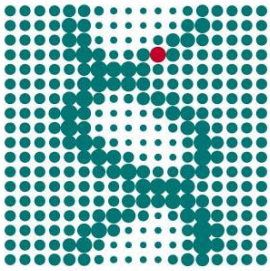


Max-Planck Institute for Molecular Genetics

Freie Universität Berlin

Identification of three novel genes for autosomal recessive intellectual disability and molecular characterization of the causative defects

DISSERTATION

Zur Erlangung des akademischen grades

Doctor rerum naturalium

(Dr. rer. nat.)

Vorgelegt von

Lia Abbasi Moheb

aus Kermanshah, Iran

Eingereicht im Fachbereich

Biologie, Chemie, Pharmazie

der

Freien Universität Berlin

Date: 08.02.2012

Diese Dissertation wurde in der Zeit vom Februar 2006 bis July 2011 am Max-Planck Institut für molekulare Genetik in Berlin, in der Abteilung von Herrn Prof. Dr. Hans-Hilger Ropers, Arbeitsgruppe Dr. Andreas W. Kuss angefertigt.

This thesis has been conducted from February 2006 till July 2011 at the Max-Planck Institute for Molecular Genetics in Berlin in the department of Prof. Dr. Hans-Hilger Ropers, Research group Familial cognitive disorders (Dr. Andreas W. Kuss)

1. Gutachter: **Prof. Dr. Hans-Hilger Ropers**
Max Planck Institut für molekulare Genetik
Ihnestr. 73, D-14195 Berlin

First Referee: **Prof. Dr. Hans-Hilger Ropers**
Max Planck Institute for molecular Genetics
Ihnestr. 73, D-14195 Berlin

2. Gutachter: **Prof. Dr. Stephan Sigrist**
Freie Universität
Takustr. 6, D-14195 Berlin

Second Referee: **Prof. Dr. Stephan Sigrist**
Free University
Takustr. 6, D-14195 Berlin

Tag der Disputation: February 08th, 2012
Day of the Disputation: February 08th, 2012

I hereby declare that the work presented in this thesis has been conducted independently and without any inappropriate support and that all sources of information, be it experimental or intellectual, are aptly referenced.

I hereby declare that this thesis has not been submitted, either in the same or a different form, to this or any other university for a degree.

Lia Abbasi Moheb
Berlin, February 2012

Contents

1 Introduction

1.1 Intelligence

Intelligence is a very general capability that, among other things, involves the ability to reason, plan, solve problems, think abstractly, comprehend complex ideas, learn quickly and learn from experience. It is not just about understanding a book or having a specific skill. By this standard, intelligence is the essence and the object of individual differences among humans [Gottfredson, 1997].
Intelligence is determined using psychometric tests and compared to a statistical norm. An intelligence quotient of 100 denotes average intelligence, which is thought to be normally distributed, with a standard deviation (SD) of 15. Individuals with an IQ of 70 (=100 minus 2 SD) are considered mentally retarded, or intellectually disabled, recently recommended as an appropriate term by the American Association on Intellectual and Developmental Disabilities (AAIDD Committe).

1.2 Intellectual disability

Based on the American Association of Mental Retardation (AAMR) 1992 definition, mental retardation, entails " significantly sub average intellectual functioning, existing concurrently with related limitations in two or more adaptive skill areas, manifesting before age 18." The definition does not provide specific IQ cutoffs. In 2002, the AAMR revised its definition to include the above but also to emphasize "practical adaptive skills" and a "multidimensional and ecological approach that reflects the interaction of the individual with the environment, and the outcomes of that interaction with regards to independence, relationships, and social contributions, participation in school and community, and personal well-being." This definition appears to incorporate social and cultural influences in the individual's functioning [Frumkin B., 2003]. Mild forms of intellectual disability are defined by an IQ between 50 and 70 and are thought to represent the lower end of the normal IQ distribution and might be the result of the interaction of many genes and environmental factors. In contrast, severe forms with an IQ <50 which have an incidence of about 0.4% might be caused by catastrophic events such as prenatal hypoxia or, more often, specific genetic factors such as chromosomal aberrations and defects in specific genes [Ropers, 2006]. Numerous studies distinguish only between mild (IQ 70-50) and severe ID (IQ <50) [Ropers and Hamel, 2005]. It is estimated that prevalence of severe and mild forms of ID in the population is approximately 1-3% [Roeleveld N. et al., 1997; Leonard and Wen, 2002; Ropers HH., 2006]. Assuming a normal distribution with an SD of 15, 2.3% of the population would be expected to have an IQ that is below 70.

To describe the severity of the ID, a more fine-grained classification of the World Health Organization (WHO) is also used:

Class	**IQ**
Profound	below 20
Severe	20-34
Moderate	35-49
Mild	50-69
Borderline deficiency	70-79

ID can also be subdivided into syndromic and non-syndromic forms. In the former, ID is accompanied by other clinical features, e.g., malformations, or dysmorphic features, neurological abnormalities, whereas patients with non-syndromic ID do not have any other abnormality. However, distinguishing between syndromic and non-syndromic forms of ID is often difficult, particularly when based on the examination of a single patient or young children [Kaufman L, 2010; Ropers HH., 2006].
Intellectual disability is one of the most important problems of health care. ID is four times more common than rheumatic heart disease and nine times more prevalent than cerebral palsy. It affects 15 times as many people as total blindness and 10 times as many children and adults as polio did before the Salk vaccine [Pediatrics for Parents, 1991].
Moreover, intellectual disability is a, or the only, clinical feature in 1715 of the known Mendelian disorders listed in OMIM (http://www.ncbi.nlm.nih.gov/Omim/searchomim).

1.3 Causes of intellectual disability

Any type of condition which affects brain development before birth, during birth or in early childhood years can give rise to intellectual disability. Hence, the underlying causes of ID are extremely heterogeneous.
ID can be caused by environmental or genetic factors and, in the majority of patients, the etiology of ID is still unknown. Finding the underlying causes of ID is essential for prognosis, management, and genetic counselling.
ID may have environmental (e.g., malnutrition during pregnancy, environmental neurotoxicity, premature birth, perinatal brain ischemia, fetal alcohol syndrome and pre or post natal infections), chromosomal (e.g., aneuploidies and microdeletion syndromes) or monogenic causes [Chelly J et al., 2006; Ropers and Hamel, 2005]. So far, genetic causes have been identified in up to 40% of cases [Ropers HH. et al., 2008; Chelly J et al., 2001].
Chromosomal causes include well known disorders such as Down syndrome (trisomy 21), which is the most common genetic form of ID [Rauch et al., 2006].
The diagnosis of small chromosomal rearrangements has been greatly facilitated by array CGH and related techniques [Pinkel D. et al., 1998; Ishkanian AS. et al., 2004]. Array CGH, is able to detect unbalanced rearrangement as small as 10 kb [Stankiewicz p. et al., 2007]. Small chromosomal deletions and duplications have been instrumental in the identification of numerous ID genes, including *SHANK2*, *SYNGAP1* and *HUWE1* [Pinto A. et al., 2010; Berkel S. et al., 2010; Froyen G. et al., 2008].

1.4 X-linked intellectual disability (XLID)

L.S. Penrose was the first to report that intellectual disability is more common in males than in females [Penrose LS., 1938]. The following studies could show the similar preponderance to recessive defects on the X chromosome. Males carrying the mutation on chromosome X will be affected, whereas in females, the normal gene copy on the second X chromosome can compensate for a mutation in an X-linked gene [Plenge RM., 2002; Amos-Landgraf, 2006]. Because of their characteristic inheritance patterns, XLID families can be easily identified, which is why XLID has been disproportionately well studied in the past.
To date, more than 90 genes have been identified in XLID and 80% of the 40 genes causing NS-ID are located on the X-chromosome [Kaufman L. et al., 2010].
For a long time, the identification of new IDX genes seemed to be an almost impossible task, given the genetic heterogeneity of this condition. However, in spite of this difficulty, the progress in genome analysis and the establishment of large collaborations between clinical and molecular research teams have led to great progress. In the syndromic types of intellectual disability, Fragile X syndrome (FRAXA) (OMIM 309550) is the most frequent type of XLMR and the best characterized. Of the total number of individuals affected with XLMR, Fragile X syndrome accounts for about 15 to 20% of cases [Chelly J et al., 2001; Ropers HH., 2006].
In addition to ID feature, these patients have specific features (long face, large ears, a high arched palate, flat feet, and macroorchidism) and behavioral abnormalities (hyperactivity, avoidance of eye contact, and repetitive speech as well as autistic features) [Hagerman RJ. et al., 2002]. At the molecular level, the disorder is due to a dynamic mutation caused by expansion of a CGG repeat located in the promoter region at 5' end of FMR1 gene [Verkerk A. et al., 1991]. The expansion of the repeat above 200 CGG results in hypermethylation of the promoter region, aberrant heterochromatinization and silencing of FMR1 gene resulting in the absence of the gene product (FMRP)[Nussbaum RL. et al., 1995; Sutcliffe JS. et al., 1992]. Premutation alleles (55-200 CGG repeats) of the fragile X mental retardation 1 (FMR1) gene are associated with autism disorder in childhood, premature ovarian failure, and the neurodegenerative disorder, fragile X-associated tremor/ataxia syndrome (FXTAS).
However, it should be noted that several genes, such as *MECP2*, *ATRX*, *SLC6A8*, *RSK2*, *OPHN1*, *ARX*, *PQBP1*, *MTC8*, *AP1S2*, *FGDY* and *JARID1C*, which were initially identified as causative genes for S-XlID, were also found to be mutated in patients with non-syndromic intellectual disability [Lisik MZ. et al., 2008]. In part, these different phenotypes may be due to allelic differences, i.e., different mutations involving the same gene, but allelic mutations cannot explain the wide clinical variability observed within many families. Here, other modulating factors must be involved.

1.5 Molecular and cellular mechanisms underlying XLID

The functional spectrum of XLID genes is wide. Many are related to signal transduction (19%) and regulation of transcription (22%); others are involved in known methabolic pathways (15%), DNA and RNA processing (6%), protein synthesis (3%), regulation of cell cycle and in protein degradation (7%) [Chiurazzi P. et al., 2008].
Defects of these fundamental processes may disproportionately affect cognition, because the

brain is particularly vulnerable, or some of the relevant genes are highly expressed in the CNS. Some XLID genes are expressed in the brain but not in neurons; for example, PLP1 encodes the proteolipid protein 1, a major component of myelin that is expressed exclusively in oligodendrocytes [Karim SA. et al., 2007].
For most of the XLID genes, so far, knowledge of the relevant pathogenic mechanisms is very limited, but for some, common pathways have been identified. Several genes have been implicated in the regulation of the actin cytoskeleton through Rho-GTPases signaling, in synaptic vesicle transport, chromatin remodeling and in the gene regulation and expression.

1.5.1 Regulation of the actin cytoskeleton through Rho-GTPases signaling

The cytoskeleton of neurons comprises three types of filamentous structures: actin microfilaments, microtubules and neurofilaments. The actin cytoskeleton is essential for morphologic differentiation, including development of specialized dendritic morphology, neurite outgrowth, establishment of cell polarity, synapse formation, synaptic plasticity and protein transport [Chechlacz et al., 2003]. In differentiated neurons, the actin cytoskeleton has a central role in dendritic plasticity, a process which is essential for learning, memory and cognition [van Rossum et al., 1999]. The members of Rho family of GTPases are small GTP-binding proteins which have a key role in signalling pathways, controlling the organization of actin cytoskeleton.So far several genes have been identified in the regulation of Rho GTPases. *OPHN1*, *PAK3*, *ARHGEF6*, *FMR1* and *FGD1* have all been implicated in XLID [Kaufman et al., 2010].
For example, *OPHN1* encodes an activator of the Rho, Rac1 and CDC42 [Billuart et al., 1998]. PAK3 is a serin thereonine protein kinase acting as a downstream effector of Rac1 and CDC42 [Allen et al., 1998]. ARHGEF6 is a Guanine nucleotide exchange factor (GEF) for Rac and CDC42 [Kaufman et al., 2010]. FGD1 is a RhoGEF and has a possible role in the stimulation of neurite outgrowth [Label et al., 2002].

1.5.2 Synaptic vesicle transport

Synaptic vesicles are located at the pre-synaptic terminal. After membrane depolarization, synaptic vesicles fuse with the pre-synaptic membrane and release neurotransmitters in the synaptic cleft. The vesicles are then recycled and filled again with neurotransmitter molecules. The correct regulation of this process is essential for correct synapse functionality and thus normal brain functioning. This process is regulated by Rab proteins, another family of GTPases [Novick et al., 1997].

1.5.3 regulation of gene expression

Gene expression is another important process that can be altered in XLID patients. Selective expression of certain genes and inhibition of others is necessary for the correct differentiation and functioning of all cell types, including neurons. Gene expression can be regulated either by the modulation of chromatin structure or by the regulation of activity of the molecules involved in the different steps of the process which lead from the gene to the mature functional protein.

1.5.4 Epigenetic regulation

The promoters of inactive genes are usually hypermethylated, while expressed genes have an open chromatin structure with unmethylated promoters. Hypoacetylation and hypermethylation can be stably maintained through mitosis and can thus be passed on to daughter cells. Mutations in genes belonging to this pathway suggest that ID could arise from an alteration of epigenetic mechanisms regulating gene expression and silencing.
MECP2 and *ATRX* are genes that function in epigenetic regulation. Mutations in both genes cause variable phenotypes and it ranges from NS-ID to Rett syndrome in the case of *MECP2* mutations and alpha thalassemia/ID syndrome caused by mutations in ATRX.
MECP2 gene (Methyl-CpG-binding Protein 2, OMIM #300005) encodes a broadly expressed nuclear protein that was originally characterized as a transcriptional repressor [D'Esposito M., et al., 1996]. MeCP2 is a member of the methyl-CpG binding protein family that has two conserved functional domains: the methyl-CpG binding domain (MBD) and the transcription repression domain (TRD) [Hendrich et al., 1998]. MeCP2 binds to methylated DNA through the MBD and effects gene silencing by imparting changes in chromatin structure via the interaction with the corepressor Sin3A and the histone deacetylase complex [Jones et al., 1998]. MeCP2 also associates with the corepressors c-Ski and N-CoR via the TRD and forms complexes with these repressors independently of its interaction with Sin3A [Kokura et al., 2001].
Tao et al. in 2004 could show that the truncation of the X-linked cyclin-dependent kinase-like 5 (*CDKL5/STK9*) gene causes intellectual disability. In 2005, mutations in *CDKL5* were identified in patients with the onset of an early seizure variant of RTT syndrome [Scala et al., 2005; Sprovieri et al., 2009]. *CDKL5* (OMIM #300203; also known as serine threonine kinase 9 (STK9)) is located in Xp22 and belongs to the serine-threonine kinase family, which shares homology with members of the mitogen-activated protein and cyclin dependent kinase (CDK) families [Montini et al., 1998]. The observation that mutations in *MECP2* and *CDKL5* cause similar phenotypes suggested that these genes may be involved in the same molecular pathway. It could be shown that the two genes have an overlapping temporal and spatial expression profile during neuronal maturation and synaptogenesis [Mari et al., 2005].

1.5.5 Chromatin remodelling

Inactive genomic regions are usually characterized by a condensed chromatin structure enriched in hypoacetylated forms of H3 and H4 histones. JARID1C is a zinc finger protein containing a PHD-finger domain. Several different mutations in *JARID1C* have been identified in families with NS-XLID. This gene contains several DNA-binding motifs which link it to transcriptional regulation and chromatin remodelling, processes that are defective in various other forms of Intellectual disability [Jensen et al., 2005].
Two other genes, *MBD5* and *BRWD3*, are also involved in chromatin maintenance. Mutations in these genes result in both S-ID and NS-ID [Kaufman et al., 2010]. Based on motif structure, BRWD3 is a chromatin modifying protein [Field et al., 2007]. It contains a bromodomain which is found in chromatin associated proteins and histone acetyltransferases. MBD5 is a methyl-CpG binding protein and its mutations are found in the autosomal dominant form of ID.

1.5.6 Transcriptional regulators

The Aristaless-related homeobox gene (*ARX*) is one of the most frequently mutated genes in a spectrum of X-chromosome phenotypes with intellectual disability (ID). At least 10 well-defined clinical entities, including Ohtahara, Partington, and Proud syndromes, X-linked infantile spasms, X-linked lissencephaly with ambiguous genitalia, X-linked myoclonic epilepsy and nonsyndromic intellectual disability, have been ascertained from among the patients with *ARX* mutations. There seems to be a genotype/phenotype correlation between the location and nature of the mutation and the severity of the phenotype [Kaufman et al., 2010].
NF-kB transcription factor regulates the expression of many neuronal genes. CC2D1A and TRAPPC9 are involved in the activation of NF-kB and affect long-term memory formation. *PQBP1* and *FMR2* are the other genes that regulate gene expression. PQBP1 is a poly glutamine binding protein that interacts with components of the spliceosome and also acts as a transcriptional repressor. Mutations in the *PQBP1* gene have been found in association with both non-syndromic and syndromic XLID.

1.6 Autosomal intellectual disability

1.6.1 Autosomal dominant intellectual disability (ADID)

With the assumption that 8-10% of moderate-to-severe forms of ID are X linked, the majority of the gene defects underlying ID must be autosomal [Ropers, 2007]. Severe dominant forms of ID are rarely familial because affected individuals do not reproduce. So far, only a few genes have been found for ADID, including *SYNGAP1*, *STXBP1*, *FOXG1* and *MEF2C*.
FOXG1 belongs to the forkhead family of transcription factors which is characterized by a distinct forkhead domain. This gene has an important role in the formation of the developing brain where it encodes a transcriptional repressor protein. The expression of *FOXG1* is restricted to brain and testis. Congenital Rett syndrome can be caused by copy-number variation in *FOXG1* [Ariani et al., 2008; Jacob FD, 2009], which expands the clinical phenotypic spectrum of *FOXG1* defect in humans.
A de novo heterozygous nonsense mutation has been identified in the *MEF2C* gene (S228X) in a patient with severe mental retardation, stereotypic movements, epilepsy, and cerebral malformation. MEF2C is involved in synaptic plasticity, learning and memory [Le Meur et al., 2010].
So far, numerous candidate genes for ADID have been identified by micro deletion screening. A recent addition to this list is *MBD5*, which was found inside of a 200 kb de novo deletion, detected by SNP microarray. *DOC8*, *CDH15* and *KIRRL3* genes were indeed recently identified by breakpoints mapping and translocations. These methods, along with candidate gene sequencing, were found to be particularly productive for the identification of ADID [Kaufman et al., 2010].

1.6.2 Autosomal recessive intellectual disability (ARID)

Functional considerations and epidemiological data suggest that the majority of the gene defects that give rise to the disease will be inherited as recessive traits. Autosomal recessive forms

of ID (ARID) may be due to mutations in hundreds if not thousands of different genes [Ropers, 2007]. Judging from the relative frequencies of syndromic and non-syndromic X-linked ID, non-syndromic forms of ARID are likely to be more common than syndromic forms.
Compared to the significant progress in XLID research, the knowledge about the molecular basis of NS-ARID is still very limited. Only six published genes causing autosomal recessive non-syndromic ID have been identified: *PRSS12* on chromosome 4q26, *CRBN* on 3p26, *CC2D1A* on 19p13, *GRIK2* on 6q16, *TRAPPC9* on chromosome 8q24.3 and *TUSC3* on 8p22 [Molinari F, 2002; Higgins JJ, 2004; Basel-Vanagaite L, 2006; Motazacker MM, 2007; Garshasbi M., 2008]. *ST3GAL3*, *ZC3H14* and *ZNF526* are still unpublished. These genes have been found in more than one family [Moheb et al., Abstract 2010; Pak et al., 2011]. Mutations in *PRSS12*, *CRBN* and *CC2D1A* genes cause a similar degree of severity of ID in all the affected members of the same family. In contrast, mutations in TUSC3 can be associated with severe ID, and the severity can vary even within the same family. Interestingly, all mutations that have been identified in these genes are protein-truncating mutations. It could be hypothisized that milder missense mutations or sequence variants in these and other genes might cause an additive effect in the pathogenesis of mild cognitive impairment. Although all these molecules are expressed in the brain, their neuronal functions are not yet completely clear.
PRSS12;
This protein is an extracellular multidomain serine protease associated with neural development and plasticity. The *PRSS12* gene contains 13 exons and is mapped to human chromosome 4q25-q26. This gene encodes a deduced 875-amino acid protein and contains a putative signal peptide, a kringle domain, 4 repeated scavenger receptor cysteine-rich (SRCR) regions, and a C-terminal serine protease domain. Recently, it has been shown that this protein specifically cleaves agrin, which is involved in the formation of filopodia on neuronal axons and dendrites. These and recent studies on the expression of neurotrypsin in live hippocampal neurons argue for an essential role of this protein in activity-dependent synapse remodelling [Ropers, 2008].
CRBN;
The protein has 442-amino acid with a 237-residue ATP-dependent protease domain and several phosphorylation sites. *CRBN* gene contains 11 exons and is identified on chromosome 3pter-p25. It is highly expressed in the brain and because of its role in cerebral development was named 'cereblon'. Higgins and his colleagues have found that the homozygous C > T nonsense mutation at nucleotide position 1,274 of a novel cDNA (1274 C > T) is involved in NS-ARID in a large family originating from Germany with 10 affected individuals. So the nonsense mutation, R419X, leading a premature stop codon in CRBN, interrupts an N-myristoylation site and eliminates a casein kinase II phosphorylation site at the C terminus [Wang Xin et al., 2008; Joseph JH et al., 2004]. Cereblon was identified as being directly associated with large conductance $Ca2^{+}$ activated K1 channels [Rotanova, 2006], which are important in the control of neuronal excitability and transmitter release [Faber et al., 2003]. It was recently stated that BKCa channel overexpression causes impairment of learning and memory in hippocampal-dependent tasks, but does not alter basal synaptic transmission or pre-synaptic release mechanisms [Hammond et al., 2006]. Therefore, assembly and surface expression of functional BKCa channels might be of importance in controlling human cognition.

CC2D1A;

In 9 consanguineous Israeli-Arab families with nonsyndromic intellectual disability from the same village, a defect in the *CC2D1A* gene was identified by Basel-Vanagaite and his coulegues (2006). He analyzed 14 candidate genes located in a haplotype-defined critical region on chromosome 19p13.12. The G408fsX437 mutation, a large genomic deletion of 3589 nucleotides, creating a truncated protein, was identified in all affected family members; parents were heterozygous for the mutation.

All of the affected were severely mentally retarded; none had autistic features or seizures, and there were no dysmorphic features. *CC2D1A* mRNA is mostly expressed in the embryonic ventricular area of the brain. It encodes an evolutionarily conserved protein that contains four DM14 domains at the N terminus and a C2 domain at the C terminus. CC2D1A is a potent activator of NF-κB. The activation of NF-κB by CC2D1A requires its C2 domain. CC2D1A activates NF-κB in a manner that depends on the ubiquitin-conjugating enzyme Ubc13, TNF receptor-associated factor TRAF2, the protein kinase TAK1, and the I-B kinase (IKK) complex. Because the NF-κB pathway is important for neural plasticity and memory, this finding may also explain why CC2D1A is indispensable for normal brain function in humans. The mutation is a deletion of 3567 nucleotides from introns 13 to16, abolishing the fourth DM14 domain and C2 domain.

Freud-1 (five' repressor element under dual repression binding protein-1) protein is the rat homologue of human CC2D1A, and negatively regulates basal 5-HT1A receptor expression in neurons via binding to the repressor element of the 5-HT1A receptor gene [Ou et al., 2003]. Freud-1 has been shown to function as a transcriptional repressor of the serotonin-1A receptor gene that binds to a novel DNA element. The DNA binding and repressor activities of Freud-1 are inhibited by calcium-calmodulin-dependent protein kinase.

GRIK2;

In an Iranian family with autosomal recessive intellectual disability (MRT6; 611092), Motazacker et al. (2007) identified a deletion removing exons 7 and 8 of the *GRIK2* gene. Loss of these exons resulted in an in-frame deletion of 84 amino acids between residues 317 and 402, close to the first ligand-binding domain (S1) in the extracellular N-terminal region of the protein. Functional studies demonstrated complete loss of function of the mutant GRIK2 protein. Further studies to elucidate the full extent of the observed mutation showed that, in addition to the 120-kb deletion removing exons 7 and 8, the mutation comprised an inversion of approximately 80 kb, including exons 9, 10, and 11, in combination with a deletion of approximately 20 kb of intron 11. Motazacker et al. (2007) predicted that at the protein level this mutation could be expected to result in the loss not only of the first ligand-binding domain but also of the adjacent transmembrane domain and the putative pore loop of GRIK2.

Glutamate receptors mediate most excitatory neurotransmission in the brain.

The predicted gene product lacks the first ligand-binding domain, the adjacent transmembrane domain, and the putative pore loop, suggesting a complete loss of function of the GLUK6 protein, which is supported by electrophysiological data. This finding provides the first proof that GLUK6 is indispensable for higher brain functions in humans, and future studies of this and other ionotropic kainate receptors will shed more light on the pathophysiology of intellectual disability [Motazacker MM et al., 2007].

TUSC3;
TUSC3 is the fifth gene that has been implicated in nonsyndromic ARID (NS-ARID). In affected members of a large consanguineous Iranian family with nonsyndromic intellectual disability, Garshasbi et al. (2008) found a homozygous 120 to 150 kb deletion on chromosome 8p22, including the first exon of the *TUSC3* gene. *TUSC3* has 11 exons spanning 224 Kbp of the genomic DNA on chromosome 8p22. According to the UniProt database, *TUSC3* encodes a predicted 348 amino acid protein with five potential transmembrane domains and seems to be involved in catalyzing the transfer of a 14 sugar oligosaccharide from dolichol to nascent protein. This reaction is the central step in the N-linked protein glycosylation pathway.
TUSC3 is assumed to encode a subunit of the ER-bound oligosaccharyltransferase (OST) complex that catalyzes a pivotal step in the protein N-glycosylation process.
As for the role of this gene in the brain, it is noteworthy that TUSC3 interacts with the alpha isoform of the catalytic subunit of protein phosphatase 1 (PPPC1A; MIM 176875)[Rual et al., 2005]. Protein phosphatase 1 has been implicated in the modulation of synaptic and structural plasticity [Munton et al., 2004] and was shown to have an impact on learning and memory in mice [Genoux et al., 2002]. It is therefore conceivable that ID in *TUSC3*-deficient patients is caused by an impairment of PPPC1A function. Zhou and Clapham (2009) hypothesized that MAGT1 and TUSC3 may function cooperatively in mediating cellular Mg^{2+} uptake.

1.7 The aim of this study

A particular straightforward strategy for the identification of genes underlying autosomal recessive disorders is homozygosity mapping in extended consanguineous families, followed by mutation screening of candidate genes.
In western civilizations, most patients with ARID are sporadic cases, due to small family sizes and low rates of parental consanguinity, and pedigrees with several affected sibs are rare, which has precluded linkage analysis and gene mapping. In contrast, up to 40% of all children are born to consanguineous parents in Iran, and large families are very common. Therefore, the structure of the Iranian population is ideal for homozygosity mapping of autosomal recessive traits. In 2003, this has prompted our group and that of H. Najmabadi (Social Welfare and Rehabilitation Sciences University, Tehran, Iran) to embark on a collaborative project aiming at the systematic identification of genes that have a role in ARID.
This collaboration revealed numerous novel loci for ARID and showed that ARID is an extremely heterogeneous disorder [Najmabadi et al., 2007]. The first 8 NS-ARID loci that were found [Najmabadi et al., 2006] were named as 'Mental retardation 4 to 11' (MRT4-11), in accordance with the nomenclature used for previously mapped NS-ARID loci (OMIM #249500, #607417, #608443). Since then, we have identified 20 additional solitary intervals for NS-ARID and 6 for syndromic forms of ARID. For some of these loci, several families with overlapping linkage intervals were found, which may indicate that, at least in the Iranian population, not all of the gene defects causing NS-ARID are extreamly rare.
However, some of these defects can be nearly as common as FMR1, which is found in 20% of the families with XLID [Kuss et al., 2010]. ID is a disabling lifelong condition with so far limited therapeutic prospects, which puts a heavy psychological as well as financial burden not only

on the affected families but also on society as a whole. Although autosomal recessive forms of ID may be responsible for a large proportion of ID cases [Ropers, 2007] only 6 genes have been implicated in NS-ARID to date. Therefore, the first objective of this project was to identify novel genes for this disorder and to study their function.
Thus, apart from providing the basis for reliable molecular diagnostics and prevention, these studies will also deepen our understanding of the underlying pathogenesis mechanisms for any form of therapy.

2 Materials and Methods

2.1 Patients, Materials and Methods

Chemical	Manufacturer
[α-32P]dCTP	Amersham Biosciences
Acrylamide (Molecular biology grade)	Sigma
Agarose	Invitrogen
Ammonium persulfate	Sigma
Ampicillin	Sigma
Aqua ad inectabilia	Baxter
Betaine	Sigma
Bradford reagent	Sigma
Bromophenol Blue	Sigma
BSA	Sigma
Chloroform	Merck
Complete, Mini Protease Inhibitor Cocktail Tablets	Roche
Diethylpyrocarbonate (DEPC)	Aldrich
DMSO	Sigma
dNTPs	Roth
DTT	Promega
EDTA	Merck
Ethanol	Merck
Ethidium bromide	Serva
First strand buffer 5x	Merk
Formaldehyde-37.0% (v/v)	Fluka Biochemika
Formamide	Fluka Biochemika
Glycerol	Roth
Glycin	Merck
HEPES	Calbiochem
Hydrogen Chloride	Merck
Isopropanol	Merck

Chemical	Manufacturer
Magnesium chloride	Merck
Methanol	Merck
Milk powder	Protifar
Oligofectamine	Invitrogen
OptiMEM	Invitrogen
PdN6	Pharmacia
SDS	Roth
Sodium Acetate	Sigma-Aldrich
Sodium Chloride	Roth
Sodium Hydroxide	Merck
Sodium Hydroxide	Sigma
Sodium perchlorate	Merck
TEMED	Gibco BRL
TRIzol reagent	Gibco BRL
Trypsin EDTA (500mg/ml Trypsin, 200mg/ml EDTA)	Cambrex
Whatman paper	Sigma
X-Gal	Appligene
β-mercaptoethanol	Whatman
Acetic acid	Merck
Tween 20	Invitrogen

2.1.1 Enzymes

Enzyme	Manufacturer
DNA polymerase 1, Klenow fragment	**USB**
Gateway LR Clonase Enzyme Mix	**Invitrogen**
MseI	**Biolabs**
PowerScript TM Reverse Transcriptase	**Clontech**
Proteinase K	**Fermentas**
Rnase-Free DNase	**Promega**
RNAsin	**Promega**
SuperTaq TM Plus	**Ambion**

2.1.2 Kit and Markers

Name	Supplier
0.24-9.5 Kb RNA ladder	Invitrogen
1 kb DNA ladder	Roth
Advantage 2 PCR Kit	Clontech
BigDye Terminatormix	Applied Biosystems
Bio-X-ACT (Bioline) PCR Kit	Bioline
Dithiothreitol (DTT)	Invitrogen
Dynabeads Oligo (dT)25 kit	DYNAL Biotech
Expand Long Template PCR system kit	Roche
Human Fetal Brain Total RNA	BD Bioscience
Hyper Ladder I	Bioline
Hyper ladder IV	Bioline
Illumina GoldenGate Genotyping Assay	Illumina
Illumina TotalPrep RNA Amplification Kit	Ambion
Lambda DNA/HindIII marker	Fermentas
Micro Spin G-50 column	Amersham
MicroSpin TM G-25 Columns	Amersham
MiniElute PCR purification kit	Qiagen
Oligo(dT)20 primer	Invitrogen
pUC Mix marker, 8	Fermentas
QIAquick Gel Extraction Kit	QIAGEN
QIAquick PCR Purification Kit	Qiagen
Random Primers	Promega
RNeasy Mini Kit	Qiagen
SDS-PAGE protein marker - High range	Sigma
Sentrix Human-6 Expression BeadChips	Illumina
Superscript ™ II Reverse Transcriptase	Invitrogen
Superscript ™ III Reverse Transcriptase	Invitrogen
Taq PCR core kit	Qiagen
Western Lightning Chemiluminescence Reagent, NL100	PerkinElmer

2.1.3 Instruments

Instrument	Manufacturer
B 5050 E incubator	Heraeus
Capillary Sequencer ABI 377	Applied Biosystems
Centrifuge 5810R	Eppendorf
Centrifuge Rotanta 46R/Rotina 4R	Hettich zentrifugen
Centrifuge Rotina 48R	Hettich zentrifugen
Clean bench Herasafe	Heraeus
CO2 water jacketed incubator	Forma Scientific
Concentrator 5301	Eppendorf
Control environment incubator shaker	New Brunswick Scientific
E.A.S.Y. 440K Gel Documentation System	Herolab
Electrophoresis power supply 2	Heathkit
Geiger Counter, Series 900 mini-monitor	Artisan Electronics Corp.
Horizontal gel apparatus Horizon® 11.14 and 20.25	Life technologies
HyperCassette BioMAX (Northern blot)	Amersham
Inverted light microscope, Eclipse TS100	Nikon
L8-70M ultracentrifuge	Beckmann
Laminar flow hood, CA/REV 6 Cleanbench	Clean Air
Mini-Gel apparatus	Bio-Rad
Multichannel pipette	Rainin

Instrument	**Manufacturer**
Phase lock gel light	Eppendorf
pH-meter	Knick
Pipett boy	Integra biosciences
Pipettes	Gilson
Power Pac 300 electrophoresis power supply	Bio-Rad
PTC-225 Tetrad and Dyad thermal cycler	Bio-Rad
REAX 2000 vortexer	Heidolph
Rnase ZapWipes	Ambion
Rotating mini hybridization oven	Appligene
Rotors TLA120.1, TLS-55, SW40	Beckmann
Scanner, Expression 1680 Pro	Epson
Sonifier cell disruptor B-30	Branson Sonic Power
Sorvall RC-5B refrigerated super speed centrifuge	Du Pont instrument
SPD 111V Speed Vac	Savant
Spectrophotometer NanoDrop ND-1000	PEQLAB
Steri-cycle CO2 incubator 371	Thermo Electron Corp.
Table centrifuge 5415C	Eppendorf
ThermoForma 758 Ultrafreezer	Thermo Electron Corp.
Thermomixer 5436	Eppendorf
TL100 ultracentrifuge	Beckmann
UV stratalinker 1800	Stratagene
UV trasilluminator	UVPinc
Western blot cassette (HyperCassette)	Amersham
Western blot Trans Blot SD	Bio-Rad
X-ray film developing machine, Curix 60	Agfa

2.1.4 Consumables

Consumable (disposable material)	**Supplier**
Adhesive PCR film	Abgene
Biomax MS X-ray film (sensitive)	Kodak
Cell culture flask (25, 75 & 100 cm2)	TTP
Cell scraper	TTP
Chromatography paper	Whatman
Disposable reaction tube 14 ml	Greiner BioOne
Disposable reaction tube 30 ml	Sarstedt
Falcon tube	Greiner BioOne
Glass coverslip	Menzel-Gläser
Hamilton syringe	Hamilton
Hybond-XL (Northern blot membrane)	Amersham
Immobilon-P transfer membrane (Western blot membrane)	Millipore
MS X-ray film	Kodak
Parafilm	Pechiney Plastic Packaging
Pasteur pipette	Roth
PCR plate (96 well)	Abgene
Pipette tip (0.1 – 10, 1-20, 20 – 200 & 1000 µl)	Biozyme
Reaction tube (1.5 & 2 ml)	Eppendorf
Scalpel	Aesculap
Serological pipette (2, 5, 10 & 25 ml)	Corning

2.1.5 Buffers and Media

Buffer/Medium	Composition
APS 10%	10% w/v APS in water, aliqouted and stored at –20°C
Blocking buffer	5% milk powder in PBST
DEPC H2O	0,1% DEPC was overnight stirred in water and autoclaved afterwards.
DMEM	From Cambrex + 10%FCS + 100U/ml Penicillin + 100µg/ml Streptomycin + 2mM L-Glutamine
DNA re-suspension buffer	0,4M Tris-HCl PH=8, 0.06M NaEDTA, 0,15M NaCl
DNA-Loading buffer	15% Ficoll, 0,25% Bromphenolblue in bidest H2O
Ethidium bromide	10 mg/ml EtBr in bidest H2O
Fetal Calf Serum (FCS)	Sigma
First strand buffer 5x	250 mM Tris-HCl (pH 8.3 at room temperature), 375 mM KCl, 15 mM MgCl2
Inverse PCR ligation buffer	50mM Tris-Hcl pH=7.4, 10mM MgCl2, 10mM DTT, 1mM ATP, 10% gelatin
Laemmli protein loading buffer (5X)	Aqueous solution containing 62.5 mM Tris HCl (pH 6.8), 5% beta-mercaptoethanol (v/v), 50% Glycerol (v/v), 2% SDS (w/v), 0.1% (w/v) Bromo phenol Blue.
Lymphoblastoid cell lines	RPMI-1640 containing 10.0% (v/v) FBS; 100 U/ml Penicillin; 68.6 µM Streptomycin; 2.00 mM L-Glu.
MOPS 10 x	0,4 M MOPS; 0,1 M NaAc; 10 mM EDTA pH 7,0
PBS 1 x buffer	137 mM NaCl; 2,7 mM KCl; 10,1 mM Na2HPO4; 1,8 mM KH2PO4
PBST buffer	1 x PBS; 1:1000 Tween 20
SDS-PAGE running buffer (1X)	196mM glycine, 0.1% SDS, 50mM Tris-HCl (pH 8.3)
Western blot Separating gel buffer	1,5 M Tris-HCl, 0,4 % SDS pH 8,8
Southern blot denaturing buffer	1.5M NaCl + 0.5M NaOH
Southern blot wash buffer	40mM Na3PO4 + 0.5% SDS
SSC 10x	3M NaCl, 0,3M Na-citrate in bidest H2O, adjust pH 7 with 1M HCl
Stacking gel buffer	0,5 M Tris-HCl, 0,4 % SDS pH 6,8
Stripping buffer	1% SDS, 20mM TRIS/HCl (pH 6.8), 1% (v/v) β-Mercaptoethanol
TAE buffer 50 x	50 mM EDTA, 5,71% v/v acetic acid, 2M Tris-HCl
TBE buffer	0.1 M Tris, 0.1 M boric acid, 2 mM EDTA
TBS	20 mM Tris, 150 mM NaCl
TBST buffer	20 mM Tris, 150 mM NaCl. 0.1% (w/v) Tween-20
TE	10 mM Tris-HCl pH 7,5; 1 mM EDTA
WB Transfer buffer 1x	5x blotting buffer: Methanol: bidest H2O 1:1:3
Western Blot Blotting buffer 5x	29,11g Tris; 14,65g Glycin; 18,75ml SDS in 1l bidest water

2.1.6 Bioinformatic databases and tools

Database	Home page
DAVID	http://david.abcc.ncifcrf.gov/
Ensembl genome browser	http://www.ensembl.org
ExonPrimer	http://ihg.gsf.de/ihg/ExonPrimer.html
ExPaSy	http://www.expasy.org/
GenBank	http://www.ncbi.nlm.nih.gov/Genbank/
University of California, Santa Cruz genome browser	http://genome.ucsc.edu/
MFOLD	http://bioweb.pasteur.fr/seqanal/interfaces/mfold-simple.html
National Center for Biotechnology Information, Bethesda, MD, USA	http://www.ncbi.nlm.nih.gov/
Online Mendelian Inheritance in Man (OMIM)	http://www.ncbi.nlm.nih.gov/Omim
Panther	http://www.pantherdb.org/
POSMED	http://omicspace.riken.jp/PosMed/
Primer3	http://frodo.wi.mit.edu/cgi-bin/primer3/primer3_www.cgi
Webcutter	http://rna.lundberg.gu.se/cutter2/

2.1.7 Patients and sampling

Collaboration with local genetic counsellors in several provinces of Iran provided families with a minimum of two intellectual disabled children. A subset of families whose pedigree patterns and clinical data seemed to be compatible with moderate to severe NS-ARID were selected and visited by experienced clinical geneticists, or invited to the Genetics Research Center in Tehran. All participants completed a standardized questionnaire that included Medical conditions of interest and questions regarding selected personal medical history. Patients and unaffected relatives photographs were taken to document physical findings. The clinical geneticists assessed the mental status of the probands by monitoring their verbal and motor abilities, by interviewing the parents about developmental milestones and, in a minority of cases, by using more sophisticated tests such as a modified version of the Wechsler Intelligence Tests for children or adults. After obtaining written consent from the parents, peripheral blood was taken from all mentally retarded individuals and their parents. Often unaffected sibs were also included, particularly in small families with closely related (first cousin) parents [Najmabadi et al., 2006]. Patients with karyotypic aberration using G-banding or some other known cause of intellectual disability were not included in this study.

2.1.8 DNA extraction

Although blood samples are the best choice for large amounts of genomic DNA, the collection of peripheral blood is not easy in some patients, especially in children and syndromic patients, who may show behavioural difficulties. Salting out method according to the standard procedure has been used to extract DNA from human blood cells [Miller et al., 1988].

2.1.9 Establishment of B-cell lines from peripheral blood mononuclear cells of patients

An EBV immortalized lymphoblastoid cell lines (LCLs) were established by the central cell culture facility for at least one of the affected individuals in each family.

2.1.10 Molecular evaluation of fragile X syndrome

FMR1 mutation is generally tested by the Southern blot method, which allows an evaluation of both the expansion and gene methylation status. PCR-based methods are also useful, especially for precisely sizing premutations or for excluding a diagnosis of fragile X, when a normal CGG repeat allele is found in a male patient [Mandel et al., 2004]. At least for one patient of each nuclear family, FMR1 gene analysis was carried out by PCR and Southern blot analysis if X-linkage could not be excluded.

2.1.11 Metabolic disorders test

Tandem mass spectrometry has been performed to exclude disorders of the amino acid, fatty acid (e.g. phenylketonuria) or organic acid metabolism [Chace et al., 2003; Wilcken et al., 2004]. Filter-dried blood of one patient per family was screened.

2.1.12 The karyotype analyses

Chromosome analysis (karyotyping) at the 450-500 G-band level can detect large genomic imbalances (losses or gains of DNA) in intellectual disable conditions such as Down, Turner and Edwards syndromes. At least one affected individual in each family standard karyotyped in order to exclude cytogenetically visible chromosomal aberrations.

2.1.13 Copy number variation analysis

Array-CGH investigations were carried out as previously published [Erdogan et al., 2006] and has been performed in group of Dr. Reinhard Ullmann, Molecular Cytogenetics Group, Max Planck Institute for Molecular Genetics, Berlin, Germany.

2.2 Methods

2.2.1 Linkage analysis methods

Identifying disease susceptibility genes is one of the major tasks in human genetics studies. Human genomes have 99.9% identity between people, but still millions of differences exist among the 3.2 billion base pairs [Kruglyak and Nickerson, 2001]. These Phenotypic variations can be caused by genetic variations and are associated with traits or diseases. Genetic markers, which are nucleotide variants with known positions, are often utitized for human disease analyses. It helps if the marker can be scored easily and cheaply using readily available material (e.g. blood cells) but it is more crucial if it be sufficiently polymorphic. Therefore a randomly

selected person has a good chance of being heterozygous. SNPs, single nucleotide polymorphisms, usually contain two alleles and because of their high quantity between the human genome, have been used widely as markers for genetic disease-mapping studies [Kruglyak, 1997; The International SNP Map Working Group, 2001]. The number of SNPs in the public database (dbSNP) increased from 2.6 million to 9.2 million with the finalizing of the first Phase of the HapMap Project (The International HapMap Consortium, 2005).
As genotyping cost has become cheaper and the process has become faster, genotyping for markers can be performed on a genome-wide scale, which produces a large amount of marker data for analysis [Gunderson et al., 2005; Syvanen et al., 2006]. Hence, statistical methods are required after numerous markers are genotyped from collected samples. Theoretical methods for linkage tests were proposed around 1930 [Fisher et al., 1935a; 1935b; Penrose et al., 1935].
Linkage analyses are used to find chromosome regions that do not recombine with a proposed disease locus. Linkage is often evaluated by the logarithm of the odds (LOD) score , the logarithm to the base 10 of the odds of linkage [Morton et al., 1955], which is the logarithm of odds of the recombination rate equal to θ estimated from the observed data with respect to the assumption that the recombination rate is 0.5. In order to determine linkage, informative meioses are required. A meiosis is informative for linkage when we can identify whether or not the gamete is recombinant [Strachan and Read 2004]. Recombination fraction values vary between 0 and 0.5. Ten meiosis are sufficient to give evidence of linkage if there are no recombinants for the given markers.
Positive LOD scores give evidence in favor of linkage, while negative LOD scores argue against linkage. LOD score 3 is the threshold for accepting linkage, with a 5% chance of error. Linkage is rejected if the LOD score is less than -2, and the values between -2 and +3 are inconclusive [Strachan and Read 2004]. Recombination rarely separates loci that lie very close together on a chromosome, because only a crossover located precisely in the small space between the two loci can create recombinants. Therefore sets of alleles on the same small chromosomal segment tend to be transmitted as a block through a pedigree. Such a block is known as a haplotype.
Linkage analysis can be performed with either two-point or multipoint estimates [Kruglyak et al., 1996]. For two-point linkage analysis, only one marker and the disease locus are considered when calculating the statistic. For multipoint linkage analysis, several markers are considered simultaneously with the disease locus. Hence, we can define the most likely position of the disease locus on the marker map. A map of the markers with distances between them is required for multipoint linkage analysis.

Parametric linkage analysis

Parametric or model-based linkage analysis is the analysis of the cosegregation of genetic loci in pedigrees. Loci that are close enough together on the same chromosome segregate together more often than do loci on different chromosomes. Loci on different chromosomes segregate together purely by chance. Each genotype for one genetic marker or locus is made up of two alleles, one inherited from each parent. Specific alleles are in gametic phase when they are coinherited from the same parent, they were present together in the gamete originating from that parent. The further apart two loci are on the same chromosome, the more likely it is that a recombination event at meiosis will break up their cosegregation. The main quantity of interest

in parametric linkage analysis is the recombination fraction θ (the probability of recombination between two loci at meiosis).
For any parametric linkage analysis, the genetic model for the disease of interest must be specified. For a simple Mendelian disease, this model comprises the mode of inheritance and frequency of disease allele. For some diseases, carrying the risk genotype does not always result in the individual being affected (incomplete penetrance).
In more complex models, only a proportion of disease cases are due to a specific major gene, resulting in some risk of disease for individuals with any disease genotype (inclusion of a sporadic rate). Model parameters must be chosen before the linkage analysis [Teare and Barrett 2005].

Non-parametric linkage analysis

For multifactorial diseases, where several genes (and environmental factors) might contribute to disease risk, there is no clear mode of inheritance. Methods to investigate linkage have therefore been developed that do not require specification of a clear mode of inheritance. Such methods are referred to as non-parametric, or model-free. The rationale is that, between affected relatives excess sharing of haplotypes that are identical by descent (IBD) in the region of a disease-causing gene would be expected, irrespective of the mode of inheritance. Various methods test whether IBD sharing at a locus is greater than expected under the null hypothesis of no linkage [Teare and Barrett 2005]. Therefore, in cases where specifying a complete genetic model is not possible, one can use a model-free, or non-parametric, method of linkage analysis. This method ignores unaffected people, and looks for alleles or chromosomal segments that are shared by affected individuals. In this study, linkage analysis has been done by Dr. Motazacker and mainly by Dr. Garshasbi, and both of the approaches were used to analyse the NS-ARID families and to determine the location of disease loci.
Individuals with rare recessive diseases in consanguineous families are likely to be autozygous for all markers in the vicinity of the disease locus. Autozygosity means homozygosity for marker alleles that are identical by descent, i.e. inherited from a recent common ancestor .

2.2.2 Prioritizing genes for mutation screening

To determine exactly the disease genes within the many susceptibility loci identified in linkage studies is difficult because these loci may contain hundreds of genes. However most of the disease genes will be involved in only a few different molecular pathways. Knowing about the relationships between the genes, can assess whether some genes functionally interact with each other, indicating a joint basis for the disease etiology. Hence some algorithms can be used to rank genes in a linkage interval in order to prioritize candidate genes for sequencing.
Prior to mutation screening in coding exons and exon-intron boundaries, the genes in each interval were ranked based on their expression patterns and functional relevance in the central nervous system by referring to the literature and/or using bioinformatics databases. For this purpose several databases such as PosMed, Prioritizer were used (bioinformatic database and tools, page 22).

2.2.3 Isolation of genomic DNA from lymphoblastoid cells

The standard protocol used in our laboratory is a modification of Sambrook's protocol for the isolation of DNA from mammalian cells [Sambrook 1989]. DNA extraction was performed based on the following protocol:
- Preparing buffer A with the following composition. After autoclaving 5ml from 20% SDS

Buffer A	For 100 ml
0,4 M Tris-HCL-buffer (pH=8)	40ml from 1 M stock
0,06 M Na-EDTA-buffer	12 ml from 0,5 M stock
0,15 M NaCl solution	15 ml from 1 M stock
	33 ml aqua dest.

(sodiumdodecylsulfat) was added to buffer. The cell pellets were transported on ice in the 50ml falcon tubes from the freezer room to the lab. 20ml of solution A, on ice, was added to each sample (the first 1 ml was added with a cut, filterless pipet tip and run it up and down several times, then added the 19ml that are left). Vortex was made until suspension appears homogenous. Then each falcon tubes has put into a holder at room temperature. 30μl RNase A (10mg/ml) was added and incubated for 60min at 37°C in a water bath. Then 5ml sodiumperchlorate was added and then mixture has shaked over head 10-15 times manually. 20ml cold chloroform under the hood was added and shaked by converting the tube 10-15 times manually followed by centrifugation for 10 min at 4000 rpm. The upper phase was removed with a glass pipette (if the upper phase is very cloudy, the chloroform extraction should be repeated). Then samples were transported in a new and labelled 50ml Falcon tube and a volume of ice-cold ethanol (100%) was added (e.g.: at 25 ml sample volume add 25 ml ethanol). The DNA transported in an Eppendorf tube containing of 1ml cold ethanol (70%) then followed by centrifugation for 1 min at 7500 U/min. The ethanol has pipetted off. To do the second washing, 500 μl of new ethanol (70%) was added and centrifuged for 1 min at 7500 rpm. The ethanol has been removed.Until the DNA pellet will be dried; tube leaved open in the Thermomixer at 50°C. Dried DNA was rehydrated in 500 ul Tris-EDTA (TE buffer) and it can be left over night at room temperature to redissolve.

2.2.4 Polymerase Chain Reaction (PCR)

In general, PCR amplifications were carried out in 50 μl reaction volumes containing 75 ng genomic DNA, 1 x reaction buffer, 10 pmol of each primer, 200 μM dNTPs and 1 unit Taq polymerase (Promega, Mannheim, Germany or Qiagen, Hilden, Germany). The following touchdown PCR profile was used.
Step 1: 96°C for 3 min followed by 20 cycles (95°C for 30 s, 65°C for 30 s) with a decrement of 0.5°C per cycle.
Step 2: 30 cycles (95°C for 30 s, 55°C for 30 s and 72°C for 30 s). The PCR was concluded by a 5 min extension at 72°C. Alternatively, a PCR profile consisting of an initial denaturation step at 96°C for 3 min followed by 30-40 cycles at 95°C for 30 s, primer sequence-dependent annealing temperature for 45 s and 72°C for 30 s, with a 5 min final extension period (72°C) was used.

2.2.5 Agarose gel electrophoresis

The specificity and the amount of the amplified products were checked by agarose gel electrophoresis before further analysis. The gel composition was 0.7-1.6% agarose (Invitrogen) in TBE buffer supplemented with 0.5μg/ml ethidium-bromide. At least 0.2 volumes of gel loading buffer containing 0.25% bromophenol blue, 0.25% xylene cyanol, and 30% glycerol were added to the nucleic acid solutions before loading into the wells. Hyperladder I, IV, VI and pUC mix 8 or Lambda DNA/EcoRI+HindIII were used as size markers. Gels were run at 100 V for 30-45 min. Nucleic acids were visualized and pictures taken using the E.A.S.Y Win32 gel documentation system.

2.2.6 Sequencing

Sequencing of samples in both directions was performed according to the Sanger sequencing method with fluorescently labeled didesoxynucleotides using the ABI 377 DNA sequencer. Quality and quantity of PCR products/clones were evaluated by agarose gel electrophoresis and spectrophotometry (if necessary, amplicon bands were excised from the agarose gel and purified using the Qiagen MiniElute PCR purification kit). The labelling reaction was carried out using the BigDye Terminatormix (Applied Biosystems) and the primers used for PCR product amplification.

The labelling reactions were carried out using the following amount of reagents shown in Table 2.1.

Thereafter, sequencing reactions were performed using the following temperature profile

Name	Amount
DNA (PCR Product)	2ng/100bp
BigDye Terminator mix (V3,1)	2µl
5X Buffer	2µl
Primer (10pmol)	1µl
H2O	Add to 10µl

Table 2.1: PCR reaction mix for sequencing reaction

shown in Table 2.2. For DNA precipitation and purifications 1μl 2%SDS was added to the

	Temperature	Time	Cycle number
Initial denaturation	96°C	1 min	1x
Denaturation	96°C	30 sec	
Annealing	50 °C	15 sec	25x
Extension	60°C	4 min	
	4°C	For ever	

Table 2.2: PCR conditions for sequencing reaction

samples and incubate at 98°C for 10 second. Then 25 μl 100% ethanol was added to each reaction and mixed thoroughly by inverting the tube followed by Centrifugation at 4000 rpm in the cool room for 60min. The supernatant carefully has been discarded by inverting the tubes and placing them on a paper towel. 150 μl 70% ethanol was added and inverted the tubes without disturbing the pellets and centrifuging has been done at 4000 rpm in the cool room for 30 min. Carefully the supernatant has been discarded by inverting the tubes and placed them on a paper towel. The washing step has been repeated. The pellet has been dried by putting the plate headfirst onto a paper towel and centrifuged just up to 4000 rpm and then stopped. The plate has been covered by an adhesive film. Sequence data were assembled and analysed using the GAP4 Contig Editor.1 or CodonCode aligner 1.6.0 beta 5 software.

2.2.7 Restriction Fragment Length Polymorphism (RFLP) Analysis

After finding mutations, which segregated with the affection status in the pedigree, a panel of healthy controls was screened for it, using direct sequencing or Restriction Fragment Length Polymorphism (RFLP) analysis. For RFLP analysis the amplicons containing the mutation were screened for restriction sites affected by the DNA damage and appropriate restriction enzymes (RE) were selected using webcutter (http://users.unimi.it/ camelot/tools/cut2.html) or other databases in a way that the number of restriction sites differed between PCR amplicons from mutation carriers and controls. DNA fragments including the position of the mutation were amplified separately for all the control individuals by PCR. Amplicons afterwards were digested using appropriate amounts of restriction enzymes. After 2-14 hours incubation at 37 °C, enzymes were inactivated by incubating the reaction mix at 80 °C For 20 minutes. Finally, digested products were separated by agarose gel electrophoresis. The following primers were used for amplification of DNA fragments in case of screening mutations in NSUN2. Restriction

Name	Sequence
NSUN2_e7_F	TGGATTGCTGCTTCTCACTG
NSUN2_e7_R	TGGATTGCTGCTTCTCACTG

Table 2.3: primers for DNA amplification in position of NSUN2 mutation in exon 7

mix was prepared according to Table 2.4 and incubated for 2 hours at 37°C then Speed-vac has been used to reduce the volume (at least 2 times) and finally Loaded on the agarose gel.

PCR product	10 µl
MseI restriction enzyme	0.1 µl
10x NEBuffer 2	5 µl
100x BSA	0.5µl
water	34.4µl

Table 2.4: Restriction mix

2.2.8 RNA extraction

Total RNA was isolated from patient lymphoblastoid cell lines using Trizol or RNeasy Mini Kit (Qiagen, Cat.#: 74104), according to the manufacturer's recommendations.

RNA extraction using Trizol

Cell pellet (5 x 107 cells) has been suspended with 10 ml Trizol reagent in a 30 ml RNase free tube then the suspension by shaking vigorously homogenized for several seconds, incubated 30 minutes at room temperature (20°C - 30°C) to be completely dissolved. 0.2 ml chloroform was added for each 1 ml of initial Trizol (2 ml), and shaked for 15 seconds and incubated for additional 2-3 min at room temperature followed by centrifugation for 20 min at 5000 RPM at 4°C. The aqueous phase was transferred to a fresh 30 ml tube or make aliquots of 550 μl in 1.5 ul eppendorf tubes. 0.5 volume of isopropanol per 1 ml of TRIZOL reagent used for initial homogenization (5 ml or 550 μl) to the aqueous phase, and was mixed well by vortexing and hold in room temperature for 5-10 min. The samples centrifuged for 10 min at 8000 RPM at 4°C (12000 g for microfuge). Remove the supernatant and add 10 ml filter sterilized 70% ethanol (500 μl for microtube) and mix well. The samples were centrifuged for 5 min. at 5000 RPM at 4°C (7500 g for microfuge). The supernatants were took off and air dried the pellet(Avoid completely drying the pellets, as this will decrease the solubility of the RNA). The RNA dissolved in 500 μl of sterile DEPC water and put it on ice for 10 min then was incubated for 5 min at 65°C using heating block or water bath. The RNA concentration was measured by Nanodrop ND-1000 Spectrophotometer (Peqlab Biotechnologie GmBH) and the quality on Agarose gel was controlled. The RNA samples can keep in the freezer (-20°C or -80°C) until further use.

2.2.9 First-strand cDNA synthesis using SuperScript TM III for RT-PCR

cDNA synthesis was performed according to the following protocol:
50-250 ng of random primers has been added to a 0.2 ml eppendorf tubes. Then 10 pg - 5 μg total RNA, 10 mM dNTP Mix (Mix: 10 mM each dATP, dGTP, dCTP and dTTP at neutral pH) were added and distilled water has been used to bring the volume to a total of 13 μl. Mixture should be heated to 65°C for 5 min and incubated on ice for at least 1 min. The contents of the tube by brief centrifugation were collected. 5 μl 5X First-Strand Buffer, 1 M DTT, Recombinant RNase Inhibitor, SuperScriptTM III RT (200 units/μl) were added, Mixed by pipetting gently up and down and incubated at 25°C for 5 min. It was incubated at 50°C for 30-60 min and followed by inactivation of the reaction by heating (70°C for 15 min). cDNAs synthesis were checked using primers for *HUWE1*, a house keeping control gene, with exon spanning primers CAAGTGAGGAAAAGGGCAAA (exon 64) and GTTCATGAGCTGCCCCAGT (exon 65) which give rise to a 568 bp amplicon.

2.2.10 Whole genome expression profiling

The Sentrix Human-6 Expression BeadChips contains six arrays on a single BeadChip, each with around 46.000 probes derived from human genes in the National Center for Bioinformatic Information (NCBI) Reference Sequence (RefSeq) and UniGene databases. 50-100 ng of

total RNA are required for the single-round in vitro transcription (IVT) reaction. Beads are assembled into 1.6 million pits, each measuring 3um in diameter, generating an average 30-fold redundancy for each sequence represented on the array. This means that each reading is taken multiple times across the array, increasing the accuracy of the measurement. Six samples can than be interrogated simultaneously on the Human-6 Expression BeadChips. Bead content design oligos that are covalently attached to beads in Human-6 Expression BeadChips contain a 29-base address concatenated to a 50-base gene-specific probe. The address is used to map and decode the array, while the probe is used to quantify expression levels of transcripts .

cRNA amplification

RNA amplification is one of the standard methods to prepare RNA samples for analysis by expression microarray techniques. The Illumina® TotalPrep RNA Amplification Kit, manufactured by Ambion, Inc. was used for generating biotinylated, amplified RNA for direct hybridization with Illumina Sentrix® arrays. The procedure consists of reverse transcription with an oligo (dT) primer bearing a T7 promoter using Array-Script, a reverse transcriptase (RT) engineered to produce higher yields of first-strand cDNA than wild type enzymes. ArrayScript catalyzes the synthesis of virtually full-length cDNA, which is the best way to ensure production of reproducible microarray samples. The cDNA then undergoes second strand synthesis and clean-up to become a template for in vitro transcription with T7 RNA Polymerase. To maximize cRNA yield, Ambion's proprietary MEGAscript® in vitro transcription(IVT) technology along with biotin UTP (provided in the kit) is used to generate hundreds to thousands of biotinylated, antisense RNA copies of each mRNA in a sample. Reverse transcription to synthesize first-strand cDNA is primed with the T7 oligo(dT) primer for synthesis of cDNA containing a T7 promoter sequence. Second-strand cDNA synthesis converts the singlestranded cDNA into a double-stranded DNA (dsDNA) template for transcription. The reaction employs DNA polymerase and RNase H to simultaneously degrade the RNA and synthesize second strand cDNA. cDNA purification removes RNA, primers, enzymes, and salts that would inhibit in vitro transcription. In vitro transcription to synthesize cRNA generates multiple copies of biotinylated cRNA from the double-stranded cDNA templates; this is the amplification and labeling step. cRNA purification removes unincorporated NTPs, salts, enzymes, and inorganic phosphate. After purification, the cRNA is ready for use with Illumina's direct hybridization array kits. While as little as 50 ng total RNA can theoretically be used to produce enough material for further hybridizations, we used 300 ng of total RNA per reaction.

Six-sample BeadChip hybridisation

Upon the completion of the cRNA amplification, RNA samples were quantified using Nanodrop ND-1000 Spectrophotometer (Peqlab Biotechnologie GmBH). 1.5 μg of cRNA sample was hybridized to the BeadChip in a multiple step procedure according to the manufacturer's instructions by our central facility. The chips were dried and scanned on the BeadArray reader.

Normalization and differential analysis algorithms

All methods of normalization aim to improve data by mathematically factoring out systematic errors among experimental groups so that their values can be compared. In the case of microarray experiments, systematic variation can result from variation in hybridization temperature, sample concentration, formamide concentration, etc. All forms of normalization achieve this result by making assumptions about the experimental samples and adjusting their values in a way that would factor out intensity changes arising from experimental variation without affecting changes based on true biological differences. The key to applying normalization effectively, therefore, is to understand the underlying assumptions of each method and deciding if they apply in the case of our experiment. Normalization is a process by which two or more populations of gene expression values from two or more samples are adjusted for easier comparison. A scaling factor is a number by which values in one population are multiplied for the sake of normalization. For example, if a normalization technique multiplies all values in Sample B by 1.5 to normalize to Sample A, we say that a scaling factor of 1.5 was applied. BeadStudio provide different methods of normalization, for our experiments the "Rank- Invariant Method" was used.

Rank-Invariant method

For most types of expression experiments, this is the most highly recommended normalization method. Rank-Invariant normalization uses a linear scaling of the populations being compared. However, unlike with averaging, the scaling factor is determined not by an average of all genes, but by only rank-invariant genes. 'Rankinvariant' genes are those whose expression values show a consistent order relative to other genes in the population. For example, a gene that is the 200th brightest gene in Sample A and the 203rd in Sample B would be considered rank-invariant and would be used to arrive at the normalization factor; a gene that goes from 200th to 10000th would not be rank-invariant and would not be used. This method is much more resistant to outliers than straight averaging and generally gives better results. However, as with averaging, if samples are very different in their behaviors, the underlying assumption of rank-invariance (the existence of a subpopulation of genes whose expression is constant across samples showing consistent ranks) will not be true and the method should not be applied.

Differentiation score

The differentiation score (diff. score) is a transformation of the p-value that provides directionality to the p value based on the difference between the average signal in the reference group vs. the comparison group.The diff. score of 13 corresponds to a p-value of 0.05, the diff. score of 20 corresponds to a p-value of 0.01, and the diff. score of 30 corresponds to a p-value of 0.001. A positive diff. score represents up-regulation, while negative represents downregulation.

2.2.11 Functional gene classification tools

Development of robust and efficient methods for analyzing and interpreting high dimension gene expression profiles continues to be a focus in computational biology. The accumulated

experiment evidence supports the assumption that genes express and perform their functions in modular fashions in cells. Therefore, several computational algorithms have emerged that use robust functional expression profiles for precise classification of complex human diseases at the modular level. In this study, two web based classification tools was used: DAVID (http://david.abcc.ncifcrf.gov/) and Ingenuity(http://ingenuity.com) .

The Database for Annotation, Visualization and Integrated Discovery (DAVID)

Grouping genes based on functional similarity can systematically enhance biological interpretation of large lists of genes derived from high throughput studies. The DAVID functional classification tool generates a gene-to-gene similarity matrix based shared functional annotation using over 75000 terms from 14 functional annotation sources like KEGG data base (Kyoto Encyclopedia of Genes and Genomes, a collection of manually drawn pathway maps representing the molecular interaction and reaction networks for: metabolism, genetic information processing, environmental information processing, cellular processes and human diseases). The DAVID clustering algorithm classifies highly related genes into functionally related groups. Tools are provide to further explore each functional gene cluster, including the listing of the "consensus terms" shared by the genes in the cluster, the display of enriched terms, and a heat map visualization of geneto- term relationships. A global view of cluster-to-cluster relationships is provided using a fuzzy heat map visualization. Summary information provided by the functional classification tool is extensively linked to DAVID Functional Annotation Tools and to external databases allowing further detailed exploration of gene and term information. The functional classification tool provides a rapid means to organize large lists of genes into functionally related groups to help unravel the biological content captured by high throughput technologies. In our case, DAVID was used to classify genes with differentiation scores smaller than - 13 in patients a compared to controls.

2.2.12 Real time PCR

SYBR green was used to monitor DNA synthesis. SYBR green is a dye that binds to double stranded DNA but not to single-stranded DNA and is frequently used in real-time PCR reactions. When it is bound to double stranded DNA it fluoresces very brightly (much more brightly than ethidium bromide). In addition the ratio of fluorescence in the presence of double-stranded DNA to the fluorescence in the presence of single-stranded DNA is much higher for SYBER green than for ethidium bromide.
Primers:
Intra-exonic primers for the regions of interest with a product size of about 90-160bp were designed using Primer3 program. The probability of secondary structure conformations for the amplicons was predicted using the M-Fold program (http://helixweb.nih.gov/nih-mfold/). Primer quality was checked by comparing the amount of product after 25, 30 and 35 amplification cycles (Table 2-5) for a normal cDNA on the agarose gel. SYBR® Green RT-PCR:
The SYBR Green PCR Master Mix is a convenient premix of all the components necessary to perform real-time PCR using SYBR® Green I Dye, except primers, template and water. Direct detection of polymerase chain reaction (PCR) product is monitored by measuring the increase

96°C	3min	1x
96°C	30sec	25x, 30x and 35x
55°C	30sec	
72°C	30sec	
72°C	10min	1x

Table 2.5: PCR Program used for checking primers

in fluorescence caused by the binding of SYBR Green dye to double-stranded (ds) DNA. The SYBR Green PCR Master Mix is supplied in a 2X concentration and contains SYBR Green I Dye, AmpliTaq Gold® DNA Polymerase, dNTPs with dUTP, and optimized buffer components. Standard curves as series of 2 fold dilutions were produced for the loading control (or

Reagent	Volume (µl)	Water (µl)	Final volume (µl)	Add to each reaction (µl)
Primers (100pmol)	8+8 (stock)	384	400	5
SYBR Green master mix	1200	-	1200	15
Master Mix			1600	20
cDNA	1.6	30.4	31	10

Table 2.6: Reaction protocol for a 96 well plate

reference gene) as well as for the gene of interest whose expression we think may change under experimental conditions. All reactions were performed in triplicate . Negative controls for each reaction were used in order to prove that primers and Taq polymerase/SYBR green PCR mixes were not contaminated. They also allowed us to determine if the primers can form primer-dimer artefacts, which are most readily seen when there is no appropriate DNA for amplification. Prior to starting the preparation of PCR plates preparation a template plate file was generated using the SDS2.1 software (AB applied biosynthesis). Experiments were performed in an ABI (PRISM 7900 HT) 96 well machine. The produced data files were analysed using the SDS 2.1 software followed by T Test and standard deviation calculations in Excel.

2.2.13 Western blotting

Cell lysate preparation

Cell lysates were prepared using the following protocol:

- Apply 3 μl lysate buffer per 20.000 cells (Total volume should be at least 100 μl to enable sonicating)
- Sonicate: 10-15 bursts (Amplitude 20-30) with the sonicator (Bandelin Sanopuls, Pro. No: 519.00002687.033)
- Denature at 95°C for 2 min, afterward vortex and spin down.

Component	Amount for 150 ml
0.1M DTT	1.5 ml
0.01% bromophenol blue	1.5 ml
10% Glycerol	17.25 ml
60mM Tris, pH6.8	9 ml
2% SDS	30 ml
Water	90.75 ml

Table 2.7: Cell lysation buffer

Separation of denatured proteins by SDS-PAGE:

For separating proteins electrophoretically, a Sodium Dodecyl Sulphate Poly Acrylamide Gel Electrophorisis (SDS-PAGE) with 10% acrylamide gels (Table 2-9) was performed, using a Bio-rad mini-apparatus (Model No: Mini-Protein® 3 cell). The method is called SDS-PAGE due to the fact that SDS, a strong anionic detergent is used to denature the proteins and a discontinuous polyacrylamide gel is used as a support medium to separate the denatured proteins according to their molecular size. The most commonly used system is also called the laemmli method after U.K. laemmli, who was the first to demonstrate this SDS-PAGE as a technique to separate proteins [Laemmli et al., 1970].

- Denature protein samples completely by first adding Laemmli protein loading buffer in 1:4 v/v (from a 5x stock of Laemmli protein loading buffer Table 2-8) and subsequently heating the mixture at 95°C for 5 minutes.
- Prepare SDS-PAGE cassettes by using a pair of clean glass plates (10 cm wide and 7 cm high) separated by a pair of spacers (0.75 mm thickness for thin gel or 1.5 mm for thick gel).
- Fill up approximately 5 cm of the cassettes with liquid separating gel mixture and allow to polymerize within the cassettes.
- Add a thin layer of water slowly to the top of separating gel layer to avoid evaporation and seep the surface separating gel smooth.
- Allow the gel to polimerize for 30 min
- Remove the water by pouring it and pipeting if necessery
- Pour stacking gel mixture on the top of the separating gel and insert a 15-well comb within.
- After polymerization of the stacking gel, remove combs slowly without disturbing the wells.
- Insert cassette into the electrophoresis chamber vertically, and fill with electrophoresis running buffer
- Load denatured protein samples into the wells using a Hamilton syringe.

- Connect the apparatus to a constant current source (10 mAmp for thin gel and 20 mAmp for thick gel) for electrophoresis. Migration of the proteins in the gel can be judged by visually monitoring the migration of the tracking dye (Bromophenolblue) in the protein-loading buffer.
- When the dye front comes close to the end of the gel, stop the electrophoresis.

Component	Amount for 150 ml
0.1M DTT	1.5 ml
0.01% bromophenol blue	1.5 ml
10% Glycerol	17.25 ml
60mM Tris, pH6.8	9 ml
2% SDS	30 ml
Water	90.75 ml

Table 2.8: Buffers and solutions required

SDS-PAGE running buffer (1X)	196mM glycine, 0.1% SDS, 50mM Tris-HCl (pH 8.3)
Laemmli protein loading buffer (5X)	62.5 mM Tris HCl (pH 6.8), 5% beta-mercaptoethanol (v/v), 50% Glycerol (v/v), 2% SDS (w/v), 0.1% (w/v) Bromo phenol Blue. Volume was made by adding water.
Separating gel mixture	10% Bis-Acrylamide (v/v), 375 mM Tris HCl (pH 8.8), 0.1% SDS (w/v), 0.1% ammonium persulfate, 0.005% TEMED in water.
Stacking gel mixture	4% Bis-Acrylamide (v/v), 125 mM Tris HCl (pH 6.8), 0.1% SDS (w/v), 0.1% ammonium persulfate, 0.005% TEMED in water.

Table 2.9: Component volumes for SDS-PAGE gels (in ml)

APS and TEMED were added just prior to pouring the gels

Western blotting analysis

Western blotting was performed according to the following protocol:

- Incubate unfixed SDS-PAGE gel shortly in a transfer buffer.
- Soak Whatman paper and nitrocellulose membranes in the same transfer buffer.
- Place the gel on the membrane.
- Place two layers of Whatman papers on both sides of the gel-membrane
- combination to make the transfer set.
- Remove air bubbles from the whole transfer-set by rolling a glass rod over it.
- Pace this combination on a transfer unit in such a way that the gel is connected to the cathode while the membrane is connected to the anode.

- Connect the apparatus to a power supply and perform electro-transfer at a constant current of 50 mAmp (for a single gel with 10x7cm) for 2 hours.
- Confirm transfer of proteins from the gel onto membrane by staining the membrane with Ponceau Red dye solution.
- Block the membrane with 5% non-fat milk dissolved in TBS-T buffer for 1 hour.
- Incubate with primary antibody in TBST buffer for 1 hour
- Wash 3 times by shaking (50 rpm) with TBST buffer for 5 min.
- Incubate with secondary antibody in TBST buffer for 1 hour.
- Wash 3 times by shaking (50 rpm) with TBST buffer for 5 min.
- Visualise signals on the membrane, according to the procedures recommended by PerkinElmer kit (western Lightning Chemiluminescence Reagent, NL100):
- Mix solutions A +B, 1:1, (2 ml is sufficient for 1 mini gel) and spread over memberane
- Incubate for 1 min at room temperature.
- Wrap in cling film.
- Expose to the Fuji Medical X-Ray films (30 sec to several mins) and develop it. In some experiments, blots were stripped by incubating the blots in a stripping buffer at 50°C for 30 minutes (shaking with 50 rpm) and re-probed again with a different primary antibody.

Transfer buffer	0.1% SDS, 20% (v/v) MeOH, 48mM TRIS7/HCl, 39 mM Glycine
Ponceau Red solution	2% (w/v) Ponceau S dye, 5% (v/v) Acetic acid
TBS-T	20 mM Tris, 150 mM NaCl. 0.1% (w/v) Tween-20
Stripping buffer	1% SDS, 20mM TRIS/HCl (pH 6.8), 1% (v/v) β-Mercaptoethanol

Table 2.10: Buffers and solutions required for Western blotting

2.2.14 Concentration measurements

Cell counting

Cells were counted using a hemacytometer: a hemacytometer is an etched glass chamber with raised sides that will hold a quartz covership exactly 0.1 mm above the chamber floor. The counting chamber is etched in a total surface area of 9 mm2. Calculation of concentration is based on the volume underneath the cover slip. One large square has a volume of 0.0001 ml (length x width x height; i.e., 0.1 cm x 0.1 cm x 0.01 cm). To fill a hemacytometer place a pipette tip filled with a well suspended mix of cells at the notch at the edge of the hemacytometer and then slowly expel some of the contents. The fluidic is then drawn into the chamber by capillary action. Staining of the cells often facilitates visualization and counting. One can either mix

cells with an equal volume of trypan blue [0.4% (W/V) trypan blue in PBS] to determine live /dead count (dead cells are blue) or kill cells with 10% fromalin and then stain e.g. with trypan blue to improve visibility of all cells. Count the number of cells in the 4 outer squares. The cell concentration is calculated as follows: Cell concentration per milliliter=total cell count in 4 squares x 2500 x dilution factor.

DNA and RNA concentration assay

DNA and RNA concentrations measured using Nanodrop ND-1000 Spectrophotometer (Peqlab Biotechnologie GmBH).

2.2.15 Cloning of wild type ZNF526 and both type of mutants

Incyte Full Length Human *ZNF526*-cDNA Clone pINCY (biocat, Clone ID: LIFESEQ2646019) has been used as the source of *ZNF526*-cDNA. To be certain about the vector obtained, a single digestion has been done using HindIII, NcoI restriction enzymes separately. Moreover a double digestion has been conducted using NotI and EcoR1 restriction enzymes, which could take out the insert. Also the entire *ZNF526* has been amplified and sequenced. pINCY vector information has been shown in figure 2.1.

Site-Direct Mutagenesis

The two mutations, G to A and C to A, were introduced separately to *ZNF526* wild type by site-directed mutagenesis, using the Stratagene QuikChange QuikChange® II XL Site-Directed Mutagenesis Kit, based on the manufacturer's instructions. The following primers were used for inserting two mutations:
Thereafter site-directed mutagenesis, plasmids has been extracted and their sequence was con-

Primer Name	Primer Sequence (5' to 3')
R>Q_g>a	5'-ccggctgtcccagcaccggcgtg-3'
R>Q_g>a_antisense	5'-cacgccggtgctgggacagccgg-3'
Q>H_g>c	5'-gtggcaagcgcttcacacacagctccaacc-3'
Q>H_g>c_antisense	5'-ggttggagctgtgtgtgaagcgcttgccac-3'

Table 2.11: Sequence of the primers employed to introduce mutations in the wild type *ZNF526*

firmed by sequencing.
For cloning full length *ZNF526* (NM_133444), the three Wild type, Mutant G>A and Mutant C>A transcript s were amplified from wild type-pINCY, Mutant G>A-pINCY and Mutant C>A-pINCY vectors. The forward primer contains BamHI sequence and also 'CC' sequence in order to facilitate the directional cloning of the amplicon into the pCMV-Tag Epitope Tagging Mammalian Expression Vectors. The reverse sequence contains XhoI digestion site and did not contain a TAA stop codon. Two expression vectors, pCMV-Tag2 and pCMV-Tag3, which contained the coding sequence for a N-terminal FLAG tag and N-terminal cMYC tag to be joined to the insert were employed (Figure 2.3). The synthetic FLAG epitope is composed of eight amino acid residues (DYKDDDDK). The c-myc epitope contains ten amino acid residues (EQKLISEEDL).

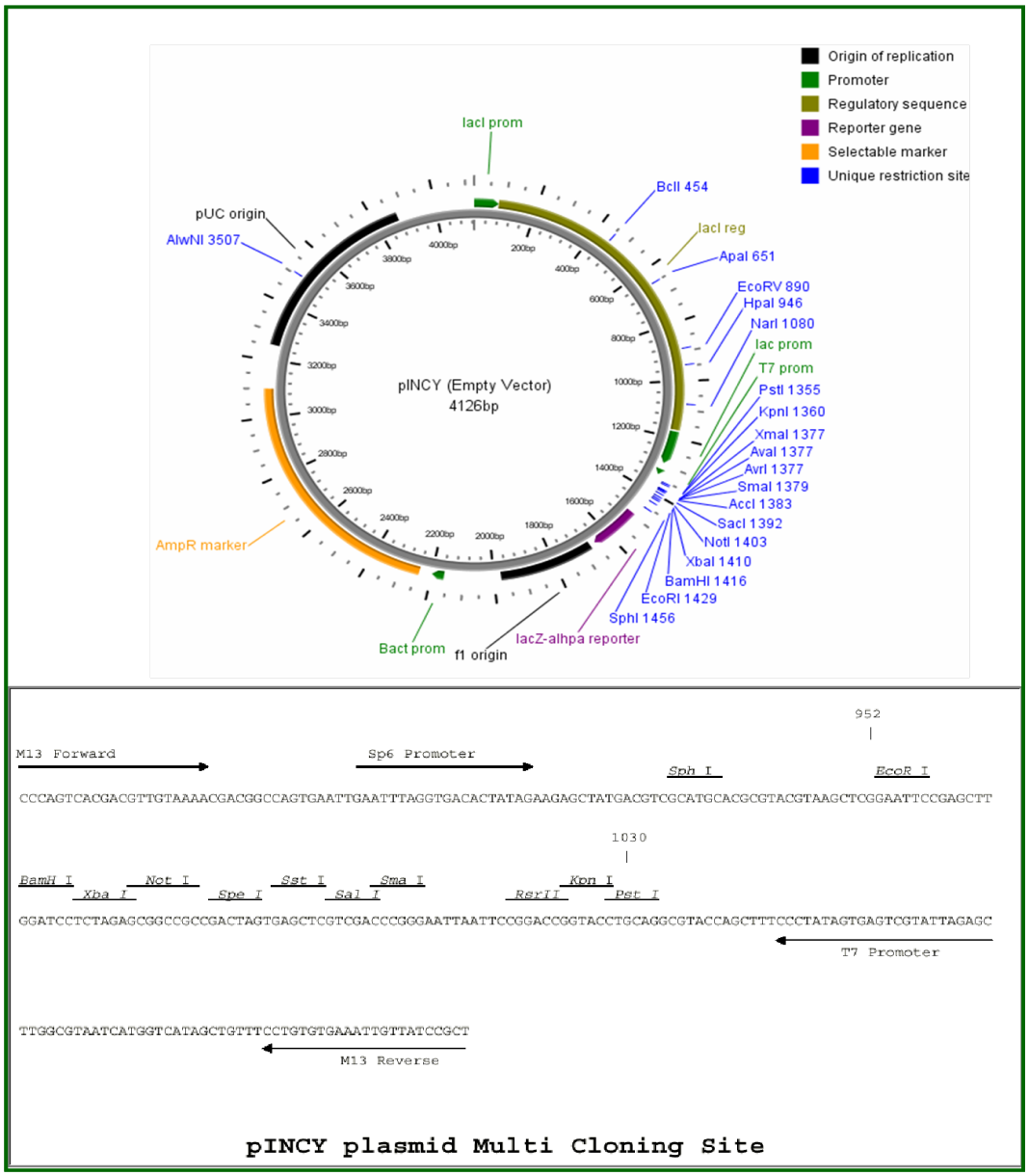


Figure 2.1: The structure of pINCY vector and the multiple cloning site.

ZNF Cloning Primers	Sequence
ZNF526-pCMV-F2	GAGCTGGATCCCATGGCAGAGGTGGTGGCTGAG
ZNF526-pCMV-R	GGTCACTCGAGTCACACGAAGGCCGTGTC

Table 2.12: The primers were used for in frame insertion of *ZNF526* to pCMV vector

These vectors contain features for expression in eukaryotic cells using (CMV) promoter. The amplification of cDNAs was done using Pfu high fidelity DNA polymerase.

Chemical transformation

100 ng of Plasmid DNA were incubated with 250 μl of chemically competent E-Coli Top10 cells for 30 min on ice. The mixture was then heat shocked at 42°C for 45 sec and chilled on ice for 2 min. The cells were afterwards incubated for 1 h in 1 ml LB medium with vigorous shaking

at 37°C before incubating them overnight at 37°C on agar plates containing the appropriate selection antibiotic.

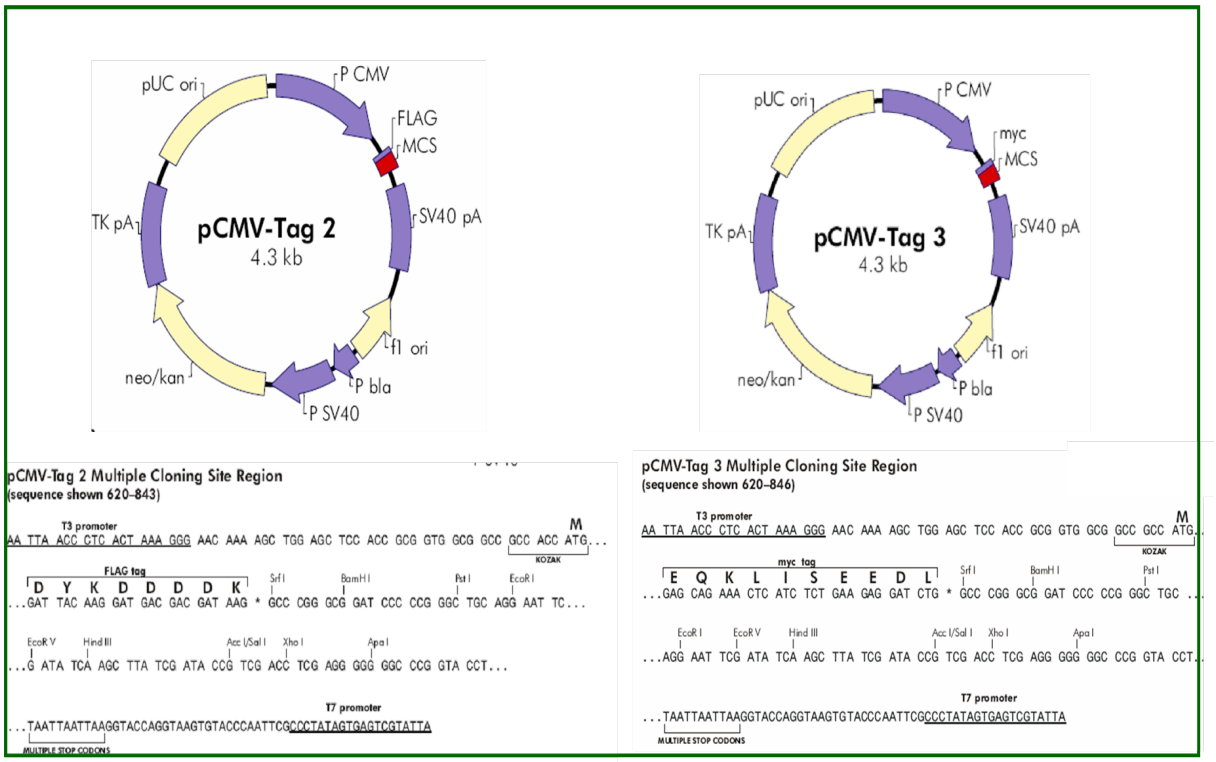


Figure 2.2: The stucture of two vectors pCMV-Tag2 and pCMV-Tag3. Tag2 contains FLAG epitope and Tag3 contains myc epitope.

Transfection

Hela cells and Neuroblastoma SH-SY5Y cells were transfected with ZNF526 wildtype (Wt) or Mutant constructs for either protein labeling and extraction or immunfluorescence studies. Nontransfected cells (Nt) were used as negative controls. For transfection, 60-80% confluent flasks were used. The following steps were performed: 83 μl transfection agent DreamFect® were mixed with 500 μl OptiMEM and pre-incubated for 5 min. Purified plasmid DNA (66 μg) was mixed with 500 μl OptiMEM and pre-incubated for 5 min. The mixture of DreamFect®-OptiMEM was then added to Construct-OptiMEM and incubated for at least 20 min. Then the final mixture was added to each flask after replacing the culture medium with fresh medium without antibiotics to reduce the stress condition in the flask as much as possible. 24 h after transfection, the medium was discarded and the cells were washed with sterile room temperature PBS prior to incubating them for another 24-48 h with fresh medium until they reached full confluence. For immunofluorescence experiments, $2x10^4$ cells were cultured for 24 h in 6-well plates on coverslips coated with poly-lysine (Poly-lysine coating provides an anchored platform for the cells and helps those cells for a better growth). For transfection purified plasmid DNA (2 μg) and 8 μl DreamFect® were separately mixed with 100 μl OptiMEM and pre-incubated for 5 min, then mixed with each other and incubated for 20 min before adding this total mixture to 2 ml fresh medium. 24 h after transfection, old medium was removed and were cultured for another 24 h with an antibiotic-free medium before they were fixed for im-

munofluorescence studies.

Immunocytochemistry

Cells were grown and transfected on glass cover slips. Two days after seeding or transfection, the cells were fixed either with 2% paraformaldehydde at room temperature (RT) or with 80% methanol in PBS (phosphate-buffered saline) at -20°C for 10 minutes, permeablized with 0.4% Triton X-100 in PBS for 5 minutes, followed by incubation with 100mM glycine dissolved in PBS for 1 hour. The cells were blocked with 5% normal goat serum or bovine serum albumin (BSA). After incubating the cells with the primary antibody for 1 hour at RT, the cells were washed three times with PBS supplemented with 0.1% Tween-20 (PBS-T). The cells were further incubated with secondary antibody diluted in PBS-T buffer. After incubation with secondary antibodies, the cells were washed three times with PBST buffer. The cover slips were finally mounted onto glass slides with fluoromount G (southern biotechnology).

Restriction digestion of dsDNA

For restriction digestion, approximately 1mg dsDNA was used in 20 ml restriction digestion mixture. The restriction-digestion mixture contains specific restriction buffer at a concentration of 1x and doubly distilled autoclaved water. Depending on the enzyme activity, restriction enzymes were added at the ratio of 1 unit/mg of DNA. In some cases more enzyme was used. Depending on the activity of the enzyme, the reaction mixture was incubated at 37°C for 3 hours or over night. For doubled restriction digestion, a compatible buffer was selected according to the manufacturer's (NEB) instruction.

Ligation of dsDNA

For ligation, restriction enzyme digested insert and vector dsDNA were first checked on an agarose gel and the bands were excised from the gel. The dsDNA was subsequently purified from the gel by using Qiagen kit according to the manufacturer's instruction. The purified insert and vector dsDNA were mixed with doubled distilled autoclaved water and ligation buffer so that the concentration of ligation buffer become 1X in the final solution. Finally T4-DNA ligase was added to the reaction mixture. The mixture was incubated over night at 16°C. For a better efficiency of ligation, insert and vector in the ligation mixture was maintained at a higher molar ratio (at least 10:1 for insert to vector). This enhances the probability of ligation reaction.

Competent E.coli cell preparation

Competent E. coli cells were prepared as mentioned below. A single colony of bacterial cells was incubated in 3 ml of Luria-Bertani (LB) broth and grown over night. One ml of this culture was added to 100 ml of LB liquid medium and incubated at 37°C to make a broth culture. The cells were grown until an OD 0.4-0.5 was reached at 600 nm. The cells were incubated for 10 min on ice and centrifuged at 3000 rpm for 5 min at 4°C. The pellet was resuspended in 30 ml of ice-cold CaCl2 (100mM) solution supplemented with 10% glycerol, distributed in tubes and stored at -80°C.

2.2.16 Generation of a stable neuroblatoma cell line

Generation of lentiviral ZNF526 wild type and mutants expression constructs

ZNF526 coding regions were amplified by PCR using pINCY plasmids containing full-length cDNAs of wild type, G>A and C>A mutants, respectively. The amplified fragments were then ligated into pLenti6/V5 TOPO vector (Invitrogen Life Technologies) following the manufacturer's protocol using the primers listed in table 2.13. The forward primers contain the BamH1

ZNF Cloning Primers	Seq
ZNF526-ORF-F	GAGCTGGATCCCACCATGGCAGAGGTGGTGGCTGAG
ZNF526-ORF-R	GGTCACTCGAGCCAAACACGAAGGCCGTGTC

Table 2.13: The primers were designed to be in frame to V5 epitope in pLenti6/V5 vector

restriction site and CACC sequence for directional cloning. The reverse primer contains XhoI restriction site and AAA sequence inserted instead of stop codon which resulted protein will be in frame with V5 tag. Each construct was validated for the correct *ZNF526* wildtype and both

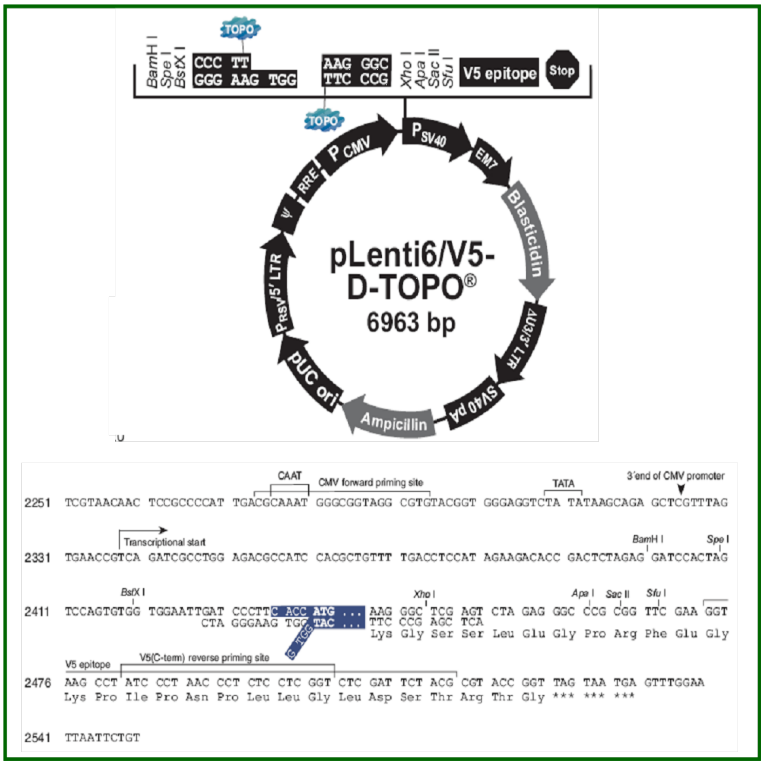


Figure 2.3: The special cloning site in pLenti6 vector.

mutations by DNA sequencing. These expression constructs that contained the *ZNF526* gene in pLenti6/V5 vector plasmid was co-transfected with the ViraPower Packaging Mix into the 293FT cell line to produce lentivirus, and the obtained lentiviral stock was transduced to the primary neurons. The ViraPower Packaging Mix contains an optimized mixture of the three packaging plasmids, pLP1, pLP2, and pLP/VSVG. These plasmids supply the helper functions as well as structural and replication proteins in trans required to produce the lentivirus. An optimized 293FT producer cell line stably expresses the SV40 large T antigen under the con-

trol of the human CMV promoter and facilitates optimal production of virus. Construction of replication-incompetent lentiviral vectors was performed using ViraPower expression system (Invitrogen), according to the recommendation of the manufacturer. Lentiviral constructs were used to infect RCN. Expression of transfected *ZNF526* was detected by antibodies against the V5-tag, expressed in these expression vectors or by GFP visualization, which represented the tag protein for the corresponding control.

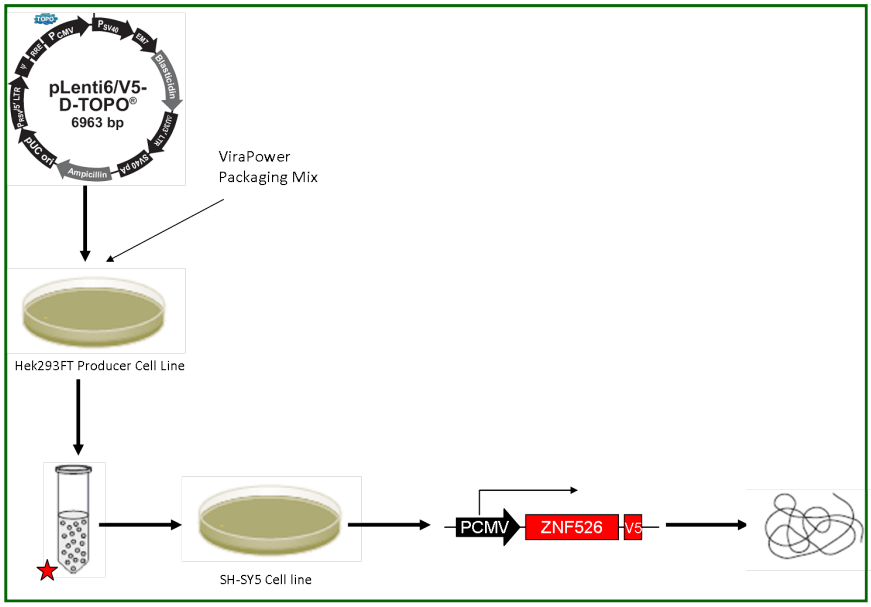


Figure 2.4: Production of pLenti-V5 expression cunstruct containing *ZNF526* Wt, R459Q and Q539H. The 293FT producer cell line cotransfected with pLenti/V5 expression constructs and the optimized packaging mix. Viral supernatant was harvested and determined the titer, thereafter the viral supernatant was added to SH-SY5Y neuronal cells and then blasticidin was used for selection of the stably transfected cells.

2.2.17 Chromatin Immunoprecipitation-Sequencing

Chromatin Immunoprecipitation Sequencing (ChIP-Seq) combines ChIP with DNA sequencing, allowing researchers to identify the binding sites of DNA-associated proteins. It can be used as a more cost effective and higher quality alternative to whole genome microarray hybridization. Chromatin immunoprecipitation (ChIP) has become an important assay for the genome-wide study of protein-DNA interactions and gene regulation. In a typical ChIP experiment, protein complexes that contact DNA are crosslinked to their binding sites, the chromatin is sheared into short fragments, and then the specific DNA fraction that interacts with the protein of interest is isolated by means of immunoprecipitation (IP). The goal of ChIP-Seq data analyses is to find those genomic regions that are enriched in a pool of specifically precipitated DNA fragments.

ChIP-Seq illustrates the power of new sequencing platforms, such as those from Solexa/ Illumina, to perform sequence census counting assays. The generic task in these applications is to identify and quantify the molecular contents of a nucleic acid sample whose genome of origin has been sequenced. The very large numbers of short individual sequence reads produced by these instruments (currently about 400,000 reads of 200 nucleotides (nt), or around 40 mil-

lion reads of 25 nt, per instrument run, depending on the platform used) are extremely well suited to making direct digital measurements of the sequence content of a nucleic acid sample. By determining a short sequence read from each of many randomly selected molecules from the sample and then informatically mapping each sequence read onto the reference genome, the identity of each starting molecule is learned, and its frequency in the sample is calculated. Sequence census assays do not require knowing in advance that a sequence is of interest as a promoter, enhancer, or RNA-coding domain, as most current microarray designs do. Below, we use the Solexa/Illumina platform, because high-read numbers contribute to high sensitivity and comprehensiveness in large genomes.

Chromatin Immunoprecipitation

ZNF526 ChIP samples were prepared from stable Neuroblasoma SH-SY5Y cells as follows. Cultures of 1 x 10^8 SH-SY5Y cells were harvested at a density of 1 x 10^6 cells/ml cells and cross-linked with 11% formaldehyde for 10 min at room temperature. Cross-linking was stopped by the addition of glycine to 2.5 M final concentration, and cells were washed twice with 2 x phosphate-buffered saline. The cell pellet was then harvested with cellscraper and suspended in DPBS buffer and centrifuged at 4°C for 5 min The cell pellet was then resuspended in 2 ml of ChIP lysis buffer I and rocked at 4°C for 10 min then centrifuge them at 1350 g for 5 min at 4°C. The cell pellet was resuspended in lysis buffer II followed the same procedure like lysis buffer I. Then the cell pellets have been resuspended in lysis buffer III and keep on ice. Samples were sonicated for 30 min with 30 μs pulses and 30 sec of resting using the Bioruptor sonicator (30sec ON/ 30sec off/high 35 min) to produce chromatin fragments of 0.5 kb on average. 50 μl of the sonicated DNA was saved as input DNA at -20°C.
1/10 volume of 10% Triton was added to the sonicated lysate and centrifuge 10 min at 4°C. Dynal beads has been washed 3 times with blocking solution by Placing the tubes on a magnet (Dynal® MPC). 100 ml V5-antibody and magnetic beads mix was added to the cell lysates and incubated a 4°C overnight.then the beads collected by MPC and washed 5 times with wash buffer after washing the pellet with buffer (TE, 50 mM NaCl) centrifuge has been done on the samples 3 min 960 g. Elution buffer was aed and incubated to 65°C for 30 min followed by centrifugation 16000 at room temperature for 1 min. Reversal of crosslink has been done by incubation of the supernatant and also the input DNA at 65°C for over night. Incubation with TE buffer and RNase A for 2 hours at 37°C has been done and followed by adding proteinase K and incubation at 2 hours in 55°C. Then purification has been done with phenol- chloroform procedure. After reversal of crosslinks and RNase treatment, ChIP DNA was purified and used directly for Solexa sequencing.

Solexa sequencing

Next generation sequencing (NGS) have been introduced as a cost-effective and fast strategy for comprehensive mutation screening and gene finding in the coding portion of the human genome. Next generation sequencing was performed using an Illumina Genome Analyzer II.

3 Very Low-Density lipoprotein Receptor Gene (*VLDLR*)

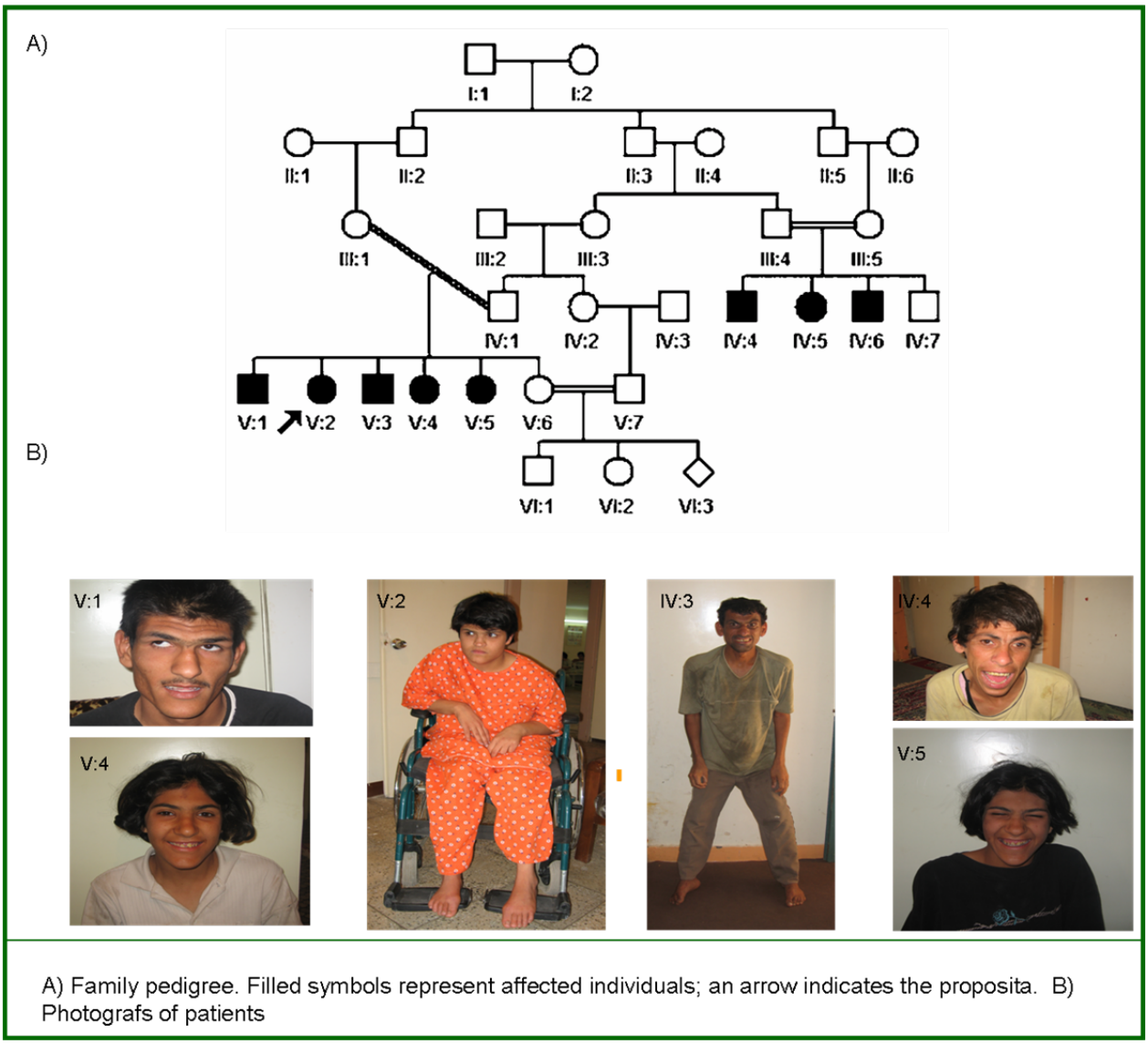
There are several examples of conditions that originally were considered to be nonsyndromic before detailed clinical investigations revealed that they have syndromic features. This illustrates that in many cases an exact discrimination between syndromic and non-syndromic forms of ID is not easy. Judging from the relative frequencies of syndromic and non-syndromic X-linked ID, syndromic forms of ARID are probably more common than non-syndromic ones. Moreover, searching for the relevant gene is usually easier in syndromic ID, as it is often guided by clinical signs suggesting specific spatio-temporal gene expression patterns and in some cases, clinical features are specific enough to establish a tentative diagnosis. This was the case in family M302, where clinical symptoms pointed to Dysequilibrium syndrome (DES, OMIM 224050). Dysequilibrium syndrome (DES, OMIM 224050) is a descriptive term that encompasses a spectrum of genetically hetereogeneous clinical conditions. It is a rare autosomal recessive non-progressive cerebellar disorder characterized by ataxia, mental retardation, cerebellar hypoplasia and some patients show strabismus, seizures and short stature. Previously, a homozygous 199 kb deletion encompassing the entire coding region of the *VLDLR* gene was detected in several DES patients from the Hutterite population in North America(Boycott 2005). As this deletion also contains an additional brain expressed gene (LOC 401491), the authors could not entirely rule out the possibility that this gene also be contributs to the phenotype. Furthermore, the size of the deletion did not allow excluding the causal involvement of intergenic regulatory elements. We now describe the detection of a homozygous stop mutation in the *VLDLR* gene in a large consanguineous Iranian family with DES, providing the first evidence that *VLDLR*deficiency is exclusively responsible for the DES phenotype.

3.1 The investigation of family M309

3.1.1 Clinical description

The family was recruited within the framework of a continuous collaborative project that aims at the elucidation of the genetic background of mental retardation in Iran. The pedigree is shown in Figure 3.1.

Eight family members (IV:4; IV:5; IV:6; V:1; V:2; V:3; V:4; V:5) suffered from moderate to severe mental retardation. They had either no speech at all or spoke just few words. Motor development was also retarded. They were able to sit independently between the ages of 12 and 24 months, but no patient could walk independently. All patients had strabismus and a body height below the 25th centile; five patients had a body height at or below the 3rd centile. Head circumferences were normal, and neither dysmorphic facial features nor other malformations



A) Family pedigree. Filled symbols represent affected individuals; an arrow indicates the proposita. B) Photografs of patients

Figure 3.1:

were observed. Seizures have not been reported for any of the patients. Unfortunately, no images of the brain by MRI or CT scan could be obtained, and putative cerebellar hypoplasia can only be inferred from the neurological defects. The clinical features are summarized in Table 3.1.

Autozygosity mapping in this family led to the identification of a single linkage interval, on

Patient	*Sex*	*Age at examination*	*MR*	*Gait ataxia*	*Strabismus*	*Height (centile)*	*OFC[a] (cm)*
IV:4	M	31	Severe	+	+	<3rd	53
IV:5	F	28	Moderate	+	+	<3rd	52
IV:6	M	15	Moderate	+	+	<3rd	52
V:1	M	27	Severe	+	+	<25th	56
V:2	F	26	Severe	+	+	10th	53
V:3	M	23	Moderate	+	+	3rd	54
V:4	F	15	Moderate	+	+	5th	52
V:5	F	15	Moderate	+	+	<3rd	50

Table 3.1: Clinical charastristics of family M309

chromosome 9 (Figure 3.2). The region on chromosome 9p24.2- 24.3, which comprised approx-

imately 3.7 Mb between the flanking markers rs1532309 and rs4131424 contains the *VLDLR* gene. As *VLDLR* has previously been implicated in the aetiology of DES2 we selected it for mu-

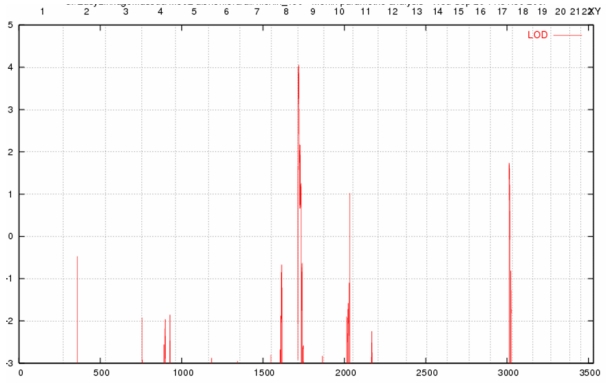
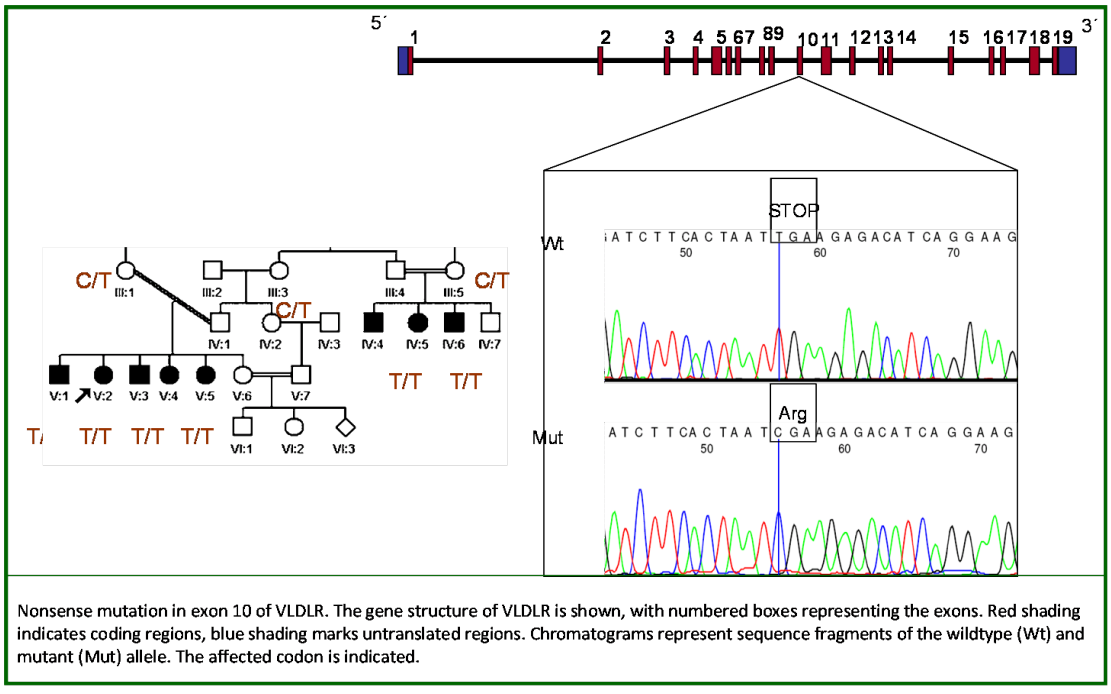


Figure 3.2: Solitary linkage interval on chromosome 9 contains VLDLR gene

tation screening. We hypothesized that a gene involved in neural development, cell positioning in brain, and cerebellar maturation could be involved in the pathogenesis of mental retardation. For *VLDLR* mutation screening, primer sequences were generated for all coding and adjacent splice site regions using the ExonPrimer software (http://ihg.gsf.de/ihg/ExonPrimer.html). PCR amplification was carried out using a standard touchdown PCR protocol. After successful amplification PCR products were sequenced on an ABI sequencer (Applied Biosystems, Foster City, CA, USA). For sequence analysis we used the CodonCode Aligner software (CodonCode Corporation, Dedham, MA, USA). A homozygous c.1342C>T nucleotide substitution was found (Figure 3.3), which introduces a premature stop codon in exon 10 (R448X).



Nonsense mutation in exon 10 of VLDLR. The gene structure of VLDLR is shown, with numbered boxes representing the exons. Red shading indicates coding regions, blue shading marks untranslated regions. Chromatograms represent sequence fragments of the wildtype (Wt) and mutant (Mut) allele. The affected codon is indicated.

Figure 3.3:

Homozygosity for the *VLDLR* mutations was perfectly coinherited with the disease and was absent in 100 healthy Iranian controls. It is located within a highly conserved region in the extracellular domain; this codon is completely conserved. *VLDLR* consists of 19 exons and encodes a 873 amino acid receptor protein of the low-density lipoprotein receptor superfamily [Strickland et al., 1995].

The low density lipoprotein (LDL) receptor family consists of structurally related endocytic receptors that include the LDL receptor, apoE receptor 2, LDL receptor-related protein (LRP also know as LRP-1), megalin (also known as gp330 or LRP- 2) and the very low density lipoprotein (VLDL) receptor. The VLDL receptor was identified by homology cloning [Takahashi et al., 1992] and is most abundant in skeletal muscle, heart, adipose tissue, and brain [Gafvels et al., 1993; Sakai et al., 1994; Webb et al., 1992]. Initially, the VLDL receptor was thought to function in the delivery of triglyceride-rich lipoproteins to peripheral tissues [Takahashi et al., 1992; Sakai et al., 1994]. However, in mice the VLDL receptor is likely not a major receptor. The chicken homologue of the human VLDL, LR8, play a critical role in mediating the transport of triglycerides into growing oocytes.

The VLDL receptor is closely related to the LDL receptor both in primary sequence as well as in the intron/exon organization of the gene [Sakai et al., 1994]. The primary structural difference between these two receptors is an additional exon in the VLDL receptor gene that encodes for the first cysteine rich class A repeat (L1) found at the aminoterminal region of the VLDL receptor.Despite the structural similarity between the LDL and VLDL receptors, their ligand binding properties differ considerably. The LDL receptor binds apoE and apoB containing lipoproteins, whereas the VLDL receptor binds to apoE-containing lipoproteins, but does not bind to LDL.

VLDLR is highly similar to LDLR in structure, but is different from LDLR in function [Strickland et al., 1995]. VLDLR contains five domains: an amino-terminal ligand-binding domain comprised of multiple cysteine-rich repeats, an epidermal growth factor (EGF) precursor homology domain, an O-linked sugar domain with a cluster of serine and threonine residues, a transmembrane domain and a cytoplasmic domain with an NPXY sequence [Chen et al., 1990]. VLDLR has two isoforms: the full length version, type I, and a version lacking an O-linked sugar region, type II [Sakai et al., 1994]. VLDLR type I is mainly distributed in heart and skeletal muscles with active fatty acid metabolism, whereas VLDLR type II is predominant in non-muscle tissue, including kidney,spleen, adrenal gland, lung, brain, testis, uterus and ovary, except liver [Takahashi et al., 1992; Webb et al., 1994]. Differences in ligand specificity and tissue distribution suggest, that the VLDLR isoforms play distinct roles in various tissues and cells. This mutation result in the partial loss of YWTD domain and loss of the o-linked sugar, transmembrane, and cytoplasmic domains. The functional loss associated with mutation at this position is likely a result of misfolding and impaired export of the receptors to the cell surface [Hobbs et al., 1992; Boycott et al., 2009].

The identification of the *VLDLR* mutation, R448X, provides molecular insight into the pathogenesis of neurodevelopmental disorders and expands the scope of diseases caused by mutations in components of the reelin pathway. In the early development of the mammalian brain, neuroblasts migrate from the ventricular zone to the. cortical plate along the pre-existing radial fibers. Neuroblasts rest on the cortical plate as an inside-out array. The layering and positioning of neuroblasts necessitate the large extracellular protein Reelin and the cytoplas-

mic mammalian Disabled (mDab1) activating tyrosine kinases. Two receptors, VLDLR and apoE receptor 2, have been reported as Reelin receptors. Reelin is a secreted glycoprotein that regulates neuronal positioning in cortical brain structures and the migration of neurons along the radial glial fiber network by binding to lipoprotein receptors VLDLR and APOER2 and the adapter protein DAB1 [Ozcelik et al.,2008].

4 Zinc Finger Protein 526

MRT11 is one of the 6 hotspot loci for autosomal recessive Intellectual disability. This locus is located on the long arm of chromosome 19q13.2-13.31 (Figure 4.1).

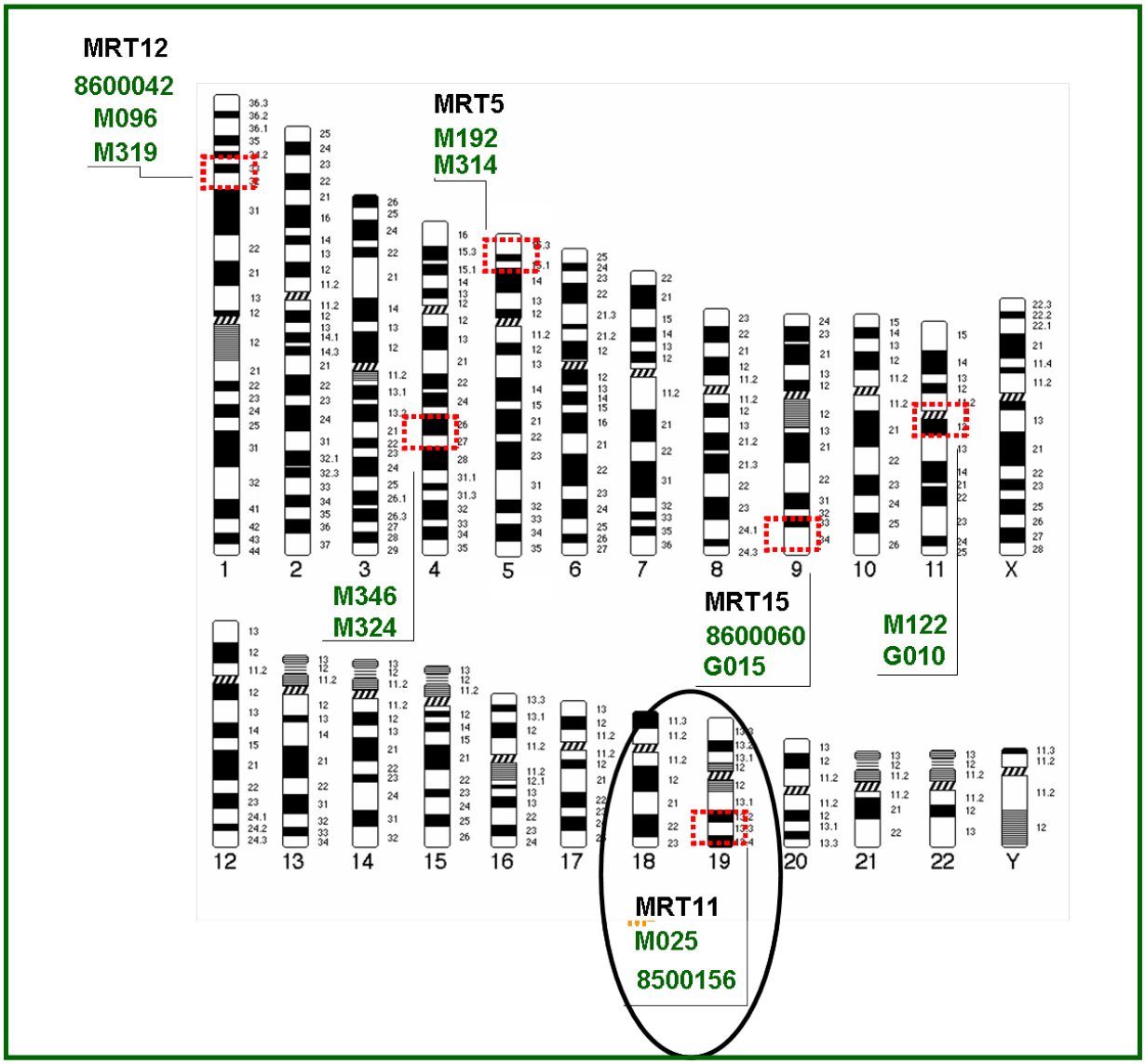
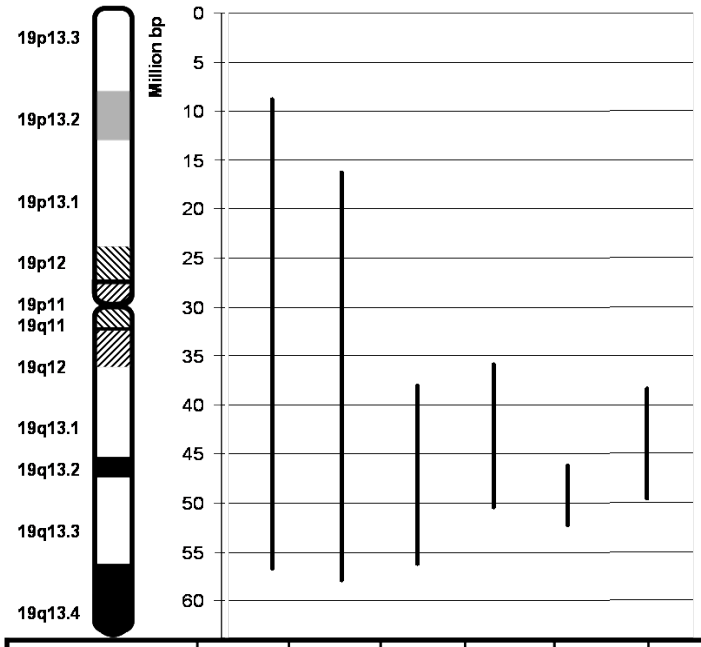


Figure 4.1: Idiograms of all human chromosomes are shown. The MRT 11 is specially labelled. Two families with solitary linkage intervals (M025, 8500156) are indicated.

The region on chromosome 19q13.2-q13.31 spans 3.5 Mbp and is common between the solitary linkage intervals of families M025 and 8500156 and one of the several linkage intervals of families M037, M196, M004 and M141 (Figure 4.2).



Families	M141	M004	M196	M037	M025	8500156
No.intervals	4	3	4	9	1	1

Figure 4.2: Chromatogram of chromosome 19 and the families with intervals overlapping the MRT11 locus. Two families, M025 and 8500156, have solitary intervals.

Families M-141 and M-196, apart from the interval overlapped with MRT11 locus, have 3 other homozygote regions in the genome. In the case of family M-141, one of the additional intervals overlaps with MRT5 locus (Figure 4.2). The highest number of intervals, containing eight homozygote regions, belongs to family M-037.

4.0.2 Clinical description

Family M025 is a large family with two family branches. In each branch two out of four offspring suffer from intellectual disability. The DNA samples were obtained from all siblings and the parents of both branches. The family is originally from Shahrekord, a city of about 130,000 inhabitants in the Western part of Iran.

The pedigree and facial aspects of the patients in family M025 are shown in Figure 4.3. The degree of ID in the affected family members ranged from moderate to severe. The patients showed no neurological problems, congenital malformations, or facial dysmorphisms. Head circumference, body height and weight were normal (Table 4.2)

4.0.3 Molecular analysis

Genotyping and linkage analysis

Affymetrix Gene chip human mapping array 50K with approximately 50,000 SNP markers has been used for whole genome SNP genotyping of all individuals in both branches. The solitary linkage interval has been identified using parametric linkage analysis of genotyping data. The maximum LOD score is 4 and the interval is located on the long arm of chromosome 19q13.2-

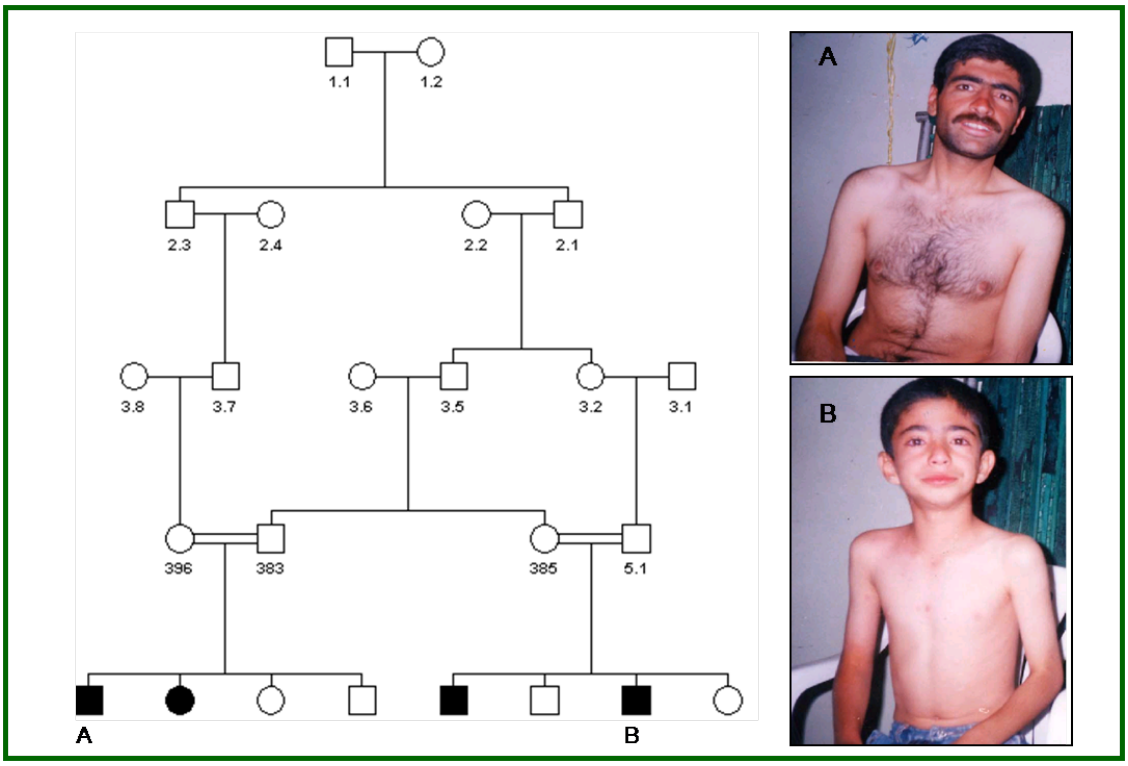


Figure 4.3: Family M025. Pedigree and clinical photographs of the patients.

Patient	Sex	Age at examination	Mental Retardation / IQ	Height	OFC[a]	Additional features
V:3	m	26 y	Severe	170 cm	55 cm	Speech problem
V:4	m	.....	...	165 cm	56 cm	.....
V:14	f	17 y	...	148 cm	53.5 cm	.....
V:16	m	9 y	moderate	...	....	Speech problem

a: OFC, occipitofrontal circumference

Table 4.1: Clinical information of the affected individuals in family M025

13.32, spanning 6.1 Mbp [Najmabadi et al., 2007].

Chromosome 19 is a gene rich chromosome; hence this small linkage interval (6.1 Mbp) contains around 156 known genes. Prior to mutation screening of the MRT11 locus, the genes have been ranked using pedestrian as well as bioinformatic programs like PosMed. The gene ranking was based on the function and the expression pattern of the genes in the central nervous system. The coding exons, exon-intron boundaries and also the untranslated regions (UTRs) of the prioritized genes have been amplified and sequenced using specific primers.

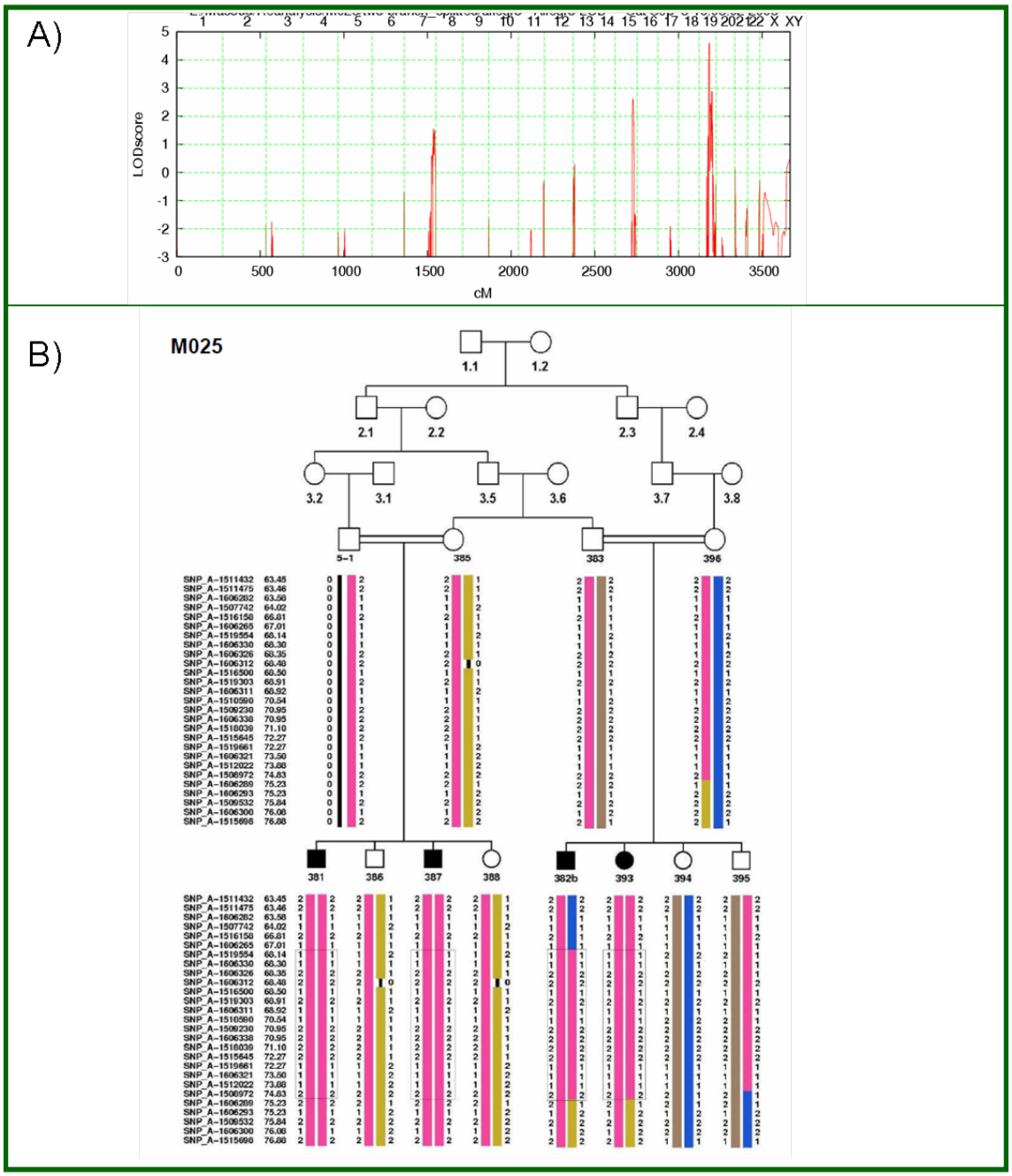


Figure 4.4: Family M025. Whole genome parametric (Allegro) linkage results (A) and haplotype of the single linkage interval with significant parametric LOD score of 4 (B)

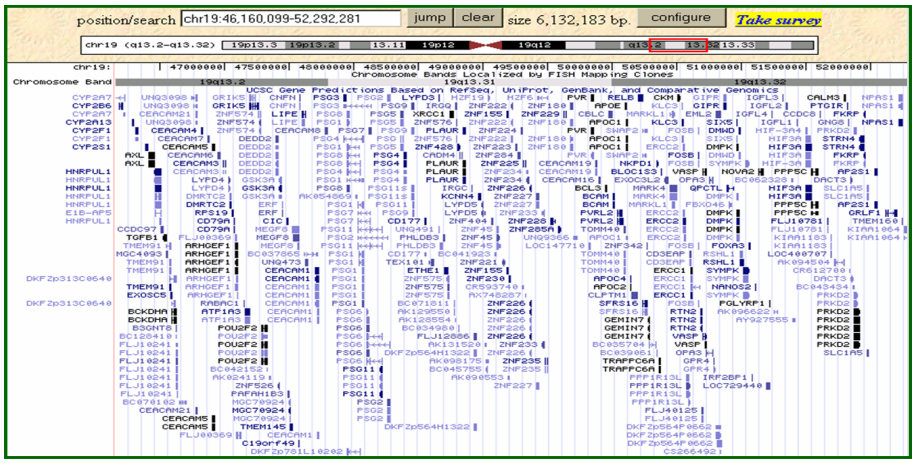


Figure 4.5: Gene content of the 6.1 Mb interval. The interval contains 156 genes (UCSC Genome Browser; http:// genome.ucsc.edu)

Identification of the mutation

Mutation screening of the genes inside the interval led to the identification of a G to A substitution (c.1224G>A) - a missense change R459Q - in ZNF526, a zinc finger protein , which co-segregated with ID in both branches of family M025 (Figure 4.6).

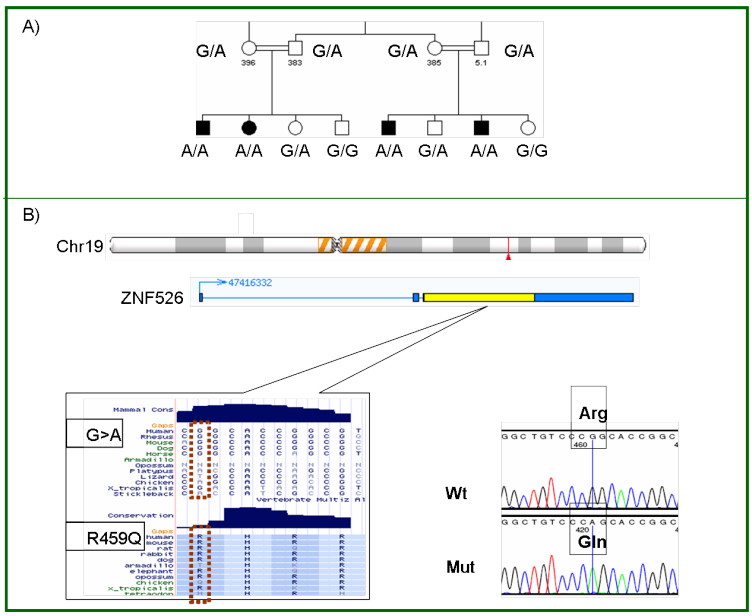


Figure 4.6: A) co-segregation of change (G>A) with ID in family M025. B) Schematic view of *ZNF526* gene; The mutation is located in a zinc finger domain highluy conserved in higher vertebrates

This variant could not be found in 120 Iranian normal individuals and 166 German controls, which were available as a control panel. The change was also not detected in 1000 genome project (63 Europian, 62 west Africans, 60 east Asians) and 200 Danish exome. Solexa sequencing of the entire interval confirmed this change as the single relevant variant in the interval. ZNF526 is a transcription factor protein containing C2H2 domains. It is located on chromosome 19 and spans 7862 nucleotides. This gene contains 3 exons with just one coding exon.

The Protein has a size of 670 amino acids with 14 C2H2 zinc finger structure.

4.1 The investigation of family 8500156

4.1.1 Clinical description

The second family with an overlapping solitary interval on the MRT11 locus is family 8500156. This family contains two branches. The core family has three intellectual disabled siblings and the small branch has just one offspring who is also affected (Figure 4.7).

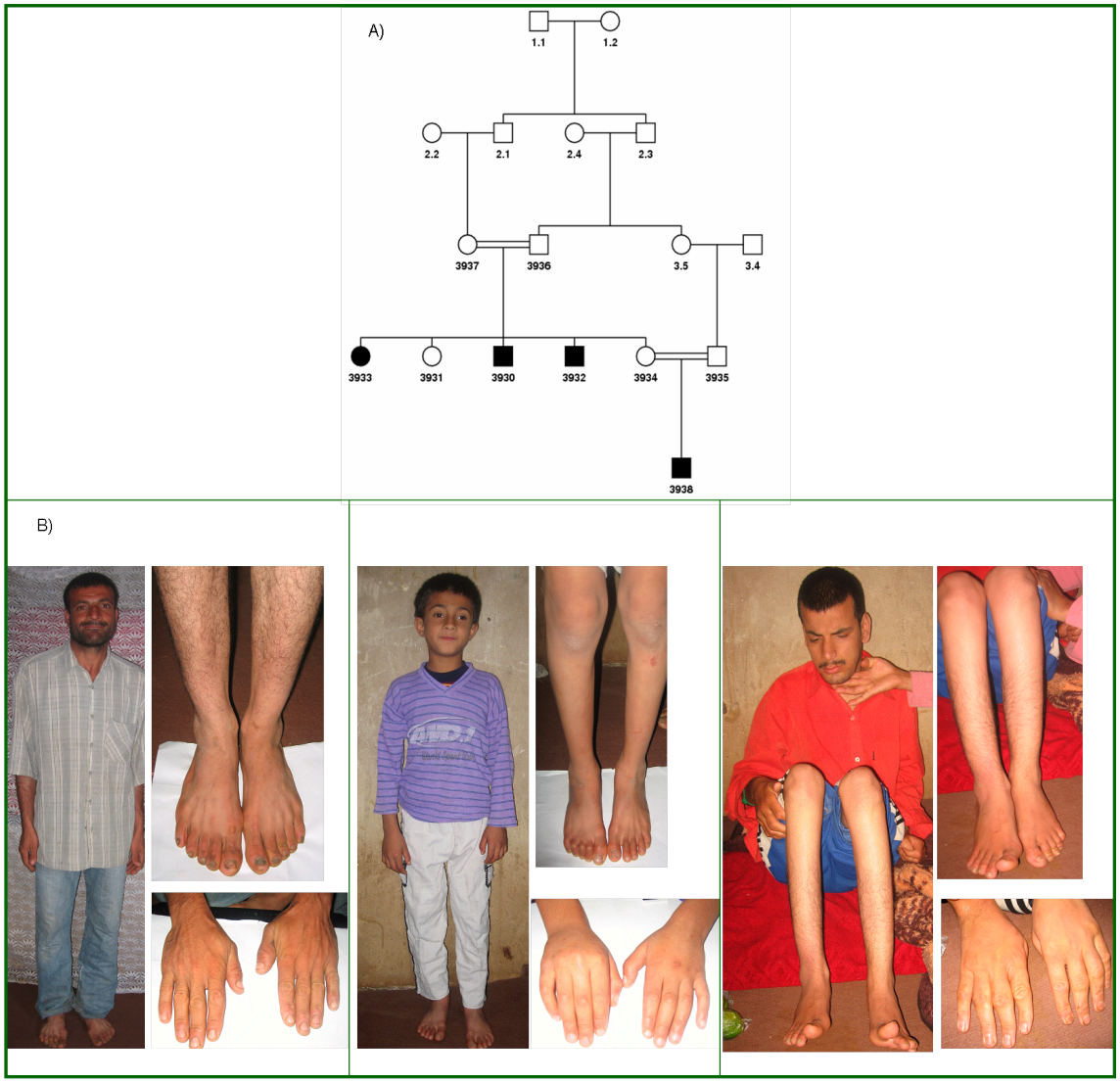


Figure 4.7: Family 8500156, A) Pedigree B) Clinical photographs of patients.

Patient	Sex	Age at examination	Mental Retardation / IQ	Height	OFC[a]	Additional features
V:1	m	33 y	48	163 cm	55 cm	---
V:3	m	24 y	22	161 cm	54.50 cm	---
V:6	f	14 y	24	150 cm	54 cm	Foot deformity
IV:1	m	10 y	54	123 cm	49 cm	---

a: OFC, occipitofrontal circumference

Table 4.2: Clinical data of the patients in family 8500156.

4.1.2 Molecular analysis

Genotyping and linkage analysis

Nine DNA samples from this family were genotyped using 250k SNP Affymetrix human array. The genotyping results of two affected and two healthy siblings, and the parents from the core branch as well as the whole small branch were used for homozygosity mapping. One single linkage interval with LOD score 3.5 was found on the long arm of chromosome 19 (MRT11), which spans 11329160 bp and is located between the two SNP markers rs11881580 and rs17727484.

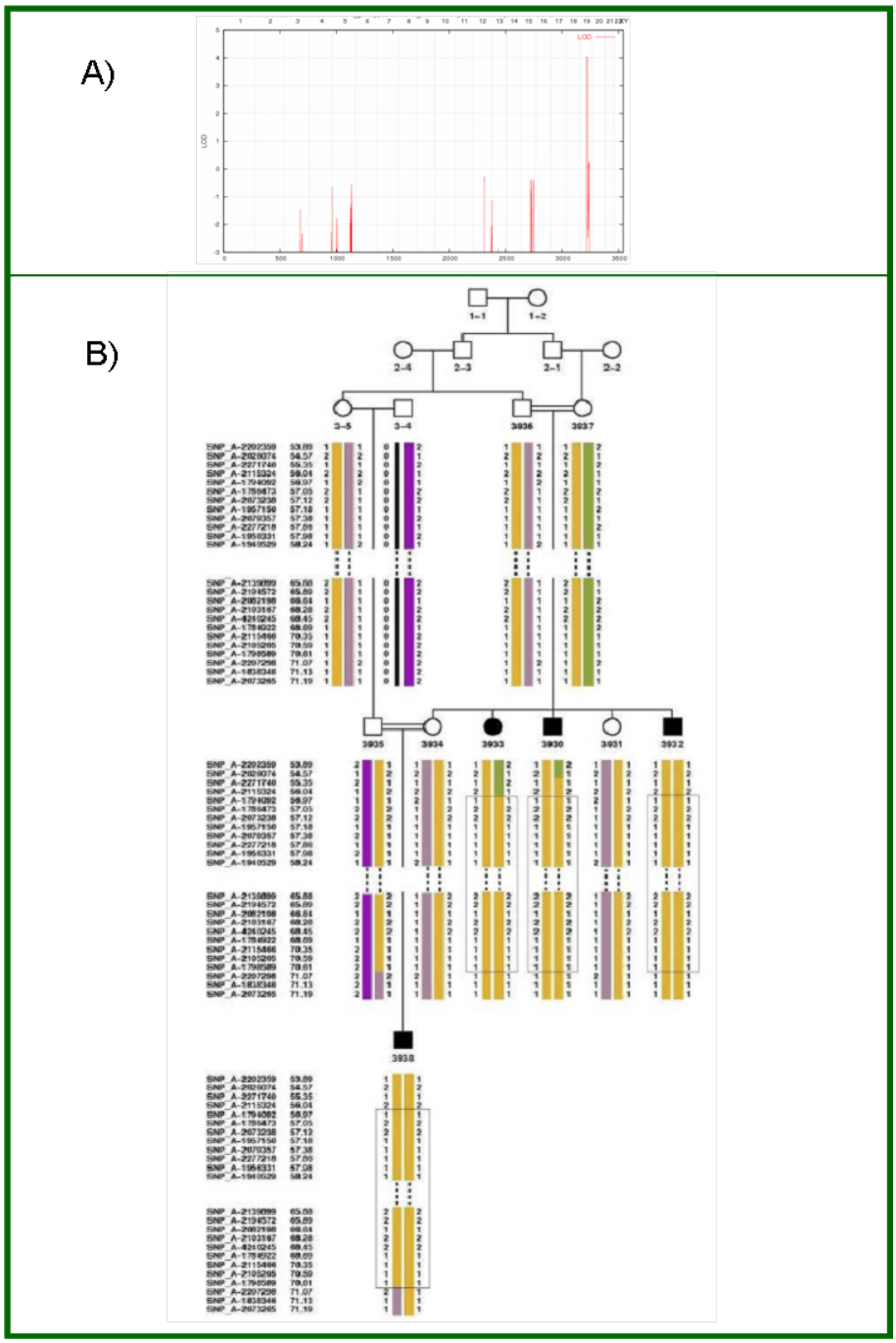


Figure 4.8: Family 8500156. Whole genome parametric (Allegro) linkage results indicate one interval on chromosome 19 (A) Haplotype of the single linkage interval (B)

4.1.3 Mutation screening

8500156 is the second family for MRT11 locus and has just one solitary linkage interval. Interestingly, Zinc finger 526, the gene mutated in family M025, is among the genes in the interval for family 8500156. *ZNF526* was sequenced using the same specific primers.

Identification of the mutation

A nucleotide transversion G to C has been found in the coding exon of *ZNF526*. This change creates a missense mutation Q539H inside a C2H2 finger domain. The mutation segregates with the disease in the family and may not be considered as a polymorphism, because it was not found in the control panels. The change was detected as the solitary putative variant in the entire interval by Solexa sequencing. cDNA position is conserved in higher vertebrate. (Figure 4.8).

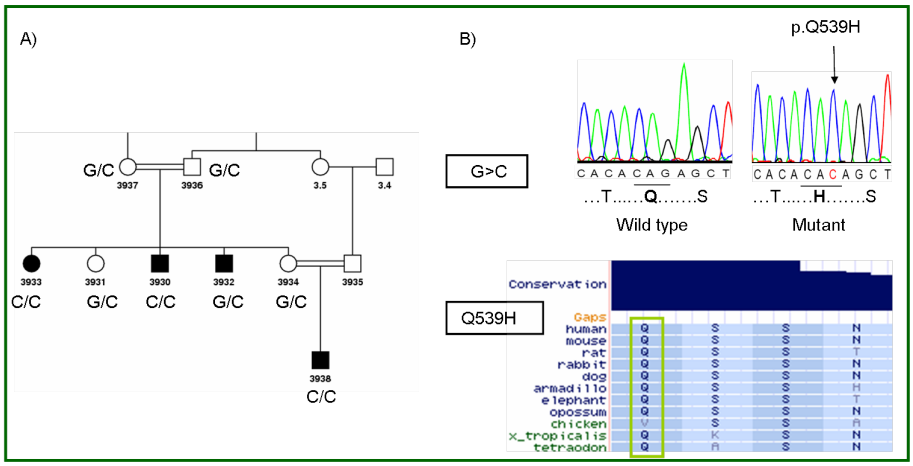


Figure 4.9: Segregation of the mutation in the affected individuals. B) Chromatograms of the control and patient; Glutamin in this position conserved in different species.

4.2 The investigation of family M037

4.2.1 Clinical description

Family M037 has two branches and this family contains three affected individuals. The proband shows hearing and speech problems, hand shaking and has an IQ less than 25. This family is from the same city as family M025.

4.2.2 Molecular analysis

DNA samples of all affected and their parents were genotyped using 10K affymetrix array SNP and linkage analysis revealed nine intervals, one of them overlapping the MRT11 locus.

4.2.3 Gene-based mutation screening

Sequencing of *ZNF526* in the affected individuals revealed the identical mutation found in family M025. A transversion G to A inside of the coding exon creates the missense mutation R459Q within one C2H2 zinc finger domain of the protein. Two of the three families in which p.R459Q was found are from the same city. It is of note, that the haplotype of family M025 turned out to be identical to the haplotype from family M037, with several accompanying linkage intervals on other chromosomes. Overall, the maximal common region between the two families spans 3.6 MbP between the two SNP markers rs2109075 and rs386569 on Chromosome 19.

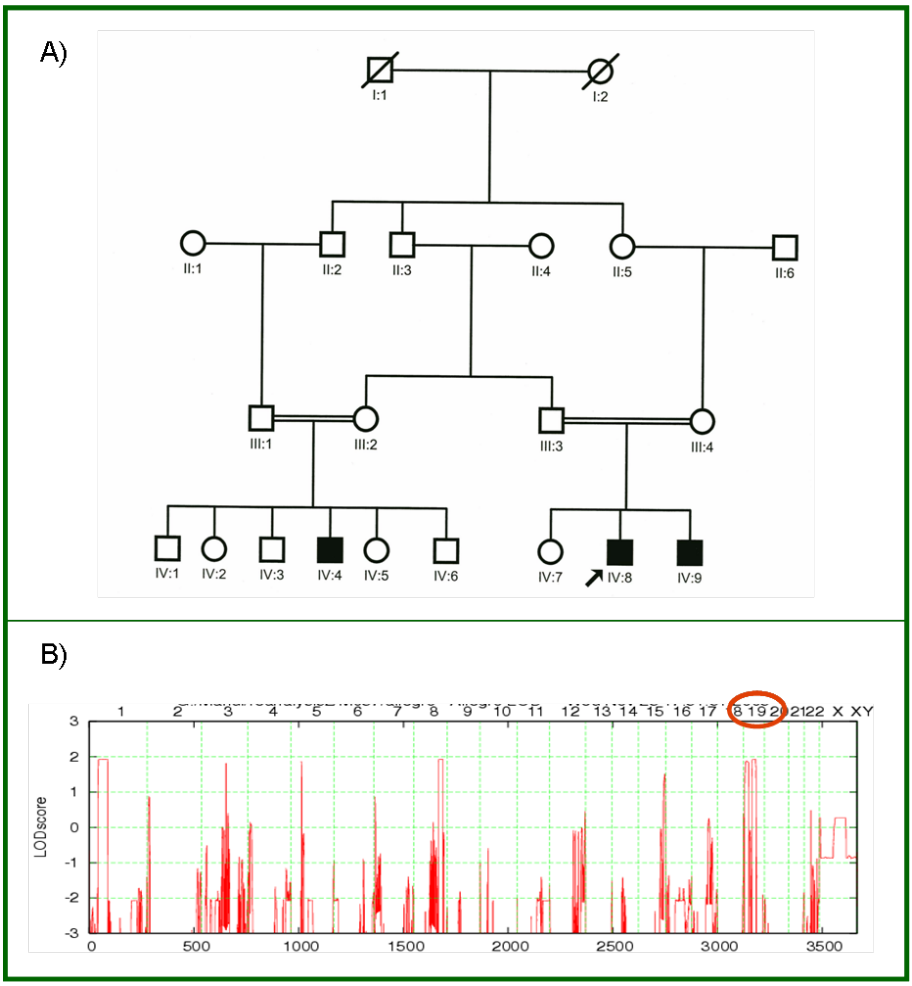


Figure 4.10: Pedigree and linkage analysis results for family M037

Therefore, an ancient founder mutation cannot be excluded and this might also indicate that the underlying mutation is evolutionarily old. Since both families live in the same town, the former possibility seems more likely, even though the families concerned are not aware of being related.

Out of six families with overlapping intervals on the MRT11 locus, missense mutations in

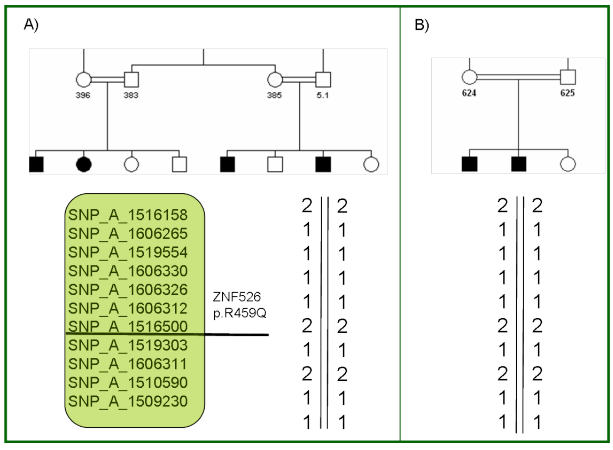
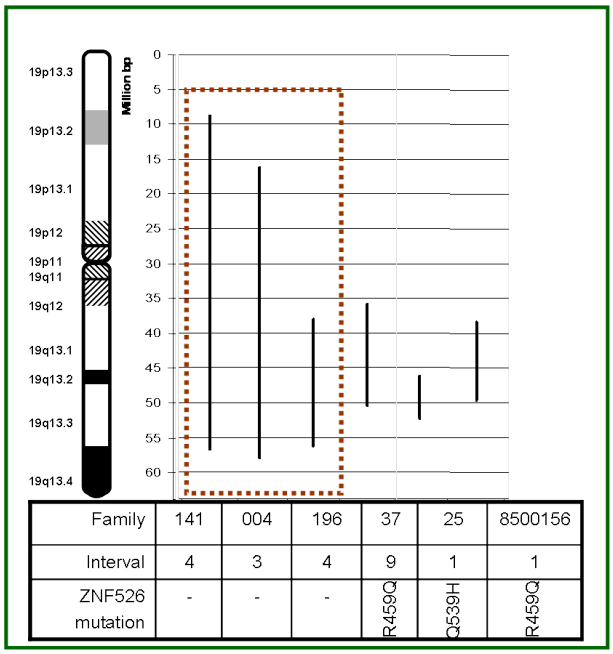


Figure 4.11: The 3.6 Mbp haplotype region for two families M025 and M037.

ZNF526 were found in three families. Two of these families showed the same haplotype and furthermore the identical missense mutation. The remaining three families have additional homozygote intervals in their genome, indicating the possibility of gene defects in these loci.



Family	141	004	196	37	25	8500156
Interval	4	3	4	9	1	1
ZNF526 mutation	-	-	-	R459Q	Q539H	R459Q

Figure 4.12: Chromatogram of chromsome 19 and the families with overlapping interval for MRT11 locus. Exons ans intron-exon boundries of ZNF526 in the families M141, M004 and M196 has no mutaion.

The pathogenic nature of both missense mutations is supported by the following arguments: (1) cosegregation with disease in the large ARID-affected families M025, M037 and 8500156; (2) absence in different control panels, including population controls; (3) conservation in several orthologs and paralogs of higher vertebrates; (4) Polyphen and Panther predictions, suggesting that p.R459Q and p.Q539H may affect protein structure or function.
There is no information available for this gene. *ZNF526* gene encodes a 670 aa protein (73 kDa) containing 14 ZNF C2H2 domains and similarly, belonging to the C2H2-type zinc-finger protein family. This protein may be involved in transcriptional regulation and it is located in the nucleus.

4.2.4 *ZNF526* expressed in the human brain

ZNF526 is thought to be ubiquitously expressed. RT-PCR was performed on fetal and adult RNA samples using specific primers. RNA was also extracted from control lymphoblastoid cell line, and RT-PCR was performed using the same primers (Appendix). Expression of the

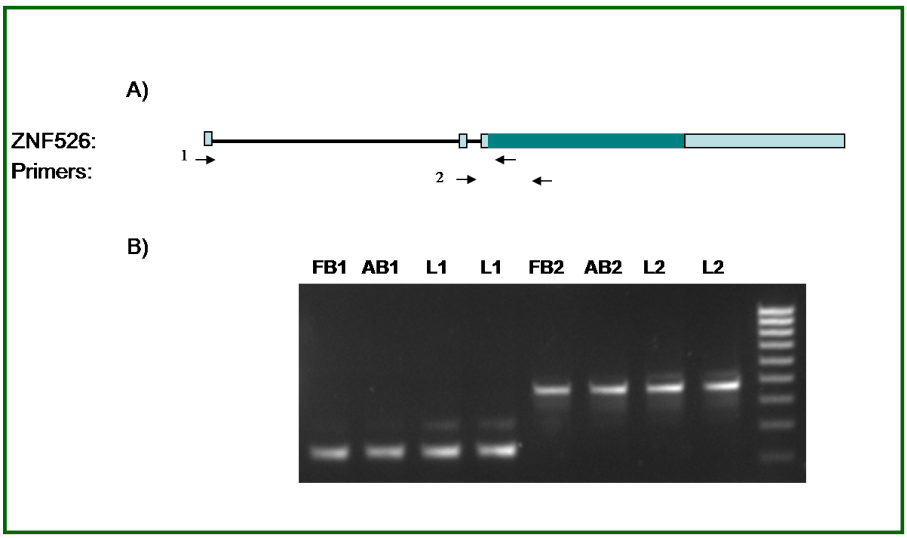


Figure 4.13: *ZNF526* expression in adult and fetal Brain and lymphoblasoid cell line. FB1, FB2;: Human Fetal Brain RNA, AB1,AB2; Human Adult Brain RNA, L1,L2; Lymphoblastoid RNA.

ZNF526 gene in developing brains suggests a possible role of this protein during development. Lymphoblastoid cell lines, the only patient tissue, are routinely available for functional studies.

4.2.5 Functional Prediction of amino acid-substituting *ZNF526* variations by different *in silico* tools

In the analysis of the mutations, two widely-used prediction programs, Polymorphism Phenotyping (PolyPhen) and Protein Analysis Through Evolutionary Relationships (Panther), were employed to predict functional significance of each of the two missense mutations in the *ZNF526* gene. Panther gives results in two prediction categories tolerated or deleterious effects while PolyPhen gives results in three categories: benign (probably lacking any phenotypic effect), possibly damaging, and probably damaging (should affect protein function). Prediction for ZNF526 mutations derived from Panther and PolyPhen both showed a significant correlation with the disease.

Neural network predictions of splice sites were performed using NetGene2. For analysis the reference sequence for *ZNF526* was altered at the respective mutated sites. NetGene2 was unable to detect the acceptor splice site or any change brought on by the mutation c.1224G>A. Netgene2 results predicted the introduction of novel exonic splicing enhancers in this mutation.

Using NetPhos2 software, a prediction analysis was performed of possible phosphorylation substrates. NetPhos 2.0 predictions indicate no introduction of novel(protein kinase C-specific) phosphorylation sites (Table 4.3).

In order to test experimentally the NetGene2 splicing prediction for G to A mutation, we es-

Mutation	Polyphen	Panther	NetGene 2	NetPhose 2.0
p.Arg459Gln	Probably damaging	deletiourus	One novel acceptor Splice site	.....
p.Gln539His	Probably damaging	deletiourus	.....	.....

Table 4.3: Evaluation of the pathogenic potential of missense mutations in the DNA binding domain of ZNF526. Overall, Polyphen and Panther predictions, sustain a pathogenic effect of p.R459Q and p.Q539H.

tablished RT-PCR using three specific primer pairs for different positions in the *ZNF526* gene. The third primer pair was designed to amplify *ZNF526*, just in case of the existence of novel splicing out, whereas only unspliced RNA from this region were amplified. RT-PCR gives rise of confirmation of no splice out.

Mutation R459Q maps to zinc finger 10 (ZF10), whereas the other *ZNF526* mutation, Q539H, is

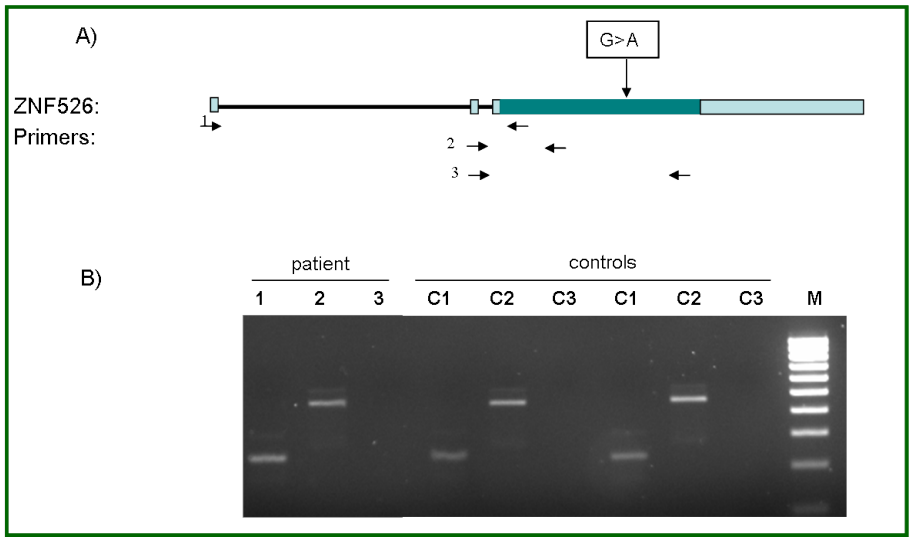


Figure 4.14: RT-PCR using patient and normal Lymphoblastoid cell line. 1, 2 and 3 are three different primer pairs. Patient and two different cntrol RNA has been used to amplify using thee three panel of primers.

located in ZF13 (Figure 4.14). The DNA binding domain of ZNF526 protein is composed of 14 C2H2-class ZFs (Figure 4.14). These types of ZFs insert into the major groove of DNA to make specific contacts with nucleotides by amino acids at positions -1, 2, 3, and 6. Both intellectual disability mutations (the ZF10-R429Q and the ZF13-Q539H) eliminate the guanine-contacting Arginine at positions 6 and glutamine at position -1, are respectively altered. Therefore, the amino acid changes for specific interaction with DNA (Figure 4.14). Thus, each of the two mutations results in a missense codon at a position predicted to be critical for ZF formation or DNA base recognition.

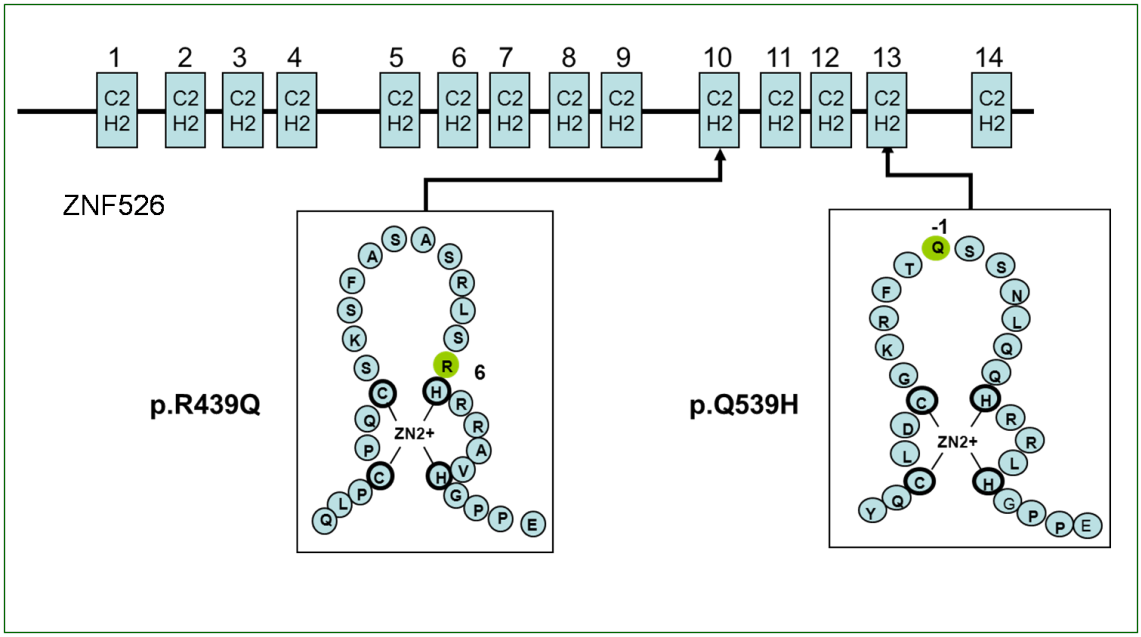


Figure 4.15: Relative position of missense changes

4.2.6 Analysis of ZNF526 subcellular localization in transfected Hela cells and neuroblastoma cell lines, SH-SY5Y

Translocation of ZNF526 to the nucleus is necessary to enable it to regulate transcription. To determine the subcellular distribution of ZNF526, we first fused the *ZNF526* cDNA in frame into vector pCMV2-Flag (Sigma) to create a Flag ZNF526 fusion protein. We detected the protein in transfected HeLa cells and Neuroblastoma cell line (SH-SY5Y) by using an anti-Flag monoclonal antibody (Sigma) and FITC-conjugated secondary antibody.
Analysis of the subcellular localization of ZNF526 by confocal microscopy showed that ZNF526 was localized predominantly to the nucleus (Figure 4.16). ZNF526 was localized to the nucleus in >95% of the cells in which ZNF526 was detected. Similar results were obtained in Neuroblastoma cells (SH-SY5Y). To determine which region is required for its nuclear localization, we examined the ZNF526 sequence for the presence of potential nuclear localization signals using pattern/motif analysis. This analysis identified a potential bipartite nuclear localization signal (bNLS) between Arg645 and Lys661, a region overlapping with ZF5. As shown in Figure 4.16, substitution of Arg459 with Glu and Glu539 with histidin did not affect the nucleolar localization of ZNF526. These data suggest that ZNF526 protein localize to the nucleus and further, that the mutations play no role in ZNF526 localization. To further analyze the localization of ZNF526, the effect of point mutations in ZNF526 in neuronal cells was examined. As shown in Figure 5.16, the ZNF526 mutants localize to the nucleus, suggesting that both changes have no major impact on ZNF526 localization in neuroblastoma cells. The same was shown for localization in Hela cells.

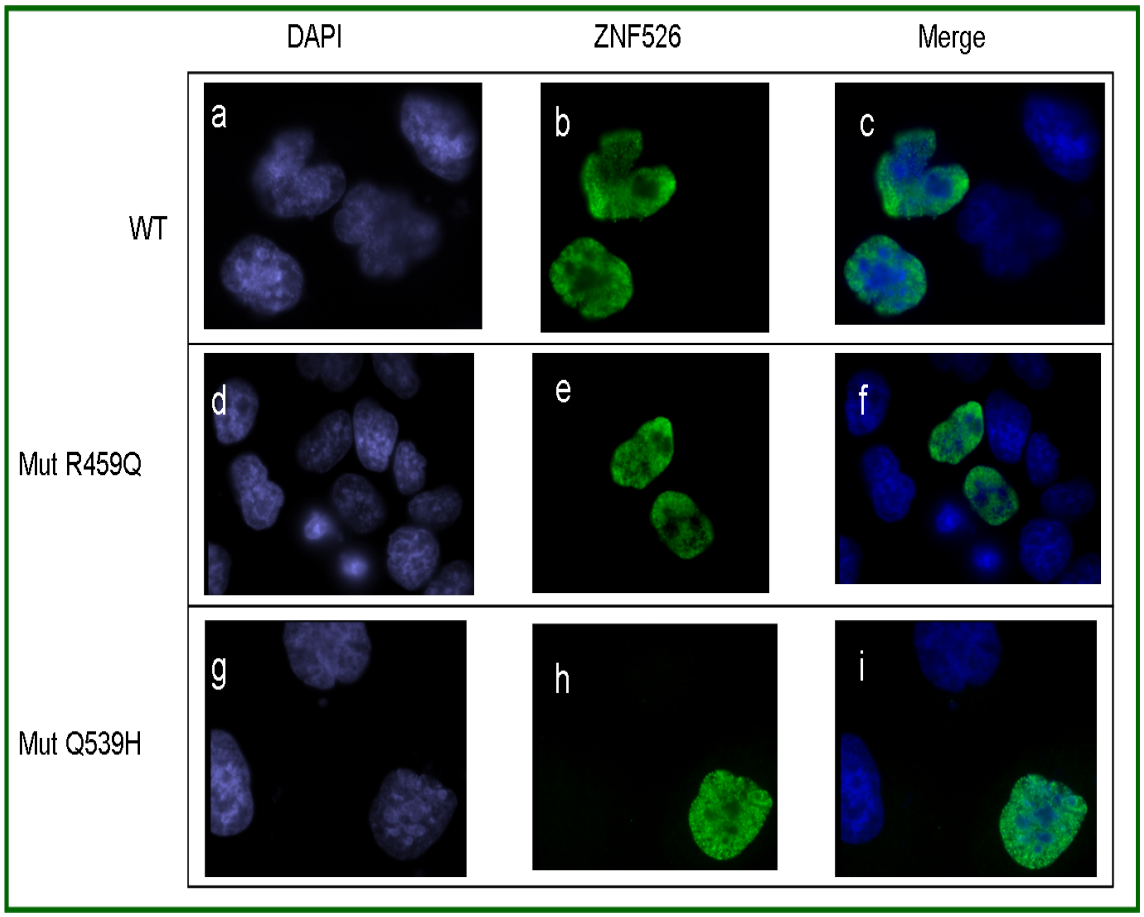


Figure 4.16: A) Subcellular localization of wild-type (WT) FLAG-ZNF526 (a-c), FLAG-ZNF526-R429Q (d-f), FLAG-ZNF526-Q539H (g-i) were examined by confocal microscopy with anti-FLAG antibody; nuclei were identified by DAPI staining. B) The localization is not affected by both type of mutations.

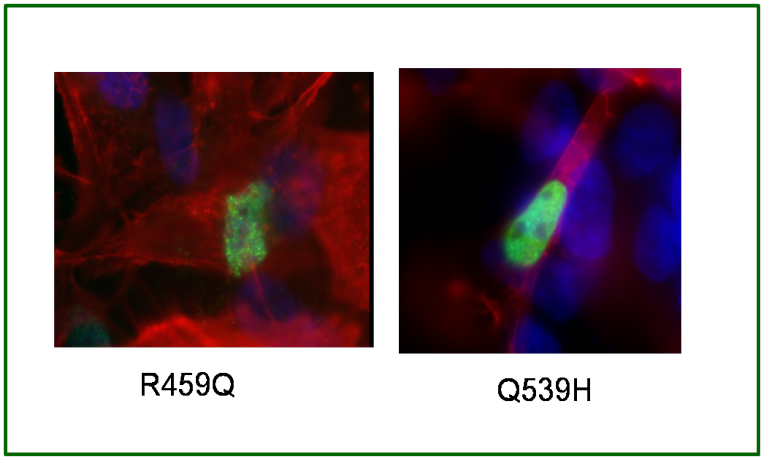


Figure 4.17: Neuroblastoma (SH-SY5Y) tranfection using PCMV2A-ZNF526 wt and mutants. anti-FITC antibody; nuclei were identified by DAPI staining and Tubulins by Phalloidin

4.2.7 Expression analysis in LCLs of intellectual disabled patients with *ZNF526* mutations

Localization of ZNF526 to the nucleus suggests transcriptional regulation via interaction with specific DNA-binding sites in the promoter regions of target genes. Hence, whole genome ex-

pression profiling was performed to investigate the effects of the *ZNF526* mutations on gene expression. Illumina Sentrix® Human-6 Expression BeadChips were employed to compare the expression levels of around 48000 transcripts from known and predicted human genes in patient lymphoblastoid cells and controls. In order to investigate the impact of the mutations on the transcriptional level, RNA was extracted from patient lymphoblastoid cell lines. RNA was isolated from two patients with the identical mutation p.R459Q, belonging to different families. RNA was also extracted from the patient carrying p.Q539H mutation and from three healthy controls.

The Differential score (Diff. score) parameters were used to determine gene expression in the Illumina BeadStudio software program. The Diff. Score is a transformation of the p-value that provides directionality to the p value based on the difference between the average signals in the reference group vs. the comparison group (see introduction). Data analyses was performed by grouping all the patients with different mutations and comparing them with the group of three controls, using the "Rank-Invariant" method of normalization and the "custom" algorithm of the bead studio software (Illumina).

This led to the identification of 2916 differentially expressed genes (diff. scores ≤ 13) for the patient from family M025, pR459Q, in comparison to the 4004 number of genes from the second patient from family M037. It was expected that a similar number of genes was differentially expressed in patients from both families, confirming the theory of common ancestor or the identical mutation. 2308 genes are differentially expressed in both patients with identical mutation. In total, 2345 genes show different expression than the control in the family 8500156, p.Q539H (Figure 4.18).

In our initial analysis, we focused on the total genes which are differentially regulated by

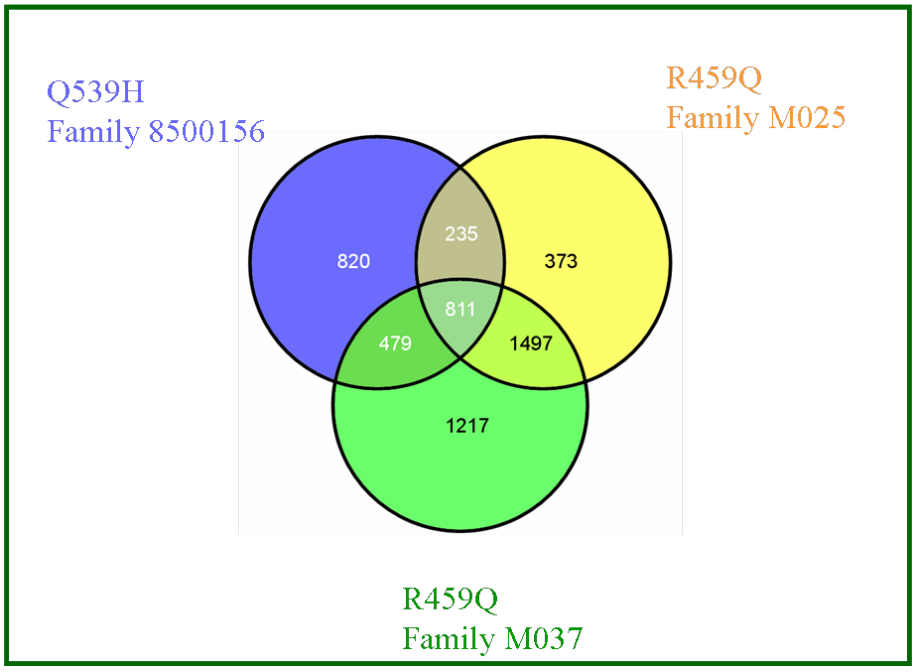


Figure 4.18: Venn Diagram showing the overlapping genes between three families with ZNF526 mutations. 811 genes are commonly differentiate expressed in all three mutated ZNF526 patients.

ZNF526 mutation. But ZNF526 can function as a transcriptional repressor as well as an activa-

tor; this results in up or down regulation of target genes. Using this approach, the resulting 811 differentially expressed genes common in all the patients were intensely analysed by Gene Ontology. These 811 differentially expressed genes have stringent diff. scores ≤ -13 (correspond to P-values ≤ 0.001). This experiment led to the identification of several promising potential targets of gene regulatory processes that involve ZNF526.

4.2.8 Functional annotation clustering of genes deregulated in ID patients with *ZNF526* mutations

In order to search for functionally related genes among the 811 genes differentially regulated in the patients, the DAVID and Ingenuity functional classification tools were employed.
The first annotation cluster with an enrichment factor (a Fisher exact test based statistical value for association between a set of genes and a specific annotation term) of 7.21 for the 628 interrogated genes out of 811 deregulated genes is shown in figure 4.19. This analysis showed highly

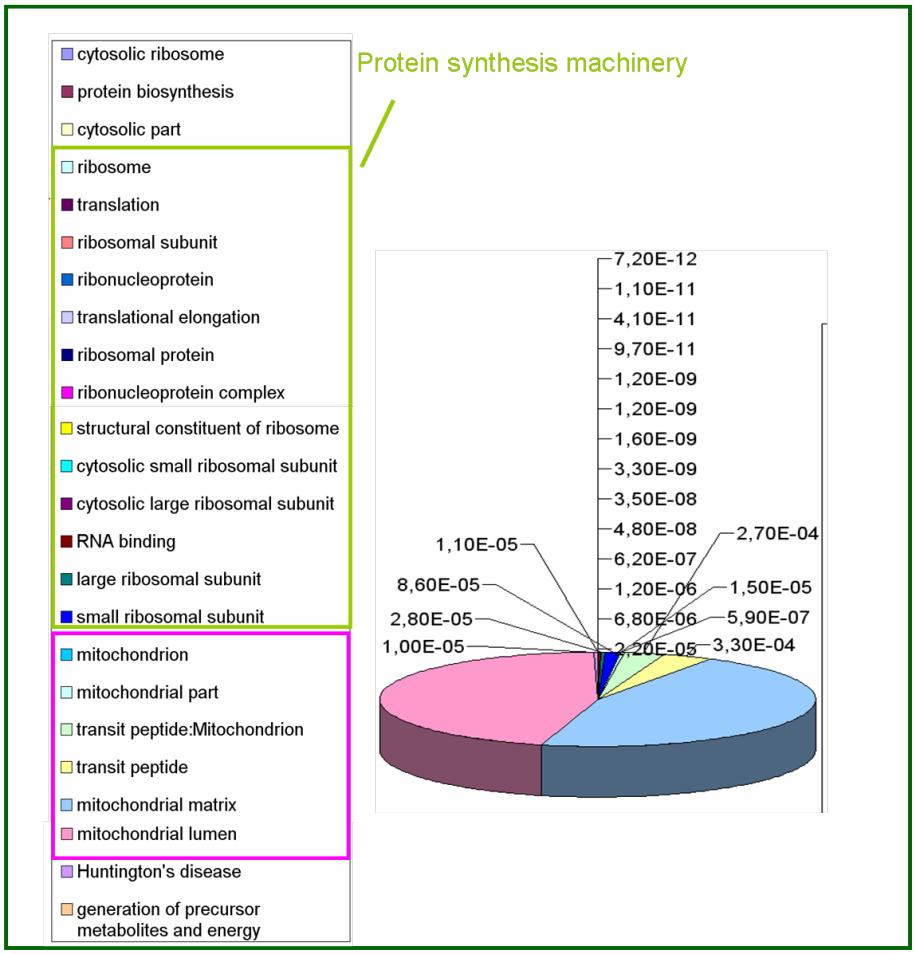


Figure 4.19: Chart representing the major biological functions and processes associated with *ZNF526* mutations. The resulting 811 differentially expressed genes were intensely analysed by Gene ontology. In cellular physiological process classification, the major part was "protein synthesis" followed by "mitochondrion disorders", "metabolites and energy metabolism".

significant (P-values $1 \times E - 12$) clustering of genes in expected pathways such as cytosolic ribosome and protein synthesis. According to the biology function classification, 21 of 628 genes

Annotation Cluster 1	Enrichment Score: 7.21			Count	P_Value
GOTERM_CC_FAT	cytosolic ribosome	RT		21	8.7E-12
SP_PIR_KEYWORDS	protein biosynthesis	RT		29	1.2E-11
GOTERM_CC_FAT	cytosolic part	RT		27	5.2E-11
SP_PIR_KEYWORDS	ribosome	RT		18	1.1E-10
GOTERM_BP_FAT	translation	RT		38	1.4E-9
GOTERM_CC_FAT	ribosomal subunit	RT		23	1.5E-9
SP_PIR_KEYWORDS	ribonucleoprotein	RT		32	1.9E-9
GOTERM_BP_FAT	translational elongation	RT		20	3.6E-9
SP_PIR_KEYWORDS	ribosomal protein	RT		24	4.0E-8
GOTERM_CC_FAT	ribonucleoprotein complex	RT		46	6.6E-8
GOTERM_CC_FAT	ribosome	RT		27	9.9E-8
KEGG_PATHWAY	Ribosome	RT		18	4.4E-7
GOTERM_MF_FAT	structural constituent of ribosome	RT		22	7.4E-7
GOTERM_CC_FAT	cytosolic small ribosomal subunit	RT		11	1.3E-6
GOTERM_CC_FAT	cytosolic large ribosomal subunit	RT		10	7.4E-6
GOTERM_MF_FAT	RNA binding	RT		49	3.0E-5
GOTERM_CC_FAT	large ribosomal subunit	RT		12	3.1E-5
GOTERM_CC_FAT	small ribosomal subunit	RT		11	9.4E-5
GOTERM_MF_FAT	structural molecule activity	RT		35	1.5E-2
Annotation Cluster 2	**Enrichment Score: 4.2**			**Count**	**P_Value**
SP_PIR_KEYWORDS	mitochondrion	RT		56	3.2E-7
GOTERM_CC_FAT	mitochondrial part	RT		45	8.5E-6
SP_PIR_KEYWORDS	transit peptide	RT		32	1.6E-4
GOTERM_CC_FAT	mitochondrial matrix	RT		19	1.9E-3
GOTERM_CC_FAT	mitochondrial lumen	RT		19	1.9E-3
Annotation Cluster 3	**Enrichment Score: 3.77**			**Count**	**P_Value**
SP_PIR_KEYWORDS	mitochondrion	RT		56	3.2E-7
GOTERM_CC_FAT	mitochondrial part	RT		45	8.5E-6
KEGG_PATHWAY	Huntington's disease	RT		24	1.0E-5
GOTERM_BP_FAT	generation of precursor metabolites and energy	RT		29	1.2E-5
SP_PIR_KEYWORDS	respiratory chain	RT		12	1.5E-5
SP_PIR_KEYWORDS	mitochondrion inner membrane	RT		20	1.5E-5
GOTERM_CC_FAT	mitochondrion	RT		68	1.9E-5
SP_PIR_KEYWORDS	ubiquinone	RT		8	3.6E-5
GOTERM_BP_FAT	oxidative phosphorylation	RT		14	5.6E-5
SP_PIR_KEYWORDS	electron transport	RT		13	8.5E-5
GOTERM_CC_FAT	respiratory chain	RT		12	9.0E-5
KEGG_PATHWAY	Parkinson's disease	RT		18	9.5E-5
KEGG_PATHWAY	Oxidative phosphorylation	RT		18	1.2E-4
GOTERM_BP_FAT	mitochondrial electron transport, NADH to ubiquinone	RT		9	1.2E-4
GOTERM_CC_FAT	NADH dehydrogenase complex	RT		9	1.3E-4
GOTERM_CC_FAT	respiratory chain complex I	RT		9	1.3E-4
GOTERM_CC_FAT	mitochondrial respiratory chain complex I	RT		9	1.3E-4
GOTERM_MF_FAT	NADH dehydrogenase activity	RT		9	1.3E-4
GOTERM_MF_FAT	NADH dehydrogenase (quinone) activity	RT		9	1.3E-4
GOTERM_MF_FAT	NADH dehydrogenase (ubiquinone) activity	RT		9	1.3E-4

Figure 4.20: Gene ontology annotation of the deregulated genes in lymphoblastoid cell lines of ID patients belonging to three families M025, M037 and 8500156

were associated with cytosolic ribosome, 29 were related to protein synthesis, and 56 were involved in mitochondrion function. For example, according to the KEGG database, 21 out of the 628 genes submitted to DAVID are involved in the ribosome with a P-value of $1.1 \times E - 12$, as shown in Figure 4.18. A highly significant enrichment of transcripts from the expected path-

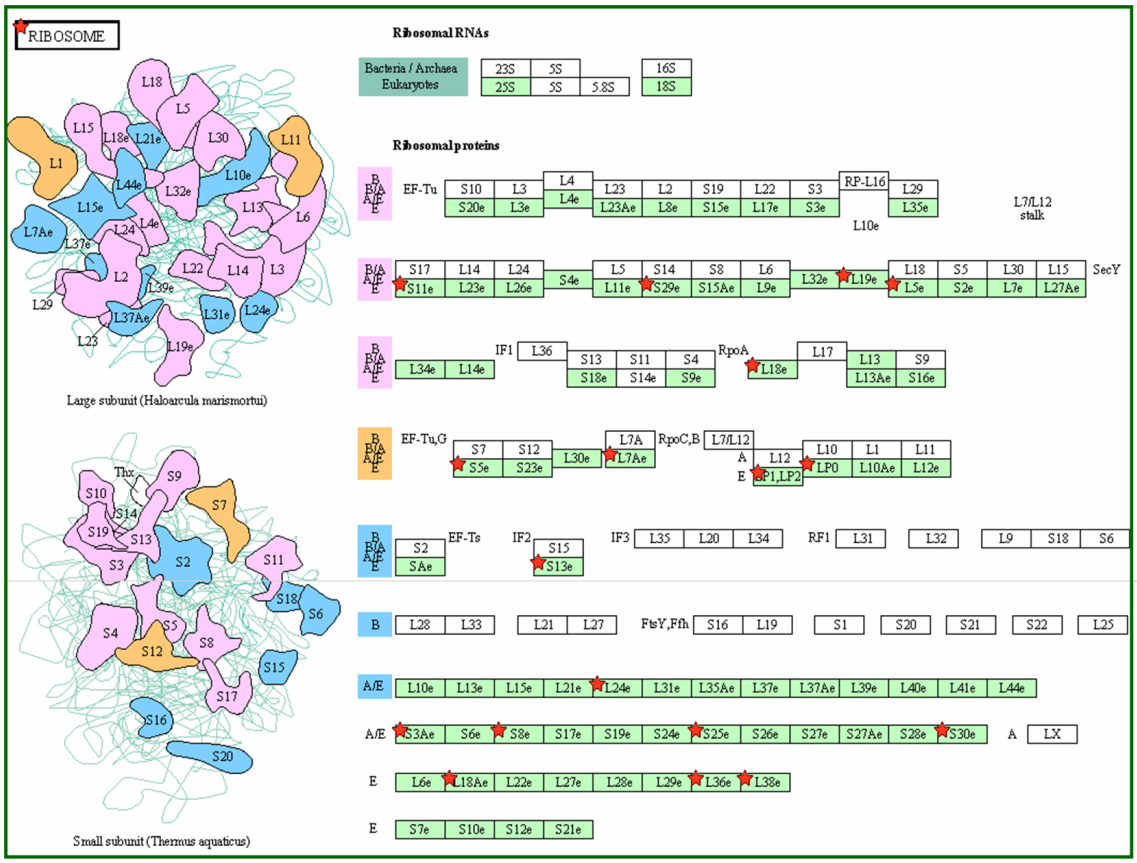


Figure 4.21: KEGG pathway for cell cycle. The 23 common genes (RPS30, RPL18, RPL18A, RPL19P12, RPL24P6, RPL36, RPL38,RPL5P1, RPL7A, RPS11P5, RPS13P2, RPS25, RP11, FTE1, RPS5, RPS8, RPL0, RPL1) between LCLs deregulated genes with diff. scores ≤ -13 (corresponding to P-values 0.05) and ribosome genes based on KEGG database are depicted by red asterisks.

ways such as huntington, chromosome segregation and DNA repair was obtained by using the Panther database. 24 genes with P-value $1.0 \times E - 10$ are involved in Huntington disease and 18 genes with P-value $9.5 \times E - 12$ were involved in Parkinson disease.

4.2.9 Generation of lentiviral *ZNF526* expression constructs and establishment of stable neuronal lines that overexpress *ZNF526* mutants

Taking advantage of a lentiviral expression system that can achieve stable gene expression, we developed a neuronal SH-SY5Y cell lines, overexpressing wild-type or R459Q or Q539H ZNF526. pLenti-LacZ has been used as a control vector, which was included in the pLenti6/V5 Directional TOPO Cloning kit.

pLenti6/V5 constructs were co-transfected with the ViraPower Packaging Mix into the 293FT cell line to produce lentivirus, and the obtained lentiviral stock was transduced to the neuroblastoma cell line. The ViraPower Packaging Mix contains an optimized mixture of the three

packaging plasmids, pLP1, pLP2, and pLP/VSVG. These plasmids supply the helper functions as well as structural and replication proteins in trans required to produce the lentivirus. An optimized 293FT producer cell line stably expresses the SV40 large T antigen under the control of the human CMV promoter and facilitates optimal production of the virus. Construction of replication-incompetent lentiviral vectors was performed using ViraPower expression system (Invitrogen), according to the recommendation of the manufacturer. Lentiviral constructs were used to infect SH-SY5Y cells. Expression of transfected pZNF526 DNAs was detected by antibodies against the V5-tag, expressed in these expression vectors or by GFP visualization, which represented the tag protein for the corresponding control empty vectors. Lentiviral production was previously described (see methods). A lentiviral expression system has been used to enforce expression of *ZNF526* wildtype and both mutants in neuroblastoma cell lines, SH-SY5Y, and then its effect on whole genome expression activation or repression was analysed.
By Western blot analysis, the expression of ectopic ZNF526 in SH-SY5Y-ZNF526-R459Q and SH-SY5Y-ZNF526-Q539H transfectants was detected (Figure 4.22), indicating the successful expression of ectopic ZNF526 wildtype or mutants in SH-SY5Y cells. Afterwards whole genome

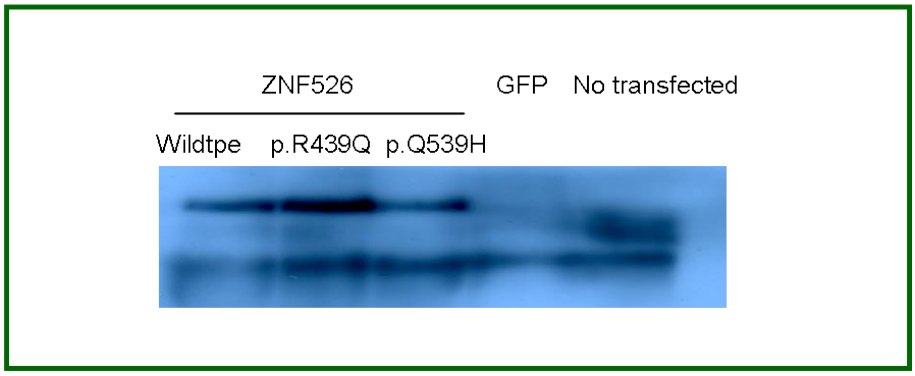


Figure 4.22: Stable overexpression of exogenous ZNF526 in human neuroblastoma cell line. SH-SY5Y cells were infected with lentiviruses carrying Wild type and both mutant p.R459Q, p.Q539H. Note that the endogenous levels of ZNF526 proteins are not visible in these Western blots due to the usage of V5 antibody to visualize the overexpressed exogenous forms of these proteins.

expression profiling (using Illumina Sentrix® Human-6 Expression BeadChip) was done in order to compare the expression levels of around 48000 known and unknown human genes in the mutant ZNF526 neuroblastoma cells and the wild type.
Functional annotation clustering using the DAVID database showed significant enrichment of common deregulated genes in expected pathways such as protein synthesis, mitochondrial dysfunction and gene expression (Figure 4.23).

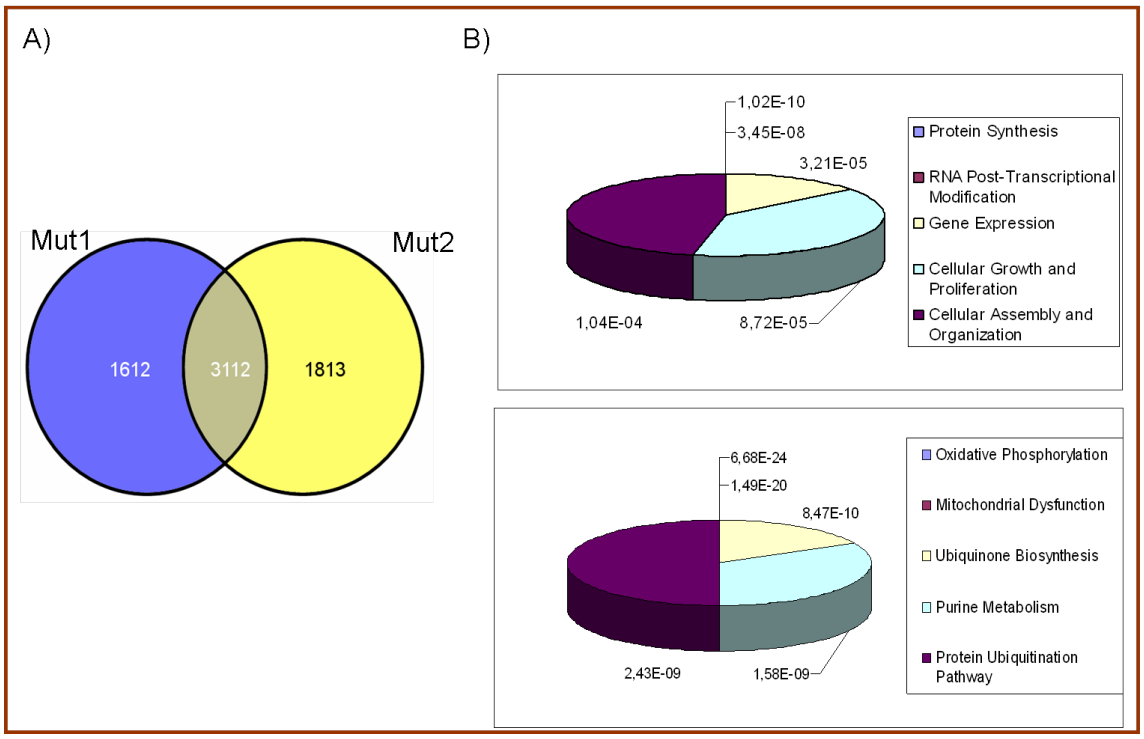
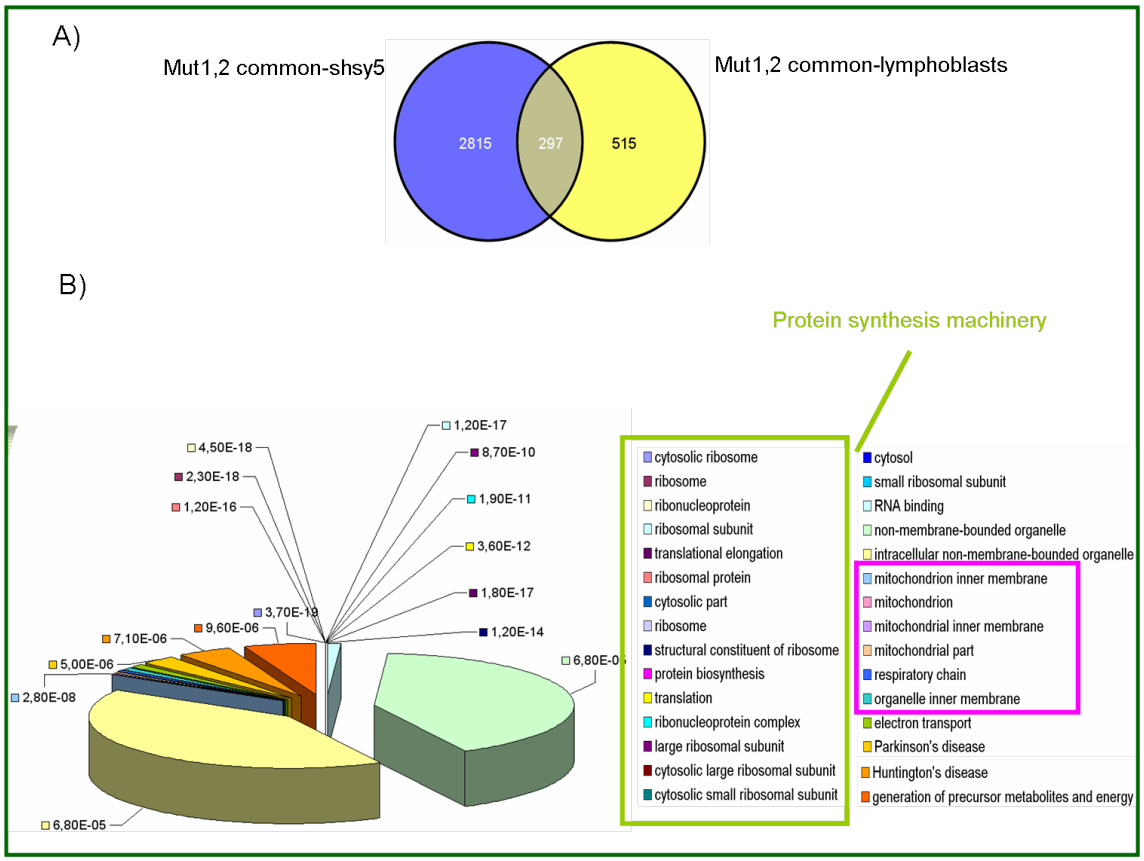


Figure 4.23: A) Venny diagram show differentially expressed genes from stable neuroblastoma cell lines expressing two different *ZNF526* mutations (Mut1: R459Q, Mut2: Q539H. B). Most of the deregulated genes are categorized to protein synthesis machinery and energy metabolism mechanisms

4.2.10 Gene ontology annotation revealed the unique nature of clustered genes

To determine the functional classification of the various gene clusters, we conducted Gene Ontology (GO) annotation using the DAVID functional classification tools, which constructs statistically significant functional profiles from a set of genes. The data file shows functional categories that are significantly enriched in the six clusters.
In order to find the common deregulated genes, the mutant neuroblastoma cell lines and patient lymphoblastoid cell lines were compared. Gene expression profiles were determined by comparing 3112 genes coming from the neuroblastoma cell lines with 811 genes resulted from the patient LCLs. The expression of 297 known genes was submitted for principal components analysis and hierarchical clustering.

The top categories of GO biologic processes in cluster A with an enrichment score of 11.55 were cytosolic ribosome, protein synthesis and translation. In cluster B the top ranked GO biological processes all relate to mitocondrion based on 30 genes (such as seven gene classes for NADH dehydrogenase), including genes involved in Alzheimer, Parkinson and Huntington diseases, and in oxidative phosphorylation.
Important genes within the expression profiles, and the related transcriptional regulatory mechanisms identified, are consistent with observations from the studies in lymphoblastoid specimens of patients and mutant *ZNF526* neuroblastoma cells tissue in independent studies, supporting the biologic and clinical relevance of genes within clusters identified in our study.



4.2.11 Molecular modelling

Zinc fingers of the Cys2His2 class constitute one of the most abundant and versatile DNA-binding motifs found in higher eukaryotes [Scot et al., 2001]. The probable primary role of C2H2 zinc finger proteins is to bind to a specific region of DNA and to participate in regulating gene expression. The fingers wrap around the major groove of the DNA and contact cognate bases through residues of the a helix at positions -1, 2, 3, 6, and 10 [Iuchi et al., 2001]. A three-dimensional modelling study of wildtype *ZNF526* and both mutatants, R459Q and Q539H , provided insights into the structure-function relationship (Figure 4.24). The effect of both mu-

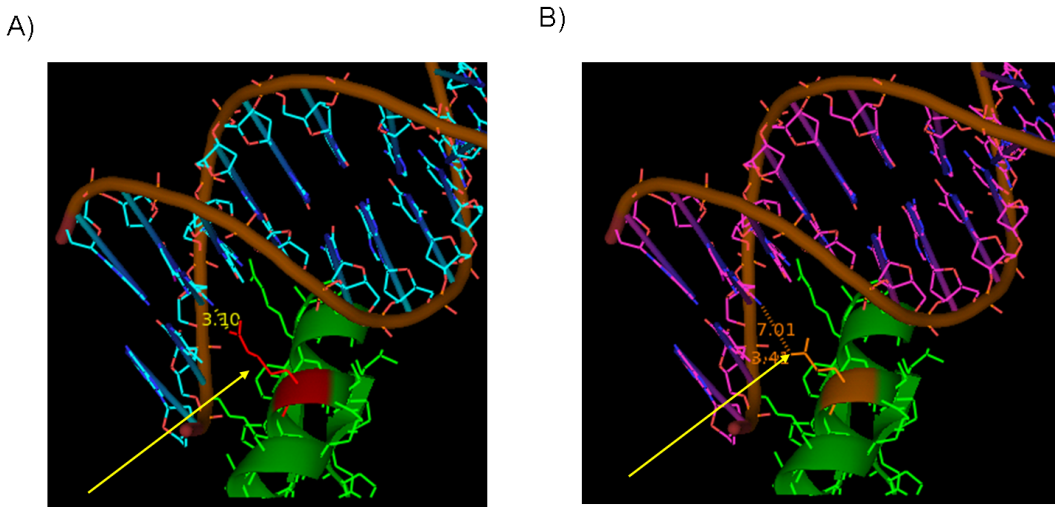


Figure 4.24: Protein modelling of wildtype ZNF526 (A) and mutant ZNF526-R459Q (B.)

tations on the DNA-binding capacity of ZNF526 was investigated using the crystal structure of a designed zinc-finger protein bound to DNA (PDB entry 1MEYC [Kim and Berg 1996; RCSB Protein Data Bank]). The sequence identity between the target and template for this region was 59%. Pymol [Brunger, 1998] has been employed to model the position of the intellectual disability associated ZNF526 mutations. Arginine at position six of the α helix was readily apparent in the 3.10 Å structure and makes a pair of hydrogen bonds to DNA. Our model predicts that the R459Q mutation increases the distance to the major groove of DNA to 7 Å. This reduces the DNA binding affinity and eventually disrupts the strong hydrophobic interaction with the primary strand. Therefore, the R459Q mutation will weaken the DNA binding of ZNF526 but nevertheless ZNF526 will still be able to recruit the transcriptional machinery to the promoter and initiate transcription.

In the case of Z13: Q539H, the mutation also decreases DNA binding, but not as significantly as Z10: R459Q. Additionally, the effect of the mutation, Q539H, appears to alter the locations of neighbouring side chains. In this case, the mutation may lead to loss of function due to both decrease in DNA affinity as well as changing in the overall protein folding (Figure 4.26).

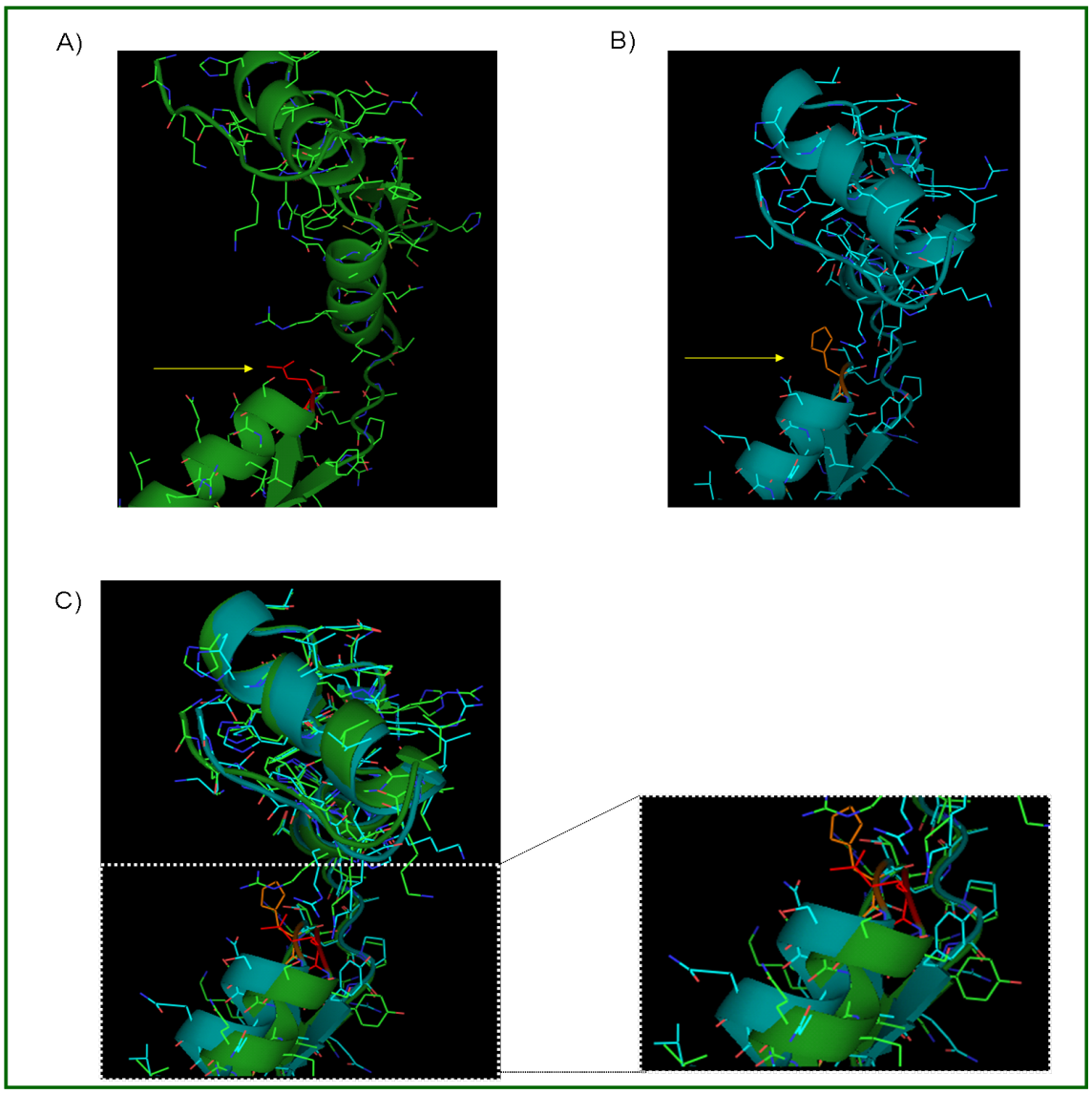


Figure 4.25: Protein modelling of the wildtype ZNF526 and mutant ZNF526-Q539H.

4.2.12 Genome wide identification of ZNF526 binding sites using CHIP-Seq

ZNF526 is predicted to bind to DNA. However, before our studies, binding of ZNF526 has not been investigated. Experimental whole genome-scale screening is needed to unveil the potential target genes for ZNF526 protein.
Through commercialization of the second generation of high throughput technologies (e.g., from Illumina-Solexa, Roche and ABI/SOLiD), ChIP-seq technology is rapidly emerging as a cost effective, global technology for high-resolution, genome-wide, unbiased mapping of protein-DNA interactions in chromatin-mapped complexes. It has been applied in global identification of in vivo binding of the insulator binding protein, CTCF [Barski et al., 2007], the neuron-restrictive silencer factor (NRSF) and STAT1 [Robertson et al., 2007; Johnson et al.,2007]. Ji et al. recently showed that ChIP-seq has better ability than ChIP-Chip in identifying TF binding regions for NRSF [24]. Therefore, we adopted the ChIP-Seq approach in our study. Identifying the ZNF526 targets would provide the information of regulatory roles of ZNF526 in intellectual disability and potential novel targets for brain development.To identify targets of ZNF526 in an unbiased genome- wide manner, the ChIP-seq experiments were performed on the four stable ZNF526-neuroblastoma stable cell lines, expressing wild type, R439Q and Q539H mutants as well as GFP as binding control. Cross-linking, chromatin isolation, sonication and immunoprecipitation using an anti-V5 antibody were performed as previously described [Barrera et al., 2008; Li et al., 2003]. Also other ChIP replicates were performed using IgG antibody for creating a negative control. The ChIP experiments were performed also on Green Fluorescent Protein (GFP) neuroblastoma stable cell lines as negative control protein for DNA binding. Two independent libraries of ChIP DNA were further sheared by sonication, end-repaired, ligated to sequencing adapters and amplified by PCR as previously described [Blow., 2008]. Gel-purified amplified ChIP DNA between 300 and 500 bp was sequenced on the Illumina Genome Analyzer II platform to generate 36-bp reads. Both Sequence libraries were aligned to the human reference genome (hg18) using Bowtie (bowtie-bio.sourceforge.net).
After appropriate quality filtering for each sample, between approximately 2.4 and 5.6 million aligned reads to single sites were used to identify regions of the genome. Sequence reads that map to multiple sites in the genome were removed from subsequent analysis. This eliminates sequences in simple repeats, some complex repeats, and 36-nt segments that are not unique by chance. The location of each remaining unique sequence read in the genome was recorded. The resulting sequence read distribution was processed with a Chip-seq peak locator. The algorithm finds a local concentration of sequence hits and, within that location, calls a peak. MACS, Model- based Analysis of Chip-Seq data [Zhang et al., 2008], has been used to analyze and compare the sequence reads of each wildtype and both of the mutants to their negative IgG controls with a high-confidence fold-enrichment (mfold) score of 8. For the ZNF526 protein 2476 peaks were found. FDR (false discovery rate) of 17% and a binding site length of 223bp represent that this protein is able to bind DNA and act as repressor or activator. 4440 peaks and 2965 peaks were found for R459Q and Q539R mutants, respectively. FDR (false discovery rate) for the first mutation is 98% and for the second mutant is 100%. This means that the binding of both mutant ZNF56 proteins is not different from their IgG controls. This indicates that the number of significant mutant peaks was not simply due to the specific binding of defected protein, and the ZNF526 mutant proteins do not bind specifically to the genome.

ZNF526 DNA	Total reads	High qualify reads	No. mapped reads	No. Peaks	MACS mfold=8
ZNF526 wt_V5	7.021.624	5.723.589	2.323.687	545	2476 d=223 FDR=17%
ZNF526 wt_IgG	10.584.136	9.087.459	1.411.456	165	
ZNF526 mut1_V5	18.252.040	15.455.154	4.892.653	1549	4440 d=184 FDR=98%
ZNF526 mut_IgG	17.520.817	14.438.325	5.381.932	993	
ZNF526 mut2_V5	12.966.672	11.285.157	5.943.671	1126	2965 d=199 FDR=100%
ZNF526 mut2_IgG	27.939.998	24.463.074	11.364.756	2411	
GFP_V5	30.369.413	25.830.015	12.294.780	1367	978 d=42 FDR=7%
GFP_IgG	17.745.844	12.673.592	892.997	274	

4.2.13 Categorization of the ZNF526 binding regions identified by ChIP-Seq based on their relative positions to genes

Previous studies of site-specific transcription factors have shown that some factors (e.g., E2F1) have a strong preference for binding near core promoters (18), whereas other factors (19, 20) show a more promiscuous binding pattern with respect to the start sites of transcription. To determine the location preference for ZNF526 binding, these binding regions have been mapped to the human genome. The mapping was done with the gene annotations from the UCSC genome annotation database for the human genome (hg18, Build 36.1). A histogram location of ZNF526 binding sites relative to transcription start site of genes is shown in figure 4.27. Most of the ZNF526 binding regions were identified around the TSS sites Figure 2.7).

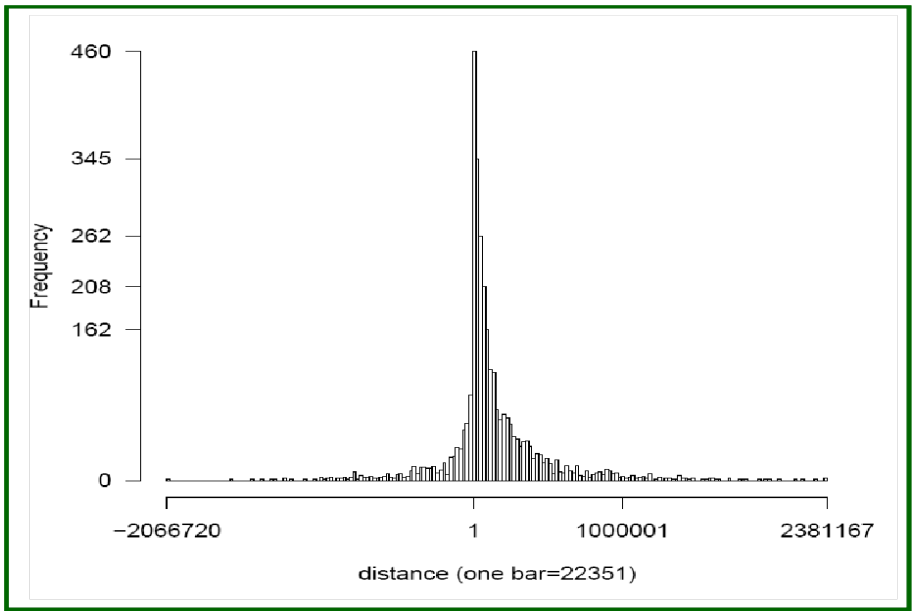


Figure 4.26: The ZNF526 binding sites are plotted with respect to the transcription start site.

4.2.14 Gene ontology and cascade analysis

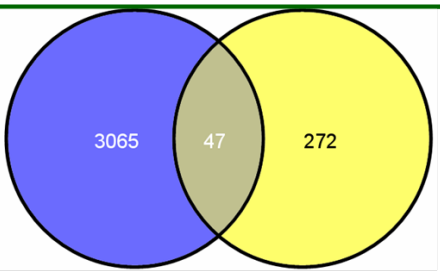
We carried out functional analysis on the 186 genes where the binding site of ZNF526 locates 2 kb upstream and downstream of their transcription start site. The significantly enriched categories for the target genes derived from the ZNF526 ChIP-Seq belong to ribosome biogenesis, rRNA processing and rRNA metabolic processes. Among the categories of ZNF526 candidate target genes, there was also enrichment for methyltransferase and methylation, as well as functional categories of DNA binding and transcriptional regulatory activity. Nuclear lumen, RNA binding, rRNA processing and methyltransferase activity are the top functional groups. Down syndrome susceptibility and neural tube defects are the top diseases associated with these genes. Thus, our findings suggest a dysregulation of methylation and protein synthesis as well as DNA binding control in the neuroblastoma ZNF526 mutant cells. The detailed GO results can be found in the figure 4.27.

Annotation Cluster 1	Enrichment Score: 1.7			Count	P_Value
GOTERM_CC_FAT	nuclear lumen	RT		21	2.3E-3
GOTERM_CC_FAT	nucleolus	RT		12	9.7E-3
GOTERM_CC_FAT	intracellular organelle lumen	RT		21	2.1E-2
GOTERM_CC_FAT	organelle lumen	RT		21	2.6E-2
GOTERM_CC_FAT	membrane-enclosed lumen	RT		21	3.2E-2
GOTERM_CC_FAT	nucleoplasm	RT		10	1.7E-1
Annotation Cluster 2	**Enrichment Score: 1.69**			**Count**	**P_Value**
GOTERM_BP_FAT	RNA processing	RT		11	7.6E-3
GOTERM_MF_FAT	RNA binding	RT		13	9.0E-3
GOTERM_BP_FAT	ribonucleoprotein complex biogenesis	RT		6	1.1E-2
GOTERM_BP_FAT	ncRNA processing	RT		6	1.3E-2
GOTERM_BP_FAT	ribosome biogenesis	RT		5	1.3E-2
GOTERM_BP_FAT	ncRNA metabolic process	RT		6	2.9E-2
GOTERM_BP_FAT	rRNA processing	RT		4	3.1E-2
GOTERM_BP_FAT	rRNA metabolic process	RT		4	3.5E-2
SP_PIR_KEYWORDS	ribosome biogenesis	RT		3	4.7E-2
SP_PIR_KEYWORDS	rrna processing	RT		3	7.1E-2
Annotation Cluster 3	**Enrichment Score: 1.6**			**Count**	**P_Value**
INTERPRO	C2 calcium-dependent membrane targeting	RT		5	1.7E-2
SMART	C2	RT		5	2.2E-2
INTERPRO	C2 membrane targeting protein	RT		4	4.1E-2
Annotation Cluster 4	**Enrichment Score: 1.44**			**Count**	**P_Value**
GOTERM_CC_FAT	nucleolus	RT		12	9.7E-3
GOTERM_CC_FAT	intracellular non-membrane-bounded organelle	RT		28	1.9E-2
GOTERM_CC_FAT	non-membrane-bounded organelle	RT		28	1.9E-2
GOTERM_CC_FAT	cytoskeleton	RT		11	5.1E-1
Annotation Cluster 5	**Enrichment Score: 1.09**			**Count**	**P_Value**
SP_PIR_KEYWORDS	methyltransferase	RT		5	2.1E-2
GOTERM_BP_FAT	biopolymer methylation	RT		3	9.4E-2
GOTERM_BP_FAT	methylation	RT		3	1.1E-1
GOTERM_BP_FAT	one-carbon metabolic process	RT		3	2.0E-1
Annotation Cluster 6	**Enrichment Score: 1.04**			**Count**	**P_Value**
SP_PIR_KEYWORDS	nucleus	RT		51	1.3E-4
SP_PIR_KEYWORDS	dna-binding	RT		23	1.4E-2
GOTERM_MF_FAT	transcription regulator activity	RT		17	1.2E-1
GOTERM_MF_FAT	DNA binding	RT		24	1.2E-1
GOTERM_MF_FAT	transcription factor activity	RT		12	1.3E-1
SP_PIR_KEYWORDS	transcription regulation	RT		19	2.0E-1
SP_PIR_KEYWORDS	Transcription	RT		18	3.2E-1
GOTERM_BP_FAT	regulation of transcription, DNA-dependent	RT		16	3.4E-1

Figure 4.27: Functional categories of the ZNF526 target genes in neuroblastoma stable cell lines.

4.2.15 Correlation of ZNF526 binding with gene expression data

To answer the question whether ligand-independent ZNF526 binding to the genomic regions exerts functional consequences through changes in the expression level of the targeted genes, we integrated the ChIP-Seq data with the gene expression profiling data. As the majority of peak calls are near the transcription start sites, ZNF526 effects on gene regulation are, on average, exerted over longer distances. 319 genes (10kb upstream and downstream of the transcription start site ZNF526 binding regions) identified by ChIP-Seq have corresponding genes on the array we used for expression profiling analysis. Comparing the expression profiling data with the ChIP-Seq data, we identified a total of 47 genes for which the genes were changed. To elucidate which genes are deregulated by ZNF526, Venn Diagram analysis was performed and subsequently GO analysis was conducted. GO analysis (Figure 4.28) revealed



Annotation Cluster 1	Enrichment Score: 1.21			Count	P_Value
GOTERM_BP_FAT	protein complex assembly	RT		6	1.4E-2
GOTERM_BP_FAT	protein complex biogenesis	RT		6	1.4E-2
GOTERM_BP_FAT	macromolecular complex assembly	RT		6	4.1E-2
GOTERM_BP_FAT	macromolecular complex subunit organization	RT		6	5.2E-2
GOTERM_BP_FAT	positive regulation of cell proliferation	RT		3	3.4E-1
GOTERM_BP_FAT	regulation of cell proliferation	RT		4	4.0E-1
Annotation Cluster 2	**Enrichment Score: 1.14**			**Count**	**P_Value**
GOTERM_MF_FAT	RNA binding	RT		6	2.6E-2
GOTERM_BP_FAT	RNA processing	RT		5	7.1E-2
GOTERM_BP_FAT	ncRNA processing	RT		3	1.0E-1
GOTERM_BP_FAT	ncRNA metabolic process	RT		3	1.4E-1
Annotation Cluster 3	**Enrichment Score: 0.95**			**Count**	**P_Value**
GOTERM_MF_FAT	RNA binding	RT		6	2.6E-2
SP_PIR_KEYWORDS	ribonucleoprotein	RT		3	1.4E-1
GOTERM_CC_FAT	ribonucleoprotein complex	RT		3	3.7E-1
Annotation Cluster 4	**Enrichment Score: 0.51**			**Count**	**P_Value**
GOTERM_BP_FAT	negative regulation of transcription	RT		4	1.5E-1
GOTERM_BP_FAT	negative regulation of gene expression	RT		4	1.8E-1
GOTERM_BP_FAT	negative regulation of nucleobase, nucleoside, nucleotide and nucleic acid metabolic process	RT		4	1.8E-1

Figure 4.28: Gene ontology of the neuroblastoma deregulated ZNF526 target genes.

that ZNF526 associated genes are mostly involved in protein synthesis machinery, which is consistent with our previous results. These results include the ZNF526 categories for protein biogenesis, RNA binding and ribonucleoprotein structural constituents of ribosome and transcription and DNA binding maintenance. ZNF526 has significant enrichment of genes that are involved in basic ribosome biogenesis, particularly those related to rRNA processing. Most of the overlapping genes (between the neuronal expression microarray and ChIP-Seq targets)

were down-regulated by ZNF526, showing that ZNF526 works as an inhibitor in most situations. A complete listing of all candidate target genes and significantly enriched functional categories for all transcription factors is provided in the appendix.

4.2.16 ZNF526 protein interactors

An interaction protein of ZNF526 has been identified as part of systematic collaboration with Prof. Matthias Mann (and Dr. A. Hyman and N. Hubner, Max planck institute of Biochemistry, Martinsfried, Germany) who established a high-throughput method for exploration of protein function in mammals. In this approach, a gene of interest in its genomic context is tagged with a construct containing, e.g., GFP [Kittler et al., 2005]. The BAC transgene can then be stably transfected into mammalian cell lines of choice (Hela cells). This allows for expression of the tagged protein at endogenous levels and ensures cell type specific processing and regulation. Interacting proteins were identified by using the GFP tag to pull down protein complexes from transfected cells. Subsequently, these dissociated, fragmented into peptides and then mass spectrometry was performed to recognize the relevant interactors. BAC-based transgenesis in Hela cell cultures, together with quantitative interaction proteomics, revealed PRKIR as an interaction partner for ZNF526. The experiment was done in triplicate and only one highly significant interactor was detected. PRKRIR (52 kDa repressor of the inhibitor of the protein kinase) indicates similar average intensity to ZNF526, 7,33639 and 7,37432 respectively.

4.2.17 Derivation of a binding motif for ZNF526

The results of the genome location experiment of ZNF526 were further analyzed using a motif-finding algorithm (MEME), which examines the ChIP-Seq selected sequence and searches for DNA sequence motifs representing the protein-DNA interaction sites. Given the output from MACS, we reduced the size of the peak regions to 60 bp by only taking the +30 bp to -30 bp regions around the peak summit. We then downloaded the repeat-masked sequences corresponding to these peak regions from the UCSC Genome Browser. After having filtered out redundant sequences, we ran MEME on the resulting 2418 sequences of length 60 bp. Two motifs from MEME search were derived from the most frequent motifs found in ZNF526-bound promoters (Figure 4.30).
Motif 1: Results of Fisher's exact test, which are highly significant, indicate the Meme motif 1 is significantly enriched in the 2418 peak sequences compared to the random sequences. This motif appears in approximately 99% of the promoters belonging to the overlap of differentially expressed genes in mut1 and mut2 in neuroblastoma stable cell line differentiated genes and in 100% of the promoters belonging to the overlap of differentially expressed genes in mut1 and mut2 in patients lymphoblastoid cells. Motif 2: The 1136 genes with at least 1 peak containing the Meme motif are given in the appendix. The Meme motif 2 is significantly enriched in the 2418 peak sequences compared to the random sequences. This motif appears in approximately 98-100% of the promoters of differentially expressed genes.

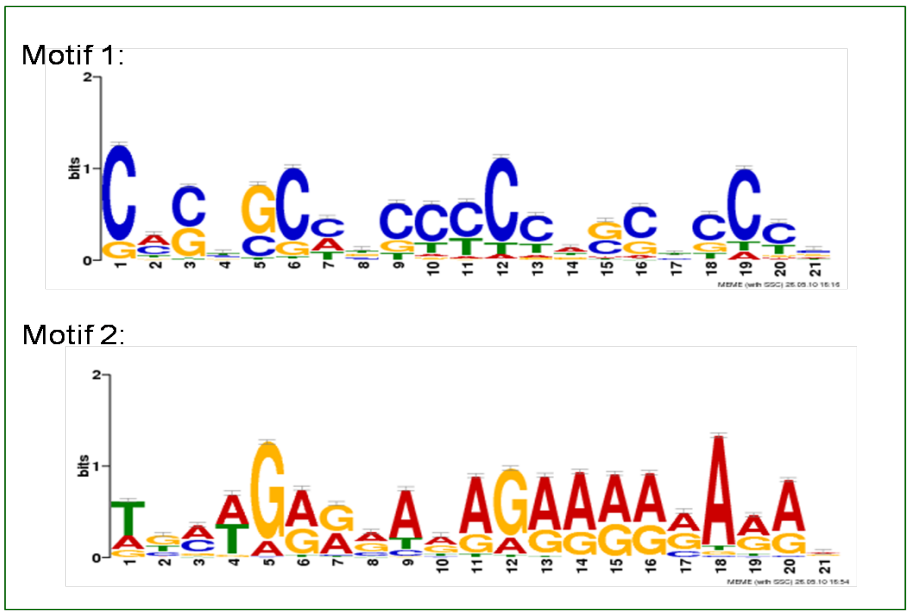


Figure 4.29: Two significant motifs have been indicated in the peak regions resulted from ChIP-Seq

4.3 Transcription factors

Transcription factors are sequence-specific DNA binding proteins, which mediate changes in gene expression patterns in response to a specific challenge. They can act both as activators or repressors of gene transcription and exert this function by binding to discrete DNA motifs within and flanking a gene in a region of active chromatin, thereby determining the activity of the gene's promotor. Some of the DNA motifs are conserved and share common consensus DNA sequences which are recognized by transcription factor families, which themselves are often expressed in a tissue-specific, developmental or stimulus-specific manner. Members of one protein family share a conserved DNA-binding region, which allows binding to a specific consensus DNA motif. Transcription factors can enhance their transactivation potency by operating in a combinatorial fashion, forming hetero- or homodimers or interacting with other (multimeric) transcription factor complexes, which are bound to DNA nearby. Transcription factors are composed of different domains that serve distinct functions.

4.3.1 Transcription factor classes

Depending on whether transcription factors are induced by cellular stimulation or are present in quiescent cells in the absence of any external stimulus, one can discriminate between inducible transcription factors, for instance the Jun, Fos and Krox proteins which are encoded by immediate-early genes, and constitutive transcription factors, e.g., the basic leucine zipper proteins CREB or ATF-1, which are already bound to DNA and are activated by phosphorylation [Herdegen and Leah, 1998]. Another group of transcription factors is classified into families of related proteins which share peptide domains involved in either specific DNA sequence recognition or dimerization with related proteins [He and Rosenfeld, 1991; Latchman, 1993; Struhl, 1991]. Being characterized by a common DNA binding domain, members of a specific family of transcription factors hence recognize similar DNA motifs in promoter and enhancer regions

of downstream target genes. In the following section, six examples for such transcription factor families are described in more detail .

POU-domain proteins:

The name POU was derived from four transcription factors which were first recognized to make up this protein family. These are the mammalian factors Pit1 and the octamer binding proteins Oct1 and Oct2, as well as the protein Unc86 from *C.elegans* [Herr et al., 1988]. POU domain transcription factors are a superfamily of homeodomain proteins that regulate cell differentiation and proliferation. Based entirely on structural homology, the POU-domain factors have been divided into six distinct subclasses. In vertebrates, proteins of the POU-III, POU-IV, and POU-VI classes are expressed predominantly in the nervous system [Jerry et al.,1998]. A prominent feature of POU proteins is a 147-156 amino acid POU domain including a 75-82-residue POU-specific segment and a 60-62 residue POU homeodomain [Jean, 1996]. Both domains are linked by a variable and poorly conserved spacer region. POU family members recognize a defined DNA sequence motif, known as the octamer element [He et al., 1989; Herr et al., 1988; Rosenfeld, 1991; Verrijzer and Van der Vliet, 1993; Wegner et al., 1993].

Basic helix-loop-helix (bHLH) proteins

The basic helix-loop-helix (bHLH) domain is a highly conserved amino acid motif that defines a group of transcription factors.The helices mediate dimerization with homologous or heterologous bHLH proteins.The bHLH domain consists of 50-60 amino acids that form two distinct segments: a stretch of 10-15 predominantly basic amino acids (the basic region) and a section of roughly 40 amino acids predicted to form two amphipathic α-helices separated by a loop of variable length (the helix-loop-helix region) [Pires., 2010]. The adjacent basic region fits in the major groove of the DNA and is required for DNA binding to the E-box hexamer CANNTG [Murre et al., 1989a; Murre et al., 1989b]. The nucleotides of this DNA recognition motif vary in different subfamilies of bHLH transcription factors, which are grouped on the basis of closer sequence similarities in the bHLH domain. The HLH domain promotes dimerization, allowing the formation of homodimeric or heterodimeric complexes between different family members [Murre et al. 1989a; Kadesh 1993]. The two basic domains brought together through dimerization bind specific hexanucleotide sequences. The bHLH proteins involved in neurogenesis include *Drosophila* Atonal and other 'proneural' proteins. In vertebrates, Mash-1, Math-1 and the neurogenins are important in the initial determination of neurons, whereas Nero-D, NeuroD2, MATH-2 and others are differentiation factors. The bHLH transcription factors dHAND and eHAND are important in cardiac development in vertebrates [Jones et al., 2004].

Homeobox transcription factors

Homeobox genes comprise a large and diverse group of genes, most of which are thought to act as transcription factors. They are characterised by possession of one or more homeobox sequences of 180 base pairs (or longer) encoding homeodomain peptides that fold into helix-loop-helix-turn-helix domains [Zhong et al., 2011]. Genes that code for members of the homeodomain family of transcription factors contain a common sequence of about 180 bp (the

homeobox) that encodes a DNA-binding protein sequence known as the homeodomain. The structure of the 61-amino acid homeodomain is constituted by three α-helices (helix-turn-helix) and is highly conserved across species [Kappen and Ruddle, 1993]. The best known homeobox genes are Hox genes, usually arranged into gene clusters and known to play pivotal roles in specification of cell identity along the developing anteroposterior body axis in the embryos of bilaterian animals [Zhong et al., 2011].

ETS-domain transcription factors

This family of transcription factors seems unique to the metazoan lineage. So, members of this family probably function in new regulatory cascades that are specific to metazoa [Sharrocks, 2001]. EST-domain transcription factors share a highly homologous 80-90-amino acid DNA binding domain, the ETS domain, which was first discovered in oncogenes cloned from retroviruses [Karim et al., 1990]. The conserved region, termed the ETS domain, corresponds to the DNA-binding domain of these proteins [Shore, 1998]. Minimal polypeptides (86 amino acids) that encompass the ETS domain bind DNA with high degrees of both affinity and specificity Outside the conserved DNA-binding domain, ETS factors reveal a pronounced divergence and can be grouped into subclasses based on additional homologous regions unique to particular members of the ETS family [Janknecht and Nordheim, 1993; Wasylyk et al., 1993].

Forkhead proteins

Forkhead/winged helix proteins are characterized by a conserved 100 amino acid domain called the forkhead box.The DNA-binding domain folds into a variant of the helix-turn-helix motif and is made up of three helices and two characteristic large loops [Klaus, 2000]. In 1998, a standardized nomenclature for all chordate winged helix/forkhead transcription factors was introduced and Fox (Forkhead box) was adopted as the unified symbol for these proteins [Kaestner, 2000]. DNA binding of Forkhead proteins relies on interactions between the third helix in the Forkhead box, the so-called "recognition helix" H3, and DNA bases within the major groove of double-stranded DNA.The optimal DNA-binding site for the FOXO members, for example, has been determined as TTGTTTAC [Boudewijn, 2003].The forkhead family in humans consists of 39 distinct members, which have been divided into 19 subgroups

Zinc finger proteins

The class of genes encoding multiple zinc finger (ZF) domains is very large in mammals and appears to have expanded on the primate lineage. Zinc fingers are small protein domains in which zinc plays a structural role contributing to the stability of the domain. Zinc fingers are structurally diverse and are present among proteins that perform a broad range of functions in various cellular processes, such as replication and repair, transcription and translation, metabolism and signaling, cell proliferation and apoptosis. Zinc fingers typically function as interaction modules and bind to a wide variety of compounds, such as nucleic acids, proteins and small molecules [Krishna, 2003]. Zinc finger proteins consist of a motif in which either four cysteine residues or two cysteine and two histidine ligands coordinate a zinc ion (C2-C2 type or C2-H2 type). Differential use of the two residues gives rise to several types of zinc

finger, such as C2H2, C2HC, C2C2, C2HC C2C2, and C2C2 C2C2 [Iuchi et al., 2001]. Among them, C2H2 is known as the classical zinc finger and is often described as CX2-4CX12HX2-6H, to show the intervals between the zinc-binding residues. C2H2 zinc finger proteins probably comprise the largest family of regulatory proteins in mammals. Most zinc fingers bind to a cognate DNA. In most proteins, this domain is present in multiple copies as regular tandem arrays with invariant cysteine and histidine residues in each repeat. Every single motif hence forms an independently folded structural domain organized around a zinc ion, which is essential for correct folding. The polypeptide fold consists of α helix and two β-strands that are arranged in a hairpin structure [Lee et al., 1989]. In addition to DNA, many of the proteins also bind to RNA or protein, and some bind to RNA only. The vast majority of zinc finger proteins can be classified as Krüppel-like, which are a further subgroup of the C2H2-zinc finger proteins and which exhibit homology to the *Drosophila melanogaster* segmentation gene product, Krüppel. Members of this subgroup are termed Krüppel-like factors (KLFs), and many KLFs exhibit tissue-selective expression and wide-ranging regulatory functions. The name 'zinc finger' reflects two facts, the first being that the domain requires one or more zinc molecules to stabilize its structure, and the second that the two-dimensional structure of the first identified zinc finger motif (C2H2) resembles a finger, as shown in Figure 4.31 [Albert et al, 1994]. Zinc

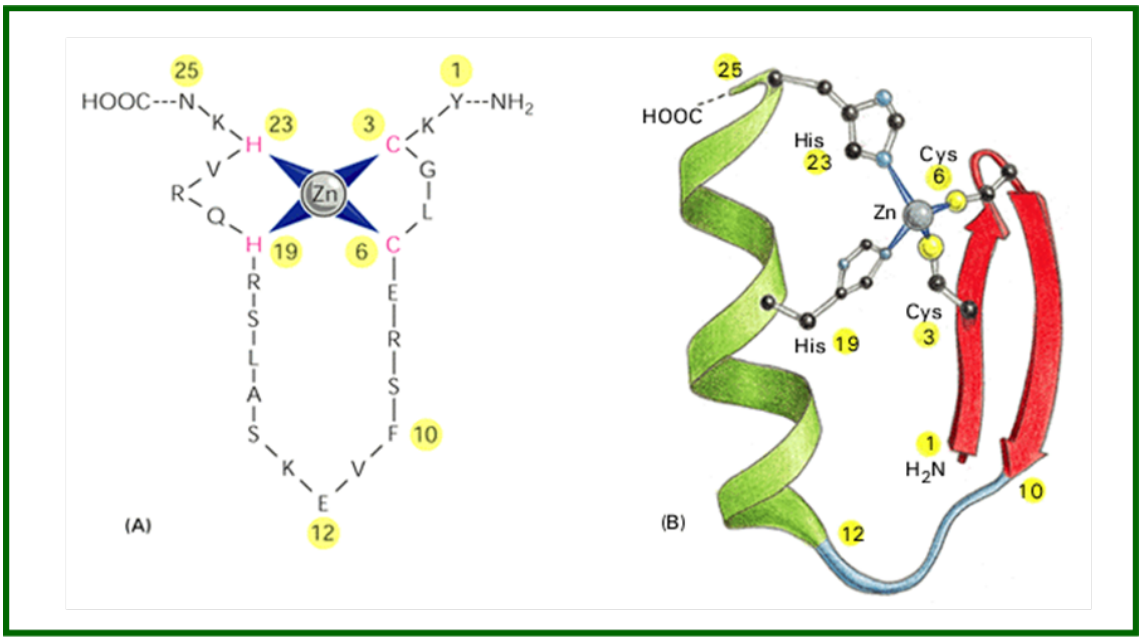


Figure 4.30: (A) Schematic picture of a Zinc Finger (B) The Three Dimensional Structure of a Zinc Finger

finger proteins are believed to be one of the most common classes of proteins in humans (3-4% of human genes encode proteins containing zinc finger domains).

The C2H2 (Kruppel-type) zinc finger motif

The C2H2 (Krüppel-like) zinc finger protein family is one of the largest families of mammalian transcription factors. Its members are characterized by the presence of individual or clustered C2H2 zinc finger motifs (CX2-4CX3FX5LX2HX3-4H), which are linked by a highly conserved H/C linker (TGE(K/R)P(Y/F)X) [Hongling et al., 2007]. This motif was first identified in the

transcription factor, TFIIIA, of *Xenopus laevis* [Looman et al, 2002]. The motif consists of two short beta strands followed by an α helix. The four conserved residues serve to hold one end of the antiparallel $\beta - sheet$ to one end of the $\alpha - helices$. The latter is known as the 'recognition helix' as it mediates DNA binding through non-covalent interactions between three of its amino acid residues and three adjacent bases within the DNA major groove. Secondary interactions involve the phosphate-backbone [Klug et al., 1999]. C2H2 zinc finger proteins contain from 1 to more than 30 fingers. Based on classification scheme by S. Iuchi (2001), most of the proteins can be classified into one of three groups: triple-C2H2, multiple-adjacent-C2H2, and separated-paired-C2H2 finger proteins.

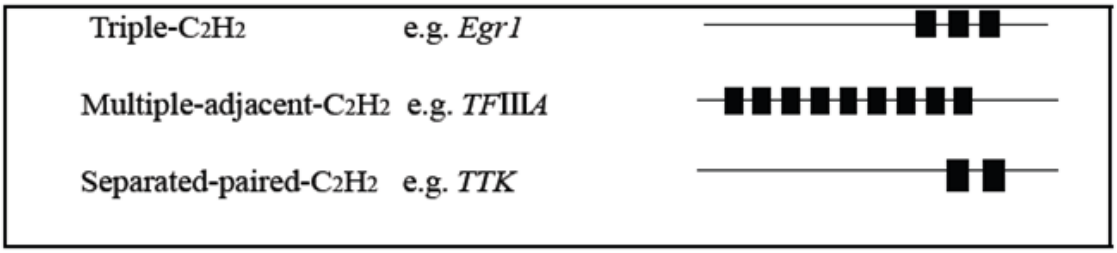


Figure 4.31: (The Three Proposed Classes of C2H2 Zinc Finger Proteins [Reproduced from Iuchi, 2001]

The C2H2 zinc finger gene family is very diverse. In addition to the zinc finger motif, most of these genes also encode conserved regulatory regions for subcellular localization, DNA binding, or gene expression. These include the kruppel-associated box (KRAB), the SCAN domain, and the BTB/POZ domain [Sander et al, 2003].

KRAB Domain

KRAB/C2H2 zinc finger proteins make up approximately one third of the various zinc finger proteins found in the human genome. Many usually function as transcriptional repressors, and may play important roles in diverse developmental and pathological processes [Hongling et al., 2007]. The KRAB domain comprises about 50-75 amino acids, located near the amino terminal of most eukaryotic kruppel-type zinc fingers. It consists of two subregions, A and B, which are predicted to fold into two amphipathic $\alpha - helices$ enriched in charged amino acids. The domain is only found in tetrapod vertebrate genomes [Collins et al, 2001].
The KRAB domain is tetrapod-specific and has a strong transcriptional repressor function when tethered to the target DNA by a DNA-binding domain, such as the C2H2 motif. For instance, the human zinc finger protein 74, ZNF74, consists of a truncated KRAB A domain and 12 different C2H2 zinc finger motifs. The role of the C2H2 motifs is to target the protein to the nuclear matrix, as well as RNA binding. ZNF74 also interacts with RNA polymerase, suggesting its function as both a transcriptional and post-transcriptional repressor [Urrutia, 2003]. The function of KRAB zinc finger proteins remains unknown. However, the fact that KRAB containing proteins are vertebrate-specific suggests that they have evolved recently, and that their key roles lie in some aspects of vertebrate development, such as the immune and nervous system,including haematopoietic cell development and differentiation, and myeloid and

lymphoid differentiation (Collins et al, 2001).

The SCAN Domain

The SCAN domain is a conserved region of 84 residues found predominantly in zinc finger DNA-binding proteins in vertebrates. This leucine-rich region is predicted to contain a high degree of $\alpha - helices$. It is located at the N terminus when it is part of a zinc finger-containing transcription factor [James R. 2002]. The SCAN domain is always present in a single copy within all of the SCAN-containing proteins [Williams et al, 1999]. Like the KRAB domain, the SCAN1 domain is a highly conserved vertebrate-specific protein domain. The name for the SCAN domain was derived from the first letters of the names of four proteins initially found to contain this domain (SRE-ZBP,CTfin51, AW-1 (ZNF174), and Number 18 cDNA or ZNF197). Concerning the biological function of the SCAN-containing proteins, some members of the family could play a role in the transcriptional regulation of growth factors, genes involved in lipid metabolism and cell differentiation genes. For instance, ZNF174 acts as repressor for PDGF (Platelet-derived Growth Factor-B chain) and TGF (Transforming Growth Factor) promoter reporter expression. Most of the SCAN zinc finger genes are found in clusters on human chromosomes, including 3p21 (e.g., ZNF20), 6p21.3 (e.g., ZNF165, ZNF192, ZNF193) and 16p13.3 (e.g., ZNF174, ZNF213). These locations are of particular interest since they are frequently disrupted in a variety of cytogenetic abnormalities [Sander et al, 2003].

The BTB/POZ Domain

In the C2H2 class of zinc fingers, the third associated modules is the pox virus and zinc finger (POZ) domain (5), which is also known as the BTB domain (Broad-Complex, Tramtrack, and Bric-a-brac)[Collins, 2001].One of the most important members of this protein subfamily that has been widely used in structural and functional studies on the BTB domain is the promyelocytic leukaemia zinc finger (PLZF) protein. The fact that the loss or alteration of the function of many zinc finger proteins can lead to serious human diseases shows they have important and versatile functions.In general, they possess a DNA-binding domain for binding to specific promotor or enhancer DNA regions of a target gene, a transactivation domain for interaction with the basal transactivation complex and sometimes an additional dimerization domain in order to form homo-or heterodimers with other transcription factors. Transcription factors often act sequentially in that they regulate the expression of other transcription factors, which in turn regulate the expression of downstream effector genes.

5 NOL1/NOP2/Sun domain family member 2 protein

The positions of six hotspot loci for the autosomal recessive intellectual disability are shown in Figure 5.1. One of these hotspot loci is MRT5, which is located on 5p15.32- p15.2 and spans 5.6

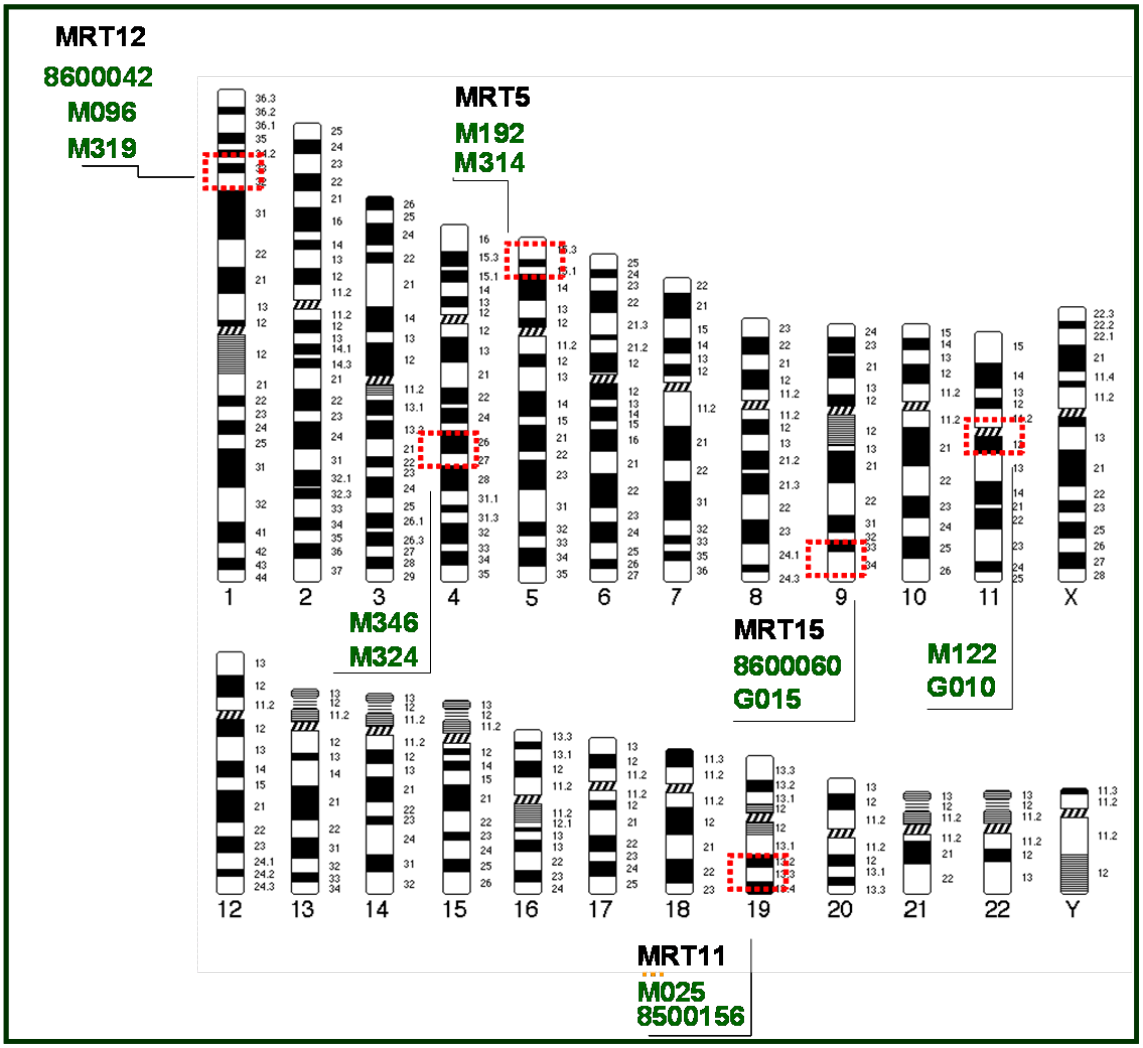


Figure 5.1: Whole genome view of the hotspot regions for autosomal recessive intellectual disability. Six regions have been found so far. Only standalone families are shown for each region.

Mbp on the short arm of the chromosome. Four families have overlapping linkage intervals in this locus, two of which, M192 and M314, were solitary and the other two have significant LOD scores for this region (figure 5.2). Family M141 has three additional homozygote regions on chromosomes 6, 19 and 22. (Table 5.1). Family 8500157 also has an additional interval on chromosome 7. This indicates that in these two families there is the possiblity of gene defects

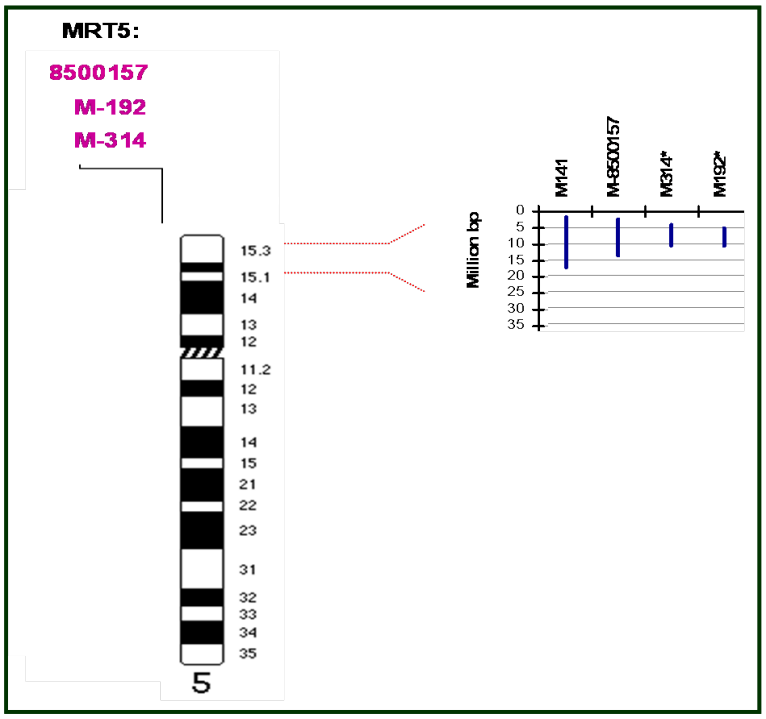


Figure 5.2: chromatogram of chromosome 5 and the MRT5 locus. The two families, M314 and M192, with solitary intervals are indicated by asterisks. The smallest interval size belongs to the family M192, 5641748 bp and the largest size belongs to family M141, 15767678 bp.

existing in the regions other than on chromosome 5.

Family	Chromosome	Start	End	Length of interval[bp]
M192	Chr05	5144766	10786514	5641748
M314	Chr05	4007570	107786776	6779206
8500157	Chr05	2203273	13822974	11619701
	Chr07	78542314	80517630	1975316
M141	Chr05	1677089	17444767	15767678
	Chr06	93549023	97375947	3826924
	Chr19	8829549	56763284	47933735
	Chr22	25544435	46443598	20899163

Table 5.1: Homozygous regions identified in the families with overlapping interval on the locus MRT5

5.1 The investigation of family M192

5.1.1 Clinical description

The pedigree and facial aspects of the patients in family M192 are shown in Figure 5.3. The degree of ID in the affected family members is moderate (Table 5.2). The patients show spas-

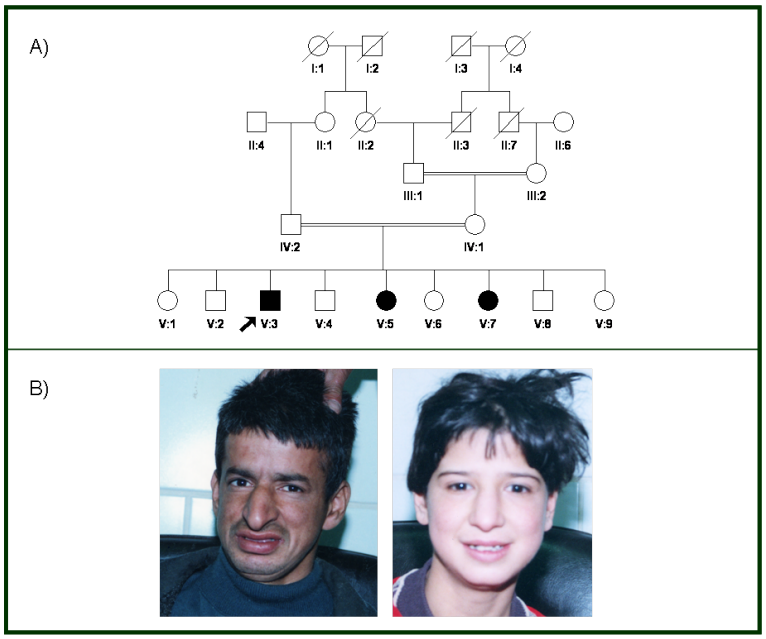


Figure 5.3: Family M192. A) Pedigree, B) Clinical photographs of the two patients.

ticity and facial dysmorphisms like long nose and pointed chin. Patient V:3 has borderline microcephaly (Head circumference; 53cm)(Table 5.2).

Patient	Sex	Age at examination	Mental Retardation / IQ	Height	OFC[a]	Additional features
V:3	m	29 y	moderate	166 cm (10th cent.)	53 cm (3rd cent.)	IQ:40, spasticity,long nose, increased DTR
V:5	f	28 y	moderate	147 cm (th cent.)	55 cm (50th cent.)	IQ:40, spasticity,long nose,pointed chin,increased DTR
V:7	f	17 y	moderate	148 cm (<3rd cent.)	56 cm (50-75th cent.)	IQ: 50,spasticity,long nose,pointed chin,increased DTR

a: OFC, occipitofrontal circumference

Table 5.2: The clinical characteristics of the patients of family M192.

5.1.2 Molecular analysis

Individuals 3641, 3642, 3643, 3644, 3645, 3646 and 3647 were genotyped using the Human Mapping 10K (SNP) Array (Affymetrix). Parametric linkage analysis based on 10000 markers with high quality scores was carried out (Dr. Motazacker), which revealed any pathogenic chromosomal rearrangement. The interval is located on Chr.5 p15-31 between SNP markers rs1824938 and rs60701 and spans 5.6 Mbp with a maximum LOD score 3 (Figure 5.4). This region spans 5.6 Mb and contains 25 genes (Figure 5.5). Prior to mutation screening of coding exons and exon-intron boundaries, genes in the MRT5 interval were ranked according to their expression patterns and functional relevance in the central nervous system and also using the software available for gene prioritization such as PosMed. The mutation screening for all of the known genes as well as all of the mRNAs and ESTs in the interval was performed by specific

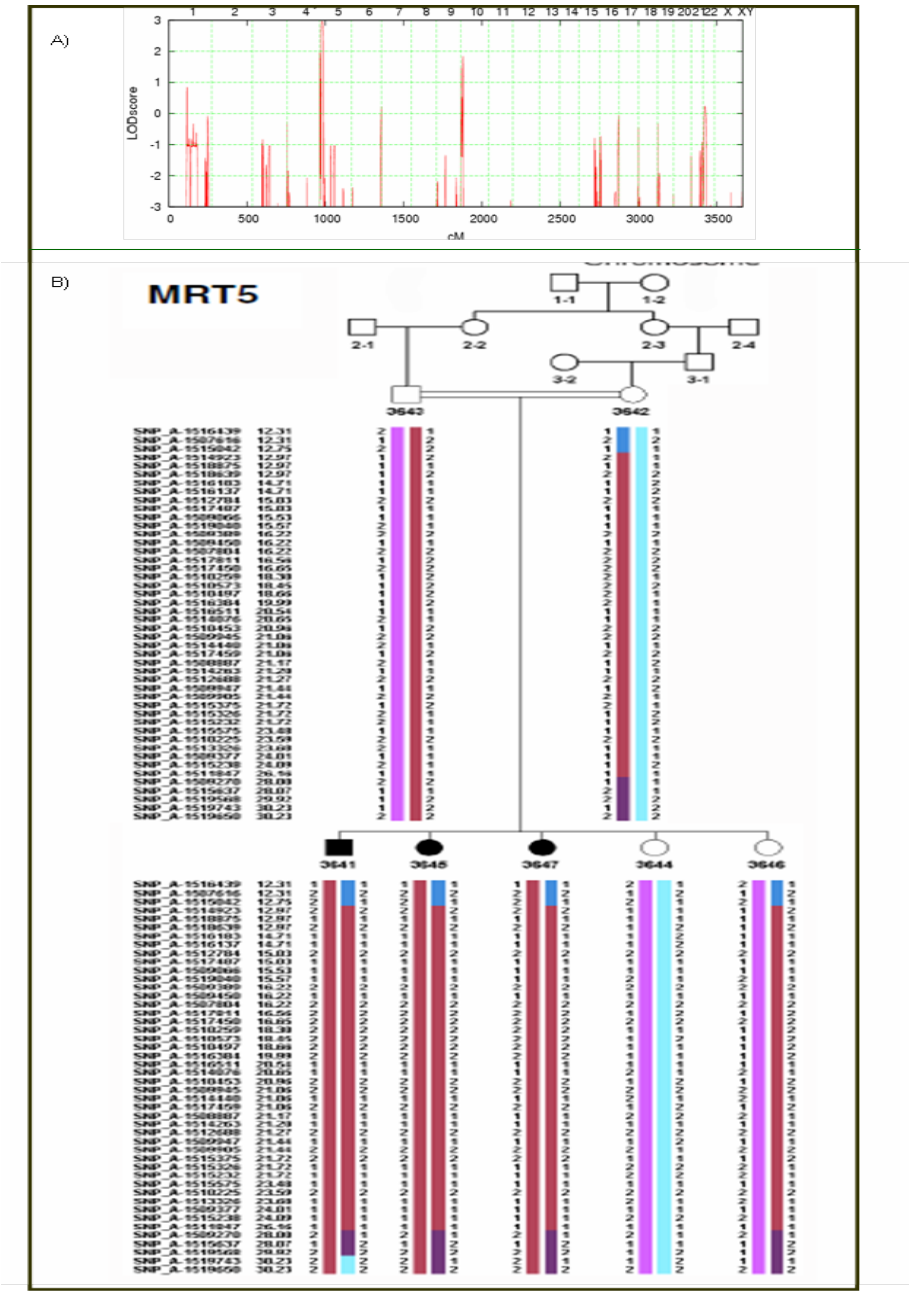


Figure 5.4: Family M192. A) Pedigree, B) The haplotype of the affected and parents

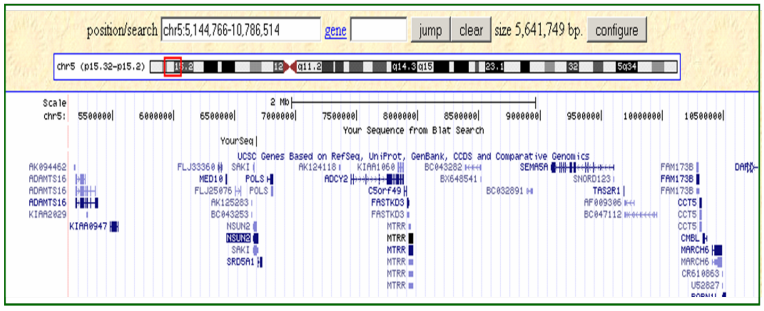


Figure 5.5: Gene content of the 5.4 Mb region which segregates in the linkage analysis. The region presents 25 genes (UCSC Genome Browser http://genome.ucsc.edu)

primers. The list of the primers and their sequence and location is available in the appendix.

5.1.3 Identification of the mutation

Mutation screening of the genes in the interval led to the identification of a Q227X mutation, a non-synonymous change (G679A: GTT>ATT), in a highly conserved region of exon 7 of *NSUN2* gene, which co-segregated with MR in family M192 (Figure 5.6). In order to rule out the pos-

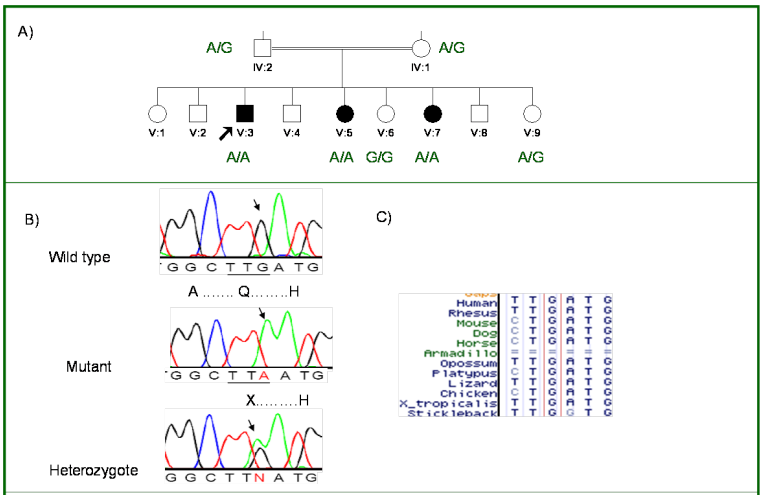


Figure 5.6: Gene content of the 5.4 Mb region which segregates in the linkage analysis. the region are present 25 genes (UCSC Genome Browserhttp://genome.ucsc.edu

sibility that the change found in *NSUN2* might be a common polymorphism, it could be excluded in the 384 Iranian control chromosomes as well as in 540 German control chromosomes (healthy blood donors). NSUN2, NOL1/NOP2/Sun domain family member 2, encodes a nucleolar RNA methyltransferase. *NSUN2* is expressed in the central nervous system and brain as well as in most other tissues. The methyltransferase domain of NSUN2 contains residues 100 - 427 and the position of the nonsense mutation found in the ID patient is located in the residue 227, which indicate that NSUN2 protein in the patient will be a truncated protein with loss of the methyltransferase domain.

5.1.4 *NSUN2* expression study

To check for *NSUN2* expression, total RNA from EBV-transformed lymphoblastoid cell lines (LCLs) of a patient and two controls were extracted and cDNA was generated. This cDNA was used to perform PCR with specific primers for exon 2 - exon 3 and exon 6 - exon 7 junction. (Appendix). All PCR products were found to be present in the controls but not in the patient, proving the complete absence of a NSUN2 transcript in homozygous carriers for stop codon mutation (Figure 5.7).

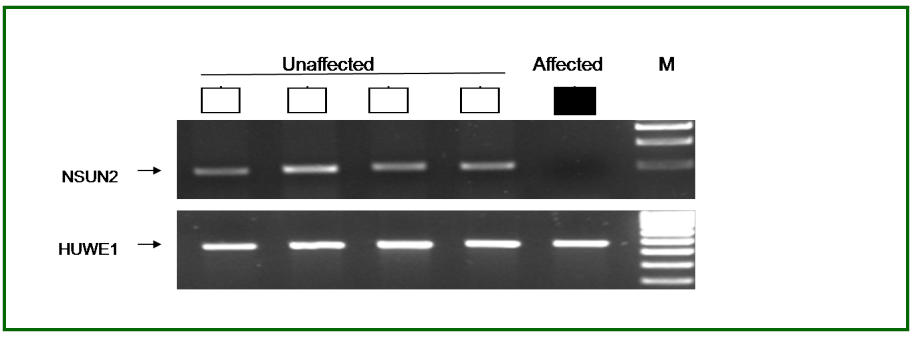


Figure 5.7: RT-PCR results with cDNA derived from RNA preparations of 1 patient (V:3) and four control lymphoblastoid cell lines. Using a sequence of primer pairs specific for amplicons covering four consecutive exons , the NSUN2 transcript was detected in the control but could not be amplified in patient cDNA. The results of an 1% agarose gel eletrophoresis of 5 μl from a 25 μl RT-PCR reaction are shown.

5.1.5 Real Time quantitative validation experiment

This result was substantiated by quantitative PCR, using lymphoblastoid cell line-derived cDNA from four healthy individuals as well as four patients (Figure 5.8). Experiments were performed using junction primers exon 3 and 4 (forward primer: GTTGGTATCCTGAAGAACTTGC) and reverse primer: CCAGGTGCTGCACACATATC). For normalization, exon 4 to 5 of the GAPDH gene (forward primer: AAGTGTGACGTGGACATCCG and reverse primer: GATCCACATCTGCTGGAAGG) was amplified in the same experiment.

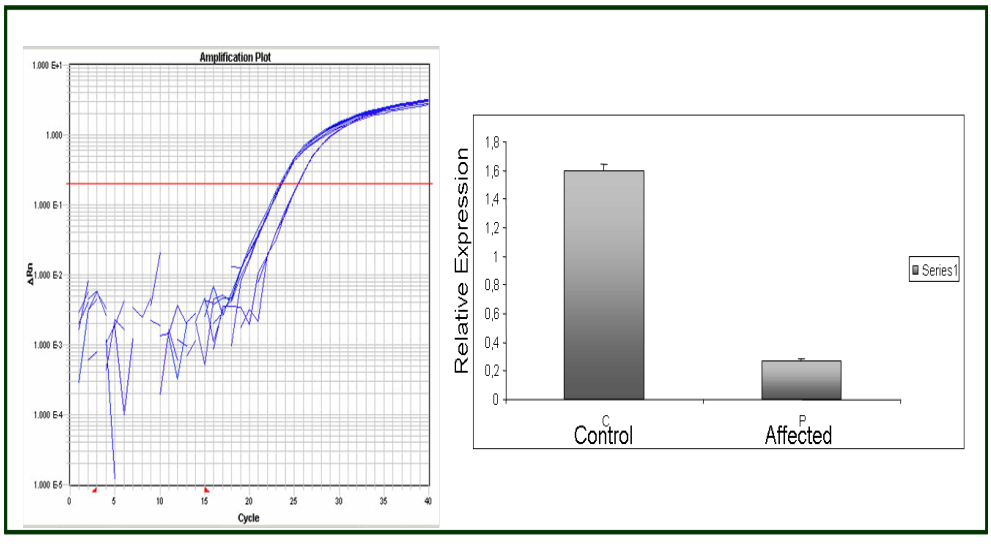


Figure 5.8: Quantitative PCR for *NSUN2* using lymphoblastoid-derived cDNA from patients V:3 as well as from healthy individuals compared to *GAPDH* expression levels. Error bars represent the standard deviation [n=3).

5.2 The investigation of family M314

5.2.1 Clinical description

The second family with overlapped interval on MRT5 locus is family M314. This family is a large consanguinity family containing two branches. Most patients have an IQ of less than 50.

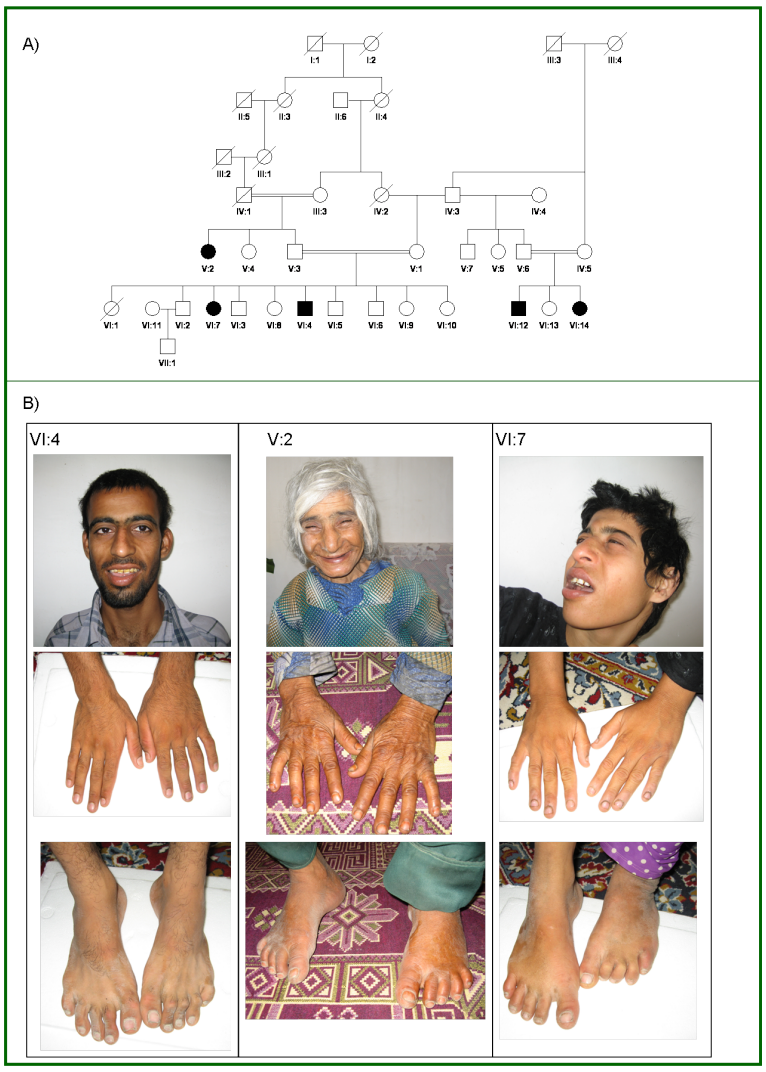


Figure 5.9: Family M314. A) Pedigree, B) Clinical photographs of the patients.

Similar symptoms like severe ID and spasticity and also microcephaly are visible in all of the patients (Table 5.3).

5.2.2 Molecular investigation

Seven DNA samples from this family were genotyped using 50K SNP Affymetrix human array. The genotyping results of two affected and two other healthy siblings and the parents from the core branch and the whole small branch were used for homozygosity mapping.

Patient	Sex	Age at examination	Mental Retardation / IQ	Height	OFC[a]	Additional features
V:1	f	61y	35	152 cm	45cm(<3rd cent.)	long nose,pointed chin,spasticity
VI:4	f	27 y	50	161 cm (10-25th cent.)	52 cm (3rd cent.)	long nose, pointed chin,spasticity, autism
VI:7	m	22y	39	168 cm (25th cent.)	51.5 cm (<3rd cent.)	long nose,pointed chin,spasticity

a: OFC, occipitofrontal circumference

Table 5.3:

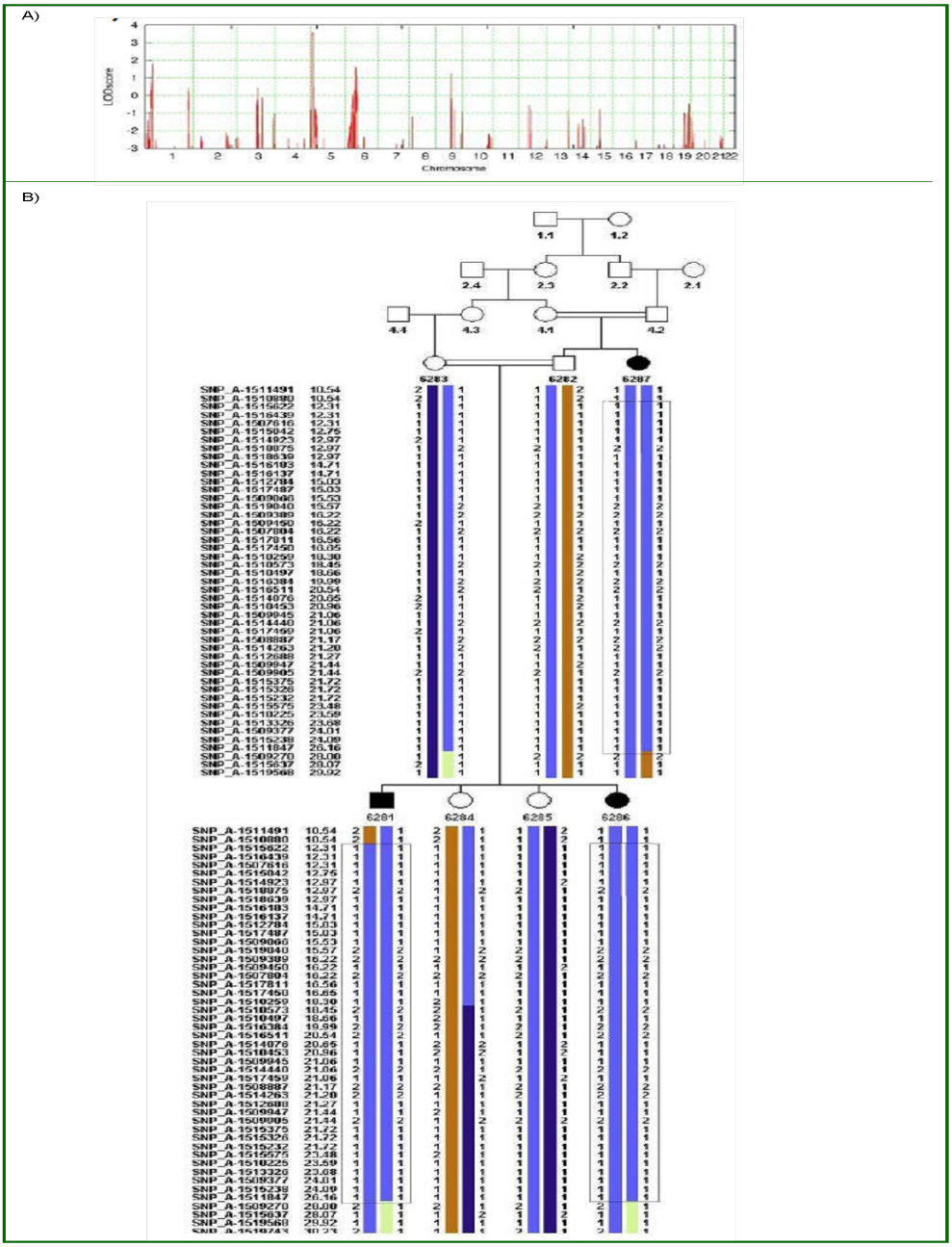


Figure 5.10: A) Whole genome parametric linkage plot of family M314. B) Haplotype for the region on chromosome 5p15.32-p15.2. The homozygous region in all three affected children is indicated by rectangles. The first adjacent heterozygous SNPs from both flanking sides of the interval are shown.

One single Linkage interval with LOD scores 3.5 was found on the short arm of chromosome 5 (MRT5), which spans 5.6 Mbp. These results revealed that family M314 is the second family with a solitary interval in the MRT5 locus.

Mutation screening

NSUN2 is one of the 25 annotated genes (Figure 5.5) in the 6.7 Mbp of MRT5, located on chromosome 5q16.1 - q21. Since this is the second family found for MRT5 locus, *NSUN2* was the first candidate gene to be checked for the defect in this family. PCR and sequencing of *NSUN2* in this family was carried out using the same primers covering the exons and exon-intron boundaries for this gene (Appendix). A T to G transversion was found at nucleotide 3482 in the splicing region, eleven bases upstream from the canonical acceptor splice site of intron 5 (Figure 5.11). This mutation segregates through the affected individuals within the family. To exclude the possibility of the rare genomic polymorphism in this location, 318 Iranian control chromosomes as well as 512 German normal chromosomes for this change have been sequenced and it was indicated that this variant is very unlikely to represent polymorphic alterations in the population. Figure 5.11 shows the sequence of the splicing region in which the point mutation was identified.

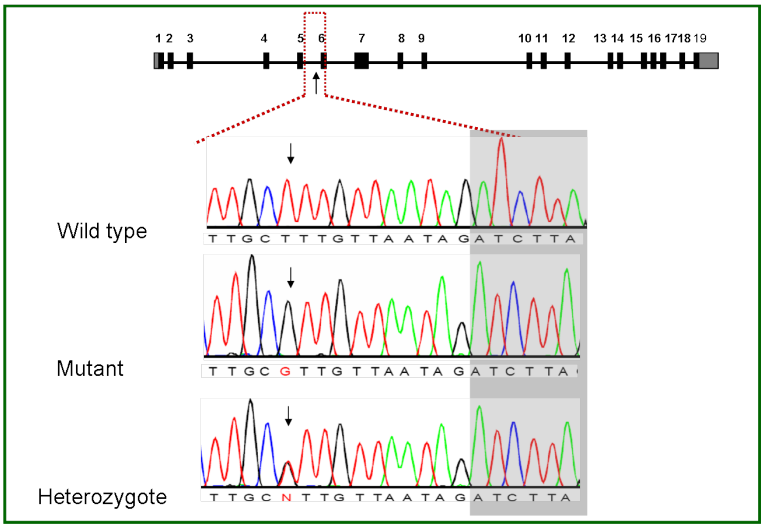


Figure 5.11: A) Schematic view of the *NSUN2* main transcript. The black arrow shows the position of the intronic change T>G. B) Genomic sequence of intron 5 and beginning of exon 6. The gray box indicates the beginning of exon 6.

5.2.3 Impression of the mutation on the *NSUN2* transcript

In order to interpret the pathogenic effect of this variant, *NSUN2* mRNA expression studies, including the search for alternative transcripts, were initiated. Fragment analysis of RT-PCR products spanning exons 4 to 8 showed that the full length transcript is depleted and the second alternative transcript was more prevalent in patients carrying transversion T to G than in controls (Figure 5.12). This mutation inserts an extra purine in the polypyrimidine tract of the acceptor site of exon 6. The change of T>G can affect binding of essential splicing factors such as U2AF65, which directly contacts the polypyrimidine tract [Zamore et al., 1992]. U2AF65

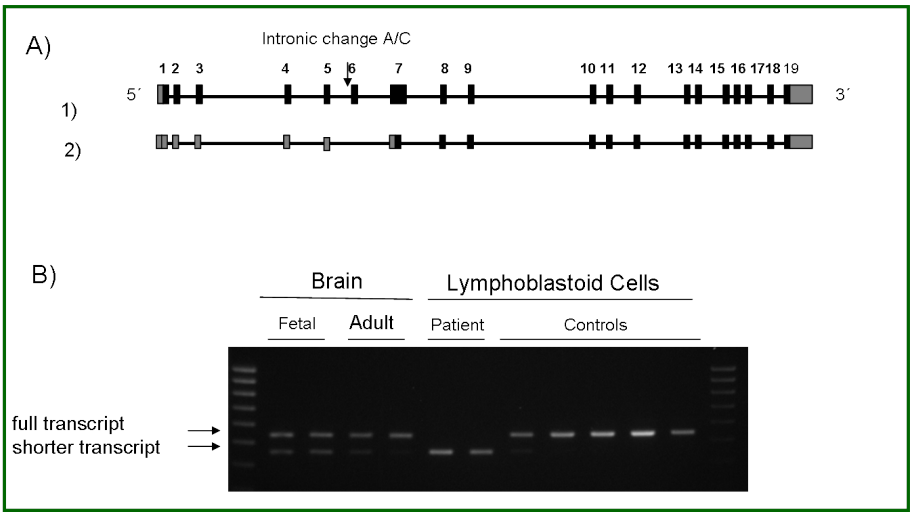


Figure 5.12: The main effect of T>G is skipping of exon 6 as presented by fragment analysis. The full length transcript, representing an out of frame skipping of exon 6 (*NSUN2-Del6*).

is the large subunit of the U2 small nuclear ribonucleoprotein particle (snRNP) uxiliary factor (AF). U2AF is a prerequisite for the association of the U2 snRNP with the lariat branch point region during assembly of the spliceosome [Singh et al., 1995].

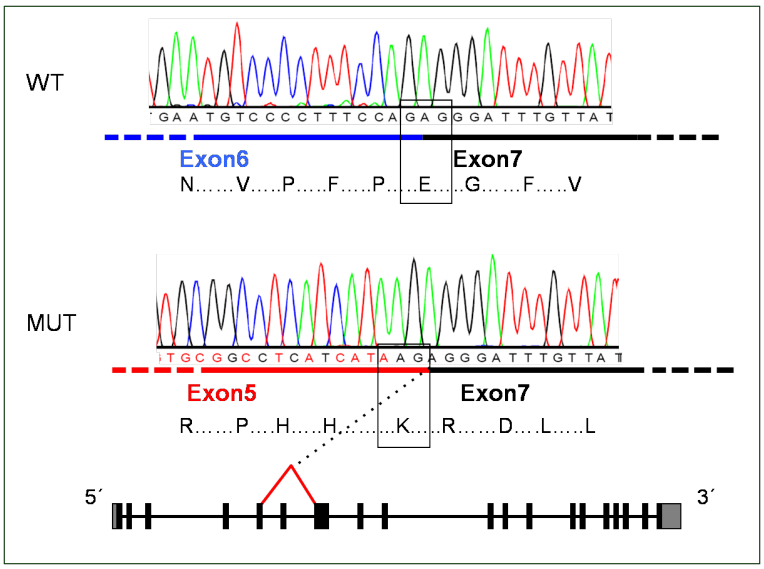


Figure 5.13: Wild type and mutant proteins with the related protein sequence are shown. Mutant NSUN2 protein is out of frame due to splicing out of exon 6.

5.3 The investigation of family G013

5.3.1 Clinical description

The third family contains two mentally retarded children and 4 healthy siblings from the first cousin consanguinity marriage. The clinical features of the family are shown in the following table. The Boy also shows cryptorchidism and epilepsy.

Patient	Sex	Age at examination	Mental Retardation / IQ	Height	OFC[a]
IV:4	f	9 y	moderate	110 cm (<3rd cent.)	48 cm (<3rd cent.)
IV:5	m	6 y	severe	99 cm (<3rd cent.)	47.5 cm (<3rd cent.)

a: OFC, occipitofrontal circumference

5.3.2 Molecular analysis

Blood DNA samples of both parents and the affected individuals IV: 4 and IV: 5 as well as all healthy siblings were genotyped using affymetrix SNP array 250k. The genotyping data has been analyzed to find homozygote regions in the genome. The linkage analysis revealed one homozygous region between SNP markers SNP-A-2286514 and SNP-A-4200250. This linkage interval spans 11.4 Mbp on the chromosome 5p15.32-p15.1, which overlaps with families M192 and M314 on the MRT5 locus.

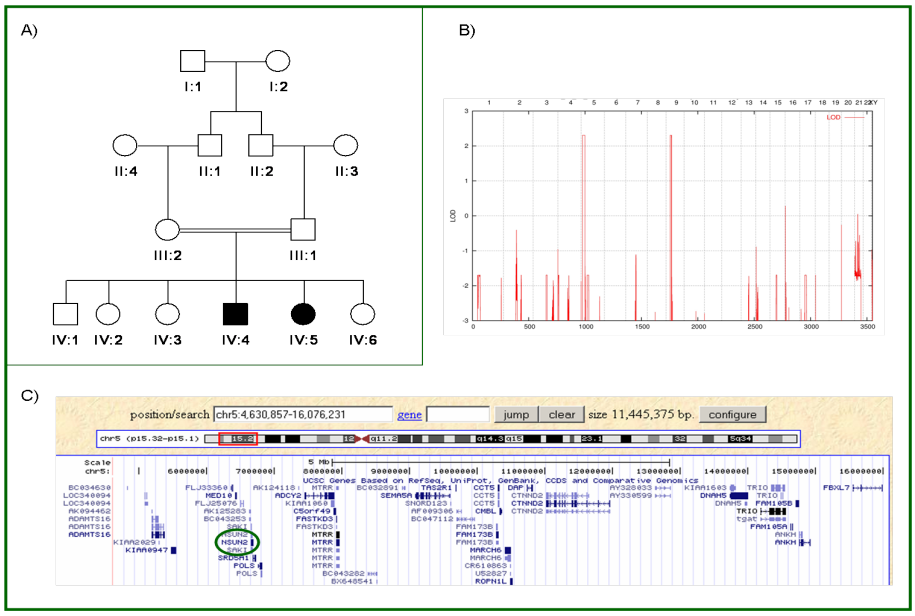


Figure 5.14: The pedigree of family G013 and parametric linkage result. The genomic interval region has been shown.

Identification of mutation

Sequencing of *NSUN2* in this family, which revealed a nonsense mutation Q372X, G114A, has been identified in the exon 11 of this gene. The mutation segregates in the family with intellectual disability. The healthy 180 Iranian controls did not show that this variant is a normal SNP.

To determine whether any of the four splice variants were indeed expressed in the brain, RT-PCR was performed with the RNA available from human adult and fetal brains. With one pair of primers (forward primer in exon seven and the reverse primer in junction of exon eight and nine), a major band corresponding to the full length cDNA and one less abundant, smaller

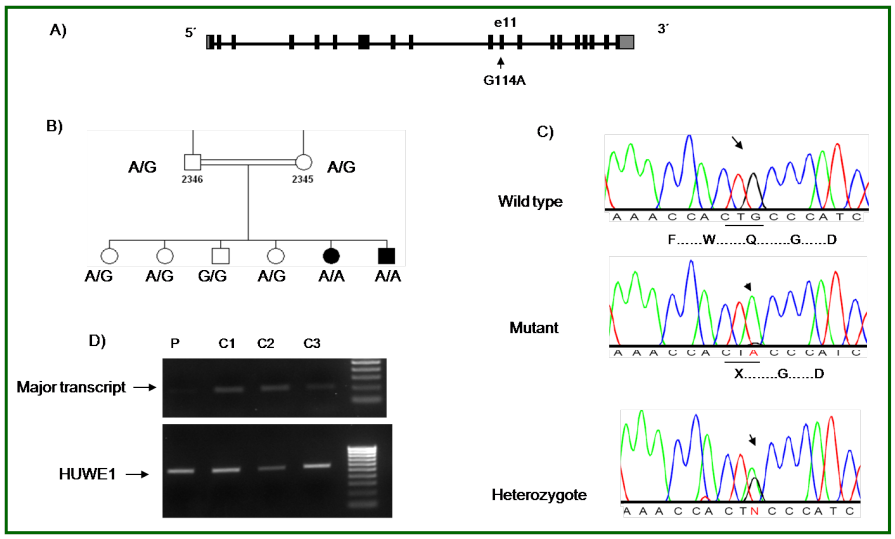


Figure 5.15: A) The schematic view of *NSUN2* main transcript and the position of the mutation.B) Co-segregation of the mutation with ID in the family. C) Chromatograms of the mutation area; D) RT-PCR using cDNA from one patient and 3 controls. *HUWE1* gene was used as a house keeping gene for the control of transcription level.

fragment in the fetal brain was detected. The major splice variant is expressed in fetal as well as adult brains. Most importantly, we could show that the main transcript is expressed in brain tissues which are important in learning and memory like the hypocampus and the temporal lobe (figure 5.16). The third and fourth splice variant results in expression in both brain tissues analysed.

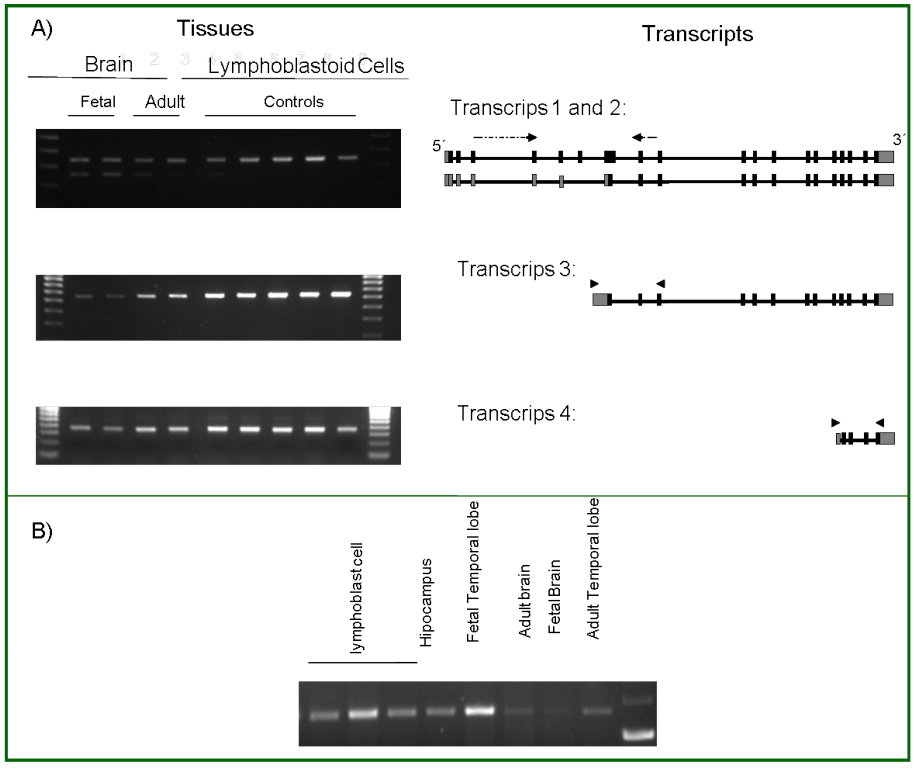


Figure 5.16: The expression pattern of *NSUN2* transcripts in different human tissues

5.3.3 Exploring human NSUN2 function using *Drosophila melanogaster*

Drosophila melanogaster is an ideal experimental organism for the investigation of the genetic background of human diseases. High percentages, about 75% of known human disease genes, have observable analogues with the genetic code of flies [Reiter et al., 2001]. After comparison of *NSUN2* with *D. melanogaster* genome on the flybase webpage (http://flybase.org) with the Basic Local Alignment Search Tool (BLAST), a homologue sequence was obtained. *CG6133* is the *D. melanogaster* ortholog of the human *NSUN2* gene, showing 44% similarity. The exons which contain all three mutations and the methyltransferase domain show 67% identity (Figure 5.17). Based on these results, we will refer from here on to the gene *CG6133* as the Drosophila NSUN2 gene (dNSUN2). dNSUN2 is located on the *D. melanogaster* X chromosome at position 4A5-4A6 and has one known transcript.

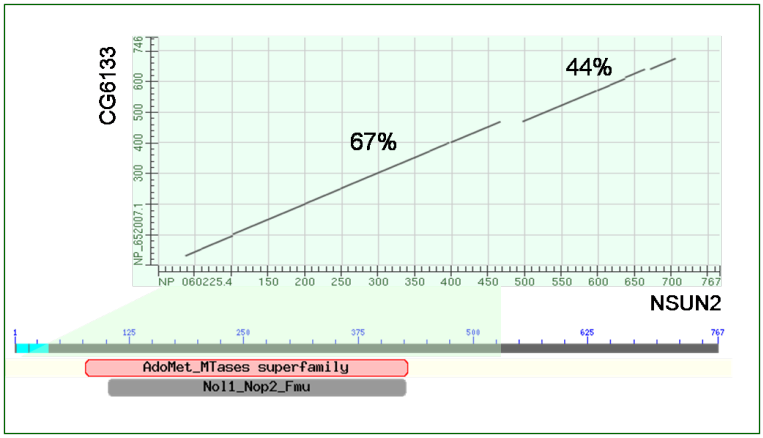


Figure 5.17: *CG6133* (*dNSUN2*) indicates high identity in the methyltransferase domain with human *NSUN2*, location: X: 4,005,189 - 4,007,767; Predicted MW (kDa) 84.2; Length (aa): 746.

5.3.4 Development of the *dNSUN2* deficient *Drosophila melanogaster*

This study has been performed in Prof. Sigrist's lab (Institute of Biology, Genetics, Free university of Berlin Germany) as part of a collaboration. Sara Mertel has performed and managed the drosophila related experiments. The FLP-FRT recombination is a technique to manipulate the genome of *D. melanogaster* with site-specific deletions based on the large array of transposon insertions containing FRT sites. These manipulations include the controlled loss of gene function by meitotic recombination with high precision. It uses the flippase recombination enzyme (FLP) to recombine sequences between two flipplase recognition target (FRT) sites, which must be oriented in the same direction. The existing collection of *D. melanogaster* stocks containing insertions with FRT-sites allowed the assortment of two parental stocks carrying the required piggyBac elements for this specific deletion. The exact positions of the insertions flanking *CG6133* on the X chromosome, while element *pBac(RB)e02478* is located upstream in the minus orientation and element *pBac(WH)dgt4(f06125)* is located downstream in the plus orientation, imply a precise deletion of *CG6133* after a successful excision of the region between the FRT-sites. One should consider that the second element is within the 5'UTR of the dgt4, the downstream gene of CG6133. Therefore *dNSUN2* is of mutant flies ($dNSUN2^{ex1}$/y males),

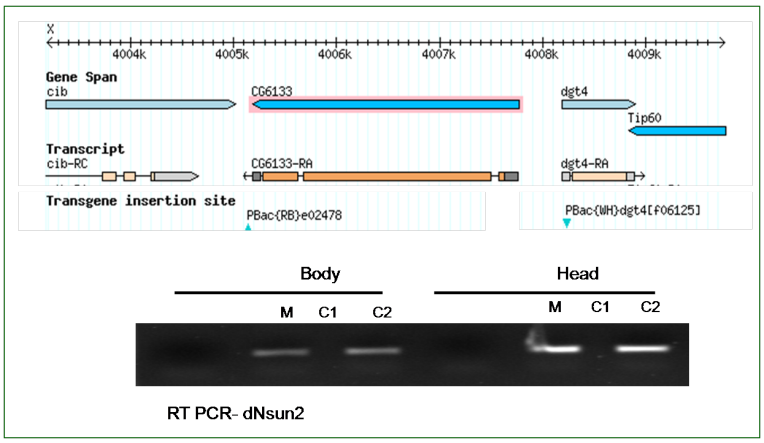


Figure 5.18: Chromosomal deletion of dNSUN2 using FlP-FRT recombination technique

eliminating the complete *dNSUN2* locus and partially deleting the 5' dgt4 locus. The genetic crosses allowed a recovery of deletions in four generations. The confirmation of the deletion can then be provided by PCR analysis with the complete set of specific designed primers (not shown). In order to screen for the loss of *CG6133* expression and to confirm the proper deletion, RT-PCR was applied for the head and the drosophila body separately (Figure 5.18). The forward primer contains the exon-intron boundaries, specifically binding to the cDNA sequence. Since *dNSUN2* is completely deleted in the created mutants, the genetic defects are expected to be very similar to those of the null mutations of *hs-NSUN2*, which were found in ID affected individuals. By creating a *dNSUN2* mutant by specific deletion of *dNSUN2*, ethical unobjectionable and large-scale behavioural experiments could be undertaken to check for ID similar phenotype.

Behavioural screening of mutant flies

Drosophila manifest a wide range of experience-dependent behaviours. This behavioural plasticity has been used both to identify and to characterize genetic mutants [Restifo, 2005; Greenspan, 1995]. Molecular analysis of these genes has revealed that mechanisms of learning and memory are shared among vertebrates and invertebrates. They can learn associations across sensory modalities and remember them for varying amounts of time, depending on the nature and duration of the training [Restifo, 2005].

Disruption of dNSUN2 specifically impairs short term memory in *Drosophila. melanogaster*

To investigate whether associative learning and memory are affected by dNSUN2 deletion, we screened short term memory after odor training in adult *Drosophila*. To rigorously analyse the impact of *dNSUN2* on short-term memory formation, we utilized the Drosophila negatively reinforced olfactory-learning paradigm [Tully and Quinn 1985]. In this paradigm, flies learn to associate an odorant (conditioned $stimulus^{+}$, CS^{+}) with an electric shock (unconditioned stimulus, US), such that the odorant becomes more aversive. The strength of the association and subsequent performance can be modulated by varying the number of pairings between the odorant and electric shock [Beck et al. 2000]. A single pairing of the CS^{+} and the US provides a

modest amount of learning and memory formation, whereas 12 shocks paired to a single odorant will provide a performance plateau [Tully and Quinn 1985; Beck et al. 2000].
We investigated short-term olfactory memory using two genotypes of flies, mutant dNsun2 and W1118 (as background for the experiments). Flies were trained for negative olfactory conditioning by pairing an aversive odor presentation with electrical shocks (12 shocks, 120 volts). Two aversive odors, 3-octanol (OCT) and 4- methylcyclohexanol (MCH), were used. Short term memory (STM) was assessed three minutes after training. Consistent with previous studies, training and testing were performed using dim red light. The number of male flies is determined and the memory performance (performance index = PI) is calculated as the number of flies that avoid the shock-paired odor minus the number of flies that avoid the non-shock-paired odor, divided by the total number of flies.
In our studies, mutant flies demonstrated significantly reduced STM for olfactory learning. We report that depletion of *dNSUN2* is a substantial decrement in associative learning and memory retrieval or stability.

Rescue of the dNSUN2 line phenotype

To gain insights into the memory defect results from disruption of *dNSUN2* and independent of the partial deletion within the *dgt4* gene (the part of 5′ end of *dgt4* overlaps with deletion), the rescue test was conducted employing *UAS/Gal4* system. In Drosophila, the most widely used system for generating spatially restricted transgene expression is based on the yeast GAL4 protein and its target upstream activating sequence (UAS). Gal4 is used to target expression of *UAS-dNSUN2* in specific neuroanatomical regions associated with memory [Brand et al 1993]. The recued flies ($dNSUN2^{ex1}$/Y; +/+; *UAS-dNSUN2/elav-Gal4*) displayed completely normal memory after odor training. This memory type was impaired as expected in the control flies ($dNSUN2^{ex1}$/Y; +/+; *UAS-dNSUN2*/+). These flies are used as negative controls for UAS promotor. Gal4 is not expressed, because of the lack of elav promoter. Therefore UAS promotor is not activated and *NSUN2* does not express.
These results implicate the learning defects of *dNSUN2* mutant was rescued and established that the learning defects of *dNSUN2* mutants are due to disruption of the *dNSUN2* gene by itself.
Collectively, the results of the behavioural analyses indicate that *dNSUN2* is essential for olfactory learning and memory retrieval or stability. These data clearly define a specific, physiological function for *dNSUN2* during memory formation.

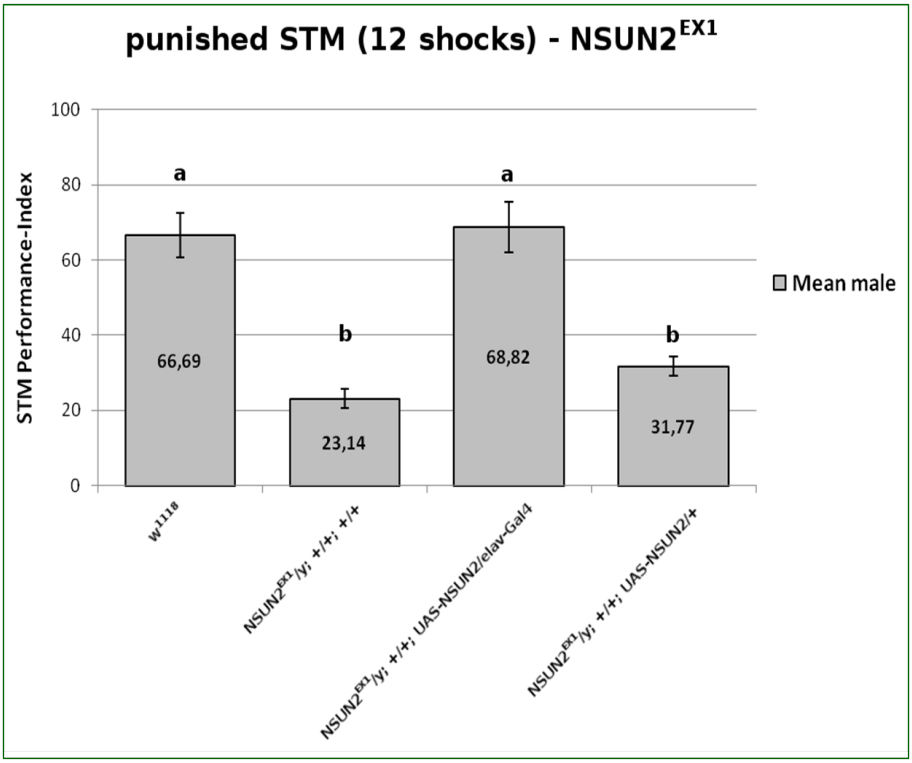


Figure 5.19: *dNSUN2* flies exhibited a significant reduction in short-term olfactory memory. Wild type flies showed normal learning. This learning defect is rescued by transgenes. (Number of male flies and the mean performance are plotted.)

5.4 NSUN2 is known as a tRNA methyltransferase protein

In 1935, after Crick suggested an adaptor molecule should be in the cell which is involved in the translation of genetic messages, transfer RNA (tRNA) molecules entered in the molecular biology research. tRNA is a very small nucleic acid that contains just 74 to 90 nucleotides.
Its structure is in the shape of a cloverleaf, with the loop at the bottom containing the anticodon (Figure 5.20). The anticodon is a sequence of three bases that pairs with the three bases of the codon on messenger RNA in a way that translation, i.e., the synthesis of a protein, can occur [Quigley and Rich, 1976]. A chain of amino acids is formed as each tRNA anticodon pairs with its corresponding codon on the mRNA. Hence, the proper protein is formed as the tRNA follows the rules of the genetic code in pairing with the mRNA [Nelson and Cox, 2005].
Thereafter the presence of such a molecule was experimentally verified and further it became clear that once the amino acid is attached to the cognate tRNA, it is the tRNA that dictates the translocation of the amino acid in the protein synthesizing machinery. Thus tRNA plays a pivotal role in information transfer in the cell. Over the years, the tRNA molecules turned out to be one of the widely studied and well investigated components of the cell. Now the primary structure of more than 180 tRNA molecules from various sources is known.
In general, tRNAs with the appropriate anticodons are expressed in proportion to the codons appearing in mRNAs. The two types of RNA with the highest level of modified nucleosides, rRNA and tRNA, participated in genetic translation. The modifications that happen on the tRNA help to enhance stability and translational efficiency. Methylation reactions explain the great number of these posttranscriptional modifications [Ünal et al., 2004]. tRNA methyltrans-

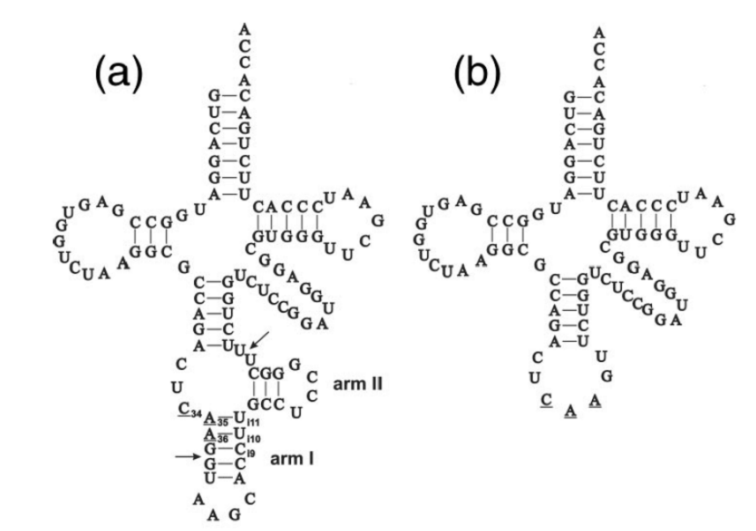


Figure 5.20: Cloverleaf structure of tRNALeu CAA precursor containing intron (a), mature tRNA (b)

ferases (Trms) are enzymes that modify tRNA along the whole length of the tRNA containing or surrounding the anticodon.
Saccharomyces cerevisiae (budding yeast) is the simplest eukaryotic organism used to study post-transcriptional modifications [Drubin et al. 2005]. *S. cervisiae* has 17 tRNA methyltransferase genes that modify tRNA.

5.4.1 tRNA modifications

Eukaryotic cytoplasmic tRNAs maturate in several processes containing: the processing of 5′ and 3′ ends, intron splicing in the case of intron-containing pre-tRNAs, transport from the nucleus to the cytoplasm and multiple nucleoside modifications that occur in the nucleus and in the cytoplasm. Transfer RNA from all organisms has modified nucleosides, which are conjugates of the four nucleosides: adenosine (A), guanosine (G), uridine (U) and cytidine (C). Almost 80 different modified nucleosides have been described in tRNAs from different organisms [Rozenski et al., 1999]. The wobble position (position 34) and the position adjacent to the anticodon (position 37) are two positions in tRNA which are frequently modified [Auffinger and Westhof, 1998; Björk, 1998]. The attendance of modified bases in RNA was first detected in transfer RNA molecules, and methylation was one of the earliest detected modifications [Borek and Srinivasan, 1966; Borek, 1963]. Since then, a large variety of modified bases have been detected in tRNA. These results have been summarised in several review articles [Morton et al., 1984; Srinivasan and Borek, 1964; Nishimura, 1972; Nau, 1976; Feldman, 1977; Bjork, 1983]. The nature and position of modified bases in tRNA are species-specific. Thus there are several bases which are exclusively found in eukaryotes or prokaryotes. Thiolation, for instance, is found only in prokaryotes, whereas methylation of cytosine is restricted to eukaryotes. Also the abundance of specific modified bases at particular positions inside the cloverleaf structure of tRNA differs clearly between eukaryotes and prokaryotes [Morton et al., 1984].
The two procedures that are closely related throughout the maturation of tRNA precursors are decreasing the size and modification. There are three classes of intron-containing tRNA

genes in the human nuclear genome: tRNA-Tyr (8 genes with intron lengths ranging from 16 to 21 bp), tRNA-Leu (5 genes with intron lengths ranging from 22 to 25 bp) and tRNA-Arg (1 gene containing a 15bp intron) (Genomic tRNA database, http//lowelab.ucsc.edu/GtRNAdb) [Brzezicha et al, 2006]. The classic characterization of nuclear intron-containing tRNA genes (Figure 5.20) is that, in all cases, introns are located one nucleotide downstream from the anticodon. The genes which code transfer RNA-Leu in yeast and vertebrates contain introns. -sytosine- which is the first base of the yeast anti-codon sequence in tRNA-Leu, is methylated to 5-methylcytosine (m5C). Methylation relies in yeast is completely on the intron.
The first functional evidence of the importance of m5C34 in tRNA Leu came from the Abelson laboratory [Brzezicha et al, 2006; Johnson et al, 1983]. The construction of mutant yeast tRNA leu genes without introns resulted in the production of mature tRNA molecules without appropriate modified bases. And it has been shown that the decreased suppressor activity of these mutant tRNAs is related to the absence of $m5C^{34}$.
Methylation of $m5C^{34}$ is essential to stabilization of anticodon-codon pairing which leads to the correct coding of mRNA. The nucleotide sequence around the modified position and also the structure of intron-containing prolongate anticodon stem are important for the methylation of $m5C^{34}$ [Brzezicha et al, 2006].
Actually, many tRNA species have a 2'-O-methylated nucleoside at the first position of the anticodon [Satoh, 2000; Murasugi and Takemura et al., 1978; Rozenski et al., 1999], probably for a better codon-anticodon interaction. The modification has never been found in the second or third position of the anticodon of the tRNA species of any organism [Satoh et al., 2000; Sprinzl et al., 1998; Rozenski et al., 1999]. This may mean that the modifications at these positions do not improve the codon-reading activities of natural tRNA species.
Transfer RNA modification might serve as a fine control mechanism for modulating translational efficiency. Thus the presence of modified nucleosides seems to be necessary for the subtle tuning and coordination of tRNA function as well as specific interactions with several protein factors [Brahmachari, 1984].

5.4.2 Enzymes of tRNA modification/methylation

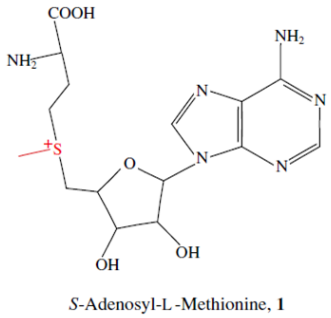
The known enzymes transferring methyl groups from SAM to nucleic acids belong to the SPOUT and MTase super families, the latter containing a Rossmann fold for binding the cofactor [Bujnicki et al., 2004; Motorin et al., 2009].

S-adenosylmethionine (SAM-e) as methyl donor

S-adenosylmethionine (SAMe) is an intermediate metabolite of the essential amino acid methionine. SAM-e functions as a methyl donor in many biological reactions and is a precursor for polyamines. Participation of SAM-e in the methylation procedure helps the growth of tissues and the repair of cells, keeping phospholipids in the cell membrane. It plays a role in the immune system, maintains cell membranes, and helps produce and break down brain chemicals such as serotonin, melatonin, and dopamine as well as vitamin B12. SAM-e also participates in the making of genetic material, known as DNA, and cartilage. Low amounts of folate (vitamin B9) in the body may lead to reduced levels of SAM-e [Bottiglieri, 2002; Crellin et al,

1993]. SAM-e is found in every living cell, with its greatest concentrations located in the brain and liver. Methylation of DNA is critical in the biological phenomenon known as gene silencing. Gene silencing helps suppress genes that may give rise to cancer or those that may carry information for endogenous retroviruses.
Methylation of RNA, particularly tRNA, is similarly important in safeguarding the form and function of these molecules in protein synthesis.
SAMe is the methyl donor for phosphatidylethanolamine in the formation of phosphatidylcholine (PC) [Hirata et al, 1978]. PC is a major component of cell membranes and is vital for maintenance of cellular membrane fluidity, important in sustaining the bioenergetics and information-processing functions of cells. SAMe is also involved in the methylation of histones, major elements in chromosomal structure. This methylation is believed to play a key role in the regulation of DNA transcription, the process by which RNA is formed. SAMe's importance in the body is further emphasized by the fact that it is also the methyl donor for the synthesis of epinephrine (adrenaline), creatine, melatonin, glutathione, the polyamines spermine and spermidine, and the amino acids L-cysteine, all of which play vital roles in human health [Bottiglieri, 2002; Gerner et al, 2004] SAMe may be beneficial to mild depression as well as seasonal and postpartum depression. SAMe appears to have some beneficial effects against arthritic inflammation [Bressa et al, 1994; Bradley et al, 1994]. Further studies with a double-blind design are needed to confirm this preliminary indication that SAMe is a relatively safe and fast-acting antidepressant.

The S-adenosylmethionine (SAM) dependent methyltransferases use SAM, the second most



S-Adenosyl-L-Methionine, **1**

Figure 5.21: Chemical structure of AdoMet. The sulfonium and the S-bound methyl group are highlighted in red.

commonly used enzymatic cofactor after ATP. Aberrant levels of SAM have been linked to many abnormalities, including Alzheimer's, depression, Parkinson's, multiple sclerosis, liver failure and cancer.

5.4.3 Modification of tRNAs

tRNA modification shows a lot of diversity which requires a whole battery of enzymes with singular base and site specificity. In *E. coli* alone six different methylases have been reported by Hurwitz et al. in 1964. Since the modifying enzymes are in limited quantity in the cell, extraction of them is difficult. In addition, the lack of suitable tRNA substrates for each enzyme and

the instability of the enzymes have delayed the characterization of the enzymes. RNA: m^5C-MTases have a large protein family that can be separated to several different groups, based upon sequence similarity and RNA substrate specificity. These groups are shown in figure 5.23.

Enzymes that introduce m^5C34 in the intron containing pre-tRNA Leu have been identified

Enzyme family	Enzyme name	Other names	Organism	Life domain	Accession	Identification type	RNA substrate	Modified position(s)	X-ray structure	Reference
RsmB family										
RsmB/Nol1	RsmB	Fmu/Fmv	*Escherichia coli*	B	AP_004502	Biochemical	16S rRNA	967	1SQF, 1SQG	(68,70)
RsmB/Nol1	P120	NSUN1	*Homo sapiens*	E	P46087	NO	Unknown	NA		(92)
RsmB/Nol1	Nop2		*Saccharomyces cerevisiae*	E	YNL061W	NO	Unknown	NA		(87)
RsmF/YebU family										
YebU			*Haloferax volcanii*	A	HVO_1594	Bioinformatics	tRNA	39/40/48/49		(75)
YebU			*Pyrococcus abyssi*	A	PAB1947	Biochemical	tRNA	Multiple (mostly 49)		(79)
YebU	aTrm4		*Pyrococcus horikoshi*	A	PH1374	Bioinformatics	Unknown	NA	1IXK	(78)
YebU	RsmF	YebU	*Escherichia coli*	B	P76273	Biochemical	16S rRNA	1407	2FRX	(71,72)
YebU	hTrm4	NSUN2/Misu	*Homo sapiens*	E	NM_017755	Biochemical	pre-tRNA^{Leu}	34		(86)
YebU	FLJ22609	NSUN3	*Homo sapiens*	E	Q9H649	NO	Unknown	NA		(65)
YebU	MGC22920	NSUN4	*Homo sapiens*	E	Q96CB9	NO	Unknown	NA		(65)
YebU	Trm4	Ncl1	*Saccharomyces cerevisiae*	E	YBL024W	Biochemical	tRNA and pre-tRNA	48/49 tRNA 34/40 pre-tRNA		(80,81)
Dnmt2 family										
DNMT2	Dnmt2		*Drosophila melanogaster*	E	Q9U6H7	Biochemical	tRNA	38		(102,107)
	Dnmt2	trdmt1	*Drosophila rerio*	E	Q588C1	Biochemical	tRNA	38?		(122)
DNMT2	Dnmt2		*Homo sapiens*	E	O14717	Biochemical	tRNA	38	1G55	(100,102)
DNMT2	Dnmt2		*Mus musculus*	E	O55055	Bioinformatics				
DNMT3	pmt1	pmt1	*Saccharomyces pombe*	E	P40999	Bioinformatics				
RlmI family										
COG1092	RlmI	YccW	*Escherichia coli*	B	P75876	Biochemical	23S rRNA	1962	3C0K	(73,74)
Ynl022 family										
Ynl022c	WBSCR20A	NSUN5A	*Homo sapiens*	E	Q96P11	NO	Unknown	NA	2B9E	
Ynl022c	WBSCR20B	NSUN5B	*Homo sapiens*	E	Q3KNT7	NO	Unknown	NA		
Ynl022c	WBSCR20C	NSUN5C	*Homo sapiens*	E	Q63ZY6	NO	Unknown	NA		
Ynl022c	WBSCR22		*Homo sapiens*	E	O43709	NO	Unknown	NA		
Ynl022c		NSUN7	*Homo sapiens*	E	Q8NE18	NO	Unknown	NA		
Ynl022c	Ynl022c		*Saccharomyces cerevisiae*	E	YNL022C	NO	Unknown	NA		
NSUN6 family										
NSUN6		NSUN6	*Homo sapiens*	E	Q8TEA1	NO	Unknown	NA		

Abbreviations: A: archaea; B: bacteria; E eukaryota; NA not analyzed.

Figure 5.22: RNA: m^5C-MTases protein families [Brzezicha et al., 2006].

in yeast. The first human m^5C methylase Dnmt2 has been reported very recently and is responsible for the modification of C38 in tRNA Asp in mice, *Drosophila melanogaster* and *Arabidopsis thaliana* [Brzezicha et al., 2006]. Three kinds of methyltransferases (MTases) create 5-methylpyrimidine in nucleic acids and form m^5U in RNA, m^5C in RNA and m^5C in DNA. The DNA: m^5C MTases have been extensively studied by crystallographic, biophysical, biochemical and computational methods [Janusz et al., 2004]. DNA and RNA are different with consideration to the number of modifications. Only three modified bases are typically found in DNA: m^5C, N4-methylcytosine (m^5C) and N6-methyladenine (m^6A). RNA MTases in comparison to the DNA MTases which are well-characterized remain poorly described. The RNA: m^5C MTases are a fascinating group of RNA modification enzymes, for which some useful information has been obtained by separate structural, biochemical and evolutionary studies [Janusz et al., 2004].

After characterization of the first, 16s rRNA:m^5CC967 MTase RsmB (previously called Sun or Fmu) from *E. coli*, homologous sequences were found and extra subfamilies for paralogous RNA: m^5C MTase were predicted. From these putative m5C MTases, two eukaryotic pro-

teins were identified: a multisite specific tRNA: m^5C MTase Trm4p and rRNA MTase Nop2p. RsmB (Fmu) and Trm4p protein family have wide distribution; however, they are the only proteins with biochemically confirmed RNA m^5C methyltransferase activity [Tscherne et al., 1999; Janusz et al., 2004]. One interesting feature of RNA: m^5C MTases is that they relate to RNA: m^5CU MTases and DNA: m^5C MTases, which are two distinct classes of enzymes that generate 5-methylpyrimidine in nucleic acids. The enzymatic mechanism of DNA: m^5C methylation has been extensively investigated by crystallography, mutagenesis, biophysical methods and molecular dynamics simulations [Janusz et al., 2004; Huang et al. 2003].

5.4.4 m^5C MTase family

So far, 260 homologs of RNA: m^5C MTases has been identified. In the multiple sequence alignment a frequent domain was found and utilized to create a phylogenic tree of the RNA: m^5C MTase family [Janusz et al., 2004].
The new human RNA: m^5C MTases candidates FLJ22609 and MGC22960 are considered to be most closely related to the Trm4p lineage (including the human Trm4p ortholog FLJ20303). The orthologe protein of the yeast protein Nop2p is the human protein Nol1p [Janusz et al., 2004].
NSUN2 is hTRM4, a human gene that encodes a methyltransferase (MTase) involved in the formation of m^5CC34 in tRNA Leu and acts at the level of the intron containing tRNA precursor. This protein is localized in the nucleoplasm and nucleolus [Brzezicha et al., 2006].

Classes of eukaryotic m^5C MTases

Four different groups of RNA: m^5C-MTases, which show distinction in their sequence and RNA substrate specificity, have been found in eukaryotes. Only three of them were detected in *S. cerevisiae*, but all four groups are exhibited in most eukaryotic genomes.

- Yeast and human Trm4 (Ncl1)

The *S. cerevisiae* Ncl1 protein was first described as a nuclear protein with unknown function (Wu et al., 1998; Motorin et al,. 2010). Later on, the involvement of this protein in the modification of yeast tRNAs and some tRNA precursors at positions 34, 40, 48 and 49 has been identified. Positions 34 and 40 are modified only in the intron-containing pre-tRNA and are specific for tRNALeu(CAA) and tRNAPhe(GAA), respectively [Motorin et al,. 2010]. Based upon this observation, Ncl1 (encoded by ORF YBL024) was named Trm4 for tRNA-specific MTase 4. Sequence similarity with yeast Trm4 also allowed the cloning and characterization of a human homologue called hTrm4 or Misu/NSUN2 [Brzezicha et al., 2006].

- Yeast and human Nop2/p120

The yeast *S. cerevisiae* nucleolar protein Nop2 (encoded by ORF YNL061) functions in the biogenesis of 60S ribosomal subunit and in the maturation of 26S rRNA. This specific operation of Nop2, which is not methylation transferase activity, is apparently significant for the viability of yeast cells. P120, a human proliferation associated protein, shows high similarity to yeast nop2 [Motorin et al,. 2010].

- Higher eukaryotic Dnmt2 and homologs

A new addition to the eukaryotic RNA: m^5C-MTase family is the Dnmt2-related proteins. The first Dnmt2-like protein (pmt1) was described in the fission yeast *S. pombe* and recognized by its substantial sequence homologies to DNA: m^5C-MTases. In light of the findings discussed above, it appears reasonable to assume that Dnmt2 is predominantly a tRNA-MTase and that the weak and highly distributive DNA-MTase activity is a consequence of a secondary enzyme activity with potentially little biological relevance [Motorin et al,. 2010].

5.4.5 NSUN Family

The family of NSUN/NOP2/NOL1 related proteins in humans contains nine different members (*NSUN1* to *NSUN7*, genes *NSUN5A, B* and *C* probably resulted from recent gene duplication). Several mRNA splicing isoforms were identified and some of these variants don't have important catalytic or RNA-binding domains. Most of these genes are highly conserved in mammals.

In most cases, these NSUN proteins retained the putative m^5C-MTase domain bearing two catalytic cysteines. However, there is a lack of knowledge on the activity or the substrate specificity of these potential m5C-MTases so far. The encoded proteins, NSUN5 A, B and C, may function as a DNA methyltransferase in the nucleus. This genes are deleted in Williams syndrome, a multisystem developmental disorder caused by the deletion of contiguous genes at 7q11.23 [Doll et al., 2001]. One of the more intriguing potential functions of NSUN7 is in spermatozoal protein translation. A mutation within this gene cause reduced sperm motility and infertility in male mice [Harris et al,. 2007].

5.4.6 Subcellular localization of eukaryotic m^5C MTases

The subcellular localization of RNA modification enzymes can provide important indications about their biological functions. Motorin et al. in the figure 5.24 gives an overview of the current knowledge on the subcellular localization of methylated RNAs and modification enzymes. Most m^5C-MTases are predicted to be nuclear or nucleolar proteins which corresponds well to their functions in tRNA and rRNA processing. The localization of different m^5C-MTases in eukaryota is not known completely. Nop2 and also human p120 are nucleolar and nucleolar proteins. Dnmt2 has either a nuclear or a cytoplasmic subcellular localization. This could indicate the complex functions of Dnmt2.

5.4.7 *NSUN2* encodes yeast *TRM4* orthologue

Yeast Trm4p enzyme introduces m^5Cinto yeast pre-tRNA Leu in an intron-dependent manner. EST sequences of putative human orthologue shows 35% identity to the yeast sequence. This gene is called NSUN according to HUGO Gene Nomenclature Committee and contains 19 exons and encompasses 34000nt [Brzezicha et al., 2006].

5.4.8 NSUN2 catalyses the formation of m^5C34 in human pre-tRNALeu

Brzezicha and his collegues could investigate NSUN2 activity in yeast. They could show that the activity of hTrm4 relies completely on a specific sequence surrounding the cytosine to be

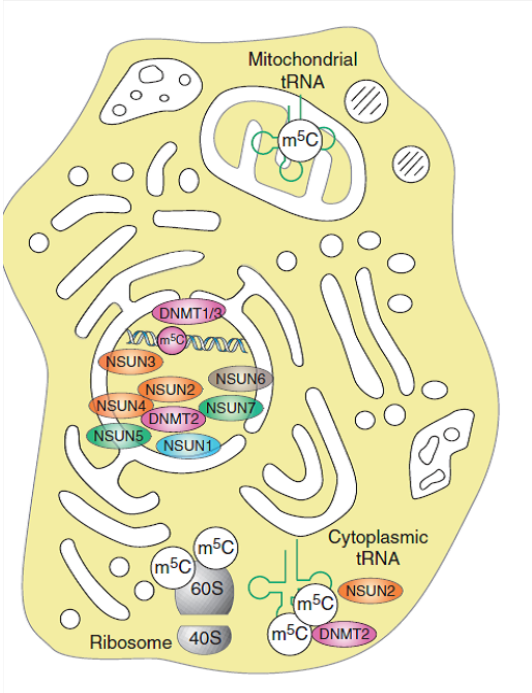


Figure 5.23: Subcellular localization of m5C residues in RNAs and m^5C-MTases in higher eukaryota [Motorin et al,. 2010].

modified and on the intron structure. The nucleotide sequence that surrounds C34 in human tRNA precursor is more limiting for m^5C34 formation than in the yeast. If the adenosine which is the middle base in anticodon sequence is replaced with Uridine, cytosine in position 34 doesn't methylate the tRNA precursorin human cells.
Furthermore, human MTase responsible for the introduction of a methyl group at the anticodon wobble cytosine (*hTrm4*) requires a consensus nucleotide sequence C/A/U32-U/A33-C34-A35-A36-G37 for m5C34 formation. Thus, all nucleotides upstream from the intron sequence in the anticodon loop are necessary for pre tRNA-Leu modification.
Yeast TRM4p is also required for C48 and C49 methylation in different yeast tRNAs. In contrast, NSUN2 is not able to introduce a methyl group at position C48/49 of any tRNA tested [Brzezicha et al., 2006].
NSUN2 has no methyltransferase activity in unmethylated DNA. In contrast, NSUN2 has a methyltransferase activity in hemimethylated DNA.

5.4.9 NSUN2 is involved in the regulation of nucleolar architecture and nucleic acids metabolism during mitosis

The evolutionary conservation of NSUN2 suggests that it plays important biological roles next to methylation of RNAs [Shiho Sakita et al., 2007]. This fact can be supported by high expression of NSUN2 in cancer cells. During the mitotic phase, the nucleolus disassembles in higher eukaryotic cells. Phosphorylation of components of the rDNA transcription machinery at the beginning of mitosis can induce nucleolar disassembly i.e., cyclin-dependent kinase (CDK) repress the rDNA transcription [Heix et al., 1998; Sirri et al., 2000]. The RNA synthesis machinery is inactive at nucleolar-organizing regions [Stoykova et al., 1985; Shiho Sakita et al., 2007].

CDK1-Cyclin B phosphorylates NPM1 and nucleolin [Peter et al., 1990], and phosphorylation of nucleolar components are involved in the disassembly. Inhibition of CDK1-Cyclin B during mitosis is sufficient to cause the resumption of rDNA transcription, but it is not sufficient to restore proper processing of RNA biogenesis or total rebuilding of the nucleolar processing machinery [Sirri et al., 2002]. Therefore, repression of the nucleolar assembly is regulated. Aurora-B is a conserved protein kinase essential for the segregation of eukaryotic chromosomes and it forms the mitotic passenger protein complex with inner centromere protein (INCENP), Survivin, and Borealin/Dasra [Shiho Sakita et al., 2007; Carmena and Earnshaw, 2003; Gassmann et al., 2004].

5.4.10 NSUN2 is phosphorylated during mitosis by Aurora-B

NSUN2 is a novel substrate of Aurora-B. The Aurora-B phosphorylation site of NSUN2 is conserved among vertebrates. Three phosphorylation sites have been identified on NSUN2 (Ser456, Ser593, and Ser743). The latest result and those of Beausoleil et al. indicate that phosphorylation of NSUN2 on Ser139 is mediated by Aurora-B and that this is one of many sites phosphorylated *in vivo*.

5.4.11 Nucleolar disassembly-reassembly processes under Aurora-B inhibition

During mitosis, NSUN2 is phosphorylated by Aurora-B to suppress its methyltransferase activities and the association of NSUN2 with NPM1 is inhibited. NSUN2 is rapidly phosphorylated by Aurora-B and probably occurs in parallel with nucleolar disassembly during mitosis. CDK1-Cyclin B is a major regulator of this disassembly [Sirri et al., 2002]. When ribosome disassembles, nucleolar proteins may be phosphorylated by multiple kinases at multiple sites. Thr4, Thr199, Thr219, Thr234, and Thr237 in NPM1 are known as CDK1-Cyclin B phosphorylation sites [Peter et al., 1990; Okuwaki et al., 2002]. The interaction of NPM1 with nucleolin is inhibited during mitosis [Liu and Yung, 1999]. NSUN2 is associated with nucleolar proteins NPM1 and nucleolin. Phosphorylation of NSUN2 inhibits this interaction during mitosis [Shiho Sakita et al., 2007]. During mitosis, NSUN2 is phosphorylated by Aurora-B to suppress its methyltransferase activities and the association of NSUN2 with NPM1 is inhibited [Shiho Sakita et al., 2007].

6 Discussion

6.1 Identification of a nonsense mutation in the very low-density lipoprotein receptor gene (*VLDLR*) in an Iranian family with dysequilibrium syndrome

We report the first dysequilibrium syndrome (DES) family outside the Hutterite population and our results show a deleterious mutation that exclusively affects *VLDLR* in patients with clinical features that are indistinguishable from previously reported cases with a complete lack of *VLDLR*. Whereas the size of the earlier reported deletion could not exclude an additional contribution to the genotype from neighbouring genes, our finding clearly demonstrates that *VLDLR* deficiency alone is sufficient to cause the human DES phenotype [Moheb et al., 2008]. Since our finding, six mutations in *VLDLR* have been identified in 5 families [Boycott et al., 2009].

Two isoforms of the protein are known, the full-length version (type I) and a version lacking an O-linked sugar region (type II) [Sakai et al., 1994]. Type I is mainly distributed in heart and skeletal muscles with active fatty acid metabolism, whereas VLDLR type II is predominant in non-muscular tissue, including kidney, spleen, adrenal gland, lung, brain, testis, uterus and ovary, but not in the liver [Takahashi 1995, Webb 1994]. Like the deletion carriers in the Hutterite family [Boycott et al., 2005], our patients can be considered to lack a functional VLDLR transcript, since the stop mutation we report affects both isoforms. The differences in tissue distribution together with divergent ligand specificity between the two isoforms suggest that they play distinct roles in various tissues and cells. As the DES phenotype comprises mostly central nervous system features, it has to be assumed that differences in compensation of VLDLR deficiency are responsible for the functional integrity of the other tissues [Moheb et al., 2008].

In the brain, VLDLR is part of the reelin, which contributes to the correct regulation of neuronal migration [Arcangelo et al., 1995; Rice et al., 2001; Tissir et al., 2003]. Two lipoprotein receptors, VLDLR and Apolipoprotein E receptor-2 (APOER2), are involved in the reelin (RELN) signalling pathway, which is an evolutionarily highly conserved pathway, and result in phosphorylation of the intracellular adaptor protein Disabled-1 (DAB1) [Hiesberger et al., 1999; Boycott et al., 2009]. This phosphorylation step acts as the 'master switch' and activates an intracellular signalling cascade that allows neuroblasts to get to the crucial parts needed to form ordered cortical layers, for example, by guiding neuroblast migration in the developing cerebral cortex and cerebellum [Hong et al., 2000; Boycott et al., 2009]. The cerebellum is a complex neurological structure, containing more than half of the brain's total number of neurons [Turkmen et al., 2009]. Cerebellar networks show long-term synaptic plasticity, which indicates that experience-dependent adaptive and learning processes are a salient feature of cerebellar function. Because the stop codon mutation is located in the extracellular domain of VLDLR (Figure

6.1), the encoded mutant proteins could not be inserted into the membrane and could not function as receptors for reelin. Mutations which cause loss of function in the human RELN gene

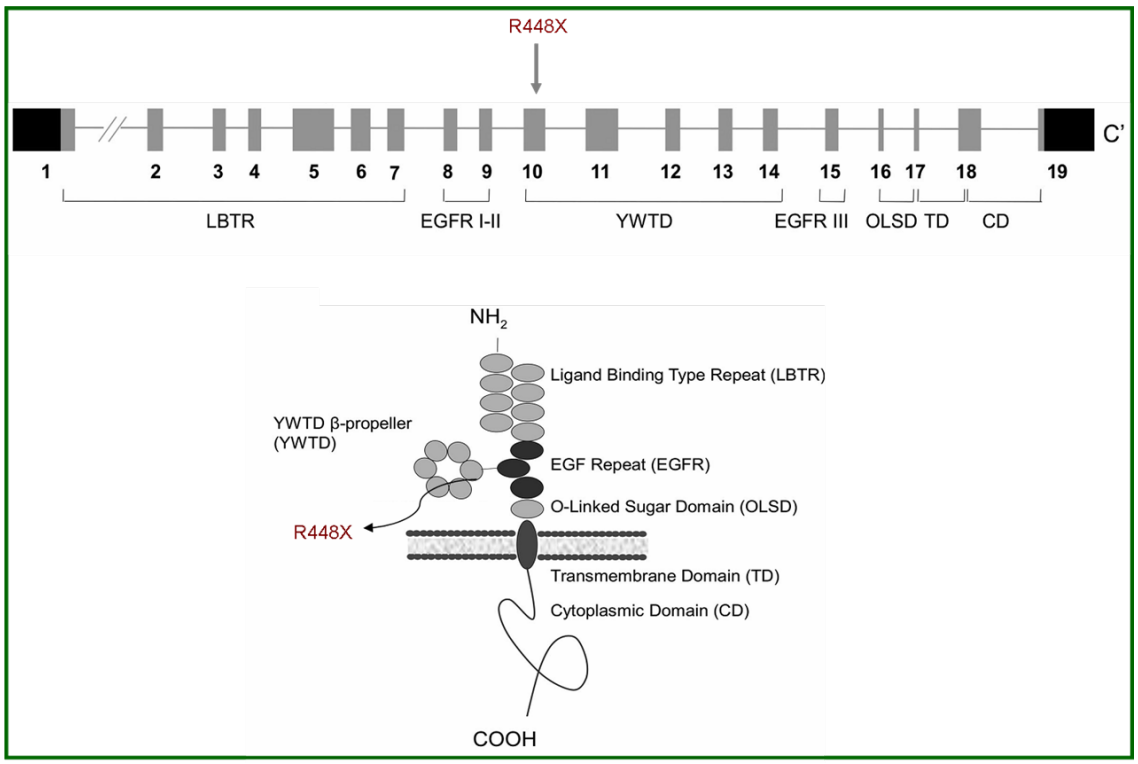


Figure 6.1: The schematic position of the mutation inside the VLDLR protein.

are the reason for an autosomal recessive syndrome composed of severe cerebellar hypoplasia together with lissencephaly [Boycott et al., 2009]. Patients with RELN and VLDLR mutations have less cortical thickening, absence of a cell-sparse zone and profound cerebellar hypoplasia. As discussed previously by Boycott et al., a possible explanation for the DES phenotype might therefore be that neurons in the cortex fail to distribute normally after reaching their assigned layer, as observed in VLDLR-deficient mice.

This type of disequilibrium syndrome typically leads to either delay in learning to walk or even disability to walk. The families reported by Ozcelik et al., 2008, as well as by Turkmen et al., 2008, have been found to be associated with quadrupedal mobility in humans, although not in all affected individuals. Given the variable incidence of quadrupedalism in individuals with mutations in the same gene, it is possible that environmental factors during development "either internal or external" contribute to this particular phenotypic outcome [Humphrey et al., 2008]. However, Boycott proposed that a behavioural adaptation to the severe orthostatic instability can easily explain quadrupedal movement [Boycott et al., 2009].

In addition to lipoproteins, the number of ligands is increasing for the LDL receptor-related proteins and essential signal transduction and modulator functions in embryonic development, synaptic transmission and in the maintenance of vascular integrity are now becoming apparent. DES is the first human lipoprotein receptor abnormality syndrome due to VLDLR muations. More recently, mutations in LRP2 (megalin) are responsible for Donnai-Barrow syndrome, which is defined by agenesis of the corpus callosum, ocular anomalies, developmental delay, congenital diaphragmatic hernia, facial dysmorphism, and sensorineural hearing loss.

We suggest that VLDLR-deficiency in the brain, at a key stage of development, leads to abnormal formation of neural structures.

6.2 Mutations in the NSUN2 gene cause autosomal recessive intellectual disability in Middle Eastern populations with elevated frequency

The three mutations led to the discovery of a novel gene that was not previously suspected to play a role in human cognition. The identification of two nonsense mutations and one splice out mutation, leading another premature protein, is an important step towards elucidating the actual cause of the brain defects observed in the patients.
NSUN2 is the first SUN-domain-containing protein to be characterized in vertebrates [Frye and Watt, 2006], and is well conserved from bacteria to human. Function of NSUN2 in humans is not yet known.
This protein is a novel mammalian RNA methyltransferase, which was independently discovered in a study of myc-induced proliferation mediators in epithelial cells [Frye and Watt, 2006]. The original working name of the protein was "substrate of Aurora-B kinase" (SAKI; accession no. AB255451), and by Frye and Watt (2006) was named myc-induced SUN-domain-containing protein (MISU; accession no. DQ490066). The protein is now called "NSUN2". NSUN2 contains an NOL1/NOP2/ sun domain. This domain is found in archaeal, bacterial, and eukaryotic proteins. It is possible that more insight into the function of NSUN2 can be derived from the functions of other homologs. Homology studies revealed that orthologs of NSUN2 are in organisms as distant as yeast (*S. cervisiae*). The nucleolar protein Nop2 is identified to be a RNA m5C methyltransferase in *S. cerevisiae* [Katz et al., 2003] and is necessary for the production and the maturation of rRNA [Shiho Sakita et al., 2007; De Beus et al., 1994; Hong et al., 1997].
Another protein related to Nop2 in yeast is Ncl1 [Wu et al., 1998]. However, Ncl1 is not essential for cell survival. Ncl1, which is also named Trm4, has tRNA m5C methyltransferase activity, and catalyzes the methylation of other RNA molecules [Motorin and Grosjean, 1999]. Hence, it probably plays an essential role in translation.

- The possible molecular causes of ID by NSUN2 impairment:
It is of note that among the four NSUN2 transcripts, only the full length transcript has a methyltransferase function [Frye and Watt, 2006]. Two of our mutations affect the main transcript and the third mutation is also located within the methylation domain indicating, most likely, that intellectual disability in patients is due to lack of enzymatic function and not to spindle stability regulation.

6.2.1 Impairment of binding of NSUN2 to SAM

Based on the studies in lower organisms, TRM4p uses s-adenoslymethionin (SAM) for methylation of tRNA, rRNA and hemimethylated DNA. Since *Nsun2* is a human homolog of TRM4p of the yeast, impairment of binding to SAM due to the loss of function of NSUN2 may have consequences in the ID phenotype. This is one of the assumptions of FTSJ1 impairment as well. *FTSJ1*

gene has a methyltransferase domain. One hypothesis regarding the involvement of FTSJ1 in cognitive function is based on its capacity to bind S-adenosylmethionine (SAM)[Ramser et al., 2004]. SAM is the major methyl donor for cellular methyltransferase reactions, and low SAM concentration in cerebrospinal fluid or plasma was observed in several neuropsychiatric and neurological disorders, including autism, depression, brain ischaemia, and dementia [James et al., 2004, Ramser et al., 2004].

6.2.2 Impairment of epigenetic regulation of gene expression

NSUN2 methylates hemimethylated DNA as well as tRNA. This elevates the likelihood that NSUN2 is implicated in alterations not only of RNA methylation patterns but also of genomic methylation patterns. Alterations in DNA methylation patterns can cause changes in gene transcription patterns and can also promote mutational events [Siedlecki et al., 2006; Jones and Baylin, 2002; Robertson, 2005]. Aberrant changes in the expression patterns of genes caused by cytosine methylation are called epigenetic mutations, or epimutations.
These epigenetic modifications can cause specific changes in brain functions [Zschocke et al., 2002; Zhao et al., 2003; Hong et al 2003]. Differentiation of oligodendrocytes in neonatal cortical progenitor depends on histone deacetylation. Histone deacetylase activity can be stopped by trichostatin A (TSA), which inhibits the progression of progenitors to mature oligodendrocyte [Marin-Husstege et al., 2002].
Promoter hypermethylation causes gene inactivation and is associated with many diseases, including cancers and neurodevelopmental disorders like ICF (immunodeficiency, centromeric instability and facial anomalies), Rett syndromes, and genomic imprinting deficiencies (Jones and Baylin, 2002). Fragile X (FMR) and ATRX syndrome are other intellectual disability disorders that result from abnormal DNA methylation machinery [Robertson et al., 2000].
Aberrations in genomic imprinting also contribute to diseases. Beckwith-Wiedemann (BWS), Prader-Willi (PWS) and Angelman syndrome (AS) are some examples. [Costello et al., 2001]. PWS and AS patients also have intellectual disability.
One abiding question concerns the reason intellectual disability is such a prominent feature of diseases linked to mutations in chromatin proteins. The trivial explanation is that brain function depends on the integrated actions of a diverse array of genes and therefore is most sensitive to the perturbation of mass gene expression.
Mammalian DNA methyltransferase-2 (Dnmt2), in addition to DNA, methylates tRNA and is a relatively new addition to the eukaryotic RNA: m5C-MTase family, suggesting eukaryotic DNA methyltransferases were derived from ancestral RNA methyltransferases rather than prokaryotic restriction DNA methyltransferases [Goll et al., 2006]. Thus, eukaryotic RNA or DNA methyltransferases possibly have broad substrate specificity in nucleic acids, even though their sequence and their organization of catalytic motifs are characteristic of RNA or DNA. It will be fascinating to explore whether NSUN2 plays a similar role.
Another, currently more speculative role for RNA methylation is the regulation of epigenetic inheritance patterns. While it is clear that most phenotypic traits are inherited through DNA, there is also evidence for RNA-dependent inheritance of certain phenotypes [Rassoulzadeganet al., 2006; Motorin et al., 2010].
Recent developments in the field of RNA methylation open a large perspective for further anal-

ysis of this important biological process. Enzymatic activity, catalytic mechanisms and RNA-recognition specificity of the corresponding RNA: m5C-MTases clearly require better characterization. The biological role of m5C residues in RNA has been largely ignored and further studies using genetically engineered models are needed to investigate this point.

6.2.3 The subcellular localization of RNA modification enzymes can provide important indications about their biological functions

The localization of NSUN2 in nucleoli is consistent with recent evidence demonstrating that the majority of tRNA is processed in nuleoli in yeast [Bertrand et al., 1998; Thompson et al., 2003]. The majority of the nucleolar proteins have functions in ribosome biogenesis, including synthesis and processing of rRNA. Therefore, the nucleolus becomes the place for almost all of the RNA editing, increasing the possibility that the positions of tRNA, rRNA, and mRNA are spatially coordinated . Frye and others could show that NSUN2 knockdown in Myc-induced cells, inhibits increasing the size of the nucleus [Frye and Watt, 2006].
Interestingly, NSUN2 distribution has never been described for RNA MTases, except for ATRX [Frye and Watt, 2006; De La Fuente 2004]. The ATRX protein is another ID related protein, which contains conserved domains, including a plant homeodomain-like zinc finger domain shared with de novo methyltransferases (DNMT3A/B and 3L) and a switch/sucrose nonfermenting family ATPase domain. ATRX is a centromeric heterochromatin binding protein and methylate repetitive DNA sequence.

6.2.4 NSUN2 might play a role in cell cycle regulation

The methyltransferase activities of NSUN2 are suppressed by Aurora-B-provoked phosphorylation during normal mitosis. Phosphorylation of NSUN2 by aurora B suppresses its methylpherase activity and dissociates it from the nucleolar protein nucleoposmin (NPM1) [Sakita-Suto et al., 2007]. Therefore phosphorylation of NSUN2 by a central mitotic kinase (Aurora B) increases the possibility that NSUN2 might play a role in cell cycle regulation.
Moreover, NSUN2 associates with and regulates the localization of nucleolar and spindle-associated protein (NuSAP), an essential spindle assembly factor. Additionally, NSUN2 and its bound RNA partner are required for proper spindle assembly and chromosome segregation.

6.2.5 NSUN2 defects can impair tRNA modification and impact on codon usage

tRNAs are the class of RNA molecules with the highest number of nucleoside modifications. However, only a few mammalian tRNA methylases have been characterized [Frye and Watt, 2006; Armengaud et al., 2004]. The different modifications can affect the maintenance of tRNA structural integrity, translational efficiency, or fidelity [Frye and Watt, 2006].
Since tRNA is a major participant in protein synthesis, tRNA has a potential role in the cell's damage response. tRNA methyltransferases (Trms) are enzymes that modify tRNA along the entire length of the anticodon and the region around it. Since modification systems may be important for stress signalling, Trms have a potential, and perhaps crucial role, in enhancing the synthesis of proteins that participate in damage response. If a gene that codes for a specific

tRNA methyltransferase (like NSUN2) is absent, the cell cannot perform the functions of the missing Trm. Hence, the modification which normally the Trm catalyze, does not occur and the cell may not be able to respond to damage.
Codon usage in mammals is known to have dramatic effects on the translation rate, especially during cell differentiation. The existence of systematic tissue-specific codon usages raises the important possibility that human tissues may differ in their relative tRNA abundances and that these differences may modulate the expression of the appropriate proteins. Genes selectively expressed in one human tissue can often be distinguished from genes expressed in another tissue purely on the basis of their synonymous codon usage. It is possible that defects in modification of tRNA-leu(CAA) leads to a different expression pattern of genes in the brain, because of different levels of tRNA abundance.
Despite the fact that *NSUN2* is ubiquitously expressed in a broad range of human tissue, loss of function mutation of *NSUN2* caused intellectual disability without apparent morphologic abnormlities in the brain or other non-neurologic symptoms, suggesting that mutations in the *NSUN2* gene only affect cognitive functions in humans. It is possible that the activity of NSUN2 is most critical during brain development, which is supported by a relatively high expression of NSUN2 in the fetal brain. It is also plausible that brain structures are more sensitive to defects in the translation machinery than other organs. Homologous methyltransferases modify untranslated RNAs and thereby play critical roles in protein translation. The association of NSUN2 with intellectual disability highlights the importance of this process, especially in brain development and cognitive processes.
Further investigation of the function of the *NSUN2* gene in the human brain will elucidate the pathoetiological basis of cognitive dysfunction present in patients with ARID.

6.2.6 NSUN2 impacts on short term memory in *Drosophila. melanogaster*

By using *Drosophila* as a simple model organism, we demonstrated that NSUN2 mediates learning and short-term memory. Our observations, that NSUN2 loss-of-function mutants displayed learning defects, suggest that precise regulation of NSUN2 is necessary to maintain optimum learning. Therefore, a new gene, *NSUN2*, has been identified for memory formation in *Drosophila*, the common fruit fly.
Because *Drosophila* learning genes are known to be conserved in higher organisms including humans, mutations in this genes often provide new insights into human brain disorders. For example, the *Drosophila* gene known as dunce provided clues to the genetics of the devastating psychiatric condition of schizophrenia. Recent studies have revealed that the human version of the dunce gene is a susceptibility determinant for schizophrenia. In a similar way, any new learning gene identified in *Drosophila* may provide new clues to genes involved in human neurological or psychiatric disorders. Genetic screens, which identify suppressors of nebula in the learning and memory pathway, will also provide insights into the underlying mechanism of mental retardation.
Despite the hundreds of millions of years of evolutionary distance that separate Homo sapiens and *Drosophila melanogaster*, ID-associated molecules are remarkably well conserved between the two species. 87% of nearly 300 molecularly identified human ID genes [Inlow and Restifo,

2004] have been identified to be highly similar to *Drosophila*, using comparative sequence analysis [Restifo, 2005].
Furthermore, the extent and type of amino acid sequence similarity between the human ID genes and their fruit fly counterparts suggest 76% similarity in the two species' biological functions [Inlow and Restifo, 2004]. The best characterized *Drosophila* ID gene counterpart is *Drosophila* fragile X mental retardation 1 (*dfmr1*). Five of the fruit fly counterparts of human ID genes are already known to be essential for normal learning and memory (*FlNA*, *FMRI*, *GNASI*, *NF*I, *RSK2*). Human clinical disorders resulted by these genes are Periventricular heterotopia, Fragile X syndrome, Albright hereditary osteodystrophy, Neurofibromatosis 1 and Coffin-Lowry syndrome [Restifo, 2005].
Enthusiasm for using Drosophila in the study of ID is bolstered by the recent development of fly models of neurodegenerative disorders and successful drug treatment to normalize behavioural and anatomical phenotypes [Bonini and Fortini, 2003].
The recent successful pharmacological rescue of *dfmr1* phenotypes and the recent report of successful treatment of *dfmr1* mutants take the fly model of FXS one giant step forward. When *dfmr1* mutant larvae are fed a diet supplemented with one of several mGluR antagonists, the resulting adult flies show normal memory in the courtship conditioning assay and normal mushroom body morphology [McBride et al., 2005]. With the recent report of successful treatment of *dfmr1* mutants, there is an extraordinary opportunity to use the Drosophila system as a drug-screening tool for ID.
It is still early in the process of making connections between *Drosophila* memory and learning genes and the pathology of human disease, but it's already clear that many of these genes will provide important conceptual information and potential insights into human brain disorders. In addition, there is every reason to believe that their gene products will one day become the target of new drugs to enhance cognition. Uncovering the new genes and their signalling pathways helps bring us that much closer to this goal.

6.3 Two independent mutations in the *ZNF526* gene cause unspecific autosomal recessive intellectual disability

To date, only six loci have been implicated in non-syndromic ARID (NS-ARID) and only two show more than one mutation in independent families. Here we report on patients with this disorder from three families, belonging to our growing cohort of more than 300 large consanguineous Iranian families, where we found independent mutations in the DNA-binding zinc finger gene, *ZNF526*. One of these changes was observed in two separate families with non-syndromic ID, but haplotype analysis revealed that these families must be distantly related. In these families, no other potentially disease-causing and co-segregating change could be identified, which confirmed that impairment of *ZNF526* gene regulatory function is responsible for the ID observed in homozygous mutation carriers. ZNF526 is a transcription factor protein, containing fourteen Cys2-His2 type zinc finger domains.
In this study, we showed that ZNF526 localizes to the nucleus in a number of cell types (Hela cells and neuroblastoma stable cells). Due to the fact that *ZNF526* encodes a transcription factor, which typically switches on cascades of other genes, translocation of ZNF526 to the

nucleus is an essential step in the pathway by which transcription is regulated. We further demonstrated that both the R459Q and Q539H mutant proteins were found to be expressed and to localize to the nucleus.
Since wild type *ZNF526* and both mutants were sequestered in the nucleus, we investigated the potential effects of these two proteins on cellular RNA expression. An Illumina gene microarray was employed to hybridize cDNA from patient cells and neuroblastoma transduced with both native ZNF526 and all the mutants. Patients lymphoblastoid cell lines (LCLs) are most frequently used for RNA expression profiling experiments and they have provided useful and biologically meaningful information; however, they do not entirely recapitulate the expression pattern of the primary tissue of interest. Hence, it was reasonable to choose the SH-SY5Y, neuroblastoma, cell line as a model.
Patients from different families with the identical mutation indicate a similar gene expression pattern. This can be due to either the identical mutation in *ZNF526* or similar haplotype of the region. An important observation is that the expression patterns of patient lymphoblastoid cells showed specific changes, which could be recapitulated in ZNF526-deficient neuroblastoma cells. Therefore these results support the biological and clinical relevance of genes within the clusters identified in our study. The down regulation of cellular genes by both mutant proteins is also of great importance. This impairment of ZNF526 function was suggested by *in silico* protein modelling.
The expression profiles presented here identified the evidence for the involvement of functionally impaired ZNF526 in protein synthesis in non-syndromic intellectual disability. The substantial evidence for the role of this gene in mitochondria, oxidative phosphorylation and NADH dehydrogenase, and also the comprehensive annotation of the genes involved in several neuronal disorders like huntington and alzheimer pathways, add significantly to the understanding of the possible role of this gene in NSID. The current study provides a comprehensive, global view of the gene expression in ZNF526 patients and provides insights into the pathogenesis of *ZNF526* mutations.
A complete picture of the regulatory functions of ZNF526 was possible from the data derived from ChIP-Seq.
This allowed us to confirm binding of ZNF526 to DNA *in vivo*. Moreover, we found the vast majority of binding sites to be located near TSSs. However, the genome-wide distribution of binding sites indicates that studies aiming to identify the effect of this transcription factor on specific genes should not be limited to promoter regions.
ZNF526 specific target genes are involved in fundamental biological processes and are significantly enriched for ribosome biogenesis, rRNA processing, DNA binding and transcriptional regulatory activity. Comparison of the expression profiling data together with the ChIP-Seq data, revealed that the unique nature of clustered genes significantly enriched in protein synthesis, mitochondrial dysfunction, energy metabolism and gene regulation.
It is known that the metabolic rate in the brain is 7.5 times higher than the average rate in the rest of the body [Guyton 1993; Whitehead et al., 2005]. High metabolic demand in the brain supports pumping of ions across neuronal membranes during action potentials, and metabolism is primarily oxidative. Mitochondria are the principal sites for oxidative phosphorylation, and are most numerous in the heart, brain and skeletal muscle cells [Whitehead et

al., 2005]. Interestingly, previous studies have suggested higher mitochondrial translation activity in the brain versus other tissues [Kimberly et al., 2006]. Because the brain has the highest energy demand of any tissue, mitochondrial diseases cause a variety of neurological problems, including intellectual disability, seizures, developmental regression, gastrointestinal problems, and lack of coordination. Additionally, biosynthesis of proteins is essential for growth and continued maintenance of the entire neuron, including axons, dendrites, and synaptic terminals, and it is clearly one of the important biochemical processes underlying adaptive changes in the nervous system. Therefore, disruption in either energy metabolism or protein synthesis can lead to insufficient brain function.

Intuitively, it would be attractive to propose that complex or metabolically active tissues, such as the brain, utilize more TFs than simple tissues [Vaquerizas et al., 2009]. Moreover, humans and primate differences correlate with metabolic changes, as evidenced by the relative upregulation of energy-related genes and metabolites in the human brain. While the mechanisms underlying these evolutionary changes have not been elucidated, altered activities of key transcription factors (TFs) could play a pivotal role [Nowick et al., 2009].

It makes *ZNF526* a fascinating gene for further investigation on intellectual impairment, because we showed ZNF526 protein interacts with PRKRIR, a repressor of the inhibitor of protein kinase or Dap4 (Death-associated protein 4). The initial picture we have of the experimentally determined ZNF526/PRKRIR interactome has been drawn with mass spectrometry in the hela cell environment.

PRKIR is a transcription factor, and recently it has been shown to belong to the 3,593 human genes which have been added to the human TF repertoire by duplication during the last 35-40 million years of primate history [Nowick and Stubbs 2010]. Given the fact that the TFs within human segmental duplications are likely to hold important keys to human evolutionary history, the impact of the PRKRIR and ZNF526 as regulatory targets of primate proteins likely provides a major resource for primate-specific functions.

PRKRIR is the upstream regulator of interferon-induced serine/threonine protein kinase r (pkr) and may block the pkr- inhibitory function of p58ipk, resulting in restoration of kinase activity and suppression of cell growth and apoptosis. Putative protein interactors have been described (p58IPK, DNAJC3, PKR, E2F6, EIF2S1, NLRC5, STK4, TP53, MST1). It is of note that the interacting proteins, PKR, TP53 and EIFs, are implicated in neurodegenerative disorders and intellectual disbility [Peel et al., 2001; Saccucci et l., 2007; Froyen et al., 2008].

ZNF genes comprise a large family of transcription factors, that underwent rapid evolution in primates and humans [Shen et al., 2010; Emerson et al., 2009]. It has been postulated that they may have evolved to play an important role in the function of complex systems such as nervous or immune systems [Collins et al., 2001]. Genes responsible for human-specific phenotypes may have been under altered selective pressures in human evolution and thus exhibit changes in substitution rate and pattern at the protein sequence level. The C2H2 Zinc-finger family grew at several evolutionary stages, especially with the appearance of vertebrates, and most substantially during the emergence of mammals and primates [Juan et al., 2009].

It is observed that transcriptional misregulation has been associated with a diverse set of diseases, including neuronal malformations and developmental syndromes [Engelkamp et al., 1998; Jimenez-Sanchez et al., 2001]. A third of human developmental disorders have been at-

tributed to dysfunctional TFs [Boyadjiev et al., 2001]. Briefly, to evaluate the overall impact of TFs in human diseases, 164 TFs have been identified so far to be directly responsible for 277 diseases or syndromes [Juan M t al., 2009]. Among these, a significant proportion is related to developmental defects, highlighting the importance of TFs during the early stages of development [Juan et al., 2009; Boyadjiev et al., 2000]. The misregulation of *TF* genes themselves also has important implications for more complex systems. A number of zinc finger proteins, autosomal and X-linked, have been found in NS-ID. These genes, harbouring both missense and nonsense mutations, cause NS-ID. Examples are ZNF41, ZNF81 and ZNF674 zinc finger proteins that belong to the Kruppel-associated box (KRAB) family. The KRAB/C2H2 zinc finger proteins make up approximately one third of the various zinc finger proteins found in the human genome. Many of them function as transcriptional repressors [Sander et al, 2003; Looman e al., 2002].
The genes *ZNF711* and *ZC3H14* [Pak et al., 2011] are more examples of zinc finger genes implicated in the development of NS-ID. Since these genes are involved in NS-ID, it is possible that their protein products target the regulation of specific neuronal genes that are involved in cognitive development, learning or memory formation, resulting in an NS-ID phenotype.
Descriptions of discrete expression patterns of transcription factors during nervous system development as well as analyses of mutant phenotypes reveal the regulatory role of region and cell-type specific transcription factors in morphogenesis and differentiation of the vertebrate nervous system [reviewed by Bang and Goulding, 1996]. Therefore, our observation of *ZNF526* expression in the fetal brain suggests that ZNF526 might be involved in brain development, and is therefore an excellent candidate for further investigation.
ZNF526 can be one of the ideal candidate genes for cognitive investigation because I) ZNF526 defects belong to the more common causes of NS-ARID, at least in the Iranian population, II) ZNF526 belongs to ZNF C2H2 proteins which have recently evolved in higher vertebrates, and III) ZNF526 intracts with PRKRIR (recently evolved gene), which more likely has an impact on primate-specific functions. To understand human cognition and the intellect, the identification of the molecular and biological causes of NS-ID is essential [Berg et al., 1998; Kaufman et al., 2010]. Since the only feature of NS-ID is intellectual impairment, genes that, when mutated, cause NS-ID are likely related to the processes of learning and memory. Understanding the molecular pathways in which these genes are involved will be important to reveal the processes of formation and evolution of normal intellectual capabilities. Additionally, understanding the genetics of a complex disease like NS-ID may also aid us in treating or relieving symptoms of NS-ID in certain cases and in genetic counselling of families with affected individuals, particularly where consanguinity is involved [Kaufman et al., 2010; Modell and Darr 2002].

7 Outlook

Given the high heterogeneity of the NS-ID phenotypes, our previous studies reveal that the elucidation of molecular causes of ARID should be speeded up, otherwise many years will be required to find these several hundred underlying genes (assuming thousands). Identification of most or all of these genes is a prerequisite for early diagnosis, prevention and, eventually, therapy of ID. Therefore we are now trying to combine homozygosity mapping, targeted exon enrichment and next-generation of sequencing systems for this purpose. It should be entirely feasible to develop a diagnostic test for mutations in all known ID genes. This will be particularly important for non-syndromic forms of ID, which cannot be distinguished clinically.
Hopefully, by finding more and more genes with the help of these powerful and fast methods, it will be possible to bring an end to many genetic diseases by performing universal carrier screening, combined with preimplantation genetic diagnosis for carrier couples who want biological children.
The identification of the underlying genes for all six hot spot ARID loci will be interesting and may indicate that common molecular causes of NS-ARID do exist, and also the most frequent ones may well account for a significant percentage of the patients. These findings will be instrumental in the identification of the underlying genes.
Furthermore, the functional characterization of the novel genes will aid our understanding of the molecular pathways and processes involved in neurodevelopment and cognition. These studies will shed more light on the pathogenetic mechanisms underlying disorders of brain development and function.
Since the only feature of NS-ID is intellectual impairment, elucidation of the molecular pathways in which these genes are involved will be important for understanding the biochemical and regulatory substrates of human intelligence as well as its evolution.
Moreover, obtaining knowledge about these genes and understanding their involved pathways may aid us in treating or relieving symptoms of NS-ID in certain cases. ID remains the most common medical condition for which no specific drug treatments are available. With the recent report of successful treatment of dfmr1 mutants, there is an extraordinary opportunity to use the Drosophila system as a drug-screening tool for ID. Also the achievement of generating induced pluripotent stem cells has recently opened new doors of hope to look for the treatments.

8 References

Akira S, et al. (2000). Effects of anticodon 29-O-methylations on tRNA codon recognition in an Escherichia coli cell-free translation. RNA; 6: 680-686.

Algeri S, et al. (1979). Changes in rat brain noradrenaline and serotonin after administration of S-adenosylmethionine. In: Zappia V, Usdin E, Salvatore S, eds. Biochemical and pharmacological roles of adenosylmethionine and the central nervous system. New York: Pergamon Press: 81-7.

Allen, K.M, et al. (1998). PAK3 mutation in nonsyndromic X-linked mental retardation. Nat. Genet; 20: 25-30.

Amos-Landgraf JM, et al. (2006). X chromosomeinactivation patterns of 1,005 phenotypically unaffected females. Am J Hum Genet; 79: 493-499.

Ariani F, et al. (2008). FOXG1 is responsible for the congenital variant of Rett syndrome. Am J Hum Genet; 83(1): 89-93.

Armengaud, J., et al. (2004). N2-methylation of guanosine at position 10 in tRNA is catalyzed by a THUMP domain-containing, S-adenosylmethionine-dependent methyltransferase, conserved in Archaea and Eukaryota. J. Biol. Chem; 279: 37142?37152.

Auffinger P, et al. (1998). Location and distribution of modified nucleotides in tRNA In: Grosjean H, Benne R, eds The modification and editing of RNA. New York: ASM: Press 569-577.

Barhamachari and Ramakrishnan. (1984). Modified bases in transfer RNA. J. Biosci; 6(5): 757-770.

Barrera, L. O. et al. (2008). Genome-wide mapping and analysis of active promoters in mouse embryonic stem cells and adult organs. Genome Res; 18: 46-59

Barski A, et al. (2007). High-resolution profiling of histone methylations in the human genome. Cell; 129: 823-837.

Bartosz Brzezicha, et al. (2006). Identification of human tRNA:m5C methyltransferase catalysing intron-dependentm5C formation in the first position of the anticodon of the pre-tRNA(CAA). Nucleic Acids Research; 34(20): 6034-6043

Basel-Vanagaite L, et al. (2006). The CC2D1A, a member of a new gene family with C2 domains, is involved in autosomal recessive non-syndromic mental retardation. J Med Genet; 43: 203-10.

Beck CDO, et al. (2000). Learning performance of Drosophila after repeated conditioning trials with discrete stimuli. J. Neurosci; 20: 2944-53

Berkel S, et al. (2010). Mutations in the SHANK2synaptic scaffolding gene in Autism Spectrum Disorder and mental retardation. Nat Genet.

Bertrand, E., et al. (1998). Nucleolar localization of early tRNA processing.Genes Dev; 12: 2463-2468.

Billuart, P. et al. (1998). Oligophrenin 1 encodes a rho-GAP protein involved in X-linked mental retardation. Pathol. Biol; 46: 678

Blow, M. J. et al. (2008). Identification of the source of ancient remains through genomic sequencing. Genome Res. 18: 1347-1353

Bonini NM, et al. (2003). Human neurodegenerative disease modeling using Drosophila.Annu Rev Neurosci; 26: 627-656.

Bottiglieri T. (1996). Folate, vitamin B12, and neuropsychiatric disorders. Nutr Rev; 54: 382-90.

Bottiglieri T. (2002). S-Adenosyl-L-methionine (SAMe): from the bench to the bedside?molecular basis of a pleiotrophic molecule1 Am J Clin Nutr;76(suppl): 1151S-7S

Boycott KM, et al. (2009). Mutations in VLDLR as a cause for autosomal recessive cerebellar ataxia with mental retardation (dysequilibrium syndrome). J Child Neurol; 24(10): 1310-5.

Boycott KM, et al. (2005). Homozygous deletion of the very low-density lipoprotein receptor gene causes autosomal recessive cerebellar hypoplasia with cerebral gyral simplification. Am J Hum Genet; 77: 477-483.

Bradley JD, et al. (1994). A randomized, double blind, placebo controlled trial of intravenous loading with S-adenosylmethionine (SAM) followed by oral SAM therapy in patients with knee osteoarthritis. J Rheumatol; 21: 905-11.

Brand AH, et al. (1993). Development; 118: 401-415.

Bressa GM. (1994). S-Adenosyl-L-methionine (SAMe) as antidepressant: meta-analysis of clinical studies. Acta Neurol Scand; 154: 7-14.

Brunger, A.T., et al.(1998). Crystallography and NMR system: a new software suite for macromolecular structure determination. Acta Crystallogr. D Biol. Crystallogr; 54: 905-921.

Bujnicki,J.M., et al. (2004). Sequence-structure-function studies of tRNA:m5C methyltransferase Trm4p and its relationship to DNA:m5C and RNA:m5U methyltransferases. Nucleic Acids Res; 32: 2453-2463.

Carmena, M., and Earnshaw, W. C. (2003). The cellular geography of aurora kinases. Nat. Rev. Mol. Cell Biol; 4: 842-854.

Chace DH., et al. (2003). Mass spectrometry-based diagnostics: the upcoming revolution in disease detection has already arrived. Clin Chem; 49 (7): 1227-8.

Chechlacz M, Gleeson JG. (2003). Is mental retardation a defect of synapse structure and function? Pediatr Neurol; 29: 11-17.

Chelly J, et al. (2006). Genetics and pathophysiology of mental retardation. Eur J Hum Genet; 14(6): 701-13.

Chelly J, et al. (2001) Monogenic causes of X-linked mental retardation. Nat Rev Genet; 2: 669-680

Chiurazzi P, et al. (2008). XLMR genes: update 2007. Eur J Hum Genet; 16: 422-434.

Christensen J, et al. (2007). RBP2 belongs to a family of demethylases, specific for tri-and dimethylated lysine 4 on histone 3. Cell; 128: 1063-76.

Cloos PA, et al. (2008). Erasing the methyl mark: histone demethylases at the center of cellular differentiation and disease. Genes Dev.; 22: 1115-40.

Costello, J. F., and Plass, C. (2001). Methylation matters. J Med Genet; 38: 285-303.

Crellin R, et al. (1993). Folates and psychiatric disorders. Clinical potential. Drugs; 45: 623-36.

D'Arcangelo G, et al. (1995). A protein related to extracellular matrix proteins deleted in the mouse mutant reeler. Nature; 374: 719-723.

Dawn Teare M, Barrett JH. (2005). Genetic linkage studies. Lancet; 366(9490): 1036-44.

De Beus, E., et al. (1994). Yeast NOP2 encodes an essential nucleolar protein with homology to a human proliferation marker. J. Cell Biol; 127: 1799-1813.

De La Fuente, R., et al. (2004). ATRX, a member of the SNF2 family of helicase/ATPases, is required for chromosome alignment and meiotic spindle organization in metaphase II stage mouse oocytes. Dev.Biol; 272: 1-14.

Doll,A. et al. (2001). Characterization of two novel genes, WBSCR20 and WBSCR22, deleted in Williams-Beuren syndrome. Cytogenet. Cell Genet; 95: 20-27.

Erdogan F. et al (2006). Impact of low copy repeats on the generation of balanced and unbalanced chromosomal aberrations in mental retardation. Cytogenet Genome Res; 115: 247?253

Faber ES, et al. (2003). Calcium-activated potassium channels: multiple contributions to neuronal function. Neuroscientist; 9(3): 181-94.

Field M, et al. (2007). Mutations in the BRWD3 gene cause X-linked mental retardation associated with macrocephaly. Am J Hum Genet; 81: 367-74.

Fisher RA (1935a). The detection of linkage with dominant abnormalities. Ann Eugen; 6: 187-201.

Fisher RA (1935b). The detection of linkage with recessive abnormalities. Ann Eugen; 6: 339-351.

Frumkin. (2003). Mental Retardation, A Primer to Cope with Expert Testimony National Legal Aid and Defender Association Cornerstone.

Frye, M., et al. (2006). The RNA methyltransferase Misu (NSUN2) mediates Myc-induced proliferation and is upregulated in tumors. Curr. Biol; 16: 971-981.

Froyen G., 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(2): 432-43

Garshasbi M, et al. (2008). A defect in the TUSC3 gene is associated with autosomal recessive mental retardation. Am J Hum Genet; 82: 1158-64.

Gassmann, R., et al. (2004). Borealin: a novel chromosomal passenger required for stability of the bipolar mitotic spindle. J. Cell Biol.; 166: 179-191.

Genoux D, et al. (2002). Protein phosphatase 1 is a molecular constraint on learning and memory. Nature; 418 (6901): 970-5.

Gerner EW, et al. (2004). Polyamines and cancer: old molecules, new understanding. Nat Rev Cancer; 4: 781-792.

Gottfredson, Linda S. (1997). Mainstream science on intelligence: An editorial with 52 signatories, history, and bibliography. Intelligence; 24(1): 13-23

Greenspan RJ. (1995). Flies, genes, learning, and memory. Neuron; 15: 747-750.

Gregg, R. G., et al. (2003). Positional cloning of the young mutation identifies an essential role for the Brahma chromatin remodeling complex in mediating retinal cell differentiation. Proc Natl Acad Sci USA; 100: 6535-40.

Gunderson KL, et al. (2005) A genome-wide scalable SNP genotyping assay using microarray technology. Nat Genet; 37(5): 549-54.

Guyton AC. (1991). Textbook of Medical Physiology 8th edition. Philadelphia: W.B. Saunders Company.

Hagerman RJ et al. (2002). The physical and behavioural phenotype in: Hagerman RJ. Fragile X syndrome: Diagnosis, Treatment and Research. Johns Hopkins University Press.

Hammond RS, et al.(2006). Small-conductance Ca2+-activated K+ channel type 2 (SK2) modulates hippocampal learning, memory, and synaptic plasticity. J Neurosci; 26(6): 1844-53.

Harris T, et al. (2007). Sperm Motility Defects and Infertility in Male Mice with a Mutation in Nsun7, a Member of the Sun Domain-Containing Family of Putative RNA Methyltransferases Biology of reproduction; 77: 376-382.

Hendrich B, Bird A. (1998). Identification and characterization of a family of mammalian methyl-CpG binding proteins. Mol Cell Biol; 18: 6538-6547.

Hiesberger T, et al. (1999). Direct binding of Reelin to VLDL receptor and ApoE receptor 2 induces tyrosine phosphorylation of disabled-1 and modulates tau phosphorylation. Neuron; 24(2): 481-489.

Higgins JJ, et al. (2004). A mutation in a novel ATP-dependent lon protease gene in a kindred with mild mental retardation. eurology.; 63: 1927-31.

Hirata F, et al. (1978). Identification and properties of two methyltransferases in conversion of phosphatidylethanolamine to phosphatidylcholine. Proc Natl Acad Sci USA; 75: 1718-21.

Hobbs HH, et al. (1992).Molecular genetics of the LDL receptor gene in familial hypercholesterolemia. Hum Mutat; 1(6): 445-466.

Hong SE, et al. (2000). Autosomal recessive lissencephaly with cerebellar hypoplasia is associated with human RELN mutations. Nat Genet; 26(1): 93-96.

Hong, C., et al. (2003). The contribution of genetic and epigenetic mechanisms to gene silencing in oligodendrogliomas. Cancer Res; 63: 7600-5.

Huang,N., et al. (2003). Proteinfacilitated base in DNA by cytosine-5 methyltransferase. Proc. Natl Acad. Sci. USA; 100: 68-73.

Humphrey N, et al. (2008). Genes and quadrupedal locomotion in humans. Proc Natl Acad Sci USA; 105: E26.

Inlow JK, Restifo LL. (2004). Molecular and comparative genetics of mental retardation. Genetics; 166: 835-881.

International HapMap Consortium (2003). The International HapMap project. Nature 426:689-796.

Ishkanian AS, et al. (2004). A tiling resolution DNA microarray with complete coverage of the human genome. Nat Genet; 36: 299-303.

Iuchi, et al. (2001). Three classes of C2H2 zinc finger proteins. CMLS, Cell. Mol. Life Sci; 58: 625-635.

Iwase S, et al. (2007). The X-linked mental retardation gene SMCX/JARID1C defines a family of histone H3 lysine 4 demethylases. Cell.; 128: 1077-88.

Ramser J, et al. (2004). splice site mutation in the methyltransferase gene FTSJ1 in Xp11.23 is associated with non-syndromic mental retardation in a large Belgian family (MRX9). J Med Genet; 41: 679-683.

Janusz M. et al. (2004). Sequence-structure function studies of tRNA:m5C methyltransferase Trm4p and its relationship to DNA:m5C and RNA:m5U methyltransferases Nucleic Acids Research; 32(8): 2453-2463

Johnson DS, et al. (2007). Genome-wide mapping of in vivo protein-DNA interactions. Science; 316: 1497-1502.

Johnson,P.F. and Abelson,J. (1983). The yeast tRNATyr gene intron is essential for correct modification of its tRNA product. Nature; 30: 681-687.

Jones PA, Baylin SB (2002). The fundamental role of epigenetic events in cancer. Nat Rev Genet 3: 415-428.

Jones PL, et al. (1998). Methylated DNA and MeCP2 recruit histone deacetylase to repress transcription. Nat Genet; 19: 187-191.

Juan M. Vaqueriza , et al. (2009). A census of human transcription factors: function, expression and evolution Luscombe APRIl; Vol 10.

Karim SA, et al. (2007). PLP overexpression perturbs myelin protein composition and myelination in a mouse model of Pelizaeus-Merzbacher disease. Glia; 55: 341-351.

Katz, J. E., et al. (2003). Automated identification of putative methyltransferases from genomic open reading frames. Mol. Cell Proteomics; 2: 525-540.

Kaufman L. and Ayub M. and Vincent John B. (2010). The genetic basis of non-syndromic intellectual disability: a review. J Neurodevelop Disord; 2: 182-209

Kim CA, et al. (1996). A 2.2 A°resolution crystal structure of a designed zinc finger protein bound to DNA. Nat Struct Biol; 3: 940-945

Kimberly A. et al. (2006). Tissue-Specific Differences in Human Transfer RNA Expression. PLoS Genetics; 2: 12

Kokura K, et al. (2001). The Ski protein family is required for MeCP2-mediated transcriptional repression. J Biol Chem; 276: 34115-34121.

Kruglyak L, et al. (1997). The use of a genetic map of biallelic markers in linkage studies. NatGenet; 17: 21-24.

Kruglyak L, et al. (1996). Parametric and nonparametric linkage analysis: a unified multipoint approach. Am J Hum Genet; 58: 1347-1363.

Kruglyak L, Nickerson DA (2001). Variation is the spice of life. Nat Genet; 27: 234-236.

Laemmli UK. (1970). Cleavage of structural proteins during the assembly of the head of bacteriophage T4. Nature; 227(5259): 680-5.

Leonard, H. and Wen, X. (2002). The epidemiology of mental retardation: challenges and opportunities in the new millennium. Ment Retard Dev Disabil Res Rev 8(3): 117-34.

Li, Z. et al. (2003). Aglobal transcriptional regulatory role for c-Myc in Burkitts lymphoma cells. Proc. Natl Acad. Sci. USA; 100: 8164-8169

Lisik MZ, et al. (2008). X-linked mental retardation. Med Sci Monit; 14: RA221-229.

Liu, H. T., and Yung, B. Y. (1999). In vivo interaction of nucleophosmin/B23 and protein C23 during cell cycle progression in HeLa cells. Cancer Lett; 144, 45-54.

Mandel JL, Biancalana V. (2004). Fragile X mental retardation syndrome: from pathogenesis to diagnostic issues. Growth Horm IGF Res 14 Suppl A: S158-65.

Mari F, et al. (2005). CDKL5 belongs to the same molecular pathway of MeCP2 and it is responsible for the early-onset seizure variant of Rett syndrome. Hum Mol Genet; 14: 1935-1946.

Marin-Husstege, et al. (2002). Histone deacetylase activity is necessary for oligodendrocyte lineage progression. J Neurosci; 22: 10333-45.

McBride SMJ, et al. (2005). Pharmacological rescue of synaptic plasticity, courtship behavior and mushroom body defects in a Drosophila model of Fragile X syndrome. Neuron; 45: 753-764.

Miller SA, et al. (1988). A simple salting out procedure for extracting DNA from human nucleated cells. Nucleic Acids Res; 16(3): 1215.

Moheb LA, et al. (2008). Identification of a nonsense mutation in the very low-density lipoprotein receptor gene (VLDLR) in an Iranian family with dysequilibrium syndrome. Eur J Hum Genet; 16: 270-273.

Molinari F, et al. (2002). Truncating neurotrypsin mutation in autosomal recessive nonsyndromic mental retardation. Science; 298: 1779-81.

Montini E, et al. (1998). Identification and characterization of a novel serine-threonine kinase gene from the Xp22 region. Genomics; 51: 427-433.

Morton NE and MacLean CJ. (1974). Analysis of family resemblance. 3. Complex segregation of quantitative traits. Am J Hum Genet; 26(4): 489-503

Motazacker MM, et al. (2007). A defect in the ionotropic glutamate receptor 6 gene (GRIK2) is associated with autosomal recessive mental retardation. Am J Hum Genet; 81: 792-798.

Motorin, Y., et al. (1999). Multisite-specific tRNA: m5C-methyltransferase (Trm4) in yeast Saccharomyces cerevisiae: identification of the gene and substrate specificity of the enzyme. RNA; 5: 1105-1118.

Munton RP, et al. (2004). The role of protein phosphatase-1 in the modulation of synaptic and structural plasticity. FEBS Lett; 567(1): 121-8.

MurasugiA, et al. (1978). Nucleotide sequence of leucine transfer RNA 1 from Candida (Torulopsis) utilis. J Biochem; 83: 1029-1038

Najmabadi H, et al. (2006). Homozygosity mapping in consanguineous families reveals extreme heterogeneity of non-syndromic autosomal recessive mental retardation and identifies 8 novel gene loci. Hum Genet.

Nowick K, et al. (2009). Differences in human and chimpanzee gene expression patterns define an evolving network of transcription factors in brain PNAS; 29 (106): 52.

Novick P, et al. (1997). The diversity of Rab proteins in vesicle transport. Curr Opin Cell Biol; 9: 496-504.

Nussbaum RL, et al. (1995). The fragile X syndrome in the metabolic and molecular bases of inherited disease. McGraw Hill Inc; 795-810.

Okuwaki, M., et al. (2002). The RNA binding activity of a ribosome biogenesis factor, nucleophosmin/B23, is modulated by phosphorylation with a cell cycle-dependent kinase and by association with its subtype. Mol. Biol. Cell; 13: 2016-2030.

Ou XM, et al. (2003). Freud-1: A neuronal calcium-regulated repressor of the 5-HT1A receptor gene. J Neurosci; 23(19): 7415-25.

Ozcelik T, et al. (2008). Mutations in the very low-density lipoprotein receptor VLDLR cause cerebellar hypoplasia and quadrupedal locomotion in humans. Proc Natl Acad Sci USA; 105(11): 4232-6.

Pak C, et al. (2011). Mutation of the conserved polyadenosine RNA binding protein, ZC3H14/dNab2, impairs neural function in Drosophila and humans. Proc Natl Acad Sci USA; 26; 108(30): 12390-5

Pawel Siedlecki1 and Piotr Zielenkiewicz. (2006). Mammalian DNA methyltransferases. Acta biochimica polonica; 53(2): 245-256

Pediatrics for Parents. (1991). Introduction to mental retardation. February 1.

Penrose LS. (1935). The detection of autosomal linkage in data which consist of pairs of brothers and sisters of unspecified parentage. Ann Eugen; 6: 133-138.

Penrose, L. (1938). A Clinical and Genetic Study of 1280 Cases of Mental Defect, HMSO, London.

Peter, M., et al. (1990). Identification of major nucleolar proteins as candidate mitotic substrates of cdc2 kinase. Cell; 60: 791-801.

Pinkel D, et al. (1998). High resolution analysis of DNA copy number variation using comparative genomic hybridization to microarrays. Nat Genet; 20: 207-211.

Pinto D, et al. (2010). Functional impact of global rare copy number variation in autism spectrum disorders. Nature. [Epub ahead of print].

Plenge RM, et al. (2002). Skewed X-chromosome inactivation is a common feature of X-linked mental retardation disorders. Am J Hum Genet; 71: 168-173.

Ramser J, et al. (2004). splice site mutation in the methyltransferase gene FTSJ1 in Xp 11.23 is associated with non-syndromic mental retardation in a large Belgian family (MRX9). J Med Genet; 41: 679-683.

Rassoulzadegan,M., et al. (2006). RNA-mediated non-mendelian inheritance of an epigenetic change in the mouse. Nature; 441: 469-474.

Rauch, A., et al. (2006). Diagnostic yield of various genetic approaches in patients with unexplained developmental delay or mental retardation. Am J Med Genet; 140(19): 2063-74.

Restifo L. (2005). Mental retardation genes in drosophila: new approaches to understanding and treating developmental brain disorders. Mental retardation and developmental disabilities research reviews; 11: 286-294.

Rice DS, Curran T. (2001). Role of the reelin signaling pathway in central nervous system development. Annu Rev Neurosci; 24: 1005-1039.

Robertson G, et al. (2007). Genome-wide profiles of STAT1 DNA association using chromatin immunoprecipitation and massively parallel sequencing. Nat Methods.

Robertson KD. (2005). DNA methylation and human disease. Nat Rev Genet; 6: 597-610.

Robertson, K. D., et al. (2000). DNMT1 forms a complex with Rb, E2F1 and HDAC1 and represses transcription from E2F-responsive promoters. Nat Genet; 25: 338-42.

Robertson, K. D., and Wolffe, A. P. (2000). DNA methylation in health and disease. Nat Rev Genet; 1: 11-9.

Roeleveld, N., et al. (1997). The prevalence of mental retardation: a critical review of recent literature. Dev Med Child Neurol; 39(2): 125-32.

Ropers HH, Hamel BC. (2005). X-linked mental retardation. Nat Rev Genet; 6(1): 46-57.

Ropers HH. (2008). Genetics of intellectual disability. Curr Opin Genet Dev.

Ropers, HH. (2006). X-linked mental retardation: many genes for a complex disorder. Curr Opin Genet Dev 16(3): 260-9.

Rotanova TV, et al. (2006). Slicing a protease: structural features of the ATP-dependent Lon proteases gleaned from investigations of isolated domains. Protein Sci 15(8): 1815-28.

Rozenski J, et al. (1999). The RNA Modification Database: 1999 update. Nucleic Acids Res; 27: 196-197

Rual JF, et al. (2005). Towards a proteome-scale map of the human protein-protein interaction network. Nature; 437 (7062): 1173-8.

Sakai J, et al. (1994). Structure, chromosome location, and expression of the human very lowdensity lipoprotein receptor gene. J Biol Chem; 269: 2173-2182.

Sambrook, J., et al. (1989). Molecular Cloning: A Laboratory Manual. 2nd ed. New York: Cold Spring Harbor Laboratory Press, Cold Spring Harbor.

Scala E, et al. (2005). CDKL5/STK9 is mutated in Rett syndrome variant with infantile spasms. J Med Genet; 42: 103-107.

Scot A. Wolfe, et al.(2001). Beyond the ?Recognition Code?: Structures of Two Cys2His2 Zinc Finger/TATA Box Complexes. Structure; 9: 717-723.

Shiho Sakita-Suto, et al. (2007). Aurora-B Regulates RNA Methyltransferase NSUN2 Molecular Biology of the Cell; 18: 1107-1117.

Sirri, V., et al. (2002). Cyclin-dependent kinases govern formation and maintenance of the nucleolus. J. Cell Biol; 156: 969-981.

Sprinzl M, et al. (1998). Compilation of tRNA sequences and sequences of tRNA genes+ Nucleic Acids Res; 26: 148-153

Sprovieri T, et al. (2009). A novel mutation in the Xlinked cyclin-dependent kinase-like 5 (CDKL5) gene associated with a severe Rett phenotype. Am J Med Genet A; 149A: 722-725.

Stankiewicz P, et al. (2007). Use of array CGH in the evaluation of dysmorphology, malformations, developmental delay, and idiopathic mental retardation. Curr Opin Genet Dev; 17: 182-192.

Stoykova, A. S., et al. (1985). Ribosome biogenesis and nucleolar ultrastructure in neuronal and oligodendroglial rat brain cells. J. Neurochem; 45; 1667-1676.

Strachan, T. and Read, A. P. (2004). Human Molecular Genetics, Chapter 13., Garland Science.

Strickland DK, et al. (1995). LDL receptor-related protein: a multiligand receptor for lipoprotein and proteinase catabolism. FASEB J; 9: 890-898.

Sutcliffe JS, et al. (1992). DNA methylation represses FMR-1 transcription in fragile X syndrome. Hum Mol Genet; 1: 397-400.

Syvanen AC . (2006). Toward genome-wide SNP genotyping. Nat Genet; 37 Suppl: S5-10.

Tabor HK, et al. (2002). Opinion: Candidate-gene approaches for studying complex genetic traits: practical considerations. Nat Rev Genet; 3(5): 391-7.

Tahiliani M, et al. (2007). The histone H3K4 demethylase SMCX links REST target genes to X-linked mental retardation. Nature.;447:601-5.

Takahashi S, et al. (1995). Enhancement of the binding of triglyceride-rich lipoproteins to the very low-density lipoprotein receptor by apolipoprotein E and lipoprotein lipase. Biol Chem; 270: 15747-15754.

Thompson, M., et al. (2003). Nucleolar clustering of dispersed tRNA genes. Science; 302: 1399-1401.

Tissir F, Goffinet AM. (2003). Reelin and brain development. Nat Rev Neurosci; 4: 496-505.

Tscherne,J.S., et al. (1999) Puri cation, cloning and characterization of the 16S RNA m5C967 methyltransferase from Escherichia coli. Biochemistry; 38: 1884-1892.

Tully T, Quinn WG. (1985). Classical conditioning and retention in normal and mutant Drosophila melanogaster. J. Comp. Physiol; 157: 263-77

Turkmen S, et al. (2008). Cerebellar hypoplasia, with quadrupedal locomotion, caused by mutations in the very low-density lipoprotein receptor gene. Eur J Hum Genet; 16(9): 1070-4.

van Rossum D, Hanisch UK. (1999). Cytoskeletal dynamics in dendritic spines: direct modulation by glutamate receptors? Trends Neurosci; 22: 290-295.

Verkerk A, 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.

Webb JC, et al. (1994). Characterization and tissue-specific expression of the human very low density lipoprotein (VLDL) receptor mRNA. Hum Mol Genet; 3: 531-537.

Wilcken B., et al. (2004). Screening of newborns for metabolic disorders with mass spectrometry. Jama; 291 (12): 1444.

Whitehead and Douglas L Crawford (2005). Variation in tissue-specific gene expression among natural populations Genome Biology; 6: R13.

Wu, P., et al. (1998). NCL1, a novel gene for a non-essential nuclear protein in Saccharomyces cerevisiae. Gene; 220: 109-117.

Tabor HK, et al. (2002). Opinion: Candidate-gene approaches for studying complex genetic traits: practical considerations. Nat Rev Genet; 3(5): 391-7.

Yannick Blanchard, et al. (2007). The promoter of the rat 5- eductase type 1 gene is bidirectional and Sp1-dependent Molecular and Cellular Endocrinology; 264: 171-183.

Yuri Motorin, et al. (2010). 5-methylcytosine in RNA: detection, enzymatic formation and biological functions. Nucleic Acids Research; 38 (5): 1415-1430.

Zhang,Y., et al. (2008). Model-based analysis of ChIP-Seq (MACS). Genome Biol; 9: R137.

Zhao, X., et al. (2003). Mice lacking methyl-CpG binding protein 1 have deficits in adult neurogenesis and hippocampal function. Proc Natl Acad Sci USA; 100: 6777-82.

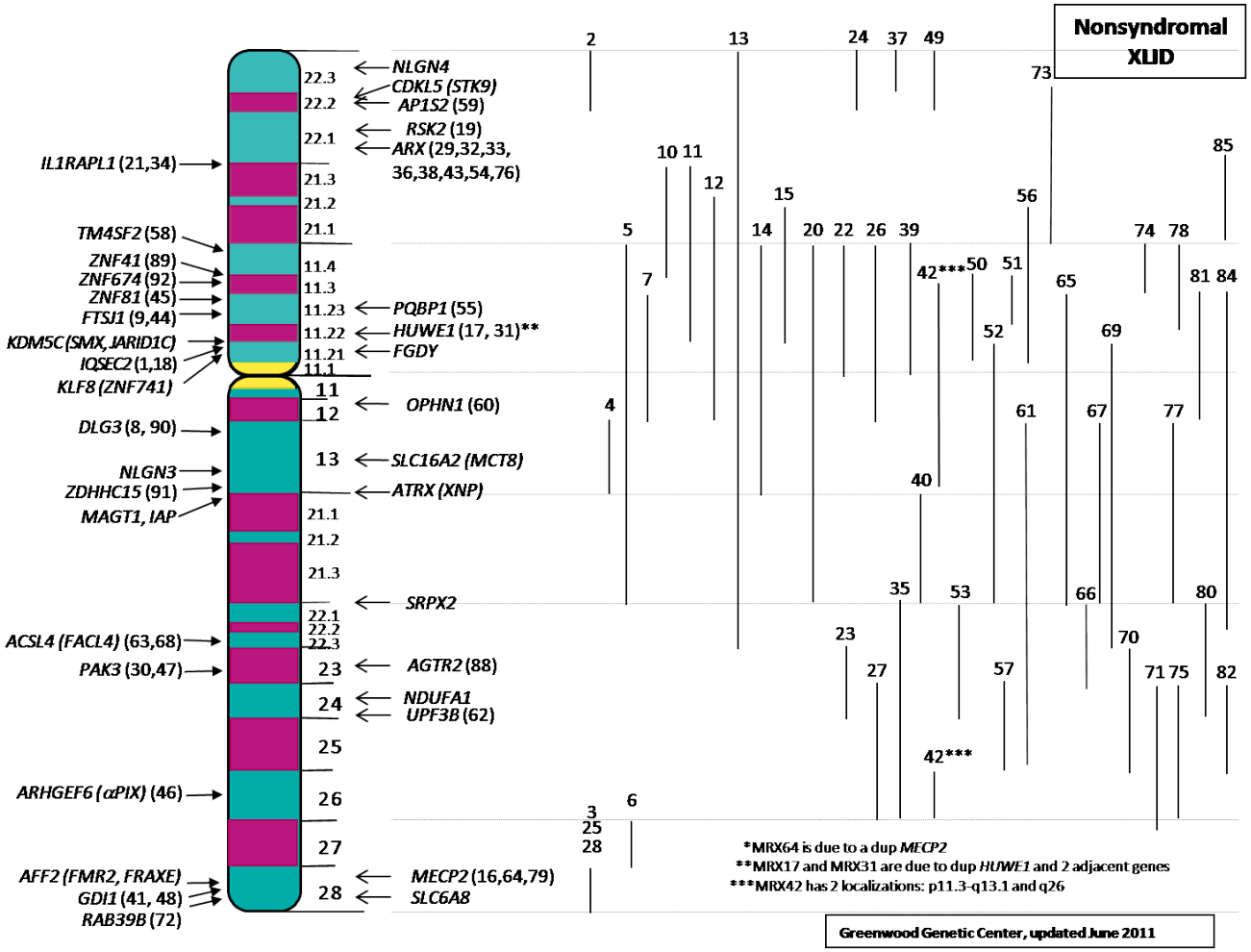
Zhou, H., et al. (2009). Mammalian MagT1 and TUSC3 are required for cellular magnesium uptake and vertebrate embryonic development. Proc. Nat. Acad. Sci; 106: 15750-15755.

Zschocke, et al. (2002). Estrogen receptor alpha-mediated silencing of caveolin gene expression in neuronal cells. J Biol Chem; 277: 38772-80.

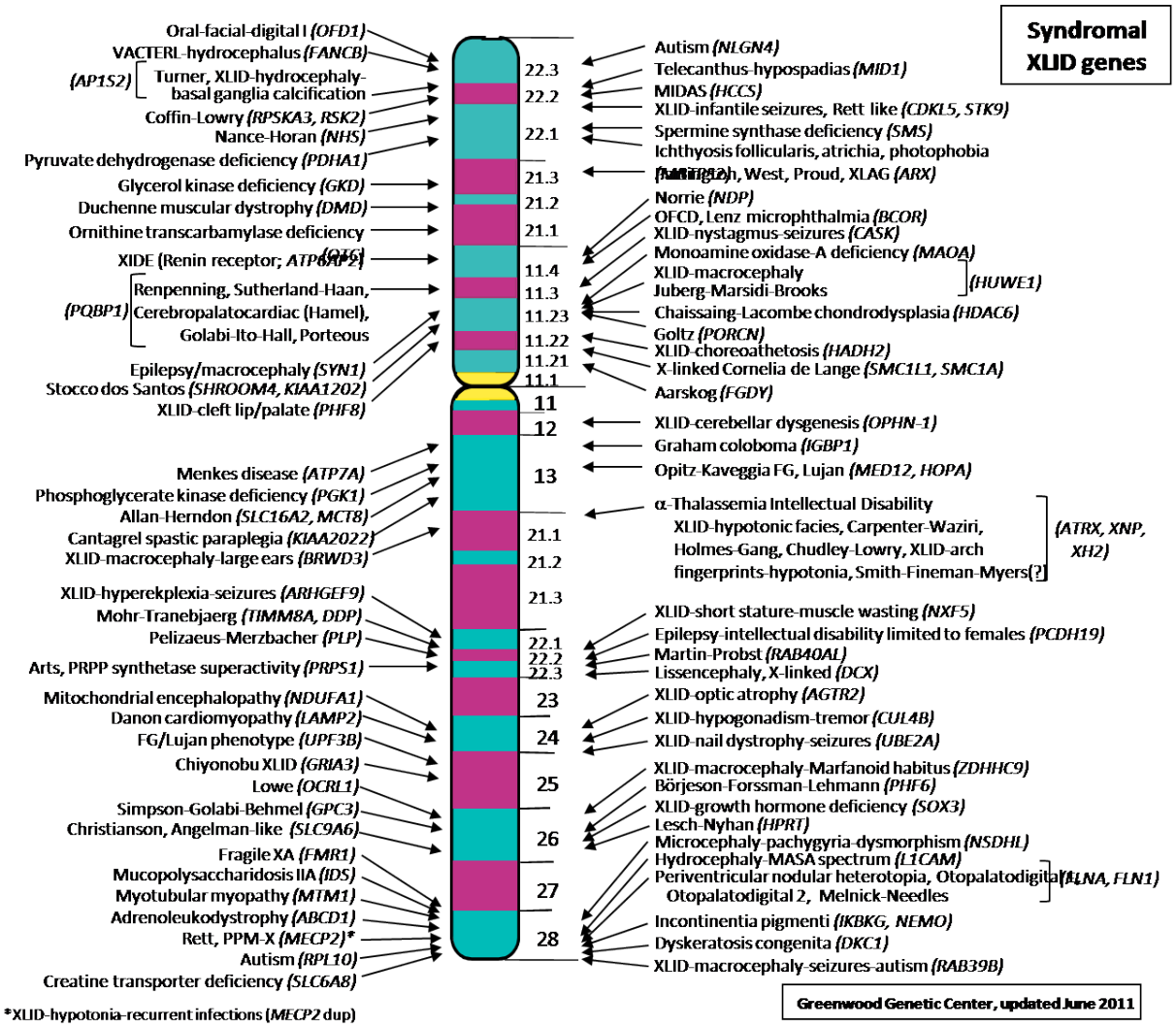
9 Supplementary data

9.1 Appendix A

Schematic view of chromosome X; Genes underlying of Non syndromic X-linked ID have been implicated.



Schematic view of chromosome X; Genes underlying of syndromic X-linked ID have been implicated.



9.2 Appendix B

List of the primers for amplification of NSUN2 gene	
Oligo Name	**Seq**
NSUN2_e3_F	ACTCTCCATGCAAAAGTTGG
NSUN2_e4_F	GCACAAAGTTGTCAACAGACG
NSUN2_e5_F	TGTTGCCTTTGAAACAATGC
NSUN2_e6_F	ATTCTGTGATTTTGGGGCAG
NSUN2_e7_F	TGGAGGTCGCATATTTCAGG
NSUN2_e8_F	TGCAGTTGGTGAAAATGGAG
NSUN2_e9_F	TGTGACATAAGGAGGTGCTTC
NSUN2_e10_F	GGGTTTGCCTTCCCTGG
NSUN2_e11_F	ATGGATACTCCTGCCCTGTG
NSUN2_e12_F	TTCTCTTGTATCATCTTTGCGG
NSUN2_e13_F	GATGGATTTTGTGAACGTGG
NSUN2_e14_F	GAAAGGAAGAATGTGCTTTTGG
NSUN2_e15_F	TGGAGGAGGAGAATAAGATGTTTAG
NSUN2_e16_F	GGCATCCTTACAGGAGATGG
NSUN2_e17_F	AGCATAGCCCCTGAGAGCAG
NSUN2_e18_F	TTCCAGATCGTCCCTACTGC
NSUN2_e19_F	CAGTTGCATATTACATCTCATTCG
NSUN2_e1_2_R	CGCCTTTAACCTCCAGCATC
NSUN2_e3_R	TGTTTGCAAAGTATACCCAAGC
NSUN2_e4_R	TGTCCACCTACTGCTTCTAGTG
NSUN2_e5_R	ACAAACAGGCATGACACTGG
NSUN2_e6_R	ACGGTGGTAGGCAAGATGTC
NSUN2_e7_R	GGCTCACGCTATCTTTCTCC
NSUN2_e8_R	AAGATCGGATGCTCACTTGC
NSUN2_e9_R	CCATACACATATGAAAAGCCC
NSUN2_e10_R	TCCGAAATTGATTCTAATTGGC
NSUN2_e11_R	ACCAGGGATGGGGCTTAAC
NSUN2_e12_R	GCAAGAAGCTCCTCTTTCCTC
NSUN2_e13_R	ACCAAGAAGTGGCTCTGGG

List of NSUN2 promoter primers	
NSUN2- promoter	**Sequence**
NSUN2 pr1_1F	GGATCTGGCAGGGAGAGG
NSUN2 pr1_1R	CTGCCGCTCTACCAGTACG
NSUN2 pr_2F	CCCGTAGTGGACGAGGAAC
NSUN2 pr_2R	CCGCCGCCCTATATGTTG
NSUN2 pr_3F	ACTCGCTGGCGTACTGGTAG
NSUN2 pr_3R	AGAACCCTTTCTGCAGAGTCC
NSUN2 pr_4F	GCCGTACACTGAGTTCGTCTG
NSUN2 pr_4R	CAGCCACAAAGGCGTCTC
NSUN2 pr_5F	AGGGTTCTCAAGTCAGGCTTC
NSUN2 pr_5R	ACAGGGAAAACAGCGAACAG
NSUN2 pr_6F	AGGGTTCTCAAGTCAGGCTTC
NSUN2 pr_6R	GGCCAGAACTCTAGCCCTAC
NSUN2 pr_7F	CTTTGGGCTTATTCCTGGAC
NSUN2 pr_7R	AGCCCTGAGGAAGGAAAGAG
NSUN2 pr_8F	CGGGAGGTGTAGGGCTAGAG
NSUN2 pr_8R	GCTCGAACAGCTTGTTCTCC
NSUN2 pr_9F	CGGGAGGTGTAGGGCTAGAG
NSUN2 pr_9R	CACTGCCTCCCTACCTTTTG

list of the primers employed for amplification of NSUN2 in Real time PCR and RT-PCR

List of NSUN2 primers for Real Time PCR	
real-Nsun2_F	TCCATTTCACCTGACACTGC
real-Nsun2_R	TATCTGGAGCCTGGGTATGC

List of primers for NSUN2 RT-PCR	
NSUN2-RT-PCR	Sequence
NSUN2-TR1-F	GTTGGTATCCTGAAGAACTTGC
NSUN2-TR1-R	CCAGGTGCTGCACACATATC
NSUN2-TR2-F	TTGAAGTTCCACAGCCACTG
NSUN2-TR2-R	CGCAATAACAAATCCCTCTTATG
NSUN2-TR3-F	CCCTGGGTTCAAATCCTAATC
NSUN2-TR3-R	GGACGAGATGAAAGCTGACTC
NSUN2-TR4-F	TGCTTTACTGTCCTCCAAACG
NSUN2-TR4-R	GACCATCTCCTGTAAGGATGC

9.3 Appendix C

List of the primers for amplification of VLDLR gene	
Oligo Name	Seq
VLDR_e1_F	GGAAAGGAGGGCCGAAg
VLDR_e2_F	TCCCCATCCATGGGTATTAG
VLDR_e3_F	ATGTGAAGCTAGGGCTGTGG
VLDR_e4_F	GAGCAGCAGCTTTGCATTG
VLDR_e5_F	AAAGATCAATGTATTAGATTTTGGGAC
VLDR_e6_7_F	GCATGGCATGTTCAGTTCTC
VLDR_e8_9_F	AATAGCCTGGGTTTTAAATGTG
VLDR_e10_F	GTGGTTTAGAAAGACCTTGCC
VLDR_e11_F	AAAAGTCCATTCTCCAAGCTC
VLDR_e12_F	GAATGCCTTGAGTTTTCTGC
VLDR_e13_F	GGGGAAAGAACCGTGAAAG
VLDR_e14_F	TGATGACCTTAGAAATGGACTTG
VLDR_e15_F	AAGGCTTTATATATACTAGGCACCG
VLDR_e16_F	ACAGCTAGCCATGCTGGAAC
VLDR_e17_F	ATTGGATGCAAAGGTTTTGG
VLDR_e18_F	TGGATTCCTGAACGTTATTACC
VLDR_e1_R	GTGGGCAAACGGAGACCTAC
VLDR_e2_R	TTTCAGAATCTCAATACATTTGCAG
VLDR_e3_R	CCTAAACTTTAAGGCAGGCAAAG
VLDR_e4_R	GATTAGTTTCCCCTGCCCAC
VLDR_e5_R	AAACACCATGGAATGAAGCC
VLDR_e6_7_R	GGGGTCCACACAGTGTTAGC
VLDR_e8_9_R	GCTGTACCTTTTAGAGGCACTTAC
VLDR_e10_R	CCCACAGATCACTTCCCAAG
VLDR_e11_R	TGAGAAAGCTCCTGACCTACAC
VLDR_e12_R	CATCTGCTAGTCCAGAATTCACC
VLDR_e13_R	GCCATGTTCAGCTGCTCTTC
VLDR_e14_R	TTCTGTGTCATGCTGCTTCC
VLDR_e15_R	AATCCAAGCTTGTGTCTTATTATATTG
VLDR_e16_R	GCTGTAGATACAAAGAGTGTGGC
VLDR_e17_R	CATGAAAGGGGTAGCTCCAG
VLDR_e18_R	CTCCTTTCTGAAAAGTCTCACTTC
VLDR_e19_1_R	TCTCTGGGATATTTACAGATGGC
VLDR_e19_2_R	GAACACTTGGTACTACTCAGGATAGG

9.4 Appendix D

list of the primers employed for amplification of ZNF526 in normal PCR and RT-PCR

List of the primers for ZNF526 amplification	
Oligo Name	Seq
ZNF526_e3_1_F	CCTTGGTTTCCTGCTGTCTC
ZNF526_e3_2_F	CAGCTCATCCTCTCCCCTG
ZNF526_e3_3_F	ATGGCAGAGGTCGGTGATG
ZNF526_e3_4_F	CACACTGCCAACCCATTG
ZNF526_e3_5_F	ATACGGGTGCACGTCCCTAC
ZNF526_e3_1_R	GAACAGCTCCTGGCAGTCC
ZNF526_e3_2_R	CCTGCTGAGGGCTGGTG
ZNF526_e3_3_R	AGCTGTTGCTCCTGTGGG
ZNF526_e3_4_R	CTGGCCGCAAGTGCAAC
ZNF526_e3_5_R	AGAGGTACCCACACTGGCTG

znf526-RT PCR primers			
Oligo Name	Seq	size in cDNA	size in Genomic DNA
5UTR-RT-ZNF526F	GGGGCTGCAGGAGATAAGAG	243 bp	no amplification
5UTR-RT-ZNF526R	CTTCTGTTGACATCTCCATCATC		
UTR2-e3-1-ZNF526F	GGATCACGGAAGACTGAAGC	479 bp	560 bp
UTR2-e3-1-ZNF526R	TTGCAGACTCTCGGAGGTG		
UTR2-e3-2RT-ZNF526F	CAGTGGATTAAGACTTCCTGAGC	565 bp	2041 bp
UTR2-e3-2RT-ZNF526R	ATGTCTCAATGATGGCGATG		

9.5 Appendix E

List of all the genes with diff. scores less than -30 (corresponding to P-values less than 0.001) for LCLs of 3 patients with 5 different ZNF526 mutations as one group in comparison with a group of 3 controls.

TargetID	Signal_X	Signal_Y	Detection_X	Detection_Y	Diff_Score	Symbol	Accession
ILMN_18416	1490,9	215,4	1	1	-224,151	VWCE	NM_152718.1
ILMN_6957	2824,3	431,8	1	1	-217,3689	PIP5K2A	NM_005028.3
ILMN_34340	7926,2	1637,8	1	1	-199,5105	LOC440359	XM_496143.2
ILMN_29238	3654,1	881,5	1	1	-178,803	WDR40A	NM_015397.1
ILMN_30165	1354	220	1	1	-173,5919	C16orf35	NM_012075.1
ILMN_4949	6191,7	1791,7	1	1	-156,413	ADIPOR1	NM_015999.2
ILMN_26284	462,2	92,6	1	0,9993	-152,1569	SLC38A5	NM_033518.1
ILMN_3232	616,7	93,3	1	1	-146,046	TRIM58	NM_015431.2
ILMN_14908	2204	443,6	1	1	-145,4297	EPB42	NM_000119.1
ILMN_10771	4874	1347	1	1	-138,1157	UBXD1	NM_025241.1
ILMN_14240	1900,3	512,7	1	1	-135,4214	C1orf128	NM_020362.2
ILMN_25234	307,2	58,1	1	0,9985	-133,568	RAB3IL1	NM_013401.2
ILMN_24336	3776,8	1122,4	1	1	-124,522	ALS2CR2	NM_018571.4
ILMN_8032	803,1	129,7	1	0,9985	-123,9562	TSPAN5	NM_005723.2
ILMN_512	1452,9	404,6	1	1	-121,9523	ATP6V0C	NM_001694.2
ILMN_20041	5133,4	1686,6	1	1	-121,5419	MKRN1	NM_013446.2
ILMN_11426	1465,6	152,1	1	1	-121,3369	KRT1	NM_006121.2
ILMN_24468	361,2	94,4	1	0,9993	-121,317	FKBP8	NM_012181.3
ILMN_5882	290,8	68,6	1	0,9985	-120,749	LOC402055	NM_001013694.1
ILMN_36916	1337,9	405,1	1	1	-119,9071	LOC389599	XM_372002.3
ILMN_12148	2922,7	808,8	1	1	-119,6006	BCL2L1	NM_138578.1
ILMN_22520	599,3	195,4	1	1	-116,7123	IGF2BP2	NM_006548.4
ILMN_24019	7827,9	2181,3	1	1	-115,7911	ASCC2	NM_032204.3
ILMN_2178	1407,7	514,8	1	1	-114,8415	PSMF1	NM_178578.1
ILMN_27643	347,2	80,1	1	0,9993	-114,493	C18orf10	NM_015476.2
ILMN_32201	2741,3	1024,1	1	1	-112,4895	LOC653778	XM_929667.1
ILMN_7706	2056,7	687,4	1	1	-111,8101	FAM46C	NM_017709.2
ILMN_5646	354,9	80,5	1	0,9993	-111,4676	IFIT1L	NM_001010987.1
ILMN_5039	4336,4	1458,9	1	1	-110,8074	GSPT1	NM_002094.1
ILMN_37027	5121,8	2008	1	1	-110,6849	LOC654103	XM_939368.1
ILMN_18191	10368,9	3547,2	1	1	-109,76	GUK1	NM_000858.4
ILMN_18317	2526,8	657,2	1	1	-107,8249	LYL1	NM_005583.3
ILMN_123803	253	63,2	0,9993	0,9978	-106,3649		AW205071
ILMN_18639	449,6	109,6	1	0,9993	-105,9308	MBNL3	NM_133486.1
ILMN_125852	471,8	160,6	1	1	-104,4261		AW977936
ILMN_25413	522	184,1	1	1	-102,9606	EIF2C2	NM_012154.2
ILMN_44953	204,5	27,8	0,9993	0,9978	-102,4863	LOC645416	XM_928457.1
ILMN_78819	189,1	38,2	0,9993	0,9978	-100,8264		DB336496
ILMN_34398	746,9	238,4	1	1	-100,6764	LOC642464	XM_926243.1
ILMN_1052	342,5	108,9	1	0,9993	-100,3431	TMOD1	NM_003275.1
ILMN_137328	1927	649,4	1	1	-99,1126	DNAJB2	NM_006736.4
ILMN_9039	18295,9	5573,3	1	1	-97,4765	CSDA	NM_003651.3
ILMN_14004	131,5	12,1	0,9978	0,9934	-95,8303	POLL	NM_013274.2
ILMN_20020	176,9	36,2	0,9985	0,9971	-94,6458	KLF1	NM_006563.2
ILMN_7718	399,1	142	1	1	-93,6207	PRSS36	NM_173502.2
ILMN_7921	1392,7	574,2	1	1	-92,7753	CRISPLD2	NM_031476.1
ILMN_71751	355,7	125,3	1	1	-90,6576		AI291290

TargetID	Signal_X	Signal_Y	Detection_X	Detection_Y	Diff_Score	Symbol	Accession
ILMN_86291	421,6	147	1	1	-89,1965		BX956905
ILMN_73255	559,3	141	1	1	-88,0913		BF434110
ILMN_23890	863,9	318,1	1	1	-87,3246	FAM104A	NM_032837.1
ILMN_87904	573,9	215,6	1	1	-87,0291		BU622030
ILMN_99206	562,7	201,2	1	1	-86,4535		BI462288
ILMN_14936	370	84,4	1	0,9985	-86,3021	FCGR3A	NM_000569.6
ILMN_112873	462,5	174,5	1	1	-85,4447		AI798610
ILMN_138633	696,7	282,3	1	1	-85,3427	ITGAM	NM_000632.2
ILMN_22401	4712,8	1712,2	1	1	-84,6089	HAGH	NM_005326.3
ILMN_107678	413,5	157,8	1	1	-83,2129		BM280411
ILMN_15785	14295,3	4378,6	1	1	-82,828	C19orf22	NM_138774.2
ILMN_23134	156,2	33,5	0,9985	0,9978	-82,3336	SPATA20	NM_022827.2
ILMN_139056	1698,2	772,6	1	1	-82,0661	LKAP	NM_014647.1
ILMN_19531	779,9	291,8	1	1	-81,7131	MED25	NM_030973.2
ILMN_36724	269,2	81,2	0,9993	0,9985	-80,6247	LOC645284	XM_932788.1
ILMN_137926	364,2	142,3	1	1	-79,8043	SEC14L1	NM_003003.1
ILMN_76396	415,3	168,1	1	1	-79,5688		CR607514
ILMN_35260	287,6	104,5	1	0,9993	-79,0399	LOC647319	XM_930388.1
ILMN_138087	209,2	52,2	0,9993	0,9985	-78,9278	LOC653133	XM_926881.1
ILMN_28542	2397,8	648	1	1	-78,6298	FBXO7	NM_012179.3
ILMN_27797	521,4	74,1	1	0,9978	-78,5125	TNS1	NM_022648.3
ILMN_138988	10793,6	2416,6	1	1	-78,4305	SLC25A37	NM_016612.1
ILMN_2491	3964	1871	1	1	-78,2396	PCBP2	NM_031989.2
ILMN_23078	422,4	160,7	1	1	-77,7476	RPIA	NM_144563.1
ILMN_11894	4862,2	1836,3	1	1	-77,5526	GLRX5	NM_016417.2
ILMN_23877	2185,8	1044,5	1	1	-77,2785	HK1	NM_033500.1
ILMN_9078	470,7	137,1	1	0,9993	-77,2222	C21orf7	NM_020152.2
ILMN_110256	1179,7	551,1	1	1	-77,0577		AK097979
ILMN_29773	1313,3	223,5	1	1	-76,821	EPB49	NM_001978.1
ILMN_2600	553,1	233,2	1	1	-76,4324	CREBBP	NM_004380.1
ILMN_30156	3938,7	1915,7	1	1	-75,8076	POLR2A	NM_000937.2
ILMN_4207	291,6	66,6	1	0,9985	-75,7314	MMD	NM_012329.2
ILMN_14781	198,7	61,4	0,9993	0,9985	-75,302	NCAM1	NM_181351.1
ILMN_24367	361,9	133,7	1	1	-74,6515	CD82	NM_001024844.1
ILMN_16716	194,7	59,9	0,9993	0,9985	-74,5398	TPD52L3	NM_001001874.1
ILMN_19409	1198,7	559,3	1	1	-74,5224	PRDX5	NM_012094.3
ILMN_21964	4983,7	2180,6	1	1	-74,1078	PIM1	NM_002648.2
ILMN_21880	144,9	32,8	0,9978	0,9978	-74,0405	SOX12	NM_006943.2
ILMN_18391	237,2	83,4	0,9993	0,9993	-73,7879	ZNF226	NM_001032372.1
ILMN_27883	241,8	86,4	0,9993	0,9993	-73,1735	TAAR2	NM_001033080.1
ILMN_127862	210,5	69,6	0,9993	0,9985	-73,1028		BU940418
ILMN_4250	320,5	127,9	1	1	-72,6628	ZNF653	NM_138783.1
ILMN_4109	370,6	143,4	1	1	-72,4628	CRAT	NM_000755.2
ILMN_18451	324,2	130,9	1	1	-72,3998	COQ4	NM_016035.1
ILMN_102192	579,8	234,2	1	1	-72,2493		BC015367
ILMN_19730	1943,4	469,7	1	1	-71,2218	E2F2	NM_004091.2
ILMN_139034	746	254,8	1	1	-70,2588	KCNJ10	NM_002241.2
ILMN_4175	1161	568,9	1	1	-69,9883	NOTCH1	NM_017617.2
ILMN_136712	296,6	110,1	1	0,9993	-69,6062		DR980498

TargetID	Signal_X	Signal_Y	Detection_X	Detection_Y	Diff_Score	Symbol	Accession
ILMN_8819	286,3	114,6	1	1	-69,0512	NEO1	NM_002499.1
ILMN_26461	711,3	314,5	1	1	-68,9399	FFAR2	NM_005306.1
ILMN_130435	261,5	99,9	0,9993	0,9993	-68,9033		AA385015
ILMN_17925	2987,2	1219,4	1	1	-68,4337	YPEL3	NM_031477.3
ILMN_24372	643,2	260,1	1	1	-68,3498	ZNF180	NM_013256.1
ILMN_16460	408,4	128,3	0,9999	0,9999	-68,3049	IQWD1	NM_018442.2
ILMN_15879	268,7	95,5	1	1	-67,7766	HMBS	NM_001024382.1
ILMN_5447	180,4	40,4	0,9985	0,9978	-67,2696	FECH	NM_000140.2
ILMN_132513	262,8	103,8	0,9993	0,9993	-67,2526		DB368382
ILMN_114931	389,2	156	1	1	-67,2436		BM683326
ILMN_6719	125,6	18,5	0,9978	0,9971	-66,6753	NR3C1	NM_001018076.1
ILMN_86352	249,8	85,3	0,9993	0,9993	-66,6633		BX109473
ILMN_14121	297,2	114,1	1	0,9993	-66,5742	ELSPBP1	NM_022142.3
ILMN_6069	322	137	1	1	-66,545	MARVELD1	NM_031484.1
ILMN_2922	221,1	81,4	0,9993	0,9993	-66,5159	UNQ846	NM_207316.1
ILMN_28928	289,6	119,4	1	1	-66,4557	CYB561D1	NM_182580.1
ILMN_11047	351,6	154	1	1	-65,85	EXOSC7	NM_015004.2
ILMN_2016	193,6	54,6	0,9993	0,9985	-65,4038	SAC	NM_018417.2
ILMN_77002	326,9	133,3	1	1	-65,3061		R94482
ILMN_8714	355,8	157,2	1	1	-65,1806	PGRMC1	NM_006667.2
ILMN_8545	386	174,8	1	1	-64,4239	YIPF6	NM_173834.2
ILMN_80542	371,6	140,3	1	1	-64,381		BF508732
ILMN_23277	578,3	234,8	1	1	-63,7359	RFX3	NM_002919.2
ILMN_16550	261	106,7	0,9993	1	-63,6438	GPC6	NM_005708.2
ILMN_103222	217,7	82,5	0,9993	0,9993	-63,4133		DB338803
ILMN_4191	638,5	274,2	1	1	-63,393	HNRPUL1	NM_144732.1
ILMN_110932	178,9	60,6	0,9985	0,9985	-62,9122		CN283701
ILMN_3537	389,8	110,4	1	1	-62,6282	BAT3	NM_080703.1
ILMN_10380	1820,5	332,1	1	1	-62,6188	SLC4A1	NM_000342.1
ILMN_86083	194,1	69,8	0,9993	0,9985	-62,5932		BG024455
ILMN_1425	1648,5	672,9	1	1	-62,3046	MLLT7	NM_005938.1
ILMN_9989	187,3	66,1	0,9993	0,9985	-62,2554	RAD17	NM_133343.1
ILMN_25685	153,2	44,4	0,9985	0,9978	-62,0035	SIPA1L3	NM_015073.1
ILMN_138838	239,8	96,6	0,9993	0,9993	-61,9085	KRTHB6	XM_937714.1
ILMN_118825	169,7	56,4	0,9985	0,9985	-61,5973		BM669906
ILMN_4479	398	165,4	1	1	-61,473	MTRF1	NM_004294.2
ILMN_26367	200,6	52,4	0,9993	0,9978	-61,4659	GLUL	NM_001033056.1
ILMN_123842	486,2	192,8	1	1	-61,4116		BX387364
ILMN_110259	145,7	41,4	0,9985	0,9978	-61,1767		AW593356
ILMN_127989	204,1	76,8	0,9993	0,9985	-61,1684		CF552427
ILMN_89485	265,9	112,8	0,9993	1	-60,7346		BX538009
ILMN_14315	141,8	24,1	0,9978	0,9956	-60,4241	MAP2K3	NM_145109.1
ILMN_10263	319,5	87,6	1	0,9978	-59,9654	TIAM2	NM_012454.3
ILMN_110855	642,6	258,4	1	1	-59,887		BX101351
ILMN_80310	175,3	53,1	0,9985	0,9978	-59,8848		BX115904
ILMN_18833	310,7	127	1	1	-59,6939	GBGT1	NM_021996.3
ILMN_122627	182,6	65,9	0,9985	0,9978	-59,3719		AA725270
ILMN_104040	223,1	89,7	0,9993	0,9993	-59,2994		CA438159
ILMN_30047	499,8	220,9	1	1	-58,9913	C16orf51	NM_015421.1

TargetID	Signal_X	Signal_Y	Detection_X	Detection_Y	Diff_Score	Symbol	Accession
ILMN_11721	214	83,8	0,9993	0,9993	-58,6057	ZNF583	NM_152478.1
ILMN_128307	192,4	72,5	0,9993	0,9985	-58,5536		AK125005
ILMN_11604	212,2	64,1	0,9993	0,9978	-58,5425	SESN3	NM_144665.2
ILMN_24076	551,9	161,8	1	1	-58,3884	CHPT1	NM_020244.2
ILMN_114365	195,1	74,3	0,9993	0,9985	-58,3546		BQ935122
ILMN_83494	172,8	61	0,9985	0,9985	-58,35		BG149710
ILMN_31076	161,4	54,5	0,9985	0,9978	-57,9173	LOC146177	XM_370944.4
ILMN_2835	1002,5	474,8	1	1	-57,8714	MTMR3	NM_153050.1
ILMN_82888	845,9	236,6	1	1	-57,8659		BX436104
ILMN_139315	330,6	153,6	1	1	-57,7441	SEC8L1	NM_021807.2
ILMN_17049	150,4	6,4	0,9985	0,9971	-57,6891	UBE2O	NM_022066.2
ILMN_138329	297,3	126,7	1	1	-57,4844	ELOF1	NM_032377.2
ILMN_14490	1337,1	300,4	1	1	-57,474	RNF10	NM_014868.3
ILMN_24695	174,5	56,5	0,9985	0,9985	-57,4324	KREMEN2	NM_024507.2
ILMN_29690	301,7	94,5	1	1	-57,4237	HIPK2	NM_022740.2
ILMN_125730	329,5	135,4	1	1	-57,381		BG190571
ILMN_13532	137,3	25,8	0,9978	0,9964	-57,286	GMPPB	NM_013334.2
ILMN_22169	121,4	30	0,9978	0,9978	-57,1558	BRD4	NM_058243.1
ILMN_9526	1796,6	979,4	1	1	-57,0274	YY1AP1	NM_139119.1
ILMN_33340	12141,7	3274,8	1	1	-56,9179	LOC643904	XM_927169.1
ILMN_7219	345,2	142,5	1	1	-56,8907	CAPN6	NM_014289.2
ILMN_129872	189,9	72,6	0,9993	0,9985	-56,8796		AA972156
ILMN_9861	311,1	48,1	1	0,9985	-56,7396	HDC	NM_002112.1
ILMN_30319	198,1	67,5	0,9993	0,9985	-56,6536	H2-ALPHA	NM_080386.1
ILMN_81638	254,1	110,9	0,9993	0,9993	-56,5102		BX108670
ILMN_132161	239,1	102,4	0,9993	0,9993	-56,2527		CR984787
ILMN_6970	1094,5	329,8	1	1	-55,8941	MYL4	NM_002476.2
ILMN_107859	170,7	62,1	0,9985	0,9985	-55,6358		BE646244
ILMN_33406	149,9	26,3	0,9985	0,9971	-55,5115	LOC643933	XM_927191.1
ILMN_9527	242	105,2	0,9993	0,9993	-55,2992	NUDT15	NM_018283.1
ILMN_28284	211,9	87,3	0,9993	0,9993	-55,2686	RNF11	NM_014372.3
ILMN_21352	210,4	86,4	0,9993	0,9993	-55,2365	SLC26A2	NM_000112.2
ILMN_24485	367,4	134,5	1	1	-55,1179	NALP1	NM_001033053.1
ILMN_124464	175,3	65,4	0,9985	0,9985	-55,1066		AA913861
ILMN_29645	6386,2	2806,5	1	1	-55,042	NUCB1	NM_006184.3
ILMN_124375	202	81,6	0,9993	0,9985	-55,0306		AW169912
ILMN_7530	519,8	257,6	1	1	-55,0288	SNW1	NM_012245.2
ILMN_21611	122,4	32,9	0,9978	0,9978	-54,9005	CHRNB3	NM_000749.2
ILMN_15175	526,4	219,9	1	1	-54,8193	CENTD2	NM_139181.1
ILMN_2951	2806,3	1291,3	1	1	-54,7326	YBX1	NM_004559.2
ILMN_125966	529,3	161,4	1	1	-54,5375		BE675125
ILMN_128325	282,4	117,9	1	0,9993	-54,3791		BF673116
ILMN_29794	228,7	82,4	0,9993	0,9993	-54,3578	RNF123	NM_022064.2
ILMN_11854	442,5	177,2	1	1	-54,3028	IRX1	NM_024337.3
ILMN_74989	148	39,3	0,9985	0,9978	-54,1434		AI286189
ILMN_115042	444,1	174,4	1	1	-53,8012		R38710
ILMN_20495	426,3	200,7	1	1	-53,1321	ZMAT2	NM_144723.1
ILMN_5069	962,8	404,2	1	1	-52,8875	PRR5	NM_001017530.1
ILMN_6530	1287,1	501,1	1	1	-52,7907	C1orf24	NM_022083.1

9.6 Appendix F

List of all the putative ZNF526 target genes (corresponding to 10 kbp up stream and down stream of transcription start site).

fold_enrichment	FDR(%)	accession No.	name2	distance to txStart
17,18	19,75	NM_139169	TRUB1	-9942
27,49	15,66	NM_001005483	OR4K5	-9896
30,98	19,12	NR_021485	C3orf50	-9829
22,91	19,49	NM_153211	TTC39C	-9584
28,63	17,14	NM_000074	CD40LG	-9496
27,49	17,35	NM_024109	C16orf68	-9375
23,24	17,11	NM_006133	DAGLA	-9147
17,18	17,82	NM_000526	KRT14	-9125
23,24	19,04	NM_018368	LMBRD1	-8899
30,98	19,29	NM_001380	DOCK1	-8216
22,91	17,17	NM_001162894	KIAA0040	-8197
46,48	11,34	NM_000787	DBH	-8061
28,63	10,67	NM_181538	GJC3	-8015
23,24	17,87	NM_003403	YY1	-7998
24,54	14,72	NM_002891	RASGRF1	-7958
24,54	15,26	NM_015852	ZNF117	-7847
38,73	15,93	NR_027068	EFCAB10	-7834
30,98	15,87	NM_207446	FAM174B	-7739
17,18	14,62	NM_057165	COL6A3	-7468
22,91	15,56	NM_198464	T-SP1	-7373
27,49	19,4	NM_001130037	ELMOD1	-7293
27,49	17,39	NM_015039	NMNAT2	-7200
30,98	18,67	NM_002001	FCER1A	-7156
30,98	18,55	NM_001553	IGFBP7	-7058
30,98	16,34	NM_031461	CRISPLD1	-6870
38,73	15,93	NM_005671	UBXN8	-6546
30,98	14,21	NM_005574	LMO2	-6514
23,24	18,51	NM_016010	FAM164A	-6079
38,73	15,93	NM_138705	CALML6	-6006
19,63	19,31	NM_007356	LAMB4	-5695
38,73	11,34	NM_000174	GP9	-5622
21,48	18,38	NM_052942	GBP5	-5466
20,62	19,48	NM_016205	PDGFC	-5389
22,91	19,88	NM_005655	KLF10	-5186
17,18	18,61	NM_176882	TAS2R40	-4962
38,73	15,45	NM_004079	CTSS	-4881
38,73	11,34	NM_004147	DRG1	-4753
22,91	16,08	NM_001349	DARS	-4502
17,18	19,57	NM_018436	ALLC	-4451
21,48	18,38	NM_153270	KLHL34	-4378
17,18	19,4	NM_033131	WNT3A	-3984
23,24	18,86	NM_153605	CRYBG3	-3928

fold_enrichment	FDR(%)	accession_No.	name2	distance to tsst
23,24	18,01	NM_058222	TECTB	-3645
19,63	19,97	NM_002485	NBN	-3643
22,91	18,76	NM_002070	GNAI2	-3371
17,18	19,13	NM_182678	UBE2E3	-3002
21,48	12,46	NM_005082	TRIM25	-2945
20,62	19,95	NM_001161726	PPP2R5C	-2638
38,73	15,93	NM_004688	NMI	-2581
30,98	17,36	NM_033334	NR6A1	-2528
22,91	19,77	NM_153277	SLC22A6	-2505
46,48	11,34	NM_207440	C13orf35	-2464
38,73	15,93	NM_003382	VIPR2	-2436
22,91	19,83	NM_017418	DEC1	-2412
38,73	12,71	NM_021732	AVPI1	-2286
20,62	19,25	NM_002587	PCDH1	-2241
21,48	14,32	NM_005400	PRKCE	-1988
23,24	19,15	NM_016653	ZAK	-1966
30,98	14,56	NM_152278	TCEAL7	-1853
13,74	19,55	NM_201517	H2AFV	-1830
22,91	19,47	NM_198923	MRGPRD	-1783
22,91	17,43	NM_001131062	SAP30L	-1702
30,98	17,48	NM_007324	ZFYVE9	-1671
30,98	17,33	NR_026876	STL	-1651
38,73	15,93	NM_001004698	OR2W5	-1603
38,73	15,93	NM_002076	GNS	-1587
17,18	10,73	NM_004575	POU4F2	-1373
17,18	19,97	NM_032943	SYTL2	-1369
22,91	17,41	NM_001042482	TPK1	-1367
23,24	17,47	NM_181533	ABHD12B	-1336
17,18	17,72	NM_001079669	TMTC4	-1330
17,18	18,62	NM_017814	TMEM161A	-1309
23,24	17,89	NM_173554	C10orf107	-1266
20,62	19,74	NM_194247	HNRNPA3	-1189
30,98	18,66	NM_001164443	ANKRD31	-1104
34,36	17,72	NM_001034023	DMAP1	-1098
22,91	17,83	NM_001034850	FAM134B	-1024
20,04	14,59	NM_033026	PCLO	-1003
23,24	17,51	NM_002597	PDC	-942
17,18	16,61	NM_005985	SNAI1	-915
38,73	15,93	NM_080666	WDR89	-898
38,73	13,13	NM_017838	NHP2	-862
17,18	14,62	NM_032036	IFI27L2	-831
28,63	15,37	NM_001010000	ARHGAP28	-819
17,18	19,95	NM_001010898	SLC6A17	-795
22,91	19,79	NM_022036	GPRC5C	-727
30,98	17,63	NM_001017920	DAPL1	-724
24,54	15,26	NM_199290	NACA2	-704
23,24	17,57	NM_006943	SOX12	-601
10,31	19,67	NM_030576	LIMD2	-558
21,48	18,38	NM_006449	CDC42EP3	-553
27,49	19,79	NM_014857	RABGAP1L	-536
23,24	17,45	NR_027135	ZSCAN18	-531
38,73	13,51	NM_015199	ANKRD28	-458
46,48	11,04	NM_004405	DLX2	-444
23,24	18,6	NM_001017926	ZHX1	-420
38,73	15,93	NM_001907	CTRL	-346
30,98	18,38	NM_198977	ARHGEF1	-333
27,49	17,5	NM_001143885	PPP1R12A	-309
30,98	17,66	NM_025203	C2orf44	-253
21,48	19,49	NM_001039958	MESP2	-173
23,24	19,1	NM_153426	PITX2	-89
27,49	17,25	NM_015102	NPHP4	-11

fold_enrichment	FDR(%)	accession_No.	name2	distance to txStart
30,92	0	NM_001006	RPS3A	0
44,67	0	NM_032120	C7orf64	0
40,09	9,48	NM_001129728	PLEKHG4	0
40,09	9,48	NM_001129729	PLEKHG4	0
40,09	9,48	NM_001129731	PLEKHG4	0
27,49	10,28	NM_002807	PSMD1	0
34,36	10,37	NM_016167	NOL7	0
40,09	10,37	NM_014888	FAM3C	0
40,09	10,37	NM_001040020	FAM3C	0
38,73	10,42	NM_005679	TAF1C	0
38,73	10,42	NM_139353	TAF1C	0
46,48	10,71	NM_058219	EXOSC6	0
28,63	11,35	NM_001171167	CAMTA2	0
28,63	11,35	NM_001171166	CAMTA2	0
28,63	11,35	NM_015099	CAMTA2	0
28,63	11,35	NM_001171168	CAMTA2	0
38,73	11,61	NM_004685	MTMR6	0
46,48	11,76	NM_001098414	ZNF621	0
46,48	11,76	NM_198484	ZNF621	0
30,98	13,18	NM_001143852	TCHP	0
30,98	14,13	NM_177985	ARL5A	0
30,98	14,13	NM_001037174	ARL5A	0
30,98	14,13	NM_012097	ARL5A	0
30,98	14,68	NM_006298	ZNF192	0
30,98	14,84	NM_016173	HEMK1	0
21,48	15,05	NM_025054	VCPIP1	0
30,98	15,3	NM_024105	ALG12	0
38,73	15,91	NM_014165	NDUFAF4	0
38,73	15,93	NM_004894	C14orf2	0
38,73	15,93	NM_001127393	C14orf2	0
38,73	15,93	NM_138792	LEO1	0
38,73	15,93	NM_019109	ALG1	0
38,73	15,93	NM_001154	ANXA5	0
34,36	16,53	NM_033414	ZNF622	0
23,24	16,55	NM_001105247	ARMC5	0
23,24	16,55	NM_024742	ARMC5	0
19,63	17,22	NM_001166226	CEP120	0
19,63	17,22	NM_153223	CEP120	0
17,18	17,23	NM_020770	CGN	0
22,91	17,42	NM_003447	ZNF165	0
14,73	17,44	NM_032924	ZNF3	0
14,73	17,44	NM_017715	ZNF3	0
30,98	17,48	NM_032194	RPF2	0
30,98	17,72	NM_005654	NR2F1	0
30,98	17,84	NM_000392	ABCC2	0
27,49	17,96	NM_000990	RPL27A	0
23,24	18,27	NM_000254	MTR	0
27,49	18,32	NM_001118890	GLRX	0
27,49	18,32	NM_002064	GLRX	0
25,77	18,38	NM_001135187	AGFG1	0
25,77	18,38	NM_004504	AGFG1	0
25,77	18,38	NM_001135188	AGFG1	0
25,77	18,38	NM_001135189	AGFG1	0
25,77	18,38	NM_001145445	SFRS15	0
25,77	18,38	NM_020706	SFRS15	0
25,77	18,38	NM_001145444	SFRS15	0
38,73	18,57	NM_016100	NAA20	0
38,73	18,57	NM_181528	NAA20	0
20,62	18,62	NM_003675	PRPF18	0
20,62	18,62	NM_024685	BBS10	0
17,18	18,62	NM_032848	C12orf52	0
17,18	18,62	NM_000987	RPL26	0
24,54	19,43	NM_001098783	RPP14	0
24,54	19,43	NM_007042	RPP14	0
21,48	19,49	NR_030178	MIR497	0
12,78	19,72	NM_016424	LUC7L3	0

fold_enrichment	FDR(%)	accession No.	name2	distance to txStart
20,62	14,62	NM_001114091	CDC27	22
38,73	15,93	NR_033182	LOC100130872-SPON2	64
38,73	11,11	NR_026780	C6orf208	99
21,48	19,49	NM_001083913	C10orf26	115
20,62	17,72	NM_004333	BRAF	120
30,98	17,48	NM_152359	CPT1C	140
30,98	15,73	NM_003810	TNFSF10	143
30,98	17,48	NM_001136052	CPT1C	146
22,91	19,3	NM_024979	MCF2L	157
38,73	15,93	NR_024569	LOC100130872	180
21,48	19,9	NM_152637	METTL7B	226
34,36	17,72	NM_005322	HIST1H1B	237
30,98	15,12	NM_001001921	OR5AS1	239
30,98	18,71	NM_001099788	ICAM2	245
30,98	18,71	NM_001099786	ICAM2	245
30,98	18,71	NM_001099789	ICAM2	245
30,98	18,71	NM_001099787	ICAM2	245
17,18	19,8	NM_001989	EVX1	255
27,49	18,02	NM_001010874	TECRL	354
23,24	17,47	NM_002854	PVALB	421
23,24	17,43	NR_028295	NCRNA00029	473
17,18	14,62	NM_001145306	CDK6	480
22,91	18,7	NM_207102	FBXW12	565
22,91	18,7	NM_001159929	FBXW12	565
22,91	18,7	NM_001159927	FBXW12	565
23,24	17,61	NM_003703	NOP14	646
38,73	11,34	NM_207360	ZC3H12D	677
30,98	17,44	NM_003145	SSR2	704
23,24	19,04	NM_015148	PASK	710
30,98	16,08	NM_148962	OXER1	728
17,18	18,86	NM_022074	FAM111A	733
17,18	18,86	NM_198847	FAM111A	733
38,73	15,93	NM_021575	AP2S1	764
38,73	15,93	NM_004069	AP2S1	764
14,73	19,41	NM_001102454	ZNF219	826
40,09	9,48	NM_001129727	PLEKHG4	937
34,36	12,5	NM_201383	PLEC	998
17,18	19,03	NM_003648	DGKD	1009
34,36	15,32	NM_014906	PPM1E	1046
22,91	19,88	NM_005099	ADAMTS4	1071
23,24	18,63	NM_001134878	KIF9	1115
23,24	18,63	NM_022342	KIF9	1115

fold_enrichment	FDR(%)	accession No.	name2	distance to txStart
19,63	18,42	NM_020946	DENND1A	4281
19,63	18,42	NM_024820	DENND1A	4281
20,62	17,89	NM_015131	WDR43	4286
14,73	19,41	NM_001146683	C14orf176	4305
30,98	17,37	NM_000136	FANCC	4327
17,18	19,74	NM_001040875	GDAP1	4375
17,18	19,74	NM_018972	GDAP1	4410
23,24	18,66	NR_027450	FAM186B	4435
23,24	18,66	NM_032130	FAM186B	4458
22,91	17,41	NM_182904	P4HA3	4537
38,73	15,93	NM_030812	ACTL8	4647
17,18	19,9	NM_002700	POU4F3	4744
30,98	8,51	NM_020802	KIAA1377	4782
30,98	15,99	NM_001024382	HMBS	4813
30,98	17,45	NR_004422	HPVC1	5004
17,18	19,55	NM_001025069	ARPP21	5028
17,18	19,55	NM_001025068	ARPP21	5028
20,62	18,5	NM_206909	PSD3	5096
23,24	19	NM_001098797	TOX2	5106
28,63	18,04	NM_032507	PGBD1	5209
14,73	19,45	NM_021183	RAP2C	5211
30,98	18,39	NR_027867	PDCD6IP	5306
38,73	15,93	NR_002775	RPLP0P2	5493
30,98	17,11	NM_012083	FRAT2	5604
17,18	19,25	NM_001677	ATP1B1	5622
17,18	19,25	NM_001001787	ATP1B1	5622
38,73	15,93	NM_198464	T-SP1	5726
23,24	19,19	NM_005312	RAPGEF1	5740
30,98	18,56	NM_012328	DNAJB9	5743
23,24	17,81	NM_004475	FLOT2	5822
38,73	15,93	NM_000912	OPRK1	5837
17,18	19,83	NM_173632	ZNF776	5907
21,48	18,38	NM_006909	RASGRF2	5940
23,24	17,83	NM_001005487	OR13G1	6005
23,24	19,09	NM_012339	TSPAN15	6155
28,63	18,04	NM_001042585	CD1E	6173
20,62	16,63	NM_012319	SLC39A6	6197
20,62	16,63	NM_001099406	SLC39A6	6197
17,18	17,8	NM_004466	GPC5	6208
21,48	19,49	NM_001006628	CHRM2	6272
21,48	19,49	NM_001006629	CHRM2	6281
30,98	18,7	NM_020931	KIAA1586	6372
21,48	19,49	NM_001006626	CHRM2	6396
30,98	17,97	NM_032599	FAM71F1	6422
21,48	19,49	NM_001006632	CHRM2	6433

fold_enrichment	FDR(%)	accession No.	name2	distance to txStart
21,48	19,49	NM_000739	CHRM2	6433
17,18	19,25	NM_012395	CDK14	6445
28,63	15,37	NM_001081492	KRT80	6507
28,63	15,37	NM_182507	KRT80	6507
30,98	18,7	NM_004006	DMD	6808
30,98	16,63	NM_001077621	VPS37D	6817
23,24	19,18	NM_001146197	LOC643677	6839
21,48	19,49	NM_001006627	CHRM2	6866
21,48	19,49	NM_001006630	CHRM2	6866
22,91	17,42	NM_022779	DDX31	6893
22,91	17,42	NM_138620	DDX31	6893
17,18	18,62	NM_138635	H2AFV	6912
17,18	18,62	NM_201436	H2AFV	6912
17,18	18,62	NM_012412	H2AFV	6912
17,18	18,62	NM_201517	H2AFV	6912
17,18	18,62	NM_201516	H2AFV	6912
17,18	19,69	NM_021216	ZNF71	6987
30,98	17,11	NM_001136202	ISOC2	7075
30,98	17,11	NM_001136201	ISOC2	7075
30,98	17,11	NM_024710	ISOC2	7075
30,98	16,06	NM_198846	SIGLEC6	7125
30,98	16,06	NM_198845	SIGLEC6	7125
30,98	16,06	NM_001245	SIGLEC6	7125
17,18	16,03	NM_004594	SLC9A5	7136
30,98	14,09	NM_153615	RGL4	7177
12,89	19,96	NR_029701	MIR146A	7187
38,73	16,39	NM_001005463	EBF3	7311
34,36	17,72	NM_015462	NOL11	7311
38,73	14,32	NM_031371	ARID4B	7318
38,73	14,32	NM_016374	ARID4B	7318
20,62	19,44	NM_022118	RBM26	7330
22,91	17,45	NM_021622	PLEKHA1	7742
22,91	15,69	NM_002528	NTHL1	7783
38,73	11,34	NR_023382	ZNF815	7820
23,24	17,81	NM_002754	MAPK13	7825
20,62	18,95	NM_001080419	UNK	7855
30,98	15,99	NM_000190	HMBS	7923
14,73	19,8	NM_001559	IL12RB2	7928
30,98	15,22	NM_012367	OR2B6	8006
27,49	19,7	NM_000938	POLR2B	8040
38,73	8,94	NM_001143667	ZBED5	8045
30,98	16,63	NM_173834	YIPF6	8127
38,73	11,11	NM_032312	YIPF4	8154
23,24	19,25	NM_004115	FGF14	8226
23,24	17,98	NM_001048230	ADORA1	8294

fold_enrichment	FDR(%)	accession No.	name2	distance to txStart
13,74	19,71	NM_005229	ELK1	8297
13,74	19,71	NM_001114123	ELK1	8297
38,73	11,11	NR_026781	C6orf122	8372
17,18	19,28	NM_198570	VWC2	8375
17,18	18,62	NM_024325	ZNF343	8423
38,73	11,34	NM_003996	GPX5	8435
38,73	11,34	NM_001509	GPX5	8435
30,98	18,65	NM_001945	HBEGF	8473
23,24	17,44	NM_003396	WNT9B	8643
30,98	14,65	NM_003492	TMEM187	8668
30,98	18,54	NM_005530	IDH3A	8811
19,63	19,47	NM_002995	XCL1	8878
38,73	11,34	NM_001143905	C12orf65	8889
38,73	11,34	NM_014702	KIAA0408	8905
21,48	18,38	NR_026705	VTRNA1-3	9099
38,73	15,93	NM_005722	ACTR2	9166
38,73	15,93	NM_001005386	ACTR2	9166
38,73	11,34	NM_152269	C12orf65	9350
34,36	12,5	NM_201380	PLEC	9350
17,18	18,52	NM_144664	FAM76B	9571
23,24	18,86	NM_005038	PPID	9820
23,24	18,23	NM_001172646	PLCB4	9851
23,24	18,93	NM_031907	USP26	9857
23,24	17,92	NM_002557	OVGP1	9922
27,49	15,53	NM_181808	POLN	9925
17,18	19,55	NR_026661	MGC14436	9999
17,18	19,55	NR_026662	MGC14436	10002

9.7 Appendix G

Molecular and cellular function as well as involvement of ZNF526 putative target genes (the distance of 10 Kbp to transcription start site).

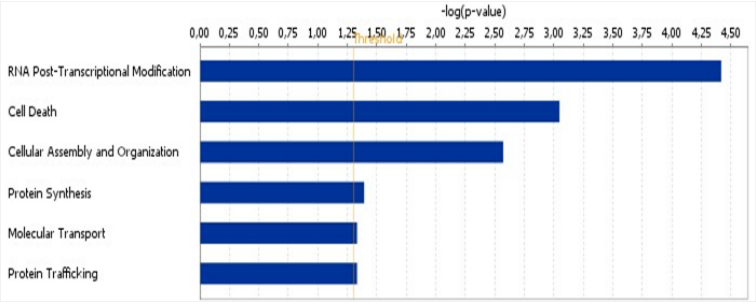


Figure 9.1: Molecular and cellular functions of ZNF526 target genes.

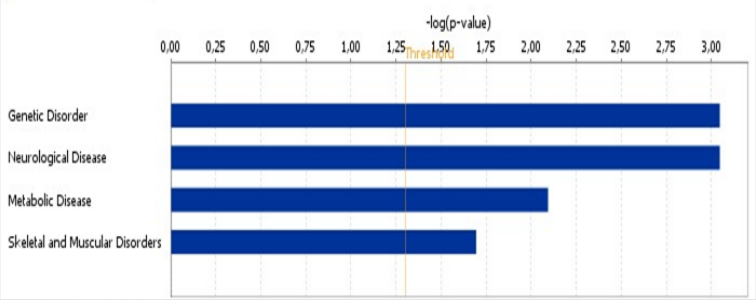


Figure 9.2: Disease and disorders - ZNF526 target genes.

10 Acknowledgements

This dissertation would not have been possible without the guidance and help of several individuals who in one way or another contributed and extended their valuable assistance in the preparation and completion of this study.
First and foremost, I would like to express my utmost gratitude to Prof. Hans-Hilger Ropers whose sincerity and encouragement I will never forget. I am deeply grateful to him for letting me join his lab and to have the opportunity to work on this fantastic project. His wide knowledge and his logical way of thinking have been of great value to me. His understanding and personal guidance have provided a good basis for the present thesis. Prof. Ropers has been my inspiration as I overcome the hurdles in the completion of this research work.
It is difficult to overstate my gratitude to Prof. Najmabadi, with whoes enthusiasm and great efforts I have been able to join the MPI in Berlin. He supported me throughout my thesis period with encouragement, sound advice, the clinical information and patient materials from such a big cohort of families, and lots of good ideas and follow ups.
I am indebted to Prof. Andreas Kuss whose help, organizational management and stimulating suggestions helped me during the research and writing of the manuscripts.
I owe my most sincere gratitude to Professor Stephan Sigrist, my second advisor.The investigation of drosophila as an animal model was done in his lab (Institute of Biology, Genetics, Free university of Berlin Germany). I wish to extend my special thanks to Sara Mertel for all her efforts to perform and manage all of the drosophila related experiments.
I am also indebted to the members of the supervisory committee.
I would like to show my deepest gratitude to Dr. Kimia Kahrizi, who not only prepared all the clinical data but also gave me friendly support throughout this study.
I would like to thank USWR(University of Welfare and Rehabilitation Sciences, Tehran, Iran) authorities for giving me this opportunity to study this project in MPI, Germany.
Many thanks to Dr. Andreas Tzschach for his essential assistance in reviewing the patient files of this study and collecting the clinical data.
I warmly thank Dr. Lars Jensen for his advice and friendly help. His extensive discussions around my work and interesting explorations in operations have been very helpful for this study.
I am grateful to Prof. Matthias Mann and Dr. A. Hyman and N. Hubner for mass-spectrometry experiments performed at the Max planck institute of Biochemistry, Martinsfried Germany.
I thank Ms. Marriane Schlicht and Bettina Lipkowitz for their contribution in the lab.
I have furthermore to thank our secretary, Mrs. Gabriele Eder, for great help in visa applications and all the different administration tasks.
I would like to thank Mrs. Susanne Freier for her help in cell culture and also for being accommodating in Germany at the beginning.
I am indebted to my friends for providing a stimulating and beautiful environment in which to

learn and grow. I am especially grateful to Joanna Walczak-Stulpa for helping me get through the difficult times and for the unforgettable times we had together. I also wish to thank Lucia Puettmann for kindly assisting me in many different ways and for all her personal guidance and also for revising the English of my dissertation. I also would give my thanks to Agnes Zecha for all the emotional support and her way of thinking, which have been of great value to me.
I wish to express my warm and sincere thanks to all of the professors,Dr. Yousef Shafeghati, Dr. Mina Ohadi, Dr. Maryam Neishabouri, Dr. Keyhani, Dr. Behjati and Dr. Mohammad Khalil Javan, at the Genetics research center, university of Welfare and Rehabilitation Sciences, Tehran (USWR).
I thank Masoud Garshasbi for Linkage analysis on the families and especially for friendly discussions. I also thank Dr. Hau Hu and Robert Weissmann for preparing CHIP-Seq data and Sara Behrens for working on the expression data for motif finding.
I wish to thank all of my previous and current lab member and all the other colleagues, PhD students and technicians in our department: Mahdi Motazacker, Georg Lienke, Marion Amende, Achim Salamon, Pamela Kepper, Hannalore Madle , Vanessa Sukow, Nadine, Corina Menzel, Artur Moradian and Julia Hossir for providing such a nice and friendly environment in the lab and all their help during this period in MPI.
I appreciate that my colleagues from USWR have been involved in my research work. I want to thank them for all their help, support, interest. Especially, I warmly thank Mrs. Khadijeh jalalvand, Ms. Susan Banihashemi, Ms. Saghar Ghasemi Firouzabadi, Mrs. Sedigheh Abedini, S. Sanaz Arzhangi, Ms. Nastaran Arabinejad, Mrs. Marzieh Mohseni and Mrs. Nourusali.
I wish to thank Roxana Kariminejad for the insights she has shared.
Many thanks to Ms. Tamara Safari for ordering primers and all the other materials regarding the sequencing and Ms. Pamela Kepper for doing the Illumina expression and genotyping hybridizations.
I take advantage of this opportunity to thank Dr. Vera Kalscheuer, Dr. Rheinhard Ullmann, and Dr. Tim Hucho for the scientific discussions and inputs during our regular Friday seminars.
I am deeply thankful to all of the patients and their respective families, for without their agreements it was not possible at all to conduct such a study.

Lastly, and most importantly, I wish to thank my parents, my sister and her family for their understanding and endless love, throughout the duration of my studies. I owe my loving thanks to my husband whose patient love enabled me to complete this work and support me spiritually throughout my study. To them I dedicate this thesis.

11 Summary

Intellectual disability (ID), one of the most complex disorders, has a worldwide prevalence of approximately 2% and is a frequent cause of severe disability. Therefore this disorder constitutes a major burden not only on the affected families but also on society. It has become clear that X-linked forms account for only ten percent of ID cases, which means that the vast majority of the underlying genetic defects must be autosomal, but it has so far received considerably less attention. A particularly straightforward strategy for the identification of genes underlying autosomal recessive disorders is homozygosity mapping in extended consanguineous families, followed by mutation screening of candidate genes. In Western societies, where most of the research takes place, its investigation has been hampered by infrequent parental consanguinity and small family sizes. Therefore, to unravel the molecular basis of ARID in a systematic fashion as a prerequisite for diagnosis, counselling and therapy, we focused on large consanguineous Iranian families with several mentally retarded children. During the course of our investigations into the autosomal recessive causes of intellectual disability (ARID,) we have identified numerous new loci for this condition.

However, no more than six hotspot loci for unspecific or non-syndromic autosomal recessive intellectual disability (NS-ARID) have been identified, which may indicate that, at least in the Iranian population, not all of the gene defects causing NS-ARID are extremely rare and the possible existence of common molecular causes for NS-ARID have not been ruled out.

The work presented here is part of this large project to shed more light on the molecular causes of ARID. In this study the investigation of two out of these 6 hot spot loci led to the identification of underlying gene defects.

One of these involves the linkage intervals of two Iranian families with several NS-ARID patients overlapping on Chr19q13.2-q13.31. Two different missense mutations with high pathogenicity scores were detected in *ZNF526*, which encodes a krueppel-type zinc finger protein. One of these changes was observed in DNA samples collected from two distinct families with non-syndromic ID, but closer inspection revealed that these families, which live in the same city in the Northwestern part of Iran, share a common haplotype and thus must be distantly related.

Each mutation affects a functional domain of ZNF526 and both alter the protein conformation, causing a putative functional impairment as suggested by in silico protein modelling. A decrease in DNA affinity was confirmed by CHIP-seq, and array-based gene expression studies showed specific changes in the expression patterns of patient lymphoblastoid cells, which could be recapitulated in *ZNF526*-deficient neuroblastoma cells. Functional annotation showed significant enrichment of the deregulated genes in pathways that play a role in protein synthesis, mitochondrial dysfunction, energy metabolism and gene regulation. We could implicate that ZNF526 protein interacts with PRKRIR, which is a transcription factor (TF) that has been added to human TF repertoire recently in the primate history. Therefore ZNF526 and PRKRIR together are particularly promising candidates in investigating the development and evolution

of higher brain function in primates.
This study also resolved the underlying gene defect of MRT5 and reported three deleterious mutations in *NSUN2*. These were found in two independent consanguineous Iranian families and one Turkish family with several patients suffering from non-syndromic ARID. *NSUN2* encodes a methyltransferase, which catalyzes the intron-dependent formation of 5-methylcytosine at C34 of tRNA-leu(CAA). Hence all mutations lead to a loss of NSUN2 protein function in homozygous mutation carriers and in all likelihood cause the patient phenotype.
In order to gain further evidence for an involvement of NSUN2 in cognitive functions, we studied Drosophila mutants that lack the *NSUN2* ortholog. These experiments revealed a marked learning impairment in mutant flies, which clearly underscores the relevance of NSUN2 in higher brain functions.

Furthermore, this study was the first report on a mutation in patients with dysequilibrium syndrome that affects *VLDLR* exclusively, confirming the central role of the very low-density lipoprotein receptor in the aetiology of this condition. The mutations in this gene have been found to be associated with quadrupedal mobility in other families, but not in our patients.

In summary, our results show that both *NSUN2* and *ZNF526* belong to the still few genes known to carry NS-ID-causing mutations in independent families, which suggests that defects in either gene belong to the more common causes of NS-ARID, at least in the Iranian population. Further studies are necessary to identify the disease causing mutations in the other 4 hot spot identified loci and to determine the contribution of the affected genes to the complex processes of human cognition. These studies will be greatly facilitated by exome enrichment and next generation sequencing (NGS), which have recently been introduced as a cost-effective and fast strategy for comprehensive mutation screening and disease-gene identification in the coding portion of the human genome.

12 Zusammenfassung

Mentale Retardierung (MR), eine der komplexesten Erkrankungen, hat eine weltweite Prävalenz von etwa 2% und ist ein häufiger Grund schwerster Behinderung. Aus diesem Grund ist diese Erkrankung sowohl für die Familie als auch für die Gesellschaft eine enorme Belastung. Es hat sich gezeigt, dass genetische Defekte auf dem X-Chromosom nur für 10% aller Fälle von MR verantwortlich sind, daher muss die überwältigende Mehrheit der genetischen Defekte auf den Autosomen kodiert sein. Die Erforschung der autosomalen MR hat bislang jedoch wesentlich weniger Aufmerksamkeit erhalten.
Die Strategie zur Identifizierung der an autosomal rezessiven Erbkrankheiten beteiligten Gene umfasst ‚Homozygosity Mapping, in großen blutsverwandten Familien mit anschließendem Sequenzieren der Kandidatengene um Mutationen aufzuspüren. In westlichen Gesellschaften, wo ein Großteil der Erforschung von MR stattfindet, ist diese Strategie aufgrund von kleinen Familien und seltener Blutsverwandtschaft der Eltern kaum einsetzbar. Um die molekularen Ursachen autosomal rezessiver MR (ARMR) systematisch zu entschlüsseln, um als Grundvoraussetzung für Diagnose, Beratung und Therapie zu dienen, haben wir den Fokus auf große, blutsverwandte iranische Familien gelegt. In unseren Forschungen zu den molekularen Ursachen von ARMR haben wir neue ARMR-Loci identifiziert, doch sind darunter sechs Hot Spot-Loci für unspezifische oder nicht syndromale ARMR (NS-ARMR). Daraus kann geschlossen werden, dass zumindest in der iranischen Bevölkerung nicht alle Gendefekte selten sind und dass die Möglichkeit gemeinsamer molekularer Ursachen von NS-ARMR nicht ausgeschlossen werden kann.
Diese Arbeit ist Teil des großen Projekts die molekularen Ursachen von ARMR zu erforschen. Zwei der sechs Hot Spots für ARMR wurden untersucht und die zugrunde liegenden Gendefekte identifiziert.
Die Kopplungsintervalle von zwei iranischen Familien mit mehreren von NS-ARMR betroffenen Patienten überlappen auf Chr19q13.2-q13.31. Im Gen *ZNF526* wurden zwei Missense-Mutationen mit hoher Pathogenitätsvoraussage identifiziert. *ZNF526* kodiert ein Krüppel-Zinkfinger-Protein. Eine dieser Mutationen wurde in einer weiteren Familie mit NS-ARMR identifiziert, die in der gleichen Stadt im Nordwesten des Irans wohnt. Weitere Analysen ergaben einen gemeinsamen Haplotyp beider Familien, die daher entfernt miteinander verwandt sein müssen.
Beide Mutationen betreffen funktionelle Domänen von ZNF526. *In silico*-Proteinmodellierung zeigte eine Veränderung der Proteinkonformation, welche wahrscheinlich die Funktion des Proteins behindert. Eine Minderung der DNA-Affinität wurde anhand von Chip-seq bestätigt. Spezifische Veränderung des Genexpressionsmusters in Lymphoblasten der Patienten wurde anhand von Arrays gezeigt. Dieser Befund konnte in *ZNF526*-defizienten Neuroblastomzellen rekapituliert werden.
Die Annotation der Genfunktionen zeigte eine Anreicherung der deregulierten Gene in Singal-

und Stoffwechselwegen, die eine Rolle in der Proteinsynthese, mitochondrialer Dysfunktion, Energiemetabolismus und Genregulation spielen.
Des Weiteren konnten wir die Interaktion von ZNF526 und PRKRIR zeigen, einem Transkriptionsfaktor, welcher sehr spät in der Primatenevolution entstand. Diese beiden Proteine sind daher sehr viel versprechende Kandidaten zur Erforschung der Entwicklung und Evolution höherer Gehirnfunktionen von Primaten.
In dieser Arbeit haben wir auch die Gendefekte in MRT5 aufgedeckt und drei schädliche Mutationen in NSUN2 identifiziert. Diese Mutationen wurden in zwei nicht miteinander verwandten blutsverwandten iranischen Familien sowie in einer türkischen Familie identifiziert. Die Patienten leiden unter NS-ARMR. NSUN2 kodiert eine Methyltransferase, welche die vom Intron abhängige Bildung von 5-Methylcytosin an das C34 der tRNA-leu(CAA) katalysiert. Alle Mutationen führen zum Verlust der Proteinfunktion und verursachen mit aller Wahrscheinlichkeit den Phänotyp der Patienten.
Um weitere Beweise für die Beteiligung von *NSUN2* an kognitiven Funktionen zu erhalten, wurden Drosophila-Mutaten untersucht, denen das NSUN2-Ortholog fehlt. Die Mutanten zeigten deutliche Behinderung des Lernens, was deutlich die Relevanz von NSUN2 für höhere Gehirnfunktionen betont.
Des Weiteren konnte diese Studie erstmals Mutationen in VLDLR (very low-density lipoprotein receptor) in Patienten mit Dysäquilibrium-Syndrom identifizieren und die zentrale Rolle des Rezeptors in der Verursachung der Krankheit bestätigen. Mutationen in VLDLR wurden auch mit vierfüßiger Fortbewegung in einigen Familien assoziiert, die jedoch in unseren Patienten nicht vorhanden ist.
Unsere Ergebnisse zeigen, dass sowohl *NSUN2* als auch *ZNF526* zu den wenigen Genen gehören, die NS-ARMR verursachende Mutationen in mehreren voneinander unabhängigen Familien tragen. Daraus lässt sich schließen, dass beide Gene zu den häufiger von Mutationen betroffenen Genen bei NS-ARMR gehören, zumindest in der iranischen Bevölkerung. Weiter Studien sind nötig, um die krankheitsverursachenden Mutationen in den anderen vier Hot Spots zu identifizieren und die Beteiligung der betroffenen Gene an der humanen Kognition zu erschließen. Diese Untersuchungen werden durch die Exom-Anreicherung und das Next Generation Sequencing (NGS) erleichtert, welche kürzlich als kosteneffiziente und schnelle Strategien für das Mutationsscreening und die Identifizierung von Krankheitsgenen in den kodierenden Regionen des humanen Genoms eingeführt wurden.