University of Helsinki

Identification of novel mutations in *DLG3* and *GRIA3* in X-linked intellectual disability in Finnish families

Pro Gradu

Faculty of Biosciences

Department of Biological and Environmental Sciences

Division of Genetics

Fatma Doagu

2012

HELSINGIN YLIOPISTO
HELSINGFORS UNIVERSITET
UNIVERSITY OF HELSINKI

Tiedekunta – Fakultet – Faculty	Laitos – Institution– Department
Faculty of Biosciences	Department of Biological and Environmental Sciences

Tekijä – Författare – Author
Fatma Doagu

Työn nimi – Arbetets titel – Title
Identification of novel mutations in *DLG3* and *GRIA3* in X-linked intellectual disability in Finnish families

Oppiaine – Läroämne – Subject
Genetics

Työn laji – Arbetets art – Level	Aika – Datum – Month and year	Sivumäärä – Sidoantal – Number of pages
Master's thesis	November-2012	71

Tiivistelmä – Referat – Abstract

Intellectual disability (ID) is a clinically diverse and genetically heterogeneous disorder characterized by central nervous system defects of varying severity resulting in substantial impairment of intellectual and adaptive functioning as expressed in conceptual (IQ<70), social and practical adaptive skills diagnosed before 18 years of age.

The condition is referred to as non-syndromic when ID is the only clinical feature and syndromic when ID is accompanied by specific other features, for example, Down syndrome.

Intellectual disability is one of the largest unsolved problems of health care with a prevalence of 2-3% in the population. There is a 30-40% excess of male versus female patients in ID which refers to over-representation of X chromosomal defects causing ID.

In this study, exome sequencing of the X chromosome was applied in order to identify genes and their mutations in two Finnish families with intellectual disability of unknown cause. The mutations were identified using Agilent Sure select array that covers almost 93% of the coding region of the chromosome. Exome sequencing resulted in 11 variations in total. Segregation of these variants was studied using PCR, ExoSAP-IT purification protocol and BigDye® Terminator v3.1 Cycle Sequencing Kit. Eventually, two novel mutations were identified: one for each family. Both mutations reside in genes that have previously been shown to cause X-linked intellectual disability. Both of the mutations were absent in over 120 control DNA samples.

In one family with three affected males, a novel splice mutation was identified in discs large homolog 3 *(DLG3)*, which encodes synapse-associated protein 102 (SAP102). The mutation is located at the splice site in intron 1 (500+1 G>C) and its effect on protein function needs to be analysed at the RNA-level using cDNA-sequencing. The clinical phenotype of the three affected brothers is mild to moderate intellectual disability.

In the other family with three severely affected male patients, a novel mutation in exon 12 was identified on glutamate receptor, ionotropic, AMPA 3 *(GRIA3)* resulting in amino acid glycine (GGG) changing to arginine (CGG) at codon 630 (G630R). *GRIA3* belongs to AMPA receptors implicated in the regulation of several biological processes.

Our findings elucidate the power of exome sequencing in the diagnosis of rare, genetically heterogeneous disorders like intellectual disability. The results obtained will help in assessing the prognosis of the disease, in estimating the risk of the disorder to other family members, and in facilitating the development of future therapies for these devastating disorders. The results also further confirm the role of *DLG3* and *GRIA3* in human cognitive development.

Avainsanat – Nyckelord – Keywords
X-linked ID, exome sequencing, gene, *DLG3*, *GRIA3*, mutation

Ohjaaja tai ohjaajat – Handledare – Supervisor or supervisors
Irma Järvelä

Säilytyspaikka – Förvaringställe – Where deposited
Viikki Science Library

Muita tietoja – Övriga uppgifter – Additional information

Tiedekunta – Fakultet – Faculty	Laitos – Institution – Department
Biotieteellinen tiedekunta	Perinnöllisyystieteen laitos

Tekijä – Författare – Author

Fatma Doagu

Työn nimi – Arbetets titel – Title

Identification of novel mutations in *DLG3* and *GRIA3* in X-linked intellectual disability in Finnish families

Oppiaine – Läroämne – Subject

Perinnöllisyystiede

Työn laji – Arbetets art – Level	Aika – Datum – Month and year	Sivumäärä – Sidoantal – Number of pages
Pro Gradu	Marraskuu-2012	71

Tiivistelmä – Referat – Abstract

Henkinen kehitysvammaisuus (KV) on kliinisesti monimuotoinen ja geneettisesti heterogeeninen sairaus, jolle on ominaista eriasteiset keskushermoston vauriot aiheuttaen olennaisia henkisiä ja adaptiivisen toiminnan kuten käsitteellisen ilmaisun (ÄÖ<70), sosiaalisten ja arkipäivän toimintojen häiriöitä. Kehitysvammaisuutta diagnosoidaan ennen 18 vuoden ikää. Älyllistä kehitysvammaisuutta luokitellaan ei-syndromiseen, kun älyllisen tason lasku on ainoa kliininen ominaisuus ja syndroomaiseen, kun oireyhtymään liittyy muitakin piirteitä, kuten Downin syndrooma. Henkinen kehitysvammaisuus, jota esiintyy 2-3 % väestöstä, on yksi terveydenhuollon suurimmista ratkaisemattomista ongelmista. Sitä esiintyy 30–40% enemmän miehillä kuin naisilla, mikä viittaa X-kromosomin yliedustukseen näiden häiriöiden aiheuttajana.

Tässä tutkimuksessa on sovellettu X-kromosomin eksomisekvensointia mahdollisesti mutatoituneiden geenien tutkimisessa kahdessa suomalaisessa perheessä, missä esiintyy tuntemattomasta syystä johtuvaa kehitysvammaisuutta. Menetelmänä käytettiin Agilentin Sure select array:tä, joka kattaa n. 93 % X- kromosomin proteiinia koodaavasta alueesta. Tuloksena saatiin 2 muutosta ensimmäisessä perheessä ja 9 muutosta toisessa perheessä. Näiden variaatioiden segregaatiota tutkittiin PCR:n, ExoSAP IT-puhdistusprotokollan ja BigDye ® Terminator v3.1 Cycle-sekvensointikitin avulla. Kahden mutaation todettiin segregoituvan yhdessä kehitysvammaisuuden kanssa: yksi kussakin perheessä. Molemmat löydetyt mutaatiot ovat geeneissä, jotka on aikaisemmin yhdistetty X-kromosomaalisesti periytyvään kehitysvammaisuuteen eli XLID:iin. Kumpaakaan mutaatiota ei löytynyt yli 120 verenluovuttaja kontrollin X-kromosomeista.

Perheessä, missä on kolme lievästä keskivaikeaan kehitysvammaista miestä, todettiin uusi silmukointimutaatio discs large homolog 3 *(DLG3)* geenissä. *DLG3* koodaa guanylaatti perheeseen kuuluvaa SAP102 proteiinia. Mutaatio sijaitsee ensimmäisen eksoni-introni liitoksen kohdalla (500+1 G>C) ja sen vaikutusta proteiinin toimintaan on analysoitava RNA-tasolla cDNA-sekvensoinnin avulla. Toisessa perheessä, missä on kolme vakavasti kehitysvammaista miespotilasta, löytyi mutaatio eksonissa 12 glutamate receptor, ionotropic, AMPA 3 *(GRIA 3)* geenissä aiheuttaen glysiinin (GGG) vaihtumista arginiiniksi (CGG) kodonissa 630 (G630R). *GRIA3* kuuluu AMPA reseptoreihin ja osallistuu usean biologisen prosessin säätelyyn.

Tämä tutkimus osoittaa, että eksomisekvensointi on tehokas työkalu selvitettäessä harvinaisia, geneettisesti heterogeenisiä sairauksia kuten tuntemattomasta syystä johtuvaa kehitysvammaisuutta. Toisaalta tutkimuksen aikana saadut tulokset auttavat arvioimaan taudin ennustetta ja esiintymisriskiä muilla perheenjäsenillä ja kehittämään uusia hoitokeinoja näille tuhoisille häiriöille. Tulokset vahvistavat myös entisestään *DLG3* ja GRIA3 roolia ihmisen kognitiivisessa kehityksessä.

Avainsanat – Nyckelord – Keywords

Kehitysvammaisuus, XLID, *DLG3*, *GRIA3*, eksomisekvensointi, geeni, mutaatio

Ohjaaja tai ohjaajat – Handledare – Supervisor or supervisors

Irma Järvelä

Säilytyspaikka – Förvaringställe – Where deposited Viikin tiedekirjasto

Muita tietoja – Övriga uppgifter – Additional information

CONTENTS

ABBREVIATIONS

AAIDD	American Association on Intellectual and Developmental Disabilities
AD	Autosomal dominant
AMPA	α-amino-3-hydroxy-5-methyl-4-isoxazolpropionate
AR	Androgen-receptor gene
AR	Autosomal recessive
BF	Blocking factor
BLAST	Basic Local Alignment Search Tool
BLAT	The Blast-Like Alignment Tool
CGH	Comparative genomic hybridization
CNP	Copy Number Polymorphism
CNS	Central nervous system
CNV	Copy Number Variation
DLG3	Discs, large homolog 3 (Drosophila)
DMD	Dystrophin
ERCC6L	Excision repair cross-complementing rodent repair deficiency, complementation group 6-like
ESX1	ESX homeobox 1
F8	Coagulation factor VIII, procoagulant component
FDH	Focal Dermal Hypoplasia = Goltz syndrome
FIMM	Finnish Institute of Molecular Medicine
GAII	Genome Analyzer II
GluRs	Glutamate receptors: iGluRs=ionotropic GluRs, mGluRs=metabotropic GluRs
GPR112	G protein-coupled receptor 112
GRIA3	Glutamate receptor, ionotropic, AMPA 3
HADH2	Hydroxysteroid (17-beta) dehydrogenase 10 (HSD17B10)
HUWE1	HECT, UBA and WWE domain containing 1, E3 ubiquitin protein ligase

ICD-10	International Classification of Diseases, 10th Revision
ID	Intellectual disability
IQ	Intelligent quotient
LINE	Long interspersed nuclear elements
LRCH2	Leucine-rich repeats and calponin homology (CH) domain containing 2
MAGUK	Membrane-associated guanylate kinases
MECP2	Methyl CpG binding protein 2 (Rett syndrome)
MR	Mental retardation
MSY	Male specific Y
NGS	Next Generation Sequencing
NMDA	N-methyl-D-aspartate
NS-ARID	Non-syndromic autosomal recessive intellectual disability
NS-XLID	Non-syndromic X-linked intellectual disability
NTD	Neural tube defects
OR13H1	Olfactory receptor, family 13, subfamily H, member 1
PAH	Phenylalanine hydroxylase
PAK3	P21 protein (Cdc42/Rac)-activated kinase 3
PHF8	PHD finger protein 8
PKU	Phenylketonuria
PLXNA3	Plexin A3
PORCN	Porcupine homolog (Drosophila)
PSD	Postsynaptic density
SAP	Synapse-associated protein
SOX3	(Sex determining region Y)-box 3
SRY	Sex determining region Y
S-XLID	Syndromic X-linked intellectual disability
Syn	Synonymous
TAF1	TAF1 RNA polymerase II, TATA box binding protein (TBP)-associated factor
VINELAND	Vineland adaptive behavioural scales

WAIS	Wechsler Adult Intelligence Scales
WHO	World Healthcare Organization
WISC	Wechsler Intelligence Scales for Children
XCI	X chromosome inactivation
X*ic*	X-inactivation center
XIST	X (inactive)-specific transcript (non-protein coding)
XLID	X-linked intellectual disability
XLMR	X-linked mental retardation
ZBTB33	Zinc finger and BTB domain containing 33

1 INTRODUCTION

Intellectual disability (ID) also referred to as mental retardation (MR), is one of the largest unsolved problems of health care with a prevalence of 2-3% in the population. ID is a clinically diverse and genetically heterogeneous disorder characterized by central nervous system defects of varying severity resulting in substantial impairment of intellectual and adaptive functioning. It is defined by intelligent quotient (IQ) below 70 scores and limitations in social and adaptive skills diagnosed before the age of 18 years. ID is usually inherited in Mendelian fashion and most of the cases are caused by unique single nucleotide variations in each family [van Bokhoven 2011]. Clinically, the condition is referred to as non-syndromic when ID is the only clinical feature and syndromic when ID is accompanied by dysmorphic, neurological and/or metabolic abnormalities.

The 30-40% excess in male versus female patients has led to the hypothesis of over-representation of X chromosomal defects causing ID. Thus, to date over 90 genes have been identified to underlie X-linked intellectual disability (XLID) which is responsible for 10-12% of males with ID. Approximately 40 of these genes have been found in non-syndromic XLID.

In the past, studying these disorders was very challenging and time consuming. However, due to enormous and rapid development of methods in molecular genetics, such as sequencing and gene-based diagnostics, identifying genes underlying ID and Mendelian diseases in general has become quite straightforward. For instance, while in 1973 sequencing of small DNA fragments, only 24 base pairs could be achieved, nowadays entire genomes are being sequenced with high (mutation detecting) resolution and rapidly decreasing costs. Thus next generation sequencing has become practically available for every laboratory and research group. At the same time the ‘personalized medicine’ approach is starting to establish its role in solving health problems.

The aim of this thesis was to identify genetic causes behind non-syndromic intellectual disability present in two large Finnish families with three affected male patients in each family, suggesting an X-linked mode of inheritance. In

order to achieve this goal, one of the newest methods in genetics, exome sequencing of the X chromosome, was conducted.

Initially this project was started in 2005. At that time 88 evenly spaced X chromosomal microsatellite markers were genotyped in the Finnish families to construct common haplotypes of the patients [Peippo et al. 2007]. Genotypes were generated using standard fluorescent detection based semi-automatic technology [Kong et al. 2002]. Haplotypes are shown in Table 1-1. The haplotype shared in family D172 was 120 Mb long containing 39 known XLID genes and hundreds of other genes. The area shared in family D174 was 98 Mb including 27 known XLID genes and dozens of other genes. Consequently, the next step would have been sequencing of all genes in the area individually in family members. However, the task would have been analogous to looking for a needle in a haystack, considering the time and resources needed. For this reason, the study was postponed until the year 2010 when improved and faster methods such as exome sequencing became available.

Table 1-1: Common haplotypes of the affected male patients on the X chromosome in two studied Finnish families, D172 and D174. The black area shows the size of haplotypes and the amount of genes associated to XLID located on this area.

Markers	**cM**		
DXS9903	9,7		
DXS7103	18,4		
DXS7108	19,6		
DXS999	32,8		
DXS1229	35,8		
DXS8027	43,4		
DXS9896	46,1		
DXS1214	47,4		
DXS1067	49		
DXS8030	87,3		
DXS1221	88,9		
DXS1209	89,9		
DXS1196	91		
DXS8077	99,6		
DXS6799	101,8		
DXS8110	110,3		
DXS1072	112,6		
DXS8064	119,7		
DXS1001	123,8		
DXS8093	132,9		
DXS8044	134,3		
DXS8013	148,6		
DXS984	149,8		
DXS8106	158,7		
DXS8061	190,3		
DXS1073	193,7		
Family		D172	174
Area size (Mb)		120	98
Number of XLMR genes		39	27

2 REVIEW OF THE LITERATURE

2.1 Definition and classification of intellectual disability

Intellectual disability (ID), also referred to as mental retardation (MR) or as cognitive impairment [Schalock et al. 2007] is a manifestation of genetically and clinically heterogeneous disorders characterized by significant limitations both in intellectual functioning and in adaptive behaviour, which covers social and practical skills in daily life. They are conditions of medical, educational and social importance. The severity of ID is classified as mild, moderate, severe and profound ID. There exist three different aspects in approaching ID. Healthcare personnel seek the aetiology of ID and the possibility of medical intervention or surgery to minimize the effects of the disorder. Educators are more concerned with school success and academic achievement. Finally, the public uses the label "poor adaptive skills" to describe a person with ID, since the majority of adults with mild ID can live independently [McDermott et. al Handbook 2007].

Intellectual disability is defined in numerous ways. Here I briefly present two widely used definitions: the *International Statistical Classification of Diseases and Related Health Problems* 10th Revision (ICD-10) definition, coordinated by World Health Organization (WHO) and the *American Association on Intellectual and Developmental Disabilities* (AAIDD) classifications. The previous definition, also applied in Finland, concentrates on the level and medical aspect of ID, while the latter is a practical tool to determine whether the person is eligible to receive extra healthcare service. ICD-10 classification of ID states intellectual disability as "a condition of incomplete development of the mind, which is especially characterized by impairment of skills manifested during the developmental period, skills which contribute to the overall level of intelligence, i.e. cognitive, language, motor, and social abilities. Disability can occur with or without any other mental or physical condition" [WHO 2010]. AAIDD [Schalock et al. 2010] alternatively defines ID as "a disability characterized by significant limitations both in intellectual functioning and adaptive behaviour as expressed in conceptual, social and practical adaptive skills". However, both of these classifications, along with all other categorizations, state the onset for intellectual

disability before the age of 18 years. They also divide intellectual disability into four categories based on IQ (Table 2-1): mild, moderate, severe and profound ID. ICD-10 is stricter than AAIDD.

Table 2-1: Classification of ID based on ICD-10 classification.

ID	Severity level	IQ-range	Description
F70	Mild ID	50-69	- acquire language with some delay - some learning difficulties in school - adults able to work and maintain social relationships
F71	Moderate ID	35-49	- developing comprehension and language use is slow - most learn to develop some degree of independence in self-care - mental age of adults 6-9 years
F72	Severe ID	20-34	- limitations in self-care and motor skills - continuous need of support - mental age of adults 3-6 years
F73	Profound ID	<20	- severe limitations in self-care, continence, communication and mobility - mental age of adults below 3 years

Both of the mentioned classifications apply the following standardized IQ tests in diagnosing ID: the Wechsler Adult Intelligence Scales (WAIS) and the Wechsler Intelligence Scales for Children (WISC) [Ropers 2010]. In addition, clinicians and health care personnel conduct other psychological tests for diagnosing ID. For example in Finland, Vineland adaptive behavioural scales, second edition (VINELAND II)[Sparrow et al. 2005] is applied to measure the personal and social skills needed for everyday living and the Bender Gestalt Test [Koppitz 1975] is applied to evaluate visual-motor functioning and visual perception skills.

In clinical genetics, intellectual disability is defined as either syndromic or non-syndromic based on the presence or absence of additional features such as

dysmorphic, neurological, and/or metabolic abnormality. In syndromic intellectual disability the phenotype is accompanied by dysmorphic, neurological, radiological, biological and/or metabolic features. In non-syndromic ID, cognitive impairment is the only manifestation of the disease [Chelly & Mandel 2001; Tarpey et al. 2004]. Based on recent molecular studies, the phenotype-genotype boundaries between the two aforementioned forms of ID are fading.

In this thesis, I will mainly focus on non-syndromic mode of ID (NSID).

2.2 Prevalence of intellectual disability

Prevalence of intellectual disability deviates substantially based on aetiology, behavioural characteristics and degree of ability. However, it is widely considered that the prevalence of mild to moderate ID (IQ<70) is 2-3% and moderate to severe ID (IQ<50) 0.3-0.5% of the population in developed countries [Raymond 2006; McDermott et al. 2007; Ropers 2010]. Therefore, there are 75 000-100 000 people with mild ID and 15 000-25 000 people with severe ID living in Finland. In developing countries prevalence of ID tends to be even higher, resulting from aetiological factors such as malnutrition, parental consanguinity, cultural deprivation and poor healthcare [McDermott et al. 2007; Ropers 2010].

Epidemiological studies have repeatedly shown a 30-50% excess of ID in males over females. This has led to the hypothesis that over-representation of male ID may be due to X-linked genes. Nevertheless Chelly and Mandel (2001) reached the conclusion that less than 10% of male ID can be explained by XLID.

2.3 The X chromosome

The X chromosome is biologically and genetically unique due to the evolution of sex chromosomes. It differs widely from autosomes in the rates of gene divergence, recombination, gene movement between chromosomes and the pattern of gene expression. This chromosome also holds a special place in the history of medical genetics for two reasons. Firstly a large number of disorders have been associated with the X chromosome since a recessive mutation for a gene lacking a homologue on the Y chromosome directly revealed a phenotypic consequence in males. In fact, although X chromosome contains only 4% of all human genes, almost 10% of diseases having a mendelian pattern of inheritance

is assigned to the X chromosome. At the same time recognizing an X-linked mode of inheritance seems relatively easy when verifying a disease in a pedigree [OMIM; Ross et al. 2005; Vicoso & Charlesworth 2006].

2.3.1 Evolution of the human X chromosome

At present, it is widely believed that the sex chromosomes, in humans the X and Y chromosomes, have evolved from an ancient, homologous pair of freely recombining autosomes within the last 300 million years, arising independently several times over the course of evolution. One hypothesis suggests that the first step in differentiation of these chromosomes from autosomes took place by occurrence of a male sex determination mutation, changing the X-linked sex determining region Y-box 3 (SOX3) into its homologue, sex determining region on the Y chromosome (SRY). Additionally, the X-Y recombination has been restricted for an elusive reason, by at least four different events leading to the accumulation of mutations and further resulting in the degeneration of the Y chromosome [Lahn et al. 2001; Singh & Petrov 2007].

Eventually the evolutionary events have brought about the sex chromosomes to differentiate morphologically and genetically. The X chromosome has maintained the original, functional elements and remains conserved between species in large regions, including its long arm. Meanwhile, the Y chromosome has lost almost all traces of the original ancestral feature, including the genes it once shared with the X chromosome. In fact, the Y chromosome contains a functional homologue for only 54 genes annotated on the X chromosome. On the other hand, approximately 15 protein-coding genes located on the Y chromosome lack homologues on the X chromosome. In addition, the euchromatic part of the X chromosome is six times longer than that of the Y chromosome, which makes Y considerably smaller than the X chromosome. The X and Y chromosomes provide the only mechanism of sex determination in humans [Charlesworth 1991; Ross et al. 2005; Singh & Petrov 2007].

2.3.2 The X chromosome sequence

The mammalian (especially human) X chromosome differs biologically from other chromosomes in the rate of gene divergence and traffic, recombination, gene movement between chromosomes and the pattern of gene expression. The human X chromosome is over 155 Mb long, representing approximately 5% of the haploid genome and containing nearly 1900 genes [VEGA]. While the male specific Y (MSY) chromosome is 65 Mb, representing approximately 2% of the haploid genome [Waters & Robinson 2008] and contains barely 460 genes [VEGA]. The X chromosome contains short pseudo-autosomal regions at both telomeric ends. These recombine with cognate regions on the Y chromosome. On the other hand it has some features which are unique in the human genome. For instance females inherit paternal and maternal X chromosomes while males inherit only the maternal one. Further, gene expression is randomly silenced on one of the female X chromosomes in early development. The inactivated chromosome remains inactive in somatic tissues thereafter but it is reactivated in the female germ line [Ross et al. 2005; Tarpey et al. 2009].

Genes of the X chromosome contain over 820 protein-coding genes, nearly 770 pseudo genes and 265 processed transcripts [VEGA]. The density and length of the genes on this chromosome (when excluding pseudo genes) is among the lowest compared to other chromosomes, the mean gene length being 49 kilo-bases. However, dystrophin (*DMD*), the largest gene known in the human genome, is located on Xp21.1 [Ross et al. 2005]. Nearly 153 000 single-nucleotide polymorphisms (SNPs) have been identified to date. The X chromosome differentiates from other chromosomes in the amount of repetitive sequences. While the genome contains 45% of interspersed repeats in average, the euchromatic sequence contains 56%. It has remarkable enrichment (29% of the sequence) for long interspersed nuclear elements (LINEs) of the L1 family while the average on the genome being only 17%.

The human X chromosome is enriched for genes responsible for brain functioning, and intelligence and also contains a high number of genes expressed in the brain, explaining the large number of X-linked intellectual disability (XLID) syndromes. Other gene groups over-represented on this chromosome are involved in early spermatogenesis, sex and reproduction while it lacks genes

involved in late spermatogenesis [Vicoso & Charlesworth 2006; Delbridge et al. 2008].

2.3.3 X chromosome inactivation

In mammalian females' somatic cells, the major part of one X chromosome is inactivated randomly; implicating the paternally-derived X is silenced in some cells and the maternally-derived X in the others. Thus, females are considered to be mosaic for the genes on the X chromosome. The purpose of X chromosome inactivation (XCI) is to ensure that males and females have equal expression of X-linked genes, since males usually inherit one X chromosome and females two. However, approximately 15% of the genes escape X-inactivation, resulting in them being expressed from both, the active and inactive X chromosomes. Many of the escaping genes lack the homologue on the Y chromosome. The silencing process is an epigenetic change occurring in the early stage of embryonic development and is passed on to all daughter cells, making it tissue specific. However, this inactivation is not transmitted to offspring [Racchi et al. 1998; Carrel & Willard 2005].

The multifunctional domain called the X-inactivation center (X*ic*) is the key element to the silencing process. The X*ic,* located on the long arm of the X chromosome, has evolved from an ancestral vertebrate gene cluster in placental mammals undergoing several rearrangements during the course of evolution. It contains the X (inactive)-specific transcript (non-protein coding) *(XIST)* gene which is very crucial inactivation-wise. *XIST* is different from other genes since it does not code a protein, but a more than 17 kb un-translated RNA and it is expressed only from the inactive X chromosome after the inactivation process. In each embryonic cell, the silencing event is initiated by *XIST* starting to produce RNA which eventually resides on the X chromosome, spreading the inactivation signal up and down the chromosome. At the same time many chromatin modifications such as DNA methylation and histone modifications are required to maintain the stability of the inactive state of the chromosome [Chow et al. 2005; Chang et al. 2006; Wutz & Gribnau 2007].

In addition to activating the silencing process, X*ic* ensures that just one X chromosome remains active per diploid genome. The hypothesis is that an

autosomal blocking factor (BF) assists it in this task, by protecting one X chromosome per diploid genome from silencing. The mechanism and the nature of the BF are still unknown. On the other hand, the question of how the random X-inactivation is actually established or if it is entirely random remains also unanswered. However, recent progress in the field suggests the involvement of multiple regulatory systems including epigenetic modifications [Wutz & Gribnau 2007; Waters & Robinson 2008].

2.4 Aetiology of intellectual disability

Aetiologically ID is a very heterogeneous disorder involving environmental, socio-demographic and genetic factors. These factors, which will be discussed next, can result alone or together in mild to severe intellectual disability. Environmental factors cause mostly mild ID while genetic factors play a larger role in causing severe ID. In fact genetic causes account for 20% of mild ID and 65% of moderate to severe ID [Chiurazzi et al. 2008; van Bokhoven 2011].

2.4.1 Environmental factors

McDermott et al. (2007) describe three main nutritional factors causing ID to some extent in their handbook of intellectual and developmental disabilities. Nutritional deficit of folic acid causes folate-sensitive neural tube defects (NTD), which contribute to a small proportion of ID [OMIM 601634]. Phenylketonuria (PKU) [OMIM 261600], occurring in Caucasian populations, is a rare defect of amino acid metabolism caused by mutations in the phenylalanine hydroxylase (PAH) gene. Although PKU is a genetic defect it can be treated by proper diet and thus ID resulting it can be prevented. PKU does not appear in Finland. A third nutritional cause of intellectual disability is iodine deficiency during pregnancy causing maternal and fetal hypothyroidism, referred to as cretinism [Delong 1993]. The author's stress that ID cases caused by the nutritional factors mentioned can be prevented by strict diet during pregnancy, infancy and early childhood [McDermott et al. Handbook 2007].

Environmental causes behind intellectual disability also include infection during pregnancy, prematurity and low birth weight, birth trauma and asphyxia, fetal stroke, a number of environmental chemicals and compounds such as alcohol exposure, smoking and lead exposure during pregnancy and infancy, postnatal

infections, traumatic brain injury during childhood and deprivation in childhood and parental consanguinity [Drews et al. 1995; Najmabadi et al. 2011; van Bokhoven 2011]. Parental consanguinity increases the genetic risk of ID. However, it may also recon among environmental factor to some extent since parental consanguinity occurs just in some cultures.

2.4.2 Genetic causes of intellectual disability

Genetically, intellectual disability is a very complex disorder. Causative factors include genomic disorders, chromosome aneuploidies, structural chromosomal abnormalities, gene defects and single nucleotide changes [Chiurazzi et al. 2008; van Bokhoven 2011]. The vast variety of causative factors is the reason why, even today, so little is known about ID and why over 40% of cases remain without etiological diagnoses [Knight et al. 1999; Ropers 2010]. However, recently rapid progress has taken place in the field, since large scale gene finding is becoming more of a reality.

Up to now, defects in 450 genes have been associated to ID and related cognitive disorders (CDs), such as autism. Of the identified genes 50 express non syndromic ID, while the rest manifest ID combined with other clinical features. Described genes are involved in synaptic plasticity, Ras and Rho GTPase signalling and some of them modify chromatin structure epigenetically. And more are coming. After all, it is estimated that the true number of ID genes might rise to 1500-2000 [van Bokhoven 2011].

Of all of the single genes underlying intellectual disability, mutation of the gene underlying fragile X syndrome is considered to be the most frequent cause of ID and autism. It is also the first gene proven to cause the classical ID phenotype. Fragile X syndrome, the prevalence being 1 in 5000-6000 males, occurs by silencing of the *FMR1* gene, resulting from expansion of the CGG repeat at the 5' end of the gene on the X chromosome [Verkerk et al. 1991].

2.4.2.1 Chromosomal aberrations behind ID

Aberrations in chromosomal number and structure (Table 2-2) are the most frequent cause of intellectual disability accounting for up to 40% of severe and 10-20% of mild ID cases [Göstason et al. 1991]. For instance, Down's syndrome, caused by trisomy 21, is the ultimate reason for ID. Trisomies 18 and 13 are less frequent but result in a more severe phenotype. Other chromosomal abnormalities consist of X chromosomal aneuploidies and a range of cytogenetically balanced and unbalanced translocations. On the other hand, large chromosomal microdeletions, which are detectable by conventional karyotyping or fluorescence in situ hybridization (FISH), are an important cause of ID. They include a number of conditions with recognizable clinical features such as Prader-Willi and Angelman, Williams-Beuren, Smith-Magenis, Miller-Dieker and DiGeorge syndromes [Chelly et al. 2006; Ropers 2010; van Bokhoven 2011].

2.4.2.2 Subtelomeric rearrangements

The human genome contains gene-rich subtelomeric regions with high-level sequence polymorphism. Previously, it was suggested that imbalances in these areas would explain a substantial number of unexplained cases of intellectual disability. In 1995, after studying 99 persons with ID using FISH, Flint et al. estimated that 6% of idiopathic intellectual disability is caused by cryptic subtelomeric rearrangements. However, in 2006 three clinical cytogenetic laboratories conducted subtelomere FISH analysis for 11,688 individuals with developmental disabilities. As the result, they concluded that only 2.5% of ID is due to subtelomeric rearrangements [Ravnan et al. 2006]. A year later Ballif et al. reported a remarkably similar result using targeted comparative genomic hybridization (CGH) array. In the study of almost 7,000 cases with ID, authors observed a rate of 2.4% of subjects with clinically significant imbalances in subtelomeric regions. These two recent studies have established the yield of ID, caused by subtelomeric rearrangements, to be 2.5% of patients with ID after routine karyotype analysis.

Table 2-2: List of some of the most frequent chromosomal aberrations causing intellectual disability. * Average IQ is 10-15 points lower than boys with normal karyotype. Prevalence data obtained from Genetics Home Reference and OMIM. Entries made on 20.08.2012.

Chromosomal abnormality	Syndrome	Intellectual disability	Prevalence in newborns	Reference
Trisomy 21	Down's syndrome	mild	1/750-1/1000	Lejeune et al. 1959
Trisomy 18	Edwards syndrome	severe	1/5000	Crider et al. 2008
Trisomy 13	Patau syndrome	severe	1/16000	Crider et al. 2008
45X	Turner's syndrome	increased risk of mild ID	1/2500 females	Bondy CA 2007
47XXX	Triple X syndrome	increased risk of mild ID	1/1000 females	Tartaglia et al. 2010
47XXY	Klinefelter syndrome	mild *	1/500-1000 males	Abramsky & Chapple 1997
47XYY	XYY-syndrome	mild *	1/1000 males	Abramsky & Chapple 1997
5p deletion	Cri du chat	severe	1/20000-50000	Gersh et al. 1995
Paternal 15q11.2 deletion	Prader-Willi syndrome	mild to moderate	1/(10000-30000)	Robinson et al. 1991
Maternal 15q11-q13 deletion	Angelman's syndrome	severe	1/(12000-20000)	Van Buggenhout & Fryns. 2009
15q13.3 microdeletion	15q13.3 microdeletion syndrome	mild to moderate	1/40000	Sharp et al. 2008
17p11.2 deletion	Smith-Magenis	mild to moderate	1/25000	Gropman et al. 2006
7q11.23 deletion	Williams-Beuren syndrome	mild to moderate	1/(7500-20000)	Carrasco et al. 2005
17q21.31 microdeletion	17q21.31 microdeletion syndrome	moderate to severe	1/16000	Koolen et al. 2006

2.4.2.3 Copy number variation (CNV)

In recent years, copy number variations (CNVs) or copy number polymorphisms (CNPs) have taken their place in the large range of diversity forms described in the human genome which includes: chromosome anomalies, insertion-deletion polymorphisms, variable number of repetitive sequence and single nucleotide changes. CNVs are 1 kb-3 Mb deletions or duplications of DNA fragments which are widely distributed throughout the genome. Currently, it is estimated that up to 12% of a phenotypically healthy individuals genome is covered by copy number variants. CNVs can occur both somatically and meiotically, thus having appreciable impact on phenotype, gene expression and natural selection. It is widely hypothesized that CNVs have an important role in the evolution of the human genome and the diversity within the human population [Iafrate et al. 2004; Sebat et al. 2004; Redon et al. 2006].

Despite the fact that most CNVs are benign, new studies have increasingly associated CNVs with diseases like cancer, mental illness, developmental disorders, childhood obesity [Frank et al. 2007; Glessner et al. 2010; Ikeda et al. 2010; Moreno-De-Luca et al. 2010]. This has led to a dilemma in clinical cytogenetics, since science is still unable to entirely distinguish pathogenic CNVs from benign ones [Lee et al. 2007]. However progress has been made and some strategies have been found for interpreting the pathogenic status of CNVs. For instance, a CNV is probably pathogenic, if it is a deletion, expansion or alteration of a CNV inherited from a parent, or it is identical with the CNV of an affected parent. Additionally, de novo CN-deletions over 400 kb in size are always genetically important [Rodriguez-Revenga et al. 2007; Miller et al. 2010].

Change in chromosomal structure can generate a copy number change, leading to a junction between two formerly separated DNA sequences [Hastings et al. 2009]. Occurrence of the junction shows the way structural change has taken place. This usually happens by one of two mechanisms. The recurrent end-points such as non-allelic homologous recombination (NAHR), means that CNVs arise by homologous recombination between repeated sequences. On the other hand, the non-recurrent end-points such as microhomology-mediated events mean that the CNVs occur at sites of limited homology of 2-15 base pairs [Hastings et al. 2009].

Traditional cytogenetic methods such as fluorescence in situ hybridization (FISH) and G-banded karyotyping have been used in clinical genetic testing for decades. However, they detect CNVs very poorly. In recent years advances have been made by conducting high-resolution microarrays on platforms targeting the entire human genome. For instance, chromosomal microarray (CMA) performs a similar function to G-banded karyotyping but at a more sensitive and higher resolution as well as having the ability to detect significant CNVs. However, currently no technology can capture all the variation present in the genome by itself [Miller et al. 2010; Vissers et al. 2010].

On the whole, to date at least 19 recurrent CNVs have been identified to cause ID although, some overlap with other cognitive disorders such as autism, epilepsy and schizophrenia. Most of these genomic intervals are due to NAHR [Bauters et al. 2009; Ropers 2010].

The disease-causing CNVs are continuously collected in databases such as ECARUCA (http://www.ecaruca.net) and DECIPHER (http://decipher.sanger.ac.uk/) in order to help clinicians, healthcare personnel and scientists.

2.4.2.4 Mendelian disorders

Intellectual disability is usually caused by single gene defects inherited in Mendelian fashion. Most commonly, mutations disrupt the protein-coding sequence. It is estimated that major cases of ID are transmitted as an *autosomal recessive* (AR) trait. Alternatively, mutations passed down by an *autosomal dominant* (AD) manner always result in severe ID and are so called *de novo* mutations. Additionally, *X chromosome linked* (XL) genes and CNVs are responsible in 10% of ID cases.

2.4.2.4.1 Autosomal dominant intellectual disability

Autosomal dominant intellectual disability (ADID) is a genetically difficult research topic because persons with the aforementioned disorder are usually severely disabled and do not reproduce. However recent studies have indicated a "*de novo* paradigm" for neurodevelopment and psychiatric diseases including intellectual disability. It is calculated that on average, new-borns have 50-100

new mutations in genome [Lynch 2009]. These spontaneous germ line mutations can have serious consequences when occurring at the transcription level. *de novo* copy number variation (CNV) and point mutations occur in the germline and patients who carry these CNVs have negative family history for the traits they express.

Vissers et al. (2010) found unique non-synonymous *de novo* point mutations in nine genes while sequencing the exomes of ten patients with non-syndromic intellectual disability and their healthy parents. Authors further investigated these mutations and concluded six of them to be pathogenic based on gene function, mutation impact and evolutionary conservation. They concluded that up to 25% of ID cases may be explained by CNVs in the future.

Meanwhile, there exist wide ranges of autosomal dominant disorders associated to ID like neurofibromatosis, tuberous sclerosis and myotonic dystrophy (DM1) occurring in infancy [Ropers 2010].

2.4.2.4.2 Autosomal recessive intellectual disability

Autosomal recessive gene defects are the most important cause behind severe, non-syndromic cognitive disorder [Najmabadi et al. 2011]. However, so far revealing the role of heterogeneous autosomal recessive intellectual disability (ARID) has been difficult due to two facts. Firstly, genetic research has centred in developed countries where parental consanguinity is rare. Secondly, big family sizes are infrequent in Western societies. Nevertheless, in consequence of international research collaboration and fall in costs of next generation sequencing (NGS), large genetic studies have taken place in developed countries. Parental consanguinity and large families are very common in North Africa, Middle East and South Asia. For instance, in Iran, 40% of the families are consanguineous and almost 70% of the population is aged 30 or younger [Ropers 2010; Najmabadi et al. 2011].

Estimates about the eventual number of genes underlying ARID vary from the hundreds to the thousands. Nonetheless due to slow progress in finding these genes in the past, until 2010 just nine genes had been published as being causative to non-syndromic autosomal recessive intellectual disability (NS-

ARID). These genes encode for ionotropic glutamate receptors, signal transductions and transcriptional regulation [Basel-Vanagaite 2007; Ropers 2010]. Thus far, the most remarkable step in this area of human genetics has been taken by a large international collaboration between German and Iranian research groups in 2011. The large-scale study of Najmabadi et al. (2011) resulted in revealing disease-causing variants in 50 novel candidate genes by deep sequencing 136 consanguineous families with ARID. In addition, the authors found mutations in 23 known ID genes. The genes described are involved with neuron- or brain-specific functions, epigenetic, histone structure and modification, housekeeping, cell growth and regulation of translation and transcription.

2.4.2.4.3 X-linked intellectual disability

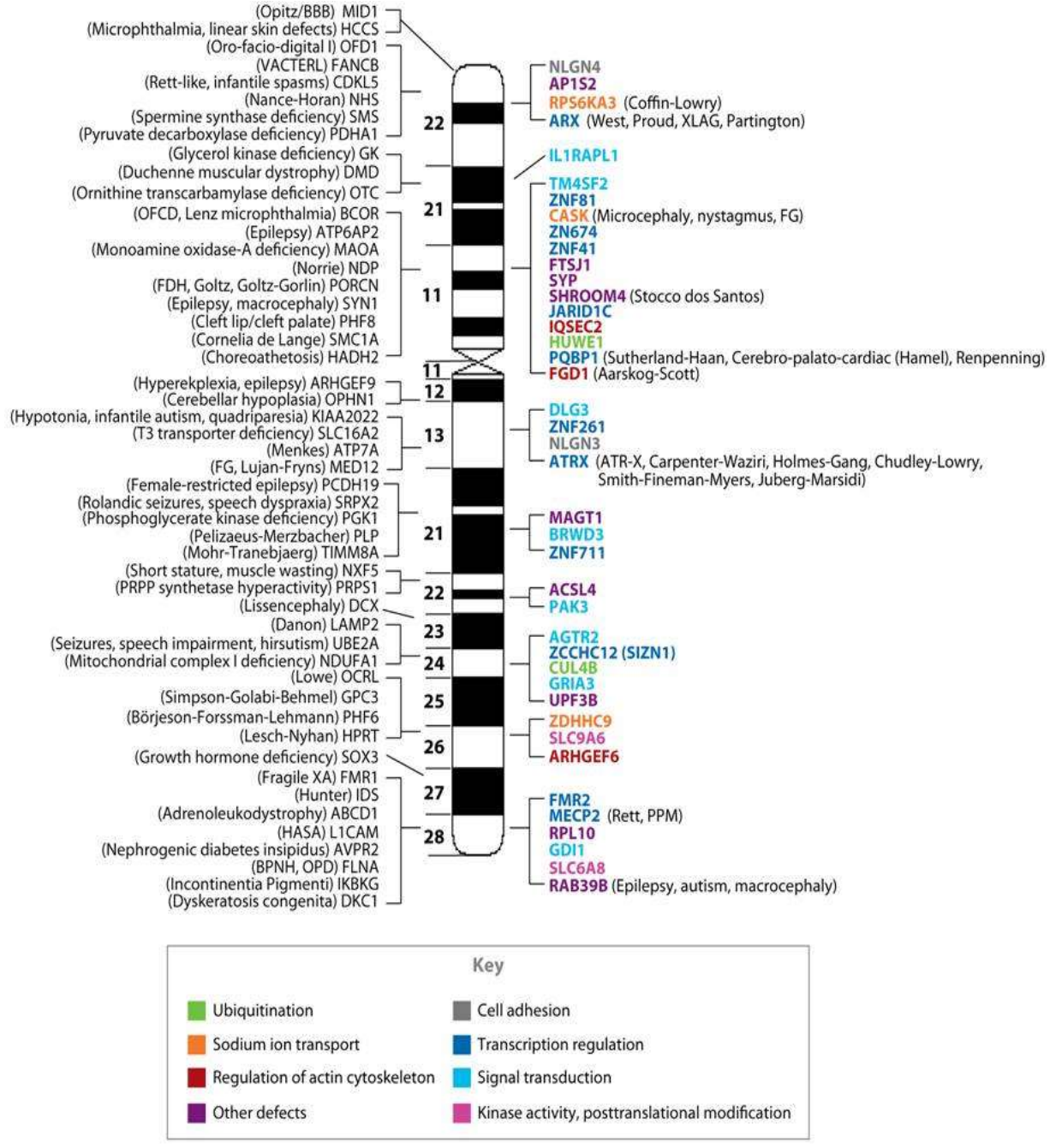
In X-linked intellectual disability (XLID), which is very heterogeneous, the affected gene is maternally inherited. As males inherit only the maternal X chromosome, they are at higher risk to inherit the defect from the carrier mother and usually express a more severe phenotype. This is one reason for the excess of ID in males compare to females. Alas, females can express the defect in a recessive manner due to skewed X-inactivation [Skuse 2005]. Currently, it is widely evaluated that 12-15% of ID will be explained by gene defects and CNV occurring in the X chromosome [Najmabadi et al. 2011; Ropers 2010]. At the same time, these defects are considered to be the most frequent cause of moderate to severe ID in males. So far nearly 100 genes have been published as being a cause of ID when mutated (Figure 2-1). From these genes, 41 underlie non-syndromic XLID (NS-XLID) [Tarpey et al. 2009; van Bokhoven 2011; Vera Kalscheuer personal communication 2012]. However, the line between syndromic XLID (S-XLID) and NS-XLID is sometimes inconstant. For instance, changes in *MECP2* (OMIM 300005) and *ATRX* (OMIM 300032) cause both S-XLID and NS-XLID depending on the mutation [Kaufman et al. 2010; Tarpey et al. 2009].

Meanwhile, regardless of the substantial amount of XLID genes published internationally, thus far only 5 genes have been published as associating with ID in the Finnish population, in addition to fragile X (Table 2-3).

Universally, the identified XLID genes affect general intelligence, social-cognition and emotional regulation, fitting with the proposal that a high density of genes for cognitive ability reside on the X chromosome [Turner 1996]. Biologically these genes influence regulation at all levels of the cell and they are involved in common pathways including RhoGTPase pathways, Reelin-Dab1 pathway, Ras-MAPK pathway and Notch signalling pathway. In general ID genes encode proteins important for early and adult neuronal development and differentiation like synaptic plasticity and synaptic vesicle cycling, transcription regulation, signal transduction, metabolism, cell adhesion, regulation of actin cytoskeleton, chromatin remodelling, membrane modification (Figure 2-1) [Basel-Vangaiate et al. 2007; Chiurazzi et al. 2008; Vaillend et al. 2008]. At the same time, appearing of epilepsy, autistic spectrum disorders and schizophrenia in combination with ID is relatively common. Which explaines many common pathways shared by mentioned defects [Guilmatre et al. 2009; von Bokhoven 2011].

Table 2-3: X-linked genes associated with ID in Finnish families.

Gene/Region	Mutation	Disorder	Reference
HUWE1 and *HADH2* region	Microscopic duplication	Moderate ID	Froyen et al. 2008
MECP2	P127L, R306C, T158M, in total 14 mutations	Rett syndrome	Auranen et al. 2001
PAK3	c.1337G>C	ID	Peippo et al. 2007
PHF8	F279S	XLMR and cleft palate	Koivisto et al. 2007
PORCN	R365Q 150 kb deletion	FDH/Goltz syndrome	Froyen et al. 2009



Ropers HH. 2010.
Annu. Rev. Genomics Hum. Genet. 11:161–87

Figure 2-1: X chromosome showing the genes associated to ID thus far. On the right are the genes underlying NS-XLID. Colour signifies functional category of the gene. Picture is adapted from Ropers (2010).

At least two international consortiums exist to unite clinical and molecular expertise of X-linked intellectual disability worldwide. The European XLMR Consortium (Euro-MRX, http://www.euromrx.com/index.htm), created in 1996, is a collaboration between five European institutes. The main agendas of this consortium are; to identify genes involving XLID, to improve genetic counselling, and to improve DNA diagnostics. Additionally, the consortium aims to simplify the exchange of knowledge and materials between participants and third party researchers. To date, Euro-MRX has collected clinical and molecular data from more than 600 families leading to the identification of 17 novel genes underlying XLID. The Genetics of Learning Disability (GOLD) established in 2001, on the other hand, is a collaboration of 8 large research groups worldwide. The aim of GOLD is to identify new genes and mutations associated with ID and to better understand the mechanism behind them. GOLD has already recruited 800 families throughout Europe, USA and Australia.

2.5 Identification of XLID genes: walking briefly through history

During the past century major developments have taken place in the field of genetics, despite the bumpy and long road. Initially, this resulted in revolutionary inventions which shaped the development of genetics in two different directions, clinically and academically; new approaches have been introduced for identifying causes behind genetic disorders like karyotyping in cytogenetics, the study of the structure, function and evolution of chromosomes. On the other hand, the era of genomes, genotyping, sequencing and developing next generation sequencing (NGS) tools has emerged.

Thomas Hunt Morgan associated chromosomes and the genes residing on them to inheritance back in 1911 [Morgan 1911], yet it took nearly a half century before Watson and Cricks discovery of the molecular structure of DNA and the proving of it to be a double helix, which landed them a Nobel price [Watson & Crick 1953]. Three years later two scientific groups independently established the normal diploid human cell as containing 46 chromosomes [Tjio & Levan 1956; Ford & Hamerton 1956]. A breakthrough which afterwards resulted in the development

of new methods for visualising chromosome structure and organization, leading quickly to the realization that chromosomes differed from each other in length and the position of their centromere. Eventually, starting from 1959, gain or loss in chromosome number was associated with diseases: trisomy 21 was shown to cause Down syndrome [Lejeune et al. 1959], loss of sex chromosome was shown to cause Turner syndrome (45, X) [Ford et al. 1959] and an extra sex chromosome was shown to be the cause behind Klinefelter syndrome (47, XXY) [Jacobs & Strong 1959]. Later, in 1963, Lejeune et al. discovered the first inherited deletion syndrome, Cri du Chat.

On the other hand, Crick et al. (1961) demonstrated that the genetic code for proteins consisted of triplets of nucleotides. Subsequently, this enabled Gilbert and Maxam to sequence a 24 base pair long fragment of DNA as pioneers in 1973. In 1983, Kary Mullis and his colleagues invented polymerase chain reaction (PCR), one of the most revolutionary discoveries in the history of molecular biology and genetics [Saiki et al. 1985]. The list of inventions and methods such as first- and next-generation sequencing tools, linkage analysis and array technology is on-going [Pettersson et al. 2009; Trachtenberg et al. 2012]. Only two decades after the tiny sequencing fragment of 24 bp, Fleischmann et al. (1995) had the ability to sequence the entire genome of an organism; the almost two million bp long *Haemophilus influenzae* genome consisting of a single, circular chromosome. Further on, in the beginning of the new millennium, the first draft of the human genome was published by two independent research groups [Lander et al. 2001; Venter et al. 2001]. To date, the genomes of thousands of organisms have been sequenced and made publicly available for research. Additionally, thousands of genes have been associated to diseases.

The ground-breaking advances described above, and many others not mentioned for lack of time, have enabled new scientific approaches for identifying pathogenic genes, CNV and mutations to evolve almost daily, in addition to the widely varying list of techniques and the costs required. The most frequently used mechanisms in finding causative genes and mutations behind ID include cytogenetic methods such as chromosome banding [Seabright 1971], fluorescence in situ hybridization (FISH) [Langer-safer et al. 1982] and comparative genomic hybridization (CGH) [Kallioniemi et al. 1992; de Vries et al. 2005], genetic

mapping like linkage analysis [Lathrop et al. 1985], genomic arrays [Lugtenberg et al. 2007; McMullan et al. 2009], next generation sequencing [Mardis 2008]. I will now briefly introduce exome sequencing, as it was the capture method used in this study.

2.5.1.1 Exome sequencing

Exome sequencing is a method for targeted sequencing, capturing and analysing the protein-coding region of the genome, the exome. It is based on the construction of a shotgun library from genomic DNA, exon enrichment by hybridization, parallel sequencing and data analysis (Figures 2-2 & 2-3). Ng et al. (2009) developed the aforementioned protocol by sequencing the exomes of 12 unrelated individuals, since seeking rare and novel variants at the genome-wide scale in large cohorts was resource demanding, time consuming and not cost-effective. Thereafter, hundreds of disease causing genes and mutations has been identified by exome sequencing [Hedges et al. 2009; Sun et al. 2010; Pugh et al. 2012]. One might even say that exome sequencing has been an answer to prayers for research of Mendelian disorders including ID [Tarpey et al. 2009; Bamshad et al. 2011; Gilissen et al. 2011; Topper et al. 2011]. Since most causative mutations in monogenic disorders likely disrupt protein-coding sequences, it is also a good method for detecting germline variants resulting in rare or common diseases. Additionally, this method is increasingly establishing its position as a diagnostic tool [Choi 2009]. But again, still some limitations and technical challenges have to be overcome in order to use this method in everyday diagnostics. All of the protein-coding exons in the genome are not known thus, the capture probes available target only known exons. Exome sequencing generates large amounts of sequencing data which have to be managed properly and captured variants interpreted correctly. Furthermore, the sensitivity and specificity of the detection of deletions, duplications and tandem repeats has to be improved [Ku et al. 2012].

Many commercial exome enrichment platforms are available. However three major ones include Agilent's SureSelect Human All Exon 50Mb, Roche/Nimblegen's SeqCap EZ Exome Library v2.0 and Illumina's TruSeq Exome Enrichment. Each platform is scalable to 96-plex robotic automation using biotinylated oligonucleotide baits. Although differing in their target choice,

bait lengths and density, each platform demonstrates a high level of efficiency covering 93-98 % of the exome. Additionally, the aforementioned platforms apply a different molecule for capture; Nimblegen and Illumina apply DNA while Agilent uses RNA [Clark 2011].

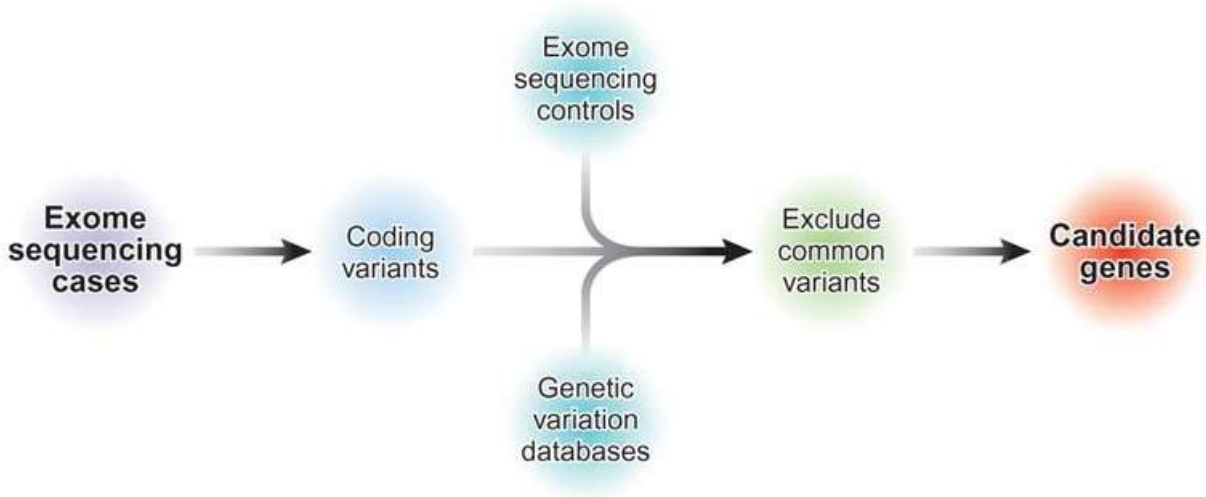


Figure 2-2: Simplified illustration of a genetic study using exome sequencing and filtering strategies when searching for candidate genes for a disorder. Figure adapted from Biesecker 2010.

A schematic in-house pipeline describing the bioinformatics needed for analysing the raw sequencing data resulting from whole genome/exome sequencing is illustrated in Figure 4-1.

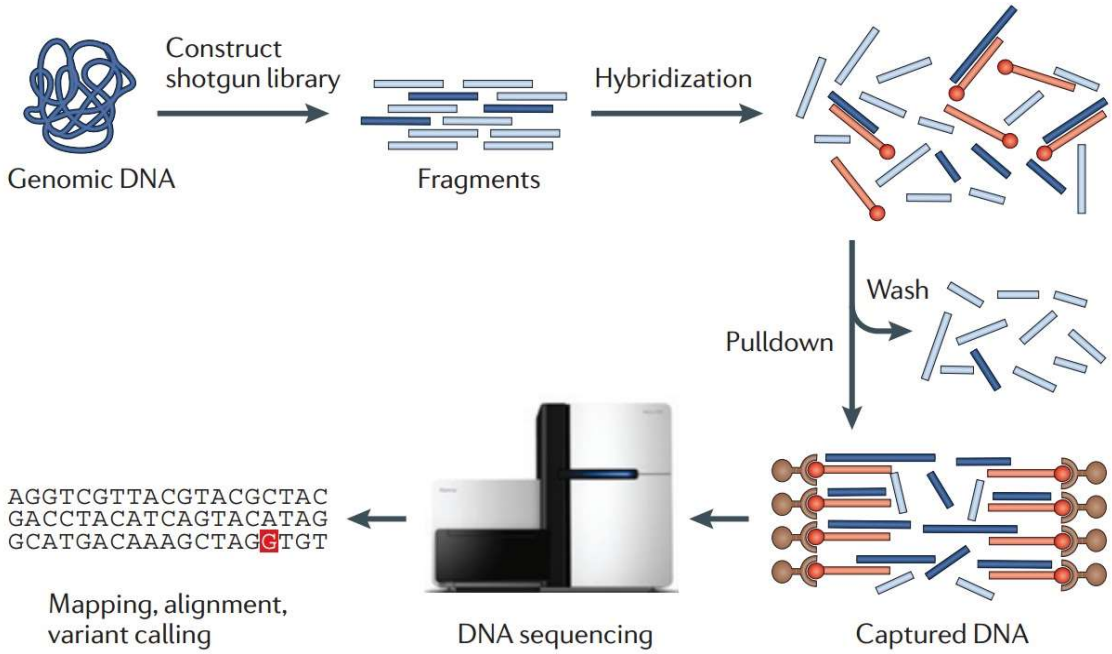


Figure000: Schematic illustration showing the basic steps required for exome sequencing. In vitro shotgun library is constructed of genomic DNA flanking the fragments by adaptors. Then the sequences corresponding to exons (dark blue fragments) are enriched by hybridization capture, followed by washing, elution, additional amplification and pulling down the captured DNA. Thereafter, DNA sequencing is performed using one of the available sequencing platforms such as Illumina and the resulting raw data is analysed using bioinformatics (Figure 4-1). Figure is adapted from Bamshad et al. 2011.

2.6 Pre-mRNA splicing

Genetic diseases result in wide-ranging syndromes and phenotypic variability. One of the significant causes underlying this variability is mutations (i.e. missense, nonsense or frameshift mutations) affecting pre-mRNA splicing, causing various defects in humans [Faustino & Cooper 2003; Krawczak et al. 2009]. The diseases caused by a mutation affecting the canonical splice site include familial dysautonomia [Slaugenhaupt et al. 2001], Menkes disease [Møller et al. 2000], Hutchinson-Gliford progeria syndrome [Eriksson et al. 2003].

Briefly, human genes, like other mammalian genes, consist of short protein coding regions, the exons, being interrupted by much larger non-coding regions, the introns. To form mature RNA, the first step of protein synthesis, the introns have to be removed from functionally differing exons. This process is called pre-mRNA splicing and is regulated by complex macromolecular machinery, the spliceosome, which includes 5 small RNAs and dozens of polypeptides [Cartegni et al. 2002].

Introns include several conserved sequences which are crucial in the splicing process: the GT and AG dinucleotides at the 5' (donor) and 3' (acceptor) exon-intron junctions, the A-branch site and the pyrimidine rich site [Keller & Noon 1984]. The splicing process begins by the spliceosome precisely recognizing the donor splicing site and binding nearby to it, followed by cleavage of the intron at the 5'-site and formation of a lariat at the A-branch site (Figure 2-4). This is followed by the intron being cleaved at the 3'-site, resulting in the two exons being ligated together. The splicing process ends with the removal of the spliced mRNA from the spliceosome and degradation of the intron. During the splicing process, the spliceosome undergoes several conformational changes which are regulated at different levels [Nilsen 2003].

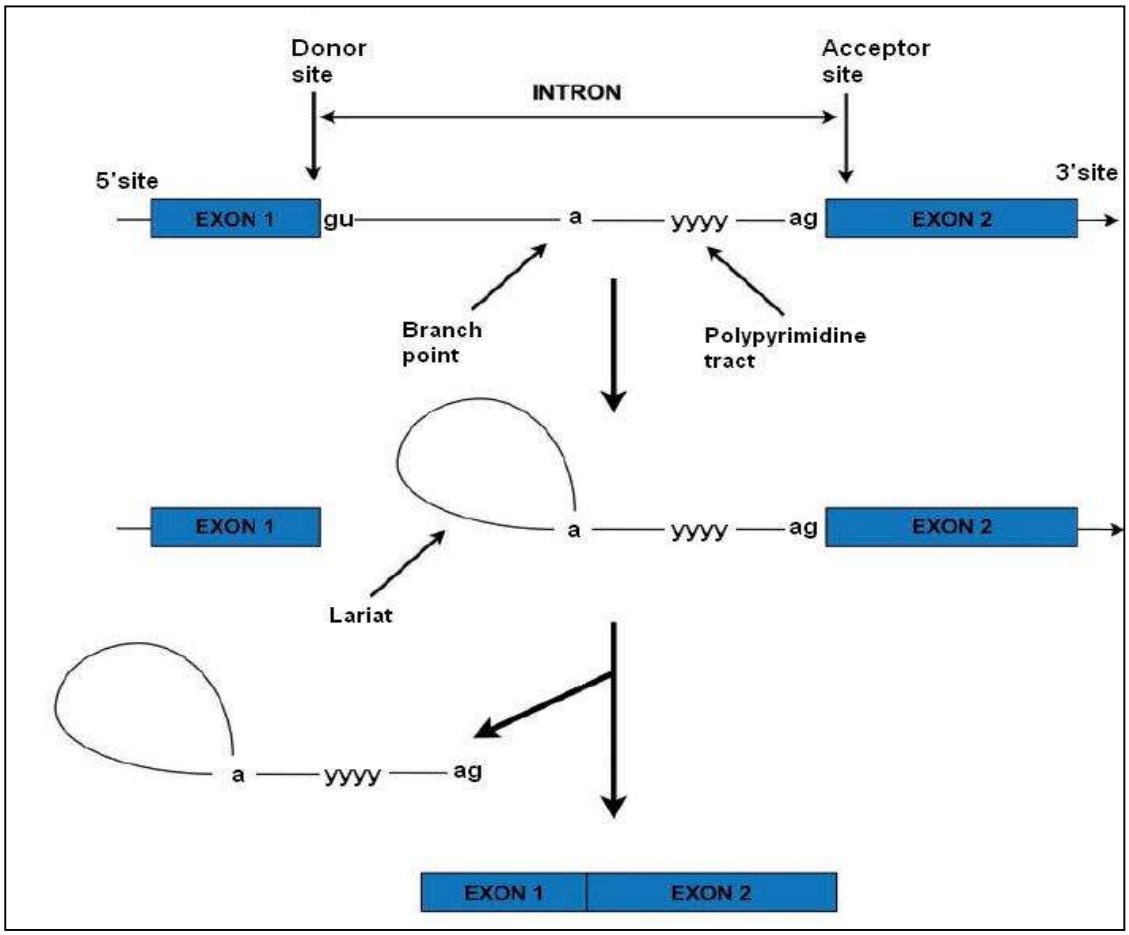


Figure 2-2: Pre-mRNA splicing simplified. First, donor and acceptor sites are recognized, the lariat structure is created and joined at the branch point. Second, the lariat intron is released and the exons are ligated together. Figure is a rewrite from Douglas and Wood (2003).

For a long time it was believed that the number of genes correlates with the number of proteins. However, after human genome sequencing was completed and the number of genes was proven to be around thirty thousand [Lander et al. 2001; Venter et al. 2001], the questions regarding genomic complexity became more inevitable than ever. Amongst other mechanisms investigated as potential sources of this genomic complexity, alternative splicing received major attention and nowadays has established its position as the underlying cause. Since the number of genes was much lower than the number of proteins and thus indicated that the number of expressed sequences, the mRNA forms to be much higher than the actual gene number. Nowadays, it is estimated over 90% of human protein-coding genes express more than one mRNA variant, derived through alternative splicing and 80% of these splicing variants cause changes in the encoded protein product [Pan et al. 2009]. Additionally, Lopez-Bigas et al. (2005) hypothesized that up to 50% of pathogenic mutations affect splicing at some level. These

mutations (Figure 2-5) can result in complete or partial skipping of an exon or intron, introducing a new splice site or truncating the protein product entirely, changing the protein isoform by disrupting splicing elements or splicing machinery, mis regulation of mRNA by affecting splicing regulators [Faustino & Cooper 2003].

Understanding mechanisms and effects of splicing mutations may lead to answers to many human diseases while also leading to development of new therapeutic clinical applications [Douglas & Wood 2011]. Although, some therapeutic approaches utilizing splicing already exist, such as antisense oligonucleotides, trans-splicing and small molecule drug therapy [Wang & Cooper 2007].

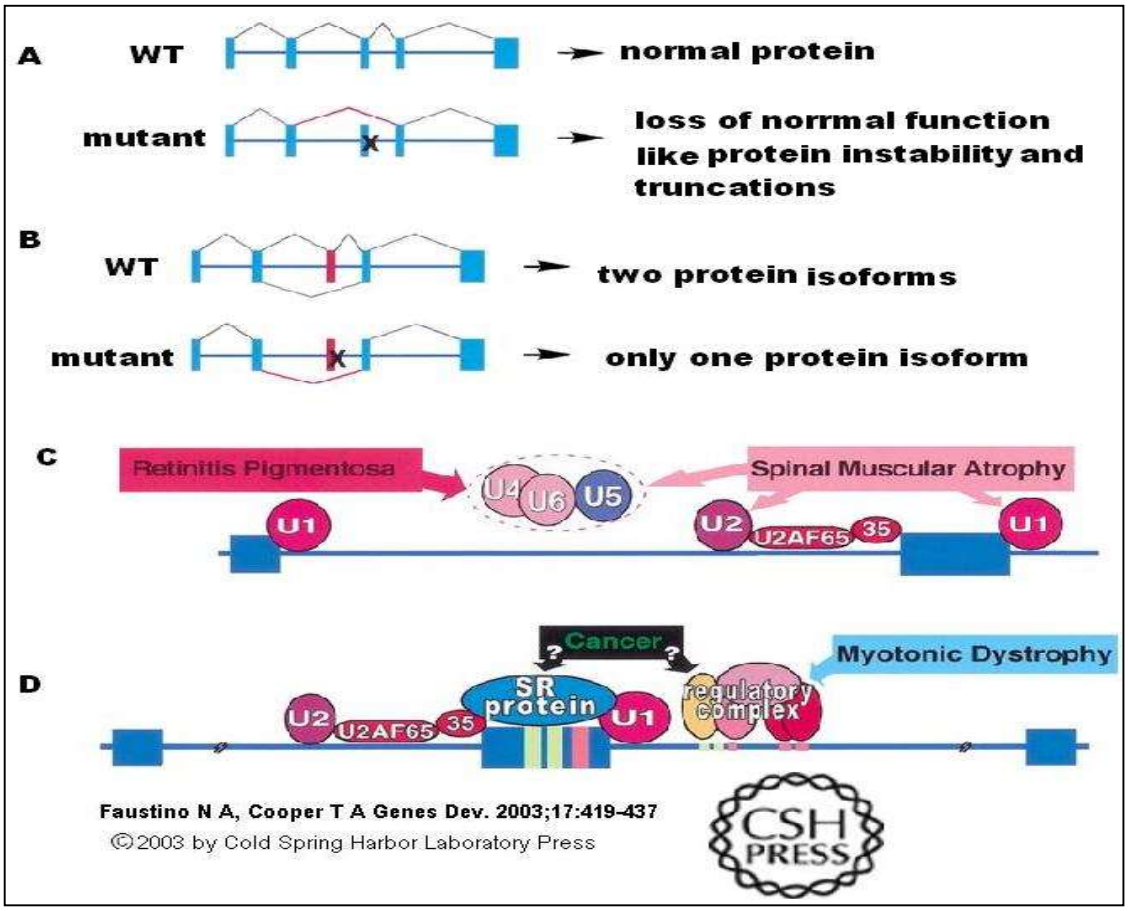


Figure 2-3: Classes of known disease causing pre-mRNA splicing defects. (A) Mutations that disrupt classical splicing signals caused by a primary defect in pre-mRNA splicing. The result is expression of unnatural mRNAs, truncated or unstable proteins and a shift in the reading frame. (B) Mutations disrupting alternative splicing. (C) Mutations disrupting the basal splicing machinery. (D) Mutations affecting splicing regulation: myotonic dystrophy and several forms of cancer arise this way. Picture is adapted from Faustino and Cooper (2003).

3 AIM OF THE STUDY

To identify mutations in Finnish families with multiple affected male patients applying exome (coding region) sequencing for the X chromosome.

4 MATERIALS AND METHODS

4.1 Family material

Two large Finnish pedigrees (Figures 5-1 & 5-3) in which at least three males have a non-syndromic intellectual disability (NS-ID) participated in this study. The families have been diagnosed at the Department of Medical Genetics, Family Welfare Federation, Helsinki. Despite extensive clinical studies, chromosomal analysis, Fragile-X and metabolic profiling, the underlying cause has remained unknown. The clinicians responsible for assembling the patients and postulating the NS-ID diagnosis were docent Mirja Somer, MD Kristiina Avela and MD Maarit Peippo.

The study has approval from the Ethical Committee of the Helsinki University Central Hospital (HUS) and its laboratories (HUSLAB). Written informed consent was obtained from all participating individuals and/or their parents. DNA from anonymous Finnish blood donors was used as control samples.

4.1.1 Clinical findings

In the first family (D172), there are three affected male patients who all have mild to moderate ID. In the second family (D174), the patients have severe ID with autistic features and other dysmorphic features such as brachycephaly, deep set eyes and prominent supraorbital ridges. They appear to have epilepsy, short stature, malposition of the feet and behavioural problems such as self-injury and aggressive outbursts.

4.2 Laboratory methods

4.2.1 DNA and RNA isolation

10 ml of peripheral blood from the participants was collected in ethylene diamine tetra-acetic acid (EDTA) vials and stored at -20°C. DNA was extracted from the blood samples using a non-enzymatic DNA extraction method. The HUSLAB protocol MP025 with a slight modification; Igepal was used to disrupt cell membranes instead of nonident. The purified DNA samples were stored at -20°C.

The QIAGEN PAXgene blood miRNA system was applied for RNA extraction. This procedure provides a complete solution for stabilization and purification of high-quality total RNA >18 nucleotides.

Whole blood from the affected members in family D172 was collected in PAX-gene Blood RNA tubes. After 72 hours of incubating at room temperature (RT), they were stored at -20°C. The PAXgene blood miRNA Kit was used for purifying RNA.

NanoDrop spectrophotometer ND-1000 was utilised to measure the concentration of the extracted DNA and RNA samples. The Agilent 2100 Bioanalyzer, conjunction with the Agilent RNA 6000 Nano Kit [Schroeder et al. 2006], was used for integrity and quality control of the extracted RNA samples. RNA samples were stored at -80°C.

4.2.2 Exome sequencing

A DNA sample of the index patient in each family was sent to Max Planck Institute for Molecular Genetics in Berlin, Germany where exome sequencing of the X chromosome was performed using Agilent SureSelect Human X chromosome kit and single-read 76 nt NGS on the Illumina Genome Analyzer II (GAII) sequencer to identify genetic variants.

In this method [Najmabadi et al. 2011] exons from homozygous intervals were enriched with custom-made Agilent SureSelect DNA capture arrays. Afterwards, they were sequenced on the Illumina GAII which yields 76-bp single reads and covers >98% of the targeted exons. After applying many bioinformatics tools to process the raw data (Figure 4-1), the resulting sequence was aligned to the human reference genome (hg18) to predict the variants. At this stage, low-quality sequence reads were removed and different algorithms were used for detecting reads that could not be aligned to the reference genome, such as smaller insertions and deletions. In addition, all sequence variants were filtered against dbSNP, the 1000 Genomes Project, exomes of 200 Danish control individuals and the in-house database in Max Planck Institute to remove presumably non-pathogenic changes that have been reported previously. For detecting all pathogenic changes, that have been described before, the sequence data was

filtered against the OMIM catalogue and the Human Gene Mutation Database [HGMD].

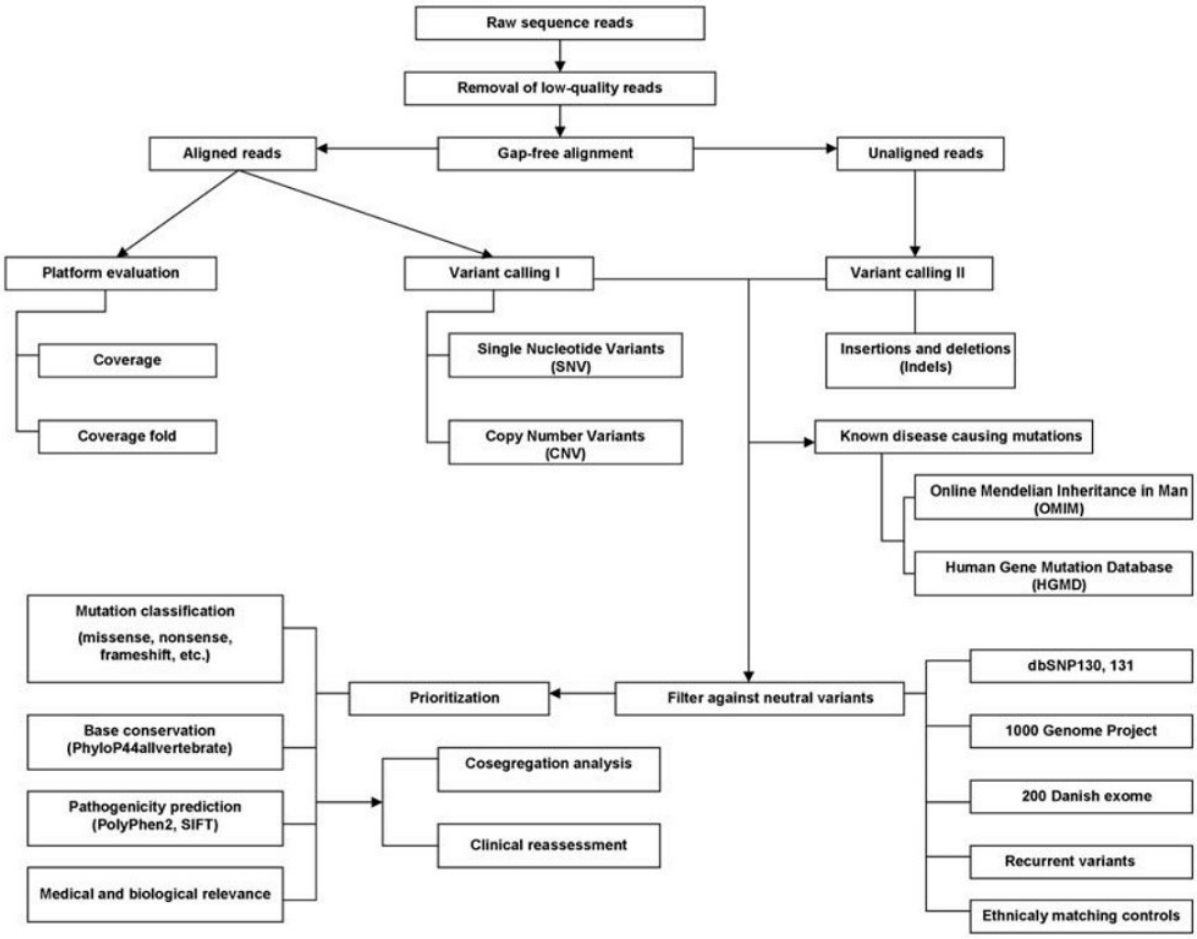


Figure 4-1: Schematic illustration of processing the data received from the Illumina Genome analyser. The workflow describes the in-house pipeline developed in Max Planck institute for analysing the raw sequence reads resulting from whole genome/ -exome sequencing. The pipeline includes the bioinformatics steps such as variant calling, filtering and prioritization. The scheme is adapted from Najmabadi et al. (2011).

4.2.3 Segregation analysis

In order to identify the disease-causing mutation, the segregation of the identified variants was analysed in the families. To verify the frequency of the variants, DNA from about 130 anonymous blood donors were analysed. Conventional Sanger sequencing was performed to analyse whether the mutation segregated with ID in the studied families.

4.2.4 Genotyping and Sequencing

Genotyping was performed by conducting polymerase chain reaction (PCR) in 15 µl volume using different sets of primers (Table 4-1). The PCR reaction mixture for patient samples consisted of 24ng of genomic DNA, 6 pmol of each primer, 3 nmol of each nucleotide, 0.3 U of Dynazyme II polymerase and 1.5 µl of 10x optimized Dynazyme buffer. For control samples with unknown concentration, 1 µl of DNA was analysed using the above protocol. Conditions for some of the genes are shown in appendix. Polymerase enzyme, nucleotides and reaction buffer were ordered from Thermo SCIENTIFIC (Helsinki, Finland) and the primers from Oligomer (Espoo, Finland).

PCR and sequencing reactions were conducted with a 2720 Thermal Cycler, Applied Biosystems or C1000 Thermal cycler, Bio-Rad. When using the former thermal cycler, amplification conditions for PCR reactions were as follows: first 5 min of denaturation at 95°C (enzyme was added during this incubation to increase specificity of the PCR product), then 35 cycles of DNA amplification containing (30 sec denaturation at 95°C, 30 sec of annealing at primer specific temperature and 30 sec of extension at 72°C), then 10 min of elongation at 72°C and finally cooling down to 10°C. When applying Bio-Rad's machine, the same temperatures and steps were used for amplification except, much less time was needed for each step and the enzyme was added to the master mix.

Amplified samples were separated by conducting electrophoresis on a 1.5% agarose gel with ethidium bromide to verify the success of the reactions. Size marker O'RangeRuler 100 bp DNA Ladder (Fermentas Life Sciences, USA) was used to verify the size of amplified products.

The PCR products were purified using USB® ExoSAP-IT® PCR Product Cleanup (Affymetrix, USA) according to manufactorer's protocol. Since ExoSAP-IT® contains both exonuclease I and shrimp alkaline phosphatase, no buffer exchange is needed. The protocol proceeds as follows: solution containing 2 µl of ExoSAP-IT® and 5 µl of PCR product is, first incubated at 37°C for 15 min to degrade remaining primers and nucleotides, then at 80°C for 15 min to inactivate ExoSAP-IT®.

Following clean-up, PCR products were sequenced using the BigDye® Terminator v3.1 Cycle Sequencing Kit (Applied Biosystems, USA) according to manufacturer's protocol. Reaction volume of 10 µl included (1.8 µl of 5x sequencing buffer, 0.65 µl of selected primer, 0.35 µl of BigDye® enzyme, 2 µl of purified DNA and 5.2 µl of MQ). The sequencing reaction was performed as follows: 25 cycles of (10 sec denaturation at 96°C, 5 sec primer annealing at primer specific temperature and 4 min of extension at 60°C). All of the patients were sequenced in forward (5'-3') and reverse (3'-5') directions and the control samples in just one direction. Following this, the products were taken to the sequencing laboratory at FIMM (Finnish Institute of Molecular Medicine) to conduct capillary sequencing as an outsourcing service. There, the products were further purified from excess dye terminators with DTR v3 filter plates (Edge Biosystems, Gaithersburg, MD) and resolved on an ABI 3730 capillary sequencer.

The sequencing data was retrieved from the user account portal from FIMM and was analysed using Sequencher 4.10.1 software (Gene Codes Corporation, USA) and compared to control sequences obtained from healthy individuals or from the following databases: NCBI, EMBL-EBI and UCSC.

4.2.5 Detection of inactivated X chromosome

The human androgen-receptor gene (*AR*) [OMIM 313700], which is located on the X chromosome contains a highly polymorphic trinucleotide, CAG repeat on its first exon. Allen et al. (1992) showed that the methylation of the *HpaII* site near this repeat correlates with X-inactivation. Thus, the methylation status of this site was characterized in the females of both families, when DNA was available, using FastDigest® *HpaII* restriction enzyme (Fermentas Life Sciences, USA).

10 µl (5 µg) of genomic DNA was digested with 5 µl of FastDigest® enzyme. The reaction, total volume 50 µl, additionally contained 30 µl of nuclease-free water and 5 µl of 10x FastDigest® buffer. After gentle mixing, the reaction was incubated at 37°C for 5 min to activate the enzyme and then heated for 5 min at 65°C to inactivate the enzyme. After this, two PCR reactions were conducted, one for digested and one for undigested DNA using the primer set for the *AR* gene (Table 4-1) and the protocol described above. On the inactive X chromosome, the

methylated DNA is resistant to *HpaII* digestion, allowing it to be PCR amplified, meaning that product is obtained only from the inactive X chromosome. Thus, the results can be analysed by comparing the digested and undigested products on the agarose gel.

4.2.6 Bioinformatics

Primer designing tool, Primer3 [http://frodo.wi.mit.edu/primer3/] was used for designing the primers. Sequence comparisons and searches were performed with NCBI-BLAST [http://blast.ncbi.nlm.nih.gov/Blast.cgi], Ensembl-BLAST/BLAT [http://www.ensembl.org/Multi/blastview] and EMBL-EBI's T-Coffee Sequence Alignment tool [http://www.ebi.ac.uk/Tools/msa/tcoffee/].

Evolutionary conservation of the 5 amino acid residues (4 described in Wu et al. 2007 and one from this study) in 9 organisms were studied by obtaining correct protein sequences from the open source software BIOMART. Afterwards, the sequences were blasted by protein-protein blast (blastp) and aligned using NCBI's Constraint-based Multiple Protein Alignment Tool (COBALT).

Table 4-1: Primers used in the study. The same primers were used for both PCR and Sanger sequencing.

GENE	FORWARD (5'→3')	REVERSE (5'→3')	T(°C)
AR	GCTGTGAAGGTTGCTGTTCCTCAT	TCCAGAATCTGTTCCAGAGCGTGC	58
DLG3	CCAGAGTGCACCTGTACCAA	CTAATGGAGACCCCCAAACA	55
ERCC6L	AGAAGAAGGGGTGGAGGAAA	CCAACCAACTGTTCACCAGA	58
ESX1	CACCCATGGTCCCTATGC	GGCCAGTGTGAGGCACAT	58
F8	GGGCATATGCTCCAGTACTTC	GCTTGGTTTGATTTCCCAAG	58
GPR112	GCACAAAGACAACAAAAATGG	TCGTAACAGCCGACTGAGATT	58
GRIA3	CTTAGATCTGGCCCCTCTGG	TCTCCACAGTCAGGAAAGCA	55
LRCH2	GGAGACCTTCCTTAGTCAAGC	ACAAACCTGTGCTGGTGGTA	63
OR13H1	CCTGTTCTGATACCAGCCTCA	TGGAGTGAGCGAATCCTTATG	63
PLXNA3	CAGTGCAGCCTCCTCTGTTA	TCCTGGAAGGAGCACTCG	63
TAF1	TCTTGATGACCCAGAATCTGTC	CACAGCCCTGTAGCATGTTG	60

5 RESULTS

5.1 Identified variants

Exome sequencing of the X chromosome of the index patients resulted in two potential variants for family D172 and a total of 9 variants for family D174 (Table 5-1). All candidate genes were analysed for segregation in the respective families. To study the rarity of the variants, the frequency in anonymous blood donors was analysed using Sanger sequencing. The UCSC Genome Browser was used to convert the positions of given variants to hg19 annotation, since they were given using annotation hg18. Additionally, genetic databases such as Ensembl, NCBI, VEGA and dbSNP were screened to locate the found variants, discover information concerning the genes involved and deciding the genes of interest.

The X chromosome inactivation pattern evaluation in the females of both families showed no correlation with the phenotype since product was obtained from one X chromosome in all cases.

5.1.1 Family D172

Two potentially causative variants were found in family D172 (Table 5-1, Figure 5-1). One is in the zinc finger and BTB domain containing 33 (*ZBTB33*) gene located on Xq23, and the other in the disc-large homolog 3 (*DLG3*) gene positioned on Xq13.1. The first variant was ruled out since it was a common SNP (dbSNP: rs192480559) although validation data is not available in the database. On the other hand, this gene has not been associated to ID before.

A novel splice mutation in the disc-large homolog 3 (*DLG3*) gene was found to be the best candidate gene in family D172. Based on the position, the potential splice site mutation located on the first exon-intron break point (500+1 G>C) of *DLG3* (Figure 6-2) co-segregated in the family and was absent in 124 control samples. Graphic illustration of the capillary sequencing result is shown in Figure 5-2. The effect of the splice site on the function of the gene has yet to be determined by cDNA sequencing.

Table 5-1: Results from exome sequencing, showing 11 variants in total. Synonymous means that the base change does not affect the amino acid. XLID means the gene is known ID gene.

Family	Gene	Position UCSC hg19	Position on X	Mutation	Variation
D172	*DLG3*	X:69665409-69665409	q13.1	synonymous	XLID
	ZBTB33	X:119389090-119389090	q24	I607S	rs192480559
D174	*TAF1*	X:70680560-70680560	q13.1	N1789S	-
	ERCC6L	X:71425175-71425175	q13.1	G1148R	rs78660817
	ESX1	X:103495171-103495171	q22.2	P320R	rs191202058
	LRCH2	X:114414073-114414073	q23	E261K	rs111460344
	GRIA3	X:122561802-122561802	q25	G630R	XLID
	OR13H1	X:130678678-130678678	q26.2	G211W	-
	GPR112	X:135426958-135426958	q26.3	F365I	-
	PLXNA3	X:153693126-153693126	q28	P653L	rs139336954
	F8	X:154158200-154158200	q28	K1289E	-

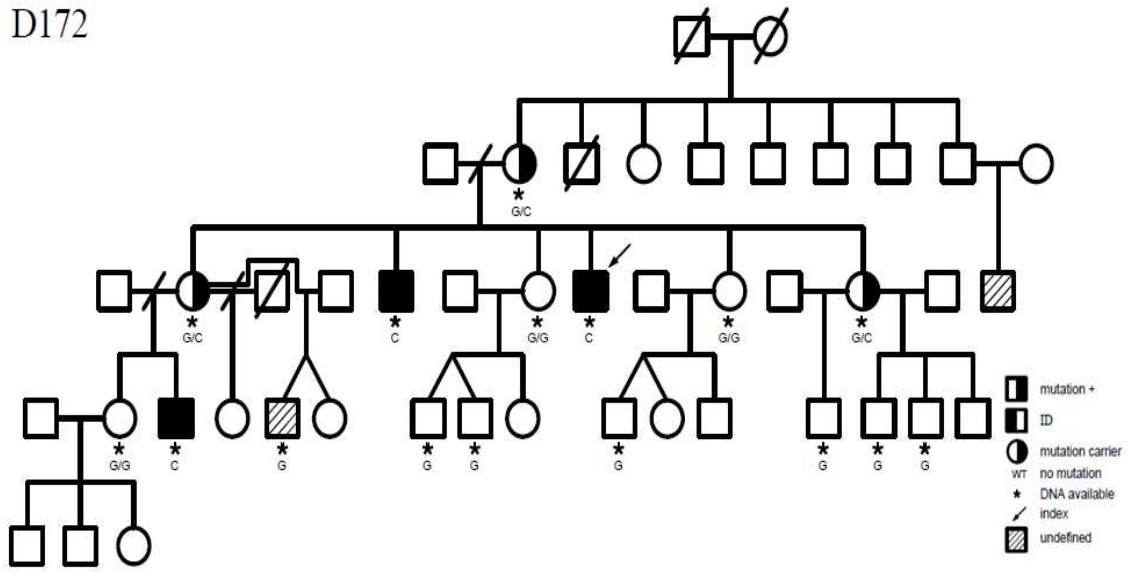


Figure 5-1: Pedigree for family D172 showing segregation of the mutation. Undefined means males have some delay in development, but they have different phenotype than the males affected with ID.

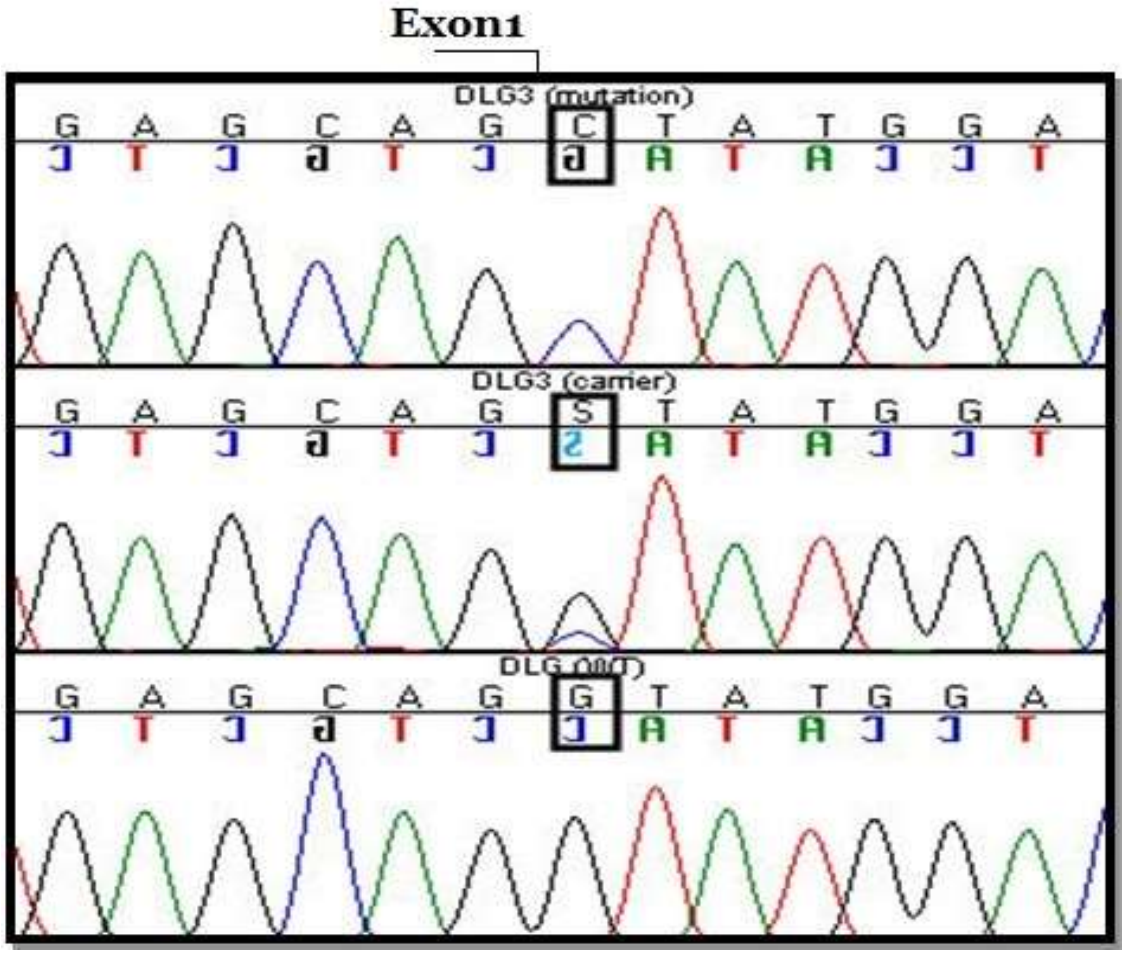


Figure 5-2: Sequence chromatogram showing the splice mutation in exon-intron breakpoint (the splice donor site) 500+1 G>C in *DLG3* for family D172. The chromatogram of the index patient is shown above, carrier in the middle and wild type below. The mutation is indicated by the rectangle.

5.1.2 Family D174

A total of 9 variants were reported in the family D174 (Table 5-1). One variant in each of the following genes: TAF1 RNA polymerase II, TATA box binding protein (TBP)-associated factor (*TAF1*), excision repair cross-complementing rodent repair deficiency, complementation group 6-like (*ERCC6L*), ESX homeobox 1 (*ESX1*), leucine-rich repeats and calponin homology (CH) domain containing 2 (*LRCH2*), olfactory receptor, family 13, subfamily H, member 1 (*OR13H1*), G protein-coupled receptor 112 (*GPR112*), plexin A3 (*PLXNA3*), coagulation factor VIII, procoagulant component (*F8*) and glutamate receptor, ionotropic, AMPA 3 (*GRIA3*). Despite finding a dbSNP-entry for some of the aforementioned variants (Table 5-1), all of them were sequenced. The mutations in genes *ERCC6L, F8, GPR112, LRCH2, OR13H1* and *PLXNA3* did not segregate with ID in the family. By contrast, the variants in four genes; *LRCH2, TAF1, ESX1* and *GRIA3* co-segregated in the studied family (Figure 5-3). The variants were found in the affected persons but absent in healthy males while the mothers of the affected male patients were carriers. However, putative mutations in two of the genes turned out to be common SNPs, since *TAF1* was found in 3.5% of control samples and *LRCH2* in 3.2% of patients manifesting autism (in-house samples used). Additionally, three carriers of the variant in *LRCH2* were found after sequencing 90 blood donor samples. The mutation in *ESX1* gene is located at 27bp sequence which normally repeats 12 times in the gene [Guan 2005]. It is very unlikely that the variant would be pathogenic. On the other hand, *ESX1* has not been associated to any disease to date. Thus, it was ruled out as being a common SNP without further sequencing. The variant in last gene was absent in 135 control samples. Thus, the novel missense mutation G>C at the 5' site of exon 12 (Figure 6-3) changing glycine at codon 630 to arginine (G630R) in *GRIA3* is the best candidate gene to cause ID in family D174. Results of capillary sequencing are illustrated graphically in Figure 5-4.

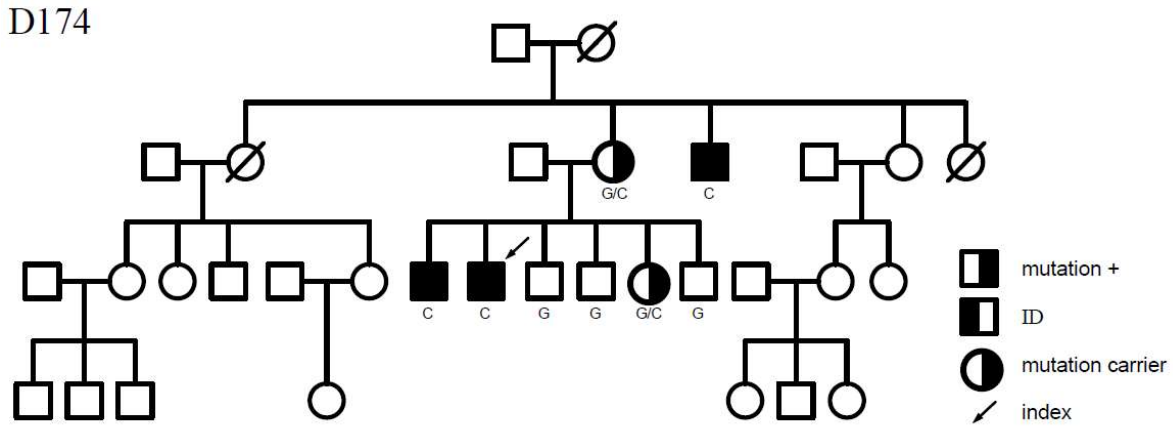


Figure 5-3: Pedigree for the family D174 showing segregation of the mutation.

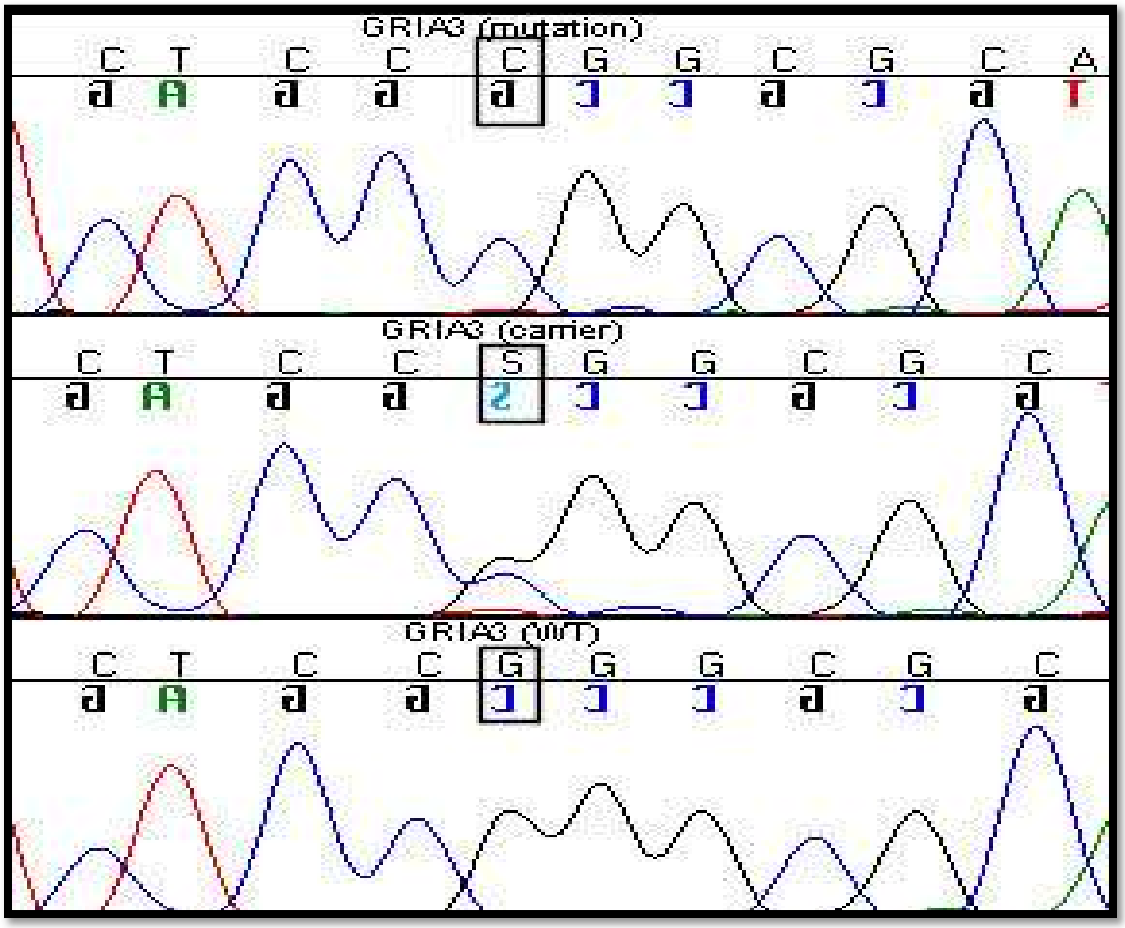


Figure 5-4 Sequence chromatogram showing the novel mutation 2180 G>C in *GRIA3* for family D174. The chromatogram of the index patients is shown above, carrier in the middle and wild type below. The mutation is indicated by the rectangle.

6 DISCUSSION

During this study, novel mutations in *DLG3* and *GRIA3* were identified in two Finnish families with X-linked intellectual disability by applying exome sequencing. Both of these genes have previously been shown to cause XLID. However, this is the first time that mutations in the aforementioned genes are associated to ID in the Finnish population. Finding causes underlying ID is not just important for healthcare but also for the families involved and the society overall. Despite the fact that rapid advances have taken place in identifying ID genes in recent years, majority of patients is far from having a molecular diagnosis.

6.1 A novel mutation in discs, large homolog 3, Drosophila (*DLG3*)

During this study we have found a novel splice mutation in the *DLG3* gene (590+1 G>C) and infer it to be the causative mutation in family D172 with mild intellectual disability (Figure 6-2). Previously, a total of five mutations have been reported in *DLG3*. Tarpey et al. (2004) found two truncating mutations, two splicing mutations and one single nucleotide insertion in *DLG3* in four families while screening 329 families with moderate to severe XLID in which at least two males were affected. Zanni et al. (2010), on the other hand, identified a splice site mutation in one family while studying 300 families with moderate to severe ID in which at least two males were affected. More detailed descriptions of previously identified mutations in *DLG3* are illustrated in Table 6-1 and the position of mutations on the gene including the novel mutation found during this study is illustrated in Figure 6-2. Functional studies and mRNA analysis have to be carried out to discover the impacts of identified mutation on the protein product.

Table 6-1: Mutations described in *DLG3* to date.

Mutation	Change	Disorder	Reference
IVS6-1 G → A	Splice site mutation, frameshift, stop codon at position 357	Moderate-Severe ID	Zanni et al. 2010
Frameshift mutation causing stop codon at position 377	Removes 54% of the normally translated protein	Moderate-Severe ID	Tarpey et al. 2004
1218+5G → A causing frameshift	Splice site mutation, causes frameshift by dropping 26 bp from exon six	Moderate-Severe ID	Tarpey et al. 2004
Single-nucleotide insertion 1325insC-X	Single nucleotide insertion on exon 7	Moderate ID	Tarpey et al. 2004
1535+1G → A causing stop codon	Splice site mutation, Causes stop codon and drops exon 8 out of transcript	Moderate ID	Tarpey et al. 2004
1606 C → G	Nonsense mutation, Causes a stop codon	Severe ID	Tarpey et al. 2004

DLG3 is the drosophila large homolog 3 gene, positioned on the positive strand of the X chromosome at location q13.1. The genomic size of the gene is over 60 kbp, consisting of 19 protein coding exons [UCSC]. *DLG3* encodes synapse-associated protein (SAP) 102 isoform a, which is a member of the membrane-associated guanylate kinase (MAGUK) protein family [Tarpey et al. 2004]. Other members of SAPs include SAP90; also called postsynaptic density (PSD) 95, SAP97 and PSD93. SAP102 shares a similar structure with the aforementioned proteins. All contain three PDZ domains consisting of 90 amino acid repeats in the amino-terminal part, followed by an src homology domain; the SH3 domain and a carboxyl-terminal guanylate kinase-like domain (Figure 6-1) [Müller et al. 1996].

Like all SAP proteins, SAP102 is expressed comprehensively in the brain, playing a crucial role in early development. It is found in axons and dendrites being additionally abundant in the cytoplasm and synapses. SAP102 is crucial for early postnatal brain development since it is expressed in almost all synapses at a high level during the first six months of development and decreases expression after this period. The expression of other SAP proteins such as PSD95 and SAP97 on the other hand is low after birth and increases after six months of age.

Furthermore, SAP102 is involved in signal transduction, especially during synapse maturation by PDZ domains interacting directly with NR2 subunits of postsynaptic glutamate receptors, in particular the N-methyl-D-aspartate (NMDA) receptor and forming protein complexes. In addition, PDZ domains of SAP102 also interact with SynGAP, a GTPase activating protein [Sans et al. 2000; Zanni et al. 2010].

The NMDA receptor is one of the essential neurotransmitter receptors in the brain. It plays a key role in functions such as brain plasticity by regulating excitatory synaptic transmission, learning, memory and cognitive development, signalling pathways and synaptic maturation and development. NMDA receptors have been associated to long-term potentiation and long-term depression [Zito & Scheuss 2009]. McCullumsmith et al. (2007) showed significant reduction of RNA of NMDA and SAP 102 in the hippocampus of individuals suffering from bipolar disorder.

To date, all of the mutations found in *DLG3* (Figure 6-1 & 6-2) introduce a premature stop codon between the second and third PDZ domain [Zanni et al. 2010] and within or before the third PDZ domain [Tarpey et al. 2004]. Identification of the aforementioned mutations linked *DLG3* as the first ID gene directly to the NMDA receptor and the involved trafficking. However, despite the fact that this connection explains the underlying cause of ID occurring in studied males, yet little is known about the pathways and pathophysiology involved.

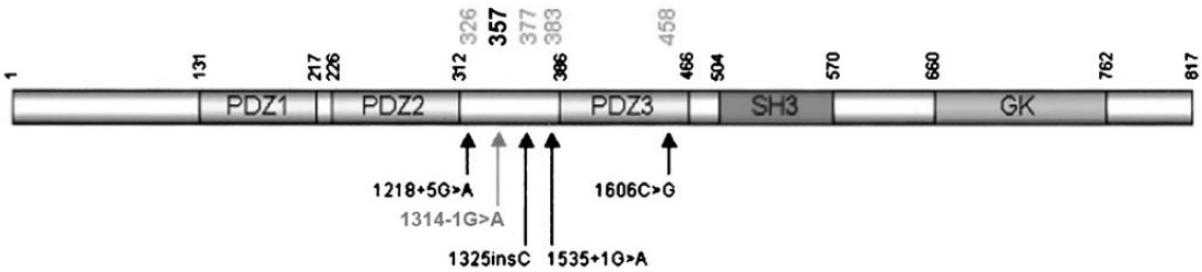


Figure 6-1: Protein domains of SAP102 with positions of mutations identified to date. Adapted from Zanni et al. (2010)

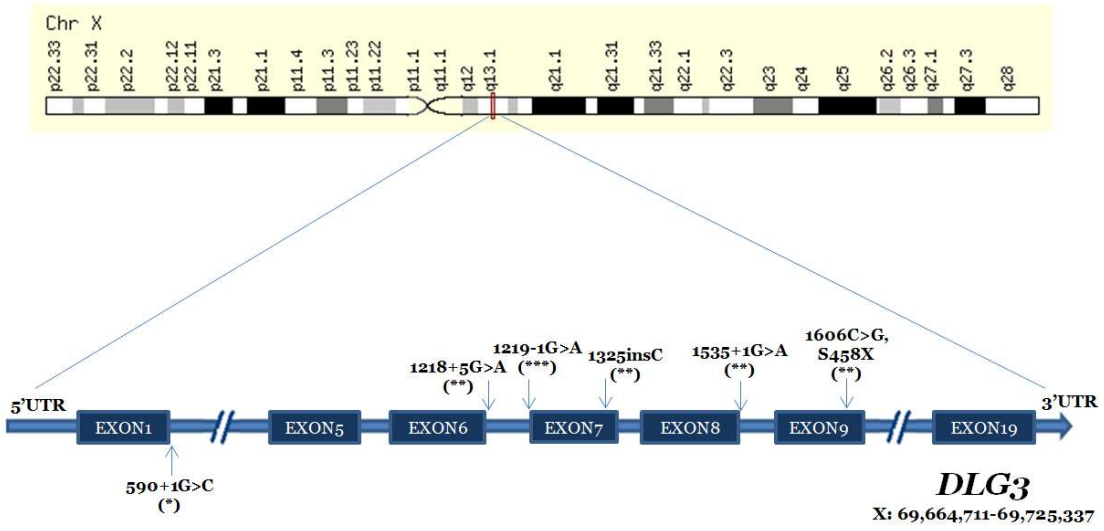


Figure 6-2: The position of the *DLG3* gene on the X chromosome and the mutations and amino acid changes on this gene associated to ID so far. The numbers are based on the transcript *DLG3*-001 from Ensembl genome browser, GRCh37 (hg annotation 19). X=stop codon. (*)=present study, (**) = Tarpey et al. 2004, (***) = Zanni et al. 2010. For example, 590+1 G>C means: 1 base from the donor site of exon 1 on intron, guanine changes to cytosine. Mutation nomenclature discussed in den Dunnen & Antonarakis (2000). Picture is not in scale.

6.2 A novel mutation in glutamate receptor, ionotrophic, AMPA (*GRIA3*)

In the second family, D174, a novel missense mutation (G>C) leading to glycine at codon 630 changing to arginine (G630R) in glutamate receptor, ionotrophic AMPA 3 (*GRIA3*) was shown to completely segregate with XLID in the pedigree (Figure 5-3). Almost 307 kbp long *GRIA3* is located on the long arm of X chromosome at position q25. The most abundant transcript of the gene contains 16 exons [Gécz et al., 1999; Wu et al., 2007]. It encodes ionotropic glutamate receptor AMPA subunit 3 (iGluR3) and has been previously linked to X-linked intellectual disability. However this is the first time *GRIA3* mutation is found in the Finnish population. To date a few genetic and functional studies have shown mutations in *GRIA3* to be associated with mild to moderate cognitive impairment in humans. In 1999 Gécz et al. characterized the gene and linked it to both XLID and bipolar affective disorder. Wu et al. (2007) described one whole gene deletion and four missense mutations in *GRIA3* when sequencing 400 males with XLID. Additionally, to date three duplications (Chiyonobu et al., 2007; Bonnet et al., 2009, 2011) have been reported in *GRIA3* to underlie the intellectual disability in affected individuals. There exists a wide variation in the severity of the learning

disability among the *GRIA3* patients reported previously although seven of the eight are moderately intellectually disabled. However, affected males carrying missense mutation G630R in the Finnish family, exhibit a severe intellectual disability with behavioural disturbances (Figure 6-3).

The genomic organization of *GRIA3* gene was determined as the first human glutamate receptor subunit gene. Comparing gene structure of human *GRIA3* to other genomes indicates it to be highly conserved during evolution. For instance, humans share almost 100% identity of exon-intron boundaries with mouse *GluR2* and a high degree of conservation with *GRIA3* [Köhler et al. 1994]. All of the amino acid residues associated to XLID in *GRIA3* to date including the one recognized during present study show evolutionary conservation (Table 6-2).

Glutamate receptors (GluRs) are divided in ionotropic (iGluRs) containing cation-specific ion channels and metabotropic (mGluRs) receptors. The iGluRs are subdivided into N-methyl-D-aspartate (NMDA) receptor channels, kainate and α-amino-3-hydroxy-5-methyl-4-isoxazolpropionate (AMPA). All have in common a large extracellular N-terminus and four hydrophobic membrane segments (Figure 6-4). Additionally they contain several evolutionary distinct modules.

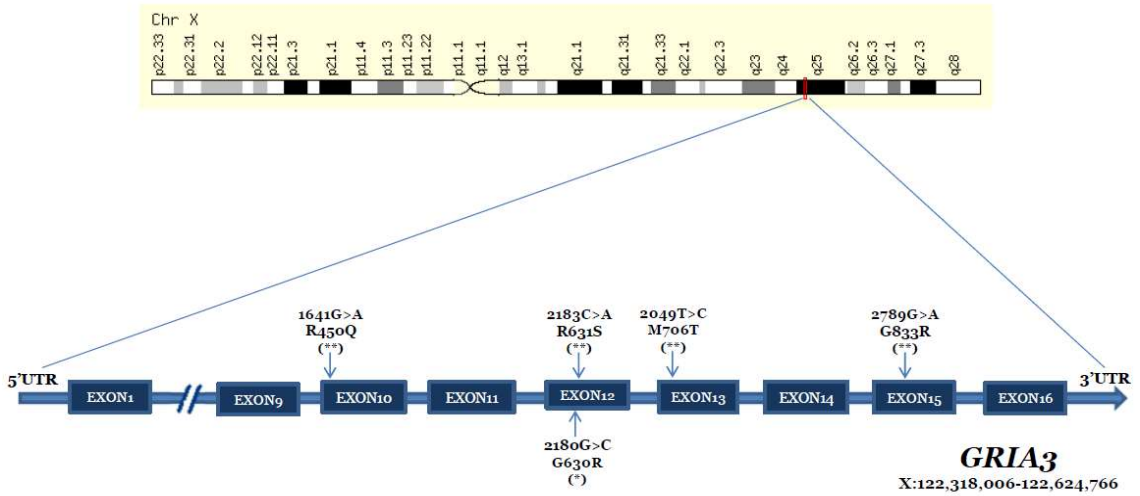


Figure 6-3: Location of the point mutations and the amino acid changes in the *GRIA3* gene found to date. The numbers are based on the transcript *GRIA3*-201 from Ensembl genome browser, GRCh37 (hg annotation 19). (*) = present study, (**) = Wu et al. 2007. Mutation nomenclature discussed in den Dunnen & Antonarakis (2000). Picture is not in scale.

Table 6-2: Evolutionary conservation of the amino acid residues associated to ID to date using COBALT. Modified from Wu et al. (2007).

Species	Amino Acid Residues				
H.sapiens	R450	G630	R631	M706	G833
E.caballus	R	G	R	M	G
P.troglodytes	R	G	R	M	G
C.porcellus	R	G	R	M	G
C.familiaris	R	G	R	M	G
R.norvegicus	R	G	R	M	G
M.musculus	R	G	R	M	G
D.rerio	K	G	R	M	G
G.gallus	K	G	R	M	G

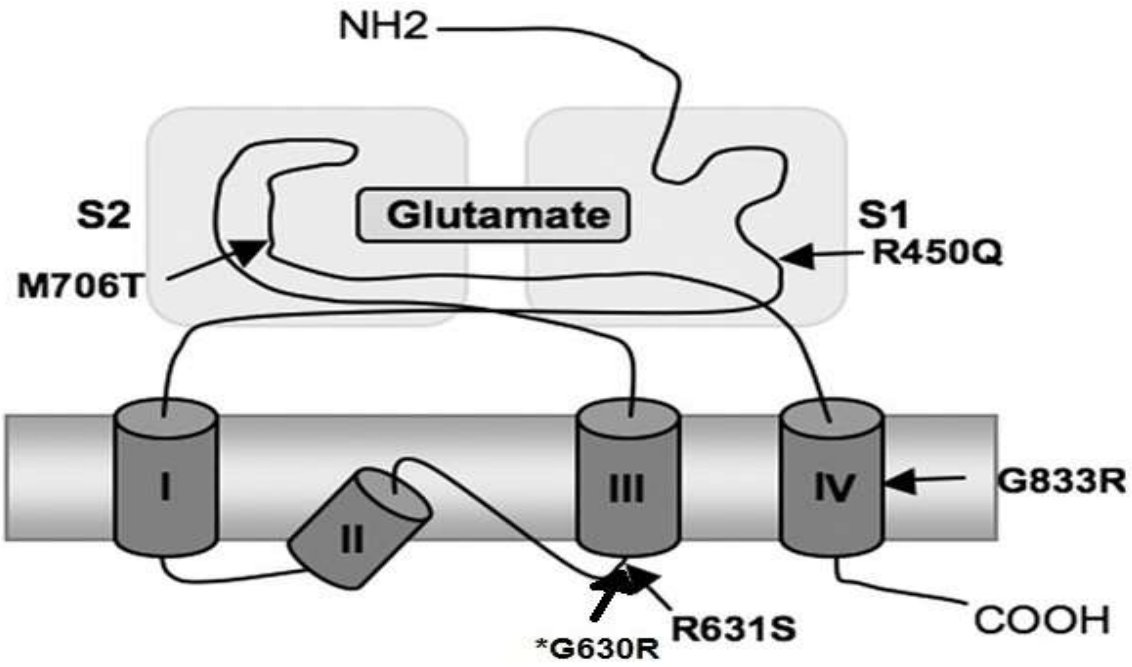


Figure 6-4: Location of iGluR3 missense variants in males with XLID. Schematic diagram shows the location of the variant in iGluR3 found in the current study (*) and four missense mutations described in Wu et al. (2007).

The AMPA receptor has four subunits, GluR1 - GluR4 which are about 900 amino acids long sharing 68-73 % sequence identity. Glutamate excites most neurons of the central nervous system (CNS) of mammals and their receptors are closely involved in the mammalian brain function. They mediate most of excitatory neurotransmission in brain and are involved in plastic changes in synaptic transmission. However in some cases glutamate can be very toxic. For instance, ischemia, head trauma or epileptic seizure results in excessive activation of GluRs and thus leads to the death of central neurons in the brain [Ozawa et al. 1998; Gécz et al. 1999].

In the long term, results achieved during this study will facilitate carrier testing and prenatal diagnosis in the families. They can be used in genetic counselling leading to new forms of family support by society. Furthermore, these results can lead to functional studies and eventually to drug discovery.

Actually, the impacts of this study have already been shown. For instance a young woman in family D174 was proven to be carrier of the mutation and thus she has already attended genetic counselling. In family D172, on the other hand, a young woman planning pregnancy was confirmed not to carry the pathogenic mutation. In addition these families have reached an answer for questions asked more than two decades ago about the causes underlying ID occurring in the pedigrees. Some research laboratories want that the mutations identified in a research laboratory have to be confirmed in an accredited service laboratory. In our laboratory this has not been possible. Thus we have given the result to genetic counsellors provided by a sentence that the study has been performed in a scientific laboratory.

In conclusion, new mutations found in *DLG3* and *GRIA3* further confirm and extend the importance of these genes in cognitive development such as intellectual disability. The results achieved show that mutations in aforementioned genes are pathogenic resulting in moderate to severe ID and thus underlying the XLID in affected males in the studied families. In addition, present study proves exome sequencing to be an excellent method for studying genetic causes underlying XLID.

7 FUTURE PROSPECTS

Development of whole exome and genome sequencing methods has substantially improved the identification of new gene mutations in patients. In severe intellectual disability it is likely that the causative mutation is located on the protein coding site of the genome, which represents just 1.5% of the whole genome. Based on the preliminary results by Vissers et al. (2010) the number of *de novo* events in ID seems to be larger than previously thought. If so, these mutations do not have a recurrence risk to the other siblings in the families. However if the mutation has occurred in one of the parent's germ-line, all siblings are at risk.

The high frequency of *de novo* mutations can be explained by recent evolutionary studies. Lynch (2009) discovered that in the human genome *de novo* mutations take place more frequently than previously estimated. According to his statement, an average new-born is born with 50-100 new mutations of which 0.86% are novel amino acid altering mutations. A mutation can cause a disease when located on a functionally important site of a gene. Recent studies indicate that *de novo* mutations are important causes of ID. Although showing the variability of the human genome, it has made the development of genetic tests very challenging. In practice, genetic tests are available only for the most common genetic syndromes, such as fragile X and RETT syndrome. Whole exome sequencing enables us to identify individual mutations very accurately. Thus, exome sequencing is the most accurate diagnostic method in genetically heterogeneous ID.

One of the future aspects in this field is to go beyond the mutations, to investigate the effect of mutations on the function of the protein, or the changes in regulatory elements. After finding these changes, the next step could be developing a cure to prevent or minimize the consequences of mutations, such as drug or dietary changes, as has been done with PKU. Before, PKU resulted in severe ID, but nowadays the phenotypic consequences are prevented by specific dietary restrictions.

One of the major issues in the field of genetics in the future, specifically concerning whole genome and exome sequencing will be the ethics. Every

genome has hundreds of silent mutations that may be benign and remain without any effect throughout a lifetime. How do we interpret these mutations? Or, what do we do when a study results in a finding we were not searching for? For instance, imagine a situation when a young man comes to a genetic counsellor for testing of a genetic disease running in his family. Since whole genome testing is cost effective, the clinician decides to apply it. As a result, it turns out that the person has a mutation associated to Parkinson's disease. In this situation the clinician has an ethical dilemma for doing the right thing concerning the patient: inform the patient about the mutation and a disease he might get in the future, or just ignore the result, since he sought treatment for something else. The aforementioned situation is in the case when the test is done properly and under the supervision of an expert. What then, when a person orders a genetic test via the internet commercially and as a result, the person seems to have a mutation associated to breast cancer. Without proper knowledge and expertise, misunderstanding the results cannot be avoided. On the other hand, at what point does a person have the right to be informed when a relative is diagnosed as a carrier of a genetic disease? Or, when information concerning a genetic defect should be available to employers and insurers [Reinders 2003]? In addition, when should the results from a genetic study be returned? If not, how should they be managed? These and many more questions have to be answered by researchers and policy makers to develop a framework and guidelines for managing data after conducting a genetic study [Bamshad et al. 2011].

A fall in the costs of high-throughput sequencing technology has made whole genome and whole exome sequencing available for researchers, healthcare professionals, and through commercial applications, also for citizens. This has raised questions concerning managing, storing and interpreting the sequencing data produced since one genome creates an enormous amount of data, let alone hundreds or thousands genomes. In any case, clinicians and researchers in the field of genetics have a great task in avoiding misinterpreting the enormous information brought to them by whole genome and exome sequencing. However, there exists light at the end of the tunnel. New tools and data management pipelines are being invented constantly, in addition to the swiftly developing field of bioinformatics.

8 ACKNOLEDGEMENTS

This thesis project was carried out at the Department of Medical Genetics, University of Helsinki.

I want to thank my supervisor docent Irma Järvelä for giving me the opportunity to participate in this interesting project and allowing me to be part of her lab. I am most grateful and very obliged to her for outstanding guidance, support and patience during this project.

I am very grateful to Päivi Onkamo for her guidance, comments and support on this project.

I am thankful to MD Maarit Peippo, Docent Mirja Somer and MD Kristiina Avela for clinical diagnoses, contacting the families, collecting samples and family data. I also thank them for sharing the clinical data of the patients. I also want to thank the Sigrid Jusélius Foundation for funding this project and all of the family members who participated in this study. Without them this thesis work would not have been possible.

I am most grateful to Minna Ahvenainen for her support, assistance during laboratory work and her remarkable job in editing this thesis.

I want to express my gratitude to all of my co-workers at Irma Järvelä's group for the positive and encouraging working environment, support, feedback and technical help. I am also most thankful to my family and friends for their enormous support, encouragement and care.

9 REFERENCES

9.1 Articles

Abramsky L., Chapple J. 1997, "47,XXY (Klinefelter syndrome) and 47,XYY: estimated rates of and indication for postnatal diagnosis with implications for prenatal counselling", *Prenat. Diagn.* 17: 363-368.

Allen R.C., Zoghbi H.Y., Moseley A.B., Rosenblatt H.M., Belmont J.W. 1992, "Methylation of HpaII and HhaI sites near the polymorphic CAG repeat in the human androgen-receptor gene correlates with X chromosome inactivation", *Am J Hum Genet.* 51: 1229–1239.

Auranen M., Vanhala R., Vosman M., Levander M., Varilo T., Hietala M., Riikonen R., Peltonen L., Järvelä I. 2001, "MECP2 gene analysis in classical rett syndrome and in patients with rett-like features", *Neurology* 56: 611-617.

Ballif B.C., Sulpizio S.G., Lloyd R.M., Minier S.L., Theisen A., Bejjani B.A., Shaffer L.G. 2007, "The clinical utility of enhanced subtelomeric coverage in array CGH", *Am J Med Genet A.* 143:1850-1857.

Bamshad M.J. 2011, "Exome sequencing as a tool for Mendelian disease gene discovery", *Nat Rev Genet,* 12: 745-755.

Basel-Vanagaite L. 2007, "Genetics of autosomal recessive non-syndromic mental retardation: recent advances", *Clinl Genet.* 72: 167-174.

Bauters M., Weuts A., Vandewalle J., Nevelsteen J., Marynen P. 2009, "Detection and validation of copy number variation in X-linked mental retardation", *Cytogenet Genome Res.* 123: 44-53.

Biesecker L.G. 2010, "Exome sequencing makes medical genomics a reality", *Nature genet.* 42:13-14.

Bondy C.A. 2007, "Care of girls and women with Turner syndrome: a guideline of the Turner syndrome study group", *J Clin Endocrinol Metab.* 92, 10-25.

Bonnet C., Leheup B., Béri M., Philippe C., Grégoire M.J., Jonveaux P. 2009, "Aberrant *GRIA3* transcripts with multi-exon duplications in a family with X-linked mental retardation", *Am J Med Genet A.* 149: 1280-1289.

Carrasco X., Castillo S., Aravena T., Rothhammer P., Aboitiz F. 2005, "Williams syndrome: pediatric, neurologic, and cognitive development", *Pediatr Neurol.* 32: 166-172.

Carrel L., Willard H.F. 2005, "X-inactivation profile reveals extensive variability in X-linked gene expression in females", *Nature* 434: 400-404.

Cartegni L., Chew S.L., Krainer A.R. 2002," Listening to silence and understanding nonsense: exonic mutations that affect splicing", *Nat Rev Genet*. 3: 285-298.

Chang S.C., Tucker T., Thorogood N.P., Brown C.J. 2006, "Mechanisms of X-chromosome inactivation", *Frontiers in Bioscience* 11: 852-866.

Charlesworth B. 1991, "The evolution of sex chromosomes", *Science* 251:1030-1033.

Chelly J., Mandel J.L. 2001 "Monogenic causes of X-linked mental retardation", *Nat Rev Genet*. 2: 669-680.

Chelly J., Khelfaoui M., Francis F., Chérif B., Bienvenu T. 2006, "Genetics and pathophysiology of mental retardation", *Eur J Hum Genet*. 14: 701-713.

Chiurazzi P, Schwartz CE, Gecz J, Neri G.2008, "XLMR genes: update 2007", *Eur J Hum Gene*. 16:422-434.

Chiyonobu T, Hayashi S, Kobayashi K, Morimoto M, Miyanomae Y. 2007, "Partial tandem duplication of GRIA3 in a male with mental retardation", *Am J Med Genet A*. 143A: 1448-1455.

Choi M. 2009, "Genetic diagnosis by whole exome capture and massively parallel DNA sequencing", *Proceedings of the National Academy of Sciences of the United States of America,* 106:19096-19101.

Chow J.C., Yen Z., Ziesche S.M., Brown C.J. 2005 "Silencing of the mammalian X chromosome", *Annu Rev Genomics Hum Genet*. 6: 69-92.

Clark M.J. 2011, "Performance comparison of exome DNA sequencing technologies", *Nat Biotechnol,* 29: 908-914.

Crick F.H., Barnett L., Brenner S., Watts-Tobin R.J. 1961, "General nature of the genetic code for proteins", *Nature* 192: 1227-1232.

Crider K.S., Olney R.S., Craqan J.D. 2008, "Trisomies 13 and 18: Population prevalences, characteristics, and prenatal diagnosis, metropolitan Atlanta, 1994-2003", *Am J Med Genet A*. 146: 820-826.

de Vries B.B., Pfundt R., Leisink M., Koolen D.A., Vissers L.E et al. 2005, "Diagnostic genome profiling in mental retardation", *Am J Hum Genet*. 77: 606-616.

Delbridge M.L., McMillan D.A., Doherty R.J., Deakin J.E., Graves J.A. 2008, "Origin and evolution of candidate mental retardation genes on the human X chromosome (MRX)", *BMC Genomics* 9:65.

DeLong G.R. 1993, "Effects of nutrition on brain development in humans", *Am J Clin Nutr*. 57: 286S-290S.

den Dunnen J. T., Antonarakis S.E. 2000, "Mutation nomenclature extensions and suggestions to describe complex mutations: a discussion", *Hum Mutat*. 15: 7-12.

Donnai D., Kamiloff-Smith A. 2000, "Williams syndrome: from genotype through to the cognitive phenotype", Am J Med Genet. 97:164-171.

Douglas A.G.L., Wood M.J.A. 2011, "RNA splicing: disease and therapy", *Brief Funct Genomics* 10: 151-164.

Drews C.D., Murphy C.C, Yeargin-Allsopp M., Decouflé P. 1996, "The relationship between idiopathic mental retardation and maternal smoking during pregnancy", *Pediatrics* 97: 547-553.

Drews C.D., Yeargin-Allsopp M., Decouflé P., Murphy C.C. 1995, "Variation in the influence of selected sociodemographic risk factors for mental retardation", *Am J Public Health* 85: 329-334.

Eriksson M., Brown W.T., Gordon L.B., Glynn M.W., Singer J., Scott L., Erdos M.R., Robbins C.M. 2003, "Recurrent de novo point mutations in lamin A cause Hutchinson-Gilford progeria syndrome", *Nature* 423: 293-298.

Faustino N.A., Cooper T.A. 2003, "Pre-mRNA splicing and human disease", *Genes Dev.* 17: 419-437.

Fisher J.A., Burn J., Alexander F.W., Gardner-Medwin D. 1987, "Angelman (happy puppet) syndrome in a girl and her brother", *J Med Genet.* 24: 294-298.

Fleischmann R.D. 1995, "Whole-genome random sequencing and assembly of Haemophilus influenzae Rd", *Science,* 269:496.

Flint J., Wilkie A.O., Buckle V.J., Winter R.M., Holland A.J., McDermid H.E. 1995, "The detection of subtelomeric chromosomal rearrangements in idiopathic mental retardation", *Nat Genet.* 9: 132-140.

Ford C.E., Hamerton J.L. 1956, "The chromosomes of man", *Nature.* doi:10.1038/ 1781020a0.

Ford C.E., Jones K.W., Polani P.E., De Almeida J.C., Briggs J.H. 1959 "A sex-chromosome anomaly in a case of gonadal dysgenesis (Turner's syndrome)", *Lancet.* 4:711-713.

Frank B., Bermejo J.L., Hemminki K., Sutter C., Wappenschmidt B. 2007, "Copy number variant in the candidate tumor suppressor gene *MTUS1* and familial breast cancer risk", *Carcinogenesis.* 28:1442-1445.

Froyen G., Corbett M., Vandewalle J., Jarvela I., Lawrence O et al. 2008, "Submicroscopic duplications of the hydroxysteroid dehydrogenase HSD17B10 and the E3 ubiquitin ligase HUWE1 are associated with mental retardation", *Am J Hum Genet.* 82: 432-443.

Froyen G., Govaerts K., Van Esch H., Verbeeck J., Tuomi M.L et al. 2009, "Novel PORCN mutations in focal dermal hypoplasia", *Clin Genet.* 79: 535-543.

Gécz J., Barnett S., Liu J., Hollway G., Donnelly A et al. 1999, "Characterization of the human glutamate receptor subunit 3 gene (*GRIA3*), a candidate for bipolar disorder and nonspecific X-linked mental retardation", *Genomics* 62: 356-368.

Gersh M., Goodart S.A., Pasztor L.M., Harris D.J., Weiss L., Overhauser J. 1995, "Evidence for a distinct region causing a cat-like cry in patients with 5p deletions", *Am J Hum Gene.* 56: 1404-1410.

Gilbert W., Maxam A. 1973, "The nucleotide sequence of the lac operator", *Proc Natl Acad Sci U S A.* 70: 3581-3584.

Gilissen C. 2011, "Unlocking Mendelian disease using exome sequencing", *Genome Biol,* 12, 228.

Glessner J.T., Bradfield J.P., Wang K., Takahashi N., Zhang H et al. 2010, "A genome-wide study reveals copy number variants exclusive to childhood obesity cases", *Am J Hum Genet.* 87: 661-666.

Göstason R., Wahlström J., Johannisson T., Holmqvist D. 1991, "Chromosomal aberrations in the mildly mentally retarded ", *J Ment Defic Res.* 35: 240-246.

Gropman A.L., Duncan W.C., Smith A.C. 2006, "Neurologic and developmental features of the Smith-Magenis syndrome (del 17p11.2)", *Pediatr Neurol.* 34: 337-350.

Guan H.H. 2005, "Identification of two novel allelic variants of ESX1L in the human placenta: lack of an association with intrauterine growth restriction", *Placenta,* 26:766-772.

Guilmatre A., Dubourg C., Mosca A.L., Legallic S., Goldenberg A., Drouin-Garraud V., Layet V., Rosier A. 2009, "Recurrent rearrangements in synaptic and neurodevelopmental genes and shared biologic pathways in schizophrenia, autism, and mental retardation", *Arch Gen Psychiatry* 66: 947-956.

Hastings P. J., Lupski J.R., Rosenberg S.M., Ira G. 2009, "Mechanisms of change in gene copy number", *Nat Rev Genet.* 10: 551-564.

Hedges DJ, Burges D, Powell E, Almonte C, Huang J, 2009, "Exome sequencing of a multigenerational human pedigree", *PLoS One.* 4(12):e8232.

Iafrate A.J., Feuk L., Rivera M.N., Listewnik M.L., Donahoe P.K., Qi Y., Scherer S.W., Lee C. 2004, "Detection of large-scale variation in the human genome", *Nat Genet.* 36: 949-951.

Ikeda M., Aleksic B., Kirov G., Kinoshita Y., Yamanouchi Y et al. 2010, "Copy number variation in schizophrenia in the Japanese population", *Biol Psychiatry* 67: 283-286.

Jacobs P.A., Strong J.A. 1959, "A case of human intersexuality having a possible XXY sex-determining mechanism", *Nature.* 31: 302-303.

Kallioniemi A., Kallioniemi O.P., Sudar D., Rutovitz D., Gray J.W et al. 1992 "Comparative genomic hybridization for molecular cytogenetic analysis of solid tumors", *Science* 258: 818-821.

Kaufman L., Ayub M., Vincent J.B. 2010, "The genetic basis of non-syndromic intellectual disability: a review", *J Neurodev Disord*. 2:182-209.

Keller E.B., Noon W.A. 1984, "Intron splicing: a conserved internal signal in introns of animal pre-mRNAs", *Proc Natl Acad Sci U S A,* 81: 7417-7420.

Knight S.J., Regan R., Nicod A., Horsley S.W., Kearney L et al. 1999, "Subtle chromosomal rearrangements in children with unexplained mental retardation", *Lancet* 354: 1676-1681.

Köhler M., Kornau H.C., Seeburg P.H. 1994, "The organization of the gene for the functionally dominant α-amino-3- hydroxy-5-methyl-isoxazole-4-propionic acid receptor subunit GluR-B", *J Biol Chem*. 269: 17367-17370.

Koivisto A.M., Ala-Mello S., Lemmelä S., Komu H.A., Rautio J., Järvelä I. 2007, "Screening of mutations in the PHF8 gene and identification of a novel mutation in a finnish family with XLMR and cleft lip/cleft palate", *Clin Genet*. 72: 145-149.

Kong A., Gudbjartsson D.F., Sainz J., Jonsdottir G.M., Gudjonsson S.A. 2002, "A high-resolution recombination map of the human genome", *Nat Genet*. 31:241-7.

Koolen D.A., Vissers L.E., Pfundt R., de Leeuw N., Knight S.J et al. 2006, "A new chromosome 17q21.31 microdeletion syndrome associated with a common inversion polymorphism", *Nat Genet*. 38: 999-1001.

Krawczak M., Thomas N.S., Hundrieser B., Mort M., Wittig M et al. 2007, "Single base-pair substitutions in exon-intron junctions of human genes: nature, distribution, and consequences for mRNA splicing", *Hum Mutat*. 28: 150-158.

Ku C.S., Cooper D.N., Polychronakos C., Naidoo N., Wu M. 2012, "Exome sequencing: dual role as a discovery and diagnostic tool", *Ann Neurol,* 71: 5-14.

Lahn B. T., Pearson N.M., Jegalian K. 2001, "The human Y chromosome, in the light of evolution", *Nat Rev Genet* 2: 207-216.

Lander E.S., Linton L.M., Birren B., Nusbaum C., Zody M.C et al. 2001, "Initial sequencing and analysis of the human genome", *Nature* 409: 860-921.

Langer-Safer P., Levine M., Ward D.C. 1982, "Immunological methods for mapping genes on Drosophila polytene chromosomes", *Proc Natl Acad Sci U S A*. 79:4381-4385.

Lathrop G.M., Lalouel J.M., Julier C., Ott J. 1985, "Multilocus linkage analysis in humans: Detection of linkage and estimation of recombination", *Am J Hum Genet*. 37:482-498.

Lee C., Iafrate A.J., Brothman A.R. 2007, "Copy number variations and clinical cytogenetic diagnosis of constitutional disorders", *Nat Genet.* 39:48-54.

Lejeune J., Gautier M., Turpin R. 1959, "Study of somatic chromosomes from 9 mongoloid children. | Etude des chromosomes somatiques de neuf enfants mongoliens", *C R Hebd Seances Acad Sci.* 248:1721-1722.

Lejeune J., Lafourcade J., Berger R., Vialatte J., Boeswillwald M. 1963, "Trois cas de délétion partielle du bras court d'un chromosome 5" *CR Acad Sci* 257: 3098–3102.

López-Bigas N., Audit B., Ouzounis C., Parra G., Guigó R. 2005, "Are splicing mutations the most frequent cause of hereditary disease?", *FEBS Lett.* 579: 1900-1903.

Lugtenberg D., Veltman J.A., van Bokhoven H. 2007, "High-resolution genomic microarrays for X-linked mental retardation", *Genet Med.* 9:560-565.

Lynch M. 2010, "Rate, molecular spectrum, and consequences of human mutation", *Proc Natl Acad Sci U S A.* 107:961-968.

Mamanova L., Coffey A.J., Scott C.E., Kozarewa I., Turner E.H. 2010, "Target-enrichment strategies for next-generation sequencing", *Nat Methods.* 7:111-118.

Mardis E.R. 2008, "The impact of next-generation sequencing technology on genetics", *Trends Genet.* 24: 133-141.

McCullumsmith R.E., Kristiansen L.V., Beneyto M., Scarr E., Dean B. 2007, "Decreased NR1, NR2A, and SAP102 transcript expression in the hippocampus in bipolar disorder", *Brain Res.* 1127: 108-118.

McMullan D.J., Bonin M., Hehir-Kwa J.Y., de Vries B.B., Dufke A et al. 2009, "Molecular karyotyping of patients with unexplained mental retardation by SNP arrays: a multicenter study", *Hum Mutat.* 30: 1082-1092.

Miller D.T., Adam M.P., Aradhya S., Biesecker L.G., Brothman A.R et al.2010, "Consensus statement: chromosomal microarray is a first-tier clinical diagnostic test for individuals with developmental disabilities or congenital anomalies", *Am J Hum Genet.* 86: 749-764.

Møller L.B., Tümer Z., Lund C., Petersen C., Cole T et al.2000, "Similar splice-site mutations of the ATP7A gene lead to different phenotypes: classical menkes disease or occipital horn syndrome", *Am J Hum Genet.* 66:1211-1220.

Moreno-De-Luca D., Mulle J.G., Kaminsky E.B., Sanders S.J., Myers S.M et al. 2010, "Deletion 17q12 is a recurrent copy number variant that confers high risk of autism and schizophrenia", *Am J Hum Genet.* 88:618-630.

Morgan T.H. 1911, "Chromosomes and associative inheritance", *Science* 34: 636-638.

Müller B.M., Kistner U., Kindler S., Chung W.J., Kuhlendahl S. 1996, "SAP102, a novel postsynaptic protein that interacts with NMDA receptor complexes in vivo", *Neuron.* 17: 255-265.

Najmabadi H., Hu H., Garshasbi M., Zemojtel T., Abedini S.S et al. 2011, "Deep sequencing reveals 50 novel genes for recessive cognitive disorders", *Nature* 478: 57-63.

Ng S.B., Turner E.H., Robertson P.D., Flygare S.D., Bigham A.W. 2009, "Targeted capture and massively parallel sequencing of 12 human exomes", *Nature*. 461: 272-276.

Nilsen T.W. 2003, "The spliceosome: the most complex macromolecular machine in the cell?", *Bioessays* 25: 1147-1149.

Ozawa S., Kamiya H., Tsuzuki K. 1998, "Glutamate receptors in the mammalian central nervous system", *Prog Neurobiol*. 54: 581-618.

Pan Q., Shai O., Lee L.J., Frey B.J., Blencowe B.J. 2008"Deep surveying of alternative splicing complexity in the human transcriptome by high-throughput sequencing", *Nat Genet*. 40:1413-1415.

Peippo M., Koivisto A.M., Särkämö T., Sipponen M., Von Koskull H et al. 2007, "PAK3 related mental disability: further characterization of the phenotype", *Am J Med Genet A*. 143A: 2406-2416.

Pettersson E., Lundeberg J., Ahmadian A. 2009, "Generations of sequencing technologies", *Genomics*. 93: 105-111.

Pugh T.J, Weeraratne S.D., Archer T.C., Pomeranz Krummel D.A., Auclair D. 2012, "Medulloblastoma exome sequencing uncovers subtype-specific somatic mutations", *Nature*. 488: 106-110.

Racchi O., Mangerini R., Rapezzi D., Rolfo M., Gaetani G.F., Ferraris A.M. 1998, "X chromosome inactivation patterns in normal females", *Blood Cells Mol Dis*. 24: 439-47

Ravnan J.B., Tepperberg J.H., Papenhausen P., Lamb A.N., Hedrick J. 2006, "Subtelomere FISH analysis of 11 688 cases: an evaluation of the frequency and pattern of subtelomere rearrangements in individuals with developmental disabilities", *J Med Genet*. 43: 478-489.

Raymond F.L. 2006, "X linked mental retardation: a clinical guide", *J Med Genet* 43: 193-200.

Redon R., Ishikawa S., Fitch K.R., Feuk L., Perry G.H et al. 2006, "Global variation in copy number in the human genome", *Nature* 444: 444-454.

Reinders, H.S. 2003, "Introduction to intellectual disability, genetics and ethics J Intellect Disabil Res. 47:501-4.

Robinson W.P., Bottani A., Xie Y.G., Balakrishman J., Binkert F et al. 1991, "Molecular, cytogenetic, and clinical investigations of Prader-Willi syndrome patients", *Am J Hum Genet*. 49: 1219-1234.

Rodriguez-Revenga L., Mila M., Rosenberg C., Lamb A., Lee C. 2007, "Structural variation in the human genome: the impact of copy number variants on clinical diagnosis", *Genet Med*. 9: 600-606.

Ropers H.H. 2010, "Genetics of Early Onset Cognitive Impairment", *Annu Rev Genomics Hum Genet* 22:161-187.

Ross M.T., Grafham D.V., Coffey A.J., Scherer S., McLay K et al.2005, "The DNA sequence of the human X chromosome", *Nature* 434: 325-337.

Saiki R.K., Scharf S., Faloona F., Mullis K.B., Horn G.T. 1985, "Enzymatic amplification of betaglobin genomic sequences and restriction site anal ysis for diagnosis of sickle cell anemia", *Science*. 20:1350-1354.

Sans N, Petralia RS, Wang YX, Blahos J 2nd, Hell JW, Wenthold RJ. 2000, "A developmental change in NMDA receptor-associated proteins at hippocampal synapses", *J Neurosci,* 20: 1260-1271.

Schalock R.L., Luckasson R.A., Shogren K.A., Borthwick-Duffy S., Bradley V et al. 2007, "The Renaming of Mental Retardation: Understanding the Change to the Term Intellectual Disability", *Intellec Dev Disabil*. 45: 116-124.

Schroeder A., Mueller O., Stocker S., Salowsky R., Leiber M. 2006, "The RIN: an RNA integrity number for assigning integrity values to RNA measurements", *BMC Mol Biol*. doi: 10.1186/1471-2199-7-3.

Seabright M. 1971, "Arapid banding technique for human chromosomes", *Lancet*. 30: 971-972.

Sebat J., Lakshmi B., Troge J., Alexander J., Young J et al. 2004, "Large-scale copy number polymorphism in the human genome", *Science* 23: 525-528.

Sharp A.J., Mefford H.C., Li K., Baker C., Skinner C et al. 2008, "A recurrent 15q13.3 microdeletion syndrome associated with mental retardation and seizures", *Nat Genet*. 40: 322-328.

Singh N.D., Petrov D.A. 2007, "Evolution of gene function on the X chromosome versus the autosomes", *Genome Dyn*. 3: 101-118.

Skuse D.H. 2005, "X-linked genes and mental functioning", *Hum Mol Genet*. 15: 27-32.

Slaugenhaupt S.A., Blumenfeld A., Gill S.P., Leyne M., Mull J et al. 2001, "Tissue-specific expression of a splicing mutation in the *IKBKAP* gene causes familial dysautonomia", *Am J Hum Genet* 68: 598-605.

Sun Y., Almomani R., Aten E., Celli J., van der Heijden J. 2010, "Terminal Osseous Dysplasia Is Caused by a Single Recurrent Mutation in the FLNA Gene", *Am J Hum Genet*, 87:146-153.

Tarpey P., Parnau J., Blow M., Woffendin H., Bignell G. 2004, "Mutations in the *DLG3* Gene Cause Nonsyndromic X-Linked Mental Retardation", *Am J Hum Genet.* 75:318-324.

Tarpey P.S., Smith R., Pleasance E., Whibley A., Edkins S et al. 2009, "A systematic, large-scale resequencing screen of X-chromosome coding exons in mental retardation", *Nat Genet.* 41:535-543.

Tartaglia N.R., Howell S., Sutherland A., Wilson R., Wilson L. 2010, "A review of trisomy X (47, XXX)", *Orphanet J Rare Dis.* 5: 1-9

Tjio J. H., Levan A. 1956, "The chromosome number of man", *Hereditas* 42: 1–6.

Topper S., Ober C., Das S. 2011, "Exome sequencing and the genetics of intellectual disability", *Clin Genet.* 80: 117-126.

Trachtenberg A.J., Robert J.H., Abdalla A.E., Fraser A., He S.Y. et al. 2012, "A primer on the current state of microarray technologies", *Methods Mol Biol.* 802: 3-17.

Turner G. 1996, "Intelligence and the X chromosome", *Lancet,* 347: 1814-1815.

Vaillend C., Poirier R., Laroche S. 2008, "Genes, plasticity and mental retardation", *Behav Brain Res.* 192: 88-105.

van Bokhoven H. 2011, "Genetic and epigenetic networks in intellectual disabilities", *Annu Rev Genet.* 45: 81–104.

van Buggenhout G., Fryns J.P. 2009, "Angelman syndrome (AS, MIM 105 830)", *Eur J Hum Genet.* 17: 1367-1373.

Venter J.C., Adams M.D., Myers E.W., Li P.W., Mural R.J., Sutton G.G., Smith H.O., Yandell M et al. 2001, "The sequence of the human genome", *Science* 291: 1304-1351.

Verkerk A.J., Pieretti M., Sutcliffe J.S., Fu Y.H., Kuhl D.P., Pizzuti A., Reiner O., Richards S et al. 1991, "Identification of a gene (FMR-1) containing a CGG repeat coincident with a breakpoint cluster region exhibiting length variation in fragile X syndrome", *Cell* 65: 905-914.

Vicoso B., Charlesworth B. 2006, "Evolution on the X chromosome: unusual patterns and processes", *Nat Rev Genet.* 7: 645-653.

Vissers L.E., de Ligt J., Gilissen C., Janssen I., Steehouwer M et al. 2010, "A de novo paradigm for mental retardation", *Nat Genet.* 42: 1109-1112.

Vissers L.E., de Vries B.B., Veltman J.A. 2010, "Genomic microarrays in mental retardation: from copy number variation to gene, from research to diagnosis", *J Med Genet* 47: 289-297.

Wang G.S., Cooper T.A. 2007, "Splicing in disease: disruption of the splicing code and the decoding machinery", *Nat Rev Genet*. 8: 749-761.

Waters P.D and Robinson T.J. 2008, "Evolution of the Mammalian X Chromosome". In: eLS. John Wiley & Sons Ltd, Chichester. http://www.els.net [doi: 10.1002/9780470015902.a0020739]

Watson J.D., Crick F.H.C. 1953, "Molecular structure of nucleic acids: a structure for deoxyribose nucleic acid", *Nature* 171: 737-738.

Wu Y., Arai A.C., Rumbaugh G., Srivastava A.K., Turner G et al. 2007, "Mutations in ionotropic AMPA receptor 3 alter channel properties and are associated with moderate cognitive impairment in humans", *Proc Natl Acad Sci U S A*. 104: 18163-18168.

Wutz A., Gribnau J. 2007, "X inactivation Xplained", *Curr Opin Genet Dev*. 17: 387-393.

Zanni G., van Esch H., Bensalem A., Saillour Y., Poirier K et al. 2010, "A novel mutation in the *DLG3* gene encoding the synapse-associated protein 102 (SAP102) causes non-syndromic mental retardation", *Neurogenetics* 11: 251-255.

Zito K., Scheuss V. 2009, "NMDA receptor function and physiological modulation", In: Encyclopedia of Neuroscience 6: 1157-1164.

9.2 Books, manual and tests

Koppitz, Elizabeth M.1964. The Bender Gestalt Test for young children. Volume I.

McDermott S, Durkin MS, Schupf N, et al. Epidemiology and etiology of mental retardation. In: Jacobson JW, Mulick JA, Rojahn J, editors. Handbook of Intellectual and Developmental Disabilities. New York: Springer, 2007. P. 3-40.

Schalock R.L., Borthwick-Duffy S.A., Bradley V.J., Buntinx W.H., Coulter D.L. et al. Intellectual Disability: Definition, Classification, and Systems of Supports. The 11th edition of the AIDD definition manual 2010.

Sparrow S. S., Cicchetti D. V. and Balla A. D. 2005. Vineland Adaptive Behavior Scales, Second Edition (Vineland II) A measure of adaptive behaviour from birth to adulthood.

World Human Organization 2010, International Statistical Classification of Diseases and Related Health Problems 10th Revision (ICD-10).

9.3 Electronic databases

dbSNP. 2012. http://www.ncbi.nlm.nih.gov/projects/SNP/. NCBI: Short Genetic Variations; visited 20.10.2012.

EMBL-EBI. 2012. http://www.ebi.ac.uk/. European Molecular Biology Laboratory's European Bioinformatics Institute; visited 20.10.2012.

Ensembl. 2012. http://www.ensembl.org/index.html. Ensembl Genome Browser; visited 20.09.2012.

Euro-MRX. 2012. http://www.euromrx.com/en/consortium.html. European Mental Retardation Consortium; visited 28.09.2010

HGNC. 2012. http://www.genenames.org/. HUGO Gene Nomenclature Committee; visited 05.11.2012

NCBI. 2012. http://www.ncbi.nlm.nih.gov/. The National Centre for Biotechnology Information; visited 24.09.2012

OMIM. 2012. http://omim.org/. Online Mendelian Inheritance in Man; visited 22.09.2012

UCSC. 2012. http://genome.ucsc.edu/. UCSC Genome Browser University of California Santa Cruz; visited 22.09.2012

VEGA. 2012. http://vega.sanger.ac.uk/index.html/. The Vertebrate Genome Annotation (VEGA) database; visited 23.09.2012

10 APPENDIX

Tables show the conditions in which the remaining genes for family D174 were analysed.

Precision melt supermix (Bio-Rad)	5
Primer For (XuM)	1
Primer Rev (XuM)	1
H2O	1
TOTAL	8

Primers	Concentration	Annealing Temp.	Product Melt Temp.	For. Anneal Temp. Sequencing	Rev. Anneal Temp. Sequencing
ERCC6L	2uM	57,8	82,6	57,8	57,8
LRCH2	20uM	62,6	77,6	60,0	60,0
OR13H1	2uM	62,6	79,0	60,0	60,0
GPR112	2uM	57,8	77,6	57,8	57,8
PLXNA3	10uM	62,6	85,6	60,0	60,0
F8	5uM	57,8	77,0	57,8	57,8

Primers	Lower Normalisation Bar	Upper Normalisation Bar	Reference Cluster
LRCH2	74,7–76,2	79,0–80,5	A/G

PCR Cycling Protocol - ERCC6L,GPR112,F8			
Pre-PCR	95°C	2:00	
Denature	95°C	0:10	40 Cycles
Anneal	57,8°C	0:30	
Extension	72°C	0:30	
Melt Curve	95°C	0:30	
	60°C	1:00	
	72°C to 87,6°C 0,2°C increments	0:10/step	
Hold	4°C	∞	

PCR Cycling Protocol - LRCH2,OR13H1,PLXNA3			
Pre-PCR	95°C	2:00	
Denature	95°C	0:10	40 Cycles
Anneal	62,6°C	0:30	
Extension	72°C	0:30	
Melt Curve	95°C	0:30	
	60°C	1:00	
	72,6°C to 90,6°C 0,2°C increments	0:10/step	
Hold	4°C	∞	