

# Identification of novel mutations and In-Silico analysis of *MCPH1* gene in Pakistani family with microcephaly

Zunaira Riaz\*, Saba Irshad\*, Tanveer Ahmed\*\*\*\*, Farman Ali\*\*\*\*

\* School of Biochemistry and Biotechnology, University of the Punjab, 54000-Lahore, Pakistan

\*\* Laboratory of Metabolism and Cell Fate, Guangzhou Institutes of Biomedicine and Health, Chinese Academy of Sciences, 510530 Guangzhou, China

\*\*\* Cadson College of Pharmacy, University of the Punjab, Lahore, Pakistan

\*\*\*\*Clinical Research Institute, the Second Affiliated Hospital &Academy of Integrative Medicine, Fujian University of Traditional Chinese Medicine, Fuzhou, Fujian350122, China.

\*For Correspondence

Saba Irshad\*<sup>1</sup>

## Abstract

Primary microcephaly (MCPH) is a congenital, static, and non-progressive neurodevelopmental disorder associated with reduced head circumference (< 4 standard deviations). About 28 known genes are associated with the MCPH. The study was carried out to probe molecular basis and genetic variants involved with MCPH in an affected Pakistani family to better understand the etiology and prevalence of the disorder. The individuals of the ascertained Pakistani family presented primary microcephaly, along with intellectual disability, speech disorder, and motor delay. By ensuring ethical compliance and patient consent, blood samples were collected from affected individuals. DNA was extracted using the salting out method followed by whole-exome sequencing and Sanger sequencing to identify causative genetic variants or mutations. In silico studies were performed to predict the effect of mutations on the structure of target proteins. Two missense allelic variants (NM\_024596.5: c.139G>C and NM\_024596.5: c.211G>C) of *MCPH1* gene were detected in a Pakistani family. In silico analysis was performed to evaluate the effect of the mutant protein. The mutation in genes affects the activities of proteins NM\_024596: p. Val47Leu and NM\_024596: p. Val71Leu respectively by disruption in protein structure. The mutations were predicted to have higher pathogenicity scores and have significantly influenced the prevalence of MCPH. We reported two novel genetic variants for the first time from the Pakistani population causing MCPH. Mutations in *MCPH1* gene are one of the major causes of MCPH in the populations where consanguine marriages are common. The novel mutations identified in this study will help to understand the

etiology of the disorder and the mechanisms of mutated proteins.

**Keywords:** primary microcephaly, *MCPH1*, exome mapping, mutation, proteins, in silico.

## 1. INTRODUCTION

Microcephaly is the neurological disorder in which reduced head size is the most prominent phenotypic expression in the individuals suffering from the neurodevelopmental disorder of microcephaly. The static abnormal brain development leading to small head size is referred to as congenital microcephaly or primary microcephaly. While the postnatal progressive neurodegeneration leading to abnormal growth of the brain is called degenerative microcephaly or secondary microcephaly [1, 2]. Primary-microcephaly (MCPH) is rare, nonprogressive, and congenital neuro-developmental syndrome. It was frequently connected with the considerably lower occipitofrontal circumference of the head (-4 SD) at the time of birth [1]. Individuals with MCPH are categorized by lesser brain-size, along with varying amount of motor-delay, and intellectual-disability [2]. However, congenital microcephaly shows wide clinical and genetic heterogeneity [3].

Worldwide about 1.3-150/100,000 persons were reported to have microcephaly [4]. The occurrence of MCPH is advanced in the Subcontinental and Arab inhabitants, where the rate of consanguine is higher [5, 6]. In Pakistan especially in northern regions due to diverse ethnicity consanguineous marriages are common with microcephaly cases occurrence of 1/100,000 [7]. The genetically heterogeneous pathogenesis of MCPH can be estimated as there are 28 loci (MCPH1-MCPH28) that have been recognized to possess genes related with primary-microcephaly

such as *MCPH1*, *WDR62*, *CDK5RAP2*, *CASC5*, *ASPM*, *CENPJ*, *STIL*, *CEP135*, *CEP152*, *ZNF335*, *PHC1*, *CDK6*, *CENPE*, *SASS6*, *MFS2A*, *ANKLE2*, *CIT*, *WDFY3*, *COPB2*, *KIF14*, *NCAPD2*, *NCAPD3*, *NCAPH*, *NUP37*, *MAP11*, *LMNB1*, *LMNB2*, and *RRP7A* [8, 9]. Among these genes the greatest prevalent cause of the MCPH are mutations in *ASPM*, *WDR62* and *MCPH1* genes that accounts for the 68.5% 14.1% and 8% of the total MCPH cases, respectively [10, 11].

With the advent and advances in next-generation sequencing technologies, nowadays it has been an easy approach to identify the genetic-variations that inspire the phenotypic consequence of primary casual agents [12-14]. The mutations can influence the variability in phenotypic expression and are important for diagnosis and therapy of the disease [15]. Such deleterious/insertion mutations are being frequently observed in neurodevelopmental disorders such as microcephaly [16-19]. Several studies reported that abnormal development due to mutations in *microcephalin* genes controlling neuronal development via cellular and physiologically crucial pathways are the causes of microcephaly [20-24].

Among the major causes of primary microcephaly, gene mutation-*MCPH1* are the third foremost cause of primary-microcephaly [25]. *MCPH1*-gene (*MCPH1* locus) maps to the chromosomal location of 8p23.1, inherited in autosomal recessive mode. *MCPH1* with 14-exon-encodes for the protein constituting 835 amino-acids [19]. It is involved in the regulation of cellular processes crucial for normal brain development such as chromosomal condensation, G2/M check-point, the cell-cycle capture, and DNA-damage reaction [26-28].

In this study, we described two-novel mutations involving the *MCPH1* genes that were found associated with primary microcephaly in consanguineous Pakistani families. We identified two missense variants Val47Leu and Val71Leu in *MCPH* protein. *In-silico* studies the consequence of the alterations is predicted on the functioning of protein, different developmental pathways and the disease pathogenicity were performed. It provided a detailed insight into the ligand-protein interactions (NAD-MCPH), hydrogen bond count and binding affinities that suggested considerable pathological role of the identified mutations. This study will further explain the molecular basis and genetic variants associated with the development of this disorder in the Pakistani population.

## 2. MATERIALS AND METHODS

Families pretentious with primary microcephaly were recognized and taken into consideration from distant ranges of Pakistan. From the ascertained individuals who agreed to contribute to the current study, written informed consents were obtained in compliance with legislation. To identify novel genetic variants and to check their effect on the prevalence of MCPH in

Pakistan, (IV:2, IV:3, IV:6) individuals from family A were recognized and engaged. Affected individuals were inspected by medical doctors for clinical-assessment phenotypically and by taking head circumference. Family history was taken, and consanguine marriages were confirmed by several interviews and questionnaires.

### 2.1. Ethical compliance

To perform this study, approval has been acquired from the Ethics Committee, the Institutional Review Board of the, PU, Lahore (Reference # 1237-02). Data of the pretentious individuals were occupied and Photographs and pedigrees were processed in compliance with local legislation and the Helsinki declaration.

### 2.2. Family ascertainment and Sampling

Three individuals presenting the symptoms of MCPH of the family from the remote area of Sindh, Pakistan, were subjected to phenotypic examination by expert consultants. Medical information of patients comprising head-circumference, age, sex, mental retardation, and other associated disorders such as speech and motor delay was collected. Complete family history and written informed consent of the affected individuals were obtained from their guardians. Blood samples 5 ml were together in vacutainers from the affected individuals. The labeled vacutainers were hatched at 4-5°C until DNA-extraction.

### 2.3. DNA extraction

DNA has been extracted employing salting-out protocol reported by Helms [29]. The extent and class of extracted-DNA was evaluated via Agar-gel electrophoresis and the Nano-drop method using a spectrophotometer. DNA staining was performed using fluorescent dye ethidium bromide and DNA bands were obtained on UV gel transilluminator documentation system (using SynGene bioimaging system).

### 2.4. Mapping of *MCPH1* targets

Whole exome-sequencing was achieved on removed DNA samples of pretentious individuals with phenotypic abnormalities using Agilent v6 enrichment kit and Illumina HiSeq 4000 sequencing platform. For the alignment of the sequence reads human-genome build-37 was used for standard genome. A series of available algorithms together with Burrows-Wheeler aligner tool (BWA), SAM tools, PICARD, and the Genome Analysis Toolkit (GATK) was implemented for exome analysis in a sequential manner. The attained genomic data was sorted and prioritized for the gene variants via the VARBANK database and gene analysis tool kit previously reported protocols [30, 31].

### 2.5. Analysis of co-segregation

Mutations of *MCPH1* as a likely reason of MCPH related with medical symptoms in these affected individuals were endorsed by performing Sanger technique. The primers (forward-primer 5'CTGCATTTTGTCTACAGGTTTCA3' and reverse-primer 5'TGGTTCTGCGATCTGTGAAAA3') has been

intended for MCPH consuming Primer3 software. The bidirectional sequencing was performed on PCR products using kit (Big Dye terminator sequencing) version 3.1. The obtained data were analyzed by DNASTAR (lasergene).

## 2.6. In silico analysis

The reactivity of the identified mutations was forecast with the help of in silico analysis. The *in silico* studies were performed using different software and databases as follows:

### 2.6.1. Screening of NAD and MCPH targets

For selection of NAD-targets, these databases with different scoring standards that is PharmMapper-Server [32, 33] and Swiss-Target Prediction [34, 35] has been used. Targets with a Norm fit value of  $>0.5$  and  $>0$  in PharmMapper Server and SwissTargetPrediction, respectively, were selected as targets for NAD. The names of genes were unified using Uniprot and duplication was removed. The MCPH targets were accessed from Online Mendelian Inheritance in Man (OMIM) database [36] and GeneCards database [37]. The common targets of the NAD and MCPH were mapped using Venn diagrams. The common targets of the NAD and MCPH were analyzed using Jvenn [38].

### 2.6.2. Protein-protein interaction data analysis

The protein-protein interaction (PPI) data limited to humans was obtained from the Search Tool for Recurring Instances of Neighboring Genes (STRING) database [39]. Protein-protein interactions were predicted using STRING database that provide access to  $> 20$  billion predicted interactions of 67.6 million known proteins from 14094 organisms. The database retrieved from <https://string-db.org/> predict direct as well as indirect assessment of proteins such as physical and functional interactions. Three confidence level thresholds 0.4, 0.7, and 0.9 were used in the STRING database. For the construction of the protein interaction network (PIN), we selected the PPI data with a 0.7 confidence score.

### 2.6.3. MCODE, GO and KEGG pathway analysis

PPI information has been introduced to Cytoscape version 3.6.0. from the STRING catalogue and further managed, examined, and envisaged by removing detached nodes. A system analyzer-tool Cytoscape has been used to find the topology parameter of degree value Furthermore, the PPI-network of joint goals was analyzed by the Maximal Clique Centrality (MCC) to find out hub genes using CytoHubba plugin [40]. It was followed by the visualization of the connections between the hub genes, targets, and pathways. Molecular Complex Detection Algorithm (MCODE) was used to explore the protein interaction network [41]. Gene ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway analysis was performed using Database for Annotation, Visualization and Integrated Discovery (DAVID) (6.8) [42]. The

results were visualized using Dotchart and Bubble tools in the Hplot platform.

### 2.6.4. SNPs data analysis

Furthermore, The National Center of Biological Information (NCBI) SNP data base was used for the data collection and identification of the deleterious single nucleotide variations of gene sequences (SNPs). The MCPH1 gene variations along with the IDs were collected from NCBI. Eleven insilico damaging SNP prediction tools including Sorting Intolerant from Tolerant (SIFT) polymorphism phenotyping v2 (Polyphen-2), Combined Annotation-Dependent Depletion (CADD), MutationAssessor, REVEL and MetaR, were employed to identify and predict the most damaging SNP. The prediction of the SNPs using SIFT prediction tool uses PSIBlast algorithm [43] for the identification of the tolerated or altered substitution at every site of the sequence by considering sequence homology as well as physical properties of the amino acids [44, 45]. PolyPhen-2 utilizes the Naïve Bayes algorithm that analyze and determine the significance of the mutation or allele change on the functioning by using probabilistic classifier [46, 47]. Moreover, the prediction by PolyPhen-2 also considers the phylogenetic and structural properties, as well as number of sequences that characterize the mutant substitution [48, 49]. In addition to above CADD is an integrative annotation tool that based on the 60 genomic features, score the human single nucleotide mutants, insertions and deletions [50]. After scoring of the mutants it can prioritize and analyses casual variants and highly penetrant contributors to severe genetic disorders [50]. It also compute the weighted average of the scores obtained by SIFT, polyphen-2 and MutantAssessor [51]. Further, by obtaining different sequence alignments the functional consequences of the mutations were investigated using MutationAssessor based on the sequence conservation [52- 54]. PROVEAN is a prediction tool that analyze and predict the effect of variation in protein sequence on the functioning of protein [55, 56]. The functional effects if the nsSNPs were determined in the protein sequence by application of the delta alignment scores according to the reference of protein sequence and variant version [57]. I-mutant3.0 predict the stability change upon single-site mutation based on protein sequence and structure by DDC value (kcal/mol) using Support Vector Machine algorithm (SVM) [58]. Using I-mutant3.0. The DDG value along with the reliability index (RI value) of the mutant were calculated. SNAP2 is based on the neural network that estimates impact of mutation of the function of the protein by using the sequence alignment and structural features of the protein [59, 60].

### 2.6.5. Conservation analysis

The conservation analysis of the native MCPH1 gene was performed by the ConSurf web tool [61] which scrutinizes the evolutionary pattern of the amino acids or nucleic acids of the

macromolecular substitutions among homologous sequences to disclose the regions of structural and functional importance [62, 63]. To calculate the conservation scores from protein sequence data Bayesian calculation method was used. For further analysis only those nsSNPs related to MCPH1 were considered that were in highly conserved region. A conservation score 1-4 was considered as variable, while 5-6 and 7-9 scores were considered as intermediate and conserved scores, respectively.

#### 2.6.6 Structural analysis

The homology modeling of the native protein was performed using transcript of reference sequence using MCPH1 on Swiss Model Platform. The model has QMEAN z score of -0.38 and sequence identity of 99.49%. For performing homology modeling to generate a human MCPH1 protein we selected X-ray crystal structure of MCPH1 protein from Protein Data Bank having PDB code 3SZM as template. For the confirmation of the positions of the SNPs and 12 mutant protein models, UCSF Chimera platform was used which is a comprehensive tool developed for computation, visualization and analysis of molecular structure parameters. The structures of native and mutant variants were energy minimized by conjugate gradient mode for 3D structure optimization, using GROMACS force field as default in NOMAD-Ref server [64].

The RMSD and total H-bonds for each structure were accessed using UCSF Chimera platform. RMSD values were checked by superimposing native and mutant protein structures. Protein-protein complexes were visualized analyzed and processed by making intermolecular interaction maps using bioCOMPLEXES CONTACT MAPS (COCOMAPS) application [65]. The homology model serves as input file in PDB format. Using this analysis interactions between monomers, interaction interfaces on chains were observed.

Virtual molecular docking was performed using Autodock Vina (1.1.2) platform [66]. For the removal of ligands, verification of the protein structure and the removal of water molecules from protein structures obtained from Protein Data Bank (PDB) the PyMol software (1.7.2.1) was used [67, 68]. MCPH1 and MCPH1 mutants (ligands and proteins) were modified in PDBQT format by AutoDockTools (1.5.6) prior to completing docking methodology [69-70]. For MCPH1 and MCPH1 mutant the active pockets: grid center -32, 9, 75 and -34, 16, 54, were selected, respectively. The Lamarckian genetic algorithm was employed to for identical docking terms. Further, the affinity binding outcomes of less than 0 kcal/mol suggested the ligands binding to the receptors. While the affinity binding scores of less than equal to -5.0 kcal/mol indicated the ligand's effective clumping to the binding sites of receptors. To represent the docking outcomes UCSF Chimera platform was used.

### 3. RESULTS

#### 3.1. Clinical findings

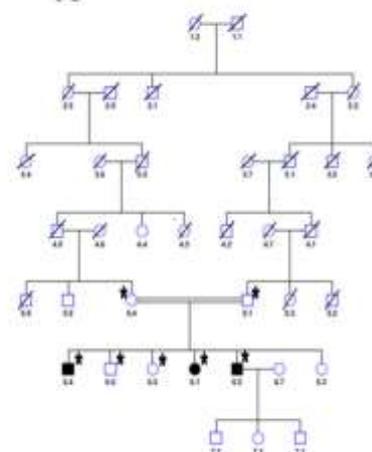
Sequence variants were mapped in Pakistani consanguineous family from Sindh province of Pakistan with three affected individuals (2 males and 1 female). Patients with reduced head circumference (3 SD) and mental health problem were recruited in accordance with the diagnostic criteria. Parents of the affected members of the family were normal with consanguineous marriage. An autosomal recessive pattern of inheritance was

B



observed from pedigree. Affected individuals of the family had microcephaly since birth. Pedigree and pictures of affected members of the family are represented (see Figure 1). Patients exhibited clinical features of microcephaly with a decreased head circumference of -8 to -11.4 SD. Clinical features of the ascertained affected individuals such as age, gender, head circumference (4 SD), mental abnormality, behavioral pattern, and any other linked disorder i.e. delay in speech etc. were noted.

A

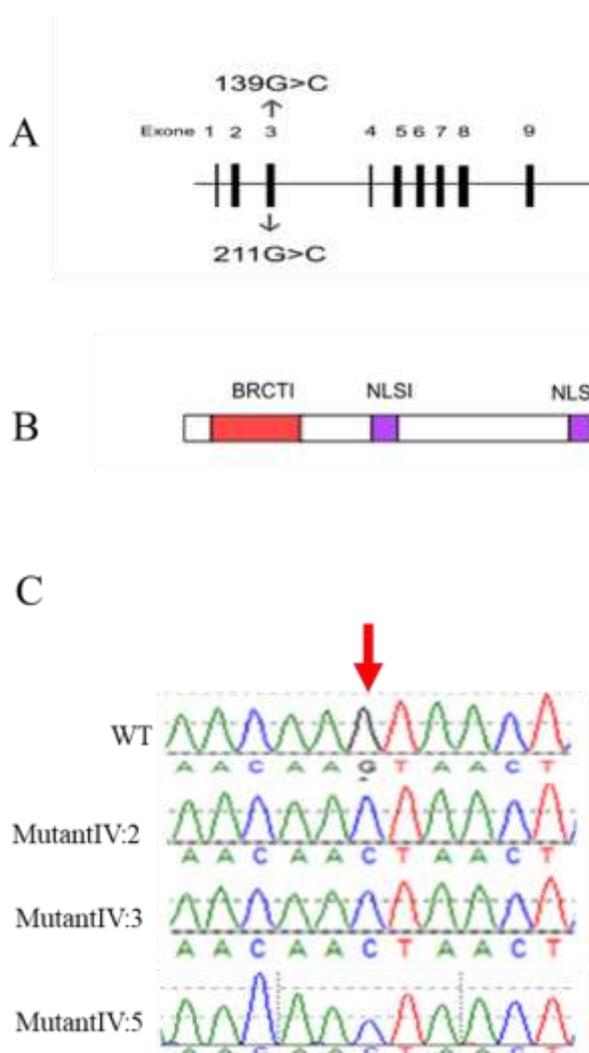


**Figure 1. (A)** Pedigree of the affected family **(B)** Pictures of the microcephaly affected individuals.

#### 3.2. Whole Exome Sequencing

After whole exome sequencing, data was filtered and analyzed for causative variants of

clinical phenotypes observed in patients. Two missense variants (NM\_024596.5: c.139G>C and NM\_024596.5: c.211G>C) were identified in the *MCPH1* gene responsible for symptoms of primary microcephaly. The gene is located on chromosome 8 in the chromosomal location of 8p23.1. These two variants in exon 3 of the transcript (NM\_024596.5) substitutes amino acid valine to leucine at position 47<sup>th</sup> and 71<sup>th</sup> of the translated protein product ([NP-00000342924](http://www.ncbi.nlm.nih.gov/clinvar/variant/NP/00000342924) : p. Val47Leu and p. Val71Leu) (see Figure 2). The pathogenicity score predicted by SIFT (<http://sift.jcvi.org>), PolyPhen-2 (<http://genetics.bwh.harvard.edu/pph2>) and CADD software in ClinVar (<https://cadd.gs.washington.edu/>) for (NM\_024596.3: c.139G>C) were found to be 0.05, 0.09 and 26 and for (NM\_024596.3: c.211G>C) scores were 0, 1 and 25 respectively posing deleterious effects of mutations on protein function.



**Figure 2.** (A) Exon and intron organization of *MCPH1* gene transcript. Position of mutations c.139G>C and c.211G>C is indicated within exon 3 of the transcript. (B) Location of functional domains is given with predicted MCPH1 protein. (C & D) Alignment of sequence chromatograms of individuals showing segregation of missense variant NM\_024596.3: c.139G>C and

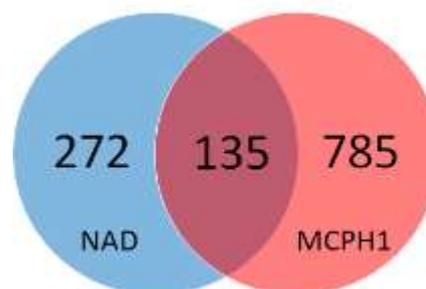
NM\_024596.3: c.211G>C (D) in affected and normal individuals (wild type) of the family.

### 3.2.1. Segregation analysis

The autosomal recessive mode of inheritance of the *MCPH1* gene variants (NM\_024596.5: c.139G>C and (NM\_024596.5: c.211G>C) in the affected family was confirmed by employing sanger sequencing. The results were compared with the reference sequence (NC\_000008.11) in GenBank. All affected family members of the family were homozygous for these two variants (see Figure 2C&D).

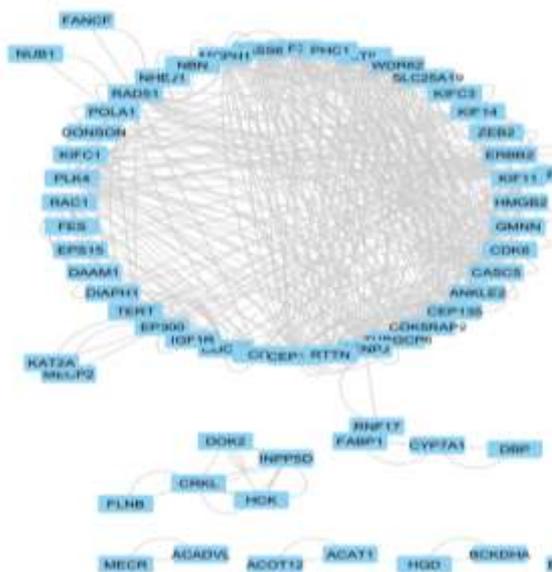
### 3.2.2. Targets of NAD and MCPH and PPI-network construction

NAD targets were obtained through PharmMapper Server and SwissTargetPrediction databases. After the removal of duplicates entities, 407 targets of NAD were obtained and 920 targets of MCPH were obtained from GeneCards and OMIM database. Further, 135 common targets of the NAD and MCPH were found using Jvenn. These common targets were considered as potential targets for further in silico processings. Figure 3 exhibited the targets of NAD and MCPH1. The overlapping region represents the common targets of both NAD and MCPH1.



**Figure 3.** Common targets of NAD and MCPH.

Common targets of NAD and MCPH1(135) were imported to the STRING database. By keeping the homo sapiens as the organisms of selection, the PPI data with 97 nodes and 532 edges was obtained at a confidence level of 0.70. Then, for analysis and visualization of the protein interaction network, PPI data was exported to Cytoscape software. The Cytoscape exhibited the network of all the genes that are involved (see Figure 4). Cluster analysis of PPI network showed the following statistics, Cluster 1; score 11.833, nodes 13, edges 146, Cluster 2: score 3.273, nodes 12, edges 36, Cluster 3: score.3.0, 3 nodes, 6 edges.



**Figure 4.** Cluster analysis of PPI network.

### 3.2.3. MCODE analysis

MCODE analysis was performed for functional analysis of PIN. In a network, MCODE locates clusters (regions with lots of connections). In our collected PPI data sets, we automatically forecast protein complexes. Clustering of the PPI network was done in 3 modules with three scores. For KEGG analysis, we selected three modules with a score  $\geq 3.0$ . The score of module 1 was 11.833 and the PIN contains 13 nodes and 146 edges. The score of module 2 was 3.273 and the PIN contains 12 nodes and 36 edges. The score of module 3 was 3.0 and the PIN contains 3 nodes and 6 edges. It uses the edge similarity centrality measure to identify complexes. Given that a greater percentage of shortest distance will run across edges that connect various complexes, these edges have higher centrality values. The edges of the networks are ranked using the edge similarity centrality scores, then the most centre edges are removed, and the process repeats until no edges were left. The edges that were removed are considered to constitute a single complex (see Figure 5).

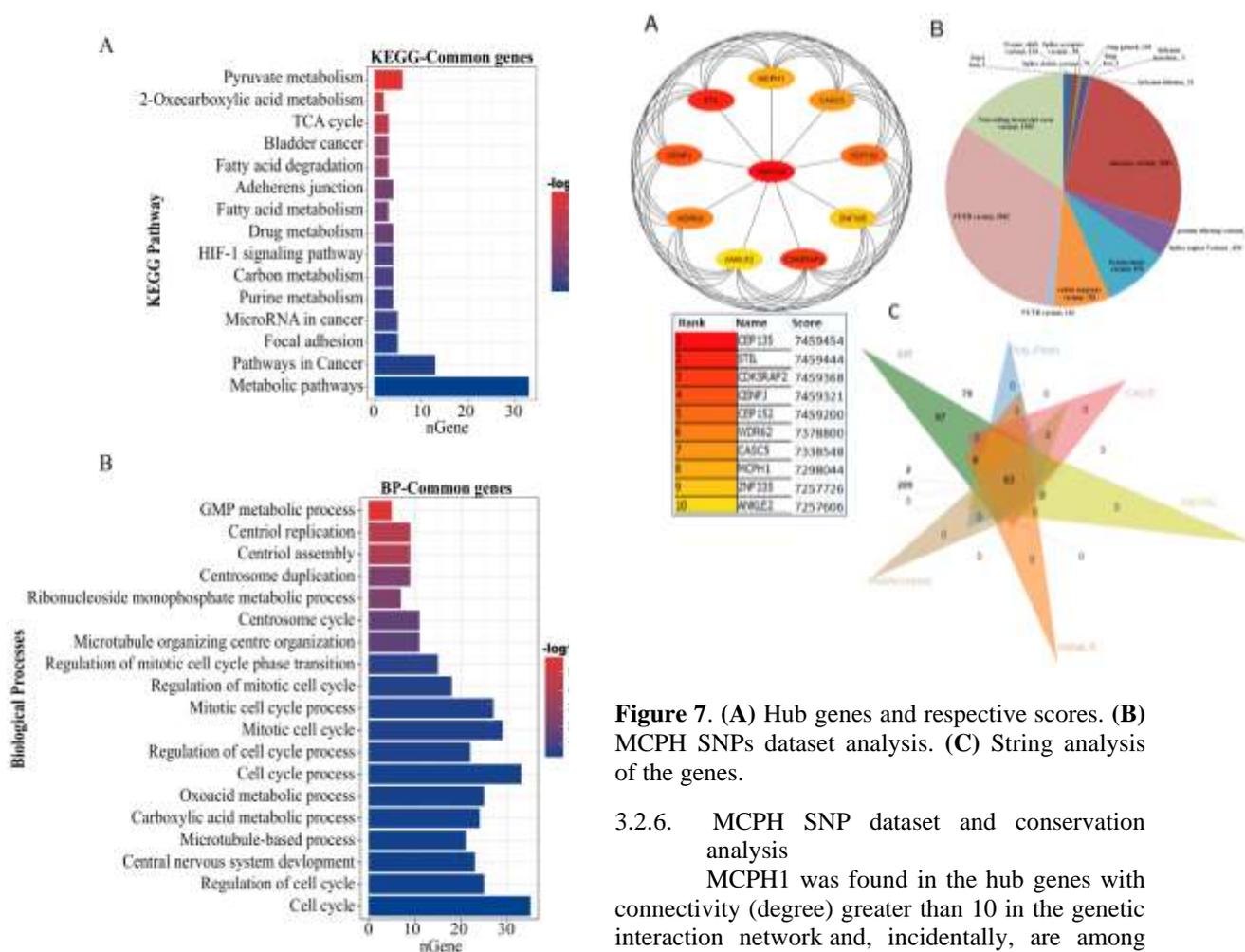


**Figure 5.** PPI network common targets.

### 3.2.4. GO terms and KEGG pathway enrichment analysis

The biological functions of the common targets were revealed by gene ontology enrichment analysis. 421 GO biological terms were obtained from the DAVID database and to visualize top GO and KEGG terms the slid chart diagrams were applied. The KEGG analysis shows the involvement of genes in some pathways such as pyruvate metabolism, 2-oxycarboxylic acid metabolism, TCA cycle, bladder cancer, fatty acid degradation, adherens junction, fatty acid metabolism, drug metabolism, HIF-1 signaling pathway, carbon metabolism, purine metabolism, microRNA in cancer, focal adhesion, pathways in cancer and metabolic pathways. However, the biological processes (BP) common gene analysis revealed that common targets were involved in the regulation of mitotic cell cycle phase transition, regulation of mitotic cell cycle and processes, oxoacid metabolic processes, carboxylic acid metabolic processes, microtubule based processes central nervous system development, regulation of cell cycle etc. While the cellular component (CC) and molecular function (MF) analysis indicated the significant involvement of common targets in Kinesin complex, centrioles, Azurophilic granule lumen, PML body, spindle pole, mitotic spindle, microbody, Peroxisome, extrinsic component of plasma membrane, spindle, centrosome, microtubule organizing center, mitochondrial matrix, membrane microdomain, membrane raft, microtubule cytoskeleton and nucleolus. These results indicated that common target genes are involved in several cellular and biological processes simultaneously (see Figure 6A-D).

technique will provide fresh perspectives on crucial regulatory networks and peptide drug targets.



**Figure 6.** (A) Analysis of GO and KEGG common genes in the network. (B) Analysis of GO and BP common genes in the network. (C) Analysis of GO and CC common genes in the network. (D) Analysis of GO and MF common genes in the network.

### 3.2.5. Identification of Hub genes

To analyze the network and identify the hub genes we used CytoHubba and we found a cluster of 10 hub genes with 45 edges. The following 10 hub genes with the respective scores were identified: CEP135 (7459454), STIL (7459444), CDK5RAP2 (7459368), CNEPJ (7459321), CEP152 (7459200), WDR62 (7378800), CASC5 (7338548), MCPH1 (7298044), ZNF335 (7257726), and ANKLE2 (7257606) (see Figure 7). The 11 topological analysis techniques offered by CytoHubba. It does the computations for all eleven ways in a single step. Additionally, researchers can create a novel analysis method by combining cytoHubba with other plugins. For experimental biologists, the networks and sub-networks captured by this topological analytic

**Figure 7.** (A) Hub genes and respective scores. (B) MCPH SNPs dataset analysis. (C) String analysis of the genes.

### 3.2.6. MCPH SNP dataset and conservation analysis

MCPH1 was found in the hub genes with connectivity (degree) greater than 10 in the genetic interaction network and, incidentally, are among the top 10% genes of highest connectivity. Thus, MCPH1 was selected for SNP analysis and total 8131 SNPs were identified in different regions of the MCPH1. Further missense mutations in MCPH1 were identified using homology sequence tools. Of the total SNPs 63 common missense pathogenic SNPs were found (see Figure 7B). From the identified 14 missense SNPs (R752G, A757V, D711G, A702V, K186E, D8A, R485W, V464G, P367Q, P92L, D8N, S102I, V47I, V71I, V47L, and V71L) 10 were found to have highly conserved region and were further processed for structural analysis. While the other 4 SNPs (R752G, R485W, P367Q, and S102I) were excluded because they were predicted with lower conservation score. In addition to that, MCPH1, V47L and V71L variants were included for structure stability analysis (see Figure 8).



Figure 8. Conservation analysis of MCPH1.

Structural analysis

3.2.7. Modeling of MCPH structure

We modeled 3D structure of human MCPH using X-ray crystal structure as template on Swiss model platform. The figure 9 showed the constructed model for MCPH. The generated model was further used to construct 12 mutant variants (see Figure 9).

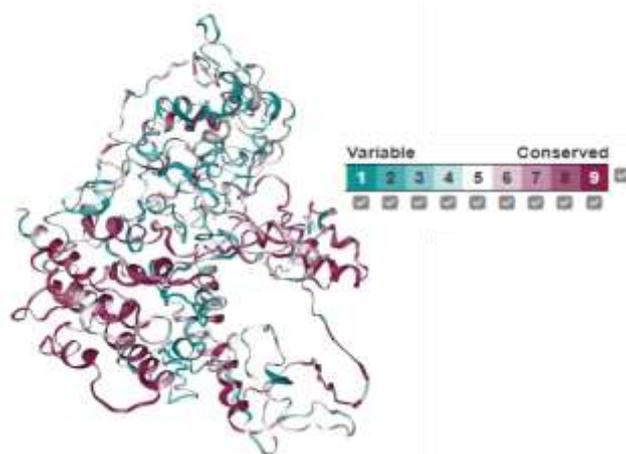


Figure 9. Structure of native MCPH protein.

3.2.8. RMSD and H-bond assessment

The RMS deviation score of the native protein as well as 12 mutant variants were accessed (as shown in Table 1). The higher deviation of mutant variants structures from native protein structure as indicated by increasing RMSD score, significantly change the activity of protein. The mutants V71L and P92L exhibited highest RMSD scores of 0.250 and 0.138 respectively. A757V, D711G, K186E, D8A and D8N showed the RMSD of 0.129. A702V and V47I exhibited the RMSD of 0.131. While V464G have the RMSD of 0.127.

Native protein showed H-bond count of 737. The mutant variants exhibited different number of H-bonds as compared to native protein. All the variants showed decreased number of H bonds than native protein. The lowest H-bonds were exhibited by V71I with the count of 698 (as shown in Table 1). A757V, D711G, and A702V have the H-bond count of 728, 723, and 732, respectively. However, K186E and D8A have the H-count of 721 while, V464G and D8N possessed 726 and 725 of this bond type. P92L and V47L, both have the H-count of 735. V47I and V71L have the count of 729 and 736 for H-bonds.

**Table 1:** RMSD and total number bonds of native protein and variants.

SR #	VARIANTS	RMSDA	H-BONDS
1	Native protein	0	737
2	A757V	0.129	728
3	D711G	0.129	723
4	A702V	0.131	732
5	K186E	0.129	721
6	D8A	0.129	724
7	V464G	0.127	726
8	P92L	0.138	735
9	D8N	0.129	725
10	V47I	0.131	729
11	V71I	0.000	698
12	V47L	0.136	735
13	V71L	0.250	736

Moreover, for the missense 3D analysis and identification of structural damage molecular modeling of the native and mutant variants was performed (as shown in Table 2). Among 12 mutant variants, structural damage was observed only in K186E. The substitution present in K186E leads to the shrinkage of the cavity volume by 102.6 Å<sup>3</sup> (as shown in Table 2, Entry 4).

**Table 2:** Missense 3D analysis, structural damage identification.

Sr #	Variants	WT	MT	Structural damage
1	A757V			Structural damage was not observed.
2	D711G			Structural damage was not observed.

3	A702V			Structural damage was not observed.
4	K186E			The substitution leads to the contraction of cavity volume by 102.6 Å <sup>3</sup>
5	D8A			Structural damage was not observed.
6	V464G			Structural damage was not observed.
7	P92L			Structural damage was not observed.
8	D8N			Structural damage was not observed.
9	V47I			Structural damage was not observed.

10	V71I			Structural damage was not observed.
11	V47L			Structural damage was not observed.
12	V71L			Structural damage was not observed.

### 3.3. Interaction analysis

Interaction analysis of native and mutant variants was accessed using COCOMAPS to evaluate the interactions and between amino acids and interfaces. A number of interaction parameters were evaluated such as number of interacting residues, hydrophilic-hydrophobic interactions, hydrophilic-hydrophilic interactions and hydrophobic-hydrophobic interactions. The 12 mutant proteins showed variably different interaction parameters as compared to native protein (as shown in Table 3).

The number of interacting residues observed in molecule 1 (as shown in Table 3, entry 2) and molecule 2 (as shown in Table 3, Entry 2) were 835, which were found same for the all mutated protein as that of native protein. In addition to above the native protein showed 7495 number of hydrophilic-hydrophilic interactions but the variants V47L showed higher value of 7597 and V71L showed the significantly lower value of 7301, respectively. Simultaneously, for hydrophobic-hydrophobic interactions native protein observed 3635 while the mutant complex V47L showed relatively lower value of 3509 and V71L showed comparatively higher value of 3690 interactions (as shown in Table 3). The simultaneous variation in number of hydrophilic-hydrophilic interactions and number hydrophobic-hydrophobic interactions showed that due to V47L and V71L mutations, conformational change in MCPH1 protein has occurred.

**Table 3:** Analysis of interactions between MCPH native protein and 12 mutant variants complexes.

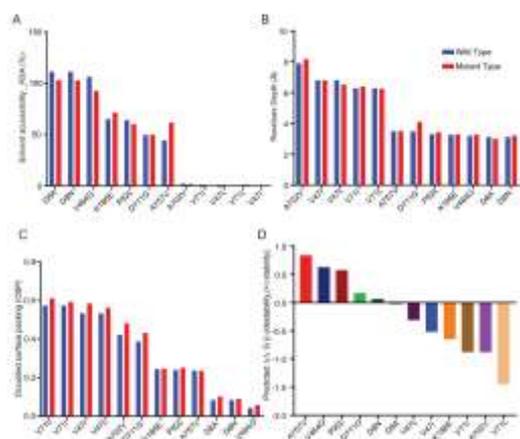
Interaction	Native	A757V	D711G	A702V	K186E	D88A	D88N	V464G	P92L	V47I	V71L	V71L
Number of interacting residues Molecule1	835	835	835	835	835	835	835	835	835	835	835	835
Number of interacting residues Molecule2	835	835	835	835	835	835	835	835	835	835	835	835
Number of hydrophilic-hydrophobic interaction	7447	6637	7451	7436	7368	7453	7444	7442	7472	6988	7507	7301
Number of hydrophobic-hydrophobic interaction	3635	3182	3629	3656	3658	3627	3622	3635	3652	3662	3652	3690

### 3.4. Effect of mutations on protein stability

The outcome of mutations on the steadiness of protein was retrieved employing structure-based tactics *via* Site-direct mutator. The amount of a specific residue's total area that is accessible to solvents in a protein is known as residual solvent accessibility, and it is correlated with the spatial arrangement and packaging of the residues. The highest solvent accessibility (>100 %) was shown

by D8A, D8N and V464G for wild-type MCPH1. These residues also showed high accessibility for mutant protein too ( $> = 100\%$ ). The lowest accessibility for both wild and mutant protein was shown by A720V, V71I, V47L, V71L and V47I.

The amount of a specific residue's total area that is accessible to solvents in a protein is known as residual solvent accessibility, and it is correlated with the spatial configuration and packaging of the residues. The results indicated that V47I (mutant). All the residues however, had the residue depth equal to the native protein. D711G (mutant), on the other hand, showed second maximum residue depth than any other residue (Figure 10). Similar to buried surface, occluded surface packaging (OSP) is more susceptible to packing geometry. The molecular surface's occluded interface bounds are extended outwards until they overlap the nearby van der Waals surface. The results displayed that the mutant protein's residues have more OSP than the native protein. That means, mutant residues are more sensitive to the structural geometry (Figure 10 (C)). The results at the end, displayed the stability of the protein upon mutation (pseudo $\Delta\Delta G$ ). The negative and positive predicted values show the instability and stability of the protein (Figure 10 (d)).



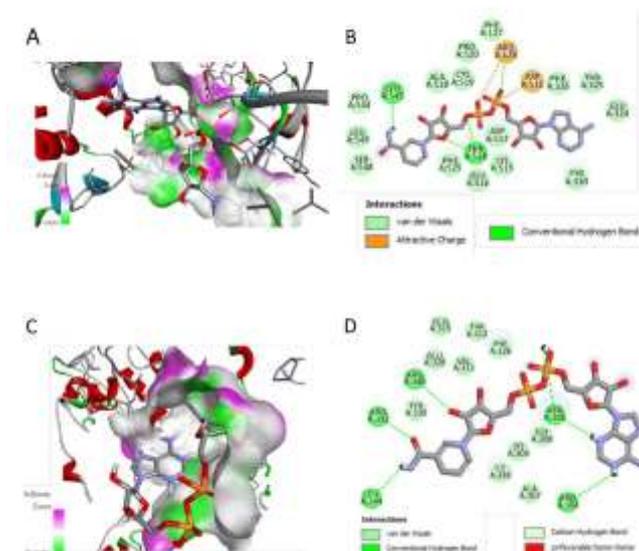
**Figure 10.** Effect of mutations on protein stability. The **A** graph represents the solvent accessibility of residues of both with and mutant-protein. The **B** graph represents the residual depth of both proteins. The **C** graph represents the occluded surface packing of residues of both with and mutant-protein. The graph **D** pseudo $\Delta\Delta G$  predictability of the residues of both with and mutant protein.

### 3.5. NAD Docking with MCPH1

Due to the identified pathogenic SNPs, the dynamic conformational changes in the structure of MCPH1 protein were observed. The effect of these modifications was further verified by using molecular docking approach taking MCPH1 wild type as well as mutant structure and docking it against NAD ligand. The binding energies of the MCPH1 wild type and mutant type with NAD ligand were found to be -8.1kcal/mol and -7.5kcal/mol, respectively. The docking

conformations of the MCPH1 wild type and mutant type with NAD ligand are shown in figure 11. The binding energy analysis showed that due to mutations in the structure of MCPH1, the binding affinity of the NAD with MCPH1 was reduced. This decrease in binding affinity indicates that SNPs have a significant pathological role in microcephaly disease.

Upon structural-analysis of the docked structure of wild and mutant-protein with ligand NAD, it was revealed that the ligand fit inside the binding pocket of both proteins (see Figure 11). That means that the proteins surfaces were accessible to the ligand which fit the ligand inside it to carry out the interactions. The docked complexes exhibited the interacting hypothesis for both wild and mutant MCPH1. For MCPH1-wild the interacting pattern were observed as ARG126, ASP531, LEU545, and THR528. Similarly, for MCPH1-mutant, the interacting residues were ASN305, LEU344, and ARG304. Here, it can be observed that translocation in the binding pocket occur which allow the NAD to interact to another binding site (see Figure 11).



**Figure 11.** (A) Molecular docking of NAD with MCPH1 wild. (B) 2D structural confirmation of NAD ligand with interacting residues. (C) 3D protein (MCPH1 mutant-ligand (NAD)) complex. (D) 2D structural confirmation of NAD ligand with interacting residues and MCPH1 mutant type.

## 4. DISCUSSION

Although Primary-microcephaly is exceptional disorder however it is still prevalent in the regions where consanguineous-marriages are collective. This genetic and nonprogressive neuro-developmental syndrome is considered by phenotypic-expressions of reduced-head size and occipitofrontal-circumference at the time of birth which is at times associated with varying degree of intellectual disability, and motor relay. This study was performed to find out the root causes of primary microcephaly inherited as autosomal

recessive disorder in Pakistani family with consanguineous marriage. In the current study for the identification of the responsible gene whole exome sequencing was performed which identified the two missense variants NM\_024596.5: c.139G>C and NM\_024596.5: c.211G>C located at chromosome 8 and exon 3 (8p23.1.). The pathogenicity scores of the identified variants predicted by SIFT revealed the deleterious effects if the mutations on the functioning of protein. Further, the Sanger-sequencing established the autosomal-recessive style of tradition of the variations of MCPH1 gene. We assembled the PIN of common-genes and screened-out 10 hub-genes, CEP135, STIL, CDK5RAP2, CNEPJ, CEP152, WDR62, CASC5, MCPH1, ZNF335, and ANKLE2. The GO and KEGG analysis revealed the pathogenicity of the MCPH1 mutants and involvement of numerous biological and cellular-processes in the diseases. The MCPH1 was found to have connectivity higher than 10 in the gene connectivity network thus it was subjected to the SNP analysis. From the total 8131 SNPs, 63 common missense pathogenic SNPs were identified. Furthermore, from the identified SNPs 10 missense SNPs were found with high conservation score and were processed for structural analysis. Among the 12 mutant variants structural damage was observed in K186E that results in the shrinking of the cavity volume. Due to the mutations V47L and V71L a significant variation in the hydrophobic and hydrophilic interactions was observed. Finally, molecular docking of the MCPH1 wild type and mutant protein against NAD ligands was performed for verification of the conformational changes, missense 3D analysis and identification of structural damage. The binding energy of the MCPH1 wild type with NAD ligand was found to be -8.1kcal/mol while the binding energy of the mutated MCPH1 with the NAD ligand was found to be -7.5kcal/mol. Hence it is obvious that mutations effected the binding potential of NAD ligand with the MCPH1 protein that indicates the significant pathological role of the MCPH1 mutations in the prevalence of the primary microcephaly disease. In the current study we used network pharmacology and molecular docking studies to predict and analyses the potential mutations and their effects on the function of the MCPH1 protein but these could be further verified and tested through biological studies.

## 5. CONCLUSION

Mutational analysis of non-syndromic primary microcephaly family of Pakistani origin revealed two novel variants in *MCPH1* gene with autosomal recessive mode of inheritance. Consanguineous marriage in the family is the root cause of the primary microcephaly. The reported genetic variants in *MCPH1* gene have significant pathogenic effect that has not been reported yet, thus adding to the mutational spectra of microcephaly genes.

## ACKNOWLEDGEMENTS

The authors would like to thank the affected patients and their families for their valuable participation in this study. Authors are thankful to Muhammad Ansar (Jules-Gonin Eye Hospital, Ophthalmology Department of University of Lausanne) for performing whole exome sequencing analysis.

## CONFLICTS OF INTEREST

The authors declare no conflict of interest.

## AUTHOR CONTRIBUTIONS

Saba Irshad and Zunaira Riaz designed the study. Zunaira Riaz acquired and analyzed the data and made a final draft of the study. Tanveer Ahmed collected the Family, prepared the figures and revised the manuscripts. Farman Ali performed the in silico analysis and revised the manuscript.

## ETHICAL COMPLIANCE

This study was approved by the Ethical Committee of the University of the Punjab, Lahore, Pakistan.

## INFORMED CONSENT

After the approval from the ethical committee, written informed consents from the parents of affected individuals was obtained for photograph and blood samples collection.

## DATA AVAILABILITY STATEMENT

Supporting data for the study will be available on request. Phenotypes of the individuals presented with MCPH are obscured due to ethical constraints and are attached as photographs in the supplementary data file.

## REFERENCES

- [1] Jayaraman, D., Bae, B.I., Walsh, C.A., 2018. The genetics of primary microcephaly. *Annu. Rev. Genomics. Hum. Genet.* 19, 177-200. doi: 10.1146/annurev-genom-083117-021441.
- [2] Faheem, M., Naseer, M.I., Rasool, M., Chaudhary, A.G., Kumosani, T.A., Ilyas, A.M., Pushparaj, P., Ahmed, F., Algahtani, H.A., Al-Qahtani, M.H., Saleh J.H., 2015. Molecular genetics of human primary microcephaly: an overview. *BMC Med. Genom.* 8, S4. <https://doi.org/10.1186/1755-8794-8-S1-S4>.
- [3] Asif, M., Abdullah, U., Nurnberg, P., Tinschert, S., Hussain, M. S., 2023. Congenital Microcephaly: A Debate on Diagnostic Challenges and Etiological Paradigm of the Shift from Isolated/Non Syndromic to Syndromic Microcephaly. *Cells.* 12(4), 642; <https://doi.org/10.3390/cells12040642>.
- [4] Muhammad, F., Mahmood Baig, S., Hansen, L., Sajid Hussain, M., Anjum Inayat, I., Aslam, M., Anver Qureshi, J., Toilat, M., Kirst, E., Wajid, M., Nürnberg, P., Eiberg, H., Tommerup, N., Kjaer, K. W., 2009. Compound heterozygous ASPM mutations in Pakistani MCPH families. *Am. J. Med. Genet. A.* 149, 926-930. <https://doi.org/10.1002/ajmg.a.32749>.
- [5] Gilmore, E.C., Walsh, C.A., 2013. Genetic causes of microcephaly and lessons for neuronal development. *Wiley Interdiscip. Rev. Dev. Biol.* 2, 461-478. <https://doi.org/10.1002/wdev.89>.
- [6] Ahmad, I., Baig, S.M., Abdulkareem, A.R., Hussain, M.S., Sur, I., Toliat, M.R., Nürnberg, G., Dalibor, N., Moawia, A., Waseem, S.S., Asif, M., Nagra, H., Sher, M., Khan, M., Hassan, I., Rehman, S.U., Thiele, H., Altmüller, J., Noegel, A.A., Nürnberg, P., 2017. Genetic heterogeneity in Pakistani microcephaly families revisited. *Clin. Genet.* 92, 62-68. <https://doi.org/10.1111/cge.12955>.
- [7] Woods, C.G., Bond, J., Enard W., 2005. Autosomal recessive primary microcephaly (MCPH): a review of clinical, molecular, and evolutionary findings. *Am. J. Hum Genet.* 76, 717-728. <https://doi.org/10.1086/429930>.
- [8] Siskos, N., Stylianopoulou, E., Skavdis, G., Grigoriou, M.E., 2021. Molecular Genetics of Microcephaly Primary Hereditary: An Overview. *Brain Sci.* 11, 581. <https://doi.org/10.3390/brainsci11050581>.
- [9] Farooq, M., Lindbæk, L., Krogh, N., Doganli, C., Keller, C., Mönnich, M., Gonçalves, A.B., Sakthivel, S., Mang, Y., Fatima, A., Andersen, V.S., Hussain, M.S., Eiberg, H., Hansen, L., Kjaer, K.W., Gopalakrishnan, J., Pedersen, L.B., Møllgård, K., Nielsen, H., Baig, S.M., Larsen, L.A., 2020. RRP7A links primary microcephaly to dysfunction of ribosome biogenesis, resorption of primary cilia, and neurogenesis. *Nat. Commun.* 11, 5816. <https://doi.org/10.1038/s41467-020-19658-0>
- [10] Zaqout, S., Morris-Rosendahl, D., Kaindl, A.M., 2017. Autosomal Recessive Primary Microcephaly (MCPH): An Update. *Neuropedia.* 48, 135-142. <https://doi.org/10.1055/s-0037-1601448>.
- [11] Naveed, M., Kazmi, S.K., Amin, M., Asif, Z., Islam, U., Shahid, K., Tehreem, S., 2018. Comprehensive review on the molecular genetics of autosomal recessive primary microcephaly (MCPH). *Genet. Res.* 100, e7. <https://doi.org/10.1017/S0016672318000046>.
- [12] Hou, J., van Leeuwen, J., Andrews, B.J., Boone, C., 2018. Genetic Network Complexity Shapes Background-Dependent Phenotypic Expression. *Trends Genet.* 34, 578-586. <https://doi.org/10.1016/j.tig.2018.05.006>.
- [13] Saleem, M., Alam, S., Kainat, I., Iftikhar, R., Ijaz, A., 2021. Single Cell Sequencing, Its Application, and Future Challenges. *Current Trends in OMICS.* 1, 56-75. <https://doi.org/10.32350/cto.11.05>.
- [14] Bilgüvar, K., Oztürk, A.K., Louvi, A., Kwan, K.Y., Choi, M., Tatli, B., Yalnizoğlu, D., Tüysüz, B., Çağlayan, A.O., Gökben, S., Kaymakçalan, H., Barak, T., Bakircioğlu, M., Yasuno, K., Ho, W., Sanders, S., Zhu, Y., Yilmaz, S., Dinçer, A., Johnson, M.H., Bronen, R.A., Koçer, N., Per, H., Mane, S., Pamir, M.N., Yalçinkaya, C., Kumandaş, S., Topçu, M., Ozmen, M., Sestan, N., Lifton, R.P., State, M.W., Günel, M., 2010. Whole-exome sequencing identifies recessive WDR62 mutations in severe brain malformations. *Nature*, 467, 207-210. <https://doi.org/10.1038/nature09327>.
- [15] Rahit, K., Tarailo-Graovac, M., 2020. Genetic Modifiers, and Rare Mendelian Disease. *Genes*, 11, 239. <https://doi.org/10.3390/genes11030239>.
- [16] Niemi, M.E., Martin, H.C., Rice, D.L., Gallone, G., Gordon, S., Kelemen, M., McAloney, K., McRae, J., Radford, E.J., Yu, S., 2018. Common genetic variants contribute to the risk of rare severe neurodevelopmental disorders. *Nature*, 562, 268-271. <https://doi.org/10.1038/s41586-018-0566-4>.
- [17] Poulton, C.J., Schot, R., Seufert, K., Lequin, M.H., Accogli, A., Annunzio, G.D., Villard, L., Philip, N., de Co, R., Catsman-Berrevoets, C., Grasshoff, U., Kattentidt-Mouravieva, A., Calf, H., de Vreugt-Gronloh, E., van Unen, L., Verheijen, F.W., Galjart, N., Morris-Rosendahl, D.J., Mancini, G.M. 2014. Severe presentation of WDR62 mutation: Is there a role for modifying genetic factors? *Am. J. Med. Genet. Part A.* 164A, 2161-2171. <https://doi.org/10.1002/ajmg.a.36611>.
- [18] Snedeker, J., Gibbons Jr, W.J., Paulding, D.F., Abdelhamed, Z., Prows, D.R., Stottmann, R.W.,

2019. Gpr63 is a modifier of microcephaly in Ttc21b mouse mutants. *PLoS Genet.* 15, e1008467. <https://doi.org/10.1371/journal.pgen.1008467>.
- [19] Yu, T.W., Mochida, G.H., Tischfield, D.J., Sgaier, S.K., Flores-Sarnat, L., Sergi, C.M., Topçu, M., McDonald, M.T., Barry, B.J., Felie, J.M., Sunu, C., Dobyns, W.B., Folkerth, R.D., Barkovich, A.J., Walsh, C.A., 2010. Mutations in WDR62, encoding a centrosome associated protein, cause microcephaly with simplified gyri and abnormal cortical architecture. *Nat. Genet.* 42, 1015-1020. <https://doi.org/10.1038/ng.683>.
- [20] Jackson, A.P., Eastwood, H., Bell, S.M., Adu, J., Toomes, C., Carr, I.M., Roberts, E., Hampshire, D.J., Crow, Y.J., Mighell, A.J., Karbani, G., Jafri, H., Rashid, Y., Mueller, R.F., Markham, A.F., Woods, C.G., 2002. Identification of microcephalin, a protein implicated in determining the size of the human brain. *Am. J. Hum. Genet.* 71, 136-142. <https://doi.org/10.1086/341283>.
- [21] Sahu, M.R., Mondal, A.C., 2021. Neuronal Hippo signaling: From development to diseases. *Dev. Neurobiol.* 81, 92-109. <https://doi.org/10.1002/dneu.22796>.
- [22] Iftikhar, R., Zahoor, A.F., Irfan, M., Rasul, A., Rao, F., 2021. Synthetic Molecules Targeting YAP Activity as Chemotherapeutics against Cancer. *Chem. Biol. Drug. Des.* 98, 1025-1037. <https://doi.org/10.1111/cbdd.13960>.
- [23] Schellino, R., Boido, M., Vercelli, A., JNK Signaling Pathway Involvement in Spinal Cord Neuron Development and Death. *Cells* 2019, 8, 1576. <https://doi.org/10.3390/cells8121576>.
- [24] Xu, X., Lee, J., Stern, D.F., 2004. Microcephalin is a DNA damage response protein involved in regulation of CHK1 and BRCA1. *J. Biol. Chem.* 279, 34091-34094. <https://doi.org/10.1074/jbc.C400139200>.
- [25] Yi, Y.G., Lee, D.W., Kim, J., Jang, J.H., Lee, S.M., Jang, D.H., 2019. Two novel mutations (c. 883-4\_890del and c. 1684C> G) of WDR62 gene associated with autosomal recessive primary microcephaly: a case report. *Front. Pediatr.* 7, 457. <https://doi.org/10.3389/fped.2019.00457>.
- [26] Barbelanne, M., Tsang, W.Y., 2014. Molecular and cellular basis of autosomal recessive primary microcephaly. *Biomed. Res. Int.* 2014, 547986. <https://doi.org/10.1155/2014/547986>.
- [27] Cox, J., Jackson, A.P., Bond, J., Woods, C.G., 2006. What primary microcephaly can tell us about brain growth? *Trends. Mol. Med.* 12, 358-66. <https://doi.org/10.1016/j.molmed.2006.06.006>.
- [28] Gruber, R., Zho, Z., Sukchev, M., Joerss, T., Frappart, P.O., Wang, Z.Q., 2011. MCPH1 regulates the neuroprogenitor division mode by coupling the centrosomal cycle with mitotic entry through the Chk1-Cdc25 pathway. *Nat. Cell. Biol.* 13, 1325-1334. <https://doi.org/10.1038/ncb2342>.
- [29] Helms, C., 1990. Salting out Procedure for Human DNA extraction. The Donis-Keller Lab-Lab Manual Homepage.
- [30] Ansar, M., Riazuddin, S., Sarwar, M.T., Makrythanasis, P., Paracha, S.A., Iqbal, Z., Khan, J., Assir, M.Z., Hussain, M., Razzaq, A., Polla, D.L., Taj, A.S., Holmgren, A., Batool, N., Misceo, D., Iwaszkiewicz, J., de Brouwer, A.P.M., Guipponi, M., Hanquinet, S., Antonarakis, S.E., 2018. Biallelic variants in LINGO1 are associated with autosomal recessive intellectual disability, microcephaly, speech and motor delay. *Genet. Med.* 20, 778-784. <https://doi.org/10.1038/gim.2017.113>.
- [31] Riazuddin, S., Hussain, M., Razzaq, A., Iqbal, Z., Shahzad, M., Polla, D. L., Song, Y., van Beusekom, E., Khan, A. A., Tomas-Roca, L., Rashid, M., Zahoor, M. Y., Wissink-Lindhout, W. M., Basra, M. A. R., Ansar, M., Agha, Z., van Heeswijk, K., Rasheed, F., Van de Vorst, M., ...Riazuddin, S., 2020. Correction: Exome sequencing of Pakistani consanguineous families identifies 30 novel candidate genes for recessive intellectual disability. *Mol. Psychiatry*, 25, 3101-3102. <https://doi.org/10.1038/s41380-018-0128-z>
- [32] Wang, X., Shen, Y., Wang, S., Li, S., Zhang, W., Liu, X., Lai, L., Pei, J., Li, H., 2017. PharmMapper 2017 update: a web server for potential drug target identification with a comprehensive target pharmacophore database. *Nucleic. Acids. Res.* 45, W356-W360. <https://doi.org/10.1093/nar/gkx374>.
- [33] Liu, X., Ouyang, S., Yu, B., Liu, Y., Huang, K., Gong, J., Zheng, S., Li, Z., Li, H., Jiang, H., 2010. PharmMapper server: a web server for potential drug target identification using pharmacophore mapping approach. *Nucleic. Acids. Res.* 38, W609-W614. <https://doi.org/10.1093/nar/gkq300>.
- [34] Gfeller, D., Michielin, O., Zoete, V., 2013. Shaping the interaction landscape of bioactive molecules. *Bioinformatics*, 29, 3073-3079. <https://doi.org/10.1093/bioinformatics/btt540>.
- [35] Daina, A., Michielin, O., Zoete, V., 2019. SwissTargetPrediction: updated data and new features for efficient prediction of protein targets of small molecules. *Nucleic. Acids. Res.* 47, W357-W364. <https://doi.org/10.1093/nar/gkz382>.
- [36] Hamosh, A., Scott, A.F., Amberger, J.S., Bocchini, C.A., McKusick, V.A., 2005. Online Mendelian Inheritance in Man (OMIM), a knowledgebase of human genes and genetic disorders. *Nucleic. Acids. Res.* 33, D514-D517. <https://doi.org/10.1093/nar/gki033>.

- [37] Stelzer, G., Rosen, N., Plaschkes, I., Zimmerman, S., Twik, M., Fishilevich, S., Stein, T.I., Nudel, R., Lieder, I., Mazor, Y., Kaplan, S., Dahary, D., Warshawsky, D., Guan-Golan, Y., Kohn, A., Rappaport, N., Safran, M., Lancet, D., 2016. The GeneCards Suite: From Gene Data Mining to Disease Genome Sequence Analyses. *Curr. Protoc. Bioinformatics.* 54, 1-33. <https://doi.org/10.1002/cpbi.5>.
- [38] Bardou, P., Mariette, J., Escudie, F., Djemiel, C., Klopp, C., 2014. Jvenn: an interactive Venn diagram viewer. *BMC Bioinformatics*, 15, 293. <https://doi.org/10.1186/1471-2105-15-293>.
- [39] Szklarczyk, D., Gable, A. L., Lyon, D., Junge, A., Wyder, S., Huerta-Cepas, J., Simonovic, M., Doncheva, N.T., Morris, J.H., Bork, P., Jensen, L.J., Mering, C.V., 2019. STRING v11: protein-protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. *Nucleic Acids. Res.* 47, D607-D613. <https://doi.org/10.1093/nar/gky1131>.
- [40] Chin, C.H., Chen, S.H., Wu, H.H., Ho, C. W., Ko, M.T., Lin, C.Y., 2014. cytoHubba: identifying hub objects and sub-networks from complex interactome. *BMC Syst. Biol.* 8, S11. <https://doi.org/10.1186/1752-0509-8-S4-S11>.
- [41] Bader, G.D. Hogue, C.W. 2003. An automated method for finding molecular complexes in large protein interaction networks. *BMC Bioinformatics*. 4, 2. <https://doi.org/10.1186/1471-2105-4-2>.
- [42] Huang da, W., Sherman, B.T., Lempicki, R.A., 2009. Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. *Nucleic Acids. Res.* 37, 1-13. <https://doi.org/10.1093/nar/gkn923>.
- [43] Kumar, P., Henikoff, S., Ng, P.C., 2009. Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm. *Nat. Protoc.* 4, 1073-1081. <https://doi.org/10.1038/nprot.2009.86>.
- [44] Dakal, T.C., Kala, D., Dhiman, G., Yadav, V., Krokhotin, A., Dokholyan, N.V., 2017. Predicting the functional consequences of nonsynonymous single nucleotide polymorphisms in IL8 gene. *Sci. Rep.* 7, 6525. <https://doi.org/10.1038/s41598-017-06575-4>.
- [45] Ng, P.C., Henikoff, S., 2006. Predicting the effects of amino acid substitutions on protein function. *Annu. Rev. Genom. Hum. Genet.* 7, 61-80. <https://doi.org/10.1146/annurev.genom.7.080505.115630>.
- [46] Ali Mohamoud, H.S., Manwar Hussain, M.R., El-Harouni, A.A., Shaik, N.A., Qasmi, Z.U., Merican, A.F., Baig, M., Anwar, Y., Asfour, H., Bondagii, N., Al-Aama, J.Y., 2014. First Comprehensive in Silico Analysis of the Functional and Structural Consequences of SNPs in Human GalNAc-T1 Gene. *Comput. Math. Methods Med.* 2014, 904052. <https://doi.org/10.1155/2014/904052>.
- [47] Adzhubei, I., Jordan, D.M., Sunyaev, S.R., 2013. Predicting functional effect of human missense mutations using PolyPhen-2. *Curr. Protoc. Hum. Genet.* 76. <https://doi.org/10.1002/0471142905.hg0720s76>.
- [48] Nailwal, M., Chauhan, J.B., 2017. In silico analysis of non-synonymous single nucleotide polymorphisms in human DAZL gene associated with male infertility. *Syst. Biol. Reprod. Med.* 2017, 63, 248-258. <https://doi.org/10.1080/19396368.2017.1305466>.
- [49] Rentzsch, P., Witten, D., Cooper, G.M., Shendure, J., Kircher, M., 2019. CADD: Predicting the deleteriousness of variants throughout the human genome. *Nucleic Acids Res.* 47, D886-D894. <https://doi.org/10.1093/nar/gky1016>.
- [50] Adzhubei, I.A., Schmidt, S., Peshkin, L., Ramensky, V.E., Gerasimova, A., Bork, P., Kondrashov, A.S., Sunyaev, S.R., 2010. A method and server for predicting damaging missense mutations. *Nat. Methods.* 7, 248-249. <https://doi.org/10.1038/nmeth0410-248>.
- [51] Ali, H.M., 2019. Identification and analysis of pathogenic nsSNPs in human LSP1 gene. *Bioinformatics*, 15, 621-626. <https://doi.org/10.6026/97320630015621>.
- [52] Hepp, D., Gonçalves, G.L., de Freitas, T.R.O., 2015. Prediction of the damage-associated non-synonymous single nucleotide polymorphisms in the human MC1R gene. *PLoS ONE*, 10, e0121812. <https://doi.org/10.1371/journal.pone.0121812>.
- [53] Reva, B., Antipin, Y., Sander, C., 2011. Predicting the functional impact of protein mutations: Application to cancer genomics. *Nucleic Acids Res.* 2011, 39, e118.
- [54] Frousios, K., Iliopoulos, C.S., Schlitt, T., Simpson, M.A., 2013. Predicting the functional consequences of non-synonymous DNA sequence variants—Evaluation of bioinformatics tools and development of a consensus strategy. *Genomics.* 102, 223-228. <https://doi.org/10.1016/j.ygeno.2013.06.005>.
- [55] Choi, Y., Chan, A.P., 2015. PROVEAN web server: A tool to predict the functional effect of amino acid substitutions and indels. *Bioinformatics.* 31, 2745-2747. <https://doi.org/10.1093/bioinformatics/btv195>.
- [56] Choi, Y., Sims, G.E., Murphy, S., Miller, J.R., Chan, A.P., 2012. Predicting the functional effect of amino acid substitutions and indels. *PLoS ONE.* 7, e46688. <https://doi.org/10.1371/journal.pone.0046688>.
- [57] Mustafa, H.A., Albkrye, A.M.S., AbdAlla, B.M., Khair, M.A.M., Abdelwahid, N., Elnasri, H.A., 2020. Computational determination of human PPAR $\gamma$  gene: SNPs and prediction of their effect on protein functions of diabetic patients. *Clin. Transl. Med.* 9, 7. <https://doi.org/10.1186/s40169-020-0258-1>.

- [58] Zhang, M., Huang, C., Wang, Z., Lv, H., Li, X., 2020. In silico analysis of non-synonymous single nucleotide polymorphisms (nsSNPs) in the human GJA3 gene associated with congenital cataract. *BMC Mol. Cell Biol.* 21, 12. <https://doi.org/10.1186/s12860-020-00252-7>.
- [59] Bromberg, Y., Yachdav, G., Rost, B., 2008. SNAP predicts effect of mutations on protein function. *Bioinformatics.* 24, 2397-2398. <https://doi.org/10.1093/bioinformatics/btn435>.
- [60] Saleh, M.A., Solayman, M., Paul, S., Saha, M., Khalil, M.I., Gan, S.H., 2016. Impacts of Nonsynonymous Single Nucleotide Polymorphisms of Adiponectin Receptor 1 Gene on Corresponding Protein Stability: A Computational Approach. *Biomed. Res. Int.* 2016, 9142190. <https://doi.org/10.1155/2016/9142190>
- [61] Ashkenazy, H., Abadi, S., Martz, E., Chay, O., Mayrose, I., Pupko, T., Ben-Tal, N., 2016. ConSurf 2016: an improved methodology to estimate and visualize evolutionary conservation in macromolecules. *Nucleic acids res.* 44, W344-W350. <https://doi.org/10.1093/nar/gkw408>.
- [62] Glaser, F., Pupko, T., Paz, I., Bell, R.E., Bechor-Shental, D., Martz, E., Ben-Tal, N., 2003. ConSurf: Identification of functional regions in proteins by surface-mapping of phylogenetic information. *Bioinformatics.* 19, 163-164. <https://doi.org/10.1093/bioinformatics/19.1.163>.
- [63] Ashkenazy, H., Abadi, S., Martz, E., Chay, O., Mayrose, I., Pupko, T., Ben-Tal, N., 2016. ConSurf 2016: An improved methodology to estimate and visualize evolutionary conservation in macromolecules. *Nucleic Acids Res.* 44, W344-W350. <https://doi.org/10.1093/nar/gkw408>.
- [64] Lindahl, E., Azuara, C., Koehl, P., Delarue, M., 2006. NOMAD-Ref: visualization, deformation and refinement of macromolecular structures based on all-atom normal mode analysis. *Nucleic. Acids. res.*, 34, W52-W56. <https://doi.org/10.1093/nar/gkl082E>.
- [65] Vangone, A., Spinelli, R., Scarano, V., Cavallo, L., Oliva, R., 2011. COCOMAPS: a web application to analyze and visualize contacts at the interface of biomolecular complexes. *Bioinformatics.* 27, 2915-2916. <https://doi.org/10.1093/bioinformatics/btr484>.
- [66] Trott, O., Olson, A.J., 2010. AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. *J. Comput. Chem.* 31, 455-461. <https://doi.org/10.1002/jcc.21334>.
- [67] Seeliger, D., de Groot, B.L., 2010. Ligand docking and binding site analysis with PyMOL and Autodock/Vina. *J. Comput. Aided. Mol. Des.* 24, 417-422. <https://doi.org/10.1007/s10822-010-9352-6>.
- [68] Ali, F., Yao, Z., Li, W., Sun, L., Lin, W., Lin, X., 2018. In-silico prediction and modeling of the quorum sensing LuxS protein and inhibition of AI-2 biosynthesis in *Aeromonas hydrophila*. *Molecules.* 23, 2627. <https://doi.org/10.3390/molecules23102627>.
- [69] Morris, G.M., Huey, R., Olson, A.J., 2008. Using autodock for ligand-receptor docking. *Curr. Protoc. Bioinformatics,* 24, 8. <https://doi.org/10.1002/0471250953.bi0814s24>.
- [70] Lill, M.A., Danielson, M.L., 2011. Computer-aided drug design platform using PyMOL. *J. Comput. Aided. Mol. Des.* 25, 13-19. <https://doi.org/10.1007/s10822-010-9395-8>.

College of Pharmacy, University of the Punjab, Lahore, Pakistan

**Fourth Author** –Farman Ali, Clinical Research Institute, the Second Affiliated Hospital &Academy of Integrative Medicine, Fujian University of Traditional Chinese Medicine, Fuzhou, Fujian350122, China

**Correspondence Author** – Saba Irshad, PhD, School of Biochemistry and Biotechnology, University of the Punjab, 54000-Lahore, Pakistan

#### AUTHORS

**First Author** – Zunaira Riaz, PhD Scholar, School of Biochemistry and Biotechnology, University of the Punjab, 54000-Lahore, Pakistan

**Second Author** – Saba Irshad, PhD, School of Biochemistry and Biotechnology, University of the Punjab, 54000-Lahore, Pakistan

**Third Author** – Tanveer Ahmed, Laboratory of Metabolism and Cell Fate, Guangzhou Institutes of Biomedicine and Health, Chinese Academy of Sciences, 510530 Guangzhou, China and Cadson