Predictive value of focal adhesion-related genes for prognosis and response to immunotherapy in glioma patients

Qiankun Ji  
The Second Affiliated Hospital of Nanchang University

Zewei Tu  
The Second Affiliated Hospital of Nanchang University

Yuan Jiang  
The Second Affiliated Hospital of Nanchang University

Xinjun Liu  
Yingtan People's Hospital

Xingen Zhu  
(ndefy89006@ncu.edu.cn)  
The Second Affiliated Hospital of Nanchang University

Kai Huang  
The Second Affiliated Hospital of Nanchang University

Jingying Li  
The Second Affiliated Hospital of Nanchang University

Research Article

Keywords: Focal adhesion-related genes, Tumor microenvironment, Immune checkpoint, Prognosis, Immunotherapy

Posted Date: August 22nd, 2022

DOI: https://doi.org/10.21203/rs.3.rs-1966119/v1

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

Background

Focal adhesions (FAs) allow cells to contact the extracellular matrix, helping to maintain tension and enabling signal transmission in cell migration, differentiation, and apoptosis. FAs are also implicated in changes in the tumor microenvironment (TME) that lead to malignant progression and drug resistance. However, there are still few studies on the comprehensive analysis of focal adhesion-related genes (FARGs) in glioma.

Methods

Expression data and clinical information of glioma samples were downloaded from public databases. Two distinct subtypes were identified based on FARGs using an unsupervised consensus clustering algorithm. A scoring system with outstanding clinical application value was constructed using different algorithms. RT-qPCR, western blotting, and immunohistochemistry staining were used to detect the expression of COL1A2. Cell lines stably knockdown and overexpressing COL1A2 were constructed by plasmids transfection. Wound healing assay, clone formation, and transwell assay were used for mechanism exploration. A cell co-culture experiment was used to study the effect of Jurkat cells on the viability of GBM cell lines after knockdown and overexpression of COL1A2.

Results

The prognosis outcome and TME characteristics are significantly different between the two subtypes. The nine-genes scoring system can not only effectively predict the prognosis outcome of glioma patients, but also contribute to predicting the immunotherapy response of glioma patients. COL1A2 is up-regulated in glioblastoma (GBM) tissues and GBM cell lines and is closely related to the prognosis outcome of glioma patients. COL1A2 is involved in the regulation of the biological behavior in GBM and may be a key molecule in the changes in TME of GBM.

Conclusion

FARGs is involved in the tumorigenesis and progression of glioma. The novel molecular subtypes and scoring system identified based on FARGs have outstanding clinical application value. It is of great significance to explore the molecular mechanism of FARGs in glioma for finding new therapeutic targets.

1. Introduction

Glioma has the highest incidence rate (70%) among intracranial primary tumors and is characterized by high malignancy and rate of recurrence [1]. In addition to conventional treatment, immunotherapy is also
an option for glioma patients[2; 3; 4]. However, the treatment effect of these regimens is not optimistic, recurrence is almost inevitable [5], and the median overall survival (OS) is hardly more than 1.5 years [6; 7]. Besides intra-tumoral heterogeneity, genetic and epigenetic factors also influence patients’ response to treatment. Identifying and elucidating the intrinsic molecular mechanism of glioma provides a theoretical basis for developing more effective therapeutic schemes.

FAs allow cells to contact the extracellular matrix (ECM), helping to maintain cell tension and enabling signal transmission [8; 9] in cell migration, differentiation, and apoptosis. FA proteins are classified according to their functions [10] as integrins, FA kinases, paxillin, etc. Some of these have been testified linked to cancer progression [11; 12; 13]. However, the predictive value of FA proteins for prognosis and treatment response in cancer is not known.

TME is closely related to tumorigenesis and malignant progression [14]. The change in TME, which reduces the adhesion of tumor cells, has an important impact on the drug resistance of tumors [15] and can cause metastasis [16; 17] by promoting epithelial-mesenchymal transition [18; 19; 20; 21]. Among the immune cells of glioma, macrophages account for the highest proportion, about 30% [22], and are an important component of TME in glioma. Immune checkpoint blockades (ICBs) can prevent immune escape by modulating the function of T cells [23; 24]. At present, CTLA-4, PD-1, and PD-L1 are the main immunotherapeutic targets for patients with advanced tumors, although they do not always prolong the survival time of patients [23; 25]. The failure of tumor immunotherapy is thought to be related to an immunosuppressive TME and low tumor mutational burden (TMB) [26; 27]. Up to now, glioma patients have neither effective immunotherapeutic targets nor biomarkers that can effectively predict immunotherapeutic response.

In this paper, we not only identified two distinct molecular subtypes but also constructed a scoring system with outstanding clinical application value. In addition, we screened the gene COL1A2, which is up-regulated in GBM and closely related to the poor prognosis of glioma patients. In addition, we screened and studied the gene COL1A2 and found that it not only participates in the regulation of biological behavior of GBM but also may be a key molecule in the TME of glioma.

2. Materials And Methods

2.1 Data sources

The flow diagram of our study was shown in Figure1. RNA-seq data, somatic mutation data, and corresponding clinical annotation of glioma patients were downloaded from the TCGA database, CGGA database, GEO database, UCSC Xena portal, GlioVis website, and NIH website. Before comparing and analyzing the gene expression data from different platforms, we performed transcripts per kilobase million (TPM) values transformation and robust multichip averaging (RMA) values transformation on RNA-seq data and microarray data, respectively. Patients with incomplete clinical information (including OS, OS status, gender, age, and WHO grade) will not be included in the clinical correlation analysis.
Demographics and clinical information of 1813 glioma patients are shown in Table S1. Somatic mutation data were processed with the R package “maftools” [28]. The genomic identification algorithm of important targets in tumors is used to process CNV data [29]. We searched 199 FARGs from the Molecular Signatures Database and the details of these FARGs can be found in Table S2. In addition, the IMvigor210 cohort treated with the PD-L1 inhibitor was also included in this study to confirm our findings.

2.2 Acquisition of GBM tissue samples

The current research was authorized by the Ethics Committee of The Second Affiliated Hospital of Nanchang University (NCUSAH) and informed consent was obtained from the patients. The surgical specimens of GBM patients were removed to liquid nitrogen immediately after surgical excision.

2.3 Unsupervised consensus clustering based on FARGs

Different glioma cohorts have different sequencing methods and gene annotation information. To facilitate follow-up analysis, we took the intersection of 199 FARGs in three independent glioma cohorts and finally obtained 192 FARGs. A novel molecular classification was identified based on the 192 FARGs extracted from the TCGA cohort by unsupervised consensus clustering with 1000 iterations using the “ConsensusClusterPlus” package [30]. The optimal value of molecular classification should consider not only the rate of increase of the area under the cumulative distribution function (CDF) curves but also consider the correlation between different molecular classifications.

2.4 Gene set variation analysis (GSVA) between different subtypes

After downloading “c2.cp.kegg.v7.2.symbols” from the Molecular Signatures Database, GSVA between different subtypes was performed using the R package “GSVA” [31]. Differences in the relative activity of functional pathways between different subtypes were explored by using the R package “limma” [32].

2.5 Immune-related analysis between different subtypes

Different types of cell components in tumor tissues can be estimated by calculating TME-related scores using the R package “ESTIMATE” [33]. The content of different types of immune cells was reckoned using MCPCOUNTER, CIBERSORT, and QUANTISEQ algorithms [34; 35].

Enrichment scores of 29 immune signatures were determined by using a single sample gene set enrichment analysis (ssGSEA), which to some extent represents the immune activity within the tumor. [36]. The analyzed immune checkpoint proteins (ICPs) were selected from a previous study [37]. The above immune-related characteristics were analyzed in different subtypes.
2.6 Construction of a scoring system and verification of its prognostic value

We screened 175 FARGs with prognostic significance from 192 FARGs using uni-Cox regression analysis. 175 FARGs were further screened by lasso algorithm and multi-Cox regression analysis, and nine FARGs were finally obtained to construct a scoring system: risk score = sum (gene expression × coefficient). High and low-risk groups were classified according to the median value of the risk score. The prognostic value of the scoring system was evaluated by using Kaplan–Meier (K–M) survival curves (R package “survminer”), receiver operating characteristic (ROC) curves (R package “survivalROC”), uni- and multi-Cox regression analysis (R package “survival” and “forestplot”). The different distribution of different clinicopathological characteristics between high and low-risk groups was analyzed by R package “limma”.

2.7 Genomic variation analysis between high and low-risk groups

The mutation type and frequency of genes between high and low-risk groups were analyzed by the R packages “maftool” [28] and “GenVisR” [38]. The relationship between TMB, as a marker for the therapeutic efficacy of anti-PD-1 antibodies in other cancers [39; 40], and risk score was also analyzed by R package “ggpubr”. As genomic alterations can impact tumor immunity and immune infiltration patterns [41; 42], we compared amplifications and deletions between high and low-risk groups and visualized the results as circle graphs using the R package “RCircos” [43].

2.8 Immune-related analysis of risk score

At present, only a sub-fraction of patients achieved long-lasting clinical benefits from ICBs treatment. To effectively predict the response of tumor patients to ICBs, the TIDE algorithm was developed mainly to model two primary mechanisms of tumor immune evasion: inducing the dysfunction of cytotoxic T lymphocytes (CTLs) in tumors and preventing the infiltration of CTLs in tumor tissues [44]. T cell dysfunction was identified by measuring the interaction between each gene and the infiltration level of CTL to influence patient survival. TIDE algorithm explored the association between gene expression data and markers of T cell dysfunction. Tumor samples that were highly positively correlated with markers of T cell dysfunction were identified as non-responders and otherwise as responders. Finally, the TIDE algorithm and an unsupervised subclass mapping method [45] were used together to forecast the response of glioma patients to anti–PD-1 and anti–CTLA-4 immunotherapy.

2.9 Cell culture, transfection, and activation
Normal human astrocytes (NHA) and GBM cell lines (U251, LN229, U118, and U87) were obtained from the Shanghai Institute of Biosciences and Cell Resources Center. Jurkat cells were purchased from The Global Bioresource Center. Except that Jurkat cells were cultured with RPMI-1640 medium (Gibco, United States), other cell lines were cultured with MEM (Gibco, United States) or DMEM (Gibco, United States). 50 IU/ml penicillin, 50 μg/ml streptomycin, and fetal bovine serum (FBS, Gibco, United States) were added to the culture medium. The COL1A2 knockdown and overexpression plasmids were constructed by the Sheweisi Biotechnology Company (Tianjin, China). U6-MCS-SV40 polyA-CMV-EGFP-SV40-NeoR was used to construct COL1A2 overexpression plasmid, and its three target sequences are shCOL1A2-1, 5'-GGTGTAAGCGGTGGTGGTTAT-3', shCOL1A2-2, 5'-GCACTATGGATGCTATCAAAG-3' and shCOL1A2-3, 5'-GCAACAGCAGGTTCACCTTACA-3' respectively. CMV-MCS-EF1a-ZsGreen1-SV40-Neomycin was used to construct COL1A2 overexpression plasmid, and its upstream and downstream primers are “AGCTGGCTAGGGTTCTCGAGGCCACCATGCTCAGCTTTGTGGATACGC" and “GTCTTTTTATTTGGCCGGTGTTTATTTTGAAACAGACTGGGCCAAATG" respectively. Lipofectamine 3000 (Thermo Fisher, L3000075, United States) was used to transfect plasmids when the cell adhesion density reached about 80%. Two methods were used to activate Jurkat cells. Method 1: Jurkat cells (1 × 10^6/ml) were stimulated by adding 20 ng/ml phorbol 12-myristate 13-acetate (PMA, Acme, P33390-1mg) and 500 ng/ml ionomycin (70-CS0002, MultiSciences, China). Method 2: Jurkat cells were stimulated by adding 5μg/ml of anti-CD3 antibody (17617-1-AP, Proteintech, China), which was coated in a 6-well plate with the 5μg/ml of anti-CD28 antibody (65099-1-Ig, Proteintech). The expression of CD8 protein was used to detect whether Jurkat cells were activated, and the expression of CD69 protein was used to detect whether Jurkat cells proliferated. Jurkat and GBM cells were co-cultured (Jurkat: GBM cells = 1:1) in 96-well plates with a total volume of 200 μl.

2.10 RT-qPCR and western blotting

The primer sequences of COL1A2 and GAPDH (internal reference gene) purchased from Ribobio (Guangzhou, China) are as follows: COL1A2 F primer: 5'-GGCCCTCAAGGGTTCTCAAGGG-3', R primer: 5'-CACCTCAGGCTCTCCAGGG-3'; GAPDH F primer: 5'-GTGAGATGGTGTTCTCTGAG-3', R primer: 5'-CAACCATGATATGCGCTTCTGAG-3'. The primary antibodies include anti-COL1A2 (Servicebio, GB13022-2, 1:3000 dilution), anti-CD8 (Proteintech, 66868-1-lg, 1:4000 dilution), anti-CD69 (Proteintech, 10803-1-AP, 1:1500 dilution) and anti-tubulin (Proteintech, 11224-1-AP, 1:8000 dilution). The secondary antibodies included goat anti-rabbit IgG (Abcam, ab6721 1:5000 dilution) and goat anti-mouse IgG (Abcam, ab6789 1:10000 dilution). The extraction and detection methods of protein and total RNA in different cell lines were completely consistent with those described in our previous study [46]. Here, we will describe the extraction steps of tissue protein in detail. First, the tissues were weighed and sheared into 2ml EP tubes. RIPA and PMSF were mixed in a ratio of 100:1 to prepare tissue lysates. 1ml of lysate was added to every 50mg of sheared tissue and homogenized with a homogenizer, and then lysed on ice for 30 minutes. Finally, the mixture of lysate and tissue was centrifuged with a low-temperature high-speed centrifuge at 4 °C and 12000 rpm for 10 min. The subsequent steps are the same as the extraction of cellular proteins. Both RT-qPCR and western blotting assays were repeated three times.
2.11 Immunohistochemical and immunofluorescence staining

First, eight pairs of GBM and adjacent tissues were fixed with 10% formalin for seven days, and then these tissues were embedded with paraffin and sectioned. After dewaxing and dehydration, the tissue sections were treated with 3% hydrogen peroxide for 10 minutes. Subsequently, the tissue sections were blocked with 5% BSA and then incubated overnight with a primary antibody against COL1A2 (Servicebio, GB13022-2, 1:1000 dilution) at 4°C. Then the tissue sections were treated with the corresponding secondary antibodies at room temperature for one hour. Finally, DAB staining, target molecule detection, and hematoxylin re-staining were carried out in turn. For immunofluorescence staining, tissue sections were immune-stained overnight with primary antibodies against COL1A2 (Servicebio, GB13022-2, 1:1000 dilution) and CD8 (Proteintech, 1:400) at 4°C, and then incubated with fluorochrome-conjugated antibodies. DAPI was added as a nuclear counterstain.

2.12 Wound healing assay

When the adhesion concentration of U251 and U87 cells in the six-well plates reached 85-90%, the tip of a 200μl sterile spear cuts through the cell layer at the bottom of the plate to form an artificial wound. After washing the non-adherent cells with phosphate-buffered saline (PBS), the adherent cells were further cultured in a serum-free medium. At 0 and 24 hours, wound closure was photographed using an inverted Leica microscope.

2.13 Transwell assay

Transwell chambers (Corning, USA) were utilized to assay the invasion and migration of GBM cell lines. First, the Matrigel (Corning, 356234, USA) with a concentration of 500 μg/ml was used to cover the upper chambers. 200μl serum-free medium and 600μl medium containing 10% FBS were added to the upper chamber and the lower chamber, respectively, and 8×10⁴ transfected cells were seeded in the upper chamber. After 24 hours of incubation at 37°C and 5% CO₂ for 24 hours, the non-invasive cells on the upper chambers were removed with a microcell scraper. The following operations were carried out at room temperature. The cells at the bottom of the upper chamber were fixed with 4% ice-precooled formaldehyde (Solarbio, P1110, China) for 30 minutes and then washed twice with PBS, and then the chambers were placed in 0.1% crystal violet (Solarbio, G1075, China) for 20 minutes. Finally, the chambers were washed with PBS, dried, and photographed with a microscope (Leica Microsystems, D-35578). In addition, we also performed the migration assay in the same way as the invasion assay, except that the upper chamber was not covered with Matrigel. Both invasion and migration assays were repeated three times.
2.14 Tumor cell viability assay

In the light of the instructions offered by the manufacturer, the lentivirus expressing luciferase designed by Sheweisi Biotechnology Company (Tianjin, China) was transfected into U251 and U87 cells and screened with neomycin. Then COL1A2 knockdown and overexpression plasmids were transfected into cells stably expressing luciferase, respectively. When these cells grew to about 90%, they have inoculated with Jurkat cells in a ratio of 1:1 to 96 well plates with a white and clear bottom (total 1.0×10^4 cells/ml). Co-cultured cells were taken at different time points (8, 16, 24, 36, 48 h), the supernatant was discarded, washed with PBS, cell lysate was added, and then D-luciferin potassium salt (D1009, UEIlandy, China) was added. After the above operations are completed, the fluorescence intensity is detected using a multifunctional microplate reader. The experiment was repeated three times.

2.15 Statistical analyses

Spearman and Pearson analysis methods were used to determine correlations between different grouping variables. The unpaired Student’s t-test and Mann–Whitney U-test were used to analyze normally and non-normally distributed data, respectively. The Wilcoxon rank-sum test was used to compare two groups or categories; for more than two groups or categories, the Kruskal–Wallis test was used. Survival was assessed by K–M analysis with the log-rank test. Cox regression analyses were carried out to evaluate the prognostic value and stability of the risk prediction model. ROC curves were used to assess the prognostic value of 1-, 3-, and 5-year OS. TIIC infiltration level was analyzed using three algorithms (CIBERSORT, QUANTISEQ, and MPACOUNTER). All statistical tests were two-sided and P<0.05 was considered statistically significant.

3. Results

3.1 Molecular classification based on FARGs in TCGA cohort

First, we show the workflow of this study in Fig. 1. According to the flow chart, 192 FARGs were obtained from the intersection of TCGA, CGGA, and GSE16011 cohorts (Fig. 2A). A novel molecular classification was identified by using an unsupervised clustering algorithm based on the expression profiles of the 192 FARGs extracted from the TCGA cohort. According to the area under the CDF curves, probably approximately correct algorithm and correlations between clusters, we determined the optimal number of clusters is two (k = 2) (Fig. 2B–E). Principal components analysis (PCA) can effectively distinguish glioma patients with the FARG1 subtype and FARG2 subtype (Fig. 2F).

3.2 Basic analysis between different subtypes in the TCGA cohort
The OS and progression-free survival (PFS) of glioma patients, low-grade glioma (LGG) patients, and GBM patients with FARG1 subtype were significantly longer than those with FARG2 subtype (Fig. 2G, H). Except for gender, there are dramatic differences in the distribution of the other four clinicopathological characteristics between FARG1 and FARG1 subtypes. In addition, most FARGs are differentially expressed between the two subtypes (Fig. 2I). We simultaneously performed GSVA in TCGA, CGGA, and GSE16011 cohorts and identified 167 functional pathways with different relative activity between the two subtypes (Fig. 2J and Table S3), of which 65 functional pathways were more active in FARG1 subtype, such as inositol phosphate metabolism, glycerol lipid metabolism and lysine degradation, and the other 102 functional pathways were more active in FARG2 subtypes, such as cell lung cancer, pancreatic cancer, and primary immunodeficiency.

### 3.3 Immune-related analysis between different subtypes in TCGA cohort

Given our above finding, we further investigated whether different subtypes have different immunological characteristics. The results revealed that there were remarkable differences in most immune signatures (except for dendritic cells and mast cells) and all four TME-related scores among the two subtypes (Fig. 3A, C–F). Compared with the FARG1 subtype, the glioma tissue of the FARG2 subtype contains more immune cells and stromal cells. We also estimated the content of immune cells using CIBERSORT, MCPCOUNTER, and QUANTISEQ algorithms, and found that the content of almost all immune cells was different between the two subtypes (Fig. 3B and Table S4). Moreover, we also found that except for TNFSF9, other immune checkpoints (ICPs) were differentially expressed between the two subtypes of FARG1 and FARG2 (Fig. 3G). Considering the effect of gene mutations on tumorigenesis and progression, we found that the TBM was dramatically different between the two subtypes (Fig. 3H).

### 3.4 Construction of a nine-gene scoring system

To establish a scoring system, 175 FARGs were screened from 192 FARGs using uni-Cox regression analysis (Table S5). These genes were further screened using the LASSO algorithm, and then 17 FARGs were obtained (Fig. 4A, B). After that, multi-Cox regression analysis in both directions was performed and yielded nine FARGs (CDC42, COL1A2, ELK1, ERBB2, FLNC, PDGFA, PTEN, SPP1, and THBS4) (Fig. 4C). The nine genes were used to construct a scoring system according to the following equation: risk score = 0.298 × (CDC42 expression) + 0.103 × (COL1A2 expression) + 0.408 × (ELK1 expression) + 0.237 × (ERBB2 expression) + 0.111 × (FLNC expression) + 0.179 × (PDGFA expression) − 0.577 × (PTEN expression) + 0.259 × (SPP1 expression) + 0.138 × (THBS4 expression).

### 3.5 Prognostic analysis of scoring system

Since there are significant differences in the prognosis outcomes between GBM and LGG patients, IDH mutant and wild-type glioma patients, we should not only explore the prognostic value of the scoring system in pan-gliomas but also explore its prognostic value among different grades and different DH
mutation types. If the risk score of glioma patients was higher than the median value, they were classified as a high-risk group, otherwise, they were classified as a low-risk group. K–M survival curves showed that in the three cohorts, the OS of patients with pan-gliomas (Fig. 4D and Fig. S1A, B), LGG (Fig. S2A, E, I), GBM (Fig. S2B, F, J), IDH mutation (Fig. S3A and S4A, G) and IDH wildtype (Fig. S3B and S4B, H) in the low-risk group tended to be longer than that of patients in the high-risk group. Moreover, the risk score can effectively predict the 1, 3, and 5-year OS rate of patients with pan-gliomas (Fig. 4F and Fig. S1C, D), LGG (Fig. S2C, G, K), GBM (Fig. S2D, H, L), IDH mutation (Fig. S3C and S4C, I) and IDH wildtype (Fig. S3D and S4D, J) in the three cohorts (Table S6). In addition, in the TCGA cohort, patients with pan-gliomas, IDH mutant and IDH wildtype in the low-risk group also tended to have longer PFS than patients in the high-risk group (Fig. 4E and S3G, H), and the risk score can also effectively predict the 1, 3, 5-year PFS rate of patients with pan-gliomas, IDH mutant and IDH wildtype (Fig. 4G, and S3I, J, and Table S6). Next, we explored the differential distribution of five clinicopathological characteristics and the differential expression of nine FARGs between the two risk groups in the three cohorts. The results showed that except for gender, there were significant differences in the distribution of the other four clinicopathological characteristics and the expression of nine FARGs between different risk groups (Fig. 4H and S1E, G). Finally, we explored whether the scoring system can be used as an independent prognostic factor for patients with pan-gliomas, LGG, GBM, IDH mutant, or wildtype. The results showed that the independent prognostic value of risk score was superior to that of age, gender, grade, IDH status, and 1p19q status in TCGA, CGGA, and GSE16011 cohorts (Fig. 4I, J, and S1F, H, and S3E, F, K, L, and S4 E, F, K, L).

3.6 Genomic variation analysis between high and low-risk groups

Because genomic variation can impact tumor immunity and immune cell infiltration patterns [47; 48], we explored the association between genomic variation and risk score in this study. The results revealed that the mutation frequency of the top-16 genes with the highest mutation frequency differed significantly between the two risk groups (Fig. 5A, B). In the high and low-risk groups, the genes with the highest mutation frequency were TP53 (37%) and IDH1(92%), respectively. There are also obvious differences in the distribution of mutation and wildtype of six well-known genes (TP53, PTEN, IDH1, EGFR, ATRX, TTN) between high and low-risk groups (Fig. 5C). Besides, we found that glioma patients with low-risk scores inclined to have a lower TMB (Fig. 5D). Moreover, patients in the low-risk group had a lower frequency of CNV, either amplification or deletion (Fig. 5E, F).

3.7 Immune-related analysis between high and low-risk groups

By performing GSVA in TCGA, CGGA, and GSE16011 cohorts, we found that the relative activities of 156 functional pathways were different between the two risk groups: the relative activity of 53 functional pathways increased in the low-risk group, mainly focusing on metabolic-related pathways, such as “butanoate metabolism”, “propanoate metabolism” and “glyoxylate and dicarboxylate metabolism”; the
relative activity of other 103 functional pathways increased in the high-risk group, including immune-related pathways (Fig. 5G and Table S7). Next, we began an in-depth analysis of the relationship between immunologic characteristics and risk scores. First of all, we found that the TME-related scores calculated by the ESTIMATE algorithm were not only highly correlated with the risk score (Fig. S5C-E) but also significantly different between the high and low-risk groups in TCGA, CGGA, and GSE16011 cohorts (Fig. 6A and Fig. S5A, B). Then, we also found that the enrichment scores of most immune signatures calculated by ssGSEA were different between the two risk groups in the three cohorts (Fig. 6A and Fig. S5A, B). Considering that immune cells play a key role in tumorigenesis, tumor progression, and tumor immunity, we further explored the association between the scoring system and immune cells and immunotherapy. The content of most immune cells evaluated by the CIBERSORT algorithm was not only correlated with the risk score (Fig. S6A-C) but also significantly different among different risk groups (Fig. 6B and S6D, E). The results of the analysis between ICPs and risk scores (Fig. 6C and S7A-E) prompted us to continue to explore whether the scoring system can effectively predict the immunotherapy response of glioma patients. There was a significant difference in the risk score between the responder and non-responder groups predicted by the TIDE algorithm (Fig. 6D and S8A, D). The proportion of responders in the high-risk group is much higher than those in the low-risk group, which indicates that glioma patients in the high-risk group are more likely to benefit from immunotherapy (Fig. 6E and S8B, E). The subclass mapping analysis indicated that glioma patients in the high-risk group were more sensitive to anti-PD-1 therapy (Bonferroni P = 0.016, 0.032, 0.039, and in GSE16011, TCGA, and CGGA cohorts, respectively) (Figs. 6F and Fig. S8C, F). If the Bonferroni corrected P > 0.05, the P-value will not be marked in the subgroup map, but from the color comparison of the corresponding patches in the subgroup map, it can be seen that the P-value of anti-CTLA4 therapy in the high-risk group is lower than that in the low-risk group, indicating that glioma patients in the high-risk group are more likely to respond to anti-CTLA4 therapy (Figs. 6F and Fig. S8C, F).

3.8 Validation of clinical application value of scoring system in IMvigor210 (mUC) cohort

To detect the stability and effectiveness of the scoring system, we selected the IMvigor210 (mUC) cohort to further validate the clinical application value of the scoring system. K-M survival curves indicated that the OS of patients with a low-risk score was longer than those with a high-risk score (Fig. 6H). Moreover, the risk score can effectively predict the 1, 3, and 5-year OS rates of patients in the IMvigor210 (mUC) cohort (Fig. 6G). Then, the differential expression of PDL1 (CD274) between high and low-risk groups was also explored. The results showed that the expression level of PDL1 in the high-risk group was higher than that in the low-risk group (Fig. 6I). Furthermore, the proportion of CR/PR and PD/SD in the high-risk group is higher than that in the low-risk group (Fig. 6J, K).

3.9 COL1A2 is up-regulated in GBM

First, compared with the other eight FARGs, the mRNA expression of COL1A2 in GBM tissues was significantly higher than that in normal brain tissues (NBT), which was verified on the GEPIA website and
three independent GBM cohorts (Fig. 7A, B). Then, immunohistochemical staining of eight pairs of GBM tissues and adjacent tissues showed that COL1A2 was significantly up-regulated in GBM tissues (Fig. 7C, D). In addition, at the tissue protein level, the expression COL1A2 in GBM tissues was also significantly higher than that in corresponding adjacent tissues (Fig. 7E, F). Finally, the K-M survival curves of COL1A2 in four independent GBM cohorts (Rembrandt, GSE16011, TCGA, and CGGA cohorts) showed that the prognosis outcomes of GBM patients in the high expression group were significantly shorter than that of GBM patients in the low expression group (Fig. 7G).

### 3.10 COL1A2 promotes the proliferation, migration, and invasion of GBM cells in vitro

By comparing the differential expression of COL1A2 between different cell lines at the transcription level and protein level, we found that the expression level of COL1A2 was the highest in U87 cells and the lowest in U251 cells (Fig. 8A, B). Then, we constructed knockdown and overexpression plasmids of COL1A2 to further explore the effect of COL1A2 on the biological behavior of GBM in vitro. The result of RT-qPCR and western blotting showed that the constructed knockdown and overexpression plasmids could significantly reduce or increase the expression of COL1A2 (Fig. 8C-F). The colony formation assay showed that COL1A2 significantly enhanced cell proliferation in GBM (Fig. 8G). In addition, COL1A2 promoted cell migration and invasion in GBM examined by wound healing and transwell assay (Fig. 8H, I). These results suggest that COL1A2 plays an important role in the progression of GBM.

### 3.11 Detection of the viability of GBM cells in vitro

In eight cases of GBM tissues, we found that the tissues with high expression of COL1A2 were accompanied by high expression of CD8. On the contrary, the tissues with low expression of COL1A2 were accompanied by low expression of CD8 (Fig. 9A), which to some extent suggested that the expression of COL1A2 might induce the infiltration of CD8 T cells or the infiltration of CD8 T cells might induce the expression of COL1A2. Then, we studied whether COL1A2 played a role in the process of Jurkat cells acting on GBM cells in vitro. Both methods, PMA, and ionomycin or anti-CD3 and anti-CD28 antibodies, can effectively activate Jurkat cells into CD8+ Jurkat cells (Fig. 9B). CD8 T cells transiently overexpress CD69 in the early stage of activation, which is considered to be a costimulatory signal of T cell proliferation. However, we did not detect the expression of CD69 protein in Jurkat cells stimulated by anti-CD3 and anti-CD28 antibodies (Fig. 9C), so we utilized PMA and ionomycin to stimulate Jurkat cells. The results of the cell viability assay showed that activated Jurkat cells could effectively inhibit the viability of U87 (Fig. 9D) and U251 (Fig. 9E) cells. With the increase of COL1A2 expression, the lethality of activated Jurkat cells to GBM cells decreases, which may be related to COL1A2 promoting the proliferation of GBM cells.

### 4. Discussion
The prognosis of glioma patients remains poor. As such, it is critical to identify new prognostic biomarkers and develop a more effective treatment. Although there have been multiple studies on immunotherapy, especially by ICB in GBM patients, the results of phase III clinical trials have not been satisfactory compared to other tumors [49; 50]. Neither anti-CTLA-4 antibody alone nor in combination with anti-PD-1 antibody yielded long-term survival benefits for glioma patients [51]. Numerous factors can affect the response of glioma to immunotherapy including factors in the TME and TMB. The latter is a biomarker for evaluating the therapeutic effect of anti-PD-1 antibody [39]; and converting the immunosuppressive TME to one that is immunostimulatory is an effective treatment strategy [52; 53]. In this study, we developed a novel molecular classification for glioma patients based on FARGs. Not only the OS and PFS outcomes of glioma patients are significantly different between different subtypes, but also the characteristics of TME are significantly different between different subtypes. In addition, a scoring system was constructed based on nine FARGs (SPP1, THBS4, ERBB2, ELK1, COL1A2, PTEN, CDC42, FLNC, and PDGFA). The expression levels of these FARGs differed between different WHO grades. Except for PTEN, all were identified as risk genes for glioma. PTEN regulates signaling pathways related to cell growth and survival [54; 55] and cell metabolism [56]. COL1A2 participates in collagen synthesis [57] but has been implicated in the immune response [58]. ELK1 is a transcription factor that activates target genes via some protein kinase/regulatory kinase pathways [59; 60]. ERBB2 is a marker gene in breast cancer. SPP1 is known to be overexpressed in many malignant tumors including glioma [61; 62; 63; 64; 65] and regulates cell growth, proliferation, apoptosis, and migration [66]. THBS4 is a member of the thrombospondin protein family and plays important roles in wound healing and tissue repair [67; 68; 69; 70], intracellular migration, adhesion, and proliferation [71; 72; 73]. FLNC is an action-binding filamin protein that regulates actin reorganization-dependent processes such as differentiation, migration, and proliferation of cells [74]. As a Rho family GTPase, CDC42 plays a key role in the activation of signaling cascades that regulate cell adhesion, cytoskeletal composition, proliferation, and migration and is therefore important for the malignant transformation of tumors [75].

The clinical application value of the scoring system has been verified in TCGA, CGGA, and GSE16011 cohorts. The scoring system can not only effectively predict the prognosis outcome of glioma patients, but also can be used as an independent prognostic factor for glioma patients. TMB can predict the immunotherapeutic response of some types of tumors and has a close relationship with the scoring system. There are also significant differences in the distribution of mutation types of some classic genes between high and low-risk groups. Moreover, the TME between high and low-risk groups is also significantly different, which will directly affect the immunotherapeutic response of glioma patients. In addition to verifying the clinical application value of the scoring system in different glioma cohorts, we also found that the scoring system has an important clinical application value in the imvigor210 (MUC) cohort.

To further study the role of FARGs in the progression of GBM, COL1A2 was screened and further studied. It was found that COL1A2 was not only up-regulated in GBM tissues, but also could promote the progression of GBM, and the specific mechanism needed to be further studied. It has been suggested that the poor response to immunotherapy in glioma patients is attributable to immunosuppression in the brain
[27], however, the specific mechanism is still unclear at present. COL1A2 was still taken as the research object. It was found that it may promote the infiltration of CD8 T cells, but at the same time may inhibit the lethality of CD8 T cells on tumor cells. Although FARGs have been extensively studied from three aspects: bioinformatics, immunohistochemical staining, and in vitro experiments, this study has certain limitations due to the lack of in-depth mechanism research and in vitro experimental verification.

In summary, we identified two distinct subtypes based on FARGs. More importantly, we established a scoring system with great clinical application value, which can not only effectively predict the prognosis outcome of glioma patients, but also predict the immunotherapy response for glioma patients. In addition, we screened COL1A2 and verified its involvement in the progression of GBM cells in vitro. It is expected that this study can contribute to the diagnosis and treatment of glioma patients.

Declarations

Ethical Approval

This study was carried out in accordance with the principles of the Declaration of Helsinki. Approval was granted by the Ethics Committee of The Second Affiliated Hospital of Nanchang University (NCUSAH). Informed consent was obtained from the patients included in the study. Written informed consent for publication was also obtained from the patients enrolled in the study.

Competing interests

The authors state that the study was conducted without any direct or indirect conflict of interest.

Authors’ contributions

Xingen Zhu, Kai Huang, and Jingying Li designed this study and directed the writing of the article. Qiankun Ji did all the experiments of this study and participated in writing the article. Zewei Tu and Yuan Jiang performed the data analysis and the figures were plotted. Xinjun Liu performed the data acquisition and critical reading of the manuscript. All authors have carefully read and consented to the final manuscript.

Funding

This research was funded by the National Natural Science Foundation (grant nos. 82002660 and 81960456), Key Research and Development projects in Jiangxi, 20212BBG73021, and Jiangxi Training Program for academic and technical leaders of major disciplines - Young talents program, 20212BCJ23023.
Availability of data and materials


References

24. Neoadjuvant PD-1 Blockade in Resectable Lung Cancer; Nivolumab and Ipilimumab in Advanced Melanoma; Overall Survival with Combined Nivolumab and Ipilimumab in Advanced Melanoma; Prolonged Survival in Stage III Melanoma with Ipilimumab Adjuvant Therapy; Combined Nivolumab and Ipilimumab or Monotherapy in Untreated Melanoma; Combined Nivolumab and Ipilimumab or Monotherapy in Untreated Melanoma; Nivolumab and Ipilimumab versus Ipilimumab in Untreated Melanoma; Rapid Eradication of a Bulky Melanoma Mass with One Dose of Immunotherapy; Genetic Basis for Clinical Response to CTLA-4 Blockade; Genetic Basis for Clinical Response to CTLA-4 Blockade in Melanoma; Nivolumab plus Ipilimumab in Advanced Melanoma; Safety and Tumor Responses with Lambrolizumab (Anti-PD-1) in Melanoma; Hepatotoxicity with Combination of Vemurafenib and Ipilimumab. N Engl J Med 379 (2018) 2185.


**Figures**

**Figure 1**

Flow diagram of this study.
**Figure 2**

Novel molecular classification based on FARGs in TCGA cohort. (A) The intersection of 199 FARGs between TCGA, CGGA, and GSE16011 cohorts. (B-D) Consensus clustering matrix based on 192 FARGs for k=2, k=3, and k=4. (E) The optimal number of consensus clustering matrices is determined by the cumulative distribution function of the unsupervised consensus clustering algorithm. (F) Obvious differences in the transcriptomes between FARG1 and FARG2 subtypes were analyzed by principal component analysis (PCA). (G, H) K-M survival curves of glioma patients. (I) The distributions of five clinicopathological characteristics and 199 FARGs between FARG1 and FARG2 subtypes. (J) GSVA between FARG1 and FARG2 subtypes. Red and blue represent the relative activation and inhibition of the pathways, respectively. *P<0.05, **P<0.01, ***P<0.001.

**Figure 3**

Immune-related analysis between FARG1 and FARG2 subtypes in the TCGA cohort. (A) Difference analysis of 29 immune signatures and four types of TME-related scores between FARG1 and FARG2 subtypes. (B) The difference in the content of immune cells was calculated by CIBERSORT, MCPCOUNTER, and QUANTISEQ algorithms between the two subtypes. (C–F) Difference analysis of TME-related scores between the two subtypes. (G) Difference analysis of ICPs between two subtypes. (H) Difference analysis of TMB between the two subtypes. *P<0.05, **P<0.01, ***P<0.001, ****p < 0.0001.

**Figure 4**

Construction of scoring system and analysis of its prognostic value in the TCGA cohort. (A) LASSO coefficient profiles of 192 FARGs. (B) Selection of the penalty parameter (λ) in the LASSO model via 1,000 cross-validations. The vertical dotted line passing through the red dot is drawn at the optimal value. (C) Nine FARGs screened by multi-Cox regression analysis were used to construct the scoring system. (D, E) The K-M survival curves showed that the OS and PFS of glioma patients in the low-risk group were longer than those in the high-risk group. (F, G) ROC curves show that risk score can effectively predict the 1,3,5-year OS and PFS for glioma patients. (H) The distributions of five clinicopathological characteristics and nine FARGs between high and low-risk groups. (I, J) Uni- and multi-Cox regression analysis of risk score and five clinicopathological characteristics. *P<0.05, **P<0.01, ***P<0.001.

**Figure 5**

Comparisons of genomic variations and functional annotations between high and low-risk groups in the TCGA cohort. (A, B) Waterfall plots of the top-16 mutated genes. The genetic alteration types are listed at
the bottom of the waterfall plot. The mutation frequencies of genes are listed on the right side of the waterfall Plot. (C) Differential distribution of mutation and wildtype of six well-known genes between the two risk groups. (D) Comparison of TMB between high and low-risk groups. (E) Circular diagram of chromosome amplification and deletion between high and low-risk groups. (F) CNV frequencies (amplification or deletion) were lower in the low-risk group than that in the high-risk score group. (G) GSVA between FARG1 and FARG2 subtypes. Red and blue represent the relative activation and inhibition of the pathways, respectively.

Figure 6

Immune-related analysis in the TCGA and IMvigor210 cohorts. (A) Difference analysis of immune signatures and four types of TME-related scores between high and low-risk groups in the TCGA cohort. (B) The difference in infiltration level of immune cells between high and low-risk groups. (C) Differential expression of six well-known ICPs between high and low-risk groups. (D) The difference in risk score between high and low-risk groups was stratified according to response to immunotherapy. (E) The difference in the distribution of responders and non-responders between the two risk groups. (F) The subgroup map predicted the response to ICB therapy between the two risk groups. (G) ROC curves verified the accuracy of the risk score in predicting the OS for patients in the IMvigor210 cohort. (H) The K-M survival curves of OS in patients between the high and low-risk groups. (I) Differential expression of CD274(PDL1) between high and low-risk groups. (J, K) Differences in the distribution of patients with and without response to ICI immunotherapy between high and low-risk groups. CR, complete response; PR, partial response; SD, stable disease; PD, progressive disease. CR/PR was identified as responders, and SD/PD was identified as non-responders. *P<0.05, **P<0.01, ***P<0.001, ****P<0.0001.

Figure 7

Basic analysis of COL1A2. (A) Summary of the differential expression analysis of nine FARGs in three independent GBM cohorts and the GEPIA website. Red represents up-regulated genes in GBM tissue, while blue represents a down-regulated gene in GBM tissue. Grey represents no statistical difference. (B) Differential expression analysis of COL1A2 between GBM tissues and NBTs in TCGA (HG-UG133A), Kamoun and Rembrandt cohorts, and GEPIA website. (C) Immunohistochemical analysis of COL1A2 protein in eight pairs of GBM tissues and adjacent nontumor tissues. (D) The IHC scores of COL1A2 in GBM tissues and matched adjacent nontumor tissues. (E, F) The protein expression level of COL1A2 in GBM tissues was significantly higher than that in corresponding adjacent tissues. (G) K-M survival curves of OS in GBM patients between the high and low COL1A2 expression groups. *P < 0.05, **P < 0.01.
Figure 8

COL1A2 promotes the proliferation, migration, and invasion of GBM cells in vitro. (A, B) Differential expression of COL1A2 at transcriptional (A) and protein levels (B) between different GBM cell lines. (C, D) The efficiency of COL1A2 knock-down plasmid at transcriptional (C) and protein levels (D) was verified in U87 cells. (E, F) The efficiency of COL1A2 overexpression plasmid at transcriptional (E) and protein levels (F) was verified in U251 cells. In (A-F), data are presented as the mean ± SD. X-axis: four different cell lines. Y-axis: expression level of genes. (G) The cell proliferation of U87 and U251 cells were detected by a colony formation assay. (H, I) The migration and invasion ability of U87 and U251 cells were detected by wound healing and Transwell assay. The representative photographs were photographed under a microscope. *P<0.05, **P<0.01, ***P<0.001.

Figure 9

COL1A2 inhibits the effect of activated Jurkat cells on the viability of GBM cells in vitro. (A) Immunofluorescence staining of the nucleus, COL1A2, and CD8 in eight cases of GBM tissues. (B, C) Jurkat cells were stimulated by PMA and ionomycin, or by anti-CD3 and anti-CD28 antibodies. After 24 hours of stimulation, the expression of CD8 and CD69 proteins were detected. (D, E) The viability of U87 (D) and U251 (E) cells were analyzed by measuring luciferase activity. ***P<0.001.

Supplementary Files

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

- FigureS1.png
- FigureS2.png
- FigureS3.png
- FigureS4.png
- FigureS5.png
- FigureS6.png
- FigureS7.png
- FigureS8.png
- TableS1.xlsx
- TableS2.xlsx
- TableS3.xlsx
- TableS4.xlsx
- TableS5.xlsx
• TableS6.xlsx
• TableS7.xlsx