Linkage Analysis and Gene Hunt in Families with Inherited Disorders

By

SARA MUMTAZ

Department of Animal Sciences
Faculty of Biological Sciences
Quaid-i-Azam University
Islamabad
2017

Linkage Analysis and Gene Hunt in Families with Inherited Disorders

A Dissertation submitted to the Department of Animal Sciences,
Quaid-i-Azam University, Islamabad, in partial fulfillment
of the requirements for the degree of

Doctor of Philosophy

In

Human Genetics

By

SARA MUMTAZ

Islamabad

Presented to

Department of Animal Sciences
Faculty of Biological Sciences
Quaid-i-Azam University
Islamabad
2017

بسم الله الرحمن الرحيم

DECLARATION

I hereby declare that I have worked on my thesis “Linkage Analysis and Gene Hunt in Families with Inherited Disorders” independently and the work presented here is original. This thesis has not been submitted in the current or a similar form to any other university.

Sara Mumtaz

Islamabad, 2017

CERTIFICATE

This is to certify that this dissertation entitled "Linkage Analysis and Gene Hunt in Families with Inherited Disorders" submitted by Sara Mumtaz is accepted in its present form by the Department of Animal sciences, Quaid-i-Azam University, Islamabad, as satisfying the dissertation requirements for the degree of Doctor of Philosophy in human genetics.

Supervisor:

_____
Dr. Sajid Malik
Department of Animal Sciences
Quaid-i-Azam University
Islamabad.

Chairperson:

_____
Prof. Dr. Sarwat Jahan
Department of Animal Sciences
Quaid-i-Azam University
Islamabad.

External Examiner: _____

Dated:……

Dedicated

To

My Loving

Parents and Teachers

In the name of Allah the most Merciful and Beneficent

Allah will exalt those who believe among you and

those who have knowledge to high ranks.

(The Holy Quran)

Golden Saying of Prophet ﷺ

One hour's meditation on the work of the Creator

is better than seventy years of prayers.

(Hadith)

ACKNOWLEDGEMENTS

All the praises for **Almighty Allah**, Who created us as Muslim and blessed us with knowledge to differentiate between right and wrong. Many thanks to Him as He blessed us with the **Holy Prophet Hazrat Muhammad,** (Peace Be Upon Him), for whom the whole universe is created and who enabled us to worship only one Allah. He (PBUH) brought us out of darkness and enlightened the way of Heaven.

I feel great pleasure in expressing my special thanks to my research supervisor Dr. Sajid Malik whose guidance, encouragement, valuable suggestions and discussions enabled me to complete this tedious work. I am thankful to the Dean Faculty of Biological Sciences, Prof. Dr. Wasim Ahmad and the Chairperson Prof. Dr. Sarwat Jahan, for their support and providing necessary research facilities. Special thanks are due to all the teachers for being a source of inspiration and enlightenment for me.

I am highly indebted to my co-supervisor Prof. Dr. Aslihan Tolun and her entire team at the Molecular Genetics Laboratory, Bogazici University Istanbul, Turkey. Dr. Tolun very generously allowed me to visit her research laboratory to carry out necessary experimental and analytical work. She guided me throughout my studies and provided substantial contribution to my thesis. She is unique in her scientific understanding, support, work ethics and hospitality.

I am very thankful to Prof. Dr. Teepu Siddique and his team at the Northwestern University Chicago, USA, for giving me training in state of art molecular genetic techniques and facilitating me during my stay in USA. I would also extend my thanks to the Higher Education Commission of Pakistan, for providing me funding for this visit through IRSIP scheme.

I also owe my recognition to my sweet friends, my lab fellows and special thanks to my juniors who helped a great deal in collecting information regarding the families on which this thesis is based. Many thanks for the supporting staff of the Department of Animal Sciences. Finally, I would like to express my appreciation to all participants/families for their volunteer participation in the study; without their participation this study would have been impossible.

I have no words to acknowledge the sacrifices, efforts, prayers, guidance, support, encouragement and firm dedication of my Sweet Parents. Their endless prayers contributed to the successful completion of this research project.

Immense thanks from the core of my heart to my loving sisters and brother, Dr. Awais Ahmad, cute nephew Abdullah.

May Allah bless us all with true knowledge to severe the humanity.

Sara Mumtaz

Contents

S.No. Title **Page**

List of Tables

List of Figures

List of Abbreviations

AD	autosomal dominant
ADPCZ	Autosomal Dominant Postaxial Polydactyly, Camptodactyly and Zygodactyly (Family D)
ALS	amyotrophic lateral sclerosis
APA	American Psychiatric Association
APS	ammonium peroxydisulfate
AR	autosomal recessive
BAM	Binary Alignment Map
bHLH	Basic helix-loop-helix protein
CEPH	cephalometric radiograph
CNVs	copy number variations
CT	computed tomography
DGV	Database of Genomic Variants
ExAC	Exome Aggregation Consortium
GCPS	Greig cephalo-polysyndactyly syndrome
GUI	graphical user interface
HapMap	Haplotype map project
HCiE	Homozygosity Comparison in Excel
Het	heterozygous
Hom	homozygous
IBD	identical by descent
ID	intellectual disability
IDSD	Intellectual Disability with Speech Problem and Dysmorphic Facial Features (Family A)

Indel	insertion-deletion
IRB	Institutional Review Board
LOD	logarithm of odds
MCPH	autosomal recessive primary microcephaly (Family B)
MHC	major histocompatibility complex
MIDSB	Microcephaly, Intellectual Disability, Short Stature and Brachydactyly (Family C)
MOPD	microcephalic osteodysplastic primordial dwarfism
MRI	magnetic resonance imaging
MSSD	mesoaxial synostotic syndactyly (Family G, H)
NGS	Next Generation Sequencing
OMIM	Online Mendelian Inheritance in man
OPG	oro-pantomograms
PCR	polymerase chain reaction
PHS	Pallister-Hall syndrome
RFLPs	restriction fragment length polymorphisms
SDS	sodium dodecylsulfate
SLS	Seckel-like Syndrome (Family F)
SNPs	single nucleotide polymorphisms
SPD	synpolydactyly
SSCP	single-strand conformation polymorphism
STRs	short tandem repeats
TAOH	Tooth Agenesis manifested as Oligodontia/Hypodontia (Family E)
TBE	Tris/Borate/EDTA
TEMED	N,N,N,N-tetramethylethylenediamine
TMJ	temporo-mandibular joint

TNF	tumor necrosis factor
UTR	untranscribed region
VNTRs	variable number of tandem repeats
WES	whole-exome sequencing
WHO	World Health Organization

ABSTRACT

Mendelian diseases have been difficult to study due to various limitations including difficulty in ascertainment of subjects/families with rare phenotypes, inadequate data, and laborious molecular methods. However, the recent advances in genetic technologies have made it possible to accurately identify the molecular bases of such disorders.

By studying such disorders and identifying their underlying molecular causes, it becomes possible to control recurrence of such disorders by proper genetic counselling for affected families. Furthermore, the discovery of each new gene of a genetic disorder tells us about basic cellular processes, which is important for basic science. Linkage analysis employing genome-wide SNP markers by using large pedigrees is a powerful tool for locus identification. After locus identification, exome sequencing and Sanger sequencing are used for finding the pathogenic mutation in the gene. Large families suitable for genetic studies are relatively abundant in Pakistan, because of its cultural setup that promotes large families, consanguineous marriages, and stable communities.

In the framework of this study, eight Pakistani families afflicted with rare diseases were explored for disease gene loci and causative mutations by using SNP-based linkage analysis and exome sequencing and/or Sanger sequencing. The diseases were Intellectual Disability with Speech Problem and Dysmorphic Facial Features (IDSD), Autosomal Recessive Primary Microcephaly (MCPH), Microcephaly,

Intellectual Disability, Short Stature and Brachydactyly (MIDSB), Autosomal Dominant Postaxial Polydactyly, Camptodactyly and Zygodactyly (ADPCZ), Tooth agenesis manifested as Oligodontia/Hypodontia (TAOH), Seckel-like syndrome (SLS), and Syndactyly type IX (MSSD).

The first family (IDSD) was with syndromic intellectual disability manifesting with impaired intelligence along with other problems. Three affected siblings and their cousin were subjected to linkage mapping and exome sequencing, and a novel homozygous missense *PDIA3* mutation was identified in these patients. The second family (MCPH) was suffering from autosomal recessive primary microcephaly (MCPH). It is characterized by a marked decrease in the brain volume and nonprogressive intellectual disability. However, the architecture of the brain remains normal. In this family linkage analysis was performed and the disease locus was found on chromosome 1q31 harboring *ASPM* gene of primary microcephaly. This family was not studied further as many mutations were already reported in this gene from the Pakistani population. The third family (MIDSB) was a consanguineous family suffering from yet another syndrome, primary microcephaly and a unique combination of skeletal, limb and skin abnormities. The disease gene locus was mapped to 18q11.2, and a homozygous *RBBP8* mutation was detected. The family is the fifth family reported with *RBBP8* mutation. The fourth family (ADPCZ) was a large autosomal dominant family with a highly variable polydactyly phenotype. We were expecting a new condition since the postaxial polydactyly manifested with a combination of synpolydactyly I or syndactyly type II. After subjecting the family to linkage analysis and whole-exome sequencing (WES), a novel mutation was discovered in *GLI3*. The fifth family (TAOH) was a large autosomal dominant family

with tooth agenesis. Whole genome SNP genotyping yielded several candidate loci. Mutation screening in the *WNT10A* gene did not reveal any pathogenic variants. It was concluded that there is still an uncharacterized gene in one of the candidate intervals. In the sixth family with Seckel-like syndrome (SLS), the whole genome SNP genotyping excluded all the known loci for hypodontia/oligodontia as the candidate loci segregating with the phenotype. There were several regions that were homozygous among the affected subjects but no possibly damaging variant in a gene relevant to the phenotype was found. It was concluded that there is still an uncharacterized gene in one of the candidate homozygous regions. In seventh and eighth families with syndactyly type IX (MSSD), two novel homozygous mutations were detected in *BHLHA9*.

In conclusion, in five families causative mutations were identified. In three families there were several regions segregating with the phenotypes and no pathogenic variants were detected in any gene relevant to the malformations. This research adds useful data to the scientific literature and furthers our understanding of the studied disorders. The scientific findings of this study would be highly beneficial for the recruited families in genetic counseling and genetic testing.

1 INTRODUCTION

Mendelian diseases are single-gene or monogenic disorders with full penetrance. They are individually rare, yet collectively affect millions of people in the world (WHO, 2005). It has been estimated that each year eight million children are born with some kind of genetic disorder around the globe. Birth defects, including a considerable number of Mendelian disorders, are considered a frequent reason of death in the early years of life. A huge number of individuals who survive severe birth defects need lifelong medical attention, specialized care, and costly symptomatic therapy. Thus, these disorders cause a high social and economic burden on the affected individual and his family as well as on the entire society (Angelis et al. 2015; Chong et al. 2015). Monogenic disorders have been described mostly in a well-known genetic catalogue called Online Mendelian Inheritance in Man (OMIM).

The diagnosis of Mendelian disorders by conventional diagnosis tests is not very successful and many patients remain without a correct diagnosis. In the normal practice, clinical geneticists rely upon clinical, radiographic and biopsy results along with this some additional tests such as metabolic tests, karyotyping and candidate-gene tests (Chong et al. 2015). This insufficiency in an accurate diagnosis can adversely affect patients, including the failure of potential treatment identification, failure to estimate the recurrence risk in subsequent pregnancies and failure to provide preventive guidance and prognosis. However, next generation sequencing and robust computational approaches together have revolutionized the discovery of rare variants

underlying genetic disorders. Now we are able to perform the task of disease variant detection with a much faster speed and in a more accurate way (Biesecker, 2010).

The identification of a disease-causing mutation not only helps in connecting clinical phenotype with genotype for a correct diagnosis but also gives a straightforward way to study altered proteins and their functions. Thus, we can get a deep understanding of the basic biological pathways and we can apply this knowledge to understand the development of the disease, which helps in the subsequent management of the disorder. The investigation of rare causal variants of Mendelian phenotypes has shown a basic relationship between human phenotypes and gene function (Chong et al. 2015).

Another advantage of identification of genes responsible for monogenic disorders is that we can apply this knowledge to find the biological pathways that are shared in Mendelian and common disorders. In some common disorders such as hypertension, diabetes and autism, a small portion of subjects has, in fact, monogenic diseases that share the phenotype with a common, complex disease. Causative genetic variants in such a small subset of patients are often helpful in understanding the general mechanisms that are involved in the development of those diseases. Many drugs for common disorder fail in the clinical trials, and the development of new drugs is limited due to a lack of knowledge about the basic, relevant biological pathway. Thus, studying families with rare Mendelian diseases caused by mutations with broad effects that intensify or lessen the risk of related common disorders can straightforwardly tell us about the fundamental relationship among genes, pathways, and common diseases. In this way, it is beneficial for the identification of targets that

likely cause more favorable effects and slighter adverse effects in therapeutics (Chong et al. 2015).

Pakistan has a high prevalence of hereditary diseases compared to the rest of the world. The most likely reason for this high prevalence is the common practice of consanguineous marriages, overlapping generations, marriage at young ages, large families, and well-adapted communities that present multigenerational family trees with a variety of rare hereditary disorders (Wahab and Ahmad, 1996; Peltonen et al. 2000).

The present research is intended to report various genetic disorders we detected in our population, to investigate their inheritance patterns, to characterize their phenotypes by common clinical procedures, to identify the disease gene loci by using SNP linkage analysis, and to find the causative mutations in each family by either candidate gene approach or exome sequencing. Our ultimate goals are to effectively use our research findings for the improvement of clinical diagnosis by using molecular analysis and to give genetic counseling for suffering families. A brief review of the literature on the anomalous conditions investigated and the molecular genetic techniques used in the study are presented below.

1.1 Family A: Intellectual Disability with Speech Problem and Dysmorphic Facial Features (IDSD)

Intellectual disability (ID), also frequently referred as early-onset cognitive impairment or mental retardation, has a great socio-economic burden (Ropers, 2010). Its distinguishing features include significantly impaired intellect and adaptive behavior with onset during developmental periods. Adaptive skills are reduced in three main domains, i.e., conceptual, social and practical abilities (APA, 2013). Based on severity levels, ID is classified as mild, moderate, severe and profound. Another classification of ID is syndromic and non-syndromic. In the syndromic entity, ID is accompanied by other clinical features, and in the non-syndromic form ID is the only clinical presentation. However, there is no straightforward way to classify into syndromic and non-syndromic as ID can accompany a variety of psychiatric behavioral and neurological disorders (Kaufman et al. 2010; Chiurazzi and Pirozzi, 2016).

Gene defects are a very common cause of ID, but their dissection is difficult due to the vast clinical and genetic heterogeneity. In the recent years, many genes responsible for ID have been discovered due to the progress in SNP-based genotyping, next-generation sequencing and functional studies. This has not only made family-based studies more fruitful but also shed light on biological pathways and diagnosis of such disorders, and offered possible therapeutic means (Vissers et al. 2016).

About 2,000 genetic factors underlying ID have been documented in the OMIM database. Even though autosomal genes make up the vast majority of those genes, the X-linked types have been studied most extensively, mainly due to robust dissection of inheritance pattern and genetic linkage. In Western populations, it appears that autosomal dominant (AD) types are more frequent than autosomal recessive (AR) types, whereas in the Muslim communities AR entities are common, due to the high frequency of parental consanguinity (Najmabadi et al. 2011). Several studies have reported mutations in numerous different genes in families with ID. However, due to the heterogeneous nature of ID, it is expected that a large number of loci and genes remain to be elucidated (Ropers, 2010).

We studied three Pakistani siblings and their cousin afflicted with a severe syndrome manifesting with ID, speech problem and dysmorphic facial features which we abbreviate as "IDSD". Through linkage mapping and exome sequencing we identified a homozygous missense *PDIA3* mutation in the patients. Previously, Gonzalez-Perez et al. (2015) showed that heterozygous *PDIA3* mutations cause amyotrophic lateral sclerosis (ALS), a neurodegenerative disorder. To the best of our knowledge, the present work is the first report of a neurodevelopment disorder caused by *PDIA3* mutation.

1.2 Family B: Autosomal Recessive Primary Microcephaly (MCPH)

Autosomal recessive primary microcephaly (MCPH) is a rare inherited disorder. It affects the development of the brain and is clinically diagnosed by an obvious small head size (≤ -2 SD) and non-progressive ID as compared to the average for age, sex, and ethnicity. However, the architecture of the brain remains normal (Kaindl 2014; Woods and Basto, 2014). Primary microcephaly is distinguished from secondary microcephaly, as former occurs during the prenatal period of life while the later develops after birth either due to an environmental insult or a genetic factor that affects postnatal life (Woods and Parker, 2013). In the development of MCPH genetic etiology plays an important role.

An underlying genetic heterogeneity is clearly evident by various studies, and mutations in various genes on different chromosomes have been reported in MCPH. However, according to Hu et al. (2014) mutations were not found in a large number of families suffering from MCPH that indicates further genetic heterogeneity. To date pathogenic mutations were observed in 17 genes, namely, microcephalin 1 (*MCPH1;* OMIM 607117), WD-repeat-containing protein 62 (*WDR62;* OMIM 613583), cyclin-dependent kinase 5 regulatory subunit-associated protein 2 (*CDK5RAP2;* OMIM 608201), cancer susceptibility candidate 5 (*CASC5;* OMIM 609173), centromeric protein J (*CENPJ*; OMIM 609279), abnormal spindle-like microcephaly-associated protein (*ASPM*, OMIM 605481), SCL/TAL1-interrupting locus (*STIL*; OMIM 181590), centrosomal protein 152 kD (*CEP152*; OMIM 613529), centrosomal protein

135 kD (*CEP135*; OMIM 611423), polyhomeotic-like 1 (*PHC1;* OMIM 602978), zinc finger protein 335 (*ZNF335*; OMIM 610827), cyclin-dependent kinase 6 (*CDK6*; OMIM 603368), SAS-6 centriolar assembly protein (*SASS6*; OMIM 609321), centromere protein E *(CENPE*; OMIM 117143), and major facilitator superfamily domain containing 2A (*MFSD2A*; OMIM 614397) (Table 1.1). According to recent studies, in the majority of cases, mitotic events are disrupted in neuronal precursor cells that lead to a decrease in the number of neurons (Faheem et al. 2015).

In the family investigated in this study all patients had microcephaly that manifested at birth and ID running in the family in an autosomal recessive fashion. In the light of these symptoms, autosomal recessive primary microcephaly was diagnosed. We performed linkage analysis, and subsequently, the disease locus was found as the locus on chromosome 1 for primary microcephaly *ASPM* gene. Since *ASPM* is a very large gene (27 exons) and many mutations have been reported in this gene with the majority from the Pakistani population, we decided not to continue with the study. We shall try sequencing this large gene in the future.

Table 1.1. Genetic heterogeneity in primary microcephaly.

OMIM	**Type***	**Locus**	**Genes**	**Reference**
251200	MCPH1	8p23.1	*MCPH1*	Jackson et al. 2002
604317	MCPH2	19q13.1-13.2	*WDR62*	Nicholas et al. 2010
604804	MCPH3	9q33.2	*CDK5RAP2*	Guernsey et al. 2010
604321	MCPH4	15q15-q21	*CEP152*	Jamieson et al.1999
608716	MCPH5	1q31.3	*ASPM*	Bond et al. 2002
608393	MCPH6	13q12.12	*CENPJ*	Bond et al. 2005
612703	MCPH7	1p33	*STIL*	Kumar et al. 2009
614673	MCPH8	4q12	*CEP135*	Hussain et al. 2012
614852	MCPH9	15q21.1	*CEP152*	Guernsey et al. 2010
615095	MCPH10	20q13.12	*ZNF335*	Yang et al. 2012
615414	MCPH11	12p13.31	*PHC1*	Awad et al. 2013
616080	MCPH12	7q21.2	*CDK6*	Hussain et al. 2013
616051	MCPH13	4q24	*CENPE*	Mirzaa et al. 2014
616402	MCPH14	1p21.2	*SASS6*	Khan et al. 2014
616486	MCPH15	1p34.2	*MFSD2A*	Alakbarzade et al. 2015
616681	MCPH16	12q24.33	*ANKLE2*	Yamamoto et al. 2014
617090	MCPH17	12q24.23	*CIT*	Shaheen et al. 2016

* all types are autosomal recessive

1.3 Family C: Microcephaly, Intellectual Disability, Short Stature and Brachydactyly (MIDSB)

Primary microcephaly is generally an isolated entity. However, in a few instances, it is witnessed with other abnormalities where it is called as a syndrome. In Seckel syndrome 2 (OMIM 606744), a sub-type primordial dwarfism, microcephaly is associated with proportionate dwarfism manifesting at birth, bird-like facial appearance, intellectual disability, minor hand deformities, cardiac murmurs and birth spots (Borglum et al. 2001). Another similar condition is Jawad syndrome (OMIM 251255), which has been described in a consanguineous Pakistani family and autosomal recessive inheritance. Its cardinal features include microcephaly, mental retardation, and hand and nail deformities. Both Seckel syndrome 2 and Jawad syndrome were localized at 18p11.31-q11.2 and pathogenic mutations were found in *RBBP8* (Hassan et al. 2008; Qvist et al. 2011). Further, together with a heterozygous deletion in *NRXN1*, it was proposed that another homozygous mutation caused a complex microcephaly syndrome characterized by microcephaly, mental retardation, epilepsy and diabetes (Agha et al. 2014). Several other microcephaly syndromes are known. For example, in Filippi syndrome (OMIM 272440) microcephaly occurs with skeletal anomalies, intellectual impairment, convulsions, and hair and testicular problems, and the underlying mutation was found in *CKAP2L* at 2q13 (Hussain et al. 2014). In the case of severe combined immunodeficiency with microcephaly, growth retardation, and sensitivity to ionizing radiation (SCID; OMIM 611291), microcephaly occurs together with severe combined immunodeficiency along with

other congenital defects in a group of patients, and the causative mutations are in *NHEJ1* located at 2q35 (Buck et al. 2006).

As exemplified by the above-mentioned syndromes, microcephaly often occurs together with various other abnormalities. Molecular genetic studies have revealed different loci and pathogenic mutations in different genes in those syndromes. This demonstrates the clinical and genetic heterogeneity.

We present a consanguineous Pakistani family afflicted with yet another syndrome, microcephaly and a unique combination of skeletal, limb and dermatological abnormities. We mapped the disease gene locus to 18q11.2 and detected the same *RBBP8* mutation as in the Jawad syndrome family. The family is the fifth family with the recessive *RBBP8* mutation.

1.4 Family D: Autosomal Dominant Postaxial Polydactyly, Camptodactyly and Zygodactyly (ADPCZ)

Limb and digit anomalies have a prevalence of approximately six in 10,000 live births, and upper limbs are more often affected than lower limbs. Isolated polydactyly (OMIM 603596) is a frequently occurring digit anomaly. Usually, it appears as preaxial wherein the extra digit is attached to hallux, and postaxial means the extra digit is attached to the fifth toe. A less frequent polydactyly type is mesoaxial which is the presence of the extra digit in three central digits (Malik, 2014;

Malik et al. 2014). Sporadic polydactyly cases, mostly unilateral, are more common as compared to familial bilateral cases (Orioli and Castilla, 1999).

Polydactyly has a high phenotypic variability. As in the postaxial polydactyly type A (OMIM 174200), the extra digit is more developed as compared to type B in which extra digit is just a small rudimentary nubbin (Temtamy and McKusick, 1978; Malik, 2014).

Genetic heterogeneity is also evident in polydactyly, as genetic mapping studies have identified different loci in different affected families. For instance, Zguricas et al. (1999) localized preaxial polydactyly II at 7q36 (OMIM 174500), and postaxial polydactyly was located at 13q (OMIM 60208) by Akarsu et al. (1997) and at 19p (OMIM 607324) by Zhao et al. (2002). In other studies, postaxial types were mapped to 7p14 (OMIM 174200) and causal mutations were identified in *GLI3* (glioma-associated oncogene family zinc) (Radhakrishna et al. 1997; Wieczorek et al. 2010).

A large Pakistani family with a highly variable polydactyly phenotype was investigated in this study. We were anticipating a new condition since the postaxial polydactyly manifested with a combination of synpolydactyly I or syndactyly type II (OMIM 186000). Eventually, a novel mutation was discovered in *GLI3* by linkage analysis and whole-exome sequencing (WES).

1.5 Family E: Tooth Agenesis Manifesting as Oligodontia/Hypodontia (TAOH)

Tooth agenesis is the congenital absence of one or more permanent teeth (Pemberton et al. 2005). Selective tooth agenesis without associated systemic disorders (OMIM 106600) is divided into two types: hypodontia, the agenesis of fewer than 6 teeth, and oligodontia, the agenesis of six or more permanent teeth. In both cases, third molars (wisdom teeth) are not included.

Tooth agenesis is the most common developmental anomaly of the human dentition occurring in ~25% of the population (Garib et al. 2009). The most affected tooth is the third molar with a prevalence of agenesis of 20.7% (Garib et al. 2009). Excluding third molars, the prevalence of tooth agenesis is ~ 4.3-7.8%, and the most frequently missing teeth are mandibular second premolars, followed by the maxillary lateral incisors and maxillary second premolars. It has been noticed that only one or two permanent teeth are missing in the majority of patients (76-83%) with tooth agenesis (Polder et al. 2004). Song et al. (2009) has reported that the prevalence becomes progressively smaller as the number of missing teeth increases. Agenesis of more than two teeth occurs in ~1% of the population (Polder et al. 2004). Oligodontia (OMIM 604625) affects ~0.1-0.3% of the population (Polder et al. 2004). It is important to draw attention to the ethnical differences in the prevalence of tooth agenesis. Epidemiological studies have demonstrated a lower prevalence of agenesis in African Americans as compare to European Americans, while the Asians have a high prevalence of agenesis. There is also gender bias in the prevalence of tooth

agenesis; in general, women have more severe phenotypes than men (Polder et al. 2004).

Genetics probably represents the primary etiological factor of tooth agenesis (Garib et al. 2010), as the prevalence of agenesis is higher in families of affected patients (Mossey, 1999). So far, many gene defects have been reported that cause tooth agenesis either as an isolated anomaly (non-syndromic) or as a part of multiple congenital anomalies (syndromic) (Song et al. 2009). Non-syndromic tooth agenesis has wide phenotypic heterogeneity and is classified as sporadic and familial, which can be inherited in an autosomal dominant, autosomal recessive, or X-linked fashion (Burzynski and Escobar, 1983). Both hypodontia and oligodontia exhibit incomplete penetrance and variable expressivity.

Dominant mutations in different genes such as *MSX1*, *PAX9* and *AXIN2* have been found in families with non-syndromic tooth agenesis (Vastardis et al. 1996; Stockton et al. 2000; Lammi et al. 2004) (Table 1.2). However, mutations in these genes were detected in only a few affected individuals suggesting that there may be other unidentified genes responsible for this group of diseases (Scarel et al. 2000; Frazier-Bowers et al. 2002).

Table 1.2. Hereditary conditions with tooth agenesis (STHAG).

OMIM	Type	Inheritance	Cardinal features	Locus	Gene	Reference
106600	STHAG1	AD	Agenesis of third molars, upper lateral incisors or lower second premolars	4p16.2	*MSX1*	Vastardis et al. 1996
602639	STHAG2	AR	Agenesis of third molars, upper lateral incisors or lower second premolars	16q13.3	-	Ahmad et al. 1998
604625	STHAG3	AD	Agenesis of maxillary, mandible 2nd premolars and mandibular central incisors	14q12	*PAX9*	Stockton et al. 2000
150400	STHAG4	AD	Absence of premolar and pegged or missing maxillary lateral and upper lateral incisors	2q35	*WNT10A*	Nieminen et al. 1995
610926	STHAG5	AD	Oligodontia	10q11.2-q21	-	Liu et al. 2001
613097	STHAG6	AR	Absence of 6 or more teeth including third molars	11q12	*LTBP3*	Noor et al. 2009
616724	STHAG7	AD	Oligodontia	12p13.2	*LRP6*	Massink et al. 2015
617073	STHAG8	AD	Oligodontia	12q13.12	*WNT10B*	Yu et al. 2016
313500	STHAG Xl	XD	Oligodontia or hypodontia	Xq12-q13.1	*EDA*	Han et al. 2008

AD=Autosomal dominant; AR=Autosomal recessive; XD= X-linked dominant.

A large Pakistani family with tooth agenesis was studied. The phenotype in this family was highly variable and was segregating in an autosomal dominant fashion. SNP genotyping and subsequent model-based linkage analyses were performed in order to localize the responsible gene. A few potential candidate regions were identified. In one of the candidate intervals encompassing 2q35 locus (STHAG4), sequence analyses of *WNT10A* did not yield any pathogenic variant.

1.6 Family F: Seckel-like Syndrome (SLS)

Seckel syndrome (OMIM 210600) is an extremely rare inherited disorder reported by Seckel in 1960. He described this novel disease on the basis of two cases he had studied in Chicago as well as 13 cases on nanocephalic dwarfs reported in the literature over a 200-year period (Seckel, 1960). The disorder is characterized by growth delay prior to birth resulting in low birth weight. Growth delay continues after birth resulting in proportionate short stature, microcephaly and mental retardation. This syndrome is associated with an unusual "beak-like" protrusion of nose. Other facial features may include abnormally large eyes, a narrow face, malformed ears and an unusually small jaw.

In 1967 McKusick and colleagues documented this condition in three out of 11 siblings in a family and suggested that inheritance was autosomal recessive. Goodship et al. (2000) studied two consanguineous families with Seckel syndrome from the same village in Pakistan who were not known to be related to each other. They allocated the first Seckel syndrome (*SCKL1*) gene. Other loci for Seckel syndrome have been mapped on different chromosomes (Table 1.3). This confirms the genetic heterogeneity of Seckel syndrome. In addition, a number of Seckel-like syndromes have been identified, particularly microcephalic osteodysplastic primordial dwarfism (*MOPD*) types I–III (Bass et al. 1975; Majewski and Spranger, 1976; Majewski and Goecke, 1982; Majewski et al. 1982a, 1982b). These anomalies can be

differentiated from Seckel syndrome on the clinical and radiographic grounds (Table 1.3).

In the present study, a family with Seckel-like syndrome was investigated. The clinical presentation in the affected subjects was concordant with bird-headed dwarfism and Seckel syndrome but with additional skeletal abnormalities. SNP genotyping and linkage analyses revealed that this condition does not map to the already known loci for Seckel syndrome or bird-headed dwarfism, and there is yet an unknown locus for the malformation segregating in this family.

Table 1.3. Seckel syndrome and its differential diagnosis.

OMIM	Type*	Phenotype	Locus	Gene	Reference
210600	SCKL1	Proportionate short stature, growth retardation, microcephaly, mental retardation, craniofacial anomalies	3q22-q24	*ATR*	Goodship et al. 2000
606744	SCKL2	Proportionate short stature, growth retardation, microcephaly, mental retardation, craniofacial anomalies, 11 pairs of ribs	18p11-q11	*RBBP8*	Qvist et al. 2011
608664	SCKL3	Proportionate short stature, microcephaly with mental retardation	14q	-	Kilinc et al. 2003
613676	SCKL4	Proportionate short stature, growth retardation, microcephaly, craniofacial anomalies, 11 pairs of ribs	13q12	*CENPJ*	Al-Dosari et al. 2010
613823	SCKL5	Microcephaly, high nasal bridge, beaked nose, retrognathia	15q21.1	*CEP152*	Kalay et al. 2011
614728	SCKL6	Microcephaly, speech delay, cognitive delay, short stature	3q22.2	*CEP63*	Sir et al. 2011
614851	SCKL7	severe pre- and postnatal growth retardation, microcephaly, developmental delay	14q22.1	*NIN*	Dauber et al. 2012
615807	SCKL8	Short stature, characteristic facial features, severe microcephaly, global developmental delay, intellectual disability	10q21.3	*DNA2*	Shaheen et al. 2014
616777	SCKL9	Intrauterine growth retardation, microcephaly, beaked nose, failure to thrive, early death	3p21.31	*TRAIP*	Harley et al. 2016
210710	MOPDI	Disproportionate Short stature with short limbs, growth	2q14.2-	*RNU4AT*	Leutenegger et al.

		retardation, microcephaly with mental retardation, carniofacial anomalies	q14.3	*AC*	2006
210720	MOPDII	Disproportionate short stature, delayed bone age, growth retardation, microcephaly with mental retardation, craniofacial anomalies, truncal obesity	21q22.3	*PCNT*	Rauch et al. 2008
210730	MOPDIII	Proportionate dwarfism, protruded sternum and ribs, growth retardation, microcephaly with mental retardation, craniofacial anomalies	-	-	Majewski et al. 1982b

MOPD = microcephalic osteodysplastic primordial dwarfism
* all types are autosomal recessive

1.7 Families G and H: Syndactyly type IX, Mesoaxial Synostotic Syndactyly (MSSD)

Syndactyly type IX, mesoaxial synostotic syndactyly (MSSD; Malik-Percin type; OMIM 609432) is an extremely rare autosomal recessive non-syndromic digit anomaly (Percin et al. 1998; Malik et al. 2005b). Of the nine well-characterized non-syndromic syndactylies it is the only recessive type except for Cenani-Lenz syndactyly (syndactyly type VII; OMIM 212780) (Malik et al. 2005a; Malik, 2012). The type IX syndactyly demonstrates a distinctive combination of clinical features which include mesoaxial osseous synostosis at the metacarpal level, reduction of one or more phalanges, hypoplasia of distal phalanges of preaxial and postaxial digits, clinodactyly of 5^{th} fingers and preaxial fusion of toes (Malik et al. 2004). The heterozygous carriers do not show any symptoms.

It has been previously established that syndactyly type IX is genetically distinct from other well-characterized syndactyly types and through autozygosity mapping a novel locus on chromosome 17p13.3 has been identified (Malik et al. 2005b). However, the underlying gene and the causative mutation remain to be elucidated. In addition, any similar phenotype with known genetic etiology has not been reported in model organisms. It is speculated that a key developmental gene with a pivotal role in the late stages of limb morphogenesis, particularly in digit development and separation might be involved in this anomaly.

We studied two families (in a pool of six families) of Pakistani origin with type IX syndactyly phenotype. In these families, linkage analyses resulted in a significant reduction of the MSSD locus at 17p13.3 and mutations were detected in *BHLHA9*, encoding a bHLH protein. Two different *BHLHA9* mutations were identified in the present two families. These results were part of the data published in Malik et al. (2014).

1.8 Genetic approaches to families with hereditary anomalies

1.8.1 Linkage analysis

Genome-wide linkage mapping has been tremendously successful in the identification of the causative genes for highly penetrant monogenic diseases (Jimenze-Sanchez, 2001). Linkage is the tendency of two loci on the same chromosome to be inherited together. Two loci on the same chromosome that are close enough to each other will segregate together; otherwise, crossing over in meiosis takes place often between homologous chromosomes and new combinations of alleles are created. Recombination frequency (theta; θ) represents the frequency of a single crossover between two loci in a meiosis event. The probability of a recombination to take place increases as the distance between two loci increases. The maximum value for θ is 0.5, which denotes that the two loci are so far apart that they behave as if they are on different chromosomes.

The recombination events can be observed using various types of genetic markers, including restriction fragment length polymorphisms (RFLPs), microsatellites or short tandem repeats (STRs), variable number of tandem repeats (VNTRs), and single nucleotide polymorphisms (SNPs). SNP arrays have improved information content as it can be processed with higher throughput. Cost-effectiveness of SNP genotyping arrays with high marker density and accessibility of information

about genetic markers through online databases facilitate mapping of recombination events and localization of disease genes with high efficiency. Hence, it is an outstanding tool for linkage analysis (Middeldrop et al. 2007).

Linkage analysis can be either nonparametric (model-free) or parametric (model-based). Nonparametric linkage analysis is applied to studies on complex traits which are due to interactions of two or more genes and may also be affected by environmental factors. A nonparametric method searches a shared haplotype in affected individuals presumably harboring the disease gene and is identical by descent from a common ancestor, regardless of the mode of inheritance. A parametric method uses the genetic information about the disease-allele frequency, penetrance, and mutation rates to assess the mode of inheritance. The statistical test that is used in parametric linkage analysis is LOD (logarithm of odds) score analysis (Ott et al. 2015).

1.8.2 LOD score calculations

LOD score analysis discriminates between the null hypothesis of no linkage ($\theta = 0.5$) and the alternative hypothesis of linkage ($\theta<0.5$) (Terwilliger and Ott, 1994). In other words, it compares the likelihood of two loci being actually linked with the likelihood of observing the same event purely by chance. Odds ratio denotes the ratio of the probability of observing the distribution of two traits in a pedigree under linkage at θ, to the probability of observing it under the hypothesis of no linkage ($\theta = 0.5$) (Risch, 1992). The logarithm of odds ratio Z (θ) is calculated for a range of θ

values, and the value that gives the maximum LOD score is reported as the maximum likelihood estimate. A LOD score of 3 is the threshold used in human linkage analysis to accept linkage. The linkage is excluded for -2 LOD score and LOD score values between -2 and +3 are considered inconclusive (Terwilliger and Ott, 1994; Strachan and Read, 2004).

LOD score analysis is either two-point or multipoint. Two-point LOD score analysis calculates the co-segregation of each marker with the disease locus, whereas multipoint LOD score analysis calculates co-segregation of each marker with more than two loci at a time and outputs LOD scores for all possible positions on a given map of genetic markers. Several open source softwares for two-point and multipoint LOD score calculations are available for computerized linkage analysis. Different softwares require different commands and input formats, and easyLINKAGE package converts all input information into a standard form to create a more user-friendly platform (Hoffman and Lindner, 2005).

1.8.3 Homozygosity mapping

Homozygosity mapping is based on the fact that affected individuals in a consanguineous family or inbred population are likely to receive a shared haplotype that has descended from a common ancestor, in the homozygous state (Lander and Botstein, 1987). This method is very efficient for mapping autosomal recessive diseases in consanguineous families. Usually, the frequency of such a disease is so low that it can only arise from inbreeding, and affected individuals in the family are

homozygous. This method is also part of the parametric linkage analysis and is widely used in gene localization studies in consanguineous families with two or more affected sibs. Linkage softwares are employed on the genotype data with an aim to detect homozygous intervals possibly identical by descent (IBD). A segment of a chromosome is IBD if it is inherited it from a common ancestor without recombination.

1.8.4 Next generation sequencing

In the past few years, Next Generation Sequencing (NGS) technologies have revolutionized research in the field of genomics with the aim to provide faster, cheaper and more accurate genomic sequencing information (von Bubnoff 2008; Behjati and Tarpey 2013; Sener et al. 2016). Traditional Sanger sequencing is performed by DNA amplification, and the final sequence obtained represents a population of DNA templates (Sener et al. 2016). However, NGS has the power to sequence single DNA molecules. Short sequence copies (40-400 bp), called reads, are generated in billions in a very short time (Pettersson et al. 2009). This increase in throughput pays off with the compromised accuracy of each short read generated by NGS, which is very low compared to Sanger sequencing. Therefore, multiple reads are generated for the same region, a concept called sequencing depth, with the aim of obtaining more reliable sequence information. The more deeply covered a region is, the more accurate the sequence is. Although deep coverage conveys most of the information on a region, validation via Sanger sequencing of the strong candidate variants or mutations reported in the NGS results is necessary (Ku et al. 2012). NGS

technologies have a very broad field of application. In addition to the sequencing of a whole genome and whole exome, NGS platforms can analyze whole transcriptomes (RNA-seq), genome-wide profiles of epigenetic marks (methyl-seq), chromatin immunoprecipitation assays (ChIP-seq) and mitochondrial genome sequence (Pareek et al. 2011).

1.8.5 Whole exome sequencing

For sequencing of genomic DNA, the region to be analyzed can be restricted as in exome sequencing by custom designed targeted capture sequencing. Whole-exome sequencing (WES) is a strategy in which coding sequences of a genome are selectively sequenced. Exons, flanking regions as well as UTR sequences are captured with probes in either liquid or solid phases and subjected to next generation sequencing platforms (Warr et al. 2015). Recently, a reduction in the cost and an enhancement in the accuracy of WES have made it a powerful technique to identify all known protein-coding genes for sequence-level mutations that have accelerated mutation detection. Since around 85 percent of causative mutations reside in protein coding regions, it is worthwhile to investigate mutations in Mendelian disorder through WES (Choi et al. 2009). In addition, WES is considered as an efficient tool in diseases with genetic heterogeneity (Biesecker, 2010). So far, WES has been successfully used to find causative variants of several rare diseases.

An important challenge in WES is the evaluation of the raw data to find the causative mutation. Alignment of exome sequencing reads to the reference genome and subsequent variant calling generates a large number of variants differing from the reference genome. This is particularly challenging as many benign missense and nonsense variants leading to protein truncation are estimated to be present and also for many variants information is absent in databases such as dbSNP, HapMap and 1000 genomes (Lalonde et al. 2010, Ng et al. 2010). In order to fulfill this deficiency, several filters need to be applied in the evaluation of the candidate variants. A recent database, ExAC (Exome Aggregation Consortium) reports the frequency of rare variants in various populations across the globe. At this point, linkage analysis provides great help by assigning certain regions as candidate loci. Thus, it helps in the narrowing down the target regions. In addition, the candidate gene should be involved in a pathway related to the pathogenesis of the disease and should be expressed in the relevant tissues. The overview of approaches used to confer candidate variants generated by exome sequencing data are given in Fig. 1.1.

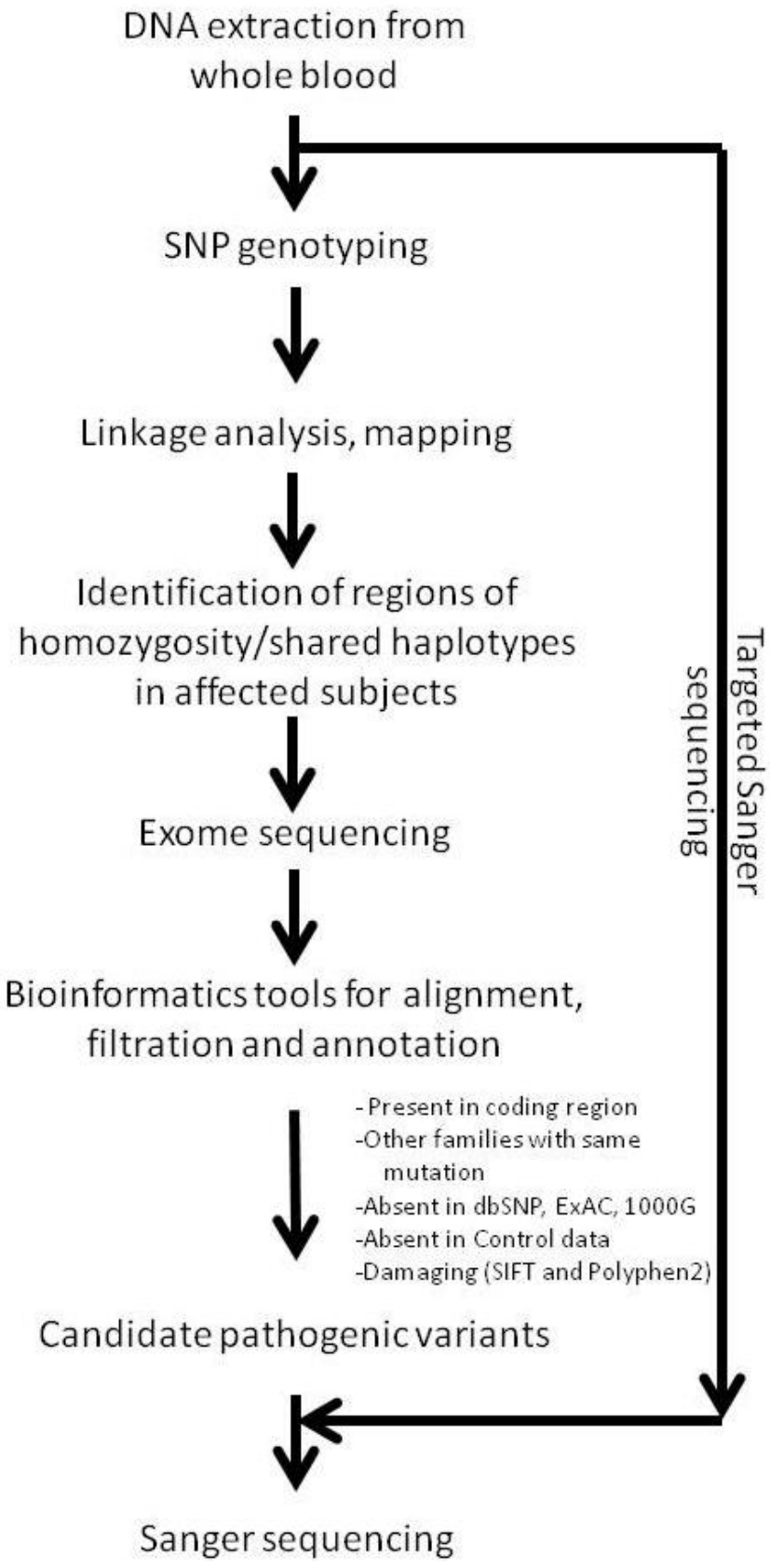


Figure 1.1. The overview of strategy in disease gene hunt used in the study.

1.9 A need for improvement in the diagnosis and genetic studies for heritable diseases in Pakistan

In spite of a high prevalence of genetic disorders in Pakistan, the area of medical genetics is still deficient in the deserved attention. This for instance, necessitates establishing new Medical Genetics Departments in teaching hospitals and medical colleges. It will enhance future progress for the diagnosis and management of these disorders. In addition, highly prevalent genetic disorders such as intellectual disability, Down syndrome, neural tube defects, thalassemia and muscular dystrophies need to be included in the community control programs.

2 AIMS AND OBJECTIVES

The study was intended to identify disease loci and causative mutations in families suffering from inherited diseases. Neurological diseases such as intellectual disability, autosomal recessive primary microcephaly, and microcephaly with a combination of skeletal and ectodermal defects are rare autosomal recessive disorders, whereas polydactyly is an autosomal dominant disease with a relatively high frequency in all populations. In all families, the purpose was first to map the disease loci via linkage analysis using genotyping data generated by SNP genome scan. After the identification of the loci, the aim was to identify the disease genes by evaluating the variants listed in the exome sequencing file within the candidate loci and validating them by Sanger sequencing.

3 SUBJECTS

Eight Pakistani families with congenital inherited disorders of skeletal, skin or intellectual disability were recruited after obtaining written informed consent for all participating members of the study families. Research protocol approval was obtained from the Institutional Review Board (IRB) of Quaid-i-Azam University, Islamabad and Boğaziçi University Institutional Review Board for Research with Human Participants. The families were afflicted with different heritable diseases and were studied with the aim of identifying disease loci and causative mutations.

3.1 Family A: Intellectual Disability with Speech Problem and Dysmorphic Facial Features (IDSD)

This family is large with two branches; in one branch two sibs are affected and in the other four sibs are affected (Fig. 3.1). Peripheral blood samples were obtained from 4 affected and 12 unaffected members of the family. For clinical evaluations, all participants underwent physical and neurological examination by a local physician. MRI of one affected subject was obtained.

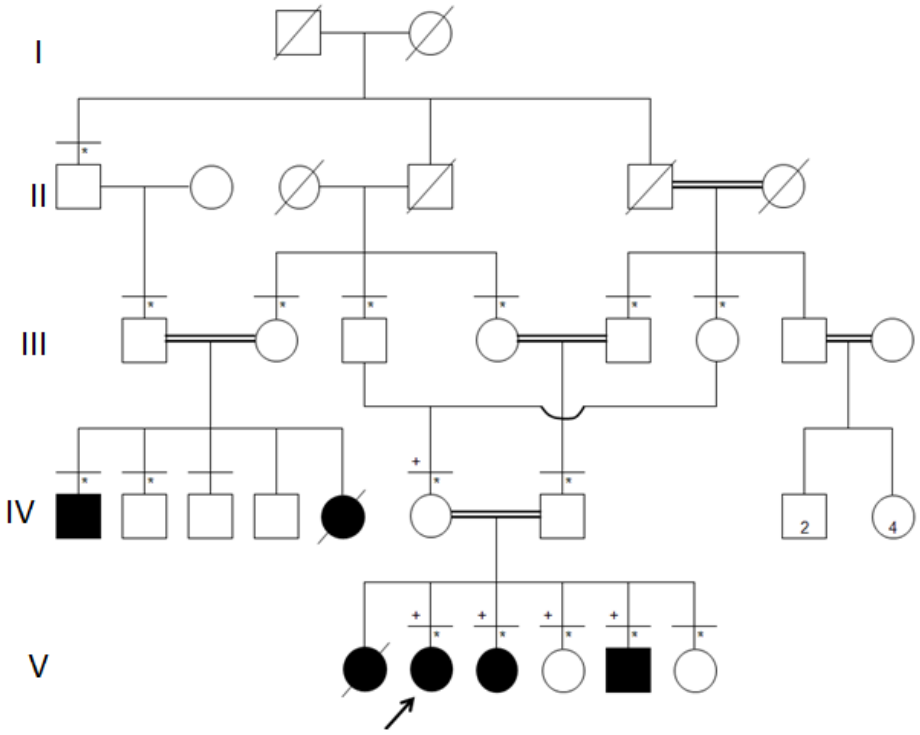


Figure 3.1. **Pedigree of IDSD Family A.** Circles represent females, squares males, black filled symbols patients, and white symbols unaffected subjects. The index case is indicated with an arrow. Asterisks (*) denote DNA availability, and plus signs (+) indicate DNA samples subjected to SNP genotyping. Subjects physically examined are indicated with a horizontal line above the symbol.

3.2 Family B: Autosomal Recessive Primary Microcephaly (MCPH)

In this consanguineous family, a total of four subjects (2 males and 2 females) are afflicted with primary microcephaly. They were photographed, and physical examination and behavioral attributes were assessed with the help of local physicians.

Blood samples of four affected individuals and two unaffected subjects were obtained in EDTA tubes (Fig. 3.2).

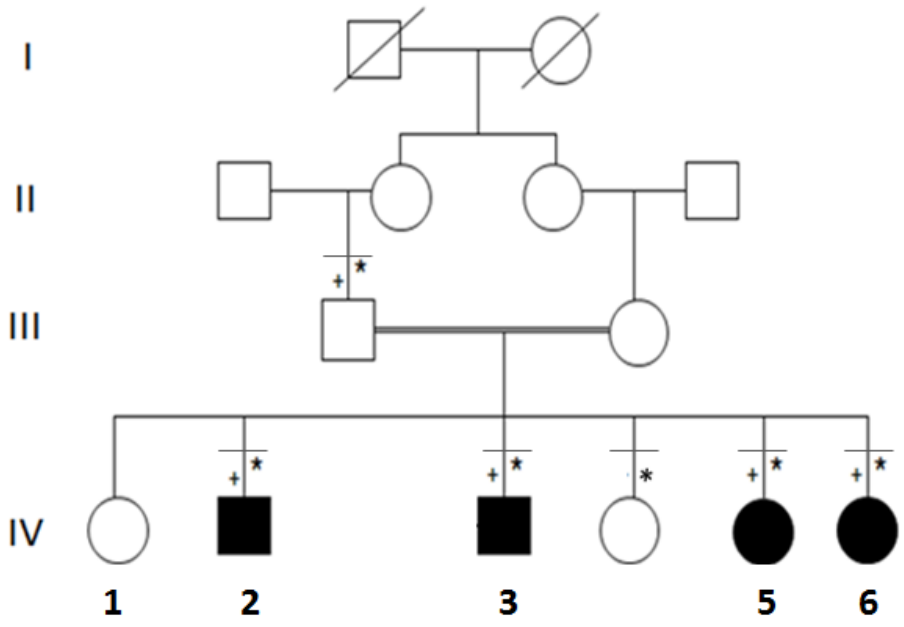


Figure 3.2. **Pedigree of MCPH Family with Microcephaly.** DNA samples subjected to SNP genotyping are indicated with +, and DNA availability is indicated with *. Subjects physically examined are indicated with a horizontal line above the symbol.

3.3 Family C: Microcephaly, Intellectual Disability, Short Stature and Brachydactyly (MIDSB)

The consanguineous family was afflicted with microcephaly combined with skeletal and skin problems. According to family elders 7 individuals were affected (3 males and 4 females) in two branches of the family (Fig. 3.3). All affected subjects were born to consanguineous parents. Clinical examinations of affected sibs were performed with the help of local physicians. Peripheral blood samples were obtained from four affected and two unaffected sibs and the parents in one branch of the family. One affected individual underwent X-ray imaging and cranial computed tomography (CT) scanning.

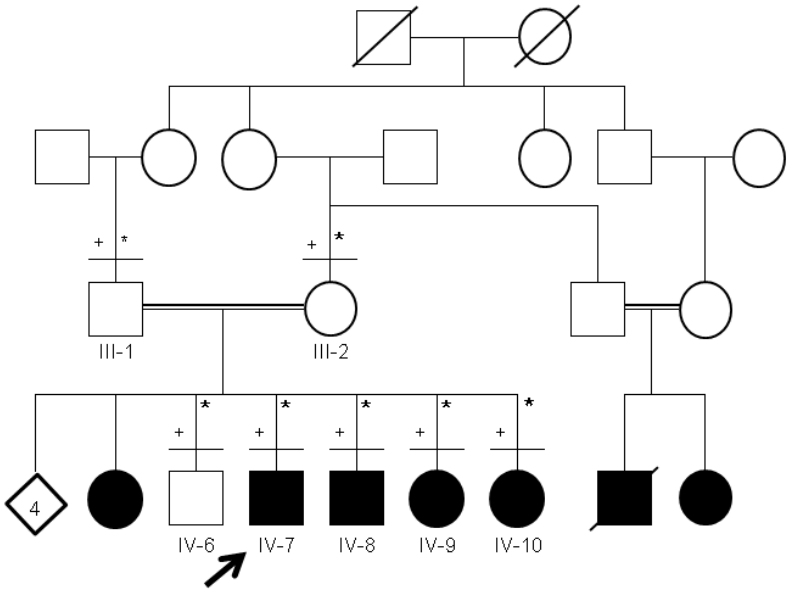


Figure 3.3. **Pedigree diagram for MIDSB Family with Syndromic Microcephaly.** DNA samples subjected to SNP genotyping are indicated with +, and DNA availability is indicated with *. Subjects physically examined are indicated with a horizontal line above the symbol.

3.4 Family D: Autosomal Dominant Postaxial Polydactyly, Camptodactyly and Zygodactyly (ADPCZ)

The large family was afflicted with variable polydactyly. The pedigree of the family containing four generations was drawn with the help of senior family members (Fig. 3.4). A total of 21 affected relatives, 14 males and 7 females, were known, and the anomaly was known to run in three generations. There was no skip generation, and all affected subjects had at least one affected parent, strongly suggesting an autosomal dominant mode of inheritance. The available affected and unaffected subjects underwent a thorough physical examination with the help of a local physician, and clinical information including photographs and X-rays were obtained.

Peripheral blood samples were obtained from 9 affected and 5 unaffected members of the family.

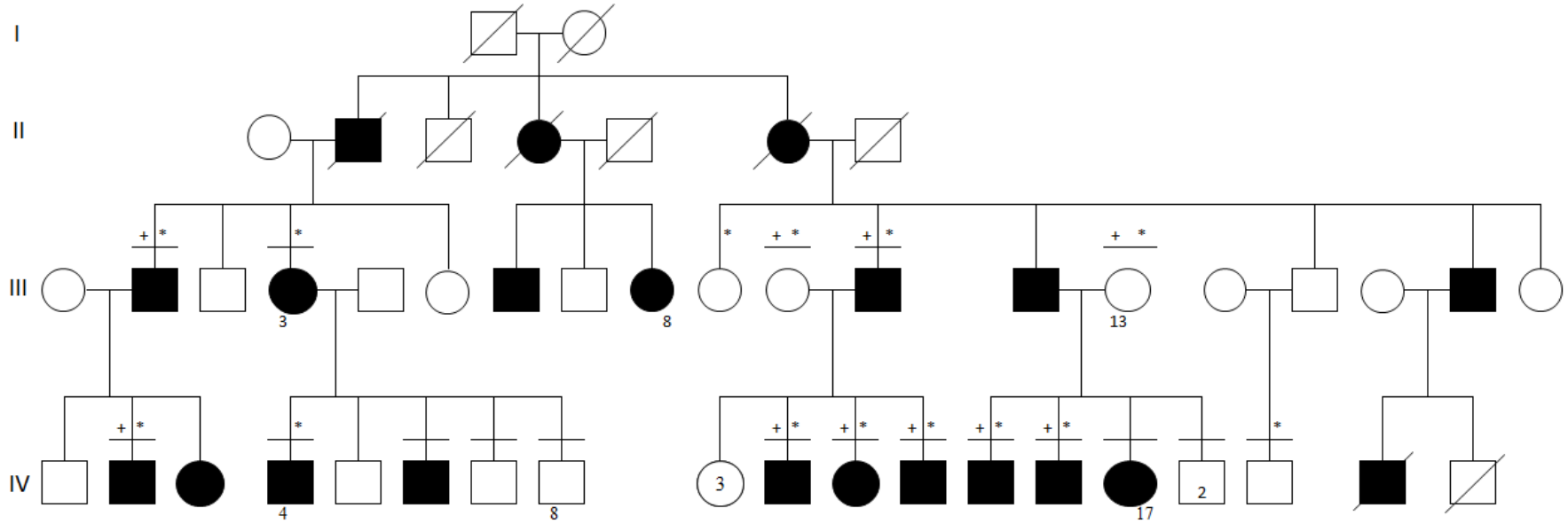


Figure 3.4. **Pedigree of the Family with Polydactyly.** DNA samples subjected to genome scan are indicated with +, and DNA availability is indicated with *. Subjects physically examined are indicated with a horizontal line above the symbol.

3.5 Family E: Tooth Agenesis Manifesting as Oligodontia/ Hypodontia (TAOH)

A large family from the suburbs of Islamabad was studied. The field visits were coordinated with the help of a coworker. A detailed pedigree was constructed in several visits to the family. All the information was checked and rechecked by interviewing various family members particularly the elders of the family. A few branches of the family were settled in Khyber Pakhtunkhwa. They were visited in the respective towns to obtain information. Nine affected and ten unaffected subjects were physically examined. There was a hereditary dental problem in several generations. Standard oral examination was conducted and dental scheme/dentition was recorded. Oropantomograms (OPG) and cephalometric radiographs (Ceph) were obtained for four affected and two unaffected subjects.

The affected individuals had several missing teeth in both the upper and lower jaws causing wide spaces between the existing ones, i.e., diastema. It was mainly an aesthetic problem and became apparent when they talk. Some of the affected subjects had artificial teeth put in for cosmetic purposes. They had otherwise no problem in eating and chewing. Additionally, no other associated abnormality of hair, nail and skin was observed in the affected subjects. All the affected individuals were of normal intelligence and productive members of the society.

The pedigree comprised five generations. All the marriages in pedigree were non-consanguineous (Fig. 3.5). Thirteen subjects (6F, 7M) were found to be affected in five consecutive generations. In all sibships, the affected subjects had at least one affected parent (I-1 x I-2; II-1 x II-2). This situation strongly suggested an autosomal dominant mode of inheritance.

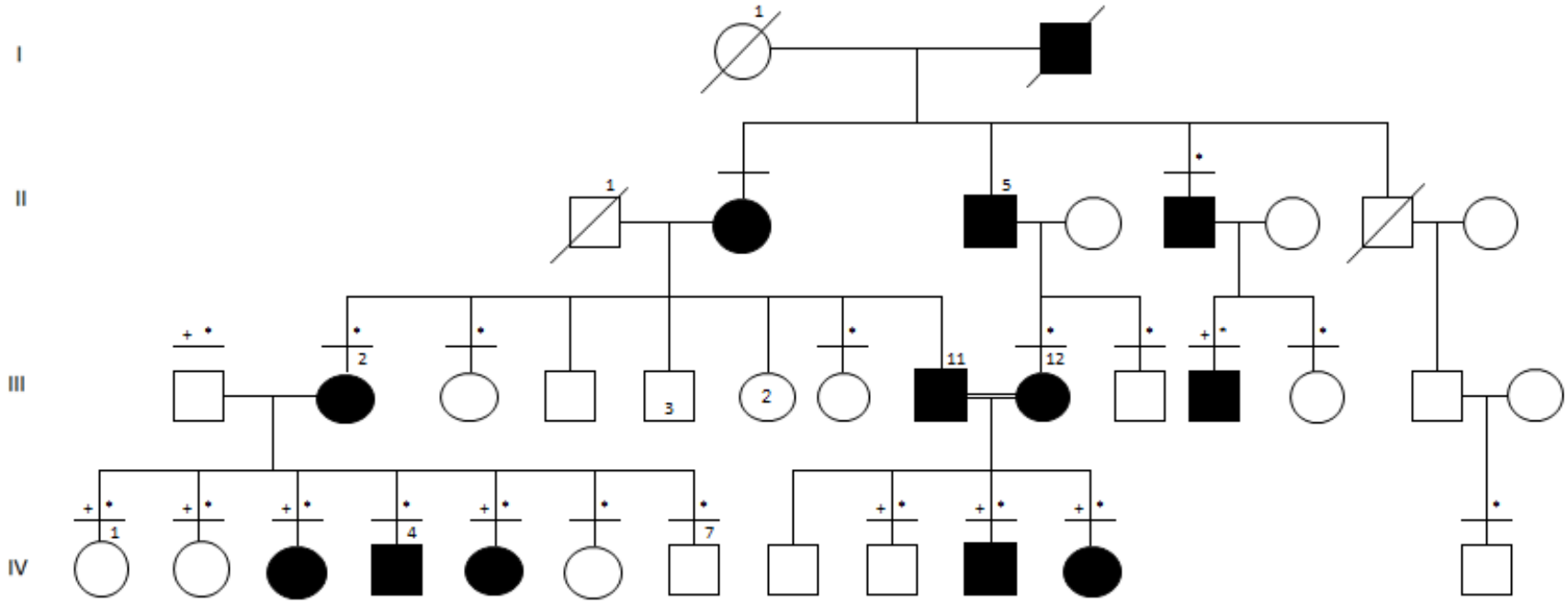


Figure 3.5. **Pedigree of the Family with tooth agenesis.** DNA samples subjected to genome scan are indicated with +, and DNA availability is indicated with *. Subjects physically examined are indicated with a horizontal line above the symbol.

3.6 Family F: Seckel-like Syndrome (SLS)

The family is settled in Federal Capital Islamabad. A detailed pedigree was constructed with the help of elders of the family and all the information was confirmed by interviewing various family subjects belonging to different loops of the pedigree. All the phenotypic information and clinical detail were recorded on standard data sheets. Photographs of three affected and one unaffected subject were obtained. Radiographs were obtained from one affected male and one normal female subject.

The pedigree comprised four generations. The affected subjects were observed to have a skeletal anomaly affecting the limbs, thorax and skull (Fig. 3.6). Three sibs were afflicted with this anomaly. Generally, the affected subjects had a thin build and weak physique but were otherwise self-confident and were relatively independent in their daily activities. All three affected individuals had normal intelligence and were getting formal education. The transmission pattern of this abnormality was consistent with an autosomal recessive mode of inheritance. One of the female sib in the family was deaf and mute while another sib had a curvature of backbone (kyphoscoliosis), but without any typical symptoms of Seckel syndrome. Hence, both of the latter two subjects were considered unaffected for the Seckel-like syndrome.

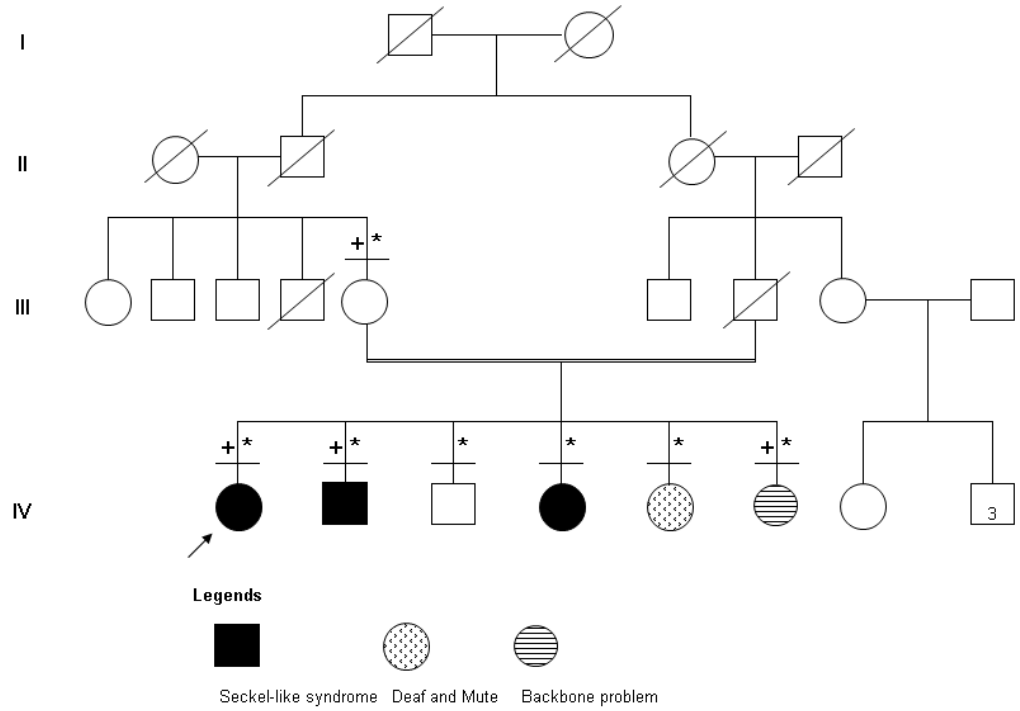


Figure 3.6. **Pedigree of the Family with Seckel-like syndrome.** DNA samples subjected to genome scan are indicated with +, DNA availability is indicated with *, and subjects underwent physical examination are indicated by a horizontal bar above symbols.

3.7 Family G and H: Syndactyly type IX, Mesoaxial Synostotic Syndactyly (MSSD)

Family G originates from southern Pakistan. The parents were first-cousins, and the single affected female was the first in a sibship of five (Fig. 3.7A). There was no previous history of limb or any other anomaly in the family. The pedigree was suggestive of autosomal recessive inheritance. The clinical data of the subject was obtained with the help of a local physician. The family was not willing to provide photographs, hence, line-drawing of the hands was obtained. Peripheral blood samples of the affected subject and her father were obtained for the molecular study.

Family H originates from northern part of Pakistan. The pedigree was constructed with the help of the elderly family members. There were two affected subjects (one female, one male) in the fourth generation, and the female had been deceased (Fig. 3.7B). The male affected subject underwent a clinical examination with the help of local physicians. Blood sample of only the affected subject was available for molecular study.

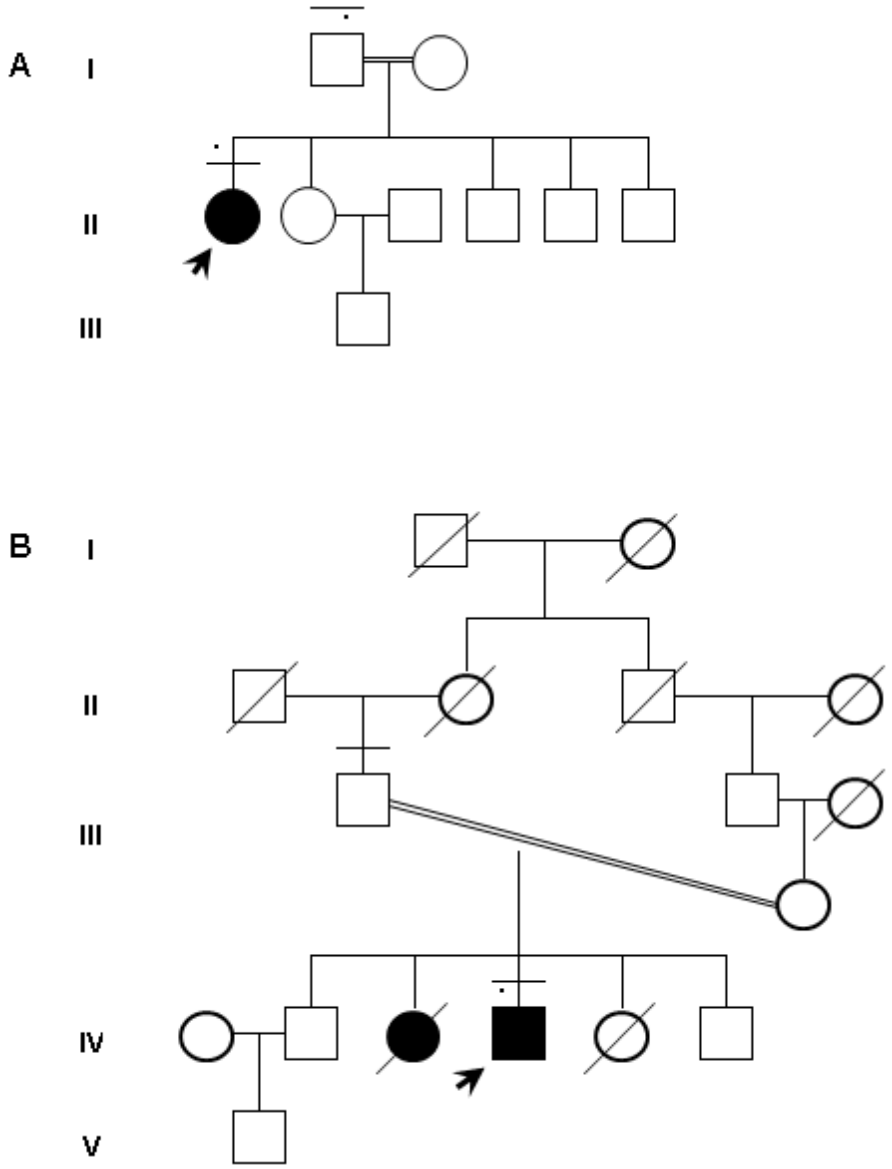


Figure 3.7. **Pedigrees of the MSSD Families. A.** Family G. **B.** Family H. Individuals subjected to molecular analyses are indicated with *. Subjects physically examined are indicated with a horizontal line above the symbol.

3.2 Chemicals

For this study, all liquid and solid chemicals were purchased from Sigma (USA), Merck (Germany), Carlo Erba (Italy), Riedel-de-Haen (Germany) or Biochar (Germany).

3.3 Solutions and buffers

3.3.1 Processing of peripheral blood samples for DNA extraction

Solutions	**Reagents**
Cell Lysis Buffer	10 mM $KHCO_3$, 155 mM NH_4Cl, 0.1 mM Na_2EDTA (pH 7.4)
Nucleus Lysis Buffer	400 mM NaCl, 2 mM Na_2EDTA, 10 mM Tris (pH 8.2)
Proteinase K Enzyme	20 mg/ml Proteinase K in dH_2O (Sigma, USA)
Ammonium Acetate	7.5 M CH_3COONH_4 in dH_2O
Sodium dodecylsulfate (SDS)	10% SDS (w/v) in dH_2O
Ethanol	Absolute Ethanol
TE buffer	20 mM Tris-HCl, 1 mM EDTA (pH 8.0)

3.2.2 Polymerase Chain Reaction (PCR)

Solutions	**Reagents**
Agarose	2% Agarose (Sigma, USA) in 0.5 X TBE buffer
10X TBE Buffer	20 mM EDTA, 0.89 M Boric Acid, 0.89 M Trizma base (pH 8.3)
6 X Loading Buffer	10 mM Tris-HCl (pH 7.6), 50% Glycerol, 60 mM EDTA, 2.5 mg/ml Bromophenol Blue
Ethidium Bromide	10 mg/ml in dH_2O

3.3.3 Agarose gel electrophoresis

Solutions	**Reagents**
Agarose	2% Agarose (Sigma, USA) in 0.5 X TBE buffer
10X TBE Buffer	20 mM EDTA, 0.89 M Boric Acid, 0.89 M Trizma base (pH 8.3)
6 X Loading Buffer	10 mM Tris-HCl (pH 7.6), 50% Glycerol, 60 mM EDTA, 2.5 mg/ml Bromophenol Blue
Ethidium Bromide	10 mg/ml in dH_2O

3.3.4 Single Strand Conformational Polymorphism (SSCP) analysis

Solutions	**Reagents**
40% Acrylamide	40% Acrylamide-bisacrylamide (37.5:1) (Stock) in dH_2O
8% Acrylamide	8 or 10% Acrylamide-bisacrylamide (37.5:1, non-denaturing) in 0.6 X TBE buffer
Glycerol	8% Glycerol in gel solution
APS	10% Ammonium Peroxydisulfate
TEMED	N,N,N,N-tetramethylethylenediamine
10X Sample Buffer	95 per cent formamide, 20 mM EDTA, 0.05% xylene cyanol, 0.05 per cent bromophenol blue

3.3.5 Silver staining

Solutions	Reagents
Staining Buffer	0.1% $AgNO_3$ in dH_2O
Developing Buffer	1.5% NaOH, 0.01% $NaBH_4$, 0.015% formaldehyde in dH_2O

3.4 Enzyme

Taq DNA polymerase was purchased from Roche (Germany) or Fermantas, Dream Taq (Lithuania) supplied with PCR buffer.

3.5 Oligonucleotide primers

Primer3 software was used for oligonucleotide primer designing, and OligoCalc was used for evaluation of secondary structures and self-annealing sites (Table 3.1). Primers were purchased from Macrogen. Lyophilized primers were liquefied in 1 ml distilled H_2O, and 10 μM dilutions were prepared to use in PCR reactions.

3.6 DNA molecular weight markers and commercial kits

Name	**Company**
Lambda DNA/*Hind*III marker	Fermentas
pUC19 DNA/*Msp*I marker	Fermentas
GC-Rich PCR System	Roche
Dream Taq PCR kit	Fermentas
Gel Extraction kit	Qiagen
PCR Purification kit	Qiagen

3.7 Equipment and apparatus

Item	**Company**
Electronic Balance	Precisa
AMB430T Autoclave	Astell
MiniSpin Plus Centrifuge	Eppendorf
Universal 16R Centrifuge	Hettich
Allegra X-22R Centrifuge	Beckman Coulter
J2-MC Centrifuge	Beckman Coulter
-20°C Deep Freezer	Arcelik
-80°C Ultra Freezer	Thermo Scientific
Gel Doc Documentation System	BioRad
Horizontal DNA Electrophoresis Equipment	BioRad
DCode Universal Mutation Detection System	BioRad
Orbital Incubator	Gallenkamp
MR3001 Magnetic Stirrer	Heidolph
Micropipettes	Gilson
Rotamax 120 Minishaker	Heidolph
KD 200 Oven	Nuve
Fotoforce 250 Electrophoresis Power Supply	Fotodyne

P250A Power Supply	Sigma-Aldrich
Power Pac Model 3000	BioRad
NanoDrop 1000 Spectrophotometer	Thermo Scientific
PTC-200 Thermal Cycler	MJ Research
T100 ThermalCycler	BioRad
480 LightCycler	Roche
Reax vortex mixer	Heidolph
Lab Dancer Vario Vortex	Roth
Grant LTD 6G Thermostatic Water Bath	Grant
Ultra-Pure Water Purification system	Hach-Lange

3.8 Electronic databases

Table 3.1. Electronic databases and tools used in the study.

Name	Description	Homepage
Online Mendelian Inheritance In Man (OMIM)	A database of human diseases and traits, and human genes.	http://www.ncbi.nlm.nih.gov/Omim/
NCBI Genome Resources	A database of sequences and maps of reference genomes.	http://www.ncbi.nlm.nih.gov/genome/
UCSC Genome Browser	A database of sequences and maps of a large number of reference genomes.	http://genome.ucsc.edu/
ENSEMBL	A database of sequences and maps of a large number of reference genomes.	http://www.ensembl.org/
NCBI UniGene	Expression data of a wide collection of genes in various tissues.	http://www.ncbi.nlm.nih.gov/unigene/
UniProt (Universal Protein Resource)	A comprehensive resource of protein sequence and functional information.	http://www.uniprot.org/
Gene Cards	A database unifying the information about human genes on genomic, proteomic and functional level.	http://www.genecards.org
Polymorphism Phenotyping	Predicts the structural and functional effects of amino acid	http://genetics.bwh.harvard.edu/pph2/

(PolyPhen-2)	substitutions on human proteins.	
SIFT (Sorting Intolerant From Tolerant)	Calculates position specific scores for amino acids. The scores range from 0 to 1, which represents the effect from damaging to neutral	http://sift.jcvi.org/
Molecular Modelling & Bioinformatics (MMB)	Predicts the pathological effects of nonsynonymous mutations. Mutations with scores greater that 0.5 are pathological.	http://mmb2.pcb.ub.es:8080/PMut
SNPs3D	Uses structure and sequence analysis to assess functional effects of non-synonymous SNPs.	http://www.snps3d.org/
Biology Workbench	An online tool package for storing and analyzing DNA sequences. Contains Primer3 software.	http://workbench.sdsc.edu
OligoCalc (Oligonucleotide Properties Calculator)	Calculates the melting properties and predicts potential hairpin structures or self-annealing sites of primers.	http://www.basic.northwestern.edu/biotools/oligocalc.html
Human Splicing Finder 2.4.1	An online tool that calculates consensus values of potential splice sites and searches for branch points in given sequences. It also predicts the effect of a sequence change on pre-mRNA splicing.	http://www.umd.be/HSF/
Database of Genomic Variants (DGV)	A database collecting structural variations in the human genome. Genomic alterations affecting fragments of DNA that are larger than 1kb are annotated as structural variations. Alterations in 100bp-1kb in size are annotated as InDels.	http://projects.tcag.ca/variation/

3.9 Bioinformatics tools

For analysis of exome sequencing data, quite a few bioinformatics programs were used that are listed in Table 3.2. These programs are open source and easily accessible on the internet.

Table 3.2. Bioinformatics tools employed to analyze exome sequencing data.

Name	**Description**	**Home page**
BWA (Burrows-Wheeler Aligner)	A program for aligning short nucleotide sequences against a reference sequence, such as the human genome.	http://bio-bwa.sourceforge.net/
SAMTools	A program for manipulating alignments in the SAM (Sequence Alignment/Map) format, including sorting, indexing, and variant calling.	http://samtools.sourceforge.net/
GATK (Genome Analysis Tool Kit)	A software library of next-generation sequencing data analysis tools, including depth of coverage analysis, SNP/indel calling and local realignment.	http://www.broadinstitute.org/gsa/wiki/index.php/The_Genome_Analysis_Toolkit
ANNOVAR	A program performing functional annotations of genetic variants obtained from next-generation sequencing data. The annotations can be gene-based, region-based or filter-based.	http://www.openbioinformatics.org/annovar/
Picard	A program comprising of Java-based command-line utilities for manipulating SAM files.	http://picard.sourceforge.net/
BEDTools	A suite of utilities for comparing genomic features such as finding feature overlaps and computing	http://code.google.com/p/bedtools/

	coverage in next-generation sequencing data.	
BamView	An interactive program for visualizing read alignments in BAM files.	http://bamview.sourceforge.net/
cnvPartition	cnvPartition was developed by Illumina and is available as a plug-in in the Beadstudio software. We used default parameters of cnvPartition; i.e., gen call threshold =0.15; call rate ≥95%; confidence threshold 35; minimum probe count 3.	https://www.illumina.com/documents/products/technotes/technote_cnv_algorithms.pdf
CNVision	A program for merging, annotating and visualizing CNVs, using SNP genotyping data obtained from Illumina GenomeStudio.	http://futo.cs.yale.edu/mw/index.php/

4 METHODS

4.1 Processing of human peripheral blood samples for genomic DNA extraction

For the extraction of genomic DNA, peripheral blood samples in sterile K_2EDTA tubes were processed. Pre-chilled 30 ml cell lysis buffer was poured onto 10 ml of blood and mixed by gently inverting the tube several times. This mixture was kept at 4°C for 10 minutes (min) to ensure osmotic swelling and cell lysis. After this incubation, the tube was centrifuged at 5,000 revolutions per minute (rpm) at 4°C. For harvesting the pellet of white blood cell nuclei, the supernatant was carefully thrown away. The pellet was suspended in cold 10 ml of cell lysis buffer and subjected to centrifugation at 4°C for 10 min, and the subsequent supernatant was poured out. The pellet was again suspended in 0.3 ml nucleus lysis buffer per ml of initial blood volume (i.e., 3 ml for 10 ml blood), and the tube was vortexed gently to entirely dissolve the pellet. To digest the residual proteins, 50 µl of Proteinase K (25 mg/ml) and 80 µl of 10% SDS were added, and the sample was mixed by gentle inversion. This mixture was kept in an incubator at 37°C overnight. After incubation, 2.8 ml of NH_4Ac (0.5 M) was added and mixed by vigorous shaking. Centrifugation was done at 12,000 rpm for 20 min at 4°C to separate the precipitated proteins in the form of a pellet. The supernatant was poured into a 50 ml tube, and two volumes of ice-cold ethanol were added. DNA was precipitated out by gentle mixing. The floating white

thread of DNA was fished out with a micropipette tip and transferred to a clean tube. The tube was left open until the ethanol evaporated. DNA was dissolved by adding TE buffer (400 µl) and stored in the freezer.

4.2 Linkage analysis

Linkage analysis is a statistical test that associates disease genes with a locus by calculating LOD (logarithm of odds) scores. For performing parametric linkage analysis, the family pedigree, the SNP genome scan data of family members included in the study and the genetic map of the markers used are required. The analyses performed for each family are described below.

4.2.1 SNP genotyping

A genome scan was performed at Macrogen Inc (South Korea) with Illumina Human OmniExpress-24 BeadChip including 713, 014 SNP markers at a median marker spacing of 4.5 kb. For handling genotyping data, Illumina Genome Studio v.1.02 was used. In IDSD family (Family A), a genome scan was performed for three affected (1 brother and 2 sisters) and an unaffected sister and one unaffected parent. In the second family (Family B) with Autosomal Recessive Primary Microcephaly (MCPH), a genome scan was performed for one parent and four affected sibs. In the third family (Family C) with Syndromic Microcephaly, a genome scan was performed for four affected (2 brothers and 2 sisters) and an

unaffected brother. In the fourth family (Family D) with polydactyly, a genome scan was carried out on 10 subjects (8 affected and 2 unaffected). In the fifth family (Family E) with tooth agenesis, a genome scan was carried out on five affected and four unaffected subjects. In the sixth family (Family F) with the Seckel-like syndrome, a genome scan was carried out on two affected and two unaffected subjects. In the seventh and eights families (Family G, Family H), direct Sanger sequencing was performed.

4.2.2 Parametric linkage analysis, homozygosity mapping and haplotype analysis

For parametric LOD score calculations, easyLINKAGE package version 5.08 was used for all families. This package is an open source GUI (graphical user interface) that comprises the programs shown in Table 4.1.

Table 4.1. The programs included in easyLINKAGE software.

Program	Description
PedCheck	Detects genotyping errors and checks the pedigree file for Mendelian incompatibilities.
SimWalk2	Calculates parametric LOD scores and constructs haplotypes for any size of pedigree, including complex families (bit size > 19).
SuperLink	Calculates two-point parametric LOD scores.
Allegro	Calculates both two-point and multipoint parametric LOD scores in small pedigrees (bit size <19) and constructs haplotypes.
GeneHunter	Calculates multipoint parametric LOD scores and constructs haplotypes in small pedigrees (bit size <19).
SPLink	Performs linkage analysis using affected sib pairs, with the method of maximum likelihood.

easyLINKAGE package is available at:

(http://nephrologie.uniklinikum-leipzig.de/nephrologie.site,postext,easylinkage,a_id,372.html)

Homozygosity mapping was performed in autosomal recessive families by Homozygosity Mapper (http://www.homozygositymapper.org) and homozygous regions >100 kb in patients were detected. For the identification of homozygous

regions identical by descent (IBD), haplotypes were constructed by ALLEGRO. In addition, Homozygosity Comparison in Excel (HCiE) that was developed in the Turkish laboratory was also used (Cetinkaya, 2010). HCiE is an MS Excel macro, in which three different SNP genotypes (AA, BB or AB) in the subjects are differently colored, and the markers are sorted in the order of their locations. This allows the user to detect shared homozygosity among the affected individuals as well as to observe whether the subjects carry the same genotype within the homozygous regions. In the case of dominant inheritance, haplotypes are constructed to detect shared chromosomal intervals segregating in the affected individuals.

Multipoint LOD score calculations were performed via program ALLEGRO, GeneHunter and Marline. In the case of families with autosomal recessive inheritance, disease frequency was set to 0.0001 and full penetrance was assumed. Markers were selected at 0.01 cM spacing and used in sets of 30. The physical and genetic maps of markers were provided by MacroGen Inc. Haplotypes were constructed using program ALLEGRO. Homozygosity was confirmed at the candidate loci by HCiE (Cetinkaya, 2010). In the case of polydactyly, an autosomal dominant model with a disease allele frequency of 0.001 and a penetrance of 99% were assumed. Multipoint LOD score calculations were performed with a marker window size of 30 and 70. The LOD score intervals with >2.0 scores were examined more closely and re-evaluated with iterated models of reduced penetrance and varying marker window sizes. Haplotypes were constructed to identify shared chromosomal intervals segregating in the affected individuals.

4.3 Copy number variation (CNV) analysis

It is possible to detect copy number variations (CNVs) using high-density SNP arrays (McCarroll et al. 2008). Based on this finding a PERL based program called Combined_CNV has been developed by Chris Mason and Stephan Sander at the State Lab, Yale University and provided to our laboratory of Dr. Tolun. We used this program for our study families to detect duplication and deletions. Combined_CNV (http://futo.cs.yale.edu) uses three CNV detection algorithms, namely, PennCNV (Wang et al. 2007), QuantiSNP (Colella et al. 2007) and GNOSIS, for investigating high-density SNP genotyping data. CNV detection relies on the LogR ratio values, which is logarithm (base 2) ratio of the normalized intensity (R value) for the SNP tested, divided by the expected normalized intensity. Combined_CNV program takes Final Reports generated by Illumina GenomeStudio as input files and outputs annotated files containing the genomic locations and flanking SNPs of CNVs. The program also annotates CNVs previously reported in Database of Genomic Variants (DGV).

4.4 Whole exome sequencing

4.4.1 Targeted exome capture

Exome sequencing was performed at MacroGen Inc (South Korea) or Yale Center for Genome Analysis as a paid service. DNA sample of only one affected subject from each family was subjected to exome capture with Illumina TruSeq Exome Capture kit (Illumina, USA), followed by massively parallel paired-end sequencing with Illumina HiSeq 2000 (Illumina, USA). The features of TruSeq Exome Capture kit are given in Table 4.2.

Table 4.2. Features of Illumina TruSeq Exome Capture kit.

Features	**No.**
Capture target	62 Mb
Number of target genes	20,794
Number of target exons	201,121
Number of probes	340,427
Size of probes	95 bp

Obtaining sequence information from exome sequencing occurs in various steps that generate a list of variants within or near the coding regions of the genome. First, Illumina TruSeq DNA Sample Preparation kit (Illumina, USA) was used for the preparation of a sample library. For this procedure, 5-10 µg of genomic DNA was used as starting material. The aim of this step is the generation of multiplexed paired-end sequencing libraries by adding adapter sequences to the DNA fragments. In this step, firstly the genomic DNA is sheared and overhangs are converted to blunt ends to obtain double stranded DNA fragments of an average of 300-400 bp in size. In order to prevent ligation of blunt fragments to one another, the 3' ends are sealed with a single "A" nucleotide. Multiple indexing adapters having a "T" nucleotide at the 3' end are used for ligation of blunt fragments. After the ligation of the adapters, biotinylated capture probes of targeted regions are hybridized with blunt fragments, and unbound fragments are washed off. Streptavidin beads pull out the biotinylated capture target, and the fragments are hybridized onto the flow cell, which is a solid surface containing forward and reverse primers in spatially separated clusters. With these primers, the captured target is amplified by bridge PCR in clusters, and subsequent sequencing reaction takes place.

4.4.2 Analysis of the exome sequencing file

Macrogen service included bioinformatics analysis of exome sequencing data with standard parameters by using programs such as BWA and SAMTools. An annotated list of variants was obtained from the company in an MS Excel file. However, we also performed bioinformatics analysis with more relaxed parameters to elude false negative calls. Paired-end ".fastq" files of raw data produced after sequencing was obtained from the company.

Human reference genome sequence (assembly GRCh37/hg19) was downloaded from UCSC Genome Bioinformatics Site (described in Table 3.1). Individually assembled chromosomes were concatenated to each other and indexed. Paired-end reads were aligned to the reference genome with program BWA and the final alignment was generated in SAM (Sequence Alignment/Map) format. The SAM file was converted to BAM (Binary Alignment/Map) format using SAMTools. The resulting file was sorted, indexed and subjected to SNP calling with SAMTools, which outputs a list of nucleotides that were different from the reference genome. The list of variants were annotated with ANNOVAR, which denotes chromosomal location, dbSNP ID, the type of the region of the nucleotide change (exonic, intronic, UTR, non-coding RNA or intergenic) and the type of the exonic nucleotide change (synonymous, non-synonymous, frameshift, stopgain or stoploss) and whether a nucleotide change is a splicing mutation. For

visualizing the alignment, BamView program was used. The command lines for the steps are given in Table 4.3.

Table 4.3. Command lines used in bioinformatics analysis of exome sequencing data and their purposes.

Purpose	**Command**
Concatenating all chromosomes of the reference genome	cat *.fa > all_chr.fasta
Indexing the reference genome	bwa index –a bwtsw all_chr.fasta
Aligning paired-end reads to reference genome (performed for each .fastq file separately)	bwa aln –n 0.01 –t 8 all_chr.fasta file1.fastq > file1.sai
Generating alignment in the SAM (Sequence Alignment/Map) format	bwa sampe all_chr.fasta file1.sai file2.sai file1.fastq file2.fastq > file.sam
Converting SAM to BAM (Binary Alignment/Map) format	samtools view –bS file.sam > file.bam
Sorting the BAM file	samtools sort file.bam
Indexing the BAM file	samtools index file.sorted.bam
SNP calling	samtools pileup –vcf all_chr.fasta file.sorted.bam > file.pileup.txt
Converting the file to a format compatible with ANNOVAR	convert2annovarl.pl file.pileup.txt –outfile file_annovar.pileup.txt
Region based annotation	annotate_variation.pl –buildver hg19 file_annovar.pileup.txt humandb

Filter based annotation	annotate_variation.pl –filter –dbtype snp131 file_annovar.pileup.txt
Visualization of the alignment	java –mx152m –jar BamView_v1.1.8.jar –a file.sorted.bam –r all_chr.fasta
Computing the coverage for each exon included in the target capture using BEDTools	bamToBed –i file.bam > file.bed coverageBed –a file.bed –b targetedRegions.bed > file_coverage.txt

4.5 Candidate genes and mutation screening

Among the variants detected by exome sequence analysis, the mutations within loci that stood out as the best candidates were selected. In selecting the candidate mutations, the variants that had high coverage and high-quality scores as well as predicted adverse effect of the mutation on the protein were considered with priority. The threshold for sufficient coverage was assumed as four reads, and the minimum quality score accepted was 40. Variants not fitting into these conditions were regarded more likely as false positives. Nonetheless, all variants were taken into consideration. The variants that were reported in ExAC database were eliminated if their frequency is equal to or more than 0.005, and the remaining rare and novel variants were prioritized with respect to severity of the mutation, expression in relevant tissues, gene function and indication in animal models.

Upon identifying a rare deleterious mutation in the exome sequencing results, the candidate mutation was validated by Sanger sequencing. Mutation screening for the entire family and for at least 200 controls was carried out by either Sanger sequencing or single-strand conformation polymorphism (SSCP).

4.5.1 PCR amplifications

Sites of all strong candidate variants resulting from the exome sequence filtering were subjected to PCR amplification in relevant individuals for subsequent DNA sequencing. PCR for each primer pair was carried out using 1X PCR buffer, 0.2 µM of each dNTP, 0.4 µM of each primer, 30 ng genomic DNA, 0.2 U Taq DNA polymerase and water adding up to a total volume of 25 µl. PCR conditions and primer sequences are given in Table 4.4.

Table 4.4. Sequences, PCR product sizes and PCR conditions for primer pairs amplifying the candidate variants obtained from exome sequence results.

Gene / Region	**Primer Sequence (5'→ 3')**	**Product Size (bp)**	**Buffer and additives**	**Annealing Temperature (°C)**
PDIA3 **(IDSD Family)**				
Exon 2 (c.170G>A)	F: GGAAGTGTCTACTAGCTCAAAGG R: CCATGTAACAAAGCTGAGACAAC	272	Buffer A	60
RBBP8 **(MCPH Microcephaly Family)**				
Exon 11 (c.1808_1809delTA)	F: GGAATGCATCATCCTTCAGC R: GCTCATGAGAACCAGCACTCT	245	Buffer A	58
GLI3 **(ADPCZ Polydactyly Family)**				
Exon 15 (c.3635delG)	F: GCCTTACAGGGCTGTTCAT R: TCCTCCTCCAAGCTCAAGT	236	Buffer A	59
BHLHA9 **(MSSD Syndactyly Type IX Family)**				
Exon 1 (c.218G>C ; c.211A>G)	F:GAGGGAGATGGATGGACGGGA R:GCAGTCCCGGGTTTTATAGTC	970	Buffer A	60

4.5.2 Analysis of PCR products

For all PCR reactions, the efficiency of amplification and the size of the products were checked on agarose gels. PCR product (5 µl) was mixed with loading dye (1 µl) and loaded on a 2% agarose gel containing ethidium bromide. For electrophoresis 0.5X TBE was used as running buffer at 10 volts/cm for 20 minutes. To visualize DNA fragments, UV light was used in BIORAD Universal Hood II and Quantity One 4.6.9 software.

4.5.3 DNA sequence analysis

All regions to be analyzed with Sanger sequencing were amplified by PCR in 50-µl reactions. Total volume per sample was 50-200 µl, depending on the efficiency of the reaction. These samples were sent to MacroGen Inc. (South Korea), where the sequence analysis was performed for all our projects.

4.5.4 SSCP analysis for mutation screening

Polyacrylamide gels with crosslinking ratios of 2%, with or without 8% glycerol, were prepared. Glass plates of sizes 20 cm x 20 cm were assembled

using 0.75 mm spacers. Six ml of 40% acrylamide:bisacrylamide stock (37.5:1) was mixed with 1.8 ml of 10 X TBE, and the volume was adjusted to 30 ml by dH_2O. For gels with glycerol, 2.4 ml of glycerol was added and volume was adjusted to 30 ml. Then, 10% APS (250 µl) and TEMED (25 µl) were added and were gently mixed, and the solution was poured between the glass plates. A comb with 20-well comb was inserted into the gel, and the gel was left to polymerize for at least one hour. After polymerization, the gel was cooled at 4°C for at least half an hour before use. In Table 4.4, the sequences of the primers used in PCR amplifications are given. PCR products and sample buffer were mixed with equal volumes, and DNA was denatured at 95°C for 5 min. Then it was immediately placed on ice. Appropriate volume (7-12 µl) of all samples were loaded onto the gel and subjected to electrophoresis at a constant power of 10 to 30 Watts in 0.6X TBE buffer at 4°C in BioRad DCode Universal Mutation detection system. The duration of the electrophoresis was adjusted according to the length of the PCR product. After electrophoresis, silver nitrate was used for staining gels and visualizing DNA. Samples with aberrant migration patterns were subjected to Sanger sequencing.

4.5.5 Silver staining

Glass gel plates were gently separated, letting the gel remain on one of them. The gel was detached from the glass by lifting with a piece of filter paper and transferred to a tray containing 300 ml of staining buffer. After incubation in staining buffer for 10 minutes, the gel was soaked in development buffer until

bands appeared (Kavaslar et al. 2000). In the case of inefficient staining, the procedure was repeated with an extensive washing step in between. Lastly, the images of gels were taken with a scanner, and the gel was stored between protector sheets.

5 RESULTS

In this section, clinical description and results of the genetic analyses are presented.

5.1 Family A: Intellectual Disability with Speech Problem and Dysmorphic Facial Features (IDSD)

5.1.1 Clinical description

Four affected and 14 unaffected subjects were examined in this family. The patients had severe mental retardation and were unable to perform simple daily life tasks and self-care. Learning abilities were too severely impaired to allow attending school. Language was poorly developed, and only the parents could understand their speech and signs. The affected subject V-1 had somewhat better verbal skills. Early milestones were delayed including crawling and walking. The patients also had attention deficit, self-stimulating behavior, poor eye contact, enuresis, and encopresis. Occasionally, they exhibited aggressive behavior in the presence of strangers but remained peaceful with the family members. According to the physiologist, mental retardation was severe to profound. The patients had no ataxia or epileptic behavior. Dysmorphic features observed in the affected subjects included large and protruding ears and micrognathia (Fig. 5.1). They had remarkable dermatoglyphics with several missing palmer creases and straight distal and proximal creases.

According to the family, the patients had uneventful birth histories. MRI of patient V-2, a 14-year old female, revealed normal brain structures. There was no

evidence of intra-/extra-axial mass, hemorrhage, or recent or remote infract was observed. Gray and white matter distribution was normal. The ventricular system was unremarkable. No abnormal enhancement was evident on post-gadolinium images. Both hippocampi appeared normal and symmetrical. The optic chiasm and pituitary gland were normal. The 7th and 8th nerve complexes were normal and symmetrical, bilaterally. Basilar cisterns appeared normal. No signal abnormality was seen in the skull base, and there were no cervico-medullary junction abnormalities. Major intra-cranial vessels showed normal flow void. Hence, a normal brain MRI was concluded. Patient photographs and the MRI image are given in Fig. 5.1.

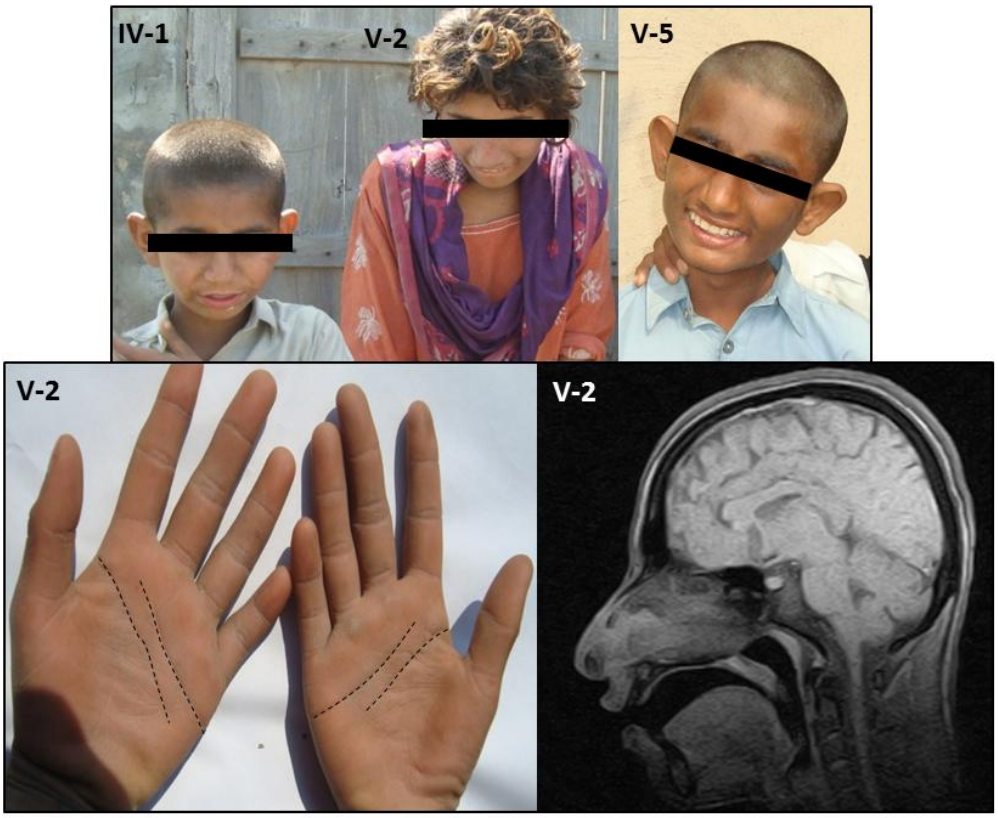


Figure 5.1. Photos of three affected subjects, dermatoglyphics and the MRI of one affected subject.

5.1.2 Multipoint linkage analysis and haplotype examination

SNP genome scan data of five family members were subjected to linkage analysis, and multipoint LOD scores were calculated under an autosomal recessive disease model with full penetrance. A maximum LOD score of 2.5 was obtained in various regions in many chromosomes. Multipoint LOD score graphics for all chromosomes are given in Fig. 5.2. The LOD score did not reach the critical value of 3 due to the small size of the family. The loci with LOD scores of 2.5 and >1-Mb size were selected as the strongest candidates. The locations and sizes of the candidate loci are given in Table 5.1. The four large intervals and other small regions were subjected to haplotype inspection, and the majority of regions were eliminated because the shared homozygosity in the patients was not due to haplotypes that were identical by descent. These analyses led to two strong homozygous regions, one at 6p25.2-p24.3 (nucleotides 2,687,729-7,557,048) with 4,869,319 bp and the other at 15q21.1-26.3 (nucleotides 43,336,670-45,095,902) with 1,759,232 bp. The haplotypes at the two loci are given in Tables 5.2 and 5.3. It was hypothesized that the causative mutation resided in one of these regions.

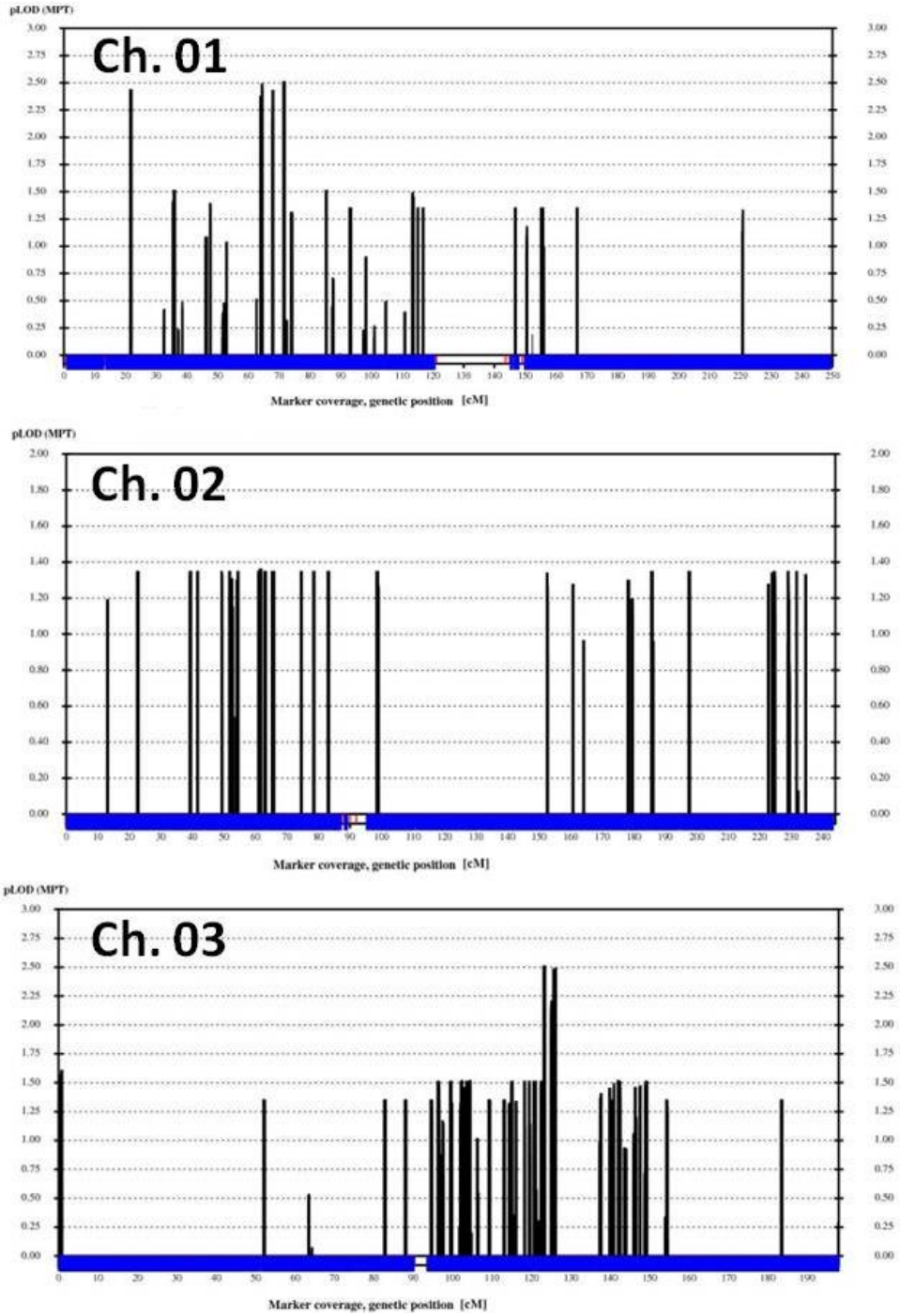


Figure 5.2. Multipoint LOD score results for autosomal recessive Intellectual Disability for all autosomal chromosomes and PAR regions.

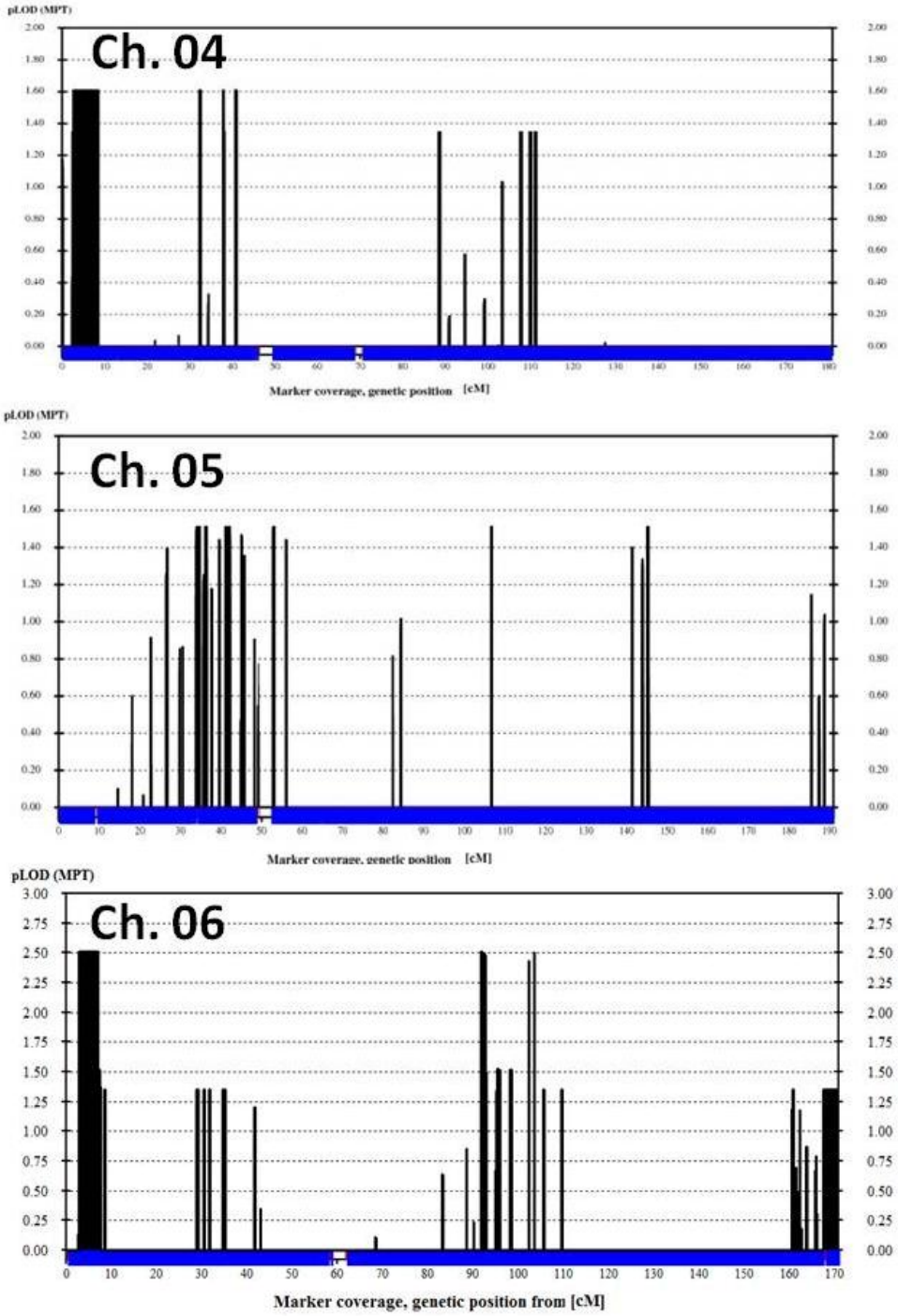


Fig. 5.2. Continued…

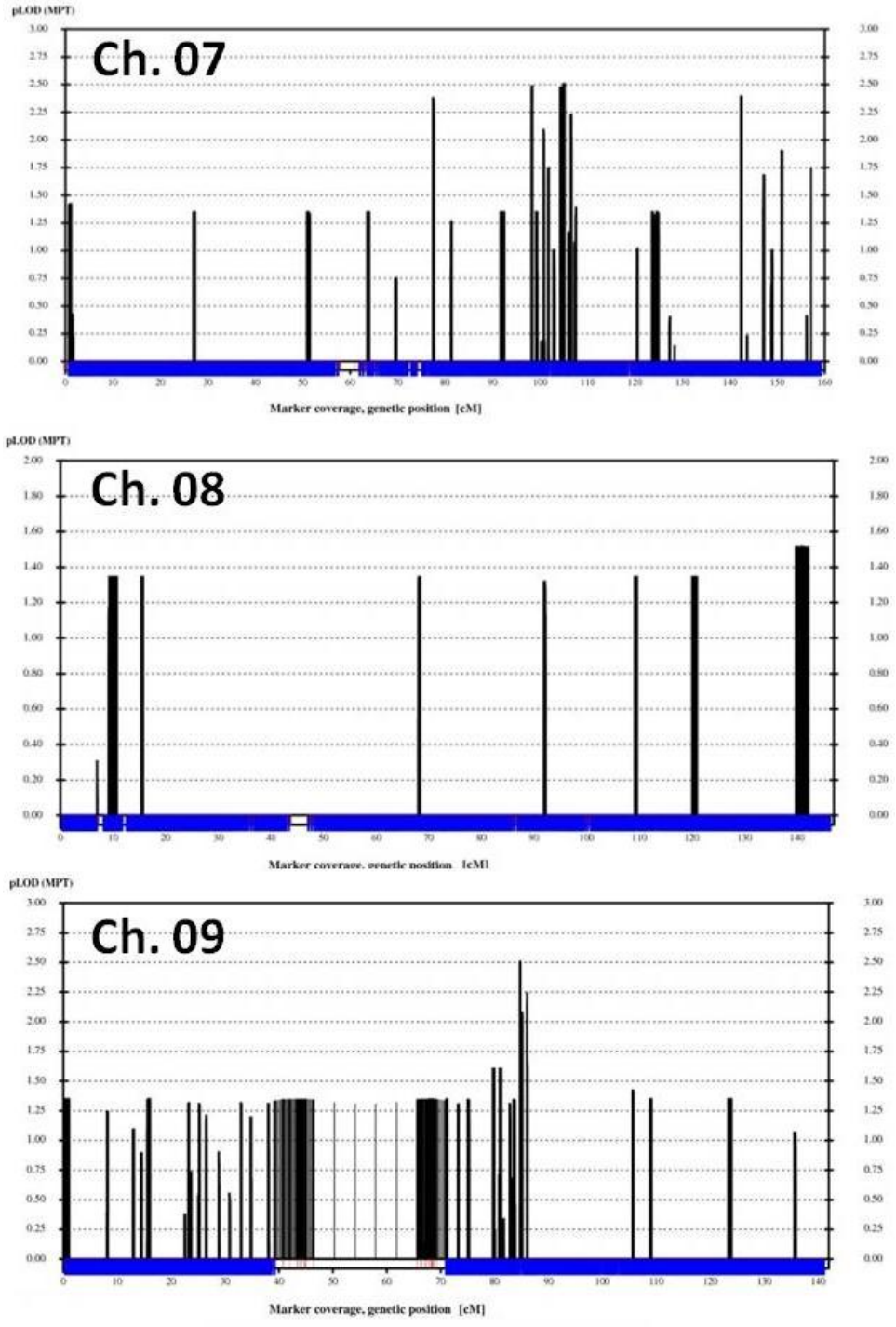


Fig. 5.2. Continued…

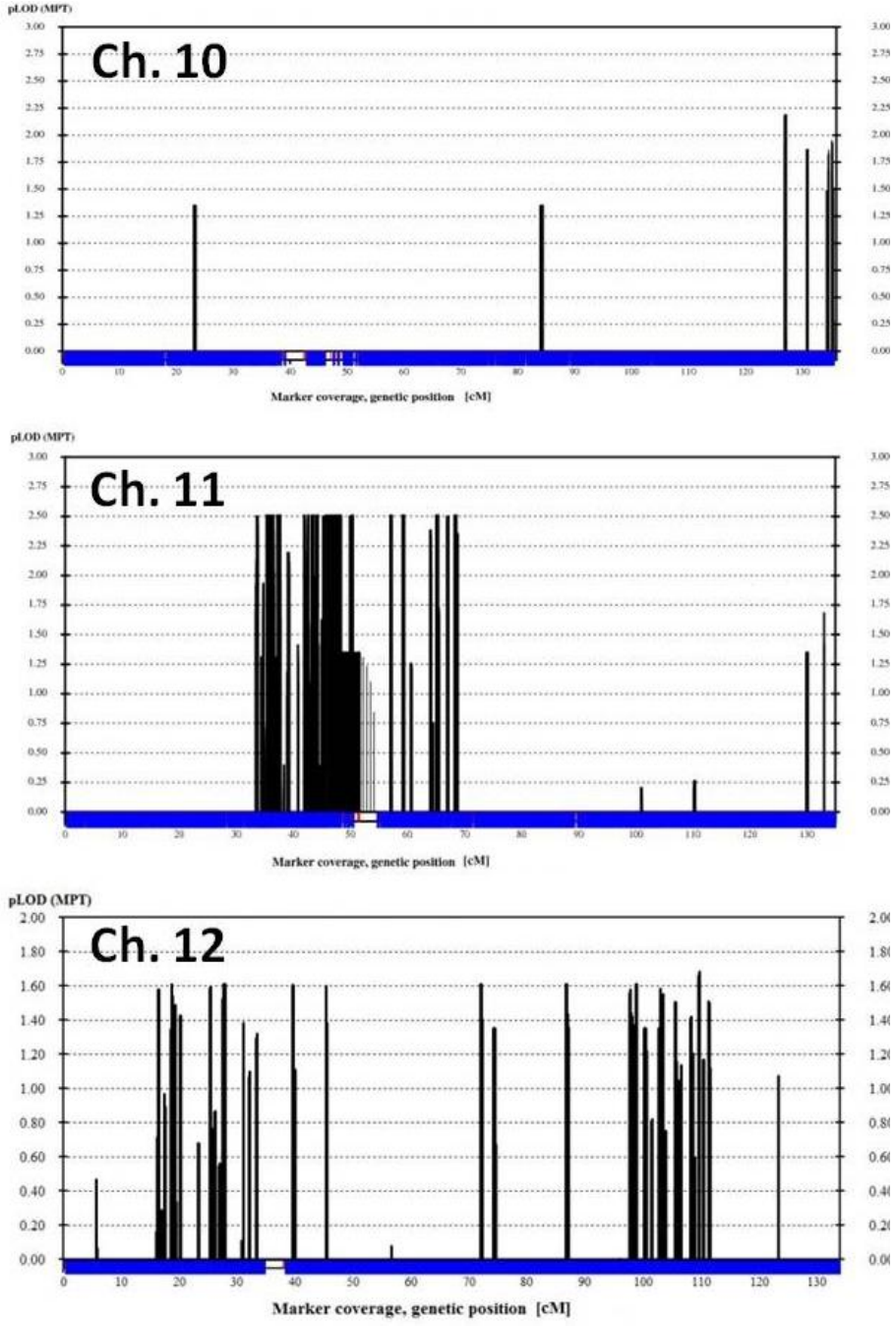


Fig. 5.2. Continued…

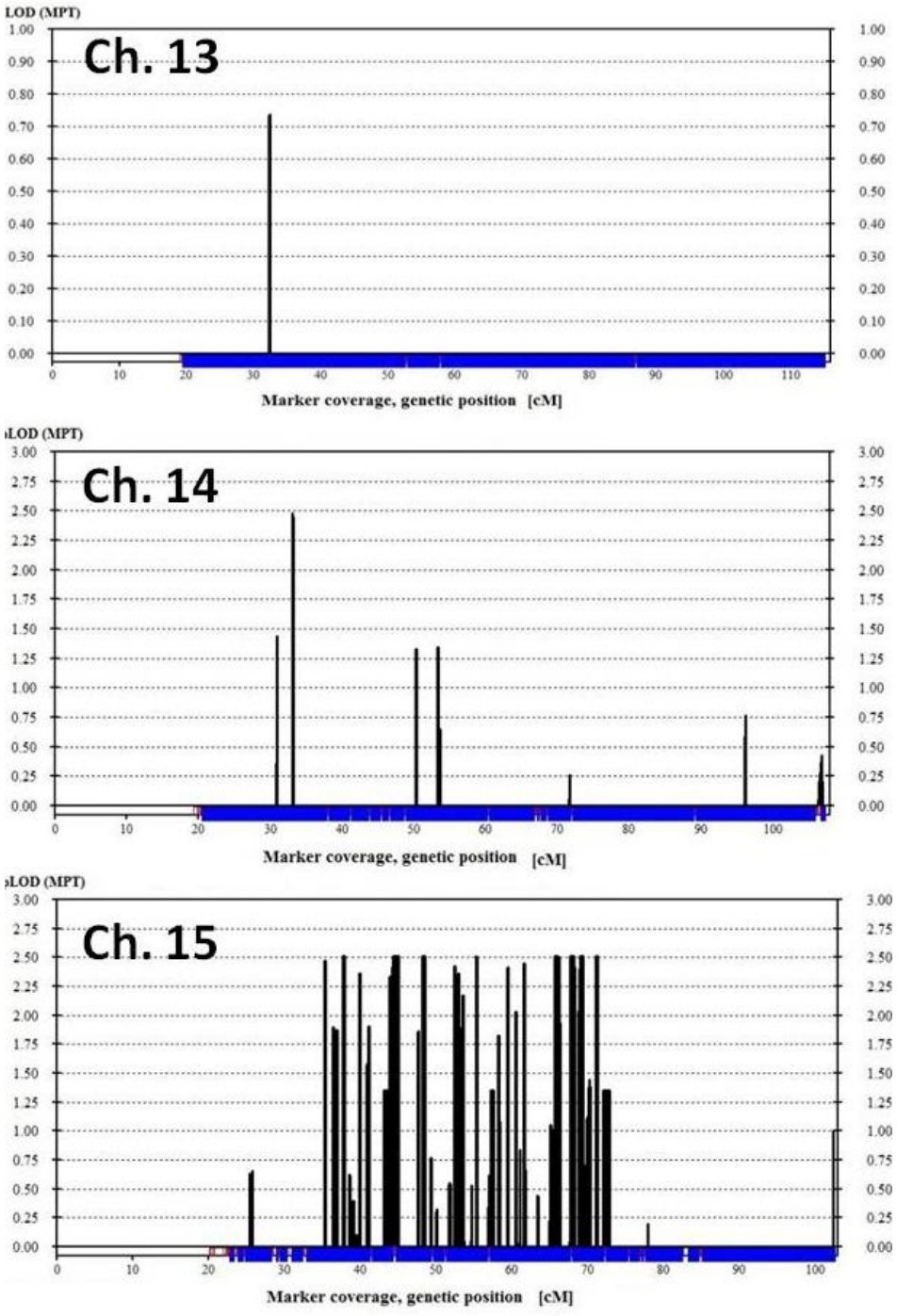


Fig. 5.2. Continued…

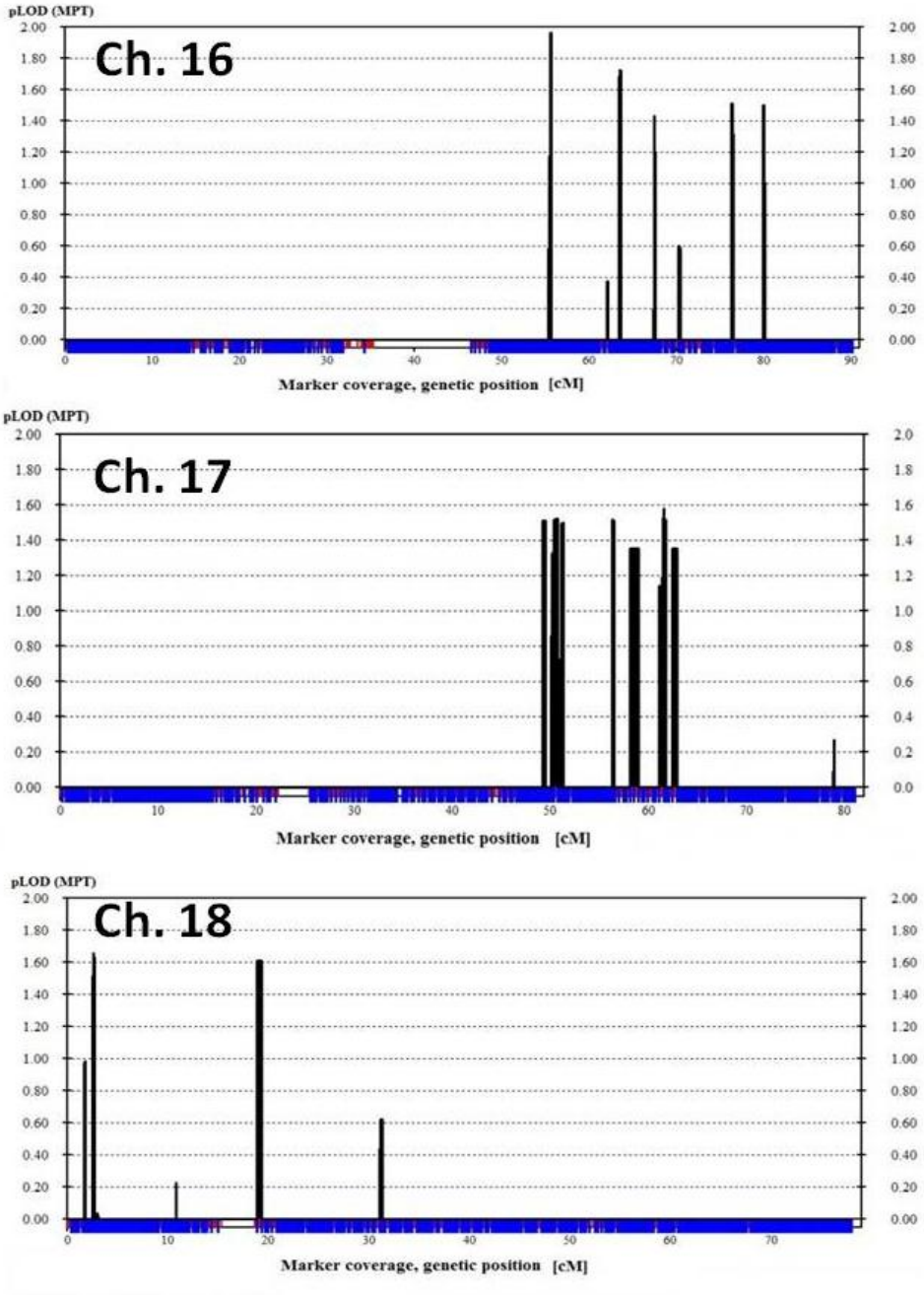


Fig. 5.2. Continued…

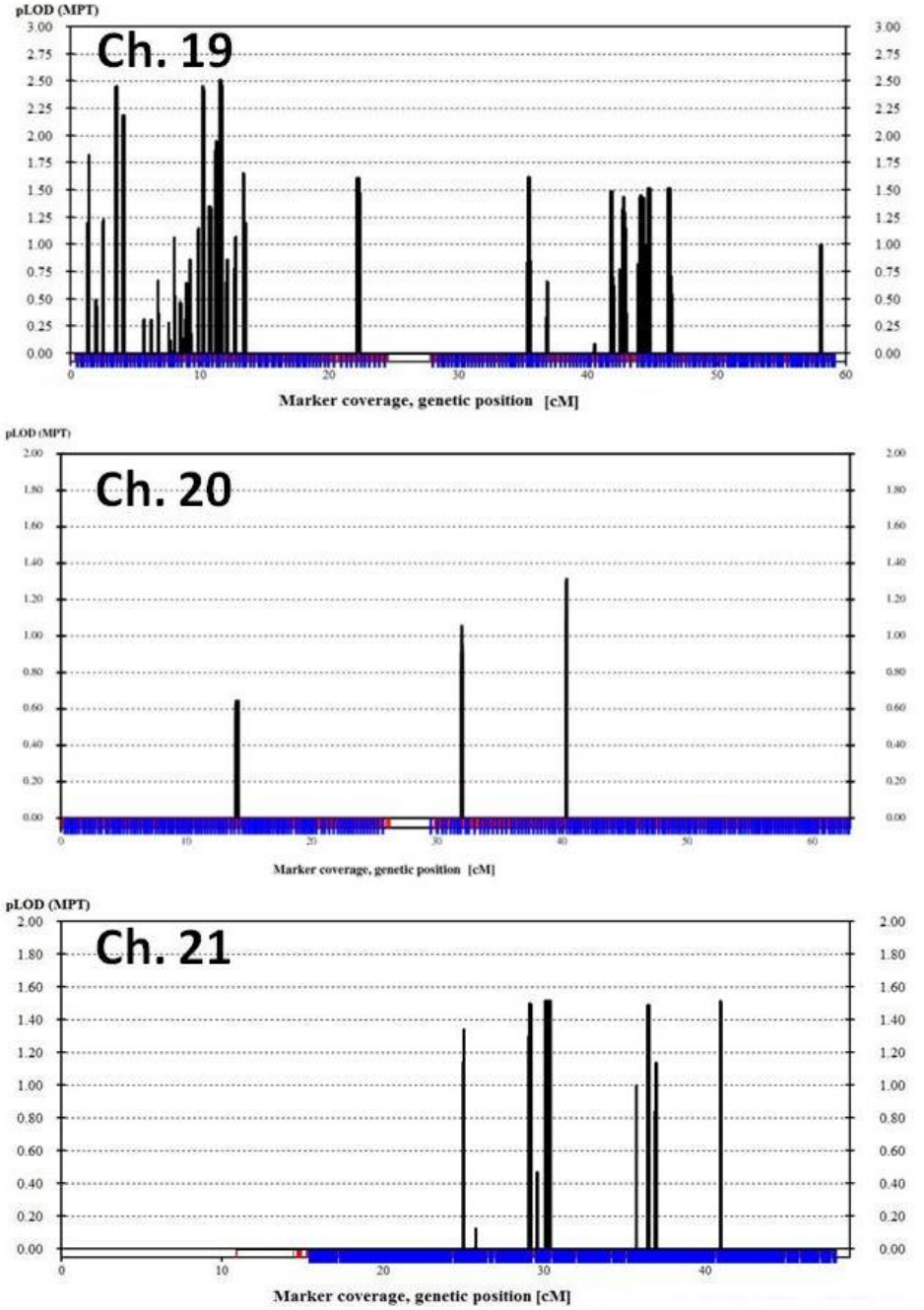


Fig. 5.2. Continued…

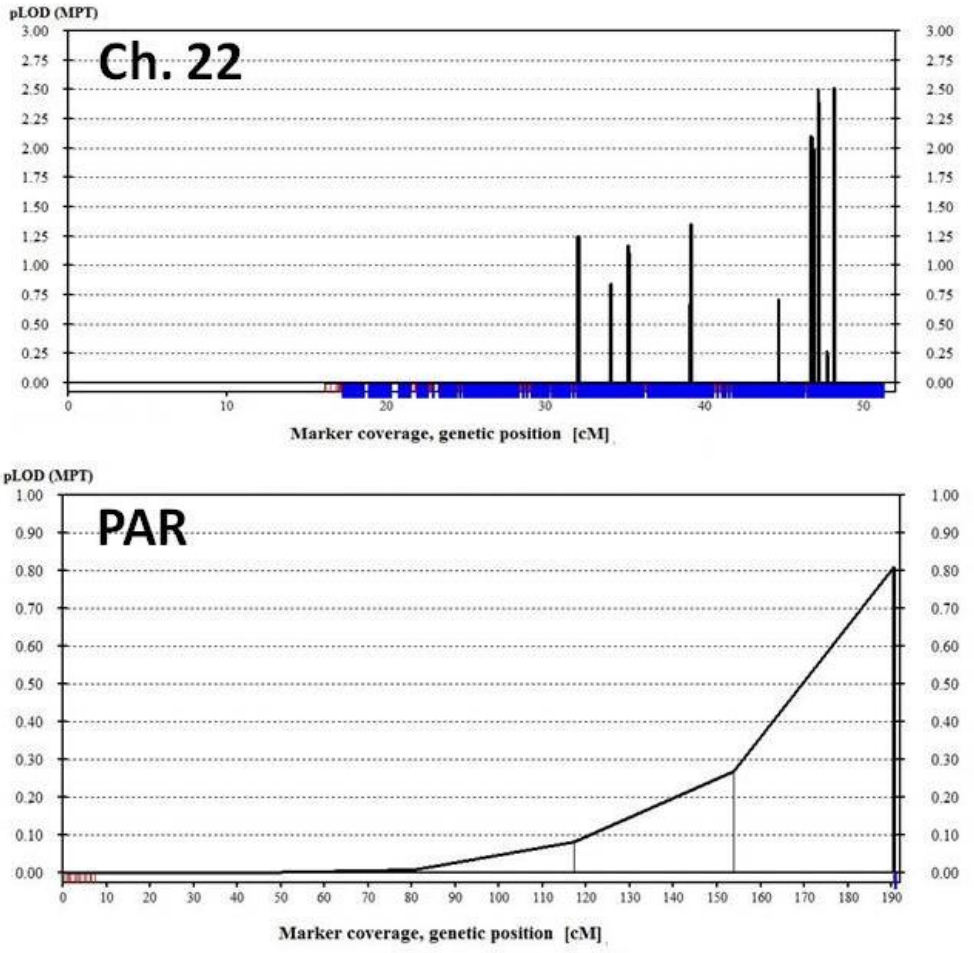


Fig. 5.2. Continued…

Table 5.1. Candidate loci >1 Mb obtained by various linkage programs in the IDSD Family.

Chr	**Flanking SNPs**	**Start**	**Stop**	**Length (bp)**	**LOD Score**	**ALLEGRO**	**GENEHUNTER**	**MERLIN**	**IBD**
5	rs1400794	3,974,744	9,209,831	5,235,087	2.5	Yes	Yes	No	No
6	rs2126243	2,687,729	7,557,048	4,869,319	2.5	Yes	Yes	Yes	Yes
11	rs7124891	46,245,361	54,794,237	8,548,876	2.5	Yes	Yes	Yes	No
15	rs6493076	43,336,670	45,095,902	1,759,232	2.5	Yes	Yes	Yes	Yes

Table 5.2. Haplotypes constructed with selected SNP markers at 6p25.2-p24.3 for IDSD Family. The genotypes of the father were deduced by the linkage software.

Marker ID	Position (MB)	Affected daughter 1		Affected daughter 2		Affected son		Unaffected daughter		Unaffected Father		Unaffected Mother	
rs879098	2.51	1	2	2	2	1	2	1	2	1	2	1	2
rs6918808	2.61	1	1	1	1	1	1	1	1	1	1	2	1
rs4382335	2.74	2	2	2	2	2	2	1	2	1	2	1	2
rs405616	2.84	2	2	2	2	2	2	1	2	1	2	2	2
rs9503375	3.05	1	1	1	1	1	1	2	1	2	1	1	1
rs13206421	3.15	1	1	1	1	1	1	2	1	2	1	2	1
rs4959235	3.36	2	2	2	2	2	2	1	2	1	2	2	2
rs9405647	3.57	2	2	2	2	2	2	1	2	1	2	1	2
rs12212080	3.68	2	2	2	2	2	2	1	2	1	2	1	2
rs659699	3.78	2	2	2	2	2	2	1	2	1	2	1	2
rs6938648	4.01	1	1	1	1	1	1	2	1	2	1	2	1
rs4605929	4.15	1	1	1	1	1	1	2	1	2	1	2	1
rs9378854	4.28	2	2	2	2	2	2	1	2	1	2	1	2
rs885185	4.4	1	1	1	1	1	1	2	1	2	1	2	1
rs12206701	4.62	1	1	1	1	1	1	2	1	2	1	2	1
rs3812183	4.72	1	1	1	1	1	1	2	1	2	1	2	1
rs12198622	4.86	1	1	1	1	1	1	2	1	2	1	2	1
rs1246996	5.06	2	2	2	2	2	2	1	2	1	2	1	2
rs6904616	5.42	1	1	1	1	1	1	2	1	2	1	2	1
rs11243016	5.52	2	2	2	2	2	2	1	2	1	2	1	2
rs9504549	5.84	1	1	1	1	1	1	1	1	1	1	2	1
rs10458090	5.95	2	2	2	2	2	2	1	2	1	2	1	2
rs9502396	6.05	2	2	2	2	2	2	1	2	1	2	1	2
rs1993554	6.29	2	2	2	2	2	2	1	2	1	2	2	2
rs12196470	6.39	1	1	1	1	1	1	1	1	1	1	2	1
rs1535705	6.49	2	2	2	2	2	2	1	2	1	2	2	2
rs115560	6.69	1	1	1	1	1	1	1	1	1	1	2	1
rs2876031	6.8	1	1	1	1	1	1	2	1	2	1	1	1
rs9502516	6.91	2	2	2	2	2	2	1	2	1	2	2	2
rs592377	7.06	1	1	1	1	1	1	1	1	1	1	2	1
rs7769554	7.37	2	2	2	2	2	2	2	2	2	2	1	2
rs11756237	7.47	1	1	1	1	1	1	1	1	1	1	2	1
rs9791321	7.57	1	2	1	1	1	2	1	1	1	1	2	1

Table 5.3. Haplotypes constructed with selected SNP markers at 15q21.33 for IDSD Family. The genotypes of the father were deduced through the software.

Marker ID	Position (MB)	Affected daughter 1		Affected daughter 2		Affected son		Unaffected daughter		Unaffected Father		Unaffected Mother	
rs1197547	42.91	1	2	1	2	1	2	1	1	1	2	2	1
rs12594951	42.93	2	2	2	2	2	2	1	2	1	2	2	2
rs16957063	42.98	1	1	1	1	1	1	1	1	1	1	1	1
rs12594483	43.02	2	2	2	2	2	2	1	2	1	2	2	2
rs10220718	43.10	2	2	2	2	2	2	2	2	2	2	2	2
rs3803341	43.24	2	2	2	2	2	2	2	2	2	2	2	2
rs2412752	43.34	2	2	2	2	2	2	1	2	1	2	1	2
rs513071	43.48	1	1	1	1	1	1	1	1	1	1	1	1
rs555001	43.55	2	2	2	2	2	2	2	2	2	2	2	2
rs1549525	43.63	1	1	1	1	1	1	1	1	1	1	1	1
rs690262	43.70	2	2	2	2	2	2	2	2	2	2	2	2
rs689931	43.82	1	1	1	1	1	1	1	2	2	1	2	1
rs2927071	43.92	1	1	1	1	1	1	1	1	2	1	1	1
rs8040336	44.05	1	1	1	1	1	1	2	2	2	1	2	1
rs11858713	44.10	2	2	2	2	2	2	2	2	2	2	2	2
rs8023508	44.18	1	1	1	1	1	1	1	1	1	1	1	1
rs1077421	44.24	2	2	2	2	2	2	2	2	1	2	2	2
rs12910886	44.46	1	1	1	1	1	1	1	1	1	1	1	1
rs2615258	44.50	1	1	1	1	1	1	2	2	2	1	2	1
rs2696089	44.53	2	2	2	2	2	2	2	2	2	2	2	2
rs8024461	44.58	1	1	1	1	1	1	1	2	2	1	1	1
rs17586255	44.73	1	1	1	1	1	1	1	2	2	1	2	1
rs2556560	44.82	1	1	1	1	1	1	1	2	2	1	1	1
rs2556568	44.85	1	1	1	1	1	1	1	2	2	1	1	1
rs7165146	44.95	1	1	1	1	1	1	1	2	2	1	1	1
rs1901530	45.01	2	2	2	2	2	2	1	2	1	2	2	2
rs2167595	45.02	1	1	1	1	1	1	1	2	2	1	1	1
rs2470908	45.03	2	2	2	2	2	2	1	2	1	2	2	2
rs2444004	45.04	1	1	1	1	1	1	2	2	2	1	2	1
rs3100139	45.05	1	1	1	1	1	1	1	2	2	1	1	1
rs2444009	45.05	2	2	2	2	2	2	1	2	1	2	2	2
rs2462043	45.07	1	1	1	1	1	1	2	2	2	1	2	1
rs2049333	45.09	2	2	2	2	2	2	1	2	1	2	2	2
rs17518970	45.10	1	2	1	2	1	2	1	2	2	1	2	1

5.1.3 CNV analysis

SNP genotyping data were subjected to deletion and duplication analysis by using cnvPartition (v3.2.0) CNV Analysis Plug-in for GenomeStudio. Default parameters were used. No deletion or duplication was detected within the candidate loci.

5.1.4 Exome sequencing and evaluation of the variants

Exome sequencing was performed for affected individual V-2 (Table 5.4). The resultant file with sequencing data obtained from the commercial exome service was subjected to bioinformatics analysis for detecting rare variants in the candidate genes.

Table 5.4. Exome sequencing analysis summary in IDSD Family in subject V-2.

Exome sequence file features	**No**
Read length (bp)	101.0
Target regions (bp)	51,189,318
% coverage of target regions (more than 1X)	99%
% coverage of target regions (more than 10X)	96.8%
Mean read depth of target regions	59.3
Number of total SNPs found	65,445
Number of coding SNPs found	19442
Number of indels found	6030
Number of coding indels found	485

First, the deleterious mutations listed in the exome sequence results in the two strongest candidate regions were investigated. No known gene involved in nonsyndromic or syndromic intellectual disability was seen in either of the regions. Instead, a strong candidate exonic variant c.G170A (p.C57Y) in exon 2 of *PDIA3* was found at 15q21.1-26.3 (GRCh37/hg19) (Table 5.5). No other exonic or splicing variant was detected in the candidate regions fitting the criteria of candidate mutation. As evident in Tables 5.5 and 5.6, all other exonic, splicing and even UTR variants had high frequencies. However, intronic variants in *MYLK4* and *UBR1* with no reported frequency were observed in the candidate regions, but both were eliminated because they were intronic. The candidate variant at Chr15: 44046024 in *PDIA3* was a G to A transition that led to the substitution of cysteine for tyrosine, a missense nonsynonymous change. The variant was not found in dbSNP or EVS, or in the South Asian samples in ExAC database that contains many Pakistani exomes. It was also absent in other exome files in our laboratory. Further analysis of the candidate variant was done by realignment of the raw reads to the reference genome (hg19/GRCh37 assembly), allowing approximately 2% base-pair mismatch in read alignments. The alignments were visualized and validated via program BamView. Sanger sequencing was performed to validate the variant and ascertain its segregation with the disease in the family members. Affected sibs were homozygous and mother heterozygous for the variant, whereas unaffected sib and cousins were wild type (Fig. 5.3). A total of 200 Pakistani control samples were screened for the *PDIA3* mutation and the mutation was not found in any of them.

Table 5.5. Variants in the exome sequence output in the candidate homozygous region 15q21.1-26.3 (nucleotides 43,336,670-45,095,902).

Location	**Ref Base**	**Obs Base**	**Hom Het**	**SNP Quality**	**Total Depth**	**Region**	**Gene**	**SNP ID**	**1000G Freq**
43330175	A	T	hom	86	20	intronic	*UBR1*	rs7166596	0.965
43352041	G	A	hom	192	35	intronic	*UBR1*	rs4924704	0.806
43398108	A	-	hom	214	25	intronic	*UBR1*	.	.
43453056	C	G	hom	124	21	intronic	*TMEM62*	rs11070392	0.321
43478113	G	C	hom	222	68	intronic	*CCNDBP1*	rs540587	0.733
43478511	C	T	hom	222	79	intronic	*CCNDBP1*	rs530118	0.432
43483537	T	C	hom	51	8	intronic	*CCNDBP1*	rs495286	0.840
43568999	C	T	hom	175	8	intronic	*TGM7*	rs530683	0.515
43784000	C	T	hom	144	68	intronic	*TP53BP1*	rs555252	0.957
43821766	G	C	hom	222	75	intronic	*MAP1A*	rs117078346	0.005
43897729	T	C	hom	187	16	intronic	*STRC*	rs139626815	0.003
43931245	A	T	hom	221	52	intronic	*CATSPER2*	rs555127142	0.0006
43941125	G	C	hom	122	17	upstream	*CATSPER2*	rs2470122	0.974
44038899	C	T	hom	159	36	exonic	*PDIA3*	rs2411284	0.518
44046024	**G**	**A**	**hom**	**222**	**36**	**exonic**	***PDIA3***	**.**	**.**
44053617	C	T	hom	130	24	intronic	*PDIA3*	rs8040336	0.335
44061802	C	T	hom	215	68	exonic	*PDIA3*	rs1053492	0.444
44068160	C	G	hom	222	74	intronic	*ELL3*	rs3844075	0.843
44089259	A	G	hom	183	43	ncRNA	*SERF2-C15ORF63*	rs116954300	0.002
44090854	T	C	hom	222	32	ncRNA	*SERF2-C15ORF63*	rs654276	0.837
44091351	A	T	hom	222	114	ncRNA	*SERF2-C15ORF63*	rs656473	0.968
44120559	T	G	hom	222	70	exonic	*WDR76*	rs678084	0.818
44903285	C	A	hom	31.3	5	intronic	*SPG11*	rs12909507	0.978
44961828	A	G	hom	222	48	intronic	*PATL2*	rs527998052	0.0004
44966389	T	C	hom	222	37	exonic	*PATL2*	rs8026845	0.985
45028710	T	C	hom	198	36	UTR5	*TRIM69*	rs2470908	0.902

45047134	G	A	hom	222	109	exonic	*TRIM69*	rs2470911	0.647
45047573	C	T	hom	222	51	exonic	*TRIM69;TRIM69*	rs3100139	0.779
45048757	T	A	hom	202	31	intronic	*TRIM69*	rs3759881	0.947
45059390	C	T	hom	56.1	7	intronic	*TRIM69*	rs6493127	0.954
45060014	A	T	hom	61	18	UTR3	*TRIM69*	rs3100143	0.836

Table 5.6: Variants in the exome sequence of the candidate homozygous region 6p25.2-p24.3 (nucleotides 2,687,729-7,557,048).

Location	**Ref Base**	**Obs Base**	**Hom Het**	**SNP Quality**	**Total Depth**	**Region**	**Gene**	**SNP ID**	**1000G Freq**
2685904	T	C	het	29	10	Intronic	*MYLK4*	rs10498657	0.264
2692874	AAAGG	-	hom	6.36	2	intronic	*MYLK4*	.	.
2766144	T	C	hom	222	11	exonic	*WRNIP1*	rs17851692	0.286
2766465	T	C	hom	222	13	exonic	*WRNIP1*	rs8333	0.400
2766727	A	G	hom	100	14	intronic	*WRNIP1*	rs3757094	0.277
2766779	G	A	Hom	18.8	2	intronic	*WRNIP1*	rs3757093	0.241
2784571	G	A	Hom	222	32	exonic	*WRNIP1*	rs160703	0.160
2833739	A	G	Hom	36.3	5	UTR3	*SERPINB1*	rs15286	0.728
2834265	C	T	Hom	155	12	intronic	*SERPINB1*	rs386915	0.686
2836266	T	A	Hom	173	104	intronic	*SERPINB1*	rs385955	0.523
2838046	A	G	Hom	177	60	intronic	*SERPINB1*	rs316339	0.680
2838102	G	A	Hom	222	114	intronic	*SERPINB1*	rs316340	0.414
2838248	G	A	Hom	31	10	intronic	*SERPINB1*	rs316341	0.664
2854367	G	A	Hom	222	33	upstream	*MIR4645*	rs412303	0.321
2890283	A	G	Hom	41.1	6	UTR3	*SERPINB9*	rs318486	0.646
3077141	T	C	Hom	195	69	exonic	*RIPK1*	rs2272990	0.853
3085477	A	C	Hom	191	52	intronic	*RIPK1*	rs909966	0.079
3118947	A	G	Hom	222	40	UTR5	*BPHL*	rs2231355	0.985
3138194	T	C	Hom	213	34	intronic	*BPHL*	rs1534998	0.724
3152822	C	G	Hom	222	98	UTR3	*BPHL*	rs2231371	0.066
3157798	G	A	Hom	101	7	upstream	*TUBB2A*	rs909964	0.613
3157809	C	T	Hom	90	8	upstream	*TUBB2A*	rs909965	0.571
3255327	G	A	Hom	214	93	intergenic	*TUBB2B,PSMG4*	rs9503497	0.701
3255340	A	C	Hom	200	90	intergenic	*TUBB2B,PSMG4*	rs9503498	0.695
3259463	G	T	Hom	42.1	7	intronic	*PSMG4*	rs9501962	0.713
3264101	A	G	Hom	222	70	intronic	*PSMG4*	rs4959785	0.632
3264487	T	C	Hom	222	60	exonic	*PSMG4*	rs4959786	0.638
3264502	G	C	Hom	222	65	exonic	*PSMG4*	rs4959787	0.635
3264582	G	A	Hom	222	58	intronic	*PSMG4*	rs11757564	0.644
3264671	T	C	Hom	108	11	intronic	*PSMG4*	rs13193012	0.717
3267710	T	C	Hom	174	105	intronic	*PSMG4*	rs9328157	0.814
3285286	A	G	Hom	170	42	intronic	*SLC22A23*	rs10793841	0.563
3287118	T	C	Hom	222	20	exonic	*SLC22A23*	rs6910086	0.870
3439598	-	A	Hom	214	30	intronic	*SLC22A23*	rs58636652	.
3738271	C	-	Hom	214	38	intronic	*PXDC1*	rs3215783	.

3739193	T	C	Hom	63	3	intronic	*PXDC1*	rs4959871	0.365
4020528	A	G	Hom	150	32	intergenic	*FAM50B,PRPF4B*	rs11751316	0.024
4021554	A	G	Hom	222	26	upstream	*PRPF4B*	rs11752006	0.999
4033065	A	G	Hom	154	14	intronic	*PRPF4B*	rs6924389	0.898
4068716	A	C	Hom	7.59	2	UTR3	*C6orf146*	rs593291	0.612
4069166	C	T	Hom	222	121	exonic	*C6orf146*	rs595413	0.555
4936074	T	G	Hom	181	26	intronic	*CDYL*	rs1539004	0.960
4999193	A	G	Hom	204	10	intronic	*RPP40*	rs396839	0.564
5003952	T	C	hom	69.1	7	intronic	*RPP40*	rs4960056	0.949
5086070	A	G	hom	222	17	exonic	*PPP1R3G*	rs667752	1.000
5086211	G	C	hom	44.5	4	exonic	*PPP1R3G*	rs584962	1.000
5139012	C	A	hom	165	28	intronic	*LYRM4*	rs463054	1.000
5260936	A	C	hom	222	30	exonic	*LYRM4*	rs2224391	0.369
5261136	G	A	hom	222	66	UTR5	*LYRM4*	rs2224392	0.234
5431340	A	G	hom	219	56	exonic	*FARS2*	rs11243011	0.177
5545654	C	T	hom	62.1	6	intronic	*FARS2*	rs7749599	0.924
5999464	C	G	hom	171	28	intronic	*NRN1*	rs3749860	0.540
6145692	C	T	hom	212	13	UTR3	*F13A1*	rs1050783	0.140
6145726	A	G	hom	119	25	UTR3	*F13A1*	rs1050782	0.506
6152007	C	T	hom	63	8	intronic	*F13A1*	rs5980	0.177
6152137	C	G	hom	222	37	exonic	*F13A1*	rs5988	0.174
6152211	C	G	hom	214	27	intronic	*F13A1*	rs2274394	0.173
6167621	G	A	hom	222	29	intronic	*F13A1*	rs435048	0.981
6174786	G	C	hom	52	3	intronic	*F13A1*	rs380058	0.978
6305477	C	A	hom	222	58	intronic	*F13A1*	rs2295753	0.313
6320808	T	G	hom	111	37	intronic	*F13A1*	rs2815822	0.908
6321128	T	C	hom	169	53	upstream	*F13A1*	rs1024231	0.760
6588881	A	C	hom	125	18	ncRNA	*LY86-AS1*	rs977785	0.699
6649980	-	AGG	hom	214	33	intronic	*LY86*	rs3062400	0.800
7211751	G	A	hom	180	32	intronic	*RREB1*	rs1334577	0.242

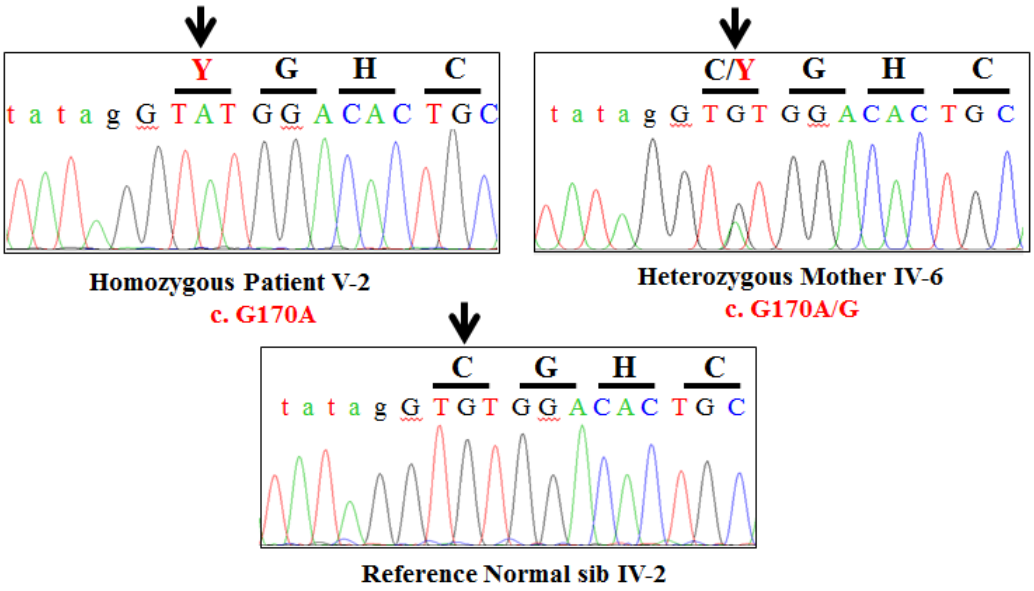


Figure 5.3. Chromatograms showing mutation *PDIA3* c.G170A in an affected subject and the heterozygous mother, and the reference sequence of an unaffected sibling.

At the protein level, c.G170A mutation causes the substitution of cysteine with tyrosine at position 57 (p.Cys57Tyr). Online mutation prediction tools (PolyPhen2, SIFT and MutationTaster) predicted the variant as pathogenic. The outputs of online tools are given in Table 5.7. At the protein level, the substitution of cysteine for tyrosine would be expected to change the protein structure, as cysteine is involved in disulfide bond formation that is important in maintaining the 3-D structure of the protein. Gene Homology showed that the residue is in a region of 40 amino acids that are totally conserved across species. A high expression of this gene was found in many tissues including the brain. Besides this, cysteine at position 57 lies in a stretch of seven amino acids that are highly conserved in mammals (Figs. 5.4, 5.5).

Table 5.7. Prediction of the effect of *PDIA3* c.170G>A (p.C57Y) mutation by online tools.

Online Tool	Score	Prediction
Polyphen2	0.995/1.00	Probably damaging
MutPred	0.995/1.00	Deleterious
MutationTaster	Model: simple_aae, prob: 1	Disease causing

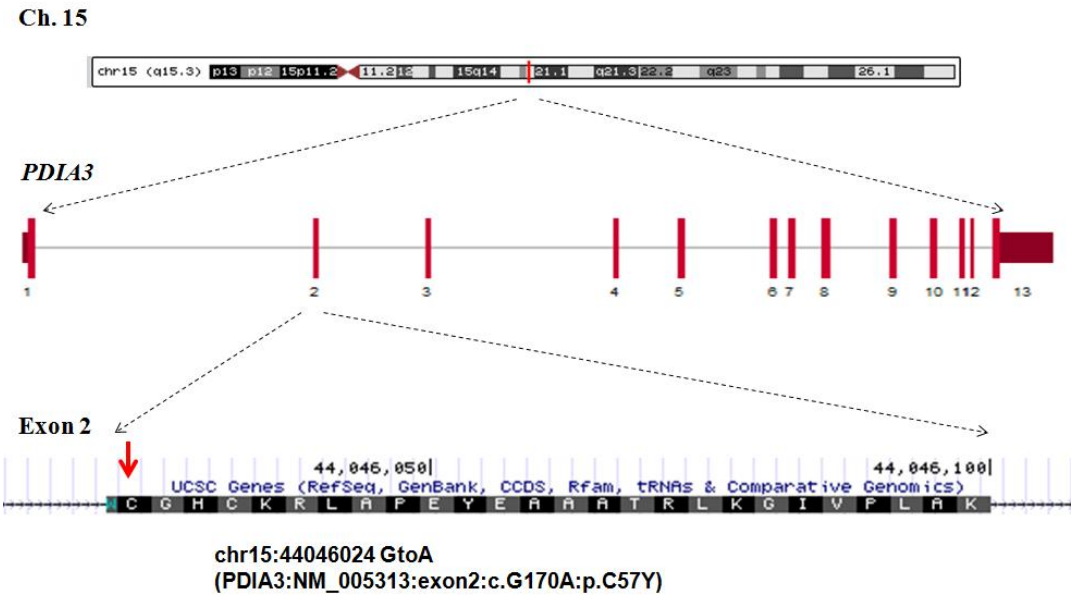


Figure 5.4. Genomic organization and location of mutation of *PDIA3* at Chr. 15q15.3.

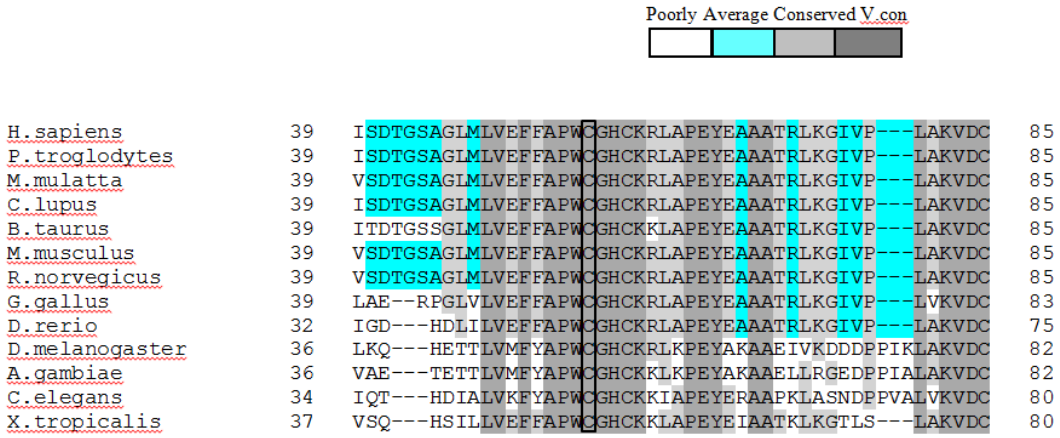


Figure 5.5. Multiple sequence alignment of partial PDIA3 amino acid sequence across 13 species. Residue Cys57 is boxed.

In summary, we found two candidate loci for IDSD in the present family, and exome sequencing revealed a novel candidate variant in *PDIA3*. The variant was assessed as the only very strong candidate because it was novel, residing in a highly conserved protein region and assessed as pathogenic by online mutation prediction tools. It was validated by Sanger sequencing and was suggested as the disease-causing mutation by population screening of 200 controls. The variant was not found in the ExAC database. No homozygous deletion or duplication was found within the candidate loci by CNV analysis.

5.2 Family B: Autosomal Recessive Primary Microcephaly (MCPH)

5.2.1 Clinical description

Four affected and two unaffected subjects were examined in this family. The patients have primary microcephaly with mild intellectual disability (ID). A total of four subjects (2 males and 2 females) were affected with the malformation. No other associated anomaly was observed in any affected individual.

The subjects had small skulls (Fig. 5.6). The condition was congenital and non-progressive. The affected subjects were able to perform routine activities. The male affected subjects were attending a shop with their father and helped him in day-to-day activities. They had a good understanding of money. They had normal speech and hearing but had no formal schooling. The family denied the presence of any features like seizures, developmental delay, ataxia, problems with movement and balance, vision problem, or hearing problems.

Figure 5.6. The affected subjects IV-2 and IV-3 (left to right) in the Family with MCPH.

5.2.2 Multipoint linkage analysis

Genome scan data obtained by SNP genotyping was subjected to linkage analysis, and multipoint LOD scores were calculated assuming an autosomal recessive inheritance model with full penetrance. Multipoint LOD score graphics for all chromosomes are given in Fig. 5.7. A maximal LOD score of 3 was obtained at many loci, but only two were >1 Mb. After haplotype inspection, the strongest candidate turned out to be at 1q25.3-q32.1 (nucleotides 185,518,666-204,614,475), 19.10 Mb in size (Fig. 5.8). This region was investigated by UCSC genome browser and MapViewer. It harbors one very important gene *ASPM* (OMIM 60816) (nucleotides 197,053,256-197,115,823) with a total of 28 exons, and more than 39 mutations have been reported in this gene in families afflicted with microcephaly. Most likely it is the gene responsible for the disease in the family. Therefore, the study was terminated and no attempt was made to further check *ASPM* for mutation.

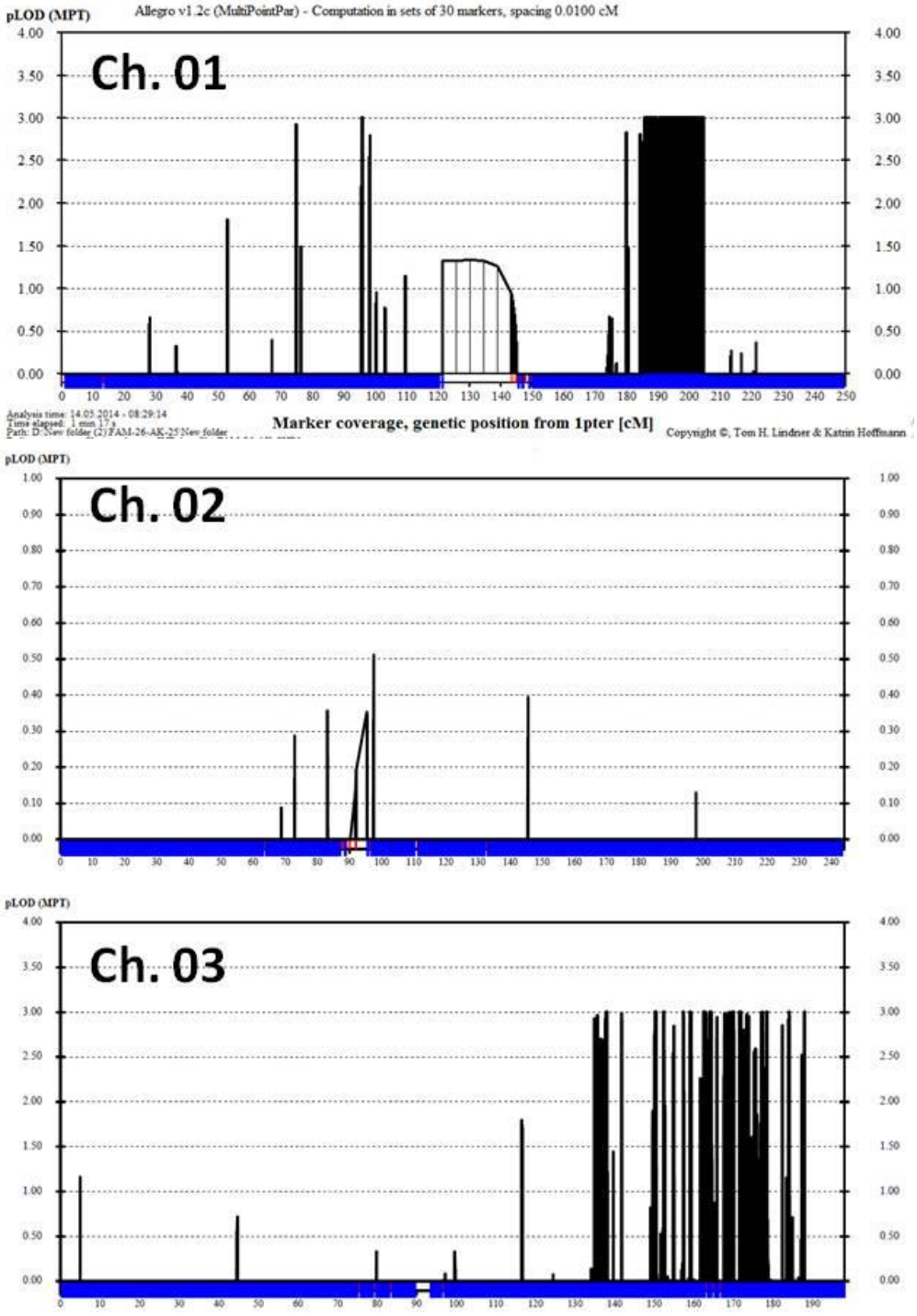


Figure 5.7. Multipoint LOD score results for autosomal recessive microcephaly for all autosomal and X chromosomes.

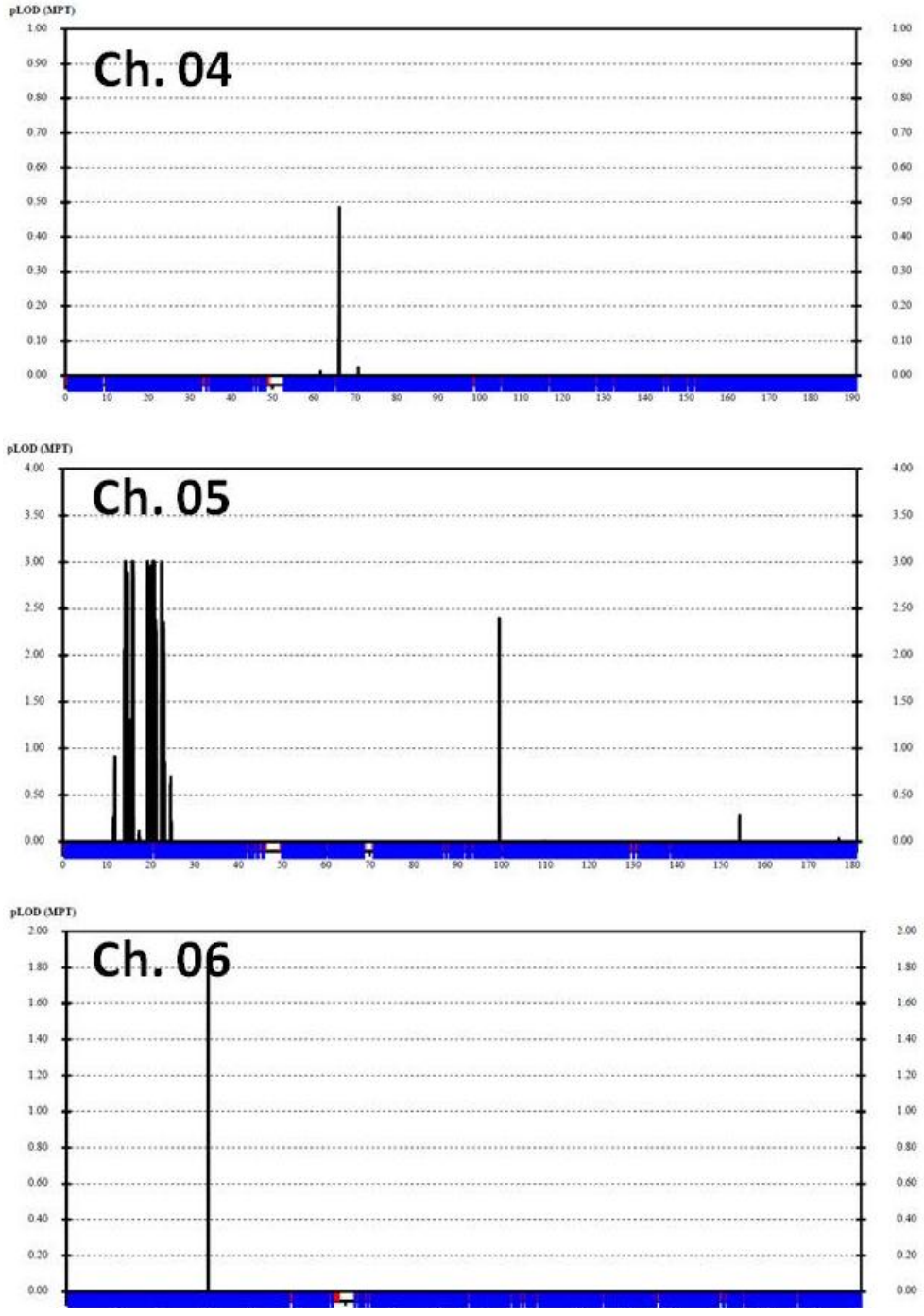


Fig. 5.7. Continued…

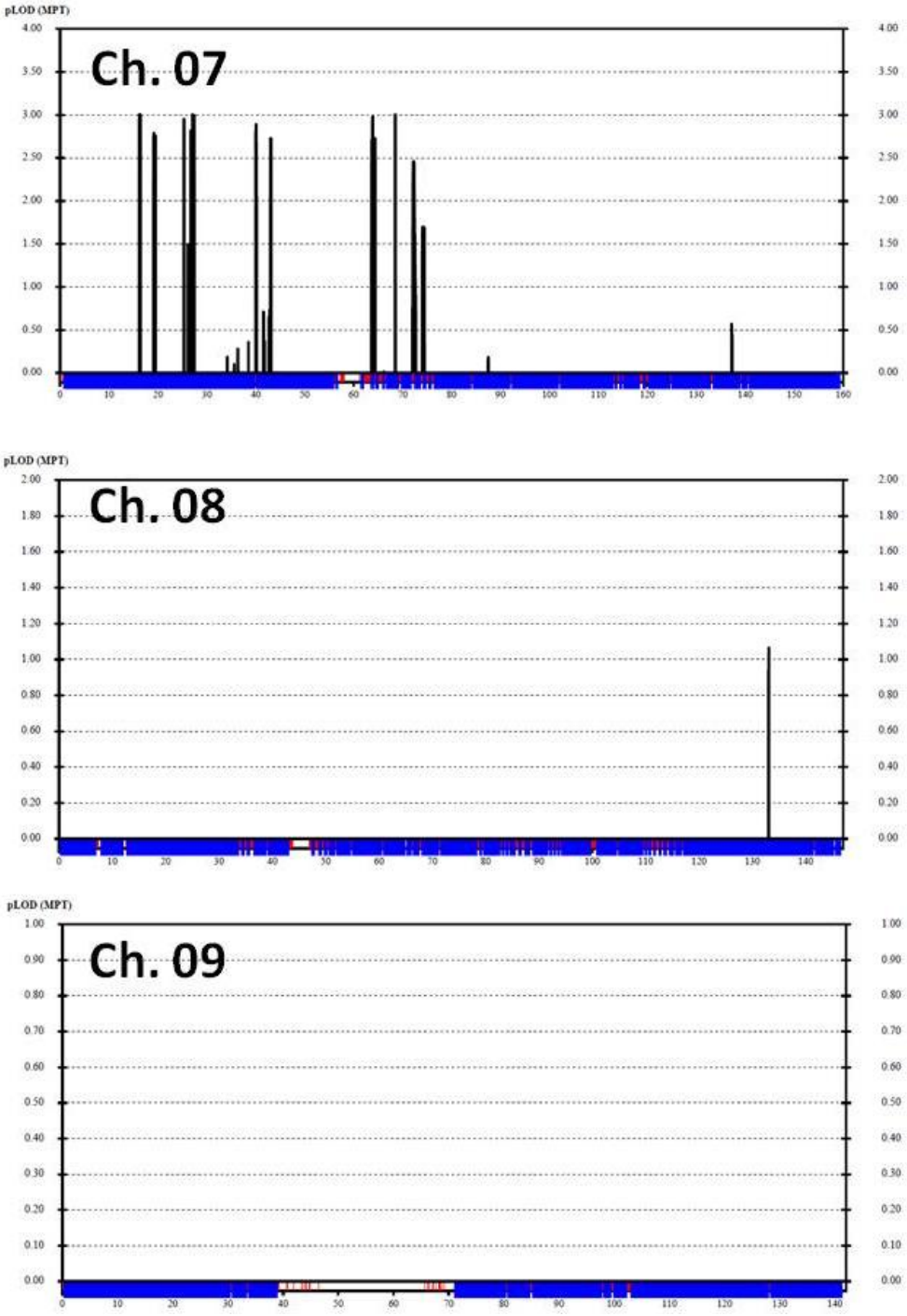


Fig. 5.7. Continued…

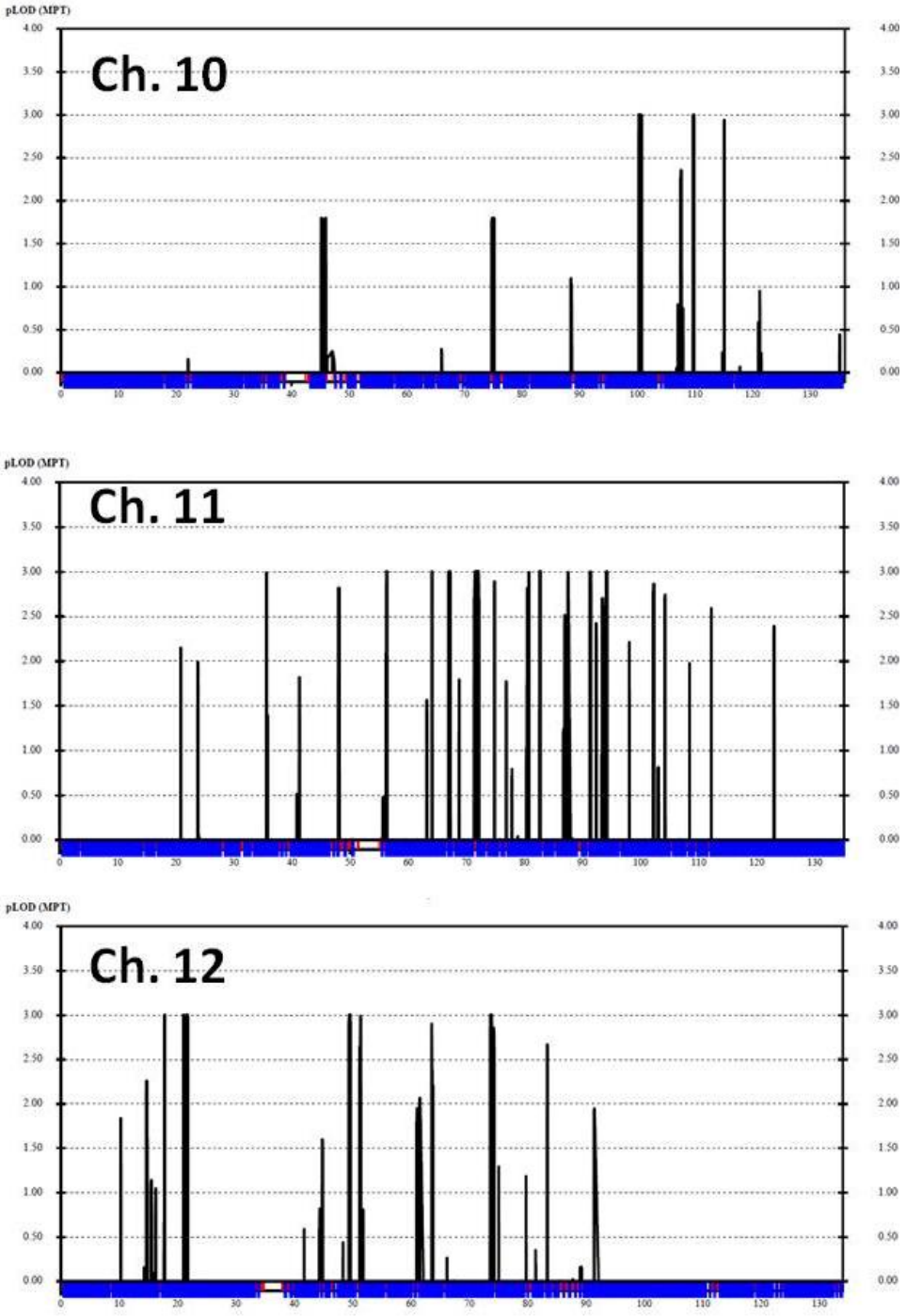


Fig. 5.7. Continued…

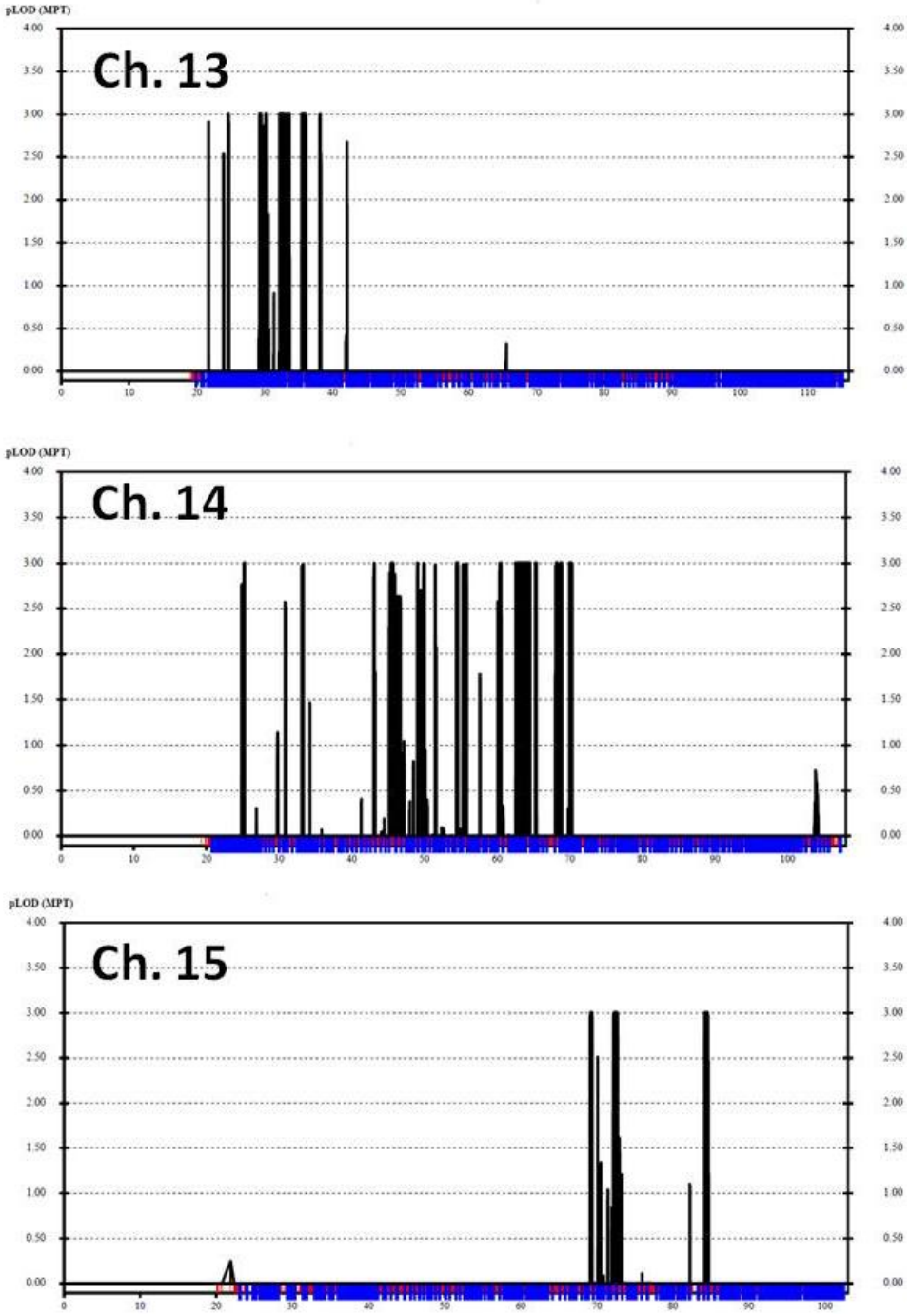


Fig. 5.7. Continued…

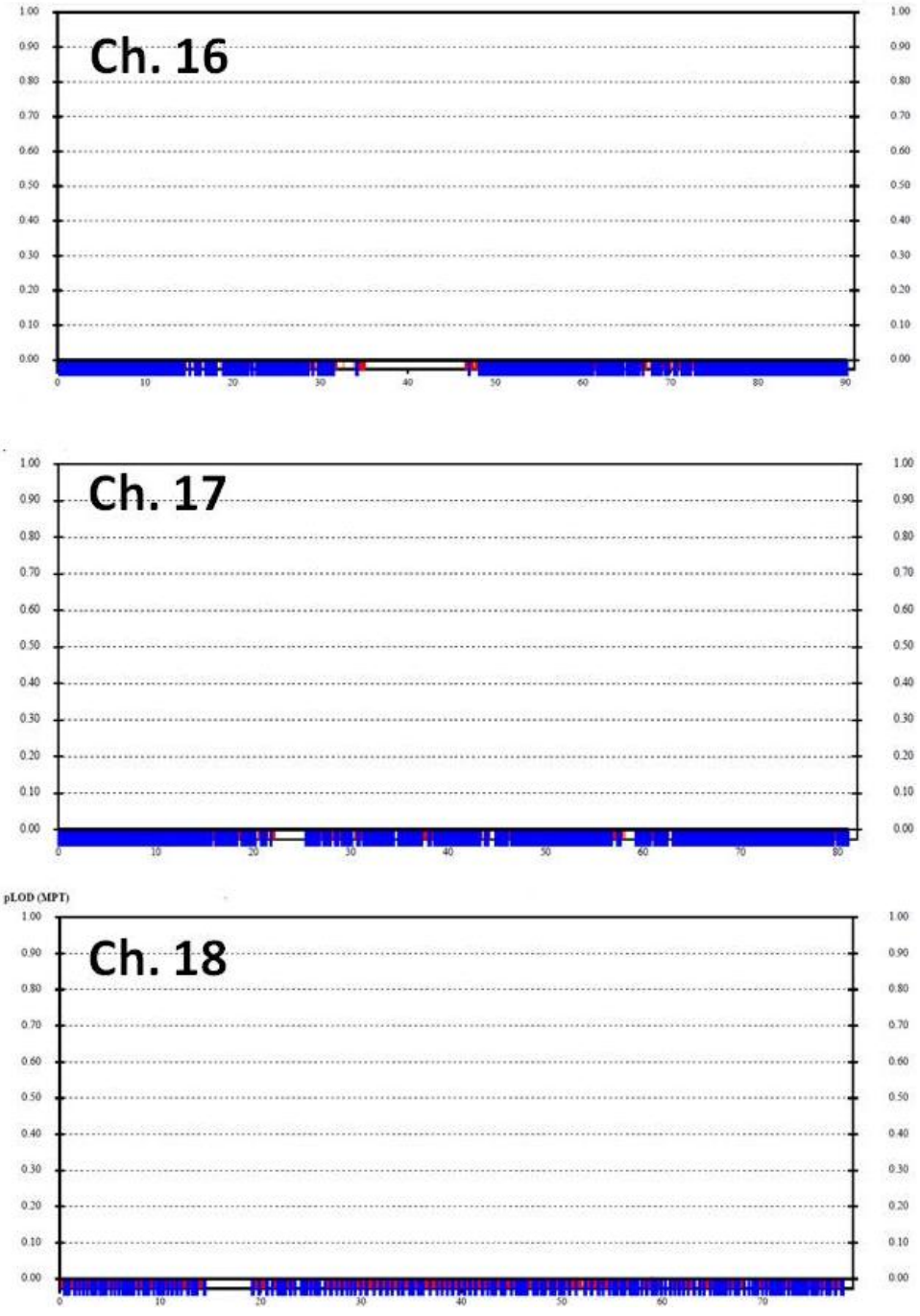


Fig. 5.7. Continued…

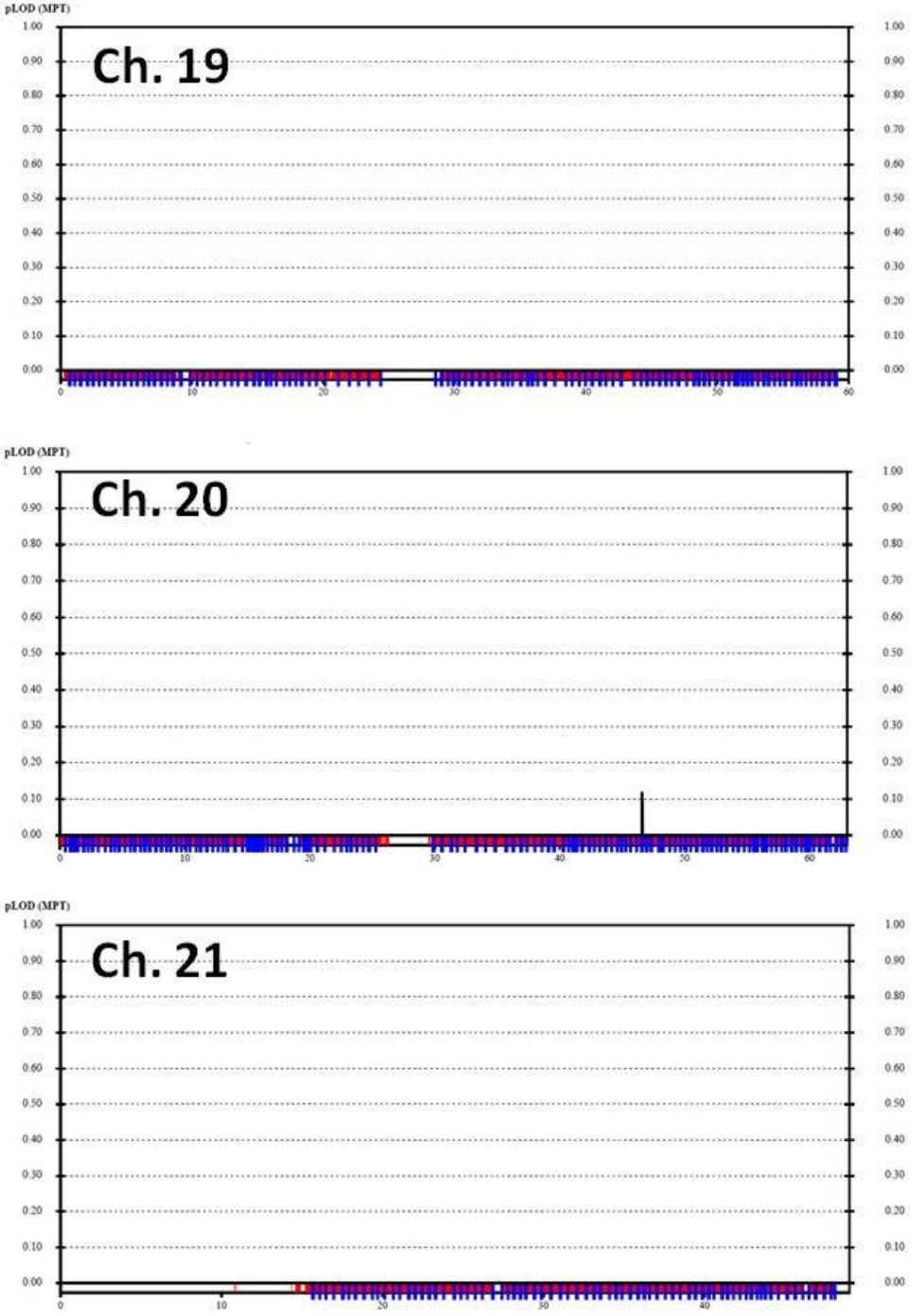


Fig. 5.7. Continued…

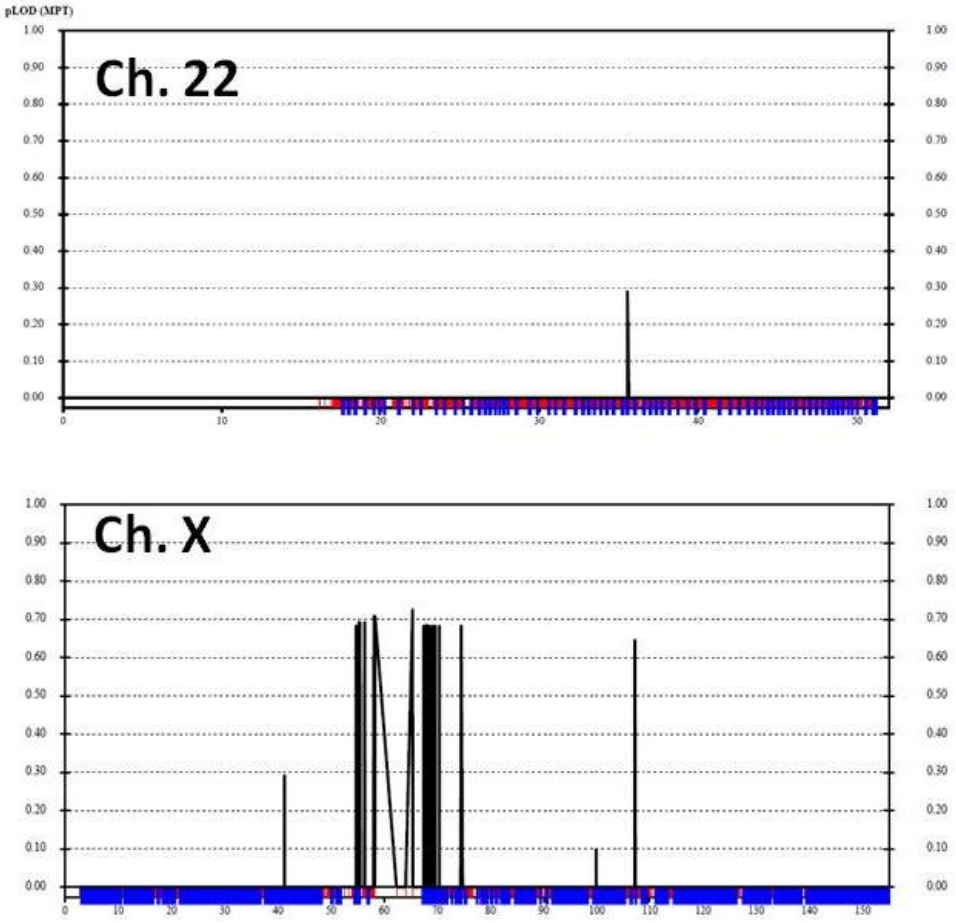


Fig. 5.7. Continued…

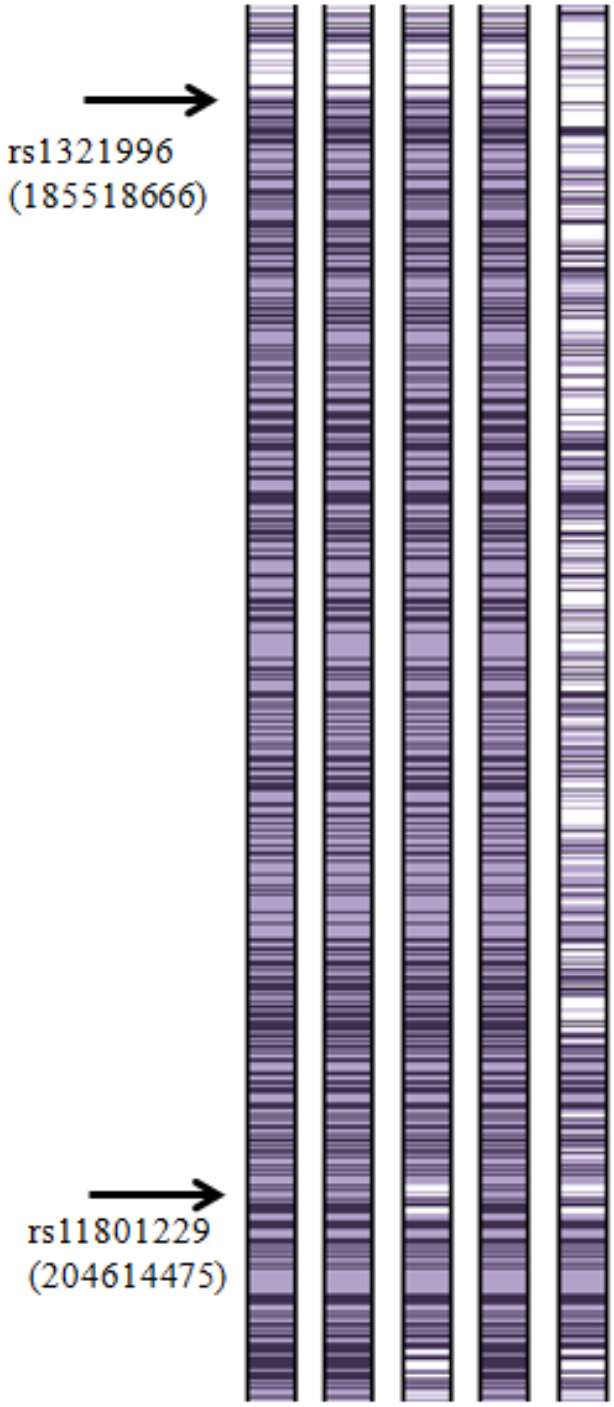


Figure 5.8. **Genotypes of the sibs based on SNP markers at 1q31 constructed by Homozygosity Comparison in Excel (HCiE) for microcephaly Family.** First four columns are for four affected sibs and the last column is for an unaffected sib. White bands are heterozygous genotypes.

5.3 Family C: Microcephaly with Intellectual Disability, Short Stature, and Brachydactyly (MIDSB)

5.3.1 Clinical description

Four affected and three unaffected subjects were examined in this family. In this kindred, microcephaly accompanied intellectual disability, skeletal anomalies, and skin phenotypes. Microcephaly and intellectual disability are the major features while nail dystrophy and minor bone anomalies are minor features. Microcephaly was congenial and non-progressive. Intellectual disability was mild in nature and the subjects could perform simple tasks, undertake self-care, and were good at communication. However, their learning abilities were compromised, so they were not able to attend school. The physical assessment showed disproportionate short stature, and skeletal, limb and skin anomalies (Table 5.8). Anthropometric and physical measurements are presented in Table 5.9. The propositus IV-7 (Fig. 5.9) was a jolly fellow of 20 years of age. He used to wander the streets with his affected younger brother (IV-8 in Fig. 5.9). His radiographic evaluation showed reduced skull diameter, protruding maxilla, prominent nose, and malocclusion of teeth (Fig. 5.9, d). A CT scan of the brain depicted diffuse thickening of skull bones, which was suggestive of hyperostosis. However, white–gray matter differentiation was unremarkable (Fig. 5.9, e,f). Limbs were short with a marked reduction of zeugopods (Fig. 5.9, g,h). In the upper limbs, there was bowing of radii and reduced spaces between the radii and ulnae. Further, bilateral crowding of carpals, short and hypoplastic metacarpals and fusion of the middle carpals was also observed. The middle segments of the

phalanges were short (brachydactyly A1/B1) (Fig. 5.9). Dermatological findings included atrophy of finger nails but not toe nails. In lower limbs, the tibiae and fibulae were short. In the left foot, severe clubbing was evident (Fig. 5.9). There were scars on the finger knuckles.

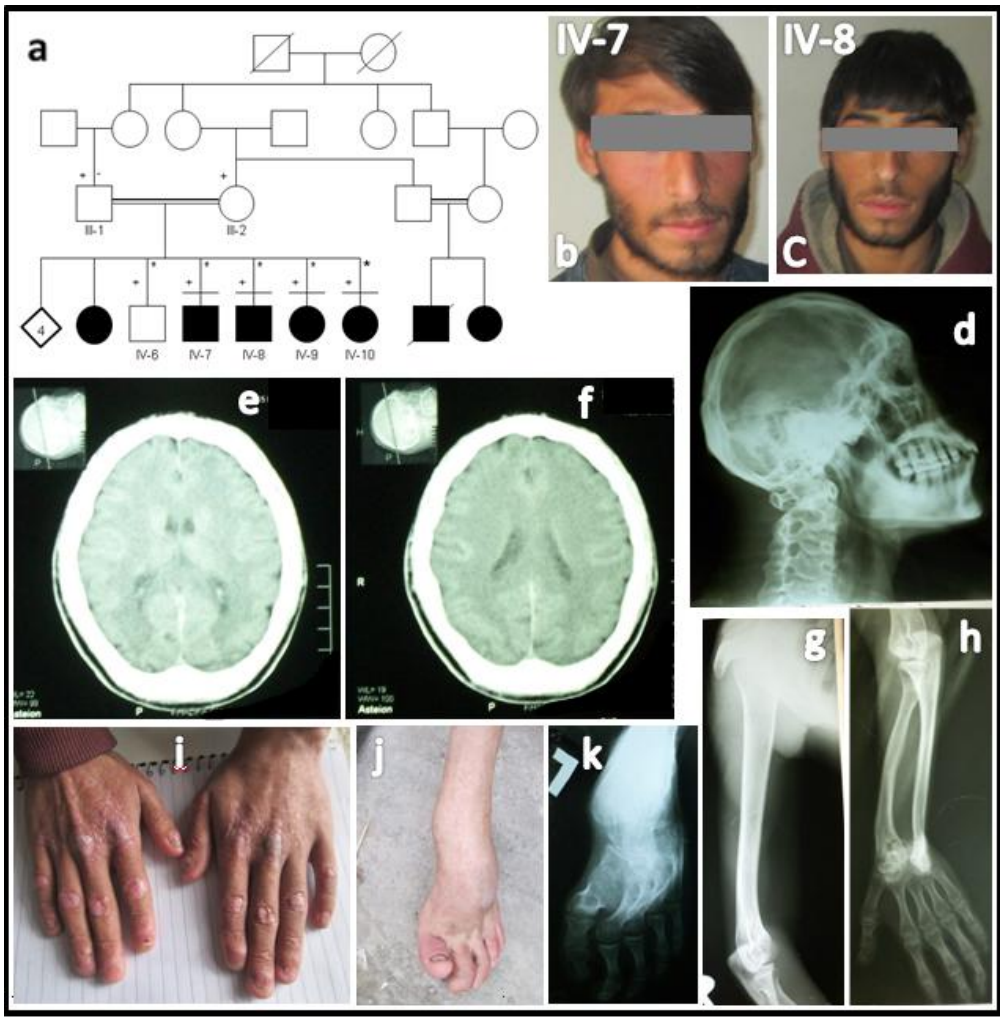


Figure 5.9. Pedigree (**a**) and phenotype (**b-k**) of the MIDSB Family. (**b, c**) Photographs of affected individuals IV-7 and IV-8. (**d–f**) Skull radiograph and CT scan of IV-7. (**g–k**). Photographs and roentgenograms of the limbs of IV-7.

Table 5.8. Summary of the clinical features in the affected siblings.

Parameter	**Individual**			
	IV-7	**IV-8**	**IV-9**	**IV-10**
Gender, age	M, 22	M, 20	F, 17	F, 15
Structural defects				
Microcephaly	+	+	+	+
Disproportionate short stature	+	+	+	+
Protruding upper jaw	+	+	+	+
Prominent nose	+	+	+	+
Anonychia	+	+	+	+
Brachydactyly, swelling of the phalangeal joints	+	+	+	+
Knuckle pads	++	++	+	+
Talipes	+	–	–	–
Behavioral and intellectual features				
Mild intellectual disability	+	+	+	+
Psychosis	+	+	+	+
Slow developmental history	++	+	++	+
Attention deficit	+	+	+	+
Mood instability	+	+	+	+
Friendly behaviour with strangers	++	++	+	+
No sense of social responsibility	++	++	++	++
Strong imaginative power	++	++	+	+
Aggressive behaviour	++	+	+	-
Temper tantrums	++	+	+	+
Roentgenographic observations				
Narrow spaces between vertebrae	+	NI	NI	NI
Crowding/missing carpals	+	NI	NI	NI
Short and narrowly spaced radii/ulnae and tibiae/fibulae	+	NI	NI	NI
Short middle phalanges in hands	+	NI	NI	NI
Short metacarpals	+	NI	NI	NI

+= present; ++= severe phenotype; –= absent; NI= no radiographic investigation.

Table 5.9. Anthropometric measurements of affected individuals.

Variable (in cm)	**Individual**			
	IV-7	**IV-8**	**IV-9**	**IV-10**
Head circumference	48	42	40	42
Neck circumference	30	27	26	28
Chest circumference	74	69	69	68
Standing height	152	137	134	139
Sitting height	84	76	74	77
Arm span	142	126	122	128

5.3.2 Multipoint linkage analysis and haplotype examination

Multipoint linkage analysis was carried out using SNP genome scan data. An autosomal recessive model with a disease allele frequency of 0.0001 was assumed. Homozygosity Mapper was employed to detect shared homozygosity regions >100 kb in the affected sibs. The data were aligned in MS-Excel to confirm shared homozygosity among the affected sibs and to detect homozygous intervals possibly identical by descent (IBD). IBD regions were further confirmed by constructing haplotypes by ALLEGRO. Linkage analysis showed 3 large regions of shared homozygosity across the whole genome, all yielding LOD scores >3.0: approximately 4 Mb at 13q13.3-32.2 (nucleotides 94,799,057-98,832,895), 7.5 Mb at 18p11.21-q11.2 (nucleotides 13,271,319-20,796,658), and 5.8 Mb at 18q21.33-22.1 (nucleotides 59,414,200-65,193,147) (Fig. 5.10). These regions were also detected as long stretches of shared homozygosity among the affected individuals but not in their unaffected sib. In these regions, the homozygous genotypes were validated as IBD by haplotype inspection (Figs. 5.11-5.13).

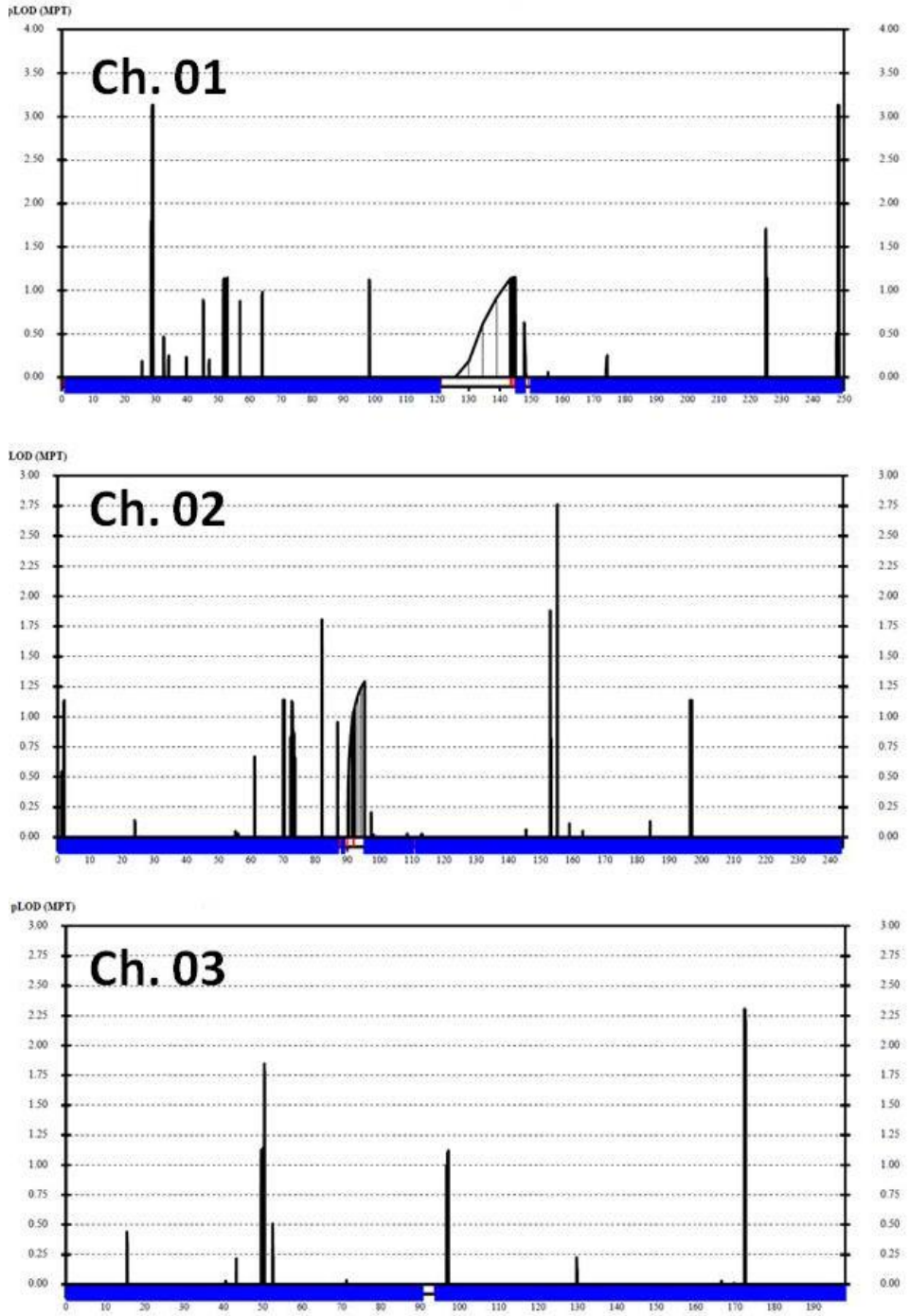


Figure 5.10. Multipoint LOD scores results for all autosomal chromosomes and PAR regions in the MIDSB microcephaly Family.

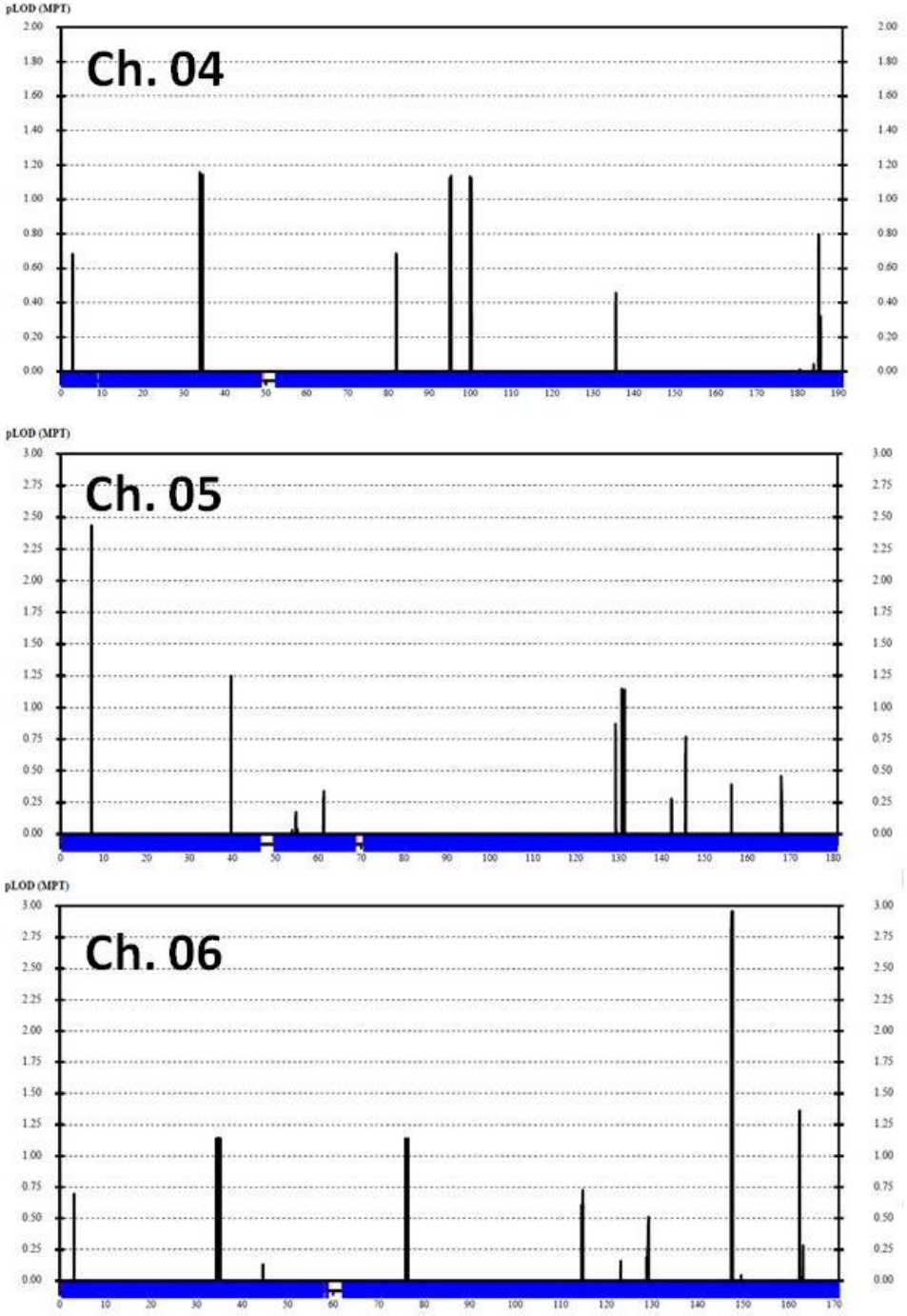


Fig. 5.10. Continued…

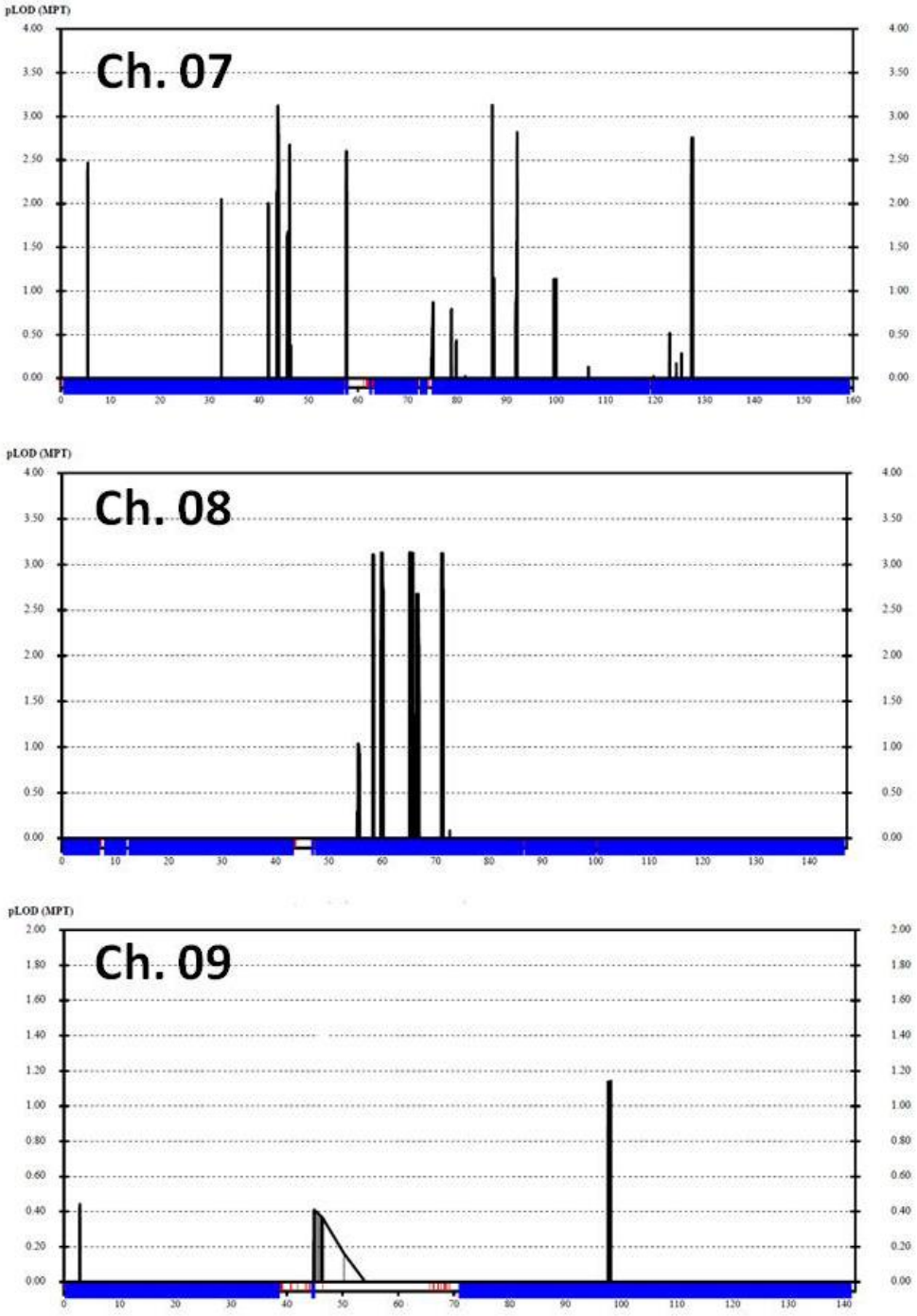


Fig. 5.10. Continued…

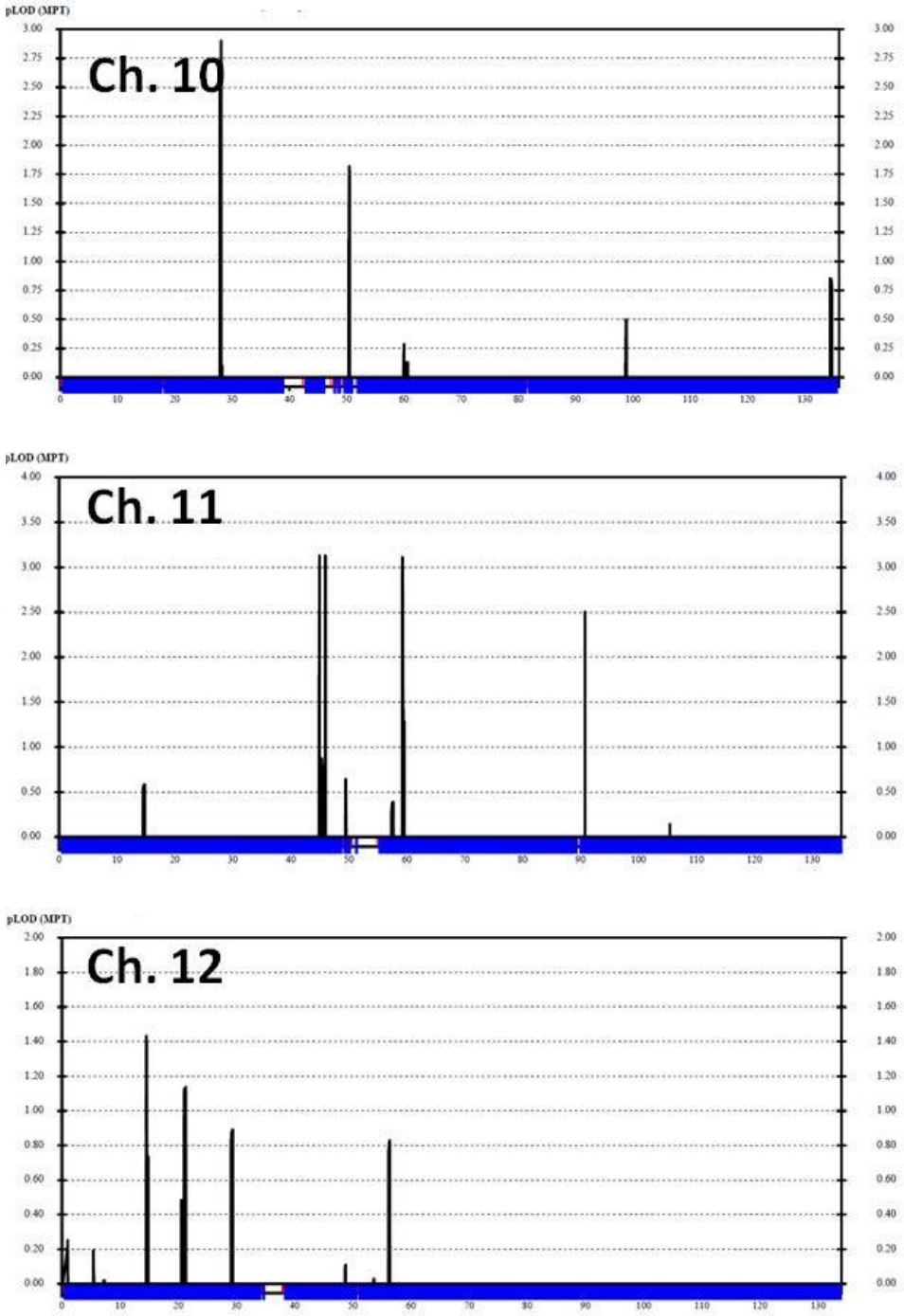


Fig. 5.10. Continued…

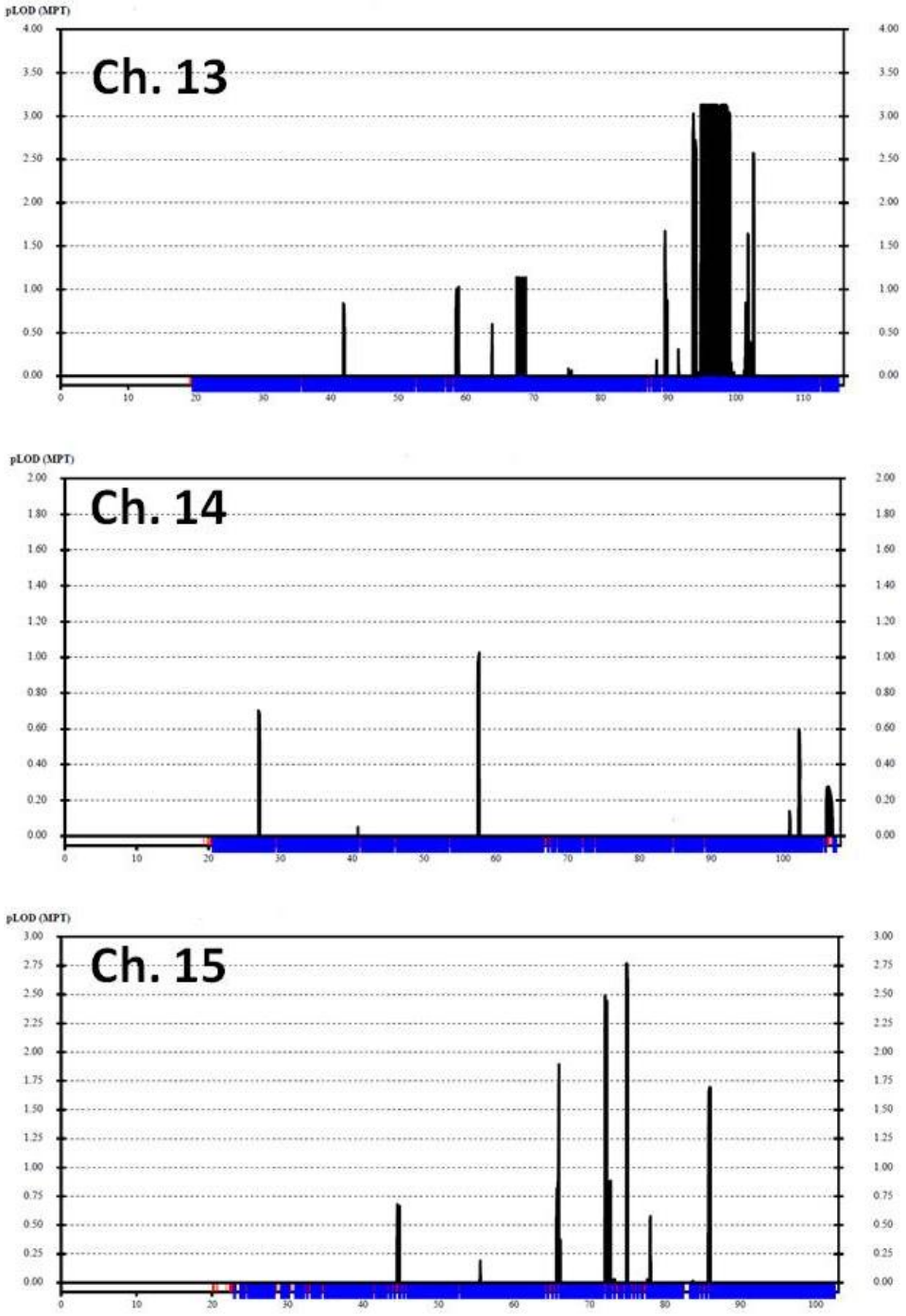


Fig. 5.10. Continued…

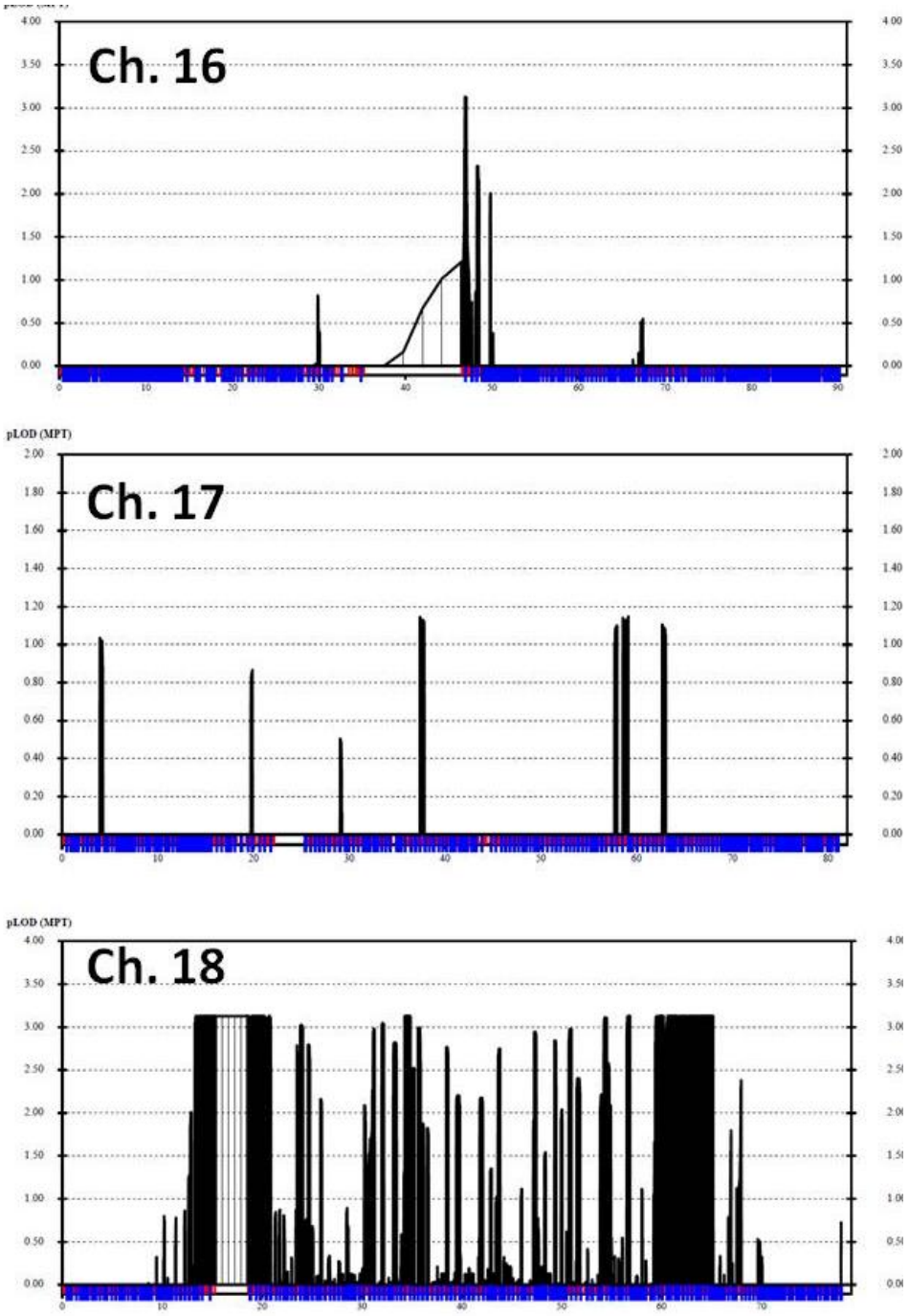


Fig. 5.10. Continued…

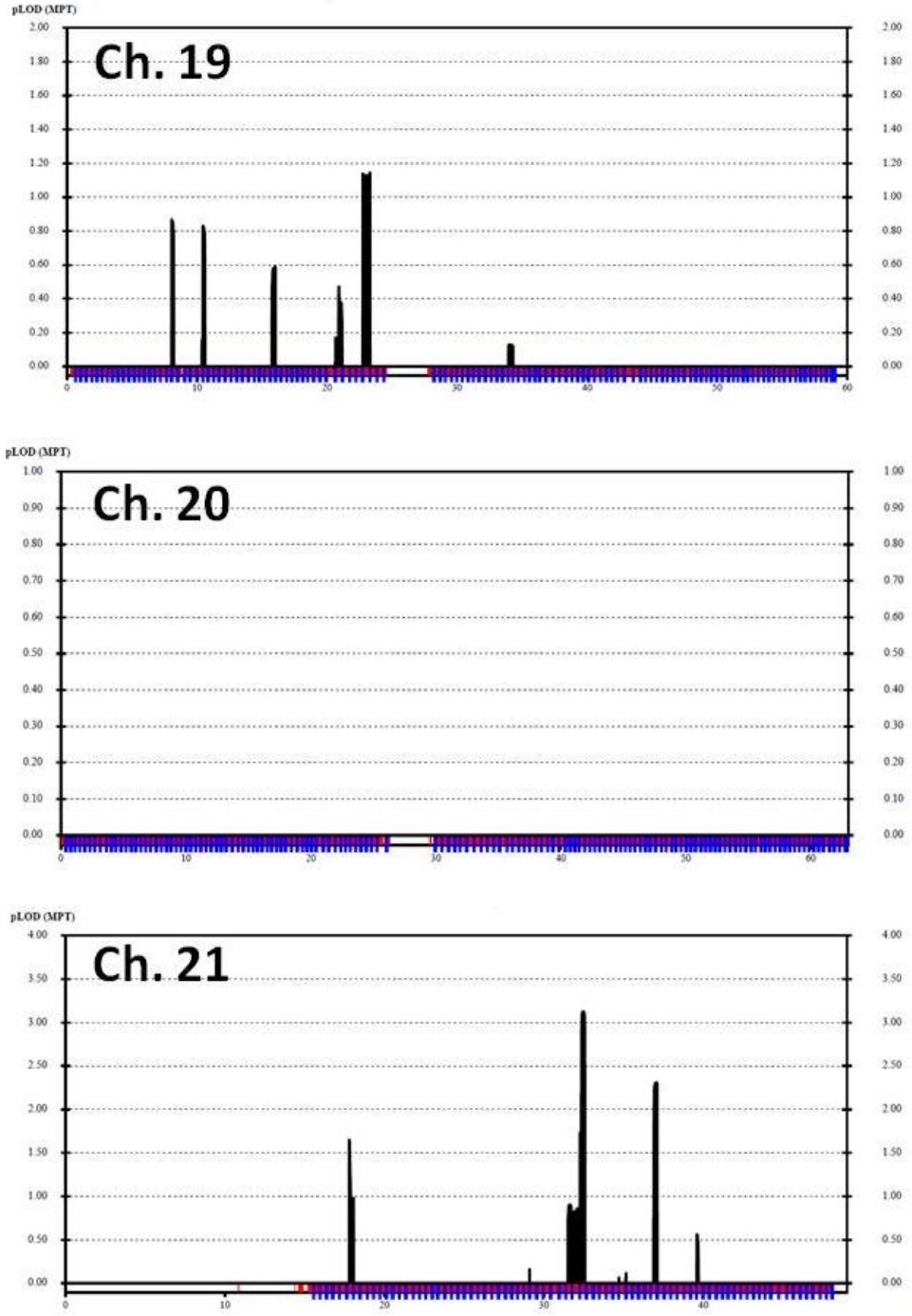


Fig. 5.10. Continued…

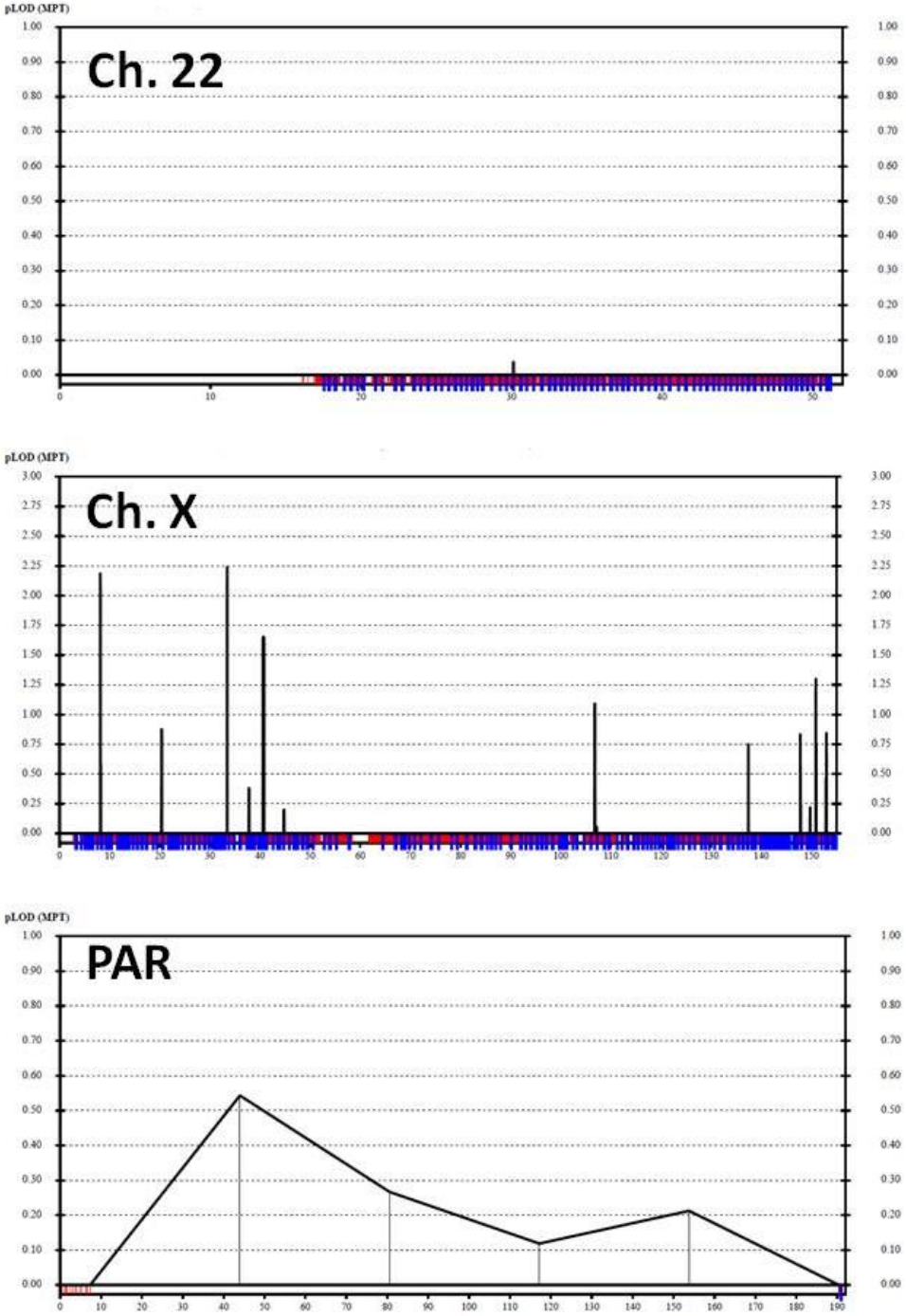


Fig. 5.10. Continued…

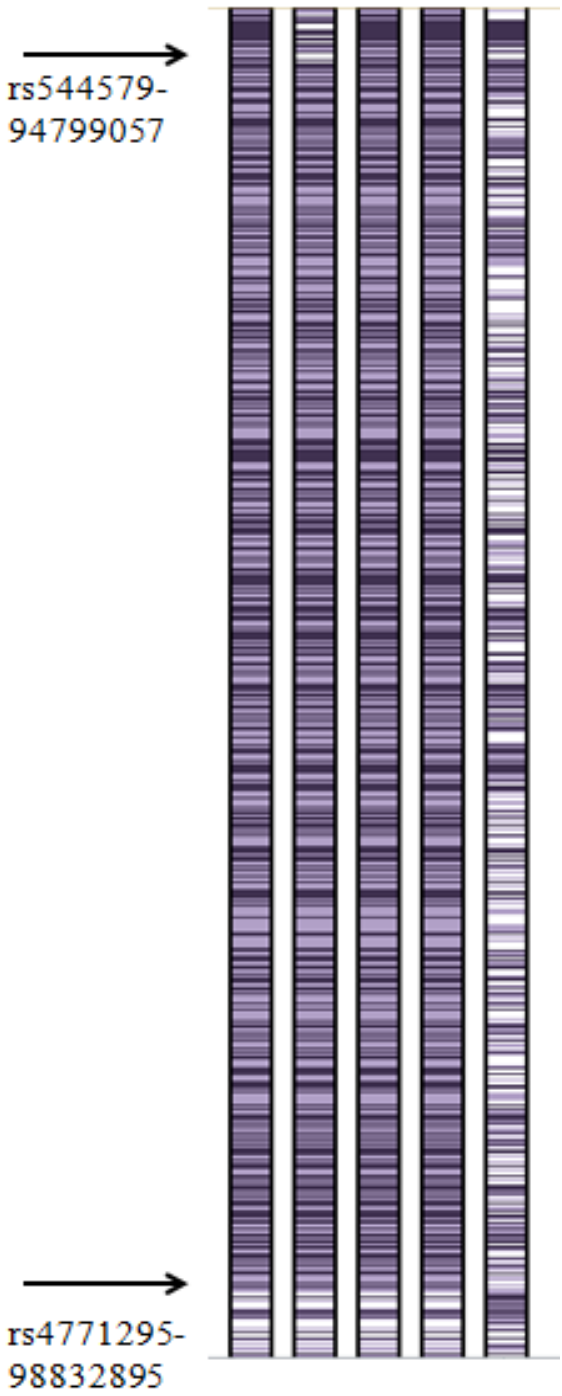


Figure 5.11. **The haplotype of SNP markers at 13q13.3-32.2 by Homozygosity Comparison in Excel (HCiE) for syndromic microcephaly Family.** First four columns are for 4 affected sibs and the last column is for the unaffected mother. White bands are heterozygous genotypes.

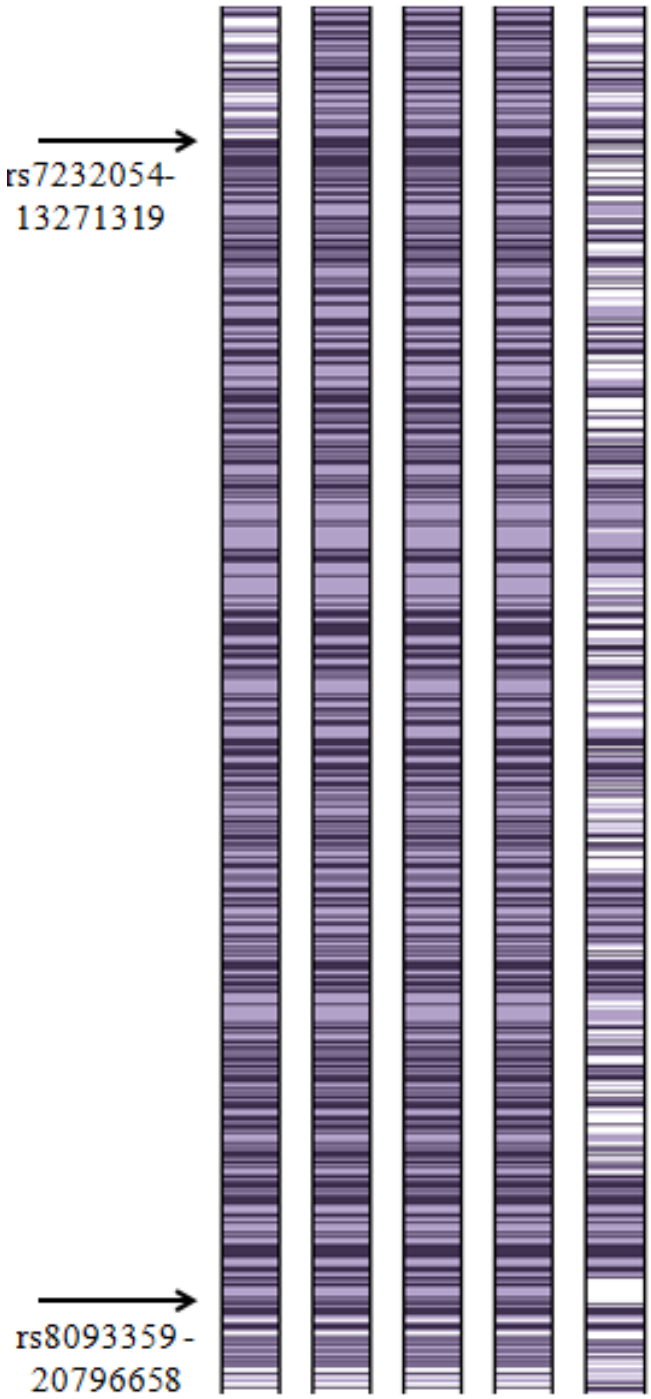


Figure 5.12. **The haplotype of SNP markers at 18p11.21-q11.2 by Homozygosity Comparison in Excel (HCiE) for syndromic microcephaly Family.** First four columns are for 4 affected sibs and the last column is for the unaffected mother.

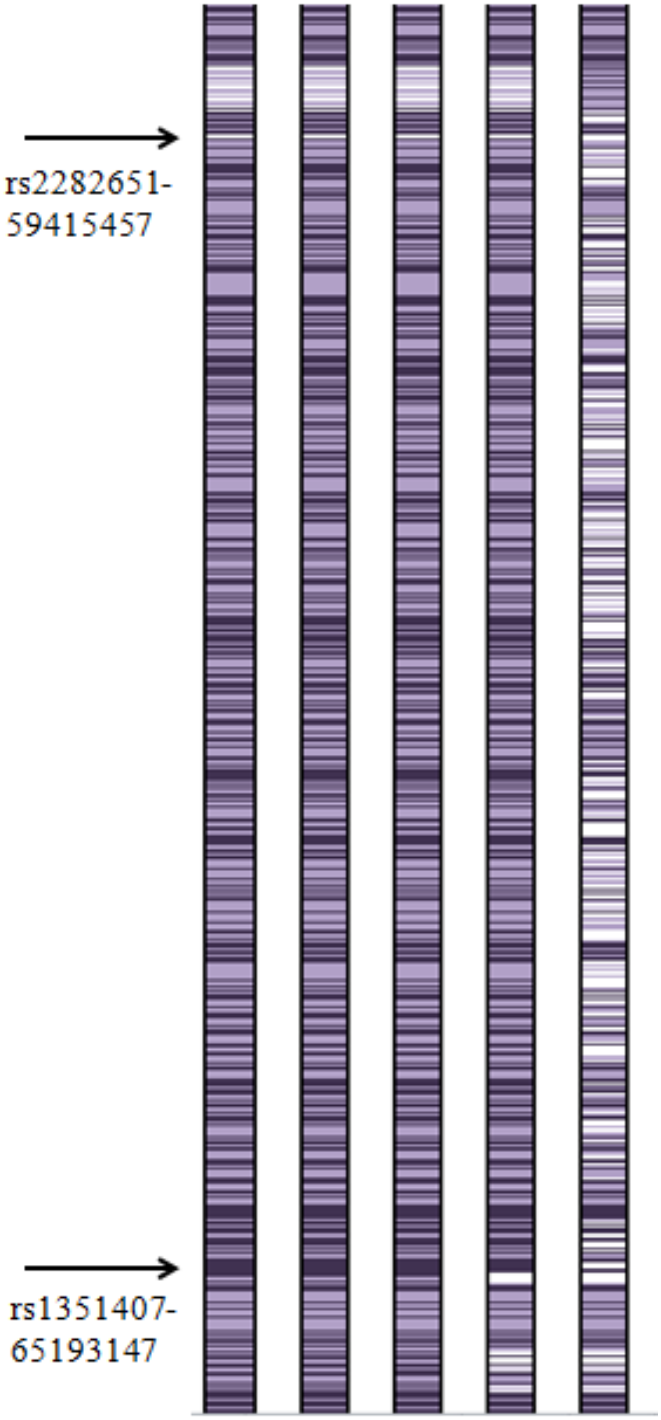


Figure 5.13. **The haplotype of SNP markers 18q21.33-q22.1 by Homozygosity Comparison in Excel (HCiE) for syndromic microcephaly Family.** First four columns are for 4 affected sibs and the last column is for the unaffected mother.

5.3.3 CNV analysis

SNP genotyping data were used to search for deletions and duplications by using cnvPartition (v3.2.0) CNV Analysis Plug-in for GenomeStudio (default parameters were used). No deletion or duplication was detected within the candidate regions.

5.3.4 Exome sequencing and evaluation of the variants

DNA sample of one patient (V-2) was exome sequenced (summary in Table 5.10). We obtained a resultant file of bioinformatics analysis of sequencing data from the company and then repeated the analysis.

Table 5.10. Summary of exome sequencing analysis of subject V-2 from the MIDSB Family with syndromic microcephaly.

Feature	**No.**
Read length (bp)	101.0
Target regions (bp)	51,189,318
% Coverage of target regions (more than 1X)	99.1
% Coverage of target regions (more than 10X)	97.5
Mean read depth of target regions	58.3
Number of total SNPs found	66,049
Number of coding SNPs found	19,576
Number of indels found	6,204
Number of coding indels found	470

The data for the regions of shared homozygosity were evaluated first for deleterious mutations. No candidate variant was found in intervals 13q13.3-32.2 and 18q21.33-22.1 (Tables 5.11, 5.12). In the third region (18p11.21-q11.2) (GRCh37/hg19), a two-base pair deletion in *RBBP8* exon 11 (c.1808_1809delTA) was detected and validated by Sanger sequencing (Fig. 5.14; Table 5.13). In order not to overlook a possible second homozygous pathogenic mutation, all of the rare/novel possibly deleterious mutations in the whole genome were filtered but none was in a region of homozygosity shared by the affected siblings. Thus, the possibility that the unusual features of the syndrome could be due to a recessive mutation in another gene was eliminated. No other exonic or splicing variant was detected in the candidate regions fitting the criteria of mutation. As evident in Tables 5.11-5.13, all other exonic and splicing even UTR variants were observed with high frequency. However, intronic variants were found without any reported frequencies in candidate regions; but, they were eliminated because they were intronic. Candidate Variant at Chr1:20573598 in RBBP8 was a two-base pair deleterious frame-shift mutation (p.Ile603Lysfs*7). It was also absent in other exome files generated simultaneously in our lab. Further analysis of the candidate variant was done by realignment of the raw reads to the reference genome (hg19/GRCh37 assembly), allowing approximately 2% base-pair mismatch in read alignments. The alignments were visualized and validated via program BamView. Sanger sequencing was performed to validate the variant in the family members. Affected individuals were homozygous and carrier mother heterozygous for the variant whereas unaffected sib and cousins were wild type (Fig. 5.3). None of the 100 Pakistani control samples tested carried this mutation.

Table 5.11. Variants in the exome sequence variant list in the candidate homozygous region 13q13.3-32.2 (nucleotides 94,799,057-98,832,895).

Location	**Ref Base**	**Obs Base**	**Hom Het**	**SNP Quality**	**Total Depth**	**Region**	**Gene**	**SNP ID**	**1000G Freq.**
92796964	G	A	hom	51.5	4	intronic	*GPC5*	rs12429836	0.34
95034763	G	A	hom	222	32	exonic	*GPC6*	rs2274020	0.14
95096505	G	A	hom	38	8	intronic	*DCT*	rs7339198	0.67
95097956	C	T	hom	159	68	intronic	*DCT*	rs755684	0.63
95118735	T	C	hom	110	9	intronic	*DCT*	rs12864325	0.13
95118741	C	T	hom	105	10	intronic	*DCT*	rs12877849	0.13
95227126	GAC	-	het	85.5	10	intronic	*TGDS*	.	.
95230877	G	A	hom	123	52	intronic	*TGDS*	rs16949940	0.27
95254382	A	G	hom	66.3	5	intronic	*GPR180*	rs9556404	0.62
95275576	C	G	hom	154	29	intronic	*GPR180*	rs4296139	0.57
95363884	G	C	hom	122	11	exonic	*SOX21*	rs9556411	0.19
95364337	T	G	hom	222	61	UTR5	*SOX21*	rs2253604	0.81
95673791	A	C	hom	169	72	UTR3	*ABCC4*	rs3742106	0.40
95673939	A	G	hom	140	24	intronic	*ABCC4*	rs9524765	0.93
95714976	C	T	hom	157	32	exonic	*ABCC4*	rs57270423	0.78
95715152	GAAA	-	hom	129	11	intronic	*ABCC4*	.	.
95726541	A	G	hom	222	79	exonic	*ABCC4*	rs1189466	0.89
95727780	T	C	hom	222	47	exonic	*ABCC4*	rs1678339	0.88
95768362	T	C	hom	203	15	intronic	*ABCC4*	rs2296652	0.06
95829870	T	C	hom	138	8	intronic	*ABCC4*	rs1751005	0.79
95847020	A	G	hom	130	29	intronic	*ABCC4*	rs2274403	0.51
95861804	A	G	hom	135	17	exonic	*ABCC4*	rs899494	0.82
95862896	A	G	hom	4.61	2	intronic	*ABCC4*	rs899496	0.73
95863143	A	G	hom	145	12	intronic	*ABCC4*	rs899497	0.73
95899354	A	G	hom	222	49	intronic	*ABCC4*	rs4148437	0.24
95900071	C	T	hom	203	34	intronic	*ABCC4*	rs9556466	0.24
96205359	A	C	hom	12.2	3	intronic	*CLDN10*	.	0.99
96258386	C	T	hom	185	19	intronic	*DZIP1*	rs9556491	0.11
96285672	C	T	hom	117	10	intronic	*DZIP1*	rs3742151	0.12
96294267	G	T	het	4.13	11	exonic	*DZIP1*	.	.
96506664	A	G	hom	175	53	exonic	*UGGT2*	rs11070154	0.34
96529383	GCCT	-	hom	157	5	intronic	*UGGT2*	.	.
96540019	C	T	hom	23.2	6	intronic	*UGGT2*	rs12871416	0.13
96540141	T	C	hom	170	28	intronic	*UGGT2*	rs9525075	0.97
96540143	T	C	hom	169	28	intronic	*UGGT2*	rs9525076	0.97
96540204	T	G	hom	181	22	exonic	*UGGT2*	rs12876018	0.36
96555149	C	T	hom	222	86	exonic	*UGGT2*	rs33949518	0.13
96599271	C	T	hom	222	24	intronic	*UGGT2*	rs7986005	0.38
96599433	A	C	hom	111	39	intronic	*UGGT2*	rs9556512	0.36
96599445	C	T	hom	222	37	intronic	*UGGT2*	rs9556513	0.38
96636170	G	C	hom	222	44	intronic	*UGGT2*	rs9562008	0.34

96638636	A	C	hom	209	30	exonic	*UGGT2*	rs816142	0.86
96638651	C	T	hom	222	25	exonic	*UGGT2*	rs12863903	0.10
96641695	T	G	hom	66.1	7	intronic	*UGGT2*	rs555120	0.97
96705739	T	C	hom	101	12	upstream	*UGGT2*	rs2277419	0.41
97079618	A	G	hom	21	3	intronic	*HS6ST3*	rs9556587	0.32
98017211	A	G	hom	53	9	intronic	*MBNL2*	rs3825422	0.22
98640170	A	C	hom	39	11	intronic	*IPO5*	.	.
98828260	T	A	hom	40	3	UTR3	*RNF113B*	rs774690	0.79
98829176	T	C	hom	222	46	exonic	*RNF113B*	rs628778	0.79
98860749	A	C	het	10.4	5	intronic	*FARP1*	rs285109	0.41
98896776	C	T	het	139	65	exonic	*FARP1*	rs7318267	0.40

Table 5.12. Variants in the exome sequence variant list in the candidate homozygous region 18q21.33-q22.1 (nucleotides 59,414,200- 65,193,147).

Location	Ref Base	Obs Base	Hom Het	SNP Quality	Total Depth	Region	Gene	SNP ID	1000G Freq.
59780393	T	A	hom	132	33	exonic	PIGN	rs3862712	0.08
59807617	-	T	hom	11.8	33	intronic	PIGN	.	.
59814324	G	C	hom	174	37	exonic	PIGN	rs9320001	0.83
59929009	G	A	hom	72.5	4	intronic	KIAA1468	rs9945414	0.15
59947118	G	A	hom	219	68	intronic	KIAA1468	rs2980981	0.71
59958678	G	A	hom	78	11	intronic	KIAA1468	rs2980983	0.36
60021761	C	T	hom	222	41	exonic	TNFRSF11A	rs35211496	0.08
60025472	TT	-	hom	53.5	49	intronic	TNFRSF11A	.	.
60027149	T	A	hom	209	32	intronic	TNFRSF11A	rs6567271	0.61
60027171	C	T	hom	219	38	intronic	TNFRSF11A	rs6567272	0.61
60027241	C	T	hom	222	41	exonic	TNFRSF11A	rs1805034	0.61
60027361	G	A	hom	133	5	intronic	TNFRSF11A	rs9653064	0.57
60034074	-	C	het	51.5	10	intronic	TNFRSF11A	.	.
60036083	A	G	hom	222	79	exonic	TNFRSF11A	rs8092336	0.98
60051941	C	T	hom	128	9	intronic	TNFRSF11A	rs56231704	0.12
60191428	G	A	hom	203	94	exonic	ZCCHC2	rs7229802	0.72
60206800	G	T	hom	38	3	intronic	ZCCHC2	rs55722786	0.14
60232405	C	T	hom	143	23	intronic	ZCCHC2	rs4940561	0.99
60237388	A	G	hom	211	53	exonic	ZCCHC2	rs8096750	0.82
60241732	C	T	hom	222	173	exonic	ZCCHC2	rs3744926	0.46
60243876	G	A	hom	123	17	UTR3	ZCCHC2	rs11557713	0.35
60562298	A	T	hom	222	50	exonic	PHLPP1	rs624821	0.50
60572448	T	-	hom	41.5	63	intronic	PHLPP1	rs3217528	.
60985879	T	C	hom	209	140	exonic	BCL2	rs1801018	0.28
61022634	T	C	hom	222	95	intronic	KDSR	rs1809319	0.57
61022791	C	T	hom	222	77	exonic	KDSR	rs2003149	0.56
61060873	G	A	hom	42.3	5	intronic	VPS4B	rs898890	0.61
61066463	A	G	hom	222	61	intronic	VPS4B	rs2276317	0.61
61067959	T	C	hom	222	33	intronic	VPS4B	rs952245	0.52
61170467	G	A	hom	39.3	5	intronic	SERPINB5	rs1840557	0.53
61170500	T	C	hom	110	13	intronic	SERPINB5	rs1455557	0.55
61170721	T	C	hom	222	58	exonic	SERPINB5	rs1455556	0.58
61170782	A	G	hom	222	59	exonic	SERPINB5	rs1455555	0.40
61170925	T	C	hom	222	105	exonic	SERPINB5	rs8093128	1.00
61259508	C	T	hom	155	16	intronic	SERPINB13	rs78722153	0.02
61262367	C	T	hom	174	16	exonic	SERPINB13	rs17071357	0.05
61309139	G	A	hom	86	13	intronic	SERPINB4	rs2518060	0.12
61326505	T	A	hom	156	23	intronic	SERPINB3	rs8089032	0.97
61326597	A	C	hom	222	27	intronic	SERPINB3	rs8085479	0.99
61326602	A	G	hom	163	28	intronic	SERPINB3	rs8085482	0.99
61326609	T	G	hom	198	28	intronic	SERPINB3	rs8089201	0.99

61326630	-	T	hom	127	13	intronic	SERPINB3	rs11424286	.
61377579	A	C	hom	110	15	exonic	SERPINB11	rs1395268	0.85
61378743	G	T	hom	112	20	intronic	SERPINB11	rs1506424	0.44
61379838	T	G	hom	222	47	exonic	SERPINB11	rs4940595	0.44
61383354	C	T	hom	222	39	exonic	SERPINB11	rs17071550	0.16
61387312	G	A	hom	222	72	exonic	SERPINB11	rs1506418	0.44
61387333	T	A	hom	183	61	exonic	SERPINB11	rs1506419	0.44
61387389	G	A	hom	73	25	exonic	SERPINB11	rs1506420	0.44
61390316	C	A	hom	154	53	exonic	SERPINB11	rs1395265	0.75
61390332	C	T	hom	222	62	exonic	SERPINB11	rs1395266	0.75
61390361	T	C	hom	213	75	exonic	SERPINB11	rs1395267	0.78
61390570	T	G	hom	178	77	exonic	SERPINB11	rs953695	0.74
61390649	C	T	hom	222	32	UTR3	SERPINB11	rs953696	0.74
61391467	-	TCT	het	194	10	downstream	SERPINB11	.	.
61391506	TCT	-	hom	214	17	downstream	SERPINB11	rs144622175	0.76
61391697	A	G	hom	53	10	downstream	SERPINB11	rs12326861	0.75
61393362	G	A	hom	49.1	7	intergenic	SERPINB11	rs12962610	0.75
61393386	G	C	hom	96	13	intergenic	SERPINB11	rs9950640	0.75
61393442	T	C	hom	182	43	intergenic	SERPINB11	rs9965604	0.78
61393644	T	A	hom	222	27	intergenic	SERPINB11	rs1893563	0.75
61394140	T	C	hom	37.3	5	intergenic	SERPINB11	rs8097905	0.75
61394143	C	G	hom	44.3	5	intergenic	SERPINB11	rs8093673	0.75
61394300	C	T	hom	222	52	intergenic	SERPINB11	rs8097334	0.74
61394403	A	G	hom	101	16	intergenic	SERPINB11	rs8098025	0.75
61463784	G	T	hom	31	10	intronic	SERPINB7	rs1701648	0.40
61465810	A	G	hom	170	9	intronic	SERPINB7	rs1720858	0.61
61468306	G	C	hom	184	21	intronic	SERPINB7	rs2689399	0.63
61575120	G	T	hom	67.1	7	intergenic	SERPINB2	rs1916661	0.11
61647069	G	A	hom	222	52	exonic	SERPINB8	rs1944270	0.34
61654463	A	G	hom	222	57	exonic	SERPINB8	rs3826616	0.55
61765052	G	A	hom	95	11	ncRNA_intronic	LINC00305	rs75049510	0.03
63476810	C	T	hom	166	9	intronic	CDH7	rs11151216	0.95
63489378	C	T	hom	214	42	exonic	CDH7	rs8097752	0.90
63491797	A	G	hom	102	16	intronic	CDH7	rs974080	0.25
63511425	CG	-	hom	39.2	2	intronic	CDH7	.	.
63530016	A	G	hom	222	109	exonic	CDH7	rs2291343	0.68
65178120	-	AC	hom	214	13	UTR3	DSEL	rs56128444	.
65179829	G	A	hom	187	64	exonic	DSEL	rs2279269	0.47
66358403	G	T	hom	165	13	intronic	TMX3	rs309228	0.24

Table 5.13. Variants in the exome sequence variant list in the candidate homozygous region 18p11.21-q11.2 (nucleotides 13,271,319- 20,796,658).

Location	**Ref Base**	**Obs Base**	**Hom Het**	**SNP Quality**	**Total Depth**	**Region**	**Gene**	**SNP ID**	**1000G Fre**
13643179	C	A	hom	156	21	intronic	*C18orf1*	rs11080665	0.57
13666768	A	G	hom	62.1	7	intronic	*FAM210A*	rs1149356	0.64
13732053	A	G	hom	152	11	intronic	*RNMT*	rs4797806	0.88
13732059	C	T	hom	33.1	7	intronic	*RNMT*	rs4797807	0.88
13762011	C	A	hom	222	46	UTR3	*RNMT*	rs4461147	0.91
14105923	T	C	hom	106	35	exonic	*ZNF519*	rs2159940	0.98
14779969	C	G	hom	141	54	exonic	*ANKRD30B*	rs9675365	0.46
14837178	T	-	hom	48.5	41	intronic	*ANKRD30B*	rs56391489	.
18539976	A	G	hom	222	61	intronic	*ROCK1*	.	.
18983766	C	T	hom	222	31	intronic	*GREB1L*	rs2117317	0.55
19020160	G	A	hom	188	26	intronic	*GREB1L*	rs9948725	0.42
19024150	T	G	hom	90	15	intronic	*GREB1L*	rs4368249	0.39
19034556	T	C	hom	170	26	intronic	*GREB1L*	rs3794938	0.36
19048666	A	G	hom	222	30	intronic	*GREB1L*	rs8083036	0.36
19075808	C	T	hom	75	9	intronic	*GREB1L*	rs3764502	0.28
19079537	A	G	hom	63.1	6	intronic	*GREB1L*	rs8087553	0.99
19079853	A	G	Hom	222	83	exonic	*GREB1L*	rs4800747	0.99
19100517	T	C	Hom	103	5	intronic	*GREB1L*	rs1395251	0.99
19100762	-	CTT	Hom	201	9	exonic	*GREB1L*	rs10654340	.
19110466	T	-	Het	67.7	6	intronic	*ESCO1*	.	.
19119801	C	T	Hom	98	20	intronic	*ESCO1*	rs1818274	0.87
19119992	TA	-	Hom	160	19	intronic	*ESCO1*	rs58118596	.
19153494	C	T	Hom	222	98	exonic	*ESCO1*	rs973730	0.87
19154143	G	A	Hom	222	73	exonic	*ESCO1*	rs13381941	0.52
19192255	C	T	Hom	164	37	upstream	*SNRPD1*	rs2850556	0.54
19208851	A	G	Hom	66	11	intronic	*SNRPD1*	rs2847117	0.54
19282519	-	GTTTT	Hom	139	5	intronic	*ABHD3*	rs61602023	.
19408950	C	T	Hom	151	17	intronic	*MIB1*	rs9989532	0.94
19423964	T	-	Hom	214	27	intronic	*MIB1*	rs3216119	.
20555219	T	C	Hom	175	31	intronic	*RBBP8*	rs2336916	0.92
20564808	CAT	-	Hom	196	12	intronic	*RBBP8*	rs34354995	.
20573598	**TA**	**-**	**Hom**	**214**	**89**	**exonic**	***RBBP8***	**.**	**.**
20964769	A	G	hom	151	10	intronic	*TMEM241*	rs8094375	0.982
20965034	G	A	het	57	36	intronic	*TMEM241*	rs62095860	0.198
21017836	T	C	hom	200	24	UTR5	*TMEM241*	rs7226498	0.924
21043106	T	C	het	225	52	intronic	*RIOK3*	rs72884533	0.350

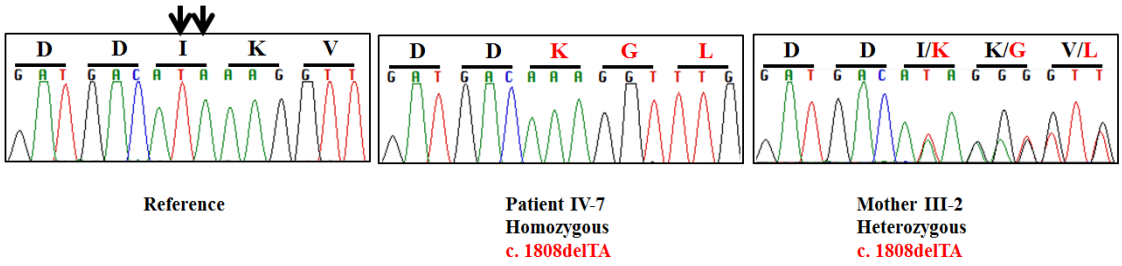


Figure 5.14. Electrophoretograms showing *RBBP8* c.1808_1809delTA.

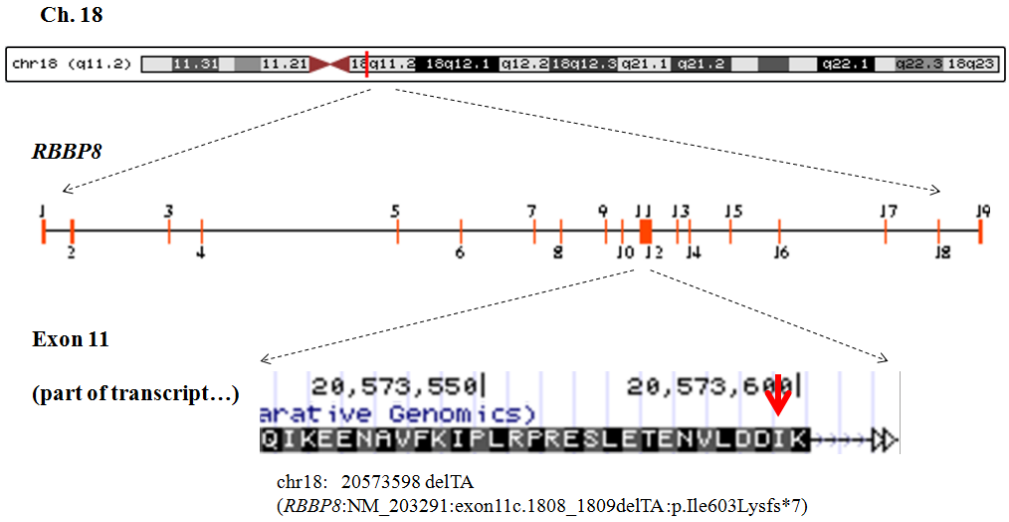


Figure 5.15. Genomic organization and location of mutation of *RBBP8* at chr18q11.2.

In conclusion, an inbred Pakistani family with homozygous *RBBP8* mutation c.1808_1809delTA (p.Ile603Lysfs*7) and afflicted with microcephaly, mild intellectual disability and a different combination of skeletal, limb and dermatological defects is presented.

The results obtained in this family have been published in Am J Med Genet:

Mumtaz S, Yildiz E, Jabeen S, Khan A, Tolun S, Malik S. Expanding the Spectrum of RBBP8 Morphopathies: Microcephaly, Intellectual Disability, Short Stature, Brachydactyly and Talipes. Am J Med Genet A. 167A(12):3148-52, 2015. doi:10.1002/ajmg.a.37299.

5.4 Family D: Autosomal Dominant Postaxial Polydactyly, Camptodactyly and Zygodactyly (ADPCZ)

5.4.1 Clinical description

In the present family, 21 members (14 males, 7 females) were found to be afflicted with postaxial polydactyly (Fig. 5.16). Twelve affected and seven unaffected subjects were examined. The phenotype was bilateral in most of the subjects and both hands and feet were involved. The additional digits at the postaxial axis of the autopods were completely developed in the majority of the affected subjects, hence, compatible with postaxial type A phenotype. In a few of the subjects, the extra digit was only evident as a small knob (Fig. 5.16). Therefore, both postaxial type A and type B phenotypes were seen in the family. Additionally, the affected subjects had various other digital anomalies such as camptodactyly of certain fingers, wide-spaced 1st and 2nd toes, and webbing of postaxial toes. Three of the subjects exhibited hypoplasia of the third toe. Four of the latter symptoms, i.e., camptodactyly of certain fingers, wide-spaced 1st and 2nd toes, webbing of postaxial toes, and hypoplasia of the third toe, are not characteristics of postaxial polydactyly and overlap with polydactyly type II or polysyndactyly. The supernumerary digits were inconvenient for the majority of subjects as they cause hindrance while performing daily activities such as putting on clothes. The summary of the clinical findings in the family are presented in Table 5.14.

Figure 5.16. Photographs and radiographs of the affected members of ADPCZ Family with postaxial polydactyly.

Table 5.14. Phenotypic features of members of ADPCZ Family with polydactyly.

Variable/ features	IV-6	IV-13	III-4	IV-15	III-12	IV-2	IV-16	IV-4	III-2	IV-17	Concordance
Gender, age	M, 2	F, 20	F, 25	M, 15	M, 50	M, 9	M, 2	M, 5	M, 38	F, 7	
Polydactyly in right hand	B	B, minor protuberance	B, minor protuberance	A	B, minor protuberance	A	A	B	A	A	10/10
Polydactyly in left hand	A	B, minor protuberance	A	A	B, minor protuberance	A	A	B	-	A	9/10
Polydactyly in right foot	-	A	A	A	-	A	A	A	A	A	8/10
Polydactyly in left foot	-	A	A	A	A	A	A	A	A	A	9/10
Camptodactyly	2-3 fingers	Fifth fingers	-	Fifth fingers	-	Fifth fingers	Fifth fingers	4-5th fingers	Fifth fingers	Fifth fingers	8/10
Wide space between hallux and second toe	-	Yes	Yes	Yes	Yes	-	Yes	-	Yes	-	6/10
Zygodactyly	-	Partial	-	-	-	Nearly complete	-	-	Nearly complete	-	3/10
Third toe hypoplastic	-	-	Right	-	Bilateral	-	-	-	Bilateral	-	3/10
Other features	Bifid fifth finger		6th toe larger than 5th		Fusion of 5-6 toes (left)	Fusion of 3-6 toes			over-riding last toe		

A= postaxial type A; B= postaxial type B; ++ = predominant feature; + = minor feature

5.4.2 Multipoint linkage analysis and haplotype examination

On the SNP genome scan data of 10 (8 affected, 2 normal) family members, linkage analysis were implemented and multipoint LOD scores were calculated with an autosomal dominant disease model by assuming disease allele frequency of 0.001 and a penetrance of 99%. The LOD score calculations were performed in several rounds. At the beginning, modified pedigree with fewer subjects were used for linkage analyses due to computational restrictions. The marker window sizes of 40, 50 and 100 were used. All the genotyped individuals were included in the subsequent rounds of calculations and a marker window size of 30 and 70 was adopted. The LOD scores intervals with >2.0 scores were examined more closely and re-evaluated with iterated models of reduced penetrances and varying marker window sizes. Haplotypes were constructed to identify shared chromosomal intervals segregating in the affected individuals.

There were many regions with LOD scores >2. The regions emerging with scores >2 repeatedly in different models were evaluated critically. There were two regions >1 Mb yielding LOD scores ≥2: on chromosomes 6 and 7 which were further evaluated by constructing haplotypes to assess whether the haplotype shared by the affected subjects was possibly identical by descent (IBD) from a common ancestor.

A region on chromosome 6 was evident at 127 cM and yielded a maximum LOD score of 2.63 (Fig. 5.17). Haplotypes were constructed at this locus in order to

observe any shared chromosomal interval among the affected subjects. However, we could not witness any common interval segregating in the affected subjects at this interval.

On chromosome 7, an interval between 40 and 50 cM showed significant maximal LOD scores in different rounds of analyses. Haplotype construction showed a common haplotype amongst all the affected individuals and in none of the normal subjects.

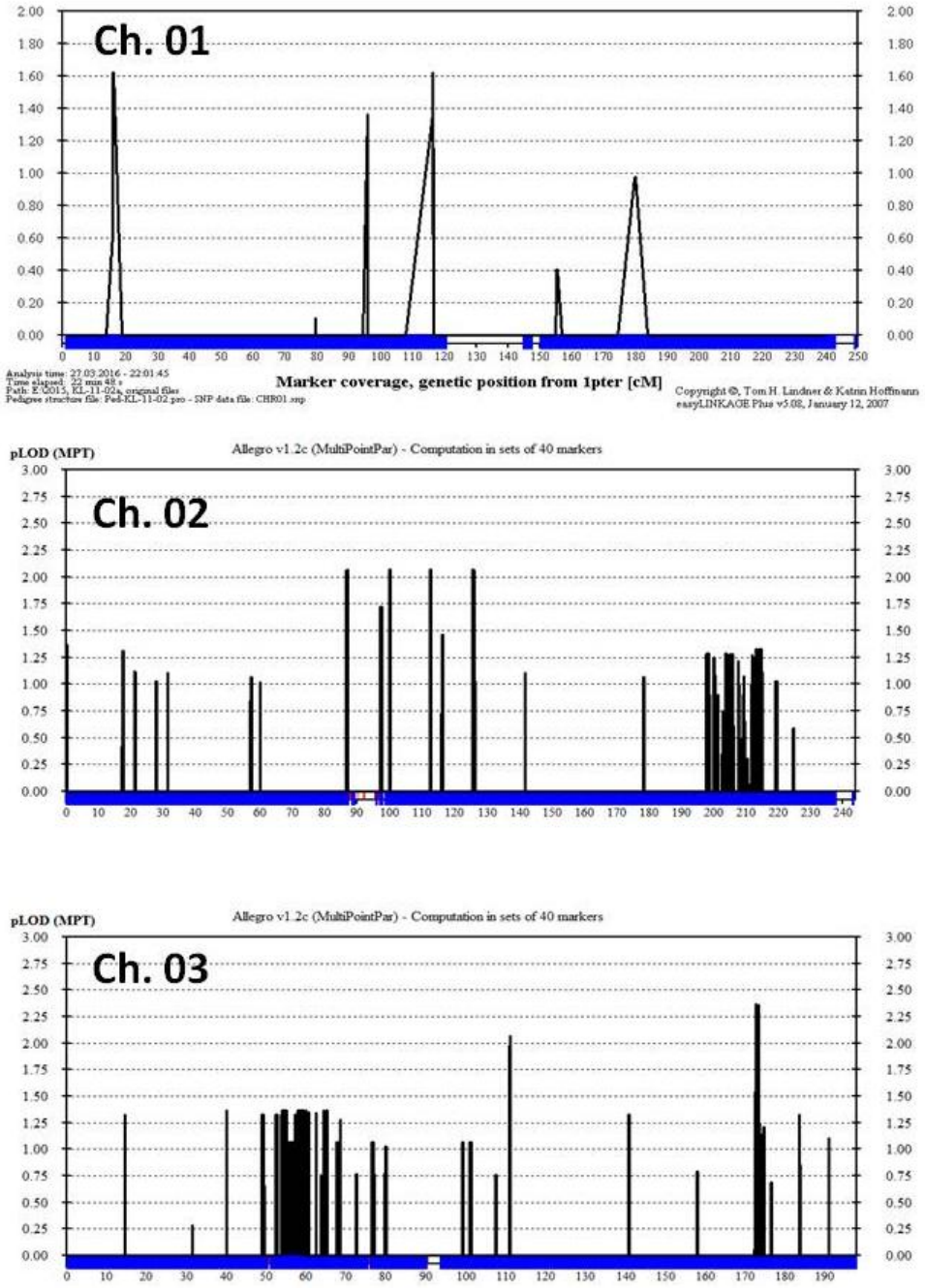


Figure 5.17. Multipoint LOD score calculations for autosomes and PAR regions in polydactyly Family.

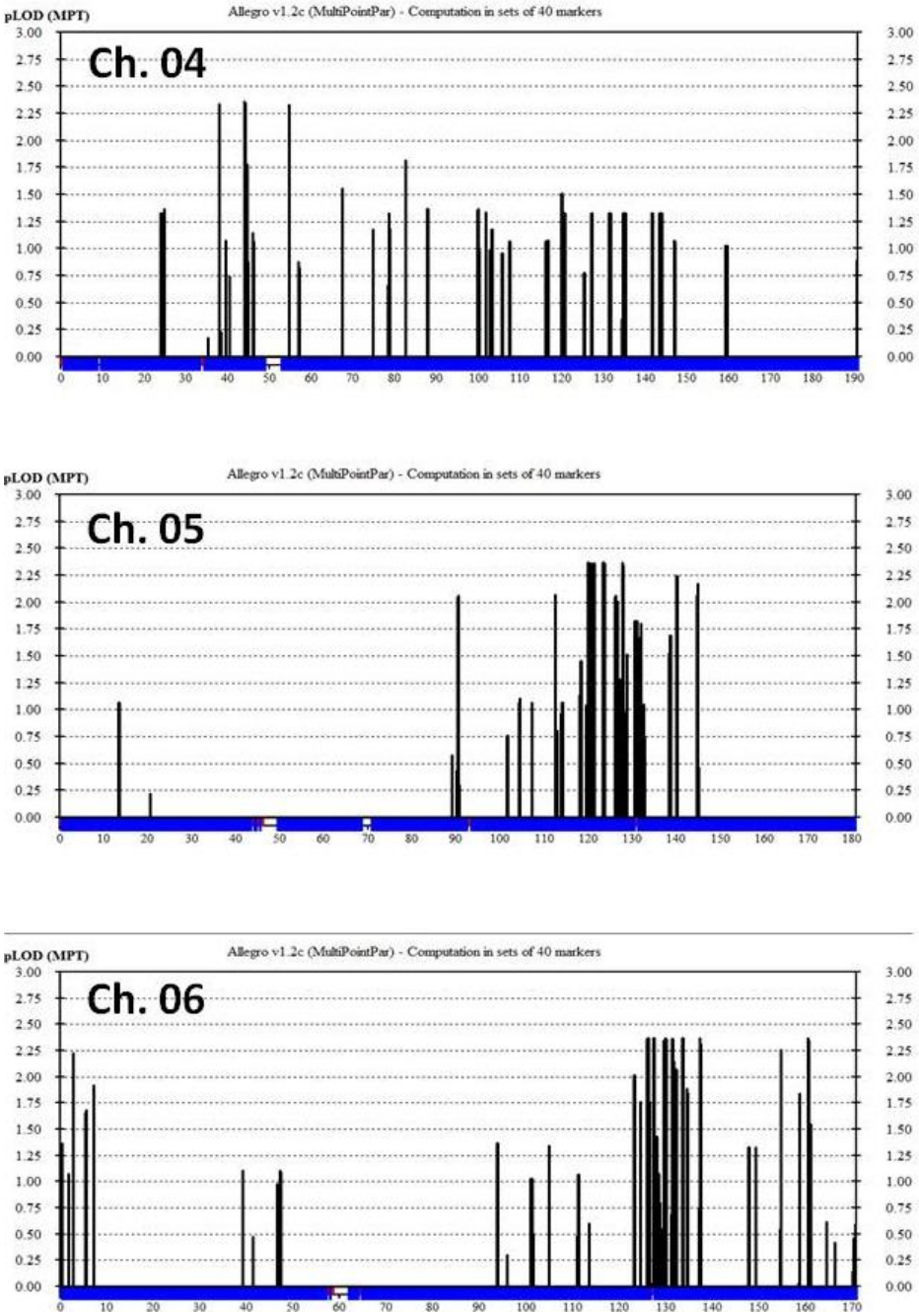


Fig. 5.17. Continued…….

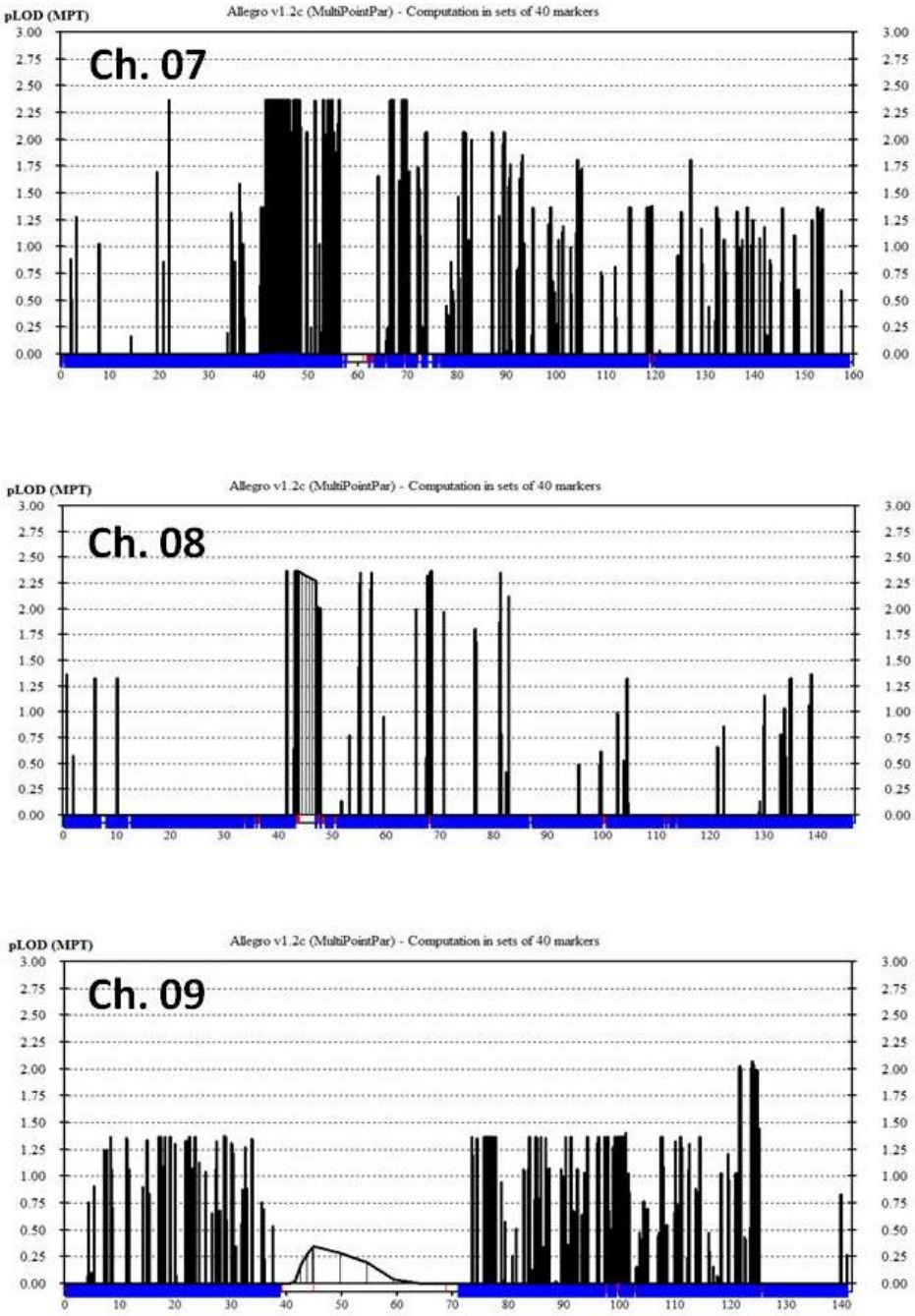


Fig. 5.17. Continued…….

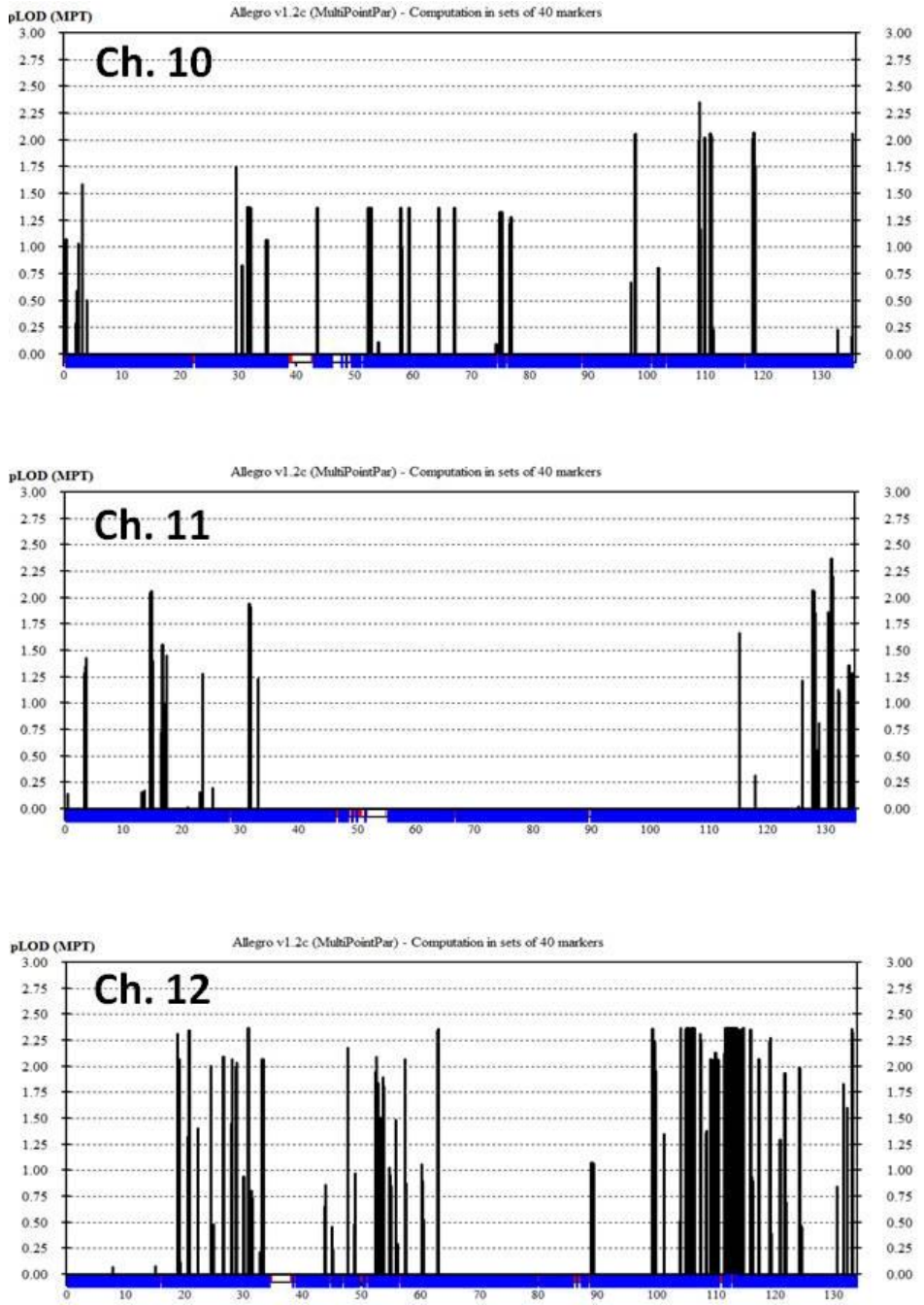


Fig. 5.17. Continued…….

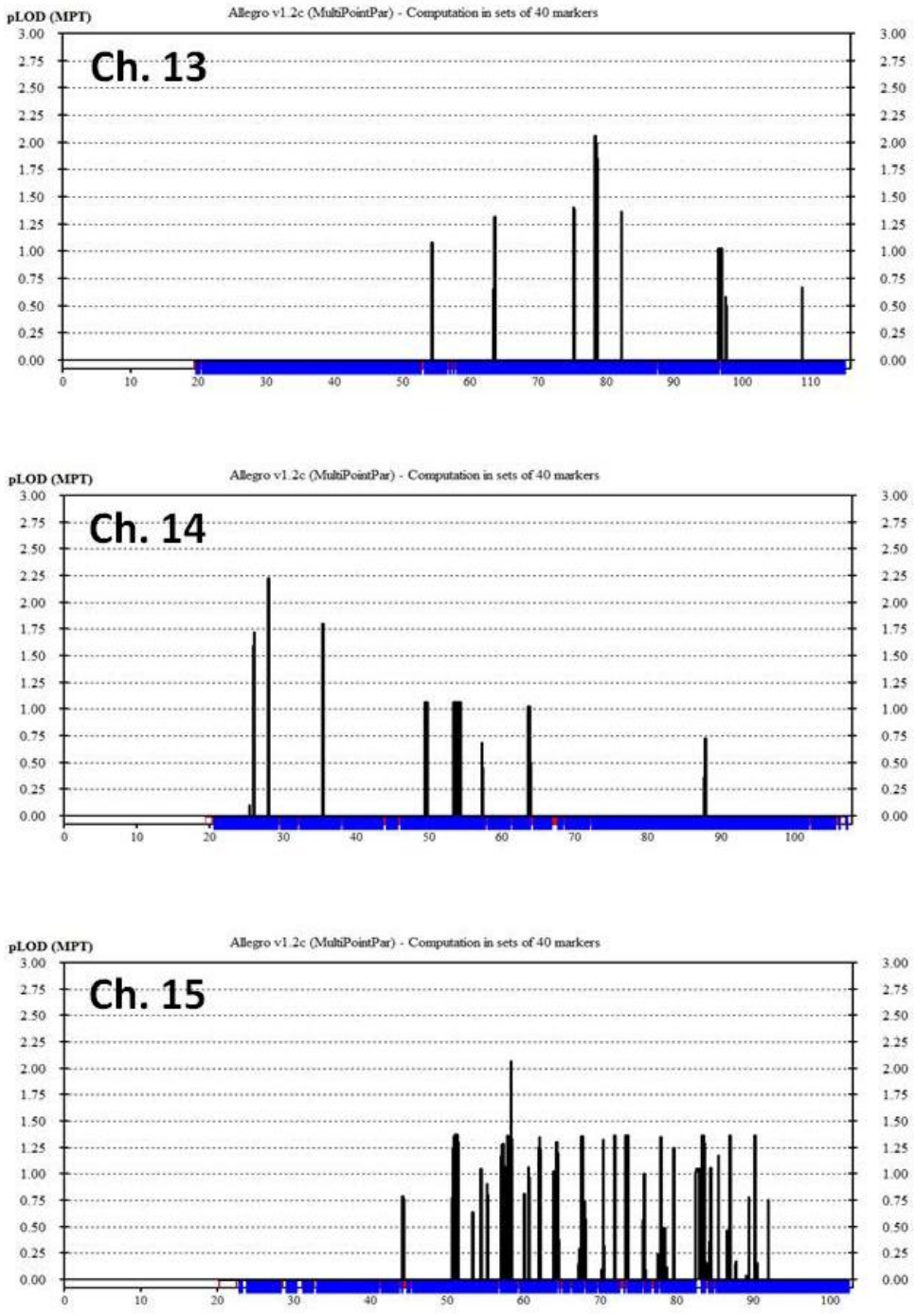


Fig. 5.17. Continued…….

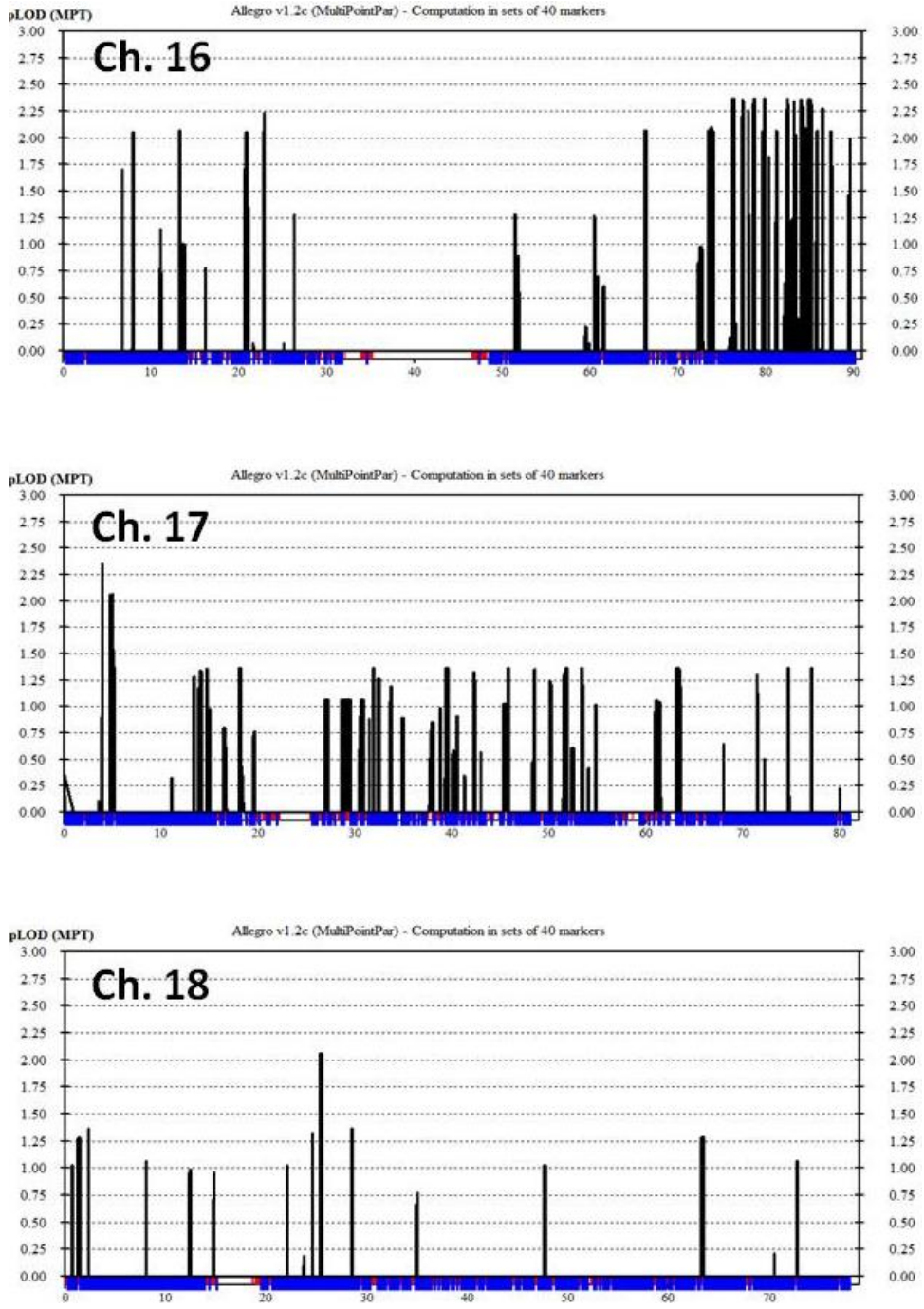


Fig. 5.17. Continued…….

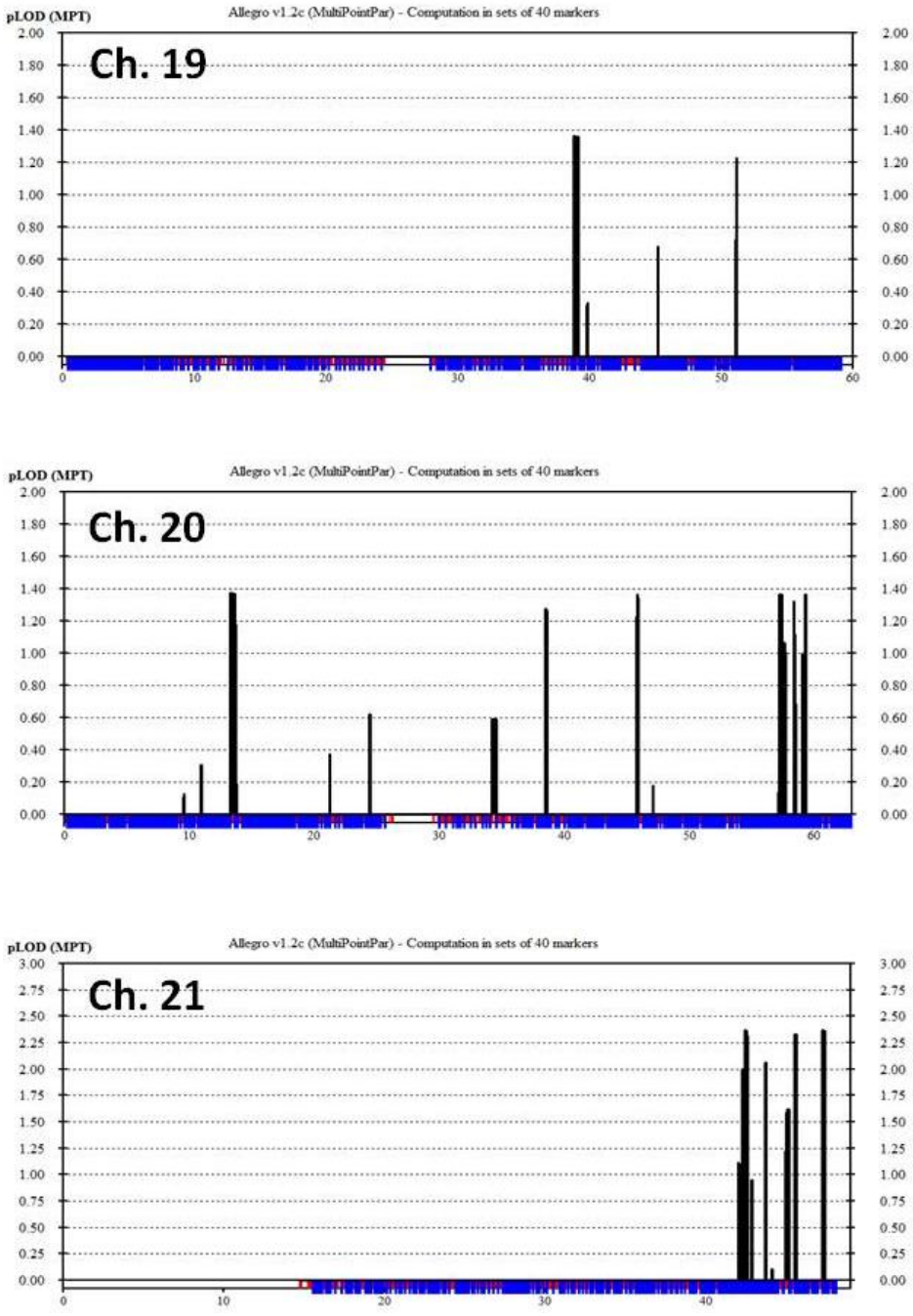


Fig. 5.17. Continued…….

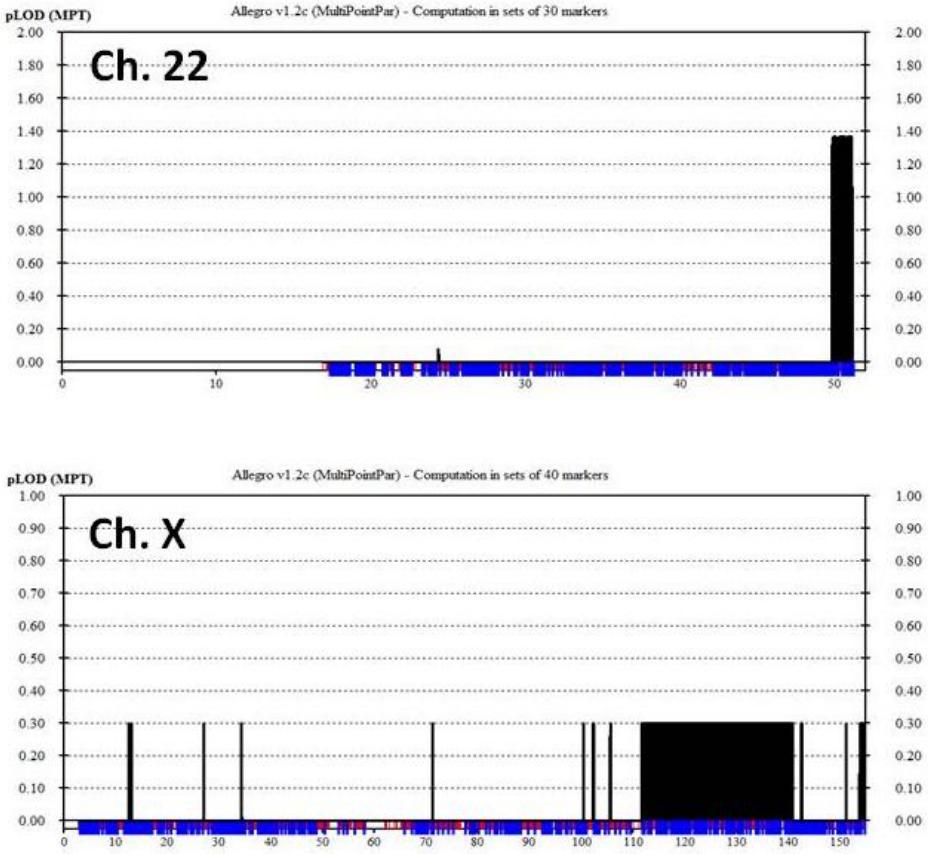


Fig. 5.17. Continued…….

5.4.3 CNV analysis

SNP genotyping data were subjected to deletion and duplication analysis by using cnvPartition (v3.2.0) CNV Analysis Plug-in for GenomeStudio. No deletion or duplication was detected within the candidate loci.

5.4.4 Exome sequencing and evaluation of the variants

Affected individual IV-1 was subjected to exome sequencing with Illumina TruSeq Exome Capture. Alignment of the reads, variant calling and annotation of the variants were performed with standard parameters in the lab.

The strongest candidate region obtained on chromosome 7p14.1 via linkage analysis and haplotype segregation was investigated for the regions of the candidate variants. All coding or splice variants were analyzed on the priority that was compatible with an autosomal dominant mode of inheritance and change seen should be pathological. The variations were filtered through the 1000 Genomes, ExAC, and EVS databases to exclude mutations already observed in the population. A unique frameshift mutation at position c.3635delG (p.Gly1212Alafs*18) in *GLI3* was observed in the region in the heterozygous state (Table 5.15). It was not found in 1000 Genomes, ExAC or EVS databases. It was validated by Sanger sequencing (Fig. 5.18). SSCP analysis showed the segregation of this variation in all the affected subjects and none of the unaffected individuals in the family. This variation was not

observed in 100 control individuals from the Pakistani population (Fig. 5.19). The online prediction tools showed that this candidate variant leads to a truncated protein (p.Gly1212Alafs*18) (Figs. 5.20, 5.21). The mutant protein was deduced to lack 342 of the 1570 native amino acids.

Another novel exonic variant, *GLI3* c.3630delT (location 42005041) was also found in the region in the heterozygous state. However, it was incompatible with autosomal dominant mode of inheritance of the disease (Table 5.15).

Table 5.15. Variants in the exome sequence variant list in the candidate region 7p14.1-p12.3 (nucleotides 40,085,765-49,983,829).

Location	Ref Base	Obs Base	Hom Het	SNP Quality	Total Depth	Region	Gene	SNP ID	1000G Fre
40085765	C	T	het	67	5	intronic	*CDK13*	rs9639817	0.166
40104400	CC	-	het	32	2	intronic	*CDK13*		
40127711	T	-	het	257	18	intronic	*CDK13*		
40220412	T	C	het	27	15	intronic	*C7orf10*	rs7778061	0.544
40220507	C	G	het	161	23	intronic	*C7orf10*	rs2036041	0.481
40228024	C	T	het	29	12	intronic	*C7orf10*		
40228033	-	AA	het	83	13	intronic	*C7orf10*		
40229072	G	A	het	45	70	intronic	*C7orf10*	rs75765039	0.442
40229114	AT	-	het	102	110	intronic	*C7orf10*		
40356551	CAG	-	het	90	8	intronic	*C7orf10*		
40416727	CC	-	het	81	3	intronic	*C7orf10*	rs61619065	
40416729	-	TT	het	29	3	intronic	*C7orf10*		
40416803	T	-	het	33	3	intronic	*C7orf10*		
40487846	CCT	-	het	32	2	intronic	*C7orf10*		
41739535	-	C	het	154	13	ncRNA	*LOC285954*		
41739539	CCT	-	het	178	14	ncRNA	*LOC285954*		
41739538	-	T	het	114	14	ncRNA	*LOC285954*		
41739541	-	TG	het	203	14	ncRNA	*LOC285954*		
41739543	-	T	het	54	14	ncRNA	*LOC285954*		
41739544	-	A	het	90	13	ncRNA	*LOC285954*		
41739545	-	A	het	170	13	ncRNA	*LOC285954*		
41739546	-	A	het	146	13	ncRNA	*LOC285954*		
41739553	-	T	het	23	13	ncRNA	*LOC285954*		
41739557	-	A	het	47	15	ncRNA	*LOC285954*		
41739560	A	-	het	67	17	ncRNA	*LOC285954*		
41790055	GT	-	het	32	2	ncRNA	*LOC285954*		
42005036	**C**	**-**	**het**	**337**	**20**	**exonic**	***GLI3***		
42005041	A	-	het	87	18	exonic	*GLI3*		
42153343	C	T	het	87	12	intronic	*GLI3*	rs6463090	0.152
42187740	CACACA	-	het	245	7	intronic	*GLI3*		
42187772	A	-	het	67	16	intronic	*GLI3*		
42187777	ACAG	-	het	215	17	intronic	*GLI3*		0.0664
42187777	AG	-	het	207	19	intronic	*GLI3*		
42187781	CT	-	het	175	21	intronic	*GLI3*		
42212210	C	T	het	62	22	intronic	*GLI3*	rs76999720	0.047
42951515	G	A	het	79	18	exonic;splicing	*C7orf25;C7orf25*	rs647117	0.713
43190479	T	G	het	87	32	intronic	*HECW1*	rs12702015	0.468
43287362	AG	-	het	29	3	intronic	*HECW1*		0.7131
43287397	AGG	-	het	26	4	intronic	*HECW1*		
43287610	AGAG	-	het	23	5	intronic	*HECW1*		0.0581
43356366	TT	-	het	105	3	intronic	*HECW1*		
43356401	CTC	-	het	29	3	intronic	*HECW1*		
43400652	-	TT	het	90	8	intronic	*HECW1*		
43470378	-	TA	het	32	2	intronic	*HECW1*		
43484310	T	C	het	133	53	exonic	*HECW1*	rs73098706	0.283
43495878	-	T	het	353	26	intronic	*HECW1*	rs3214631	0.431
43496138	T	C	het	44	17	intronic	*HECW1*	rs2304331	0.167
43505626	G	-	het	20	3	intronic	*HECW1*		
43505631	AGG	-	het	105	3	intronic	*HECW1*		
43506015	C	A	het	56	14	intronic	*HECW1*	rs2304330	0.425
43506158	T	C	het	127	9	exonic	*HECW1*	rs2304328	0.432
43519174	C	A	het	54	15	intronic	*HECW1*		
43527943	GA	-	het	172	6	intronic	*HECW1*		
43540513	A	G	het	28	9	intronic	*HECW1*	rs78224160	0.09
43540887	A	C	het	183	26	exonic	*HECW1*	rs17339479	0.107
43548516	C	T	het	36	26	intronic	*HECW1*	rs2304323	
43548832	T	C	het	31	11	intronic	*HECW1*	rs74368380	0.16

43622940	GCC	-	het	48	22	exonic	*STK17A*		
43731021	-	GG	het	29	3	intronic	*C7orf44*		
43731066	-	GGT	het	23	5	intronic	*C7orf44*		
43804380	-	CAC	het	32	2	intronic	*BLVRA*		
43827704	C	T	het	38	14	intronic	*BLVRA*		
43843205	-	C	het	98	22	intronic	*BLVRA*		
43908866	C	T	het	23	7	intronic	*MRPS24,URGCP-MRPS24*	rs2730608	0.39
43961766	C	T	het	38	4	intronic	*URGCP*		
43998387	TC	-	het	32	2	ncRNA	*POLR2J4*		
44003598	T	A	het	30	4	ncRNA	*POLR2J4*	rs7805884	0.286
44033938	CG	-	het	172	6	ncRNA	*POLR2J4*		
44033944	AA	-	het	172	6	ncRNA	*POLR2J4*		
44033967	GGAAGG	-	het	20	6	ncRNA	*POLR2J4*		
44033971	GG	-	het	96	6	ncRNA	*POLR2J4*		
44033985	AA	-	het	99	5	ncRNA	*POLR2J4*		
44034007	AA	-	het	99	5	ncRNA	*POLR2J4*		
44034017	AG	-	het	99	5	ncRNA	*POLR2J4*		
44034050	AGA	-	het	32	2	ncRNA	*POLR2J4*		
44035069	C	T	het	35	3	ncRNA	*POLR2J4*		
44056177	CT	-	het	101	43	ncRNA	*POLR2J4*	rs66631939	
44080507	C	T	het	33	25	ncRNA	*FLJ35390*		0.032
44080531	C	T	het	43	29	ncRNA	*FLJ35390*	rs76964317	0.054
44080617	-	TCC	het	80	26	ncRNA	*FLJ35390*		
44080619	-	CTT	het	142	26	ncRNA	*FLJ35390*		
44115310	-	C	het	49	6	intronic	*POLM*		
44115316	-	TTTTT	het	23	5	intronic	*POLM*		
44115317	-	TT	het	99	5	intronic	*POLM*		
44143800	CTG	-	het	32	2	upstream	*AEBP1*		
44143811	CTA	-	het	29	3	upstream	*AEBP1*		
44143826	CCG	-	het	23	5	upstream	*AEBP1*		
44148553	A	G	het	100	23	exonic	*AEBP1*	rs2595701	0.768
44153614	C	G	het	88	59	exonic	*AEBP1*	rs61736256	0.028
44153780	A	G	het	95	72	exonic	*AEBP1*	rs13928	0.355
44158278	-	TT	het	102	4	intronic	*POLD2*		
44163284	GG	-	het	20	6	upstream	*POLD2*		
44178829	A	G	het	20	6	intronic	*MYL7*	rs882019	0.516
44185088	G	A	het	123	62	intronic	*GCK*	rs2908274	0.356
44185741	-	C	het	27	196	intronic	*GCK*		
44185744	-	C	het	123	199	intronic	*GCK*		
44185747	-	C	het	463	200	intronic	*GCK*		
44185874	-	C	het	56	83	intronic	*GCK*		
44185879	-	CC	het	95	83	intronic	*GCK*		
44185883	-	C	het	43	86	intronic	*GCK*		
44185885	-	C	het	106	86	intronic	*GCK*		
44185891	-	C	het	45	84	intronic	*GCK*		
44185898	-	C	het	26	87	intronic	*GCK*		
44185900	-	C	het	30	85	intronic	*GCK*		
44185901	-	G	het	121	85	intronic	*GCK*		
44185909	-	A	het	65	88	intronic	*GCK*		
44185955	A	G	het	228	69	intronic	*GCK*	rs76323047	0.08
44189321	A	G	het	87	67	intronic	*GCK*	rs2268574	0.607
44190468	T	G	het	53	88	intronic	*GCK*	rs2268573	0.63
44191190	T	A	het	48	3	intronic	*GCK*	rs6971410	0.209
44245853	A	T	het	70	7	intronic	*YKT6*	rs917793	0.223
44282369	T	C	het	37	9	intronic	*CAMK2B*	rs35452037	0.252
44298163	-	TAT	het	23	5	intronic	*CAMK2B*		
44365291	-	G	het	24	4	upstream	*CAMK2B*		
44558269	A	-	het	32	21	intronic	*NPC1L1*		
44579180	G	C	het	110	45	exonic	*NPC1L1*	rs2072183	0.243
44605733	CCAC	-	het	121	10	intronic	*DDX56*		
44605731	-	AG	het	121	10	intronic	*DDX56*		
44605732	-	GGAGC	het	79	10	intronic	*DDX56*		
44605740	CA	-	het	160	10	intronic	*DDX56*		
44606098	C	T	het	106	11	exonic	*DDX56*	rs6656	0.769
44608718	C	T	het	73	14	intronic	*DDX56*	rs2289051	0.161

44612347	G	C	het	55	17	intronic	*DDX56*	rs217374	0.792
44619302	T	C	het	48	17	intronic	*TMED4*	rs217361	0.787
44620267	A	T	het	24	6	intronic	*TMED4*		
44620277	A	G	het	31	6	intronic	*TMED4*		
44620284	A	G	het	46	6	intronic	*TMED4*		
44620290	A	C	het	27	6	intronic	*TMED4*		
44620312	G	A	het	23	7	intronic	*TMED4*		
44620324	T	C	het	21	8	intronic	*TMED4*		
44620395	-	C	het	98	8	intronic	*TMED4*		
44620398	T	C	het	21	7	intronic	*TMED4*		
44620408	T	C	het	39	9	intronic	*TMED4*		
44620836	C	A	het	34	18	exonic	*TMED4*	rs8580	0.804
44664458	AAG	-	het	242	8	intronic	*OGDH*		0.0203
44713691	G	A	het	37	5	intronic	*OGDH*	rs740040	0.21
44714695	G	-	het	67	5	intronic	*OGDH*		0.4966
44737922	G	A	het	43	16	intronic	*OGDH*	rs3213662	0.301
44795859	T	C	het	51	19	exonic	*ZMIZ2*		
44805118	-	C	het	30	19	exonic	*ZMIZ2*		
44836314	A	G	het	100	12	UTR5	*PPIA*	rs6850	0.284
44839616	C	T	het	180	23	intronic	*PPIA*	rs6463247	0.49
44840864	G	A	het	88	20	intronic	*PPIA*	rs10249442	0.49
44841247	AG	-	het	140	5	UTR3	*PPIA*		0.2318
44880685	A	-	het	172	10	intronic	*H2AFV*		0.0328
44887623	C	-	het	21	7	UTR5	*H2AFV*		
44924627	A	C	het	45	14	exonic	*PURB*	rs6966024	0.65
45001343	-	T	het	72	12	downstream	*MYO1G*		0.2492
45001359	-	C	het	148	14	downstream	*MYO1G*		
45001364	AG	-	het	135	13	downstream	*MYO1G*		
45001375	G	A	het	25	12	downstream	*MYO1G*		0.398
45001542	-	GCTGT	het	80	10	downstream	*MYO1G*		
45001546	GA	-	het	78	12	downstream	*MYO1G*		
45001571	A	-	het	43	16	downstream	*MYO1G*		
45001576	-	AG	het	142	16	downstream	*MYO1G*		
45001584	G	-	het	76	15	downstream	*MYO1G*		
45039279	-	CCCGC	het	26	4	upstream	*CCM2*		
45039280	-	GG GCTCCC	het	26	4	upstream	*CCM2*		
45039762	-	C	het	175	5	ncRNA	*CCM2*		
45040063	-	T	het	166	4	ncRNA	*CCM2*		
45040066	GCA	-	het	166	4	ncRNA	*CCM2*		
45040065	-	GG	het	128	4	ncRNA	*CCM2*		
45040077	CAGCT	-	het	26	4	ncRNA	*CCM2*		
45040076	C	-	het	25	4	ncRNA	*CCM2*		
45040078	-	T	het	96	5	ncRNA	*CCM2*		
45040088	-	T	het	68	7	ncRNA	*CCM2*		
45040090	-	G	het	67	7	ncRNA	*CCM2*		
45040094	-	C	het	57	7	ncRNA	*CCM2*		
45628474	TGTGTG	-	het	105	3	intronic	*ADCY1*		
45650418	-	GGC	het	29	3	intronic	*ADCY1*		
45650577	-	GA	het	29	3	intronic	*ADCY1*		
45724696	T	C	het	29	9	intronic	*ADCY1*	rs2461106	0.15
45743128	T	G	het	213	31	intronic	*ADCY1*	rs2247685	0.41
45954681	AC	-	het	102	4	intronic	*IGFBP3*		0.0444
45957173	A	-	het	33	3	intronic	*IGFBP3*	rs56296462	0.5714
47317510	G	A	het	139	62	UTR3	*TNS3*	rs3750163	0.289
47319723	C	T	het	199	21	intronic	*TNS3*	rs4720590	0.708
47319852	C	T	het	29	21	intronic	*TNS3*	rs6971754	0.204
47323154	A	G	het	90	13	intronic	*TNS3*	rs3829000	0.706
47323456	A	G	het	228	137	exonic	*TNS3*	rs3750164	0.69
47342710	G	A	het	149	300	exonic	*TNS3*		
47407888	G	A	het	114	31	intronic	*TNS3*	rs2255749	0.428
47408029	A	G	het	228	65	exonic	*TNS3*	rs2255744	0.637
47408974	T	A	het	164	47	exonic	*TNS3*	rs1534136	0.645
47428599	G	-	het	61	7	intronic	*TNS3*		
47428633	AG	-	het	20	6	intronic	*TNS3*		
47428651	AG	-	het	23	5	intronic	*TNS3*		

47428775	-	AGC	het	102	4	intronic	*TNS3*	rs71003396	
47453509	CAG	-	het	27	29	intronic	*TNS3*		0.0091
47467864	C	T	het	82	98	intronic	*TNS3*	rs1799370	0.375
47474713	-	TT	het	589	19	intronic	*TNS3*		0.3522
47479363	A	G	het	163	42	intronic	*TNS3*	rs1810137	0.233
47513110	A	T	het	41	3	intronic	*TNS3*	rs6963440	0.232
47520850	T	C	het	65	113	intronic	*TNS3*	rs1543035	0.188
47568718	C	T	het	194	30	UTR5	*TNS3*	rs10261525	0.256
47577564	CA	-	het	29	3	intronic	*TNS3*		
47853397	G	A	het	25	14	intronic	*C7orf69,PKD1L1*	rs4595031	0.18
47859241	-	A	het	52	10	UTR3	*C7orf69*		
47894326	A	G	het	149	20	intronic	*PKD1L1*	rs13228166	0.513
47894952	-	T	het	58	8	intronic	*PKD1L1*	rs11441954	0.0336
47898222	-	G	het	404	17	intronic	*PKD1L1*		0.0486
47898223	-	T	het	273	17	intronic	*PKD1L1*		
47898561	C	T	het	48	19	intronic	*PKD1L1*	rs60527187	0.502
47898597	T	C	het	95	14	intronic	*PKD1L1*	rs2053985	0.123
47917087	C	T	het	51	21	exonic;splicing	*PKD1L1;PKD1L1*	rs72601626	0.042
47971626	G	A	het	77	18	exonic	*PKD1L1*	rs885338	0.139
47979716	A	C	het	61	6	intronic	*PKD1L1*	rs62447089	0.14
48008666	A	G	het	64	12	ncRNA	*HUS1*	rs3176582	0.852
48008738	A	G	het	123	15	ncRNA	*HUS1*	rs3176581	0.139
48016442	G	A	het	22	14	ncRNA	*HUS1*	rs2307252	0.176
48035446	C	T	het	44	7	intronic	*SUN3*	rs13230023	0.291
48046714	G	A	het	124	27	intronic	*SUN3*	rs1574242	0.436
48058152	T	C	het	120	18	intronic	*SUN3*	rs1915963	0.38
48058220	A	-	het	149	28	intronic	*SUN3*		
48068258	G	C	het	85	5	intronic	*SUN3*	rs7357251	0.35
48081095	G	T	het	200	32	exonic	*C7orf57*	rs10951942	0.348
48083339	A	T	het	43	8	intronic	*C7orf57*	rs6978838	0.456
48100016	T	C	het	20	5	UTR3	*C7orf57*	rs7790252	0.29
48237721	G	C	het	24	12	intronic	*ABCA13*	rs28637820	0.21
48237802	C	T	het	92	22	intronic	*ABCA13*	rs2188	0.347
48237824	C	T	het	49	33	intronic	*ABCA13*	rs17060	0.207
48242694	C	T	het	25	4	intronic	*ABCA13*		0.119
48242884	C	G	het	24	5	intronic	*ABCA13*	rs10240647	0.167
48285485	C	T	het	72	31	exonic	*ABCA13*	rs1880738	0.346
48287803	TT	-	het	796	26	intronic	*ABCA13*	rs4024052	0.0474
48313757	G	A	het	62	28	exonic	*ABCA13*	rs77190804	0.116
48313881	T	C	het	73	45	exonic	*ABCA13*	rs17712299	0.113
48326123	CAGA	-	het	36	4	intronic	*ABCA13*		
48397830	GGGA	-	het	32	2	intronic	*ABCA13*		
48412112	T	-	het	27	11	intronic	*ABCA13*		
48443634	G	A	het	90	7	intronic	*ABCA13*		
48495099	C	A	het	51	5	intronic	*ABCA13*	rs13243674	0.16
48506416	G	C	het	139	6	intronic	*ABCA13*	rs4917151	0.662
48506416	G	A	het	139	6	intronic	*ABCA13*		
48626500	T	G	het	31	8	intronic	*ABCA13*	rs7789709	0.699
48626645	C	G	het	61	52	intronic	*ABCA13*	rs7778411	0.702
48634113	A	G	het	27	3	intronic	*ABCA13*	rs75933127	0.067
48654761	G	A	het	80	10	intronic	*ABCA13*	rs10243913	0.415
48964322	G	T	het	44	15	ncRNA	*CDC14C*	rs11767646	0.507
48965000	G	A	het	68	36	ncRNA	*CDC14C*	rs10259692	0.393
49815390	C	G	het	50	8	exonic	*VWC2*	rs769604	0.434
49815474	C	T	het	52	12	exonic	*VWC2*	rs73347631	0.037
49815493	C	G	het	70	16	exonic	*VWC2*	rs61390655	0.041
49842561	C	T	het	95	12	intronic	*VWC2*	rs73108844	0.087
49903022	-	T	het	144	6	intronic	*VWC2*		
49983829	G	A	het	41	3	intronic	*ZPBP*		

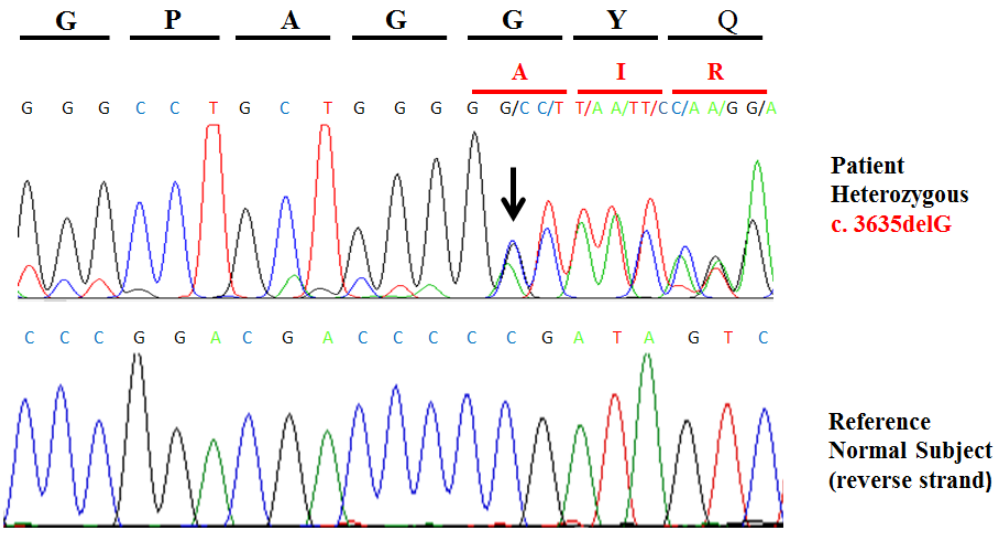
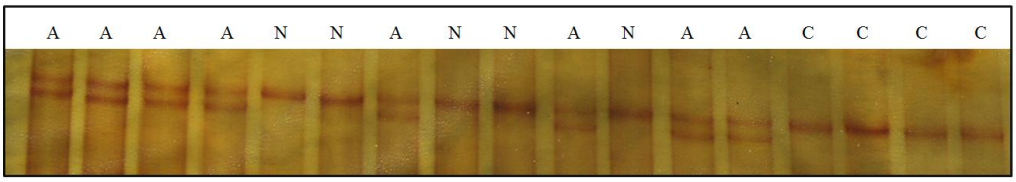


Figure 5.18. Chromatograms showing mutation *GLI3* c.3635delG in an affected heterozygous subject and the reference sequence of the unaffected subject.

.....

Figure 5.19. **SSCP results of polydactyly Family and population screen of *GLI3* c.3635delG** in 8 Affected (A), 5 normal (N) individuals and 4 control (C).



representatives from the population.

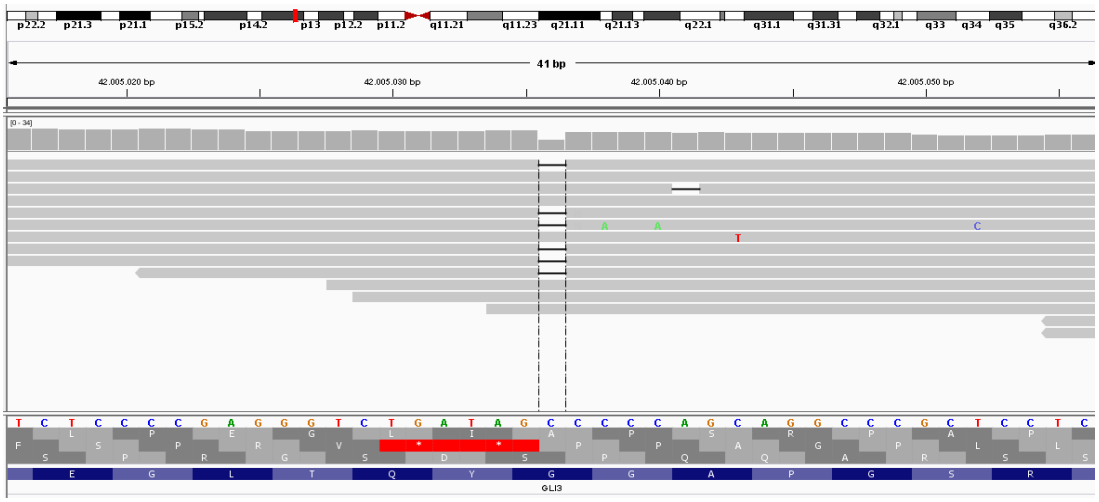


Figure 5.20. The sequence reads around the mutation as visualized with IGV (Integrative Genome Viewer).

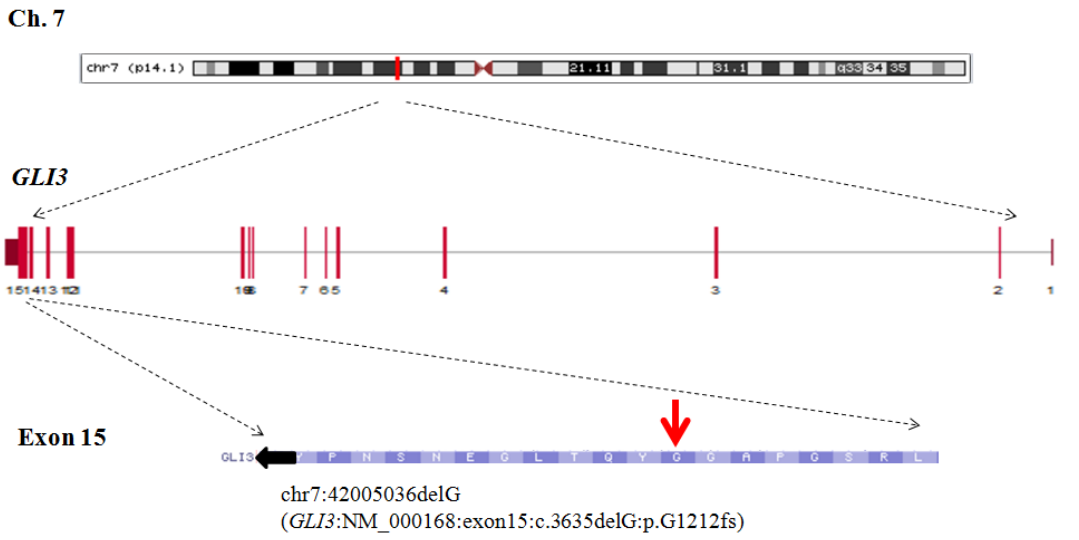


Figure 5.21. Genomic organization and location of mutation of *GLI3* at chr7p14.1.

The results obtained in polydactyly famiy have been compiled and the manuscript is submitted in Eur J Med Genet:

Mumtaz S, Yildiz E, Lal K, Tolun A, Malik S. Complex Postaxial Polydactyly Type A/B with Camptodactyly, Hypoplastic Third Toe, Zygodactyly and Other Digit Anomalies Caused by a Novel GLI3 Mutation. Eur J Med Genet. 2017; 60(5):268-274. doi: 10.1016/j.ejmg.2017.03.004

5.5 Family E: Tooth Agenesis Manifesting as Oligodontia/Hypodontia (TAOH)

5.5.1 Clinical description

Ten affected and 11 unaffected subjects were physically examined. In this family, the affected subjects had several missing teeth in maxilla and mandible. Generally, lateral incisors and first premolars were absent in both upper and lower jaws. The existing teeth were malformed, disfigured, misaligned and depicted remarkable diastema. For instance, the central incisors were peg shaped; labial frenum was extraordinarily grown, and there was prominent median diastema. Diastema was also significant in lateral incisors and canines. Other rare finding was rampant carries which was present in several teeth in one subject (IV-7), while one other subject (IV-6) had carries in lower first molars (clinical manifestations are presented in Table 5.16).

The existing teeth had normal enamel, colour, and no signs of hypominearalization. Additionally, there was no history of pain, sensitivity or erosion. The eruption time was normal for both deciduous and permanent teeth. There was no other facio-cephalic abnormality and the occlusion appeared normal. Temporo-mandibular joint (TMJ) was normal. The affected subjects had no difficulty in eating and chewing, and they were good at all sorts of food.

Table 5.16. Features of the members of TAOH Family with tooth agenesis.

Features/variable	IV-7	IV-4	IV-5	IV-3	III-2	II-2	III-3	III-1	IV-6	IV-2	IV-1
Affection status	A	A	A	A	A	A	A	A	A	A	A
Age (year) / gender	6	12	13	16	35	56	35	45	7	18	20
Nails: texture, shape, color	N	N	N	N	N	N	N	N	N	N	N
Skin: texture, elasticity	N	N	N	N	N	N	N	N	N	N	N
Excessive sweating on palms	+	-	-	-	-	-	-	-	-	-	-
Palmoplentor dermatoglypics	N	N	N	N	N	N	N	N	N	N	N
Thinning of scalp hair	-	-	-	-	+	+	-	-	-	-	-
Dentition	N	Ab	Ab	Ab	N	N	N	N	N	N	N
Dental carries	+	-	-	-	-	-	-	-	+	-	-
Associated anomaly	-	-	-	-	Back pain	-	Back pain	-	-	-	Back pain

A= Affected; N = Normal; + = feature present; - = feature absent; Ab = Abnormal;

5.5.2 Clinical features of selected affected subjects

Affected subject III-2. Upper laterals were missing. Lower dentition was normal in number and alignment. In the Ceph (cephalometric) radiographs, there was normal occlusion and normal bite in the anterior and posterior teeth.

Affected subject IV-3. Lateral incisors were absent in the upper jaw. In the lower jaw, there were deciduous laterals and the canines were missing. There was remarkable median diastema in the lower jaws (Fig. 5.22 A, B). Labial franum was well-developed particularly in the upper jaw. Lower central incisors were peg shaped. In the OPG, the lateral canine was evident in the right upper jaw which was impacted and failed to erupt/exfoliate. Both upper and lower incisors were gross and hypertrophic. The incisors appeared distinct in the occlusion line and might have over-grown due to the absence of opposing tooth in the occlusion.

Affected subject IV-4

There was mixed dentition. Laterals were missing in upper and lower jaws. Additionally, upper premolars were missing. OPG and the Ceph radiographs revealed impacted upper permanent canines which failed to erupt. In the lower jaw, the right canine was impacted but the left erupted. Additionally, lower left premolar also appeared impacted. Occlusion was normal and the posterior teeth showed well occluded. However, anterior teeth depicted overbite. Facial symmetry appeared unremarkable.

Affected subject IV-5

There was mixed dentition with four missing teeth in upper and lower jaws. OPG revealed that lateral incisors and deciduous and permanent canines were missing in the upper jaw (Fig. 5.22 C, D). Frenum was remarkably grown. Likewise, lateral incisors and permanent canines were missing in the lower jaw. The remaining incisors were peg shaped. In the Ceph radiograph, occlusion and facial symmetry appeared normal. However, upper incisors were protruding and there was a sign of overbite in the anterior teeth. The detail of dentition and missing teeth of the selected affected subjects is presented in Table 5.17.

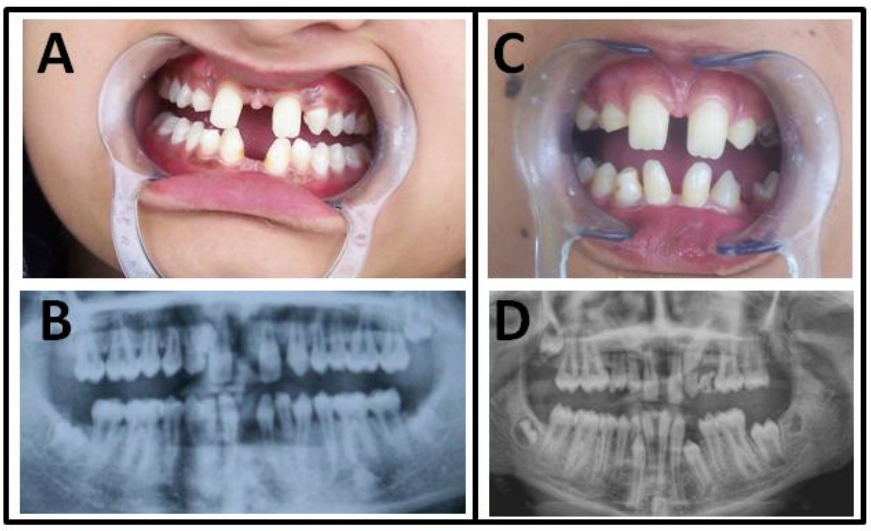


Figure 5.22. **Photograph and OPG of affected subjects IV-3 (A-B) and IV-5 (C-D).** A-B. Peg shaped incisors and a remarkable diastema. Lateral incisors and canines are missing. Canine in the right upper jaw is impacted. C-D. Peg shaped incisors and prominent diastema. Mixed dentition and missing of lateral incisors and canines in upper jaw, and missing lateral incisors and canines in the lower jaw.

Table 5.17. Clinical features in the subjects of TAOH Family.

Subject ID	Age (yr)	Status	Dentition	Present dentition	Missing teeth
III-02	35	Affected	Permanent	7 6 5 4 | 1 3 4 5 6 7 7 6 5 4 3 2 1 | 1 2 3 4 5 6 7	1 | 2
IV-3	16	Affected	Mixed	7 6 5 4 3 1 | 1 3 4 5 6 7 7 6 5 4 b 1 | 1 b 4 5 6 7	2 | 2 3 | 3
IV-4	13	Affected	Mixed	7 6 e c 1 | 1 c e 6 7 7 6 5 4 c 1 | 1 3 5 6 7	5 4 2 | 2 4 5 2 | 2
IV-5	12	Affected	Mixed	7 6 e d 1 | 1 d e 6 7 7 6 e d c 1 | 1 c d e 6 7	3 2 | 2 3 3 2 | 2 3
IV-7	6	Normal	Deciduous	e d c b a | a b c d e e d c b a | a b c d e	

The literature survey and the review of reported cases in the OMIM database showed that the clinical symptoms in this family were consistent with selective tooth agenesis (OMIM 106600). The phenotypic presentation in the family was suggestive of the presence of two subtypes of tooth agenesis: a) hypodontia with agenesis of fewer than 6 teeth; and b) oligodontia with agenesis of six or more permanent teeth.

5.5.3 Targeted sequencing of *WNT10A*

Since STHAG4 locus accounts for most of the dominantly segregating tooth agenesis types, we initially selected *WNT10A* (OMIM 606268) for mutation analysis through direct Sanger sequencing. However, sequence analyses of *WNT10A* did not yield any pathogenic variant segregating with the phenotype in this family. We therefore, launched SNP genotyping.

5.5.4 Multipoint linkage analysis and haplotype examination

The SNP genome scan data of five affected and four unaffected family members were subjected to linkage analysis. Multipoint LOD scores were calculated under an autosomal dominant disease model with 90% penetrance. Due to the computational constraints of the linkage software, the pedigree was analyzed in two parts. The analyses were iterated to reach a plausible model. Large intervals with a LOD score of >1.2 were further scrutinized under various penetrance estimates. These analyses yielded several intervals co-segregating with

the phenotype. Among those, the following regions were important: chromosome 2p16.1-q14.2 (60-120 Mb; LOD=1.25), 6p22-p21.1 (25-45 Mb; LOD=2.0), 10q26.11-26.3 (120-140 Mb; LOD=2.0) and 12q21.1-21.31 (75-85 Mb; LOD=1.2), and 15q12-13.3 (27-33 Mb; LOD=1.25). Other regions also prominent in these analyses were on chromosome 5q11.2 (52-58 Mb; LOD=1.25), 11q12.1-13.3 (58-70 Mb; LOD=1.3), 15q12-13.3 (27-33 Mb; LOD=1.25), and 17q23.1-24.1 (58-63 Mb; LOD 1.25). Multipoint LOD score graphics generated with a pedigree structure comprising four affected and two unaffected subjects are given in Fig. 5.23.

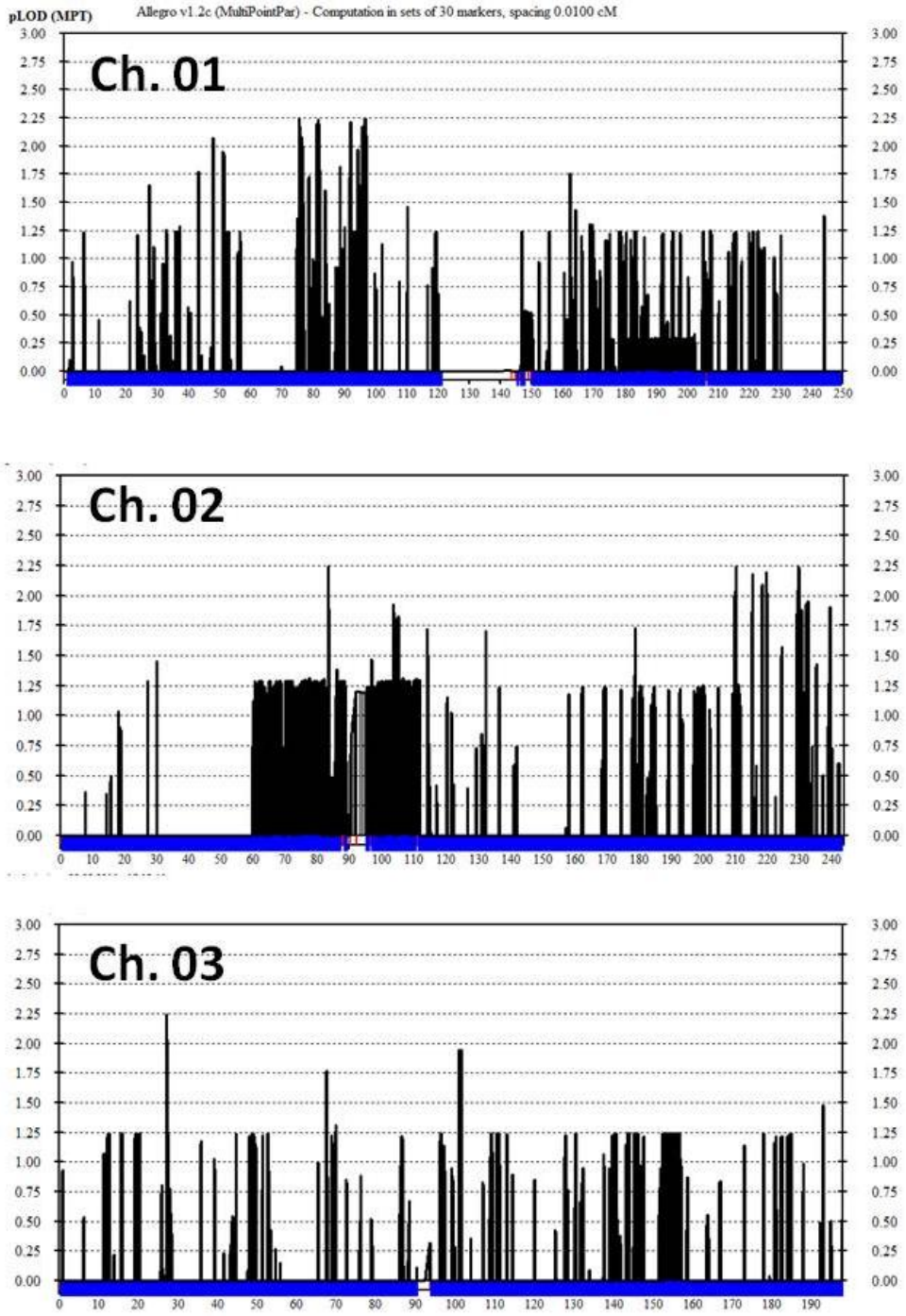


Figure 5.23. Multipoint LOD score calculations for autosomes and X-chromosome in Oligodontia/Hypodontia Family.

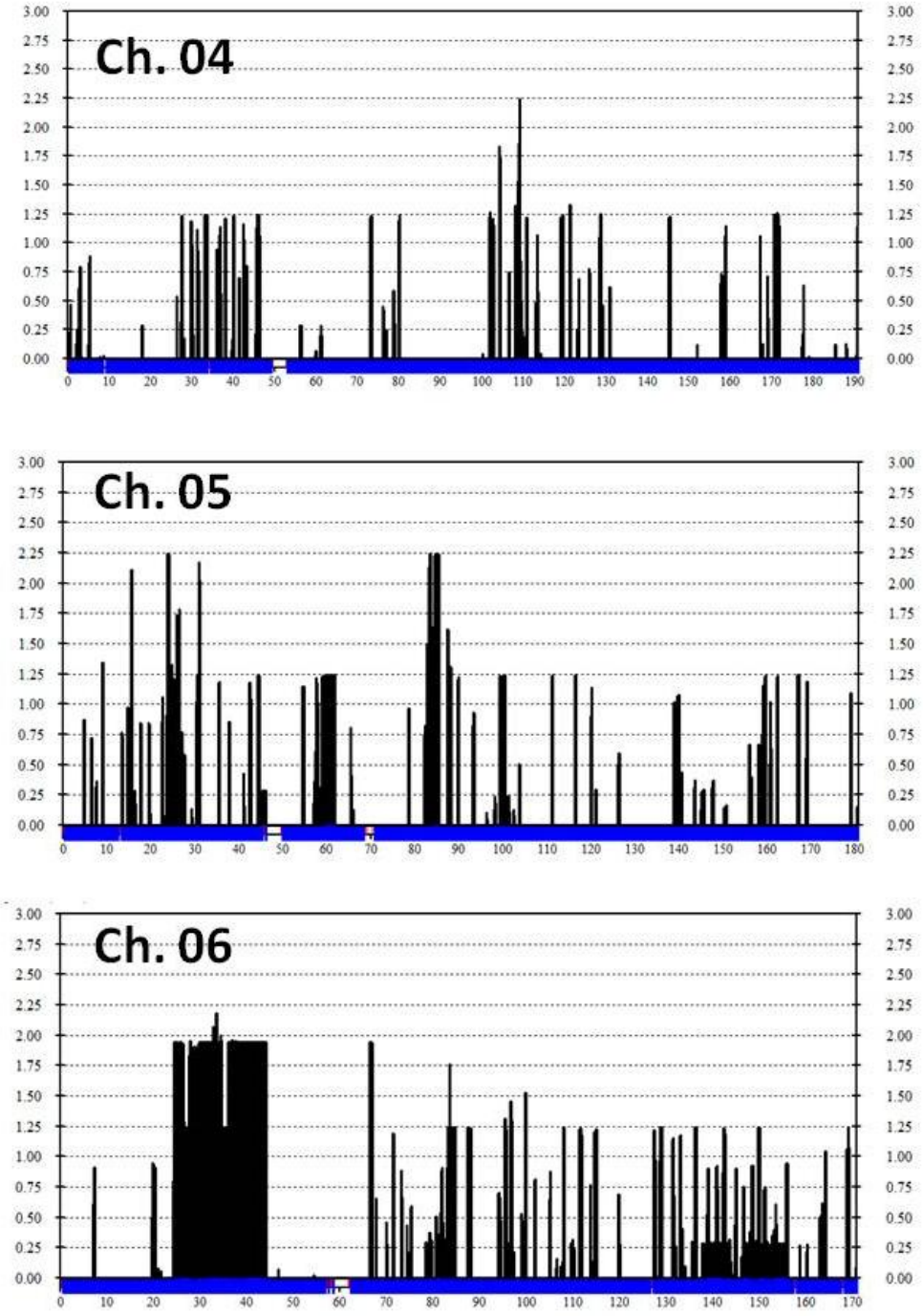


Fig. 5.23. Continued…….

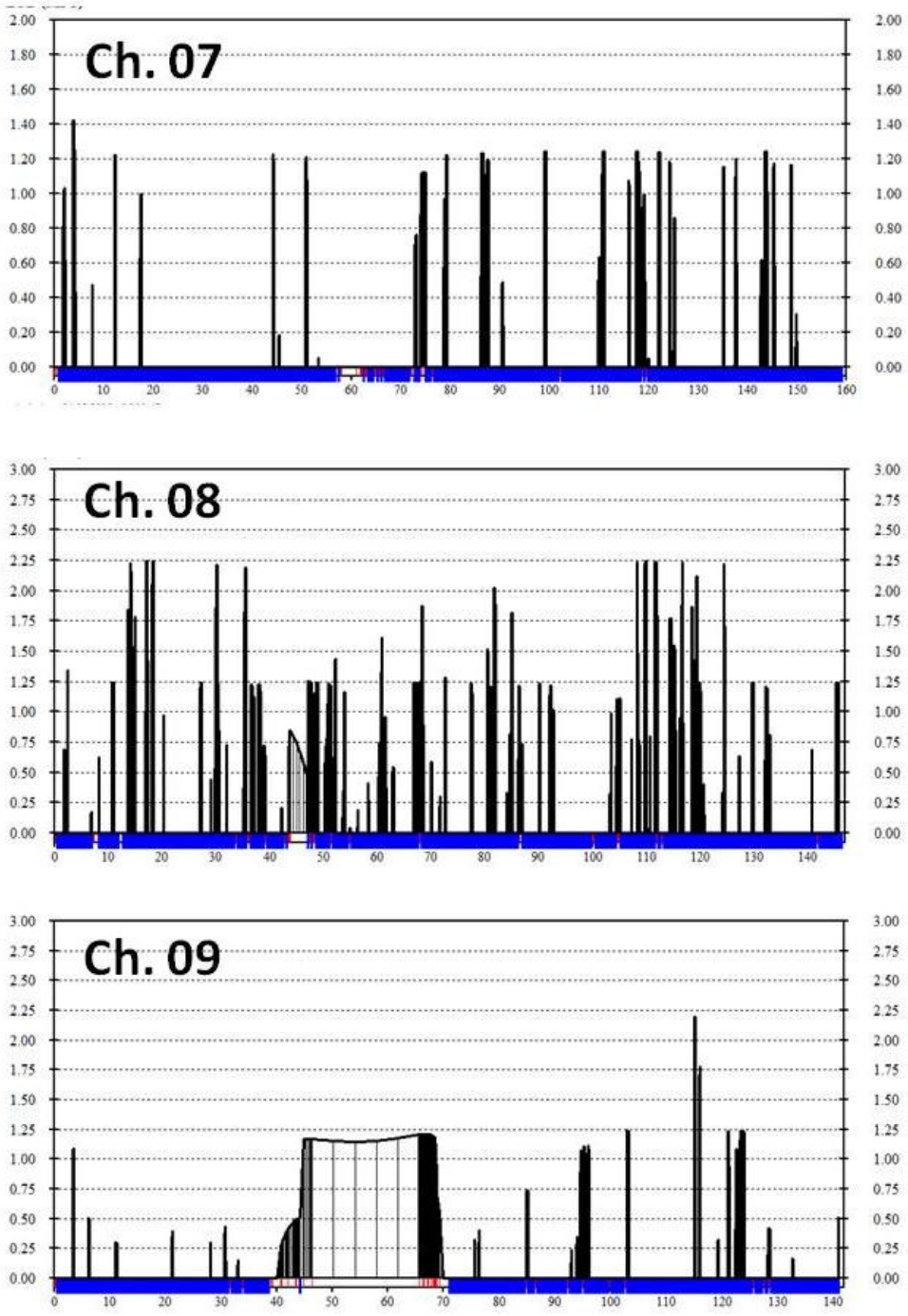


Fig. 5.23. Continued…….

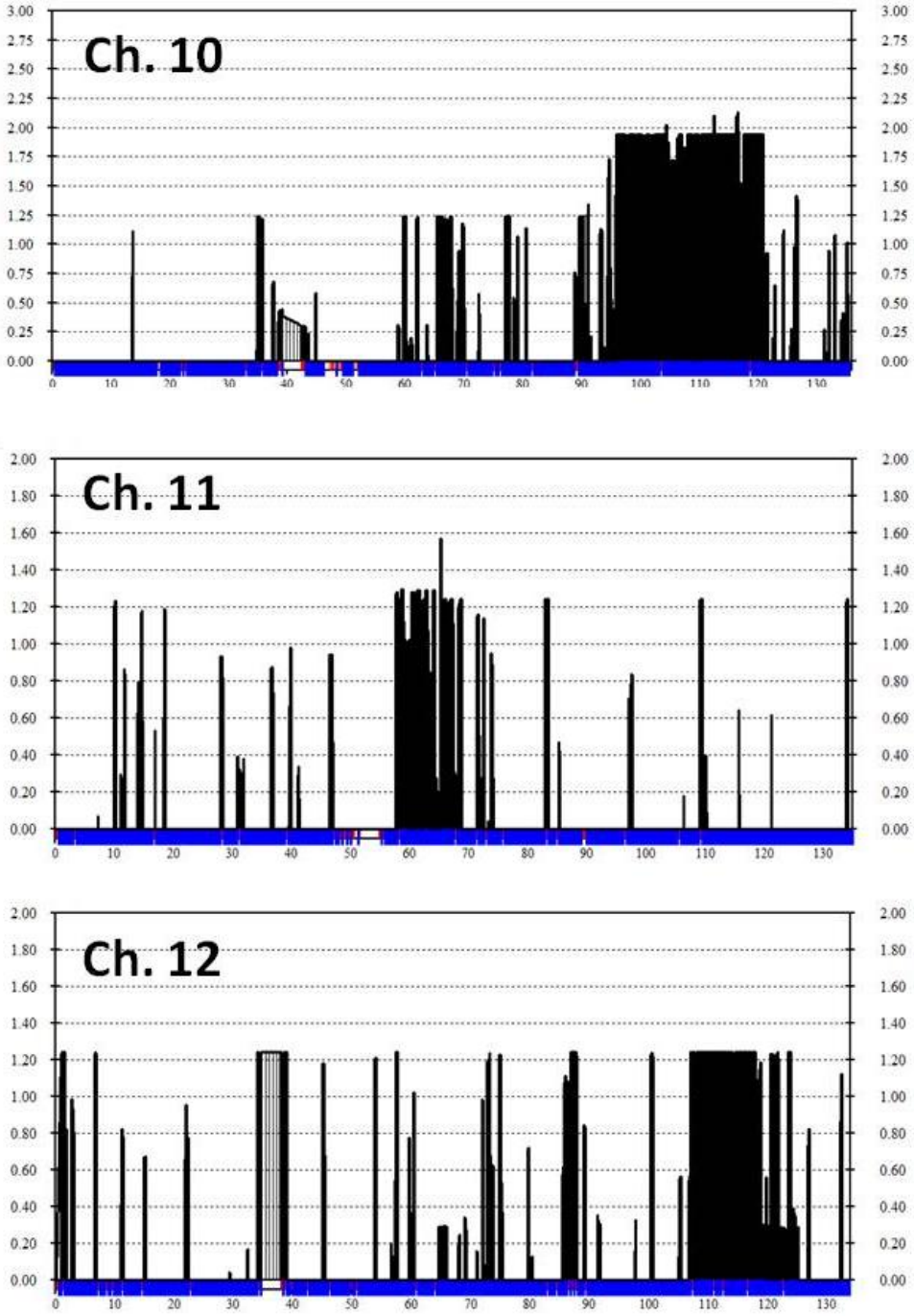


Fig. 5.23. Continued…….

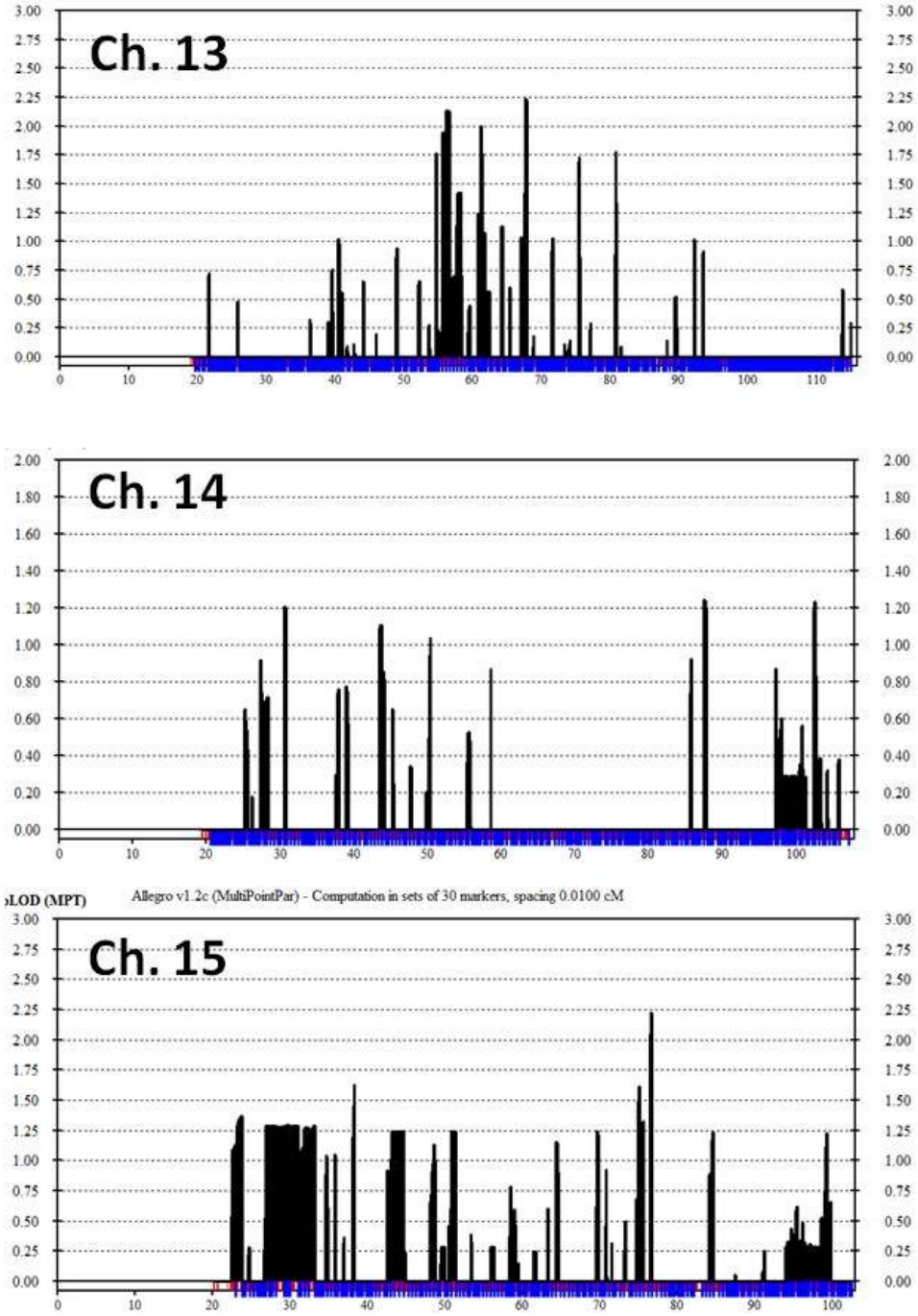


Fig. 5.23. Continued…….

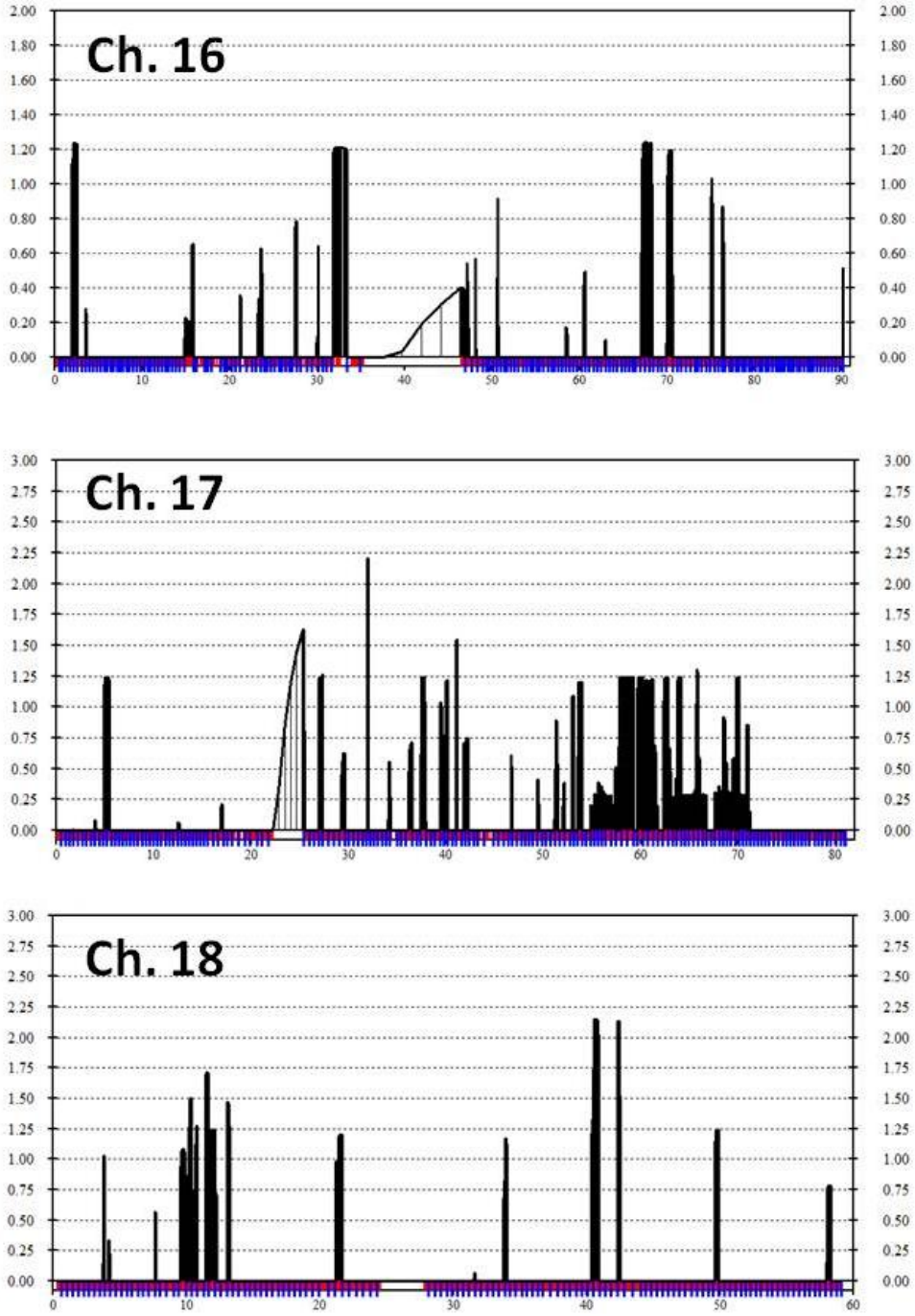


Fig. 5.23. Continued…….

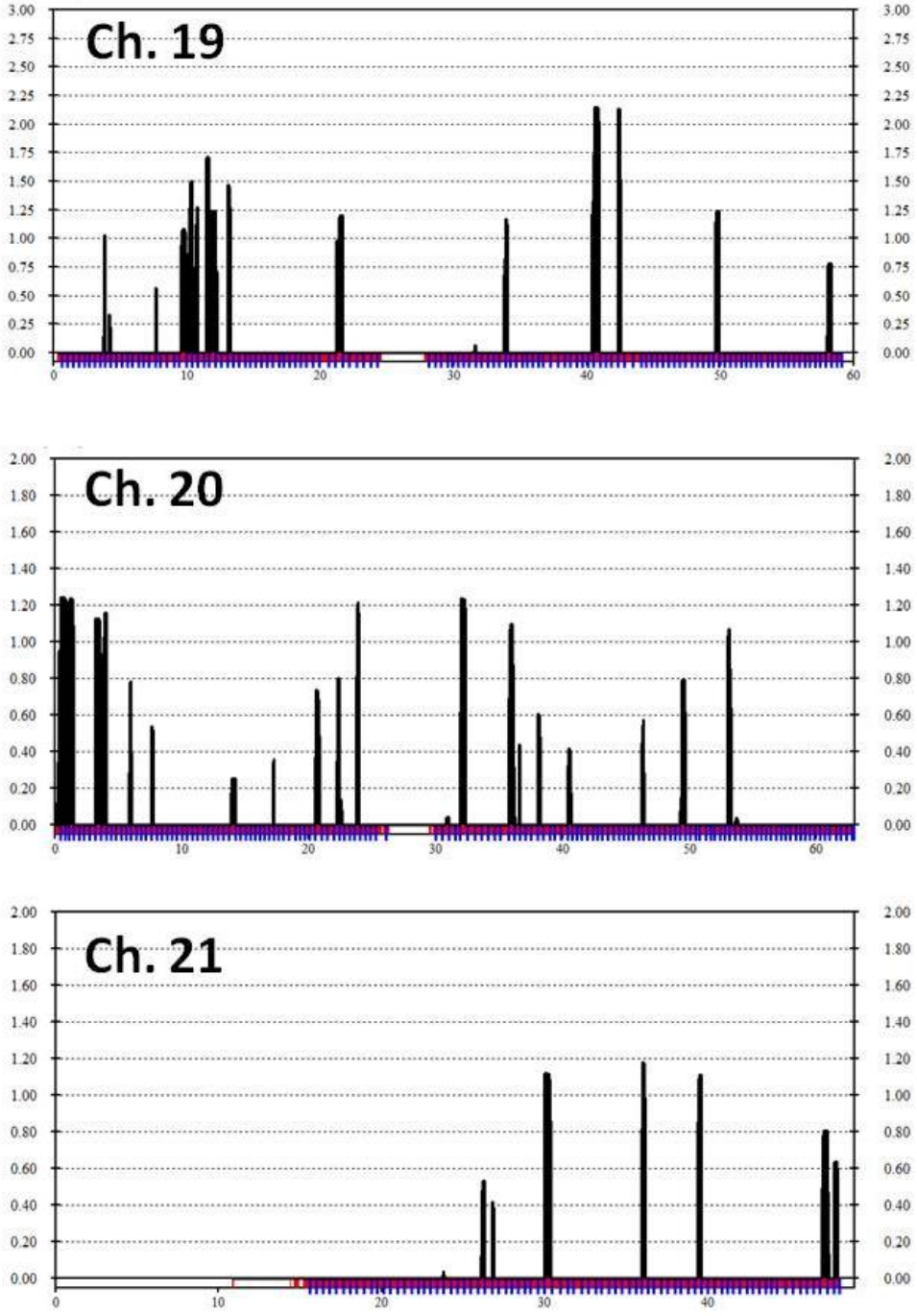


Fig. 5.23. Continued…….

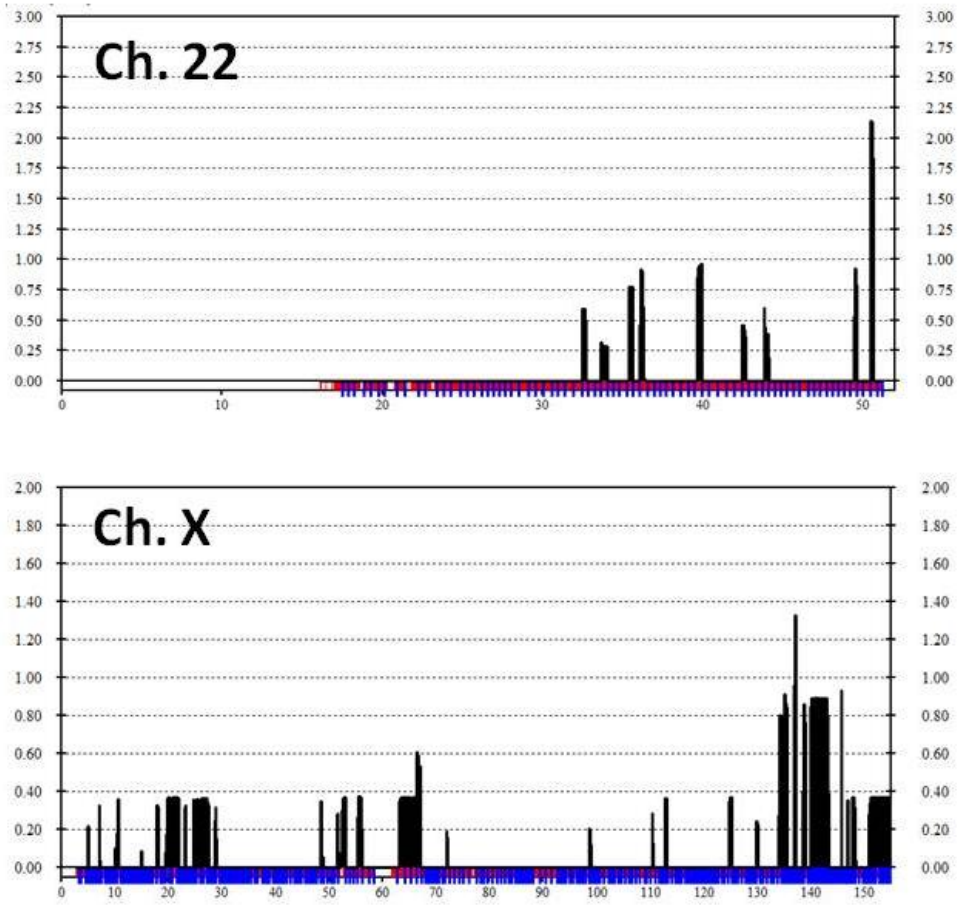


Fig. 5.23. Continued…….

The multipoint LOD score analyses revealed that the anomaly in the present family did not link to any known locus for tooth agenesis. The candidate regions were scrutinized through genome databases for the presence of putative genes with involvement in tooth/ectodermal development. It was observed that in one of those regions, locus 2p16.1-q14.2 harbored *EDAR* (ectodysplasin A receptor) which is known to be implicated in ectodermal dysplasia 10B hypohydrotic/hair/tooth type (ECTD10B; OMIM 224900). However, this condition segregates as an autosomal recessive condition. This region also encompassed *FBLN7* (fibulin 7) which is known to be involved in skin development. Furthermore, the locus on chromosome 6p22-p21.1 harbored *TNF* (tumor necrosis factor), which in addition to cancer, is also know to be involved in tooth genesis. Exome sequence data of affected subject IV-3 was available. However, no pathogenic variant was detected in the candidate genes present in the candidate intervals. SNP genotyping data were subjected to deletion/duplication analysis by using cnvPartition (v3.2.0) CNV Analysis Plug-in for GenomeStudio. No deletion or duplication was detected within the candidate loci.

In this family, even though we did not reach a threshold of LOD score of 3.0, several potentially interesting regions appeared as large blocks segregating with the phenotype. Due to the computational constraints, we could not incorporate the data of all genotyped subjects in the linkage calculations. It was concluded that there is still an uncharacterized gene on the candidate locus segregating with the phenotype.

5.6 Family F: Seckel-like Syndrome (SLS)

5.6.1 Clinical description, anthropometric measurements

Six sibs and their unaffected mother were physically examined in this kindred. The three affected siblings had a unique skeletal anomaly affecting the limbs, trunk and the skull. The limbs were disproportionate with small proximal and medial segments, but almost normal distal segments (hands, feet) (Fig. 5.24). Thinning of bones and joint dislocation were also noticeable. Others features included craniofacial abnormality such as large bulging eyes, beak-like protrusion of the nose, malformed ears, tooth deformity, unusually small jaw (micrognathia), and narrow face (Fig. 5.24 a,b). In addition, muscle wasting was evident in all (Fig. 5.24d). Siblings IV-1 and IV-2 had fragile bone and had undergone multiple surgical corrections after traumatic limb fractures (Fig. 5.25). According to their mother, they had intra-uterine growth retardation, were very small with low birth weight and had severe disproportionate short stature (summary of clinical finding in Table 5.18). Anthropometric measurements of the affected subjects are presented in Table 5.19.

Figure 5.24. **Photographs of affected subjects in SLS Family. a**: Front view of subject IV-2 showing narrow face, low-set ears and small jaw. **b**: Lateral view of subject IV-1 showing protruding nose, small jaw and thinning of hair. **c**: Right arm of subject IV-I showing normal distal segment but short middle segment, and abnormal elbow joint. **d**: Legs of subject IV-1 showing abnormality in knee joints, prominent veins, and muscle wasting.

Figure 5.25. **Radiographs of subject IV-11 in SLS Family. a**: In hand, cylindrical phalangeal elements are evident. There is lack of maturity in the digital elements and low bone age. **b**: Arm showing malformation in the distal head of humerus and dislocation of proximal radial head. **c**: Lower limbs with thin tibiae (surgical manipulation is evident in the right), and hypoplastic fibulae with marked decalcification. **d**: Left foot with crowding of tarsal bones and decalcification of phalangeal bones.

Table 5.18. Features of the members of SLS Family.

Subject ID	IV-1	1V-2	IV-4	IV-5	IV-6
Phenotype/variables	Seckel-like	Seckel-like	Seckel-like	Deaf and mute	Back bone problem
Growth					
Short stature, disproportionate	+	+	+	-	-
Intrauterine growth retardation	+	+	+	-	-
Slow postnatal growth	+	+	+	-	-
Thin and lean physique	+	+	+	-	-
Cranio-facial					
Longitudinal skull	+	+	+	-	-
Dysplastic ear	+	+	+	-	-
Prominent eyes	+	+	+	-	-
Sparse eyebrows and eyelashes	+	+	+	-	-
Nose: pointed, large	+	+	+	-	-
Irregular dentition	+	+	+	-	-
High pitch voice	+	+	+	-	-
Skeletal/limbs					
Prominent sternum	+	+	+	-	-
Dysplastic joints	+	+	+	-	-
Weak long bones	+	+	+	-	-
Muscles wasting	+	+	+	-	-
Limb shortening	+	+	+	-	-
Delayed bone age	+	+	+	-	-
Ectodermal					
Thin dermis with prominent veins	+	+	+	-	-
Thinning of scalpal hair	+	+	+	-	-
Transverse palmer creases	-	-	-	-	-

+= feature present; - = feature absent

Table 5.19. Anthropometric measurements of subjects in SLS Family.

Subject ID	IV-1	1V-2	IV-4	IV-5	IV-6
Phenotype/variable	Seckel-like	Seckel-like	Seckel-like	Deaf and mute	Backbone curvature (kyphoscoliosis)
Age (yrs)	22	21	16	15	10
Weight (kg)	**20**	**26**	**16**	44	16
Standing height (cm)	**113**	**137.2**	**99**	152.5	111.7
Sitting height (cm)	**61**	**70**	**68.6**	52.5	**49**
Arm span (cm)	**126.5**	**135**	**119.5**	161.7	125.3
Head circumference (cm)	**48.2**	**48.8**	**49**	54.8	47.7
Neck circumference (cm)	30.8	28.5	24.8	32.5	26.5
Chest circumference (cm)	71	69.5	61.5	81.2	59.4
Leg length (cm) R/L	**50.3/49.5**	**78.7/77.5**	**52.7/54.5**	90.5/90.5	**61.5/63**

* values in bold-face show significantly low estimate, when compared to controls for age and gender.

5.6.2 SNP genotyping and haplotype examination

SNP genome scan data of two affected (IV-1, IV-2) and two unaffected (III-1 and IV-6) subjects were available. Given the small family size and only two affected genotyped individuals, the LOD score calculations were not fruitful. Genotype data were aligned through Excel and the haplotypes were visually inspected in order to identify large stretches of homozygous regions (≥1 Mb) unique to the affected subjects. These analyses revealed that in this family there were a total of 51 regions across the genome that depicted homozygosity in stretches of ≥1 Mb. However, only 10 regions demonstrated homozygosity shared by only the affected subjects (Table 5.20). These regions did not overlap with any of the known locus implicated in Seckel syndrome or Seckel-like syndromes (Table 1.3). Five of these candidate regions harbored genes that are known to be involved in skeletal/dwarfism or microcephalic phenotypes (Table 5.20).

In order to identify putative pathogenic variant, the affected subject IV-1 was subjected to whole exome analyses. However, no potentially pathogenic variant could be detected among the genes harbored in the candidate loci. It was concluded that the potential gene might be present in the currently uncharacterized area within the candidate intervals, or there is yet a gene with unknown function.

Table 5.20. Homozygous regions segregating among the affected subjects in SLS Family.

Chr.	Start SNP	End SNP	Start nucleotide	End nucleotide	Interval size (Mb)	Candidate genes in the interval
1	rs16834780	rs472967	237,227,726	244,150,083	6.92	*FH*; fumarate hydratase; *AKT3*; v-akt murine thymoma viral oncogene homolog 3
2	rs1592	rs749873	135,722,143	136,817,088	1.09	*RAB3GAP1*; Warburg Micro Syndrome 1; WARBM1
4	rs7695536	rs3919827	57,955,234	67,067,806	9.11	
7	rs964823	rs1557658	54,621,134	93,966,036	39.34	*PSPH*; phosphoserine phosphatase; *KCTD7*; potassium channel tetramerization domain containing 7; *AUTS2*; autism susceptibility candidate 2; *POR, P450* (cytochrome) oxidoreductase; *PCLO*, piccolo presynaptic cytomatrix protein; *PEX1*, peroxisomal biogenesis factor 1; *CDK6*, cyclin-dependent kinase 6
9	rs3802386	rs1687392	95,879,087	117,099,956	21.22	*FANCC*, Fanconi anemia; *ERCC6L2*, excision repair cross-complementation group 6-like 2; *FKTN*, fukutin,
10	rs16930755	rs4746237	75,379,685	76,551,502	1.17	
12	rs1795963	rs11060180	122,281,575	123,303,586	1.02	
16	rs17215447	rs12933371	6,214,147	8,256,357	2.04	
16	rs152745	rs7187187	23,366,422	59,427,237	36.06	*PALB2*; *TUFM*; *TBX6*, T-box 6; *ORC6*, origin recognition complex, subunit 6; *SALL1*, spalt-like transcription factor 1; *COQ9*, coenzyme Q9,
X	rs5942915	rs11152658	109,549,026	110,620,035	1.07	

5.7 Family G and H: Syndactyly type IX, Mesoaxial Synostotic Syndactyly (MSSD)

5.7.1 Clinical description

The affected subjects in both families had remarkable phenotypic similarities, i.e., they had four fingers, bilaterally. In both hands, 3rd and 4th fingers were replaced by a single thick digit. The fifth fingers had clinodactyly. In the feet there was variable cutaneous webbing of toes 1-4.

Family G: One affected subject and her unaffected father were physically examined. Affected subject (II-1) was 24-years of age at the time of examination. She was a university student. There was no other anomaly except the involvement of upper autopods. She had four digits in both hands, i.e., the mesoaxial digits were missing, bilaterally (Fig. 5.26 A,B). In both hands and particularly in the right, there was a wide gap between second and "fourth" fingers, giving an evidence of absent third digit. The existing fingers were relatively short, hypoplastic and gave an impression of radial deviation. The fifth fingers were demonstrating clinodactyly, more pronounced in the left hand. Fingertips were tapering with hypoplastic nail plates. In the second and fourth digits the nails were reduced to small islands. The feet were unremarkable.

Family H: There were two affected subjects in this kindred (1M, IF), of which one had been deceased. The available affected subject and his unaffected father were physically examined. The index subject was a 23 years old male (IV-4), who was a university student. The subject was third in the sibship of five and his parental marriage type was ‘first-cousin-once-removed.’ The index subject was observed to have four digits in both hands and one of the mesoaxial digit was omitted, bilaterally (Fig. 5.26 C,D). The thumbs were short and the index fingers depicted valgus deviations at their terminal phalanx. Additionally, the left index finger had symphalangism. The third and fourth fingers were replaced by a single digit, which demonstrated broad proximal phalanx, hypoplastic and club shaped terminal phalanx, and reduced nail, bilaterally. The fifth fingers were dysplastic and exhibited clinodactyly. At the palmer side, characteristic dermatoglyphic changes were evident, i.e., triradii was grossly distorted and inter-digital flexion creases were less remarkable. The phenotype in the feet was characterized by hallux varus, partial zygodactyly between second and third toes, and shortening of fourth toes, bilaterally. Generally, the nails of digits 2-4 were hypoplastic. There was no anomaly of any other organ-system in the subject.

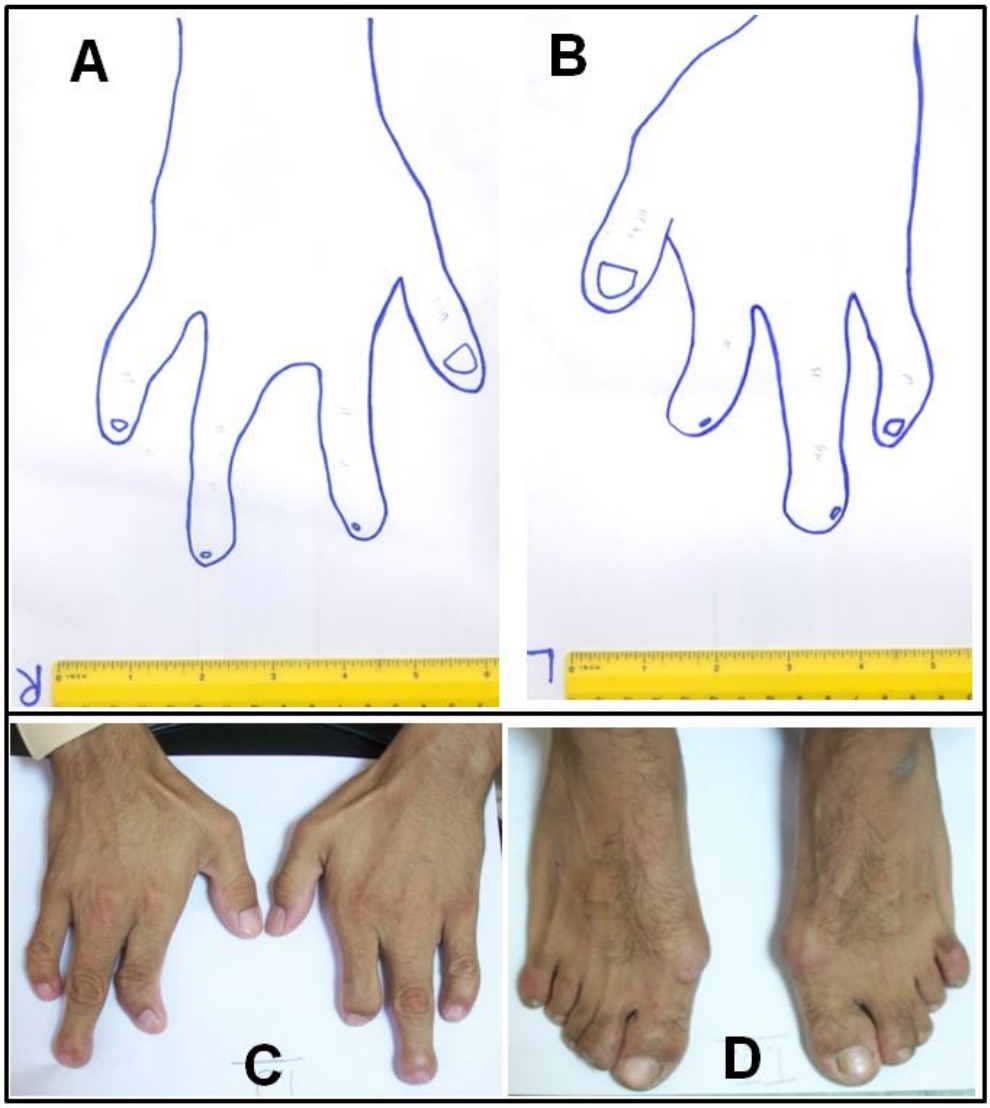


Figure 5.26. Phenotype of the index subjects in Family G (A-B, line drawing) and Family H (C-D).

5.7.2 Sequencing of *BHLHA9*

Owing to the remarkable similarity of the phenotypes segregating in these families with type IX syndactyly (MSSD), we decided to directly conduct mutation screening of *BHLHA9.* To screen for pathogenic mutation in the coding sequences of *BHLHA9* (NM_001164405.1) all the exons and their boundaries were PCR amplified. The index subjects in both families were subjected to mutation screening.

In Family G, a homozygous substitution c.218G>C (p.R73P) was detected in exon 1. In Family H, a homozygous substitution c.211A>G (p.N71D) in exon 1 was detected in the affected subject (Fig. 5.27). Both these variants were absent in the control panel and were not reported in the ExAC and 1000-genome databases. Exon 1 of *BHLHA9* determines a highly conserved basic helix-loop-helix protein. Two of the mutations observed in these families are positioned in close neighborhood and exchange critical, conserved amino acids of the DNA-binding region of the basic domain (Fig. 5.28).

These variants were predicted to be pathogenic and disease causing by the online prediction tools (Table 5.21). For instance, both of these variants were predicted to be damaging by SIFT online tool. PolyPhen-2 predicted them as probably damaging (N71D: score 0.998; sensitivity: 0.27; specificity: 0.99; R73P: score 0.999; sensitivity: 0.14; specificity: 0.99). MutationTaster predicted both variants as disease causing. In MutPred, the probability of deleterious mutation was 0.766 and 0.831 for N71D and R73P, respectively (Table 5.21).

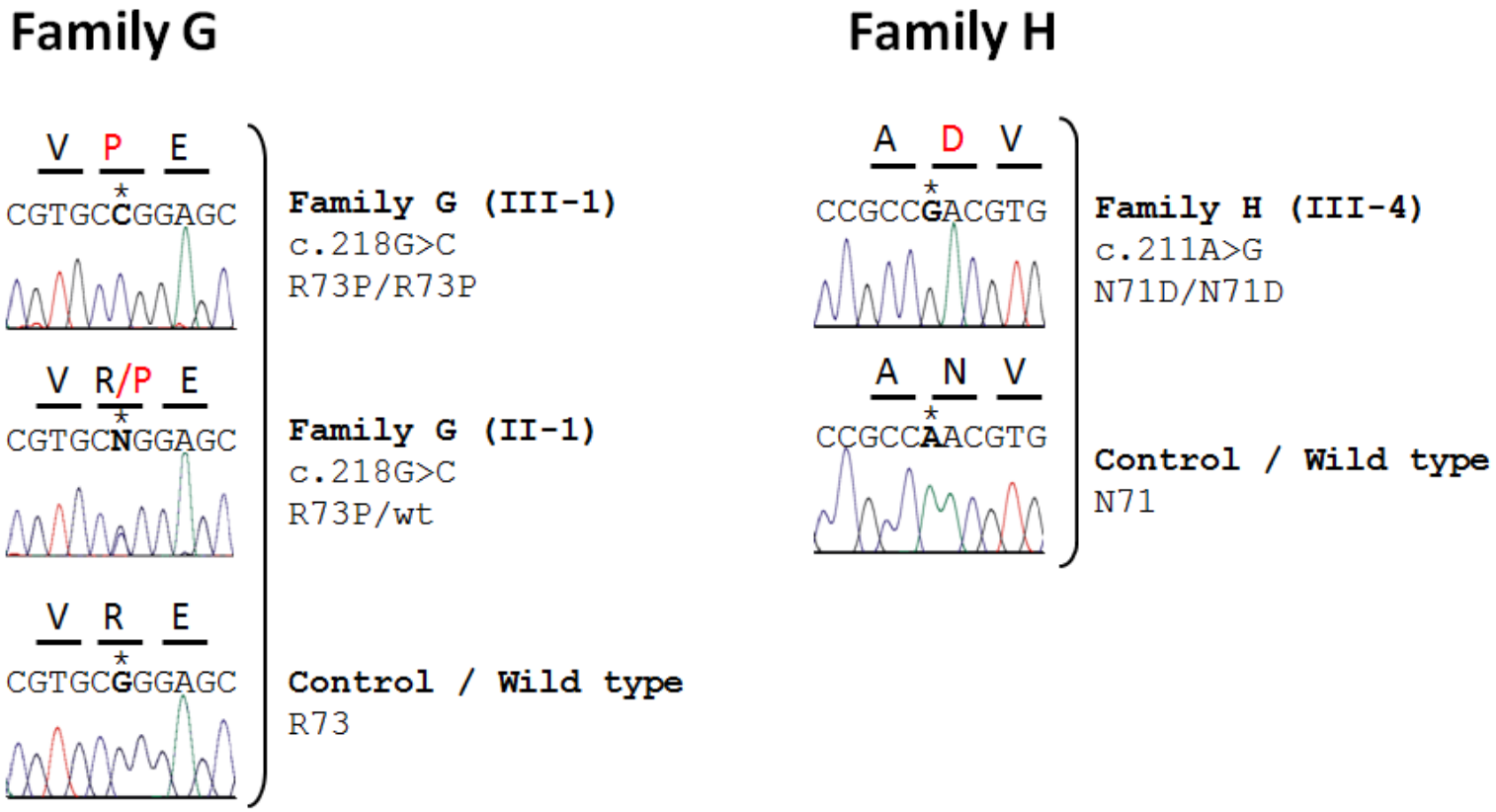


Figure 5.27. **Chromatograms showing mutation c.218G>C in Family G and c.211A>G in Family H.**

Mutant amino acids are depicted in Red.

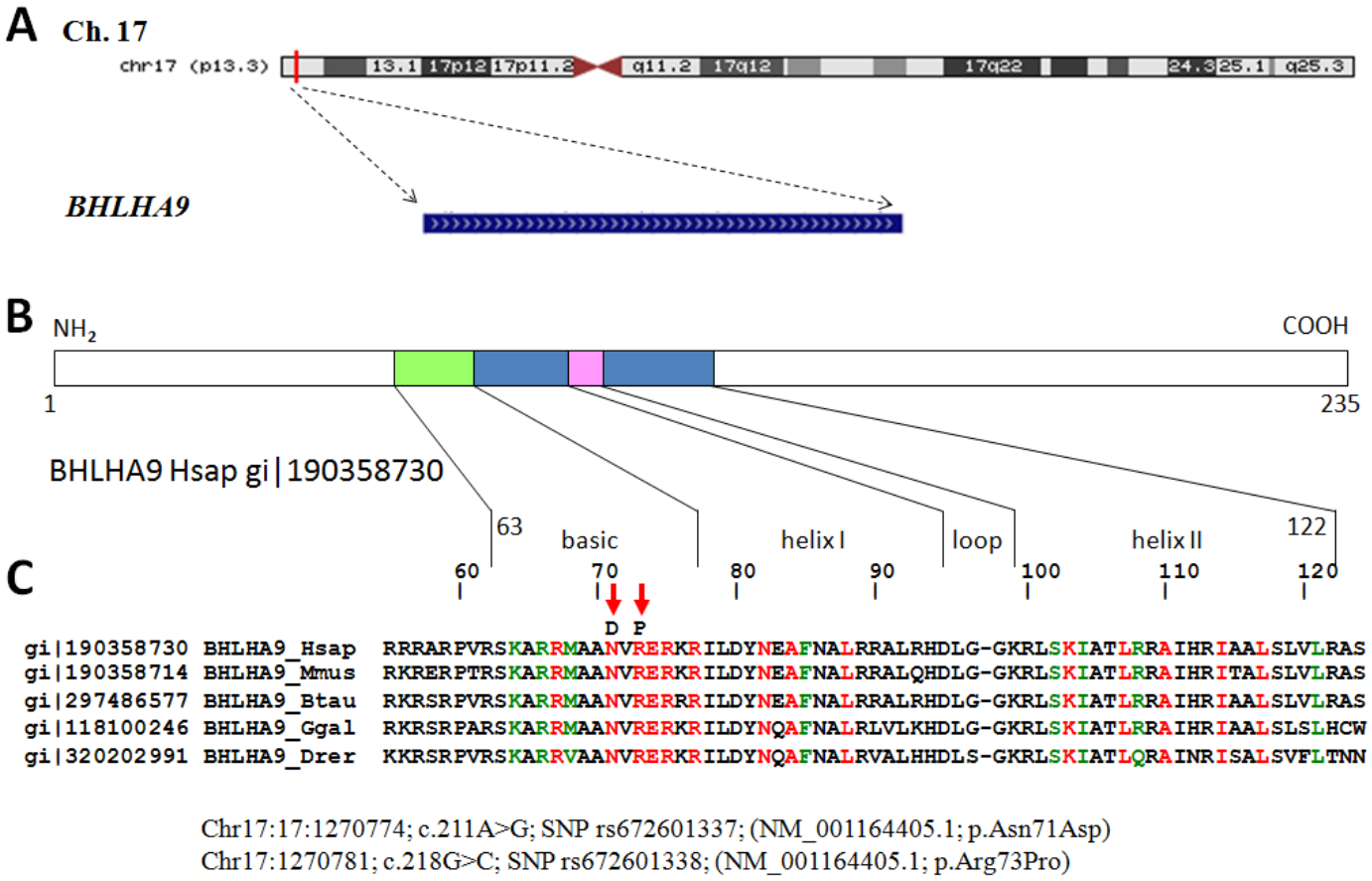


Figure 5.28. (A) **Genomic organization and location of mutation of *BHLHA9* at chr17p13.1.** *BHLHA9* has a single exon. (B) Model of the human BHLHA9 protein. (C) Mapping of MSSD-causing missense mutations to the basic nucleic acid binding site enclosing the E-box specificity site DNA-binding region of BHLHA9, projected onto an amino acid sequence alignment of orthologous Bhlha9 sequences from various vertebrates (Mmus = Mus musculus, Btau = Bos taurus, Ggal = Gallus gallus, Drer = Danio rerio).

Table 5.21. Consequences of the missense mutations in the first exon of *BHLHA9*, as depicted by the computational tools.

Family	**Mutation**	**Prediction tools**				
		SIFT:	**PolyPhen2**	**MutationTaster**	**MutPred**	**MutPred**
G	c.218G>C; p. R73P	DAMAGING - Low confidence predictions with Median conservation above 3.25	PROBABLY DAMAGING with a score of 0.999 (sensitivity: 0.14; specificity: 0.99)	disease causing	Probability of deleterious mutation 0.831	Gain of ubiquitination at K76 (P = 0.0303) Loss of helix (P = 0.0558) Gain of glycosylation at R73 (P = 0.0591) Loss of MoRF binding (P = 0.0921) Gain of methylation at K76 (P = 0.1203)
H	c.211A>G; p. N71D	DAMAGING - Low confidence predictions with Median conservation above 3.25	PROBABLY DAMAGING with a score of 0.998 (sensitivity: 0.27; specificity: 0.99)	disease causing	Probability of deleterious mutation 0.766	Gain of ubiquitination at K76 (P = 0.0314) Loss of methylation at R67 (P = 0.0831) Gain of helix (P = 0.132) Loss of loop (P = 0.2897) Loss of stability (P = 0.3857)

SIFT (http://sift.jcvi.org/www/SIFT_enst_submit.html), PolyPhen-2 (http://genetics.bwh.harvard.edu/pph2/), MutationTaster (http://www.mutationtaster.org/), MutPred (http://mutpred.mutdb.org/)

The results obtained in families G and H have been published in Am J Hum Genet:

Malik S, Percin FE, Bornholdt D, Albrecht B, Percesepe A, Koch MC, Landi A, Fritz B, Khan R, Mumtaz S, Akarsu NA, Grzeschik KH. Mutations affecting the BHLHA9 DNA-binding domain cause MSSD, mesoaxial synostotic syndactyly with phalangeal reduction, Malik-Percin type. Am J Hum Genet. 95(6):649-59, 2014. doi: 10.1016/j.ajhg.2014.10.012

6 DISCUSSION

In this thesis, high-density SNP genotyping arrays and exome sequencing were applied to solve the genetic basis of monogenic familial disorders by identifying the disease loci, genes, and mutations. These robust techniques have accelerated the field of genetics by decreasing the cost and increasing the efficiency in disease gene identification to the extent that now these have replaced traditional methods of the disease gene hunt using microsatellite-based genotyping, screening of known disease genes for mutations (candidate gene approach), exclusion mapping, and SSCP, even in clinical settings. Whole genome scan (SNP genotyping) in family members can be used to detect a haplotype that may be harboring the disease gene. In addition, it is possible to analyze copy number variations (CNVs) with this approach that further reduces the cost of additional molecular tests. The majority of scientific reports have shown that mutations responsible for most of the genetic disorders are in the coding regions of the genome. Hence, after the identification of disease gene locus, exome sequencing is the method of choice.

In the present research, I employed linkage analyses and mutation screening in eight Pakistani families afflicted with different monogenic disorders. In all families, first linkage analysis was performed on whole genome scan SNP data and candidate loci were identified. Model-based LOD score analyses were performed in order to detect the loci segregating with the phenotypes in these families. In the next step, in three families, whole exome sequencing was performed for one affected family

member to find the underlying mutation(s). Generally, a subject with most severe phenotype in the family was selected for exome sequencing. The exome data were analyzed and mutations were found in all of the three families. Variants generated by exome sequencing were prioritized by their chromosomal location, population frequency, predicted effect on protein function, and expression of the gene in relevant tissues. In the fourth family, a very strong candidate gene at the identified locus was hypothesized to be responsible for the disease and thus the study was terminated.

6.1 Family A: Intellectual Disability with Speech Problem and Dysmorphic Facial Features (IDSD)

Family A was afflicted with a recessive intellectual disability, language problem, and dysmorphic facial features. A total of six family members were diagnosed with the disease. The disease gene locus was identified at 15q21.1-26.3, and exome sequencing revealed that the disease-causing mutation is the novel, homozygous missense mutation *PDIA3* p.Cys57Tyr. The mutation was validated by Sanger sequencing and was found to be segregating with the disease. The substitution of a cysteine residue (a sulfur-containing amino acid involved in disulfide bond formation that is important for the folding and stability of proteins) with tyrosine (an aromatic, partially hydrophobic amino acid remains buried in protein hydrophobic cores) is predicted to adversely affect the protein structure by the online prediction

tools employed. Online tools such as Mutation Taster, SIFT, PolyPhen2 and MutPred predicted the variant as damaging.

We also investigated the conservation of the protein sequence across species. All 505 amino acids showed absolute identity (100%) in human and chimpanzee and 99% in human and rhesus monkey which had 504 amino acids. The mutated residue Cys57 is within a stretch of eight amino acids that are fully conserved in all eukaryotic organisms explored and within a stretch of 19 amino acids that are conserved in full in bird (jungle fowl) and mammals including human, chimpanzee, rhesus monkey, wolf, house mouse and rat except cattle that differs in one amino acid which is similar to the organisms of lower phyla.

Protein disulfide-isomerase A3 (*PDIA3*) encodes ERp57, a 505-amino acid disulfide that is an isomerase enzyme. This endoplasmic reticulum protein interacts with other proteins for modulating the folding of recently synthesized proteins (Santana-Codina et al. 2013). It has also been demonstrated that due to its disulfide isomerase activity, the protein enhances disulfide bond formation in the glycoprotein substrates. The gene is highly expressed in many tissues including the brain.

The role of *PDIA3* is evident in the late onset nervous system disorder amyotrophic lateral sclerosis (ALS), a neurodegenerative disorder. Gonzalez-Perez et al. (2015) reported rare, heterozygous gene variants in ALS patients. However, to the best of our knowledge, no studies were published to show a role of *PDIA3* in a

neurodevelopmental disorder. To find out the role of *PDIA3* in the nervous system, Castillo et al. (2015) conducted a study in which they produced transgenic mice with overexpression of *PDIA3* in the nervous system and found that the protein has a role in axonal regeneration after an induced injury. The authors claimed a functional role of the protein in axonal regeneration and an effect on peripheral nerve regeneration.

In addition, the role of *PDIA3* has been established by studying the mouse model in the assembling process of the peptide-loading complex of major histocompatibility complex (MHC). In *PDIA3*-deficient cells, the interaction between MHC class I molecules and the loading complex was found to be short-lived (Garbi et al. 2006). It is also a candidate protein for a variety of cancers (Dihazi et al. 2013).

Here, we propose that alteration of conserved residue Cys57 may decrease the activity of this protein and impair its role in the early stage of axon development that may lead to ID. Therefore, it was concluded that the *PDIA3* mutation detected underlies the ID in the presented family. The next target of our research team is to generate animal and cell models for this mutation to find its effect at the cellular and biochemical levels.

6.2 Family B: Autosomal Recessive Primary Microcephaly (MCPH)

Family B was afflicted with autosomal recessive primary microcephaly (MCPH). In the family, four sibs were diagnosed as MCPH. DNA samples of four affected and one unaffected family members were subjected to whole genome scan. The strongest locus found by linkage analysis was 1q25.3-q32.1 that harbors *ASPM* (OMIM 60816) with 28 exons, and more than 39 reported mutations. Since this gene has been reported previously with several mutations in Pakistani MCPH families as well, we concluded that this is the most likely gene responsible for the disease in the family and it would not be possible to publish this work. So the DNA of an affected family member was not subjected to exome sequencing or Sanger sequencing for the gene, and the study was terminated.

In MCPH, a significant cognitive decline occurs and brain size develops only to one-third of the adult size. The disease exhibits high genetic heterogeneity, and studies have suggested that it is possibly a primary disorder of neurogenic mitosis (Faheem et al. 2015). With the advancement in new generation genetic technologies, there is an improvement in gene identification and discovery of new genes involved also in MCPH. It is very helpful to detect the recurrence of the disorder by prenatal testing, to perform a differential diagnosis of the disorder by postnatal testing, and by carrier testing to provide help in decision making for consanguineous families that are suffering from the disorder for generations (in which the disease is known to occur).

That would eventually enable us to reduce the incidence of MCPH in highly consanguineous populations. These discoveries along with good neuroimaging techniques also facilitated the establishment of exact genotype-phenotype correlations (Faheem et al. 2015). Those studies have increased our understanding of the underlying cellular mechanism as each new gene discovery helps to uncover one relevant cellular pathway. Therefore, with the study of more families afflicted with MCPH, we would be able to discover new genes that will help in resolving fundamental cellular pathways involved in neuronal production which would eventually increase our knowledge of disease mechanisms.

6.3 Family C: Microcephaly, Intellectual Disability, Short Stature and Brachydactyly (MIDSB)

In Family C, genetic analyses led to the conclusion that *RBBP8* mutation underlies the disease. The work has been published in the American Journal of Medical Genetics (Mumtaz et al. Am J Med Genet A. 167A (12):3148-52, 2015) with the title "RBBP8 syndrome with microcephaly, intellectual disability, short stature and brachydactyly". In the literature, there were four families reported with homozygous mutations in *RBBP8*. The family we studied is thus the fifth family reported with *RBBP8* mutation. Therefore, due to the rarity of this malformation, it is

still not possible to establish a phenotype–genotype correlation for *RBBP8* mutations. The homozygous truncating mutation detected in this kindred was previously reported in a family with Jawad syndrome with clinical features of microcephaly, learning disability, anonychia congenita, syndactyly and polydactyly (Hassan et al. 2008). Other families reported with *RBBP8* mutation also had diverse phenotypes (Table 6.1). The SCKL2 patients had intellectual disability, microcephaly, and short stature along with cardiac and other problems (Qvist et al. 2011). SCKL2 and Jawad type manifest with intellectual disability, microcephaly, anonychia, minor facial anomalies, epilepsy, and diabetes mellitus, but the latter two symptoms had been attributed to a heterozygous deletion in *NRXN1* (Agha et al. 2014). The patients in the fourth family with *RBBP8* mutation had the major clinical findings of primordial dwarfism and cranio-facial anomalies but without ID (Shaheen et al. 2014). The clinical features of the patients in the present family partly overlap with all four aforementioned conditions, yet the syndrome we studied is distinct from those. As in the other *RBBP8* mutation disorders, our patients have intellectual disability, microcephaly, developmental delay and facial anomalies. They also have short stature as in SCKL2 but the appendicular skeleton is more affected in our family and dwarfism is of disproportionate type. Other dissimilarities include anomalies of limbs and cervical vertebrae (Table 6.1). Similarly, the new disorder is distinct from Jawad syndrome where there is no short stature or brachydactyly but instead polydactyly or syndactyly, which are not in the present family. It can be said that, unlike in SCKL2 and Jawad, our patients do not have epilepsy, diabetes mellitus, high-pitched voice or hyperhidrosis. Hence, this syndrome can be referred as an atypical SCKL2. Our findings allow all these entities to be grouped as "*RBBP8* syndromes". The

phenotypic similarities in all these families establish that *RBBP8* mutation plays a role in the disease development.

RBBP8 encodes retinoblastoma-binding protein 8 that is mainly nuclear. This protein is among a group of proteins that interact with retinoblastoma protein, which controls cell division. Studies have shown that the protein interacts with BRCA1 during regulation of transcription, cell cycle checkpoint and DNA repair mechanisms. It is also thought that this protein may act directly as a tumor suppressor (Sartori et al. 2007). To test this hypothesis, an experiment was conducted by Qvist et al. (2011) on cell lines derived from SCKL2 patients. Their results showed that at the cellular level the mutation causes the impairment of mechanisms involved in DNA damage. However, the detailed mechanisms through which *RBBP8* may affect brain morphogenesis and the development of other organ-systems remain to be elucidated.

Table 6.1. Clinical manifestation of *RBBP8* mutation.

Malformation	SCKL2	Jawad	Complex microcephaly*	Primordial dwarfism	Present family
OMIM	606744	251255	–	–	–
Family origin	Iraq	Pakistan	Pakistan	Saudi Arabia	Pakistan
No. of affected individuals	4	7	2	1	7
Cardinal features					
Microcephaly	Yes	Yes	Yes	No	Yes
Short stature	Proportionate type	No	Mild	Proportionate type	Disproportionate type
Intellectual disability	Yes	Yes	Yes	–	Yes
Minor facial anomalies	Yes	Yes	Yes	Yes	Yes
Prominent nose	Yes	Yes	Yes	No	Yes
Short anterior cranial base and maxillary length	Yes	–	–	Yes	Yes
High-pitched voice	Yes	No	Yes	–	No
Cardiac murmur	Yes	–	No	–	No
Birth spots	Yes	No	No	–	No
Hyperhidrosis	No	–	Yes	–	No
Epilepsy*	No	–	Yes	–	No
Obesity/diabetes*	No	No	Yes	-	No

Hyperhidrosis	-	-	Yes	-	No
Developmental delay	Yes	Yes	Yes	Yes	Yes
Limb anomalies					
Pigmented spots on hands/feet skin	No	Yes	Yes	–	Only on hands
Polydactyly	No	Yes	No	–	No
Syndactyly	No	Yes	No	–	No
Clinodactyly	Yes	Yes	No	–	No
Brachydactyly	No	No	No	No	Yes
Slender extremities	Yes	No	No	–	No
Anonychia	No	Yes	Yes	–	Yes
Swelling over the phalangeal joints	Yes	Yes	No	–	Yes
Absent ossification centers of distal/middle epiphyses in 5th finger	Yes	–	–	–	No
Mutation	c.T2347+53G	c.1808_1809 delTA	c.A919G; deletion in *NRXN1*	c.C298T	c.1808_1809 delTA
Reference	(2)	(4)	(5)	(6)	Present study

–, not reported.

*Features attributed to *NRXN1* deletion mutation.

6.4 Family D: Autosomal Dominant Postaxial Polydactyly, Camptodactyly and Zygodactyly (ADPCZ)

Family D was diagnosed with postaxial polydactyly. However, there were some additional finding in the family such as camptodactyly and zygodactyly that are typically seen in syndactyly type II or synpolydactyly (SPD; OMIM-186000) (Malik and Grzeschik, 2008; Malik, 2014). Due to these overlapping clinical manifestations, it is suspected that the causative mutation could be in one of the synpolydactyly (2q31, 22q13.31, and 14q11.2-q12) or polydactyly (7p14.1, 7q36, 13q21-32 and 19.p13.1-13.2) loci. So, linkage analysis was performed in search of the disease locus. The locus was identified as 7p14.1 that harbors a large candidate gene *GLI3* responsible for preaxial and postaxial A1 and B polydactyly. We proceed for exome sequencing in order to search for a mutation in this gene and also possibly in any other gene at the candidate loci yielding LOD scores >2.5. Exome analysis revealed a novel mutation in *GLI3*; however, no other variant fitting the criteria of mutation was found in any linked region. Our manuscript "Complex Postaxial Polydactyly Type A/B with Camptodactyly, Hypoplastic Third Toe, Zygodactyly and Other Digit Anomalies Caused by a Novel *GLI3* Mutation" has been submitted in Am J Med Genet.

GLI3 (GLI family zinc finger 3; OMIM 165240) encodes a zinc finger transcription factor that binds directly to the target gene's promotor region and directs its activity (Vortkamp et al. 1995). Its role is also evident in sonic hedgehog pathway.

Various mutations in *GLI3* in syndromic and non-syndromic digital abnormalities including polydactyly called *GLI3* morphopathies have been reported (Radhakrishna et al. 1999). From the literature review, it is evident that mutations in *GLI3* mostly cause Pallister-Hall syndrome (PHS), Greig cephalopolysyndactyly syndrome (GCPS) and isolated polydactyly, either postaxial type A/B (MIM 174200) or crossed polydactyly type I (MIM174700) (Radhakrishna et al. 1999; Cheng et al. 2011).

We attempted to investigate a genotype-phenotype correlation for *GLI3* diseases by considering also the reported mutations. Jamsheer et al. (2012) proposed that GCPS and PHS syndromes are produced by mutations in the first and second third part of the GLI3 protein, respectively. However, we observed that the mutations leading to polydactyly type I were mostly in the central part of GLI3 protein and while in postaxial polydactyly type A/B the first third of GLI3 protein was found to be harboring the majority of mutations (Fig. 6. 1).

For example, mutation p.Arg643* in an Indian with postaxial polydactyly A1/B, fusion of toes and zygodactyly (Radhakrishna et al. 1999), was also found in a Chinese family with crossed polydactyly type I, preaxial polydactyly, preaxial webbing and broad hallux (Wang et al. 2014). It was proposed that different manifestations of the same mutation may be attributed to a modifier gene (Cheng et al. 2006). In order to investigate for any modifier variants in the presented family, all regions with LOD scores >2.5 were analyzed for a second gene having a deleterious variant fitting the criteria of mutation, but no such variant was found. Findings in the present study are also contradictory to those described by Johnston et al. (2005). The

authors proposed that mutations from nucleotides 1998 to 3481 in *GLI3* lead primarily to PHS. However, in the presented family, *GLI3* mutation was in c.2884delG but without any clinical overlap with PHS. For instance, features like hypothalamic hamartoma, imperforate anus, and deformed laryngeal cleft are found in PHS but not found in any member of this family. Thus, our findings do not agree with the proposed phenotype-genotype correlation for *GLI3* limb-specific mutations.

In conclusion, a novel *GLI3* mutation was identified in the presented family with a variable phenotype that further expands the phenotypic spectrum of *GLI3* mutation.

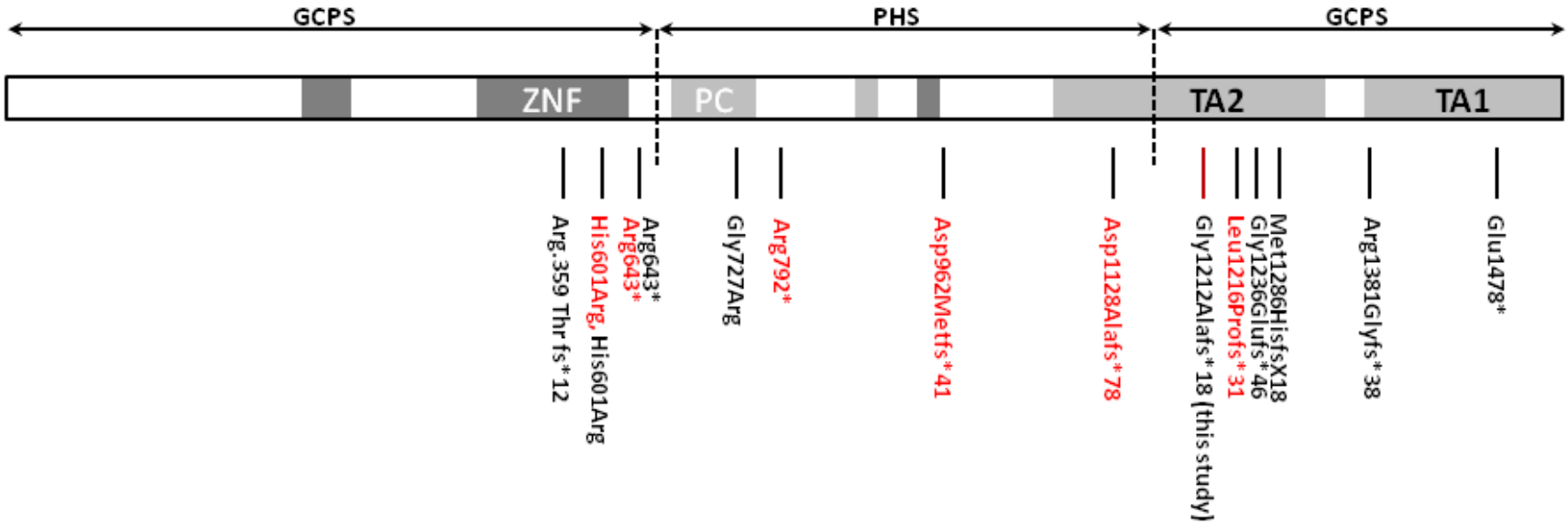


Figure 6.1. **Schematic representation of GLI3 domains and mutations causing isolated limb malformations.** Mutations causing postaxial types A1/B are in black, and mutations causing crossed polydactyly type I are in red.

6.5 Family E: Tooth Agenesis Manifesting as Oligodontia/Hypodontia (TAOH)

The pedigree of this family provided a convincing evidence for an autosomal dominant mode of inheritance. The most commonly reported mode of inheritance for this anomaly is autosomal dominant, but also autosomal recessive, X-linked, polygenic or multifactorial models of inheritances have been reported (Suarez and Spence, 1974; Brook, 1984; Peck et al. 1993).

In this family, the mean number of missing teeth in the five affected individuals was five (range 2-8 teeth), which is less than the mean number of missing teeth described in other studies. For instance, in the study by Song et al. (2009) the mean number of missing teeth was 14 with a range of 8-18. In another study by Lexner et al. (2007) the mean number of missing teeth was 22 with a range of 14-28. The four affected individuals in the present family had missing central or lateral incisors. This observation is consistent with the phenotype reported in four patients by Song et al. (2009). It was noticeable that two of the participants had congenitally missing maxillary and mandibular canines, which is very rare in isolated oligodontia with other genetic defects (Kim et al. 2006). In addition, first premolars were observed to be absent in only one subject which was also present in patients reported by Song et al. (2009).

In this family shapes and positions of the existing teeth were also abnormal in association with missing teeth. In the previous studies different types of tooth

agenesis were seen to be frequently associated with other dental anomalies, including microdontia (Garn and Lewis, 1970), ectopias and infraocclusion of deciduous molars (Baccetti, 1998; Garib et al. 2009), delayed dental development (Baba-Kawano et al. 2002), conical and tapered incisors (Lexner et al. 2007), and generalized enamel hypoplasia (Baccetti, 1998). The most likely explanation is that it may be due to same underlying genetic defect, i.e., clinical variability by a single gene which results in more than one dental anomaly but genetic heterogeneity cannot be excluded. Additionally, Song et al. (2009) reported cases in which the missing teeth were distributed extensively in both dentitions but the shapes and the sizes of the residual teeth were observed to be normal.

Genetic analyses in this family yielded several interesting regions which segregated among the affected subjects, however, a LOD score of 3.00 was not obtained on any chromosome due to the small number of individuals included in the calculations. Due to time and computational constraints, the linkage software are restricted to execute the data of 5-6 genotyped subjects. Adding more non-founders and genotyped subjects increases the computational time on an exponential scale which makes the analyses more trivial. It is quite likely that adding three additional genotyped subjects (IV-9, IV-10, IV-11) as a part of secondary pedigree in the linkage calculations would improve the results. The exome data analyses did not reveal any potential pathogenic variant in a known gene in the candidate intervals. It is quite likely that there is still an uncharacterized gene on the candidate locus segregating with the phenotype.

6.6 Family F: Seckel-like Syndrome (SLS)

The family presented here exhibited a severe skeletal disorder. The subjects were born to consanguineous parents. The anomaly was running in an autosomal recessive inheritance pattern. All the cases of Seckel and Seckel-like syndromes reported to date had the same mode of inheritance (OMIM 210600).

Typical manifestations of Seckel syndrome are a profound prenatal and postnatal growth retardation, bird-headed profile with a beaked nose, micrognatia, lobeless ears, crowded teeth, delayed bone age and elbow dislocation (Seckel 1960). Additionally, microcephaly with mental retardation, proportionate short stature with the normal ratio between the upper and lower body, have also been observed. The later mentioned features were absent in the studied family. Conversely, muscle wasting is not a feature of Seckel syndrome but was observed in the family under study. Another Seckel-like syndrome called Majewski osteodysplastic primordial dwarfism type II (MOPD II) is characterized by the absence of mental retardation and disproportionate growth retardation (Majewski et al. 1982a). But its features include shortening of the distal and middle limb segments, which was not observed in the family we studied. Instead, disproportionate short stature was unique, as all the affected persons had normal hands and feet while all other long bones were short. Thus, many clinical features in this family were not similar with the reported families.

SNP genotyping revealed that this condition does not map to any of the already known loci for Seckel syndrome or bird-headed dwarfism and there is yet an unknown locus for the malformation in this family. We could not detect any potentially pathogenic variant among the genes present at the candidate loci. It was concluded that the putative gene might be residing in a region within the candidate intervals not analyzed or in a genomic region whose sequence is yet unknown.

6.7 Family G and H: Syndactyly type IX, Mesoaxial Synostotic Syndactyly (MSSD)

Syndactyly type IX or MSSD demonstrates a distinctive combination of features in hands and feet and is unique among the syndactyly types in its characteristic clinical presentation and autosomal recessive inheritance. Hand and foot abnormalities may include hypoplastic thumbs, clinodactyly of second and fifth fingers, syndactyly, and shortened or absent middle phalanges (Malik et al. 2004). The phenotypic features in the affected families suggest a developmental anomaly in which a gene is likely involved in the reduction of the mesoaxial fingers in the hands. Mapping studies by our group had revealed that the MSSD locus is at 17p13.3 (Malik et al. 2005). The study of the present two families (among six families) showed that

homozygous missense mutations in *BHLHA9* are associated with syndactyly type IX (MSSD).

BHLHA9 contains a single exon. BHLHA9 belongs to the basic helix-loop-helix (bHLH) proteins which act as transcription factors and regulate embryonic development. These proteins are characterized by a basic domain that binds to E-box DNA and a helix-loop-helix domain that facilitates interactions with other protein subunits. BHLHA9 further belongs to class F bHLH transcription factors and have an additional COE (collier/OLF/EBF) domain that contributes to dimerization and DNA binding (Jones, 2004). Amino acids N71 and R73, which were mutated in the studied two families, are identical in the BHLHA9 orthologues in vertebrates. This region of the transcription factor is highly conserved in evolution. With a sequence of 235 amino acids, BHLHA9 is small and contains in addition to the bHLH domain a potentially functionally relevant proline-rich carboxyl terminus.

Schatz et al. (2014) observed that human and mouse BHLHA9 proteins contained 235 and 231 amino acids, respectively. In the embryonic stages of the mouse, they observed that BHLHA9 was expressed in the limb buds in day 9.5. Also, the expression started as a crescent in the distal hand and foot plates and narrowed down towards the distal-most regions covering the developing digits.

Mutations in *BHLHA9* are known to cause other hereditary conditions besides MSSD. For instance, Klopocki et al. (2012) reported a family with split-hand/foot

malformation characterized by severe malformations of the distal limbs affecting the central rays of hands and/or feet. The affected subjects also had a long bone deficiency. Through high-resolution array comparative genomic hybridization, microduplications at 17p13.3 encompassing *BHLHA9* was detected. The authors concluded that genomic duplications encompassing *BHLHA9* are associated with SHFLD and non-Mendelian inheritance characterized by a high degree of non-penetrance with sex bias. Further, in a consanguineous Indian family with complex camptosynpolydactyly, Phadke et al. (2016) identified a homozygous missense mutation in *BHLHA9* that segregated with the phenotype. This mutation was not observed in the variant databases or ethnically matched control subjects.

In conclusion, the identification of mutations in these families verifies the crucial role of *BHLHA9* in limb development. These data would be useful for the respective families in genetic counseling and risk estimation for future pregnancies.

6.8 Conclusion

The summary of the scientific findings of this study is presented in Table 6.2.

In the present study, several families with various inherited genetic disorders were identified. In the consanguineous families, a mutation is a common recent ancestor has been inherited in a homozygous fashion because the mutation is likely not new The mutation could arise spontaneously or could be a cause of environmental insult in germ cells that was transferred to next generation. Latest molecular genetic techniques such as SNP genotyping and exome sequencing were found useful for the identification of the pathogenic mutations in these families. So, it is concluded that these techniques could help in the diagnosis of such disorders.

Table 6.2. Summary of the scientific findings of this study.

Family	Malformation	Inheritance	Localization	Mutation detected	Conclusion
Family A	Intellectual Disability with Speech Problem and Dysmorphic Facial Features (IDSD)	AR	15q15.3	*PDIA3*	PDIA3 is known to be involved in certain types of cancers. Hence, functional assays are required to prove the genetic data.
Family B	Autosomal Recessive Primary Microcephaly (MCPH)	AR	Several loci including 1q25.3-q32.1	-	*ASPM* is a likely candidate. Many mutations are known in this gene. Study was terminated
Family C	Microcephaly, Intellectual Disability, Short Stature and Brachydactyly (MIDSB)	AR	18q11.2	*RBBP8*	Results published in Am J Med Genet.
Family D	Autosomal Dominant Postaxial Polydactyly, Camptodactyly and Zygodactyly (ADPCZ)	AD	7p14.1	*GLI3*	Manuscript submitted in Eur J Med Genet.
Family E	Tooth agenesis manifested as Oligodontia/Hypodontia (TAOH)	AR	Several regions including 2q35 (STHAG4)	-	Mutation analyses in *WNT10A* did not yield any pathogenic variant. There might be an unidentified mutation in a candidate region or uncharacterized region.
Family F	Seckel-like syndrome (SLS)	AR	Several regions were homozygous	-	No mutation was detected in the gene present in the candidate regions
Family G	Syndactyly type IX (MSSD)	AR	-	*BHLHA9*	Results published in Am J Hum Genet.
Family H	Syndactyly type IX (MSSD)	AR	-	*BHLHA9*	

6.9 Future direction

A high prevalence of genetic disorders causes a great socioeconomic burden on the individual families and also as a whole on the entire economy of the country. Because a large percentage of the population live below the poverty line, the majority of suffering families are unable to afford costly treatment and rehabilitation of affected family members, and complete recovery is usually not possible. The subjects with severe types of genetic disorders remain highly disadvantaged in the society and have meager opportunities of health care, schooling, and personal development. For the betterment of those families, it is suggested that Health Department of the Ministry of Health should initiate immediate measures to prevent the recurrence of genetic disorders, such as:

- By print and electronic media special programs should be initiated for the awareness of public about the association of consanguineous marriages and hereditary genetic defects/disorders.

- Free prenatal diagnosis and screening should be provided to citizens in state hospitals by trained medical geneticists to avoid the recurrence of a disorder in the future generation.

- There are very few research institutions conducting relevant research in the country. More institution and funding should be allocated for planned research in this area. Every institution should be monitored periodically (on a yearly basis) to show that its output is directly beneficial to our public.

- At the same time, there should be international collaborations to conduct research at an advanced level so that it can be possible in the near future to establish NGS technologies in the country.

REFERENCES

Agha Z, Iqbal Z, Azam M, Siddique M, Willemsen MH, Kleefstra T, Zweier C, de Leeuw N, Qamar R, van Bokhoven H. A complex microcephaly syndrome in a Pakistani family associated with a novel missense mutation in *RBBP8* and a heterozygous deletion in *NRXN1*. Gene 2014; 538:30–35.

Ahmad W, Brancolini V, Faiyaz ul Haque M, Lam H, ul Haque S, Haider M, Maimon A, Aita VM, Owen J, Brown D, Zegarelli DJ, Ahmad M, Ott J, Christiano AM. A locus for autosomal recessive hypodontia with associated dental anomalies maps to chromosome 16q12.1. Am J Hum Genet 1998; 62:987-991

Akarsu AN, Ozbas F, Kostakoglu N. Mapping of the second locus of postaxial polydactyly type A (PAP-A2) to chromosome 13q21-q32. Am J Hum Genet 1997; 61(Suppl):A265.

Alakbarzade V, Hameed A, Quek DQ, Chioza BA, Baple EL, Cazenave-Gassiot A, Nguyen LN, Wenk MR6, Ahmad AQ, Sreekantan-Nair A, Weedon MN, Rich P, Patton MA, Warner TT, Silver DL, Crosby AH. A partially inactivating mutation in the sodium-dependent lysophosphatidylcholine transporter MFSD2A causes a non-lethal microcephaly syndrome. Nat Genet 2015; 47(7):814-7.

Al-Dosari MS, Shaheen R, Colak D, Alkuraya FS. Novel *CENPJ* mutation causes Seckel syndrome. J Med Genet 2010; 47:411-414

Angelis A, Tordrup D, Kanavos P. Socio-economic burden of rare diseases: A systematic review of cost of illness evidence. Health Policy 2015; 119(7):964-79.

APA. American Psychiatric Association. Diagnostic and statistical manual of mental disorders. 5th ed. Arlington, VA. American Psychiatric Publishing. 2013.

Awad S, Al-Dosari MS, Al-Yacoub N, Colak D, Salih MA, Alkuraya FS, Poizat C: Mutation in PHC1 implicates chromatin remodeling in primary microcephaly pathogenesis. Hum Mol Genet 2013; 22(11):2200-2213.

Baba-Kawano S, Toyoshima Y, Regalado L, Sado B, Nakasima A. Relationship between congenitally missing lower third molars and late formation of tooth germs. Angle Orthod 2002; 72(2):112-7.

Baccetti T. A controlled study of associated dental anomalies. Angle Orthod 1998; 68(3):267-74.

Bass H, Smith L, Sparkes R, Gycpes M. Case report 33. Syndrome Ident 1975; 3:12–14.

Behjati S, Tarpey PS. What is next generation sequencing? Arch Dis Child Educ Pract Ed 2013; 98(6):236-8.

Biesecker LG. Exome sequencing makes medical genomics a reality. Nat Genet 2010; 42(1):13-4.

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

Bond J, Roberts E, Springell K, Lizarraga SB, Scott S, Higgins J, Hampshire DJ, Morrison EE, Leal GF, Silva EO, Costa SM, Baralle D, Raponi M, Karbani G, Rashid Y, Jafri H, Bennett C, Corry P, Walsh CA, Woods CG. A centrosomal mechanism involving *CDK5RAP2* and *CENPJ* controls brain size. Nat Genet 2005; 37(4):353-5.

Borglum AD, Balslev T, Haagerup A, Birkebaek N, Binderup H, Kruse TA, Hertz JM. A new locus for Seckel syndrome on chromosome 18p11.31-q11.2. Eur J Hum Genet 2001; 9:753–757.

Brook AH. A unifying aetiological explanation for anomalies of human tooth number and size. Arch Oral Biol 1984; 29(5):373–8.

Buck D, Malivert L, de Chasseval R, Barraud A, Fondanèche MC, Sanal O, Plebani A, Stéphan JL, Hufnagel M, le Deist F, Fischer A, Durandy A, de Villartay JP, Revy P. Cernunnos, a novel nonhomologous end-joining factor, is mutated in human immunodeficiency with microcephaly. Cell 2006; 124(2):287-99.

Burzynski NJ, Escobar VH. Classification and genetics of numeric anomalies of dentition. Birth Defects Orig Artic Ser 1983; 19*:*95–106

Castillo V, Oñate M, Woehlbier U, Rozas P, Andreu C, Medinas D, Valdés P, Osorio F, Mercado G, Vidal RL, Kerr B, Court FA, Hetz C. Functional role of the disulfide isomerase *ERp57* in axonal regeneration. PLoS One 2015; 10(9):e0136620.

Çetinkaya M. Gene hunt in four inherited diseases. Boğaziçi University, Istanbul. PhD Thesis. 2010.

Cheng B, Dong Y, He L, Tang W, Yu H, Lu J, Xu L, Zheng B, Li K, Xiao C. Crossed polydactyly type I caused by a point mutation in the GLI3 gene in a large Chinese pedigree. J Clin Lab Anal 2006; 20:133–138.

Cheng F, Ke X, Lv M, Zhang F, Li C, Zhang X, Zhang Y, Zhao X, Wang X, Liu B, Han J, Li Y, Zeng C, Li S. A novel frame-shift mutation of GLI3 causes non-syndromic and complex digital anomalies in a Chinese family. Clin Chim Acta 2011; 412:1012–1017.

Chiurazzi P, Pirozzi F. Advances in understanding - genetic basis of intellectual disability. F1000Res 2016; 7:5.

Choi M, Scholl UI, Ji W, Liu T, Tikhonova IR, Zumbo P, Nayir A, Bakkaloglu A, Ozen S, Sanjad S, Nelson-Williams C, Farhi A, Mane S, Lifton RP. Genetic diagnosis by whole exome capture and massively parallel DNA sequencing Proc Nat Acad Sci 2009; 106(45):19096-19101.

Chong JX, Buckingham KJ, Jhangiani SN, Boehm C, Sobreira N, Smith JD, Harrell TM, McMillin MJ, Wiszniewski W, Gambin T, Coban Akdemir ZH, Doheny K, Scott AF, Avramopoulos D, Chakravarti A, Hoover-Fong J, Mathews D, Witmer PD, Ling H, Hetrick K, Watkins L, Patterson KE, Reinier F, Blue E, Muzny D, Kircher M, Bilguvar K, López-Giráldez F, Sutton VR, Tabor HK, Leal SM, Gunel M, Mane S, Gibbs RA, Boerwinkle E, Hamosh A, Shendure J, Lupski JR, Lifton RP, Valle D, Nickerson DA; Centers for Mendelian

Genomics, Bamshad MJ. The Genetic Basis of Mendelian Phenotypes: Discoveries, Challenges, and Opportunities. Am J Hum Genet 2015; 97(2):199-215.

Colella S, Yau C, Taylor JM, Mirza G, Butler H, Clouston P, Bassett AS, Seller A, Holmes CC, Ragoussis J. QuantiSNP: an Objective Bayes Hidden-Markov Model to detect and accurately map copy number variation using SNP genotyping data. Nucleic Acids Res 2007; 35(6):2013-2025.

Dauber A, Lafranchi SH, Maliga Z, Lui JC, Moon JE, McDeed C, Henke K, Zonana J, Kingman GA, Pers TH, Baron J, Rosenfeld RG, Hirschhorn JN, Harris MP, Hwa V. Novel microcephalic primordial dwarfism disorder associated with variants in the centrosomal protein ninein. J Clin Endocrinol Metab 2012; 97(11):E2140-51.

Dihazi H, Dihazi GH, Bibi A, Eltoweissy M, Mueller CA, Asif AR, Rubel D, Vasko R, Mueller GA. Secretion of ERP57 is important for extracellular matrix accumulation and progression of renal fibrosis, and is an early sign of disease onset. J Cell Sci 2013; 126(Pt 16):3649-63.

Faheem M, Naseer MI, Rasool M, Chaudhary AG, Kumosani TA, Ilyas AM, Pushparaj P, Ahmed F, Algahtani HA, Al-Qahtani MH, Saleh Jamal H. Molecular genetics of human primary microcephaly: an overview. BMC Med Genomics 2015; 8(Suppl 1):S4.

Frazier-Bowers SA, Scott MR, Cavender A, Mensah J, D'Souza RN. Mutational analysis of families affected with molar oligodontia. Connect Tissue Res 2002; 43:296–300.

Garbi N, Tanaka S, Momburg F, Hämmerling GJ. Impaired assembly of the major histocompatibility complex class I peptide-loading complex in mice deficient in the oxidoreductase ERp57. Nat Immunol 2006; 7(1):93-102.

Garib DG, Alencar BM, Ferreira FV, Ozawa TO. Associated dental anomalies: The orthodontist decoding the genetics which regulates the dental development disturbances. Dental Press J Orthod 2010; 15(2):138-157.

Garib DG, Peck S, Gomes SC. Increased occurrence of dental anomalies in patients with second premolar agenesis. Angle Orthod 2009; 79(3):436-41

Garn SM, Lewis AB. The gradient and pattern of crown size reduction in simple hypodontia. Angle Orthod 1970; 40:51–58.

Gonzalez-Perez P, Woehlbier U, Chian RJ, Sapp P, Rouleau GA, Leblond CS, Daoud H, Dion PA, Landers JE, Hetz C and Brown RH Identification of rare protein disulfide isomerase gene variants in amyotrophic lateral sclerosis patients. Gene 2015; 566(2):158-165.

Goodship J, Gill H, Carter J, Jackson A, Splitt M, Wright M. Autozygosity mapping of a Seckel syndrome locus to chromosome 3q22.1-q24. Am J Hum Genet 2000; 67:498-503.

Guernsey DL, Jiang H, Hussin J, Arnold M, Bouyakdan K, Perry S, Babineau-Sturk T, Beis J, Dumas N, Evans SC, Ferguson M, Matsuoka M, Macgillivray C, Nightingale M, Patry L, Rideout AL, Thomas A, Orr A, Hoffmann I, Michaud JL, Awadalla P, Meek DC, Ludman M, Samuels ME. Mutations in

centrosomal protein CEP152 in primary microcephaly families linked to MCPH4. Am J Hum Genet 2010; 87(1):40-51.

Han D, Gong Y, Wu H, Zhang X, Yan M, Wang X, Qu H, Feng H, Song S. Novel *EDA* mutation resulting in X-linked non-syndromic hypodontia and the pattern of *EDA*-associated isolated tooth agenesis. Eur J Med Genet 2008; 51:536-546

Harley ME, Murina O, Leitch A, Higgs MR, Bicknell LS, Yigit G, Blackford AN, Zlatanou A, Mackenzie KJ, Reddy K, Halachev M, McGlasson S, Reijns MA, Fluteau A, Martin CA, Sabbioneda S, Elcioglu NH, Altmüller J, Thiele H, Greenhalgh L0, Chessa L, Maghnie M, Salim M, Bober MB, Nürnberg P, Jackson SP, Hurles ME, Wollnik B, Stewart GS, Jackson AP. TRAIP promotes DNA damage response during genome replication and is mutated in primordial dwarfism. Nat Genet 2016; 48(1):36-43.

Hassan MJ, Chishti MS, Jamal SM, Tariq M, Ahmad W. A syndromic form of autosomal recessive congenital microcephaly (Jawad syndrome) maps to chromosome 18p11.22-q11.2. Hum Genet 2008; 123:77–82.

Hoffmann K, Lindner TH. easyLINKAGE-Plus--automated linkage analyses using large-scale SNP data. Bioinformatics 2005; 21(17):3565-3567.

Hu H, Suckow V, Musante L, Roggenkamp V, Kraemer N, Ropers HH, Hübner C, Wienker TF, Kaindl AM. Previously reported new type of autosomal recessive primary microcephaly is caused by compound heterozygous *ASPM* gene mutations. Cell Cycle 2014; 13(10):1650-1.

Hussain MS, Baig SM, Neumann S, Nürnberg G, Farooq M, Ahmad I, Alef T, Hennies HC, Technau M, Altmüller J, Frommolt P, Thiele H, Noegel AA, Nürnberg P. A truncating mutation of *CEP135* causes primary microcephaly and disturbed centrosomal function. Am J Hum Genet 2012; 90(5):871-8.

Hussain MS, Baig SM, Neumann S, Peche VS, Szczepanski S, Nürnberg G, Tariq M, Jameel M, Khan TN, Fatima A, Malik NA, Ahmad I, Altmüller J, Frommolt P, Thiele H, Höhne W, Yigit G, Wollnik B, Neubauer BA, Nürnberg P, Noegel AA. CDK6 associates with the centrosome during mitosis and is mutated in a large Pakistani family with primary microcephaly. Hum Mol Genet 2013; 22(25):5199-214.

Hussain MS, Battaglia A, Szczepanski S, Kaygusuz E, Toliat MR, Sakakibara S, Altmüller J, Thiele H, Nürnberg G, Moosa S, Yigit G, Beleggia F, Tinschert S, Clayton-Smith J, Vasudevan P11, Urquhart JE, Donnai D, Fryer A, Percin F, Brancati F, Dobbie A, Smigiel R, Gillessen-Kaesbach G, Wollnik B, Noegel AA, Newman WG, Nürnberg P. Mutations in *CKAP2L*, the human homolog of the mouse *Radmis* gene, cause Filippi syndrome. Am J Hum Genet 2014; 95(5):622-32.

Jackson AP, Eastwood H, Bell SM, Adu J, Toomes C, Carr IM, Roberts E, Hampshire DJ, Crow YJ, Mighell AJ, Karbani G, Jafri H, Rashid Y, Mueller RF, Markham AF, Woods CG. Identification of microcephalin, a protein implicated in determining thesize of the human brain. Am J Hum Genet 2002; 71(1):136-42.

Jamieson CR, Govaerts C, Abramowicz MJ. Primary autosomal recessive microcephaly: homozygosity mapping of MCPH4 to chromosome 15. Am J Hum Genet 1999; 65(5):1465-9.

Jamsheer A, Sowińska A, Trzeciak T, Jamsheer-Bratkowska M, Geppert A, Latos-Bieleńska1 A. Expanded mutational spectrum of the *GLI3* gene substantiates genotype–phenotype correlations. J Appl Genet 2012; 53:415–422.

Jimenze-Sanchez G, Childs B, Valle D. Human disease genes. Nature 2001; 409:853-855.

Johnston JJ, Olivos-Glander I, Killoran C, Elson E, Turner JT, Peters KF, Abbott MH, Aughton DJ, Aylsworth AS, Bamshad MJ, Booth C, Curry CJ, David A, Dinulos MB, Flannery DB, Fox MA, Graham JM, Grange DK, Guttmacher AE, Hannibal MC, Henn W, Hennekam RC, Holmes LB, Hoyme HE, Leppig KA, Lin AE, Macleod P, Manchester DK, Marcelis C, Mazzanti L, McCann E, McDonald MT, Mendelsohn NJ, Moeschler JB, Moghaddam B, Neri G, Newbury-Ecob R, Pagon RA, Phillips JA, Sadler LS, Stoler JM, Tilstra D, Walsh Vockley CM, Zackai EH, Zadeh TM, Brueton L, Black GC, Biesecker LG. Molecular and clinical analyses of Greig cephalopolysyndactyly and Pallister–Hall syndromes: robust phenotype prediction from the type and position of *GLI3* mutations. Am J Hum Genet 2005; 76:609–622.

Jones S. An overview of the basic helix-loop-helix proteins. Genome Biol 2004; 5(6):226.

Kaindl AM. Autosomal recessive primary microcephalies (MCPH). Eur J Paediatr Neurol 2014; 18(4):547-8.

Kalay E, Yigit G, Aslan Y, Brown KE, Pohl E, Bicknell LS, Kayserili H, Li Y, Tüysüz B, Nürnberg G, Kiess W, Koegl M, Baessmann I, Buruk K, Toraman B, Kayipmaz S, Kul S, Ikbal M, Turner DJ, Taylor MS, Aerts J, Scott C, Milstein K, Dollfus H, Wieczorek D, Brunner HG, Hurles M, Jackson AP, Rauch A, Nürnberg P, Karagüzel A, Wollnik B. CEP152 is a genome maintenance protein disrupted in Seckel syndrome. Nat Genet 2011; 43(1):23-6.

Kaufman L, Ayub M, Vincent JB. The genetic basis of non-syndromic intellectual disability: a review. J Neurodev Disord 2010; 2(4):182-209.

Kavaslar GN, Onengüt S, Derman O, Kaya A, Tolun A. The novel genetic disorder microhydranencephaly maps to chromosome 16p13.3-12.1. Am J Hum Genet 2000; 66(5):1705-9.

Khan MA, Rupp VM, Orpinell M, Hussain MS, Altmüller J, Steinmetz MO, Enzinger C, Thiele H, Höhne W, Nürnberg G, Baig SM, Ansar M, Nürnberg P, Vincent JB, Speicher MR, Gönczy P, Windpassinger C. A missense mutation in the PISA domain of HsSAS-6 causes autosomal recessive primary microcephaly in a large consanguineous Pakistani family. Hum Mol Genet 2014; 23(22):5940-9.

Kilinc MO, Ninis VN, Ugur SA, Tuysuz B, Seven M, Balci S, Goodship J, Tolun A. Is the novel *SCKL3* at 14q23 the predominant Seckel locus? Eur J Hum Genet 2003; 11:851-857.

Kim JW, Simmer JP, Lin BP, Hu JC. Novel *MSX1* frameshift causes autosomal-dominant oligodontia. J Dent Res 2006; 85:267–271.

Klopocki E1, Lohan S, Doelken SC, Stricker S, Ockeloen CW, Soares Thiele de Aguiar R, Lezirovitz K, Mingroni Netto RC, Jamsheer A, Shah H, Kurth I,

Habenicht R, Warman M, Devriendt K, Kordass U, Hempel M, Rajab A, Mäkitie O, Naveed M, Radhakrishna U, Antonarakis SE, Horn D, Mundlos S. Duplications of BHLHA9 are associated with ectrodactyly and tibia hemimelia inherited in non-Mendelian fashion. J Med Genet 2012; 49(2):119-25.

Ku CS, Wu M, Cooper DN, Naidoo N, Pawitan Y, Pang B, Iacopetta B, Soong R. Exome versus transcriptome sequencing in identifying coding region variants. Expert Rev Mol Diagn 2012; 12(3):241-51.

Kumar A, Girimaji SC, Duvvari MR, Blanton SH. Mutations in STIL, encoding a pericentriolar and centrosomal protein, cause primary microcephaly. Am J Hum Genet 2009; 84(2):286-90.

Lalonde E, Albrecht S, Ha KC, Jacob K, Bolduc N, Polychronakos C, Dechelotte P, Majewski J, Jabado N. Unexpected allelic heterogeneity and spectrum of mutations in Fowler syndrome revealed by next-generation exome sequencing. Hum Mutat 2010; 31(8):918-23.

Lammi L, Arte S, Somer M, Jarvinen H, Lahermo P, Thesleff I. Mutations in *AXIN2* cause familial tooth agenesis and predispose to colorectal cancer. Am J Hum Genet 2004; 74:1043–1050.

Lander ES, Botstein D. Homozygosity Mapping: A way to map human recessive traits with the DNA of inbred children. Science 1987; 236(4808):1567-1570.

Leutenegger AL, Labalme A, Genin E, Toutain A, Steichen E, Clerget-Darpoux F, Edery P. Using genomic inbreeding coefficient estimates for homozygosity mapping of rare recessive traits: application to Taybi-Linder syndrome. Am J Hum Genet 2006; 79:62-66.

Lexner MO, Bardow A, Hertz JM, Nielsen LA, Kreiborg S. Anomalies of tooth formation in hypohidrotic ectodermal dysplasia. Int J Paediatr Dent 2007; 17:10–18

Liu W, Wang H, Zhao S, Zhao W, Bai S, Zhao Y, Xu S, Wu C, Huang W, Chen Z, Feng G, He L. The novel gene locus for agenesis of permanent teeth maps to chromosome 10q11.2. J Dent Res 2001; 80:1716-1720

Majewski F, Goecke T. Studies of microcephalic primordial dwarfism. Approach to a delineation of the Seckel syndrome. Am J Med Genet 1982; 12: 7–21

Majewski F, Ranke M, Schinzel A. Studies of microcephalic primordial dwarfism II: the osteodysplastic type II of primordial dwarfism. Am J Med Genet 1982a; 12:23–35

Majewski F, Spranger J. A new (brachymelic) type of primordial dwarfism. Monatsschr Kinderheilkd 1976; 124:499–503

Majewski F, Stoeckenius M, Kemperdick H. Studies of microcephalic primordial dwarfism. III. An intrauterine dwarf with platyspondyly and anomalies of pelvis and clavicles—osteodysplastic primordial dwarfism type III. Am J Med Genet 1982b; 12:37–42

Malik S, Ahmad W, Grzeschik K-H, Koch MC. A simple method for characterising syndactyly in clinical practice. Genet Couns 2005a; 16:229-238.

Malik S, Arshad M, Amin-Ud-Din M, Oeffner F, Dempfle A, Haque S, Koch MC, Ahmad W, Grzeschik KH. A novel type of autosomal recessive syndactyly: clinical and molecular studies in a family of Pakistani origin. Am J Med Genet A 2004; 126A:61-67.

Malik S, Grzeschik KH: Synpolydactyly: clinical and molecular advances. Clin Genet 2008; 73:113–120.

Malik S, Percin FE, Ahmad W, Percin S, Akarsu NA, Koch MC, Grzeschik KH. Autosomal recessive mesoaxial synostotic syndactyly with phalangeal reduction maps to chromosome 17p13.3. Am J Med Genet A 2005b; 134:404-408.

Malik S, Ullah S, Afzal M, Lal K, Haque S. Clinical and descriptive genetic study of polydactyly a Pakistani experience of 313 cases/families. Clin Genet 2014; 85:482–486.

Malik S. Polydactyly: phenotypes, genetics and classification. Clin Genet 2014; 85:203–212.

Malik S. Syndactyly: phenotypes, genetics and current classification. Eur J Hum Genet 2012; 20:817-824.

Malik S, Percin FE, Bornholdt D, Albrecht B, Percesepe A, Koch MC, Landi A, Fritz B, Khan R, Mumtaz S, Akarsu NA, Grzeschik KH. Mutations affecting the BHLHA9 DNA-binding domain cause MSSD, mesoaxial synostotic syndactyly with phalangeal reduction, Malik-Percin type. Am J Hum Genet 2014; 95(6):649-59.

Massink MP, Créton MA, Spanevello F, Fennis WM, Cune MS, Savelberg SM, Nijman IJ, Maurice MM, van den Boogaard MJ, van Haaften G. Loss-of-function mutations in the WNT co-receptor LRP6 cause autosomal-dominant oligodontia. Am J Hum Genet 2015; 97(4):621-6.

Massink MP, Créton MA, Spanevello F, Fennis WM, Cune MS, Savelberg SM, Nijman IJ, Maurice MM, van den Boogaard MJ, van Haaften G. Loss-of-Function Mutations in the WNT Co-receptor LRP6 Cause Autosomal-Dominant Oligodontia. Am J Hum Genet 2015; 97(4):621-6.

McCarroll SA, Kuruvilla FG, Korn JM, Cawley S, Nemesh J, Wysoker A, Shapero MH, de Bakker PI, Maller JB, Kirby A, Elliott AL, Parkin M, Hubbell E, Webster T, Mei R, Veitch J, Collins PJ, Handsaker R, Lincoln S, Nizzari M, Blume J, Jones KW, Rava R, Daly MJ, Gabriel SB, Altshuler D. Integrated detection and population-genetic analysis of SNPs and copy number variation. Nat Genet 2008; 40(10):1166-74.

McKusick VA, Mahloudji M, Abboptt MH, Lindenberg R, Kepas D. Seckel's bird-headed dwarfism. N Engl J Med 1967; 86:277-279.

Middeldorp A, Jagmohan-Changur S, Helmer Q, van der Klift HM, Tops CM, Vasen HF, Devilee P, Morreau H, Houwing-Duistermaat JJ, Wijnen JT, van Wezel T. A procedure for the detection of linkage with high density SNP arrays in a large pedigree with colorectal cancer. BMC Cancer 2007; 7:6.

Middeldorp A1, Jagmohan-Changur S, Helmer Q, van der Klift HM, Tops CM, Vasen HF, Devilee P, Morreau H, Houwing-Duistermaat JJ, Wijnen JT, van Wezel

T. A procedure for the detection of linkage with high density SNP arrays in a large pedigree with colorectal cancer. BMC Cancer 2007; 7:6.

Mirzaa GM, Vitre B, Carpenter G, Abramowicz I, Gleeson JG, Paciorkowski AR, Cleveland DW, Dobyns WB, O'Driscoll M. Mutations in CENPE define a novel kinetochore-centromeric mechanism for microcephalic primordial dwarfism. Hum Genet 2014; 133(8):1023-39.

Morton NE. Sequential tests for the detection of linkage. Am J Hum Genet 1995; 7(3):277-318.

Mossey PA. The heritability of malocclusion: part 2. The influence of genetics in malocclusion. Br J Orthod 1999; 26(3):195-203

Mumtaz S, Yıldız E, Jabeen S, Khan A, Tolun A, Malik S. *RBBP8* syndrome with microcephaly, intellectual disability, short stature and brachydactyly. Am J Med Genet A 2015; 167A(12):3148-52.

Najmabadi H, Hu H, Garshasbi M, Zemojtel T, Abedini SS, Chen W, Hosseini M, Behjati F, Haas S, Jamali P, Zecha A, Mohseni M, Püttmann L, Vahid LN, Jensen C, Moheb LA, Bienek M, Larti F, Mueller I, Weissmann R, Darvish H, Wrogemann K, Hadavi V, Lipkowitz B, Esmaeeli-Nieh S, Wieczorek D, Kariminejad R, Firouzabadi SG, Cohen M, Fattahi Z, Rost I, Mojahedi F, Hertzberg C, Dehghan A, Rajab A, Banavandi MJ, Hoffer J, Falah M, Musante L, Kalscheuer V, Ullmann R, Kuss AW, Tzschach A, Kahrizi K, Ropers HH. Deep sequencing reveals 50 novel genes for recessive cognitive disorders. Nature 2011; 478(7367):57-63.

Ng SB, Buckingham KJ, Lee C, Bigham AW, Tabor HK, Dent KM, Huff CD, Shannon PT, Jabs EW, Nickerson DA, Shendure J, Bamshad MJ. Exome sequencing identifies the cause of a Mendelian disorder. Nat Genet 2010; 42(1):30-55.

Nicholas AK, Khurshid M, Désir J, Carvalho OP, Cox JJ, Thornton G, Kausar R, Ansar M, Ahmad W, Verloes A, Passemard S, Misson JP, Lindsay S, Gergely F, Dobyns WB, Roberts E, Abramowicz M, Woods CG. *WDR62* is associated with the spindle pole and is mutated in human microcephaly. Nat Genet 2010; 42(11):1010-4.

Nieminen P, Arte S, Pirinen S, Peltonen L, Thesleff I. Gene defect in hypodontia: exclusion of *MSX1* and *MSX2* as candidate genes. Hum Genet 1995; 96:305-308

Noor A, Windpassinger C, Vitcu I, Orlic M, Rafiq MA, Khalid M, Malik MN, Ayub M, Alman B, Vincent JB. Oligodontia is caused by mutation in *LTBP3*, the gene encoding latent TGF-beta binding protein 3. Am J Hum Genet 2009; 84:519-523

OMIM. Online Mendelian Inheritance in Man, OMIM®. McKusick-Nathans Institute of Genetic Medicine, Johns Hopkins University, Baltimore, MD. 2016. http://omim.org/

Orioli IM, Castilla EE. Thumb/hallux duplication and preaxial polydactyly type I. Am J Med Genet 1999; 82:219-224.

Ott J, Wang J, Leal SM. Genetic linkage analysis in the age of whole-genome sequencing. Nat Rev Genet 2015; 16(5):275-84.

Pareek CS, Smoczynski R, Tretyn A. Sequencing technologies and genome sequencing. J Appl Genet 2011; 52(4):413-35.

Peck L, Peck S, Attia Y. Maxillary canine-first premolar transposition, associated dental anomalies and genetic basis. Angle Orthod 1993; 63(2):99-109.

Peck S, Peck L, Kataja M. Concomitant occurrence of canine malposition and tooth agenesis: evidence of orofacial genetic fields. Am J Orthod Dentofacial Orthop 2002; 122(6):657–6.

Peltonen L, Palotie A, Lange K. Use of population isolates for mapping complex traits. Nat Rev Genet 2000; 1(3):182-90.

Pemberton TJ, Das P, Patel PI. Hypodontia: genetics and future perspectives. Braz J Oral Sci 2005; 4:695–706.

Percin EF, Percin S, Egilmez H, Sezgin I, Ozbas F, Akarsu AN. Mesoaxial complete syndactyly and synostosis with hypoplastic thumbs: an unusual combination or homozygous expression of syndactyly type I? J Med Genet 1998; 35:868-874.

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

Phadke SR, Kar A, Bhowmik AD, Dalal A. Complex Camptosynpolydactyly and Mesoaxial synostotic syndactyly with phalangeal reduction are allelic disorders. Am J Med Genet A 2016; 170(6):1622-5.

Polder BJ, Vant Hof MA, Van der Linden FP, Kuijpers-Jagtman AM. A meta-analysis of the prevalence of dental agenesis of permanent teeth. Community Dent Oral Epidemiol 2004; 32:217–226.

Qvist P, Huertas P, Jimeno S, Nyegaard M, Hassan MJ, Jackson SP, Børglum AD. *CtIP* mutations cause Seckel and Jawad syndromes. PLoS Genet 2011; 7:e1002310.

Radhakrishna U, Bornholdt D, Scott HS, Patel UC, Rossier C, Engel H, Bottani A, Chandal D, Blouin JL, Solanki JV, Grzeschik KH, Antonarakis SE. The phenotypic spectrum of GLI3 morphopathies includes autosomal dominant preaxial polydactyly type-IV and postaxial polydactyly type-A/B; No phenotype prediction from the position of GLI3 mutations. Am J Hum Genet 1999; 65:645–655.

Radhakrishna U, Wild A, Grzeschik KH, Antonarakis SE. Mutation in *GLI3* in postaxial polydactyly type A. Nat Genet 1997; 17:269–271.

Rauch A, Thiel CT, Schindler D, Wick U, Crow YJ, Ekici AB, van Essen AJ, Goecke TO, Al-Gazali L, Chrzanowska KH, Zweier C, Brunner HG. Mutations in the pericentrin (*PCNT*) gene cause primordial dwarfism. Science 2008; 319:816-819.

Risch N. Genetic Linkage: Interpreting Lod Scores. Science 1992; 255(5046):803-804.

Ropers HH. Genetics of early onset cognitive impairment. Annu Rev Genomics Hum Genet 2010; 11:161-87.

Santana-Codina N, Carretero R, Sanz-Pamplona R, Cabrera T, Guney E, Oliva B, Clezardin P, Olarte OE, Loza-Alvarez P, Méndez-Lucas A, Perales JC, Sierra

A. A transcriptome-proteome integrated network identifies endoplasmic reticulum thiol oxidoreductase (*ERp57*) as a hub that mediates bone metastasis. Mol Cell Proteomics 2013; 12(8):2111-25

Sartori AA, Lukas C, Coates J, Mistrik M, Fu S, Bartek J, Baer R, Lukas J, Jackson SP. Human CtIP promotes DNA end resection. Nature 2007; 450:509–514.

Scarel RM, Trevilatto PC, di Hipolito OJ, Camargo LE and Line SR. Absence of mutations in the homeodomain of the MSX1 gene in patients with hypodontia. Am J Med Genet 2000; 92:346–349.

Schatz O, Langer E, Ben-Arie N. Gene dosage of the transcription factor Fingerin (bHLHA9) affects digit development and links syndactyly to ectrodactyly. Hum Mol Genet 2014; 23(20):5394-401.

Seckel HPG. Bird-Headed Dwarfs. Studies in Developmental Anthropology Including Human Dimensions. Springfield, I11. Charles C Thomas. 1960.

Sener EF, Canatan H, Ozkul Y. Recent advances in autism spectrum disorders: applications of whole exome sequencing technology. Psychiatry Investing 2016; 13(3):255-64.

Shaheen R, Faqeih E, Ansari S, Abdel-Salam G, Al-Hassnan ZN, Al-Shidi T, Alomar R, Sogaty S, Alkuraya FS. Genomic analysis of primordial dwarfism reveals novel disease genes. Genome Res 2014; 24:291–299.

Shaheen R, Hashem A, Abdel-Salam GM, Al-Fadhli F, Ewida N, Alkuraya FS. Mutations in *CIT*, encoding citron rho-interacting serine/threonine kinase, cause severe primary microcephaly in humans. Hum Genet 2016; 135(10):1191-7.

Shelby MD, Bishop JB, Mason JM, Tindall KR. Fertility, reproduction, and genetic disease: studies on the mutagenic effects of environmental agents on mammalian germ cells. Environ Health Perspect 1993; 100:283-291

Shwayder T. Pediatic nail disease. Semin Dermatol 1995; 14:21–26

Sir JH, Barr AR, Nicholas AK, Carvalho OP, Khurshid M, Sossick A, Reichelt S, D'Santos C, Woods CG, Gergely F. A primary microcephaly protein complex forms a ring around parental centrioles. Nat Genet 2011; 43(11):1147-53.

Song S, Han D, Qu H, Gong Y, Wu H, Zhang X, Zhong N, Feng H. *EDA* gene mutations underlie non-syndromic oligodontia. J Dent Res 2009; 88:126-130

Stockton DW, Das P, Goldenberg M, D'Souza RN, Patel PI. Mutation of *PAX9* is associated with oligodontia. Nat Genet 2000; 24:18–19

Strachan T, Read A. Human Molecular Genetics 3. Garland Publishing. 2004.

Suarez BK, Spence MA. The genetics of hypodontia. J Dent Res 1974; 53(4):781–5

Temtamy SA, McKusick VA: Polydactyly. In: The genetics of hand malformations. New York, NY: Alan R Liss. 1978; 364–439.

Terwilliger JD, Ott J. Handbook of Human Genetic Linkage, John Hopkins University Press, Baltimore. 1994.

Vastardis H, Karimbux N, Guthua SW, Seidman JG, Seidman CE. A human *MSX1* homeodomain missense mutation causes selective tooth agenesis. Nat Genet 1996; 13:417–421

Vissers LE, Gilissen C, Veltman JA. Genetic studies in intellectual disability and related disorders. Nat Rev Genet 2016; 17(1):9-18

von Bubnoff A. Next-generation sequencing: the race is on. Cell 2008; 132:721–723.

Vortkamp A, Gessler M, Grzeschik KH. Identification of optimized target sequences for the GLI3 zinc finger protein.DNA Cell Biol 1995; 14(7):629-34.

Wahab A, Ahmad M. Biosocial perspective of consanguineous marriages in rural and urban Swat, Pakistan. J Biosoc Sci 1996; 28(3):305-313.

Wang K, Li M, Hadley D, Liu R, Glessner J, Grant SF, Hakonarson H, Bucan M. PennCNV: an integrated hidden Markov model designed for high-resolution copy number variation detection in whole-genome SNP genotyping data. Genome Res 2007; 17(11):1665-74.

Wang Z, Wang J, Li Y, Geng J, Fu Q, Xu Y, Shen Y. Novel frame-shift mutations of GLI3 gene in non-syndromic postaxial polydactyly patients. Clin Chim Acta 2014; 433:195–199.

Warr A, Robert C, Hume D, Archibald A, Deeb N, Watson M. Exome Sequencing: Current and Future Perspectives. G3 (Bethesda) 2015; 5(8):1543-50.

WHO. World Health Organization. Control of genetic diseases. Geneva 2005; 55-77.

Wieczorek D, Pawlik B, Li Y, Akarsu NA, Caliebe A, May KJ, Schweiger B, Vargas FR, Balci S, Gillessen-Kaesbach G, Wollnik B. A specific mutation in the distant sonic hedgehog (SHH) cis-regulator (ZRS) causes Werner mesomelic syndrome (WMS) while complete ZRS duplications underlie Haas type polysyndactyly and preaxial polydactyly (PPD) with or without triphalangeal thumb. Hum Mutat 2010; 31:81–89.

Woods CG, Basto R. Microcephaly. Curr Biol 2014; 24(23):1109-11.

Woods CG, Parker A. Investigating microcephaly. Arch Dis Child 2013; 98(9):707-13.

Yamamoto S, Jaiswal M, Charng WL, Gambin T, Karaca E, Mirzaa G, Wiszniewski W, Sandoval H, Haelterman NA, Xiong B, Zhang K, Bayat V, David G, Li T, Chen K, Gala U, Harel T, Pehlivan D, Penney S, Vissers LE, de Ligt J, Jhangiani SN, Xie Y, Tsang SH, Parman Y, Sivaci M, Battaloglu E, Muzny D, Wan YW, Liu Z, Lin-Moore AT, Clark RD, Curry CJ, Link N, Schulze KL, Boerwinkle E, Dobyns WB, Allikmets R, Gibbs RA, Chen R, Lupski JR, Wangler MF, Bellen HJ. A drosophila genetic resource of mutants to study mechanisms underlying human genetic diseases. Cell 2014; 159(1):200-14.

Yang YJ, Baltus AE, Mathew RS, Murphy EA, Evrony GD, Gonzalez DM, Wang EP, Marshall-Walker CA, Barry BJ, Murn J, Tatarakis A, Mahajan MA, Samuels HH, Shi Y, Golden JA, Mahajnah M, Shenhav R, Walsh CA. Microcephaly gene links trithorax and *REST/NRSF* to control neural stem cell proliferation and differentiation. Cell 2012; 151(5):1097-1112.

Yu P, Yang W, Han D, Wang X, Guo S, Li J, Li F, Zhang X, Wong SW, Bai B, Liu Y, Du J, Sun ZS, Shi S, Feng H, Cai T. Mutations in *WNT10B* are identified in individuals with oligodontia. Am J Hum Genet 2016; 99(1):195-201.

Zguricas J, Heus H, Morales-Peralta E, Breedveld G, Kuyt B, Mumcu EF, Bakker W, Akarsu N, Kay SP, Hovius SE, Heredero-Baute L, Oostra BA, Heutink P.

Clinical and genetic studies on 12 preaxial polydactyly families and refinement of the localisation of the gene responsible to a 1.9 cM region on chromosome 7q36. J Med Genet 1999; 36(1):32-40.

Zhao H, Tian Y, Breedveld G, Huang S, Zou Y, Jue Y, Chai J, Li H, Li M, Oostra BA, Lo WH, Heutink P. Postaxial polydactyly type A/B (PAP-A/B) is linked to chromosome 19p13.1-13.2 in a Chinese kindred. Eur J Hum Genet 2002; 10:162–166.

Thesis

ORIGINALITY REPORT

%8	%6	%5	%2
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	www.ncbi.nlm.nih.gov Internet Source	%2
2	jdr.sagepub.com Internet Source	%1
3	Nomura, S.. "Development of a novel nano-Invader DNA chip system", Journal of Biochemical and Biophysical Methods, 20070801 Publication	<%1
4	www.imfedk.org Internet Source	<%1
5	Submitted to Yonsei University Student Paper	<%1
6	Le Hellard, S.. "Haplotype Analysis and a Novel Allele-Sharing Method Refines a Chromosome 4p Locus Linked to Bipolar Affective Disorder", Biological Psychiatry, 20070315 Publication	<%1
7	fodbold.ru Internet Source	<%1