

# **Molecular Genetic Analysis of Autosomal Recessive Primary Microcephaly in Pakistani Kindreds**

Inaugural-Dissertation

zur

Erlangung des Doktorgrades

der Mathematisch-Naturwissenschaftlichen Fakultät

der Universität zu Köln



vorgelegt von

**Muhammad Sajid Hussain**

aus Muzaffar Garh (Pakistan)

Köln, 2011

Berichtersteller/in: Prof. Dr. Peter Nürnberg

Prof. Dr. Angelika A. Noegel

Tag der mündlichen Prüfung: 04-04-2011

Die vorliegende Arbeit wurde in der Zeit von November 2007 bis Februar 2011 unter Anleitung von Prof. Dr. Peter Nürnberg und Prof. Dr. Angelika A. Noegel am Cologne Center for Genomics (CCG) und im Institut für Biochemie I der Medizinischen Fakultät der Universität zu Köln, angefertigt.

*Dedicated to*

*My*

*Loving Parents*

## Acknowledgements

With deep regards and profound respect, I owe this opportunity to express my deep sense of gratitude and indebtedness to my supervisor **Prof. Dr. Peter Nürnberg** for his inspiring guidance, constructive criticism and valuable suggestions throughout the project work. I most gratefully acknowledge his constant encouragement and help in different ways to complete this project successfully.

A few lines are too short to make a complete account of my deep appreciation and hearties gratitude to **Prof. Dr. Angelika Anna Noegel** for her kindness and supervision for the successful completion of this project. I would like to thank her with immense pleasure for her valuable guidance and constant encouragements which I have received during last two years. I also apologize to keep her busy for reading during weekends and holidays.

I feel profound privilege to thank my co-supervisor **Dr. Shahid Mahmood Baig**, National Institute for Biotechnology and Genetic Engineering (NIBGE), Faisalabad, Pakistan for presenting an opportunity to work in Human Molecular Genetics (HMG) laboratory. His constant encouragement and friendly nature are highly acknowledged.

I am also thankful to all patients and their family members for their participation and cooperation for this study. All colleagues in NIBGE are also acknowledged, particularly Dr. Muhammad Farooq for his valuable suggestions and Aysha Azher for her support at the preliminary stages of this project.

Sincere regards and thanks to Dr. Ursula Euteneuer for fruitful discussion.

I express my profound gratitude to my colleagues in CCG particularly Dr. Muhammad Reza Toliat, Elisabeth Kirst, Nina Dalibor, Holger Trucks, Thomas Alef and Dr. Birgit Budde for their valuable suggestions, guidance and help. I am really thankful to Dr. Birgit Budde and Dr. Kathryn Stemshorn for devoting their valuable time to proofread my thesis. I sincerely thank Dr. Wilfried Gunia for his quick response to provide his IT support. My sincere gratitude is to Gudrun Nürnberg for her mathematical assistance to complete this project.

I would also like to hearty thanks my colleagues in Biochemistry I, particularly Martin Technau, and Sascha Neumann for their guidance, Vivek Shahaji Peche for helping to produce results of immunohistochemistry and Martina Munck for her technical assistance.

For administrative help, I say thanks to Nicole Riedel and Dörte Püsche.

My sincere gratitude is to Dr. Aadesh P. Singh for his moral support, helping me to edit and making the final shape of the thesis. I would also like to thank my friend Muhammad Ilyas and his wife Shaista Ilyas for proofreading the thesis and moral support during this study. I am also thankful to my colleague Ilknur Sur for her devotion to proofread my thesis. I would also like to thank all my friends specially, Muhammad Usman, Ghazanfar Ali, Hussain Askari, Dr. Ramzan Khan, Dr. Ali Ahmed Naz and Dr. Hasnain Raza for their moral support and wonderful company. My acknowledgement will not be completed, if I didn't show my sincerity to my best friend Abid Mahmood Alvi, whose inspiration helped me to complete thesis.

I would also like to give a reverence honor and heartfelt thanks to my school teachers and college professors, Muhammad Nawaz Arif, Sajjad Hussain, Ghulam Mustafa junior, Ikhtlaq Ahmed, Ghulam Mustafa senior, Khawaja Ashiq Hussain and Prof. Haq Nawaz for their valuable attention and inspiration.

I remember with gratitude my whole family members who were always a source of strength, support and inspiration. But I express my humble gratitude and indebtedness to my parents for their endless love, constant prayers, patience, tolerance, sacrifice, and encouragement throughout my life. Their encouragement, motivation and prayers enabled the successful completion of this project.

Last but not the least, thanks to funding agencies like Higher Education Commission (HEC) of Pakistan for giving the opportunity to initiate this work in Pakistan and the German Academic Exchange Service (DAAD) to accomplish this project in Germany.

## Table of Contents

<b>1</b>	<b>Introduction.....</b>	<b>1</b>
1.1	Overview of Pakistani population.....	1
1.2	Microcephaly.....	2
1.2.1	Primary microcephaly.....	2
1.2.1.1	Epidemiology and clinical features of primary microcephaly.....	3
1.2.1.2	Molecular genetics of primary microcephaly.....	4
1.3	Functions of MCPH proteins.....	5
1.3.1	Microcephalin.....	6
1.3.2	WD repeat-containing protein 62 (WDR62).....	7
1.3.3	CDK5 regulatory subunit-associated protein 2 (CDK5RAP2).....	7
1.3.4	Centrosomal protein of 152 kDa (CEP152).....	9
1.3.5	Abnormal spindle-like microcephaly-associated protein (ASPM).....	9
1.3.6	Centromere protein J (CENPJ).....	12
1.3.7	SCL-interrupting locus protein (SIL).....	13
1.4	Role of MCPH proteins in neurogenesis.....	13
<b>2</b>	<b>Objectives of the Thesis .....</b>	<b>15</b>
<b>3</b>	<b>Material and Methods.....</b>	<b>16</b>
3.1	Subjects.....	16
3.2	Pedigrees construction and mode of inheritance analysis.....	16
3.3	Blood sampling.....	17
3.4	DNA extraction.....	17
3.4.1	Solutions used for DNA extraction.....	18
3.5	DNA quantification.....	19
3.6	RNA isolation.....	19
3.6.1	Blood.....	19
3.6.2	Human primary fibroblasts.....	19
3.7	Polymerase chain reaction (PCR) and agarose gel electrophoresis.....	19
3.8	Reverse transcriptase PCR (RT-PCR).....	20
3.9	Genotyping with short tandem repeat (STR) markers.....	20
3.10	DNA sequencing.....	22
3.10.1	Sanger sequencing.....	22
3.10.1.1	Cleaning of PCR Product with Exo-SAP.....	23
3.10.1.2	BigDye® terminator cycle sequencing.....	23
3.10.2	Pyrosequencing.....	24
3.10.2.1	Primer design for pyrosequencing.....	24
3.10.2.2	Preparation of PCR product for pyrosequencing.....	24
3.10.2.3	Solution used for pyrosequencing.....	24
3.11	Genome-wide linkage analysis.....	24
3.12	Mutation detection.....	25

3.13	Bacterial culture.....	25
3.13.1	Storing the bacterial culture.....	26
3.14	Mammalian cell lines and cell culture.....	26
3.15	Cell line establishment from patient biopsy.....	26
3.16	Fibroblast synchronization at G0/G1 by serum starvation.....	27
3.17	<i>CDK6</i> and <i>CEP135</i> GFP constructs.....	27
3.17.1	<i>CDK6</i> GFP constructs.....	27
3.17.2	<i>CEP135</i> GFP constructs.....	27
3.17.3	Sequencing of entry and expression clones.....	27
3.18	Plasmid purification.....	28
3.19	Transfection of GFP tagged clones.....	28
3.20	Knockdown of <i>CDK6</i> .....	28
3.20.1	Primer design.....	28
3.20.2	Oligonucleotides annealing.....	29
3.20.3	Digestion of pSHAG-1 vector.....	29
3.20.4	Cloning procedure.....	30
3.20.5	Plasmid purification.....	31
3.20.6	Transfection.....	31
3.21	Preparation of protein lysates from eukaryotic cells.....	31
3.22	SDS-polyacrylamide gel electrophoresis (SDS-PAGE).....	31
3.22.1	Solutions used for SDS-PAGE.....	32
3.23	Staining of polyacrylamide gels with Coomassie-Brilliant-Blue R 250.....	32
3.24	Protein transfer to membranes (Western blot).....	32
3.25	Immunodetection of proteins bound to the membrane.....	33
3.26	Immunocytochemistry.....	34
3.26.1	Immunostaining of microtubules.....	34
3.26.2	Immunostaining of $\gamma$ -tubulin.....	34
3.26.3	Solutions used for Immunocytochemistry.....	34
3.27	Immunohistochemistry of paraffin-embedded embryo sections.....	36
3.28	Confocal microscopy.....	36
3.29	Antibodies used for Immunofluorescence and western blotting.....	36
<b>4</b>	<b>Results.....</b>	<b>38</b>
4.1	Genotyping of primary microcephaly families.....	38
4.2	MCPH1 linked family.....	39
4.3	MCPH2 linked families.....	42
4.4	MCPH4 linked families.....	46
4.4.1	MCP8 family region on chromosome 1.....	49
4.5	MCPH5 linked families.....	49
4.5.1	Haplotypes of MCPH5 linked families.....	49
4.5.2	Abnormal Spindle-like Microcephaly associated ( <i>ASPM</i> ) Sequencing.....	58

4.6	MCPH6 linked family.....	62
4.7	Excluded families.....	63
4.8	Excluded family MCP4.....	63
4.8.1	MCPH8 locus identification.....	64
4.8.2	Candidate genes in the MCPH8 locus.....	67
4.8.3	Candidate gene sequencing in the MCPH8 locus.....	67
4.8.4	<i>CDK6</i> as the causative gene of the MCPH8 locus.....	69
4.8.5	Structure of <i>CDK6</i> .....	70
4.8.6	Evolutionary conservation of Alanine .....	70
4.8.7	Sub-cellular localization of <i>CDK6</i> .....	71
4.8.8	Cell cycle synchronization of <i>CDK6</i> patient fibroblasts.....	73
4.8.8.1	Misshapen nuclei.....	73
4.8.8.2	Reduced cell proliferation.....	73
4.8.8.3	Increased centrosome-nucleus distance.....	74
4.8.9	<i>CDK6</i> knockdown.....	75
4.8.9.1	Normal microtubule network in control.....	75
4.8.9.2	Misshapen nuclei.....	76
4.8.9.3	Reduced cell proliferation.....	76
4.8.9.4	Disorganised microtubules.....	77
4.8.10	Overexpression of <i>CDK6</i> .....	78
4.8.10.1	Localization of GFP-tagged <i>CDK6</i> .....	78
4.8.10.2	Disorganised microtubules.....	78
4.8.10.3	Multiple centrosomes.....	79
4.9	Excluded family MCP63.....	80
4.9.1	MCPH9 locus identification.....	81
4.9.2	Candidate genes of the MCPH9 locus.....	85
4.9.3	Sequencing the candidate genes in the MCPH9 locus.....	85
4.9.4	<i>CEP135</i> as the causative gene of MCPH9.....	86
4.9.5	Structure of <i>CEP135</i> .....	86
4.9.6	<i>CEP135</i> in wildtype primary fibroblasts.....	87
4.9.7	<i>CEP135</i> patient primary fibroblasts.....	87
4.9.7.1	Centrosome number abnormalities .....	88
4.9.7.2	Absence of centrosomes.....	90
4.9.7.3	Disorganised microtubules.....	91
4.9.7.4	Misshapen nuclei.....	91
4.9.8	Overexpression of <i>CEP135</i> .....	93
4.9.8.1	Overexpression of <i>CEP135</i> leads to a disorganised microtubule system.....	93
4.9.8.2	Abnormal number of centrosomes.....	95
4.9.9	<i>CDK6</i> and <i>CEP135</i> expression in mouse neuroepithelium.....	95
4.10	Excluded family MCP50.....	97

4.11	Excluded family MCP53.....	99
4.12	Excluded family MCP67.....	101
<b>5</b>	<b>Discussion.....</b>	<b>105</b>
5.1	Mutational spectra of known MCPH genes.....	106
5.2	<i>WDR62</i> , causative gene of the MCPH2 locus.....	109
5.3	<i>CEP152</i> and MCPH4 locus.....	110
5.4	Identification of novel loci.....	110
5.5	<i>CDK6</i> as novel MCPH gene.....	111
5.6	<i>CEP135</i> as novel MCPH gene.....	114
5.7	Cdk6 and Cep135 expression in the neuroepithelium of the mouse cerebral cortex.....	115
5.8	Final conclusions and outlook.....	115
<b>6</b>	<b>Abstract.....</b>	<b>117</b>
<b>7</b>	<b>Zusammenfassung.....</b>	<b>119</b>
<b>8</b>	<b>References.....</b>	<b>121</b>
<b>9</b>	<b>Abbreviations.....</b>	<b>130</b>
	<b>Appendices.....</b>	<b>132</b>
	<b>Erklärung .....</b>	<b>146</b>
	<b>Curriculum Vitae .....</b>	<b>147</b>
	<b>Lebenslauf.....</b>	<b>148</b>

# 1. Introduction

## 1.1 Overview of the Pakistani population

Pakistan is inhabited by more than 170 million people ranking it at place six among the world's most populated countries. Different religions are practiced in Pakistan with the Islam being the major religion (96.3%). The rest of 3.7% is shared by Hinduism (1.6%), Christianity (1.6%) and others such as Sikhs, Parsis, Buddhists, Jews, Bahais, and Animists (Library of Congress, <http://lcweb2.loc.gov/frd/cs/profiles/Pakistan.pdf>). Pakistan is a multilingual country with more than sixty languages being spoken. The main national language is Urdu but English is the official language and is used in official business and governmental legal issues. Other major languages spoken in Pakistan are Punjabi, Saraiki, Pashto, Sindhi, and Balochi along with a lot of other local languages. The Pakistani population is divided into six main ethnic groups, Punjabis (44.15%), Pashtuns (15.42%), Sindhis (14.1%), Seraikis (10.53%), Muhajirs (7.57%), and Balochs (3.57%). Other, smaller ethnic groups (4.66%) mainly found in the northern parts of the country are Kashmiris, Hindkowans, Kalash, Burusho, Brahui, Khowar, Shina, and Turwalis. (<http://www.brookings.edu/~media/Files/Programs/FP/pakistan%20index/index.pdf>).

The majority of the Pakistani population (~70%) lives in villages facing severe educational, economic, and health problems. People marry early and have many children. Due to the social, cultural and ethnic customs, consanguineous marriages are very common in the Pakistani society. Usually several members of related families marry in a reciprocal way. This results in large family sizes with multiple loops of consanguinity. Consequently, many children are born with genetic defects. Especially the frequency of recessive genetic diseases is dramatically increased over the usual prevalence observed in western countries due to homozygous mutations. The poor socio-economic conditions in combination with a high rate of illiteracy make the situation even worse. Birth control is not popular in Pakistan and not at all applied in rural areas, thus even couples having already a large number of afflicted children continue to reproduce in the hope for a normal one. Diseases that are frequently observed are beta-thalassemia, inherited deafness and microcephaly (Baig et al., 2008, Malik et al., 2006, Baig et al., 2011, Pattison et al., 2000, Roberts et al., 1999). No doubt, the high incidence of disabled children is a major political health issue in Pakistan and needs to be solved in the coming decades. On the other hand, the current situation also gives molecular geneticists good opportunities to identify and characterize the genes that result in the disease phenotype when mutated. This is well documented for autosomal recessive primary microcephaly (MCPH) in which six out of nine genetic loci (seven loci previously described and two from the present study) were identified in consanguineous families of Pakistani origin (Woods et al., 2005).

## 1.2 Microcephaly

The term microcephaly is composed of two words, “micro” is derived from the Greek word “*mikros*”, small, and “cephaly” from a Greek word “*Kephale*”, head. The small head size is the prominent phenotype of the patients afflicted with microcephaly.

There are two types of microcephaly; primary and secondary microcephaly. Primary microcephaly is a static anomaly in which the growth of the brain is reduced during pregnancy (develops at prenatal stages), whereas in secondary microcephaly the brain size is normal at birth but subsequently fails to grow normally due to progressive neurodegeneration. In contrast to primary microcephaly, the secondary microcephaly develops at postnatal stages (Woods, 2004). There are some environmental factors like; intrauterine infections, drugs taken during pregnancy; prenatal radiation exposure, maternal phenylketonuria; and birth asphyxia forming the cause of the rare secondary microcephaly (Jackson et al., 1998). Congenital infection with *Toxoplasma*, overconsumption of alcohol during pregnancy and Rubenstein Taybi syndrome are some factors that must be excluded for diagnosis of autosomal recessive primary microcephaly, because these factors also cause microcephaly with mental retardation (Woods et al., 2005). The autosomal recessive primary microcephaly has only a genetic cause without environmental factors.

### 1.2.1 Primary microcephaly

Autosomal recessive primary microcephaly (MCPH) is defined as a neurodevelopmental disorder characterized by reduced head circumference ( $> -3SD$ ), due to a small but structurally normal brain, and mild to severe intellectual disability with normal height (except in minor cases where the MCPH1 gene is involved), weight and neurological functions (Cox et al., 2006, Nicholas et al., 2009). The small head circumference (HC) is the result of reduced cerebral cortex size which leads to a simplified gyral pattern but the cortex is of normal thickness (Desir et al., 2008, Figure 1.1).

As the patients have a significantly reduced brain growth without any affect on brain structure, it will be instructive to investigate the mechanism of neurogenesis in this disorder and MCPH can be considered a model disease for the identification of novel genes which have a role in prenatal neurogenesis.

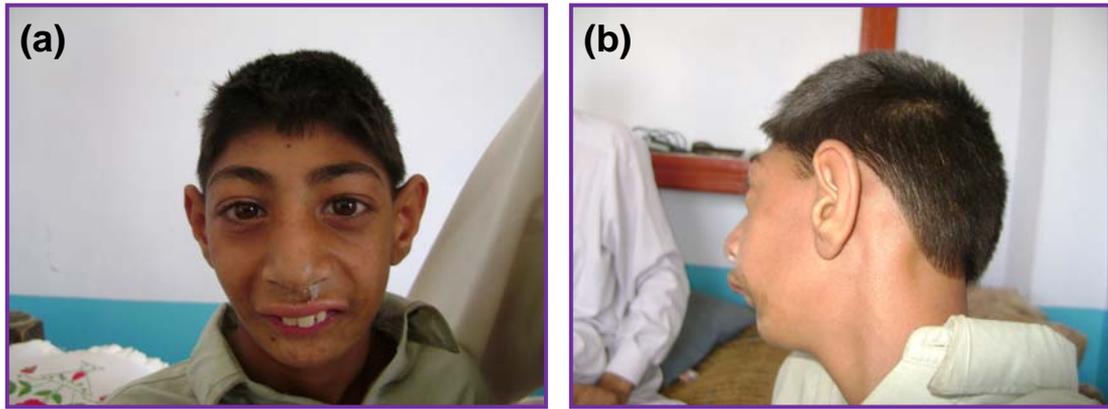
### **1.2.1.1 Epidemiology and clinical features of primary microcephaly**

The incidence of MCPH is 1/250,000 in the Netherlands, 1/30,000 in Japan and 1/2,000,000 in Scotland (Van den Bosch, 1959, Komai et al., 1955, Tolmie et al., 1987). But the incidence of microcephaly from genetic as well as environmental causes was estimated to be 1/93,000 (Van den Bosch 1959). MCPH is highly prevalent in the northern Pakistani population with an incidence of nearly 1 in 10,000. This shows the high occurrence in the Indian subcontinent as compared to the other parts of the world. The incidence is greater in Asian and Arab populations where a high rate of consanguineous marriages occurs (Woods et al., 2005).

The inheritance pattern of MCPH is autosomal recessive. All the reported families afflicted with this disorder have multiple or few afflicted individuals with unafflicted consanguineous parents (Muhammad et al., 2009).

Primary microcephaly is likely to be genetic in aetiology. The patients have learning difficulties but no environmental, metabolic, cytogenetic, developmental malformations or significant neurological deficits. The microcephaly which has only the genetic aetiology is also known as true microcephaly, microcephaly vera, recessive microcephaly or primary microcephaly. As MCPH is a disorder of foetal brain growth (the only significant clinical feature), microcephaly is apparent at birth with a sloping forehead as an obvious feature, The head circumference (HC) is significantly less than expected for an individual of similar age and sex, and the patients are mentally retarded. Brain size measurement can also be done with a more accurate technique like volumetric nuclear magnetic resonance (NMR) scanning, but HC measurement remains the most common and simple method for evaluating gross brain size. To determine the skull circumference reliably it is important to carry out the measurement from the forehead to the occipital prominence at the back of the head, which closely correlates with reduced brain volume. Reduced cranial size observed in microcephalic patients strongly correlated with mental retardation (Cox et al., 2006).

In MCPH, the reduced size of the skull is due to a reduced central nervous system and the cerebral cortex is the structure which is most strongly affected. Magnetic resonance imaging (MRI) and computerized tomography (CT) scans of the MCPH patient's brain show a normal architecture of the brain but of reduced size, which is especially evident in the cerebral cortex. The structure of the cerebral cortex is simplified, and there is a slight reduction in the volume of the white matter, consistent with the small size of the brain (Barkovich et al., 2001, Bond et al., 2002).



**Figure 1.1:** A MCPH patient; front view (a) and side view (b) with typical features of reduced head and sloping forehead.

### 1.2.1.2 Molecular genetics of primary microcephaly

Although all MCPH patients reported to date have indistinguishable features, this disorder is genetically heterogeneous. Autozygosity mapping techniques using microsatellite markers or SNPs throughout the human genome in large consanguineous MCPH families identified seven loci, MCP1-MCP7 (Kaindl et al., 2010). Further five gene loci (MCPH8-MCPH12) have been identified during this study. So, the number of identified MCPH loci increased to 12.

The MCPH1 locus was identified on chromosome 8p22-pter, MCPH2 on 19q13.1-q13.2, MCPH3 on 9q34, MCPH4 on 15q15-q21, MCPH5 on 1q31, MCPH6 on 13q12.2 and MCPH7 on 1p32.3-p33 (Jackson et al., 1998, Roberts et al., 1999, Moynihan et al., 2000, Jamieson et al., 1999, Pattison et al., 2000, Leal et al., 2003, Kumar et al., 2009). Of these seven gene loci, three (MCPH4, MCPH6 and MCPH7) were identified in families from Morocco, north eastern Brazil and India respectively. MCPH1, MCPH2, MCPH3 and MCPH5 were mapped in Pakistani families. Studies conducted to date suggest that *MCPH5* is the most common locus in all populations (Thornton and Woods, 2009). The possibility of further MCPH locus (loci) exists because 18 northern Pakistani MCPH families out of 56, 10 Pakistani families out of 33, 81 Iranian families out of 112 do not show linkage to any of the known seven MCPH loci (Roberts et al., 2002, Gul et al., 2006b, Darvish et al., 2010). MCPH heterogeneity studies have been performed in a northern Pakistani population and MCPH5 was confirmed as the most common locus, accounting for linkage of nearly half of the amassed families (24/56 families). MCPH2 accounted for 10 families (14%); MCPH1 and MCPH3 were each associated with two families (4%), whereas MCPH4 linkage has not been found in the Pakistani population (Roberts et al., 2002). MCPH6 and MCPH7 loci had not been identified at the time of the heterogeneity investigation.

For these loci the following genes have been identified, *Microcephalin/BRIT1* was identified at the MCPH1 locus, *WDR62* (WD repeat domain 62) at MCPH2, *CDK5RAP2/Cep215* (Cyclin Dependent Kinase 5 Regulatory Associated Protein 2, Centrosomal Protein of 215 kDa) at MCPH3, *CEP152* (Centrosomal Protein 152 kDa) at MCPH4, *ASPM* (Abnormal Spindle-like Microcephaly-associated) at MCPH5, *CENPJ* (Centromere Protein J) at MCPH6 and *STIL/SIL* (SCL/TAL1 interrupting locus) at MCPH7 (Jackson et al., 2002, Yu et al., 2010, Nicholas et al., 2010, Bond et al., 2005, Guernsey et al., 2010, Bond et al., 2002, Kumar et al., 2009) (Table 1.1).

**Table 1.1: Overview of all identified MCPH loci with respect to their protein coding genes, ethnicity and chromosomal location.** The table is reproduced from (Thornton and Woods 2009), only the genes functions are omitted and ethnicity is added. The data of MCPH2 and MCPH4 loci have been added (Yu et al., 2010 and Nicholas et al., 2010, Guernsey et al., 2010).

Locus	Chromosome Location	Ethnicity	Gene	Protein	Cellular Localization
<i>MCPH1</i>	8p23	Pakistani	<b><i>MCPH1/Microcephalin/BRIT1</i></b> (BRCT inhibitor of telomerase I)	Microcephalin/ BRIT1	Nucleus/chromatin; centrosome
<i>MCPH2</i>	19q13.12- q13.2	Pakistani	<b><i>WDR62*</i></b>	WDR62	cytoplasmic at Interphase, Spindle pole throughout the cell cycle
<i>MCPH3</i>	9q33.2	Pakistani	<b><i>CDK5RAP2/CEP 215</i></b>	CDK5RAP2/ Cep215	Centrosomal throughout cell cycle; midbody at cytokinesis
<i>MCPH4</i>	15q15-q21	Moroccan	<b><i>CEP152</i></b>	CEP152	Centrosome
<i>MCPH5</i>	1q31.3	Pakistani	<b><i>ASPM</i></b>	ASPM	Pericentrosomal at mitotic spindle poles; midbody at cytokinesis; cytoplasmic at Interphase,
<i>MCPH6</i>	13q12.12	Brazilian	<b><i>CENPJ/ CPAP</i></b>	CENPJ/ CPAP	Centrosomal throughout cell cycle; midbody at cytokinesis
<i>MCPH7</i>	1p33	Indian	<b><i>STIL/SIL</i></b>	SIL/STIL	Pericentrosomal at mitotic spindle poles

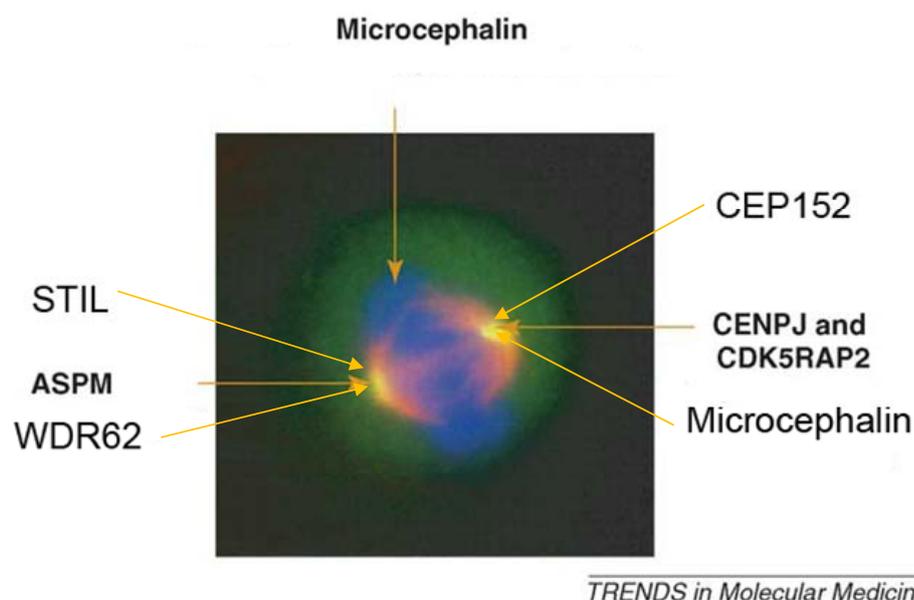
\* The genes with HGNC name are highlighted in bold with most common alternatives.

### 1.3 Functions of MCPH proteins

MCPH is a genetically heterogeneous disorder with seven loci identified so far. The causative genes of these loci have been identified by using a positional cloning strategy. In this section I will briefly introduce each MCPH gene and the possible role of the encoded proteins in neurogenesis.

### 1.3.1 Microcephalin

Microcephalin was the first causative gene of MCPH identified in two consanguineous families originating from the Mirpur area of Pakistan. A single homozygous c.74C>G (p.S25\*) nonsense mutation was identified in the second exon of the MCPH1 gene. MCPH1 encodes microcephalin, a 835 amino acid protein (Jackson et al., 2002). Microcephalin contains three breast cancer 1 (BRCA1) C-terminal (BRCT) domains. One BRCT domain is found at the N-terminus and the other two at the C-terminus of the protein. DNA repair and cell cycle checkpoint proteins also have BRCT domains (Huyton et al., 2000). The role of microcephalin has been elucidated in DNA damage induced cellular responses by the C-terminal BRCT domains in human embryonic kidney cells (Xu et al., 2004). This protein also plays a role in the regulation of chromosome condensation (Trimborn et al., 2004). The role of microcephalin in the etiology of MCPH might be due to defects of normal cell cycle regulation in neuronal progenitors or excessive apoptosis due to failure in DNA repair during neurogenesis (Jackson et al., 2002). Centrosomal location of microcephalin was reported in U2OS cells (Zhong et al., 2006). Previously the nuclear and chromatin associated distribution of microcephalin was reported (Xu et al., 2004, Lin et al., 2005). The centrosomal localization was confirmed in chicken DT40 cells, where it has been shown to localize on centrosomes by using centrosomal and pericentriolar markers like  $\gamma$ -tubulin, centrin, Aurora-A and Nedd1 (Figure 1.2). The centrosomal localization and cell cycle regulatory function is mediated by the N-terminal BRCT domain (Jeffers et al., 2008).



**Figure 1.2: Localization of all identified MCPH proteins.** Microcephalin has a centrosomal as well as a nuclear and chromatin localization. CDK5RAP2, CENPJ and CEP152 are core centrosomal proteins, while ASPM, WDR62 and STIL have a spindle pole localization. The image is taken from (Cox et al., 2006) and was modified to add the localization of newly identified MCPH proteins.

### **1.3.2 WD repeat-containing protein 62 (WDR62)**

The MCPH2 locus was identified in two multi-afflicted consanguineous families of Northern Pakistani origin on chromosome 19q13.1-q13.2 (Roberts et al., 1999). Despite discovery of six other MCPH genes since then, the gene responsible for MCPH2 has escaped discovery for a long time. Recently the MCPH2 gene has been identified by using classical and neo-classical reverse genetics. First the WD repeat domain 62 protein gene (*WDR62*) was found to be mutated in patients afflicted with microcephaly and cortical abnormalities (Bilguvar et al., 2010). These afflicted patients belong to consanguineous Turkish families and manifested with microcephaly, moderate to severe mental retardation, and cortical malformation including pachygyria with cortical thickening, microgyria, lissencephaly, hypoplasia of the corpus callosum, schizencephaly, and in one instance, cerebellar hypoplasia. Later on two independent groups reported *WDR62* as a causative gene of the MCPH2 locus (Yu et al., 2010, Nicholas et al., 2010). The *WDR62* gene consists of 32 exons and encodes a 1,523 amino acid protein containing at least 15 WD repeats.

Immunohistochemistry analysis and in situ hybridization in human and mice revealed *WDR62* expression in neural progenitors within the ventricular and subventricular zones and predominantly in the nucleus in neuronal cells, which was confirmed by immunofluorescence microscopy analysis. It was also reported that this protein is not associated with the centrosome during mitosis (Bilguvar et al., 2010). However confocal microscopy analysis in HeLa cells demonstrated a spindle pole localization of *WDR62* protein during mitosis and weak cytoplasmic staining was observed during interphase. Spindle pole localization of *WDR62* was confirmed by overexpression of GFP-tagged wild-type and mutant c.1313G>A (p.Arg438His) protein in HeLa cells. The wild-type *WDR62* protein was located at the spindle pole whereas the mutant protein was not. Immunohistochemistry also showed *Wdr62* expression in mouse cerebral cortex neuroepithelium from early to late neurogenesis (Nicholas et al., 2010). An independent study also confirmed the spindle pole localization of *WDR62* in HeLa cells during mitosis and perinuclear localization during interphase, suggesting a localization to the Golgi apparatus, which was confirmed by colocalization with the Golgi marker GM130. Recombinant HA-tagged *WDR62* protein in HeLa cells showed a distribution closely matching the localization CEP170 and surrounding LIS1. This pattern of localization was similar to that of other identified primary microcephaly proteins (Yu et al., 2010).

### **1.3.3 CDK5 regulatory subunit-associated protein 2 (CDK5RAP2)**

*CDK5RAP2* which is also known as *CEP215* was identified as the third causative gene of MCPH in a family of Pakistani origin. The *CDK5RAP2/Cep215* protein localizes in HeLa

cells to the spindle pole. Mouse Cdk5rap2 was also reported to be expressed in the the neuroepithelium lining the lateral ventricles of the forebrain (Bond et al., 2005). At the centrosome CDK5RAP2 colocalized with  $\gamma$ -tubulin (Fong et al., 2008).

Studies of the *Drosophila* centrosomin (cnn) that is the orthologue of human *CDK5RAP2* showed an interaction of centrosomin with the  $\gamma$ -tubulin ring complexes at the centrosome which are responsible for the formation of the microtubules that extend from the mitotic spindle (Li and Kaufman, 1996, Megraw et al., 1999, Terada et al., 2003). The *Drosophila* cnn mutant also exhibits reduced cell numbers in the central and peripheral nervous system (Li and Kaufman, 1996). Cnn is also required for the role of the centrioles in maintaining a stable connection both to the pericentriolar matrix (PCM) during centrosome maturation, and to astral microtubule arrays generated by the centrosome during mitosis. Without cnn, centrioles become displaced to the PCM periphery before losing contact and migrating randomly in the cytoplasm (Lucas and Raff, 2007). The depletion of either polo kinase or cnn severely affected the centrosome maturation (Dobbelaere et al., 2008). One of the conserved motifs of cnn (cnn motif 1) which is found at its N-terminus is necessary for the proper recruitment of  $\gamma$ -tubulin, D-TACC (the homolog of vertebrate transforming acidic coiled-coil proteins (TACC), and Minispindles (Msps) to embryonic centrosomes but is not required for assembly of other centrosome components including Aurora A kinase and CP60. Centrosome separation and centrosomal satellite formation are severely disrupted in Cnn Motif 1 mutant embryos (Zhang and Megraw, 2007).

Overexpression of full length CDK5RAP2 wild-type protein leads to the formation of clusters of CDK5RAP2 with variable sizes both at the centrosomes and in the cytoplasm. This also resulted in the increased accumulation of pericentrin (PCNT),  $\gamma$ -tubulin and Cep250 and could nucleate microtubules without centrioles (Fong et al., 2008). The siRNA screens identified CDK5RAP2 as one of the few proteins required for centrosome cohesion (Graser et al., 2007), and Knockdown of CDK5RAP2 led to centrosome splitting and reduced PCNT localisation at the centrosome, although other proteins tested localized normally. CDK5RAP2 localisation to the centrosome is also partially dependent on PCNT (Graser et al., 2007, Haren et al., 2009).

Thus, on the basis of investigations of CDK5RAP2 orthologues which showed that they regulate centrosome maturation, recruitment to and strengthening of the PCM at the centrioles, and might also regulate centrosome cohesion, whereas in their absence centrosomes fail to mature, cannot efficiently organize microtubules, and generation of astral microtubules is reduced, it can be speculated that similar defects in humans might

lead to spindle positioning defects that cannot be tolerated in neuroepithelial progenitors (Thornton and Woods, 2009).

#### **1.3.4 Centrosomal protein of 152 kDa (CEP152)**

Another primary microcephaly gene responsible for MCPH4 locus that waited 11 years for its discovery is the *CEP152* gene. It was identified in three families of an Eastern Canadian subpopulation. In the first family one missense variant p.Q265P was identified and the second family was compound heterozygous for the missense mutation plus a second, premature-termination mutation (p.R987\*) which resulted in the truncation of CEP152 protein. Expression of GFP fused wild-type and mutant (p.R987\*) CEP152 protein in human U2OS osteosarcoma-derived cells revealed a centrosomal localization for the wild-type protein whereas the pR987\* mutant was not detected at the centrosome. By contrast, centrosomal localization was not disturbed in the point mutant (p.Q265P) CEP152 construct (Guernsey et al., 2010).

CEP152 was originally identified as a core component of the centrosome by direct proteomic characterization of the human centrosome by protein correlation profiling (Andersen et al., 2003). CEP152 is the mammalian ortholog of the *Drosophila* gene *asterless*, in which mutations cause arrest of embryogenesis in the fly and male infertility. The protein product of *asterless* (Asl) directly localizes to centrioles (Varmark et al., 2007). Polo-like-kinase 4 (Plk4) plays a role in centriole assembly and duplication. The centriolar protein *asterless* (Asl; human orthologue CEP152) provides a conserved molecular platform, of which the amino terminus interacts with the cryptic Polo box of Plk4 whereas the carboxy terminus interacts with the centriolar protein Sas4. So *asterless* is a scaffolding protein for the onset of centriole assembly (Dzhindzhev et al., 2010).

Recently CEP152 was found to be a genome maintenance protein, of which disruption causes seckel syndrome. Immunofluorescence analysis of patient fibroblasts having mutations in CEP152 revealed the presence of multiple nuclei without astral microtubules. There were also multiple and fragmented centrosomes seen and partially depolymerized microtubules together with micronuclei in addition to a main nucleus (Kalay et al., 2011).

#### **1.3.5 Abnormal spindle-like microcephaly-associated protein (ASPM)**

The fifth gene responsible for the MCPH5 locus is *ASPM*. Homozygous mutations in this gene are the most common cause of MCPH (Bond et al., 2002, Bond et al., 2003, Pichon et al., 2004). The ASPM protein has been predicted to contain an N-terminal microtubule-binding domain, two calponin homology domains (common in actin-binding proteins), 81 Ile–Gln (IQ) repeat motifs which are predicted to undergo a conformational change when bound to calmodulin, and a C-terminal region of unknown function (Bond et al., 2002).

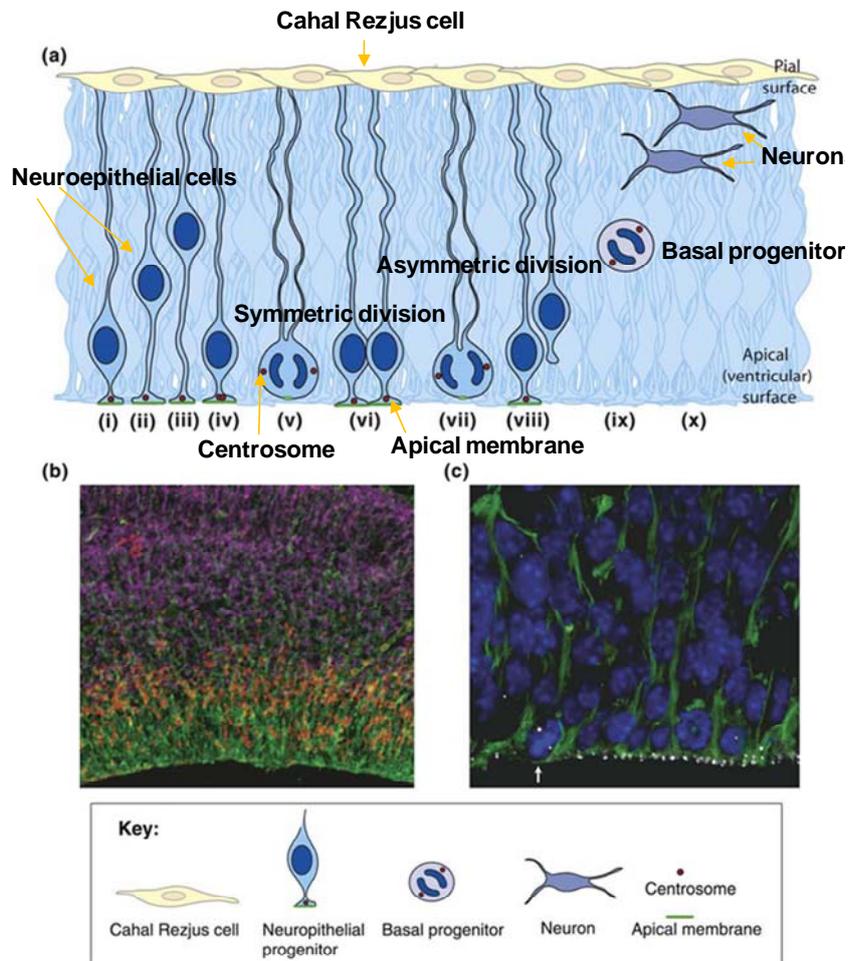
ASPM is also a member of a novel family of proteins that contain ASH domains (ASPM, SPD-2 and Hydin), which are commonly present in ciliary proteins. Based on the presence of ASH domains, the ciliary function of ASPM in embryogenesis has been uncovered, that is, it is involved in ciliary mediated movement of cerebrospinal fluid required for guidance of young migrating neurons (Ponting, 2006). But this function of the ASPM protein is not the cause of disturbed neurogenesis in MCPH because loss of ciliary functions of ASPM result in neuron-migration abnormalities not neuron number reduction (Cox et al., 2006).

Down regulation of endogenous ASPM by siRNA has been performed which resulted in the decrease of endogenous BRCA1 protein levels. ASPM localizes to the centrosome during interphase and to the spindle poles from prophase through telophase. These findings indicate that ASPM may be involved in mitotic spindle function possibly through regulation of BRCA1 (Zhong et al., 2005). ASPM localization was also confirmed in mouse neuroepithelial cells at the spindle pole throughout the cell cycle with only decreased intensity at telophase. In this study there was no centrosomal localization of ASPM during interphase reported (Fish et al., 2006). Immunostaining of HT1080 cells revealed that ASPM is localized at the spindle poles during mitosis. This finding suggests that MCPH is the consequence of impairment in mitotic spindle regulation in cortical progenitors due to mutations in ASPM (Kouprina et al., 2005).

The *ASPM* gene is the putative human ortholog of the *Drosophila melanogaster* 'abnormal spindle' gene (*asp*). Recessive mutants of *asp* lead to larval lethality or infertility with dividing neuron progenitors unable to conclude asymmetric cell division (Gonzalez et al., 1990). The *asp* protein is required for microtubule organisation of the mitotic spindle poles and the central spindle in mitosis and meiosis (Gonzalez et al., 1990, do Carmo et al., 2001). Based on the data of *asp*, the role of ASPM during neurogenesis might be to organize microtubules at the spindle pole during mitosis and at the central spindle during cytokinesis.

Among all identified MCPH proteins, ASPM is the only protein for which a role in neurogenesis has been directly assessed. The ventricular zone (VZ) of mouse neuroepithelium where proliferative dividing progenitor cells are found showed a high *Aspm* mRNA expression at the onset of neurogenesis and during the early stages decreasing progressively as neurogenesis proceeds. *Aspm* protein localizes also to mitotic spindle poles but not overlapping with  $\gamma$ -tubulin. Knockdown of *Aspm* in the mouse neuroepithelium by using RNAi technique did not affect cell-cycle progression or block mitosis. The only feature was a detachment of the centrosome during telophase. Instead, loss of *Aspm* resulted in an increased deviation of the cleavage plane in proliferative

neuroepithelial progenitors, causing almost 50% of them to bypass the apical membrane and resulting in unequal inheritance of this domain by the daughter cells (Figure 1.3). As a result of this bypass, an increase in neuron-like progeny was observed. This suggests that loss of *Aspm* disrupts the alignment of the cleavage plane, causing an increase in non-neuroepithelial progeny and depleting the progenitor pool prematurely (Fish et al., 2006).



TRENDS in Genetics

**Figure 1.3. The developing mouse neuroepithelium.** (a) Neuroepithelial cells (blue) are connected with apical (ventricular) and pial (basal) surfaces. The nuclei (dark blue) migrate basally during G1, cells (i,ii) at a basal position (iii); and migrate again apically during G2 (iv). The centrosomes (red circles) remain at the apical membrane (green). Mitosis occurs at the apical surface, where the centrosomes now form the spindle poles. Symmetrical division leads to the production of two identical neuroepithelial cells (v,vi). By contrast, asymmetrical division (vii) leads to the production of one neuroepithelial cell, whereas the other daughter detaches from the membrane (viii) and becomes either a basal progenitor (ix) or a neuron. Basal progenitors (ix) lack processes and polarity and predominantly divide terminally to produce two neurons (x). (b) Immunofluorescence in E14 mouse neuroepithelium showing the apical progenitors and their processes (nestin, green), basal progenitors (*Tbr2*, red) and neurons (*bIII Tubulin*, purple). (c) Immunofluorescence in E12 mouse neuroepithelium showing the apical progenitors and their processes (nestin, green), the nuclei of the apical progenitors (DAPI, blue) and the centrosomes ( $\gamma$ -tubulin, white). Note the arrow pointing to the cell in metaphase. The image is taken from (Thornton and Woods, 2009).

### 1.3.6 Centromere protein J (CENPJ)

The gene of the MCPH6 locus is *CENPJ* which encodes centromere-associated protein J (Bond et al., 2005). *CENPJ* was initially named centrosomal protein 4.1- associated protein (*CPAP*) and, despite its name, *CENPJ* is a centrosomal protein and interacts with the nonerythrocytes 4.1 protein 135 splice variant (4.1R-135) and is associated with the  $\gamma$ -tubulin complex (Hung et al., 2000). The centrosomal localization of this protein was confirmed by immunofluorescence analysis in HeLa cells (Bond et al., 2005). There are two main domains in *CENPJ*, a microtubule-binding domain (MBD) and a microtubule-destabilising domain (MDD). The microtubule-destabilizing motif (MDD) of *CENPJ* inhibits microtubule nucleation from the centrosome and is also involved in the depolymerization of taxol-stabilized microtubules (Hung et al., 2004). Knockdown of *CENPJ* in HeLa cells by siRNA demonstrated a cell cycle arrest in more than 40% of the cells with multiple spindle poles and induced apoptosis (Cho et al., 2006). Therefore, it has been speculated that *CENPJ* might control the production of centrosomal microtubules during neurogenic mitosis (Bond et al., 2005).

On the basis of data from *SAS-4* of *C. elegans*, which is the probable homologue of *CENPJ*, the novel role of *CENPJ* in centriole biogenesis has been elucidated, as the protein product of *SAS-4* is responsible for centriole duplication (Leidel and Gonczy, 2005). Another study performed in *Drosophila* indicated that loss of *DSas-4* resulted in loss of centrioles during embryonic development and slow mitotic spindle assembly and 30% abnormal asymmetric divisions of larval neuroblasts (Basto et al., 2006). A Knockout study of *sas4* showed abnormal spindle formation and DNA segregation defects due to loss of centrioles (Rodrigues-Martins et al., 2008).

Later on the direct role of *CENPJ* in centriole biogenesis was also described by different studies. Through siRNA-mediated depletion and immunoelectron microscopy directed towards individual centrosomal proteins it was found that *CENPJ*, *Plk4*, *hSas-6*, *Cep135*,  $\gamma$ -tubulin, and *CP110* were required at different stages of procentriole formation and were associated with different centriolar structures. *CENPJ* together with *CEP135* formed a core structure within the proximal lumen of both parental and nascent centrioles (Kleylein-Sohn et al., 2007). Another study showed that *CENPJ* is required for centrosome duplication in cycling human cells and its overexpression resulted in the formation of abnormal centrioles which were longer and in the production of more than one procentriole. These abnormal centrioles lead to multipolar spindle assembly and cytokinesis defects (Kohlmaier et al., 2009). Knockdown of *CENPJ* inhibited centrosome duplication and its overexpression induced the formation of elongated procentriole-like structures (Tang et al., 2009).

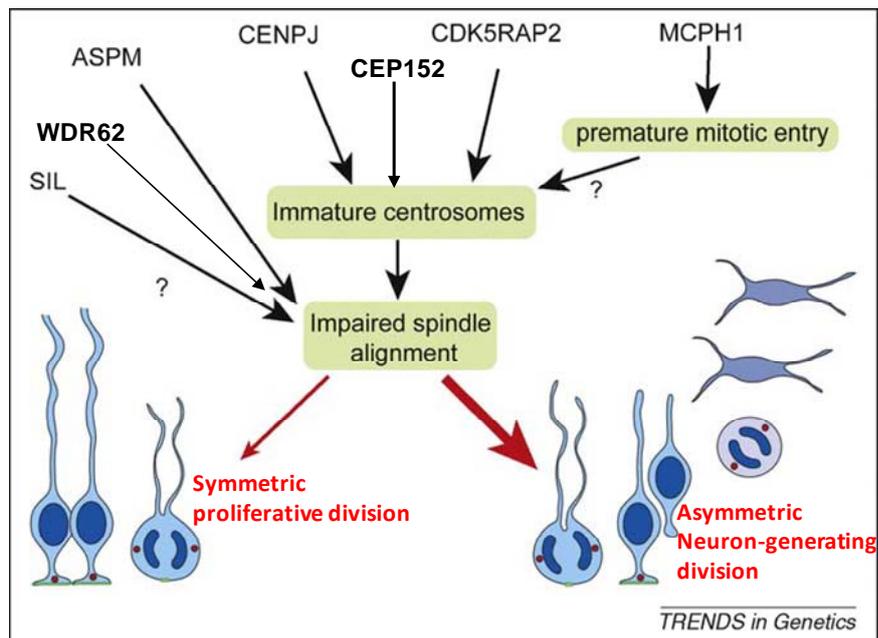
### 1.3.7 SCL-interrupting locus protein (SIL)

*STIL* was identified as the causative gene of the MCPH7 locus in Indian MCP families. In mouse the expression of this protein has been shown in subventricular neuroepithelial cells at embryonic day 14.5 (E14.5) (Kumar et al., 2009). Antibodies generated against human SIL revealed its localization in HeLa cells to the poles of the mitotic spindle in metaphase cells, but it was not detected during anaphase. The localization of SIL to the spindle pole is similar to the one of ASPM. Knockdown of SIL by shRNA in HeLa cells showed disorganized mitotic spindles in dividing cells. Localization and knockdown of SIL suggested the role of SIL in organizing the mitotic spindle (Pfaff et al., 2007). Striking developmental anomalies like pericardial swelling, midline neural tube defects, failure of neural tube closure and holoprosencephaly are the main phenotypes observed in SIL knockout mice. These phenotypes were observed between embryonic day E7.5 to E8.5, and the mutant homozygote die *in utero* at E10.5 (Izraeli et al., 1999).

## 1.4 Role of MCPH proteins in neurogenesis

Before discussing the role of MCPH proteins in neurogenesis, a brief description of neurogenesis in mouse is described. During mouse brain development, apical neuroepithelial (NE) cells are the first progenitor cells and have centrosomes which remain at the apical membrane during G1, S and G2 phase of the cell cycle. These NE progenitor cells are attached to the apical (ventricular) and pial surfaces, have an apico-basal polarity, and the orientation of the spindle is dependent on the positioning of the centrosomes. Early in neurogenesis, NE progenitor cells undergo symmetric proliferative division, producing the same types of NE cells, which stay attached with the apical and pial plasma membrane resulting in the increase of the progenitor pool. This symmetric division is due to the parallel alignment of spindles to the neuroepithelium within progenitor cells. As neurogenesis progresses, NE cells undergo asymmetric neurogenic division due to the alignment of the spindle perpendicular to the neuroepithelium. After asymmetric division one daughter cell stays attached with the apical surface and has the NE cell fate while the other daughter cell detaches from the apical surface and takes on the basal progenitor cell fate of the neuron. This basal progenitor cell is unpolarized and divides symmetrically to produce two neurons (Figure 1.3). Asymmetric cell division in NE cells maintains the progenitor pool cell numbers while producing committed precursors such as neuronal committed precursors. Progenitors give rise to various cell types such as neurons and glial cells (astrocytes, oligodendrocytes and microglia) that further differentiate and migrate. The cortical size depends not only on the generation, maturation and migration of cells, but also on the survival of generated cells. Also, the balance between symmetric proliferative division and asymmetric neurogenic division of

NE cells is considered to be important for the eventual numbers of neurons (Thornton and Woods, 2009, Kaindl et al., 2010).



**Figure 1.4: A model of MCPH protein functions during neurogenesis.** Loss of MCPH1 results in premature mitotic entry which affects centrosome maturation. Disruption of CDK5RAP2 or CENPJ has direct effects on centrosome maturation. Loss of CENPJ results in loss of the centriole. These immature centrosomes reduced PCM accumulation and can affect spindle orientation during division by formation of fewer astral microtubules. ASPM localizes to the spindle pole of the centrosomes and directly regulates spindle positioning. SIL and WDR62 also have spindle pole localization, suggesting a role for these proteins (SIL and WDR62) like ASPM. Impaired spindle alignment can result in symmetric divisions and reducing the progenitor pool, which ultimately results in fewer neurons. Image is taken from (Thornton and Woods, 2009).

Notable features of all identified MCPH proteins are their centrosomal localization and expression in progenitor cells of the mouse neuroepithelium. Defects in MCPH proteins can be associated with a delay in centrosome maturation or disrupt spindle orientation directly which affects the daughter centriole and centrosome, reducing the recruitment of pericentriolar matrix (PCM) and accessory proteins, and decreasing microtubular nucleation. This can result in the production of two unequal centrosomes and also altered spindle positioning. A defective spindle orientation resulting from the loss of MCPH proteins could then lead to an increase in NE cells producing neurogenic progeny and reducing the proportion of symmetric divisions thereby depleting the progenitor pool and limiting the total number of neurons that can be generated (Figure 1.4) (Thornton and Woods, 2009).

Identification of further proteins involved in MCPH will improve our knowledge of neurogenesis and the pathomechanism of this disorder.

## 2. Objectives of the Thesis

The overall aim of this doctoral thesis was the exploration and molecular characterization of the genes responsible for autosomal recessive primary microcephaly (MCPH).

Although seven MCPH genes have already been identified, there are still a large number of families which do not carry mutations in anyone of the known genes. Identifying the unknown genetic lesions in these families will on the one hand pave the way to a comprehensive diagnostic management of this highly heterogeneous disease including prenatal diagnosis and preconceptional carrier screening, on the other hand it will help to further elucidate the molecular mechanisms causing the characteristically disturbed brain development, i. e. to describe new signaling pathways and define essential interactions of various gene products in prenatal brain development especially in the cerebral cortex.

MCPH is a disorder of neurogenic mitosis and so far all identified MCPH proteins are expressed in neuronal progenitors. Thus, the identification of genes responsible for MCPH and the understanding of their functions might have direct relevance to other clinical spheres such as stem-cell therapy of neurological disorders.

The specific aims of this doctoral study were:

1. To analyse Pakistani families afflicted with primary microcephaly.
2. To screen for mutations in known *MCPH* genes to add to the mutational spectra of these genes.
3. To identify novel *MCPH* loci and the underlying causative gene variants.
4. To uncover the functions of the novel genes and their role in the etiology of MCPH.

### **3. Material and Methods**

#### **3.1. Subjects**

Thirty families of autosomal recessive primary microcephaly (MCPH) with 1-7 afflicted individuals were ascertained from remote locations of Pakistan. All families were consanguineous and diagnosed clinically with the prominent phenotype of a sloping forehead which is a defining feature of this disease, normal height and weight, mild to moderate mental retardation without any other neurological finding, such as spasticity, seizures, or progressive cognitive decline. The mental functions of the patients were assessed by standardized questionnaires in all family members. Clinical data in terms of head circumference (HC) was collected which is the common and simple method for evaluating gross brain size (Woods et al., 2005). Head circumference (HC) of at least three standard deviations (SDs) below the expected mean for age and sex was observed. The head circumference and intelligence level of all parents of microcephaly families were normal. Clinical data also indicated that the disease was present at birth in all thirty analyzed families of microcephaly. The phenotype of this disease was documented by photos (Figure 3.1).

#### **3.2. Pedigrees construction and mode of inheritance analysis**

The family history and mode of inheritance of all MCPH families was worked out by pedigree analysis which was constructed in the field by standard symbols. Pedigree data was given by family members. Blank squares and circles represent the normal male and female respectively while filled squares and circles represent the afflicted family members. A horizontal line between a circle and square is a marriage line. Children are represented by drawing a vertical line down from the marriage line. Fraternal twins are denoted by a pair of diagonal lines connected to the parents' offspring line, so that they each touch the horizontal line at the same point. In addition, a horizontal line between the two children is used to denote identical twins. A diagonal line across a circle or square indicates a deceased individual. An arrow pointing at a particular afflicted individual is added to indicate proband (also propositus for male and propoita for female).

In the laboratory, the pedigrees were drawn by using HaploPainter v.1.043 (Thiele and Nurnberg, 2005). This software is equipped with an intuitive graphical interface and is powerful enough to draw even complex consanguineous pedigrees. Filled symbols represent afflicted individuals and open symbols non-afflicted ones. Persons with an unknown afflicted status are shown in grey with a question mark inside. A diagonal slash indicates that the person is deceased. After pedigree constructions, the mode of inheritance of all 30 MCPH families was analyzed as autosomal recessive.



**Figure 3.1:** Clinical features of afflicted individuals of microcephaly families.

### **3.3. Blood sampling**

After informed consent, the peripheral blood samples of all available afflicted individuals, their parents and one or two normal, healthy siblings were collected and stored in 10 ml BD vacutainer® (BD-Plymouth, UK) having Ethylenediaminetetraacetic acid (EDTA) as anticoagulant.

### **3.4. DNA extraction**

DNA was extracted by using the phenol-chloroform method, a standard protocol of DNA extraction from blood. 0.75 ml of whole blood was mixed with same amount of solution A

in a 1.5 ml microcentrifuge tube (Eppendorf) and was incubated for 10 minutes at room temperature. The supernatant was discarded after 1 minute of centrifugation at 13,000 rpm. The pellet was resuspended in 400  $\mu$ l of Solution A and the sample was centrifuged for 1 minute at 13,000 rpm. The supernatant was discarded and the nuclear pellet was resuspended in 400  $\mu$ l of Solution B, 25  $\mu$ l of 10% SDS (13  $\mu$ l of 20% SDS), and 8  $\mu$ l of Proteinase K (20 mg/ml). Samples were incubated overnight at 37 or 65°C in a water bath for 3 hours. 0.5 ml of freshly prepared equal volume of solution C (250  $\mu$ l) and solution D (250  $\mu$ l) was added and the sample was centrifuged for 10 minutes at 13,000 rpm. The upper aqueous phase was collected in a new 1.5 ml tube and mixed with an equal quantity of solution D. This step of upper phase collection, solution D addition and centrifugation was repeated. To the collected aqueous phase, 55  $\mu$ l of 3M Na-acetate (pH 6) and an equal volume of isopropanol (500  $\mu$ l, stored at -20°C) was added. The tube was inverted gently to precipitate the DNA. The sample was centrifuged at 13,000 rpm for 10 minutes. The solvent phase was removed without disturbing the DNA pellet. To the tube containing the pelleted DNA, 350  $\mu$ l of 70% ethanol (stored at -20°C) was added and then the tube was centrifuged at 13,000 rpm for 10 minutes. Ethanol was removed and the DNA pellet was dried in a vacuum dryer. DNA was dissolved in 250  $\mu$ l of double distilled deionized water.

### **3.4.1. Solutions used for DNA extraction**

#### ***Solution A (pH 7.5) 500ml***

0.32 M Sucrose	54.72 g
10 mM Tris-HCl (pH 7.5)	5 ml (1M)
5 mM MgCl <sub>2</sub>	2.5 ml (1M)

The above mentioned chemicals were dissolved in 400 ml of dist. water and pH was adjusted to 7.5. A final volume of 495 ml was prepared by adding water. After autoclaving, the solution was cooled at room temperature. Finally 5 ml of Triton X-100 (1% v/v) was added.

#### ***Solution B, 500ml***

10 mM Tris-HCl (pH 7.5)	5 ml
400 mM NaCl	40 ml
2 mM EDTA (pH 8.0)	2 ml

The volume of the solution was brought to 500 ml and autoclaved.

#### ***Solution C (saturated phenol pH 8.0)***

Phenol (Sigma-Aldrich) was heated to 68°C in a water bath. The 100 ml of heated phenol were mixed with 100 mg of hydroxyquiniline (0.1%) and 100 ml of 0.5M Tris-HCl (pH 8.0).

Stirring was performed for 15 minutes. The upper aqueous phase was discarded. 0.1 M Tris-HCl (pH 8.0) of equal volume was added and stirred for 15 minutes. Again the upper phase was removed. This step was repeated until the pH of phenol reached 7.8. The pH of phenol was measured within the upper phase just above the phenolic layer. Then, 10 ml of Tris-HCl were added containing 200  $\mu$ l of  $\beta$ -mercaptoethanol. The solution was stored at 4°C and is stable for 1 month.

#### ***Solution D***

Isoamyl alcohol + chloroform (1:24)

### **3.5. DNA Quantification**

DNA was quantified by NanoDrop ND-1000 Spectrophotometer (Thermo Scientific) and a 2 ng/  $\mu$ l dilution of DNA was prepared for amplification with a specific set of primers.

### **3.6. RNA isolation**

#### **3.6.1. Blood**

The peripheral blood was collected in PAXgene Blood RNA Tubes to isolate cellular RNA from whole blood. RNA was isolated by using the PAXgene Blood RNA kit (PreAnalytiX).

#### **3.6.2. Human primary fibroblasts**

Total RNA was isolated from human fibroblast cells by using the RNeasy Mini kit (Qiagen)

### **3.7. Polymerase chain reaction (PCR) and agarose gel electrophoresis**

Polymerase chain reaction (PCR) was performed with patient and control DNAs using different PCR profiles and mastermix compositions depending upon the size of nucleic acid to amplify and the nature of primers (sequencing oligos or microsatellites). The primers can be found in appendix 1 to appendix 11. Different types of DNA polymerases were used to get different products. For sequencing oligos and microsatellite markers amplification with DNA of microcephalic patients, InnuTaq DNA Polymerase (AnalytikJena) was used, *PfuUltra*<sup>TM</sup> High-Fidelity DNA Polymerase (Stratagene) was employed to amplify wild-type and mutant *CDK6* from whole cDNA and Takara PrimeSTAR<sup>TM</sup> HS DNA Polymerase (TaKaRa, USA) was used for CEP135 wild-type and mutant constructs. Mastermixes were run on DNA Engine Tetrad 2 Thermal Cyclers, Version 2.0 (Bio-Rad). Different PCR profiles and mastermix compositions are mentioned in each respective section. 0.7 to 2% agarose (Invitrogen) was melted in TAE buffer and the gel was run in horizontal gel electrophoresis chambers (Horizon<sup>®</sup> 11.14) to resolve the nucleic acids. Quantification and size determination of the nucleic acids was done by HyperLadder I (BIOLINE) and 1 Kb DNA Ladder (Invitrogen). Ethidium bromide (0.1  $\mu$ g/ml

final concentration) was used to stain the nucleic acids. Images of the gels were taken by Alpha Innotech Multi Image® Light Cabinet (Biozym).

***TAE buffer, pH 8.3***

40 mM Tris base  
20 mM Acetic acid  
2 mM EDTA

***Gel loading dye (Bromophenol Blue)***

***For Bromophenol Blue production***, 40g of sucrose and 0.25g of Bromophenol Blue were dissolved in 80ml of dist. H<sub>2</sub>O. Afterwards the volume was increased to 100ml.

**3.8. Reverse transcriptase PCR (RT-PCR)**

cDNA was synthesized from whole RNA by using SuperScript III reverse transcriptase enzyme (Invitrogen). The procedure is as follows, 9µl RNase free ddH<sub>2</sub>O was mixed with 1µl dNTPs (10 mM), 1 µl Oligo d(T)18-Primers (50 µM) and 2µl RNA (665 ng/ µl) respectively in a nuclease-free microcentrifuge tube. The mixture was incubated for 5 minutes at 65°C to get rid of secondary structures of the RNA. Then, 4 µl 5X First-Strand Buffer, 2µl DTT (0.1 M) and 1µl SuperScript RT (200 U/µl) were added. The samples were incubated at 50°C for 50 minutes and 70°C for 15 minutes in a thermal cycler.

**3.9. Genotyping with short tandem repeat (STR) markers**

For linkage analysis, PCR (Polymerase Chain Reaction) based genotyping, was performed using highly polymorphic fluorescently labeled microsatellite (STR) markers. For genotyping, the microsatellite markers were selected for each locus from the Marshfield (<http://www.bli.uzh.ch/BLI/Projects/genetics/maps/marsh.html>) genetic map, and were labeled with fluorescent dyes e.g. FAM, TET, HEX and NED at 5' end of forward primers. Some markers of chromosome 19 were selected from Human STR Finder (<http://portal.ccg.uni-koeln.de/cgi-bin/repfinder.pl>) and designed by Primer3 v. 0.4.0 (Rozen and Skaletsky, 2000). The mastermix composition to amplify the microsatellite markers is listed in table 3.1.

**Table 3.1: Mastermix composition to amplify DNA of microcephaly patients with microsatellite (STR) markers**

	Final Concentration	Mastermix for one reaction (µl)	Mastermix for 8 reactions (µl)
DNA (2ng/µl)		4.00	
Innu Taq (5U/µl)	2.5U	0.05	0.40
dNTP's (10mM)	100 µM each	0.10	0.80
Primer-F (10µM)	0.25 µM	0.25	2.00
Primer-R (10µM)	0.25 µM	0.25	2.00
10xPCR-Buffer II (10X)	1.5Mg	1.00	8.00
ddH <sub>2</sub> O		4.35	34.8
<b>Total Volume</b>		<b>10.00</b>	<b>48.00</b>

**Table 3.2: PCR profiles to amplify the microsatellite markers of microcephaly.**

MS			TDM		
1	94°C	3 min.	1	95°C	5 min.
2	94°C	30 Sec.	2	95°C	15 Sec.
3	61°C	45 Sec.	3	65°C	30 Sec.
4	72°C	1 min.	<b>Decrease by 0.5°C every cycle</b>		
<b>Step 2-4 repeat 2 more times</b>			4	72°C	30 Sec.
5	94°C	30 Sec.	<b>Cycle to step 2 for 19 more times</b>		
6	59°C	45 Sec.	5	95°C	15 Sec.
7	72°C	1 min.	6	55°C	30 Sec.
<b>Step 5-7 repeat 2 more times</b>			7	72°C	30 Sec.
8	94°C	30 Sec.	<b>Cycle to step 7 for 19 more times</b>		
9	57°C	45 Sec.	8	72°C	10 min.
10	72°C	1 min.	9	4°C	Forever
<b>Step 8-10 repeat 2 more times</b>					
11	94°C	30Sec.			
12	55°C	45Sec.			
13	72°C	1min.			
<b>Step 11-13 repeat 2 more times</b>					
14	72°C	20 min.			
15	4°C	Forever			

## **PCR programmes**

Two different PCR profiles were used to amplify the patient and control DNAs for microsatellite markers. These profiles are given in table 3.2.

Amplified markers were electrophoresed on the ABI 3730 DNA Analyzer using standard fragment analysis protocols according to the manufacturer's specifications (Applied Biosystems, Foster City, CA). Fragment-length analysis was performed using the ABI GeneMapper® v.3.5 analysis packages. Haplotypes were constructed by HaploPainter v.1.043 (Thiele and Nurnberg, 2005) and EasyLinkage Plus v5.02 (Lindner and Hoffmann, 2005). If all afflicted individuals were homozygous and parents were heterozygous for specific markers of a region, the family was called to be linked with that particular locus. If all afflicted individuals and parents were heterozygous for specific markers of a region, the family was excluded for that particular locus.

### **3.10. DNA sequencing**

#### **3.10.1. Sanger sequencing**

Sanger sequencing was used to sequence different candidate genes. Different sets of primers were employed to amplify with desired DNA of microcephaly families. The mastermix composition is listed below in table 3.3.

**Table 3.3: Mastermix composition to amplify DNA of microcephaly patients for sequencing.**

	Mastermix for one reaction (µl)	Mastermix for eight reactions (µl)
DNA (2 ng/µl)	3.00	--
Taq (5 U/µl)	0.12	0.96
dNTP's (10 mM)	0.30	2.40
Primer-F (10 µM)	0.30	2.40
Primer-R (10 µM)	0.30	2.40
10xPCR-Buffer II (10x)	1.50	12.00
ddH <sub>2</sub> O	9.48	75.84
<b>Total volume</b>	<b>15.00</b>	<b>96.00</b>

Most of the DNA was amplified with a PCR profile named TDM and 60°C. TDM profile is mentioned in table 3.2 and 60°C in table 3.4a.

**Table 3.4: PCR profiles used for Sanger sequencing. (a)** For patient DNA amplification with an annealing temperature of 60°C. **(b)** ExoSAP reaction to clean up the PCR product ultimately used for Sanger sequencing. **(c)** Thermal cycler profile for BigDye® Terminator v1.1 Cycle sequencing.

<b>(a) For patient DNA Amplification</b>			<b>(b) For ExoSAP reaction</b>		
1	95°C	5 min.	1	37°C	30 min.
2	95°C	30 sec.	2	85°C	15 min.
3	60°C	30 sec.	3	4°C	Hold
4	72°C	30 sec.			
<b>Step 2-4 repeat 44 more times</b>			<b>(c) For Big dye reaction</b>		
5	72°C	5 min.	1	96°C	10 sec.
6	4°C	Forever	2	55°C	5 sec.
			3	60°C	4 min.
			Step 1-3 repeat 44 more times		

### 3.10.1.1. Cleaning of PCR product with Exo-SAP

Residual nucleotides from PCR products were removed with the help of the enzymes, Exonuclease I (20U/μl, New England Biolabs) and the Shrimp Alkaline Phosphatase (SAP) (1U/μl, Promega). The former enzyme removes leftover primers, while the later removes the unused dNTPs (Table 3.4b and Table 3.5a).

**Table 3.5: Mastermix composition used for Sanger sequencing. (a)** For PCR product clean up by ExoSAP **(b)** For BigDye® Terminator v1.1 Cycle sequencing kit.

<b>(a) ExoSAP mastermix ( μl)</b>		<b>(b) BigDye mastermix ( μl)</b>	
PCR product	8	Water	5.25
Exonuclease I (20U/μl)	0.075	BigDye® Terminator v1.1	0.50
SAP (1U/ μl)	0.300	5X BigDye sequencing buffer	2.00
Water	1.625	Primer (10μM)	0.25
<b>Total Volume</b>	<b>10.00 μl</b>	PCR template (cleaned)	2.00
		<b>Total Volume</b>	<b>10.0 μl</b>

### 3.10.1.2. BigDye® terminator cycle sequencing

Sequencing was performed using the BigDye® Terminator v1.1 Cycle sequencing kit (Applied Biosystem) together with an ABI3730/3730xl automated DNA sequencer (Applied Biosystem). Reaction mixture composition for BigDye® Terminator v1.1 is given above in table 3.5b. The thermal cycler programme was 96°C for 10 seconds 55°C for 5

seconds and 60°C for 4 minutes with a total of 32 cycles (Table 3.4c). Sequences were analysed by DNA Star (Lasergene) and Mutation Surveyor (SoftGenetics).

### **3.10.2. Pyrosequencing**

The polymorphic nature of pathogenic mutations (substitutions or deletions) found in patient DNAs was ruled out by sequencing the control samples of the same ethnic group by pyrosequencing which is, widely applicable and has the advantages of accuracy, flexibility, parallel processing, and easy automation. This method is based on the sequencing by synthesis principle and relies on the detection of pyrophosphate release on nucleotide incorporation.

#### **3.10.2.1. Primer design for pyrosequencing**

PCR primers were designed by the PSQ Assay Design program v.1.0.6 (Qiagen, Hilden, Germany). The forward primers used to amplify the product for pyrosequencing were Biotin labeled at the five prime ends (Appendix 8, Appendix 9).

#### **3.10.2.2. Preparation of PCR products for pyrosequencing**

10 µM of sequencing primer (0.36 µl for one reaction) was mixed with 1x annealing buffer (11.64 µl for one reaction). PCR products were mixed with 70 µl of streptavidin sepharose (GE Healthcare) and shaken for 6 minutes. The PCR product and streptavidin sepharose mix was added to a solution containing sequencing primer and annealing buffer with the help of the PyroMark Q96 Vacuum Prep Workstation. This primer-hybridization mix was incubated at 85°C for 2 minutes and pyrosequencing was done according to manufacturer's instructions on a PSQ HS96A instrument (Qiagen, Hilden, Germany). Pyrosequencing was conducted using PyroMark Gold Q96 Reagents (Qiagen, Hilden, Germany). This kit contains enzyme (DNA Polymerase, ATP sulfurylase, luciferase and apyrase), dNTPs (dATP, dCTP, dGTP and dTTP) and the substrate, adenosine 5' phosphosulfate (APS) and luciferin. Pyrosequencing data was analysed by Pyro Q-CpG V.1.0.9 analysis software (Qiagen, Hilden, Germany).

#### **3.10.2.3. Solutions used for pyrosequencing**

##### ***1x annealing buffer (1x AB), pH 7.6***

20mM Tris-Acetate

2mM Mg-Acetate

### **3.11. Genome-wide linkage analysis**

Excluded consanguineous MCPH families were subjected to genome-wide linkage analysis by using the Affymetrix GeneChip® Human Mapping 250K SNP Array. Genotypes were called by the GeneChip® DNA Analysis Software (GDAS v3.0,

Affymetrix). Sample genders were verified by counting heterozygous SNPs on the X chromosome. Relationship errors were evaluated with the help of the program Graphical Relationship Representation (Abecasis et al., 2001). The program PedCheck was applied to detect Mendelian errors (O'Connell and Weeks, 1998) and data for SNPs with such errors was removed from the data set. Non-Mendelian errors were identified by using the program MERLIN (Abecasis et al., 2002) and unlikely genotypes for related samples were deleted. Linkage analysis was performed assuming autosomal recessive inheritance, full penetrance, consanguinity and a disease gene frequency of 0.0001. The multipoint LOD score was calculated using the program ALLEGRO (Gudbjartsson et al., 2000). Haplotypes were reconstructed with ALLEGRO and presented graphically with HaploPainter (Thiele and Nurnberg, 2005). All data handling was performed using the graphical user interface ALOHOMORA (Ruschendorf and Nurnberg, 2005).

### 3.12. Mutation detection

All the genes residing in minimum linkage intervals of excluded families (MCP4, MCP50, MCP53, MCP63 and MCP67) were screened by NCBI database and prioritized by using Endeavour (Aerts et al., 2006) and GeneWanderer (Kohler et al., 2008). Endeavour is a software used to prioritize the candidate genes of a specific disease, based on a set of training genes while *GeneWanderer* is a computational method to prioritise a set of candidate genes by probability to be involved in a particular disease or phenotype. Highly ranked gene structures were searched using ENSEMBL (<http://www.ensembl.org>) and UCSC (<http://genome.ucsc.edu>) genome databases. All exons and the intron-exon boundaries of highly ranked genes were first sequenced in two afflicted individuals and one parent and one normal sibling. Mutations were found by analyzing the sequence using DNA Star (Lasergene) and Mutation Surveyor (SoftGenetics).

### 3.13. Bacterial culture

The solution of Luria Bertani (LB) media prepared with deionised water was sterilised by autoclaving at 120°C. The final concentration of antibiotics used for LB media was 100 µg/ml of Ampicillin and 50 µg/ml Kanamycin.

#### ***LB medium (1L) (pH 7.0)***

Tryptone enzymatic digest from Casein (Fluka, BioChemika)	10 g
Yeast extract (Fluka, BioChemika)	5 g
NaCl	5 g

#### ***LB agar plates (1L) (pH 7.0)***

Tryptone enzymatic digest from Casein (Fluka, BioChemika)	10 g
Yeast extract (Fluka, BioChemika)	5 g
NaCl	5 g

### 3.13.1. Storing the bacterial culture

1.0 ml overnight bacterial culture was mixed with 1.0 ml glycerol (room temperature and stored at -80°C).

### 3.14. Mammalian cell lines and cell culture

**COS7** cell line was obtained by immortalizing a CV-1 cell line derived from kidney cells of the African green monkey with a version of the SV40 genome that can produce large T antigen but has a defect in genomic replication (Jensen et al., 1964, Gluzman, 1981).

**HaCaT** is a Human skin keratinocyte cell line (Boukamp et al., 1988).

**Primary fibroblast cell lines** were established from biopsies of patients carrying mutations in *CDK6* and *CEP135* and control.

All cell lines were cultured in Dulbecco's Modified Eagle's Medium (DMEM, PAA supplemented with 10% fetal bovine serum (FBS, Biochrom), L-Glutamine (PAN Biotech) and antibiotics (Penicillin/Streptomycin, PAN Biotech)). Cells were cultivated at 37°C in a humidified incubator supplied with 5% CO<sub>2</sub>.

#### **Composition of media used for mammalian cell culture**

DMEM (High Glucose (4.5 g/L) with L-Glutamine)	500 ml
FBS	60 ml
L-Glutamine (200 mM, 100X)	6 ml
Penicillin/Streptomycin (10.00 IE (µl)/ml)	6 ml

**Note:** For CEP135 primary fibroblasts, 15% FBS was used.

### 3.15. Cell line establishment from patient biopsies

The biopsy samples were taken from one afflicted individual of each MCP4 (individual MCP4-3) and MCP63 (individual MCP63-1) family and transported in DMEM with more additives like Amphotericin B and Gentamycin. Tissue samples were cleaned with antiseptic agent (Betaisodonna, Mundipharma) and incubated overnight with Dispase II (1.5 U/ml, Roche) diluted in PBS at 4°C to separate the epidermis from the dermis. The dermis was incubated in DMEM at 37°C. After one week, primary fibroblast cells were seen crawling from the dermis. The number of primary fibroblasts was increased by further culturing. Primary fibroblasts established from the MCP63 patient biopsy grew very slowly. To enhance growth, DMEM with 15% FBS was used.

### **3.16. Fibroblast synchronization at G0/G1 by serum starvation**

CDK6 patient and control fibroblasts (passage 4) were synchronized at G0/G1 by serum starvation. The medium was exchanged by two washes with PBS and then DMEM without FBS was added. After 48 hours, the medium was replaced by DMEM containing 15% FBS. The cells were then processed for immunofluorescence after 24, 26, 28 and 30 hours.

### **3.17. CDK6 and CEP135 GFP constructs**

Gateway® Technology (Invitrogen) was used to clone wild-type and mutant *CDK6* and *CEP135* cDNA.

#### **3.17.1. CDK6 GFP constructs**

mRNA from *CDK6* wild-type (NM\_001145306.1) and mutant (c.589G>A) primary fibroblasts was isolated and converted into cDNA (procedure is described in section 3.8) and the *CDK6* sequences cloned into the entry vector pENTR™/TEV/DTOPO® by using the pENTR™ Directional TOPO® Cloning Kit to get the entry clones. Expression clones of wild-type and mutant *CDK6* were achieved by performing an LR recombination reaction (Gateway® LR Clonase™ II Enzyme Mix, Invitrogen) between the entry clone and a Gateway® destination vector pcDNA-DEST53 (Invitrogen) which has an N-terminal cycle 3 GFP tag.

#### **3.17.2. CEP135 GFP constructs**

*CEP135* wild-type transcript (NM\_025009.3) isolated from wild-type primary fibroblasts was converted into cDNA and cloned into entry vector (procedure is the same as described in section 3.8). The *CEP135* mutation (c.970 delC) was produced in the wild-type entry clone by using the QuikChange® II Site-Directed Mutagenesis Kit (Stratagene). The QuikChange Primer Design tool (Stratagene) was used to design the mutagenesis primers (Table 3.6). The mutant entry vector was propagated in XL1-Blue competent *E. coli* cells instead of SURE 2 cells. Both wild-type and mutant expression clones were acquired as listed for the *CDK6* GFP constructs.

#### **3.17.3. Sequencing of entry and expression clones**

Sequencing of entry clones was done by using vector specific (M13 sequencing primers) and internal primers, while for expression clones, the primers were designed within the vector and the inserts. The primers are listed in appendix 11 and the protocol for sequencing is described in section 3.10.

**Table 3.6: Primers used for amplification of *CDK6*, *CEP135* and for site directed mutagenesis of *CEP135*.**

Primers used to amplify <i>CDK6</i>	
<b><i>CDK6</i>-cDNA-F</b>	CACCATGGAGAAGGACGGCCTG
<b><i>CDK6</i>-cDNA-R</b>	GCAGAGCCTGTCCAGAAGAC
Primers used to amplify <i>CEP135</i>	
<b><i>CEP135</i>-cDNA-F</b>	CACCTTAGAAGACGAGATGACTACAGC
<b><i>CEP135</i>-cDNA-R</b>	CCATGTTTCATGATTTTCAGCA
Primers for <i>CEP135</i> site directed mutagenesis	
<b><i>CEP135-del970-F</i></b>	ACTCTGCCAAGAATTAAGTAAA TAGATAGTTAGCACAGCAGT
<b><i>CEP135-del970-R</i></b>	ACTGCTGTGCTAACTATCTATTTTC AGTTAATTCTTGGCAGAGT

### 3.18. Plasmid purification

Correct clones identified by sequencing were cultured overnight in 100 ml LB medium with 100 µl Kanamycin (50 mg/ml). Plasmids were extracted by PureYield™ Plasmid Midiprep System (Promega).

### 3.19. Transfection of GFP-tagged clones

The plasmids (expression clones) (10 µg/µl) were used for transfection of COS7 cells (80% confluent and trypsinized one day before transfection) by electroporation. The COS7 cells were incubated on ice for 5 minutes and taken up in 189 µl normal cell culture medium. 200 µl (11 µl plasmid DNA plus 189 µl cells) were filled into a pre-chilled cuvette (0.4 mm gap) and incubated on ice for 15 minutes. Cells were electroporated using a GENE PULSER II (BIO-RAD) and spread in prewarmed cell culture medium. Immunofluorescence and protein was extracted on the third day, 72 hour after transfection.

### 3.20. Knockdown of *CDK6*

*CDK6* gene silencing was performed by shRNA (short hairpin RNA) technique.

#### 3.20.1. Primer design

To knockdown *CDK6*, two sets of primers were designed by taking 31 nucleotides from each of exon 4 (5'-GATCTAAAACACAGAACATTCTGGTGACCA-3') and 7

(5'-TGAGAAGTTTGTAACAGATATCGATGAACTA-3') of the *CDK6* gene along with a control sequence (5'-TTAAAAGAGCTAGTTTAACTAATGGGAACTG-3'). This control sequence is a randomized sequence of *CDK6*. The Hairpin Protocol V2.1 ([http://hannonlab.cshl.edu/protocols/BseRI-BamHI\\_Strategy.pdf](http://hannonlab.cshl.edu/protocols/BseRI-BamHI_Strategy.pdf)) was used to design the primers. The primers are listed in appendix 6.

### 3.20.2. Oligonucleotides annealing

The synthesized forward and reverse single-stranded oligonucleotides (100  $\mu$ M) were annealed using the following conditions.

#### **PCR mastermix**

Primer A-F (100 $\mu$ M)	4 $\mu$ l
Primer A-R (100 $\mu$ M)	4 $\mu$ l
5X Annealing Buffer	2 $\mu$ l

#### **PCR profile**

95°C for 30 second.  
72°C for 2 minutes  
37°C for 2 minutes  
25°C for 2 minutes  
4°C for Hold

This PCR ultimately yields double stranded oligos at a concentration of 50  $\mu$ M (assuming 100% theoretical annealing) which are ready for ligation into the pSHAG-1 vector. The oligos are diluted with TE Buffer at the ratio of 1:100 (2  $\mu$ l oligo + 198  $\mu$ l TE buffer)

#### **5x Annealing buffer (50 ml)**

50 mM Tris/HCl, pH 8.0	2.5 ml
250 mM NaCl	2.5 ml
5 mM EDTA	500 $\mu$ l

### 3.20.3. Digestion of pSHAG-1 vector

3  $\mu$ g of pSHAG-1 vector (Paddison et al., 2002) were linearized using *BseRI* and *BamHI* that generate ends compatible with the target oligonucleotide sequence. Two control digestion, samples were digested with one enzyme each and analysed by gel electrophoresis (Table 3.7).

**Table 3.7: Mastermix composition of pSHAG-1 and control sample digestion**

pSHAG-1-digestion		Control digestion		
<b>Vector (pSHAG-1)</b>	3µl (1µg/µl)		<i>Bam</i> HI	<i>Bse</i> RI
<b><i>Bam</i>HI</b>	2µl (20u/µl)	pSHAG-1	0.2µl (1µg/µl)	0.2µl (1µg/µl)
<b><i>Bse</i>RI</b>	2µl (10u/µl)	<i>Bam</i> HI	0.5µl (20u/µl)	
<b>Buffer 2</b>	10µl	<i>Bse</i> RI		0.5µl (20u/µl)
<b>BSA (NeBiolabs)</b>	1µl (100X)	Buffer 2	2µl	2µl
<b>Water</b>	82µl	BSA	0.2µl (100X)	0.2µl (100x 10mg/ml)
<b>Total Volume</b>	<b>100µl</b>	Water	16.6µl	16.6µl
		<b>Total Volume</b>	<b>20µl</b>	<b>20µl</b>

The mixture was incubated for 4 hours at 37°C. The digested product was analyzed by agarose gel electrophoresis and was purified by High Pure PCR Product Purification Kit (Roche).

#### 3.20.4. Cloning procedure

Double stranded oligonucleotides (0.5 µM) were ligated into linearized the pSHAG-1 Vector (15ng/µl) as described below and incubated overnight at 15-17°C (Table 3.8).

**Table 3.8: Mastermix composition to clone ds shRNA oligos into to pSHAG-1**

pSHAG-1 Vector (15ng/ul)	4µl
Insert (ds shRNA Oligos)	1.3µl
T4 Ligase Buffer	4µl
T4 DNA Ligase (1U/ µl)	1µl
Nuclease Free Water	9.7µl
<b>Total Volume</b>	<b>20µl</b>

10 µl of ligation reaction was mixed with *E. coli* GT116. Incubation was conducted on ice for 15 minutes. The cells were shocked by heating for 90 seconds at 42°C. 1 ml of SOC medium was added followed by 1 hour shaking at 37°C. 50 and 100 µl of this mix were spread on LB agar plates containing Kanamycin. The plates were incubated at 37°C

overnight. Colonies (6-10) were picked and plasmids were extracted (PureLink™ Quick Plasmid Miniprep Kit) and analysed by sequencing.

### **3.20.5. Plasmid purification**

Correct clones were cultured overnight in 100 ml LB medium containing 100 µl Kanamycin (50 mg/ml). Plasmid DNA was isolated using PureYield™ Plasmid Midiprep System (Promega).

### **3.20.6. Transfection**

The plasmids (4 µg/µl) were used for transfection of HaCaT (80% confluent) cells by using the Amaxa Cell line Nucleofector (R) Kit V (LONZA). Transfection was carried out twice with three days interval. Immunofluorescence analysis and preparation of protein lysates was done on the third day after the second transfection.

### **3.21. Preparation of protein lysates from eukaryotic cells**

Cells were washed twice and scraped into ice cold 1x phosphate buffered saline (PBS) plus protease inhibitor (DTT, Benzamidine and PMSF at 1 mM each). After centrifugation at 15,000 rpm at 4°C the pellet was resuspended in lysis buffer (50 mM Tris/HCl, pH 7.5, 150 mM NaCl, 1% Nonidet P-40, 0.5% Na-desoxycholat, 0.1% SDS, Proteinase Inhibitor Cocktail (PIC, Sigma) and protease inhibitors DTT, Benzamidine and PMSF). The amount of lysis buffer should be 2-4 fold volume of the pellet. The solution was passed several times through 0.4x19 mm needles (27Gx3/4", Nr.20, BD Microlane TM3) and was incubated on ice for 15 minutes. The sample was denatured in 5x SDS sample buffer at 95°C for 5 minutes.

### **3.22.SDS-polyacrylamide gel electrophoresis (SDS-PAGE)**

Proteins were resolved according to their molecular weight by using denaturing SDS-PAGE. The MiniPROTEAN® (Biorad) system was used for gel casting and for electrophoresis. The composition of the resolving and stacking gel is mentioned in table 3.9. The gel was polymerized for 30 minutes at room temperature. To smoothen the surface of the gel and to remove any air bubbles at the top of the resolving gel, Isopropanol was overlayed. The stacking gel was applied to the top of the resolving gel after complete removal of isopropanol, and left to polymerise for 30 minutes. The gels were assembled and placed into the Biorad Mini-PROTEAN® Gel Chamber and 1× SDS running buffer was added. The samples were loaded into the wells of the gel together with PageRuler™ Plus Prestained Protein Ladder (Fermentas) and Low Molecular Weight (LMW) marker (GE Healthcare, UK ) and were run at 80-120 V until the Bromophenol blue dye of the samples reached the bottom of the gel.

### 3.22.1. Solutions used for SDS-PAGE

<b>5X SDS-sample buffer (10 ml)</b>	<b>1X SDS-running buffer (pH 8.3)</b>	
0.125M Tris/HCl (pH 6.8)	2.5 ml [0.5M]	25 mM Tris-base
4% SDS	4 ml [10%]	192 mM Glycine
20% Glycerol	2 ml	0.1% SDS
10% $\beta$ -mercaptoethanol	1 ml	
Bromophenol blue	0.005% (w/v)	

**Table 3.9: Composition of SDS-PA gels for both the resolving and stacking gel.**

	Resolving gel		Stacking gel
	10%	12%	4%
1.5M Tris/HCl (pH 8.8)	2.5ml	2.5ml	
0.5M Tris/HCl (pH 6.8)			2.00ml
s30% Polyacrylamide	3.34ml	4.0ml	1.33ml
10% SDS	100 $\mu$ l	100 $\mu$ l	100 $\mu$ l
Water	4.0ml	3.40ml	6.56ml
10%APS	22 $\mu$ l	22 $\mu$ l	30 $\mu$ l
TEMED	10 $\mu$ l	10 $\mu$ l	16 $\mu$ l

### 3.23. Staining of polyacrylamide gels with Coomassie-Brilliant-Blue R 250

Proteins separated on polyacrylamide gels were stained with Coomassie-Brilliant-Blue R 250 solution for 20 minutes at room temperature. The excess staining was removed with destaining solution.

#### **Coomassie-Blue stain solution (1L)**

Brilliant-Blue R250 (Sigma)	2.5 g (Final conc. 0.25%)
Ethanol absolute	500 ml (Final conc. 50%)
Acetic acid	100 ml (Final conc. 10%)

#### **Destain solution (1L)**

Acetic acid	70 ml (Final conc. 7%)
Ethanol absolute	200 ml (Final conc. 20%)

### 3.24. Protein transfer to membranes (Western blot)

Wet blotting was performed to transfer the proteins from gels to a nitrocellulose membrane (PROTRAN<sup>R</sup>, Germany). The nitrocellulose membrane was soaked in blotting buffer together with thick blotting papers for 15 minutes. The gel and membrane were

sandwiched between sponge and paper and all were clamped tightly together after ensuring that no air bubbles had formed between the gel and the membrane. The order of gel and membrane from anode to cathode was blotting paper, membrane, gel and blotting paper. The sandwich is submerged in transfer buffer. Transfer was at 4°C overnight, and an electrical field of 10-15 V was applied depending on the size of protein. The negatively-charged proteins travel towards the positively-charged electrode, but the membrane stops them, binds them, and prevents them from continuing on.

***Wet blot buffer (1L)***

Glycine (193 mM)	14 g
Tris-base (25 mM)	3 g
Absolute ethanol (20%)	50 ml

**3.25. Immunodetection of proteins bound to the membrane**

The nitrocellulose membrane was blocked with 5% milk powder (non-fat dry milk) solution prepared in 1X TBS-T buffer for one hour to prevent unspecific binding of proteins. After blocking, the membrane was incubated with primary antibody diluted in TBS-T for overnight at 4°C. After washing three times each for 5 minutes with 1X TBS-T buffer, the membrane was incubated with an appropriate peroxidase-coupled secondary antibody (1:10,000 dilution) for one hour. Afterwards, the membrane was washed three times each for 5 minutes the blots were developed using the enhanced chemiluminescence (ECL) system. The membrane was exposed to an X-ray film and developed by using developer and fixer solutions.

***10x Tris-buffered saline (TBS) (500ml)***

1 M Tris/HCl (pH 7.4)	100 ml
5 M NaCl	137 ml

The volume was increased to 500 ml by adding water

***1x Tris-buffered saline-Tween 20 (TBS-T) 500 ml***

10XTBS Buffer	50 ml
Tween 20 (Sigma)	5.0 ml
Water	445 ml

The solution was pipet up and down to dissolve Tween-20

***Enhanced chemiluminescence (ECL) detection solution (20ml)***

Luminol (0.25 M)	200 µl
p-Cumaric acid (0.09 M)	89 µl
1M Tris/HCl (pH 8.5)	2 ml

The volume was increased to 20 ml by adding water

Add 6.1 µl of 30% H<sub>2</sub>O<sub>2</sub> directly before use.

### **3.26. Immunocytochemistry**

Patients and control primary fibroblasts were cultured on 12 mm coverslips and fixed with 3% paraformaldehyde. Permeabilization was done by 0.5% Triton X-100 in 1X PBS buffer for 4 minutes at room temperature. Subsequently the fixed cells were treated three times with 1X PBS. Blocking was done for 15 minutes with blocking buffer (1x PBG: PBS containing 5% BSA and 0.45% fish gelatine). Primary antibodies were diluted in blocking buffer and incubation was done overnight at 4°C. The following antibodies were used: mouse monoclonal CDK6 antibody (Abcam, ab54576) 1:200, rabbit polyclonal anti CEP135 (Abcam, ab75005) 1:500, rabbit polyclonal anti CEP135 (Bethyl Laboratories, Inc. A302-250A) 1:500, rabbit polyclonal anti Pericentrin (Abcam, ab4448) 1:300, mouse monoclonal anti- $\gamma$ -tubulin (Sigma, GTU-88) 1:300, rat monoclonal anti Y/L1/2 (Kilmartin et al., 1982) 1:20. After incubation with 1X PBS three times for 5 minutes, secondary antibodies (1:1,000 diluted in blocking buffer) were added for one hour at room temperature. As secondary antibodies, Alexa Fluor 568 goat anti-mouse IgG (Invitrogen, A11004), Alexa Fluor 647 donkey anti-rabbit IgG (Invitrogen, A31573), Alexa Fluor 647 donkey anti-mouse IgG (Invitrogen, A31571), Alexa Fluor 488 goat anti-rat IgG (Invitrogen, A11006) and Alexa Fluor 568 goat anti-rabbit IgG (Invitrogen, A11011). DNA was detected with 4',6-Diamidino-2'-phenylindole (DAPI) (Sigma, D9564). Finally the cells were mounted on glass slides with gelvatol. Images were taken by confocal microscopy (Leica, LSM TCS SP5).

#### **3.26.1. Immunostaining of microtubules**

For staining microtubules, cells were fixed in 3% paraformaldehyde prepared in tubulin stabilization buffer for 15 minutes and permeabilized in 0.5% Triton X-100 in 1X tubulin stabilization buffer. All washing steps before blocking were performed with 1X tubulin stabilization buffer. Blocking and staining steps are as described in "immunocytochemistry".

#### **3.26.2. Immunostaining of $\gamma$ -tubulin**

For  $\gamma$ -tubulin detection, the cells were permeabilized with prechilled absolute methanol for 10 minutes at -20°C. Blocking and staining steps are same as described in immunocytochemistry.

#### **3.26.3. Solutions used for immunocytochemistry**

##### ***3% Paraformaldehyde***

3 g PFA were dissolved in 80 ml 1x PBS by heating this solution up to 50°C, while stirring 0.4M NaOH was added drop wise until PFA was dissolved. The pH was adjusted to pH 6.1 and the volume was filled up to 100 ml with H<sub>2</sub>O.

### ***Gelvatol***

Poly(vinyl alcohol) 87-89% hydrolyzed,  $[-CH_2CHOH-]_n$ , MW 13-23,000 (Sigma Aldrich, P8136). 4.8 g gelvatol in 24 ml glycerol (50%) was added. After stirring gently for several hours, 24 ml Tris/HCl (0.2M), pH 8.5, was added. The solution was heated to 50°C for 10 minutes with gentle stirring, until it was clarified. Centrifugation was conducted at 5,000 g for 15 minutes in 50 ml-falcon tubes, then 1.3 g of DABCO (=1,4-diazobicyclo-[2.2.2]-octane) (Sigma Aldrich, D2522) was added. Aliquots were prepared in 2 ml-syringes and stored at -20°C.

### ***10x Triton X-100***

2.5 ml of Triton X-100 was dissolved in 20 ml of 1x PBS. As Triton X-100 is very viscous, it takes some time to dissolve. After dissolving, the falcon tube was centrifuged to determine the exact quantity of the solution and the volume adjusted to 25 ml by adding more 1x PBS.

### ***0.5% Triton X-100***

2.5 ml of 10x Triton X-100 were dissolved in 47.5 ml of 1x PBS. Both stock and working dilution were stored at 4°C.

**Note:** For microtubule staining, Paraformaldehyde and Triton X-100 were prepared in 1x-tubulin stabilization buffer instead of PBS (Phosphate Buffered Saline).

### ***1x Tubulin-stabilisation buffer (TSB) (1L) (pH 6.1)***

Hank's Buffer (10x)	100 ml
1 mM MES	10 ml (0.1 M)
2 mM EGTA	20 ml (0.1 M)
2 mM MgCl <sub>2</sub>	2 ml (1.0 M)

The pH was adjusted with HCl (20%).

## Hank's buffer

Table 3.10: Composition of Hank's Buffer

	1x (g)	5x (g)	10x (g)
137mM NaCl	4.0	20	40
5mM KCl	0.186	0.93	1.86
1.1mM Na <sub>2</sub> HPO <sub>4</sub>	0.078	0.93	0.78
0.4mM KH <sub>2</sub> PO <sub>4</sub>	0.0272	0.136	0.272
5mM Glucose	0.45	2.25	4.5
4mM NaHCO <sub>3</sub>	0.168	0.84	1.68

### 3.27. Immunohistochemistry of paraffin-embedded embryo sections

Wild-type mouse embryos at E11.5 and E15.5 were fixed with 4% paraformaldehyde overnight. Embryos were washed with PBS (without Ca/Mg), stored in 70% ethanol and embedded in paraffin. Paraffin-embedded sections (6  $\mu$ m) were de-paraffinized and heat-mediated antigen retrieval was performed. Sections were blocked with PBG for 2 hours at room temperature. Primary mouse-monoclonal CDK6 (Abcam, ab54576), rabbit-polyclonal CEP135 (Abcam, ab75005) and chicken anti-nestin (Neuromics, CH23001) antibodies diluted in PBG were incubated with the sections overnight at 4°C. Sections were washed with PBS, and incubated with fluorescent conjugated secondary antibodies (Invitrogen-Molecular Probes) for 1 hour. Nuclei were visualized with 4', 6-Diamidino-2'-phenylindole (DAPI). After 3 washes with PBS, sections were mounted and imaged with a Leica confocal microscope.

### 3.28. Confocal microscopy

The images were taken by confocal laser scanning microscopy (Leica, LSM TCS SP5) with 60 and 100X objectives. Images were analysed by *Leica-TCS-SP5/LAS-AF-Lite\_2.0.2\_2038* and processed using Photoshop.

### 3.29. Antibodies used for immunofluorescence and western blotting

Antibodies	Source	Dilution
<b>Primary antibodies for immunofluorescence</b>		
Mouse monoclonal CDK6 antibody	(Abcam, ab54576)	1:200
Rabbit polyclonal anti CEP135	(Abcam, ab75005)	1:500
Rabbit polyclonal anti CEP135 (Bethyl Laboratories, Inc. A302-250A)		1:500
Rabbit polyclonal anti Pericentrin	(Abcam, ab4448)	1:300
Mouse monoclonal anti- $\gamma$ -tubulin	(Sigma, GTU-88)	1:300

Rat monoclonal anti Y/L1/2	(Kilmartin et al., 1982)	1:20
Chicken anti-nestin	(Neuromics, CH23001)	1:200

***Secondary antibodies for immunofluorescence***

Alexa Fluor 568 goat anti-mouse IgG	(Invitrogen, A11004)	1:10,000
Alexa Fluor 647 donkey anti-rabbit IgG	(Invitrogen, A31573)	1:10,000
Alexa Fluor 647 donkey anti-mouse IgG	(Invitrogen, A31571)	1:10,000
Alexa Fluor 488 goat anti-rat IgG	(Invitrogen, A11006)	1:10,000
Alexa Fluor 568 goat anti-rabbit IgG	(Invitrogen, A11011)	1:10,000

***Primary antibodies for western blotting***

Mouse monoclonal CDK6 antibody	(Abcam, ab54576)	1:500
Rabbit polyclonal anti CEP135	(Abcam, ab75005)	1:500
Mouse monoclonal anti- $\beta$ -Actin	(Sigma, A5316)	1:10,000

***Secondary antibodies for western blotting***

Anti-mouse IgG peroxidase conjugated	(Sigma, A4416)	1:10,000
Anti-rabbit IgG peroxidase conjugated	(Sigma, A6154)	1:10,000

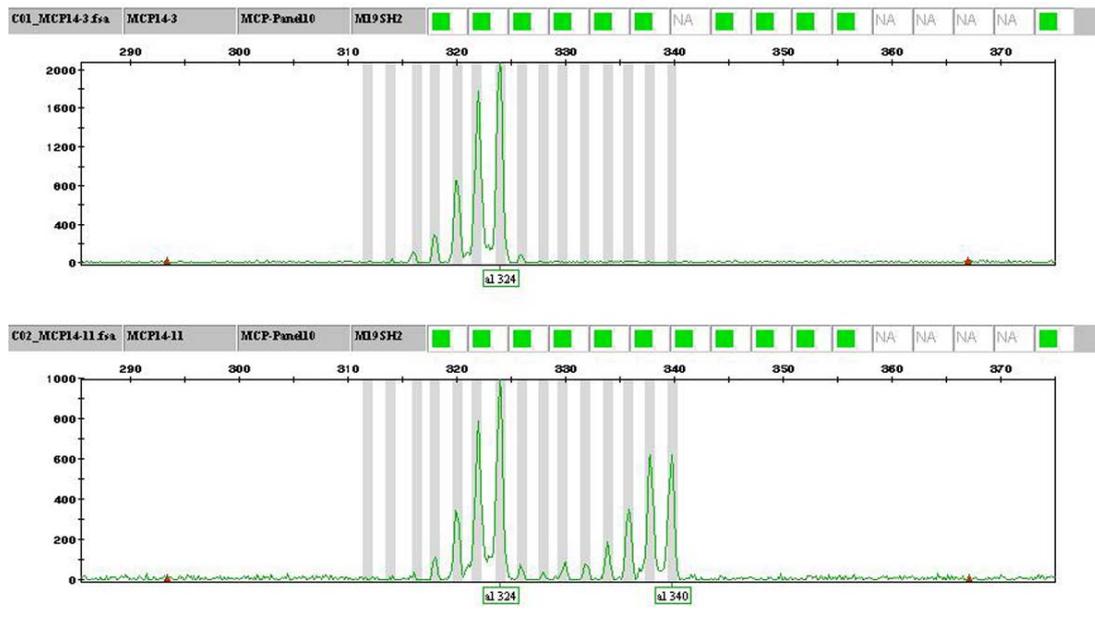
***DNA staining agent***

4',6-Diamidino-2'-phenylindole (DAPI)	(Sigma, D9564)	
---------------------------------------	----------------	--

## 4. Results

### 4.1. Genotyping of primary microcephaly families

PCR based genotyping of 30 autosomal recessive primary microcephaly families was performed by using highly polymorphic fluorescently labelled microsatellite markers of each known MCPH locus (Figure 4.1).

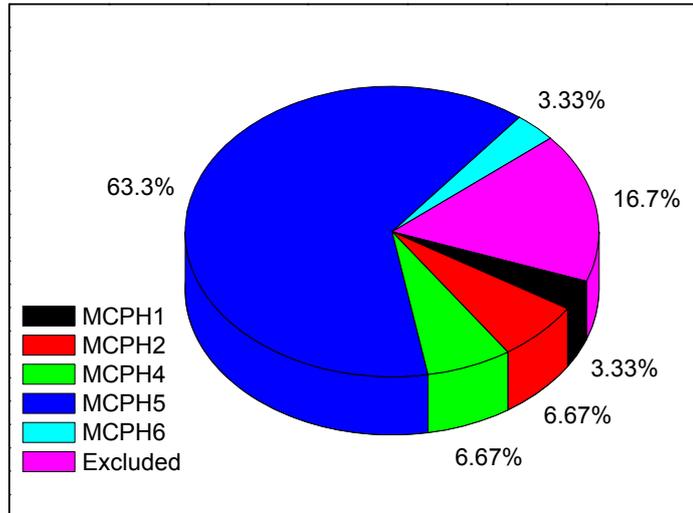


**Figure 4.1:** Electropherograms of the M19SH2 marker analyzed within the MCP14 family with the ABI GeneMapper® v.3.5. MCP14-3 is the patient and shows homozygosity for the M19SH2 marker with allele size 324 bp, while MCP14-11 is a heterozygous normal sib having one mutant allele, size 324 bp and one normal allele, size 340 bp.

Homozygosity mapping of 30 families revealed linkage of 19 families to MCPH5, of one family to MCPH1, of two families to MCPH2, of two to MCPH4 and of one to the MCPH6 locus. Five families were not linked to any of the known MCPH loci (Table 4.1). MCPH heterogeneity clearly shows that the majority of families were linked to MCPH5 locus (Figure 4.1).

**Table 4.1: Linkage analysis of thirty primary microcephaly families.**

	MCPH1	MCPH2	MCPH4	MCPH5	MCPH6	Excluded	Total
<b>No. of families</b>	<b>1</b>	<b>2</b>	<b>2</b>	<b>19</b>	<b>1</b>	<b>5</b>	<b>30</b>



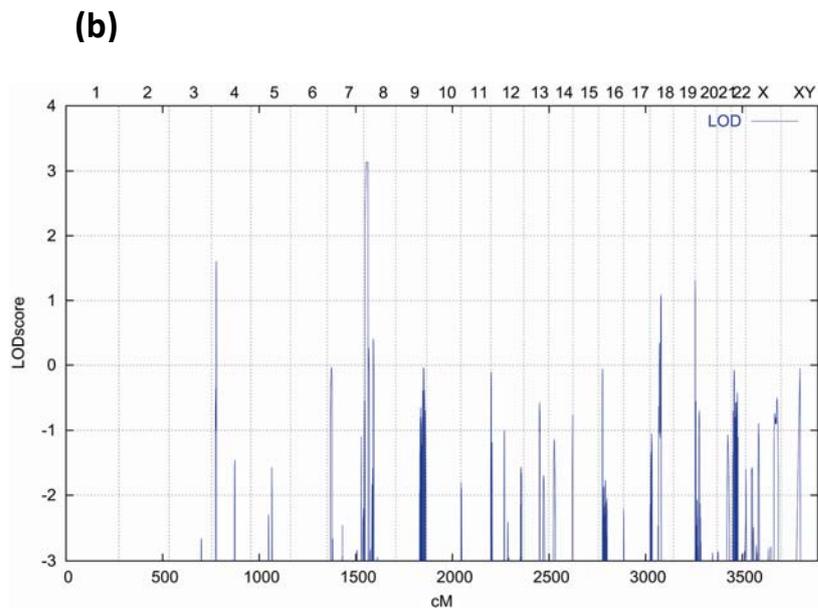
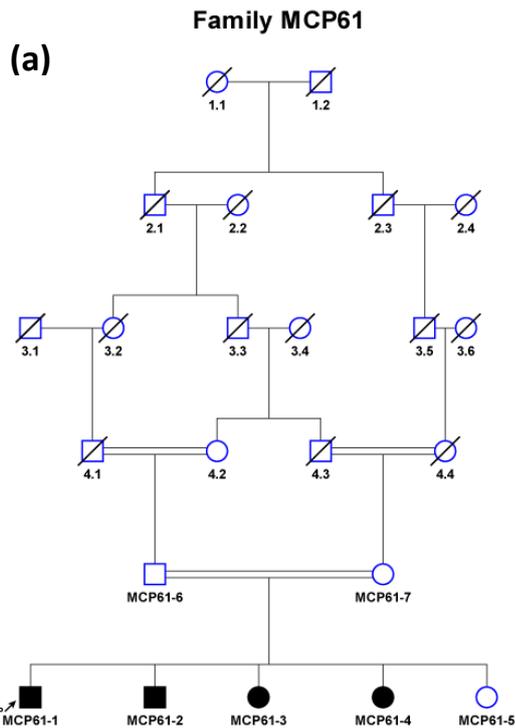
**Figure 4.2:** A pie chart illustrating the linkage status of MCPH families. Among 30 primary microcephaly families, 63.3% (blue colored) were linked with MCPH5 locus, 3.33% (black) with MCPH1, 6.67% (red) with MCPH2, 6.67% (green) with MCPH4, 3.33% (cyan) with MCPH6 and 16.7% (magenta) were excluded with all known MCPH loci.

#### 4.2. MCPH1 linked family

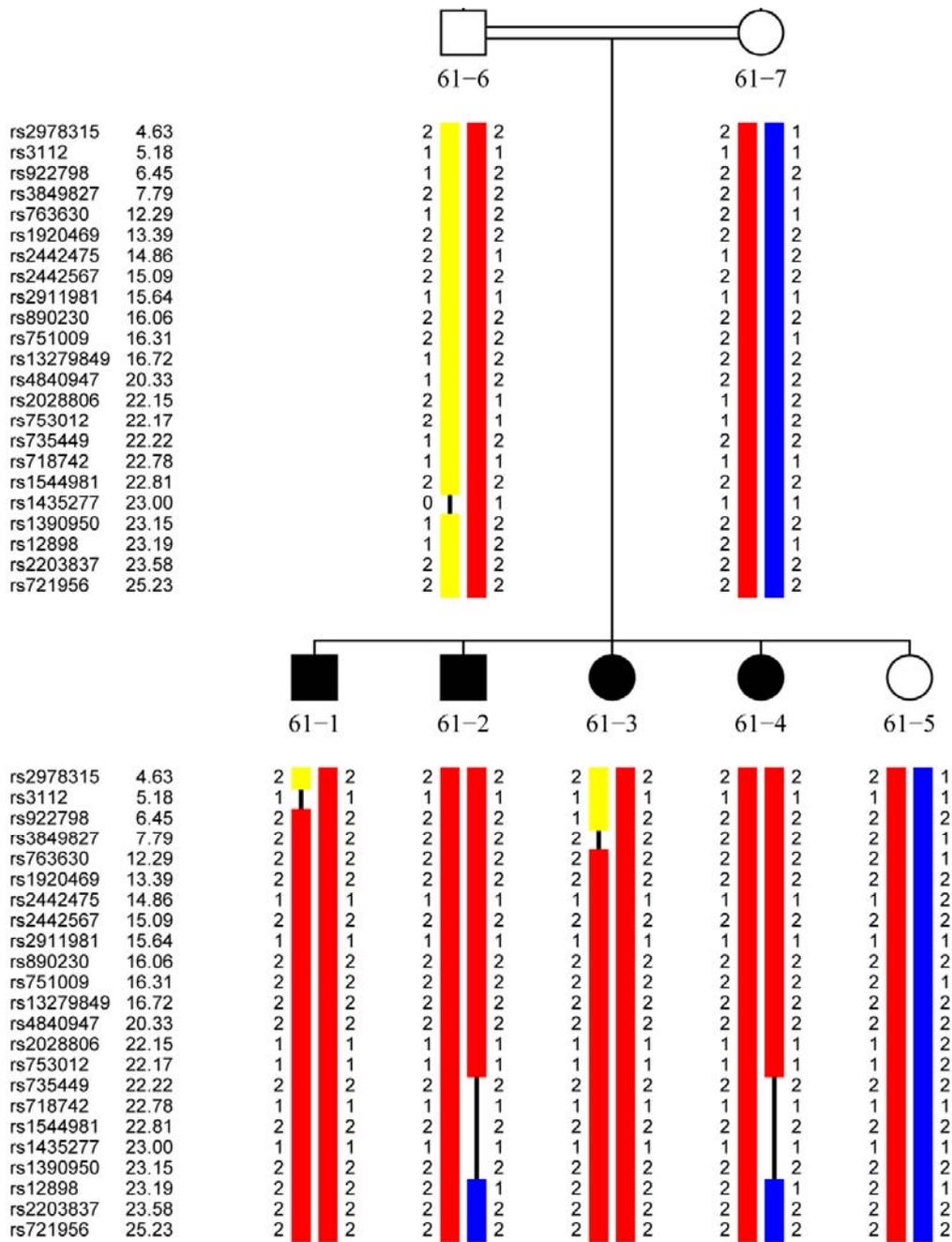
The MCP61 family is a large consanguineous family with four afflicted members (Figure 4.3a). All four patients with normal parents and one normal sib were subjects of this study.

Genome-wide linkage analysis was performed prior to linkage or exclusion with microsatellite markers in family MCP61. Genome-wide screens for regions of linkage with Affymetrix GeneChip(R) Human Mapping 250K SNP Array (version 2.0) show the linkage to the MCPH1 locus. Parametric linkage analysis showed the peak on chromosome 8 with maximum possible LOD score of 3.13 (Figure 4.3b).

This region maps between SNPs rs922798 (6.45cM and 3,481,618bp) and rs12898 (23.19cM and 11,701,198bp) and results in a 8.21 Mb region (Figure 4.4). The MCPH1 gene also resides in this region. The 14 exons of *MCPH1* were sequenced in all afflicted members, one normal sister and both parents of the MCP61 family. One novel mutation (c.1178delG) was found in this family and could not be detected in 200 healthy control chromosomes of the same ethnic group (Figure 4.5). Deletion of a single nucleotide G (c.1178delG) causes a frameshift incorporating 49 amino acids before resulting in a stop codon (p.R393Sfs\*50).

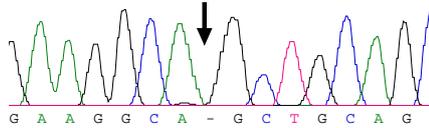


**Figure 4.3: Pedigree and genome-wide graphical view of the MCP61 family. (a)** Pedigree of family MCP61 with four afflicted sibs and one normal sib along with the parents, showing an recessive mode of inheritance. **(b)** Genome-wide graphical view of the MCP61 locus and the result of parametric linkage analysis showing LOD scores (3.13) on chromosome 8.

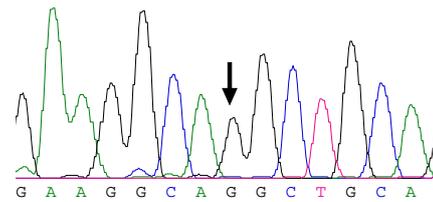


**Figure 4.4:** The region of chromosome 8 showing the haplotype of the MCP63 family linked with the MCPH1 locus (*MCPH1* gene).

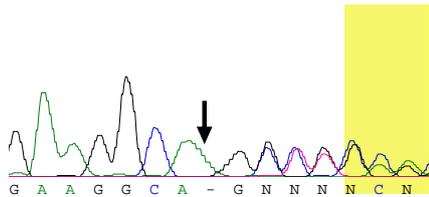
### Homozygous mutant c.1178delG (MCP61-1)



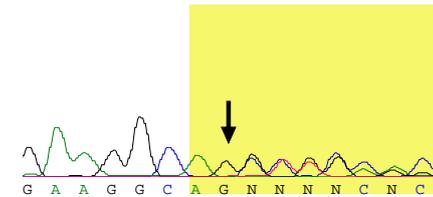
### wild type



### Heterozygous mutant c.1178delG (MCP61-6)



### Heterozygous mutant c.1178delG (MCP61-7)

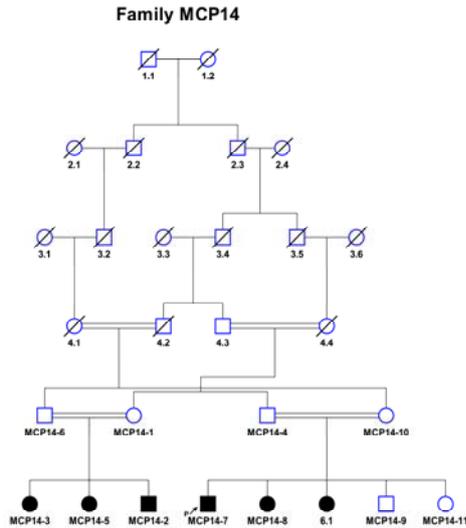


**Figure 4.5: Sequencing chromatograms of c.1178delG.** This mutation was identified in exon 8 of *MCPH1*. Black arrows indicate the position of mutations where G is deleted. Lower chromatograms represent the heterozygous parents (MCP61-6 and MCP61-7).

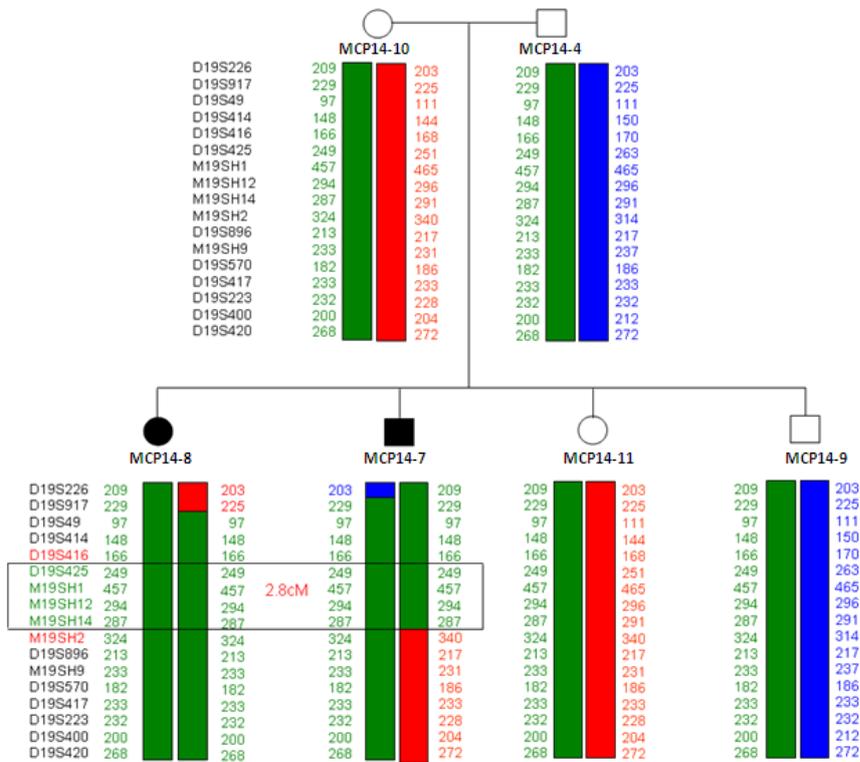
### 4.3. MCPH2 linked families

Two primary microcephaly families, MCP14 (Figure 4.6) and MCP59 (Figure 4.9), were linked with the MCPH2 locus which was originally mapped at chromosome 19q13.1-13.2 between the polymorphic markers D19S416 and D19S420. This 7.6 cM region was resolved in the MCP14 family. To achieve this, highly polymorphic microsatellite markers were selected by using the Human STR Finder (<http://portal.ccg.uni-koeln.de/cgi-bin/repfinder.pl>) and by designing primers with Primer3 v. 0.4.0 (Rozen and Skaletsky, 2000) within the region of the already described locus. The linkage interval was narrowed from 7.6 cM to 2.8 cM, which is 1/3 of the reported region. The previously reported region has its upper boundary at marker D19S416 and the lower at marker D19S420. I have detected heterozygosity at marker M19SH2 (61.5cM) in the MCP14-7 afflicted individual which defines the lower region of the MCPH2 locus at M19SH2 (61.5cM). So the new minimum linkage interval is defined by the markers D19S416 (58.7cM) and M19SH2 (61.5cM) at chromosome 19q13.11-19q13.12 which is 2.8 cM (NCBI Human Genome Build 36.3) (Figure 4.7).

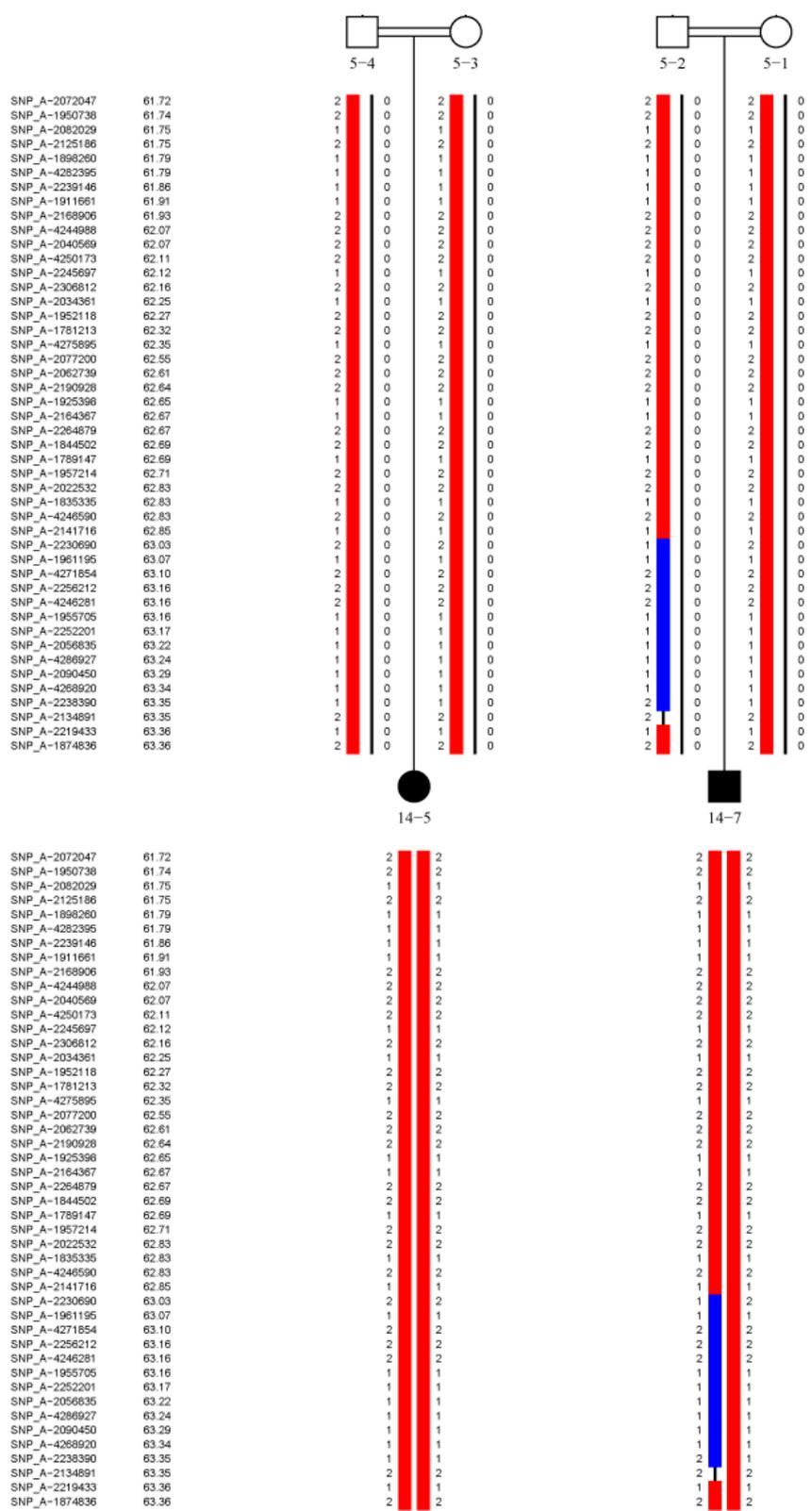
To confirm this reduced linkage interval, homozygosity mapping was performed with the help of SNP marker of chromosome 19 which define the correct reduced linkage interval of the MCPH2 locus mapped between SNP\_A-2299150 (rs3786913, physical position 38,678,914bp) and SNP\_A-2230690 (rs2271844, physical position 41,592,850bp) (NCBI Human Genome Build 36.3) (Figure 4.8). This 2.91 Mb region is the reduced linkage interval of the MCPH2 locus.



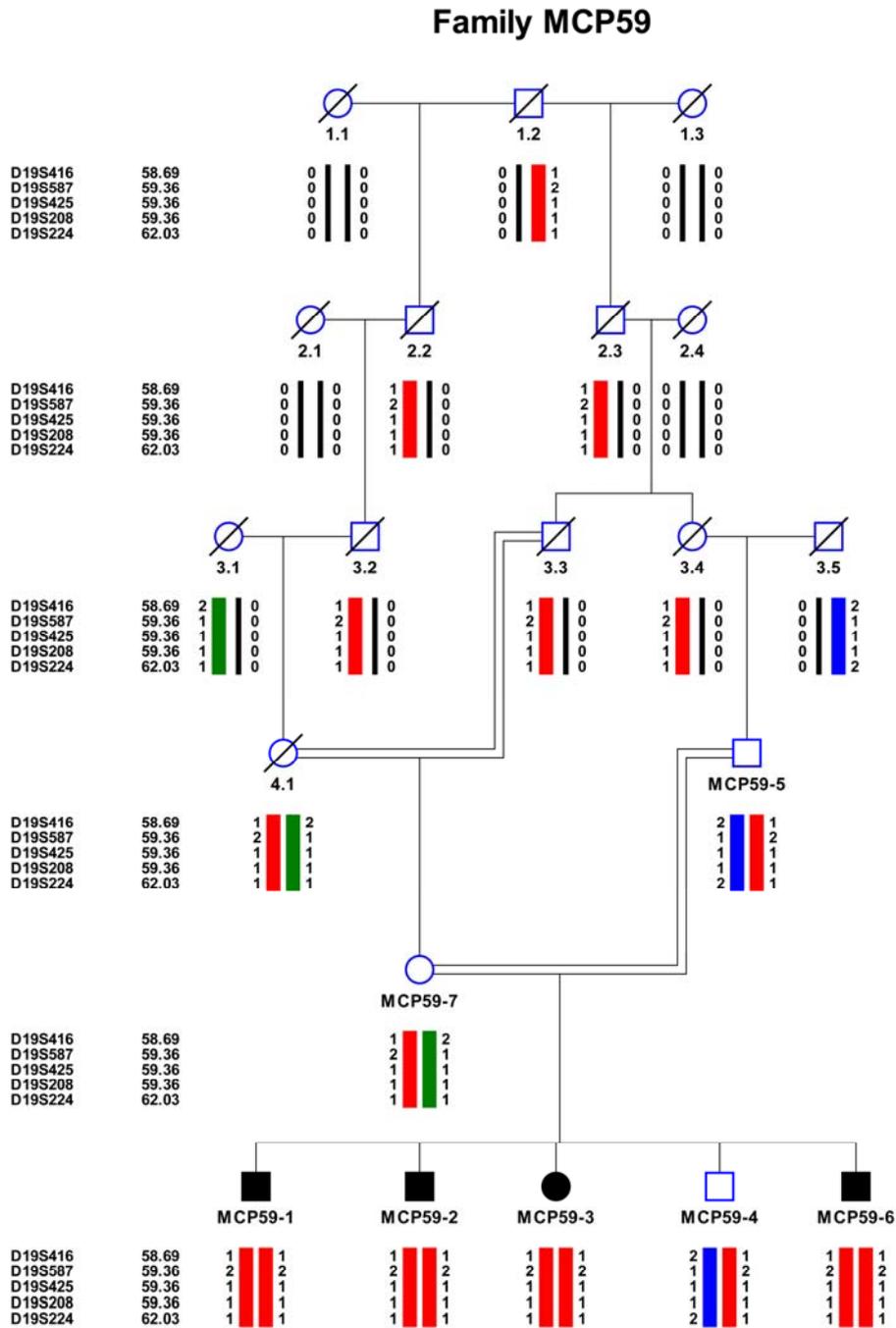
**Figure 4.6: Pedigree of the MCPH2 linked family.** MCP14 is a large consanguineous family having six afflicted individuals.



**Figure 4.7: Haplotype of the family MCP14 linked to the MCPH2 locus.** The already defined locus (between markers D19S416 and D19S224 which was 7.6 cM) was narrowed to 2.8 cM between markers D19S416 and M19SH2. The data for the second loop of this family is not shown here.



**Figure 4.8: Haplotype analysis of family MCP14 with the MCPH2 locus.** Marker names and positions are displayed on the left side of each generation. The disease haplotype fragments are indicated with red boxes.



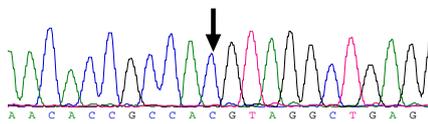
**Figure 4.9: Pedigrees of family MCP59 with haplotypes of the MCPH2 locus.** Marker names and positions are displayed on the left side of each generation. The disease haplotype fragments are indicated with red boxes.

There are 89 genes in this 2.91 Mb region (NCBI Human Genome Build 36.3). Among these 89, the following genes were selected and sequenced: myelin associated glycoprotein (*MAG*), sodium channel, voltage-gated, type I, beta (*SCN1B*), amyloid beta (A4) precursor-like protein 1 (*APLP1*), carbohydrate (N-acetylgalactosamine 4-0) sulfotransferase 8 (*CHST8*), glucose phosphate isomerase (*GPI*), CAP-GLY domain containing linker protein 3 (*CLIP3*), potassium channel tetramerisation domain containing 15 (*KCTD15*), and ubiquitin-like modifier activating enzyme 2 (*UBA2*). No mutation was

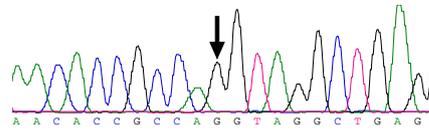
found in any of the above mentioned genes except for one substitution (IVS7+13C>T) in the intronic region of *MAG* in the MCP14 family. Family MCP59 was negative for any kind of mutation in *MAG*.

Recently *WDR62* (position 41,237,623 bp to 41,287,852 bp NCBI Human Genome Build 36.3) has been identified as a causative gene of the MCPH2 locus (Yu et al., 2010; Nicholas et al., 2010). I sequenced our MCPH2 linked families for *WDR62* which comprises 32 exons. The sequencing data demonstrates the identification of two novel mutations in the *WDR62* gene (Figure 4.10). In the MCP14 family a missense mutation (c.332G>C) replaces C with G which results in the change of Arginine to Threonine (p.R111T) in exon3 of *WDR62*. In MCP59, the nonsense mutation c.3503G>A introduces a premature stop codon p.W1168\* in exon 29 of *WDR62*.

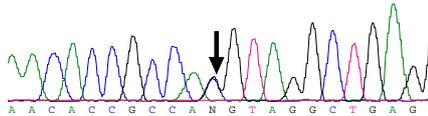
**Homozygous mutant c.332G>C (MCP14-7)**



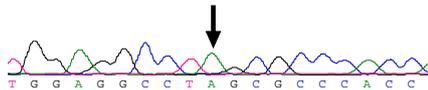
**Wild type**



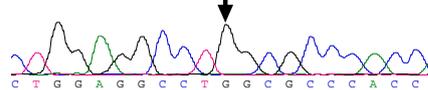
**Heterozygous mutant c.332G>GC (MCP14-1)**



**Homozygous mutant c.3503G>A (MCP59-1)**



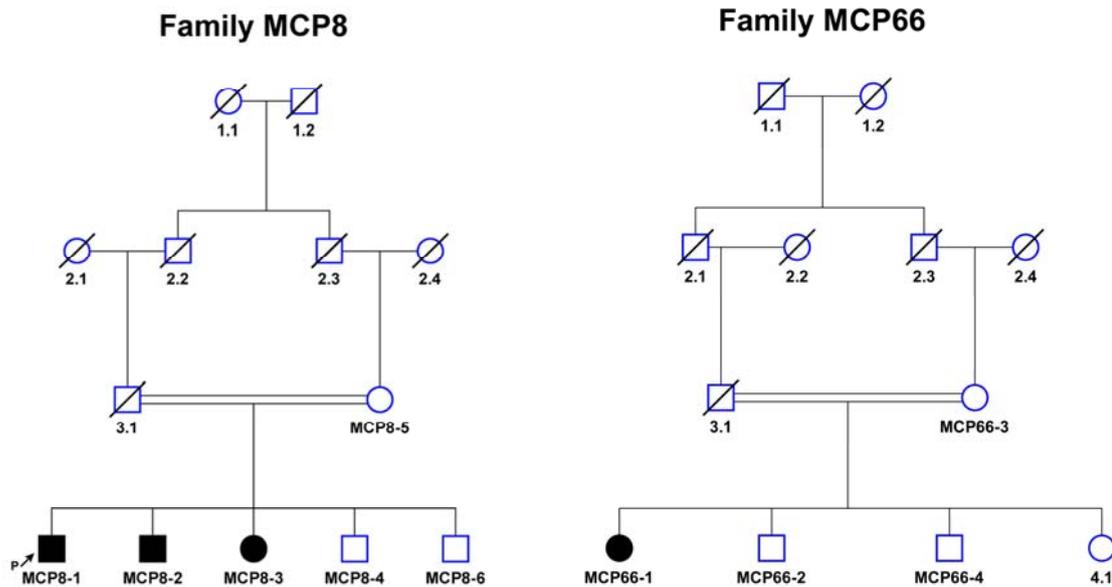
**Wild type**



**Figure 4.10: Sanger electropherograms for the *WDR62* gene showing variants c.332G>C and c.3503G>A in MCP14 (upper traces) and in MCP59 (lower traces) respectively. Black arrows indicate the position of the mutations.**

#### 4.4. MCPH4 linked families

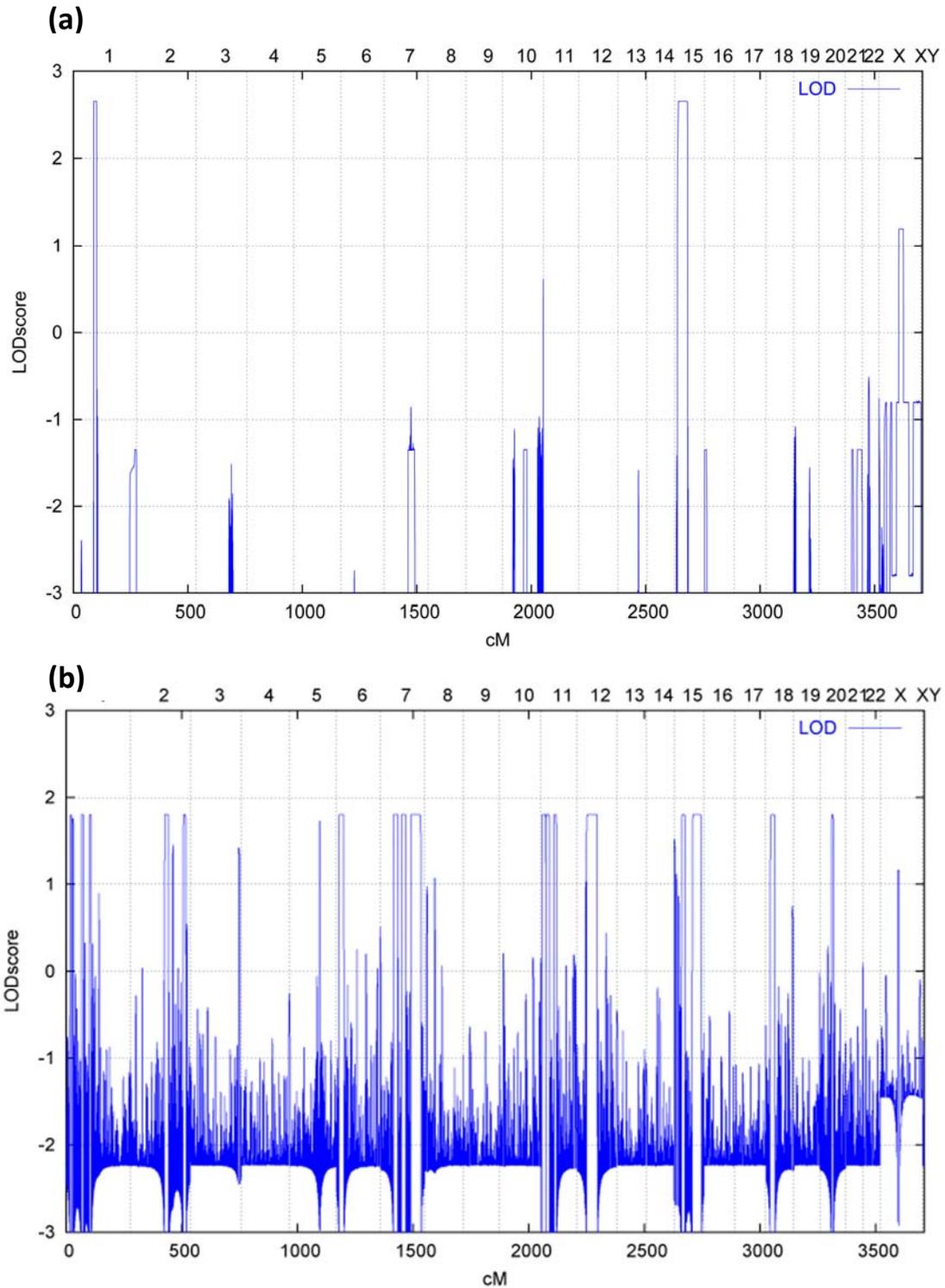
Two families MCP8 and MCP59 were linked with the MCPH4 locus. Both families are small and consanguineous (Figure 4.11).



**Figure 4.11: Pedigrees of the MCPH4 linked families.** MCP8 and MCP66 contain three and one afflicted individual respectively and show recessive modes of inheritance.

Homozygosity mapping of the MCP8 and the MPC66 families with microsatellite markers supported the linkage of these families to the MCPH4 locus which was further confirmed by genome-wide linkage analysis with the Affymetrix GeneChip(R) Human Mapping 250K SNP Array (version 2.0) (Figure 12).

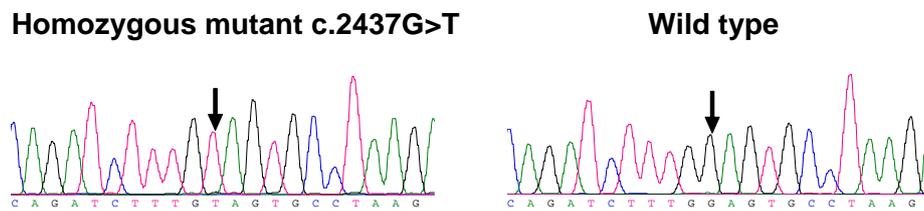
The already reported MCPH4 locus mapped on chromosome 15q15-q21 within 19 cM observed between markers ACTC and D15S98. Genome-wide linkage data of the MCP8 and MPC66 families did not reduce the linkage interval. Genome-wide analysis of three further families from Belgium and Denmark reduced this 19 cM critical linkage interval to 2.08 Mb mapped between rs936532 (37,585,407 bp) and rs1819454 (39,669,675 bp). This newly defined region of 2.08 Mb on chromosome 15q14-15q15.1 harboured 59 genes (NCBI Human Genome Build 36.3). Candidate genes were selected for sequencing which have mitotic functions, centrosomal or spindle pole localization, embryonic brain expression or play a role in neurogenesis. Among these 59 genes, *BUB1B*, *PAK6*, *INOC1*, *ITPKA*, *NUSAP1*, *OIP5* and *THBS1* were sequenced. Tubulin gamma complex associated protein 4 (*TUBGCP4*) which is located in the larger 19 cM region was also sequenced. No mutation was found in these genes.



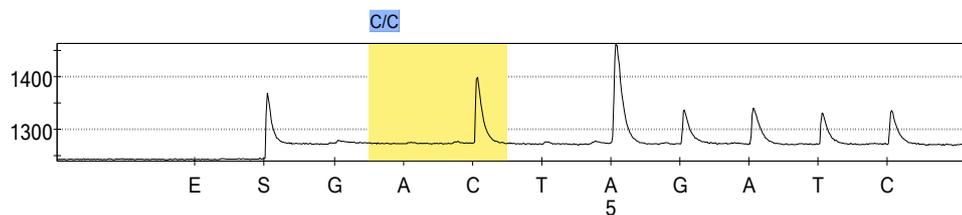
**Figure 4.12: Genome-wide graphical view of the MCPH4 families.** (a) Parametric linkage analysis (Allegra software) showing a large and wider peak (LOD = 2.6) on chromosome 15q13.1-15q22.2 for the MCPH4 locus. This locus in the MCP8 family is mapped between SNP\_A-2263685 (rs768547, position: 25,774,538bp, 15.62cM) and SNP\_A-4290104 (rs11636064 position 57,660,153bp 59.68cM). (b) The region of family MCP66 mapped between SNP\_A-1816205 (rs11856348, position 35,979,748bp, 37.42cM) and SNP\_A-2091012 (rs17663920 position 51,954,110bp, 51.87cM) on chromosome 15q14-15q21.3.

Only one substitution c.2437G>T was found in the leukocyte receptor tyrosine kinase (*LTK*) which results in a missense mutation p.E813\* creating a stop codon (Figure 4.13).

This substitution was only found in the MCP8 family and could not be observed in the MCP66 and other families from Denmark and Belgium. However, c.2437G>T was not found in 768 Pakistani healthy control chromosomes (Figure 4.14).



**Figure 4.13** Sequencing chromatograms illustrate the c.2437G>T (p.E813\*) mutation in the MCP8 family (*LTK* gene). Black arrows indicate the position of mutation where G is replaced by T which introduces a stop codon.



**Figure 4.14:** Pyrogram traces obtained by Pyrosequencing analysis of the c.2437G>T mutation in exon 20 of Pakistani control samples which were negative for the c.2437G>T mutation in *LTK*.

#### 4.4.1. MCP8 family region on chromosome 1

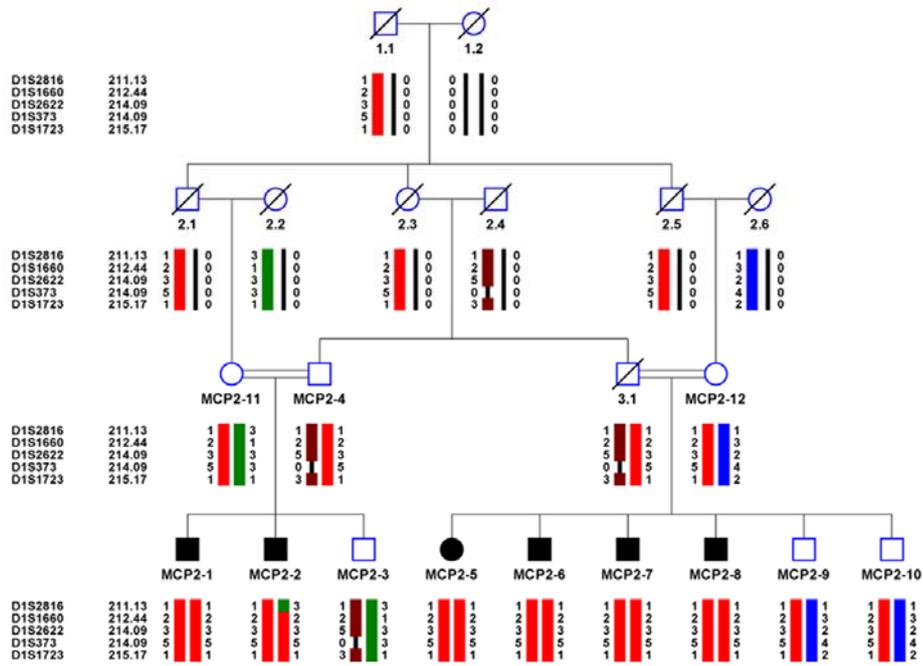
Homozygosity mapping of the MCP8 family by genome-wide linkage analysis with the Affymetrix GeneChip(R) Human Mapping 250K SNP Array (version 2.0) revealed a narrow peak on chromosome 1p31.3-1p31.1 (Figure 4.12). This region is mapped between SNP\_A-1829175 (rs855325 position 63,837,368bp) and SNP\_A-2270971 (rs6658302 position 77,681,573bp). 95 genes reside within this 13.84 Mb region (NCBI Human Genome Build 36.3).

### 4.5. MCPH5 linked families

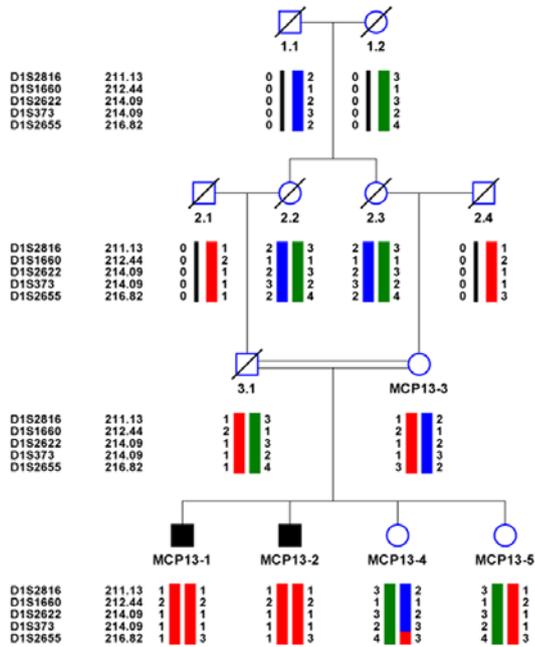
#### 4.5.1. Haplotypes of MCPH5 linked families

The pedigrees and haplotypes of 19 families linked with the MCPH5 locus (*ASPM* gene) were constructed with the help of HaploPainter v.1.043 (Thiele et al., 2004) and EasyLinkage Plus v5.02 (Lindner et al., 2005) respectively (Figure 4.15). For small pedigrees and large sets of markers, Two-/MultiPoint Para-/NonParametric Linkage Analysis was performed with the help of Allegro v1.2c and for larger pedigrees and small sets of markers, Multipoint Linkage Analysis was performed with the help of SimWalk v2.91.

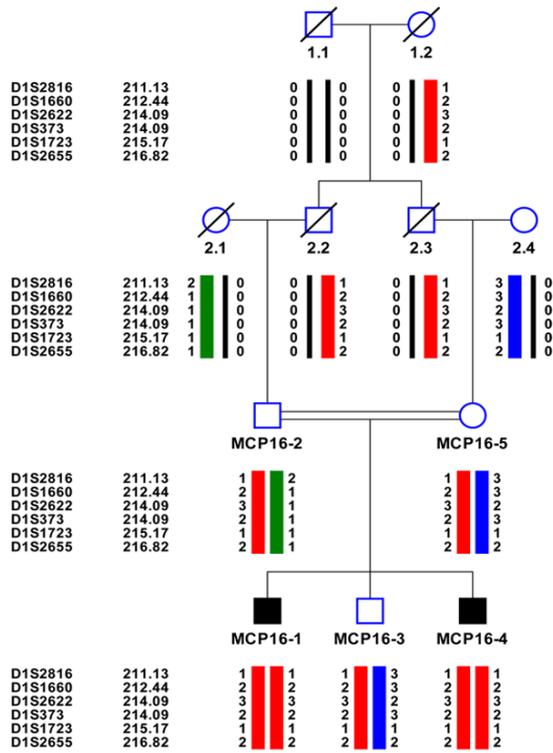
### Family MCP2



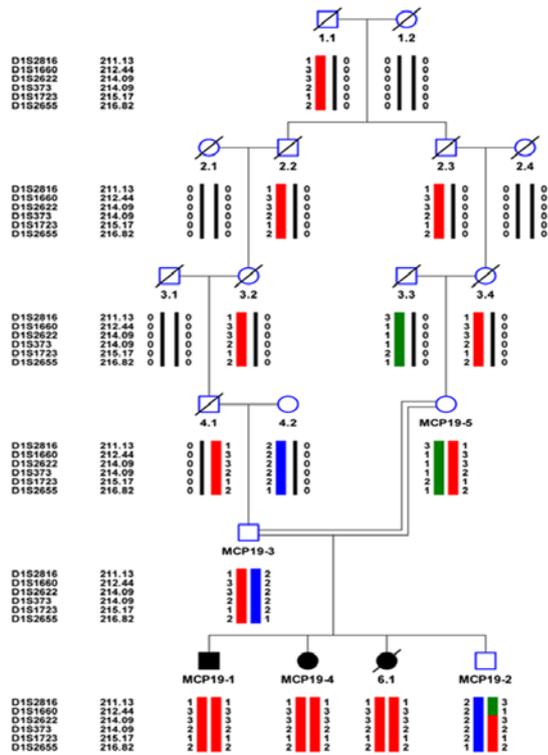
### Family MCP13



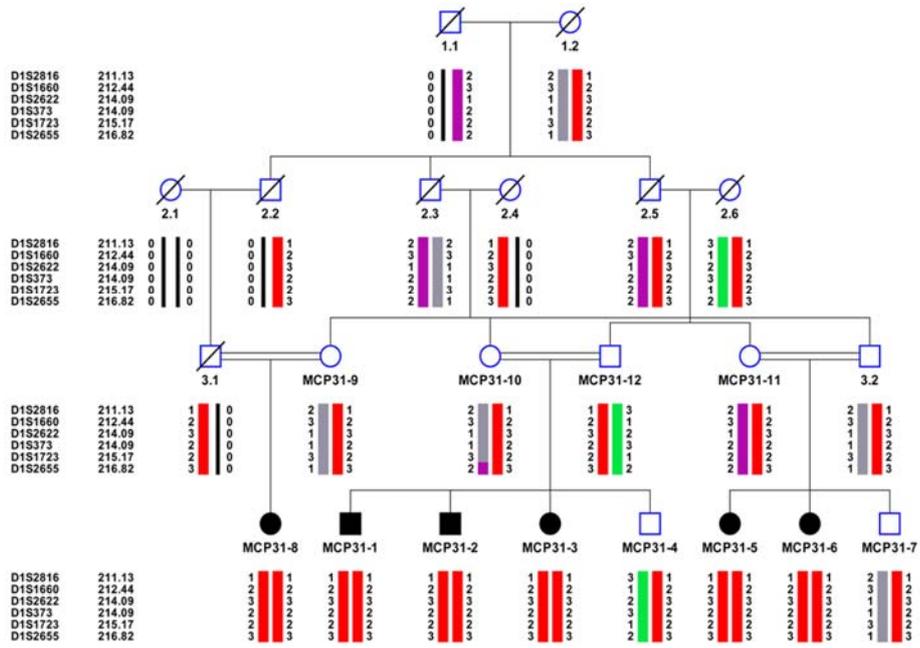
### Family MCP16



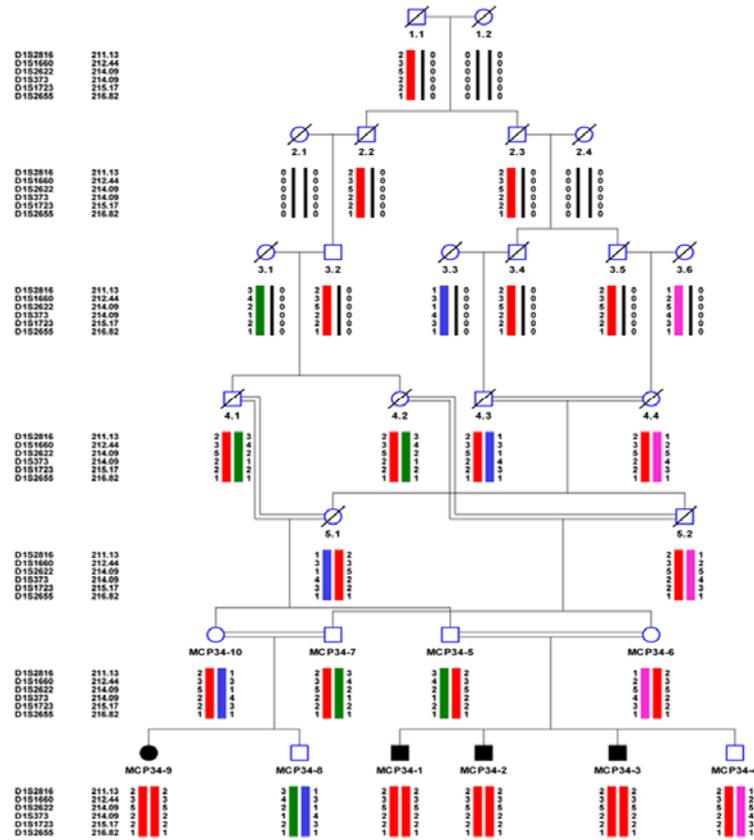
### Family MCP19



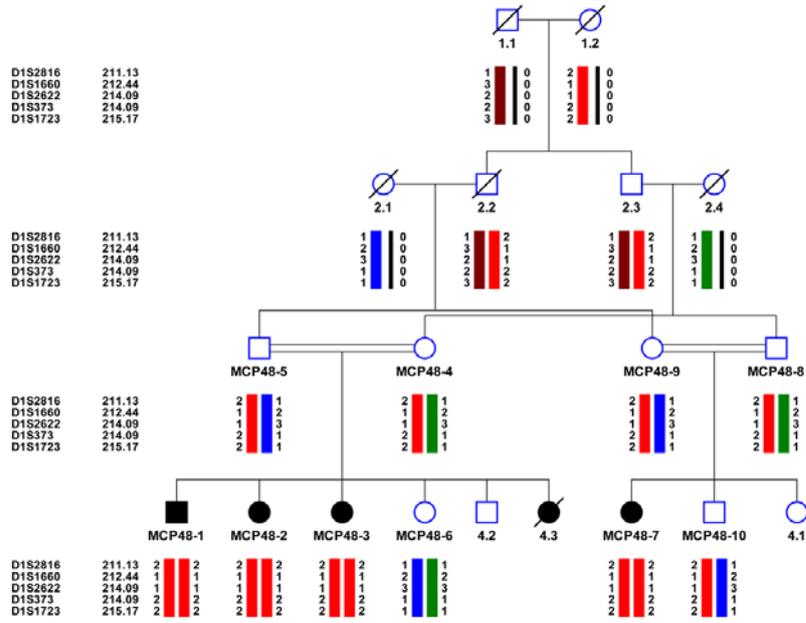
### Family MCP31



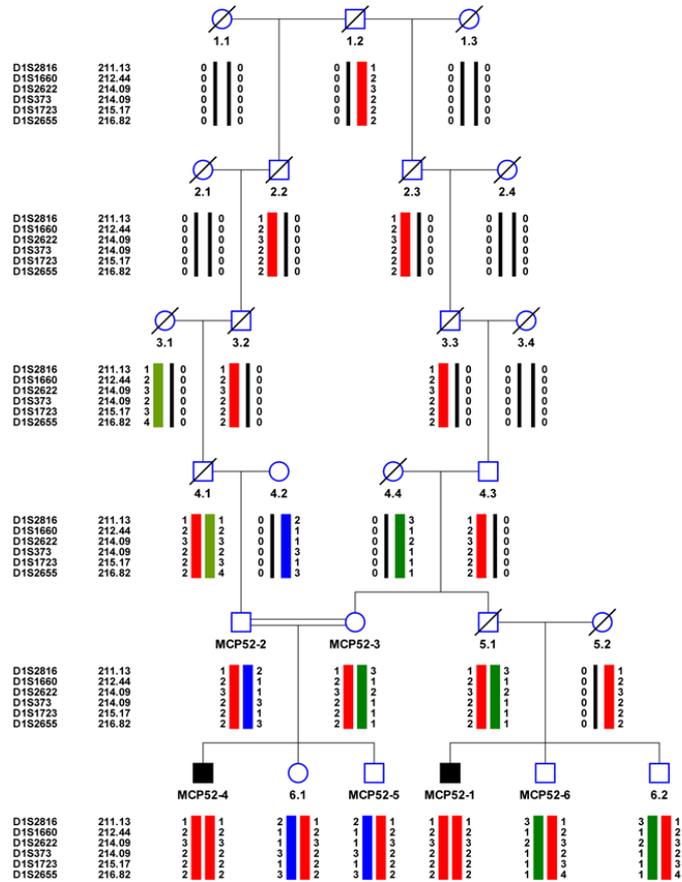
### Family MCP34



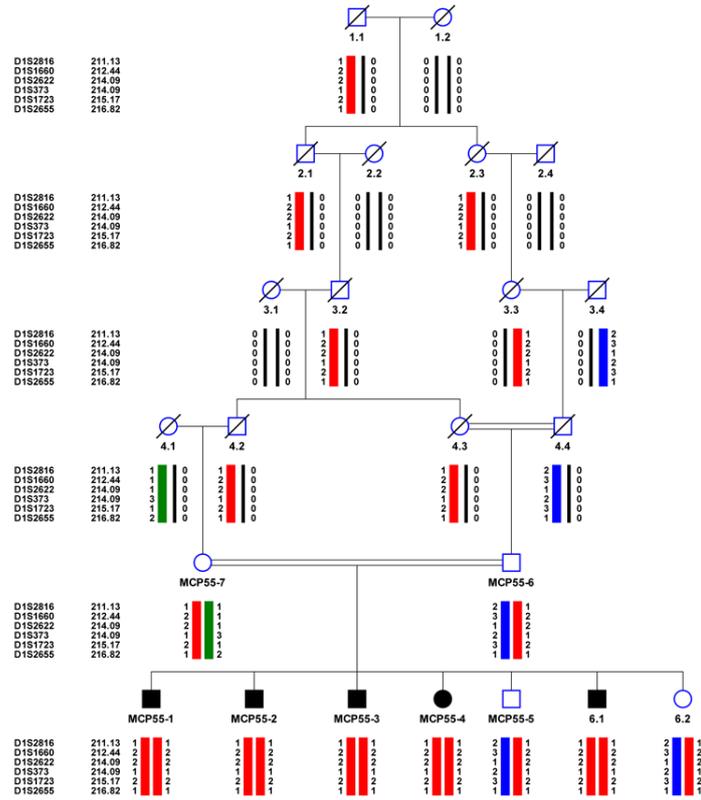
### Family MCP48



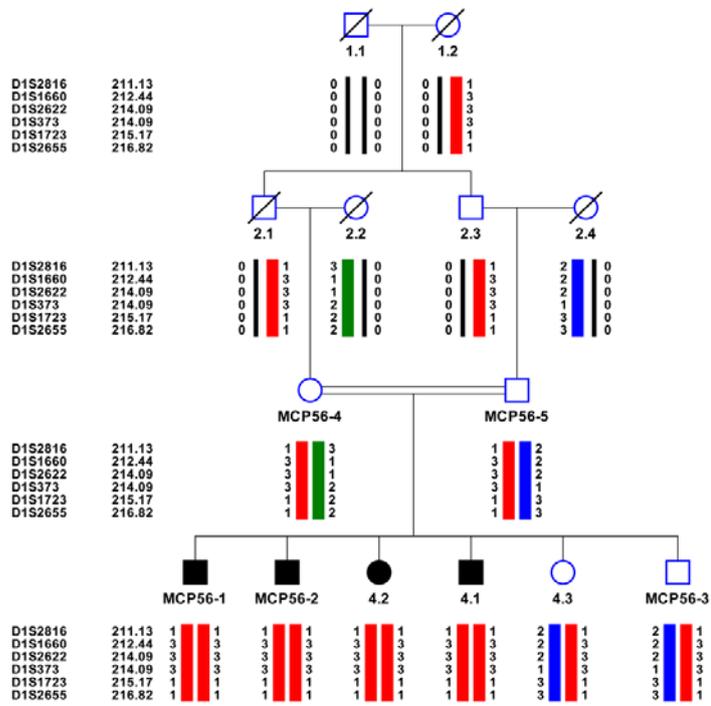
### Family MCP52



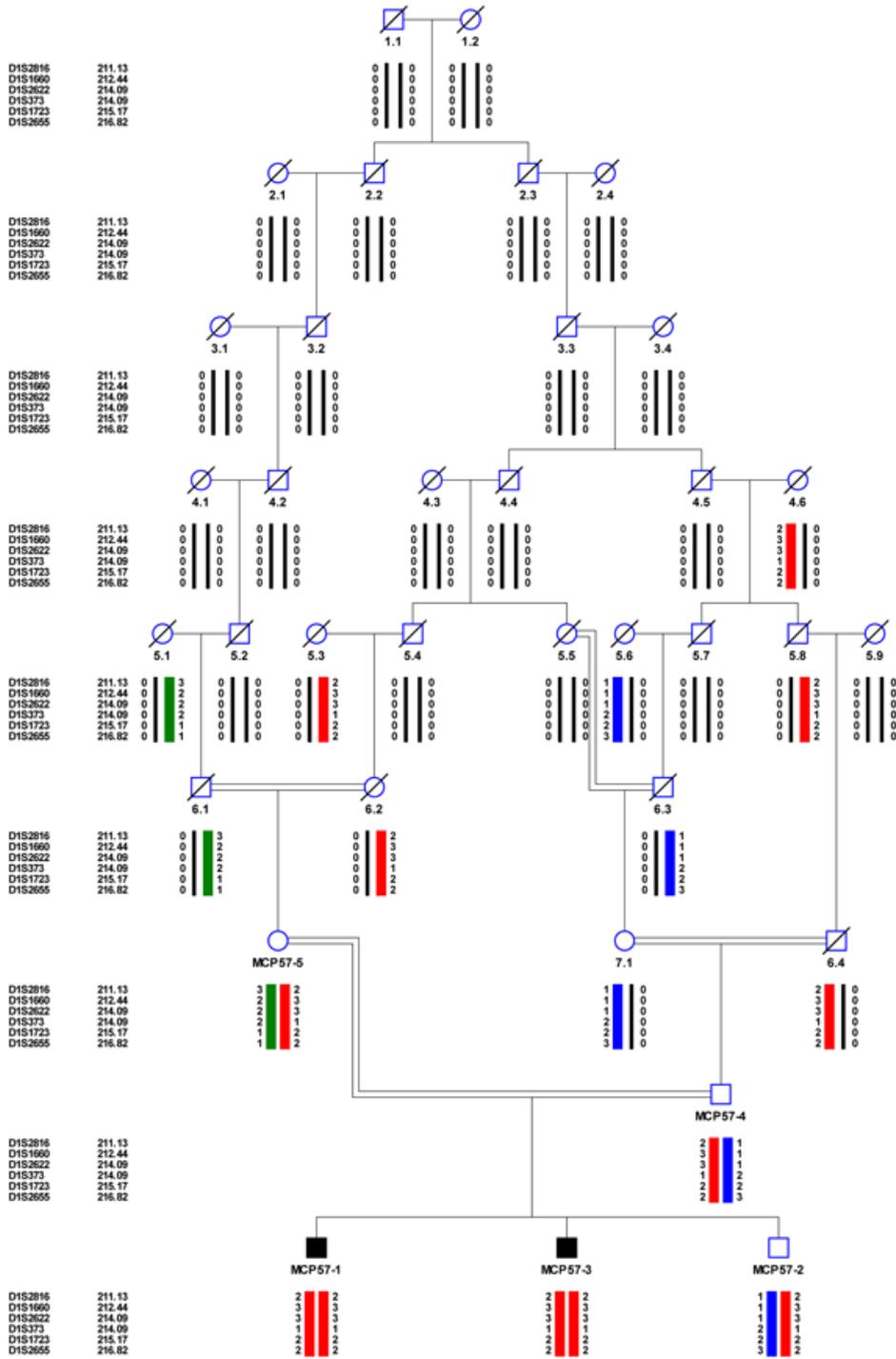
### Family MCP55



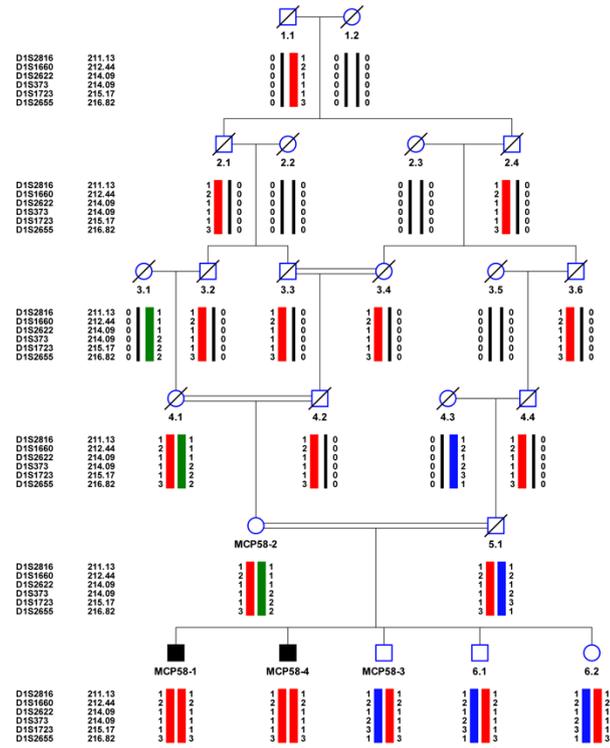
### Family MCP56



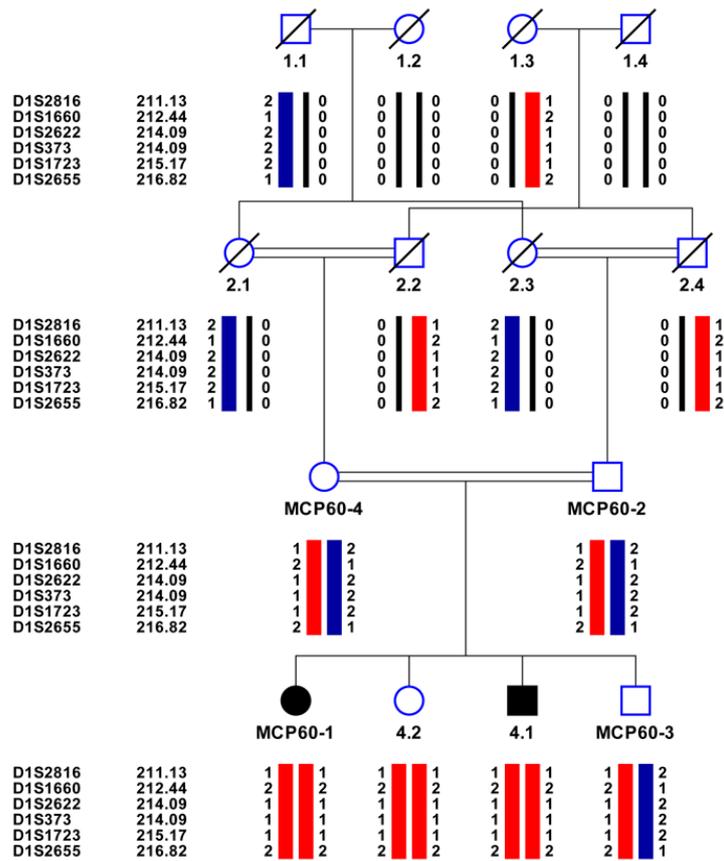
# Family MCP57



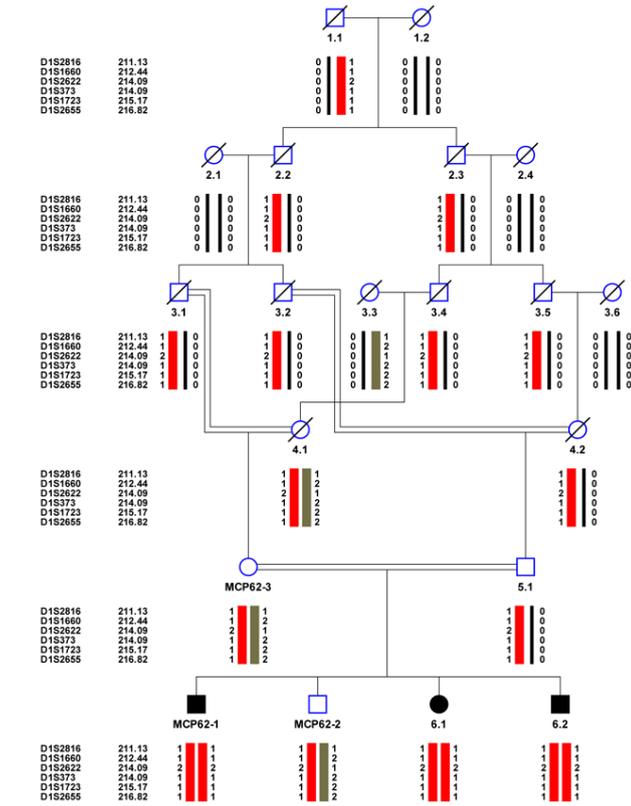
### Family MCP58



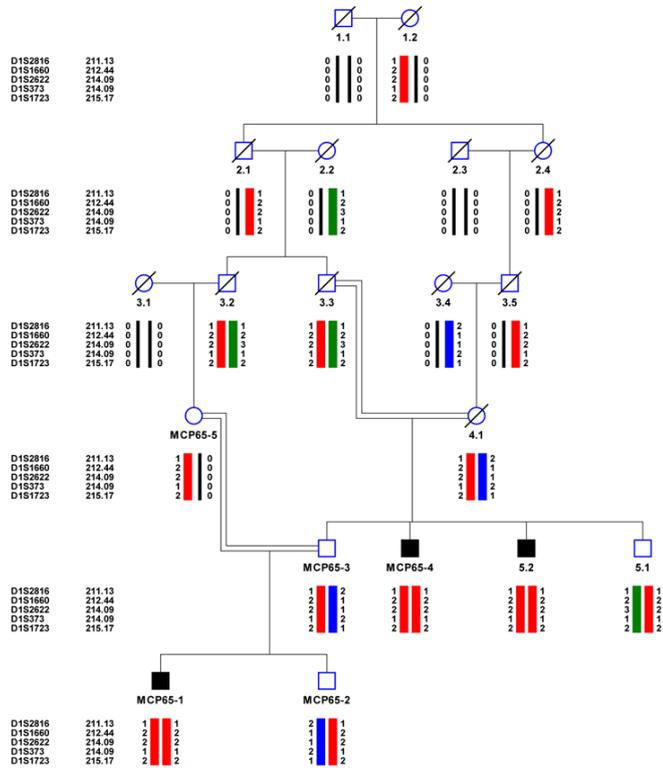
### Family MCP60



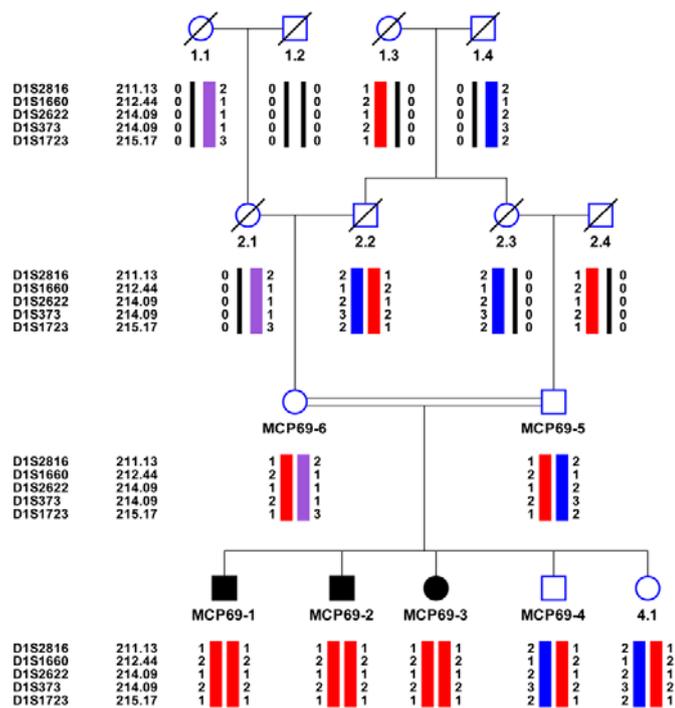
### Family MCP62



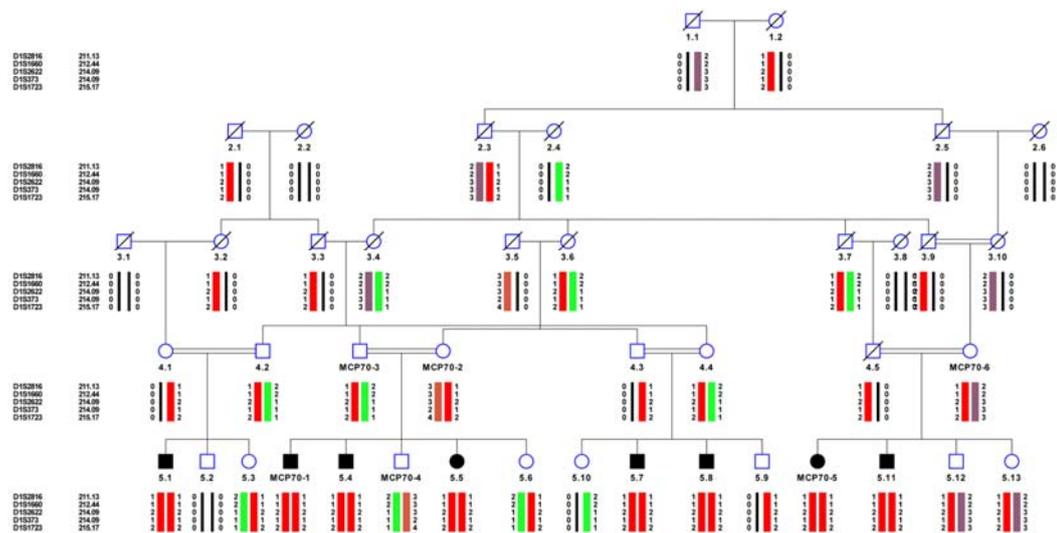
### Family MCP65



### Family MCP69



### Family MCP70



**Figure 4.15: Pedigrees and haplotype analysis of the MCPH5 families.** 19 families were linked with the MCPH5 locus. Locus specific microsatellite markers were selected and genotyped within all microcephaly families. Marker names and positions are displayed on the left side of each generation. Their disease haplotype fragments are indicated with red boxes.

#### 4.5.2. Abnormal Spindle-like Microcephaly associated (*ASPM*) Sequencing

The Abnormal Spindle-like Microcephaly associated (*ASPM*) gene was identified as the causative gene for the MCPH5 locus (Bond J et al., 2002). The *ASPM* gene is the human

ortholog of the *Drosophila melanogaster* 'abnormal spindle' gene (*asp*), which is essential for normal mitotic spindle function in embryonic neuroblasts. *ASPM* is a large gene with 28 exons having an ORF (Open Reading Frame) of 10.4 kb. It spans 62 kb of genomic sequence and encodes a protein with 3,477 amino acids. The mutation in the *ASPM* gene is the major cause of primary microcephaly in the Pakistani population (Bond et al., 2003).

MCPH5 linked families were subjected to a screen of the *ASPM* gene via direct DNA sequencing. To sequence the *ASPM* gene, a set of 43 primer pairs was designed using Primer3 v. 0.4.0 (Rozen and Skaletsky, 2000), which covers all 28 exons and exon-intron boundaries of this gene. Two afflicted and one normal individual of each MCPH5 linked family with one normal control were sequenced using Sanger sequencing.

Five previously reported (p.W1326\*, p.Y3164\*, p.R3244\*, p.K1862E, and p.Y3263\*) and eight novel mutations (p.L333fs, p.Q2890\*, p.Y3163\*, p.R1019\*, c.3977G>A (p.W1326\*), p.R1327\*, p.N2734Lfs\*16, and p.L2837Mfs\*35) were identified in the *ASPM* gene. The results are summarized in table 4.2.

**Table 4.2: Novel and known mutations in the *ASPM* gene.** 19 MCPH5 linked families demonstrated eight novel and five known mutations. Novel mutations in families MCP2, 13, 16 and 19 were published with (Muhammad et al., 2009). Novel mutations are highlighted red.

Family ID	cDNA mutation	Protein Mutation	Exon	Status of Mutation
MCP2	c.1002delA	p.L333fs	3	Novel
MCP13	c.9492T>G	p.Y3163*	23	Novel
MCP16	c.8668C>T	p.Q2890*	18	Novel
MCP19	c.3055C>T	p.R1019*	11	Novel
MCP31	c.3978G>A	p.W1326*	17	Known
MCP34	c.3978G>A	p.W1326*	17	Known
MCP48	c.3977G>A	p.W1326*	17	Novel
MCP52	c.3978 G>A	p.W1326*	17	Known
MCP54	c.9730C>T	p.R3244*	24	Known
MCP55	c.5584A>G	p.K1862E	18	Known+
MCP55	c.9492T>G	p.Y3164*	23	Known++
MCP56	c.3978 G>A	p.W1326*	17	Known
MCP57	c.3978G>A	p.W1326*	17	Known
MCP58	c.3979C>T	p.R1327*	17	Novel
MCP60	c.9789T>A	p.Y3263*	24	Known
MCP62	c.8668C>T	p.Q2890*	18	Novel+++

MCP64	c.3978 G>A	p.W1326*	17	Known
MCP65	c.3978G>A	p. W1326*	17	Known
MCP69	c.8200-8201delAA	p. N2734Lfs*16	18	Novel
MCP70	c.8507-8508delAG	p. L2837Mfs*35	18	Novel

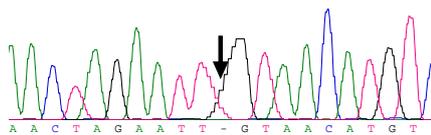
<sup>+</sup>This could be the rare polymorphism or might have association with the diseased variant (p.Y3164\*).

<sup>++</sup>The mutation c.9492T>G (p.Tyr3164\*) is in cis to the missense mutation c.5584A>G (p.Lys1862Glu). The latter had previously been found in one MCPH family in homozygous state and deemed to be pathogenic (Darvish et al., 2010).

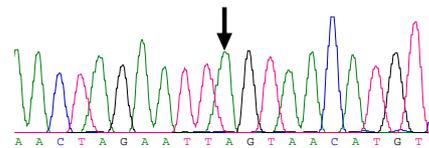
<sup>+++</sup> p.Q2890\* mutation was identified in two different families, MCP16 and MCP62.

A nonsense mutation (c.3977G>A, p.W1326\*) was observed in family MCP48 which changes nucleotide G at position 3977 of the cDNA instead of position 3978 which has already been reported (Figure 4.16). All of these mutations (except c.5584A>G, p.K1862E) are truncating the protein so it is likely that they all result in nonsense mediated mRNA decay.

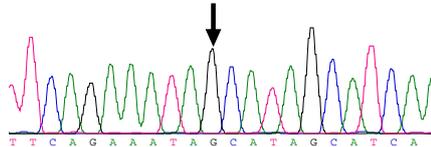
**Homozygous mutant c.1002delA**



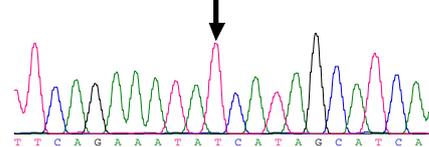
**wild type**



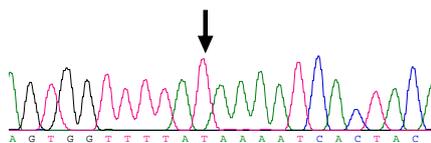
**Homozygous mutant c.9492T>G**



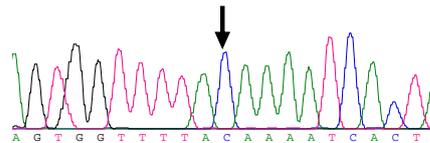
**wild type**



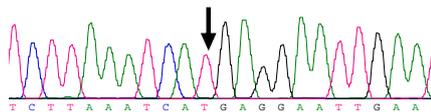
**Homozygous mutant c.8668C>T**



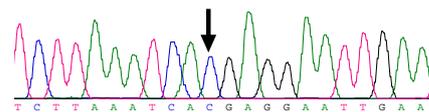
**wild type**



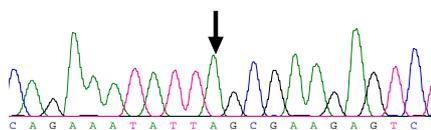
**Homozygous mutant c.3055C>T**



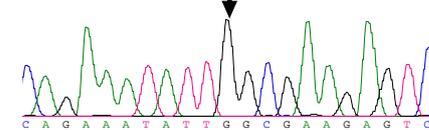
**wild type**

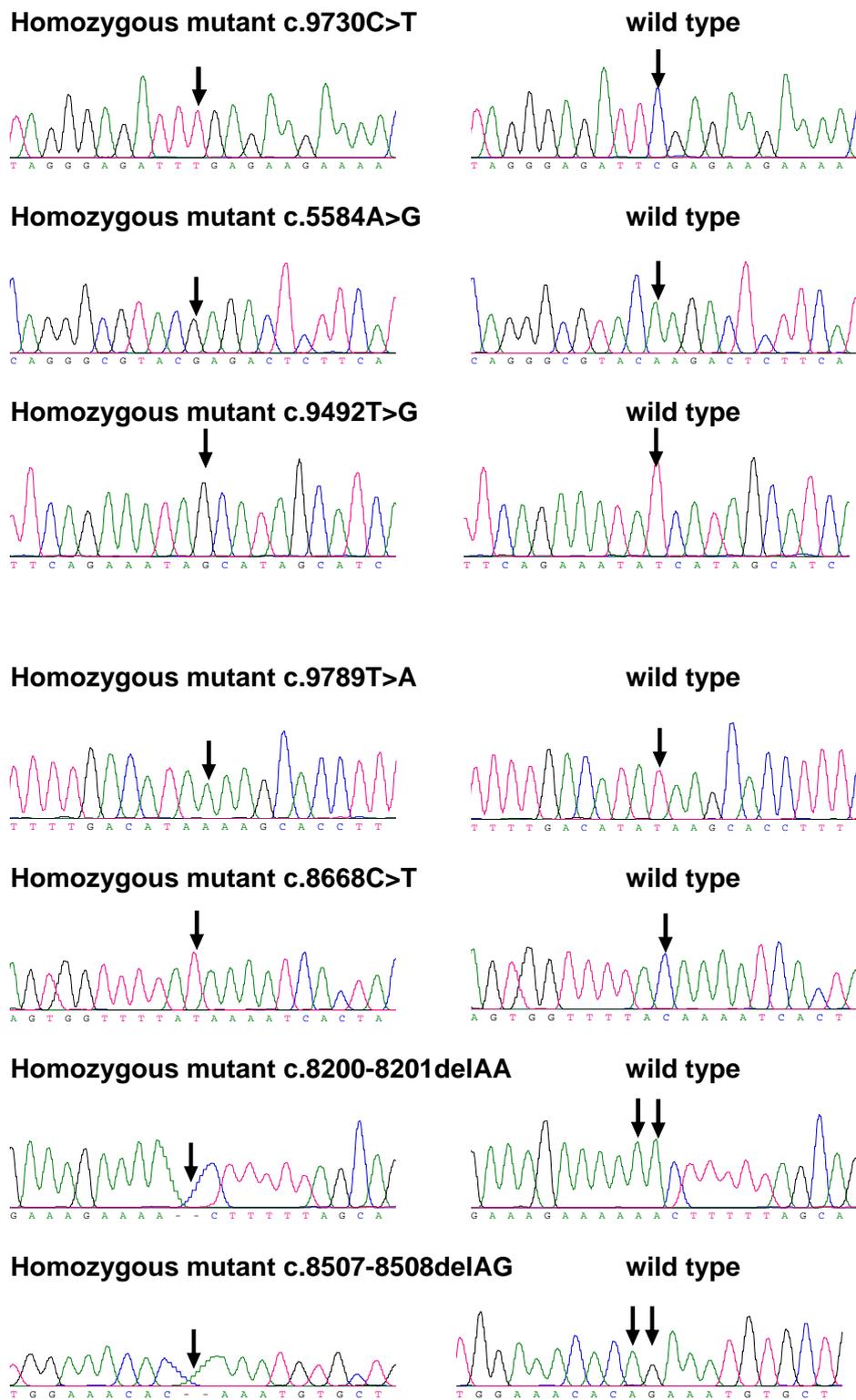


**Homozygous mutant c.3977G>A**



**wild type**

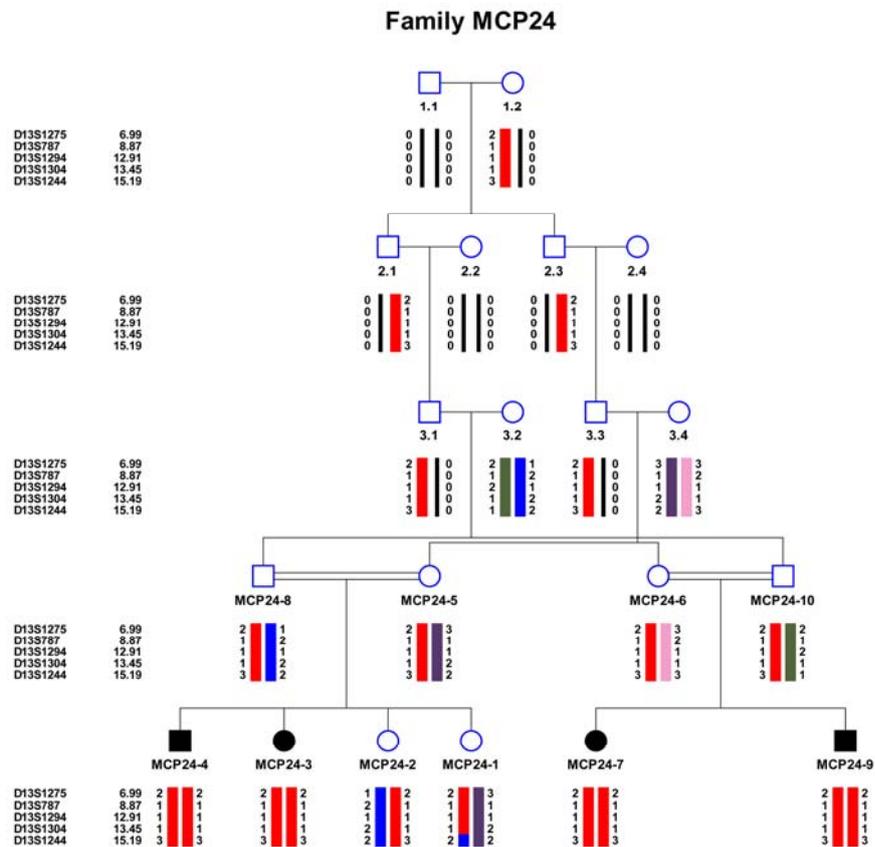




**Figure 4.16: Chromatograms of *ASPM* mutations in MCPH5 families.** Representative Sanger traces illustrate different novel mutations identified in *ASPM*. Mutant chromatograms are depicted on the left while wild-type chromatograms are found on the right. Black arrows indicate the position of the mutation.

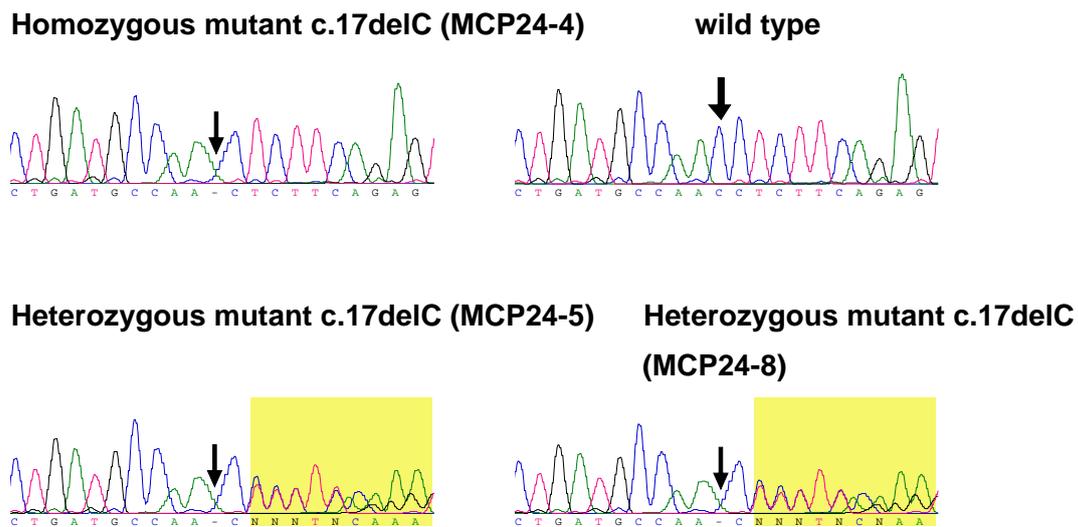
#### 4.6. MCPH6 linked family

One family (MCP24) was linked to the MCPH6 locus (*CENPJ* gene) and was excluded to be linked to all other known MCPH loci. The family consists of four afflicted individuals from consanguineous parents (Figure 4.17).



**Figure 4.17: Haplotype analysis of the MCP24 family linked with the MCPH6 locus.** Marker names and positions (cM) are displayed on the left side of each generation. The disease haplotype fragments are indicated with red boxes.

The causative gene for the MCPH6 locus is the centromere associated Protein J (*CENPJ*; MIM609279), also known as centrosomal protein 4.1- associated protein (*CPAP*). This gene contains 17 exons and spans 39.8 kb on human chromosome 13q12.2 (Bond et al., 2005). To sequence the 17 exons and exon-intron boundaries of the *CENPJ* gene, primers were designed using Primer3 v. 0.4.0 (Rozen and Skaletsky, 2000). The *CENPJ* gene was sequenced in family MCP24 (linked to the MCPH6 locus). Sequencing analysis identified one known homozygous deletion c.17delC (Figure 4.18). The single base deletion causes a frameshift that incorporates two amino acids and then introduced stops codon (p.T6fs\*3).



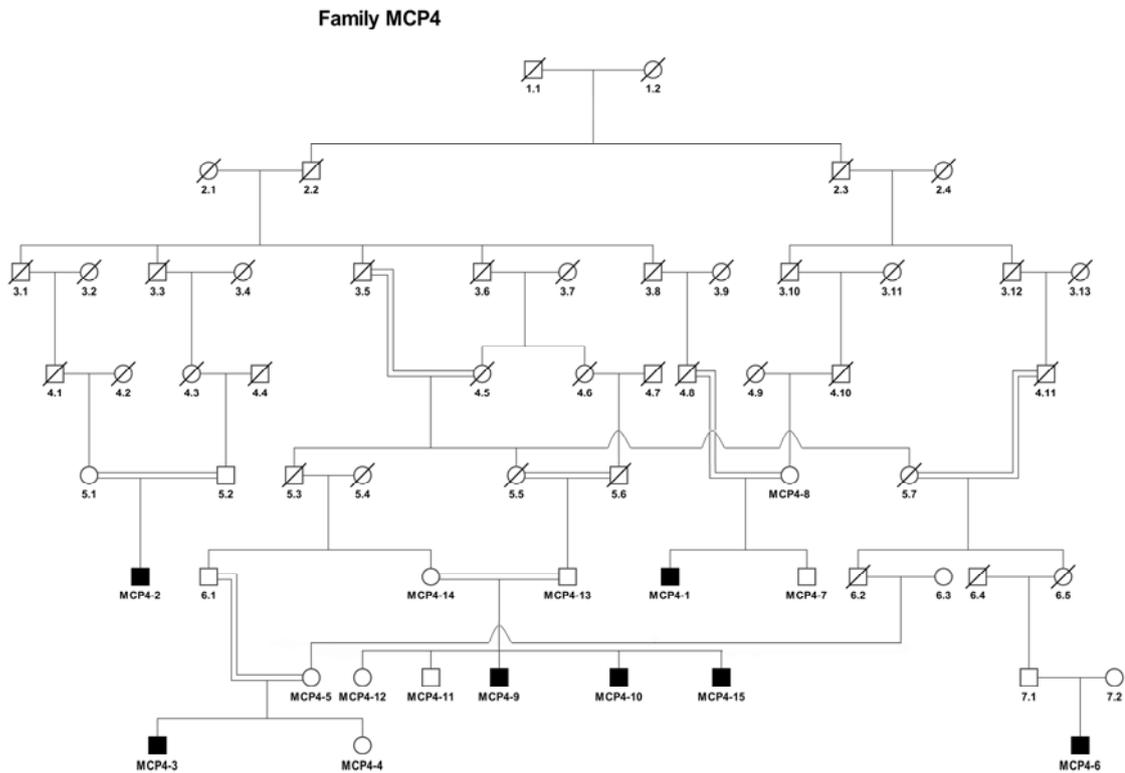
**Figure 4.18: Sequencing chromatograms of the c.17delC mutation identified in exon 2 of *CENPJ*.** Black arrows indicate the position of the mutation where C is deleted. Lower chromatograms represent the heterozygous parents (MCP24-5 and MCP24-8).

#### 4.7. Excluded families

PCR based Genotyping of 30 autosomal recessive primary microcephaly families revealed exclusion of five families. These families were not linked to any of the known MCPH loci and were a unique source for the identification of further genes underlying this disorder. These families were MCP4, MCP50, MCP53, MCP63 and MCP67.

#### 4.8. Excluded family MCP4

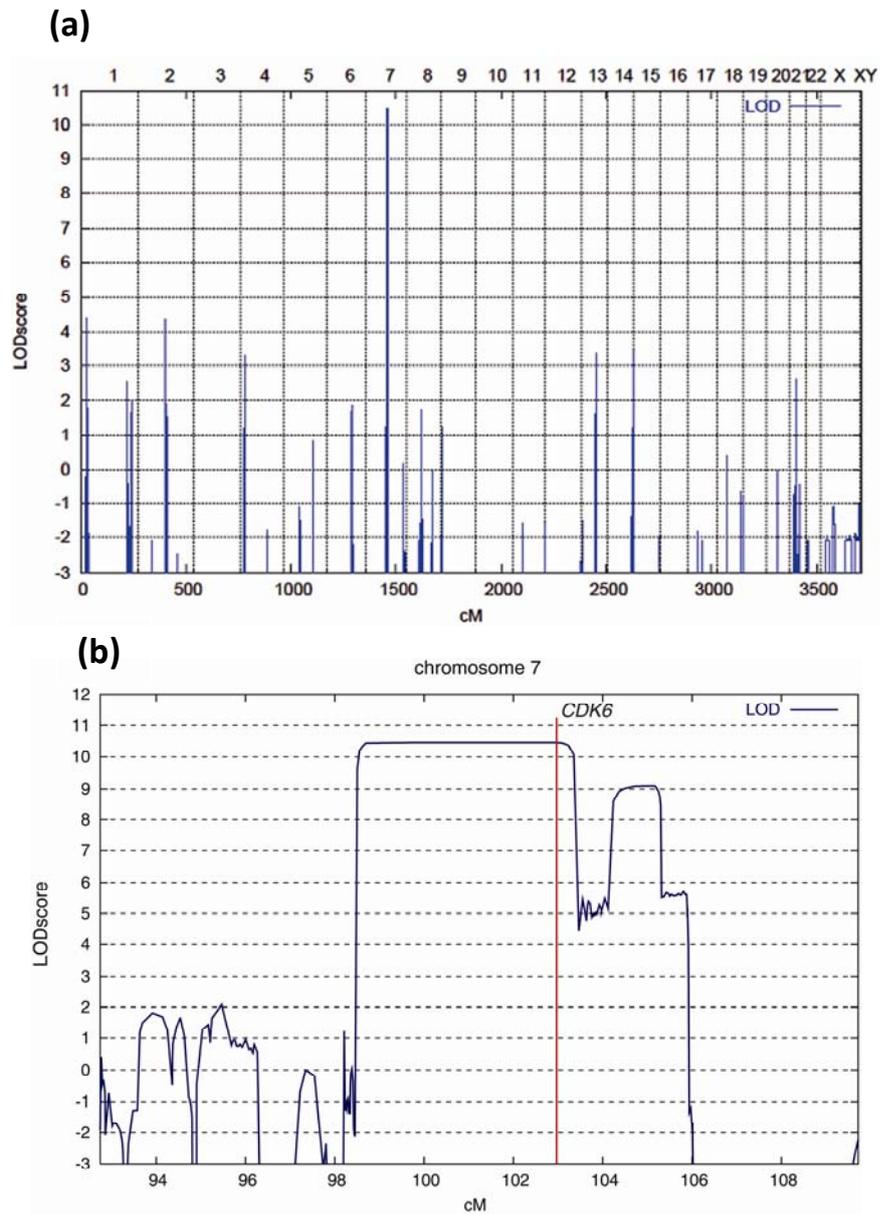
MCP4 is a large consanguineous family with seven afflicted individuals (Figure 4.19) and was ascertained from remote regions of Pakistan. The patients have the typically sloping forehead and are mentally retarded without any other abnormalities. After exclusion of linkage with all known seven MCPH loci, the family was subjected to a genome-wide screen for regions of linkage with the Affymetrix GeneChip(R) Human Mapping 250K SNP Array (version 2.0) on all seven afflicted children (MCP4-1, MCP4-2, MCP4-3, MCP4-6, MCP4-9, MCP4-10 and MCP4-15), two unaffected siblings (MCP4-11 and MCP4-12), and the parents (MCP4-13 and MCP4-14)



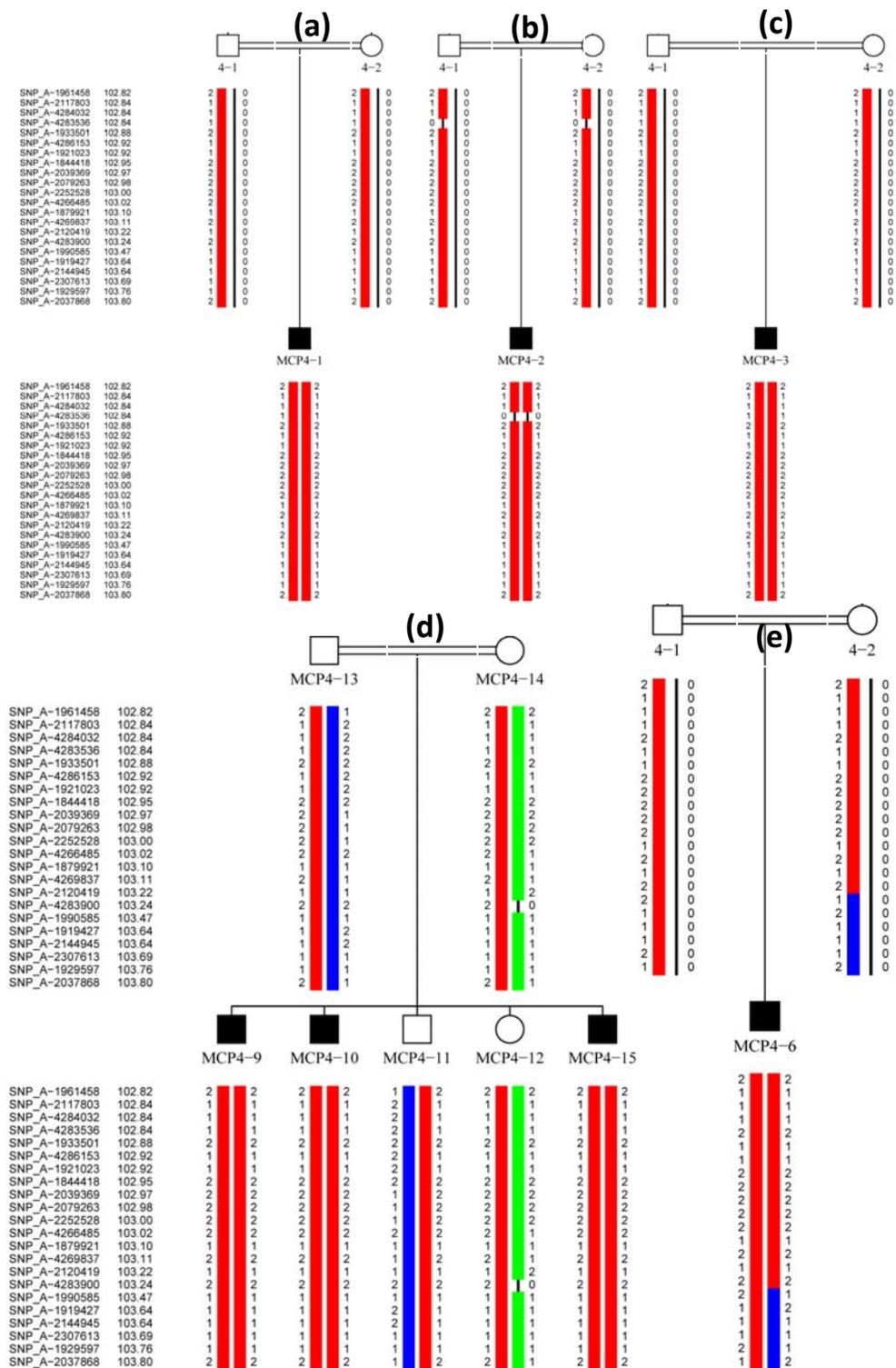
**Figure 4.19: Pedigree of the MCP4 family with the MCPH8 locus.** The members denoted with MCP were subjected to the genome screen.

#### 4.8.1. MCPH8 locus identification

Genome-wide linkage analysis revealed the identification of one novel gene locus MCPH8 on chromosome 7q21.11-q21.3. The genome-wide graph shows a single clear peak on chromosome 7q21.11-q21.3 with a LOD score of 10.47 which reaches the theoretical maximum of primary microcephaly families as shown in figure 4.20. This locus maps between SNP\_A-1844880 (rs17159934, position 84,908,672bp) and SNP\_A-1990585 (rs42511 position 92,662,815bp) and comprises a region of 7.75 Mb (Figure 4.21).



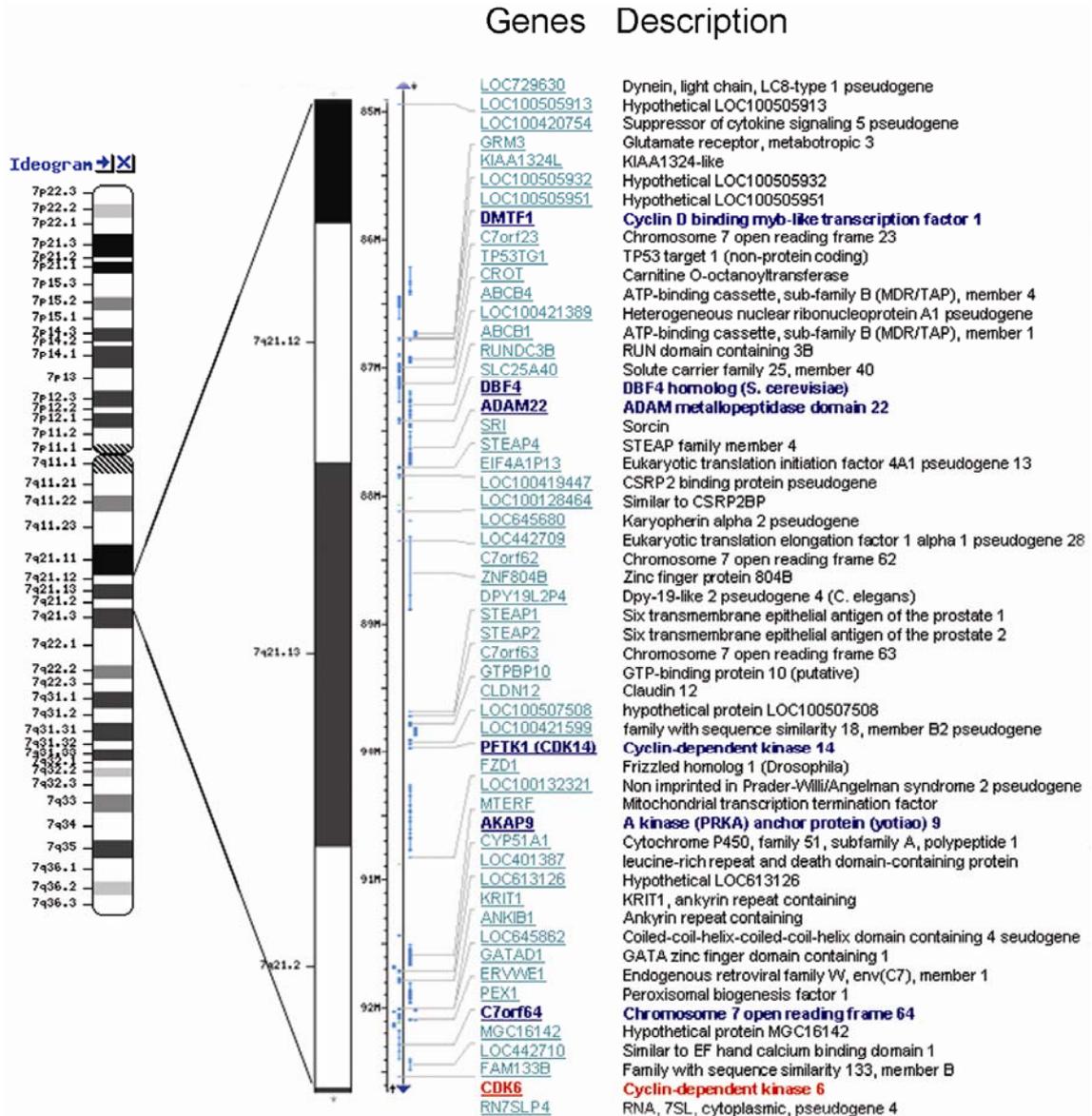
**Figure 4.20. Genome-wide graphical view of the MCPH8 locus. (a)** Parametric linkage analysis (Allegro software) showing a single peak (LOD = 10.47) on chromosome 7q21.11-q21.3 for the MCPH8 locus. The calculated LOD score (10.47) reaches the theoretical maximum LOD score of primary microcephaly families. **(b)** Results of the parametric linkage analysis showing LOD scores on chromosome 7. The red line shows the position of *CDK6* (causative gene of MCPH8 locus).



**Figure 4.21: Haplotypes of the MCP4 family.** Haplotype analysis of the MCP4 family linked to chromosome 7q21.11-q21.3. Due to large family size, the haplotype is divided into five parts (a to e), each part represent one loop of the family. Black symbols represent afflicted individuals. The critical interval was defined by the recombinations observed for markers from SNP\_A-1844880 (rs17159934, physical position 84,908,672bp) to SNP\_A-1990585 (rs42511 physical position 92,662,815bp) comprising 7.75 Mb region. Here only the region near *CDK6* is shown.

#### 4.8.2. Candidate genes in the MCPH8 locus

A bioinformatical analysis of the MCPH8 (7.75 Mb) region was performed to seek the candidate genes. 49 known and predicted genes (NCBI Human Genome Build 36.3) were found in this region (Figure 4.22).



**Figure 4.22: G-banded chromosome 7 showing the position of the MCPH8 locus along with the position of the genes on the long arm 7q21.11-q21.3.** The locus was mapped between SNP\_A-1844880 (rs17159934, physical position 84908672bp) and SNP\_A-1990585 (rs42511 physical position 92662815bp) (NCBI Human Genome Build 36.3). This 7.75 Mb region harbours 49 genes. Blue and red coloured genes were sequenced because they have the highest expression in human brain and have a role in the cell cycle.

#### 4.8.3. Candidate genes sequencing in the MCPH8 locus

As autosomal recessive primary microcephaly is a disorder of neurogenic mitosis I selected those genes for sequencing which have mitotic functions, centrosomal or spindle

pole localization, embryonic brain expression or a role in neurogenesis. Among these 49 genes, Cyclin D binding myb-like transcription factor 1 (*DMTF1*), DBF4 homolog (*DBF4*), ADAM metallopeptidase domain 22 (*ADAM22*), PFTAIRE protein kinase 1 (*PFTK*) also known as cyclin-dependent kinase 14 (*CDK14*), A kinase (PRKA) anchor protein (yotiao) 9 (*AKAP9*), Hypothetical protein DKFZp564O0523 (*DKFZP564O0523*) also known as chromosome 7 open reading frame 64 (*C7orf64*) and cyclin-dependent kinase 6 (*CDK6*) genes were considered the strongest candidate genes. *AKAP9* and *Pericentrin* (*PCNT*) are orthologs of the yeast Spc110 protein and a mutation in *Pericentrin* is the cause of Microcephalic osteodysplastic primordial dwarfism-type Majewski II (MOPD-II) (Rauch et al., 2008) which has an overlapping phenotype with MCPH. So *AKAP9* was also considered the strongest candidate gene. Moreover, all 49 genes residing in the MCPH8 region were prioritized by the disease gene prioritizing tools Endeavour (Aerts et al., 2006) and GeneWanderer (Kohler et al., 2008) (Figure 4.23, Figure 4.24).

Global prioritization	Annotation EnsemblEst	Annotation GeneOntology	Annotation Interpro	Annotation Kegg	Annotation Swissprot	Blast	CisRegModule	Expression SonEtAl	Motif	Precalculated Ouzounis	Precalculated Prospectr	Test
CDK6	SRI	CDK6	CDK6	CDK6	PFTK1	AKAP9	CDK6	DMTF1	FZD1	CLDN12	CDK6	ADAM22
AKAP9	PEX1	AKAP9	AKAP9	FZD1	CDK6	CROT	STEAP2	MTERF	DMTF1	ABCB1	ADAM22	AKAP9
ADAM22	FZD1	SRI	SRI	ABCB4	SRI	PEX1	AKAP9	SRI	SRI	ABCB4	GRM3	PFTK1
SRI	AKAP9	PEX1	PEX1	GRM3	ADAM22	PFTK1	PFTK1	PFTK1	ABCB4	PEX1	ABCB1	DMTF1
DMTF1	DMTF1	FZD1	FZD1	ABCB1	CLDN12	CDK6	ADAM22	PEX1	CDK6	ADAM22	ABCB4	FZD1
PFTK1	ADAM22	DMTF1	DMTF1	AKAP9	ABCB1	SRI	CLDN12	AKAP9	ADAM22	FZD1	CROT	GRM3
CLDN12	STEAP2	ADAM22	ADAM22	SRI	AKAP9	ADAM22	MTERF	CLDN12	ABCB1	DMTF1	STEAP2	CDK6
FZD1	CLDN12	STEAP2	STEAP2	PEX1	ABCB4	CLDN12	FZD1	ADAM22	CROT	MTERF	SRI	SRI
PEX1	MTERF	CLDN12	CLDN12	DMTF1	MTERF	ABCB1	GRM3	GRM3	GRM3	PFTK1	CLDN12	MTERF
ABCB1	CROT	MTERF	MTERF	ADAM22	PEX1	ABCB4	DMTF1	STEAP2	PFTK1	STEAP2	FZD1	STEAP2
ABCB4	CDK6	CROT	CROT	STEAP2	FZD1	MTERF	CROT	CROT	PEX1	CDK6	PFTK1	CLDN12
CROT	ABCB4	ABCB4	ABCB4	CLDN12	GRM3	FZD1	PEX1	ABCB4	MTERF	SRI	AKAP9	PEX1
MTERF	PFTK1	PFTK1	PFTK1	MTERF	DMTF1	GRM3	SRI	CDK6	AKAP9	AKAP9	DMTF1	CROT
GRM3	GRM3	GRM3	GRM3	CROT	STEAP2	DMTF1	ABCB1	ABCB1	CLDN12	CROT	MTERF	ABCB4
STEAP2	ABCB1	ABCB1	ABCB1	PFTK1	CROT	STEAP2	ABCB4	FZD1	STEAP2			

**Figure 4.23:** Prioritized candidate genes of the MCPH8 locus on chromosomes 7 by Endeavour. *CDK6* is at the highest rank. Green colour represents the *CDK6* which has got the highest rank.

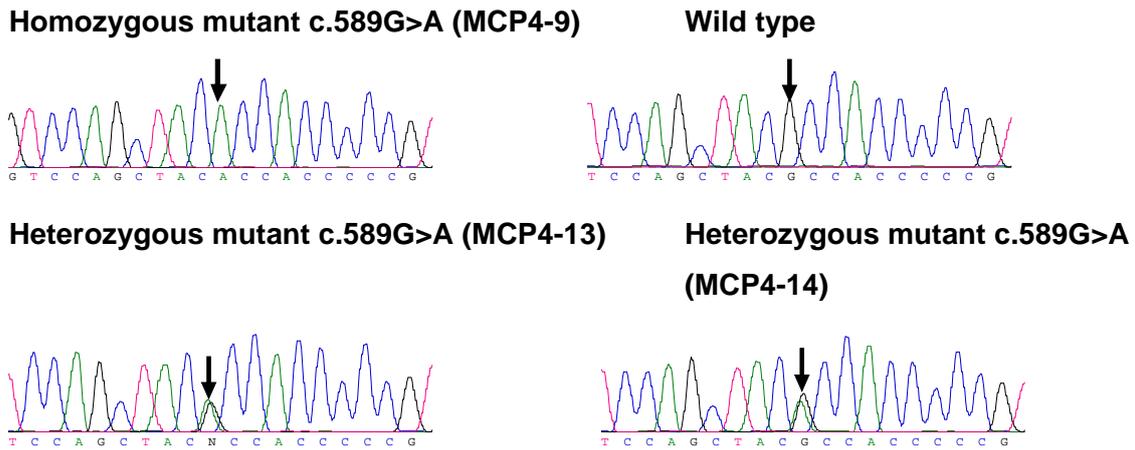
Rank	Gene Symbol	Score	Start	End
1	<u>DKEZP5640.0523</u>	1.197	91996042	92004778
2	<u>AKAP9</u>	0.05532	91408127	91577924
3	<u>CDK6</u>	0.02480	92072170	92301147
4	<u>ABCB1</u>	0.01983	86970883	87180499
5	<u>CYP51A1</u>	0.01910	91579401	91601945
6	<u>DMTF1</u>	0.008254	86619859	86663583
7	<u>ADAM22</u>	0.005694	87401637	87664389
8	<u>FZD1</u>	0.002848	90731718	90736067
9	<u>GRM3</u>	0.002269	86111165	86332127
10	<u>SRI</u>	0.002241	87672367	87694243
11	<u>KRIT1</u>	0.001227	91666218	91713349
12	<u>PEX1</u>	0.0009363	91954272	91995780
13	<u>ABCB4</u>	0.0008893	86869296	86947683
14	<u>PFTK1</u>	0.0007322	90176647	90677839
15	<u>STEAP1</u>	0.0007275	89621624	89632076
16	<u>GTPBP10</u>	0.0006239	89813925	89853309
17	<u>ERWE1</u>	0.0006175	91935630	91938673
18	<u>TP53AP1</u>	0.0006004	86792476	86812766
19	<u>STEAP2</u>	0.0005945	89678935	89704864
20	<u>DBF4</u>	0.0005793	87343479	87376791
21	<u>CROT</u>	0.0005676	86812946	86867044
22	<u>MTERF</u>	0.0002409	91339956	91347951
23	<u>CLDN12</u>	0.0001741	89870731	89883203
24	<u>GATAD1</u>	6.303e-05	91914700	91926677

**Figure 4.24:** Prioritized genes of the MCPH8 locus on chromosome 7 by GeneWanderer. *CDK6* is at position three.

All selected genes were sequenced in MCP4 family members. No mutation was found in all of these genes except *CDK6*.

#### 4.8.4. *CDK6* as the causative gene of the MCPH8 locus

As Cyclin-dependent kinase 6 (*CDK6*) has got the highest rank among 49 genes, prioritized by Endeavour and GeneWanderer (Figure 4.23 and 4.24), this Cyclin-dependent kinase 6 (*CDK6*) was sequenced within three afflicted individuals of the MCP4 family with one control. A homozygous single base substitution c.589G>A in exon 5 was found. Later on the complete family was sequenced bi-directionally to confirm this change. This mutation was found in all afflicted individuals and segregated in a recessive pathogenic way. The parents were heterozygous carriers (Figure 4.25) and the mutation in *CDK6* c.589G>A was not found in 768 Pakistani healthy control samples. The substitution in *CDK6* (c.589G>A) was also not found in 760 German healthy control samples.



**Figure 4.25: Sequencing chromatograms illustrate the c.589G>A (p.Ala197Thr) mutation in the MCPH8 family (*CDK6* gene).** Black arrows indicate the position of the mutation where G is replaced by A which causes a change from alanine to threonine. Carrier status of both parents (MCP4-13 and MCP4-14) is also shown.

#### 4.8.5. Structure of *CDK6*

There are five transcripts of *CDK6* but only two are protein coding. The longest transcript (ENST00000265734) consists of 8 exons that spread over 11,612 bps length and code a protein of 326 residues. Exon 8 is the largest one and contains the C-terminus of the protein and a long non-coding region as shown in figure 4.26.



**Figure 4.26: Graphical representation of the *CDK6* gene.** Exons are drawn according to scale (1Kb = 1cm), while introns are just shown as artificial lines. Black boxes represent the non coding exons while white ones depict coding regions. The position of the mutation (c.589G>A) in exon five is shown.

#### 4.8.6. Evolutionary conservation of Alanine

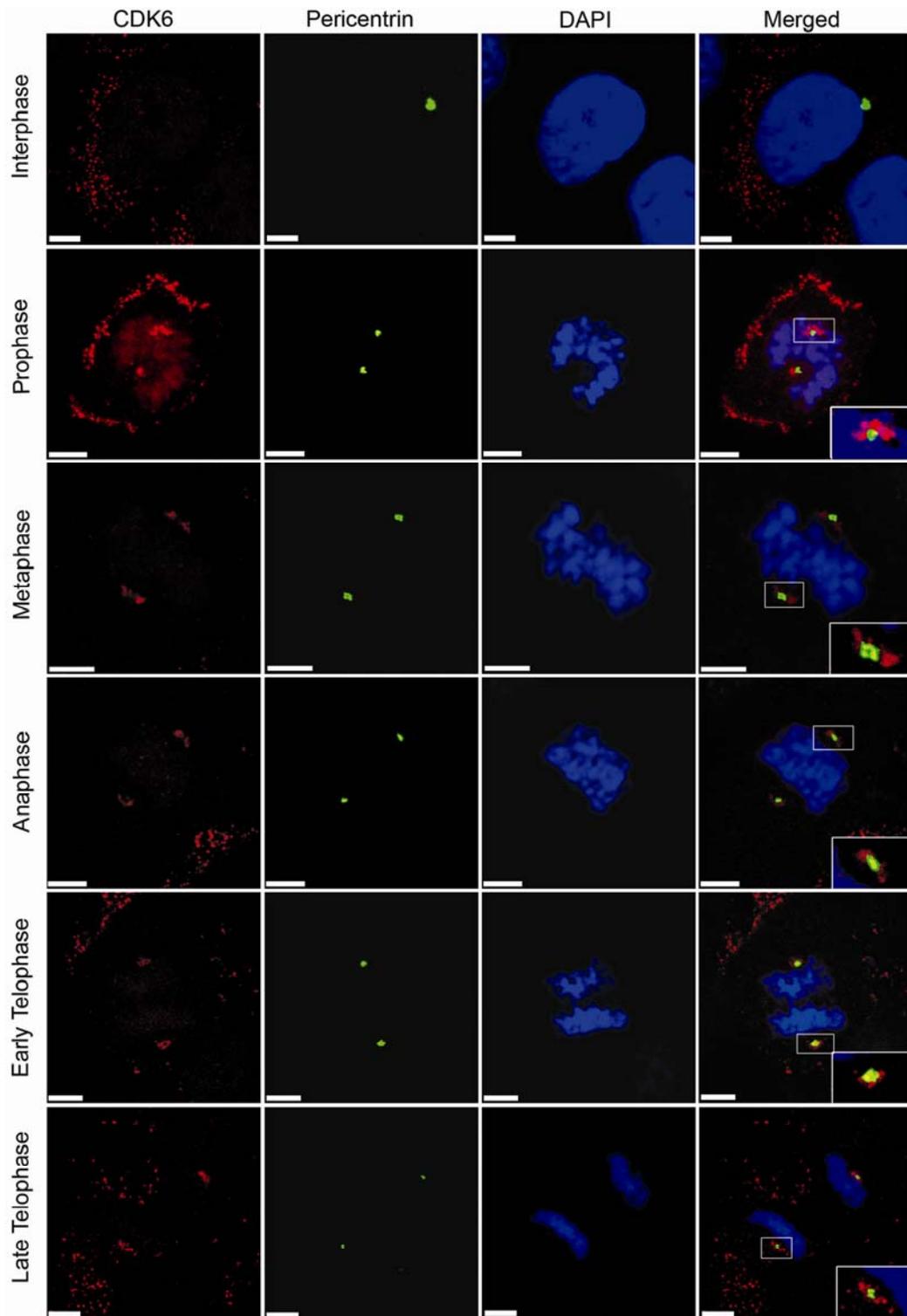
The missense mutation (c.589G>A) in *CDK6* replaces an alanine with a threonine (p.A197T). Alanine in this position is highly conserved in primates and some other animals but is not conserved in other CDKs except *CDK4* which is considered the closest homolog of *CDK6* (Figure 4.27).

Mus	VTLWYRAPEVLLQSSY <b>A</b> TPVDLWSVGCIFAEMF
Rattus	VTLWYRAPEVLLQSSY <b>A</b> TPVDLWSVGCIFAELF
Homo	VTLWYRAPEVLLQSSY <b>A</b> TPVDLWSVGCIFAEMF
Macaca	VTLWYRAPEVLLQSSY <b>A</b> TPVDLWSVGCIFAEMF
Pan	VTLWYRAPEVLLQSSY <b>A</b> TPVDLWSVGCIFAEMF
Canis	VTLWYRAPEVLLQSSY <b>A</b> TPVDLWSVGCIFAEMF
Felis	VTLWYRAPEVLLQSSY <b>A</b> TPVDLWSVGCIFAEMF
Loxodonta	VTLWYRAPEVLLQSSY <b>A</b> TPVDLWSVGCIFAEMF
Monodelphis	VTLWYRAPEVLLQSSY <b>A</b> TPVDLWSVGCIFAEMF
Gallus	VTLWYRAPEVLLQSSY <b>A</b> TPVDLWSVGCIFAEMF
Xenopus	VTLWYRAPEVLLQSSY <b>A</b> TPVDLWSVGCIFAEMF
	*****:*
Human CDK6	ALTSVVVTLWYRAPEVLLQSS-Y <b>A</b> TPVDLWSVGCIFAEMFRR-KPLFRGSSD
Human CDK4	ALTPVVVTLWYRAPEVLLQST-Y <b>A</b> TPVDMWSVGCIFAEMFRR-KPLFCGNSE
Human CDK2	TYTHEVVTLWYRAPEILLGCKYY <b>S</b> TAVDIWSLGCIFAEMVTR-RALFPGDSE
Human CDK3	TYTHEVVTLWYRAPEILLGSKFY <b>T</b> TAVDIWSIGCIFAEMVTR-KALFPGDSE
Human CDK5	CYSAEVVTLWYRPPDVLFGAKLY <b>S</b> TSIDMWSAGCIFAELANAGRPLFPGNDV
Human CDK10	PMTPKVVTLWYRAPELLLGT <b>T</b> TTQ <b>T</b> TSIDMWAVGCILAELLAH-RPLLPGTSE
Human CDK9	RYTNRVVTLWYRPELLLGERDY <b>G</b> PPIDLWGAGCIMAEMWTR-SPIMQGNTE
Human CDK8	DLDPVVVTFWYRAPELLLGARHY <b>T</b> KAIDIWAIGCIFAELLTS-EPIFHCRCQE

**Figure 4.27: Cross species alignment of CDK6 sequences by ClustalW:** Multiple Sequence Alignment (2.0.5). The amino acid Alanine (A, blue) is conserved in all species but is not conserved in different human CDKs (A, red).

#### 4.8.7. Sub-cellular localization of CDK6

Endogenous and transfected CDK6 was reported to localize in the cytoplasm mainly in the cortical region and in cellular protrusions and in the nucleus (Kohrt et al., 2009). I studied the subcellular distribution of CDK6 by immunocytochemistry and confocal microscopy in the human keratinocyte (HaCaT) cell line during the cell cycle. This revealed a cytoplasmic and nuclear localization of CDK6 during interphase, while a strong accumulation at the spindle pole (centrosome) was observed during mitosis from prophase to telophase as shown in figure 4.28.



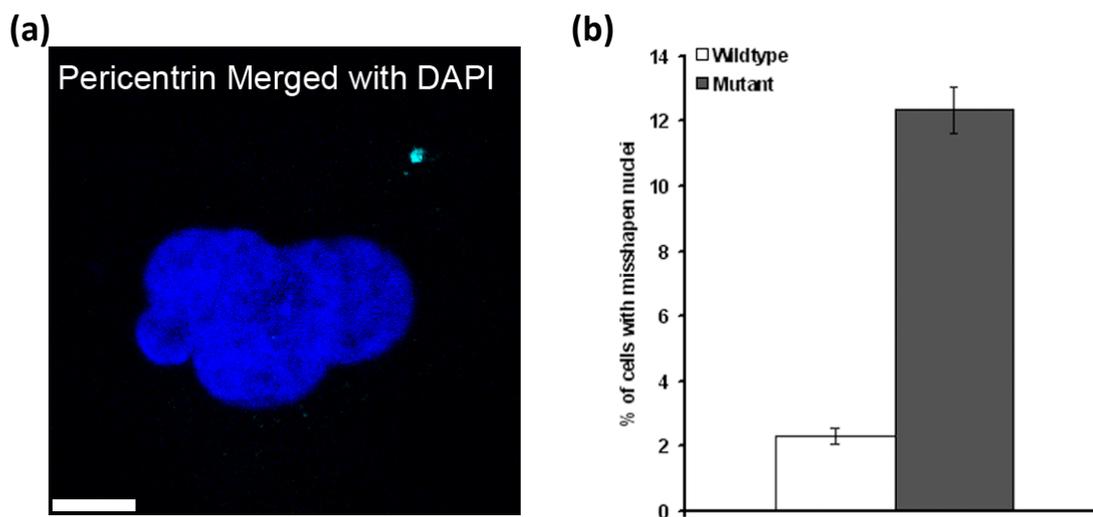
**Figure 4.28: Subcellular localization of CDK6 in HaCaT cells.** CDK6 localizes to the cytoplasm and nucleus during interphase and to the centrosome (spindle pole) during prophase to telophase. Confocal microscopy was performed in HaCaT cells stained for CDK6 (red), centrosomal marker Pericentrin (green) and DAPI (blue) as a DNA marker. Scale bar, 5  $\mu$ m. The insets show centrosomes at higher magnification.

#### 4.8.8. Cell cycle synchronization of CDK6 patient fibroblasts

Fibroblasts from a CDK6 patient (MCP4-3) and control fibroblasts (passage 4) were synchronized at G0/G1 by serum starvation. The cells were then processed for immunofluorescence after 24, 26, 28 and 30 hours. Immunocytochemistry and confocal microscopy of synchronized mutant fibroblasts showed the following phenotypes of misshapen nuclei, reduced cell proliferation and increased centrosome-nucleus distance, which were not observed in the control primary fibroblasts.

##### 4.8.8.1. Misshapen nuclei

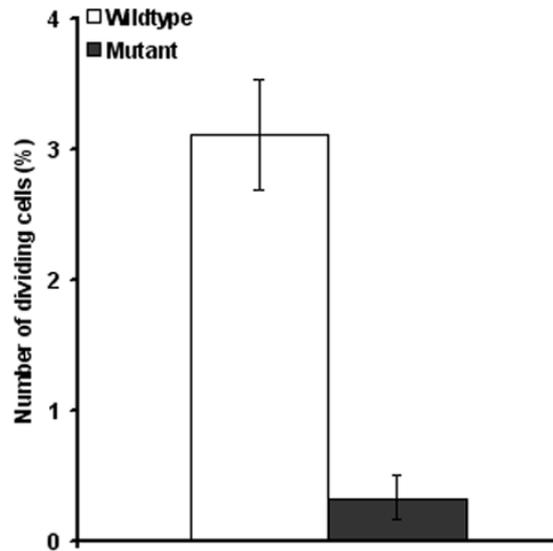
Normally, fibroblast nuclei have an oval shape. In primary fibroblasts of the CDK6 patient MCP4-3, I frequently observed nuclei that were no longer oval. Instead, they had different shapes and were lobulated (Figure 4.29a and 4.31b). The statistical analysis revealed that ~12.3 % of CDK6 patient fibroblasts had misshapen nuclei as compared to controls where I found abnormal nuclear shapes in ~2.3% of all cells as shown in figure 4.29b.



**Figure 4.29: CDK6 patient primary fibroblasts with misshapen nuclei and statistical analysis.** (a) Confocal microscope images of CDK6 mutant primary fibroblasts showing dysmorphic nuclei. The cells were stained for Pericentrin (turquoise) and DAPI (blue) harbouring dysmorphic nuclei. Scale bar, 5  $\mu\text{m}$ . (b) Histogram showing the presence of dysmorphic nuclei in wild type and CDK6 mutant primary fibroblasts. 900 cells were analysed each. Error bars represent s.e.m.  $P$  value,  $1.83477\text{E}-06$  (Student's  $t$ -test).

##### 4.8.8.2. Reduced cell proliferation

I also noted a slower growth for mutant fibroblasts in culture. When I analysed cell proliferation by determining the number of dividing CDK6 mutant and control primary fibroblasts that had been synchronised at G0/G1, I found ~3 % of control cells in different stages of the cell cycle whereas this number was significantly reduced in the mutant with only ~0.3 % of dividing cells confirming an altered proliferation rate (Figure 4.30).

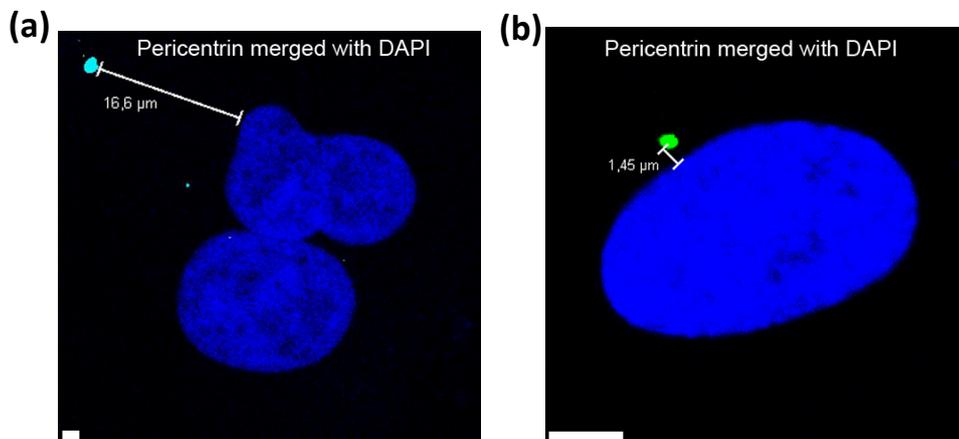


**Figure 4.30: Statistical data of aberrant cell division in wild-type and CDK6 mutant primary fibroblasts synchronized at G0/G1.** ~3% of control cells were found dividing as compared to the mutant cells where only ~0.3% of cells were dividing confirming an altered proliferation rate. 900 cells were counted. Error bars represent s.e.m. *P* value, 1.2968E-04 (Student's *t*-test).

#### 4.8.8.3. Increased centrosome-nucleus distance

Centrosomes detected with CDK6 as well as with pericentrin specific antibodies of synchronized patient primary fibroblasts were frequently detached from nuclei and exhibited a significantly increased distance as shown in figure 4.31a and table 4.3. In contrast, the control synchronized primary fibroblast cells exhibited normal nuclear shape and the centrosomes were not detached from the nucleus (Figure 4.31b).

Statistical data taken from 50 cells each showed a significant difference in the nucleus-centrosome distance measured in patient versus control fibroblasts (Table 4.3).



**Figure 4.31: Confocal microscope images of CDK6 patient and control primary fibroblasts showing the nucleus stained with DAPI and the centrosome.** (a) Mutant cells were stained for Pericentrin (turquoise) and DAPI (blue). The cell exhibits an increased centrosome-nucleus distance. The nucleus of this cell has also the characteristic of misshapen nuclei. (b) Wild-type primary fibroblast stained for Pericentrin (green) and DAPI (blue). The centrosome-nucleus distance was determined by Lecia Microsystem LAS AF lite. Scale bar, 5  $\mu$ m.

**Table 4.3: Centrosomes-nucleus distance measured in CDK6 patient and control primary fibroblast cells.** Results are shown in terms of mean  $\pm$  s.e.m and the *P* values (Student's *t*-test) which are calculated with control versus mutant data.

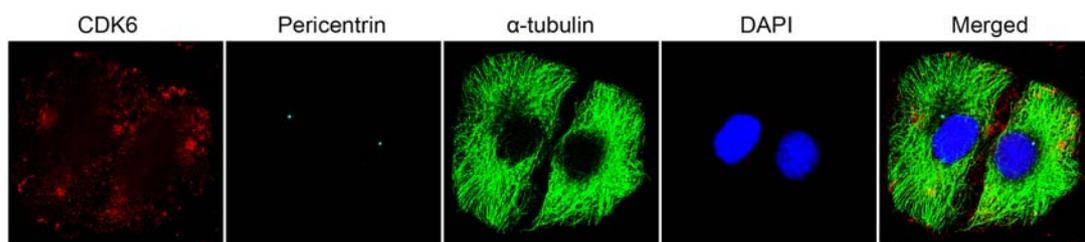
Centrosome-Nucleus distance ( $\mu$ m) (N=50)	
Wild-type primary fibroblast	Mutant primary fibroblast
3.4514 $\pm$ 0.6062	5.1646 $\pm$ 0.2661
<i>P</i> value	0.008286

#### 4.8.9. CDK6 knockdown

To obtain independent support for a role of CDK6 in cell proliferation and at the centrosome, I carried out Knockdown experiments in HaCaT cells using shRNA. The efficiency of the Knockdown was first tested by western blotting. A faint band of 37 kDa was detected in Knockdown cells while in control (transfected HaCaT cells with construct having scrambled sequence of *CDK6*), a strong signal was observed. To confirm equal protein loading,  $\beta$ -actin (42 kDa) was detected (Figure 4.33b).

##### 4.8.9.1. Normal microtubule network in control

I first analysed the microtubule network in control HaCaT cells. These cells *have* a normal oval nucleus and the microtubule form a regular network extending into the cell periphery (Figure 4.32).



**Figure 4.32: Control HaCaT cells with normal microtubules.** CDK6 expression was observed mostly in the cytoplasm of cells. Cells have normal microtubules and a regular shape of the nucleus. Cells were stained with CDK6 (red), Pericentrin (turquoise),  $\alpha$ -tubulin (green) and DAPI (blue). Scale bar, 5  $\mu$ m.

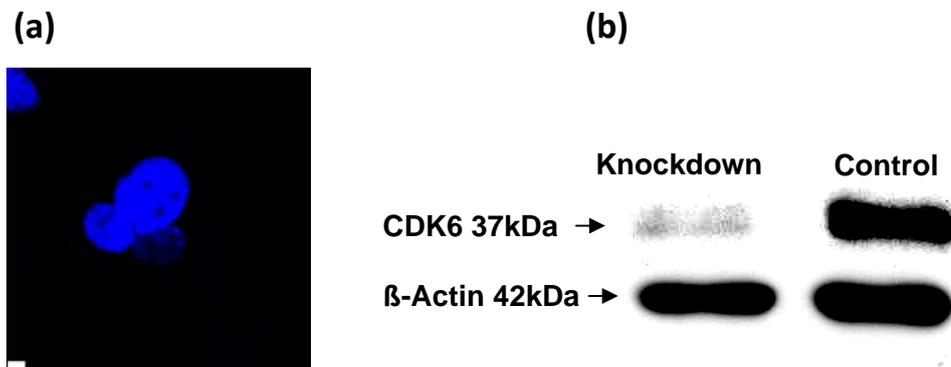
In the Knockdown I observed several changes as misshapen nuclei, reduced proliferation rate and disorganised microtubules.

#### 4.8.9.2. Misshapen nuclei

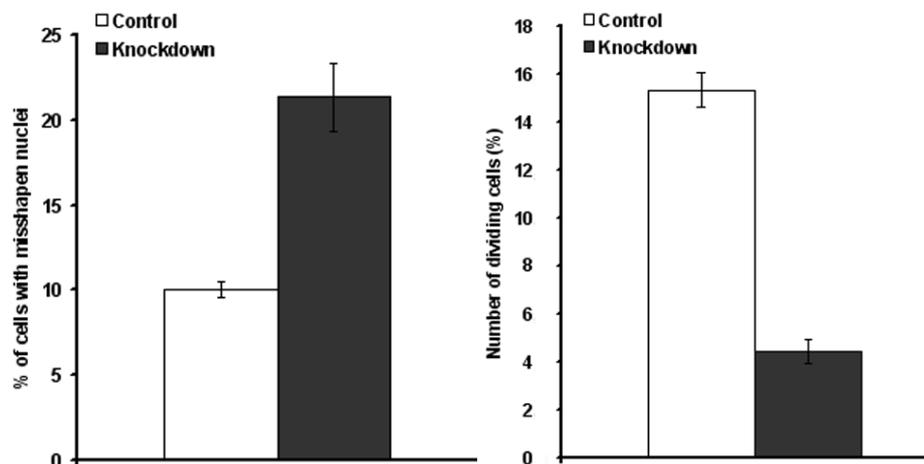
In CDK6 knockdown HaCaT cells, I found that ~21 % of all cells had dysmorphic nuclei as compared to 10 % in control cells (Figure 4.33a and 4.34a).

#### 4.8.9.3. Reduced cell proliferation

As observed for patient fibroblasts cell proliferation was also affected by reduced levels of CDK6 protein. When I analysed the percentage of dividing cells in knockdown HaCaT cells I found that 4.45 % of the cells were dividing vs 15.3 % of control cells which were observed at different stages of the cell cycle (Figure 4.34b).



**Figure 4.33: CDK6 mutant primary fibroblasts with western blot of CDK6 knockdown. (a)** Confocal microscope image of CDK6 knockdown HaCaT cell stained for DAPI (blue) harbouring dysmorphic nuclei. Scale bar, 5  $\mu$ m. **(b)** Western blot analysis of CDK6 knockdown and control cells showing the reduced amount of CDK6. The blots was probed with  $\beta$ -actin (42 kDa) as loading control.

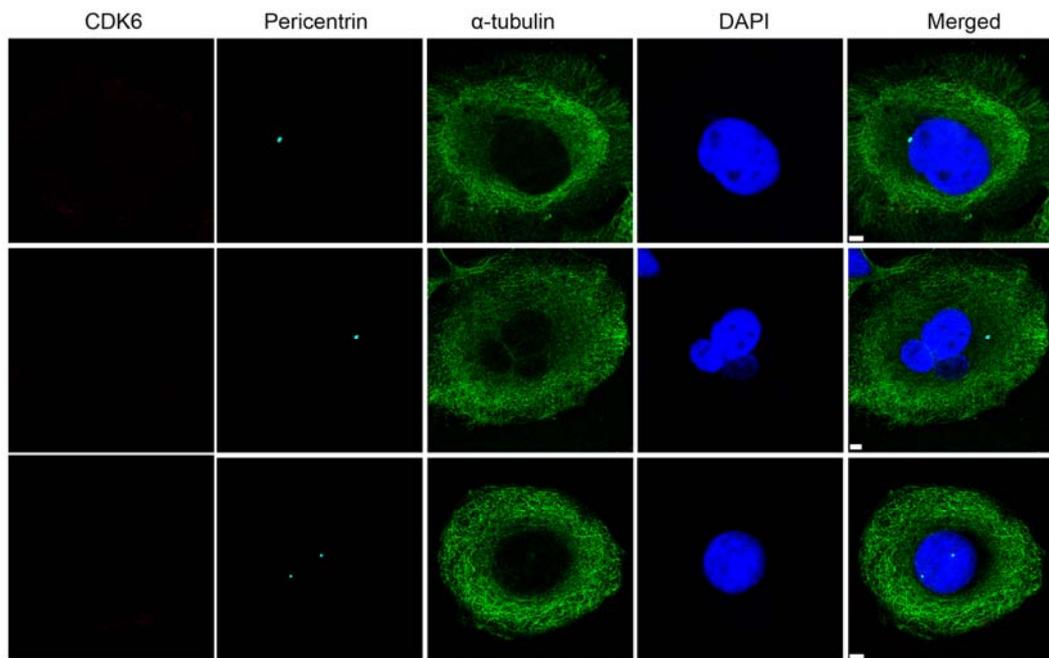


**Figure 4.34: Statistical Analysis. (a)** Statistical data of dysmorphic nuclei in CDK6 knockdown and control HaCaT cells. 900 cells each were analysed. Error bars represent

s.e.m. *P* value, 4.72E-04 (Student's *t*-test). **(b)** Histogram representing the statistical analysis of aberrant cell division in Knockdown cells. Data is taken from 900 cells. Error bars represent s.e.m. *P* value, 1.02517E-06 (Student's *t*-test).

#### 4.8.9.4. Disorganised microtubules

A prominent phenotype noticed in CDK6 knockdown HaCaT cells was the disorganised microtubule network (Figure 4.34). Normally microtubules extend from the centrosome out into the cytosol in form of long fibers. In CDK6 Knockdown cells I observed that the area around the nucleus was less populated by microtubules. Instead microtubules formed a fine filamentous meshwork near the periphery of the cells.



**Figure 4.35: CDK6 knockdown HaCaT cells.** Cells were stained for CDK6 to demonstrate the efficiency of the Knockdown (red), Pericentrin (turquoise),  $\alpha$ -tubulin (green) and DAPI (Blue). The cells showed disorganised microtubules. Scale bar, 5  $\mu$ m.

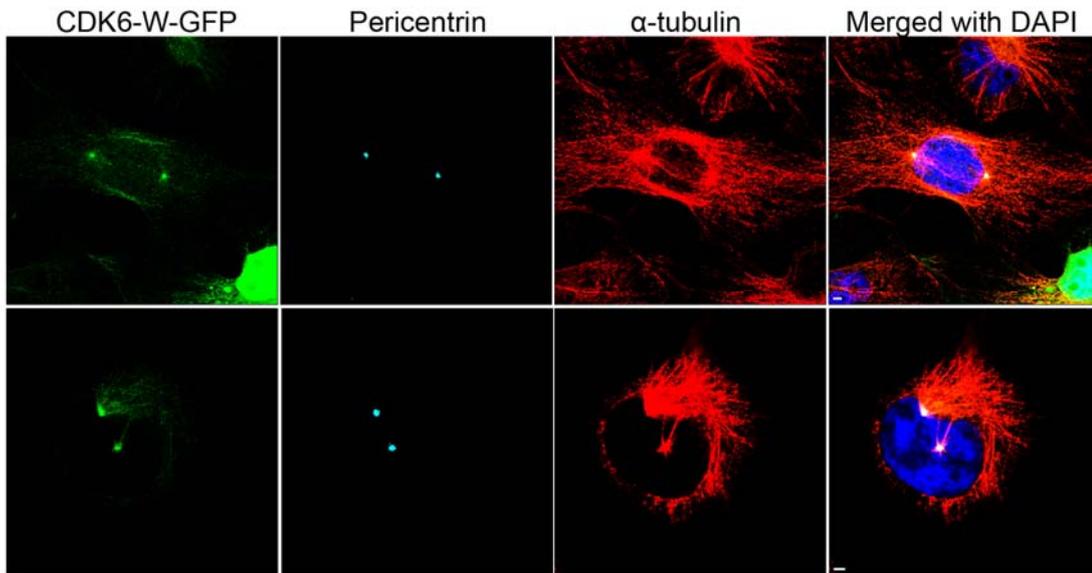
#### 4.8.10. Overexpression of CDK6

To check the ectopic expression of wild-type and mutant (c.589G>A) CDK6 protein, both versions were expressed with a GFP-tag in COS7 cells. I observed a reduced cell division rate for both wild-type and mutant transfected cells so, only few mitotic cells were detected. I also analysed the localization of the proteins and their effect on centrosomes and microtubule organisation

##### 4.8.10.1. Localization of GFP-tagged CDK6

During interphase CDK6 wild-type GFP-tagged protein localizes to the cytoplasm as well as the nucleus and associates with centrosomes in dividing cells resembling the observations with primary fibroblasts (Figure 4.36). The mutant CDK6 protein also

localizes to cytoplasm and in the nucleus in a same fashion like wild type, but I did not find dividing cells in this case to observe the localization of mutant CDK6-GFP (Figure 4.37b).



**Figure 4.36: Localization of GFP-tagged CDK6 in dividing cell.** COS7 cells were transfected with GFP-tagged wild-type CDK6 (green). Staining was for Pericentrin (turquoise),  $\alpha$ -tubulin (red) and DAPI (blue). Only few mitotic cells were found in these transfected cells. Scale bar, 10  $\mu$ m.

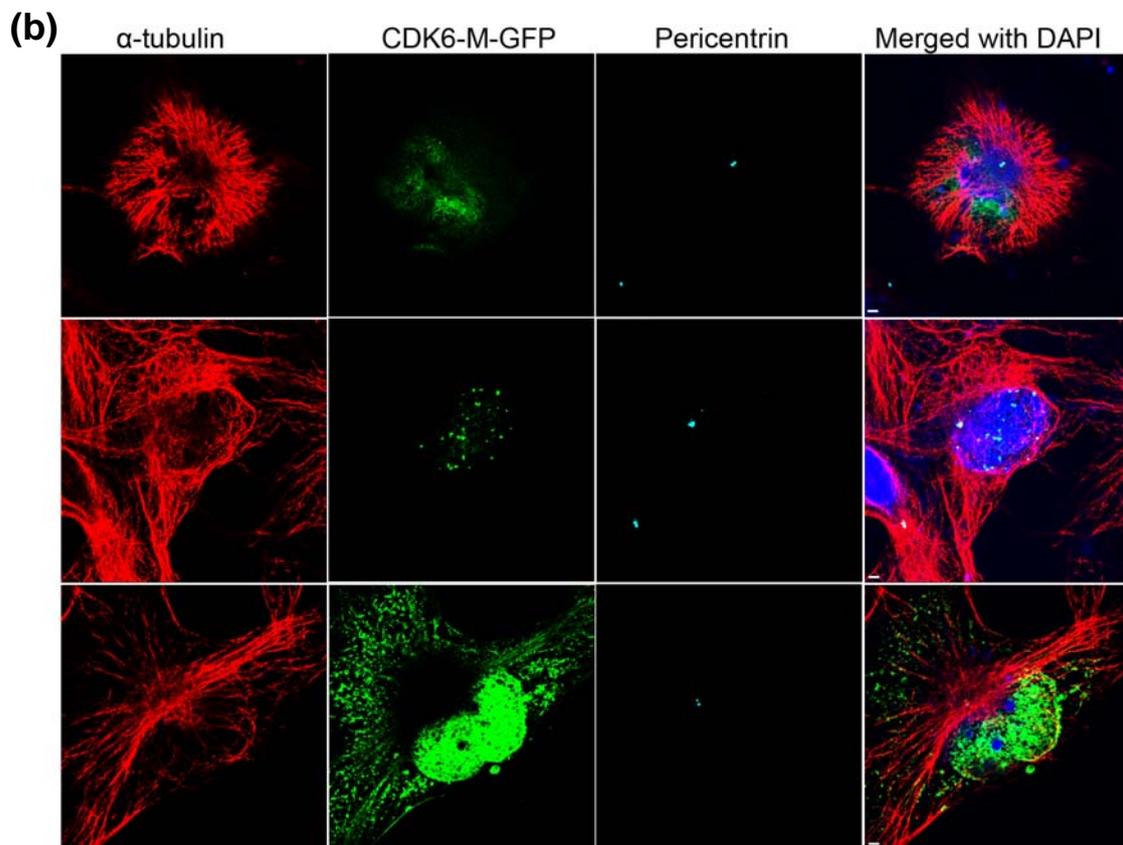
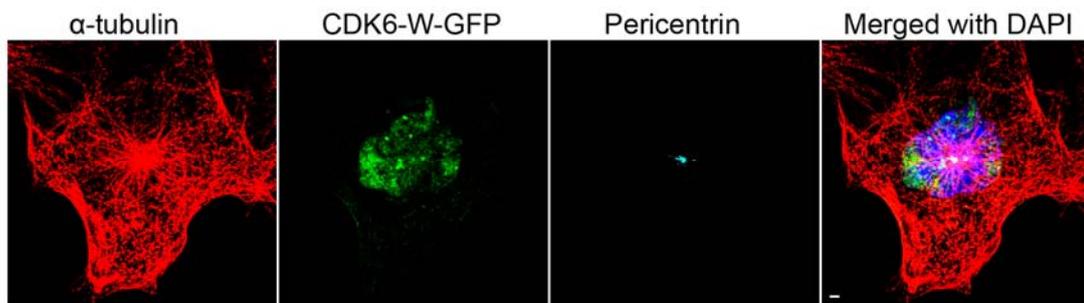
#### 4.8.10.2. Disorganised microtubules

Overexpression of both wild-type and mutant GFP tagged CDK6 resulted in an abnormal appearance of the microtubules. The pattern appeared irregular with a strong accumulation around the periphery in cells overexpressing wild-type CDKs whereas in the case of the mutant protein I observed shorter and in some cases fewer microtubules. Some of the transfected cells contained GFP-positive dots which appeared in the cytoplasm as well as in the nucleus of the cells. The size and number of these dots was variable (Figure 4.37).

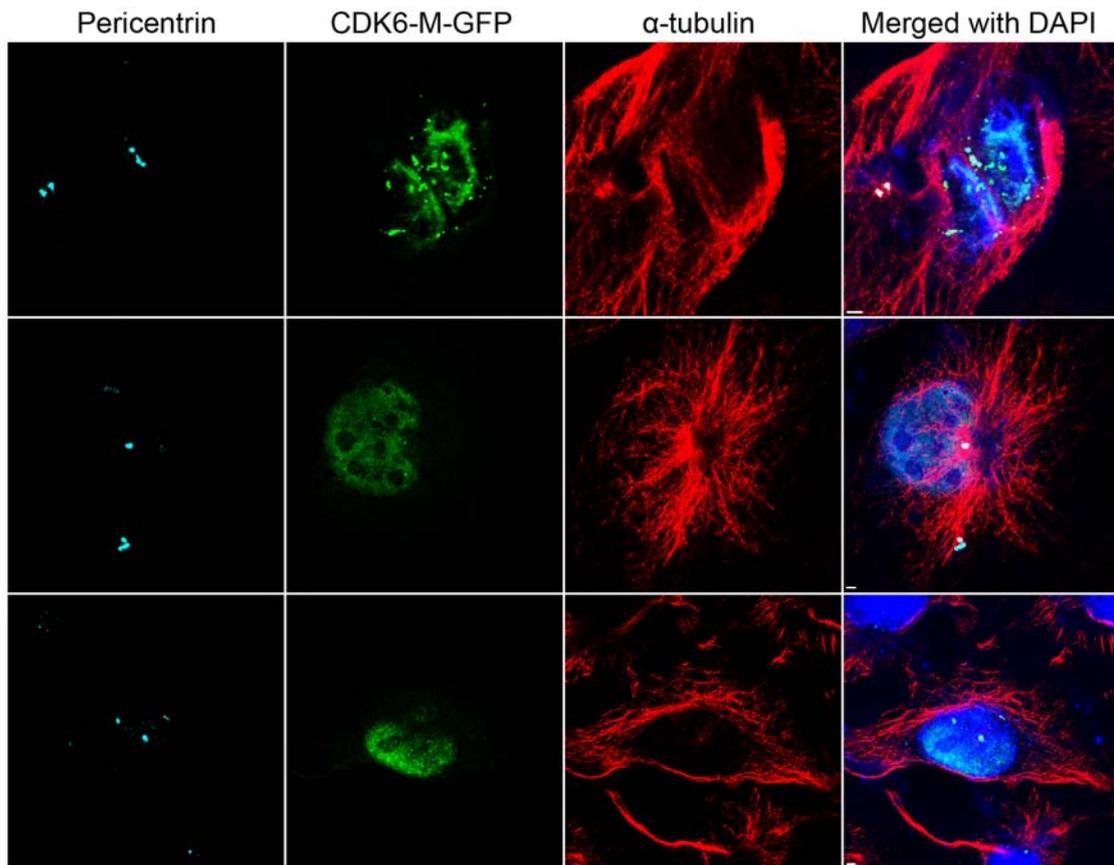
#### 4.8.10.3. Multiple centrosomes

A distinguishing feature noticed in cells expressing mutant CDK6 was the presence of multiple centrosomes. This was not observed in cells overexpressing wild-type protein. The centrosome number was increased from 3 to 5, furthermore, they appeared fragmented. Some were closely attached and others were found apart from each other. Interestingly these centrosomes frequently did not serve as microtubule organising centers and the cells exhibited abnormal microtubule networks (Figure 4.38).

**(a)**



**Figure 4.37: Disorganised microtubule system in CDK6 transfected cells.** Confocal microscopy images of cells overexpressing GFP-tagged wild-type CDK6 **(a)** and mutant CDK6 **(b)**. GFP fusion of CDK (green), Pericentrin (turquoise),  $\alpha$ -tubulin (red) and DAPI (blue). Scale bar, 10  $\mu$ m.

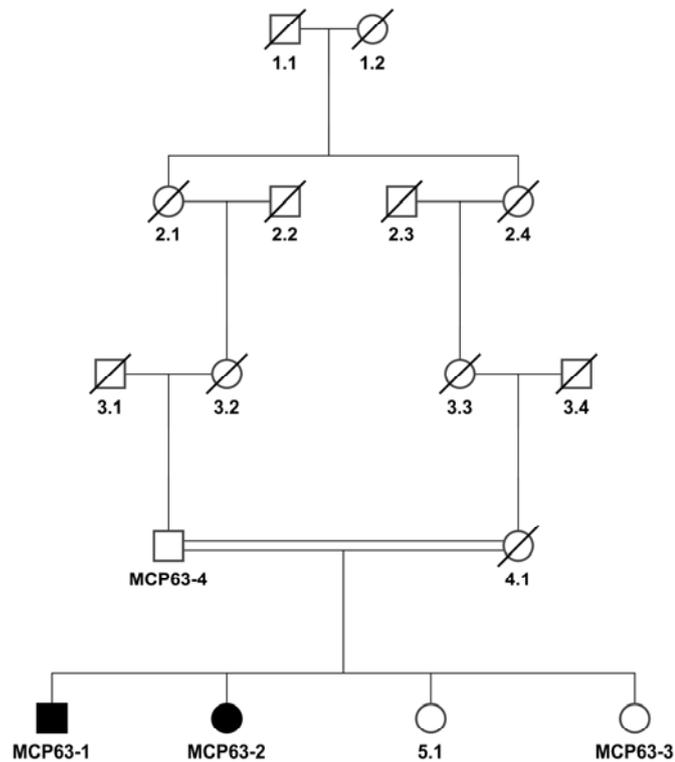


**Figure 4.38: Multiple numbers of centrosomes in cells expressing mutant CDK6.** COS7 cells were transfected with GFP-tagged mutant CDK6. Confocal microscopy images demonstrate multiple numbers of centrosomes. Some cells have three while others have five centrosomes. The cells were stained for Pericentrin (turquoise),  $\alpha$ -tubulin (red) and DAPI (blue). Mutant CDK6-GFP is shown in green. The cells with multiple centrosomes also have a disorganised microtubule network. Scale bar, 5  $\mu$ m.

#### 4.9. Excluded family MCP63

MCP63 is a consanguineous family with two afflicted individuals (Figure 4.39) and was ascertained from the northern region of Pakistan. The patients have a typical sloping forehead and are mentally retarded without any other abnormalities. After exclusion of linkage with all known seven MCPH loci, the family was subjected to a genome-wide screen for regions of linkage with the Affymetrix GeneChip(R) Human Mapping 250K SNP Array (version 2.0) on the two afflicted children (MCP63-1 and MCP63-2), one normal individual (MCP63-3), and the normal father (MCP63-4).

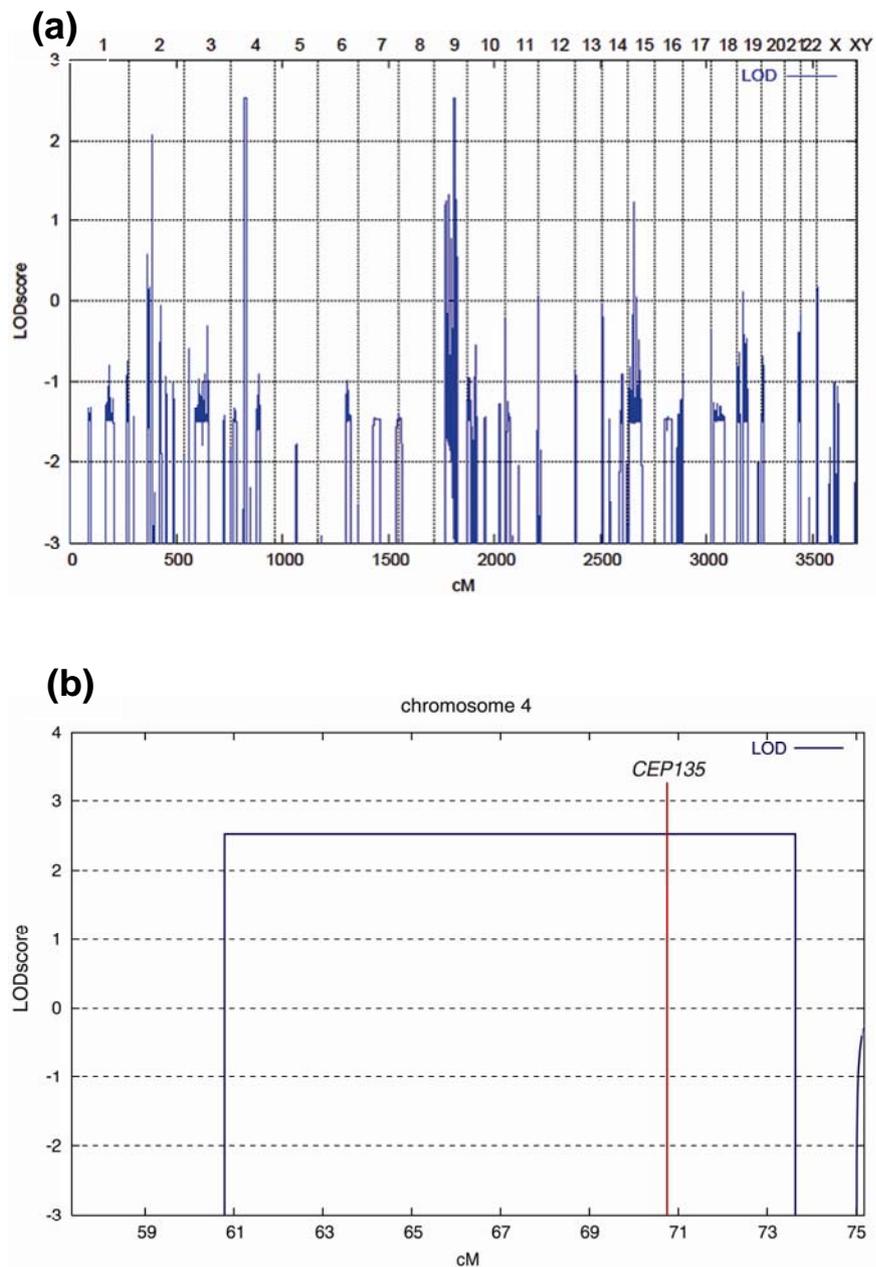
## Family MCP63



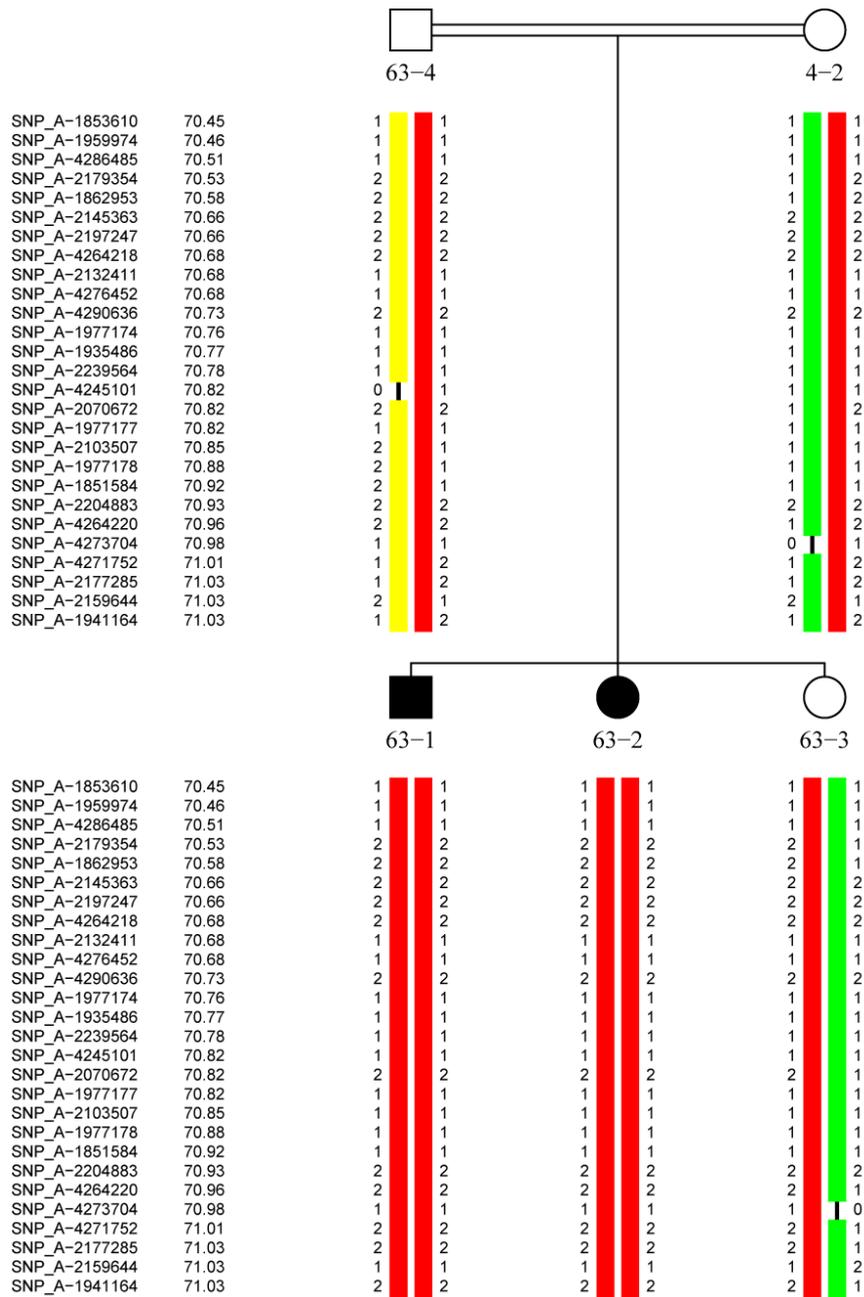
**Figure 4.39:** Pedigree of family MCP63 having the novel MCPH9 locus. Four individuals were subjected to this study.

### 4.9.1. MCPH9 locus identification

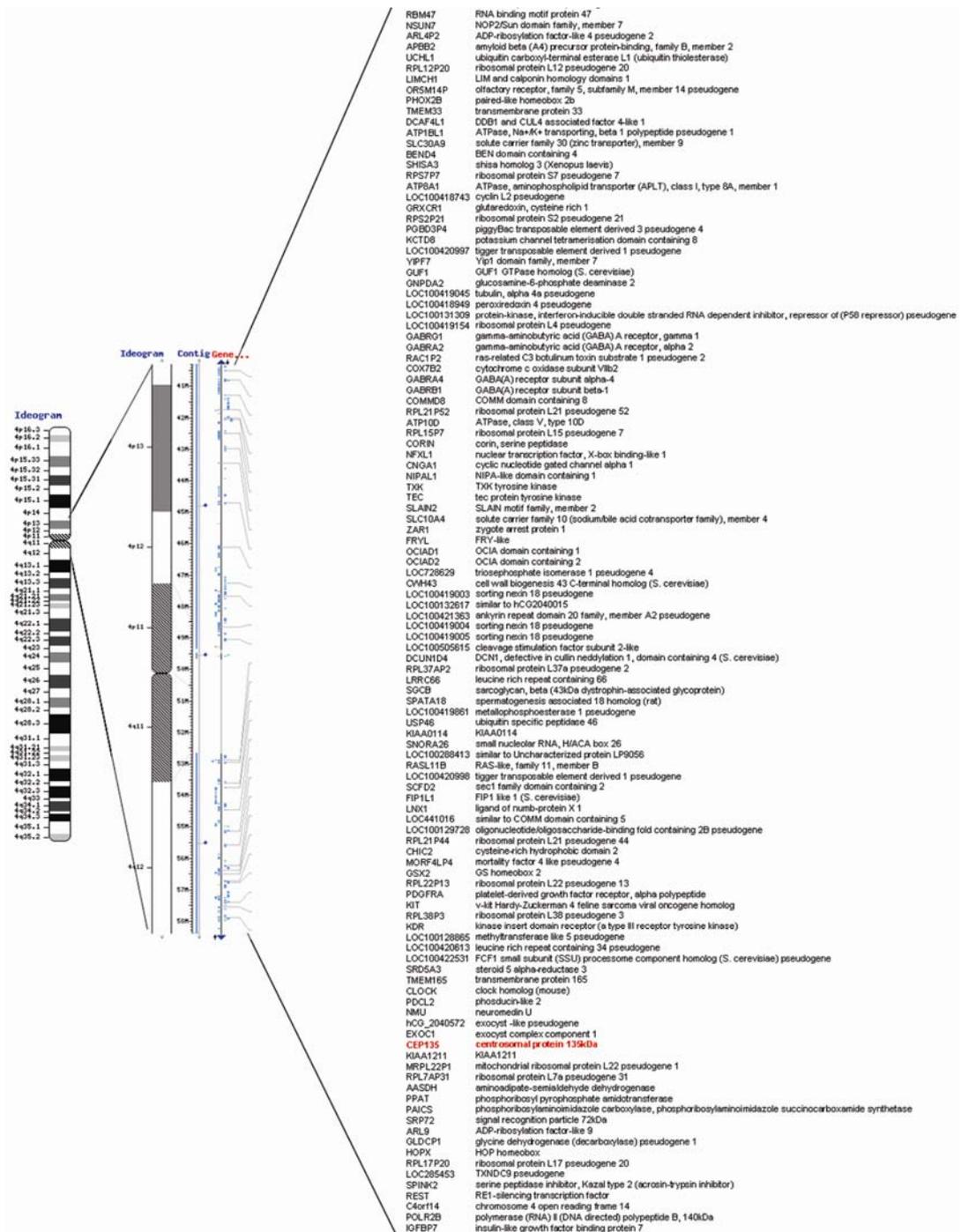
Genome-wide linkage analysis of the MCP63 family revealed the identification of one novel gene locus MCPH9 which mapped on chromosome 4p14-4q12 with LOD score 2.53 between SNP\_A-2154951 (rs12498424, position 40,326,233bp) and SNP\_A-1894332 (rs13134527 position 58,396,887bp) comprising 18.07 Mb and 12.84 cM (Figure 4.40a and 4.41). Genome-wide analysis of the MCP63 family also revealed two narrow peaks on chromosome number 9 which were also taken into consideration. One peak maps between SNPs rs7044445, (physical position 91,463,981bp) and rs10993082(physical position 95,980,440 bp) on chromosome 9q22.31-9q22.32 and the other one maps between SNPs rs10993188 (physical position 96,201,711bp) and rs642771 (physical position 100,162,135bp) on chromosome 9q22.32-9q31.1. But the causative gene was found only in the 4p14-4q12 region which is the actual MCPH9 locus (Figure 4.40b).



**Figure 4.40: Genome-wide graphical view of the MCPH9 locus.** A genome-wide screen for regions of linkage was performed with the Affymetrix 250K SNP GeneChip. A marker panel of 126,000 SNPs was used (a) Parametric linkage analysis (Allegro software) showing the MCPH9 locus mapped in two afflicted and two normal individuals. The highest LOD score (2.58) was obtained on chromosome 4p14-4q12. (b) Genome-wide view of the MCPH9 locus on chromosome 4 with the position of CEP135 which is shown by the red line.



**Figure 4.41: Haplotype analysis of the MCP63 family linked to chromosome 4p14-4q12.** Black symbols represent afflicted individuals. The critical interval was defined by the recombinations observed between markers SNP\_A-2154951 (rs12498424, physical position 40,326,233bp) and SNP\_A-1894332 (rs13134527 physical position 58,396,887bp) comprising 18.07 Mb and 12.84 cM. Here only the region near *CEP135* is shown.



**Figure 4.42: G-banded chromosome 4 showing the position of the MCPH9 locus, along with the position of CEP135 on short arm 4p14-4q12.** The locus was mapped between SNP\_A-2154951 (rs12498424, physical position 40,326,233bp) and SNP\_A-1894332 (rs13134527, physical position 58,396,887bp). This 18.07 Mb (12.84 cM) region harbours 116 genes (NCBI Human Genome Build 36.3). Some genes with hypothetical proteins are skipped. The CEP135 gene was selected for sequencing because it has the centrosomal localization and plays a role in the cell cycle. Deletion of a single base c.970delC was found which causes frameshift incorporating one amino acid and then causing a stop codon (p.Q324Sfs\*2). This mutation was not found in 768 healthy controls of Pakistani origin.

#### 4.9.2. Candidate genes of the MCPH9 locus

Bioinformatic analysis of the MCPH9 (18.07 Mb) region was performed to seek the candidate genes. This region on chromosome 4p14-4q12 harbours 116 known and hypothetical genes (NCBI Human Genome Build 36.3) (Figure 4.42).

#### 4.9.3. Sequencing the candidate genes in the MCPH9 locus

A role in mitosis and/or in neurogenesis, centrosomal or spindle pole localization or embryonic brain expression was the selection criterion to seek the candidate genes of the MCPH9 locus. Moreover, the genes were prioritized by tools like Endeavour (Aerts et al., 2006) (Figure 4.43).

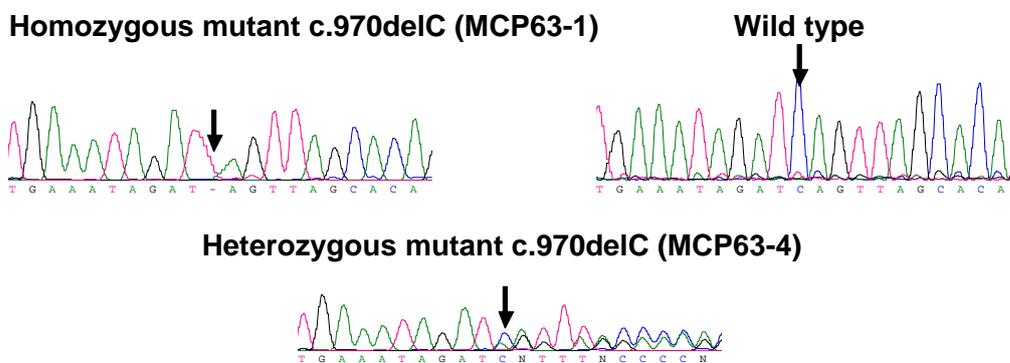
Global prioritization	Annotation Ensembl	Annotation GeneOntology	Annotation Keeg	Annotation Swissprot	Blast	CisRegModule	Expression SonEtA1	Motif	Precalculated Ouzomis	Precalculated Prosper	Text
CEP135	KIT	KIT	KIT	COX7B2	CEP135	KIT	SPINK2	KIT	CLOCK	PDGFRA	CEP135
KIT	NMU	NMU	COX7B2	GABRB1	SRP72	CLOCK	NMU	SCFD2	PDGFRA	GABRG1	SLAIN2
PAICS	DCUN1D4	DCUN1D4	PAICS	NPAL1	EXOC1	REST	REST	DCUN1D4	KIT	GABRA4	GUF1
GUF1	COX7B2	COX7B2	TEC	IGFBP7	NSUN7	TMEM165	COMMD8	TXK	KDR	ATP10D	GABRG1
NSUN7	SPATA18	SPATA18	SRP72	C4orf14	ATP8A1	KCTD8	SLAIN2	COX7B2	PHOX2B	KIT	ATP10D
PDGFRA	COMMD8	ARL9	PPAT	KCTD8	SLAIN2	RASL11B	NFXL1	PDGFRA	GUF1	KDR	NPAL1
SLAIN2	ARL9	PAICS	UCHL1	GUF1	CCDC4	NFXL1	CLOCK	TMEM33	UCHL1	GABRA2	SCFD2
CLOCK	PAICS	SLC10A4	KDR	TXK	GUF1	GABRA2	NSUN7	SLC30A9	SGCB	SLC10A4	NSUN7
LNK1	SLC10A4	NPAL1	GABRG1	TEC	CLOCK	FIP1L1	PPAT	REST	ZAR1	POLR2B	SPATA18
ATP10D	NPAL1	TEC	LNK1	SLC30A9	PDGFRA	CHIC2	GABRA2	YIPF7	SRP72	ATP8A1	TMEM33
REST	NSUN7	OC1AD2	POLR2B	ZAR1	COX7B2	NMU	USP46	RASL11B	NSUN7	PPAT	WDR21B
C4orf14	TEC	NFXL1	GABRA2	UCHL1	UCHL1	SLC10A4	OC1AD1	ATP10D	C4orf14	TEC	TMEM165
EXOC1	OC1AD2	CHIC2	GNPDA2	SRP72	PDCL2	PDGFRA	SGCB	IGFBP7	CORIN	FRYL	C4orf14
GABRB1	NFXL1	IGFBP7	PDGFRA	SPATA18	PAICS	LNK1	ARL9	NPAL1	OC1AD1	C4orf14	TEC
TXK	CHIC2	SRP72	CLOCK	PAICS	NFXL1	PHOX2B	GUF1	OC1AD1	GABRG1	GABRB1	FIP1L1
KCTD8	IGFBP7	C4orf14	GABRA4	SCFD2	GABRB1	ATP10D	TXK	CEP135	PAICS	IGFBP7	RASL11B
SRP72	SRP72	APBB2	GABRB1	NSUN7	APBB2	CORIN	SLC30A9	CORIN	PPAT	RASL11B	UCHL1
DCUN1D4	C4orf14	FIP1L1	TXK	CLOCK	KIT	SPATA18	LNK1	ARL9	REST	KCTD8	USP46
SLC30A9	APBB2	PPAT	NMU	ATP8A1	TEC	APBB2	SRP72	PDCL2	SPINK2	CNGA1	GABRA4
CHIC2	FIP1L1	UCHL1	DCUN1D4	ATP10D	NPAL1	COX7B2	PAICS	NSUN7	EXOC1	PHOX2B	REST
RASL11B	WDR21B	KDR	SPATA18	SGCB	IGFBP7	UCHL1	GABRB1	GABRB1	SLC10A4	APBB2	GNPDA2
GABRG1	PPAT	CNGA1	ARL9	CNGA1	C4orf14	TEC	FIP1L1	CNGA1	RASL11B	GUF1	EXOC1
COX7B2	UCHL1	KCTD8	SLC10A4	DCUN1D4	KCTD8	NPAL1	CHIC2	SRP72	SLC30A9	FIP1L1	DCUN1D4
ATP8A1	KDR	SGCB	NPAL1	LNK1	TXK	SCFD2	ATP8A1	ATP8A1	ATP10D	PAICS	PAICS
KDR	CNGA1	GABRG1	OC1AD2	CHIC2	SLC30A9	CNGA1	CEP135	LNK1	YIPF7	SLC30A9	LNK1
GABRA4	KCTD8	SPINK2	NFXL1	FIP1L1	ZAR1	KDR	POLR2B	ZAR1	COMMD8	NMU	GABRA2
CCDC4	SGCB	FRYL	CHIC2	FRYL	SPATA18	GABRG1	APBB2	EXOC1	LNK1	CORIN	POLR2B
NPAL1	YIPF7	USP46	IGFBP7	EXOC1	SCFD2	GABRA4	RASL11B	TMEM165	POLR2B	TXK	PHOX2B
POLR2B	GABRG1	GUF1	C4orf14	CEP135	ATP10D	ARL9	GNPDA2	SGCB	DCUN1D4	NPAL1	GABRB1
OC1AD1	SPINK2	LNK1	APBB2	CCDC4	SGCB	TMEM33	NPAL1	PAICS	OC1AD2	TMEM165	IGFBP7
IGFBP7	FRYL	CORIN	FIP1L1	PDGFRA	CNGA1	OC1AD2	EXOC1	PPAT	GABRA2	USP46	OC1AD1
UCHL1	SLAIN2	POLR2B	CNGA1	REST	DCUN1D4	COMMD8	KDR	COMMD8	TMEM165	UCHL1	TXK
SPINK2	USP46	ATP8A1	KCTD8	KDR	LNK1	YIPF7	C4orf14	CCDC4	GNPDA2	SCFD2	SLC10A4
TEC	CEP135	EXOC1	SGCB	NFXL1	CHIC2	CEP135	ATP10D	TEC	KCTD8	LNK1	OC1AD2
PPAT	GUF1	GABRA2	SPINK2	PHOX2B	FIP1L1	SRP72	DCUN1D4	CHIC2	GABRA4	DCUN1D4	NFXL1
TMEM165	CCDC4	GNPDA2	FRYL	APBB2	FRYL	EXOC1	IGFBP7	GABRA4	CEP135	CHIC2	SRP72
FIP1L1	LNK1	TMEM165	USP46	CORIN	REST	NSUN7	KIT	KCTD8	TXK	EXOC1	CHIC2
GABRA2	CORIN	PDGFRA	GUF1	WDR21B	KDR	ATP8A1	TMEM165	POLR2B	USP46	ARL9	APBB2
SPATA18	POLR2B	SLC30A9	CORIN	KIT	PHOX2B	SLAIN2	UCHL1	APBB2	FRYL	CLOCK	COX7B2
SCFD2	ATP8A1	SCFD2	ATP8A1	PPAT	CORIN	CCDC4	TMEM33	SLAIN2	NMU	SLAIN2	PDCL2
USP46	EXOC1	CLOCK	EXOC1	GABRG1	WDR21B	GUF1	TEC	USP46	COX7B2	NSUN7	CORIN

Figure 4.43: Prioritized genes of the MCPH9 locus by Endeavour. Green color represents CEP135. The position of CEP135 is noticeable at the highest rank.

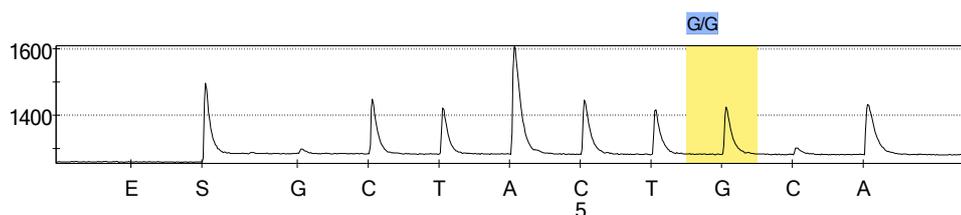
The centrosomal Protein 135 kDa (*CEP135*) on chromosome 4p14-4q12 and cell division cycle 14 homolog B (*CDC14B*) on chromosome 9q22.32-9q22.33 were considered the strongest candidates and were thus sequenced in MCP63 family members (Figure 4.43).

#### 4.9.4. *CEP135* as the causative gene of MCPH9

Sequence analysis showed no mutation in *CDC14B* but a homozygous single base deletion c.970delC in exon 8 of the *CEP135* gene (Figure 4.44). The mutation segregates in a recessive pathogenic way. The parents are heterozygous carriers and this deletion was not found in 768 Pakistani healthy control chromosomes (Figure 4.45). Deletion of a single base (c.970delC) of *CEP135* results in a frameshift which leads to a stop codon after one amino acid (p. Gln324Serfs\*2).



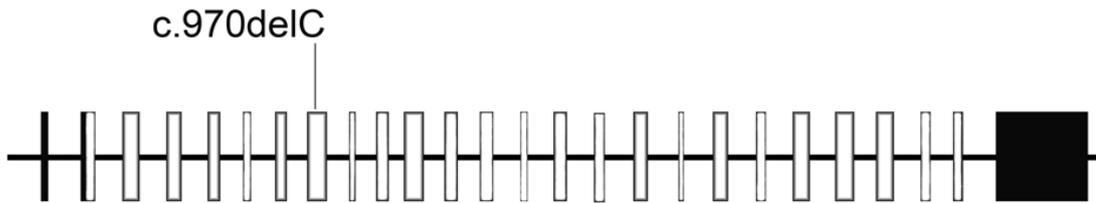
**Figure 4.44: Sequencing chromatograms of the MCPH9 patient with wild type that illustrate the c.970delC (p. Gln324Serfs\*2) mutation in the MCPH8 family (*CEP135* gene) by sanger sequencing.** Black arrows indicate the position of the mutation where C is deleted which results in a frameshift that leads to a stop codon after one amino acid (p. Gln324Serfs\*2). The mutant sequence is shown along with the wild type and one of the heterozygous parents. Mutation (c.970delC) was found in Pakistani family.



**Figure 4.45: Pyrogram with control samples from Pakistan.** Pyrogram traces obtained by Pyrosequencing analysis of the c.970delC mutation in exon 8 of Pakistani control samples which were negative for the c.970delC mutation in *CEP135*.

#### 4.9.5. Structure of *CEP135*

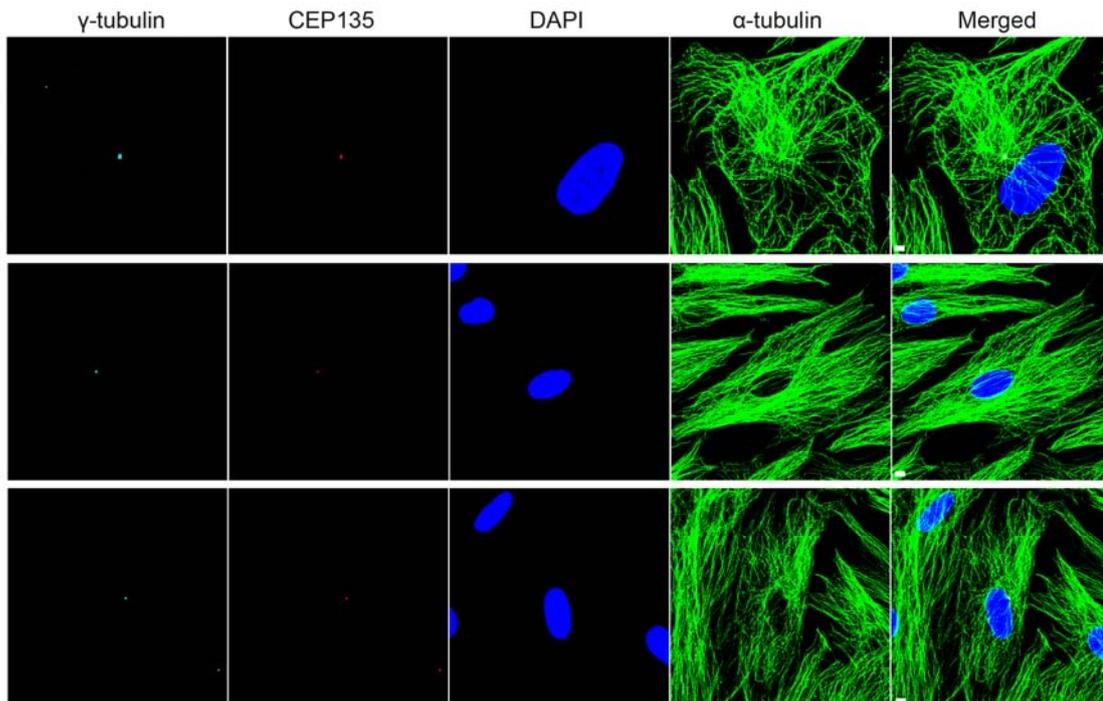
The *CEP135* gene is composed of 26 exons (Figure 4.46). The 3,423-bp ORF encodes a 1,140 amino acids protein. It gives rise to five transcripts of variable lengths (only two protein coding). The longest transcript (ENST00000257287) with 4,570 bps length codes for a protein of 1,140 amino acids.



**Figure 4.46:** The position of the *CEP135* mutation in MCPH9. The human *CEP135* gene consists of 26 exons which are drawn according to scale (1kb = 1cm), while introns are shown as artificial lines. Black boxes represent the non coding exons while white ones represent coding exons. Above exon 8 the position of mutation c.970delC is indicated.

#### 4.9.6. CEP135 in wild-type primary fibroblasts

I first studied the CEP135 staining in control fibroblasts. For this I used antibodies that were generated against the C-terminal region of human CEP135. These antibodies recognized a single spot in non dividing cells which overlapped with the  $\gamma$ -tubulin stained structure defining the centrosome or microtubule organising center. Microtubules radiated from this structure into the cytoplasm in a regular pattern (Figure 4.47).



**Figure 4.47: Wild-type primary fibroblast cells.** Immunofluorescence staining and confocal microscopy images of wild-type primary fibroblasts with a single centrosome and well organised microtubules. Antibodies were directed against  $\gamma$ -tubulin (white), CEP135 (red),  $\alpha$ -tubulin (green) and DAPI (blue) for DNA staining. Scale bar, 5  $\mu$ m.

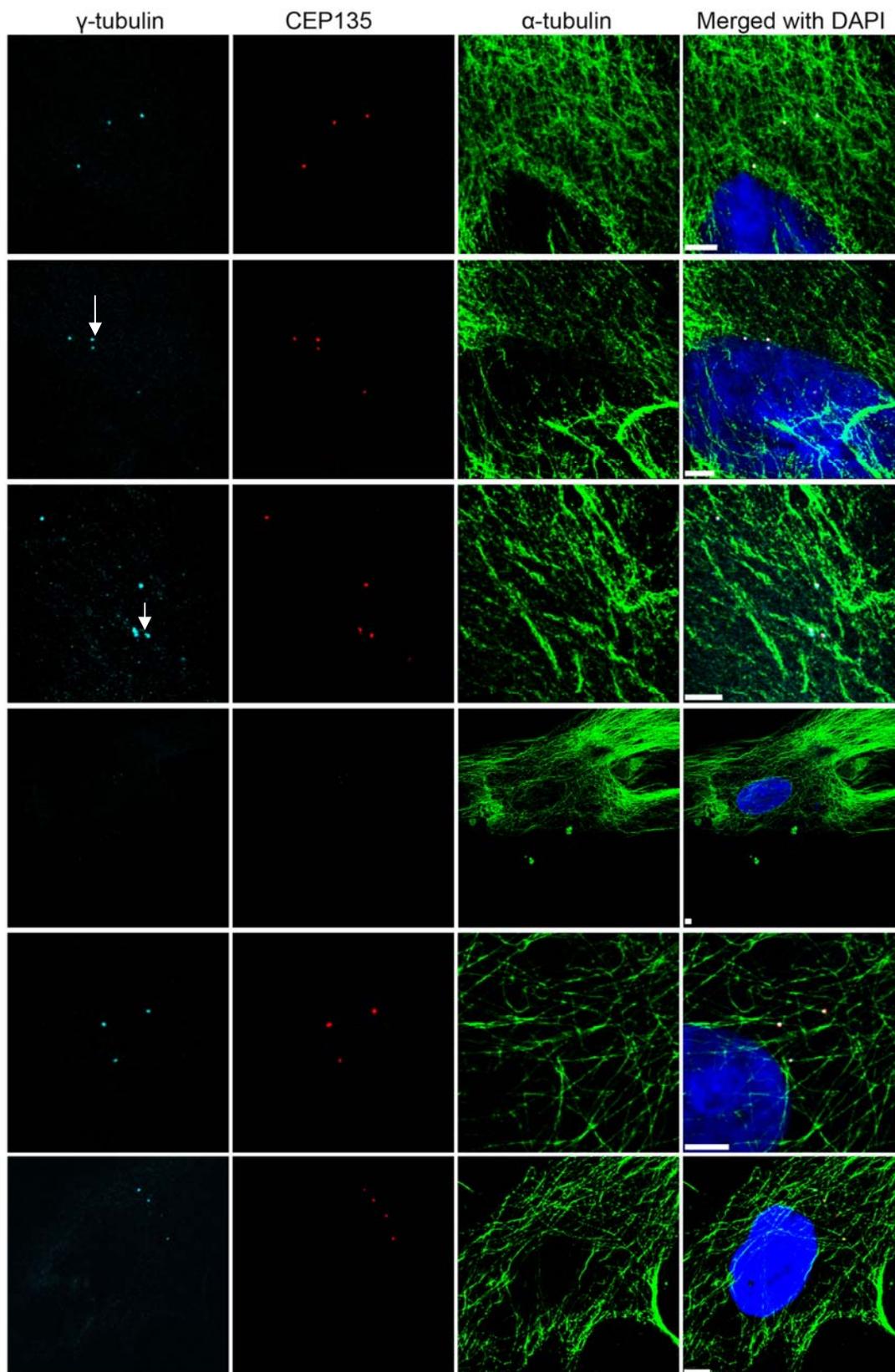
#### 4.9.7. CEP135 patient primary fibroblasts

When I analysed CEP135 patient primary fibroblasts I noted various alterations with regard to the centrosome, microtubules, cell division and nuclear shape. Most

remarkably, the CEP135 antibody still recognized the  $\gamma$ -tubulin stained centrosome although I had assumed that the mutation would lead to a premature stop in the protein.

#### **4.9.7.1. Centrosome number abnormalities**

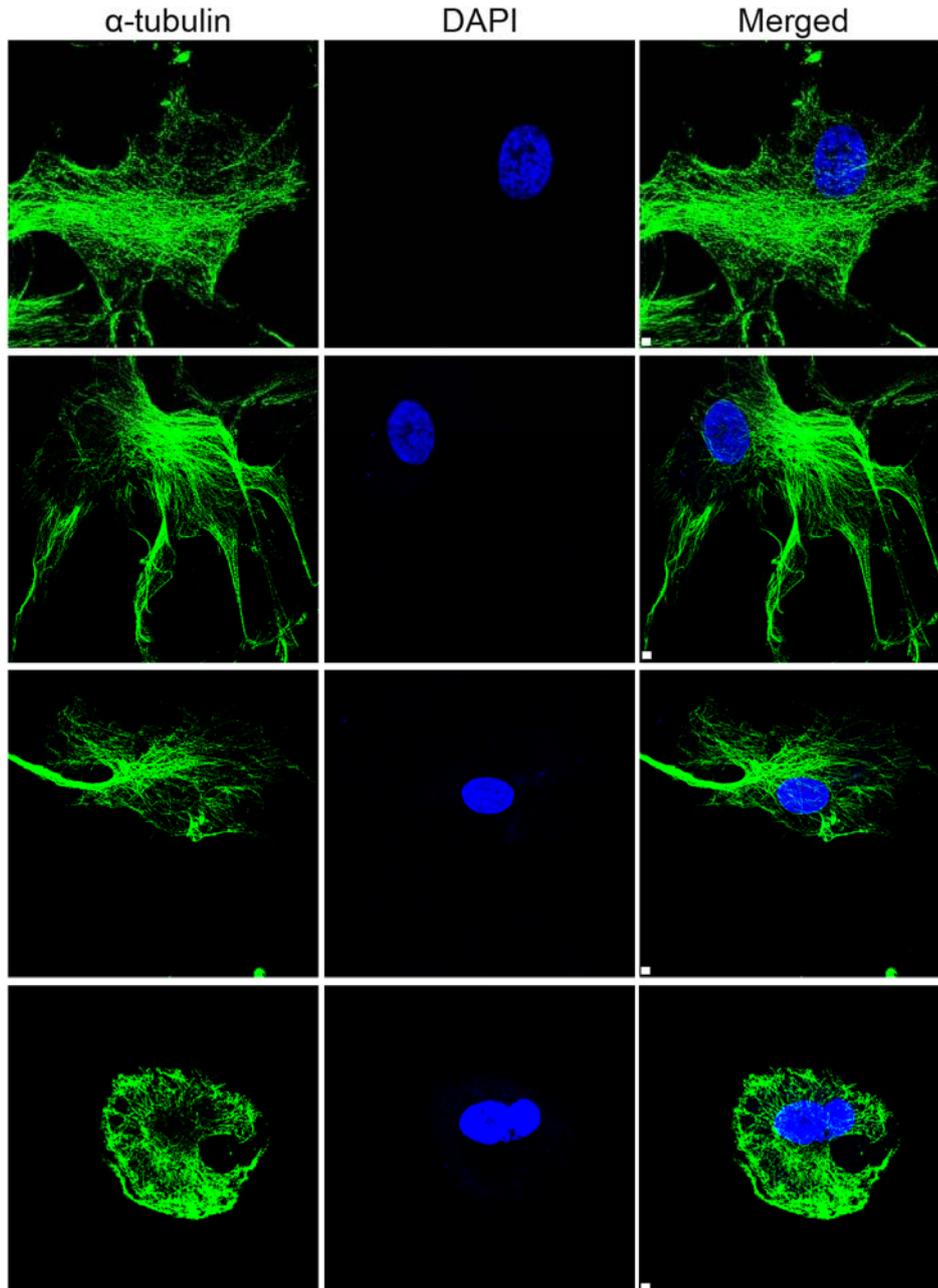
The cells harboured 3, 4 or 5 centrosomes with a frequency of 18.67 % (Table 5.4). The centrosomes also appeared fragmented. The abnormal centrosome number strengthens the role of CEP135 in centriole biogenesis. Interestingly the cells with multiple centrosomes also show the disorganised microtubules (Figure 4.48).



**Figure 4.48: CEP135 mutant primary fibroblasts with centrosomal abnormalities.** Centrosomes are variable in number (3, 4 or 5) and fragmented. The arrowheads show the multiple and fragmented centrosomes. Microtubules of these cells also have an abnormal appearance. Scale bar, 5  $\mu$ m.

#### 4.9.7.2. Absence of centrosomes

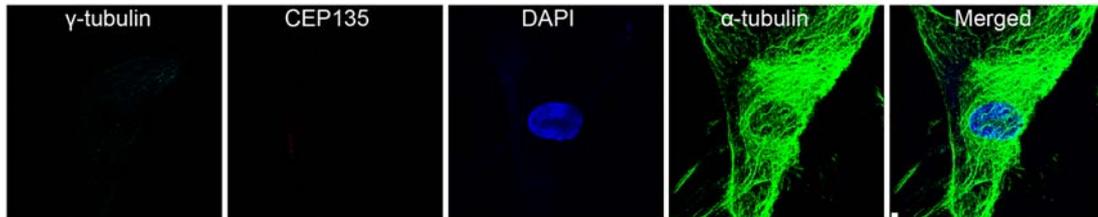
Another prominent feature of CEP135 patient primary fibroblasts was the complete loss of centrosomes (Figure 4.49). About 22 % of mutant primary fibroblasts completely lacked centrosomes whereas this was never observed in WT (Table 5.4). Interestingly, the majority of the cells with misshapen nuclei were also deficient of centrosomes (Figure 4.51).



**Figure 4.49: Absence of centrosomes in CEP135 mutant primary fibroblasts.** No signal for CEP135 and gamma tubulin was observed in ~22 % of cells. These cells also show disorganised microtubules paralleled by an abnormal cell shape. Scale bar, 5  $\mu$ m.

#### 4.9.7.3. Disorganised microtubules

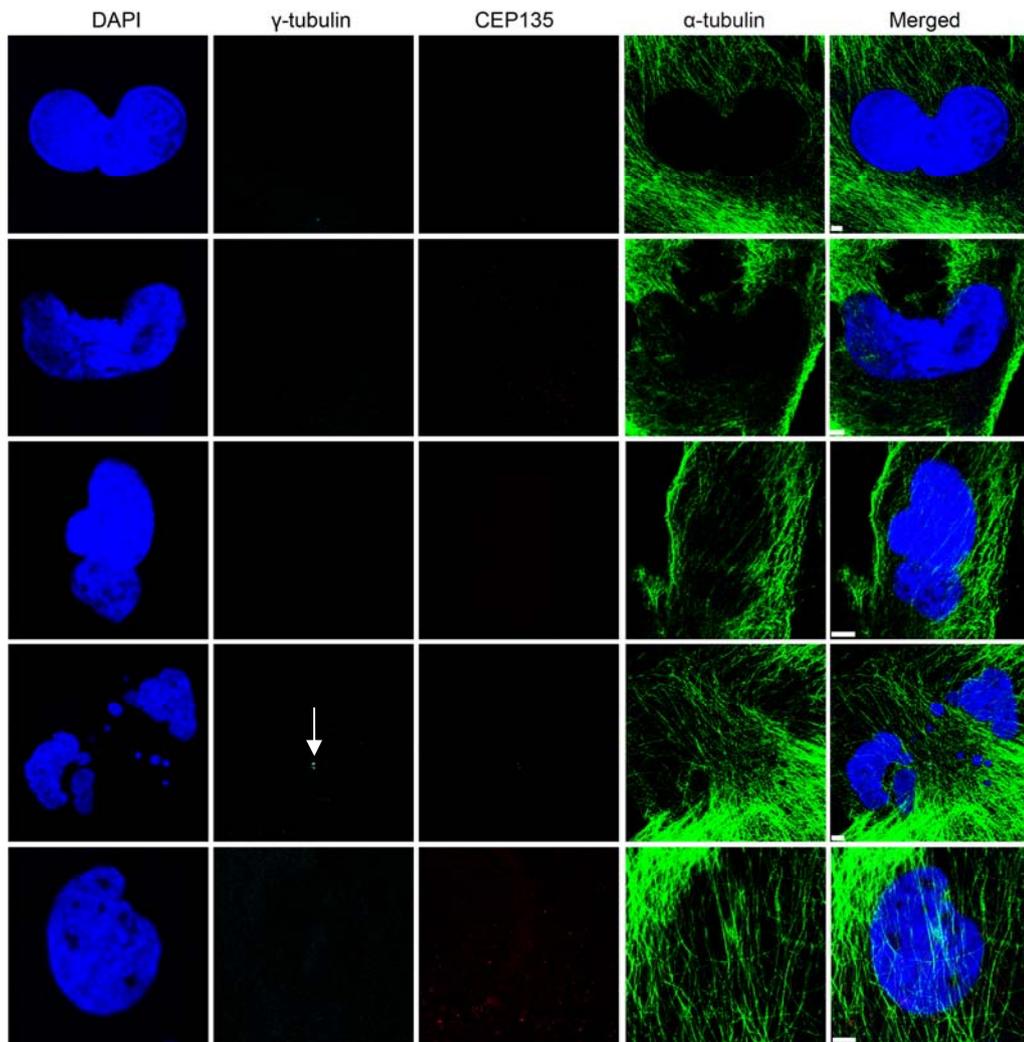
A further feature was a disorganised microtubule system which was accompanied by cell shape changes and was observed in 55.33 % of the cells. (Figure 4.48, 4.49, 4.50, Table 5.4). Disorganisation of the microtubule network points out the importance of the centrosome as the microtubule organising center.



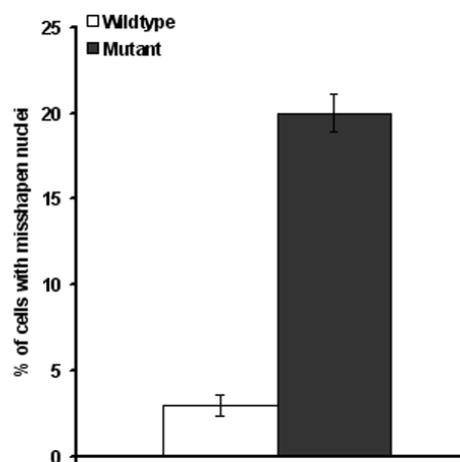
**Figure 4.50: CEP135 mutant primary fibroblasts with disorganised microtubules.** Confocal microscopy images of CEP135 mutant primary fibroblasts showing a disorganised microtubule network. Cells also lack centrosomes resulting in the absence of  $\gamma$ -tubulin and CEP135 staining. Antibodies were directed against  $\gamma$ -tubulin (white), CEP135 (red),  $\alpha$ -tubulin (green) and DAPI (blue) for DNA staining. Scale bar, 5  $\mu$ m.

#### 4.9.7.4. Misshapen nuclei

Mutant cells had an increased number of misshapen and fragmented nuclei (Figure 4.51). Often nuclei were highly lobulated and fragmentation led to DAPI stained material in the cytoplasm. A statistical analysis showed that these changes were significant. About 20 % of mutant fibroblasts harboured misshapen nuclei whereas in control fibroblasts this number was only ~3% (Figure 4.52).



**Figure 4.51: CEP135 mutant primary fibroblasts with misshapen nuclei.** Dysmorphic nuclei often appeared like broken ones. Only few cells with dysmorphic nuclei features had centrosomes (arrowhead), while others did not have centrosomes and consequently no CEP135 staining. Another obvious feature of these cells was the occurrence of disorganised microtubules. Scale bar, 5  $\mu$ m.



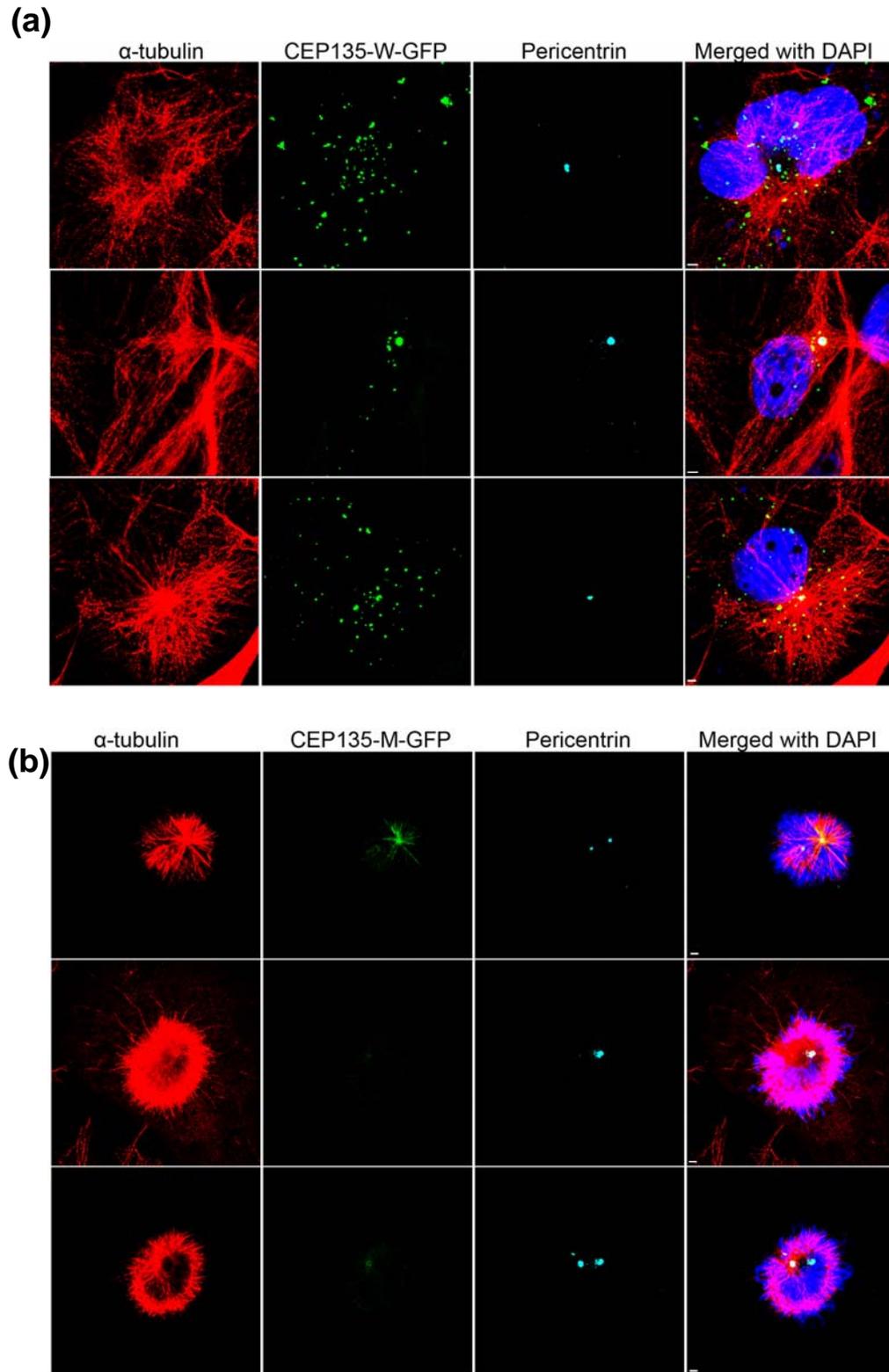
**Figure 4.52: Graphical representation showing the number of CEP135 mutant primary fibroblasts with misshapen nuclei.** 300 cells were counted for wild type and mutant. Error bars represent s.e.m. *P* value, 3.44E-3 (Student's *t*-test).

#### **4.9.8. Overexpression of CEP135**

COS7 cells were transfected with GFP-tagged wild-type and mutant CEP135. The wild-type exogenous CEP135 localized to the centrosome and resulted in the formation of spots outside of the centrosome. These spots were variable in number and size and were also recognized by polyclonal antibodies directed against CEP135 (Figure 4.55). But staining with pericentrin indicates only one spot. So the extra spots are not centrosomes but only exogenous CEP135 (Figure 4.53a). Furthermore I noted unusual phenotypes in the COS7 cells expressing GFP-CEP135. I have observed very weak mutant exogenous CEP135 expression (Figure 4.53b).

##### **4.9.8.1. Overexpression of CEP135 leads to a disorganised microtubule system.**

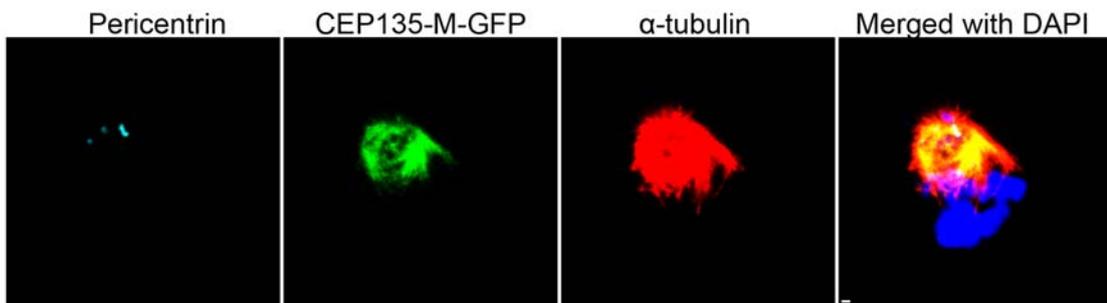
GFP-tagged wild-type and mutant CEP135 overexpression led to abnormal microtubule systems. In wild type, the immunofluorescence dots were seen but in mutant cells the weak expression of GFP was observed. I have noticed only one spot, stained for pericentrin (Figure 4.53).



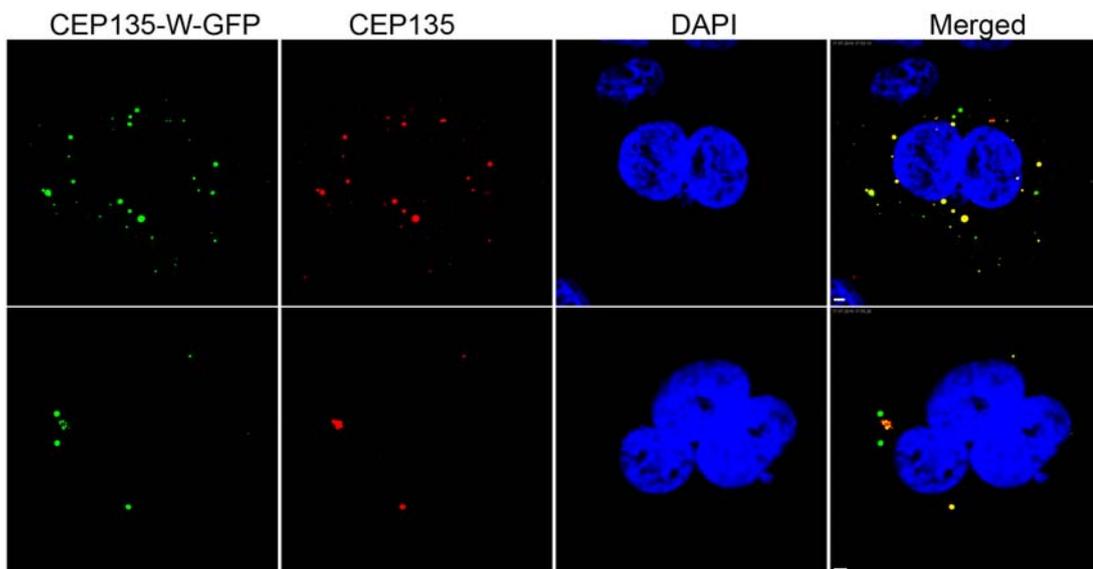
**Figure 4.53: Abnormal microtubules in CEP135 transfected cells.** Confocal microscopy images of CEP135 wild type **(a)** and mutant **(b)** transfected COS7 cells showing the phenotype of disorganised microtubules. The wild-type cells also show several dots with variable size and number. The cells were stained with CEP135-GFP (green),  $\alpha$ -tubulin (red) Pericentrin (turquoise) and DAPI (blue) for DNA staining. Scale bar, 10  $\mu$ m.

#### 4.9.8.2. Abnormal number of centrosomes

GFP-tagged CEP135 mutant transfected cells also demonstrated multiple centrosomes (3 centrosomes observed). This feature was only observed in mutant transfected cells and was also observed in CEP135 mutant primary fibroblast cells (Figure 4.54).



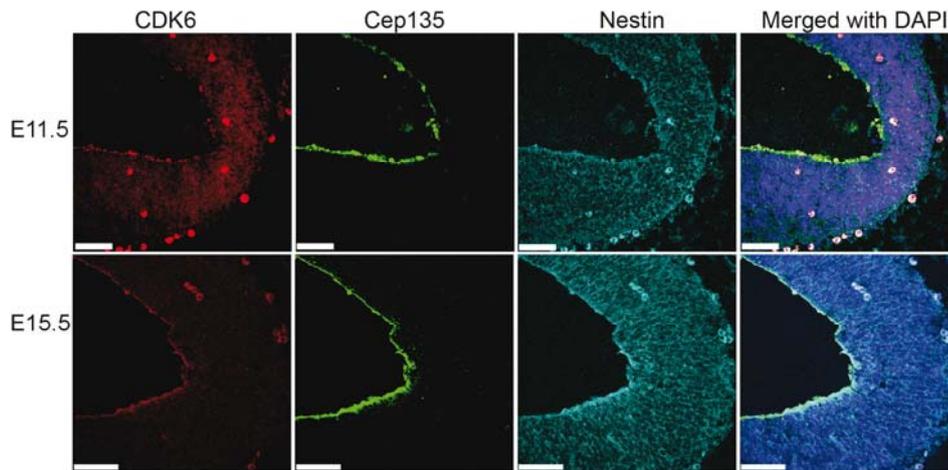
**Figure 4.54: Abnormal number of centrosomes in CEP135 mutant cells.** COS7 cells were transfected with GFP-tagged mutant CEP135 (CEP135-M-GFP, green) and stained for  $\alpha$ -tubulin (red) Pericentrin (turquoise) and DAPI (blue) for DNA staining. Scale bar, 10  $\mu$ m.



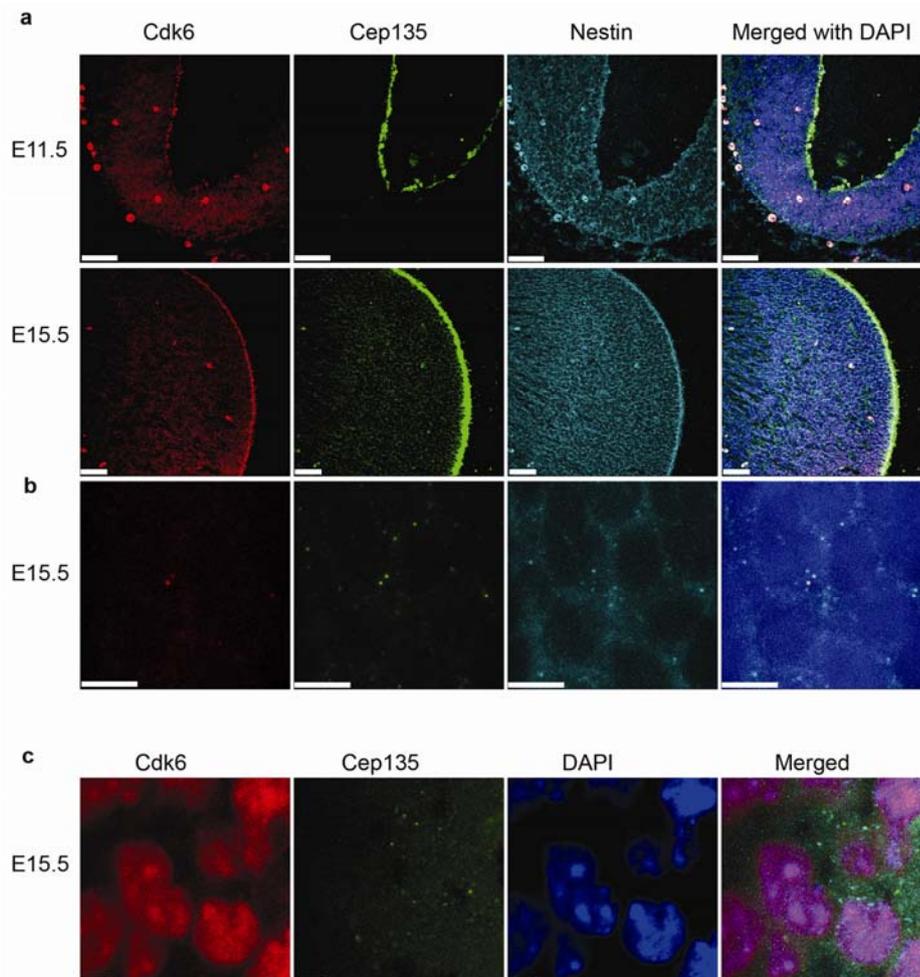
**Figure 4.55: COS7 cells with CEP135 immunofluorescence dots.** These dots were variable in size and number. The cells were stained with CEP135-GFP (green), CEP135 (red) and DAPI (blue) for DNA staining. Scale bar, 10  $\mu$ m.

#### 4.9.9. CDK6 and CEP135 expression in mouse neuroepithelium

It is assumed that in MCPH neurogenesis is afflicted during development. Therefore I examined the expression profile of CDK6 and CEP135 during mammalian cerebral cortical neurogenesis. Immunofluorescence studies were performed using mouse sections obtained at embryonal stages E11.5 and E15.5. At both stages I observed a strong staining for Cep135 and Cdk6 in the developing neuroepithelium of the cerebral cortex (Figure 4.56).



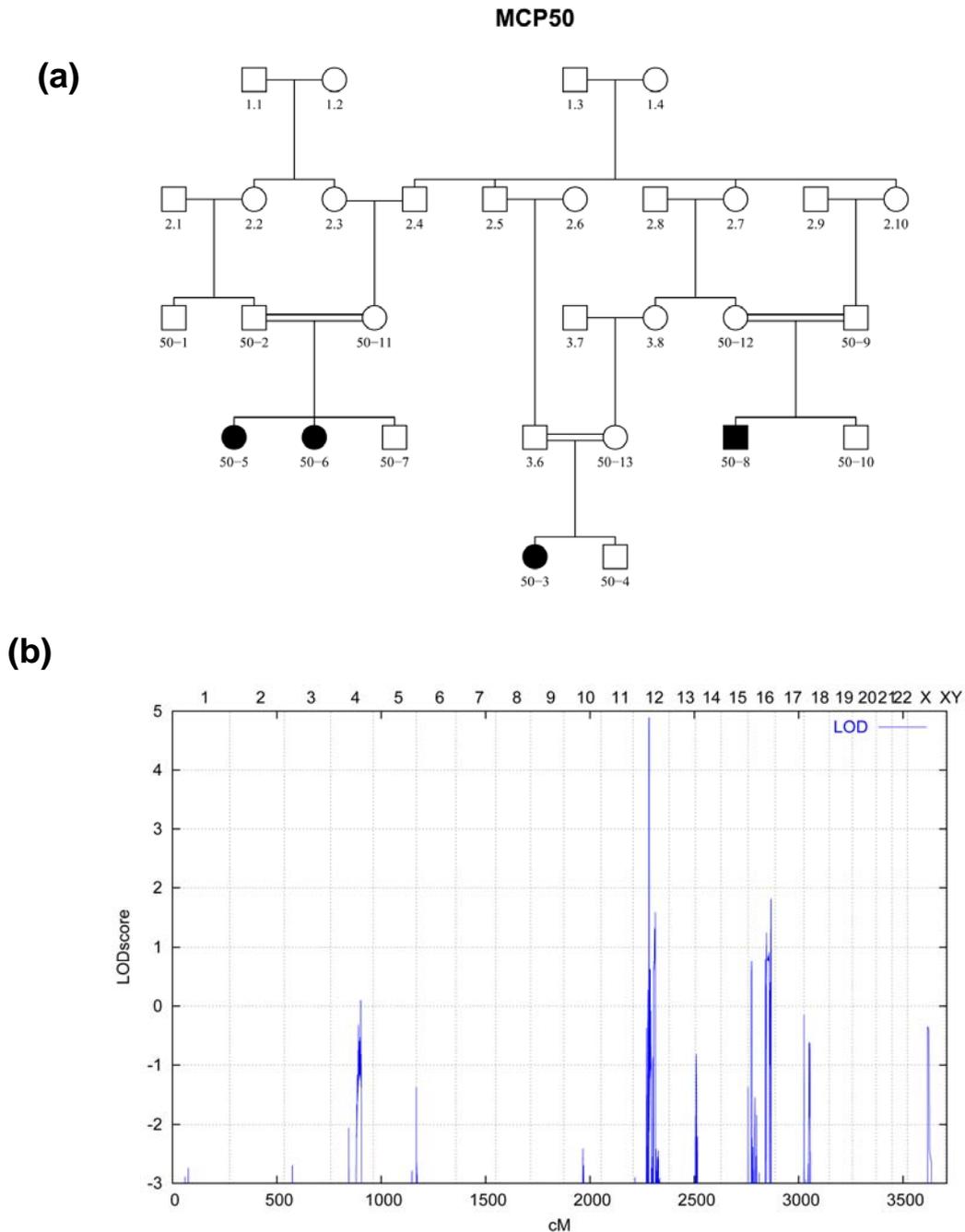
**Figure 4.56:** Confocal microscopy images of endogenous Cdk6 and Cep135 in mouse embryonic brain neuroepithelium. Embryos were taken from E11.5 and E15.5. Antibodies were used against Cdk6 (red), Cep135 (green) and Nestin (turquoise). Scale bar, 50  $\mu\text{m}$ .



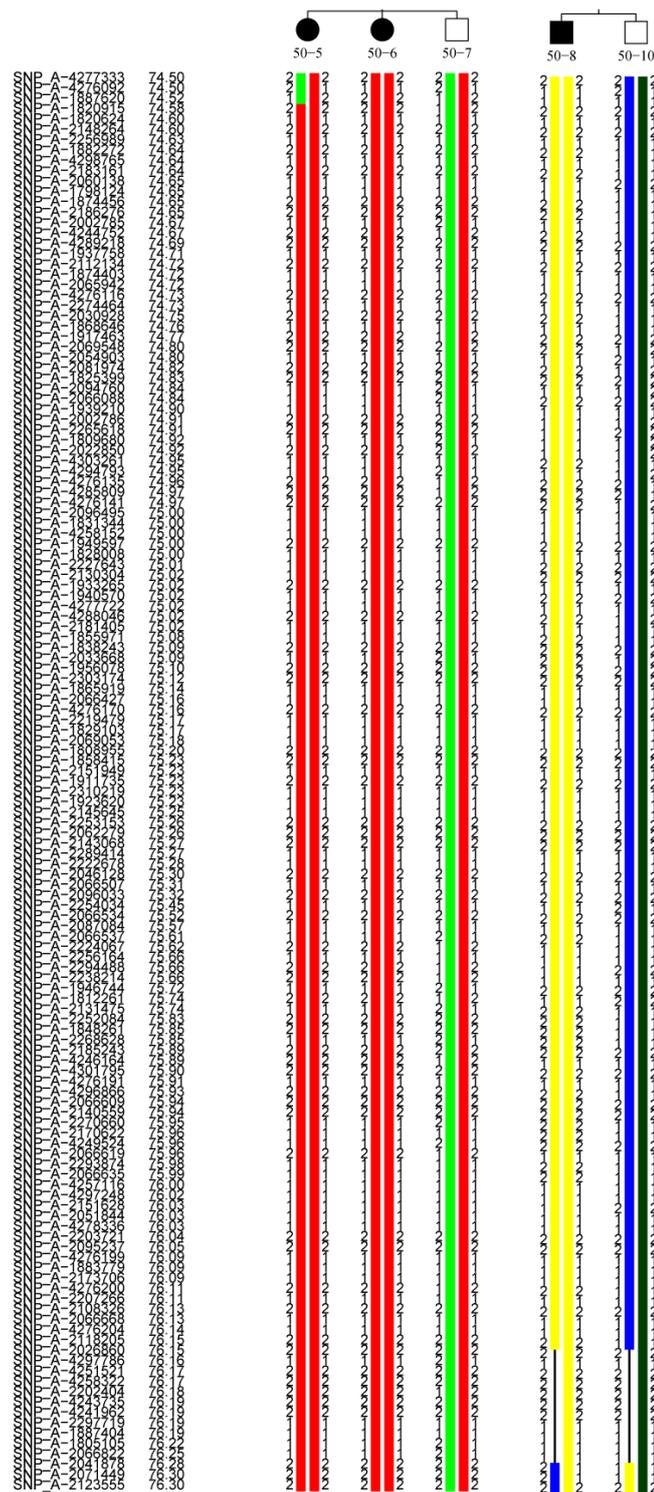
**Figure 4.56: Cdk6 and Cep135 expression in the neuroepithelium of the developing mouse brain.** (a) Expression at stage E11.5 and 15.5, Cdk6 (red), Cep135 (green), Nestin (turquoise). Scale bar, 50  $\mu\text{m}$ . (b) Localization of Cdk6 to the centrosome in the mouse neuroepithelium during E15.5. Cdk6 staining (red) is merged with Cep135 staining (green) which is a bona fide centrosomal protein. Scale bar, 5  $\mu\text{m}$ . (c) Nuclear Cdk6 localization in cells of the mouse neuroepithelium at E15.5. Scale bar, 5  $\mu\text{m}$ .

#### 4.10. Excluded family MCP50

MCP50 is a large consanguineous family with three different loops. All afflicted individuals have consanguineous parents (Figure 4.57a).



**Figure 4.57: Pedigree of family MCP50 and Genome-wide graphical view of the MCPH10 locus. (a)** A family tree showing the pedigree of family MCP50 afflicted with primary microcephaly. **(b)** Parametric linkage analysis (Allegro software) showing the MCPH10 locus mapped in the MCP50 family. The highest LOD score (4.9) was obtained on chromosome 12q14.1-12q14.2.



**Figure 4.58: Haplotype analysis of the MCP50 family linked to chromosome 12q14.1-12q14.2.** Black symbols represent afflicted individuals. The critical interval was defined by the recombinations observed for markers SNP\_A-1887620 (rs17629041, physical position 59,900,744bp and 74.52cM) to SNP\_A-2041878 (rs1463821, physical position 61,689,142bp and 76.28cM).

PCR based genotyping of microsatellite markers resulted in the exclusion of linkage of this family with all known MCPH loci. So this family was subjected for genome-wide mapping using the 250K SNP Array (Affymetrix) and obtained a single maximum (LOD)

score of 4.9 for a region between rs17629041 (physical position 59,900,744bp, 74.52cM) and rs1463821 (physical position 61,689,142bp, 76.28cM) on chromosome 12q14.1-12q14.2 (NCBI Human Genome Build 36.3) (Figure 4.57b, Figure 4.58).

Bioinformatic analysis of this region revealed 13 genes (NCBI Human Genome Build 36.3). Among these, Family with sequence similarity 19 (chemokine (C-C motif)-like), member A2 (*FAM19A2*), Ubiquitin specific peptidase 15 (*USP15*), MON2 homolog (*S. cerevisiae*) (*MON2*) and Protein phosphatase 1H (PP2C domain containing) (*PPM1H*) were selected as the strongest candidates and sequenced within MCP50 individuals. The MCP50 family was negative for any pathogenic change in these four selected genes.

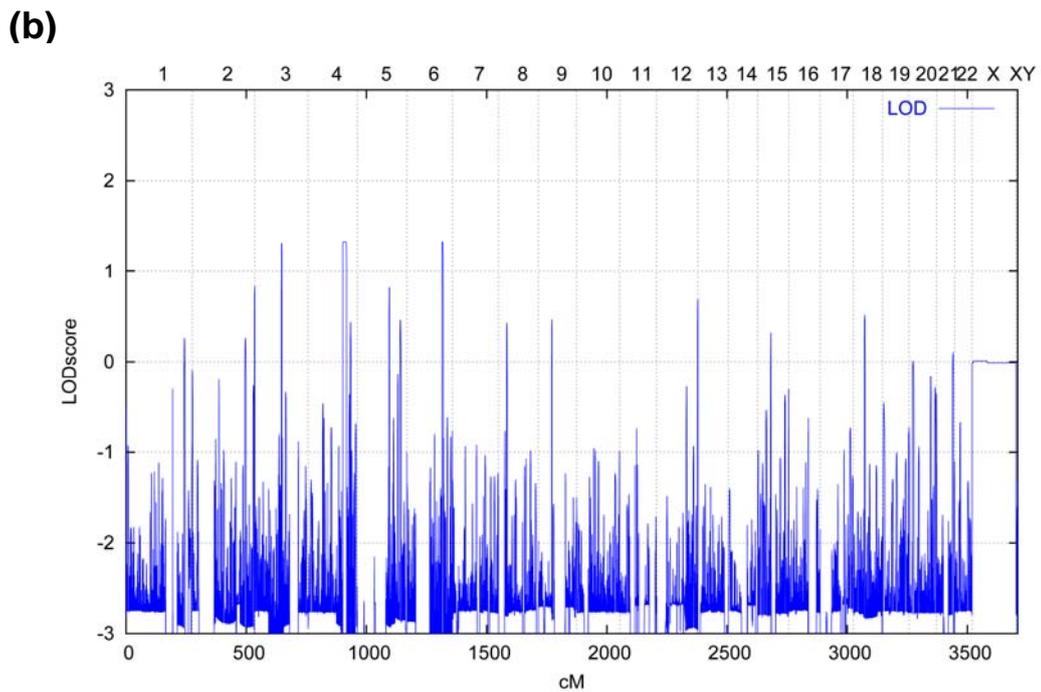
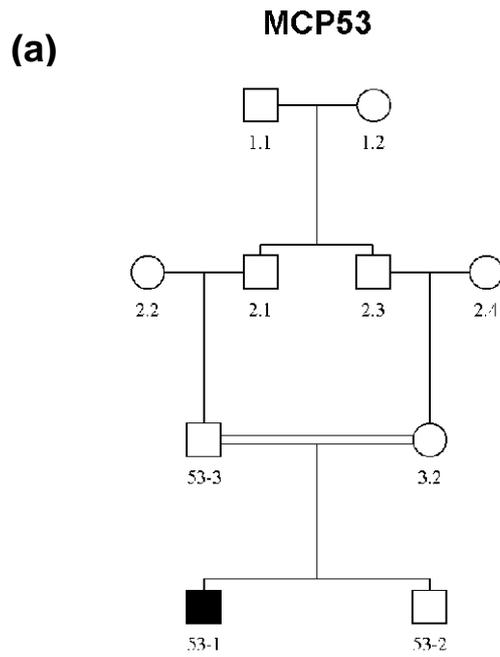
#### **4.11. Excluded family MCP53**

MCP53 is small consanguineous family with one afflicted individual (Figure 4.59a). After exclusion of linkage of the MCP53 family with microsatellite markers of all known MCPH loci, genome-wide linkage analysis was performed which demonstrated three different peaks on chromosomes 3, 4 and 6 with the maximum possible LOD score of 1.33 (Figure 4.59b).

On chromosome 3, the minimum linkage interval is defined by the SNP\_A-1889563 (rs832042 physical position: 99,227,425bp, 111.01cM) and SNP\_A-2101206 (rs17353809 physical position: 103,498,894bp, 113.22cM). This region of 4.27 Mb on chromosome 3q12.1-3q12.3 harbours 60 genes (NCBI Human Genome Build 36.3).

A wider peak which is observed on chromosome 4 is defined from SNP\_A-1911424 (rs13101536 physical position 150,514,489bp, 144.72cM) to SNP\_A-2288389 (rs4690796 physical position 166,092,275bp, 159.87cM). There are 92 genes which reside in this 15.57 Mb region on chromosome 4q31.23-4q32.3.

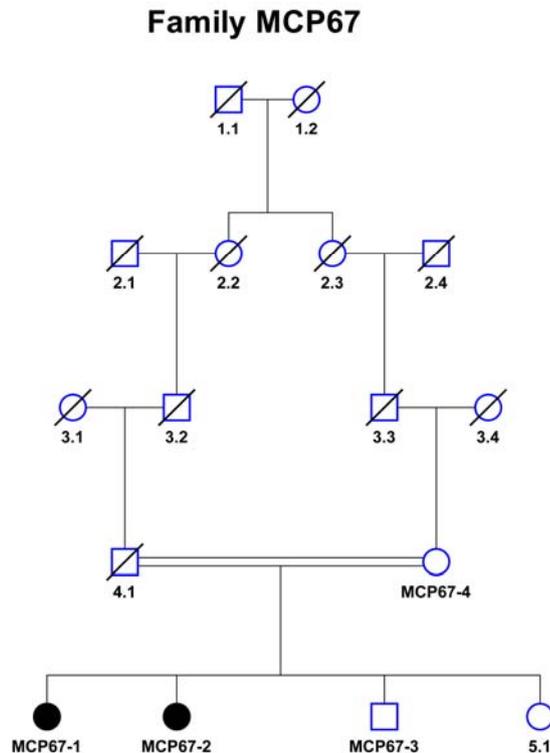
The region of the third peak on chromosome 6 (6q24.2-6q24.3) is between SNP\_A-2271440 (rs4896697 physical position 144,467,714bp, 146.69cM) and SNP\_A-1782436 (rs6570879 physical position 149,035,821bp 151.26cM) and comprises 20 genes.



**Figure 4.59: Pedigree and Genome-wide graphical view of the MCPH11 locus. (a)** Pedigree of family MCP53 having one afflicted individual. **(b)** Parametric linkage analysis (Allegro software) showing the MCPH11 locus mapped in MCP53 family. The highest three peaks can be seen on chromosomes 3, 4 and 6.

#### 4.12. Excluded family MCP67

MCP63 is a small consanguineous family with three afflicted individuals (Figure 4.60). PCR based genotyping with microsatellite markers, specific for each MCPH loci, excluded this family from linkage to all known MCPH loci.

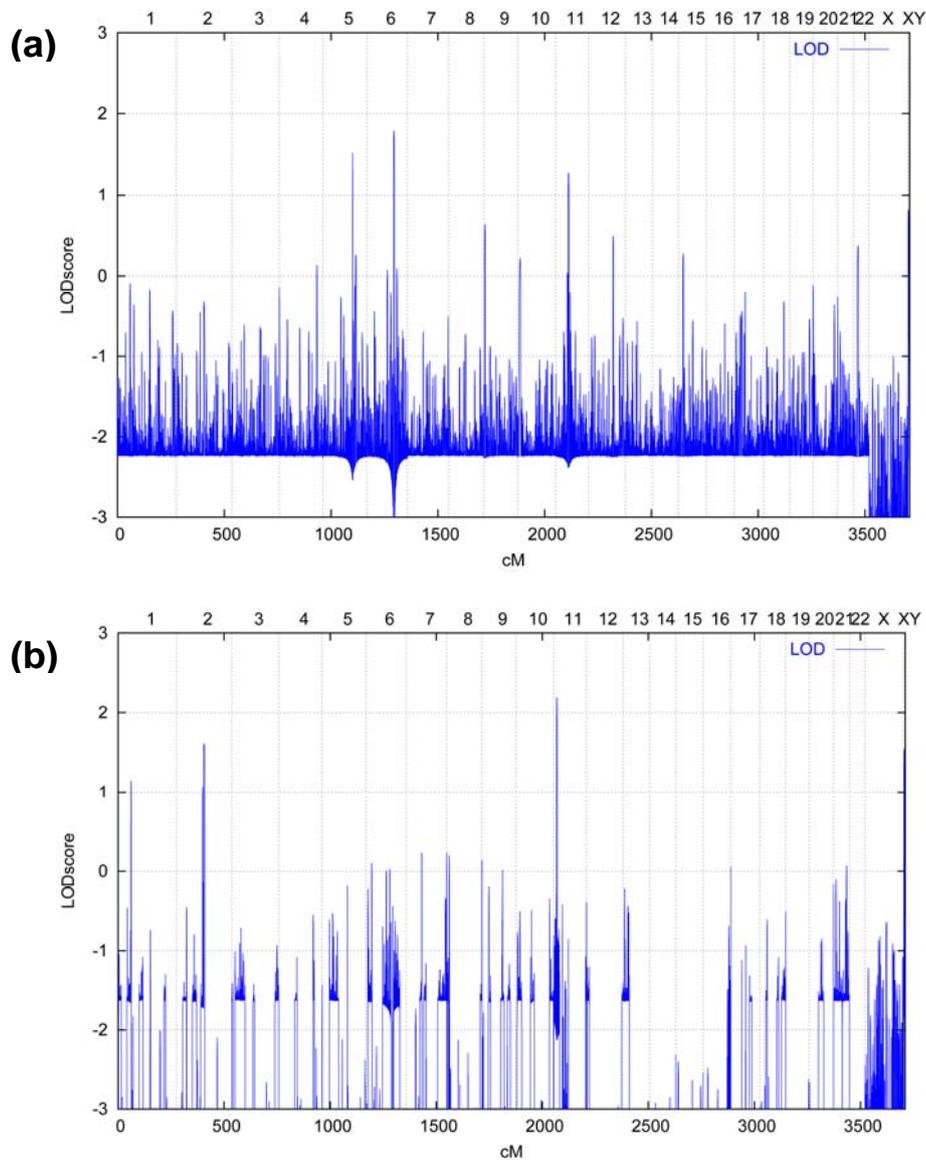


**Figure 4.60: Family tree of MCP67 with two afflicted individuals.** Only four individuals were screened. The father of these patients is deceased.

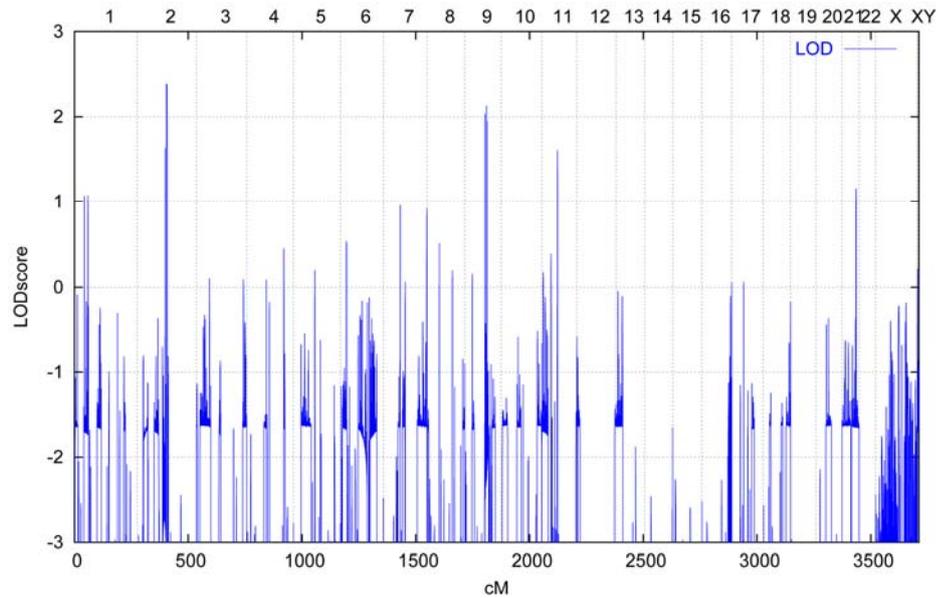
Genome-wide linkage analysis was performed on the MCP67 family with a reduced marker panel of 20K, which shows a peak on chromosome 6 which maps between SNP\_A-4282597 (rs12206695 position 126,297,240bp 126.69cM) and SNP\_A-2181427 (rs9388518 position 127,145,478bp 127.27cM) along with 6q22.32 (Figure 4.60a, Figure 4.62a). Eight genes were found in this small 0.84 Mb (0.58cM) linkage interval (NCBI Human Genome Build 36.3).

Analysis with an increased marker panel of 32K and 50K was conducted, which resulted in a peak on chromosome 11 with a LOD score of approximately 2.2 and a region on chromosome 2 with a LOD score of approximately 2.4 (Figure 4.60b, Figure 4.61). These three regions on chromosomes 2, 6 and 11 were further confirmed by linkage analysis with the Affymetrix GeneChip(R) Human Mapping 250K SNP Array (version 2.0). The region on chromosome 11p15.4 mapped between SNP\_A-4286081 (rs4929918 position 8,795,236bp 15:14cM) and SNP\_A-2048279 (rs7931872 position 9,098,638bp and 15.98cM) (Figure 4.62b). The region on chromosome 2q14.2 mapped between SNP\_A-

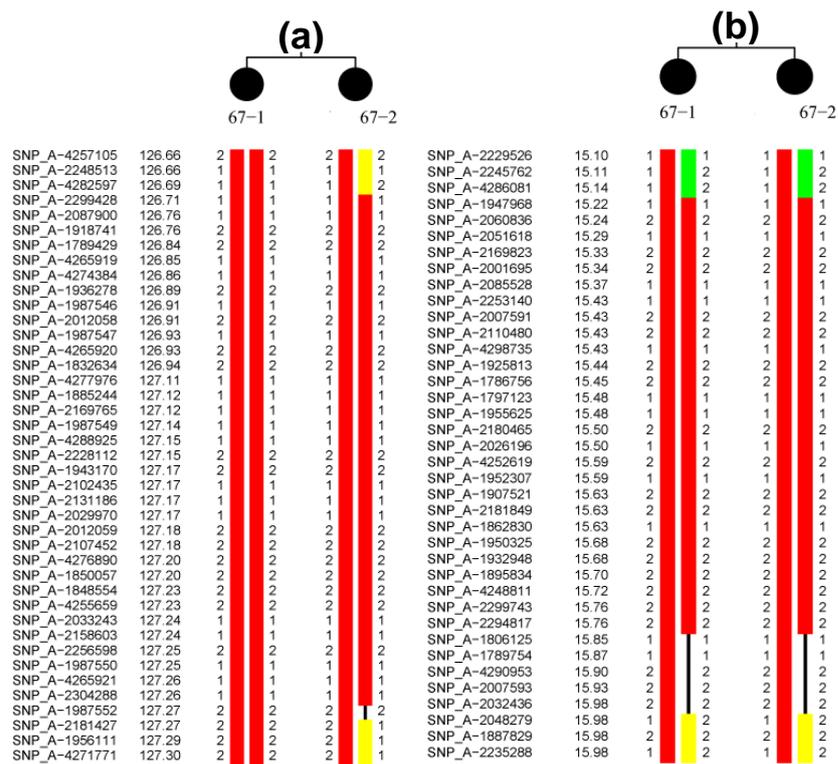
2164389 (rs6541729 position 121,096,831bp and 130.91cM) and SNP\_A-4253620 (rs12616919 position 121,627,677bp and 132.91cM) (Figure 4.63). Eight genes were found in the 0.30 Mb (0.84 cM) region on chromosome 11p15.4 and one gene was found on chromosome 2q14.2 in the 0.53 Mb (2cM) region (NCBI Human Genome Build 36.3).



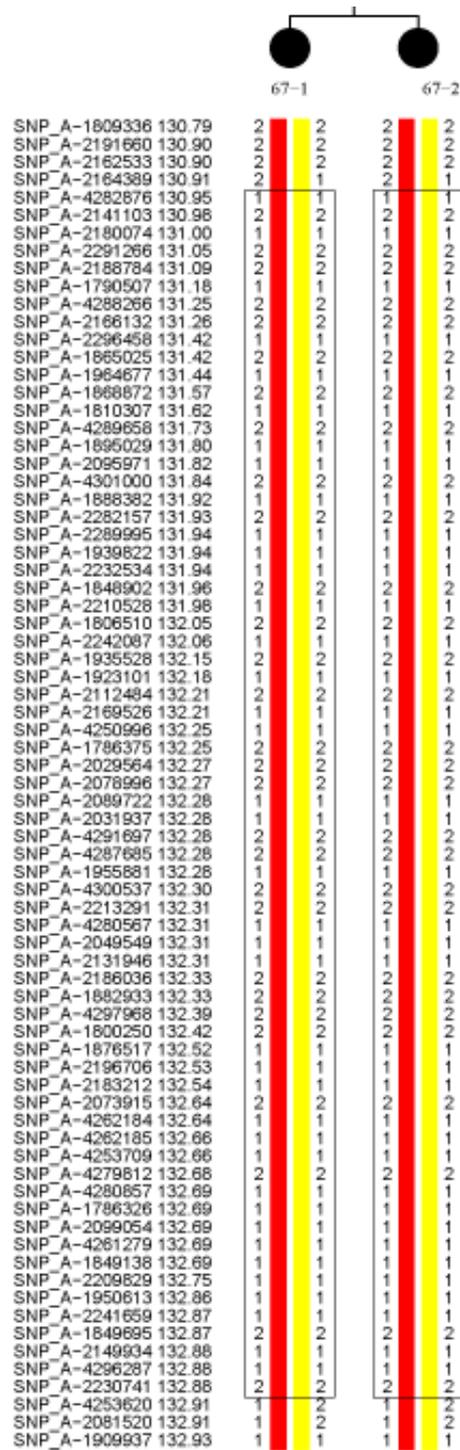
**Figure 4.60: Genome-wide graphical view of the MCPH12 locus.** (a) Parametric linkage analysis (Allegro software) with a marker panel of 20K, showing the MCPH12 locus on chromosome 6q22.32 which was mapped in the MCP67 family. (b) Parametric linkage analysis (Allegro software) with a marker panel of 32K, showing the MCPH12 locus mapped on chromosome 11p15.4 in the MCP67 family.



**Figure 4.61: Genome-wide graphical view of the MCPH12 locus.** Parametric linkage analysis (Allegro software) with a marker panel of 50K, showing the MCPH12 locus mapped on chromosome 2q14.2 in the MCP67 family.



**Figure 4.62: Haplotype analysis of MCP67 family.** (a) On left the haplotype on chromosome 6q22.32 is described which maps between SNP\_A-4282597 (rs12206695 position 126,297,240bp 126.69cM) and SNP\_A-2181427 (rs9388518 position 127,145,478bp 127.27cM) with a small linkage interval of 0.84 Mb (0.58 cM). (b) On the right the haplotype on chromosome 11p15.4. is depicted. The critical interval (0.30 Mb, 0.84 cM) is mapped between SNP\_A-4286081 (rs4929918 position 8,795,236bp 15:14cM) and SNP\_A-2048279 (rs7931872 position 9,098,638bp and 15.98cM).



**Figure 4.63: Haplotype analysis of the MCP67 family linked to chromosome 2q14.2.** The linkage interval (0.53 Mb, 2 cM) is mapped between SNP\_A-2164389 (rs6541729 position 121,096,831bp and 130.91cM) and SNP\_A-4253620 (rs12616919 position 121,627,677bp and 132.91cM).

## 5. Discussion

Autosomal recessive primary microcephaly (MCPH) is a rare congenital, genetically and clinically heterogeneous disease. Patients with MCPH typically exhibit reduced head size and mental retardation, but usually there are no further neurological findings or malformations. Also, the macroscopic organization of the brain is preserved. MCPH therefore is considered a consequence of a reduced brain volume, which is evident particularly within the cerebral cortex and thus results to a large part from a reduction of grey matter (Kaindl et al., 2010). Although the MCPH brain is small, it exhibits the normal six-layer neuronal architecture, suggesting that MCPH results from a decrease in neural progenitor cell number (Bond and Woods, 2006).

There has been a dramatic expansion in brain size during mammalian evolution with a thousand fold increase in cortical surface area between mouse and man (Rakic, 1995). Morphometric studies have revealed a scaling of the brain relative to body size during mammalian evolution with disproportionate enlargement of the cerebral cortex (Northcutt and Kaas, 1995), (Clark et al., 2001). In primates, this might have been driven by increased social rather than environmental complexity and their ability to execute complex cognitive functions such as speech (Finlay and Darlington, 1995). Expansion progressed through an increase in surface area rather than an increase in cortical thickness (Rakic, 1995), (Northcutt and Kaas, 1995). This has been accommodated topologically by folding of the cerebral cortex to form convolutions, so that there is a transition from the smooth cortical surface seen in rodents to one with multiple folds in humans (Ponting and Jackson, 2005).

The genes responsible for MCPH are specific regulators of brain size and have evolved under strong positive selection in the lineage leading to *Homo sapiens*. One genetic variant of *Microcephalin* in modern humans, which arose ~37,000 years ago, increased in frequency too rapidly to be compatible with neutral drift. This indicates that it has spread under strong positive selection, although the exact nature of the selection is unknown. The finding that an important brain gene has continued to evolve adaptively in anatomically modern humans suggests the ongoing evolutionary plasticity of the human brain. It also makes *Microcephalin* an attractive candidate locus for studying the genetics of human variation in brain-related phenotypes (Evans et al., 2005). One genetic variant of *ASPM* in humans arose merely about 5800 years ago and has since swept to high frequency under strong positive selection. These findings, especially the remarkably young age of the positively selected variant, suggest that the human brain is still undergoing rapid adaptive evolution (Mekel-Bobrov et al., 2005).

MCPH is a primary disorder of neurogenic mitosis and not due to neural migration, neural apoptosis, or neural function. This is the case because all known MCPH genes are expressed in the neuroepithelium, and phenotypic features and brain scans suggest that MCPH patients have a small brain that functions normally for its size (Woods et al., 2005). The functions of MCPH genes during neurogenic mitosis are unknown except for *ASPM*, so elucidation of the genes responsible for MCPH and understanding the pathomechanisms leading to MCPH is of high importance not only for our understanding of physiologic brain development (particularly of cortex formation), but also for both genetic counseling and prenatal diagnosis. So I aimed to identify novel genes responsible for causing the phenotype of MCPH and to explore their role during cell division.

### **5.1. Mutational spectra of known MCPH genes**

Molecular genetic analysis of thirty primary microcephaly families ascertained from different regions of Pakistan resulted in the linkage of 19 (63.3 %) families to the MCPH5 locus, one (3.33 %) was found mutated at the *MCPH1* gene, one (3.33 %) in *CENPJ* (MCPH6), two (6.67 %) in *WDR62* (MCPH2), two (6.67 %) were linked to the MCPH4 locus and five were excluded from any linkage with MCPH loci which accounts for 16.7 % of all families (Figure 4.2). Our data is in line with previous results which reported the MCPH5 (*ASPM*) mutations as the major common cause of MCPH in the Pakistani and other populations of the world (Bond et al., 2002, Roberts et al., 2002, Gul et al., 2006b, Kumar et al., 2004, Nicholas et al., 2009, Muhammad et al., 2009, Kousar et al., 2010). Recently a study carried out in 112 Iranian MCP families detected homozygosity of 8 families at MCPH1, 13 at MCPH5, 3 at MCPH2, five at MCPH6 and two at MCPH7, while the remaining 81 families were not linked to any of the seven known loci (Darvish et al., 2010). The *ASPM* linked families in this study account for 13.3 % which is considerably lower prevalence than within the Pakistani's population, where frequencies between 43 % and 55 % were reported (Gul et al., 2006b, Roberts et al., 2002). In the Indian population however, 33 % *ASPM* linked families were reported (Kumar et al., 2004). As already reported in the study carried out in Iranian population (Darvish et al., 2010), there is a high prevalence of *ASPM* linked families in the Pakistani population as compared to other parts of the world.

To date, 95 different mutations have been reported in *ASPM* (Bond et al., 2002, Bond et al., 2003, Kumar et al., 2004, Shen et al., 2005, Gul et al., 2006b, Gul et al., 2007, Nicholas et al., 2009, Muhammad et al., 2009, Kousar et al., 2010, Passemard et al., 2009, Darvish et al., 2010, Halsall et al., 2010). This study has increased the number to 99, by identifying eight further novel mutations in *ASPM* (Table 4.2). Among these eight novel mutations are, four (c.1002delA, c.9492T>G, c.3055C>T and c.8668C>T) that I

have already contributed to an earlier publication (Muhammad et al., 2009). One of our novel mutations (c.3977G>A) is interesting in the sense that it introduced a change of nucleotide G into A at position 3977, but an already reported variant (c.3978G>A) constitutes a change at nucleotide position 3978. Both variants result in the introduction of a stop codon (p.W1326\*). In family MCP55 I have found one missense variant c.5584A>G which is not reported as a SNP in the database. This variant introduces glutamic acid instead of lysine (p.K1862E) in exon 18 of *ASPM*. On completion of the *ASPM* screening in the MCP55 family, I observed a further already reported nonsense mutation (p.Y3164\*) in exon 23 of *ASPM*. The mutation c.9492T>G (p.Tyr3164\*) is in cis to the missense mutation c.5584A>G (p.Lys1862Glu). The latter had previously been found in one MCPH family in homozygous state and deemed to be pathogenic (Darvish et al., 2010). This suggests that the missense variant (p.K1862E) is infact a rare variant that is not associated with the phenotype of primary microcephaly. Except for one missense mutation (c.9539A>C, p.Q3180P) which still has to be tested functionally, all of the other mutations identified so far in *ASPM* result in premature stop codons leading to truncated protein products due to resistance of the mRNA to nonsense mediated mRNA decay (NMD) (Gul et al., 2006b, Kouprina et al., 2005).

*Microcephalin/MCPH1* was identified as the first causative gene of MCPH (Jackson et al., 2002). To date twelve mutations are reported in *MCPH1* (Jackson et al., 2002, Trimborn et al., 2004, Trimborn et al., 2005, Garshasbi et al., 2006, Farooq et al., 2010, Darvish et al., 2010). All reported variants in *MCPH1* result in a premature truncation except three missense variants p.Thr27Arg, p.His49>Gln and p.Ser72>Leu which may have an affect on the protein structure by changing the amino acids (Trimborn et al., 2005, Darvish et al., 2010). I identified the variant c.1178delG which causes a frameshift incorporating 49 amino acids before introducing a stop codon (p.R393Sfs\*50). This increases the number of MCPH1 mutations to 13 (Table 5.1). Mutations in MCPH1 are the rare cause of primary microcephaly in the Pakistani population, because out of twelve already identified MCPH1 mutations, only one (c.74C>G) was reported in the Pakistani population (Jackson et al., 2002). Other populations screened for mutations in MCPH1 are Caucasians, Lebaneses and Iranians. Out of twelve known MCPH1 mutations, eight were identified in the Iranian population (Garshasbi et al., 2006, Darvish et al., 2010).

**Table 5.1: Summary of reported *MCPH1* mutations**

No.	cDNA	Protein	Affected Exon	References
1	c.74C>G	p.S25*	2	Jackson et al., 2002
2	427insA	Thr143Asnfs*5+	5	Trimborn et al., 2004
3	c.80C>G	p.T27R	2	Trimborn et al., 2005
4	150-200Kb deletion++	Truncated protein	1 to 6	Garshasbi et al., 2005
5	c.302C>G	S101*	4	Farooq et al., 2010
6	del exon 4	Truncated protein	4	Darvish et al., 2010
7	c.566_567insA	p.Asn189fs	6	Darvish et al., 2010
8	del exon 2&3	Truncated protein	2 and 3	Darvish et al., 2010
9	c.436+1G>T	Truncated protein	5	Darvish et al., 2010
10	c.147C>G	p.His49>Gln	3	Darvish et al., 2010
11	del exon 3	Truncated protein	3	Darvish et al., 2010
12	c.215C>T	p.Ser72>Leu	3	Darvish et al., 2010
13	c.1178delG	p.R393Sfs*50+++	8	This study

+ Insertion resulted in a frameshift incorporating 4 amino acids followed by a stop codon

++ Deletion of approximately 150-200 kb, encompassing the promoter and first 6 exons

+++ Deletion of a single nucleotide G causes a frameshift incorporating 49 amino acids followed by a stop codon (p.R393Sfs\*50).

In *CENPJ*, the causative gene of the MCPH6 locus, four mutations have been reported in families from Iran, Brazil and Pakistan (Bond et al., 2005, Darvish et al., 2010, Gul et al., 2006a). I have identified one already reported mutation (p.Thr6fs\*3) in *CENPJ* increasing its prevalence in the Pakistani society (Table 5.2).

**Table 5.2: Summary of reported *CENPJ* mutations**

No.	cDNA	Protein	Affected Exon	References
1	c.17delC	p.Thr6fs*3	1	Bond et al., 2005, present study
2	c.2462C>T	p.Thr821Met	7	Darvish et al., 2010
3	c.3243delTCAG	p.S1081fs	11	Gul et al., 2006a
4	c.3704A>T	p.E1235V	16	Bond et al., 2005

The MCPH2 locus was identified in 1999 but the causative gene of this locus (*WDR62*) was only recently reported by two independent groups (Yu et al., 2010, Nicholas et al., 2010). They have reported 10 novel pathogenic variants in *WDR62*. Among these are,

three missense mutations, while the other seven are truncating the WDR62 protein. I have identified two novel variants in *WDR62* which increased the total number of different mutations to 12 (Table 5.3).

Mutations in the *WDR62* gene have also been reported in patients afflicted with microcephaly and cortical abnormalities. One missense (W224S) and four nonsense (Q470\*, E526\*, G1280Afs\*21, V1402Gfs\*12) mutations were identified (Bilguvar et al., 2010).

**Table 5.3: Summary of reported *WDR62* mutations**

No.	cDNA mutation	Protein mutation	Affected Exon	References
1	c.193G>A	p.Val65Met	2	Yu et al., 2010
2	c.363delT	p.Asp112Mfs*5	4	Yu et al., 2010
3	c.1043+1G>A	p.Ser348Rfs*63	8	Yu et al., 2010
4	c.2867+4_c.2867+7delGGTG	p.Ser956Cfs*38	23	Yu et al., 2010
5	c.3839_3855delGCCAAGAGC CTGCCCTG	p.Gly1280Afs*21	30	Yu et al., 2010
6	c.1313G>A	p.Arg438His	10	Nicholas et al., 2010
7	c.1531G>A	p.Asp511Asn	11	Nicholas et al., 2010
8	c.3232G>A	p.Ala1078Thr	27	Nicholas et al., 2010
9	c.3936dupC	p.Val1314Rfs*18	30	Nicholas et al., 2010
10	c.4241dup	p.Leu1414Lfs*41	31	Nicholas et al., 2010
11	c.332G>C	p.R111T	3	This study
12	c.3503G>A	p.W1168*	29	This study

## 5.2. *WDR62*, the causative gene of the MCPH2 locus

Among 30 MCP families analyzed during this study, two (MCP14 and MCP59) were linked to the MCPH2 locus. The linkage interval was reduced from the originally reported 7.6cM to 2.8cM by using highly polymorphic microsatellite markers. Sequencing analysis of 8 candidate genes residing in this small region did not show any pathogenic mutation. Later on one group reported *WDR62* as causative gene of the MCPH2 locus. The position of the *WDR62* gene (41,237,623bp to 41,287,852bp according to NCBI Human Genome Build 36.3) is however outside of the linkage interval defined by family MCP14 (38,760,490 to 41,220,124bp) (Figure 4.7). Unfortunately, I have observed heterozygosity of two different markers (D19S224 and M19SH2) adjacent to this region. Interestingly homozygosity mapping based exclusively on the SNP markers of chromosome 19 defined the correct linkage interval of the MCPH2 locus between

rs3786913 (physical position 38,678,914bp) and rs2271844 (physical position 41,592,850 bp) with *WDR62* included (Figure 4.8). Sequencing analysis of MCP14 and 59 revealed two novel homozygous mutations in the *WDR62* gene (Table 5.3). Obviously the homozygosity analysis via STR markers did not provide reliable data and should be avoided in future studies as a method for reducing any linkage interval defined by SNP data.

### **5.3. *CEP152* and the MCPH4 locus**

The MCPH4 locus was originally identified on chromosome 15q15-q21 with a 19 cM minimal critical region mapped between markers ACTC and D15S98 (Jamieson et al., 1999). Homozygosity mapping of five MCP families (further three families from Denmark and Belgium) reduced the linkage interval to 2.08 Mb between rs936532 (37,585,407 bp) and rs1819454 (39,669,675 bp). I have identified one missense variant (c.2437G>T, p.E813\*) in the leukocyte receptor tyrosine kinase (*LTK*) gene residing in a newly defined region. I was however unable to find any further pathogenic variant in any one of the four other MCPH4 linked families. Recently one group from Canada claimed to have identified *CEP152* as the causative gene of the MCPH4 locus (Guernsey et al., 2010). Sequencing results of the *CEP152* gene excluded any pathogenic variant in our MCPH4 linked families which is in line with the fact that *CEP152* is not located in the reduced 2.08 Mb critical interval. These results indicate that either *CEP152* is not the causative gene of the MCPH4 locus, so there might be the possibility of another causative gene at the MCPH4 locus, or in our MCPH4 linked families the mutation might be located in the regulatory elements of *CEP152*. It also might be the case that the leukocyte receptor tyrosine kinase (*LTK*) gene is the MCPH4 causative gene because the variant c.2437G>T was not found in 768 Pakistani healthy control chromosomes and mutations in further MCPH4 linked families might be located in the regulatory elements of *LTK*. However, the mutation in *LTK* might also represent a non-pathogenic nonsense variant which seem to be quite common in human genomes (Yngvadottir et al., 2009).

### **5.4. Identification of novel loci**

Five families could not be linked to any of the known MCPH loci and subsequently lead to the discovery of 5 new gene loci, MCPH8-MCPH12, situated on different chromosomes. For two of the five new loci, namely MCPH8 on chromosome 7q21-q22 (LOD score 10.47) and MCPH9 on chromosome 4p14-4q12 (LOD score 2.53), the causative genes could be identified in families MCP4 and MCP63, respectively. The MCPH10 locus was mapped on chromosome 12q14.1-12q14.2 in family MCP50 (LOD score 4.9). The MCPH11 locus is less precisely defined so far. It should be situated in one of the following three chromosomal regions, 3q12.1-3q12.3, 4q31.23-4q32.3, and 6q24.2-

6q24.3. The 15.57 Mb region mapped on chromosome 4q31.23-4q32.3 would be most interesting to focus on in future whole-exome sequencing of MCP53. Nevertheless, the regions on chromosomes 3 and 6 should also be taken into account when scrutinizing the whole-exome sequencing data for pathogenic variants. The mapping data of the MCPH12 locus are also still ambiguous. Three chromosomal regions will have to be considered, 2q14.2, 6q22.32, or 11p15.4. The region on chromosome 6q22.32 seems to be most promising because it is the largest homozygous region found in family MCP67.

### **5.5. CDK6 as novel MCPH gene**

In family MCP4, cyclin-dependent kinase 6 (*CDK6*) was identified as the causative gene of the MCPH8 locus, where a homozygous single base substitution c.589G>A was found in exon 5. The missense mutation (c.589G>A) in *CDK6* changes an alanine with a threonine (p.A197T). Alanine is strictly conserved in this position in vertebrates and even up to the sea urchin homolog but is not conserved in other human CDKs except for *CDK4* which is considered the closest homolog of *CDK6* (Figure 4.27).

The cyclin-dependent kinase CDK6, a protein of 326 amino acids, together with CDK4 and cyclin D act in concert with CDK2 and cyclin E to regulate the G1 phase and the G1/S transition of the cell cycle. Active CDK4/6 phosphorylate the retinoblastoma tumor suppressor protein (pRb), releasing the transcription factor E2F which induces the transcription of genes including Cyclins A, Cyclin E, DNA polymerase and Thymidine kinase ultimately allowing the cells to enter S-Phase, whereas the INK4 family of inhibitors prevent CDK4/6 activity (Meyerson and Harlow, 1994). CDK6 also plays a role in the differentiation of a variety of cell types. CDK6-null mice are viable, develop normally, display normal organogenesis and cell proliferation in most tissues; however hematopoiesis was impaired (Malumbres et al., 2004).

In non-dividing cells CDK6 was reported to localize in the nucleus and to the cytoplasm with particular enrichment in cytoplasmic protrusions (Kohrt et al., 2009). I found that CDK6 is not only present in the cytosol and in the nucleus but localises also to the spindle poles (centrosome) during mitosis from prophase to telophase as exemplary shown for HaCaT cells (Figure 4.28). Spindle pole localization of CDK6 is similar to the localization of other MCPH proteins like ASPM, STIL and WDR62 (Bond et al., 2005, Pfaff et al., 2007, Bond and Woods, 2006, Nicholas et al., 2010).

Analysis of CDK6 patient primary fibroblasts revealed defects in the growth rate of the cells, aberrant nuclear shape as well as an increased centrosome-nucleus distance as compared to control primary fibroblasts. Suppression of CDK6 by shRNA resulted in similar defects, with regard to nuclear shape and growth rate. Furthermore the microtubule system was disorganised. Overexpression of mutant CDK6 produced

multiple centrosomes and also resulted in disorganised microtubules which confirmed its role in microtubule dynamics (see Table 5.4 for statistical analysis).

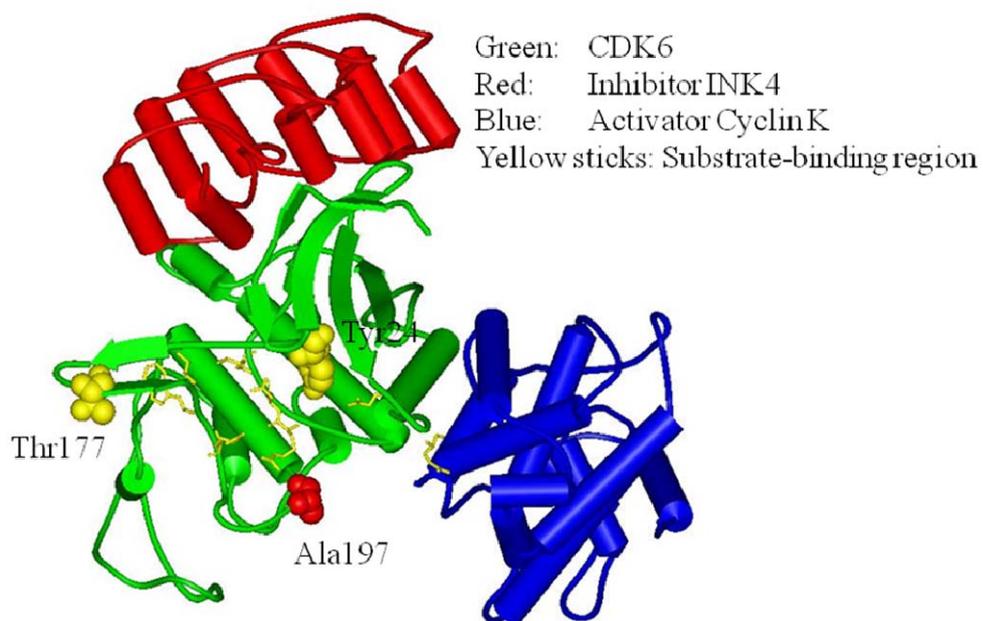
A disorganised microtubule system was also reported for other MCPH proteins, where depletion of known MCPH proteins or their homologs led to the production of abnormal microtubules (Thornton and Woods, 2009). Furthermore, the centrosome abnormalities of CDK6 patient fibroblasts might also be the result of perturbed centrosome duplication as the Rb/E2F pathway regulates centrosome duplication (Meraldi et al., 1999) and the phosphorylation of Rb protein is mediated by CDK6-Cyclin D. Cyclin D1 and D2 are regulatory subunits of CDK6 and are expressed in the neural progenitor cells of the ventricular zone (VZ) and subventricular zone (SVZ), respectively (Meyerson and Harlow, 1994), (Glickstein et al., 2009). Mice lacking Cyclin D2 show reduced proliferation and enhanced cell cycle exit in embryonic cortical progenitors and were microcephalic. It was also demonstrated that Cyclin D2 is used for radial glial cells (RGCs) transition to intermediate progenitor cells (IPCs) and is important for expansion of the IPC number (Glickstein et al., 2009). CDK6 activity is observed at the mid G1 phase of the cell cycle and is responsible for G1 progression and G1 to S phase transition. These are the stages where centrosomes and subsequently genomic DNA are duplicated.

**Table 5.4: Statistical analysis of CDK6 mutant and control primary fibroblasts and HaCaT control and CDK6 knockdown cells.**

Phenotype	Mutant/ Knockdown	Control	P Value	N
Dividing primary fibroblasts after synchronization at G0/G1	0.34 ± 0.17	3.11 ± 0.42	3.32E-04	900
Misshapen nuclei in primary fibroblasts after synchronization at G0/G1	12.34 ± 0.71	2.34 ± 0.24	1.83477E-06	900
Dividing cells after CDK6 knockdown	4.45 ± 0.5	15.3 ± 0.69	1.02517E-06	900
Cells with misshapen nuclei after CDK6 knockdown	21.34 ± 2.02	10 ± 0.5	4.72E-04	900
Centrosome-nucleus distance (µm) measured in control and mutant primary fibroblasts	5.1646 ± 0.2661	3.4514 ± 0.6062	8.286E-03	50

Results are shown in terms of mean ± s.e.m and the corresponding *P* values (Student's *t*-test) which are calculated with control versus mutant data. N is the number of cells analysed.

An interaction between known MCPH proteins and transcription factor E2F has already been reported. Microcephalin promotes the expression of CHK1 and BRCA1 through interaction with transcription factor E2F1 and enhancement of their transcription as well as other E2F targeted genes involved in DNA repair and apoptosis such as RAD51, DDB2, TOPBP1, p73 and caspases (Yang et al., 2008). STIL expression is also regulated by transcription factor E2F, and it is required for E2F induced transition through mitosis (Erez et al., 2008). Based on these results I speculate that the expression of STIL is regulated by CDK6, because active CDK6 protein complex regulates transcription factor E2F and there might be a common pathway between microcephalin and CDK6 to regulate the expression of E2F.



**Figure 5.1: Three-dimensional structure of CDK6 with its activator cyclin K and inhibitor INK4.** Red balls (Ala 197) indicate the position of the mutation (p.A197T). Phosphorylation of Thr177 activates the enzyme, phosphorylation of Tyr24 leads to inactivation.

The consequences of the A197T mutation in the patient are currently unclear as this region is not directly implicated in cyclin or INK binding nor is the ATP-binding site or the catalytic center directly affected (Figure 5.1). Based on the data obtained with the patient's fibroblasts, CDK6 knockdown, and overexpression of wild-type and mutant CDK6, I propose that the A197T mutation may lead to reduced cell proliferation and also affects the correct functioning of the centrosome in microtubule organisation and its positioning near the nucleus.

## 5.6. *CEP135* as novel MCPH gene

The gene on chromosome 4p14-4q12 encoding the centrosomal protein 135 kDa (*CEP135*) was mutated in family MCP63, where a homozygous single base deletion c.970delC in exon 8 of this gene was found. The *CEP135* gene is composed of 26 exons (Fig. 2d). The 3,423-bp ORF encodes a 1,140 amino acids protein. CEP135 was previously identified as a centrosomal component by proteomic analysis and localizes around the centriolar surface as well as within the proximal lumen of the centrioles (Andersen et al., 2003, Kleylein-Sohn et al., 2007). Knockdown of CEP135 caused premature centrosome splitting and disorganized interphase and mitotic spindle microtubules (Kim et al., 2008, Ohta et al., 2002). A role in procentriole formation has been uncovered where CEP135 and CENPJ (MCPH6 protein) form a core structure within the proximal lumen of both parental and nascent centrioles (Kleylein-Sohn et al., 2007).

I observed in CEP135 patient primary fibroblasts alterations with regard to the centrosome, microtubules, cell division and nuclear shape (Table 5.5). The centrosomes were either multiple in number or completely missing in the patient cells which is consistent with its proposed scaffolding role during centriole biogenesis (Kleylein-Sohn et al., 2007) and the premature centrosome splitting and disorganized interphase and mitotic spindle microtubules after CEP135 (Kim et al., 2008, Ohta et al., 2002) and confirm CEP135's role in centriole biogenesis. CEP135 patient's fibroblasts also exhibit slow growth rate and I was unable to detect any single dividing cell, even though the disorganised microtubule network was only observed during interphase. Overexpression of wild-type and mutant CEP135 also leads to disorganised microtubules and the mutant protein led to the production of multiple centrosomes. Similar results have been reported previously where ectopic expression of the CEP135 mutants (one mutant coding only 1-658 and other coding 648 to 1141 amino acids) proteins also caused centrosome splitting in association with a reduction of the levels of centrosomal C-NAP1 (Kim et al., 2008). The abnormal centrosome number strengthens the role of CEP135 in centriole biogenesis, whereas a disorganisation of the microtubule network points out its role at the centrosome as microtubule organising center.

**Table 5.5: Statistical analysis of CEP135 mutant and control primary fibroblasts.**

<b>Phenotype</b>	<b>Mutant</b>	<b>Control</b>	<b>P Value</b>	<b>N</b>
Primary fibroblasts having misshapen nuclei	20 ± 1.15	3 ± 0.58	3.44E-03	300
Primary fibroblasts with multiple centrosomes	18.67 ± 2.73	0	0.0207	300
Primary fibroblasts without centrosomes	22 ± 2.31	0	0.0441	300
Primary fibroblasts with disorganised microtubules	55.33 ± 2.60	0	0.0022	300

Results are shown in terms of mean ± s.e.m with corresponding *P* values (Student's *t*-test). N is the number of cells analysed.

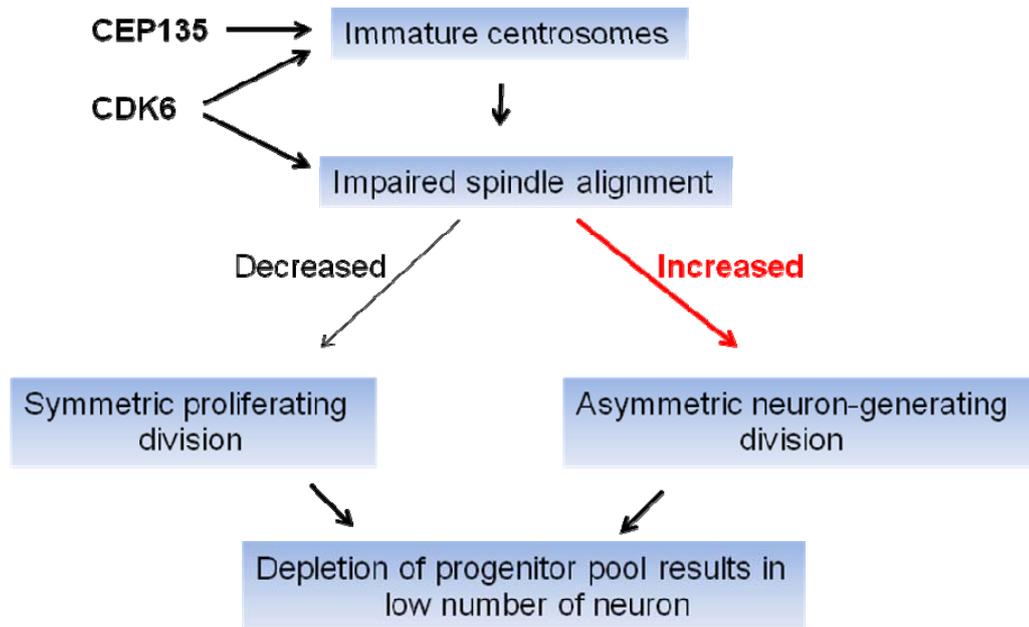
### **5.7. Cdk6 and Cep135 expression in the neuroepithelium of the mouse cerebral cortex**

Most known MCPH proteins are expressed in the ventricular zone (VZ) of the mouse neuroepithelium, particularly in progenitors undergoing proliferative divisions (Thornton and Woods, 2009). Our study also shows a strong staining for CEP135 and CDK6 in the developing neuroepithelium of the cerebral cortex. Detection of CEP135 and CDK6 at E11.5 and E15.5 shows that both proteins are expressed in the neuroepithelium of the developing cortex during neurogenesis. This localization is reminiscent of that of WDR62, a spindle pole protein expressed in neuronal precursor cells undergoing mitosis in the mammalian embryonic neuroepithelium. Mutations in this protein presumably decrease the number of neuronal precursor cells giving rise to microcephaly (Nicholas et al., 2010, Yu et al., 2010). Our studies suggest that CEP135 and CDK6 are important components of the spindle pole and might be involved in controlling the position of the central spindle and regulation of cytokinesis.

### **5.8. Final conclusions and outlook**

Brain size is largely determined by the relative rates of proliferation and cell death, and proteins encoded by MCPH genes have been shown to play a prominent role in several key steps of developmental processes that could be responsible for a marked reduction of the cortical area, such as cell cycle progression, DNA and spindle checkpoints as well as centrosome and spindle apparatus integrity. Some common pathways have been identified for MCPH proteins such as the interaction of MCPH1 and STIL with transcription factor E2F (Kaindl et al., 2010).

In this work I propose a novel role for CDK6 at the centrosome during particular cell cycle phases. Together with the implication of the centrosomal component CEP135 in MCPH this strengthens the role of centrosomes and spindle poles in the etiology of MCPH. In line with previous suggestions I speculate that mutations in these proteins can affect proliferation of apical neuronal precursor cells and/or can create an imbalance of symmetric and asymmetric cell division resulting in depletion of the progenitor pool which may be the cause of reduced neuron production.



**Figure 5.2: A model of CDK6 and CEP135 functions during neurogenesis.** Loss of CDK6 might cause premature mitotic entry which affects centrosome maturation. Loss of CEP135 is suggested to results in immature centrosomes or a loss of the centriole. The immature centrosomes have a reduced PCM accumulation and can affect spindle orientation during division due to formation of fewer astral microtubules. CDK6 like ASPM localizes to the spindle pole of the centrosome and directly regulates spindle positioning. Impaired spindle alignment can result in asymmetric divisions and reduction of the progenitor pool, which ultimately results in fewer neurons. The idea was expanded from Thornton and Woods, 2009.

Elucidation of further MCPH genes and their pathomechanisms will not only increase our knowledge of this Mendelian disease but also enhance our understanding of the normal brain development and of the evolution of the brain development. Moreover, the identification of novel loci, genes and their pathogenic variants in Pakistani families as analyzed in this study will help to design improved strategies of genetic counseling in Pakistan and will also be of great importance to establish prenatal diagnosis and carrier screening in this country.

## 6. Abstract

Autosomal recessive primary microcephaly (MCPH) is a rare genetic disorder in which the afflicted individuals have head circumference more than 3 SDs below the age- and sex-related mean. The reduced head circumference is due to a small but architecturally normal cerebral cortex. MCPH is characterized by a pronounced heterogeneity with seven loci, designated MCPH1-7, have already been identified. The underlying genetic defects were found in the following seven genes, *MCPH1*, *WDR62*, *CDK5RAP2*, *CEP152*, *ASPM*, *CENPJ*, and *STIL/SIL*. The incidence of this disorder is highest in Pakistan (Woods et al., 2005).

Here, I ascertained thirty families with MCPH from various regions of Pakistan. Homozygosity mapping revealed linkage in 19 families to the MCPH5 locus, in 2 to MCPH2, in 2 to MCPH4, in 1 to MCPH1, in 1 to MCPH6, and in 5 families linkage to all known MCPH loci was excluded. Families linked to the MCPH1, MCPH2, MCPH5, and MCPH6 loci were also subjected to direct genomic sequencing of the corresponding genes, i.e. *MCPH1*, *WDR62*, *ASPM*, and *CENPJ*, respectively. This revealed one, two, and nine novel mutations in *MCPH1*, *WDR62*, and *ASPM*, respectively. Genome-wide linkage analysis in the 5 families previously excluded to be linked to any of the known loci resulted in 5 different new gene loci, MCPH8-MCPH12, situated on different chromosomes. For two of the five new loci, namely MCPH8 on chromosome 7q21-q22 (LOD score 10.47) and MCPH9 on chromosome 4p14-4q12 (LOD score 2.53), the causative genes could be identified. Positional candidate gene sequencing revealed mutations in *CDK6* (c.589G>A, p.A197T) at the MCPH8 locus and in *CEP135* (c.970delC, p.Gln324Serfs\*2) at the MCPH9 locus as the most likely pathogenic variants. These variants were not found in 768 chromosomes from healthy Pakistani controls.

These two novel MCPH proteins cyclin-dependent kinase 6 (CDK6) and a centrosomal protein of 135kDa (CEP135) presented as transient or permanent components of the centrosome. Cdk6 and Cep135 showed a high expression level in the developing neuroepithelium of the mouse cerebral cortex of E11.5 and E15.5 embryos. In human cell lines, the localization of CDK6 at the spindle pole was observed. Primary fibroblasts of the patient with the *CDK6* mutation failed to grow normally and showed an aberrant nuclear shape as well as centrosome-nucleus distance. CDK6 suppression by shRNA mimicked the defects in cell proliferation, nuclear shape, and microtubule organization. Likewise, overexpression of mutant CDK6 resulted in the production of multiple centrosomes and disorganised microtubules. Primary fibroblasts of the patient with the *CEP135* mutation showed multiple and fragmented centrosomes, a disorganised microtubule system, misshapen and fragmented nuclei, and sometimes a complete loss

of centrosomes. Altered levels of wild-type and mutant CEP135 protein by overexpression caused disorganization of microtubules, while overexpression of mutant CEP135 showed also multiple centrosomes observed before in the patient's primary fibroblasts.

Based on the data on CDK6, I propose that mutation p.A197T may lead to a reduced cell proliferation and may also affect the correct functioning of the centrosome in microtubule organisation and its positioning near the nucleus. The abnormal centrosome number associated with mutant CEP135 strengthens its role in centriole biogenesis, whereas a disorganisation of the microtubule network points to its role at the centrosome as a microtubule organising center. The data obtained lend further support to the hypothesis that the exquisite control of the cleavage furrow orientation in mammalian neural precursor cell mitosis, controlled in great part by the centrosomes and spindle poles, is critical in the etiology of MCPH (Fish et al., 2006, Thornton and Woods, 2009).

## 7. Zusammenfassung

Autosomal rezessive primäre Mikrozephalie (MCPH) ist eine seltene, genetisch bedingte Erkrankung, bei der die Betroffenen einen verringerten Kopfumfang von mindestens drei Standardabweichungen unter dem für das Alter und Geschlecht durchschnittlichen Normalwert haben. Der reduzierte Kopfumfang ist auf eine kleinere, aber architektonisch normale Hirnrinde zurückzuführen. MCPH ist durch eine ausgeprägte genetische Heterogenität gekennzeichnet. Bisher sind sieben Loci bekannt, die als MCPH1-7 bezeichnet werden. Die zugrundeliegenden genetischen Defekte wurden in den folgenden sieben Genen gefunden: *MCPH1*, *WDR62*, *CDK5RAP2*, *CEP152*, *ASPM*, *CENPJ* und *STIL/SIL*. Die Inzidenz dieser Erkrankung ist in Pakistan am höchsten (Woods et al., 2005).

In der vorliegenden Arbeit wurden dreißig Familien mit MCPH aus verschiedenen Regionen Pakistans untersucht. Homozygotie-Kartierungen ergaben, dass die Erkrankung in 19 Familien mit dem MCPH5-Lokus, in zwei Familien mit dem MCPH2-Lokus, in zwei weiteren Familien mit dem MCPH4-Lokus, in einer Familie mit dem MCPH1-Lokus und in einer anderen Familie mit dem MCPH6 gekoppelt ist. In fünf Familien wurden alle bekannten MCPH-Loci ausgeschlossen. In den Familien, bei denen eine Kopplung mit dem MCPH1-, MCPH2-, MCPH5- oder MCPH6-Lokus vorlag, wurden die entsprechenden Gene, d. h. *MCPH1*, *WDR62*, *ASPM* und *CENPJ*, genomisch sequenziert. Dabei konnten in den Genen *MCPH1*, *WDR62* und *ASPM* eine, zwei bzw. neun neue Mutationen gefunden werden. Eine genomweite Kopplungsanalyse in den 5 Familien, in denen zuvor eine Kopplung zu bekannten Loci ausgeschlossen worden war, resultierte in 5 verschiedenen neuen Loci, MCPH8-MCPH12, die sich auf verschiedenen Chromosomen befinden. Für zwei der fünf neuen Loci, MCPH8 auf Chromosom 7q21-q22 (LOD-Score 10,47) und MCPH9 auf Chromosom 4p14-4q12 (LOD-Score 2,53), konnten die ursächlichen Gene identifiziert werden. Positionelle Kandidatengensequenzierungen offenbarten Mutationen in *CDK6* (c.589G> A, p.A197T, MCPH8-Lokus) und in *CEP135* (c.970delC, p.Gln324Serfs\*2, MCPH9-Lokus) als die jeweils wahrscheinlichsten pathogenen Varianten. Diese Varianten waren in 768 Chromosomen gesunder pakistanischer Kontrollen nicht nachweisbar.

Die zu den neuen MCPH-Genen korrespondierenden Proteine, die Cyclin-abhängige Kinase 6 (CDK6) und das Zentrosomale Protein 135 kDa (CEP135), stellen vorübergehende oder dauerhafte Komponenten des Zentrosoms dar. In Mausembryonen (Stadien E11.5 und E15.5) zeigen die Proteine Cdk6 und Cep135 eine hohe Expression im sich entwickelnden Neuroepithel der Großhirnrinde. In humanen Zelllinien konnte erstmalig die Lokalisation von CDK6 am Spindelpol beobachtet werden. Primäre

Fibroblasten des Patienten mit der CDK6-Mutation wuchsen sehr schlecht und zeigten eine abweichende Kernform sowie einen extremen Zentrosom-Kern-Abstand. Die CDK6-Inhibierung durch shRNA ahmte die Proliferationschwäche der Zellen, die veränderte Kernform und die gestörte Mikrotubuli-Organisation nach. Ebenso führte die Überexpression von mutiertem CDK6 zur Bildung von multiplen Zentrosomen und desorganisierten Mikrotubuli. Primäre Fibroblasten des Patienten mit der CEP135-Mutation zeigten multiple und fragmentierte Zentrosomen, ein desorganisiertes Mikrotubuli-System, deformierte und fragmentierte Kerne sowie manchmal einen vollständigen Verlust der Zentrosomen. Eine Überexpression von Wild-typ und mutiertem CEP135-Protein führte zu einer Desorganisation der Mikrotubuli, während bei der Überexpression von mutiertem CEP135 zusätzlich multiple Zentrosomen wie zuvor in den primären Fibroblasten des Patienten zu beobachten waren.

Basierend auf den CDK6-Daten kann vermutet werden, dass die Mutation p.A197T zu einer verringerten Zellproliferation führt und auch die korrekte Funktion der Zentrosomen für die Mikrotubuli-Organisation und ihre Positionierung in der Nähe des Kerns beeinträchtigt ist. Die abnorme Zentrosomenzahl, die mit dem mutierten CEP135 assoziiert ist, weist auf die Rolle dieses Proteins in der Zentriol-Biogenese hin, während die Desorganisation der Mikrotubuli auf seine Rolle am Zentrosom als Mikrotubuli-organisierendes Zentrum hindeutet.

Die gewonnenen Daten bestätigen erneut die Hypothese, dass die exquisite Kontrolle der Teilungsfurchenorientierung während der Mitose der neuralen Vorläuferzellen, die in starkem Maße von den Zentrosomen und den Spindelpolen kontrolliert wird, für die Ätiologie der MCPH entscheidend ist (Fish et al., 2006, Thornton and Woods, 2009).

## 8. References:

- Abecasis,G.R., Cherny,S.S., Cookson,W.O., and Cardon,L.R. (2001). GRR: graphical representation of relationship errors. *Bioinformatics*. 17, 742-743.
- Abecasis,G.R., Cherny,S.S., Cookson,W.O., and Cardon,L.R. (2002). Merlin--rapid analysis of dense genetic maps using sparse gene flow trees. *Nat. Genet.* 30, 97-101.
- Aerts,S., Lambrechts,D., Maity,S., Van,L.P., Coessens,B., De,S.F., Tranchevent,L.C., De,M.B., Marynen,P., Hassan,B., Carmeliet,P., and Moreau,Y. (2006). Gene prioritization through genomic data fusion. *Nat. Biotechnol.* 24, 537-544.
- Andersen,J.S., Wilkinson,C.J., Mayor,T., Mortensen,P., Nigg,E.A., and Mann,M. (2003). Proteomic characterization of the human centrosome by protein correlation profiling. *Nature* 426, 570-574.
- Baig,S.M., Din,M.A., Hassan,H., Azhar,A., Baig,J.M., Aslam,M., Anjum,I., Farooq,M., Hussain,M.S., Rasool,M., Nawaz,S., Qureshi,J.A., and Zaman,T. (2008). Prevention of beta-thalassemia in a large Pakistani family through cascade testing. *Community Genet.* 11, 68-70.
- Baig,S.M., Koschak,A., Lieb,A., Gebhart,M., Dafinger,C., Nurnberg,G., Ali,A., Ahmad,I., Sinnegger-Brauns,M.J., Brandt,N., Engel,J., Mangoni,M.E., Farooq,M., Khan,H.U., Nurnberg,P., Striessnig,J., and Bolz,H.J. (2011). Loss of Ca(v)1.3 (CACNA1D) function in a human channelopathy with bradycardia and congenital deafness. *Nat. Neurosci.* 14, 77-84.
- Barkovich,A.J., Kuzniecky,R.I., and Dobyns,W.B. (2001). Radiologic classification of malformations of cortical development. *Curr. Opin. Neurol.* 14, 145-149.
- Basto,R., Lau,J., Vinogradova,T., Gardiol,A., Woods,C.G., Khodjakov,A., and Raff,J.W. (2006). Flies without centrioles. *Cell* 125, 1375-1386.
- Bilguvar,K., Ozturk,A.K., Louvi,A., Kwan,K.Y., Choi,M., Tatli,B., Yalnizoglu,D., Tuysuz,B., Caglayan,A.O., Gokben,S., Kaymakcalan,H., Barak,T., Bakircioglu,M., Yasuno,K., Ho,W., Sanders,S., Zhu,Y., Yilmaz,S., Dincer,A., Johnson,M.H., Bronen,R.A., Kocer,N., Per,H., Mane,S., Pamir,M.N., Yalcinkaya,C., Kumandas,S., Topcu,M., Ozmen,M., Sestan,N., Lifton,R.P., State MW, and Gunel,M. (2010). Whole-exome sequencing identifies recessive WDR62 mutations in severe brain malformations. *Nature* 467, 207-210.
- Bond,J., Roberts,E., Mochida,G.H., Hampshire,D.J., Scott,S., Askham,J.M., Springell,K., Mahadevan,M., Crow,Y.J., Markham,A.F., Walsh,C.A., and Woods,C.G. (2002). ASPM is a major determinant of cerebral cortical size. *Nat. Genet.* 32, 316-320.
- Bond,J., Roberts,E., Springell,K., Lizarraga,S.B., Scott,S., Higgins,J., Hampshire,D.J., Morrison,E.E., Leal,G.F., Silva,E.O., Costa,S.M., Baralle,D., Raponi,M., Karbani,G., Rashid,Y., Jafri,H., Bennett,C., Corry,P., Walsh,C.A., and Woods,C.G. (2005). A centrosomal mechanism involving CDK5RAP2 and CENPJ controls brain size. *Nat. Genet.* 37, 353-355.
- Bond,J., Scott,S., Hampshire,D.J., Springell,K., Corry,P., Abramowicz,M.J., Mochida,G.H., Hennekam,R.C., Maher,E.R., Fryns,J.P., Alswaid,A., Jafri,H., Rashid,Y., Mubaidin,A., Walsh,C.A., Roberts,E., and Woods,C.G. (2003). Protein-

- truncating mutations in ASPM cause variable reduction in brain size. *Am. J. Hum. Genet.* **73**, 1170-1177.
- Bond,J. and Woods,C.G. (2006). Cytoskeletal genes regulating brain size. *Curr. Opin. Cell Biol.* **18**, 95-101.
- Boukamp,P., Petrussevska,R.T., Breitkreutz,D., Hornung,J., Markham,A., and Fusenig,N.E. (1988). Normal keratinization in a spontaneously immortalized aneuploid human keratinocyte cell line. *J. Cell Biol.* **106**, 761-771.
- Cho,J.H., Chang,C.J., Chen,C.Y., and Tang,T.K. (2006). Depletion of CPAP by RNAi disrupts centrosome integrity and induces multipolar spindles. *Biochem. Biophys. Res. Commun.* **339**, 742-747.
- Clark,D.A., Mitra,P.P., and Wang,S.S. (2001). Scalable architecture in mammalian brains. *Nature* **411**, 189-193.
- Cox,J., Jackson,A.P., Bond,J., and Woods,C.G. (2006). What primary microcephaly can tell us about brain growth. *Trends Mol. Med.* **12**, 358-366.
- Darvish,H., Esmaeeli-Nieh,S., Monajemi,G.B., Mohseni,M., Ghasemi-Firouzabadi,S., Abedini,S.S., Bahman,I., Jamali,P., Azimi,S., Mojahedi,F., Dehghan,A., Shafeghati,Y., Jankhah,A., Falah,M., Soltani Banavandi,M.J., Ghani-Kakhi,M., Garshasbi,M., Rakhshani,F., Naghavi,A., Tzschach,A., Neitzel,H., Ropers,H.H., Kuss,A.W., Behjati,F., Kahrizi,K., and Najmabadi,H. (2010). A clinical and molecular genetic study of 112 Iranian families with primary microcephaly. *J. Med. Genet.* **47**, 823-828.
- Desir,J., Cassart,M., David,P., Van,B.P., and Abramowicz,M. (2008). Primary microcephaly with ASPM mutation shows simplified cortical gyration with antero-posterior gradient pre- and post-natally. *Am. J. Med. Genet. A* **146A**, 1439-1443.
- do Carmo,A.M., Tavares,A., and Glover,D.M. (2001). Polo kinase and Asp are needed to promote the mitotic organizing activity of centrosomes. *Nat. Cell Biol.* **3**, 421-424.
- Dobbelaere,J., Josue,F., Suijkerbuijk,S., Baum,B., Tapon,N., and Raff,J. (2008). A genome-wide RNAi screen to dissect centriole duplication and centrosome maturation in *Drosophila*. *PLoS. Biol.* **6**, e224.
- Dzhindzhev,N.S., Yu,Q.D., Weiskopf,K., Tzolovsky,G., Cunha-Ferreira,I., Riparbelli,M., Rodrigues-Martins,A., Bettencourt-Dias,M., Callaini,G., and Glover,D.M. (2010). Asterless is a scaffold for the onset of centriole assembly. *Nature* **467**, 714-718.
- Erez,A., Chaussepied,M., Castiel,A., Colaizzo-Anas,T., Aplan,P.D., Ginsberg,D., and Izraeli,S. (2008). The mitotic checkpoint gene, SIL is regulated by E2F1. *Int. J. Cancer* **123**, 1721-1725.
- Evans,P.D., Gilbert,S.L., Mekel-Bobrov,N., Vallender,E.J., Anderson,J.R., Vaez-Azizi,L.M., Tishkoff,S.A., Hudson,R.R., and Lahn,B.T. (2005). Microcephalin, a gene regulating brain size, continues to evolve adaptively in humans. *Science* **309**, 1717-1720.
- Farooq,M., Baig,S., Tommerup,N., and Kjaer,K.W. (2010). Craniosynostosis-microcephaly with chromosomal breakage and other abnormalities is caused by a truncating MCPH1 mutation and is allelic to premature chromosomal condensation

- syndrome and primary autosomal recessive microcephaly type 1. *Am. J. Med. Genet. A* 152A, 495-497.
- Finlay,B.L. and Darlington,R.B. (1995). Linked regularities in the development and evolution of mammalian brains. *Science* 268, 1578-1584.
- Fish,J.L., Kosodo,Y., Enard,W., Paabo,S., and Huttner,W.B. (2006). Aspm specifically maintains symmetric proliferative divisions of neuroepithelial cells. *Proc. Natl. Acad. Sci. U. S. A* 103, 10438-10443.
- Fong,K.W., Choi,Y.K., Rattner,J.B., and Qi,R.Z. (2008). CDK5RAP2 is a pericentriolar protein that functions in centrosomal attachment of the gamma-tubulin ring complex. *Mol. Biol. Cell* 19, 115-125.
- Garshasbi,M., Motazacker,M.M., Kahrizi,K., Behjati,F., Abedini,S.S., Nieh,S.E., Firouzabadi,S.G., Becker,C., Ruschendorf,F., Nurnberg,P., Tzschach,A., Vazifehmand,R., Erdogan,F., Ullmann,R., Lenzner,S., Kuss,A.W., Ropers,H.H., and Najmabadi,H. (2006). SNP array-based homozygosity mapping reveals MCPH1 deletion in family with autosomal recessive mental retardation and mild microcephaly. *Hum. Genet.* 118, 708-715.
- Glickstein,S.B., Monaghan,J.A., Koeller,H.B., Jones,T.K., and Ross,M.E. (2009). Cyclin D2 is critical for intermediate progenitor cell proliferation in the embryonic cortex. *J. Neurosci.* 29, 9614-9624.
- Gluzman,Y. (1981). SV40-transformed simian cells support the replication of early SV40 mutants. *Cell* 23, 175-182.
- Gonzalez,C., Saunders,R.D., Casal,J., Molina,I., Carmena,M., Ripoll,P., and Glover,D.M. (1990). Mutations at the asp locus of *Drosophila* lead to multiple free centrosomes in syncytial embryos, but restrict centrosome duplication in larval neuroblasts. *J. Cell Sci.* 96 ( Pt 4), 605-616.
- Graser,S., Stierhof,Y.D., and Nigg,E.A. (2007). Cep68 and Cep215 (Cdk5rap2) are required for centrosome cohesion. *J. Cell Sci.* 120, 4321-4331.
- Gudbjartsson,D.F., Jonasson,K., Frigge,M.L., and Kong,A. (2000). Allegro, a new computer program for multipoint linkage analysis. *Nat. Genet.* 25, 12-13.
- Guernsey,D.L., Jiang,H., Hussin,J., Arnold,M., Bouyakdan,K., Perry,S., Babineau-Sturk,T., Beis,J., Dumas,N., Evans,S.C., Ferguson,M., Matsuoka,M., Macgillivray,C., Nightingale,M., Patry,L., Rideout,A.L., Thomas,A., Orr,A., Hoffmann,I., Michaud,J.L., Awadalla,P., Meek,D.C., Ludman,M., and Samuels,M.E. (2010). Mutations in centrosomal protein CEP152 in primary microcephaly families linked to MCPH4. *Am. J. Hum. Genet.* 87, 40-51.
- Gul,A., Hassan,M.J., Hussain,S., Raza,S.I., Chishti,M.S., and Ahmad,W. (2006a). A novel deletion mutation in CENPJ gene in a Pakistani family with autosomal recessive primary microcephaly. *J. Hum. Genet.* 51, 760-764.
- Gul,A., Hassan,M.J., Mahmood,S., Chen,W., Rahmani,S., Naseer,M.I., Dellefave,L., Muhammad,N., Rafiq,M.A., Ansar,M., Chishti,M.S., Ali,G., Siddique,T., and Ahmad,W. (2006b). Genetic studies of autosomal recessive primary microcephaly in 33 Pakistani families: Novel sequence variants in ASPM gene. *Neurogenetics.* 7, 105-110.

- Gul,A., Tariq,M., Khan,M.N., Hassan,M.J., Ali,G., and Ahmad,W. (2007). Novel protein-truncating mutations in the ASPM gene in families with autosomal recessive primary microcephaly. *J. Neurogenet.* 21, 153-163.
- Halsall,S., Nicholas,A.K., Thornton,G., Martin,H., and Geoffrey,W.C. (2010). Critical consequences of finding three pathogenic mutations in an individual with recessive disease. *J. Med. Genet.* 47, 769-770.
- Haren,L., Stearns,T., and Luders,J. (2009). Plk1-dependent recruitment of gamma-tubulin complexes to mitotic centrosomes involves multiple PCM components. *PLoS. One.* 4, e5976.
- Hung,L.Y., Chen,H.L., Chang,C.W., Li,B.R., and Tang,T.K. (2004). Identification of a novel microtubule-destabilizing motif in CPAP that binds to tubulin heterodimers and inhibits microtubule assembly. *Mol. Biol. Cell* 15, 2697-2706.
- Hung,L.Y., Tang,C.J., and Tang,T.K. (2000). Protein 4.1 R-135 interacts with a novel centrosomal protein (CPAP) which is associated with the gamma-tubulin complex. *Mol. Cell Biol.* 20, 7813-7825.
- Huyton,T., Bates,P.A., Zhang,X., Sternberg,M.J., and Freemont,P.S. (2000). The BRCA1 C-terminal domain: structure and function. *Mutat. Res.* 460, 319-332.
- Izraeli,S., Lowe,L.A., Bertness,V.L., Good,D.J., Dorward,D.W., Kirsch,I.R., and Kuehn,M.R. (1999). The SIL gene is required for mouse embryonic axial development and left-right specification. *Nature* 399, 691-694.
- Jackson,A.P., Eastwood,H., Bell,S.M., Adu,J., Toomes,C., Carr,I.M., Roberts,E., Hampshire,D.J., Crow,Y.J., Mighell,A.J., Karbani,G., Jafri,H., Rashid,Y., Mueller,R.F., Markham,A.F., and Woods,C.G. (2002). Identification of microcephalin, a protein implicated in determining the size of the human brain. *Am. J. Hum. Genet.* 71, 136-142.
- Jackson,A.P., McHale,D.P., Campbell,D.A., Jafri,H., Rashid,Y., Mannan,J., Karbani,G., Corry,P., Levene,M.I., Mueller,R.F., Markham,A.F., Lench,N.J., and Woods,C.G. (1998). Primary autosomal recessive microcephaly (MCPH1) maps to chromosome 8p22-pter. *Am. J. Hum. Genet.* 63, 541-546.
- Jamieson,C.R., Govaerts,C., and Abramowicz,M.J. (1999). Primary autosomal recessive microcephaly: homozygosity mapping of MCPH4 to chromosome 15. *Am. J. Hum. Genet.* 65, 1465-1469.
- Jeffers,L.J., Coull,B.J., Stack,S.J., and Morrison,C.G. (2008). Distinct BRCT domains in Mcph1/Brit1 mediate ionizing radiation-induced focus formation and centrosomal localization. *Oncogene* 27, 139-144.
- Jensen,F.C., Girardi,A.J., Gilden,R.V., and Koprowski,H. (1964). Infection of human and simian tissue cultures with rous sarcoma virus. *Proc. Natl. Acad. Sci. U. S. A* 52, 53-59.
- Kaindl,A.M., Passemard,S., Kumar,P., Kraemer,N., Issa,L., Zwirner,A., Gerard,B., Verloes,A., Mani,S., and Gressens,P. (2010). Many roads lead to primary autosomal recessive microcephaly. *Prog. Neurobiol.* 90, 363-383.
- Kalay,E., Yigit,G., Aslan,Y., Brown,K.E., Pohl,E., Bicknell,L.S., Kayserili,H., Li,Y., Tuysuz,B., Nurnberg,G., Kiess,W., Koegl,M., Baessmann,I., Buruk,K., Toraman,B.,

- Kayipmaz,S., Kul,S., Ikbal,M., Turner,D.J., Taylor,M.S., Aerts,J., Scott,C., Milstein,K., Dollfus,H., Wieczorek,D., Brunner,H.G., Hurles,M., Jackson,A.P., Rauch,A., Nurnberg,P., Karaguzel,A., and Wollnik,B. (2011). CEP152 is a genome maintenance protein disrupted in Seckel syndrome. *Nat. Genet.* 43, 23-26.
- Kilmartin,J.V., Wright,B., and Milstein,C. (1982). Rat monoclonal antitubulin antibodies derived by using a new nonsecreting rat cell line. *J. Cell Biol.* 93, 576-582.
- Kim,K., Lee,S., Chang,J., and Rhee,K. (2008). A novel function of CEP135 as a platform protein of C-NAP1 for its centriolar localization. *Exp. Cell Res.* 314, 3692-3700.
- Kleylein-Sohn,J., Westendorf,J., Le,C.M., Habedanck,R., Stierhof,Y.D., and Nigg,E.A. (2007). Plk4-induced centriole biogenesis in human cells. *Dev. Cell* 13, 190-202.
- Kohler,S., Bauer,S., Horn,D., and Robinson,P.N. (2008). Walking the interactome for prioritization of candidate disease genes. *Am. J. Hum. Genet.* 82, 949-958.
- Kohlmaier,G., Loncarek,J., Meng,X., McEwen,B.F., Mogensen,M.M., Spektor,A., Dynlacht,B.D., Khodjakov,A., and Gonczy,P. (2009). Overly long centrioles and defective cell division upon excess of the SAS-4-related protein CPAP. *Curr. Biol.* 19, 1012-1018.
- Kohrt,D.M., Crary,J.I., Gocheva,V., Hinds,P.W., and Grossel,M.J. (2009). Distinct subcellular distribution of cyclin dependent kinase 6. *Cell Cycle* 8, 2837-2843.
- KOMAI,T., KISHIMOTO,K., and OZAKI,Y. (1955). Genetic study of microcephaly based on Japanese material. *Am. J. Hum. Genet.* 7, 51-65.
- Kouprina,N., Pavlicek,A., Collins,N.K., Nakano,M., Noskov,V.N., Ohzeki,J., Mochida,G.H., Risinger,J.I., Goldsmith,P., Gunsior,M., Solomon,G., Gersch,W., Kim,J.H., Barrett,J.C., Walsh,C.A., Jurka,J., Masumoto,H., and Larionov,V. (2005). The microcephaly ASPM gene is expressed in proliferating tissues and encodes for a mitotic spindle protein. *Hum. Mol. Genet.* 14, 2155-2165.
- Kousar,R., Nawaz,H., Khurshid,M., Ali,G., Khan,S.U., Mir,H., Ayub,M., Wali,A., Ali,N., Jelani,M., Basit,S., Ahmad,W., and Ansar,M. (2010). Mutation analysis of the ASPM gene in 18 Pakistani families with autosomal recessive primary microcephaly. *J. Child Neurol.* 25, 715-720.
- Kumar,A., Blanton,S.H., Babu,M., Markandaya,M., and Girimaji,S.C. (2004). Genetic analysis of primary microcephaly in Indian families: novel ASPM mutations. *Clin. Genet.* 66, 341-348.
- Kumar,A., Girimaji,S.C., Duvvari,M.R., and Blanton,S.H. (2009). Mutations in STIL, encoding a pericentriolar and centrosomal protein, cause primary microcephaly. *Am. J. Hum. Genet.* 84, 286-290.
- Leal,G.F., Roberts,E., Silva,E.O., Costa,S.M., Hampshire,D.J., and Woods,C.G. (2003). A novel locus for autosomal recessive primary microcephaly (MCPH6) maps to 13q12.2. *J. Med. Genet.* 40, 540-542.
- Leidel,S. and Gonczy,P. (2005). Centrosome duplication and nematodes: recent insights from an old relationship. *Dev. Cell* 9, 317-325.
- Li,K. and Kaufman,T.C. (1996). The homeotic target gene centrosomin encodes an essential centrosomal component. *Cell* 85, 585-596.

- Lin,S.Y., Rai,R., Li,K., Xu,Z.X., and Elledge,S.J. (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*, 15105-15109.
- Lindner,T.H. and Hoffmann,K. (2005). easyLINKAGE: a PERL script for easy and automated two-/multi-point linkage analyses. *Bioinformatics*. *21*, 405-407.
- Lucas,E.P. and Raff,J.W. (2007). Maintaining the proper connection between the centrioles and the pericentriolar matrix requires *Drosophila* centrosomin. *J. Cell Biol.* *178*, 725-732.
- Malik,S., Abbasi,A.A., Ansar,M., Ahmad,W., Koch,M.C., and Grzeschik,K.H. (2006). Genetic heterogeneity of synpolydactyly: a novel locus SPD3 maps to chromosome 14q11.2-q12. *Clin. Genet.* *69*, 518-524.
- Malumbres,M., Sotillo,R., Santamaria,D., Galan,J., Cerezo,A., Ortega,S., Dubus,P., and Barbacid,M. (2004). Mammalian cells cycle without the D-type cyclin-dependent kinases Cdk4 and Cdk6. *Cell* *118*, 493-504.
- Megraw,T.L., Li,K., Kao,L.R., and Kaufman,T.C. (1999). The centrosomin protein is required for centrosome assembly and function during cleavage in *Drosophila*. *Development* *126*, 2829-2839.
- Mekel-Bobrov,N., Gilbert,S.L., Evans,P.D., Vallender,E.J., Anderson,J.R., Hudson,R.R., Tishkoff,S.A., and Lahn,B.T. (2005). Ongoing adaptive evolution of ASPM, a brain size determinant in *Homo sapiens*. *Science* *309*, 1720-1722.
- Meraldi,P., Lukas,J., Fry,A.M., Bartek,J., and Nigg,E.A. (1999). Centrosome duplication in mammalian somatic cells requires E2F and Cdk2-cyclin A. *Nat. Cell Biol.* *1*, 88-93.
- Meyerson,M. and Harlow,E. (1994). Identification of G1 kinase activity for cdk6, a novel cyclin D partner. *Mol. Cell Biol.* *14*, 2077-2086.
- Moynihan,L., Jackson,A.P., Roberts,E., Karbani,G., Lewis,I., Corry,P., Turner,G., Mueller,R.F., Lench,N.J., and Woods,C.G. (2000). A third novel locus for primary autosomal recessive microcephaly maps to chromosome 9q34. *Am. J. Hum. Genet.* *66*, 724-727.
- Muhammad,F., Mahmood,B.S., Hansen,L., Sajid,H.M., Anjum,I., I, Aslam,M., Anver,Q.J., Toilat,M., Kirst,E., Wajid,M., Nurnberg,P., Eiberg,H., Tommerup,N., and Kjaer,K.W. (2009). Compound heterozygous ASPM mutations in Pakistani MCPH families. *Am. J. Med. Genet. A* *149A*, 926-930.
- Nicholas,A.K., Khurshid,M., Desir,J., Carvalho,O.P., Cox,J.J., Thornton,G., Kausar,R., Ansar,M., Ahmad,W., Verloes,A., Passemard,S., Misson,J.P., Lindsay,S., Gergely,F., Dobyns,W.B., Roberts,E., Abramowicz,M., and Woods,C.G. (2010). WDR62 is associated with the spindle pole and is mutated in human microcephaly. *Nat. Genet.* *42*, 1010-1014.
- Nicholas,A.K., Swanson,E.A., Cox,J.J., Karbani,G., Malik,S., Springell,K., Hampshire,D., Ahmed,M., Bond,J., Di,B.D., Fichera,M., Romano,C., Dobyns,W.B., and Woods,C.G. (2009). The molecular landscape of ASPM mutations in primary microcephaly. *J. Med. Genet.* *46*, 249-253.
- Northcutt,R.G. and Kaas,J.H. (1995). The emergence and evolution of mammalian neocortex. *Trends Neurosci.* *18*, 373-379.

- O'Connell,J.R. and Weeks,D.E. (1998). PedCheck: a program for identification of genotype incompatibilities in linkage analysis. *Am. J. Hum. Genet.* 63, 259-266.
- Ohta,T., Essner,R., Ryu,J.H., Palazzo,R.E., Uetake,Y., and Kuriyama,R. (2002). Characterization of Cep135, a novel coiled-coil centrosomal protein involved in microtubule organization in mammalian cells. *J. Cell Biol.* 156, 87-99.
- Paddison,P.J., Caudy,A.A., Bernstein,E., Hannon,G.J., and Conklin,D.S. (2002). Short hairpin RNAs (shRNAs) induce sequence-specific silencing in mammalian cells. *Genes Dev.* 16, 948-958.
- Passemard,S., Titomanlio,L., Elmaleh,M., Afenjar,A., Alessandri,J.L., Andria,G., de Villemeur,T.B., Boespflug-Tanguy,O., Burglen,L., Del,G.E., Guimiot,F., Hyon,C., Isidor,B., Megarbane,A., Moog,U., Odent,S., Hernandez,K., Pouvreau,N., Scala,I., Schaer,M., Gressens,P., Gerard,B., and Verloes,A. (2009). Expanding the clinical and neuroradiologic phenotype of primary microcephaly due to ASPM mutations. *Neurology* 73, 962-969.
- Pattison,L., Crow,Y.J., Deeble,V.J., Jackson,A.P., Jafri,H., Rashid,Y., Roberts,E., and Woods,C.G. (2000). A fifth locus for primary autosomal recessive microcephaly maps to chromosome 1q31. *Am. J. Hum. Genet.* 67, 1578-1580.
- Pfaff,K.L., Straub,C.T., Chiang,K., Bear,D.M., Zhou,Y., and Zon,L.I. (2007). The zebra fish *cassiopeia* mutant reveals that SIL is required for mitotic spindle organization. *Mol. Cell Biol.* 27, 5887-5897.
- Pichon,B., Vankerckhove,S., Bourrouillou,G., Duprez,L., and Abramowicz,M.J. (2004). A translocation breakpoint disrupts the ASPM gene in a patient with primary microcephaly. *Eur. J. Hum. Genet.* 12, 419-421.
- Ponting,C. and Jackson,A.P. (2005). Evolution of primary microcephaly genes and the enlargement of primate brains. *Curr. Opin. Genet. Dev.* 15, 241-248.
- Ponting,C.P. (2006). A novel domain suggests a ciliary function for ASPM, a brain size determining gene. *Bioinformatics.* 22, 1031-1035.
- Rakic,P. (1995). A small step for the cell, a giant leap for mankind: a hypothesis of neocortical expansion during evolution. *Trends Neurosci.* 18, 383-388.
- Rauch,A., Thiel,C.T., Schindler,D., Wick,U., Crow,Y.J., Ekici,A.B., van Essen,A.J., Goecke,T.O., Al-Gazali,L., Chrzanowska,K.H., Zweier,C., Brunner,H.G., Becker,K., Curry,C.J., Dallapiccola,B., Devriendt,K., Dorfler,A., Kinning,E., Megarbane,A., Meinecke,P., Semple,R.K., Spranger,S., Toutain,A., Trembath,R.C., Voss,E., Wilson,L., Hennekam,R., de,Z.F., Dorr,H.G., and Reis,A. (2008). Mutations in the pericentrin (PCNT) gene cause primordial dwarfism. *Science* 319, 816-819.
- Roberts,E., Hampshire,D.J., Pattison,L., Springell,K., Jafri,H., Corry,P., Mannon,J., Rashid,Y., Crow,Y., Bond,J., and Woods,C.G. (2002). Autosomal recessive primary microcephaly: an analysis of locus heterogeneity and phenotypic variation. *J. Med. Genet.* 39, 718-721.
- Roberts,E., Jackson,A.P., Carradice,A.C., Deeble,V.J., Mannan,J., Rashid,Y., Jafri,H., McHale,D.P., Markham,A.F., Lench,N.J., and Woods,C.G. (1999). The second locus for autosomal recessive primary microcephaly (MCPH2) maps to chromosome 19q13.1-13.2. *Eur. J. Hum. Genet.* 7, 815-820.

- Rodrigues-Martins,A., Riparbelli,M., Callaini,G., Glover,D.M., and Bettencourt-Dias,M. (2008). From centriole biogenesis to cellular function: centrioles are essential for cell division at critical developmental stages. *Cell Cycle* 7, 11-16.
- Rozen,S. and Skaletsky,H. (2000). Primer3 on the WWW for general users and for biologist programmers. *Methods Mol. Biol.* 132, 365-386.
- Ruschendorf,F. and Nurnberg,P. (2005). ALOHOMORA: a tool for linkage analysis using 10K SNP array data. *Bioinformatics.* 21, 2123-2125.
- Shen,J., Eyaid,W., Mochida,G.H., Al-Moayyad,F., Bodell,A., Woods,C.G., and Walsh,C.A. (2005). ASPM mutations identified in patients with primary microcephaly and seizures. *J. Med. Genet.* 42, 725-729.
- Stoler-Poria,S., Schweiger,A., Lerman-Sagie,T., Malinger,G., and Lev,D. (2010). [Microcephaly diagnosed during pregnancy]. *Harefuah* 149, 37-40, 62.
- Tang,C.J., Fu,R.H., Wu,K.S., Hsu,W.B., and Tang,T.K. (2009). CPAP is a cell-cycle regulated protein that controls centriole length. *Nat. Cell Biol.* 11, 825-831.
- Terada,Y., Uetake,Y., and Kuriyama,R. (2003). Interaction of Aurora-A and centrosomin at the microtubule-nucleating site in *Drosophila* and mammalian cells. *J. Cell Biol.* 162, 757-763.
- Thiele,H. and Nurnberg,P. (2005). HaploPainter: a tool for drawing pedigrees with complex haplotypes. *Bioinformatics.* 21, 1730-1732.
- Thornton,G.K. and Woods,C.G. (2009). Primary microcephaly: do all roads lead to Rome? *Trends Genet.* 25, 501-510.
- Tolmie,J.L., McNay,M., Stephenson,J.B., Doyle,D., and Connor,J.M. (1987). Microcephaly: genetic counselling and antenatal diagnosis after the birth of an affected child. *Am. J. Med. Genet.* 27, 583-594.
- Trimborn,M., Bell,S.M., Felix,C., Rashid,Y., Jafri,H., Griffiths,P.D., Neumann,L.M., Krebs,A., Reis,A., Sperling,K., Neitzel,H., and Jackson,A.P. (2004). Mutations in microcephalin cause aberrant regulation of chromosome condensation. *Am. J. Hum. Genet.* 75, 261-266.
- Trimborn,M., Richter,R., Sternberg,N., Gavvovidis,I., Schindler,D., Jackson,A.P., Prott,E.C., Sperling,K., Gillissen-Kaesbach,G., and 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, 496.
- Van den,B. (1959). Microcephaly in the Netherlands: a clinical and genetical study. *Ann. Hum. Genet.* 23, 91-116.
- Varmark,H., Llamazares,S., Rebollo,E., Lange,B., Reina,J., Schwarz,H., and Gonzalez,C. (2007). Asterless is a centriolar protein required for centrosome function and embryo development in *Drosophila*. *Curr. Biol.* 17, 1735-1745.
- Woods,C.G. (2004). Human microcephaly. *Curr. Opin. Neurobiol.* 14, 112-117.
- Woods,C.G., Bond,J., and Enard,W. (2005). Autosomal recessive primary microcephaly (MCPH): a review of clinical, molecular, and evolutionary findings. *Am. J. Hum. Genet.* 76, 717-728.

- Xu,X., Lee,J., and Stern,D.F. (2004). Microcephalin is a DNA damage response protein involved in regulation of CHK1 and BRCA1. *J. Biol. Chem.* 279, 34091-34094.
- Yang,S.Z., Lin,F.T., and Lin,W.C. (2008). MCPH1/BRIT1 cooperates with E2F1 in the activation of checkpoint, DNA repair and apoptosis. *EMBO Rep.* 9, 907-915.
- Yngvadottir,B., Xue,Y., Searle,S., Hunt,S., Delgado,M., Morrison,J., Whittaker,P., Deloukas,P., and Tyler-Smith,C. (2009). A genome-wide survey of the prevalence and evolutionary forces acting on human nonsense SNPs. *Am. J. Hum. Genet.* 84, 224-234.
- Yu,T.W., Mochida,G.H., Tischfield,D.J., Sgaier,S.K., Flores-Sarnat,L., Sergi,C.M., Topcu,M., McDonald,M.T., Barry,B.J., Felie,J.M., Sunu,C., Dobyns,W.B., Folkerth,R.D., Barkovich,A.J., and Walsh,C.A. (2010). Mutations in WDR62, encoding a centrosome-associated protein, cause microcephaly with simplified gyri and abnormal cortical architecture. *Nat. Genet.* 42, 1015-1020.
- Zhang,J. and Megraw,T.L. (2007). Proper recruitment of gamma-tubulin and D-TACC/Msps to embryonic *Drosophila* centrosomes requires Centrosomin Motif 1. *Mol. Biol. Cell* 18, 4037-4049.
- Zhong,X., Liu,L., Zhao,A., Pfeifer,G.P., and Xu,X. (2005). The abnormal spindle-like, microcephaly-associated (ASPM) gene encodes a centrosomal protein. *Cell Cycle* 4, 1227-1229.
- Zhong,X., Pfeifer,G.P., and Xu,X. (2006). Microcephalin encodes a centrosomal protein. *Cell Cycle* 5, 457-458.

## 9. Abbreviations

Asp	Abnormal spindle gene
<i>ASPM</i>	Abnormal Spindle-like Microcephaly-associated
Asl	Asterless
APS	Adenosine 5' Phosphosulfate
BRCT	Breast cancer 1(BRCA1) C-terminal
<i>CDK5RAP2</i>	Cyclin Dependent Kinase 5 Regulatory Associated Protein 2
<i>CDK6</i>	Cyclin-dependent Kinase 6
<i>CEP135</i>	Centrosomal protein 135 kDa
<i>CENPJ</i>	Centromere Protein J
<i>CEP152</i>	Centrosomal Protein of 152 kDa
<i>CPAP</i>	Centrosomal protein 4.1- associated protein
Cnn	Centrosomin
CT	Computerized tomography
DAPI	4', 6-Diamidino-2'-phenylindole
DMEM	Dulbecco's Modified Eagle's Medium
ECL	Enhanced chemiluminescence
FBS	Fetal bovine serum
HC	Head circumference
IPCs	Intermediate progenitor cells
IQ	Isoleucine-Glutamine
LB	Luria Bertani
LMW	Low Molecular Weight
LOD	Logarithm of the odds
LTK	Leukocyte receptor tyrosine kinase
MCPH	Autosomal recessive primary microcephaly
MBD	Microtubule-binding domain
MRI	Magnetic resonance imaging
MCPH1	Autosomal recessive primary microcephaly locus 1
MCPH2	Autosomal recessive primary microcephaly locus 2
MCPH3	Autosomal recessive primary microcephaly locus 3
MCPH4	Autosomal recessive primary microcephaly locus 4
MCPH5	Autosomal recessive primary microcephaly locus 5
MCPH6	Autosomal recessive primary microcephaly locus 6
MCPH7	Autosomal recessive primary microcephaly locus 7
MCPH8	Autosomal recessive primary microcephaly locus 8

MCPH9	Autosomal recessive primary microcephaly locus 9
MCPH10	Autosomal recessive primary microcephaly locus 10
MCPH11	Autosomal recessive primary microcephaly locus 11
MCPH12	Autosomal recessive primary microcephaly locus 12
MDD	Microtubule-destabilising domain
NCBI	National Center for Biotechnology Information
NE	Neuroepithelial
NMR	Nuclear magnetic resonance
ORF	Open Reading Frame
PBS	Phosphate buffered saline
PCR	Polymerase chain reaction
PCM	Pericentriolar matrix
PCNT	Pericentrin
PIC	Proteinase Inhibitor Cocktail
Plk4	Polo-like-kinase 4
pRB	Retinoblastoma tumor suppressor protein
RT-PCR	Reverse transcriptase PCR
SAP	Shrimp Alkaline Phosphatase
SD	Standard deviation
SDS-PAGE	SDS-polyacrylamide gel electrophoresis
shRNA	short hairpin RNA
STR	Short tandem repeat
SVZ	Subventricular zone
STIL/SIL	SCL/TAL1 interrupting locus
TBS	Tris-buffered saline
VZ	Ventricular zone
<i>WDR62</i>	WD repeat domain 62

## Appendices

### Appendix 1

Microsatellite markers spanning the identified interval of six MCPH loci with additional information of amplification condition, fluorescent dyes, repeat unit size and physical positions. Physical positions in term of base pair is taken from NCBI Human Genome Build 36.3 and in term of Centimorgan (cM) are taken from Marshfield genetic map. In marker M19SH1, M stands for microsatellite, 19 is chromosome number, SH is the name of marker developer (in this case SH means Sajid Hussain) and 1 stands for 1st primer of this region.

Locus	Marker	PCR Prog	Dye	Repeat unit	PCR product size (bp)	Position (cM)	Physical position (bp)
<b>MCPH1</b>	D8S518	MS	Fam	Di	229-253	5.6	4475013-4475263
	D8S1798	MS	Tet	Di	145-165	6.7	5090694-5091016
	D8S1742	MS	Fam	Di	130-150	7.7	6201415-6201550
	D8S277	MS	Fam	Di	148-180	8.34	6504133-6504278
	D8S1819	MS	Fam	Di	207-223	10.0	6737377-6737597
	D8S1825	MS	Hex	Di	127-143	15.4	8962276-8962408
<b>MCPH2</b>	D19S226	MS	Ned	Di	191-233	42.3	14494401-14494643
	D19S917	MS	Tet	Di	216-240	43.3	16222045-16222265
	D19S49	MS	Hex	Di	102-126	50.8	26783346-26783462
	D19S414	MS	Hex	Di	125-171	54.0	36606611-36606773
	D19S416	MS	Fam	Di	165-185	58.7	38760490-38760654
	D19S425	MS	Fam	Di	252-280	59.4	40185772-40186036
	M19SH1	TDM	Fam	Di	464	60.0	41038399-41038449
	M19SH12	MS	Fam	Di	295	61.0	41130340-41130382
	M19SH14	MS	Hex	Di	293	61.0	41166101-41166127
	D19S224	MS	Tet	Di	240-262	61.5	41219912-41220169
	M19SH2	MS	Hex	Di	337	61.5	41220086-41220124
	D19S896	MS	Fam	Di	194-220	62.0	42170482-42170679
	M19SH9	MS	Fam	Di	237	62.0	42195509-42195555
	D19S570	MS	Fam	Di	186-210	62.0	42419583-42419782
	D19S420	MS	Fam	Di	251-267	66.3	48500639-48500901
	<b>MCPH3</b>	D9S302	TDM	Fam	Tetra	232-316	123.3
D9S934		MS	Hex	Tetra	206-230	128	120135575-120135798
D9S1872		TDM	Fam	Di	111-135	129.7	120829322-120829427
D9S1850		MS	Ned	Di	218-230	132.1	122490084-122490307

	D9S1682	MS	Hex	Di	200-208	132.1	124033006-124033207
	D9S1881	MS	Fam	Di	220-236	135.9	126019300-126019529
	D9S1825	MS	Ned	Di	127-145	136.5	126927953-126928083
	D9S1821	TDM	Hex	Di	168-174	137.4	128369402-128369573
	D9S1819	TDM		Di	228-236	136.5	129241188-129241414
<b>MCPH4</b>	D15S1042	MS	Ned	Di	177-202	32.6	34050903-34051090
	D15S1012	MS	Fam	Di	160-174	36.0	36794835-36795008
	D15S1044	MS	Fam	Di	183-195	39.7	37456267-37456453
	D15S146	MS	Tet	Di	217-227	39.7	37911344-37911564
	D15S968	MS	Tet	Di	204-235	40.3	38432560-38432785
	D15S641	MS	Fam	Tri	90-114	39.7	39631223-39631327
	D15S537	MS	Hex	Tetra	185	40.3	42602780-42602964
	D15S659	MS	Hex	Tetra	166-210	43.5	44161300-44161483
	D15S1006	MS	Fam	Di	208-228	44.9	45334255-45334468
	D15S126	TDM	Fam	Di	188-218	45.6	47089563-47089750
	D15S1016	MS	Hex	Di	239-267	47.3	51320121-51320369
	D15S962	MS	Fam	Di	282-294	47.9	54362360-54362643
	D15S648	MS	Hex	Tri	242-260	47.9	55165601-55165839
<b>MCPH5</b>	D1S2816	MS	Fam	Di	248-252	211.1	194917168-194917415
	D1S1660	MS	Fam	Tetra	248-252	212.4	196877970-196878197
	D1S1660	MS	Fam	Di	165-189	214.8	198381195-198381375
	D1S373	MS	Hex	Tetra	283-330	214.0	198521109-198521419
	D1S1723	MS	Hex	Di	167-181	215.1	199657511-199657681
	D1S2655	MS	Fam	Di	224-260	216.8	200831893-200832140
<b>MCPH6</b>	D13S1275	MS	Ned	Di	198-214	7cM	
	D13S787	MS	Hex	Tetra	231-271	8.87	23278738-23278989
	D13S742	MS	Hex	Tetra	364	10.7	24180814-24181171
	D13S283	MS	Fam	Di	128-153	11.5	24498985-24499113
	D13S1294	MS	Hex	Di	247-273	12.9	25374909-25375159
	D13S221	MS	Ned	Di	318-366	12.91	25474866-25475100
	D13S1304	MS	Hex	Di	149-165	13.45	26268064-26268222
	D13S1244	MS	Fam	Di	92-108	15.19	27144977-27145072

## Appendix 2

MCPH1 Sequencing Oligo used to sequence *MCPH1* gene in MCP61 microcephaly family.

<b>MCPH1 Sequencing Oligo</b>	
MCPH1-Ex1F	CTCCCGCCTCACCTACAGAG
MCPH1-Ex1R	AGCAGTACGGGGGAGGAAG
MCPH1-Ex2F	GTCTCAAACCCCTGACTTCG
MCPH1-Ex2R	TCCTCCTCCTCACTCAATGC
MCPH1-Ex3F	TTCACCGTTGTAAGTGAACAG
MCPH1-Ex3R	TCCCCCTAGGCAGAGTTAGG
MCPH1-Ex4F	GGGAAGTTTGATTTATACTGACTTTTG
MCPH1-Ex4R	ACCACAGGCTTTTCCATTTC
MCPH1-Ex5F	TTGCCAGTTCACATACAGTGC
MCPH1-Ex5R	TATTATGGCTCCCAGCCAAG
MCPH1-Ex6F	CAACATCACTGCCTGTGGAG
MCPH1-Ex6R	CAAAGCCAGCCATGAAATAAG
MCPH1-Ex7F	CAGGCAAGTTGACTTTAAGATCC
MCPH1-Ex7R	GAGCAATGGATTCTGTAGCAAG
MCPH1-Ex8-1F	CCCTTAAGTGGAATGAGAAGAAC
MCPH1-Ex8-1R	CAAACGATACTTCTCTTCAAACG
MCPH1-Ex8-2F	TCATCACCATCTTTCCTCACC
MCPH1-Ex8-2R	AACTGAGCAGGGCTTGATG
MCPH1-Ex8-3F	CTGGAGGCTCTTAGCTGTGG
MCPH1-Ex8-3R	CTTTTCAGACCAACCGCTTC
MCPH1-Ex8-4F	TTCGAGTTGCGTGACTTCTG
MCPH1-Ex8-4R	TCATTGGAATGATCAAATGCAC
MCPH1-Ex9F	TTTATTTAAAAGGCCTATAACTTCCTG
MCPH1-Ex9R	GCAAACAACGCTTTCAGTTTC
MCPH1-Ex10F	GCTTTGGGGACAGTATCTGAG
MCPH1-Ex10R	CTGCCTAAACAACCCAGAG
MCPH1-Ex11F	TTTATTTCCCAGGTTTCAAAG
MCPH1-Ex11R	GACCAACAGGAGGAAAGACG
MCPH1-Ex12F	TTCATAACATTTATGCAACATGAAG
MCPH1-Ex12R	AAAAGGTGTTAAGTTCTGTGAATGTC

MCPH1-Ex13F	TGTCATCATCTTCTCTGGATTCTC
MCPH1-Ex13R	CGTCTGCTAACAGCAAGGAG
MCPH1-Ex14F	TTGTATTGAATTTTCGTTTCACG
MCPH1-Ex14R	TCTTAAGAACCACAACACATGG

### Appendix 3

Sequencing oligonucleotides used to find pathogenic variant in MCPH5 (*ASPM*) linked families.

<b>ASPM Sequencing Oligos</b>	
ASPM-Ex1-1-F	GATTCCGCTACCGTTCTCAC
ASPM-Ex1-1-R	AAGGAGACCTGCAGAAGTGG
ASPM-Ex1-2-F	GAATTCACCTCCACGACCTC
ASPM-Ex1-2-R	AAGAGCCACCCACAGTTATTG
ASPM-Ex2-F	TTCTTATAATTAAGCAGATAGGGTAGG
ASPM-Ex2-R	GCAATATTACTTCAAGCCTGTTATC
ASPM-Ex3-1-F	ATGTTTTAACCATTCTGTGATTTACTC
ASPM-Ex3-1-R	ACGTACAGAGAGTGGCAAGC
ASPM-Ex3-2-F	CCCCCAACAGAAAACAATTC
ASPM-Ex3-2-R	GGGATAAAATAGGATTAAGTACTGACTC
ASPM-Ex3-3-F	AGCCTGTGCATTTGGAATC
ASPM-Ex3-3-R	ACCCACTGCACTGTTGAGAC
ASPM-Ex3-4-F	GAAGGCCACCTGTACCAGAG
ASPM-Ex3-4-R	AGGAAATGTACCCAGCAAATAAG
ASPM-Ex4-F	TCATAGCATTAGTGCGTGGAG
ASPM-Ex4-R	TTCTTCCAGGCTGTTATTCAAC
ASPM-Ex5-F	TTCAGTGTTTTAAAGATGGTATTGC
ASPM-Ex5-R	GCTAATGAACAGGGAATTATGC
ASPM-Ex6-F	TGAAATTGCATTTTATTGCTG
ASPM-Ex6-R	AAAAACACAAAATCTCTTGAATG
ASPM-Ex7-F	TTTCCCACTGATATACTCTCCTTG
ASPM-Ex7-R	AAAAGATGAATCAAGCTAACCTAATG
ASPM-Ex8-F	TTATGGTCTGCCATTCCTCAG
ASPM-Ex8-R	GGAATGAGGGTGGAGGAAG
ASPM-Ex9-F	TTGCTACCCTACACTTTTTGTTTTAC

ASPM-Ex9-R	GATGGGACTCACCAGACAGG
ASPM-Ex10-F	CAGAATGATTTGGAGGATTTG
ASPM-Ex10-R	GACATAACATTGATGTACCACTTCC
ASPM-Ex11-F	ATGGAGCAACAGAGTTTTGAG
ASPM-Ex11-R	AAAAGAAAAGGTTGTCCATTAGC
ASPM-Ex12-F	AGAGGTAAACTTATTCTCCTCACTG
ASPM-Ex12-R	GTTACTGGGGCAAAATAAACC
ASPM-Ex13-F	TCCATTTTCAGGCACTTTATTTTC
ASPM-Ex13-R	AAGAAATTCAAGAAAAGACTTCAGG
ASPM-Ex14-F	CTGCCTTTATTATTCACAGGTATTC
ASPM-Ex14-R	CTATTATGCAGGAGGAAAGGAG
ASPM-Ex15-F	TCCCTTTTGGTTACTTTTCC
ASPM-Ex15-R	AAAAACCCACAAAAGATAAAAAC
ASPM-Ex16-F	AATTCATAATAAAGTCTCAGAAGATG
ASPM-Ex16-R	ATACTCATACCTCCCCAACC
ASPM-Ex17-F	TTCAAAGATGAACACAAGAAG
ASPM-Ex17-R	ACCATAACGAGCTTTTTCAGG
ASPM-Ex18-1-F	TGAATTGGCTACAGGTATATCAATG
ASPM-Ex18-1-R	CTTGTGATTGCATTTTACGTTG
ASPM-Ex18-2-F	GTTACAATTCAGAGGCATTGG
ASPM-Ex18-2-R	CGCTCAATCTTTCCTTTCAG
ASPM-Ex18-3-F	AGAAAAGATTTTCGGTGCTTTC
ASPM-Ex18-3-R	AGACTGCAGCACAAATGACAG
ASPM-Ex18-4-F	TCATTTCCGAGCTTATATTTTTG
ASPM-Ex18-4-R	AATACTGCCGAGCCTTTCTC
ASPM-Ex18-5-F	AGGATACCTTGTCCGAAAGC
ASPM-Ex18-5-R	AAATCACAGCTGCCTTTGTC
ASPM-Ex18-6-F	ACGCCAGCTAATCAAACAAC
ASPM-Ex18-6-R	CTTTTGTTGCACATGCATTC
ASPM-Ex18-7-F	TGTGGAAGGGAAAAACACTGAG
ASPM-Ex18-7-R	ATAAAAGTGGCTGCCCTGTG
ASPM-Ex18-8-F	AAGACAGCAATTAATCCAATCTG
ASPM-Ex18-8-R	GAATGAGAGTTGCGGCTATATG
ASPM-Ex18-9-F	CAGCAAAGATACTGGGCAATG
ASPM-Ex18-9-R	ATGCTGCCTCTGCAGTTTTG
ASPM-Ex18-10-F	GGGCTGCTACTTTCATCCAG

ASPM-Ex18-10-R	GCATGTTTCCAAGTCTGAAATG
ASPM-Ex18-11-F	AAGGGCTGCAGTTCTCATTG
ASPM-Ex18-11-R	AATAATAATGGCAGCCTGGTG
ASPM-Ex18-12-F	CTTGTGTTTCAGGCAGGTTTTG
ASPM-Ex18-12-R	GGCTGCCATCTTTTCCTCTG
ASPM-Ex18-13-F	GCTGCTTTTAGAGGCATGAAAG
ASPM-Ex18-13-R	TTGGTCAAAGAAAGACTCACAG
ASPM-Ex19-F	ATGCAATTTATTTTTGTCTGCAC
ASPM-Ex19-R	CAGTACGAGAGATGTATTTTTGTTCC
ASPM-Ex20-F	ACTTCTTTCGTGTGCGTGTG
ASPM-Ex20-R	TGTGTGAAATAAATGCATACTTAGGTC
ASPM-Ex21-F	TGAAACATGGCATTCTTAGAC
ASPM-Ex21-R	TCAGTGCTCTTGTCACTTACC
ASPM-Ex22-F	TGATTAAGTCTAGGAGGCAGATACTTG
ASPM-Ex22-R	AACCCTCAAATCATTGTAACAGC
ASPM-Ex23-F	TGGACTGGAAAGGTTTTTGC
ASPM-Ex23-R	TGAGTTATTCTACCGGCTAATGC
ASPM-Ex24-25-F	GATGTAGATATAAATAGAAAACATTGG
ASPM-Ex24-25-R	CAGGGGCATATTTGTTGAC
ASPM-Ex26-F	AAAGTCCTTTCGACTTGCTG
ASPM-Ex26-R	TTTATCCGTGCAAAAAGCAG
ASPM-Ex27-F	GCGACAGAGCAAGAGAGACC
ASPM-Ex27-R	TGTTGTTTCTCCACTGAAAAGC
ASPM-Ex28-F	AAGGTGAAGGTAAGTCTAAAGAGAAG
ASPM-Ex28-R	TTTTATGAAAATTCCTCAACTTACC

#### Appendix 4

Sequencing oligonucleotides used to find pathogenic variant in MCPH6 (*CENPJ*) linked family.

<b>CENPJ Sequencing Oligos</b>	
CENPJ-Ex1-F	CGCCTACGTGACCACTG
CENPJ-Ex1-R	GAACGAAGCCACTGAACTGC
CENPJ-Ex2-F	GCCCAATTGCTTTGCTAGTT
CENPJ-Ex2-R	TTTCAGAGTGAAGGGGAAAAA
CENPJ-Ex3-F	TTTTAGATTTTCGGTGGCTGA

CENPJ-Ex3-R	GCAGCTCCTTCTTGTTCCAG
CENPJ-Ex4-F	TGGTTTTGTTAAACACCATTTCC
CENPJ-Ex4-R	AACACGGCAACACTCCATCT
CENPJ-Ex5-F	CCGGCCTCAAGTGTCTTAAT
CENPJ-Ex5-R	GGCAGCTTGACATATCTCCTC
CENPJ-Ex6-7F	AGGTGTGGAGAGGGATTGTG
CENPJ-Ex6-7R	CCGCTGGAGTTGCTGTCTAT
CENPJ-Ex7b-F	AACAACCATTTTTAAAACGAGGAG
CENPJ-Ex7b-R	TCCTTTTCCTTTTCTCGTTCC
CENPJ-Ex7c-F	AAGAGCCCCATAAGGGAGAC
CENPJ-Ex7c-R	TGTGGGCTATCCTCTCTGCT
CENPJ-Ex7d-F	GCACTGACTCTGAGGAACAGC
CENPJ-Ex7d-R	TTGAGCAATAAGGAAAGTTAGGATTT
CENPJ-Ex8-F	AAATTCCCCTGATTTCTCA
CENPJ-Ex8-R	TTGACTACCCTGGGTCATTTT
CENPJ-Ex9-F	GTAGGCACCCAATGAATGCT
CENPJ-Ex9-R	TTGCTGAATGAATGAATAATGGA
CENPJ-Ex10-F	TCGTCTTAAATGTCGTCTGTTGA
CENPJ-Ex10-R	TTCACAGTCACTCCCCACTG
CENPJ-Ex11-F	CTTGTTTGCTTTGGCCTGTT
CENPJ-Ex11-R	GGCGGAATAAGATGATGCAC
CENPJ-Ex12-13-F	CCTGTGTCGGAGAACTACAAA
CENPJ-Ex12-13-R	TCTGAACGAGAAATGGCAAC
CENPJ-Ex14-15-16-F	CGAGTTTGCTACTCCTCACG
CENPJ-Ex14-15-16-R	AATATGGGTGTCACATATCATCAG
CENPJ-Ex17a-F	GCAGAAACACTGTAGAAAGTTACAA
CENPJ-Ex17a-R	AAAAATGAAGATGCCACAGG
CENPJ-Ex17b-F	TGTACATTTACTGTGGATTCTGTTTA
CENPJ-Ex17b-R	CCATGATTACTACCTGTTGTGTG

## Appendix 5

Sequencing oligonucleotides used to find pathogenic variant in MCPH8 (*CDK6*) linked family.

<b>CDK6 Sequencing Oligos</b>	
CDK6-Ex1F	GCACACTCTGGGCTGCTC
CDK6-Ex1R	TACGAAGCCTCCATCGCTAC
CDK6-Ex2-1F	AAAGGAGGAGGAGGGGAGTC
CDK6-Ex2-1R	GGGGTGCCTCAACTAGC
CDK6-Ex2-2F	GGGGAGCAAGTAAAGCTAGACC
CDK6-Ex2-2R	TTTCTGGGCCTGAGGATTC
CDK6-Ex3F	TTTTCTATTATGAGACCAAATTGAGC
CDK6-Ex3R	GGCAATTTTCATAAGGAAAGAAAG
CDK6-Ex4F	GCAGCCACAAAATGCTAAGG
CDK6-Ex4R	GAGACATGGAAGAGGGACAGAC
CDK6-Ex5F	GGAGTTCTTTCTGAACTACATCTTGTC
CDK6-Ex5R	AATGCCATGCTGCCACTC
CDK6-Ex6F	TTTTGCTTCCTGTGAAACAATG
CDK6-Ex6R	GCATGTCAGAGGAAAGTCACTG
CDK6-Ex7F	TTTGCCTTTCTCTGGAAATTG
CDK6-Ex7R	ACTGATATTTGTGCCACCAC
CDK6-Ex8F	CCCCTGTAGATGCGCTTAAC
CDK6-Ex8R	TGAACATAAAGCTGCAATCACTC
CDK6-EX8-2F	CCTTCCAGCTGCTGTCTTC
CDK6-EX8-2R	TTCTTGTCGTTTACAAAAGTAACAC
CDK6-EX8-3F	AGTGGGTAATCCAGGCACAG
CDK6-EX8-3R	GGGAAGAAAGGAATTTCAAACC
CDK6-EX8-4F	TCTTTTGGAAAATAGCCTGTAAGC
CDK6-EX8-4R	CCTGTTAGGTTTGCAGAATCG
CDK6-EX8-5F	TGTCTATCTCCCGGCACTTC
CDK6-EX8-5R	GCCAACTGAGCACTAAACCAC
CDK6-EX8-6F	GCCTGGGCATTGTAAACAG
CDK6-EX8-6R	ACCTGACCTTGAATGCTGTG
CDK6-EX8-7F	GGAGATGGGCCATACAAATG
CDK6-EX8-7R	TTCACTCAAATGCAACAACACTACAC
CDK6-EX8-8F	GCACAAACATTCTGGCATTTC
CDK6-EX8-8R	CCTTTGCCATGGGATTTTC

CDK6-EX8-9F	TGTTTGAATTTTCATAGTAGGACCAG
CDK6-EX8-9R	TCAAGATGGATAGTACATTCTGGAG
CDK6-EX8-10F	TGAAAGCTTTAGAAAGGCATGAG
CDK6-EX8-10R	TTTCTTTCCAGGGTAAATATTCTGTAG
CDK6-EX8-11F	TCCAAAAGCAGGCTTTAATTG
CDK6-EX8-11R	GTGGCATAGGAAACACATGC
CDK6-EX8-12F	AGAAGCTGTCCCAAGCAAAC
CDK6-EX8-12R	AAACATTTTGAAAAGCTGATTGG
CDK6-EX8-13F	CCAAAATAATCCTTTGGTCTCTC
CDK6-EX8-13R	GCGCAGAAAATAAAGTTAGAATG
CDK6-EX8-14F	AGGCTCTTTTCGAGCATCTG
CDK6-EX8-14R	AGGGAATTTGAAAGCAGCAG
CDK6-EX8-15F	AAAATTTTGAAGTACCCAGTTACC
CDK6-EX8-15R	AAAATACTGCAATATCCTTCCCTAC
CDK6-EX8-16F	TGAGCATAACAACCTTAGGAGTGTG
CDK6-EX8-16R	GTTTTGAGCACGAGCAAATG
CDK6-EX8-17F	TCCAGTTAAAGCCCCTATGC
CDK6-EX8-17R	CCAGGCATATCTTTCACCATC
CDK6-EX8-18F	CTGTCTACCCCGCATCTCTC
CDK6-EX8-18R	CGGAATTTTTCTAGGGACCAC
CDK6-EX8-19F	TCAAATGTAGGTTTTTCTCAAAGG
CDK6-EX8-19R	TTTCACAGGACTAGTGTGATTCAG
CDK6-EX8-20F	CTACCCCCAGGAACAGCAC
CDK6-EX8-20R	TACGAGGGCACTACATCACG
CDK6-EX8-21F	CTTAATTTAGAACTGTGCCAGGTG
CDK6-EX8-21R	CCCTATGGCAATGCAGAAAC
CDK6-EX8-22F	GGTTAAAGGGCAAGTGTTTGG
CDK6-EX8-22R	CCAGCATCAGGAACCATCTC
CDK6-EX8-23F	TGCTCTCCATCTTCTCCAC
CDK6-EX8-23R	CAGGCCACTGTGGTAACTCTC
CDK6-EX8-24F	GTTCTGCAGACCCAAGAAGC
CDK6-EX8-24R	AAACCCTGGGTCACAAAGC
CDK6-EX8-25F	CCTTCCTATCTGAGGAGGAGAATAG
CDK6-EX8-25R	AAAAGAAAATAGCCAGGAGAGTAATTC
CDK6-EX8-26F	TTGGTGTGAAATATCACTGTCC
CDK6-EX8-26R	CATGCCCGGCTAGTTTTTC

## Appendix 6

Oligonucleotide sequences used to knockdown the expression of CDK6.

CDK6-shRNA-1-F

TCACCAGAATGTTCTGTGGTTTTAGATCGAAGCTTGGATTTAAAACCATAGAACATTC  
TGGTGGCCATTTTTT

CDK6-shRNA-1-R

GATCAAAAAATGGCCACCAGAATGTTCTATGGTTTTAAATCCAAGCTTCGATCTAAAA  
CCACAGAACATTCTGGTGACG

CDK6-shRNA-2-F

TTCATCGATATCTGTTACAACTTCTCAGAAGCTTGTGAGAAGTTTGTGACAGATATT  
GATGAGCTATTTTTT

CDK6-shRNA-2-R

GATCAAAAAATAGCTCATCAATATCTGTCACAACTTCTCACAAGCTTCTGAGAAGTT  
TGTAACAGATATCGATGAACG

CDK6-shRNA-C-F

TTCCATTAGTTAACTAGCTCTTTTAAGAAGCTTGTTAGAAGAGTTAGTTTAATTAAT  
GGAACTGTTTTT

CDK6-shRNA-C-R

GATCAAAAAACAGTTCCCATTAATTAACCTAACTCTTCTAACAAGCTTCTTAAAAGAG  
CTAGTTTAACTAATGGGAACG

## Appendix 7

Sequencing oligonucleotides used to find pathogenic variant in MCPH9 (*CEP135*) linked family.

<b>CEP135 sequencing Oligos</b>	
CEP135-Ex1F	CTGAGGTTTGGCATTGCTG
CEP135-Ex1R	GCCTGGCCTCACTTCCTC
CEP135-Ex2F	CCAGAAGGCTACCTACCTGAAG
CEP135-Ex2R	AGAGTTCTGCCGTGTGGAAG
CEP135-Ex3F	CAGCTGCCCCCATTATAAATC
CEP135-Ex3R	CACACCATCCAGCTGTAGGAC
CEP135-Ex4F	TGAAGATTCTCTTTTGTGCTTAGTG
CEP135-Ex4R	GCATGAAAATCCAATCCAAAG
CEP135-Ex5F	TTGATAACCATGTCTTGTTGAGG
CEP135-Ex5R	AGTGCCGGTGCAAAGTTC
CEP135-Ex6F	ATTGATTTGGGGACCCTTTG
CEP135-Ex6R	AAAGCCACCTAAGCCATTCC

CEP135-Ex7F	TGAATGTTAATCTCTTGCATACATTG
CEP135-Ex7R	TTCTATCAAATTTTTCCCATTGC
CEP135-Ex8F	CACTGCACATGGTCAGTTTTG
CEP135-Ex8R	CACTTCAGAAAACCTGAATTACC
CEP135-Ex9F	TCACTGGTAAGAGGGAACGTG
CEP135-Ex9R	CATGTTGATGTGGCTGGATG
CEP135-Ex10F	CCACCCTCAAAGAAGCATTG
CEP135-Ex10R	TGGTTTTTACTAATGGTAAGCAC
CEP135-Ex11F	CCTCTTCAATAGATTTGGTACCTACAG
CEP135-Ex11R	CAGGATTTTTCAAAGCAAGTACG
CEP135-Ex12F	AAAGGCACTTGTGTTCTTTTTG
CEP135-Ex12R	CATAGATAGGCTAGAAAAGGGAAAG
CEP135-Ex13F	GCGAAGAGAGTGACGGTAGC
CEP135-Ex13R	GGCCATACATAAACAAAGTGGTC
CEP135-Ex14F	TGTAAAAGATGAGAACGTATTGTGC
CEP135-Ex14R	GCTTGAAGCTGGAGTTGGAG
CEP135-Ex15F	CACCCATTCTGACTCTTAGAACC
CEP135-Ex15R	CAAAGCTCCCAGGTGAGTC
CEP135-Ex16F	AGATGGAACCATTTTTGGAGAAG
CEP135-Ex16R	ATGAGCCTCCTCATCCTGTG
CEP135-Ex17F	GAACTTGAATCAGCCCAAGC
CEP135-Ex17R	CCAAGAGATAGGGGTTGCAG
CEP135-Ex18F	AAGGGTTGGTCAAACATGC
CEP135-Ex18R	CAAGCTGGAGACTGAGAAAATG
CEP135-Ex19F	TTAATAGGTCTCTTAATGCAGCAG
CEP135-Ex19R	AATACCCTGATAAACACCATTGTC
CEP135-Ex20F	GCTGTGATCACGCCATTG
CEP135-Ex20R	TGCAGGTACCACACTAAAATCTATG
CEP135-Ex21F	TGCCTTTTTCTATTTTTCCCATAG
CEP135-Ex21R	AATTGGCCAGGTGTGGTG
CEP135-Ex22F	AAATAGTAGGCAGCTTCCTGAGAG
CEP135-Ex22R	AGATGCCTGAGGGAGAATTG
CEP135-Ex23F	CCCAAATGAAACAAGTGATAGG
CEP135-Ex23R	TGAGCTCAGGCAATCTACCC
CEP135-Ex24F	GAGAATTCTTGTTGAAAGGTCTTG
CEP135-Ex24R	GCTATTATCTTCCCCAACACTTG

CEP135-Ex25F	AGAAGCAAGATGCTCCTCAC
CEP135-Ex25R	TATACCCACATAACTACTATCACCAAG

### Appendix 8

Oligonucleotide sequences used to screen the control individuals of Pakistani population for variant (c.970delC p.Gln324Serfs\*2) in *CEP135* with pyrosequencing.

CEP135-Ex8-DelC-F	BIOTIN-TGGAAACCAAGGAAACAGTGACAT
CEP135-Ex8-DelC-R	TTTGCTTCCCCAAGCTCTTTATC
CEP135-Ex8-DelC-S	CTTTCCAAGTCTGTG

### Appendix 9

Oligonucleotide sequences used to screen the control individuals of Pakistani population for variant (c.2437G>T p.E813\*) in *LTK* with pyrosequencing.

LTK-ex20.GT-F	BIOTIN-CTGGGCTGGGGAACAGAT
LTK-ex20.GT-R	ACTCAGTTCCTGGGGCTGTG
LTK-ex20.GT-S	GGGGTCTTAGGCACT

### Appendix 10

Sequencing oligonucleotides used to find pathogenic variants in *MCPH2* (*WDR62*) linked families.

WDR62-AMP.1-F	GCCCCATTGGTTCTAAGC
WDR62-AMP.2-F	GGGTGTTGAATGTAGCAGGAC
WDR62-AMP.3-F	ATTTGTGGGCTTTTCTGGTG
WDR62-AMP.4-F	AGGCAGAAGAGGGCAGAGTC
WDR62-AMP.5-F	GACTTGGAGTGGGGACGAG
WDR62-AMP.6-F	CCATTTGGAAGAGCTTCTTAGAG
WDR62-AMP.7-F	TGGGTGTCTTAGACCCAAGG
WDR62-AMP.8-F	GATGGATGGCCACAAATACC
WDR62-AMP.9-F	TTGCACGATACCTGTTTATCTTTAAC
WDR62-AMP.10-F	CTCTGCATCACTGCCTCTTG
WDR62-AMP.11-F	ATTGGTGAAACTCCCAGCTC

WDR62-AMP.12-F	TTCAACACGGCTCCCTTTAC
WDR62-AMP.13-F	GAACCAAAGACCCCTTCTCC
WDR62-AMP.14-F	TTGATTGATGGCTCTTGCAG
WDR62-AMP.15-F	TCAGCTTGAGATGGGGTTTC
WDR62-AMP.16-F	AGAGTCCCATGTGCTGTGC
WDR62-AMP.17-F	CAGGAGCATCTGGAGAATGG
WDR62-AMP.18-F	CACATGGAGCTTAGCACAGG
WDR62-AMP.19-F	CTTCTGACTTCTGGGTGTGG
WDR62-AMP.20-F	CAGATGGGCTGTGTGAAATG
WDR62-AMP.21-F	CTGCTGGCATGGTTCCTG
WDR62-AMP.22-F	AGGGGGCATTGGAGGAG
WDR62-AMP.23-F	TACCATCCCTCCTCCAGATG
WDR62-AMP.24-F	CCCACTGAGCCTGGAAGAC
WDR62-AMP.25-F	CTGCTCGTACCCTGTGTCTG
WDR62-AMP.26-F	CCGGATGTCCAGCCTAGC
WDR62-AMP.27-F	GGGACCCAGAGAAGTGTTC
WDR62-AMP.1-R	GAACAACCTGAGGGGACTGG
WDR62-AMP.2-R	GAGCTGGGACACAAATCCAG
WDR62-AMP.3-R	CGTCTCCTCTGGTCACTGC
WDR62-AMP.4-R	TGCGTCCTCTCATGTAGCTC
WDR62-AMP.5-R	GAGAGTAGGCCCAGAAGCAG
WDR62-AMP.6-R	AACCTGTGTGGTCAATGCTG
WDR62-AMP.7-R	TGGGTATTTGTGGGATGGAC
WDR62-AMP.8-R	AGAGGGTACTGCTGCTTGG
WDR62-AMP.9-R	TTGCCCTCATGGAACCTAAC
WDR62-AMP.10-R	CTCAGCTCCCCAGAGACG
WDR62-AMP.11-R	TATTCACAGCCCCTCATTCC
WDR62-AMP.12-R	AACTCCCAGCTGTCATCCAC
WDR62-AMP.13-R	CGGCCCCAGTCATTATTCTG
WDR62-AMP.14-R	ATTCCTTCCCAGCTTCTTGG
WDR62-AMP.15-R	GGCACAGGTCTCAGGTGTG
WDR62-AMP.16-R	ACTCCCACAATCCAAACTGG
WDR62-AMP.17-R	TCAGCTGTGAGTGTGCTAAGG
WDR62-AMP.18-R	GCTGGGATTACAGGTATCAGC
WDR62-AMP.19-R	ACTGAGCCTGGAAGGATGAC
WDR62-AMP.20-R	GGAGTGGGAGTCAGCCTAAG

WDR62-AMP.21-R	CCACTGGGAAGGCAGAGAG
WDR62-AMP.22-R	CAGAATCCTCAGGCAGCAG
WDR62-AMP.23-R	CGTCTTCCAGGCTCAGTGG
WDR62-AMP.24-R	CTCTCCACTCACCCCTCCTG
WDR62-AMP.25-R	CACAGAACTGATGGCTAGGG
WDR62-AMP.26-R	CGAATGAATGAATGGCACAG
WDR62-AMP.27-R	TCTGGGCATCACCTTCTACC

## Appendix 11

Sequencing oligonucleotides used to confirm the sequences of *CDK6* (a) and *CEP135* (b) constructs with entry vector (pENTR™ Directional TOPO® Cloning Kits, Invitrogen) (c) and destination vector (d).

(a)

CDK6-Seq-F1	CGTGGTCAGGTTGTTTGATG
CDK6-Seq-R1	AATTGGTTGGGCAGATTTTG

(b)

CEP135-Seq-F1	TTGGAGAAAGAGAGCAAAGC
CEP135-Seq-R1	AATTCCACCACAACCCTCTC
CEP135-Seq-F2	CAAAAATGAAAACTCTGCCAAG
CEP135-Seq-R2	TCTTCTTTTTCTTGCACTGCTG
CEP135-Seq-F3	GACTCCAACATATAATACAGCGAAG
CEP135-Seq-R3	TTTTATGGAAAGTCGGTGCTG

(c)

M13 Forward	GTAAAACGACGGCCAG
M13 Reverse	CAGGAAACAGCTATGAC

(d)

DEST53-F	GATGGAAACATTCTCGGACAC
DEST53-R	GCTTAATGCGCCGCTACAG

## Erklärung

Ich versichere, daß ich die von mir vorgelegte Dissertation selbstständig angefertigt, die benutzten Quellen und Hilfsmittel vollständig angegeben und die Stellen der Arbeit – einschließlich Tabellen, Karten und Abbildungen, die anderen Werken in Wortlaut oder dem Sinn nach entnommen sind, in jedem Einzelfall als Entlehnung kenntlich gemacht habe; daß diese Dissertation noch keiner anderen Fakultät oder Universität zur Prüfung vorgelegen hat; daß sie abgesehen von unten angegebenen Teilpublikationen noch nicht veröffentlicht worden ist sowie, daß ich eine solche Veröffentlichung vor Abschluß des Promotionsverfahrens nicht vornehmen werde.

Die Bestimmungen dieser Promotionsordnung sind mir bekannt. Die von mir vorgelegte Dissertation ist von **Prof. Dr. Peter Nürnberg** betreut worden.

Köln, im Februar 2011

---

**Muhammad Sajid Hussain**

### Teilpublikationen:

**Hussain MS**, Bakhtiar SM, Muhammad F, Inayat IA, Janzen E, Toliat MR, Eiberg H, Kjaer KW, Tommerup N, Noegel AA, Nürnberg P, Baig SM, Hansen L.(2012). Genetic heterogeneity in Pakistani microcephaly families. **Clinical Genetics**. 9999(999A).

**Hussain M.S**, Baig S.M, Neumann S, Nürnberg G, Farooq M, Ahmad I, Wajid M, Alef T, Hennies H-C, Technau M, Altmüller J, Frommolt P, Thiele H, Noegel A.A, Nürnberg P. (2012). A Truncating Mutation of CEP135 is associated with primary microcephaly and disturbed centrosomal function. **Am. J. Hum. Genet.** 90, 1–8.

Muhammad F, Mahmood Baig S, Hansen L, **Sajid Hussain M**, Anjum Inayat I, Aslam M, Anver Qureshi J, Toilat M, Kirst E, Wajid M, Nürnberg P, Eiberg H, Tommerup N, Kjaer KW. (2009). Compound heterozygous ASPM mutations in Pakistani MCPH families. **Am. J. of Med. Genet. A.** 149A(5), 926-30.

## Curriculum Vitae

Name Muhammad Sajid Hussain  
Date of birth 10. March 1981  
Place of birth Muzaffar Garh, Pakistan  
Nationality Pakistani

### Education

2007- 2011 **Ph.D student** at the Cologne Center for Genomics (CCG), University of Cologne, Germany.  
Area of Research: Human molecular genetics.  
**Ph.D thesis title**  
“Genetic analysis of autosomal recessive primary microcephaly in Pakistani kindreds”.

2007-2009 **DAAD sandwich scholarship** from June 2007 to September 2009.

2005-2006 M.Phil (Master of Philosophy) course work in Biotechnology from NIBGE (National Institute for Biotechnology and Genetic Engineering), Quaid-e-Azam University Islamabad, Pakistan.

2003-2004 **Teaching experience**  
One and half year teaching experience, as Lecturer in Biology in Superior Science Higher Secondary School Muzaffar Garh (Punjab)

2000-2003 **M.Sc (Master of Science)** in Botany, Bahauddin Zakariya University Multan.

1998-2000 **B.Sc (Bachelor of Science)**, Bahauddin Zakariya University Multan.

1996-1998 **F.Sc (Faculty of Science)** Board of Intermediate and Secondary Education D.G Khan.

1994-1996 **Secondary School Certificate** (Matriculation Examination) Board of Intermediate and Secondary Education Dera Ghazi Khan.

## Lebenslauf

Name Muhammad Sajid Hussain  
Geburtsdatum 10. März 1981  
Geburtsort Muzaffar Garh, Pakistan  
Staatsangehörigkeit Pakistani

### Bildung

2007- 2011 **Doktorand** am Cologne Center for Genomics (CCG),  
Universität zu Köln, Deutschland  
**Dissertationsthema;**  
“Molecular genetic analysis of autosomal recessive primary  
microcephaly in Pakistani kindreds”.

2007-2009 **DAAD-Stipendium** von Juni 2007 bis September 2009.

2005-2006 **M.Phil (Master of Philosophy)** im Bereich der Biotechnologie am  
NIBGE (National Institute for Biotechnology and Genetic  
Engineering), Quaid-e-Azam University Islamabad, Pakistan.

2003-2004 **Lehrerfahrung:** Anderthalb Jahre Lehrerfahrung als Dozent im  
Fach Biologie an der Superior Science Higher Secondary School  
Muzaffar Garh (Punjab, Pakistan).

2000-2003 **M.Sc (Master of Science)** in Botanik, Bahauddin Zakariya  
University Multan.

1998-2000 **B.Sc (Bachelor of Science)**, Bahauddin Zakariya University  
Multan.

1996-1998 **F.Sc (Faculty of Science)** Board of Intermediate and Secondary  
Education D.G Khan.

1994-1996 **Secondary School Certificate (Matriculation Examination)**  
Board of Intermediate and Secondary Education Dera Ghazi Khan.