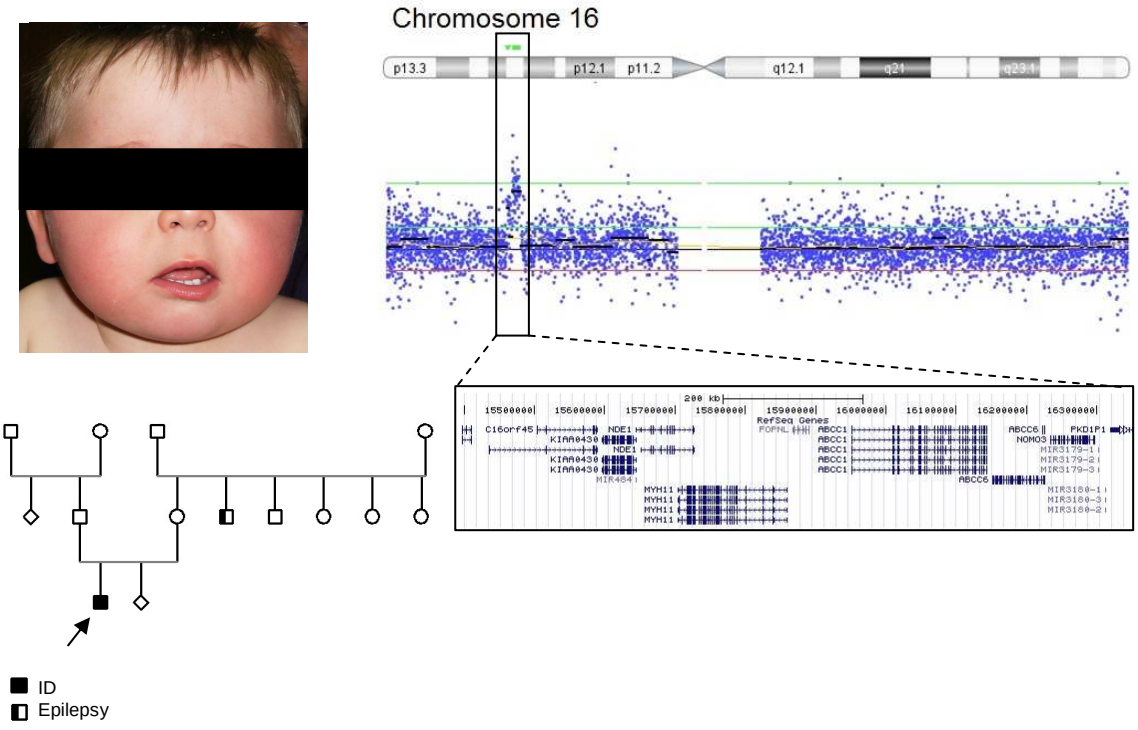


Figure 2.32 - Facial appearance, schematic representation of the genomic alteration and pedigree representation for patient 26. (A) The facial appearance of the patient reveals his brachycephaly, round face, small nose and short neck. **(B)** The genomic alteration is represented by the green line above the 16p13.11 cytoband where it is possible to observe the increased mean ratio of the probes contained in that region. The UCSC genes present in the genomic region altered in the patient in also represented. **(C)** The patient is represented by the arrow. The patient has aunt (mother´s sister) with epilepsy.

In contrast to patient 25 who had no quantitative criteria for ID, patient 26 is moderately intellectually handicapped. This patient presents dysmorphisms, autism and hypotony.

One important difference between patient 24 and 26, is the fact that patient 26 has two more genes possibly covered by the duplication: *LOC339047* and *NOMO3* (nodal modulator 3). For the *LOC339047* gene there is no entrance in OMIM or PubMed yet. *NOMO3* is a gene with a metal-binding domain with high similarity with NOMO1, a protein that is part of a protein complex that

participates in the Nodal signaling pathway, an important pathway for differentiation during development.
Similarly to the previous patient, *NED1* gene is also covered by the duplication, being a good candidate for the increased head circumference of the patient and, consequently, ID phenotype.

Ts-CoExp analysis retrieved the same results as in patient 25.

A summary of the alteration, genes involved and genes pinpointed by the TS-CoExp tool is represented in table 2.32.

Table 2.32 - Summary of the alteration in patient 26 and genes involved. In bold are represented the most likely significant genes for the phenotype based on their function.

Patient Number	Event	Number of Genes Involved	Number of Probes Involved	Genes	TS-CoExp
#26	16p13.11(15,384,327 -16,351,469)X3	11	54	*ABCC1, ABCC6, C16orf45, C16orf63, KIAA0430, LOC339047, MIR484, MPV17L, MYH11,* ***NDE1****,* ***NOMO3***	*ABCC1, NDE1, KIAA0430*

Once again, the fact that the mother carries the same alteration and her sister has epilepsy raises the hypothesis of the presence of this alteration in more members of the family revealing the incomplete penetrance of the alteration.

Challenges in the interpretation of aCGH results

1. CNVs present in healthy parents

Copy number variations of large dimensions and/or encompassing possibly interesting genes were found to be carried by healthy parents.

In patient 27, two alterations of unknown significance were found by aCGH. One of these alterations was a 0.07 Mb deletion in region 4q24 (chr4:107,322,522-107,395,454) (figure 2.33A) and a 0.17 Mb deletion in 7p21.1 (chr7:16,449,856-16,625,033) (figure 2.33B).

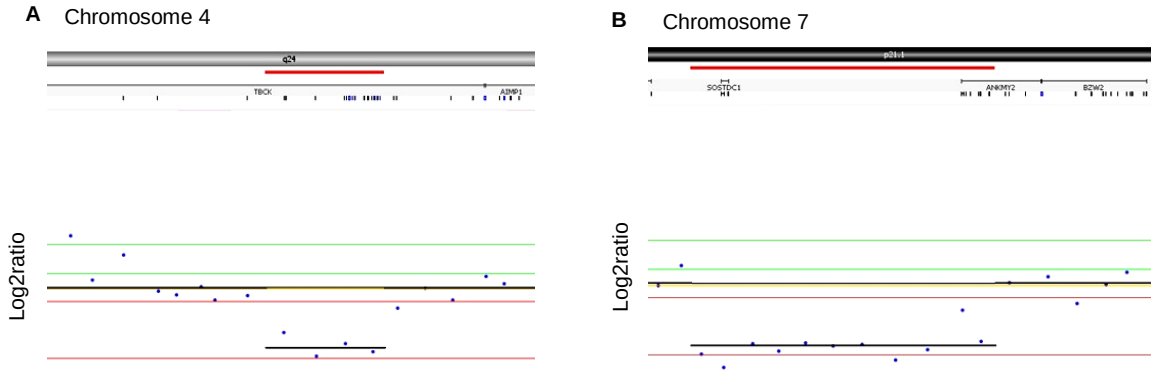


Figure 2.33 – Representation of the genomic alterations found in patient 27. In (A) is represented the deletion found in 4q24 while in (B) is represented the deletion found in 7p21.1.

Both alterations could possibly cause the disruption of the genes involved, however they were found to be inherited from the healthy mother. The confirmation of the presence of the alterations in the mother excluded them as causative of pathology which ended in a negative case for pathogenic CNVs.

In patient 28, a 0.1 Mb deletion was found in 13q14.3 (chr13:51,413,362-51,515,663) (figure 2.34) covering three genes.

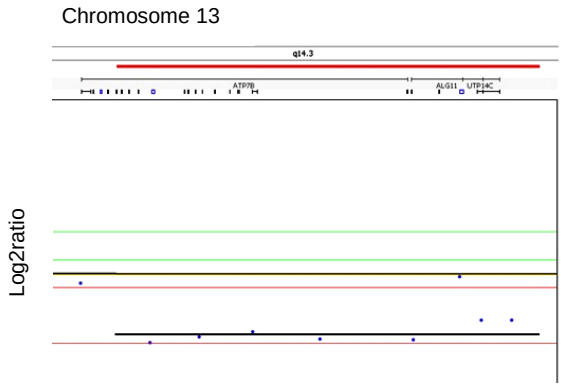


Figure 2.34 - Representation of the 13q14.3 deletion found in patient 28.

One of those genes was the *ATP7B* (ATPase, Cu(2+)-transporting, beta polypeptide), the gene mutated in Wilson disease[13]. In this case, qPCR analysis for *ATP7B* gene showed the presence of the alteration in the father. Thus we found that father and son are heterozigous carriers of a pathogenic variant causing autosomal recessive disorder, raising genetic counselling issues, such as if the finding of genomic information not requested by the genomic test should be given to the patient or not.

In patient 29, three alterations of unknown significance were find by aCGH: a 0.1 Mb duplication at 4q13.2 (chr4:69,038,815-69,170,341), a 0.5 Mb duplication at 4q13.3 (chr4:70,766,959-71,285,452) region and a 0.2 Mb duplication at 4q21.1 (chr4:76,987,928-77,215,148) (figure 2.35A).

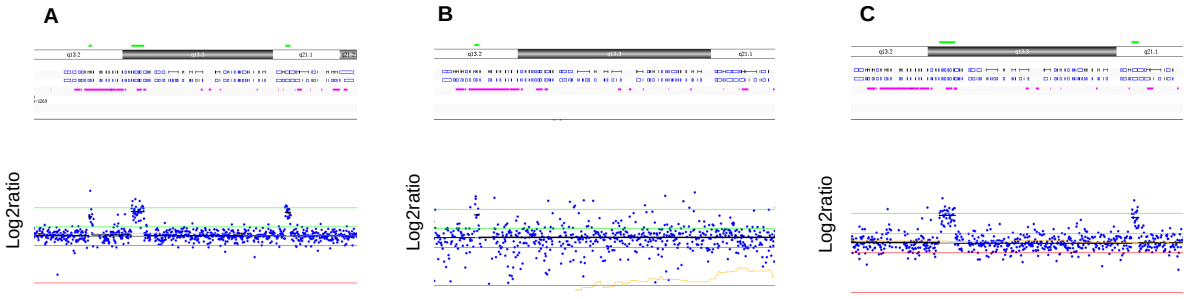


Figure 2.35 - Representation of the genomic alterations found in patient 29. In (A) is represented the duplications found in 4q13.2, 4q13.3 and 4q21.1 present in the index case. In (B) is represented the 4q13.2 duplication found in the mother and in (C) are represented the 4q13.3 and 4q21.1 duplications found in the father.

aCGH analysis in the parents revealed the presence of the first alteration in the mother (figure 2.35B), while the other two were present in the father (figure 2.35C). The 4q13.2 duplication alone (present in the healthy mother) has criteria for classification as a benign CNV since it is present in more than 3 healthy controls. However, this is not the case for 4q13.3 and 4q21.1 duplications (both present in the father). Together, these duplications encompassed 23 genes, for which no significant CNVs were reported in DGV.

In this case, the patient inherited both CNVs from each parent ending out with a chromosome 4 with 3 distinct alterations. We cannot conclude that these two alterations are benign because when alone they may not cause the phenotype, but together they may account for the disease in the child. This is one case for which, even though we have data from the parents, no clear conclusions will be achieved.

[13] Wilson disease is an autosomal recessive disorder characterized by dramatic build-up of intracellular hepatic copper with subsequent hepatic and neurologic abnormalities.

In patient 30, a 0.7 Mb duplication at 19p13.3-p13.2 (chr19:6,283,716-6,944,284) and a 0.6 Mb duplication at 19p13.2 (chr19:7,028,066-7,633,437) were found (figure 2.36) encompassing a total of 34 genes together. Exactly the same alteration was found in the mother, for whom a suspicion of mild ID was raised. From the paternal side of the family there is also history of ID and deaths during childhood and teenage years (patients suspected to have ID). A cousin of the patient has optical atrophy and diabetes.

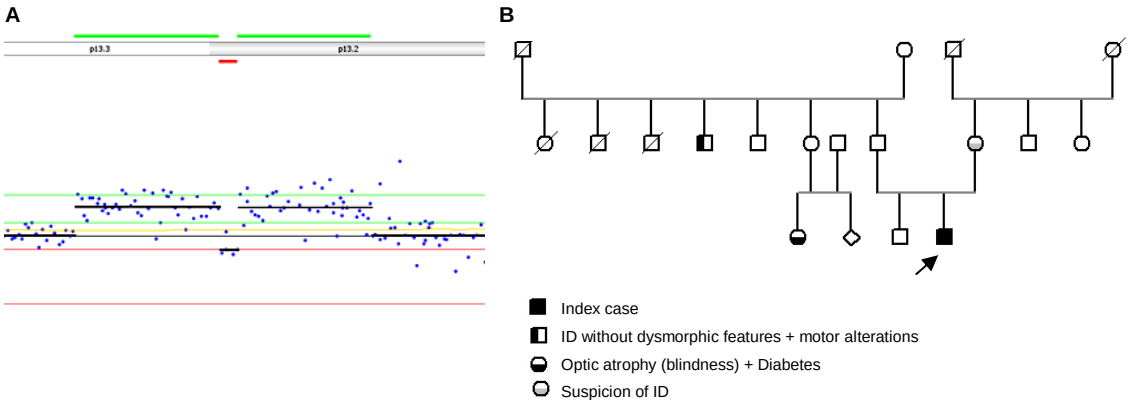


Figure 2.36 - Representation of the genomic alterations found in patient 30. In (A) is represented the duplications found in 4q13.2, 4q13.3 and 4q21.1 present in the index case. In (B) is represented the 4q13.2 duplication found in the mother and in (C) are represented the 4q13.3 and 4q21.1 duplications found in the father.

Once again, in this case we cannot clearly conclude about the significance of the alteration is present in the mother also suspected to have a milder form of ID. Adequate psychometric evaluation needs to be performed in the mother in order to assess her IQ. Another important aspect to take into account is that this alteration can also be a susceptibility locus that together with any other alteration inherited from the father can account for the more severe phenotype in the child while by itself has a lighter phenotype (the case of the mother).

The summary of these four cases as well as the genes involved in the alterations is represented in table 2.33.

Table 2.33 – Summary of the alterations described above as well as the genes involved in each case.

Patient	Alteration	Cytoband	Location	Inheritance	Genes
A	Deletion	4q24	chr4:107,322,522-107,395,454	Mother	*TBCK*
	Deletion	7p21.1	chr7:16,449,856-16,625,033	Mother	*ANKMY2, SOSTDC1*
B	Deletion	13q14.3	chr13:51,413,362-51,515,663	Father	*ALG11, ATP7B, UTP14C*
C	Duplication	4q13.2	chr4:69,038,815-69,170,341	Mother	*TMPRSS11E, UGT2B17*
	Duplication	4q13.3	chr4:70,766,959-71,285,452	Father	*C4orf35, C4orf40, C4orf7, CSN1S1, CSN1S2A, CSN1S2B, CSN2, CSN3, HTN1, HTN3, ODAM, SMR3A, SMR3B, STATH*
	Duplication	4q21.1	chr4:76,987,928-77,215,148	Father	*ART3, CXCL10, CXCL11, CXCL9, NAAA, PPEF2, SDAD1*
D	Duplication	19p13.3-p13.2	chr19:6,283,716-6,944,284	Mother	*ACER1, ALKBH7, C3, CD70, CLPP, CRB3, DENND1C, EMR1, EMR4P, GPR108, GTF2F1, KHSRP, MIR220B, PSPN, SH2D3A, SLC25A23, SLC25A41, TNFSF14, TNFSF9, TRIP10, TUBB4, VAV1*
	Duplication	19p13.2	chr19:7,028,066-7,633,437	Mother	ARHGEF18, C19orf45, INSR, KIAA1543, LOC100128573, LOC100131801, MCOLN1, PCP2, PEX11G, PNPLA6, STXBP2, XAB2, ZNF358, ZNF557

The surprising discovery for these four cases highlights the crucial importance of confirmation in the parents and, consequently, of the collection of parents' samples and clinical information.

CNVs of unknown significance

The increasing resolution achieved by aCGH platforms nowadays allows a deeper look at the alterations present in the patients, many times not easy to interpret. In spite of the adopted strategies for CNV classification, doubts in the application of the criteria many times arise and fitting certain alterations in each one of the categories established is often difficult.

For 27 cases in the present cohort alterations were found that are still raising doubts concerning their clinical significance. Whenever a case like this is found, it is necessary to delineate a confirmatory strategy for the case and parents. The decision of the approach to use will be taken based on the (I) the number of alterations that are necessary to confirm, (II) the nature of the alterations (deletion or duplication, size, mean probes ratio, if there are described controls) and (III) the number and nature of the affected genes.

A large portion of these alterations are:

- Deletions or duplications for which CNVs have been described in control cases, but in less than 3 controls and/or not covering the entire region.
- Presence of deletions or duplications not described in the control database, for which a large number of the reciprocal alterations (duplications or deletions, respectively) are present in controls. Given the comparative nature of the method, and given that the reference samples is actually a pool of DNAs from healthy individuals, the excessive presence of one type of CNVs in the reference samples may lead to the observation of a false reciprocal CNVs in our test sample often with dubious signal. In these cases, the confirmation of the alteration by another method (real time quantitative PCR) is often necessary. In figure 2.37 there is represented an example of the previous described situation.

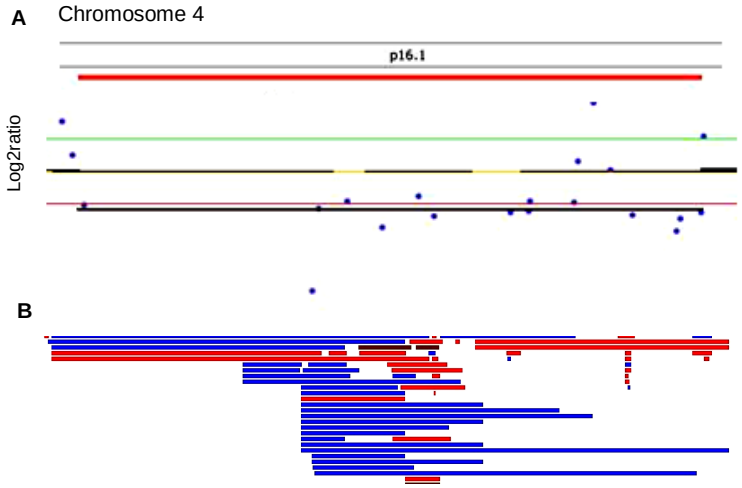


Figure 2.37 – Schematic representation of a dubious mean probes ratio. An example of an alteration present in 4p16.1 cytoband where the probe mean ration is near the lower treshold to call an aberration **(A).** In image **(B)** there is represented the described CNVs in control in healthy individuals in DGV for the same region (blue lines for the duplications and red lines for the deleltions). Notive the increase number of duplications (blue lines) described when compared with the deletions (red lines).

- In genomic imbalances for which controls have a deletion described while the patient has duplication, if the breakpoints affect the same genes, it is more likely that the duplication in the patient is not causing the disease, since deletions that lead to haploinsufficiency of the encompassed genes are not pathogenic as well. However, if the situation is the other way around, pathogenicity is more likely and a confirmatory test (using qPCR, for instance) is required.

In cases like the ones described above, the genomic analysis for the parents may not always be helpful for the determination of its significance. We cannot forget that a large portion of mutations (either point mutations or genomic imbalances) occur de novo during development and their absence in the parents may not give us its real significance for the phenotype.

Other type of alterations that may lead to necessity of confirmatory techniques is when a alteration is found in a region not covered by CNVs described in controls and encompasses genes. The nature of the gene may give us clues about its possible significance for the disease displayed by the patient. In this work, special attention was given to genes that are expressed in the nervous system (particularly in brain) and that, based on their function, can play important roles in any of the pathways described to be involved in the pathogenesis of ID.

The most significant alterations found in these 27 patients as well as the genes involved and associated function are briefly described in table 2.34.

Table 2.34 – Summary of the cases still in study and for which confirmation with another technique or assessment of the parents is required in order to conclude about the clinical significance of the aCGH findings. The symbol * indicates the gene that could be disrupted by the breakpoint in the alteration (duplication or deletion).

Patient Code	Event	Chromosome Region	Cytoband	Genes Involved	Possible most relevant gene for ID	Function/Phenotype associated	Confirmation strategy (ongoing)
389	Duplication	chrX:70,352,533-70,433,152	Xq13.1	*GJB1*, NONO*, ZMYM3*	*ZMYM3*	ID	qPCR; Evaluate parents
453	Duplication	chrX:70,352,533-70,433,152	Xq13.1	*GJB1*, NONO*, ZMYM3*	*ZMYM3*	ID	qPCR; Evaluate parents
403	Duplication	chr3:53,118,008-53,329,390	3p21.1	*DCP1A*, MIR1303, PRKCD, RFT1*, TKT*	*PRKCD, TKT*	Brain function	qPCR; Evaluate parents
412	Duplication	chr7:151,439,244-151,528,655	7q36.1	*GALNT11*, MLL3**	*MLL3*	Brain function	qPCR; Evaluate parents
421	Duplication	chr16:73,949,117-74,057,944	16q23.1	*CFDP1*, TMEM170A*	*CFDP1*	Craniofacial development function	qPCR; Evaluate parents
423	Duplication	chr1:61,470,344-61,899,438	1p31.3	*NFIA**	*NFIA*	Central nervous system function	aCGH; Evaluate parents
	Deletion	chr9:118,317,581-118,573,731	9q33.1	*ASTN2*, TRIM32*	*TRIM32*	Ubiquitin ligase E3; Obesity, ID and dysmorphic features	aCGH; Evaluate parents
443	Duplication	chr10:73,517,597-73,583,426	10q22.1	*ASCC1*, SPOCK2**	*ASCC1*	Component of the extra-cellular matrix highly expressed in brain	?
455	Deletion	chrX:122,924,947-122,981,041	Xq25	*STAG2**	*STAG2*	Interacts with cohesin proteins	qPCR; Evaluate parents
464	Duplication	chr3:48,439,971-48,549,239	3p21.31	*ATRIP, CCDC51, CCDC72, PFKFB4*, PLXNB1*, SHISA5, TREX1*	*PLXNB1*	Semaphorin receptor	qPCR; Evaluate parents
479	Duplication	chr2:106,632,846-106,910,178	2q12.2-q12.3	*ST6GAL2**	*ST6GAL2*	?	?
480	Duplication	chr2:182,411,985-182,561,778	2q31.3	*PPP1R1C*, SSFA2*	*PPP1R1C*	Cell cycle function	qPCR; Evaluate parents

515	Duplication	chr4:174,658,836-174,718,746	4q34.1	*HAND2, NBLA00301*	*HAND2*	Heart function	qPCR; Evaluate parents
563	Deletion	chr17:60,712-155,836	17p13.3	*RPH3AL**	*RPH3AL*	Exocytosis and Cytoskeleton function	qPCR; Evaluate parents
572	Deletion	chr7:1,998,754-2,047,386	7p22.3	*MAD1L1**	*MAD1L1*	Chromosomal instability	aCGH; Evaluate parents
	Duplication	chr16:10,638,423-10,751,290	16p13.13	*NUBP1*, TEKT5**	*TEKT5*	Structural component of microtubules	aCGH; Evaluate parents
575	Duplication	chr1:36,547,812-36,690,552	1p34.3	*C1orf113*, FAM176B, LSM10, OSCP1*, STK40*	Not determined	Not determined	aCGH; Evaluate parents
	Deletion	chr2:130,497,826-130,676,027	2q21.1	*CCDC74B, FAM128B, LOC440905*, POTEF, SMPD4, TUBA3E**	*TUBA3E, FAM128B*	Tubulin or tubulin association	aCGH; Evaluate parents
	Deletion	chr16:2,521,392-2,638,612	16p13.3	*CEMP*1, FLJ42627*, LOC652276, PDPK1*	*PDPK1*	Phosphorylation	aCGH; Evaluate parents
	Deletion	chr16:29,048,939-29,247,323	16p11.2	*RUNDC2C**	Not determined	Not determined	aCGH; Evaluate parents
607	Deletion	chr5:1,220,200-1,335,461	5p15.33	*SLC6A18, SLC6A19, TERT**	*SLC6A18, SLC6A19*	Neurotransmitter transporters	aCGH; Evaluate parents
	Deletion	chr22:49,210,215-49,329,334	22q13.33	*ADM2, LMF2, MIOX, NCAPH2, ODF3B, SAPS2, SBF1, SCO2, TYMP**	*SBF1*	Myotubulin family member	aCGH; Evaluate parents
658	Deletion	chr16:367,675-488,401	16p13.3	*DECR2, LOC100134368, NME4, RAB11FIP3*, TMEM8A**	*RAB11FIP3*	Rab GTPase	aCGH; Evaluate parents
787	Deletion	chr17:78,070,140-78,294,730	17q25.3	*FN3K*, FN3KRP, FOXK2*, RAB40B, WDR45L*	*FOXK2, RAB40B*	Transcription regulation and spontaneous occlusion of the circle of Willis	qPCR; Evaluate parents (?)
639	Duplication	chr6:90,479,080-90,629,922	6q15	*CASP8AP2*, MDN1**	Not determined	Not determined	qPCR; Evaluate parents

659	Duplication	chr5:177,335,291-177,389,093	5q35.3	*FAM153C*, PROP1*	PROP1	Transcription regulation	qPCR; Evaluate parents
669	Duplication	chr17:74,618,880-74,966,215	17q25.3	*HRNBP3**	HRNBP3	Neuronal function	qPCR; Evaluate parents
732	Duplication	chr18:75,287,551-75,353,034	18q23	*NFATC1**	NFATC1	Not determined	qPCR; Evaluate parents
743	Duplication	chr1:103,836,427-104,103,459	1p21.1	*AMY1A, AMY1B, AMY1C, AMY2A, AMY2B, LOC648740, RNPC3**	Not determined	Not determined	aCGH; Evaluate parents
	Duplication	chr17:45,956,061-46,132,442	17q21.33	*ABCC3, ANKRD40*, CACNA1G, EPN3, MYCBPAP*, SPATA20*	Not determined	Not determined	aCGH; Evaluate parents
	Duplication	chrX:1,333,567-1,382,414	Xp22.33	*CSF2RA**	Not determined	Not determined	aCGH; Evaluate parents
745	Duplication	chr4:92,412,048-92,542,935	4q22.1	*FAM190A**	*FAM190A*	Not determined	aCGH; Evaluate parents
	Duplication	chr7:154,904,327-154,995,882	7q36.3	*CNPY1*, EN2*	*EN2*	Autism	aCGH; Evaluate parents
	Duplication	chr22:29,558,944-29,617,330	22q12.2	*OSBP2**	*OSBP2*	Neuronal function	aCGH; Evaluate parents
763	Deletion	chrX:32,177,045-32,421,416	Xp21.1	*DMD**	*DMD*	Duchene muscular dystrophy	qPCR; Evaluate parents
776	Deletion	chr7:81,647,142-81,725,143	7q21.11	*CACNA2D1**	*CACNA2D1*	Not determined	qPCR; Evaluate parents

General discussion

We used aCGH to investigate DNA copy number imbalances in 130 patients with idiopathic ID. All the patients included in this study were previously studied for the routinely performed laboratory tests such as high-resolution G-banding karyotype, study of subtelomeric rearrangements (by MLPA or FISH), other fluorescence in situ hybridization studies when a specific syndrome was suspected, metabolic screening, and fragile X testing when clinically indicated (FRAXA and FRAXE).

Using the same platform and consensual workflow criteria, a total of 23 genomic imbalances that were very likely to account for the phenotype were found in 26 patients.

The distribution based on the size of the genomic alterations is described in table 2.35.

Table 2.35 – Number of alterations found distributed by their size.

Size of imbalance	Number of cases	Percentage of cases (%)
>10 Mb	1	4%
5-10 Mb	4	15%
1-5 Mb	10	39%
<1 Mb	11	42%
TOTAL	26	100%

Of all the alterations found in the 26 described patients, 40% were alterations below 1 Mb of size and would for sure be missed in a conventional karyotype.

If we consider only this set of patients, we achieved a diagnostic yield of 20% in patients that were extensively studied without success before. However, this number can even be increased with the determination of the clinical significance of the remaining 27 cases of less clear pathogenicity. It is also important to highlight that, when used as a routine diagnostic tool, the criteria for patient selection are less stringent and one might be able to achieve higher detection rates. Another important factor for the detection rate is the array resolution and design. The platform used here allows a mean resolution of 17 Kb, achieving resolutions as high as 11 Kb in regions known to be associated with development disorders/congenital abnormalities. This particular array (AMADID 023363) was designed by the LLC Dutch consortium, for its implementation as a first-tier diagnostic test in the study of children with ID and/or congenital abnormalities in the Netherlands. Consequently, the array is enriched in probes in regions known to be associated with ID/DD, allowing a higher detection rate

The largest alteration was found in patient 9 and has 10.2 Mb size. The most likely scenario is that this alteration would be found by conventional karyotyping in a retrospective analysis. This would

also be possible for the four alterations found between 5 and 10 Mb size. However, the majority of the chromosomal changes found in the study (80%) were below 5 Mb in size, a dimension already critical for detection by conventional karyotyping, revealing once again the large number of explainable cases missed by this technique.

When the patients carry de novo relatively large alterations, it is reasonable to assume that the imbalance is causing the disease. On the other hand, when the alteration or at least a very similar one is present in affected and non-affected members of the family, these findings have high impact for genetic counseling. This situation occurs for patients 24, 25 and 26. The alterations described for these 3 cases are present in regions of susceptibility for developmental disorders and were described to be present either in affected, mildly affected and healthy individuals [286], [296]. The presence of more affected individuals in the family raises the hypothesis of the alteration being present not only the parent, but also in other siblings, something that needs to be further analyzed (the case of susceptibility loci).

Still in the same topic, are the cases were the alteration (or at least a very similar one) is also found in one of the parents and a suspicion is raised if the progenitor is in fact unaffected. Many times the correct assessment of the carrier parent is only performed after the discovery of the alteration and not rarely it is found that the father or mother are not absolutely healthy but display a milder form of ID or behavior alterations. This possibility is present in some of the cases reported in this work (such as patient 7).

Other interesting alterations were found in patients 1 and 2, in whom the imbalance size seems to correlate with the severity of the disease. Although both patients have criteria for classification and *MECP2* duplication syndrome, patient 1 has three more genes covered by the duplication for which point mutations are known to be involved in X-linked ID. If the increased severity of the disease in patient 1 is due to the genomic imbalance of any of those genes individually, or to the combination of all of them remains unknown.

Only 6 patients presented alterations fully concordant with previous described syndromes in Decipher. The remaining 18 patients presented alterations not previously described at all or only partially overlapping with syndrome regions. These alterations may represent novel pathogenic genomic imbalances. Another important contribution of these new previously non-reported alterations is the identification of the genes encompassed by them, which raises the need to clarify their contribution for the etiology of ID.

In spite of the use of established consensus criteria for evaluation of the genomic alterations described, we should keep in mind that genotype-phenotype correlations can be challenging in some cases, such as when (I) the inheritance of the CNVs cannot be determined (in cases for which parents samples are not available, there is incomplete penetrance of an alteration, that can be present in both affected and unaffected individuals). Also, (II) a small alteration can be pathogenic even if it involves genes of unknown or apparently not relevant function, (III) an alteration can be pathogenic even if it doesn´t cover genes, (IV) a deletion can be pathogenic by

unmasking a recessive mutation in the other allele and (V) a CNV can be pathogenic only in combination with mutations in other genes [113]. All these aspects should be taken into account since, the correct interpretation of the results is crucial for the physician in order to perform a correct counseling and patient management.

Regarding the still inconclusive cases, special attention needs to be given to the duplication effects when this is not covering the totality of a gene. The most straightforward way of interpreting the duplications pathogenicity is when it encompasses the entire gene. In this case, the most likely way to cause the disease is if the affected gene(s) are dosage sensitive and either by their deregulation or downstream targets/interactors deregulation, leading to disease. On the other hand, whenever a duplication doesn´t cover the totality of the gene, a set of possible consequences needs to be confirmed. If located in tandem, the alteration may lead to the disruption of the gene (either by encoding a mutant protein that is degraded or has altered functions) causing a similar consequence to a deletion of the gene. The other possible outcome, is that if the duplicated segment is located somewhere else in the genome. In this case, it is less likely that the duplication could account for the disease, unless the duplicated genomic portion is able to encode for a protein that as noxious function in the same pathway as the original one, leading also to the disease. Either way, whenever a duplication that may disrupt a gene is found in a patient, FISH studies and possible mRNA analysis are required in order to better understand its functional consequences.

Other important aspect to take into account is that when a large rearrangement if found, usually dozens of genes are covered by the alteration but not all of them will be dosage sensitive and relate to the phenotype. An important aspect of the clinical interpretation of this type of genomic alterations is the look for the key gene(s) that may be causing the phenotypes. This type of search can be performed by two ways: (I) look for the described function for each of the genes involved in the alteration (a laborious and time consuming strategy that is subjected to human errors and knowledge of the molecular mechanisms behind the etiology of ID) or (II) by using bioinformatics tools that help in the prioritization the most interesting genes for each phenotypic alteration. By the analysis of either ways for each of the patients described, we believe that the most reliable method for avoiding false positive results is the use of both strategies together. Another important improvement that can be made to this approach is the use of more bioinformatics tools that not take into account only the co-expression patterns of the genes but also its function, tissue expression, species conservation and so on. This type of approach will allow not only the discovery of new genes that account for ID but also genes that are related with other phenotypic features often seen in these patients (such as congenital malformations and behavior alterations).

In the recent years, new conceptual insights have been acquired regarding the fundamental constitution and function of the genome elements [298, 299, 300]. The existence of long-range regulatory functional domains is for a long time supported by genomic functional studies and their importance

in the regulation of differential gene expression at several stages of development is undeniable [299]. Consequently, alterations in regions that don´t encompass genes may have severe consequences for the regulation of crucial elements for development and account as well for disease [298, 301].

For the described alterations (regarding patients 1-30) a total of 13 patients (43%) presented genomic imbalances that also covered miRNAs. Whenever a large genomic imbalance is found in a patient, special attention is given to the genes involved, namely their function, expression pattern and interacting proteins. However, the presence of miRNAs in those regions can also lead to the de-regulation of target mRNAs, crucial for disease. In spite of the advances in the clarification of miRNAs' role in DD spectrum disease, no significant conclusions can be made for diagnosis purposes so far regarding deletions/duplications of those genomic elements.

This work revealed a 20% yield of detection in patients with idiopathic ID and for which no explanation was possible to be given to the parents so far. With the integration in this study, the parents of at least 26 children will have now the possibility of genetic counseling. Besides the fact that for a set of patients the ID could finally be explained, a larger experience was taken from this work. This study revealed the importance of the parent's analysis in order to conclude about the clinical significance of the alterations found and that copy number variations in regions not previously described as imbalance do not invariably imply pathogenicity. Conversely, in the diagnostic service front, this technique also requires caution when asserting that a CNV is not pathogenic just because is reported in healthy parents. The parents should always be further evaluated in this situation in order to search for undetermined or minimal clinical symptoms. The genomic imbalance may also have an incomplete penetrance or milder expression of in the parents (or even in very few control individuals), and be expressed in a more severe way in the child due to a more complex mechanism (a variation in other allele or even another CNVs that also predispose to the disease). Indeed, false positive results that require parental testing comprise the weaknesses of this technique since they increase the costs of the test in a 3x fold.
Another important aspect that may lead to limitations in the interpretation of the results is the lack of a CNVs study in control Portuguese cases. Since the genetic background of the Portuguese population is different from the one present in the DGV database and the one constituted by the Dutch parents, there may be CNVs in the Portuguese cases that, despite being benign and have relatively high recurrence rates, are not described in the other databases and may lead to false/dubious alterations among our patients.

Despite all the limitations still found by the users of aCGH as a diagnostic tool of children with ID, this approach is the currently recommended by international guidelines and is already used as a first-tier test in the USA and several European countries [114, 148, 302]. We believe that this pioneer study in Portugal gave a crucial contribute to the implementation of this technique as a first test used in the diagnostic of children with ID/DD and congenital abnormalities, either by the technical

formation as well as by the preparation and self-learning skills acquired regarding the interpretation of the results.

Final remarks

In the last 40 years conventional karyotyping has been the most used methodology to study human chromosomes, allowing increasingly subtle abnormalities of the human genome to be discovered and linked to specific diseases. The improvements in the conventional cytogenetics (such as higher resolution and different FISH approaches) quickly proved to be insufficient and soon the need was felt for a technique that allowed the possibility to identify genomic imbalances across the entire genome at a much higher resolution and less labour-intensivily than conventional karyotyping methodologies. Array CGH appeared as a molecular cytogenetic approach designed to provide genome-wide screening of the amplification and deletion of specific chromosome regions.

The impact of aCGH in the practice of Medical Genetics in the recent years has been transformative and a even better understanding of the full clinical spectrum of these disorders will be achieved as the use of array CGH in the clinical becomes more prevalent and as correlations of these clinical findings with the genomic lesions are increasingly made. The use of aCGH in Clinical Genetics has allowed researchers and clinicians to (I) expand the phenotype of previous existing conditions, (II) identify reciprocal products of known conditions, (III) determine the genomic alterations in known conditions and (IV) discover many new syndromes. In addition to cancer and neurological conditions, there is a clear indication that constitutional chromosomal alterations detected by aCGH may affect many organ systems and that CNVs may play important roles in many multifactorial conditions, such as Crohn´s disease, immunodeficiency disorders, psoriasis and isolated congenital heart diseases [303, 304, 305, 306].

Although our understanding of CNVs and their role in health and disease is improving, it is still common to identify alterations of unclear clinical significance that cause considerable interpretive difficulties for the laboratory and create significant counseling dilemmas for the clinician.

Contemporary cytogenetics has already integrated array technologies into clinical use in many countries around the world. Well-designed arrays for clinical use allow easy interpretation and are likely to provide diagnosis in a substantial number of currently undiagnosed cases of human genetic disorders. Molecular karyotyping has already transformed the practice of both Medical Genetics and Cytogenetics laboratories ushering in the new era of genomic medicine.

Our main proposal with this work is to highlight the advantages of aCGH technology for the diagnosis of ID patients and its adoption as the first-tier test to be used in Portugal for this condition.

References

1. Muller RF, Young ID. *Emery´s Elements of Medical Genetics*. 10th ed. Churchill Livingstone; 2002.

2. Stankiewicz P, Lupski JR. Structural Variation in the Human Genome and its Role in Disease. *Annual Review of Medicine*. 2010.

3. Kumar D. Disorders of the genome architecture : a review. *Journal of Medical Genetics*. 2009;(2008):69-76.

4. Ae O, Ga M, Microarray-based YB. Microarray-based comparative genomic hybridization and its applications in human genetics. *American Journal of Human Genetics, The*. 2004;(4):488-495.

5. Patel DR, Greydanus DE, Jr JLC, Pratt HD. Developmental Disabilities Across the Lifespan. *YMDA*. 2010;56(6):305-397. Available at: http://dx.doi.org/10.1016/j.disamonth.2010.02.001.

6. Heller T, Factor A. Family support and intergenerational caregiving: report from the State of the Science in Aging with Developmental Disabilities Conference. *Disability and health journal*. 2008;1(3):131-5.

7. WHO. The ICD-10 Classification of Mental and Behavioural Disorders. *International Classification of Mental and Behavioural Disorders: Clinical description and diagnositic guidelines*.1-267.

8. Raymond FL, Tarpey P. The genetics of mental retardation. *Review Literature And Arts Of The Americas*. 2006;15(2):110-116.

9. Strømme P. Prevalence of psychiatric diagnoses in children with mental retardation : data from a population-based study. *Mental Retardation*. 2000:266-270.

10. Raymond FL. X linked mental retardation: a clinical guide. *Journal of Medical Genetics*. 2006;(November 2008).

11. Castellvı S, Mila M. Genes Responsible for Nonspecific Mental Retardation. *Molecular Genetics and Metabolism*. 2001;108:104 -108.

12. Lespinasse J, Gimelli S, Paoloni-giacobino A. European Journal of Medical Genetics Characterization of an interstitial deletion 6q13 – q14 . 1 in a female with mild mental retardation , language delay and minor dysmorphisms. *European Journal of Medical Genetics*. 2009;52:49-52.

13. Brouwer APMD, Yntema HG, Kleefstra T, et al. Mutation Frequencies of X-linked Mental Retardation Genes in Families from the EuroMRX Consortium. *Human Mutation*. 2007;953.

14. Chelly J. Breakthroughs in molecular and cellular mechanisms underlying X-linked mental retardation. *Review Literature And Arts Of The Americas*. 1999;8(10):1833-1838.

15. Chelly J. MRX Review. *American Journal of Medical Genetics*. 2000;366:364-366.

16. Veltman JA. Genomic microarrays in mental retardation : from CNV to gene , from research to diagnosis. *Journal of Medical Genetics*. 2009.

17. Cle OLA. Screening of ARX in mental retardation families : consequences for the strategy of molecular diagnosis. *October*. 2006:39-46.

18. Grønskov K, Hallberg A, Brøndum-nielsen K. Mutational analysis of the FMR1 gene in 118 mentally retarded males suspected of fragile X syndrome : absence of prevalent mutations. *Human Molecular Genetics*. 1998;2:440-445.

19. Patsalis PC, Sismani C, Hettinger JA, et al. Molecular Screening of Fragile X (FRAXA) and FRAXE Mental Retardation Syndromes in the Hellenic Population of Greece and Cyprus : Incidence , Genetic Variation , and Stability. *American Journal of Medical Genetics*. 1999;190(January):184-190.

20. Ropers H-hilger, Hoeltzenbein M, Kalscheuer V, et al. Nonsyndromic X-linked mental retardation : where are the missing mutations ? *Trends in genetics*. 2003;19(6):316-320.

21. Vaillend C, Poirier R, Laroche S. Genes , plasticity and mental retardation. *Behavioural Brain Research*. 2008;192:88-105.

22. Skovgaard AM. Mental health problems and psychopathology in infancy and early childhood. An epidemiological srudy. Dan Med Bull. 2010 Oct; 57(10):B4193

23. Chiurazzi P, Schwartz CE, Gecz J, Neri G. REVIEW XLMR genes : update 2007. *European Journal of Human Genetics*. 2008:422-434.

24. Haydar TF . Advanced microscopic imaging methods to investigate cortical development and the etiology of mental retardation. *Ment Retard Dev disabil Rev*. 2005;11(4):303-16.

25. Pomeroy SL, Kim JYH. Biology and Pathobiology of Neuronal development. *Mental Retardation and Developmental Disabilities Research Reviews*. 2000;46:41-46.

26. Diaz AL, Gleeson JG. The molecular and genetic mechanisms of neocortex development. *Neurosciences*. 2010;36(3):503-512.

27. Clowry G, Molnár Z, Rakic P. Renewed focus on the developing human neocortex. *Journal of anatomy*. 2010;217(4):276-88.

28. Price DJ, Kennedy H, Dehay C, et al. The development of cortical connections. *European J Neurosci*. 2007;23(4):910-20.

29. Crino PB, Eberwine J. Cellular and molecular basis of cerebral dysgenesis. *Journal of Neuroscience Research*. 1997;50(6):907-916.

30. Leventer RJ, Mills PL, Dobyns WB. X-linked malformations of cortical development. *American journal of medical genetics*. 2000;97(3):213-20.

31. Guerrini R, Parrini E. Neuronal migration disorders. *Neurobiology of disease*. 2010;38(2):154-66.

32. Guerreiro MM. Malformations of cortical development. *Arquivos de neuro-psiquiatria*. 2009;67(2B):570-4.

33. Golden J a. Cell migration and cerebral cortical development. *Neuropathology and applied neurobiology*. 2001;27(1):22-8.

34. Adachi Y, Poduri a, Kawaguch a, et al. Congenital Microcephaly with a Simplified Gyral Pattern: Associated Findings and Their Significance. *AJNR. American journal of neuroradiology*. 2011:1-7.

35. Squier W, Jansen A. Abnormal development of the human cerebral cortex. *Journal of anatomy*. 2010;217(4):312-23..

36. Williams CA, Dagli A, Battaglia A. Genetic Disorders Associated With Macrocephaly. *Brain*. 2008;2037:2023 - 2037.

37. Verrotti A, Spalice A, Ursitti F, et al. New trends in neuronal migration disorders. *European journal of paediatric neurology : EJPN : official journal of the European Paediatric Neurology Society*. 2010;14(1):1-12.

38. Wynshaw-Boris A, Pramparo T, Youn YH, Hirotsune S. Lissencephaly: mechanistic insights from animal models and potential therapeutic strategies. *Seminars in cell & developmental biology*. 2010;21(8):823-30.

39. Nijholt IM. Molecular mechanisms of synaptic plasticity: implications for immediate early gene expressionand learning. *Cancer research*. 2002;68(24):192.

40. Shen K, Cowan CW. Guidance molecules in synapse formation and plasticity. *Cold Spring Harbor perspectives in biology*. 2010;2(4):a001842.

41. Chechik G, Meilijson I, Ruppin E. Neuronal regulation: A mechanism for synaptic pruning during brain maturation. *Neural computation*. 1999;11(8):2061-80.

42. White T, Su S, Schmidt M, Kao C-yen, Sapiro G. Brain and Cognition The development of gyrification in childhood and adolescence. *Brain and Cognition*. 2010;72(1):36-45.

43. Henriques-coelho T, Fortunato JS, Leite-moreira AF. Sinaptogénese : do cone de crescimento à sinapse neuromuscular. *Sinapse*. 2001;3:95-121.

44. Hotulainen P, Hoogenraad CC. Actin in dendritic spines: connecting dynamics to function. *The Journal of Cell Biology*. 2010;189(4):619-629.

45. Nishimura YV, Sekine K, Chihama K, et al. Dissecting the factors involved in the locomotion mode of neuronal migration in the developing cerebral cortex. *The Journal of biological chemistry*. 2010;285(8):5878-87.

46. Ma L, Wang DD, Zhang TY, et al. Region-specific involvement of BDNF secretion and synthesis in conditioned taste aversion memory formation. *J Neurosci*. 2011;31(6):2079-90.

47. Egea J, Klein R. Bidirectional Eph-ephrin signaling during axon guidance. *Trends in cell biology*. 2007;17(5):230-8.

48. Yang GY, Luo ZG. Implication of Wnt signaling in neuronal polarization. *Dev Neurobiol*. 2011;71(6):495-507.

49. Maier V, Jolicoeur C, Rayburn H, et al. Semaphorin 4C and 4G are ligands of Plexin-B2 required in cerebellar development. *Mol Cell Neurosci*. 2011;46(2):419-31.

50. Zagrebelsky M. Nogo-A stabilizes the architecture of hippocampal neurons. *J Neurosci*. 2010;30(40):13220-34.

51. Anon. Synapses , Receptor Targeting Neuronal synapses Heterogeneity within a single postsynaptic cell. *Receptor*.

52. Spronsen MV, Hoogenraad CC. Synapse Pathology in Psychiatric and Neurologic Disease. *Organization*. 2010;(March):207-214.

53. Lin RC, Scheller RH. Mechanisms of synaptic vesicle exocytosis. *Annu. Rev. Cell Dev. Biol.* 2000;16:19-49.

54. Chechlacz M. Is mental retardation a defect of synapse structure and function? *Pediatric Neurology*. 2003;29(1):11-17.

55. Humeau Y, Gambino F, Chelly J, Vitale N. X-linked mental retardation: focus on synaptic function and plasticity. *Journal of neurochemistry*. 2009;109(1):1-14.

56. Bohlen OV. Dendritic spine abnormalities in mental retardation. *Cell and Tissue Research*. 2010:317-323.

57. Svitkina T, Lin W-H, Webb DJ, et al. Regulation of the Postsynaptic Cytoskeleton: Roles in Development, Plasticity, and Disorders. *The Journal of neuroscience : the official journal of the Society for Neuroscience*. 2010;30(45):14937-14942.

58. Nadif Kasri N, Van Aelst L. Rho-linked genes and neurological disorders. *Pflügers Archiv : European journal of physiology*. 2008;455(5):787-97.

59. Hamdan FF, Piton A, Gauthier J, et al. De novo *STXBP1* mutations in mental retardation and nonsyndromic epilepsy. *Annals of Neurology*. 2009;65(6):748-53.

60. Bienvenu, T., Des Portes, V., Saint Martin, A., *et a*l. Non-specific X-linked semidominant mental retardation by mutations in a Rab GDP-dissociation inhibitor. *Hum. Molec. Genet*. 1998;(7):1311-1315.

61. D'Adamo, P., Gulisano, M., Oostra, B. A., Chelly, J., Toniolo D. GDI is responsible for X-linked mental retardation. *Am. J. Hum. Genet*. 1997;61.

62. Garcia CC, Blair HJ, Seager M, Coulthard A, Tennant S, Buddles M, Curtis A GJ. Identification of a mutation in synapsin I, a synaptic vesicle protein, in a family with epilepsy. *J Med Genet*. 2004;41(3)(183-6).

63. Battini, R., Chilosi, A., Mei, D., Casarano, M., Alessandri, M. G., Leuzzi, V., Ferretti, G., Tosetti, M., Bianchi, M. C., Cioni G. Mental retardation and verbal dyspraxia in a new patient with de novo creatine transporter (*SLC6A8*) mutation. *Am. J. Med. Genet*. 2007;143A:1771-1774.

64. Bizzi, A., Bugiani, M., Salomons, G. S., Hunneman, D. H., Moroni, I., Estienne, M., Danesi, U., Jakobs, C., Uziel G. X-linked creatine deficiency syndrome: a novel mutation in creatine transporter gene SLC6A8. *Ann. Neurol*. 2002;52:227-231.

65. Tarpey P, Parnau J, Blow M, *et al*. Mutations in the DLG3 gene cause nonsyndromic X-linked mental retardation. *Am J Hum Genet*. 2004;75(2)(318-24).

66. Gauthier J, Siddiqui TJ, Huashan P, et al. Truncating mutations in NRXN2 and NRXN1 in autism spectrum disorders and schizophrenia. *Human genetics*. 2011..

67. Wu, Y., Arai, A. C., Rumbaugh, G., Srivastava, A. K., Turner, G., Hayashi, T., Suzuki, E., Jiang, Y., Zhang, L., Rodriguez, J., Boyle, J., Tarpey P. Mutations in ionotropic AMPA receptor 3 alter channel properties and are associated with moderate cognitive impairment in humans. *Proc. Nat. Acad. Sci.* 2007;104:18163-18168.

68. Durand, C. M., Betancur, C., Boeckers, T. M., Bockmann, J., Chaste, P., Fauchereau, F., Nygren, G., Rastam, M., Gillberg, I. C., Anckarsater, H., Sponheim, E., Goubran-Botros H. Mutations in the gene encoding the synaptic scaffolding protein SHANK3 are associated with autism spectrum disorders. *Nature genetics*. 2007;39:25-27.

69. Berkel, S., Marshall, C. R., Weiss, B., Howe, J., Roeth, R., Moog, U., Endris, V., Roberts, W., Szatmari, P., Pinto, D., Bonin, M., Riess, A., Engels, H., Sprengel, R., Scherer, S. W., Rappold GA. Mutations in the SHANK2 synaptic scaffolding gene in autism spectrum disorder and mental retardation. *Nature Genet.* 2010;42:489-491.

70. Melko M, Bardoni B. The role of G-quadruplex in RNA metabolism: involvement of FMRP and FMR2P. *Biochimie*. 2010;92(8):919-26.

71. Lagali PS, Corcoran CP, Picketts DJ. Hippocampus development and function: role of epigenetic factors and implications for cognitive disease. *Clinical genetics*. 2010;78(4):321-33.

72. Zukin RS, Richter JD, Bagni C. Signals, synapses, and synthesis: how new proteins control plasticity. *Frontiers in neural circuits*. 2009;3(October):14.

73. Hansen KF, Sakamoto K, Wayman G a, Impey S, Obrietan K. Transgenic miR132 Alters Neuronal Spine Density and Impairs Novel Object Recognition Memory. *PloS one*. 2010;5(11):e15497.

74. Corbin R, Olsson-Carter K, Slack F. The role of microRNAs in synaptic development and function. *BMB reports*. 2009;42(3):131-5.

75. Yujing Li, Li Lin APJ. The MicroRNA Pathway and Fragile X Mental Retardation Protein. *Biochim Biophys Acta*. 2008;11(4):702-705.

76. Wu H, Tao J, Chen PJ, et al. Genome-wide analysis reveals methyl-CpG-binding protein 2-dependent regulation of microRNAs in a mouse model of Rett syndrome. *PNAS*. 2010;107(42).

77. Djagaeva I, Doronkin S. Dual regulation of dendritic morphogenesis in Drosophila by the COP9 signalosome. *PloS one*. 2009;4(10):e7577.

78. Tarpey PS, Raymond FL, Meara SO, et al. Mutations in CUL4B , Which Encodes a Ubiquitin E3 Ligase Subunit , Cause an X-linked Mental Retardation Syndrome Associated with Aggressive Outbursts , Seizures , Relative Macrocephaly , Central Obesity , Hypogonadism , Pes Cavus , and Tremor. *Journal of Human Genetics*. 2007;80(February):345-352.

79. Zou, Y., Liu, Q., Chen, B., Zhang, X., Guo, C., Zhou, H., Li, J., Gao, G., Guo, Y., Yan, C., Wei, J., Shao, C., Gong Y. Mutation in CUL4B, which encodes a member of cullin-RING ubiquitin ligase complex, causes X-linked mental retardation. *Am. J. Hum. Genet*. 2007;80:561-566.

80. Froyen, G., Corbett, M., Vandewalle, J., Jarvela, I., Lawrence, O., Meldrum, C., Bauters, M., Govaerts, K., Vandeleur, L., Van Esch, H., Chelly, J., Sanlaville D. Submicroscopic duplications of the hydroxysteroid dehydrogenase HSD17B10 and the E3 ubiquitin ligase HUWE1 are associated with mental retardation. *Am. J. Hum. Genet*. 2008;82:432-443.

81. Hegde A. The ubiquitin-proteasome pathway and synaptic plasticity. *Learn Mem*. 2010;17(7):314-327.

82. Huttenlocher PR. Dendritic development and mental defect. *Neurology*. 1970;20(4):381.

83. Allan L, Karen L, Brøndum-nielsen BN-pedersen K. High-throughput analysis of Fragile X (CGG) n alleles in the normal and premutation range by PCR amplification and automated capillary electrophoresis. *Cell*. 1997:564-568.

84. Larsen LA, Armstrong JSM, Grønskov K, et al. Haplotype and AGG-Interspersion Analysis of FMR1 (CGG) n Alleles in the Danish Population : Implications for Multiple Mutational Pathways Towards Fragile X Alleles. *American Journal of Medical Genetics*. 2000;106(March):99-106.

85. Fatemi SH, Folsom TD. The Role of Fragile X Mental Retardation Protein in Major Mental Disorders. *Neuropharmacology*. 2010;(November):1-6.

86. Krueger DD, Bear MF. Toward Fulfilling the Promise of Molecular Medicine in Fragile X Syndrome. *Annual review of medicine*. 2010;(November 2010):1-19.

87. Johnston MV, Jeon OH, Pevsner J, Blue ME, Naidu S. Neurobiology of Rett syndrome: a genetic disorder of synapse development. *Brain & development*. 2001;23 Suppl 1:S206-13.

88. Schule B, Armstrong D. Severe congenital encephalopathy caused by MECP2 null mutations in males: central hypoxia and reduced neuronal dendritic structure. *Clinical Genetics*. 2008;74(2):116-126.

89. Johnston M. Clinical disorders of brain plasticity. *Brain and Development*. 2004;26(2):73-80.

90. Dityatev A, Rusakov D a. Molecular signals of plasticity at the tetrapartite synapse. *Current opinion in neurobiology*. 2011:1-7.

91. Dityatev A, Schachner M, Sonderegger P. The dual role of the extracellular matrix in synaptic plasticity and homeostasis. *Nature reviews. Neuroscience*. 2010.

92. Dityatev A, Seidenbecher CI, Schachner M. Compartmentalization from the outside: the extracellular matrix and functional microdomains in the brain. *Trends in neurosciences*. 2010.

93. Eroglu C, Barres BA, Stevens B. Glia as Active Participants in the Development and Function of Synapses. *Development*. 2008.

94. Buysse K. Molecular karyotyping: a powerful tool for the study of genomic defects in patients with mental retardation. PhD thesis. Ghent University, Faculty of medicine and health sciences, Center for Medical Genetics. Belgium, 2009

95. Bae JS, Cheong HS, Park BL, et al. Genome-wide profiling of structural genomic variations in Korean HapMap individuals. *PloS one*. 2010;5(7):e11417.

96. Feuk L. Inversion variants in the human genome : role in disease and genome architecture. *Genetics*. 2010.

97. Perry GH, Dominy NJ, Claw KG, et al. variation. *October*. 2008;39(10):1256-1260.

98. Redon R, Ishikawa S, Fitch KR, et al. Global variation in copy number in the human genome. *Nature*. 2006;444(7118):444-54.

99. Li J, Yang T, Wang L, et al. Whole Genome Distribution and Ethnic Differentiation of Copy Number Variation in Caucasian and Asian Populations. *Distribution*. 2009;4(11):1-7.

100. Zhang F, Gu W, Hurles ME, Lupski JR. Copy number variation in human health, disease, and evolution. *Annual review of genomics and human genetics*. 2009;10:451-81.

101. Schuster-Böckler B, Conrad D, Bateman A. Dosage sensitivity shapes the evolution of copy-number varied regions. *PloS one*. 2010;5(3):e9474.

102. Smith AJD, Tsalenko A, Sampas N, et al. Array CGH analysis of copy number variation identifies 1284 new genes variant in healthy white males : implications for association studies of complex diseases. *Access*. 2007;16(23):2783-2794.

103. Zepeda-mendoza CJ, Lemus T, Yáñez O, et al. Identical repeated backbone of the human genome. *BMC Genomics*. 2010.

104. Merikangas AK, Corvin AP, Gallagher L. Copy-number variants in neurodevelopmental disorders : promises and challenges. *Trends in genetics*. 2009.

105. Lee JA, Carvalho CMB, Lupski JR. A DNA Replication Mechanism for Generating Nonrecurrent Rearrangements Associated with Genomic Disorders. *Cell*. 2007:1235-1247.

106. Zhang F, Khajavi M, Connolly AM, et al. The DNA replication FoSTeS / MMBIR mechanism can generate genomic , genic and exonic complex rearrangements in humans. *Nature Publishing Group*. 2009;41(7):849-853.

107. Conrad DF, Pinto D, Redon R, et al. Origins and functional impact of copy number variation in the human genome. *Nature*. 2010;464(7289):704-712.

108. Ou Z, Stankiewicz P, Xia Z, et al. Observation and prediction of recurrent human translocations mediated by NAHR between nonhomologous chromosomes. *Genome Research*. 2011;2:33-46.

109. Lee JA, Carvalho CM LJ. A DNA replication mechanism for generating nonrecurrent rearrangements associated with genomic disorders. *Cell*. 2007;1312(7):1235-1247.

110. Gökçümen Ö, Lee C. Copy number variants (CNVs) in primate species using array-based comparative genomic hybridization. *Methods*. 2009;49(1):18-25.

111. Mkrtchyan H, Gross M, Hinreiner S, et al. The human genome puzzle - the role of copy number variation in somatic mosaicism. *Current genomics*. 2010;11(6):426-31..

112. Bishop DVM. Which neurodevelopmental disorders get researched and why? *PloS one*. 2010;5(11):e15112.

113. Hehir-kwa JY, Wieskamp N, Webber C, et al. Accurate Distinction of Pathogenic from Benign CNVs in Mental Retardation. *PLoS Computational Biology*. 2010;6(4).

114. Miller DT, Adam MP, Aradhya S, et al. Consensus Statement : Chromosomal Microarray Is a First-Tier Clinical Diagnostic Test for Individuals with Developmental Disabilities or Congenital Anomalies. *The American Journal of Human Genetics*. 2010;86(5):749-764.

115. Article R. New Horizons of Genomic in situ Hybridization. *Cytogenetic and Genome Research*. 2010:368-375.

116. Shaffer LG, Bejjani BA. A cytogeneticist ' s perspective on genomic microarrays. *Human Reproduction*. 2004;10(3):221-226.

117. Universit F, April BR. A Rapid Banding Technique for Routine Use in Human and Comparative Cytogenetics. *Humangenetik*. 1972;353:349-353.

118. Devi J, Ko JM, Seo BB. FISH and GISH : Modern cytogenetic techniques. *Indian Journal of Biotechnology*. 2005;4(July):307-315.

119. Dória S, Lima V, Carvalho B, et al. Application of touch FISH in the study of mosaic tetraploidy and maternal cell contamination in pregnancy losses. *Genome*. 2010:657-662.

120. Liehr T, Starke H, Weise A, Lehrer H, Claussen U. Multicolor FISH probe sets and their applications. *Cancer*. 2004:229-237.

121. Shaffer LG, Lupski JR. CHROMOSOMAL REARRANGEMENTS IN HUMANS. 2000:297-329.

122. Yunis J. High resolution of human chromosomes. *Science*. 1976;191(4233):1268-70.

123. Ragoussis J. Genotyping Technologies for Genetic Research. *Annual Review of Genomics and Human Genetics*. 2009.

124. Ravel TJLD, Devriendt K. What ' s new in karyotyping ? The move towards array comparative genomic hybridisation (CGH). *European Journal of Pediatrics*. 2007:637-643.

125. Bogdanowicz J, Pawłowska B, Ilnicka a, et al. Subtelomeric rearrangements in Polish subjects with intellectual disability and dysmorphic features. *Journal of applied genetics*. 2010;51(2):215-7.

126. Kearney L. Molecular cytogenetics. *Best Practice & Research Clinical Haematology*. 2001;14(3):645-668.

127. Hutchins GG, Grabsch HI. Molecular pathologu – the future? *The Surgeon*. 2009;7(6):366-377.

128. Shaffer LG, Theisen A, Bejjani B a, et al. The discovery of microdeletion syndromes in the post-genomic era: review of the methodology and characterization of a new 1q41q42 microdeletion syndrome. *Genetics in medicine : official journal of the American College of Medical Genetics*. 2007;9(9):607-16.

129. Rooms L, Reyniers E, Wuyts W, et al. Multiplex ligation-dependent probe amplification to detect subtelomeric rearrangements in routine diagnostics. *Clinical Genetics*. 2006:58-64.

130. Abeid L. Subtelomeric Microduplications in Three Sisters with Moderate Mental Retardation. *Biochemical Genetics*. 2010:909-914.

131. Yue Y, Grossmann B, Holder SE, Haaf T. De novo t(7;10)(q33;q23) translocation and closely juxtaposed microdeletion in a patient with macrocephaly and developmental delay. *Human genetics*. 2005;117(1):1-8.

132. Saugier-veber P, Goldenberg A, Drouot N, et al. Simple detection of genomic microdeletions and microduplications using QMPSF in patients with idiopathic mental retardation. *European Journal of Human Genetics*. 2006:1009-1017.

133. Bejjani B a, Shaffer LG. Clinical Utility of Contemporary Molecular Cytogenetics. *Annual Review of Genomics and Human Genetics*. 2008;9(1):71-86.

134. Koolen D A. Clinical and molecular delineation of the 17q21.31 microdeletion syndrome. *J. Med. Genet*. 2010: 45:710-720

135. Shaffer LG, Bejjani B a. A cytogeneticist's perspective on genomic microarrays. *Human reproduction update*. 2004;10(3):221-6.

136. Heller MJ. DNA Microarray Technologies: Devices, Systems and Applications. *Annu. Rev. Biomed. Eng.* 2002;4:129-53.

137. Albertson DG, Pinkel D. Genomic microarrays in human genetic disease and cancer. *Review Literature And Arts Of The Americas*. 2003;12(2):145-152.

138. Ylstra B, Ijssel PVD, Carvalho B, Brakenhoff RH, Meijer GA. SURVEY AND SUMMARY BAC to the future ! or oligonucleotides : a perspective for micro array comparative genomic hybridization (array CGH). *Complexity*. 2006;34(2):445-450.

139. Neill NJ, Torchia BS, Bejjani BA, Shaffer LG, Ballif BC. Comparative analysis of copy number detection by whole-genome BAC and oligonucleotide array CGH. *Cytogenetics*. 2010:1-14.

140. Tijssen M, Vosse SJ, Meijer GA, Ylstra B, Buffart TE. Across Array Comparative Genomic Hybridization : A Strategy to Reduce Reference Channel Hybridizations. *Cancer*. 2008;1004(May):994-1004.

141. Wieringen WNV, Wiel MAVD, Ylstra B. Normalized, Segmented or Called aCGH Data? *Cancer Informatics*. 2007:321-327.

142. Wiel MAVD, Brosens R, Eilers PHC, et al. Smoothing waves in array CGH tumor profiles. *Bioinformatics*. 2009;25(9):1099-1104.

143. Wiel MAVD, Picard F, Wieringen WNV. Preprocessing and downstream analysis of microarray DNA copy number profiles. *Briefings in Bioinformatics*. 2010.

144. Gai X, Perin JC, Murphy K, et al. CNV Workshop : an integrated platform for high- throughput copy number variation discovery and clinical diagnostics. *BMC Bioinformatics*. 2010:1-9.

145. Galasso C, Lo-Castro A, El-Malhany N, Curatolo P. "Idiopathic" mental retardation and new chromosomal abnormalities. *Italian journal of pediatrics*. 2010;36(1):17.

146. Vissers LELM, Vries BB a de, Veltman J a. Genomic microarrays in mental retardation: from copy number variation to gene, from research to diagnosis. *Journal of medical genetics*. 2010;47(5):289-97.

147. Ropers HH. Genetics of early onset cognitive impairment. *Annual review of genomics and human genetics*. 2010;11(X):161-87.

148. Manning M, Facmg MS, Hudgins L. Array-based technology and recommendations for utilization in medical ge- netics practice for detection of chromosomal abnormalities. *Genetics in Medicine*.6-9.

149. Salman M, jhanwar SC, ostrer H. Will the new cytogenetics replace the old cytogenetics ? *Clinical Genetics*. 2004;(Table 1):265-275.

150. Metzker ML. Sequencing technologies — the next generation. *Nature Reviews Genetics*. 2009;11(1):31-46.

151. Schuster SC. Next-generation sequencing transforms today ' s biology. *Most*. 2008;5(1):16-18.

152. Heller M. DNA Microarrays technologies: devices, systems and applications. Annu. Rev. Biomed. Eng. 2002.

153. Jovanovic L, Delahunt B, Mciver B, Eberhardt NL, Grebe SKG. Thyroid Gland Clonality Revisited. *Journal of Clinical Endocrinology & Metabolism*. 2003:3284-3291.

154. Edelmann L, Hirschhorn K. Clinical utility of array CGH for the detection of chromosomal imbalances associated with mental retardation and multiple congenital anomalies. *Annals of the New York Academy of Sciences*. 2009;1151:157-66.

155. Carvalho, C. M. B., Zhang, F., Liu, P., Patel, A., Sahoo, T., Bacino, C. A., Shaw, C., Peacock, S., Pursley, A., Tavyev, Y. J., Ramocki, M. B., Nawara, M., Obersztyn, E., Vianna-Morgante, A. M., Stankiewicz, P., Zoghbi, H. Y., Cheung, S. W., Lupski JR. Complex rearrangements in patients with duplications of *MECP2* can occur by fork stalling and template switching. *Hum. Molec. Genet*. 2009;18:2188-2203.

156. Giannandrea, M., Bianchi, V., Mignogna, M. L., Sirri, A., Carrabino, S., D'Elia, E., Vecellio, M., Russo, S., Cogliati, F., Larizza, L., Ropers, H.-H., Tzschach A. Mutations in the small GTPase gene RAB39B are responsible for X-linked mental retardation associated with autism, epilepsy, and macrocephaly. *Am. J. Hum. Genet*. 2010;86:185-195.

157. Rosenberg, E. H., Almeida, L. S., Kleefstra, T., DeGrauw, R. S., Yntema, H. G., Bahi, N., Moraine, C., Ropers, H.-H., Fryns, J.-P., DeGrauw, T. J., Jakobs, C., Salomons, G. SRosenberg, E. H., Almeida, L. S., Kleefstra, T., DeGrauw, R. S., Yntema, H. G., B GS. High prevalence of SLC6A8 deficiency in X-linked mental retardation. *Am. J. Hum. Genet*. 2004;75:97-105.

158. Braissant O, Henry H, Béard E UJ. Creatine deficiency syndromes and the importance of creatine synthesis in the brain. *Amino Acids*. 2011;40(5):1315-24.

159. Vandewalle, J., Van Esch, H., Govaerts, K., Verbeeck, J., Zweir, C., Madrigal, I., Mila, M., Pijkels, E., Fernandez, I., Kohlase, J., Spaich, C., Rauch, A., Fryns, J.-P., Marynen, P., Froyen G. Dosage-dependent severity of the phenotype in patients with mental retardation due to a recurrent copy-number gain at Xq28 mediated by an unusual recombination. *Am. J. Hum. Genet.* 2009;85:809-822.

160. Fox, J. W., Lamperti, E. D., Eksioglu, Y. Z., Hong, S. E., Feng, Y., Graham, D. A., Scheffer, I. E., Dobyns, W. B., Hirsch, B. A., Radtke, R. A., Berkovic, S. F., Huttenlocher, P. R., Walsh CA. Mutations in filamin 1 prevent migration of cerebral cortical neurons in human periventricular heterotopia. *Neuron*. 1998;21:1315-1325.

161. Parrini E, Rivas IL, Toral JF, Pucatti D, Giglio S, Mei D GR. In-frame deletion in FLNA causing familial periventricular heterotopia with skeletal dysplasia in males. *Am J Med Genet A*. 2011;27(3-4):207-16.

162. Serrano EM, Ricofort RD, Zuo J, Ochotny N, Manolson MF HL. Regulation of vacuolar H(+)-ATPase in microglia by RANKL. *Biochem Biophys Res Commun*. 2009;389(1):193-7.

163. Hedera, P., Alvarado, D., Beydoun, A., Fink JK. Novel mental retardation-epilepsy syndrome linked to Xp21.1-p11.4. *Ann. Neurol*. 2002;51:45-50.

164. Pehlivan T, Pober BR, Brueckner M, et al. *GATA4* haploinsufficiency in patients with interstitial deletion of chromosome region 8p23.1 and congenital heart disease. *American journal of medical genetics*. 1999;83(3):201-6.

165. Arceci RJ, King AA, Simon MC, Orkin SH WD. Mouse GATA-4: a retinoic acid-inducible GATA-binding transcription factor expressed in endodermally derived tissues and heart. *Mol Cell Biol.* 1993;13(4):2235-2246.

166. Takash W, Cañizares J, Bonneaud N, Poulat F, Mattéi MG, Jay P BP. SOX7 transcription factor: sequence, chromosomal localisation, expression, transactivation and interference with Wnt signalling. *Nucleic Acids Res.* 2001;29(21):4274-83.

167. Wat MJ, Shchelochkov OA, Holder AM, Breman AM, Dagli A, Bacino C, Scaglia F, Zori RT, Cheung SW, Scott DA KS. Chromosome 8p23.1 deletions as a cause of complex congenital heart defects and diaphragmatic hernia. *Am J Med Genet A*. 2009;149A(8):1661-77.

168. Ballarati L, Cereda A, Caselli R, Selicorni A, Recalcati MP, Maitz S, Finelli P, Larizza L GD. Genotype-phenotype correlations in a new case of 8p23.1 deletion and review of the literature. *Eur J Med Genet*. 2011;54(1):55-9.

169. Yu S, Fiedler S, Stegner A GW. Genomic profile of copy number variants on the short arm of human chromosome 8. *Eur J Hum Genet*. 2010;18(10):1114-20.

170. Hollox EJ, Barber JC, Brookes AJ AJ. Defensins and the dynamic genome: what we can learn from structural variation at human chromosome band 8p23.1. *Genome Res*. 2008;18(11):1686-97.

171. Baynam G, Goldblatt J WI. Deletion of 8p23.1 with features of Cornelia de Lange syndrome and congenital diaphragmatic hernia and a review of deletions of 8p23.1 to 8pter? A further locus for Cornelia de Lange syndrome. *Am J Med Genet A.* 2008;146A(12):1565-70.

172. Esch, F. S., Keim, P. S., Beattie, E. C., Blacher, R. W., Culwell, A. R., Oltersdorf, T., McClure, D., Ward PJ. Cleavage of amyloid beta peptide during constitutive processing of its precursor. *Science*. 1990;248:1122-1124.

173. Berquin, I. M., Cao, L., Fong, D., Sloane BF. Identification of two new exons and multiple transcription start points in the 5-prime-untranslated region of the human cathepsin-B-encoding gene. *Gene*. 1995;159:143-149.

174. Roessler, B. J., Bell, G., Heidler, S., Seino, S., Becker, M., Palella TD. Cloning of two distinct copies of human phosphoribosylpyrophosphate synthetase cDNA. *Nucleic Acids Res.* 1990;18.

175. Roessler, B. J., Palella, T. D., Heidler, S., Becker MA. Identification of distinct PRPS1 mutations in two patients with X-linked phosphoribosylpyrophosphate synthetase superactivity. *Clin. Res.* 1991;39.

176. Stevens, S. J. C., Smeets, E. E. J. G. L., Van Den Broek, N., Droog, R. P., Breukels, M. A., Albrechts, J. C. M., Rauh-van Delst, M., Traa, E., Lennarts, M., Janssen, J. W. H., Engelen JJ. Partial monosomy 8p/trisomy 8q in a newborn infant due to a maternal three-way translocation: clinical and cytogenetic comparison with San Luis Valley syndrome. *Am. J. Med. Genet.* 2010;152A:2123-2126.

177. Appel, S., Reichwald, K., Zimmermann, W., Reis, A., Rosenthal, A., Hennies HC. Identification and localization of a new human myotubularin-related protein gene, MTMR8, on 8p22-p23. *Genomics*. 2001;75:6-8.

178. Comes N, Buie LK BT. Evidence for a role of angiopoietin-like 7 (*ANGPTL7*) in extracellular matrix formation of the human trabecular meshwork: implications for glaucoma. *Genes Cells.* 2011;16(2):243-32.

179. Yuan, Z.-R., Wang, R., Solomon, J., Luo, X., Sun, H., Zhang, L., Shi Y. Identification and characterization of survival-related gene, a novel cell survival gene controlling apoptosis and tumorigenesis. *Cancer Res.* 2005;65:10716-10724.

180. Cahill, D. P., Da Costa, L. T., Carson-Walter, E. B., Kinzler, K. W., Vogelstein, B., Lengauer C. Characterization of MAD2B and other mitotic spindle checkpoint genes. *Genomics*. 1999;58:181-187.

181. Vilhais-Neto, G. C., Maruhashi, M., Smith, K. T., Vasseur-Cognet, M., Peterson, A. S., Workman, J. L., Pourquie O. Rere controls retinoic acid signalling and somite bilateral symmetry. *Nature*. 2010;463:953-957.

182. Vulcani-Freitas TM, Saba-Silva N, Cappellano A, Cavalheiro S TS. PRAME gene expression profile in medulloblastoma. *Arq Neuropsiquiatr.* 2011;69(1):9-12.

183. Otto, E. A., Hurd, T. W., Airik, R., Chaki, M., Zhou, W., Stoetzel, C., Patil, S. B., Levy, S., Ghosh, A. K., Murga-Zamalloa, C. A., Van Reeuwijk, J., Letteboer SJF. Candidate exome capture identifies mutation of *SDCCAG8* as the cause of a retinal-renal ciliopathy. *Nature Genet*. 2010;42:840-850.

184. Nakatani, K., Sakaue, H., Thompson, D. A., Weigel, R. J., Roth RA. Identification of a human Akt3 (protein kinase B gamma) which contains the regulatory serine phosphorylation site. *Biochem. Biophys. Res. Commun.* 1999;257:906-910.

185. Boland, E., Clayton-Smith, J., Woo, V. G., McKee, S., Manson, F. D. C., Medne, L., Zackai, E., Swanson, E. A., Fitzpatrick, D., Millen, K. J., Sherr, E. H., Dobyns, W. B., Black GCM. Mapping of deletion and translocation breakpoints in 1q44 implicates the serine/threonine kinase AKT3 in postnatal microcephaly and agenesis of the corpus callosum. *Am. J. Hum. Genet.* 2007;81:292-303.

186. Encinas, J. A., Kikuchi, K., Chedotal, A., De Castro, F., Goodman, C. S., Kimura T. Cloning, expression, and genetic mapping of Sema W, a member of the semaphorin family. *Proc. Nat. Acad. Sci.* 1999;96:2491-2496.

187. Joshi, R., Gilligan, D. M., Otto, E., McLaughlin, T., Bennett V. Primary structure and domain organization of human alpha and beta adducin. *J. Cell Biol.* 1991;115:665-675.

188. Gilligan, D. M., Lozovatsky, L., Gwynn, B., Brugnara, C., Mohandas, N., Peters LL. Targeted disruption of the beta adducin gene (Add2) causes red blood cell spherocytosis in mice. *Proc. Nat. Acad. Sci.* 1999;96:10717-10722.

189. Britton, S., Freeman, T., Vafiadaki, E., Keers, S., Harrison, R., Bushby, K., Bashir R. The third human FER-1-like protein is highly similar to dysferlin. *Genomics*. 2000;68(313-321).

190. Marchese, A., Sawzdargo, M., Nguyen, T., Cheng, R., Heng, H. H. Q., Nowak, T., Im, D-S., Lynch, K. R., George, S. R., O'Dowd BF. Discovery of three novel orphan G-protein-coupled receptors. *Genomics*. 1999;56:12-21.

191. Sumiyama, K., Washio-Watanabe, K., Saitou, N., Hayakawa, T., Ueda S. Class III POU genes: generation of homopolymeric amino acid repeats under GC pressure in mammals. *J. Molec. Evol.* 1996;43:170-178.

192. Kunita R, Otomo A IJ. Identification and characterization of novel members of the CREG family, putative secreted glycoproteins expressed specifically in brain. *Genomics*. 2002;80(5):456-60.

193. Koc, E. C., Burkhart, W., Blackburn, K., Moseley, A., Spremulli LL. The small subunit of the mammalian mitochondrial ribosome: identification of the full complement of ribosomal proteins present. *J. Biol. Chem.* 2001;276:19363-19374.

194. Shimizu K, Watanabe K, Yamashita H, Abe M, Yoshimatsu H, Ohta H, Sonoda H SY. Gene regulation of a novel angiogenesis inhibitor, vasohibin, in endothelial cells. *Biochem Biophys Res Commun.* 2005;327(3):700-6.

195. Suzuki, N., Matsunaga, T., Nagasumi, K., Yamamura, T., Shihara, N., Moritani, T., Ue, H., Fukushima, M., Tamon, A., Seino, Y., Tsuda, K., Yasuda K. alpha(2B)-adrenergic receptor deletion polymorphism associates with autonomic nervous system activity in young healthy Japanese. *J. Clin. Endocr. Metab.* 2003;88:1184-1187.

196. Wang C-Y, Yang P, Shi J-D, et al. Molecular cloning and characterization of the mouse Acdp gene family. *BMC Genomics*. 2004;5(1):7.

197. Qin Y, Yao L, King EE, et al. Germline mutations in TMEM127 confer susceptibility to pheochromocytoma. *Nature Genet*. 2010;42:229-233.

198. Nevea E, Svenssona K, Fuxea J, Pettersson RF. VIPL, a VIP36-like membrane protein with a putative function in the export of glycoproteins from the endoplasmic reticulum. *Experimental Cell Research*. 2000;288(1):70-83.

199. Croft M, Turnpenny P. Deletion 3q22.1-q23 with blepharophimosis, ptosis and epicanthus inversus and an Albright hereditary osteodystrophy-like brachydactyly phenotype. *Clin Dysmorphol.* 2008;17(3):189-91.

200. Crisponi, L., Deiana, M., Loi, A., Chiappe, F., Uda, M., Amati, P., Bisceglia, L., Zelante, L., Nagaraja, R., Porcu, S., Ristaldi, M. S., Marzella R. The putative forkhead transcription factor FOXL2 is mutated in blepharophimosis/ptosis/epicanthus inversus syndrome. *Nature Genet*. 2001;27:159-166.

201. Batlle, E., Bacani, J., Begthel, H., Jonkeer, S., Gregorieff, A., Van De Born, M., Malats, N., Sancho, E., Boon, E., Pawson, T., Gallinger, S., Pals, S., Clevers H. EphB receptor activity suppresses colorectal cancer progression. *Nature*. 2005;435:1126-1130.

202. Cowan, C. A., Yokoyama, N., Bianchi, L. M., Henkemeyer, M., Fritzsch B. EphB2 guides axons at the midline and is necessary for normal vestibular function. *Neuron*. 2000;26:417-430.

203. Wanschers BF, Van De Vorstenbosch R, Schlager MA, Splinter D, Akhmanova A, Hoogenraad CC, Wieringa B FJ. A role for the Rab6B Bicaudal-D1 interaction in retrograde transport in neuronal cells. *Exp Cell Res.* 2007;313(17):3408-20.

204. Komatsu, M., Chiba, T., Tatsumi, K., Iemura, S., Tanida, I., Okazaki, N., Ueno, T., Kominami, E., Natsume, T., Tanaka K. A novel protein-conjugating system for Ufm1, a ubiquitin-fold modifier. *EMBO*. 2004;23:1977-1986.

205. Zheng M, Gu X, Zheng D, Yang Z, Li F, Zhao J, Xie Y, Ji C MY. UBE1DC1, an ubiquitin-activating enzyme, activates two different ubiquitin-like proteins. *J Cell Biochem.* 2008;104(6):2324-34.

206. De Baere E, Speleman F, Van Roy N, De Paepe A ML. Assignment of the cellular retinol-binding protein 1 gene (RBP1) and of the coatomer beta subunit gene (COPB2) to human chromosome band 3q23 by in situ hybridization. *Cytogenet Cell Genet.* 1998;82:226-7.

207. Smits P, Saada A, Wortmann SB, Heister AJ, Brink M, Pfundt R, Miller C, Haas D, Hantschmann R, Rodenburg RJ, Smeitink JA VDHL. Mutation in mitochondrial ribosomal protein MRPS22 leads to Cornelia de Lange-like phenotype, brain abnormalities and hypertrophic cardiomyopathy. *Eur J Hum Genet.* 2010;19(4):394-9.

208. Wanschers BF, Van De Vorstenbosch R, Schlager MA, Splinter D, Akhmanova A, Hoogenraad CC, Wieringa B FJ. A role for the Rab6B Bicaudal-D1 interaction in retrograde transport in neuronal cells. *Exp Cell Res.* 2007;313:3406-20.

209. Horiuchi, K., Weskamp, G., Lum, L., Hammes, H.-P., Cai, H., Brodie, T. A., Ludwig, T., Chiusaroli, R., Baron, R., Preissner, K. T., Manova, K., Blobel CP. Potential role for ADAM15 in pathological neovascularization in mice. *Molec. Cell. Biol.* 2003;23:5614-5624.

210. Primakoff, P., Myles DG. The ADAM gene family: surface proteins with adhesion and protease activity. *Trends Genet.* 2000;16:83-87.

211. Ting MJ, Day BW, Spanevello MD BA. Activation of ephrin A proteins influences hematopoietic stem cell adhesion and trafficking patterns. *Exp Hematol.* 2010;38(11):1087-98.

212. Wang X, Yeh S, Wu G, Hsu CL, Wang L, Chiang T, Yang Y, Guo Y CC. Identification and characterization of a novel androgen receptor coregulator ARA267-alpha in prostate cancer cells. *J Biol Chem.* 2001;276(44):40417-23.

213. Sotos JF, Dodge PR,Muirhead D, Crawford JD. Cerebral gigantism in childhood: a syndrome of excessively rapid growth and acromegaic features and a nonprogressive neurologic disorder. *N Engl J Med.* 1984;271:109-16.

214. Nagai, T., Matsumoto, N., Kurotaki, N., Harada, N., Niikawa, N., Ogata, T., Imaizumi, K., Kurosawa, K., Kondoh, T., Ohashi, H., Tsukahara, M., Makita, Y., Sugimoto, T., Sonoda, T., Yokoyama, T., Uetake, K., Sakazume, S., Fukushima, Y., Naritomi K. Sotos syndrome and haploinsufficiency of NSD1: clinical features of intragenic mutations and submicroscopic deletions. *J. Med. Genet.* 2003;40:285-289.

215. Tahvanainen, J., Kallonen, T., Lahteenmaki, H., Heiskanen, K. M., Westermarck, J., Rao, K. V. S., Lahesmaa R. PRELI is a mitochondrial regulator of human primary T-helper cell apoptosis, STAT6, and Th2-cell differentiation. *Blood*. 2009;11:1268-1277.

216. Erdman, R. A., Shellenberger, K. E., Overmeyer, J. H., Maltese WA. Rab24 is an atypical member of the Rab GTPase family: deficient GTPase activity, GDP dissociation inhibitor interaction, and prenylation of Rab24 expressed in cultured cells. *J. Biol. Chem.* 2000;275:3848-3856.

217. Chen, H., Kalaitsidaki, M., Warren, A. C., Avramopoulos, D., Antonarakis SE. A novel zinc finger cDNA with a polymorphic pentanucleotide repeat (ATTTT)n maps on human chromosome 19p. *Genomics*. 1993;15:621-625.

218. Lehman A, Du Souich C, Chai D, Eydoux P, Huang J, Fok A, Avila L, Swingland J, Delaney A, McGillivray B, Goldowitz D, Argiropoulos B, Kobor M BC. 19p13.2 microduplication causes a Sotos syndrome-like phenotype and alters gene expression. *Clin Genet.* 2010.

219. Chang, T.-C., Yamashita, A., Chen, C.-Y. A., Yamashita, Y., Zhu, W., Durdan, S., Kahvejian, A., Sonenberg, N., Shyu A-B. UNR, a new partner of poly(A)-binding protein, plays a key role in translationally coupled mRNA turnover mediated by the c-fos major coding-region determinant. *Genes Dev.* 2004;18:2010-2023.

220. Kuan, C.-Y., Yang, D. D., Roy, D. R. S., Davis, R. J., Rakic, P., Flavell RA. The Jnk1 and Jnk2 protein kinases are required for regional specific apoptosis during early brain development. *Neuron*. 1999;22:667-676.

221. Wijk SJ, De Vries SJ, Kemmeren P, Huang A, Boelens R, Bonvin AM TH van. A comprehensive framework of E2-RING E3 interactions of the human ubiquitin-proteasome system. *Mol Syst Biol.* 2009;295(5).

222. Jackson, A., Panayiotidis, P., Foroni L. The human homologue of the Drosophila tailless gene (TLX): characterization and mapping to a region of common deletion in human lymphoid leukemia on chromosome 6q21. *Genomics*. 1998;50:34-43.

223. Shi, Y., Lie, D. C., Taupin, P., Nakashima, K., Ray, J., Yu, R. T., Gage, F. H., Evans RM. Expression and function of orphan nuclear receptor TLX in adult neural stem cells. *Nature*. 2004;427:78-83.

224. Monaghan, A. P., Bock, D., Gass, P., Schwager, A., Wolfer, D. P., Lipp, H.-P., Schutz G. Defective limbic system in mice lacking the tailless gene. *Nature*. 1997;390:515-517.

225. Ouled Amar Ben Cheikh, B., Baulac, S., Lahjouji, F., Bouhouche, A., Couarch, P., Khalili, N., Regragui, W., Lehericy, S., Ruberg, M., Benomar, A., Heath, S., Chkili, T., Yahyaoui, M., Jiddane, M., Ouazzani, R., LeGuern E. A locus for bilateral occipital polymicrogyria maps to chromosome 6q16-q22. *Neurogenetics*. 2009;10:35-42.

226. Abu-Issa, R., Eichele, G., Youssoufian H. Expression of the Fanconi anemia group A gene (Fanca) during mouse embryogenesis. *Blood*. 1999;94:818-824.

227. Smallwood, P. M., Munoz-Sanjuan, I., Tong, P., Macke, J. P., Hendry, S. H. C., Gilbert, D. J., Copeland, N. G., Jenkins, N. A., Nathans J. Fibroblast growth factor (FGF) homologous factors: new members of the FGF family implicated in nervous system development. *Proc. Nat. Acad. Sci.* 1996;93:9850-9857.

228. Haft, C. R., De La Luz Sierra, M., Barr, V. A., Haft, D. H., Taylor SI. Identification of a family of sorting nexin molecules and characterization of their association with receptors. *Molec. Cell. Biol.* 1998;18:7278-7287.

229. Vervoort, V. S., Viljoen, D., Smart, R., Suthers, G., DuPont, B. R., Abbott, A., Schwartz CE. Sorting nexin 3 (SNX3) is disrupted in a patient with a translocation t(6;13)(q21;q12) and microcephaly, microphthalmia, ectrodactyly, prognathism (MMEP) phenotype. *J. Med. Genet.* 2002;39:893-899.

230. Pleiser S, Rock R, Wellmann J, Gessler M KE. Expression patterns of the mouse Spir-2 actin nucleator. *Gene Expr Patterns.* 2010;10(7-8):345-350.

231. Titomanlio L, Giurgea I, Baumann C, et al. A locus for sacral/anorectal malformations maps to 6q25.3 in a 0.3Mb interval region. *European Journal of Human Genetics*. 2006;14:971-974.

232. Nagamani S, Erez A, Eng C, et al., with Sandesh. Interstitial deletion of 6q25.2–q25.3: a novel microdeletion syndrome associated with microcephaly, developmental delay, dysmorphic features and hearing loss. *European Journal of Human Genetics*. 2009;17:573-581.

233. Nemoto, Y., Wenk, M. R., Watanabe, M., Daniell, L., Murakami, T., Ringstad, N., Yamada, H., Takei, K., De Camilli P. Identification and characterization of a synaptojanin 2 splice isoform predominantly expressed in nerve terminals. *J. Biol. Chem.* 2001;276:41133-41142.

234. Schimenti JC, Reynolds JL PA. Mutations in Serac1 or Synj2 cause proximal t haplotype-mediated male mouse sterility but not transmission ratio distortion. *Proc Natl Acad Sci*. 2005;102(9):3342-7.

235. Ronshaugen M, McGinnis N, Inglis D, Chou D, Zhao J MW. Structure and expression patterns of Drosophila TULP and TUSP, members of the tubby-like gene family. *Mech Dev*. 2002;117:209-15.

236. Yarar, D., Waterman-Storer, C. M., Schmid SL. SNX9 couples actin assembly to phosphoinositide signals and is required for membrane remodeling during endocytosis. *Dev. Cell*. 2007;13:43-56.

237. Boeckel, T., Dierks, A., Vergopoulos, A., Bahring, S., Knoblauch, H., Muller-Myhsok, B., Baron, H., Aydin, A., Bein, G., Luft, F. C., Schuster H. A new mutation in the elastin gene causing supravalvular aortic stenosis. *Am. J. Cardiol.* 1999;83:1141-1143.

238. Bennett, M. K., Calakos, N., Scheller RH. Syntaxin: a synaptic protein implicated in docking of synaptic vesicles at presynaptic active zones. *Science*. 1992;257:255-259.

239. Yuan-Qing Wu, Bassem A. Bejjani, Lap-Chee Tsui, Ariane Mandel LR, Osborne ALGS. Refinement of the Genomic Structure of STX1A and Mutation Analysis in Nondeletion Williams Syndrome Patients. *Am J Med Genet.* 2002;109:121-124.

240. Cairo, S., Merla, G., Urbinati, F., Ballabio, A., Reymond A. WBSCR14, a gene mapping to the Williams-Beuren syndrome deleted region, is a new member of the Mlx transcription factor network. *Hum. Molec. Genet.* 2001;10:617-627.

241. Lu, X., Meng, X., Morris, C. A., Keating MT. A novel human gene, WSTF, is deleted in Williams syndrome. *Genomics*. 1998;54:241-249.

242. Pérez Jurado LA, Wang YK, Peoples R, Coloma A, Cruces J FU. A duplicated gene in the breakpoint regions of the 7q11.23 Williams-Beuren syndrome deletion encodes the initiator binding protein TFII-I and BAP-135, a phosphorylation target of BTK. *Hum Mol Genet.* 1998;7:325-34.

243. Morris, C. A., Mervis, C. B., Hobart, H. H., Gregg, R. G., Bertrand, J., Ensing, G. J., Sommer, A., Moore, C. A., Hopkin, R. J., Spallone, P. A., Keating, M. T., Osborne, L., Kimberley, K. W., Stock AD. GTF2I hemizygosity implicated in mental retardation in Williams syndrome: genotype-phenotype analysis of five families with deletions in the Williams syndrome region. *Am. J. Med. Genet.* 2003;123A:45-59.

244. Dai, L., Bellugi, U., Chen, X.-N., Pulst-Korenberg, A. M., Jarvinen-Pasley, A., Tirosh-Wagner, T., Eis, P. S., Graham, J., Mills, D., Searcy, Y., Korenberg JR. Is it Williams syndrome? GTF2IRD1 implicated in visual-spatial construction and GTF2I in sociability revealed by high resolution arrays. *Am. J. Med. Genet.* 2009;149A:302-314.

245. Funakoshi T, Maeshima K, Yahata K, Sugano S, Imamoto F IN. Two distinct human POM121 genes: requirement for the formation of nuclear pore complexes. *FEBS Lett.* 2007;581(25):4910-6.

246. Mersman, D. P., Du, H.-N., Fingerman, I. M., South, P. F., Briggs SD. Polyubiquitination of the demethylase Jhd2 controls histone methylation and gene expression. *Genes Dev.* 2009;23:951-962.

247. Jensen, L. R., Amende, M., Gurok, U., Moser, B., Gimmel, V., Tzschach, A., Janecke, A. R., Tariverdian, G., Chelly, J., Fryns, J.-P., Van Esch, H., Kleefstra, T., Hamel, B., Moraine, C., Gecz, J., Turner, G., Reinhardt, R., Kalscheuer, V. M., Ropers, H.-H S. Mutations in the JARID1C gene, which is involved in transcriptional regulation and chromatin remodeling, cause X-linked mental retardation. *Am. J. Hum. Genet.* 2005;76:227-236.

248. Hayashi, K., Yano, H., Hashida, T., Takeuchi, R., Takeda, O., Asada, K., Takahashi, E., Kato, I., Sobue K. Genomic structure of the human caldesmon gene. *Proc. Nat. Acad. Sci.* 1992;89:12122-12126.

249. Mayanagi T SK. Diversification of caldesmon-linked actin cytoskeleton in cell motility. *Cell Adh Migr.* 2011;5:150-9.

250. Laramie JM, Wilk JB, Williamson SL, Nagle MW, Latourelle JC, Tobin JE, Province MA, Borecki IB MR. Polymorphisms near EXOC4 and LRGUK on chromosome 7q32 are associated with Type 2 Diabetes and fasting glucose; the NHLBI Family Heart Study. *BMC Med Genet*. 2008;9:46.

251. Iwakoshi, M., Okamoto, N., Harada, N., Nakamura, T., Yamamori, S., Fujita, H., Niikawa, N., Matsumoto N. Loss-of-function mutations in Euchromatin histone methyl transferase 1 (EHMT1) cause the 9q34 subtelomeric deletion syndrome. *Am. J. Hum. Genet.* 2006;79:370-377.

252. Kramer JM, Kochinke K, Oortveld MA, Marks H, Kramer D, De Jong EK, Asztalos Z, Westwood JT, Stunnenberg HG, Sokolowski MB, Keleman K, Zhou H, Van Bokhoven H SA. Epigenetic regulation of learning and memory by Drosophila EHMT/G9a. *PLoS Biol.* 2011;9.

253. Dyson, E., Sucov, H. M., Kubalak, S. W., Schmid-Schonbein, G. W., DeLano, F. A., Evans, R. M., Ross, J., Jr., Chien KR. Atrial-like phenotype is associated with embryonic ventricular failure in retinoid X receptor alpha -/- mice. *Proc. Nat. Acad. Sci.* 1995;92:7386-7390.

254. Mohn, A. R., Gainetdinov, R. R., Caron, M. G., Koller BH. Mice with reduced NMDA receptor expression display behaviors related to schizophrenia. *Cell*. 1999;98:427-436.

255. Muller, J., Ory, S., Copeland, T., Piwnica-Worms, H., Morrison DK. C-TAK1 regulates Ras signaling by phosphorylating the MAPK scaffold, KSR1. *Molec. Cell*. 2001;8:983-993.

256. Chia CY, Lim CW, Leong WT LM. High expression stability of microtubule affinity regulating kinase 3 (MARK3) makes it a reliable reference gene. *IUBMB Life*. 2010;62(3):200-203.

257. Chernajovsky, Y., Brown, A., Clark J. Human kinesin light (beta) chain gene: DNA sequence and functional characterization of its promoter and first exon. *DNA Cell Biol.* 1996;15:965-974.

258. Jennings, M. D., Pavitt GD. eIF5 has GDI activity necessary for translational control by eIF2 phosphorylation. *Nature*. 2010;465:378-381.

259. Goldschmidt-Reisin, S., Kitakawa, M., Herfurth, E., Wittmann-Liebold, B., Grohmann, L., Graack H-R. Mammalian mitochondrial ribosomal proteins: N-terminal amino acid sequencing, characterization, and identification of corresponding gene sequences. *J. Biol. Chemm.* 1998;273:34828-34836.

260. Mangale, V. S., Hirokawa, K. E., Satyaki, P. R. V., Gokulchandran, N., Chikbire, S., Subramanian, L., Shetty, A. S., Martynoga, B., Paul, J., Mai, M. V., Li, Y., Flanagan, L. A., Tole, S., Monuki ES. Lhx2 selector activity specifies cortical identity and suppresses hippocampal organizer fate. *Science*. 2008;319:304-309.

261. Lavdas, A. A., Grigoriou, M., Pachnis, V., Parnavelas JG. The medial ganglionic eminence gives rise to a population of early neurons in the developing cerebral cortex. *J. Neurosci.* 1999;99:7881-7888.

262. Gong, S., Zheng, C., Doughty, M. L., Losos, K., Didkovsky, N., Schambra, U. B., Nowak, N. J., Joyner, A., Leblanc, G., Hatten, M. E., Heintz N. A gene expression atlas of the central nervous system based on bacterial artificial chromosomes. *Nature*. 2003;425:Gong, S., Zheng, C., Doughty, M. L., Losos, K., Di.

263. Schumacher, J. M., Artzt, K., Braun RE. Spermatid perinuclear ribonucleic acid-binding protein binds microtubules in vitro and associates with abnormal manchettes in vivo in mice. *Biol. Reprod.* 1998;59:69-76.

264. Zhu, Y., Paszty, C., Turetsky, T., Tsai, S., Kuypers, F. A., Lee, G., Cooper, P., Gallagher, P. G., Stevens, M. E., Rubin, E., Mohandas, N., Mentzer WC. Stomatocytosis is absent in “stomatin”-deficient murine red blood cells. *Blood*. 1999;93:2404-2410.

265. Zardo, G., Tiirikainen, M. I., Hong, C., Misra, A., Feuerstein, B. G., Volik, S., Collins, C. C., Lamborn, K. R., Bollen, A., Pinkel, D., Albertson, D. G., Costello JF. Integrated genomic and epigenomic analyses pinpoint biallelic gene inactivation in tumors. *Nature Genet*. 2002;32:453-458.

266. Fulp CT, Cho G, Marsh ED, Nasrallah IM, Labosky PA GJ. Identification of Arx transcriptional targets in the developing basal forebrain. *Hum Mol Genet*. 2008;17(23):3740-60.

267. Jeong, D., Kim, J. M., Cha, H., Oh, J. G., Park, J., Yun, S.-H., Ju, E.-S., Jeon, E.-S., Hajjar, R. J., Park WJ. PICOT attenuates cardiac hypertrophy by disrupting calcineurin-NFAT signaling. *Circ. Res.* 2008;102:711-719.

268. Kao SC, Wu H, Xie J, Chang CP, Ranish JA, Graef IA CG. Calcineurin/NFAT signaling is required for neuregulin-regulated Schwann cell differentiation. *Science*. 2009;323:651-4.

269. Hegi, M. E., Diserens, A.-C., Gorlia, T., Hamou, M.-F., De Tribolet, N., Weller, M., Kros, J. M., Hainfellner, J. A., Mason, W., Mariani, L., Bromberg, J. E. C., Hau, P., Mirimanoff, R. O., Cairncross, J. G., Janzer, R. C., Stupp R. MGMT gene silencing and benefit from temozolomide in glioblastoma. *New Eng. J. Med.* 2005;352:997-1003.

270. Muncke, N., Jung, C., Rudiger, H., Ulmer, H., Roeth, R., Hubert, A., Goldmuntz, E., Driscoll, D., Goodship, J., Schon, K., Rappold G. Missense mutations and gene interruption in PROSIT240, a novel TRAP240-like gene, in patients with congenital heart defect (transposition of the great arteries). *Circulation*. 2003;108:2843-2850.

271. Dias, D. C., Dolios, G., Wang, R., Pan Z-Q. CUL7: A DOC domain-containing cullin selectively binds Skp1-Fbx29 to form an SCF-like complex. *Proc. Nat. Acad. Sci.* 2002;99:16601-16606.

272. Inohara, N., Ding, L., Chen, S., Nunez G. Harakiri, a novel regulator of cell death, encodes a protein that activates apoptosis and interacts selectively with survival-promoting proteins Bcl-2 and Bcl-X(L). *EMBO J*. 1997;16:1686-1694.

273. Brinkmann U. CAS, the human homologue of the yeast chromosome-segregation gene CSE1, in proliferation, apoptosis, and cancer. *Am. J. Hum. Genet.* 1998;62:509-513.

274. Moro E, Maran C, Slongo ML, Argenton F, Toppo S OM. Zebrafish spata2 is expressed at early developmental stages. *Int J Dev Biol.* 2007;51:241-6.

275. Rape, M., Kirschner MW. Autonomous regulation of the anaphase-promoting complex couples mitosis to S-phase entry. *Nature*. 2004;432:588-595.

276. Townsley, F. M., Aristarkhov, A., Beck, S., Hershko, A., Ruderman JV. Dominant-negative cyclin-selective ubiquitin carrier protein E2-C/UbcH10 blocks cells in metaphase. *Proc. Nat. Acad. Sci.* 1997;94:2362-2367.

277. Sheen, V. L., Ganesh, V. S., Topcu, M., Sebire, G., Bodell, A., Hill, R. S., Grant, P. E., Shugart, Y. Y., Imitola, J., Khoury, S. J., Guerrini, R., Walsh CA. Mutations in ARFGEF2 implicate vesicle trafficking in neural progenitor proliferation and migration in the human cerebral cortex. *Nature Genet.* 2004;36:69-76.

278. Capon, F., Bijlmakers, M.-J., Wolf, N., Quaranta, M., Huffmeier, U., Allen, M., Timms, K., Abkevich, V., Gutin, A., Smith, R., Warren, R. B., Young HS. Identification of ZNF313/RNF114 as a novel psoriasis susceptibility gene. *Hum. Molec. Genet.* 2008;17:1938-1945.

279. Menard, C., Hein, P., Paquin, A., Savelson, A., Yang, X. M., Lederfein, D., Barnabe-Heider, F., Mir, A. A., Sterneck, E., Peterson, A. C., Johnson, P. F., Vinson, C., Miller FD. An essential role for a MEK-C/EBP pathway during growth factor-regulated cortical neurogenesis. *Neuron*. 2002;36:597-610.

280. Zhong, W., Feng, H., Santiago, F. E., Kipreos ET. CUL-4 ubiquitin ligase maintains genome stability by restraining DNA-replication licensing. *Nature*. 2007;80:561-566.

281. Ju, T., Cummings RD. Chaperone mutation in Tn syndrome. *Nature*. 2005;437.

282. Danon, M. J., Oh, S. J., DiMauro, S., Manaligod, J. R., Eastwood, A., Naidu, S., Schliselfeld LH. Lysosomal glycogen storage disease with normal acid maltase. *Neurology*. 2004;31:51-57.

283. Tanaka, Y., Guhe, G., Suter, A., Eskelinen, E.-L., Hartmann, D., Lullmann-Rauch, R., Janssen, P. M. L., Blanz, J., Von Figura, K., Saftig P. Accumulation of autophagic vacuoles and cardiomyopathy in LAMP-2-deficient mice. *Nature*. 2000;406:902-906.

284. Shi, B., Hsu, H.-L., Evens, A. M., Gordon, L. I., Gartenhaus RB. Expression of the candidate MCT-1 oncogene in B- and T-cell lymphoid malignancies. *Blood*. 2003;102:297-302.

285. Prosniak, M., Dierov, J., Okami, K., Tilton, B., Jameson, B., Sawaya, B. E., Gartenhaus RB. A novel candidate oncogene, MCT-1, is involved in cell cycle progression. *Cancer Res.* 1998;58:4233-4237.

286. Brunetti-Pierri N, Berg JS, Scaglia F, et al. Recurrent reciprocal 1q21.1 deletions and duplications associated with microcephaly or macrocephaly and developmental and behavioral abnormalities. *Nature genetics*. 2008;40(12):1466-71.

287. Mefford HC, Sharp AJ, Baker C, et al. Recurrent rearrangements of chromosome 1q21.1 and variable pediatric phenotypes. *The New England journal of medicine*. 2008;359(16):1685-99.

288. Kramps, T., Peter, O., Brunner, E., Nellen, D., Froesch, B., Chatterjee, S., Murone, M., Zullig, S., Basler K. Wnt/Wingless signaling requires BCL9/legless-mediated recruitment of pygopus to the nuclear beta-catenin-TCF complex. *Cell*. 2002;109:47-60.

289. Li J, Zhou G, Ji W, Feng G, Zhao Q, Liu J, Li T, Li Y, Chen P, Zeng Z, Wang T, Hu Z, Zheng L, Wang Y, Shen Y, He L SY. Common variants in the BCL9 gene conferring risk of schizophrenia. *Arch Gen Psychiatry.ç*. 2011;68(3):232-40.

290. Gelb, B. D., Zhang, J., Cotter, P. D., Gershin, I. F., Desnick RJ. Physical mapping of the human connexin 40 (GJA5) flavin-containing monooxygenase 5, and natriuretic peptide receptor A genes on 1q21. *Genomics*. 1997;39:409-411.

291. Simon, A. M., Goodenough, D. A., Paul DL. Mice lacking connexin40 have cardiac conduction abnormalities characteristic of atrioventricular block and bundle branch block. *Curr. Biol.* 1998;8:295-298.

292. Polyakov, A. V., Shagina I. A., Khlebnikova, O. V., Evgrafov OV. Mutation in the connexin 50 gene (GJA8) in a Russian family with zonular pulverulent cataract. *Clin. Genet.* 2001;60:476-478.

293. Thornton, C., Snowden, M. A., Carling D. Identification of a novel AMP-activated protein kinase beta subunit isoform that is highly expressed in skeletal muscle. *J. Biol. Chem.* 1998;273:12443-12450.

294. Ahel, D., Horejsi, Z., Wiechens, N., Polo, S. E., Garcia-Wilson, E., Ahel, I., Flynn, H., Skehel, M., West, S. C., Jackson, S. P., Owen-Hughes, T., Boulton SJ. Poly(ADP-ribose)-dependent regulation by DNA repair by the chromatin remodeling enzyme ALC1. *Science*. 2009;325:1240-1243.

295. Ma, N.-F., Hu, L., Fung, J. M., Xie, D., Zheng, B.-J., Chen, L., Tang, D.-J., Fu, L., Wu, Z., Chen, M., Fang, Y., Guan X-Y. Isolation and characterization of a novel oncogene, amplified in liver cancer 1, within a commonly amplified region at 1q21 in hepatocellular carcinoma. *Hepatology*. 2008;47:503-510.

296. Hannes FD, Sharp a J, Mefford HC, et al. 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. *Journal of medical genetics*. 2009;46(4):223-32.

297. Gennuso, F., Fernetti, C., Tirolo, C., Testa, N., L'Episcopo, F., Caniglia, S., Morale, M. C., Ostrow, J. D., Pascolo, L., Tiribelli, C., Marchetti B. Bilirubin protects astrocytes from its own toxicity by inducing up-regulation and translocation of multidrug resistance-associated protein 1 (Mrp1). *Proc. Nat. Acad. Sci.* 2004;101:2470-2475.

298. Gingeras TR. Origin of phenotypes: genes and transcripts. *Genome research*. 2007;17(6):682-90.

299. Thurman RE, Day N, Noble WS, Stamatoyannopoulos J a. Identification of higher-order functional domains in the human ENCODE regions. *Genome research*. 2007;17(6):917-27.

300. Washietl S, Pedersen JS, Korbel JO, et al. Structured RNAs in the ENCODE selected regions of the human genome. *Genome research*. 2007;17(6):852-64.

301. Koch CM, Andrews RM, Flicek P, et al. The landscape of histone modifications across 1% of the human genome in five human cell lines. *Genome research*. 2007;17(6):691-707.

302. Li MM, Andersson HC. Clinical Application of Microarray-Based Molecular Cytogenetics: An Emerging New Era of Genomic Medicine. *The Journal of Pediatrics*. 2009;155(3):311-317.

303. Roberts RL, Gallo LM, Barclay ML, Gómez-García M, Cardeña C, Merriman TR, Gearry RB MJ. Independent replication of an association of CNVR7113.6 with Crohn's disease in caucasians. *Inflamm Bowel Dis.* 2011.

304. Orange JS, Glessner JT, Resnick E, Sullivan KE, Lucas M, Ferry B, Kim CE, Hou C, Wang F, Chiavacci R, Kugathasan S, Sleasman JW, Baldassano R, Perez EE, Chapel H, Cunningham-Rundles C HH. Genome-wide association identifies diverse causes of common variable immunodeficiency. *J Allergy Clin Immunol.* 2011;127(6):1360-1367.

305. AA. AR. Molecular genetics of Psoriasis (Principles, technology, gene location, genetic polymorphism and gene expression). *Int J Health Sci (Qassim).* 2010;4(2):103-127.

306. F Erdogan, L A Larsen, L Zhang, Z Tümer, N Tommerup, W Chen, J RJacobsen, M Schubert, J Jurkatis, A Tzschach H-HRARU. High frequency of submicroscopic genomic aberrations detected by tiling path array comparative genome hybridisation in patients with isolated congenital heart disease. *J. Med. Genet.* 2008;45:704-709.

307. Vermeesch JR. Array CGH : Opening New Horizons. *New Horizons*. 2009:421-437.

Appendix

Supplementary tables

Table S1 – Genes used for PRC amplification of HUMARA locus.

Direction	Sequence 5'→3'	Size	TmºC	Modifications	Cycling conditions	
Forward	TCCAGAATCTGTTCCAGAGCGTGC	24	69	6-FAM (5')	96ºC – 10'	
Reverse	GCTGTGAAGGTTGCTGTTCCTCAT	24	66	None	96ºC – 45''' 67ºC – 30'' 72ºC – 45''	40 X
					72 ºC – 15' 4ºC - ∞	

Table S2 – OMIM entrances used as training set for the gene prioritization in Ts-CoExp web tool.

OMIM entrance	Title
#300419	MENTAL RETARDATION, X-LINKED, WITH OR WITHOUT SEIZURES, ARX-RELATED; MRXARX
#309500	RENPENNING SYNDROME 1; RENS1
#300055	MENTAL RETARDATION, X-LINKED, SYNDROMI C 13; MRXS13; MENTAL RETARDATION, X-LINKED 79; MRX79
#303350	MASA SYNDROME
#309510	PARTINGTON X-LINKED MENTAL RETARDATION SYNDROME; PRTS
#300624	FRAGILE X MENTAL RETARDATION SYNDROME; PRIMARY OVARIAN INSUFFICIENCY, FRAGILE X-ASSOCIATED, INCLUDED
#224050	CEREBELLAR ATAXIA, MENTAL RETARDATION, AND DYSEQUILIBRIUM SYNDROME 1; CAMRQ1
#300260	LUBS X-LINKED MENTAL RETARDATION SYNDROME; MRXSL
#300354	MENTAL RETARDATION, X-LINKED, WITH SHORT STATURE, HYPOGONADISM, AND ABNORMAL GAIT
#300749	MENTAL RETARDATION AND MICROCEPHALY WITH PONTINE AND CEREBELLAR HYPOPLASIA; MENTAL RETARDATION, X-LINKED, CASK-RELATED, INCLUDED
#300220	MENTAL RETARDATION, X-LINKED, SYNDROMIC 10; MRXS10
#309530	MENTAL RETARDATION, X-LINKED 1; MRX1; MENTAL RETARDATION, X-LINKED 18, INCLUDED; MRX18, INCLUDED
#309580	MENTAL RETARDATION-HYPOTONIC FACIES SYNDROME, X-LINKED, 1; MRXHF1
#300486	MENTAL RETARDATION, X-LINKED, WITH CEREBELLAR HYPOPLASIA AND DISTINCTIVE FACIAL APPEARANCE
#300523	ALLAN-HERNDON-DUDLEY SYNDROME; AHDS
#309548	MENTAL RETARDATION, X-LINKED, ASSOCIAT ED WITH FRAGILE SITE FRAXE
#141750	ALPHA-THALASSEMIA/MENTAL RETARDATION SYNDROME, DELETION-TYPE
#300558	MENTAL RETARDATION, X-LINKED 30; MRX30
#300243	MENTAL RETARDATION, X-LINKED, SYNDROMIC, CHRISTIANSON TYPE
#300352	CREATINE DEFICIENCY SYNDROME, X-LINKED
#300143	MENTAL RETARDATION, X-LINKED 21; MRX21
#607417	MENTAL RETARDATION, AUTOSOMAL RECESSIV E 2; MRT2
#300659	MENTAL RETARDATION, X-LINKED 93; MRX93
#300263	SIDERIUS X-LINKED MENTAL RETARDATION SYNDROME
#300705	MENTAL RETARDATION, X-LINKED 17; MRX17
#241080	HYPOGONADISM, ALOPECIA, DIABETES MELLITUS, MENTAL RETARDATION, AND EXTRAPYRAMIDAL SYNDROME
#309549	MENTAL RETARDATION, X-LINKED 9; MRX9
#613227	CEREBELLAR ATAXIA, MENTAL RETARDATION, AND DYSEQUILIBRIUM SYNDROME 3; CAMRQ3
#600176	MICROCEPHALY, CORTICAL MALFORMATIONS, AND MENTAL RETARDATION
#300706	MENTAL RETARDATION, X-LINKED, SYNDROMIC, TURNER TYPE
#300387	MENTAL RETARDATION, X-LINKED 63; MRX63
#300123	MENTAL RETARDATION, X-LINKED, WITH PAN HYPOPITUITARISM; MENTAL RETARDATION, X-LINKED, WITH ISOLATED GROWTH HORMONE DEFICIENCY, INCLUDED; MRGH, INCLUDED
#300434	STOCCO DOS SANTOS X-LINKED MENTAL RETARDATION SYNDROME
#612292	BIRK-BAREL MENTAL RETARDATION DYSMORPHISM SYNDROME
#300534	MENTAL RETARDATION, X-LINKED, SYNDROMIC, JARID1C-RELATED
#235730	MOWAT-WILSON SYNDROME
#305450	OPITZ-KAVEGGIA SYNDROME; OKS
#300088	EPILEPSY, FEMALE-RESTRICTED, WITH MENTAL RETARDATION; EFMR
#249500	MENTAL RETARDATION, AUTOSOMAL RECESSIVE 1; MRT1
#239300	HYPERPHOSPHATASIA WITH MENTAL RETARDATION
#206700	ANIRIDIA, CEREBELLAR ATAXIA, AND MENTAL RETARDATION
#300630	MENTAL RETARDATION, X-LINKED 59; MRX59
#610954	PITT-HOPKINS SYNDROME; PTHS
#608443	MENTAL RETARDATION, AUTOSOMAL RECESSIVE 3; MRT3
#300436	MENTAL RETARDATION, X-LINKED 46; MRX46
#304400	DEAFNESS, X-LINKED 2; DFNX2; PERILYMPHATIC GUSHER DURING STAPES SURGERY, INCLUDED
#606612	MUSCULAR DYSTROPHY-DYSTROGLYCANOPATHY (CONGENITAL WITH OR WITHOUT MENTAL RETARDATION), TYPE B, 5; MDDGB5
#300676	MENTAL RETARDATION, X-LINKED, SYNDROMIC 14; MRXS14
#300271	MENTAL RETARDATION, X-LINKED 72; MRX72
#613155	MUSCULAR DYSTROPHY-DYSTROGLYCANOPATHY (CONGENITAL WITH MENTAL RETARDATION), TYPE B, 1; MDDGB1
#612447	SKELETAL DEFECTS, GENITAL HYPOPLASIA, AND MENTAL RETARDATION
#608840	MUSCULAR DYSTROPHY-DYSTROGLYCANOPATHY CONGENITAL WITH MENTAL RETARDATION), TYPE B, 6; MDDGB6

#309583	MENTAL RETARDATION, X-LINKED, SNYDER-ROBINSON TYPE
#180849	RUBINSTEIN-TAYBI SYNDROME 1; RSTS1
#156200	MENTAL RETARDATION, AUTOSOMAL DOMINANT1; MRD1
#613192	MENTAL RETARDATION, AUTOSOMAL RECESSIVE 13; MRT13
#136630	MENTAL RETARDATION, FRA12A TYPE
#604278	RENAL TUBULAR ACIDOSIS, PROXIMAL, WITH OCULAR ABNORMALITIES AND MENTAL RETARDATION
#611093	MENTAL RETARDATION, AUTOSOMAL RECESSIVE 7; MRT7
#300643	ROLANDIC EPILEPSY, MENTAL RETARDATION, AND SPEECH DYSPRAXIA, X-LINKED; RESDX
#610003	CEROID LIPOFUSCINOSIS, NEURONAL, 8, NORTHERN EPILEPSY VARIANT
#606858	PREMATURE CHROMOSOME CONDENSATION WITH MICROCEPHALY AND MENTAL RETARDATION
#300578	CHROMOSOME Xp11.3 DELETION SYNDROME
#300495	AUTISM, SUSCEPTIBILITY TO, X-LINKED 2; AUTSX2 MENTAL RETARDATION, X-LINKED, INCLUDED
#300472	CORPUS CALLOSUM, AGENESIS OF, WITH MENTAL RETARDATION, OCULAR COLOBOMA, AND MICROGNATHIA
#194072	WILMS TUMOR, ANIRIDIA, GENITOURINARY A NOMALIES, AND MENTAL RETARDATION SYNDROME; WAGR
#613670	MENTAL RETARDATION WITH LANGUAGE IMPAIRMENT AND AUTISTIC FEATURES
#613156	MUSCULAR DYSTROPHY-DYSTROGLYCANOPATHY
#613151	MUSCULAR DYSTROPHY-DYSTROGLYCANOPATHY
#300799	MENTAL RETARDATION, X-LINKED, SYNDROMIC, ZDHHC9-RELATED
#300498	MENTAL RETARDATION, X-LINKED 45; MRX45
#300194	AMME COMPLEX
#612780	SEIZURES, SENSORINEURAL DEAFNESS, ATAXIA, MENTAL RETARDATION, AND ELECTROLYTE IMBALANCE
#612652	MENTAL RETARDATION, JOINT HYPERMOBILITY, AND SKIN LAXITY, WITH OR WITHOUT METABOLIC ABNORMALITIES
#309520	LUJAN-FRYNS SYNDROME
#300423	MENTAL RETARDATION, X-LINKED, WITH EPILEPSY; XMRE
#177980	PTERYGIA, MENTAL RETARDATION, AND DISTINCTIVE CRANIOFACIAL FEATURES
#612580	MENTAL RETARDATION, AUTOSOMAL DOMINANT3; MRD3
#300699	MENTAL RETARDATION, X-LINKED 94; MRX94
#300210	MENTAL RETARDATION, X-LINKED 58; MRX58
#610156	MENTAL RETARDATION, TRUNCAL OBESITY, RETINAL DYSTROPHY, AND MICROPENIS
#608747	INSULIN-LIKE GROWTH FACTOR I DEFICIENCY
#241410	HYPOPARATHYROIDISM-RETARDATION-DYSMORPHISM SYNDROME; HRD
#164280	FEINGOLD SYNDROME
#613671	MENTAL RETARDATION, ANTERIOR MAXILLARY PROTRUSION, AND STRABISMUS; MRAMS
#613152	MUSCULAR DYSTROPHY-DYSTROGLYCANOPATHY(CONGENITAL WITHOUT MENTAL RETARDATION), TYPE B, 4; MDDGB4
#612581	MENTAL RETARDATION, AUTOSOMAL DOMINANT4; MRD4
#611092	MENTAL RETARDATION, AUTOSOMAL RECESSIVE 6; MRT6
#300803	MENTAL RETARDATION, X-LINKED, ZNF711-RELATED
#300802	MENTAL RETARDATION, X-LINKED, SYP-RELATED
#300716	MENTAL RETARDATION, X-LINKED 95; MRX95
#225060	CLEFT LIP/PALATE-ECTODERMAL DYSPLASIA SYNDROME; CLPED1; OROFACIAL CLEFT 7, INCLUDED; OFC7, INCLUDED
#606937	SPINOCEREBELLAR ATAXIA, AUTOSOMAL RECESSIVE 5; SCAR5
#613443	MENTAL RETARDATION, STEREOTYPIC MOVEMENTS, EPILEPSY, AND/OR CEREBRAL MALFORMATIONS
#612621	MENTAL RETARDATION, AUTOSOMAL 5; MRD5 DOMINANT
#300831	CK SYNDROME
#119540	CLEFT PALATE, ISOLATED; CPI; CLEFT PALATE, ISOLATED, AND MENTAL RETARDATION, INCLUDED
#609579	SCAPHOCEPHALY, MAXILLARY RETRUSION, AND MENTAL RETARDATION
#300577	MENTAL RETARDATION, X-LINKED 91; MRX91
#612469	WILMS TUMOR, ANIRIDIA, GENITOURINARY ANOMALIES, MENTAL RETARDATION, AND OBESITY SYNDROME; WAGRO
#308100	ICHTHYOSIS, X-LINKED; XLI
#301040	ALPHA-THALASSEMIA/MENTAL RETARDATION SYNDROME, NONDELETION TYPE, X-LINKED; ATRX
#218040	COSTELLO SYNDROME
#303600	COFFIN-LOWRY SYNDROME; CLS
#608636	CHROMOSOME 15q11-q13 DUPLICATION SYNDROME; AUTISM, SUSCEPTIBILITY TO, 4, INCLUDED; AUTS4, INCLUDED
#300672	EPILEPTIC ENCEPHALOPATHY, EARLY INFANTILE, 2
#257300	MOSAIC VARIEGATED ANEUPLOIDY SYNDROME; MVA
#216550	COHEN SYNDROME; COH1

#194050	WILLIAMS-BEUREN SYNDROME; WBS
#105830	ANGELMAN SYNDROME; AS; ANGELMAN SYNDROME CHROMOSOME REGION, INCLUDED; ANCR, INCLUDED
#612001	CHROMOSOME 15q13.3 DELETION SYNDROME
#123450	CRI-DU-CHAT SYNDROME
#300257	DANON DISEASE
#606777	GLUT1 DEFICIENCY SYNDROME 1; GLUT1DS1
#601224	POTOCKI-SHAFFER SYNDROME
#300049	HETEROTOPIA, PERIVENTRICULAR, X-LINKED DOMINANT; HETEROTOPIA, PERIVENTRICULAR NODULAR, WITH FRONTOMETAPHYSEAL DYSPLASIA, INCLUDED
#259730	OSTEOPETROSIS, AUTOSOMAL RECESSIVE 3; OPTB3
#606407	HYPOTONIA-CYSTINURIA SYNDROME; HOMOZYGOUS 2p21 DELETION SYNDROME, INCLUDED
#310200	MUSCULAR DYSTROPHY, DUCHENNE TYPE; DMD
#307000	HYDROCEPHALUS DUE TO CONGENITAL STENOSIS OF AQUEDUCT OF SYLVIUS; HSAS; HYDROCEPHALUS, X-LINKED, WITH CONGENITAL IDIOPATHIC INTESTINAL PSEUDOOBSTRUCTION, INCLUDED
#242860	IMMUNODEFICIENCY-CENTROMERIC INSTABILITY-FACIAL ANOMALIES SYNDROME
#609308	MUSCULAR DYSTROPHY-DYSTROGLYCANOPATHY(LIMB-GIRDLE), TYPE C, 1; MDDGC1
#605899	GLYCINE ENCEPHALOPATHY; GCE; HYPERGLYCINEMIA, TRANSIENT NEONATAL, INCLUDED; TNH, INCLUDED
#601808	CHROMOSOME 18q DELETION SYNDROME
#261600	PHENYLKETONURIA; PKU;
#216360	COACH SYNDROME
#191100	TUBEROUS SCLEROSIS 1; TSC1
#190685	DOWN SYNDROME; TRISOMY 21, INCLUDED
#147920	KABUKI SYNDROME
#122470	CORNELIA DE LANGE SYNDROME 1; CDLS1
#117550	SOTOS SYNDROME
#236670	MUSCULAR DYSTROPHY-DYSTROGLYCANOPATHY (CONGENITAL WITH BRAIN AND EYE ANOMALIES), TYPE A, 1; MDDGA1
#612475	CHROMOSOME 1q21.1 DUPLICATION SYNDROME
#607872	CHROMOSOME 1p36 DELETION SYNDROME
#600118	WARBURG MICRO SYNDROME; WARBM
#308350	EPILEPTIC ENCEPHALOPATHY, EARLY INFANTILE, 1
#253280	MUSCULAR DYSTROPHY-DYSTROGLYCANOPATHY (CONGENITAL WITH BRAIN AND EYE ANOMALIES), TYPE A, 3; MDDGA3
#235510	HENNEKAM LYMPHANGIECTASIA-LYMPHEDEMA SYNDROME
#212720	MARTSOLF SYNDROME
#208400	ASPARTYLGLUCOSAMINURIA; AGU
#145410	OPITZ GBBB SYNDROME, AUTOSOMAL DOMINANT
#115150	CARDIOFACIOCUTANEOUS SYNDROME
#301900	BORJESON-FORSSMAN-LEHMANN SYNDROME; BFLS
#188025	THROMBOCYTOPENIA, PARIS-TROUSSEAU TYPE; TCPT
#612474	CHROMOSOME 1q21.1 DELETION SYNDROME, 1.35-MB
#609425	CHROMOSOME 3q29 DELETION SYNDROME
#608363	CHROMOSOME 22q11.2 DUPLICATION SYNDROME
#310600	NORRIE DISEASE; ND
#300004	CORPUS CALLOSUM, AGENESIS OF, WITH ABNORMAL GENITALIA
#266265	CONGENITAL DISORDER OF GLYCOSYLATION, TYPE IIc; CDG2C
#256850	GIANT AXONAL NEUROPATHY 1; GAN1
#250800	METHEMOGLOBINEMIA DUE TO DEFICIENCY OF METHEMOGLOBIN REDUCTASE; METHEMOGLOBINEMIA, TYPE I, INCLUDED
#248800	MARINESCO-SJOGREN SYNDROME; MSS
#248510	MANNOSIDOSIS, BETA A, LYSOSOMAL
#236200	HOMOCYSTINURIA DUE TO CYSTATHIONINE BETA-SYNTHASE DEFICIENCY; HYPERHOMOCYSTEINEMIA, THROMBOTIC, CBS-RELATED, INCLUDED
#216400	COCKAYNE SYNDROME, TYPE A; CSA
#192430	VELOCARDIOFACIAL SYNDROME
#182212	SHPRINTZEN-GOLDBERG CRANIOSYNOSTOSIS SYNDROME
#146390	CHROMOSOME 18p DELETION SYNDROME
#606232	CHROMOSOME 22q13.3 DELETION SYNDROME
#300815	CHROMOSOME Xq28 DUPLICATION SYNDROME
#251200	MICROCEPHALY, PRIMARY AUTOSOMAL RECESSIVE, 1; MCPH1
#162200	NEUROFIBROMATOSIS, TYPE I; NF1
#103050	ADENYLOSUCCINASE DEFICIENCY

#613026	CHROMOSOME 19q13.11 DELETION SYNDROME
#612513	CHROMOSOME 2p16.1-p15 DELETION SYNDROME
#612164	EPILEPTIC ENCEPHALOPATHY, EARLY INFANTILE, 4
#611603	LISSENCEPHALY 3; LIS3
#610443	CHROMOSOME 17q21.31 DELETION SYNDROME
#610253	KLEEFSTRA SYNDROME
#609625	CHROMOSOME 10q26 DELETION SYNDROME
#608716	MICROCEPHALY, PRIMARY AUTOSOMAL RECESSIVE, 5; MCPH5
#607855	MUSCULAR DYSTROPHY, CONGENITAL MEROSIN-DEFICIENT, 1A; MDC1A; MUSCULAR DYSTROPHY, CONGENITAL, DUE TO PARTIAL LAMA2 DEFICIENCY, INCLUDED
#607426	COENZYME Q10 DEFICIENCY
#309000	LOWE OCULOCEREBRORENAL SYNDROME; OCRL
#300801	CHROMOSOME Xp11.23-p11.22 DUPLICATION SYNDROME
#300673	ENCEPHALOPATHY, NEONATAL SEVERE, DUE TO MECP2 MUTATIONS
#300337	HYPOMELANOSIS OF ITO; HMI
#276710	TYROSINEMIA, TYPE III
#270400	SMITH-LEMLI-OPITZ SYNDROME; SLOS
#269700	LIPODYSTROPHY, CONGENITAL GENERALIZED, TYPE 2; CGL2
#246560	SPLIT-HAND/FOOT MALFORMATION 3; SHFM3
#238970	HYPERORNITHINEMIA-HYPERAMMONEMIA-HOMOCITRULLINURIA SYNDROME
#219000	FRASER SYNDROME CRYPTOPHTHALMOS-SYNDACTYLY SYNDROME, INCLUDED
#213700	CEREBROTENDINOUS XANTHOMATOSIS
#210600	SECKEL SYNDROME 1; SCKL1
#200990	ACROCALLOSAL SYNDROME; ACLS
#182290	SMITH-MAGENIS SYNDROME; SMS;: SMITH-MAGENIS CHROMOSOME REGION, INCLUDED; SMCR, INCLUDED
#176270	PRADER-WILLI SYNDROME; PWS; PRADER-WILLI SYNDROME CHROMOSOME REGION, INCLUDED; PWCR, INCLUDED
#175700	GREIG CEPHALOPOLYSYNDACTYLY SYNDROME; GCPS
#160900	DYSTROPHIA MYOTONICA 1
#158900	FACIOSCAPULOHUMERAL MUSCULAR DYSTROPHY1A; FSHMD1A
#300830	CHROMOSOME Xp22 DELETION SYNDROME
#612736	GUANIDINOACETATE METHYLTRANSFERASE DEFICIENCY
#609307	SPINOCEREBELLAR ATAXIA 27; SCA27
#609192	LOEYS-DIETZ SYNDROME, TYPE 1A; LDS1A
#602481	MIGRAINE, FAMILIAL HEMIPLEGIC, 2; FHM2
#601803	PALLISTER-KILLIAN SYNDROME; PKS; HEXASOMY 12p, MOSAIC, INCLUDED
#601675	TRICHOTHIODYSTROPHY, PHOTOSENSITIVE; TTDP; ICHTHYOSIFORM ERYTHRODERMA WITH HAIR A BNORMALITY AND MENTAL AND GROWTH RETARDATION, INCLUDED
#312920	SPASTIC PARAPLEGIA 2, X-LINKED; SPG2
#312750	RETT SYNDROME; RTT
#312170	PYRUVATE DECARBOXYLASE DEFICIENCY; LACTIC ACIDEMIA, THIAMINE-RESPONSIVE, INCLUDED
#311200	OROFACIODIGITAL SYNDROME I; OFD1
#308205	ICHTHYOSIS FOLLICULARIS, ATRICHIA, AND PHOTOPHOBIA SYNDROME
#300322	LESCH-NYHAN SYNDROME; LNS; HPRT DEFICIENCY, NEUROLOGIC VARIANT, INCLUDED
#300067	LISSENCEPHALY, X-LINKED, 1; LISX1; SUBCORTICAL LAMINAR HETEROTOPIA, X-LINKED, INCLUDED; SCLH, INCLUDED
#278250	WRINKLY SKIN SYNDROME; WSS
#276880	UROCANASE DEFICIENCY
#276600	TYROSINEMIA, TYPE II
#270700	SPASTIC PARAPLEGIA 15, AUTOSOMAL RECESSIVE; SPG15
#256540	GALACTOSIALIDOSIS; GSL
#253800	MUSCULAR DYSTROPHY-DYSTROGLYCANOPATHY (CONGENITAL WITH BRAIN AND EYE ANOMALIES), TYPE A, 4; MDDGA4
#250850	METHIONINE ADENOSYLTRANSFERASE DEFICIENCY
#248600	MAPLE SYRUP URINE DISEASE
#248500	MANNOSIDOSIS, ALPHA B, LYSOSOMAL; ALPHA-MANNOSIDOSIS, TYPE II, INCLUDED
#234050	TRICHOTHIODYSTROPHY, NONPHOTOSENSITIVE 1; TTDN1
#230650	GM1-GANGLIOSIDOSIS, TYPE III
#220110	MITOCHONDRIAL COMPLEX IV DEFICIENCY
#218000	AGENESIS OF THE CORPUS CALLOSUM WITH PERIPHERAL NEUROPATHY; ACCPN
#214800	CHARGE SYNDROME

#214450	GRISCELLI SYNDROME, TYPE 1; GS1
#207800	ARGININEMIA
#204000	LEBER CONGENITAL AMAUROSIS 1; LCA1
#194190	WOLF-HIRSCHHORN SYNDROME; WHS
#158170	CHROMOSOME 9p DELETION SYNDROME
#147791	JACOBSEN SYNDROME; JBS
#613675	CHROMOSOME 17q11.2 DELETION SYNDROME, 1.4-Mb; NF1 MICRODUPLICATION SYNDROME, INCLUDED
#613436	AUTISM, SUSCEPTIBILITY TO, 17; AUTS17
#611913	CHROMOSOME 16p11.2 DELETION SYNDROME, 593-KB; AUTISM, SUSCEPTIBILITY TO, 14, INCLUDED; AUTS14, INCLUDED
#181500	SCHIZOPHRENIA; SCZD
#613174	CHROMOSOME 5p13 DUPLICATION SYNDROME
#613161	BETA-UREIDOPROPIONASE DEFICIENCY
#613038	PITUITARY HORMONE DEFICIENCY, COMBINED, 1; CPHD1
#612936	CEREBRAL PALSY, SPASTIC QUADRIPLEGIC, 3; CPSQ3
#612626	CHROMOSOME 15q26-qter DELETION SYNDROME
#612582	CHROMOSOME 6pter-p24 DELETION SYNDROME
#612379	CONGENITAL DISORDER OF GLYCOSYLATION, TYPE Iq; CDG1Q
#612285	JOUBERT SYNDROME 9; JBTS9
#611584	WAARDENBURG SYNDROME, TYPE 2E; WS2E
#610688	JOUBERT SYNDROME 6; JBTS6
#610532	LEUKODYSTROPHY, HYPOMYELINATING, 5
#610031	POLYMICROGYRIA, ASYMMETRIC
#606854	POLYMICROGYRIA, BILATERAL FRONTOPARIETAL
#605259	SPINOCEREBELLAR ATAXIA 13; SCA13
#604369	SIALURIA, FINNISH TYPE
#604360	SPASTIC PARAPLEGIA 11, AUTOSOMAL RECESSIVE; SPG11
#604168	CONGENITAL CATARACTS, FACIAL DYSMORPHISM, AND NEUROPATHY
#600513	EPILEPSY, NOCTURNAL FRONTAL LOBE, TYPE
#600462	MYOPATHY, LACTIC ACIDOSIS, AND SIDEROBLASTIC ANEMIA 1; MLASA1
#600430	BRACHYDACTYLY-MENTAL RETARDATION SYNDROME; BDMR; CHROMOSOME 2q37 DELETION SYNDROME, INCLUDED
#309900	MUCOPOLYSACCHARIDOSIS TYPE II
#309400	MENKES DISEASE
#304800	DIABETES INSIPIDUS, NEPHROGENIC, X-LINKED
#304100	CORPUS CALLOSUM, PARTIAL AGENESIS OF, X-LINKED
#300661	PHOSPHORIBOSYLPYROPHOSPHATE SYNTHETASE SUPERACTIVITY GOUT, PRPS-RELATED, INCLUDED
#300653	PHOSPHOGLYCERATE KINASE 1 DEFICIENCY
#300166	MICROPHTHALMIA, SYNDROMIC 2; MCOPS2
#275350	TRANSCOBALAMIN II DEFICIENCY
#271980	SUCCINIC SEMIALDEHYDE DEHYDROGENASE DEFICIENCY
#268300	ROBERTS SYNDROME; RBS
#266510	REFSUM DISEASE, INFANTILE FORM
#266150	PYRUVATE CARBOXYLASE DEFICIENCY
#252150	MOLYBDENUM COFACTOR DEFICIENCY
#250940	METHYLCOBALAMIN DEFICIENCY, cblG TYPE
#239500	HYPERPROLINEMIA, TYPE I; HPI
#236792	L-2-HYDROXYGLUTARIC ACIDURIA
#235800	HISTIDINEMIA
#234100	HALLERMANN-STREIFF SYNDROME; HSS
#230400	GALACTOSEMIA
#229050	FOLATE MALABSORPTION, HEREDITARY
#227650	FANCONI ANEMIA; FA
#222300	WOLFRAM SYNDROME 1; WFS1
#213300	JOUBERT SYNDROME; JBTS
#182601	SPASTIC PARAPLEGIA 4, AUTOSOMAL DOMINANT; SPG4
#179613	RECOMBINANT CHROMOSOME 8 SYNDROME
#158350	COWDEN DISEASE; CD; DYSPLASTIC GANGLIOCYTOMA OF THE CEREBELLUM, INCLUDED

#153480	BANNAYAN-RILEY-RUVALCABA SYNDROME; BRRS
#147950	KALLMANN SYNDROME 2; KAL2
#133540	COCKAYNE SYNDROME, TYPE B; CSB
#124200	DARIER-WHITE DISEASE; DAR
#107480	TOWNES-BROCKS SYNDROME; TBS
#101200	APERT SYNDROME
#613154	MUSCULAR DYSTROPHY-DYSTROGLYCANOPATHY (CONGENITAL WITH BRAIN AND EYE ANOMALIES), TYPE A, 6; MDDGA6
#613150	MUSCULAR DYSTROPHY-DYSTROGLYCANOPATHY (CONGENITAL WITH BRAIN AND EYE ANOMALIES), TYPE A, 2; MDDGA2
#300679	CHROMOSOME Xp21 DELETION SYNDROME
#190350	TRICHORHINOPHALANGEAL SYNDROME, TYPE I; TRPS1
#170100	PROLIDASE DEFICIENCY
#613254	TUBEROUS SCLEROSIS 2; TSC2
#612463	PSEUDOPSEUDOHYPOPARATHYROIDISM; PPHP
#612462	PSEUDOHYPOPARATHYROIDISM, TYPE IC; PHP1C
#612079	ALOPECIA, NEUROLOGIC DEFECTS, AND ENDOCRINOPATHY SYNDROME
#612073	MITOCHONDRIAL DNA DEPLETION SYNDROME 5 (ENCEPHALOMYOPATHIC WITH METHYLMALONIC ACIDURIA); MTDPS5
#611936	CHROMOSOME 3q29 DUPLICATION SYNDROME
#611726	EPILEPSY, PROGRESSIVE MYOCLONIC 3; EPM
#610651	XERODERMA PIGMENTOSUM, COMPLEMENTATION GROUP B; XPB
#610474	CAMPTODACTYLY, TALL STATURE, AND HEARING LOSS SYNDROME
#610198	3-@METHYLGLUTACONIC ACIDURIA, TYPE V
#610188	JOUBERT SYNDROME 5; JBTS5
#610168	LOEYS-DIETZ SYNDROME, TYPE 1B; LDS1B
#609152	HYPERTHYROIDISM, NONAUTOIMMUNE
#609029	EMANUEL SYNDROME
#608688	AICAR TRANSFORMYLASE/IMP CYCLOHYDROLAS E DEFICIENCY
#608328	WEILL-MARCHESANI SYNDROME, AUTOSOMAL DOMINANT
#608233	HERMANSKY-PUDLAK SYNDROME 2; HPS2
#608091	JOUBERT SYNDROME 2; JBTS2
#608033	ENCEPHALOPATHY, ACUTE NECROTIZING 1, SUSCEPTIBILITY TO; ANE1
#607596	PONTOCEREBELLAR HYPOPLASIA TYPE 1
#607432	LISSENCEPHALY 1; LIS1; SUBCORTICAL LAMINAR HETEROTOPIA, INCLUDED; SCLH, INCLUDED
#607330	LATHOSTEROLOSIS
#606812	FUMARASE DEFICIENCY
#606369	EPILEPTIC ENCEPHALOPATHY, LENNOX-GASTA UT TYPE
#603513	CEREBRAL PALSY, SPASTIC QUADRIPLEGIC, 1; CPSQ1
#602849	MUENKE SYNDROME
#602668	DYSTROPHIA MYOTONICA 2; DM2
#601186	MICROPHTHALMIA, SYNDROMIC 9; MCOPS9
#600721	D-2-@HYDROXYGLUTARIC ACIDURIA 1
#600002	EIKEN SKELETAL DYSPLASIA
#500001	LEBER OPTIC ATROPHY AND DYSTONIA
#312080	PELIZAEUS-MERZBACHER DISEASE; PMD
#311300	OTOPALATODIGITAL SYNDROME, TYPE I; OPD 1
#311250	ORNITHINE TRANSCARBAMYLASE DEFICIENCY, HYPERAMMONEMIA DUE TO VALPROATE SENSITIVITY, INCLUDED
#309801	MICROPHTHALMIA, SYNDROMIC 7; MCOPS7
#308050	CONGENITAL HEMIDYSPLASIA WITH ICHTHYOS IFORM ERYTHRODERMA AND LIMB DEFECTS
#305600	FOCAL DERMAL HYPOPLASIA; FDH
#305400	FACIOGENITAL DYSPLASIA
#302960	CHONDRODYSPLASIA PUNCTATA 2, X-LINKED DOMINANT; CDPX2
#300607	HYPEREKPLEXIA AND EPILEPSY
#300438	17-@BETA-HYDROXYSTEROID DEHYDROGENASE X DEFICIENCY
#300240	HOYERAAL-HREIDARSSON SYNDROME; HHS
#300209	SIMPSON-GOLABI-BEHMEL SYNDROME, TYPE 2
#277600	WEILL-MARCHESANI SYNDROME, AUTOSOMAL RECESSIVE
#277400	METHYLMALONIC ACIDURIA AND HOMOCYSTINURIA, cblC TYPE

#274600	PENDRED SYNDROME; PDS
#274270	DIHYDROPYRIMIDINE DEHYDROGENASE DEFICIENCY5-@FLUOROURACIL TOXICITY, INCLUDED
#274000	CHROMOSOME 1q21.1 DELETION SYNDROME, 200-KB; TETRAPHOCOMELIA-THROMBOCYTOPENIA SYNDROME, INCLUDED
#271930	STRIATONIGRAL DEGENERATION, INFANTILE; SNDI
#271245	MITOCHONDRIAL DNA DEPLETION SYNDROME 7(HEPATOCEREBRAL TYPE); MTDPS7
#269150	SCHINZEL-GIEDION MIDFACE RETRACTION SYNDROME
#266130	GLUTATHIONE SYNTHETASE DEFICIENCY
#261640	HYPERPHENYLALANINEMIA, BH4-DEFICIENT, A; HPABH4A
#261630	HYPERPHENYLALANINEMIA, BH4-DEFICIENT, C; HPABH4C
#261540	PETERS-PLUS SYNDROME
#257200	NIEMANN-PICK DISEASE, TYPE A
#256731	CEROID LIPOFUSCINOSIS, NEURONAL, 5; CLN5
#256730	CEROID LIPOFUSCINOSIS, NEURONAL, 1; CL
#256710	ELEJALDE DISEASE
#256600	NEURODEGENERATION WITH BRAIN IRON MULATION 2A; NBIA2A ACCU
#256000	LEIGH SYNDROME; LS
#253220	MUCOPOLYSACCHARIDOSIS TYPE VII
#252600	MUCOLIPIDOSIS III ALPHA/BETA
#251450	DESBUQUOIS DYSPLASIA; DBQD
#251260	NIJMEGEN BREAKAGE SYNDROME
#245400	MITOCHONDRIAL DNA DEPLETION SYNDROME 9(ENCEPHALOMYOPATHIC TYPE WITH METHYLMALONIC ACIDURIA); MTDPS9
#245348	PYRUVATE DEHYDROGENASE E2 DEFICIENCY
#243800	JOHANSON-BLIZZARD SYNDROME; JBS
#242600	IMINOGLYCINURIA
#239510	HYPERPROLINEMIA, TYPE II; HPII
#238700	HYPERLYSINEMIA
#233910	HYPERPHENYLALANINEMIA, BH4-DEFICIENT, B; HPABH4B; DYSTONIA, DOPA-RESPONSIVE, WITH OR WITHOUT HYPERPHENYLALANINEMIA, AUTOSOMAL RECESSIVE, INCLUDED
#231550	ACHALASIA-ADDISONIANISM-ALACRIMA SYNDROME;
#230350	GALACTOSE EPIMERASE DEFICIENCY
#230000	FUCOSIDOSIS
#228000	FARBER LIPOGRANULOMATOSIS
#222748	DIHYDROPYRIMIDINASE DEFICIENCY
#221750	PITUITARY HORMONE DEFICIENCY, COMBINED, 3; CPHD3
#219200	CUTIS LAXA, AUTOSOMAL RECESSIVE, TYPE IIA;
#218600	BALLER-GEROLD SYNDROME; BGS
#215700	CITRULLINEMIA, CLASSIC
#215100	RHIZOMELIC CHONDRODYSPLASIA PUNCTATA, TYPE 1; RCDP1
#214100	ZELLWEGER SYNDROME; ZS
#212066	CONGENITAL DISORDER OF GLYCOSYLATION, TYPE IIa; CDG2A
#210720	MICROCEPHALIC OSTEODYSPLASTIC PRIMORDIAL DWARFISM, TYPE II
#210200	3-@METHYLCROTONYL-CoA CARBOXYLASE 1 DEFICIENCY
#207900	ARGININOSUCCINIC ACIDURIA
#193700	ARTHROGRYPOSIS, DISTAL, TYPE 2A; DA2A
#187600	THANATOPHORIC DYSPLASIA, TYPE I; TD1; ACHONDROPLASIA, SEVERE, WITH DEVELOPMENTAL DELAY AND ACANTHOSIS NIGRICANS, INCLUDED; SADDAN, INCLUDED
#164200	OCULODENTODIGITAL DYSPLASIA; ODDD
#163950	NOONAN SYNDROME 1; NS1
#157170	HOLOPROSENCEPHALY 2; HPE2
#151100	LEOPARD SYNDROME 1
#150230	TRICHORHINOPHALANGEAL SYNDROME, TYPE II; TRPS2
#148820	WAARDENBURG SYNDROME, TYPE 3; WS3
#136760	FRONTONASAL DYSPLASIA 1; FND1
#135100	FIBRODYSPLASIA OSSIFICANS PROGRESSIVA; FOP
#115470	CAT EYE SYNDROME; CES
#113620	BRANCHIOOCULOFACIAL SYNDROME; BOFS
#104290	ALTERNATING HEMIPLEGIA OF CHILDHOOD
#103580	PSEUDOHYPOPARATHYROIDISM, TYPE IA; PHP1A

#101600	PFEIFFER SYNDROME
#613564	CHROMOSOME 2p12-p11.2 DELETION SYNDROME
#612345	CHROMOSOME 2q31.2 DELETION SYNDROME
#610042	CORTICAL DYSPLASIA-FOCAL EPILEPSY SYNDROME
#312870	SIMPSON-GOLABI-BEHMEL SYNDROME, TYPE 1; SGBS1
#300422	FG SYNDROME 4; FGS4
#253200	MUCOPOLYSACCHARIDOSIS TYPE VI
#229100	FORMIMINOTRANSFERASE DEFICIENCY
#209900	BARDET-BIEDL SYNDROME; BBS
#176200	PORPHYRIA VARIEGATA
#110100	BLEPHAROPHIMOSIS, PTOSIS, AND EPICANTHUS INVERSUS; BPES
#106210	ANIRIDIA; AN
#613180	POLYMICROGYRIA WITH OPTIC NERVE HYPOPLASIA
#613179	PURINE NUCLEOSIDE PHOSPHORYLASE DEFICIENCY
#613090	BARTTER SYNDROME, TYPE 4B
#613068	NEURODEGENERATION DUE TO CEREBRAL FOLATE TRANSPORT DEFICIENCY
#612716	DYSTONIA, DOPA-RESPONSIVE, DUE TO SEPIAPTERIN REDUCTASE DEFICIENCY
#612233	LEUKODYSTROPHY, HYPOMYELINATING, 4
#610127	CEROID LIPOFUSCINOSIS, NEURONAL, 10; CLN10
#609975	HYPERINSULINEMIC HYPOGLYCEMIA, FAMILIAL, 4; HHF4
#609528	CEREBRAL DYSGENESIS, NEUROPATHY, ICHTHYOSIS, AND PALMOPLANTAR KERATODERMA SYNDROME
#609460	GOLDBERG-SHPRINTZEN MEGACOLON SYNDROME
#609241	SCHINDLER DISEASE, TYPE I; ALPHA-N-ACETYLGALACTOSAMINIDASE DEFICIENCY, TYPE III, INCLUDED
#608782	PYRUVATE DEHYDROGENASE PHOSPHATASE DEFICIENCY
#608629	JOUBERT SYNDROME 3; JBTS3
#608393	MICROCEPHALY, PRIMARY AUTOSOMAL RECESSIVE, 6; MCPH6
#608093	CONGENITAL DISORDER OF GLYCOSYLATION, TYPE Ij; CDG1J
#607341	FOCAL CORTICAL DYSPLASIA OF TAYLOR; FCDT
#607095	ANAUXETIC DYSPLASIA
#606762	HYPERINSULINEMIC HYPOGLYCEMIA, FAMILIAL, 6; HHF6
#604804	MICROCEPHALY, PRIMARY AUTOSOMAL RECESSIVE, 3; MCPH3
#604292	ECTRODACTYLY, ECTODERMAL DYSPLASIA, AND CLEFT LIP/PALATE SYNDROME 3; EEC3
#604004	MEGALENCEPHALIC LEUKOENCEPHALOPATHY WITH SUBCORTICAL CYSTS; MLC
#602522	BARTTER SYNDROME, TYPE 4A
#602485	HYPERINSULINEMIC HYPOGLYCEMIA, FAMILIAL, 3; HHF3
#602473	ENCEPHALOPATHY, ETHYLMALONIC
#601815	PHOSPHOGLYCERATE DEHYDROGENASE DEFICIENCY
#601776	EHLERS-DANLOS SYNDROME, MUSCULOCONTRAC TURAL TYPE
#601678	BARTTER SYNDROME, ANTENATAL, TYPE 1
#598500	
#500007	CYCLIC VOMITING SYNDROME; CVS
#308300	INCONTINENTIA PIGMENTI; IP
#307030	HYPERGLYCEROLEMIA
#305620	FRONTOMETAPHYSEAL DYSPLASIA; FMD
#304120	OTOPALATODIGITAL SYNDROME, TYPE II; OPD2
#302350	NANCE-HORAN SYNDROME; NHS
#300496	AUTISM, SUSCEPTIBILITY TO, X-LINKED 3; AUTSX3
#300425	AUTISM, SUSCEPTIBILITY TO, X-LINKED 1; AUTSX1
#300373	OSTEOPATHIA STRIATA WITH CRANIAL SCLEROSIS; OSCS
#277590	WEAVER SYNDROME
#277410	METHYLMALONIC ACIDURIA AND HOMOCYSTINU RIA, cblD TYPE
#275630	CHANARIN-DORFMAN SYNDROME; CDS
#275100	HYPOTHYROIDISM, CONGENITAL, NONGOITROUS, 4; CHNG4
#274900	THYROID DYSHORMONOGENESIS 5; TDH5
#274800	THYROID DYSHORMONOGENESIS 4; TDH4
#274700	THYROID DYSHORMONOGENESIS 3; TDH3

#274400	THYROID DYSHORMONOGENESIS 1; TDH1
#270550	SPASTIC ATAXIA, CHARLEVOIX-SAGUENAY TYPE; SACS
#270200	SJOGREN-LARSSON SYNDROME; SLS
#268400	ROTHMUND-THOMSON SYNDROME; RTS
#268310	ROBINOW SYNDROME, AUTOSOMAL RECESSIVE; RRS
#266100	EPILEPSY, PYRIDOXINE-DEPENDENT; EPD
#264470	PEROXISOMAL ACYL-CoA OXIDASE DEFICIENCY
#260400	SHWACHMAN-DIAMOND SYNDROME; SDS
#259770	OSTEOPOROSIS-PSEUDOGLIOMA SYNDROME; OPPG
#256800	INSENSITIVITY TO PAIN, CONGENITAL, WITH ANHIDROSIS; CIPA
#256550	NEURAMINIDASE DEFICIENCY SIALIDOSIS, TYPE I, INCLUDED
#256450	HYPERINSULINEMIC HYPOGLYCEMIA, FAMILIAL, 1; HHF1
#255800	SCHWARTZ-JAMPEL SYNDROME, TYPE 1; SJS1
#252940	MUCOPOLYSACCHARIDOSIS TYPE IIID
#252930	MUCOPOLYSACCHARIDOSIS TYPE IIIC
#252920	MUCOPOLYSACCHARIDOSIS TYPE IIIB
#252900	MUCOPOLYSACCHARIDOSIS TYPE IIIA
#252650	MUCOLIPIDOSIS IV
#252605	MUCOLIPIDOSIS III GAMMA
#247200	MILLER-DIEKER LISSENCEPHALY SYNDROME;
#245349	PYRUVATE DEHYDROGENASE E3-BINDING PROTEIN DEFICIENCY
#245150	KEUTEL SYNDROME
#242100	ICHTHYOSIFORM ERYTHRODERMA, CONGENITAL, NONBULLOUS, 1; NCIE1
#241200	BARTTER SYNDROME, ANTENATAL, TYPE 2
#240800	HYPOGLYCEMIA, LEUCINE-INDUCED; LIH
#237300	CARBAMOYL PHOSPHATE SYNTHETASE I DEFICIENCY, HYPERAMMONEMIA DUE TO
#230500	GM1-GANGLIOSIDOSIS, TYPE I
#225500	ELLIS-VAN CREVELD SYNDROME; EVC
#225250	HYPOTHYROIDISM, CONGENITAL, NONGOITROUS, 5; CHNG5
#222765	RHIZOMELIC CHONDRODYSPLASIA PUNCTATA, TYPE 2; RCDP2
#219500	CYSTATHIONINURIA
#214150	CEREBROOCULOFACIOSKELETAL SYNDROME 1; COFS1
#210900	BLOOM SYNDROME; BLM
#208050	ARTERIAL TORTUOSITY SYNDROME; ATS
#207410	ANTLEY-BIXLER SYNDROME WITHOUT GENITAL ANOMALIES OR DISORDERED STEROIDOGENESIS; ABS2
#204200	CEROID LIPOFUSCINOSIS, NEURONAL, 3; CLN3
#204100	LEBER CONGENITAL AMAUROSIS 2; LCA2
#203750	ALPHA-METHYLACETOACETIC ACIDURIA
#202370	ADRENOLEUKODYSTROPHY, AUTOSOMAL NEONATAL FORM
#175780	PORENCEPHALY, FAMILIAL
#166780	OTOFACIOCERVICAL SYNDROME
#150250	LARSEN SYNDROME, AUTOSOMAL DOMINANT; LRS
#146000	HYPOCHONDROPLASIA; HCH
#129850	EDINBURGH MALFORMATION SYNDROME
#125800	DIABETES INSIPIDUS, NEPHROGENIC, AUTOSOMAL
#124000	MITOCHONDRIAL COMPLEX III DEFICIENCY
#123500	CROUZON SYNDROME
#120330	PAPILLORENAL SYNDROME
#118450	ALAGILLE SYNDROME 1; ALGS1
#109400	BASAL CELL NEVUS SYNDROME; BCNS
#105650	DIAMOND-BLACKFAN ANEMIA; DBA
#101400	SAETHRE-CHOTZEN SYNDROME; SCS
#613729	CHROMOSOME 7q11.23 DELETION SYNDROME, DISTAL, 1.2-MB
#613604	CHROMOSOME 16p12.2-p11.2 DELETION SYNDROME, 7.1- TO 8.7-MB
#612313	CHROMOSOME 2q32-q33 DELETION SYNDROME
#611881	GLYCOGEN STORAGE DISEASE XII; GSD12

#610883	POTOCKI-LUPSKI SYNDROME; PTLS
#609334	CHROMOSOME 18 PERICENTRIC INVERSION
#607208	DRAVET SYNDROME; EPILEPSY, INTRACTABLE CHILDHOOD, WITH GENERALIZED TONIC-CLONIC SEIZURES
#605021	MYOCLONIC EPILEPSY, FAMILIAL INFANTILE; FIME
#312000	PANHYPOPITUITARISM, X-LINKED; PHPX
#304150	OCCIPITAL HORN SYNDROME; OHS
#302950	CHONDRODYSPLASIA PUNCTATA 1, X-LINKED RECESSIVE; CDPX1
#300623	FRAGILE X TREMOR/ATAXIA SYNDROME; FXTAS
#300448	ALPHA-THALASSEMIA MYELODYSPLASIA SYNDROME; ATMDS
#300376	MUSCULAR DYSTROPHY, BECKER TYPE; BMD
#300215	LISSENCEPHALY, X-LINKED, 2; LISX2
#269160	SCHIZENCEPHALY
#217095	CONOTRUNCAL HEART MALFORMATIONS; CTHM
#212112	CARDIOMYOPATHY, DILATED, WITH HYPERGON ADOTROPIC HYPOGONADISM
#208900	ATAXIA-TELANGIECTASIA; AT
#185500	SUPRAVALVULAR AORTIC STENOSIS; SVAS
#182900	SPHEROCYTOSIS, TYPE 1; SPH1
#181450	ULNAR-MAMMARY SYNDROME; UMS
#175100	ADENOMATOUS POLYPOSIS OF THE COLON; APC
#133701	EXOSTOSES, MULTIPLE, TYPE II
#125370	DENTATORUBRAL-PALLIDOLUYSIAN ATROPHY; DRPLA
#121200	EPILEPSY, BENIGN NEONATAL, 1; EBN1
#613732	JMP SYNDROME
#613700	SUPERNUMERARY DER(22)t(8;22) SYNDROME
#613638	CHROMOSOME 19p13.13 DELETION SYNDROME
#613444	CHROMOSOME 16p11.2 DELETION SYNDROME, 220-KB
#613406	CHROMOSOME 15q24 DELETION SYNDROME
#613153	MUSCULAR DYSTROPHY-DYSTROGLYCANOPATHY (CONGENITAL WITH BRAIN AND EYE ANOMALIES), TYPE A, 5; MDDGA5
#612718	ARGININE:GLYCINE AMIDINOTRANSFERASE DEFICIENCY
#612337	CHROMOSOME 1q43-q44 DELETION SYNDROME
#612126	GLUT1 DEFICIENCY SYNDROME 2; GLUT1DS2
#611588	MUSCULAR DYSTROPHY-DYSTROGLYCANOPATHY(LIMB-GIRDLE), TYPE C, 4; MDDGC4
#611560	JOUBERT SYNDROME 7; JBTS7
#611209	CONGENITAL DISORDER OF GLYCOSYLATION, TYPE IIg; CDG2G
#611102	DEAFNESS, SENSORINEURAL, AND MALE INFERTILITY
#609757	WILLIAMS-BEUREN REGION DUPLICATION SYNDROME
#607014	HURLER SYNDROME
#605289	SPLIT-HAND/FOOT MALFORMATION 4; SHFM4
#600850	SCHIZOPHRENIA 4; SCZD4
#600651	FRAGILE SITE 11B; FRA11B
#500002	MITOCHONDRIAL MYOPATHY WITH DIABETES
#305100	ECTODERMAL DYSPLASIA, HYPOHIDROTIC, X- LINKED; XHED
#305000	DYSKERATOSIS CONGENITA, X-LINKED; DKC
#300615	BRUNNER SYNDROME
#300321	FG SYNDROME 2; FGS2
#300048	INTESTINAL PSEUDOOBSTRUCTION, NEURONAL, CHRONIC IDIOPATHIC, X-LINKED
#300000	OPITZ GBBB SYNDROME, X-LINKED
#278700	XERODERMA PIGMENTOSUM, COMPLEMENTATION GROUP A; XPA
#269000	SC PHOCOMELIA SYNDROME
#257320	LISSENCEPHALY 2; LIS2
#255200	MYOPATHY, CENTRONUCLEAR, AUTOSOMAL RECESSIVE
#253250	MULIBREY NANISM
#249700	LANGER MESOMELIC DYSPLASIA
#249420	FRANK-TER HAAR SYNDROME; FTHS
#242900	IMMUNOOSSEOUS DYSPLASIA, SCHIMKE TYPE
#230200	GALACTOKINASE DEFICIENCY

#212065	CONGENITAL DISORDER OF GLYCOSYLATION, TYPE Ia; CDG1A
#208920	ATAXIA, EARLY-ONSET, WITH OCULOMOTOR APRAXIA AND HYPOALBUMINEMIA; EAOH
#194070	WILMS TUMOR 1; WT1
#182600	SPASTIC PARAPLEGIA 3, AUTOSOMAL DOMINANT; SPG3A
#180860	SILVER-RUSSELL SYNDROME; SRS
#180500	AXENFELD-RIEGER SYNDROME, TYPE 1; RIEG1
#143465	ATTENTION DEFICIT-HYPERACTIVITY DISORDER; ADHD
#130070	EHLERS-DANLOS SYNDROME, PROGEROID FORM
#120200	COLOBOMA, OCULAR
#113300	BRACHYDACTYLY, TYPE E1; BDE1
#107250	ANTERIOR SEGMENT MESENCHYMAL DYSGENESIS; ASMD
#613728	SPINOCEREBELLAR ATAXIA, AUTOSOMAL RECESSIVE 10; SCAR10
#613684	RUBINSTEIN-TAYBI SYNDROME 2; RSTS2
#613676	SECKEL SYNDROME 4; SCKL4
#613674	VESICOURETERAL REFLUX 3; VUR3
#613615	SENIOR-LOKEN SYNDROME 7; SLSN7
#613612	CONGENITAL DISORDER OF GLYCOSYLATION, TYPE IIi; CDG2I
#613603	CHROMOSOME 4q32.1-q32.2 TRIPLICATION SYNDROME
#613533	CHROMOSOME 17q21.31 DUPLICATION SYNDROME
#613477	EPILEPTIC ENCEPHALOPATHY, EARLY INFANTILE, 5
#613470	HEMOLYTIC ANEMIA, NONSPHEROCYTIC, DUE TO GLUCOSE PHOSPHATE ISOMERASE DEFICIENCY
#613458	CHROMOSOME 16p13.3 DUPLICATION SYNDROME
#613457	CHROMOSOME 14q11-q22 DELETION SYNDROME
#613456	FRONTONASAL DYSPLASIA 3; FND3
#613454	RETT SYNDROME, CONGENITAL VARIANT
#613451	FRONTONASAL DYSPLASIA 2; FND2
#613402	MICROCEPHALY, SEIZURES, AND DEVELOPMEN TAL DELAY; MCSZ
#613224	NOONAN SYNDROME 6; NS6
#613215	CHROMOSOME 17p13.3 DUPLICATION SYNDROME
#613204	MUSCULAR DYSTROPHY, CONGENITAL, DUE TO INTEGRIN ALPHA-7 DEFICIENCY
#613159	NEPHRONOPHTHISIS-LIKE NEPHROPATHY 1; NPHPL1
#613158	MUSCULAR DYSTROPHY-DYSTROGLYCANOPATHY(LIMB-GIRDLE), TYPE C, 2; MDDGC2
#613157	MUSCULAR DYSTROPHY-DYSTROGLYCANOPATHY (LIMB-GIRDLE), TYPE C, 3; MDDGC3
#612951	LEUKOENCEPHALOPATHY, CYSTIC, WITHOUT MEGALENCEPHALY
#612940	CUTIS LAXA, AUTOSOMAL RECESSIVE, TYPE IIB; ARCL2B
#612900	CEREBRAL PALSY, SPASTIC QUADRIPLEGIC, 2; CPSQ2
#612863	CHROMOSOME 6q24-q25 DELETION SYNDROME
#612530	CHROMOSOME 1q41-q42 DELETION SYNDROME
#611182	CONGENITAL DISORDER OF GLYCOSYLATION, TYPE IIh; CDG2H
#611040	MICROPHTHALMIA, ISOLATED 5; MCOP5
#610828	HOLOPROSENCEPHALY 7; HPE7
#610759	CORNELIA DE LANGE SYNDROME 3; CDLS3
#610733	NOONAN SYNDROME 4; NS4
#610623	CATARACT, POSTERIOR POLAR, 4; CTPP4
#610549	DIABETES MELLITUS, INSULIN-RESISTANT, WITH ACANTHOSIS NIGRICANS
#610543	CHROMOSOME 16p13.3 DELETION SYNDROME
#609924	AMINOACYLASE 1 DEFICIENCY
#609136	PERIPHERAL DEMYELINATING NEUROPATHY, CENTRAL DYSMYELINATION, WAARDENBURG SYNDROME, AND HIRSCHSPRUNG DISEASE; PCWH
#608808	TRANSPOSITION OF THE GREAT ARTERIES, DEXTRO-LOOPED 1; DTGA1
#608804	LEUKODYSTROPHY, HYPOMYELINATING, 2
#608594	LIPODYSTROPHY, CONGENITAL GENERALIZED, TYPE 1; CGL1
#608540	CONGENITAL DISORDER OF GLYCOSYLATION, TYPE Ik; CDG1K
#608149	UNIPARENTAL DISOMY, PATERNAL, CHROMOSOME 14
#607906	CONGENITAL DISORDER OF GLYCOSYLATION, TYPE Ii; CDG1I
#607734	CHARCOT-MARIE-TOOTH DISEASE, DEMYELINATING, TYPE 1F; CMT1F
#607483	BASAL GANGLIA DISEASE, BIOTIN-RESPONSIVE

#607364	BARTTER SYNDROME, TYPE 3
#607326	SMITH-MCCORT DYSPLASIA; SMC
#607155	MUSCULAR DYSTROPHY-DYSTROGLYCANOPATHY(LIMB-GIRDLE), TYPE C, 5; MDDGC5
#607136	SPINOCEREBELLAR ATAXIA 17; SCA17
#607091	CONGENITAL DISORDER OF GLYCOSYLATION, TYPE IId; CDG2D
#606217	ATRIOVENTRICULAR SEPTAL DEFECT, SUSCEP TIBILITY TO, 2; AVSD2
#604321	MICROCEPHALY, PRIMARY AUTOSOMAL RECESSIVE, 4; MCPH4
#603903	SICKLE CELL ANEMIA
#602579	CONGENITAL DISORDER OF GLYCOSYLATION, TYPE Ib; CDG1B
#602081	SPEECH-LANGUAGE DISORDER 1; SPCH1
#602014	HYPOMAGNESEMIA 1, INTESTINAL; HOMG1
#601536	ATHABASKAN BRAINSTEM DYSGENESIS SYNDROME; ABDS
#600316	DEAFNESS, AUTOSOMAL RECESSIVE 3; DFNB3
#600309	ATRIOVENTRICULAR SEPTAL DEFECT; AVSD
#520000	DIABETES-DEAFNESS SYNDROME, MATERNALLY TRANSMITTED
#311360	PREMATURE OVARIAN FAILURE 1; POF1
#311070	CHARCOT-MARIE-TOOTH DISEASE, X-LINKED RECESSIVE, 5; CMTX5
#310400	MYOTUBULAR MYOPATHY 1; MTM1
#310300	EMERY-DREIFUSS MUSCULAR DYSTROPHY, 1; EDMD1
#308930	LEIGH SYNDROME, X-LINKED
#304110	CRANIOFRONTONASAL SYNDROME; CFNS
#303100	CHOROIDEREMIA; CHM
#302800	CHARCOT-MARIE-TOOTH DISEASE, X-LINKED DOMINANT, 1; CMTX1
#301835	ARTS SYNDROME; ARTS
#301830	SPINAL MUSCULAR ATROPHY, X-LINKED 2; SMAX2
#301050	ALPORT SYNDROME, X-LINKED; ATS
#300804	JOUBERT SYNDROME 10; JBTS10
#300590	CORNELIA DE LANGE SYNDROME 2; CDLS2
#300555	DENT DISEASE 2
#300500	ALBINISM, OCULAR, TYPE I; OA1
#300068	ANDROGEN INSENSITIVITY SYNDROME; AIS
#278780	XERODERMA PIGMENTOSUM, COMPLEMENTATION GROUP G; XPG
#278760	XERODERMA PIGMENTOSUM, COMPLEMENTATION GROUP F; XPF
#271900	CANAVAN DISEASE
#271700	SPONDYLOPERIPHERAL DYSPLASIA
#270800	SPASTIC PARAPLEGIA 5A, AUTOSOMAL RECESSIVE; SPG5A
#269920	INFANTILE SIALIC ACID STORAGE DISORDER
#267430	RENAL TUBULAR DYSGENESIS; RTD
#267300	RENAL TUBULAR ACIDOSIS, DISTAL, WITH PROGRESSIVE NERVE DEAFNESS
#266900	SENIOR-LOKEN SYNDROME 1; SLSN1
#262000	BJORNSTAD SYNDROME; BJS
#261000	INTRINSIC FACTOR DEFICIENCY; IFD
#258870	GYRATE ATROPHY OF CHOROID AND RETINA; GACR
#258315	OMODYSPLASIA 1; OMOD1
#257220	NIEMANN-PICK DISEASE, TYPE C1; NPC1
#255310	MYOPATHY, CONGENITAL, WITH FIBER-TYPE DISPROPORTION; CFTD
#253700	MUSCULAR DYSTROPHY, LIMB-GIRDLE, TYPE2C; LGMD2C
#253260	BIOTINIDASE DEFICIENCY; MULTIPLE CARBOXYLASE DEFICIENCY, LATE- ONSET
#253230	MUCOPOLYSACCHARIDOSIS TYPE VIII
#252700	MUCOPOLYSACCHARIDOSES, UNCLASSIFIED TYPES
#252500	MUCOLIPIDOSIS II ALPHA/BETA
#252010	MITOCHONDRIAL COMPLEX I DEFICIENCY
#251290	BAND-LIKE CALCIFICATION WITH SIMPLIFIE D GYRATION AND POLYMICROGYRIA; BLCPMG
#250100	METACHROMATIC LEUKODYSTROPHY
#247100	LIPOID PROTEINOSIS OF URBACH AND WIETHE
#246700	CHYLOMICRON RETENTION DISEASE; CMRD

#237900	HYPERBILIRUBINEMIA, TRANSIENT FAMILIAL NEONATAL
#236250	HOMOCYSTINURIA DUE TO DEFICIENCY OF N(5,10)-METHYLENETETRAHYDROFOLATE REDUCTASE ACTIVITY
#234500	HARTNUP DISORDER
#234200	NEURODEGENERATION WITH BRAIN IRON ACCUMULATION 1; NBIA1
#231670	GLUTARIC ACIDEMIA I
#231300	GLAUCOMA 3, PRIMARY CONGENITAL, A; GLC3A
#225400	EHLERS-DANLOS SYNDROME, TYPE VI; EDS6
#225300	SPLIT-HAND/FOOT MALFORMATION 6; SHFM6
#222700	LYSINURIC PROTEIN INTOLERANCE; LPI
#215400	CHORDOMA, SUSCEPTIBILITY TO; CHDM
#210210	3-@METHYLCROTONYL-CoA CARBOXYLASE 2 DEFICIENCY
#203700	MITOCHONDRIAL DNA DEPLETION SYNDROME 4 A (ALPERS TYPE); MTDPS4A
#203100	ALBINISM, OCULOCUTANEOUS, TYPE IA; OCA1A
#201000	CARPENTER SYNDROME
#194080	DENYS-DRASH SYNDROME; DDS
#190351	TRICHORHINOPHALANGEAL SYNDROME, TYPE III; TRPS3
#188570	THYROID HORMONE RESISTANCE, GENERALIZED, AUTOSOMAL DOMINANT; GRTH
#188400	DIGEORGE SYNDROME; DGS
#184250	SPONDYLOEPIMETAPHYSEAL DYSPLASIA, STRUDWICK TYPE
#183900	SPONDYLOEPIPHYSEAL DYSPLASIA CONGENITA; SEDC
#176000	PORPHYRIA, ACUTE INTERMITTENT
#173100	ISOLATED GROWTH HORMONE DEFICIENCY, TYPE II; IGHD2
#172800	PIEBALD TRAIT; PBT
#169400	PELGER-HUET ANOMALY; PHA
#167200	PACHYONYCHIA CONGENITA, TYPE 1; PC1
#164400	SPINOCEREBELLAR ATAXIA 1; SCA1
#163800	SICK SINUS SYNDROME 2, AUTOSOMAL DOMINANT; SSS2
#162400	NEUROPATHY, HEREDITARY SENSORY AND AUTONOMIC, TYPE IA; HSAN1A
#162210	NEUROFIBROMATOSIS, FAMILIAL SPINAL
#157640	PROGRESSIVE EXTERNAL OPHTHALMOPLEGIA WITH MITOCHONDRIAL DNA DELETIONS, AUTOSOMAL DOMINANT, 1; PEOA1
#154700	MARFAN SYNDROME; MFS
#148210	KERATITIS-ICHTHYOSIS-DEAFNESS SYNDROME, AUTOSOMAL DOMINANT
#147250	SOLITARY MEDIAN MAXILLARY CENTRAL INCISOR; SMMCI
#147060	HYPER-IgE RECURRENT INFECTION SYNDROME, AUTOSOMAL DOMINANT
#146510	PALLISTER-HALL SYNDROME; PHS
#146255	HYPOPARATHYROIDISM, SENSORINEURAL DEAFNESS, AND RENAL DISEASE
#146200	HYPOPARATHYROIDISM, FAMILIAL ISOLATED; FIH
#142945	HOLOPROSENCEPHALY 3; HPE3
#142900	HOLT-ORAM SYNDROME; HOS
#142623	HIRSCHSPRUNG DISEASE, SUSCEPTIBILITY TO, 1; HSCR1
#141500	MIGRAINE, FAMILIAL HEMIPLEGIC, 1; FHM1
#137580	GILLES DE LA TOURETTE SYNDROME; GTS
#136610	FRAGILE SITE 2q11
#136570	CHROMOSOME 16p12.1 DELETION SYNDROME, 520-KB; FRAGILE SITE 16p12, INCLUDED
#135400	HYPERTRICHOSIS TERMINALIS, GENERALIZED, WITH OR WITHOUT GINGIVAL HYPERPLASIA
#133700	EXOSTOSES, MULTIPLE, TYPE I
#130650	BECKWITH-WIEDEMANN SYNDROME; BWS
#128235	DYSTONIA 12; DYT12
#123150	JACKSON-WEISS SYNDROME; JWS
#121800	CORNEAL DYSTROPHY, CRYSTALLINE, OF SCHNYDER
#120430	COLOBOMA OF OPTIC NERVE; OPTIC NERVE HEAD PITS, BILATERAL CONGENITAL, INCLUDED
#119600	CLEIDOCRANIAL DYSPLASIA; CCD
#118600	CHONDROCALCINOSIS 2; CCAL2
#118400	CHERUBISM
#116920	LEUKOCYTE ADHESION DEFICIENCY, TYPE I; LAD
#112500	BRACHYDACTYLY, TYPE A1; BDA1

#108300	STICKLER SYNDROME, TYPE I; STL1
#104300	ALZHEIMER DISEASE; AD
#101000	NEUROFIBROMATOSIS, TYPE II; NF2

Table S3. 1 – CNVs found for patient 1.

Chromosome Region	Event	Length	Genes	Probes	Cytoband	Gene Symbols
chr1:72,532,314-72,590,954	Homozygous Copy Loss	58641	0	3	p31.1	
chr1:147,255,457-147,753,238	CN Gain	497782	2	21	q21.1	*FCGR1C, LOC388692*
chr1:194,975,299-195,078,812	CN Gain	103514	3	10	q31.3	*CFH, CFHR1, CFHR3*
chr1:246,667,935-246,745,152	CN Gain	77218	3	6	q44	*OR2T2, OR2T3, OR2T5*
chr1:246,745,152-246,882,848	CN Loss	137697	7	9	q44	*OR2G6, OR2T10, OR2T11, OR2T27, OR2T29, OR2T34, OR2T35*
chr2:242,462,319-242,486,257	CN Loss	23939	1	4	q37.3	*C2orf85*
chr4:34,448,772-34,500,505	CN Loss	51734	0	4	p15.1	
chr4:69,038,815-69,170,341	CN Gain	131527	2	10	q13.2	*TMPRSS11E, UGT2B17*
chr4:70,195,060-70,269,497	CN Loss	74438	1	5	q13.2	*UGT2B28*
chr5:745,872-874,147	CN Loss	128276	2	12	p15.33	*TPPP, ZDHHC11*
chr5:68,866,395-68,960,939	CN Gain	94545	7	5	q13.2	*GTF2H2, GTF2H2B, GTF2H2C, GTF2H2D, LOC647859, LOC653391, OCLN*
chr6:0-193,938	CN Gain	193939	0	6	p25.3	
chr6:327,656-666,208	CN Gain	338553	3	20	p25.3	*EXOC2, HUS1B, IRF4*
chr8:7,121,948-7,803,775	CN Loss	681828	27	25	p23.1	*DEFB103A, DEFB103B, DEFB104A, DEFB104B, DEFB105A, DEFB105B, DEFB106A, DEFB106B, DEFB107A, DEFB107B, DEFB109P1B, DEFB4, FAM66B, FAM90A10, FAM90A13, FAM90A14, FAM90A18, FAM90A19, FAM90A20, FAM90A5, FAM90A7, FAM90A8, FAM90A9, LOC349196, SPAG11A, SPAG11B, ZNF705G*
chr10:38,856,419-39,080,121	CN Loss	223703	0	5	p11.1	
chr14:105,493,945-105,680,979	CN Loss	187035	1	13	q32.33	*ADAM6*
chr15:19,329,731-20,235,151	CN Loss	905421	6	28	q11.2	*LOC727924, NF1P1, OR4M2, OR4N3P, OR4N4, POTEB*
chr16:32,437,818-33,691,406	CN Loss	1253589	3	42	p11.2	*SLC6A10P, TP53TG3, TP53TG3B*
chr22:18,708,904-19,069,275	CN Gain	360372	2	12	q11.21	*PI4KAP1, RIMBP3*
chrX:152,001,572-154,591,854	High Copy Gain	2590283	75	217	q28	*ABCD1, ARHGAP4, ATP2B3, ATP6AP1, AVPR2, BCAP31, BGN, BRCC3, CLIC2, CTAG1A, CTAG1B, CTAG2, DKC1, DNASE1L1, DUSP9, EMD, F8, F8A1, F8A2, F8A3, FAM3A, FAM50A, FAM58A, FLNA, FUNDC2, G6PD, GAB3, GDI1, H2AFB1, H2AFB2, H2AFB3, HAUS7, HCFC1, IDH3G, IKBKG, IRAK1, L1CAM, LAGE3, LOC100132963, MAGEA1, MECP2, MIR718, MPP1, MTCP1, MTCP1NB, NAA10, NCRNA00204, NCRNA00204B, OPN1LW, OPN1MW, OPN1MW2, PDZD4, PLXNA3, PLXNB3, PNCK, RAB39B, RENBP, RPL10, SLC10A3, SLC6A8, SNORA36A, SNORA56, SNORA70, SRPK3, SSR4, TAZ, TEX28, TKTL1, TMEM187, TMLHE, TREX2, UBL4A, VBP1, ZFP92, ZNF275*
chrX:154,591,854-154,881,207	CN Gain	289354	3	43	q28	*IL9R, SPRY3, VAMP7*

Table S3. 2 . CNVs found for patient 2.

Chromosome Region	Event	Length	Genes	Probes	Cytoband	Gene Symbols
chr1:72,532,314-72,590,954	Homozygous Copy Loss	58641	0	3	p31.1	
chr2:89,327,057-89,895,566	CN Gain	568510	0	16	p11.2	
chr5:70,043,543-70,555,896	CN Loss	512354	12	10	q13.2	*GTF2H2, GTF2H2B, GTF2H2C, GTF2H2D, LOC647859, LOC653391, NAIP, OCLN, SERF1A, SERF1B, SMN1, SMN2*
chr6:229,040-327,656	CN Loss	98617	1	7	p25.3	*DUSP22*
chr8:7,214,279-7,942,536	Homozygous Copy Loss	728258	23	24	p23.1	*DEFB103A, DEFB103B, DEFB104A, DEFB104B, DEFB105A, DEFB105B, DEFB106A, DEFB106B, DEFB107A, DEFB107B, DEFB109P1B, DEFB4, FAM66E, FAM90A10, FAM90A13, FAM90A14, FAM90A18, FAM90A19, FAM90A7, FAM90A8, FAM90A9, SPAG11A, SPAG11B*
chr8:39,349,059-39,508,365	High Copy Gain	159307	2	10	p11.23 - p11.22	*ADAM3A, ADAM5P*
chr10:46,370,252-46,573,606	CN Gain	203355	4	16	q11.22	*GPRIN2, LOC728643, PPYR1, SYT15*
chr12:9,519,693-9,606,536	Homozygous Copy Loss	86844	0	6	p13.31	
chr14:18,436,931-19,490,518	CN Gain	1053588	9	36	q11.1 - q11.2	*OR11H12, OR4K1, OR4K2, OR4K5, OR4M1, OR4N2, OR4Q3, P704P, POTEG*
chr14:105,602,384-105,637,917	Homozygous Copy Loss	35534	0	3	q32.33	
chr16:32,019,584-33,441,199	CN Loss	1421616	4	47	p11.2	*HERC2P4, SLC6A10P, TP53TG3, TP53TG3B*
chr16:34,328,350-34,611,145	CN Gain	282796	2	17	p11.2 - p11.1	*LOC146481, LOC283914*
chr22:14,433,473-14,831,547	CN Gain	398075	2	4	q11.1	*OR11H1, POTEH*
chr22:17,036,669-17,240,532	CN Gain	203864	2	5	q11.21	*GGT3P, USP18*
chrX:152,783,739-153,255,487	High Copy Gain	471749	16	57	q28	*ARHGAP4, AVPR2, FLNA, HCFC1, IRAK1, L1CAM, MECP2, MIR718, NAA10, OPN1LW, OPN1MW, OPN1MW2, RENBP, TEX28, TKTL1, TMEM187*

Table S3. 3 – CNVs found for patient 3

Chromosome Region	Event	Length	Genes	Probes	Cytoband	Gene Symbols
chr1:16,725,089-16,987,680	CN Loss	262592	5	15	p36.13	*CROCCL1, ESPNP, MST1P2, MST1P9, NBPF1*
chr1:159,786,360-159,874,395	CN Loss	88036	4	6	q23.3	*FCGR2C, FCGR3A, FCGR3B, HSPA7*
chr1:246,800,021-246,863,559	Homozygous Copy Loss	63539	3	5	q44	*OR2T10, OR2T11, OR2T34*
chr2:89,064,440-89,895,566	CN Loss	831127	0	24	p11.2	
chr2:137,030,336-137,200,793	CN Gain	170458	0	11	q22.1	
chr4:3,863,079-4,243,929	CN Loss	380851	2	11	p16.2	*LOC348926, OTOP1*
chr4:34,448,772-34,500,505	Homozygous Copy Loss	51734	0	4	p15.1	
chr4:69,063,649-69,207,266	CN Gain	143618	2	11	q13.2	*UGT2B15, UGT2B17*
chr5:726,908-933,798	CN Loss	206891	3	20	p15.33	*BRD9, TPPP, ZDHHC11*
chr5:69,138,408-70,245,600	CN Loss	1107193	9	18	q13.2	*GTF2H2B, GTF2H2C, GTF2H2D, LOC100170939, LOC653391, SERF1A, SERF1B, SMN1, SMN2*
chr6:29,951,068-30,013,265	CN Loss	62198	3	4	p21.33	*HCG2P7, HCG4P6, HLA-H*
chr6:79,027,678-79,098,580	CN Gain	70903	0	4	q14.1	
chr8:7,026,686-12,529,929	CN Loss	5503244	70	275	p23.1	*AMAC1L2, BLK, C8orf12, C8orf74, CLDN23, CTSB, DEFB103A, DEFB103B, DEFB104A, DEFB104B, DEFB105A, DEFB105B, DEFB106A, DEFB106B, DEFB107A, DEFB107B, DEFB109P1, DEFB109P1B, DEFB130, DEFB134, DEFB135, DEFB136, DEFB4, ERI1, FAM167A, FAM66A, FAM66B, FAM66D, FAM66E, FAM86B1, FAM86B2, FAM90A10, FAM90A13, FAM90A14, FAM90A18, FAM90A19, FAM90A20, FAM90A5, FAM90A7, FAM90A8, FAM90A9, FDFT1, FLJ10661, GATA4, LOC157627, LOC349196, LOC392196, MFHAS1, MIR124-1, MIR1322, MIR548I3, MIR597, MIR598, MSRA, MTMR9, NEIL2, PINX1, PPP1R3B, PRAGMIN, RP1L1, SOX7, SPAG11A, SPAG11B, T-SP1, TDH, TNKS, USP17L2, XKR6, ZNF705D, ZNF705G*
chr8:39,349,059-39,508,365	Homozygous Copy Loss	159307	2	10	p11.23 - p11.22	*ADAM3A, ADAM5P*
chr11:70,990,916-71,304,747	CN Loss	313832	3	14	q13.4	*DEFB108B, FAM86C, LOC100133315*
chr14:18,436,931-19,490,518	CN Loss	1053588	9	36	q11.1 - q11.2	*OR11H12, OR4K1, OR4K2, OR4K5, OR4M1, OR4N2, OR4Q3, P704P, POTEG*
chr14:21,678,281-21,794,360	CN Loss	116080	0	7	q11.2	
chr14:105,602,384-105,637,917	Homozygous Copy Loss	35534	0	3	q32.33	
chr15:18,643,606-19,221,090	CN Gain	577485	4	19	q11.2	*BCL8, GOLGA6L6, GOLGA8C, LOC646214*
chr15:19,699,696-20,154,336	CN Gain	454641	4	19	q11.2	*LOC727924, OR4M2, OR4N3P, OR4N4*
chr15:32,511,473-32,641,620	CN Loss	130148	2	11	q14	*GOLGA8A, GOLGA8B*
chr17:31,449,032-31,890,587	CN Loss	441556	10	11	q12	*CCL3L1, CCL3L3, CCL4, CCL4L1, CCL4L2, TBC1D3B, TBC1D3C, TBC1D3F, TBC1D3G, TBC1D3H*
chr17:42,963,132-43,029,201	CN Loss	66070	1	5	q21.32	*NPEPPS*

chr19:10,693,339-10,885,518	CN Loss	192180	5	12	p13.2	*C19orf38, CARM1, DNM2, MIR199A1, TMED1*
chr22:14,433,473-14,831,547	CN Loss	398075	2	4	q11.1	*OR11H1, POTEH*

Table S3. 4 - CNVs found for patient 4

Chromosome Region	Event	Length	Genes	Probes	Cytoband	Gene Symbols
chr1:8,516,261-15,269,259	CN Loss	6752999	86	364	p36.23 - p36.21	*AADACL3, AADACL4, AGTRAP, ANGPTL7, APITD1, C1orf127, C1orf158, C1orf187, C1orf200, CA6, CASZ1, CLCN6, CLSTN1, CORT, CTNNBIP1, DFFA, DHRS3, ENO1, EXOSC10, FBXO2, FBXO44, FBXO6, GPR157, H6PD, HNRNPCL1, KIAA1026, KIAA2013, KIF1B, LOC440563, LOC649330, LZIC, MAD2L2, MASP2, MFN2, MIIP, MIR34A, MTHFR, MTOR, NMNAT1, NPPA, NPPB, PDPN, PEX14, PGD, PIK3CD, PLOD1, PRAMEF1, PRAMEF10, PRAMEF11, PRAMEF12, PRAMEF13, PRAMEF14, PRAMEF15, PRAMEF16, PRAMEF17, PRAMEF18, PRAMEF19, PRAMEF2, PRAMEF20, PRAMEF21, PRAMEF22, PRAMEF3, PRAMEF4, PRAMEF5, PRAMEF6, PRAMEF7, PRAMEF8, PRAMEF9, PRDM2, PTCHD2, RBP7, RERE, SLC25A33, SLC2A5, SLC2A7, SNORA59A, SNORA59B, SPSB1, SRM, TARDBP, TMEM201, TNFRSF1B, TNFRSF8, UBE4B, UBIAD1, VPS13D*
chr1:103,965,250-104,103,459	CN Loss	138210	4	4	p21.1	*AMY1A, AMY1B, AMY1C, AMY2A*
chr1:142,453,717-142,952,778	CN Loss	499062	4	13	q21.1	*FLJ39739, LOC100130000, LOC100286793, PPIAL4G*
chr1:194,979,783-195,097,982	CN Gain	118200	3	10	q31.3	*CFH, CFHR1, CFHR3*
chr3:163,987,281-164,108,596	High Copy Gain	121316	0	7	q26.1	
chr4:34,448,772-34,500,505	CN Gain	51734	0	4	p15.1	
chr4:69,038,815-69,165,944	CN Loss	127130	2	9	q13.2	*TMPRSS11E, UGT2B17*
chr5:742,057-900,209	CN Loss	158153	2	16	p15.33	*TPPP, ZDHHC11*
chr5:68,892,660-69,177,814	CN Loss	285155	7	12	q13.2	*GTF2H2, GTF2H2B, GTF2H2C, GTF2H2D, GUSBP3, LOC100272216, LOC653391*
chr5:70,312,515-70,375,604	Homozygous Copy Loss	63090	5	4	q13.2	*GTF2H2, GTF2H2B, GTF2H2C, GTF2H2D, NAIP*
chr5:70,375,604-70,671,221	CN Loss	295618	7	7	q13.2	*GTF2H2, GTF2H2B, GTF2H2C, GTF2H2D, LOC647859, NAIP, OCLN*
chr6:79,027,678-79,098,580	CN Gain	70903	0	4	q14.1	
chr8:39,349,059-39,508,365	Homozygous Copy Loss	159307	2	10	p11.23 - p11.22	*ADAM3A, ADAM5P*
chr9:39,115,053-39,251,004	CN Gain	135952	1	7	p13.1	*CNTNAP3*
chr9:43,413,941-43,827,369	CN Gain	413429	1	8	p11.2	*FAM75A6*
chr11:55,119,157-55,178,810	CN Gain	59654	3	3	q11	*OR4C11, OR4P4, OR4S2*
chr12:11,398,916-11,449,885	CN Gain	50970	2	4	p13.2	*PRB1, PRB2*
chr12:17,267,054-17,361,877	CN Gain	94824	0	6	p12.3	
chr14:18,555,363-19,244,829	CN Loss	689467	2	20	q11.1 - q11.2	*P704P, POTEG*
chr15:21,120,476-21,227,165	CN Loss	106690	0	5	q11.2	
chr17:31,449,032-31,762,515	CN Loss	313484	8	9	q12	*CCL3L1, CCL3L3, CCL4, CCL4L1, CCL4L2, TBC1D3B, TBC1D3C, TBC1D3F*
chr17:41,768,841-42,128,214	CN Gain	359374	5	7	q21.31 - q21.32	*ARL17A, ARL17B, LRRC37A, LRRC37A2, NSF*

chr22:22,650,373-22,676,684	CN Loss	26312	3	3	q11.23	*DDT, GSTT2, GSTTP1*
chr22:22,676,684-22,729,016	CN Gain	52333	4	4	q11.23	*GSTT1, GSTTP1, GSTTP2, LOC391322*

Table S3. 5 - CNVs found for patient 5

Chromosome Region	Event	Length	Genes	Probes	Cytoband	Gene Symbols
chr1:1,612,252-1,686,796	CN Loss	74545	6	5	p36.33	*CDK11A, CDK11B, LOC728661, MMP23A, NADK, SLC35E2*
chr1:103,961,733-104,117,636	CN Gain	155904	4	7	p21.1	*AMY1A, AMY1B, AMY1C, AMY2A*
chr1:194,993,686-195,078,812	CN Gain	85127	2	8	q31.3	*CFHR1, CFHR3*
chr1:241,618,630-241,805,298	CN Loss	186669	2	11	q43 - q44	*AKT3, SDCCAG8*
chr1:246,667,935-246,781,770	CN Gain	113836	4	8	q44	*OR2G6, OR2T2, OR2T3, OR2T5*
chr1:246,781,770-246,863,559	CN Loss	81790	4	6	q44	*OR2T10, OR2T11, OR2T29, OR2T34*
chr3:163,987,281-164,108,596	High Copy Gain	121316	0	7	q26.1	
chr3:174,728,013-174,778,763	CN Gain	50751	1	4	q26.31	*NLGN1*
chr4:69,038,815-69,170,341	CN Gain	131527	2	10	q13.2	*TMPRSS11E, UGT2B17*
chr5:68,991,456-70,312,515	CN Gain	1321060	11	25	q13.2	*GTF2H2B, GTF2H2C, GTF2H2D, GUSBP3, LOC100170939, LOC653391, NAIP, SERF1A, SERF1B, SMN1, SMN2*
chr6:244,003-327,656	CN Loss	83654	1	6	p25.3	*DUSP22*
chr6:32,548,396-32,600,341	Homozygous Copy Loss	51946	1	4	p21.32	*HLA-DRB5*
chr6:32,600,341-32,634,334	CN Loss	33994	2	4	p21.32	*HLA-DRB5, HLA-DRB6*
chr7:143,509,786-143,693,789	CN Gain	184004	8	8	q35	*ARHGEF5, ARHGEF5L, CTAGE4, OR2A1, OR2A20P, OR2A42, OR2A7, OR2A9P*
chr8:39,349,059-39,508,365	Homozygous Copy Loss	159307	2	10	p11.23 - p11.22	*ADAM3A, ADAM5P*
chr9:38,701,870-38,942,543	CN Loss	240674	0	14	p13.1	
chr12:9,519,693-9,606,536	Homozygous Copy Loss	86844	0	6	p13.31	
chr14:105,954,649-106,023,194	High Copy Gain	68546	1	5	q32.33	*LOC100133469*
chr16:0-19,827	CN Loss	19828	1	3	p13.3	*LOC100288778*
chr16:68,697,986-68,774,654	CN Loss	76669	2	5	q22.1	*CLEC18C, PDPR*
chr17:31,378,710-31,541,638	CN Loss	162929	4	11	q12	*CCL18, CCL3, CCL4, TBC1D3B*
chr19:20,391,455-20,502,094	CN Loss	110640	1	8	p12	*ZNF826*
chr22:18,708,904-19,014,218	CN Gain	305315	2	9	q11.21	*PI4KAP1, RIMBP3*

chrX:40,705,283-40,826,725	CN Gain	121443	0	6	p11.4
chrX:62,356,295-62,418,619	CN Gain	62325	0	4	q11.1

Table S3. 6 - CNVs found for patient 6

Chromosome Region	Event	Length	Genes	Probes	Cytoband	Gene Symbols
chr1:72,532,314-72,590,954	Homozygous Copy Loss	58641	0	3	p31.1	
chr1:103,961,733-104,103,459	CN Loss	141727	4	5	p21.1	*AMY1A, AMY1B, AMY1C, AMY2A*
chr1:194,993,686-195,078,812	CN Loss	85127	2	8	q31.3	*CFHR1, CFHR3*
chr2:70,748,414-74,540,348	CN Loss	3791935	49	241	p13.3 - p13.1	*ACTG2, ADD2, ALMS1, ALMS1P, ANKRD53, ATP6V1B1, BOLA3, C2orf7, C2orf78, C2orf81, CCT7, CD207, CLEC4F, CYP26B1, DCTN1, DGUOK, DUSP11, DYSF, EGR4, EMX1, EXOC6B, FBXO41, FIGLA, INO80B, LOC100189589, MCEE, MOBKL1B, MPHOSPH10, MTHFD2, NAGK, NAT8, NAT8B, NOTO, OR7E91P, PAIP2B, RAB11FIP5, RTKN, SFXN5, SLC4A5, SMYD5, SPR, STAMBP, TET3, TEX261, TPRKB, VAX2, WBP1, WDR54, ZNF638*
chr2:74,558,382-74,840,026	CN Loss	281645	13	24	p13.1	*AUP1, C2orf65, CCDC142, DOK1, DQX1, HTRA2, LBX2, LOC151534, LOXL3, PCGF1, SEMA4F, TLX2, TTC31*
chr3:120,201,772-120,294,810	High Copy Gain	93039	1	5	q13.32	*IGSF11*
chr3:163,987,281-164,108,596	Homozygous Copy Loss	121316	0	7	q26.1	
chr4:48,953,993-49,276,565	CN Gain	322573	0	3	p11	
chr4:69,038,815-69,170,341	CN Gain	131527	2	10	q13.2	*TMPRSS11E, UGT2B17*
chr5:68,892,660-70,696,119	CN Loss	1803460	15	41	q13.2	*GTF2H2, GTF2H2B, GTF2H2C, GTF2H2D, GUSBP3, LOC100170939, LOC100272216, LOC647859, LOC653391, NAIP, OCLN, SERF1A, SERF1B, SMN1, SMN2*
chr6:193,938-327,656	CN Loss	133719	1	9	p25.3	*DUSP22*
chr6:29,951,068-30,013,265	CN Loss	62198	3	4	p21.33	*HCG2P7, HCG4P6, HLA-H*
chr6:32,563,029-32,600,341	Homozygous Copy Loss	37313	1	3	p21.32	*HLA-DRB5*
chr6:32,600,341-32,703,052	CN Loss	102712	3	10	p21.32	*HLA-DRB1, HLA-DRB5, HLA-DRB6*
chr6:79,027,678-79,131,904	CN Loss	104227	0	6	q14.1	
chr7:143,548,913-143,667,377	CN Loss	118465	6	6	q35	*CTAGE4, OR2A1, OR2A20P, OR2A42, OR2A7, OR2A9P*
chr8:7,705,464-8,124,705	CN Loss	419242	17	13	p23.1	*DEFB103A, DEFB103B, DEFB104A, DEFB104B, DEFB105A, DEFB105B, DEFB106A, DEFB106B, DEFB107A, DEFB107B, DEFB109P1B, DEFB4, FAM66E, FLJ10661, MIR548I3, SPAG11A, SPAG11B*
chr8:39,349,059-39,508,365	Homozygous Copy Loss	159307	2	10	p11.23 - p11.22	*ADAM3A, ADAM5P*
chr8:145,486,966-145,761,617	CN Gain	274652	24	28	q24.3	*ADCK5, C8ORFK29, CPSF1, CYHR1, DGAT1, FBXL6, FOXH1, GPR172A, GPT, HSF1, KIAA1688, KIFC2, LRRC14, LRRC24, MFSD3, MGC70857, MIR1234, MIR939, NFKBIL2, PPP1R16A, RECQL4, SCRT1, SLC39A4, VPS28*
chr9:38,769,965-38,942,543	CN Gain	172579	0	10	p13.1	
chr9:41,791,678-43,613,184	CN Loss	1821507	10	55	p12 - p11.2	*ANKRD20A2, ANKRD20A3, AQP7P3, FAM75A6, FAM95B1, FOXD4L2, FOXD4L4, KGFLP2, LOC642929, MGC21881*
chr9:65,222,385-65,792,243	CN Gain	569859	4	6	q12	*FAM74A2, FAM74A4, FAM75A5, FAM75A7*

chr10:39,008,638-39,080,121	CN Loss	71484	0	2	p11.1	
chr14:19,304,869-19,507,271	CN Gain	202403	5	12	q11.2	*OR4K1, OR4K2, OR4K5, OR4M1, OR4N2*
chr14:105,136,197-105,194,518	High Copy Gain	58322	0	5	q32.33	
chr14:105,194,518-105,330,429	CN Gain	135912	0	10	q32.33	
chr15:18,643,606-20,154,336	CN Loss	1510731	11	52	q11.2	*BCL8, CXADRP2, GOLGA6L6, GOLGA8C, LOC646214, LOC727924, NF1P1, OR4M2, OR4N3P, OR4N4, POTEB*
chr15:21,989,766-22,270,122	CN Gain	280357	0	16	q11.2	
chr16:32,437,818-33,691,406	CN Loss	1253589	3	42	p11.2	*SLC6A10P, TP53TG3, TP53TG3B*
chr17:40,937,706-41,021,360	CN Loss	83655	1	3	q21.31	*LRRC37A4*
chr19:20,374,751-20,421,451	CN Loss	46701	1	3	p12	*ZNF826*
chr19:20,421,451-20,502,094	Homozygous Copy Loss	80644	0	6	p12	
chr20:60,310,050-60,343,063	CN Gain	33014	2	4	q13.33	*ADRM1, LAMA5*
chr20:62,390,214-62,435,964	CN Loss	45751	0	4	q13.33	

Table S3. 7 - CNVs found for patient 7

Chromosome Region	Event	Length	Genes	Probes	Cytoband	Gene Symbols
chr1:147,255,486-148,058,610	CN Loss	803125	8	30	q21.1 - q21.2	*FCGR1A, FCGR1C, HIST2H2BF, HIST2H3D, LOC388692, LOC728855, PPIAL4A, PPIAL4C*
chr1:194,979,812-195,098,012	CN Loss	118201	3	10	q31.3	*CFH, CFHR1, CFHR3*
chr1:200,721,902-200,794,272	CN Loss	72371	1	5	q32.1	*PPP1R12B*
chr1:241,114,655-241,213,134	CN Loss	98480	0	5	q43	
chr2:87,872,712-88,037,639	CN Gain	164928	2	6	p11.2	*RGPD1, RGPD2*
chr2:89,017,454-89,176,704	CN Loss	159251	0	6	p11.2	
chr2:101,122,697-105,631,450	CN Loss	4508754	24	197	q11.2 - q12.2	*C2orf29, C2orf49, CREG2, FHL2, GPR45, IL18R1, IL18RAP, IL1R1, IL1R2, IL1RL1, IL1RL2, LOC150568, MAP4K4, MFSD9, MRPS9, POU3F3, RFX8, RNF149, SLC9A2, SLC9A4, SNORD89, TBC1D8, TGFBRAP1, TMEM182*
chr2:194,430,401-194,546,889	CN Loss	116489	0	6	q32.3	
chr3:65,489,585-65,574,328	CN Gain	84744	1	5	p14.1	*MAGI1*
chr3:75,849,278-75,886,078	CN Gain	36801	1	3	p12.3	*ZNF717*
chr3:152,985,044-153,033,954	CN Loss	48911	2	3	q25.1	*AADAC, LOC201651*
chr3:196,848,458-196,944,322	CN Loss	95865	3	6	q29	*MIR570, MUC20, SDHAP2*
chr4:34,448,802-34,500,534	CN Gain	51733	0	4	p15.1	
chr4:48,818,175-49,276,624	CN Loss	458450	0	9	p11	
chr4:68,978,217-69,170,363	CN Loss	192147	2	12	q13.2	*TMPRSS11E, UGT2B17*
chr5:68,866,425-68,913,120	CN Loss	46696	7	3	q13.2	*GTF2H2, GTF2H2B, GTF2H2C, GTF2H2D, LOC647859, LOC653391, OCLN*
chr6:244,029-327,686	CN Loss	83658	1	6	p25.3	*DUSP22*
chr6:32,548,425-32,600,368	CN Loss	51944	1	4	p21.32	*HLA-DRB5*
chr6:79,027,708-79,098,610	CN Loss	70903	0	4	q14.1	
chr7:143,503,240-143,693,818	CN Gain	190579	8	9	q35	*ARHGEF5, ARHGEF5L, CTAGE4, OR2A1, OR2A20P, OR2A42, OR2A7, OR2A9P*
chr8:7,303,319-8,141,220	CN Loss	837902	25	24	p23.1	*DEFB103A, DEFB103B, DEFB104A, DEFB104B, DEFB105A, DEFB105B, DEFB106A, DEFB106B, DEFB107A, DEFB107B, DEFB109P1B, DEFB4, FAM66E, FAM90A10, FAM90A13, FAM90A14, FAM90A18, FAM90A19, FAM90A7, FAM90A8, FAM90A9, FLJ10661, MIR548I3, SPAG11A, SPAG11B*
chr8:15,986,485-16,062,660	CN Loss	76176	1	5	p22	*MSR1*
chr8:39,349,089-39,508,394	High Copy Gain	159306	2	10	p11.23 - p11.22	*ADAM3A, ADAM5P*
chr10:46,132,024-47,327,989	CN Loss	1195966	16	42	q11.22	*ANTXRL, ANXA8, ANXA8L1, ANXA8L2, BMS1P1, BMS1P5, FAM25B, FAM25C, FAM25G, FAM35B, FAM35B2, GPRIN2, LOC642826, LOC728643, PPYR1, SYT15*

chr10:47,521,106-47,911,496	CN Loss	390391	6	6	q11.22	*ANXA8, ANXA8L1, FAM25B, FAM25C, FAM25G, LOC642826*
chr10:81,053,624-81,631,465	CN Loss	577842	4	10	q22.3	*LOC642361, LOC650623, SFTPA1, SFTPA2*
chr10:88,911,213-89,220,524	CN Loss	309312	5	10	q23.2	*FAM22A, FAM22D, FAM35A, LOC439994, LOC728190*
chr11:337,945-470,027	CN Gain	132083	5	11	p15.5	*ANO9, B4GALNT4, PKP3, PTDSS2, SIGIRR*
chr11:134,395,794-134,452,384	CN Gain	56591	0	6	q25	
chr12:0-54,932	CN Loss	54933	2	4	p13.33	*FAM138D, IQSEC3*
chr12:31,170,817-31,239,425	CN Loss	68609	0	5	p11.21	
chr14:18,436,960-19,244,853	CN Gain	807894	3	21	q11.1 - q11.2	*OR11H12, P704P, POTEG*
chr14:106,219,129-106,256,008	CN Gain	36880	0	3	q32.33	
chr15:18,561,649-20,154,365	CN Loss	1592717	11	52	q11.2	*BCL8, CXADRP2, GOLGA6L6, GOLGA8C, LOC646214, LOC727924, NF1P1, OR4M2, OR4N3P, OR4N4, POTEB*
chr15:32,469,503-32,641,649	CN Loss	172147	2	13	q14	*GOLGA8A, GOLGA8B*
chr15:82,611,372-82,737,224	CN Loss	125853	1	6	q25.2	*LOC388152*
chr16:22,377,731-22,631,044	CN Loss	253314	5	11	p12.1	*LOC100132247, LOC100190986, LOC641298, LOC653786, NPIPL3*
chr17:41,521,693-42,128,243	CN Loss	606551	6	21	q21.31 - q21.32	*ARL17A, ARL17B, KIAA1267, LRRC37A, LRRC37A2, NSF*
chr19:20,374,780-20,502,123	CN Loss	127344	1	7	p12	*ZNF826*
chr22:14,433,473-14,831,576	CN Gain	398104	2	5	q11.1	*OR11H1, POTEH*
chr22:17,036,698-17,294,519	CN Loss	257822	4	9	q11.21	*DGCR6, GGT3P, PRODH, USP18*

Table S3. 8 - CNVs found for patient 8

Chromosome Region	Event	Length	Genes	Probes	Cytoband	Gene Symbols
chr1:102,431,081-102,633,897	CN Loss	202817	0	11	p21.1	
chr1:142,117,366-142,453,717	CN Loss	336352	0	4	q12 - q21.1	
chr1:194,993,686-195,097,982	CN Gain	104297	2	9	q31.3	*CFHR1, CFHR3*
chr2:91,273,082-91,625,698	CN Loss	352617	2	11	p11.1	*FKSG73, GGT8P*
chr2:96,098,910-97,594,697	CN Gain	1495788	24	88	q11.2	*ADRA2B, ANKRD23, ANKRD36, ANKRD36B, ANKRD39, ARID5A, ASTL, CIAO1, CNNM3, CNNM4, DUSP2, FAHD2B, FAM178B, FER1L5, ITPRIPL1, KIAA1310, LMAN2L, LOC285033, NCAPH, NEURL3, SEMA4C, SNRNP200, STARD7, TMEM127*
chr2:130,418,126-130,467,710	CN Loss	49585	1	4	q21.1	*RAB6C*
chr2:131,811,198-131,860,657	CN Loss	49460	1	5	q21.1	*LOC150786*
chr3:75,752,964-75,905,213	CN Loss	152250	3	12	p12.3	*FRG2C, MIR1324, ZNF717*
chr3:163,987,281-164,108,596	High Copy Gain	121316	0	7	q26.1	
chr4:34,448,772-34,500,505	CN Gain	51734	0	4	p15.1	
chr4:68,978,187-69,170,341	CN Gain	192155	2	12	q13.2	*TMPRSS11E, UGT2B17*
chr4:70,227,242-70,269,497	CN Loss	42256	0	3	q13.2	
chr5:742,057-839,349	CN Gain	97293	1	9	p15.33	*TPPP*
chr5:70,245,600-70,375,604	CN Gain	130005	9	5	q13.2	*GTF2H2, GTF2H2B, GTF2H2C, GTF2H2D, NAIP, SERF1A, SERF1B, SMN1, SMN2*
chr6:193,938-327,656	CN Loss	133719	1	9	p25.3	*DUSP22*
chr6:32,548,396-32,600,341	Homozygous Copy Loss	51946	1	4	p21.32	*HLA-DRB5*
chr6:32,600,341-32,623,614	CN Gain	23274	1	3	p21.32	*HLA-DRB5*
chr6:32,623,614-32,681,379	CN Loss	57766	2	5	p21.32	*HLA-DRB1, HLA-DRB6*
chr6:79,027,678-79,098,580	CN Gain	70903	0	4	q14.1	
chr6:168,076,174-168,333,368	CN Gain	257195	4	18	q27	*FRMD1, HGC6.3, KIF25, MLLT4*
chr7:38,284,089-38,349,754	CN Loss	65666	0	5	p14.1	
chr7:124,186,384-124,227,026	High Copy Gain	40643	1	3	q31.33	*GPR37*
chr7:142,066,473-142,209,840	CN Loss	143368	3	11	q34	*PRSS1, PRSS2, TRY6*
chr7:142,947,514-143,049,236	CN Gain	101723	2	4	q35	*FAM115C, LOC441294*
chr8:12,270,256-12,366,919	Homozygous Copy Loss	96664	3	3	p23.1	*DEFB109P1, FAM66A, FAM86B2*

chr8:12,366,919-12,554,253	CN Loss	187335	0	4	p23.1	
chr8:39,349,059-39,508,365	High Copy Gain	159307	2	10	p11.23 - p11.22	*ADAM3A, ADAM5P*
chr9:39,999,379-40,265,103	CN Loss	265725	0	5	p13.1 - p12	
chr9:65,525,874-65,792,243	CN Loss	266370	0	4	q12	
chr10:38,948,353-39,080,121	CN Loss	131769	0	4	p11.1	
chr10:45,969,999-46,573,606	CN Loss	603608	10	31	q11.21 - q11.22	*BMS1P1, BMS1P5, FAM35B, FRMPD2L1, GPRIN2, LOC728643, PPYR1, PTPN20A, PTPN20B, SYT15*
chr10:47,169,331-47,451,593	CN Loss	282263	3	4	q11.22	*ANTXRL, ANXA8L2, FAM21B*
chr10:48,381,494-49,040,567	CN Loss	659074	7	9	q11.22	*FAM25B, FAM25C, FAM25G, FRMPD2, LOC399753, PTPN20A, PTPN20B*
chr11:18,851,036-18,925,523	CN Loss	74488	1	5	p15.1	*MRGPRX1*
chr11:55,119,157-55,178,810	CN Gain	59654	3	3	q11	*OR4C11, OR4P4, OR4S2*
chr12:8,208,895-8,273,494	CN Loss	64600	3	4	p13.31	*FAM66C, FAM90A1, ZNF705A*
chr12:31,170,795-31,292,064	CN Gain	121270	0	8	p11.21	
chr14:18,436,931-19,490,518	CN Loss	1053588	9	36	q11.1 - q11.2	*OR11H12, OR4K1, OR4K2, OR4K5, OR4M1, OR4N2, OR4Q3, P704P, POTEG*
chr14:40,669,880-40,734,612	CN Loss	64733	0	3	q21.1	
chr14:105,602,384-105,693,838	CN Loss	91455	0	9	q32.33	
chr15:18,561,627-20,154,336	CN Loss	1592710	11	55	q11.2	*BCL8, CXADRP2, GOLGA6L6, GOLGA8C, LOC646214, LOC727924, NF1P1, OR4M2, OR4N3P, OR4N4, POTEB*
chr16:22,377,707-22,631,014	CN Loss	253308	5	11	p12.1	*LOC100132247, LOC100190986, LOC641298, LOC653786, NPIPL3*
chr16:32,096,628-32,582,495	CN Loss	485868	0	17	p11.2	
chr16:33,230,071-33,720,310	CN Loss	490240	0	26	p11.2	
chr16:34,345,125-34,600,842	CN Loss	255718	2	15	p11.2 - p11.1	*LOC146481, LOC283914*
chr16:58,848,869-59,019,408	CN Loss	170540	0	9	q21	
chr16:68,697,986-68,774,654	CN Gain	76669	2	5	q22.1	*CLEC18C, PDPR*
chr17:31,459,116-31,511,535	CN Loss	52420	0	5	q12	
chr17:41,577,131-41,768,841	CN Gain	191711	3	11	q21.31	*ARL17B, KIAA1267, LRRC37A*
chr19:17,330,586-17,377,389	CN Gain	46804	2	4	p13.11	*BST2, PLVAP*
chr19:20,067,301-20,148,413	CN Loss	81113	2	5	p12	*ZNF486, ZNF90*
chr19:20,374,751-20,502,094	CN Loss	127344	1	9	p12	*ZNF826*
chr22:14,433,473-14,629,187	CN Loss	195715	0	2	q11.1	

chr22:17,036,669-17,240,532	CN Loss	203864	2	5	q11.21	*GGT3P, USP18*

Table S3. 9 - CNVs found for patient 9

Chromosome Region	Event	Length	Genes	Probes	Cytoband	Gene Symbols
chr1:16,746,886-16,923,881	CN Loss	176996	4	10	p36.13	*CROCCL1, ESPNP, MST1P2, NBPF1*
chr1:72,532,314-72,590,954	Homozygous Copy Loss	58641	0	3	p31.1	
chr1:194,993,686-195,059,281	CN Loss	65596	2	6	q31.3	*CFHR1, CFHR3*
chr2:89,244,042-89,895,566	CN Loss	651525	0	19	p11.2	
chr3:132,898,329-132,993,320	Homozygous Copy Loss	94992	1	6	q22.1	*CPNE4*
chr3:132,993,320-136,601,728	CN Loss	3608409	23	223	q22.1 - q22.2	*ACAD11, ACPP, AMOTL2, ANAPC13, BFSP2, C3orf36, CCRL1, CDV3, CEP63, CPNE4, DNAJC13, EPHB1, KY, NCRNA00119, NPHP3, RAB6B, RYK, SLCO2A1, SRPRB, TF, TMEM108, TOPBP1, UBA5*
chr3:136,601,728-136,808,917	Homozygous Copy Loss	207190	0	11	q22.2	
chr3:136,808,917-143,101,242	CN Loss	6292326	41	382	q22.2 - q23	*A4GNT, ACPL2, ARMC8, ATP1B3, BPESC1, C3orf72, CEP70, CLDN18, CLSTN2, COPB2, DBR1, DZIP1L, ESYT3, FAIM,* ***FOXL2****, GRK7, IL20RB, MRAS, MRPS22, MSL2, NCK1, NMNAT3, PCCB, PIK3CB, PISRT1, PPP2R3A, PRR23A, PRR23B, PRR23C, RASA2, RBP1, RBP2, RNF7, SLC25A36, SOX14, SPSB4, STAG1, TMEM22, TRIM42, TXNDC6, ZBTB38*
chr3:163,987,281-164,108,596	High Copy Gain	121316	0	7	q26.1	
chr3:196,885,475-196,942,769	CN Loss	57295	3	4	q29	*MIR570, MUC20, SDHAP2*
chr4:68,978,187-69,170,341	CN Loss	192155	2	12	q13.2	*TMPRSS11E, UGT2B17*
chr6:212,083-327,656	CN Loss	115574	1	8	p25.3	*DUSP22*
chr6:168,076,174-168,333,368	CN Gain	257195	4	18	q27	*FRMD1, HGC6.3, KIF25, MLLT4*
chr7:100,750,082-100,924,707	CN Gain	174626	2	11	q22.1	*EMID2, RABL5*
chr8:7,121,948-8,104,612	CN Loss	982665	29	30	p23.1	*DEFB103A, DEFB103B, DEFB104A, DEFB104B, DEFB105A, DEFB105B, DEFB106A, DEFB106B, DEFB107A, DEFB107B, DEFB109P1B, DEFB4, FAM66B, FAM66E, FAM90A10, FAM90A13, FAM90A14, FAM90A18, FAM90A19, FAM90A20, FAM90A5, FAM90A7, FAM90A8, FAM90A9, LOC349196, MIR548I3, SPAG11A, SPAG11B, ZNF705G*
chr9:38,769,965-38,942,543	CN Gain	172579	0	10	p13.1	
chr17:41,521,663-41,768,841	CN Gain	247179	3	14	q21.31	*ARL17B, KIAA1267, LRRC37A*
chrX:154,690,941-154,846,576	CN Loss	155636	1	25	q28	*VAMP7*

Table S3. 10 - CNVs found for patient 10

Chromosome Region	Event	Length	Genes	Probes	Cytoband	Gene Symbols
chr1:16,778,535-17,218,740	CN Gain	440206	9	25	p36.13	*ATP13A2, CROCC, CROCCL1, ESPNP, MFAP2, MST1P2, MST1P9, NBPF1, SDHB*
chr1:72,532,314-72,609,367	Homozygous Copy Loss	77054	0	4	p31.1	
chr1:153,239,357-153,361,162	CN Gain	121806	6	9	q21.3 - q22	*ADAM15, DCST1, DCST2, EFNA3, EFNA4, ZBTB7B*
chr4:70,170,112-70,285,238	CN Gain	115127	1	7	q13.2	*UGT2B28*
chr5:69,585,286-70,043,543	CN Gain	458258	4	5	q13.2	*GTF2H2B, GTF2H2C, GTF2H2D, LOC653391*
chr5:70,043,543-70,375,604	CN Loss	332062	10	6	q13.2	*GTF2H2, GTF2H2B, GTF2H2C, GTF2H2D, LOC653391, NAIP, SERF1A, SERF1B, SMN1, SMN2*
chr5:176,544,003-176,734,904	CN Gain	190902	6	14	q35.3	*LMAN2, MXD3, NSD1, PRELID1, RAB24, RGS14*
chr6:32,609,419-32,659,791	CN Loss	50373	2	4	p21.32	*HLA-DRB1, HLA-DRB6*
chr8:143,526,840-143,626,016	CN Loss	99177	1	8	q24.3	*BAI1*
chr9:65,222,385-65,955,080	CN Loss	732696	4	7	q12	*FAM74A2, FAM74A4, FAM75A5, FAM75A7*
chr10:45,969,999-46,477,067	CN Loss	507069	8	24	q11.21 - q11.22	*BMS1P1, BMS1P5, FAM35B, FRMPD2L1, GPRIN2, PTPN20A, PTPN20B, SYT15*
chr10:46,477,067-46,573,606	Homozygous Copy Loss	96540	2	7	q11.22	*LOC728643, PPYR1*
chr10:48,381,494-49,040,567	CN Loss	659074	7	9	q11.22	*FAM25B, FAM25C, FAM25G, FRMPD2, LOC399753, PTPN20A, PTPN20B*
chr12:19,358,692-19,474,911	CN Gain	116220	1	7	p12.3	*PLEKHA5*
chr14:18,149,473-19,244,829	CN Gain	1095357	3	27	q11.1 - q11.2	*OR11H12, P704P, POTEG*
chr15:18,561,627-20,154,336	CN Loss	1592710	11	55	q11.2	*BCL8, CXADRP2, GOLGA6L6, GOLGA8C, LOC646214, LOC727924, NF1P1, OR4M2, OR4N3P, OR4N4, POTEB*
chr16:83,706,547-84,115,992	CN Gain	409446	0	22	q24.1	
chr17:19,439,761-19,486,821	CN Loss	47061	0	3	p11.2	
chr17:41,577,131-42,142,705	CN Gain	565575	6	19	q21.31 - q21.32	*ARL17A, ARL17B, KIAA1267, LRRC37A, LRRC37A2, NSF*
chr19:9,468,575-9,748,852	CN Gain	280278	6	17	p13.2	*ZNF121, ZNF426, ZNF560, ZNF561, ZNF562, ZNF846*
chr19:59,929,792-60,074,245	CN Gain	144454	8	6	q13.42	*KIR2DL1, KIR2DL3, KIR2DL4, KIR2DS4, KIR3DL1, KIR3DL2, KIR3DL3, KIR3DP1*
chr22:19,386,371-19,792,906	CN Gain	406536	14	33	q11.21	*AIFM3, CRKL, FLJ39582, LOC400891, LZTR1, MGC16703, P2RX6, P2RX6P, PI4KA, SERPIND1, SLC7A4, SNAP29, THAP7, TMEM191A*
chr22:23,983,829-24,247,592	CN Loss	263764	2	19	q11.23	*IGLL3, LRP5L*

Table S3. 11 - CNVs found for patient 11

Chromosome Region	Event	Length	Genes	Probes	Cytoband	Gene Symbols
chr1:147,255,457-147,553,575	CN Gain	298119	1	16	q21.1	*LOC388692*
chr1:246,682,708-246,781,770	CN Gain	99063	4	7	q44	*OR2G6, OR2T2, OR2T3, OR2T5*
chr1:246,781,770-246,882,848	CN Loss	101079	6	7	q44	*OR2T10, OR2T11, OR2T27, OR2T29, OR2T34, OR2T35*
chr2:132,594,362-132,762,389	CN Loss	168028	2	8	q21.2	*MIR663B, NCRNA00164*
chr2:242,507,916-242,683,523	CN Loss	175608	1	17	q37.3	*LOC728323*
chr3:101,814,939-101,906,244	CN Gain	91306	1	6	q12.2	*GPR128*
chr3:163,987,281-164,108,596	High Copy Gain	121316	0	7	q26.1	
chr4:34,429,336-34,500,505	CN Gain	71170	0	5	p15.1	
chr4:69,038,815-69,165,944	CN Gain	127130	2	9	q13.2	*TMPRSS11E, UGT2B17*
chr5:745,872-874,147	CN Loss	128276	2	12	p15.33	*TPPP, ZDHHC11*
chr5:32,135,149-32,197,234	CN Gain	62086	2	4	p13.3	*GOLPH3, PDZD2*
chr5:179,345,860-180,237,221	CN Gain	891362	12	57	q35.3	*CNOT6, FLT4, GFPT2, LOC729678, MAPK9, MGAT1, MIR340, OR2Y1, RASGEF1C, RNF130, SCGB3A1, ZFP62*
chr6:212,083-343,443	CN Loss	131361	2	9	p25.3	*DUSP22, IRF4*
chr6:79,027,678-79,116,521	CN Loss	88844	0	5	q14.1	
chr8:7,245,105-7,830,098	CN Loss	584994	21	21	p23.1	*DEFB103A, DEFB103B, DEFB104A, DEFB104B, DEFB105A, DEFB105B, DEFB106A, DEFB106B, DEFB107A, DEFB107B, DEFB4, FAM90A10, FAM90A13, FAM90A14, FAM90A18, FAM90A19, FAM90A7, FAM90A8, FAM90A9, SPAG11A, SPAG11B*
chr8:11,976,320-12,548,010	CN Loss	571691	9	13	p23.1	*DEFB109P1, DEFB130, FAM66A, FAM66D, FAM86B1, FAM86B2, LOC392196, USP17L2, ZNF705D*
chr8:39,332,477-39,508,365	CN Loss	175889	2	11	p11.23 - p11.22	*ADAM3A, ADAM5P*
chr9:38,848,109-42,219,258	CN Loss	3371150	13	93	p13.1 - p12	*CNTNAP3, FAM74A1, FAM74A3, FAM75A1, FAM75A2, FAM75A3, FAM75A5, FAM75A7, KGFLP2, LOC653501, MGC21881, ZNF658, ZNF658B*
chr9:43,131,618-47,055,569	CN Loss	3923952	5	68	p11.2 - p11.1	*FAM27A, FAM27C, FAM75A6, KGFLP1, LOC642929*
chr9:65,222,385-65,955,080	CN Loss	732696	4	7	q12	*FAM74A2, FAM74A4, FAM75A5, FAM75A7*
chr11:76,562,765-76,611,534	CN Loss	48770	2	3	q13.5	*GDPD4, MYO7A*
chr12:46,240,128-46,335,778	High Copy Gain	95651	0	5	q13.11	
chr14:105,602,384-105,637,917	Homozygous Copy Loss	35534	0	3	q32.33	
chr14:105,818,109-106,012,840	CN Loss	194732	1	15	q32.33	*LOC100133469*
chr15:19,796,307-20,044,857	CN Loss	248551	4	15	q11.2	*LOC727924, OR4M2, OR4N3P, OR4N4*
chr15:82,558,171-82,881,345	CN Loss	323175	3	14	q25.2	*GOLGA6L5, LOC388152, UBE2Q2P1*
chr16:68,697,986-68,752,812	High Copy Gain	54827	1	4	q22.1	*PDPR*
chr16:72,921,969-73,017,673	High Copy Gain	95705	2	6	q22.3	*CLEC18B, LOC283922*
chr17:41,555,382-41,713,328	CN Loss	157947	1	11	q21.31	*KIAA1267*
chr17:42,994,344-43,029,201	CN Loss	34858	1	3	q21.32	*NPEPPS*
chr21:45,629,020-45,951,495	CN Gain	322476	3	21	q22.3	*COL18A1, NCRNA00175, SLC19A1*

Table S3. 12 - CNVs found for patient 12

Chromosome Region	Event	Length	Genes	Probes	Cytoband	Gene Symbols
chr1:16,725,089-17,166,694	CN Loss	441606	6	24	p36.13	*CROCC, CROCCL1, ESPNP, MST1P2, MST1P9, NBPF1*
chr1:141,494,201-141,663,976	CN Gain	169776	0	6	q12	
chr1:159,748,177-159,827,104	CN Gain	78928	4	7	q23.3	*FCGR2A, FCGR2C, FCGR3A, HSPA6*
chr1:194,975,299-195,078,812	CN Loss	103514	3	10	q31.3	*CFH, CFHR1, CFHR3*
chr2:242,507,916-242,951,149	CN Loss	443234	1	20	q37.3	*LOC728323*
chr3:163,987,281-164,108,596	High Copy Gain	121316	0	7	q26.1	
chr3:196,933,249-196,956,726	CN Loss	23478	1	3	q29	*MUC20*
chr4:48,953,993-49,276,565	CN Loss	322573	0	3	p11	
chr4:69,038,815-69,170,341	CN Loss	131527	2	10	q13.2	*TMPRSS11E, UGT2B17*
chr4:70,141,356-70,269,497	CN Gain	128142	1	8	q13.2	*UGT2B28*
chr6:32,563,029-32,600,341	Homozygous Copy Loss	37313	1	3	p21.32	*HLA-DRB5*
chr6:32,600,341-32,647,235	CN Gain	46895	2	5	p21.32	*HLA-DRB5, HLA-DRB6*
chr6:32,647,235-32,681,379	CN Loss	34145	1	3	p21.32	*HLA-DRB1*
chr6:108,537,896-108,829,534	CN Gain	291639	3	18	q21	*LACE1, NR2E1, SNX3*
chr8:39,349,059-39,508,365	High Copy Gain	159307	2	10	p11.23 - p11.22	*ADAM3A, ADAM5P*
chr9:6,792,781-6,902,081	CN Loss	109301	1	6	p24.1	*KDM4C*
chr9:17,639,085-17,725,970	CN Gain	86886	1	5	p22.2	*SH3GL2*
chr9:39,081,057-39,251,004	CN Gain	169948	1	8	p13.1	*CNTNAP3*
chr11:55,119,157-55,178,810	CN Loss	59654	3	3	q11	*OR4C11, OR4P4, OR4S2*
chr13:62,664,941-62,752,472	CN Loss	87532	0	5	q21.31	
chr14:18,436,931-19,507,271	CN Loss	1070341	9	38	q11.1 - q11.2	*OR11H12, OR4K1, OR4K2, OR4K5, OR4M1, OR4N2, OR4Q3, P704P, POTEG*
chr14:105,602,384-105,637,917	Homozygous Copy Loss	35534	0	3	q32.33	
chr14:105,637,917-105,889,976	CN Loss	252060	0	23	q32.33	
chr15:18,856,397-19,441,979	CN Gain	585583	7	23	q11.2	*BCL8, CXADRP2, GOLGA6L6, GOLGA8C, LOC646214, NF1P1, POTEB*
chr16:488,401-617,378	CN Loss	128978	7	10	p13.3	*C16orf10, C16orf11, NHLRC4, PIGQ, RAB11FIP3, RAB40C, SOLH*
chr16:54,345,532-54,383,408	CN Loss	37877	1	3	q12.2	*CES4*
chr16:54,383,408-54,441,898	CN Gain	58491	2	4	q12.2	*CES1, CES7*
chr16:88,395,085-88,444,115	CN Gain	49031	2	4	q24.3	*FANCA, SPIRE2*
chr17:31,449,032-31,890,587	CN Loss	441556	10	11	q12	*CCL3L1, CCL3L3, CCL4, CCL4L1, CCL4L2, TBC1D3B, TBC1D3C, TBC1D3F, TBC1D3G, TBC1D3H*
chr17:40,937,706-41,036,432	CN Gain	98727	2	4	q21.31	*LOC644172, LRRC37A4*
chr17:41,521,663-41,648,920	CN Gain	127258	1	9	q21.31	*KIAA1267*
chr19:20,357,873-20,502,094	CN Loss	144222	1	10	p12	*ZNF826*
chr21:9,896,630-10,034,406	CN Loss	137777	1	5	p11.2 - p11.1	*TPTE*

chr22:22,659,866-22,792,316	CN Gain	132451	5	9	q11.23	*CABIN1, GSTT1, GSTTP1, GSTTP2, LOC391322*
chrX:1,333,567-1,390,376	CN Gain	56810	1	8	p22.33	*CSF2RA*
chrX:83,129,544-83,184,995	CN Gain	55452	0	3	q21.1	
chrX:137,447,004-137,563,727	CN Gain	116724	2	7	q26.3	*FGF13, LOC158696*

Table S3. 13 - CNVs found for patient 13

Chromosome Region	Event	Length	Genes	Probes	Cytoband	Gene Symbols
chr1:72,532,314-72,590,954	Homozygous Copy Loss	58641	0	3	p31.1	
chr1:194,979,783-195,064,797	CN Loss	85015	3	8	q31.3	*CFH, CFHR1, CFHR3*
chr2:88,975,437-89,895,566	CN Loss	920130	0	29	p11.2	
chr5:103,341,259-103,428,149	CN Loss	86891	0	5	q21.2	
chr6:32,563,029-32,634,334	Homozygous Copy Loss	71306	2	7	p21.32	*HLA-DRB5, HLA-DRB6*
chr6:32,634,334-32,666,536	CN Loss	32203	2	3	p21.32	*HLA-DRB1, HLA-DRB6*
chr6:156,054,446-158,724,482	CN Loss	2670037	7	140	q25.3	*ARID1B, GTF2H5, SERAC1, SNX9, SYNJ2, TULP4, ZDHHC14*
chr7:76,269,619-76,451,792	CN Gain	182174	1	9	q11.23	*PMS2L11*
chr8:144,959,805-145,133,898	CN Loss	174094	8	12	q24.3	*EPPK1, MIR661, MIR937, NRBP2, PARP10, PLEC, PUF60, SCRIB*
chr14:18,741,463-19,490,518	CN Gain	749056	7	23	q11.1 - q11.2	*OR4K1, OR4K2, OR4K5, OR4M1, OR4N2, OR4Q3, P704P*
chr14:105,602,384-105,854,362	CN Gain	251979	0	23	q32.33	
chr15:28,271,861-28,589,267	CN Loss	317407	2	11	q13.2	*CHRFAM7A, DKFZP434L187*
chr16:31,908,350-32,582,495	CN Gain	674146	1	24	p11.2	*HERC2P4*
chr16:34,328,350-34,611,145	High Copy Gain	282796	2	17	p11.2 - p11.1	*LOC146481, LOC283914*
chr18:1,706,128-1,828,669	CN Loss	122542	0	7	p11.32	
chr20:27,100,000-28,266,113	CN Loss	1166114	1	6	q11.1	*FRG1B*
chr20:60,516,643-60,570,962	CN Loss	54320	3	4	q13.33	*C20orf166, C20orf200, MIR1-1*

Table S3. 14 - CNVs found for patient 14

Chromosome Region	Event	Length	Genes	Probes	Cytoband	Gene Symbols
chr1:147,255,457-147,526,118	CN Loss	270662	0	15	q21.1	
chr1:194,993,686-195,078,812	Homozygous Copy Loss	85127	2	8	q31.3	*CFHR1, CFHR3*
chr3:196,826,201-196,942,769	CN Gain	116569	3	6	q29	*MIR570, MUC20, SDHAP2*
chr4:68,872,534-69,207,266	CN Loss	334733	4	21	q13.2	*TMPRSS11E, UGT2B15, UGT2B17, YTHDC1*
chr5:769,844-827,999	High Copy Gain	58156	0	5	p15.33	
chr5:827,999-900,209	CN Gain	72211	1	8	p15.33	*ZDHHC11*
chr5:68,892,660-69,300,554	CN Gain	407895	7	14	q13.2	*GTF2H2, GTF2H2B, GTF2H2C, GTF2H2D, GUSBP3, LOC100272216, LOC653391*
chr5:69,585,286-69,823,666	High Copy Gain	238381	4	4	q13.2	*GTF2H2B, GTF2H2C, GTF2H2D, LOC653391*
chr7:72,091,520-72,241,581	CN Gain	150062	5	6	q11.23	*GTF2IP1, LOC100093631, PMS2L2, SPDYE8P, STAG3L3*
chr7:72,379,797-72,783,852	CN Loss	404056	11	28	q11.23	*BAZ1B, BCL7B, DNAJC30, FKBP6, FZD9, MLXIPL, STX1A, TBL2, TRIM50, VPS37D, WBSCR22*
chr8:39,349,059-39,439,185	High Copy Gain	90127	2	6	p11.23	*ADAM3A, ADAM5P*
chr9:94,114,736-94,296,418	CN Gain	181683	5	13	q22.31	*ASPN, CENPP, NOL8, OGN, OMD*
chr10:133,163,835-133,323,898	CN Loss	160064	0	8	q26.3	
chr12:9,519,693-9,606,536	High Copy Gain	86844	0	6	p13.31	
chr14:19,122,778-19,490,518	CN Gain	367741	6	15	q11.2	*OR4K1, OR4K2, OR4K5, OR4M1, OR4N2, OR4Q3*
chr14:105,602,384-105,637,917	Homozygous Copy Loss	35534	0	3	q32.33	
chr15:19,074,651-19,777,939	CN Gain	703289	5	22	q11.2	*BCL8, CXADRP2, LOC646214, NF1P1, POTEB*
chr15:19,777,939-20,154,336	High Copy Gain	376398	4	18	q11.2	*LOC727924, OR4M2, OR4N3P, OR4N4*
chr16:32,848,712-33,548,689	CN Gain	699978	2	27	p11.2	*TP53TG3, TP53TG3B*
chr16:34,310,061-34,625,973	CN Gain	315913	2	19	p11.2 - p11.1	*LOC146481, LOC283914*
chr17:33,379,598-33,681,159	CN Loss	301562	2	5	q12	*TBC1D3, TBC1D3F*
chr19:48,342,188-48,449,551	CN Loss	107364	3	7	q13.31	*PSG4, PSG5, PSG9*
chr21:9,896,630-10,198,129	CN Gain	301500	6	14	p11.2 - p11.1	*BAGE, BAGE2, BAGE3, BAGE4, BAGE5, TPTE*
chr22:37,683,134-37,724,108	Homozygous Copy Loss	40975	2	4	q13.1	*APOBEC3A, APOBEC3B*

Table S3. 15 - CNVs found for patient 15

Chromosome Region	Event	Length	Genes	Probes	Cytoband	Gene Symbols
chr1:120,339,508-120,431,061	CN Loss	91554	1	7	p12	*NOTCH2*
chr1:143,921,320-144,079,918	CN Loss	158599	2	9	q21.1	*NBPF10, NOTCH2NL*
chr1:146,886,639-148,058,580	CN Loss	1171942	11	36	q21.1 - q21.2	*FCGR1A, FCGR1C, HIST2H2BF, HIST2H3D, LOC388692, LOC645166, LOC728855, NBPF16, PPIAL4A, PPIAL4C, PPIAL4E*
chr1:221,119,130-221,263,305	CN Loss	144176	1	9	q41	*DISP1*
chr2:89,265,770-89,895,566	CN Loss	629797	0	16	p11.2	
chr2:97,134,097-97,288,036	CN Loss	153940	1	8	q11.2	*ANKRD36*
chr2:132,157,969-132,284,054	CN Loss	126086	2	11	q21.1 - q21.2	*C2orf27A, C2orf27B*
chr4:69,008,654-69,117,422	CN Loss	108769	2	7	q13.2	*TMPRSS11E, UGT2B17*
chr5:742,057-900,209	CN Gain	158153	2	12	p15.33	*TPPP, ZDHHC11*
chr5:68,866,395-70,671,221	CN Loss	1804827	15	42	q13.2	*GTF2H2, GTF2H2B, GTF2H2C, GTF2H2D, GUSBP3, LOC100170939, LOC100272216, LOC647859, LOC653391, NAIP, OCLN, SERF1A, SERF1B, SMN1, SMN2*
chr5:175,557,661-175,689,861	CN Loss	132201	1	9	q35.2	*C5orf25*
chr6:193,938-327,656	CN Loss	133719	1	9	p25.3	*DUSP22*
chr7:61,060,605-61,374,646	CN Loss	314042	0	4	q11.1 - q11.21	
chr7:72,052,516-72,359,842	CN Loss	307327	12	14	q11.23	*GTF2IP1, GTF2IRD2P1, LOC100093631, NCF1B, NSUN5, NSUN5C, NSUN5P2, PMS2L2, POM121, SPDYE8P, STAG3L3, TRIM74*
chr7:143,596,593-143,693,789	CN Loss	97197	6	4	q35	*ARHGEF5, CTAGE4, OR2A1, OR2A20P, OR2A42, OR2A9P*
chr8:7,121,948-8,124,705	CN Loss	1002758	30	31	p23.1	*DEFB103A, DEFB103B, DEFB104A, DEFB104B, DEFB105A, DEFB105B, DEFB106A, DEFB106B, DEFB107A, DEFB107B, DEFB109P1B, DEFB4, FAM66B, FAM66E, FAM90A10, FAM90A13, FAM90A14, FAM90A18, FAM90A19, FAM90A20, FAM90A5, FAM90A7, FAM90A8, FAM90A9, FLJ10661, LOC349196, MIR548I3, SPAG11A, SPAG11B, ZNF705G*
chr8:11,934,392-12,554,253	CN Loss	619862	9	16	p23.1	*DEFB109P1, DEFB130, FAM66A, FAM66D, FAM86B1, FAM86B2, LOC392196, USP17L2, ZNF705D*
chr8:39,349,059-39,508,365	High Copy Gain	159307	2	9	p11.23 - p11.22	*ADAM3A, ADAM5P*
chr8:137,744,855-137,920,248	CN Loss	175394	0	8	q24.23	
chr9:42,403,900-42,625,244	CN Loss	221345	1	8	p11.2	*FAM95B1*
chr10:38,856,419-39,080,121	CN Loss	223703	0	5	p11.1	
chr10:46,907,108-47,077,489	CN Loss	170382	0	5	q11.22	
chr10:81,342,341-81,667,828	CN Loss	325488	3	9	q22.3	*LOC642361, LOC650623, SFTPA1*
chr11:4,180,454-4,346,983	CN Loss	166530	1	8	p15.4	*OR52B4*
chr14:18,149,473-18,640,884	CN Loss	491412	2	11	q11.1	*OR11H12, POTEG*
chr14:18,704,526-19,528,037	CN Loss	823512	8	29	q11.1 - q11.2	*OR4K1, OR4K15, OR4K2, OR4K5, OR4M1, OR4N2, OR4Q3, P704P*
chr14:40,669,880-40,734,612	CN Loss	64733	0	3	q21.1	
chr15:19,101,155-19,796,307	CN Gain	695153	6	22	q11.2	*BCL8, CXADRP2, LOC646214, LOC727924, NF1P1, POTEB*
chr16:381,902-488,401	CN Loss	106500	4	10	p13.3	*DECR2, LOC100134368, NME4, RAB11FIP3*
chr16:31,389,482-31,419,934	CN Gain	30453	3	5	p11.2	*C16orf58, SLC5A2, TGFB1I1*
chr16:32,079,878-33,691,406	CN Loss	1611529	3	55	p11.2	*SLC6A10P, TP53TG3, TP53TG3B*
chr16:34,345,125-34,656,892	CN Loss	311768	2	19	p11.2 - p11.1	*LOC146481, LOC283914*
chr17:31,459,116-31,762,515	CN Loss	303400	7	8	q12	*CCL3L1, CCL3L3, CCL4L1, CCL4L2, TBC1D3B, TBC1D3C, TBC1D3F*
chr17:41,052,171-41,334,928	CN Gain	282758	7	16	q21.31	*C17orf69, CRHR1, IMP5, LOC100128977, LOC100130148, MAPT, MGC57346*
chr17:41,577,131-41,768,841	CN Gain	191711	3	10	q21.31	*ARL17B, KIAA1267, LRRC37A*

chr19:0-213,756	CN Loss	213757	6	3	p13.3	*FAM138A, FAM138C, FAM138F, FLJ45445, LOC375690, OR4F17, hsa-mir-1302-2*
chr21:12,300,000-13,397,475	CN Loss	1097476	1	5	q11.1 - q11.2	*C21orf99*
chr22:23,983,829-24,247,592	CN Loss	263764	2	15	q11.23	*IGLL3, LRP5L*

Table S3. 16 - CNVs found for patient 16

Chromosome Region	Event	Length	Genes	Probes	Cytoband	Gene Symbols
chr1:2,311,143-2,437,924	CN Gain	126782	5	9	p36.32	*MORN1, PANK4, PEX10, PLCH2, RER1*
chr1:72,532,314-72,590,954	Homozygous Copy Loss	58641	0	3	p31.1	
chr1:103,905,063-104,103,459	CN Gain	198397	6	9	p21.1	*AMY1A, AMY1B, AMY1C, AMY2A, AMY2B, LOC648740*
chr1:120,594,254-120,815,510	CN Loss	221257	3	7	p12 - p11.2	*FAM72B, FCGR1B, HIST2H2BA*
chr1:147,255,457-147,430,850	Homozygous Copy Loss	175394	0	8	q21.1	
chr1:147,430,850-147,503,409	CN Loss	72560	0	5	q21.1	
chr1:147,526,118-148,032,481	CN Loss	506364	7	15	q21.1 - q21.2	*FCGR1A, FCGR1C, HIST2H2BF, LOC388692, LOC728855, PPIAL4A, PPIAL4C*
chr1:194,993,686-195,078,812	CN Loss	85127	2	8	q31.3	*CFHR1, CFHR3*
chr1:246,667,935-246,726,094	CN Loss	58160	3	5	q44	*OR2T2, OR2T3, OR2T5*
chr2:91,095,408-91,376,821	CN Loss	281414	2	11	p11.1	*GGT8P, LOC654342*
chr3:163,987,281-164,108,596	Homozygous Copy Loss	121316	0	7	q26.1	
chr4:69,063,649-69,170,341	CN Loss	106693	1	9	q13.2	*UGT2B17*
chr4:191,050,769-191,127,447	CN Loss	76679	1	8	q35.2	*FRG1*
chr5:745,872-900,209	High Copy Gain	154338	2	15	p15.33	*TPPP, ZDHHC11*
chr5:68,877,031-69,300,554	CN Loss	423524	9	15	q13.2	*GTF2H2, GTF2H2B, GTF2H2C, GTF2H2D, GUSBP3, LOC100272216, LOC647859, LOC653391, OCLN*
chr6:29,951,068-30,013,265	CN Loss	62198	3	4	p21.33	*HCG2P7, HCG4P6, HLA-H*
chr7:0-171,033	CN Gain	171034	0	3	p22.3	
chr7:1,345,917-1,488,005	CN Gain	142089	2	9	p22.3	*INTS1, MICALL2*
chr7:132,827,191-134,903,411	.	2076221	15	131	q33	*AGBL3, AKR1B1, AKR1B10, AKR1B15, BPGM, C7orf49, CALD1, CNOT4, EXOC4, LRGUK, NUP205, SLC35B4, STRA8, TMEM140, WDR91*
chr7:143,509,786-143,667,377	CN Gain	157592	7	7	q35	*ARHGEF5L, CTAGE4, OR2A1, OR2A20P, OR2A42, OR2A7, OR2A9P*
chr8:39,349,059-39,526,316	CN Loss	177258	2	11	p11.23 - p11.22	*ADAM3A, ADAM5P*
chr9:138,350,119-138,387,209	CN Gain	37091	3	3	q34.3	*CARD9, DNLZ, GPSM1*
chr9:139,219,916-139,267,322	CN Gain	47407	7	5	q34.3	*C9orf169, C9orf173, FAM166A, NDOR1, RNF208, SLC34A3, TUBB2C*
chr14:18,555,363-19,490,518	CN Loss	935156	8	33	q11.1 - q11.2	*OR4K1, OR4K2, OR4K5, OR4M1, OR4N2, OR4Q3, P704P, POTEG*
chr15:18,467,080-19,777,939	CN Gain	1310860	7	39	q11.2	*BCL8, CXADRP2, GOLGA6L6, GOLGA8C, LOC646214, NF1P1, POTEB*
chr16:1,754,188-1,778,414	CN Gain	24227	6	4	p13.3	*EME2, MAPK8IP3, MRPS34, NME3, NUBP2, SPSB3*
chr16:33,266,719-33,485,002	CN Gain	218284	0	17	p11.2	
chr17:31,449,032-31,541,638	CN Loss	92607	2	7	q12	*CCL4, TBC1D3B*
chr20:27,100,000-28,266,113	CN Loss	1166114	1	6	q11.1	*FRG1B*
chr22:14,433,473-14,831,547	CN Loss	398075	2	4	q11.1	*OR11H1, POTEH*
chr22:17,036,669-17,240,532	CN Loss	203864	2	5	q11.21	*GGT3P, USP18*

Table S3. 17 - CNVs found for patient 17

Chromosome Region	Event	Length	Genes	Probes	Cytoband	Gene Symbols
chr1:17,077,550-17,131,728	CN Loss	54179	1	4	p36.13	*CROCC*
chr1:72,532,314-72,590,954	Homozygous Copy Loss	58641	0	3	p31.1	
chr2:75,302,024-76,665,026	CN Gain	1363003	3	82	p13.1 - p12	*C2orf3, FAM176A, MRPL19*
chr2:110,037,747-110,667,961	CN Gain	630215	7	21	q13	*LIMS3, LIMS3-LOC440895, LOC151009, LOC440895, MALL, NCRNA00116, NPHP1*
chr3:163,987,281-164,108,596	High Copy Gain	121316	0	7	q26.1	
chr4:34,448,772-34,500,505	CN Gain	51734	0	4	p15.1	
chr4:69,008,654-69,170,341	CN Gain	161688	2	11	q13.2	*TMPRSS11E, UGT2B17*
chr4:70,151,754-70,345,475	CN Loss	193722	1	11	q13.2	*UGT2B28*
chr5:769,844-857,045	CN Gain	87202	1	8	p15.33	*ZDHHC11*
chr5:12,845,443-13,146,867	CN Loss	301425	0	8	p15.2	
chr6:193,938-244,003	CN Loss	50066	1	3	p25.3	*DUSP22*
chr6:244,003-343,443	CN Gain	99441	2	7	p25.3	*DUSP22, IRF4*
chr6:32,563,029-32,634,334	Homozygous Copy Loss	71306	2	7	p21.32	*HLA-DRB5, HLA-DRB6*
chr6:32,634,334-32,711,050	CN Loss	76717	2	7	p21.32	*HLA-DRB1, HLA-DRB6*
chr6:115,364,925-115,464,319	High Copy Gain	99395	0	5	q22.1	
chr6:115,464,319-115,567,351	CN Gain	103033	0	5	q22.1	
chr7:43,970,243-44,035,484	High Copy Gain	65242	3	4	p13	*POLR2J4, RASA4P, SPDYE1*
chr7:45,762,511-45,827,273	CN Gain	64763	1	4	p13	*13-Set*
chr7:143,509,786-143,693,789	CN Loss	184004	8	8	q35	*ARHGEF5, ARHGEF5L, CTAGE4, OR2A1, OR2A20P, OR2A42, OR2A7, OR2A9P*
chr8:7,214,279-7,942,536	CN Loss	728258	23	24	p23.1	*DEFB103A, DEFB103B, DEFB104A, DEFB104B, DEFB105A, DEFB105B, DEFB106A, DEFB106B, DEFB107A, DEFB107B, DEFB109P1B, DEFB4, FAM66E, FAM90A10, FAM90A13, FAM90A14, FAM90A18, FAM90A19, FAM90A7, FAM90A8, FAM90A9, SPAG11A, SPAG11B*
chr8:39,349,059-39,508,365	Homozygous Copy Loss	159307	2	10	p11.23 - p11.22	*ADAM3A, ADAM5P*
chr9:134,757,732-140,273,252	CN Gain	5515521	135	371	q34.13 - q34.3	*ABCA2, ABO, ADAMTS13, ADAMTSL2, AGPAT2, ANAPC2, ARRDC1, BRD3, C8G, C9orf116, C9orf139, C9orf140, C9orf142, C9orf163, C9orf167, C9orf169, C9orf172, C9orf173, C9orf37, C9orf69, C9orf7, C9orf75, C9orf86, C9orf96, CACNA1B, CAMSAP1, CARD9, CEL, CELP, CLIC3, COBRA1, COL5A1, DBH, DNLZ, DPP7, EDF1, EGFL7, EHMT1, ENTPD2, ENTPD8, EXD3, FAM157B, FAM163B, FAM166A, FAM69B, FBXW5, FCN1, FCN2, FLJ40292, FUT7, GBGT1, GFI1B, GLT6D1, GPSM1, GRIN1, GTF3C5, INPP5E, KCNT1, KIAA0649, KIAA1984, LCN1, LCN10, LCN12, LCN15, LCN6, LCN8, LCN9, LCNL1, LHX3, LOC100131193, LOC100289341, LOC26102, LRRC26, MAMDC4, MAN1B1, MED22, MIR126, MIR602, MRPL41, MRPS2, NACC2, NCRNA00094, NDOR1, NELF, NOTCH1, NOXA1, NPDC1, NRARP, OBP2A, OBP2B, OLFM1, PAEP, PHPT1, PMPCA, PNPLA7, PTGDS, QSOX2, RALGDS, REXO4, RNF208, RNU6ATAC, RPL7A, RXRA, SARDH, SDCCAG3, SEC16A, SLC2A6, SLC34A3, SNAPC4, SNHG7, SNORA17, SNORA43, SNORD24, SNORD36A, SNORD36B, SNORD36C, SOHLH1, SSNA1, SURF1, SURF2, SURF4, SURF6, TMEM141, TMEM203, TMEM8C, TRAF2, TSC1, TUBB2C, TUBBP5, UAP1L1, UBAC1, VAV2, WDR5, WDR85, ZMYND19*

chr10:81,172,311-81,570,873	CN Loss	398563	3	5	q22.3	*LOC650623, SFTPA1, SFTPA2*
chr14:19,304,869-19,490,518	CN Loss	185650	5	10	q11.2	*OR4K1, OR4K2, OR4K5, OR4M1, OR4N2*
chr14:21,641,899-22,013,947	CN Loss	372049	0	23	q11.2	
chr14:102,028,863-103,648,365	CN Gain	1619503	22	105	q32.31 - q32.33	*AMN, ANKRD9, ASPG, BAG5, C14orf153, C14orf2, C14orf73, CDC42BPB, CKB, EIF5, KLC1, MARK3, PPP1R13B, RCOR1, SNORA28, TDRD9, TECPR2, TNFAIP2, TRAF3, TRMT61A, XRCC3, ZFYVE21*
chr14:104,175,876-105,602,384	CN Gain	1426509	24	106	q32.33	*ADAM6, ADSSL1, AHNAK2, AKT1, BRF1, BTBD6, C14orf79, C14orf80, CDCA4, CRIP1, CRIP2, GPR132, INF2, JAG2, KIAA0125, KIAA0284, MGC23270, MTA1, NUDT14, PACS2, PLD4, SIVA1, TMEM121, ZBTB42*
chr14:105,602,384-105,637,917	High Copy Gain	35534	0	3	q32.33	
chr14:105,637,917-105,765,096	CN Gain	127180	0	13	q32.33	
chr14:105,954,649-106,012,840	CN Loss	58192	1	4	q32.33	*LOC100133469*
chr15:18,717,290-18,902,988	CN Gain	185699	0	6	q11.2	
chr15:19,074,651-20,154,336	High Copy Gain	1079686	9	40	q11.2	*BCL8, CXADRP2, LOC646214, LOC727924, NF1P1, OR4M2, OR4N3P, OR4N4, POTEB*
chr16:33,230,071-33,345,494	CN Gain	115424	0	8	p11.2	
chr16:33,345,494-33,535,805	CN Loss	190312	0	13	p11.2	
chr16:34,328,350-34,611,145	CN Gain	282796	2	17	p11.2 - p11.1	*LOC146481, LOC283914*
chr17:21,285,538-21,447,302	CN Loss	161765	1	9	p11.2	*C17orf51*
chr17:35,072,144-35,116,013	CN Gain	43870	5	5	q12	*ERBB2, PGAP3, PNMT, STARD3, TCAP*
chr22:17,036,669-17,240,532	CN Gain	203864	2	5	q11.21	*GGT3P, USP18*
chr22:18,708,904-19,044,001	CN Gain	335098	2	10	q11.21	*PI4KAP1, RIMBP3*
chrX:44,334,447-44,423,226	High Copy Gain	88780	0	5	p11.3	
chrX:114,131,819-114,286,055	CN Loss	154237	2	11	q23	*IL13RA2, LRCH2*
chrX:148,485,054-148,559,717	High Copy Gain	74664	1	6	q28	*TMEM185A*

Table S3. 18 - CNVs found for patient 18

Chromosome Region	Event	Length	Genes	Probes	Cytoband	Gene Symbols
chr1:16,778,535-16,987,680	CN Gain	209146	5	13	p36.13	CROCCL1, ESPNP, MST1P2, MST1P9, NBPF1
chr1:246,781,770-246,863,559	CN Loss	81790	4	6	q44	OR2T10, OR2T11, OR2T29, OR2T34
chr2:89,064,440-89,895,566	CN Loss	831127	0	24	p11.2	
chr2:90,979,707-91,625,698	CN Loss	645992	3	23	p11.2 - p11.1	FKSG73, GGT8P, LOC654342
chr4:69,063,649-69,207,266	CN Gain	143618	2	11	q13.2	UGT2B15, UGT2B17
chr5:745,872-857,045	CN Loss	111174	2	10	p15.33	TPPP, ZDHHC11
chr6:32,563,029-32,634,334	Homozygous Copy Loss	71306	2	7	p21.32	HLA-DRB5, HLA-DRB6
chr7:0-171,033	CN Gain	171034	0	3	p22.3	
chr7:143,509,786-143,705,809	CN Gain	196024	8	9	q35	ARHGEF5, ARHGEF5L, CTAGE4, OR2A1, OR2A20P, OR2A42, OR2A7, OR2A9P
chr8:7,026,686-7,803,775	CN Loss	777090	27	29	p23.1	DEFB103A, DEFB103B, DEFB104A, DEFB104B, DEFB105A, DEFB105B, DEFB106A, DEFB106B, DEFB107A, DEFB107B, DEFB109P1B, DEFB4, FAM66B, FAM90A10, FAM90A13, FAM90A14, FAM90A18, FAM90A19, FAM90A20, FAM90A5, FAM90A7, FAM90A8, FAM90A9, LOC349196, SPAG11A, SPAG11B, ZNF705G
chr8:39,349,059-39,526,316	CN Loss	177258	2	11	p11.23 - p11.22	ADAM3A, ADAM5P
chr9:38,748,218-38,887,219	CN Gain	139002	0	9	p13.1	
chr9:122,564,885-122,884,841	CN Gain	319957	7	21	q33.2	C5, FBXW2, LOC253039, LOC402377, PHF19, PSMD5, TRAF1
chr9:122,884,841-126,227,440	High Copy Gain	3342600	45	211	q33.2 - q33.3	C9orf45, CEP110, CRB2, DAB2IP, DENND1A, GGTA1, GPR21, GSN, LHX2, LHX6, LOC100129034, MIR548D1, MIR600, MIR601, MORN5, MRRF, NDUFA8, NEK6, OR1B1, OR1J1, OR1J2, OR1J4, OR1K1, OR1L1, OR1L3, OR1L4, OR1L6, OR1L8, OR1N1, OR1N2, OR1Q1, OR5C1, PDCL, PSMB7, PTGS1, RAB14, RABGAP1, RBM18, RC3H2, SNORD90, STOM, STRBP, TTLL11, ZBTB26, ZBTB6
chr12:7,885,718-8,020,685	CN Gain	134968	2	8	p13.31	SLC2A14, SLC2A3
chr12:9,519,693-9,580,901	Homozygous Copy Loss	61209	0	4	p13.31	
chr12:31,129,983-31,239,395	CN Loss	109413	1	8	p11.21	DDX11
chr15:18,985,008-20,154,336	CN Gain	1169329	11	42	q11.2	BCL8, CXADRP2, GOLGA6L6, GOLGA8C, LOC646214, LOC727924, NF1P1, OR4M2, OR4N3P, OR4N4, POTEB
chr16:22,377,707-22,615,315	CN Loss	237609	5	10	p12.1	LOC100132247, LOC100190986, LOC641298, LOC653786, NPIPL3
chr17:31,449,032-31,762,515	CN Loss	313484	8	9	q12	CCL3L1, CCL3L3, CCL4, CCL4L1, CCL4L2, TBC1D3B, TBC1D3C, TBC1D3F
chr17:41,535,964-42,142,705	CN Loss	606742	6	21	q21.31 - q21.32	ARL17A, ARL17B, KIAA1267, LRRC37A, LRRC37A2, NSF
chr22:43,812,038-43,961,595	CN Loss	149558	1	8	q13.31	NUP50

Table S3. 19 - CNVs found for patient 19

Chromosome Region	Event	Length	Genes	Probes	Cytoband	Gene Symbols
chr1:12,867,822-13,013,869	CN Loss	146048	6	6	p36.21	PRAMEF10, PRAMEF22, PRAMEF4, PRAMEF6, PRAMEF7, PRAMEF8
chr1:16,860,419-16,943,714	CN Gain	83296	1	5	p36.13	ESPNP
chr1:194,979,783-195,078,812	CN Loss	99030	3	9	q31.3	CFH, CFHR1, CFHR3
chr1:246,781,770-246,882,848	CN Gain	101079	6	7	q44	OR2T10, OR2T11, OR2T27, OR2T29, OR2T34, OR2T35
chr2:110,194,966-110,427,222	CN Gain	232257	3	12	q13	MALL, NCRNA00116, NPHP1
chr4:48,818,145-48,953,993	CN Loss	135849	0	5	p11	
chr4:69,008,654-69,106,362	CN Loss	97709	2	6	q13.2	TMPRSS11E, UGT2B17
chr6:229,040-327,656	CN Loss	98617	1	7	p25.3	DUSP22
chr7:57,806,031-57,931,397	CN Loss	125367	0	4	p11.1	
chr7:143,509,786-143,693,789	CN Loss	184004	8	8	q35	ARHGEF5, ARHGEF5L, CTAGE4, OR2A1, OR2A20P, OR2A42, OR2A7, OR2A9P
chr8:39,349,059-39,508,365	CN Loss	159307	2	10	p11.23 - p11.22	ADAM3A, ADAM5P
chr10:131,264,691-131,920,458	CN Loss	655768	3	40	q26.3	EBF3, GLRX3, MGMT
chr14:18,436,931-18,640,884	CN Gain	203954	2	7	q11.1	OR11H12, POTEG
chr14:19,244,829-19,490,518	CN Gain	245690	6	13	q11.2	OR4K1, OR4K2, OR4K5, OR4M1, OR4N2, OR4Q3
chr15:18,822,832-18,902,988	High Copy Gain	80157	0	4	q11.2	
chr15:19,033,127-19,480,163	CN Gain	447037	6	19	q11.2	BCL8, CXADRP2, GOLGA8C, LOC646214, NF1P1, POTEB
chr15:19,796,307-20,154,336	High Copy Gain	358030	4	17	q11.2	LOC727924, OR4M2, OR4N3P, OR4N4
chr15:32,469,479-32,656,070	CN Loss	186592	2	14	q14	GOLGA8A, GOLGA8B
chr17:31,459,116-31,599,722	CN Loss	140607	5	7	q12	CCL3L1, CCL3L3, CCL4L1, CCL4L2, TBC1D3B
chr17:41,577,131-41,768,841	CN Gain	191711	3	11	q21.31	ARL17B, KIAA1267, LRRC37A
chr18:35,207,077-35,309,059	CN Loss	101983	1	6	q12.2	LOC647946
chr20:62,390,214-62,435,964	CN Loss	45751	0	4	q13.33	
chr22:14,433,473-14,875,192	CN Gain	441720	2	5	q11.1	OR11H1, POTEH
chr22:17,272,553-17,391,813	CN Gain	119261	5	11	q11.21	DGCR10, DGCR5, DGCR6, DGCR9, PRODH

Table S3. 20 - CNVs found for patient 20

Chromosome Region	Event	Length	Genes	Probes	Cytoband	Gene Symbols
chr1:147,095,314-147,503,409	CN Loss	408096	1	15	q21.1	*LOC645166*
chr1:174,497,709-174,552,884	CN Gain	55176	0	3	q25.2	
chr1:194,993,686-195,078,812	CN Loss	85127	2	8	q31.3	*CFHR1, CFHR3*
chr2:91,273,082-91,625,698	CN Gain	352617	2	11	p11.1	*FKSG73, GGT8P*
chr3:131,276,416-131,385,677	CN Gain	109262	2	6	q21.3	*ALG1L2, LOC729375*
chr3:163,987,281-164,108,596	CN Gain	121316	0	7	q26.1	
chr3:196,848,435-196,956,726	CN Loss	108292	3	7	q29	*MIR570, MUC20, SDHAP2*
chr4:4,033,732-4,174,459	CN Gain	140728	0	3	p16.2	
chr4:34,448,772-34,500,505	CN Loss	51734	0	4	p15.1	
chr4:68,978,187-69,165,944	CN Loss	187758	2	11	q13.2	*TMPRSS11E, UGT2B17*
chr4:70,195,060-70,269,497	CN Loss	74438	1	5	q13.2	*UGT2B28*
chr6:79,027,678-79,098,580	CN Gain	70903	0	4	q14.1	
chr7:38,262,476-38,336,127	CN Loss	73652	1	6	p14.1	*TARP*
chr7:61,060,605-61,574,690	CN Gain	514086	0	8	q11.1 - q11.21	
chr7:97,344,237-97,453,585	CN Gain	109349	2	7	q21.3	*MGC72080, OCM2*
chr8:7,214,279-8,124,705	CN Gain	910427	25	28	p23.1	*DEFB103A, DEFB103B, DEFB104A, DEFB104B, DEFB105A, DEFB105B, DEFB106A, DEFB106B, DEFB107A, DEFB107B, DEFB109P1B, DEFB4, FAM66E, FAM90A10, FAM90A13, FAM90A14, FAM90A18, FAM90A19, FAM90A7, FAM90A8, FAM90A9, FLJ10661, MIR548I3, SPAG11A, SPAG11B*
chr8:11,976,320-12,139,755	CN Gain	163436	5	5	p23.1	*FAM66D, FAM86B1, LOC392196, USP17L2, ZNF705D*
chr8:12,366,919-12,554,253	CN Gain	187335	0	4	p23.1	
chr8:39,349,059-39,508,365	Homozygous Copy Loss	159307	2	10	p11.23 - p11.22	*ADAM3A, ADAM5P*
chr10:38,856,419-39,080,121	CN Gain	223703	0	5	p11.1	
chr11:3,436,246-3,539,381	CN Gain	103136	0	3	p15.4	
chr12:113,989,883-115,926,066	CN Loss	857445	8	53	q24.21 - q24.22	*C12orf49, FBXW8, HRK, MAP1LC3B2, MED13L, MIR620, NCRNA00173, RNFT2*
chr14:18,520,876-19,490,518	CN Loss	969643	8	34	q11.1 - q11.2	*OR4K1, OR4K2, OR4K5, OR4M1, OR4N2, OR4Q3, P704P, POTEG*
chr14:43,993,122-44,089,370	CN Loss	96249	1	6	q21.3	*FSCB*
chr15:18,717,290-18,902,988	CN Loss	185699	0	6	q11.2	
chr15:19,101,155-20,154,336	CN Loss	1053182	9	39	q11.2	*BCL8, CXADRP2, LOC646214, LOC727924, NF1P1, OR4M2, OR4N3P, OR4N4, POTEB*
chr16:31,878,102-32,546,468	CN Gain	668367	1	24	p11.2	*HERC2P4*
chr16:33,345,494-33,548,689	CN Gain	203196	0	14	p11.2	
chr16:34,328,350-34,611,145	CN Gain	282796	2	17	p11.2 - p11.1	*LOC146481, LOC283914*
chr17:31,511,535-31,903,419	CN Loss	391885	9	6	q12	*CCL3L1, CCL3L3, CCL4L1, CCL4L2, TBC1D3B, TBC1D3C, TBC1D3F, TBC1D3G, TBC1D3H*
chr17:40,937,706-41,021,360	CN Gain	83655	1	3	q21.31	*LRRC37A4*
chr17:41,521,663-41,648,920	CN Gain	127258	1	9	q21.31	*KIAA1267*

chr19:47,927,296-48,058,449	CN Gain	131154	3	10	q13.31	*PSG10, PSG3, PSG8*
chr22:17,036,669-17,240,532	CN Gain	203864	2	5	q11.21	*GGT3P, USP18*
chr22:22,667,137-22,729,016	CN Gain	61880	4	5	q11.23	*GSTT1, GSTTP1, GSTTP2, LOC391322*
chrX:103,063,574-103,207,921	CN Gain	144348	4	10	q22.2	*H2BFM, H2BFWT, H2BFXP, TMSB15B*
chrX:140,498,089-140,604,390	CN Gain	106302	2	7	q27.2	*SPANXA1, SPANXA2*
chrX:154,043,595-154,085,916	CN Loss	42322	0	4	q28	

Table S3. 21 - CNVs found for patient 21

Chromosome Region	Event	Length	Genes	Probes	Cytoband	Gene Symbols
chr1:16,886,363-16,943,714	CN Loss	57352	1	4	p36.13	*ESPNP*
chr3:196,885,475-196,956,726	CN Loss	71252	3	6	q29	*MIR570, MUC20, SDHAP2*
chr4:34,448,772-34,500,505	CN Gain	51734	0	4	p15.1	
chr4:69,038,815-69,207,266	CN Loss	168452	3	12	q13.2	*TMPRSS11E, UGT2B15, UGT2B17*
chr5:754,690-874,147	CN Loss	119458	1	11	p15.33	*ZDHHC11*
chr5:68,866,395-70,452,813	CN Gain	1586419	15	37	q13.2	*GTF2H2, GTF2H2B, GTF2H2C, GTF2H2D, GUSBP3, LOC100170939, LOC100272216, LOC647859, LOC653391, NAIP, OCLN, SERF1A, SERF1B, SMN1, SMN2*
chr6:29,951,068-30,013,265	CN Loss	62198	3	4	p21.33	*HCG2P7, HCG4P6, HLA-H*
chr6:170,767,027-170,899,992	CN Loss	132966	0	4	q27	
chr8:7,319,013-7,775,352	CN Gain	456340	18	14	p23.1	*DEFB104A, DEFB104B, DEFB105A, DEFB105B, DEFB106A, DEFB106B, DEFB107A, DEFB107B, FAM90A10, FAM90A13, FAM90A14, FAM90A18, FAM90A19, FAM90A7, FAM90A8, FAM90A9, SPAG11A, SPAG11B*
chr8:12,233,164-12,529,929	High Copy Gain	296766	3	6	p23.1	*DEFB109P1, FAM66A, FAM86B2*
chr8:39,332,477-39,473,281	CN Loss	140805	2	9	p11.23	*ADAM3A, ADAM5P*
chr9:2,917,236-3,001,419	CN Loss	84184	0	4	p24.2	
chr9:39,131,028-39,410,420	CN Gain	279393	3	13	p13.1	*CNTNAP3, FAM75A1, FAM75A2*
chr11:38,952,791-38,996,778	CN Gain	43988	0	3	p12	
chr12:9,519,693-9,591,347	Homozygous Copy Loss	71655	0	5	p13.31	
chr12:31,185,572-31,239,395	CN Loss	53824	0	4	p11.21	
chr12:85,034,910-85,215,631	Homozygous Copy Loss	180722	1	11	q21.31 - q21.32	*MGAT4C*
chr12:85,215,631-85,399,218	CN Loss	183588	1	11	q21.32	*MGAT4C*
chr12:114,893,119-115,188,686	CN Gain	295568	2	19	q24.21	*MED13L, MIR620*
chr16:33,740,983-33,798,908	High Copy Gain	57926	0	3	p11.2	
chr16:34,328,350-34,611,145	High Copy Gain	282796	2	17	p11.2 - p11.1	*LOC146481, LOC283914*
chr17:33,338,413-33,649,919	CN Loss	311507	2	4	q12	*TBC1D3, TBC1D3F*
chr17:41,521,663-41,713,328	CN Loss	191666	1	13	q21.31	*KIAA1267*
chr18:0-131,060	CN Gain	131061	0	5	p11.32	
chr19:0-572,272	CN Gain	572273	19	33	p13.3	*BSG, C19orf20, C2CD4C, CDC34, FAM138A, FAM138C, FAM138F, FLJ45445, GZMM, HCN2, LOC375690, MADCAM1, MIER2, ODF3L2, OR4F17, POLRMT, PPAP2C, SHC2, THEG*
chrX:90,045,791-90,178,305	CN Gain	132515	0	7	q21.31	
chrX:149,223,089-149,345,498	CN Gain	122410	0	7	q28	

Table S3. 22 - CNVs found for patient 22

Chromosome Region	Event	Length	Genes	Probes	Cytoband	Gene Symbols
chr1:147,408,706-147,526,118	CN Loss	117413	0	8	q21.1	
chr2:87,237,617-87,750,596	CN Gain	512980	1	26	p11.2	*NCRNA00152*
chr3:163,987,281-164,108,596	High Copy Gain	121316	0	7	q26.1	
chr4:0-68,743	CN Loss	68744	2	5	p16.3	*ZNF595, ZNF718*
chr4:48,921,615-49,276,565	CN Loss	354951	0	4	p11	
chr4:69,038,815-69,185,257	CN Loss	146443	2	11	q13.2	*TMPRSS11E, UGT2B17*
chr5:0-140,194	CN Loss	140195	0	8	p15.33	
chr6:193,938-327,656	CN Loss	133719	1	9	p25.3	*DUSP22*
chr6:102,775,955-102,844,521	CN Loss	68567	0	4	q16.3	
chr8:7,214,279-7,942,536	CN Loss	728258	23	24	p23.1	*DEFB103A, DEFB103B, DEFB104A, DEFB104B, DEFB105A, DEFB105B, DEFB106A, DEFB106B, DEFB107A, DEFB107B, DEFB109P1B, DEFB4, FAM66E, FAM90A10, FAM90A13, FAM90A14, FAM90A18, FAM90A19, FAM90A7, FAM90A8, FAM90A9, SPAG11A, SPAG11B*
chr8:39,349,059-39,508,365	CN Loss	159307	2	10	p11.23 - p11.22	*ADAM3A, ADAM5P*
chr9:41,983,501-42,393,913	CN Loss	410413	3	19	p12	*ANKRD20A2, ANKRD20A3, KGFLP2*
chr12:9,519,693-9,606,536	High Copy Gain	86844	0	6	p13.31	
chr15:19,138,672-20,258,362	CN Loss	1119691	10	39	q11.2	*BCL8, CXADRP2, GOLGA8DP, LOC646214, LOC727924, NF1P1, OR4M2, OR4N3P, OR4N4, POTEB*
chr16:32,914,009-33,871,999	CN Loss	957991	2	38	p11.2	*TP53TG3, TP53TG3B*
chr17:31,459,116-32,016,320	CN Loss	557205	13	18	q12	*CCL3L1, CCL3L3, CCL4L1, CCL4L2, GGNBP2, MYO19, PIGW, TBC1D3B, TBC1D3C, TBC1D3F, TBC1D3G, TBC1D3H, ZNHIT3*
chr17:41,577,131-41,768,841	CN Gain	191711	3	11	q21.31	*ARL17B, KIAA1267, LRRC37A*
chr20:27,100,000-28,266,113	CN Loss	1166114	1	6	q11.1	*FRG1B*
chr20:42,717,234-48,284,251	CN Loss	5567018	88	363	q13.12 - q13.13	*ACOT8, ARFGEF2, B4GALT5, C20orf123, C20orf165, C20orf199, CD40, CDH22, CEBPB, CSE1L, CTSA, DBNDD2, DDX27, DNTTIP1, ELMO2, EYA2, KCNB1, KCNK15, KCNS1, LOC100131496, LOC100240726, LOC284749, MATN4, MIR1259, MMP9, NCOA3, NCOA5, NEURL2, PABPC1L, PCIF1, PI3, PIGT, PLTP, PREX1, PTGIS, RBPJL, RIMS4, RNF114, SDC4, SEMG1, SEMG2, SLC12A5, SLC13A3, SLC2A10, SLC35C2, SLC9A8, SLPI, SNAI1, SNORD12, SNORD12B, SNORD12C, SNX21, SPATA2, SPINLW1, SPINT3, SPINT4, STAU1, STK4, SULF2, SYS1, SYS1-DBNDD2, TMEM189, TMEM189-UBE2V1, TNNC2, TOMM34, TP53RK, TP53TG5, UBE2C, UBE2V1, WFDC10A, WFDC10B, WFDC11, WFDC12, WFDC13, WFDC2, WFDC3, WFDC5, WFDC6, WFDC8, WFDC9, WISP2, YWHAB, ZMYND8, ZNF334, ZNF335, ZNFX1, ZSWIM1, ZSWIM3*
chr21:10,003,548-10,198,129	CN Gain	194582	6	10	p11.1	*BAGE, BAGE2, BAGE3, BAGE4, BAGE5, TPTE*
chr22:48,295,275-48,356,493	CN Loss	61219	0	4	q13.33	

Table S3. 23 - CNVs found for patient 23

Chromosome Region	Event	Length	Genes	Probes	Cytoband	Gene Symbols
chr1:16,725,089-16,817,477	CN Gain	92389	2	5	p36.13	*CROCCL1, NBPF1*
chr1:16,817,477-16,962,625	High Copy Gain	145149	4	8	p36.13	*CROCCL1, ESPNP, MST1P2, MST1P9*
chr1:16,962,625-17,025,806	CN Gain	63182	1	3	p36.13	*MST1P9*
chr1:103,918,406-103,961,733	CN Gain	43328	2	3	p21.1	*AMY2A, AMY2B*
chr1:103,961,733-104,103,459	CN Loss	141727	4	5	p21.1	*AMY1A, AMY1B, AMY1C, AMY2A*
chr1:159,773,725-159,912,273	CN Loss	138549	5	10	q23.3	*FCGR2B, FCGR2C, FCGR3A, FCGR3B, HSPA7*
chr1:194,993,686-195,037,455	CN Loss	43770	1	4	q31.3	*CFHR3*
chr2:132,121,275-132,284,054	CN Loss	162780	2	13	q21.1 - q21.2	*C2orf27A, C2orf27B*
chr3:163,987,281-164,108,596	High Copy Gain	121316	0	7	q26.1	
chr4:69,038,815-69,165,944	CN Loss	127130	2	9	q13.2	*TMPRSS11E, UGT2B17*
chr5:794,486-866,296	CN Loss	71811	1	7	p15.33	*ZDHHC11*
chr5:68,991,456-69,300,554	CN Loss	309099	2	10	q13.2	*GUSBP3, LOC653391*
chr5:69,300,554-69,451,454	CN Gain	150901	5	5	q13.2	*LOC653391, SERF1A, SERF1B, SMN1, SMN2*
chr6:29,951,068-30,013,265	CN Loss	62198	3	4	p21.33	*HCG2P7, HCG4P6, HLA-H*
chr7:0-171,033	CN Gain	171034	0	3	p22.3	
chr7:76,255,603-76,451,792	CN Gain	196190	1	10	q11.23	*PMS2L11*
chr7:143,509,786-143,667,377	CN Gain	157592	7	7	q35	*ARHGEF5L, CTAGE4, OR2A1, OR2A20P, OR2A42, OR2A7, OR2A9P*
chr8:7,214,279-7,942,536	CN Loss	728258	23	24	p23.1	*DEFB103A, DEFB103B, DEFB104A, DEFB104B, DEFB105A, DEFB105B, DEFB106A, DEFB106B, DEFB107A, DEFB107B, DEFB109P1B, DEFB4, FAM66E, FAM90A10, FAM90A13, FAM90A14, FAM90A18, FAM90A19, FAM90A7, FAM90A8, FAM90A9, SPAG11A, SPAG11B*
chr8:12,270,256-12,366,919	Homozygous Copy Loss	96664	3	3	p23.1	*DEFB109P1, FAM66A, FAM86B2*
chr8:39,349,059-39,508,365	High Copy Gain	159307	2	10	p11.23 - p11.22	*ADAM3A, ADAM5P*
chr9:38,848,109-41,858,396	CN Loss	3010288	11	76	p13.1 - p12	*CNTNAP3, FAM74A1, FAM74A3, FAM75A1, FAM75A2, FAM75A3, FAM75A5, FAM75A7, LOC653501, ZNF658, ZNF658B*
chr9:43,361,452-43,613,184	CN Loss	251733	1	5	p11.2	*FAM75A6*
chr9:65,426,303-65,955,080	CN Loss	528778	0	6	q12	
chr14:18,149,473-19,490,518	CN Loss	1341046	9	40	q11.1 - q11.2	*OR11H12, OR4K1, OR4K2, OR4K5, OR4M1, OR4N2, OR4Q3, P704P, POTEG*
chr14:40,669,880-40,734,612	Homozygous Copy Loss	64733	0	3	q21.1	
chr14:105,602,384-105,637,917	CN Gain	35534	0	3	q32.33	
chr15:19,796,307-20,235,151	CN Loss	438845	4	18	q11.2	*LOC727924, OR4M2, OR4N3P, OR4N4*
chr16:14,942,869-15,031,536	CN Loss	88668	2	8	p13.11	*NPIP, PDXDC1*
chr16:34,328,350-34,611,145	CN Loss	282796	2	17	p11.2 - p11.1	*LOC146481, LOC283914*
chr16:68,697,986-68,774,654	CN Gain	76669	2	5	q22.1	*CLEC18C, PDPR*
chr17:41,577,131-41,713,328	CN Gain	136198	1	10	q21.31	*KIAA1267*
chr17:42,963,132-43,029,201	CN Loss	66070	1	5	q21.32	*NPEPPS*
chr22:14,433,473-15,135,777	CN Loss	702305	2	7	q11.1	*OR11H1, POTEH*
chr22:17,036,669-17,240,532	CN Gain	203864	2	5	q11.21	*GGT3P, USP18*
chrX:89,937,438-90,084,709	CN Loss	147272	0	6	q21.31	
chrX:119,476,634-119,789,009	High Copy Gain	312376	4	19	q24	*C1GALT1C1, CUL4B, LAMP2, MCTS1*

Table S3. 24 - CNVs found for patient 24

Chromosome Region	Event	Length	Genes	Probes	Cytoband	Gene Symbols
chr1:16,860,419-16,923,881	CN Loss	63463	2	4	p36.13	*AL035288, ESPNP*
chr1:16,923,881-17,025,806	CN Gain	101926	3	5	p36.13	*KIAA0445, MST1P9, MSTP9*
chr1:26,465,912-26,498,891	CN Gain	32980	5	3	p36.11	*CCDC21, CEP85, SH3BGRL3, UBXD5, UBXN11*
chr1:120,354,903-120,594,254	CN Gain	239352	3	11	p12	*N2, NOTCH2, NOTCH2NL*
chr1:143,698,897-144,079,918	CN Gain	381022	12	26	q21.1	*AK000887, BX647792, DQ586480, KIAA1245, KIAA1693, NBPF10, NBPF14, NBPF20, NBPF8, NOTCH2NL, PDE4DIP, SEC22B*
chr1:144,577,524-147,095,314	CN Gain	2517791	43	93	q21.1	*ACP6, AF131738, AF380582, AK023548, AK023809, AK091688, AK309393, BC036212, BC041003, BC110832, BCL9, CHD1L, CR610404, CR617196, DKFZp434H1228, DKFZp451B1418, DQ576969, FAM108A3, FMO5, GJA5, GJA8, GPR89A, GPR89B, GPR89C, KIAA1693, LOC200030, LOC728989, NBPF1, NBPF10, NBPF11, NBPF12, NBPF14, NBPF15, NBPF16, NBPF20, NBPF24, PDIA3P, PDZK1P1, PPIAL4D, PPIAL4E, PPIAL4F, PRKAB2, hPACPL1*
chr1:147,095,314-147,753,238	CN Loss	657925	8	23	q21.1	*AK310441, AK311729, BC023516, FCGR1B, FCGR1C, LOC388692, LOC645166, NBPF20*
chr1:147,753,238-147,887,475	CN Gain	134238	5	3	q21.1	*AB074166, AK307112, AK309334, BC005949, BC094809*
chr1:194,979,783-195,051,521	CN Gain	71739	3	6	q31.3	*CFH, CFHR1, CFHR3*
chr1:246,667,935-246,707,859	CN Loss	39925	2	4	q44	*OR2T2, OR2T3*
chr3:163,987,281-164,108,596	High Copy Gain	121316	1	7	q26.1	*BC073807*
chr4:48,868,216-49,276,565	CN Gain	408350	1	7	p11	*DQ593719*
chr5:70,312,515-70,375,604	CN Loss	63090	2	4	q13.2	*GTF2H2, NAIP*
chr6:32,563,029-32,600,341	CN Loss	37313	2	3	p21.32	*HLA-DRB1, HLA-DRB5*
chr6:79,027,678-79,098,580	CN Loss	70903	0	4	q14.1	
chr7:143,548,913-143,705,809	CN Gain	156897	9	8	q35	*ARHGEF5, BC040701, FKSG35, LOC728377, OR2A1, OR2A20P, OR2A42, OR2A7, OR2A9P*
chr8:7,214,279-7,638,263	CN Loss	423985	18	12	p23.1	*DEFB103A, DEFB103B, DEFB104A, DEFB104B, DEFB105A, DEFB105B, DEFB106A, DEFB106B, DEFB107A, DEFB107B, DEFB4, DEFB4B, FAM90A18, FAM90A19, FAM90A7, FAM90A8, HE2, SPAG11B*
chr8:7,638,263-7,741,568	Homozygous Copy Loss	103306	2	6	p23.1	*FAM90A10, FAM90A9*
chr8:7,741,568-7,942,536	CN Loss	200969	9	6	p23.1	*DEFB103A, DEFB4, DEFB4A, FAM66E, HE2, LOC100132396, SPAG11A, SPAG11B, ZNF705D*
chr8:39,349,059-39,526,316	CN Loss	177258	3	11	p11.23 - p11.22	*ADAM3A, ADAM5P, tMDC*
chr9:38,769,965-38,942,543	CN Gain	172579	0	10	p13.1	
chr9:39,061,678-39,251,004	CN Loss	189327	2	9	p13.1	*CNTNAP3, KIAA1714*
chr9:43,361,452-43,827,369	CN Loss	465918	3	9	p11.2	*AK054645, CNTNAP3B, FAM75A6*
chr9:44,722,011-46,339,951	CN Gain	1617941	9	33	p11.2	*AK131029, BC134347, CR615666, DQ594366, DQ594428, FAM27A, FAM27C, FAM27E3, LOC100132167*
chr9:65,222,385-65,792,243	CN Gain	569859	2	6	q12	*FAM74A2, FAM74A4*
chr12:9,519,693-9,591,347	High Copy Gain	71655	0	5	p13.31	
chr12:55,616,671-55,667,718	CN Loss	51048	1	4	q13.3	*RDH16*
chr14:19,244,829-19,490,518	CN Loss	245690	7	13	q11.2	*OR11H2, OR4K1, OR4K2, OR4K5, OR4M1, OR4N2, OR4Q3*
chr14:21,583,358-22,028,382	CN Loss	445025	31	28	q11.2	*AK093552, AK125397, AK310110, AV1S3A1T, AV25S1, AV2S1A1, AV30S1, AV4S1, TCR V alpha 8.1, J alpha IGRJa07, TCR-[alpha] V 33.1, TCR-alpha, TCRA, TCRAV14.1a, TCRAV8.1a, TCRAVN1, TCRD, TCRDV2, TCRVA13, TRA@, TRAC, TRD, TRD@, V alpha immunoglobulin, X61074, av27s1, hADV23S1, hADV29S1, hADV36S1, hADV38S2, hDV102S1, hDV103S1*

chr15:19,033,127-19,583,601	CN Loss	550475	14	21	q11.2	*A26B1, AL832227, BC047459, CXADRP2, DQ582025, DQ582260, DQ786202, GOLGA8C, GOLGA8E, LOC646214, M84131, NBEAP1, NF1P1, POTEB*
chr16:34,328,350-34,611,145	CN Gain	282796	6	17	p11.2 - p11.1	*BC023607, BC038761, BC045579, LOC100130700, LOC146481, LOC283914*
chr16:69,397,161-69,724,827	CN Gain	327667	6	21	q22.1 - q22.2	*DKFZp434D0513, DKFZp434I0850, DKFZp761B238, HYDIN, HYDIN2, KIAA1864*
chr17:41,555,382-41,768,841	CN Loss	213460	7	12	q21.31	*ARL17, ARL17B, ARL17P1, KIAA1267, LOC51326, LOC644246, LRRC37A*
chr17:41,768,841-42,128,214	CN Gain	359374	9	7	q21.31 - q21.32	*ARL17, ARL17A, ARL17B, ARL17P1, LOC51326, LRRC37A, LRRC37A2, NSF, NSFP1*
chr18:1,896,940-1,965,541	CN Loss	68602	0	4	p11.32	
chr18:25,844,492-26,393,582	CN Loss	549091	1	33	q12.1	*MIR302F*
chr19:61,855,379-61,922,492	CN Gain	67114	3	4	q13.43	*AK095199, LOC147670, ZNF835*
chr22:37,683,134-37,724,108	CN Loss	40975	2	4	q13.1	*APOBEC3A, APOBEC3B*
chrX:154,881,207-154,913,754	CN Gain	32548	3	6	q28	*IL9R, WASH1, interleukin 9 receptor*

Table S3. 25 - CNVs found for patient 25

Chromosome Region	Event	Length	Genes	Probes	Cytoband	Gene Symbols
chr1:72,532,314-72,590,954	Homozygous Copy Loss	58641	0	3	p31.1	
chr1:195,037,455-195,157,595	CN Loss	120141	2	9	q31.3	*CFHR1, CFHR4*
chr4:69,008,654-69,170,341	CN Loss	161688	2	11	q13.2	*TMPRSS11E, UGT2B17*
chr4:70,170,112-70,333,309	CN Loss	163198	1	9	q13.2	*UGT2B28*
chr4:190,715,999-190,919,040	CN Loss	203042	0	12	q35.2	
chr5:754,690-900,209	CN Loss	145520	1	14	p15.33	*ZDHHC11*
chr5:68,938,168-70,671,221	CN Gain	1733054	15	38	q13.2	*GTF2H2, GTF2H2B, GTF2H2C, GTF2H2D, GUSBP3, LOC100170939, LOC100272216, LOC647859, LOC653391, NAIP, OCLN, SERF1A, SERF1B, SMN1, SMN2*
chr6:57,304,911-57,774,413	CN Gain	469503	1	18	p11.2	*PRIM2*
chr6:57,917,805-58,142,695	CN Gain	224891	0	5	p11.2	
chr7:143,548,913-143,596,593	Homozygous Copy Loss	47681	6	3	q35	*CTAGE4, OR2A1, OR2A20P, OR2A42, OR2A7, OR2A9P*
chr7:143,596,593-143,667,377	CN Loss	70785	5	3	q35	*CTAGE4, OR2A1, OR2A20P, OR2A42, OR2A9P*
chr8:12,233,164-12,366,919	CN Gain	133756	3	4	p23.1	*DEFB109P1, FAM66A, FAM86B2*
chr8:39,349,059-39,508,365	Homozygous Copy Loss	159307	2	10	p11.23 - p11.22	*ADAM3A, ADAM5P*
chr10:48,381,494-49,040,567	CN Gain	659074	7	9	q11.22	*FAM25B, FAM25C, FAM25G, FRMPD2, LOC399753, PTPN20A, PTPN20B*
chr14:105,602,384-105,650,001	CN Gain	47618	0	4	q32.33	
chr15:18,467,080-19,074,651	CN Gain	607572	2	17	q11.2	*GOLGA6L6, GOLGA8C*
chr15:19,796,307-20,154,336	CN Loss	358030	4	17	q11.2	*LOC727924, OR4M2, OR4N3P, OR4N4*
chr16:14,748,555-15,045,422	CN Gain	296868	5	17	p13.11	*ABCC6P2, NOMO1, NPIP, NTAN1, PDXDC1*
chr16:15,391,681-16,215,845	CN Gain	824165	9	51	p13.11	*ABCC1, ABCC6, C16orf45, C16orf63, KIAA0430, MIR484, MPV17L, MYH11, NDE1*
chr17:31,459,116-31,762,515	CN Loss	303400	7	8	q12	*CCL3L1, CCL3L3, CCL4L1, CCL4L2, TBC1D3B, TBC1D3C, TBC1D3F*
chr17:41,648,920-42,128,214	CN Gain	479295	5	12	q21.31 - q21.32	*ARL17A, ARL17B, LRRC37A, LRRC37A2, NSF*
chr17:78,632,251-78,774,742	CN Loss	142492	1	4	q25.3	*METRNL*
chr22:18,708,904-18,851,705	CN Gain	142802	2	6	q11.21	*PI4KAP1, RIMBP3*
chr22:23,983,829-24,247,592	High Copy Gain	263764	2	19	q11.23	*IGLL3, LRP5L*

Table S3. 26 - CNVs found for patient 26

Chromosome Region	Event	Length	Genes	Probes	Cytoband	Gene Symbols
chr1:72,532,314-72,590,954	Homozygous Copy Loss	58641	0	3	p31.1	
chr1:120,406,782-120,528,618	CN Gain	121837	1	6	p12	*NOTCH2*
chr1:146,795,981-147,494,816	CN Gain	698836	4	19	q21.1	*LOC645166, NBPF15, NBPF16, PPIAL4E*
chr1:150,775,626-150,864,838	CN Gain	89213	5	6	q21.3	*LCE3A, LCE3B, LCE3C, LCE3D, LCE3E*
chr1:195,008,432-195,078,812	CN Loss	70381	2	7	q31.3	*CFHR1, CFHR3*
chr2:132,229,559-132,331,391	CN Gain	101833	2	6	q21.2	*C2orf27A, C2orf27B*
chr3:75,390,841-75,784,380	CN Gain	393540	2	16	p12.3	*FAM86D, MIR1324*
chr3:163,987,281-164,108,596	Homozygous Copy Loss	121316	0	7	q26.1	
chr3:196,826,201-197,001,270	CN Gain	175070	4	11	q29	*MIR570, MUC20, MUC4, SDHAP2*
chr4:8,982,874-9,383,286	CN Gain	400413	3	21	p16.1	*DEFB131, LOC650293, MIR548I2*
chr5:68,938,168-70,043,543	CN Gain	1105376	11	25	q13.2	*GTF2H2B, GTF2H2C, GTF2H2D, GUSBP3, LOC100170939, LOC100272216, LOC653391, SERF1A, SERF1B, SMN1, SMN2*
chr5:70,043,543-70,413,976	CN Loss	370434	12	7	q13.2	*GTF2H2, GTF2H2B, GTF2H2C, GTF2H2D, LOC647859, LOC653391, NAIP, OCLN, SERF1A, SERF1B, SMN1, SMN2*
chr6:29,951,068-29,994,193	CN Loss	43126	2	3	p21.33	*HCG2P7, HLA-H*
chr6:32,563,029-32,600,341	High Copy Gain	37313	1	3	p21.32	*HLA-DRB5*
chr6:32,600,341-32,725,877	CN Gain	125537	4	12	p21.32	*HLA-DQA1, HLA-DRB1, HLA-DRB5, HLA-DRB6*
chr6:79,027,678-79,098,580	CN Gain	70903	0	4	q14.1	
chr7:5,286,433-5,325,459	CN Gain	39027	2	4	p22.1	*SLC29A4, TNRC18*
chr7:43,970,243-44,035,484	CN Gain	65242	3	4	p13	*POLR2J4, RASA4P, SPDYE1*
chr7:101,892,363-102,138,397	CN Gain	246035	8	8	q22.1	*LRWD1, POLR2J, POLR2J2, POLR2J3, RASA4, SPDYE2, SPDYE2L, UPLP*
chr7:143,548,913-143,667,377	CN Loss	118465	6	6	q35	*CTAGE4, OR2A1, OR2A20P, OR2A42, OR2A7, OR2A9P*
chr8:11,953,787-12,601,257	CN Gain	647471	9	17	p23.1	*DEFB109P1, DEFB130, FAM66A, FAM66D, FAM86B1, FAM86B2, LOC392196, USP17L2, ZNF705D*
chr8:145,674,701-145,708,648	High Copy Gain	33948	4	4	q24.3	*GPT, MFSD3, PPP1R16A, RECQL4*
chr11:3,357,125-3,583,730	CN Gain	226606	1	8	p15.4	*LOC650368*
chr12:0-54,903	CN Gain	54904	2	4	p13.33	*FAM138D, IQSEC3*
chr14:18,750,780-19,507,271	CN Gain	756492	7	24	q11.1 - q11.2	*OR4K1, OR4K2, OR4K5, OR4M1, OR4N2, OR4Q3, P704P*
chr14:105,095,304-105,361,965	CN Gain	266662	0	20	q32.33	
chr15:19,922,429-20,154,336	CN Loss	231908	0	8	q11.2	
chr15:42,919,752-43,044,136	CN Gain	124385	1	8	q21.1	*C15orf43*
chr15:76,751,572-76,825,339	CN Gain	73768	0	5	q25.1	
chr16:0-36,949	CN Gain	36950	1	5	p13.3	*LOC100288778*
chr16:617,378-687,501	CN Gain	70124	11	11	p13.3	*C16orf13, FAM195A, FBXL16, JMJD8, RAB40C, RHBDL1, RHOT2, STUB1, WDR24, WDR90, WFIKKN1*
chr16:14,695,541-15,031,536	CN Gain	335996	5	18	p13.12 - p13.11	*ABCC6P2, NOMO1, NPIP, PDXDC1, PLA2G10*
chr16:15,384,327-16,351,469	CN Gain	967143	11	54	p13.11	*ABCC1, ABCC6, C16orf45, C16orf63, KIAA0430, LOC339047, MIR484, MPV17L, MYH11, NDE1, NOMO3*

chr17:0-80,275	CN Gain	80276	1	7	p13.3	*RPH3AL*
chr17:25,839,797-25,930,310	CN Gain	90514	3	8	q11.2	*GOSR1, LRRC37B2, TBC1D29*
chr17:41,521,663-41,703,816	CN Gain	182154	1	12	q21.31	*KIAA1267*
chr17:41,768,841-42,142,705	CN Loss	373865	5	8	q21.31 - q21.32	*ARL17A, ARL17B, LRRC37A, LRRC37A2, NSF*
chr17:69,753,244-69,901,172	CN Gain	147929	5	8	q25.1	*BTBD17, DNAI2, GPR142, KIF19, TTYH2*
chr17:77,286,848-77,401,037	CN Gain	114190	3	7	q25.3	*DYSFIP1, P4HB, SLC25A10*
chr18:0-117,644	CN Gain	117645	0	4	p11.32	
chr19:48,029,654-48,470,723	CN Gain	441070	9	30	q13.31	*PSG1, PSG10, PSG11, PSG2, PSG4, PSG5, PSG6, PSG7, PSG9*
chr20:25,691,766-25,830,492	CN Gain	138727	1	5	p11.21 - p11.1	*FAM182B*
chr21:13,555,699-13,647,953	CN Gain	92255	0	3	q11.2	
chr21:13,647,953-13,861,001	CN Loss	213049	0	8	q11.2	
chr22:16,941,475-17,316,844	CN Gain	375370	6	19	q11.21	*DGCR6, GGT3P, PEX26, PRODH, TUBA8, USP18*
chr22:18,670,410-19,056,269	CN Gain	385860	3	14	q11.21	*DGCR6L, PI4KAP1, RIMBP3*
chrX:1,376,768-1,592,621	CN Gain	215854	6	26	p22.33	*ASMTL, CSF2RA, IL3RA, NCRNA00105, P2RY8, SLC25A6*

Table S3. 27 - CNVs found for patient 27

Chromosome Region	Event	Length	Genes	Probes	Cytoband	Gene Symbols
chr1:16,886,363-16,943,714	CN Gain	57352	1	4	p36.13	*ESPNP*
chr1:141,494,201-141,652,073	CN Loss	157873	0	5	q12	
chr1:147,255,457-147,482,142	CN Gain	226686	0	11	q21.1	
chr1:167,483,633-167,531,051	Homozygous Copy Loss	47419	1	3	q24.2	*NME7*
chr1:195,008,432-195,078,812	CN Loss	70381	2	7	q31.3	*CFHR1, CFHR3*
chr1:246,781,770-246,863,559	CN Loss	81790	4	6	q44	*OR2T10, OR2T11, OR2T29, OR2T34*
chr2:89,244,042-89,895,566	CN Loss	651525	0	19	p11.2	
chr3:163,987,281-164,108,596	High Copy Gain	121316	0	7	q26.1	
chr4:54,609-111,638	CN Gain	57030	2	5	p16.3	*ZNF595, ZNF718*
chr4:34,448,772-34,500,505	CN Loss	51734	0	4	p15.1	
chr4:69,063,649-69,185,257	CN Gain	121609	1	10	q13.2	*UGT2B17*
chr4:107,322,522-107,395,454	CN Loss	72933	1	4	q24	*TBCK*
chr5:742,057-894,896	CN Gain	152840	2	15	p15.33	*TPPP, ZDHHC11*
chr5:68,960,939-70,312,515	CN Loss	1351577	12	26	q13.2	*GTF2H2B, GTF2H2C, GTF2H2D, GUSBP3, LOC100170939, LOC100272216, LOC653391, NAIP, SERF1A, SERF1B, SMN1, SMN2*
chr5:70,312,515-70,375,604	CN Gain	63090	5	4	q13.2	*GTF2H2, GTF2H2B, GTF2H2C, GTF2H2D, NAIP*
chr5:70,375,604-70,696,119	CN Loss	320516	7	8	q13.2	*GTF2H2, GTF2H2B, GTF2H2C, GTF2H2D, LOC647859, NAIP, OCLN*
chr5:97,834,591-97,950,837	CN Loss	116247	0	7	q21.1	
chr6:32,563,029-32,600,341	High Copy Gain	37313	1	3	p21.32	*HLA-DRB5*
chr6:32,600,341-32,711,050	CN Gain	110710	3	11	p21.32	*HLA-DRB1, HLA-DRB5, HLA-DRB6*
chr7:16,449,856-16,625,033	CN Loss	175178	2	11	p21.1	*ANKMY2, SOSTDC1*
chr7:61,060,605-61,374,646	CN Loss	314042	0	4	q11.1 - q11.21	
chr8:39,349,059-39,508,365	High Copy Gain	159307	2	10	p11.23 - p11.22	*ADAM3A, ADAM5P*
chr9:45,086,142-45,569,831	CN Loss	483690	0	9	p11.2	
chr14:18,149,473-19,490,518	CN Loss	1341046	9	40	q11.1 - q11.2	*OR11H12, OR4K1, OR4K2, OR4K5, OR4M1, OR4N2, OR4Q3, P704P, POTEG*
chr15:19,583,601-20,258,362	CN Loss	674762	5	22	q11.2	*GOLGA8DP, LOC727924, OR4M2, OR4N3P, OR4N4*
chr16:2,583,264-2,680,019	CN Gain	96756	4	8	p13.3	*FLJ42627, KCTD5, LOC652276, PDPK1*
chr16:32,096,628-32,650,001	CN Loss	553374	2	18	p11.2	*TP53TG3, TP53TG3B*

Table S3. 28 - CNVs found for patient 28

Chromosome Region	Event	Length	Genes	Probes	Cytoband	Gene Symbols
chr1:17,077,550-17,203,213	CN Gain	125664	3	8	p36.13	*ATP13A2, CROCC, MFAP2*
chr1:194,993,686-195,051,521	CN Gain	57836	1	5	q31.3	*CFHR3*
chr4:34,429,336-34,500,505	CN Loss	71170	0	5	p15.1	
chr4:63,644,174-63,790,657	CN Gain	146484	0	8	q13.1	
chr4:69,038,815-69,170,341	CN Loss	131527	2	10	q13.2	*TMPRSS11E, UGT2B17*
chr5:69,661,014-70,452,813	CN Loss	791800	12	12	q13.2	*GTF2H2, GTF2H2B, GTF2H2C, GTF2H2D, LOC647859, LOC653391, NAIP, OCLN, SERF1A, SERF1B, SMN1, SMN2*
chr6:256,093-327,656	CN Loss	71564	1	5	p25.3	*DUSP22*
chr6:32,563,029-32,634,334	CN Loss	71306	2	7	p21.32	*HLA-DRB5, HLA-DRB6*
chr8:12,270,256-12,580,432	CN Gain	310177	3	8	p23.1	*DEFB109P1, FAM66A, FAM86B2*
chr8:39,349,059-39,508,365	Homozygous Copy Loss	159307	2	10	p11.23 - p11.22	*ADAM3A, ADAM5P*
chr10:135,088,908-135,230,109	CN Gain	141202	3	12	q26.3	*CYP2E1, LOC619207, SYCE1*
chr11:55,119,157-55,178,810	Homozygous Copy Loss	59654	3	3	q11	*OR4C11, OR4P4, OR4S2*
chr13:51,413,362-51,515,663	CN Loss	102302	3	8	q14.3	*ALG11, ATP7B, UTP14C*
chr15:19,699,696-20,235,151	CN Loss	535456	4	20	q11.2	*LOC727924, OR4M2, OR4N3P, OR4N4*
chr15:30,378,732-30,674,063	CN Gain	295332	2	7	q13.3	*FAM7A1, FAM7A2*
chr17:41,768,841-42,128,214	CN Gain	359374	5	7	q21.31 - q21.32	*ARL17A, ARL17B, LRRC37A, LRRC37A2, NSF*
chr19:59,929,792-60,074,245	CN Gain	144454	8	6	q13.42	*KIR2DL1, KIR2DL3, KIR2DL4, KIR2DS4, KIR3DL1, KIR3DL2, KIR3DL3, KIR3DP1*
chr21:9,896,630-10,198,129	CN Gain	301500	6	14	p11.2 - p11.1	*BAGE, BAGE2, BAGE3, BAGE4, BAGE5, TPTE*
chr22:22,676,684-22,729,016	CN Loss	52333	4	4	q11.23	*GSTT1, GSTTP1, GSTTP2, LOC391322*

Table S3. 29 - CNVs found for patient 29

Chromosome Region	Event	Length	Genes	Probes	Cytoband	Gene Symbols
chr1:0-688,026	CN Loss	688027	13	8	p36.33	FAM138A, FAM138C, FAM138F, LOC100132062, LOC100132287, LOC100133331, LOC100288778, MIR1977, OR4F16, OR4F29, OR4F3, OR4F5, WASH5P, hsa-mir-1977, hsa-mir-1302-2
chr1:72,532,314-72,590,954	Homozygous Copy Loss	58641	0	3	p31.1	
chr3:163,987,281-164,108,596	High Copy Gain	121316	0	7	q26.1	
chr4:0-54,609	CN Loss	54610	2	4	p16.3	ZNF595, ZNF718
chr4:69,038,815-69,170,341	CN Gain	131527	2	10	q13.2	TMPRSS11E, UGT2B17
chr4:70,766,959-71,285,452	CN Gain	518494	14	35	q13.3	C4orf35, C4orf40, C4orf7, CSN1S1, CSN1S2A, CSN1S2B, CSN2, CSN3, HTN1, HTN3, ODAM, SMR3A, SMR3B, STATH
chr4:76,987,928-77,215,148	CN Gain	227221	7	18	q21.1	ART3, CXCL10, CXCL11, CXCL9, NAAA, PPEF2, SDAD1
chr5:742,057-866,296	CN Loss	124240	2	12	p15.33	TPPP, ZDHHC11
chr5:68,877,031-70,696,119	CN Loss	1819089	15	42	q13.2	GTF2H2, GTF2H2B, GTF2H2C, GTF2H2D, GUSBP3, LOC100170939, LOC100272216, LOC647859, LOC653391, NAIP, OCLN, SERF1A, SERF1B, SMN1, SMN2
chr6:193,938-327,656	CN Loss	133719	1	9	p25.3	DUSP22
chr6:29,951,068-30,013,265	CN Loss	62198	3	4	p21.33	HCG2P7, HCG4P6, HLA-H
chr6:30,013,265-30,072,018	CN Gain	58754	2	4	p21.33	HCG9, HLA-A
chr6:32,563,029-32,634,334	CN Loss	71306	2	7	p21.32	HLA-DRB5, HLA-DRB6
chr7:38,262,476-38,336,127	CN Loss	73652	1	6	p14.1	TARP
chr7:61,060,605-61,197,061	CN Loss	136457	0	2	q11.1 - q11.21	
chr8:7,121,948-7,942,536	CN Gain	820589	28	27	p23.1	DEFB103A, DEFB103B, DEFB104A, DEFB104B, DEFB105A, DEFB105B, DEFB106A, DEFB106B, DEFB107A, DEFB107B, DEFB109P1B, DEFB4, FAM66B, FAM66E, FAM90A10, FAM90A13, FAM90A14, FAM90A18, FAM90A19, FAM90A20, FAM90A5, FAM90A7, FAM90A8, FAM90A9, LOC349196, SPAG11A, SPAG11B, ZNF705G
chr8:39,332,477-39,526,316	CN Loss	193840	2	12	p11.23 - p11.22	ADAM3A, ADAM5P
chr8:146,240,172-146,274,826	CN Loss	34655	1	3	q24.3	C8orf33
chr12:9,519,693-9,591,347	Homozygous Copy Loss	71655	0	5	p13.31	
chr14:19,244,829-19,490,518	CN Loss	245690	6	13	q11.2	OR4K1, OR4K2, OR4K5, OR4M1, OR4N2, OR4Q3
chr14:21,678,281-22,042,277	CN Loss	363997	0	23	q11.2	
chr14:106,079,453-106,137,200	CN Loss	57748	0	5	q32.33	
chr15:18,775,860-19,480,163	CN Gain	704304	7	27	q11.2	BCL8, CXADRP2, GOLGA6L6, GOLGA8C, LOC646214, NF1P1, POTEB
chr16:31,840,621-33,691,406	CN Loss	1850786	4	64	p11.2	HERC2P4, SLC6A10P, TP53TG3, TP53TG3B
chr17:31,459,116-31,511,535	CN Loss	52420	0	5	q12	
chr17:41,555,382-41,768,841	CN Loss	213460	3	12	q21.31	ARL17B, KIAA1267, LRRC37A
chr17:41,768,841-42,128,214	CN Gain	359374	5	7	q21.31 - q21.32	ARL17A, ARL17B, LRRC37A, LRRC37A2, NSF
chr17:78,638,125-78,774,742	CN Loss	136618	1	3	q25.3	METRNL
chr20:62,379,348-62,435,964	CN Loss	56617	0	5	q13.33	
chr21:14,190,683-14,248,706	CN Loss	58024	1	4	q11.2	C21orf81
chr21:46,908,131-46,944,323	CN Loss	36193	1	3	q22.3	PRMT2
chr22:22,676,684-22,729,016	CN Gain	52333	4	4	q11.23	GSTT1, GSTTP1, GSTTP2, LOC391322

chrX:7,773,321-8,217,918	CN Gain	444598	3	27	p22.31	MIR651, PNPLA4, VCX2, hsa-mir-651
chrX:154,043,595-154,085,916	CN Loss	42322	0	4	q28	

Table S3. 30 - CNVs found for patient 30

Chromosome Region	Event	Length	Genes	Probes	Cytoband	Gene Symbols
chr1:103,905,063-103,965,250	CN Gain	60188	3	5	p21.1	AMY2A, AMY2B, LOC648740
chr1:194,993,686-195,078,812	CN Gain	85127	2	8	q31.3	CFHR1, CFHR3
chr3:198,840,959-198,876,555	CN Gain	35597	0	3	q29	
chr4:69,038,815-69,170,341	CN Loss	131527	2	10	q13.2	TMPRSS11E, UGT2B17
chr5:290,282-334,827	CN Gain	44546	2	4	p15.33	PDCD6, SDHA
chr5:69,661,014-70,245,600	CN Gain	584587	6	5	q13.2	GTF2H2B, GTF2H2C, GTF2H2D, LOC653391, SERF1A, SERF1B
chr6:32,576,756-32,711,050	CN Loss	134295	3	13	p21.32	HLA-DRB1, HLA-DRB5, HLA-DRB6
chr7:143,509,786-143,693,789	CN Gain	184004	8	8	q35	ARHGEF5, ARHGEF5L, CTAGE4, OR2A1, OR2A20P, OR2A42, OR2A7, OR2A9P
chr8:12,139,755-12,366,919	CN Gain	227165	4	5	p23.1	DEFB109P1, DEFB130, FAM66A, FAM86B2
chr8:39,349,059-39,508,365	Homozygous Copy Loss	159307	2	10	p11.23 - p11.22	ADAM3A, ADAM5P
chr10:67,871,783-68,019,393	CN Gain	147611	1	8	q21.3	CTNNA3
chr12:9,519,693-9,591,347	Homozygous Copy Loss	71655	0	5	p13.31	
chr14:18,555,363-19,490,518	CN Loss	935156	8	33	q11.1 - q11.2	OR4K1, OR4K2, OR4K5, OR4M1, OR4N2, OR4Q3, P704P, POTEG
chr14:27,289,441-27,363,658	CN Gain	74218	0	5	q12	
chr16:22,377,707-22,615,315	CN Loss	237609	5	10	p12.1	LOC100132247, LOC100190986, LOC641298, LOC653786, NPIPL3
chr16:34,328,350-34,611,145	CN Gain	282796	2	17	p11.2 - p11.1	LOC146481, LOC283914
chr16:72,921,969-73,030,768	CN Gain	108800	2	7	q22.3	CLEC18B, LOC283922
chr17:31,449,032-31,511,535	High Copy Gain	62504	1	6	q12	CCL4
chr17:41,577,131-41,768,841	CN Loss	191711	3	11	q21.31	ARL17B, KIAA1267, LRRC37A
chr17:41,768,841-42,128,214	High Copy Gain	359374	5	7	q21.31 - q21.32	ARL17A, ARL17B, LRRC37A, LRRC37A2, NSF
chr19:6,283,716-6,944,284	CN Gain	660569	22	42	p13.3 - p13.2	ACER1, ALKBH7, C3, CD70, CLPP, CRB3, DENND1C, EMR1, EMR4P, GPR108, GTF2F1, KHSRP, MIR220B, PSPN, SH2D3A, SLC25A23, SLC25A41, TNFSF14, TNFSF9, TRIP10, TUBB4, VAV1
chr19:6,944,284-7,028,066	CN Loss	83783	6	5	p13.2	FLJ25758, MBD3L2, MBD3L3, MBD3L4, MBD3L5, ZNF557
chr19:7,028,066-7,633,437	CN Gain	605372	14	39	p13.2	ARHGEF18, C19orf45, INSR, KIAA1543, LOC100128573, LOC100131801, MCOLN1, PCP2, PEX11G, PNPLA6, STXBP2, XAB2, ZNF358, ZNF557
chr19:8,698,817-8,822,409	CN Gain	123593	4	8	p13.2	MBD3L1, MUC16, OR2Z1, ZNF558
chr21:45,191,792-45,278,798	CN Gain	87007	2	6	q22.3	C21orf70, NCRNA00162
chr22:48,228,521-48,341,402	CN Loss	112882	0	7	q13.33	