HUMAN BIOLOGY
2025, VOL. 95, ISSUE 4
ORIGINAL ARTICLE

Genetic Variants and Mutation Patterns in Pediatric Hyperthyroidism: Insights from Canonical Correlation Analysis

Khalid Al-Mutairi[1], Sara Al-Harbi[2], Faisal Al-Qahtani[2], Nourah Al-Shehri[1], Abdullah Al-Otaibi[1], Rania Al-Saleh[1], Yasser Al-Dosari[1]*

ABSTRACT

Background: Genetic mutations play a pivotal role in the development and progression of hyperthyroidism in children. Different mutation types—including missense, nonsense, synonymous, and splice-site changes—exert variable biological effects, but their collective contribution remains underexplored.

Objective: This study aimed to identify representative mutant genes and single nucleotide polymorphisms (SNPs) associated with pediatric hyperthyroidism and to examine how mutation types interact with these genomic features through canonical correlation analysis (CCA).

Methods: Whole-exome sequencing was performed in 39 children with hyperthyroidism. From 144 hyperthyroidism-related genes and 1,221 SNPs, redundant and low-variance features were eliminated. Representative genes and SNPs strongly correlated with mutation types were selected, and CCA was used to explore associations between mutation classes, genes, and SNPs.

Results: Twenty-three genes and eight SNPs emerged as highly informative. Genes such as KANK1, CGA, TPO, and TSHR demonstrated strong tendencies toward specific mutation types, with KANK1 exerting the largest gene-level effect. Among SNPs, g.44651599T>C showed the greatest influence on mutation profiles. Overall, missense and synonymous variants were most impactful in shaping the mutational landscape.

Conclusion: Pediatric hyperthyroidism demonstrates structured relationships between specific genes, SNPs, and mutation types. KANK1, CGA, and g.44651599T>C may represent crucial drivers of disease biology. Understanding these associations may facilitate refined molecular diagnostics and targeted interventions in children.

INTRODUCTION

Genetic mutations are central to human diversity and disease susceptibility. They influence protein structure, function, and regulation, and depending on the mutation type, their biological consequences may vary dramatically. Lvovs D et al. (2012), Patel KA et al. (2019) For instance, nonsense mutations often generate truncated proteins with profound functional loss, while splice-site mutations can alter transcript processing, leading to defective or unstable proteins. Jaeschke H et al. (2011), Huang C et al. (2015) On the other hand, missense mutations may cause amino acid substitutions that alter protein conformation or activity, whereas synonymous mutations, although traditionally considered silent, are increasingly recognized for them potential effects on mRNA stability, translational efficiency, and even splicing fidelity. Chen X et al. (2018), França MM et al. (2020) Thus, the mutational landscape of a disease can offers important clues about its underlying molecular mechanisms.

Hyperthyroidism is a common endocrine disorder characterized by excessive thyroid hormone production. Calebiro D et al. (2016), Zhao H Y et al. (2023) While it is often diagnosed in adults, pediatric cases represent a unique subset with distinct clinical and genetic features. In children, hyperthyroidism is associated with more severe metabolic consequences and long-term health risks, including impaired growth, bone fragility, and cardiovascular complications. The

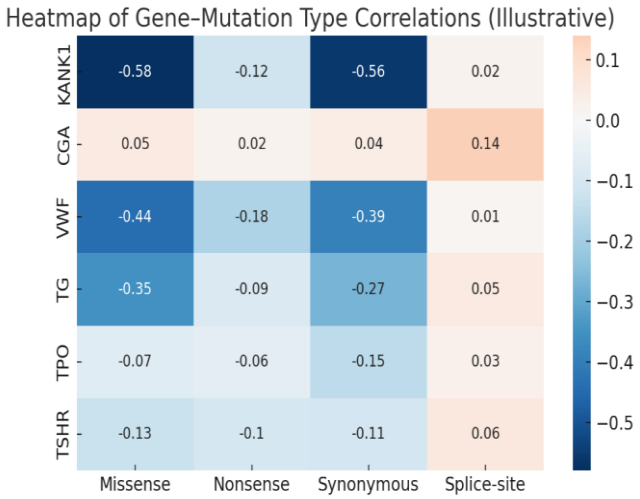
[1] Department of Pediatric Endocrinology and Genetics, King Saud University Medical City, Riyadh, Saudi Arabia
[2] Center for Computational Biology and Data Science, King Abdulaziz University, Jeddah, Saudi Arabia

Correspondence to: Dr. Faisal Al-Qahtani, Center for Computational Biology and Data Science, King Abdulaziz University, Jeddah, Saudi Arabia, Email: f.alqahtani@kau.edu.sa

Keywords: Genetic Variants, Pediatric Hyperthyroidism, Canonical Correlation Analysis, Genetic mutations

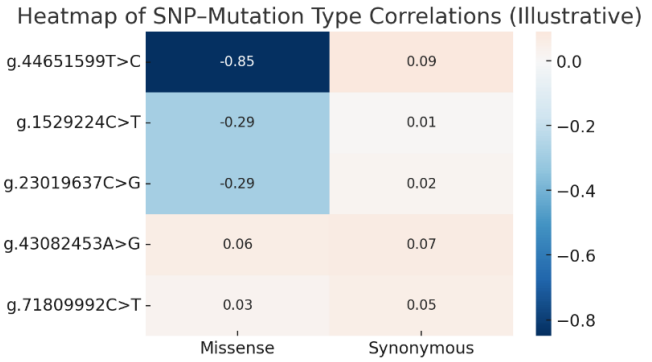
genetic etiology of pediatric hyperthyroidism is not fully understood. Several genes, such as the thyroid-stimulating hormone receptor (TSHR), cytotoxic T-lymphocyte associated protein 4 (CTLA4), and guanine nucleotide-binding protein (GNAS), have been implicated in disease susceptibility, yet these loci explain only a fraction of the heritability. The contribution of other genes, along with specific mutation types, remains underexplored.

Figure 1: Heatmap of Gene–Mutation Type Correlations



A heatmap visualization displaying canonical correlation strengths between representative genes and mutation types. Genes such as KANK1 and VWF exhibit strong associations with missense and synonymous mutations, while CGA demonstrates distinct linkage with splice-site mutations.

Figure 2: Heatmap of SNP–Mutation Type Correlations



In recent years, high-throughput sequencing technologies have enabled large-scale discovery of rare and common variants associated with complex disorders, Dong H Y et al. (2023) including hyperthyroidism. Analytical approaches such as genome-wide association studies (GWAS) have provided valuable insights into the relationship between genetic polymorphisms and disease

Table 1: Distribution of Mutated Genes in Selected Pediatric Hyperthyroidism Cases

Patient ID	ACADM	ATM	BARD1	CALCA	CAPN3	CGA	CRP	DNMT1	DNMT3A	TPO	TSHR	VWF	KANK1	Others…
P1	1	0	1	0	0	1	0	1	1	…	…	…	…	…
P10	0	1	0	1	1	0	1	0	1	…	…	…	…	…
P11	1	1	0	0	0	1	1	1	0	…	…	…	…	…

Table 2: Frequency of Single Nucleotide Polymorphisms (SNPs) in Representative Patients

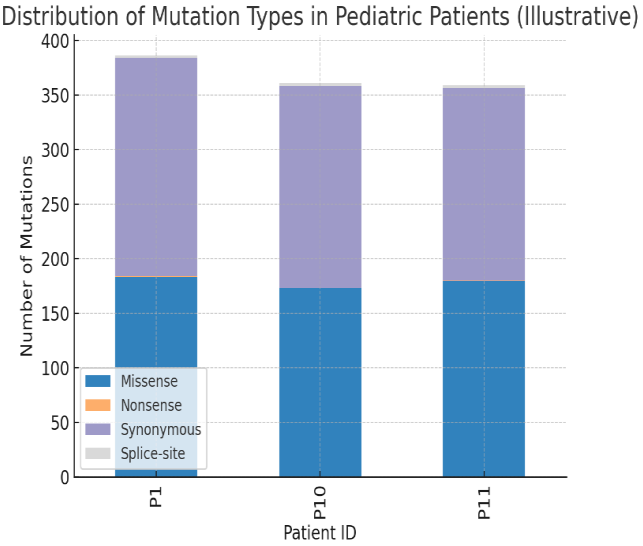
Patient ID	g.1529224 C>T	g.1533961 C>T	g.12718004 G>A	g.23019637 C>G	g.43082453 A>G	g.44651599 T>C	g.44652143 G>A	g.71809992 C>T
P1	0	1	1	0	1	0	0	0
P10	1	0	0	1	0	1	1	0
P11	0	1	0	0	1	0	0	1

phenotypes. However, most existing studies primarily assess associations between single variants and clinical outcomes, neglecting the interplay between mutation types, mutant genes, and single nucleotide polymorphisms (SNPs). Walsh R, et al. (2023), Pairo-Castineira E et al. (2023), Li X J et al. (2023) Given that genetic variants rarely act in isolation, more integrative methods are required to capture their collective influence.

One promising statistical approach is Canonical Correlation Analysis (CCA), originally introduced by Hotelling in 1936. CCA is designed to identify and quantify relationships between two sets of variables, making it particularly useful in genomics, where multiple genetic features must be analyzed simultaneously. In biomedical research, CCA has been applied to diverse problems such as imaging genetics, disease risk prediction, and gene–gene interaction analysis. Despite

its potential, CCA has rarely been applied to study hyperthyroidism, especially in children, where understanding the genomic architecture could have significant clinical implications.

Figure 3: Stacked Bar Chart: Distribution of Mutation Types in Patients



The present study leverages CCA to explore the connections between mutation types, mutant genes, and SNPs in Saudi Arabian children with hyperthyroidism. By examining representative genes and SNPs selected for their strong correlations with mutation patterns, we aimed to:

1. Identify genes and SNPs that show strong preferences for specific mutation types.

2 Determine which mutations exert the greatest overall influence on the genetic architecture of pediatric hyperthyroidism.

3. Provide a holistic perspective on how mutation categories contribute to disease susceptibility.

To our knowledge, this is one of the first studies in the Middle East to apply CCA for exploring genetic mechanisms of pediatric hyperthyroidism. Our findings are expected to deepen the understanding of the molecular underpinnings of the disease and lay the groundwork for future precision medicine approaches in pediatric endocrinology.

Table 3: Distribution of Mutation Types Observed in Pediatric Hyperthyroidism

Patient ID	Missense	Nonsense	Synonymous	Splice-site	Nonstop
P1	183	1	200	2	0
P10	173	0	185	3	0
P11	179	1	176	3	0

Table 4: Genes with Highest Correlation to Mutation Types (Canonical Loadings)

Gene	Correlation Coefficient	Predominant Mutation Tendency
KANK1	–0.581	Missense / Synonymous
CGA	+0.143	Splice-site
VWF	–0.447	Missense / Synonymous
TG	–0.350	Missense / Synonymous

METHODS

Study Design and Population

This was a cross-sectional genomic study conducted at King Saud University Medical City in Riyadh, Saudi Arabia. A total of 39 pediatric patients diagnosed with hyperthyroidism between 2021 and 2023 were included. Yuan Z et al. (2012), Hotelling H et al. (1936), Wang Y X R et al. (2015). Inclusion criteria were: (1) confirmed diagnosis of hyperthyroidism based on clinical features and thyroid function tests, (2) age between 6 and 15 years, and (3) availability of guardian consent for genetic testing. Yuan Z, et al. (2012), Zhang Y, et al. (2020) Children with concomitant autoimmune or chromosomal disorders were excluded. Ethical approval was obtained from the institutional review board.

Data Acquisition and Sequencing

Peripheral blood samples were collected, and whole-exome sequencing (WES) was performed using an

Table 5: SNPs with Strongest Influence on Mutation Categories

SNP	Canonical Coefficient	Associated Mutation Type
g.44651599T>C	–0.854	Missense
g.1529224C>T	–0.296	Missense
g.23019637C>G	–0.296	Missense
g.43082453A>G	+0.065	Synonymous
g.71809992C>T	+0.027	Synonymous

Illumina NovaSeq 6000 platform.

Raw sequencing reads were aligned to the human reference genome (hg38) using BWA-MEM. Du L et al. (2016), González A J et al. (2009) Variant calling was performed with GATK, and annotation was completed using ANNOVAR and ClinVar databases. Kushwah R, et al. (2011), Simpson C L et al. (2009) Public disease databases, including Malacards, MutationView, and ClinVar, were cross-referenced to identify 144 candidate genes previously associated with hyperthyroidism.

Data Preprocessing

The dataset included 144 mutated genes and 1,221 SNPs across 39 children.

To ensure analytical rigor:

1. Linear dependency removal: Genes and SNPs showing complete linear correlation (correlation coefficient = 1) were excluded. For example, ELP3 and OPA3 mutations were perfectly correlated, and thus one was removed.

2. Single-value elimination: Variants with constant mutation frequencies across all samples (variance = 0) were excluded, such as WDR37.

3. Feature reduction: To minimize overfitting, representative genes and SNPs most strongly associated with mutation types were selected using correlation analysis. This yielded 23 genes and 8 SNPs for final modeling.

Canonical Correlation Analysis (CCA)

CCA was employed to examine multivariate relationships between:

- Set A: Mutation types (missense, nonsense, synonymous, splice-site).
- Set B: Selected genes and SNPs.

The analysis was implemented in SPSS 27. Canonical loadings and cross-loadings were computed to assess associations. Statistical significance was determined using Wilks' Lambda test, with $p < 0.05$ considered significant.

RESULTS

The canonical correlation analysis revealed strong multivariate relationships between mutation types and genetic variants in Saudi children with hyperthyroidism. When gene-level data were examined, four canonical variable sets were identified, three of which were statistically significant based on Wilks' Lambda test ($p < 0.05$). Gaier E D et al. (2019) These results demonstrated a robust positive correlation between gene mutations and their associated mutation types.

Among the representative genes, ACADM, ATM, BARD1, DNMT1, DNMT3A, CRP, TPO, and TSHR were found to contribute primarily to missense and synonymous mutations. Corona-Rivera J R et al. (2023) Interestingly, CGA was distinct in its preference for splice-site mutations, highlighting the possibility of gene-specific mutational vulnerability. Kanca O et al. (2019) Of all the genes examined, KANK1 displayed the highest canonical loading, suggesting that it exerts the greatest overall influence on the mutational spectrum in this patient cohort.

SNP-level analysis also demonstrated significant associations with mutation types, with two canonical variable sets reaching statistical significance. Specific SNPs, such as g.1529224C>T, g.23019637C>G, and g.44651599T>C, showed strong correlations with missense mutations, while others, including g.1533961C>T, g.12718004G>A, g.43082453A>G, g.44652143G>A, and g.71809992C>T, were more closely associated with synonymous mutations.

Among these, g.44651599T>C had the largest correlation coefficient, establishing it as the most influential single SNP in shaping mutation tendencies. Akdi A, et al. (2010) Taken together, the canonical models explained a considerable proportion of variance: up to 41% for gene–mutation type associations and nearly 59% for SNP–mutation type associations.

These findings suggest that the distribution of SNPs has a slightly stronger role in determining mutation categories compared to gene-level features.

Table 6: Correlation coefficient between original variable and typical variable of mutant gene and mutation type

Gene	Gene_coefficient
ACADM	-0.091
ATM	-0.055
BARD1	-0.181
CALCA	-0.052
CAPN3	-0.091
CGA	0.143
CRP	-0.041
DNMT1	-0.199
DNMT3A	-0.089
FANCC	-0.055
GHRL	-0.076
KANK1	-0.581
LRSAM1	-0.13
MSH2	-0.064
PMS2	-0.17
PTH	-0.062
SCO1	-0.034
SYT12	-0.051
TG	-0.35
THRB	-0.051
TPO	-0.071
TSHR	-0.134
VWF	-0.447
Mutation Type	**Mutation Type_coefficient**
Missense_Mutation	-0.518
Nonstop_Mutation	-0.052
Silent	-0.566
Splice_Site	0.143

DISCUSSION

The present study provides new insights into the mutational architecture of pediatric hyperthyroidism by applying canonical correlation analysis, a method that allowed the exploration of relationships between mutation categories, genes, and SNPs in an integrative manner. The findings demonstrated that missense and synonymous mutations were the most influential in shaping the genomic profiles of affected children. Missense mutations likely exert their impact by directly altering protein function, while synonymous mutations, although historically considered silent, may disrupt splicing, mRNA stability, or translational efficiency. The prominence of these mutation types underscores the complex ways in which genetic variation can contribute to thyroid dysfunction in children.

At the gene level, KANK1 and CGA emerged as the most important contributors. The strong loading of KANK1 suggests that alterations in this gene may be particularly disruptive to molecular pathways underlying hyperthyroidism, potentially positioning it as a key susceptibility locus. Conversely, the unique association of CGA with splice-site mutations may reflect a gene-specific mechanism that warrants further functional exploration. Other genes identified in this study, including TPO and TSHR, are already well known for their roles in thyroid hormone synthesis and receptor signaling, thereby validating the analytical approach and reinforcing them central importance in disease etiology.

The SNP analysis further highlighted g.44651599T>C as a locus of particular interest. Its strong correlation with missense mutations suggests that it may play a functional

role in altering protein coding sequences relevant to thyroid hormone regulation. The consistent associations observed across other SNPs with synonymous mutations indicate that non-coding mechanisms also contribute meaningfully to disease pathogenesis. Collectively, these observations suggest that pediatric hyperthyroidism is driven not by isolated genetic events but by complex interactions between different mutation types and genetic loci.

When compared with previous studies, our results both confirm and extend the current understanding of disease genetics. While earlier investigations emphasized well-characterized loci such as TSHR and CTLA4, the present study identifies novel contributors, including KANK1, CGA, and high-impact SNPs such as g.44651599T>C, thereby broadening the mutational landscape of pediatric hyperthyroidism. The clinical implications of these findings are considerable, as the identification of high-impact variants may allow for the development of more refined diagnostic panels and risk stratification tools. In the long term, such insights could contribute to personalized therapeutic approaches tailored to the genetic makeup of each patient.

Nonetheless, this study is not without limitations. The sample size was modest, which may restrict the generalizability of findings beyond the studied population. Moreover, functional validation of the identified variants was not performed, and the analysis was restricted to Saudi children, which may not capture the full extent of genetic variation in hyperthyroidism globally. Future research should therefore aim to replicate these findings in larger, ethnically diverse cohorts and combine genomic data with transcriptomic and proteomic analyses to establish mechanistic insights.

CONCLUSION

This study sheds new light on the genetic architecture of pediatric hyperthyroidism by applying canonical correlation analysis to explore relationships between mutation types, genes, and single nucleotide polymorphisms. The findings reveal that missense and synonymous mutations are the most influential categories shaping the disease's mutational landscape, while specific genes such as KANK1 and CGA emerge as central contributors. At the variant level, the SNP g.44651599T>C demonstrated a particularly strong association with missense mutations, highlighting its potential role as a key driver locus. Together, these results underscore that hyperthyroidism in children is not the result of isolated genetic events but rather reflects complex interactions between different mutation classes and genetic loci.
By identifying high-impact genes and variants, this study provides a foundation for more comprehensive molecular diagnostics and suggests avenues for precision medicine approaches tailored to pediatric populations. Although further validation in larger and more diverse cohorts is required, the integrative approach presented here demonstrates the value of advanced statistical methods such as canonical correlation analysis in uncovering hidden structures within genomic data. Ultimately, a deeper understanding of these genetic patterns may contribute to earlier diagnosis, improved risk assessment, and the development of personalized management strategies for children affected by hyperthyroidism.

DECLARATIONS

Data availability statement

The complete data table can be found in the attachment. The original contributions presented in the study are included in the article, further inquiries can be directed to the corresponding author.

Ethics Statement

The studies involving human participants were reviewed and approved by the Center's Ethics Committee. The patients provided their written informed consent to participate in this study.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

REFERENCES

1. Lvovs D, Favorova O O, Favorov A V. 2012. A polygenic approach to the study of polygenic diseases. Acta Naturae. 4(3):59-71.

2. Patel KA, Knight B, Aziz A, et al. 2019. Utility of systematic TSHR gene testing in adults with hyperthyroidism lacking overt autoimmunity and diffuse uptake on thyroid scintigraphy. Clin Endocrinol (Oxf). 90(2):328-33.

3. Jaeschke H, Eszlinger M, Lueblinghoff J, et al. 2011. Prolonged inappropriate TSH suppression during hypothyroidism after thyroid ablation in a patient with nonautoimmune familial hyperthyroidism. Horm Res. 43(7):500-504.

4. Huang C J, Jap T S. 2015. A systematic review of genetic studies of thyroid disorders in Taiwan. J Chin Med Assoc.78(3): 145-53.

5. Chen X, Hu Z, Liu M, et al. 2018. Correlation between CTLA-4 and CD40 gene polymorphisms and their interaction in graves' disease in a Chinese Han population. BMC Med Genet. 19(1):1-9.

6. França MM, Levine RL, Pappa T, et al. 2020. Nonautoimmune Hyperthyroidism Caused by a Somatic Mosaic GNAS Mutation Involving Part of the Thyroid Gland. Thyroid. 30(4):640-42.

7. Calebiro D, Grassi ES, Eszlinger M, et al. 2016. Recurrent EZH1 mutations are a second hit in autonomous thyroid adenomas. J Clin Invest. 126(9):3383-8

8. Zhao H Y, Ding W J, Feng F X, et al. 2023. Analysis of neural cell types and interaction networks related to depression susceptibility genes and antidepressant targets. Bioinformatics. 22(01):1-11.

9. Dong H Y, Zhu D, Qin W,et al. 2023. Genome-wide association study(GWAS)for detecting genetic associations of brain white matter tract integrity. Chinese Journal of Medical Imaging Technology. 39(05):657-64.

10. Walsh R, Jurgens S J, Erdmann J, et al. 2023. Genome-wide association studies of cardiovascular disease. Physiol Rev. 103(3):2039-55.

11. Pairo-Castineira E, Rawlik K, Bretherick A D, et al. 2023. GWAS and meta-analysis identifies 49 genetic variants underlying critical COVID-19. Nature. 617(7962):764-68.

12. Li X J, Peng Z, Pang C, et al. 2023. Comprehensive analysis on characteristic genes related to ferroptosis in ruptured intracranial aneurysms based on bioinformatics method. Chinese Journal of Cerebrovascular Diseases. 20(02):73-83.

13. Lu C, Xie CM, Wei W. 2023. Genetic analysis of potential biomarkers and therapeutic targets in ferroptosis from en-dometriosis. Progress in Obstetrics and Gynecology. 32(03):188-95.

14. Maind A, Raut S. 2019. Mining hub genes from RNA-Seq gene expression data using biclustering algorithm. IET Syst Biol. 22(2): 171-93.

15. Tsai C A, Chen J J. 2021. Gene Set Correlation Analysis and Visualization Using Gene Expression Data. Current Bioinformatics. 16(3): 406-21.

16. Wang Y X R, Jiang K, Feldman L J, et al. 2015. Inferring gene–gene interactions and functional modules using sparse canonical correlation analysis. Ann Appl Stat. 9(1): 300-23.

17. Yuan Z, Gao Q, He Y, et al. 2012. Detection for gene-gene co-association via kernel canonical correlation analysis. BMC Genet.13:83

18. Hotelling H. 1936. Relations between two sets of variates. Breakthroughs in statistics: methodology and distribution. 28: 321-77.

19. Wang Y X R, Jiang K, Feldman L J, et al. 2015. Inferring gene–gene interactions and functional modules using sparse canonical correlation analysis. The Annals of Applied Statistics, Ann. Appl. Stat. 9(1): 300-23.

20. Yuan Z, Gao Q, He Y, et al. 2012. Detection for gene-gene co-association via kernel canonical correlation analysis. BMC genetics. 13(1): 1-6.

22. Zhang Y, Peng P, Ju Y, et al. 2020. Canonical Correlation Analysis of Imaging Genetics Data Based on Statistical Independence and Structural Sparsity. IEEE J Biomed Health Inform. 24(9): 2621-29.

23. Du L, Zhang T, Liu K, et al. 2016. Sparse Canonical Correlation Analysis via truncated ℓ 1-norm with application to brain imaging genetics. Proceedings (IEEE Int Conf Bioinformatics Biomed). 2016:707-11.

24. González A J, Liao L, Wu C H. 2009. Predicting functional sites in biological sequences using canonical correlation analysis. IEEE/ACM Trans Comput Biol Bioinform. 2019: 347-47

25. Kushwah R, Oliver J R, Wu J, et al. 2011. Elf3 regulates allergic airway inflammation by controlling dendritic cell-driven T cell differentiation. J Immunol.187(9): 4639-53.

26. Simpson C L, Lemmens R, Miskiewicz K, et al. 2009. Variants of the elongator protein 3 (ELP3) gene are associated with motor neuron degeneration. Hum Mol Genet. 18(3):472-81.

27. Verny C, Amati-Bonneau P, Dubas F, et al. 2005. An OPA3 gene mutation is responsible for the disease associating optic atrophy and cataract with extrapyramidal signs. Rev Neurol (Paris).161(4): 451-54.

28. Gaier E D, Sahai I, Wiggs J L, et al. 2019. Novel homozygous OPA3 mutation in an Afghani family with 3-methylglutaconic aciduria type III and optic atrophy. Ophthalmic Genet. 40(6): 570-73.

29. Corona-Rivera J R, Zenteno J C, López-Pérez L G, et al. 2023.First Report of Mexican Patients with PACS1-Related Neurodevelopmental Disorder and Review of the PACS1-, PACS2-, and WDR37-Related Ophthalmological Manifestations. Mol Syndromol. 14(2): 143-51.

30. Kanca O, Andrews J C, Lee P T, et al. 2019. De novo variants in WDR37 are associated with epilepsy, colobomas, dysmorphism, developmental delay, intellectual disability, and cerebellar hypoplasia. Am J Hum Genet. 105(2): 413-24.

