Bioinformatics analysis of potential core genes for pilocytic astrocytoma

siqi zhang (zhangsiqi199802@163.com)
Shantou University Medical College https://orcid.org/0000-0002-1382-0708
Hongwu Zeng
Shenzhen Children's Hospital

Research Article

Keywords: Bioinformatics, Pilocytic astrocytoma, DEGs

Posted Date: March 10th, 2022

DOI: https://doi.org/10.21203/rs.3.rs-1414043/v1

Abstract

Background

Pilocytic astrocytoma (PA) is the most common primary brain tumor of childhood. Due to its complicated pathogenesis, the choice and timing of adjuvant therapy after tumor treatment are controversial. This study explored and identified potential biomarkers for PA.

Materials and methods

To identify potential core genes in PA that may provide new therapeutic insights, we analyzed three gene chips (GSE50161, GSE66354, GSE86574) screened from the GEO database. Differentially expressed genes (DEG) from the tissues of PA and normal brain were screened using GEO2R. To determine the functional annotation and pathway of DEG, Gene Ontology (GO) and KEGG pathway enrichment analysis were conducted using DAVID database. Protein interactions of DEG were visualized using PPI network on Cytoscape software. Next, 10 Hub nodes were screened from the differentially expressed network using DMNC algorithm on CytoHubba software and subsequently identified as Hub genes. Finally, the relationship between Hub genes and the prognosis of PA patients was described using GEPIA2 survival analysis web tool.

Results

A total of 37 up-regulated and 144 down-regulated genes were identified through microarray analysis. Amongst the 10 Hub genes selected, SLC12A5 and RAB3C are associated with poor prognosis in a Pilocytic astrocytoma based on the survival analysis.

Conclusion

Our study suggests that low expression of SLC12A5 and RAB3C are associated with poor prognosis in PA patients, whether they can be used as a new therapeutic target for PA.

1. Introduction

Pilocytic astrocytomas are the most common primary brain tumor of childhood, accounting for 70–85% of all cerebellar astrocytomas and 15% of all pediatric brain tumors [1]. The term pilocytic refers to the elongated hair-like projections from the neoplastic cells, while the presence of Rosenthal fibers and hyalinization of blood vessels are also characteristic features [2]. However, there can be heterogeneity in the pathological presentation within the same tumor. Since that we need more information, especially the genetic ones related to diagnosis as well as assessment. Recent genomics and proteomics

advancements have enabled the identification of prominent molecular biomarkers. In addition, the availability of free online bioinformatics tools has facilitated the basic theoretical knowledge of cellular immunotherapy and molecular targeted therapy [3]. A sizable number of oncogene microarray results that exhibit variability can be retrieved from the online database. On this basis, a series of screening and statistical processing of the gene data can be used in the identification of potential core genes for PA.

2. Materials And Method

2.1 Data filtering

Gene chips from the GEO database (https://www.ncbi.nlm.nih.gov/geo/) were screened. Using the keywords “pilocytic astrocytoma” to search on the GEO DataSets database, a total of 38084 samples were exited, while three gene expression datasets (GSE50161, GSE66354, GSE86574) were obtained from GEO (Affymetrix GPL570 platform, Human Genome U133 Plus 2.0 Array). They contained information from the tissues of both PA and normal brain.

2.2 Detection of DEGs

GEO2R web tool (https://www.ncbi.nlm.nih.gov/geo/geo2r/) was used to analyze and compare PA and normal brain tissue samples of the three chips. DEGs were determined by using the adjusted P-value < 0.0001 and $|logFC| \geq 5$ as the screening criteria. While genes with $logFC \geq 5$ were determined as up-regulated genes, those with $logFC \leq -5$ were identified as down-regulated genes. The overlapping portions between the up-regulated and down-regulated genes of the three chips were identified using Venn web tool (bioinformatics.psb.ugent.be/webtools/Venn/).

2.3 DEG Enrichment Analysis by GO and KEGG Pathways

Gene Ontology (GO) database was used for identifying target genes and defining their associated functions. Genes are divided into three categories based on gene function, namely cellular component (CC), molecular function (MF), and biological process (BP). Using the KEGG database, various pathways related to the genes have been identified. Both GO and KEGG databases contain functional information about each gene. Enrichment analysis that integrates these functions based on calculations was conducted using DAVID database tool (https://david.ncifcrf.gov/) to determine the functions and pathways enriched by DEGs. The selection criteria were based on $P < 0.001$ with gene counts > 10.

2.4 A Construction of the PPI Network and Modules

The PPI network information was obtained from DEGs via the internet database Search Tool for the Retrieval of Interacting Genes (STRING; http://string-db.org) (version 11.5). Subsequently, a clear illustration of the Protein Interaction Network was demonstrated using Cytoscape software (https://cytoscape.org/) with CytoHubba, which is a plug-in that uses the DMNC algorithm to screen the hub genes, i.e., the Hub nodes where the top 10 linkage degrees in the differential expression network were calculated.

2.5 Survival analysis of 10 Hub genes

GEPIA (http://gepia.cancer-pku.cn/) is an online tool for profiling tumors and expression of the normal gene in The Cancer Genome Atlas (TCGA) database and performing interactive analyses. Survival curves for each Hub gene in LGG patients were plotted using the GEPIA2 online survival analysis tool (http://gepia2.cancer-pku.cn/#survival) and grouped by the median. The hazard ratios were calculated using Cox PH (Proportional Hazards) Model, with 95% confidence intervals represented as dashed lines and months as the axis units. A P value of 0.05 is considered statistically significant.

3. Results

3.1 Screening of DEGs in PA

The current investigation includes 40 PA and 36 normal brain samples. Using the GEO2R online tool to obtain data, we extracted 224, 227, and 481 DEGs from GSE50161, GSE66354, and GSE86574. Following that, Venn diagram software was used to evaluate and detect the common genes between the three datasets. The findings revealed that the three GES had 181 similar DEGs in LGG samples, including 37 upregulated genes (logFC > 5) and 144 genes that were downregulated (logFC < -5) as shown in Fig. 1.

3.2 Enrichment analysis in PA

A functional and pathway enrichment analysis of all 181 common DEGs was performed with DAVID to evaluate their biological classification. Our analysis revealed that DEGs were mainly enriched in chemical synaptic transmission, nervous system development, and calcium ion binding of BP and MF, respectively. In terms of CC, DEGs were found to be mainly enriched in cell junction, synaptic vesicle membrane, synapse, plasma membrane, axon, neuronal cell body, and dendrite. While the KEGG pathway analysis showed that DEGs were mainly enriched in the nicotine addiction pathway, retrograde endocannabinoid signaling pathway, GABAergic synapse pathway, and morphine addiction pathway. The details were shown in Table 1 and Fig. 2.

Table 1
GO and KEGG pathway enrichment analysis of DEGs

Category	Term	Description	Count	P-value
BP term	GO:0007268	Chemical synaptic transmission	19	8.59E-12
BP term	GO:0007399	Nervous system development	13	1.46E-05
CC term	GO:0030054	Cell junction	34	1.33E-20
CC term	GO:0030672	Synaptic vesicle membrane	11	5.80E-11
CC term	GO:0045202	Synapse	14	1.06E-08
CC term	GO:0005886	Plasma membrane	69	6.52E-08
CC term	GO:0030424	Axon	14	1.19E-07
CC term	GO:0043025	Neuronal cell body	14	6.07E-06
CC term	GO:0030425	Dendrite	13	5.59E-05
MF term	GO:0005509	Calcium ion binding	23	5.29E-07
KEGG_Pathway	hsa05033	Nicotine addiction	11	3.01E-12
KEGG_Pathway	hsa04723	Retrograde endocannabinoid signaling	14	1.08E-11
KEGG_Pathway	hsa04727	GABAergic synapse	11	7.85E-09
KEGG_Pathway	hsa05032	Morphine addiction	11	1.54E-08

Table 2
Genes in the top 10 DMNC scores

Gene symbol	Gene description	MCC score
UNC13C	Unc-13 homolog C	0.865371546
GABRA2	Gamma-aminobutyric acid (GABA) A receptor alpha 2	0.865371546
SLC12A5	Solute carrier family 12 member 5	0.828054247
TMEM151B	Transmembrane protein 151B	0.811203617
RIMS1	Regulating synaptic membrane exocytosis protein 1	0.793825472
STX1A	Syntaxin 1A	0.779924797
CALY	Calcyon neuron-specific vesicular protein	0.77915908
PVALB	Parvalbumin alpha	0.770550985
RAB3C	RAB3C, member RAS oncogene family	0.768716331
BEST3	Vitelliform macular dystrophy 2-like protein 3	0.768338555

3.3 Development of the PPI Network and Module Analysis.

DEGs built a PPI network through the STRING database, which was visualized using Cytoscape software (version3.9.0), with 138 nodes and 819 edgese (Fig. 3a). The top 10 Hub genes with DMNC scores were calculated using CytoHubba plugin of Cytoscape (Fig. 3b). Our results showed that Unc-13 homolog C (UNC13C) was the most outstanding gene with DMNC = 0.865371546, followed by Gamma-aminobutyric acid A receptor alpha 2 (GABRA2; DMNC = 0.865371546), Solute carrier family 12 member 5 (SLC12A5; DMNC = 0.828054247), Transmembrane protein 151B (TMEM151B; DMNC = 0.811203617), Regulating synaptic membrane exocytosis protein 1 (RIMS1; DMNC = 0.793825472), Syntaxin 1A (STX1A; DMNC = 0.779924797), Calcyon neuron-specific vesicular protein (CALY; DMNC = 0.77915908), Parvalbumin alpha (PVALB; DMNC = 0.770550985), RAB3C, member RAS oncogene family (RAB3C; DMNC = 0.768716331) and Vitelliform macular dystrophy 2-like protein 3 (BEST3; DMNC = 0.768338555).

3.4 Analysis of Hub DEGs via the GEPIA

To explore the relationship between the top 10 Hub genes and prognosis in PA patients, we plotted the survival curves for each Hub gene in LGG and normal brain samples using the GEPIA2 online survival analysis tool (Fig. 4). The overall survival with four DEGs high/low expression in LGG patients showed significant difference statistically. SLC12A5: Logrank p = 0.0094, HR (high) = 0.62, p(HR) = 0.01, n(high) = 257, n(low) = 257. TMEM151B: Logrank p = 1.2e-05, HR (high) = 0.44, p(HR) = 2e-05, n(high) = 257, n(low) = 257. RAB3C: Logrank p = 1.2e-05, HR (high) = 0.44, p(HR) = 1.9e-05, n(high) = 257, n(low) = 257. BEST3: Logrank p = 0.00062, HR (high) = 0.54, p(HR) = 0.00076, n(high) = 257, n(low) = 257.

4. Discussion

Pilocytic astrocytomas are the most common primary brain tumor of childhood, accounting for 70–85% of all cerebellar astrocytomas and 15% of all pediatric brain tumors [1]. The term pilocytic refers to the elongated hair-like projections from the neoplastic cells, while the presence of Rosenthal fibers and hyalinization of blood vessels are also characteristic features [2]. However, there can be heterogeneity in the pathological presentation within the same tumor. Since that we need more information especially the genetic ones related to diagnosis as well as assessment. It is commonly recognized that mutations in BRAF are considered to have a poor prognosis, and the targeted therapies using inhibitors have made a significant impact on the treatment of PAs [4]. Although large-scale genomic profiling efforts have identified most of the pathogenic mutations that are convergent on the MAPK/ERK signaling pathway, it's now realized that these events are not sufficient to lead to gliomagenesis and clinical progression [5].

To identify more helpful prognostic biomarkers in PAs, we employed bioinformatics tools and analyzed three gene expression datasets (GSE50161, GSE66354, GSE86574). This study included 40 PAs and 36 normal brain tissue samples. We detected 181 often altered DEGs (P-value < 0.0001 and |logFC|> 5) by using GEO2R and Venn tools, including 37 upregulations (logFC > 5) and 144 downregulations (logFC < -5). After constructing a PPI network by STRING and Cytoscape software, the top 10 Hub genes with

DMNC scores were calculated. In addition, through GEPIA2 analysis ($P < 0.001$), we found that four of ten DEGs (SLC12A5, TMEM151B, RAB3C and BEST3) were significantly associated with poor prognosis.

Solute carrier (SLC) family transporters utilize an electrochemical potential difference or an ion gradient for transporting their substrates across biological membranes. More than 300 SLC transporters have been identified and are expressed in key tissues such as the kidney, liver, intestine, and brain, playing crucial roles in maintaining body homeostasis [6]. Among them, SLC12A5 gene subfamily is a type of K+-Cl- cotransporter that maintains neuronal chloride homeostasis, which is related to various central and peripheral nervous system diseases [7]. A Pan-Cancer analysis found that SLC12A5 was distinctly associated with methyltransferases, mismatch repair proteins, TMB, and MSI in human cancers [8], especially a negative association between SLC12A5 expressions and most immunosuppressive molecules was found in LGG. Our study discovered that the downregulations of SLC12A5 are associated with poor OS, which suggests that SLC12A5 may mediate cancer progression by affecting the immune infiltrate in malignancies. Besides, with the results of an analysis in GBM patients that SLC12A5 was one of their core genes [9], we suppose that an indicator of tumor progression may be.

Rab proteins are GTPases involved in all stages of vesicular transport and membrane fusion in mammalian cells. Rab3 isoforms (Rab3A, Rab3B, Rab3C, and Rab3D) are expressed almost exclusively in neurons and secretory cells [10]. One study discovered that RAB3C overexpression promotes tumor metastasis and is associated with poor prognosis in colorectal cancer, through modulating the ability of cancer cells to release IL-6 through exocytosis and activate the JAK2-STAT3 signaling pathway [11]. However, we found a poor prognostic correlation with RAB3C overexpressed based on our bioinformatics analysis, which is worth pursuing further research on it.

5. Conclusion

Our bioinformatics analysis was based on microarray screening of DEGs between PA samples and normal brain tissues from GEO database. Our analysis identified 10 possible PA hub genes, which are UNC13C, GABRA2, SLC12A5, TMEM151B, RIMS1, STX1A, CALY, PVALB, RAB3C, BEST3. Final survival analysis revealed that SLC12A5, TMEM151B, RAB3C and BEST3 lower expression were associated with a poorer prognosis in LGG patients, although further validation is required. Overall, SLC12A5 and RAB3C may be suitable prognostic markers for PA. Our findings may provide new insights into the potential development of therapeutics for PA.

Declarations

Declaration of Conflicting Interest

The authors declared no potential conflicts of interest with respect to the research, authorship, and publication of this article.

Funding

This article is supported by the Sanming Project of Medicine in Shenzhen (Grant No. SZSM202011005)

References

[1] AlRayahi J, Zapotocky M, Ramaswamy V, et al. Pediatric Brain Tumor Genetics: What Radiologists Need to Know. Radiographics. 2018;38(7):2102-2122. doi:10.1148/rg.2018180109

[2] Louis DN, Perry A, Wesseling P, et al. The 2021 WHO Classification of Tumors of the Central Nervous System: a summary. Neuro Oncol. 2021;23(8):1231-1251. doi:10.1093/neuonc/noab106

[3] van de Haar J, Canisius S, Yu MK, et al. Identifying Epistasis in Cancer Genomes: A Delicate Affair. Cell. 2019;177(6):1375-1383. doi:10.1016/j.cell.2019.05.005

[4] Plant-Fox AS, O'Halloran K, Goldman S. Pediatric brain tumors: the era of molecular diagnostics, targeted and immune-based therapeutics, and a focus on long term neurologic sequelae. Curr Probl Cancer. 2021;45(4):100777. doi:10.1016/j.currproblcancer.2021.100777

[5] Milde T, Rodriguez FJ, Barnholtz-Sloan JS, et al. Reimagining pilocytic astrocytomas in the context of pediatric low-grade gliomas. Neuro Oncol. 2021;23(10):1634-1646. doi:10.1093/neuonc/noab138

[6] Liu X. SLC Family Transporters. Adv Exp Med Biol. 2019;1141:101-202. doi:10.1007/978-981-13-7647-4_3

[7] Yang Z, Gong W, Zhang T, Gao H. Molecular Features of Glioma Determined and Validated Using Combined TCGA and GTEx Data Analyses. Front Oncol. 2021;11:729137. Published 2021 Sep 30. doi:10.3389/fonc.2021.729137

[8] Jiang Y, Liao HL, Chen LY. A Pan-Cancer Analysis of SLC12A5 Reveals Its Correlations with Tumor Immunity. Dis Markers. 2021;2021:3062606. Published 2021 Sep 29. doi:10.1155/2021/3062606

[9] Li C, Pu B, Gu L, et al. Identification of key modules and hub genes in glioblastoma multiforme based on co-expression network analysis. FEBS Open Bio. 2021;11(3):833-850. doi:10.1002/2211-5463.13078

[10] Raffaniello RD. Rab3 proteins and cancer: Exit strategies. J Cell Biochem. 2021;122(10):1295-1301. doi:10.1002/jcb.29948

[11] Chang YC, Su CY, Chen MH, et al. Secretory RAB GTPase 3C modulates IL6-STAT3 pathway to promote colon cancer metastasis and is associated with poor prognosis. Mol Cancer. 2017;16(1):135. Published 2017 Aug 7. doi:10.1186/s12943-017-0687-7

Figures

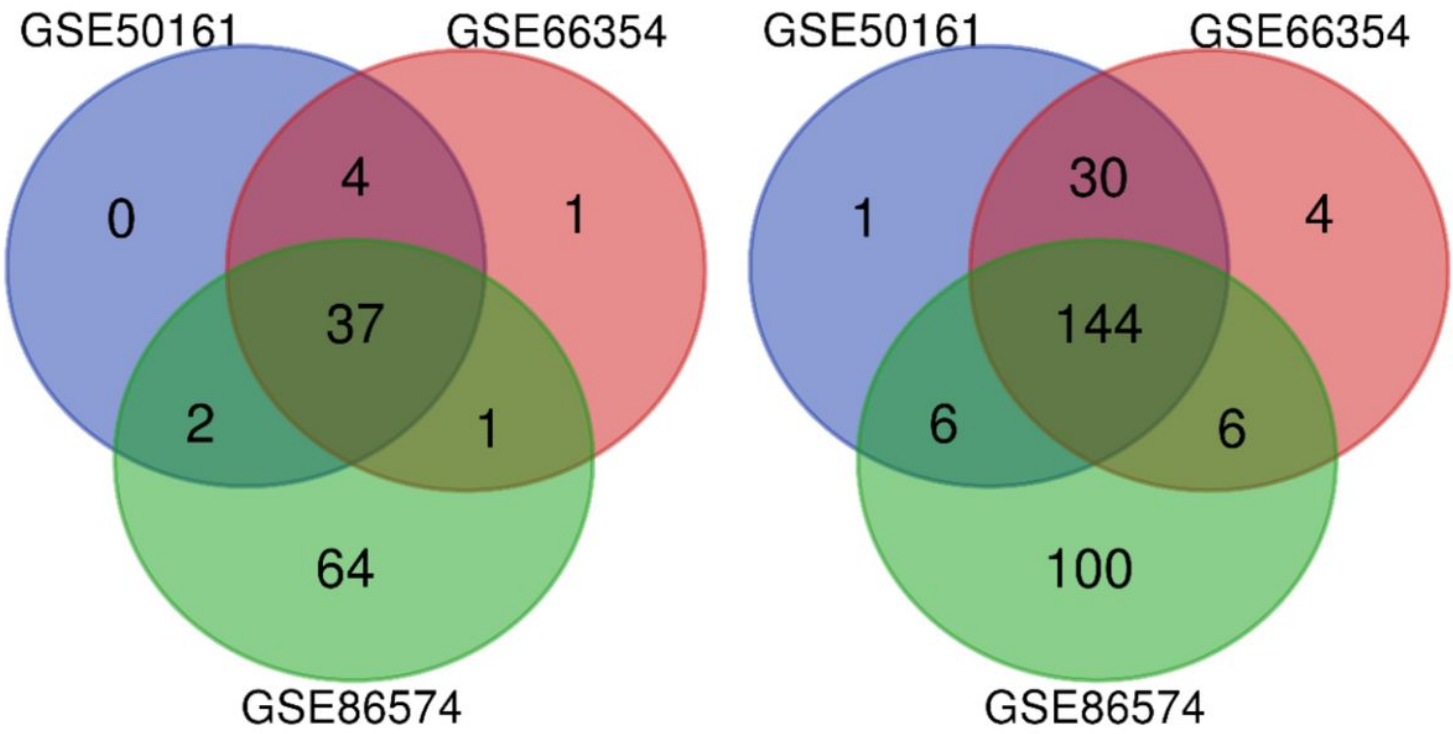

Figure 1

The Venn diagram software identified 181 common genes among the three datasets (GSE50161, GSE66354, and GSE86574). Distinct colors in the figure display different datasets. (a) In the three datasets, 37 DEGs were upregulated (logFC > 5). (b) In the three datasets, 144 DEGs were downregulated (logFC < -5).

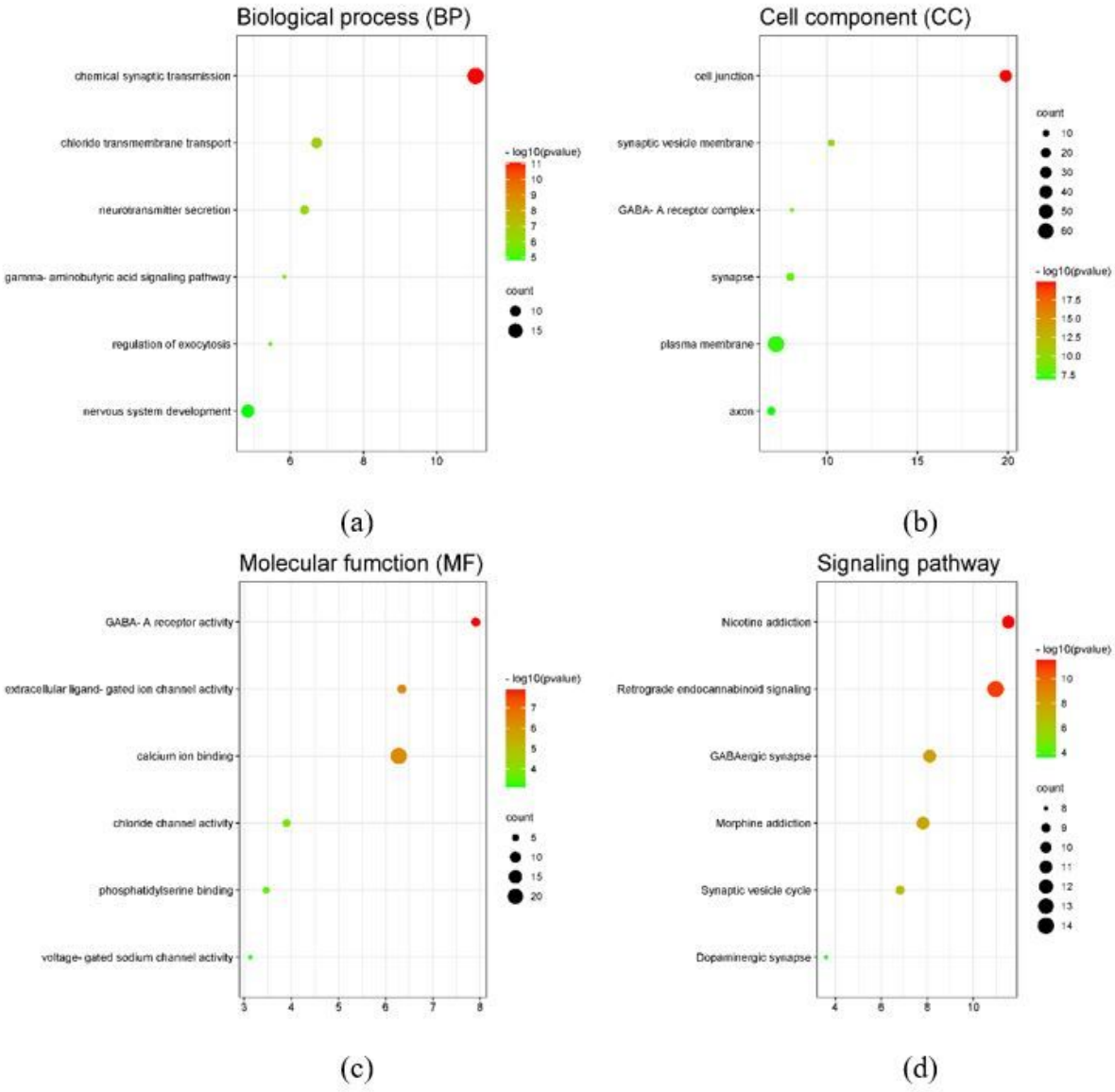

Figure 2

Functional assessment for common upregulated DEGs of three datasets by DAVID: (a) top 6 BP results; (b) top 6 CC results; (c) top 6 MF results; (d) top 6 KEGG results.

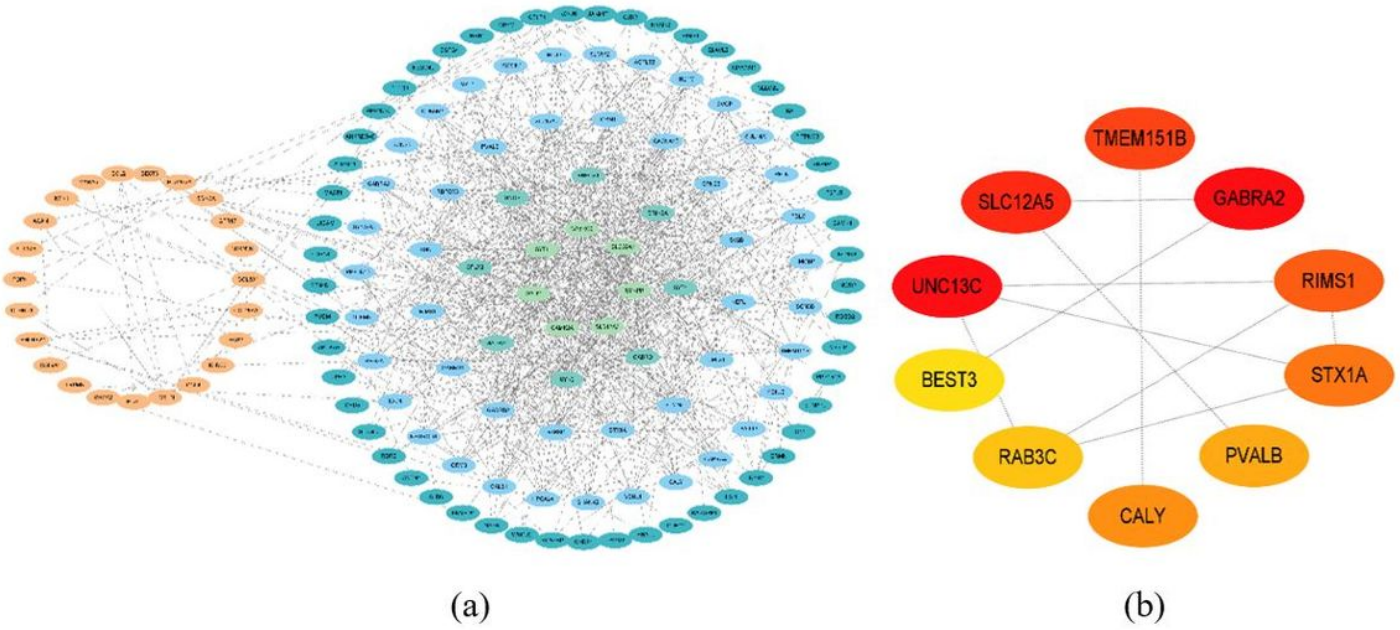

Figure 3

Protein Interaction Network of DEGs by Cytoscape: (a) Orange dots on the left indicate up-regulated genes, blue and green dots on the right indicate down-regulated genes; (b) Genes in the top 10 DMNC scores, the deeper one ranks higher.

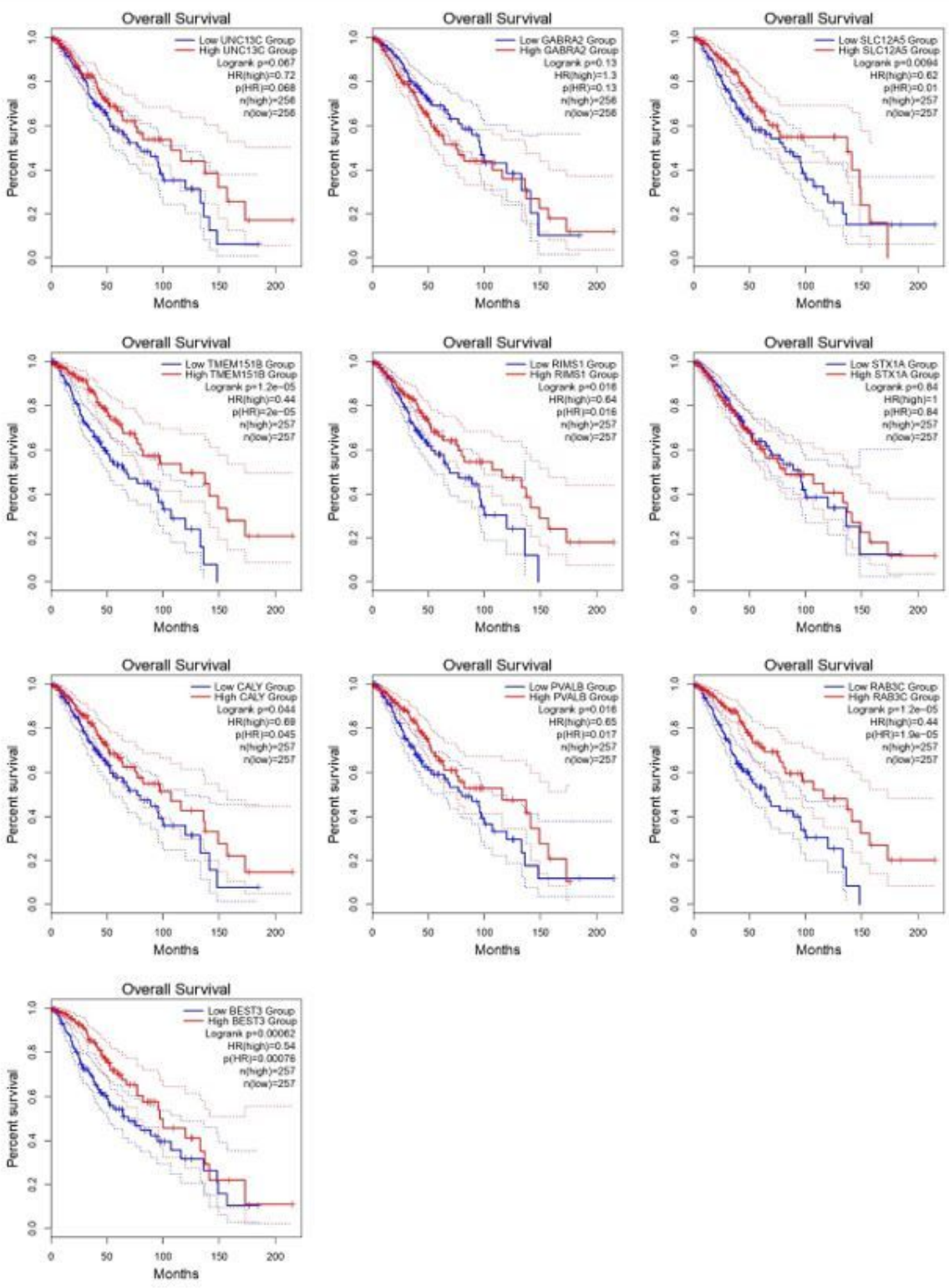

Figure 4

The overall survival curve of the top 10 Hub genes in PA patients mapped using GEPIA2.