

Integrative Analysis of Dna Methylation and Gene Expression Profiles to Explore Potential Biomarkers of Glioblastoma

Yazdan Rahmati

Iran University of Medical Sciences: Tehran University of Medical Sciences

Sajad Najafi

Shahid Beheshti University of Medical Sciences: Shaheed Beheshti University of Medical Sciences

Mohammad-Reza Alivand (✉ mohammadreza_alivand@yahoo.com)

Tabriz University of Medical Sciences

Research

Keywords: glioblastoma, GEO, methylation, gene expression, biomarker, TCGA

Posted Date: March 6th, 2021

DOI: <https://doi.org/10.21203/rs.3.rs-272827/v1>

License:  This work is licensed under a Creative Commons Attribution 4.0 International License. [Read Full License](#)

Abstract

Glioblastoma multiform (GBM) is the most common, most invasive, and malignant type of primary brain tumor with poorly prognosis and poorly survival rate. Using GSE22891 the expression and methylation status of same GBM patients was evaluated to explore key epigenetic genes in GBM. Using $|log2FC| > 1$ and $FDR < 0.05$ as the threshold, DEGs including 4910 downregulated and 2478 upregulated were screened and by $|log2FC| > 0.2$ and $p.value < 0.05$, 3223 DMCs were detected. By merging the results of DEGs and DMCs, 643 genes were selected for network analysis by WGCNA, and based on expression values three modules and by methylation values, one module was selected. Using STRING and Cytoscape, PPI network of genes of all modules were constructed separately. According to the PPI network, core genes were picked out. The expression status of core genes was evaluated using GSE77043, GSE42656, GSE30563, GSE22891, GSE15824, and GSE122498, and 50 genes were validated. The methylation status of 50 genes was explored using GSE50923, GSE22891, and GSE36245, and finally, 12 hub genes including ARHGEF7, RAB11FIP4, PPP1R16B, OLFM1, CLDN10, BCAT1, C1QB, C1QC, IFI16, NUP37, PARP9, and PCALF were selected. Using GEPIA database, the expression and survival plot, and using cBioportal the scatterplot of methylation versus expression of 12 hub genes were extracted based on TCGA. To determine the diagnostic values of the hub genes, the ROC curve and the area under the curve (AUC) were extracted based on GSE22891.

Introduction

According to the world health organization (WHO) classification, gliomas are staged into four grades based on histopathologic features. Glioblastoma multiformes (GBM) is in grade IV of this classification and accounts for about seventy percent of glioma malignancies. Glioblastoma is the most common, most invasive and malignant type of primary brain tumors which poor prognosis is predicted for the patients despite performing therapeutical strategies (surgery, radiotherapy, and chemotherapy) for whom a 10% five-year survival and median survival of 12 to 15 months is reported (1). The glioblastoma is divided into two primary and secondary subtypes according to the origin of the malignancy. Over ninety percent of glioblastomas are of the primary type, which rapidly develops and affects older people. A secondary form of low-grade astrocytoma progresses gradually to glioblastoma and affects younger people (2-5).

The main causative factors resulting in pathogenesis of human various cancers including GBM are epigenetic phenomena. Since development of gene microarray and RNA-seq technologies, the aberrant expression and methylation profiles of GBM both in transcriptome and genome levels have been reported. By employment of the registered expression results in the Gene Expression Omnibus (GEO) web database, Bo et al. (1) identified 431 differentially expressed genes (DEGs) in GBM tissues compared to healthy samples. According to *in situ* studies among these DEGs, 69 genes were known to be associated with considerable prognosis in the patients with GBM. In another experiment using the gene expression profile of GEO series (GSE) GSE50161, of a total 486 DEGs, upregulated CDK1, CCNB1 and CDC20 were showed to be associated with poorly survival in GBM patients (6) Despite many studies and identification of several sets of DEGs, our understanding of the biological mechanisms in pathogenesis of the disease requires further studies.

Epigenetic mechanisms, and in particular DNA hypermethylation, have been shown to play a crucial role in the cancer development and invasion studied in many types of cancers. Hypermethylation of promoter in the CpG island region of DNA repair and tumor suppressor genes result in the silencing of these genes which eventually affects various biological pathways. This is important in many types of cancers including gliomas. The term CpG island methylator phenotype (CIMP) refers to extensive methylation of the CpG islands that were primarily identified in colorectal cancer (CRC) and then in other types of cancer. Many studies have reported hypermethylation of promoter-associated CpG islands in human glioblastoma (2, 3, 7-10). However, the valuable prognostic rate of the mentioned aberrant methylated markers in various subtypes of GBM and the complexity of DNA methylation roles of different gene regions are still controversial issue and it requires more verification in larger independent cohorts of patients with GBM.

Molecular mechanism studies using *in situ* computations has been recognized as a brilliant method in cancer research. Bioinformatics experiments not only can help the molecular pathogenesis of tumors to be explored, but also provides identification of biomarkers for early diagnosis, treatment, and prediction of prognosis in cancers. In the present study weighted gene co-expression network analysis (WGCNA) was used to analyze the relationship between gene sets and phenotypic traits (11, 12). WGCNA is a systems biology procedure that can describe the correlation between genes in high throughput data mining approaches like microarray and next generation sequencing (NGS). WGCNA can constitute clusters from highly correlated genes (modules), find the correlation between modules and also between modules and external clinical data.

In this study, we evaluated the effect of methylation on gene expression by employment of an expression array, and then both upregulated-hypomethylated and downregulated-hypermethylated genes were selected for network analysis using WGCNA. For further validation of the detected genes, their expressions were evaluated using six expression arrays including GEO series GSE77043, GSE42656, GSE30563, GSE22891, GSE15824, and GSE122498. Methylation patterns were checked by application of microarray data from GEO series GSE50923, GSE22891, and GSE36245 arrays and also two datasets of TCGA database (13) and Glioblastoma Multiforme (TCGA, Firehose Legacy). The survival data and box plot expression of the detected genes were evaluated by cBioPrtal and GEPIA, respectively (14) (15). We also used receiver operating characteristic (ROC) curve analysis to evaluate the prognostic power of the genes based on the data from GEO series GSE22891.

Methods

Selecting dataset and explore of differentially expressed genes (DEGs)

GEO series GSE22891 expression array data was recaptured from the GEO database (<https://www.ncbi.nlm.nih.gov/geo/>) that has evaluated gene expression in 40 GBM samples and 6 healthy controls. This dataset also has assessed the DNA methylation patterns in 56 GBM and 4 healthy samples. Expression and methylation data of GSE22891 array were normalized via the quantile normalization function in the limma package (16). Using the aggregate function in the S4 vectors package, which gives an average measure for the probes of each gene, only a single measure for each gene in the expression array

was considered. Limma package was used for detecting the DEGs by the cutoff value > 1 and P-value < 0.05 .

Preprocessing and detecting the DMCs

To explore the CpG islands with the most difference between GBM and control groups, the minfi (17) and limma packages were utilized. First, the methylation β -value of the CpG site was downloaded. Then, it was converted to the methylation M-value (log ratio of beta values) of the CpG. The differentially methylated CpGs (DMCs) between healthy and GBM samples were screened by utilizing the moderated t-tests provided in the limma R -package, under settings of cutoff > 0.2 and P-value < 0.05 to detect the CpG sites.

WGCNA and key modules identification

By merging the DEGs and DMCs of the GSE22891, a list of upregulated-hypomethylated and downregulated-hypermethylated genes with their expression and methylation values was considered for construction of WGCNA. Two WGCNAs were created. Constructing a WGCNA is necessarily accompanied by selecting the soft thresholding power β , which through creating co-expression similarity leads to adjacency calculation. At this point, we intended to demonstrate that utilizing the pickSoftThreshold function has the capacity to both analyze the topology network and assist in selecting an appropriate soft-thresholding power for providing a scale-free topology fit index that reaches values above 0.9. After the calculation of adjacencies, to minimize the effect of noise and spurious associations, adjacency results transformed into Topological Overlap Matrix (TOM). The results of TOM considered as input to produce dendrogram of genes, cutreeDynamic function used for branch cutting. Minimum module size of 30 and the module detection sensitivity deep Split 2 in block wise Consensus Modules function were used for network construction.

Identification of clinically significant modules and functional annotation

The correlation between, and association of individual genes and the clinical data were measured through defining Gene Significance (GS) as the absolute value. For each module, a quantitative criterion of module membership (MM) was considered as the correlation between the eigengene module and gene expression profile through which similarity of all genes on the array was measured and determined to every module. Utilizing the GS and MM, it is possible to explore the interesting module(s) encompassing genes with the great significance of both clinical data and module membership.

Protein-Protein interaction (PPI) Network Construction and enrichment

We constructed the protein-protein interaction (PPI) network of the selected module using STRING (string-db.org) and cytoscape (18). ClueGO v2.5.3 (19) was utilized to perform pathway enrichment analysis of the genes, and the most important signaling pathways based upon KEGG database were detected (20).

Validation of detected genes by other datasets

After construction of PPI network, hub genes were detected and their expression and methylation values were evaluated in other independent datasets. GEO series GSE77043, GSE42656, GSE30563, GSE22891, GSE15824, and GSE122498 were utilized to explore aberrant expression between GBM and healthy controls and also GEO series GSE50923, GSE22891, and GSE36245 were used to consider the methylation status of CpGs related to the hub genes.

Hub gene validation by GEPIA and cBioportal

Overall survival (OS) analysis of the hub genes was performed using the Kaplan-Meier plotter in cBioportal (14). Hazard ratio (HR), 95% confidence intervals (CI) and log-rank P-values were calculated. DNA Methylation versus gene expression of the hub genes were explored using cBioportal. The online tool GEPIA (15) accompanied by data sourced from the Cancer Genome Atlas (TCGA) was used to validate the expression of these hub genes in the GBM.

Comparison of diagnostic values of the detected hub genes using ROC curve

To determine the diagnostic values of the hub genes, we plotted ROC curve and the area under the curve (AUC). We considered P-values of < 0.05 for statistical difference based on GEO seies GSE22891.

Results

Differentially expressed genes (DEGs) screening

Normalization boxplot can be useful for determining if the selected samples are suitable for differential expression analysis. We normalized expression dataset by quantile normalization function in the limma package (Fig. 1A). Uniform manifold approximation and projection (UMAP) was used for visualizing how samples are related to each other (Fig. 1B). Vooma plot was utilized to evaluate if there is a lot of variation in the data (Fig. 1C). Using $|log2FC| > 1$ and $FDR < 0.05$ as the threshold, DEGs including 4910 downregulated and 2478 upregulated were screened (Fig. 1D).

Differentially methylated CpGs (DMCs) screening

In the current study, GSE22891 that includes 56 GBM and 4 control Samples downloaded, normalized and analyzed and finally, DMCs identified. Considering the applied platform in methylation array (Illumina HumanMethylation27 BeadChip), which its Coverage distributed across gene regions. Using cutoff $|log2FC| > 0.2$ and $pvalue < 0.05$, 3223 DMCs were detected. Normalization boxplot (Fig. 2A), Uniform manifold approximation and projection (UMAP) plot (Fig. 2B), Vooma plot (Fig. 2C), and the DMCs of this dataset were shown as volcano plot (Fig. 2D), which has 556 hypomethylated CGs and 886 hypermethylated CGs. By merging DEGs and DMCs, 643 genes were identified and upregulated- hypomethylated and downregulated-hypermethylated were selected as final genes for network analysis using WGCNA (Fig. 3).

Weighted co-expression network construction and key modules identification

Expression and methylation values of final genes considered as input for the WGCNA package. According to **pick Soft Threshold** function power of $\beta = 9$ and 8 were selected for expression and methylation respectively, for providing scale-free topology fit index that reaches values above 0.9, for both datasets. By hierarchical clustering for two separate network analysis, 15 modules were identified for expression values (Fig. 4A) and 18 modules were identified for methylation values (Fig. 5A). These modules exist different sizes in term of the number of genes and labeled by the different colors which show in (Fig. 4B and 5B). Green (correlation= 0.6, P.Value= 3.2e-22) (Fig. 4C), blue (correlation= 0.51, P.Value= 2e-46) (Fig. 4D), and turquoise (correlation= 0.96, P.Value= 1e-200) (Fig. 4E) modules were selected for expression network and yellow module (correlation= 0.41, P.Value= 6.1e-11) (Fig. 5C) for methylation network.

Protein-Protein interaction (PPI) Network Construction

Genes of detected modules were used for PPI network construction in STRING database and were visualized in Cytoscape. PPI network of green (Fig. 6A), blue (Fig. 7A), and turquoise (Fig. 8A) modules identified some genes as core genes related to expression data. This figure was used for genes of yellow module related to methylation data (Fig. 9A). We enriched genes of each module separately, and analyzed them through the Cytoscape plug-in ClueGo based upon KEGG database, genes of green module have role in oocyte meiosis, cell cycle, cellular senescence, p53 signaling pathway, human t-cell leukemia virus 1 infection, mismatch repair, DNA replication, and complement and coagulation cascades (Fig. 6B). For blue module human t-cell leukemia virus 1 infection, central carbon metabolism of cancer, cellular senescence, leukocyte transendothelial migration, age-rage signaling pathway of diabetic complications, and ribosome biogenesis in eukaryotes were the most important signaling pathways (Fig. 7B). Synaptic vesicle cycle, insulin secretion, endocrine and other factor-regulated calcium reabsorption, cAMP signaling pathway, and GnRH secretion were the most dysregulated signaling pathways for genes of turquoise module (Fig. 8B). We also performed gene enrichment for yellow module and NF- κ B signaling, chemokine signaling pathway, pancreatic secretion, basal cell carcinoma, and B cell receptor signaling pathway were detected as most important signaling pathways (Fig. 9B).

Validation of detected genes by other datasets

The expression and methylation status of all genes of PPI networks were evaluated by other separate datasets. The expression status of these genes were evaluated using GSE77043, GSE42656, GSE30563, GSE22891, GSE15824, and GSE122498 and was shown as heatmap (Fig. 10), and the methylation values of all genes were evaluated by GSE50923, GSE22891, and GSE36245 and were visible as heatmap (Fig. 11). The expression evaluation validated 50 gens including 32 upregulated and 18 downregulated genes that seems play a significant role in GBM. The methylation status of mentioned 50 genes approved 12 of them including ARHGEF7, RAB11FIP4, PPP1R16B, OLFM1, CLDN10, BCAT1, C1QB, C1QC, IFI16, NUP37, PARP9, and PCALF as final gene list.

Hub gene validation by GEPIA and cBioportal

12 hub genes were validated by GBM data of GEPIA database. The expression level of all 12 hub genes in TCGA is similar to our results, BCAT1, C1QB, C1QC, IFI16, NUP37, PCALF, and PARP9 are upregulated and

ARHGEF7, CLDN10, RAB11FIP4, PPP1R16B, and OLFM1 are downregulated (Fig. 12). Among them, ARHGEF7, RAB11FIP4, PPP1R16B, OLFM1, CLDN10, BCAT1, C1QB, C1QC, IFI16, NUP37, and PARP9 were negatively associated, while PCALF was positively associated with the overall survival of GBM patients (Fig. 13). Scatterplot of methylation status versus expression status of all 12 hub genes showed negative correlation between GBM and healthy control samples based on TCGA data (Fig. 14).

Comparison of diagnostic values of detected hub genes using ROC curve

A ROC curve was obtained to verify the potent diagnostic performance of 12 hub genes based on the GSE22891 dataset. The AUC showed that ARHGEF7, RAB11FIP4, PPP1R16B, OLFM1, CLDN10, BCAT1, C1QB, C1QC, IFI16, NUP37, PARP9, and PCALF indicated excellent diagnostic efficiency for tumor and normal tissues (Fig. 15).

Discussion

GBM is the most common central nervous system tumor that begins in brain tissue and originates from astrocyte cells (21). The subtypes have been identified for this disease based on gene expression changes including the TP53 (p53), the epidermal growth factor receptor (EGFR), and IDH1 (Isocitrate dehydrogenase 1). Also other genetic alterations such as RB and the PI3K/AKT pathways (22-24) play an important role in GBM. In addition to genetic abnormalities, gliomas also exhibit various types of epigenetic alterations, which are expressed as inherited changes in gene expression, which can be further investigated, such as post-translational changes in histones, and methylation in the CpG islands of DNA sequences. Methylation changes have been observed in the DNA sequences of this disease and many other tumors, including hypermethylation in the CpG islands of DNA, or hypomethylation in specific genes (25, 26). DNMTs are a family of enzymes including DNMT1, DNMT3A and -3B in mammals that play a significant role in the maintenance of DNA methylation patterns. Somatic mutations and altered expression of DNMT genes have been identified in various types of cancers (25-28). However, their significant effect on gene methylation and gene expression pattern remains unclear (29).

In our study, an integrative analysis of methylation and expression data on the same patients using GSE22891 was performed, and genes that were hypomethylated- upregulate and hypermethylated- downregulate were selected for network analysis. By WGCNA, significant modules including green, blue, and turquoise using expression values and yellow module using methylation value were detected. Using STRING and Cytoscape, PPI networks of mentioned modules were constructed and core genes were selected. For validation, expression values of core genes were evaluated using separate datasets including GSE77043, GSE42656, GSE30563, GSE22891, GSE15824, and GSE122498. Due to explore the effect of methylation on the expression of selected genes, their methylation was evaluated by GSE50923, GSE22891, and GSE36245, and 12 hub genes were selected for further consideration. Using GEPIA database, box plot expression and survival data of 12 hub genes based on TCGA database were obtained. By cBioportal, Scatterplot of methylation versus expression status of all 12 hub genes was obtained. ROC curve was used to explore the diagnostic performance of all mentioned genes.

The main genes with expression changes in our signature containing a set of 12 genes are summarized according to the literature in Table 1. We found discordant changes in the expression of a few genes including *ARHGEF7*, *OLFM1*, *PPP1R16B*, and *IFI16*, while other were shown to be altered in GBM samples compared to normal controls in concordant to the past findings in other studies (Table 1). All of the genes in the signature except for *ARHGEF7* with less changes but also significant showed highly significant alterations in their expression levels compared to control samples. Among them, *ARHGEF7* (also called beta Pix or *Cool1*) encodes a guanine-nucleotide exchange factor (GEF) has been found to be upregulated in glioblastoma cell lines playing an essential role in the tumorigenesis of invasive cells and its knockdown was revealed to reverse the promoting impacts on tumor cell proliferation and invasion (30). Methylation status of *ARHGEF7* has been studied in several cell types of GBM and in combination with others is suggested as a signature in classification, prognosis and treatment of GBM disease (31). *BCAT1* encoded protein plays a role in the metabolism of branched-chain amino acids (BCAAs), known to be up-regulated in GBM samples in the past studies (32) and also as a result enhances tumor progression in GBM cells (33). Other genes like *C1QB* and *C1QC* which are involved in complement cascade, *KIAA0101* (or *PCALF*) in cell proliferation, and *PARP9* performing in DNA damage repair have been known in literature to be up-regulated in GBM and promote tumorigenesis in consistent with our results (34-37) while no witness was found for *NUP37* as an encoding gene for Nucleoporin 37; a member of nuclear pore complex (NPC) which also was shown to be upregulated in the signature. On the other side, *CLDN10* encoding claudin 10 in Tight junction interactions and *RAB11FIP4* located on chromosome 17q11.2 and encodes for Rab11 family-interacting protein 4 which plays role in the regulation of endocytic traffic, both were shown to be down-regulated in the signature in consistent with the past findings (38-40), although a controversial result also has been reported for the latter (41).

In conclusion, by integrated analysis of methylation and expression profiles, we identified 12 genes that can be regarded as prognostic biomarkers of GBM. Further experimental analysis is needed to prove the biological importance and functions of the reported genes.

Abbreviations

GBM, Glioblastoma; GEO, Gene Expression Omnibus; GSE, GEO series; TCGA, The Cancer Genome Atlas; WGCNA, Weighted Gene Co-expression Network Analysis; GSA, Global Screening Array; GSE, Gene Series; FC, Fold change; FDR, False Discovery Rate; ROC curve, receiver operating characteristic curve; DMC, Differentially methylated CpG; DEG, differentially expressed genes; CRC, Colorectal Cancer; CIMP, CpG island methylator phenotype; TOM, Topological Overlap Matrix, NGS, Next Generation Sequencing; MM, Module Membership; AUC, Area Under the Curve

Declarations

Ethics approval and consent to participate: Regarding the approved codes of Ethics Committee at Iran University of Medical Sciences as well as the declaration of Helsinki, the written informed consent forms were signed for each sampling.

Consent for publication: Not Applicable

Availability of data and materials: The data that support the findings of this study are available on request from the corresponding author. The data are not publicly available due to privacy or ethical restrictions.

Competing interests: The authors declare no conflict of interest.

Funding: This study was funded by Iran University of medical sciences, grant number 96-03-30-30735.

Authors' contributions: Yazdan Rahmati designed the study. Mohammad Reza Alivand and Sajad Najafi wrote the first draft of the manuscript. Yazdan Rahmatia and Mohammad Reza Alivand supervised the study. All authors read and approved the final version of the manuscript.

Acknowledgements: We would like to express our gratitude to personnel of medical genetic laboratory.

References

1. Lee DH, Ryu H-W, Won H-R, Kwon SH. Advances in epigenetic glioblastoma therapy. *Oncotarget*. 2017;8(11):18577.
2. Hill VK, Shinawi T, Ricketts CJ, Krex D, Schackert G, Bauer J, et al. Stability of the CpG island methylator phenotype during glioma progression and identification of methylated loci in secondary glioblastomas. *2014;14(1):506*.
3. Smith AA, Huang Y-T, Eliot M, Houseman EA, Marsit CJ, Wiencke JK, et al. A novel approach to the discovery of survival biomarkers in glioblastoma using a joint analysis of DNA methylation and gene expression. *2014;9(6):873-83*.
4. Holland EC. Glioblastoma multiforme: the terminator. *Proceedings of the National Academy of Sciences*. 2000;97(12):6242-4.
5. Chen R, Smith-Cohn M, Cohen AL, Colman H. Glioma Subclassifications and Their Clinical Significance. *Neurotherapeutics : the journal of the American Society for Experimental NeuroTherapeutics*. 2017;14(2):284-97.
6. Zhang Y, Xia Q, Lin J. Identification of the potential oncogenes in glioblastoma based on bioinformatic analysis and elucidation of the underlying mechanisms. *Oncology reports*. 2018;40(2):715-25.
7. Etcheverry A, Aubry M, De Tayrac M, Vauleon E, Boniface R, Guenot F, et al. DNA methylation in glioblastoma: impact on gene expression and clinical outcome. *2010;11(1):701*.
8. Hughes LA, Melotte V, De Schrijver J, De Maat M, Smit VT, Bovée JV, et al. The CpG island methylator phenotype: what's in a name? *2013;73(19):5858-68*.
9. Turcan S, Rohle D, Goenka A, Walsh LA, Fang F, Yilmaz E, et al. IDH1 mutation is sufficient to establish the glioma hypermethylator phenotype. *2012;483(7390):479*.
10. Noushmehr H, Weisenberger DJ, Diefes K, Phillips HS, Pujara K, Berman BP, et al. Identification of a CpG island methylator phenotype that defines a distinct subgroup of glioma. *2010;17(5):510-22*.

11. Shi Q, Shen L, Gan J, He L, Lin J, Guo S, et al. Integrative analysis identifies DNMTs against immune-infiltrating neutrophils and dendritic cells in colorectal cancer. 2019;14(4):392-404.
12. Wang Z, Chen X, Zhang D, Cao Y, Zhang L, Tang WJAon. PYCARD Gene Plays a Key Role in Rapidly Progressive Glomerulonephritis: Results of a Weighted Gene Co-Expression Network Analysis. 2018;48(3):193-204.
13. Network CGAR. Comprehensive genomic characterization defines human glioblastoma genes and core pathways. *Nature*. 2008;455(7216):1061.
14. Gao J, Aksoy BA, Dogrusoz U, Dresdner G, Gross B, Sumer SO, et al. Integrative analysis of complex cancer genomics and clinical profiles using the cBioPortal. *Science signaling*. 2013;6(269):pl1-pl.
15. Tang Z, Li C, Kang B, Gao G, Li C, Zhang Z. GEPIA: a web server for cancer and normal gene expression profiling and interactive analyses. *Nucleic acids research*. 2017;45(W1):W98-W102.
16. Smyth GK. Limma: linear models for microarray data. *Bioinformatics and computational biology solutions using R and Bioconductor*: Springer; 2005. p. 397-420.
17. Aryee MJ, Jaffe AE, Corrada-Bravo H, Ladd-Acosta C, Feinberg AP, Hansen KD, et al. Minfi: a flexible and comprehensive Bioconductor package for the analysis of Infinium DNA methylation microarrays. *Bioinformatics*. 2014;30(10):1363-9.
18. Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, et al. Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome research*. 2003;13(11):2498-504.
19. Bindea G, Mlecnik B, Hackl H, Charoentong P, Tosolini M, Kirilovsky A, et al. ClueGO: a Cytoscape plug-in to decipher functionally grouped gene ontology and pathway annotation networks. *Bioinformatics*. 2009;25(8):1091-3.
20. Kanehisa M, Araki M, Goto S, Hattori M, Hirakawa M, Itoh M, et al. KEGG for linking genomes to life and the environment. *Nucleic acids research*. 2007;36(suppl_1):D480-D4.
21. Bleeker FE, Molenaar RJ, Leenstra S. Recent advances in the molecular understanding of glioblastoma. *Journal of neuro-oncology*. 2012;108(1):11-27.
22. Verhaak RG, Hoadley KA, Purdom E, Wang V, Qi Y, Wilkerson MD, et al. Integrated genomic analysis identifies clinically relevant subtypes of glioblastoma characterized by abnormalities in PDGFRA, IDH1, EGFR, and NF1. *Cancer cell*. 2010;17(1):98-110.
23. Bleeker FE, Lamba S, Zanon C, Molenaar RJ, Hulsebos TJ, Troost D, et al. Mutational profiling of kinases in glioblastoma. *BMC cancer*. 2014;14(1):718.
24. Hayden EC. Genomics boosts brain-cancer work. *Nature Publishing Group*; 2010.
25. Pojo M, Costa BM. Molecular hallmarks of gliomas. *Molecular targets of CNS tumors*: IntechOpen; 2011.
26. Garami M. *Molecular targets of CNS tumors*: BoD–Books on Demand; 2011.
27. Roy DM, Walsh LA, Chan TA. Driver mutations of cancer epigenomes. *Protein & cell*. 2014;5(4):265-96.
28. Ibrahim AE, Arends MJ, Silva A-L, Wyllie AH, Greger L, Ito Y, et al. Sequential DNA methylation changes are associated with DNMT3B overexpression in colorectal neoplastic progression. *Gut*. 2011;60(4):499-508.

29. Shi Q, Shen L, Gan J, He L, Lin J, Guo S, et al. Integrative analysis identifies DNMTs against immune-infiltrating neutrophils and dendritic cells in colorectal cancer. *Epigenetics*. 2019;14(4):392-404.
30. Connor K, Murray DW, Jarzabek MA, Tran NL, White K, Dicker P, et al. Targeting the RhoGEF β PIX/COOL-1 in Glioblastoma: Proof of Concept Studies. *Cancers (Basel)*. 2020;12(12):3531.
31. Pangeni RP, Zhang Z, Alvarez AA, Wan X, Sastry N, Lu S, et al. Genome-wide methylomic and transcriptomic analyses identify subtype-specific epigenetic signatures commonly dysregulated in glioma stem cells and glioblastoma. *Epigenetics*. 2018;13(4):432-48.
32. Chaumeil MM, Larson PE, Woods SM, Cai L, Eriksson P, Robinson AE, et al. Hyperpolarized [1-13C] glutamate: a metabolic imaging biomarker of IDH1 mutational status in glioma. *Cancer Res.* 2014;74(16):4247-57.
33. Tönjes M, Barbus S, Park YJ, Wang W, Schlotter M, Lindroth AM, et al. BCAT1 promotes cell proliferation through amino acid catabolism in gliomas carrying wild-type IDH1. *Nature Medicine*. 2013;19(7):901-8.
34. Mangogna A, Belmonte B, Agostinis C, Zacchi P, Iacopino DG, Martorana A, et al. Prognostic Implications of the Complement Protein C1q in Gliomas. *Front Immunol*. 2019;10:2366-.
35. Mitchell D, Chintala S, Fetcko K, Henriquez M, Tewari BN, Ahmed A, et al. Common Molecular Alterations in Canine Oligodendrogloma and Human Malignant Gliomas and Potential Novel Therapeutic Targets. *Front Oncol*. 2019;9:780-.
36. Li G, Luna C, Gonzalez P. miR-183 Inhibits UV-Induced DNA Damage Repair in Human Trabecular Meshwork Cells by Targeting of KIAA0101. *Investigative Ophthalmology & Visual Science*. 2016;57(4):2178-86.
37. Xu H, Chai S, Wang Y, Wang J, Xiao D, Li J, et al. Molecular and clinical characterization of PARP9 in gliomas: A potential immunotherapeutic target. *CNS Neuroscience & Therapeutics*. 2020;26(8):804-14.
38. Yang L, Zeng W, Sun H, Huang F, Yang C, Cai X, et al. Bioinformatical Analysis of Gene Expression Omnibus Database Associates TAF7/CCNB1, TAF7/CCNA2, and GTF2E2/CDC20 Pathways with Glioblastoma Development and Prognosis. *World Neurosurgery*. 2020;138:e492-e514.
39. Mamoor S. Differential expression of RAB11FIP2 and RAB11FIP4 in glioblastoma tumors. 2020.
40. Yu K, Yang H, Lv Q-I, Wang L-c, Tan Z-I, Zhang Z, et al. Construction of a competitive endogenous RNA network and analysis of potential regulatory axis targets in glioblastoma 2020.
41. Li L, Liu X, Ma X, Deng X, Ji T, Hu P, et al. Identification of key candidate genes and pathways in glioblastoma by integrated bioinformatical analysis. *Exp Ther Med*. 2019;18(5):3439-49.
42. Chaumeil MM, Larson PEZ, Woods SM, Cai L, Eriksson P, Robinson AE, et al. Hyperpolarized [1-¹³C] Glutamate: A Metabolic Imaging Biomarker of IDH1 Mutational Status in Glioma. *Cancer Research*. 2014;74(16):4247-57.
43. Zhang C, Han L, Zhang A, Yang W, Zhou X, Pu P, et al. Global changes of mRNA expression reveals an increased activity of the interferon-induced signal transducer and activator of transcription (STAT) pathway by repression of miR-221/222 in glioblastoma U251 cells. *Int J Oncol*. 2010;36(6):1503-12.
44. Pollo B, Mazibrada J, Calatozzolo C, Cacciatore F, Spinello S, Grgenti V, et al. P06.15 IFI16 expression in gliomas and its potential role in immuno-surveillance. *Neuro Oncol*. 2016;18(Suppl 4):iv31-iv.

45. Zhang M, Lv X, Jiang Y, Li G, Qiao Q. Identification of aberrantly methylated differentially expressed genes in glioblastoma multiforme and their association with patient survival. *Exp Ther Med*. 2019;18(3):2140-52.
46. Hopp L, Willscher E, Löffler-Wirth H, Binder H. Function Shapes Content: DNA-Methylation Marker Genes and their Impact for Molecular Mechanisms of Glioma. *Journal of Cancer Research Updates*. 2015;4:127-48.
47. Shai RM, Reichardt JK, Ya-Hsuan H, Kremen TJ, Liau LM, Cloughesy TF, et al. Robustness of gene expression profiling in glioma specimen samplings and derived cell lines. *Molecular brain research*. 2005;136(1-2):99-103.
48. Mamoor S. Differential expression of protein phosphatase PPM1B and multiple protein phosphatase regulatory subunits in glioblastoma. 2020.

Table

Table 1: Function of final genes

Gene Symbol	Location On chromosome	General Protein Function	Expression Change in GBM	Role in GBM tumorigenesis	Methylation Status in GBM
<i>ARHGEF7</i>	13q34	A guanine-nucleotide exchange factor (GEF) for GTP-binding proteins	Up-regulated (30)	Promotion (30)	Hypomethylated (31)
<i>BCAT1</i>	12p12.1	Branched-chain amino acids (BCAAs) metabolism	Up-regulated (42)	Promotion (33)	Hypermethylated <i>BCAT1</i> promotor 1 and hypomethylated promotor 2 (33)
<i>C1QB</i>	1p36.12	Regulation of complement cascade	Up-regulated (34)	Promotion (34)	N/A
<i>C1QC</i>	1p36.12	Regulation of complement cascade	Up-regulated (34)	Promotion (34)	N/A
<i>CLDN10</i>	13q32.1	Tight junction interactions	Down-regulated (38)	N/A	N/A
<i>IFI16</i>	1q23.1	Transcriptional regulation, cellular proliferation, autophagy regulation, immunomodulation, and inflammation	Down-regulated (43)	Suppression (44)	Hypomethylated (45)
<i>KIAA0101 (PCALF)</i>	15q22.31	Regulation of cell proliferation and DNA repair during replication	Up-regulated (35)	Promotion (36)	Hypermethylated in IDH (46)
<i>NUP37</i>	12q23	Assembly of a functional nuclear pore complex (NPC) and chromosome segregation during mitosis	N/A	N/A	N/A
<i>OLFM1</i>	9q34.3	Axonal growth regulation	Up-regulated (47)	N/A	N/A
<i>PARP9</i>	3q21.1	DNA damage repair	Up-regulated (37)	Promotion (37)	N/A
<i>PPP1R16B</i>	20q11.23	Regulation of angiogenesis and proliferation of endothelial cells	Up-regulated (48)	Promoting (48)	N/A

<i>RAB11FIP4</i>	17q11.2	Regulation of endocytic traffic	Down-regulated (39, 40) or Up-regulated (41)	N/A	N/A
------------------	---------	---------------------------------	--	-----	-----

Figures

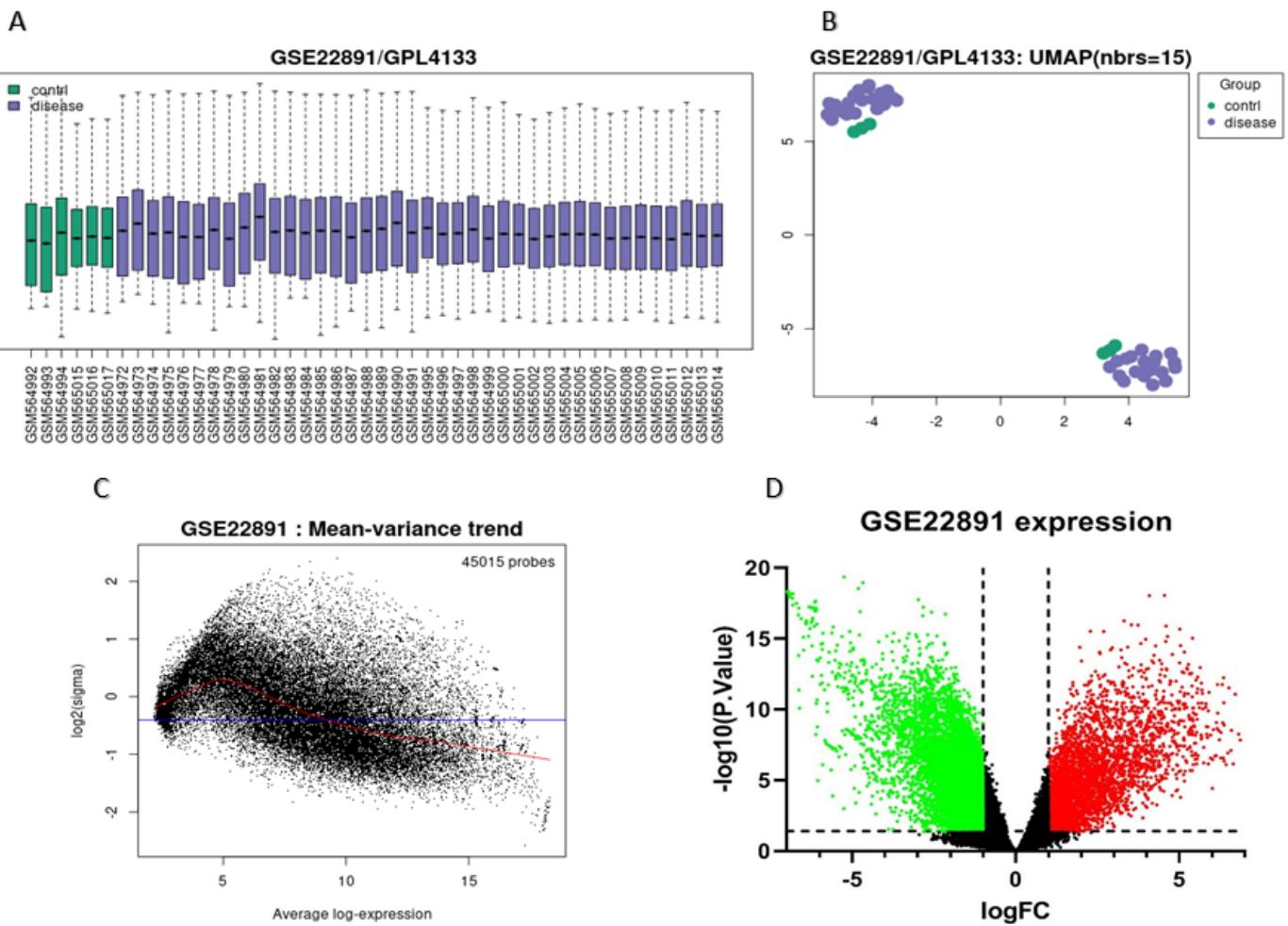


Figure 1

Identification of DEGs in GSE22891. (A) The distribution of the values of the selected Samples through boxplot, Green represented healthy controls. Blue represented GBM. (B) Uniform Manifold Approximation and Projection (UMAP) for visualizing how Samples are related to each other. The number of nearest neighbors used in the calculation is indicated in the plot. (C) The mean-variance relationship of the expression data, after fitting a linear model using Vooma plot. (D) Visualization of differentially expressed genes. Green represented downregulated Genes. Red represented upregulated Genes.

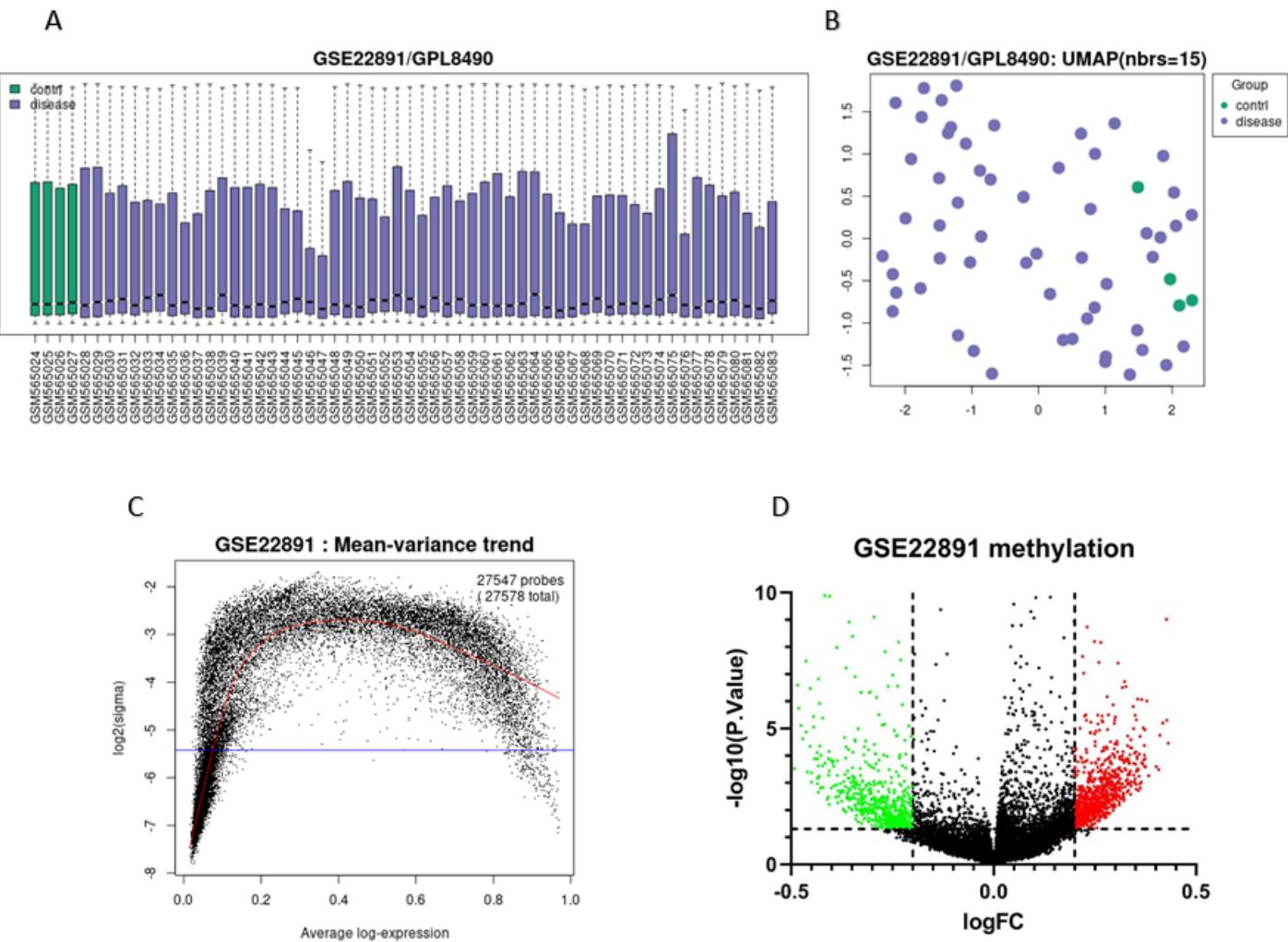


Figure 2

Identification of DMCs in GSE22891. (A) The distribution of the values of the selected Samples through boxplot, Green represented healthy controls. Blue represented GBM. (B) Uniform Manifold Approximation and Projection (UMAP) for visualizing how Samples are related to each other. The number of nearest neighbors used in the calculation is indicated in the plot. (C) The mean-variance relationship of the expression data, after fitting a linear model using Vooma plot. (D) Visualization of differentially methylated CpGs. Green represented downregulated CpGs. Red represented upregulated CpGs.

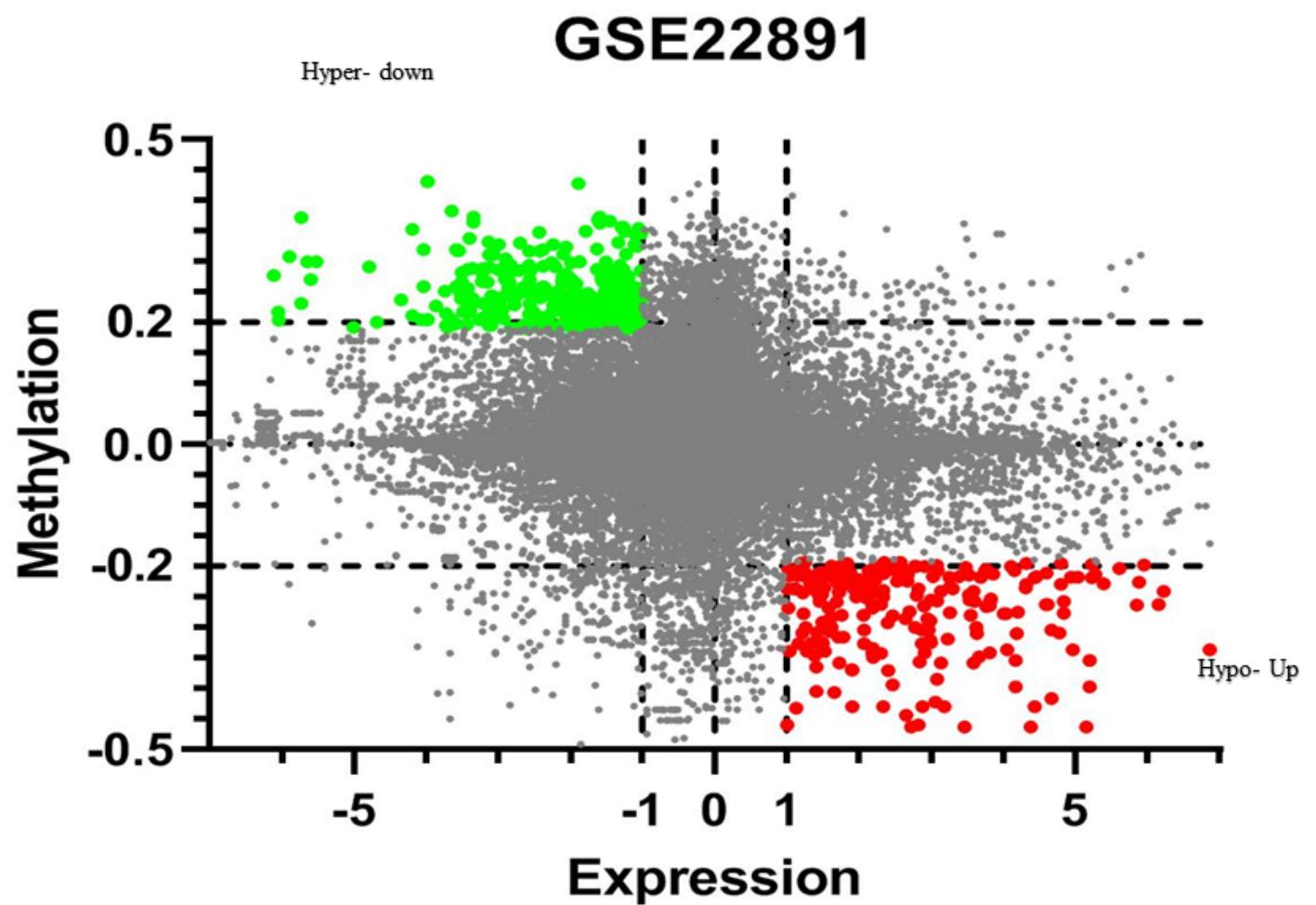


Figure 3

Scatterplot of differentially methylated gene bodies with differential gene expression between healthy and GBM samples. Hypermethylated-downregulated genes are green. Hypomethylated-upregulated genes are red.

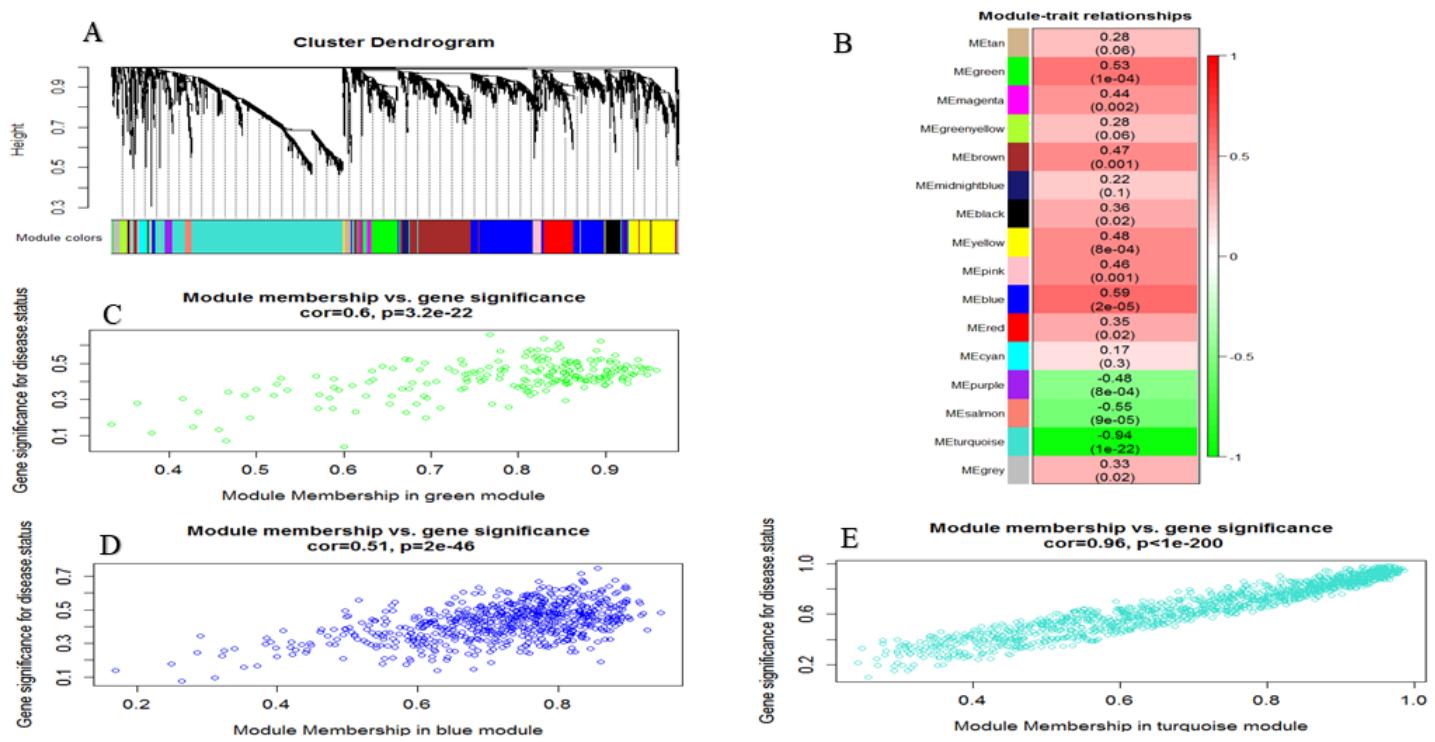


Figure 4

WGCNA data for GSE22891. (A) Clustering dendrogram of DEGs with dissimilarity based on topological overlap, together with assigned module colors. Each color represents a module in the constructed gene co-expression network. (B) Module-trait associations for DEGs. Each row corresponds to a module eigengene, column to a trait. Each cell contains the corresponding correlation and p-value. The table is color-coded by correlation according to the color legend. (C)(D)(E) Scatterplot of Gene Significance (GS) for disease gene significance vs. Module Membership (MM) of green, blue, and turquoise modules, respectively.

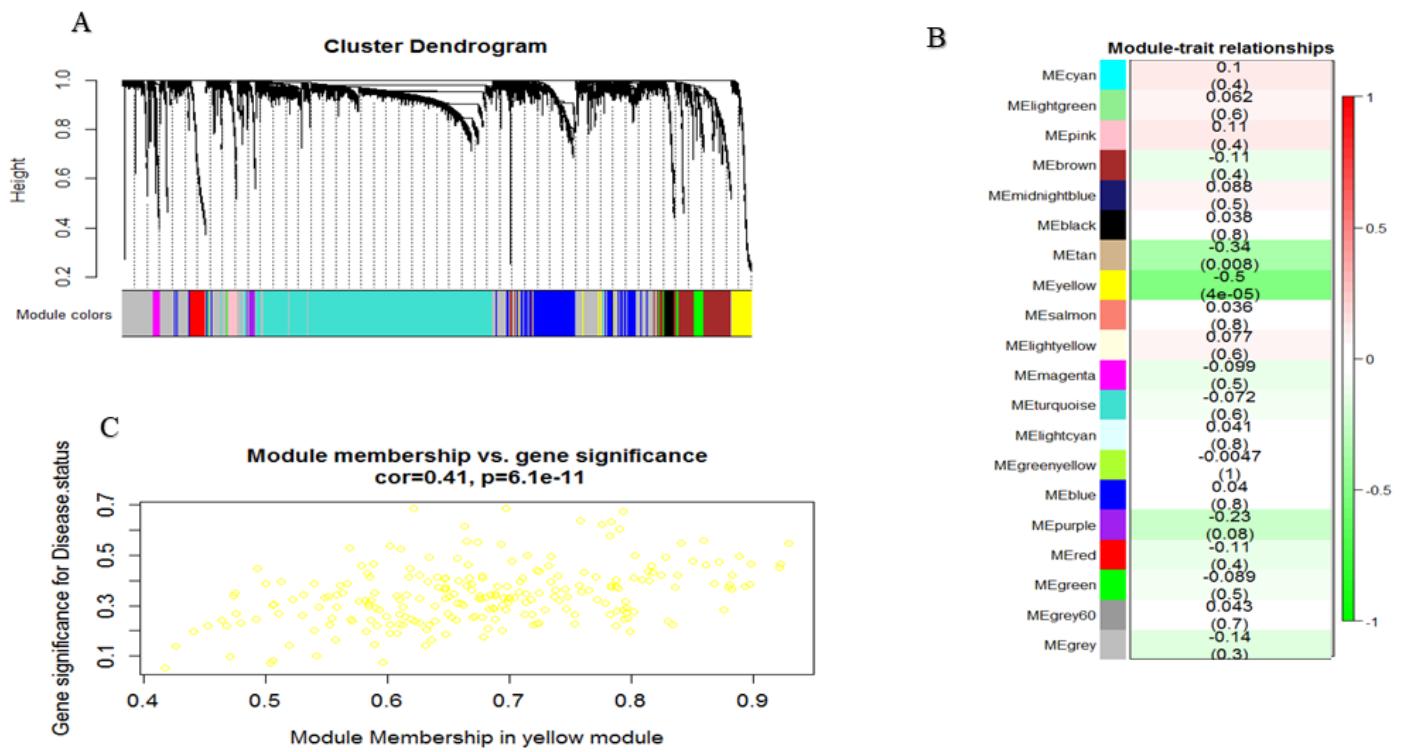


Figure 5

WGCNA data for GSE22891. (A) Clustering dendrogram of DEGs with dissimilarity based on topological overlap, together with assigned module colors. Each color represents a module in the constructed gene co-expression network. (B) Module-trait associations for DEGs. Each row corresponds to a module eigengene, column to a trait. Each cell contains the corresponding correlation and p-value. The table is color-coded by correlation according to the color legend. (C) Scatterplot of Gene Significance (GS) for disease gene significance vs. Module Membership (MM) of yellow module.

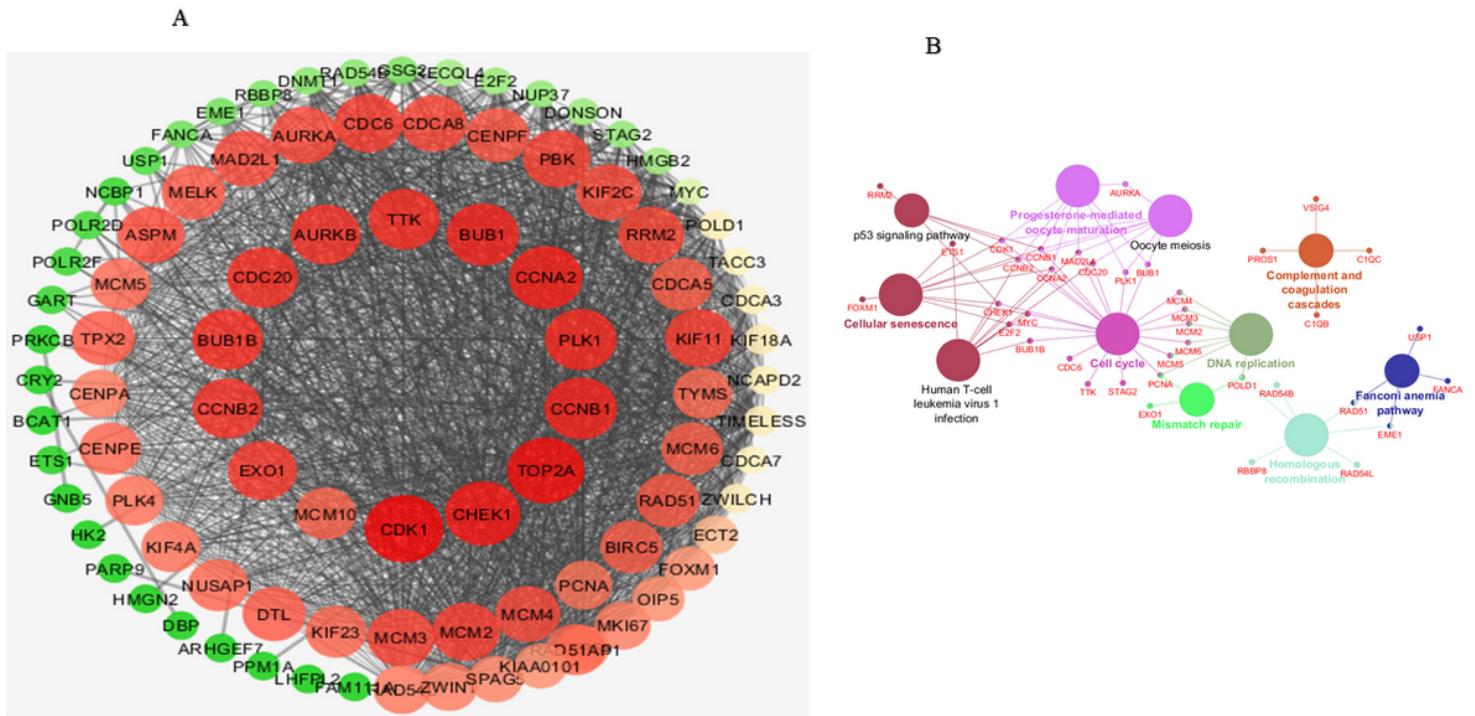


Figure 6

(A) Protein-protein interaction network of the green module genes. Showing genes with the highest number of interactions. The gradual change in the size indicates the connectivity degree. The gradual change in the color indicates the betweenness centrality. (B) Enriched KEGG pathways using the ClueGo plugin of Cytoscape. Functionally grouped networks based on KEGG database of Genes with terms as nodes linked based on their κ score level (≥ 0.3).

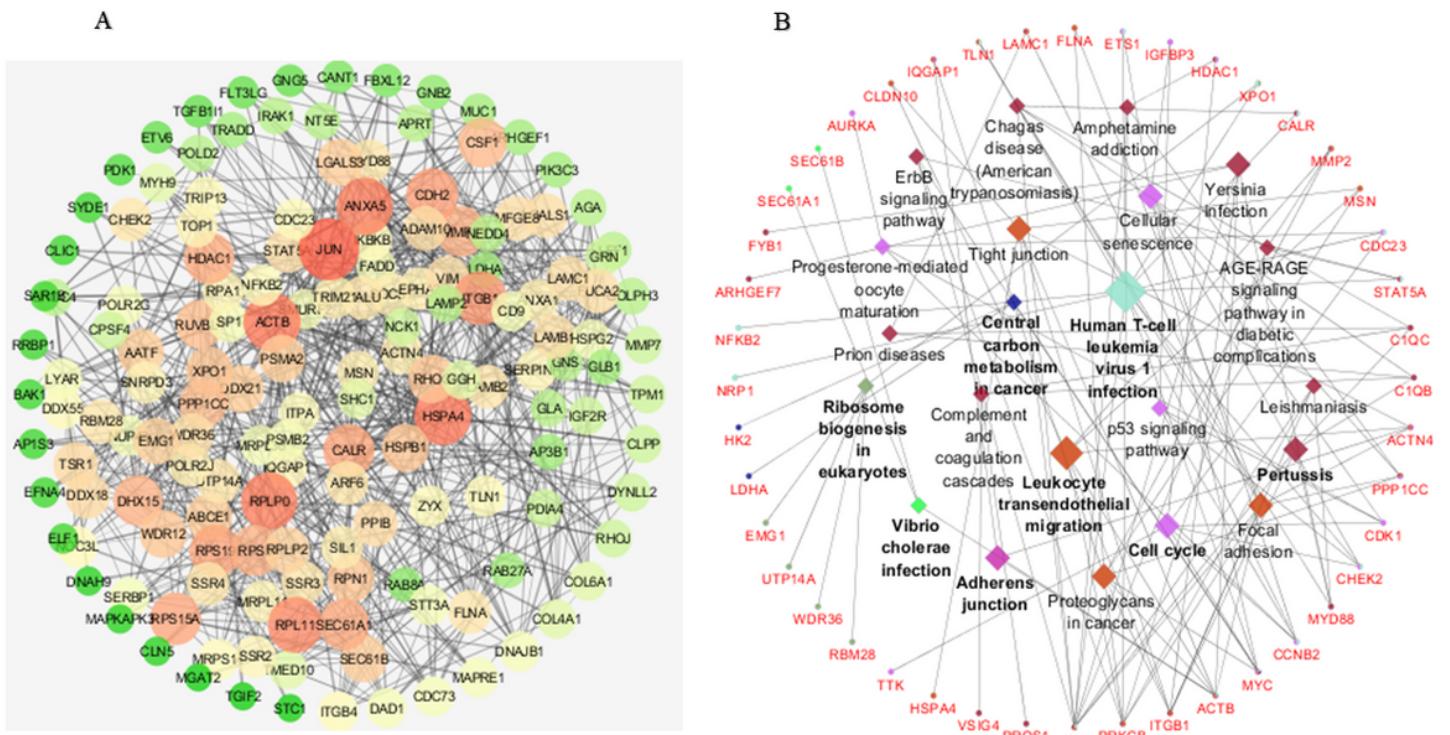


Figure 7

(A) Protein-protein interaction network of the blue module genes. Showing genes with the highest number of interactions. The gradual change in the size indicates the connectivity degree. The gradual change in the color indicates the betweenness centrality. (B) Enriched KEGG pathways using the ClueGo plugin of Cytoscape. Functionally grouped networks based on KEGG database of Genes with terms as nodes linked based on their κ score level (≥ 0.3).

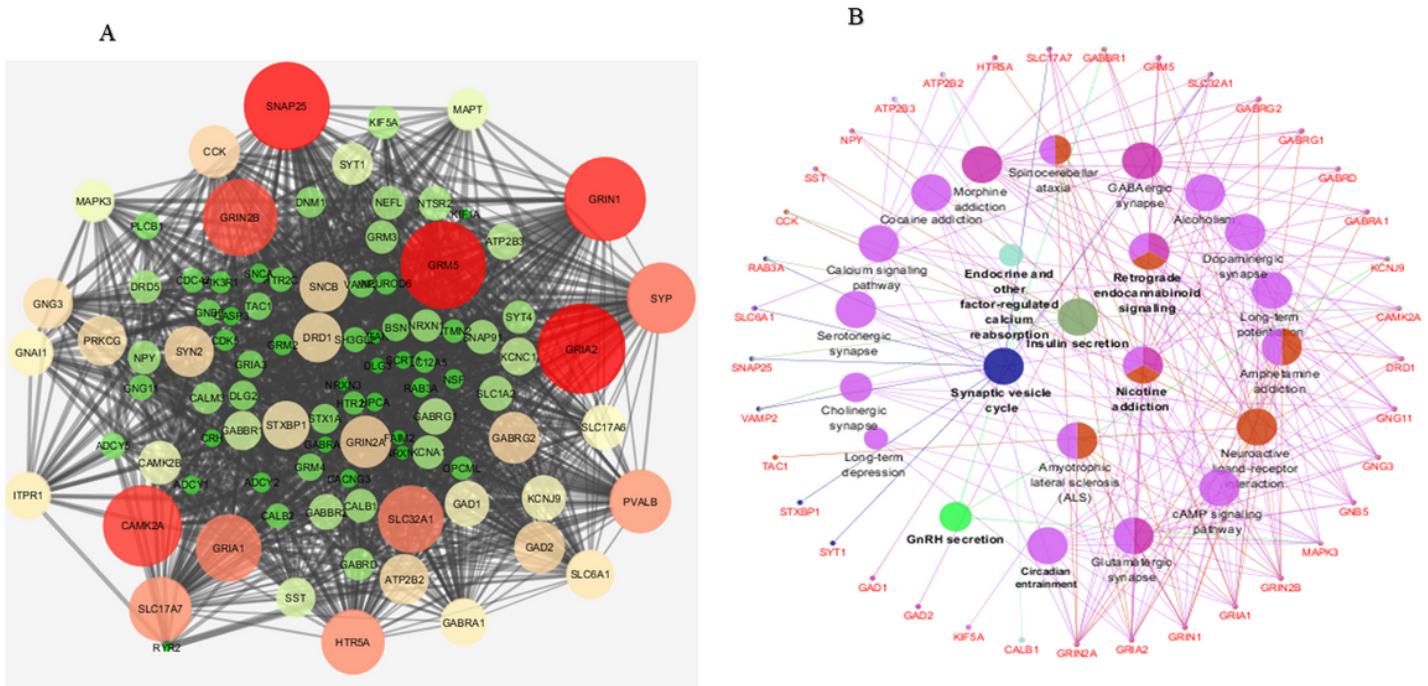


Figure 8

(A) Protein-protein interaction network of the turquoise module genes. Showing genes with the highest number of interactions. The gradual change in the size indicates the connectivity degree. The gradual change in the color indicates the betweenness centrality. (B) Enriched KEGG pathways using the ClueGo plugin of Cytoscape. Functionally grouped networks based on KEGG database of Genes with terms as nodes linked based on their κ score level (≥ 0.3).

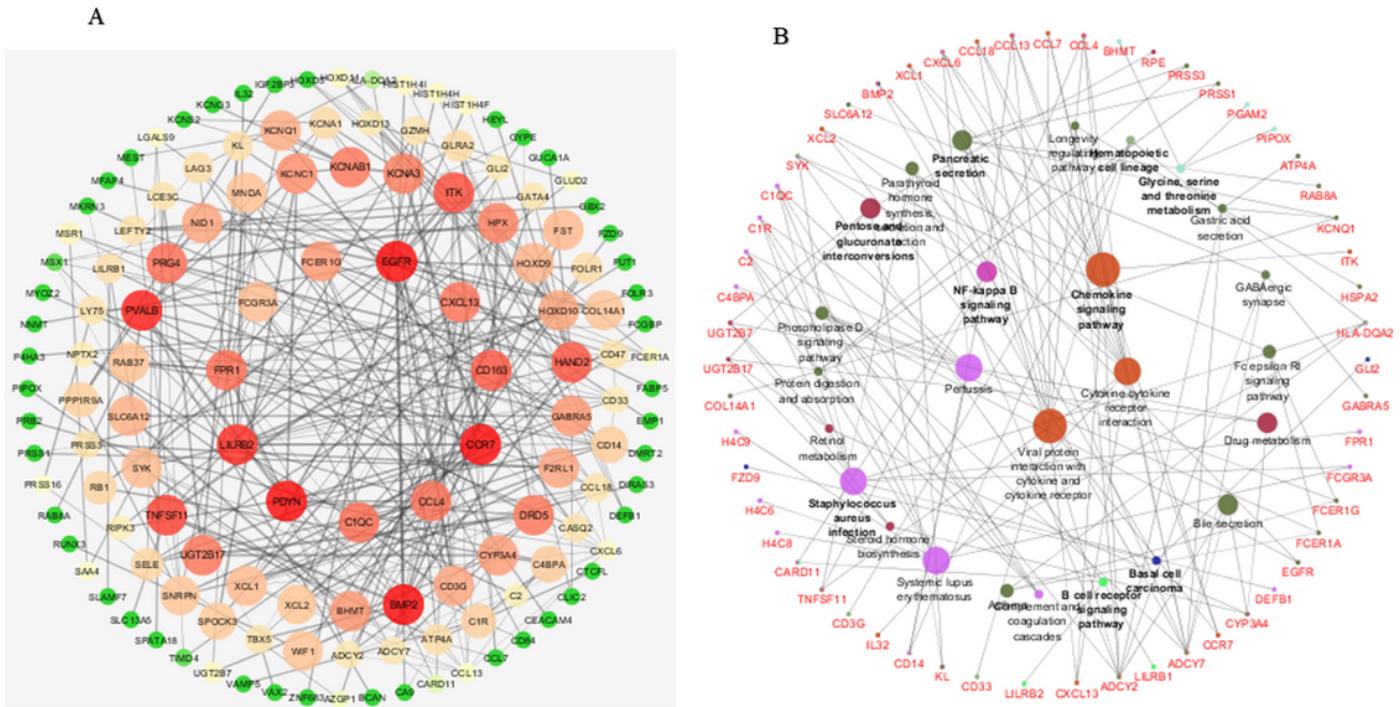


Figure 9

(A) Protein-protein interaction network of the yellow module genes. Showing genes with the highest number of interactions. The gradual change in the size indicates the connectivity degree. The gradual change in the color indicates the betweenness centrality. (B) Enriched KEGG pathways using the ClueGo plugin of Cytoscape. Functionally grouped networks based on KEGG database of Genes with terms as nodes linked based on their κ score level (≥ 0.3).

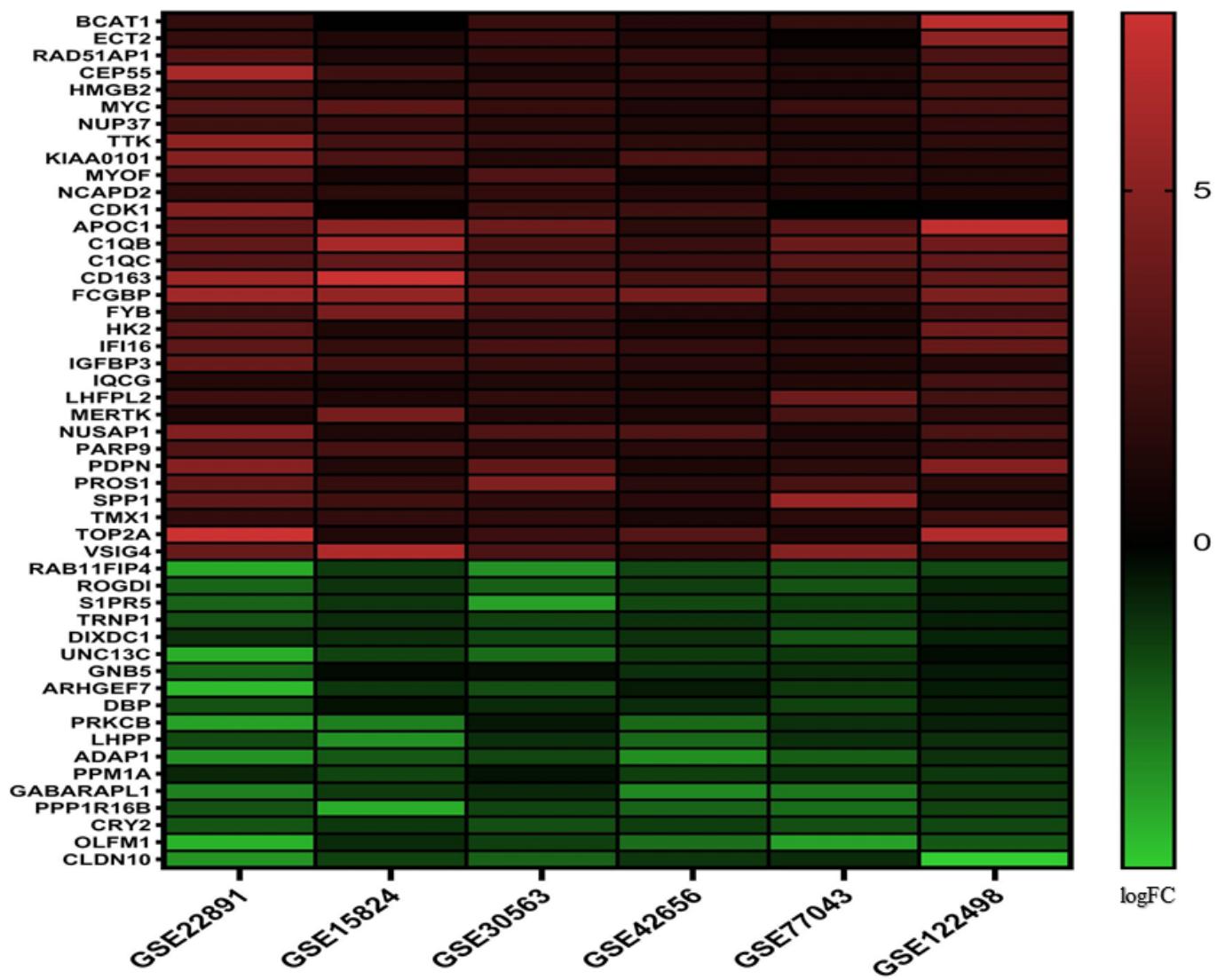


Figure 10

The heatmap of expression values of 50 hub genes using GSE77043, GSE42656, GSE30563, GSE22891, GSE15824, and GSE122498.

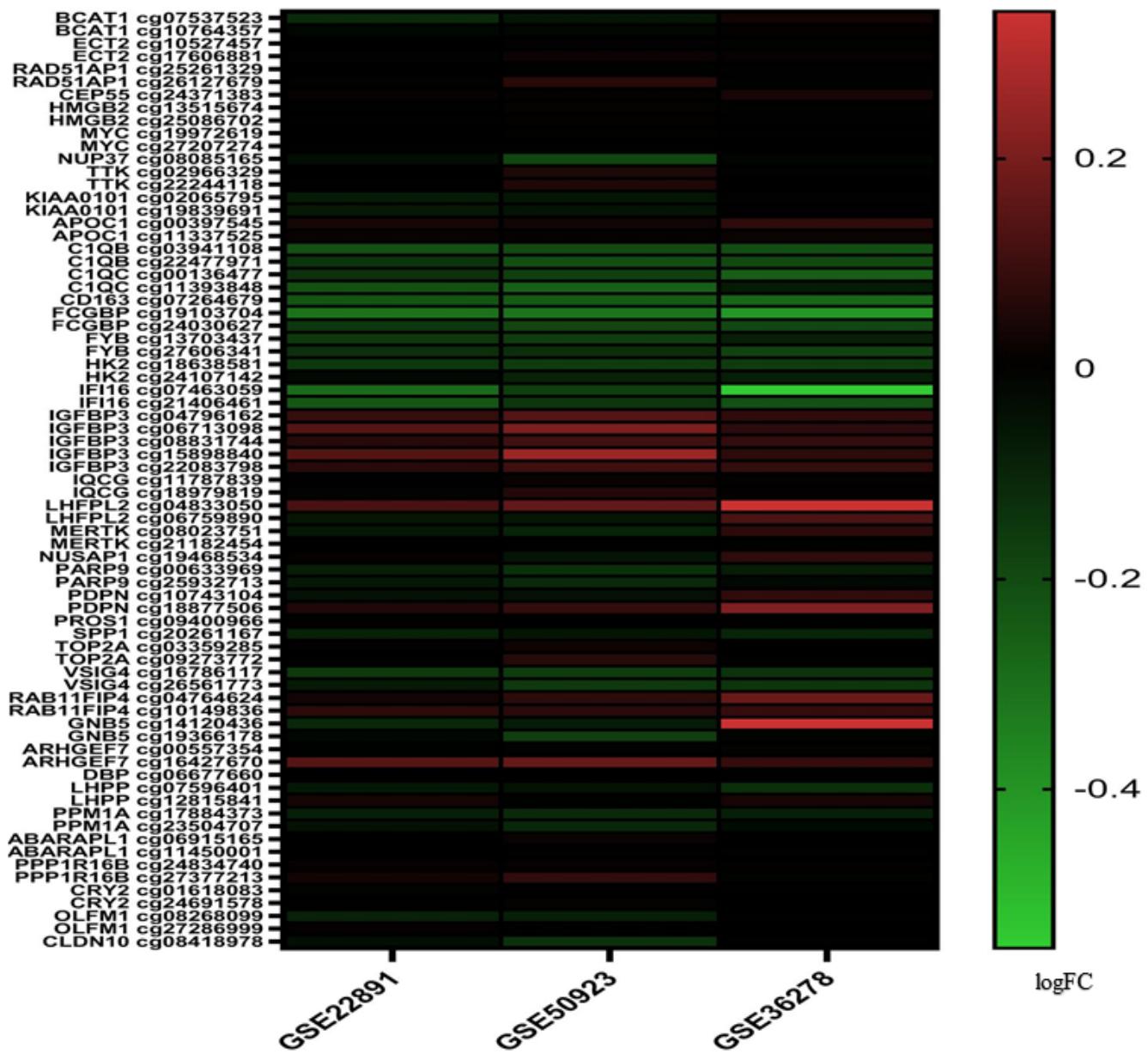


Figure 11

The heatmap of methylation values of 50 hub genes using GSE50923, GSE22891, and GSE36245.

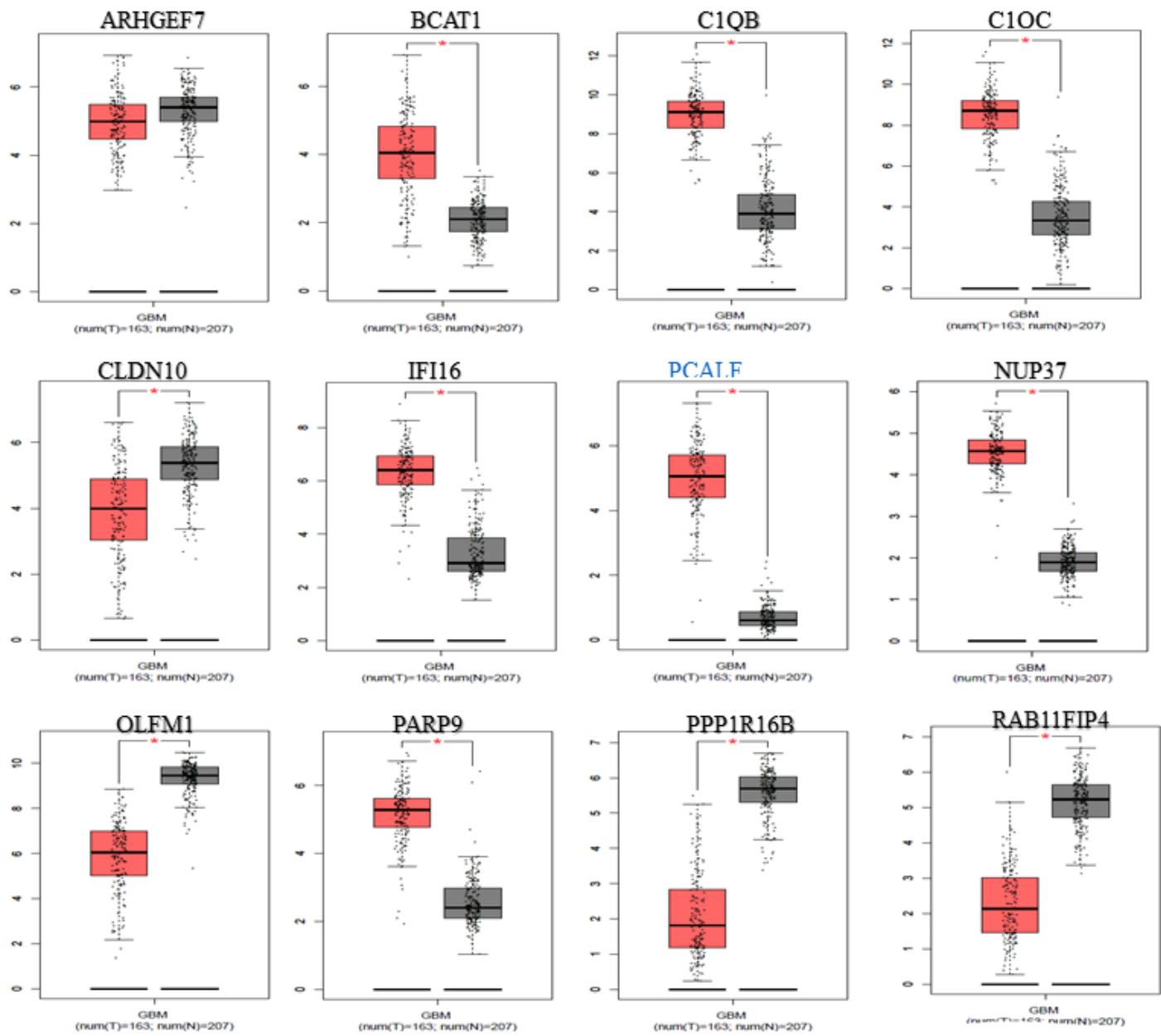


Figure 12

Boxplot expression of 12 hub genes including ARHGEF7, RAB11FIP4, PPP1R16B, OLFM1, CLDN10, BCAT1, C1QB, C1QC, IFI16, NUP37, PARP9, and PCALF in GBM.

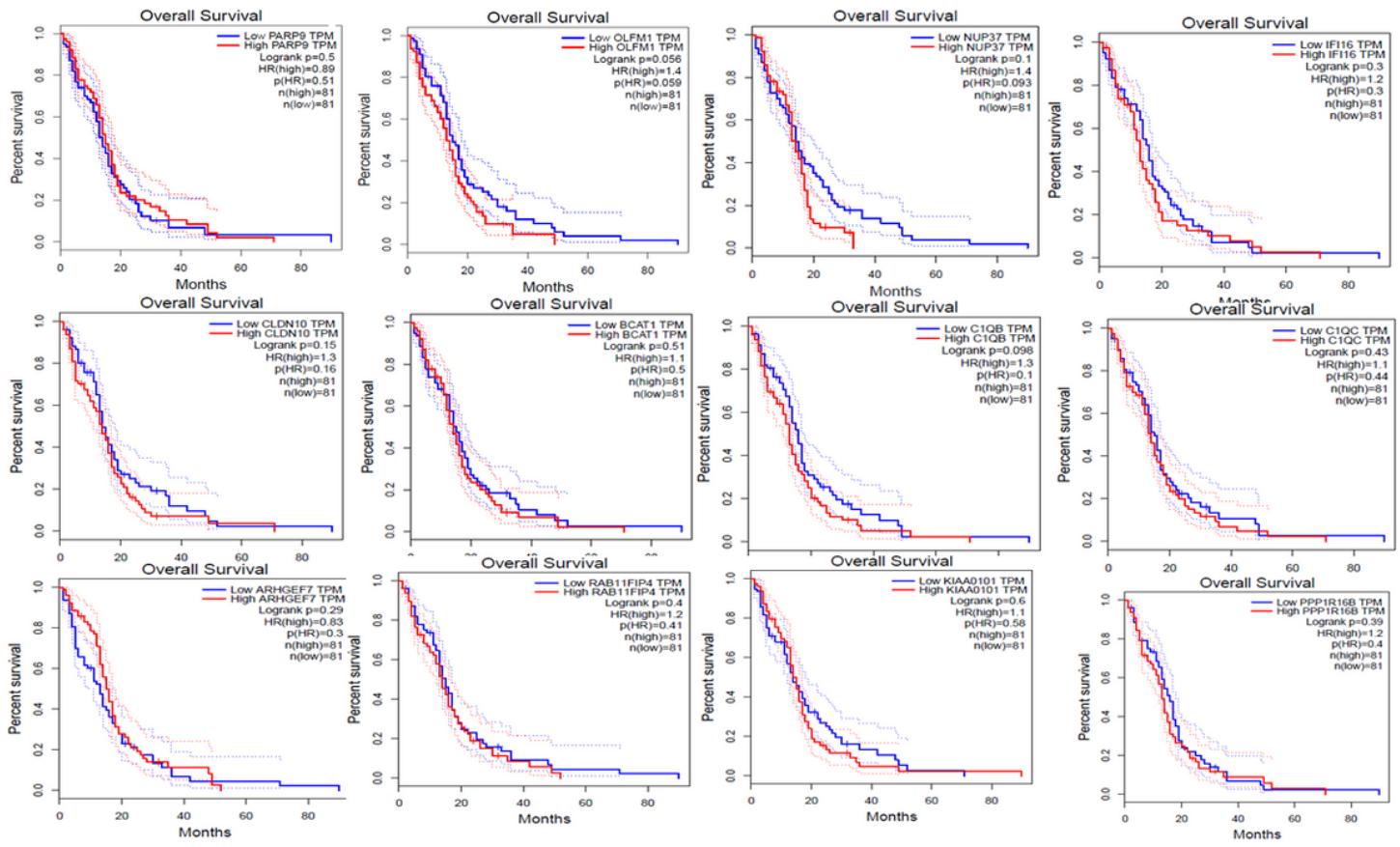


Figure 13

Overall survival analysis of 12 hub genes including ARHGEF7, RAB11FIP4, PPP1R16B, OLFM1, CLDN10, BCAT1, C1QB, C1QC, IFI16, NUP37, PARP9, and PCALF in GBM.

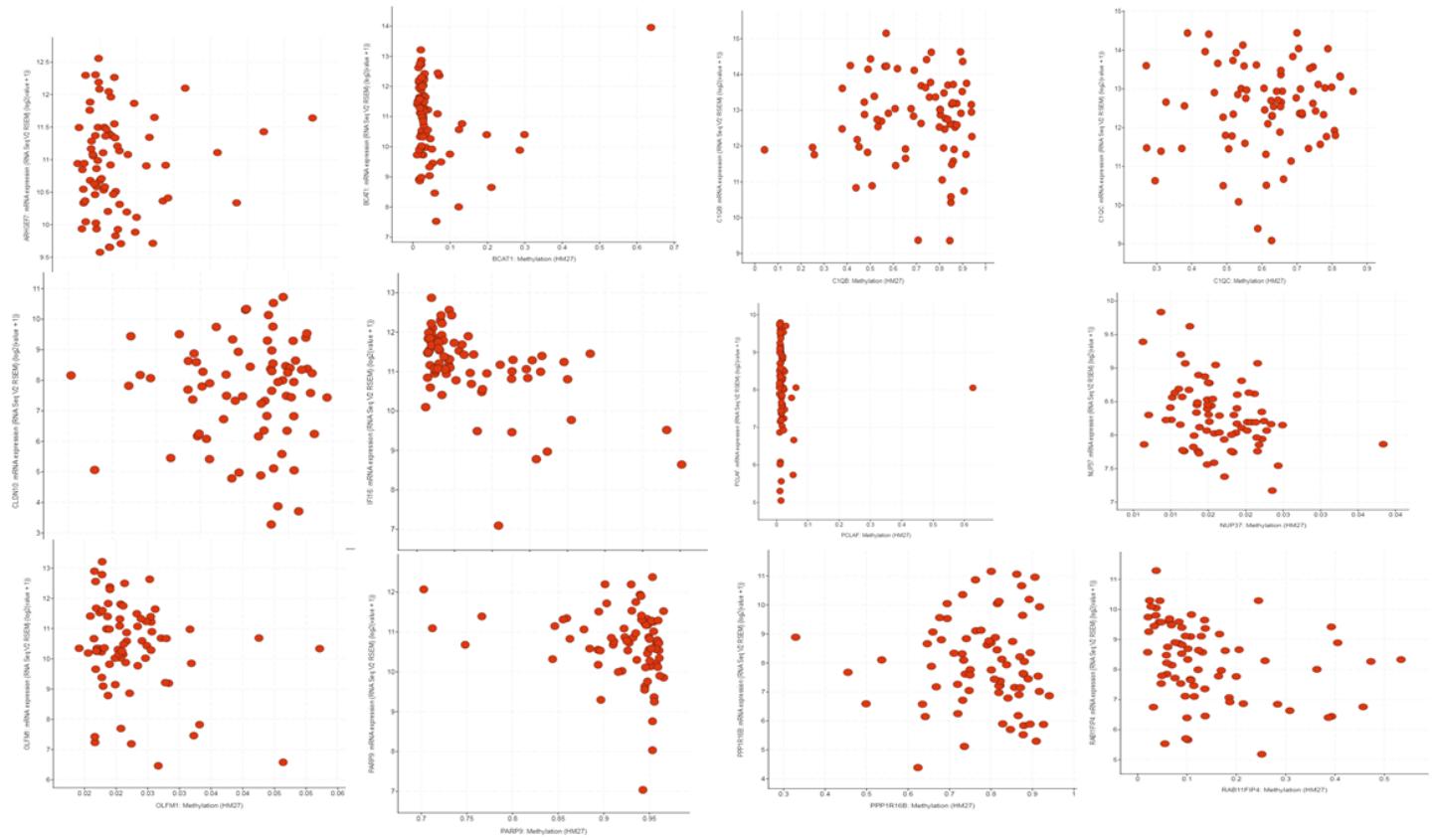


Figure 14

Scatterplot of methylation status versus expression status of all 12 hub genes using cBioportal database.

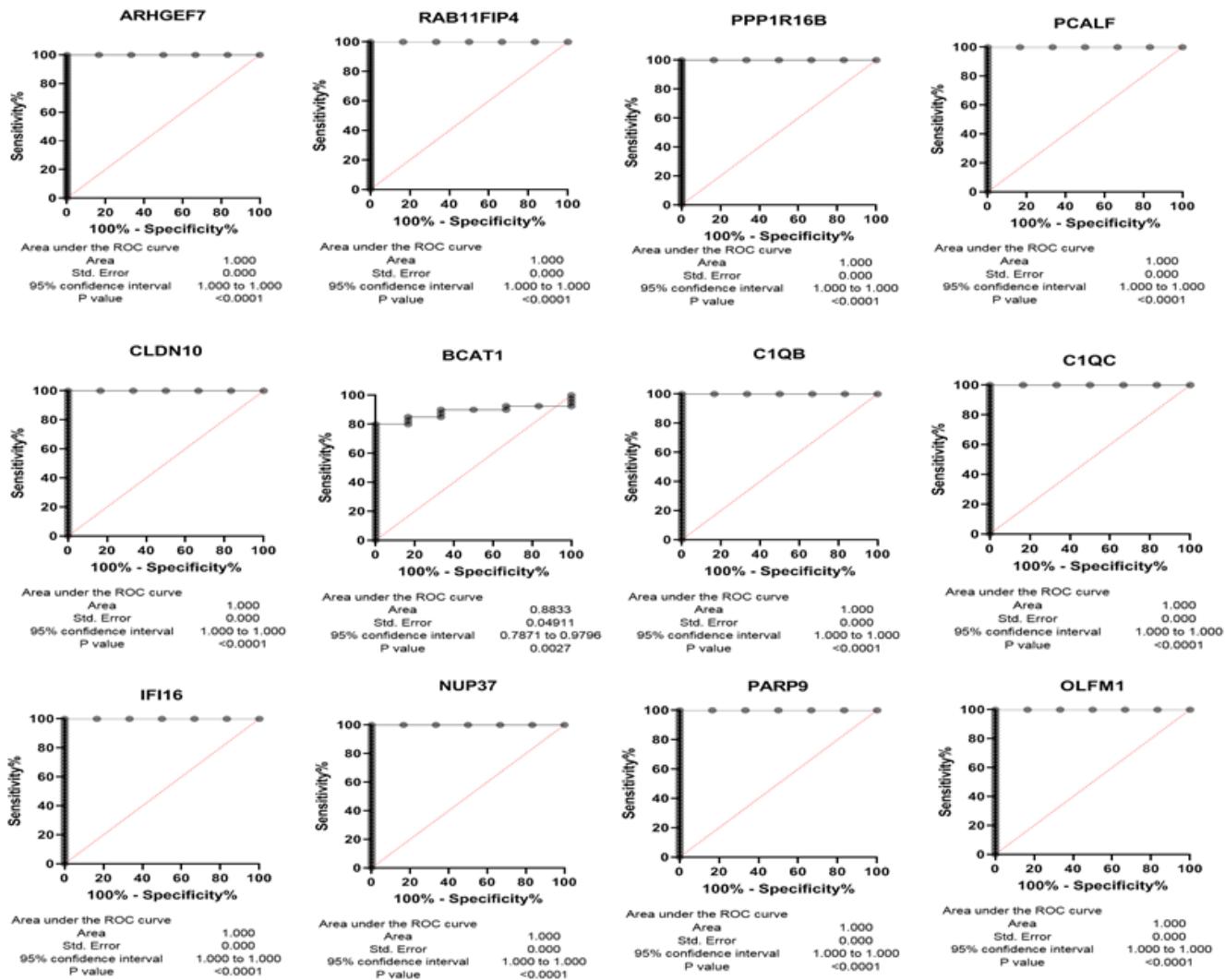


Figure 15

ROC curve analysis of 12 hub genes including ARHGEF7, RAB11FIP4, PPP1R16B, OLFM1, CLDN10, BCAT1, C1QB, C1QC, IFI16, NUP37, PARP9, and PCALF in GBM based on GSE22891 dataset. Receiver operating characteristic (ROC) curve and area under the curve (AUC) statistics are used to evaluate the capacity to discriminate GBM from healthy controls with excellent specificity and sensitivity.