Original Article

Integrative Whole Exome Sequencing and Machine Learning Analysis of MCPH1, ERCC2, CENPJ, and ERCC6 Variants in Pakistani Families with Primary Micro-cephaly

Bilal Ahmad,[1] Muhammad Qasim,[2] Muhammad Asif,[3] Muhammad Tariq[4]

[1-3]Department of Bioinformatics & Biotechnology, Government College University Faisalabad, Faisalabad, Pakistan; [4]National Institute for Biotechnology and Genetic Engineering College, Pakistan Institute of Engineering and Applied Sciences (NIBGE-C, PIEAS), Faisalabad, Pakistan

Abstract

Background: Primary microcephaly (MCPH) is a genetically and clinically diverse condition characterized by small head size, structural brain abnormalities, and non-progressive intellectual impairment. To date, variations in 30 genes have been associated with MCPH.

Objective: The study aims to identify the genetic variants of MCPH in the Pakistani population, where consanguineous marriages are common, and to explore the functional relationship of MCPH with other neurodevelopmental disorders (NDDs) such as autism spectrum disorder (ASD), intellectual disability (ID), and developmental delay (DD).

Methods: Whole-exome sequencing (WES) and Sanger sequencing were applied to identify genetic variants in MCPH patients. The functional relationship between MCPH and other NDD genes was explored using DGH-GO software, employing hierarchical clustering. This cross-sectional study was conducted from September 2023 to October 2024.

Results: We identified two novel variants, ERCC2 (c.2255G>A) and ERCC6 (c.1178C>T), and two already reported variants, MCPH1 (c.1254delT) and CENPJ (c.18delC). Machine learning analysis revealed a significant functional overlap between MCPH and other neurodevelopment disorders (NDDs), affected genes.

Conclusion: Our study expands the mutational spectrum of MCPH and supports shared genetic etiology between MCPH and other NDDs. These findings provide a deeper understanding of the genetic underpinnings and comorbidities of MCPH, guiding future research toward effective therapeutic strategies.

Received: 08-04-2025 | **1st Revision:** 04-08-2025 | **2nd Revision:** 10-09-2025 | **Accepted:** 29-09-2025

Corresponding Author | Dr. Muhammad Qasim, Professor, Department of Bioinformatics & Biotechnology, Government College University Faisalabad, Faisalabad, Pakistan **Email:** qasemawan@gmail.com

Keywords | Primary microcephaly; Neurological disorders; Pakistani families; Whole-exome sequencing; Sanger sequencing; Machine learning.

How to cite: Ahmad B, Qasim M. Asif M. Tariq M. Integrative Whole Exome Sequencing and Machine Learning Analysis of MCPH1, ERCC2, CENPJ, and ERCC6 Variants in Pakistani Families with Primary Micro-cephaly. Ann King Edw Med Univ.2025;31(3): 347-354

Introduction

Microcephaly (MCPH) is a genetically heterogeneous disorder; at least 28 different genes have already been implicated in the etiology of MCPH, most of which are expressed in the cerebral cortex during the proliferation of neural precursor cells (NPCs).[1] Different variants of MCPH have been reported to co-occur with other developmental disorders. Frequently reported comorbidities associated with MCPH are autism spectrum disorder (ASD),[2] intellectual disability (ID), and developmental delays (DD), such as problems in walking, speech, and performing daily life tasks, have also

Production and Hosting by KEMU

https://doi.org/10.21649/akemu.v31i3.6167
2079-7192/

been observed in affected individuals.

Data from genomics and other omics can provide useful clues for understanding the biological mechanisms underlying clinically relevant phenotypes.[3] Instead of labeled data, analysis of the unlabeled data can offer a more effective approach. This will not only overcome the limitation of using labels, but it will also save the cost of annotation.[4] This approach is called unsupervised machine learning. We applied it to gene variants underlying microcephaly to identify functional overlap between microcephaly and other neurodevelopmental disorders, such as autism spectrum disorder, intellectual disability, and developmental delay.

We employed whole-exome sequencing (WES) to identify causative gene variants in four families affected with MCPH. Two novel and two previously reported variants were identified in four genes (ERCC2, ERCC6, MCPH1, and CENPJ) to cause MCPH. Our findings suggest that MCPH genes have overlapping functions with genes of other NDDs. Our results indicate that ERCC2 and ERCC6 are also functionally related to genes involved in multiple NDDs.

Despite the established etiology of primary microcephaly, the intersection between neurodevelopmental disorder (NDD) genes and microcephaly primary hereditary (MCPH)-related genes within the Pakistani cohort remains inadequately investigated. This study seeks to address this deficiency by employing whole-exome sequencing (WES) coupled with machine learning techniques to examine common genetic patterns.

Methods

This was a cross-sectional study in which four unrelated consanguineous families, each with multiple affected individuals, segregating primary microcephaly, were investigated between September 2023 to October 2024. Patients were selected based on diagnostic criteria such as small head circumference (HC), speech impairments, seizures, intellectual disability, and developmental delay. After obtaining written informed consent from parents/guardians, pedigrees were drawn with the help of elders in each family. Affected individuals were physically examined, their head circumference was measured, and all the clinical features were recorded.

The study was formally approved by the Institutional Review Boards (IRBs) of Government College University, Faisalabad, Pakistan (Ref. No. GCUF/ERC/307). All participants and their legal guardians provided written informed consent.

Whole-exome sequencing (WES) was performed on DNA samples from affected patients in each pedigree using the Illumina NovaSeq 6000 platform with a mean coverage of 100x. The results from the sequencing were aligned to the human reference genome assembly (GRCh37). The Genome Analysis Tool Kit (GATK, version 3.7) was used to identify the variants. Variants were filtered after they had been annotated using online sources such as the 1,000 Genomes Project and Genome Aggregation Database (gnomAD). Minor allele frequency (MAF) > 0.005 was used to exclude all variations. Homozygous and compound heterozygous variants with autosomal recessive inheritance were retained for further analysis.

Candidate variants were Sanger sequenced in the probands, and validated variants were sequenced in all the available individuals to confirm if the variants were segregating with MCPH. BigDye terminator sequencing chemistry was employed for Sanger sequencing on the Genetic Analyzer 3730 (Applied Biosystems, Foster City, CA, USA). Chromatograms were visualized using Sequencher 5.4.6 (Gene Codes Corporation).

Our in-house developed software, DGH-GO, was employed to calculate the functional similarities between the genes causing multiple disorders.[5] DGH-GO is a web application that includes unsupervised machine learning methods that may be employed to delineate the genetic heterogeneity of complex diseases. DGH-GO can be utilized to assess the shared etiology and commonalities between different complex disorders. An optimum number of clusters and validation of clustering results were defined by the Silhouette measure. Enrichr tool[6] was employed to find the enriched pathways and disease terms for the identified clusters. Enrichr is a web-based tool that allows users to analyze genes and discover their associated functional annotations, pathways, and gene ontology terms. It creates a ranked list of significantly enriched terms for input genes.

Results

We examine four Pakistani consanguineous families, one from South Waziristan and three from rural areas of Punjab, Pakistan. Patients were chosen for this study if they had met the diagnostic criteria, which include reduced head circumference, intellectual disability, and dysmorphic facial features. All families have shown an autosomal recessive mode of inheritance. The clinical profiles of the affected families are summarized in Table 1, and their pedigrees are shown in Figure 1.

Present = +, Absent = -; Intellectual Disability: (+) =

Table 1: *Clinical profiles of affected members of families A, B, C, and D.*

Family No.	Ethnic origin	Gene	Mutation	Family ID & Gender	Age (years)	Speaking Age (years)	HC (cm)	Intellectual Disability level	Hearing Impairment	Facial Dysmorphism	Seizure	Ocular anomalies
Family A	South Waziristan	*ERCC2*	c.2255G>A	IV:2 (M)	20	2	42	+	-	+	-	-
				IV:4 (M)	13	2	37	+++	+	-	-	-
				IV:7 (F)	25	1.5	43	+	-	+	++	+
Family B	Punjab	*MCPH1*	c.1254del T	IV:1 (F)	18	3	41	++	-	+	-	-
				IV:2 (M)	21	3	40	+	-	-	++	-
				IV:3 (F)	20	4	39	+++	+	-	++	-
				IV:5 (M)	23	1.5	43	+	-	-	-	+
Family C	Punjab	*CENPJ*	**c.18del C**	VI:1 (M)	15	1.5	36	++	-	+	-	+
				VI:2 (F)	17	-	38	+++	-	-	+	+
				VI:4 (M)	20	-	43	+	-	-	++	+
				VI:5 (F)	23	-	42	+	+	-	++	+
				VI:6 (M)	25	-	43	+	+	+	-	+
				VI:7 (M)	27	2	45	++	-	-	+	+
Family D	Punjab	*ERCC6*	c.1178C>T	IV:1 (F)	26	2	46	++	-	-	+	-
				IV:4 (M)	23	2	47	++	-	-	++	-

Mild, (++) = Moderate, (+++) = Severe/Profound; Seizure Frequency: (+) = Rare (<1/month), (++) = Occasional (1/month–1/week), (+++) = Frequent (>1/week); HC = Head Circumference; M = Male; F = Female.

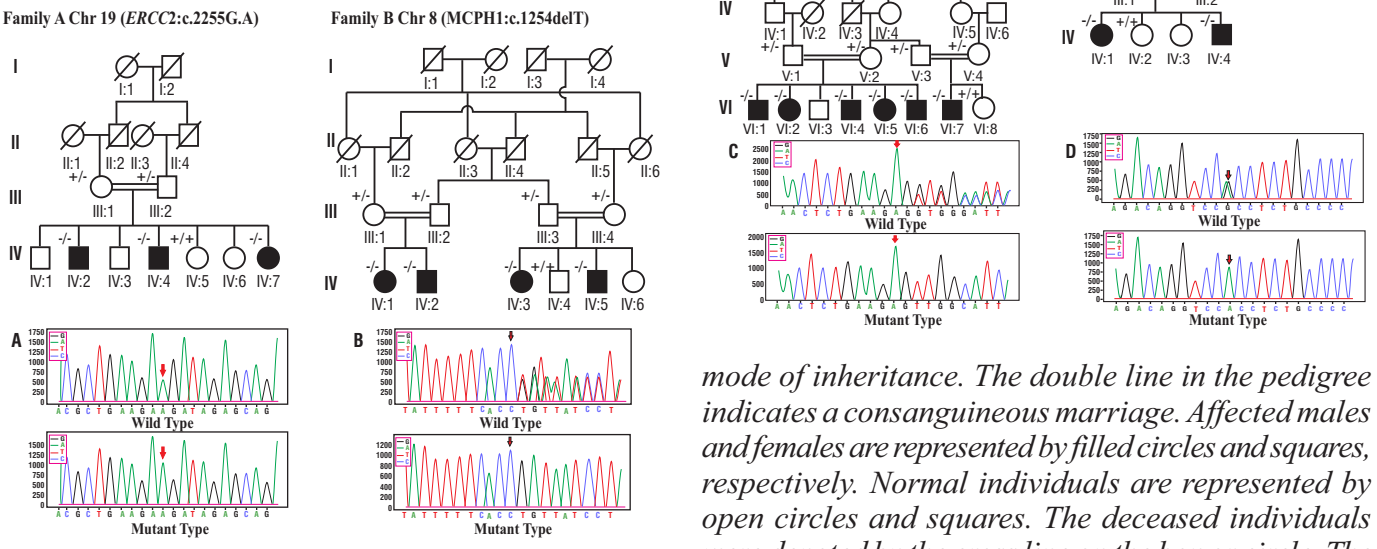

Figure 1: *Pedigrees and chromatograms of primary microcephaly families displaying an autosomal recessive mode of inheritance. The double line in the pedigree indicates a consanguineous marriage. Affected males and females are represented by filled circles and squares, respectively. Normal individuals are represented by open circles and squares. The deceased individuals were denoted by the cross line on the box or circle. The signs (+/-), (-/-), and (+/+) indicate heterozygous carriers, homozygous mutants, and homozygous wild-type indi-*

viduals, respectively.

***A:** Chromatogram ERCC2:c.2255G>A in Family A*

***B:** Chromatogram MCPH1:c.1254delT in Family B*

***C:** Chromatogram CENPJ:c.18delC in Family C*

***D:** Chromatogram ERCC6:c.1178C>T in Family D*

Family A: WES data of family A revealed a missense variant ERCC2:c.2255G>A. ERCC2 is involved in transcription initiation and nucleotide excision repair, and variation can interfere with the DNA repair mechanism,[7] affecting brain development and may cause primary microcephaly. ERCC2 encodes DNA helicase, which is essential for the nucleotide excision repair (NER) pathway, which repairs UV-induced DNA damage. Some phenotypes, such as Cockayne Syndrome, Xeroderma Pigmentosum (XP), and COFS Syndrome, can be induced due to the variations in ERCC2. These conditions may include growth failure, neurological impairment, microcephaly, and progressive neurodegeneration.[8,9] Additional studies are required for more specific insights into this association.

Family B: WES of family B revealed a frameshift variant, MCPH1:c.1254delT. MCPH1 is essential for brain growth, especially at the time of fetal stages. It is a multifunctional gene that plays roles in neuroprogenitor cell self-renewal, DNA damage response, and brain development. Sanger sequencing results showed that the single-base deletion in the MCPH1 segregates with the disorder. The variation (c.1254delT) results in damage to the Asp amino acid and loss of gene function that leads to primary microcephaly.

Family C: WES revealed a previously identified homozygous frameshift variation in CENPJ:c.18delC.[10] To maintain genomic integrity and normal cell development, CENPJ is a crucial protein engaged in several processes related to cell division and centrosome function. Variations in the CENPJ are linked with primary microcephaly, particularly in cases of the autosomal recessive mode, in which affected individuals have a small head size because of defective brain development. This gene also regulates several procedures necessary for brain development, and variations can cause microcephaly and other neurological disorders.[11]

Family D: Exome sequencing and Sanger sequencing revealed variation (c.1178C>T) in ERCC6, which encodes a DNA-binding protein required for transcription-coupled excision repair. At DNA repair sites, the encoded protein may encourage the formation of complexes by interacting with several transcription and excision repair proteins.[12] Variants in this gene have been reported to cause Cockayne Syndrome and Cerebro-oculo-facio-skeletal Syndrome 1.[13] Some severe cases of Cockayne Syndrome Type B (CSB) include neurodegeneration, developmental delays, and microcephaly.[14] While distinct disorders, Cockayne syndrome and primary microcephaly, share clinical features including intellectual disability and neurological malfunction.[15]

Silhouette analysis of the similarity matrix showed the existence of 8 clusters with a global Silhouette score of 0.32. Individual cluster validation showed that 6 out of 8 clusters were stable and compact as indicated by their average Silhouette scores (Table 2). In line with,[16] a cluster with an average Silhouette score of greater than 0.26 was considered a compact and real cluster. Silhouette value indicates the compactness of a cluster and is calculated from individual Silhouette values of all entities (genes) in a cluster. A gene with a Silhouette score less than 0 is considered an outlier (wrongly clustered). Figure 2A shows the outliers in each cluster. Clusters 1 and 4 contain a higher proportion of wrongly clustered genes and are not real clusters (Figure 2B).

Table 2: *Clustering results validations: Silhouette analysis of the similarity matrix showed the existence of 8 clusters with a global Silhouette score of 0.32.*

Cluster	Cluster size	Cluster's Silhouette	Average of Silhouette scores from all clusters
1	149	0.173	0.32
2	177	0.395	
3	135	0.279	
4	186	0.196	
5	283	0.339	
6	93	0.359	
7	134	0.477	
8	78	0.411	

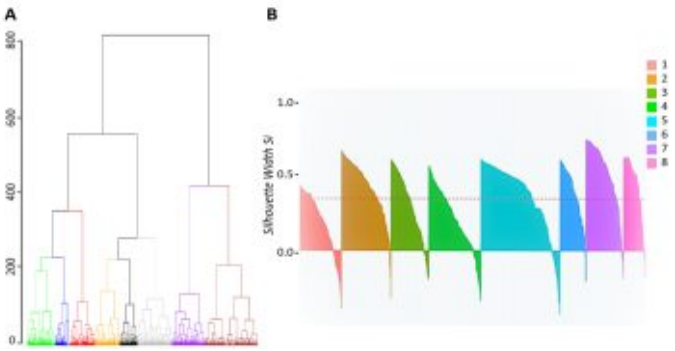

Figure 2. *Clustering analysis of genes causing multiple neurodevelopment disorders.*

(A) Graphical representation of identified clusters

(N=8). Each cluster is shown in a different color. The terminal nodes represent the genes that were clustered based on their functional similarity. (B) A graphical representation of genes that were clustered into 8 different clusters. The genes with a Silhouette score less than 0 are wrongly clustered genes. Each cluster possesses a small fraction of wrongly clustered genes. The red dotted line indicates the average Silhouette score for all clusters. Cluster 7 showed the Silhouette score compared to the scores of other clusters.

Separate enrichment analyses of genes from compact and stable clusters (2, 3, 5, 6, 7, and 8 clusters) were performed using the Enrichr web tool.[6] Table 3 contains the top 5 enriched pathways for genes from each cluster. The functional enrichment analysis of genes from cluster 2 showed significant enrichment of genes involved in specific signaling pathways.

Implicating a significant role of cluster 2 genes in cellular communication and signaling processes. Previously, studies have discovered the association of signaling pathways to neuronal growth, differentiation, and synaptic plasticity. The disruption in signaling pathways contributes significantly to NDDs etiology. The genes of cluster 2 were associated with phosphorylation bio-

Table 3: *Top five significantly enriched pathways for each compact and stable cluster.*

Cluster	IDs	Reactome pathway name	Adjusted *p*-value
2	R-HSA-162582	Signal Transduction	$7.89E^{-28}$
	R-HSA-5663202	Diseases of Signal Transduction by Growth Factor Receptors and Second Messengers	$3.79E^{-18}$
	R-HSA-9006934	Signaling by Receptor Tyrosine Kinases	$2.72E^{-16}$
	R-HSA-5684996	MAPK1/MAPK3 Signaling	$1.17E^{-12}$
	R-HSA-5683057	MAPK Family Signaling Cascades	$1.64E^{-12}$
3	R-HSA-112316	Neuronal System	$5.47E^{-50}$
	R-HSA-112315	Transmission Across Chemical Synapses	$5.68E^{-30}$
	R-HSA-6794361	Neurexins And Neuroligins	$7.34E^{-15}$
	R-HSA-112310	Neurotransmitter Release Cycle	$4.59E^{-14}$
	R-HSA-6794362	Protein-protein Interactions at Synapses	$5.96E^{-14}$
5	R-HSA-3247509	Chromatin Modifying Enzymes	$3.98E^{-40}$
	R-HSA-74160	Gene Expression (Transcription)	$1.40E^{-31}$
	R-HSA-212436	Generic Transcription Pathway	$9.62E^{-27}$
	R-HSA-73857	RNA Polymerase II Transcription	$1.17E^{-25}$
	R-HSA-8878171	Transcriptional Regulation by RUNX1	$1.64E^{-11}$
6	R-HSA-9675108	Nervous System Development	$1.54E^{-10}$
	R-HSA-422475	Axon Guidance	$3.29E^{-10}$
	R-HSA-6794362	Protein-Protein Interactions at Synapses	$6.46E^{-07}$
	R-HSA-1266738	Developmental Biology	$1.01E^{-06}$
	R-HSA-418990	Adherents Junctions Interactions	$1.08E^{-05}$
7	R-HSA-1430728	Metabolism	$1.71E^{-11}$
	R-HSA-72163	mRNA Splicing - Major Pathway	$6.84E^{-07}$
	R-HSA-72172	mRNA Splicing	$7.45E^{-07}$
	R-HSA-5668914	Diseases of Metabolism	$1.21E^{-06}$
	R-HSA-72203	Processing of Capped Intron-Containing Pre-mRNA	$6.89E^{-06}$
8	R-HSA-2514853	Condensation of Prometaphase Chromosomes	$8.61E^{-02}$
	R-HSA-68886	M Phase	$2.39E^{-01}$
	R-HSA-68877	Mitotic Prometaphase	$2.62E^{-01}$
	R-HSA-2299718	Condensation of Prophase Chromosomes	$2.99E^{-01}$
	R-HSA-8948216	Collagen Chain Trimerization	$2.99E^{-01}$

logical processes, more specifically protein phosphorylation and modification.

The overrepresentation analysis of cluster 3 genes revealed their associations with neuronal systems. This is contrary to cluster 2 genes, which were associated with signaling pathways. In line with pathways, cluster 3 genes were enriched for biological processes playing a key role in synaptic transmission.

Cluster 5 genes were different from both clusters 2 and 3 and were found to be involved in transcriptional regulation. ERCC2 and ERCC6 were found grouped in cluster[5]. Genes from cluster 6 were also linked with core pathways of NDDs, such as nervous system development. The GO enrichment analysis also showed the association of these genes with the neuron differentiation biological process. Clusters 7 and 8 genes tend to have a role in cellular metabolism and division (Table 3).

Discussion

MCPH is a rare neurodevelopmental condition characterized by a significantly small head circumference at birth as compared to normal individuals.[17] This disorder results in reduced brain size and cognitive impairment due to defective brain growth during embryogenesis.[18] In our study, Whole Exome Sequencing was employed for mutation detection. We identified four different genes; two of them (ERCC2 and ERCC6) are novel, while the other two (MCPH1 and CENPJ) are already known to cause primary microcephaly.[19] In addition, we also identified a functional overlap between MCPH and other genes causing neurodevelopment disorders (NDDs) by unsupervised machine learning.

Unsupervised machine learning analysis revealed that our identified target genes ERCC2, MCPH1, ERCC6, and CENPJ were present in clusters 3 and 5, which are associated with primary microcephaly either directly or indirectly. Cluster 5 includes (ERCC2, MCPH1, and ERCC6), which were different from cluster 3 (CENPJ). The variants of ERCC2, MCPH1, and ERCC6 of families A, B, and D, respectively, fall in Cluster 5. All affected members of family A, B, and D had intellectual disabilities. Some people who were affected had epilepsy and lowered muscle tone. Only family A had a member with polydactyly. Family A and B all had hearing loss, facial dysmorphisms, and ocular problems, while family D did not. The Reactome database pathways of ERCC2 and ERCC6 associated with cluster 5 include; Gene Expression, (Transcription) (R-HSA-74160), Generic Transcription Pathway (R-HSA-212436), Transcriptional Regulation By TP53 (R-HSA-3700989), RNA Polymerase I Transcription Initiation (R-HSA-73762), RNA Polymerase I Transcription (R-HSA-73864), and RNA Polymerase II Transcription (R-HSA-73857).[20] It means these pathways play a major role in the transcription process, and mutations in these genes (ERCC2 and ERCC6) may cause transcription to be interrupted, which can result in the disorders.[7,21] On the other hand, MCPH1 is also included in cluster 5. This gene is involved in the Condensation Of Prophase Chromosomes (R-HSA-2299718). Chromosome condensation may be disturbed as a result of the mutation in MCPH1, which may lead to primary microcephaly.[22]

The variant of CENPJ of family C is included in Cluster 3. All affected members of family C have an intellectual disability. Some affected members of this family had hearing loss, visual problems, and seizures. Facial dysmorphism was also observed in certain members of this family. The Reactome database pathways of CENPJ associated with cluster 3 consist of; Intellectual Disability, Global developmental delay, Mental and motor retardation, and Cognitive delay, which revealed a strong association with primary microcephaly (MCPH).[23]

This study supports earlier research indicating that primary microcephaly is genetically heterogeneous, particularly within a consanguineous Pakistani population. Various populations have previously linked MCPH1 and CASK to primary microcephaly,[10] the discovery of additional variants, ERCC2 and ERCC6, expands the spectrum of MCPH mutations. The pathogenesis of MCPH may be attributed to a disruption in transcription-coupled DNA repair, although ERCC2 and ERCC6 are also associated with DNA repair pathways and conditions such as Cockayne syndrome.[7] This underscores the significant functional relationship between the DNA repair process and neurodevelopment.

In our study, employing unsupervised machine learning marks a novel strategy for understanding the functional links between primary microcephaly and various neurodevelopmental disorders. We identified common biological processes and molecular functions primarily within clusters 2, 3, and 5 by utilizing gene-based clustering analysis alongside pathway enrichment data, in line with methodologies employed in recent neurodevelopmental research.[24] These clusters encompass pathways associated with synaptic signaling, transcription regulation, and neuronal differentiation. These discoveries suggest that shared genetic mechanisms may contribute to different clinical presentations of neurodevelopmental disorders, reinforcing the emerging idea of gene pleiotropy in these conditions.[25]

While our findings enhance the understanding of MCPH

genetics, disease mechanisms, and shared molecular pathways, the modest sample size may limit broader applicability. Functional validation, though beyond the present scope, would further clarify causal links. The clustering analysis revealed promising biological associations; however, additional experimental work is needed to confirm these connections. Future studies with larger cohorts and integrated transcriptomic or proteomic data will help refine the MCPH gene network and may uncover novel therapeutic targets.

Conclusion

This study explored the genetic variations associated with primary microcephaly (MCPH) in Pakistani consanguineous families using WES and identified novel and previously reported variants in ERCC2, MCPH1, CENPJ, and ERCC6. These findings add to the understanding of the diverse genetic etiology of MCPH and extend the spectrum of variations associated with the disease. Additionally, using unsupervised machine learning methods provided insights into the functional overlap between MCPH and other neurodevelopmental disorders (NDDs), thereby offering a new perspective on the genetic overlaps and potential shared disease mechanisms between these disorders.

Ethical Approval: The Institutional Review Boards (IRBs) of Government College University, Faisalabad, Pakistan approved this study vide No. (Ref. No. GCUF/ERC/307).

Data sharing statement

The Whole Exome Sequencing (WES) data generated in this study is not openly accessible due to ethical and privacy concerns. The corresponding author, Dr. Muhammad Qasim, at qasimawan@gcuf.edu.pk, may be approached to access these datasets for valid research purposes.

Conflict of Interest: The authors declare no conflict of interest.

Funding Source: None

Author's Contribution

BA: Analysis & interpretation of data, drafting of article

MQ: Conception & design, drafting of article, critically revised it for important intellectual content, final approval of the version to be published

MA: Analysis & interpretation of data, critically revised it for important intellectual content

MT: Conception & design, drafting of article, critically revised it for important intellectual content, final approval of the version to be published

References

1. Kristofova M, Ori A, Wang ZQ. Multifaceted microcephaly related gene MCPH1. Cells. 2022;11(2):275. doi:10.3390/cells11020275.
2. Prasad T, Iyer S, Chatterjee S, Kumar M. In vivo models to study neurogenesis and associated neurodevelopmental disorders—Microcephaly and autism spectrum disorder. WIREs Mech Dis. 2023;15(4):e1603. doi: 10.1002/wsbm.1603.
3. Muers M. Gene expression: Transcriptome to proteome and back to genome.. 2011;12(8):518. doi:10.1038/nrg3037.
4. Xi J, Yuan X, Wang M, Li A, Li X, Huang Q, et al. Inferring subgroup specific driver genes from heterogeneous cancer samples via subspace learning with subgroup indication. Bioinformatics. 2020;36(6):1855–63. doi: 10.1093/bioinformatics/btz793.
5. Asif M, Martiniano HFMC, Lamurias A, Kausar S, Couto FM. DGH GO: dissecting the genetic heterogeneity of complex diseases using gene ontology. BMC Bioinformatics. 2023;24(1):171. doi:10.1186/ s12859 - 023-05290-4.
6. Xie Z, Bailey A, Kuleshov MV, Clarke DJB, Evangelista JE, Jenkins SL, et al. Gene Set Knowledge Discovery with Enrichr. Curr Protoc. 2021;1(3):e90. doi: 10. 1002/ cpz1.90.
7. Singh A, Compe E, Le May N, Egly JM. TFIIH subunit alterations causing xeroderma pigmentosum and trichothiodystrophy specifically disturb several steps during transcription. Am J Hum Genet. 2015; 96(2): 194–207. doi:10.1016/j.ajhg.2014.12.012.
8. Baer S, Obringer C, Julia S, Chelly J, Capri Y, Gras D, et al. Early onset nucleotide excision repair disorders with neurological impairment: Clues for early diagnosis and prognostic counseling. Clin Genet. 2020; 98(3): 251–60. doi:10.1111/cge.13798.
9. Reunert J, van den Heuvel A, Rust S, Marquardt T. Cerebro oculo facio skeletal (COFS) syndrome caused by the homozygous Gly47Arg variant in ERCC2. Am J Med Genet A. 2021;185(3):930–6. doi:10.1002/ ajmg. a.62048.
10. Khan NM, Masoud MS, Baig SM, Qasim M, Chang J. Identification of pathogenic mutations in primary microcephaly (MCPH) related three genes CENPJ, CASK and MCPH1 in consanguineous Pakistani families. Biomed Res Int. 2022;2022:3769948. doi:10.1155/ 2022/ 37 69948.

11. Cueto-González AM, Fernández-Cancio M, Fernández-Alvarez P, García-Arumí E, Tizzano EF. Unusual context of CENPJ variants and primary microcephaly: compound heterozygosity and nonconsanguinity in an Argentinian patient. Hum Genome Var. 2020; 7(1):20. doi:10.1038/s41439-020-0105-3.
12. Zhang X, Horibata K, Saijo M, Ishigami C, Ukai A, Okuda Y, et al. Mutations in UVSSA cause UV-sensitive syndrome and destabilize ERCC6 in transcription-coupled DNA repair. Nat Genet. 2012;44(5):593–7. doi: 10.1038/ng.2228.
13. Duong NT, Anh NP, Bac ND, Quang LB, Miyake N, van Hai N, et al. Whole exome sequencing revealed a novel ERCC6 variant in a Vietnamese patient with Cockayne syndrome. Hum Genome Var. 2022;9:21. doi:10.1038/s41439-022-00200-1.
14. Sartorelli J, Travaglini L, Macchiaiolo M, Garone G, Gonfiantini MV, Vecchio D, et al. Spectrum of ERCC6-related Cockayne syndrome (type B): from mild to severe forms. Genes. 2024;15(4):508. doi: 10.3390/ genes 15040508
15. He C, Sun M, Wang G, Yang Y, Yao L, Wu Y. Two novel mutations in ERCC6 cause Cockayne syndrome B in a Chinese family. Mol Med Rep. 2017;15(6): 3957–62. doi:10.3892/mmr.2017.6487.
16. Asif M, Martiniano HFMC, Marques AR, Santos JX, Vilela J, Rasga C, et al. Identification of biological mechanisms underlying a multidimensional ASD phenotype using machine learning. Transl Psychiatry. 2020; 10(1):1–12. doi:10.1038/s41398-020-0721-1.
17. Asif M, Abdullah U, Nürnberg P, Tinschert S, Hussain MS. Congenital microcephaly: a debate on diagnostic challenges and etiological paradigm of the shift from isolated/non-syndromic to syndromic microcephaly. Cells. 2023;12(4):642. doi:10.3390/cells12040642.
18. Faheem M, Naseer MI, Rasool M, Chaudhary AG, Kumosani1 TA, Ilyas AM,et al. Molecular genetics of human primary microcephaly: an overview. BMC Med Genomics. 2015;8(1):S4. doi:10.1186/1755-8794-8-S1-S4.
19. Woods CG, Bond J, Enard W. Autosomal recessive primary microcephaly (MCPH): a review of clinical, molecular, and evolutionary findings. Am J Hum Genet. 2005;76(5):717–728. doi:10.1086/429930
20. Croft D, Mundo AF, Haw R, Milacic M, Weiser J, Wu G, et al. The Reactome pathway knowledgebase. Nucleic Acids Res. 2014;42(Database issue):D472–7. doi:10.1093/nar/gkt1102.
21. Gültekin-Zaim ÖB, Yalçın-Çakmaklı G, Çolpak Aİ, Şimşek-Kiper PÖ, Utine GE, Elibol B. Cockayne syndrome type 3 with dystonia-ataxia and clicking blinks. Mov Disord Clin Pract. 2023;10(3):S48. doi:10. 1002/mdc3.13778.
22. Yamashita D, Shintomi K, Ono T, Gavvovidis I, Schindler D, Neitzel H, et al. MCPH1 regulates chromosome condensation and shaping as a composite modulator of condensin II. J Cell Biol. 2011;194(6):841–54. doi:10.1083/jcb.201106141.
23. Rasool S, Baig JM, Moawia A, Ahmad I, Iqbal M, Waseem SS, et al. An update of pathogenic variants in ASPM, WDR62, CDK5RAP2, STIL, CENPJ, and CEP135 underlying autosomal recessive primary microcephaly in 32 consanguineous families from Pakistan. Mol Genet Genomic Med. 2020;8(9):e1408. doi: 10.1002/mgg3.1408.
24. Cuppens T, Kaur M, Kumar AA, Shatto J, Ng HT, Leclercq M, et al. Developing a cluster-based approach for deciphering complexity in individuals with neurodevelopmental differences. Front Pediatr. 2023; 11: 1171920. doi:10.3389/fped.2023.1171920.
25. Girault JB, Veatch OJ, Won H. Etiologic heterogeneity, pleiotropy, and polygenicity in behaviorally defined intellectual and developmental disabilities. J Neurodev Disord. 2024;16:8. doi:10.1186/s11689-024-09526-z.