Identification of potential biomarkers and candidate small molecule drugs in glioblastoma

Wei-cheng Lu
China Medical University First Hospital
Hui Xie
Shenyang Medical College
Ce Yuan
University of Minnesota
Jin-jiang Li
Hospital General of Northern Theater Command
Zhao-yang Li
China Medical University
An-hua Wu (wuanhua@yahoo.com)
https://orcid.org/0000-0002-3402-1884

Primary research

Keywords: Glioblastoma, differentially expressed genes, hub genes, prognosis, small molecular drugs

Posted Date: March 10th, 2020

DOI: https://doi.org/10.21203/rs.3.rs-16628/v1

Version of Record: A version of this preprint was published on August 28th, 2020. See the published version at https://doi.org/10.1186/s12935-020-01515-1.

Abstract

Background and aioms: Glioblastoma (GBM) is a common and aggressive primary brain tumor, and the prognosis for GBM patients remains poor. This study aimed to identify the key genes associated with GBM and to further analyze the prognostic factors and small molecular drugs.

Methods: Three microarray datasets (GSE111260, GSE103227, and GSE104267) were selected to identify the differential expressed genes (DEGs) between GBM and normal tissues. Then, prognosis related DEGs were screened by survival analysis, followed by functional enrichment analysis was performed. The protein-protein interaction (PPI) network was constructed to explore the hub genes associated with GBM. The prognostic of these hub genes in GBM patients were analyzed using The Cancer Genome Atlas (TCGA) database, and the protein expression level of hub genes were validated using the Human Protein Atlas (HPA) databases. Subsequently, the small molecule drugs of GBM were predicted by Connectivity Map (CMAP) database.

Results: A total of 78 DEGs related to GBM prognosis were identified, and 10 hub genes with high degree were obtained. The mRNA expression and protein expression levels of *CETN2*, *MKI67*, *ARL13B*, and *SETDB1* were overexpressed in GBM tissues, while the expression levels of *CALN1*, *ELAVL3*, *ADCY3*, *SYN2*, *SLC12A5*, and *SOD1* were down-regulated in GBM tissues. Additionally, these genes were significantly associated with the prognosis of GBM. A total of 98 small molecular drugs were predicted, among these, adiphenine and podophyllotoxin were considered as the potential drugs to treat GBM.

Conclusions: Our study provided 10 key genes for diagnosis, prognosis, and therapeutic target for GBM. These findings might contribute to a better comprehension of molecular mechanisms of GBM development, and provide new perspective for further GBM research. However, specific regulatory mechanism of these genes needed further elaboration.

Background

Glioblastoma (GBM) is a most common and aggressive malignant brain tumor, accounting for 16% of all primary brain and central nervous system neoplasms [1]. The mean survival of GBM is approximately 14.6 months, and GBM is one of the most challenging malignancies to treat due to its high heterogeneity, high recurrence rate, and diffusing invasiveness [2]. Recently, the main treatment methods of GBM are contained surgery, radiation, and chemotherapy [3]. Despite extensive efforts to explore novel therapies, the survival of GBM has not markedly improved. Therefore, it is necessary to develop effective treatment options. Currently, gene therapy, molecularly targeted therapy, and immunotherapy are promising treatment approaches [4].

Extensive studies have reported the biomarkers and drug targets for GBM treatment. Shergalis et al. indicated that genes like estrogen receptor 2 (ESR2), ELOVL fatty acid elongase 6 (ELOVL6), and iroquois homeobox 3 (IRX3) were over-expressed in patients with GBM, and these genes were related to poor survival outcomes [3]. Additionally, Jin et al. proved that aryl hydrocarbon receptor (AhR) and its ligand

kynurenine could decrease GBM cell invasion, and also inhibited GBM carcinogenesis [5]. Prior studies have demonstrated that GBM stem cells (GSCs) participant in the GBM development and resistance to therapy, and targeting these cells could be considered as a therapeutic strategy [6]. Stangeland et al. investigated the molecular targets in GSCs, such as PDZ binding kinase (PBK), centromere protein A (CENPA), and kinesin family member 15 (KIF15), suggesting these genes might be potential targets for GBM treatment [7]. Furthermore, drug therapy of GBM has also been reported. Lu et al. revealed that triple-drug therapy (bevacizumab, irinotecan, and temozolomide) had benefit effect on recurrent GBM [8]. Meanwhile, Tea et al. observed that targeting the spingolipid metabolism might be used in the treatment of GBM [9]. However, the molecular mechanism of GBM pathogenesis has not been entirely elucidated, and it is desperately required to explore novel biomarkers and small drug molecules.

Currently, the microarray gene expression research has been performed to uncover the molecular mechanism of various cancers. The mRNA data are collected from two databases, including Gene Expression Omnibus (GEO) and The Cancer Genome Atlas (TCGA). The GEO database can be applied to identify the differentially expressed genes (DEGs), to explore molecular signal and its correlation, and to analyze gene regulation network [10]. However, due to the limited samples, the analysis results of a single microarray dataset may be biased and unreliable. Hence, integrated analysis of multiple datasets can improve the accuracy and reliability of the results, thus obtain a comprehensive discovery of DEGs in tumors.

In the present study, three microarray datasets related GBM were selected for further study, the raw data of mRNA profile were downloaded from GEO database, and integrated analysis of three sets was conducted. The overlapping DEGs were identified by the intersection of three datasets. Then, the DEGs associated with GBM prognosis were screened by TCGA database. Functional enrichment analysis and protein-protein interaction (PPI) of these DEGs were performed. The hub genes that closely related to GBM prognosis were verified by using UCLCAN online tool. Subsequently, gene mutations and protein levels of hub genes were investigated. Finally, the small drug molecules of GBM were explored by connectivity map (CMAP) database. A flow chart of this study is shown in Fig. 1.

Methods And Materials

Microarray data

To screen the DEGs between GBM samples and normal samples, three microarray datasets (GSE111260, GSE103227, and GSE104267) were downloaded from the GEO database (http://www.ncbi.nlm.nih.gov/geo) [11]. In total, the three datasets included 81 GBM samples and 11 normal samples. Specifically, GSE111260 was derived from GPL5175 [HuEx-1_0-st] Affymetrix Human Exon 1.0 ST Array [transcript (gene) version], and composed of 67 GBM samples and three control samples; GSE103227 was obtained from GPL16956 Agilent-045997 Arraystar human lncRNA microarray V3 (Probe Name Version) platform, and including five GBM and five normal tissues; GSE104267 was

extracted from GPL22448 Phalanx Human lncRNA OneArray v1_mRNA platform, and contained nine GBM and three control samples.

Data pre-processing

The raw data was read by using the limma package of the R software (Version 3.34.9, http://www.bioconductor.org/packages/release/bioc/html/limma.html) [12], and then data normalization was performed by robust multi-array average (RMA) method [13,14], including background adjustment, quantile normalization, and log2 conversion. The probes were annotated with the platform annotation file, the probes that do not match the gene (gene symbol) were removed; in addition, for the different probes that mapped to the same gene, the mean value was selected as the final expression value.

DEGs analysis

DEGs between GBM and control sample in the three datasets were respectively screened by using the Limma package. $P < 0.05$ and |log Fold change (FC)| >1 were considered as the criteria for DEGs. The DEGs obtained from each dataset were analyzed by VENN analysis to observe the up- or down-regulated of the genes, which served as the basis for the gene change trend during subsequent TCGA verification.

Integration of the DEGs from three datasets

The intersection of DEGs from these microarray datasets was conducted by NetworkAnalyst 3.0 database [15] (https://www.networkanalyst.ca/NetworkAnalyst/uploads/MetaLoadView.xhtml), which could compare and analyze of DEGs generated from different studies via various statistical methods. In this study, Fisher's method, Fixed effects models, and Vote Counting were used to integrate multiple data sets. DEGs with $P < 0.001$ (both Fisher's method and Fixed effects models) and vote counts ≥ 2 were considered as shared DEGs. Meanwhile, the ComBat function of the R package sva [15] was utilized to eliminate heterogeneity between these three datasets. Additionally, the DEGs screened by the three integration methods were also analyzed by VENN, and the genes that existed in at least two methods were selected as the focus of the further analysis.

Survival analysis

The genomic data commons (GDC) GBM data in TCGA database (https://xenabrowser.net/) [16] was downloaded to obtain the prognostic related clinical information of GBM, including overall survival (OS) time and OS status. According to the shared DEGs identified from the integrated analysis, the samples with no OS time (or less than one mouth) and the DEGs with median expression level less than 0 were removed. Then, the remaining samples were divided into high expression group and low expression group, and the log-rank statistical test was performed. $P < 0.05$ as statistically significant threshold, and Kaplan-Meier (KM) method was conducted to plot OS curves.

Functional enrichment analysis of prognostic related DEGs

To investigate the biological functions and pathways involved in these prognostic related DEGs, the Gene Ontology (GO) terms and pathway analysis were performed by using metascape database [17] (http://metascape.org). P < 0.01, count > 3, and enrichment factors (ratio of observed counts to accidental exoected counts) > 1.5 were selected as the threshold of significantly enriched terms. In order to further explore the relationship between the statistically enriched terms, the kappa-statistical similarities of these terms were calculated, and the overlapping or related terms were identified, then the functional network clustering was performed. Subsequently, the interactions of terms were obtained based on the similarity of genes which enriched in each terms. Finally, the similar function integration network was constructed.

The PPI network analysis

The STRING (version: 11.0, https://string-db.org) [18] was used to predict the PPI among the prognostic genes. The species was *Homo sapiens*, and the confident interaction score more than 0.15 (low confident) was set as significant interaction. The PPI network was visualized using Cytoscape software (version: 3.6.1, http://www.cytoscape.org/) [19]. The degree of each protein node was calculated, and nodes with degree ≥ 10 were selected as hub genes.

Verification of hub genes

We used the online software UALCAN (http://ualcan.path.uab.edu/index.html) [20] to verify the hub genes identified from the PPI network. The candidate hub genes were submitted to the UALCAN database, and the TCGA data were applied to validate the relationship between the genes espression and the prognosis of GBM.

Gene mutation analysis

The cBio Cancer Genomics Portal could analyze the molecular data obtained from cancer tissues and cytology, to recognize and understand heredity, epigenetics, and gene expression. Thus, we used the CBiocancer genomics portal (https://www.cbioportal.org/) [21] to analyze the genetic mutations of the key genes between samples.

The Human Protein Atlas

The Human Protein Atlas (HPAs), composed of tissue atlas, cell atlas, and pathology atlas, were provided the data of transcriptomics and proteomics in specific human tissues. In this study, the protein immunohistochemical level of hub genes was investigated in GBM tissues and compared normal tissues by using HPA database [22], and the data were downloaded from the pathology atlas in HPA.

Drug-related pathways and candidate drugs for GBM

Small drug molecules that had synergistic or antagonistic effects with drug were analyzed by CMAP database [23]. The hub genes were compared with the small molecules in the CMAP database to obtain

the small drug molecules related to drug treatment, $P < 0.05$ as the cut-off criteria. Additionally, enrichment > 0 indicated the molecules had potential synergistic effects with drugs, enrichment < 0 revealed molecules had potential antagonistic effects with drugs.

Results

Identification of DEGs from GEO datasets analysis

The raw data from three gene expression profiles (GSE103227, GSE104267, and GSE111260) were downloaded from NCBI GEO database. The detailed information of these three datasets is listed in Table 1. DEGs between GBM samples and normal samples were screened from three studies, and visualized by PCA and volcano plots (Figure 2A and 2B). Afterwards, the number of DEGs obtained from three datasets is shown in Supplementary Table 1. Besides, an integrative analysis across three datasets was conducted, and the results are shown in Figure 2C. We observed 18 up-regulated and 6 down-regulated overlapping DEGs existed in three microarray datasets.

Integrated analysis of three GEO datasets

By employing three statistical methods, a total of 5801, 640, and 2368 DEGs were identified by Fisher's method, Fixed effects models, and Vote counting, respectively. Additionally, 613 genes were obtained by all three statistical methods (Supplementary Figure 1A). Heat map representation of the top 10 hub DEGs across different microarrays is displayed in Supplementary Figure 1B. Among these, centrin 2 (*CETN2*), marker of proliferation ki-67 (*MKI67*), ADP ribosylation factor like GTPase 13B (*ARL13B*), SET domain bifurcated histone lysine methyltransferase 1 (*SETDB1*), calneuron 1 (*CALN1*), ELAV like RNA binding protein 3 (*ELAVL3*), adenylate cyclase 3 (*ADCY3*), synapsin II (*SYN2*), solute carrier family 12 member 5 (*SLC12A5*), and superoxide dismutase 1 (*SOD1*) were selected as hub genes and subjected to further analysis and validation.

Survival analysis of hub genes

A total of 167 samples were downloaded from TCGA-GBM expression dataset. To further analyze the relationship between hub gene expression and the prognosis of GBM, the overall survival analysis of hub genes were performed (Figure 3). According to the median level of expression, the GBM patients were divided into high and low expression group. *CETN2*, *MKI67*, *ARL13B*, and *SETDB1* with lower expression level were related to a significantly longer survival time; meanwhile, high expression of *CALN1, ELAVL3, ADCY3, SYN2, ARL13B*, *SLC12A5*, and *SOD1* were associated with better overall survival among patients with GBM.

GO enrichment and KEGG pathway analysis of hub genes

Based on the GEM prognosis related genes, the functional enrichment analysis was conducted. The results of GO enrichment analysis indicated that these genes were mainly associated with behavior, sensory organ morphogenesis, and chromosome separation. Additionally, the KEGG pathway analysis

revealed that these genes were primarily involved in longevity regulating pathway, bile secretion, and hemostasis pathways (Figure 4A). Furthermore, all terms were grouped into clusters based on the similarities, and a total of 14 clusters of significantly enriched terms were obtained (Figure 4B), among these, sensory organ morphogenesis was the most enriched term.

Establishment of PPI network

In order to understand the potential relationships between prognostic related DEGs, the PPI network was constructed. The PPI network composed of 71 nodes and 214 edges (Figure 5). There were 16 DEGs with their degree > 10 were considered as hub genes. Additionally, the specific degree values of these genes are listed in Table 2. We found the degree of *ELAVL3*, HDAC2, and *CALB1* were higher than other genes.

The expression and mutation of hub genes

By analyzing the expression of the hub genes in the TCGA GBM data, we observed that the change trend of genes expression was consistent with microarray datasets. Compared with normal samples, the expression level of *MKI67*, *ARL13B*, and *SETDB1* was significantly up-regulated in GBM samples, while *ELAVL3*, *ADCY3*, *SOD1*, *CALN1*, *SYN2*, and *SLC12A5* were markedly down-regulated (Figure 6). The information of hub genes is presented in Table 3. In addition, the hub gene mutations in GBM were tested using cBioPortal. The *MKI67*, *SLC12A5*, and *SOD1* exhibited higher mutation frequencies, and the proportion respectively was 2.2, 0.7, and 0.2% (Supplementary Figure 2A). Meanwhile, approximately 3% of GBM clinical cases showed significant alterations in the 10 hub genes (Supplementary Figure 2B).

Immunohistochemical Analysis

The protein expression levels of the 10 hub genes in GBM were explored using the HPA database (Figure 7). The protein level of *MKI67* and *ARL13B* was not detected in normal tissues, while the level of these genes was medium and high in the GBM tissues. The protein level of *CETN2* was low in normal samples, while the level was high in GBM samples. Additionally, the medium protein level of *SETDB1* was observed in normal tissues, whereas the high protein level was revealed in GBM tissues. Meanwhile, the protein level of *CALN1* was medium in normal samples, while was low in the GBM samples. In brief, these results indicated that the protein levels of the hub genes were consistent with the mRNA expression levels.

Analysis of GBM-related small molecular drugs

We obtained 98 GBM-related small molecular drugs from the CMAP database according to the screening criteria of $p < 0.05$, among these, 45 drugs might play potential synergies role in the development of GBM (enrichment > 0), and 53 drugs might served repress role in the GBM progression (enrichment < 0). The most relevant top 10 drugs (with a smaller p-value) were selected and are displayed in Figure 8. The results revealed that drugs like nomegestrol (enrichment=0.975), adiphenine (enrichment=0.783), etiocholanolone (enrichment=0.775), and podophyllotoxin (enrichment=0.835) might be potential drugs for GBM treatment.

Conclusions

In summary, with the integrated bioinformatics analysis of three GBM-related gene expression profiles, 10 key genes connected with pathogenesis and prognosis of GBM were identified, including CETN2, MKI67, ARL13B, SETDB1, CALN1, ELAVL3, ADCY3, SYN2, SLC12A5, and SOD1. These hub genes might serve as novel diagnostic and treatment biomarkers of GBM, and might conduct to elucidate the molecular mechanism of the occurrence and progression of GBM. Additionally, a series of small molecule drugs of GBM were identified, such as adiphenine and podophyllotoxin, these drugs might be used for GBM therapy.

Declarations

Acknowledgements

The authors thank all members in our lab for the excellent technical help.

Authors' contributions

AHW conceived and designed this study. WCL carried out the plan and wrote this paper. HX, CY, JJL and ZYL gave advice and carried out the data analysis. All authors read and approved the final manuscript.

Funding

This work was supported by grants from Guidance Plan of Natural Science Foundation of Liaoning Province (No. 201602773; 2019-ZD-0340).

Availability of data and materials

The data that support the findings of this study are available from University of California Santa Cruz Genome Browser and GEO database.

Ethics approval and consent to participate

This work was approved by the Ethical Board of China Medical University.

Consent of publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

Author details

[1] Department of Neurosurgery, First Affiliated Hospital of China Medical University, Shenyang, Liaoning, China

[2] Department of Histology and Embryology, College of Basic Medicine, Shenyang Medical College, Shenyang, Liaoning, China

[3] Graduate Program in Bioinformatics and Computational Biology, University of Minnesota, Minneapolis, USA.

[4] Department of Neurosurgery, General Hospital of Northern Theater Command, Shenyang, Liaoning, China

[5] Department of Laboratory Animal Center, China Medical University, Shenyang, Liaoning, China

Abbreviations

GBM: Glioblastoma; ESR2: Estrogen receptor 2; ELOVL6: ELOVL fatty acid elongase 6; IRX3: Iroquois homeobox 3; AhR: Aryl hydrocarbon receptor; GSCs: GBM stem cells; PBK: PDZ binding kinase; CENPA: Centromere protein A; KIF15: Kinesin family member 15; GEO: Gene Expression Omnibus; TCGA: The Cancer Genome Atlas; PPI: Protein-protein interaction; CMAP: Connectivity map; RMA: Robust multi-array average; FC: Fold change; GDC: Genomic data commons; OS: Overall survival; KM: Kaplan-Meier; GO: Gene Ontology; HPAs: Human Protein Atlas; CETN2: Centrin 2; MKI67: Marker of proliferation ki-67; ARL13B: ADP ribosylation factor like GTPase 13B; SETDB1: SET domain bifurcated histone lysine methyltransferase 1; CALN1: Calneuron 1; ELAVL3: ELAV like RNA binding protein 3; ADCY3: Adenylate cyclase 3; SYN2: Synapsin II; SLC12A5: Solute carrier family 12 member 5; SOD1: Superoxide dismutase 1;

References

1. Pujadas E. Epigenetic Landscapes and Glioblastoma: Johns Hopkins University.
2. Alexander BM, Cloughesy TF. Adult glioblastoma. Journal of Clinical Oncology. 2017;35: 2402-2409.
3. Shergalis A, Bankhead A, Luesakul U, Muangsin N, Neamati N. Current challenges and opportunities in treating glioblastoma. Pharmacological reviews. 2018;70: 412-445.
4. Delgado-López P, Corrales-García E. Survival in glioblastoma: a review on the impact of treatment modalities. Clinical and Translational Oncology. 2016; 18: 1062-1071.
5. Jin U-H, Karki K, Cheng Y, Michelhaugh SK, Mittal S, et al. The aryl hydrocarbon receptor is a tumor suppressor–like gene in glioblastoma. Journal of Biological Chemistry. 2019; 294: 11342-11353.
6. Safa AR, Saadatzadeh MR, Cohen-Gadol AA, Pollok KE, Bijangi-Vishehsaraei K. Glioblastoma stem cells (GSCs) epigenetic plasticity and interconversion between differentiated non-GSCs and GSCs. Genes & Diseases 2015; 2: 152-163.

7. Stangeland B, Mughal AA, Grieg Z, Sandberg CJ, Joel M, et al. Combined expressional analysis, bioinformatics and targeted proteomics identify new potential therapeutic targets in glioblastoma stem cells. Oncotarget. 2015; 6: 26192.
8. Lu G, Rao M, Zhu P, Liang B, El-Nazer RT, et al. Triple-drug therapy with bevacizumab, irinotecan, and temozolomide plus tumor treating fields for recurrent glioblastoma: a retrospective study. Frontiers in neurology. 2019; 10: 42.
9. Tea MN, Poonnoose SI, Pitson SM. Targeting the Sphingolipid System as a Therapeutic Direction for Glioblastoma. Cancers. 2020; 12: 111.
10. Barrett T, Suzek TO, Troup DB, Wilhite SE, Ngau W-C, et al. NCBI GEO: mining millions of expression profiles—database and tools. Nucleic acids research. 2005; 33: D562-D566.
11. Barrett T, Troup DB, Wilhite SE, Ledoux P, Rudnev D, et al. NCBI GEO: mining tens of millions of expression profiles—database and tools update. Nucleic acids research. 2006; 35: D760-D765.
12. Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, et al. limma powers differential expression analyses for RNA-sequencing and microarray studies. Nucleic acids research. 2015; 43: e47-e47.
13. Bolstad BM, Irizarry RA, Åstrand M, Speed TP. A comparison of normalization methods for high density oligonucleotide array data based on variance and bias. Bioinformatics. 2003; 19: 185-193.
14. Irizarry RA, Hobbs B, Collin F, Beazer-Barclay YD, Antonellis KJ, et al. Exploration, normalization, and summaries of high density oligonucleotide array probe level data. Biostatistics. 2003; 4: 249-264.
15. Zhou G, Soufan O, Ewald J, Hancock RE, Basu N, et al. NetworkAnalyst 3.0: a visual analytics platform for comprehensive gene expression profiling and meta-analysis. Nucleic acids research.
16. Goldman M, Craft B, Hastie M, Repecka K, Kamath A, et al. The UCSC Xena platform for public and private cancer genomics data visualization and interpretation. bioRxiv, 2019; 326470.
17. Zhou Y, Zhou B, Pache L, Chang M, Khodabakhshi AH, et al. Metascape provides a biologist-oriented resource for the analysis of systems-level datasets. Nature communications. 2019; 10: 1-10.
18. Szklarczyk D, Morris JH, Cook H, Kuhn M, Wyder S, et al. The STRING database in 2017: quality-controlled protein–protein association networks, made broadly accessible. Nucleic acids research. 2016; gkw937.
19. Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, et al. Cytoscape: a software environment for integrated models of biomolecular interaction networks. Genome research. 2003; 13: 2498-2504.
20. Chandrashekar DS, Bashel B, Balasubramanya SAH, Creighton CJ, Ponce-Rodriguez I, et al. UALCAN: a portal for facilitating tumor subgroup gene expression and survival analyses. Neoplasia. 2017; 19: 649-658.
21. Gao J, Aksoy BA, Dogrusoz U, Dresdner G, Gross B, et al. Integrative analysis of complex cancer genomics and clinical profiles using the cBioPortal. Sci Signal. 2013; 6: pl1-pl1.
22. Uhlen M, Zhang C, Lee S, Sjöstedt E, Fagerberg L, et al. A pathology atlas of the human cancer transcriptome. Science. 2017; 357: eaan2507.

23. Subramanian A, Narayan R, Corsello SM, Peck DD, Natoli TE, et al. A next generation connectivity map: L1000 platform and the first 1,000,000 profiles. Cell. 2017; 171: 1437-1452. e1417.
24. Hanif F, Muzaffar K, Perveen K, Malhi SM, Simjee SU. Glioblastoma multiforme: a review of its epidemiology and pathogenesis through clinical presentation and treatment. Asian Pacific journal of cancer prevention: APJCP. 2017; 18: 3.
25. Yang S, Gao K, Li W. Identification of hub genes and pathways in glioblastoma by bioinformatics analysis. Oncology letters. 2019; 17: 1035-1041.
26. Dani N, Broadie K. Glycosylated synaptomatrix regulation of trans-synaptic signaling. Developmental neurobiology. 2012; 72: 2-21.
27. Taniguchi N, Kizuka Y. Glycans and cancer: role of N-glycans in cancer biomarker, progression and metastasis, and therapeutics. Advances in cancer research: Elsevier. 2015; pp. 11-51.
28. Bassoy EY, Kasahara A, Chiusolo V, Jacquemin G, Boydell E, et al. ER–mitochondria contacts control surface glycan expression and sensitivity to killer lymphocytes in glioma stem-like cells. The EMBO journal. 2017; 36: 1493-1512.
29. Lumeng JY, Wall BA, Wangari-Talbot J, Chen S. Metabotropic glutamate receptors in cancer. Neuropharmacology. 2017; 115: 193-202.
30. Krasikova YS, Rechkunova N, Maltseva E, Craescu C, Petruseva I, et al. Influence of centrin 2 on the interaction of nucleotide excision repair factors with damaged DNA. Biochemistry (Moscow). 2012; 77: 346-353.
31. Kamileri I, Karakasilioti I, Garinis GA. Nucleotide excision repair: new tricks with old bricks. Trends in genetics. 2012; 28: 566-573.
32. Huan J, Gao X, Xing L, Qin X, Qian H, et al. Screening for key genes associated with invasive ductal carcinoma of the breast via microarray data analysis. Genet Mol Res. 2014; 13: 7919-7925.
33. Tatarano S, Chiyomaru T, Kawakami K, Enokida H, Yoshino H, et al. miR-218 on the genomic loss region of chromosome 4p15. 31 functions as a tumor suppressor in bladder cancer. International journal of oncology. 2011; 39: 13-21.
34. Hou Y-Y, Cao W-W, Li L, Li S-P, Liu T, et al. MicroRNA-519d targets MKi67 and suppresses cell growth in the hepatocellular carcinoma cell line QGY-7703. Cancer letters. 2011; 307: 182-190.
35. Laible M, Schlombs K, Kaiser K, Veltrup E, Herlein S, et al. Technical validation of an RT-qPCR in vitro diagnostic test system for the determination of breast cancer molecular subtypes by quantification of ERBB2, ESR1, PGR and MKI67 mRNA levels from formalin-fixed paraffin-embedded breast tumor specimens. BMC cancer. 2016; 16: 398.
36. Wang R-j, Li J-w, Bao B-h, Wu H-c, Du Z-h, et al. MicroRNA-873 (miRNA-873) inhibits glioblastoma tumorigenesis and metastasis by suppressing the expression of IGF2BP1. Journal of Biological Chemistry. 2015; 290: 8938-8948.
37. Kong J, Wang F, Teodoro G, Cooper L, Moreno CS, et al. High-performance computational analysis of glioblastoma pathology images with database support identifies molecular and survival correlates; 2013. IEEE. pp. 229-236.

38. Győrffy B, Lanczky A, Eklund AC, Denkert C, Budczies J, et al. An online survival analysis tool to rapidly assess the effect of 22,277 genes on breast cancer prognosis using microarray data of 1,809 patients. Breast cancer research and treatment. 2010; 123: 725-731.
39. Casalou C, Faustino A, Silva F, Ferreira IC, Vaqueirinho D, et al. Arl13b Regulates Breast Cancer Cell Migration and Invasion by Controlling Integrin-Mediated Signaling. Cancers. 2019; 11: 1461.
40. Fuks F. DNA methylation and histone modifications: teaming up to silence genes. Current opinion in genetics & development. 2005; 15: 490-495.
41. Wang G, Long J, Gao Y, Zhang W, Han F, et al. SETDB1-mediated methylation of Akt promotes its K63-linked ubiquitination and activation leading to tumorigenesis. Nature cell biology. 2019; 21: 214-225.
42. Spyropoulou A, Gargalionis A, Dalagiorgou G, Adamopoulos C, Papavassiliou KA, et al. Role of histone lysine methyltransferases SUV39H1 and SETDB1 in gliomagenesis: modulation of cell proliferation, migration, and colony formation. Neuromolecular medicine. 2014; 16: 70-82.
43. Gong L, Bao Q, Hu C, Wang J, Zhou Q, et al. Exosomal miR-675 from metastatic osteosarcoma promotes cell migration and invasion by targeting CALN1. Biochemical and biophysical research communications. 2018; 500: 170-176.
44. Pignolet BS, Gebauer CM, Liblau RS. Immunopathogenesis of paraneoplastic neurological syndromes associated with anti-Hu antibodies: a beneficial antitumor immune response going awry. Oncoimmunology. 2013; 2: e27384.
45. CAUDERAY A. ELAVL3: A NEW PROTEIN PROMOTING THE GROWTH OF PRONEURAL GLIOBLASTOMA?: Université de Lausanne, Faculté de biologie et médecine. 2016.
46. Goni L, Riezu-Boj JI, Milagro FI, Corrales FJ, Ortiz L, et al. Interaction between an ADCY3 genetic variant and two weight-lowering diets affecting body fatness and body composition outcomes depending on macronutrient distribution: A randomized trial. Nutrients. 2018; 10: 789.
47. Hong S-H, Goh S-H, Lee SJ, Hwang J-A, Lee J, et al. Upregulation of adenylate cyclase 3 (ADCY3) increases the tumorigenic potential of cells by activating the CREB pathway. Oncotarget. 2013; 4: 1791.
48. Labak CM, Wang PY, Arora R, Guda MR, Asuthkar S, et al. Glucose transport: meeting the metabolic demands of cancer, and applications in glioblastoma treatment. American journal of cancer research. 2016; 6: 1599.
49. Kato S, Esumi H, Hirano A, Kato M, Asayama K, et al. Immunohistochemical expression of inducible nitric oxide synthase (iNOS) in human brain tumors: relationships of iNOS to superoxide dismutase (SOD) proteins (SOD1 and SOD2), Ki-67 antigen (MIB-1) and p53 protein. Acta neuropathologica. 2003; 105: 333-340.
50. Gao Z, Sarsour EH, Kalen AL, Li L, Kumar MG, et al. Late ROS accumulation and radiosensitivity in SOD1-overexpressing human glioma cells. Free Radical Biology and Medicine. 2008; 45: 1501-1509.
51. Xiao J, Yiqing Q. Bioinformatics analysis of the gene expression profile in Bladder carcinoma. Genetics and molecular biology. 2013; 36: 287-291.

52. Raghavan R, Hyter S, Pathak HB, Godwin AK, Konecny G, et al. Drug discovery using clinical outcome-based Connectivity Mapping: application to ovarian cancer. BMC genomics. 2016; 17: 811.
53. Gordaliza M, Castro Md, Miguel del Corral J, Feliciano AS. Antitumor properties of podophyllotoxin and related compounds. Current pharmaceutical design. 2000; 6: 1811-1839.
54. Dar RA, Brahman PK, Tiwari S, Pitre KS. Adsorptive stripping voltammetric determination of podophyllotoxin, an antitumour herbal drug, at multi-walled carbon nanotube paste electrode. Journal of Applied Electrochemistry. 2011; 41: 1311.
55. Sk UH, Dixit D, Sen E. Comparative study of microtubule inhibitors–estramustine and natural podophyllotoxin conjugated PAMAM dendrimer on glioma cell proliferation. European journal of medicinal chemistry. 2013; 68: 47-57.

Tables

Table 1 Characteristics of studies composing the gene expression compendium

Dataset	Study(Citation)	Platform	Organism	Sample (Glioblastoma)	Sample (Control)
GSE103227	chun luo□2018	Agilent-045997 Arraystar human lncRNA microarray V3	Homo sapiens	5	5
GSE104267	Jianjun Gu□2017	Phalanx Human lncRNA OneArray v1_mRNA	Homo sapiens	9	3
GSE111260	Jeanmougin Jeanmougin□2018	Affymetrix Human Exon 1.0 ST Array	Homo sapiens	67	3

Table 2 The degree value of hub genes in PPI network

Symbol	Degree	Betweenness	Closeness
ELAVL3	18	649.4292	0.479452
HDAC2	17	758.398	0.486111
CALB1	16	283.0139	0.469799
CUL3	16	345.8947	0.479452
SYN2	15	363.5061	0.44586
CIT	13	246.3135	0.434783
SHANK2	13	332.2537	0.47619
SLC12A5	12	300.8864	0.4375
SOD1	11	222.0949	0.457516
SETDB1	11	338.4737	0.457516
CALN1	10	77.96878	0.434783
CAP2	10	185.884	0.434783
ARL13B	10	185.142	0.406977
ADCY3	10	160.4279	0.414201
CETN2	10	111.9884	0.414201
MKI67	10	247.3053	0.434783

Table 3 The expression level of hub genes

Genes	Comparison	Statistical significance	TCGA type	GEO type
MKI67	Normal-vs-Primary	1.62E-12	up	up
ELAVL3	Normal-vs-Primary	7.37E-12	down	down
ADCY3	Normal-vs-Primary	4.05E-06	down	
CETN2	Normal-vs-Primary	8.01E-05	up	
SOD1	Normal-vs-Primary	1.18E-03	down	
ARL13B	Normal-vs-Primary	9.18E-03	up	up
CALN1	Normal-vs-Primary	2.39E-02	down	down
SYN2	Normal-vs-Primary	2.61E-02	down	down
SETDB1	Normal-vs-Primary	2.81E-02	up	
SLC12A5	Normal-vs-Primary	4.31E-02	down	down

Figures

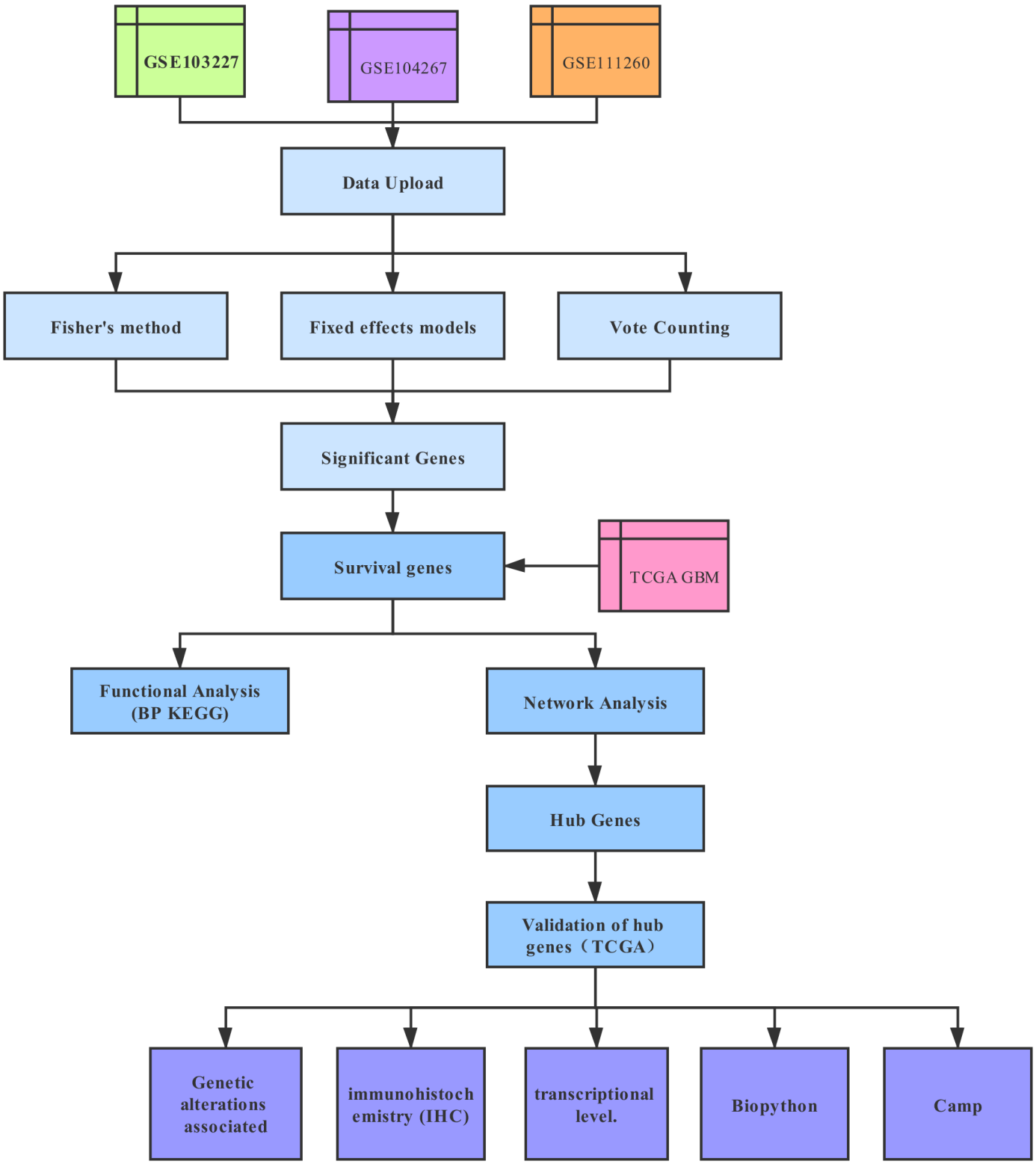

Figure 1

A flow chart of this study

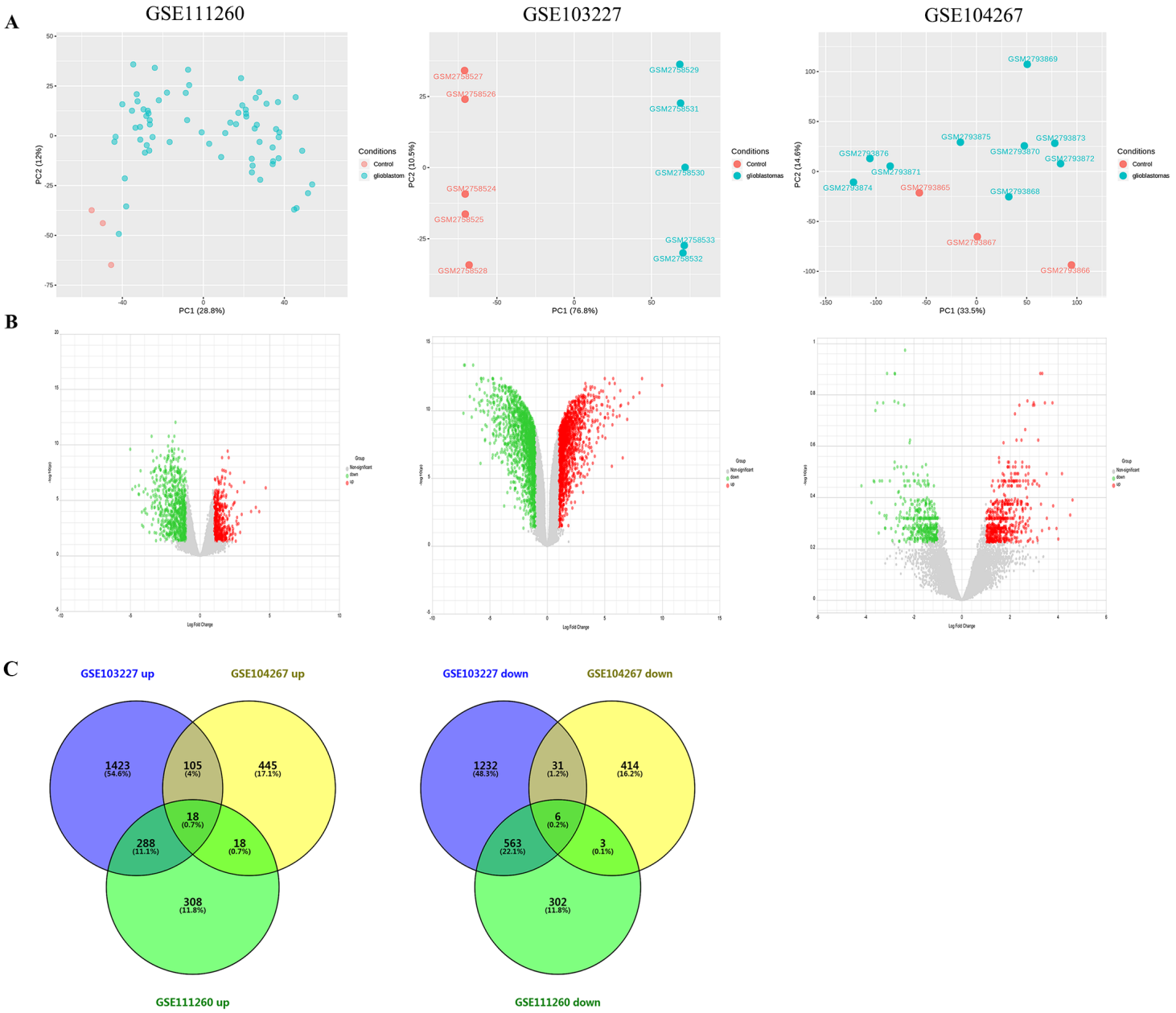

Figure 2

Identification of DEGs in three GEO datasets. A: PCA plot. Red represents control sample, and blue represents GBM sample. B: Volcano plot. Green indicates down-regulated DEGs, and red indicates up-regulated DEGs. C: VENN diagram of DEGs identified from three datasets (left: up-regulated DEGs, right: down-regulated DEGs).

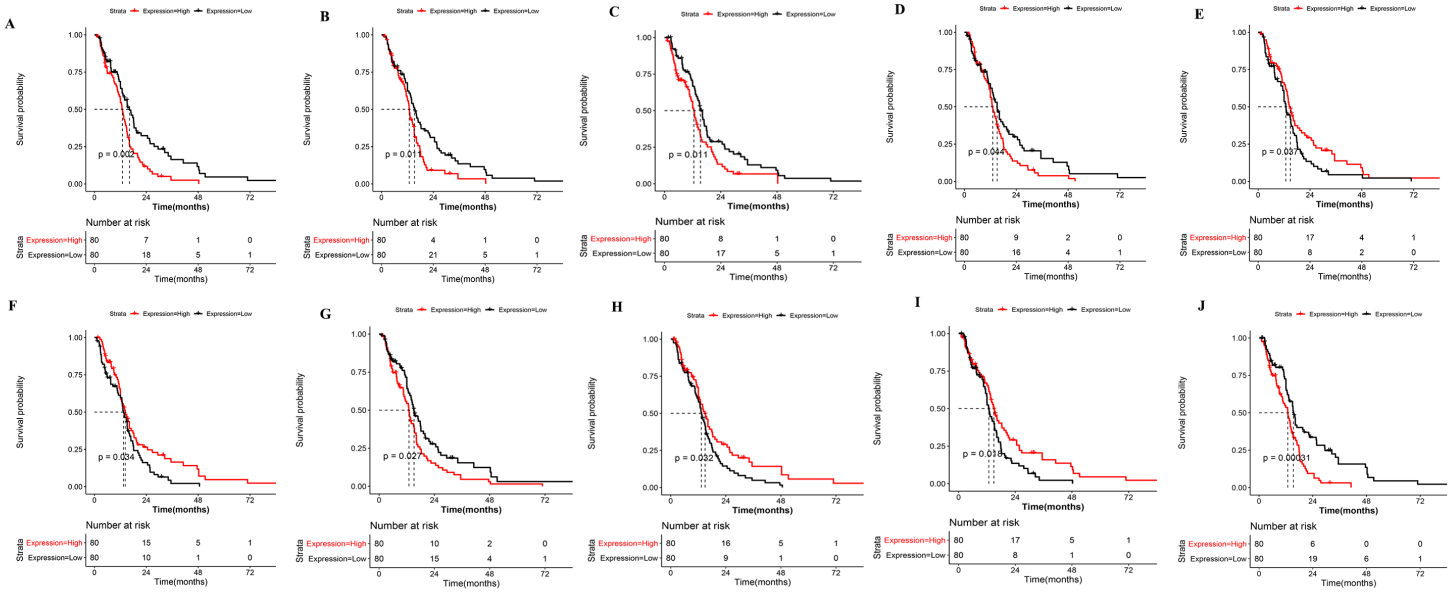

Figure 3

Survival analysis for hub genes in GBM. Kaplan-Meier plots show 10 hub genes related to overall survival rate (P<0.05). A: CETN2, B: MKI67, C: ARL13B, D: SETDB1, E: CALN1, F: ELAVL3, G: ADCY3, H: SYN2, I: SLC12A5, J: SOD1

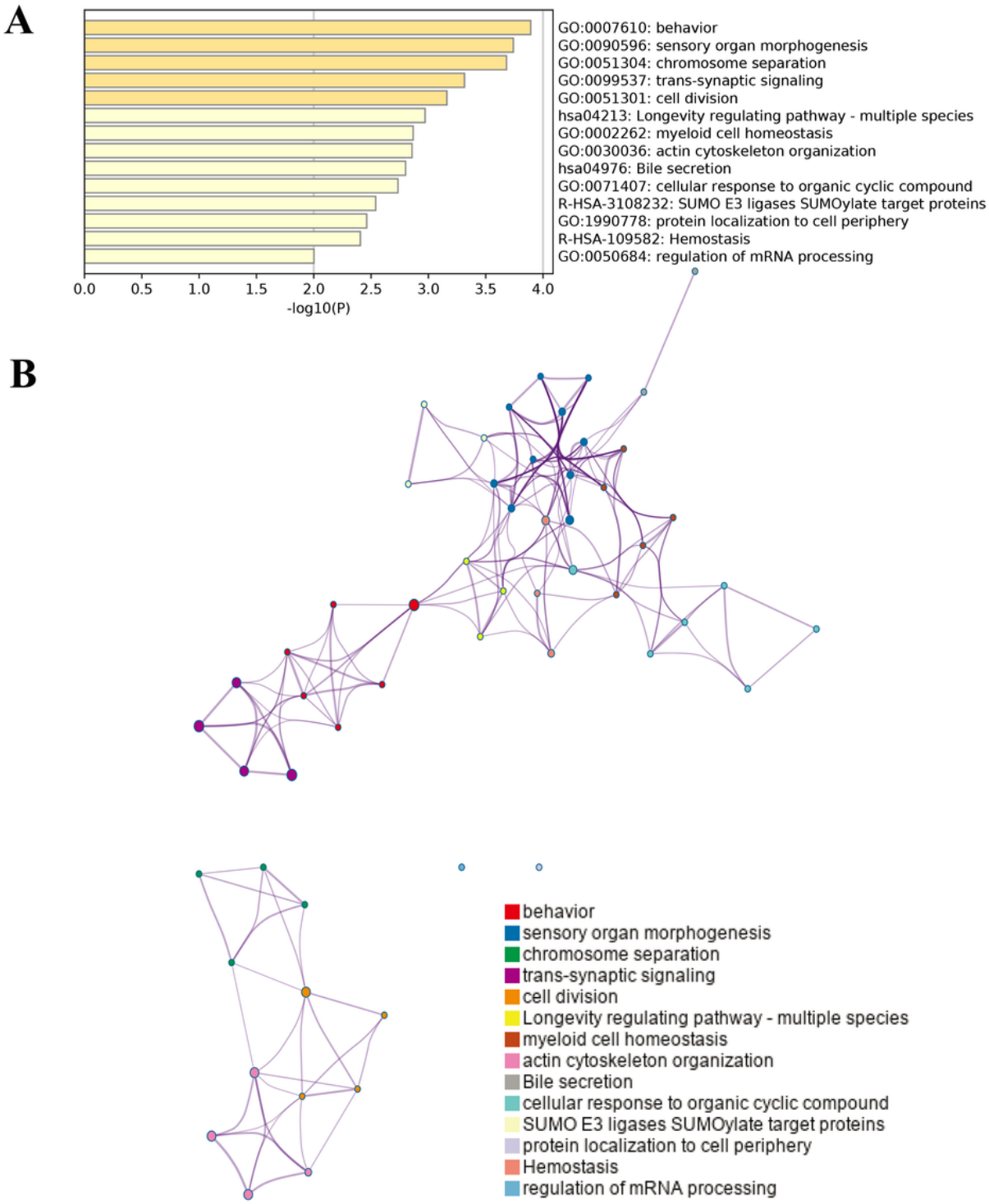

Figure 4

Functional enrichment analysis of hub genes. A: heatmap of functional enrichment results. The terms are colored by -log10(P) value. B: network of clusters of enriched terms. Each node represents one enriched term colored by cluster ID, nodes share the same cluster are typically close to each other. Terms with Kappa similarity above 0.3 were connected.

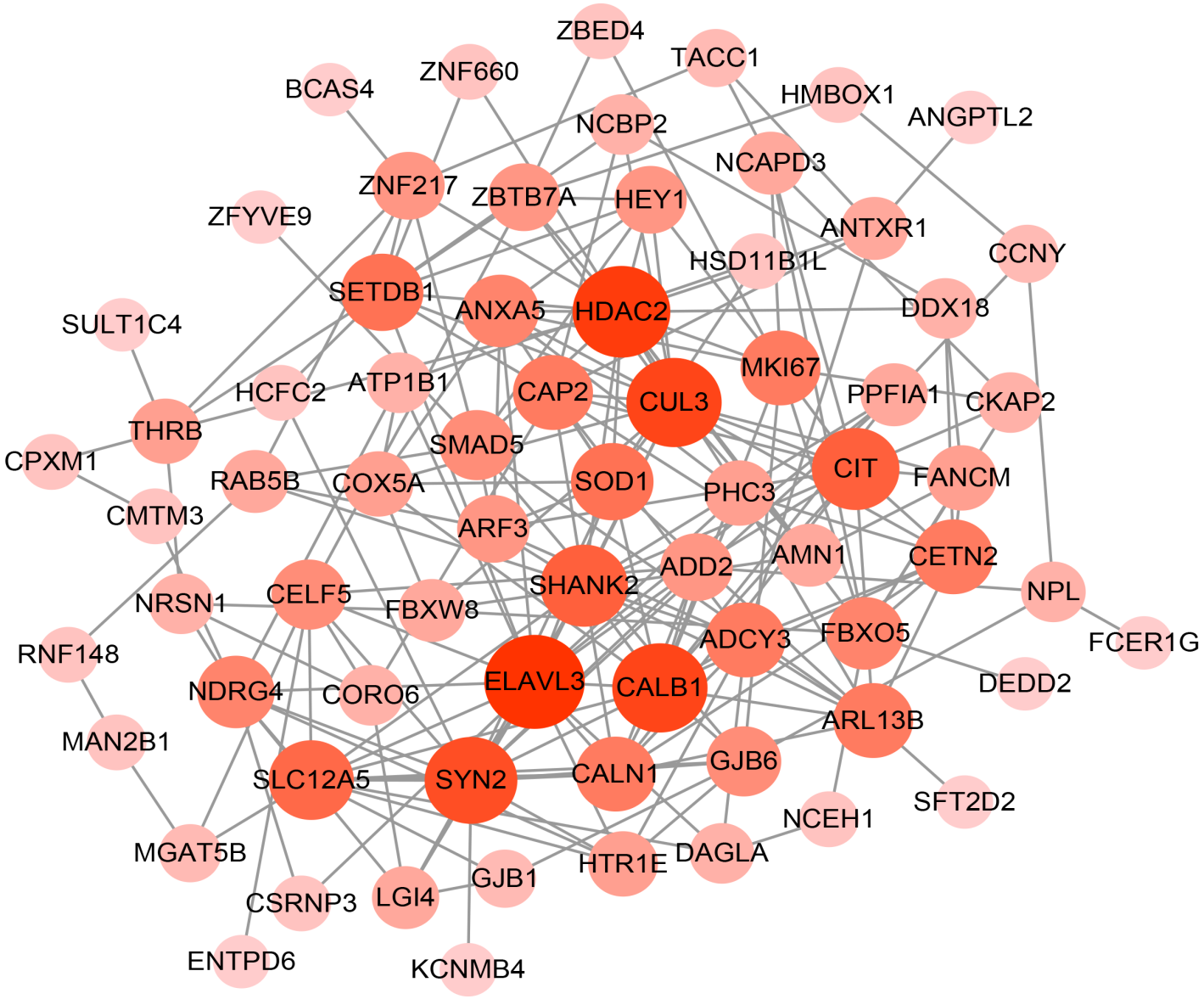

Figure 5

PPI network analysis. The color depth of nodes represents the corrected P-value. The size of nodes represents the number of genes involved.

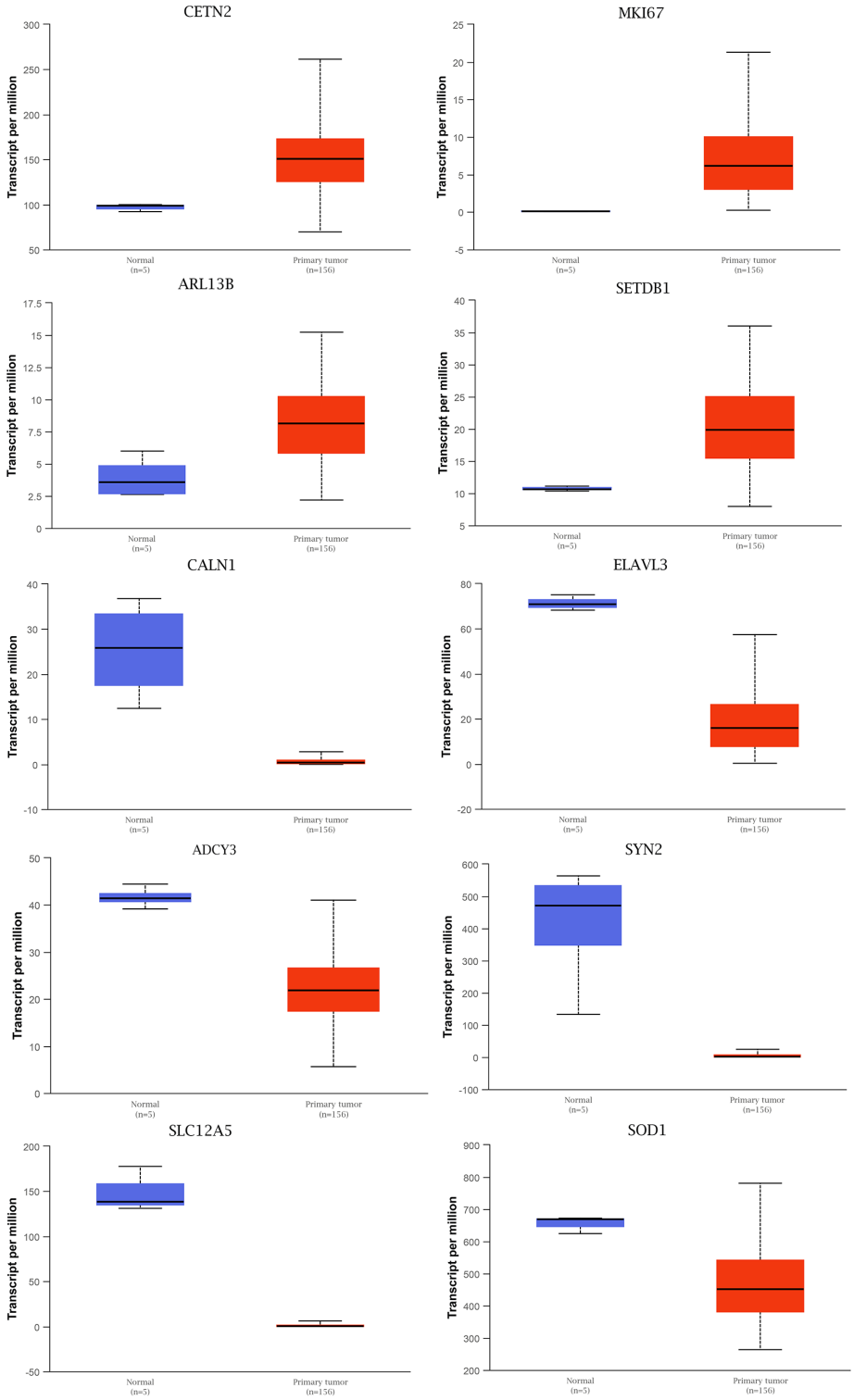

Figure 6

The expression level of hub genes between GBM and normal samples in three datasets. Blue box indicates normal tissue, and red box indicates GBM tissue.

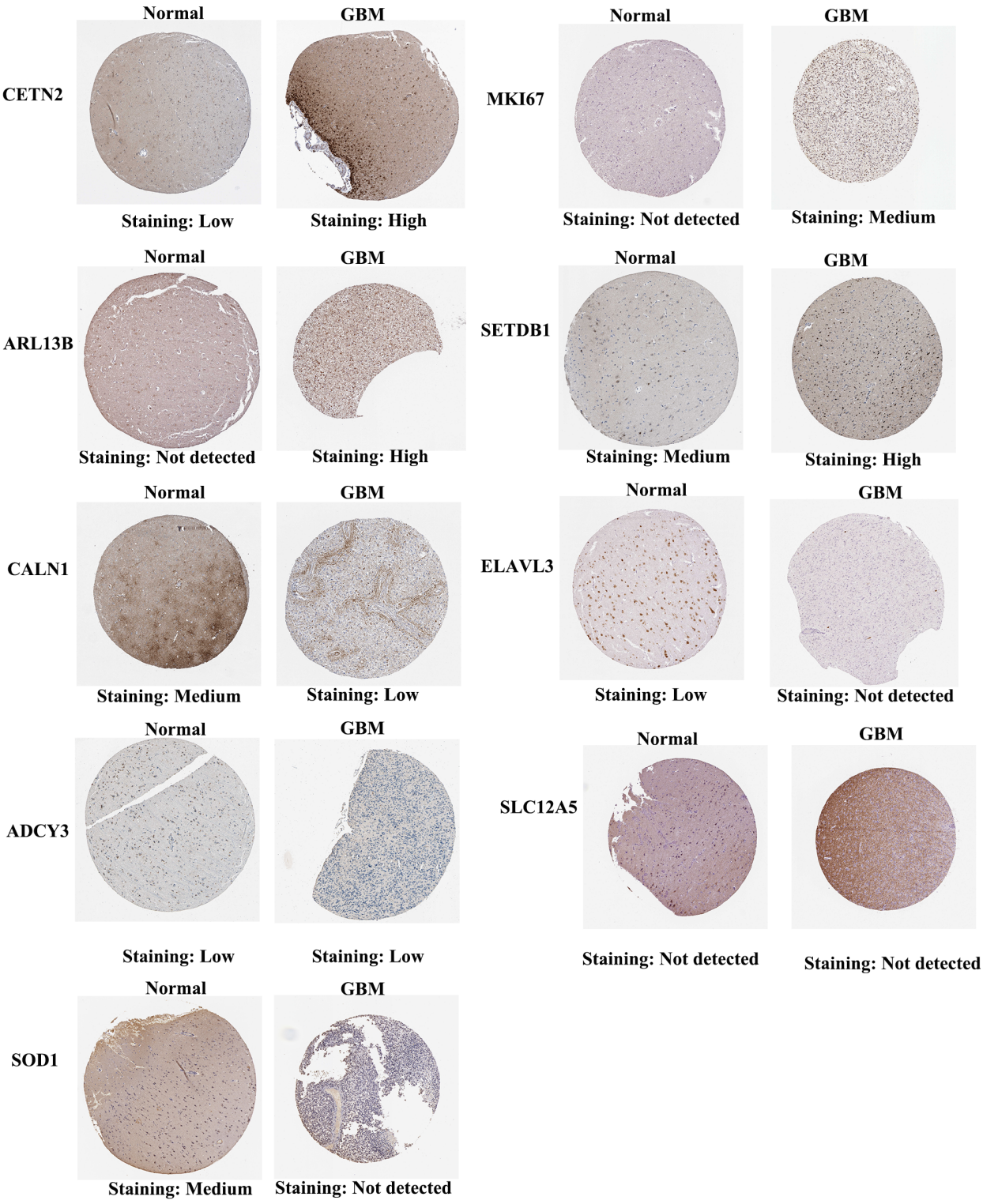

Figure 7

The protein expression level of hub genes in GBM tissues and normal tissues from HPA database

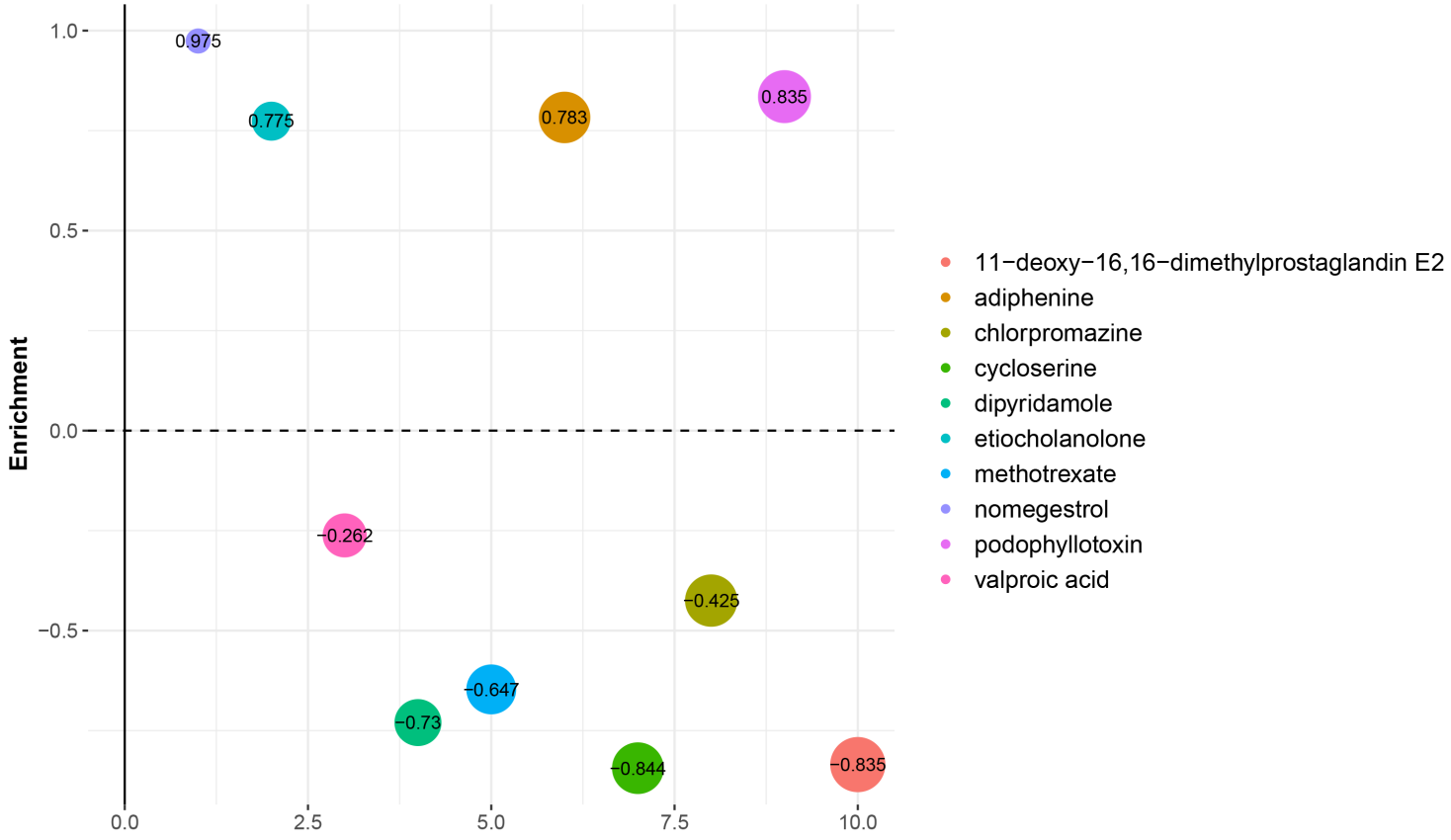

Figure 8

The top 10 small molecule drugs from the CMAP database The bubble size represents p value, the smaller the p value, the larger the bubble.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- SupplementaryFigure2.tif
- SupplementaryTable1.docx
- SupplementaryFigure1.tif