Investigation of Candidate Genes and Mechanism Associated with Immune Cell during the Progression of Glioblastoma based on Bioinformatics

Hui Xie

Department of Histology and Embryology, College of Basic Medicine, Shenyang Medical College

Xiao-hui Ding

Department of Histology and Embryology, College of Basic Medicine, Shenyang Medical College

Ce Yuan

Graduate Program in Bioinformatics and Computational Biology, University of Minnesota

Jin-jiang Li

Department of Neurosurgery, General Hospital of Northern Theater Command

Zhao-yang Li

Department of Laboratory Animal Center, China Medical University

Wei-cheng Lu (✉ 87k10b@163.com)

First Affiliated Hospital of China Medical University <https://orcid.org/0000-0001-6498-0931>

Research

Keywords: glioblastoma multiforme, differentially expressed genes, M2 Macrophage, competing endogenous RNAs, survival analysis, prognostic markers

Posted Date: December 2nd, 2020

DOI: <https://doi.org/10.21203/rs.3.rs-117485/v1>

License: This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Abstract

Background: This study aimed to investigate the molecular mechanism of immune cell infiltration during the development of glioblastoma multiforme (GBM), and explore the potential immune cell associated prognostic genes for GBM.

Methods: Gene expression data and corresponding clinical data of GBM samples (tumor group) and normal samples (normal group) in TCGA-GBM and GTEx dataset were downloaded. The differentially expression analysis was performed on samples between two groups. Based on tumor immune microenvironment analysis, the immune-related RNAs (lncRNAs and mRNAs) were further explored. Then, functional enrichment analysis, ceRNA network, risk prediction model and prognosis investigation were performed. Finally, the results of survival prognosis of key genes were tested by additional datasets.

Results: A total of 4989 up-regulated genes and 2349 down-regulated genes were revealed between tumor group and normal group. M2 macrophages accounted for the largest proportion of tumor infiltrates immune cells in GBM, and was related to the prognosis of GBM patients. Totally 168 mRNAs (KIF18B) and 5 lncRNAs were related to infiltration of M2 Macrophage, of which 25 mRNAs and 5 lncRNAs forms a ceRNA network through 37 miRNAs (eg., miR-6849-3p). These genes were mainly assembled in functions like signal release. A risk model based on 5 mRNAs (such as FOX4 and ELFN2) and lncRNA PR11-161H23.5 was constructed. Verification test showed that all 5 mRNAs were significantly associated with OS prognosis.

Conclusions: M2 Macrophage infiltration might participate in tumorigenesis of GBM via RP11-161H23.5-miR6849-3p-KIF18B ceRNA interaction. Furthermore, mRNAs such as FOX4 and ELFN2 might be potential prognostic markers for GBM patients.

Background

Glioblastoma multiforme (GBM) is a common malignant neuroepithelial tumor of the central nervous system in adults [1]. The survival of GBM following diagnosis is only 12–15 months [2]. Typically, surgery after chemotherapy and radiation therapy are common used for the treatment of GBM [3]. However, the poor effect of existing treatments on the diffusive, infiltrative, and metastatic block the clinical therapy of GBM [4].

The cancer immunotherapy are proved to be valuable in clinical trials of patients with GBM [5]. Innovative novel immunotherapies for the treatment of GBM is hopeful since targeting the immune system contribute to the GBM therapy with significant clinical benefits [6]. Actually, the immune infiltration of tumor microenvironment followed by immunotherapy in GBM has been revealed [7]. A previous study proves that T cells can infiltrate the tumor bed [8]. Kmiecik et al. showed that the survival rate of patients with GBM is closed related with the expression of CD3 + and CD8 + in immune cells [9]. Fortunately, the detail mechanism of immune cell infiltration of immune cells during the progression of tumor development can be revealed by the gene expression analysis [10]. Based on correlation analysis of

immune-related functions, immune cell infiltration abundance and patient survival, Zhao et al. showed that ADAMTSL4 gene was a novel immune-related biomarker for primary GBM multiforme [11]. Based on TCGA datasets and Cell-type Identification by Estimating Relative Subsets of RNA Transcripts (CIBERSORT) algorithm, a recent gene expression-based study determined the relative proportions of 22 infiltrative immune cells, which closely associated with the GBM prognosis [12]. However, due to the lack of detail mechanisms of immune cell-related genes in tumor, possible prediction models and subsequent data validation, the potential immune cell-related genes, as well as correlation between these genes and GBM prognosis are still unclear.

In this study, the gene expression data and clinical information of GBM tumor samples and normal tissue samples in TCGA-GBM dataset and GTEx dataset were downloaded. The differentially expression analysis was performed on samples between two groups. Then, based on tumor immune microenvironment analysis, the immune-related differentially expressed RNAs (DEGs), differentially expressed lncRNAs (DElncRNAs) and differentially expressed mRNAs (DEMs) were further investigated. The enrichment analysis was performed on DEMs, followed by mRNAs associated miRNAs were predicted. Moreover, the lncRNA-miRNA-mRNA (competing endogenous RNAs (ceRNA)) network study, risk prediction model and prognosis evaluation were performed based on key genes. Finally, the results of survival prognosis of key genes were tested by additional datasets. This study hoped to investigate the detail molecular mechanism of immune cell during the development of GBM, and explore the potential immune cell associated genes for prognosis of GBM patients.

Methods

The flow chart of experimental design in this study was listed in Supplementary Figure 1.

Microarray data

The RNA-seq data including corresponding clinical phenotype data of GBM samples in TCGA and GTEx were downloaded from UCSC database. A total of 166 GBM samples obtained from TCGA-GBM dataset were enrolled as tumor group. Meanwhile, 5 normal paracancerous tissue samples from TCGA-GBM dataset and 105 normal brain cortex samples from GTEx dataset were combined as normal group. Moreover, all transcripts per million (TPM) format data, phenotypic data and survival information data of all enrolled samples were further obtained.

Differentially expression analysis

The linear regression and empirical bayesian methods [13, 14] in limma package (3.40.6) of R [15] were used to explore DEGs between tumor group and normal group. The $P < 0.05$ and $|\log_2 \text{fold change (FC)}| > 1$ were selected as the thresholds for DEGs screening. Then, the volcano plots and clustering heatmap were constructed using ggplot2 (version: 3.2.1) [16] and using pheatmap (version: 1.0.12) [17],

respectively. Moreover, the DEMs and DElncRNAs from DEGs were explored by gene annotation based on the HUMAN Release 23 gene annotation information in GENCODE database. Finally, the GEPIA 2 database [18] (<http://gepia2.cancer-pku.cn/#index>) was used to verify the results in current differentially expression analysis.

Immune score and cytolytic activity analysis

ESTIMATE (version: 1.0.13) algorithm [19] was used to estimate the scores of stromal cells and immune cells in reactive tumors. Furthermore, the immune cytolytic score was calculated by the log-average of GZMA and PRF1 expression values in TPM. Finally, the differences of immune score and cytolytic activity between tumor and normal samples were analyzed by T test, and were visualized by Boxplot using ggpubr (version: 0.2.2) [20] software in R. Cibersort algorithm [21] was performed to reveal infiltration abundance of 22 kinds of immune cells. The calculation results of TCIA database [22] were used for verification. Furthermore, TIMER algorithm [23] was used to estimate infiltration abundance of 6 kinds of immune cells (B cells, CD4 T cells, CD8 T cells, Neutrophil, Macrophage and Dendritic cells). Meanwhile, the relationship between immune cells infiltration abundance and the overall survival (OS) was explored by log-rank test, followed by the Kaplan-Meier (K-M) survival curves construction using survival (version: 2.44-1.1) in R. Finally, based on the results of the two algorithms, the immune cells with different abundances between GBM and normal samples and related to the prognosis of GBM were selected as the focus of follow-up analysis.

Immune cells associated DEGs investigation

The Spearman correlation test between GBM and normal samples was performed on DEMs, DElncRNAs and infiltration abundance of immune cells using corrplot (version: 0.84) [24] in R. The $P < 0.05$ and correlation coefficient $|r| > 0.3$ were selected as the cutoff value for screening the immune associated DEMs and DElncRNAs.

Enrichment analysis of the immune associated DEMs

GO function [25] and KEGG pathway [26] enrichment analyses of immune associated DEMs and DElncRNAs were performed using the clusterProfiler (version: 3.12.0) software [27]. Based on Fisher's exact test, the P value < 0.05 was considered as cut-off values of significant enrichment results. The results were visualized by Bubble chart and histogram.

The miRNA prediction and co-expression investigation

The lncRNA-mRNA pairs were identified from all DElncRNAs and DEMs by Pearson correlation analysis using corrplot (version: 0.84) [24] in R. The lncRNA-mRNA pairs with $r > 0.5$ and $P < 0.01$ were enrolled for

the further investigation. Furthermore, the miRNAs prediction for DEMs which correlated with mRNAs were performed using miRWalk3.0 software (parameters: binding probability = 0.95; binding site position = 3'UTR) [28]. Then, totally 4 databases including miRWalk3.0 [28], miRDB [29], TargetScan [30] and MirTarBase [31] were used to explore the potential mRNAs-miRNAs relations. Finally, the mRNAs-miRNAs interactions that validated by at least 3 databases (must include MirTarBase) were enrolled for the further investigation.

The online tool LncBase Predicted v.2 was used to predict miRNAs targeting immune related DElncRNAs. The miRNA-lncRNA regulatory relations were screened with the threshold score > 0.7. The lncRNA-miRNA-mRNA interactions were screened by integrating miRNA-mRNA, miRNA-lncRNA and mRNA-lncRNA interactions, and the ceRNA network was visualized by Cytoscape software [32].

PPI network construction

According to STING database (version: 10.0) [33], the information of the protein interaction was explored, and PPI pairs among mRNAs in ceRNA network were predicted (minimum required interaction score = 0.4). Then, the results of these interactions were visualized by Cytoscape software [32].

Survival analysis

According to the OS and progression free survival (PFS) survival prognosis data of 165 GBM samples with survival information, the prognostic value of mRNAs and lncRNAs (from ceRNA network) on GBM patients were investigated. The mRNAs and lncRNAs significantly related to the prognosis of OS and PFS were screened by single factor Cox expression regression, and the significance p value and the prognosis correlation coefficient β value were obtained by log-rank test, followed by the Risk score calculation with:

$$\text{Risk score} = \beta_{gene1} \times expr_{gene1} + \beta_{gene2} \times expr_{gene2} + \dots + \beta_{genen} \times expr_{genen}$$

The corresponding risk score of each sample was calculated, and then with median risk score as the boundary, all samples were divided into high risk group or low risk group. The Kaplan-Meier (KM) [34] survival curve was used to assess the association between risk models and survival outcomes in GBM samples. Furthermore, the computational model for predicting clinical OS and PFS endpoints by the expression value of screened key genes in samples was constructed with svm package, followed by the area under curve (AUC) calculation based on receiver operating characteristic (ROC) curve. AUC between 0.1 and 1 was used as a numerical value to evaluate the quality of the classifier intuitively.

Prognosis verification based on external data set

A total of 4 GBM related GEO dataset including GSE42669, GSE37418, GSE7696 and GSE2817 were selected to verify the survival prognosis of the screened key genes By using SurvExpress software [35]. Furthermore, the mRNA-seq 693.RSEM-genes.20190909 profile and associated clinical data mRNAseq_693.clinical.20190909 were downloaded from Co-expressed Gene Groups Analysis (CGGA) database (<http://www.cgga.org.cn/index.jsp>). The GBM and rGBM samples were selected from the original GEO dataset, followed by the survival information extraction including OS and Censor status. Then, these data were integrated with the expression data of the screened key genes. The samples were divided into two groups with high expression and low expression according to the median expression level, followed by the log-RANK test analysis. Finally, the result was visualized by K-M survival curve.

Results

Differentially expression analysis

A total of 4989 up-regulated genes and 2349 down-regulated genes were revealed between tumor group and normal group. Among them, there were 4412 up-regulated and 1805 down-regulated mRNAs, as well as 155 up-regulated and 316 down-regulated lncRNAs between two groups (Supplementary Figure 2). The verification results based on the GEPIA2 database show that there were totally 2454 down-regulated genes and 5213 up-regulated genes. Moreover, there were 2229 intersected down-regulated genes (94.89% of GEPIA2 analysis results) and 4778 intersected up-regulated genes (95.77% of GEPIA2 analysis results), which indicating a reliable results of current differentially expression analysis.

Tumor immune score and cytolytic activity analysis

The Box plots were used to reveal the tumor immune score and cytolytic activity between tumor samples and normal samples. The result showed that the stromal score (Supplementary Figure 3A), immune score (Supplementary Figure 3B) and ESTIMATE score (Supplementary Figure 3C) in GBM group were significantly higher (all $P < 0.05$) than those in normal group. Moreover, the cytolytic activity score in GBM group was significantly higher ($P < 0.05$) than that in normal group (Supplementary Figure 3D).

Immune cell infiltration analysis based on CIBERSORT and TIMER

In the expression matrix of CIBERSORT, a total of 335 genes were included in the 547 genes of LM22 signal matrix (accounting for 61.4%). A total of 22 immune cell infiltration abundance matrices were estimated by CIBERSORT algorithm. In all 276 samples, only 66 samples were valid ($P < 0.05$), including 65 GBM samples and 1 normal sample (Figure 1A). Among these 66 samples, the infiltration ratio of M2 macrophages cells in GBM samples reached an average of 0.458 and a maximum of 0.759. The finding was confirmed in TCIA database that macrophages M2 cells had the highest average infiltration based

on 604 samples (Figure 1B). However, in the unique normal sample (GTEX-ZYY3-3126-SM-5SI9), the infiltration ratio of M2 Macrophages was only 0.109, which was much lower than the GBM sample. Due to the small sample size, the result could not be verified in this data. Furthermore, the clustering heat map of 22 immune cell infiltration abundances was showed in Supplementary Figure 4. This result showed that the infiltration degree of macrophages M2 cells in GBM samples was significantly higher than that of other immune cells.

Based on TIMER, the immunocyte infiltration histograms of GBM vs. normal showed that the infiltration degree of Macrophage cells in GBM samples were significantly higher ($P < 0.05$) than those in normal samples (Figure 1C). Meanwhile, the Violin chart of infiltration abundance showed that except for B cells ($P = 0.489$) and CD4 T cells ($P = 0.006$), the infiltration abundance between two groups in all other cells were significantly different (all $P < 0.01$). The heatmap for GBM immune cell subsets was showed in Supplementary Figure 5. Furthermore, the ROC analysis of infiltration abundance of 6 immune cells and OS showed that Macrophage cell infiltration abundance was significantly related to sample survival time (Supplementary Figure 6). The lower the infiltration abundance, the longer the overall survival of the patient. Based on the results of the two algorithms, Macrophages, which had different abundance in GBM vs. normal and associated with the prognosis of GBM, were selected as the focus for the further investigation.

Immune cell associated DEMs/DElncRNAs investigation and enrichment analysis

Correlation analysis between Macrophages M2 infiltration abundance and gene expression was performed to explore the immune cell associated DEMs and DElncRNAs. The result showed that there were 168 immune-related DEMs and 5 immune-related DElncRNAs (including RP11-161H23.5, RP11-713C5.1, AC007620.3, RP11-74E22.8 and RP11-126K1.6) were revealed as immune cell associated genes by screening with $P < 0.015$.

The GO function analysis showed that the immune-related DEMs were mainly assembled in functions like Cognition (GO:0050890; Gene: BRSK1), Signal release (GO:0099643; Gene: FOX4 etc.) and Regulation of neurotransmitter transport (GO:0051590; Gene: BRSK1) (Figure 2A). Furthermore, the KEGG pathway enrichment analysis showed that the immune cell associated DEMs were mainly enriched in pathways like Endocrine resistance (hsa01522), Viral protein interaction with cytokine and cytokine receptor (hsa04061), as well as Chemokine signaling pathway (hsa04062) (Figure 2B).

CeRNA network construction

Person correlation analysis of lncRNA-mRNA interaction revealed that totally 137 mRNAs were positively correlated with 5 lncRNAs with $r > 0.5$ and $P < 0.01$. The top-10 mRNA-lncRNA interactions according to the correlation coefficient were showed in Supplementary Table 1. The result showed that all the 10

mRNAs including RUNDC3A, BRSK1 and SPTBN2 were interacted with the same lncRNA RP11-74E22.8. Moreover, there were totally 41 mRNAs, 126 miRNAs and 140 interactions in current mRNA-miRNA relations. Furthermore, the lncRNA-miRNA interaction analysis revealed a total of 580 interactions, 5 lncRNAs and 501 miRNAs.

A total of 37 miRNAs, 25 mRNAs and 5 lncRNAs (Supplementary Table 2 and Supplementary Table 3) were included in the current ceRNA interaction network (Figure 3A). The results showed that several lncRNA-miRNA-mRNA interactions such as RP11-161H23.5-miR6849-3p-KIF18B and RP11-74E22.8-miR-339-5p-BSK1 were involved in current ceRNA network. Moreover, the PPI network constructed by 25 mRNAs in current ceRNA network was showed in Figure 3B.

Risk prediction model construction based on prognostic analysis

Based on the COX univariate regression analysis, the mRNAs including SOX4, KIFC1, ELFN2, PLP2 and MEX3A, as well as lncRNA PR11-161H23.5 were related to the prognosis of PFS. Meanwhile, the K-M survival curve analysis showed that SOX4, KIFC1, ELFN2, MEX3A, and RP11-161H23.5 were significantly correlated with prognosis. However, the result of multivariate COX regression analysis showed that only RP11-161H23.5 was associated with prognosis (Supplementary Figure 7). Then, the risk prediction model based on the expression of SOX4, KIFC1, ELFN2, PLP2 and MEX3A and PR11-161H23.5 was used to predict the clinical end point of PFS. The result showed that the AUC of the prediction model was 0.828 with a specificity of 0.758 and a sensitivity of 0.909, which could be used to predict the clinical end point of PFS (Figure 4).

According to the median value of risk score (0.08325018), 165 samples were divided into high risk (H) groups and low risk (L) groups. The result of samples sorted by risk score was showed in the Figure 5A. Meanwhile, the scatter plot of sample grouping and PFS survival time showed a shorter survival time in H group (Figure 5B). Moreover, the K-M curve analysis showed that the PFS survival time of H group was significantly lower than that of L group (Figure 5C). Furthermore, the clustering heat map of 6 gene expression levels among the sample groups showed that the expression levels of SOX4, KIFC1, ELFN2, and MEX3A were lower in the L group when compared to the H group. Meanwhile, the expression of PLP2 and RP11-161H23.5 were higher in samples of L group when compared to the H group (Figure 6). The box plot analysis for 6 gene expression showed that there was significant difference in the expression of each key gene between two groups (Figure 7).

Prognosis verification based on external data sets

Since there was no PFS survival information in the external data sets (GEO and CGGA datasets), the effect of the risk model constructed in the previous step could not be verified. Thus, only the differences of 6 key genes and OS prognosis were enrolled for current verification analysis. A total of 693 samples

were downloaded from the CGGA database. After screening, there were 249 GBM and rGBM samples. The analysis results showed that KIFC1, ELFN2 and PLP2 were significantly related to OS (Supplementary Figure 8). Furthermore, COX survival analysis was performed based on 4 GEO validation datasets. The results showed that except for RP11-161H23.5, the COX survival analysis results of the other four genes were contribute to the survival (Supplementary Figure 9).

Discussion

Although GBM-infiltrated innate immune cells are proved to be associated with the progression of disease [9], the detail mechanism of these cells and associated genes are still unclear. In this study, the bioinformatics analysis showed that compared with the normal tissues, the immune score and cytolytic activity in GBM tissues were significantly higher than those in normal tissues. Meanwhile, the proportion of M2 Macrophage infiltration in GBM was significantly higher than that in normal tissues, and was related to the prognosis of GBM patients. A total of 4989 up- and 2349 down-regulated genes were revealed between tumor group and normal group. Among them, totally 37 DEGs (KIF18B) and 5 DElncRNAs (such as lncRNA PR11-161H23.5) were related to infiltration of M2 Macrophage, and forms a ceRNA network through DEMs (such as miR-6849-3p). Signal release was one of the outstanding functions assembled by these DEGs. Furthermore, a risk model for predicting the clinical endpoint of PFS in GBM was successfully constructed based on 5 mRNAs (such as FOX4 and ELFN2) and lncRNA PR11-161H23.5. Finally, the analysis of external data sets showed that all these 5 mRNAs were significantly related to the OS prognosis in GBM patients.

Tumor associated macrophages are vital for the tumor microenvironment [36]. M2 Macrophage, also named as macrophage anti-inflammatory phenotype, is known to secrete various cytokines and promote tumor-growth [37]. A previous study proves the association between infiltration levels of M2 Macrophage and cytokines such as IL-6 in GBM [38]. An investigation for the invasiveness of GBM stem cells in the absence and presence of macrophages M2 shows that M2 Macrophage may affect cancer stem cell phenotypes via cell-cell communication [39]. Waldemar Debinski et al. indicated that more specific gene markers for M2 Macrophage were needed for detail mechanism investigation of M2 Macrophage in GBM [40]. Kinesin Family Member (KIF) 18B (KIF18B) gene is a member belong to the Kinesin family [41]. Previous studies shows that KIF family members such as KIF3A and KIF18A take part in the progression of different kinds of tumors [42, 43]. Recently, a study has proved that KIF18B is associated with the development of tumor in human [44]. Moreover, the biological function of KIF family members are commonly regulated by miRNAs such as miR-127 [45]. Meanwhile, some immune-related lncRNAs such as RP11-161H23.5 contributes to the prognosis prediction of GBM via certain lncRNA-miRNA interactions [46]. Actually, the lncRNA-miRNA-mRNA regulation take part in the malignant progression of GBM disease identification and functional characterization [47]. A previous differential correlation analysis of GBM reveals that immune ceRNA interactions can be used to predict the patient survival [48]. In this study, the immune cell infiltration analysis showed that macrophages have the highest percentage of infiltration in GBM tissue compared to all immune-related cells, and was related to the prognosis of GBM patients. Moreover, the ceRNA network showed that RP11-161H23.5-miR6849-3p-KIF18B was one of the

outstanding lncRNA-miRNA-mRNA interaction relations constructed by key genes associated with M2 Macrophage infiltration. Thus, M2 Macrophage infiltration might contribute to the GBM progression via RP11-161H23.5-miR6849-3p-KIF18B ceRNA interaction.

Due to the extremely poor prognosis, there is a pressing need for an improved understanding of the molecular biomarkers for GBM [49]. Based on a literature-based meta-analysis, a previous study shows that the PFS can be used as a surrogate endpoint for overall survival in GBM [50]. Survival prediction in patients with GBM can be realized by certain human gene variation [51]. SRY-Box Transcription Factor 4 (SOX4), a member of SOX family, is a transcription factor required for neuron survival and neurite growth [52]. Li et al. indicated that SOX4 was overexpressed in diffusely infiltrating astrocytoma and conferred poor prognosis [53]. The differentially expressed of SOX4 contribute to the overall survival and poor prognosis of various cancer [54]. A previous study shows that SOX4 can block the cell growth and induce cell cycle arrest in GBM [55]. Moreover, ELFN2 is proved to be associated with the biological function of cerebral cortex neurons [56]. The evaluation of survival following melanoma brain metastases shows that ELFN2 differentially expressed between tumor samples and controls samples [57]. It has been proved that ELFN2 has important significance for evaluating the prognosis of patients with GBM [58]. In the current study, a risk model for predicting the clinical endpoint of PFS in GBM was successfully (AUC > 0.8) constructed based on M2 Macrophage related mRNAs such as FOX4 and ELFN2. Importantly, the external datasets verifications showed that these mRNAs were significantly associated with OS prognosis in GBM patients. Thus, we speculated that mRNAs such as FOX4 and ELFN2 might be used as novel prognostic key genes for patients with GBM.

Conclusions

M2 Macrophage infiltration might play contribute to the GBM progression via RP11-161H23.5-miR6849-3p-KIF18B ceRNA interaction. Moreover, mRNAs such as FOX4 and ELFN2 might be used as novel prognostic key genes for patients with GBM.

Abbreviations

GBM: Glioblastoma multiforme; DEGs: differentially expressed RNAs; DElncRNAs: differentially expressed lncRNAs; DEMs: differentially expressed mRNAs; ceRNA: competing endogenous RNAs; TPM :transcripts per million; OS: overall survival; PFS: progression free survival; AUC :area under curve ; ROC :receiveroperating characteristic; CGGA :Co-expressed Gene Groups Analysis

Declarations

Acknowledgements

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

Authors' contributions

WCL conceived and designed this study. HX carried out the plan and wrote this paper. XHD, CY, JJJ and ZYL gave advice and carried out the data analysis.

Funding

This work was supported by grants from Guidance Plan of Natural Science Foundation of Liaoning Province (No. 201602773; 2019-ZD-0340) and Guidance Project of Science and Technology Fund of Liaoning Provincial (No. 20180550093).

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 for publication

Not applicable.

Competing interests

The authors declare that there are no conflicts of interest.

Author details

¹ Department of Histology and Embryology, College of Basic Medicine, Shenyang Medical College, Shenyang, Liaoning, China

² Graduate Program in Bioinformatics and Computational Biology, University of Minnesota, Minneapolis, USA.

³ Department of Neurosurgery, General Hospital of Northern Theater Command, Shenyang, Liaoning, China

⁴ Department of Laboratory Animal Center, China Medical University, Shenyang, Liaoning, China

⁵ Department of Neurosurgery, First Affiliated Hospital of China Medical University, Shenyang, Liaoning, China

References

1. Ostrom QT, et al. CBTRUS Statistical Report: Primary Brain and Central Nervous System Tumors Diagnosed in the United States in 2006-2010. *Neuro-oncology*. 2013. <https://doi.org/10.1093/neuonc/not151>.
2. Young RM, et al. Current trends in the surgical management and treatment of adult glioblastoma. *Ann Transl Med*. 2015. <https://doi.org/10.3978/j.issn.2305-5839.2015.05.10>.
3. Nam JY and JF de Groot. Treatment of Glioblastoma. *Journal of Oncology Practice*. 2017. <https://doi.org/10.1200/JOP.2017.025536>.
4. Gallego O. Nonsurgical treatment of recurrent glioblastoma. *Current Oncology*. 2015. <https://doi.org/10.3747/co.22.2436>.
5. Jiri Polivka, JR. et al. Advances in Experimental Targeted Therapy and Immunotherapy for Patients with Glioblastoma Multiforme. *Anticancer Research*. 2017. <https://doi.org/10.21873/anticancer.11285>.
6. Salem S. Innovative novel immunotherapies for the treatment of glioblastoma multiforme. *Open Access Theses*. 2016. https://docs.lib.purdue.edu/open_access_theses/895.
7. Sokratous GS, Polyzoidis and K Ashkan. Immune infiltration of tumor microenvironment following immunotherapy for glioblastoma multiforme. *Human vaccines & immunotherapeutics*. 2017. <https://doi.org/10.1080/21645515.2017.1303582>.
8. Walker DG. T-cell apoptosis in human glioblastoma multiforme: implications for immunotherapy. *Journal of neuroimmunology*. 2006. <https://doi.org/10.1016/j.jneuroim.2006.03.006>.
9. Kmiecik J, et al. Elevated CD3+ and CD8+ tumor-infiltrating immune cells correlate with prolonged survival in glioblastoma patients despite integrated immunosuppressive mechanisms in the tumor microenvironment and at the systemic level. 2013. <https://doi.org/10.1016/j.jneuroim.2013.08.013>.
10. Ali HR. et al. Patterns of immune infiltration in breast cancer and their clinical implications: a gene-expression-based retrospective study. *PLoS medicine*. 2016. <https://doi.org/10.1371/journal.pmed.1002194>.
11. Zhao Z, et al. ADAMTSL4, a Secreted Glycoprotein, Is a Novel Immune-Related Biomarker for Primary Glioblastoma Multiforme. *Disease markers*. 2019. <https://doi.org/10.1155/2019/1802620>.
12. Zhong QY, et al. A gene expression-based study on immune cell subtypes and glioma prognosis. *BMC cancer*. 2019. <https://doi.org/10.1186/s12885-019-6324-7>.
13. Novi Quadrianto Wray L. Buntine. Linear regression. *Encyclopedia of Machine Learning*. 2016. 1 https://doi.org/10.1007/978-1-4899-7502-7_481-1.

14. Koop G. Bayesian methods for empirical macroeconomics. *Review of Economic Analysis*. 2017.
15. Smyth GK. Limma: linear models for microarray data in Bioinformatics and computational biology solutions using R and Bioconductor. Springer. 2005. <https://doi.org/1198/jasa.2007.s179>.
16. Wickham H and MH Wickham. The ggplot package. 2007.
17. Kolde R and MR Kolde. Package 'pheatmap'. R Package, 2015.
18. Tang Z, et al. GEPIA2: an enhanced web server for large-scale expression profiling and interactive analysis. *Nucleic acids research*. 2019. <https://doi.org/1093/nar/gkz430>.
19. Yoshihara K, et al. Inferring tumour purity and stromal and immune cell admixture from expression data. *Nature communications*. 2013. <https://doi:1038/ncomms3612>.
20. Kassambara A. ggpubr: "ggplot2" based publication ready plots. R package version 0.1, 2017.
21. Newman AM, et al. Robust enumeration of cell subsets from tissue expression profiles. *Nature methods*, 2015. <https://doi.org/1038/nmeth.3337>.
22. Clark K, et al. The Cancer Imaging Archive (TCIA): maintaining and operating a public information repository. *Journal of digital imaging*, 2013. <https://doi.org/1007/s10278-013-9622-7>.
23. Li T, et al. TIMER: A web server for comprehensive analysis of tumor-infiltrating immune cells. *Cancer Research*. 2017. <https://doi.org/1158/0008-5472.CAN-17-0307>.
24. Wei T, et al. Package 'corrplot'. *Statistician*. 2017.
25. Ashburner M, et al. Gene Ontology: tool for the unification of biology. *Nat Genet*. 2000. <https://doi:10.1038/75556>.
26. Kanehisa M and S Goto. KEGG: Kyoto Encyclopedia of Genes and Genomes. *Nucleic Acids Research*. 2000. <https://doi.org/1093/nar/28.1.27>.
27. Yu G, et al. clusterProfiler: an R package for comparing biological themes among gene clusters. *Omics-a Journal of Integrative Biology*. 2012. <https://doi.org/1089/omi.2011.0118>.
28. Dweep H, et al. miRWalk–database: prediction of possible miRNA binding sites by “walking” the genes of three genomes. *Journal of biomedical informatics*. 2011. <https://doi.org/1016/j.jbi.2011.05.002>.
29. Wang X. miRDB: A microRNA target prediction and functional annotation database with a wiki interface. *RNA*. 2008. <https://doi.org/1261/rna.965408>.
30. Grimson A, et al. MicroRNA Targeting Specificity in Mammals: Determinants beyond Seed Pairing. *Molecular Cell*. 2007. <https://doi.org/1016/j.molcel.2007.06.017>.
31. Chou C-H, et al. miRTarBase update 2018: a resource for experimentally validated microRNA-target interactions. *Nucleic acids research*. 2017. <https://doi.org/1093/nar/gkx1067>.
32. Shannon P, et al. Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome research*. 2003. <https://doi.org/1101/gr.1239303>.
33. Szklarczyk D, et al. The STRING database in 2011: functional interaction networks of proteins, globally integrated and scored. *Nucleic acids research*. 2010. <https://doi.org/1093/nar/gkq973>.

34. Bland JM and DG Altman. Survival probabilities (the Kaplan-Meier method). *Bmj*, 1998. <https://doi.org/1136/bmj.317.7172.1572>.
35. Aguirre-Gamboa R, et al. SurvExpress: an online biomarker validation tool and database for cancer gene expression data using survival analysis. *PloS one*. 2013. <https://doi.org/1371/journal.pone.0074250>.
36. Susen R.M, et al. Macrophage HIF regulates tumor-suppressive Spint1 in the tumor microenvironment. *Molecular Carcinogenesis*. 2019. <https://doi.org/1002/mc.23103>.
37. Liu R, et al. Negative Immune Regulator TIPE2 Promotes M2 Macrophage Differentiation through the Activation of PI3K-AKT Signaling Pathway. *Plos One*. 2017. <https://doi.org/1371/journal.pone.0170666>.
38. Hori T and T Sasayama. P08.37 Tumor associated M2 macrophage infiltration in glioblastoma. 2016. <https://doi.org/1093/neuonc/nou188.170>.
39. Nusblat LM, MJ Carroll and CM Roth. Crosstalk between M2 macrophages and glioma stem cells. *Cellular Oncology*. 2017. <https://doi.org/1007/s13402-017-0337-5>.
40. Debinski W, et al. Functional presence of M2 macrophage markers in GBM tumor cells. *Neuro-Oncology*. 2014. <https://doi.org/1093/neuonc/nou208.66>.
41. Yan H, C Zhu and L Zhang. Kinesin family member 18B: A contributor and facilitator in the proliferation and metastasis of cutaneous melanoma. *Journal of Biochemical and Molecular Toxicology*. 2019. <https://doi.org/10.1002/jbt.22409>.
42. Lan HM, et al. Disruption of KIF3A in patient-derived glioblastoma cells: Effects on ciliogenesis, hedgehog sensitivity, and tumorigenesis. *Oncotarget*. <https://doi.org/18632/oncotarget.6854>.
43. Shichijo S, et al. A unique gene having homology with the kinesin family member 18A encodes a tumour-associated antigen recognised by cytotoxic T lymphocytes from HLA-A2+ colon cancer patients. *European Journal of Cancer*. 2005. <https://doi.org/1016/j.ejca.2005.02.025>.
44. Zhong Y, et al. Clinical Significance And Integrative Analysis Of Kinesin Family Member 18B In Lung Adenocarcinoma. *OncoTargets and therapy*. 2019. <https://doi.org/10.2147/OTT.S227438>.
45. Aguado-Fraile E, et al. miR-127 protects proximal tubule cells against ischemia/reperfusion: identification of kinesin family member 3B as miR-127 target. *PloS one*. 2012. <https://doi.org/1371/journal.pone.0044305>.
46. Zhou M, et al. An immune-related six-lncRNA signature to improve prognosis prediction of glioblastoma multiforme. *Molecular neurobiology*. 2018. <https://doi.org/1007/s12035-017-0572-9>.
47. Ren Y, et al. Aberrant ceRNA-mediated regulation of KNG1 contributes to glioblastoma-induced angiogenesis. *Oncotarget*. 2016. <https://doi.org/18632/oncotarget.12659>.
48. Chiu Y-C, et al. Differential correlation analysis of glioblastoma reveals immune ceRNA interactions predictive of patient survival. *BMC bioinformatics*. 2017. <https://doi.org/1186/s12859-017-1557-4>.
49. Nusblat LM, MJ Carroll and CM Roth. Crosstalk between M2 macrophages and glioma stem cells. *Cellular Oncology*. 2017. <https://doi.org/1007/s13402-017-0337-5>.

50. Han K, et al. Progression-free survival as a surrogate endpoint for overall survival in glioblastoma: a literature-based meta-analysis from 91 trials. *Neuro-Oncology*. 2014. <https://doi.org/10.1093/neuonc/not236>.
51. Wang L, et al. Survival prediction in patients with glioblastoma multiforme by human telomerase genetic variation. *J Clin Oncol*. 2006. <https://doi.org/10.1200/jco.2005.04.0402>.
52. Jankowski MP, et al. SRY-box containing gene 11 (Sox11) transcription factor is required for neuron survival and neurite growth. *Neuroscience*. 2006. <https://doi.org/10.1016/j.neuroscience.2006.09.010>.
53. Li L, et al. SOX4 is overexpressed in diffusely infiltrating astrocytoma and confers poor prognosis. *Neuropathology*. 2015. <https://doi.org/10.1111/neup.12212>.
54. Wang W, et al. SOX4 is associated with poor prognosis in cholangiocarcinoma. *Biochemical and biophysical research communications*. 2014. <https://doi.org/10.1016/j.bbrc.2014.08.124>.
55. Zhang J, et al. SOX4 inhibits GBM cell growth and induces G0/G1 cell cycle arrest through Akt-p53 axis. *BMC neurology*. 2014. <https://doi.org/10.1186/s12883-014-0207-y>.
56. Dunn HA, et al. ELFN2 is a postsynaptic cell adhesion molecule with essential roles in controlling group III mGluRs in the brain and neuropsychiatric behavior. *Molecular psychiatry*. 2019. <https://doi.org/10.1038/s41380-019-0512-3>.
57. Kim E, et al. Bmet-04. Evaluation of survival following stereotactic radiosurgery (SRS) for melanoma brain metastases. Oxford university press US. <https://doi.org/10.1093/neuonc/nwz212.104>.
58. Liu C, et al. LINC00470 coordinates the epigenetic regulation of ELFN2 to distract GBM cell autophagy. *Molecular Therapy*. 2018. <https://doi.org/10.1016/j.ymthe.2018.06.019>.

Figures

Figure 1

The histogram and violin chart for immune cell infiltration. A, the histogram for 66 samples of immune cell infiltration abundance explored by CIBERSORT algorithm; the X-axis represented the different samples, while the Y-axis represented the relative percent. B, the histogram for 604 samples of immune cell infiltration abundance explored by TCGA-GBM dataset. The different color represented different immune cell type. C, Infiltration histograms of 6 immune cells in Glioblastoma and normal samples explored by TIMER algorithm; the Y-axis represented the relative percent; the different color represented different types of immune cells. D, Violin chart for infiltration abundance difference of 6 immune cells in Glioblastoma and normal samples explored by TIMER algorithm; the X-axis represented the different abundance of immune cells, while the Y-axis represented the different types of immune cells.

Figure 1

The histogram and violin chart for immune cell infiltration. A, the histogram for 66 samples of immune cell infiltration abundance explored by CIBERSORT algorithm; the X-axis represented the different samples, while the Y-axis represented the relative percent. B, the histogram for 604 samples of immune cell infiltration abundance explored by TCGA-GBM dataset. The different color represented different immune cell type. C, Infiltration histograms of 6 immune cells in Glioblastoma and normal samples explored by TIMER algorithm; the Y-axis represented the relative percent; the different color represented different types of immune cells. D, Violin chart for infiltration abundance difference of 6 immune cells in Glioblastoma and normal samples explored by TIMER algorithm; the X-axis represented the different abundance of immune cells, while the Y-axis represented the different types of immune cells.

Figure 2

The function and pathway enrichment analysis for the differentially expressed mRNAs. A, the functions assembled by differentially expressed mRNAs; X-axis represented the gene ratio; Y-axis represented the different items of functions. B, the pathways enriched by differentially expressed mRNAs; X-axis represented the gene ratio; Y-axis represented the different items of pathways. MF, molecular function; BP, biological process; CC, cellular component. The deeper the color, the smaller the P value; the larger the circle, the bigger the count value.

Figure 2

The function and pathway enrichment analysis for the differentially expressed mRNAs. A, the functions assembled by differentially expressed mRNAs; X-axis represented the gene ratio; Y-axis represented the different items of functions. B, the pathways enriched by differentially expressed mRNAs; X-axis represented the gene ratio; Y-axis represented the different items of pathways. MF, molecular function; BP, biological process; CC, cellular component. The deeper the color, the smaller the P value; the larger the circle, the bigger the count value.

Figure 3

The lncRNA-miRNA-mRNA (ceRNA) interaction network and protein-protein interaction network in current study. A, the ceRNA interaction network constructed by 37 differentially expressed miRNAs (DEMs), 25 differentially expressed mRNAs (DEGs) and 5 differentially expressed lncRNA (DElncRNA). Red circular nodes represented up-regulated immune-related DEGs; Blue circular nodes represented down-regulated immune-related DEGs; Red diamond nodes represented up-regulated immune-related DElncRNA; Blue diamond nodes represent down-regulated immune-related DElncRNA; The red box outline represented a positive correlation with immune cell infiltration; The blue box outline represented negative correlation with immune cell infiltration; Green triangle nodes represented DEMs that targeting immune-related genes; Gray lines represented PPI relationships or miRNA-mRNA interactions or miRNA-lncRNA interactions. Purple lines represented protein interactions. B, the PPI network constructed by mRNAs in ceRNA network; the red circle represented the up-regulated mRNAs. The blue circle represented the down-regulated mRNAs; the line between two nodes represented interaction ration.

Figure 3

The lncRNA-miRNA-mRNA (ceRNA) interaction network and protein-protein interaction network in current study. A, the ceRNA interaction network constructed by 37 differentially expressed miRNAs (DEMs), 25 differentially expressed mRNAs (DEGs) and 5 differentially expressed lncRNA (DElncRNA). Red circular nodes represented up-regulated immune-related DEGs; Blue circular nodes represented down-regulated immune-related DEGs; Red diamond nodes represented up-regulated immune-related DElncRNA; Blue diamond nodes represent down-regulated immune-related DElncRNA; The red box outline represented a positive correlation with immune cell infiltration; The blue box outline represented negative correlation with immune cell infiltration; Green triangle nodes represented DEMs that targeting immune-related genes; Gray lines represented PPI relationships or miRNA-mRNA interactions or miRNA-lncRNA interactions. Purple lines represented protein interactions. B, the PPI network constructed by mRNAs in ceRNA network; the red circle represented the up-regulated mRNAs. The blue circle represented the down-regulated mRNAs; the line between two nodes represented interaction ration.

Figure 4

The risk prediction model constructed by the expression of mRNAs and lncRNAs. A total of 5 mRNAs including SOX4, KIFC1, ELFN2, PLP2 and MEX3A, as well as 1 lncRNA RP11-161H23.5 were enrolled in current prediction. The X-axis represented the specificity, while the Y-axis presented the sensitivity.

Figure 4

The risk prediction model constructed by the expression of mRNAs and lncRNAs. A total of 5 mRNAs including SOX4, KIFC1, ELFN2, PLP2 and MEX3A, as well as 1 lncRNA RP11-161H23.5 were enrolled in current prediction. The X-axis represented the specificity, while the Y-axis presented the sensitivity.

Figure 5

The survival time analysis based on high and low risk group. A, Graph of 165 samples sorted by risk score; the X-axis represented the number of samples, while the Y-axis represented the risk score. B, Scatter plot of sample grouping and PFS survival; the X-axis represented the number of samples, while the Y-axis

represented the time of follow up. C, K-M curve of sample with high-risk group and low-risk group; the X-axis represented the survival time, while the Y-axis represented the time of survival rate. The red line represented the high-risk group, and the blue line represented the low-risk group.

Figure 5

The survival time analysis based on high and low risk group. A, Graph of 165 samples sorted by risk score; the X-axis represented the number of samples, while the Y-axis represented the risk score. B, Scatter plot of sample grouping and PFS survival; the X-axis represented the number of samples, while the Y-axis represented the time of follow up. C, K-M curve of sample with high-risk group and low-risk group; the X-axis represented the survival time, while the Y-axis represented the time of survival rate. The red line represented the high-risk group, and the blue line represented the low-risk group.

Figure 6

The clustering heat map of the expression of 6 key genes between high-risk group and low-risk group. A total of 5 mRNAs including SOX4, KIFC1, ELFN2, PLP2 and MEX3A, as well as 1 lncRNA RP11-161H23.5 were enrolled in current clustering.

Figure 6

The clustering heat map of the expression of 6 key genes between high-risk group and low-risk group. A total of 5 mRNAs including SOX4, KIFC1, ELFN2, PLP2 and MEX3A, as well as 1 lncRNA RP11-161H23.5 were enrolled in current clustering.

Figure 7

The box plot analysis for 6 gene expression between high-risk and low-risk groups. A total of 5 mRNAs including SOX4, KIFC1, ELFN2, PLP2 and MEX3A, as well as 1 lncRNA RP11-161H23.5 were enrolled in current clustering. The X-axis represented different groups, while the Y-axis represented the expression value.

Figure 7

The box plot analysis for 6 gene expression between high-risk and low-risk groups. A total of 5 mRNAs including SOX4, KIFC1, ELF2, PLP2 and MEX3A, as well as 1 lncRNA RP11-161H23.5 were enrolled in current clustering. The X-axis represented different groups, while the Y-axis represented the expression value.

Supplementary Files

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

- [SupplementaryFigure1.tif](#)
- [SupplementaryFigure1.tif](#)
- [SupplementaryFigure2.tif](#)
- [SupplementaryFigure2.tif](#)
- [SupplementaryFigure3.tif](#)
- [SupplementaryFigure3.tif](#)
- [SupplementaryFigure4.tif](#)

- [SupplementaryFigure4.tif](#)
- [SupplementaryFigure5.tif](#)
- [SupplementaryFigure5.tif](#)
- [SupplementaryFigure6.tif](#)
- [SupplementaryFigure6.tif](#)
- [SupplementaryFigure7.tif](#)
- [SupplementaryFigure7.tif](#)
- [SupplementaryFigure8.tif](#)
- [SupplementaryFigure8.tif](#)
- [SupplementaryFigure9.tif](#)
- [SupplementaryFigure9.tif](#)
- [SupplementaryTable1.docx](#)
- [SupplementaryTable1.docx](#)
- [SupplementaryTable2.docx](#)
- [SupplementaryTable2.docx](#)
- [SupplementaryTable3.xlsx](#)
- [SupplementaryTable3.xlsx](#)