Max-Planck Institute for Molecular Genetics

Freie Universität Berlin

Identification of 31 genomic loci for autosomal recessive mental retardation and molecular genetic characterization of novel causative mutations in four genes

DISSERTATION

Zur Erlangung des akademischen grades
Doctor rerum naturalium
(Dr. Rer. Nat.)

Vorgelegt von

Masoud Garshasbi

Aus Kermanshah, Iran

Eingereicht im Fachbereich
Biologie, Chemie, Pharmazie
der
Freien Universität Berlin

Date: May 2009

Diese Dissertation wurde in der Zeit vom Februar 2004 bis May 2008 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 2004 till May 2008 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. Volker A. Erdmann**
Freie Universität
Thielallee 63, D-14195 Berlin

Second Referee: **Prof. Dr. Volker A. Erdmann**
Free University
Thielallee 63, D-14195 Berlin

Tag der Disputation: May 18th , 2009
Day of the Disputation: May 18th , 2009

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.

Masoud Garshasbi
Berlin, May 2009

1 Introduction

A key feature of the functional human brain is its ability to reason, plan, solve problems, think abstractly, comprehend complex ideas, learn quickly and from experience. This general mental capability is also referred to as "intelligence".

Intelligence can be assessed through various tasks designed to evaluate different types of reasoning. Typically this involves standardized testing with norm-referenced tests that allow to compare a proband's skills to others in his or her age group. The performance score is the Intelligence Quotient (IQ). Among the general population IQ scores are normally distributed (Fig.1) and about 2% of people have an IQ below 70, which is generally considered as the threshold for "mental retardation".

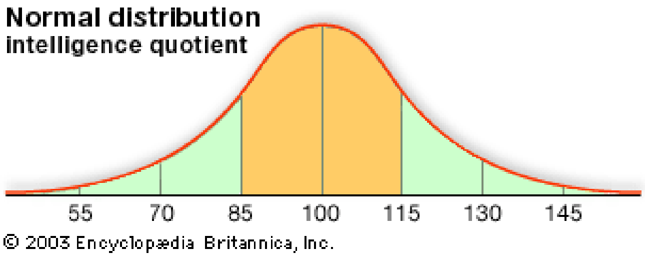


Figure 1-1: Graph of intelligence quotient (IQ) as a normal distribution with a mean of 100 and a standard deviation of 15. The shaded region between 85 and 115 (within one standard deviation of the mean) accounts for about 68 percent of the total area, hence 68 percent of all IQ scores (Britannica 2003).

1.1 Mental retardation

Based on IQ, mental retardation (MR) is subdivided into several classes. Most commonly, the WHO (World Health Organization) classification and terminology (see Table 1-1) are used (WHO 1980), but numerous studies distinguish only between mild (IQ 70–50) and severe MR (IQ <50) (Ropers and Hamel 2005).

Table 1-1: MR classes based on IQ

Terminology	Intelligence quotient
Profound	<20
Severe	20–35
Moderate	35–50
Mild	50–70
Borderline	70–85

MR is a disability, which originates before the age of 18 and is characterized by significant limitations both in intellectual functioning and in adaptive behaviour as expressed in conceptual, social, and practical adaptive skills.

Based on the definition by the American Association on Mental Retardation (AAMR 2005), adaptive behaviour is the collection of conceptual, social, and practical skills that people have learned so they can function in their everyday lives.

Some specific examples of conceptual skills are receptive and expressive language, reading and writing, money concepts and self-directions. Social skills include interpersonal, responsibility, self-esteem, gullibility (likelihood of being tricked or manipulated), naiveté, following the rules, obeying laws and avoiding victimization. Practical skills are personal activities of daily living such as eating, dressing, mobility and toileting, instrumental activities of daily living such as preparing meals, taking medication, using the telephone, managing money, using transportation and doing housekeeping activities, occupational skills and maintaining a safe environment (AAMR 2005).

Mental and behavioral disorders are common, affecting people throughout the world and usually causing severe disability (WHO The World Health Report 2001). Apart from intellectual disability, mental and behavioral disorders also include depression, substance use disorders, schizophrenia, epilepsy and Alzheimer’s disease. They constitute a major burden for the affected families but also for society.

Severe and mild forms of MR affect approximately 1-3% of the general population in the world, and health care-related costs are higher than for any other diagnosis included in the International Statistical Classification of Diseases (ICD) (Honeycutt AA 2003; Losada and Hirano 2005; Roeleveld and others 1997).

Consanguineous marriage, for which there is a cultural preference in many countries, is one important risk factor for MR and other congenital disorders. This has been amply documented by a significant excess of (both severe and mild) MR in the progeny of consanguineous matings (al-Ansari 1993; Bittles 2001; Bundey and others 1991; Durkin and others 1998; Fernell 1998; Khalid and others 2006; Kulkarni and Kurian 1990; Magnus and others 1985; Temtamy and others 1994; Yaqoob and others 1995) and finds further support in the linear correlation between the birth prevalence of congenital disorders and the coefficient of consanguinity (Bittles and Neel 1994).

Etiologically MR can result from extraordinary heterogeneous 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 syndrome) or monogenic causes (Chelly and others 2006; Ropers and Hamel 2005).

In total, the genetic cause of MR in up to 40% of cases is known so far (Ropers 2008). Conventionally, genetic forms of MR associated with clinical, radiological, metabolic or biological features are considered as syndromic MR, and those forms in which cognitive impairment represents the only manifestation of the disease are categorized as unspecific or “non-syndromic” (NS) MR. This distinction remains very useful for clinical purposes, but many recent phenotype–genotype studies and detailed clinical follow-ups of patients indicate that the boundary between syndromic and non-syndromic MR forms is often blurred, and cases of the latter are increasingly recognized as syndromic (Chelly and others 2006).

1.1.1 X-linked mental retardation (XLMR)

Based on the observation that MR is significantly (30-50%) more common in males than in females, X-linked gene defects have long been considered to be important causes of MR, (Chelly and others 2006; Ropers and Hamel 2005). Still, one should keep in mind that due to the hemizygosity of males the identification of X-linked conditions is easier, as males inevitably manifest a phenotype when harboring a mutant allele (Chiurazzi and others 2008).

The first found and most common form of X-linked mental MR (XLMR) is fragile X mental-retardation syndrome which is caused by mutations in the *FMR1* gene (for review see e.g. Chiurazzi and others 2004 or Mandel and Biancalana 2004).

Since then, clinical observations and linkage studies in many affected families revealed that XLMR is a highly heterogeneous condition. Correspondingly, mutations in more than 80 X-chromosomal genes have been found up to now (either by positional cloning or translocation breakpoint mapping methodologies), and 24 of these genes are presently considered to be implicated in NS-XLMR (Table 1-1) (Raymond and Tarpey 2006; Ropers 2006).

Table 1-1: Genes implicated in NS-XLMR (Ropers 2006).

Gene	Functions
ACSL4 (FACL4)	Long-chain fatty acid synthase; possible role in membrane synthesis and/or recycling
AGTR2	Brain-expressed angiotensin receptor 2
ARHGEF6	Integrin-mediated activation of Rac–cdc42; stimulation of neurite outgrowth
ARHGEF9	Cdc42 guanine nucleotide exchange factor; pivotal role in formation of postsynaptic glycine and GABA(A) receptor clusters
ARX[a]	Transcription factor with possible role in the maintenance of specific neuronal subtypes in the cerebral cortex and axonal guidance in the floorplate. Neuronal proliferation, differentiation of GABA-ergic neurons
ATRX[a]	DNA-binding helicase, involved in chromatin remodelling, DNA methylation and regulation of gene expression; intrinsic regulator of cortical size
DLG3	Post-synaptic scaffolding protein linked to NMDA-type glutamatergic receptors
FGD1[a]	RhoGEF; possible role in stimulation of neurite outgrowth
FMR2	Transcriptional regulator; possible role in long-term memory and enhanced long-term potentiation
FTSJ1	RNA methyltransferase, possible role in tRNA modification and translation
GDI1	Regulation of synaptosomal Rab4 and Rab5 pools; possible role in endocytosis
GRIA3	AMPA receptor GLUR3; mediates fast, synaptic transmission in central nervous system
IL1RAPL	Regulator of dense-core granule exocytosis; possible modulator of neurotransmitter release
JARID1C[a] **(SMCX)**	Role in chromatin remodelling
MECP2[a]	Transcriptional silencer of neuronal genes, role in splicing
NLGN4[a]	Postsynaptic membrane protein; involved in induction of presynaptic structures; linked to NMDA-type glutamatergic receptors
PAK3	Regulation of actin cytoskeleton; stimulation of neurite outgrowth
PQBP1[a]	Polyglutamine-binding; putative role in transcription and mRNA splicing
RPS6KA3 (RSK2)[a]	Serine–threonine protein kinase; CREB phosphorylation; role in long-term memory formation
SLC6A8[a]	Creatine transporter; required for maintenance of (phospho)creatine pools in the brain
TM4SF2	Modulation of integrin-mediated signalling; neurite outgrowth; possible role in synapse formation
ZNF41	KRAB domain-containing zinc finger protein; putative transcriptional regulator; possible involvement in chromatin remodelling
ZNF674	KRAB domain-containing zinc finger protein; related to ZNF41 and ZNF81
ZNF81	KRAB domain-containing zinc finger protein; related to ZNF41 and ZNF674

[a] Also mutated in S-XLMR.

1.1.2 Autosomal Mental Retardation (AMR)

So far, very little is known about the role of autosomal genes in MR. The reason for this is that severe autosomal dominant forms of MR (ADMR) manifest nearly always as sporadic cases since affected patients rarely reproduce. Consequently, the prevalence of autosomal dominant disease genes is dependent on the new mutation rate.

The identification of novel dominant genes has so far relied on the collecting of individuals with MR and distinct dysmorphic features in order to analyse and group together (Raymond and Tarpey 2006). However, this strategy does not work in patients without syndromic features. This is why the elucidation of autosomal dominant forms of NS-MR has lagged behind.

Functional considerations argue for autosomal recessive MR (ARMR) to be more common than ADMR, and there is reason to believe that most of the patients with 'idiopathic' MR carry autosomal recessive gene defects (Bartley and Hall 1978). However, in developed countries, due to small family sizes, most patients with ARMR appear as isolated cases, too. This is the reason why the identification of autosomal recessive disease-causing mutations has been disproportionately slow. Particularly the scarcity of large pedigrees with consanguineous marriages (which are a prerequisite for autozygosity mapping) in developed countries has hitherto hampered the identification of ARMR genes and precluded successful mapping and identification of candidate loci. In this project, however we have overcome these obstacles by investigating highly consanguineous families from Iran, where close to 40% of all children are born to consanguineous parents and families are large, as reflected by the fact that 70% of the population are below the age of 30.

Syndromic- Autosomal Recessive Mental Retardation (S-ARMR)

Quite a large number of genes has been found for syndromic forms of ARMR so far (for review see Raymond and Tarpey 2006) including several gene defects that give rise to MR with microcephaly.

Microcephaly

Microcephaly is defined by a head circumference (HC) that is more than 3 standard deviations below the age- and sex- related population mean (Ross 1977; Ross and Frias 1977). It is generally the result of perturbed neurodevelopment, which in turn leads to a disproportionate reduction in the size of the cerebral cortex. HC is a surrogate measurement of brain size as it is only imperfectly correlated with brain volume; still, it remains the most common, simple method for evaluating gross brain size (Woods and others 2005).

Microcephaly is divided into ***primary microcephaly*** (MCPH, microcephaly vera), which is present at birth, and ***secondary microcephaly,*** which develops postnatally (Woods 2004). MCPH is usually a static developmental anomaly, whereas secondary microcephaly indicates a progressive neurodegenerative condition (Woods and others 2005). Thus, MCPH is a distinct entity, further defined by the absence of other malformations or significant neurological deficits and inherited as an autosomal recessive trait (Bundey 1992). The gyral pattern is relatively normal and cortical architecture is well preserved (McCreary and others 1996; Mochida and Walsh 2001), which may explain why the only significant neurological deficit in this disorder is that of reduced cognitive abilities (Bundey 1992).

MCPH phenotype definition and clinical features

The current common clinical characteristics of MCPH are as follows:

- HC is at least three standard deviations (SD) below the age- and sex-adjusted mean and is evident at birth.
- HC usually does not vary by > 2SD between affected individuals of the same family and throughout life degree of microcephaly does not change.
- Microcephalic patients have MR from borderline (Trimborn and others 2005) to severe but no other neurological symptomes such as spasticity or progressive cognitive decline.
- Height, weight, appearance, chromosome analysis and brain scan are normal in the majority of individuals with MCPH.
- Specifically for patients with *MCPH1* mutations, cytogenetic analysis reveals an increased proportion of prophase-like cells. A reduction in height can occur, but the HC is always significantly more reduced than height. On MRI scans, some patients show evidence of periventricular neuronal heterotopias, which is suggestive of neuronal migration defects (Cox and others 2006).

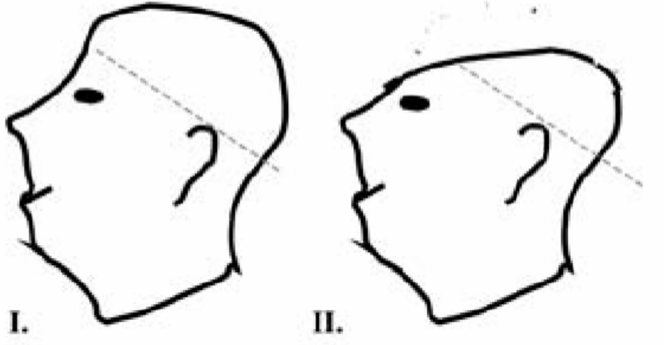


Figure 1-1: A pictorial representation of microcephaly. The line drawings show a normal individual (i) alongside a microcephalic individual (ii). Note that the reduced occipitofrontal circumference (depicted by the dashed line) of the microcephalic individual results in a sloping forehead and apparent protrusion of the face (from O'Driscoll and others 2006).

Etiology of MCPH

Etiologically, the reduction in brain size is likely to reflect a reduction in the number of neural cells generated during neurogenesis, either as a consequence of reduced proliferation or increased cell death (Mochida and Walsh 2001).

All four known MCPH genes (see below) are expressed in the neuroepithelium. Neuroepithelial cells are the primary neural progenitors from which all other central nervous system (CNS) progenitors and — directly or indirectly — all CNS neurons derive. It is likely that the MCPH genes are affecting the neurogenic mitosis either by controlling the expansion of the neural progenitor pool or involvement in the decision to switch from symmetric to asymmetric cell division (for more information see the Cox and others 2006 or discussion). Taken together this suggests that MCPH is a primary disorder of neurogenic mitosis and not one of neural migration, neural apoptosis, or neural function (Woods and others 2005).

MCPH Genes

Up to now six autosomal recessive loci (MCPH1–MCPH6) are known (Jackson and others 1998; Jamieson and others 2000; Leal and others 2003; Moynihan and others 2000; Pattison and others 2000; Roberts and others 1999) and the causative genes at four loci could be identified (Table 1-2). These are *ASPM* (Bond and others 2002), *CDK5RAP2* (Bond and others 2005), *CENPJ* (Bond and others 2005) and *MCPH1* (Microcephalin; Jackson and others 2002).

Table 1-2: An overview of primary microcephaly genes and loci (O'Driscoll and others 2006).

Locus	Clinical Features	Gene	Functional Domains	Cellular features & proposed/demonstrated function
MCPH1	Microcephaly Growth retardation	Microcephalin (BRIT1)	3 BRCT domains (C-terminal tandem)	Defective DNA damage response and cell cycle regulation. Patient cells exhibit premature chromosome condensation (PCC). Supernumerary mitotic centrosomes.
MCPH2	Primary microcephaly	Unknown		
MCPH3	Primary microcephaly	CDK5RAP2	Spindle association domain	Centrosome organization, specifically in neurons. Homology of fly centrosomin.
MCPH4	Primary microcephaly	Unknown		
MCPH5	Primary microcephaly	ASPM	2 calponin homology domains. 81 IQ repeats. ASH domain	Spindle microtubule nucleation at centrosome.
MCPH6	Primary microcephaly	CENPJ	Tep 10 domain	Centrosome association via gamma-tubulin. ortholog of fly and worm SAS4, a protein required for centriole replication.

ASPM (Abnormal Spindle-like Microcephaly-associated gene)

Mutations in *ASPM* are the most common cause of the MCPH phenotype (Bond and others 2002; Kumar and others 2004). *ASPM* is the orthologue of the Drosophila *asp* (abnormal spindle) gene. It spans ~63 kb of human genomic DNA and contains 28 exons with a

10,4 kb open reading frame (Bond and others 2002). More than 25 mutations distributed throughout the *ASPM* gene have been reported so far.

Investigations of the fetal expression pattern of the murine *Aspm* gene in the mouse brain have demonstrated that its expression is maximal in the sites of active neurogenesis and is downregulated when neurogenesis is complete, indicating that Aspm is involved in neuron production (Bond and others 2002).

Species comparison of the 3477 amino acid ASPM protein by bioinformatic means indicates that it contains one N-terminal microtubule-binding domain, two calponin homology domains (common in actin-binding proteins), 81 Ile–Gln repeat motifs, which are predicted to undergo a conformational change when bound to calmodulin, and a C-terminal region of unknown function (Bond and others 2002; Kouprina and others 2005; Rhoads and Kenguele 2005). Structural projections suggest that ASPM directly interacts with the intracellular cytoskeleton and assumes a semi-rigid rod conformation upon interactions with multiple calmodulin molecules.

ASPM mutations in humans produce a mitotic defect specific to the brain. In Drosophila, larvae with asp mutations are stillborn or infertile with dividing neuron progenitors that are unable to conclude asymmetric cell division (Ponting 2006). As ASPM is required in microtubule organization of the mitotic spindle poles and the central spindle in meiosis and mitosis, it can be hypothesized that during neurogenesis, ASPM organizes microtubules at the spindle pole during mitosis and at the central spindle during cytokinesis (Cox and others 2006).

CDK5RAP2 (Cyclin-Dependent Kinase 5 Regulatory Associated Protein 2)

Mutations in *CDK5RAP2* which spans 191kb of genomic DNA and includes 34 exons are a rare cause of MCPH. The CDK5RAP2 protein is known as a centrosomal component which localizes to the centrosomes during interphase and to the spindle poles during mitosis (Bond and others 2002; Hung and others 2004).

Centrosomin (*cnn*) is the probable Drosophila ortholog of *CDK5RAP2*. Cnn interacts with the gamma-tubulin ring complexes within the centrosome, which are responsible for the production of the microtubules that form the mitotic spindle (Terada and others 2003). Drosophila cnn mutants exhibit a gross reduction in cell number in the central and peripheral nervous system (Li and Kaufman 1996). Therefore, it is hypothesized that CDK5RAP2 affects neurogenic mitosis by reducing the availability of the microtubules that are needed to build the mitotic spindle and astral microtubule network (Cox and others 2006).

CENPJ (Centromere-associated protein J)

CENPJ has 17 exons and spans ~41 kb of genomic DNA. Like CDK5RAP2 it is also associated with the gamma-tubulin ring complex, and *in vitro* evidence suggests that CENPJ might inhibit microtubule nucleation and depolymerise microtubules (Hung and others 2004). Therefore, similar to CDK5RAP2, CENPJ might have a role in the control of centrosomal microtubule production during neurogenic mitosis (Bond and others 2005). An additional function in centriol formation for CENPJ is suggested by findings from *Caenorhabditis elegans* where (Leidel and Gonczy 2005) showed that one of the only five proteins that are essential for centriole duplication in *C. elegans* is encoded by the *SAS-4* gene, which is the probable homologue of *CENPJ*. Moreover, RNAi-mediated knock-down of *CENPJ* arrests all cells during mitosis, and many of them have multi-polar spindles (Cho and others 2006).

MCPH1 (microcephalin)

The *MCPH1*/*Microcephalin* gene is a 14-exon gene that encodes an 835-amino acid protein on chromosome 8p23 (Figure 1-2). An MCPH-causing homozygous trunicating mutation (74C>G; S25X in exon 2) of the *MCPH1*gene was found for the first time by positional cloning in two consanguineous Pakistani families with an ancestral common haplotype, and the previously uncharacterised MCPH1 protein was then named microcephalin (Jackson and others 2002).

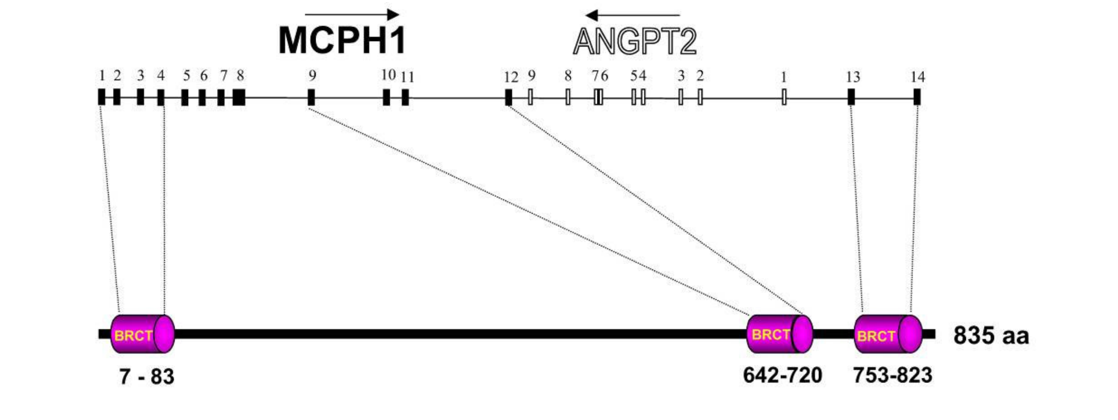


Figure 1-2: *MCPH1's* exon organization and approximate coding regions for its 3 BRCT domains of *MCPH1* (exons are represented by *black boxes*) and *ANGPT2* (exons represented *by empty boxes*). *Arrows* indicate the orientation of the genes.

It has been shown that *MCPH1* is expressed in fetal mouse brain during the period of neurogenesis, with highest expression in the ganglionic eminences and lateral ventricles, from which the neurons of the cerebral cortex are generated. However, it is also expressed at similar levels in many other fetal tissues such as brain, liver and kidney and, at lower levels, in other tissues (Jackson and others 2002).

Additionally, patients with *MCPH1* mutations have a unique cellular phenotype with premature chromosome condensation in early G2 phase and delayed decondensation after mitosis. The full length MCPH1-encoded protein microcephalin has one N-terminal and two C-terminal BRCT-domains (BRCA1 C-terminal). BRCT-domains are predominantly found in proteins involved in cell cycle checkpoint control and DNA repair. Moreover, microcephalin was identified in a genetic screen for transcriptional repressors of hTERT, the catalytic subunit of human telomerase, from which derives its alternative name BRIT1 (BRCT-repeat inhibitor of hTERT expression). Therefore, it was speculated that microcephalin might play a role in DNA damage response and checkpoint control, in addition to its role in chromosome condensation and delayed decondensation post-mitosis. The findings so far implicate microcephalin/BRIT1 as a novel regulator of chromosome condensation and link the apparently disparate fields of neurogenesis and chromosome biology (Jackson and others 2002; Trimborn and others 2004; Xu and others 2004).

Non-Syndromic Autosomal Recessive Mental Retardation (NS-ARMR)

As mentioned earlier, relatively little is known about the molecular causes of NS-ARMR. Until 2004, only one gene (*PRSS12*) was known to be directly linked to NS-ARMR (Molinari and others 2002). Since then, three more genes, *CRBN* (cereblon [MIM 609262] ; Higgins and others 2004), *CC2D1A* [MIM 610055] (Basel-Vanagaite and others 2005) and *GRIK2* (Motazacker and others 2007) have been identified; all of these were found by autozygosity mapping in highly consanguineous families.

PRSS12 (Neurotrypsin)

PRSS12 is located on chromosome 4q26 and was found in a large consanguineous Algerian family with four severely mentally retarded children. It has 13 exons, encompasses ~71 kb of genomic DNA and encodes the neuronal serine protease “neurotrypsin” (motopsin).

Neurotrypsin is highly expressed in the cerebral cortex, the hippocampus and the amygdala. Within neuronal cells, it is localized in the pre-synaptic membrane and the pre-synaptic active zone of both excitatory and inhibitory synapses.

In the MR patients, a four base-pair truncating mutation was found that results in a shortened protein, lacking a catalytic domain.

Until the age of 1.5 years, psychomotor development in the affected members of this family was normal. Only thereafter, they showed cognitive deterioration leading to severe MR.

This late-onset manifestation may indicate that neurotrypsin plays a role in adaptive synaptic function, such as synapse reorganization during later stages of neurodevelopment and adult synaptic plasticity rather than in the formation of synapses (Molinari and others 2002).

Reduced 24-h long-term memory but not short-term memory loss has been shown for mutant Drosophila flies lacking neurotrypsin orthologue (Didelot and others 2006).

CRBN (Cereblon)

The second gene found to be responsible for NS-ARMR is CRBN on chromosome 3p26.2. It was identified in a family with 10 affected individuals originating from Germany. The IQs of the patients range from 50 to 70 and are lower in men than in women (Higgins and others 2004). In contrast to the patients with neurotrypsin mutations, these patients were developmentally delayed from early childhood on. Their homozygous nonsense mutation in the CRBN gene, R419X, causes premature truncation of the protein. *CRBN* is a 11-exon gene which spans ~30kb of genomic DNA and encodes cereblon, a member of an Adenosine 5′- triphosphate-dependent Lon protease gene family coding for multi-domain enzymes that are associated with diverse functions, from proteolysis to membrane trafficking (Jo and others 2005). It has been shown that cereblon is directly associated with large conductance Ca^{2+}-activated K1 (BK_{Ca}) channels (Rotanova and others 2006), which are important in the control of neuronal excitability and transmitter release (Faber and Sah 2003).

Overexpression of BK_{Ca} channel was shown recently to cause impairment of learning and memory in hippocampal-dependent tasks (Hammond and others 2006). Therefore, assembly and surface expression of functional BK_{Ca} channels might be important in controlling human cognition.

CC2D1A (coiled-coil and C2 domain containing 1A)

The third gene causing NS-ARMR is *CC2D1A* on chromosome 19p13, which was found mutated in a large extended family of nine consanguineous branches with severe MR. The patients present with developmental delay during early childhood and are unable to speak a single word.

All affected individuals carry a large genomic deletion of 3589 nucleotides in *CC2D1A* which leads to a truncated protein (G408fsX437).

CC2D1A is highly expressed in the embryonic ventricular zone and developing cortical plate in mouse embryos, persisting into adulthood with highest expression in the cerebral cortex and hippocampus (Basel-Vanagaite and others 2006).

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 and others 2003).

Recent studies have revealed that Freud-1 also binds to an intronic repressor element in the dopamine receptor D2 gene. Both receptors function as pre-synaptic autoreceptors regulating the neurotransmission of serotonin and dopamine, respectively, and have a role in memory and behaviour (Ropers 2008).

Therefore, disruption of the CC2D1A protein is expected to cause MR via 5-HT1A serotonin and dopamin receptors.

GRIK2 (ionotropic glutamate receptor 6; *GLUR6*)

The fourth gene for NS-ARMR is *GRIK2.* It was found in a large consanguineous Iranian family with moderate to severe MR in a study which was conducted in parallel to the one presented here. *GRIK2* is a 17-exon gene spanning ~671 kbp of genomic DNA on chromosome 6. *GRIK2* encodes GLU_{K6}, a subunit of kainate receptors that is highly expressed in the brain.

A complex deletion-inversion mutation (~120 kbp deletion spanning exons 7 and 8 and an inversion of ~80 kbp, including exons 9, 10, and 11 in combination with another deletion of ~20 kb of intron 11) was found which results in an in-frame deletion of 84 aa between amino acids 317 and 402, close to the first ligand binding domain (S1) in the extracellular N-terminal region of GLU_{K6} (Motazacker and others 2007).

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 (see appendix for a copy of the article by Motazacker and others 2007).

1.2 Aim of the study

The aim of this study was to shed more light on the molecular background of autosomal recessive forms of MR. To this end, autozygosity mapping and mutation screening was performed in a large number of consanguineous families from Iran, followed by research into structural and functional properties of mutated genes and their products. These studies have also provided first data on the genetic heterogeneity of ARMR.

2 Material and methods

2.1 Materials

2.1.1 General reagents

Table 2-1: Chemicals

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
Magnesium chloride	Merck
Methanol	Merck
Milk powder	Protifar
Northern blot hybridization buffer	Ambion
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

2.1.2 Kit and Markers:

Table 2-2Kit 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
FirstChoice® RLM-RACE Kit	Ambion
GeneChip® Human Mapping 10K (Xba131 and 142) Array and Assay Kit	Affymetrix
GeneChip® Human Mapping 250K (nsp) Array and Assay Kit	Affymetrix
GeneChip® Human Mapping 50K (xba 240) Array and Assay Kit	Affymetrix
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
SMART™ RACE cDNA Amplification Kit	Clontech
Superscript ^{TM} II Reverse Transcriptase	Invitrogen
Superscript ^{TM} III Reverse Transcriptase	Invitrogen
TaKaRa LA PCR kit ver. 2.1	CHEMICON
Taq PCR core kit	Qiagen
Western Lightning Chemiluminescence Reagent, NL100	PerkinElmer

2.1.3 Enzymes

Table 2-3: Enzymes	Manufacturer
DNA polymerase 1, Klenow fragment	USB
Gateway LR Clonase Enzyme Mix	Invitrogen
Hpy 188 I	Biolabs
HpyCH4 III	Biolabs
PowerScript TM Reverse Transcriptase	Clontech
Proteinase K	Fermentas
Rnase-Free DNase	Promega
RNAsin	Promega
SuperTaq TM Plus	Ambion

2.1.4 Instruments

Table 2-4: 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
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

2.1.5 Consumables

Table 2-5: Consumables (disposable materials)	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

Table 2-5: Consumables (disposable materials)	Supplier
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

2.1.6 Software

Table 2-6: Software	
Name	**Source**
Allegro	http://www.decode.com/software/allegro
Alohomora	http://gmc.mdc-berlin.de/alohomora/
BeadStudio	http://www.illumina.com/
BioEdit v.7.0.5.2	http://www.mbio.ncsu.edu/BioEdit/bioedit.html
CodonCode	http://www.codoncode.com/
Cyrillic 2.1.3	http://www.cyrillicsoftware.com/
EasyLinkage v5.05	http://sourceforge.net/projects/easylinkage/
Endeavour v. 1. 39	http://www.esat.kuleuven.be/endeavour
FastLink	http://www.cs.rice.edu/~schaffer/fastlink.html.
GCG package	http://www.accelrys.com/products/gcg/
Gene Runner v3,05	http://www.generunner.net/
GeneHunter	http://www.broad.mit.edu/ftp/distribution/software/genehunter/
Ghostscript 8.14	http://pages.cs.wisc.edu/~ghost/doc/AFPL/get814.htm
Gnuplot	http://www.gnuplot.info/
GRR	http://www.sph.umich.edu/csg/abecasis/GRR/
GSview 4.6	http://pages.cs.wisc.edu/~ghost/gsview/get46.htm
Haplopainter	http://haplopainter.sourceforge.net/html/ManualIndex.htm
ImageQuant 5.2	http://www.mdyn.com/
Merlin	http://www.sph.umich.edu/csg/abecasis/Merlin/download
PedCheck	http://watson.hgen.pitt.edu/register
Prioritizer v1.2	http://humgen.med.uu.nl/~lude/prioritizer/download.php
SDS 2.1	http://www.appliedbiosystems.com/
STADEN package	http://portal.litbio.org/Registered/Option/staden.html
SuperLink	http://bioinfo.cs.technion.ac.il/superlink/.
Swiss-PdbViewer 3.7	http://www.expasy.org/spdbv/

2.1.7 Bioinformatic databases and tools

Table 2-7: Data 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/
Fatigo+	http://babelomics.bioinfo.cipf.es/fatigoplus/cgi-bin/fatigoplus.cgi
GenBank	http://www.ncbi.nlm.nih.gov/Genbank/

Table 2-7: Data Bioinformatic databases and tools

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/

2.2 Patients and sampling

Families with a minimum of two mentally retarded children were identified through collaboration with local genetic counsellors in several provinces of Iran. Families whose pedigree patterns and clinical data seemed to be compatible with moderate to severe NS-ARMR were selected and visited by experienced clinical geneticists, or invited to the Genetics Research Centre in Tehran. Patients and unaffected relatives were examined in a standardized way using a questionnaire, and 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, 10 ml 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 and others 2006).

2.3 DNA extraction

DNA was extracted according to the standard salting out method by (Miller and others 1988).

2.4 Immortalized Cell line preparation

At least for one of the affected individuals in each family, an EBV Immortalized cell line was established by the central cell culture facility.

2.5 Fragile X Test

At least for one patient of each nuclear family, Fragile X testing was carried out by PCR and Southern blot analysis if X-linkage could not be excluded.

2.6 Metabolic disorders Test

Filter-dried blood of one patient per family was screened by tandem mass spectrometry to exclude disorders of the amino acid, fatty acid (e.g. phenylketonuria) or organic acid metabolism (Chace 2003; Wilcken 2004).

2.7 Karyotyping

At least for one affected individual in each family standard 450 G-band karyotyping was performed in order to exclude cytogenetically visible chromosomal aberrations.

2.8 Autozygosity Mapping

Autozygosity is a term describing homozygosity for markers that are identical by descent (IBD), i.e. inherited from a (recent) common ancestor. Individuals with a rare recessive disease in a consanguineous family are likely to be autozygous for markers linked to the disease locus. If the parents are second cousins they share 1/32 of all their genes because of their common ancestry, while their children will be autozygous at 1/64 of all loci. If a child is homozygous for a particular marker allele, this can be because of autozygosity, or because a second copy of the same allele has entered the family independently. The rarer the allele is in the population, the greater the likelihood that homozygosity represents autozygosity (Strachan T 2003).

Autozygosity mapping involves locating a gene that causes a rare recessive trait by using multipoint linkage analysis to find regions of IBD that are shared among inbred affected children. The method is particularly powerful because it does not require families with multiple affected individuals but, rather, requires only unrelated affected singletons from consanguineous marriages. Even small families with multiple affected members will yield to significant results: in principle, four offspring from a first-cousin marriage suffice to obtain a LOD score of 3.0 (Kruglyak and others 1995).

Wide SNP-arrays offer many advantages over previous genotyping methods aimed to define recessive loci. In kindreds with an apparently recessive disorder, particularly such families where parental consanguinity is suspected, this approach can be used to map regions of extended homozygosity with high resolution and essentially complete genomic coverage (depending of the panel of choice).

Because all tracts of disease segregating homozygosity will be identified and all heterozygous regions/non-segregating homozygous tracts excluded, one can be confident that the region harbouring the genetic lesion underlying disease has been identified; with the caveat that the model needs to be correct (i.e., the disease must be caused by a homozygous change inherited from a relatively recent common ancestor).

The size of these regions depends on several factors:

- The degree of parental consanguinity.
- The number of informative family members.
- The relatively stochastic distribution of recombination events.

In small families where there are fewer meioses and thus more chance for variation in the length and number of homozygous tracts, the power of autozygosity mapping will generally be lower.

In populations with a low level of inbreeding, or where there is a high degree of separation between affected family members, the resolution of this kind of analysis is likely to be high, and genome-wide SNP typping will offer a great advantage over traditional genome-wide linkage analysis.

2.9 SNP Genotyping Methods

The recent discovery of millions of SNPs in the human genome and the development of DNA arrays allowing to type more than 10^6 SNPs in a single experiment has made it possible to use them as a useful linkage analysis tool in order to investigate the genetic causes of human diseases (Collins and others 1997; Matise and others 2003).

Affymetrix (Kennedy and others 2003; Matsuzaki and others 2004) and Illumina are offering different panels of markers.

In addition to speed and resolution, genome-wide SNP-assay offers one more advantage in autozygosity mapping: this technique allows the direct visualization of structural genetic mutation such as genomic deletion and duplication, which often underlie recessive disorders (Gibbs and Singleton 2006).

After having adjusted DNA concentrations and check its quality on agarose gels, whole genome SNP Genotyping was performed using different versions of the Affymetrix GeneChip® Human Mapping Array (10k, 50K or 250k) or the Illumina GoldenGate™ Assay 6k panel.

For most of the families the 10k SNP array (Affymetrix Xba142) was used. For a smaller number, panels with higher density like the Affymetrix 50k (Xba240) or 250k (Nsp) arrays were used.

Hybridizations were performed commercially at the Max-Delbrück-Centrum für Molekulare Medizin (MDC), the "Deutsches Ressourcenzentrum für Genomforschung (RZDPD)", the microarray facility in Tübingen or the "ATLAS Biolabs GmbH".

Several families were analyzed with the Illumina linkage IVb panel at the core facility of the Max-Planck Institute for Molecular Genetics.

2.9.1 Affymetrix GeneChip® Human Mapping Array

The Affymetrix GeneChip array is a high-throughput genotyping platform that uses a one-primer assay to genotype a large number of SNPs per individual on a single oligonucleotide array. The approach uses restriction digestion to fractionate the genome, followed by amplification of a specific subset of the genome containing the markers to be genotyped. The resulting reduction in genome complexity enables allele-specific hybridization to the array.

The selection of SNPs is primarily determined by computer-predicted lengths of restriction fragments containing the SNPs, and is further driven by strict empirical measurements of accuracy, reproducibility, and average call rate (Matsuzaki and others 2004).

Principles of Allele-Specific Hybridization

Allele-specific hybridization (ASH) is a way to distinguish allelic variants at the DNA level. By synthesizing probes on the array that are complementary to each of the two possible alleles at each SNP and hybridizing the target DNA to the array, it is possible to determine whether a SNP is heterozygous or homozygous for the different alleles (AB, AA, or BB) by analyzing the resulting signals from the allele-specific probes. 25-mer probes perfectly matching the A allele sequence (PMA) and the B allele sequence (PMB) are synthesized. To determine specificity in binding, a 25-mer with a single base pair mismatch at the center position for each allele (MMA and MMB) is included. To increase sensitivity, Affymetrix chips carry 40 different binding oligonucleotides, for each SNP.

Complexity Reduction Assay

Total genomic DNA is digested with an approprite restriction enzyme (depending on the type of the panel) and ligated to adaptors recognizing the cohesive four base overhangs. All fragments resulting from restriction enzyme digestion, regardless of size, are substrates for adaptor ligation. A generic primer, which recognizes the adaptor sequence, is used to amplify ligated DNA fragments, and PCR conditions are optimized to preferentially amplify fragments in the 250-1000 bp size range. The amplified DNA is then □abelled and hybridized to the GeneChip arrays. The arrays are washed and stained on a GeneChip fluidics station and scanned on a GeneChip scanner Figure 2-1 (10K GeneChip® Mapping Assay Manual).

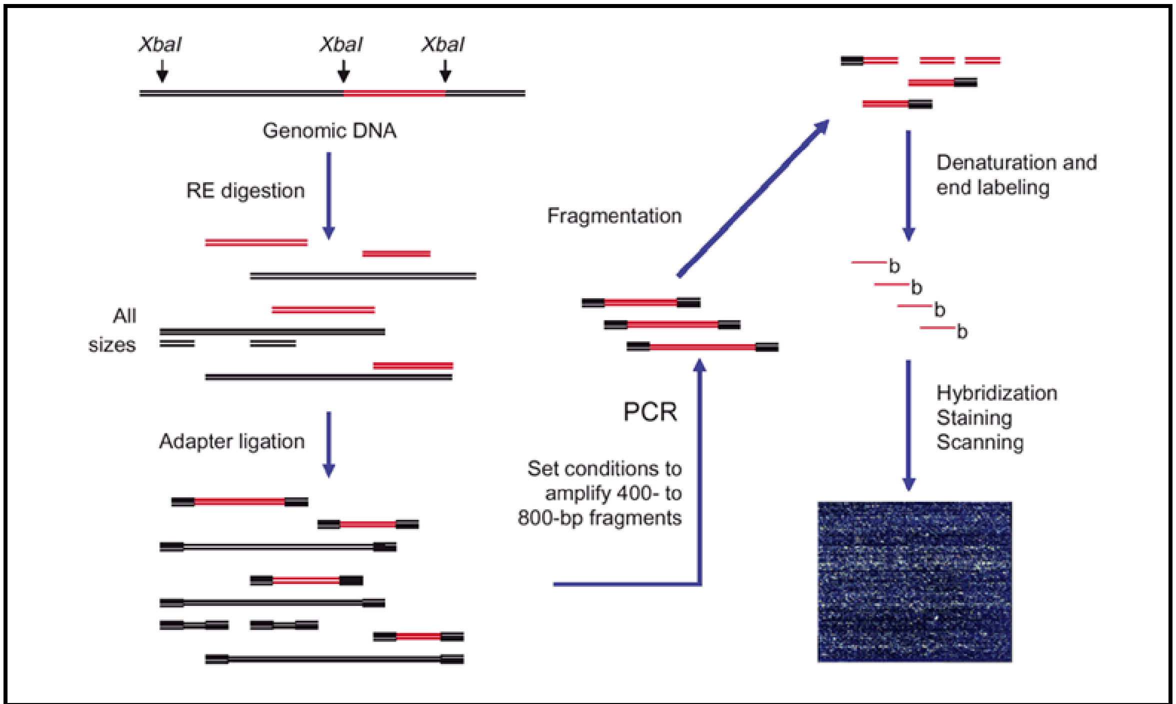


Figure 2-1: Principles of the Complexity Reduction Assay for an Affymetrix 10K array (from Affymetrix GeneChip human 10k array manual).

This assay was first developed for simultaneous genotyping of over 10,000 SNPs on a single array (GeneChip® Human Mapping 10K Array Xba 142 2.0). Later on, changing the choice of the restriction enzymes used and increasing the capacity of the high-density arrays rendered genotyping of up to 1 million SNPs possible (Manual 2005–2006 Affymetrix Inc.).

The arrays are scanned using GeneChip® Operating Software (GCOS). The resulting image file (the .dat file) is then displayed and a grid is applied to the image resulting in the automatic generation of a ".cel file". These file contains the averaged image data and is required for analysis with the GeneChip® DNA Analysis Software (GDAS). Following the generation of the .cel files, GDAS can immediately be used to generate genotyping calls (10K GeneChip® Mapping Assay Manual).

2.9.2 Illumina's GoldenGate Assay

Bead arrays

Illumina has developed a novel bead array technology. A multicore optical 'imaging' fiber is etched such that a single bead can fit into the resulting micron-sized etched wells on the tip of the fiber. Different oligonucleotide sequences are attached to each bead, and thousands of beads can be self-assembled on the fiber bundle. A subsequent decoding process is carried out to determine which bead occupies which well. Complementary oligonucleotides present in the sample bind to the beads, and bound oligonucleotides are measured by using a fluorescent label.

GoldenGate Genotyping Assay

The Illumina GoldenGate Genotyping Assay utilizes a discriminatory DNA polymerase to assay from 384 to 1536 loci simultaneously. Different steps of this assay are shown in Figure 2-3.

In this method three oligonucleotides are designed for each SNP locus. Two oligos are specific for the different alleles of the SNP site, called the Allele-Specific Oligos (ASOs). A third oligo that hybridizes several bases downstream from the SNP site is the Locus-Specific Oligo (LSO). All three oligonucleotide sequences contain regions of genomic complementarity and universal PCR primer sites; the LSO also contains a unique address sequence that targets a particular bead type. Up to 1,536 SNPs may be interrogated simultaneously in this manner. During the primer hybridization process, the assay oligonucleotides hybridize to the genomic DNA sample bound to paramagnetic particles. Because hybridization occurs prior to any amplification steps, no amplification bias can be introduced into the assay. Following hybridization, several wash steps are performed, reducing noise by removing excess and mis-hybridized oligonucleotides. Extension of the appropriate ASO and ligation of the extended product to the LSO, joins information about the genotype present at the SNP site to the address sequence on the LSO. These joined, full-length products provide a template for PCR using universal PCR primers P1, P2, and P3. Universal PCR primers P1 and P2 are Cy3- and Cy5-labeled. After downstream-processing the single-stranded, dye-labeled DNAs are hybridized to their complement bead type through their unique address sequences (see Figure 2-2, Illumina Technical Bulletin).

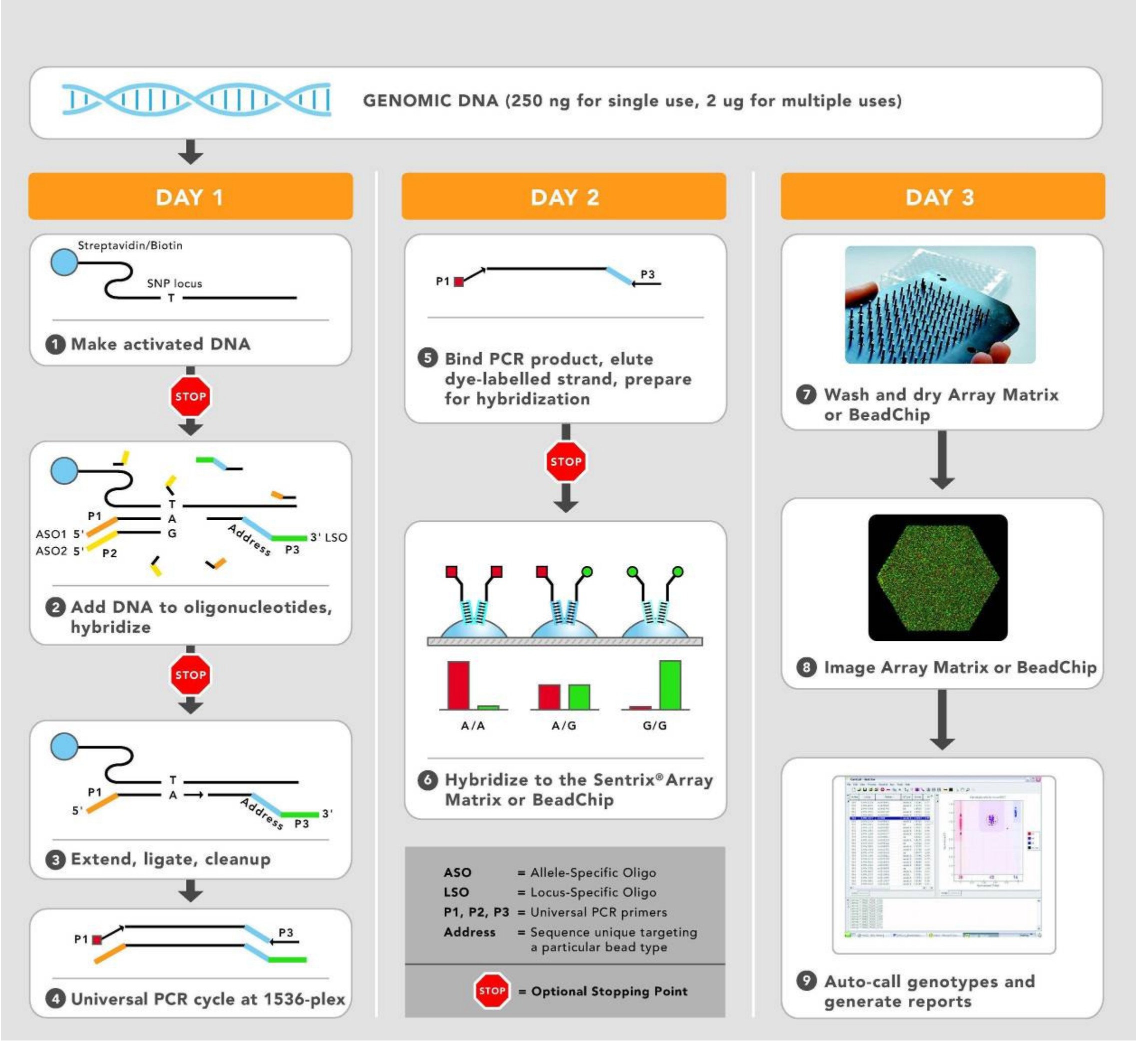

GENOMIC DNA (250 ng for single use, 2 ug for multiple uses)
DAY 1
DAY 2
DAY 3
Streptavidin/Biotin
SNP locus
T
1 Make activated DNA
STOP
P1
ASO1 5'
ASO2 5'
P2
T
A
G
Address
P3
3' LSO
2 Add DNA to oligonucleotides, hybridize
STOP
P1
5'
T
A
Address
P3
3'
3 Extend, ligate, cleanup
P1
P3
4 Universal PCR cycle at 1536-plex
P1
P3
5 Bind PCR product, elute dye-labelled strand, prepare for hybridization
STOP
A/A
A/G
G/G
6 Hybridize to the Sentrix® Array Matrix or BeadChip
ASO = Allele-Specific Oligo
LSO = Locus-Specific Oligo
P1, P2, P3 = Universal PCR primers
Address = Sequence unique targeting a particular bead type
STOP = Optional Stopping Point
7 Wash and dry Array Matrix or BeadChip
8 Image Array Matrix or BeadChip
9 Auto-call genotypes and generate reports

Figure 2-3: Different steps for performing the Illumina GoldenGate assay (from Illumina GoldenGate assay manual).

2.10 Data Conversion

The substantial number of markers on SNP arrays (>6000) raises a problem for data analysis, as most linkage programs were designed for the requirements of much smaller microsatellite marker sets (<1000). For instance, GeneHunter 2.1 is restricted to 300 markers and Simwalk2 (Davis and others 1997) to 31 only. In part, this can be overcome by using recompiled versions of the programs allowing for a higher maximum number of markers. Alternatively, the analysis may be performed with subsets of markers using a sliding window mode.

Therefore, the two following softwares were used for converting the genotyping data into appropriate linkage formats with appropriate subset sizes: ALOHOMORA (Ruschendorf and Nurnberg 2005) and/or EasyLinkage v5.05 (Hoffmann and Lindner 2005).

2.10.1. ALOHOMORA

The current version of ALOHOMORA accepts genotype data as generated by the GeneChip DNA Analysis Software (GDAS v3.0) from Affymetrix and the BeadStudio software from Illumina.

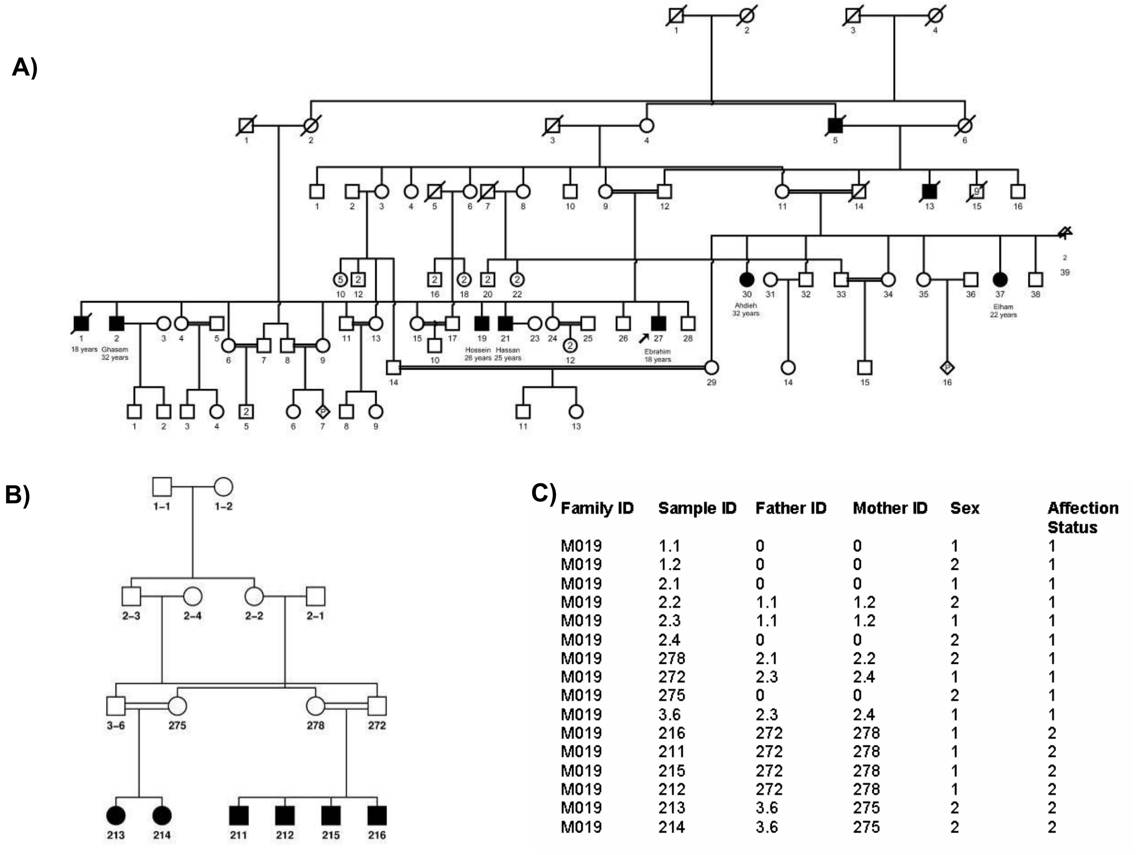
Pedigree and genotyping files were prepared to have the following column structures (see also Figure 2-4C):

Family ID Sample ID Father ID Mother ID Sex Affection

Person IDs have to match exactly to those in the marker files. Person / family IDs must be unique throughout the pedigree information file because otherwise, the files cannot be assembled appropriately prior to linkage analysis.

The pedigree file must have the name "pedfile.pro" and must be located in the same folder as the genotype file.

Considering the big size of families, in most of the cases simplified versions of the pedigrees with minimum numbers of individuals and the smallest possible loops were used. As an example, the simplified and original pedigree versions for family M019 are shown in Figure 2-4.



Family ID	Sample ID	Father ID	Mother ID	Sex	Affection Status
M019	1.1	0	0	1	1
M019	1.2	0	0	2	1
M019	2.1	0	0	1	1
M019	2.2	1.1	1.2	2	1
M019	2.3	1.1	1.2	1	1
M019	2.4	0	0	2	1
M019	278	2.1	2.2	2	1
M019	272	2.3	2.4	1	1
M019	275	0	0	2	1
M019	3.6	2.3	2.4	1	1
M019	216	272	278	1	2
M019	211	272	278	1	2
M019	215	272	278	1	2
M019	212	272	278	1	2
M019	213	3.6	275	2	2
M019	214	3.6	275	2	2

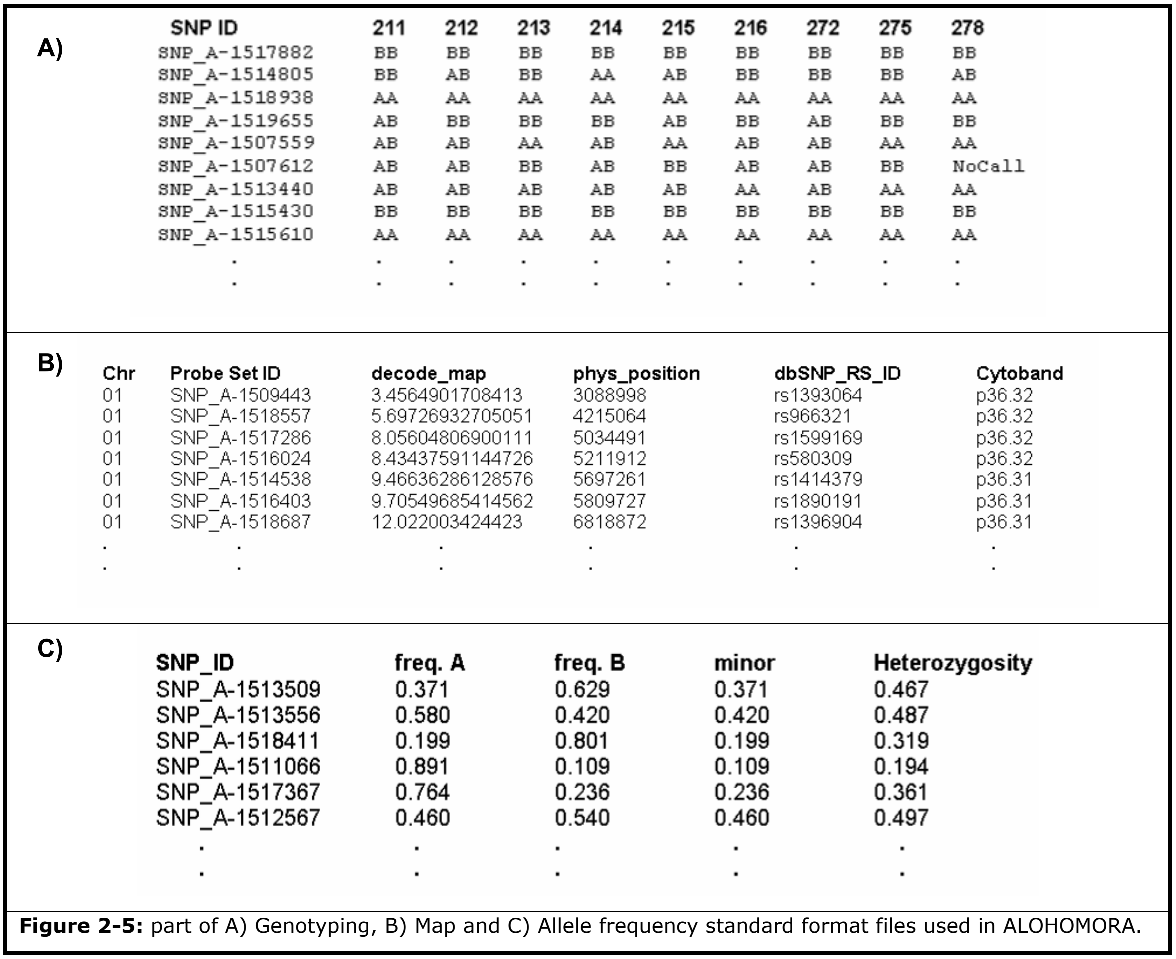
Figure 2-4: A) Original pedigree of family M019. B) Simplified version of the pedigree with maximum two consanguinity loops. C) Linkage format pedigree file.

The next requirement concerns the genotyping data file which has to have a header line describing the columns. The first column contains the SNP name and all other columns contain genotypes. The columns are tab-delimited.

```
SNP_Name    sample1     sample2     sample3     sample4
```

The genotypes must be coded A or AA for homozygotes for the first allele, B or BB for homozygotes for the second allele and AB for heterozygotes. The sample IDs in the columns header are either identical with the sample ID in the "pedfile.pro".

The preferred genetic map and the marker allele frequencies for the appropriate ethnicity have to be prepared. For our analysis we used the Caucasian population allele frequencies provided by Affymetrix and genetic marker map information provided by decode (Figure 2-5).



A)

SNP ID	211	212	213	214	215	216	272	275	278
SNP_A-1517882	BB	BB	BB	BB	BB	BB	BB	BB	BB
SNP_A-1514805	BB	AB	BB	AA	AB	BB	BB	BB	AB
SNP_A-1518938	AA	AA	AA	AA	AA	AA	AA	AA	AA
SNP_A-1519655	AB	BB	BB	BB	AB	BB	AB	BB	BB
SNP_A-1507559	AB	AB	AA	AB	AA	AB	AB	AA	AA
SNP_A-1507612	AB	AB	BB	AB	BB	AB	AB	BB	NoCall
SNP_A-1513440	AB	AB	AB	AB	AB	AA	AB	AA	AA
SNP_A-1515430	BB	BB	BB	BB	BB	BB	BB	BB	BB
SNP_A-1515610	AA	AA	AA	AA	AA	AA	AA	AA	AA
.	.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.	.

B)

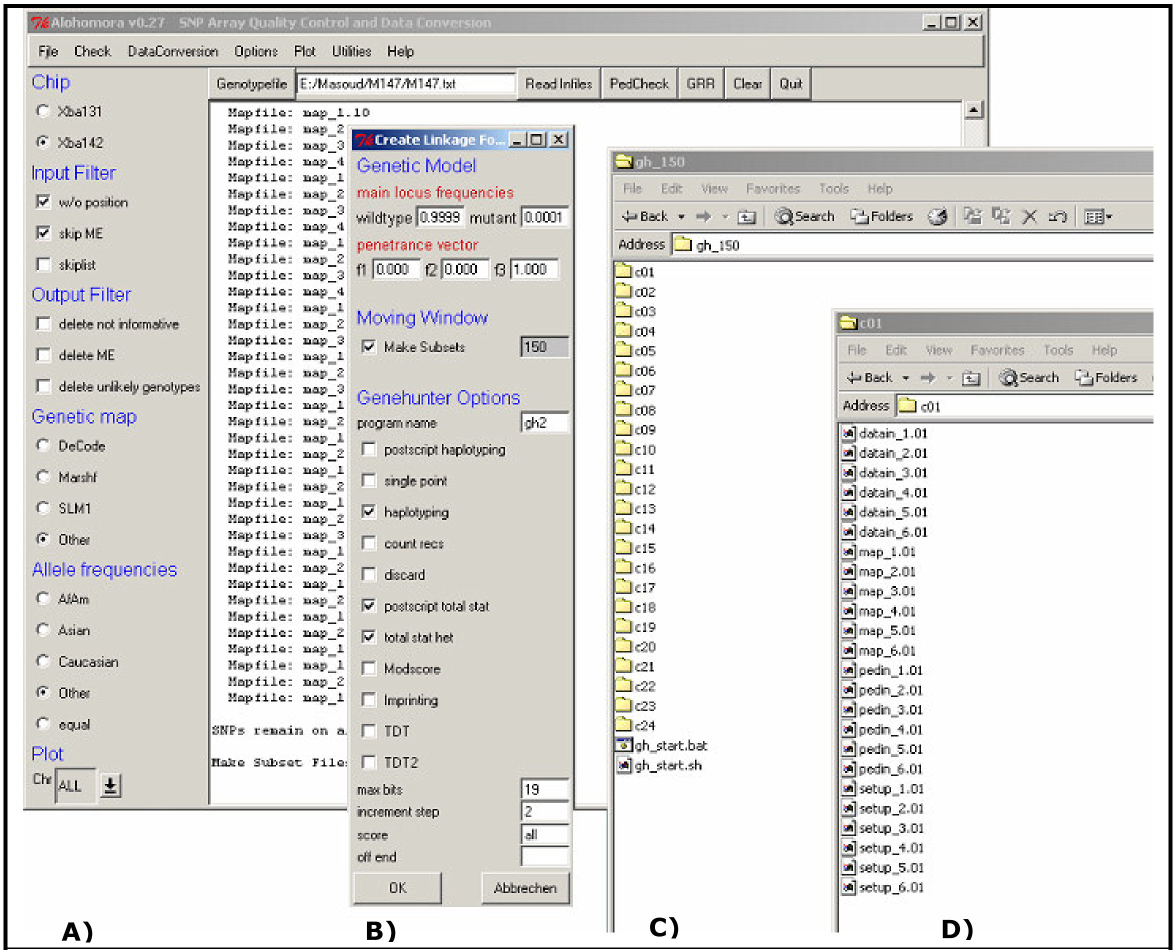
Chr	Probe Set ID	decode_map	phys_position	dbSNP_RS_ID	Cytoband
01	SNP_A-1509443	3.4564901708413	3088998	rs1393064	p36.32
01	SNP_A-1518557	5.69726932705051	4215064	rs966321	p36.32
01	SNP_A-1517286	8.05604806900111	5034491	rs1599169	p36.32
01	SNP_A-1516024	8.43437591144726	5211912	rs580309	p36.32
01	SNP_A-1514538	9.46636286128576	5697261	rs1414379	p36.31
01	SNP_A-1516403	9.70549685414562	5809727	rs1890191	p36.31
01	SNP_A-1518687	12.022003424423	6818872	rs1396904	p36.31
.	.	.	.	.	.
.	.	.	.	.	.

C)

SNP_ID	freq. A	freq. B	minor	Heterozygosity
SNP_A-1513509	0.371	0.629	0.371	0.467
SNP_A-1513556	0.580	0.420	0.420	0.487
SNP_A-1518411	0.199	0.801	0.199	0.319
SNP_A-1511066	0.891	0.109	0.109	0.194
SNP_A-1517367	0.764	0.236	0.236	0.361
SNP_A-1512567	0.460	0.540	0.460	0.497
.	.	.	.	.
.	.	.	.	.

Figure 2-5: part of A) Genotyping, B) Map and C) Allele frequency standard format files used in ALOHOMORA.

AOHOMORA creates a directory "name of the software _ size of subsets" for example "merlin_800" with subdirectories for each chromosome (c01, c02, …) and sets of linkage-format files with defined number of SNPs in each set (Figure 2-6).



Genotypefile E:/Masoud/M147/M147.txt

○ Xba131
● Xba142
☑ w/o position
☑ skip ME
☐ skiplist
☐ delete not informative
☐ delete ME
☐ delete unlikely genotypes
○ DeCode
○ Marshf
○ SLM1
● Other
○ AfAm
○ Asian
○ Caucasian
● Other
○ equal
Chr ALL

```
Mapfile: map_1.10
Mapfile: map_2
Mapfile: map_3
Mapfile: map_4
Mapfile: map_1
Mapfile: map_2
Mapfile: map_3
Mapfile: map_4
Mapfile: map_1
Mapfile: map_2
Mapfile: map_3
Mapfile: map_4
Mapfile: map_1
Mapfile: map_2
Mapfile: map_3
Mapfile: map_1
Mapfile: map_2
Mapfile: map_3
Mapfile: map_1
Mapfile: map_2
Mapfile: map_1
Mapfile: map_2
Mapfile: map_1
Mapfile: map_2
Mapfile: map_1
Mapfile: map_2
Mapfile: map_3
Mapfile: map_1
Mapfile: map_2
Mapfile: map_1
Mapfile: map_2
Mapfile: map_1
Mapfile: map_2
Mapfile: map_1
Mapfile: map_1
Mapfile: map_2
Mapfile: map_1

SNPs remain on a

Make Subset File:
```

wildtype 0.9999 mutant 0.0001
f1 0.000 f2 0.000 f3 1.000
☑ Make Subsets 150
program name gh2
☐ postscript haplotyping
☐ single point
☑ haplotyping
☐ count recs
☐ discard
☑ postscript total stat
☑ total stat het
☐ Modscore
☐ Imprinting
☐ TDT
☐ TDT2
max bits 19
increment step 2
score all
off end

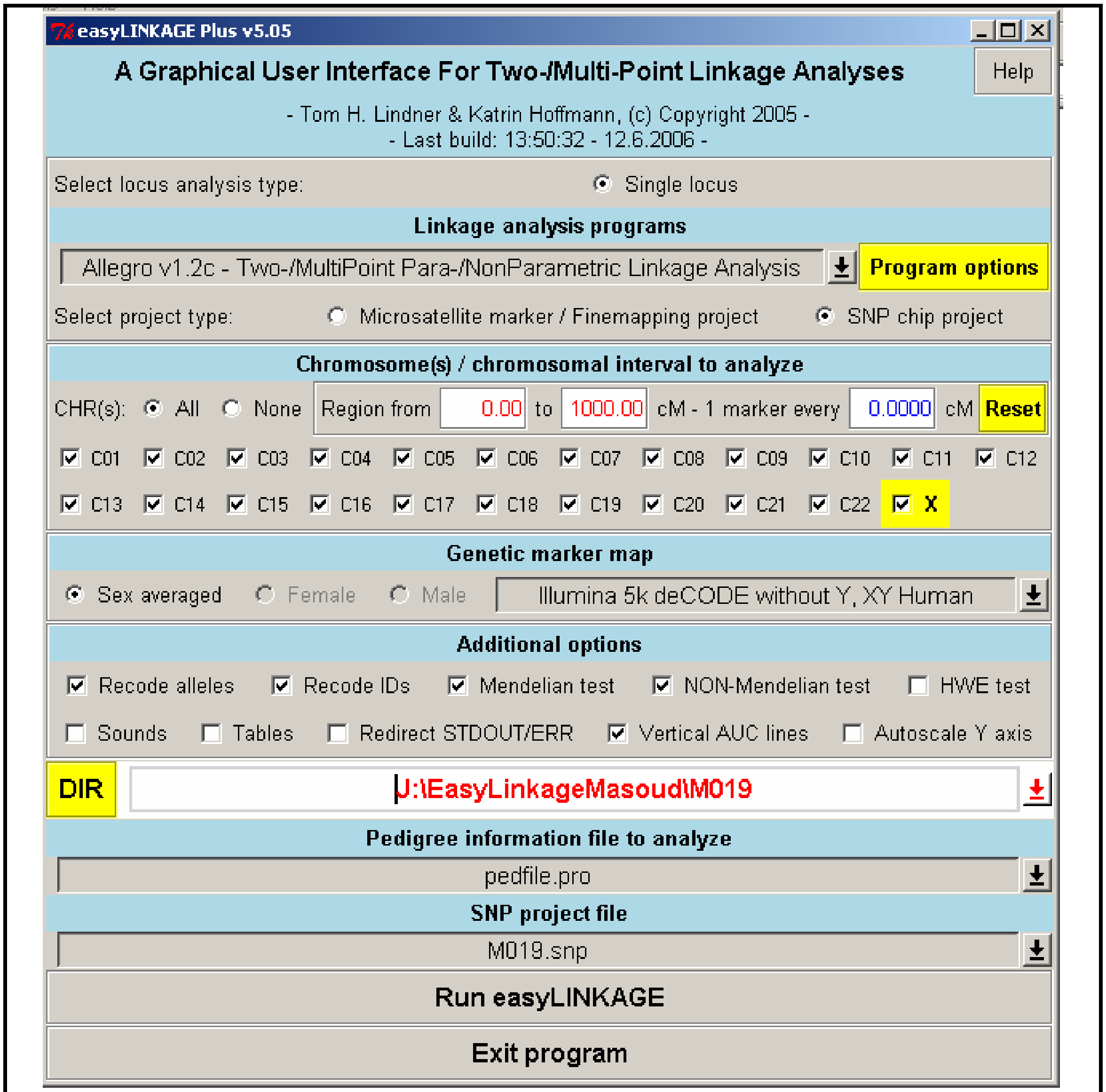
Address gh_150
c01 c02 c03 c04 c05 c06 c07 c08 c09 c10 c11 c12 c13 c14 c15 c16 c17 c18 c19 c20 c21 c22 c23 c24 gh_start.bat gh_start.sh

Address c01
datain_1.01 datain_2.01 datain_3.01 datain_4.01 datain_5.01 datain_6.01 map_1.01 map_2.01 map_3.01 map_4.01 map_5.01 map_6.01 pedin_1.01 pedin_2.01 pedin_3.01 pedin_4.01 pedin_5.01 pedin_6.01 setup_1.01 setup_2.01 setup_3.01 setup_4.01 setup_5.01 setup_6.01

Figure 2-6: A) ALOHOMORA main interface, B) GeneHunter menu: to choose size of subsets, haplotyping and several other options. C) Created directory with the name of the software and size of subsets (gh_150) with subdirectories for each chromosome (c01, c02, ...). D) Sets of linkage-format files with defined number of SNPs in each set.

2.10.2. EasyLinkage

The other software that was used for converting SNP data to linkage standard format files was EasyLinkage. The program package supports currently single-point linkage analyses (FastLink, SPLink, SuperLink), multi-point linkage analyses (GeneHunter, GeneHunter Plus, GeneHunter-Imprinting/-TwoLocus, Allegro, SimWalk, Merlin), and the simulation package SPLink, and provides genome-wide as well as chromosomal postscript plots of LOD scores, NPL scores, P values, and many other parameters. The software can analyze STRs as well as SNP chip data from Affymetrix, Illumina, or self-defined SNP data (Rockefeller Genetic Analysis Software Homepage).



easyLINKAGE Plus v5.05

A Graphical User Interface For Two-/Multi-Point Linkage Analyses

- Tom H. Lindner & Katrin Hoffmann, (c) Copyright 2005 -
- Last build: 13:50:32 - 12.6.2006 -

Select locus analysis type: ● Single locus

Linkage analysis programs

Allegro v1.2c - Two-/MultiPoint Para-/NonParametric Linkage Analysis

Select project type: ○ Microsatellite marker / Finemapping project ● SNP chip project

Chromosome(s) / chromosomal interval to analyze

CHR(s): ● All ○ None Region from 0.00 to 1000.00 cM - 1 marker every 0.0000 cM

☑ C01 ☑ C02 ☑ C03 ☑ C04 ☑ C05 ☑ C06 ☑ C07 ☑ C08 ☑ C09 ☑ C10 ☑ C11 ☑ C12

☑ C13 ☑ C14 ☑ C15 ☑ C16 ☑ C17 ☑ C18 ☑ C19 ☑ C20 ☑ C21 ☑ C22 ☑ X

Genetic marker map

● Sex averaged ○ Female ○ Male Illumina 5k deCODE without Y, XY Human

Additional options

☑ Recode alleles ☑ Recode IDs ☑ Mendelian test ☑ NON-Mendelian test ☐ HWE test

☐ Sounds ☐ Tables ☐ Redirect STDOUT/ERR ☑ Vertical AUC lines ☐ Autoscale Y axis

DIR J:\EasyLinkageMasoud\M019

Pedigree information file to analyze

pedfile.pro

SNP project file

M019.snp

Figure 2-7: Screenshot of EasyLinkage main interface: It's possible to choose the software of interest, single chromosomes or sets of chromosomes, centimorgan intervals or total genome-wide analyses, microsatellite or SNP marker projects in addition to several error check routines.

The user can perform single/two-locus analyses. Furthermore, single chromosomes, sets of chromosomes, defined centimorgan intervals can be analyzed, or total genome-wide analyses can be applied see Figure 2-7 (Hoffmann and Lindner 2005).

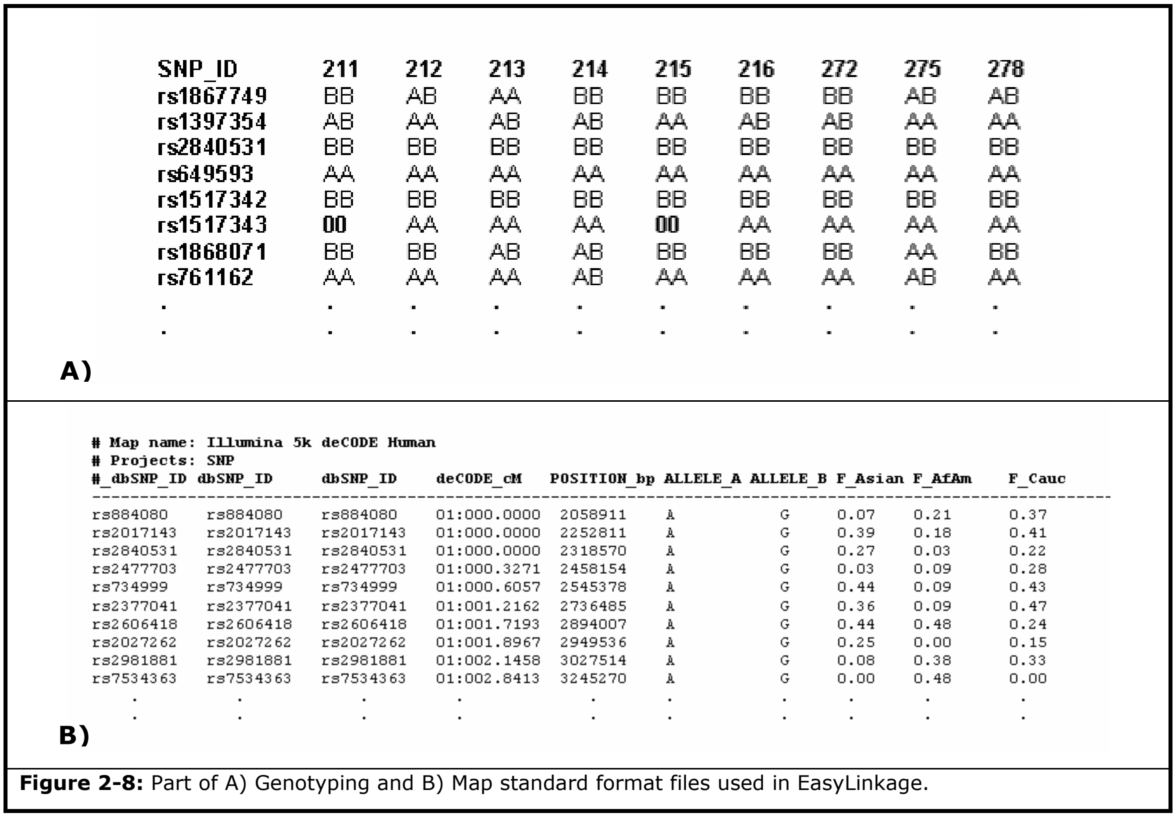
Here again the pedigree files have to be provided in the standard linkage format similar to the one that has been introduced for ALOHOMORA and must be saved as a text file with .pro extension and always a "p" at the beginning of the file name, for example "p_M019.pro".

Scanned genotyping data from the chip have to be converted to the ".txt" files using the BeadStudio software provided by Illumina.

In the genotype file the first line contains the person IDs. On the left hand-side the SNPs are listed. IDs of SNPs can be those from Affymetrix, dBSNP, or from "The SNP consortium (TSC)". Alleles must be depicted by "A" or "B", blanks as "00" (bold).

The genotyping file must be saved with .snp extension for example "M019.snp" see Figure 2-8a.

For the SNPs, the genetic map and Caucasian allele frequency file were provided from deCode and Illumina respectively and converted to the appropriate formats (Figure 2-8b).



```
SNP_ID      211  212  213  214  215  216  272  275  278
rs1867749   BB   AB   AA   BB   BB   BB   BB   AB   AB
rs1397354   AB   AA   AB   AB   AA   AB   AB   AA   AA
rs2840531   BB   BB   BB   BB   BB   BB   BB   BB   BB
rs649593    AA   AA   AA   AA   AA   AA   AA   AA   AA
rs1517342   BB   BB   BB   BB   BB   BB   BB   BB   BB
rs1517343   00   AA   AA   AA   00   AA   AA   AA   AA
rs1868071   BB   BB   AB   AB   BB   BB   BB   AA   BB
rs761162    AA   AA   AA   AB   AA   AA   AA   AB   AA
.           .    .    .    .    .    .    .    .    .
.           .    .    .    .    .    .    .    .    .
```

```
# Map name: Illumina 5k deCODE Human
# Projects: SNP
#_dbSNP_ID dbSNP_ID     dbSNP_ID     deCODE_cM     POSITION_bp ALLELE_A ALLELE_B F_Asian F_AfAm    F_Cauc
------------------------------------------------------------------------------------------------------------
rs884080    rs884080     rs884080     01:000.0000   2058911     A           G      0.07    0.21     0.37
rs2017143   rs2017143    rs2017143    01:000.0000   2252811     A           G      0.39    0.18     0.41
rs2840531   rs2840531    rs2840531    01:000.0000   2318570     A           G      0.27    0.03     0.22
rs2477703   rs2477703    rs2477703    01:000.3271   2458154     A           G      0.03    0.09     0.28
rs734999    rs734999     rs734999     01:000.6057   2545378     A           G      0.44    0.09     0.43
rs2377041   rs2377041    rs2377041    01:001.2162   2736485     A           G      0.36    0.09     0.47
rs2606418   rs2606418    rs2606418    01:001.7193   2894007     A           G      0.44    0.48     0.24
rs2027262   rs2027262    rs2027262    01:001.8967   2949536     A           G      0.25    0.00     0.15
rs2981881   rs2981881    rs2981881    01:002.1458   3027514     A           G      0.08    0.38     0.33
rs7534363   rs7534363    rs7534363    01:002.8413   3245270     A           G      0.00    0.48     0.00
    .           .            .            .             .       .           .       .       .        .
    .           .            .            .             .       .           .       .       .        .
```

Figure 2-8: Part of A) Genotyping and B) Map standard format files used in EasyLinkage.

2.11 Quality Control

With ALOHOMORA it is possible to perform useful quality control steps of the genotyping data prior to starting linkage analysis, like checking gender and relationships between family members with the help of the Graphical Representation of Relationship errors (GRR) software. Mendelian errors can be identified by the PedCheck software, non-Mendelian errors and unlikely genotypes can be detected with Merlin.

2.11.1. Gender Check

As a first quality control step, the gender of samples was checked by counting the heterozygous SNPs on the X-chromosome and comparing it to the pedigree file information.

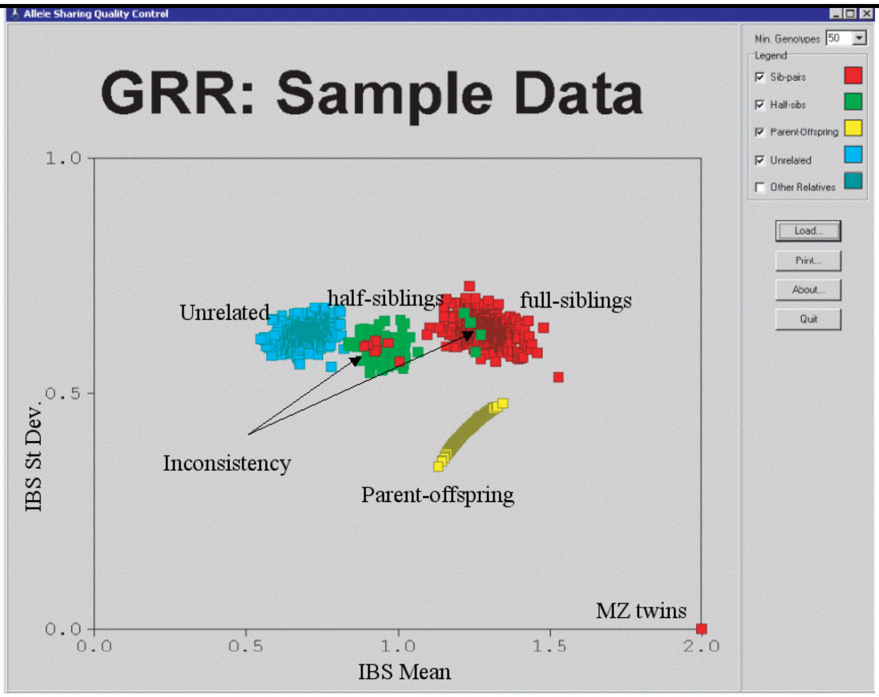
2.11.2. Graphical Representation of Relationship errors (GRR)

A common problem in genetic studies is the misspecification of relationships between DNA samples. Misspecification of relationships can lead to inaccurate or biased results; therefore in order to verify the assumed relationships between individuals in each family, the data were subjected to standard quality control using GRR.

GRR uses a simple, general approach for verifying that individuals with the same specified relationship have similar patterns of allele sharing.

The method is defined as follows: first, classify each pair of individuals according to their assumed relationship (such as sib-pairs, parent–offspring pairs, unrelated individuals, etc.). Second, calculate the mean (μi j) and variance (σi j) of identical-by-state allele sharing over a number of polymorphic loci for each pair of individuals, i and j. If the sample is homogenous, we expect each group to display a characteristic pattern of allele sharing.

For example, sib-pairs will be expected to share more alleles on average than unrelated individuals, while parent– offspring pairs (which share at least one chromosome) are expected to show less variability in allele sharing than sibpairs (which may share zero, one or two chromosomes). A convenient way to identify individuals with patterns of allele sharing that are inconsistent with their specified relationship is to colour code and plot these mean variance statistics (Figure 2-9)(Abecasis and others 2001).



Min. Genotypes 50

☑ Sib-pairs 🟥
☑ Half-sibs 🟩
☑ Parent-Offspring 🟨
☑ Unrelated 🟦
□ Other Relatives

Figure 2-9: Typical results for a genome scan in a non-inbred sample (screenshot). Several distinct clusters are present: unrelated individuals have the lowest average sharing and high variance (coloured in blue); half-siblings have higher sharing on average (coloured in green) and full-siblings have even higher sharing (coloured in red); parent– offspring pairs have a similar degree of allele sharing to sib-pairs but with lower variance (coloured in yellow). All other relative pairs are grouped together and not displayed by default. Note that some sibling and full-sibling pairs have been misclassified and appear in other clusters. A single sib-pair displays maximum average sharing (bottom right corner) and corresponds to a pair of identical twins (Abecasis and others 2001).

2.11.3. Elimination of Mendelian inconsistebcies

Prior to performaning of linkage analysis, elimination of all Mendelian inconsistencies in the pedigree data is essential. Often, identification of erroneous genotypes by visual inspection can be very difficult and time consuming. In fact, sometimes the errors are not recognized until the stage of running the linkage-analysis software. In such a case significant efforts are required to find the erroneous genotypes and to cross-reference pedigree and marker data that may have been recoded and renumbered.

PedCheck is a computer program with four error-checking algorithms, which help to identify all Mendelian inconsistencies in pedigree data and will provide them with useful and detailed diagnostic information to help resolve the errors. This program handles large data sets quickly and efficiently, accepts a variety of input formats, and offers various error-checking algorithms that match the subtlety of the pedigree error. These algorithms range from simple parent-offspring compatibility checks to a single-locus likelihood-based statistic that identifies and ranks the individuals most likely to be in error (O'Connell and Weeks 1998).

SNPs with Mendelian errors and SNPs that are not informative for any individual of a dataset can be selectively removed from the data. Therefore in this study the PedCheck program was used for detection of Mendelian errors.

2.11.4. Detection of non-Mendelian errors and unlikely genotypes

“Unlikely genotypes” are equivalent to double recombinations in a short chromosomal segment. The reasons may be genotyping errors, a wrong SNP position in the genetic map, or very rarely gene conversion.

Particular markers may not show any Mendelian problem if analyzed individually. The non-Mendelian test is Merlin based (see section 2.13.1) and depends on the allele frequency algorithm and on the haplotype assignment by Merlin. To detect improbable genotypes, Merlin calculates the likelihood of observed genotypes conditional on all recombination fractions L(G|θ), assuming that all markers are unlinked, L(G|θ=½). Merlin then marks, in turn, each genotype g as unknown and updates these likelihoods to obtain L(G\g|θ) and L(G\g|θ=½). If the information provided by g is consistent with neighboring markers, we expect that the ratio rlinked=L(G\g|θ)/L(G|θ) is small compared to runlinked=L(G\g|θ=½)/L(G|θ=½). Genotypes that provide information inconsistent with neighbouring markers, however, will cause the statistic r=rlinked/runlinked to take unusually large values (Abecasis and others 2002).

Merlin has a two step analysis run, which the first run is for error detection and non-parametric LOD score analysis. The second run makes a non-parametric LOD score analysis and a haplotyping with the cleaned data set.

2.12 Linkage analysis

The human genome contains 20-30 thousands of genes. Therefore, finding the particular gene or genes responsible for any given human disease has always been a tricky task, quite literally like finding a needle in a haystack.

“Linkage analysis” serves as a way of disease gene-hunting and genetic testing. In this approach, the aim is to find out the rough location of the gene relative to another DNA sequence with known position in the genome called a genetic marker (any polymorphic

Mendelian character that can be used to follow a chromosomal segment through a pedigree).

Linkage and linkage disequilibrium are two key concepts in this context. Two genetic loci are linked if they are transmitted together from parent to offspring more often than expected under independent inheritance. They are in linkage disequilibrium if, across the population as a whole, they are found together on the same haplotype more often than expected. In general, two loci in linkage disequilibrium will also be linked, but the reverse is not necessarily true (Dawn Teare and Barrett 2005).

2.12.1. LOD scores

Linkage is usually reported as a logarithm of the odds (LOD) score. This score was first proposed by Morton (Morton 1955). It is a function of the recombination fraction (θ) or chromosomal position measured in cM. This means that the LOD score is different depending upon which value of θ is being considered. Large positive scores are evidence for linkage (or cosegregation), and negative scores are evidence against. To calculate a LOD score a model for disease expression must be specified. This model includes the frequency of the disease allele and mode of inheritance (e.g. dominant or recessive), marker allele frequencies, and a full marker map for each chromosome. The ultimate objective of the analysis is to estimate the recombination fraction between individual markers and the disease locus (two-point) or position of the disease locus relative to a fixed map of markers where the location of each marker is assumed to be known (multipoint). The best (maximum likelihood) estimate of θ or position is that which maximises the LOD score function: the maximum LOD score.

LOD score analysis is equivalent to likelihood ratio testing, but for historical reasons, instead of natural logarithms, logs to the base 10 are used. In the linkage analysis framework, the only parameter of interest is the recombination fraction (θ) between marker and disease locus or the map position of the disease locus with respect to a fixed map of markers. The null hypothesis represents no linkage between disease and marker locus ($\theta=0{\cdot}5$), and the alternative hypothesis assumes that linkage exists ($\theta<0{\cdot}5$).

The LOD score function is then defined as:

$$\text{Recombination Fraction} = \theta = \frac{\text{Recombinant Meiosis}}{\text{Recombinant Meiosis + Non Recombinant Meiosis}}$$

$$LOD(\theta)=\log_{10}\left[\frac{Like(\theta)}{Like(\theta=\frac{1}{2})}\right]$$

The LOD score function is maximised with respect to the recombination fraction (θ) in two-point analysis (a single marker and disease locus), or map position in multipoint analysis (disease locus and at least two markers at fixed relative positions).

The value of θ which gives the maximum LOD score is the maximum likelihood estimate of θ (Dawn Teare and Barrett 2005).

2.12.2. 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—ie, 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 (Dawn Teare and Barrett 2005).

FastLink v4.1 (see section 2.13.4) and SuperLink v1.4 (see section 2.13.5) were used for two-point parametric linkage analysis in special cases.

Multipoint parametric linkage analyses were performed using Allegro v1.2 (see section 2.13.3), GeneHunter 2.1v5 (see section 2.13.2) and Merlin (see section 2.13.1). There is no limitation regarding the number of markers for 10K arrays in Allegro and Merlin but for 50k and 250k panels, subsets of 300-800 markers were used. GeneHunter however has some limitations with respect to the number of markers, therefore the analyses were performed with contiguous subsets of 50–300 markers in the way of a non-overlapping moving window except for one (ie., the last SNP in the first set is identical with the first SNP in the second set, etc.).

In case of large pedigrees, when GeneHunter dropped individuals from the analysis (due to the limit of calculation power) and both Allegro and Merlin stalled, the pedigrees were split to appropriate sizes in order to be able to run Allegro, Merlin or GeneHunter. Non-parametric and parametric LOD scores were calculated and plotted for all chromosomes.

In order to not miss the linkage signal in regions between contiguous marker sets, the data were analysed with marker sets of different sizes. For the analyses, marker allele frequencies in the Caucasian population were used.

2.12.3. 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 (Dawn 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.

Non-parametric LOD score calculations were preferably performed with Merlin (2.13.1) or GeneHunter (2.13.2) chromosome by chromosome, using all SNPs on a chromosome simultaneously for a multipoint analysis. With Merlin no limitation regarding the number of markers was observed up to 1000 SNPs (depending on pedigree size) but for larger numbers subsets of 300-800 markers were used. For GeneHunter contiguous sets of 50–300 markers (depending on pedigree size) were used.

With the help of Gnuplot software we produced a genome-wide view of each analysis.

2.12.4. Haplotyping

Sets of alleles on the same small chromosomal segment tend to be transmitted, as a block through a pedigree is known as haplotype. Haplotypes mark recognizable chromosomal segments that can be tracked through pedigrees and through populations (Strachan T 2003).

Merlin, GeneHunter and Allegro (see sections 2.13.1-2.13.3) are able to infer haplotypes however, Merlin was mostly used for this purpose and afterwards haplotypes were visualized by HaploPainter (Thiele and Nurnberg 2005).

2.13 Linkage analysis software

After converting genotyping data to proper linkage input files using ALOHOMORA and/or EasyLinkage, parametric and non-parametric multipoint linkage analysis (using software Merlin, GeneHunter and Allegro) were carried out.

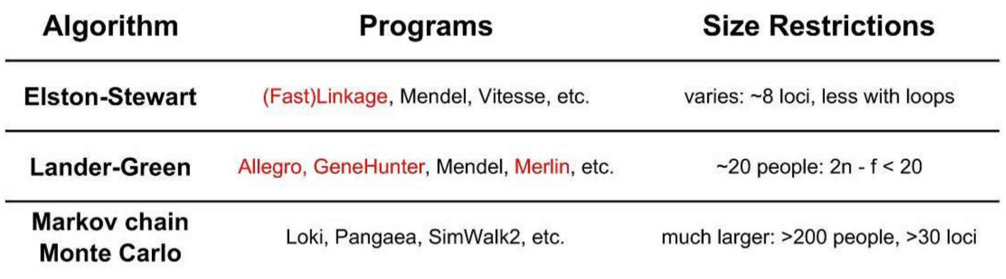
In special cases, parametric single point linkage analysis (using software FastLink, and SuperLink) was performed.

Parametric analysis was based on the assumption of an autosomal recessive mode of inheritance.

In few families, when it was not possible to rule out the X-linked mode of inheritance based on pedigree information, X-linked analysis was also performed.

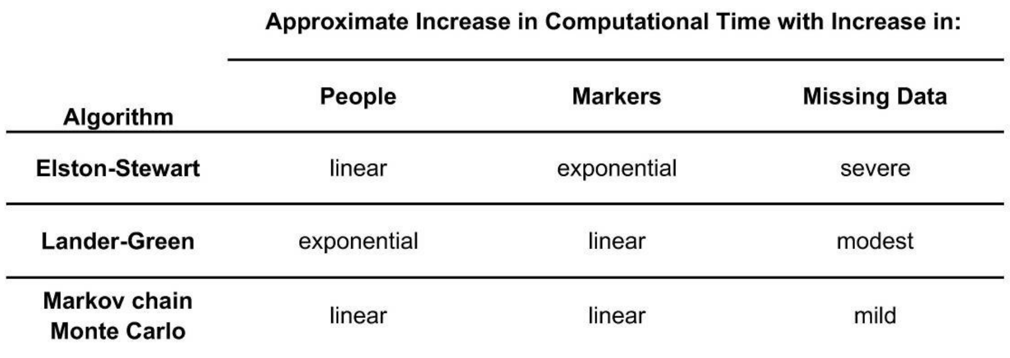
In the following, standard linkage programs that were used regularly (Merlin, GeneHunter and Allegro), or rarely (FASTLINK and SUPERLINK) will be briefly introduced.

Allegro, GeneHunter and Merlin use Lander-Green algorithm and have a pedigree size restriction of about 16 people per analysis (2n-f<16 which n and f refer to the number of non-founder and founder samples in pedigree). In this algorithm, increasing the number of individuals and markers will respectively increase the calculation time exponentially and linearly. Missing data have only a modest effect on calculation time. Comparisons to the other algorithms like Elston-stewart and Markov chain Monte Carlo are shown in Table 2-8 and Table 2-9.



Algorithm	Programs	Size Restrictions
Elston-Stewart	(Fast)Linkage, Mendel, Vitesse, etc.	varies: ~8 loci, less with loops
Lander-Green	Allegro, GeneHunter, Mendel, Merlin, etc.	~20 people: 2n - f < 20
Markov chain Monte Carlo	Loki, Pangaea, SimWalk2, etc.	much larger: >200 people, >30 loci

Table 2-8: The implemented algorithms in some of the commonly used linkage programs and their pedigree size restrictions. The softwares depicted in red were used in our analyses.



	Approximate Increase in Computational Time with Increase in:		
Algorithm	People	Markers	Missing Data
Elston-Stewart	linear	exponential	severe
Lander-Green	exponential	linear	modest
Markov chain Monte Carlo	linear	linear	mild

Table 2-9: Comparison of Elston-Stewart, Lander-Green and Markov chain Monte Carlo algorithms' approximate computational time by increasing in pedigree sizes and markers size, in addition to the amount of data missing.

2.13.1. MERLIN: Multipoint Engine for Rapid Likelihood INference

Merlin (Abecasis and others 2002) carries out single-point and multipoint analyses of pedigree data, including IBD and kinship calculations, non-parametric and variance component linkage analyses, error detection and information content mapping. For multipoint analyses in dense maps, Merlin allows the user to impose constraints on the number of recombinants between consecutive markers. Merlin estimates haplotypes by finding the most likely path of gene flow or by sampling paths of gene flow at all markers jointly. It can also list all possible non-recombinant haplotypes within short regions. Finally, Merlin provides swap-file support for handling very large numbers of markers as well as gene-dropping simulations for estimating empirical significance levels (Abecasis and others 2002).

2.13.2. GeneHunter

GeneHunter provides wide range of analyses for performing linkage and disequilibrium analyses. It can perform very rapid extraction of complete multipoint inheritance information from pedigrees of moderate size. This information is then used for the exact computation of multipoint LOD scores, non-parametric linkage statistics as well as a wide range of sib pair analyses and a new variance components analysis. In addition, several transmission disequilibrium test (TDT) analyses are also available for searching for association/disequilibrium in addition to linkage. Quick calculations involving dozens of markers, even in pedigrees with inbreeding and marriage loops, are possible with GeneHunter (Kruglyak and others 1996).

2.13.3. Allegro

Allegro can do both classical parametric linkage analysis and analysis based on allele sharing models. In addition, Allegro estimates total number of recombinations between markers, computes posterior IBD sharing probabilities, reconstructs haplotypes and does two types of simulation. Thus Allegro includes the basic functionality of the well known GeneHunter program (Kruglyak and others 1996). It can analyse pedigrees of moderate size, and it can handle many markers (as opposed to programs such as Linkage and FastLink, which do parametric analysis of large pedigrees, but only with a few markers). The biggest advantages of Allegro over GeneHunter are the allele sharing models that it provides and a much shorter execution time (Gudbjartsson and others 2000).

2.13.4. FASTLINK

FastLink is a faster version of the existing genetic linkage analysis programs in LINKAGE 5.1.

The core of the LINKAGE package is a series of programs for maximum likelihood estimation of recombination rates, calculation of LOD score tables, and analysis of genetic risks. The analysis programs are divided into two groups. The first group can be used for general pedigrees with marker and disease loci. Programs in the second group are for three-generation families and codominant marker loci, and are primarily intended for the construction of genetic maps from data on reference families (FASTLINK Home Page).

2.13.5. SuperLink

SuperLink is a computer program that performs exact genetic linkage analysis with input-output relationships similar to those in standard genetic linkage programs.
Some features of SuperLink are (SuperLink home page):

- Analysis of general pedigrees (many individuals, inbreeding loops, many markers, etc.).
- Analysis of two-locus traits.
- Analysis of autosomal or sex-linked traits.
- Maximum-likelihood Haplotyping analysis.
- Analysis of complex traits

2.14 Copy number analysis

Identification and detection of DNA copy number changes could be a powerful tool in studying of genetic diseases. High resolution, whole genome, SNP arrays provide new ways of detecting chromosomal imbalances by enabling researchers to analyze copy number alterations, Loss of heterozygosity (LOH)- loss of normal function of one allele of a gene in which the other allele was already inactivated-, and genotypes in a single experiment.
Recently several different softwares have been developed to identify genome-wide chromosomal gains and losses using high density, oligonucleotide, array-based SNP genotyping methods and the Whole Genome Sampling Assay (WGSA).
We used the Affymetrix GeneChip® Chromosome Copy Number Analysis Tool 4.0 (CNAT 4.0) integrated into the Affymetrix GeneChip® Genotyping Analysis Software (GTYPE) and Copy Number Analyzer for Affymetrix GeneChip (CNAG2.0) to perform copy number analysis.
Copy number analyses were performed in a non-paired way (samples without a paired reference from the same individual) for all the affected members
In case of observing a variation (especially for the regions with the high LOD scores) its segregation in the family were investigated.

2.15 Prioritizing genes for mutation screening

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 bioinformatic databases.

For this purpose several databases such as PosMed (PosMed home page), Prioritizer (Prioritizer home page) and in some cases ENDEAVOUR (Aerts and others 2006) were used.

2.16 Isolation of genomic DNA from lymphoblastoid cells

DNA extraction was performed based on the following protocol:

- Make buffer A with the following composition:

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.

After autoclaving add 5ml 20% -SDS (Sodiumdodecylsulfat)

- Transport the cell pellets in 50ml falcon tubes on ice from the freezer room to the lab
- Add 20ml from solution A (see above) on ice to each sample: (First add 1 ml with a cut, filterless pipet tip and let it run up and down several times, then add the 19ml that are left)
- Vortex, until suspension appears homogenous
- Put the falcon tubes into a holder at room temperature
- Add 30µl RNase A (10mg/ml)
- Incubate for 60min at 37°C in a water bath
- Add 5ml sodiumperchlorate (Natriumperchlorat)
- Shake it over head 10-15 times (not more!) manually
- Add 20ml cold chloroform under the hood
- Shake by converting the tube 10-15 times manually
- Centrifugate for 10 min at 4000U/min
- Remove the upper phase with a glass pipette and the pipette boy [If the upper phase is very cloudy (a lot of protein debris) repeat the chloroform extraction]

- Transport in a new and labelled 50ml Falcon tube
- Add a volume of ice-cold ethanol (100%) [e.g.: at 25 ml sample volume add 25 ml ethanol]
- Capture the DNA with an expendable diluting loop [The DNA precipitate after converting the tube several times]
- Transport the DNA in an eppendorf tube with 1ml cold ethanol (70%)
- Centrifuge for 1 min at 7500 U/min
- Pipette off the ethanol
- Add 500µl of new ethanol (70%)
- Centrifuge for 1 min at 7500 U/min
- Remove the ethanol
- Leave tube open in the Thermomixer at 50°C, until the DNA pellet is dry
- Add 500µl Tris-EDTA- (TE-) buffer
- Leave it over night at room temperature to re-dissolve

2.17 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 U Taq polymerase (Promega, Mannheim, Germany or Qiagen, Hilden, Germany). The following touchdown PCR profile was used.

Step1: 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.

Step2: 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.

2.18 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 FF, and 30% glycerol were added to the nucleic acid solutions before loading into the wells. HyperLadder I, IV, 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.

2.19 Sequencing

The original PCR products were either purified using the Qiaquick PCR Purification Kit (Qiagen, Hilden, Germany) or they were directly sequenced in both directions using the ABI 377 DNA sequencer.

The labelling reactions were carried out using the following amount of reagents shown in Table 2-10.

Table 2-10: PCR reaction mix for sequencing reaction

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

Thereafter, sequencing reactions were performed using the following temperature profile shown in Table 2-11.

Table 2-11: PCR conditions for sequencing reaction

	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	

DNA precipitation and purifications were done based on the following protocol:

- Add 1µl 2%SDS and incubate at 98°C for 10 second
- Add 25 µl 100% EtOH to each reaction and mix thoroughly by inverting the tube
- Centrifugate at 4000 rpm in the cool room for 60
- Carefully discard the supernatant by inverting the tubes and placing them on a paper towel
- Add 150 µl 70% EtOH and invert the tubes without disturbing the pellets
- Centrifuge at 4000 rpm in the cool room for 30 min
- Carefully discard the supernatant by inverting the tubes and placing them on a paper towel

- Repeat the washing step
- Dry the pellet by putting the plate headfirst onto a paper towel and centrifuge just up to 4000 rpm and then stop
- Put an adhesive film on the plate and wrap it with Aluminum foil (if it is going to be shipped somewhere else)
- Base calling by putting samples in the sequencing machine.

Sequence data were assembled and analysed using the GAP4 Contig Editor.1 or CodonCode aligner 1.6.0 beta 5 software.

2.20 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 (Table 2-12) were used for amplification of DNA fragments in case of screening mutations in *CA8* and *CYP7B1*.

Table 2-12: Primers for DNA amplification in position of *CA8* and *CYP7B1* mutations

Name	Sequence
MG407_CA8_7F	TCAGGATTGTTATTAATTCACTTGC
MG408_CA8_7R	CAAAACAGACATTTTCCTTTCTCAG
MG665_CYP_4F	GACAAGATGGTGCAGCAGTG
MG666_CYP_4R	TGCAAATCTAATCAGTGTAATAAACG

- Master-mixes for Hpy188I and HpyCH4 III digestion were prepared as follows:

Table 2-13: Enzyme Master Mix	
10X buffer	0.1 µl
Restriction enzyme	0.1 µl
Water	Up to (2µl X samples)

- Restriction mix was prepared according to Table 2-14:

Table 2-14: Restriction mix	
PCR Product	5 µl
10X Buffer	5.5 µl
Water	42.5 µl
Enzyme Master Mix	2 µl

- Incubate for 2 hours at 37°C
- Speed-vac to reduce the volume (at least 2 times)
- Loading on the agarose gel

2.21 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

- Suspend Cell pellet (5 x 10^7 cells) with 10ml Trizol reagent in a 30ml RNase free tube.
- Homogenize the suspension by shaking vigorously for several seconds. Incubate 30 minutes at room temp (20-30 °C) to be completely dissolved.
- Add 0.2 ml chloroform for each 1 ml of initial Trizol (2ml). Shake for 15 seconds, and incubate for additional 2-3 min. at room temp.
- Centrifuge the samples for 20 min. at 5000 RPM at 4°C.
- Transfer the aqueous phase to a fresh 30ml tube or make aliquots of 550 µl in 1.5 µl eppendorf tubes.
- Add 0.5 volume of isopropanol per 1 ml of TRIZOL reagent used for initial homogenization (5ml or 550µl) to the aqueous phase, mix well by vortexing and hold in room temperature for 5-10min.
- Centrifuge the samples for 10 min. at 8000 RPM at 4°C (12000g for microfuge).

- Remove the supernatant and add 10 ml filter sterilized 70% ethanol (500µl for microtube) and mix well.
- Centrifuge the samples for 5 min. at 5000 RPM at 4°C (7500g for microfuge).
- Take off the supernatant and air dry the pellet. Avoid completely drying the pellets, as this will decrease the solubility of the RNA.
- Dissolve the RNA in 500 µl of sterile DEPC water and put it on ice for 10 min then incubate for 5 min at 65°C using heating block or water bath.
- Measure the RNA concentration by Nanodrop ND-1000 Spectrophotometer (Peqlab Biotechnologie GmBH) and check the quality on Agarose gel.
- Keep the RNA in the freezer (-20 or –80 C) until further use.

2.22 First-Strand cDNA Synthesis Using SuperScript ™ III for RT-PCR

cDNA synthesis was performed according to the following protocol:

- Add 50-250ng of random primers to a 0.2ml eppendorf tubes.
- Add 10pg – 5µg total RNA.
- Add 1µl 10mM dNTP Mix (Mix: 10mM each dATP, dGTP, dCTP and dTTP at neutral pH)
- Add distilled water to bring the volume to a total of 13µl.
- Heat mixture to 65°C for 5 min and incubate on ice for at least 1 min.
- Collect the contents of the tube by brief centrifugation.
- Add: 5µl 5X First-Strand Buffer,
  1µl 0.1M DTT,
  1µl RNaseOUTTM Recombinant RNase Inhibitor
  1µl of SuperScript™ III RT (200 units/µl)
- Mix by pipetting gently up and down and incubate tube at 25°C for 5 min.
- Incubate at 50°C for 30-60 min.
- Inactivate the reaction by heat (70°C for 15 min).

cDNAs synthesis were checked using using primers for HUWE1, a control gene located on chromosome X with exon spanning primers CAAGTGAGGAAAAGGGCAAA (exon64) and GTTCATGAGCTGCCCCAGT (exon65) which give rise to a 568bp amplicon.

2.23 Rapid Amplification of cDNA Ends (RACE)

Rapid amplification of cDNA ends (RACE) is a polymerase chain reaction-based technique, which facilitates the cloning of full-length cDNA sequences when only a partial cDNA sequence is available. It can be used to obtain the 5' end (5' RACE-PCR) or 3' end (3' RACE-PCR) of mRNA.

The protocols for 5' or 3' RACE differ slightly. 5' RACE-PCR begins using mRNA as a template for a first round of cDNA synthesis using an anti-sense oligonucleotide primer that recognizes a known sequence in the gene of interest; the primer is called a gene specific primer (GSP) and copies the mRNA template in the 3' to the 5' direction to generate a specific single stranded cDNA product. Following first strand cDNA synthesis, the enzyme terminal deoxynucleotidyl transferase (TdT) is used to add a homopolymeric tail (i.e. a string of identical nucleotides) to the 5' end of the cDNA. A PCR reaction is then carried out, which uses a second anti-sense gene specific primer (GSP2) that binds to the known sequence, and a sense general universal primer (UP) that binds the homopolymeric tail added to the 5' ends of the cDNAs, to amplify a cDNA product from the 5' end.

3' RACE-PCR uses the natural polyA tail that exists at the 3' end of all mRNAs for priming during reverse transcription. Therfore, this method does not require the addition of nucleotides by TdT. First strand cDNAs are generated using an Oligo-dT-adaptor primer that complements the polyA stretch and adds a special adaptor sequence to the 3' end of each cDNA. PCR is then used to amplify 3' cDNA from a known region in the specific cDNA using a sense GSP, and an anti-sense primer complementary to the added adaptor sequence.

SMART™ RACE

5′ RACE was done using SMART™ RACE cDNA Amplification Kit (Clontech, Cat. No. 634914), based on the manufacture manual.

RLM-RACE

As a alternative method FirstChoice® RLM-RACE Kit (Ambion, Cat. No: AM1700) have been used for 5′ RACE.

In this method, total or poly(A) selected RNA is treated with Calf Intestine Alkaline Phosphatase (CIP) to remove free 5'-phosphates from molecules such as ribosomal RNA, fragmented mRNA, tRNA, and contaminating genomic DNA. The cap structure found on intact 5' ends of mRNA is not affected by CIP. The RNA is then treated with Tobacco Acid Pyrophosphatase (TAP) to remove the cap structure from full-length mRNA, leaving a 5'-monophosphate. A 45 base RNA Adapter oligonucleotide is ligated to the RNA population using T4 RNA ligase. The adapter cannot ligate to dephosphorylated RNA because these molecules lack the 5'-phosphate necessary for ligation. During the ligation reaction, the majority of the full length, decapped mRNA acquires the adapter sequence as its 5' end. A random-primed reverse transcription reaction and nested PCR then amplifies the 5' end of a specific transcript.

All the 5′ RACE reactions were performed with 4 different Gene-Specific Primers (GSPs) inside 3′ UTR of MCPH1.

Table 2-15: Primers used for 5′ RACE in MCPH1

Primer Name:	**Sequence:**
MCPH1_RACE_1R	TGACCTCACTGGCCTGTGGTGACTG
MCPH1_RACE_2R	AGAGACAGGGTTTCGCCATGTTGGC
MCPH1_RACE_3R	CACAATGTCCACTGGCCGCTTTTTG
MCPH1_RACE_4R	TGCAGTGAGCCAAGGTTGCAGTGAA

Gene-Specific Primers (GSPs) should be:

- 23–28 nt
- 50–70% GC
- Tm ≥65°C; best results are obtained if Tm >70°C (enables the use of touchdown PCR)

2.24 Whole genome expression profiling

Introduction

The Sentrix Human-6 Expression BeadChips contains six arrays on a single BeadChip, each with >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 3μm 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 (Figure 2-10).

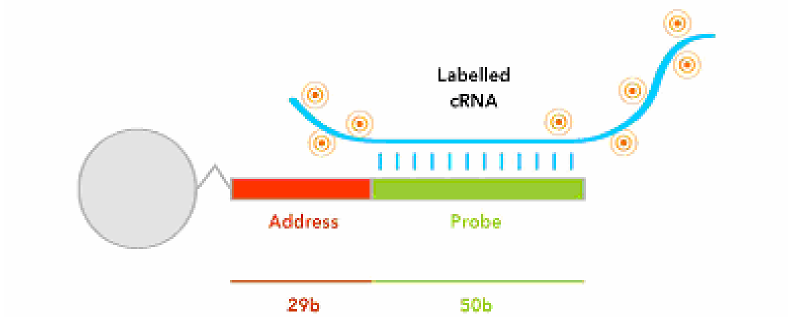


Figure 2-10: BeadChips design: contain a 29-base address concatenated to a 50-base gene-specific probe (from Illumina Sentrix® Human-6 Expression BeadChip bulttein)

Content Sources

The Human-6 Expression BeadChips contain content from a variety of public data sources (Table 2-16).

Table 2-16: CONTENT SOURCES	
Curated RefSeq (Release 4 and Build 34)	19.730
Genome Annotation RefSeq (Release 4 and Build 34)	6.368
Gnomon (Build 34)	9.576
Unigene-163	11.622
Total	47.296

Controls

Every array on each Human-6 Expression BeadChip includes >1000 bead types as controls for every experiment.

The controls allow all steps in the process to be monitored carefully, using the following parameters (Illumina Technical Bulletin):

- Sample quality
- Labeling reaction success
- Hybridization stringency
- Signal Generation

2.24.1. 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 single-stranded 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.

2.24.2. 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.

2.24.3. Expression Data Analysis

BeadStudio

For the analysis of expression data the BeadStudio software package included with Illumina® Gene Expression System was applied, which is a tool for analyzing gene expression data from scanned microarray images collected from the Illumina BeadArray

Reader. Resulting BeadStudio files can be used by most standard gene expression analysis programs. BeadStudio executes two types of data analysis:

- Gene Analysis:
  Quantifying gene expression signal levels
- Differential Analysis:
  Determining if gene expression levels have changed between two experimental groups.

One can perform these analyses on individual samples or on groups of samples treated as replicates.

BeadStudio reports experiment performance based on built-in controls that accompany each experiment.

In addition, BeadStudio provides scatter plotting and dendrogram tools, facilitating quick, visual means for exploratory analysis.

Experiment Creation & Analysis

Using the intensity files produced by the BeadArray Reader, BeadStudio's Gene Analysis tool produces output files containing:

- Probe and gene lists
- Associated hybridization intensities (normalized or raw)
- Information about the system controls

If desired, BeadStudio's Differential Analysis tool can produce output files determining the probability that a gene's signal has changed between two samples or groups of samples.

Using these output files, BeadStudio's Data Visualization tools can create more sophisticated plotting analyses such as Scatter Plots, Cluster Analysis Dendrograms, and Control Summary Graphs. To produce the BeadStudio output files, an experiment have to be defined. In a BeadStudio experiment, the used samples and their grouping (sample sets that can be compared against each other for the purpose of identifying gene expression differences) have to be defined.

To define experiment, first groups have to be specified, then samples have to be assigned to them. In the simplest experiment, each group will have only one sample. However, if experiment includes replicate samples, they can be assigned to the same group.

Within a group, BeadStudio will average the values for each gene across the samples, and its algorithms will automatically take advantage of the replicates' statistical power to provide more sensitive determination of detection and differential expression.

We did differential analysis by comparing all the patients as a group against all the controls as another group. We also analyzed each patient separately in comparison with the group of controls.

Normalization & 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. 'Rank-invariant' 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.

Differential Expression Algorithms

Beadstudio uses the following 3 algorithms to compare a group of samples (referred to as the condition group) to a reference group.

- Illumina custom
- Mann-Whitney
- T-test

Here, the Illumina Custom algorithm was applied for doing differential expression analysis.

Illumina Custom Algorithm

This model assumes that target signal intensity (I) is normally distributed among replicates corresponding to some biological condition. The variation has three components: sequence specific biological variation (σ_{bio}), non-specific biological variation (σ_{neg}), and technical error (σ_{tech}).

$$I = N(\mu, \sigma)$$
$$\sigma = \sqrt{\sigma_{tech}^2 + \sigma_{neg}^2 + \sigma_{bio}^2}$$
$$\sigma_{tech} = a + b < I >$$

Variation of non-specific signal σ_{neg} is estimated from the signal of negative control sequences (using median absolute deviation). For (σ_{tech}), two sets of parameters (a_{ref},b_{ref}) and (a_{cond},b_{cond}) for reference and condition groups are estimated respectively.(σ_{tech}) is estimated using iterative robust least squares fit which reduces influence of highly variable genes. This implicitly assumes that the majority of genes do not have high biological variation among replicates. When this assumption does not hold, technical error by some averaged biological variation will be overestimated (BeadStudio User Guide).

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 formula is:

Diff. score = 10*sgn (μ_{ref}-μ_{cond})*log10(p)

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.

2.24.4. Selection of candidates for validation by a second method

Several of the differentially expressed genes were chosen for validation by a second method. Candidates were selected by looking through the literature and/or using some of bioinformatic tools like ENDEAVOUR.

ENDEAVOUR is a software application for the computational prioritization of `test genes', based on a set of `training genes'. The ranking of a test gene is based on its similarity with the training genes, using (currently) the following information sources:

- MEDLINE abstracts and LocusLink textual descriptions
- Gene Ontology annotation
- Interpro protein domains
- BIND protein interactions
- KEGG pathways
- EST-based expression data
- Microarray expression data (also microarray data sets in local mysql databases)
- Transcription factor binding sites (TFBS)
- Cis-regulatory modules (combinations of TFBSs)
- Sequence similarity by BLAST (Aerts and others 2006)

Genes with following criteria’s were used as a training set:

- Genes involved in DNA repair,
- Genes involved in cell cycle control
- Genes with BRCT domains

At the end, genes that commonly appeared within different categories (especially if they were promising according to the based literature approach as well) were selected as candidates.

2.25 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 were used: DAVID (http://david.abcc.ncifcrf.gov/) and Panther (http://www.pantherdb.org/tools/).

2.25.1. 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 gene-to-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 -30 in patients a compared to controls.

2.25.2. Panther

Panther expression analysis tools can be used for microarray data interpretation. Multiple gene lists can be mapped to PANTHER molecular function and biological process categories, as well as to biological pathways. Its pathway visualization tool will display experimental results on detailed diagrams of the relationships between genes/proteins in known pathways.

2.26 Northern Blotting

2.26.1. DEPC treated water

Treatment with diethylpyrocarbonate (DEPC) is the most common method to remove RNases from solutions. 1ml of 0.1 % DEPC was added to 2000 ml aqua bidest., mixed thoroughly, and let set at room temperature for 1 h. Then the water was sterilised by autoclaving and cooled to room temperature prior to use.

2.26.2. Poly-A RNA extraction

Poly-A+ RNAs, obtained from 100 µg total RNA by using Dynabeads oligo-dT25 (Dynal Biotech, Oslo, Norway).

2.26.3. Probe Preparation

The gene-specific probes with an average size between 500 and 1000 were PCR amplified from genomic DNA or cDNA. All probes were designed to hybridize to at least 300 bases of the respective RefSeq cDNA. The specificity of the probes was checked by BLAST alignment.

Amplified probes were purified using QIAquick PCR Purification Kit, (Qiagen, Cat. # 28106). Qualities of the purified probes were checked by running on the 1% agarose gel and measuring the absorbance by Nanodrop ND-1000 Spectrophotometer (Peqlab Biotechnologie GmBH).

List of the primers used for probe amplification is shown in Table 2-17.

Table 2-17: Northern blotting probe amplification primers

Probe Name	Primer sequence	Size	Type
LCK forward	CAACTCATGAGGCTGTGCTG	513bp	Genomic
LCK reverse	CAAGGAGGAGCACACAGAGG		
STAT1 forward	GCTCCCTCTCTGGAATGATG	553bp	cDNA
STAT1 reverse	TTCAGCTGTGATGGCGATAG		
PTEN forward	GCACTTTCCCGTTTTATTCC	766bp	Genomic
PTEN reverse	AGCACATGAAGCATCCACAG		
DUSP4 forward	ACCTCGCAGTTCGTCTTCAG	757bp	Genomic
DUSP4 reverse	CTACGGTGCTCAGCTGTTTG		
ANXA11 forward	TGACTGGTGGCTCACTTCTG	721bp	Genomic
ANXA11 reverse	TTTCCAGACCATTCCAGAGC		
PSAT1 forward	CGGGCCTCTCTGTATAATGC	800bp	Genomic
PSAT1 reverse	GGGAGGGGGTACAACTCTTG		
NCOA1 forward	TCAGTCAAGCTGTCCAGAACC	544bp	cDNA
NCOA1 reverse	TGAAGAATGGCTGCAGATTG		
PLCG2 forward	TCTGCGCTTTGTGGTTTATG	529bp	cDNA
PLCG2 reverse	ATGGCAGGCTTGAAGAAAAG		
HK1 forward	ACCAGACGGTGAAGGAACTG	766bp	Genomic
HK1 reverse	AAGACACATTTCCGCAGGAC		
EGR2 forward	CACTGCTTTTCCGCTCTTTC	775bp	Genomic
EGR2 reverse	CCTCCTTATTCTGGCTGTGC		
PHGDH forward	AGGAGATCATTGGCTGTTCC	805bp	Genomic
PHGDH reverse	GGCCAGCAGGTAGGAGTAAG		
NK4 forward	ACAGACCCTGAATGGTGCTC	653bp	Genomic
NK4 reverse	TGTGAAAACGGACTAATACGG		
FLJ31978 forward	GCAGGAGTTTGTTCATCTGG	588bp	Genomic
FLJ31978 reverse	GCTTTTGCCTTTCAAACTGG		
MG3111_MCPH_Big13F	ACAAGGGGAGAGAACAAGCA	2062bp	Genomic
MG3112_MCPH_Big13R	CAGCTCAGCTCCTACCACCT		
MG3113_MCPH _Big12F	TCAGATCTGCGGAGTGTATCA	1643bp	Genomic
MG3114_MCPH_Big12R	TTGCAAGGAAGTTCAGAGTCC		
MG3115_MCPH _Big11F	GATGCTGGCTCTGTCCCTAC	1344bp	Genomic
MG3116_MCPH _Big11R	CCACCAATGCAAATGAACAG		
MG3117_MCPH _Big8F	AGCCTACCAAATGGCTCACT	2105bp	Genomic
MG3118_MCPH _Big8R	TTTCACACTTTCTCTATGATACAATCG		
MCPH1_N_Prob_F	ATGTCGTCATCCAGGTTGTG	624bp	cDNA
MCPH1_N_Prob_R	CGCCAGTTCCTTCTCTTCAC		

2.26.4. Chemical Transformation

Chemical transformation was performed based on the following protocol:

- Thaw 50 µl of chemically competent E-Coli TOP10 (DH5 alpha, XL-1 blue and …) cells on ice
- Add 5 ng plasmid DNA
- Incubate on ice for 1/2 hour
- Heatshock by incubating tube for 45 seconds at 42°C (not more!) in water bath.
- Place on ice for 2-3 minutes
- Add 1 ml SOC or LB medium
- Incubate tube at 37°C for 1 hour (set shaker at 200 rpm)
- Spread 100 and 900 µl of the transformation mix on LB plates containing appropriate antbiotics (pGEM-T easy carries an ampicilin resistance gene)
- Blue /white selection is possible using pGEM-T easy. Use X-gal on plates.

AXI Plates: Amp: 25µg/ml; X-gal: 40µg/ml; IPTG 0.07 mM.

2.26.5. Northern blot analysis

- Boil 4 g agarose and 300 ml DEPC H_20 in micorwave
- Add MOPS and formaldehyde (Table 2-18) under fume hood

Table 2-18	
2.2 M formaldehyde	72 ml from stock (37%)
1 x MOPS	40 ml MOPS (10x)

- After gel polymerisation for approximately 1 hour, add running buffer (Table 2-19) to the tray.

Table 2-19: Gel running buffer (2 liter)	
1 x MOPS	200ml MOPS (10x)
2.2 M formaldehyde	360 ml from stock (37%)
H2O up to 2 liters	1440 ml

- Pre-run gel for 10 min with 50 V.
- Flush the wells (with 1000 ml pipet) before loading Poly-A RNA samples.
- Denature Poly-A RNA samples and 4µl marker for 5 min at 65°C followed by transferring on ice and loading on the gel.
- Run gel overnight (about 16 hours) at 50 V (31mA),
- Wash with DEPC water for 20 min (for removing formaldehyde).
- Take a picture using a UV-sensitive ruler (align with gel pockets).
- After photography equilibrate gel by washing in 10 x SSC for 20 minute.
- Make blotting sandwich in the following order: 2 x whatman paper + gel + Hybond N+ membrane + 1 whatman paper + paper towels (Figure 2-11:).
- Put a plate and approximately 0.5 kg weight on top of paper towel pack.
- Let it transfer overnight in 10X SSC.
- Disassemble the sandwich the day after and mark position of gel pockets on the membrane using a needle.
- Soak membrane in 10x SSC for a few minutes and thereafter dry on paper towels (RNA side up).
- Crosslink membrane by UV using the auto-crosslink settings of a Stratalinker (Stratagene, La Jolla, CA, USA).

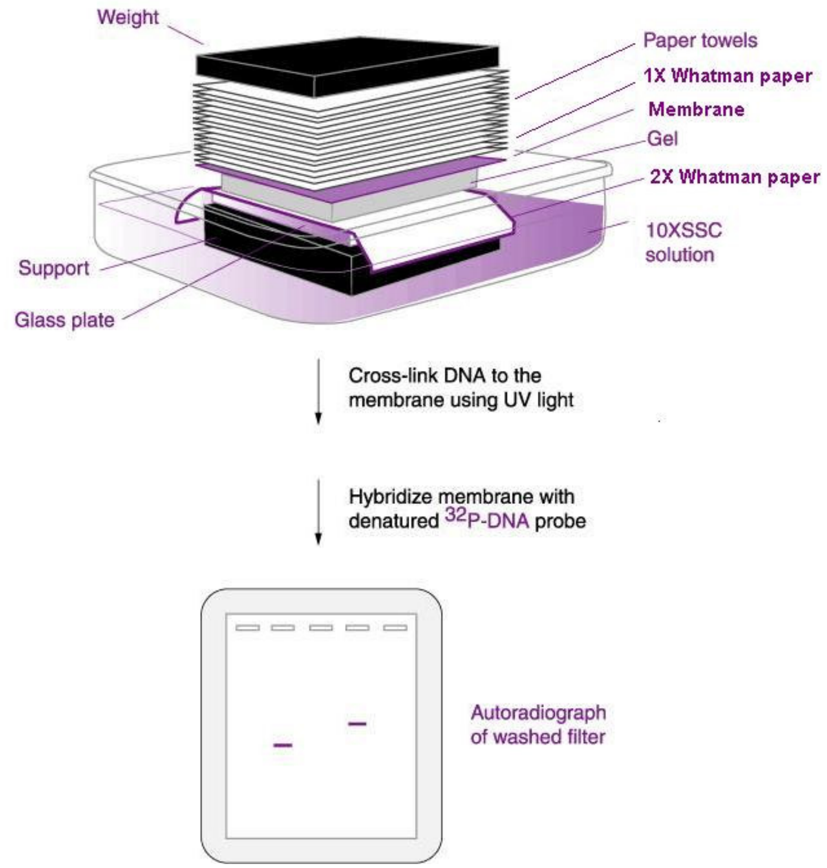

Weight
Paper towels
1X Whatman paper
Membrane
Gel
2X Whatman paper
10XSSC solution
Support
Glass plate
Cross-link DNA to the membrane using UV light
Hybridize membrane with denatured 32P-DNA probe
Autoradiograph of washed filter

Figure 2-11: Northern blotting assembly. (Figure taken and modified from the University of Arizona website, Department of Biochemistry & Molecular Biophysics, Professor Roger L. Miesfeld)

5X OLB Buffer preparation

Solution O: (1.25mM TrisHCL pH 8.0 and 125 mM MgCL2x6H2O)	
TrisHCL 1M	125 µl
MgCL2x6H2O	2.5 g
aqua bidest.	Up to 100 ml

Autoclave the solution

Solution A	
Solution O	5 ml
dNTP (100mM each)	50 µl
Beta mercaptoethanol	90 µl

Solution B: (2M HEPES pH 6.6)

- Solve 47.6 g HEPES in 50 ml aqua bidest.
- Adjust the pH to 6.6 using KOH (1M)
- Add aqua bidest. up to 100 ml

- Sterilize by filtering

Solution C:

Solve 50 OD of Pd $(N)_6$ in 7.5 ml TE (pH7.5).

5X OLB:

- Add 5ml of solution A
- Add 12.5 ml of solution B
- Add 7.5 ml of solution C
- Mix well
- Sterilize by filtering

Probe labelling

- Label probes with 32[P]dCTP using Klenow enzyme and random hexamer primers as follow:

5X OLB	4µl
Probe	20 ng
Water	Up to final volume of 20 µl

- Incubate the mix for 5 min at 97°C,
- Transfer on ice and spin shortly
- Add the following reagents:

α32 dCTP	2-4µl (depending on the amount of radioactivity)
Klenow enzyme	1.5 µl

- Incubate in 37°C for 45 minute.
- Meanwhile preheat hybridisation UltraHyb buffer (Ambion, Austin, TX, USA) at 45°C.
- After homogenizing thoroughly, pour approximately 2 or 3 ml of hybridisation buffer (depending on the size of the tube) in an appropriate tube and insert membrane in such a way that the RNA side face the lumen of the tube.
- Close the lid tightly and preheat for approximately 20 min.
- After 45 min incubation of the probe, purify the labelled fragments using illustraTM Microspin G-50 Columns (GE Healthcare Life Sciences, Product code: 27-5330-01) based on the provided manufacturer's protocol.
- Add the purified labelled probes to the preheated membrane with UltraHyb buffer and hybridise overnight at 42°C in a rotating glass tube.

- Wash membrane with washing buffer (2x SSC and 0.1x SDS in DEPC water) three times by shaking slowly for 20 minutes.
- Expose the Northern blots to Fuji Medical X-Ray films at -80°C for 4 h up to 4 days (depending on the amount of radioactivity on the blot) or analyse using a Storm 820 imaging system (APBiotech, Piscataway, NJ, USA).
- To control for RNA loading, re-probe blots with a β-actin probe (BioChain, Hayward, CA, USA).

In case of high background the blots were washed with stringent buffer (0.2x SSC and 0.1x SDS in DEPC water) and re-exposed to X-Ray films again afterwards.
In order to re-hybridise membranes with another probe, the blots were rinsed in boiled stripping buffer (0.01x SSC and 0.1x SDS in one liter DEPC water) and were left to reach to the room temperature.

2.27 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 2-20) for a normal cDNA on the agarose gel.

Table 2-20: PCR Program used for checking primers.

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

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 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, Passive Reference, and optimized buffer components.

Table 2-21: Reaction protocol for a 96 well plate

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

Standard curves as series of 2 fold dilutions were produced for the loading control (or 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 (Figure 2-12).

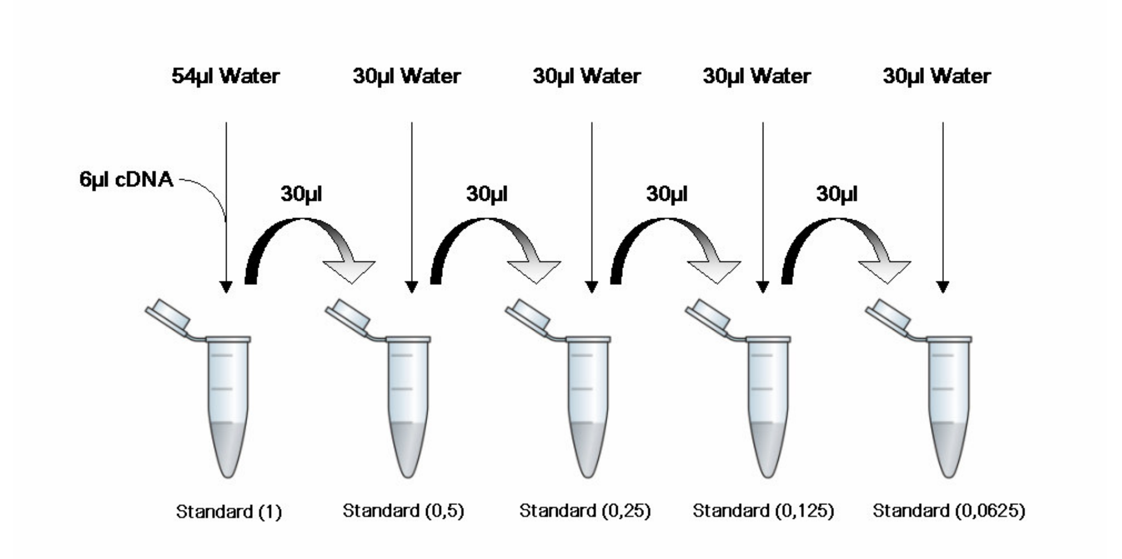


Figure 2-12: Standard curve dilutions preparation

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 (Table 2-22) was generated using the SDS2.1 software (AB applied biosynthesis). Experiments were performed in an ABI (PRISM 7900 HT) 96 well machine.

	1	2	3	4	5	6	7	8	9	10	11	12
A	Patient 1	Patient 2	Patient 3	Patient 4	Patient 5	Patient 6	Patient 7	Patient 8	Patient 9	Control 1	Control 2	Control 3
	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**
B	Control 4	Control 5	Control 6	Control 7	Control 8	NTC	NTC	S (1)	S (0.5)	S (0.25)	S (0.125)	S (0.0625)
	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**
C	Patient 1	Patient 2	Patient 3	Patient 4	Patient 5	Patient 6	Patient 7	Patient 8	Patient 9	Control 1	Control 2	Control 3
	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**
D	Control 4	Control 5	Control 6	Control 7	Control 8	NTC	NTC	S (1)	S (0.5)	S (0.25)	S (0.125)	S (0.0625)
	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**
E	Patient 1	Patient 2	Patient 3	Patient 4	Patient 5	Patient 6	Patient 7	Patient 8	Patient 9	Control 1	Control 2	Control 3
	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**
F	Control 4	Control 5	Control 6	Control 7	Control 8	NTC	NTC	S (1)	S (0.5)	S (0.25)	S (0.125)	S (0.0625)

Table 2-22: PCR Plate: there are triplicate reactions for 9 patients, 9 controls, 2 negative controls (NTC) and standard curve dilutions (S).

The produced data files were analysed using the SDS 2.1 software followed by T Test and standard deviation calculations in EXCEL.

2.28 Western blotting

2.28.1. Cell lysate preparation

Cell lysates were prepared using the following protocol:

- Prepare cell lysation buffer (Table 2-23)

Table 2-23: Cell lysation buffer

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

- Apply 3µl 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.

2.28.2. 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 2-25) was performed, using a Biorad 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 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 2-24) 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 (Table 2-24) 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 (table 2-24), 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 (Bromophenol blue) in the protein-loading buffer.
- When the dye front comes close to the end of the gel, stop the electrophoresis.

Table 2-24: 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 2-25: Component volumes for SDS-PAGE gels (in ml)

Stocks	10ml of 10% separating gel	5ml of 4% stacking gel
H2O	4.1	3.075
1.5M Tris-HCl, PH 8.8	2.5	---
0.5M Tris-HCl, PH 6.8	---	1.25
20% (w/v) SDS	0.05	0.025
Acrylamide/Bis-acrylamide (30%/0.8% v/v)	3.3	0.67
10% (w/v) APS*	0.05	0.025
TEMED*	0.005	0.005

- APS and TEMED were added just prior to pouring the gels.

2.28.3. 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 (Figure 2-13).
- 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.

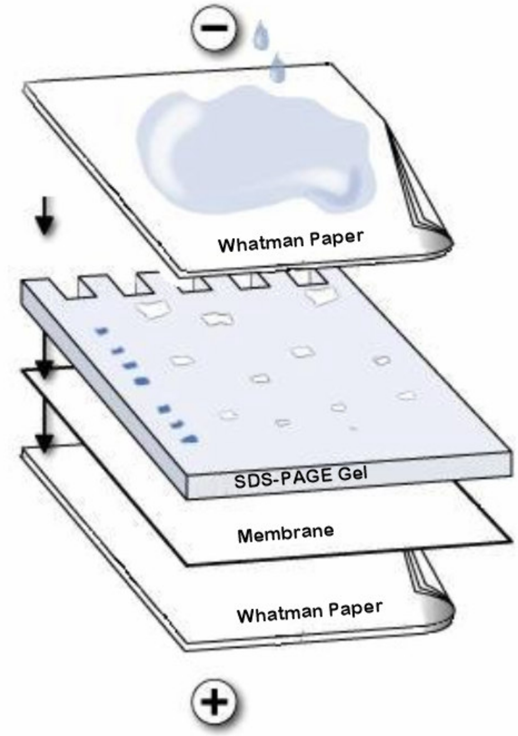


Figure 2-13: Western blotting assembly (Figure taken and modified from Millipore website: http://www.millipore.com/immunodetection/id3/westernblottingprotocols).

- Connect the apparatus to a power supply and perform electro-transfer at a constant current of 50 mAmp (for a single gel with 10X7 cm) 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.

Table 2-26: Buffers and solutions required for Western blotting:	
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

2.29 Knockdown expriments

Experiments were performed using the following protocol:

Day 1: Plate 200.000 U2OS cells per well in a 6 well falcon plate and add 2 ml DMEM.

Day 2: **Transfection:**

- Prepare following solutions (A and B), mix gently and incubate at RT for 5 min

Solution A	Solution B
10 μl of specific siRNA (20μM) + 202 μl OptiMEM	4 μl Oligofectamin + 56 μl OptiMEM

- Combine solution A and B (212μl +60μl) and mix gently.
- Incubate for 25 min at RT to let the complexes form.
- Slowly add 272 µl of the appropriate siRNA/Oligofectamine complexes per well of the cells.
- Incubate cells over-night in cell culture incubator

Day 3 (24h later): Replace medium and repeat the transfection procedure.

Day 4: (24h later): Process cells

The used siRNAs sequences are depicted in Table 2-27.

Table 2-27: siRNA sequences	
siRNA	**Sequence**
MCPH1-2	CTCTCTGTGTGAAGCACCTT
MCPH1-Xu4	GATGATGTACCTATTCTCTTA
MCPH1-Xu1	AAAGGAAGTTGGAAGGATCCA
CAP-G	GTCTCATGAAGCAAACAGCTT
Control	TTCTCCGAACGTGTCACGTTT

In order to have enough cells for RNA extraction and check point assay performing all of the reactions were performed in duplicate as follow:

Day 1: Plate 12 x 500.000 U2OS cells in 25 cm² flasks and add 5 ml DMEM.

Day 2: Transfection:

- Prepare the following solutions (A and B)

Reaction A:

❑ Add 55 μl of each specific siRNA (20μM) [MCPH1-2, MCPH1-Xu1, MCPH1-Xu4, CAP-G (Condensin I), Control-siRNA, Mock]

❑ Add 1111 μl OptiMEM

❑ Mix gently

❑ Incubate at room temperature for 5 min

Reaction B:

❑ Add 22 μl Oligofectamin (for making Mastermix: 145.2μl Oligofectamin)

❑ Add 308 μl OptiMEM (for making master mix 2032.8 μl OptiMEM)

❑ Mix gently

❑ Incubate at RT for 5 min

- Combine solution A and B (1166 μl +330 μl) and mix gently
- Incubate for 25 min at room temperature to let the complexes form.
- Slowly add 935 µl of the appropriate siRNA/Oligofectamine complexes per flask.
- Incubate cells over-night in cell culture incubator

Day 3: Replace the medium and repeat the transfection procedure.

Day 4:

- Extract RNA from a part of the cells (at least 1 milion cells using RNAeasy Kit)
- Feed rest of the cells and incubate them for one more day

Day 5:

- Extract RNA from a part of the cells (at least 1 milion cells using RNAeasy Kit)
- Perform checkpoint assay on the rest of the cells

2.30 Radiation assay

- Irradiate cell with 4 Gy,
- After 1.45 h trypsinize cells
- Transfer 1.000.000 LCL's to a Falcon (blue cap, 15 ml)
- Centrifuge: 10 min, 500 g, RT
- Discard the supernatant (leave approx. 0.1 ml)

Fixation:

- Add 1 ml 2% PFA (in 1xPBS) and resuspend. Proceed slowly

- Incubate in a water bath, 37°C 10 min.
- Centrifuge: 10 min, 900 rpm
- Discard the supernatant (with a pipette!!) up to 0.1 ml and resuspend
- Place on ice (with an angle of 45°)

Permeabilization:

- Add 0.9 ml of Methanol 100% (final concentration 90%)
- Incubate for 30 min on ice (-4°C).
- Store cells over the weekend at -20°C
- Sending the cells on dry ice to Würzburg

Staining: (this part performed by our collaborators in Würzburg):

- Centrifuge: 500 g, 5 min, RT
- Wash in 3 ml of 5% BSA (in PBS)
- Centrifuge, discard completely the supernatant (with a pipette)
- Add 0.1 ml 5% BSA (in PBS)
- Incubate 10 min at RT
- Add 1.5 µl ɑ-H3-P (#9708, Alexa Fluor 488, Cell Signaling). Incubate 1h at RT (in a dark chamber)
- Add DAPI (final concentration: 2 µg/ml)
- Incubate 30 min at 4°C in a dark chamber
- Make the measure in the Flow Cytometer

2.31 Concentration Measurements

2.31.1. 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 mm^2.

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.

2.31.2. DNA and RNA concentration assay

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

2.31.3. Protein assay

Bradford assay

A starting 10 µg/ml BSA solution was prepared by diluting a 5 mg/ml BSA stock solution 1:500 in the working buffer (specific for each experiment, Table 2-23), which had been previously diluted 1:1000 in bidest water (2 µl 5 mg/ml BSA + 998 µl 1:1000 buffer).

BSA standard solutions were prepared by diluting the 10 µg/ml BSA solution in 1:1000 working buffer as indicated in Table 2-28.

Table 2-28: Standard BSA curve preparation

BSA standard curve (Name-µg/ml)	BSA 10 µg/ml (µl)	1:1000 buffer (µl)
S6-10	---	---
S5-7.5	150	50
S4-5	100	100
S3-4	80	120
S2-2	40	160
S1-1	20	180

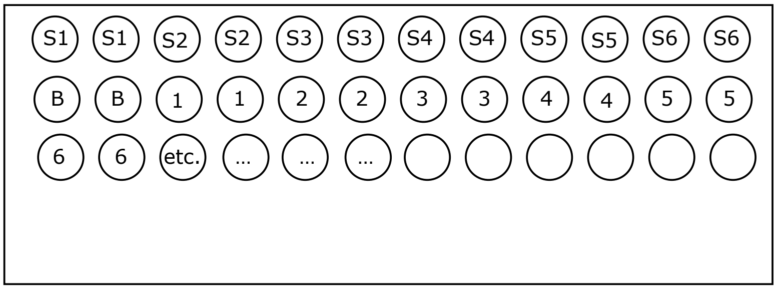


Figure 2-14: Position of samples in falcon plate.

2 x 80 µl of each sample, a blank (only 1:1000 buffer) and the standards (S) were placed into a Falcon microtiter plate as shown in Figure 2-14. Then 20 µl of Bradford reagent were added with a multi-channel pipette and the samples were extensively mixed by pippeting up and down. The reaction was allowed to proceed for 1-5 min. Bubbles were carefully removed by taping the plate or using pippet tip and the absorbance at 595 nm of each probe was measured in an anthos 2020 spectrophotometer (anthos), which also calculated the protein concentration by correlation to the BSA standard curve.

BCA assay

BSA standard solutions were prepared by diluting BSA in a range of 1µg to 10 µg in 10 µl of the buffer used for initial extraction of the sample proteins. 1 µg /µl BSA were diluted as follow Table 2-29:

Table 2-29: Standard BSA curve preparation

Sample volume	1µl					
BSA (1µg/µl)	1µl	2 µl	4µl	5 µl	7.5 µl	10µl
Extraction buffer	9	8	6	5	2.5	0
Total protein content	1µg	2µg	4µg	5µg	7.5µg	10µg
Protein concentration	0.1 µg/µl	0.2µg/µl	0.4µg/µl	0.5µg/µl	0.75µg/µl	1µg/µl
Dilution factor	10	5	2.5	2	1/0.75	1
Sample protein conc.	1µg/µl	1µg/µl	1µg/µl	1µg/µl	1µg/µl	1µg/µl

Standard BSA curve preparation

- Pipette the samples into the wells of 96 well plates.
- Calculate the total amount of reagent needed depending on the number of the samples (total volume of the final detection solution needed for one well of a 96 well plate is 200 µl).
- Mix 50 parts of reagent 2 with 1 part of reagent 1 (BCA Protein Assay Kit, Thermo Scientific, Prod. No. 23250).
- Mix well and add 200 µl to the each sample.
- Incubate the 96 well for 30 min at 37 degree in moisture chamber (simply a box laid out with wet paper towels).

Before measure the protein concentrations, wait 5 min for the samples to cool down to room temperature.

2.32 Sequence Logos

A CA8 protein multiple alignment from Human (*Homo sapiens*; NP 004047), dog (*Canis familiaris*; XP 544094), cow (*Bos taurus*; NP 001077159), horse (*Equus caballus*; XP 001496523), mouse (*Mus musculus*; NP 031618), rat (*Rattus norvegicus*; NP 001009662), opossum (*Monodelphis domestica*; XP 001368351), chicken (*Gallus gallus*; XP 419221), frog (*Xenopus (Silurana*) tropicalis; NP 001011213), trout (*Oncorhynchus mykiss*; NP 001118116), zebra_sh (*Danio rerio*; NP 001017571), and sea urchin (*Strongylocentrotus purpuratus*; XP 795365) was prepared with ClustalX (Thompson and others 2002) and sequence logos were prepared using texshade (Beitz 2000). This analysis was performed by Dr. Peter Nick Robinson from Charite-Universitätsmedizin Berlin.

3 Results

3.1. Linkage results

For this study, 135 Iranian families with a minimum of two mentally retarded children were selected for linkage analysis after exclusion of Fragile-X syndrome, large chromosomal aberrations and fatty-, amino or organic acid metabolic disorders.
For these families, whole genome SNP typing using the Human Affymetrix Mapping 10k, 50k, 250k or Illumina 6k SNP array was performed with DNA from all affected individuals, the parents and (if available) a maximum of two healthy siblings. The genotype data from each family were then used for linkage analysis.
For a given family, intervals were considered as potential sites of the causative mutation if the corresponding LOD scores were less than one unit lower than the highest peak observed ('one LOD down method'), and if all affected members were homozygous carriers of the same SNP haplotype. 16 families were excluded from the analysis or postponed because of inconsistencies, lacking linkage results or missing samples.
86 families yielded more than one linkage interval, reflecting the limited size of most of these families and/or high degree of consanguinity (Appendix-A).

In first-cousin marriages with two affected children -the minimum size of the families considered in this study- parametric linkage analysis typically resulted in 5–6 intervals, each with the maximally attainable LOD score of 1.8, but in many of them, the number of peaks could be reduced to about 3 by including all available unaffected siblings or other family members.

To study the genomic distribution of these intervals, and to get a first impression as to the number and localization of genetic defects causing NS-ARMR in the Iranian population, a method was used that had been employed previously to assess the distribution of gene defects underlying non-syndromic X-linked MR on the human X chromosome (Ropers and others 2003).

Table 3- 1: Number of analyzed families

	Families with solitary intervals	Families with more than one intervals	Excluded	Total
Our study	33	86	16	135
Another study in our lab	7	34	7	48
Total	40	120	23	183

The curve shown in Figure 3-1 results from the superposition of weighted linkage intervals from 78 consanguineous families with NS-ARMR including 37 families from this analysis and 41 families analysed in the context of another study conducted in parallel in our lab (Motazacker Thesis; 2008).

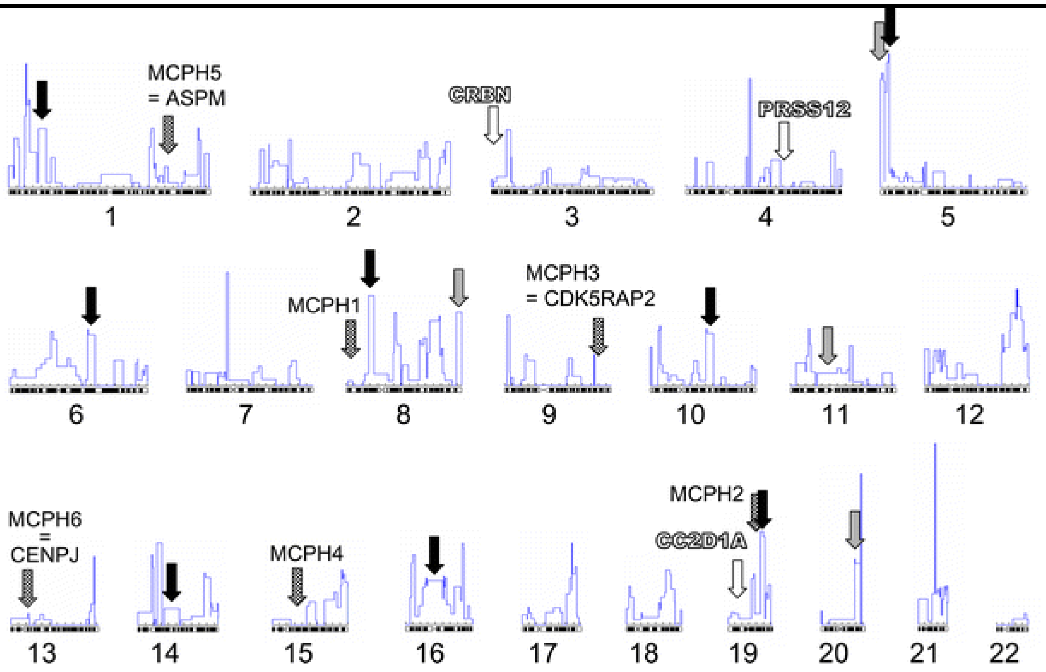


Figure 3-1: Distribution of linkage intervals reveals heterogeneity of NS-ARMR and defines novel MRT loci. The curve results from superposition of weighted linkage intervals from 78 consanguineous families with NS-ARMR, as previously described for families with X-linked MR (Ropers and others 2003). Large ARMR families with single co-segregating intervals are represented by rectangles of different shape, but identical surface, or 'weight'. Black arrows: single intervals with LOD scores >3 (= novel MRT loci), grey arrows: single intervals with LOD scores between 2 and 3. In small families with several co-segregating haplotypes, their cumulative length was used to normalize their weight. For this calculation, linkage intervals were disregarded if their LOD scores were more than 1 unit lower than those of the highest LOD score observed in that family. The surface under the curve is a parameter for the proportion of gene defects mapping to the relevant genome segment. Empty arrows: cloned MRT genes, checkered arrows: loci for ARMR with microcephaly (Najmabadi and others 2006).

Large ARMR families with single co-segregating intervals, as exemplified by the interval at the distal end of chromosome 8q in Figure 3-1, are represented by rectangles of different shape, but identical surface, or 'weight'. For families with more than one interval, the lengths of all physical genome intervals (as defined by the distance between the closest SNP markers flanking the respective haplotypes) co-segregating with ARMR were added up and their total 'weight' was normalized in accordance with the fact that each family represents one single mutation. Thereafter, the weighted linkage intervals from all families (including the 41 families that were analyzed in parallel study) were graphically superposed. The resulting curve reflects the genomic distribution of the mutations causing ARMR in 78 families. Interval positions for the individual families are listed in appendix-A.

The surface under the curve, covering a given region of the genome, is a parameter for the proportion of gene defects mapping to the respective genome segment. Single intervals from individual large families appear as bars of different height, but all with identical surfaces under the curve.

Given the considerable number and length of linkage intervals in small families, it is not surprising that some of these overlapped. Still, it appears that their co-localization on some chromosomes is not completely random. For example, apparent clustering of linkage intervals was observed on chromosomes 16 and 19, which may indicate that these regions carry genes that are mutated in more than one family.

3.1.1. Linkage analysis results of 135 Iranian Families

The data presented above are merely the result of the first round of studies into the genetic causes of ARMR; later on we increased the number of families to 183 [in addition to 41 families from (Motazacker Thesis; 2008)]. 135 families out of 183 were analysed in the course of this study which linkage analysis results for 119 out of them (16 were excluded from the analysis) are presented in appendix-A.

3.1.2. Identification of 31 new mental retardation (MRT) loci

In 33 out of 135 families, single autosomal linkage intervals with LOD scores above 2 were identified (Table 3-2 and Fig a-1 to Fig a-33 in Appendix-B) 31 of which were novel.

18 of the 31 families with novel solitary linkage intervals showed LOD scores above 3 (Morton 1998), (Table 3-2 and Fig S-1 to Fig S-33 in Appendix-B). They can therefore be considered as new Mental Retardation (MRT) loci, but also the remaining 13 with LOD scores between 2 and 3 are still likely to contain the causative mutation, as they represent the only genomic regions that co-segregate with the disease. Therefore they have the highest probability of harboring the causative mutation in the respective family.

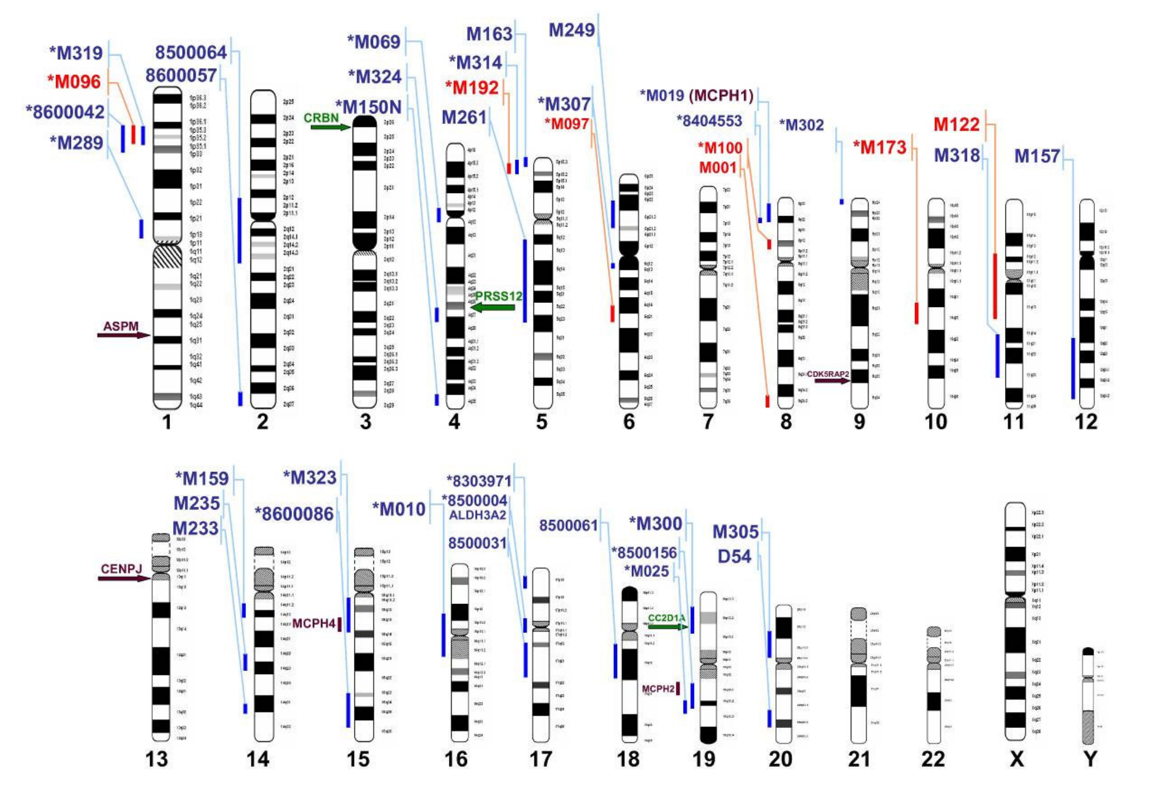


Figure 3-2: Whole genome view presentation of all the 40 found solitary linkage intervals. Regions with LOD scores above three are depicted by asterisks. Intervals shown in red (7) are from the other study that has been conducted in our lab parallel to this study. Locations for all known genes for ARMR are shown by green arrows. Locations of 4 known genes and 2 loci for primary microcephaly are shown by brown arrow and bars.

5 out of these 31 novel loci have already been published (MRT9–11; Najmabadi and others 2006) together with the less significant intervals of families M163 and D54.

Linkage analysis results and haplotypes of the 33 families with solitary peaks are shown in Fig S-1 to Fig S-33 in Appendix-B. Parametric (LOD score) and non-parametric (NPL) LOD scores as calculated by the Merlin (Abecasis and others 2002), Allegro (Gudbjartsson and others 2000) or GeneHunter (Kruglyak and others 1996) software are shown in genome-wide linkage plots of the analyzed families. Inferred haplotypes using GeneHunter and Merlin programs were visualized using the Haplopainter software (Thiele and Nurnberg 2005).

Table 3-2: Physical position and flanking SNP markers for standalone families:

Family	Chr.	Start SNP	Position [bp]	End SNP	Position [bp]	Length [Mbp]	LOD Score
8404553 (3) (4)	8	rs613566	12944303	rs11203893	17579537	4.6	6.3
M010 (1) (3)	16	rs724466	22705353	rs3901517	48948887	26.2	5.2
M019 (4)	8	rs718122	3146756	rs725944	14168140	11.0	4.2
M319 (3)	1	rs724321	38896190	rs1934393	48981205	10.0	4.2
M302 (4)	9	rs1532309	593192	rs4131424	4325918	3.7	4.1
M025 (1)	19	rs2109075	46844069	rs8101149	52292281	5.4	4.0
8500156	19	rs11881580	38339219	rs17727484	49668378	11.3	4.0
8600086	15	rs936227	72919012	rs4238484	90871916	17.9	3.9
8600042	1	rs16825353	38907775	rs207149	55579847	16.7	3.7
M314	5	rs2008927	4007570	rs60701	10786776	6.7	3.5
M300	19	rs4807852	6189183	rs731617	16526857	10.3	3.4
M289 (2)	1	rs10494061	107511901	rs1938250	120495870	12.9	3.4
M324	4	rs843571	121862949	rs318539	130052650	8.2	3.3
M159 (1)	14	rs1998463	26578858	rs243286	31700826	6.2	3.2
M323	15	rs1966109	21847665	rs614215	40159940	18.3	3.2
8600004	17	rs17648393	17195117	rs9303660	28444843	11.2	3.1
M150N	4	rs1158815	188094186	qter	191263063	3.2	3.1
8303971	17	rs1367950	3565047	rs1826925	8644145	5.0	3.1
M069	4	rs728293	47052440	rs1105434	57488508	10.4	3.0
M307	6	rs7753225	65492346	rs10498840	66715234	1.2	3.0
M163 (1)	5	pter	1	rs2115289	5019895	5.0	2.8
M249	6	rs911361	20523032	rs1885615	39446413	18.9	2.7
M318	11	rs1391221	85632927	rs1880206	114226220	28.5	2.7
8500031	17	rs7502685	33901979	rs9913816	52662679	18.7	2.7
8500061	18	rs11081675	26834675	rs4940195	43812253	16.9	2.7
M157	12	rs725421	86220791	rs2013160	126641435	40.4	2.7
M261	5	rs9291745	61327157	rs6866438	120718305	59.3	2.5
M233	14	rs1958843	86796902	rs2402074	91363271	4.5	2.5
M305	20	rs1028846	10960899	rs1888610	22854446	11.9	2.5
M235	14	rs731700	58334879	rs178384	79252594	20.9	2.5
D54 (1)	20	rs756529	47444415	rs728504	55768572	8.3	2.4
8500064	2	rs13033902	77117862	rs294669	123280342	46.1	2.4
8600057	2	rs13017098	240751883	qter	242951149	2.2	2.5

[1] These families were included in (Najmabadi and others 2006) as new or potential loci.

[2] Consanguinity is not clear, analysis were performed by assuming first cousin degree of consanguinity

[3] Result of split-pedigree-analysis

[4] Published as Garshasbi et al. (2008), Garshasbi et al (2006) and Abbasi-Moheb et al. (2007).

3.1.3. Overlapping regions of autozygosity

By superposing all the intervals for 183 families, we found 3 overlapping regions among families with solitary intervals on chromosomes 1, 5 and 19.

The region on chr1p34.3-p34.1 spans 7.1Mbp and is common between the solitary linkage intervals of families M096, 8600042 and M319 and one of the several linkage intervals of families M037 and M8500320 (Figure 3-3).

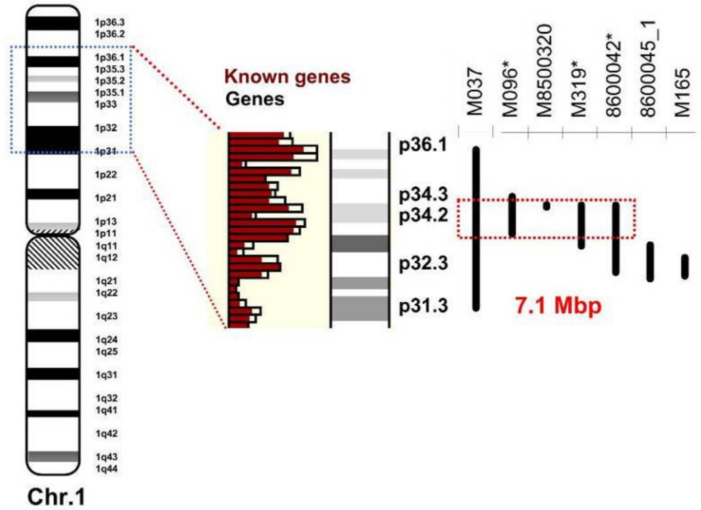


Figure 3-3: Overlapping region on chromosome 1. Three families with solitary linkage interval (M096, M319 and 8600042) are indicated by asterisks. The red rectangle shows the overlapping region (7.1 Mbp at chromosome 1p34.3-p34.1) between the solitary linkage interval of these 3 families and one of several intervals of two additional families with more than one linkage interval.

On chromosome 5p a total of six families show overlapping linkage intervals four of which were solitary and three had a significant LOD score. In three different parts of this region (depicted as A, B and C) at least 3 of the four families with solitary intervals show overlaps (Figure 3-4).

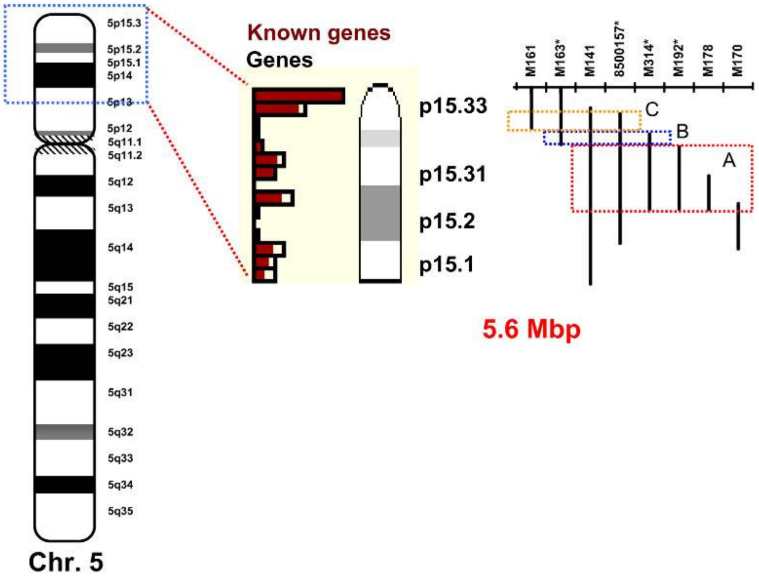


Figure 3-4: Three separate overlapping regions on chromosome 5. Four families with solitary linkage interval (M163, 8500157, M314 and M192) are indicated by asterisks. The regions covered by rectangle A (5.6 Mbp at chr5p15.32-p15.2) and B (1.1 Mbp at chr5p15.33-p15.32) show an overlap between the solitary linkage interval of three families and one of several intervals of an additional families. Rectangle C (1.8 Mbp at chr5p15.33) marks the overlap between the solitary interval of two families and one of several intervals of an additional families.

Finally two families with solitary linkage intervals (M025 and 8500156) and 4 families with more than one linkage interval overlapping a region of 3.5 Mbp on chromosome 19q13.2-q13.31.

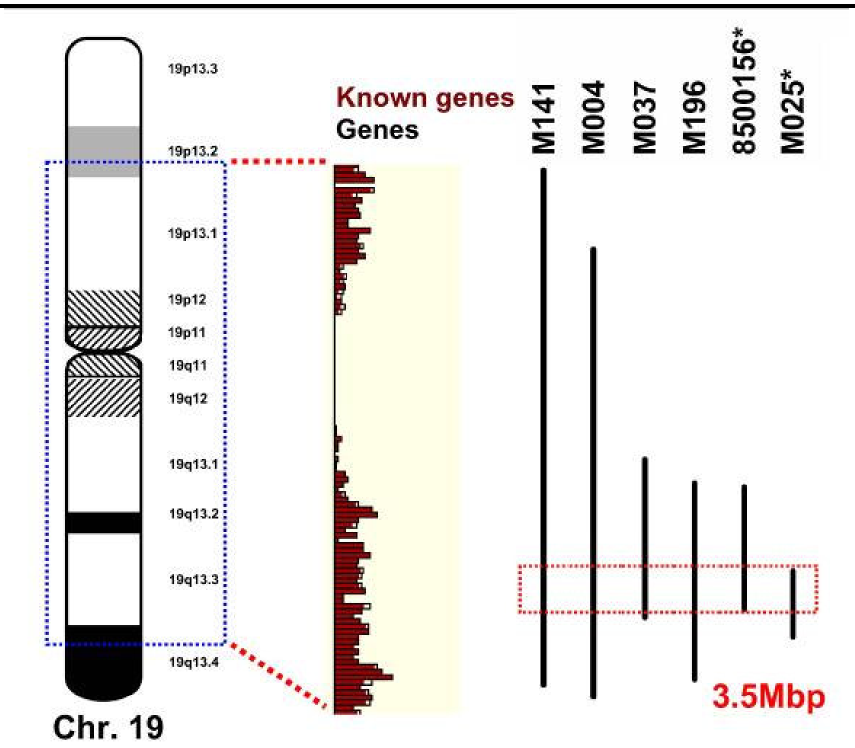


Figure 3-5: Overlapping region on chromosome 19. Two families with solitary linkage interval (M025 and 8500356) are indicated by asterisks. The red rectangle shows the overlapping region (3.5 Mbp at chromosome 19q13.2-q13.31) between the solitary linkage intervals of these 2 families and one of several intervals of four additional families with more than one linkage interval.

3.1.4. Overlapping intervals with known ARMR genes

One of the linkage intervals of family M346 (2 adjacent intervals on chromosome 4) encompassed the neurotrypsin precursor *PRSS12* (previously identified gene for ARMR (Molinari and others 2002). However, sequencing of the exons and exon-intron boundaries of this gene did not reveal a mutation.

The linkage intervals of families M225 and M300 (solitary interval with LOD score 3.4) on chromosome 19 (Figure 3-6) encompassed *CC2D1A,* previously implicated in NS-ARMR by Basel-Vanagaite and others (2006), but also for this gene mutation screening in families did not reveal any sequence changes in the exons and exon-intron boundaries.

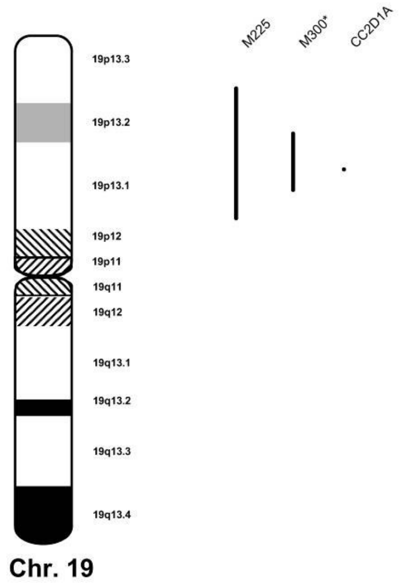


Figure 3-6: Linkage intervals overlapping *CC2D1A*. Linkage intervals for families M225 (with 2 additional intervals) and M300 (solitary interval with LOD score 3.4) which they overlap with the *CC2D1A* (previously known gene for NS-ARMR) are indicated.

3.2. Mutation Screening

Mutation screening was performed for some of the families with solitary linkage intervals. In addition a few selected candidate genes (based on the phenotype or functions of the genes) were investigated in families with several linkage intervals.

Prior to screening, the genes in a given linkage interval were ranked based on their expression patterns and functional relevance in the central nervous system according to the available literature information. In some cases prioritization was performed with the help of softwares and databases such as PosMed (PosMed home page), Endeavour (Aerts and others 2006) and Prioritizer (Prioritizer home page). Table 3-3 shows the list of genes for which mutation screenings were performed.

Table 3-3: Genes that were screened for mutations

Family	GENES
8303971	*SHBG* and *MPDU1*
8401973	*A2BP1*
8404553	*TUSC3, MSR1, FGF20, FFHA2, ZDHHC2, CNOT7, VPS37A, MTMR7, SLC7A2, PDGFRL, MTUS1, SGCZ, c8orf58* and *DLC1*
8600004	*ALDH3A2*
8600013	*KCNG1* and *DPM1*
D54	*DPM1*
M005	*FLT1, KIAA0774, SLC7A1, UBL3, KATNAL1, HMGB1, LOC728437* and *c13orf12*
M017N	*BDKRB1* and *BDKRB2*
M019	*MCPH1*

Table 3-3: Genes that were screened for mutations

Family	GENES
M069	*GABRB1, REST, NMU, AASDH, CORIN, KIT, OCIAD2, PDGFRA, CNGA1, CEP135, CHIC2, CLOCK, COMMD8, EXOC1, SRP72, TPARL, SCL10A4, TEC, SGCB, SCFD2, RASL11B, PPAT, PDCL2, HOP, ATP10D, DCUN1D4, GSH2, FLJ21511, FIP1L1, KDR, NPAL1, OCIAD1, PAICS, LOC402176, SRD5A2L2, USP46, LOC339977, NFXL1, SPATA18, TXK, ZAR1, LNX, SPINK2* and *ARL9*
M096	*SF3A3, NDUFS5, ZMPSTE, CAP1, IPO13, ELOVL1, FHL3, SIAT6* and *DMAP1*
M104	*DPAGT1*
M107	*CA8, RAB2A, TOX, ASPH, GRIK1, NCAM2, OLIG2* and *OLIG1*
M122	*FUCT1*
M147	*DPM1*
M150N	*FRG1, ZFP42, FLJ25801 (TRIML1), FLJ3, AY494056* and *AK095968*
M159	*SCFD1, FOXG1B, COCH, NUBPL, KIAA1333, c14orf126, STRN3, ARHGAP5, AP4S1, PRKD1, HECTD1, AKAP6* and *NAPS*
M163	*IRX1, IRX2, IRX4, SLC6A3, MRPL36, NDUFS6, TERT, SLC6A18, SLC6A19, CEI, CRR9, FLJ12443, SLC12A7, SEC6L1, SDHA, TPPP, SLC9A3, AHRR* and *Cep72*
M165	*MPI*
M180	*MPDU1*
M190	*CA8, RAB2A, TTPA, GGH, ASPH, BHLHB5, CHPP, ARMC1, MGC34646, FAM77D, PDE7A, COPS5, CRH, RRS1, TRIM55, VCPIP1, SGKL, DNAJC5B, ZNF537, ZNF536, c19orf2, DPY19L3, CCNE2, POP4* and *UQCRFS1*
M196	*CPT1C*
M225	*CC2D1A*
M233	*TTC8, GALC, GPR65, KCNK10, SPATA7, ZC3H14, PTPN21, EML5, FOXN3, TDP1, C14orf143, KCNK13, C14orf102, PSMC1, CALM1, TTC7B, RPS6KAS, C14orf159, GPR68, SMEK1, PP8961, CCDC88C, CATSPERB* and *TC2N*
M282	*MCPH1*
M300	*CC2D1A*
M307	*EGFL11*
M346	*PRSS12*
M347	*CHL1* and *CNTN6*

Apart from several SNPs and silent changes, so far no sequence changes were found in all of the screened genes except for 2 mutations in *MCPH1*, and one each in *TUSC3*, *CA8, ALDH3A2* and *CYP7b1* respectively. The latter, however, turned out to be a polymorphism.

3.3. Identification of a R237Q mutation in exon 7 of *CA8*

The investigation of family M107 - four affected sibs out of six as a result of a first cousin marriage- led to the identification of two regions with a maximum LOD score of 2.4 on chromosome 8 and 21 (Figure 3- 7 and Figure 3-9). One of the affected and one of the unaffected siblings had died owing to unknown causes. The affected persons in this family have ataxia, dysarthric speech, and mild mental retardation.

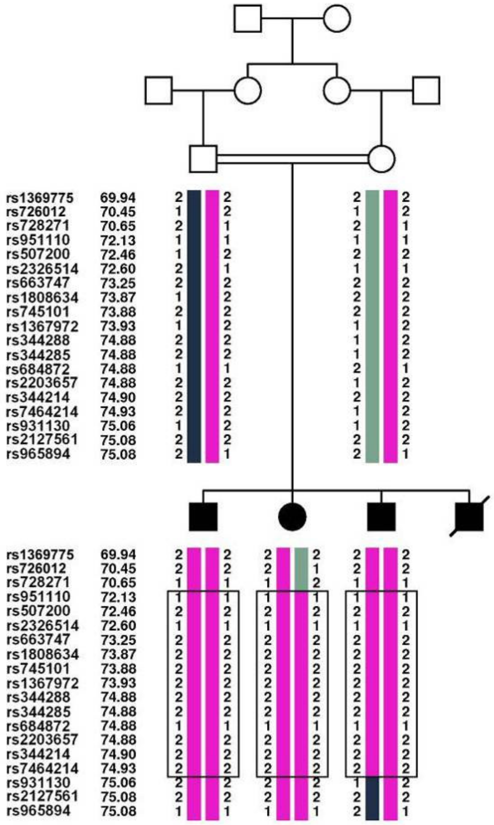


Figure 3- 7: Haplotype for the region on chromosome 8q12.1-q12.3 in family M107. The homozygous region in all the three affected children is indicated by rectangles. The first three adjacent heterozygous SNPs from both flancking sides of the interval are shown.

Magnetic resonance cerebral imaging performed on the younger surviving affected brother revealed normal appearing cerebral and cerebellar anatomy.

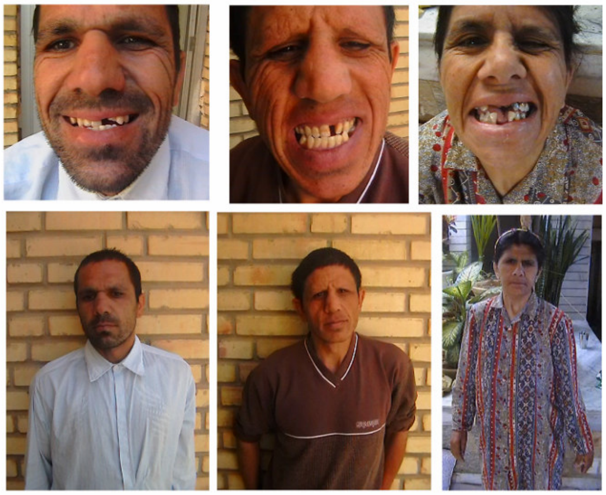


Figure 3- 8: Clinical photographs of the three living patients in family M107

The region on chromosome 8 was screened for mutations because at an early stage of our study it was the first example of a common region occurring between two families with only 2 peaks each (Figure 3-9).

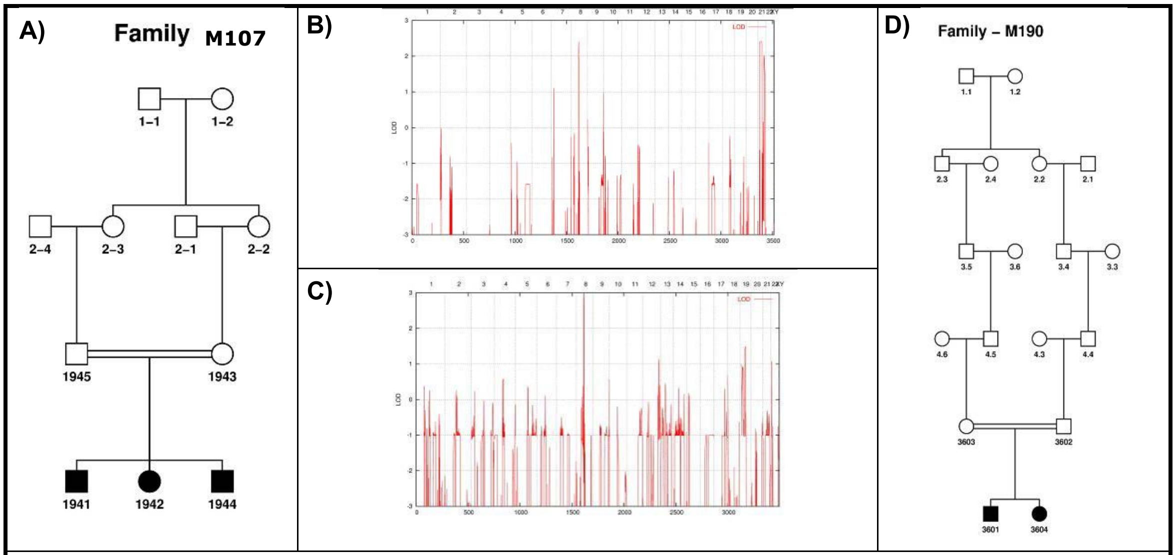


Figure 3-9: Pedigree and whole genome parametric linkage plots of families M107 (A and B) and M190 (C and D) with a common region on chromosome 8.

This led to the identification of a R237Q mutation -a non-synonymous change (CGA>CAA) - in a highly conserved region of carbonic anhydrase VIII (*CA8*), which cosegregated with MR in family M107 (Figure 3-10).

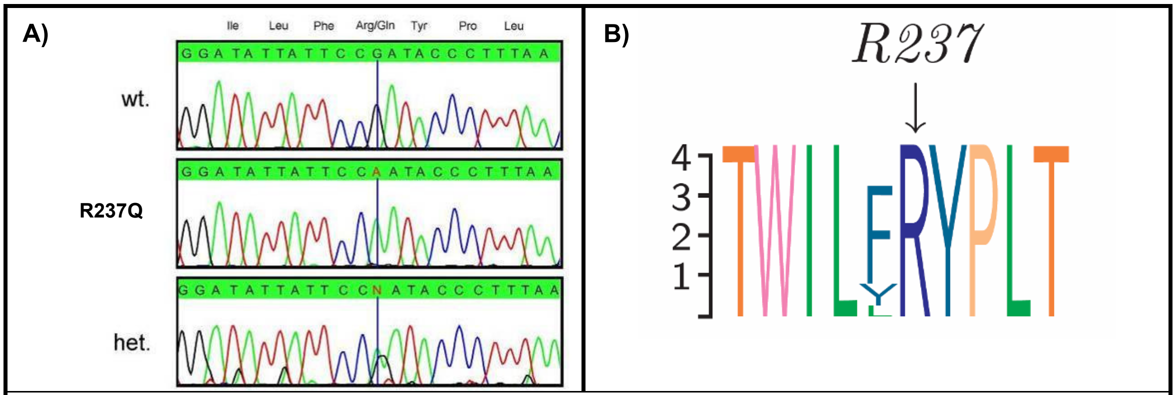
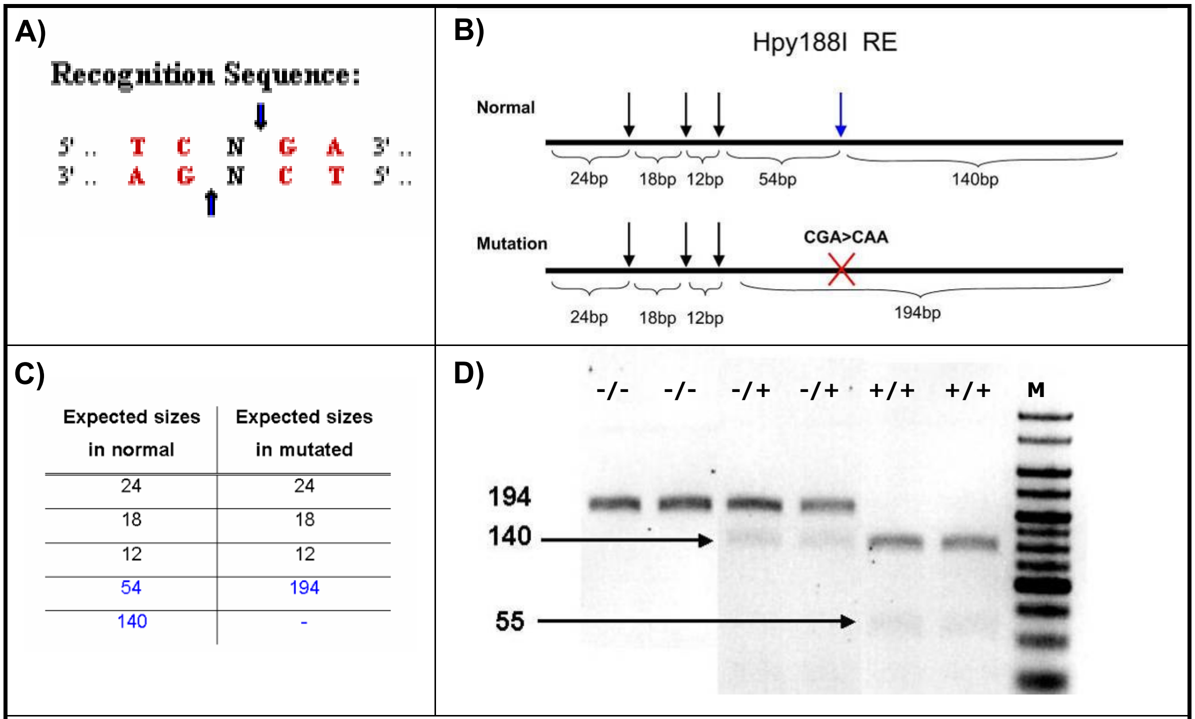


Figure 3-10: Sequence chromatogram and logo of alignment of the R237Q mutation in *CA8*. A) Chromatograms for the region of mutation in one control (wt), one patient (homozygous for R237Q) and one parent (het) are shown. B) Sequence logo of an alignment of *CA8* orthologs from 12 species from human to sea urchin showing the sequence affected by R237Q. Sequence logos were prepared using texshade (Beitz 2000).

In order to rule out the possibility that the change found in *CA8* might be a common polymorphism an RFLP assay using the Hpy188I restriction enzyme was designed (Figure

3-11) for checking controls and used to exclude this change in 384 Iranian control chromosomes as well as in 540 German control chromosomes (healthy blood donors).



Expected sizes in normal	Expected sizes in mutated
24	24
18	18
12	12
54	194
140	-

Figure 3-11: RFLP experiment for mutation in CA8. A) Hpy188I recognition site. B) Schematic representation of the Hpy188I recognition sites in the vicinity of the CGA>CAA change in exon 7 of *CA8*. The recognition site that is destroyed by the mutation is indicated by a blue arrow. C) Expected fragment sizes after restriction, for mutated and normal alleles. D) Restriction pattern for homozygous and heterozygous forms of the mutated and normal allele. The samples loaded in the first two lanes are homozygous for the mutated allele (-/-), the next two are heterozygous (-/+) and the last two before the marker lane (M) are homozygous for the normal allele (+/+).

3.4. Identification of a c.1107 +1delGTA mutation in *ALDH3A2*

Autozygosity mapping in family 8600004 (Figure 3-12) with five patients including a triplet (1 pair of monzygotic twins and one heterozygoute) led to one solitary region on chromosome 17p11.2-q11.2 with a maximum parametric LOD score of 3.2 (Figure 3-13).

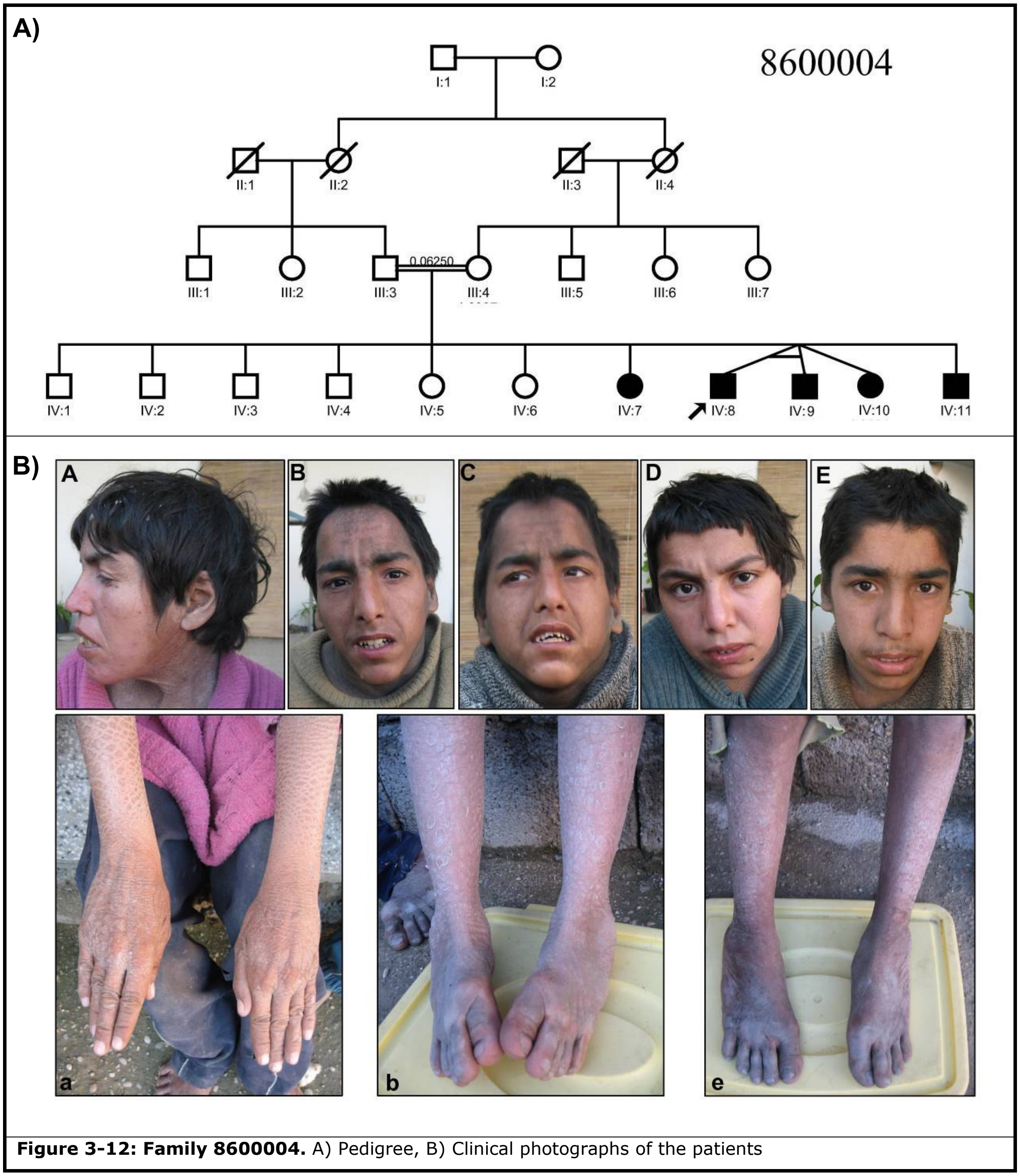


Figure 3-12: Family 8600004. A) Pedigree, B) Clinical photographs of the patients

This region included the Aldehyde dehydrogenase 3A2 isoform 2 (*ALDH3A2*) which is the responsible gene for Sjogren-Larsson syndrome (SLS).

SLS is an autosomal recessive neurocutaneous disorder characterized by a combination of severe MR, spastic di- or tetraplegia and congenital ichthyosis (increased keratinization). Ichthyosis is usually evident at birth, neurologic symptoms appear in the first or second year of life. Most patients have an IQ of less than 60. Additional clinical features include glistening white spots on the retina, seizures, short stature and speech defects (Gordon 2007). As we could see similar symptoms like severe MR, ichthyosis (hyperkeratosis), short stature and spastic paraplegia in our patients (Figure 3-12) we screened *ALDH3A2* for mutations and found a c.1107 +1delGTA mutation, which deletes the first three nucleotides after exon 7 (see Figure 3-14).

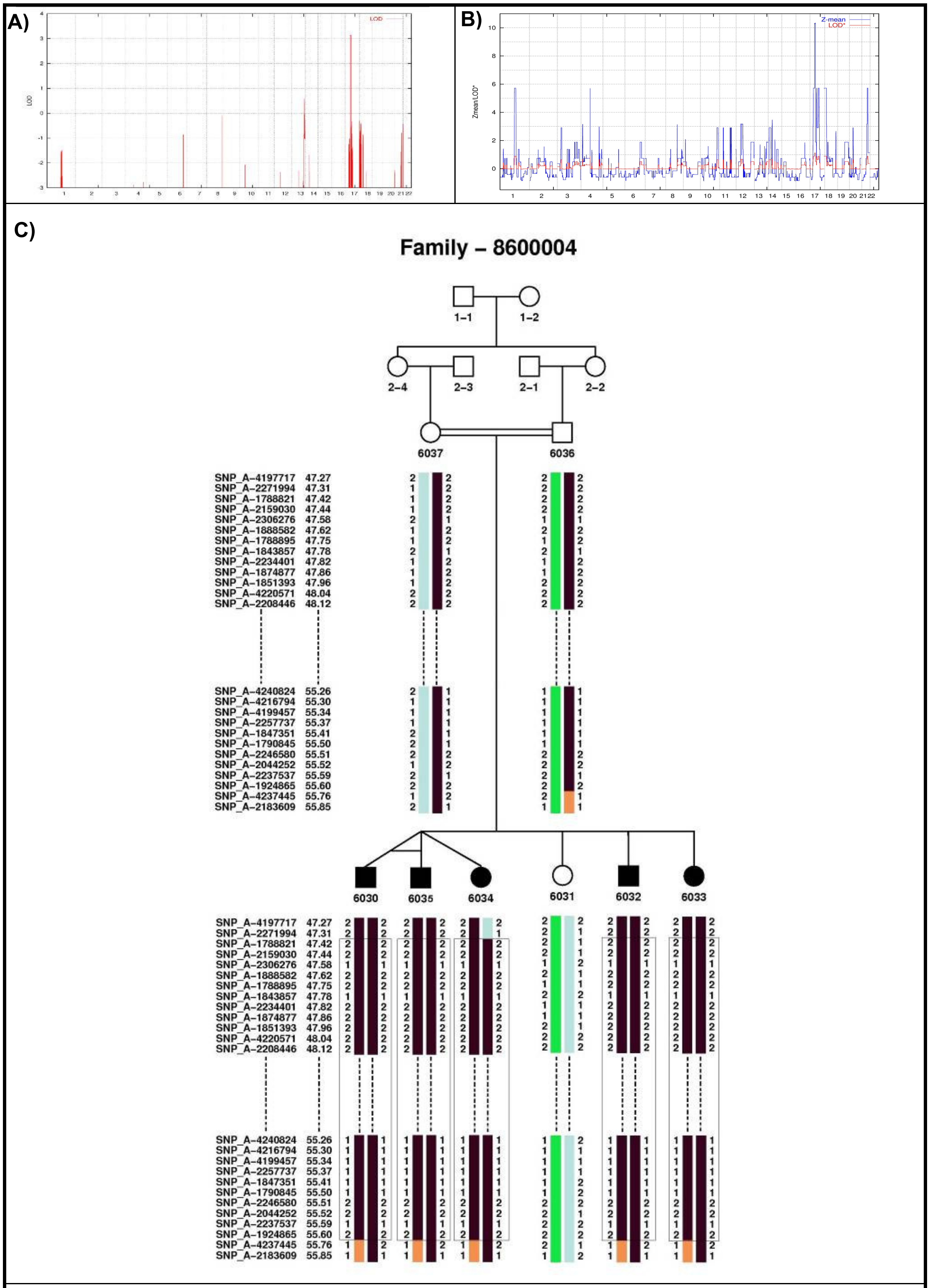


Figure 3-13: Family 8600004. Whole genome parametric (A) and non-parametric (B) linkage results (Merlin software) and haplotype of the only linkage interval with significant parametric LOD score of 3.13 (C). About 10 subsequent markers from both ends of the interval and the first adjacent heterozygous SNPs are shown.

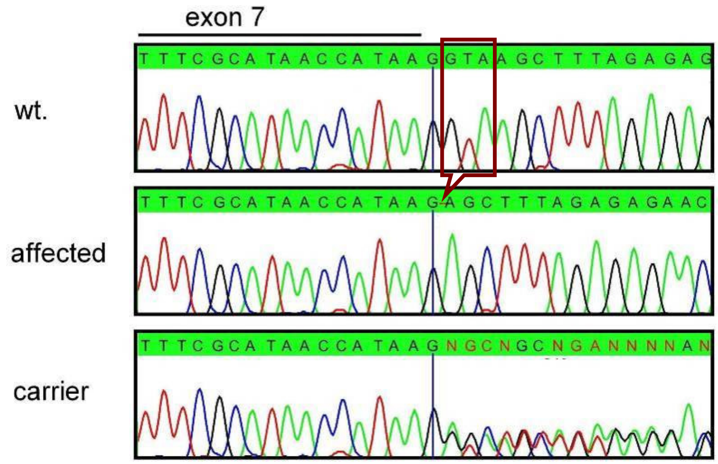


Figure 3-14: Sequence chromatogram of the c.1107 +1delGTA mutation in *ALDH3A2*. Chromatograms for the region containing the mutation in one control (Control), one patient (Affected) and one parent (Carrier) are shown.

This change destroys the donor-splicing site for exon 7 which can lead to the skipping of exon 7 or retention of intron 7.

3.5. Genomic deletion in *TUSC3*

The pedigree and facial aspects of the patients in family 8404553 are shown in Figure 3-15. The degree of MR in the affected family members ranged from moderate to severe (Table 3-4). The patients showed no neurological problems, congenital malformations, or facial dysmorphisms. Head circumference, body height and weight were normal (Table 3-4). MRI scans for two of the patients (IV:5 and IV:7) were performed, which revealed no apparent morphological abnormalities.

Table 3-4: Clinical information of 8404553

Patient	Sex	Age at examination	Mental Retardation / IQ	Height (centile)	OFC (cm)
IV:2	F	21 y	Moderate (35-40)	147 cm	54.5 cm
IV:1	M	21 y	Severe (20-30)	162 cm	54 cm
IV:5	M	8 y	Moderate (40-49)	119 cm	50 cm
IV:7	F	29 y	Moderate (30-40)	151 cm	55 cm
IV:9	F	26 y	Moderate (35-40)	149 cm	51.5 cm
IV:14	F	17 y	Moderate (40-49)	156 cm	54.5 cm

OFC, occipitofrontal circumference;

3.5.1. Genotyping and linkage analysis

Individuals III:1, III:2, III:9, III:10, IV:1-IV:3, IV:5, IV:7, IV:9, IV:13 and IV:14 were genotyped using the Human Mapping 250K (Nsp) Array (Affymetrix).

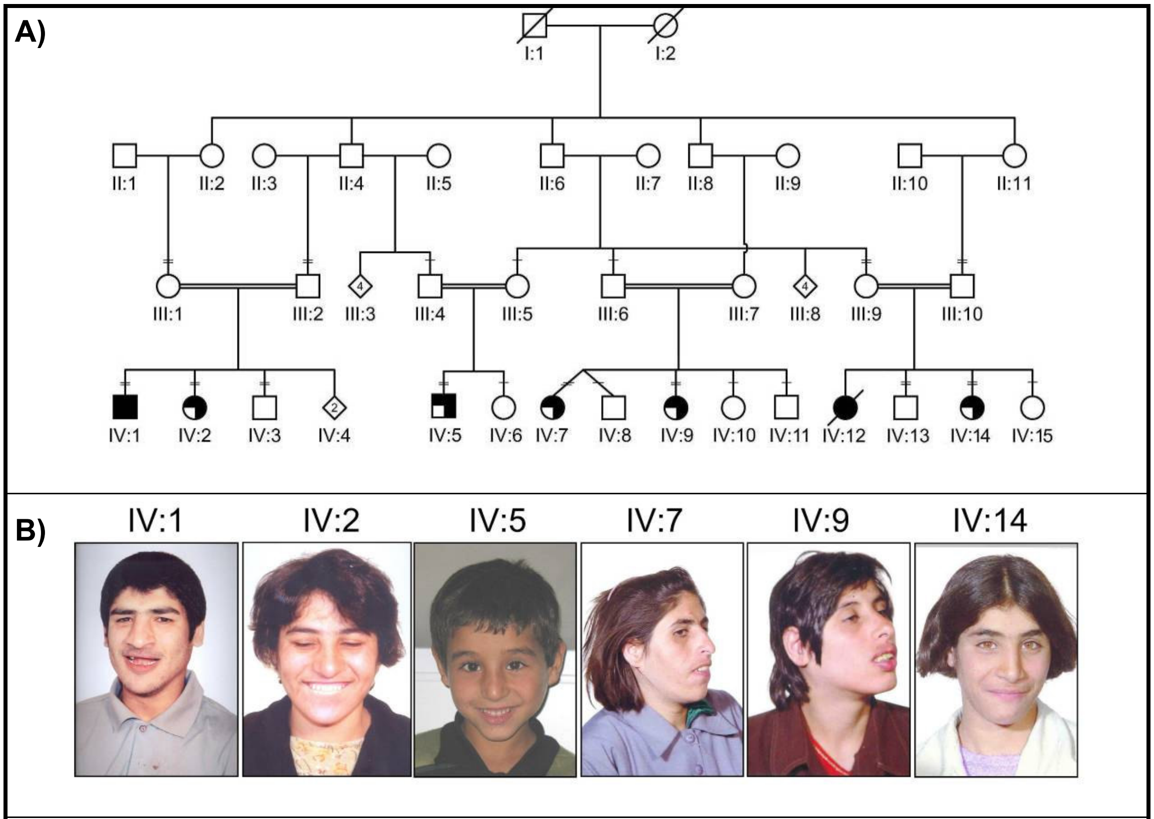


Figure 3-15: Family 8404553. A) Pedigree. Filled symbols indicate severe MR and three quarter filled symbols depict moderate MR. B) Facial aspects of affected family members.

Parametric linkage analysis based on ~50000 markers with high quality scores was carried out which revealed a 4.6 Mb interval containing 14 genes on Chr. 8p22 between rs613566 and rs11203893 with a LOD score of 6.26 in all four branches of the pedigree (Fig S-25 in appendix-B).

3.5.2. Copy number analysis

In parallel, a non-allele specific DNA copy number analysis was performed using the complete set of 250 000 SNPs and the allele specific signal intensity of the markers for each patient in a non-paired mode with two different programs: Copy Number Analyzer for Affymetrix GeneChip (CNAG2.0) (Nannya and others 2005) and the CNAT 4.0 tool (Affymetrix). This led to the identification of 16 markers within the linkage interval that were not called in the patients (Figure 3-16), indicative of a homozygous deletion of approx. 120-150 Kb including the first exon of the *TUSC3* gene.

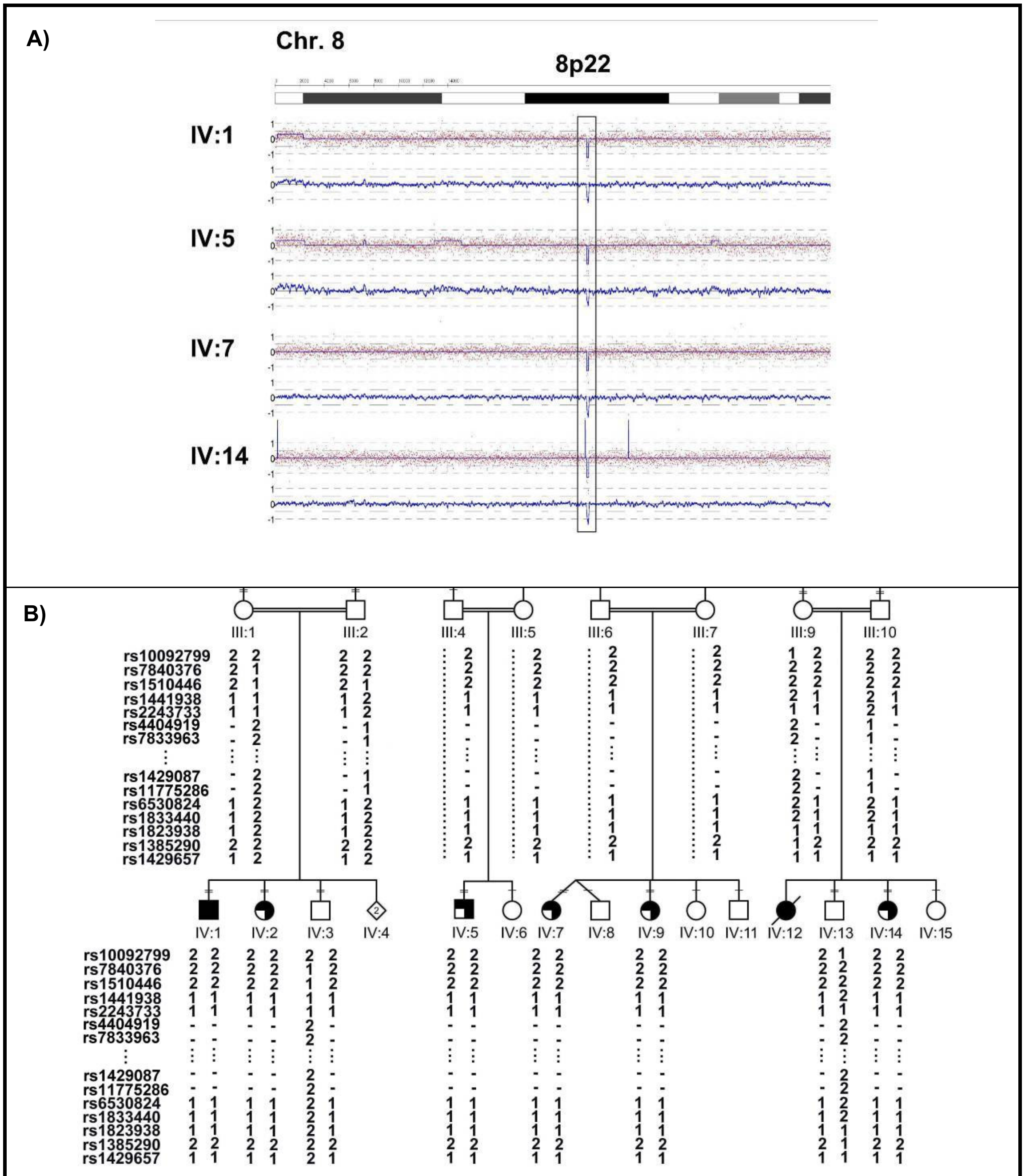


Figure 3-16: Copy number analysis and haplotyping of MR patients from family 8404553. A) Non-paired DNA copy number anaylsis results (CNAG2 tool for copy number variations): Copy number state and log2 ratios for Nsp array SNP markers present inside the first ~30 Mb of chromosome 8 are displayed, showing a 120 Kb homozygous deletion of 8p22 comprising the first exon of *TUSC3* B) Haplotypes of the deletion region: The markers bordering the deletion are shown, revealing that all the affected members are homozygous for the same haplotype while parents and healthy sibs are heterozygous carriers.

By PCR amplification (forward primer: TTGGGTACACCTCCCAGATG; reverse primer: ATCCCAACCCATCATGTCAC) and sequencing of the junction fragment, the exact borders of the deletion were defined (Figure 3-17 A). We found out that 121595 bp (between positions 15347852 and 15469447, NCBI genome build 36.1) were homozygously deleted in all patients. Heterozygous carriers were identified by PCR-amplification of the junction fragment and a PCR product specific for the normal allele (forward primer: TACTTGTGAAAATAACCTGCCATT; reverse primer: TCTCACCAAAATGGTCCACA). We could show that all parents of patients were heterozygous for the deletion. In contrast we did not find homozygous nor heterozygous deletion carriers among 192 unrelated healthy Iranian individuals that were screened as controls (Figure 3-17 B).

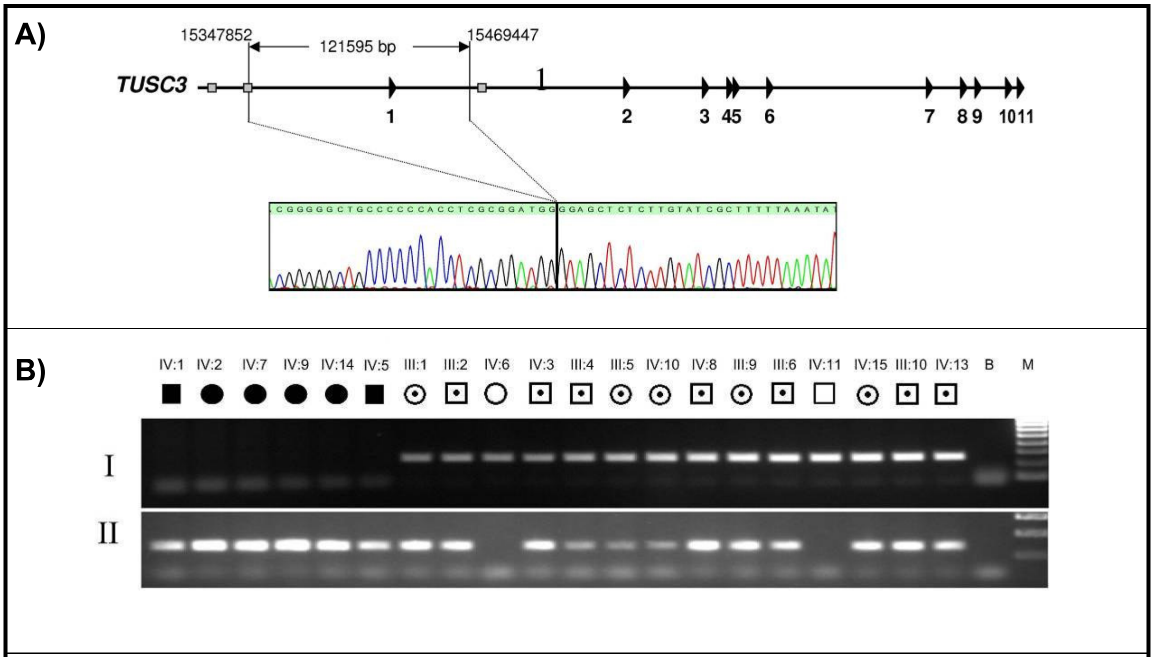


Figure 3-17: Deletion encompassing the first exon of *TUSC3* in MR patients from family 8404553. A) Schematic representation of *TUSC3*: Arrowheads represent exons and grey boxes mark the positions of simple tandem repeats that could be causatively involved in the genesis of the deletion. The positions of the borders of the 121595 bp encompassing deletion (based on NCBI Build 36 .1) are indicated. The sequence chromatogram shows part of the PCR amplicon covering the junction of the deletion borders. B) Co-segregation analysis by PCR: Results of a deletion specific PCR (I) and a deletion spanning PCR (II) are shown for all the available family members. All homozygous carriers show only amplification of the junction fragment (II). Heterozygous carriers show both amplicons and non-carriers show only amplicon I.

3.5.3. *TUSC3* expression study

To check for *TUSC3* expression, total RNA from EBV-transformed lymphoblastoid cell lines (LCLs) of a patient and two controls was extracted and cDNA was generated. This cDNA was used to perform PCRs with a series of primer combinations for one or several adjacent exon sequences together covering the entire gene (Appendix-C).

All PCR products were found to be present in the controls but not in the patients, proving the complete absence of a *TUSC3* transcript in homozygous deletion carriers (Figure 3-18).

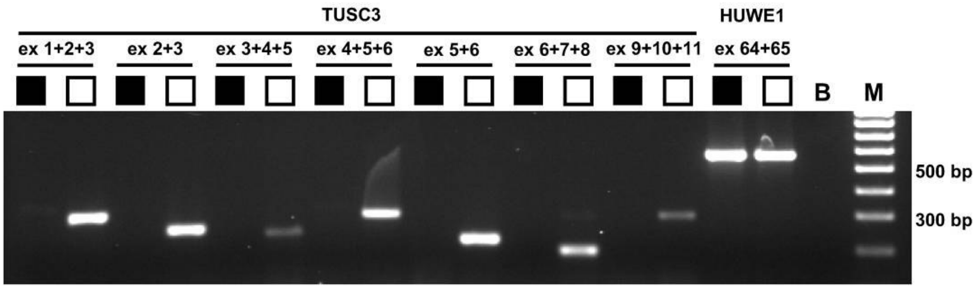


Figure 3-18: RT-PCR results from an experiment with cDNA derived from RNA preparations of 1 patient (IV:5) and 1 control lymphoblastoid cell line sample. Using a sequence of primer pairs specific for amplicons covering 2 to 3 consecutive exons each, the complete *TUSC3* transcript was detected in the control but could not be amplified from patient cDNA. The results of an agarose gel electrophoresis of 5 µl from a 25µl RT-PCR reaction are shown. Patient and control products for a specific amplicon (the exons covered by each amplicon are indicated) were loaded on neighbouring lanes in ascending order of the amplified exons. As positive control a PCR specific for the X-chromosomal *HUWE1* gene was performed (lanes 15 and 16: ex64+65). Filled squares represent the patient and open squares the control, "B" marks the lane loaded with the negative control and "M" indicates the marker lane (HyperLadder IV, Bioline).

This result was substantiated by quantitative PCR, using blood-derived cDNA from four healthy individuals as well as four patients (Figure 3-19).

Experiments were performed using primers covering exon 2 and 3 (forward primer: TAAAGGCACCACCTCGAAAC and reverse primer: TCATTAGCTTGCCTGCACAC).

For normalization, exon 4 to 5 of the Beta-actin gene (forward primer: AAGTGTGACGTGGACATCCG and reverse primer: GATCCACATCTGCTGGAAGG) were amplified in the same experiment.

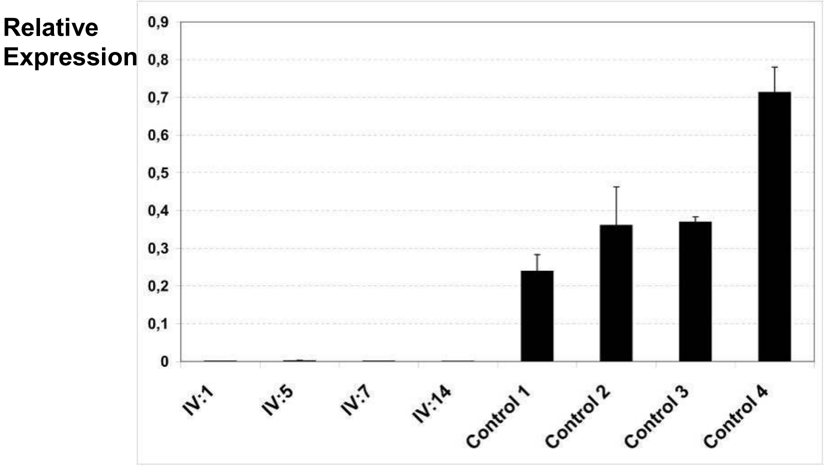


Figure 3-19: Quantitative PCR for *TUSC3* using blood-derived cDNA from four patients as well as from four healthy individuals compaired to Beta-actin expression levels.

3.6. Genomic deletion encompassing exon 4 of *MCPH1*

By Microsatellite analysis our colleagues at the Genetics Research Center (GRC) of the Welfare and Rehabilitation Sciences University (USWR) in Iran showed linkage between microcephaly and the *MCPH1* locus in the affected individuals of family M282 (Figure 3-20) by genotyping the five following microsatellite markers in all family members: D8S1798, D8S1099, D8S1742, D8S277, D8S561 and D8S1819. Three of these markers (D8S1099, D8S277 and D8S1819) were informative and co-segregated with the disease in the pedigree.

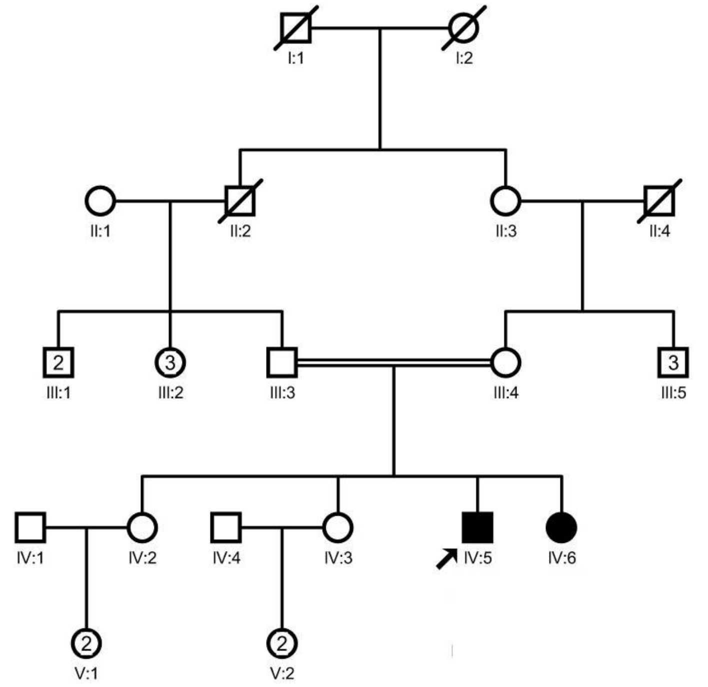


Figure 3-20: Pedigree of family M282

For mutation screening we tried to amplify all the coding exons and exon-intron boundaries of *MCPH1* by specific PCRs. Interestingly, in the patients we could amplify all the exons except for exon 4. This was found in parents, all healthy sibs and controls but not in the affected members, which indicated a homozygous deletion of exon 4 in this family (Figure 3-21).

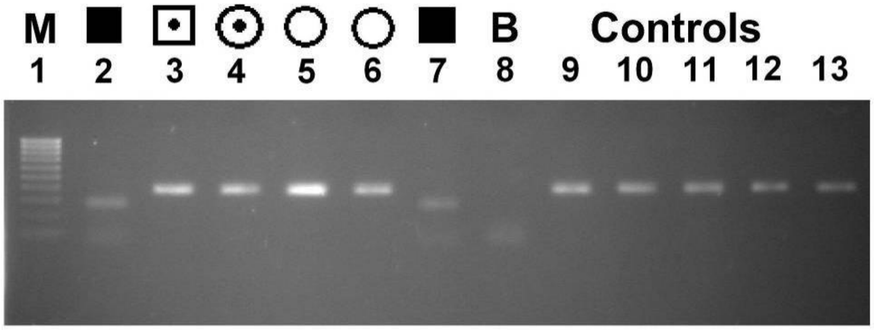


Figure 3-21: **PCR amplification for exon 4 of *MCPH*1 in family M282.** In contrast to the parents (lanes 3 and 4), all healthy sibs (lanes 5 and 6) and controls (lanes 9-13) amplification of exon 4 was not possible in affected individuals (lanes 2 and 7) of family M282.

3.7. Genomic deletion encompassing exon 1-9 of *MCPH1* gene

Further Investigations involved four males and two females between 18 and 32 years of age with moderate MR (IQ between 35 and 70), as well as their unaffected parents and siblings from another large consanguineous Iranian family M019 (see Figure 3-22). Prior to sampling, the individuals were clinically examined (for data see Table 3-5).

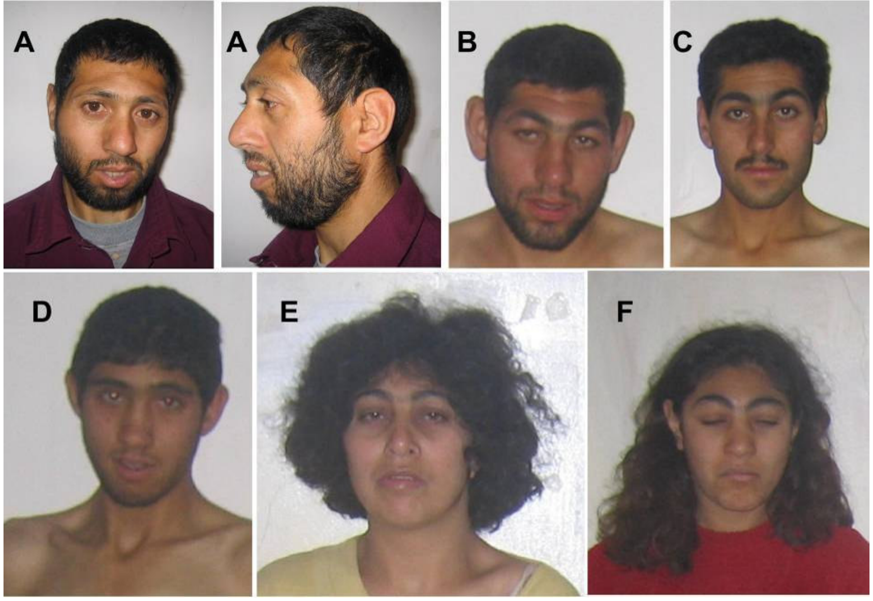


Figure 3-22: Affected members of family M019. A) IV:2, B) IV:3, C) IV:4, D) IV:5, E) IV:7, F) IV:8

Table 3-5: M019 clinical data

Pedigree ID	Age (year)	Head circumference (cm)	Body height (cm)	IQ
IV:2	32	50 (−3 SD)	166 (−2.5 SD)	50
IV:3	26	50 (−3 SD)	167 (−2.5 SD)	51
IV:4	25	50 (−3 SD)	167 (−2.5 SD)	51
IV:5	18	50 (−3 SD)	170 (−2 SD)	70
IV:7	32	49 (−3 SD)	151 (−3 SD)	50
IV:8	22	49 (−3 SD)	150 (−3 SD)	52

For each individual, genotyping was performed with the Affymetrix Human Mapping 10 K Array Version 2 (Kennedy and others 2003). Non-parametric and parametric multipoint linkage analysis yielded a single significant peak between the markers rs718122 and rs725944 (8p22–8p23.2) on the short arm of chromosome 8 with LOD Scores of 4.2 and 22 respectively (for linkage andhaplotype results see the Fig S-2 in appendix B).

In both analyses a discontinuity involving rs1057187 was observed at the centre of this linkage interval (Figure 3-19), pointing to heterozygosity for this marker, while at the same time the affected individuals were homozygous for 40 other tested SNPs from this region. Examination of the raw genotyping data revealed that in the mentally retarded individuals, rs1057187 along with the two flanking markers, rs725438 and rs1868551, failed to yield any hybridisation signal.

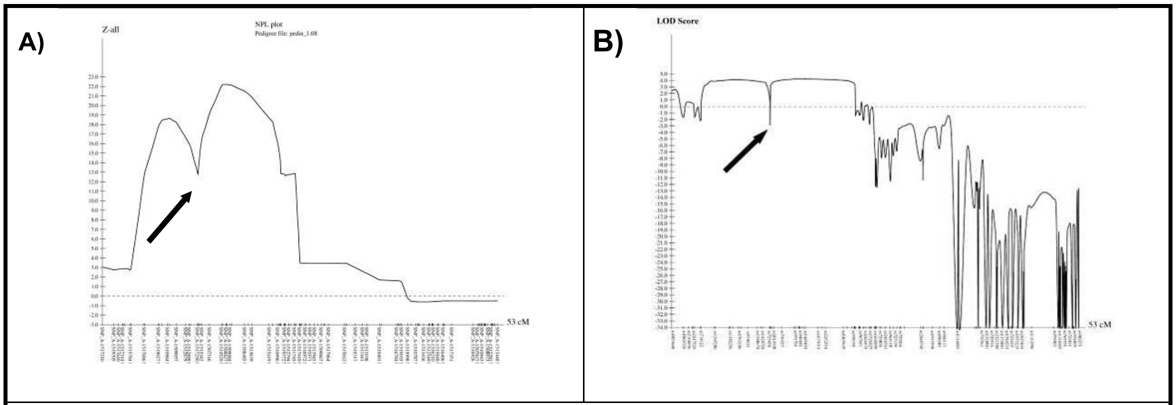


Figure 3-23: Results of linkage analysis for family M019. Non-parametric (A) and parametric (B) LOD scores for the markers in the region of the linkage interval on chromosome 8 showing a discontinuity in the middle of the flat ‘peak region’ around marker SNP_A-1517719.

The latter were not informative in this family, but in the case of rs1057187, the parents of five affected children appeared to be homozygous carriers of different alleles. Upon closer inspection, however, it turned out that these parents are in fact heterozygous for the deletion and different rs1057187 alleles (Figure 3-24).

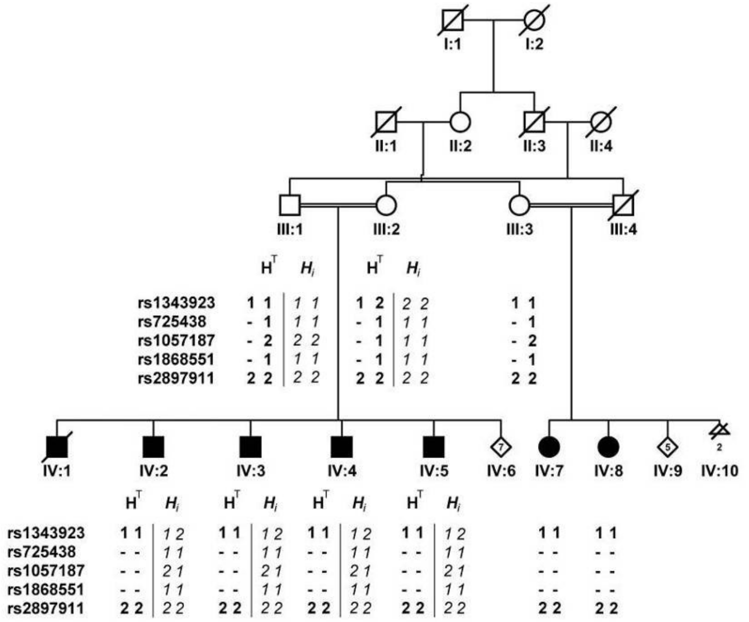

I:1
I:2
II:1
II:2
II:3
II:4
III:1
III:2
III:3
III:4
H^T H_i H^T H_i
rs1343923 1 1 | 1 1 1 2 | 2 2 1 1
rs725438 - 1 | 1 1 - 1 | 1 1 - 1
rs1057187 - 2 | 2 2 - 1 | 1 1 - 2
rs1868551 - 1 | 1 1 - 1 | 1 1 - 1
rs2897911 2 2 | 2 2 2 2 | 2 2 2 2
7
5
2
IV:1
IV:2
IV:3
IV:4
IV:5
IV:6
IV:7
IV:8
IV:9
IV:10
H^T H_i H^T H_i H^T H_i H^T H_i
rs1343923 1 1 | 1 2 1 1 | 1 2 1 1 | 1 2 1 1 | 1 2 1 1 1 1
rs725438 - - | 1 1 - - | 1 1 - - | 1 1 - - | 1 1 - - - -
rs1057187 - - | 2 1 - - | 2 1 - - | 2 1 - - | 2 1 - - - -
rs1868551 - - | 1 1 - - | 1 1 - - | 1 1 - - | 1 1 - - - -
rs2897911 2 2 | 2 2 2 2 | 2 2 2 2 | 2 2 2 2 | 2 2 2 2 2 2

Figure 3-24: M019 Family pedigree. For each individual, based on rs1057187 and its neighboring markers, the true haplotypes (HT) and the haplotypes inferred by the analytical software (*H i*) are shown.

Several primer pairs from the relevant region failed to yield PCR products in four patients indicating that there was a deletion segregating in this family. The size (50–480 kb) of the deleted DNA segment was inferred from the distance between the three deleted markers (50 kb) and the distance between the closest flanking SNPs that showed hybridization signals on the chip (480 kb) (Figure 3-29).

To confirm these findings, PCR experiments were performed using specific primers for a 170 bp DNA stretch between the informative marker (rs1057187) and its 3’ neighbour (rs1868551). These experiments yielded amplicons of the expected size for all the parents and healthy siblings examined, whereas no PCR product could be obtained from DNA of the affected individuals (Figure 3-25).

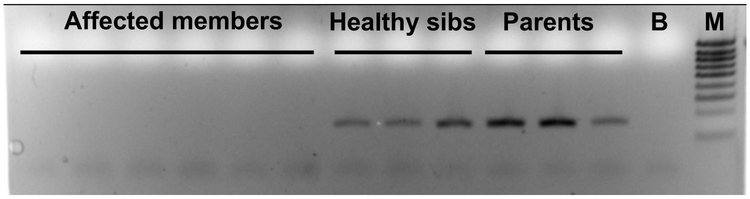


Figure 3-25: PCR amplification of genomic DNA in affected and unaffected individuals from family M019. The first 6 lanes relate to 6 affected individuals and show no amplification signal. The next 3 lanes relate to the DNA of healthy siblings where PCR is giving rise to the 170bp product. The next three lanes belong to healthy parents and the final lane before the marker is negative blank control.

The region of interest encompassed *MCPH1* and Angiopoietin 2 encoding gene (*ANGPT2*) which is located in the opposite direction and inside the intron 12 of *MCPH1*. In order to check the extent to which *MCPH1* and *ANGPT2* were affected by the deletion, we performed a series of PCR experiments and were able to obtain specific amplicons for exons 7–9 of *MCPH1* as well as for the terminal exon of *ANGPT2* but not for *MCPH1* exon 6 (Primer sequences are listed in Appendix-D). Thus the observed deletion in the mentally retarded individuals did not affect *ANGPT2* but comprised the first six exons of *MCPH1* and at least 26 kbp of the upstream region, as determined by the missing SNP markers (Figure 3-29).

In collaboration with the group of Dr. Ullmann at the Max Planck Institute for Molecular Genetics in Berlin, a sub-megabase resolution array CGH (Erdogan and others 2006) was performed for one of the patients of family M019. This revealed that the deletion was confined to two BAC clones mapping to 8p23.1 (April 2003 Genome Browser version: 6232904–6582980 bp), indicating that the deletion spanned 150–200 kbp. The genomic segment covered by these two BACs contains all 11 exons of *MCPH1* as well as the inversely arranged *ANGPT2*, which is situated within intron 9 of *MCPH1* (Figure 3-26).

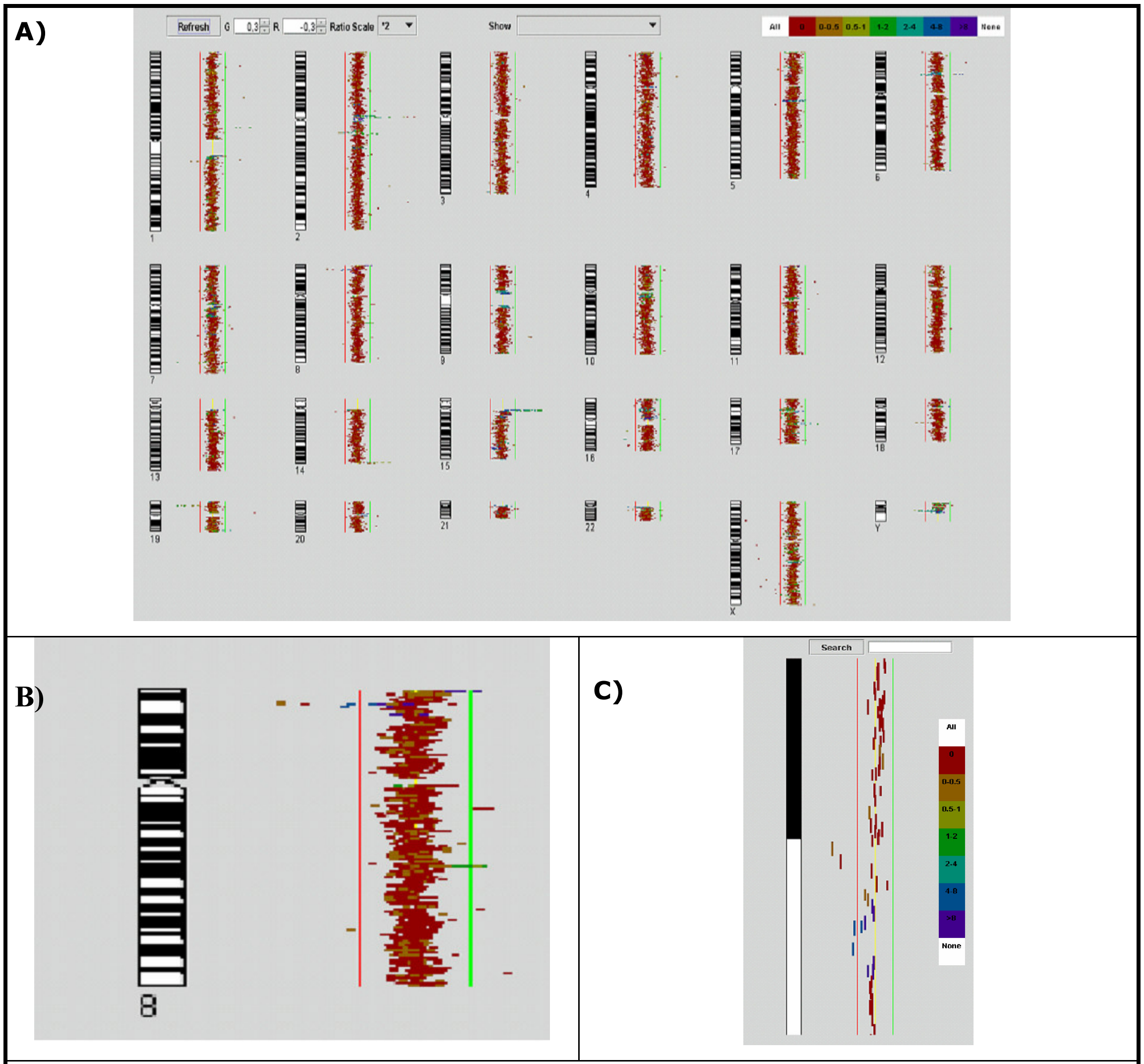


Figure 3-26: Array CGH results for one affected member of family M019. (A) Genomic view of high-resolution array CGH. (B and C) Two deleted BACs in position 8p23.1.

It has been shown before that in addition to microcephaly another hallmark of the phenotype of patients with *MCPH1* mutations is the occurrence of a high number of cells (10–15%) with prophase-like chromosomes and low-quality metaphase G-banding in routine karyotype analyses (Trimborn and others 2004). Both features were found in all mentally retarded members of family M019 (Figure 3-27).

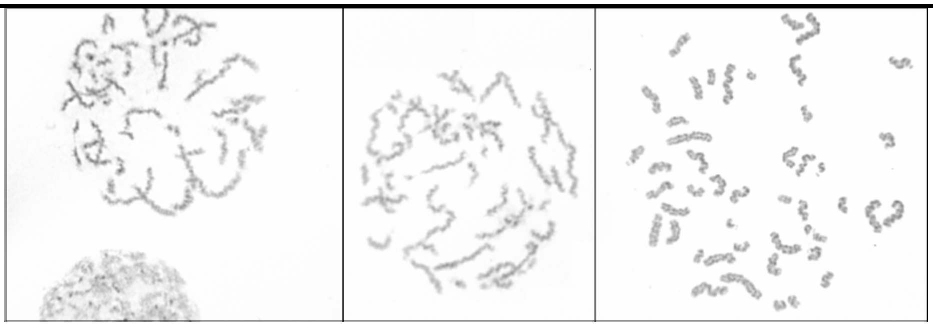


Figure 3-27: Examples of metaphase chromosome preparations from affected members of family M019. They show defective chromosome condensation and poor Giemsa banding.

3.7.1. Expression analysis of *MCPH1* in LCLs from microcephalic patients with *MCPH1* mutations

In order to investigate the impact of the deletion on the transcriptional level, RNA was extracted form patient LCLs. In addition, RNA was isolated from 3 other patients with S25X, T143NfsX5 and W75R mutations in *MCPH1* respectively (provided kindly by Dr. Trimborn and Prof. Dr. Neitzel from the Charité Universitätsmedizin Berlin) as well as 5 healthy controls. After isolating poly A RNA (Dynabeads® Oligo (dT)25; DYNAL BIOTECH) Northern blotting was performed using a probe covering exon 11 to 14 of *MCPH1* (forward primer: ATGTCGTCATCCAGGTTGTG and reverse primer: CGCCAGTTCCTTCTCTTCAC). Three different transcripts with approximate sizes of 2.9, 3.4 and 4kb were detected in the controls. Except in case of the patient with the S25X mutation the results indicated nonsense-mediated decay (NMD) for all the investigated patients with *MCPH1* mutations (Ex1_9del, T143NfsX5 and W75R (Figure 3-28).

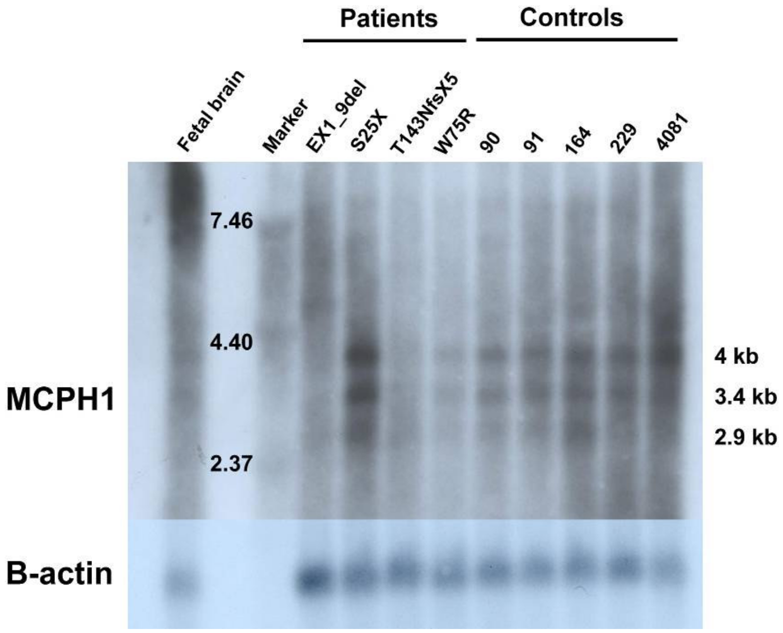


Figure 3-28: Northern blotting using a probe from exon 11-14 of MCPH1. The First lane is RNA from fetal brain. The first four lanes after the marker belong to four patients with four different mutations of *MCPH1* followed by 5 healthy controls. At the bottom the same blot is shown after stripping and re-probing with beta-actin as a control for sample loading. Patients with Ex1_9del, T143NfsX5 and W75R mutations showed NMD for all the three different detected transcripts.

The 835 amino acid MCPH1/microcephalin protein is predicted to contain three breast cancer 1 C-terminal (BRCT) domains. One BRCT domain is present at the N-terminus and two at the C-terminus of the protein (Figure 3-29).

BRCT motifs are commonly found in DNA damage response proteins, particularly those functioning as mediators in the signalling response providing provocative although circumstantial evidence that MCPH1 might function in a DNA damage response pathway (Jackson and others 2002).

All of the known *MCPH1* mutations occur in the first part of the gene, and in all the cases the last two BRCT domains remain intact.

Like the published mutations for *MCPH1*, the two deletion mutations that we found (deletion of exon 1 to 9 and deletion of exon 4) also affect only the first BRCT domain (Figure 3-29).

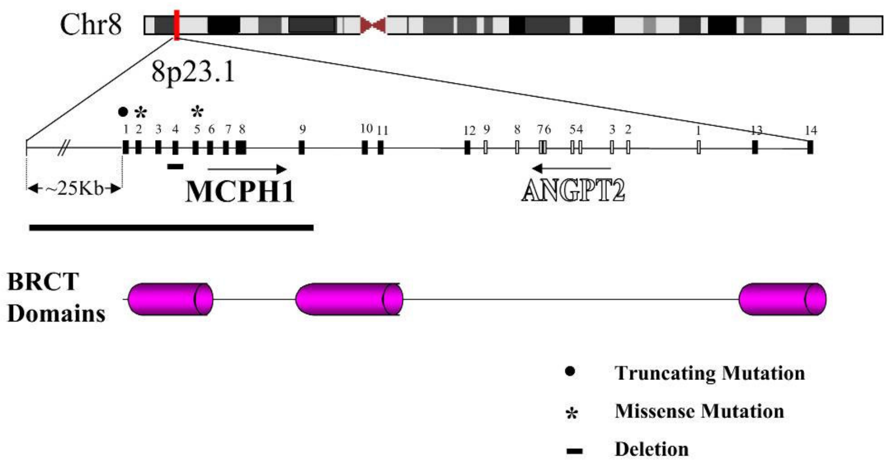


Figure 3-29: Schematic representation of the location of *MCPH1* (exons represented by black boxes) and *ANGPT2* (exons represented by empty boxes). Arrows indicate the orientation of the genes. Both of the 2 deletion mutations found by us and all of the other three published mutations found so far in *MCPH1* (S25X, 427insA and c.80C>G) are indicated at their relevant positions. Approximate positions of the three BRCT domains are depicted by pink cylinders at the bottom.

These observations may argue that functions of the first BRCT domain differs from that of the other ones.

To study whether there is any transcription for the distal part of *MCPH1* in patients (especially in the patient with the EX1_9del mutation), RNA was extracted from the lymphoblastoid cell lines (LCLs) of the patient with the EX1_9del mutation and RT-PCR was performed for one of the patients and 6 controls. The primers we used were suitable to amplify a fragment containing *MCPH1* exons 4 to 8 (forward primer: GAATCATTGTTCCCTGCAGC and reverse primer: TTACTGAGGAACTCCTGGGTC) or the 3'UTR region of *MCPH1* (forward primer: GAGTGCAATGGCACAATCTC and Reverse primer: GATCGAGTCTAAGCCAAGAA).

In this way, the 3'UTR of *MCPH1* could be amplified both in the patient and the controls. In contrast, exons 4 to 8 could only be amplified in the controls, indicating that the distal part of the *MCPH1* in patients with the EX1_9 del mutation is still transcriptionally active (Figure 3-30).

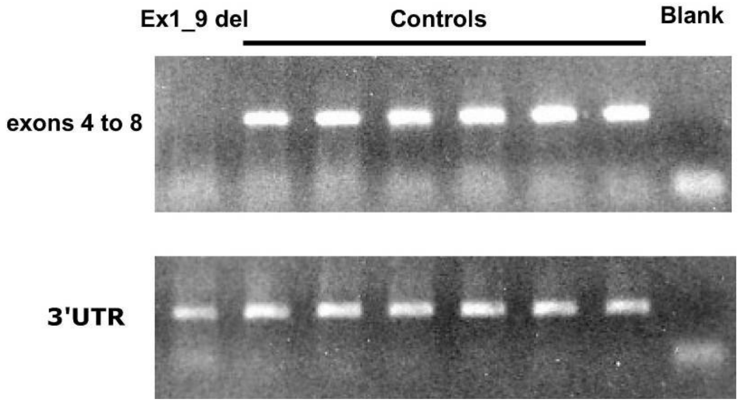


Figure 3-30: RT-PCR using primers belonging to exon 4 to 8 and 3'UTR of *MCPH1* in the patient with the Ex1_9 del mutation and in 6 controls.

Interestingly the RT-PCRs perfomed on control cDNA revealed a new *MCPH1* isoform without exon 13, which was confirmed by sequencing.

The results of RT-PCR for 3UTR in patients with exon 1-9 deletion strengthened the hypothesis that the patients retain a small *MCPH1* transcript which contains the last two BRCT domains. Such a transcript might even be present in healthy individuals.

In order to investigate this possibility, we performed Rapid Amplification of cDNA Ends (RACE) on cDNA from a patient with the Ex1_9del mutation, using four gene specific primers (GSPs) located in the 3'UTR of MCPH1 (Figure 3-31).

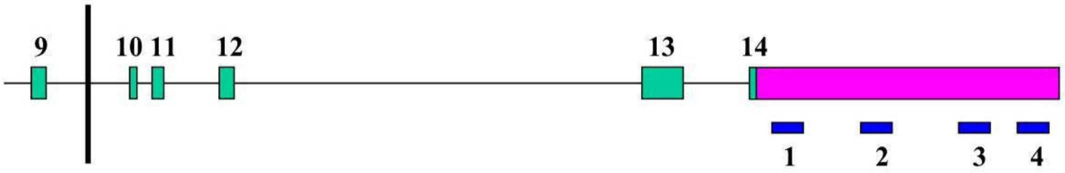


Figure 3-31: Approximate positions of the four gene specific primers (GSPs) in MCPH1's 3'UTR that were used for 5' RACE.

However, a complete and reliable characterization of the putative small transcript was not possible, which is most probably due to extremely low expression levels in lymphoblastoeid cell lines (LCLs).

3.7.2. Radiation assay

Published evidence suggests that MCPH1 plays a role in DNA repair (for review see the O'Driscoll and others 2006).

To check DNA repair in microcephalic patients with *MCPH1* mutations, a radiation assay was performed in collaboration with Dr. Trimborn and Prof. Dr. Neitzel (Charité Universitätsmedizin Berlin) using LCLs of the patient with the Ex1_9del from this study along with LCLs from patients with different *MCPH1* mutations (S25X and T27R). As positive control cells from patients with ATM (Ataxia telangiectasia mutated) mutations were used, since it is known that patients with ATM mutations have defects in checkpoint arrest. Additionally, LCLs from two healthy individuals served as negative controls.

In contrast to the patient cells with *ATM* mutaion, all the patient cells with *MCPH1* mutations as well as the control cells behaved normally in response to 1 and 4 Gy of radiation. Evaluation of mitotic index one and two hours after exposure showed that the cell division rate in the patient cells decreased dramatically (Figure 3-32).

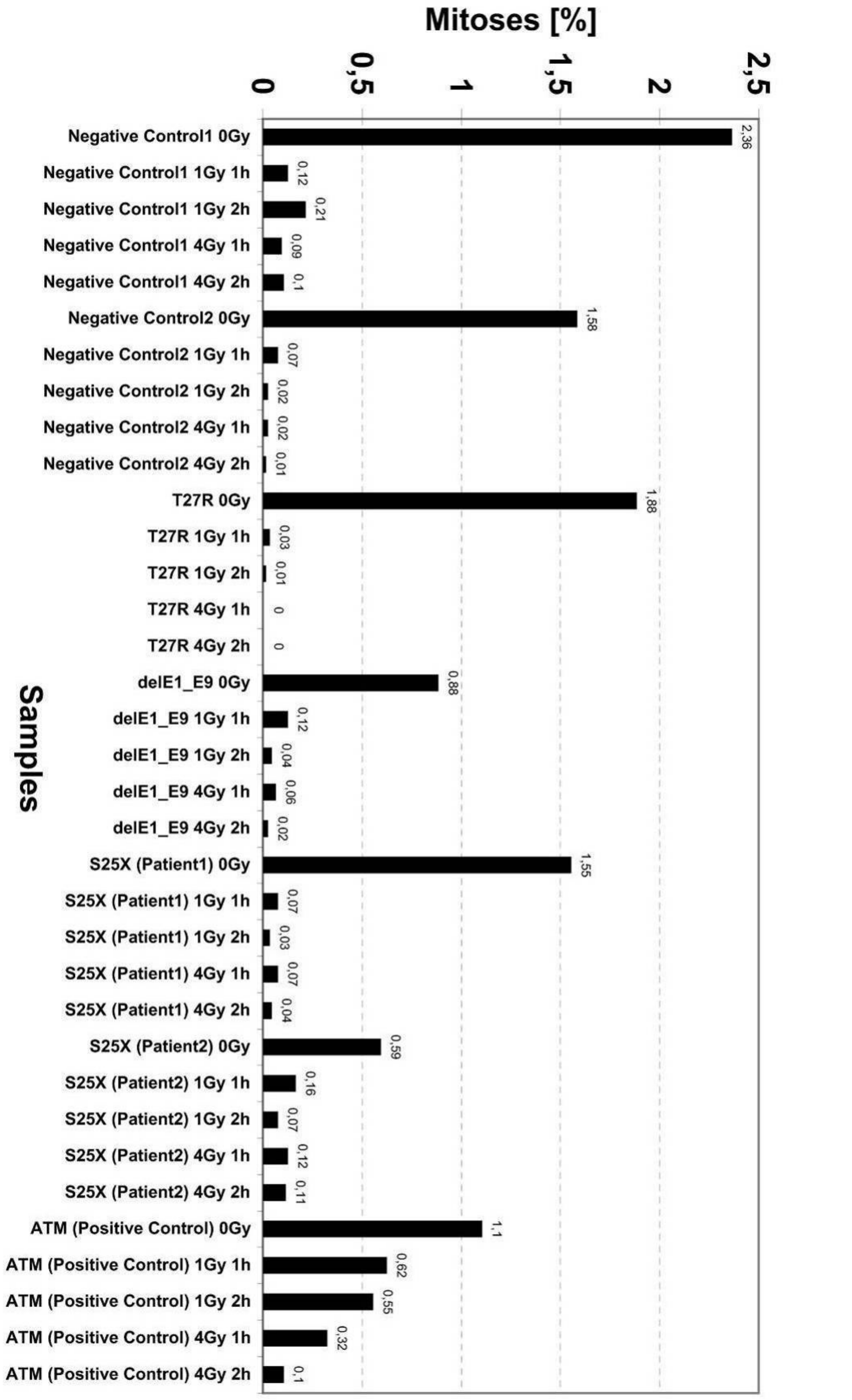


Figure 3-32: Radiation assay for patients with *MCPH1* and *ATM* mutations. Mitotic index were measured for two healthy samples (negative controls), patients with T27R, Ex1_9del, S25X mutations of *MCPH1* and one patient with mutation of *ATM* (positive control) at one and two hours after 1 and 4Gy of exposures.

3.7.3. Whole genome expression profiling on LCLs from microcephalic patients with *MCPH1* mutations

It has been repeatedly reported that *MCPH1* would regulate protein and transcript levels of other genes such as *hTERT*, *BRCA1* and *CHK1* (Lin and Elledge 2003; Lin and others 2005; Xu and others 2004), so that, make this believe that microcephalin/BRIT1 may function in transcriptional regulation. Therefore, we performed whole genome expression profiling to investigate the effects of the *MCPH1* mutations on the expression of other gene. Therefore Illumina Sentrix® Human-6 Expression BeadChips were employed to compare the expression levels of ~48000 transcripts from known and predicted human genes in patient cells with *MCPH1* mutations and controls.

In a first set of expriments, four patients with different mutations of *MCPH1* (EX1_9 del, S25X, T143NfsX5 and W75R) and 8 controls were studied.

This experiment led to the identification of several promising potential targets of gene regulatory processes that involve MCPH1.

At the beginning we compared *MCPH1* expression levels obtained by Northern blotting (Figure 3-28) with the array results (Table 3-6) and found that the array-results confirmed the observations made in the Northern blot experiments.

Table 3-6: Illumina *MCPH1* expression data for 4 patients with MCPH1 mutations

Patients	**Ex1_9del**	**S25X**	**T143NfsX5**	**W75R**
Differentiation score*	-14.84	1.20	-7.29	-1.04

* The 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 upregulation, while negative represents downregulation.

To see if *MCPH1* deficiency has an impact on the expression levels of other genes known to be responsible for microcephaly, Northern blotting with *ASPM* and *CDK5RAP2* specific probes was performed in addition to the array analysis.

Based on the Illumina gene expression array data only the patient with EX1_9del showed downregulation for *ASPM*, which was substantiated by Northern blotting as well (Figure 3-33).

Abnormal spindle-like, microcephaly-associated (ASPM)

Patients	Ex1_9del	S25X	T143NfsX5	W75R
Differentiation score	-22.48	3.57	-3.23	-1.01

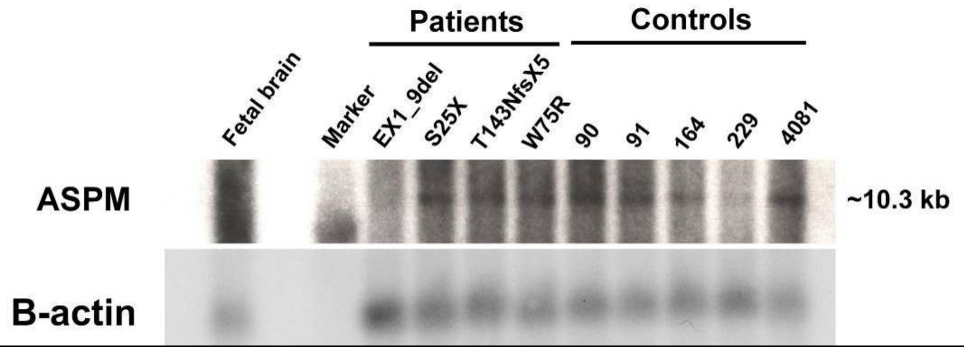


Figure 3-33: Northern blot with *ASPM* specific probe. The first lane contains RNA from fetal brain. The first four lanes after the marker lane belong to the four patients with different mutations in *MCPH1* (Ex1_9del, S25X, T143NfsX5 and W75R). These are followed by the 5 controls. After stripping, the blot was re-probed with a beta-actin specific probe in order to control for sample loading. The diff. scores for the same patients in comparison to control on an Illumina array are presented in the table above. Compatible with the array data only patients with the Ex1_9del mutation show downregulation for the only detected transcript with the approximate size of 10.3kbp.

In case of *CDKA5RAP2*, array data showed downregulation in patients with the EX1_9del and W75R mutations but also in one of the controls (229), which was confirmed by Northern blotting (Figure 3-34).

CDK5 regulatory subunit associated protein 2 (CDK5RAP2)

Patients	Ex1_9del	S25X	T143NfsX5	W75R	229
Differentiation score	-15.69	17.60	7.16	-26.35	-42.2

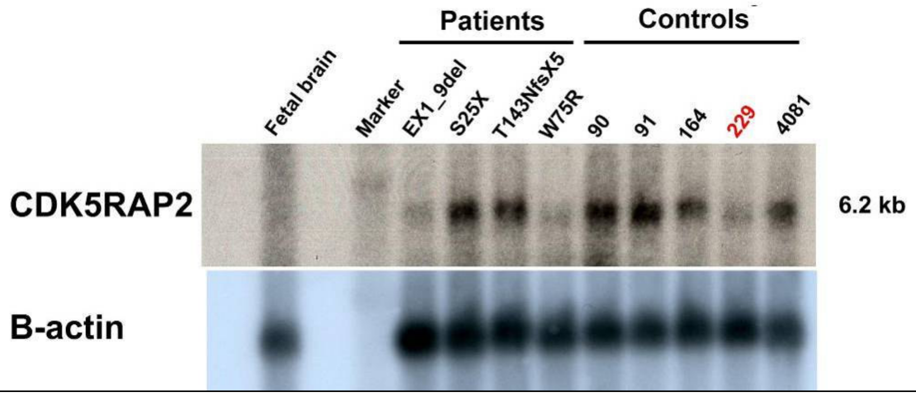


Figure 3-34: Northern blot with *CDK5RAP2* specific probe. The first lane contains RNA from fetal brain. The first four lanes after the marker lane belong to the four patients with different mutations of *MCPH1* (Ex1_9del, S25X, T143NfsX5 and W75R). These are followed by the 5 controls. After stripping, the blot was re-probed with a beta-actin specific probe in order to control for sample loading. The diff. scores for the same patients in comparison to the control on an Illumina array are presented in the table above. Compatible with the array data patients with Ex1_9del and W75R mutations in addition to one of the controls (229) show downregulation for the only detected transcript with the approximate size of 6.2kbp.

To see if the observed decrease in *CDKA5RAP2* transcription has any effect on its protein levels, Western blotting was performed with total cell protein lysate from lymphoblastoid cell lines (LCLs) of patients with *MCPH1* mutations (one sample with the Ex1_9del, two samples form each of S25X, T143NfsX5 and one sample from each W75R and T27R mutations) and controls. However, no differences between patients and controls were observed (Figure 3-35).

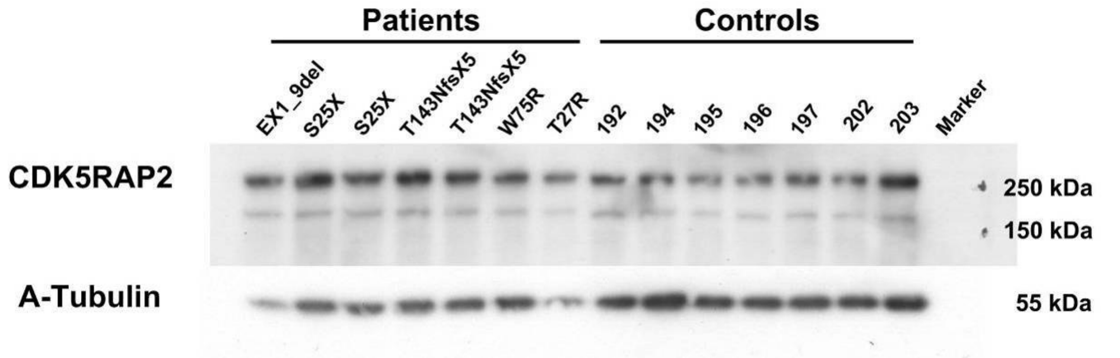


Figure 3-35: Western blotting result for CDK5RAP2. The first 7 lanes were loaded with total cell lysate from patients with five different mutations in *MCPH1* (Ex1_9del, S25X, T143NfsX5, W75R and T27R) followed by 7 healthy controls. The same blot, after stripping was re-probed with an α-Tubulin specific antibody to control for sample loading. In contrast to the RNA levels, no differences for both variants of CDK5RAP2 in protein level were observed between patients and controls.

To check the validity of the micro-array data, several additional genes with deregulation patterns were checked by Northern blotting. These genes were selected either manually according to their function based on available literature or using bioinformatic tools like ENDEAVOUR software (Aerts and others 2006).

All the genes known to be involved in cell cycle, DNA repair or chromosome segregation were considered as training sets and fed to ENDEAVOUR software separately. Thereafter those genes among our list with downregulation in patient cells with *MCPH1* mutations that could be assigned to each of these categories by ENDEAVOUR were extracted. The common genes that appeared in several categories were selected for Northern blotting. Primers used for amplification of proper probes are shown in table 2-17.

Comparison of expression profiling array and Northern blotting results for *EGR2, DUSP4, LCK, PHGDH, HK1, PSAT1, STAT1, PLCG2, PTEN, NK4, ANXA11* and *FLJ31978* are presented in appendix-E.

Furthermore, to check the reliability of the array data in case of up-regulated genes, Northern blotting was performed for *FLJ31978*. Both micro array data and Northern blotting showed up-regulation for the patients with the Ex1_9del, S25X and W75R mutations but not the patients with the T143NfsX5 mutation in comparison to controls (Appendix-F).

Thus, the Northern blotting results for all the selected genes were compatible with the array results. Therefore it was decided to expand the number of patient samples in order to reduce the chance of cell line specific changes due to the EBV-transformation and to decrease alterations, which probably are not biological consequences of MCPH1 mutations.

For this purpose additional patients with the same mutations were included in the analysis. In total 8 patients with 5 different mutations of *MCPH1* (two patients with EX1_9 del, two with S25X, two with T143NfsX5, one patient with W75R and one with T27R) were examined. Furthermore one parent (heterozygous for T143NfsX5) and 9 additional controls were investigated.

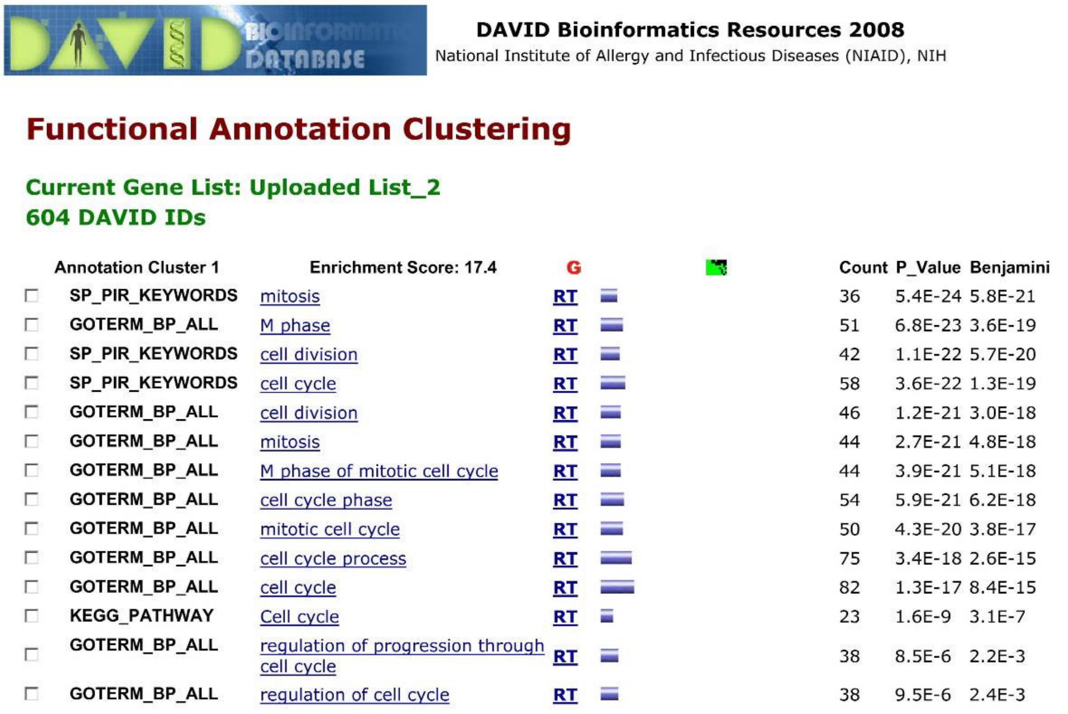
The Diff. Score parameters were used to determine gene expression in the Illumina BeadStudio software program. The Diff. Score is a transformation of the p-value that provides directionality to the p value based on the difference between the average signals in the reference group vs. the comparison group.

Data analyses were performed by grouping all the patients with different mutations and comparing them with the group of 9 controls, using the "Rank-Invariant" method of normalization and the "custom" algorithm of the bead studio software (Illumina). This led to the identification of 675 downregulated genes (diff. scores ≤-13). The 79 downregulated genes with stringent (ie. ≤-30) diff. scores corresponding to P-values ≤0.001 are shown in appendix-G.

We decided to consider only the downregulated genes for further analysis, given the reported role of microcephalin/BRIT1 in transcriptional activation (Yang and others 2008).

Functional annotation clustering of genes downregulated in MR patients with *MCPH1* mutations:

In order to search for functionally related genes among the 675 genes which were downregulated in the patients, the DAVID and Panther 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 17.4 for the 604 interrogated genes out of 675 downregulated genes is shown in Figure 3-36. This analysis showed highly significant (P-values ≤ $\times e^{-24}$) clustering of genes in expected pathways such as cell cycle control and regulation of mitosis.



DAVID Bioinformatics Resources 2008

National Institute of Allergy and Infectious Diseases (NIAID), NIH

Functional Annotation Clustering

Current Gene List: Uploaded List_2

604 DAVID IDs

	Annotation Cluster 1	Enrichment Score: 17.4	G	Count	P_Value	Benjamini
□	SP_PIR_KEYWORDS	mitosis	RT	36	5.4E-24	5.8E-21
□	GOTERM_BP_ALL	M phase	RT	51	6.8E-23	3.6E-19
□	SP_PIR_KEYWORDS	cell division	RT	42	1.1E-22	5.7E-20
□	SP_PIR_KEYWORDS	cell cycle	RT	58	3.6E-22	1.3E-19
□	GOTERM_BP_ALL	cell division	RT	46	1.2E-21	3.0E-18
□	GOTERM_BP_ALL	mitosis	RT	44	2.7E-21	4.8E-18
□	GOTERM_BP_ALL	M phase of mitotic cell cycle	RT	44	3.9E-21	5.1E-18
□	GOTERM_BP_ALL	cell cycle phase	RT	54	5.9E-21	6.2E-18
□	GOTERM_BP_ALL	mitotic cell cycle	RT	50	4.3E-20	3.8E-17
□	GOTERM_BP_ALL	cell cycle process	RT	75	3.4E-18	2.6E-15
□	GOTERM_BP_ALL	cell cycle	RT	82	1.3E-17	8.4E-15
□	KEGG_PATHWAY	Cell cycle	RT	23	1.6E-9	3.1E-7
□	GOTERM_BP_ALL	regulation of progression through cell cycle	RT	38	8.5E-6	2.2E-3
□	GOTERM_BP_ALL	regulation of cell cycle	RT	38	9.5E-6	2.4E-3

Figure 3-36: DAVID functional annotation clustering for 604 downregulated genes with diff. scores ≤–13 (corresponding to P-values ≤0.05) when comparing expression profiles of 8 patients with 5 different mutations in *MCPH1* (EX1_9 del, S25X, T143NfsX5, W75R and T27R mutations) as one group in comparison to a group of 9 controls. The most significant P-values (P_value and Benjamini) yielded by DAVID for downregulated genes classified to different annotation terms based on different databases are shown in the last two columns.

For example, according to the KEGG database 23 out of the 604 genes submitted to DAVID are involved in the cell cycle with a P-value of $1.6 \times e^{-9}$, as shown in Figure 3-37.

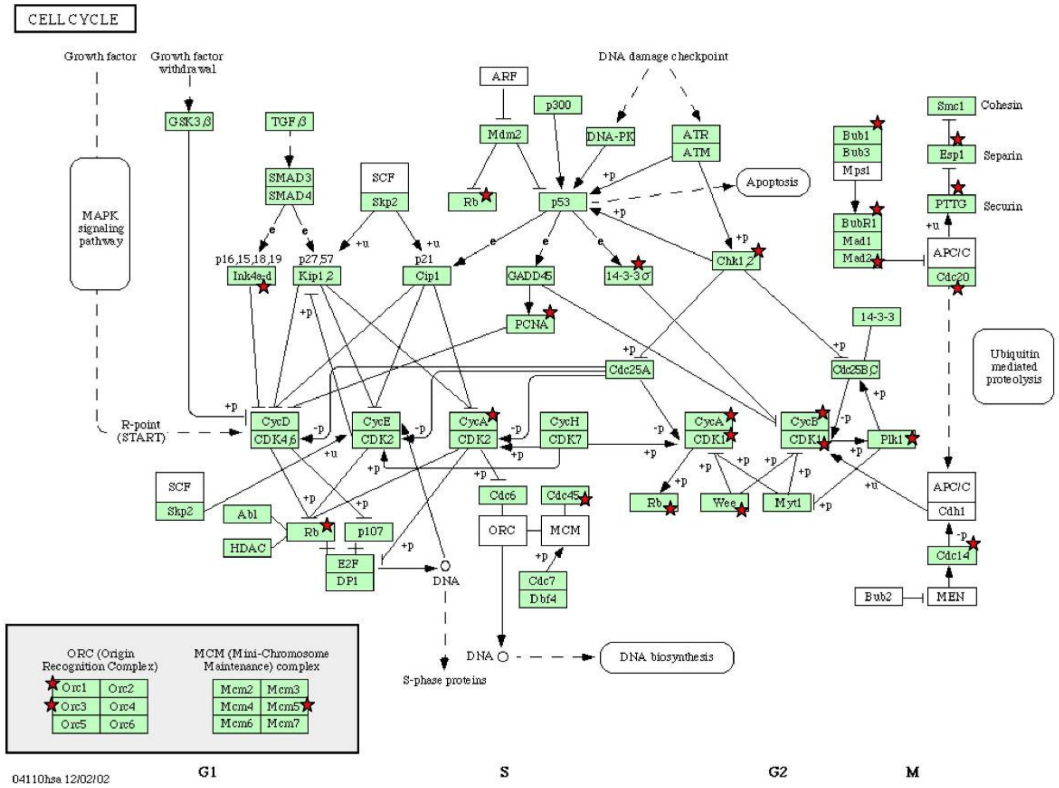
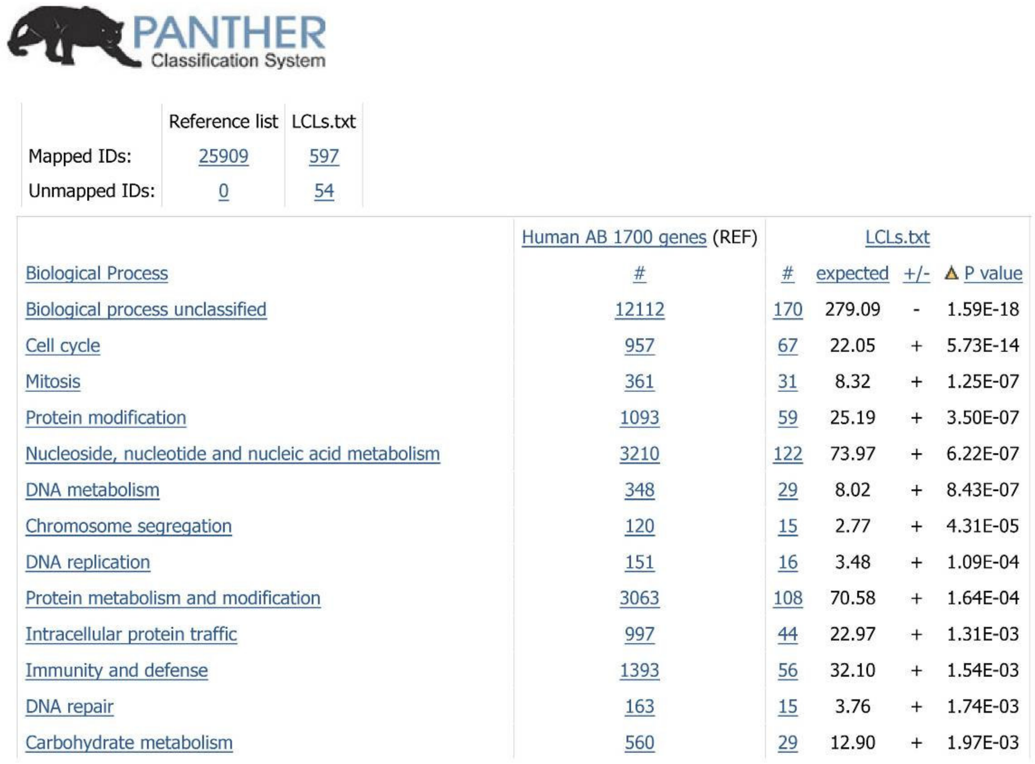


Figure 3-37: KEGG pathway for cell cycle. The 23 common genes (*WEE1 ORC3L, BUB1B, PLK1, MAD2L1, CDC45L, ESPL1, CDC2, PTTG1, PTTG2, WEE1, ORC1L, CCNA2, RB1, MCM5, CHEK1, CHEK2, YWHAG, PCNA, CDKN2C, CDC14B, CDC20, BUB1* and *CCNB1*) between LCLs downregulated genes with diff. scores ≤-13 (corresponding to P-values ≤0.05) and cell cycle genes based on KEGG database are depicted by red asterisks.

A highly significant enrichment of transcripts from the expected pathways such as cell cycle, mitosis, chromosome segregation and DNA repair was obtained by using the Panther database (Figure 3-38).



	Reference list	LCLs.txt
Mapped IDs:	25909	597
Unmapped IDs:	0	54

	Human AB 1700 genes (REF)	LCLs.txt			
Biological Process	#	#	expected	+/-	▲ P value
Biological process unclassified	12112	170	279.09	-	1.59E-18
Cell cycle	957	67	22.05	+	5.73E-14
Mitosis	361	31	8.32	+	1.25E-07
Protein modification	1093	59	25.19	+	3.50E-07
Nucleoside, nucleotide and nucleic acid metabolism	3210	122	73.97	+	6.22E-07
DNA metabolism	348	29	8.02	+	8.43E-07
Chromosome segregation	120	15	2.77	+	4.31E-05
DNA replication	151	16	3.48	+	1.09E-04
Protein metabolism and modification	3063	108	70.58	+	1.64E-04
Intracellular protein traffic	997	44	22.97	+	1.31E-03
Immunity and defense	1393	56	32.10	+	1.54E-03
DNA repair	163	15	3.76	+	1.74E-03
Carbohydrate metabolism	560	29	12.90	+	1.97E-03

Figure 3-38: Panther biological process classification results for the downregulated genes with diff. scores ≤-13 (corresponding to P-values ≤0.05) when comparing expression profiles of 8 patients with 5 different *MCPH1* mutations (EX1_9 del, S25X, T143NfsX5, W75R and T27R mutations) as one group to a group of 9 controls. The most significant P-values for the over-presentation of downregulated genes classified to each annotation term by Panther in comparision to the random expected numbers are shown in the last column.

3.7.4. Expression profiling in *MCPH1* RNAi depleted cells

In order to have an additional system for studying the expression profile of MCPH1 interaction partners (in collaboration with Dr. Trimborn and Prof. Dr. Neitzel, Charité-Universitätsmedizin Berlin) *MCPH1* was knocked down in U2OS cell lines using 3 different siRNAs (against exon 6, exon 7 and exon 8).

Three types of controls for the experiments were used: First, a siRNA against *hCAP-G*, which is known that doesn’t interact with *MCPH1*. Second, a non-silencing siRNA, and finally MOCK (includes all the reactions but any siRNA).

Transfection efficiency was monitored by checking premature chromosome condensation in metaphase preparations and *MCPH1* expression levels were determined using real time PCR (Figure 3-39), showing more than ~70% reduction in *MCPH1* expression for all of the three types of siRNAs.

Afterwards whole genome expression profiling (using Illumina Sentrix® Human-6 Expression BeadChip) was done in order to compare the expression levels of ~48000 known and unknown human genes in *MCPH1* RNAi depleted cells and controls 48 and 72 hours after transfection.

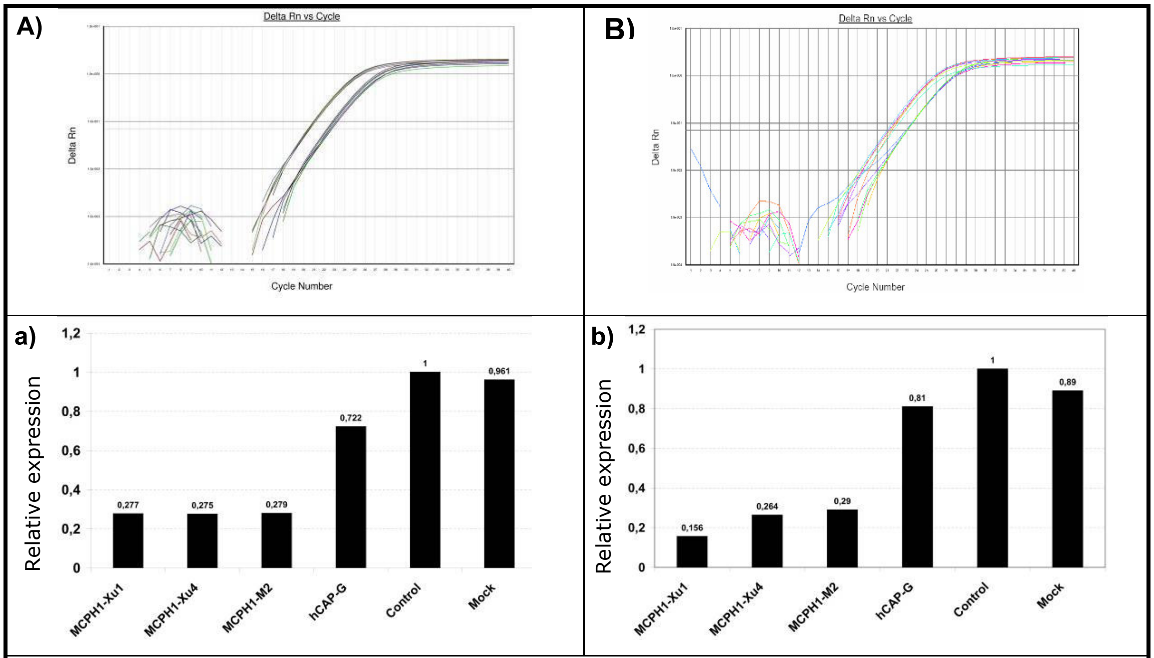


Figure 3-39: Expression analysis of *MCPH1* by quantitative real-time PCR after siRNA treatments. At both 48 hours (A and a) and 72 hours after transfection (B and b) all the three different siRNAs against *MCPH1* (MCPH1-Xu1, MCPH1-Xu4 and MCPH1-M2) reduced expression levels of *MCPH1* by more than 70%.

This experiment led to the identification of several promising potential targets for expression regulation that involves *MCPH1* (Figure 3-40).

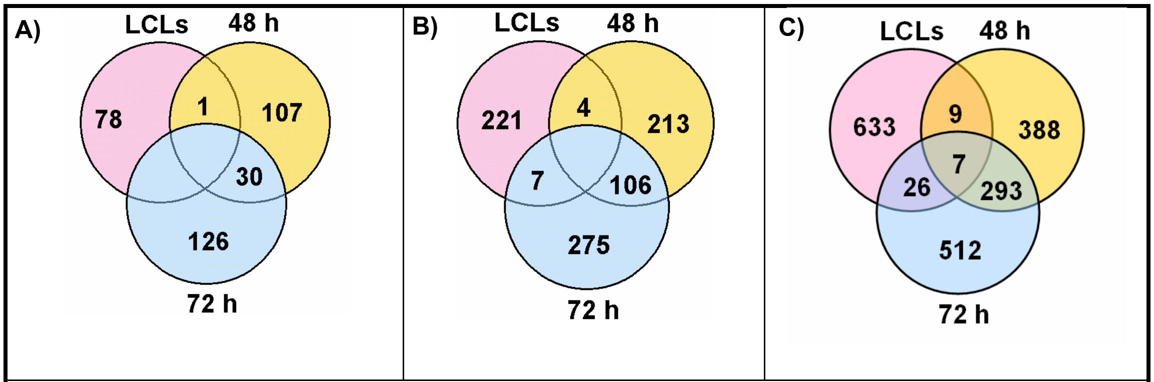
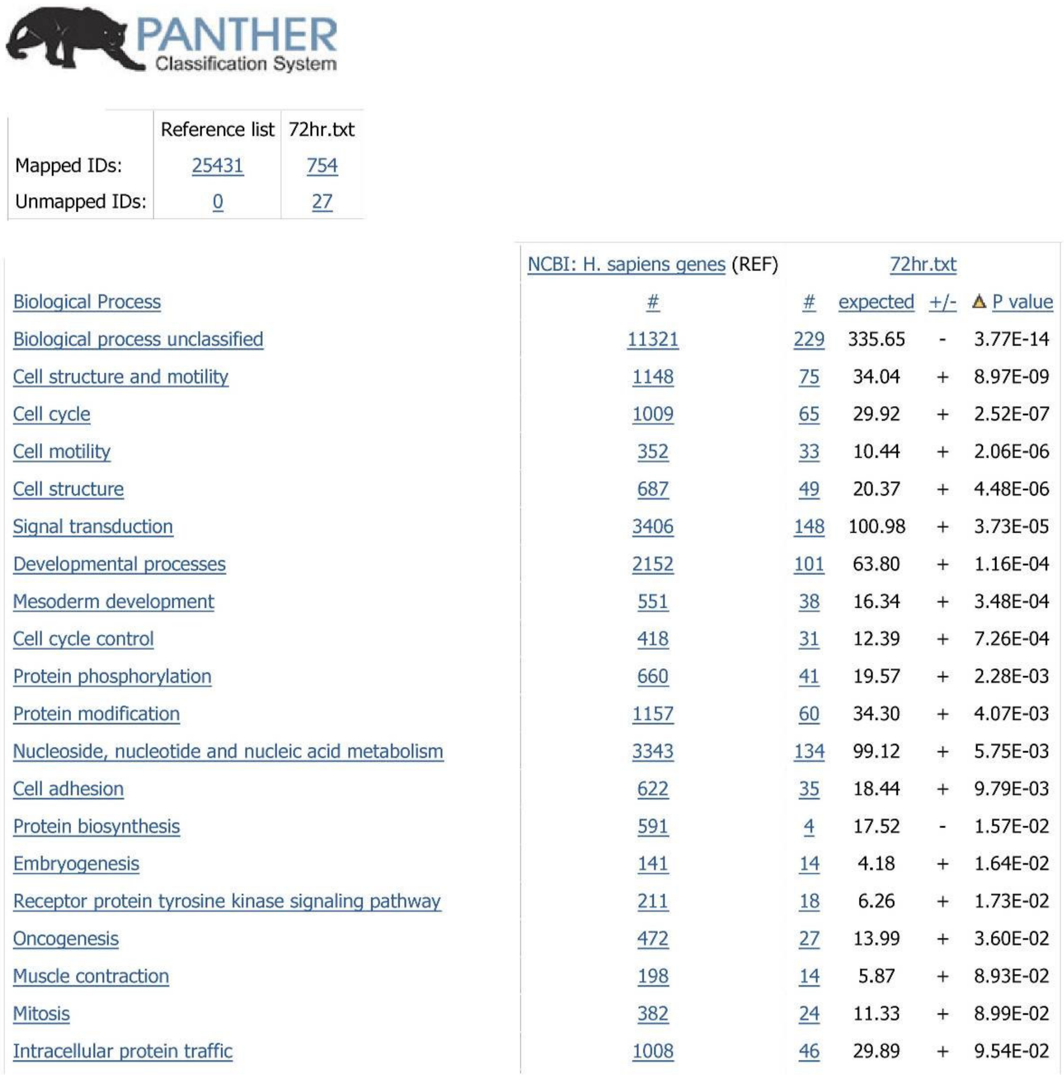


Figure 3-40: Venn diagram presentation of common downregulated genes between LCLs of patients with *MCPH1* mutations (depicted by pink circle), 48 (depicted by yellow circle) and 72 hours (depicted by blue circle) after siRNA transfection in U2OS cells with three diff. score thresholds of ≤-30 (corresponding to P-values ≤0.001) [A], ≤-20 (corresponding to P-values ≤0.01) [B] and ≤-13 (corresponding to p-values ≤0.05) [C].

Biological process classification was performed for genes that were downegulated 48 and 72 hours after transfection in U2OS cells treated with 3 different types of siRNAs for *MCPH1* knock down as one group in comparison to the 3 different controls as reference group. Genes that were significantly downregulated in (diff. scores ≤−13; corresponding to P-values ≤0.05) 72 hours after transfection could be classified into a number of relevant pathways with significant P-values (Figure 3-41). This was not observed for data obtained 48 hour after transfection (data not shown). This might indicate that the influence of MCPH1 on the regulation of other genes is mediated by other proteins and therefore the changes in the expression level of other relevant genes in response to the RNAi knock down of *MCPH1* needs more time.



	Reference list	72hr.txt
Mapped IDs:	25431	754
Unmapped IDs:	0	27

	NCBI: H. sapiens genes (REF)	72hr.txt			
Biological Process	#	#	expected	+/-	▲ P value
Biological process unclassified	11321	229	335.65	-	3.77E-14
Cell structure and motility	1148	75	34.04	+	8.97E-09
Cell cycle	1009	65	29.92	+	2.52E-07
Cell motility	352	33	10.44	+	2.06E-06
Cell structure	687	49	20.37	+	4.48E-06
Signal transduction	3406	148	100.98	+	3.73E-05
Developmental processes	2152	101	63.80	+	1.16E-04
Mesoderm development	551	38	16.34	+	3.48E-04
Cell cycle control	418	31	12.39	+	7.26E-04
Protein phosphorylation	660	41	19.57	+	2.28E-03
Protein modification	1157	60	34.30	+	4.07E-03
Nucleoside, nucleotide and nucleic acid metabolism	3343	134	99.12	+	5.75E-03
Cell adhesion	622	35	18.44	+	9.79E-03
Protein biosynthesis	591	4	17.52	-	1.57E-02
Embryogenesis	141	14	4.18	+	1.64E-02
Receptor protein tyrosine kinase signaling pathway	211	18	6.26	+	1.73E-02
Oncogenesis	472	27	13.99	+	3.60E-02
Muscle contraction	198	14	5.87	+	8.93E-02
Mitosis	382	24	11.33	+	8.99E-02
Intracellular protein traffic	1008	46	29.89	+	9.54E-02

Figure 3-41: Panther biological process classification results for the downregulated genes with diff. scores ≤−13 (corresponding to P-value ≤0.05) when comparing U2OS MCPH1 depleted cells using three different siRNAs with 3 different controls at 72 hours after transfection. The most significant P-values for the over-presentation of downregulated genes classified to each annotation term by Panther in comparison to the random expected numbers are shown in the last column.

In order to find the common genes (which most probably they have to be related to the functions of *MCPH1*) between the two different types of experiments -ie, LCLs and RNAi expriments- and different genes (they may explain the differences between the results of patient cells and *in vitro* studies) the downregulated genes with diff. scores ≤-13 in patient LCLs and *MCPH1* depleted U2OS cells at 72 hours after transfection were compared. This revealed 33 common genes.

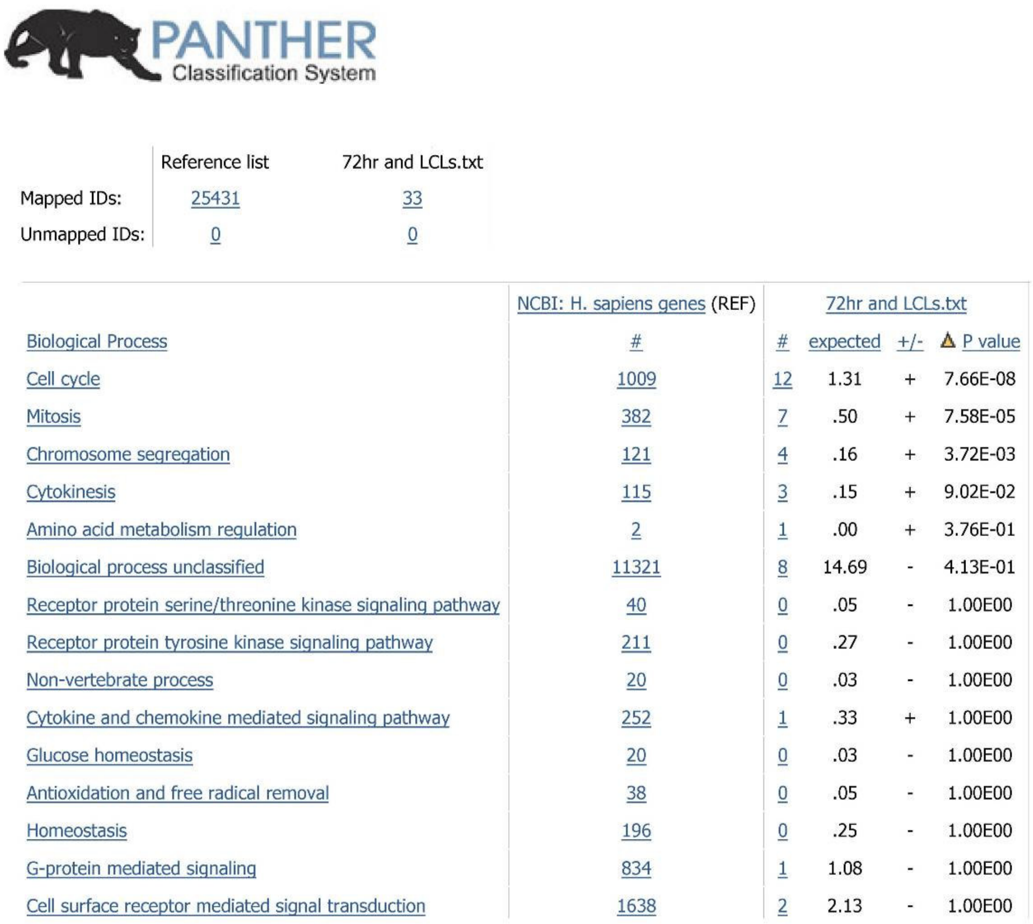
Diff. scores for all of these genes in patient LCLs, 48 and 72 hours after siRNA transfection are listed in (Table 3-7), and those that are downregulated in all three categories are typeset in boldface.

Table 3-7: Common downregulated genes between patients LCLs and MCPH1 knock down cells at 72 hours after siRNA transfection.

Symbol	LCLs diff. scores	72 hr after transfection diff. scores	48 hr after transfection diff. scores
CENPA	-35.1217	-23.5155	-5.858
PLK4	-34.6664	-22.7645	4.1992
KIF11	-32.3268	-13.7236	**-13.2579**
MELK	-31.1405	-13.973	-1.8945
PARP9	-30.2662	-13.2186	-3.2086
IQGAP3	-28.7916	-47.2051	-1.8608
TUBA1	-28.6824	-14.5899	-6.3779
KIF23	-25.0371	-24.764	1.4691
CDCA2	-24.9985	-22.5005	2.3953
PRC1	-24.0274	-13.1223	0.2642
IMPA2	**-23.0703**	**-18.5035**	**-22.8916**
PKM2	-21.9523	-30.4276	-0.3601
C17orf53	-21.6486	-14.9102	-1.6423
APOBEC3B	**-20.1362**	**-13.7452**	**-18.0433**
FLJ40629	-18.9561	-17.9409	1.68
ESPL1	-18.7878	-15.7337	-2.9096
BCKDK	-18.0679	-13.0987	-9.3826
LAG3	**-17.8591**	**-44.6367**	**-68.7041**
TYMS	-17.5165	-16.4167	-4.682
PPME1	-17.0378	-22.8081	-5.4711
EBP	**-17.0118**	**-33.9949**	**-16.1807**
SHMT1	**-16.8087**	**-53.8862**	**-20.6537**
FLJ20364	-16.4627	-15.4024	-2.924
CDC20	-15.9258	-19.6191	-10.4007
NCKIPSD	-15.2783	-14.0189	-12.9848
GPSM2	-14.8067	-18.8618	-0.8315
C1orf112	-14.7392	-16.4556	-1.2776
YIF1B	-14.6376	-18.2579	-4.5508
PLK1	-14.6329	-38.2993	-4.2306
DACT1	**-14.4151**	**-47.236**	**-17.5173**
KIF20A	-14.0945	-19.0953	-8.7368
RFC5	-13.4457	-16.828	-6.9323
UHRF1	-13.0536	-13.8466	1.4105

Functional annotation clustering using DAVID database showed significant enrichment of these genes in expected pathways such as cell cycle, mitotic cell cycle, mitosis, and cell division (see Fig S-46 in appendix-H).

Repeating this analysis with the Panther database led to similar results (Figure 3-42).



	Reference list	72hr and LCLs.txt
Mapped IDs:	25431	33
Unmapped IDs:	0	0

	NCBI: H. sapiens genes (REF)	72hr and LCLs.txt			
Biological Process	#	#	expected	+/-	P value
Cell cycle	1009	12	1.31	+	7.66E-08
Mitosis	382	7	.50	+	7.58E-05
Chromosome segregation	121	4	.16	+	3.72E-03
Cytokinesis	115	3	.15	+	9.02E-02
Amino acid metabolism regulation	2	1	.00	+	3.76E-01
Biological process unclassified	11321	8	14.69	-	4.13E-01
Receptor protein serine/threonine kinase signaling pathway	40	0	.05	-	1.00E00
Receptor protein tyrosine kinase signaling pathway	211	0	.27	-	1.00E00
Non-vertebrate process	20	0	.03	-	1.00E00
Cytokine and chemokine mediated signaling pathway	252	1	.33	+	1.00E00
Glucose homeostasis	20	0	.03	-	1.00E00
Antioxidation and free radical removal	38	0	.05	-	1.00E00
Homeostasis	196	0	.25	-	1.00E00
G-protein mediated signaling	834	1	1.08	-	1.00E00
Cell surface receptor mediated signal transduction	1638	2	2.13	-	1.00E00

Figure 3-42: Panther functional annotation clustering for common downregulated genes with diff. scores ≤–13 (corresponding to P-value of ≤0.05) between LCLs and 72 hours after siRNA transfection. The most significant P-values for the over-presentation of downregulated genes classified to each annotation term by Panther in comparision to the random expected numbers are shown in the last column.

Interestingly, from the 33 common downregulated genes with diff. scores ≤–13 (corresponding to P-values ≤0.05) between patient LCLs with *MCPH1* mutations and U2OS *MCPH1* depleted cells (72 hours after siRNA transfection) 12 were classified as belonging to the cell cycle pathway by the Panther database with a highly significant P-value of 7.6 e^{-8} (Table 3-8).

Table 3-8: list of genes classified in cell cycle by Panther among common downregulated genes with diff. scores ≤–13 (corresponding to P-values of 0.05) between LCLs and U2OS MCPH1 depleted cells 72 hours after siRNA transfection.

Symbol	Name	PANTHER Biological Process
ESPL1	extra spindle poles like 1	Mitosis
IQGAP3	IQ motif containing GTPase activating protein 3	Intracellular signaling cascade; Cell cycle control
RFC5	replication factor C (activator 1) 5	DNA replication; DNA replication
CDC20	CDC20 cell division cycle 20 homolog	Proteolysis; Cell cycle control
UHRF1	ubiquitin-like, containing PHD and RING finger domains, 1	Nucleoside, nucleotide and nucleic acid transport; Transport; Other cell cycle process; Cell proliferation and differentiation
TUBA1	tubulin, alpha 1	Intracellular protein traffic; Chromosome segregation; Cell structure; Cell motility
KIF11	kinesin family member 11	Protein targeting and localization; Chromosome segregation
PRC1	protein regulator of cytokinesis 1	Cytokinesis

Table 3-8: list of genes classified in cell cycle by Panther among common downregulated genes with diff. scores ≤−13 (corresponding to P-values of 0.05) between LCLs and U2OS MCPH1 depleted cells 72 hours after siRNA transfection.

Symbol	Name	PANTHER Biological Process
KIF23	kinesin family member 23	Chromosome segregation
KIF20A	kinesin family member 20A	Intracellular protein traffic; Protein targeting and localization; Meiosis; Cytokinesis; Chromosome segregation; Cell proliferation and differentiation; Cell structure
GPSM2	G-protein signalling modulator 2	G-protein mediated signaling; Cytokinesis
PLK4	polo-like kinase 4	Protein phosphorylation; Embryogenesis; Cell cycle

Furthermore, recently regulation of *CHK1* and *BRCA1* through interaction of *MCPH1* with the transcription factor E2F1 was described. There It is shown that MCPH1 regulates other E2F target genes involved in DNA repair and apoptosis such as *RAD51, DDB2, TOPBP1, p73, P107, APAF1*, caspase 3 and 7 (Yang and others 2008).

By comparing these findings with our data, even though the data from the patient cell lines and MCPH1 RNAi depleted cells were not compatible with each other (which was expected according to the previous reports) but still it was possible to see some similarities between these data and our findings in patient cells or RNAi data separately.

For example, at both of 48 and 72 hour transfection some isoforms of BRCA1, E2F2 and CASP2 were downregulated in all the three different siRNAs expriments (Appendix-I). But our data from the patient cells showed downregulation for *CHEK1, CHEK2, RAD51* and *CASP7* and upregulation for *CASP10* (Appendix-J).

4 Discussion:

MR is the costliest socioeconomic disease condition in developed societies. So far, almost 300 different gene defects are known to give rise to MR (Inlow and Restifo 2004), but their total number may run into the thousands, and most of them are still unknown.

MR is extremely heterogeneous and can be due to environmental factors (malnutrition during pregnancy, environmental neurotoxicity, premature birth, perinatal brain ischemia, fetal alcohol syndrome, pre- or post-natal infections), genetic factors (including chromosomal abnormalities such as aneuploidies, microdeletion syndromes and gene mutations) or it can be due to a combination of genetic and non-genetic factors (multifactorial inheritance). The precise cause, however, is found only in about 50% of cases with moderate to severe MR, and in an even lower proportion of individuals with mild MR. Apart from numeric chromosomal aberrations such as trisomy 21, that account for about 1.2/1000 live births, compelling evidence is suggesting that sub-telomeric rearrangements, as a group, may account for 5–7% of syndromic forms of MR (Chelly and others 2006).

The strategy of choice for the identification of genes underlying autosomal recessive disorders is homozygosity mapping in extended consanguineous families, followed by mutation screening of candidate genes.

Prior to these studies, one affected individual in each family was subjected to tandem mass spectrometry to exclude disorders of the amino acid, fatty acid (e.g. phenylketonuria) or organic acid metabolism. Furthermore karyotyping and Southern blotting was performed to exclude cytogenetically visible chromosomal aberrations and fragile X syndrome respectively. Patients from several families were found to have known metabolic disorders, chromosomal aberrations or fragile X syndrome. This was so helpful as helped us to detect several families with metabolic disorders, chromosomal abnormalities or fragile X which they were excluded from further analysis.

In the remaining families, Affymetrix 10k SNP-arrays were employed to map recessive gene defects. In addition to that at least one affected individual per family was analysed by high-resolution array CGH in order to detect small micro deletion or duplications which are not visible by routine chromosome analysis. We did this because they seem to be a major cause of MR. Recent publications have established that between 5 an 10% of patients with MR carry causative deletions or duplications (Shaw-Smith and others 2004; Vissers and others 2003).

This is reflected by our results, as the majority of the mutations found in our cohort of Iranian MR families so far were microdeletions (2 deletions in *MCPH1* and one in *TUSC3* in this study and a microdeletion in *GRIK2*; Motazacker and others 2007).

Later on by emerging denser SNP arrays as they can also be employed for detecting copy number variation we employed them for our genotypings. By this ability to view and score structural genomic variation there was no need anymore to do array CGH analysis.

After genotyping the first crucial step is to exclude sample swaps or related mishaps during different steps of sample collection and processing. By employing different methods of quality control such as checking relationship between the samples or their gender we were able to find several inconsistencies. Those that we were not able to solve them, were excluded from any further analyses. This is particularly important as most of the times, analysing with the inclusion of these problematic samples yields to the completely wrong result which it is almost impossible to be recognized later on. Obviously the rest of experiments that will be built on such a result are worthless without knowing.

In most of the families, whole genome SNP typing showed several different linkage peaks, reflecting the limited size of these families and/or a high degree of inbreeding.

Individuals whose parents are related are expected to be homozygous for a portion of their autosomal genome. The more closely the parents are related, the larger this portion will be. The probability that the offspring of first cousins will be homozygous by descent for one of the four great-grand-parental alleles at a given locus is 1/16. These homozygous regions could occur due to autozygosity or because of a second copy of the same allele that has entered the family independently. The rarer the allele is in the population, the greater the likelihood that homozygosity represents autozygosity. Therefore for an infinitely rare allele, a homozygous affected child born to first cousins generates a LOD score of Log_{10} (16) = 1.2. If there are three other affected sibs who are also homozygous for the same rare allele, the LOD score is 3.01 (=Log_{10} (16×4×4×4); because the chance that a sib would have inherited the same pair of parental haplotypes even if they have nothing to do with the disease is 1 in 4).

With the probability of 1/16, the expected commulative length of shared homozygous intervals will be equal to 200 Mbp for the 3.2 billion bp genome of a child born to first cousin consanguinity. Similarly the cumulative size of shared homozygous regions among two, three and four sibs of a first cousin marriage will be expected to be 50, 12.5 and 3.125 Mbp respectively. In practice, our findings show a similar picture as well, but with bigger sizes for some of the families which this pretty well reflects the high background of consanguinity in those cases.

This is fitting also with a recent publication where they show that in first-cousin offspring, the average size of individual homozygous segments is in the order of 20 cM but in populations where prolonged parental inbreeding has led to a background level of homozygosity that is ~5% higher than the value predicted by simple models of consanguinity, and therefore the average size of homozygous intervals was found to be 26 cM (Woods and others 2006).

In small inbred families, genotyping healthy sibs can be very helpful for excluding homozygous intervals that are not relevant for the disease. Therefore whenever available, at least 2 healthy sibs were included in the analysis of each small family.

In theory, the probability to find mutations in each of the intervals of multiple peak families with the same LOD scores is equal but with a better chance for the bigger intervals because it is rather unlikely that a big region of genomic DNA remains intact through the generations without being interrupted by recombinations. Small regions of the genome can escape from recombination and remain homozygous completely by chance. Furthermore, in some cases very small intervals in families can occur due to a failure to detect heterozygous positions as a consequence of the detection limit of the genotyping array that has been used. Also one has to keep in mind that all humans are related, if we go back far enough. Due to this fact, two supposedly unrelated individuals may also share some small ancestral haplotypes. An increased probability for larger linkage intervals to be really conserved and to carry homozygous disease causing mutations was also observed in our cohort of families: in all the families with multiple peaks where we were able to identify the causative mutations, it was found inside the biggest interval of the respective family, while mutation screening in very small intervals was often not successful. One example for this was family M005 where screening for all of the genes in the only very small interval of homozygosity (1.8 mbp) yielded no mutations. However, even though this is probably due to the reason explained above (random homozygosity) it might also mean that the causative mutation lies outside the protein coding regions in this interval.

Another example was family M307 where autozygosity mapping using the 50k affymetrix SNP array led to the identification of 2 small regions (0.5 and 1.2 Mbp) on both sides close to the centromere of chromosome 6. Amplification of the whole region of the first interval at the upstream of the centromere (0.5 Mbp in size) by overlapping long range PCR amplicons (5 to 10 Kbp) followed by Solexa sequencing led to the identification of several heterozygous variants which had not been detected by the SNP array (appendix-K). Sequencing of the coding exons and exon–intron boundaries of the only gene (*EGFL11*) in the second interval at the downstream of the centromere revealed one heterozygous change in the middle of the interval. Thus these results can be considered as an example for a shared small ancestral haplotype with several recent polymorphisms that were not detected due to resolution limits of the employed array.

Yet another example was family M150N where autozygosity mapping revealed only one small region on chromosome 5. Later on *FMR1* turned out to be mutated in this family, although the pedigree was not suggestive for X-linked inheritance. This means that even in case of families with X-linked problems there is a chance to find regions of homozoygosity on the autosome chromosomes. The other point will be that the pedigree based discrimination of the X-linked families from the autosomal forms has to be handled very carefully.

Therefore, one has to be cautious about very small solitary intervals especially in families with only one branch as these intervals might be evolutionarily conserved ancestral haplotypes.

Finally, it is also possible that very small apparently homozygous regions are identical by state (IBS) but not identical by descent (IBD); particularly if these regions are only defined by few informative markers. This is, however, rather unlikely for large haplotypes with many informative markers.

In total linkage analysis and homozygosity mapping in 183 families (41 from another study conducted in parallel to the one presented here) that fulfilled our selection criteria revealed novel solitary linkage intervals in 38 families.

23 out of 38 novel solitary linkage intervals had significant LOD scores of above 3 and therefore represent novel gene loci for ARMR. In 15 families with single linkage intervals, the LOD scores are too low (between 2 and 3) to formally prove that these sites represent additional MR loci. Still this is likely for most of them if we accept that the condition is monogenic and autosomal recessive, then inevitably the mutation must be somewhere in the genome. Therefore the only autozygous region in a given family will still be the most likely site of the disease-causing mutation, even if the LOD score does not reach the canonical value of 3 because the size of the family is limiting.

Finding 38 families with novel solitary intervals which they do not coindide considerably is the first suggestive data in such a scale for the high heterogeneity of ARMR at least in the Iranian population.

The heterogeneity rate is even more if those families with two or more small intervals [so that the commulative length of these intervals is even physically smaller than the only one region in the stand-alone families and moreover, the mutation must be somewhere in one of these intervals (see the above argument)] will be considered as their roughly coincidation won’t be more than 10 % as well.

This heterogeneity of NS-ARMR is not surprising as the brain is a tissue where most of the genes are expressed, and therefore theoretically any defect in one of these genes could potentially lead to a perturbation of the networks in which the gene is involved and cause MR.

The first 8 NS-ARMR loci that were found (Najmabadi and others 2006) were named as ‘Mental Retardation 4 to 11’ (MRT4–11), in accordance with the nomenclature used for previously mapped NS-ARMR loci (OMIM #249500, #607417, #608443).

Since then, we have identified 20 additional solitary intervals for NS-ARMR and 6 for syndromic forms of ARMR, as reported here.

Even though some of the new intervals include known genes for NS-ARMR (*PRSS12*, *CRBN*, *CC2D1A* and *GRIK2*), no second mutation was found in any of these genes. This indicates that none of the previously found genes play an important role in NS-ARMR. On the other hand and partially contradicting our previous conclusion that NS-ARMR is extremely heterogenous, we have eventually found three loci on chromosomes 1p34.3-p34.1, 5p15.32-p15.2 and 19q13.2-q13.31 respectively, where solitary autozygous regions of three, four and two independent families overlap. These loci also coincide with homozygous regions in other families with more than one linkage interval.

Furthermore, the distribution of homozygous intervals in the genome (including those from families with more than one autozygous region) is far from being even, with several regions showing conspicuous clustering of linkage intervals. These findings argue strongly for the existence of NS-ARMR genes that are involved in more than one family.

By identifying a deletion inactivate *TUSC3* gene in an Iranian family and the independent report of a *TUSC3* mutation by a French group (Molinari and others 2008), we have discovered the first gene for NS-ARMR which is involved in more than one family.

TUSC3 is believed to be the ortholog of the yeast Ost3 protein that was initially identified as a 34-kD subunit in the yeast oligosaccharyltransferase (OST) complex (Kelleher and Gilmore 1994; Kelleher and Gilmore 2006; MacGrogan and others 1996). It is expressed in a wide range of human tissues, including the brain. *TUSC3* has 11 exons spanning ~224 Kbp of the genomic DNA on chromosome 8p22. According to the UniProtKB database, *TUSC3* encodes a predicted 348-amino acid protein with five potential transmembrane domains (Figure 4-1) 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.

We found a deletion mutation with the size of 120Kbp including exon 1 and the promoter region of *TUSC3*; thereby affecting several predicted functional protein domains, as well as almost the entire portion of this gene that shows homology to other, possibly functionally related genes (Figure 4-1).

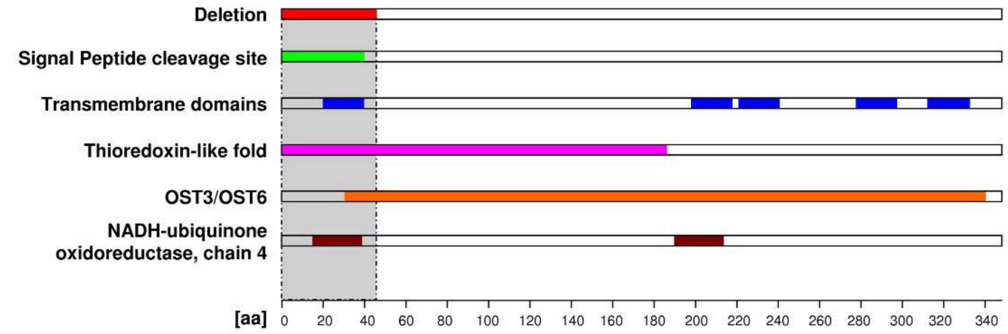


Figure 4-1: Schematic representation of predicted functional domains in the *TUSC3* gene product: The 348 amino acid TUSC3 protein is shown as an open box. Different functional domains are indicated. Differently colored shading marks their extent and position within the protein. The deletion encompasses the first 46 amino acids and affects all of the predicted functional domains (grey shading).

Unlike other patients with congenital disorders of glycosylation (CDG) which are characterized by ataxia, seizures, retinopathy, liver fibrosis, coagulopathies, dysmorphic features and ocular abnormalities (Jaeken and Matthijs 2007), our patients only present with non-syndromic MR. An explanation for the conspicuous absence of additional symptoms in our patients may be the presence of a closely related gene on Xq21.1, which encodes the Implantation-Associated Protein precursor (IAP/MAGT1). MAGT1 is also assumed to be involved in N-glycosylation through its association with N-oligosaccharyl transferase (Kelleher and others 2003). It might thus be able to partly compensate for the loss of TUSC3, probably in a tissue specific manner. Our finding that affected individuals show no aberrant glycosylation of serum transferrin (as determined by isoelectric focussing) is in keeping with this speculation.

However, the assumption that TUSC3 plays a role in protein glycosylation is solely based on its 20% sequence similarity with the yeast *Ost3* gene (MacGrogan and others 1996). Indeed, the normal glycosylation patterns seen in serum of TUSC3 deficient patients may argue against a central role of this protein in the glycosylation process. Similarly, the fact that none of these patients has a history of cancer casts doubt on the original assumption that TUSC3 acts as tumor suppressor (Bova and others 1996; MacGrogan and others 1996). As to 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) (Rual and others 2005). Protein phosphatase 1 has been implicated in the modulation of synaptic and structural plasticity (Munton and others 2004) and was shown to have an impact on learning and memory in mice (Genoux and others 2002). It is therefore conceivable that MR in TUSC3 deficient patients is caused by an impairment of PPPC1A function. This opens up interesting perspectives for future studies into the function of *TUSC3*.

Syndromic autosomal recessive mental retardation

There are several examples of conditions that originally were considered to be non-syndromic 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 MR is not easy. Judging from the relative frequencies of syndromic and non-syndromic X-linked MR, syndromic forms of ARMR are probably more common than are non-syndromic ones.

Moreover, searching for the relevant gene is usually easier in syndromic MR, as it is often guided by clinical signs suggesting specific spatio-temporal gene expression patterns and sometimes, clinical features are specific enough to establish a tentative diagnosis. This was the case in family 8600004, where clinical symptoms pointed to Sjoegren-Larsson Syndrome (SLS) [MIM:270200]. SLS is caused by defects in Aldehyde dehydrogenase 3A2 isoform 2 *(ALDH3A2)* (Gordon 2007). Presence of characteristic symptoms such as severe MR, ichthyosis (hyperkeratosis), short stature and spastic paraplegia in our patients prompted us to screen this gene for mutations, which led to the identification of a splice site mutation.

ALDH3A2 encodes fatty aldehyde dehydrogenase (FALDH) which catalyzes the oxidation of long-chain aliphatic aldehydes to fatty acids and acts on a variety of saturated and unsaturated aliphatic aldehydes between 6 and 24 carbons in length. It is likely that the biochemical pathogenesis of SLS originates from accumulation of lipid substrates that cannot be metabolized by FALDH and/or their diversion into other metabolic products; or deficiency of critical fatty acid products of FALDH (Rizzo 2007).

In other families clinical findings were very helpful for prioritizing positional candidate genes for mutation screening, as in a family with dysequilibrium syndrome a nonsense mutation in the very low-density lipoprotein receptor gene (*VLDLR*) was found (Moheb and others 2008) and in two families (M152 and M179) with Cohen syndrome where we found mutations in the *COH1* gene (Seifert and others 2008).

In a family with ataxia and MR (M107) we could identify a R237Q mutation in exon 7 of carbonic anhydrase VIII (CA8). The protein encoded by this gene shows a high sequence similarity with other known carbonic anhydrase genes, but lacks carbonic anhydrase activity (i.e., the reversible hydration of carbon dioxide). Instead, Hirota and others (2003) have shown to interact with the inositol 1,4,5-trisphosphate (IP3) receptor1 (IP3R1) an intracellular IP3-gated Ca^{2+} channel that is located on intracellular Ca^{2+} stores. This is one of several factors that modulate the ability of ITPR1 to rapidly release calcium stores from the endoplasmatic reticulum in Ca^{2+} signalling. Modulation of intracellular calcium is important for a number of cerebellar functions such as long-term depression (Aiba and others 1994). Western blot analysis has revealed that *CA8* is expressed exclusively in Purkinje cells of the cerebellum, in which IP3R1 is abundantly expressed.

The cerebellum is a complex neurological structure, containing more than half of the brain's total number of neurons. Cerebellar networks show long-term synaptic plasticity, which indicates that experience-dependent adaptive and learning processes are a salient feature of cerebellar function. Most afferent information enters the cerebellum via climbing fibers (CF) and mossy fibers, which excite the Purkinje cells indirectly through the parallel fiber (PF) pathway. CA8 inhibits IP3 binding to IP3R1 by reducing the affinity of the receptor for IP3 (Hirota and others 2003). Therefore, we speculate that the consequences of a CA8 mutation may involve improper modulation of the ITPR1 with resultant functional and/or developmental defects in the cerebellum.

Additionally in this family, the clue to the identification of this gene defect was the absence of *CA8* gene transcription in the cerebellum of the lurcher mutant in mice, which gives rise to neurological defects (Kelly and others 1994), and the waddles (wdl) mouse with a 19-bp deletion in exon 8 of the carbonic anhydrase-related protein VIII gene (Car8), which results in ataxia and appendicular dystonia as most conspicuous signs (Jiao and others 2005). The wdl phenotype is very similar to the ataxia observed in our patients, which is a strong indication that the observed *CA8* mutation is indeed causative.

In parallel to this finding our collaborators in Charite-Universitätsmedizin Berlin identified another mutation in *CA8* in an Iraqi family with ataxia and mild mental retardation (Turkmen and others 2009).

In this family the healthy parents were first cousins, and four of eight sibs were affected. The parents claimed that the affected persons never learned to crawl on their knees as most infants do, but ambulated from infancy on with their legs held straight with a "bear-like" gait. They also claimed that attempts to teach the children to walk on two legs with crutches or other supports failed. They walked with straight legs, placing weight on the palms of their hands. Although the affected members were able to walk on two legs for several steps, they tended to tumble into a quadrupedal position quickly, complained of lack of balance and occasionally fell from a sitting position.

So far mutations in *CA8* and *VLDLR* (Ozcelik and others 2008; Turkmen and others 2008) have been found to be associated with quadrupedal locomotion in humans, although not in all affected individuals. Given the variable incidence of quadrupedalism in individuals with mutations in the same gene, it is probable that contextual factors during development - either internal or external - contribute to this particular phenotypic outcome (Humphrey and others 2008). As one possibility, we note that ataxia associated with mutations at all two loci is congenital and also associated with cerebral defects, which are not generally a feature of other hereditary ataxias, such as Joubert syndrome (Joubert and others 1999) or AVED (Cavalier and others 1998) Thus, perhaps it is only when congenital ataxia is coupled to a certain kind of malfunction of the cerebral cortex that individuals are likely to remain walking on all fours.

It is easier to deal with syndromic forms of MR where still no causative genes are known than with pure non-syndromic forms. But sometimes even the presence of very specific clinical signs is not enough to identify the gene defect. We have reported two examples of this kind: Family M069 with MR, cataract, coloboma and kyphosis (Kahrizi and others 2008, in press) and families 8402061 and 8508395 with alopecia–MR syndrome (Tzschach and others 2008) where in both cases the genes harbouring the causative mutations are still waiting to be found.

Some syndromic forms of ARMR are also genetically heterogeneous which hampers pinpointing the gene carrying the causative mutation in such cases. An example is primary microcephaly, one of the important clinical features which is always accompanied by MR. For this condition 4 genes and at least two loci have been found so far (MCPH1-6), and linkage to the different loci has to be checked prior to mutation screening. This can be done by genotyping several markers specific for the regions instead of whole genome genotyping.

MCPH1

In the course of this study, we discovered two submicroscopic deletions on chromosome 8p in the microcephalin gene *(MCPH1)* in two different families with microcephalic patients.

One of them was a deletion encompassing the first 8 exons of *MCPH1* as well as more than 25 kbp of the 5′ flanking region. The extent of this deletion, which is the first of its kind in *MCPH1*, suggests that the truncated gene is no longer functional, and that this mutation is the primary cause of the MR seen in the affected subjects. This assumption was corroborated by earlier studies, where *MCPH1* had been found to be mutated in three other families with MR, significant microcephaly and short stature (Jackson and others 2002; Neitzel and others 2002) two of which share an ancestral 8p23 haplotype (Jackson and others 2002).

The second deletion mutation that we found later, removes only exon 4 of *MCPH1* in another Iranian family with microcephaly.

In our family with the exon 1-9 deletion, only the affected females showed distinctly shorter stature than their unaffected relatives, and head circumferences of all patients were between 49 and 50 cm, i.e. only 3 SD below the age- and sex-specific mean. Thus, compared to the previously reported patients with MCPH1 mutations (Jackson and others 2002; Neitzel and others 2002), the phenotype observed in our family was surprisingly mild, which is particularly striking since in this family, half of the *MCPH1* gene as well as the promoter region are deleted and one would expect a rather stronger effect. However more cases with only mild microcephaly (−3SD) due to the mutations in the *MCPH1* gene were described later. For example a family with a homozygous missense mutation (c.80C > G, Thr27Arg) has been reported with mild microcephaly (−3SD) and MR (intelligence quotient = 74) (Trimborn and others 2005).

Altogether and according to a recent definition of primary microcephaly (Woods and others 2005), it is even questionable whether the observed reductions of the head circumference in patients with the Ex1_9del mutation in *MCPH1* (as well as the other reported mutation with mild microcephaly) are severe enough to justify their classification as primary microcephaly. Indeed further cases will have to be studied in order to gain comprehensive insight into the range of symptoms that are characteristic for patients with *MCPH1* mutations.

Additionally we observed premature chromosome condensation (PCC; poor chromosome banding and an excess of prophase-like cells on cytogenetic analysis of peripheral blood) in all of our patients which is in line with the previous finding that PCC syndrome (Neitzel and others 2002) is the allelic form of primary microcephaly caused by *MCPH1* mutations (Trimborn and others 2004). It has been shown by analysis of patient cells with *MCPH1* mutations that PCC occurs because chromosomes condense within an intact nuclear envelope during G2 and post mitotic decondensation is delayed. Both findings strongly suggest that loss of MCPH1 function causes aberrant regulation of chromosome condensation (Neitzel and others 2002; Trimborn and others 2004). Furthermore It has been shown that patients with PCC syndrome also have periventricular neuronal heterotopias, suggesting that *MCPH1* mutations might be associated with neuronal migration phenotypes (Trimborn and others 2004). Additionally the *MCPH1* gene product, microcephalin, is expressed in fetal brain, liver and kidney, and at lower levels in other fetal and adult tissues. Moreover, in situ hybridization studies have shown that in the mouse, the orthologous *Mcph1* gene is expressed in the developing forebrain during neurogenesis.

Microcephalin has 3 BRCT motifs which are commonly found in DNA damage response proteins, particularly in those functioning as mediators in the signaling response. This is circumstantial yet provocative evidence that MCPH1 might function in a DNA damage response pathway (Jackson and others 2002).

In addition to that there is also experimental evidence that suggests that MCPH1 has a role in DNA repair following ionizing radiation damage: Hemagglutinin (HA) tagged- and endogenous MCPH1 colocalise with MDC1 and gamma-H2AX irradiation induced foci (IRIF) in response to irradiation (IR), suggesting that MCPH1 localises to the sites of DNA damage (Lin and others 2005; Xu and others 2004). Contrary to this, however, the patient cells with different mutations in *MCPH1* didn't show any evidence of DNA damage response deficiency to IR (Trimborn and others 2004). Therefore we were interested to check this in our patients with the exon 1-9 deletion. We performed IR on our patient cells with the exon 1-9 deletion as well as patients with S25X and T27R mutations. As a positive control ATM (Ataxia telangiectasia mutated) patient cells were used, since it is known that patients with ATM mutations have defects in checkpoint arrest (for review see the O'Driscoll and others 2006). In contrast to the patient cells with *ATM* mutation the mitotic index in response to radiation exposure decreased dramatically in patient cells with Ex1-9_del mutation in *MCPH1* and the control cells, meaning that LCLs of patients with *MCPH1* mutations are checkpoint proficient, as the cells react normally by stopping the cell cycle in order to give enough time to the DNA repair system to correct the DNA-aberrations that were caused by radiation.

There are even more incompatibilities between the results from the patient cells and MCPH1 in vitro studies. For example by studies employing siRNA it has been concluded that MCPH1 is required for the formation of damage response foci and additionally functions to transcriptionally regulate Chk1 and Brca1, hence acting as a crucial DNA damage regulator. This has been described by Xu and others 2004 who found that siRNA knock down of *MCPH1* is accompanied by co-knock down at the transcriptional level of Brca1 and Chk1 via an unknown mechanism. But in contrast to this it has been shown that patient cells with *MCPH1* mutations express normal protein levels of Chk1 and Brca1, and Chk1 is phosphorylated normally after DNA damage (Alderton and others 2006).

Considering these differences, the very mild microcephaly in the patients with the ex1-9 deletion and the fact that all of the mutations that have been found so far in *MCPH1* affect only the first part of the gene (including only the first BRCT domain), it is probable that the first BRCT domain has different functions from the last two domains. Therefore, it is plausible to assume the existence of an additional transcription site for the second part of *MCPH1*, which might encode a smaller isoform containing the last two BRCT domains. This small putative isoform of MCPH1 protein might play an important role in rescuing the patient cells from a complete loss of MCPH1 function.

In addition to defects in DNA damage responses, indefinite proliferation is another specialized cellular quality implicated in the development of cancer. It has been shown that 90% of human cancer cells show elevated telomerase activity resulting from reactivation of the expression of the catalytic subunit hTERT (Varon and others 1998).

Knowing that *MCPH1* was originally identified as a repressor of the transcriptional activity of hTERT, a potential role for MCPH1 in cellular immortalization and consequently in tumorigenesis would be expected (Lin and Elledge 2003). Moreover, the chromosome region including *MCPH1* is found to be frequently deleted in several malignancies such as breast, (Miller and others 2003; Thor and others 2002) ovarian (Pribill and others 2001) and prostate cancer (DeMarzo and others 2003). These characteristics in addition to the roles of MCPH1 in controlling the critical activities of DNA repair and checkpoint control, can be seen to support the idea that MCPH1 is a potential tumor suppressor gene.

However, patients with defective *MCPH1* show no signs of developing malignancies or cancer predisposition. This is even more striking in our patients with the ex1_9 deletion, some of whom are now in their forth decade of life without having developed any from of cancer. Furthermore, these patients are apparently able to marry and even reproduce normally (see pedigree of the family M019).

Thus, one of the questions still waiting to be answered concerns these differences between MCPH1 siRNA knock down cell lines and patient cell lines with *MCPH1* mutations. The first explanation that one might propose would be a siRNA off-target effect. One of our aims to compare expression profiling of the patient and MCPH1 siRNA treated cells together was to tackle this question. In fact several genes were found to be deregulated in each set which might be the cause for these differences.

Another explanation for the differing observations from patient cells and the previous *in-vitro* results might be the presence of a putative small transcript in patient cells, but as we were not able to characterize one in our patient cells with the Ex1-9_del mutation this speculation remains to be proved.

Although we were able to show the presence of transcription for the 3'UTR of *MCPH1* with RT PCR but we were not able to characterize the complete version of this putative transcript by RACE experiments. This can be due to the very low level of *MCPH1* expression in LCLs. The mechanism behind the role of this putative transcript could be more complex, it might for example include extra coding regions apart from the currently annotated exons for *MCPH1*, or it might act in a tissue specific manner, or both.

But the main question regarding the functions of MCPH1 still is its role in the development of the human nervous system and its involvement in the determination of brain size. One way of learning more about this is to study the influence of MCPH1 on the expression of other genes.

Furthermore It is repeatedly reported that *MCPH1* would regulate protein and transcript levels of other genes such as *hTERT*, *BRCA1* and *CHK1* (Lin and Elledge 2003; Lin and others 2005; Xu and others 2004). Therefore, it is speculated that microcephalin/BRIT1 may function in transcriptional regulation. Very recently an interaction of microcephalin/BRIT1 with the transcription factor E2F1 was described (Yang and others 2008). In this study, the C-terminal BRCT-domains were identified to be crucial for E2F1 binding and activation. Therefore, we investigated the effects of the MCPH1 mutations on gene expression profiling to find out more in this regard. The preliminary results of comparing the expression profiles of patient and healthy control cells were promising. Checking several of the significantly deregulated genes by Northern blotting showed a high rate of accuracy for the micro-array data. There were several promising candidates among the deregulated genes, which persuaded us to expand and repeat the analysis with another system like *MCPH1* siRNA knock down cells.

The high numbers of deregulated genes in patient cells with different mutations in *MCPH1* and *MCPH1* siRNA knock down cells might be considered as a good reason in favor of its role as a transcription factor.

Functional annotation of the downregulated genes common between the two different approaches showed highly significant ($P \leq 0.001$) enrichment in relevant pathways such as cell cycle, mitosis and DNA damage response. This might be a good indication that biological reasons for the expression perturbation of such particular genes must exist. Further characterization of the involvement of *MCPH1* in these pathways can be very valuable for a better understanding of the mechanism and pathogenesis of primary microcephaly.

By performing RNAi expression profiling at two different times after transfection (48 and 72 hr) we noticed that the function of de-regulated genes after 72 hr of transfection are more relevant to the expected pathways for *MCPH1*. This might indicate that the influence of MCPH1 on the regulation of other genes is mediated indirectly via other proteins and this is why the changes in the expression level of other relevant genes in response to the RNAi knock down of *MCPH1* needs more time.

Alos it has to be noted that there are several different probes for some of the genes on the array and the detection quality for all of these probes due to various reasons such as different expression levels of these transcript variants or technical and hybridization differences won’t be equal. Therefore we might see differences in the patterns of different probes of one gene but in such cases considering only those probes with higher detection rates will be more realistic.

The most distinct role of MCPH1 in checkpoint regulation seems to be at the G2 to M transition phase (Trimborn and others 2004) which was corroborated by our expression profiling results, which show that almost ¾ of all the downregulated cell cycle genes (based on the KEGG database) are belonging to the G2 or M phase of the cell cycle (Figure 3-37).

It has previously been shown that MCPH1 is required for the activity or the expression of *ATR*, *ATM*, *BRCA1* and *Chk1* (Lin and others 2005; Rai and others 2006), and our expression profiling data show that *Chk1* is also downregulated in patient cells.

Moreover, it has previously been suggested by a study showing that MCPH1 cells have impaired degradation of Cdc25A, identical to that observed in ATR-Seckel cells, both in unperturbed cell growth as well as following UV irradiation that MCPH1 functions downstream of Chk1 in the ATR pathway. These findings suggest that MCPH1 acts to regulate Cdc25A (Alderton and others 2006). The regulation of Cdc25A activity and stability is still poorly understood. However, there is clear evidence that Chk1 phosphorylates Cdc25A at multiple sites that can regulate both its activity and ubiquitin-dependent degradation (Boutros and others 2006).

It is known that entry into mitosis is dependent on the activation of the Cdk1-cyclin B1 complex by regulatory phosphorylation. It has also been shown that the levels of inhibitory Tyr15 phosphorylated Cdk1 (pY15-Cdk1) observed in cells released following synchronisation at the G1/S boundary, decreased rapidly in MCPH1 cell lines as compared to control cells (Figure 4-2). Therefore it has been proposed that the regulation of mitotic entry by MCPH1 is both ATR dependent (another pathway known to be involved in DNA damage response), in controlling Cdc25A degradation, and ATR independent, in regulating the Cdk1-cyclin B kinase activity (Alderton and others 2006).

Although, our expression profiling data didn’t show downregulation of the abovementioned genes, they still revealed downregulation for many other genes with similar functions in cell cycle control such as *CDC45L*, *CDC2*, *CDKN2C*, *CDC14B* and *CDC20* in line with the available literature data regarding involvement of *MCPH1* in cell cycle control.

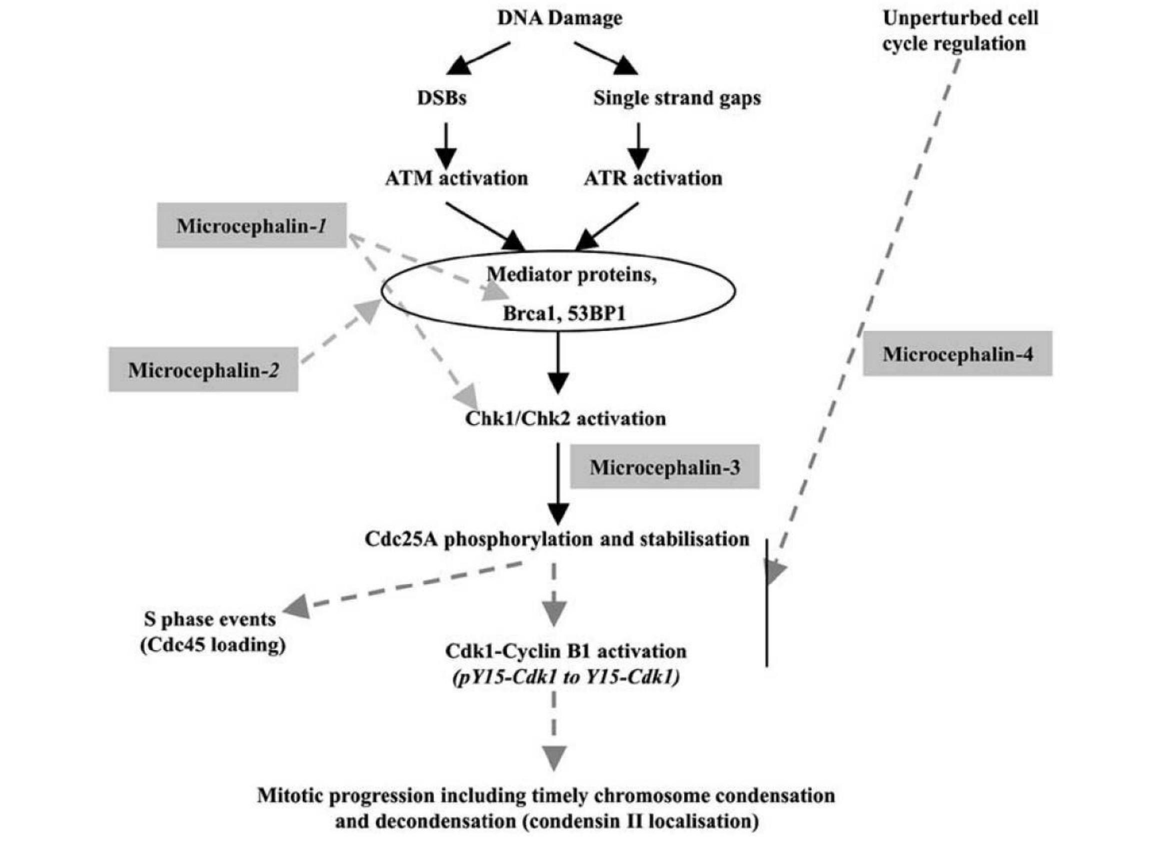

DNA Damage
Unperturbed cell cycle regulation
DSBs
Single strand gaps
ATM activation
ATR activation
Microcephalin-*1*
Mediator proteins, Brca1, 53BP1
Microcephalin-*2*
Microcephalin-4
Chk1/Chk2 activation
Microcephalin-3
Cdc25A phosphorylation and stabilisation
S phase events (Cdc45 loading)
Cdk1-Cyclin B1 activation
(pY15-Cdk1 to Y15-Cdk1)
Mitotic progression including timely chromosome condensation and decondensation (condensin II localisation)

Figure 4-2: Impacts of MCPH1 deficiency. One function of MCPH1 (depicted by microcephalin-1 box) suggested by siRNA studies is the transcriptional regulation of Brca1 and Chk1. A second function of MCPH1 (depicted by microcephalin-2 box) suggested by siRNA studies is the formation of MDC1, 53BP1, p-ATM and NBS1 foci. γ-H2AX foci form normally, however. Studies on MCPH1 cell lines have exposed an MCPH1 function that cannot be attributed to an impact on Brca1 or Chk1 expression in the DNA damage response that is downstream of Chk1 activation but impacts upon Cdc25A stabilization (depicted by microcephalin-3 box). This could represent a role in facilitating Chk1 phosphorylation of Cdc25A. MCPH1 can interact with Chk1. Finally, MCPH1 has a function that does not overlap with the DNA damage response in regulating entry into mitosis via the regulation of Y15-Cdk1 phosphorylation (depicted by microcephalin - 4 box), revealed by studies on patient cell lines (Figure and content taken from O'Driscoll and others 2006).

It has been shown that condensin II localizes to the nucleus in patient cells with *MCPH1* mutations and in some cases binds to the central chromosomal axis, even though condensin I is still in the cytoplasm (Trimborn and others 2006). In contrast, condensin II is cytoplasmically localized in normal G2 cells (Trimborn and others 2006). Although nuclear localization in some cells can be observed, its enrichment in the chromatid axis is rare. This suggests that a consequence of MCPH1 deficiency is the premature binding of condensin II to chromatin. Furthermore, it is shown that siRNA depletion of condensin II subunits is able to alleviate the PCC phenotype as well as the delayed post-mitotic decondensation phenotype (Trimborn and others 2006). It is also shown that simultaneous depletion of condensin II and MCPH1 by siRNA in Hela cells prevents PCC. Nevertheless, knock down of condensin I doesn’t impact upon the PCC phenotype (Trimborn and others 2006).

We observed no significant changes in the expression levels of the condensin I and II indicating that MCPH1 does not influence their transcription. However, we saw downregulation of 15 genes (*BUB1, BUB1B, CDCA1, CENPE, CHL1, CSPG6, KIF11, KIF20A, KIF23, MAD2L1, PTTG1, PTTG2, RAD21, SMC2L1* and *TUBA1*) which are known to be involved in chromosome segregation and maintenance. This finding can be very important, as another feature observed in MCPH1 cell lines is the presence of mitotic cells with supernumerary centrosomes in up to 30% of the cells, suggesting that MCPH1 is probably involved in regulation of centrosome stability (Alderton and others 2006).

This function can have important consequences especially in the neuroepithelial cells as they are the primary neural progenitors from which all other CNS progenitors and — directly or indirectly — all CNS neurons derive. Prior to neurogenesis, the neural tube wall is consisting of only single-layered neuroepithelial cells which are extended from the apical (ventricular) surface to the basal lamina (apical–basal polarity) (Huttner and Kosodo 2005).

Neural progenitor cells can have symmetric divisions in which the mitotic spindle is in the plane of the neuroepithelium and yield two neural progenitor cells, or asymmetric divisions, which occur when the mitotic spindle is oriented perpendicular or oblique to the neuroepithelium and give rise to one postmitotic neuron and one progenitor cell (Figure 4-3).

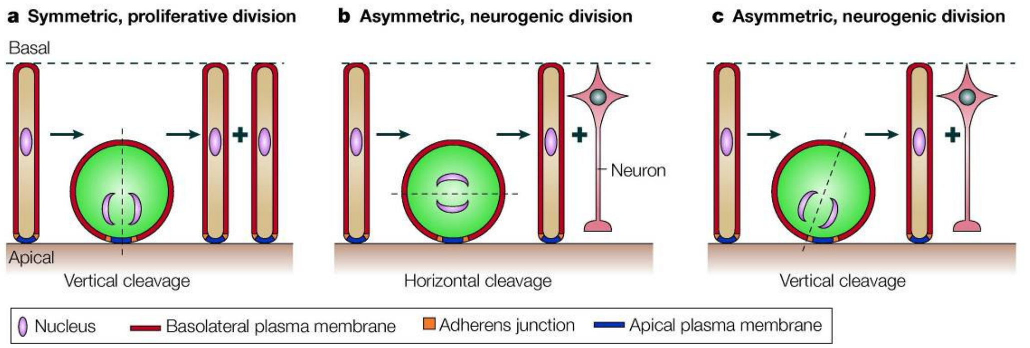

a Symmetric, proliferative division
b Asymmetric, neurogenic division
c Asymmetric, neurogenic division
Basal
Apical
Neuron
Vertical cleavage
Horizontal cleavage
Vertical cleavage
Nucleus
Basolateral plasma membrane
Adherens junction
Apical plasma membrane

Figure 4-3: a) Symmetric versus asymmetric division of neuroepithelial and radial glial cells. The figure summarizes the relationship between apical–basal polarity, cleavage-plane orientation and the symmetric, proliferative versus asymmetric, neurogenic division of neuroepithelial and radial glial cells. a) Vertical cleavage results in a symmetric, proliferative division. B) Horizontal cleavage results in an asymmetric, neurogenic division. C) Vertical cleavage results in an asymmetric, neurogenic division (Figure and content taken from Gotz and Huttner 2005).

Therefore, spindle pole orientation can play a critical determining role in the normal brain development. In other words, brain development might be highly sensitive to any perturbation in the maintenance of centrosome stability and function.

Thus MCPH1 might play a central role in brain growth during development by regulating centrosome stability and correct spindle pole orientation (O'Driscoll and others 2006). This effect can be mediated by some of the deregulated genes found by our expression profiling analysis that are involved in different aspects of chromosome segregation and maintenance.

Finally observing several common downregulated genes between both approaches (studying patient cells with *MCPH1* mutations and MCPH1 RNAi depleted cells, see Table 3-8) is siginificantly meaningful especially that the function of these genes seems to be relevant to the function of *MCPH1*. For example CDC20 (cell division cycle protein 20) is one of these genes which appears to act as a regulatory protein interacting with several other proteins at multiple points in the cell cycle and it is required for two microtubule-dependent processes, nuclear movement prior to anaphase and chromosome separation. This is the same in case of several members of Kinesin family members (*KIF11, KIF23* and *KIF20A*) which are mainly involved in chromosome segregation and cell proliferation. Therefore involvement of MCPH1 in the cell cycle or chromosome condensation probably happens in concert with the proteins encoded by these genes.

Outlook:

Our results showed a high genetic heterogeneity for NS-ARMR by revealing many new loci, some of which are very big and include many genes. This makes it impossible to screen all of the genes by routine ways of sequencing. However, emerging the next-generation sequencing systems and oligonucleotide arrays should greatly facilitate mutation screening in the nearest future.

There is enough reason to believe that a significant proportion of the genetic variation causing or predisposing for disease involves non-coding sequences and there is no doubt that these methods will revolutionize genotype–phenotype comparisons in man, but at the same time, they will greatly aggravate the problem of how to make sense of all these newly uncovered genetic variation. Therefore, recognizing clinically relevant changes, in a sea of functionally neutral sequence variants, will be a considerable challenge which can only be met by studying very large cohorts of clinically well-characterized patients (Ropers 2008).

We are now trying to adjust the currently available next-generation of sequencing systems for this purpose, which are based on DNA fragmentation and massively parallel clonal amplification of these fragments, followed by multiplex pyrosequencing (454-Roche) or stepwise incorporation of fluorescent dye–labeled nucleotides (Solexa-Illumina) and visualization by sensitive detection systems.

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 of genetic diseases by performing universal carrier screen combined with preimplantation genetic diagnosis for carrier couples who want biological children.

Another important issue will be the functional characterization of the novel genes that will be found and their relevance for MR. For this purpose working on the different aspects of these genes is necessary. This can provide important insights into the molecular basis of brain function and broaden the basis to identify compounds that can prevent the development of symptoms. In the attempt to elucidate the functions of these genes employing the advancements in the field of stem cell re-programming can be very useful.

But the most ultimate goal will be the treatment of genetic disease. The objective with most of the present treatments is not to correct the error in the DNA, but rather to prevent the development of the symptoms. The achievement of generating induced pluripotent stem cells has recently opened new doors of hope to look for the treatments based on the first approach.

With respect to Microcephaly which is one of the most prevalent conditions accompanied by MR, studying large cohorts and big families are necessary to find more genes. Further characterization of the cellular mechanisms behind them can help a lot in understanding the structure and function of the brain. Specifically in case of *MCPH1* developing an animal model and generating functional antibodies seems to be essential. Finding interacting partners of MCPH1 by other methods such as yeast two hybrid system can be very helpful to know more about the pathways in which it is involved.

5 References:

10K GeneChip® Mapping Assay Manual. ©2003-2004 Affymetrix, Inc. Doc. No: 701441 Rev. 2.

BeadStudio User Guide. Data analysis software for use with Illumina gene expression products, Doc. No. 11179632 Rev. B.

FASTLINK Home Page. http://www.cs.rice.edu/~schaffer/fastlink.html.

Illumina Technical Bulletin. http://www.illumina.com/General/pdf/Whole%20GenomeExpressionTechnicalBulletin.pdf.

Prioritizer home page. http://humgen.med.uu.nl/~lude/prioritizer/download.php.

Rockefeller Genetic Analysis Software Homepage. http://linkage.rockefeller.edu/soft/#e.

AAMR. 2005. Definition of mental retardation. American Association on Mental Retardation.

Abecasis GR, Cherny SS, Cookson WO, Cardon LR. 2001. GRR: graphical representation of relationship errors. Bioinformatics 17(8):742-3.

Abecasis GR, Cherny SS, Cookson WO, Cardon LR. 2002. Merlin--rapid analysis of dense genetic maps using sparse gene flow trees. Nat Genet 30(1):97-101.

Aiba A, Kano M, Chen C, Stanton ME, Fox GD, Herrup K, Zwingman TA, Tonegawa S. 1994. Deficient cerebellar long-term depression and impaired motor learning in mGluR1 mutant mice. Cell 79(2):377-88.

al-Ansari A. 1993. Etiology of mild mental retardation among Bahraini children: a community-based case control study. Ment Retard 31(3):140-3.

Alderton GK, Galbiati L, Griffith E, Surinya KH, Neitzel H, Jackson AP, Jeggo PA, O'Driscoll M. 2006. Regulation of mitotic entry by microcephalin and its overlap with ATR signalling. Nat Cell Biol 8(7):725-33.

Bartley JA, Hall BD. 1978. Mental retardation and multiple congenital anomalies of unknown etiology: frequency of occurrence in similarly affected sibs of the proband. Birth Defects Orig Artic Ser 14(6B):127-37.

Basel-Vanagaite L, Attia R, Yahav M, Ferland RJ, Anteki L, Walsh CA, Olender T, Straussberg R, Magal N, Taub E and others. 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(3):203-10.

Beitz E. 2000. TEXshade: shading and labeling of multiple sequence alignments using LATEX2 epsilon. Bioinformatics 16(2):135-9.

Bittles A. 2001. Consanguinity and its relevance to clinical genetics. Clin Genet 60(2):89-98.

Bittles AH, Neel JV. 1994. The costs of human inbreeding and their implications for variations at the DNA level. Nat Genet 8(2):117-21.

Bond J, Roberts E, Mochida GH, Hampshire DJ, Scott S, Askham JM, Springell K, Mahadevan M, Crow YJ, Markham AF and others. 2002. ASPM is a major determinant of cerebral cortical size. Nat Genet 32(2):316-20.

Bond J, Roberts E, Springell K, Lizarraga SB, Scott S, Higgins J, Hampshire DJ, Morrison EE, Leal GF, Silva EO and others. 2005. A centrosomal mechanism involving CDK5RAP2 and CENPJ controls brain size. Nat Genet 37(4):353-5.

Bond J, Woods CG. 2006. Cytoskeletal genes regulating brain size. Curr Opin Cell Biol 18(1):95-101.

Boutros R, Dozier C, Ducommun B. 2006. The when and wheres of CDC25 phosphatases. Curr Opin Cell Biol 18(2):185-91.

Bova GS, MacGrogan D, Levy A, Pin SS, Bookstein R, Isaacs WB. 1996. Physical mapping of chromosome 8p22 markers and their homozygous deletion in a metastatic prostate cancer. Genomics 35(1):46-54.

Brezun JM, Daszuta A. 1999. Depletion in serotonin decreases neurogenesis in the dentate gyrus and the subventricular zone of adult rats. Neuroscience 89(4):999-1002.

Britannica Oeo. 2003. http://original.britannica.com/eb/article-13355/human-intelligence.

Bundey S, Alam H, Kaur A, Mir S, Lancashire R. 1991. Why do UK-born Pakistani babies have high perinatal and neonatal mortality rates? Paediatr Perinat Epidemiol 5(1):101-14.

Cavalier L, Ouahchi K, Kayden HJ, Di Donato S, Reutenauer L, Mandel JL, Koenig M. 1998. Ataxia with isolated vitamin E deficiency: heterogeneity of mutations and phenotypic variability in a large number of families. Am J Hum Genet 62(2):301-10.

Chace DH. 2003. Mass spectrometry-based diagnostics: the upcoming revolution in disease detection has already arrived. Clin Chem 49(7):1227-8; author reply 1228-9.

Chelly J, Khelfaoui M, Francis F, Cherif B, Bienvenu T. 2006. Genetics and pathophysiology of mental retardation. Eur J Hum Genet 14(6):701-13.

Chiurazzi P, Schwartz CE, Gecz J, Neri G. 2008. XLMR genes: update 2007. Eur J Hum Genet.

Chiurazzi P, Tabolacci E, Neri G. 2004. X-linked mental retardation (XLMR): from clinical conditions to cloned genes. Crit Rev Clin Lab Sci 41(2):117-58.

Cho JH, Chang CJ, Chen CY, Tang TK. 2006. Depletion of CPAP by RNAi disrupts centrosome integrity and induces multipolar spindles. Biochem Biophys Res Commun 339(3):742-7.

Collins FS, Guyer MS, Charkravarti A. 1997. Variations on a theme: cataloging human DNA sequence variation. Science 278(5343):1580-1.

Cox J, Jackson AP, Bond J, Woods CG. 2006. What primary microcephaly can tell us about brain growth. Trends Mol Med 12(8):358-66.

Davis S, Sobel E, Marinov M, Weeks DE. 1997. Analysis of bipolar disorder using affected relatives. Genet Epidemiol 14(6):605-10.

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

DeMarzo AM, Nelson WG, Isaacs WB, Epstein JI. 2003. Pathological and molecular aspects of prostate cancer. Lancet 361(9361):955-64.

Didelot G, Molinari F, Tchenio P, Comas D, Milhiet E, Munnich A, Colleaux L, Preat T. 2006. Tequila, a neurotrypsin ortholog, regulates long-term memory formation in Drosophila. Science 313(5788):851-3.

Durkin MS, Hasan ZM, Hasan KZ. 1998. Prevalence and correlates of mental retardation among children in Karachi, Pakistan. Am J Epidemiol 147(3):281-8.

Erdogan F, Chen W, Kirchhoff M, Kalscheuer VM, Hultschig C, Muller I, Schulz R, Menzel C, Bryndorf T, Ropers HH and others. 2006. Impact of low copy repeats on the generation of balanced and unbalanced chromosomal aberrations in mental retardation. Cytogenet Genome Res 115(3-4):247-53.

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

Fernell E. 1998. Aetiological factors and prevalence of severe mental retardation in children in a Swedish municipality: the possible role of consanguinity. Dev Med Child Neurol 40(9):608-11.

Garshasbi M, Hadavi V, Habibi H, Kahrizi K, Kariminejad R, Behjati F, Tzschach A, Najmabadi H, Ropers HH, Kuss AW. 2008. A defect in the TUSC3 gene is associated with autosomal recessive mental retardation. Am J Hum Genet 82(5):1158-64.

Genoux D, Haditsch U, Knobloch M, Michalon A, Storm D, Mansuy IM. 2002. Protein phosphatase 1 is a molecular constraint on learning and memory. Nature 418(6901):970-5.

Gibbs JR, Singleton A. 2006. Application of genome-wide single nucleotide polymorphism typing: simple association and beyond. PLoS Genet 2(10):e150.

Gordon N. 2007. Sjogren-Larsson syndrome. Dev Med Child Neurol 49(2):152-4.

Gotz M, Huttner WB. 2005. The cell biology of neurogenesis. Nat Rev Mol Cell Biol 6(10):777-88.

Gudbjartsson DF, Jonasson K, Frigge ML, Kong A. 2000. Allegro, a new computer program for multipoint linkage analysis. Nat Genet 25(1):12-3.

Hammond RS, Bond CT, Strassmaier T, Ngo-Anh TJ, Adelman JP, Maylie J, Stackman RW. 2006. Small-conductance Ca2+-activated K+ channel type 2 (SK2) modulates hippocampal learning, memory, and synaptic plasticity. J Neurosci 26(6):1844-53.

Higgins JJ, Pucilowska J, Lombardi RQ, Rooney JP. 2004. A mutation in a novel ATP-dependent Lon protease gene in a kindred with mild mental retardation. Neurology 63(10):1927-31.

Hirota J, Ando H, Hamada K, Mikoshiba K. 2003. Carbonic anhydrase-related protein is a novel binding protein for inositol 1,4,5-trisphosphate receptor type 1. Biochem J 372(Pt 2):435-41.

Hoffmann K, Lindner TH. 2005. easyLINKAGE-Plus--automated linkage analyses using large-scale SNP data. Bioinformatics 21(17):3565-7.

Humphrey N, Mundlos S, Turkmen S. 2008. Genes and quadrupedal locomotion in humans. Proc Natl Acad Sci U S A 105(21):E26.

Hung LY, Chen HL, Chang CW, Li BR, Tang TK. 2004. Identification of a novel microtubule-destabilizing motif in CPAP that binds to tubulin heterodimers and inhibits microtubule assembly. Mol Biol Cell 15(6):2697-706.

Huttner WB, Kosodo Y. 2005. Symmetric versus asymmetric cell division during neurogenesis in the developing vertebrate central nervous system. Curr Opin Cell Biol 17(6):648-57.

Inlow JK, Restifo LL. 2004. Molecular and comparative genetics of mental retardation. Genetics 166(2):835-81.

Jackson AP, Eastwood H, Bell SM, Adu J, Toomes C, Carr IM, Roberts E, Hampshire DJ, Crow YJ, Mighell AJ and others. 2002. Identification of microcephalin, a protein implicated in determining the size of the human brain. Am J Hum Genet 71(1):136-42.

Jackson AP, McHale DP, Campbell DA, Jafri H, Rashid Y, Mannan J, Karbani G, Corry P, Levene MI, Mueller RF and others. 1998. Primary autosomal recessive microcephaly (MCPH1) maps to chromosome 8p22-pter. Am J Hum Genet 63(2):541-6.

Jaeken J, Matthijs G. 2007. Congenital disorders of glycosylation: a rapidly expanding disease family. Annu Rev Genomics Hum Genet 8:261-78.

Jamieson CR, Fryns JP, Jacobs J, Matthijs G, Abramowicz MJ. 2000. Primary autosomal recessive microcephaly: MCPH5 maps to 1q25-q32. Am J Hum Genet 67(6):1575-7.

Jiao Y, Yan J, Zhao Y, Donahue LR, Beamer WG, Li X, Roe BA, Ledoux MS, Gu W. 2005. Carbonic anhydrase-related protein VIII deficiency is associated with a distinctive lifelong gait disorder in waddles mice. Genetics 171(3):1239-46.

Jo S, Lee KH, Song S, Jung YK, Park CS. 2005. Identification and functional characterization of cereblon as a binding protein for large-conductance calcium-activated potassium channel in rat brain. J Neurochem 94(5):1212-24.

Joubert M, Eisenring JJ, Robb JP, Andermann F. 1999. Familial agenesis of the cerebellar vermis: a syndrome of episodic hyperpnea, abnormal eye movements, ataxia, and retardation. 1969. J Child Neurol 14(9):554-64.

Kahrizi K, Hossein Najmabadi, Roxana Kariminejad, Payman Jamali, Mahdi, Malekpour MG, H.-Hilger Ropers, Andreas W. Kuss, Andreas, Tzschach A. 2008. An autosomal recessive syndrome of severe mental retardation, cataract, coloboma and kyphosis maps to the pericentromeric region of chromosome 4. Eur J Hum Genet in press.

Kanehisa M, Goto S, Kawashima S, Okuno Y, Hattori M. 2004. The KEGG resource for deciphering the genome. Nucleic Acids Res 32(Database issue):D277-80.

Kelleher DJ, Gilmore R. 1994. The Saccharomyces cerevisiae oligosaccharyltransferase is a protein complex composed of Wbp1p, Swp1p, and four additional polypeptides. J Biol Chem 269(17):12908-17.

Kelleher DJ, Gilmore R. 2006. An evolving view of the eukaryotic oligosaccharyltransferase. Glycobiology 16(4):47R-62R.

Kelleher DJ, Karaoglu D, Mandon EC, Gilmore R. 2003. Oligosaccharyltransferase isoforms that contain different catalytic STT3 subunits have distinct enzymatic properties. Mol Cell 12(1):101-11.

Kelly C, Nogradi A, Walker R, Caddy K, Peters J, Carter N. 1994. Lurching, reeling, waddling and staggering in mice--is carbonic anhydrase (CA) VIII a candidate gene? Biochem Soc Trans 22(3):359S.

Kennedy GC, Matsuzaki H, Dong S, Liu WM, Huang J, Liu G, Su X, Cao M, Chen W, Zhang J and others. 2003. Large-scale genotyping of complex DNA. Nat Biotechnol 21(10):1233-7.

Kouprina N, Pavlicek A, Collins NK, Nakano M, Noskov VN, Ohzeki J, Mochida GH, Risinger JI, Goldsmith P, Gunsior M and others. 2005. The microcephaly ASPM gene is expressed in proliferating tissues and encodes for a mitotic spindle protein. Hum Mol Genet 14(15):2155-65.

Kruglyak L, Daly MJ, Lander ES. 1995. Rapid multipoint linkage analysis of recessive traits in nuclear families, including homozygosity mapping. Am J Hum Genet 56(2):519-27.

Kruglyak L, Daly MJ, Reeve-Daly MP, Lander ES. 1996. Parametric and nonparametric linkage analysis: a unified multipoint approach. Am J Hum Genet 58(6):1347-63.

Kulkarni ML, Kurian M. 1990. Consanguinity and its effect on fetal growth and development: a south Indian study. J Med Genet 27(6):348-52.

Kumar A, Blanton SH, Babu M, Markandaya M, Girimaji SC. 2004. Genetic analysis of primary microcephaly in Indian families: novel ASPM mutations. Clin Genet 66(4):341-8.

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

Leal GF, Roberts E, Silva EO, Costa SM, Hampshire DJ, Woods CG. 2003. A novel locus for autosomal recessive primary microcephaly (MCPH6) maps to 13q12.2. J Med Genet 40(7):540-2.

Lehner B, Fraser AG. 2004. A first-draft human protein-interaction map. Genome Biol 5(9):R63.

Leidel S, Gonczy P. 2005. Centrosome duplication and nematodes: recent insights from an old relationship. Dev Cell 9(3):317-25.

Li K, Kaufman TC. 1996. The homeotic target gene centrosomin encodes an essential centrosomal component. Cell 85(4):585-96.

Lin SY, Elledge SJ. 2003. Multiple tumor suppressor pathways negatively regulate telomerase. Cell 113(7):881-9.

Lin SY, Rai R, Li K, Xu ZX, Elledge SJ. 2005. BRIT1/MCPH1 is a DNA damage responsive protein that regulates the Brca1-Chk1 pathway, implicating checkpoint dysfunction in microcephaly. Proc Natl Acad Sci U S A 102(42):15105-15109.

Losada A, Yokochi T, Hirano T. 2005. Functional contribution of Pds5 to cohesin-mediated cohesion in human cells and Xenopus egg extracts. J Cell Sci 118(Pt 10):2133-41.

MacGrogan D, Levy A, Bova GS, Isaacs WB, Bookstein R. 1996. Structure and methylation-associated silencing of a gene within a homozygously deleted region of human chromosome band 8p22. Genomics 35(1):55-65.

Magnus P, Berg K, Bjerkedal T. 1985. Association of parental consanguinity with decreased birth weight and increased rate of early death and congenital malformations. Clin Genet 28(4):335-42.

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

Manual GMKA. 2005–2006 Affymetrix Inc. P/N 701930 Rev. 3.

Matise TC, Sachidanandam R, Clark AG, Kruglyak L, Wijsman E, Kakol J, Buyske S, Chui B, Cohen P, de Toma C and others. 2003. A 3.9-centimorgan-resolution human single-nucleotide polymorphism linkage map and screening set. Am J Hum Genet 73(2):271-84.

Matsuzaki H, Loi H, Dong S, Tsai YY, Fang J, Law J, Di X, Liu WM, Yang G, Liu G and others. 2004. Parallel genotyping of over 10,000 SNPs using a one-primer assay on a high-density oligonucleotide array. Genome Res 14(3):414-25.

McCreary BD, Rossiter JP, Robertson DM. 1996. Recessive (true) microcephaly: a case report with neuropathological observations. J Intellect Disabil Res 40 (Pt 1):66-70.

Miller BJ, Wang D, Krahe R, Wright FA. 2003. Pooled analysis of loss of heterozygosity in breast cancer: a genome scan provides comparative evidence for multiple tumor suppressors and identifies novel candidate regions. Am J Hum Genet 73(4):748-67.

Miller SA, Dykes DD, Polesky HF. 1988. A simple salting out procedure for extracting DNA from human nucleated cells. Nucleic Acids Res 16(3):1215.

Mochida GH, Walsh CA. 2001. Molecular genetics of human microcephaly. Curr Opin Neurol 14(2):151-6.

Moheb LA, Tzschach A, Garshasbi M, Kahrizi K, Darvish H, Heshmati Y, Kordi A, Najmabadi H, Ropers HH, Kuss AW. 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(2):270-3.

Molinari F, Foulquier F, Tarpey PS, Morelle W, Boissel S, Teague J, Edkins S, Futreal PA, Stratton MR, Turner G and others. 2008. Oligosaccharyltransferase-subunit mutations in nonsyndromic mental retardation. Am J Hum Genet 82(5):1150-7.

Molinari F, Rio M, Meskenaite V, Encha-Razavi F, Auge J, Bacq D, Briault S, Vekemans M, Munnich A, Attie-Bitach T and others. 2002. Truncating neurotrypsin mutation in autosomal recessive nonsyndromic mental retardation. Science 298(5599):1779-81.

Morton NE. 1955. Sequential tests for the detection of linkage. Am J Hum Genet 7(3):277-318.

Motazacker MM, Rost BR, Hucho T, Garshasbi M, Kahrizi K, Ullmann R, Abedini SS, Nieh SE, Amini SH, Goswami C and others. 2007. A defect in the ionotropic glutamate receptor 6 gene (GRIK2) is associated with autosomal recessive mental retardation. Am J Hum Genet 81(4):792-8.

Moynihan L, Jackson AP, Roberts E, Karbani G, Lewis I, Corry P, Turner G, Mueller RF, Lench NJ, Woods CG. 2000. A third novel locus for primary autosomal recessive microcephaly maps to chromosome 9q34. Am J Hum Genet 66(2):724-7.

Mulder NJ, Apweiler R, Attwood TK, Bairoch A, Barrell D, Bateman A, Binns D, Biswas M, Bradley P, Bork P and others. 2003. The InterPro Database, 2003 brings increased coverage and new features. Nucleic Acids Res 31(1):315-8.

Munton RP, Vizi S, Mansuy IM. 2004. The role of protein phosphatase-1 in the modulation of synaptic and structural plasticity. FEBS Lett 567(1):121-8.

Najmabadi H, Motazacker MM, Garshasbi M, Kahrizi K, Tzschach A, Chen W, Behjati F, Hadavi V, Nieh SE, Abedini SS and others. 2006. Homozygosity mapping in consanguineous families reveals extreme heterogeneity of non-syndromic autosomal recessive mental retardation and identifies 8 novel gene loci. Hum Genet.

Nannya Y, Sanada M, Nakazaki K, Hosoya N, Wang L, Hangaishi A, Kurokawa M, Chiba S, Bailey DK, Kennedy GC and others. 2005. A robust algorithm for copy number detection using high-density oligonucleotide single nucleotide polymorphism genotyping arrays. Cancer Res 65(14):6071-9.

Neitzel H, Neumann LM, Schindler D, Wirges A, Tonnies H, Trimborn M, Krebsova A, Richter R, Sperling K. 2002. Premature chromosome condensation in humans associated with microcephaly and mental retardation: a novel autosomal recessive condition. Am J Hum Genet 70(4):1015-22.

O'Connell JR, Weeks DE. 1998. PedCheck: a program for identification of genotype incompatibilities in linkage analysis. Am J Hum Genet 63(1):259-66.

O'Driscoll M, Jackson AP, Jeggo PA. 2006. Microcephalin: a causal link between impaired damage response signalling and microcephaly. Cell Cycle 5(20):2339-44.

Ou XM, Lemonde S, Jafar-Nejad H, Bown CD, Goto A, Rogaeva A, Albert PR. 2003. Freud-1: A neuronal calcium-regulated repressor of the 5-HT1A receptor gene. J Neurosci 23(19):7415-25.

Ozcelik T, Akarsu N, Uz E, Caglayan S, Gulsuner S, Onat OE, Tan M, Tan U. 2008. Mutations in the very low-density lipoprotein receptor VLDLR cause cerebellar hypoplasia and quadrupedal locomotion in humans. Proc Natl Acad Sci U S A 105(11):4232-6.

Pattison L, Crow YJ, Deeble VJ, Jackson AP, Jafri H, Rashid Y, Roberts E, Woods CG. 2000. A fifth locus for primary autosomal recessive microcephaly maps to chromosome 1q31. Am J Hum Genet 67(6):1578-80.

Ponting CP. 2006. A novel domain suggests a ciliary function for ASPM, a brain size determining gene. Bioinformatics 22(9):1031-5.

Pribill I, Speiser P, Leary J, Leodolter S, Hacker NF, Friedlander ML, Birnbaum D, Zeillinger R, Krainer M. 2001. High frequency of allelic imbalance at regions of chromosome arm 8p in ovarian carcinoma. Cancer Genet Cytogenet 129(1):23-9.

Rai R, Dai H, Multani AS, Li K, Chin K, Gray J, Lahad JP, Liang J, Mills GB, Meric-Bernstam F and others. 2006. BRIT1 regulates early DNA damage response, chromosomal integrity, and cancer. Cancer Cell 10(2):145-57.

Raymond FL, Tarpey P. 2006. The genetics of mental retardation. Hum Mol Genet 15 Spec No 2:R110-6.

Rhoads A, Kenguele H. 2005. Expression of IQ-motif genes in human cells and ASPM domain structure. Ethn Dis 15(4 Suppl 5):S5-88-91.

Rizzo WB. 2007. Sjogren-Larsson syndrome: molecular genetics and biochemical pathogenesis of fatty aldehyde dehydrogenase deficiency. Mol Genet Metab 90(1):1-9.

Roberts E, Hampshire DJ, Pattison L, Springell K, Jafri H, Corry P, Mannon J, Rashid Y, Crow Y, Bond J and others. 2002. Autosomal recessive primary microcephaly: an analysis of locus heterogeneity and phenotypic variation. J Med Genet 39(10):718-21.

Roberts E, Jackson AP, Carradice AC, Deeble VJ, Mannan J, Rashid Y, Jafri H, McHale DP, Markham AF, Lench NJ and others. 1999. The second locus for autosomal recessive primary microcephaly (MCPH2) maps to chromosome 19q13.1-13.2. Eur J Hum Genet 7(7):815-20.

Robertson G, Bilenky M, Lin K, He A, Yuen W, Dagpinar M, Varhol R, Teague K, Griffith OL, Zhang X and others. 2006. cisRED: a database system for genome-scale computational discovery of regulatory elements. Nucleic Acids Res 34(Database issue):D68-73.

Roeleveld N, Zielhuis GA, Gabreels F. 1997. The prevalence of mental retardation: a critical review of recent literature. Dev Med Child Neurol 39(2):125-32.

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

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

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

Ropers HH, Hoeltzenbein M, Kalscheuer V, Yntema H, Hamel B, Fryns JP, Chelly J, Partington M, Gecz J, Moraine C. 2003. Nonsyndromic X-linked mental retardation: where are the missing mutations? Trends Genet 19(6):316-20.

Rotanova TV, Botos I, Melnikov EE, Rasulova F, Gustchina A, Maurizi MR, WLODawer A. 2006. Slicing a protease: structural features of the ATP-dependent Lon proteases gleaned from investigations of isolated domains. Protein Sci 15(8):1815-28.

Rual JF, Venkatesan K, Hao T, Hirozane-Kishikawa T, Dricot A, Li N, Berriz GF, Gibbons FD, Dreze M, Ayivi-Guedehoussou N and others. 2005. Towards a proteome-scale map of the human protein-protein interaction network. Nature 437(7062):1173-8.

Ruschendorf F, Nurnberg P. 2005. ALOHOMORA: a tool for linkage analysis using 10K SNP array data. Bioinformatics 21(9):2123-5.

Seifert W, Holder-Espinasse M, Kuhnisch J, Kahrizi K, Tzschach A, Garshasbi M, Najmabadi H, Walter Kuss A, Kress W, Laureys G and others. 2008. Expanded mutational spectrum in cohen syndrome, tissue expression, and transcript variants of COH1. Hum Mutat.

Shaw-Smith C, Redon R, Rickman L, Rio M, Willatt L, Fiegler H, Firth H, Sanlaville D, Winter R, Colleaux L and others. 2004. Microarray based comparative genomic hybridisation (array-CGH) detects submicroscopic chromosomal deletions and duplications in patients with learning disability/mental retardation and dysmorphic features. J Med Genet 41(4):241-8.

Steemers FJ, Gunderson KL. 2005. Illumina, Inc. Pharmacogenomics 6(7):777-82.

Stelzl U, Worm U, Lalowski M, Haenig C, Brembeck FH, Goehler H, Stroedicke M, Zenkner M, Schoenherr A, Koeppen S and others. 2005. A human protein-protein interaction network: a resource for annotating the proteome. Cell 122(6):957-68.

Strachan T R, A. P. 2003. Human Molecular Genetics, 3rd ed. 674 p.

Temtamy SA, Kandil MR, Demerdash AM, Hassan WA, Meguid NA, Afifi HH. 1994. An epidemiological/genetic study of mental subnormality in Assiut Governorate, Egypt. Clin Genet 46(5):347-51.

Terada Y, Uetake Y, Kuriyama R. 2003. Interaction of Aurora-A and centrosomin at the microtubule-nucleating site in Drosophila and mammalian cells. J Cell Biol 162(5):757-63.

Thiele H, Nurnberg P. 2005. HaploPainter: a tool for drawing pedigrees with complex haplotypes. Bioinformatics 21(8):1730-2.

Thompson JD, Gibson TJ, Higgins DG. 2002. Multiple sequence alignment using ClustalW and ClustalX. Curr Protoc Bioinformatics Chapter 2:Unit 2 3.

Thor AD, Eng C, Devries S, Paterakos M, Watkin WG, Edgerton S, Moore DH, 2nd, Etzell J, Waldman FM. 2002. Invasive micropapillary carcinoma of the breast is associated with chromosome 8 abnormalities detected by comparative genomic hybridization. Hum Pathol 33(6):628-31.

Trimborn M, Bell SM, Felix C, Rashid Y, Jafri H, Griffiths PD, Neumann LM, Krebs A, Reis A, Sperling K and others. 2004. Mutations in microcephalin cause aberrant regulation of chromosome condensation. Am J Hum Genet 75(2):261-6.

Trimborn M, Richter R, Sternberg N, Gavvovidis I, Schindler D, Jackson AP, Prott EC, Sperling K, Gillessen-Kaesbach G, Neitzel H. 2005. The first missense alteration in the MCPH1 gene causes autosomal recessive microcephaly with an extremely mild cellular and clinical phenotype. Hum Mutat 26(5):496.

Trimborn M, Schindler D, Neitzel H, Hirano T. 2006. Misregulated chromosome condensation in MCPH1 primary microcephaly is mediated by condensin II. Cell Cycle 5(3):322-6.

Turkmen S, Hoffmann K, Demirhan O, Aruoba D, Humphrey N, Mundlos S. 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.

Türkmen S, Gue G, Garshasbi M, Hofmann K, Alshalah A, Kahrizi K, Tzschach A, Kuss AW, Najmabadi H, Ropers HH, Humphrey N, Mundlos S, Robinson PN. 2009. CA8 mutations cause a novel syndrome characterized by ataxia and mild mental retardation with predisposition to quadrupedal gait. 20. Jahrestagung der Deutschen Gesellschaft für Humangenetik, Eurogress Aachen, 1-3 Aprill 2009.

Tzschach A, Bozorgmehr B, Hadavi V, Kahrizi K, Garshasbi M, Motazacker MM, Ropers HH, Kuss AW, Najmabadi H. 2008. Alopecia-mental retardation syndrome: clinical and molecular characterization of four patients. Br J Dermatol.

Varon R, Vissinga C, Platzer M, Cerosaletti KM, Chrzanowska KH, Saar K, Beckmann G, Seemanova E, Cooper PR, Nowak NJ and others. 1998. Nibrin, a novel DNA double-strand break repair protein, is mutated in Nijmegen breakage syndrome. Cell 93(3):467-76.

Vissers LE, de Vries BB, Osoegawa K, Janssen IM, Feuth T, Choy CO, Straatman H, van der Vliet W, Huys EH, van Rijk A and others. 2003. Array-based comparative genomic hybridization for the genomewide detection of submicroscopic chromosomal abnormalities. Am J Hum Genet 73(6):1261-70.

WHO. 1980. International classification of impairments, disabilities and handicaps. (World Health Organization, Geneva).

WHO. The World Health Report 2001. mental health: new understanding, new hope.

Wilcken B. 2004. Screening of newborns for metabolic disorders with mass spectrometry. Jama 291(12):1444; author reply 1444-5.

Wingender E, Chen X, Hehl R, Karas H, Liebich I, Matys V, Meinhardt T, Pruss M, Reuter I, Schacherer F. 2000. TRANSFAC: an integrated system for gene expression regulation. Nucleic Acids Res 28(1):316-9.

Woods CG. 2004. Human microcephaly. Curr Opin Neurobiol 14(1):112-7.

Woods CG, Bond J, Enard W. 2005. Autosomal recessive primary microcephaly (MCPH): a review of clinical, molecular, and evolutionary findings. Am J Hum Genet 76(5):717-28.

Woods CG, Cox J, Springell K, Hampshire DJ, Mohamed MD, McKibbin M, Stern R, Raymond FL, Sandford R, Malik Sharif S and others. 2006. Quantification of homozygosity in consanguineous individuals with autosomal recessive disease. Am J Hum Genet 78(5):889-96.

Xu X, Lee J, Stern DF. 2004. Microcephalin is a DNA damage response protein involved in regulation of CHK1 and BRCA1. J Biol Chem 279(33):34091-4.

Yang SZ, Lin FT, Lin WC. 2008. MCPH1/BRIT1 cooperates with E2F1 in the activation of checkpoint, DNA repair and apoptosis. EMBO Rep 9(9):907-15.

Yaqoob M, Bashir A, Tareen K, Gustavson KH, Nazir R, Jalil F, von Dobeln U, Ferngren H. 1995. Severe mental retardation in 2 to 24-month-old children in Lahore, Pakistan: a prospective cohort study. Acta Paediatr 84(3):267-72.

Yasuno F, Suhara T, Nakayama T, Ichimiya T, Okubo Y, Takano A, Ando T, Inoue M, Maeda J, Suzuki K. 2003. Inhibitory effect of hippocampal 5-HT1A receptors on human explicit memory. Am J Psychiatry 160(2):334-40.

Yunis K, Mumtaz G, Bitar F, Chamseddine F, Kassar M, Rashkidi J, Makhoul G, Tamim H. 2006. Consanguineous marriage and congenital heart defects: a case-control study in the neonatal period. Am J Med Genet A 140(14):1524-30.

6 Supplementary data

6.1 Appendix - A

Table S-1: Linkage intervals in all the families

	Family	Chromosome	Start	End	Length
1	**M003**	16	18065860	51653024	33587164
		3	111068414	119454521	8386107
		6	163829542	170914574	7085032
2	**M004**	11	86018126	86479274	461148
		19	16213403	57869570	41656167
		7	91537285	108240427	16703142
3	**M005**	3	68818296	73438043	4619747
		5	64886978	79484510	14597532
		12	11940929	20833416	8892487
		12	22992221	26583494	3591273
4	**M007L**	12	97754292	115798038	18043746
		4	187393384	191731959	4338575
		9	81420987	86939429	5518442
5	**M007R**	21	30537890	33352236	2814346
		10	12766482	34466736	21700254
		12	50938088	66484625	15546537
		14	86355943	96160652	9804709
		15	49274478	59139245	9864767
		6	166961336	170914576	3953240
		8	2390378	8584258	6193880
		8	22493085	64494827	42001742
6	**M008**	1	172689984	179153668	6463684
		18	43986452	61180205	17193753
		19	57313456	60695892	3382436
		7	1	7223707	7223706
7	**M010**	16	23004204	50484515	27480311
8	**M012L**	15	1	25005399	25005398
		17	32304466	55317193	23012727
		2	227188513	235834145	8645632
		21	43201250	46976097	3774847
		3	6482290	27002462	20520172
		9	21525576	29694495	8168919
		6	158396429	166836243	8439814
9	**M012R**	1	182773283	187355001	4581718
		19	1	8829549	8829548
		17	32304466	55317193	23012727
		17	72357083	77773470	5416387
		2	117422094	133953217	16531123
		10	53115688	62408160	9292472
		12	127935939	132078379	4142440
10	**M017**	1	7145191	13459518	6314327
		1	171092175	177327853	6235678
11	**M019**	8	3146756	14134135	10987379
12	**M021**		???		
13	**M025**	19	46160099	52292281	6132182
14	**M056**	1	19417982	22380638	2962656
		12	97019296	99614649	2595353
		15	45591329	57955770	12364441
		7	128444650	138136267	9691617
15	**M064**	1	190155586	206197802	16042216
		2	1	7125905	7125904
		1	234461830	242481793	8019963
16	**M069**	4	47273225	57709293	10436068
17	**M088**	1	20062430	22380638	2318208
		2	125249479	144892787	19643308
18	**M103**	2	23314413	23685963	371550
		8	105665819	118153391	12487572
19	**M107**	21	21748820	36117298	14368478
		8	59645979	63667057	4021078
20	**M108**	10	121667176	130702443	9035267
		17	65153675	75236653	10082978

Table S-1: Linkage intervals in all the families

	Family	Chromosome	Start	End	Length
		21	1	33346030	33346029
		3	53415287	133199938	79784651
		3	154610302	186246297	31635995
21	**M110**	10	26096564	43053450	16956886
	(2013)	12	21587716	23051233	1463517
		16	83272419	86086476	2814057
		22	42878000	48760841	5882841
		3	150835697	174510186	23674489
		6	44381505	88713362	44331857
22	**M110**	1	151893200	159272524	7379324
	(2017)	1	173572009	206197802	32625793
		2	10162390	34192114	24029724
		2	34239768	34517685	277917
		2	54739686	69722351	14982665
		2	123161810	133952983	10791173
		5	13931581	23080796	9149215
		5	29145855	33101281	3955426
		5	33818197	39236795	5418598
23	**M123**	20	57277661	58901435	1623774
		21	26321270	28060803	1739533
24	**M142**	12	108357888	117476805	9118917
		2	23685963	44895495	21209532
25	**M144**	12	115339048	130375660	15036612
		4	11662043	13704596	2042553
26	**M146**	5	54982253	56781260	1799007
		6	129585626	140272407	10686781
		19	53558026	61011337	7453311
27	**M147**	7	41305773	43886690	2580917
		7	47002343	78932787	31930444
		8	70774872	94352595	23577723
		12	125274270	130375660	5101390
		17	43472314	52787908	9315594
		20	11892445	49145368	37252923
28	**M149**	1	1	12221853	12221852
		1	113087863	149380131	36292268
		2	228043462	235834145	7790683
29	**M150**	8	23920355	52857562	28937207
		8	98238495	105665819	7427324
		8	120554005	124668763	4114758
		12	12721181	19879644	7158463
		18	20840916	46526392	25685476
30	**M150N**	4	188553082	191411218	2858136
	M152	1	186272652	201087653	14815001
		2	229842774	236010878	6168104
		5	119023683	124544737	5521054
		8	77870512	117872603	40002091
31	**M153**	1	238330027	243931905	5601878
		2	12078062	17194641	5116579
		16	8495208	12634228	4139020
32	**M154**	6	32773006	74604508	41831502
		15	85324892	92788371	7463479
		16	53911502	61359442	7447940
33	**M156**	9	114628322	114892989	264667
		12	116700000	124400000	7700000
		17	60350346	68847559	8497213
34	**M157**	12	86199128	126428581	40229453
35	**M159**	14	26578858	31700826	6201680
36	**M163**	5	1	5019633	5019632
37	**M164**	1	178625450	182046893	3421443
		10	52593593	72056306	19462713
		12	106753295	126779462	20026167
		15	94248752	99604449	5355697
		17	1	6758576	6758575
		21	43914978	46813969	2898991
		4	91430139	96547452	5117313
		6	38821276	86977719	48156443
38	**M165**	15	69405826	85324892	15919066
		1	51946762	56475802	4529040
39	**M169**	1	19868655	34462324	14593669
		4	74135822	76225029	2089207

Table S-1: Linkage intervals in all the families

	Family	Chromosome	Start	End	Length
40	**M177**	3	24508243	25207120	698877
		8	102932948	122096140	19163192
		2	157765045	158094413	329368
41	**M183**	16	50658865	54881532	4222667
		18	4326748	10142858	5816110
		18	53981883	66044576	12062693
42	**M190**	8	60684518	68699546	8015028
		19	33155427	37741783	4586356
43	**M198**	3	62049014	69708151	7659137
		11	62812227	79708448	16896221
44	**M225**	13	83098651	87474013	4375362
		19	3542840	20152513	16609673
45	**M226**	4	25489589	64668181	39178592
		6	45618052	108415602	62797550
46	**M233**	14	86797152	91363521	4566369
47	**M235**	14	58334879	79252844	20917715
48	**M239**	18	61129618	63347340	2217722
		18	64954930	65507016	552086
		18	29857021	31201178	1344157
		11	22023838	24533759	2509921
		9	75365803	77132893	1767090
49	**M248**	17	45542883	61966895	16424012
		13	29247237	35129488	5882251
		9	77658489	81456712	3798223
		5	16531200	77818845	61287645
50	**M249**	6	17455552	39446663	21991111
51	**M251**	1	118053780	156711071	38657291
		2	151020482	151317011	296529
		10	30105182	31953821	1848639
		10	34532194	35681577	1149383
		18	72210260	75586155	3375895
52	**M252**	3	41586232	73194663	31608431
		3	112655821	114916790	2260969
		3	115786843	120801097	5014254
		6	116042452	116781531	739079
		15	95944662	100338915	4394253
		18	25613138	26407442	794304
53	**M254**	3	151215520	152450553	1235033
		3	161568749	163337023	1768274
		8	98916138	99304253	388115
		8	104015036	106059630	2044594
		9	124428147	128740358	4312211
		15	47022543	60591780	13569237
54	**M261**	5	61327157	120723042	59395885
55	**M263**	10	93716070	105835217	12119147
		10	109803462	110188074	384612
		2	80226413	80538740	312327
		2	81893624	83211736	1318112
		6	115701966	116244187	542221
		14	26493657	26682620	188963
		18	4182281	4477509	295228
56	**M269**	13	30483928	43168894	12684966
		3	125349784	125966523	616739
57	**M289**	1	107511901	120495870	12983969
58	**M300**	19	9315423	16527107	7211684
59	**M302**	9	593192	4325668	3732476
60	**M304**	10	66796694	71360481	4563787
		10	86519759	87179215	659456
		7	39163879	41279517	2115638
		1	217721015	218481495	760480
61	**M305**	20	10961149	22854696	11893547
62	**M307**	6	51676830	52220925	544095
63	**M314**	5	4007570	10786776	6779206
64	**M318**	11	76728612	114226470	37497858
65	**M319**	1	38896190	48981205	10085015
66	**M323**	15	21847915	40160190	18312275
67	**M324**	4	122001354	130191055	8189701
68	**M331**	6	33586474	44421400	10834926
		1	109245821	114433503	5187682

Table S-1: Linkage intervals in all the families

	Family	Chromosome	Start	End	Length
69	**M346**	4	80058957	132225967	52167010
70	**8207040**	12	108357863	113987683	5629820
		13	96715213	100233193	3517980
		16	29272335	90041932	60769597
		4	89167664	102524025	13356361
		4	142129426	157810146	15680720
71	**8600042***	1	38907775	55579847	16672072
72	**8305358**	11	7362108	20044296	12682188
		15	90504391	100256656	9752265
73	**8307307**	2	45583739	52294689	6710950
		2	171472223	206829958	35357735
		6	0	7456900	7456900
		7	39891935	49039749	9147814
		16	83254156	86086476	2832320
74	**8307998**	4	104710169	116523708	11813539
		16	6103801	12781401	6677600
75	**8401214**	4	96959569	103310859	6351290
		5	52042975	53314739	1271764
		12	96708635	121619921	24911286
		18	73670156	76117153	2446997
76	**8401811**	7	12766030	21527471	8761441
		8	28202672	30917928	2715256
		10	94744240	108785365	14041125
		20	17077244	42801349	25724105
		13	75833297	78642739	2809442
		12	128752792	132449811	3697019
77	**8401973**	13	37071816	39040280	1968464
		16	6050548	8435410	2384862
78	**8303971**	17	3565047	8644145	5079098
79	**D54**	20	48696430	57020587	8324157
80	**8404553**	8	12944303	17579537	4635234
81	**8500031**	17	33901979	52662679	18760700
82	**8500032**	20	15943223	30419769	14476546
		13	100531649	101334508	802859
		6	106923556	108135726	1212170
		16	86128352	86790993	662641
83	**8500036**	11	123986930	125985566	1998636
		6	43495925	85699654	42203729
84	**8500058**	17	74205146	qter	32163440
		14	51114664	56467237	5352573
		11	132605443	133674016	1068573
85	**8500059**	18	58960763	63862253	4901490
		3	145877335	150092774	4215439
86	**8500061**	18	26834675	43812253	16977578
87	**8500064**	2	77117862	123280342	46162480
88	**8500156***	19	38339219	49668378	11329159
89	**8500157**	5	2203273	13822974	11619701
		7	78542314	80517630	1975316
90	**8500194**	2	159883240	168718543	8835303
		2	177073122	195575344	18502222
		2	211967419	222673543	10706124
		11	23496972	25797805	2300833
91	**8500234**	11	34412684	69338136	34925452
		1	167202400	175143539	7941139
92	**8500235**	1	21615324	37349801	15734477
		3	99698640	107415425	7716785
		4	123837191	127248513	3411322
		13	30176440	33290909	3114469
		22	46556372	49691432	3135060
93	**8500302**	1	114218960	120003821	5784861
		1	203951975	230048894	26096919
		4	146579862	156328628	9748766
		9	123918003	128318676	4400673
		21	1	19011109	19011108
94	**8500306**	2	106752670	108951212	2198542
		6	127363178	137713134	10349956
		6	143841086	147412758	3571672
		16	54304447	60636703	6332256
95	**8500313**	2	137146498	147854552	10708054

Table S-1: Linkage intervals in all the families

	Family	Chromosome	Start	End	Length
		14	1	21400897	21400896
96	**8500314**	8	105437584	106470472	1032888
97	**8500318**	4	116044453	117400685	1356232
		4	59713	2916279	2856566
98	**8500320**	1	38782583	39276465	493882
		1	30486063	34066888	3580825
		6	157853047	159881478	2028431
		11	12439144	14008775	1569631
		13	22536210	25966373	3430163
		17	76225231	78774742	2549511
99	**M257**	14	88718897	96927688	8208791
100	**8600057***	2	240751883	242951149	2199266
101	**M332**	6	125293073	136392392	11099319
		12	130653713	132349534	1695821
		7	131493578	132725020	1231442
		7	136210788	137652308	1441520
102	**M347**	3	1	1576470	1576469
		5	88364920	90386754	2021834
		1	161707727	163083580	1375853
103	**M203**	6	145021020	145385768	364748
		12	82336904	82740523	403619
104	**8600004***	17	17195117	28444843	11249726
105	**8600005**	2	155289884	156537450	1247566
		4	84797731	94741372	9943641
		4	96596018	97191397	595379
		4	142217482	156850941	14633459
		7	141274408	155427591	14153183
		12	4258345	5511618	1253273
106	**8600011**	2	105179382	105714131	534749
		5	121475977	145918691	24442714
107	**8600013**	6	119341282	120526827	1185545
		12	82689827	83389184	699357
		20	48630405	49791124	1160719
108	**8600041**	11	1	7184263	7184262
		3	126410481	127203704	793223
		5	71880756	72726216	845460
109	**8600043**	6	81735706	82279077	543371
		6	83644948	83739609	94661
		5	65968099	66291082	322983
		1	205568150	206301726	733576
110	**5600045_1**	19	6912182	34916304	28004122
		19	37813569	38922921	1109352
		1	48827467	57201177	8373710
111	**8600045_2**	3	59715459	60439116	723657
		11	19939756	23377994	3438238
112	**8600273**	10	3768250	17370186	13601936
		18	6628009	26201590	19573581
		19	58784434	59506905	722471
		21	25008485	29908100	4899615
		21	30073388	30970253	896865
		21	31675026	32393779	718753
		7	1	792020	792019
113	**8600277**	5	149119259	159250036	10130777
		5	58538099	60913396	2375297
		8	50451575	51651167	1199592
		3	185957520	187455785	1498265
		2	193055632	195053028	1997396
		2	195660744	196824022	1163278
		9	101497186	101707078	209892
114	**8600012**	6	25532985	44405750	18872765
		9	112925472	126780259	13854787
		10	63617334	80623399	17006065
		20	60819065	62435964	1616899
115	**8600059**	9	25953989	66266543	40312554
		9	21523705	21801783	278078
		7	103431429	103894652	463223
		7	110520703	111410982	890279
		7	125535796	126322656	786860
		3	3996579	4353453	356874
		4	118711450	119827074	1115624

Table S-1: Linkage intervals in all the families

	Family	Chromosome	Start	End	Length
		12	100686206	101510129	823923
		13	63500690	64062182	561492
		13	57132954	58143886	1010932
		21	45147909	45602169	454260
116	**8600074**	7	37947572	38695370	747798
		14	63447838	76688610	13240772
		15	23847489	24489581	642092
117	**8600086***	15	72919012	90871916	17952904
118	**8600162**	2	173272129	213293183	40021054
		14	19489520	22084152	2594632
		14	46498239	51151567	4653328
119	**8600058**	17	12280053	16741855	4461802
		17	9349119	9644200	295081
		17	10070677	10506938	436261
		11	88591430	89682437	1091007

6.2 Appendix - B

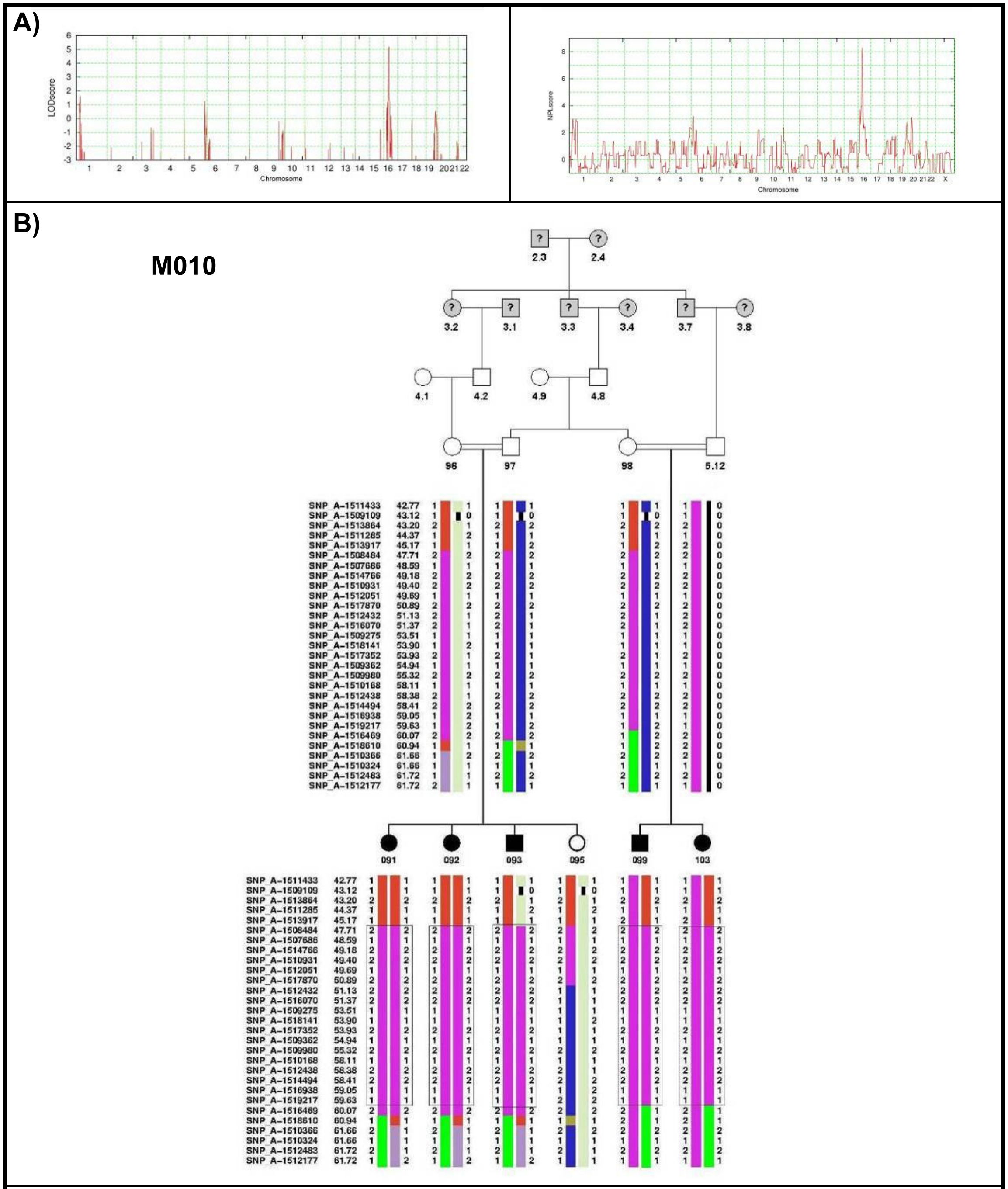


Fig S-1: Family M010. Whole genome parametric (Allegro) and non-parametric (GeneHunter) linkage results (A) and haplotype of the only linkage interval with significant parametric LOD score of 5.2 (B)

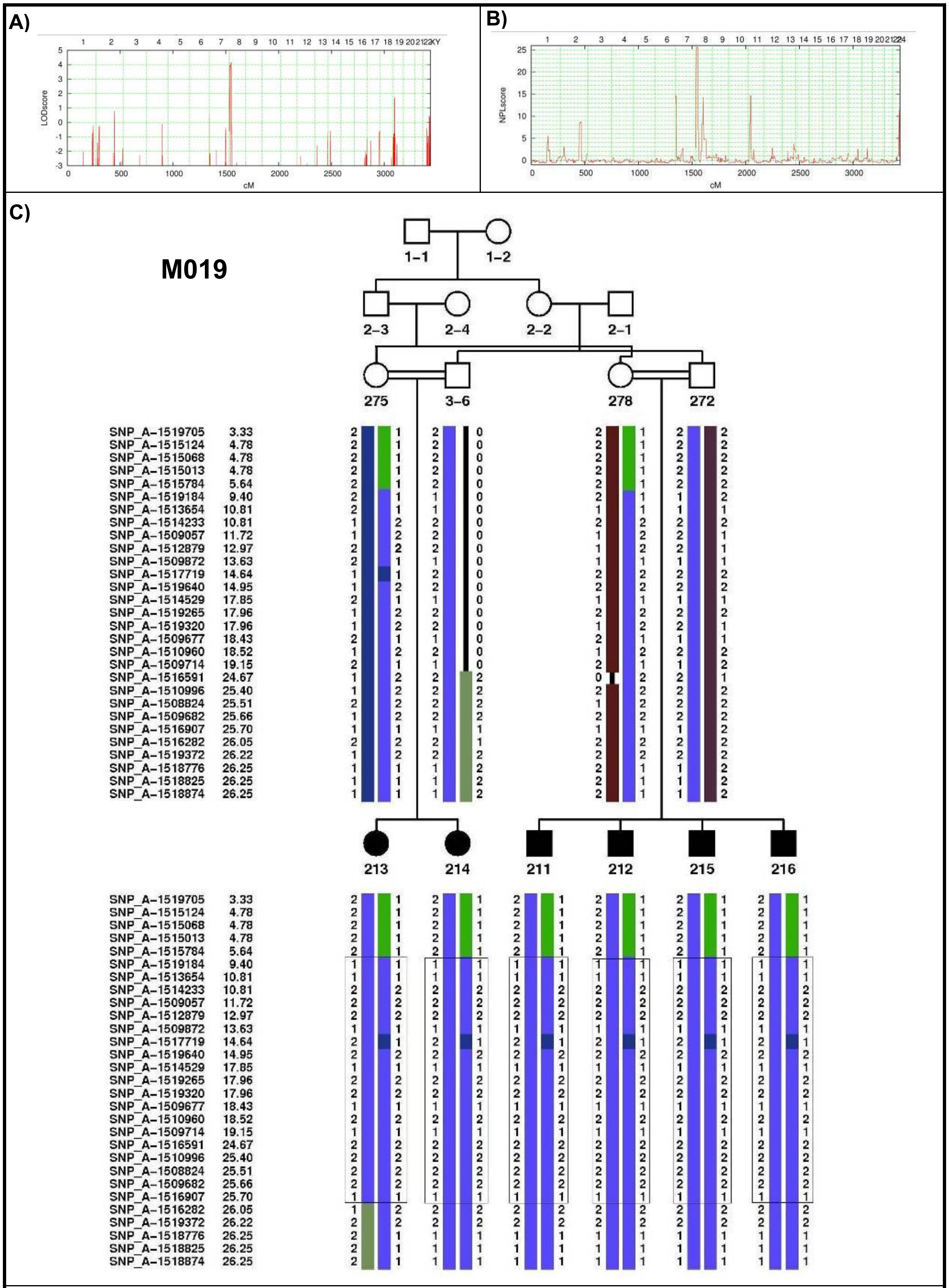


Fig S-2: Family M019 linkage results. Whole genome parametric (A) and non-parametric (B) linkage results (Allegro and GeneHunter) and haplotype of the only linkage interval on chromosome 8 with significant parametric LOD score of 4.2. All the markers are homozygous in this region except for SNP_A-1517719.

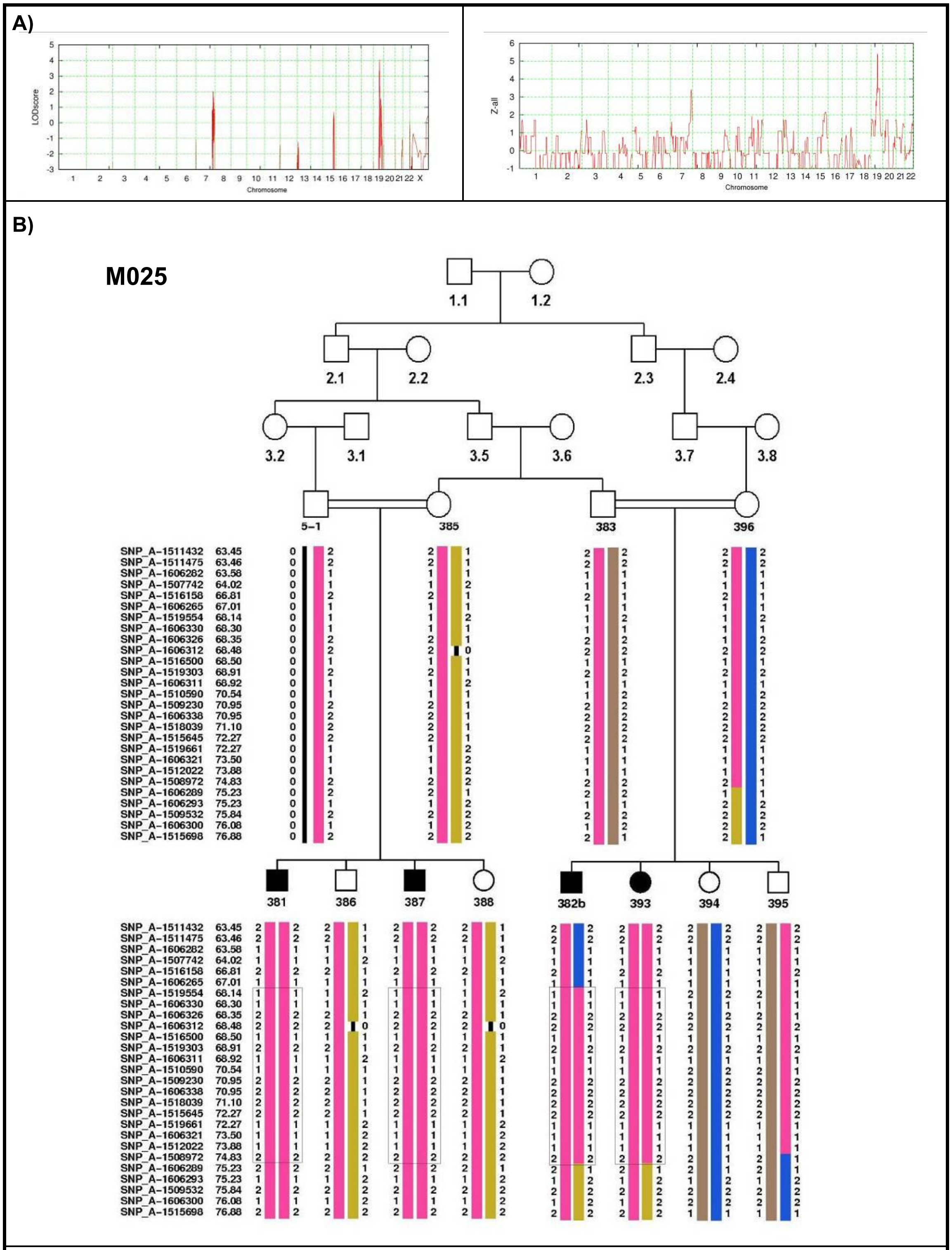


Fig S-3: Family M025. Whole genome parametric (Allegro) and non-parametric (GeneHunter) linkage results (A) and haplotype of the only linkage interval with significant parametric LOD score of 4 (B)

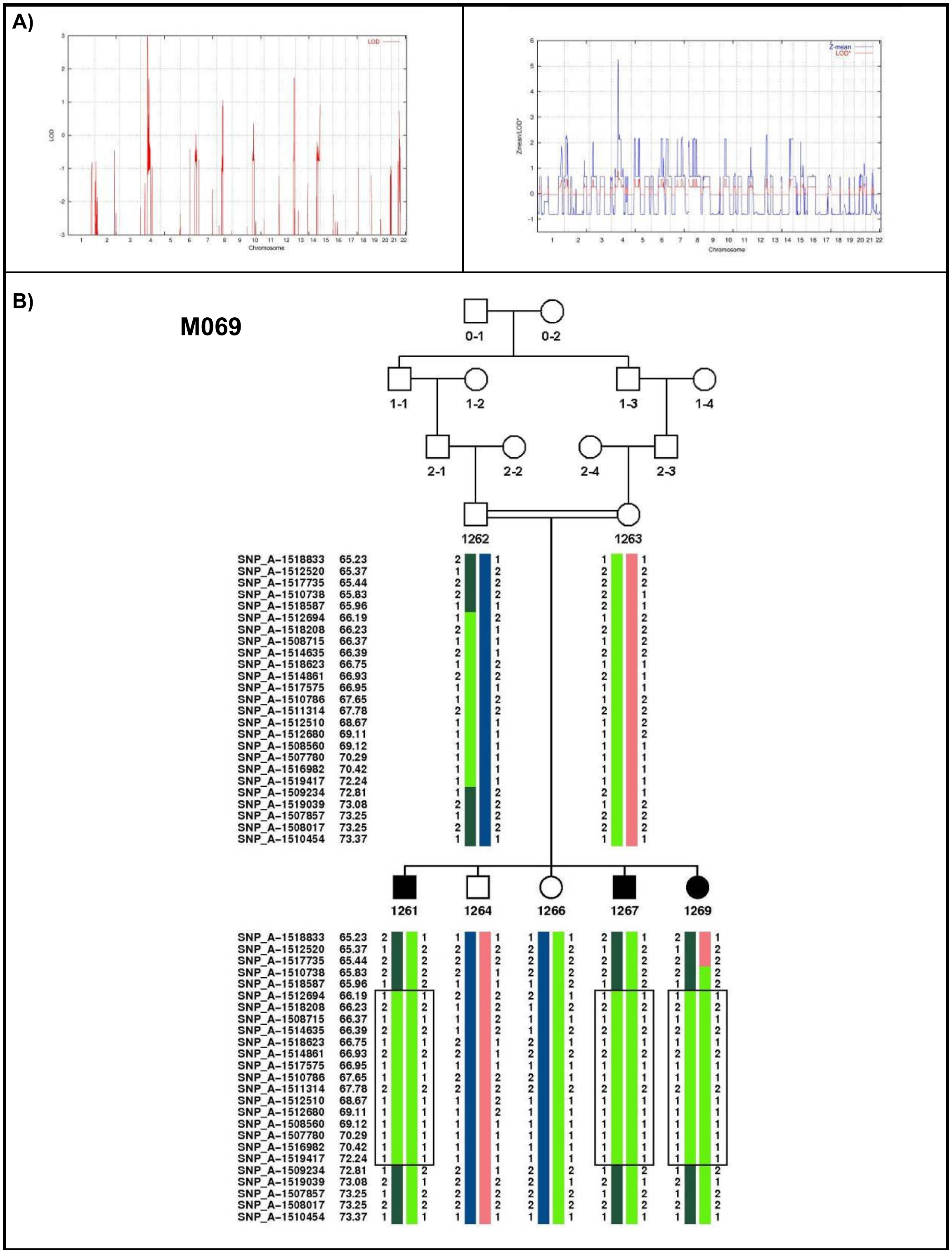
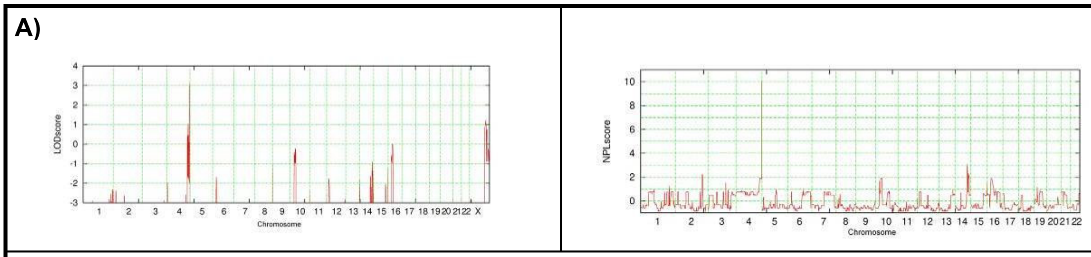


Fig S-4: Family M069. Whole genome parametric (Merlin) and non-parametric (Merlin) linkage results (A) and haplotype of the only linkage interval with significant parametric LOD score of 3 (B)

A)



B)

M150N

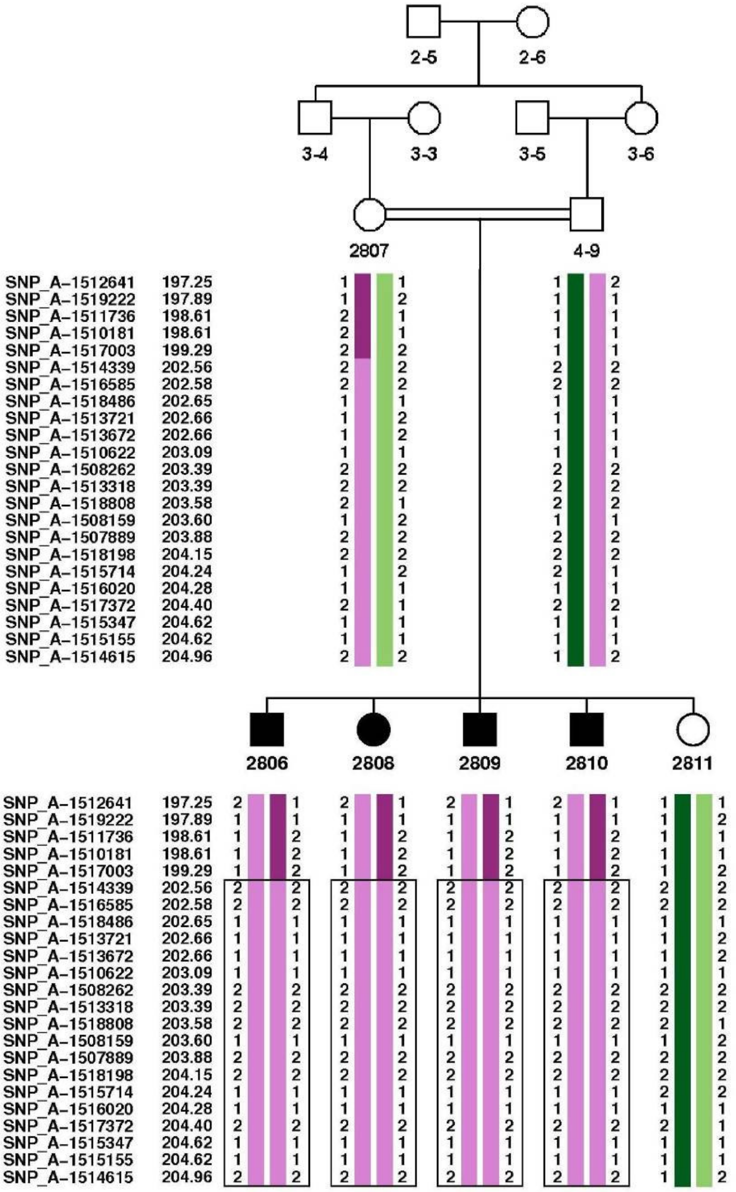

2-5
2-6
3-4
3-3
3-5
3-6
2807
4-9
SNP_A-1512641 197.25 1 1 1 2
SNP_A-1519222 197.89 1 2 1 1
SNP_A-1511736 198.61 2 1 1 1
SNP_A-1510181 198.61 2 1 1 1
SNP_A-1517003 199.29 2 2 1 1
SNP_A-1514339 202.56 2 2 2 2
SNP_A-1516585 202.58 2 2 2 2
SNP_A-1518486 202.65 1 1 1 1
SNP_A-1513721 202.66 1 2 1 1
SNP_A-1513672 202.66 1 2 1 1
SNP_A-1510622 203.09 1 1 1 1
SNP_A-1508262 203.39 2 2 2 2
SNP_A-1513318 203.39 2 2 2 2
SNP_A-1518808 203.58 2 1 2 2
SNP_A-1508159 203.60 1 2 1 1
SNP_A-1507889 203.88 2 2 2 2
SNP_A-1518198 204.15 2 2 2 2
SNP_A-1515714 204.24 1 2 2 1
SNP_A-1516020 204.28 1 1 1 1
SNP_A-1517372 204.40 2 1 2 2
SNP_A-1515347 204.62 1 1 1 1
SNP_A-1515155 204.62 1 1 1 1
SNP_A-1514615 204.96 2 2 1 2
2806
2808
2809
2810
2811
SNP_A-1512641 197.25 2 1 2 1 2 1 2 1 1 1
SNP_A-1519222 197.89 1 1 1 1 1 1 1 1 1 2
SNP_A-1511736 198.61 1 2 1 2 1 2 1 2 1 1
SNP_A-1510181 198.61 1 2 1 2 1 2 1 2 1 1
SNP_A-1517003 199.29 1 2 1 2 1 2 1 2 1 2
SNP_A-1514339 202.56 2 2 2 2 2 2 2 2 2 2
SNP_A-1516585 202.58 2 2 2 2 2 2 2 2 2 2
SNP_A-1518486 202.65 1 1 1 1 1 1 1 1 1 1
SNP_A-1513721 202.66 1 1 1 1 1 1 1 1 1 2
SNP_A-1513672 202.66 1 1 1 1 1 1 1 1 1 2
SNP_A-1510622 203.09 1 1 1 1 1 1 1 1 1 1
SNP_A-1508262 203.39 2 2 2 2 2 2 2 2 2 2
SNP_A-1513318 203.39 2 2 2 2 2 2 2 2 2 2
SNP_A-1518808 203.58 2 2 2 2 2 2 2 2 2 1
SNP_A-1508159 203.60 1 1 1 1 1 1 1 1 1 2
SNP_A-1507889 203.88 2 2 2 2 2 2 2 2 2 2
SNP_A-1518198 204.15 2 2 2 2 2 2 2 2 2 2
SNP_A-1515714 204.24 1 1 1 1 1 1 1 1 2 2
SNP_A-1516020 204.28 1 1 1 1 1 1 1 1 1 1
SNP_A-1517372 204.40 2 2 2 2 2 2 2 2 2 1
SNP_A-1515347 204.62 1 1 1 1 1 1 1 1 1 1
SNP_A-1515155 204.62 1 1 1 1 1 1 1 1 1 1
SNP_A-1514615 204.96 2 2 2 2 2 2 2 2 1 2

Fig S-5: Family M150N. Whole genome parametric (Allegro) and non-parametric (GeneHunter) linkage results (A) and haplotype of the only linkage interval with significant parametric LOD score of 3.1 (B)

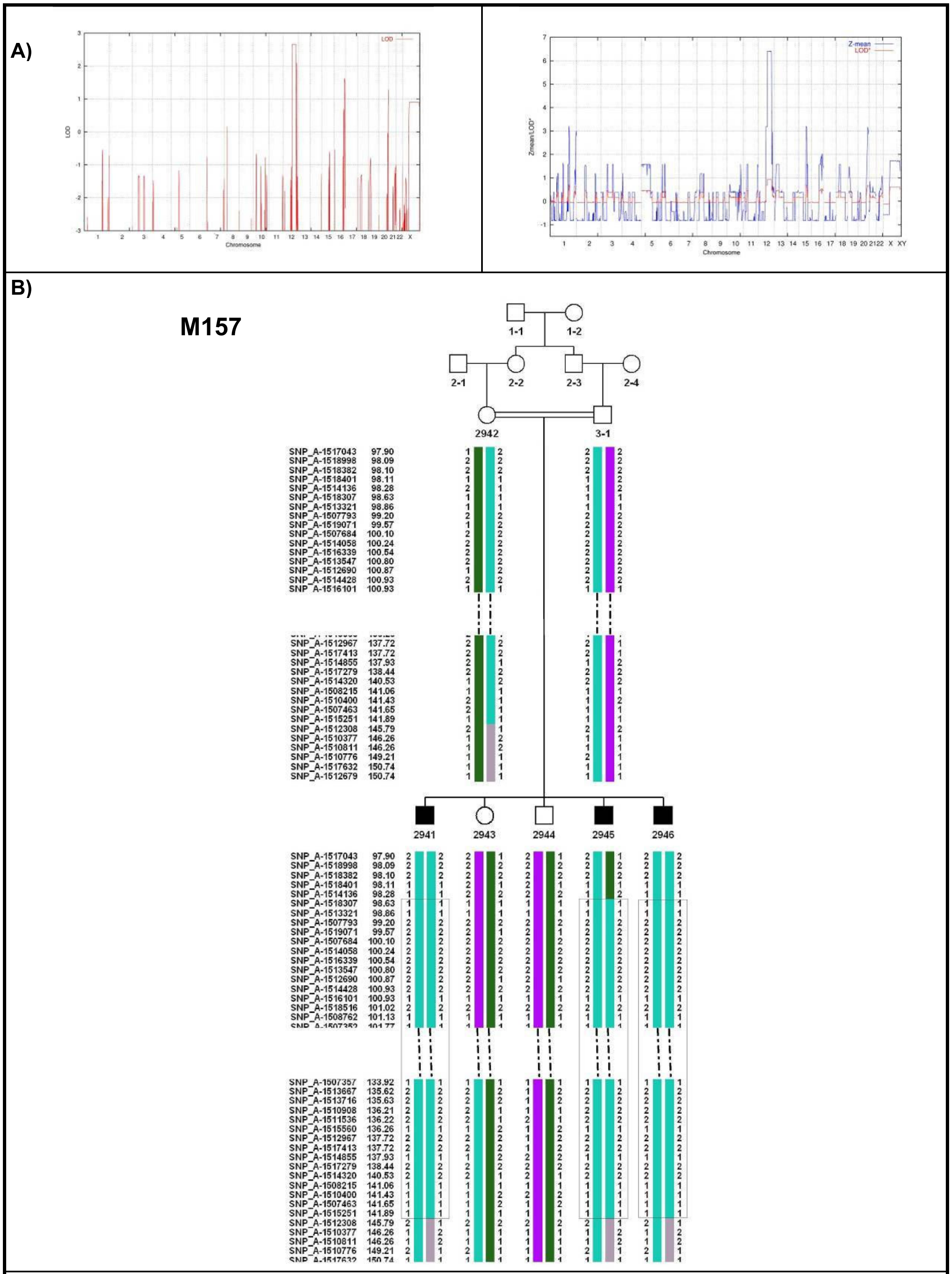


Fig S- 6: Family M157. Whole genome parametric (Merlin) and non-parametric (Merlin) linkage results (A) and haplotype of the only linkage interval with significant parametric LOD score of 2.6 (B)

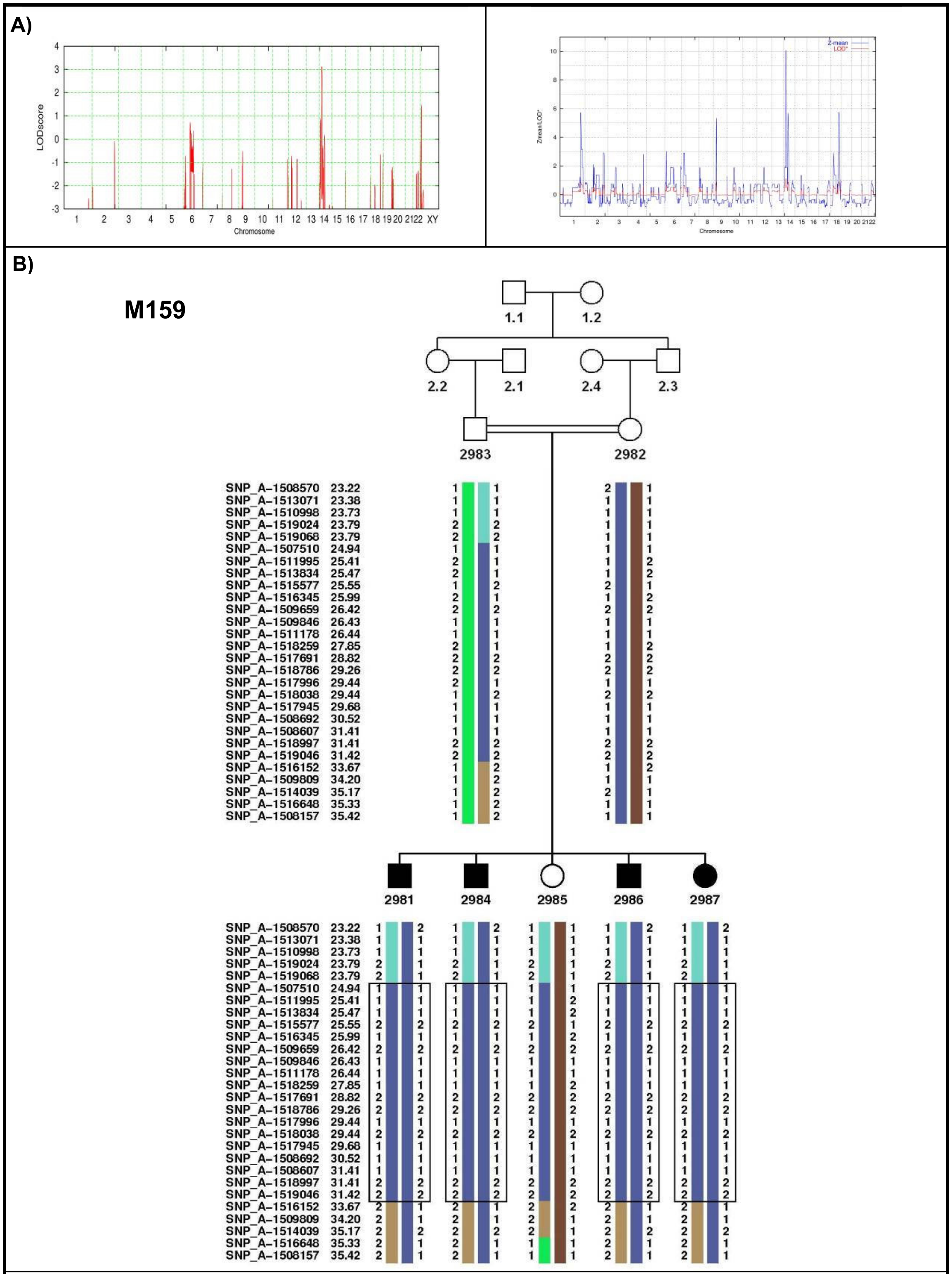


Fig S-7: Family M159. Whole genome parametric (Allegro) and non-parametric (Merlin) linkage results (A) and haplotype of the only linkage interval with significant parametric LOD score of 3.2 (B)

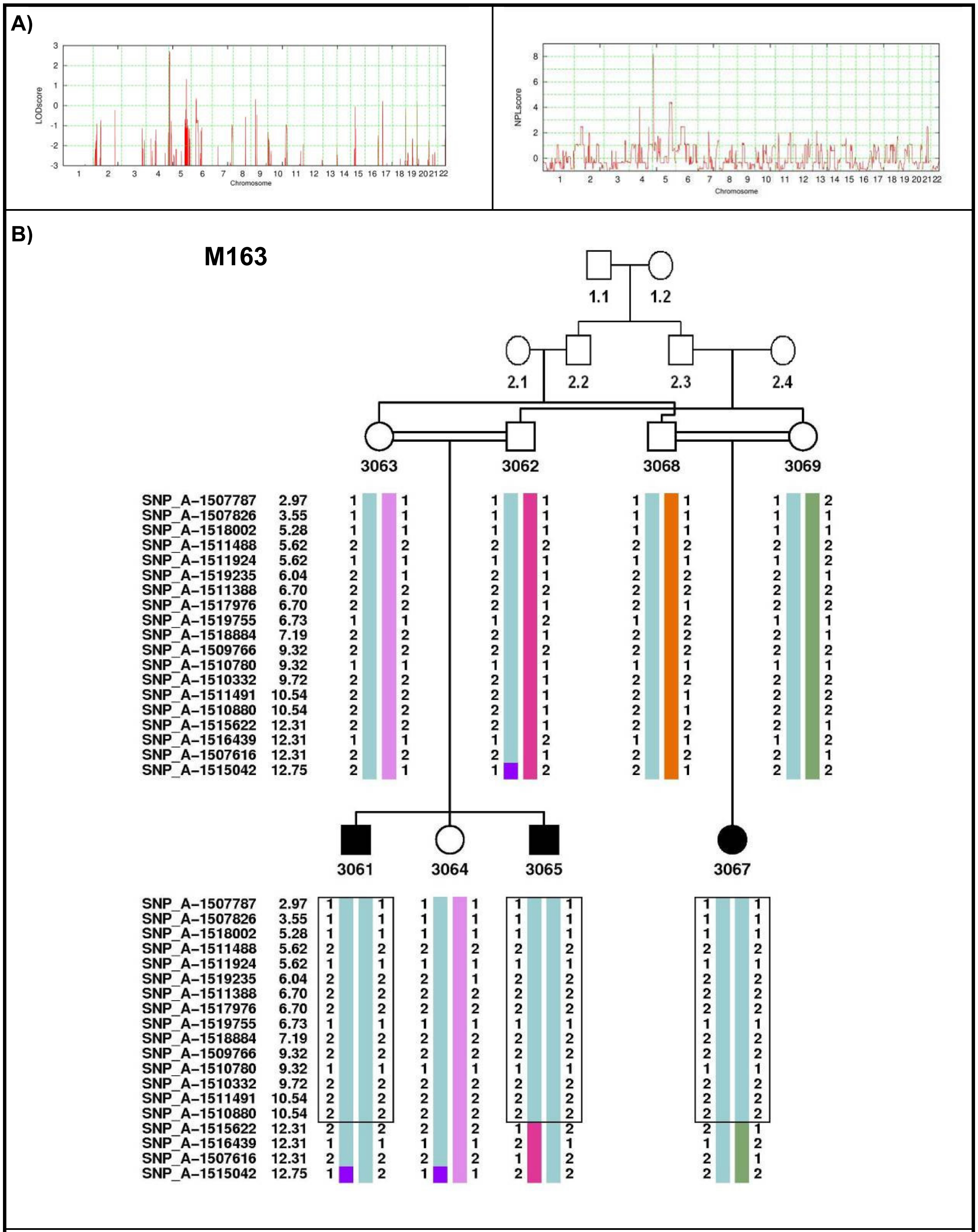


Fig S-8: Family M163. Whole genome parametric (Allegro) and non-parametric (GeneHunter) linkage results (A) and haplotype of the only linkage interval with significant parametric LOD score of 2.8 (B).

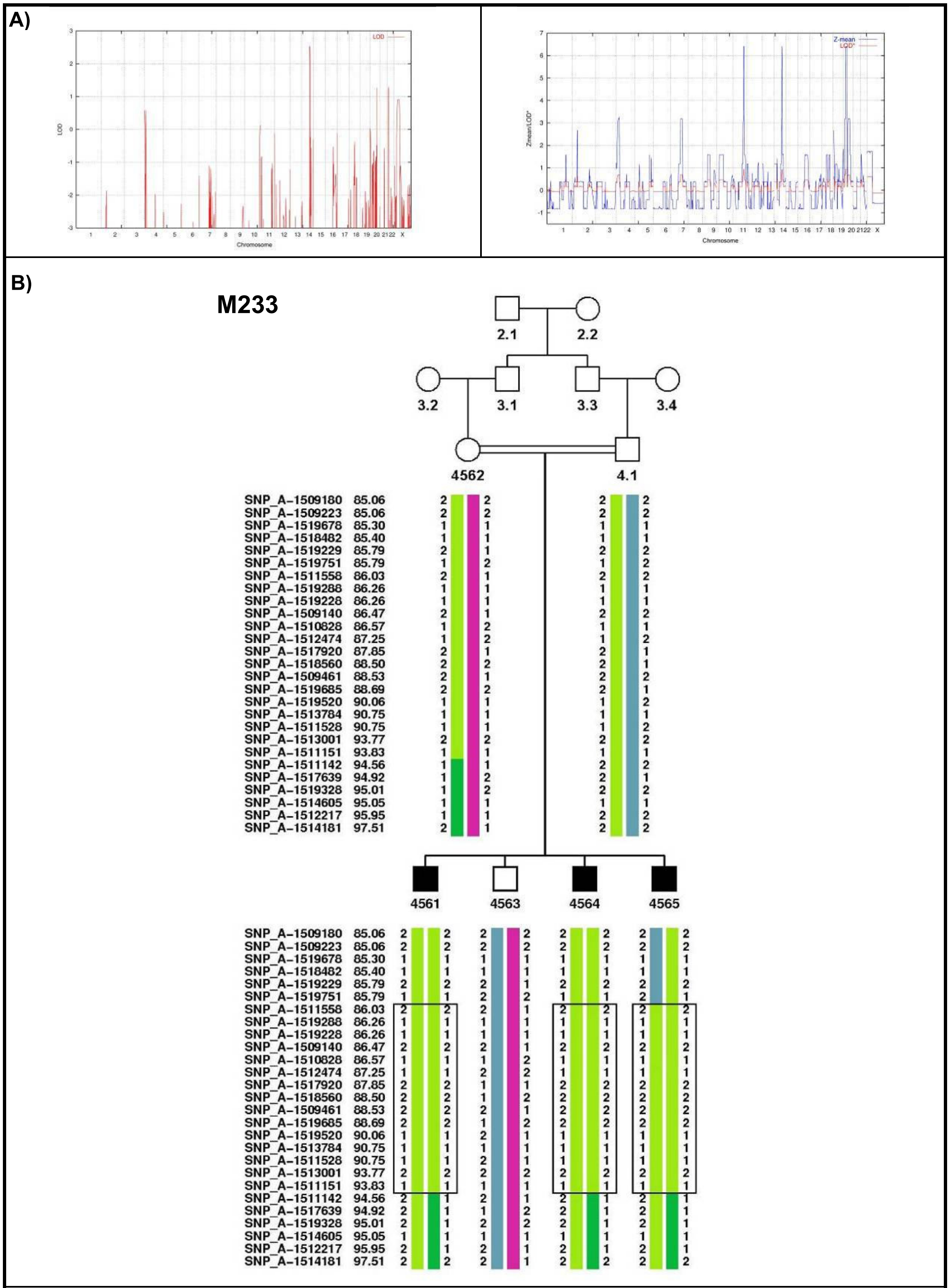


Fig S-9: Family M233. Whole genome parametric (Merlin) and non-parametric (Merlin) linkage results (A) and haplotype of the only linkage interval with significant parametric LOD score of 2.5 (B).

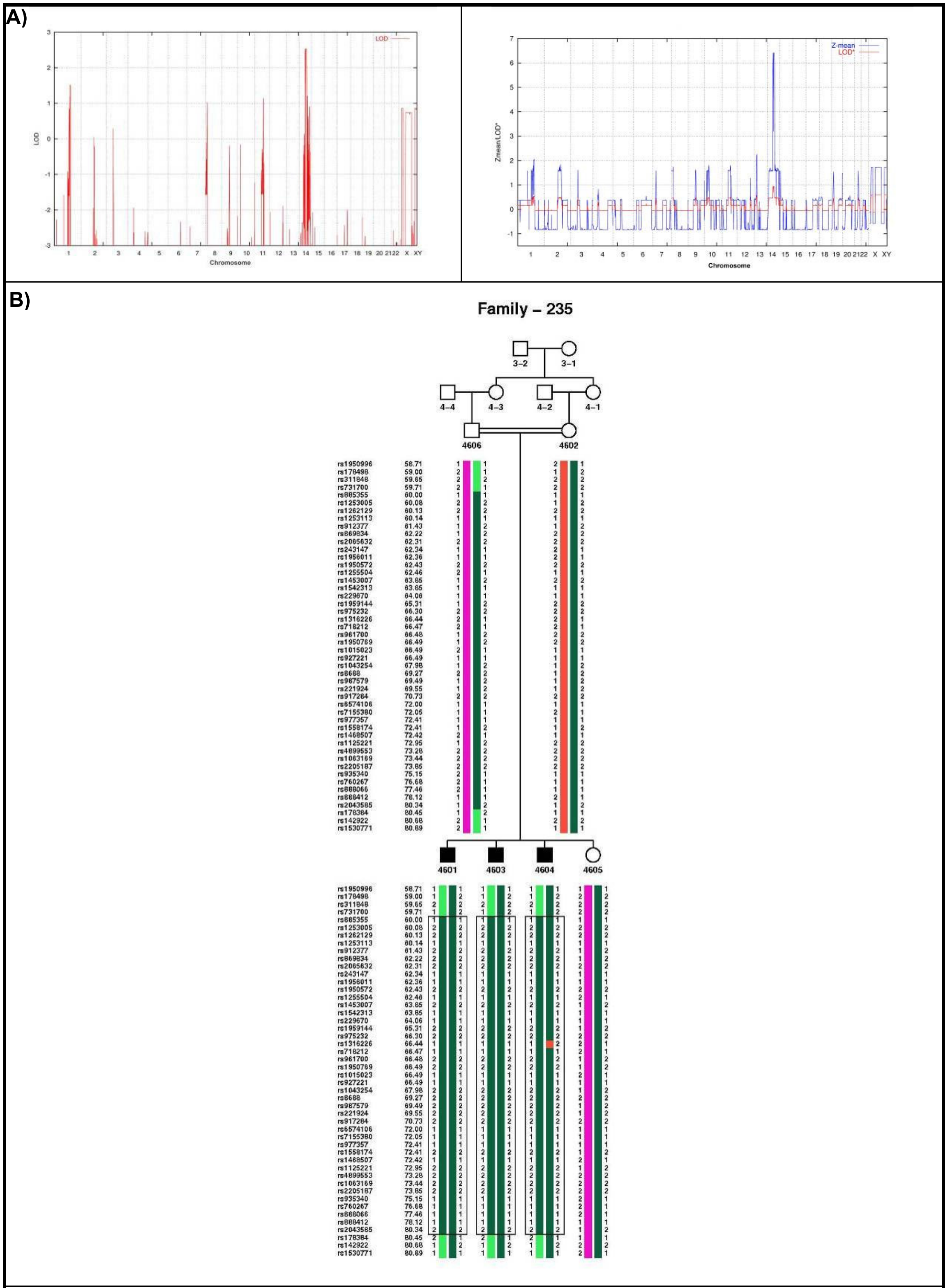
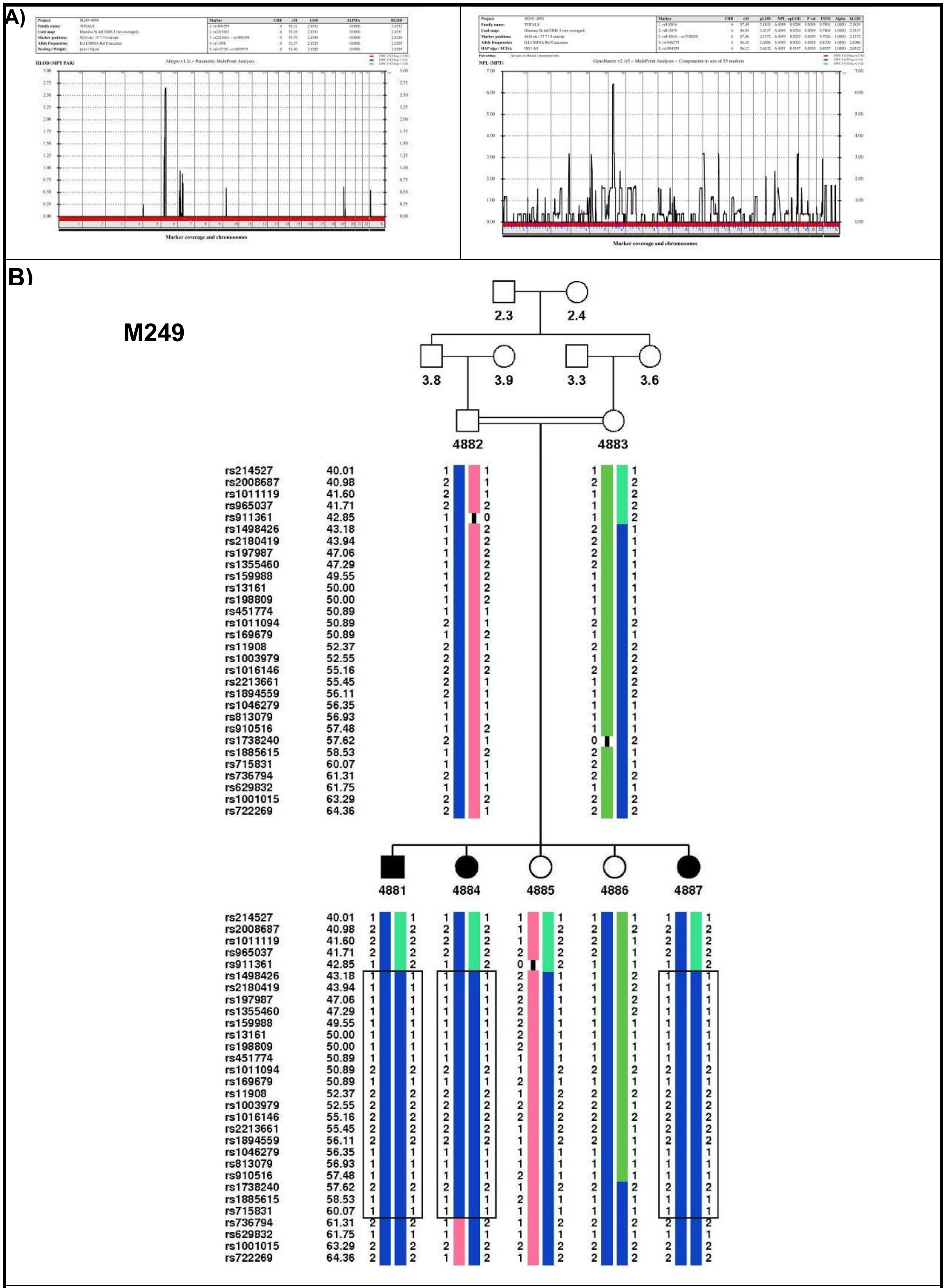


Fig S-10: Family M235. Whole genome parametric (Merlin) and non-parametric (Merlin) linkage results (A) and haplotype of the only linkage interval with significant parametric LOD score of 2.5 (B).


A)
HLOD (MPT PAR)
Allegro v1.2c – Parametric MultiPoint Analyses
Marker coverage and chromosomes
NPL (MPT)
GeneHunter v2.1r5 – MultiPoint Analyses – Computation in sets of 55 markers
Marker coverage and chromosomes
B)
M249
2.3
2.4
3.8
3.9
3.3
3.6
4882
4883
4881
4884
4885
4886
4887

		4882		4883	
rs214527	40.01	1	1	1	1
rs2008687	40.98	2	1	2	2
rs1011119	41.60	2	1	1	2
rs965037	41.71	2	2	1	2
rs911361	42.85	1	0	1	2
rs1498426	43.18	1	2	2	1
rs2180419	43.94	1	2	2	1
rs197987	47.06	1	2	2	1
rs1355460	47.29	1	2	2	1
rs159988	49.55	1	2	1	1
rs13161	50.00	1	2	1	1
rs198809	50.00	1	2	1	1
rs451774	50.89	1	1	1	1
rs1011094	50.89	2	1	2	2
rs169679	50.89	1	2	1	1
rs11908	52.37	2	1	2	2
rs1003979	52.55	2	2	1	2
rs1016146	55.16	2	2	2	2
rs2213661	55.45	2	1	1	2
rs1894559	56.11	2	1	1	2
rs1046279	56.35	1	1	1	1
rs813079	56.93	1	1	1	1
rs910516	57.48	1	2	1	1
rs1738240	57.62	2	1	0	2
rs1885615	58.53	1	2	2	1
rs715831	60.07	1	1	2	1
rs736794	61.31	2	1	2	2
rs629832	61.75	1	1	1	1
rs1001015	63.29	2	2	2	2
rs722269	64.36	2	1	2	2

		4881		4884		4885		4886		4887	
rs214527	40.01	1	1	1	1	1	1	1	1	1	1
rs2008687	40.98	2	2	2	2	1	2	2	2	2	2
rs1011119	41.60	2	2	2	2	1	2	2	1	2	2
rs965037	41.71	2	2	2	2	2	2	2	1	2	2
rs911361	42.85	1	2	1	2	0	2	1	1	1	2
rs1498426	43.18	1	1	1	1	2	1	1	2	1	1
rs2180419	43.94	1	1	1	1	2	1	1	2	1	1
rs197987	47.06	1	1	1	1	2	1	1	2	1	1
rs1355460	47.29	1	1	1	1	2	1	1	2	1	1
rs159988	49.55	1	1	1	1	2	1	1	1	1	1
rs13161	50.00	1	1	1	1	2	1	1	1	1	1
rs198809	50.00	1	1	1	1	2	1	1	1	1	1
rs451774	50.89	1	1	1	1	1	1	1	1	1	1
rs1011094	50.89	2	2	2	2	1	2	2	2	2	2
rs169679	50.89	1	1	1	1	2	1	1	1	1	1
rs11908	52.37	2	2	2	2	1	2	2	2	2	2
rs1003979	52.55	2	2	2	2	2	2	2	1	2	2
rs1016146	55.16	2	2	2	2	2	2	2	2	2	2
rs2213661	55.45	2	2	2	2	1	2	2	1	2	2
rs1894559	56.11	2	2	2	2	1	2	2	1	2	2
rs1046279	56.35	1	1	1	1	1	1	1	1	1	1
rs813079	56.93	1	1	1	1	1	1	1	1	1	1
rs910516	57.48	1	1	1	1	2	1	1	1	1	1
rs1738240	57.62	2	2	2	2	1	2	2	2	2	2
rs1885615	58.53	1	1	1	1	2	1	1	1	1	1
rs715831	60.07	1	1	1	1	1	1	1	1	1	1
rs736794	61.31	2	2	1	2	1	2	2	2	2	2
rs629832	61.75	1	1	1	1	1	1	1	1	1	1
rs1001015	63.29	2	2	2	2	2	2	2	2	2	2
rs722269	64.36	2	2	1	2	1	2	2	2	2	2

Fig S-11: Family M249. Whole genome parametric (Allegro) and non-parametric (GeneHunter) linkage results (A) and haplotype of the only linkage interval with significant parametric LOD score of 2.65 (B).

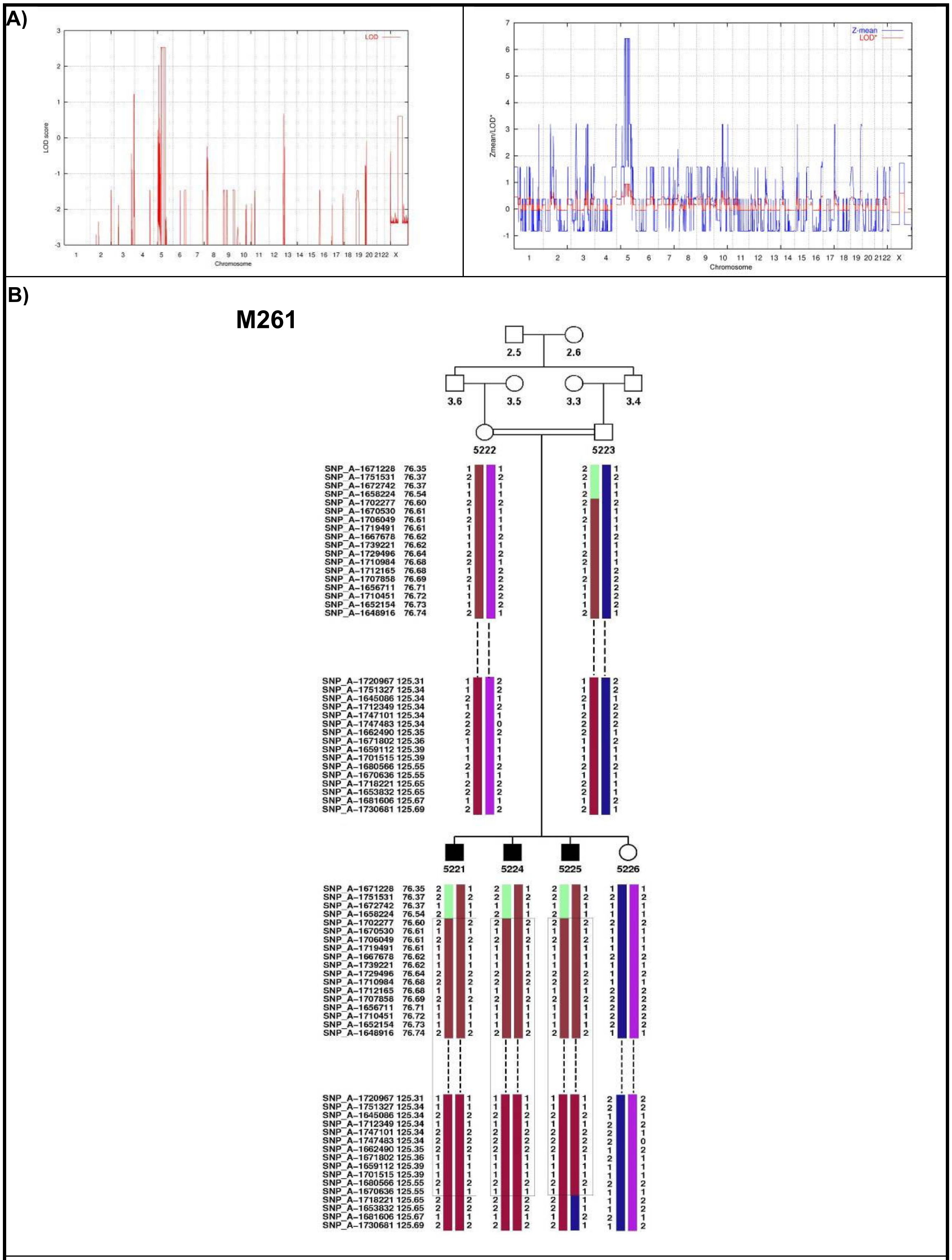


Fig S-12: Family M261. Whole genome parametric (Merlin) and non-parametric (Merlin) linkage results (A) and haplotype of the only linkage interval with significant parametric LOD score of 2.53 (B). Due to the high number of markers just several markers in the both flanking sites of the interval are shown, and markers in between were excluded from the figure.

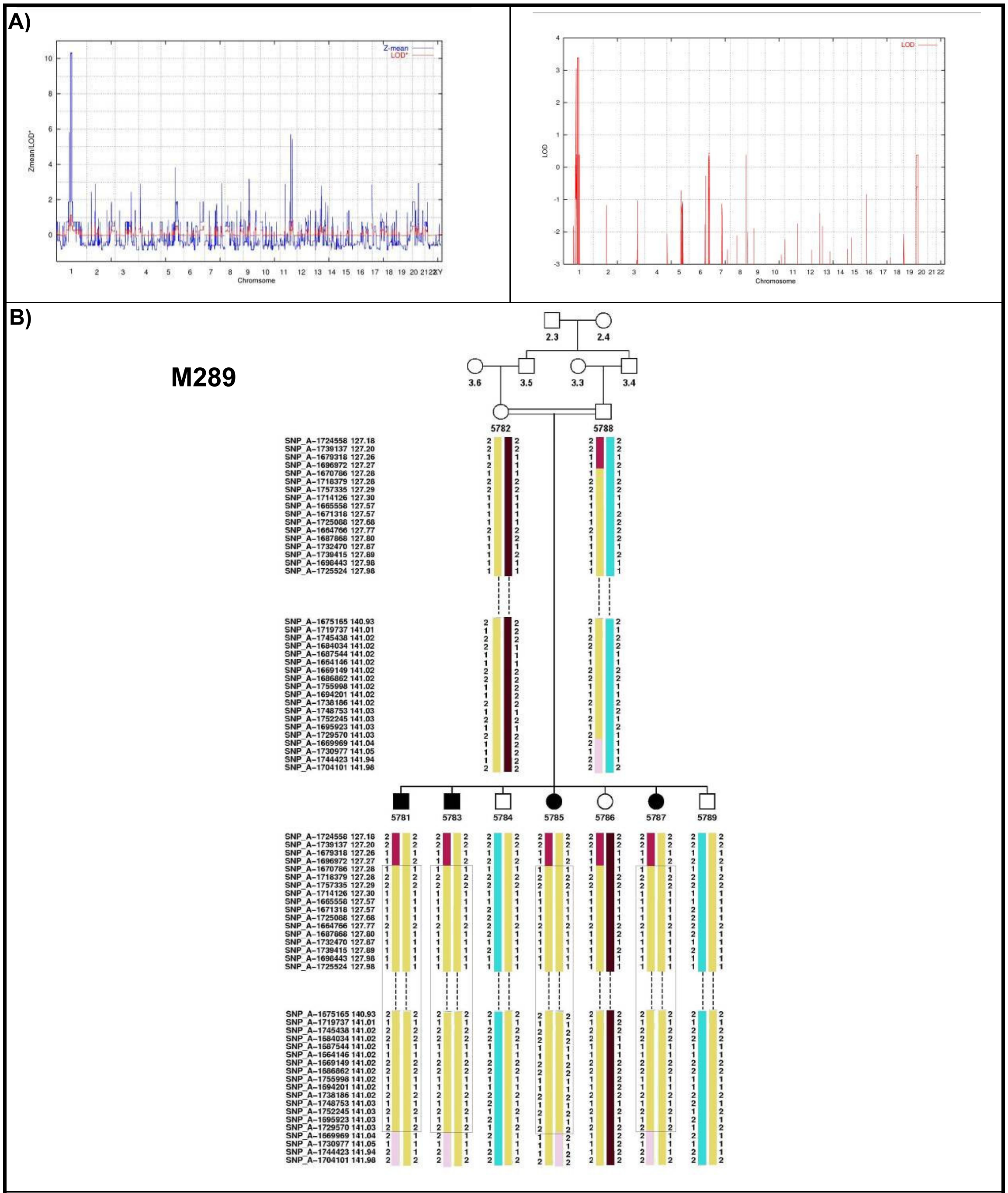


Fig S-13: Family M289. Whole genome parametric (Merlin) and non-parametric (Merlin) linkage results (A) and haplotype of the only linkage interval with significant parametric LOD score of 3.38 (B). Degree of consanguinity was not clear, analysis were done with the assumption of first cousin degree of consanguinity. Due to the high number of markers just several markers in the both flanking sites of the interval are shown, and markers in between were excluded from the figure.

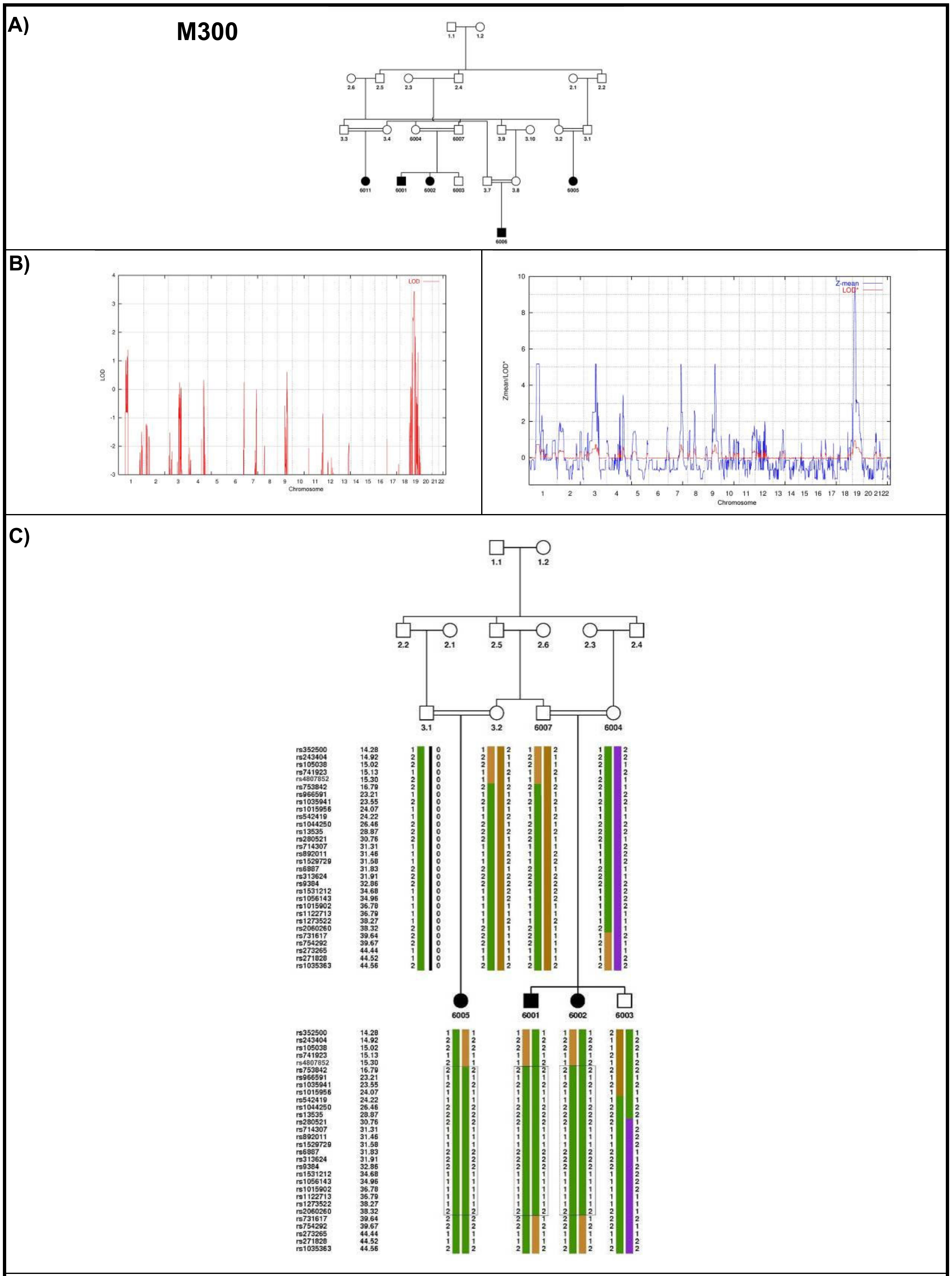


Fig S-14: Family M300. The whole pedgree (A) Whole genome parametric (Merlin) and non-parametric (Merlin) linkage results (B) and haplotype of the only linkage interval with significant parametric LOD score of 3.4 for the branches those were compatible with eachother (C).

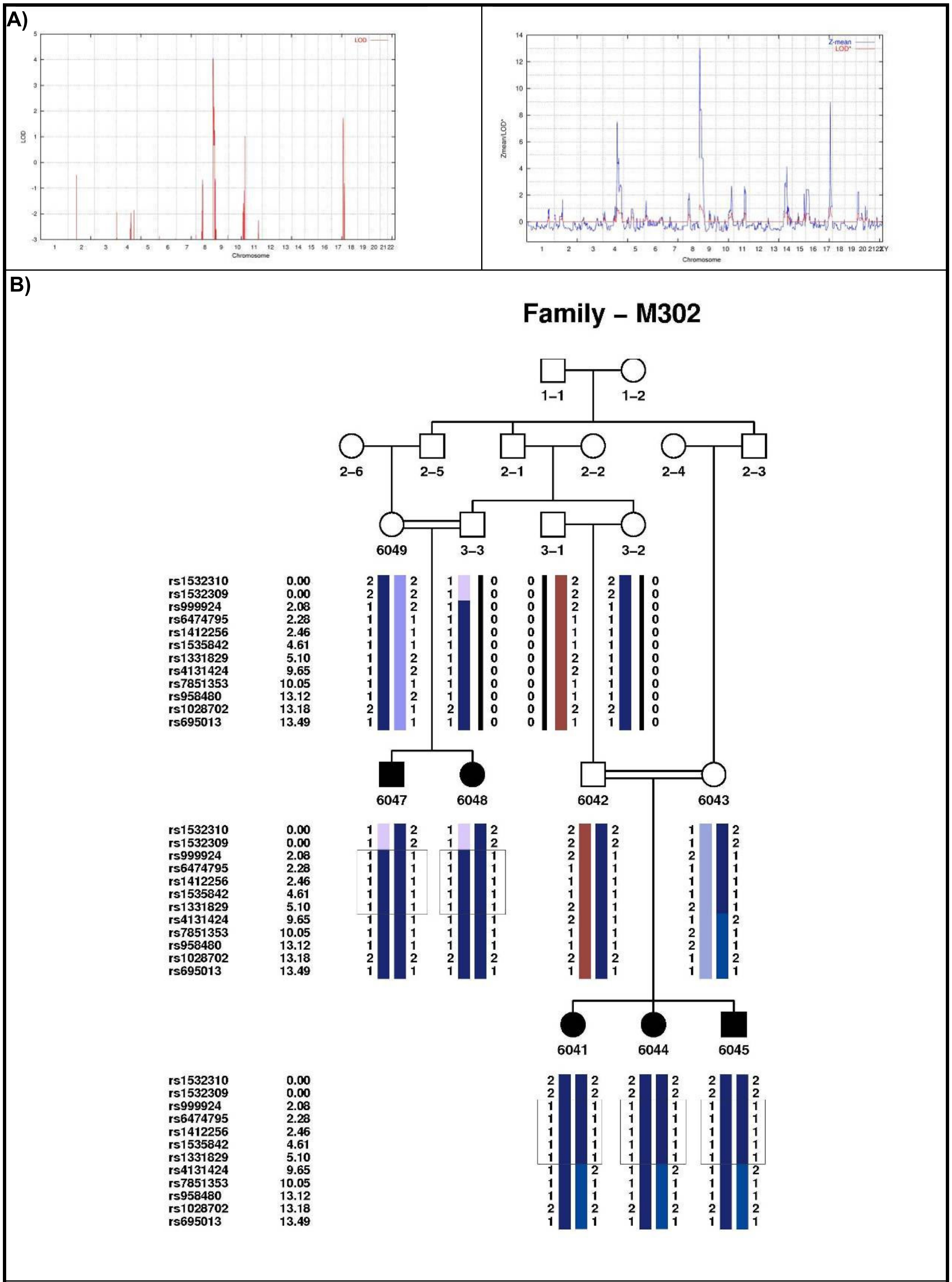


Fig S-15: Family M302. Whole genome parametric (Merlin) and non-parametric (Merlin) linkage results (A) and haplotype of the only linkage interval with significant parametric LOD score of 4.1 (B).

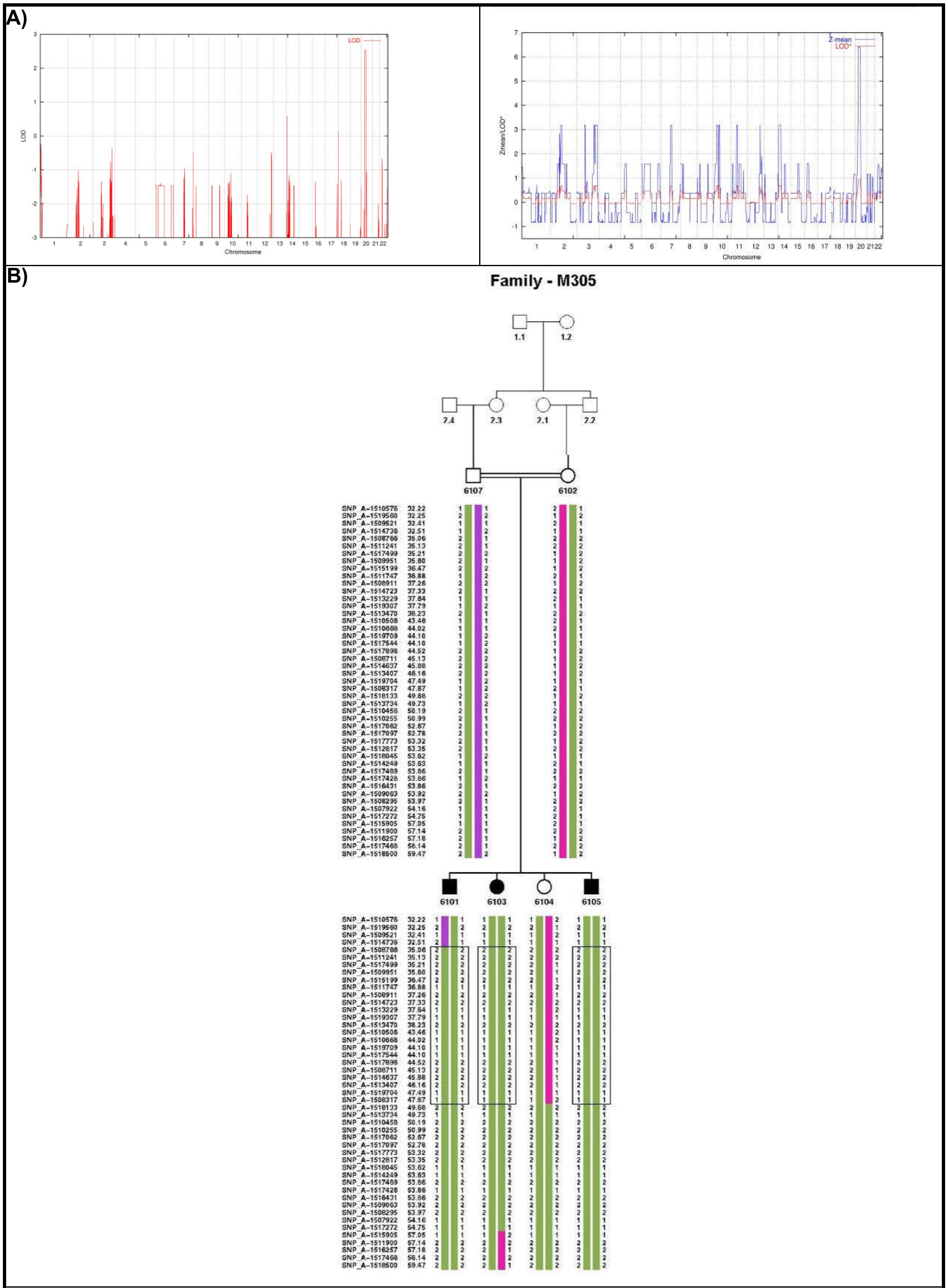


Fig S-16: Family M305. Whole genome parametric (Merlin) and non-parametric (Merlin) linkage results (A) and haplotype of the only linkage interval with significant parametric LOD score of 2.5 (B).

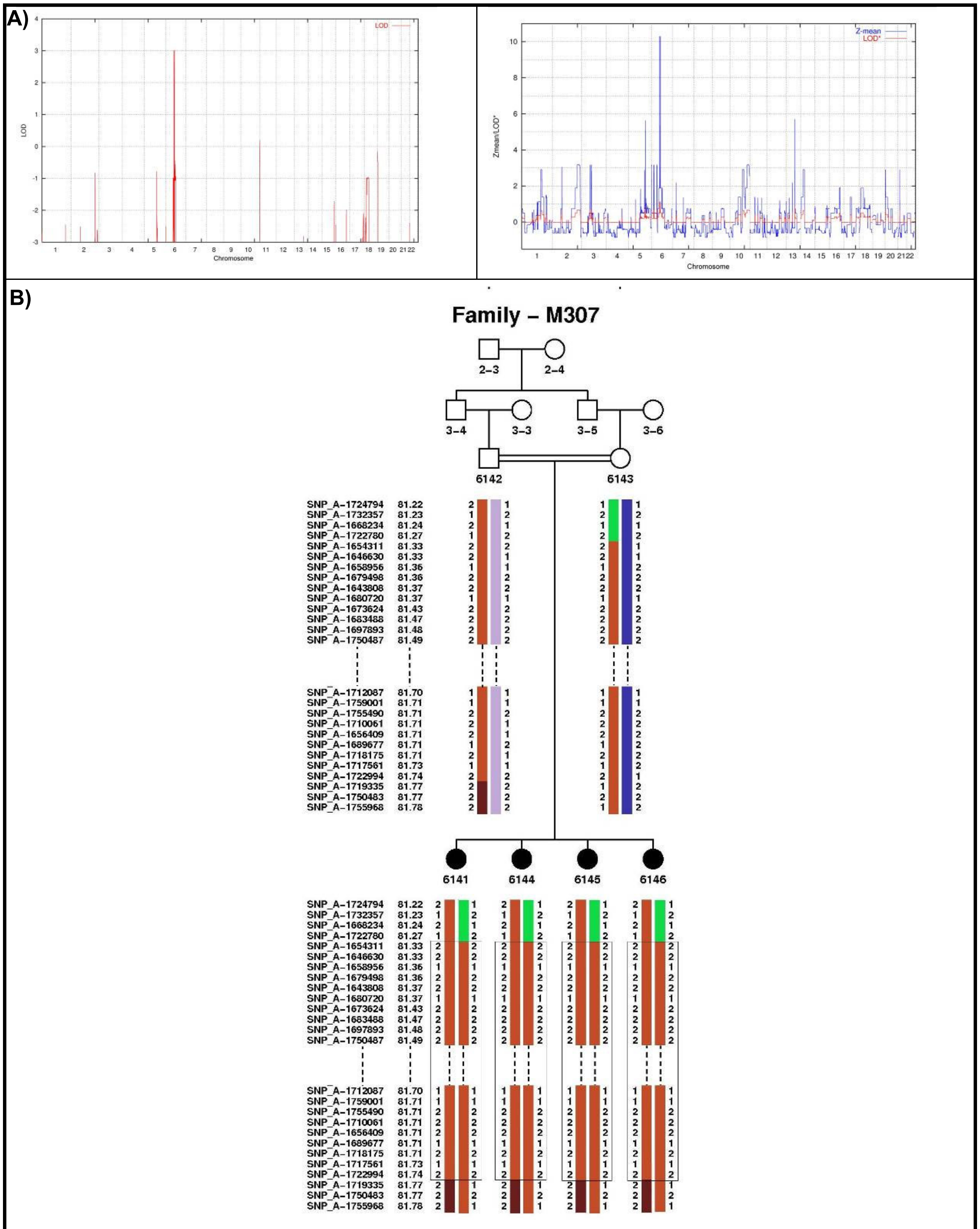


Fig S-17: Family M307. Whole genome parametric (Merlin) and non-parametric (Merlin) linkage results (A) and haplotype of the only linkage interval with significant parametric LOD score of 3 (B). Due to the high number of markers just several markers in the both flanking sites of the interval are shown, and markers in between were excluded from the figure.

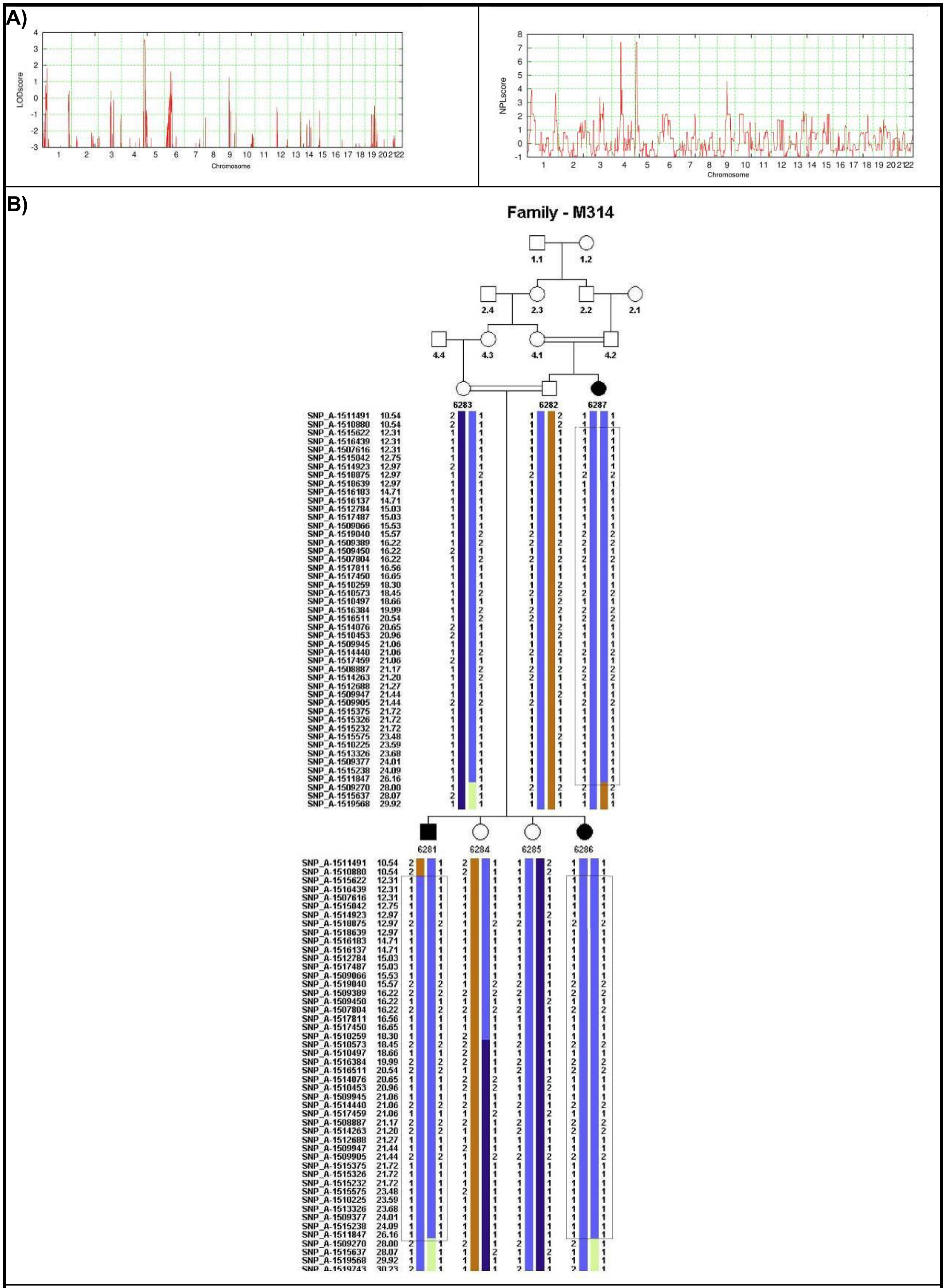


Fig S-18: Family M314. Whole genome parametric (GeneHunter) and non-parametric (GeneHunter) linkage results (A) and haplotype of the only linkage interval with significant parametric LOD score of 3.5 (B).

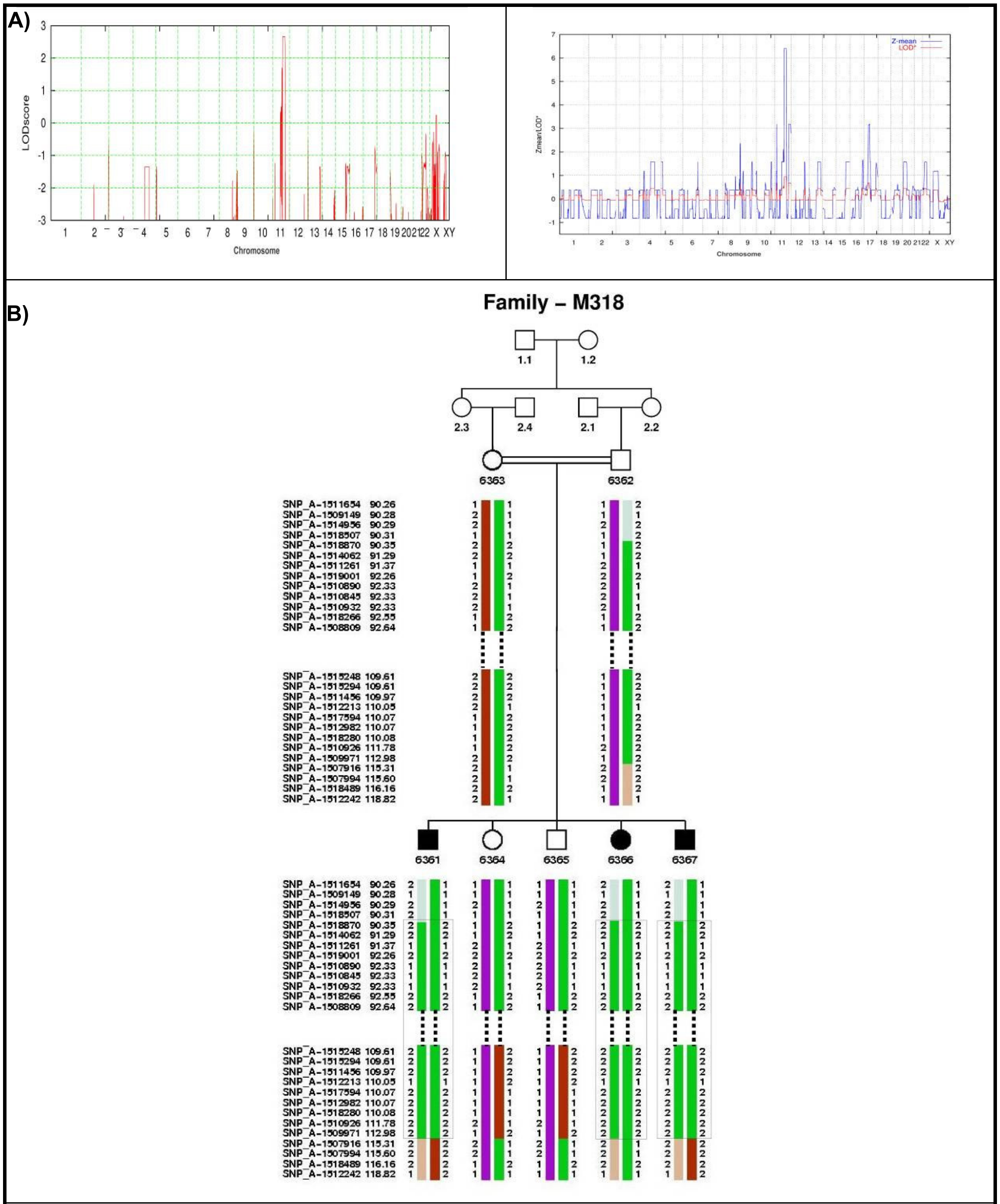


Fig S-19: Family M318. Whole genome parametric (Merlin) and non-parametric (Merlin) linkage results (A) and haplotype of the only linkage interval with significant parametric LOD score of 2.65 (B). Due to the high number of markers just several markers in the both flanking sites of the interval are shown, and markers in between were excluded from the figure.

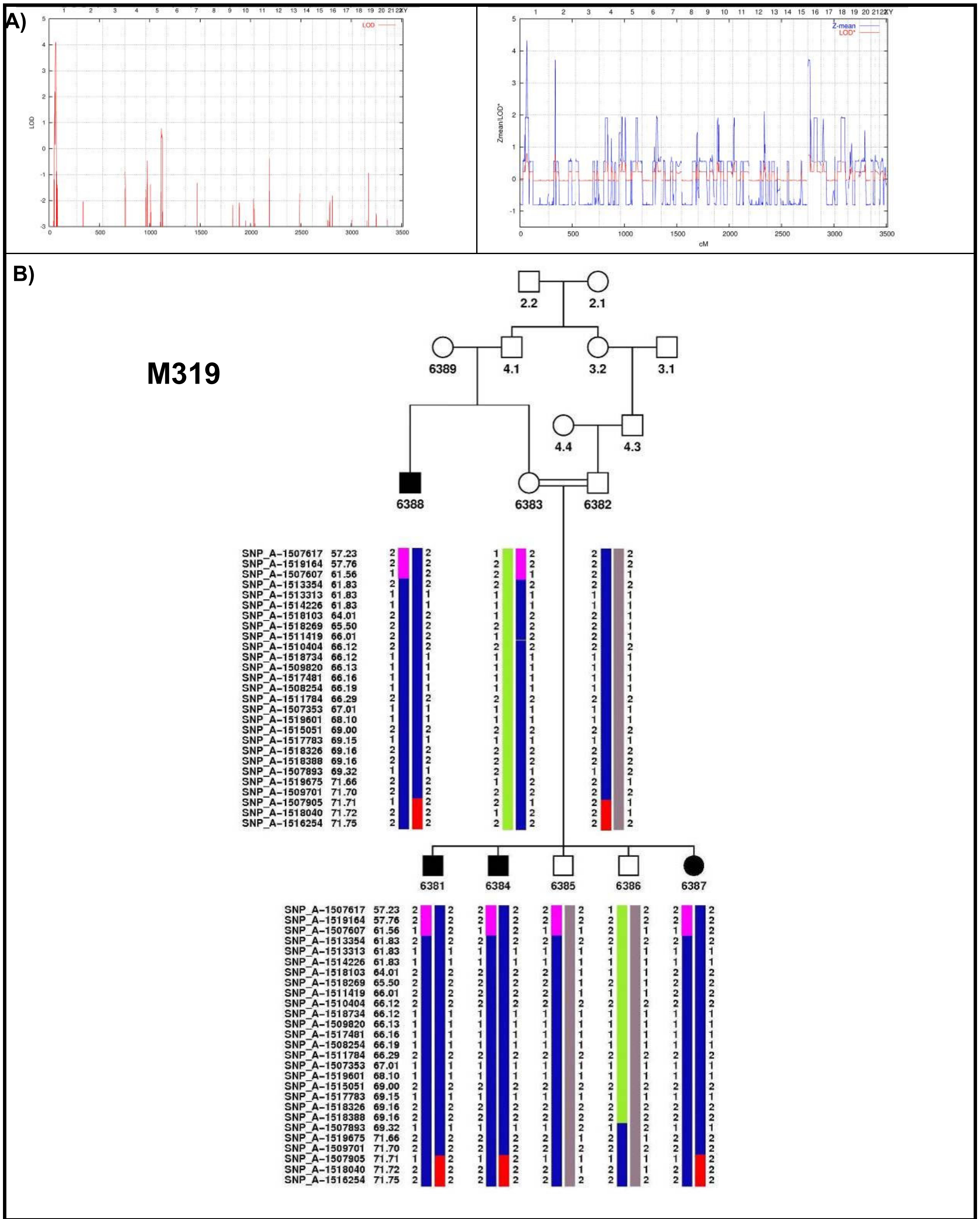


Fig S-20: Family M319. Whole genome parametric (Merlin) and non-parametric (Merlin) linkage results (A) and haplotype of the only linkage interval with significant parametric LOD score of 4.1 (B).

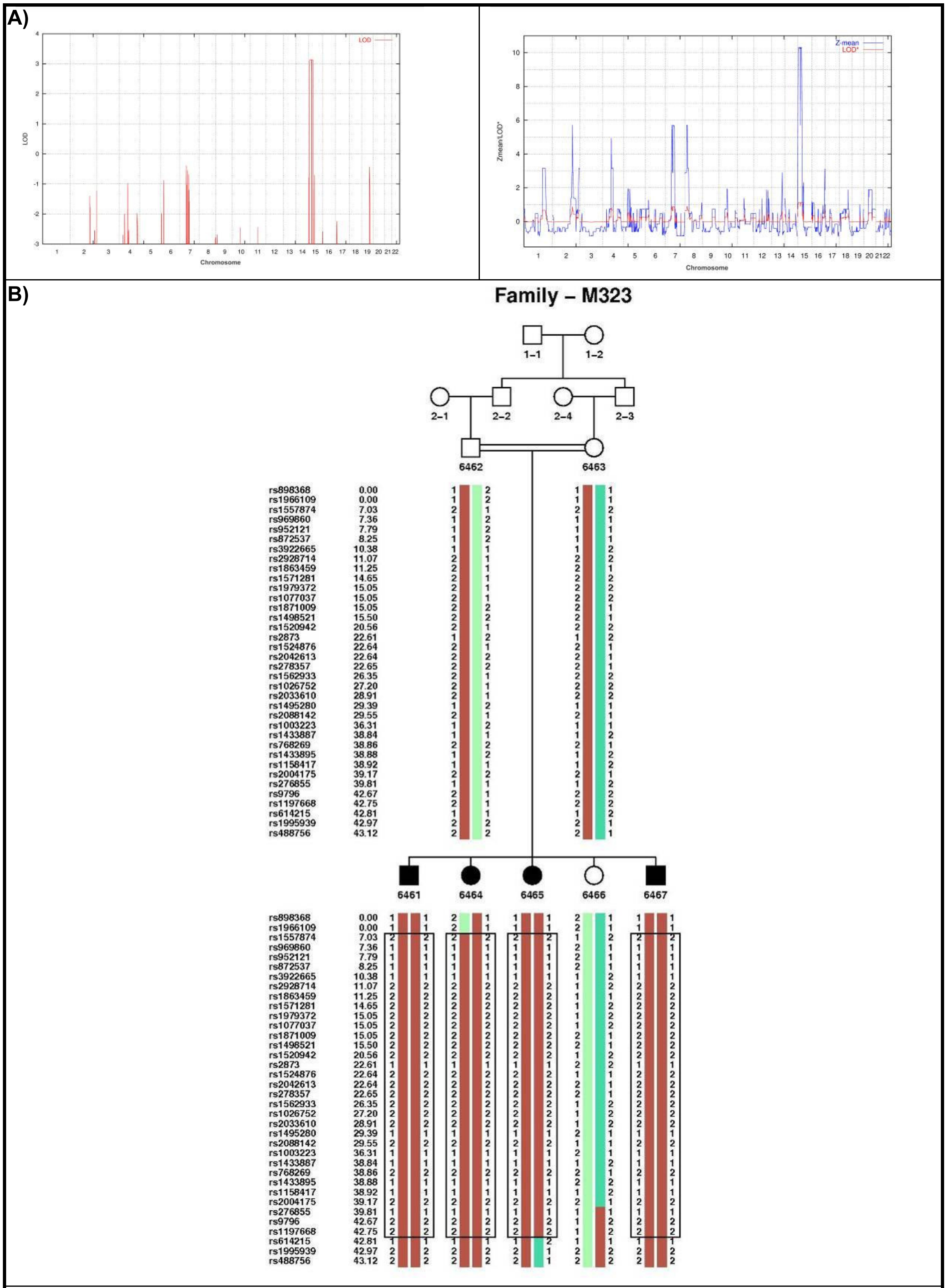
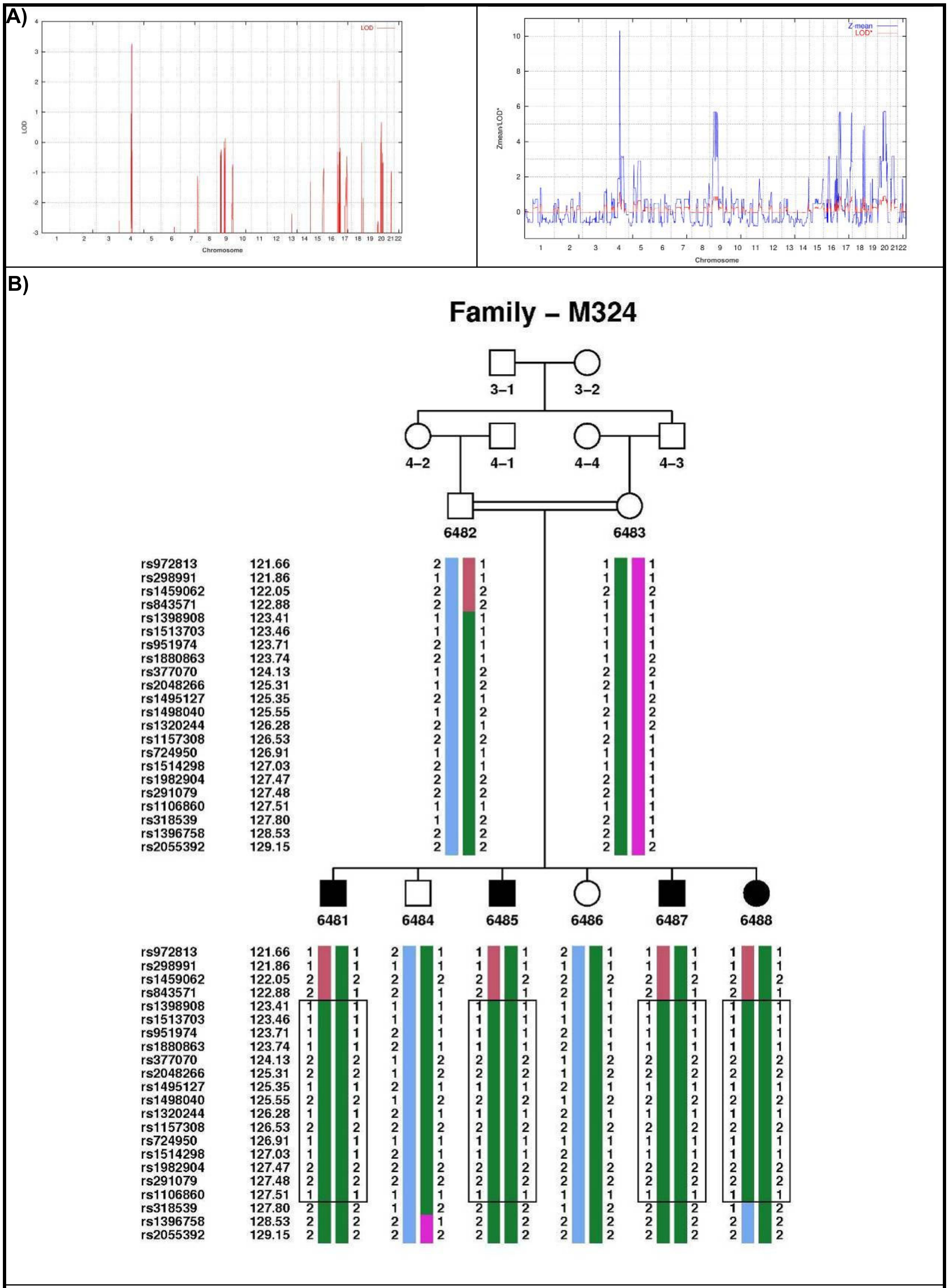


Fig S-21: Family M323. Whole genome parametric (Merlin) and non-parametric (Merlin) linkage results (A) and haplotype of the only linkage interval with significant parametric LOD score of 3.2 (B).



		6482		6483	
rs972813	121.66	2	1	1	1
rs298991	121.86	1	1	1	1
rs1459062	122.05	2	2	2	2
rs843571	122.88	2	2	1	1
rs1398908	123.41	1	1	1	1
rs1513703	123.46	1	1	1	1
rs951974	123.71	2	1	1	1
rs1880863	123.74	2	1	1	2
rs377070	124.13	1	2	2	2
rs2048266	125.31	1	2	2	1
rs1495127	125.35	2	1	1	2
rs1498040	125.55	1	2	2	2
rs1320244	126.28	2	1	1	2
rs1157308	126.53	2	2	2	1
rs724950	126.91	1	1	1	1
rs1514298	127.03	2	1	1	1
rs1982904	127.47	2	2	2	1
rs291079	127.48	2	2	2	1
rs1106860	127.51	1	1	1	1
rs318539	127.80	1	2	2	1
rs1396758	128.53	2	2	2	1
rs2055392	129.15	2	2	2	2

		6481		6484		6485		6486		6487		6488	
rs972813	121.66	1	1	2	1	1	1	2	1	1	1	1	1
rs298991	121.86	1	1	1	1	1	1	1	1	1	1	1	1
rs1459062	122.05	2	2	2	2	2	2	2	2	2	2	2	2
rs843571	122.88	2	1	2	1	2	1	2	1	2	1	2	1
rs1398908	123.41	1	1	1	1	1	1	1	1	1	1	1	1
rs1513703	123.46	1	1	1	1	1	1	1	1	1	1	1	1
rs951974	123.71	1	1	2	1	1	1	2	1	1	1	1	1
rs1880863	123.74	1	1	2	1	1	1	2	1	1	1	1	1
rs377070	124.13	2	2	1	2	2	2	1	2	2	2	2	2
rs2048266	125.31	2	2	1	2	2	2	1	2	2	2	2	2
rs1495127	125.35	1	1	2	1	1	1	2	1	1	1	1	1
rs1498040	125.55	2	2	1	2	2	2	1	2	2	2	2	2
rs1320244	126.28	1	1	2	1	1	1	2	1	1	1	1	1
rs1157308	126.53	2	2	2	2	2	2	2	2	2	2	2	2
rs724950	126.91	1	1	1	1	1	1	1	1	1	1	1	1
rs1514298	127.03	1	1	2	1	1	1	2	1	1	1	1	1
rs1982904	127.47	2	2	2	2	2	2	2	2	2	2	2	2
rs291079	127.48	2	2	2	2	2	2	2	2	2	2	2	2
rs1106860	127.51	1	1	1	1	1	1	1	1	1	1	1	1
rs318539	127.80	2	2	1	2	2	2	1	2	2	2	1	2
rs1396758	128.53	2	2	2	1	2	2	2	2	2	2	2	2
rs2055392	129.15	2	2	2	2	2	2	2	2	2	2	2	2

Fig S-22: Family M324. Whole genome parametric (Merlin) and non-parametric (Merlin) linkage results (A) and haplotype of the only linkage interval with significant parametric LOD score of 3.3 (B).

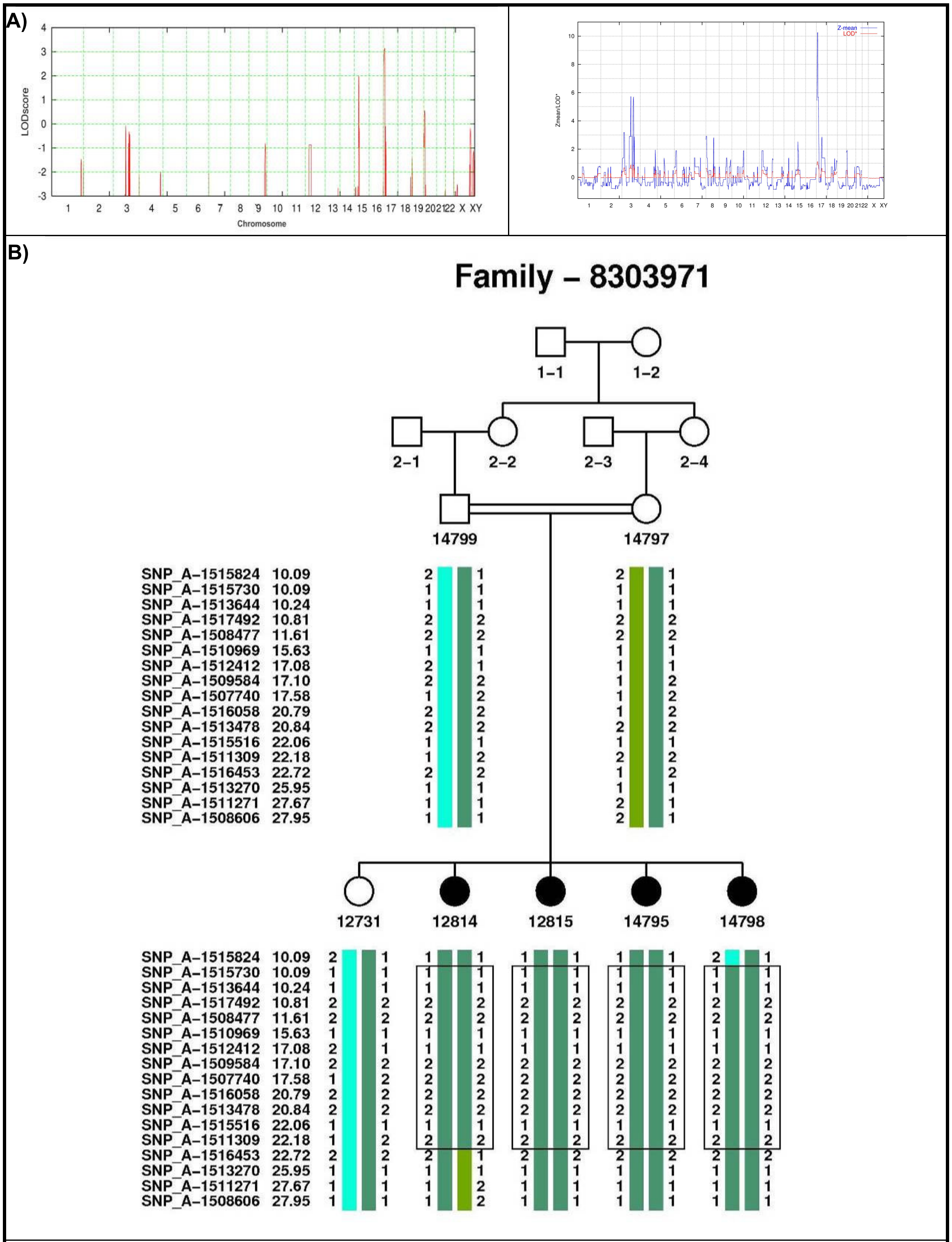


Fig S-23: Family 8303971. Whole genome parametric (Allegro) and non-parametric (Merlin) linkage results (A) and haplotype of the only linkage interval with significant parametric LOD score of 3.1 (B).

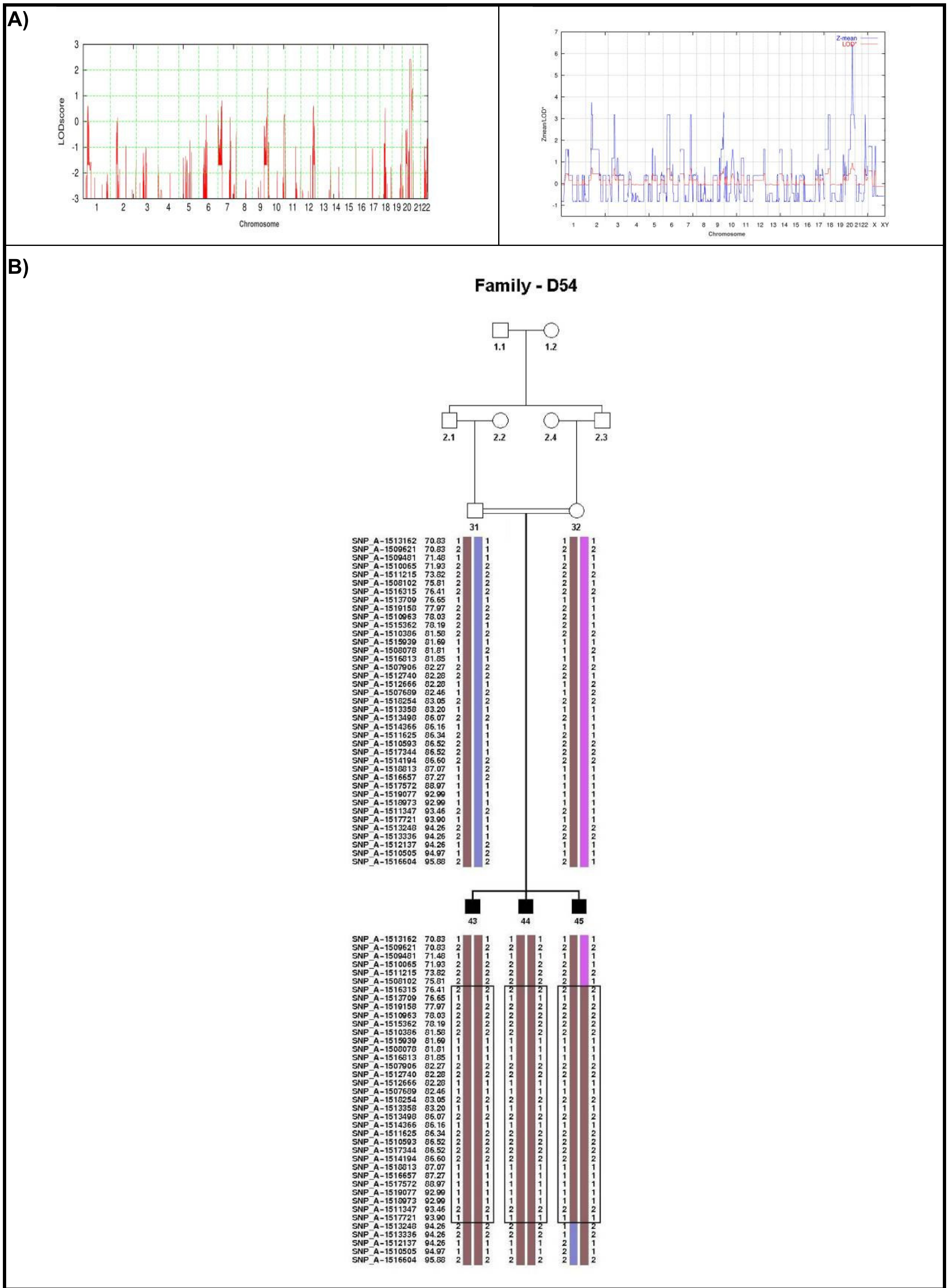


Fig S-24: Family D54. Whole genome parametric (GeneHunter) and non-parametric (Merlin) linkage results (A) and haplotype of the only linkage interval with significant parametric LOD score of 2.4 (B).

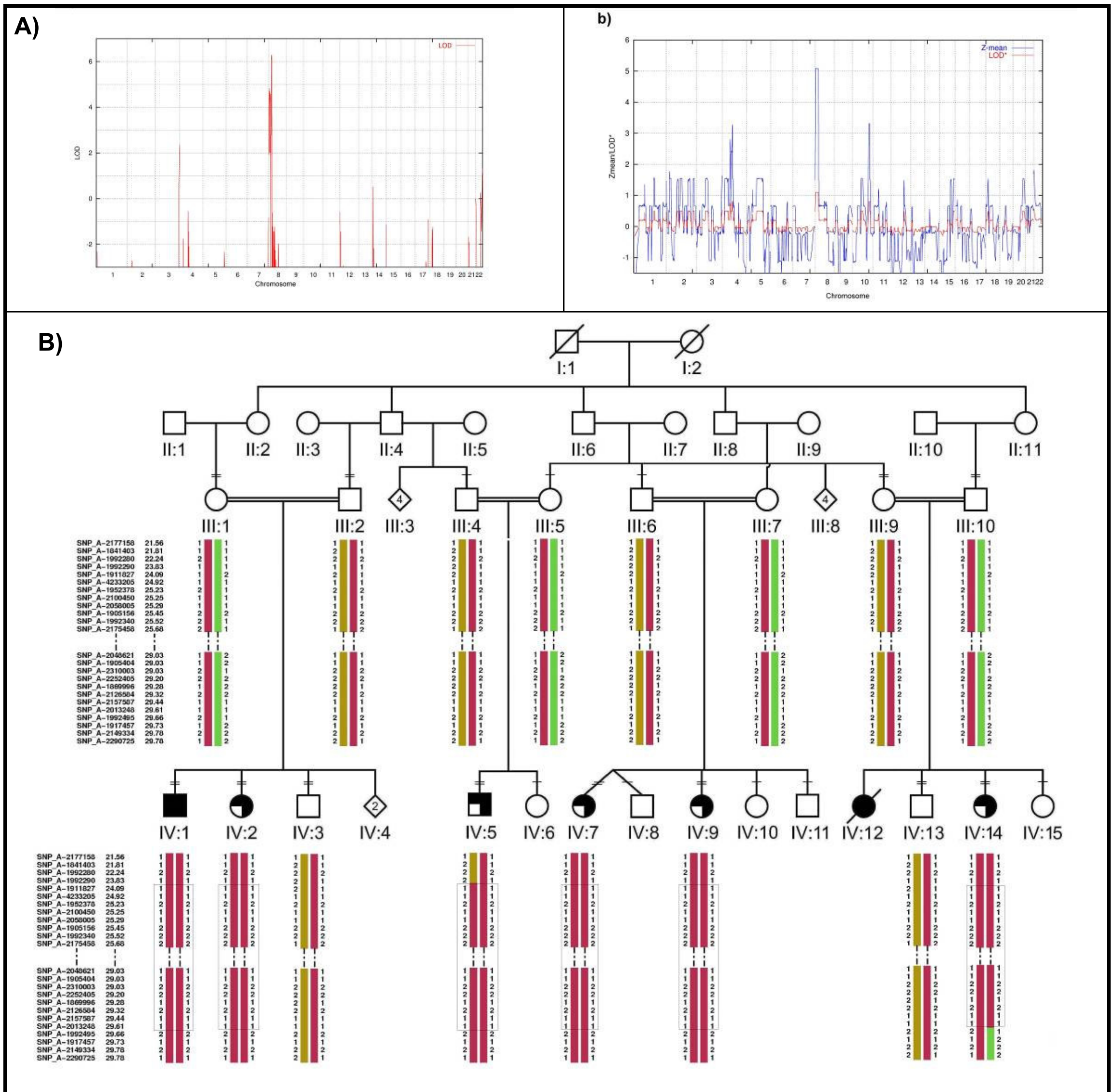


Fig S-25: Family 8404553: Whole genome parametric and non-parametric (Merlin) linkage results (A) and haplotype of the only linkage interval with significant parametric LOD score of 6.26. Due to the high number of markers just several markers in the both flanking sites of the interval are shown, and markers in between were excluded from the figure.

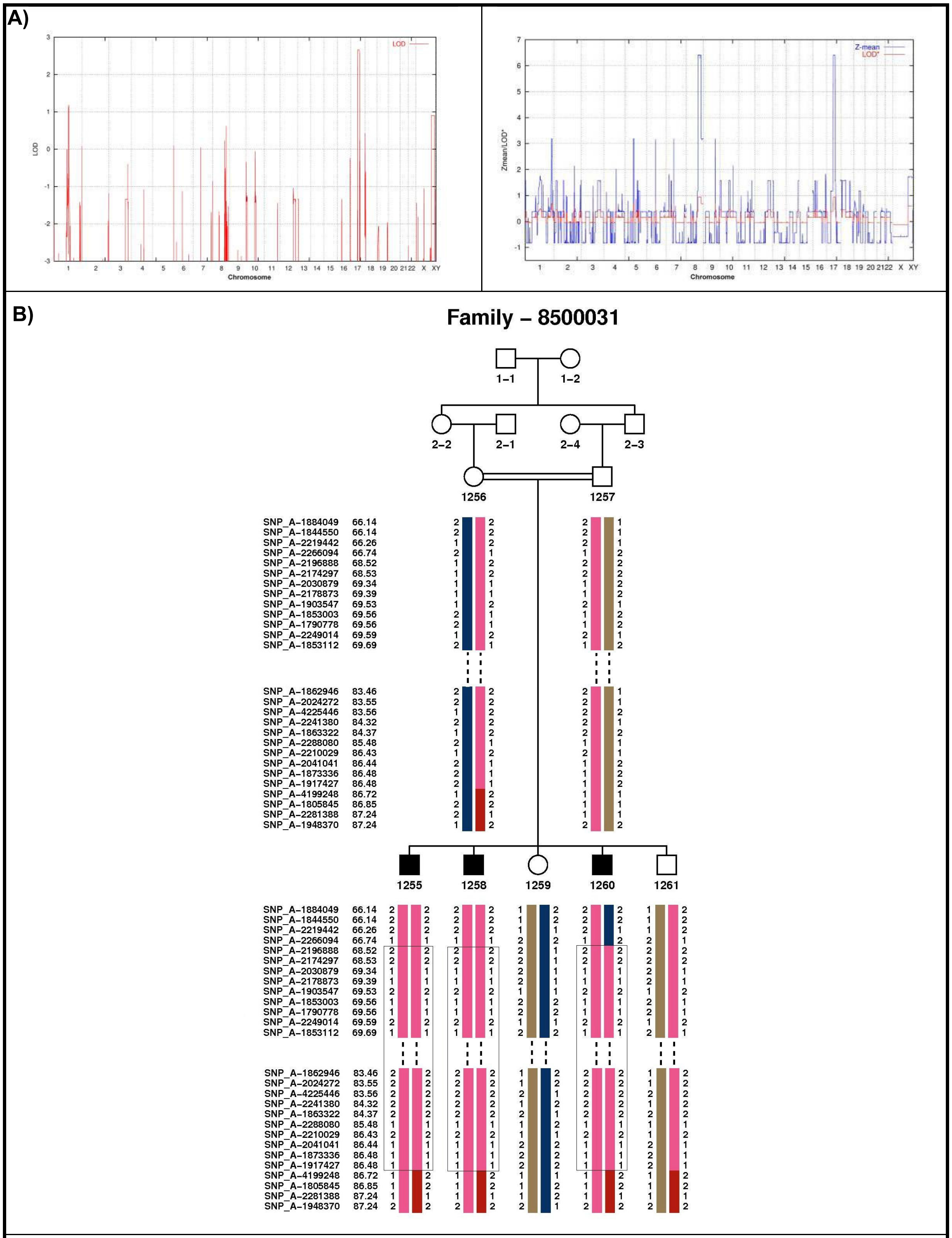


Fig S-26: Family 8500031. Whole genome parametric (Merlin) and non-parametric (Merlin) linkage results (A) and the haplotype of the only linkage interval with significant parametric LOD score of 2.65 (B). Due to the high number of markers just several markers in the both flanking sites of the interval are shown, and markers in between were excluded from the figure.

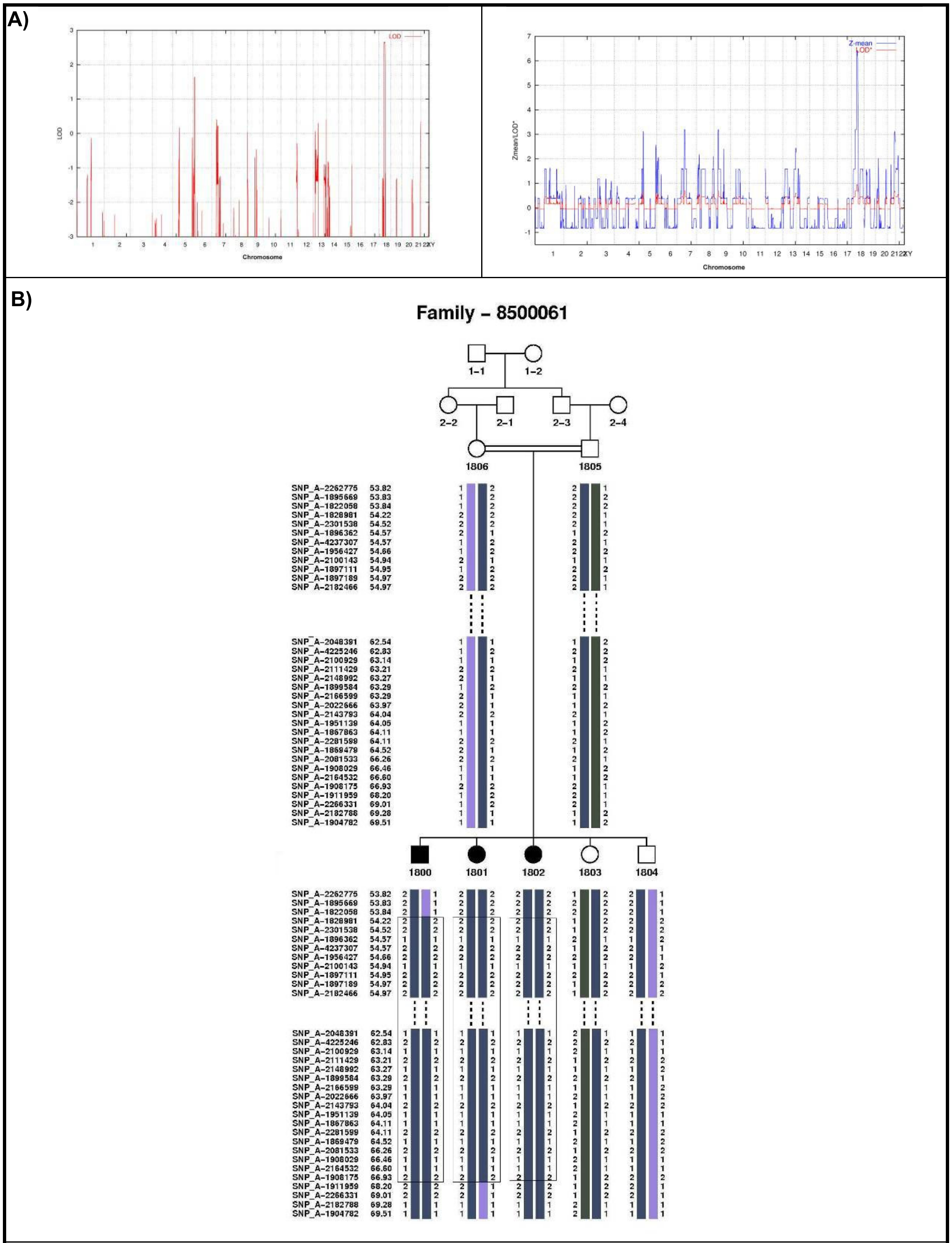


Fig S-27: Family 8500061. Whole genome parametric (Merlin) and non-parametric (Merlin) linkage results (A) and haplotype of the only linkage interval with significant parametric LOD score of 2.65 (B). Due to the high number of markers just several markers in the both flanking sites of the interval are shown, and markers in between were excluded from the figure.

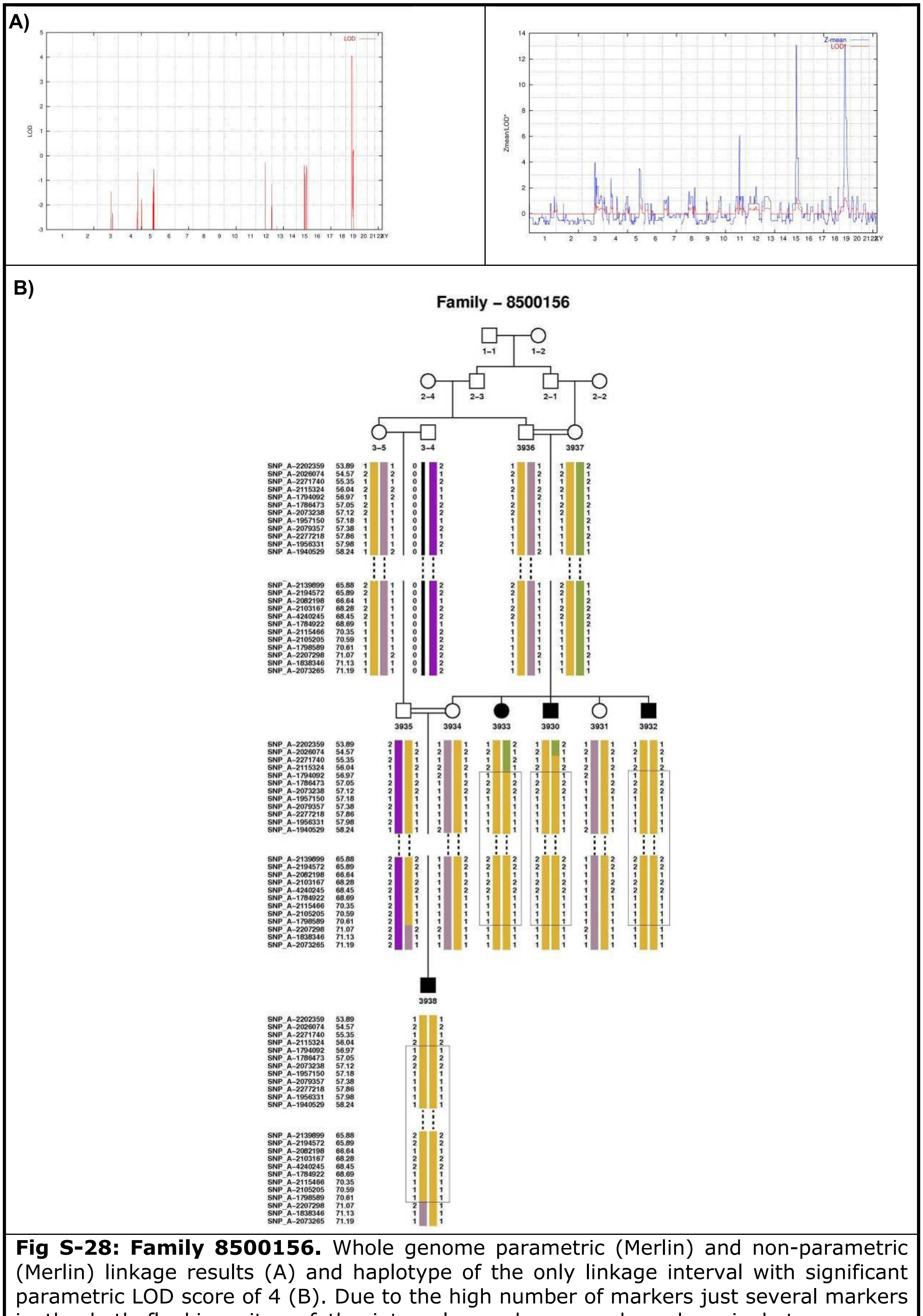


Fig S-28: Family 8500156. Whole genome parametric (Merlin) and non-parametric (Merlin) linkage results (A) and haplotype of the only linkage interval with significant parametric LOD score of 4 (B). Due to the high number of markers just several markers in the both flanking sites of the interval are shown, and markers in between were excluded from the figure.

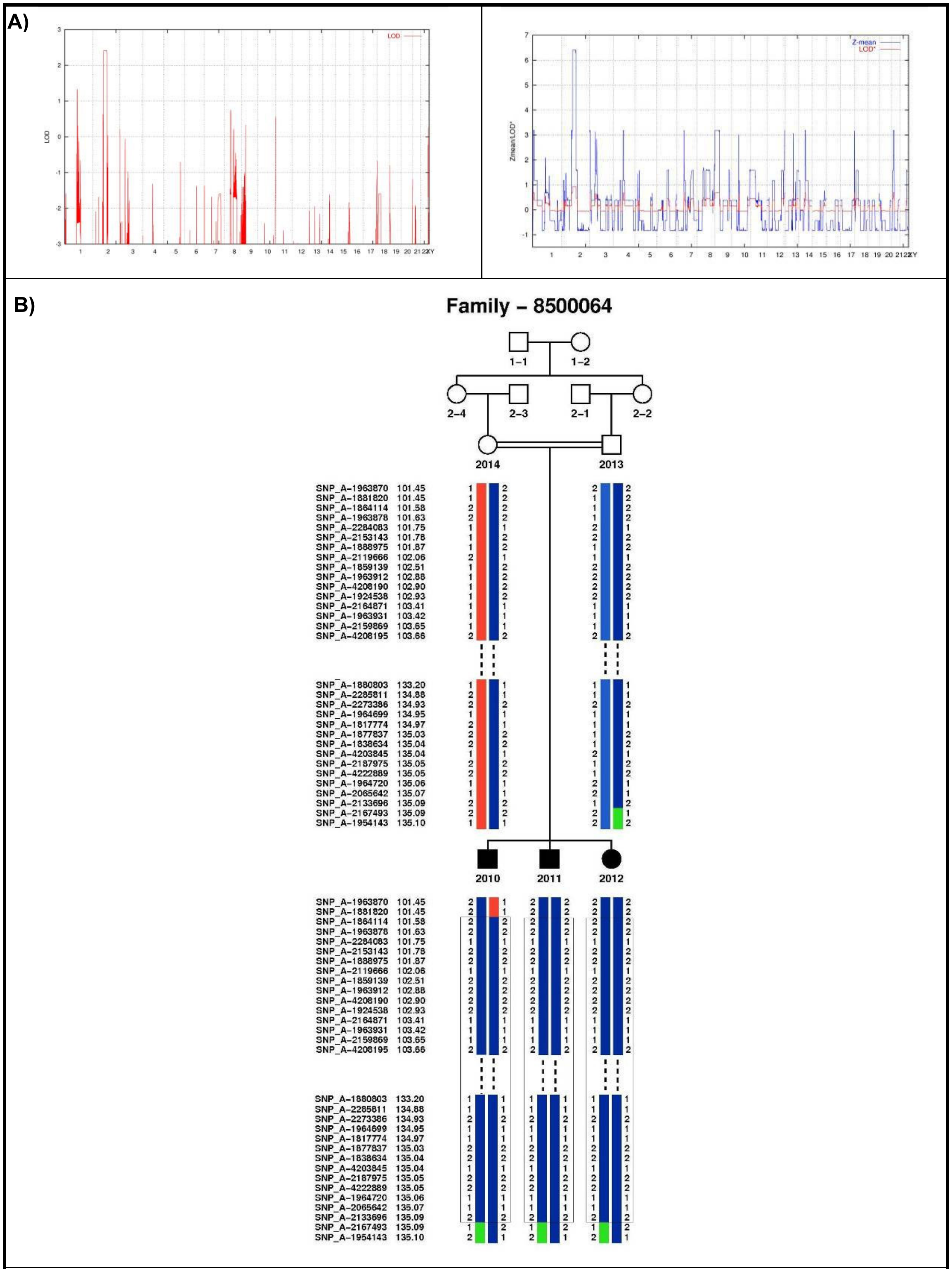


Fig S-29: Family 8500064. Whole genome parametric (Merlin) and non-parametric (Merlin) linkage results (A) and haplotype of the only linkage interval with significant parametric LOD score of 2.4 (B). Due to the high number of markers just several markers in the both flanking sites of the interval are shown, and markers in between were excluded from the figure.

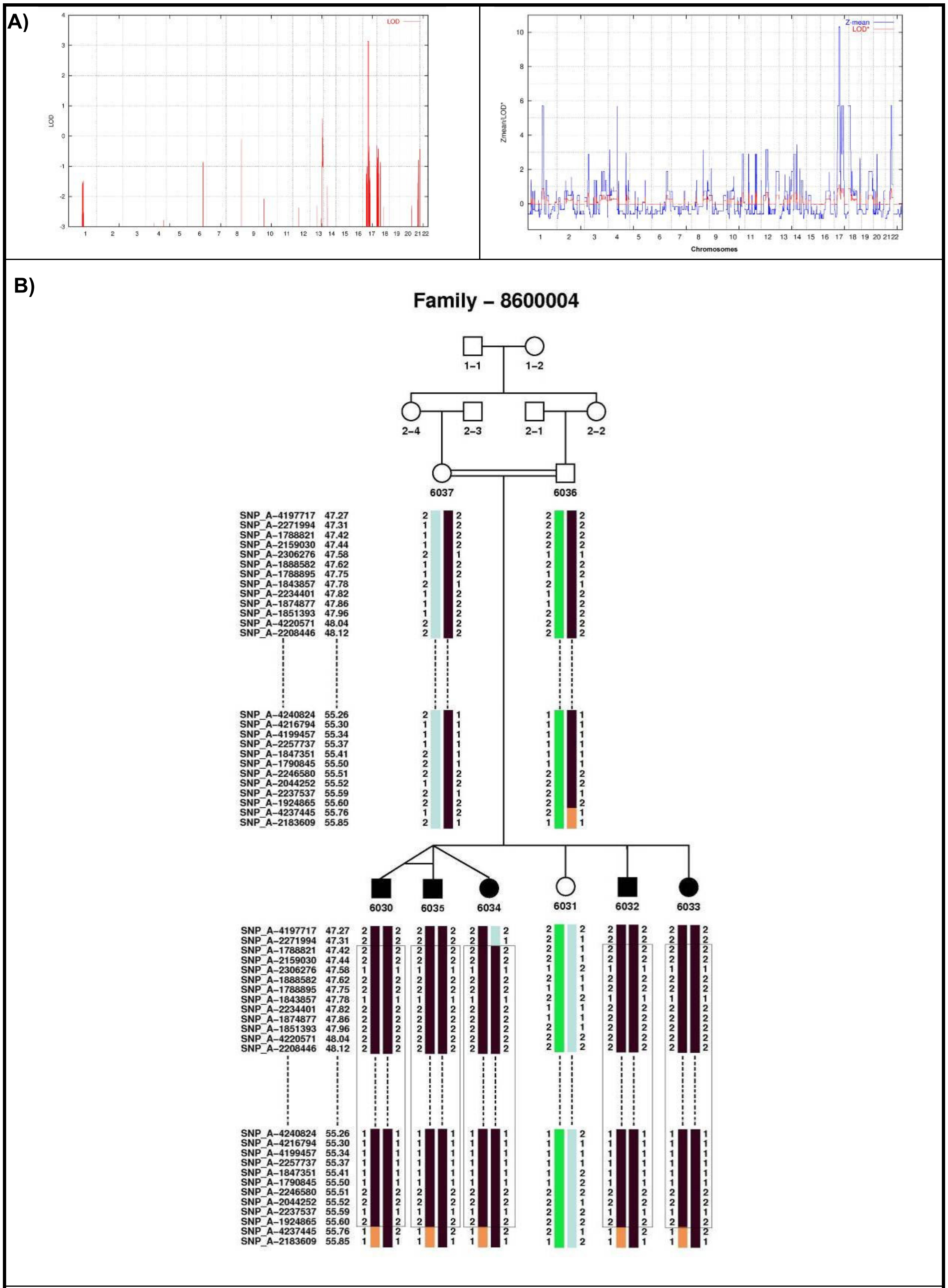


Fig S- 30: Family 8500004. Whole genome parametric (Merlin) and non-parametric (Merlin) linkage results (A) and haplotype of the only linkage interval with significant parametric LOD score of 3.13 (B). Due to the high number of markers just several markers in the both flanking sites of the interval are shown, and markers in between were excluded from the figure.

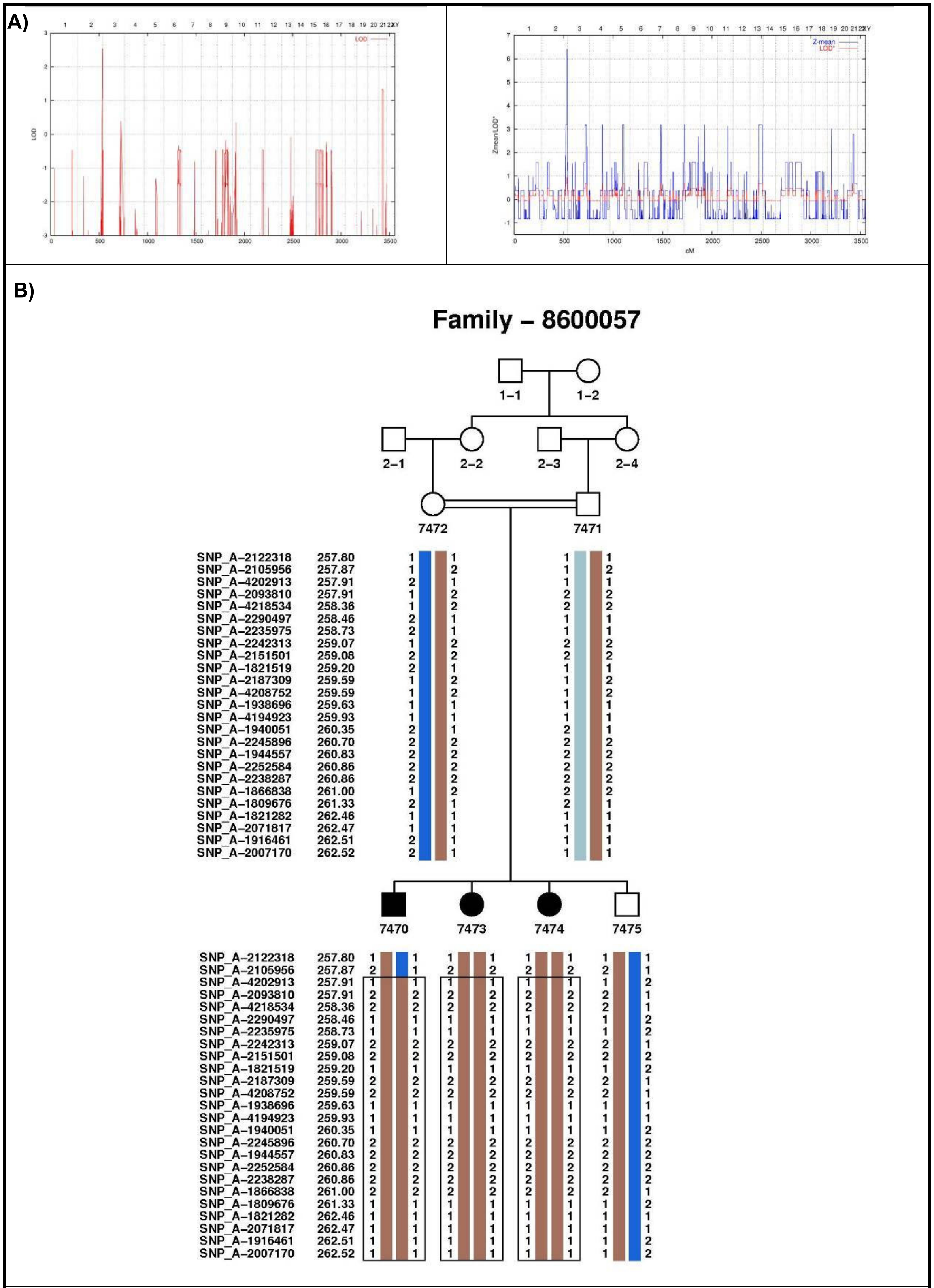


Fig S-31: Family 8500057. Whole genome parametric (Merlin) and non-parametric (Merlin) linkage results (A) and haplotype of the only linkage interval with significant parametric LOD score of 2.53 (B).

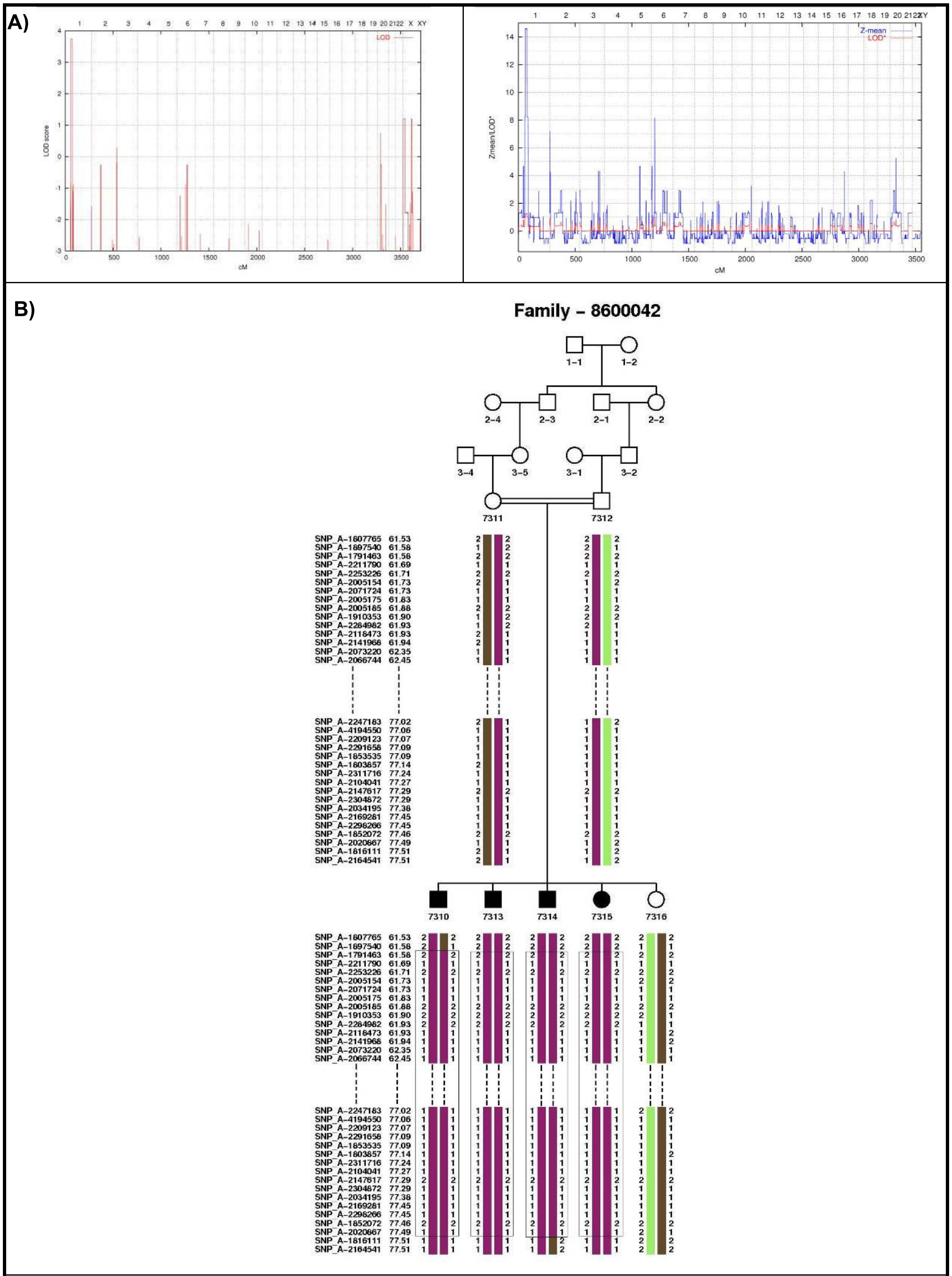


Fig S-32: Family 8500042. Whole genome parametric (Allegro) and non-parametric (Merlin) linkage results (A) and haplotype of the only linkage interval with significant parametric LOD score of 3.73 (B). Due to the high number of markers just several markers in the both flanking sites of the interval are shown, and markers in between were excluded from the figure.

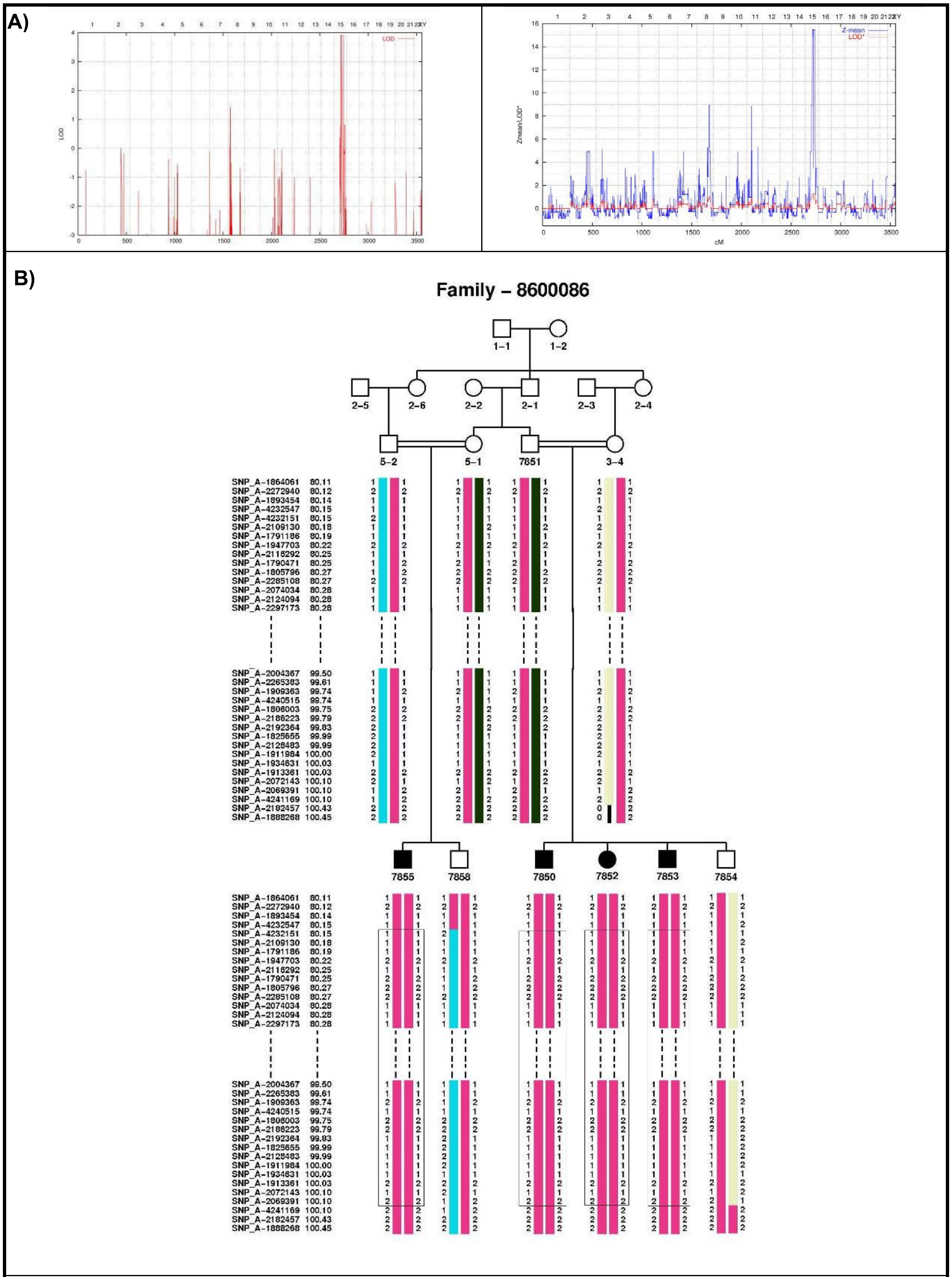


Fig S- 33: Family 8500086. Whole genome parametric (Merlin) and non-parametric (Merlin) linkage results (A) and haplotype of the only linkage interval with significant parametric LOD score of 3.9 (B). Due to the high number of markers just several markers in the both flanking sites of the interval are shown, and markers in between were excluded from the figure.

6.3 Appendix - C

Table S-2: TUSC3 RT-PCR primers

Primer name	Sequence	Size
MG4406_TUSC_RT_5+6_F	GGGTTTTCAGACCACCCAAC	225
MG4407_TUSC_RT_5+6_R	CTTGTCCATTGTGTGGGTTC	
MG4408_TUSC_RT_4+5+6_F	TTCCTCCAAAAGGCAGACC	300
MG4409_TUSC_RT_4+5+6_R	GTCCACGGATATGGTTCCAC	
MG4410_TUSC_RT_3+4+5_F	AGGGGACAGACGTTTTTCAG	238
MG4411_TUSC_RT_3+4+5_R	AAACCTCCAACAAGCGACAC	
MG4412_TUSC_RT_2+3+4_F	TAAAGGCACCACCTCGAAAC	247
MG4413_TUSC_RT_2+3+4_R	CTGCCTTTTGGAGGAAAATG	
MG4414_TUSC_RT_2+3_F	GACGCTCAATCTTCCGAATG	249
MG4415_TUSC_RT_2+3_R	TGTTGAGCTGCTGAAAAACG	
MG4416_TUSC_RT_9+10+11_F	GGCTATCCTTATAGTGATCTGGAC	300
MG4417_TUSC_RT_9+10+11_R	AACACTTTGATATTTCCTTTGTAGATT	
MG4418_TUSC_RT_1+2+3_F	ACCGGATGCTCTGTCAGTCT	220
MG4419_TUSC_RT_1+2+3_R	CTTGCCTACGGCGTGAAG	

6.4 Appendix - D

Table S-3: List of the primers for proving the deletion

Primer	Sequence (5′–3′)
170 bp between rs1057187 and rs1868551	
MG169_MCPH1_F	AGT GGG GTT CAG CAT GAG AG
MG170_MCPH1_R	AAA GTT CGA CCA CCT TGA TGA
MCPH1_exon 9_F	TTG CTT AAG TTG TAT TTG GTC CAT
MCPH1_exon 9_R	TTC ATT GAC CCA GAG AAG AAC A
MCPH1_exon 10_F	ACA GTT TAT TTC TGT GGG AAA AAT
MCPH1_exon 10_R	GCC TAA AGG CAC CCA GAA TTA
MCPH1_exon 11_F	GGC ATG TGC AAC AAA GTC AT
MCPH1_exon 11R	CCT CAG GGT GAC CCA CTC TA
MCPH1_exon 12_F	GCG GAG TGT ATC ACT TTT TGC
MCPH1_exon 12_R	GCA AAC TGC ATT TAC CAT CG
ANGPT2 exon specific	
ANGPT2_exon 9_F	GCA TGT GGT CCT TCC AAC TT
ANGPT2_exon 9_R	CTC AGG TGG ACT GGG ATG TT

6.5 Appendix – E

EGR2 (Early Growth Response 2)

Patients	Ex1_9del	S25X	T143NfsX5	W75R
Differentiation score	-91.84	-115.59	-115.95	-81.47

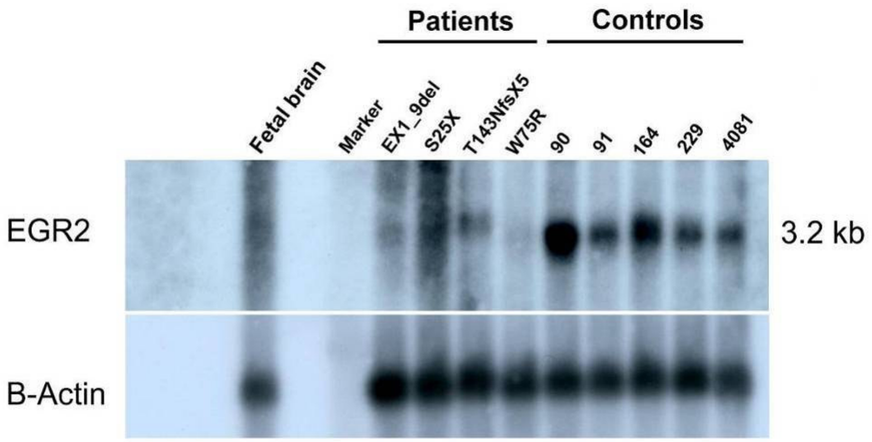


Fig S-34: Northern blot with EGR2 specific probe. The first lane contains RNA from fetal brain. The first four lanes after the marker lane belong to the four patients with different mutations in MCPH1 (Ex1_9del, S25X, T143NfsX5 and W75R). These are followed by the 5 controls. After stripping, the blot was re-probed with a beta-actin specific probe in order to control for sample loading. The diff. scores for the same patients in comparison to control on an Illumina array are presented in the table above. Compatible with the array data all of the patients showed strong downregulation for the only detected transcript with the approximate size of 3.2kbp.

LCK (Lymphocyte-specific protein tyrosine kinase)

Patients	Ex1_9del	S25X	T143NfsX5	W75R
Differentiation score	-39.08	-60.84	-36.12	-14.3

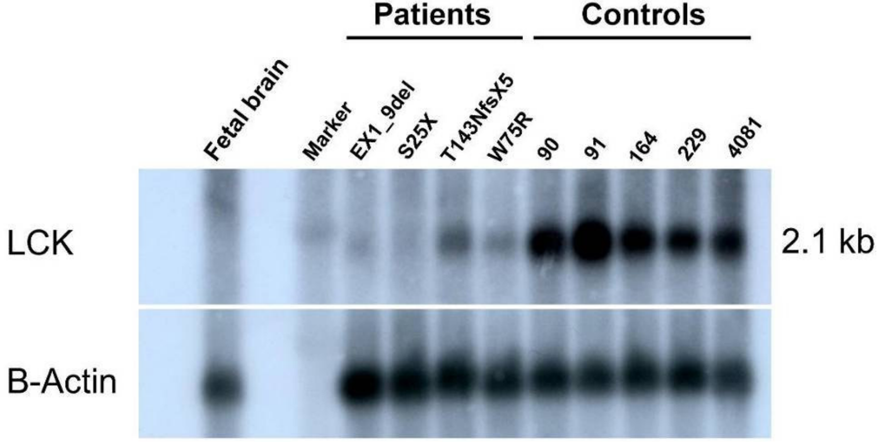


Fig S-35: Northern blot with *LCK* specific probe. The first lane contains RNA from fetal brain. The first four lanes after the marker lane belong to the four patients with different mutations in *MCPH1* (Ex1_9del, S25X, T143NfsX5 and W75R). These are followed by the 5 controls. After stripping, the blot was re-probed with a beta-actin specific probe in order to control for sample loading. The diff. scores for the same patients in comparison to control on an Illumina array are presented in the table above. Compatible with the array data all of the patients showed strong downregulation for the only detected transcript with the approximate size of 2.1kbp.

DUSP4 (Dual Specificity Phosphatase 4)

Patients	Ex1_9del	S25X	T143NfsX5	W75R
Differentiation score	-32.99	-38.57	-43.66	-50.6

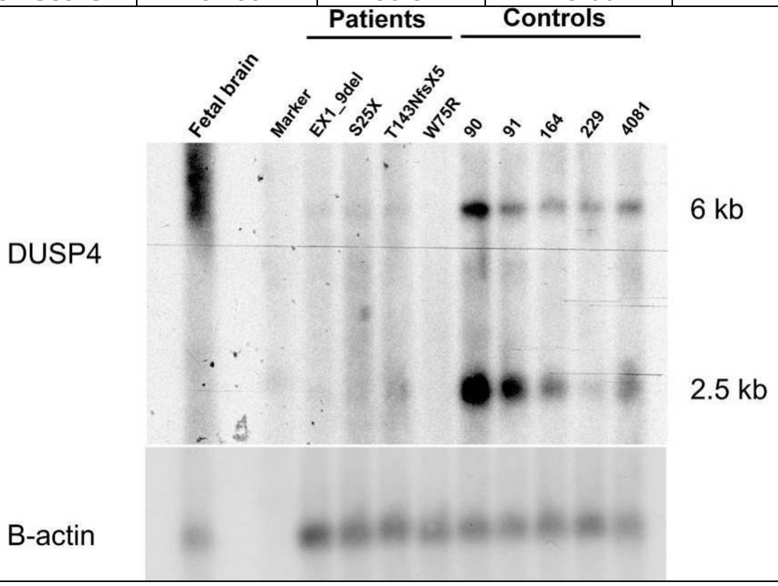


Fig S-36: Northern blot with *DUSP4* specific probe. The first lane contains RNA from fetal brain. The first four lanes after the marker lane belong to the four patients with different mutations in *MCPH1* (Ex1_9del, S25X, T143NfsX5 and W75R). These are followed by the 5 controls. After stripping, the blot was re-probed with a beta-actin specific probe in order to control for sample loading. The diff. scores for the same patients in comparison to control on an Illumina array are presented in the table above. Compatible with the array data all of the patients showed strong downregulation for the both detected transcripts with the approximate size of 2.5kbp and 6kbp.

PHGDH (Phosphoglycerate Dehydrogenase)

Patients	Ex1_9del	S25X	T143NfsX5	W75R	229
Differentiation score	-12.57	-10.59	-14.91	-14.28	-13.54

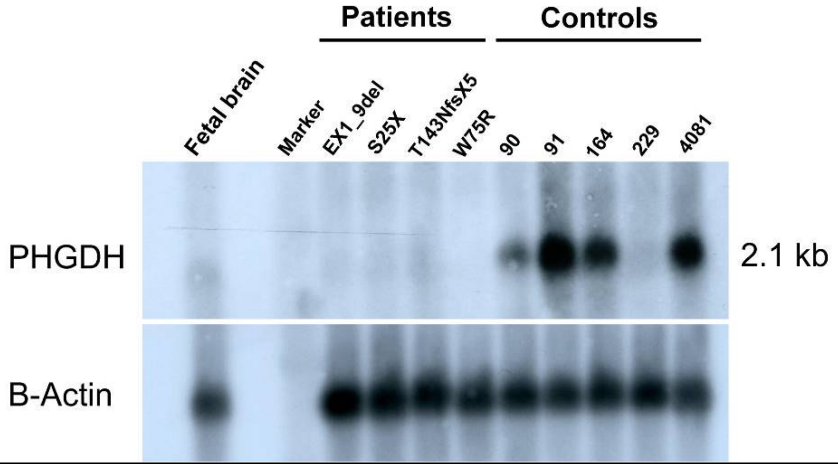


Fig S-37: Northern blot with *PHGDH*.specific probe. The first lane contains RNA from fetal brain. The first four lanes after the marker lane belong to the four patients with different mutations in *MCPH1* (Ex1_9del, S25X, T143NfsX5 and W75R). These are followed by the 5 controls. After stripping, the blot was re-probed with a beta-actin specific probe in order to control for sample loading. The diff. scores for the same patients in comparison to control on an Illumina array are presented in the table above. Compatible with the array data all of the patients and one of the controls (229) showed strong downregulation for the only detedcted transcript with the approximate size of 2.1kbp.

HK1 (Hexokinase 1 isoform HKI-td)				
Patients	**Ex1_9del**	**S25X**	**T143NfsX5**	**W75R**
Differentiation score	-35.79	-107.74	-76.53	-64.02

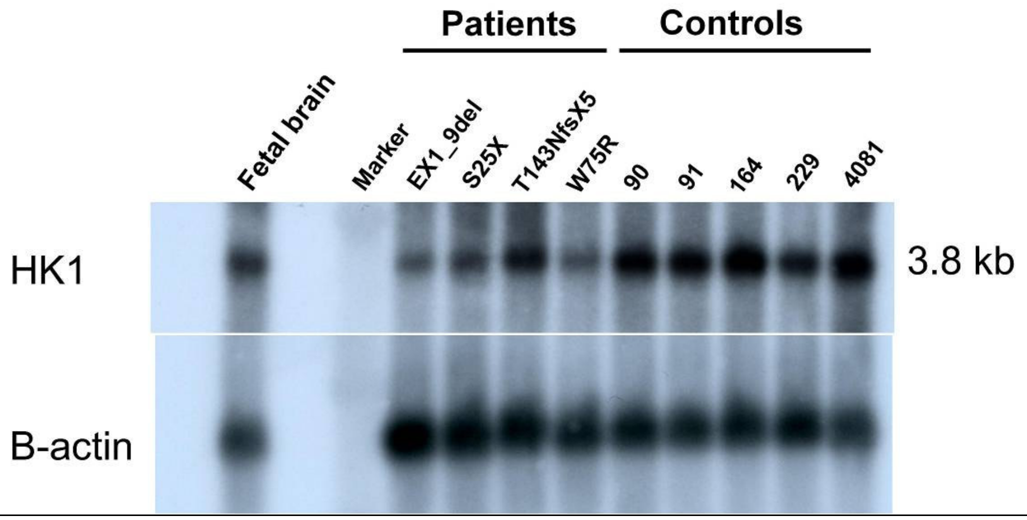


Fig S-38: Northern blot with *HK1* specific probe. The first lane contains RNA from fetal brain. The first four lanes after the marker lane belong to the four patients with different mutations in *MCPH1* (Ex1_9del, S25X, T143NfsX5 and W75R). These are followed by the 5 controls. After stripping, the blot was re-probed with a beta-actin specific probe in order to control for sample loading. The diff. scores for the same patients in comparison to control on an Illumina array are presented in the table above. Compatible with the array data all of the patients showed strong downregulation for the only detected transcript with the approximate size of 3.8kbp.

PSAT1 (Phosphoserine aminotransferase isoform 1)					
Patients	**Ex1_9del**	**S25X**	**T143NfsX5**	**W75R**	**229**
Differentiation score	-38.97	-27.48	-36.11	-58.18	-50.43

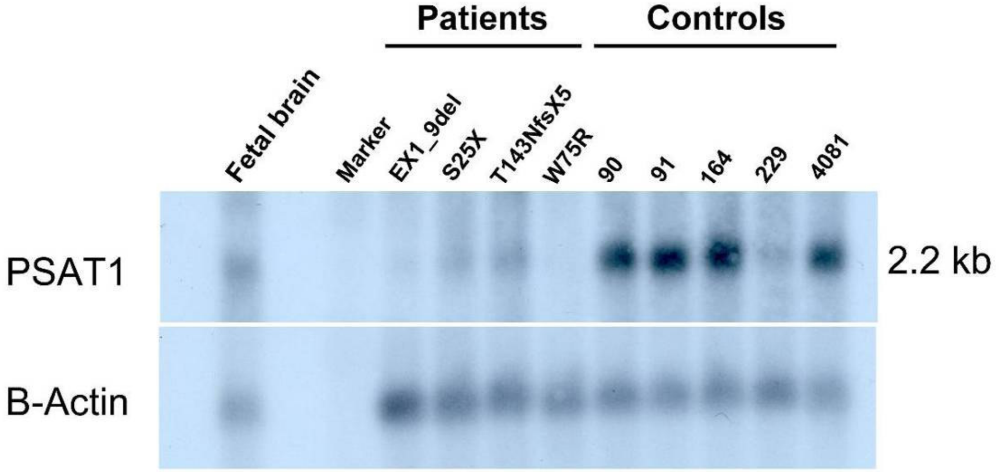


Fig S-39: Northern blot with PSAT1specific probe. The first lane contains RNA from fetal brain. The first four lanes after the marker lane belong to the four patients with different mutations in MCPH1 (Ex1_9del, S25X, T143NfsX5 and W75R). These are followed by the 5 controls. After stripping, the blot was re-probed with a beta-actin specific probe in order to control for sample loading. The diff. scores for the same patients in comparison to control on an Illumina array are presented in the table above. Compatible with the array data all of the patients and one of the controls (229) showed strong downregulation for the only detected transcript with the approximate size of 2.2kbp.

STAT1 (Signal transducer and activator of transcription)

Patients	Ex1_9del	S25X	T143NfsX5	W75R
Differentiation score	-24.41	-92.6	-198.32	-68.94

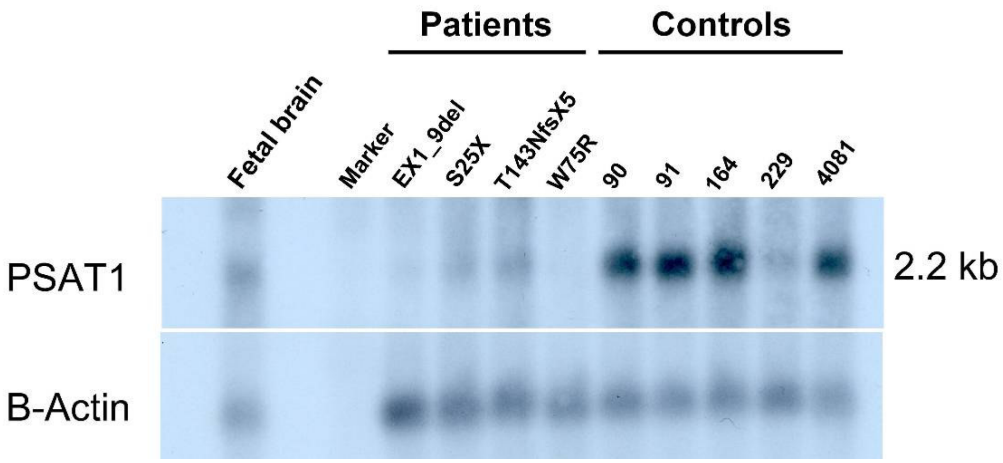


Fig S-40: Northern blot with *STAT1* specific probe. The first lane contains RNA from fetal brain. The first four lanes after the marker lane belong to the four patients with different mutations in *MCPH1* (Ex1_9del, S25X, T143NfsX5 and W75R). These are followed by the 5 controls. After stripping, the blot was re-probed with a beta-actin specific probe in order to control for sample loading. The diff. scores for the same patients in comparison to control on an Illumina array are presented in the table above. Compatible with the array data T143NfsX5 and W75R showed strong downregulation for both detected transcripts with the approximate sizes of 2.8 and 4.4kbp but S25X didn’t show downregulation as is expected by array data.

PLCG2 (Phospholipase C, gamma 2)

Patients	Ex1_9del	S25X	T143NfsX5	W75R
Differentiation score	-85.35	-128.4	-19.31	-36.67

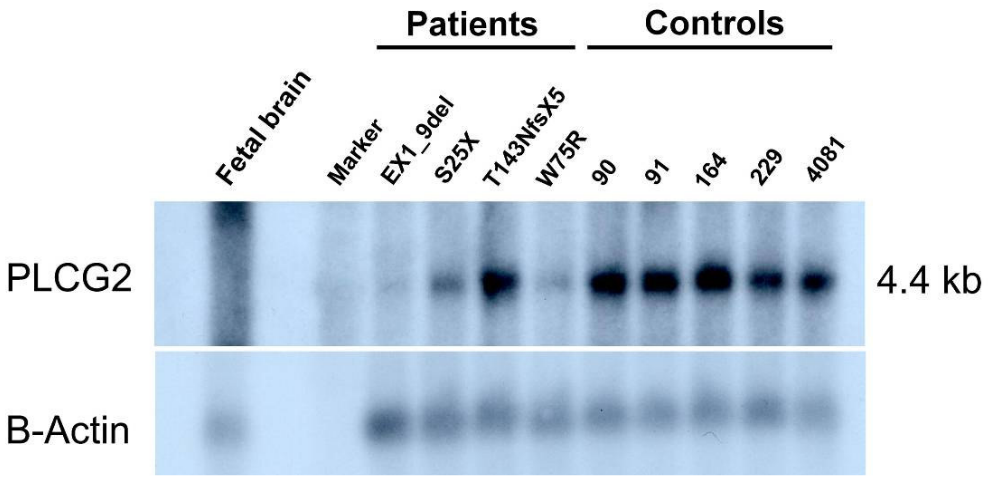


Fig S-41: Northern blot with *PLCG2* specific probe. The first lane contains RNA from fetal brain. The first four lanes after the marker lane belong to the four patients with different mutations in *MCPH1* (Ex1_9del, S25X, T143NfsX5 and W75R). These are followed by the 5 controls. After stripping, the blot was re-probed with a beta-actin specific probe in order to control for sample loading. The diff. scores for the same patients in comparison to control on an Illumina array are presented in the table above. Compatible with the array data 3 of the patients (Ex1_9del, S25X, T143NfsX5 and W75R) showed strong downregulation for the only detected transcript with the approximate size of 4.4kbp.

PTEN (Phosphatase and tensin homolog)

Patients	Ex1_9del	S25X	T143NfsX5	W75R
Differentiation score	-20.95	-21.54	-64.68	-18.88

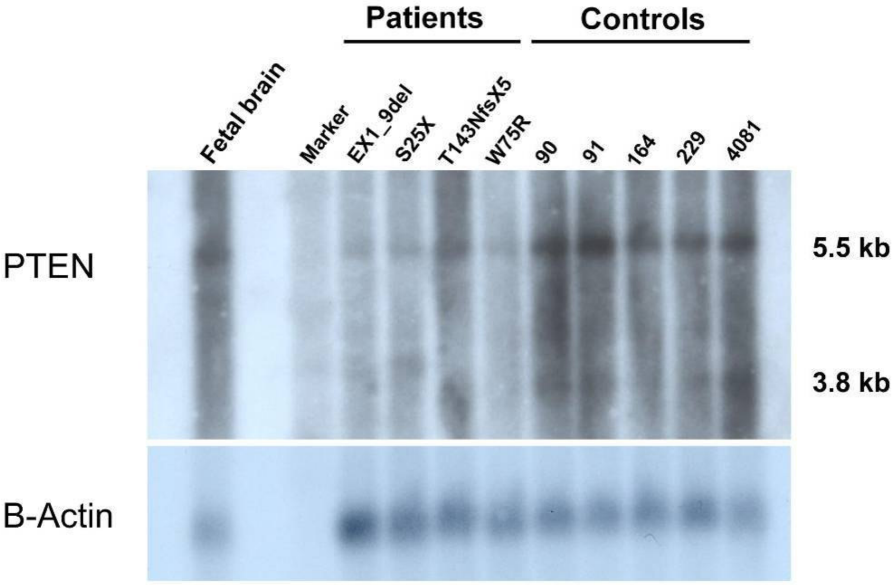


Fig S-42: Northern blot with *PTEN*.specific probe. The first lane contains RNA from fetal brain. The first four lanes after the marker lane belong to the four patients with different mutations in *MCPH1* (Ex1_9del, S25X, T143NfsX5 and W75R). These are followed by the 5 controls. After stripping, the blot was re-probed with a beta-actin specific probe in order to control for sample loading. The diff. scores for the same patients in comparison to control on an Illumina array are presented in the table above. Compatible with the array data all of the patients showed strong downregulation for both transcripts with the approximate sizes of 3.8kbp and 5.5kbp.

NK4 (Natural Killer cell transcript 4)

Patients	Ex1_9del	S25X	T143NfsX5	W75R
Differentiation score	-8.00	-15.12	-17.57	-14.2

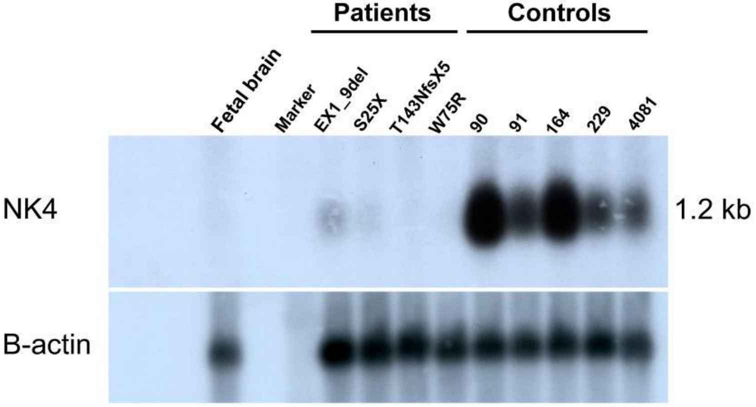


Fig S-43: Northern blot with NK4 specific probe. The First lane contains RNA from fetal brain. The first four lanes after the marker lane belong to the four patients with different mutations in MCPH1 (Ex1_9del, S25X, T143NfsX5 and W75R). These are followed by the 5 controls. After stripping, the blot was re-probed with a beta-actin specific probe in order to control for sample loading. The diff. scores for the same patients in comparison to control on an Illumina array are presented in the table above. Compatible with the array data all of the patients showed strong downregulation for the only transcript with the approximate size of 1.2kbp.

ANXA11 (Annexin A11)

Patients	Ex1_9del	S25X	T143NfsX5	W75R	229
Differentiation score	-37.35	-7.85	-58.16	-46.61	-8.5

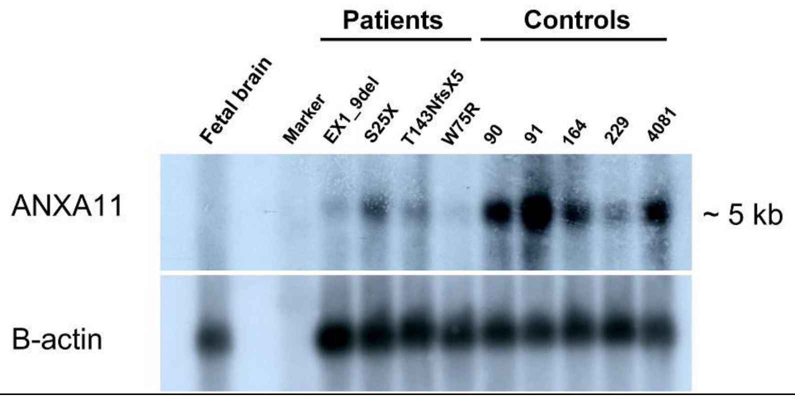


Fig S-44: Northern blot with ANXA11 specific probe. The first lane contains RNA from fetal brain. The first four lanes after the marker lane belong to the four patients with different mutations in *MCPH1* (Ex1_9del, S25X, T143NfsX5 and W75R). These are followed by the 5 controls. After stripping, the blot was re-probed with a beta-actin specific probe in order to control for sample loading. The diff. scores for the same patients in comparison to control on an Illumina array are presented in the table above. Compatible with the array data all of the patients and one of the controls (229) showed strong downregulation for the only detected transcript with the approximate size of 5kbp.

6.6 Appendix - F

FLJ31978

Patients	Ex1_9del	S25X	T143NfsX5	W75R
Differentiation score	371.33	371.33	-24.34	371.33

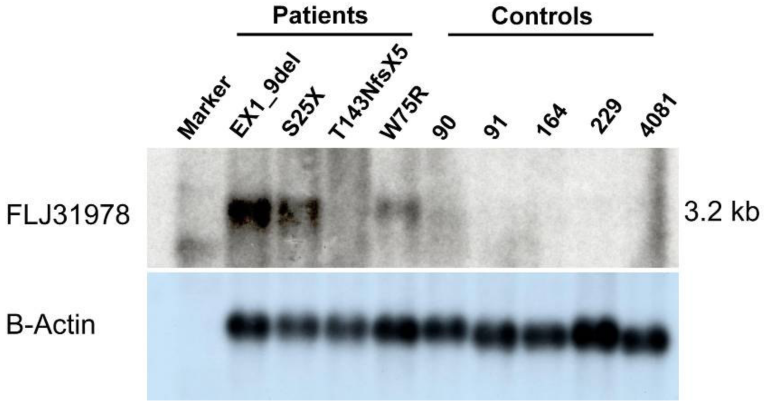


Fig S-45: Northern blot with *FLJ31978* specific probe. The first lane contains RNA from fetal brain. The first four lanes after the marker lane belong to the four patients with different mutations in *MCPH1* (Ex1_9del, S25X, T143NfsX5 and W75R). These are followed by the 5 controls. After stripping, the blot was re-probed with a beta-actin specific probe in order to control for sample loading. The diff. scores for the same patients in comparison to control on an Illumina array are presented in the table above. Compatible with the array data 3 of the patients (Ex1_9del, S25X and W75R) showed strong upregulation for the only transcript with the approximate size of 3.2kbp.

6.7 Appendix - G

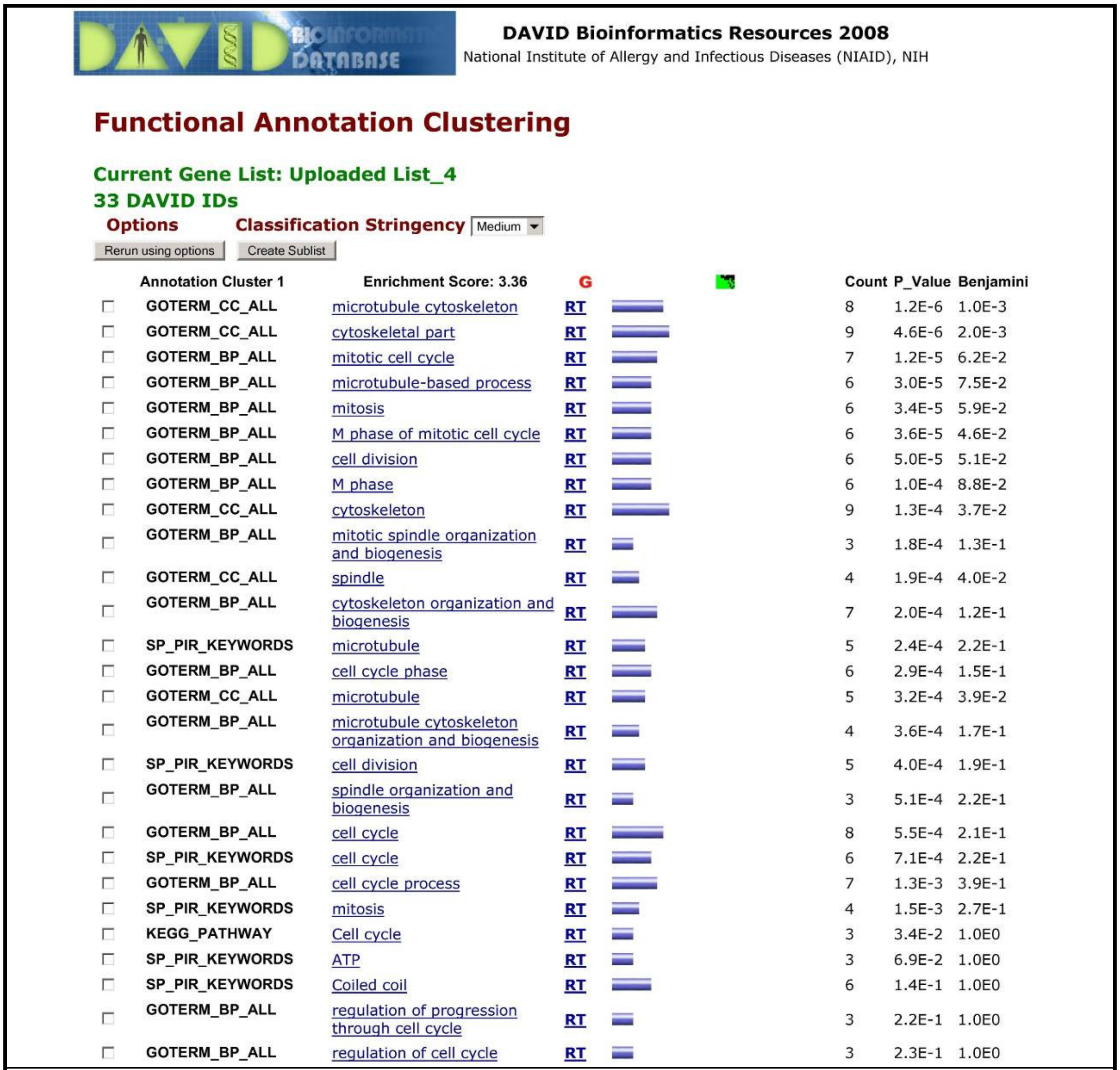
Table S-4: List of all the genes with diff. scores ≤ –30 (corresponding to P-values ≤ 0.001) for LCLs of 8 patients with 5 different *MCPH1* mutations (EX1_9 del, S25X, T143NfsX5, W75R and T27R) as one group in comparison with a group of 9 controls.

Symbol	Signal_X	Signal_Y	Detection_X	Detection_Y	Diff. Score	Accession
EGR2	399.4	117	1	1	**-100.0617**	NM_000399.2
NFATC1	314.6	161.1	1	1	**-65.6367**	NM_172390.1
GFI1	95.6	38.5	1	0.998	**-61.363**	NM_005263.1
HK1	1504.1	931.6	1	1	**-59.7894**	NM_000188.1
KIAA0830	278.1	110.9	1	1	**-58.0433**	XM_290546.1
MASTL	288.4	118.2	1	1	**-57.1399**	NM_032844.1
PLCG2	1056.2	649.9	1	1	**-56.7128**	NM_002661.1
BIK	179.5	48.2	1	0.9987	**-54.6041**	NM_001197.3
ISG20	3401	2016.4	1	1	**-53.1294**	NM_002201.4
CDC2	261	120	1	1	**-53.0125**	NM_001786.2
WEE1	154	86.1	1	0.9993	**-50.3392**	NM_003390.2
TLR9	60.7	21.7	0.9993	0.9888	**-50.1989**	NM_138688.1
JFC1	203.3	107.8	1	1	**-48.6705**	NM_032872.1
SEC24D	558.1	279.5	1	1	**-47.0721**	NM_014822.1
LMO4	467.3	252.9	1	1	**-46.4252**	NM_006769.2
BAG2	222.2	137	1	1	**-44.4926**	NM_004282.2
LLT1	1244.3	439.5	1	1	**-44.2712**	NM_013269.1
OGG1	90.5	45.9	1	0.998	**-43.8788**	NM_016827.1
DKFZp434K1210	937.1	532.2	1	1	**-43.4201**	NM_017606.2
NCOA1	80.2	31.3	1	0.9974	**-42.6961**	NM_147223.1
PSMB8	786.7	500.4	1	1	**-42.4316**	NM_148919.2
CCNA2	367.4	137	1	1	**-42.404**	NM_001237.2
RFC3	104.7	46.5	1	0.9987	**-42.0286**	NM_181558.1
UMPK	289.3	186	1	1	**-41.8724**	NM_012474.3
LOC161577	87.2	28	1	0.9967	**-41.5563**	NM_198524.1
BUB1	554.8	297.5	1	1	**-41.3903**	NM_004336.1
DKFZP727G051	189.7	99.5	1	0.9993	**-41.0399**	XM_045308.6
SQLE	507.5	320	1	1	**-40.891**	NM_003129.2
SMC2L1	138.4	65.6	1	0.9993	**-40.4442**	NM_006444.1
LAP3	1385.7	887	1	1	**-39.9151**	NM_015907.2
UBE2J1	1259.1	823.3	1	1	**-39.7225**	NM_016336.2
AURKB	1717.3	1033.1	1	1	**-39.4817**	NM_004217.1
TOPK	370.1	194.5	1	1	**-39.3099**	NM_018492.2
CDC2	710.3	404.4	1	1	**-39.0595**	NM_033379.2
BZRP	331.9	156.7	1	1	**-38.8465**	NM_000714.3
SWAP70	560	271.3	1	1	**-37.7632**	NM_015055.1
LOC126208	120.6	66.9	1	0.9993	**-37.3885**	XM_058999.7
LOC389386	143.6	87.3	1	0.9993	**-37.3485**	XM_371818.1
C8FW	335.4	145.8	1	1	**-36.6834**	NM_025195.2
LYAR	317	202.2	1	1	**-36.4625**	NM_017816.1
XTP1	95.4	38.9	1	0.998	**-36.3081**	NM_018369.1
RGS20	356.6	164.3	1	1	**-36.1538**	NM_003702.2
LOC159090	318.3	214.4	1	1	**-35.7591**	NM_145284.3
EAF2	189.1	98.2	1	0.9993	**-35.6546**	NM_018456.4
ALAS1	340.7	223	1	1	**-35.6404**	NM_000688.4
NCOA1	90.3	48.6	1	0.9987	**-35.5586**	NM_147233.1
BLNK	200.4	119.7	1	1	**-35.3023**	NM_013314.2
GMDS	603.3	412	1	1	**-35.1742**	NM_001500.2
CENPA	238.6	137.7	1	1	**-35.1217**	NM_001809.2
MGC29814	809	567.3	1	1	**-34.9124**	NM_182565.2
PLK4	307	164.1	1	1	**-34.6664**	NM_014264.2
ORC1L	140.4	73.6	1	0.9993	**-34.3205**	NM_004153.2
PECI	357.7	191.2	1	1	**-34.1978**	NM_006117.2
ALDOA	5368.3	3848.3	1	1	**-33.8529**	NM_184041.1
CAST	204.7	126.2	1	1	**-33.1422**	NM_173061.1
AKR1B1	762.9	536.7	1	1	**-32.8416**	NM_001628.2
CDCA2	129	68	1	0.9993	**-32.814**	XM_351774.1
ITGA4	74.3	31.8	1	0.9967	**-32.6629**	NM_000885.2
GFPT1	303.2	208	1	1	**-32.6027**	NM_002056.1
KIF11	238.3	122.7	1	1	**-32.3268**	NM_004523.2
CDCA1	120.3	59.1	1	0.9993	**-32.0748**	NM_031423.2
C10orf3	464.4	281.2	1	1	**-32.0136**	NM_018131.3
ZNF288	162.5	73	1	0.9993	**-31.6259**	NM_015642.1
DKFZp762E1312	285.2	162.4	1	1	**-31.4425**	NM_018410.2

Table S-4: List of all the genes with diff. scores ≤ –30 (corresponding to P-values ≤ 0.001) for LCLs of 8 patients with 5 different *MCPH1* mutations (EX1_9 del, S25X, T143NfsX5, W75R and T27R) as one group in comparison with a group of 9 controls.

Symbol	Signal_X	Signal_Y	Detection_X	Detection_Y	Diff. Score	Accession
MELK	240.4	140.1	1	1	**-31.1405**	NM_014791.2
EPB41L2	143.7	58.8	1	0.9993	**-31.1199**	NM_001431.1
KIAA0342	122.7	70.8	1	0.9993	**-31.096**	XM_047357.4
ARL5	318.4	208	1	1	**-31.038**	NM_177985.1
MAD2L1	435.8	257.1	1	1	**-30.9747**	NM_002358.2
OSR2	96.4	43.5	1	0.998	**-30.8836**	NM_053001.1
LOC116228	189.9	125.5	1	1	**-30.7826**	NM_198076.2
PCMT1	963.1	695.5	1	1	**-30.6418**	NM_005389.1
IFI44	1554.3	855.9	1	1	**-30.3158**	NM_006417.2
LPXN	3738.2	2646.4	1	1	**-30.2883**	NM_004811.1
BAL	451.8	236.7	1	1	**-30.2662**	NM_031458.1
CHEK2	145.1	93.7	1	0.9993	**-30.1159**	NM_145862.1
BZRP	3012.2	1805	1	1	**-30.0741**	NM_007311.2
BM039	359.7	209.9	1	1	**-30.0167**	NM_018455.3
PPIL5	291.9	172.6	1	1	**-30.0031**	NM_152329.3

6.8 Appedix – H



DAVID Bioinformatics Resources 2008
National Institute of Allergy and Infectious Diseases (NIAID), NIH

Functional Annotation Clustering

Current Gene List: Uploaded List_4
33 DAVID IDs

Options **Classification Stringency** Medium

	Annotation Cluster 1	Enrichment Score: 3.36	G	Count	P_Value	Benjamini
□	GOTERM_CC_ALL	microtubule cytoskeleton	RT	8	1.2E-6	1.0E-3
□	GOTERM_CC_ALL	cytoskeletal part	RT	9	4.6E-6	2.0E-3
□	GOTERM_BP_ALL	mitotic cell cycle	RT	7	1.2E-5	6.2E-2
□	GOTERM_BP_ALL	microtubule-based process	RT	6	3.0E-5	7.5E-2
□	GOTERM_BP_ALL	mitosis	RT	6	3.4E-5	5.9E-2
□	GOTERM_BP_ALL	M phase of mitotic cell cycle	RT	6	3.6E-5	4.6E-2
□	GOTERM_BP_ALL	cell division	RT	6	5.0E-5	5.1E-2
□	GOTERM_BP_ALL	M phase	RT	6	1.0E-4	8.8E-2
□	GOTERM_CC_ALL	cytoskeleton	RT	9	1.3E-4	3.7E-2
□	GOTERM_BP_ALL	mitotic spindle organization and biogenesis	RT	3	1.8E-4	1.3E-1
□	GOTERM_CC_ALL	spindle	RT	4	1.9E-4	4.0E-2
□	GOTERM_BP_ALL	cytoskeleton organization and biogenesis	RT	7	2.0E-4	1.2E-1
□	SP_PIR_KEYWORDS	microtubule	RT	5	2.4E-4	2.2E-1
□	GOTERM_BP_ALL	cell cycle phase	RT	6	2.9E-4	1.5E-1
□	GOTERM_CC_ALL	microtubule	RT	5	3.2E-4	3.9E-2
□	GOTERM_BP_ALL	microtubule cytoskeleton organization and biogenesis	RT	4	3.6E-4	1.7E-1
□	SP_PIR_KEYWORDS	cell division	RT	5	4.0E-4	1.9E-1
□	GOTERM_BP_ALL	spindle organization and biogenesis	RT	3	5.1E-4	2.2E-1
□	GOTERM_BP_ALL	cell cycle	RT	8	5.5E-4	2.1E-1
□	SP_PIR_KEYWORDS	cell cycle	RT	6	7.1E-4	2.2E-1
□	GOTERM_BP_ALL	cell cycle process	RT	7	1.3E-3	3.9E-1
□	SP_PIR_KEYWORDS	mitosis	RT	4	1.5E-3	2.7E-1
□	KEGG_PATHWAY	Cell cycle	RT	3	3.4E-2	1.0E0
□	SP_PIR_KEYWORDS	ATP	RT	3	6.9E-2	1.0E0
□	SP_PIR_KEYWORDS	Coiled coil	RT	6	1.4E-1	1.0E0
□	GOTERM_BP_ALL	regulation of progression through cell cycle	RT	3	2.2E-1	1.0E0
□	GOTERM_BP_ALL	regulation of cell cycle	RT	3	2.3E-1	1.0E0

Fig S-46: DAVID functional annotation clustering for common downregulated genes with diff. scores ≤–13 (corresponding to P-values ≤0.05) between LCLs of patients with *MCPH1* mutations and U2OS *MCPH1* depleted cells at 72 hours after transfection. The most siginificant P-values (P_value and Benjamini) yealded by DAVID for downregulated genes classified to different annotation terms based on different databases are shown in the last two columns.

6.9 Appendix - I

Table S-5: Expression level of previously known interacting partners of MCPH1 in MCPH1 RNAi depleted cells

Gene symbol	Accession No.	After 48 hr	After 72 hr
CHEK1	NM_001274.2	-5.1243	-8.6398
CHEK2	NM_145862.1	-2.488	-2.862
CHEK2	NM_007194.2	1.6634	0.3965
BRCA1	NM_007295.1	-2.6477	2.2168
BRCA1	NM_007298.1	-4.3937	-2.9404
BRCA1	NM_007306.1	-23.1246	-15.29
BRCA2	NM_000059.1	-3.133	-2.8497
TOPBP1	NM_007027.2	-3.0549	-11.425
RAD51	NM_002875.2	-3.738	-4.4059
RAD51	NM_133487.1	-1.7736	-2.2917
DDB2	NM_000107.1	-0.7801	-1.2231
TP73	NM_005427.1	0.7234	0.6119
E2F1	NM_005225.1	-1.0804	-4.6255
E2F2	NM_004091.2	-20.2245	-39.849
E2F3	NM_001949.2	0.4624	3.222
E2F4	NM_001950.3	-3.4169	-6.9807
E2F5	NM_001951.2	4.1532	11.354
E2F6	NM_198257.1	10.627	6.7327
E2F7	NM_203394.1	5.9688	0.5232
E2F8	NM_024680.2	0.3417	-6.0162
APAF1	NM_181869.1	-2.134	-6.7957
CASP1	NM_033292.1	-1.9912	-1.9088
CASP1	NM_033295.1	1.9255	0.8316
CASP10	NM_032974.1	-0.2201	-3.2977
CASP12	NR_000035.1	0.5527	1.0016
CASP14	NM_012114.1	-0.7568	0.435
CASP2	NM_001224.3	-16.9438	-24.337
CASP3	NM_032991.1	5.466	-0.0057
CASP3	NM_004346.2	-4.2973	1.8049
CASP4	NM_001225.2	1.8628	69.081
CASP4	NM_033306.1	1.2561	10.674
CASP4	NM_033307.1	1.4378	-0.5874
CASP5	NM_004347.1	-0.1762	0.8287
CASP6	NM_001226.2	-4.8798	-2.5171
CASP6	NM_032992.1	-3.6136	-0.739
CASP7	NM_033340.1	11.9287	5.3629
CASP7	NM_033340.1	1.7781	-3.2926
CASP8	NM_001228.2	1.1772	-0.6635
CASP8	NM_033356.1	1.4574	-0.0182
CASP8	NM_033357.1	1.8226	-0.7409
CASP8	NM_033358.1	0.5029	5.5887
TERT	NM_198254.1	3.6732	-1.5416

6.10 Appendix - J

Table S-6: Expression level of previously known interacting partners of MCPH1 in patient cells

Symbol	Accession	Diff_Score	Detection_controls	Detection_patients
CHEK1	**NM_001274.2**	**-24.6375**	**1**	**1**
CHEK2	**NM_145862.1**	**-30.1159**	**1**	**0.9993**
CHEK2	NM_007194.2	-3.4797	0.3929	0.1945
BRCA1	NM_007295.1	5.7968	0.9993	0.998
BRCA1	NM_007298.1	-0.3356	1	1
BRCA1	NM_007306.1	-2.3602	0.8444	0.6724
BRCA2	NM_000059.1	-6.3705	0.9974	0.9796
TOPBP1	NM_007027.2	-11.5411	1	1
RAD51	**NM_002875.2**	**-20.6571**	**0.9993**	**0.9974**
RAD51	NM_133487.1	-6.4827	0.878	0.5498
DDB2	NM_000107.1	8.5352	1	1
TP73	NM_005427.1	-9.5609	0.849	0.2861
E2F1	NM_005225.1	-6.1054	0.9993	0.9934
E2F2	NM_004091.2	-0.395	1	1
E2F3	NM_001949.2	-0.6411	1	1
E2F4	NM_001950.3	-8.0954	0.9993	0.9934
E2F5	NM_001951.2	-12.2702	1	1
E2F6	NM_198257.1	-1.912	1	0.9993
E2F7	NM_203394.1	-12.7853	1	1
APAF1	NM_181869.1	3.102	1	1
CASP1	NM_033292.1	2.3862	0.8134	0.8873
CASP1	NM_033295.1	4.2689	0.5419	0.8009
CASP10	NM_032974.1	0.192	0.9097	0.8939
CASP10	**NM_032974.1**	**16.6551**	**1**	**1**
CASP14	NM_012114.1	-1.5163	0.2808	0.1523
CASP2	NM_001224.3	-5.72	0.0692	0.0092
CASP2	NM_032984.2	-9.6615	0.9993	0.998
CASP3	NM_032991.1	-9.1229	1	1
CASP3	NM_004346.2	4.174	0.8682	0.9433
CASP4	NM_001225.2	-7.9397	1	1
CASP4	NM_033306.1	-5.2748	0.8207	0.5366
CASP4	NM_033307.1	-1.9578	0.9664	0.9136
CASP5	NM_004347.1	8.4686	0.998	0.998
CASP6	NM_001226.2	0.4295	0.9993	0.998
CASP6	NM_032992.1	7.6548	1	1
CASP7	**NM_033340.1**	**-14.9385**	**0.9993**	**0.9987**
CASP7	NM_033340.1	2.7093	0.0092	0.0514
CASP8	NM_001228.2	-1.7374	0.5405	0.4272
CASP8	NM_033356.1	-1.5411	0.499	0.3441
CASP8	NM_033357.1	-3.2843	0.0053	0
CASP8	NM_033358.1	-4.6398	0.9374	0.7713
CASP8AP2	NM_012115.2	1.8327	1	0.9993
CASP9	NM_001229.2	2.4088	0.8134	0.8701
CASP9	NM_032996.1	1.2059	1	1
TERT	NM_198254.1	3.2394	0.0382	0.1898

6.11 Appendix - K

Table S-7: Identified variants by solexa sequencing in of family M307

position	SNPs	No. of reads	Allele A read	Allele C read	Allele T read	Allele G read
51734454	SNP G C/G	21	A:	C:14	T:	G:
51676830	SNP T C/T	598	A:3	C:257	T:	G:
51678420	SNP A A/G	873	A:1	C:4	T:	G:383
51679289	SNP G A/G	818	A:809	C:2	T:2	G:
51679385	No SNP	497	A:	C:	T:252	G:
51680474	SNP G A/G	751	A:384	C:	T:	G:
51681835	No SNP	512	A:1	C:	T:278	G:1
51682688	SNP G G/T	331	A:	C:1	T:164	G:
51683284	SNP G C/G	514	A:1	C:258	T:1	G:
51686866	SNP G C/G	469	A:4	C:178	T:1	G:
51687554	SNP C C/T	973	A:6	C:	T:465	G:1
51688806	SNP T A/T	969	A:496	C:	T:	G:
51698037	SNP T A/G	867	A:	C:397	T:	G:2
51699533	SNP T A/T	1251	A:575	C:1	T:	G:2
51699547	SNP G C/G	1370	A:2	C:663	T:2	G:
51701834	SNP G A/G	328	A:138	C:1	T:1	G:
51703424	SNP G A/G	330	A:321	C:	T:5	G:
51703812	No SNP	424	A:2	C:215	T:2	G:
51705517	SNP C A/G	327	A:1	C:	T:152	G:1
51706073	No SNP	965	A:450	C:3	T:3	G:
51706365	SNP A C/T	936	A:	C:2	T:2	G:449
51706689	SNP G A/G	649	A:276	C:	T:	G:
51707528	SNP C G/T	961	A:476	C:1	T:3	G:4
51709083	SNP T C/T	1024	A:5	C:488	T:	G:3
51711271	No SNP	157	A:	C:	T:68	G:1
51729998	SNP G A/G	28	A:20	C:	T:	G:
51759596	No SNP	981	A:	C:2	T:	G:501
51760294	SNP A A/G	1316	A:	C:3	T:6	G:630
51768409	No SNP	485	A:1	C:208	T:	G:1
51768970	No SNP	500	A:	C:2	T:1	G:225
51771447	No SNP	34	A:	C:14	T:	G:
51771838	No SNP	51	A:1	C:	T:19	G:
51772171	SNP T A/T	77	A:35	C:	T:	G:
51774937	No SNP	88	A:	C:1	T:	G:39
51802465	SNP A A/T	127	A:	C:	T:65	G:
51802711	SNP G A/G	920	A:452	C:	T:2	G:
51802865	SNP T C/T	612	A:1	C:344	T:1	G:
51803172	SNP C C/G	643	A:5	C:	T:1	G:339
51803529	SNP T A/T	525	A:263	C:2	T:	G:1
51803606	SNP A A/T	559	A:	C:3	T:293	G:1
51804015	SNP A A/G	698	A:	C:2	T:	G:328
51804033	SNP G A/G	462	A:198	C:1	T:1	G:
51804227	SNP A A/G	704	A:	C:1	T:1	G:380
51804314	SNP A A/T	834	A:	C:	T:417	G:1
51804541	SNP G A/G	620	A:340	C:	T:1	G:
51804634	SNP G G/T	661	A:2	C:	T:337	G:
51804715	SNP A A/C	556	A:	C:316	T:1	G:2
51804795	SNP A A/T	563	A:	C:2	T:317	G:3
51804828	SNP A A/G	609	A:	C:	T:	G:307
51805139	SNP T A/T	576	A:276	C:	T:	G:1
51805526	SNP T C/T	553	A:2	C:287	T:	G:
51805617	SNP T C/T	500	A:1	C:174	T:	G:1
51805630	SNP A A/G	502	A:	C:	T:1	G:159
51805637	SNP T C/T	468	A:2	C:162	T:	G:
51805926	SNP T G/T	500	A:	C:1	T:	G:238
51806174	SNP T C/T	516	A:3	C:258	T:	G:
51806397	SNP T C/T	674	A:2	C:323	T:	G:2
51806776	SNP C C/T	503	A:1	C:	T:242	G:
51806939	SNP T C/T	842	A:11	C:406	T:	G:5
51807363	SNP G A/G	544	A:287	C:	T:3	G:
51807712	SNP G G/T	445	A:	C:3	T:211	G:
51808517	SNP C C/T	595	A:1	C:	T:264	G:1
51809017	SNP C C/T	516	A:8	C:	T:231	G:1
51809443	SNP G A/G	453	A:216	C:	T:1	G:
51810041	SNP A A/G	230	A:	C:	T:	G:101

Table S-7: Identified variants by solexa sequencing in of family M307

position	SNPs	No. of reads	Allele A read	Allele C read	Allele T read	Allele G read
51810463	SNP T C/T	398	A:8	C:168	T:1	G:
51810878	SNP T A/T	504	A:243	C:	T:	G:3
51810929	No SNP	660	A:3	C:338	T:	G:
51810971	SNP T A/T	525	A:260	C:4	T:	G:1
51811945	SNP A A/G	540	A:	C:	T:1	G:258
51812845*	SNP G A/G	631	A:307	C:3	T:1	G:
51813583	SNP C C/T	24	A:	C:	T:17	G:
51813655	SNP T C/T	42	A:1	C:32	T:	G:
51818496	No SNP	24	A:	C:18	T:	G:
51819522	SNP C C/T	33	A:	C:	T:19	G:
51819632	SNP C C/T	32	A:	C:	T:32	G:
51819872	SNP A A/C	49	A:	C:25	T:	G:
51819941	SNP C C/T	30	A:	C:	T:13	G:
51820451	SNP A A/T	61	A:	C:	T:23	G:
51821674	No SNP	25	A:	C:	T:10	G:1
51821675	No SNP	27	A:2	C:	T:9	G:
51823261	SNP A A/G	49	A:	C:	T:	G:49
51823293	SNP C C/T	31	A:	C:	T:30	G:
51824122	SNP C C/T	541	A:2	C:	T:537	G:1
51824674	SNP C C/T	619	A:	C:1	T:612	G:
51825336	SNP G A/G	502	A:496	C:1	T:1	G:
51825413	SNP G A/G	429	A:425	C:1	T:	G:
51825485	SNP C C/T	234	A:1	C:	T:231	G:2
51825518	SNP G C/G	669	A:2	C:656	T:9	G:
51825889	SNP T C/T	549	A:5	C:543	T:	G:
51825991	SNP T C/T	352	A:2	C:348	T:	G:
51826235	SNP G A/G	429	A:424	C:2	T:	G:
51826637	SNP G A/G	673	A:668	C:1	T:1	G:
51826816	SNP C C/G	386	A:	C:	T:2	G:383
51827033	SNP A A/T	457	A:	C:3	T:447	G:3
51827165	SNP T C/T	414	A:2	C:409	T:	G:1
51827254	SNP C C/T	342	A:1	C:	T:333	G:4
51827368	SNP G G/T	300	A:	C:1	T:298	G:
51827524	SNP C C/T	315	A:1	C:	T:309	G:1
51827710	SNP G A/G	219	A:213	C:	T:1	G:
51827785	SNP T G/T	202	A:	C:	T:	G:200
51828167	SNP C C/T	538	A:	C:	T:534	G:
51828314	SNP G A/G	779	A:772	C:	T:2	G:
51841783*	SNP T G/T	200	A:	C:	T:	G:98
51841856	SNP G A/G	92	A:91	C:	T:	G:
51841864	No SNP	91	A:	C:	T:37	G:
51841910	No SNP	128	A:1	C:67	T:	G:
51841966	SNP A A/C	155	A:	C:153	T:	G:
51842122	SNP C C/T	221	A:	C:	T:99	G:
51842610	SNP G A/G	220	A:218	C:1	T:1	G:
51843220	SNP G A/C	202	A:1	C:1	T:200	G:
51843572	SNP T C/T	184	A:1	C:183	T:	G:
51843812	SNP T A/T	151	A:150	C:	T:	G:1
51844134	SNP G C/G	188	A:1	C:187	T:	G:
51844178	SNP C C/T	136	A:2	C:	T:134	G:
51844234	SNP A A/G	185	A:	C:	T:	G:184
51844715	No SNP	194	A:97	C:	T:2	G:
51844777	No SNP	99	A:	C:	T:1	G:39
51844967	SNP T A/T	124	A:119	C:1	T:	G:1
51845136	SNP G A/G	91	A:43	C:	T:	G:
51845656	No SNP	388	A:5	C:297	T:	G:
51845946	SNP G G/T	127	A:1	C:	T:126	G:
51846128	No SNP	344	A:171	C:1	T:	G:
51846316*	SNP A A/G	112	A:	C:	T:1	G:109
51846379	SNP G A/G	83	A:81	C:	T:	G:
51853574	No SNP	21	A:6	C:1	T:	G:
51857292	SNP C C/T	268	A:	C:	T:252	G:1
51857625	SNP GGGT -/GGGT	21	A:20	C:	T:	G:
51857871	SNP T C/T	283	A:	C:281	T:	G:1
51858156	SNP G A/G	322	A:152	C:	T:2	G:
51858522	SNP A A/T	255	A:	C:3	T:249	G:1
51859050	No SNP	272	A:1	C:128	T:	G:1
51859294	SNP G A/G	353	A:347	C:3	T:	G:
51860200	No SNP	311	A:	C:	T:144	G:1

Table S-7: Identified variants by solexa sequencing in of family M307

position	SNPs	No. of reads	Allele A read	Allele C read	Allele T read	Allele G read
51860765	SNP T C/T	357	A:	C:174	T:	G:1
51860841	SNP C C/T	340	A:1	C:	T:169	G:
51861056	SNP C C/T	383	A:3	C:	T:191	G:1
51861410	No SNP	221	A:	C:3	T:111	G:
51863136	SNP G C/T	450	A:446	C:2	T:	G:
51863181	SNP C A/G	499	A:2	C:	T:251	G:1
51864405	SNP C A/C	503	A:253	C:	T:3	G:4
51864610	SNP C C/T	572	A:	C:	T:566	G:
51864973	SNP C C/T	614	A:	C:	T:612	G:
51865456	SNP C C/T	584	A:2	C:	T:282	G:
51865465	No SNP	594	A:269	C:1	T:	G:1
51866364	SNP G A/G	645	A:328	C:1	T:	G:
51879537*	SNP C A/C	766	A:333	C:	T:1	G:
51879739	SNP C C/T	758	A:5	C:	T:376	G:
51879740	SNP G C/T	753	A:363	C:	T:	G:1
51882337	SNP T A/T	610	A:277	C:1	T:	G:
51887597	SNP A A/G	423	A:	C:	T:2	G:419
51888271	SNP C C/G	546	A:	C:	T:2	G:267
51888272	SNP A A/G	536	A:	C:	T:5	G:530
51899050	SNP G G/T	488	A:1	C:3	T:481	G:
51900596	SNP C A/C	425	A:419	C:	T:1	G:3
51901385	No SNP	81	A:39	C:1	T:	G:
51902272	SNP - -/T	222	A:	C:1	T:82	G:
51902273	No SNP	213	A:	C:1	T:102	G:
51902274	No SNP	208	A:72	C:	T:	G:
51902413	SNP T C/T	421	A:4	C:417	T:	G:
51902423	SNP G A/G	373	A:371	C:	T:	G:
51903111	No SNP	170	A:82	C:	T:	G:
51903113	No SNP	169	A:	C:60	T:	G:
51903836	SNP G C/G	815	A:3	C:806	T:4	G:
51904058	SNP G A/G	950	A:937	C:3	T:5	G:1
51904107	SNP T C/T	647	A:3	C:640	T:1	G:
51904710	SNP G A/G	91	A:91	C:	T:	G:
51905356	SNP T C/T	203	A:	C:201	T:	G:
51906423	SNP T C/T	190	A:	C:189	T:	G:1
51906712	SNP G A/G	189	A:188	C:	T:	G:
51906719	SNP T A/T	205	A:90	C:1	T:	G:
51907125	SNP T C/T	155	A:	C:155	T:	G:
51907861	SNP T G/T	177	A:	C:	T:	G:177
51908404	SNP A A/G	166	A:	C:	T:2	G:162
51908511	SNP G G/T	198	A:2	C:1	T:194	G:
51909409	SNP A A/C	223	A:	C:222	T:	G:
51909553	No SNP	29	A:4	C:24	T:	G:
51909584	No SNP	52	A:23	C:	T:	G:
51909739	SNP T C/T	256	A:	C:255	T:	G:1
51910182	SNP G A/G	226	A:226	C:	T:	G:
51910303	SNP C C/T	198	A:	C:	T:195	G:3
51910455	SNP G A/G	245	A:241	C:1	T:2	G:
51910628	SNP G A/G	191	A:184	C:5	T:	G:
51910646	SNP G A/G	311	A:307	C:1	T:1	G:
51910736	SNP T C/T	307	A:1	C:297	T:	G:1
51910768	SNP T C/T	251	A:1	C:249	T:	G:
51910883	SNP G A/G	81	A:79	C:	T:	G:1
51910891	SNP - -/CTG	47	A:	C:	T:	G:33
51911300	SNP C A/C	112	A:111	C:	T:1	G:
51911349	SNP C C/T	42	A:	C:	T:42	G:
51911403	SNP T C/T	24	A:	C:23	T:	G:1
51911411	SNP C C/T	72	A:	C:	T:71	G:
51911433	SNP A A/G	217	A:	C:	T:1	G:214
51911780	SNP T C/T	185	A:1	C:182	T:	G:
51911811	SNP T C/T	161	A:1	C:157	T:	G:1
51911831	No SNP	163	A:	C:	T:77	G:
51911923	SNP T C/T	93	A:	C:93	T:	G:
51911972	SNP A A/G	141	A:	C:	T:3	G:138
51913513	SNP G A/G	139	A:139	C:	T:	G:
51913633	SNP T A/T	207	A:204	C:1	T:	G:1
51914170	SNP G A/G	123	A:105	C:2	T:3	G:
51914308	SNP A A/G	114	A:	C:	T:1	G:111

Table S-7: Identified variants by solexa sequencing in of family M307

position	SNPs	No. of reads	Allele A read	Allele C read	Allele T read	Allele G read
51914354	SNP A A/C	206	A:	C:206	T:	G:
51914634	SNP C C/G	134	A:	C:	T:1	G:130
51914799*	SNP T C/T	299	A:	C:135	T:	G:
51915048	SNP T C/T	563	A:2	C:555	T:	G:1
51916467	SNP T C/T	595	A:1	C:591	T:	G:
51916700	SNP G A/G	428	A:422	C:	T:3	G:
51916963	SNP C C/T	84	A:	C:1	T:83	G:
51917037	SNP C C/G	145	A:1	C:	T:	G:144
51917040	SNP C A/C	183	A:181	C:	T:	G:
51917576	SNP T A/C	303	A:	C:	T:	G:300
51917703	SNP G C/T	525	A:520	C:	T:1	G:
51919748	SNP T A/T	727	A:724	C:1	T:	G:1
51919774	SNP A A/G	553	A:	C:2	T:1	G:546
51921169	SNP A A/C	502	A:	C:498	T:1	G:1
51921573	SNP A A/G	715	A:	C:	T:2	G:710
51922175	SNP C C/T	441	A:3	C:	T:431	G:
51922176	SNP A A/G	464	A:	C:2	T:2	G:460
51922734	SNP T C/T	751	A:4	C:739	T:	G:1
51922979	No SNP	107	A:52	C:	T:	G:1
51923575	SNP G A/G	657	A:644	C:3	T:2	G:
51923804	SNP C A/G	659	A:3	C:	T:648	G:
51925163*	SNP C C/G	723	A:5	C:	T:7	G:350
51928851	SNP T C/T	1310	A:8	C:1294	T:	G:2
51928940	SNP A A/G	864	A:	C:	T:8	G:854
51929007	SNP G A/G	710	A:700	C:1	T:3	G:
51929048	SNP T C/T	745	A:3	C:730	T:	G:1
51929600	SNP G C/G	1279	A:3	C:613	T:3	G:
51930702*	No SNP	1088	A:9	C:530	T:	G:1
51931976	SNP T A/G	87	A:	C:86	T:	G:
51933244	SNP C C/G	59	A:	C:	T:	G:59
51933581	SNP T A/G	55	A:	C:55	T:	G:
51934173	SNP G A/G	70	A:27	C:	T:	G:
51934307	No SNP	83	A:	C:44	T:	G:
51934741	SNP T A/G	64	A:	C:64	T:	G:
51936437	SNP T C/T	47	A:	C:47	T:	G:
51937849	SNP T A/T	144	A:80	C:	T:	G:
51940453	SNP C C/T	104	A:	C:	T:59	G:
51941008	SNP A A/G	103	A:	C:1	T:1	G:101
51941194	SNP T A/T	154	A:153	C:	T:	G:
51941979	SNP C A/C	100	A:99	C:1	T:	G:
51941997	SNP A A/C/G	73	A:	C:	T:	G:73
51942239	SNP G C/T	72	A:72	C:	T:	G:
51942256	SNP A C/T	70	A:	C:	T:	G:70
51943226	SNP C A/G	66	A:	C:	T:64	G:1
51943443	SNP C G/T	391	A:369	C:2	T:1	G:1
51945503	SNP C C/T	937	A:2	C:	T:931	G:1
51946222	SNP A A/G	813	A:	C:1	T:5	G:806
51946331	SNP T A/G	780	A:7	C:769	T:	G:
51947434	SNP G A/G	794	A:790	C:1	T:	G:
51947509	SNP A A/C	521	A:	C:510	T:2	G:
51947777	SNP A A/T	745	A:	C:1	T:741	G:1
51948274	SNP T C/T	769	A:4	C:760	T:	G:1
51948721	SNP A A/G	202	A:	C:	T:6	G:165
51985675	SNP C C/T	291	A:	C:	T:291	G:
51986013	SNP A C/T	284	A:	C:	T:4	G:280
51986430	SNP A A/G	224	A:1	C:1	T:	G:220
51986744	SNP A A/G	306	A:	C:1	T:	G:303
51987457	SNP A A/G	86	A:	C:	T:1	G:83
51987867	SNP T G/T	215	A:	C:	T:	G:213
51988305	SNP A A/G	207	A:	C:	T:1	G:205
51988750	SNP C A/G	320	A:4	C:	T:311	G:1
51989121	SNP G A/G	231	A:224	C:3	T:	G:
51989667	SNP G G/T	315	A:1	C:1	T:312	G:
51990034	SNP C C/G	239	A:	C:	T:4	G:235
51990323	No SNP	178	A:104	C:	T:	G:
51990639	SNP A C/T	169	A:	C:1	T:1	G:167
51990981	SNP C A/G	262	A:5	C:	T:255	G:1
51992001	SNP G C/T	149	A:145	C:1	T:	G:

Table S-7: Identified variants by solexa sequencing in of family M307

position	SNPs	No. of reads	Allele A read	Allele C read	Allele T read	Allele G read
51992088	SNP G C/T	209	A:204	C:	T:1	G:
51992410	SNP C C/T	228	A:2	C:	T:222	G:1
51993310	SNP C A/G	178	A:	C:	T:178	G:
51995378	SNP T A/C	226	A:	C:	T:	G:225
52002398	SNP A A/G	1881	A:	C:2	T:7	G:1861
52002539	SNP G G/T	1623	A:4	C:9	T:1599	G:
52002914	SNP A A/G	1404	A:1	C:	T:5	G:1398
52003955	SNP A A/G	1701	A:1	C:	T:5	G:1690
52004376	SNP C C/T	1047	A:	C:	T:1036	G:2
52004528	SNP G A/C	1408	A:5	C:11	T:1382	G:
52005524	SNP C A/G	1365	A:1	C:	T:1355	G:3
52006114*	No SNP	1384	A:8	C:	T:591	G:4
52006223	SNP T A/T	1148	A:1141	C:2	T:	G:2
52006676	SNP A C/T	701	A:	C:1	T:1	G:699
52006818	SNP G C/T	582	A:561	C:2	T:3	G:2
52006864	SNP A A/G	1102	A:	C:1	T:4	G:1093
52006950	SNP A C/T	649	A:	C:1	T:4	G:619
52007783	SNP A C/T	651	A:1	C:1	T:4	G:644
52009401	SNP G C/G	67	A:1	C:66	T:	G:
52009409	SNP G C/T	82	A:82	C:	T:	G:
52010487	SNP T A/G	100	A:	C:99	T:	G:
52013231	SNP A C/T	80	A:	C:	T:	G:80
52013320	SNP G A/C	73	A:	C:	T:71	G:
52013775	SNP T A/T	346	A:345	C:	T:	G:1
52016869	SNP T A/G	255	A:	C:255	T:	G:
52018967	No SNP	167	A:55	C:	T:	G:
52025509	SNP G A/G	619	A:615	C:1	T:1	G:
52032733	SNP A A/G	392	A:	C:	T:	G:391
52034073	SNP G A/G	401	A:399	C:1	T:	G:
52037435	SNP T G/T	405	A:2	C:1	T:	G:399
52039520	SNP C C/T	562	A:2	C:	T:554	G:
52041671	SNP T A/G	399	A:3	C:393	T:	G:
52042692	SNP C C/T	665	A:1	C:	T:659	G:1
52042937	SNP A A/C	682	A:	C:673	T:1	G:1
52043020	SNP T G/T	591	A:1	C:	T:	G:585
52067961	SNP A A/G	50	A:	C:	T:1	G:49
52068566	SNP T G/T	112	A:	C:	T:	G:112
52068749	SNP G A/G	108	A:107	C:1	T:	G:
52069121	SNP C C/G	33	A:	C:	T:1	G:32
52069538	SNP C C/T	114	A:	C:	T:111	G:1
52069751	SNP A A/G	75	A:	C:	T:	G:73
52071402	SNP T C/T	142	A:	C:142	T:	G:
52095855	SNP C C/T	748	A:1	C:	T:736	G:5
52097931	SNP T A/C	693	A:3	C:	T:1	G:686
52101385	SNP A C/T	570	A:	C:2	T:	G:567
52105114	SNP A G/T	427	A:	C:425	T:	G:
52105898*	No SNP	1467	A:43	C:	T:14	G:504
52105902	No SNP	31	A:5	C:7	T:4	G:
52111584	No SNP	25	A:1	C:13	T:4	G:
52111585	No SNP	154	A:93	C:7	T:	G:7
52111615	No SNP	21	A:	C:	T:9	G:
52119363	No SNP	64	A:	C:3	T:16	G:15
52130671	SNP C C/G	448	A:1	C:	T:2	G:444
52130791	SNP A A/C	951	A:	C:935	T:1	G:3
52133273	SNP A A/G	742	A:	C:3	T:4	G:734
52135656	SNP T A/T	814	A:808	C:	T:	G:3
52135686	SNP T C/T	687	A:10	C:674	T:	G:1
52136525	SNP G A/G	623	A:610	C:3	T:3	G:
52137237	No SNP	77	A:	C:27	T:	G:
52166121	SNP A A/C	142	A:	C:142	T:	G:
52166168	SNP T G/T	97	A:1	C:	T:	G:95
52166598	SNP A A/C	64	A:	C:63	T:	G:
52166719	SNP G A/G	92	A:43	C:	T:	G:
52166789	SNP C C/T	49	A:1	C:	T:47	G:1
52167032	SNP G A/G	31	A:29	C:	T:	G:
52167716	No SNP	108	A:62	C:	T:	G:
52168050	SNP A A/G	152	A:	C:	T:	G:151
52168319	SNP G C/G	113	A:	C:112	T:	G:
52168397	SNP C C/T	74	A:	C:	T:74	G:

Table S-7: Identified variants by solexa sequencing in of family M307

position	SNPs	No. of reads	Allele A read	Allele C read	Allele T read	Allele G read
52168841	SNP A A/T	192	A:	C:1	T:190	G:
52168895	SNP C C/T	105	A:	C:	T:105	G:
52169043	No SNP	30	A:19	C:	T:	G:
52169044	SNP T A/T	31	A:14	C:	T:	G:
52169092	SNP A A/G	65	A:	C:	T:	G:65
52169173	No SNP	60	A:36	C:	T:	G:
52169222	SNP G A/G	62	A:59	C:	T:	G:
52169460	No SNP	23	A:23	C:	T:	G:
52169660	SNP C C/T	43	A:	C:	T:33	G:2
52169816	SNP C A/C	64	A:35	C:	T:	G:
52169947	No SNP	40	A:14	C:	T:	G:
52170158	SNP C A/C	98	A:96	C:	T:	G:1
52170159	SNP G A/G	93	A:93	C:	T:	G:
52170193	No SNP	104	A:50	C:	T:	G:
52170315*	SNP G C/G	173	A:	C:86	T:1	G:1
52170544	SNP G A/G	126	A:126	C:	T:	G:
52170573	SNP T C/T	122	A:	C:120	T:	G:
52170575	SNP A A/G	122	A:	C:	T:2	G:119
52170637	SNP T C/T	85	A:	C:85	T:	G:
52170862	SNP T C/T	88	A:2	C:85	T:	G:1
52171013	SNP A A/G	84	A:	C:	T:	G:83
52171060	SNP C C/T	94	A:	C:	T:94	G:
52186746	No SNP	27	A:12	C:	T:	G:
52186987	No SNP	21	A:	C:	T:8	G:
52210599	SNP C A/G	99	A:	C:	T:42	G:1
52210858	SNP C A/G	164	A:	C:	T:75	G:1
52213516	SNP G A/G	147	A:68	C:2	T:	G:
52213626	SNP G G/T	85	A:	C:	T:47	G:

Variants that are marked by asterisks were verified by Sanger sequencing.

7 Acknowledgements

I owe my loving thanks to my family especially my father, my brother and sisters and their families as well as a number of relatives for all of their support throughout the time from my childhood till now. Also I would like to dedicate this thesis to the memory of my mother who left me alone very soon.

I owe my most sincere gratitude to Prof. Hans-Hilger Ropers for not only his important support and scientific supervision during this period, but also providing a very well-infrastructured environment that made it possible to work on this project. My special thank goes for his moral support and care during the time of my stay in hospital and for all of his efforts to help me in order to not lose too much time and be able to finish my project on time. I warmly thank him for letting me to work in the lab of Prof. Dr. Heidemarie Neitzel as well as Dr. Peter Nick Robinson from the department of Prof. Stefan Mundlos in Charité University Hospital. I am also deeply grateful for letting me to go to MDC to work with Franz Rüschendorf for a short time (in this juncture I'll thank him for all the things that I learnt from him in the field of linkage analysis) as well as attending to the Annual Short Course on Medical and Experimental Mammalian Genetics held in the Jackson Laboratory.

I wish to express my warm and sincere thanks to Prof. Hossein Najmabadi for a very fruitful time that I had with him and his group during my master at the university of social welfare and rehabilitation sciences in Tehran and afterward because of his supports and collaboration that I have been able to join the MPI in Berlin and have the opportunity to work on this fantastic project. I would like to express my cordial appreciation for providing all the clinical information and patient materials from such a big cohort of interesting families which I have to admit that without all of his management, wisdom and follow ups was not possible at all.

Gratitude and appreciation is also extended to Dr. Andreas W. Kuss for all the helps, organizational management, scientific discussion and understandings in the lab which was really important to work in a friendly environment. My deepest thank for reading this manuscript critically and doing all the corrections.

Not being alone at the beginning of coming out from Iran especially for the first time is very important, that's why my special thank goes to my old friend Mohammad Mahdi Motazacher from the master period till now for all the helps and all the good times that we have had together both in lab and outside lab, Many thanks indeed for all that Mahdi!.

I would like to show my deepest gratitude to Mrs. Susanne Freier for not only her helps in cell culture and keeping the cells that I needed alive but her sincere helps in preliminary official organizations such as finding flat, registering in bank and police, translating the very difficult German official letters and making phone calls in behalf of me specially at the start of my joining that I was not able to talk any word in German (except for schönes wockenende!).

Many thanks to Dr. Farkhondeh Behjati, for approaching me to the field of mental retardation and initiating this project by motivating me and some other students to recruit X-lined families as well as her encouragements during my stay at genetics research center in Iran.

Many Thanks from our clinicians both in Tehran and Berlin Dr. Kimia Kahrizi and Dr. Andreas Tzschach for all their efforts and follow ups to characterize the patients clinically.

I am deeply grateful to all of the patients and their respected families as without their collaborations it was not possible at all to conduct such a study. I hope that the results of this study in the near future can open new doors of hope to help them somehow or at least prevent from giving birth of more children with such problems. In this juncture I have to confess that the strongest motivation for me all the time through this study was the attitude to help these families, especially that I had the opportunity to visit some of them personally. Thanks for opening my eyes and helping me to be meaningful.

Special acknowledgment to Dr. Lars Riff Jensen for his friendly helps with the lab works and the scientific discussions whenever that he was needed.

My sincere thanks are due to the Dr. Stephen Lenzner my former group leader for all the guidance and valuable advices, even though that I only had the chance to work with him for a very short time.

I wish to extend my special thanks to Dr. Marc Trimborn from Charité University Hospital Berlin for helping me in doing the siRNA experiments and working on MCPH1 and Dr. Gao Guo for helping to perform Flp-in T-Rex and Bac-to-Bac experiments.

I would like to thank Dr. Andrew Jackson (Western General Hospital, Edinburgh) for providing us cell line from their patients with *MCPH1* mutations.

I wish to thank Dr. Chandan Goswami for his helps and discussions mostly on the long stops very late in the afternoon while passing their lab before going home.

I thank Dr. Diego Walther for providing the knock-out MCPH1 mice and special thank for surgery of the mice and taking out their brain and other organs needed for RNA extraction even though that I have not presented the data here.

Many thanks to Ms. Tamara Safari for ordering primers and all the other materials regarding to the sequencing and Ms. Pamela Kepper for doing the Illumina expression and genotyping hybridizations.

I would like to thank the genotyping facility centers in RZPD, Tubingen, Atlas biolabs and MDC and many thanks from Prof. Dr. Guerters and Dr. O. Blankenstein for mass-spectrometry experiments performed at the Institute for Experimental Pediatric Endocrinology in Charite Universitaetmedizin, Berlin, Germany.

I take the advantage of this opportunity to thank all of my previous and current lab member and colleagues: Lia Abbasi Moheb, Sahar Esmaeeli Nieh, Agness Zecha, Lucia Püttmann, Joanna Walczak, Bettina Lipkowitz, Marion Amende, Melanie Wndehack, Bartek Budney and Achim Salamon for providing such a nice and friendly environment in the lab.

I wish to thank all the other colleagues, PhD students and technicians in our department for all their great helps during this period in MPI, many thanks to Olivier Hagens, Fikret Erdoghan, Hannalore Madle, Vanessa Sukow, Nadine, Nils Rademacher, Corina Menzel, Artur moradian and Julia Kuhn.

Many thanks to Ms. Marriane Schlicht and Georg Lienke for their brilliant job in preparing and performing sequencing reactions and all of the other students that had been working with me during this time such as Petra Pop Ristova, Annika Quast, Alexander Reichenbach, Julia Moser and Katharina Albers.

It is a great pleasure to show my profound gratitude toward Dr. Vera Kalscheuer, Dr. Rheinhard Ullmann, Dr. Wei Chen and Dr. Tim Hucho for the scientific discussions and inputs especially during our regular Friday seminars.

I am deeply indebted to all of my former professors at the Genetics research center (Professor Mohammad Ali Mowlavi, Dr. Yousef Shafeghati, Dr. Mina Ohadi, Dr. Maryam Neishabouri, Dr. Keyhani and Dr. Mohamad Khalil Javan) at USWR.

I warmly thank all of those colleagues in the genetics research center of USWR and Kariminejad-Najmabadi Pathology & Genetics Center who has been involved to help somehow in this project such as Ms. Susan Banihashemi, Ms. Saghar Ghasemi Firouzabadi, Mrs. Sedigheh Abedini, Mrs. Khadijeh jalalvand, Ms. Sanaz Arzhangi, Ms. Masoomeh Falah, Ms. Nastaran Arabinejad, Ms. Vale Hadavi and Mrs. Marzieh Mohseni.

My special thanks and gratitude will go to my best friend Raghu Bhushan and his wife Smita Sudheer for all the helps and discussions especially during our lunch time together everyday.

I am particularly grateful from my good friend Nima Keshvari for reading some parts of this manuscript and his comments.

I would like to thank Nejat Mahdieh one of my old friend and classmate to help with the recruitment of X-linked families which was a starting point for initiating this project.

I would like to my old friend Alireza Pouya for his help to collect clinical information and DNA samples from one of the mental retarded families which I needed to be complete very fast before my joining to the MPI.

I am also greatly indebted to two of my dear friends Nima Yazdanbakhsh and Nader Naderi for their helps and all the good times that I had with them in Berlin.

Special thank to all of the teachers and professors from whom I learned a lot especially my first Genetics professor, Dr. Farhang Haddad during my B.sc. for his nice way of teaching the basics of molecular genetics and human genetics as well as lending the text books to me and one of my old friends Modjtaba Emadi-Baygi (which I use this opportunity here and thank him for studding together and all the other helps).

This acknowledgment would not be complete without mentioning all of those colleagues from the welfare institutions in different city and provinces of Iran, especially Kermanshah city which I was personally worked with them for their corporation at the very first steps of this project. Many thanks from Mr. Kamran Alimadadi, Dr. Hilda Yazdan

and many of the nurses which sincerely helped me with the recruitment of blood samples from the families without any expectations.

8 Summary

Severe mental and behavioral disorders are common, affecting 1-3% of the world populace. They thus constitute a major burden not only for the affected families but also for society.

There is reason to believe that autosomal recessive mental retardation (ARMR) is more common than X-linked MR, but it has so far received considerably less attention. This is partly due to small family sizes and low consanguinity rates in industrialized societies, both of which have hampered gene mapping and identification, which is illustrated by the fact that until 2003, when this study was started, no more than one gene was shown to be implicated in non-syndromic ARMR (NS-ARMR). The work presented here is part of a larger project to shed more light on the molecular causes of ARMR as a prerequisite for diagnosis, counselling and therapy, focusing on large consanguineous Iranian families with several mentally retarded children. It combines clinical and molecular approaches such as patient recruitment, clinical characterization, sample collection, SNP array genotyping, whole genome linkage analysis, homozygosity mapping and finally mutation screening in a systematic fashion. Successful mutation detection is followed by functional analyses of the affected genes.

In the study presented here, the investigation of 135 families led to the identification of 31 novel genomic loci for ARMR. Contrary to previous observations, which prima facie argued against the existence of frequently mutated genes, overlapping autozygosity regions from several families could now be observed on chromosomes 1, 5 and 19. At each of these loci a minimum of two overlapping linkage intervals were solitary in the respective families and showed a LOD score of, or above, three.

Mutation screening in one of these families with NS-ARMR has led to the discovery of a new gene for NS-ARMR, *TUSC3*, where a mutation was found that leads to the loss of *TUSC3* transcript in patient cells.

Additional investigations in families with syndromic forms of ARMR revealed a new gene for ataxia and mild mental retardation. This gene, *CA8*, was found to carry a R237Q mutation, with a putatively deleterious effect on functional properties of the gene product in the affected patients.

Furthermore one novel mutation in *ALDH3A2* in patients with Sjögren-Larsson syndrome and two in the *MCPH1* gene in patients with primary microcephaly were found. Gene expression profiling, knockdown experiments and irradiation studies added more evidence on the involvement of MCPH1 in cell cycle control, DNA damage response and transcriptional regulation.

In summary, the identification of a novel gene for NS-ARMR and many new genomic intervals with a high probability for containing different genes with disease causing mutations is in keeping with previous results that indicated a high degree of genetic heterogeneity for this disorder. Still, the several overlapping loci found in this study now also indicate the presence of genes with an increased frequency of mutations in ARMR patients. Further studies are necessary to identify the disease causing mutations in these newly identified linkage intervals and to determine the contribution of the affected genes to the complex processes of human cognition. These studies will be greatly facilitated by the novel high throughput sequencing technologies, which are now available and that will allow a much increased pace for the detection of disease causing mutations.

9 Zusammenfassung

Schwere kognitive Erkrankungen und Verhaltensstörungen betreffen ca. 1-3% der Weltbevölkerung und stellen damit eine erhebliche Belastung für die betroffenen Familien, aber auch die Gesellschaft als Ganzes dar.

Es gibt Gründe die dafür sprechen, dass autosomal rezessive Formen mentaler Retardierung (ARMR) häufiger auftreten als X-chromosomal vererbte Formen geistiger Behinderung, jedoch bisher weniger Aufmerksamkeit erfahren haben. Dies liegt zum Teil daran, dass in industrialisierten Gesellschaften die dort vorherrschenden kleinen Familien und der geringe Grad an Konsanguinität in der Bevölkerung die Genkartierung behindert haben. Veranschaulicht wird dieser Sachverhalt durch die Tatsache, dass zu Beginn dieser Studie im Jahr 2003 nur ein Gen für nicht-syndromale mentale Retardierung (NS-ARMR) bekannt war. Die hier vorgestellte Arbeit ist Teil eines größer angelegten Projekts zur Aufklärung der molekularen Ursachen geistiger Behinderung in konsanguinen iranischen Großfamilien mit mehreren mental retardierten Kindern, um damit die Voraussetzungen für Diagnose, Beratung und Therapie zu verbessern. Diese Studie verbindet klinische und molekulargenetische Untersuchungsmethoden wie Patientenrekrutierung, klinische Charakterisierung, Probensammlung, SNP-array Genotypisierung, genomweite Kopplungsanalyse, Homozygotiekartierung und Mutationsanalyse auf systematische Art und Weise. Auf erfolgreiche Mutationsanalysen folgen schließlich Untersuchungen zur Funktion betroffener Gene.

In der hier vorgestellten Arbeit führte die Untersuchung von 135 Familien zur Identifizierung von 31 neuen Loci für ARMR. Im Gegensatz zu früheren Beobachtungen, welche zunächst gegen die Existenz häufig mutierter Gene sprachen, wurden nun überlappende autozygote Bereiche von mehren Familien auf den Chromosomen 1, 5 und 19 gefunden. An jedem dieser Loci waren mindestens zwei der überlappenden Intervalle die einzigen in den jeweiligen Familien und zeigten einen LOD Score von drei oder höher. Die Mutationsanalyse in einer dieser Familien mit NS-ARMR führte zur Entdeckung eines neuen Gens für NS-ARMR, *TUSC3*, in welchem eine Mutation gefunden wurde, die den Verlust des zugehörigen Transkripts in Patientenzellen zur Folge hat.

Weitere Untersuchungen von Familien mit syndromalen Fromen mentaler Retardierung brachten ein neues Gen für Ataxie mit milder geistiger Behinderung zu Tage. In diesem Gen, *CA8*, tragen die betroffenen Patienten eine R237Q Mutation mit mutmaßlich stark einschränkenden Auswirkungen auf die Funktion des Genprodukts. Des Weiteren wurde eine neue Muation im *ALDH3A2* Gen von Patienten mit Sjögren-Larsson Syndrom, sowie zwei bisher unbekannte Mutationen im *MCPH1* Gen von Patienten mit primärer Mikrozephalie gefunden. Genomweite Genexpressionsuntersuchungen, Knockdown-Experimente und Bestrahlungsversuche lieferten neue Erkenntnisse über die Beteiligung

von MCPH1 an der Zellzykluskontrolle, bei zellulären DNA-Reparatursystemen und Transkriptionsregulation.

Zusammenfassend kann gesagt werden, dass die Identifizierung eines neuen Gens für NS-ARMR und vieler neuer Kopplungsintervalle, welche mit einer hohen Wahrscheinlichkeit unterschiedliche Gene mit Krankheitsverursachenden Mutationen enthalten, mit vorangegangenen Ergebnissen, welche ein Hohes Maß an Heterogenität für ARMR nahe legen, übereinstimmen. Andererseits jedoch deuten die hier beschriebenen überlappenden Loci nun auch auf das Vorhandensein von Genen hin, welche bei ARMR-Patienten häufiger von Mutationen betroffen sind. Weitere Untersuchungen sind erforderlich, um die krankheitsverursachenden Mutationen in diesen neu identifizierten Kopplungsintervallen zu finden, und den Beitrag der betroffenen Gene zu den komplexen kognitiven Vorgängen im menschlichen Gehirn zu verstehen. Diese Studien werden durch die inzwischen zugänglichen neuen Hochdurchsatz-Sequenziertechnologien stark erleichtert, die es ermöglichen Mutationen erheblich schneller aufzuspüren als bisher.

10 List of publications

Turkmen S, Guo G, **Garshasbi M**, Hoffmann K, Alshalah AJ, Mischung C, Kuss A, Humphrey N, Mundlos S, Robinson PN. 2009. *CA8 mutations cause a novel syndrome characterized by ataxia and mild mental retardation with predisposition to quadrupedal gait.* PLoS Genet 5(5):e1000487.

Seifert W, Holder-Espinasse M, Kuhnisch J, Kahrizi K, Tzschach A, **Garshasbi M**, Najmabadi H, Walter Kuss A, Kress W, Laureys G and others. 2008. *Expanded mutational spectrum in cohen syndrome, tissue expression, and transcript variants of COH1*. Hum Mutat.

Kahrizi K, Najmabadi H, Kariminejad R, Jamali P, Malekpour M, **Garshasbi M**, Ropers HH, Kuss AW, Tzschach A. 2008. *An autosomal recessive syndrome of severe mental retardation, cataract, coloboma and kyphosis maps to the pericentromeric region of chromosome 4*. Eur J Hum Genet.

Tzschach A, Bozorgmehr B, Hadavi V, Kahrizi K, **Garshasbi M**, Motazacker MM, Ropers HH, Kuss AW, Najmabadi H. 2008. *Alopecia-mental retardation syndrome: clinical and molecular characterization of four patients*. Br J Dermatol 159(3):748-51.

Garshasbi M, Hadavi V, Habibi H, Kahrizi K, Kariminejad R, Behjati F, Tzschach A, Najmabadi H, Ropers HH, Kuss AW. 2008. *A defect in the TUSC3 gene is associated with autosomal recessive mental retardation*. Am J Hum Genet 82(5):1158-64.

Moheb LA, Tzschach A, **Garshasbi M**, Kahrizi K, Darvish H, Heshmati Y, Kordi A, Najmabadi H, Ropers HH, Kuss AW. 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(2):270-3.

Motazacker MM, Rost BR, Hucho T, **Garshasbi M**, Kahrizi K, Ullmann R, Abedini SS, Nieh SE, Amini SH, Goswami C and others. 2007. *A defect in the ionotropic glutamate receptor 6 gene (GRIK2) is associated with autosomal recessive mental retardation.* Am J Hum Genet 81(4):792-8.

Najmabadi H, Motazacker MM, **Garshasbi M**, Kahrizi K, Tzschach A, Chen W, Behjati F, Hadavi V, Nieh SE, Abedini SS and others. 2007. *Homozygosity mapping in consanguineous families reveals extreme heterogeneity of non-syndromic autosomal recessive mental retardation and identifies 8 novel gene loci*. Hum Genet 121(1):43-8.

Garshasbi M, Motazacker MM, Kahrizi K, Behjati F, Abedini SS, Nieh SE, Firouzabadi SG, Becker C, Ruschendorf F, Nurnberg P and others. 2006. *SNP array-based homozygosity mapping reveals MCPH1 deletion in family with autosomal recessive mental retardation and mild microcephaly*. Hum Genet 118(6):708-15.

Dadgar S, Hagens O, Dadgar SR, Haghighi EN, Schimpf S, Wissinger B, **Garshasbi M**. 2006. *Structural model of the OPA1 GTPase domain may explain the molecular consequences of a novel mutation in a family with autosomal dominant optic atrophy*. Exp Eye Res 83(3):702-6.

Khodayari N, **Garshasbi M**, Fadai F, Rahimi A, Hafizi L, Ebrahimi A, Najmabadi H, Ohadi M. 2004. *Association of the dopamine transporter gene (DAT1) core promoter polymorphism -67T variant with schizophrenia*. Am J Med Genet B Neuropsychiatr Genet 129B(1):10-2.

Garshasbi M, Oberkanins C, Law HY, Neishabury M, Kariminejad R, Najmabadi H. 2003. *alpha-globin gene deletion and point mutation analysis among in Iranian patients with microcytic hypochromic anemia*. Haematologica 88(10):1196-7.