Genetic epidemiology, mapping and molecular characterization of families with inherited skeletal and neurological disorders

By

Anisa Bibi

Human Genetics Lab

Department of Zoology

Faculty of Biological Sciences

Quaid-i-Azam University, Islamabad, Pakistan

2022

Genetic epidemiology, mapping and molecular characterization of families with inherited skeletal and neurological disorders

PhD Dissertation

A dissertation submitted in partial fulfilment of requirements for degree of Doctor of Philosophy in Human Genetics

Supervised by:

Professor Dr. Sajid Malik

Conducted by:

Anisa Bibi

Human Genetics Lab

Department of Zoology

Faculty of Biological Sciences

Quaid-i-Azam University, Islamabad, Pakistan

2022

Author's Declaration

I **Ms. Anisa Bibi** hereby state that my PhD thesis titled "Genetic Epidemiology, Mapping and Molecular Characterization of Families with Inherited Skeletal and Neurological Disorders" is my own work and has not been submitted previously by me for taking any degree from Quaid-i-Azam University, Islamabad, Pakistan.

At any time if my statement is found to be incorrect even after my Graduate the University has the right to withdraw my Ph.D. degree.

Ms. Anisa Bibi

Certificate of Approval

This is to certify that the research work presented in this thesis, entitled "Genetic Epidemiology, Mapping and Molecular Characterization of Families with Inherited Skeletal and Neurological Disorders" was conducted by **Ms. Anisa Bibi** under the supervision of **Prof. Dr. Sajid Malik**. No part of this thesis has been submitted anywhere else for any other degree. This thesis is submitted to the Department of Zoology of Quaid-i-Azam University, Islamabad in partial fulfillment of the requirements for the degree of Doctor of Philosophy in Field of Human Genetics.

Student Name: **Ms. Anisa Bibi** Signature:_____

Examination Committee:

a) External Examiner 1: Signature:_____

1. **Dr. Mazhar Qayyum**
   Professor (Retd.)
   Department of Zoology
   PMAS Arid Agriculture University,
   Rawalpindi

b) External Examiner 2: Signature:_____

2. **Dr. Kehkeshan Mazhar**
   Principle Scientific Officer
   Institute of Biomedical and Genetic Engineering (IBGE)
   24-Mauve Area, G-9/1, Islamabad

3. **Prof. Dr. Sajid Malik**
   Research Supervisor

Name of HOD: **Prof. Dr. Amina Zuberi** Signature:_____

Date: 25.10.2023

CHAIRPERSON
Department of Zoology
Quaid-i-Azam University
Islamabad.

DEPARTMENT OF ZOOLOGY
ISLAMABAD
DATE 26/10/23

Plagiarism Undertaking

I solemnly declare that research work presented in the thesis "Genetic Epidemiology, Mapping and Molecular Characterization of Families with Inherited Skeletal and Neurological Disorders" is solely my research work with no significant contribution from any other person. Small contribution/ help wherever taken has been duly acknowledged and that complete thesis has been written by me.

I understand the zero tolerance policy of the HEC and Quaid-i-Azam University towards plagiarism. Therefore I as an Author of the above titled thesis declare that no portion of my thesis has been plagiarized and any material used as reference is properly referred/ cited.

I undertake that if I am found guilty of any formal plagiarism in the above titled thesis even after award of Ph.D degree, HEC and the University has the right to publish my name on the HEC/University Website on which names of students are placed who submitted plagiarized thesis.

Student / Author Signature: _____

Name: **Ms. Anisa Bibi**

Dedicated To

Allah Almighty,

Holy Prophet Muhammad ﷺ

My Loving Parents, Teachers

&

Family Members

Acknowledgment

I offer my humblest and sincere thanks to Almighty ALLAH, Who bestowed me with potential and ability to make a solid contribution to already existing ocean of knowledge and for the Holy Prophet Hazrat Muhammad (PBUH), Who showed us the right path and enabled us to recognize our Creator.

This work would not have been possible without the support and encouragement of my teacher **Prof. Dr. Sajid Malik**, under whose supervision I chose this topic and began the thesis. When I registered at QAU, I was nothing but today I am able to submit this thesis. He always has been a continuous source of encouragement for me.

I am grateful to Dean Faculty of Biological Sciences; **Prof. Dr. Sarwat Jehan** and Chairperson Department of Zoology; **Prof. Dr. Amina Zuberi** for providing pre-requisites for this work and thus facilitating this task.

My sincere gratitude goes out to **Prof. Dr. Mustafa Khokha, Dr. Saquib Lakhani, Mis. Emily**, and the team at Yale University in New Haven, Connecticut, USA, for providing me with training in the current state of clinical and functional investigation using the model organism *X. tropicalis*. Ms. Emily guided me throughout my stay in USA and provided substantial contribution to my research. She is unique in her scientific understanding, support, and work ethics. I also want to express my gratitude to the Higher Education Commission of Pakistan for supporting my trip under the IRSIP program. I am highly indebted to our collaborator Prof. Dr. Aslihan Tolun and her entire team at the Molecular Genetics Laboratory, Istanbul Technical University, Istanbul, Turkey.

This study would have been impossible without all the families and subjects who gave their consent and participated in the study. I am also thankful to all those people who helped and facilitated me during the field work.

Clerical and technical assistant in scientific research is an undeniable important element. So, I am righteous in thanking Syed Mujahid Hussain Shah (Lab. assistant), Naeem Masih (Senior Clerk) and Samiullah (Junior Clerk) for their thorough and in time assistance.

I am highly thankful to my senior Dr. Rana Kamran Shabir for his help in terms of polite behavior, positive criticism and guidance whenever I need at any time at any ocCAion during the field work and lab work and he never disappointed me. Special thanks also to my friend Dr. Misbah Rasheed for her continuous support and motivation and to Dr. Sara Mumtaz for her help to become an independent researcher. Due to her moral support I come to know about my true self. My PhD and M.Phil lab fellows provided much moral support and encouragement and I remain grateful for their friendship and time spent together.

In the end I owe my gratitude to my parents and their unconditional love, prayers and encouragement. I achieved this academic task only because of their efforts and the sacrifices they made for me.

List of Tables

List of figures

List of Acronyms

Abbreviation	Elaboration
ABD	Actin-binding domains
ABE	Adenine base editor
AJK	Azad Jamun and Kashmir
bp	Base pairs
BLAST	Basic Local Alignment Search Tool
BAM	Binary Alignment Map
BD	Boomerang dysplasia
BWA	Burrows-Wheeler Aligner
CHD1	Calponin homology domain1
CHD2	Calponin homology domain2
CVS	Chorionic villus sampling
CFAP46	Cilia And Flagella Associated Protein 46
CRISPR	Clustered regularly interspaced short palindromic repeats
CI	Coefficient of inbreeding
CT scan	Computerized tomography scan
CA	Congenital abnormalities
CAKUT	Congenital anomalies of the kidney and urinary tract
CLD	Congenital limb deficiency
CIOMS	Council for International Organizations of Medical Sciences
CA9	CRISPR-associated protein 9
CBE	Cytidine base editor
DNA	Deoxy ribose nucleic acid
DSBs	Double-stranded breaks

ES	Exome sequencing
FRAS1	Extracellular Matrix Complex Subunit
FATA	Federally Administered Tribal Areas
GSDs	Genetic skeletal dysplasia conditions
GE	Genome editing
GATK	GenomeAnalysis Toolkit
GWAS	Genome-wide association studies
gDNA	Genomic DNA
GRIP1	Glutamate Receptor Interacting Protein 1
HMCN1	Hemicentin1
HDR	Homology-directed repair
HH	Homozygosity haplotype
HM	Homozygosity mapping
IBD	Identical-by-descent
ICE	Inference of CRISPR edit
UIC	Injected control
Indels	Insertions/or deletions
IDT	Integrated DNA Technologies
ID	Intellectual disability
ICD-10	International Classification of Diseases, Tenth Revision
IOD	Intraocular distance
KP	Khyber Pakhtunkhwa
KO	Knockout
MRI	Magnetic resonance imaging
MPS	Multiple pterygium syndrome

NGS	Next generation sequencing
NHEJ	Non-homology end-joining
OMIM	Online Mendelian Inheritance in Man
PCR	Polymerase chain reaction
PAM	Protospacer Adjacent Motif
RAB33B	RAS-ASSOCIATED PROTEIN RAB33B
RCHH	Region of conserved homozygosity
SAM	Sequence Alignment/Map
sgRNA	Single guide RNA
SNP	Single nucleotide polymorphism
Sox9	SRY-Box Transcription Factor 9
STROBE	Strengthening the Reporting of Observational Studies in Epidemiology
TALENs	Transcriptional activator-like effector nucleases
Twist1	Twist-related protein 1
WGS	Whole-Genome Sequencing
WHO	World Health Organization
ZFNs	Zinc finger nucleases

TABLE OF CONTECT

Abstract

Congenital and hereditary anomalies (CA) are common in Pakistan and put a tremendous impact on the morbidity and mortality of infant, young and adult populations. The present study aimed to investigate the epidemiological, clincal and molecular genetic characteristics of Pakistani families affected with CA. This multitiered study was carried out in order to: 1) estimate the burden of CA in the Hazara population of Pakistan by determining their prevalence-pattern and bio-demographic coorelates; 2) establish the prevalence-pattern, phenotypic manifestation and descriptive genetics of congenital limb deficiencies; 3) clinical and molecular diagnosis of rare malformations segregating in extended families; and 4) provide clinical and functional evidence through animal model that rare variant is causative for craniofacial anomalies. First, through a mix of modified random door-to-door and convenience sampling approach, 1,189 independent cases/families with CA were identified in the Hazara population (Chapter 2). The estimated consanguinity rate in this sample was 66%. An estimated 65% cases were sporadic in nature and 35% familial; and 70% were isolated and 30% syndromic. CA were categorized into 9 major phenotypic types. Neurological disorders were the most common (n=486; proportion=0.409;% CI=0.381-0.437), followed by limb defects (n=292), musculoskeletal defects (n=106), sensorineural/ear defects (n=101), blood disorders (n=75), eye/visual impairments (n=39), ectodermal anomalies (n=30), and congenital heart defects (n=26). This is the first study of its kind reporting prevalence-pattern of CA in Hazara population of Pakistan. Congenital limb deficiencies (CLD) are very rare anomalies and cause severe disability. In another study (Chapter 3), 141 independent subjects involving 166 limbs with CLD were recruited from Khyber Pakhtunkhwa (KP) province of Pakistan. Subjects with transverse defects were n=77 (55%), longitudinal defects n=61 (43%), and intercalary defects n=3 (2%). Upper limb deficiencies were 86% compared to lower limb deficiencies, i.e., 14%. There was a high preponderance of unilateral cases (83%), sporadic nature (92%), low parental consanguinity rate (33%), and syndromic appearance in 21% of cases, which may be indicative of the substantial role of non-genetic factors in the etiology of CLD. This is the first study reporting phenotypic pattern of CLD from Pakistan. Increased inbreeding is a common cause of the recessive mutations that lead to the appearance of rare disorders in Pakistani families. Another aim of the current research

was to locate the causal mutations in Pakistani families associated with autosomal recessive malformations. Here, four families with rare hereditary skeletal, neurological and ophthalmological disorders were recruited. DNA of the available affected and unaffected subjects was extracted from peripheral blood for molecular study. Whole genome SNP genotyping and/or exome sequencing were carried to in order to find the candidate chromosomal region/variant. (i) In the first family (Chapter 4), the patients were presented with intellectual disability, developmental delay, orofacial defects and a range of behavioral anomalies. In this family, SNP based homozygosity mapping led to the discovery of 12 homozygous intervals >1Mb, the largest (54Mb) on chromosome 5p15.2-q12.3, followed by 39Mb stretch on 3q23-q26.32 and 22Mb on 4q31.22-q32.3. Several potential candidate genes related to the phenotype were present on these regions. These findings would be very helpful in a prospective gene hunt through exome sequencing. (ii) In the second family (Chapter 5), a 13-year old boy was presented with cryptophthalmia, midface hypoplasia, agenesis of right kidney, and cutaneous syndactyly in fingers and toes but no symptoms in any other organ including lungs, anorectal system, genitalia or umbilical system were observed. Exome sequencing in this family led to the discovery of homozygous truncating variant c.1774C>T (p.Gln592Ter) in *GRIP1* gene which is known to cause Fraser syndrome 3 (FRASRS3). This case is also the oldest reported individual with FRASRS3, to our knowledge, and shows that FRASRS3 case may be milder than known and live into at least adolescence. (iii) In family 3 (Chapter 6), the patients were presented with rhizomelia, short trunk dwarfism, scoliosis, microcephaly, intellectual disability, and other clinically diverse symptoms. Exome sequencing in this family led to the discovery of a novel homozygous missense mutation c.1072C>T (p.Q358*) in *DYMECLIN* gene. This mutation was predicted to form a truncated DYMECLIN (DYM) protein which may profoundly impair its normal function. The mutation in this gene is known to cause Dyggve-Melchior-Clausen disease (DMC). The current molecular analysis described a putative DMC phenotype associated with a *DYM* gene and also confirmed the clinical condition that segregate in the family. This study findings provide a start point for future studies on the function of DYMECLIN protein. (iv) In family 4 (Chapter 7), three affected subjects were presented with short neck, disproportionately short trunk, and protruding abdomen (secondary to lordosis). Other common characteristics included kyphoscoliosis, winged scapulae, crowded ribs, and pectus carinatum, while

clinodactyly was present in one of the affected individuals. Exome analysis revealed homozygous missense mutation c.220C>T (p.Glu74*) in the coding region of *FLNB* gene segregating with the malformation. Conclusively, the study described the expanded clinical spectrum of spondylocarpotarsal synostosis (SCT) syndrome. The scientific findings of this study might be helpful to establish genotype-phenotype correlation of SCT. Lastly (Chapter 8), a family with a single female patient was presented with syndromic condition along craniofacial anomalies like, cupid's bow, small jaw, and narrow palate. Exome sequencing in this subject led to the discovery of *de novo* compound heterozygous mutations p.L605P and p.M319V in the *CFAP46* gene. In order to get an insight into the molecular function of this gene, the disease was modeled *in vivo* using CRISPR/Cas9-mediated genome editing in frog tadpoles, and the functional effects of the detected variant on candidate protein function were evaluated. The whole mount *in situ* hybridization technique was adapted to explore the role of the CFAP46 protein in embryonic development. The craniofacial anomalies induced by the CRISPR/Cas9-mediated genome editing of *CFAP46* mimicked the patient's condition. Since the non-overlapping CRISPRs were so effective, it is likely that *CFAP46* is crucial to development. Further evidence from whole-mount *in situ* hybridization indicated that *CFAP46* is essential for neural crest development and that abnormalities of the face can result from CFAP46 protein deficiency. This work presents the first evidence of human craniofacial anomalies caused by *CFAP46* mutations and provides clinical and functional support for this hypothesis. In conclusion, the findings of this study may be helpful for understanding the puzzling nature of the rare genetic disorders. This knowledge could result in cutting-edge methods for quickly diagnosing, treating or preventing rare diseases. Clinicians might benefit from the study, and it would be equally useful for other medical institutes and organizations as well as for setting up centers for genetic counseling.

1. Introduction

1.1: Congenital anomalies

Congenital anomalies (CA), also known as birth defects, are any anatomical or functional abnomalities changes from expected norm, including physiological and metabolic illnesses, that appearing at birth. These birth defects are brought on by genetic, epigenetic, environmental, or additional causes such as maternal factors and consanguinity. CA are a significant contributor to newborn and infant morbidity and mortality, whether they are present alone or as part of a syndrome (Aloui *et al.,* 2017; Oliveira and Fett-Conte, 2013).

As reported that each year, 7.9 million infants born with major CA (Ndibazza *et al.,* 2011). Worldwide studies have revealed that there are significant regional differences in the occurrence of CA at birth demonstrating the complicated relationships between genetic and environmental factors. It has been estimated that CA occur in 2-3% of newborns in the US (Sarkar *et al.,* 2013; Ajao and Adeoye, 2019). According to reports, it can range from 1.07% in Japan to 4.3% in Taiwan. CA account for 2% of births in England and 1.49% in South Africa, respectively. In Southern Beirut, Lebanon, there are 1.64% serious CA per 1,000 live births. Different social, economic, and racial factors can explain differences in the CA rates (Francine *et al.,* 2014; Ameen *et al.,* 2018;). Congenital heart problems, Down syndrome, and neural tube defects are examples of CA that frequently occur (De Vrueh *et al.,* 2014).

Known causes of CA included genetic (30–40%) and environmental (5–10%) factors, but the origins of nearly 50% of CA are still unknown (Abebe *et al.,* 2021). Additional predisposing factors include chromosomal anomalies (6%), single gene abnormalities (25%) and multifactorial (20–30%) causes (Francine *et al.,* 2014). The

most common cause of CA has been identified as consanguineous unions. CA is more common in the offspring of consanguineous couples (Francine *et al.,* 2014).

Since CA have a long-lasting impact on health and survival, CA are becoming a major worldwide health concern (WHO, 2010; WHO, 2016). In low- and middle-income countries, CA are playing a more important role in determining children's survival and health as infections and other causes of early mortality are being controlled (WHO, 2010). CA affects 1/33 of infants and is linked to 3.2 million birth-related disabilities and about 300,000 annual deaths during the first month of life (Kurdi *et al.,* 2019). The WHO has consequently emphasized the urgent need for actions to support in prevention, diagnosis, and implementation of early interventions (WHO, 2010). Many low- and middle-income countries lack access to data on the prevalence and mortality of CA. While the majority of reports concerning CA statistics coming from high-income regions. For example, the UK report 20-year survival rate of 86% in the population-based study of live births in the UK with CA (Tennant *et al.,* 2010). In the same way, in New York state, the 25-year survival rate for live births with CA was 83% (Wang *et al.,* 2011), showing a consistent increase from the 1980s (78% from 1983 to 1988) to the early 2000s (89% from 2001 to 2006). Cardiovascular malformations (51%) and chromosomal anomalies (33%) were the main causes of mortality in children born with CA. In Korea, the foetal mortality rate was 13.5/10 000 live births, and the newborn mortality rate among infants with CA was 6.8/10 000 live births (Ko *et al.,* 2017).

Interactions between genetic, epigenetic, environmental, sociocultural, socioeconomic, racial and ethnic factors seem to be involved in the geographical patterns of CA occurrence. Patterns and prevalence may change with time and geographic location due to these complex interaction between known and unknown

genetic and environmental factors (Aloui *et al.,* 2017) which should be addressed to decrease the incidence and severity of conditions seen.

1.2: Genetic Basis of Disease

A substantial proportion of morbidities are influenced by genetics to some extent. Variations in our deoxyribonucleic acid (DNA) and differences in how that DNA works (individually or in combination), as well as the environment (which includes lifestyle) contribute to the disease process (Jackson *et al.,* 2018).

The majority of these illnesses are monogenic/Mendelian and account for about 80% of all rare diseases. Mendelian/monogenic disorders are the term used to describe diseases that segregate according to the Mendelian pattern of inheritance. There are almost 400 million people throughout the world with affected by 7000 different rare genetic disorders (OMIM) making these conditions a significant health burden as a group. Despite being monogenic, more than half of all inherited mendelian disorders still lack a genetic basis (Boycott *et al.,* 2017). The phenotype is still variable even when the causal gene is known and in patients with identical causative variant (such as siblings; Kose *et al.,* 2019), which makes diagnosis and patient management challenging (Missaglia *et al.,* 2015). On the other hand, by analyzing extensive genome-wide sequencing studies researchers have found evidence of a genetic predisposition to disease in individuals who have not yet manifested the disease (Tarailo-Graovac *et al.,* 2017). Thanks to the massive genome-wide sequencing research, we now know more about the penetrance and expressivity of rare disorders as well as the potential role of genetic modifiers (Rahit *et al.,* 2020).

1.3: Genetic Factors and Rare Mendelian Disease: Role of whole genome sequencing in the identification of causative variants

Even though we have relatively little understanding of Mendelian modifiers for rare diseases, it is clear that phenotypic diversity brought on by modifiers needs to be taken into consideration and addressed (Maroilley and Tarailo-Graovac, 2019; Aubart *et al.,* 2018). For instance, the rare Mendelian illnesses Gaucher disease (GD; OMIM 230800), caused by alterations in the Glucosylceramidase Beta 1 (GBA1) gene, was first identified about 150 years ago, in 1882 (Davidson *et al.,* 2018). GD is known to have a wide spectrum of presentations even among siblings. There are significant gaps in the identification and understanding of genetic modifiers in GD as well as in other single-gene disorders, despite decades of research efforts. For GD and other rare disorders, it is crucial to fill in these gaps for diagnosistic, prognosticis, therapeutic, and overall general patient care processes. To achieve this, it is important to look at each complete genome to completely understand this rare disorder's the variable phenotypesic manifestation observed. Whole-Genome Sequencing (WGS) gives access to all the genetic variants in an individual for exploration and analysis. With significantly decreasing costs, WGS makes it feasible to get an almost complete genomic view (~98–99%) of an individual and to identify potential disease-causing variations. With advancedments in bioinformatics tools, it is now possible to identify causative variants for phenotypes more precisely and correctly than ever before (Lappalainen *et al.,* 2019; Salgado *et al.,* 2016), which also creates the possibility to examine modifier variants in the same individuals.

Two important periods in the history of gene discovery can be identified. Prior to the turn of the millennium, positional cloning was used to identify 1,300 disease-related genes during the first era of linkage analysis. The second era began with the

completion of the human genome and the discovery of next generation sequencing (NGS), accelerating the rate of disease gene discovery from years of research, required previously, to genes now identified within a week (McInerney-Leo and Duncan, 2021).

The focus in human genetics research has shifted from identifying genetic variations to the extensive characterization of associations between genetic variation and phenotypes, and, more recently, of the general mechanisms by which genetic variation influences human biology and health. This shift began about 20 years ago (Lappalainen and MacArthur, 2021), with the completion and publication of the complete human genome. The majority mechanism of action of the more than 200,000 genetic variants, identified by that genome-wide association studies (GWAS), have conclusively linked to human complex traits are still mechanistically unknown (Claussnitzer *et al.,* 2020). Furthermore, it's sometimes unclear which gene is responsible for the biological impacts of these variants because the majority of them occur in non-coding regions of the genome (Maurano *et al.,* 2012). Similarly, to this, 47% of the nearly 1 million entries in the ClinVar database (Landrum *et al.,* 2018) of variants found in people with severe genetic diseases are labeled as having either uncertain or conflicting annotations, indicating a lack of clear understanding about the impact of variants on their molecular mechanisms function and disease (ClinVar). To tackle this persistent uncertainty, more accurate and efficient tools to understand the molecular pathways by which mechanisms of genetic variations affects phenotype will be necessary.

By improving understanding of the direct inference mechanisms of variation and their likely disruptive effects on the normal gene function of key genes, such approaches will increase the accuracy of genetic diagnosis and prediction, especially

for rare/or novel disease-causing variants. These methods will help to accelerate the development of new therapeutics by highlighting not only the gene products directly responsible for illness pathogenesis but also the direction of impact, the relevant target cell types affected, and the broader biological pathways through which a gene influences involved in disease risk (Lappalainen and MacArthur, 2021).

Sequencing Genetic analysis of the parent-child trio allows for the discovery of de novo mutations that were previously not found by linkage analysis. NGS allows for the discovery of recessive disease genes in even a single affected person. In addition, both the number and size of the affected families required to locate a causal gene have decreased. Each NGS technology, including whole exome sequencing (WES) and panel sequencing, has its advantages. The majority of genes have been discovered through NGS using WES, while WGS is better at spotting copy number variations, chromosomal rearrangements, and repeat-rich regions. Panels, on the other hand, produce manageable amounts of data with no chance of unexpected results and are incredibly cost-effective, so they are frequently utilized for diagnostic purposes.

However, when there is diagnostic uncertainty, selecting the right panel can be quite difficult. Therefore, Under these conditions, WES has a greater diagnostic yield. All methods of sequencing follow the same basic principle. There are four major steps that are common in all types of NGS (Fig. 1.1). Many of the social, ethical, and legal ramifications of NGS are similar to those of genetic testing in general but are magnified by the sheer amount of data (e.g., relationship misattribution, identification of variants of uncertain significance, and genetic discrimination); some are specific to WES and WGS, (e.g., incidental or secondary findings; (McInerney-Leo and Duncan, 2021)).

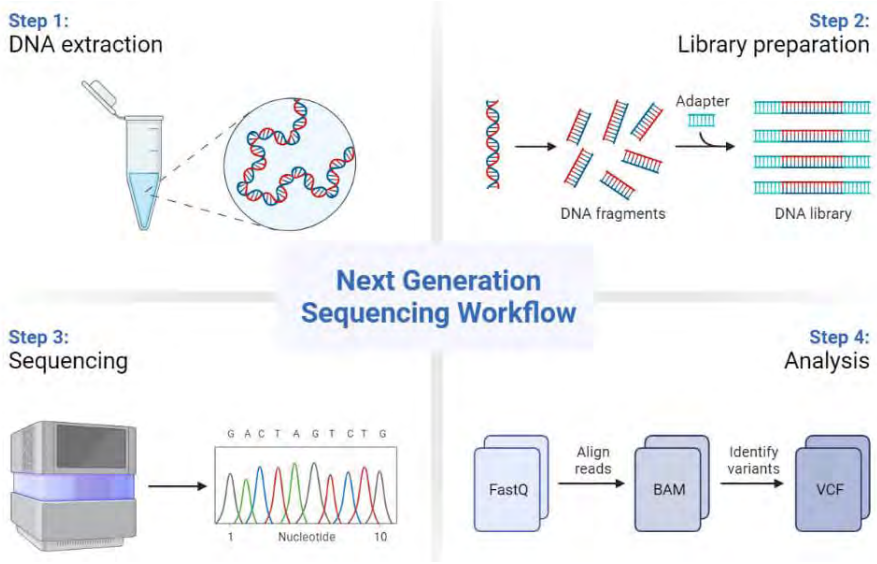


Fig 1.1: Generic overview of NGS (Source**:** https://thebiologynotes.com/next-generation-sequencing-ngs/)

1.4: Overview of the current study: Scheme of study

This was a multitired study in which various levels of sampling, experimental and analytical approaches were utilized. Initially, through an epidemiological approach the data on subjects with hereditary disorders originating from Hazara population of Pakistan were collected. First-hand data on 1,189 subjects from independent subjects/families were collected and the prevalence-pattern of congenital/hereditary anomalies were established for Hazara division of Pakistan. Secondly, 141 independent subjects with congenital limb deficiencies (CLD) were recruited from Khyber Pakhtunkhwa (KP) province of Pakistan and their phenotypic pattern was established. Further, blood samples of five families with rare phenotypes were collected, with the possibility to carry out molecular genetic analyses and functional genomic study.

Five families in which the patients had various presentations like intellectual disability, skeletal dysplasia, anophthalmia, scoliosis and blindness and craniofacial anomalies were selected for molecular study. The phenotypes in the four families were segregating autosomal recessively. Novel mutations in the candidate genes *GRIP1*, *DYM1*, and *FLNB* were found through mutation analysis in three families. While a single new gene, *CFAP46*, linked to craniofacial anomalies, segregated within a single family. Finally, in a large family with multiple affected subjects exhibiting the symptoms of intellectual disability, SNP based genotyping data were generated. Homozygosity mapping approach was adopted in order to detect regions of homozygosity shared among the affected subjects. Further, two of the affected individuals in this family was subjected to exome sequencing. Analyses of these data led to the shortlisting of rare variants which are pathologically relevant to the phenotype and also fall in the homozygous intervals detected in the SNP scan. In conclusion, a multitired approach was adopted for clinical and genetic elucidation of subjects/families with rare disorders. The overall scheme of the study is shown in Fig. 1.2.

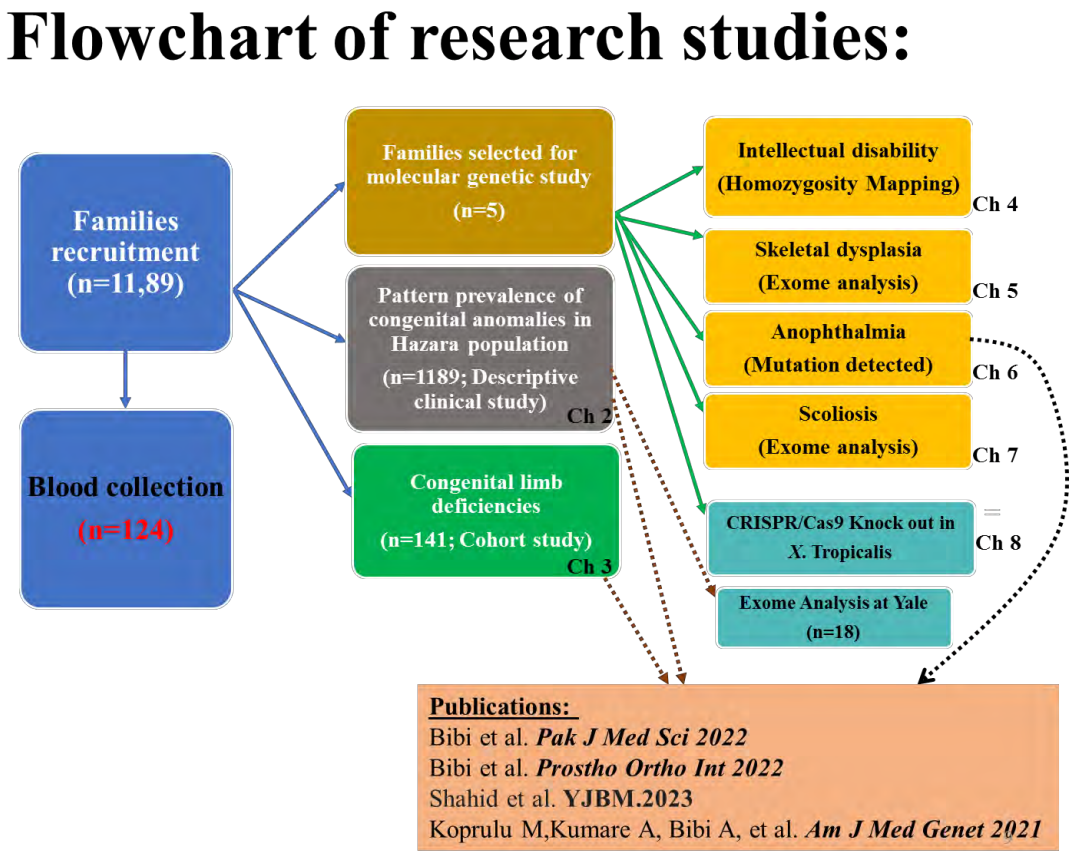


Fig 1.2: Flowchart showing the distribution of current study.

2. Clinical and genetic attributes of congenital and hereditary anomalies prevalent in Haraza population of Khyber Pakhtunkhwa, Pakistan

2.1: Abstract

In Pakistan, various factors like high rate of consanguinity, extended families and marriage at young age contribute to a high prevalence of congenital and hereditary anomalies (CA). Epidemiological studies elucidating detailed biodemographic correlates of CA, that could serve as a framework for establishing the management strategies for these anomalies, are deficient in Pakistan. To this end, the current research aimed to characterise the prevalence, clinical and genetic attributes, and biodemographic characteristics of CA in the Hazara population for which there was no data available. In a retrospective cross-sectional study design, subjects/families with CA were identified through district hospitals, community centers and through door-to-door surveys in Hazara. Without any previous knowledge of a candidate's gender, ethnicity, or caste system, a convenience sampling was used to determine their prospects for recruitment. Descriptive and phenotypic information was collected, pedigrees were examined, and demographic and parental characteristics were documented. In this study 1,189 subjects and families were recruited with nine major and at least 97 minor categories of anomalies. The most prevalent condition was neurological disorder (n=486; proportion=0.409; % Coefficient of inbreeding (CI)=0.381-0.437), which was followed by limb defects (n=292), musculoskeletal defects (n=106), sensorineural/ear defects (n=101), blood disorders (n=75), eye/visual impairments (n=39), ectodermal anomalies (n=30), and congenital heart defects (n=26). There were 64.6% sporadic cases and 35.3% familial cases in this cohort. There is significantly higher parental consanguinity in isolated and familial cases, compared to syndromic and sporadic cases (67% and 72% *vs* 65% 63%). Further, speech apraxia (n=48) and epilepsy (n=36) were the commonest associations in the syndromic case. Demography, parental consanguinity, paternal ages, and pedigree structures showed conspicuous heterogeneity among the major and minor categories of CA. In conclusion, the trend in anomalies seen in this cohort, along with the high number of sporadic cases, suggest that nongenetic etiological factors may play a significant role that may be minimized by improving healthcare as well as through genetic counseling.

2.2: Introduction

Congenital anomalies (CA) are a diverse group of conditions that affect embryonic and fetal development (Bhatti *et al.,* 2019). CA are categorized into two main types: single-gene disorders with Mendelian dominant, X-linked or recessive inheritance patterns, and multifactorial disorders involving combinations of neutral variants producing a predisposition to disorders including cardiovascular disease and asthma/eczema, among others.

The prevalence of CA ranges from 4 to 5% worldwide, with the actual numbers differring significantly between different countries and geographic areas (EUROCAT, 2015). CA management is difficult and necessitates a more comprehensive understanding of the problem (Zahra *et al.,* 2016). The prevalence of CA across different sub-populations, linguistic and ethnic groups, and socioeconomic strata within the population is variable (Cosme *et al.,* 2017).

CA are common among many populations and they cause lifelong illness (physical or mental) or early death. CA can be severe when affecting vital organ-systems and causing major disabilities, i.e., malformations of the central nervous system, heart, kidneys, etc., or can be milder when they do not seriously affect the normal life functions of the subject, i.e., polydactyly, syndactyly, albinism, etc. CA may occur as isolated abnormalities or as a part of a syndrome and involve multiple organ-systems (Leppig *et al.,* 1987; Temtamy and McKusick, 1978). CA may be found at an early age in life as soon after birth or appear later in life.

Consanguinity has relatively little effect on the prevalence of dominant and X-linked disorders, but has a large effect on recessive disorders. The impact of relatedness in the predisposition of multifactorial disorders is still unclear. Most

people carry one or two gene variants, that have no effect on their health but these can cause a recessive disorder in specific individuals (Vogel and Motulsky, 2013).

Children have a 25% chance of inheriting it from both parents and suffering from the homozygous recessive illness when the parents are heterozygous for a recessive variant. Hundreds of recessive disease variants are present in most populations and a range of disorders occurs occaionally, apparently sporadically when both parents carry the same variants. Some diseases are common, e.g. cystic fibrosis or haemoglobin disorders, but most are rare sinse the chance of both partners carrying the same variant is small. In typical Northern European populations, ~4% of all congenital/genetic disorders and 17% of single-gene disorders are recessive (Baird *et al.,* 1988). Still, they severely impact the disease spectrum and cause long-term disability as well as child mortality. The chance of both parents carrying the same recessive disease variant and having affected children increases due to consanguinity. This effect is particularly marked for rare disorders where a partner who carries the same disorder is unlikely unless they are related. Consequently, the prevalence of many rare recessive conditions and of congenital and genetic disorders is highest in communities in which consanguineous marriage is common. Still, some widespread congenital anomalies like cerebral palsy (CP), Down syndrome and neural tubes defects are not inherited in recessive patterns and are not common in consanguineous families (Modell and Darr, 2002).

According to a cross-sectional study carried out in Khyber teaching hospital in Peshawar, CA mostly affected the brain (10/1,000), heart (8/1,000), kidneys (4/1,000), and limbs. Among 1062 deliveries, 3% of newborns were born with certain types of CA, with hydrocephalus (23%), anencephaly (13%), and spina bifida (10%) as major anomalies (Khan *et al.,* 2015). Zahra *et al.,* (2016) recruited a total of 246

independent subjects with CA during an observational study on the epidemiology of CA in Kurram Tribal Agency in northwest Pakistan. which is a population badly affected by war and political instability. In those data, the sporadic and isolated occurrence of CA were more abundant than familial and syndromic occurrences. Further, neurological disorders were the most frequent, followed by musculoskeletal defects, limb anomalies, sensorineural/ear defects, ectodermal anomalies, congenital heart defects, and eye/visual impairments. Nevertheless, in-depth studies relative to the risk factors, epidemiology and distribution pattern of CA are essential for the implementation of intervention measurements (Zahra *et al.,* 2016).

In Pakistan, CA caused approximately 6-10% of perinatal deaths (Korejo *et al.,* 2007). According to the reports, the etiology of CA is unknown in 40-60% of CA, 20% have genetic and non-genetic factors, 8% are due to single-gene mutations, 6% are the result of chromosomal abnormalities, and 5% are due to maternal influences and pregnancy outcomes (Blue *et al.,* 2019). The scope of socio-demographic factors linked with CA differs in various populations, for example Costa *et al.,* (2006) observed that rate of CA was significantly associated with poor literacy and low socio-economic status.

The children from consanguineous unions have a significantly high risk of recessively inherited disorders (Shawky *et al.,* 2013). Ethnicity, language, religion, and geographic distribution, combined with a high prevalence of consanguinity, result in a genetically isolated community. In such populations, a limited, comprehensive, long, and multigenerational pedigree is expected (Hamamy, 2012), which give geneticists opportunities for linkage analysis and mapping of monogenic autosomal recessive diseases. Many isolated populations, such as in Pakistan, and certain other

communities, have played an important role in identifying novel mutations in autosomal recessive genetic disorders (Ali, 2010).

In consanguineous unions, every partner shares the genes of a common grandparent (Hamamy, 2012). In the Middle East and South Asian countries, almost 40% of all marriages are consanguineous (Abdalla and Zaher, 2013). While in Pakistan approximately 65% of marriages are commenced among blood relatives due to economic, social, and cultural reasons in different regions of Pakistan (Iqbal *et al.,* 2022). More than a few studies have been conducted in different areas of Pakistan including Rahim Yar Khan, Sargodha, Bhimber district of Mirpur division (AJK), Khyber Pakhtunkhwa and Federally Administered Tribal Areas (FATA) showed 58.46%, 56.72%, 62%, 58.3% and 22.34% consanguineous marriages respectively (Umair *et al.,* 2018). At present new interventions and better control reduce the burden of communicable diseases but meanwhile hereditary and genetic disorders appear as a major problem in healthcare systems. Most of the monogenic diseases are rare when appearing separately but as a group becomes quite common (Biesecker *et al.,* 2012).

Pakistan has major populations living in rural areas where health care is present but poorly maintained, lacking access to trained professional as well as modern equipment. In these situations, the healthcare system of Pakistan is unable to deliver proper support to the communities (Kumar and Bano, 2017). Even though considerable advancements over the last twenty years in Pakistan, infant and neonatal death rates are increasing. Shockingly, Pakistan was rated 149th among 179 countries on the Maternal Mortality Ratio Index in 2015 (Hall and Taylor, 2003). Healthcare resources, including finance and transport, are not distributed in a need-based manner in Pakistan. Moreover, the current number of healthcare professionals does not meet

the high rate of population growth. In addition, available medical staff is inexperienced, poorly paid, and deprived of the most recent skills for medical practice. However, mostly the health-care system depends upon private organizations for more advanced medical facilities, these services are very expensive, and most people cannot afford them. There are huge inequities in the availability of health-care services to the deprived population (Desa, 2015; Kumar and Bano, 2017). Recent assessment represents that 2.3% of total deaths in Pakistan occur due to CA. At this point, certain factors such as low socioeconomic status, insufficient prenatal care, maternal malnutrition and consanguinity are significantly associated with the high rate of CA. According to a study conducted in the Combined Military Hospital, Kharian, CA affected 7% of the 3,210 cases. In-depth study of the literature represents that majority of the reports on CA were hospital-based (Arfaksad and Wajahat, 2016; Hussain *et al.,* 2014). Such studies do not completely cover the subjects from remote areas, where usually birth takes place in the home with the help of local attendants (Shah *et al.,* 2010). A countrywide picture on the epidemiology of CA is fragmented due to a lack of proper surveillance and reporting.

CA are more prevalent in Pakistan compared to the rest of the world, due in part, to consanguineous marriages, overlapping generations, marriage at young ages, large families, and well-adapted communities presenting multigenerational family trees with many rare hereditary disorders (Peltonen *et al.,* 2000; Wahab and Ahmad, 1996). In Pakistan, the basic focus of the national health surveillance system is a traumatic, infectious and common disease while there is no well-adapted systematic method for the monitoring of CA (Jabeen and Malik, 2014). In this situation, there is an immense need to monitor the pattern of CA across different geographic regions,

ethnicities, and socioeconomic levels for the ascertainment of their burdens on the society and to plan efficient intervention approaches.

2.3: Subjects and Methods

2.3.1: Sampling area

A descriptive clinical and genetic epidemiological study on CA was carried out in the northwestern regions of Pakistan, in the Hazara division of Khyber Pakhtunkhwa. Hazara is located to the east of Tarbela Dam, west of Azad Jammu and Kashmir, north of Punjab, and south of Gilgit Baltistan. Hazara division encompasses seven districts including Haripur, Abbottabad, Mansehra, Upper Kohistan, Lower Kohistan, Batagram and Torghar (Fig. 2.1).

The Hazara Division is home to people of different ethnicities, including the Awan, Dhund, Gujar, Jadoon, Kashmiri, Kharral, Kohistani, Pathan, Syed, Swati, Tanoli, and Turk (Akbar *et al.,* 2016).

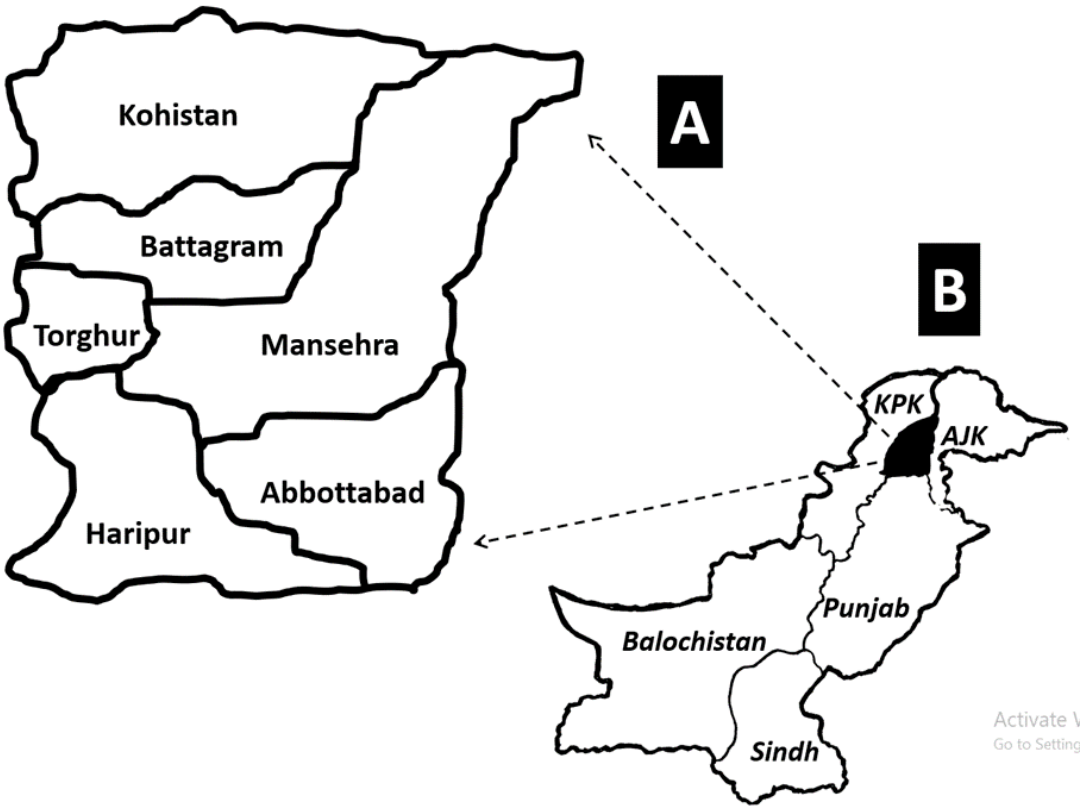


Fig 2.1: Map of Pakistan (B) with superimposed map of Hazara division (A) depicting the study districts.

2.3.2: Sampling strategy

For a picture of the epidemiology of different CA across Hazara Division, a study was carried out from July 2015-to-June 2021. Subjects with various types of CA were recruited from District Headquarter hospital of Haripur, Abbottabad and Mansehra districts. Additionaly, the subjects were also recruited by visiting public places and through door-to-door surveys. It was an observational study in which convenience sampling was conducted due to the availability of resource persons familiar with local languages and traveling and logistics options. Different sampling strategies were adapted in different time scenarios. The study was approved by the Ethical Review Committee, Quaid-i-Azam University, Islamabad.

Data were collected and documented in written form in the presence of the family head/guardian after informed written, or proper verbal, consent (where the subject was illiterate) was approved prior to the data collection. When a participant was incapable of providing consent due to reasons including: below the legal age of consent delivery, disabled due to neuromuscular defects, deaf-mute, blind, or otherwise incapable, their parent/guardian or educated elder family member provided informed written consent. Only subjects belonging to the Hazara division were included in the study and those who were not regular members of the surveyed household or not capable to provide complete information were excluded. Data for each subject were obtained on biodemographic and socio-economic variables including personal information, ethnicity, parental marital status, family type, occupation, socioeconomic status etc. With the help of a local medical officer and physician, all individuals were physically examined. Detailed clinical features were recorded and photographs of affected individuals showing phenotype were taken. Previous medical records and copies of significant test reports, including CT scans, MRI scans and X-rays, were taken and assessed by a local physician for diagnosis. Participants from the rural area were brought in for clinical examination to the local district hospital. Pedigrees are helpful for the determination of the familial segregation of deformity, affected sibship and generation with the disease, hence, three-generation pedigrees for each family were drawn. When multiple affected subjects present in one family, with a particular malformation, were considered as one mutational event, and hence, only the index subject in each family was included in the analyses. The fieldwork and data collection were performed by a qualified person trained in medical genetics together with a local guide. The guides were respectable residents of specific areas and included local lady health workers or paramedic who was familiar with the

local population. Rehabilitation or special education centers were not included in the data collection.

2.3.3: Classification of anomalies

Only congenital or hereditary anomalies were considered for data analysis, while those resulting from trauma, acquired or infectious conditions were not included. Specialist resident doctors at the District Hospitals made the primary diagnoses and only confirmed CA with congenital and/or hereditary anomalies were recruited. Recruited subjects were categorized into (i) familial or sporadic, and (ii) isolated or syndromic. Syndromic CA were identified with respect to the more severe symptoms in the following order: neurological disorders, neuromuscular defects, musculoskeletal defects, eye/visual impairments, sensorineural/ear anomalies, and limb defects. Secondary symptoms counted as associated malformation when present in addition to the primary disease. A structured questionnaire was used to record information. This was divided into three sections: section one for demographic data; section two dealt with various risk factors, including medical history, parental medical history, paternal age and consanguinity; and section three documented phenotypic details of the abnormalities. The standard coding system of the International Classification of Diseases, Tenth Revision (ICD-10), was used to define abnormalities based on primary diagnoses. Each entity was used to search the Online Mendelian Inheritance in Man (OMIM) and Orphanet databases. Limb defects and malformations were further characterized into well defined phenotypic entities and sub classified by involvement of upper and/or lower limbs, laterality, symmetry, and axis of involvement (Malik, 2014).

2.4: Results

1,189 subjects and familes were recruited having 9 major categories of CA. 678 (57%) were males and 511 (43%) females (Table 2.1). Sporadic cases were more common than familial (769(65%) *vs* 420(35%); respectively). The total number of affected subjects was 2212 with 1284 males 928 females affected (p=0.0005 for males vs females). Additional analyses of the gender-specific data revealed that neurological disorders (n=276 vs n=210), limb abnormalities (n=163 vs n=129), musculoskeletal defects (n=63 vs n=43), sensorineural/ear defects (n=66 vs n=35), blood disorders (n=48 vs 27) and congenital heart defects (n=16 *vs* n=10) were more common in males than in females (Table 2.1). Contrary to this Eye/visual impairments and ectodermal anomalies were highest in females compared to males (n=21 *vs* 18 and n=16 *vs* n=*14 vs* n=16; Table 2.1).

Table 2.1: Major categories of CA, familial/sporadic nature, and total number of affected family members

Major category	Index subject			Proportion	95% CI	Familial/sporadic nature*		Total number of affected in all families*		
	Male	Female	Total			Familial	Sporadic	Males	Female	Total
Neurological disorders	276	210	486	0.409	0.381-0.437	118	368	396	301	697
Limb defects	163	129	292	0.246	0.221-0.270	102	190	336	216	552
Musculoskeletal defects	63	43	106	0.089	0.073-0.105	54	52	136	129	265
Sensorineural/ear defects	66	35	101	0.085	0.069-0.101	52	49	137	83	220
Blood disorders	48	27	75	0.063	0.049-0.077	29	46	83	34	117
Eye/visual impairments	18	21	39	0.033	0.023-0.043	23	16	57	46	103
Ectodermal anomalies	14	16	30	0.025	0.016-0.034	22	8	70	42	112
Congenital heart defects	16	10	26	0.022	0.014-0.030	6	20	24	17	41
Others	14	20	34	0.029	0.019-0.038	14	20	45	60	105
Total	**678**	**511**	**1,189**	1.000	-	**420**	**769**	**1284**	**928**	**2212**

*****Chi-test statistics were statistically significant.

In the total sample, n=107 subjects belonged to District Haripur, n=191 from Mansehra, n=126 from Abbottabad, 31 from Kohistan and n=23 from district Batagram (Table 2.2; Fig. 2.1). Detailed distribution of index subjects and familial/sporadic cases is across the key demographic variable is presented in Table 2.2. Distribution was observed to be statistically non-significant in all variables.

Table 2.2: Demographic distribution of index subjects

Variables	Male, No.(%)	Female, No.(%)	Total, No.(%)
District			
Haripur	307 (45)	241 (47)	548 (46)
Mansehra	191 (28)	128 (25)	319 (27)
Abbotabad	126 (19)	103 (20)	229 (19)
Kohistan	31 (5)	22 (4)	53 (5)
Batagram	23 (3)	17 (3)	40 (3)
Sum	**678 (57)**	**511 (43)**	**1189 (100)**
Rural/urban origin			
Rural	401 (59)	301 (59)	702 (59)
Urban	277 (41)	210 (41)	487 (41)
Age intervals (years)			
Up to 5	213 (31)	151 (29)	364 (31)
>5-9	108 (16)	80 (16)	188 (16)
>9-19	226 (33)	148 (29)	374 (31)
>19	131 (19)	132 (26)	263 (22)
Mother tongue			
Hindko	471 (69)	370 (72)	841 (71)
Pashto	83 (12)	55 (11)	138 (11)
Punjabi	56 (8)	28 (5)	84 (7)
Urdu	39 (6)	34 (7)	73 (6)
Pahari	11 (2)	5 (1)	16 (1)
Others	20 (3)	19 (4)	39 (3)
Caste-system			
Awan	162 (24)	129 (25)	291 (24)
Pathan	66 (10)	60 (12)	126 (11)
Gujjar	67 (10)	43 (8)	110 (9)
Tanoli	37 (5)	28 (5)	65 (5)
Swati	30 (4)	30 (6)	60 (5)
Rajpoot	29 (4)	22 (4)	51 (4)
Mughal	36 (5)	12 (2)	48 (4)
Abbasi	26 (4)	19 (4)	45 (4)
Others	184 (33)	131 (33)	315 (33)
Family/household type			
Nuclear	366 (54)	256 (50)	622 (52)
Extended	312 (46)	255 (50)	567 (48)

Literacy (n=794)			
Illiterate	298 (63)	212 (66)	510 (64)
Literate	174 (37)	110 (34)	284 (36)
Literacy level (n=284)			
Primary	93 (20)	58 (18)	151 (19)
Secondary school	71 (15)	36 (11)	107 (14)
Graduate and higher	10 (2)	16 (5)	26 (3)
Economic quintile (n=1109)			
Low	305 (46)	192 (41)	497 (45)
Low-mid	278 (42)	221 (47)	499 (45)
Mid	11 (2)	2 (0)	13 (1)
High-mid	9 (1)	11 (2)	20 (2)
High	56 (9)	44 (10)	80 (7)

Chi-distribution was statistically not significant in all variables.

All the major categories were further resolved according to their nearest definitions in OMIM and ICD-10 databases and hence, at least 99 minor categories could be identified. Table 2.3 presents all the major and minor categories with their proportions, 95% CI, OMIM entries codes and ICD-10 database identifiers (Table 2.3).

Table 2.3: Major and minor categories of congenital/hereditary malformations

Major/minor categories	Frequency	Proportion	95% CI	ICD-10	OMIM
Neurological disorders	**486**	**0.409**	**0.381-0.437**		
Intellectual disability	176	0.148	0.128-0.168	F79	
Cerebral palsy	148	0.124	0.106-0.143	G80.0	
Epilepsy	41	0.034	0.024-0.045	G40	117100
Autism/low IQ	25	0.021	0.013-0.029	F84.0	
Down syndrome	18	0.015	0.008-0.022	Q90	190685
Hydrocephaly	14	0.012	0.006-0.018	G91.9	236600
Microcephaly	14	0.012	0.006-0.018	Q02	251200
Global developmental delay	13	0.011	0.005-0.017	Z13.42	618330
Spina bifida	11	0.009	0.004-0.015	Q05	182940
Ataxia	7	0.006	0.002-0.010	R27.0	160120
Migraine	5	0.004	0.001-0.008	G43	
Multiple sclerosis	4	0.003	0.000-0.007	G35	
Neuropathies	4	0.003	0.000-0.007	G60.9	162400
Macrocephaly	3	0.003	0.000-0.005	Q75.3	153470
Arnold Chiari malformation	1	0.001	-0.001-0.002	Q07.0	207950
Cystic encephalomalacia	1	0.001	-0.001-0.002		
Tremor	1	0.001	-0.001-0.002	R25.1	190300
Limb defects	**292**	**0.246**	**0.221-0.270**		
Talipes (all types)	141	0.119	0.100-0.137	Q66.0	119800
Polydactyly, postaxial	34	0.029	0.019-0.038	Q69	174200
Polydactyly, preaxial	31	0.026	0.017-0.035	Q69.1	174400, 174500
Transverse limb amputations	23	0.019	0.012-0.027	Y83.5	
Syndactyly (all types)	18	0.015	0.008-0.022	Q70	609815
Brachydactyly (all types)	10	0.008	0.003-0.014	Q68.81	113000
Clinodactyly	9	0.008	0.003-0.012	Q74.0	148520
Camptodactyly	7	0.006	0.002-0.010	Q74.0	114200
Leg length	4	0.003	0.000-0.007	M21.7	

discrepancy					
Constriction band syndrome	3	0.003	0.000-0.005	Q79.8	217100
Thumb hypoplasia/aplasia	3	0.003	0.000-0.005		188100
Clubbing of digits	2	0.002	-0.001-0.004	R68.3	119900
Hallux valgus	2	0.002	-0.001-0.004	M20.1	
Fibular hypoplasia	1	0.001	-0.001-0.002	Q73	
Macrodactyly	1	0.001	-0.001-0.002	Q74.2	155500
Radial hemimelia	1	0.001	-0.001-0.002	Q73.8	
Symphalangism	1	0.001	-0.001-0.002	Q70.9	185800
Trigger thumb	1	0.001	-0.001-0.002	M65.319	190410
Musculoskeletal defects	**106**	**0.089**	**0.073-0.105**		
Muscular dystrophy	23	0.019	0.012-0.027	G71.0	310200
Hypotonia (limbs)/myopathies	23	0.019	0.012-0.027	P94.2	300868
Dwarfisms	20	0.017	0.010-0.024	E34.3	100800
Congenital hip dysplasia	11	0.009	0.004-0.015	Q65.8	142700
Scoliosis	6	0.005	0.001-0.009	M41	181800
Kyphoscoliosis	4	0.003	0.000-0.007	M40	610170
Osteogenesis imperfecta	4	0.003	0.000-0.007	Q78.0	166200
Arthrogryposis	2	0.002	-0.001-0.004	Q74.3	108120
Carpal fusion	2	0.002	-0.001-0.004		
Exostosis	2	0.002	-0.001-0.004	Q78.6	133700
Klippel Feil syndrome	2	0.002	-0.001-0.004	Q76.1	118100
Pectus carinatum	2	0.002	-0.001-0.004	Q67.7	
DuPan syndrome	1	0.001	-0.001-0.002		228900
Genu valgum disorder	1	0.001	-0.001-0.002	M21.06	137370
Muscular torticollis	1	0.001	-0.001-0.002	M43.6	189600
Rheumatoid arthritis	1	0.001	-0.001-0.002	M06	180300
Rickets, vitamin D resistant	1	0.001	-0.001-0.002	E83.3	277440
Sensorineural/ear defects	**101**	**0.085**	**0.069-0.101**		
Deaf and mute	88	0.074	0.059-0.089	H90	304500
Microtia/deformed pinna	8	0.007	0.002-0.011	Q17.2	600674

Speech apraxia	3	0.003	0.000-0.005	R47.9	602081
Deaf only	1	0.001	-0.001-0.002		
Mute only	1	0.001	-0.001-0.002		
Blood disorders	**75**	**0.063**	**0.049-0.077**		
Thalassemia	59	0.050	0.037-0.062	D56	613985
Hemophilia	15	0.013	0.006-0.019	D66	306700
Fanconi anemia	1	0.001	-0.001-0.002	D61.09	227650
Eye/visual impairments	**39**	**0.033**	**0.023-0.043**		
Blindness	20	0.017	0.010-0.024	H53.5	216900
Squint eyes/strabismus	9	0.008	0.003-0.012	H50.9	185100
Colour blindness	3	0.003	0.000-0.005	H53.5	303800
High myopia	3	0.003	0.000-0.005	H52.10	
Night blindness	3	0.003	0.000-0.005	H53.60	310500
Anophthalmia	1	0.001	-0.001-0.002	Q11.2	251600
Ectodermal anomalies	**30**	**0.025**	0.016-0.034		
Atopic dermatitis/eczema	8	0.007	0.002-0.011	L20.9	603165
Albinism, oculocutaneous	5	0.004	0.001-0.008	E70.3	203100
Alopecia totalis	4	0.003	0.000-0.007	L63.1	203655
Psoriasis	3	0.003	0.000-0.005	L40.9	177900
Ectodermal dysplasia	2	0.002	-0.001-0.004	Q82.4	305100
Hypotrichosis	2	0.002	-0.001-0.004	H02.72	605389
Ichthyosis	2	0.002	-0.001-0.004	L85.0	242300
Alopecia areata	1	0.001	-0.001-0.002	L63.9	104000
Neurofibromatosis	1	0.001	-0.001-0.002	Q85.00	162200
Onychodystrophy	1	0.001	-0.001-0.002	L60.3	161050
Palmoplantar keratoderma	1	0.001	-0.001-0.002	L40.3	144200
Congenital heart defects	**26**	**0.022**	0.014-0.030		
Ventricular septal defect	12	0.010	0.004-0.016	Q21.0	614429
Arterial septal defect	6	0.005	0.001-0.009	Q21.1	108800
Coronary artery disease	5	0.004	0.001-0.008	I125.10	608901
Atrioventricular canal defect	2	0.002	-0.001-0.004	Q21.2	606215
Bradycardia	1	0.001	-0.001-0.002	R00.1	

Others	**34**	**0.029**	0.019-0.038		
Cleft lip/cleft pallet	8	0.007	0.002-0.011	Q37	119530
Bardet-Biedl syndrome	5	0.004	0.001-0.008	Q87.89	209900
Neonatal diabetes mellitus	4	0.003	0.000-0.007	P70.2	222100
Anomalies of kidney and urinary tract	3	0.003	0.000-0.005	Q64.9	
Celiac disease	2	0.002	-0.001-0.004	K90.0	212750
Congenital hypothyroidism	2	0.002	-0.001-0.004	E03.9	275200
Lymphedema	2	0.002	-0.001-0.004	I89.0	
Anorectal malformations	1	0.001	-0.001-0.002		107100
Bladder exstrophy	1	0.001	-0.001-0.002	Q64.10	600057
Congenital immunodeficiency	1	0.001	-0.001-0.002	D89.9	
Glucose 6-P-deficiency	1	0.001	-0.001-0.002	D55.0	305900
Hirschsprung disease	1	0.001	-0.001-0.002	Q43.1	142623
Neonatal adiposity	1	0.001	-0.001-0.002	E66.9	
Orofacial anomaly	1	0.001	-0.001-0.002	G24.4	
Undescended testes	1	0.001	-0.001-0.002	Q53.9	

The overall parental consanguinity was estimated to be 66% and it ranged from 60%-81% in the major categories of CA (P=0.07; Table 2.4). Consanguinity rates were observed to be highest in congenital heart defects (81%), sensorineural/ear defects (77%), blood disorders (72%), and ectodermal anomalies (70%), and were lowest in limb defects (60%) and neurological disorders (65%). Consanguinity was found to be significantly higher in the familial cases compared to sporadic cases(72% *vs*. 63%, respectively; P=0.004). Consanguinity among the isolated and syndromic cases was 67% and 65%, respectively (P=0.694), and among the index males and females 68% and 64%, respectively (P=0.193; Table 2.4).

In this cohort, the isolated presentation was more customary compared to the syndromic occurrence (70% *vs*. 30%; P<0.0001) (Table 4). The highest number of isolated cases were witnessed among sensorineural/ear defects (98%), followed by blood disorders (96%), ectodermal anomalies (90%), and limb defects (89%). On the other hand, syndromic cases were most conspicuous among the neurological disorders (55%), followed by musculoskeletal defects (25%) and 'Others' category (24%; P<0.0001). Furthermore, familial cases had a higher likelihood of being isolated types compared to the sporadic cases (78% *vs*. 66%; p<0.0001). The distribution of isolated/syndromic presentations among the index males and females did not differ significantly (P=0.512; Table 2.4).

Table 2.4: Parental consanguinity in various sample types

Variables	**Parental consanguinity, N (%)**		**Isolated/syndromic nature, N (%)**	
Major category	**Yes**	**No**	**Isolated**	**Syndromic**
Neurological disorders	316 (65)	170 (35)	217 (45)	269 (55)
Limb defects	176 (60)	116 (40)	259 (89)	33 (11)
Musculoskeletal defects	72 (68)	34 (32)	79 (75)	27 (25)
Sensorineural/ear defects	78 (77)	23 (23)	99 (98)	2 (2)
Blood disorders	54 (72)	21 (28)	72 (96)	3 (4)
Eye/visual impairments	27 (69)	12 (31)	31 (79)	8 (21)
Ectodermal anomalies	21 (70)	9 (30)	27 (90)	3 (10)
Congenital heart defects	21 (81)	5 (19)	23 (88)	3 (12)
Others	23 (68)	11 (32)	26 (76)	8 (24)
Total	**788 (66)**	**401 (34)**	**833 (70)**	**356 (30)**
	P=0.07		P<0.0001	
Familial/sporadic nature				
Sporadic	487 (63)	282 (37)	507 (66)	262 (34)
Familial	301 (72)	119 (28)	326 (78)	94 (22)
	P=0.004		P<0.0001	
Isolated/Syndromic				
Isolated	555 (67)	278 (33)	-	-
Syndromic	233 (65)	123 (35)	-	-
	P=0.694			
Gender of index subject				
Male	460 (68)	218 (32)	470 (69)	208 (31)
Female	328 (64)	183 (36)	363 (71)	148 (29)
	P=0.193		P=0.512	

The syndromic cases were further analyzed in order to observe the combination of associated anomalies (Table 2.5). In the overall cohort, there were at least 495 associated anomalies. Among the individual anomalies, the syndromic appearance were common in intellectual disability (n=215), cerebral palsy(n=65), talipes (n=20), epilepsy (n=17), Down syndrome (n=16), and hypotonia (n=14). The six most common associated

anomalies were speech apraxia/stuttering (n=93), epilepsy (n=55), hypotonia (n=43), talipes (n=35), squint eyes (n=27), and deaf/mute (n=22).

Analysis of the data with respect to the parity of the index subject showed most of the cases falling into the first parity (n=329), followed by second (n=250), third (n=198), and forth parity (n=127) (Table 2.6). However, the differences in the distribution of major anomalies with respect to parity were statistically not significant.

Table 2.5: Syndromic cases with the combination of associated malformations

Congenital malformation[1]	No. of cases	Associated anomalies#,*										Sum
		Speech apraxia/ stuttering	Epilepsy	Hypotonia	Talipes	Squint eyes	Deaf/ mute	Shortness of breath	Cerebral palsy	Blind	Others	
Intellectual disability	117	48	36	31	16	20	17	7	12	5	23	215
Cerebral palsy	46	29	14	0	6	5	1	1	0	0	9	65
Talipes	20	3	1	0	0	0	2	0	0	0	14	20
Epilepsy	16	2	0	5	0	0	0	4	0	1	5	17
Down syndrome	10	4	0	3	4	1	1	3	0	0	0	16
Hypotonia	10	3	0	0	6	0	0	0	0	1	4	14
Microcephaly	9	0	1	1	1	1	0	0	2	0	3	9
Hydrocephaly	8	2	1	3	1	0	0	0	0	1	3	11
Blindness	5	0	0	0	1	0	1	0	0	0	3	5
Migraine	4	0	2	0	0	0	0	0	0	1	2	5
Muscular dystrophy	4	2	0	0	0	0	0	0	0	0	3	5
Polydactyly	4	0	0	0	0	0	0	0	0	0	4	4
Total	**253**	**93**	**55**	**43**	**35**	**27**	**22**	**15**	**14**	**9**	**73**	**386**

[1] categories with four or more syndromic cases are presented; #nine most common associated anomalies are reported.
*P<0.0001

Table 2.6: Congenital malformation in association with parity

Malformation	Parity							
	1	**2**	**3**	**4**	**5**	**6**	**7 and above**	**Total**
Neurological disorders	141	102	72	48	33	28	31	455
Limb defects	91	66	45	35	21	10	10	278
Musculoskeletal defec	32	18	21	14	6	3	5	99
Sensorineural/ear defects	21	17	29	12	9	3	3	94
Blood disorde	16	19	16	5	7	7	4	74
Eye/visual impairments	6	7	6	3	3	4	2	31
Ectodermal anomal	12	7	1	4	0	1	1	26
Congenital heart defects	5	5	5	4	2	0	2	23
Others	5	9	3	2	1	1	2	23
Total	**329**	**250**	**198**	**127**	**82**	**57**	**60**	**1,103**

Pedigree analyses of the familial cases revealed that the malformations mostly segregated in just one generation (n=151), followed by segregation in two and three generations (88 and 40, respectively; Table 2.7).

Table 2.7: Disease segregating generations in Familial cases

	Generation with disease					
Malformation	**I**	**II**	**III**	**IV**	**V**	**Total**
Neurological disorders	50	20	42	1	0	113
Limb defects	32	34	22	10	1	99
Musculoskeletal defects	15	9	27	2	0	53
Sensorineural/ear defects	15	12	24	0	0	51
Blood disorders	19	2	4	0	0	25
Eye/visual impairments	8	1	7	3	1	20
Ectodermal anomalies	8	5	7	0	0	20
Congenital heart defects	3	2	0	0	0	5
Others	1	3	7	1	0	12
Total	**151**	**88**	**140**	**17**	**2**	**398**

Further, the pedigree analysis revealed that the number of affected sibships in the pedigrees was mostly single followed by two, three, four and five sibships (n=183, 84, 44, 28, 24, respectively; Table 2.8).

Table 2.8: Sibship with disease in Familial cases

	Sibship with disease							
Malformation	**1**	**2**	**3**	**4**	**5**	**6**	**7 and above**	**Total**
Neurological disorders	59	25	10	7	1	6	5	113
Limb defects	31	29	12	7	7	2	4	92
Musculoskeletal defects	20	13	12	1	6	0	0	52
Sensorineural/ear defects	27	8	4	6	3	1	0	49
Blood disorders	23	2	0	0	0	1	0	26
Eye/visual impairments	8	1	1	2	5	0	2	19
Ectodermal anomalies	9	3	2	3	1	1	1	20
Congenital heart defects	4	0	1	0	0	0	0	5
Others	2	3	2	2	1	2	1	13
Sum	**183**	**84**	**44**	**28**	**24**	**13**	**13**	**389**

2.5: Discussion

Major types of CA affect roughly 2% to 3% of all newborns, with varying frequencies in various populations (Rankin *et al.,* 2005). Depending on the type of defect, congenital malformations or birth defects might be identified soon after delivery or later. Congenital malformations are a major cause of newborn mortality and morbidity. High-income countries have built accurate monitoring frameworks to determine the prevalence of CA so that effective preventive strategies can be developed (Gul *et al.,* 2012). Congenital and inherited abnormalities are common in Pakistan. The healthcare system is unable to handle the number, providing limited, if any support for affected individuals or families. This results in significant social, economic, and psychological consequences for affected families and society generally. CA are associated with 6–9% of perinatal fatalities in Pakistan (Korejo *et al.,* 2007).

In the present study, a total of 1,189 individuals afflicted with different congenital anomalies were recruited from different areas of the Hazara division. Neurological disorders were the most prevalent (41%) then limb defects (25%) and musculoskeletal defects (9%). Intellectual disability (0.148%) and cerebral palsy (0.124%) were common across all neurological disorders. All types of mild to severe mental retardation, with any other associated anomaly, were merged into intellectual disability. Our findings support two previous studies, conducted in tertiary care hospitals under Peshawar and Kurram tribal agency, northwest Pakistan (Khan *et al.,* 2015, Zahra *et al.,* 2016), which documented neurological disorders as prominent categories of CA. Pakistan has one of the highest levels of children with intellectual disability (Mirza *et al.,* 2009) and, in our sample and an earlier study (Kaufman *et al.,* 2010), with predominantly males affected. Ageing mothers, minimal education, low

socio-economic status, rural origin, less access to healthcare and high rates of cousin marriages are risk factors for increased rates of ID in developing countries, including Pakistan. According to the birth history taken at the time of data collection, birth asphyxia and delayed crying, associated with birth asphyxia, were the two major associations with ID. ID is strongly linked to a delayed neonatal crying at the time of birth, especially in impoverished nations, including a third of cases reported in Indian research, and a third of cases in a Lahore study (Kramer, 2008; Omar and Kokab, 2019). Premarital counselling, genetic screening sessions, health education, and the provision of antenatal and perinatal care, which is a difficult task, are all likely to help prevent these risks.

In this study, males subjects were more represented than females (n=1284 *vs* 928). Male predominance among congenitally deformed babies has been shown in a number of previous studies (Omar and Kokab, 2019; Sokal *et al.,* 2014; Rankin *et al.,* 2005). Pregnancies affected by birth defects were more prevalent among male babies than females (55% *vs* 45%, respectively). In the present study, isolated and syndromic cases were more prevalent than the familial and syndromic affected subjects. These findings support the results of earlier studies and provide evidence for the contribution of environmental factors as a risk for the development of CA.

In the present study, limb defects were the second most common group of CA (n=292). Preveiously reported study in Sialkot demonstrates that limb defects are prevalent among all types of CA (47%; Bhatti *et al.,* 2019). While Zahra et al reported limb defects as the third most common type of CA (21%) after neurological (34%) and musculoskeletal defects (23%; Zahra *et al.,* 2016). According to the present estimate talipes (0.119%) were prominent following polydactyly preaxial and postaxial (0.029% and 0.026%). The majority of

limb/digit disorders do not cause severe disability, and because of their modest nature, they are underreported in epidemiological research (Lal and Malik 2015; Khan *et al.,* 2015; Gillani *et al.,* 2011) Contradictory to this study represents the greater number of limb defects causing the disability (Talipes=141, Transverse limb amputations=23, Leg length discrepancy=4, constriction band syndrome=3 and thumb hypoplasia/aplasia=3) compare to milder defects (polydactyly=65, syndactyly=18, brachydactyly=10 , clinodactyly and camptodactyly=16). However, Bhatti *et al.,* reported the amputation/reduction defects as the most prevalent limb defect in the population of Sialkot. While other studies witnessed for high preponderance for polydactyly (Lal and Malik 2015; Ullah *et al.,* 2015).

The next category represents the musculoskeletal disorders with an overwhelming number of muscular dystrophy (n=106) followed by other disorders including Hypotonia (limbs)/myopathies, dwarfism and congenital hip dysplasia which were another cause of physical disability in affected subjects. Musculoskeletal and limb defects have been identified as one of the most common types of malformations, according to studies. Pakistan has a higher burden than Western countries due to insufficient diagnostic, treatment, and rehabilitation facilities. An understanding of the different forms and nature of malformations is a first step toward developing public health policies (Ullah *et al.,* 2015).

In the Sialkot District, Azhar et al (2011) carried out interventional research on disabled subjects, focussing on surgical procedures, physiotherapy, and bracing used to help disabled people rehabilitate, but with no attempt to explain the nature and clinical spectrum of deformities. Poliomyelitis was the most common disability among the 644 people with impairments, followed by CP, skeletal dysplasia, muscular dystrophy, congenital dislocated hip, and talipes equinovarus. In the

current study we excluded cases of poliomyelitis from data collection. A general observation made during the current data collection was that affected infants with talipes are usually treated during infancy and have a surgical correction to rehabilitate them.

In the current cohort, parental consanguinity was calculated to be 66%, with congenital heart defects patients having the highest prevalence (81%). These results are consistent with a study by Zahra *et al*. (2016), in which inbreeding was the commonest finding in CA of congenital heart defects and deaf/mute individuals. Different studies conducted in various Pakistani populations reported the consanguinity rate from 57-62% (Riaz *et al.,* 2016; Hina and Malik, 2015; Jabeen and Malik, 2014). However, the greater frequency of CA in Pakistani society is largely attributed to a high percentage of paternal consanguinity (Masood *et al.,* 2011). But various categories of CA were not significantly associated with consanguineous unions in present statistics.

The syndromic cases were further analyzed in order to observe the associated anomalies. In the overall cohort, there were at least 495 distinct associated malformations. These analyses showed that the six most common associated malformations were speech apraxia/stuttering, epilepsy, hypotonia, talipes, squint eyes, and deaf/mute. These findings give the clue about genetic heterogeneity in disease etiology.

Individually, Mendelian or monogenic disorders were uncommon, but as a group were very common. Infectious/communicable diseases are reducing globally as a result of improved surveillance and rigorous treatment but inherited and genetic abnormalities are becoming a huge burden on healthcare systems (WHO, 2014). A

high prevalence of genetic abnormalities has been documented in Pakistan. Multigenerational pedigrees with numerous people displaying rare genetic disorders have emerged from a high proportion of consanguinity, overlapping generations, and stable communities (Jabeen and malik 2014; Malik *et al.,* 2014).

It could be helpful to limit known reported risk factors linked with CA, such as advanced maternal age, smoking, low education, mother illness, and exposure to radiation and smoking, by educating women generally as well as pregnant women and providing them with effective antenatal care. In addition, a number of screening methods, such as maternal serum marker determination, ultrasonography, amniocentesis, and chorionic villus collection, can be utilized to identify and manage at-risk pregnancies. Congenital anomalies are a leading cause of infant death, therefore determining the frequency and prevalence of congenital abnormalities in society is crucial (Bhatti *et al.,* 2019). Consanguinity was the most frequently reported risk factor. Since neurological problems were the most common anomaly, therefore early prenatal diagnosis is extremely beneficial in reducing perinatal mortality by allowing for the choice of early pregnancy termination. The present data set represented the frequency of congenital malformations and self-reported risk factors for congenital malformations in the Hazara division of Khyber Pakhtunkhwa (KP), Pakistan. More research is needed to evaluate interventions aimed at eliminating risk factors and reducing the prevalence of congenital abnormalities (Khan *et al.,* 2015).

3. Phenotypic spectrum and descriptive genetic study of a large cohort of congenital limb deficiencies from North-West Pakistan

3.1: Abstract

Congenital limb deficiency (CLD) is a group of very rare disorders characterized by substantial hypoplasia or the complete absence of one or more bones of limbs. CLD render a significant physical, clinical and psychological burden to affected subjects and their families. The aim of the current clinico-descriptive study was to characterize the prevalence-pattern, phenotypic manifestations, and risk factors associated with CLD in a cohort assembled from North-West Pakistani population which bears high burden of such anomalies. Through a cross-sectional study, 141 independent subjects involving 166 limbs with CLD were recruited during 2017-2021. Subjects with transverse defects were n=77 (55%), longitudinal defects n=61 (43%), and intercalary defects n=3 (2%). Among the subjects with transverse defects, terminal amputations and symbrachydactyly were n=52 and n=25, respectively; whereas among the longitudinal defects, thumb aplasia/hypoplasia was the most common presentation (n=20), followed by oligodactyly (n=18), and radial hemimelia (n=18). Upper limb deficiencies were 86% compared to lower limb deficiencies, 14%. There was a high preponderance of unilateral cases (83%), sporadic nature (92%), and low parental consanguinity rate (33%) and syndromic appearance in 21% cases, which may be indicative of the substantial role of non-genetic factors in the etiology of CLD. This study demonstrates marked heterogeneity in CLD subtypes in the involvement of limbs and associated bio-demographic variables. It is foremost important to establish a proper national registry for subjects CLD and to promote the development of molecular genetic diagnostics and therapeutic approaches and social support of such subjects.

3.2: Introduction

Congenital limb deficiency (CLD) is a general term for a range of disorders characterized by substantial hypoplasia or aplasia of one or more bones of limbs. CLD can be further divided into entities based on which anatomical segments are affected. The particular terminology that should be assigned to each limb reduction defect is still being debated (Lowry and Bedard, 2016; Gold *et al.,* 2011). Limb loss is categorized into two categories: major and minor. A transhumeral, trans-radial, transfemoral, or transtibial amputation is a major limb loss. On the other hand, amputation of the hand, digits, toes, or midfoot level is considered minor limb loss (Varma *et al.,* 2014; Tseng *et al.,* 2007).

3.2.1: Classification of CLD

Limb deficiencies have two categories by conventional nomenclature: longitudinal and transverse (Gold *et al.,* 2011). Longitudinal deficiencies, such as the lack of the radius, are aplasia or hypoplasia of a bone along the long axis of the limb. A transverse deficiency, on the other hand, is the absence of limb parts distal to a certain level across the long axis of the limb, such as a foot or a hand amputation. Transverse deficiencies are commonly referred to as preaxial (radial and tibial side), postaxial (ulnar and fibular side), and central (Wilcox *et al.,* 2015).

3.2.2: Prevalence of CLD

The prevalence of limb amputation has been reported to vary from one place to the other due to different classification schemes adopted and the methodologies employed in their ascertainment. Despite methodological differences, the overall prevalence of CLD is quite similar across studies, with most falling in the 5–7/10,000 total births range (Bedard *et al.,* 2015). CLDs vary in anatomic location, nature, and

etiology, and are strongly associated with long-term functional impairment as well as have a significant economic impact (Boonstra *et al.,* 2000).

Epidemiological studies help in the timely detection of trends in congenital limb defects and the association with other anomalies. The studies on the prevalence of CLDs from developing countries like Pakistan are missing.

3.2.3: Morbidity and impact of CLD

Limb amputations are a leading cause of disability and have a significant anatomical, physical, and financial burden on the patient (Weir *et al.,* 2010). The loss of a limb by anyone, especially in poorer nations with inadequate prosthetic facilities, has devastating economic, social, and psychological consequences for the patient and their family (Chalya *et al.,* 2012; Soomro *et al.,* 2012). Amputations of major limbs are essentially disfiguring operations with a high rate of perioperative death and morbidity, and people who have had amputations are frequently considered incomplete people (Keramat *et al.,* 2021; Masood *et al.,* 2008). Limb amputation varies in anatomic location, type, and etiology, and is frequently linked to life-long functional impairment (Boonstra *et al.,* 2000).

Like any handicap, societal and environmental issues, as well as a lack of services such as transportation, a good prosthesis, and social support, aggravate an amputee's psychological situation (McFarland *et al.,* 2010). However, in industrialized countries and for an urban part of the population in Pakistan, rehabilitation facilities for amputees are increasing daily, resulting in good functional outcomes and community integration (Rathore *et al.,* 2016). However, statistics on amputees are few, particularly in developing countries including Pakistan.

3.2.4: Etiology/risk factors for CLD

The majority of amputations are isolated and characterized as disruptive defects. Known causes of CLD included defective genes (e.g. in Fanconi anemia), chromosomal anomalies (e.g.trisomy 18), teratogenic or other environmental exposures including medications such as thalidomide and misoprostol, as well as risky prenatal diagnostic procedures including chorionic villus sampling (C*VS*; Gold *et al.,* 2011).

The current study aimed to investigate the prevalence pattern of a large cohort of CLD assembled from Khyber Pakhtunkhwa (KP) province of Pakistan and to report clinical heterogeneity and association with other congenital anomalies and socio-demographic attributes.

3.3: Subjects and Methods

3.3.1: Ethical consideration

The Ethical Review Committee of Quaid-i-Azam University gave their approval to this research. The study followed international guidelines for human subject protection, including the principles outlined in the most recent (2008) and earlier amendments to the Declaration of Helsinki of 1964, as well as the Council for International Organizations of Medical Sciences (CIOMS) International Ethical Guidelines for Biomedical Research Involves Humans (2002).

3.3.2: Data source and sample collection

A cross-sectional study focusing on subjects with various congenital anomalies was undertaken from September 2018 to March 2020. Participants were recruited by convenient sampling from different areas of Khyber Pakhtunkhwa (KP) province.

Door-to-door surveys and visits to public sites such as educational institutes, hospitals, and community centers were performed. Before collecting data, each participant or his/her parents or guardians gave written consent in any situation, the participant's assent was always obtained. Field visits were conducted with the help of a resource person or a local health visitor who assisted in the collection of documented consent and clinical data. The researchers and enumerators all had formal medical genetics expertise and all data were obtained with informed consent.

The current study followed the cross-sectional reporting guidelines of STROBE (Strengthening the Reporting of Observational Studies in Epidemiology) declaration (von Elm *et al.,* 2008). Data on demographic characteristics and limb abnormalities were recorded on a structured proforma. Local medical personnel

examined individuals with limb anomalies physically. Previous medical records were retrieved when they were available, and case histories were gathered. Individuals who were not permanent residents of the KP or gave inaccurate information were excluded from the study.

A structured questionnaire was used to collect clinical, bio-demographic, and family history data. The Pakistan Demographic and Health Survey (NIPS, 2019), as described elsewhere, was used as a standard to define demographic variables. A comprehensive pedigree was created in each case. With an inbreeding value of F=0.0156, parental marriage types up to the second cousin were considered consanguineous (Jabeen and Malik, 2014; Bittles, 2010).

3.3.3: Inclusion exclusion criteria

All patients, regardless of age or gender, who had a congenital limb amputation and gave their consent were included. Patients who refused permission or who had already undergone surgery at another facility but required stump revision were excluded from the research.

3.3.4: Definitions and classification of anomalies

Gold *et al.,* (2011) introduced a novel classification method that encompasses all possible phenotypes. Infants with limb deficiency were documented in the hospital-based Active Malformations Surveillance Program at Brigham and Women's Hospital in Boston, from 1972 to 1974 and 1979 to 2000. This classification was the expansion of an earlier classification scheme conducted in the same hospital but was based on the small number of infants. In the present study, we adapted the classification scheme presented by Gold *et al.,* 2011. Further, the Online Mendelian

Inheritance in Man (OMIM) database was also used to classify CLD, and the closest definition was found in the International Classification of Disease codes (ICD-10) (OMIM; ICD-10).

3.3.5: Statistical analysis

The distribution of anomalies was examined across the biodemographic factors, with a significant limit of $p<0.05$. Chi-test and Fisher exact tests were used to assess the independence of discrete variables. Statistical analysis was carried out using GraphPad Prism Software.

3.4: Results

3.4.1: Sample characteristics

A total of 141 subjects (91 males, 50 females) with CLD were recruited and a total of 166 limb extremities were involved. The age of subjects was 17.2±12.8 (mean±St.Dev.) years and there were 78 subjects with age up to 15 years. The majority of the subjects (83%) had a rural origin, spoke the Pushto language (60%), and belonged to poor and low socioeconomic strata (52%), and extended families (56%; Table 3.1). The district-wise distribution across gender was statistically significant ($p = 0.005$). Further, the distribution of CLD concerning mother tongue across familial sporadic categories was also statistically significant ($p =0.003$; Table 3.1).

Table 3.1: Sample characteristics and demographic distribution across gender and sporadic/familial categories

Variable	**Gender (n)**		**Sporadic/familial nature (n)**		**Total CA**	
District	**Male**	**Female**	**Sporadic**	**Familial**	**No.**	**%**
Haripur	5	13	17	1	18	13
Chitral	16	2	18	0	18	13
Charsadda	14	3	16	1	17	12
Nowshera	11	4	14	1	15	11
Kurram Agency	5	7	11	1	12	9
Bajaur Agency	9	2	11	0	11	8
Others	31	19	43	7	50	35
Sum	**91**	**50**	**130**	**11**	**141**	**100**
	P=0.001		P=0.497			
Age range (years)						
Up to 9	22	18	39	1	40	28
>9-19	37	20	52	5	57	41
>19	32	12	39	5	44	31
	P=0.237		P=0.299			
Rural/urban origin						
Rural	77	40	111	6	117	83
Urban	14	10	19	5	24	17
	P=0.485		P= 0.009			
Mother tongue						
Pashto	58	26	77	7	84	59
Hindko	10	19	28	1	29	20
Khowar	16	2	18	0	18	13
Urdu	3	2	2	3	5	4
Others	4	1	5	0	5	4
	P=0.009		P=0.003			
Literacy (age ≥5 years)						
Illiterate	26	11	34	3	37	27
Literate	64	35	91	8	99	73
	P=0.684		P=0.996			
Literacy level						
Primary	24	13	34	3	37	37

Matriculate	26	12	35	3	38	38
Secondary and higher	14	10	22	2	24	24
	P=0.720		P=0.998			
Marital status (age ≥17 yrs)						
Single	28	11	34	5	39	70
Married	12	5	16	1	17	30
	P=0.927		P=0.440			
Economic status						
Poor	39	21	56	4	60	43
Low	10	3	13	0	13	9
Mid	33	20	50	3	53	38
High	10	5	12	3	15	11
	P=0.800		P=0.174			
Family structure						
Nuclear	35	27	57	5	62	44
Extended	56	23	73	6	79	56
	P=0.080		P=0.918			
Parental consanguinity						
Yes	30	16	40	6	46	33
No	61	34	90	5	95	67
	P=0.907		P=0.106			

3.4.2: Limb phenotypes

Subjects with longitudinal defects were 43.3% (n=61), transverse defects 54.6% (n=77), and intercalary defects accounted for 2.1% (n=3) of the cases.

Among the longitudinal defects, six entities were established (Tables 3.2, 3.3). The commonest was thumb aplasia/hypoplasia (20 subjects), then oligodactyly (n=18), and radial hemimelia (n=18). In both upper and lower limbs, right and left sides were almost equally involved. Seven subjects with oligodactyly had bilateral presentations. Preaxial deficiencies were more common than preaxial types (48 *vs*. 13; Fig. 3.1).

Among the transverse defects, two broad entities, i.e., terminal amputations and symbrachydactyly were established (Tables 3.2, 3.3). Most of the subjects had deficiencies at the point of mid-hand (i.e., symbrachydactyly, n=25), followed by wrist (with or without nubbins; n=19), elbow (n=9), and forearm (n=9). Among the subjects with symbrachydactyly, adactyly and monodactyly were the most common phenotypes (n=9 each). In terminal deficiencies, the left arm was twice as involved as the right arm (30 *vs*. 16; p=0.205) (Fig. 3.1).

Intercalary defects of upper limbs were witnessed in 3 subjects (Table 3.2) but amelia was not found in this cohort. Sporadic cases were predominant (92.2%) and only 7.8% cases had familial occurrence (Table 3.2). An autosomal recessive inheritance pattern was evident in 6 of 11 familial cases. Oligodactyly had the highest presentation of familial cases (16.7%).

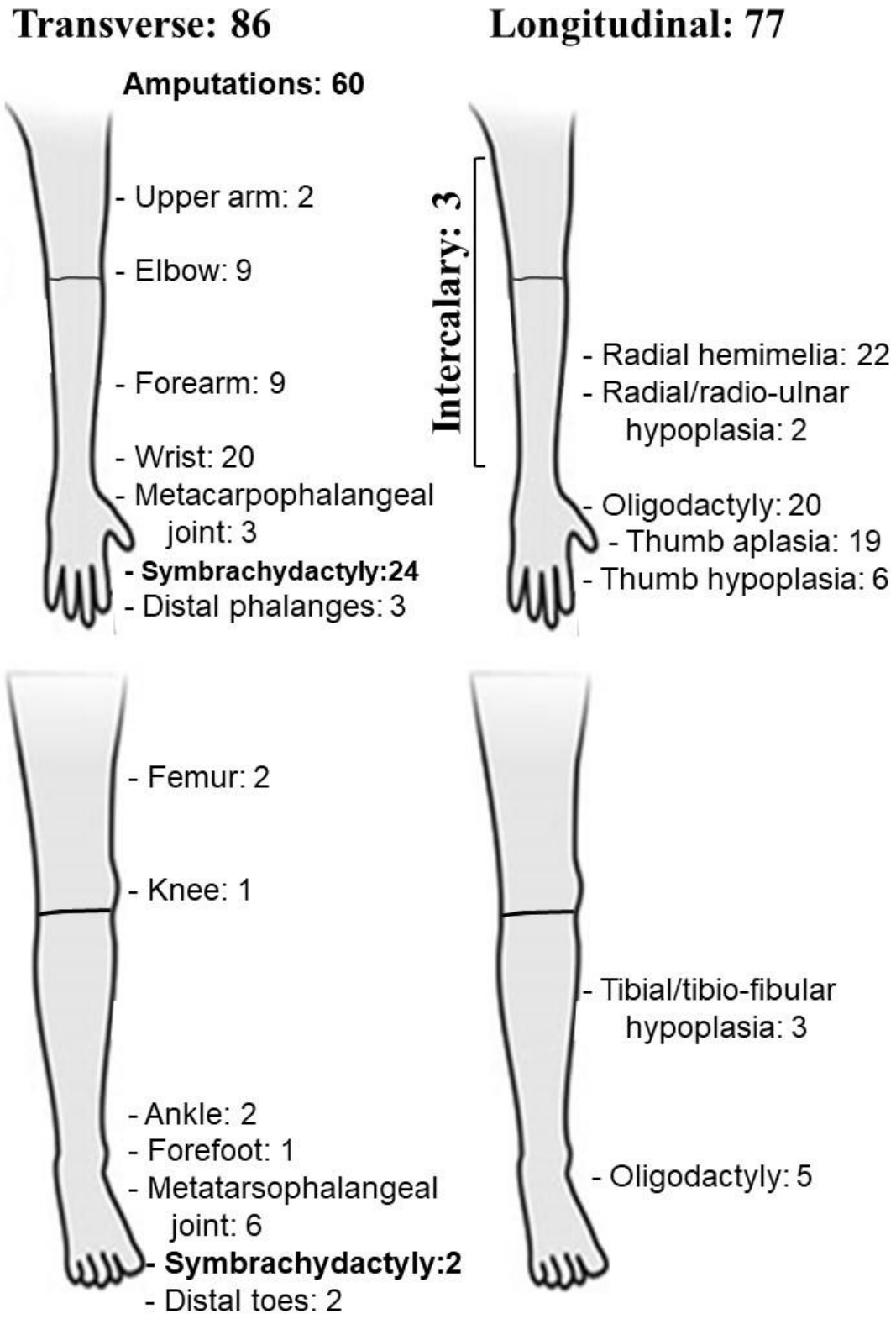


Fig 3.1: Schematic depicting the distribution and frequency of CLD in upper and lower limb segments.

Table 3.2: CLD and subtypes in gender-wise and familial/sporadic samples

Congenital limb deficiency	No.	Proportion	95% CI	Gender*		Familial/sporadic nature *	
				Male	Female	Sporadic	Familial
Longitudinal defects	**61**	**0.433**	**0.351-0.514**	**43**	**18**	**55**	**6**
Radial hemimelia	**18**	0.128	0.073-0.183	12	6	16	2
Radial/radio-ulnar hypoplasia	**2**	0.014	-0.005-0.034	2	0	2	0
Oligodactyly	**18**	0.128	0.073-0.183	13	5	15	3
Thumb aplasia	**14**	0.099	0.050-0.149	11	3	13	1
Thumb hypoplasia	**6**	0.043	0.009-0.076	3	3	6	0
Tibial/tibio-fibular hypoplasia	**3**	0.021	-0.003-0.045	2	1	3	0
Transverse defects	**77**	**0.546**	**0.464-0.628**	**47**	**30**	**73**	**4**
Terminal amputations	**52**	0.369	0.289-0.448	28	24	49	3
Symbrachydactyly	**25**	0.177	0.114-0.240	19	6	24	1
Intercalary	**3**	**0.021**	**-0.003-0.045**	**1**	**2**	**2**	**1**
Total	**141**	**1.000**	**-**	**91**	**50**	**130**	**11**

*Chi-distributions in major CLD categories and subtypes were statistically not significant

Most of the subjects (n=119; 83%) were presented with only one limb deficiency (i.e., unilateral) and there were 20 subjects (18%) with the involvement of two limbs; involvement of three and four limbs was witnessed in one subject each. The upper limb deficiencies were 86% compared to the lower limb deficiencies 14%. In the upper limb, the left side was relatively more commonly involved than the right side (79 *vs*. 63; Table 3.3).

Table 3.3: Detailed manifestations, involved limbs and laterality in CLD

Congenital limb deficiency (major and minor division)	No. of subjects	Involved limb (n)				Total limbs (n)	Laterality (n)		ICD-10
		Right arm	Left arm	Right leg	Left leg		Unilateral	Bilateral	
Longitudinal defects	**61**	**34**	**35**	**4**	**4**	**77**	**45**	**16**	
Radial hemimelia	18	11	11	0	0	22	14	4	Q71.4
Radial/radio-ulnar hypoplasia	2	1	1	0	0	2	2	0	Q71.4
Oligodactyly	18	10	10	2	3	25	11	7	
Oligodactyly, pre-axial	5	5	4	0	0	9	1	4	Q71.3
Oligodactyly, post-axial	8	2	3	1	2	8	8	0	Q71.30
Oligodactyly, central	3	2	2	0	0	4	2	1	
Oligodactyly, pre-/post-axial	2	1	1	1	1	4	0	2	Q71.30, Q71.31
Thumb aplasia	14	9	10	0	0	19	9	5	Q71.31
Thumb hypoplasia	6	3	3	0	0	6	6	0	Q71.31
Tibial/tibio-fibular hypoplasia	3	0	0	2	1	3	3	0	Q72.5
Transverse defects	**77**	**28**	**42**	**9**	**7**	**86**	**71**	**6**	
Transverse terminal amputations	52	16	30	8	6	60	47	5	
Upper arm	2	1	1	0	0	2	2	0	Q71.2
Elbow	9	1	8	0	0	9	9	0	Q71.2
Forearm	9	3	6	0	0	9	9	0	Q71.2
Wrist, no nubbins/ankle	15	8	8	1	1	18	13	2	Q71.2
Wrist, nubbins	4	1	3	0	0	4	4	0	Q71.2
Metacarpophalangeal joint	3	1	2	0	0	3	3	0	Q71.30
Distal phalanges/toes	2	1	2	1	1	5	1	1	Q71.30, Q72.30
Femur	2	0	0	2	0	2	2	0	Q72.2
Knee	1	0	0	0	1	1	1	0	Q72.2
Forefoot	1	0	0	1	0	1	1	0	Q72.3
Metatarsophalangeal joint	4	0	0	3	3	6	2	2	Q72.3

Symbrachydactyly (including midhand)	25	12	12	1	1	26	24	1	Q71.30
Symbrachy.; adactyly	9	6	3	0	0	9	9	0	Q71.30
Symbrachy.; monodactyly	9	3	5	1	1	10	8	1	Q71.30
Symbrachy.; bidactyly	4	1	3	0	0	4	4	0	Q71.30
Symbrachy.; multidigit type	3	2	1	0	0	3	3	0	Q71.30
Intercalary	**3**	**1**	**2**	**0**	**0**	**3**	3	0	Q71.1
Total	**141**	**63**	**79**	**13**	**11**	**166**	**119**	**22**	

3.4.3: Parental consanguinity

Parental consanguinity was observed in 33% of the cases (Table 3.4). Among the longitudinal defects, consanguinity in oligodactyly was observed to be 44% compared to 21% in thumb aplasia. Among the transverse defects, consanguinity in terminal amputations was 35% compared to 28% in symbrachydactyly. Among the major categories of CLD, the differences in the distribution of parental consanguinity were statistically not significant (Table 3.4).

Table 3.4: Parental consanguinity in CLD subtypes

Congenital limb deficiency	**No. of cases**	**Parental consanguinity***	
		No.	**%**
Longitudinal defects	**61**	19	31
Radial hemimelia	**18**	2	11
Radial/radio-ulnar hypoplasia	**2**	1	50
Oligodactyly	**18**	8	44
Thumb aplasia	**14**	3	21
Thumb hypoplasia	**6**	2	33
Tibial/tibio-fibular hypoplasia	**3**	3	100
Transverse defects	**77**	27	32
Terminal amputations	**52**	18	35
Symbrachydactyly	**25**	7	28
Intercalary	**3**	2	67
Total	**141**	46	33

*Chi-distributions in major CLD categories and subtypes were statistically not significant

3.4.4: Birth order

The majority of the cases (33%) fall in the first birth order, followed by second (20%) and third birth order (12%) (Table 3.5). In the overall cohort, the mean maternal and paternal ages for affected births were calculated to be 26.89±0.66 and 31.53±0.73, respectively ($p<0.0001$). Statistically significant different mean maternal and paternal ages were found for affected individuals with thumb aplasia and transverse amputations.

Table 3.5: Parity of the subjects with CLD

Congenital limb deficiency	**No. of cases**	**Birth order**					
		1st	**2nd**	**3rd**	**4th**	**5th**	**6th and higher**
Longitudinal defects	**50**	**19**	**11**	**4**	**6**	**5**	**5**
Radial hemimelia	12	5	4	0	0	2	1
Radial/radio-ulnar hypoplasia	2	1	1	0	0	0	0
Oligodactyly	16	7	2	1	3	1	2
Thumb aplasia	13	4	2	3	2	2	0
Thumb hypoplasia	4	2	1	0	0	0	1
Tibial/tibio-fibular hypoplasia	3	0	1	0	1	0	1
Transverse defects	**68**	**19**	**13**	**11**	**8**	**5**	**12**
Terminal amputations	46	11	8	9	6	3	9
Symbrachydactyly	22	8	5	2	2	2	3
Intercalary	**3**	**2**	**0**	**0**	**0**	**1**	**0**
Total	**121**	**40**	**24**	**15**	**14**	**11**	**17**

3.4.5: Syndromic cases and associated anomalies

CLD associated with any other anomaly was considered as syndromic and was observed to be 21% (n=30). The frequency of syndromic cases was ranging from 28% in longitudinal defects to 17% in transverse defects (p=0.120). The distribution of syndromic cases in various ascertainment categories was statistically not significant. The most common associations were brachydactyly (n=5), cliodactyly (n=4), low IQ (n=4), camptodactyly (n=3), and syndactyly (n=3; Table 3.6).

Table 3.6: Distribution of Isolated and syndromic cases in CLD

Variable*	**No. of cases**	**Syndromic nature (n)**	
		Isolated	**Syndromic**
Congenital limb defects			
Longitudinal defects	**61**	44	17
Radial hemimelia	**18**	13	5
Radial/radio-ulnar hypoplasia	**2**	2	0
Oligodactyly	**18**	12	6
Thumb aplasia	**14**	12	2
Thumb hypoplasia	**6**	2	4
Tibial/tibio-fibular hypoplasia	**3**	3	0
Transverse defects	**77**	64	13
Terminal amputations	**52**	42	10
Symbrachydactyly	**25**	22	3
Intercalary	**3**	3	0
Total	**141**	**111**	**30**
Gender			
Male	91	72	19
Female	50	39	11
Familial/sporadic nature			
Sporadic	130	104	26
Familial	11	7	4
Parental consanguinity			
Yes	46	38	8
No	95	73	22

*Chi-distributions in all variables were statistically not significant

3.5: Discussion

CLD renders a significant physical, clinical and psychological burden to affected subjects and their families. These subjects have a visible abnormal appearance which might negatively influence their quality of life and participation (Michielsen et al. 2010). There is no prevalence estimate for CLD in the Pakistani population. Due to their highly rare nature, this group of malformations has not gained much attention among the researchers (Riaz et al. 2014; Lal and Malik, 2015).

In the majority of the published studies on CLD, parental consanguinity and familial occurrence have not been reported (Goldfarb *et al.,* 2015). Consanguinity has been shown to have a substantial contribution to the emergence of rare recessive disorders in the Pakistani population. Consanguinity rate was observed to be 33% in the present cohort which is very low compared to the national average (i.e., 60%) as well as estimates available for the north-western populations of Pakistan (Ahmad *et al.,* 2016; Tufail *et al.,* 2017; Nawaz *et al.,* 2021). Secondly, familial cases were very low in number compared to the sporadic cases (8% *vs*. 92%, respectively). In the cohort studied by Gold *et al.,* (2011) familial occurrence was 3%. These observations give clues to the substantial role of non-genetic etiological factors for CLD. Here, the role of de novo mutation in the pathogenesis of a handful number of subjects with CLD cannot be ruled out.

In our cohort, transverse defects were relatively more common compared to longitudinal defects (55% *vs*. 43%, respectively). Likewise, Al-Worikat and Dameh (2008) studied CLD among Jordanian patients and observed that transverse and longitudinal defects were 54% and 46%, respectively. In the experience of Jain (1994), who recruited 200 Indian patients with CLD, transverse deficiencies were

observed in 75% of subjects compared to longitudinal deficiencies in 18% of patients. Gold *et al.,* (2011) recruited a total of 135 subjects with CLD; longitudinal defects were 54% and terminal transverse defects were 29%. Furthermore, the prevalence of specific CLD also differs substantially among the genders. In the present cohort, transverse terminal amputations appeared in nearly equal proportions among the male and female patients, whereas symbrachydactyly was thrice as common in males as compared to females. In the cohort studied by Jain (1994), transverse leg deficiencies were common in male patients while transverse forearm deficiencies were common in female patients.

Here, upper limb deficiencies were 86% compared to lower limb deficiencies 14%. This is consistent with the observation of Sener *et al.,* (1999). Among 417 cases with CLD, Mano *et al.,* (2018) observed that 67% of subjects had upper limbs affected, 18% had lower limbs affected, and 15% had both upper and lower limbs affected. Disparities in the phenotypic involvement of upper and lower limbs have been witnessed by several researchers. For instance, Jain (1994) demonstrated that in the upper limbs, transverse forearm deficiency (below elbow) and transverse phalangeal deficiency were the prevalent types, whereas, in the lower limbs, transverse metatarsal deficiency and transverse leg partial deficiency (below the knee) were most common. In the cohort recruited by Mano *et al.,* (2018), in the upper limbs, the transverse deficiency was the most prevalent compared to longitudinal deficiency (46% *vs*. 31%), whereas, in the lower limbs, the longitudinal deficiency was the most prevalent compared to the transverse deficiency (37% *vs*. 32%).

Among the longitudinal defects, Gold *et al.,* (2011) showed that pre-axial anomalies were more common than postaxial (29 *vs* 14), and transverse defects with metacarpophalangeal/metatarsophalangeal joints were common (16 of 39). This

phenotypic heterogeneity may be attributed to the etiological heterogeneity and genetic background of the study population.

Furthermore, the majority of the reported cohorts of CLD show the involvement of a single limb, followed by the involvement of two or more limbs. In the present study, 83% of subjects were presented with only one limb deficiency, followed by the involvement of two limbs (18%). This observation is close to Mano *et al.,* (2018), who witnessed that 67% of patients had one affected limb, 20% had two, 6% had three, and 7% had four affected limbs.

None of the patients in the present cohort had undergone any surgical correction or prosthetic intervention. It is quite alarming that adequate intervention and therapeutic options are not accessible to the majority of Pakistani patients with disabilities and particularly CLD. The quality of life of subjects with CLD can be substantially improved by various intervention measures. Mano *et al.,* (2018) showed that 67% of patients visited the hospital on the day of birth, and that patient visits decreased with increasing age. Prosthetic interventions were recommended at age 6–9 months in children with limb deficiencies (Krebs *et al.,* 1991).

In recent years, several advancements have been made by the development of high-throughput genome analyses. The genetic etiologies of CLD include single gene variants, small insertions and deletions, and copy number variants (Sun *et al.,* 2021; Gold *et al.,* 2011). Despite the technological advancements and owing to the etiological heterogeneity the molecular genetic diagnostic yield of reduction anomalies is usually low. In the experience of Sun *et al.,* (2021), 28% (7 of 25) of subjects with limb reduction anomalies were molecularly diagnosed.

3.6: Limitations of this study

Despite recruiting a large cohort, there are few limitations to this study. First, it does not cover the entire population of Pakistan and is restricted to convenient sampling. Hene, the presented number of subjects was either underestimated or overestimated. In addition, the current study does not focus on the risk factors and the genetic etiology of limb anomalies due to the lack of available diagnostic facilities.

3.7: Conclusion

The high presentation of unilateral and sporadic cases and low parental consanguinity are reminiscent of the substantial role of non-genetic factors in the etiology of the CLD in this population. It is vital to establish a proper national registry for CLD and to promote development of molecular genetic diagnostics and therapeutic approaches, genetic counseling, health and social support for affected individuals and families**.**

4. Analyses of whole genome SNP data, homozygosity mapping and identification of shared autozygous candidate intervals in a family with recessively segregating intellectual disability

4.1: Abstract

SNPs (single nucleotide polymorphisms) are the most abundant class of polymorphisms which are unevenly distributed across the genome. High throughput SNP genotyping methods have enabled the employment of these markers in variant identification and association studies. In the case of rare recessive disorders, SNP homozygosity mapping has become an important tool for the identification of candidate gene. This study was aimed to employ SNP based genotyping in an inbred Pakistani family with recessively segregating intellectual disability, in order to identify homozygous intervals shared among the affected subjects and to prioritize the gene hunt process for a prospective exome sequencing. Whole genome SNP genotyping data were generated for two affected cousins by using Illumina 730K map. Homozygous and heterozygous genotypes were manually identified through aligning data in MS-Excel and shared autozygous regions with >0.7 Mb were highlighted. The SNP data were uploaded on HomozygosityMapper online tool for comparative analysis. Identified homozygous interval by both manual method and HomozygosityMapper were tabulated and compared. Genome browser UCSC was employed to find the cytogenetic position of identified intervals and the regions were scrutinized through GeneDistiller. These analyses let to the identification of 12 homozygous intervals with >1Mb, the largest (54Mb) on chromosome 5p15.2-q12.3, followed by 39Mb stretch on 3q23-q26.32 and 22Mb on 4q31.22-q32.3. Several potential candidate genes related to the phenotype were present on these regions. Exome sequencing shall be launched in order to discover the underlying gene and variant segregating with the malformation in the family.

4.2: Introduction

4.2.1: SNP Genotyping

SNPs or single nucleotide polymorphisms affect specific locations within the genome that are found in more than 1% of the population. The average frequency of SNPs in the human genome is approximately one per 1000bp (Wang *et al.,* 1998; Brookes, 1999). SNPs are therefore the most abundant type of genetic variation, and their number exceeds 9 million in in the SNP databases (Rocha *et al.,* 2006). These are important markers, showing the association of sequence variation with phenotypic difference, expected to enhance understanding of human physiology, through understanding of the molecular basis of disease. SNPs are unevenly distributed across the genome and occur much less frequently in coding regions of the genome than in noncoding regions (Wang *et al.,* 1998). When present in noncoding region SNPs do not change encoded proteins, but serve as important markers for comparative or evolutionary genomics studies. When located in regulatory sitess of genes, SNPs can alter the encoded protein by affecting transcription. When in the coding regions, SNPs can change the structure and function of proteins that can lead to disease development or change in response to a drug or environmental toxin. Hence, SNPs are used as important markers both in genetic as well as in pharmacogenomic studies (Kim and Misra, 2007).

Identification of causative genes is important for better understanding of disease mechanisms. Using unrelated individuals, whole-genome association studies have significantly assisted in the assessment of prevalent diseases and the development of respective treatments. whereas linkage-based methods are effective for families with large number of affected individuals (Manolio, 2010; Ott, 1999).

Recently, focus has shifted to diseases that are challenging to study due to limited numbers of affected families and/or collecting sufficient number of samples. Such disorders include those having multiple rare genetic variants, genes with low penetrance, or genes with effects that manifest in later life (Hardy and Singleton, 2009; McCarthy *et al.,* 2008). To untangle genetic causes of such disorders there is a need to adopt approaches which are proving useful for small numbers of unrelated patients. The original goal of the homozygosity mapping (HM) method was to find disease-causing genes by screening patients from inbred families. This was developed and used with patients from outbred families (Lander and Botstein, 1987; Hildebrandt *et al.,* 2009; Browning *et al.,* 2010) and the effectiveness improved with the use of SNP data from genome-wide analysis (Huqun *et al.,* 2007; Miyazawa *et al.,* 2007; Huqun *et al.,* 2007; Miyazawa *et al.,* 2007). However, data correction is required because the algorithm used in HM is particularly susceptible to genotyping errors (Fukuyama *et al.,* 2010). In contrast, since it only employs a fraction of the SNP genotyping data, the homozygosity haplotype (HH) algorithm (Miyazawa *et al.,* 2007) is an imputation-free method for identifying haplotypes. There is a chance that two or more individuals who share an identical-by-descent (IBD) fragment in one or both strands of the homologous chromosomes have the same region of conserved homozygosity (RCHH). The algorithm is robust to genotyping errors, therefore little to no error correction is required (Hagiwara *et al.,* 2011).

Positional cloning has been used for the identification of disease gene. Low marker informativity as well as restricted number of meiosis limits the resolution of genetic mapping in rare diseases. Advances in high-density SNP microarrays for genotyping can overcome low marker informativity by using large numbers of markers to achieve greater coverage at finer resolution (Chiang *et al.,* 2006).

Affymetrix and Illumina SNP arrays are currently used for homozygosity mapping. Both work with different chemistry but still share several aspects (LaFramboise, 2009). SNP array-based homozygosity mapping were used in this analysis for the identification of candidate gene.

4.2.2: Advantages

A quick and inexpensive method that boosts the yield of mutation analysis from next-generation sequencing (NGS) is homozygosity mapping (HM). In cases of genetically heterogeneous disorders with aoutosomal recessive (AR) inheritance, like genodermatoses or deafness, HM reduces the requirement to sequence many genes. HM can explain the contribution of hitherto unidentified mutations, such as deep intronic or missense variants with ambiguous significance, to the pathogenesis of illness. HM is a high-throughput, genome-wide technique for finding new genes linked to disease (Hagiwara *et al.,* 2011).

4.2.3: Limitations

Since HM only works for Mendelian diseases with homozygous AR inheritance, HM is mainly used for patients born to consanguineous parents. Rare compound heterozygous mutations in consanguineous families will not be detectable in the region harboring the mutated gene by this method (Hagiwara *et al.,*2011).

In this study, SNP based genotyping was employed in an inbred Pakistani family with intellectual disability with an aim to identify homozygous intervals shared among the affected subjects and to streamline the gene hunt process in the prospective exome sequencing.

4.3: Subjects and methods

4.3.1: Study approval

This study formed part of an international collaboration and was approved by the institution review board of Quaid-i-Azam University, Islamabad, Pakistan (DAS-1070; Dated July 8, 2015). All Information was collected after taking informed and written consent according to the declaration of Helsinki II**.**

4.3.2: Family with intellectual disability

A family with intellectual disability (ID) was recruited from Southern Punjab. A family pedigree across the four generations reported was constructed by interviewing the elders of the family to confirm the information with the pedigree reviewed by their relatives. All affected individuals belong to fourth generation, including two males and three females and were product of three consanguineous couples (Fig. 4.1).

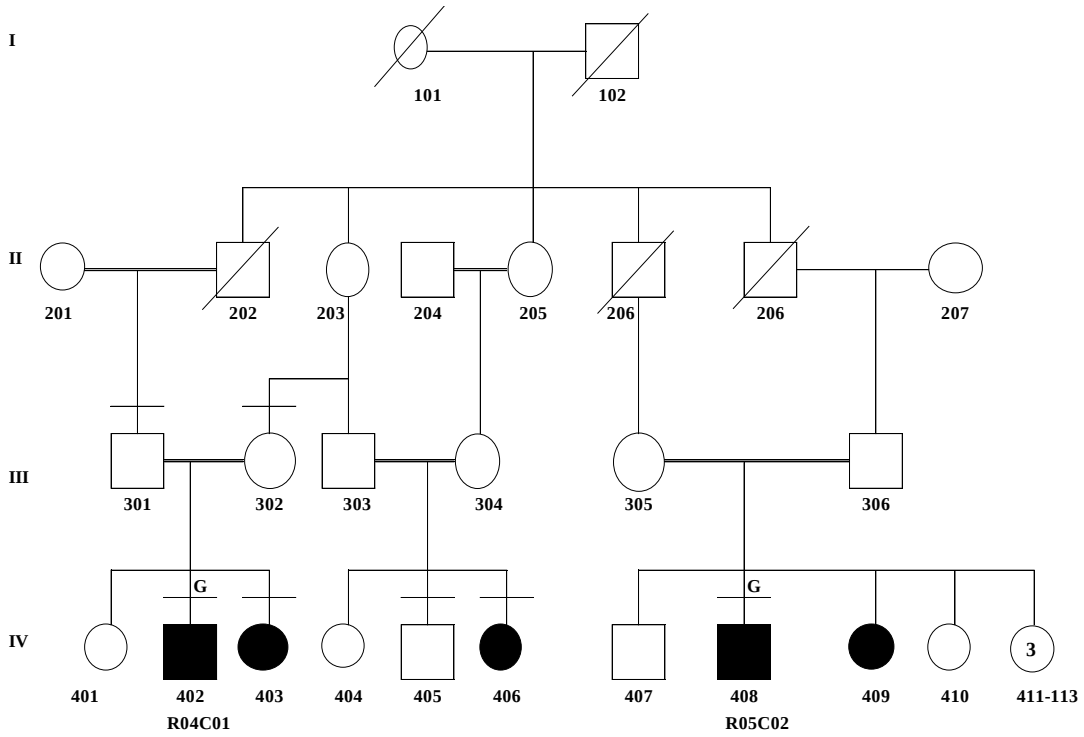


Fig 4.1: Pedigree of family with intellectual disability. Circles=females, squares=males, black filled symbols=affected, white symbols=unaffected individuals. Where physically examinations were carried out these are indicated with a horizontal line above the symbol. Symbols with 'G' denote the subjects who underwent SNP genotyping.

4.3.3: Clinical description of affected family subjects

Physical examination indicated the following consistent features in the diseased subjects including both structural and behavioral disabilities. Structural defects included developmental delay, learning late, walking late, toileting late, protruding forehead, squint eyes, digestive problems, prominent jaws and flat feet. Behavioral characters were poor self-care, poor communication, hyperactivity, mood instability, poor concept of money and social responsibility, and sensitivity to crowd (Table 4.1).

Table 4.1: Clinical features of affected subjects

	Subject ID			
Variable	**402**	**403**	**406**	**409**
Age (yrs)	25	22	29	30
Gender	M	F	F	F
Structural defects				
Developmental delay	+	+	+	+
Learning late	+	+	+	+
Walking late	+	+	+	+
Toileting late	+	+	+	+
Forehead protruding	−	+	−	+
Squint eyes	+	+	+	+
Digestive problems	+	+	+	+
Prominent jaw	+	+	+	+
Flat feet	+	+	+	+
Behavioral anomalies				
Self-care	Minor	Minor	Minor	No
Communication	Poor	Poor	−	Poor
Hyperactivity	+, periodic	+, periodic	+, periodic	+, periodic
Mood instability	+	+	+	+
Free time activity	Wandering in streets	Sitting silently	Playing with nephews	Sitting silently
Concept of money	+	−	−	−
Social responsibility	+	+	+	+
Sense of self respect	+	+	+	+
Sensitive to crowd	+	+	+	+

+, feature present; −, feature absent.

4.3.4: High throughput SNP genotyping

Peripheral blood samples from the affected and unaffected members of the relevant family have been taken and preserved in sterile K2EDTA tubes. DNA from peripheral blood sample was extracted through Phenol-chloroform extraction method. DNA sequencing were carried out with the help of local vender but all necessary steps were carried out in our lab. Whole genome SNP genotyping of two affected subjects (402, 408) was performed by using 730K map (730,000 SNP markers; Global Screening Array v2 offered by Illumina). A text file containing genotypes was imported to MS-Excel. Homozygous and heterozygous genotypes were highlighted with different color pattern. Then each chromosome was manually analyzed from upper telomere to lower telomere for the identification of >0.7 Mb homozygous intervals.

The SNP file were uploaded on HomozygosityMapper online tool for comparative analysis. Identified homozygous interval by both manual method and HomozygosityMapper were tabulated long with boundary SNPs and respective nucleotide positions. Genome Browser UCSC (https://genome.ucsc.edu) was employed to find the cytogenetic position of identified intervals. These homozygous intervals were furthers assessed through GeneDistiller (http://www.genedistiller.org) and regions harboring genes with expression in brain or involved in pathogenesis of ID were prioritized. The scheme used for the identification of homozygous intervals is represented through a flow chart (Fig. 4.2).

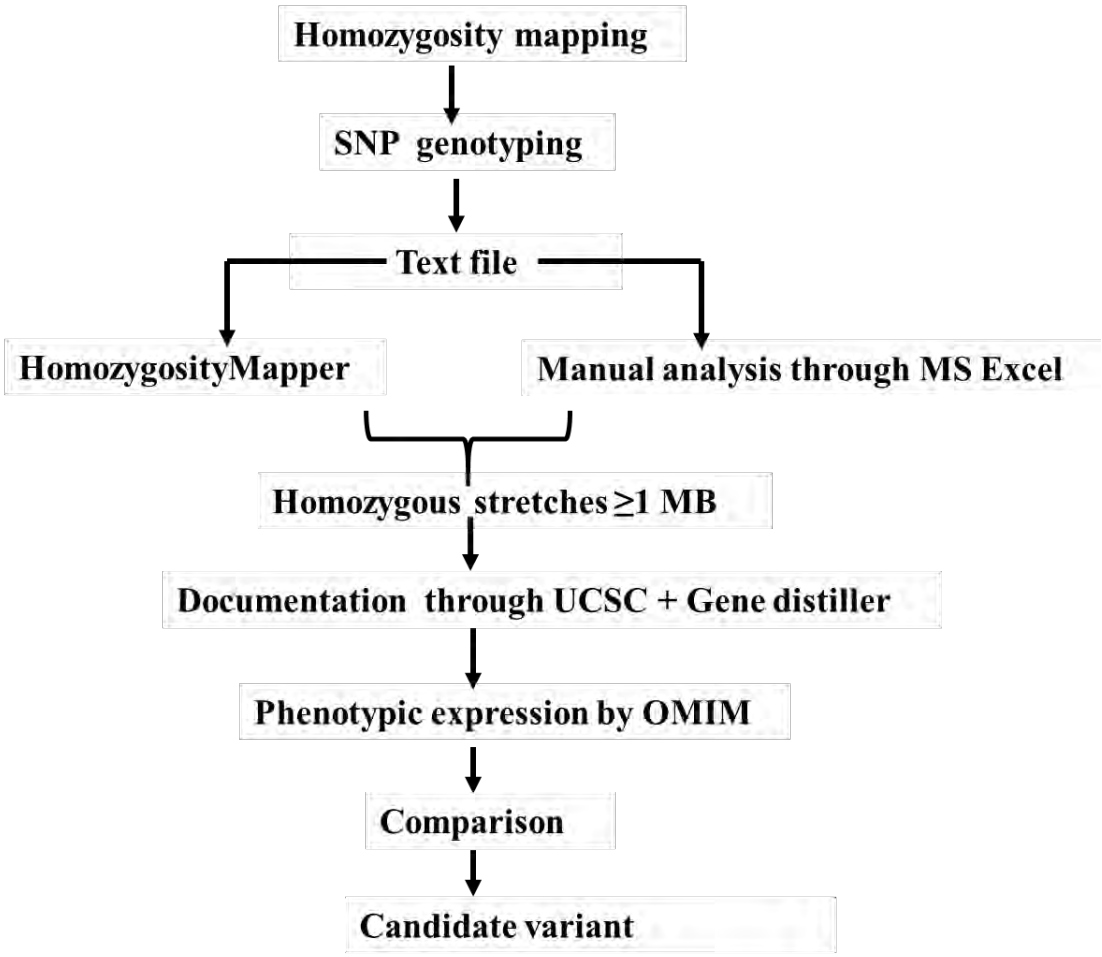


Fig 4.2: The strategy for disease gene hunt used in study.

4.4: Results

4.4.1: Identification of ≥0.7 Mb homozygous intervals in whole-genome SNP data

By using filters in MS-Excel, chromosome and size of homozygous regions were sorted in ascending order. These intervals were categorized based on size and recorded chromosome-wise in a table. Summary Tables represent full overview of the genome homozygosity (Tables 4.2; 4.3).

Table 4.2: Summary of identified homozygous intervals: (Individual 402)

	Number of homozygous intervals with size in (Mb)									
Chr. No	**0.7-0.79**	**0.8-0.89**	**0.9-0.99**	**1-1.99**	**2-2.99**	**3-3.99**	**4-4.99**	**5-5.99**	**>6**	**Total**
1	-	-	-	1	-	-	-	-	1	2
2	3	2	1	2	-	-	-	1	1	10
3	3	3	-	-	-	-	1	-	1	8
4	-	2	-	-	-	1	-	-	-	3
5	1	-	-	2	-	1	-	-	1	5
6	1	-	-	3	-	1	-	-	1	6
7	2	1	-	-	-	-	1	-	-	4
9	-	1	-	-	1	-	-	1	1	4
10	2	-	-	1	-	2	-	1	1	7
11	1	-	1	1	-	1	-	-	-	4
12	-	-	-	-	-	1	-	-	-	1
13	1	-	-	1	-	-	-	-	-	2
14	1	-	-	1	-	-	-	-	-	2
15	-	-	-	2	2	-	-	-	1	5
16	1	1	-	-	-	-	-	-	1	3
17	2	2	-	2	-	-	-	-	-	6
18	-	-	-	-	-	-	1	-	-	1
19	2	-	-	1	-	1	-	-	-	4
20	-	-	-	1	-	1	-	-	-	2
22	1		-	-	-	-	-	-	-	1
X	4	3	3	4	1	-	1	-	-	16
Y	-	-	-	2	-	1	-	-	-	3
Total	**25**	**15**	**5**	**24**	**4**	**10**	**4**	**3**	**9**	**99**
%age	**25.25**	**15.15**	**5.05**	**24.24**	**4.04**	**10.10**	**4.04**	**3.03**	**9.09**	**100.0**

Table 4.3: Summary of identified homozygous intervals: (Individual 408)

	No of homozygous intervals with size in (Mb)									
Chr. No	**0.7-0.79**	**0.8-0.89**	**0.9-0.99**	**1-1.99**	**2-2.99**	**3-3.99**	**4-4.99**	**5-5.99**	**>6**	**Total**
1	1	-	-	1	-	-	-	-	1	3
2	1	2	1	2	-	-	-	1	1	8
3	2	-	-	-	-	-	1	-	3	6
4	-	1	1	-	1	1	-	-	2	6
5	-	2	-	2	-	1	-	-	-	5
6	1	1	1	2	-	1	-	-	-	6
7	1	-	-	2	-	1	-	-	-	4
9	-	-	-	-	-	-	-	-	2	2
10	4	2	-	2	-	1	-	-	-	9
11	1	-	1	-	-	1	-	-	-	3
12	-	-	-	1	-	1	-	-	-	2
13	1	1	-	2	-	-	-	-	-	4
14	-	-	-	1	-	-	-	-	-	1
15	1	-	-	2	-	-	-	-	-	3
16	-	-	-	3	-	-	-	-	1	4
17	2	2	-	3	1	-	-	-	-	8
18	-	-	2	1	1	-	1	-	1	6
19	-	-	-	1	-	1	-	-	-	2
20	-	-	-	-	1	1	-	-	-	2
21	1	-	-	-	-	-	-	-	-	1
22	-	-	-	-	1	-	-	-	-	1
X	-	-	1	2	3	-	2	1	8	17
Y	2	2	-	4	-	-	2	-	-	10
Total	**18**	**13**	**7**	**31**	**8**	**9**	**6**	**2**	**19**	**113**
%age	**15.93**	**11.50**	**6.19**	**27.43**	**7.08**	**7.96**	**5.31**	**1.77**	**16.81**	**100.0**

4.4.2: Selection of homozygous intervals ≥1 Mb in SNP data:

For further analysis, regions having length of ≥1Mb were selected because they are larger in size and had higher likelihood of containing the causative gene. Further, the likelihood of two recombination events taking place in >1Mb region is low. The boundary SNPs and their corresponding nucleotide positions of these intervals were obtained and written separately. In this way, each chromosome was analyzed from upper telomere to lower telomere. Finally, the information or intervals was tabulated chromosome-wise (Tables 4.4).

Table 4.4: Identification of overlapping homozygous intervals of >1Mb in individuals 402 and 408

Chr. No	Start SNP	End SNP	Start bp	End bp	Difference (bp)	Difference in MB
4	rs7674086	rs10012987	49091782	52706270	3614488	3.61
5	rs28558979	rs6453529	68820348	70704077	1883729	1.88
6	rs806786	rs2235233	26259748	27387831	1128083	1.13
6	rs12661831	rs16894836	28337769	29361707	1023938	1.02
11	rs7479178	rs34944111	51566909	54794237	3227328	3.23
12	rs9705474	rs4002730	34826574	37876400	3049826	3.05
18	rs9957245	rs7239116	15102421	19486725	4384304	4.38
20	rs6107172	rs6141273	26278250	29904377	3626127	3.63

4.4.3: Identification of homozygous intervals through HomozygosityMapper

SNP genotype data were also examined for the identification of homozygous intervals on each chromosome by using HomozygosityMapper online tool. First, the data were analyzed by using the default parameters. Then selected parameters were used for the identification of homozygous intervals. There are Block Length limit with respect to number of SNP per block and block count length parameters. Initially default parameters were used then block limit changed from 500 to 8000. With increasing the block limit number of chromosomes with homozygous stretches decrease. Higher block limit count largest homozygous interval. Summary table represent full image of the genome homozygosity (Table 4.5).

Table 4.5: Distribution of homozygous stretches with respect to variable block limits (Individuals 402 and 408)

	Block length					
	3000	**4000**	**5000**	**6000**	**7000**	**8000**
Chromosome	**3**	**3**	**3**	**3**	**3**	**3**
	3	3				
	4	4	4			
	5	**5**	**5**	**5**	**5**	**5**
	6	6				
	6					
	9	9				
	10	10	10	10		
	15	15	15			
	18	18	18	18		

*Block length limit indicates number of SNPs

4.4.4: Genome-wide homozygosity score view through HomozygosityMapper (Individuals 402 and 408)

Figure 4.3 shows the homozygosity score plotted against the physical position in the genome. The program can display the whole genome, a single chromosome or a selected region on a single chromosome. Homozygosity Mapper presents the homozygosity over the entire genome as bar charts, together with the increase or loss of homozygosity in affected individuals compared to the control genotypes. Here, two cases were analyzed for shared homozygous intervals by using default parameters first then change the block length limit from 500 to 8000. As block length limit increases the chromosome wise distribution of homozygous intervals narrowed down. Interesting candidate regions are indicated by red bars in graphical view (Fig. 4.3-4.7).

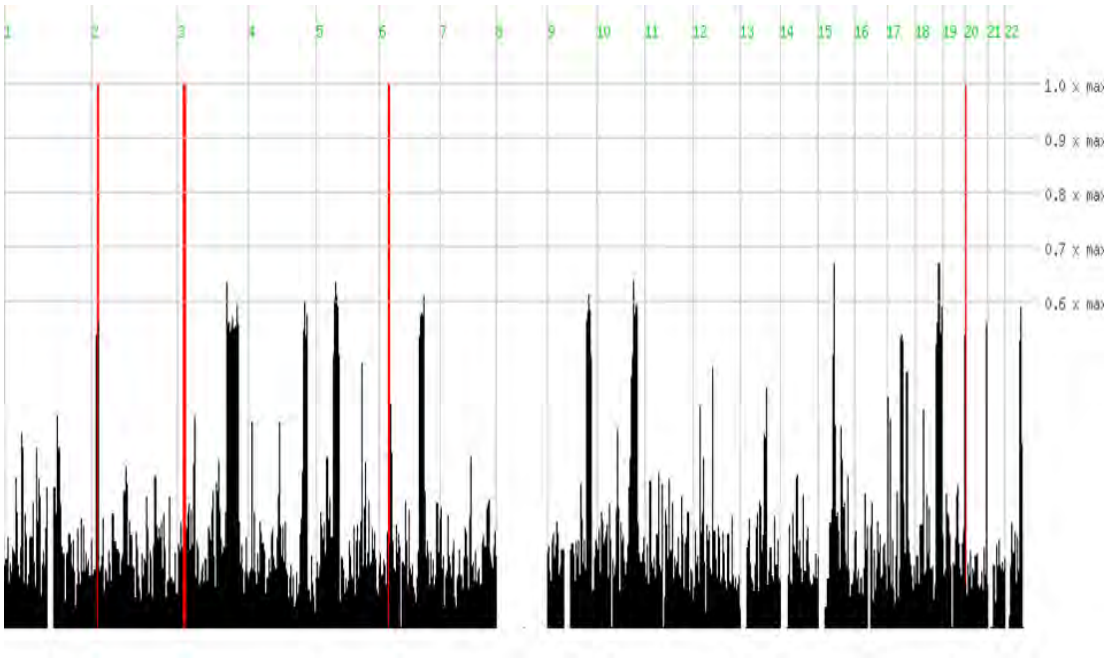


Fig 4.3: Graphical output of HomozygosityMapper showing genomewide homozygosity score. Red bars on chromosomes 2, 3, 6, and 20 show the most possible shared genomic regions.

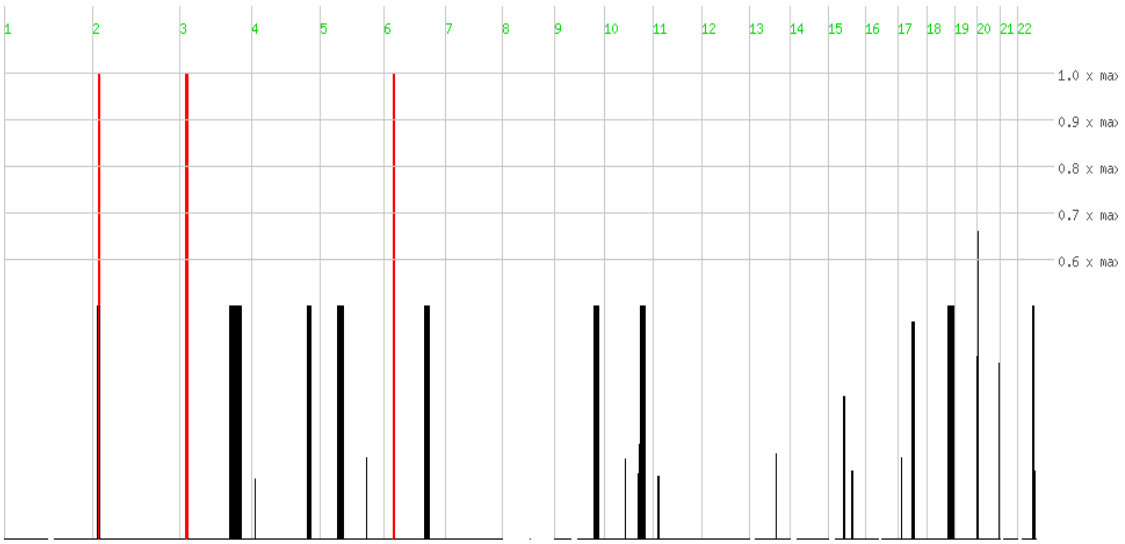


Fig 4.4: Output of HomozygosityMapper with block length limit of 1000.

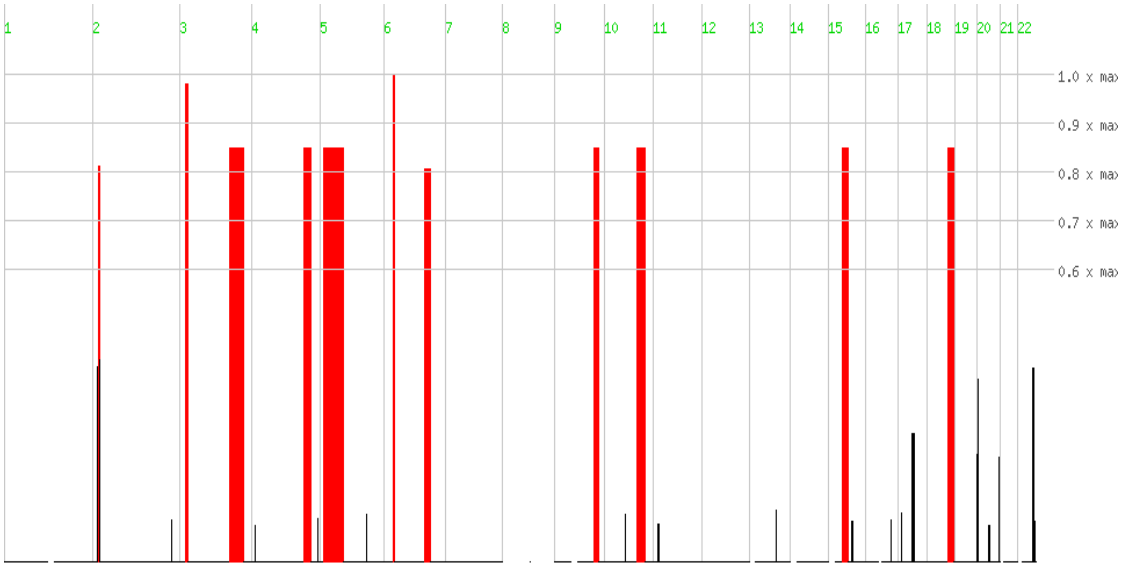


Fig 4.5: Homozygous intervals with block length limit of 3000.

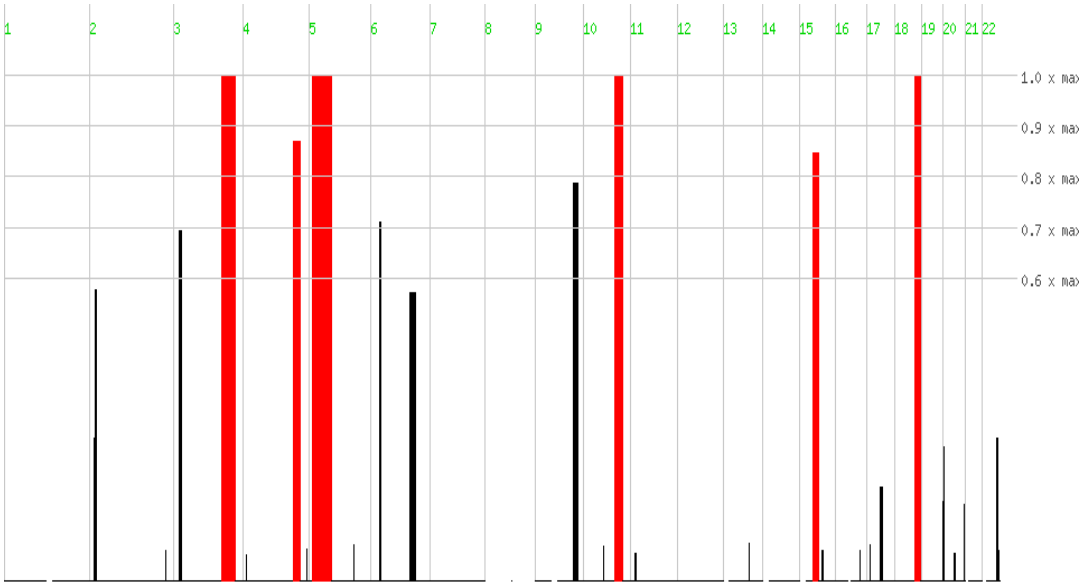


Fig 4.6: Homozygous intervals with block length limit of 5000.

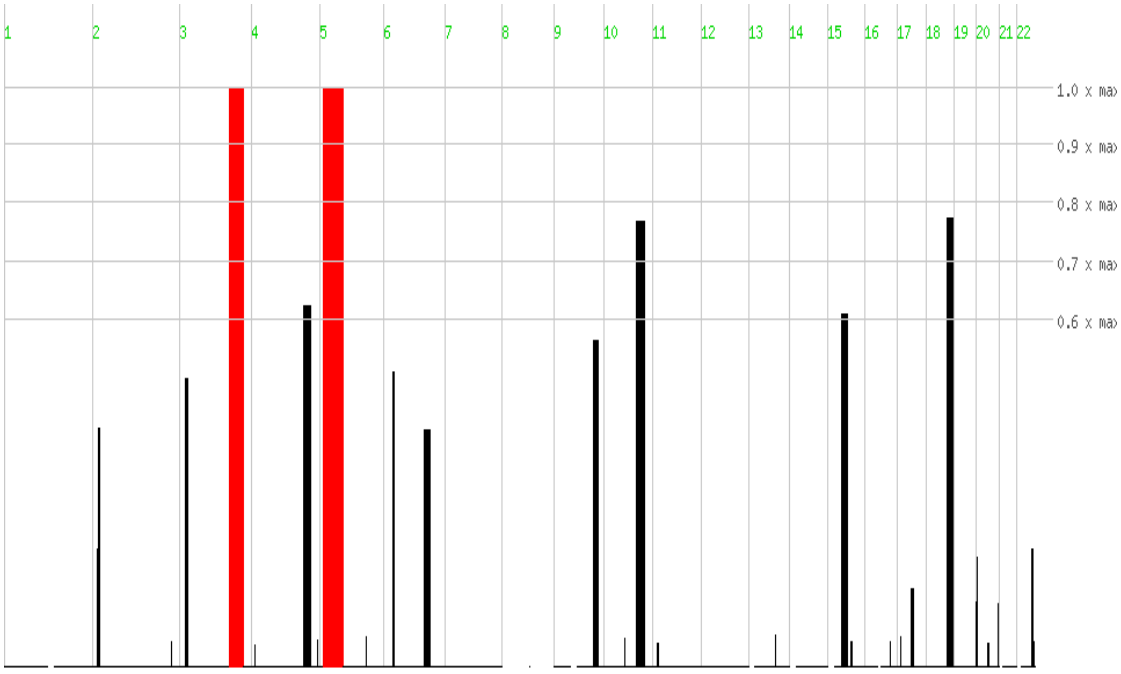


Fig 4.7: Homozygous intervals with block length limit of 7000.

4.4.5: Elucidation of candidate genes encompassed by 12 homozygous regions yielded by using GeneDistiller (http://www.genedistiller.org/) and UCSC browsers (https://genome.ucsc.edu/)

Selected homozygous intervals were further analyzed for the identification of candidate genes of phenotypically relevant syndromes. First, homozygous intervals were scrutinized by using UCSC genome browser (https://genome.ucsc.edu/; hg19) to identify their position on respective chromosome. In GeneDistiller, only chromosome number and starting and ending nucleotide positions was entered without changing the pre-defined parameters. Homozygous intervals were analyzed thoroughly by using search terms: 'intellectual disability, mental retardation, cognitive impairment, microcephaly, brain/neurological diseases.' Regions harboring disease-related genes (having an expression in the brain) were written in the separate table along with respective genes and their expression. So above mentioned genes causing the ID were selected in homozygous intervals (Table 4.6).

Table 4.6: Identification of candidate genes in 12 homozygous intervals through GeneDistiller and UCSC (Individuals 402 and 408)

Chr. No.	Start bp	End bp	Start SNP	End SNP	Size in Mb	Total genes	UCSC position	Candidate gene	Expression
2	14693546	17271863	rs6710377	rs10177601	2.6	8	p24.3-p24.2	***MYCN***	**Intellectual disability, Microcephaly**
									Nervous system development, cognitive impairment
3	16764234	23683479	rs6778843	rs12108186	6.9	32	p24.3	*RAB5A*	Expression in brain
								PLCL2	Brain development, cognitive impairment, microcephaly
3	139379385	178539486	rs11707887	rs6807583	39	302	q23-q26.32	*ZIC1*	Brain development,
								P2RY1	Nervous system development
								MBNL1	Nervous system development
								SHOX2	Nervous system development
								SERPINI1	Nervous system development
								NLGN1	**Intellectual disability**
								TBL1XR1	**Intellectual disability, cognitive impairment, microcephaly**
								ATR	Cognitive impairment
								TERC	Cognitive impairment
								SLC7A14	Cognitive impairment
								BCHE	

								CLRN1	Cognitive impairment
								SLC33A1	Global developmental delay
								GFM1	Microcephaly, motor delay
								AGTR1	Microcephaly
								IFT80	Macrocephaly
								SLC2A2	Global developmental delay
								DIPK2A	Mental retardation
4	146867376	168764988	rs6537382	rs2007133	22	144	q31.22-q32.3	*SFRP2*	Brain development
								PDGFC	Nervous system development
								GRIA2	Nervous system development
								GLRB	Nervous system development, cognitive impairment
								MAB21L2	**Intellectual disability, moderate, macrocephaly**
								DCHS2	Cognitive impairment
								LRAT	Cognitive impairment
								MSMO1	Microcephaly
								TDO2	Associated with autism
5	10924404	65086627	rs2399984	rs7711077	54	351	p15.2-q12.3	*NIPBL*	Brain development, intellectual disability
								OXCT1	Brain development
								SELENOP	Brain development

HMGCS1	Brain development
NDUFS4	Brain development, intellectual disability, Global developmental delay
KIF2A	Nervous system development, cognitive impairment, Alzheimer's disease, Microcephaly, Global developmental delay, Seizures
CTNND2	**Intellectual disability**
TRIO	Mild intellectual disability
CPLANE1	**Intellectual disability, global developmental delay**
PDE4D	**Intellectual disability,** global developmental delay, Autism
ERCC8	**Intellectual disability**, microcephaly, Seizures, severe postnatal growth retardation
NDUFAF2	**Intellectual disability,** global developmental delay, Seizures, progressive macrocephaly
ZSWIM6	**Intellectual disability**
HTR1A	Cognitive impairment, autism

								MOCS2	Microcephaly
								NADK2	Global developmental delay
								HCN1	**Intellectual disability, seizures**
								MCIDAS	Expression in brain
								GHR	Down syndrome
								SREK1IP1	Down regulated in brain of Alzheimer disease
6	26198046	29530850	rs4145878	rs362511	3.3	313	p22.2-p22.1	*ABT1*	Spinal cord motor neuron differentiation
6	114206311	129169909	rs503593	rs4487603	15	108	q21-q22.33	*NCOA7*	Expression in brain
								GJA1	**Intellectual disability,** cognitive impairment, global developmental delay, microcephaly, macrocephaly
								DSE	Cognitive impairment, delayed gross motor development
								FABP7	Down syndrome, nervous system developmental
								HEY2	Nervous system phenotype
9	109986449	122260749	rs13292544	rs10984568	12	109	q31.2-q33.1	*ASTN2*	Expression in brain
								TRIM32	Cognitive impairment
								PRPF4	Cognitive impairment
								PAPPA	Down syndrome

10	91275580	111303358	rs7087280	rs10884798	20	260	q23.31-q25.1	*CPEB3*	Expression in brain
								ARL3	Expression in brain
								NEURL1	Brain development
								KIF11	**Intellectual disability,** mild, Mental retardation, Microcephaly
								ALDH18A1	**Intellectual disability,** Microcephaly, associated with Down syndrome
								TCTN3	**Intellectual disability**, Microcephaly
								ENTPD1	**Intellectual disability**, Microcephaly
								COX15	**Intellectual disability,** Microcephaly
								PAX2	**Intellectual disability**
								TWNK	**Intellectual disability**
								SUFU	**Intellectual disability,** Macrocephaly, Cognitive impairment
								NT5C2	**Intellectual disability,** Motor delay
								CHUK	Microcephaly
								CNNM2	**Intellectual disability,** Microcephaly, Global developmental delay
								FGF8	Forebrain neuron

									development
								PITX3	Neuron development
15	38337341	55555602	rs8034546	rs4261468	17	250	q14-q21.3	*DMXL2*	Moderate expression in brain, postnatal growth retardation, Motor delay
								TMOD2	Nervous system development
								MYO5A	**Intellectual disability,** Global developmental delay, Cognitive impairment
								BUB1B	Severe. Global developmental delay, Microcephaly, Postnatal developmental delay, Cognitive impairment
								IVD	Global developmental delay, Cognitive impairment
								CHST14	Global developmental delay, Cognitive impairment, Motor delay, Intellectual disability
								NDUFAF1	Global developmental delay
								TUBGCP4	Global developmental delay
								GATM	Global developmental delay, Intellectual disability, Autism
								AP4E1	Global developmental delay, Microcephaly
								CEP152	Microcephaly, Cognitive

									impairment
								TUBGCP4	Microcephaly
								UBR1	Intellectual disability, Microcephaly
								SPG11	Mental deterioration, Intellectual disability
								CDAN1	Mild Postnatal growth retardation
								DLL4	Cognitive impairment
								ARPP19	Low levels in Down syndrome and Alzheimer's disease
18	58363142	75089915	rs7230590	rs4480888	17	95	q21.32-q23	*MBP*	Nervous system development
								RTTN	**Intellectual disability,** moderate, microcephaly
								ZNF407	**Intellectual disability**
								BCL2	Cognitive impairment
								PIGN	Macrocephaly, Global developmental delay
								TNFRSF11A	Global developmental delay
20	689537	1900961	rs6117232	rs6111988	1.2	24	p13	*SNPH*	Brain development

4.4.6: Candidate genes detected in both lines of analysis (manual and HomozygosityMapper)

Manually analyzed data as well as HomozygosityMapper results were compared for similar candidate genes involved in ID. UCSC genome browser was used to find the location of candidate gene and similar genes were sorted for specific intellectual disability through OMIM. (Tables 4.7; 4.8).

Table 4.7: Candidates genes for ID identified through HomozygosityMapper and manual analyses in individual 402

Chr. No	Cytogenetic position	Candidate genes
2	2p24.3-2p24	*MYCN*
5	5p15.2-5q12.3	*NIPBL*
		NDUFS4
		CTNND2
		CPLANE1
		PDE4D
		ERCC8
		NDUFAF2
		ZSWIM6
		HCN1
6	6q21-6q22.33	*GJA1*
10	10q23.31-10q25.1	*KIF11*
		ALDH18A1
		TCTN3
		ENTPD1
		COX15
		PAX2
		TWNK
		SUFU
		NT5C2
		CNNM2
15	15q14-15q21.3	*MYO5A*
		CHST14
		GATM
		UBR1
		SPG11

.

Table 4.8: Candidate genes for ID detected from HomozygosityMapper and manual analysis (Individual 408)

Chr.	UCSC Location	Candidate Genes	Expression/Phenotype
2	2p25.1-2p24.2	*MYCN*	Intellectual disability
3	3q23-3q26.32	*ATR*	Intellectual disability
		SLC9A9	Intellectual disability
4	4q31.2-4q32.3	*MAB21L2*	Intellectual disability
17	17q21.31-17q31.32	*KANSL1*	Intellectual disability
18	18q21.32-18q23	*RTTN*	Intellectual disability
		ZNF407	Intellectual disability

4.4.7: Shared homozygous stretches between individuals 402 and 408

HomozygosityMapper has set parameters and requirements and VCF files are uploaded against default parameters and analyzed. Both individual's (402 and 408) files were analyzed separately as well as in combined form. After analysis, a table appears that represents chromosome number with homozygous stretches, start and end SNP as well as start and end bp. After calculating the difference manually total length of homozygous stretch is obtained (Table 4.4). Those stretches with their start and end positions are indicated on ideogram schematic (Fig. 4.8).

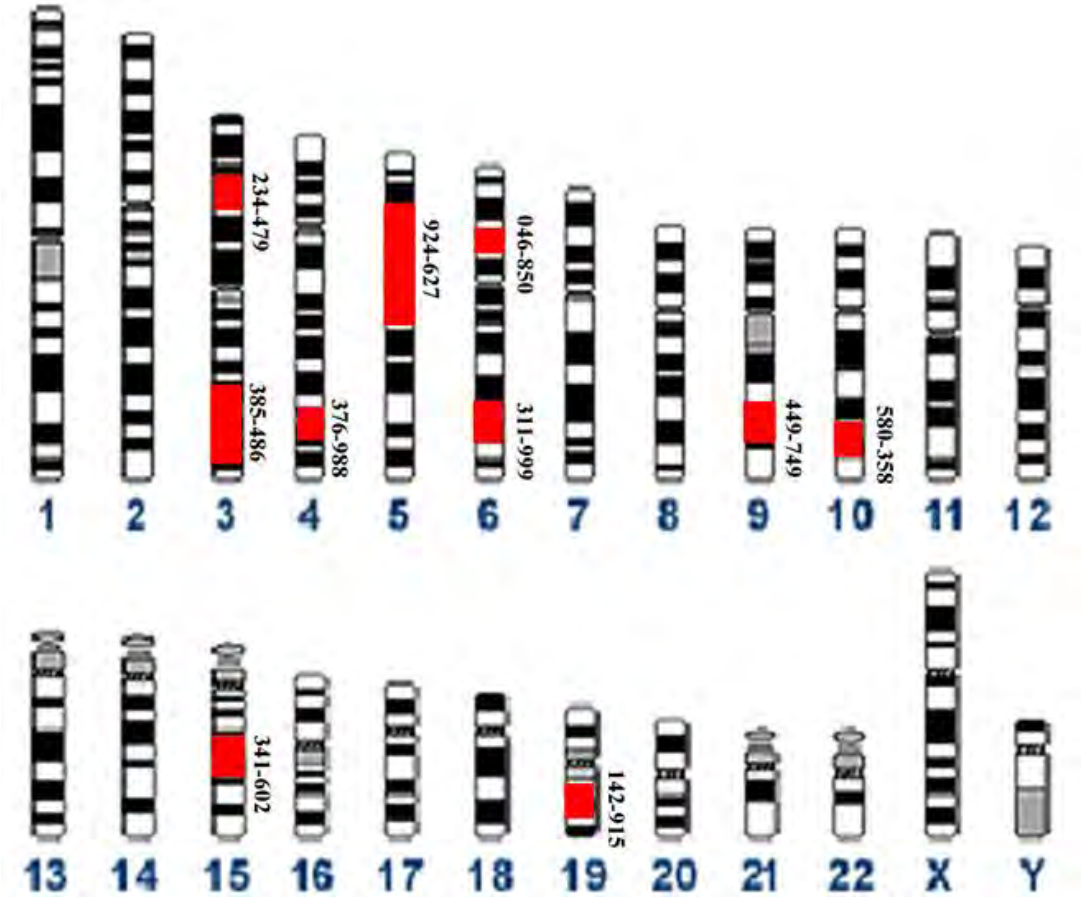


Fig 4.8: Ideogram showing the homozygous intervals (red boxes) shared between two affected individuals (402 and 408). Largest intervals were found on chromosomes 3 and 5. Numbers along the red boxes indicates the start and end positions.

4.5: Conclusion

In the present study, a family with moderate to severe intellectual disability was recruited. Homozygosity mapping was used to screen individuals before targeted mutation analysis. Both patients 402 and 408 displayed homozygosity in chromosomal regions that probably encompassed the mutant gene. Mutational analysis will be useful in genetic counselling and therapeutic interventions that are likely to become available in coming years.

5. Clinical and molecular investigations of a family with variable Fraser syndrome: microcephaly, cryptophthalmos, midface hypoplasia and tetramelic syndactyly

5.1: Abstract

The autosomal recessive disorder, Fraser syndrome 3 (FRASRS3) has bi-allelic variants in *GRIP1 and* is characterized by crptophathalmos, syndactyly and other annormalies of repiratory and urogenital tracts. Five unrelated affected (without Fraser syndrome) individuals, mostly babies or fetuses, have also been reported to date. Generally, the syndrome results from truncating variants. It is lethal in the first year of life; the exception is a 3-year-old Turkish girl, now almost 9 years old. In the current study, a 13-year old boy with a homozygous truncating variant c.1774C>T (p.Gln592Ter) in *GRIP1* gene was identified. The boy had a milder form of FRASRS3 with cryptophthalmia, midface hypoplasia, agenesis of the right kidney, cutaneous syndactyly in fingers and toes, but no symptoms in the lungs, anorectal system, genitalia, or umbilical system. To our knowledge, this case was the oldest reported individual with FRASRS3, demonstrating that FRASRS3 case may be milder than known and live into at least adolescence**.**

5.2: Introduction

Fraser syndrome (FS; OMIM 219000) is a rare autosomal recessive multi-system disorder with high clinical heterogeneity. The cardinal features of the syndrome include cryptophthalmos, laryngotracheal deformities, cutaneous syndactyly, renal agenesis, and ambiguous genitalia. Fraser syndrome may be accompanied by additional developmental anomalies such as skull ossification defects, nasal abnormalities, dysplastic ear, anorectal deformities, and umbilical abnormalities (Vogel *et al.,* 2012; van Haelst *et al.,* 2007).

Clinical examination is used to make the diagnosis, and it is based on major and minor criteria. Thomas *et al.,* (1986) described diagnostic criteria for FS for the first time. The major criteria comprise cryptophthalmos, respiratory tract anomalies, syndactyly, uncertain genitalia, and abnormalities of the urinary tract, while minor criteria encompass cranial ossification defects, congenital nose and ear deformities, anorectal malformations, and umbilical hernia. The diagnosis was based on one major with four minor criteria or two major with one minor criterion (van Haelst *et al.,* 2007; Thomas *et al.,* 1986). In another study, positive family history as well as cryptophthalmos, respiratory tract malformations, cutaneous syndactyly, ambiguous genitalia, and urinary tract abnormalities were found to be the most specific diagnostic criteria for Fraser syndrome (major criteria). Congenital anomalies like skull ossification defects, ear and nose anomalies, anorectal abnormalities, and an umbilical hernia were less specific (minor criteria; van Haelst *et al.,* 2007). There is, however, significant interfamilial phenotypic variability, ranging from minor clinical symptoms to severe abnormalities such as renal agenesis (Frasser, 1962, van Haelst *et al.,* 2007).

In contrast to significant clinical interest, epidemiological research on FS prevalence is limited. In the Spanish population, Martnez-Fras *et al.,* (1998) calculated the minimum frequency of FS to be 0.43 per 100,000 live births and 11.06 per 100,000 stillbirths (Comstock *et al.,* 2005). In the European population, the minimum estimate of FS was 0.2 cases per 100,000 births, with significant geographic disparities, such as the western half of Europe having many more cases than the rest of Europe (Barisic *et al.,* 2013). In addition, 25% of Fraser syndrome patients who survive their first year of life die, most commonly from the airway or urinary tract problems (Boyd *et al.,* 1988).

Fraser syndrome involves failure of programmed cell death and epidermal adhesion problems during embryonic development, resulting in massive lesions (Pavlakis *et al.,* 2011; Short *et al.,* 2007). Fraser syndromes, including FRASR1, FRASR2, and FRASR3 result from biallelic mutations in *FRAS1* (Fraser extracellular matrix complex subunit 1)*, FREM2* (FRAS1 Related Extracellular Matrix 2) and *GRIP1* (Gulatamate receptor interecting protein 1) genes, respectively. *GRIP1* (OMIM 604597) interacts directly with FRAS1/FREM protein complexes and is essential for their basal side cell localization (Jadeja *et al.,* 2005). Based on zebrafish studies, several new genes, having functions in basement membrane anchoring, have been proposed, including Hemicentin1 (HMCN1), Furin, and Fibrillin2, suggesting more genetic heterogeneity (Carney *et al.,* 2010).

The first *GRIP1*-related Fraser syndrome cases - two fetuses and a stillborn baby were reported in 2012 by three unrelated families with paternal consanguinity and no mutations in FRAS1 or FREM2 (Vogel *et al.,* 2012). Two of the fetuses were affected with the same homozygous splice site mutation (c.2113+1G>C), which result in the exclusion of exon 17 and thus assumed to shift the translational frame at codon

658 and result in a premature stop codon after the synthesis of 13 non-native residues. Consequently, the 471 terminal residue was deleted from the native protein of total 1128-amino acid. The remaining fetus was believed to be homozygous for c.1181_1184 del, p.(Lys394Thrfs*8) because parents were heterozygous for the same mutation. Schanze *et al.,* (2014) reported a stillborn baby with homozygous nonsense c.1860C>A, (p.(Tyr620*) mutation in exon 16 and a 3-year-old Turkish girl with compound heterozygous for nonsense c.2120C>A, (p.(Ser707*) mutation in exon 18. Due to the intragenic deletion with breakpoints possibly located inside intron "16 and 18," it was assumed that exons 17 and 18 would be deleted and frameshift would occur (Schanze *et al.,* 2014). All five reported variants were truncating. Furthermore, two single individuals with compound heterozygous missense mutations but no cryptophthalmos were detected. One of them (age unspecified) had c.1846G>A (p.Gly616Arg) and c.2750G>T (p.Arg917Leu) mutations together with congenital anomalies of the kidney and urinary tract (CAKUT) but other manifestation characteristics of FRASR3-like cryptophthalmos, syndactyly, or genital abnormalities were not present, therefore considered as isolated congenital anomalies of the kidney and urinary tract (CAKUT) (Kohl *et al.,* 2014). The other individual had c.1142G>T, (p.Ser381Ile) and c.160G>A, (p.Val54Ile) mutations without a FRASRS3 profile but with neurological symptoms along with agenesis of the corpus callosum and subependymal heterotopia (Karaca *et al.,* 2015). In summary, a total of 7 unrelated cases with 8 different *GRIP1* variants and an intragenic deletion have been reported (Table 5.1). All of the mutations are the biallelic variants that truncate *GRIP1* and cause FRASRS3 (Shanze *et al.,* 2014; Vogel *et al.,* 2012).

Table 5.1: Mutational status of *GRIP1* for patients with FRASR3 and related phenotpes

Phenotype	Genomic location	cDNA	Protein	Mutation type	Zygosity	Consequences	Ref
FRASRS3	12:66786270	c.2113+1G>C	p.(Arg658fs*13)	Frame shift	Homozygous	Donor/splice site	Vogel *et al.,* 2012
FRASRS3	12:66849203-66849206	c.1181_1184del	p.(Lys394Thrfs*8)	Frame shift	Homozygous	Exonic	Vogel *et al.,* 2012
FRASRS3	12:66,788,101	c.1860C>A	p.(Tyr620*)	Nonsense	Homozygous	Exonic	Schanze *et al.,* 2014
FRASRS3	12:66786276	c.2120C>A	p. (Ser707*)	Nonsense	Homozygous	Exonic	Schanze *et al.,* 2014
FRASRS3	**12:66,800,117**	**c.1774C>T**	**p. (Gln592Thr)**	**Missense**	**Homozygous**	**Exonic**	**Present study**
Congenital anomalies of the kidney and urinary tract	12:66788115	c.1846G>A	p.(Gly616Arg)	Missense	Compound heterozygous	Splice site	Kohl *et al.,* 2014
Congenital anomalies of the kidney and urinary tract	12:66,765,580	c.2750G>T	p.(Arg917Leu)	Missense	Compound heterozygous	Exonic	Kohl *et al.,* 2014
Agenesis of corpus callosum, subependymal heterotopia and lacking cryptophthalmos	12:66,849,245	c.1142G>T	p.(Ser381Ile)	Missense	Compound heterozygous	Exonic	Karaca *et al.,* 2015
Agenesis of corpus callosum, subependymal heterotopia and lacking cryptophthalmos	12:66,925,707	c.160G>A	p. (Val54Ile)	Missense	Compound heterozygous	Exonic	Karaca *et al.,* 2015

Fraser syndrome is a generally lethal disorder where life expectancy is less than a year. There are only 3 cases that lived over 20 years; the most notable is the 96-year-old female (Impallomeni *et al.,* 2006). The underlying genetic defects for all those cases are unknown. Despite the severe mutation, the 15.5-year-old affected boy investigated through current research is the oldest of the known *GRIP1*–related FRASRS patients and indicates a milder end of FRASRS3.

5.3: Subjects and Methods

5.3.1: Family

Family originates from upper Punjab, Pakistan, and has one affected sib and four unaffected sibs (Fig. 5.1). Parents are first-cousins and have no other affected relatives. There is no family record of congenital malformations or any miscarriages.

Clinical examination of the affected subject included X-rays of the chest, skull, and upper and lower limbs, and ultrasonography of vital organs was performed. Peripheral blood samples from the affected and unaffected members of the relevant family have been taken and preserved in sterile K2EDTA tubes in order to extract genomic DNA. All subjects gave their informed consent, and the pedigree was drawn after consulting the elder family members. The study was approved by the Ethical Review Committee of Quaid-I-Azam University (DAS-1070; Dated July 8, 2015).

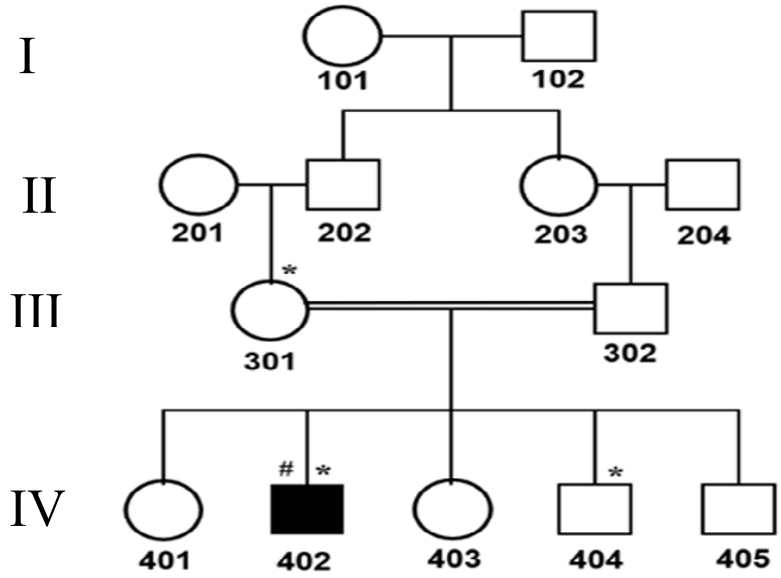


Fig 5.1: Pedigree of family, segregating autosomal recessive FRASR3. Squares are used to indicate males, while females are specified by circles. Clear and filled shapes represent normal and affected family members, respectively. Double line between male and female indicates cousin marriage. *, DNA available for genetic study; #, exome data generated.

5.3.2: Clinical findings

At the birth of the affected son, both parents were 22 years old. His prenatal history and pubertal onset were both recorded as normal. Clinical features of the affected boy are presented in Fig. 5.2. His face was elongated and triangular, his hairline was quite low, and his physique was slim and lean. With bilateral cryptophthalmos, there was dimpling of the skin covering the eyes. While eyelids, eyebrows and eyelashes were absent. The affected boy had midface hypoplasia, hypertrophied frontal and maxillary sinuses, deformed pinnae, grade1 microtia with overfolded helix, small nasal bone, bifid nose tip, and alae nasi. Hair development on the alae nasi and temples extending to the supraorbital line was noted during the onset of puberty. Upper incisors protrusion was noticeable, while the hypodontia involving lateral upper and lower incisors and upper left premolar, as well as the peg-shaped

lower central incisors were also prominent(Fig. 5.2 A-B). The lateral X-ray images revealed a prominent maxilla with overbite and malocclusion, overjet upper incisors, and retrognathia with a pointed chin. The pectoral girdle and clavicles were hypoplastic, and the humeral shafts were thin on a roentgenogram (Fig. 5.2 C-D).

An autopod radiograph showed cutaneous fusion of fingers 2-4 on the right hand and 1-5 on the left, along with bilaterally fused nails of the synostozed fingers 3-4. The toes were abnormal or absent, and the feet were flat. In the right foot, a complete fusion of toes 1-5 was observed; toes 2-4 were buried under the skin and came up with terminal phalangeal hypoplasia, leaving only toes 1 and 5 visible. Complete cutaneous fusion of toes 2-3 and toes 4-5 was noticed in left foot (Fig. 5.2 E-H).

Abdominal ultrasonography indicated hypoplasia of the right kidney and a slightly constricted gall bladder. Roentgenographic examinations of the skull indicated hypertrophied frontal and maxillary sinuses, as well as hypoplastic orbits, associated with facial hyper-divergence. Ultrasound examination revealed normal lobation of the lungs and no sign of pulmonary hyperplasia. Other symptoms like, developmental delay, cleft lip/palate or anomalies of the genital system were not observed. Patient's IQ was normal, along with a prolific voice and he was attending a local religious school.

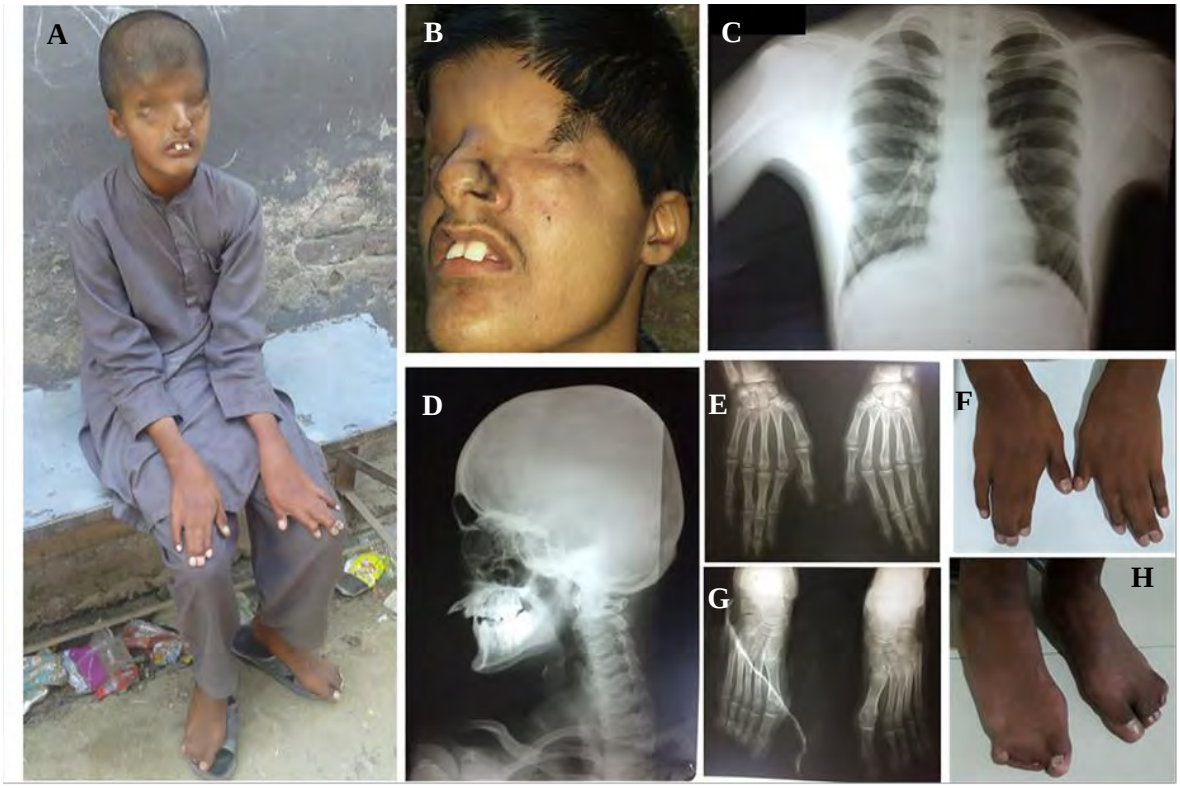


Fig 5.2: Clinical and radiographic features of affected boy. He had microcephaly, elongated and triangular face. Very low hair line, dimpling of the skin covering the eyes present with bilateral cryptophthalmos (A). Affected boy had midface hypoplasia, hypertrophied frontal and maxillary sinuses, deformed pinnae, grade1 microtia with overfolded helix, underdeveloped crus, small nasal bone, bifid nose tip, and alae nasi (B). Lungs present with normal lobation, and no pulmonary hypoplasia is present (C). Upper incisors that protrude, mandibular retrognathia with a pointed chin, malocclusion, and upper incisors that overjet (D). Bilateral symphalangism in distal phalanges of 2nd to 4rth fingers; cutaneous fusion of the 2nd to 4rth and 1st to 5th fingers in the right and left hands, respectively (E-F). Terminal phalangeal hypoplasia of the 2nd to 4rth toes; In the right foot, total fusion of the 1st to 5th toes, and complete fusion of the 2nd to 3rd and 4rth to 5th toes in the left foot (G-H).

5.3.3: Genetic findings

DNA was extracted from the samples of the mother (302), affected son (402) and unaffected son (404). DNA was extracted mannualy through phenol chloroform DNA extraction method. The affected son's exome was sequenced on the Illumina HiSeq2500 platform with the help of our collaborators at Department of Molecular Biology and Genetics, MOBGAM, Istanbul Technical University, Istanbul, Turkey. IDTxGen Exome Research Panel was used to create an exome library, extra target of 2,500 sites that were not covered completely or poorly covered by the kit were also included. WA-0.7.12-r1030 was used to align raw reads to GRCh37 (hg19) and variant calling was done with SAMtools-0.1.14. Functional annotation was done by using ANNOVAR. Non-synonymous, truncating or splicing variants with minor allele frequency >0.01 (for all populations reported in 1000 Genomes and gnomAD databases), probably homozygous (alt/tot ratio >0.6), and capable of changing the amino acid sequence were filtered from the exome file.

After applying the filtering approach to the variants in the affected son's exome file, there were 17 candidate variants that could potentially modify the protein sequence, with *GRIP1* c.1774C>T (p.Gln592Ter; NM 021150.4) being the most likely candidate (Table 5.2). While among the remaining variants those that were present either in our 5l in-house exome files or were reported in homozygous states in a variation database, in a gene where homozygous loss of function variants are listed in such a database were excluded. Pathogenicity of the candidate variants was checked by using *In-silico* tools like CADD and MutationTaster2 (Rentzsc *et al.,* 2019; Schwarz *et al.,* 2014). Segregation of identified variant with the disease was confirmed via Sanger sequencing; affected sib was homozygous while mother and unaffected son were heterozygous (Fig. 5.3). On the other hand, no rare variant was

detected in *FRAS1* or *FREM2* genes. Fig. 5.4 represents a protein schematic with the position of the previously recognized variant in the *GRIP1* gene and the location of the recently discovered variants (c.1774C>T; p.Gln592Ter) in bold. The mutation found in the current study is situated in the Pdz 5 domain of the GRIP1 protein, which contains seven Pdz domains (Pdz1-Pdz7).

Table 5.2: Exome sequence analysis revealed following list of possibly homozygous (alt/tot depth > 0.6) rare (frequency < 0.01) or novel variants

Chr	Start	End	Ref	Alt	Consequence	Gene	HGSV	Frequency (gnomAD)	Depth (alt/tot)
chr1	36859619	36859619	C	T	nonsny	*LSM10*	c.112G>A (p.Val38Met)	3.89E-05	131/131
chr6	131912594	1.32E+08	T	G	nonsny	*MED23*	c.3545A>C (p.Glu1182Ala)	novel	38/63
chr6	136603827	1.37E+08	-	G	splicing	*BCLAF1*	c.-114-2_-114-1insC (p.Xxx_Yyydelfsxx)	novel	7/7
chr6	166779466	1.67E+08	G	A	nonsny	*MPC1*	c.172C>T (p.His58Tyr)	1.68E-04	97/97
chr9	214969	214969	G	A	nonsny	*C9orf66*	c.428C>T (p.Pro143Leu)	1.37E-05	66/102
chr9	98211514	98211514	G	A	nonsny	*PTCH1*	c.3641C>T (p.Thr1214Met)	3.59E-04	138/138
chr10	72493750	72493750	C	T	nonsny	*ADAMTS14*	c.1327C>T (p.Arg443Cys)	1.08E-04	107/107
chr12	66800117	66800117	G	A	stopgain	*GRIP1*	c.1774C>T (p.Gln592Ter)	novel	65/65
chr15	52442107	52442107	C	T	nonsny	*GNB5*	c.391G>A (p.Val131Met)	1.99E-05	29/47
chr17	641142	641142	C	T	nonsny	*FAM57A*	c.263C>T (p.Ser88Leu)	2.54E-04	93/94
chr17	1628935	1628935	C	T	nonsny	*WDR81*	c.682C>T (p.Pro228Ser)	2.36E-04	153/153
chr17	3402246	3402246	C	T	nonsny	*ASPA*	c.806C>T (p.Thr269Met)	1.08E-04	91/91
chr17	6329946	6329946	C	T	nonsny	*AIPL1*	c.773G>A (p.Arg258Gln)	9.17E-05	152/152

chr17	8093133	8093133	C	G	nonsny	*BORCS6*	c.326G>C (p.Arg109Pro)	9.12E-05	123/123
chr19	48809552	48809552	C	T	nonsny	*CCDC114*	c.515G>A (p.Ser172Asn)	1.07E-04	79/79
chr22	26422611	26422611	C	A	nonsny	*MYO18B*	c.6671C>A (p.Ala2224Asp)	1.64E-04	87/142
chrX	57936451	57936452	CA	-	frameshift deletion	*ZXDA*	c.403_404delTG (p.Cys135LeufsTer42)	novel	6/6

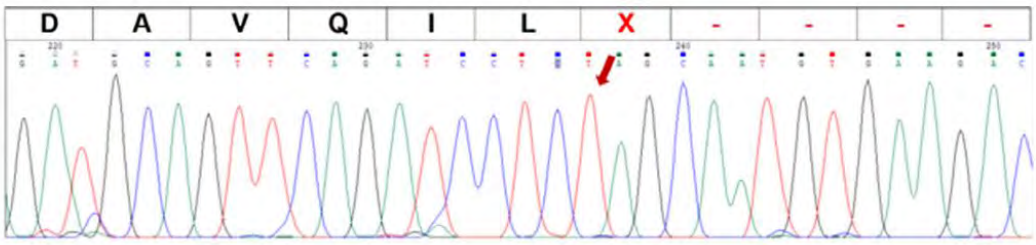
Nonsny; nonsynonymous. Red box represents the most promising variant identified in the current study

(A)

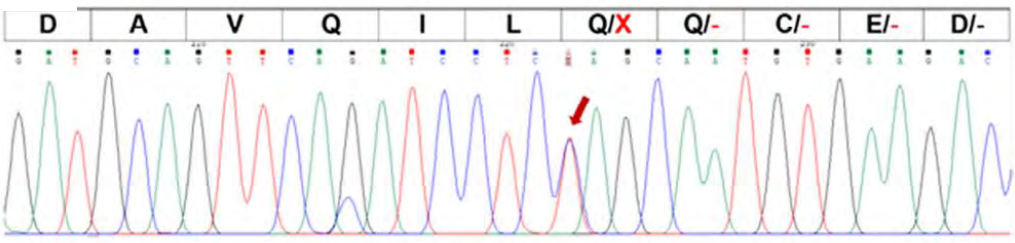
Oligo	Length	Temp. ºC	GC%	Sequence
Right primer	20	59	50	TGACAGATGGGCAGTTCGAA
Left primer	20	59	5	GAACTGGGACCCTGGAACTT

Product size: 199

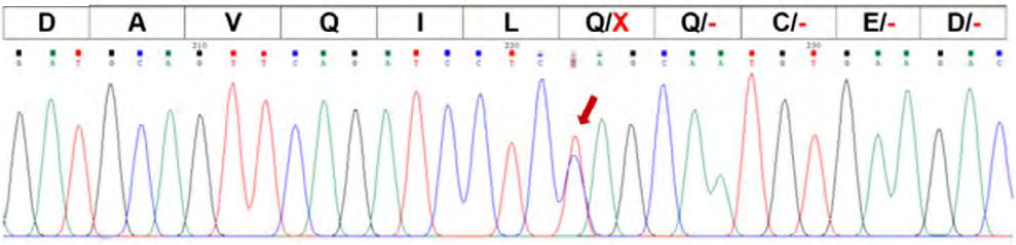
(B)



(C) **402 (Affected; +/+)**



(D) **301 (Unaffected; +/-)**



404 (Unaffected; +/-)

Fig 5.3: Sequence chromatogram of *GRIP1* c.1774C>T). The primer sequence for sanger (A). homozygous stop gain variant (GRIP1 c.1774C>T (p.Gln592Ter) in an affected boy (402) is indicated by red arrow (B). While the unaffected family members 301 and 404 are heterozygous for the wild type allele (C, D). +, indicates mutant allele, while wild type allele is represented by −.

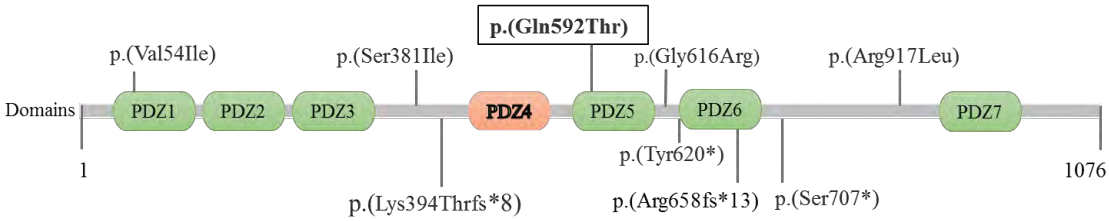


Fig 5.4: GRIP1 Protein domain structure along with reported mutations causing Fraser syndrome 3 (FRASRS3): Upper half of the figure represents missense mutations p.(Val54Ile), p.(Ser381Ile), p.(Gly616Arg) and p.(Arg917Leu); while lower half indicating frameshift and nonsense mutations p.(Lys394Thrfs*8), p.(Arg658fs*13) and p.(Tyr620*), p.(Ser707*) respectively. Mutation p.(Gln592Thr) in box marked with bold letters represents our 13-year-old index patient.

5.4: Discussion

The present study involved the clinical and molecular evaluation of a consanguineous family with a single affected boy with autosomal recessive FRASR3. He represents the first reported adolescence case of Fraser Syndrome 3, with a novel nonsense variant in *GRIP1.* A homozygous *GRIP1* variant c.1774C>T (p.Gln592Ter) was discovered by exome sequencing (Schanze *et al.,* 2014; Vogel *et al.,* 2012). The variant is novel (not found in genomic databases such as 1000 Genomes, gnomAD, or others including the ClinVar) and affects all five isoforms of the protein. Codon 592 of the 1128-residue GRIP1 protein is changed to a stop codon, resulting in premature termination and deletion of 537 native amino acids. The variant was predicted to be deleterious through *in silico* tools CADD (score 39) and MutationTaster2. Up till now, five unrelated FRASRS3 cases are documented. Four of these (three in total) were fetuses with homozygous truncation mutations. The fifth and the last case was compound heterozygous for a truncation variant and a deletion probably encompassing exons 17 and 18 and resulting shift in the translational reading frame, and truncated protein (Schanze *et al.,* 2014; Vogel *et al.,* 2012). Furthermore, two independently detected variants lacking cryptophthalmos were described, but no link to FRASR3 was discovered. Congenital kidney and urinary tract abnormalities were discovered in one of the patients (CAKUT). The other had a neurological condition characterized by agenesis of the corpus callosum and subependymal heterotopia, and one of the variants (c.160G>A) was discovered to be rather common in the population (0.03 in the Turkish population), throwing doubt on the causality (Karaca *et al.,* 2015; Kohl *et al.,* 2014). In 15 Turkish exome files in the host lab, we found one heterozygote mutation for that variant. As a result, biallelic missense *GRIP1* variants can induce different disorders, whereas all FRASRS3-causing mutations truncate the

protein, resulting in severe effects. The boy we are presenting here is the oldest of the FRASRS3 case that have been recorded so far. He exhibits the characteristic FRASRS3 presentation, fitting three of the six major and two of the five minor diagnostic criteria. Particular diagnostic features like cryptophthalmia, midface retrusion, right kidney agenesis, cutaneous webbing in hands and feet (van Haelst *et al.,* 2007). Typical malformations of the respiratory system such as abnormal larynx, tracheal atresia, pulmonary hyperplasia, abnormal lung lobation, hyperechogenic lungs as well as the specific features of genital system including the abnormally positioned anus, hypoplastic scrotum, and hypoplastic bladder and genitalia were not present. Usually FRASRS3 patients could not survive to the childhood, only one FRASRS3 case, a three-year-old girl, was reported by Schanze *et al.,* The patient had hydronephrosis, renal agenesis, finger webbing, but no cryptophthalmos. She was diagnosed with agenesis of some teeth and short roots in others at the age of nine years (Kunz *et al.,* 2020).

These findings indicate that FRASRS3 bi-allelic truncating variants can manifest with milder clinical features, including unremarkable affects on the respiratory and anorectal systems and genitalia, free of umbilical abnormalities, and living to adolescence at least. We present the oldest reported patient with FRASRS3 to our knowledge and hope that this report would aid the clinical diagnosis of new FRASRS3 case with variable features.

6. Clinical and molecular study of a family with skeletal dysplasia, microcephaly and intellectual disability

6.1: Abstract

Microcephaly and intellectual disability are hallmarks of the severe autosomal recessive spondyloepimetaphyseal dysplasia known as Dyggve-Melchior-Clausen syndrome (DMC; MIM #223800). DMC is carried on by loss-of-function mutations in the DYM gene, which encodes the Golgi protein DYMECLIN, which is known to be involved in intercellular trafficking. Patients with DMC frequently display a variety of skeletal-related symptoms, such as facial dysmorphism, short trunk dwarfism with numerous radiographic features, and proximal limb reductions. On the other hand, DMC typically reports microcephaly with mild to severe intellectual disability and frequently reported poor or absent verbal abilities. As skeletal deformities and microcephaly both manifest during development and are not identified at birth, this suggests that postnatal mechanisms were involved in the physiological mechanism causing DMC. A three generations pedigree with autosomal recessive inheritance pattern of disease was recruited from district Harpur. Physical and clinical assessment done by taking photographs and X-rays. Blood samples taken from six subjects (3 affected and 12 unaffected) for DNA extraction to perform Exome sequencing. A novel homozygous missense mutation NM_017653.3:c.1072C>T(p.Q358*) in *DYM* gene was identified via Exome sequencing. This mutation predicted to form a truncated DYMECLIN protein which may profoundly impair its normal function. Although the *DYM* is not novel gene associated with DMC but the currently identified variant is novel. In silico analysis predicted the mutation as protein damaging. The identified mutation caused an autosomal ressive phenotype previously described in OMIM database, in extended Pakistani kindred. All patients in the family

had common feature of DMC including rhizomelia, short trunk dwarfism, scoliosis, microcephaly, intellectual disability, and other clinically diverse symptoms. The current molecular analysis described a putative DMC phenotype associated with a *DYM* gene and also confirmed the clinical condition that seggregate in the family. The reported findings provide for future functional studies on the DYMECLIN protein.

6.2: Introduction

Skeletal dysplasias (osteochondrodysplasias; OCD) are generalized skeletal defects that frequently result in disproportionately small stature. Although there are exceptions, most affected individuals have disproportionately short stature with skeletal dysplasia, while others with proportionately short stature have hormonal, nutritive, or other hereditary or teratogenic causes. For instance, certain types of osteogenesis imperfecta (OI) and hypophosphatasia have body proportions that are rather normal (Krakow and Rimon, 2010). These disorders have a diverse variety of symptoms; some people have severe and neonatal mortality, whereas others are of average height with early arthropathy. These disorders have a range of neurologic, psychologic, visual, hearing, pulmonary, cardiac, and renal difficulties (Krakow and Rimoin, 2010).

Despite their rarity, genetic skeletal dysplasia conditions (GSDs) account for 5% of all congenital malformations, with an incidence of 1 in every 5000 births. They may be associated to high rates of morbidity and mortality (Sabir and Coli, 2019; Marzin and Cormier-Daire, 2020). There are 461 disorders listed in the current "Nosology and categorization of genetic skeletal disorders" (2019), with 425 (92%) having a known molecular etiology (Sabir and Irving, 2021).

Osteochodrodysplasia (OCD) are a heterogeneous group of skeletal dysplasia characterized by abnormalities in the development, patterning, texture, and maintenance of bone and cartilage (Warman *et al.,* 2011; Liu and McEntee, 2017). OCD occur due to mutations and their phenotypes continuously evolved over the years, whereas dysostosis, which is a deformity of a single or group of bones in combination caused by abnormal blastogenesis in the uterus, phenotypically remained

static throughout life (Panda *et al.,* 2014). These differences are blurring as their basic defects have been elucidated. Based on clinical, radiographic, biochemical, and molecular criteria, there appear to be about 450 different forms of OCD (Warman *et al.,* 2010). Each type of skeletal dysplasia has a different prevalence, with achondroplasia being the most common, having a frequency of 1 in 2500. Only 40% of CA with OCD show symptoms at birth, while the remaining are diagnosed later in life (Barbosa-Buck *et al.,* 2012). Among the former, the currently estimated overall prevalence is to be 2.4-7.6 per 1000 births (Barbosa-Buck *et al.,* 2012; Liu and McEntee, 2019).

Since 1970, numerous attempts have been made to classify these malformations. The most effective way for discriminating between the various skeletal ailments has been the detection of radiographic anomalies. The different components of the long bone, such as the epiphysis, metaphysis, and diaphysis, are used to classify them radiographically. The epiphyseal, metaphyseal and diaphyseal disorders are further categorized depending on whether the spine is involved (spondyloepiphyseal, spondylometaiphyseal dysplasias [SMDs], or spondyloepimetaphyseal dysplasias [SEMD]). Skeletal dysplasias can be classified into a number of disorders based on clinical and radiographic assessment (Krakow and Rimoin, 2010).

Dyggve-Melchior-Clausen Syndrome (DMC syndrome, MIM 223800) that belongs to the spondyloepimetaphyseal dysplasia (SEMD), a subgroup of osteochondrodysplasias is a rare autosomal recessive disorder (Abdullah *et al.,* 2020; Khalifa *et al.,* 2011). In 1962, DMC syndrome was first time described as a new form of dwarfism linked with intellectual disability (ID). Since then only 100 cases have been reported which account mean prevalence of 0.1 per million (Nectoux *et al.,*

2013). DMC syndrome is caused by mutations in the Dymeclin (*DYM*) gene (mapped on 18q21.1 chromosomal region; Cohn *et al.,* 2003).

DMC syndrome is characterized by progressive spondyloepimetaphysial dysplasia (SEMD), as well as microcephaly, facial dysmorphism, and different levels of ID. Short stature, short limbs and trunk, barrel-shaped chest, varus and valgus deformities of the knee, reduced joint mobility, brachydactyly, and micropenis are characteristic features. Radiological features include platyspondyly with double-humped vertebrae (scoliosis, kyphoscoliosis), metaphysis, epiphysis, hypoplastic odontoid process, anterior beaking of vertebral bodies, abnormal ossification of long bones and lacy iliac crests (Aglan *et al.,* 2009; Gupta *et al.,* 2010; Abdullah *et al.,* 2021).

Here, detailed clinical and genetic diagnosis of an extended consanguineous Pakistani family presented with dwarfism, microcephaly and ID, was performed. Clinical and radiological examination of all three affected family members reveals substantial microcephaly, ID, and other skeletal characteristics that are similar to the DMC syndrome previously documented. Exome sequencing and sanger sequencing revealed a novel homozygous nonsense mutation in exon 10 of the *DYM* (c.1072C>T, p.Gln358*).

6.3: Subjects and methods

6.3.1: Family recruitment

The family was ascertained from a remote area of Northern Pakistan. Three generation pedigree was drawn by interviewing the head of the family. Information was collected after taking the informed and written consent. The provided information about family and disease phenotype was crosschecked by interviewing other relatives. The pedigree represents an autosomal recessive mode of inheritance (Fig. 6.1). Upon assessment 3 affected individuals (1 male, 2 female) were observed in the family. All available affected members (301, 306 and 307) underwent physical examination with the help of local physicians. Blood samples of three affected and three unaffected subjects were available for molecular analyses. One affected male (301) was subjected to a radiological evaluation. Two individuals 301 and 307 were subjected to exome analyses.

In order to extract genomic DNA, peripheral blood samples from the available affected and healthy individuals of the relevant family have been stored in sterile K2EDTA tubes. All information and blood samples were obtained according to the Helsinki-II declaration and the protocol of the Ethical Review Committee of Quaid-i-Azam University Islamabad Pakistan (DAS-1070; Dated July 8, 2015).

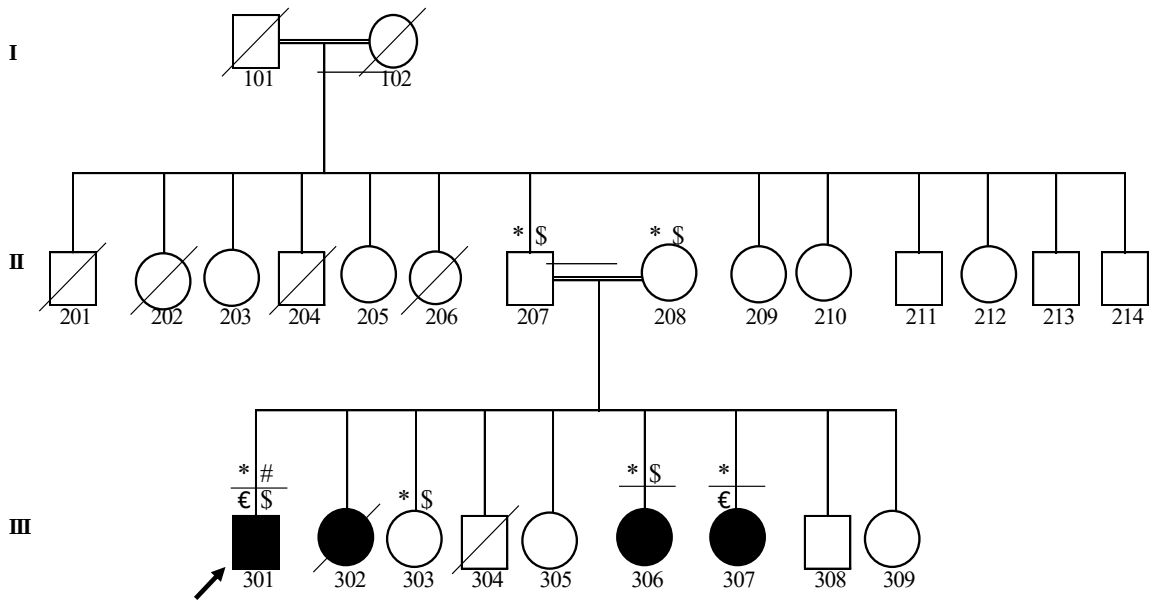


Fig 6.1: Pedigree of family afflicted with Dyggve-Melchior-Clausen like-syndrome. Symbols with above horizontal line: physically observed; Symbol with oblique line: deceased; *: blood sample obtained; #: radiographic examination performed; €: exome sequence carried out; $: Sanger sequence performed.

6.3.2: Clinical description

The consanguineous family of with four affected individuals was recruited. The affected family members were presented with severe to profound ID and marked short stature. They have small head, large ears, short trunk, protruding sternum, scoliosis, rhizomelia, broad hands with short stubby fingers, and club thumb. In the lower limbs, there is genu valgum or knock knees , waddling gait, broad feet and pes planus or flat feet (Table 6.1).

Table 6.1: Clinical characteristics of patients

Pedigree IDs	**301**	**306**	**307**
Sex, age (Years)	M, 40	F, 31	F, 29
Intellectual disability*	Profound	Profound	Severe
Short stature	+	+	+
Microcephaly	+	+	+
Large ear	+	+	-
Short trunk	+	+	+
Sternal protrusion	+	+	+
Scoliosis	+	+	-
Rhizomelia	+	+	+
Broad hands	+	+	+
Short stubby fingers	+	+	+
Club thumb	+	+	+
Genu valgum	+	+	+
Waddling gait	+	+	+
Broad feet	+	+	+
Pes planus	+	-	-

*as per criteria of the American Psychiatric Association

The roentgenograms of affected subject 301 revealed short neck, prominent mandible, and protruding and barrel-shaped chest. There was scoliosis in the thorasicolumber region. There was narrowing of the upper ribs with the absence of floating ribs. There was marked shortening of long bones of limbs with hypertrophic proximal and distal heads (Fig. 6.2). The carpals were variably short and fused with short and stubby phalanges. There were large gaps between the tibiae and fibulae. Further, pelvic girdle was hyperplastic and there was evidence of decalcification and hyperplasia (Fig. 6.2).

Fig 6.2: Phenotypic manifestion of patient 301. The conspicuous features are short stature, small head, protruding chest and genu valgum. The roentgenograms revealed shortening of long bones of limbs, scoliosis, and narrowing of cervical and thoracic vertebrae and corresponding ribs.

The affected subjects also have delayed developmental landmarks including sitting late, standing late, crawling and walking late, and speech delay. They have certain behavioral problems including speech apraxia, aggression, hyperactivity, psychosis, temper tantrum and attention deficit (Table 6.2).

Table 6.2: Developmental and behavioral features of affected subjects

Pedigree IDs	**301**	**306**	**307**
Sex, age (Years)	M, 40	F, 31	F, 29
Developmental features			
Developmental delay	+	+	+
Standing late	+	+	+
Sitting late	+	+	+
Walking late	+	+	+
Delay speech	+	+	+
Toileting late	+	+	+
Behavioral problems			
Psychosis	++	++	+
Hyperactivity	++	++	–
Aggression	++	++	+
Temper tantrum	++	++	+
Attention deficit	+	+	–
Mood instability	++	++	+
Self-care	–	–	Minor
Sense of self-respect	–	–	+
Speech apraxia	++	++	+
Friendly towards strangers	–	–	+
Head flapping	+	–	–
Hand biting	+	+	–
Sensitive to crowd	+	+	+

+, feature present; ++, severe phenotype; −, feature absent

Anthropometric measurements taken from three affected subjects, which showed delayed development. All subjects were short in height (Table 6.3).

Table 6.3: Anthropometric measurements of family members

Pedigree IDs	**Imran**	**Shugufta**	**Tehmeena**
Sex, age (Years)	M, 40	F, 31	F, 29
Standing height*	117 (<1)	127 (<1)	127 (<1)
Sitting height†	71 (<1)	71 (<1)	76 (<1)
Arm span§	123 (<1)	133 (<1)	130 (<1)
Head circumference‡	51 (<2)	46 (<3)	48 (<3)
Chest circumference	74	74	76
Weight (kg)*	25 (<1)	30 (<1)	35 (<1)

Percentiles are given in parentheses.

All measurements are in cm.; Head circumference is with respect to age and sex.
*Percentiles are from WHO Growth Reference:
http://www.who.int/growthref/who2007_height_for_age/en/

†Kelly AM, Shaw NJ, Thomas AM, Pynsent PB, Baker DJ. Growth of Pakistani children in relation to the 1990 growth standards. Arch Dis Child. 1997;77:401-5.

§With reference to height. Chen WY, Lin YT, Chen Y, Chen KC, Kuo BI, Tsao PC, Lee YS, Soong WJ, Jeng MJ. Reference equations for predicting standing height of children by using arm span or forearm length as an index. J Chin Med Assoc. 2018;81:649-56.

‡James HE, Perszyk AA, MacGregor TL, Aldana PR. The value of head circumference measurements after 36 months of age: a clinical report and review of practice patterns. J Neurosurg Pediatr. 2015;16(2):186-94.

6.3.3: Whole exome sequencing and variant selection

Through phenol chloroform method DNA was extracted from peripheral blood samples. Next the hunt of potential candidate gene(s), whole exome sequencing was performed on 301 and 306, followed by Sanger sequencing carried out on all family members available. For exome sequencing, the sample was processed through Agilent Sure Select Target Enrichment Kit as per the case manufacturer's guide through local vender. The libraries were sequenced with Illumina HiSeq2000/2500 sequencer. The BWA Enrichment application of BaseSpace (Illumina Inc. SanDiego, USA) was employed to analyze the generated reads. Sequence alignment with the reference genome was performed with Burrows-Wheeler Aligner (BWA) (Li and Durbin, 2009), and the variants were called with GenomeAnalysis Toolkit (GATK; McKenna *et al.,* 2010). The called variants were subsequently annotated with Illumina VariantStudio v2.2. Variant filtration was based on the inheritance pattern (autosomal recessive) and parental consanguinity (heterozygous parents and homozygous patients), and homozygous variants were retained. Variants shared by both affected individuals in the homozygous state were prioritized for downstream analysis. Exome sequence data was further scrutinized for homozygous variants that were further checked in the public sequence databases (dbSNP, Exome Variant Server, 1000 Genome Browser, and gnomAD). The filtered variants were further checked in the in-house data of 100 exomes of Pakistani subjects.

6.3.4: In Silico analysis

Human reference sequence GRCH38 was retrieved from UCSC Genome browser (https://genome.ucsc.edu/cgi-bin/hgGateway) Conservation of the genomic region of interest across species was visualized by HomoloGene

(https://www.ncbi.nlm.nih.gov/homologene) was used to check the conservation of genomic region of interest. Computational algorithms Mutation Taster and Protein Variation Effect Analyzer (PROVEAN) were used to predict the effect of the identified deletion on the protein.

6.4: Results

The exome filtration strategy led to the identification of homozygous variant (NM_017653.3:c.1072C>T, p.Gln358*) in *DYM* gene, shared among both affected subjects. This variant is predicted to cause a premature termination leading to a truncated protein. This variant was found to be novel and not observed in any of the public sequence databases like dbSNP, Exome Variant Server, 1000 Genome Browser, and gnomAD in the homozygous state. Sanger results were inconclusive and chromatogram of affected subjects was not of good quality and were repeated. The results were under progress until the submission of the thesis (Fig. 6.3). The primers utilized in Sanger sequencing are given below:

Oligo	**Length**	**Temp. ºC**	**GC%**	**Sequence**
Right primer	20	60.04	50	CAGTCCTTTCCCCTCATCAA
Left primer	20	59.53	45	AGGGATAGCATTTGGCGATA

Product size: 350bp

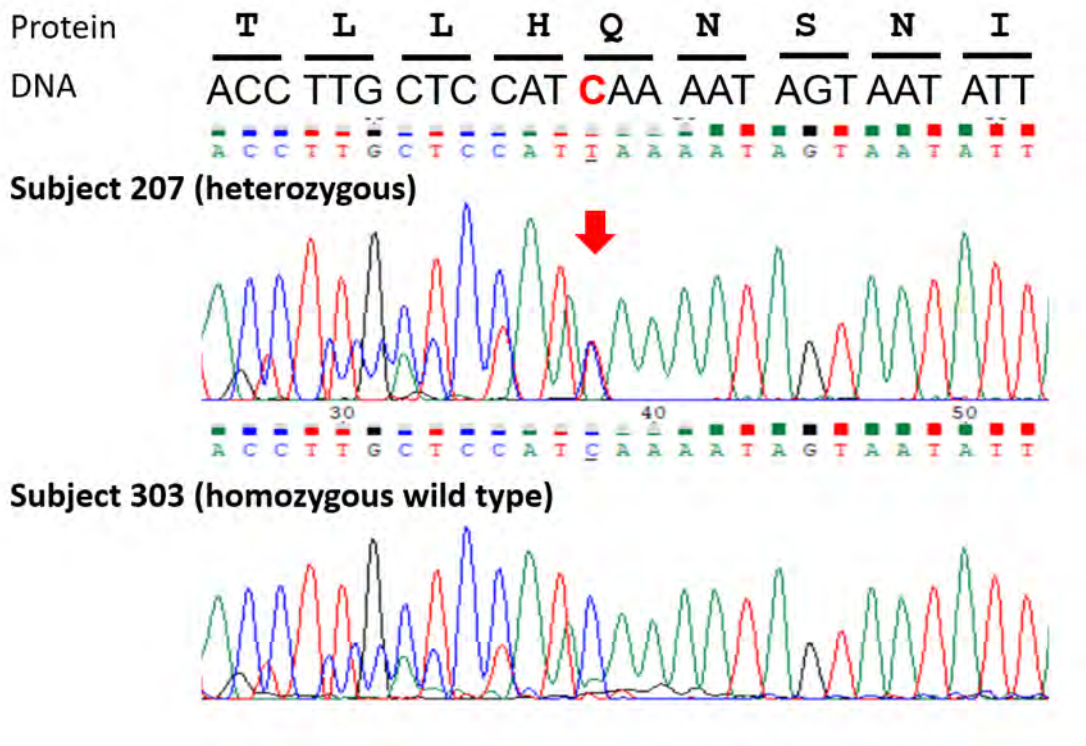


Fig 6.3: Chromatogram showing results of Sanger sequencing. (Chromatogram of affected subjects was not of good quality and were repeated. The results were under progress until the submission of the thesis).

The Q358 (p.Gln358) is highly conserved. Deep phylogenetic analyses through UCSC revealed complete conservation of this amino acid upto Zebrafish (Fig. 6.4). DYM is 669 amino acid protein and various mutation in this gene have been reported prevuiously. A schamtics dipcted the domain structure of DYM protein with previously documented mutations and the novel variant found during the current research is represented in red (Fig. 6.5).

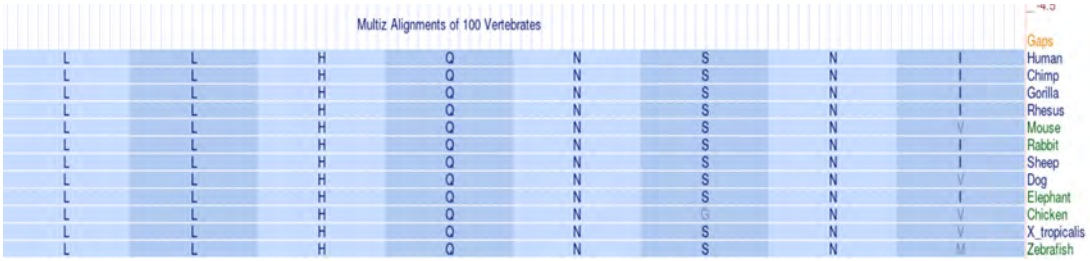


Fig 6.4: Conservation tract of UCSC depicting Vertebrate Multiz Alignment & Conservation ranging up to Zebrafish. This illustrates multiple alignments of 100 vertebrate species and measurements of evolutionary conservation using two methods (phastCons and phyloP) from the PHAST package.

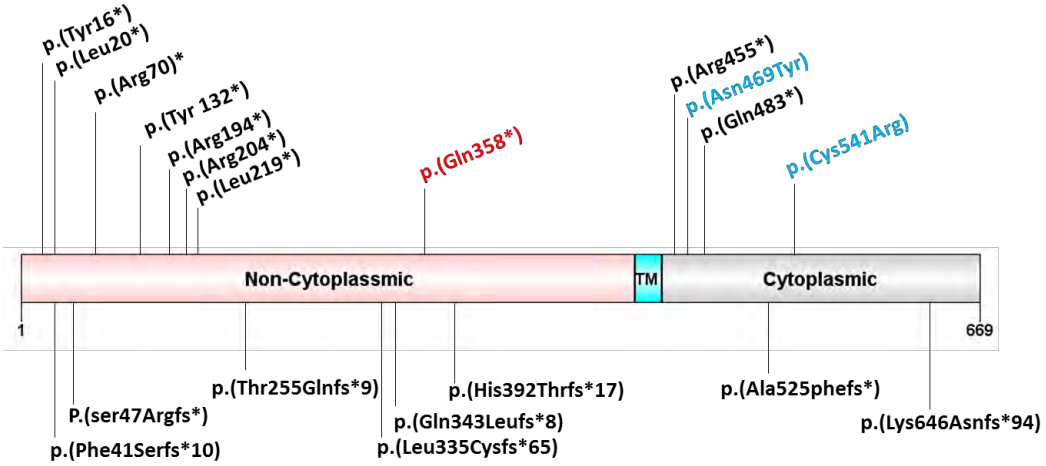


Fig 6.5: Schematic representation DYM protein with all identified homozygous recessive mutations causing autosomal recessive DMC syndrome. Upper half of the schematics shows missense mutations (blue) and nonsense mutations (black), while lower half shows the frameshift mutations. The variant identified is labelled in bold red. TM; Transmembrane domain.

6.5: Discussion

Here, an inbred Pakistani family with skeletal and neurological phenotype has been preseted. Four of the three affected subjects have been evaluated in detail and

were witnessed to have combination of symptoms like dwarfism, microcephaly and cognative impariment. The skeletal phenotype ranged from short stature, sternal protrusion, sciliosis, rhizomelia and genu valgum. The affected subjects also had severe to profound type of ID, inaddition to a ranger of developmental and behavioral anomalies. These features were concordent with DMC syndrome. One of four affected sibs had been deceased. The genetic analyses of this family led to the identification of a novel homozygous variant in *DYM* (c.1072C>T, p.Gln358*), which is corcordent with the malformation.

DMC is an early onset disorder (MIM). The initial signs of DMC syndrome usually develop between the ages of one and eighteen months, while the double-hump appearance of the vertebral bodies and the characteristic feature of the iliac crest (irregular and lacy) appear between the ages of three and four years and last until adulthood (Schorr *et al.,* 1974). In the present family, as reported by the elder family members, the symptoms in the affected subjects appreaed in early childhood and then aggreviated with the time course. Curiously, DMC syndrome has some clinical features that are similar to Morquio disease (mucopolysaccharidosis IV). However, DMC syndrome is distinguished from Morquio disease by the absence of corneal capacity and keratin sulphate excretion in urine, as well as the presence of some specific radiological findings (Khalifa *et al.,* 2011; Gaboon *et al.,* 2020).

Another disorder Smith-McCort dysplasia (SMC, MIM # 607326) has been discovered to be allelic to DMC syndrome, sharing many skeletal phenotypes, but ID and microcephaly have only been observed in later cases (Elalaoui *et al.,* 2011). In both disorders, *DYM* disruption results in abnormal cartilage histology, with abnormal chondrocyte columns containing degenerating cells and rough endoplasmic reticulum inclusions in the growth plate (Latrech, 2013). Unlike the skeletal phenotype, which is

fairly stable among patients, ID is a variable feature in DMC syndrome and was first used to distinguish DMC from SMC (Smith and McCort, 1958). At both the intrafamilial and interfamilial levels, ID ranges from mild to severe type. Some children are clearly hyperactive, have autistic features, and cannot speak, but others (even from the same family) can speak and have a moderate level of ID (Spranger *et al.,* 1976). Yet another disorder, Smith-McCort dysplasia 2 (SMC2, MIM #615222), is caused by a homozygous mutation in the *RAB33B* gene. *RAB33B* encodes a golgi protein involved in cellular trafficking (Alshammari *et al.,* 2012; Dupuis *et al.,* 2015). Importantly, missense mutations in SMC1 result in residual activity and do not result in complete DYM protein degradation. In contrast, the DMC syndrome, which is caused by a truncating mutation, is predicted to result in the complete loss of *DYM* function (Dimitrov *et al.,* 2009). Complete deletion of the DYM protein could suggest brain damage (Neumann *et al.,* 2006; Martínez-Fría *et al.,* 2007). Dupuis *et al.,* (2015) used a mouse knockout (Dym -/-) experiment and found that the DYM protein was depleted, resulting in a defect in brain development and the onset of microcephaly after birth (Dupuis *et al.,* 2015). Concordingly, in the present family, a truncating variant was detected which is likely to cause loss of function of *DYM*, thus leading to the phenotype.

The *DYM* gene is composed of 17 exons, encoding Dymeclin (DYM), a membrane protein spanning the cytosol and mature golgi membranes. This is involved in regulating the growth and function of the Golgi apparatus, as well as transport of proteins and vesicles across the organelle. Cellular localization studies have revealed that DYM is found mainly in golgi apparatus, cytosol, plasma membrane, and extracellular space with confidence scores of 5, 3, 3 and 2, respectively (GeneCards.org). STRING analyses revealed that DYM interacts with HS2ST1,

P2RY12, PNKD, ACO2 and ANOS1. Dymeclin is particularly expressed in the brain, chondrocytes, osteoblast and skin fibroblast (El Ghouzzi *et al.,* 2003; Osipovich *et al.,* 2008).

DYM protein is either entirely missing or has its expression altered in DMC syndrome. According to *In situ* hybridization research on distinct embryonic stages of human development, DYM protein expression is seen in human fetal tissues throughout the human growth processes (Dimitrov *et al.,* 2009). Dymeclin-deficient mice (Dym-/-) have been shown to develop progressive skeletal abnormalities, implying that DMC and SMC have skeletal characteristics in humans (Osipovich *et al.,* 2008). In a prospective study, it would be worthwhile to carry out genotype-phenotype correlation studies for *DYM*.

7. Clinical and genetic characterization of a family presented with a severe form of scoliosis and dwarfism

7.1: Abstract

Skeletal disorders most often result from gain of function mutations in Filamin B (FLNB). Short height, fused vertebrae, and fused carpal and tarsal bones are all symptoms of the rare autosomal recessive illness known as spondylocarpotarsal synostosis syndrome (SCT). Filamins are dimeric actin-binding proteins that control structure and function of the cytoskeleton through the three-dimensional networks of actin. Cartilage growth and condensation of developing vertebrae is mediated by FLNB-encoded filamin B. FLNB-related disorders have been classified into two groups on both clinical manifestations and genetic etiology. The first group include autosomal recessive forms of SCT while the second group include autosomal dominant syndromes including Boomerang dysplasia (BD; OMIM # 112310), Larsen syndrome (LS; OMIM # 150250), Atelosteogenesis I (AOI; OMIM # 108720) and III (AOIII; OMIM # 108721). The current study aimed to report a homozygous missense mutation in the *FLNB* gene that causes SCT syndrome affecting a large family with previously unreported clinical symptoms. This consanguineous family, recruited from North Pakistan, had six affected members, four living and two dead. Affected individuals have a short neck, disproportionately short trunk, and protruding abdomen (secondary to lordosis). Other common characteristics included kyphoscoliosis, winged scapulae, crowded ribs, and pectus carinatum. While clinodactyly is present in one of the affected individuals. Exome analysis revealed homozygous missense mutation (NM_001164317.2:(FLNB_v001):c.220C>T, p.Glu74*) in the coding region of *FLNB* as causative reason. Conclusively, the study described the expanded

clinical spectrum of SCT syndrome in a large consanguineous family. The study results might be helpful to establish genotype-phenotype correlation of the SCT.

7.2: Introduction

Congenital spine deformities like scoliosis, kyphosis, and lordosis arise due to abnormal vertebral development. The spine abnormality is usually a birth presentation or may appear in early childhood. In the affected subject with congenital curves, the anomaly tends to be rigid and resistant to correction. There is high phenotypic and genetic heterogeneities underlying this condition (Giampietro *et al.,* 2003).

Scoliosis may occur as an isolated entity, however, it may also appear as a part of well-established syndromes; for instance, muscular dystrophy, osteogenesis imperfecta, osteochondrodystrophy type of dwarfism, neurofibromatosis, Marfan syndrome, Ehlers-Danlos syndrome, Noonan syndrome, Angelman Syndrome, Prader Willi syndrome, and Rett syndrome (to name the few; OMIM; Giampietro *et al.,* 2013).

A very rare combination of scoliosis and short stature is Spondylocarpotarsal synostosis syndrome (SCT; MIM# 272460) which affects the development of bones of the whole body. SCT is characterized by disproportionately spinal deformity, scoliosis or lordosis, short stature, and synostosis of carpal and tarsal bones (Yasin *et al.,*2021; Shimizu *et al.,*2019). The newborns with the disorder have relatively normal length, but impaired growth of the torso results in marked short stature. Over the developmental course, the vertebrae are misshapen and fused resulting in lordosis and scoliosis. The patients are also presented with clubfoot, facial dysmorphism, cleft palate, dental enamel hypoplasia, joint laxity, clinodactyly, brachydactyly, and conductive hearing loss (Carapito *et al.,* 2016; OMIM). Radiologic examination reveals block vertebrae and carpal and tarsal fusion.

Langer and Moe (1975) described this disease for the first time and at least 47 CA are presented to date (Sing *et al.,*2013; Carapito *et al.,* 2016; Salain *et al.,* 2018; Yasin *et al.,* 2021). SCT segregates in an autosomal recessive inheritance pattern. Pathogenic mutations in filamin B (FLNB), myosin heavy chain 3 (MYH3), and the recently proposed Refilin A (RFLNA) have been implicated. It demonstrates both locus and allelic heterogeneity (Mangaraj *et al.,* 2017; Shimizu *et al.,* 2019). The majority of the patients reported with homozygous variants in *FLNB* (Krakow *et al.,* 2004) which encode cytoskeletal filamin B. However Carapito *et al.,* (2016) identified the association of *MYH3* gene with autosomal dominant SCT. During the study two independent families were investigated; first with typical autosomal dominant SCT while the second with both SCT and multiple pterygium syndrome (MPS; Carapito *et al.,* 2016). Shimizu *et al.,* (2019) described a patient with a typical SCT phenotype who had a novel homozygous frameshift mutation in the refilin A gene (RFLNA).

Filamins are dimeric actin-binding proteins and control the structure and function of the cytoskeleton by arranging the actin into three-dimensional networks. *FLNB*-encoded filamin B play a role in cartilage growth and condensation of developing vertebrae (Baudier *et al.,* 2018; Mangaraj *et al.,* 2017). FLNB protein is composed of two tandem calponin homology domains also called actin-binding domains (ABD) at the N-terminus and 24 filamin repeat regions. These repeats are interrupted by two hinges between 15 to 16 and 23 to 24 repeats (Salian *et al.,* 2018). Filamin connects the cell membrane with the cytoskeleton mechanically and functionally by interacting with various signaling proteins which are present in cytosol and transmembrane receptors. Additionally, filaments also plays important role in skeletal development by roles in intracellular signaling (Stossel *et al.,* 2001; Lu *et al.,* 2007). Filamin family has three isoforms in mammals: filamin A (FLNA),

filamin B (FLNB), and filamin C (FLNC). Filamin A was the earliest and considered as the most abundant and widely distributed member of this lineage (2020). Abnormal expression of filamin proteins due to polymorphism in *FLN* genes contributes to the etiology of different congenital anomalies. Polymorphisms in *FLNA* and *FLNC* affect cardiovascular, nervous, and skeletal systems (Nakamura *et al.,* 2011; Zhou *et al.,* 2010) while *FLNB* polymorphisms are involved in two groups of skeletal conditions, indicating a critical role in skeletal development (Robertson *et al.,* 1993). The first group is Spondylocarpotarsal synostosis syndrome (SCT), a recessive condition with bi-allelic missense, nonsense, or frameshift truncating mutations in *FLNB* resulting in the loss of function (Yang *et al.,* 2017; Daniele *et al.,* 2012). The second group are autosomal dominant syndromes including Boomerang dysplasia (BD; OMIM # 112310), Larsen syndrome (LS; OMIM # 150250), Atelosteogenesis I (AOI; OMIM # 108720) and III (AOIII; OMIM # 108721).

Here a large Pakistani family with four patients presenting with a skeletal dysplasia similar to SCT has been described. Detailed clinical and molecular study of this family was carried out and the molecular genetic diagnosis was made.

7.3: Subjects and Methods

The family originates from North Pakistan. They reside in a rural area and are engaged in agriculture. the study protocol was approved by the Ethical Review Committee of Quaid-i-Azam University (DAS-1070; Dated July 8, 2015). The family head provided formal consent for participation in this study. All the material was obtained according to Helsinki-II declaration.

The detailed pedigree drawn with the help of family elders revealed six affected subjects, two of whom had been deceased (Fig. 1). The three available

affected subjects and unaffected parents and sibs were physically examined with the help of local physicians and specialized doctors.

Photographs and anthropometric measurements of the three patients were obtained. Whole body roentgenograms of one patient were obtained. Peripheral blood samples of three affected and three unaffected subjects were taken for molecular study. Blood samples were preserved in sterile K2EDTA tubes in order to extract genomic DNA. DNA was extracted from peripheral blood by using the standard protocol (Phenol-chloroform method). All information and blood samples were obtained according to the Helsinki-II declaration and the protocol of the Ethical Review Committee of Quaid-i-Azam University Islamabad Pakistan (DAS-1070; Dated July 8, 2015). Two patients (404 and 406) underwent whole exome sequencing. The candidate variant in *FLNB* gene was sanger sequenced in all available samples.

7.3.1: Pedigree

There were four affected sibs in the family of which one had been deceased. There were affected sibs had unaffected parents suggesting an autosomal recessive inheritance pattern (Fig. 7.1).

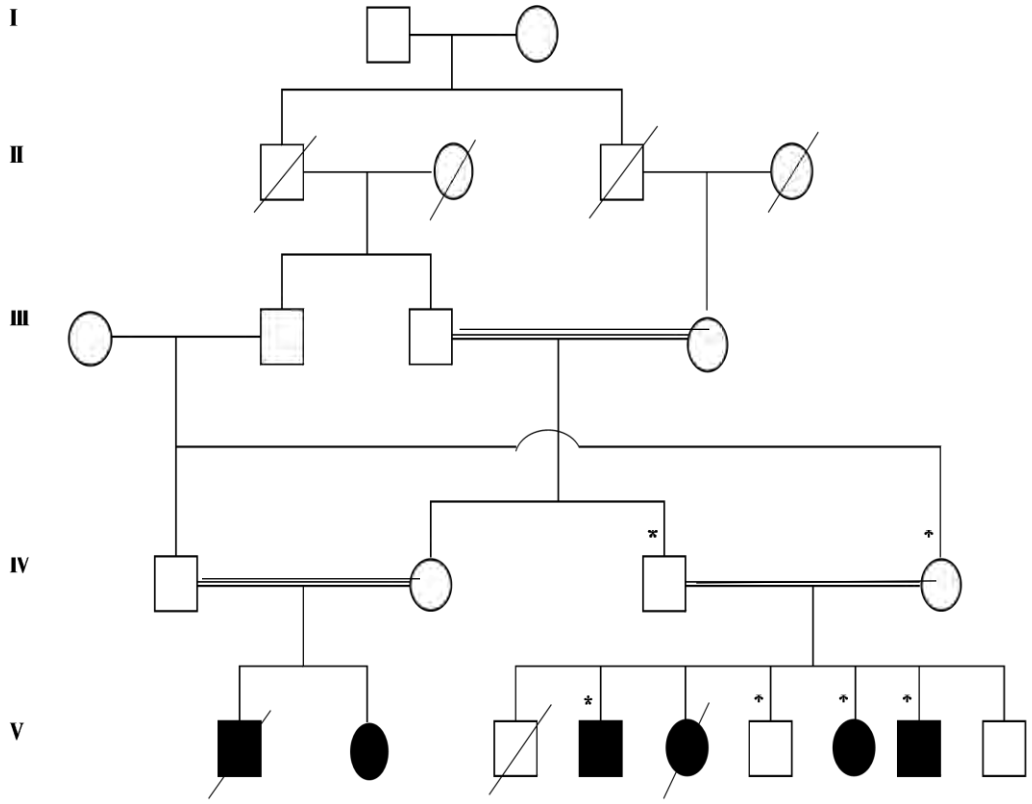


Fig 7.1: Pedigree of family segregating autosomal recessive SCT. Squares represent males, while circles indicate females. Clear shapes show normal while filled ones indicate affected members of the family. The double line between male and female represents a consanguineous union. Oblique line on the symbol represents a deceased individual. Asterisk represent individuals who participated in the study.

7.3.2: Clinical description

The patients have short stature of disproportionate type, short trunk and short neck. They have severe kyphoscoliosis, protruding sternum and abdomen, and winged scapula. Patient 504 comparatively has the most severe symptoms. According to the family, the malformation is progressive and the symptoms become prominent with growing age. On the physical examination, the patients have rounded faces with

anteverted nares and otherwise unremarkable facial features. Patients have a complaint of knee pain and backache. They have difficulty in walking and gait problems. According to the family, the deceased patients were also having the same symptoms. They all have normal IQ, hearing, and vision. All patients were attending a normal school. The clinical features of the affected individuals are presented in (Table 7.1).

Table 7.1: Phenotypic variability in patients

Variables	**404**	**406**	**407**
Sex, age (years)	M, 12	M, 10	F, 8
Short stature, disproportionate	+	+	+
Short trunk	+	+	+
Kyphoscoliosis	+	+	+
Winged scapulae	+	+	+
Crowded ribs	+	+	+
Pectus carinatum	+	+	+
Protruding abdomen (secondary to lordosis)	+	+	-
Short neck	+	+	+
Dysmorphic face	-	-	-
Face type	Round	Round	Round
Frontal bossing, mild	+	+	+
Hearing loss, conductive	–	–	–
Anteverted nares	+	+	+
Ocular findings	–	–	–
Tooth enamel hypoplasia	–	–	–
Failure of eruption of permanent teeth	N/A	?	?
Clinodactyly	+	–	–
Brachydactyly	–	–	–
Club foot	–	–	–
Flat foot	+	+	+
High arched/cleft palate	–	–	–
Anomalie of internal vital organs	–	–	–

+, feature present; –, feaure absent; ?, could not be ascertained.

Radiographic features of patient 404. The roentgenographic study revealed delayed bone age. There was thoracolumbar fusions, cervical fusion, scoliosis (thoracic and lumbar), lordosis (thoracic and lumbar) and sacral anomaly (Fig. 7.2). In the upper limbs, carpal synostosis, capitate-hamate coalition, and lunate-triquetrum

fusion was observed. There was an impression of bilateral clinodactyly. In the lower limbs, there was tarsal fusion. However, coxa vara, epiphyseal dysplasia (femur, tibia, fibula), limited elbow extension and short metacarpals (IV or V) were not witnessed (Fig. 7.2).

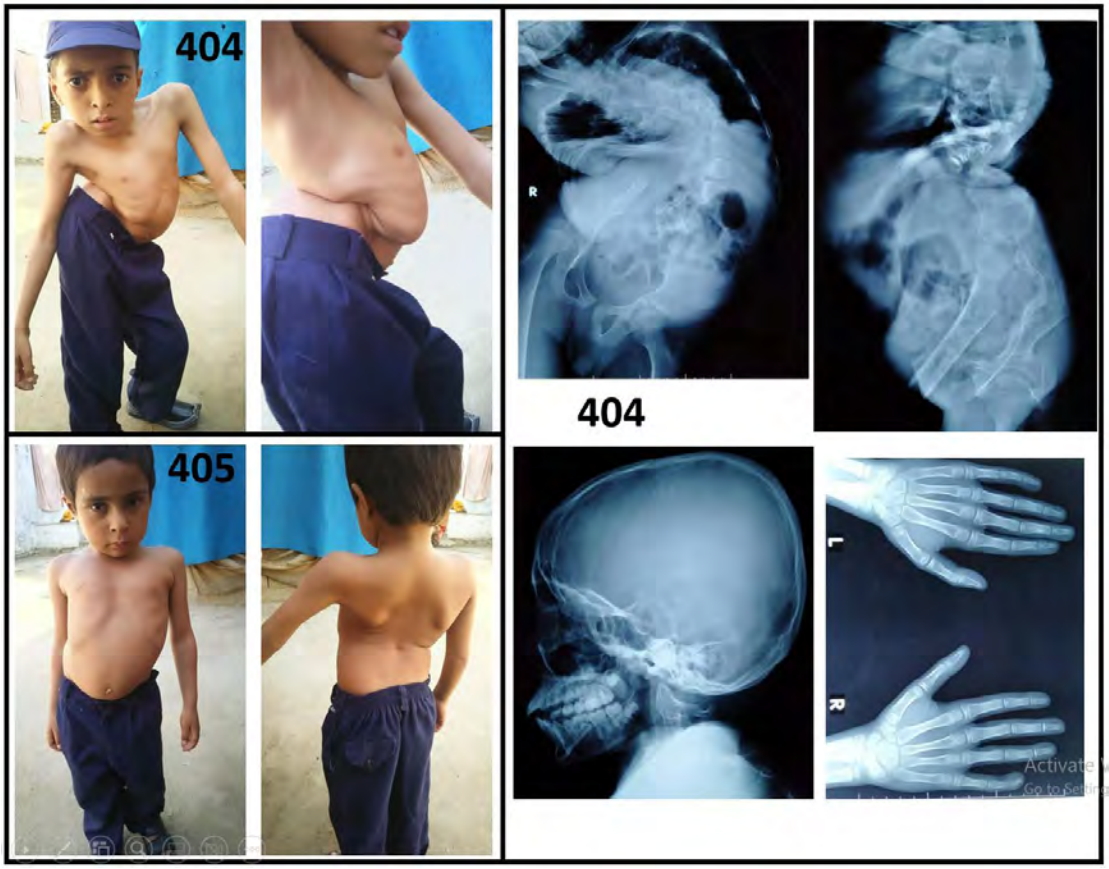


Fig 7.2: Phenotypic expression of patients 404 and 406.

7.3.3: Genetic analyses

Exome filtration identified c.220C>T in exon 1 of the *FLNB* gene that would result in premature termination of the protein (p.Gln74*; Fig. 7.3). The genomic positon of identified variant is given in (fig. 7.4).The nonsense/stopgain variant c.220C>T in exon 1 of *FLNB* gene was not found in the public sequence databases like 1000 Genomes Project, ExAC, or gnomAD. This nucleotide is highly conserved across species. Several bioinformatic prediction tools were utilized to evaluate the identified variant. MutationTaster showed the score of 6.0, PROVEAN predicted it with a score of -10.352, M-cap predict as possibly damaging and MutPred-LOF predict pathogenic with a score of 0.545. Sanger sequencing was used to validate the identified candidate variant in 4 family members. The mutation was homozygous in the affected subject, whereas it was either normal or heterozygous in the normal subjects (Fig. 7.5). All species of vertebrates shared a conservation of the deleted amino acids (Table. 7.2). A homology model of FLNB-wildtype and mutant proteins revealed that the stop gain mutation p.Q74* was located at a Calponin homology domain (or CH domain) of FLNB protein which is a family of actin-binding domains found in both cytoskeletal proteins and signal transduction protein.

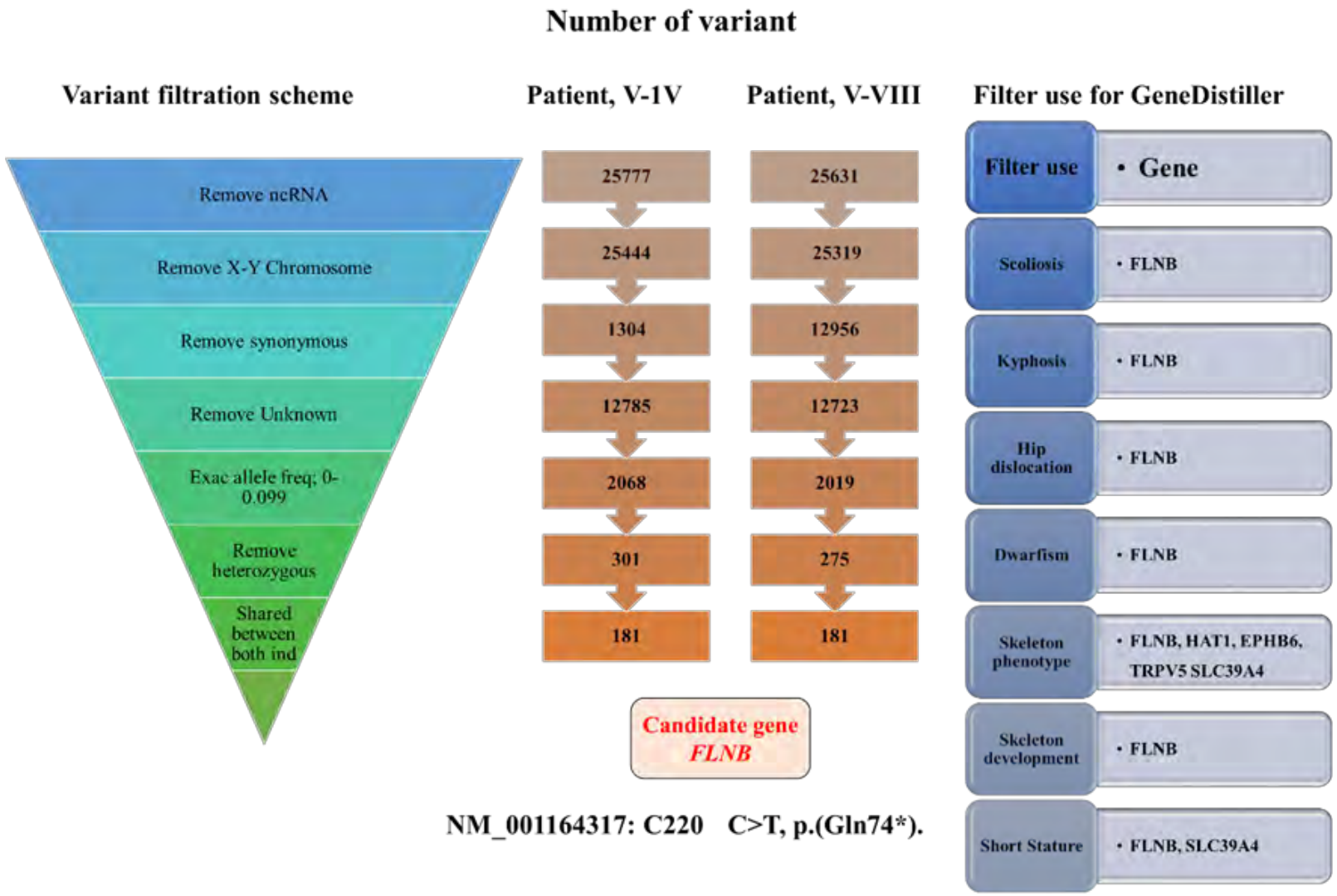


Fig 7.3: Summary of exome filtration scheme.

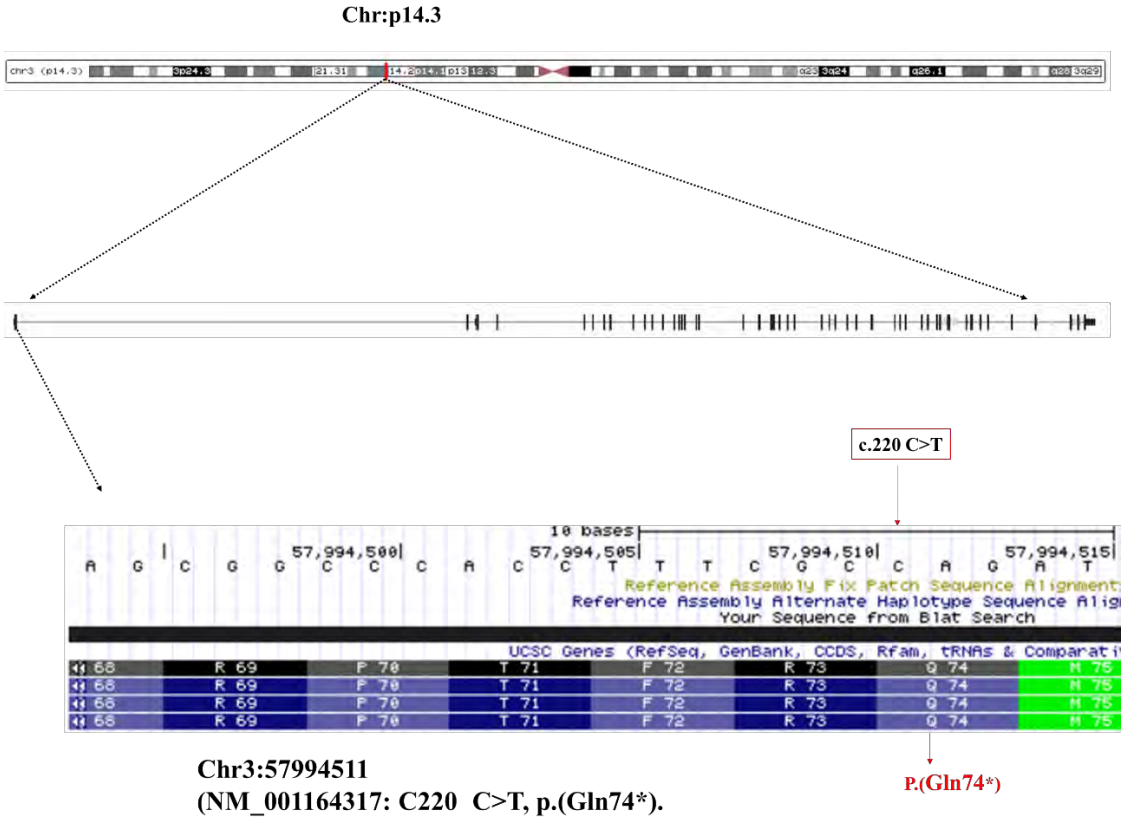


Fig 7.4: Genomic localization and location of the detected variant in *FLNB* at Chr3:p14.3. cDNA change and altered protein position is represented in red.

(A)

Oligo	Length	Temp. ºC	GC%	Sequence
Right primer	20	58	50	ACCTCAAGTGCGTGAACAAA
Left primer	20	59	45	CGATGGACACGAGCTTGATG

Product size: 330bp

(B)

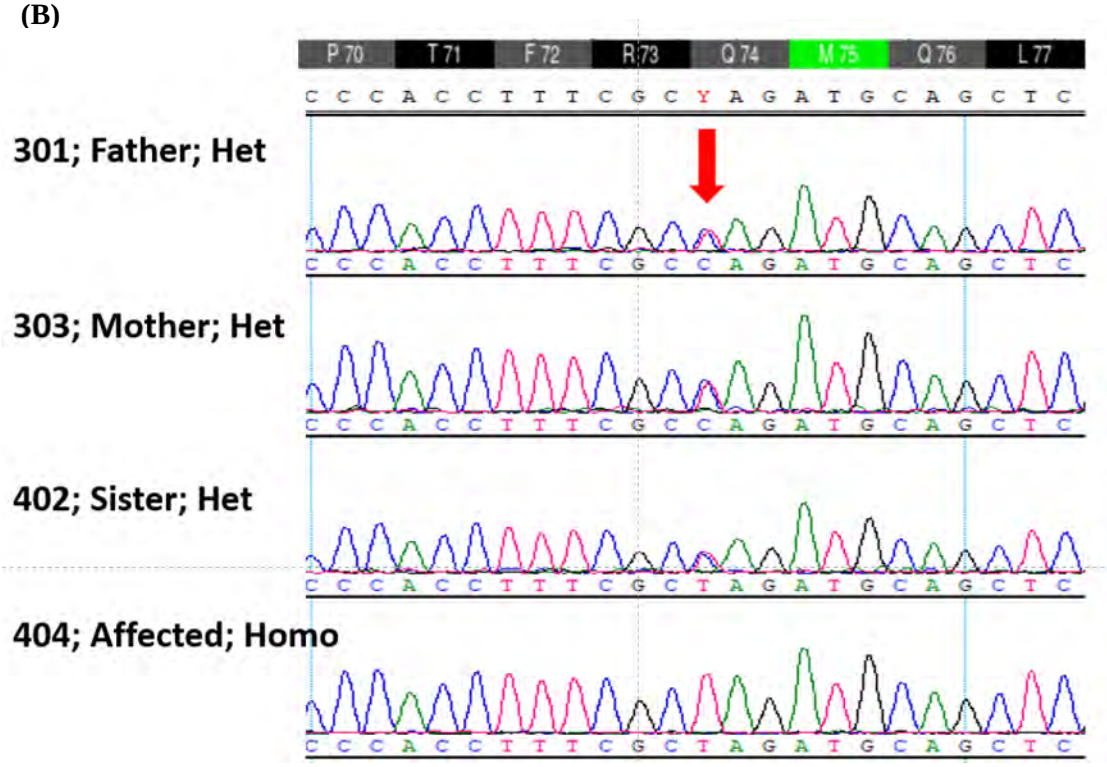


Fig 7.5: Primer sequnces for sanger sequencing (A). Chromatogram depicting the variant (c.220C>T in exon 1 of *FLNB*) segregating in the family (B).

Table 7.2: Conservation of mutated amino acid across different species

Species	**Start AA**	**Amino acid sequence**											**End AA**
H. Sapiens	69	R	P	T	F	R	Q	M	Q	L	E	N	79
Mutated	69	R	P	T	F	R	*						79
Gorilla	69	R	P	T	F	R	Q	M	Q	L	E	N	79
Mus musculus	69	R	P	T	F	R	Q	M	K	L	E	N	79
Pan troglodytes	69	R	P	T	F	R	Q	M	Q	L	E	N	79
Xenopus tropicalis	69	R	P	T	F	S	Q	M	Q	L	E	N	79
Sorex Araneus	69	R	P	T	F	R	Q	M	Q	L	E	N	79
Loxodonta Africana	69	R	P	T	F	R	Q	M	Q	L	E	N	79

AA, amino acid;

7.4 Results

FLNB-related SCT has been reported in less than 20 different families (Salian *et al.,* 2018; Table 3). Affected people have *FLNB* truncating mutations that are either homozygous or compound heterozygous. All of the identified variants are minor insertions or deletions (indels) or single nucleotide variants (SNVs), which will shorten protein translation. It has been demonstrated that several of these mutations cause nonsense-mediated decay (Yasin *et al.,* 2021) and that SCT results from the lack of filamin B protein. Schematics of the FLNB protein domain structure are shown, with previously described mutations marked in black and the variant discovered in the present study highlighted in red. The filamin protein's actin binding domain is the site of the Gln74Ter variant (Fig. 7.6).

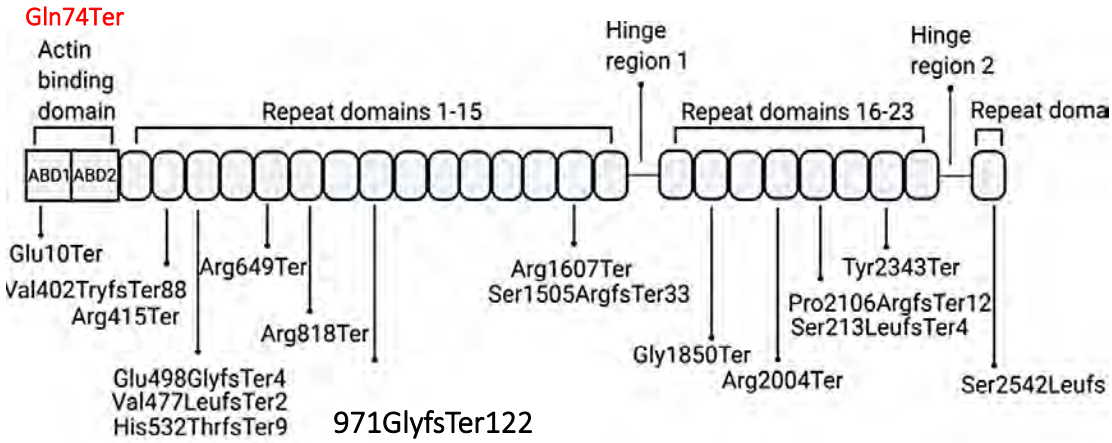


Fig 7.6: Schematic representing the already known homozygous recessive *FLNB* associated with autosomal recessive SCT. The variant detected in this study is shown in red. Acting binding domain 1 and 2 (ABD1 and ABD2 respectively). Two hinge regions, Hinge 1 and Hinge 224 separate (Yasin *et al.,* 2021). The variant detected in the present study is likely pathogenic as predicted by various online tools. (Adapted from Yasin *et al.,* 2021).

Table 7.3: Pathogenicity of the reported mutations in *FLNB*

S. No	Nucleotide	Protein consequence	Mutation type	Exon	UCSC Location	MutationTaster	M-Cap Score	Reference
1	c.1945C >T	p.(Arg649*)	Splice site	13	chr3:58094188	Disease causing (6.0)	Possibly pathogenic (0.074)	(Krakow *et al.,* 2004)
2	c.2452C > T	p.(Arg818*)	Splice site	16	chr3:58095865	Disease causing (6.0)	Possibly pathogenic (2.126)	(Krakow *et al.,* 2004)
3	c.4819C > T	p.(Arg1607*)	Splice site change	28	chr3:58121853	c.4912C>T Disease causing(6.0)	Possibly pathogenic (0.373)	
4	c.6408delC	p.(Pro2137Argfs*12)	Frameshift	39	chr3:58135902_58135902	Disease causing	Only score missense variant	
5	c.7029 T > G	p.(Tyr2343*)	Stop gain	43	chr3:58148888	Disease causing (6.0)	Not found	
6	c.4671G > A	p.(=)	Splice site	27	chr3:58118557	Disease causing		(Farrington-Rock

								et al., 2007)
7	c.5548G > T	p.(Gly1850*)	Splice site	33	chr3:58129205	Disease causing (29)	Possibly pathogenic (0.394)	
8	c.6010C > T	p.(Arg2004*)	Splice site	36	chr3:58133956	Disease causing (32)	Possibly pathogenic (0.027)	(Mitter *et al.,* 2008)
9	c.7621dupG	p.(Ser2542Leufs*82)	Frameshift					(Yang *et al.,* 2017)
10	c.28G > T	p.(Glu10*)	Stop gain	1	chr3:57994319	Disease causing (6.0)	Possibly pathogenic (0.254)	(Salian *et al.,* 2018)
11	c.429delinsCT	p.(Gln143Hisfs*2)	Frameshift					
12	c.1204delG	p.(Val402Trpfs*88)	Frameshift	8	chr3:58084494_58084494	Disease causing		
13	c.1243C > T	p.(Arg415*)	Splice site	8	chr3:58084533	Disease causing (6.0)	Possibly pathogenic (0.329)	
14	c.1493delA	p.(Glu498Glyfs*4)	Frameshift	10	chr3:58089695_58089695	Disease causing		
15	c.1592dup	p.(His532Thrfs*9)	Frameshift	10	chr3:58089794_58089795	Disease causing		

16	c.6317delC	p.(Pro2106Argfs*12)	Frameshift					
17	c.2911dupG	p.(Ala971Glyfs*122)	Frameshift	20	chr3:58107015_58107016	Disease causing		(Yasin *et al.,* 2021)
18	c.220C>T	p.(Gln74*)	Splice site	1	chr3:57994511	Disease causing (6.0)	Possibly pathogenic (0.481)	Present study

NM_001457.4, NM_001164317, (NM_001164318) Predicting the pathogenicity index in FLNB reported variant associated SCT (Mutation Accessor).

The refilin (*Rfln)* genes are limited to vertebrates only and no homologs have been found in *Drosophila* and *C. elegant* genomes. In vertebrates two orthologs of *Rfln* exist, Refilin A (RFLNA) and Refilin B (RFLNB), also known as FAM101A and FAM101B, respectively. Under TGF-β stimulation, filamins bind to RFLNs transforming connecting actin filaments into parallel bundle structures that accumulate to form perinuclear actin caps (Fig. 7.7). This is important for cell migration, differentiation, endochondral ossification and skeletal development (Baudier *et al.,* 2018; yang *et al.,* 2017).

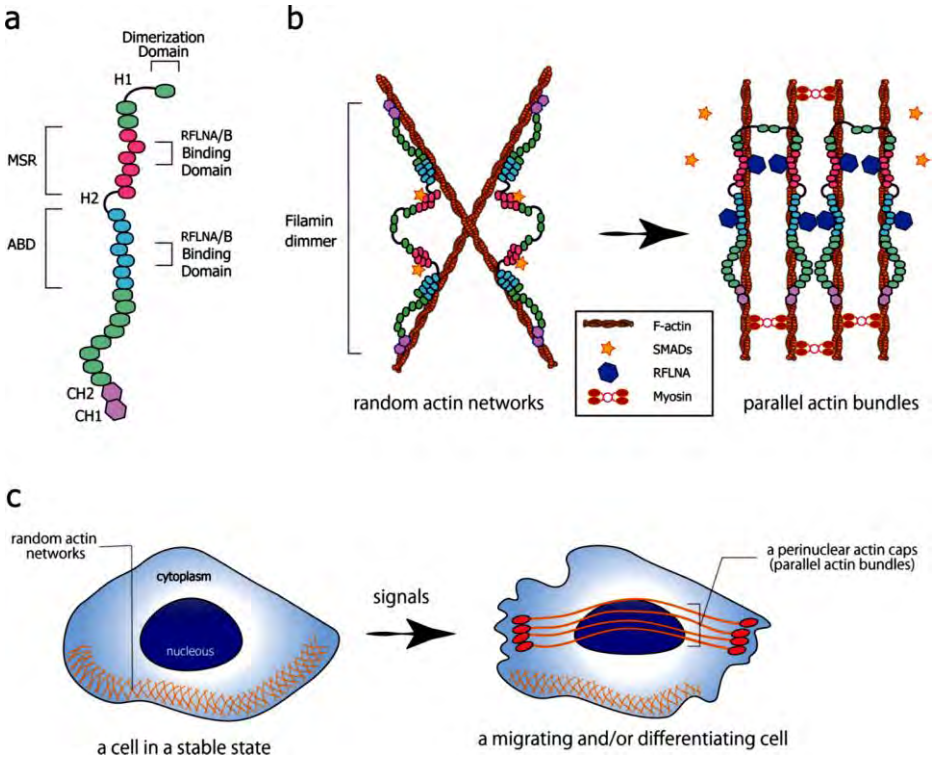


Fig 7.7: A schematic illustration of FLNB and formation of parallel actin bundles and perinuclear caps. a. monomeric chains of actin filament. b. Structure of vertebrate filamin dimer and formation of parallel actin bundles. c. parallel actin bundles accumulate and produce perinuclear actin caps. These molecular dynamics are essential of cell migration and differentiation (Adopted from Shimizu *et al.,* 2019).

7.4.1: FLNB allelic disorders

As stated earlier, mutations in *FLNB* cause five well-characterized syndrome, namely atelosteogenesis, type I; (AO1; OMIM 108720), atelosteogenesis, type III; (AO3; OMIM 108721), boomerang dysplasia; (BOOMD; OMIM 112310), Larsen syndrome; (LRS; OMIM 150250), and spondylocarpotarsal synostosis syndrome; (SCT OMIM 272460). A comparison of their clinical features is presented in (Table 7.4). Only the spondylocarpotarsal synostosis syndrome (SCT) segregates in autosomal recessive fashion.

Table 7.4: Clinical characterization of *FLNB* Gene associated with different skeletal dysplasia

Phenotype/variables	**Atelosteogenesis Type I**	**Atelosteogenesis, type III**	**Boomerang dysplasia**	**Larsen syndrome**	**Spondylocarpotarsal synostosis syndrome**	**Present family with SCT**
MIM No	108720	108721	112310	150250	272460	272460
Mode of inheritance	AD	AD	AD	AD	AR	AR
Short stature, disproportionate	-	-	+	+	+	+
Short trunk	-	-	-	-	+	+
Frontal bossing	+	+	-	-	+	+
Micrognathia	+	+	-	-	-	-
Dysmorphic face/Facia dysmorphism	-	-	-	-	+	?
Mid face hypoplasia	+	+	-	-	-	-
Hearing loss	-	-	-	+	+	?
Vision impairment	-	-	-	-	+	?
Anteverted nares	-	-	-	-	+	+
Cleft palate	+	+	-	+	-	?
Cleft lip	-	-	-	+	-	?
Dentition/enamel	-	-	-	-	+	?

anomaly						
Short Neck	+	+	-	-	+	+
Winged scapulae	-	-	-	-	+	+
Crowded ribs	-	-	-	-	+	+
Pectus carinatum	-	-	-	+	+	+
Pectus excavatum	-	-	-	+	-	?
Protruding abdomen	-	-	-	-	+	+
Foramen magnum stenosis	-	-	-	-	+	?
Platybasia	-	-	-	-	+	?
Spine						
Fusion of vertebral arches	-	-	-	-	+	?
Fusion of posterior processes	-	-	-	-	+	?
Scoliosis	-	+	-	+	+	+
Lordosis	-	-	-	+	+	+
Thoracolumbar fusions	-	-	-	-	+	?

Sacral anomaly	-	-	-	-	+	?
Coxa vera	_				+	_
Coxa velga	-	-	-	-	-	+
Epiphyseal dysplasia (femur, tibia, fibula)	-	-	-	-	+	?
Limited elbow extension	-	-	-	-	+	?
Carpal synostosis	-	-	-	-	+	+
Capitate-hamate coalition	-	-	-	-	+	+
Short metacarpals (IV or V)	-	-	-	-	+	?
Clinodactyly	-	-	-	-	+	+
Brachydactyly	+	-	-	-	+	?
Tarsal fusion (in some patients)	-	-	-	-	+	+
Flat feet	-	-	-	-	+	?
Club foot	+	+		+	+	?
Limited joint mobility	-	-	-	-	+	?
Elbow, knee and	-	-	-	-	+	

shoulder pterygium						
Single palmar crease	-	-	-	-	+	
Others	-	-	-	-		
Pterygium colli	-	-	-	-	+	**?**
Ocular findings	-	-	-	-	+	**?**

+, feature present; -, feature absent; ?, not reproted

7.5: Conclusion

This study confirms the molecular diagnosis of SCT in this family and expands the mutation spectrum of *FLNB*, though no much phenotypic variability was witnessed from the known clinical spectrum of SCT. The information gleaned through this study would be very valuable for the genetic counseling of this family.

8. *De novo* pathogenic variants in cilia and flagella associated protein 46 (*CFAP46*) cause craniofacial anomalies

8.1: Abstract

Diseases associated with *CFAP46* include Retinitis Pigmentosa 63 and Optic Atrophy 8. The gene is predicted to be located in the axoneme and is likely involved in axoneme assembly, playing a role in the central apparatus of cilium motility. Next-generation sequencing has facilitated the identification of genetic causes of various congenital anomalies so the objective of the current study was to identify candidate genes that may contribute to craniofacial anomalies. Subsequently functional data would support the functional roles of the variants identified. Candidate genes were found using exome sequencing. The disease was modeled in vivo using CRISPR/Cas9-mediated genome editing in frog tadpoles, and the functional effects of the patient variant on candidate protein function were evaluated. The whole mount *in situ* hybridization technique was then adapted to explore the role of the CFAP46 protein in embryonic development. This study describes a single female patient with a novel de novo compound heterozygous mutation p.L605P and p.M319V in the *CFAP46* gene. The patient had a number of syndromic features, including severe craniofacial anomalies. The craniofacial anomalies induced by the CRISPR/Cas9-mediated genome editing of *CFAP46* mimicked the patient's condition. Although the craniofacial anomalies comparable to patient phenotypes were the focus of the present research, gastrulation disorders with neural tube defects were also prominent. Since the non-overlapping CRISPRs were so effective, it is likely that *CFAP46* is crucial to development. Further evidence from whole-mount in situ hybridization indicates that *CFAP46* is essential for neural crest development and that abnormalities of the face can result from CFAP46 protein deficiency. This work presents the first evidence of

human craniofacial anomalies from *CFAP46* mutations with both clinical and functional support for this hypothesis.

8.2: Introduction

8.2.1: Genome Editing

Genome editing (GE) alters a particular DNA segment by base substitutions, insertions and/or deletions (indels) in the target sequences (Aglawe *et al.,* 2018). Direct targeting and modification of the genomic sequences across all eukaryotic cells are now possible; thanks to genetically engineered or bacterial nucleases. GE increased our ability to understand the genetic process underlie a specific disease by accelerating the development of more accurate models of disease conditions (Manghwar *et al.,* 2019). GE uses a variety of methods, including the use of zinc finger nucleases (ZFNs), which are targetable DNA cleavage proteins used to cut DNA sequences at any site. Transcriptional activator-like effector nucleases (TALENs), which produce double-stranded breaks (DSBs) in target sequences and, as a result, trigger DNA damage response pathways, can also be utilized in genome editing. Most recently, clustered regularly interspaced short palindromic repeat (CRISPR)/CRISPR-associated nuclease 9 (Case9) system (Li *et al.,* 2020) is proving game changing in editing the genome. CRISPR/Cas9 is an RNA-guided endonuclease that specifically targets DNA sequences via nucleotide base pairing. Although ZFNs and TALENs have been extensively employed for genetic engineering in human, animal, and plant cells since 2002 and 2011, respectively, there are still significant restrictions that prevent their efficient usage. ZFN has a low degree of specificity and frequently induces off-target mutations. ZFN and TALEN vector construction is costly and time-consuming. As a result, since 2013, the focus has been shifted to the

application of CRISPR/Cas9 and, more recently, to recently discovered CRISPR/Cas9 variants (Li *et al.,* 2020; Cornu *et al.,* 2017; Ghosh et al., 2019).

Gene therapy is a technique used to change or replace an undesirable or faulty gene in a cell. Since many human disorders are caused by genetic mutations or incorrect gene expression, thus, gene therapy usually refers to human gene therapy. Due to its tremendous potential to treat hereditary disorders, gene therapy has attracted the attention of scientific and pharmaceutical communities. Human gene therapy has been the focus of further clinical studies and research since it was initially introduced in the late 1980s and early 1990s. However, despite these efforts, the field has advanced very slowly due to a variety of obstacles, including the difficulty of accurately editing a gene (Rosenberg *et al.,* 1990). However, due to the recently discovered clustered regularly interspaced short palindromic repeats (CRISPR) and CRISPR-associated enzyme (Case) technologies, the area of gene therapy is now rapidly growing and becoming more adaptable for the treatment of human genetic disorders. CRISPR/Cas9 genome editing is uses the CRISPR/Cas9 system that bacteria or archaea employ to cut and destroy invading DNA from sources including viruses (Makarova *et al.,* 2020). Hence, gene knockout was the first and most advanced use of the CRISPR/Cas9 approach.

CRISPR/Cas9-based gene therapy can be utilized to delete, replace, or fix abnormal, disease causing genes. Although CRISPR/Cas9 technology for genome editing was first described in 2012 (Gasiunas *et al.,* 2012; Jinek *et al.,* 2012), it has evolved over the last ten years into and numerous techniques for basic and practical research. Gene knock-in/-out, base editing, and prime editing are a few of them that have demonstrated remarkable promise in gene therapy (Fig. 8.1).

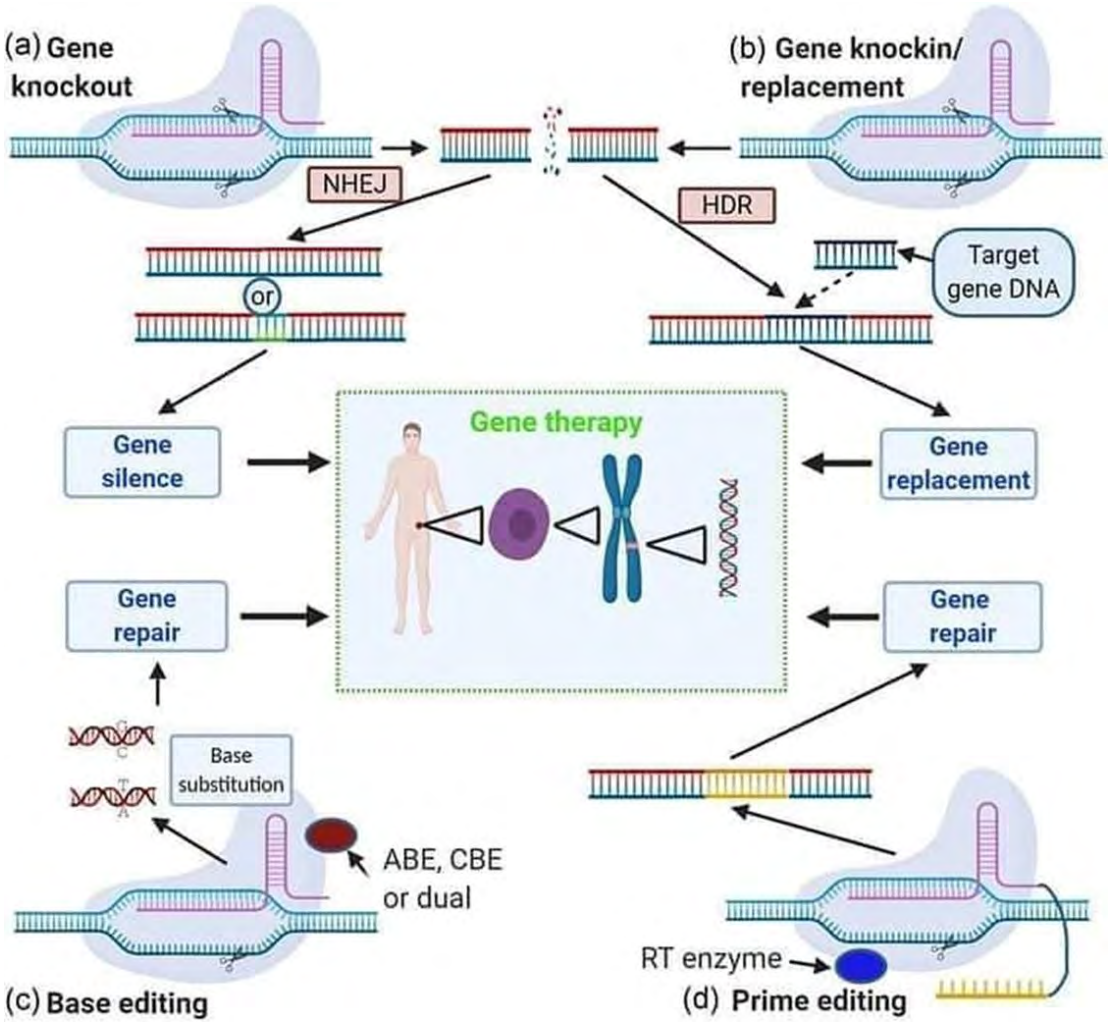


Fig 8.1: Four major strategies for CRISPR/Cas9 mediated gene therapy.

Abbreviations: HDR, Homology-directed repair; NHEJ, Non-homology end-joining; ABE, Adenine base editor; CBE, Cytidine base editor (Adapted from Zhang, 2021).

8.2.2: The Neural Crest

The neural crest (NC), a special collection of migrating pluripotent cells, differentiate into a wide range of cell types, including those that build the facial skeleton and the peripheral nervous system. The strong regulation of NC migration, specification, and differentiation during the early developmental stages is carried out

via a complex regulatory system (Fig. 8.2). Previously, NC development has been extensively studied by using different animal models including *X.tropicalis* (frog) embryos, however, the current loss-of-function assay for investigating NC formation in *Xenopus* embryos is mostly based on the knockdown technique of morpholino antisense oligonucleotides (Liu *et al.,* 2016).

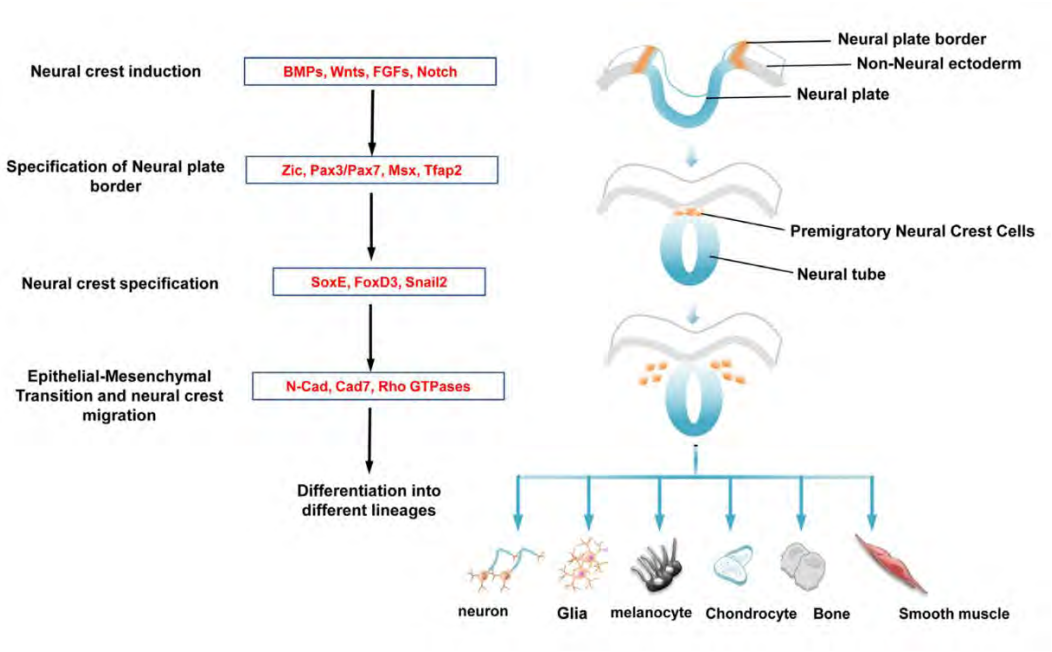


Fig 8.2: A gene regulatory network regulates neural crest development (Adapted from Liu *et al.,* 2016).

8.2.3: Craniofacial anomalies

8.2.3.1: Early embryonic development

Scientists now have a better understanding of how the craniofacial regions develop and how improper development leads to malformations due to advancements in developmental biology and human genetics (Rice, 2005; Van *et al.,* 2016).

The human face not only expresses emotions but also functions as a portal for breathing, eating, and talking. People are fascinated by their own features and notice

even the slightest variations from "normal" facial development. Sperber and Sperber (2014) observed normal embryological development of the human face and cranium. The inner cell mass of the embryonic blastocyst forms the three germ layers, ectoderm, mesoderm and endoderm. At day 20, the ectoderm differentiate into cutaneous and neural components through the formation of the neural crest and neural plate, which folds midline to produce the neural tube (Sperber and Sperber, 2014). The neural tube forms the central nervous system while neural crest forms the peripheral nervous system as well as other elements of the developing organism.

Neural crest cells are crucial in the embryological development of craniofacial structures with the ectomesenchymal tissue of the neural crest creating a distinct pluripotent tissue layer. These cells differentiate into cartilage, bone, ligaments, muscles, and arteries, and are the primary source of connective tissue throughout the body. Congenital abnormalities result from disruptions in the normal migration and differentiation of these cells, which can have serious implications. Most craniofacial deformities develop in the first 12 weeks of gestation, a crucial time for organogenesis (Bronner *et al.,* 2012).

8.2.3.2: Classification of craniofacial anomalies

Lack of understanding of the embryology and the etiology of anomalies has hindered efforts to classify them. In 1981, the Committee on Nomenclature and Classification of Craniofacial Anomalies of American Cleft Palate Association proposed a straightforward classification scheme with five categories that provides a simple overview of current knowledge of the origins, diagnosis, and treatment of the most common craniofacial anomalies. They categorized craniofacial malformations based on the variety of their etiologies, anatomical structures, and treatments. The

most common craniofacial malformations are listed below, along with their etiologies, assessments, and treatments.

1:Facial clefts/encephaloceles and dysostosis

2: Atrophy/hypoplasia

3:Neoplasia/Hyperplasia

4:Craniosynostosis

5:Unclassified

8.2.4: X. tropicalis developmental stages

For *in situ* hybridization, the specific developmental stage of *X.tropicalis* is crucial since various RNA probes only express themselves at particular developmental stages, while in the other stages, expression is absent or less prominent**.** The embryos of *X. laevis* and *X. tropicalis* are strikingly similar in morphology with similar staging using the same table (Nieuwkoop and Faber, 1994), that was created for *X. laevis*. The Zahn drawings and crowd-sourced micrographs (accessible through the community website resource Xenbase, www.xenbase.org) are two additional outstanding tools for embryo stages (Zahn *et al.,* 2017). But having a complete staging series for *X. tropicalis* as well is more useful due to some variances in embryological appearance. In 2019 Kakebeen and Wills gave the full staging series for *X.tropicalis* (Fig. 8.3).

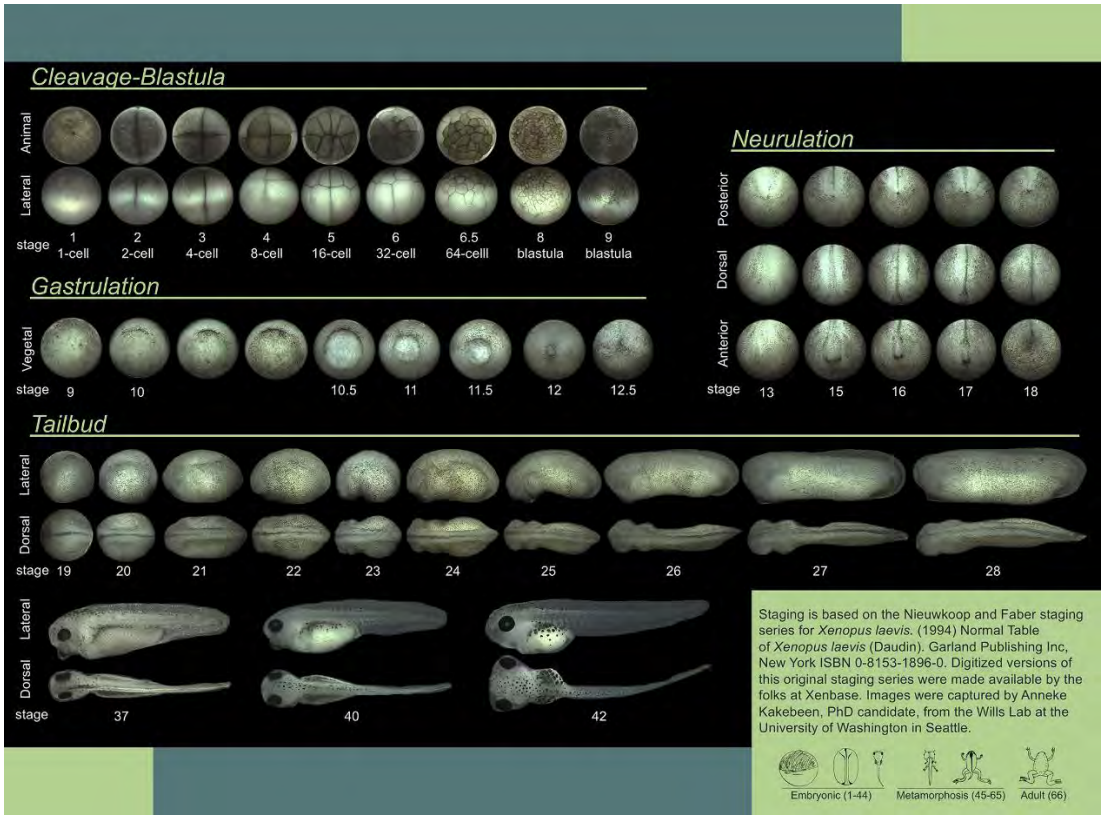


Fig 8.3: Full staging series in *X. tropicalis* from signle cell to free swimming tadpole; (adapted from, Kakebeen and Wills, 2019).

Understanding the molecular etiology of a specific developmental disorder is just the first step in figuring out whether a mutation causes a particular condition. Combining human genetics, developmental biology, and a number of other biomedical fields will result in a more comprehensive etiological picture and, as a result, more precise and effective treatment (Li *et al.,* 2020). The current study is a combination of clinical genetics, developmental biology and functional study in order to provide evidence that *CFAP46* plays important roles in the development of skull and vertebral column and loss of *CFAP46* function results in craniofacial and neural tube defects.

8.3: Subjects and Methods

Case report consent forms were obtained according to the institutional policies of Yale New Haven Hospital. A 27-year-old female subject was reported with a syndromic condition including craniofacial anomalies including cupid's bow, small jaw, and narrow palate (Fig. 8.4). According to Yale University Institutional Animal Care and Use Committee protocols, Xenopus were kept and taken care of in the Pediatrics Genomics Aquatics Facility.

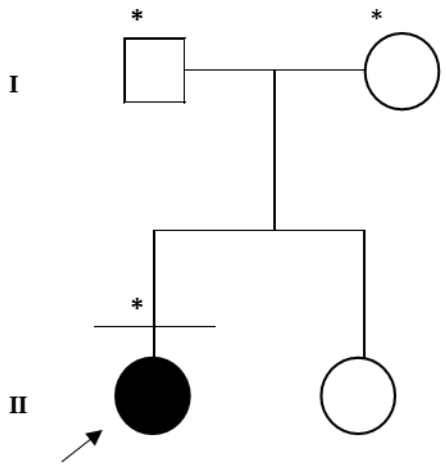


Fig 8.4: Pedigree of family showing unaffected parents and affected child.

Symbol with horizontal line above; Physically examined, symbol witth * : Exome sequences.

8.3.1: Whole exome sequencing (WES)

Each patient and their healthy parents were analyzed as trios using whole exome sequencing (WES). Standard procedures were used to isolate genomic DNA from whole blood. DNA was sheared and exome capture was performed using xGen target capture kits (Integrated DNA Technologies, Inc., Coralville, IA). 99 base pair-

end sequencing was then performed on the HiSeq 4000 platform (Illumina, San Diego, Califonia) the Yale Center for Genome Analysis used a CLIA sequencing protocol for the probands or a research protocol for parental samples as previously described (Sega *et al.,* 2019). The proband and parent samples were sequenced to a mean depth of ~100× and ~50× coverage, respectively. The sequence reads were converted to FASTQ format and were aligned to human genome build GRCh37/UCSC hg19, and analyzed for sequence variants using a custom-developed analysis tool as well as additional sequencing technology and variant interpretation protocol already described (Tanaka *et al.,* 2015). The probands of affected individuals and their parents were sequenced to a mean depth of 82.6–120.7 independent reads per targeted base across all the samples, with an average of 93.0%–94.4% of targeted bases in all the samples This provides a high-confidence identification of variants, especially de novo variants, across the exomes. After being manually screened, variants were analyzed as potential disease -causing by reviewing clinical and basic science literature on each gene, including PubMed, OMIM, HGMD, and ClinVar, as well as by using other in silico methods.

8.3.2: Genomic study

Chromosomal microarray and whole exome sequencing were used in the initial clinical investigation. Clinical exome analysis reported de novo compound heterozygous variants p.Leu605Pro and p.Met319Val in *CFAP46* (NM_001200049.3; Fig. 8.6).

8.3.3: Genome editing in *X. tropicalis*

Previous studies have claerly demonstrated the utility of *X. tropicalis* in the study of genetic and developmental disorders (Sega *et al.,* 2019; McQueen and

Pownall, 2017; Wen *et al.,* 2017; Boskov*s*ki *et al.,* 2013). The current study used the previously published CRISPR/Cas9-mediated genome editing method in *X. tropicalis* (Bhattacharya *et al.,* 2015). Three non-overlapping, independent CRISPR sgRNAs targeting exon 3, 7, and 8 respectively, of *CFAP46* were designed to generate knockouts.

►►***CFAP46-EX3*:**

CTAGCtaatacgactcactataGGACACAGAACGTACAAATCgttttagagctagaa

►►***CFAP46-Ex7*:**

CTAGCtaatacgactcactataGGGCATCCAACAAACACTCTgttttagagctagaa

►►***CFAP46-Ex8*:**

CTAGCtaatacgactcactataGGCAAAGGGCTGTGCGACTTgttttagagctagaa

Previously described *in vitro* fertilization and microinjection protocols developed in the lab (del Viso and Khokha, 2012) and available online (http:// http://khokha.medicine.yale.edu/) were used to inject one cell stage embryos. For microinjections, borosilicate glass needles were calibrated. Using previously described methods, CRISPR sgRNAs were designed for *CFAP46* (Bhattacharya *et al.,* 2015). Three non-overlapping CRISPRs including *CFAP46-EX3, CFAP46-Ex7*, and *CFAP46-Ex8* sgRNAs were injected at 400 pg/embryo with 1.6 ng of case9 protein (CP03, PNA Bio) into single-celled embryos in accordance with conventional methods (Ran *et al.,* 2013; Sega *et al.,* 2019). After injections, embryos were placed in 1/9X MR + 3% Ficoll for 1–2 hr and then transferred to 1/9X MR supplemented with 50 μg/ml of gentamycin. Confirmation of successful injections was by fluorescent lineage tracing with a Zeiss Lumar fluorescence stereomicroscope at stage 28–30. Embryos/tadpoles were further incubated at 26 °C until stage 45–46.

Uninjected controls (UIC), *CFAP46-EX3, CFAP46-Ex7*, and *CFAP46-Ex8* injected tadpoles were then fixed in 4% Paraformaldehyde (PFA) and examined for craniofacial anomalies under a microscope. Photomicrographs were obtained using a Nikon SMZ745T stereomicroscope fitted with am Excelis Accu-Scope camera.

8.3.4: Tadpole Genotyping for analysis of CRISPR cut efficiency

Purpose:

This approach is used to determine the efficiency of CRISPR in *X.tropicalis* tadpoles. The entire process includes sample selection and organization, gDNA extraction and purification, designing primers for PCR amplification and sequencing, and analyzing and understanding Inference of CRISPR edit (ICE) output data. CRISPR cutting analysis should be done once a consistent phenotype (eg, craniofacial anomalies in the Khokha Lab) of 2-3 non-overlapping sgRNAs has been observed for the same gene.

- ***Optimize injection dose***

● Standard dose is 400pg of sgRNA and 1.6ng of Case9 protein per embryo (200pg/nl sgRNA,0.8ng/nl of Case9 protein)

● Amount of sgRNA can be increased to 600pg/emb (or higher?)

● Maximum density of embryos per 100mm dish: 60?

- ***DNA extraction and purification***

CRISPR efficiency was determined through Sanger sequencing. For this purpose, DNA was extracted from *CFAP46-EX3*, *CFAP46-Ex7*, and *CFAP46-Ex8* injected and UIC animals at stages 45-46 by using the Alkaline (NAOH) Lysis

method. A total of four different DNA samples were extracted from distinct embryos, with the first three samples containing DNA from a single tadpole and the fourth sample including a pool of DNA from five tadpoles. Monarch gDNA Purification kit was used to clean up the extracted DNA and the "Tissue" recommended protocol was adopted. Pre-warm elute buffer (EB) at 60º was used.

➢ ***Sequencing Preparation***

- **Primer designing**

Primer3 Plus was used to synthesize seven primer pairs (2 pairs for each CRISPR, *CFAP46-EX3* and *CFAP46-Ex8* while 3 pairs for *CFAP46-Ex7*), and regions surrounding the target sites were amplified using primers 200 base pairs (bp) upstream and 500 base pairs (bp) downstream of the CRISPR/Cas9 cut site. Primers were designed on the basis of the following properties:

- Pairs should have a similar melting temperature
- Need a minimum 1000 bp amplicon
- Need adequate sequence before and after target site
    - 300-500bp upstream of the predicted cut site
    - 500-800bp downstream of the predicted cut site
- **BLAST primers to *X. tropicalis*genome on Xenbase**

The designed primers underwent the BLAST search and were selected on the basis of following criteria

- should only have 1 hit at correct locus
- Avoid regions with repetitive sequences.

-Optimized primers with UIC gDNA

- **Primer designing parameters**

Max size	**Min Tm**	**Max TM**	**Min CG%**	**Max CG%**	**Max PolyX**	**CG Clamp**
22 bp	55° C	65° C	45	75	3	2

- **Primer sequence**

The sequence of primer pairs used for geneotyping are given below (Table 8.1). Different combination of primers for individual CRISPR were used in order to get the single clear band on gel for genotyping.

Table 8.1: Primer pairs used for genotyping for CRISPR cut efficiency

Primer pair No	**CRISPR name**	**Primer type**	**Sequence**
1	*CFAP46-EX3*	Forward-1	TGCACCTTTTAGCCCAATTT
		Reverse-1	CCTCAGCATACACAGGCAAA
2	*CFAP46-EX3*	Forward-2	GGGGGATCAGAAGCACTCTA
		Reverse-2	CAGCATACACAGGCAAATGG
1	*CFAP46-Ex7*	Forward-1	CCCTTTGTTAGCACCTGCAT
		Reverse-1	TTAGTAAATGGGCCCCACAG
2	*CFAP46-Ex7*	Forward-2	CAGTAAATCCCTTTGTTAGCACCT
		Reverse-2	AACCTTTTTGGCTGAGAGAGC
3		Forward-3	CCCTTTGTTAGCACCTGCAT
		Reverse-3	GGCCCCACAGTGTAAGCTAGT
1	*CFAP46-Ex8*	Forward-1	TCTCCCCCATTTCCCTTAAC
		Reverse-1	AATGTTAGGCATCCCCCAGT

2	*CFAP46-Ex8*	Forward-2	CCCCATTTCCCTTAACAAGT
		Reverse-2	GCCTTAACGCCTTAAACTCG

- ***PCR cleaning and purification***

PCR products were cleaned by using Monarch PCR DNA Cleanup kit and 15 ul of EB was used for each sample elution.

- ***Sanger Sequencing***

The forward and reverse primers used for PCR were also employed for sanger sequencing. Then set up samples and plasticware according to Quintara Biosciences Protocol.

8.3.5: Inference of CRISPR edit (ICE) analysis by Synthego

Inference of CRISPR edit (ICE) is a free online software that analyzes Sanger sequencing data to determine editing efficiency, %INDELs, %knockout, and the types of mutations generated by the sgRNA/Cas9 in mosaic F0 embryos. Ideally, there should be 40-95% knock out score in phenotypic embryos. ICE provides a simple quantitative evaluation of indels produced by CRISPR in pools and clones for free online. In ICE, amplicons made from genomic DNA extracted from the modified and control (wild type) samples are compared to one another. Potential proposed editing results are fitted to the observed data by linear regression. ICE's report outputs the overall indel percentage in addition to the knockout score (the proportion of outcomes that result in a supposed knockout). Successful ICE analysis displayed following parameters.

8.3.6: Interpretation of ICE analysis

- ***Summary window***

For analyses that included multiple samples, a summary window with bar graphs of results with a list of the analyzed samples appears. The window shows basic information about edited samples, including:

Sample: The unique sample name for each sample use in analysis.

Guide target: The 17-23 bp guide RNA, excluding the PAM sequence.

PAM sequence: Protospacer Adjacent Motif (PAM) sequence for the nuclease used with sgRNA.

Indel percentage: The percentage of sequences that contain an insertion or deletion (indel) in the sample.

Model Fit (R2): The R2 value (Pearson correlation coefficient) is a measure of how well the proposed indel distribution fits the Sanger sequence data of the edited sample. The maximum R2 value is 1.0 and the sum of all individual contributions will be equal to the R2 value. For example, if the R2 is 0.84, then all of the contributions will add up to 84%. The difference between 1.0 and the R2 (e.g., 100% - 84% = 16%) represents the percentage of Sanger sequencing data that is unexplained and does not match the expected outcomes. The R2 value is critical for assessing the indel %, KO Score The higher the R2 value, the higher the confidence in the indel percentage and KO Score. An R2 value of 0.8 or more indicates a robust analysis, but R2 values below 0.8 should be considered with caution.

Knockout (KO) score: The proportion of sequences that are likely to result in functional protein knockout (frameshift mutations or indels of 21+ bp). The higher the

KO Score, the higher the percentage of sequences that result in a knockout of the target gene.

- ***Contribution tab***

Top of the contribution tab analysis status, guide target, PAM sequence, indel %, Mode fit (R2) and knockout score and all are represented for a particular edited score. The contributions tab displays all of the indels and their associated frequencies (% contribution) of a particular sample. The guide sequence, indel %, and KO Score, and other information about the sample are also presented. If the unedited control (wild type) sequence is present at a high frequency in the sample, it will be demarcated by an orange "+" sign on the far left. However, if the wild type sequence is present at a very low frequency or absent, then it will not appear as a contribution.

8.3.7*: In situ* hybridization

X. tropicalis embryos were microinjected with 400 pg of human *CFAP46-EX3* CRISPR with Case9 by injecting one cell at the two-cell stage. Tadpoles were collected at stages 14,15,16,18,19 and fixed in MEMFA (1x MEMFA salts, 1:10 formaldehyde). For staining, Snail Family Transcriptional Repressor 2 (*Snai2*), SRY-Box Transcription Factor 9 (*Sox9*) and Twist-related protein 1 (*Twist1*) RNA probes were created from an *X. tropicali s*clone. *In situ* hybridization of the injected embryos with the neural crest markers including *Snai2, Sox9* and *Twist1* RNA probes were executed on the BioLane HTI 16Vx (Intavis Bioanalytical Instruments) using standard protocols (Khokha *et al.,* 2002).

8.4: Results

8.4.1: Clinical phenotyping

The Proband is a 27 year-old female from non-consanguineous parents (Fig. 8.4). She had a range of symptoms including intellectual disability, non-verbal with seizure disorder, cataract (one senile, one non-senile), cleft palate, Pierre-Robin sequence, mediastinal teratoma, hepatic lesions, cyst at the head of pancreas, renal cysts, scoliosis, contractures of hands, osteoporosis, short stature, hypothyroidism, hypotonia, "clumsy gait". There was an impression of normocephalic phenotype, hence no brain MRI was perfomed.

Dysmorphic features: narrow facies, prominent central incisors and dental crowding, hypotelorism. long nose, flat nasal bridge, prominent cupid's bow, small jaw, narrow palate.

8.4.2: CRISPR mediated Knock out in *X. tropicalis*

Through the Exome filtration and segregation analyses by sanger sequencing, it was hypothesized that the p.L605P and p.M319V compound heterozygous variant was leading to a loss of function of the CFAP46 protein (Fig. 8.5 and 8.6). To model this, *in vivo* effect of *CFAP46* knockout was determined using CRISPR/Cas9 in *X. tropicali s*tadpoles due to their diploid genome. Initially, three independent CRISPRs were used and observe the phenotype, next only single most efficient CRISPR was chosen for the rest of the experiment.

Pos.	Gene	Gene	Variant	Intoler ance	pLI	MAF GnomAD_All	CADD	SIFT	PPH	Effect score	ClinVar/ HGMD	OMIM	Dad	Mom
10:134732863_A>G	CFAP46	cilia and flagella associated protein 46	L605P	.	0	0.0003	25.3	D	.	4	.		Ref	Het
10:134743220_T>C	CFAP46		M319V	.	0	0.0003	0.002	T	B	0	.		Het	Ref

Fig 8.5: Screenshot of table showing compound heterozygous variant in *CFAP46.*

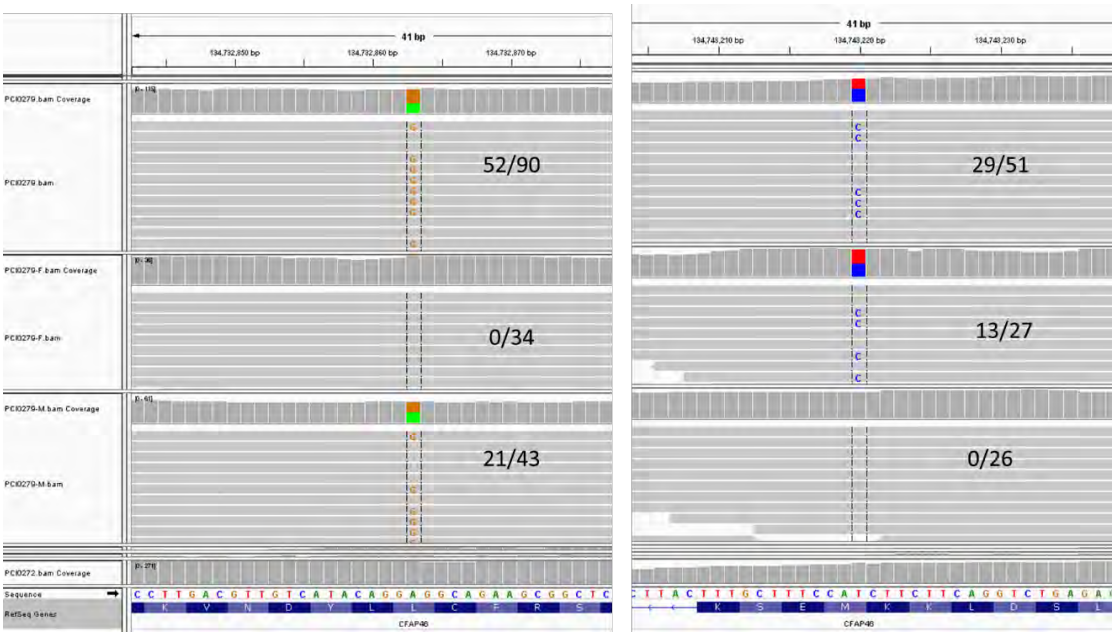


Fig 8.6: Screenshots of IGV depicting coverage of the detected variants.

8.4.3: Analysis of Sanger data to determine the CRISPR'S editing efficiency

Purpose: To identify the CRISPR with the best cut efficiency in order to choose just one CRISPR for the remaining experiments. For knockout experiments, just one CRISPR with the best cut efficiency is recommended.

8.4.4: Knock out analysis for *CFAP46-EX3*, (Reverse Primers)

For *CFAP46-EX3* two sets of primers were used for PCR reaction in order to get the single strong band. Detail of primers is given in Table 8.1. The amplified PCR product was puried by usining Monarch PCR DNA cleanup kit (5ug) and send for sanger sequencing. After getting the sanger data (.ab1 files) for *CFAP46-EX3* were run on Synthego ICE against reverse primer. The highest indel were observed as 82%, 79% and 70% while, lowesrt as 27% with knockout score of 44, 27, 78 and 66% respectively. In general, both the indel percentage as well as the knockout score was satisfactory for *CFAP46-EX3* (Fig. 8.7).

SAMPLE	GUIDE TARGET	PAM SEQUENCE	INDEL %	MODEL FIT (R^2)	KNOCKOUT-SCORE	
29_B07	GCACACAGAACGTACAAATC	AGG	70	0.84	66	View Details
AniEx3-R4_CFAP46-R_..2022-06-29_D07	GCACACAGAACGTACAAATC	AGG	79	0.92	78	View Details
AniEx3-R1_CFAP46-R_..2022-06-29_A07	GCACACAGAACGTACAAATC	AGG	27	0.96	27	View Details
AniEx3-R3_CFAP46-R_..2022-06-29_C07	GCACACAGAACGTACAAATC	AGG	82	0.93	44	View Details

Fig 8.7: ICE summary window for *CFAP46-EX3* (Reverse primer).

Further analysis shows that the contribution tab represents the inferred sequences in the edited population and their relative proportions, while t0 the in the Indel plot (Indek Distribution tab) does not specify sequence contributions. Cut sites are represented by black vertical dotted lines and wild type sequence is marked by " + " symbol on the far left (Fig. 8.8).

Fig 8.8: Contribution tab for *CFAP46-EX3* (Reverse Primer) showing indel percentage for knockout score of 78 and 66 repectively.

8.4.5: Synthego ICE analysis for *CFAP46-Ex7* (Forward Primer)

After receiving the sanger data (.ab1 files), the reverse primer for *CFAP46-Ex7* was run on ICE. In four separate samples, the indel percentage and knockout score were not comparable. Two out of four samples show the knockout scores of 75 and 49 with indels of 87% and 79% (Fig. 8.9).

Fig 8.9: ICE summary window for *CFAP46-Ex7* (Forward primer).

For *CFAP46-Ex7*, the contribution tabs show 87% indels with a KO score of 49 and 79% indels with a KO score of 75. The two samples have model fit (R2) values of 0.71 and 0.88, respectively. Only one sample with 38% indels is noteworthy because the Model Fit (R2) score shouldn't be lower than 0.88. Both contributions tabs additionally contain wild type sequence, which is shown by an orange "+" on the far left bar (Fig. 8.10).

Fig 8.10: Contribution tab for *CFAP46-Ex7* (Forward Primer) showing the indels percentage for knockout score of 75 and 49 repectively.

8.4.6: Synthego ICE Results for *CFAP46-Ex8* (Forward Primer)

Ice analysis for *CFAP46-Ex8* gives a successful analysis with forward primer. The indel percentage and KO score are very comparable for all four individual samples. The indel percentages were 55%, 79%, 95%, and 99%, respectively, with KO scores of 37, 58, 59, and 57 (fig. 8.11).

SAMPLE	GUIDE TARGET	PAM SEQUENCE	INDEL %	MODEL FIT (R^2)	KNOCKOUT-SCORE	
01_H10	GACAAAGGGCTGTGCGACTT	TGG	79	0.93	58	View Details
AniEx8-1_CFAP46-EX8-F__2022-07-01_E10	GACAAAGGGCTGTGCGACTT	TGG	95	0.95	59	View Details
AniEx8-2_CFAP46-EX8-F__2022-07-01_F10	GACAAAGGGCTGTGCGACTT	TGG	96	0.96	57	View Details
AniEx8-3_CFAP46-EX8-F__2022-07-01_G10	GACAAAGGGCTGTGCGACTT	TGG	55	0.95	37	View Details

Fig 8.11: ICE summary window for *CFAP46-Ex8* (Forward primer).

Contribution tabs for all four samples indicate Model Fit (R2) values of 0.95, 0.96, 0.95 and 0.93. All samples give more than the threshold values. This CRISPR was very efficient as declared by ICE analysis. But this was very lethal to the developing embryos and give a lot of death after few hours of injections therefore, *CFAP46-EX3* was preferably selected for the rest of the experiments (Fig. 8.12).

Fig 8.12: Contribution tab for *CFAP46-Ex8* (Forward Primer) showing indels for knockout score of 59 and 58 respectively.

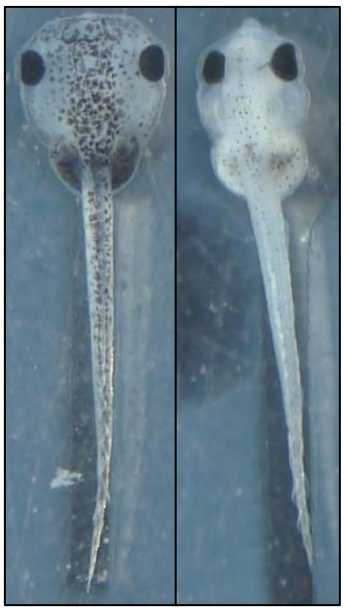
8.4.7: Phenotype with three non-overlapping CRISPRS (*CFAP46-EX3, CFAP46-Ex7, CFAP46-Ex8*)

For the current study it was hypothesized that the compound heterozygous variants p.L605P and p.M319V were loss of function of the CFAP46 protein. To model this, the *in vivo* effect of *CFAP46* knockout was initially studied using CRISPR/Cas9 in *X.tropicalis* due to its diploid genome. Uninjected controls displayed normal phenotypes during all developmental stages. In contrast, the *CFPA46*-Ex3,

CFAP46-Ex7 and *CFAP46-Ex8* CRISPR injected embryos give abnormal phenotypes at the gastrulation stage. At the gastrulation stage embryos give neural tube defects along the gastrulation defects in embryos. Gastrulation defects were highest in *CFAP46-Ex8* compared to the *CFAP46-Ex7* and *CFAP46-Ex7* (Fig.8.13).

Fig 8.13: Gastrulation and neural tube defects in *CFAP46 Knockouts.*

Both the uninjected control and CRISPR-injected embryos from the same clutch were allowed to grow till stages 45-46. Then under the microscope, they observed for craniofacial anomalies. There is a very clear demarcation between uninjected control embryos *vs* CRISPR injected. The CRISPR-injected embryos have difference in head area and interaocular distance (IOD) as compared to the uninjected control (Fig. 8.14).



UIC *CFAP*-46 Knockout

Fig 8.14: CRISPR/Cas9-mediated knockdown induces craniofacial anomalies in *X. tropicalis* tadpoles. Uninjected control (UIC) tadpole displays normal phenotype while kockout embryo showing narrow pointed face and decresased interaocular distance.

8.4.8: The measurement of head circumference for the evaluation of craniofacial anomalies using Fiji (an image processing program based on Image J2)

All three CRISPRs underwent three different experiments. Then, using Fiji by Image J, measure the head circumference at developmental stages 45–46. With the help of GraphPad Prism 9, the data were analyzed. In violion plots, the analyzed data were displayed, allowing us to compare how the data were distributed among the four groups (UIC, *CFAP46-EX3*, *CFAP46-Ex7*, *CFAP46-Ex8*; Fig. 8.15).

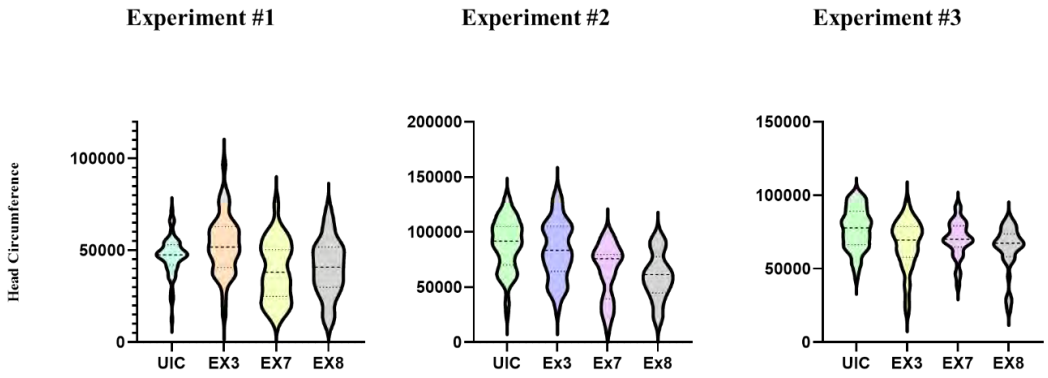


Fig 8.15: Comparison of head circumference among four data groups.

8.4.9: Measurement of interaocular distance for the assessment of craniofacial anomalies

The IOD by Fiji was measured on the same experimental group that has been examined for measurements of head circumference. The voilion plots represents the distribution of data amng four study grous (UIC *vs CFAP46-EX3*, *CFAP46-Ex7*, *CFAP46-Ex8*; Fig. 8.16).

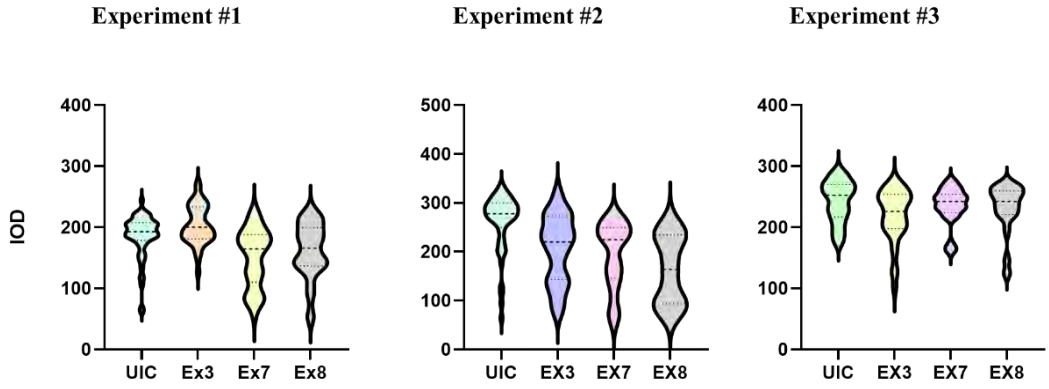


Fig 8.16: Comparison of IOD among four data groups.

8.4.10: Measurement of head circumference and IOD for the assessment of craniofacial anomalies (with single CRISPR; *CFAP46-EX3*)

After getting the CRISPR efficiency results from Synthego ICE. Only one CRISPR (*CFAP46-EX3*) was selected for the rest of the experiments. The voilion plots indicate the measurements of head circumferences and IOD of the UIC and *CFAP46-EX3* injected embryos.. There is a considerable difference in head circumference and IOD between UIC and the knockout embryos (Fig. 8.17).

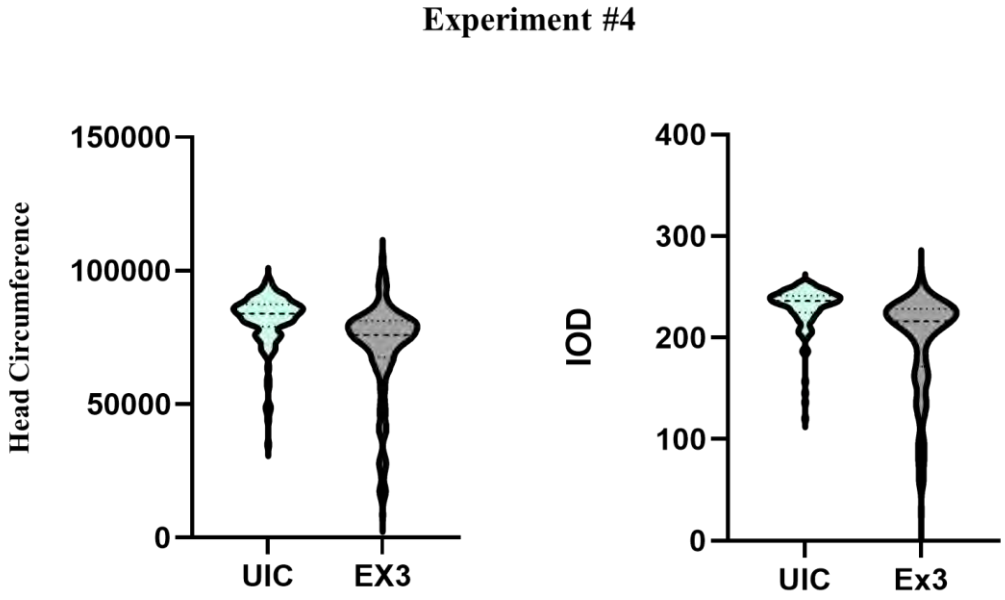


Fig 8.17: Measurement of head circumference and IOD.

8.4.11: Selection of neural crest markes (*Snai2*; *Sox9* and *Twist1*) as RNA probes for *in situ* Hybridization

The RNA seq data for various RNA probes are available from Xenbase, which gives us the ability to select a specific embryonic stage. *Snai2*, *Sox9*, and *Twist1* RNA probes were used in the current study as neural crest markers for the whole mount *in*

situ hybridization. Various *X.tripicalis* developmental stages were selected in order to examine the expression of these RNA probes. According to the RNA seq DATA *Snai2* expression start rising from stage 13 and 14, then highest at stage 22. At stage 15, 17, 19, 21 the *Snai2* also have the highest expression (Fig. 8.18). In the present study, stage 14, 16, 18 and 19 of *X.tropicalis* for three selected probes (*Snai2, Sox9* and *twist*1) were used . Both the uninjected control and CRISPR/Cas9 knockout were selected from the same clutch. For *Sox9* the RNA seq data indicates expression peaks at various stages similar as *Snai2* the *Sox9* start expression from stage 13 then rise from 14 till 15. Then there is fall in expression at stage 26. Then the second rise occurs at stages 29 and 30 (Fig. 8.19). RNA seq data for *Twist1* shows a continuous rise in expression from stage 11 to stage 27. There is a slightly lower peak for stages 30 to 32. Then again expression curves rise from stage 33 (Fig. 8.20).

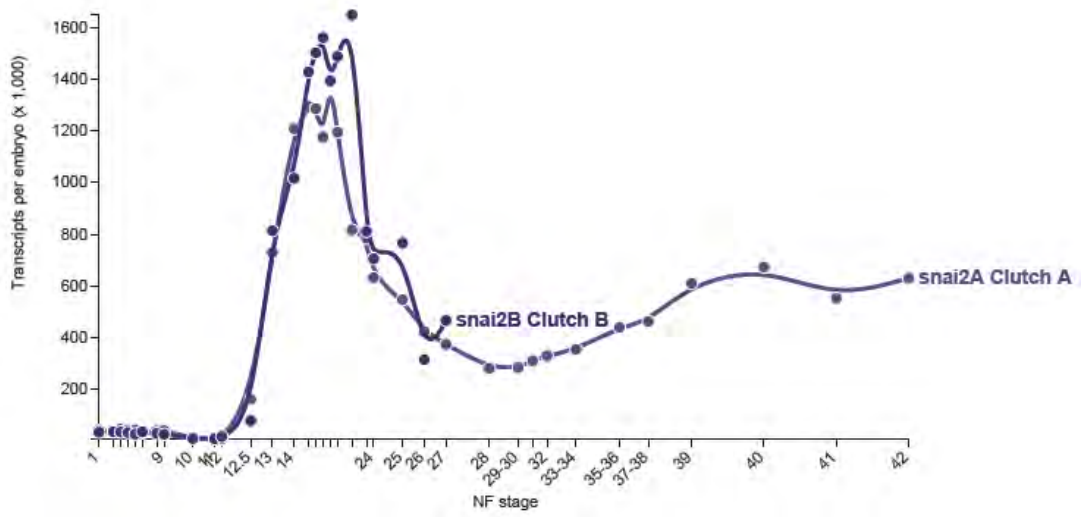


Fig 8.18: *X. tropicalis* RNA-Seq Data for *Snai2*.This image is generated by **www.xenbase.org**

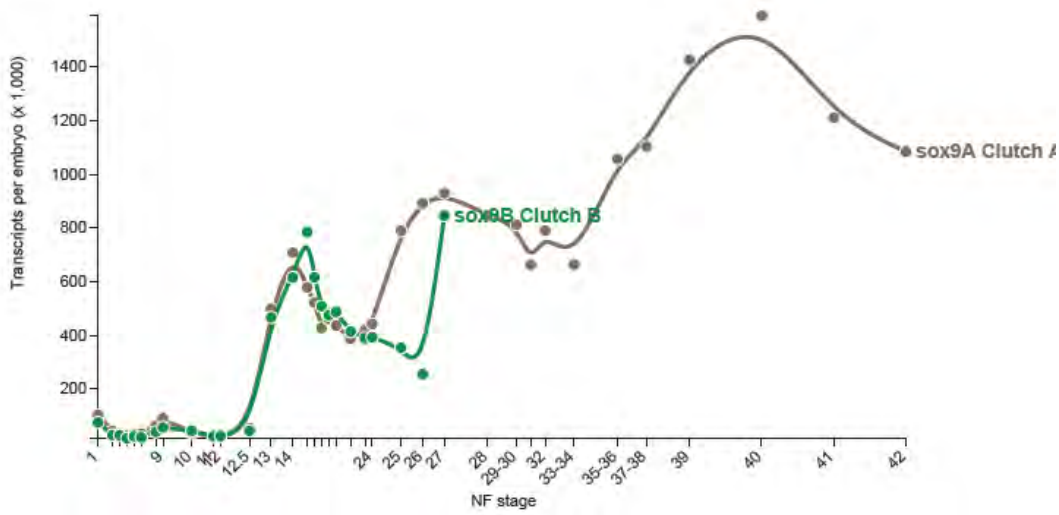


Fig 8.19: *X. tropicalis* RNA-Seq Data for *Sox9*. This image is generated by **www.xenbase.org**.

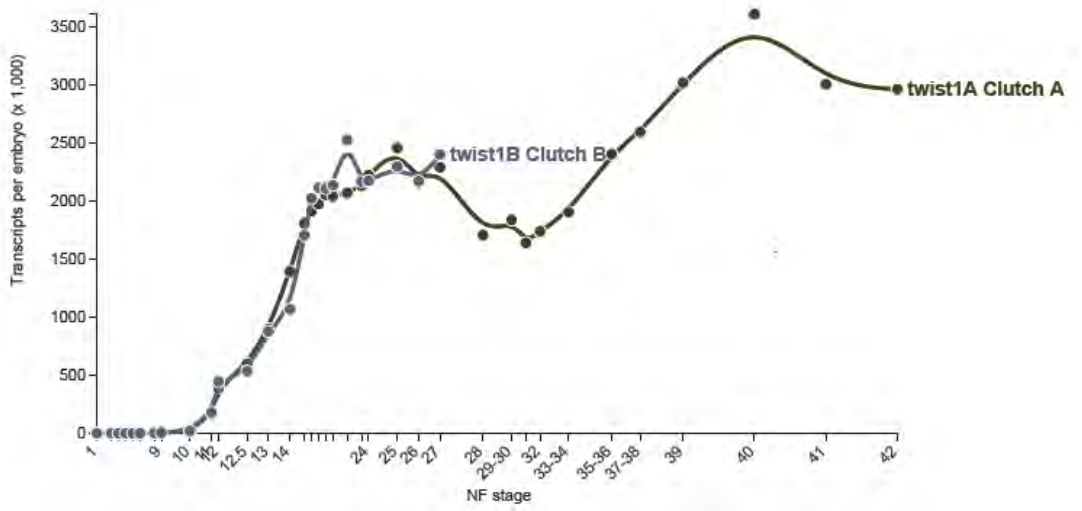


Fig 8.20: *X. tropicalis* RNA-Seq Data for *Twist1*. This image is generated by **www.xenbase.org**.

8.4.12: *In situ* hybridization for *Snai2, Sox9* and *Twist*

The embryos were fixed in MEMFA (1x MEMFA salts, 1:10 formaldehyde) at different developmental stages including 14,16,18, and 19 then underwent the whole

mount *in situ* hybridization with neural crest markers *Snai2*, *Sox9* and *Twist1*. After imaging the embryos, Fiji software was used to measure the probe developed area. GraphPad Prism 9 was used to evaluate the measured area.

8.4.13: *In situ* hybridization for *Snai2* at stages 14,16,18,19

The left graph represents the uninjected controls (UIC) embryos. Both the left and right sides of experimental embryos were measured independently with the left side injected with *CFAP46-EX3* CRISPR and right side taken as the internal, uninjected control side of experimental embryos. The right graph, on the other hand, shows the UIC and a right-side-injected embryo with a left-side internal control. There is a decline of *Snai2* in left side injected and right side injected embryos *vs* UIC and internal control. There is a prominent decline of *Snai2* in all stages including 14,16,18 and 19 (Fig. 8.21-Fig. 8.24).

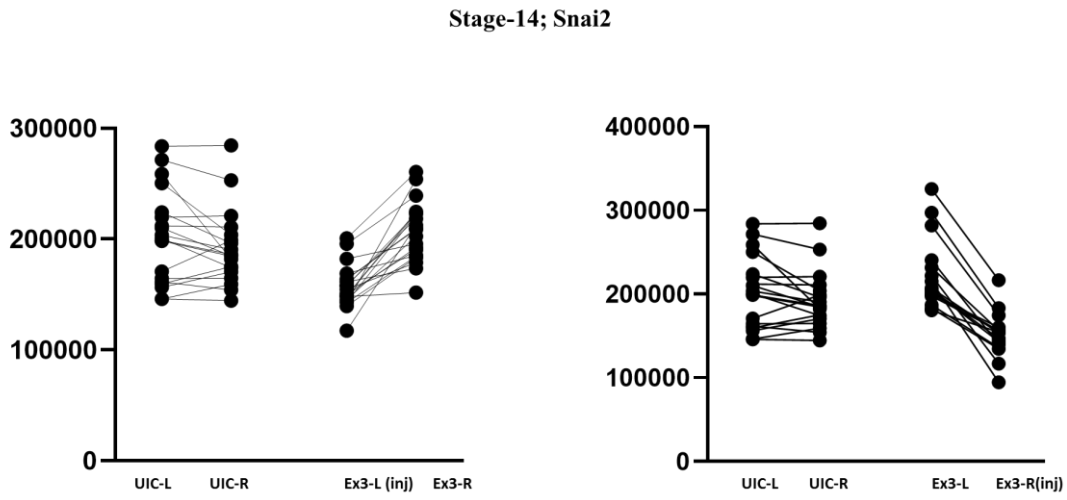


Fig 8.21: Graph showing the comparison of the Snai2 developed region at stage 14 between UIC and the left and right side of the injected embryos.

Abbreviations; Uninjected control left side (UIC-L), Uninjected control (UIC-R), *CFAP46-EX3* left side injected (Ex3-L(inj), *CFAP46-EX3* right uninjected (Ex3-R),

CFAP46-EX3 left side uninjected (Ex3-L), *CFAP46-EX3* right side injected (Ex3-R(inj).

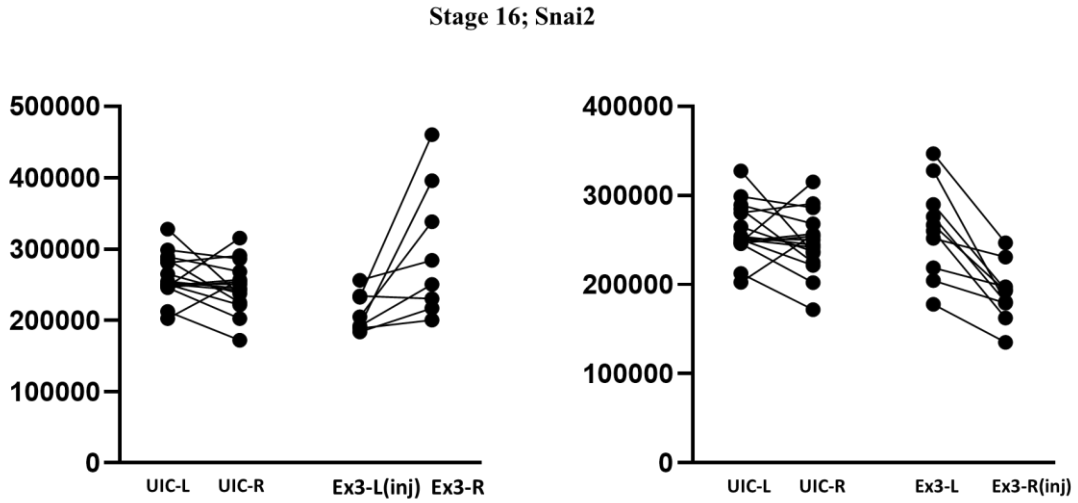


Fig 8.22: Graph showing the comparison of the *Snai2* developed region at stage 16 between UIC and the left and right side of the injected embryos.

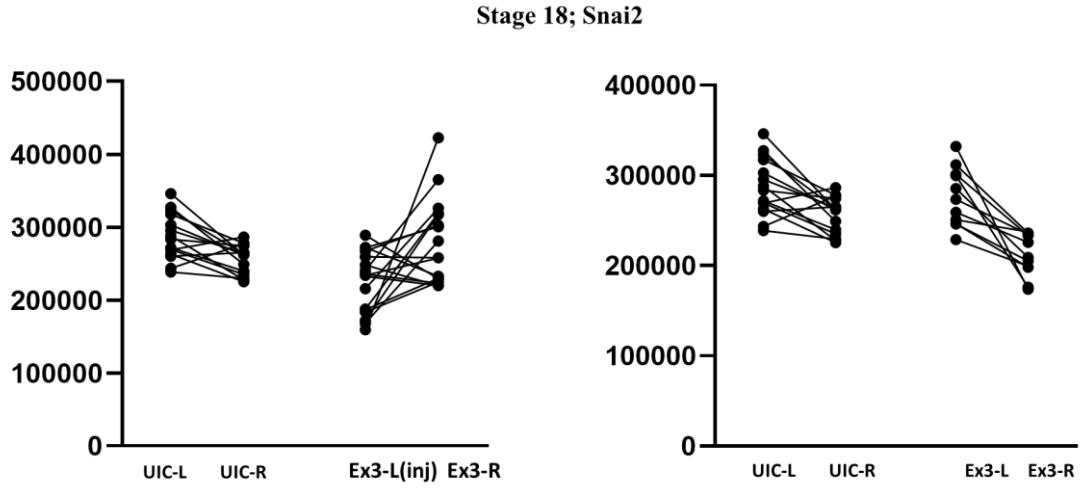


Fig 8.23: Graph showing the comparison of the *Snai2* developed region at stage 18 between UIC and the left and right side of the injected embryos.

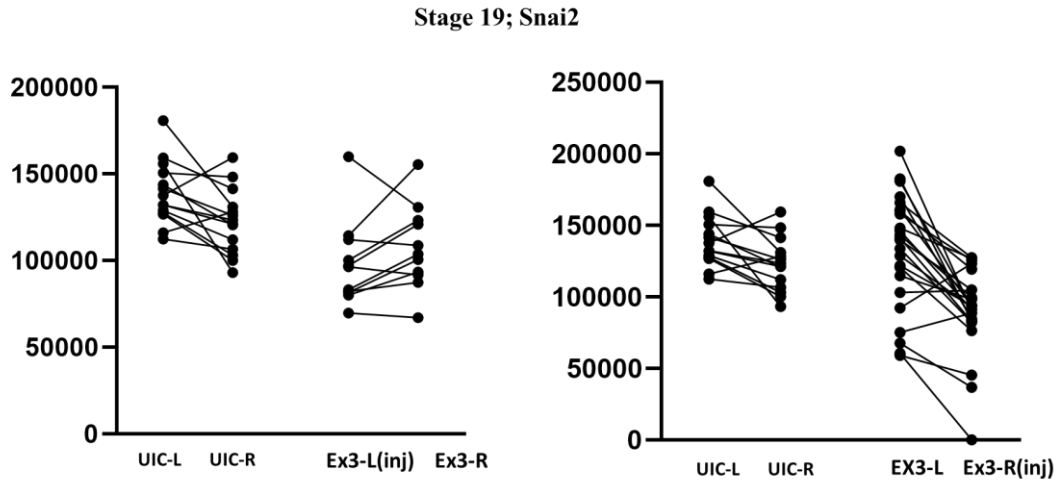


Fig 8.24: Graph showing the comparison of the *Snai2* developed region at stage 19 between UIC and the left and right side of the injected embryos**.**

8.4.14: Phenotype of *CFAP46-EX3* injected embryos on neural crest's *Snai2*, gene expression at stages 14, 16, 18, and 19

The function of *CFAP46* in neural crest and craniofacial development was evaluated through knockout of this gene by using CRISPR/Cas9. Consequences on neural crest gene expression at the different neural stages including stages 14,16,18 and 19, was also analysed. The CRISPR was designed to knockout the gene at exon 3. Upon injection of *CFAP46-EX3* CRISPR it was found that *CFAP46* was knockout in the morphant embryos. *CFAP46-EX3* knockout also resulted in a marked reduction of *Snai2* at all four stages 14,16,18, and 19 (Fig. 8.25). Along with its reduction, *Snai2*'s pattern in the left and right injected sides is altered at stage 18.

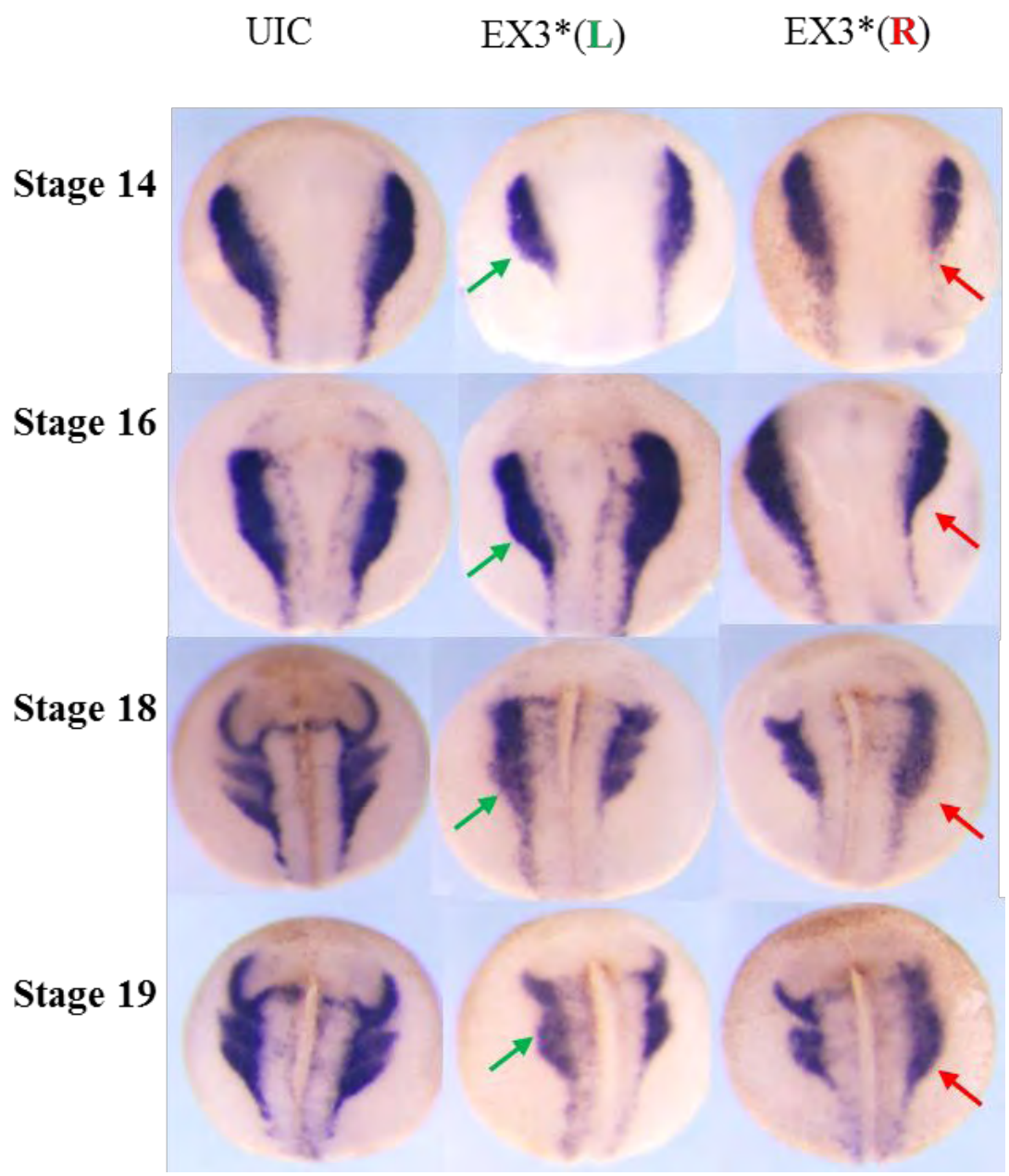


Fig 8.25: *CFAP46* knockout affects neural crest formation**.** There is reduction of *Snai2* expression at stage 14-16 while at stage 18-19 pattern of *Snai2* expression is also change along reduction in expression. UIC= Uninjected controls; Ex3*= Injected (Green; Left side: Red; Right side).

8.4.15: In situ hybridization for Sox9 at stages 14,16,18,19

The left graph represents the uninjected controls UIC, both the left and right sides of the embryos for the developed probe area were measured independently, and the left side of the embryos were injected with *CFAP46-EX3*. The right side of the injected embryo was taken as the internal control because the right side was not treated with *CFAP46-EX3*. The right graph, on the other hand, shows the UIC and a right-side-injected embryo with a left-side internal control. There is a decrease in *Sox9* in left side injected and right side injected embryos *vs* UIC and internal control (Fig. 8.26-Fig. 8.29). The pattern is variable in different developmental stages.

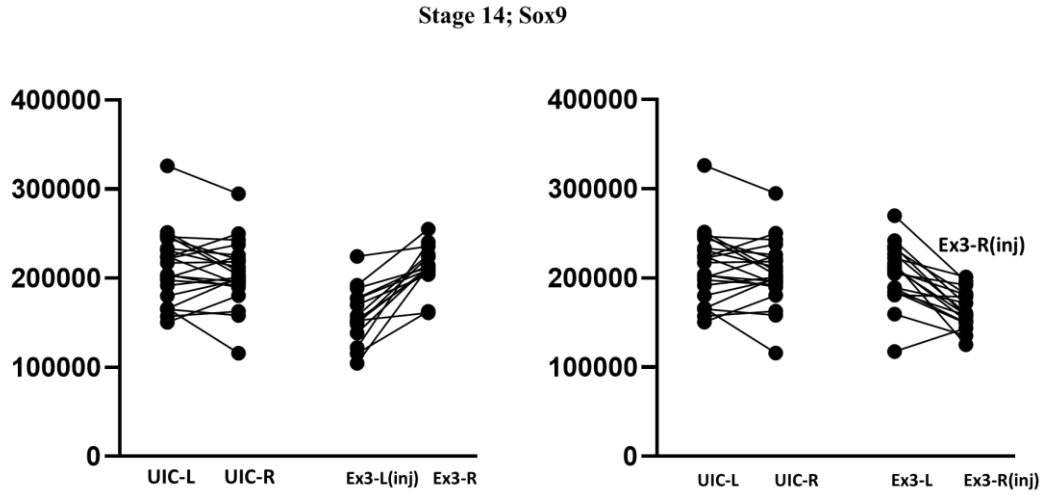


Fig 8.26: Graph showing the comparison of the *Sox9* developed region at stage 14 between UIC and the left and right side of the injected embryos.

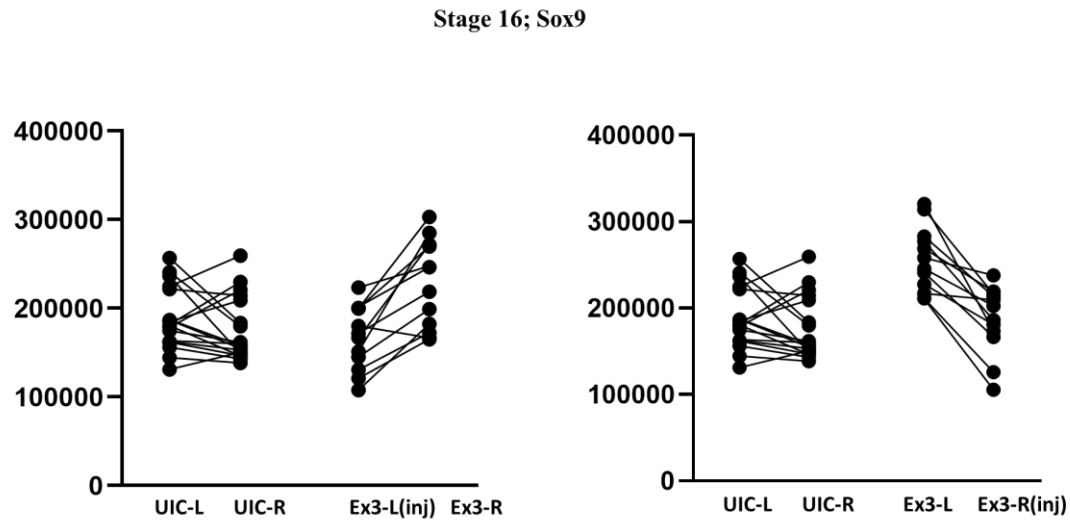


Fig 8.27: Graph showing the comparison of the *Snai2* developed region at stage 16 between UIC and the left and right side of the injected embryos.

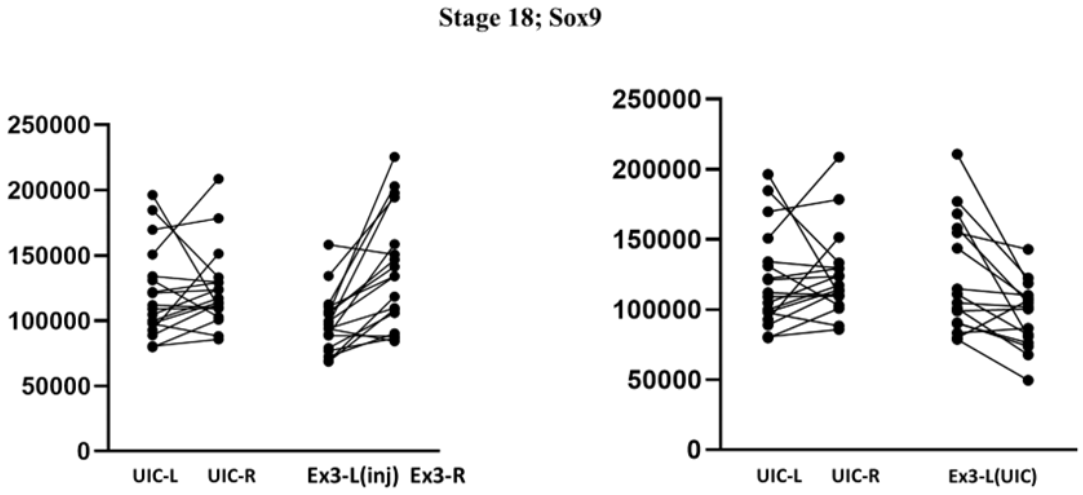


Fig 8.28: Graph showing the comparison of the *Sox9* developed region at stage 18 between UIC and the left and right side of the injected embryos.

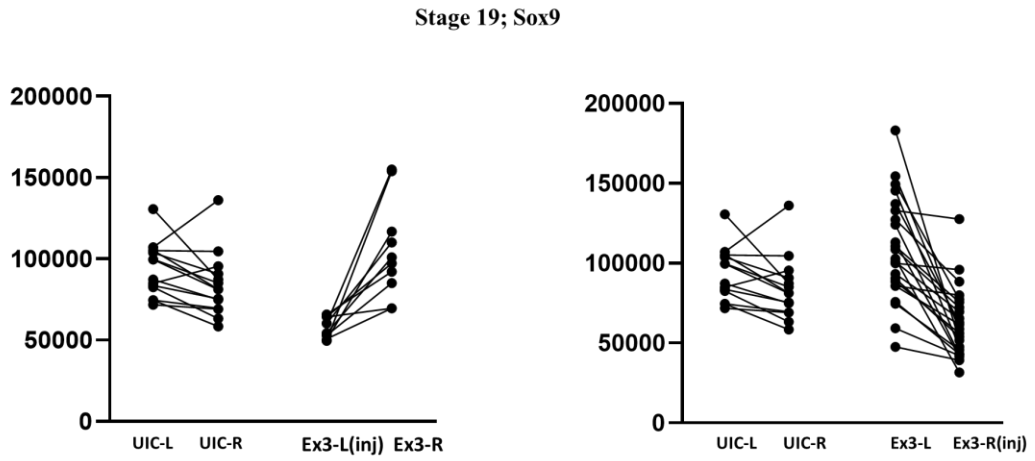


Fig 8.29: Graph showing the comparison of the *Sox9* developed region at stage 19 between UIC and the left and right side of the injected embryos.

8.4.16: Phenotype of *CFAP46-EX3* injected embryos on neural crest's *Sox9* gene expression at four stages including 14, 16, 18, and 19

Upon injection of *CFAP46-EX3* CRISPR it was found that *CFAP46* was knockout in the morphant embryos. *CFAP46-EX3* knockout also resuted in a marked reduction of *Sox9* at all four stages 14,16,18,19. However, there wasn't much of a difference between UIC and CFAP26 knockouts at stage 16 (Fig. 8.30).

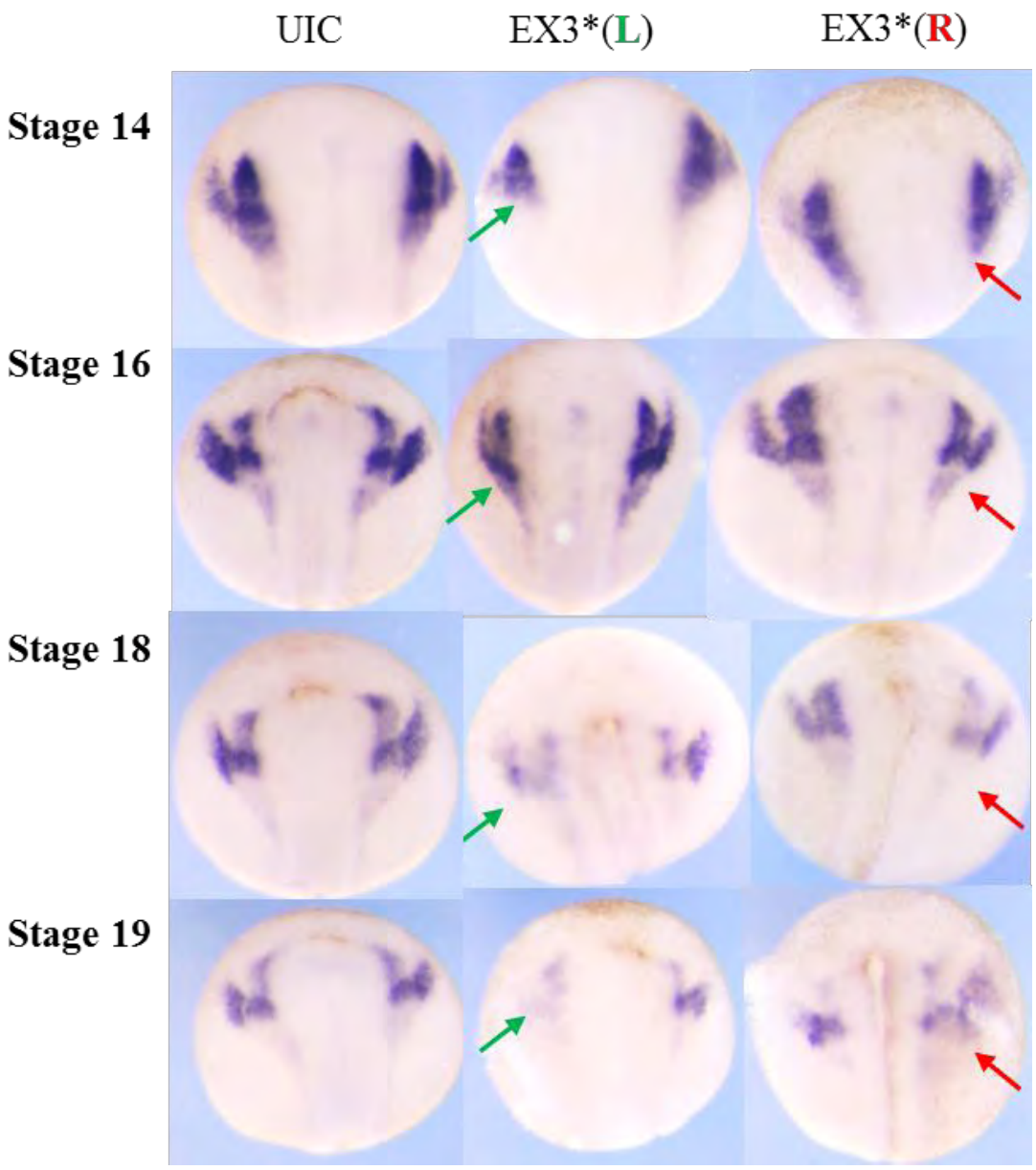


Fig 8.30: *CFAP46* knockout affects neural crest formation**.** There is a reduction of *Sox9* expression at stages 14,18 and 19 while at stages 18-19 pattern of *Snai2* expression is also changes along a reduction in expression. UIC= Uninjected controls; Ex3*= Injected (Green; Left side: Red; Right side).

8.4.17: *In situ* hybridization for *Twist1* at stage 14,16,18,19

There is a reduction in *Twist1* expression in left side injected and right side injected embryos *vs* UIC and internal control (Right side as internal control in left side injected

embryo and left side as the internal control in right side injected embryos. This reduction in *Twist1* expression is not prominent at stage 14 and 19 for left side injected embryos.While at stage 18 the right side injected embryos show no prominent decrease in *Twist1* expression (Fig. 8.31-Fig. 8.34).

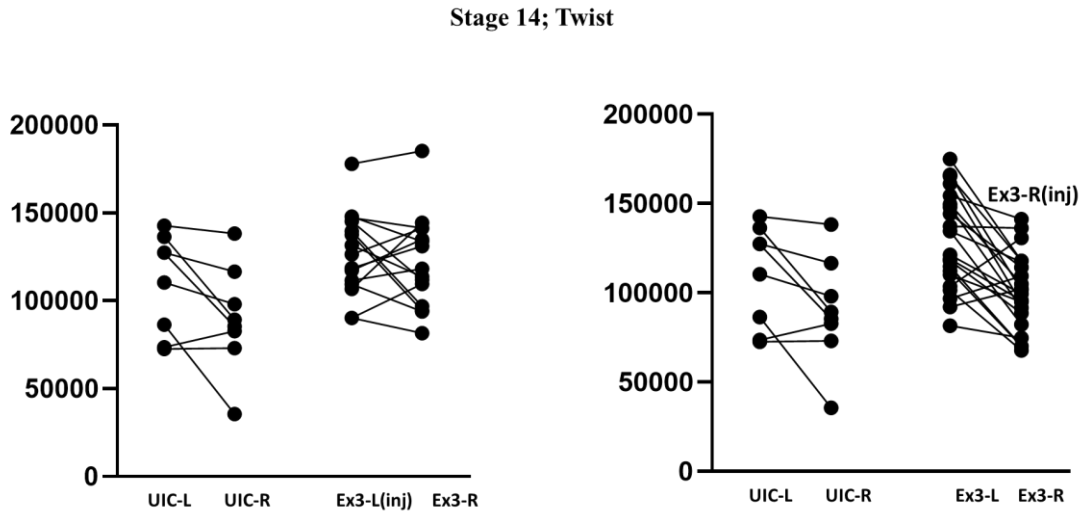


Fig 8.31: Graph showing the comparison of the *Twist1*developed region at stage 14 between UIC and the left and right side of the injected embryos.

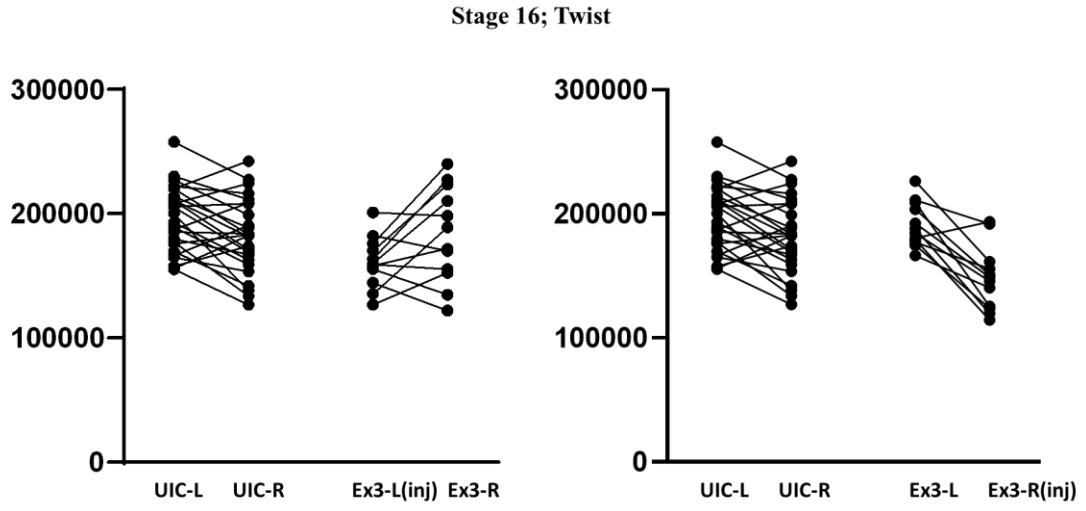


Fig 8.32: Graph showing the comparison of the *Twist1*developed region at stage 16 between UIC and the left and right side of the injected embryos.

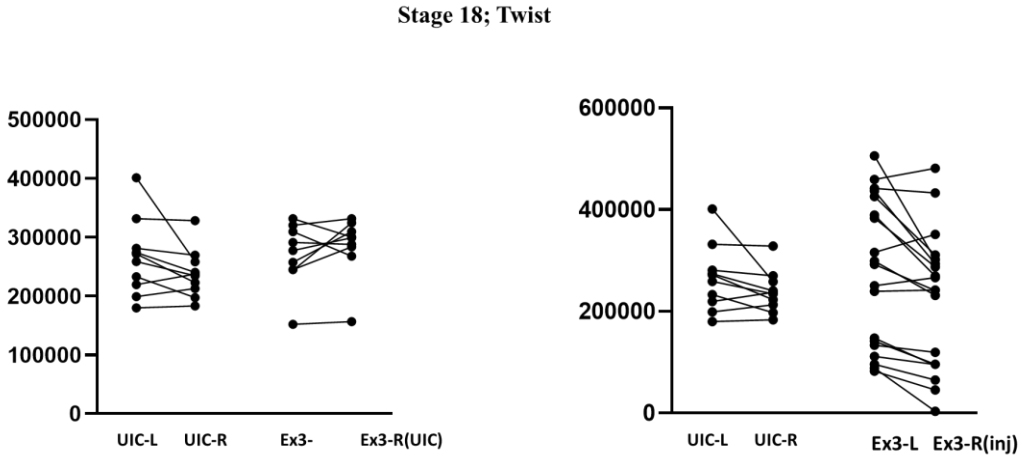


Fig 8.33: Graph showing the comparison of the *Twist1* developed region at stage 18 between UIC and the left and right side of the injected embryos.

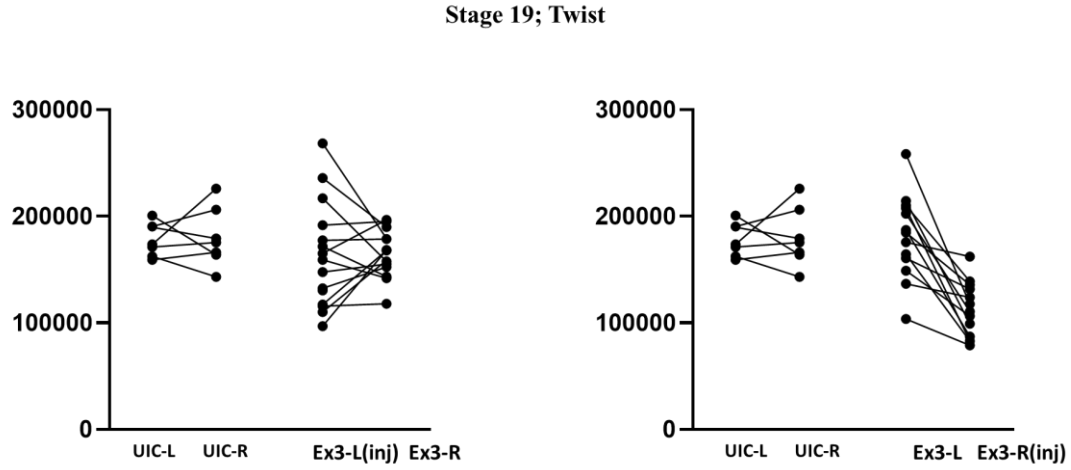


Fig 8.34: Graph showing the comparison of the *Twist1* developed region at stage 19 between UIC and the left and right side of the injected embryos.

8.4.18: Phenotype of *CFAP46-EX3* injected embryos on neural crest's *Twist*1 gene expression at four stages including 14, 16, 18, and 19

*Twist*1 at stages 14, 18, and 19 was considerably reduced as a result of the *CFAP46-EX3* knockout. There is no distinction between UIC and *CFAP46* knockout embryos at stage 16. The pattern of the *Twist1* expression also changed in stages 18 and 19 (Fig. 8.35).

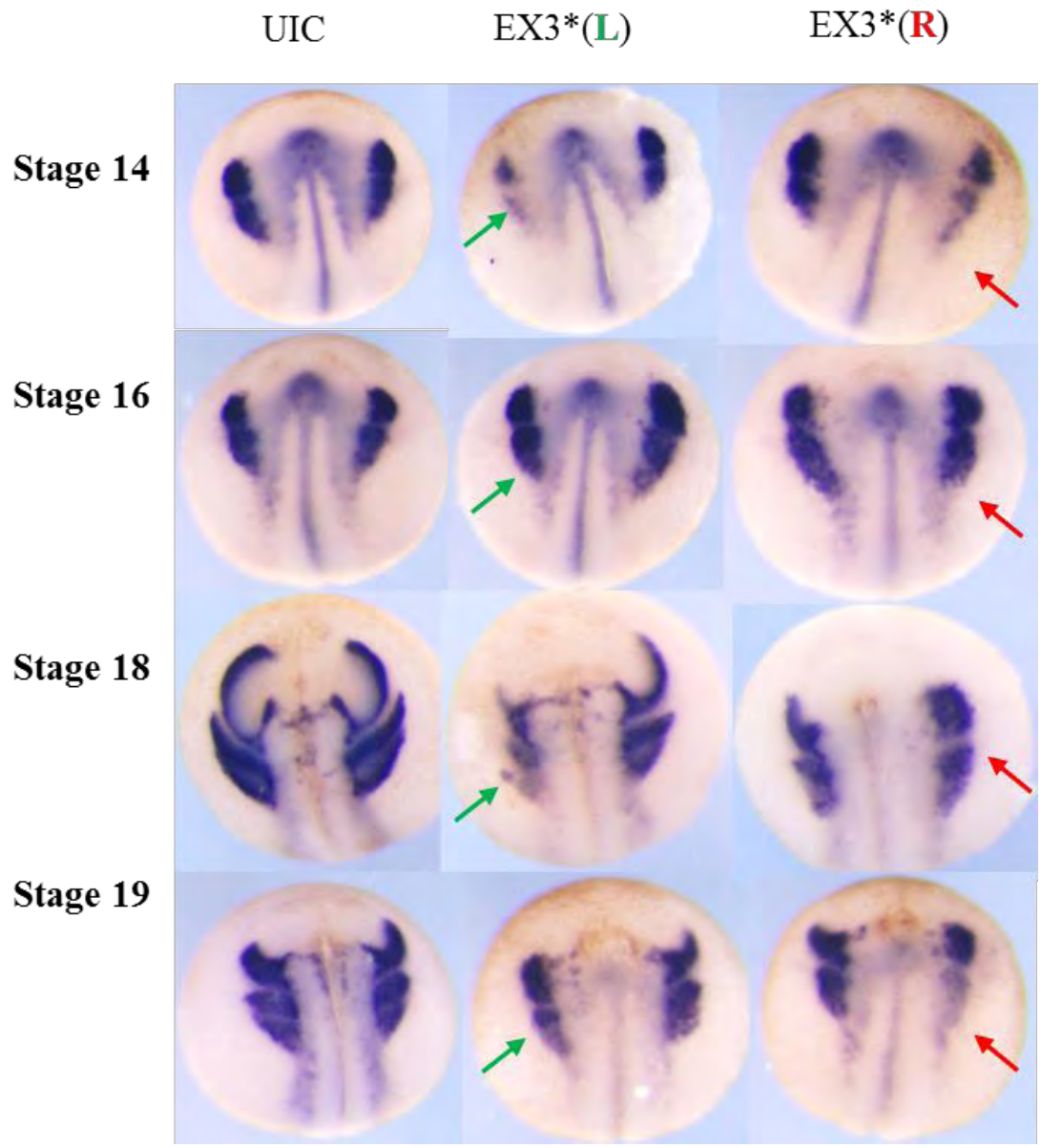


Fig 8.35: *CFAP46* knockout affects neural crest formation**.** There is reduction of *Twist*1 expression at stage 14,18 and 19, while at stage 18-19 pattern of

Twist1 expression is also change along a reduction in expression. UIC= Uninjected controls; Ex3*= Injected (Green; Left side: Red; Right side).

8.5: Discussion

This study presents a 20-year-old patient with different symptoms including intellectual disability, non-verbal with seizure disorder, cataracts (one senile, one non-senile), cleft palate, Pierre-Robin sequence, mediastinal teratoma, hepatic lesions, cyst at the head of the pancreas, renal cysts, scoliosis, contractures of hands, osteoporosis, short stature, hypothyroidism, hypotonia, "clumsy gait". He had narrow facies, prominent central incisors and dental crowding, and hypotelorism. Comprehensive molecular genetic analyses including exome sequencing, variant filtration, Sanger sequencing, and segregation analyses gave evidence that the malformation in the patient was likely caused by a compound heterozygous mutation in *CFAP46* and one frameshift mutation in the Echinoderm microtubule-associated protein-like 2 (EML2) gene.

Cilia and Flagella Associated Protein 46 (*CFAP46*) is also known as Tetratricopeptide Repeat Protein 40 (GeneCards). *CFAP46* localises at chr10q26.3 and has a size of 134,432 bases (ch10:132,808,392-132,942,823; GRCh38/hg38) (ncbi.nlm.nih.gov/gene/54777; Ayadi *et al.,* 2014). It comprises 59 exons and various transcripts are expected. The protein has a size of 2,715 amino acids with a molecular mass of 303500 Da (UniProtKB/Swiss-Prot). This gene is predicted to be located in the axoneme and is likely involved in axoneme assembly (Alliance of Genome Resources, 2022). As part of the central apparatus of the cilium, the axoneme plays a role in cilium movement (GeneCards Summary for *CFAP46*).

CFAP46 is a protein-coding gene and the phenotypes associated with this gene are Retinitis Pigmentosa 63 and Optic Atrophy 8 (GeneCards). As demonstrated in the GWAS Catalog, at least five phenotypes, i.e., visual perception measurement, BMI-adjusted waist circumference, sclerosing cholangitis, protein

measurement, and memory performance were witnessed to be strongly associated with *CFAP46* SNPs rs71481956, rs61865614, rs61861422, rs115217552, and rs76929979, respectively (GeneCards Summary).

The CFAP46 protein has been shown to localize to the cytoskeleton, cytosol, and nucleus (Confidence scores 4, 1, and 1, respectively). The STRING interaction network preview showed that it interacts with AGO2.

In order to observe the expression pattern of *CFAP46*, Ayadi *et al.,* (2014) carried out detailed studies in cancer cells and through the use of bisulfide genomic sequencing and methylation-specific PCR, witnessed that it was highly methylated in nasopharyngeal carcinoma samples compared to nontumor tissues. The authors employed RT-PCR analysis and further demonstrated that the extent of methylation was associated with transcriptional repression of the gene (Ayadi *et al.,* 2014).

The global expression studies for *CFAP46* have demonstrated that it has ubiquitous expression in a range of tissues ranging from immune, nervous, muscle, internal organs, secretory, and reproductive tissues (GeneCards). The highest expression was witnessed in the testis, stomach, thyroid, lung, kidney, and brain (HPA RNA-seq normal tissues). The study of orthologs for *CFAP46* showed that it had 96% and 78% similarity, respectively, with Chimpanzee (Pan troglodytes) and Dog (Canis familiaris) (genecards.org/cgi-bin/carddisp.pl?gene=*CFAP46*).

In this study, the detected compound heterozygous variants resulting in p.L605P and p.M319V were novel and not reported in any of the public genome variation databases, and likely underlie the phenotype in the patient. The segregation was concordant with the disease model**.**

8.6: Conclusion

In order to model the patient's craniofacial deformities and evaluate the functional impact of patient variants on candidate protein function compound heterozygous variant in *CFAP46* was chosen for the current investigation.

The reproduction of patients' craniofacial anomalies with CRISPR/Cas9 mediated knockout of *CFAP46* suggests that functional *CFAP46* is important to prevent neural tube and craniofacial defects. This hypothesis was further supported by current findings that the patient *CFAP46* variants give the craniofacial anomalies and neural tube defects. Thus it was concluded that the single nucleotide changes seen in the patient were detrimental to CFAP46 protein function.

General Discussion

Any morphological, biochemical, or functional abnormalities that manifest at birth are referred to as congenital anomalies (CA), sometimes called birth defects. Genetic, epigenetic, environmental, or other factors including consanguinity and maternal variables might cause these birth abnormalities.

The probability that both parents will have affected offspring and have the same recessive disease variations increases with consanguinity. When it comes to rare disorders, this effect is particularly noticeable because a spouse cannot share a disease unless they are related. Because of this, groups with high rates of consanguineous marriage also tend to have greater rates of rare recessive illnesses, congenital and genetic diseases, and other conditions. Consanguineous unions are very common in different regions of the world. In the Middle East and South Asian countries, almost 40% of all marriages are consanguineous (Abdalla and Zaher, 2013). While in Pakistan approximately 65% of marriages are commenced among blood relatives due to economic, social, and cultural reasons in different regions of Pakistan (Iqbal *et al.,* 2022). As a consequence, various types of severe autosomal recessive disorders (monogenetic) are common in the Pakistani population (Khan *et al.,* 2016).

At present, new interventions and better control reduce the burden of communicable diseases but meanwhile hereditary and genetic disorders appear as a major problem in healthcare systems. However genetic counseling, prenatal testing and awareness among families having history of recessive disorders can prevent transmission of genetic disorders.

The current research, which was described in the dissertation, was a multitiered study that made use of sampling at several levels as well as experimental and analytical techniques. Initially, data on individuals with genetic illnesses from Pakistan's Hazara population were gathered using an epidemiological approach. The prevalence-pattern of congenital/hereditary abnormalities was established for the Hazara division of Pakistan using first-hand data on 1,189 patients from independent subjects/families. Second, the prevalence and phenotypic pattern of 141 independent patients with congenital limb defects (CLD) were determined. These subjects were selected from the province of Khyber Pakhtunkhwa (KP) in Pakistan.

Further five families in which the patients had various presentations like intellectual disability, skeletal dysplasia, anophthalmia, scoliosis and blindness, and craniofacial anomalies, were selected for molecular study. Disorders in the families were diagnosed by medical specialists/radiologists working in government and private hospitals. Disease-causing genes in the families were searched by genotyping microsatellite markers, SNPs array, whole exome sequencing, and Sanger sequencing. Next, the family five underwent CRISPR/Cas9 mediated genome editing in order to model the patient's craniofacial anomalies and evaluate the functional impact of patient variants on candidate protein function.

Family 1 was a large family with multiple affected subjects exhibiting the symptoms of intellectual disability, SNP-based genotyping data were generated. Homozygosity mapping approach was adopted in order to detect regions of homozygosity shared among the affected subjects. Two of the affected individuals in this family were subjected to exome sequencing. Analyses of these data led to the shortlisting of rare variants that are pathologically relevant to the phenotype and also fall in the homozygous intervals detected in the SNP scan. Further in families 2, 3 and 4, mutation analyses through exome sequencing for candidate genes led to the identification of novel variants in already known genes responsible for causing skeletal deformities. These included *GRIP1* linked with Fraser syndrome 3 (FRASRS3), *FLNB* associated with Spondylocarpotarsal synostosis syndrome (SCT) and *DYM1* causing skeletal dysplasia. While in family 5, a novel gene *CFAP46* was found to be associated with craniofacial anomalies. The disease was modeled in vivo using CRISPR/Cas9-mediated genome editing in frog tadpoles, and the functional effects of the patient variant on candidate protein function were evaluated. The whole mount in situ hybridization technique was then adapted to explore the role of the CFAP46 protein in embryonic development. The patient had a number of syndromic features, including severe craniofacial anomalies. The craniofacial anomalies induced by the CRISPR/Cas9-mediated genome editing of *CFAP46* mimicked the patient's condition. Thus it was concluded that the single nucleotide changes seen in the patient were detrimental to CFAP46 protein function.

Next-generation sequencing (NGS) allows the discovery of recessive disease genes in even a single affected person. In addition, both the number and size of the affected families required to locate a causal gene have decreased. Each NGS technology, including whole exome sequencing (WES), whole genome sequencing (WGS) and panel sequencing, has its advantages. The majority of genes have been discovered through NGS using WES, while WGS is better at spotting copy number variations, chromosomal rearrangements, and repeat-rich regions. Panels, on the other hand, produce manageable amounts of data with no chance of unexpected results and are incredibly cost-effective, so they are frequently utilized for diagnostic purposes. However, when there is diagnostic uncertainty, selecting the right panel can be quite difficult. Therefore, under these conditions, WES and WGS can be considered as the most comprehensive second-tier test.

WES is less efficient in sequencing exome regions with high GC content. Moreover, WES was unable to cover the mutation discovered in the non-coding region. Consequently, the purpose of WES is not fulfilled; on the other hand, the novel PCR-free WGS offers coverage of the exome and other genomic sequences that are clinically meaningful. Because of its wider exome coverage, WGS is therefore not only superior to WES in detecting non-coding pathogenic variation, but it is the superior WES overall. With the further decrease in sequencing costs and the use of suitable virtual panels, WGS might eventually completely replace WES and other methods involving the targeted collection of sequences.

The genetic basis of many disorders is now well understood because of recent developments in NGS methods. Studies employing animal models, such as Xenopus, have considerably added to this knowledge at the cellular and molecular levels (Dubey and Saint-Jeannet, 2017). Frog embryos have long been used as a model organism for studying the early developmental disorders. The advent of transgenesis, Xenopus tropicalis and Xenopus Leavis genome sequencing, and knockdown in the last ten years have all contributed to the model's growing importance for studying early development (Pegoraro and Monsoro-Burq, 2012).

Genome editing (GE) alters a particular DNA segment by base substitutions, insertions and/or deletions (indels) in the target sequences (Aglawe *et al.,* 2018). Direct targeting and modification of the genomic sequences across all eukaryotic cells are now possible; thanks to genetically engineered or bacterial nucleases. GE increased our ability to understand the genetic process underlying a specific disease by accelerating the development of more accurate models of disease conditions (Manghwar *et al.,* 2019). Most recently, CRISPR/Cas9 system (Li *et al.,* 2020) is proving game-changing in editing the genome. Due to CRISPR/Cas9 system, the area of gene therapy is now rapidly growing and becoming more adaptable for the treatment of human genetic disorders. CRISPR/Cas9 genome editing is uses the CRISPR/Cas9 system that bacteria or archaea employ to cut and destroy invading DNA from sources including viruses (Makarova *et al.,* 2020). CRISPR/Cas9-based gene therapy can be utilized to delete, replace, or fix abnormal, disease-causing genes.

Prevalent genetic abnormalities place a huge socioeconomic load on families that impacts their social security and places a financial strain, not just on families and communities but on the whole country where healthcare systems are struggling to cope. Poverty and the low likelihood that genetic diseases can be treated to full, or even partial recovery, has life-long impacts where disability has negative effects on people's lives, preventing them from taking advantage of possibilities including basic healthcare, education, and social advancement, which would otherwise be available to them. Therefore public awareness should be initiated about close unions and its consequences in the form of hereditary disorders/defects.

In summary, a multitiered strategy was used to elucidate the clinical, genetic, and functional aspects of individuals and families with rare hereditary diseases. The discovery of new genes and pathogenic variations causing a variety of genetic illnesses promotes the development of genotypic-phenotypic associations and makes it easier to apply various genetic tests (like carrier screening). The functional genomic analysis of this study was limited to a single family; further investigation of the remaining mutations at the cellular level is necessary, in addition to examining the pathways in which a given gene functions. That will increase our comprehension and knowledge of hereditary disorders.

Future directions

High prevalence of genetic disorders place a financial strain on individuals, families, and the national economy as a whole. Therefore, to manage the rate of such diseases, strategies and policies are desperately required.

The following actions may be ordered by the health ministry or health department:

- To prevent the future spread of the defective gene pool, every citizen should receive screening and diagnoses during parental selection at government hospitals from expert geneticists.
- International partnerships are necessary in order to build next generation sequencing (NGS) technologies in the nation and undertake research at the highest levels.
- Further support the utility of *X. tropicalis* and similar animal system in the modelling of human genetic disorders.

References

Abdalla B, Zaher A (2013). Consanguineous marriages in the middle east: nature versus nurture. Open Complementary Med J. 5(1):1-11.

Abdullah, Shah PW, Nawaz S, Hussain S, Ullah A, Basit S, Ahmad W (2020). A homozygous nonsense variant in DYM underlies Dyggve-Melchior-Clausen syndrome associated with ectodermal features. *Mol Biol Rep*. 47(9):7083-7088.

Abebe, S., Gebru, G., Amenu, D., Mekonnen, Z., & Dube, L. (2021). Risk factors associated with congenital anomalies among newborns in southwestern Ethiopia: A case-control study. *PloS one*. 16(1).

Aglan MS, Temtamy SA, Fateen E, Ashour AM, Eldeeb K, Hosny GA (2009). Dyggve–Melchior–Clausen syndrome: clinical, genetic, and radiological study of 15 Egyptian patients from nine unrelated families. *J Pediatr Orthop*. 3:451–8.

Aglawe SB, Barbadikar KM, Mangrauthia SK, Madhav MS (2018). New breeding technique "genome editing" for crop improvement: applications, potentials and challenges. *3 Biotech*. 8(8):336.

Ahmad I, Rehman AU, Malik S (2016). Determinants of consanguinity and inbreeding coefficient f in Dir Lower district, north-west Pakistan: a multivariate approach. *Iran J Public Health.* 45(4):537-539.

Ajao, A. E., & Adeoye, I. A. (2019). Prevalence, risk factors and outcome of congenital anomalies among neonatal admissions in OGBOMOSO, Nigeria. *BMC pediatrics*. 19(1):88.

Akbar N, Ahmad H, Nadeem MS, Hemphill BE, Muhammad K, Ahmad W, Ilyas M (2016). HVSI polymorphism indicates multiple origins of mtDNA in the Hazarewal population of Northern Pakistan. *Genet Mol Res*. 24;15(2).

Ali G (2010). Genetic deafness in Pakistani population. *J Pak Med Assoc*. 60(6):418-9

Aloui M, Nasri K, Ben Jemaa N, Ben Hamida AM, Masmoudi A, Gaïgi SS, Marrakchi R. (2017) Congenital anomalies in Tunisia: Frequency and risk factors. *J Gynecol Obstet Hum Reprod*. 46(8):651-655.

Alshammari MJ, Al-Otaibi L, Alkuraya FS (2012). Mutation in RAB33B, which encodes a regulator of retrograde Golgi transport, defines a second Dyggve–Melchior–Clausen locus. *J. Med* Genet. 49, 455–461.

Al-Worikat AF, Dameh W (2008). Children with limb deficiencies: demographic characteristics. *Prosthet Orthot Int*. 32(1):23-8.

Ameen, S. K., Alalaf, S. K., & Shabila, N. P. (2018). Pattern of congenital anomalies at birth and their correlations with maternal characteristics in the maternity teaching hospital, Erbil city, Iraq. *BMC pregnancy and childbirth.* 18(1): 501.

Arfaksad A, Wajahat Y (2016). Frequency of fetal congenital anomalies and associated risk factors observed in pregnant women in a public sector hospital of Karachi. *Ann Abbasi Shaheed Hospital & Karachi Med Dent College*. 21(3):139-146.

Aubart M, Gazal S, Arnaud P, Benarroch L, Gross MS, Buratti J, Boland A, Meyer V, Zouali H, Hanna N, Milleron O, Stheneur C, Bourgeron T, Desguerre I, Jacob MP, Gouya L, Génin E, Deleuze JF, Jondeau G, Boileau C (2018).

Association of modifiers and other genetic factors explain Marfan syndrome clinical variability. *Eur J Hum Genet*. 26(12):1759-1772.

Ayadi W, Allaya N, Frikha H, Trigui E, Khabir A, Ghorbel A, Daoud J, Frikha M, Gargouri A, Mokdad-Gargouri R (2014). Identification of a novel methylated gene in nasopharyngeal carcinoma: TTC40. *Biomed Res Int*. 2014:691742.

Azhar MM (2011). Helping the disabled to become respectable members of the society. *Pak J Med Health Sci*. 5:720–3.

Baird PA, Anderson TW, Newcombe HB, Lowry RB (1988). Genetic disorders in children and young adults: a population study. *Am J Hum Genet*. 42(5):677-93.

Barbosa-Buck CO, Orioli IM, da Graça Dutra M, Lopez-Camelo J, CAtilla EE, Cavalcanti DP (2012). Clinical epidemiology of skeletal dysplasias in South America. *Am J Med Genet A*. 158A(5):1038-45.

Barisic I, Odak L, Loane M, Garne E, Wellesley D, Calzolari E, Dolk H, Addor MC, Arriola L, Bergman J, Bianca S, Boyd PA, Draper ES, Gatt M, Haeusler M, Khoshnood B, Latos-Bielenska A, McDonnell B, Pierini A, Rankin J, Rissmann A, Queisser-Luft A, Verellen-Dumoulin C, Stone D, Tenconi R (2013). Fraser syndrome: epidemiological study in a European population. *Am J Med Genet A*. 161A(5):102-8.

Baudier J, Jenkins ZA, Robertson SP (2018). The filamin-B-refilin axis - spatiotemporal regulators of the actin-cytoskeleton in development and disease. *J Cell Sci.* 131(8) jcs213959.

Bhattacharya D, Marfo CA, Li D, Lane M, Khokha MK (2015). CRISPR/CA9: An inexpensive, efficient loss of function tool to screen human disease genes in Xenopus. *Dev Biol*. 408(2):196-204.

Bhatti NA, Mumtaz S, Malik S (2019). Epidemiological study of congenital and hereditary anomalies in Sialkot District of Pakistan revealed a high incidence of limb and neurological disorders. *Asian Biomed*. 13(2):49-60.

Biesecker LG, Burke W, Kohane I, Plon SE, Zimmern R (2012). Next-generation sequencing in the clinic: are we ready? *Nature Reviews Genetics*. 13(11):818-824.

Bittles AH (2010). Consanguinity, genetic drift, and genetic diseases in populations with reduced numbers of founders. In Speicher MR, Antonarakis SE, Motulsky AG. (eds) Vogel and Motulsky's Human Genetics. *Berlin Heidelberg, Springer-Verlag*. 507-728.

Blue, N. R., Page, J. M., & Silver, R. M. (2019). Genetic abnormalities and pregnancy loss. *Seminars in perinatology*. 43(2):66–73.

Boonstra AM, Rijnders LJ, Groothoff JW, Eisma WH (2000). Children with congenital deficiencies or acquired amputations of the lower limbs: functional aspects. *Prosthet Orthot Int*. 24(1):19-27.

Boskovski MT, Yuan S, Pedersen NB, Goth CK, Makova S, Clausen H, Brueckner M, Khokha MK (2013). The heterotaxy gene GALNT11 glycosylates Notch to orchestrate cilia type and laterality. *Nature*. 504(7480):456-9.

Boycott KM, Rath A, Chong JX, Hartley T, Alkuraya FS, Baynam G, Brookes AJ, Brudno M, Carracedo A, den Dunnen JT, Dyke SOM, Estivill X, Goldblatt J, Gonthier C, Groft SC, Gut I, Hamosh A, Hieter P, Höhn S, Hurles ME,

Kaufmann P, Knoppers BM, Krischer JP, Macek M Jr, Matthijs G, Olry A, Parker S, Paschall J, Philippakis AA, Rehm HL, Robinson PN, Sham PC, Stefanov R, Taruscio D, Unni D, Vanstone MR, Zhang F, Brunner H, Bamshad MJ, Lochmüller H (2017). International Cooperation to Enable the Diagnosis of All Rare Genetic Diseases. *Am J Hum Genet*. 100(5):695-705.

Boyd PA, Keeling JW, Lindenbaum RH (1988). Fraser syndrome (cryptophthalmos-syndactyly syndrome): a review of eleven CA with postmortem findings. *Am J Med Genet*. 31(1):159-68.

Bronner ME, LeDouarin NM (2012). Development and evolution of the neural crest: an overview. *Dev Biol*. 366(1):2-9.

Brookes AJ (1999). The essence of SNPs. *Gene* 234:177–86.

Browning SR, Browning BL (2010). High-resolution detection of identity by descent in unrelated individuals. *Am J Hum Genet*. 86(4):526-539.

Carapito R, Goldenberg A, Paul N, Pichot A, David A, Hamel A, Dumant-Forest C, Leroux J, Ory B, Isidor B, Bahram S (2016). Protein-altering MYH3 variants are associated with a spectrum of phenotypes extending to spondylocarpotarsal synostosis syndrome. *Eur J Hum Genet*. 24(12):1746-1751.

Carney TJ, Feitosa NM, Sonntag C, Slanchev K, Kluger J, Kiyozumi D, Gebauer JM, Coffin Talbot J, Kimmel CB, Sekiguchi K, Wagener R, Schwarz H, Ingham PW, Hammerschmidt M (2010). Genetic analysis of fin development in zebrafish identifies furin and hemicentin1 as potential novel fraser syndrome disease genes. *PLoS Genet*. 6(4):1000907.

Chalya PL, Mabula JB, Dass, R.M., Ngayomela, I.H., Chandika, A.B., Mbelenge, N. and Gilyoma, J.M (2012). Major limb amputations: A tertiary hospital experience in northwestern Tanzania. *J Orthopaedic Surg Res*. *7*(1):1-6.

Chiang AP, Beck JS, Yen, H.J., Tayeh, M.K., Scheetz, T.E., Swiderski, R.E., Nishimura, D.Y., Braun, T.A., Kim, K.Y.A., Huang, J. and Elbedour, K (2006). Homozygosity mapping with SNP arrays identifies TRIM32, an E3 ubiquitin ligase, as a Bardet–Biedl syndrome gene (BBS11). *Proc Nat Acad Sci*. 103(16):6287-6292.

Claussnitzer M, Cho JH, Collins R, Cox NJ, Dermitzakis ET, Hurles ME, Kathiresan S, Kenny EE, Lindgren CM, MacArthur DG, North KN, Plon SE, Rehm HL, Risch N, Rotimi CN, Shendure J, Soranzo N, McCarthy MI (2020). A brief history of human disease genetics. *Nature.* 577(7789):179-189.

Cohn DH, Ehtesham N, Krakow D, Unger S, Shanske A, Reinker K, Powell BR, Rimoin DL (2003). Mental retardation and abnormal skeletal development (Dyggve-Melchior-Clausen dysplasia) due to mutations in a novel, evolutionarily conserved gene. *Am J Hum Genet*. 72(2):419-28.

Comstock JM, Putnam AR, Opitz JM, Pysher TJ, Szakacs J (2005). Prenatal death in Fraser syndrome. *Fetal Pediatr Pathol*. 24(4-5):223-38

Cornu TI, Mussolino C, Cathomen T (2017). Refining strategies to translate genome editing to the clinic. *Nat Med*. 23(4):415-423.

Cosme HW, Lima LS, Barbosa LG (2017). Prevalence of congenital anomalies and their associated factors in newborns in the city of SAO PAULO from 2010-2014. *Rev Paul Pediatr*. 35(1):33-38.

Costa C MdS, Gama SGNd, Leal MdC (2006). Congenital malformations in Rio de Janeiro, Brazil: prevalence and associated factors. *Cadernos de saude Publica*. 22:2423-2431.

Daniel PB, Morgan T, Alanay Y, Bijlsma E, Cho TJ, Cole T, Collins F, David A, Devriendt K, Faivre L, Ikegawa S, Jacquemont S, Jesic M, Krakow D, Liebrecht D, Maitz S, Marlin S, Morin G, Nishikubo T, Nishimura G, Prescott T, Scarano G, Shafeghati Y, Skovby F, Tsutsumi S, Whiteford M, Zenker M, Robertson SP (2012). Disease-associated mutations in the actin-binding domain of filamin B cause cytoplasmic focal accumulations correlating with disease severity. *Hum Mutat*. 33(4):665-73.

Davidson BA, Hassan S, Garcia EJ, Tayebi N, Sidransky E (2018). Exploring genetic modifiers of Gaucher disease: The next horizon. *Hum Mutat*. 39(12):1739-1751.

Dimitrov A, Paupe V, Gueudry C, Sibarita JB, Raposo G, Vielemeyer O, Gilbert T, Csaba Z, Attie-Bitach T, Cormier-Daire V, Gressens P, Rustin P, Perez F, El Ghouzzi V (2009). The gene responsible for Dyggve-Melchior-Clausen syndrome encodes a novel peripheral membrane protein dynamically associated with the Golgi apparatus. *Hum Mol Genet*. 18(3):440-53.

Dupuis N, Fafouri A, Bayot A, Kumar M, Lecharpentier T, Ball G, Edwards D, Bernard V, Dournaud P, Drunat S, Vermelle-Andrzejewski M, Vilain C, Abramowicz M, Désir J, Bonaventure J, Gareil N, Boncompain G, Csaba Z, Perez F, Passemard S, Gressens P, El Ghouzzi V (2015). Dymeclin deficiency causes postnatal microcephaly, hypomyelination and reticulum-to-

Golgi trafficking defects in mice and humans. *Hum Mol Genet*. 24(10):2771-83.

El Ghouzzi V, Dagoneau N, Kinning E, Thauvin-Robinet C, Chemaitilly W, Prost-Squarcioni C, Al-Gazali LI, Verloes A, Le Merrer M, Munnich A, Trembath RC, Cormier-Daire V (2003). Mutations in a novel gene Dymeclin (FLJ20071) are responsible for Dyggve-Melchior-Clausen syndrome. *Hum Mol Genet*. 12(3):357-64.

Elalaoui SC, Mariam T, Ilham R, Yassamine D, Abdelaziz S (2011). A recurrent mutation in Moroccan patients with Dyggve-Melchior-Clausen syndrome: Report of a new CAe and review. *Indian J Hum Genet*. 17(2):97-9.

EUROCAT. European Surveillance of Congenital Anomalies. 2015.Available from: http://www.eurocat-network.eu/

Francine R, Pascale S, Aline H (2014). Congenital anomalies: prevalence and risk factors. *Univ J Public Health*. 2:58–63.

Fukuyama, S.I., Morino, H., Miyazawa, H., Tanaka, T., Suzuki, T., Kohda, M., Kawakami, H., Okazaki, Y., Seyama, K. and Hagiwara, K (2010). A quantitatively-modeled homozygosity mapping algorithm, qHomozygosityMapping, utilizing whole genome single nucleotide polymorphism genotyping data. In BMC bioinformatics. *BioMed Central*. 11(7):5.

Gaboon NEA, Parveen A, Ahmad KA, Shuaib T, Al-Aama JY, Abdelwehab L, Arif A, Wasif N (2020). A Novel Homozygous Frameshift Variant in DYM Causing Dyggve-Melchior-Clausen Syndrome in Pakistani Patients. *Front Pediatr*. 8:383.

Gasiunas, G., Barrangou, R., Horvath, P., & Siksnys, V (2012). CA9-crRNA ribonucleoprotein complex mediates specific DNA cleavage for adaptive immunity in bacteria. *Proceedings of the National Academy of Sciences of the United States of America*. 109(39):579–2586.

Ghosh D, Venkataramani P, Nandi S, Bhattacharjee S (2019). CRISPR-CA9 a boon or bane: the bumpy road ahead to cancer therapeutics. *Cancer Cell Int*. 19:12.

Giampietro PF, Blank RD, Raggio CL, Merchant S, Jacobsen FS, Faciszewski T, Shukla SK, Greenlee AR, Reynolds C, Schowalter DB (2003). Congenital and idiopathic scoliosis: clinical and genetic aspects. *Clin Med Res*. 1(2):125-36.

Gillani S, Kazmi NH, Najeeb S, Hussain S, Raza A (2011). Frequencies of congenital anomalies among newborns admitted in nursery of Ayub Teaching Hospital Abbottabad, Pakistan. *J Ayub Med Coll Abbottabad*. 23(1):117-21.

Gold NB, Westgate MN, Holmes LB (2011). Anatomic and etiological classification of congenital limb deficiencies. *Am J Med Genet A*. 55A(6):1225-35.

Goldfarb CA, Wall LB, Bohn DC, Moen P, Van Heest AE (2015). Epidemiology of congenital upper limAab anomalies in a midwest United States population: an assessment using the Oberg, Manske, and Tonkin classification. *J Hand Surg Am*. 40(1):127-32.

Gul F, Jabin M, Khan AS (2012). Frequency of congenital malformations and associated risk factors at Liaqat Memorial Hospital, Kohat. *Khyber Med Univ J*. 4(3):119-124.

Gupta V, Kohli A, Dewan V (2010). Dyggve melchior clausen syndrome. *Indian Pediatr*. 2010. 47(11):973-5.

Hagiwara, K., Morino, H., Shiihara, J., Tanaka, T., Miyazawa, H., Suzuki, T., Kohda, M., Okazaki, Y., Seyama, K. and Kawakami, H (2011). Homozygosity mapping on homozygosity haplotype analysis to detect recessive disease-causing genes from a small number of unrelated, outbred patients. *PLoS One*. 6(9):25059.

Hall JJ, Taylor R (2003). Health for all beyond 2000: the demise of the Alma-Ata Declaration and primary health care in developing countries. *Med J Aust*. 178(1):17-20

Hamamy H (2012). Consanguineous marriages : Preconception consultation in primary health care settings. *J Community Genet*. 3(3):185-92.

Hardy, J. and Singleton, A (2009). Genomewide association studies and human disease. *New Eng J Med*. 360(17):1759-1768.

Hildebrandt, F., Heeringa, S.F., Rüschendorf, F., Attanasio, M., Nürnberg, G., Becker, C., Seelow, D., Huebner, N., Chernin, G., Vlangos, C.N. and Zhou, W (2009). A systematic approach to mapping recessive disease genes in individuals from outbred populations. *PLoS Genet*. 5(1):1000353.

Hina S, Malik S (2015). Pattern of consanguinity and inbreeding coefficient in Sargodha district, Punjab, Pakistan. *J Biosoc Sci*. 47(6):803-11.

Huqun, Izumi S, Miyazawa H, Ishii K, Uchiyama, B., Ishida, T., Tanaka, S., Tazawa, R., Fukuyama, S., Tanaka, T. and Nagai, Y (2007). Mutations in the SLC34A2 gene are associated with pulmonary alveolar microlithiasis. *Am J Respir Crit Care Med*. 175(3):263-268.

Hussain S, Asghar I, Sabir MU, Chattha MN, Tarar SH, Mushtaq R (2014). Prevalence and pattern of congenital malformations among neonates in the neonatal unit of a teaching hospital. *J Pak Med Assoc*. 64(6):629-34.

Impallomeni M, Subramanian D, Mahmood N, Illes J (2006). Fraser syndrome in a 96-year-old female. *Age Ageing*. 35(6):642-3.

Iqbal, S., Zakar, R., Fischer, F., & Zakar, M. Z. (2022). Consanguineous marriages and their association with women's reproductive health and fertility behavior in Pakistan: secondary data analysis from Demographic and Health Surveys, 1990-2018. *BMC women's health*, 22(1), 118.

Jabeen N, Malik S (2014). Consanguinity and its sociodemographic differentials in Bhimber District, Azad Jammu and Kashmir, Pakistan. *J Health Popul Nutr*. 32(2):301-13.

Jabeen N, Malik S (2014). Prevalence of Congenital Anomalies and Non-Communicable Diseases in Women of Age 12-75 Years in District Bhimber, Azad Jammu and Kashmir, Pakistan. *Iran J Public Health*. 43(1):42-9.

Jackson M, Marks L, May GHW, Wilson JB (2018). The genetic basis of disease. *Essays Biochem.* 62(5):643-723.

Jadeja S, Smyth I, Pitera JE, Taylor MS, van Haelst M, Bentley E, McGregor L, Hopkins J, Chalepakis G, Philip N, Perez Aytes A, Watt FM, Darling SM, Jackson I, Woolf AS, Scambler PJ (2005). Identification of a new gene mutated in Fraser syndrome and mouse myelencephalic blebs. *Nat Genet.* 37(5):520-5

Jain SK (1994). A study of 200 CA of congenital limb deficiencies. *Prosthet Orthot Int.* 18(3):174-9.

Jinek M, Chylinski K, Fonfara I, Hauer M, Doudna JA, Charpentier E (2012). A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science*. 337(6096):816-21.

Kakebeen A, Wills A (2019). Advancing genetic and genomic technologies deepen the pool for discovery in Xenopus tropicalis. *Dev Dyn*. 248(8):620-625.

Karaca E, Harel T, Pehlivan D, Jhangiani SN, Gambin T, Coban Akdemir Z, Gonzaga-Jauregui C, Erdin S, Bayram Y, Campbell IM, Hunter JV, Atik MM, Van Esch H, Yuan B, Wiszniewski W, Isikay S, Yesil G, Yuregir OO, Tug Bozdogan S, Aslan H, Aydin H, Tos T, Aksoy A, De Vivo DC, Jain P, Geckinli BB, Sezer O, Gul D, Durmaz B, Cogulu O, Ozkinay F, Topcu V, Candan S, Cebi AH, Ikbal M, Yilmaz Gulec E, Gezdirici A, Koparir E, Ekici F, Coskun S, Cicek S, Karaer K, Koparir A, Duz MB, Kirat E, Fenercioglu E, Ulucan H, Seven M, Guran T, Elcioglu N, Yildirim MS, Aktas D, Alikaşifoğlu M, Ture M, Yakut T, Overton JD, Yuksel A, Ozen M, Muzny DM, Adams DR, Boerwinkle E, Chung WK, Gibbs RA, Lupski JR (2015). Genes that Affect Brain Structure and Function Identified by Rare Variant Analyses of Mendelian Neurologic Disease. *Neuron*. 88(3):499-513.

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

Keramat, K.U., Shahzad, M.S., Tariq, J. and Bhutta, A.H (2021. Etiology of Amputation in Patients Presented At HHIRS Mansehra: JRCRS-2019-11-062. *Journal Riphah College of Rehabilitation Sciences*. 9(01):18-21.

Khalifa O, Imtiaz F, Al-Sakati N, Al-Manea K, Verloes A, Al-Owain M (2011). Dyggve-Melchior-Clausen syndrome: novel splice mutation with atlanto-axial subluxation. *Eur J Pediatr*. 170(1):121-6

Khan A, Zuhaid M, Fayaz M, Ali F, Khan A, Ullah R, Zafar J, Ullah H, Baloch S, Gandapur S (2015). Frequency of congenital anomalies in newborns and its relation to maternal health in a Tertiary Care Hospital in Peshawar, Pakistan. *Int J Med Students*. 3(1):19-23.

Khokha MK, Chung C, Bustamante EL, Gaw LW, Trott KA, Yeh J, Lim N, Lin JC, Taverner N, Amaya E, Papalopulu N, Smith JC, Zorn AM, Harland RM, Grammer TC (2002). Techniques and probes for the study of Xenopus tropicalis development. *Dev Dyn*. 225(4):499-510.

Kim S, Misra A (2007). SNP genotyping: technologies and biomedical applications. *Annu Rev Biomed Eng*. 9:289-320.

Ko HS, Kim DJ, Chung Y, Wie JH, Choi SK, Park IY, Park YG, Shin JC (2017). A national cohort study evaluating infant and fetal mortality caused by birth defects in Korea. *BMJ Open*. 7(11):017963.

Kohl S, Hwang DY, Dworschak GC, Hilger AC, Saisawat P, Vivante A, Stajic N, Bogdanovic R, Reutter HM, Kehinde EO, Tasic V, Hildebrandt F (2014). Mild recessive mutations in six Fraser syndrome-related genes cause isolated congenital anomalies of the kidney and urinary tract. *J Am Soc Nephrol*. 25(9):1917-22

Korejo R, Bhutta S, Noorani KJ, Bhutta ZA (2007). An audit and trends of perinatal mortality at the Jinnah Postgraduate Medical Centre, Karachi. *Parity*. 31(40):40.

Kose MD, Canda E, Kağnıcı M, Uçar SK, Onay H, Yıldırım Sozmen E, Karapınar D, Özkınay F, Çoker M (2019). Coexistence of Gaucher Disease and severe congenital neutropenia. *Blood Cells Mol Dis*. 76:1-6.

Krakow D, Rimoin DL (2010). The skeletal dysplasias. *Genet Med*. 12(6):327-41

Krakow D, Robertson SP, King LM, Morgan T, Sebald ET, Bertolotto C, Wachsmann-Hogiu S, Acuna D, Shapiro SS, Takafuta T, Aftimos S, Kim CA, Firth H, Steiner CE, Cormier-Daire V, Superti-Furga A, Bonafe L, Graham JM Jr, Grix A, Bacino CA, Allanson J, Bialer MG, Lachman RS, Rimoin DL, Cohn DH (2004). Mutations in the gene encoding filamin B disrupt vertebral segmentation, joint formation and skeletogenesis. *Nat Genet*. 36(4):405-10.

Kramer MS, Aboud F, Mironova E, Vanilovich I, Platt RW, Matush L, Igumnov S, Fombonne E, Bogdanovich N, Ducruet T, Collet JP, Chalmers B, Hodnett E, Davidovsky S, Skugarevsky O, Trofimovich O, Kozlova L, Shapiro S (2008); Promotion of Breastfeeding Intervention Trial (PROBIT) Study Group. Breastfeeding and child cognitive development: new evidence from a large randomized trial. *Arch Gen Psychiatry*. 65(5):578-84.

Krebs DE, Edelstein JE, Thornby MA (1991). Prosthetic management of children with limb deficiencies. *Phys Ther*. 71(12):920–34.

Kumar S, Bano S (2017). Comparison and analysis of health care delivery systems: Pakistan versus Bangladesh. *J Hosp Med Manage*. 3:1-7.

Kunz F, Kayserili H, Midro A, de Silva D, Basnayake S, Güven Y, Borys J, Schanze D, Stellzig-Eisenhauer A, Bloch-Zupan A, Zenker M (2020). Characteristic

dental pattern with hypodontia and short roots in Fraser syndrome. *Am J Med Genet A*. 182(7):1681-1689

Kurdi AM, Majeed-Saidan MA, Al Rakaf MS, AlHashem AM, Botto LD, Baaqeel HS, Ammari AN (2019). Congenital anomalies and associated risk factors in a Saudi population: a cohort study from pregnancy to age 2 years. *BMJ Open*. 9(9):026351.

LaFramboise T (2009). Single nucleotide polymorphism arrays: a decade of biological, computational and technological advances. *Nucleic Acids Res*. 37(13):4181-93

Lal K, Malik S (2015). Epidemiological study of congenital limb defects in individuals or families from the interior Sindh region of Pakistan. *Asian Biomed*. 9(3):325-334.

Lander ES, Botstein D (1987). Homozygosity mapping: a way to map human recessive traits with the DNA of inbred children. *Science*. 236(4808):1567-70

Landrum MJ, Lee JM, Benson M, Brown GR, Chao C, Chitipiralla S, Gu B, Hart J, Hoffman D, Jang W, Karapetyan K, Katz K, Liu C, Maddipatla Z, Malheiro A, McDaniel K, Ovetsky M, Riley G, Zhou G, Holmes JB, Kattman BL, Maglott DR (2018). ClinVar: improving access to variant interpretations and supporting evidence. *Nucleic Acids Res*. 46(D1):D1062-D1067.

Langer LO, Moe JH (1975). A recessive form of congenital scoliosis different from spondylothoracic dysplasia. *Birth Defects Orig Artic Ser*. 11(6):83-86.

Lappalainen T, MacArthur DG (2021). From variant to function in human disease genetics. *Science*. 373(6562):1464-1468.

Lappalainen T, Scott AJ, Brandt M, Hall IM (2019). Genomic Analysis in the Age of Human Genome Sequencing. *Cell*. 177(1):70-84.

Latrech H, Skiker I, Bentata Y, Alami Z, Mouhib Lav O, Oulali N, Benajiba N, Benmassoud S, El Jabri M, Gaouzi A, Gharbi MH, Chradibi A (2013). Dyggve-Melchiore-Clausen dysplasia (DMC): syndrome associated with a micropenis. P*ediatr Endocrinol Rev*. (2):181-5.

Leppig KA, Werler MM, Cann CI, Cook CA, Holmes LB (1987). Predictive value of minor anomalies. I. Association with major malformations. *The Journal of pediatrics*. 110(4):531-537.

Li H, Durbin R (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics*. 25(14):1754-60.

Li H, Yang Y, Hong W, Huang M, Wu M, Zhao X (2020). Applications of genome editing technology in the targeted therapy of human diseases: mechanisms, advances and prospects. *Signal Transduct Target Ther*. 3;5(1):1.

Liu A, McEntee J (2019). Osteochondrodysplasia. *Pediatr Rev*. 40(8):435-438.

Liu Z, Cheng TT, Shi Z, Liu Z, Lei Y, Wang C, Shi W, Chen X, Qi X, Cai D, Feng B, Deng Y, Chen Y, Zhao H (2016). Efficient genome editing of genes involved in neural crest development using the CRISPR/CA9 system in Xenopus embryos. *Cell Biosci*. 31;6:22.

Lowry RB, Bedard T (2016). Congenital limb deficiency classification and nomenclature: The need for a consensus. *Am J Med Genet A.* 170(6):1400-4.

Lu J, Lian G, Lenkinski R, De Grand A, Vaid RR, Bryce T, Stasenko M, Boskey A, Walsh C, Sheen V (2007). Filamin B mutations cause chondrocyte defects in skeletal development. *Hum Mol Genet*. 16(14):1661-75.

Makarova KS, Wolf YI, Iranzo J, Shmakov SA, Alkhnbashi OS, Brouns SJJ, Charpentier E, Cheng D, Haft DH, Horvath P, Moineau S, Mojica FJM, Scott D, Shah SA, Siksnys V, Terns MP, Venclovas Č, White MF, Yakunin AF, Yan W, Zhang F, Garrett RA, Backofen R, van der Oost J, Barrangou R, Koonin EV (2020). Evolutionary classification of CRISPR-CA systems: a burst of class 2 and derived variants. *Nat Rev Microbiol*. 18(2):67-83.

Mangaraj S, Choudhury AK, Singh M, Patro D, Baliarsinha AK (2017). Spondylocarpotarsal synostosis syndrome. A rare CAe of short stature and congenital scoliosis. *Clin CA Miner Bone Metab*. 14(2):258-261.

Manghwar H, Lindsey K, Zhang X, Jin S (2019). CRISPR/CA System: Recent Advances and Future Prospects for Genome Editing. *Trends Plant Sci*. 24(12):1102-1125.

Mano H, Fujiwara S, Takamura K, Kitoh H, Takayama S, Ogata T, Hashimoto S, Haga N (2018). Congenital limb deficiency in Japan: a cross-sectional nationwide survey on its epidemiology. *BMC Musculoskelet Disord*. 19(1):262

Manolio TA (2010). Genomewide association studies and assessment of the risk of disease. *N Engl J Med*. 363(2):166-76.

Maroilley T, Tarailo-Graovac M (2019). Uncovering Missing Heritability in Rare Diseases. *Genes (Basel).* 10(4):275.

Martínez-Frías ML, Cormier-Daire V, Cohn DH, Mendioroz J, Bermejo E, Mansilla E (2007). Síndrome de Dyggve-Melchior-Clausen: presentación de un CAo con una mutación de posible origen español [Dyggve-Melchior-Clausen

syndrome: presentation of a CAe with a mutation of possible Spanish origin]. *Med Clin (Barc).* 128(4):137-40.

Marzin P, Cormier-Daire V (2020). New perspectives on the treatment of skeletal dysplasia. *Ther Adv Endocrinol Metab*. 11:1-16.

Masood Jawaid IA, Kaimkhani GM (2008). Current indications for major lower limb amputations at Civil Hospital, Karachi. *Pak J Surg*. 24(4):228-231.

Masood SN, Jamil N, Mumtaz SN, Masood MF, Muneer S (2011). Congenital malformations in newborns of consanguineous and non-consanguineous parents. *Pak J Med Sci*. 27:1–5.

Maurano MT, Humbert R, Rynes E, Thurman RE, Haugen E, Wang H, Reynolds AP, Sandstrom R, Qu H, Brody J, Shafer A, Neri F, Lee K, Kutyavin T, Stehling-Sun S, Johnson AK, Canfield TK, Giste E, Diegel M, Bates D, Hansen RS, Neph S, Sabo PJ, Heimfeld S, Raubitschek A, Ziegler S, Cotsapas C, Sotoodehnia N, Glass I, Sunyaev SR, Kaul R, Stamatoyannopoulos JA (2012). Systematic localization of common disease-associated variation in regulatory DNA. *Science*. 337(6099):1190-5.

McCarthy MI, AbeCAis GR, Cardon LR, Goldstein DB, Little J, Ioannidis JP, Hirschhorn JN (2008). Genome-wide association studies for complex traits: consensus, uncertainty and challenges. *Nat Rev Genet.* 9(5):356-69.

McFarland LV, Choppa AJ, Betz K, Pruden JD, Reiber GE (2010) . Resources for wounded warriors with major traumatic limb loss. *J Rehab Res Dev*. 47(4).

McInerney-Leo AM, Duncan EL (2021). Massively Parallel Sequencing for Rare Genetic Disorders: Potential and Pitfalls. *Front Endocrinol (Lausanne).* 11:628946.

McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytsky A, Garimella K, Altshuler D, Gabriel S, Daly M, DePristo MA (2010). The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res.* 20(9):1297-303.

McQueen C, Pownall ME (2017). An analysis of MyoD-dependent transcription using CRISPR/CA9 gene targeting in Xenopus tropicalis embryos. *Mech Dev*. 146:1-9.

Michielsen A, Van Wijk I, Ketelaar M (2010). Participation and quality of life in children and adolescents with congenital limb deficiencies: A narrative review. *Prosthet Orthot Int*. 34(4):351-61.

Mirza I, Tareen A, Davidson LL, Rahman A (2009). Community management of intellectual disabilities in Pakistan: a mixed methods study. *J Intellect Disabil Res.* 53(6):559-70.

Missaglia S, Tasca E, Angelini C, Moro L, Tavian D (2015). Novel missense mutations in PNPLA2 causing late onset and clinical heterogeneity of neutral lipid storage disease with myopathy in three siblings. *Mol Genet Metab*. 115(2-3):110-7.

Miyazawa H, Kato M, Awata T, Kohda M, Iwasa H, Koyama N, Tanaka T, Huqun, Kyo S, Okazaki Y, Hagiwara K (2007). Homozygosity haplotype allows a genomewide search for the autosomal segments shared among patients. *Am J Hum Genet*. 80(6):1090-102.

Modell B, Darr A (2002) . Genetic counselling and customary consanguineous marriage. *Nature Reviews Genetics*. 3(3):225-229.

Nakamura F, Stossel TP, Hartwig JH (2011). The filamins: organizers of cell structure and function. *Cell Adh Migr.* 5(2):160-9.

Nawaz A, Zaman M, Malik S (2021). Consanguinity, inbreeding coefficient, fertility and birth-outcome in population of Okara district, Pakistan. *Pak J Med Sci*. 37(3):770-775.

Ndibazza, J., Lule, S., Nampijja, M., Mpairwe, H., Oduru, G., Kiggundu, M., Akello, M., Muhangi, L., & Elliott, A. M. (2011). A description of congenital anomalies among infants in Entebbe, Uganda. Birth defects research. *Part A, Clinical and molecular teratology*. 91(9): 857–861.

Nectoux E, Hocquet B, Fron D, Mezel A, Paris A, Herbaux B (2013). Unpredictability of hip behavior in Dyggve-Melchior-Clausen syndrome: a mid-term assessment of siblings. *Orthop Traumatol Surg Res*. 99(6):745-8.

Neumann LM, El Ghouzzi V, Paupe V, Weber HP, Fastnacht E, Leenen A, Lyding S, Klusmann A, Mayatepek E, Pelz J, Cormier-Daire V (2006). Dyggve-Melchior-Clausen syndrome and Smith-McCort dysplasia: clinical and molecular findings in three families supporting genetic heterogeneity in Smith-McCort dysplasia. *Am J Med Genet A*. 140(5):421-6.

Nieuwkoop PD, Faber J (1994). Normal Table of Xenopus laevis (Daudin). New York: Garland Publishing Inc.

NIPS. National Institute of Population Studies [Pakistan] and ICF (2019). Pakistan Demographic and Health Survey 2017-18. Islamabad, Pakistan, and Rockville, Maryland, USA: NIPS and ICF.

Oliveira CI, Fett-Conte AC (2013). Birth defects: Risk factors and consequences. *J Pediatr Genet*. 2(2):85-90.

Omar N, Kokab F (2019). Intellectual disability among special children and its associated factors: A CAe control study, Lahore Pakistan. *J Pak Med Assoc*. 69(5):684-689.

Osipovich AB, Jennings JL, Lin Q, Link AJ, Ruley HE (2008). Dyggve-Melchior-Clausen syndrome: chondrodysplasia resulting from defects in intracellular vesicle traffic. *Proc Natl Acad Sci U S A*. 105(42):16171-6.

Ott J (1999). Analysis of Human Genetic Linkage. Johns Hopkins University Press.

Panda A, Gamanagatti S, Jana M, Gupta AK (2014). Skeletal dysplasias: A radiographic approach and review of common non-lethal skeletal dysplasias. *World J Radiol*. 6(10):808-25.

Pavlakis E, Chiotaki R, Chalepakis G (2011). The role of Fras1/Frem proteins in the structure and function of basement membrane. *Int J Biochem Cell Biol*. 43(4):487-95.

Peltonen L, Palotie A, Lange K (2000). Use of population isolates for mapping complex traits. *Nat Rev Genet*. 1(3):182-190.

Rahit KMTH, Tarailo-Graovac M (2020). Genetic Modifiers and Rare Mendelian Disease. *Genes (Basel)*. 11(3):239.

Ran FA, Hsu PD, Wright J, Agarwala V, Scott DA, Zhang F (2013). Genome engineering using the CRISPR-CA9 system. *Nat Protoc*. 8(11):2281-2308.

Rankin J, Pattenden S, Abramsky L, Boyd P, Jordan H, Stone D, Vrijheid M, Wellesley D, Dolk H (2005). Prevalence of congenital anomalies in five British regions, 1991-99. *Arch Dis Child Fetal Neonatal Ed*. 90(5):374-479.

Rathore FA, Ayaz SB, Mansoor SN, Qureshi AR, Fahim M (2016). Demographics of Lower Limb Amputations in the Pakistan Military: *A Single Center, Three-Year Prospective Survey*. 8(4):566.

Riaz HF, Lal K, Ahmad B, Shuaib M, Naqvi SF, Malik S (2014). Study of non-syndromic thumb aplasia in six independent CA. *Pak J Med Sci*. 30(3):677-681.

Riaz HF, Mannan S, Malik S (2016). Consanguinity and its socio-biological parameters in Rahim Yar Khan District, Southern Punjab, Pakistan. *J Health Popul Nutr*. 35:14.

Rice DP (2005). Craniofacial anomalies: from development to molecular pathogenesis. *Curr Mol Med*. 5(7):699-722.

Rocha D, Gut I, Jeffreys AJ, Kwok PY, Brookes AJ, Chanock SJ (2006). Seventh international meeting on single nucleotide polymorphism and complex genome analysis: 'ever bigger scans and an increasingly variable genome'. *Hum Genet.* 119(4):451-460.

Rosenberg SA, Aebersold P, Cornetta K, Kasid A, Morgan RA, Moen R, Karson EM, Lotze MT, Yang JC, Topalian SL, et al (1990). Gene transfer into humans--immunotherapy of patients with advanced melanoma, using tumor-infiltrating lymphocytes modified by retroviral gene transduction. *N Engl J Med*. 323(9):570-8.

Sabir A, Irving M (2021). Clinical trials in skeletal dysplasia: a paradigm for treating rare diseases. *Br Med Bull*. 139(1):16-35.

Sabir AH, Cole T (2019). The evolving therapeutic landscape of genetic skeletal disorders. *Orphanet J Rare Dis*. 14(1):300.

Salgado D, Bellgard MI, Desvignes JP, Béroud C (2016). How to Identify Pathogenic Mutations among All Those Variations: Variant Annotation and Filtration in the Genome Sequencing Era. *Hum Mutat*. (12):1272-1282.

Salian S, Shukla A, Shah H, Bhat SN, Bhat VR, Nampoothiri S, Shenoy R, Phadke SR, Hariharan SV, Girisha KM (2018). Seven additional families with spondylocarpotarsal synostosis syndrome with novel biallelic deleterious variants in FLNB. *Clin Genet*. 94(1):159-164.

Sarkar, S., Patra, C., Dasgupta, M. K., Nayek, K., & Karmakar, P. R. (2013). Prevalence of congenital anomalies in neonates and associated risk factors in a tertiary care hospital in eastern India. *Journal of clinical neonatology*. 2(3): 131–134.

Schanze D, Kayserili H, Satkın BN, Altunoglu U, Zenker M (2014). Fraser syndrome due to mutations in GRIP1--clinical phenotype in two families and expansion of the mutation spectrum. *Am J Med Genet A*. 164A(3):837-840

Schorr S, Legum C, Ochshorn M, Hirsch M, Moses S, Lasch EE, El-Masri M (1977). The Dyggve-Melchio-Clausen syndrome. *AJR Am J Roentgenol*. 128(1):107-113.

Schwarz JM, Cooper DN, Schuelke M, Seelow D (2014). MutationTaster2: mutation prediction for the deep-sequencing age. *Nat Methods*. 11(4):361-362.

Sega AG, Mis EK, Lindstrom K, Mercimek-Andrews S, Ji W, Cho MT, Juusola J, Konstantino M, Jeffries L, Khokha MK, Lakhani SA (2019). De novo pathogenic variants in neuronal differentiation factor 2 (NEUROD2) cause a form of early infantile epileptic encephalopathy. *J Med Genet*. 56(2):113-122.

Sener G, Yigiter K, Bayar K, Erbahceci F (1999). Effectiveness of prosthetic rehabilitation of children with limb deficiencies present at birth. *Prosthet Orthot Int*23:130 –134.

Shah N, Rohra DK, Shams H, Khan NH (2010). Home deliveries: reasons and adverse outcomes in women presenting to a tertiary care hospital. *J Pak Med Assoc*. 60(7):555-8.

Shawky RM, Elsayed SM, Zaki ME, El-Din SM, Kamal FM (2013). Consanguinity and its relevance to clinical genetics. *Egypt J Med Hum Genet.* 14(2):157-164.

Shimizu H, Watanabe S, Kinoshita A, Mishima H, Nishimura G, Moriuchi H, Yoshiura KI, Dateki S (2019). Identification of a homozygous frameshift variant in RFLNA in a patient with a typical phenotype of Spondylocarpotarsal synostosis syndrome. *J Hum Genet*. 64(5):467-471.

Short K, Wiradjaja F, Smyth I (2007). Let's stick together: the role of the Fras1 and Frem proteins in epidermal adhesion. *IUBMB Life*. 59(7):427-35.

Sokal R, Tata LJ, Fleming KM (2014). Sex prevalence of major congenital anomalies in the United Kingdom: a national population-based study and international comparison meta-analysis. *Birth Defects Res A Clin Mol Teratol.* 100(2):79-91.

Soomro, N., Bibi, R., Ahmed, S.I., Kamran, B., Minhas, M.A. and Siddiqui, K.Y (2013). Epidemiology of amputation. *The Professional Medical Journal*. 20(02):261-265.

Sperber GH, Sperber SM. The genesis of craniofacial biology as a health science discipline. Aust Dent J. 2014 Jun;59 Suppl 1:6-12.

Spranger J, Bierbaum B, Herrmann J (1976). Heterogeneity of Dyggve-Melchior-Clausen dwarfism. *Hum Genet.* 33(3):279-87.

Stossel TP, Condeelis J, Cooley L, Hartwig JH, Noegel A, Schleicher M, Shapiro SS (2001)Lu. Filamins as integrators of cell mechanics and signalling. *Nat Rev Mol Cell Biol*. 2(2):138-45.

Sun L, Huang Y, Zhao S, Zhao J, Yan Z, Guo Y, Lin M, Zhong W, Yin Y, Chen Z, Zhang N, Zhang Y, Zhao Z, Li Q, Wang L, Dong X, Li Y, Li X, Qiu G (2021). DISCO (Deciphering Disorders Involving Scoliosis & COmorbidities) study group, Zhang TJ, Wu Z, Tian W, Wu N. Deciphering the mutational signature of congenital limb malformations. *Mol Ther Nucleic Acids*. 24:961-970.

Tanaka AJ, Cho MT, Millan F, Juusola J, Retterer K, Joshi C, Niyazov D, Garnica A, Gratz E, Deardorff M, Wilkins A, Ortiz-Gonzalez X, Mathews K, Panzer K, Brilstra E, van Gassen KL, Volker-Touw CM, van Binsbergen E, Sobreira N, Hamosh A, McKnight D, Monaghan KG, Chung WK (2015). Mutations in SPATA5 Are Associated with Microcephaly, Intellectual Disability, Seizures, and Hearing Loss. *Am J Hum Genet*. 97(3):457-64.

Tarailo-Graovac M, Zhu JYA, Matthews A, van Karnebeek CDM, Wasserman WW (2017). Assessment of the ExAC data set for the presence of individuals with

pathogenic genotypes implicated in severe Mendelian pediatric disorders. *Genet Med*. 19(12):1300-1308.

Tennant PW, Pearce MS, Bythell M, Rankin J (2010). 20-year survival of children born with congenital anomalies: a population-based study. *Lancet*. 375(9715):649-56.

Thomas IT (1986), Frias JL, Felix V, Sanchez de Leon L, Hernandez RA, Jones MC. Isolated and syndromic cryptophthalmos. *Am J Med Genet*. 25(1):85-98.

Tseng CL, Helmer D, Rajan M, Tiwari A, Miller D, Crystal S, Safford M, Greenberg J, Pogach L (2007). Evaluation of regional variation in total, major, and minor amputation rates in a national health-care system. *Int J Qual Health Care*. 19(6):368–76.

Tufail M, Rehman AU, Malik S. Determinants of consanguinity and inbreeding coefficient in the multiethnic population of Mardan, Khyber Pakhtunkhwa, Pakistan. *Asian Biomed*. 11(6):451–60.

Ullah S, Dasti JI, Malik S (2015). Descriptive epidemiology of hereditary musculoskeletal and limb defects in the isolated population of Chitral, North-West Pakistan. *Pak J Med Sci*. 31(5):1047–52.

Umair M, Ahmad F, Ullah A (2018). Whole exome sequencing as a diagnostic tool for genetic disorders in Pakistan. *Pak J Med Res*. 57(2):90-91.

van Haelst MM, Scambler PJ (2007). Fraser Syndrome Collaboration Group, Hennekam RC. Fraser syndrome: a clinical study of 59 CA and evaluation of diagnostic criteria. *Am J Med Genet A*. 143A(24):3194-203.

Van Otterloo E, Williams T, Artinger KB (2016). The old and new face of craniofacial research: How animal models inform human craniofacial genetic and clinical data. *Dev Biol*. 415(2):171-187.

Varma P, Stineman MG, Dillingham TR (2014). Epidemiology of limb loss. *Phys Med Rehabil Clin N Am.* 25(1):1-8.

Vogel F, Motulsky AG (2013). Vogel and Motulsky's Human Genetics: Problems and Approaches. *Springer Science & Business Media*. 5(1):73-74.

Vogel MJ, van Zon P, Brueton L, Gijzen M, van Tuil MC, Cox P, Schanze D, Kariminejad A, Ghaderi-Sohi S, Blair E, Zenker M, Scambler PJ, Ploos van Amstel HK, van Haelst MM (2012). Mutations in GRIP1 cause Fraser syndrome. *J Med Genet*. 49(5):303-6.

von Elm E, Altman DG, Egger M, Pocock SJ, Gøtzsche PC, Vandenbroucke JP (2008). STROBE Initiative. The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement: guidelines for reporting observational studies. *J Clin Epidemiol*. 61(4):344-9.

Wahab A, Ahmad M (1996). Biosocial perspective of consanguineous marriages in rural and urban Swat, Pakistan. *Journal of biosocial science*. 28(3):305-313.

Wang DG, Fan JB, Siao CJ, Berno A, Young P, Sapolsky R, Ghandour G, Perkins N, Winchester E, Spencer J, Kruglyak L, Stein L, Hsie L, Topaloglou T, Hubbell E, Robinson E, Mittmann M, Morris MS, Shen N, Kilburn D, Rioux J, Nusbaum C, Rozen S, Hudson TJ, Lipshutz R, Chee M, Lander ES (1998). Large-scale identification, mapping, and genotyping of single-nucleotide polymorphisms in the human genome. *Science*. 280(5366):1077-82.

Wang Y, Hu J, Druschel CM, Kirby RS (2011). Twenty-five-year survival of children with birth defects in New York State: a population-based study. *Birth Defects Res A Clin Mol Teratol*. 91(12):995-1003.

Warman ML, Cormier-Daire V, Hall C, Krakow D, Lachman R, LeMerrer M, Mortier G, Mundlos S, Nishimura G, Rimoin DL, Robertson S, Savarirayan R, Sillence D, Spranger J, Unger S, Zabel B, Superti-Furga A (2011). Nosology and classification of genetic skeletal disorders: 2010 revision. *Am J Med Genet A*. 155A(5):943-68.

Weir S, Ephraim P, Mackenzie E (2010). Effects of paediatric limb loss on healthcare utilisation, schooling and parental labour supply. *Disabil Rehabil*. 32(24):2046-55.

Wen L, Shibata Y, Su D, Fu L, Luu N, Shi YB (2017). Thyroid Hormone Receptor α Controls Developmental Timing and Regulates the Rate and Coordination of Tissue-Specific Metamorphosis in Xenopus tropicalis. *Endocrinology.* 158(6):1985-1998.

WHO. World Health Organization (2010). International Classification of Disease. CD-10 [internet]. Geneva: WHO; Available from: http://apps.who.int/classifications/icd10/browse/2010/en

WHO. World Health Organization (2010). Sixty -Third World health assembly. Care port by the Secretariat, A63/10. Geneva, Switzerland: World Health Organization.

WHO. World Health Organization (2014). Facts Sheet on Hereditary Disorders. http://www.who.int/mediacentre/ factsheets/fs370/en/ (Accessed Dec. 2021)

WHO. World Health Organization (2014). Neonatal-perinatal database and birth defects surveillance. Report of the regional review meeting, New Delhi, India, 19–21 August 2014.

WHO. World Health Organization (2016). Global health Observatory (GHO) data. Under-five mortality, 2016. Available: http://www. who. int/ gho/ child_ health/ mortality/ mortality_ under_ five_ text/ en/ (Accessed 12 Dec 2021).

Wilcox WR, Coulter CP, Schmitz ML (2015). Congenital limb deficiency disorders. *Clin Perinatol*. 42(2):281-300.

Yang CF, Wang CH, Siong H'ng W, Chang CP, Lin WD, Chen YT, Wu JY, Tsai FJ (2017). Filamin B Loss-of-Function Mutation in Dimerization Domain Causes Autosomal-Recessive Spondylocarpotarsal Synostosis Syndrome with Rib Anomalies. *Hum Mutat*. 38(5):540-547.

Yasin S, Makitie O, Naz S (2021). Spondylocarpotarsal synostosis syndrome due to a novel loss of function FLNB variant: a CAe report. *BMC Musculoskelet Disord*. 22(1):31

Zahn N, Levin M, Adams DS (2017). The Zahn drawings: new illustrations of Xenopus embryo and tadpole stages for studies of craniofacial development. *Development.* 144(15):2708-2713.

Zahra Q, Shuaib M, Malik S (2016). Epidemiology of congenital anomalies in the Kurram Tribal Agency, northwest Pakistan. *Asian Biomed.* 10(6):591-601..

Zhang C, Quan R, Wang J (2018). Development and application of CRISPR/CA9 technologies in genomic editing. *Hum Mol Genet*.27(2):79-88.

Zhou X, Tian F, Sandzén J, Cao R, Flaberg E, Szekely L, Cao Y, Ohlsson C, Bergo MO, Borén J, Akyürek LM (2007). Filamin B deficiency in mice results in skeletal malformations and impaired microvascular development. *Proc Natl Acad Sci U S A*. 104(10):3919-24

List of databases and *in silico* tools

Database	Description	Homepage
1000 Genome	The detailed catalog of human genetic variation and project designed for the sequence of 1000 anonymous individuals from different ethnic groups.	http://www.internationalgenome.org/1000-genomes-browsers
ClinVar	ClinVar aggregates information about genomic variation and its relationship to human health.	https://www.ncbi.nlm.nih.gov/clinvar/
CRISPRscan	CRISPRscan is a novel scoring algorithm from the Giraldez Lab (Yale University) that helps you select the best gRNAs.	https://www.crisprscan.org/
ENSEMBL Genome Browser	Scientific project at European Bioinformatics Institute to provide a centralized resource for studying the genome of our and different species.	https://asia.ensembl.org/index.html
Gene Cards	The database provides genomic, transcriptomic, proteomic, genetic, and functional information on all known and predicted human genes.	https://www.genecards.org/
Gene Ontology (GO)	Bioinformatics tool to unify the gene and gene products across all species.	http://geneontology.org/
GeneDistiller	The tool that is used to query, select and project genes from within a linkage interval.	http://www.genedistiller.org/
Genome Aggregation Database (gnomAD)	A resource developed by international collaborations with the goal of aggregation and harmonizing both exome and genome sequencing data.	http://gnomad.broadinstitute.org/
HGMD	The Human Gene Mutation Database (HGMD®), which is kept up to date in Cardiff by D.N. Cooper, E.V. Ball, P.D. Stenson, A.D. Phillips, and K. Evans, S. Heywood, M.J. Hayden, M.M. Chapman, M.E. Mort, L.	https://www.hgmd.cf.ac.uk/

	Azevedo, and D.S. Millar, represents an effort to compile all known (published) gene lesions responsible for human inherited disease	
HomoloGene	Tool designed for the automated detection of homologs among the annotated genes of completely sequenced eukaryotic genomes.	https://www.ncbi.nlm.nih.gov/homologene
HomozygosityMapper	A tool used to obtain the homozygosity intervals across the genome.	http://www.homozygositymapper.org/
Human Splicing Finder 3.1	An online bioinformatic tool designed to predict the splicing signals.	http://umd.be/Redirect.html
IGV	The Integrative Genomics Viewer (IGV) is a high-performance, easy-to-use, interactive tool for the visual exploration of genomic data.	https://software.broadinstitute.org/software/igv/
Mendelian Clinically Applicable Pathogenicity (M-CAP)	The first pathogenic tool is used for the classification of a rare missense variant in the human genome.	http://bejerano.stanford.edu/mcap/
Mutation Taster2	Web-based tool to evaluate the DNA sequence variants for their disease-causing potential.	http://www.mutationtaster.org/
NCBI PubMed	The database comprised more than 29M citations for biomedical literature from MEDLINE, life science journals, and online books.	https://pubmed.ncbi.nlm.nih.gov/
Online Mendelian Inheritance in Man (OMIM)	OMIM is an online continuously updating database with human genes and genetic traits, with a focus on the genotype-phenotype relationship.	https://www.ncbi.nlm.nih.gov/omim
Polymorphism Phenotyping (PolyPhen-2)	The tool that predicts the possible impact of an amino acid on the structure and function of the protein.	http://genetics.bwh.harvard.edu/pph2/

Primer3 software	A computer program that suggests PCR primers for a variety of applications e.g. creation of sequence tag sites for radiation hybrid mapping.	https://bioinfo.ut.ee/primer3-0.4.0/
Protein Variation Effect Analyzer (PROVEAN)	Software that is used to predict whether amino acids substitution or indels has an impact on the biological function of a protein.	http://provean.jcvi.org/index.php
SIFT (Sorting Intolerant from Tolerant)	A tool that predicts the potential impact of amino acid substitution on protein function.	https://sift.bii.a-star.edu.sg
STRING	STRING is a biological database and online resource for protein-protein interactions that have been observed and predicted in molecular biology. Information from a variety of sources, including experimental data, computer prediction techniques, and public text collections, is included in the STRING database.	https://string-db.org/
UCSC Genome Browser	Online genome browser that provides access to genomic sequence data of vertebrate and invertebrate species.	https://genome.ucsc.edu/
UniProt (Universal Protein Resource)	Database that provides comprehensive, high-quality, and freely accessible resources of the protein sequence.	https://www.uniprot.org/
Xenbase	Xenbase is a web-accessible resource that integrates all the diverse biological, genomic, genotype and phenotype data available from Xenopus research.	https://www.xenbase.org/

Turnitin Originality Report

Genetic epidemiology, mapping and molecular characterization of families with inherited skeletal and neurological disorders by Anisa Bibi .

From PhD (PhD DRSML)

- Processed on 14-Dec-2022 10:32 PKT
- ID: 1980872229
- Word Count: 36012

Similarity Index
17%
Similarity by Source

Internet Sources:
11%
Publications:
15%
Student Papers:
3%

sources:

1 2% match (Internet from 12-Oct-2022)
http://pjms.org.pk/index.php/pjms/article/download/5486/1377/

2 1% match (Mine Koprulu, Aneeta Kumare, Anisa Bibi, Sajid Malik, Aslıhan Tolun. " The first adolescent case of Fraser syndrome 3, with a novel nonsense variant in ", American Journal of Medical Genetics Part A, 2021)
Mine Koprulu, Aneeta Kumare, Anisa Bibi, Sajid Malik, Aslıhan Tolun. " The first adolescent case of Fraser syndrome 3, with a novel nonsense variant in ", American Journal of Medical Genetics Part A, 2021

3 1% match (Internet from 25-Sep-2022)
https://sciendo.com/article/10.1515/abm-2019-0040

4 1% match (Annalisa G Sega, Emily K Mis, Kristin Lindstrom, Saadet Mercimek-Andrews et al. "De novo pathogenic variants in neuronal differentiation factor 2 (NEUROD2) cause a form of early infantile epileptic encephalopathy", Journal of Medical Genetics, 2019)
Annalisa G Sega, Emily K Mis, Kristin Lindstrom, Saadet Mercimek-Andrews et al. "De novo pathogenic variants in neuronal differentiation factor 2 (NEUROD2) cause a form of early infantile epileptic encephalopathy", Journal of Medical Genetics, 2019

5 < 1% match (Internet from 21-Nov-2022)
https://sciendo.com/pdf/10.1515/abm-2019-0040

6 < 1% match ()
Samina Yasin, Outi Makitie, Sadaf Naz. "Spondylocarpotarsal synostosis syndrome due to a novel loss of function variant: a case report", BMC Musculoskeletal Disorders

7 < 1% match ()
Engin Deniz, Stephan Jonas, Michael Hooper, John N. Griffin, Michael A. Choma, Mustafa K. Khokha. "Analysis of Craniocardiac Malformations in using Optical Coherence Tomography", Scientific Reports

8 < 1% match ()
Udhaya Kumar S., Srivarshini Sankar, Salma Younes, Thirumal Kumar D. et al. "Deciphering the Role of Filamin B Calponin-Homology Domain in Causing the Larsen Syndrome, Boomerang Dysplasia, and Atelosteogenesis Type I Spectrum Disorders via a Computational Approach", Molecules

9 < 1% match ()
PLoS Genet. 2020 Aug 19; 16(8):e1008967

10 < 1% match (student papers from 18-Dec-2017)
Submitted to Higher Education Commission Pakistan on 2017-12-18

11 < 1% match (student papers from 31-Jul-2021)
Submitted to Higher Education Commission Pakistan on 2021-07-31