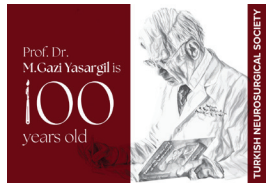
Turk Neurosurg 35(3):463-473, 2025
DOI: 10.5137/1019-5149.JTN.46886-24.2

Original Investigation

Neuro-Oncology

Received: 08.05.2024
Accepted: 11.10.2024
Published Online: 30.04.2025



Mutational and Expressional Similarities Among Paraganglioma, Low-Grade Glioma, and Glioblastoma: A Comprehensive Clustering Approach to Central Nervous System Tumors

Saliha ACAR[1], Giyasettin OZCAN[2], Eyyup GULBANDILAR[3]

[1]Anadolu University, Computer Research and Application Center, Eskisehir, Türkiye
[2]Bursa Uludag University, Faculty of Engineering, Department of Computer Engineering, Bursa, Türkiye
[3]Eskisehir Osmangazi University, Faculty of Engineering and Architecture, Department of Computer Engineering, Eskisehir, Türkiye

Corresponding author: Saliha ACAR ✉ salihaacar@anadolu.edu.tr

ABSTRACT

AIM: To compare central nervous system (CNS) tumors, such as paraganglioma, low-grade glioma (LGG), and glioblastoma (GBM), in terms of driver genes and gene expression, and to investigate the roles of common driver genes and genes with altered expression in cellular proliferation mechanisms and their interactions.

MATERIAL and METHODS: Mutation datasets for pheochromocytoma/paraganglioma, LGG, and GBM from The Cancer Genome Atlas (TCGA) database were used for driver gene prediction. Six datasets from the Gene Expression Omnibus (GEO) database were used for differential gene expression analysis. A hybrid approach combining clustering and computational biology methods was applied to identify driver genes. Gene expression analyses were repeated for two gene expression datasets for each tumor type, and the intersection of the results was taken. Protein interaction analyses, overall survival analyses, and carcinogenesis-related functional analyses were performed on the common driver genes and the genes with the most significant changes in expression.

RESULTS: ATRX, NF1, MUC16, and TTN were identified as driver gene candidates for all three tumor types. FSTL5, GABRG2, VSNL1, and LPL were found to be the genes with the most altered expression across all tumor types. Our findings suggest that, while CNS tumors with similar symptoms share molecular features, they can be more accurately differentiated through detailed investigation of the expression and mutation burden of the identified genes. This may also help accelerate the treatment planning process.

CONCLUSION: This study confirms that paraganglioma, LGG, and GBM may share common mutational and expressional gene patterns. The identified genes may serve as potential therapeutic targets in the treatment of glial and neuroendocrine tumors.

KEYWORDS: CNS, Brain tumors, Driver gene, Gene expression, GEO, Neuroendocrine tumors, TCGA

INTRODUCTION

Central nervous system (CNS) tumors are masses that affect nerve and glial cells, with a high morbidity rate, although more than half are benign. CNS tumors, which make up approximately 1.6% of all human tumors, are among the most complex types of cancer. They encompass a wide variety of tumor types that, while anatomically similar, differ in morphology, etiology, origin, molecular biology, and clinical progression (6). There are more than 120 types of CNS tumors, and malignant ones are a leading cause of death in both adults and children (2). The 5-year survival rate

Saliha ACAR : 0009-0005-9973-4980 Giyasettin OZCAN : 0000-0002-1166-5919 Eyyup GULBANDILAR : 0000-0001-5559-5281

for malignant CNS tumors across all ages is around 36%, but survival expectations vary depending on tumor characteristics (12).

Recent statistics indicate an increase in the incidence of CNS tumors. Understanding their epidemiology is crucial for early diagnosis, treatment, and the development of new therapeutic methods (8).

The literature reveals that studies in this field generally focus on the classification of CNS tumors and brain tumors, a subset of CNS tumors (3,5,11). Clinical single case reports are also available (1,13,15).

This study compares driver genes and gene expressions between paraganglioma, low-grade glioma (LGG), and glioblastoma (GBM), all types of CNS tumors. The roles of common driver genes and genes with altered expression in the cellular proliferation mechanism, as well as their interactions, were also investigated.

The World Health Organization (WHO) grades CNS tumors on a scale from 1 to 4. Grade 1 tumors grow slowly and tend not to spread, while Grade 4 tumors are the most aggressive. Paragangliomas are classified as Grade 1, 2, or 3, low-grade gliomas as Grade 1 or 2, and glioblastomas as Grade 4 tumors (17).

Non-epithelial neuroendocrine neoplasms, known as paragangliomas, predominantly produce catecholamines and secrete them into the bloodstream like hormones. These tumors have a high genetic predisposition and can be either sympathetic or parasympathetic. The term pheochromocytoma refers specifically to intraadrenal paragangliomas, which represent the classic sympathetic form (7). Gliomas are tumors that develop in the glial cells supporting the CNS, constituting 30% of all CNS tumors. Approximately 55% of glioma cases present as glioblastoma, an aggressive tumor (6).

Cancer is a multifactorial and complex disease. Genetic alterations play a key role in carcinogenesis, making comprehensive bioinformatics analyses essential for understanding factors in disease progression. The publicly available The Cancer Genome Atlas (TCGA) and Gene Expression Omnibus (GEO) databases provide access to functional genomic datasets of various types of human cancer (4,14), facilitating a range of analyses.

In the current study, data from both databases enabled a detailed examination of the relationship between gene mutations, expression changes, and various CNS malignancies. First, a domain-specific clustering method was developed for each of the tumors mentioned, allowing for driver gene predictions. Differential gene expression analyses were then performed. The intersections of predicted driver genes and significantly altered genes were identified, and their interactions were investigated. Additionally, function analyses related to carcinogenesis and expression-dependent survival studies were conducted for the relevant tumor types.

MATERIAL and METHODS

Prediction of Driver Genes

The general process for predicting cancer driver genes is outlined in Figure 1, with a detailed flow diagram provided in Figure 2. The analyses were coded in RStudio using version 4.3.0 of the R programming language and in Jupyter Notebook using version 3.10.9 of Python. These analyses were performed using somatic mutation datasets for pheochromocytoma/paraganglioma (PCPG), LGG, and GBM, all of which were downloaded from the TCGA database.

The impact value of each mutation was scaled from 1 to 5, with low-impact mutations assigned a value of 1 and high-impact mutations a value of 5. Mutations with medium impact were given values in the range of 2 to 4. This scaling was based on the SIFT and PolyPhen variables from the mutation dataset, with missing values for these variables filled using the mean value of 0.5.

Weighted mutation scores were calculated using the entropy criterion weighting method. The number of frameshift mutations, which may result in loss or gain of function, and the number of mutations altering start/stop codons were also in-

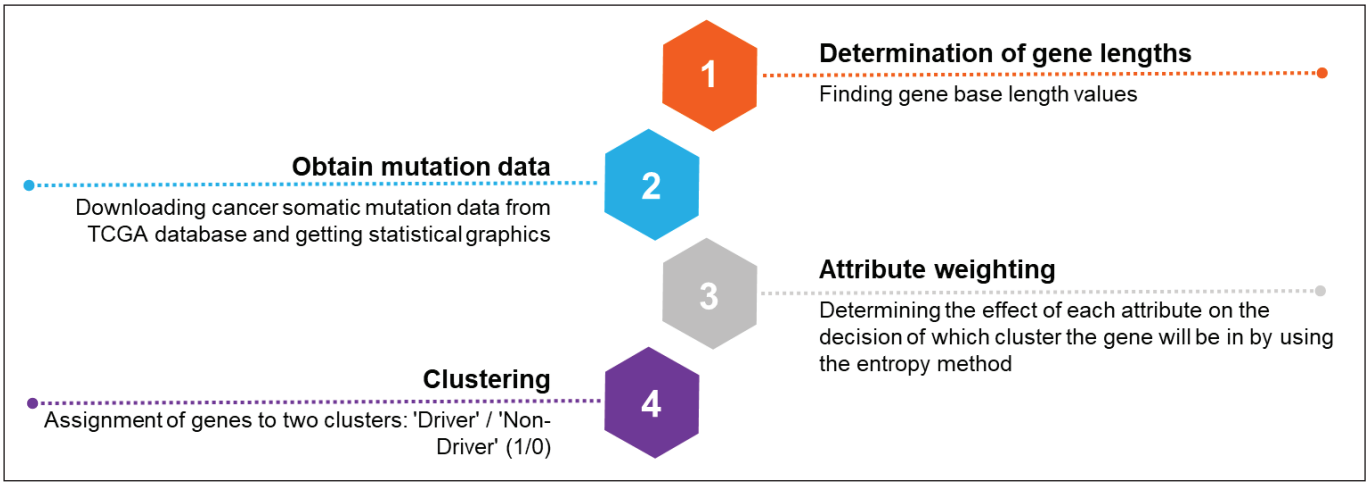


Figure 1: General process steps of driver gene prediction.

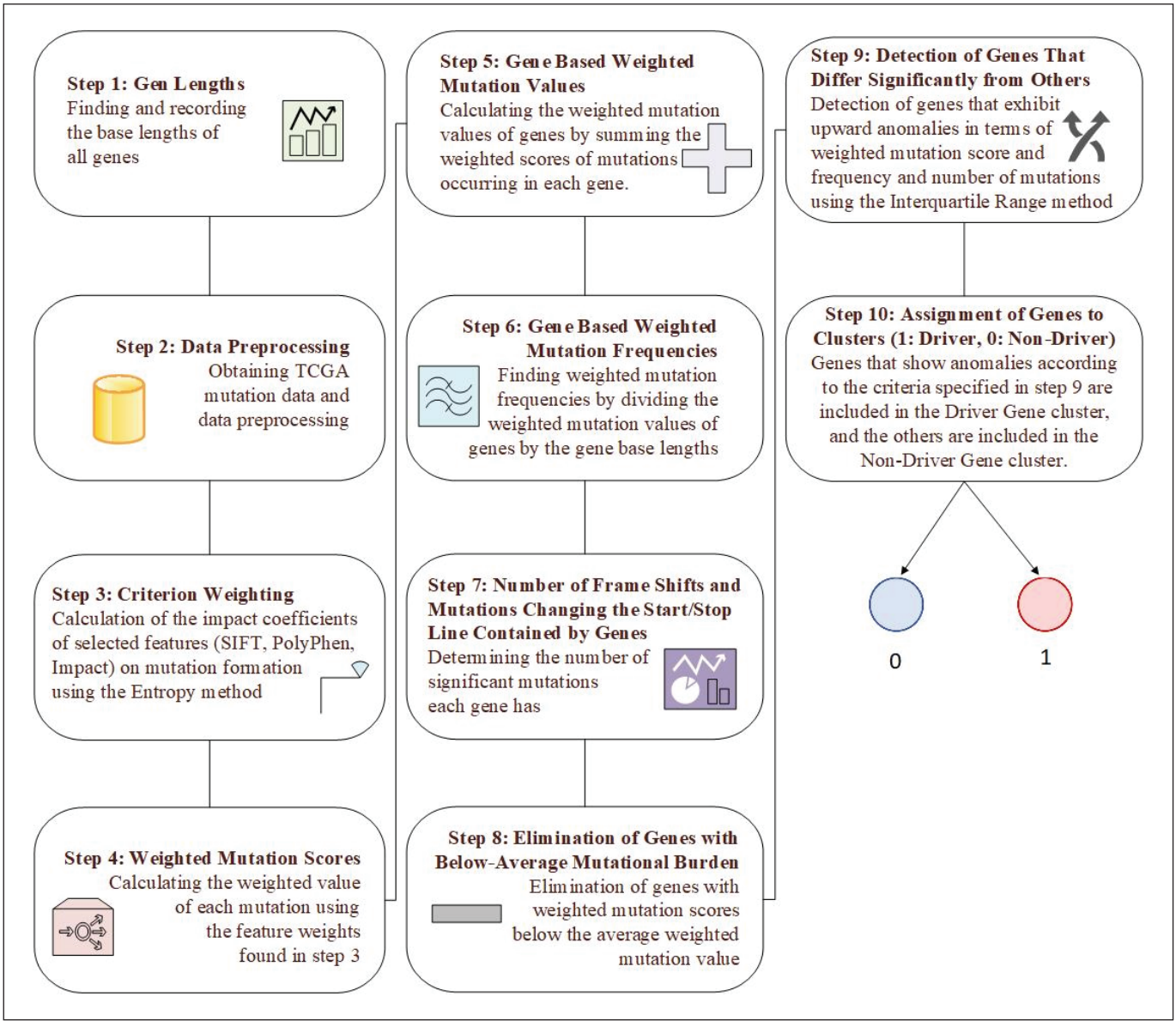


Figure 2: Detailed flow diagram of driver gene prediction method.

cluded in the analysis. After scoring all genes based on their mutations, genes with a total weighted mutation score below the average were first eliminated to narrow down the gene set potentially associated with the tumor. The Interquartile Range (IQR) method, commonly used to identify outliers in data mining, was then applied for further elimination. In descriptive statistics, the IQR measures the spread of data by covering the middle 50% of a sorted dataset, calculated as the difference between the third quartile (Q3) and the first quartile (Q1) (i.e., Q3–Q1).

Differentially Expressed Genes Analysis

For the expression analysis, datasets coded GSE50442 and GSE67066 for PCPG, GSE21354 and GSE26576 for low-grade glioma, and GSE13276 and GSE50161 for glioblastoma were downloaded from the GEO database. Data on pheochromocytoma and high-grade gliomas (other than glioblastoma) were excluded from the analysis. Coding was performed in RStudio using version 4.3.0 of the R programming language and the "limma" method. In each dataset, the top 100 genes with the most significant decrease in expression and the top 100 genes with the most significant increase in expression, compared to normal tissues, were identified. The intersection of these results across datasets was then analyzed. A flow diagram of the study process is shown in Figure 3.

Survival and Carcinogenesis-Related Function Analysis

Overall survival analysis for PCPG, LGG, and GBM tumors was conducted in RStudio using version 4.3.0 of the R programming language. Survival analyses were performed for the PCPG, LGG, and GBM cancer cohorts in relation to the expression of common driver genes and genes with

significantly altered expression identified through the driver gene prediction and differential expression analyses. At this stage, the GEPIA2 tool was utilized (10).

In addition to overall survival analysis, the carcinogenesis-related gain-of-function of the relevant genes was investigated. Using the CancerSEA database and web application (16), functional correlations between genes with the most significantly altered expression and carcinogenesis in PCPG, LGG, and GBM tumors were analyzed using datasets GSE84465 and GSE102130.

Protein–Protein Interaction Analysis

Interactions between genes associated with carcinogenesis, identified through driver gene prediction and expression analyses, were obtained using the widely known web-based STRING software (9). STRING constructs protein interaction networks based on both physical and functional associations of proteins, derived from known relationships and predictions.

RESULTS

Common Candidate Driver Genes for PCPG, LGG, and GBM

The PCPG dataset in the TCGA database contains 1,946 somatic mutation records, while the LGG and GBM datasets contain 32,780 and 55,177 records, respectively. After running the algorithm, 14 candidate driver genes were identified for PCPG, 26 for LGG, and 55 for GBM. The genes with the highest weighted mutation values are shown in Table I.

Four common driver genes were identified across all three tumors: ATRX, NF1, TTN, and MUC16. Ten common driver genes were found between LGG and GBM, including TP53, PTEN, EGFR, PIK3CA, PIK3R1, and SYNE1. Graphs showing the expression levels of these ten genes in tumor and normal tissues within TCGA cancer cohorts were generated using the GEPIA2 web application and are presented in Figures 4, 5, and 6. Analysis of these graphs shows that, except for MUC16, these genes are associated with poor prognosis in many cancer types, not just PCPG, LGG, and GBM.

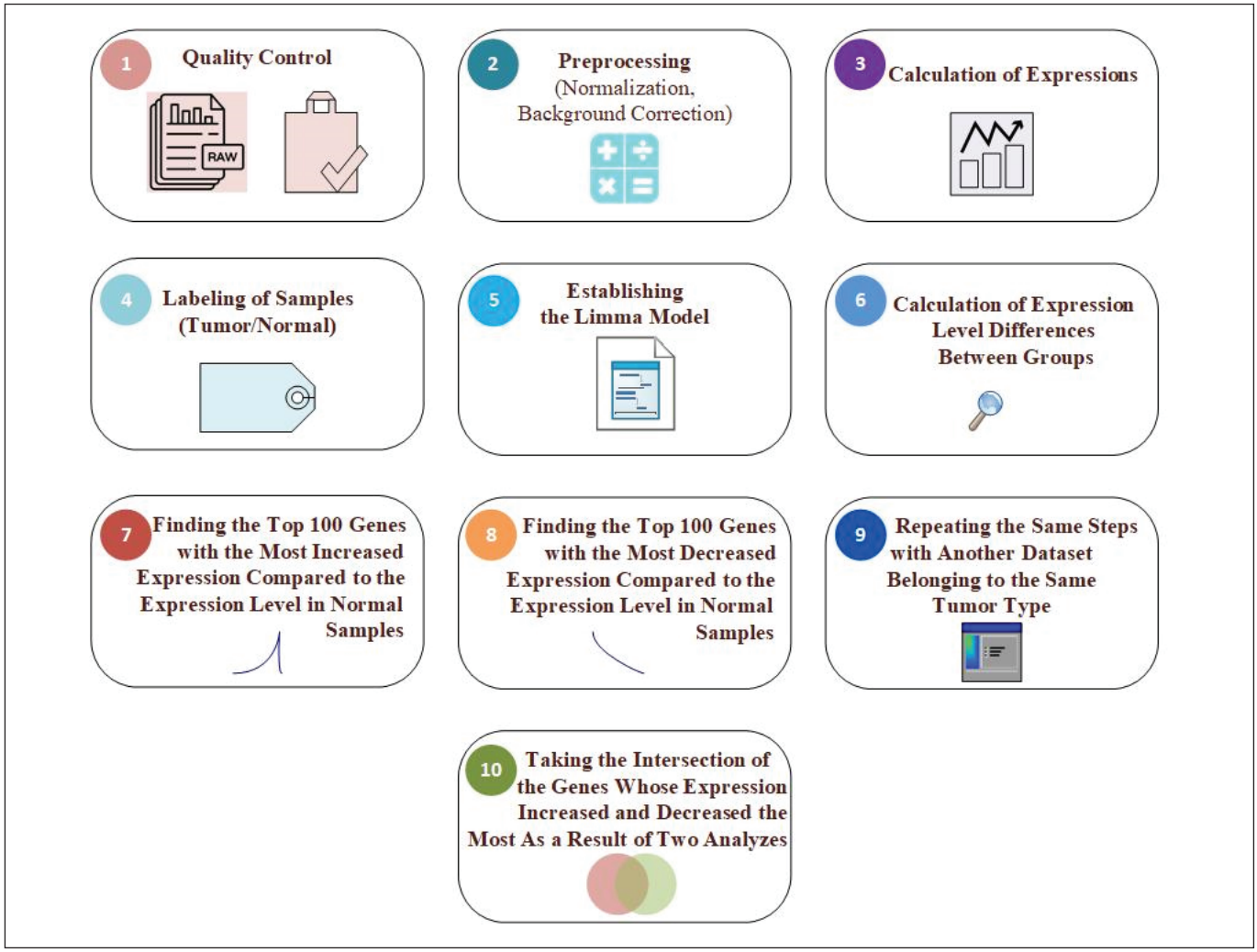


Figure 3: Flow diagram of the DEG analysis study.

Table I: Candidate 30 driver genes identified by our algorithm in PCPG, LGG and GBM

PCPG			LGG			GBM		
Gene	**TWMS**	**NFSLM**	**Gene**	**TWMS**	**NFSLM**	**Gene**	**TWMS**	**NFSLM**
NF1	25.87	16	TP53	433.19	61	TP53	250.86	61
HRAS	12.4	0	ATRX	321.96	173	PTEN	238.37	173
EPAS1	10.74	0	IDH1	286.91	0	TTN	206.12	0
ATRX	10.09	6	CIC	197.56	66	EGFR	182.43	66
RET	7.22	0	TTN	103.77	12	NF1	101.44	12
SCRIB	6.45	1	FUBP1	72.34	45	MUC16	100.81	45
MUC16	6.13	1	NF1	62.96	36	LRP2	72.39	36
CSDE1	5.83	4	NOTCH1	60.25	11	ATRX	72.14	11
TGDS	4.03	0	PIK3CA	47.51	1	RYR2	70.86	1
TTN	3.73	1	EGFR	42.07	1	RB1	64.44	1

TWMS: *Total weighted mutation score,* ***NFSLM:*** *Number of Frameshift and start/stop lost mutations.*

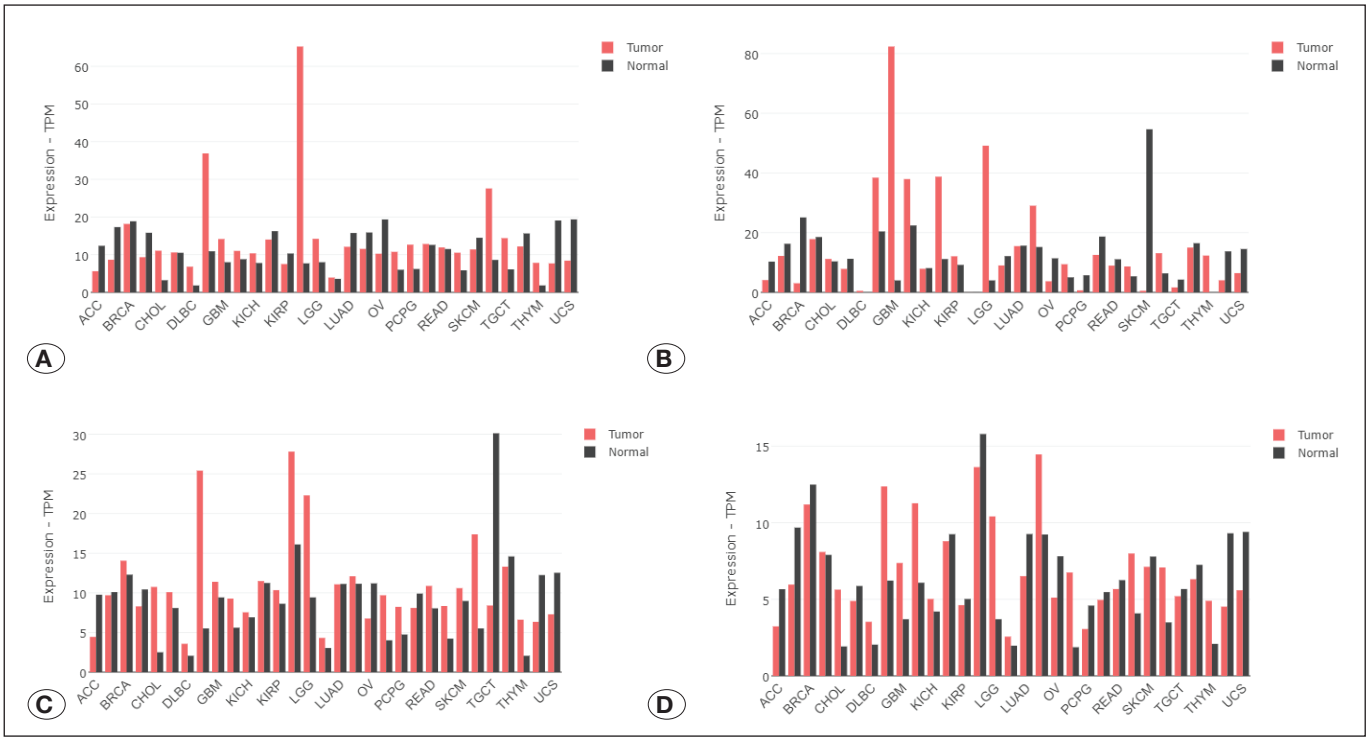


Figure 4: Expression levels of **A)** ATRX, **B)** EGFR, **C)** NF1 and **D)** PIK3CA genes in tumor and normal tissues.

Differentially Expressed Genes Analysis

The Limma method was employed to identify the top 200 genes (comprising the top 100 positively and the top 100 negatively expressed genes) whose expression changed the most in tissue samples from patients with PCPG, LGG, and GBM. Two separate expression datasets for each cancer cohort were analyzed, and the intersection of the results was determined. Common genes across PCPG, LGG, and GBM were identified, along with the intersections between PCPG-LGG, PCPG-GBM, and LGG-GBM, which were also analyzed. As a result, one gene with the highest increased expression (LPL) and three genes with the highest decreased expression (FSTL5, GABRG2, and VSNL1) were found to be common across all three tumor types (see Tables II and III). Graphs illustrating the expression levels of these genes in tumor and

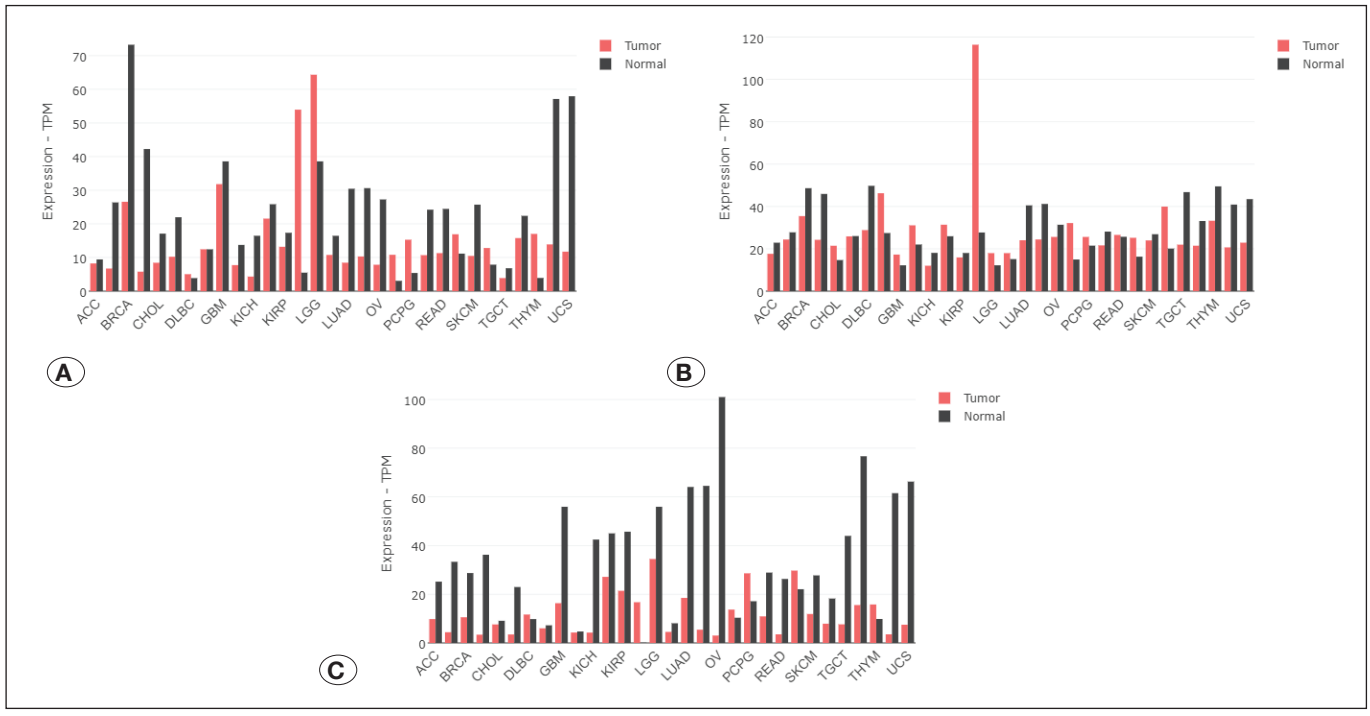


Figure 5: Expression levels of **A)** PIK3R1, **B)** PTEN and **C)** SYNE1 genes in tumor and normal tissues.

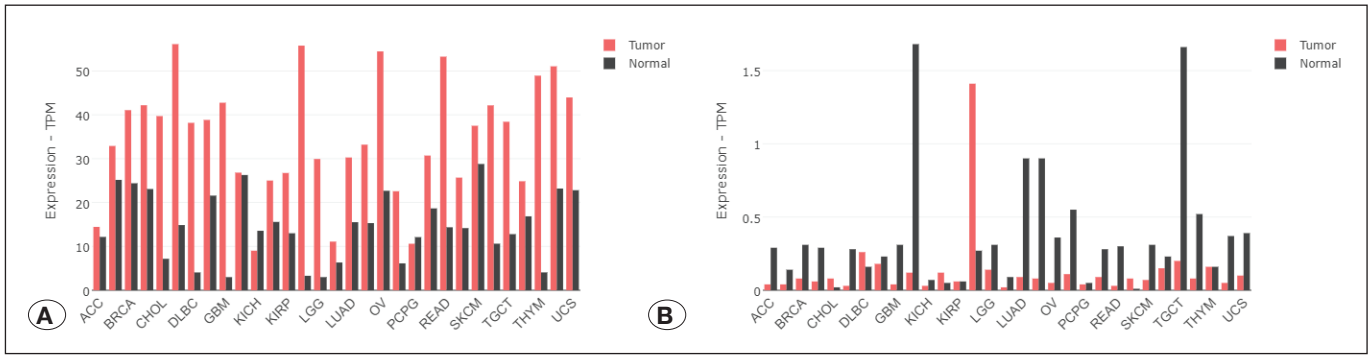


Figure 6: Expression levels of **A)** TP53 and **B)** TTN genes in tumor and normal tissues.

Table II: Common Genes with the Most Decreased Expression for PCPG, LGG and GBM

Gene	logFC_PCPG	logFC_LGG	logFC_GBM
FSTL5	-4.44	-3.66	-4.3
GABRG2	-2.13	-4.05	-5.49
VSNL1	-1.6	-4.2	-2.55

Table III: Common Genes with the Most Increased Expression for PCPG, LGG and GBM

Gene	logFC_PCPG	logFC_LGG	logFC_GBM
LPL	1.85	3.9	2.45

normal tissues from TCGA cancer cohorts were generated using the GEPIA2 web application and are presented in Figure 7. Analysis of these graphs indicates that LGG and GBM are the tumor types with the most significant expression change rates between different groups. The binary cluster intersections revealed a greater number of genes, with the highest intersection rate observed between LGG and GBM. In this cluster, 60 genes exhibited the most significant decrease in expression, while 12 genes showed the most significant increase. Overall, it has been observed that gene expressions tend to change more significantly in the direction of decrease.

Overall Survival Analysis for PCPG, LGG, and GBM Tumors

Upon examining the overall survival analysis curves for all three tumor types, it was found that PCPG had the highest overall survival rate, while GBM exhibited the lowest. Survival

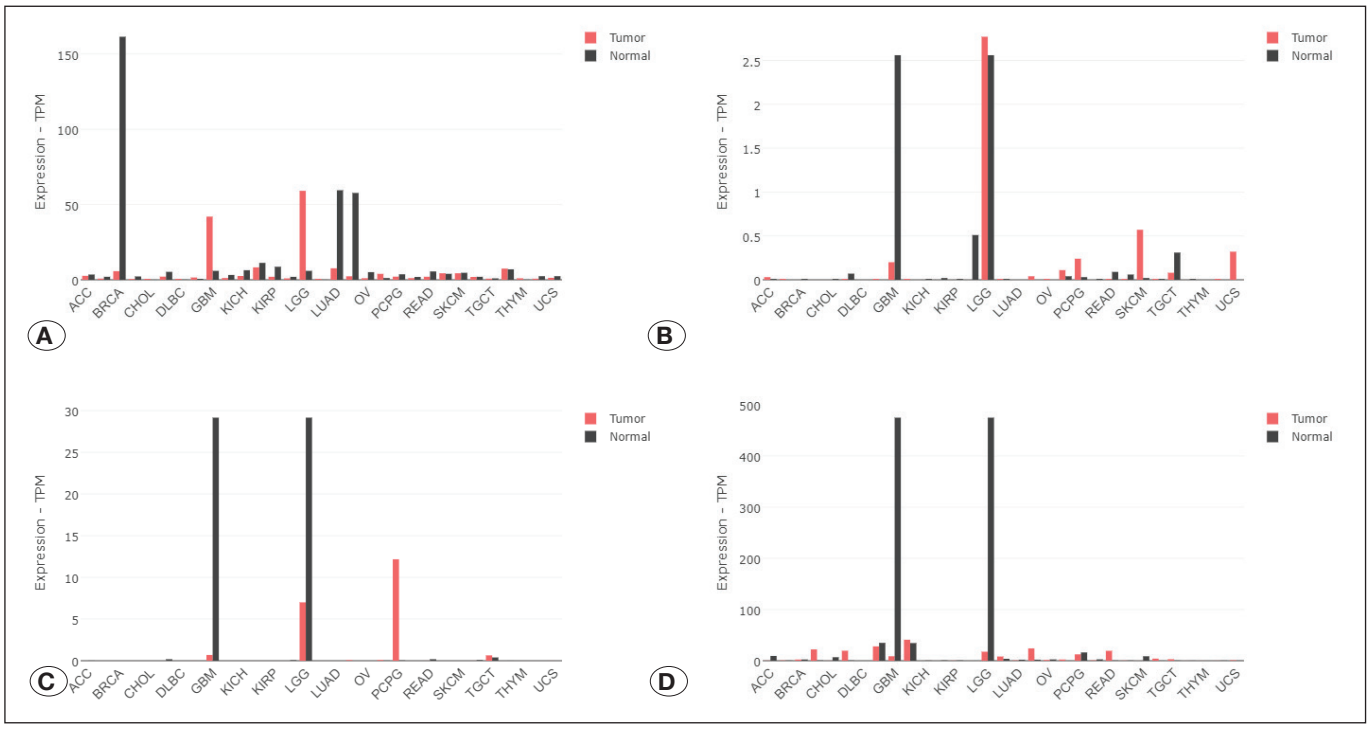


Figure 7: Expression levels of **A)** LPL, **B)** FSTL5, **C)** GABRG2 and **D)** VSNL1 genes in tumor and normal tissues.

analyses were conducted to evaluate the prognostic roles of the common genes identified in our driver gene prediction and expression analysis study within the relevant cancer cohorts. The results indicated that genes with significantly decreased expression (FSTL5, GABRG2, and VSNL1) and low expression of the candidate driver genes ATRX, NF1, and TTN are associated with lower survival rates. Conversely, high expression of the LPL gene, which showed significant increases across all three tumor types, was also associated with lower survival rates (Figure 8). Additionally, low expression of PTEN and SYNE1, common driver gene candidates for LGG and GBM, correlated with lower survival rates. High expression levels of TP53, EGFR, PIK3CA, and PIK3R1 were likewise associated with reduced survival rates (Figure 9).

Single-Cell Function Analysis for Common Genes with the Most Changed Expression Rate

Using CancerSEA, a database that shows different functional states of cancer cells at the single-cell level, we conducted functional analyses of the common genes (LPL, FSTL5, GABRG2, and VSNL1) that exhibited the most significant changes in expression in PCPG, LGG, and GBM tumors. The analysis revealed that these genes positively correlate with epithelial-mesenchymal transition (EMT), invasion, cell cycle, and metastasis functions in LGG (Figure 10A). In GBM, these correlations were even stronger, with the genes also positively linked to inflammation, angiogenesis, hypoxia, and apoptosis functions (Figure 10B).

Protein Interactions Between Candidate Driver Genes

Protein interactions among common candidate driver genes identified for PCPG, LGG, and GBM tumors, as well as those for LGG and GBM, were obtained using STRING software (Figure 11). Developed by a consortium of several academic institutions, the STRING database encompasses both physical and functional relationships of protein interactions, with data weighted and integrated using a confidence score for all interactions.

Examination of the interaction network reveals a high correlation between the genes involved, indicating strong interactions among them.

DISCUSSION

Malignant nervous system tumors are prevalent worldwide and present significant treatment challenges (7). This study utilized comprehensive bioinformatics analyses to identify common mutational and expression changes in paraganglioma, low-grade glioma, and glioblastoma, which are types of CNS tumors. Driver genes were predicted using a molecular clustering algorithm applied to somatic mutation data, while differential expression analyses were conducted using gene expression data. The intersections of the results from these analyses for each tumor type were examined. Our findings indicate that the driver genes ATRX, NF1, and TTN, which exhibit significant mutations, along with the genes LPL, FSTL5, GABRG2, and VSNL1, which show the most pronounced changes in expression, contribute to carcinogenesis by promoting cell proliferation across all three tumor types. Additionally, we observed that the common candidate driver genes identified for PCPG, LGG, and GBM, as well as those found specifically for LGG and GBM, demonstrate intense physical and functional interactions with one another. A noteworthy outcome of this study is that the expressions of common genes, which

significantly change, vary almost exclusively in nervous system tumors; conversely, the expression of common candidate driver genes has undergone substantial changes across many tumor types.

Another important finding is the positive correlation between the common genes with the most significant changes in expression and carcinogenic processes, such as epithelial-mesenchymal transition, metastasis, and invasion, particularly evident in the single-cell function analyses for LGG and GBM. Although these genes do not exhibit a significant

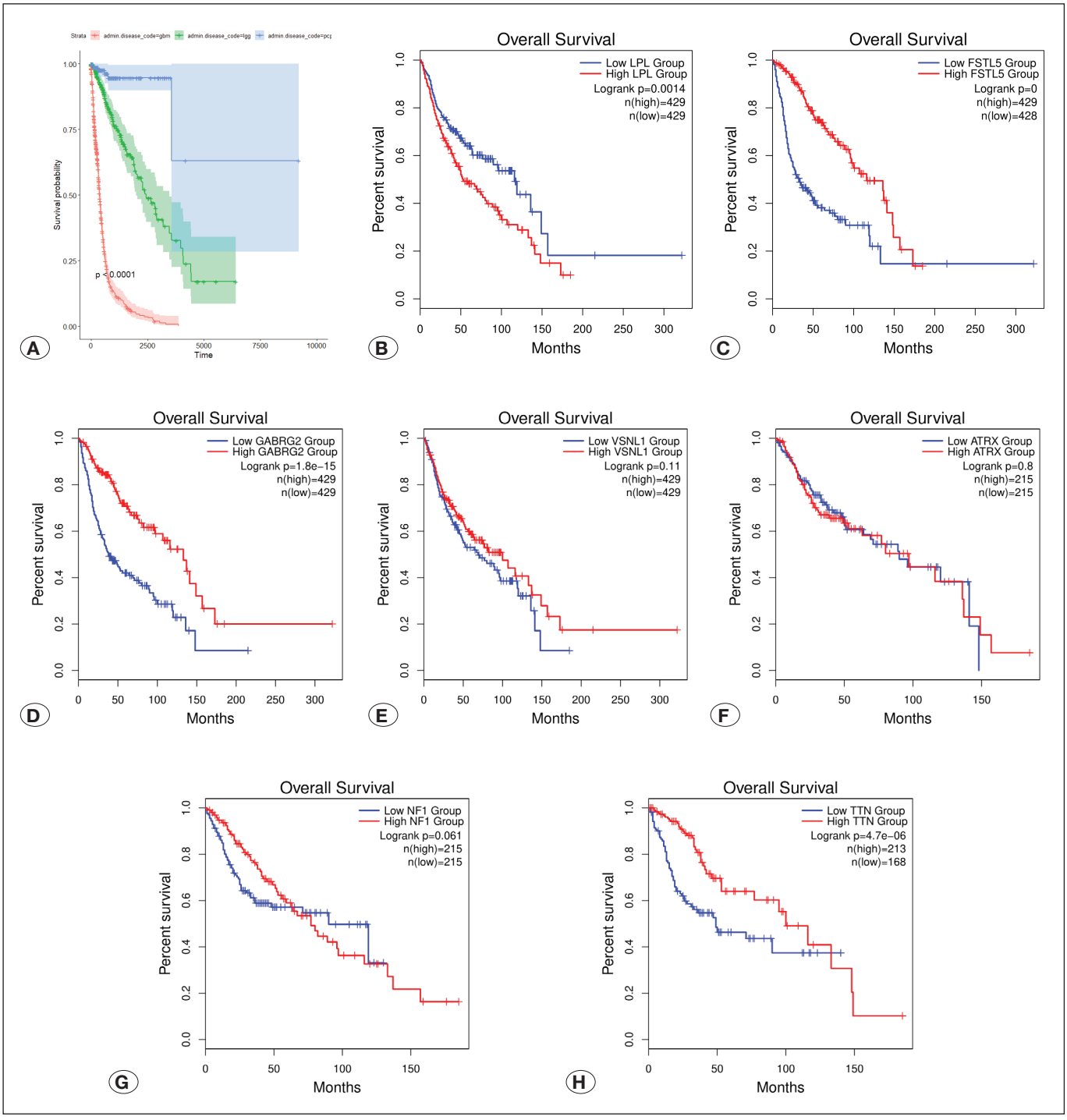


Figure 8: Overall survival curve of **A)** PCPG, LGG and GBM, **B)** LPL, **C)** FSTL5, **D)** GABRG2, **E)** VSNL1, **F)** ATRX, **G)** NF1 and **H)** TTN genes in PCPG, LGG and GBM obtained with Kaplan-Meier method.

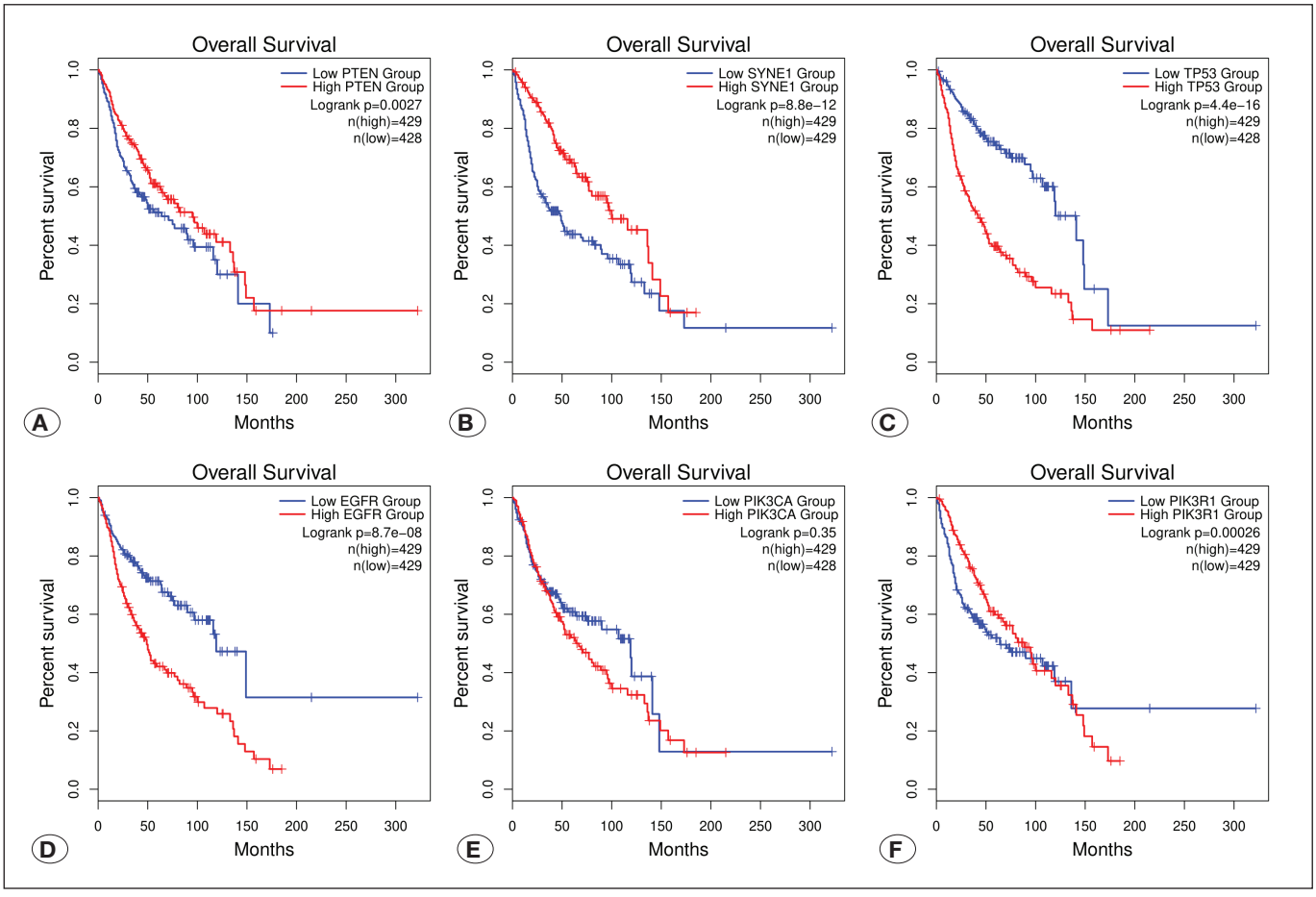


Figure 9: Overall survival curve of **A)** PTEN, **B)** SYNE1, **C)** TP53, **D)** EGFR, **E)** PIK3CA and **F)** PIK3R1 gene in LGG and GBM obtained with Kaplan-Meier method.

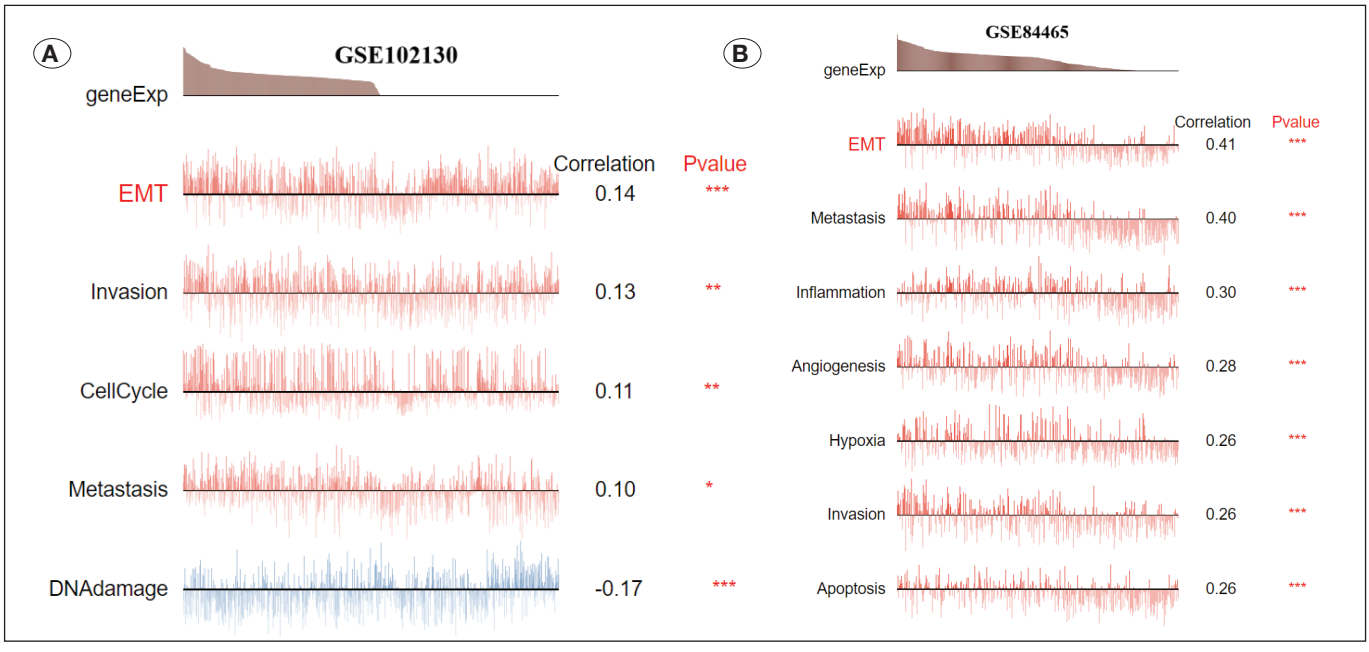


Figure 10: Single cell function analysis of LPL, FSTL5, GABRG2 and VSNL1 expressions in **A)** GSE102130 **B)** GSE84465 coded datasets (*** p<0.01, **EMT:** epithelial-mesenchymal transition).

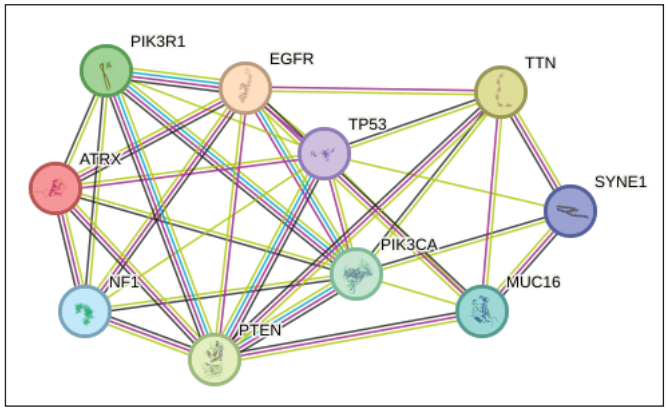


Figure 11: Protein interactions between common candidate driver genes (pink, blue edges: known interactions; green, red, purple edges: predicted interactions).

mutation burden according to the somatic mutation data we analyzed, their expression changes appear to be influenced by epigenetic mechanisms.

CONCLUSION

Our findings confirmed the hypothesis that there may be some common molecular changes in the development of PCPG, LGG, and GBM tumors with different grades and identified which genes may have prognostic value in carcinogenesis for these three tumors. It is hoped that the mutational and expression patterns revealed by this study will provide valuable information in the diagnosis and treatment of CNS tumors.

ACKNOWLEDGMENTS

The article did not receive ethical approval according to national guidelines because it did not involve living subjects (https://trdizin.gov.tr/wp-content/uploads/2022/04/TRDizin_etik_ilkeleri_akis_semasi.pdf).

Declarations

Funding: This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Availability of data and materials: The datasets generated and/or analyzed during the current study are available from the corresponding author by reasonable request.

Disclosure: The authors declare no competing interests.

AUTHORSHIP CONTRIBUTION

Study conception and design: SA
Data collection: SA
Analysis and interpretation of results: SA
Draft manuscript preparation: SA
Critical revision of the article: GO
Other (study supervision, fundings, materials, etc...): EG
All authors (SA, GO, EG) reviewed the results and approved the final version of the manuscript.

REFERENCES

1. Ada NG, Yalcin C, Orhan B, Pınar IE, Ersal T, Ozkocaman V, Ozkalemkas F, Ali R: HIV related primary central nervous system lymphoma: A case report. Turkish J Internal Medicine 3 (Supplement 1):77-78, 2021. https://doi.org/ 10.46310/tjim.876945.
2. Bicer A: Computer-aided automated grading of brain tumors on MR spectroscopy signals (Unpublished dissertation), Bilecik: Bilecik Şeyh Edebali University, 2019: 1-77
3. Chang P, Grinband J, Weinberg BD, Bardis M, Khy M, Cadena G, Su MY, Cha S, Filippi CG, Bota D, Baldi P, Poisson LM, Jain R, Chow D: Deep learning convolutional neural networks accurately classify genetic mutations in gliomas. Am J Neuroradiol 39:1201-1207, 2018. https://doi.org/ 10.3174/ajnr.A5667.
4. Clough E, Barrett T: The gene expression omnibus database. Methods Mol Biol 93-110, 2016. https://doi.org/10.1007/978-1-4939-3578-9_5.
5. Joo B, Ahn SS, An C, Han K, Choi D, Kim H, Park JE, Kim HS, Lee SK: Fully automated radiomics-based machine learning models for multiclass classification of single brain tumors: Glioblastoma, lymphoma, and metastasis. J Neuroradiol 50:388-395, 2023. https://doi.org/10.1016/j.neurad.2022.11.001.
6. Kulcu N: Investigation of gene expression changes and pathways that play a role in the progression of glial brain tumors (Unpublished dissertation), Kocaeli: Kocaeli Unıversity, 2021:1-131
7. Mete O, Asa SL, Gill AJ, Kimura N, Krijger RR, Tischler A: Overview of the 2022 WHO classification of paragangliomas and pheochromocytomas. Endocr Pathol 33:90-114, 2022. https://doi.org/10.1007/s12022-022-09704-6.
8. Ozgiray E, Caliskan KE, Cagli MS, Yurtseven T, Ertan Y, Akalin T, Sanli U, Kamer S, Anacak Y, Haydaroglu A: Ege Üniversitesi Hastanesi merkezi sinir sistemi tümörlerinin epidemiyolojik ve genel sağ kalım özellikleri. Ege J Med 58:10-18, 2019. https://doi.org/10.19161/etd.668489.
9. Szklarczyk D, Kirsch R, Koutrouli M, Nastou K, Mehryary F, Hachilif R, Gable AL, Fang T, Doncheva N, Pyysalo S, Bork P, Jensen LJ, Mering CV: The STRING database in 2023: protein–protein association networks and functional enrichment analyses for any sequenced genome of interest. Nucleic Acids Res 51:D638-D646, 2023. https://doi.org/10.1093/nar/gkac1000.
10. Tang Z, Kang B, Li C, Chen T, Zhang Z: GEPIA2: An enhanced web server for large-scale expression profiling and interactive analysis. Nucleic Acids Res 47:W556-W560, 2019. https://doi.org/10.1093/nar/gkz430.
11. Tasci E, Zhuge Y, Kaur H, Camphausen K, Krauze AV: Hierarchical voting-based feature selection and ensemble learning model scheme for glioma grading with clinical and molecular characteristics. Int J Mol Sci 23:14155, 2022. https://doi.org/10.3390/ijms232214155.
12. Thierheimer M, Cioffi G, Waite KA, Kruchko C, Ostrom QT, Barnholtz-Sloan JS: Mortality trends in primary malignant brain and central nervous system tumors vary by histopathology, age, race, and sex. J Neurooncol 162:167-177, 2023. https://doi.org/10.1007/s11060-023-04279-6.

13. Unal MM, Aslan A: Multiple spinal meningiomas: A case report of a rare entity. GMJ 32:442-447, 2021. https://doi.org/10.12996/gmj.2021.100.
14. Weinstein JN, Collisson EA, Mills GB, Shaw KRM, Ozenberger BA, Ellrott K, Shmulevich I, Sander C, Stuart JM: The cancer genome atlas pan-cancer analysis project. Nat Genet 45:1113-1120, 2013. https://doi.org/10.1038/ng.2764.
15. Yang Z, Shi Q, Bao F: A case of an unexpected posterior mediastinal functional paraganglioma: Case report and literature review. BMC Anesthesiol 20:1-5, 2020. https://doi.org/10.1186/s12871-020-01026-6.
16. Yuan H, Yan M, Zhang G, Liu W, Deng C, Liao G, Xu L, Luo T, Yan H, Long Z, Shi A, Zhao T, Xiao Y, Li X: CancerSEA: A cancer single-cell state atlas. Nucleic Acids Res 47:D900-D908, 2019. https://doi.org/10.1093/nar/gky939.
17. Zou YF, Meng LB, He ZK, Hu CH, Shan MJ, Wang DY, Yu X: Screening and authentication of molecular markers in malignant glioblastoma based on gene expression profiles. Oncol Lett 18:4593-4604, 2019. https://doi.org/10.3892/ol.2019.10804.