

Identification of collagen genes related to immune infiltration and epithelial-mesenchymal transition in glioma

Wen Yin

Xiangya Hospital Central South University

Jun Tan

Xiangya Hospital Central South University

Zhaoqi Xin

Xiangya Hospital Central South University

Quanwei Zhou

Xiangya Hospital Central South University

Yudong Cao

Xiangya Hospital Central South University

Zhaoping Wu

Xiangya Hospital Central South University

Wang Lei

Xiangya Hospital Central South University

Xingjun Jiang

Xiangya Hospital Central South University

Caiping Ren

Central South University Xiangya School of Medicine

Guihua Tang (✉ guihuatang0827@163.COM)

Hunan Provincial People's Hospital <https://orcid.org/0000-0002-7740-1289>

Primary research

Keywords: glioma, collagen gene, weighted gene co-expression network analysis (WGCNA), prognostic markers, epithelial-mesenchymal transition (EMT), immune microenvironment

Posted Date: August 21st, 2020

DOI: <https://doi.org/10.21203/rs.3.rs-44941/v2>

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

[Read Full License](#)

Version of Record: A version of this preprint was published at Cancer Cell International on May 25th, 2021. See the published version at <https://doi.org/10.1186/s12935-021-01982-0>.

Abstract

Background: Gliomas account for the majority of fatal primary brain tumors, and there is much room for research in the underlying pathogenesis, the multistep progression of glioma, and how to improve survival. In our study, we aimed to identify potential biomarkers or therapeutic targets of glioma and study the mechanism underlying the tumor progression.

Methods: We downloaded the microarray datasets (GSE43378 and GSE7696) from the Gene Expression Omnibus (GEO) database. Then, we used weighted gene co-expression network analysis (WGCNA) to screen potential biomarkers or therapeutic targets related to the tumor progression. ESTIMATE (Estimation of STromal and Immune cells in MAlignant Tumors using Expression data) algorithm and TIMER (Tumor Immune Estimation Resource) database were used to analyze the correlation between the selected genes and the tumor microenvironment. Real-time reverse transcription polymerase chain reaction was used to measure the selected gene. Transwell and wound healing assay were used to measure the cell migration and invasion capacity.

Results: We identified specific module genes that were positively correlated with the WHO grade but negatively correlated with OS of glioma. Importantly, we identified that 6 collagen genes (COL1A1, COL1A2, COL3A1, COL4A1, COL4A2, and COL5A2) could regulate the immunosuppressive microenvironment of glioma. Moreover, we found that these collagen genes were significantly involved in the epithelial-mesenchymal transition (EMT) process of glioma. Finally, taking COL5A2 as a further research object, the results showed that knockdown of COL5A2 significantly inhibited the migration and invasion of SHG44 and A172 cells.

Conclusions: In summary, our study demonstrated that collagen genes play an important role in regulating the immunosuppressive microenvironment and EMT process of glioma and could serve as potential therapeutic targets for glioma management.

1. Introduction

Gliomas are the most fatal primary brain tumors[1]. There were about 30% central nervous system (CNS) tumors diagnosed as gliomas in the United States, and even worse, this data was as high as 81% in malignant CNS tumors[2]. The World Health Organization (WHO) classified gliomas into grades I–IV based on histological characteristics, which contained increased degrees of anaplasia, undifferentiation, and infiltration[3]. Diffuse low-grade (WHO grade II) and III intermediate-grade gliomas are called lower-grade gliomas (LGGs)[4]. Because of their rapid growth and highly remarkable aggressiveness, a subset of these LGGs will develop into glioblastoma (WHO grade IV) within several months[5]. Glioblastoma (GBM) is the most common and malignant glioma, which is characterized by poor clinical prognosis and the survival time rarely exceeds 14 months[6]. ;

Unfortunately, despite some remarkable achievements, the median survival time of glioma patients has not been significantly prolonged in the past decades[5]. There is still much room for research in the

underlying pathogenesis, the multistep progression of glioma, and how to improve survival. Malignant gliomas are diffusely infiltrated into the surrounding normal brain tissue. Brain extracellular matrix (ECM) plays a crucial role in modulating and driving glioma invasion[7]. The ECM comprises numerous proteins, including collagen, proteoglycans, laminin, and fibronectin[8]. Collagen is the most important component of the ECM acting as a scaffold to provide sites for tumor cell adhesion[8]. Few studies have focused on the relationship between collagen genes and glioma; more studies are needed to elucidate the function of collagen genes in glioma progression.

Weighted gene co-expression network (WGCNA) is a powerful systems biology method to identify potential biomarkers or therapeutic targets and study the mechanism underlying the tumor progression involved[9, 10]. In this study, we performed WGCNA to identify the clusters of highly interconnected genes that correlated with WHO grades and overall survival, and we found such module and intramodular hub genes by facilitating network-based gene screening methods. Then, six collagen genes (COL1A1, COL1A2, COL3A1, COL4A1, COL4A2, and COL5A2) were selected for further validation in Oncomine, TCGA (The Cancer Genome Atlas) and CGGA (Chinese Glioma Genome Atlas) database. Further analysis revealed that the expression of the collagen genes is positively related to stromal and immune scores, immunosuppressive cell recruitment and immunosuppressive factors, and the infiltration of various immune cells. Moreover, our study also demonstrated that the collagen genes were significantly involved in the epithelial-mesenchymal transition (EMT) process, which was closely related to glioma malignancies. Finally, taking COL5A2 as a further research object, the result showed that knockdown COL5A2 could significantly inhibit the migration and invasion of glioma cells *in vitro*. These results may provide a novel understanding of collagen genes in glioma progression. At the same time, we also found that collagen genes may become promising therapeutic targets for fighting against gliomas.

2. Materials And Methods

2.1. Data Collection

The gliomas mRNA expression microarrays were downloaded from the repository browser of the Gene Expression Omnibus (GEO) database (<https://www.ncbi.nlm.nih.gov/geo/summary/>). We selected such datasets of which the presented sample information contained WHO grades and survival time. Besides, datasets with sample size less than 30 were removed to minimize bias resulted from small samples. Ultimately, datasets with the accession number of GSE43378[11] and GSE7696[12, 13] performed on the Affymetrix Human Genome U133 Plus 2.0 Array platform (GPL570) were downloaded. The raw data (CEL files and annotation files) of this platform and related sample information were also obtained. There were a total of 120 glioma samples and 4 non-tumor specimens in this study after eliminating recurrent samples.

2.2. Differentially Expressed Genes (DEGs) Screening

We applied the R package “affy”[14] for preprocessing expression microarrays. Briefly, original files were normalized using the Robust-Multi Array (RMA) method followed by logarithm transformed based on 2.

Subsequently, another R package “sva”[15] was used to adjust batch effects between different datasets and “limma”[16] package to screen differentially expressed genes (DEGs) between gliomas and non-tumor controls. The cut-off criteria of DEGs were defined as \log_2 fold-change (\log_2 FC) more than 1 or less than -1 with false discovery rate (FDR) adjusted p -value less than 0.05.

2.3. Gene Co-expression Network Construction

A weighted gene co-expression network of DEGs was constructed by using an R package “WGCNA”[9]. Outlier samples were eliminated by clustering analysis before WGCNA analysis to ensure network reliability. A suitable soft threshold power of 8 with a model fitting index $R^2 > 0.9$ was selected to maximize scale-free topology and the adjacencies matrix of topology similarity were calculated applying a power function, following, transformed to a topological overlap matrix (TOM). Then, the corresponding dissimilarity (1-TOM) was also calculated as the distance to hierarchically cluster genes, with which the modules were identified by the dynamic tree cut method and resulted to dendrogram. The module eigengene (ME) and module membership (MM) were calculated by function `moduleEigengenes` and `signedKME` respectively. ME was considered as the first principal component of a clustered module representing the gene expression profiles, and MM correlated ME with gene expression values, so it quantified the membership of a gene with respect to a given module.

2.4. Identifying Significant Modules Associated with Clinical Traits

MEs were applied to evaluate the correlations between identified modules and clinical traits including WHO grade and survival time. It was considered statistically significantly correlated when the p -value of Pearson’s correlation tests was no more than 0.05. The module was identified as the candidate that of the most outstanding correlation coefficients and significant correlations with clinical traits. For further analysis, the gene significance (GS) was defined as the correlation between individual genes and clinical traits. Generally, GS and MM were highly associated, meaning that the genes were highly important elements and significantly correlated with clinical traits.

2.5. GO Terms and KEGG Pathways

Gene ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) enrichment analysis of candidate module genes were performed to understand the biological functions[17]. GO terms which included cellular components (CC), biological processes (BP) and molecular function (MF) were achieved in the Database for Annotation, Visualization, and Integrated Discovery (DAVID), and KEGG pathway enrichment analysis was accomplished in R software using KEGG package. The enrichment analysis was considered as significant only when the FDR < 0.05.

2.6. Screening and Validation of the Collagen Genes

The co-expression network of the candidate module was imported to Cytoscape[18] software (v3.6.1), which is powerful for integrating and visualization biomolecular interaction networks. Hub genes that

highly connected intramodule were determined by cytoHubba. Genes that positively correlated with glioma grade and with a significant $P < 0.05$ in grade plots and survival analysis were defined as the “actual” hub genes. Six collagen genes were identified from the hub genes for further analysis.

Subsequently, to determine the expression level of collagen genes in different cancer types and validate the relationship between candidate collagen genes and glioma grades, online expression analysis was performed on the database Oncomine[19] (<https://www.oncomine.org/>). Simultaneously, the Chinese Glioma Genome Atlas (CGGA) dataset was used to not only validate the relationship between candidate gene expression and WHO grades, but also molecular subgroup based on the mutation status of IDH and codeletion status of 1p/19q. Moreover, GBMLGG RNA-seq datasets in The Cancer Genome Atlas (TCGA) database was download from UCSC Xena (<https://xena.ucsc.edu/>) to validate the relationship between candidate gene expression and WHO grades, and overall survival (OS) analyses were also completed in the Gene Expression Profiling Interactive Analysis (GEPIA)[20] database to display the prognostic values of these genes. Finally, the protein-level validation of the collagen genes was carried out in the Human Protein Atlas database (<https://www.proteinatlas.org/>).

2.7. Analysis of Immune Infiltration Characteristics

ESTIMATE (Estimation of STromal and Immune cells in MAlignant Tumors using Expression data) is a method that uses gene expression profile to infer the fraction of stromal and immune cells in tumor samples [21]. Three scores are generated by ESTIMATE algorithm: stromal score (reflecting the infiltration of stromal cells in tumor samples), immune score (reflecting the infiltration of immune cells in tumor samples) and estimate score (inferring tumor purity in tumor samples). According to a previous study, immunosuppressive cell recruitment factors and immunosuppressive factors were selected for further analysis [22]. Then the correlation between collagen expressions and immunosuppressive cell recruitment factor and immunosuppressive factors were analyzed by R language.

2.8. Associations between the Collagen Genes and Tumor Immune Infiltrating Cells

Tumor Immune Estimation Resource (TIMER) database (<http://cistrome.dfci.harvard.edu/TIMER/>) is an online tool for assessing the association between the specific gene(s) and tumor immune infiltrating cells[23]. 8 tumor immune infiltrating cells (B-cells, CD4+ T-cells, CD8+ T-cells, dendritic cells, macrophages, and neutrophils) were included in the TIMER database. By using the TIMER database, we explored the correlations between the collagen genes and tumor immune infiltrating cells in both LGGs and GBM.

2.9. The Correlation between Epithelial-mesenchymal Transition (EMT) Markers and the Collagen Genes

10 EMT related genes (TJP1, CDH1, CDH2, FN1, VIM, CTNNB1, TWIST1, SNAI2, SNAI1, and ACTA2) were selected from published literature [22]. Based on TCGA database, Pearson correlation coefficients between EMT markers and collagen genes were calculated. Pearson correlation coefficient reflects the degree of linear correlation between the two genes.

2.10. Protein-Protein Interaction (PPI) Network Construction

The 6 collagen genes and 10 EMT related genes were imported into the STRING database (<https://string-db.org/>), which is a web tool used to explore protein-protein interactions.

2.11. Cell Culture and Real-time Quantitative PCR (qRT-PCR)

SHG44 and A172 glioma cells were provided by Xiangya Medical School of Central South University, Changsha, China. The SHG44 and A172 cells were cultured in DMEM high glucose medium (Gibco/Thermo Fisher Scientific, Inc.) with 10% fetal bovine serum. All cells were maintained at 37 °C with a humidified atmosphere of 5% CO₂. The siRNAs against the COL5A2 gene were synthesized by RiboBio Corporation (Guangzhou, China).

Total RNA was extracted from treated SHG44 and A172 cells by Trizol lysis method. cDNA synthesis was performed by using the Thermo Scientific RevertAid First Strand cDNA Synthesis Kit (Thermo Scientific, Waltham, MA). RNA levels of COL5A2 were detected by using qRT-PCR according to the manufacturer's protocol. Expression of COL5A2 and β -actin were analyzed by the $2^{-\Delta\Delta C_t}$ method.

2.12. Wound Healing and Transwell Assays

Wound healing and transwell assay were determined as previously described[24].

2.13. Statistical Analysis

Statistical analyses were performed using GraphPad Prism 7 software. Differences in quantitative data between the two groups were compared by using Student's t-test. Each experiment was performed three times independently. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$ or **** $p < 0.0001$ was considered to be significant.

3. Results

3.1. WGCNA Identified Clinically Significant Modules

In total, 120 primary glioma samples and 4 non-tumor samples were obtained from the GSE43378 and GSE7696. After data preprocessing and normalization, 4219 DEGs were selected for analysis, including 2340 down-regulated and 1879 up-regulated in glioma samples. Expression profiling was presented in a volcano plot (**Figure 1A**).

Identified DEGs were submitted to WGCNA, which were assigned to gene co-expression modules after removing outlier samples by cluster dendrogram trees (**Figure 1B**). The co-expression network clustered to 7 modules of at least 30 genes, corresponding to 7 different colors. Module-containing genes were assigned by their intramodule connectivity and the non-clustering genes were divided into the gray module. TOM reflecting adjacencies or topological overlaps was visualized by heatmap (**Figure 1C**), the topological overlap depicted the similar commonality between nodes interconnected. It can be known

from the topological overlap heatmap that more obvious topological overlap appeared in genes within a module than across modules.

As displayed in the relationship between modules and clinical traits (**Figure 1D**), the yellow module was the most highlighted. It was not only positively correlated with WHO grade (correlation coefficient = 0.4, $p = 5e-06$) but also negative correlated with survival (correlation coefficient = -0.36, $p = 5e-05$). It was suggested that the yellow module genes played a role in promoting the development of glioma and leading to a poor prognosis. Subsequently, for further investigating the correlations between the yellow module and clinical traits, the relationships between GS and MM in the module were plotted to verify the significance of yellow module related to WHO grade (**Figure 1E**) and survival (**Figure 1F**).

3.2. Screening Collagen Genes

401 genes in the yellow module were submitted to DAVID website to enrich their potential functions. The results showed enriched BP (**Supplementary figure 1A**) was mainly focused on extracellular matrix (ECM) organization, cell adhesion, collagen fibril organization, angiogenesis, leukocyte migration, and single organismal cell-cell adhesion. MF (**Supplementary figure 1B**) mainly concentrated on ECM structural constituent, the binding of biomolecules (including protein, platelet-derived growth factor, integrin, and collagen), along with cadherin binding involved in cell-cell adhesion. With regards to CC (**Supplementary figure 1C**), a variety of enrichments were mainly involved in extracellular exosome, extracellular matrix, focal adhesion, and so on. The results of KEGG enrichment analysis (**Supplementary figure 1D**) indicated that the tumor-associated pathways in the yellow module were closely associated with ECM-receptor interaction, focal adhesion, protein processing in the endoplasmic reticulum, and PI3K-Akt signaling pathway. Based on these results, we hypothesized that genes correlated with the ECM-receptor interaction signal pathway may play an important role in the glioma progression.

We then imported the weighted co-expression networks of the yellow module to Cytoscape followed by screening the candidate hub genes, which are keynotes in the interactive network. A total of four calculation methods, including Degree, Edge percolated component (EPC), Closeness, and Radiality in cytoHubba application were employed in selecting hub genes, which were the intersection of the top 20 genes determined by these algorithms (**Table 1**). Eventually, 6 collagen genes (COL1A1, COL1A2, COL3A1, COL4A1, COL4A2, and COL5A2) in a family were identified as the candidate key genes for further verification. And the weighted co-expression network of these genes interacting within the module was visualized (**Supplementary figure 1E**).

3.3. Higher Expression of 6 Collagen Genes was Negatively Correlated with the Prognosis of Glioma Patients

To test the collagen genes, differential expression analysis among different cancers and glioma grades were executed on a cancer microarray data-mining database: Oncomine[25]. The results showed that the collagen genes were overexpressed in most types of tumors, including central nervous system cancer, breast cancer, head and neck cancer, colorectal cancer, and so on (**Supplementary figure 2A**). Further, we

utilized the Sun Brain dataset[26], a maximal sample size mRNA microarray that could be grouped by glioma grade in histology analysis. The grade plots of collagen genes (**Supplementary figure 2B**) were completed. The screened collagen genes were almost high-expression in glioma, and the expression level rises with glioma grade increases.

On the CGGA website (mRNAseq_325), collagen genes (COL1A1, COL1A2, COL3A1, COL4A1, COL4A2, and COL5A2) in WHO grades II, III and IV glioma samples were analyzed to explore the expression levels of these genes. Similarly, higher mRNA levels of these collagen genes were also found in higher grade glioma (**Supplementary figure 3A-F**). As we all know, IDH wild-type glioma patients present a shorter progression-free survival (PFS) and overall survival than mutant-type, and IDH mutant-type with 1p/19q co-deletion patients have the best prognosis[27]. Subsequently, we found that glioma patients with IDH mutant-type have lower expression levels of these collagen genes than IDH wild-type patients in both LGGs (WHO grade II and III) and GBM (WHO IV) groups (**Supplementary figure 4A-F**). Moreover, we discovered that the expression levels of 5 genes (COL1A1, COL1A2, COL3A1, COL4A1, COL4A2, and COL5A2) were lowest in LGG patients with IDH mutant-type and with 1p/19q co-deletion (**Supplementary figure 4A, C-F**). Nevertheless, the expression level of the COL1A2 gene was not statistically significant between 1p/19q co-deleted and 1p/19q non-co-deleted IDH mutant-type patients (**Supplementary figure 4B**).

From the UCSC database, level 3 RNA-Seq datasets (TCGA-GBMLGG) and their clinical information were downloaded to investigate the relationship between gene expressions and WHO grades of patients. As showed in **Figure 2A**, similar results were achieved. Besides, survival analyses were performed on the GEPIA database, which was based on The Cancer Genome Atlas (TCGA) data, to confirm the prognostic value of candidate genes. We found that high expression of collagen genes showed a significantly poor prognosis (**Figure 2B**). Moreover, we found that samples with IDH mutant-type have lower expression levels of these collagen genes. However, the expressions of the collagen genes were not significantly correlated with the 1p/19q status in the samples of the GBMLGG dataset (Figure 3A-F). In IDH wide-type diffuse gliomas, EGFR amplification or PTEN deletion always indicates a poor prognosis. Our analysis showed that IDH wide-type diffuse gliomas with EGFR amplification or PTEN deletion have higher expression of the collagen genes (Figure 3G-H). These results indicated that the expressions of the 6 collagen genes were negatively correlated with the prognosis of glioma patients, and may become an independent prognostic marker.

Moreover, immunohistochemistry (IHC) datasets retrieved from the Human Protein Atlas database were utilized to reveal the protein level of the collagen gene. In total, four collagen genes (COL1A1, COL1A2, COL4A1, and COL4A2) were retrieved (No data found for COL5A2 and COL3A1 was not detected in most of the samples) (**Supplementary figure 5A-D**).

3.4. The Collagen Gene Expressions were Correlated with Stromal and Immune Cell Infiltration in Glioma

The tumor microenvironment consists of various immune and stromal cells, which has been considered to closely correlate with patient prognosis. The correlations between collagen gene expressions and

ESTIMATE scores were examined. Results showed that collagen gene expressions were significantly positively correlated with stromal and immune scores in TCGA dataset (**Figure 4A-F**). Moreover, collagen genes also showed a significant correlation with stromal and immune scores in GEO and CGGA datasets (**Supplementary figure 6A-B**).

3.5. The Collagen Gene Expressions were Correlated with Immunosuppressive Properties in Glioma

Stromal and immune cells can secrete many factors that cultivate a chronic inflammatory and immunosuppressive intratumoral atmosphere. We found that the collagen genes were significantly positively correlated with the majority of immunosuppressive and immunosuppressive cell recruitment factors (**Figure 5A-D**). Moreover, a PPI network of collagen genes and immunosuppressive and immunosuppressive cell recruitment factors showed that they were closely correlated with each other (**Figure 5E**).

3.6. The Collagen Gene Expressions were Closely Associated with Immune Cell Infiltration in Glioma

We comprehensively explored the correlation between the collagen genes and immune cell infiltration in both LGGs and GBM by using the TIMER database. The results revealed that there was a positive correlation between the expression of COL1A1, COL1A2, COL3A1, COL4A1, and COL4A2, and the infiltration of B cells, CD8+ T cells, CD4+ T cells, macrophages, neutrophils, and dendritic cells in LGGs ($p < 0.05$) (**Figure 6A-E**). Except for in CD4+ T cells, the COL5A2 expression was positively correlated with the infiltration of the other five immune cell types (B cells, CD8+ T cells, macrophages, neutrophils, and dendritic cells; all $p < 0.05$) in LGGs (**Figure 6F**). Besides, the expression levels of 6 collagen genes were all positively correlated with the infiltration of dendritic cells (all $p < 0.05$) in GBM (**Supplementary figure 7A-F**).

3.7. The Collagen Genes were Significantly Involved in the EMT Process of Glioma

FN1, VIM, SNAI2, ACTA2, CTNNB1, TWIST1, and SNAI1 are important EMT biomarkers [22, 28]. As shown in **Figure 7A**, they were strongly or moderately positively correlated with the collagen genes (COL1A1, COL1A2, COL3A1, COL4A1, COL4A2, and COL5A2) based on the TCGA dataset. Moreover, we conducted a PPI network analysis of collagen genes and EMT related genes to explore the potential interactions among them. As expected, these collagen genes were closely correlated with the EMT related genes (**Figure 7B**).

3.8. Knockdown of COL5A2 Suppressed Migration and Invasion in Glioma Cells

To further validate the collagen genes were involved in the EMT process, we took COL5A2 for the subsequent analysis. The expression of COL5A2 in SHG44 and A172 cells was higher than normal glial cells by qRT-PCR (**Figure 8A**). To further confirm the biological function of COL5A2 on glioma progress, COL5A2 was knockdown by using small interfering RNA (siRNA) in SHG44 and A172 cells. After transfected with siCOL5A2, the expression of COL5A2 was low in SHG44 and A172 cells (**Figure 8B-C**). On account of the EMT program usually accompanied by the changes in cell migration and invasion,

then, wound healing and transwell assays were conducted to detect the migratory and invasive ability of glioma cells. The result showed that knockdown of COL5A2 leads to delayed wound healing and decreased invasive ability (**Figure 8D-E**). These findings indicated that COL5A2 promotes migration and invasion in glioma.

4. Discussion

Whenever possible, the current standard therapy of malignant glioma is surgical resection[3], which could achieve cytoreduction, relieve mass effect and provide adequate tissue for tumor histologic characterization and establishing molecular genotype[29]. Combined radiotherapy, chemotherapy, and chemoradiation therapy is also needed, but it is far from combating the tumor progression[30]. A more comprehensively prognostic classification integrating genetic and epigenetic signatures are even more important than clinical factors. This integrated pattern may be more accurate in clinical diagnosis and management decisions. Despite a deeper understanding of molecular and genetic diversity and improvements in therapeutic strategies, the current treatment effect for the majority of patients, including concomitant chemotherapy and biologically targeted agents, remain disappointing[31]. Thus, molecular characterization based on glioma progression is urgently needed for further elucidating novel and essential information, understanding underlying aggression mechanisms, and providing prospective prognostic and therapeutic biomarkers.

WGCNA is a robust method of constructing gene co-expression networks to illustrate genes expression profiles, detect modules based on similarities of expression features and recognize candidate module and hub genes related to clinical information. It has been successfully applied in a variety of tumors to explore potential crucial genes of clinical significance[32, 33]. In the present study, we identified 4219 DEGs by using bioinformatics analysis in the public microarray datasets (GSE43378 and GSE7696) between glioma samples and non-tumor samples. These DEGs were divided into 7 modules according to intramodule connectivity and yellow module was selected for further exploring because of its most prominent correlation with clinical features. Functional and enrichment analyses of the yellow module may provide a novel viewpoint to illuminate the mechanism of glioma progression. GO annotation analysis revealed that genes in the yellow module were mainly focused on ECM, indicating that ECM was likely to play crucial roles in the progression of glioma. Moreover, biological pathways demonstrated that KEGG was significantly focused on ECM-receptor interaction. These findings may help to explain the mechanism of how it influenced the progression and survival of glioma. were consistent with previous studies.

Collagens are the most significant protein component of the ECM[8]. Increased deposition of collagen alone or in combination facilitates cancer cell proliferation, migration, and metastasis by interfering with cell polarity, cell-cell adhesion and ultimately resulting in amplification of the growth factor signaling[34]. COL1A1, collagen type I alpha 1 chain, has been identified as an invasion-related gene in malignant astrocytomas[35]. COL1A2 was found significantly promoted gastric cancer cell proliferation, migration, apoptosis, and invasion[36]. In a recent study, the collagen molecules COL4A1[37] and COL3A1[38] were

proved to be essential for the development of glioma. COL4A2 was significantly higher expressed in GBM than astrocytoma, which was confirmed by RT-qPCR[39]. COL5A2 has not reported the roles in glioma progression, but they were reported to participate in various malignancies. COL5A2 was involved in osteosarcoma cells proliferation and invasion, provided new insights into cytostatic drug resistance of ovarian cancer, and represented early potential diagnostic biomarkers and therapeutic targets for bladder cancer[40-42].

Although a previous study has been found collagen genes were related to the progression and prognosis of LGGs[43], no study has been focused on the correlation between collagen genes and tumor microenvironment of glioma. The tumor microenvironment consists of inflammatory cells, stromal cells, fibroblasts, vascular endothelial cells, and ECM [44]. Collagens are important components of the ECM and play a critical role in the tumor microenvironment. In our study, the immune and stromal scores of each sample were calculated based on the ESTIMATE algorithm, and then the correlation between collagen genes and immune and stromal scores were analyzed. The result showed that collagen genes were significantly associated with immune and stromal scores. These findings confirmed that the collagen genes were significantly correlated with the immune microenvironment of glioma. Subsequent analysis showed that collagen genes were significantly positively correlated with immunosuppressive and immunosuppressive cell recruitment factors. Thus, collagen genes might be involved in regulating the immunosuppressive microenvironment of glioma. Immune cell infiltration could affect the progression and recurrence of glioma, and act as a significant determinant of prognosis and immunotherapy[45]. Based TIMER database, it was easily found that the infiltration levels of B cells, CD8+ T cells, CD4+ T cells, macrophages, neutrophils, and dendritic cells were significantly negatively correlated with the OS of LGGs, and the infiltration level of dendritic cells was significantly negatively correlated with the OS of GBM. In this study, we revealed that a significant positive correlation between the expression of the collagen genes and the infiltration of the six immune cell types (B cells, CD8+ T cells, CD4+ T cells, macrophages, neutrophils, and dendritic cells) in LGGs, and a positive correlation between the expression of the collagen genes and the infiltration of dendritic cells in GBM. That indicated that the collagen genes could not only serve as prognostic biomarkers but also reflect the immune status of glioma.

EMT, as a key factor of malignant glioma invasion enhancement, can be characterized by an increase of the expression of biomarkers, such as VIM, FN1, SNAI1, SNAI2, ACTA2, and so on [46]. We found that collagen genes were positively correlated with EMT biomarkers, thus possibly involved in the EMT process of glioma. Furthermore, PPI analysis showed that EMT related genes closely interact with the collagen genes. To further explore the function of collagen genes in promoting glioma progression, we used rarely studied COL5A2 as a further research object. The results showed that the knockdown of COL5A2 decreased the migration and invasion of glioma cells. Thus, we believe that the collagen genes are important EMT mediators in glioma.

5. Conclusion

In conclusion, the collagen genes (COL1A1, COL1A2, COL3A1, COL4A1, COL4A2, and COL5A2) can regulate the immunosuppressive microenvironment and be involved in the EMT process of glioma. Thus, collagen genes participate in the malignant progression of glioma and could serve as potential therapeutic targets for glioma management.

Abbreviations

BP: Biological process; CC: Cellular component; CGGA: Chinese Glioma Genome Atlas; CNS: Central nervous system; DEG: Differentially expressed gene; DAVID: Database for Annotation, Visualization, and Integrated Discovery; ECM: Extracellular matrix; EMT: Epithelial-mesenchymal transition; ESTIMATE: Estimation of STromal and Immune cells in MAlignant Tumors using Expression data; FDR: False discovery rate; GBM: Glioblastoma; GEO: Gene Expression Omnibus; GEPIA: Gene Expression Profiling Interactive Analysis; GO: Gene ontology; GS: Gene significance; KEGG: Kyoto Encyclopedia of Genes and Genomes; LGG: Lower-grade glioma; ME: Module eigengene; MF: Molecular function; MM: Module membership; OS: Overall survival; TCGA: The Cancer Genome Atlas; TIMER: Tumor Immune Estimation Resource; TOM: Topological overlap matrix; WGCNA: Weighted gene co-expression network analysis; WHO: World Health Organization.

Declarations

Acknowledgment

We sincerely acknowledge the public databases: GEO, Oncomine, CGGA, TCGA, and GEPIA.

Authors' contributions

Conceived and designed this study: XJJ, CPR, GHT. Wrote the manuscript: WY, JT. Data analysis and visualization: ZQX, QWZ, YDC, ZPW, LW. All authors read and approved the final manuscript.

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Funding

This study was supported by the National Natural Science Foundation of China (no.81472355) and the Hunan Provincial Science and Technology Department (no.2014FJ6006).

Availability of data and materials

The datasets used in the current study are included within the article. Other data and materials are available from the corresponding author on reasonable request.

Declaration of Competing Interest

The authors declare that there are no conflicts of interest.

References

1. Guichet PO, Masliantsev K, Tachon G, Petropoulos C, Godet J, Larrieu D, Milin S, Wager M, Karayan-Tapon L: **Fatal correlation between YAP1 expression and glioma aggressiveness: clinical and molecular evidence.** *The Journal of pathology* 2018.
2. Ostrom QT, Gittleman H, Stetson L, Virk SM, Barnholtz-Sloan JS: **Epidemiology of gliomas.** *Cancer treatment and research* 2015, **163**:1-14.
3. Omuro A, DeAngelis LM: **Glioblastoma and other malignant gliomas: a clinical review.** *Jama* 2013, **310**(17):1842-1850.
4. Patil V, Mahalingam K: **A four-protein expression prognostic signature predicts clinical outcome of lower-grade glioma.** *Gene* 2018, **679**:57-64.
5. Peng L, Fu J, Ming Y: **The miR-200 family: multiple effects on gliomas.** *Cancer management and research* 2018, **10**:1987-1992.
6. Heiland DH, Haaker G, Delev D, Mercas B, Masalha W, Heynckes S, Gabelein A, Pfeifer D, Carro MS, Weyerbrock A *et al*: **Comprehensive analysis of PD-L1 expression in glioblastoma multiforme.** *Oncotarget* 2017, **8**(26):42214-42225.
7. Ferrer VP, Moura Neto V, Mentlein R: **Glioma infiltration and extracellular matrix: key players and modulators.** *Glia* 2018, **66**(8):1542-1565.
8. Walker C, Mojares E, Del Río Hernández A: **Role of Extracellular Matrix in Development and Cancer Progression.** *International journal of molecular sciences* 2018, **19**(10).
9. Langfelder P, Horvath S: **WGCNA: an R package for weighted correlation network analysis.** *BMC bioinformatics* 2008, **9**:559.
10. Zhao H, Cai W, Su S, Zhi D, Lu J, Liu S: **Screening genes crucial for pediatric pilocytic astrocytoma using weighted gene coexpression network analysis combined with methylation data analysis.** *Cancer gene therapy* 2014, **21**(10):448-455.
11. Kawaguchi A, Yajima N, Tsuchiya N, Homma J, Sano M, Natsumeda M, Takahashi H, Fujii Y, Kakuma T, Yamanaka R: **Gene expression signature-based prognostic risk score in patients with glioblastoma.** *Cancer science* 2013, **104**(9):1205-1210.
12. Lambiv WL, Vassallo I, Delorenzi M, Shay T, Diserens AC, Misra A, Feuerstein B, Murat A, Migliavacca E, Hamou MF *et al*: **The Wnt inhibitory factor 1 (WIF1) is targeted in glioblastoma and has a tumor suppressing function potentially by induction of senescence.** *Neuro-oncology* 2011, **13**(7):736-747.

13. Murat A, Migliavacca E, Gorlia T, Lambiv WL, Shay T, Hamou MF, de Tribolet N, Regli L, Wick W, Kouwenhoven MC *et al*: **Stem cell-related "self-renewal" signature and high epidermal growth factor receptor expression associated with resistance to concomitant chemoradiotherapy in glioblastoma.** *Journal of clinical oncology : official journal of the American Society of Clinical Oncology* 2008, **26**(18):3015-3024.
14. Gautier L, Cope L, Bolstad BM, Irizarry RA: **affy—analysis of Affymetrix GeneChip data at the probe level.** *Bioinformatics (Oxford, England)* 2004, **20**(3):307-315.
15. Leek JT, Johnson WE, Parker HS, Jaffe AE, Storey JD: **The sva package for removing batch effects and other unwanted variation in high-throughput experiments.** *Bioinformatics (Oxford, England)* 2012, **28**(6):882-883.
16. Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W, Smyth GK: **limma powers differential expression analyses for RNA-sequencing and microarray studies.** *Nucleic acids research* 2015, **43**(7):e47.
17. Yin W, Tang G, Zhou Q, Cao Y, Li H, Fu X, Wu Z, Jiang X: **Expression Profile Analysis Identifies a Novel Five-Gene Signature to Improve Prognosis Prediction of Glioblastoma.** *Frontiers in genetics* 2019, **10**:419.
18. Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, Amin N, Schwikowski B, Ideker T: **Cytoscape: a software environment for integrated models of biomolecular interaction networks.** *Genome research* 2003, **13**(11):2498-2504.
19. Liu Z, Meng J, Li X, Zhu F, Liu T, Wu G, Zhang L: **Identification of Hub Genes and Key Pathways Associated with Two Subtypes of Diffuse Large B-Cell Lymphoma Based on Gene Expression Profiling via Integrated Bioinformatics.** *BioMed research international* 2018, **2018**:3574534.
20. Tang Z, Li C, Kang B, Gao G, Li C, Zhang Z: **GEPIA: a web server for cancer and normal gene expression profiling and interactive analyses.** *Nucleic acids research* 2017, **45**(W1):W98-w102.
21. Yoshihara K, Shahmoradgoli M, Martínez E, Vegesna R, Kim H, Torres-Garcia W, Treviño V, Shen H, Laird PW, Levine DA *et al*: **Inferring tumour purity and stromal and immune cell admixture from expression data.** *Nature communications* 2013, **4**:2612.
22. Su J, Ma Q, Long W, Tang H, Wu C, Luo M, Wang X, Xiao K, Li Y, Xiao Q *et al*: **LCTL Is a Prognostic Biomarker and Correlates With Stromal and Immune Infiltration in Gliomas.** *Front Oncol* 2019, **9**:1083.
23. Li T, Fan J, Wang B, Traugh N, Chen Q, Liu JS, Li B, Liu XS: **TIMER: A Web Server for Comprehensive Analysis of Tumor-Infiltrating Immune Cells.** *Cancer research* 2017, **77**(21):e108-e110.
24. Jia W, Jiang X, Liu W, Wang L, Zhu B, Zhu H, Liu X, Zhong M, Xie D, Huang W *et al*: **Effects of three-dimensional collagen scaffolds on the expression profiles and biological functions of glioma cells.** *Int J Oncol* 2018, **52**(6):1787-1800.
25. Rhodes DR, Yu J, Shanker K, Deshpande N, Varambally R, Ghosh D, Barrette T, Pandey A, Chinnaiyan AM: **ONCOMINE: a cancer microarray database and integrated data-mining platform.** *Neoplasia (New York, NY)* 2004, **6**(1):1-6.

26. Sun L, Hui AM, Su Q, Vortmeyer A, Kotliarov Y, Pastorino S, Passaniti A, Menon J, Walling J, Bailey R *et al*: **Neuronal and glioma-derived stem cell factor induces angiogenesis within the brain.** *Cancer cell* 2006, **9**(4):287-300.
27. Turkalp Z, Karamchandani J, Das S: **IDH mutation in glioma: new insights and promises for the future.** *JAMA neurology* 2014, **71**(10):1319-1325.
28. Jiang X, Tan J, Wen Y, Liu W, Wu S, Wang L, Wangou S, Liu D, Du C, Zhu B *et al*: **MSI2-TGF- β /TGF- β R1/SMAD3 positive feedback regulation in glioblastoma.** *Cancer chemotherapy and pharmacology* 2019, **84**(2):415-425.
29. Buckner J, Giannini C, Eckel-Passow J, Lachance D, Parney I, Laack N, Jenkins R: **Management of diffuse low-grade gliomas in adults - use of molecular diagnostics.** *Nature reviews Neurology* 2017, **13**(6):340-351.
30. Miyauchi JT, Tsirka SE: **Advances in immunotherapeutic research for glioma therapy.** *Journal of neurology* 2018, **265**(4):741-756.
31. Diwanji TP, Engelman A, Snider JW, Mohindra P: **Epidemiology, diagnosis, and optimal management of glioma in adolescents and young adults.** *Adolescent health, medicine and therapeutics* 2017, **8**:99-113.
32. Yuan L, Qian G, Chen L, Wu CL, Dan HC, Xiao Y, Wang X: **Co-expression Network Analysis of Biomarkers for Adrenocortical Carcinoma.** *Frontiers in genetics* 2018, **9**:328.
33. Mao Q, Zhang L, Zhang Y, Dong G, Yang Y, Xia W, Chen B, Ma W, Hu J, Jiang F *et al*: **A network-based signature to predict the survival of non-smoking lung adenocarcinoma.** *Cancer management and research* 2018, **10**:2683-2693.
34. Lu P, Weaver VM, Werb Z: **The extracellular matrix: a dynamic niche in cancer progression.** *The Journal of cell biology* 2012, **196**(4):395-406.
35. Sun S, Wang Y, Wu Y, Gao Y, Li Q, Abdulrahman AA, Liu X-F, Ji G-Q, Gao J, Li L *et al*: **Identification of COL1A1 as an invasion-related gene in malignant astrocytoma.** *Int J Oncol* 2018, **53**(6):2542-2554.
36. Ao R, Guan L, Wang Y, Wang J-N: **Silencing of COL1A2, COL6A3, and THBS2 inhibits gastric cancer cell proliferation, migration, and invasion while promoting apoptosis through the PI3k-Akt signaling pathway.** *J Cell Biochem* 2018, **119**(6):4420-4434.
37. Chen SY, Lin JS, Lin HC, Shan YS, Cheng YJ, Yang BC: **Dependence of fibroblast infiltration in tumor stroma on type IV collagen-initiated integrin signal through induction of platelet-derived growth factor.** *Biochimica et biophysica acta* 2015, **1853**(5):929-939.
38. Gao YF, Mao XY, Zhu T, Mao CX, Liu ZX, Wang ZB, Li L, Li X, Yin JY, Zhang W *et al*: **COL3A1 and SNAP91: novel glioblastoma markers with diagnostic and prognostic value.** *Oncotarget* 2016, **7**(43):70494-70503.
39. Vastrad B, Vastrad C, Godavarthi A, Chandrashekar R: **Molecular mechanisms underlying gliomas and glioblastoma pathogenesis revealed by bioinformatics analysis of microarray data.** *Med Oncol* 2017, **34**(11):182.

40. Chen H, Liu W, Zhong L, Liao D, Zhang R, Kang T, Wu Y: **NKX2-2 Suppresses Osteosarcoma Metastasis and Proliferation by Downregulating Multiple Target Genes.** *Journal of Cancer* 2018, **9**(17):3067-3077.
41. Li S, Liu X, Liu T, Meng X, Yin X, Fang C, Huang D, Cao Y, Weng H, Zeng X *et al*: **Identification of Biomarkers Correlated with the TNM Staging and Overall Survival of Patients with Bladder Cancer.** *Frontiers in physiology* 2017, **8**:947.
42. Januchowski R, Swierczewska M, Sterzynska K, Wojtowicz K, Nowicki M, Zabel M: **Increased Expression of Several Collagen Genes is Associated with Drug Resistance in Ovarian Cancer Cell Lines.** *Journal of Cancer* 2016, **7**(10):1295-1310.
43. Jiang Y, He J, Guo Y, Tao H, Pu F, Li Y: **Identification of genes related to low-grade glioma progression and prognosis based on integrated transcriptome analysis.** *J Cell Biochem* 2019.
44. Domingues P, González-Tablas M, Otero Á, Pascual D, Miranda D, Ruiz L, Sousa P, Ciudad J, Gonçalves JM, Lopes MC *et al*: **Tumor infiltrating immune cells in gliomas and meningiomas.** *Brain, behavior, and immunity* 2016, **53**:1-15.
45. Liu F, Huang J, Liu X, Cheng Q, Luo C, Liu Z: **CTLA-4 correlates with immune and clinical characteristics of glioma.** *Cancer cell international* 2020, **20**:7.
46. Meng L, Wang X, Liao W, Liu J, Liao Y, He Q: **BAF53a is a potential prognostic biomarker and promotes invasion and epithelial-mesenchymal transition of glioma cells.** *Oncol Rep* 2017, **38**(6):3327-3334.

Table

Table 1 Hub genes ranked in cytoHubba

Category	Rank methods in cytoHubba			
	Degree	EPC	Closeness	Radiality
Genes top 20	FN1	FN1	FN1	FN1
	CD44	COL4A1	COL4A1	COL4A1
	COL4A1	ANGPT2	ANXA2	ANXA2
	ANXA2	COL5A2	CD44	ANGPT2
	COL5A2	ANXA2	ANGPT2	LAMB1
	ANGPT2	COL3A1	LAMB1	COL1A1
	COL3A1	COL1A1	SERPINH1	SERPINH1
	COL1A1	CALD1	COL5A2	CALD1
	CLIC1	CD44	COL1A1	CD44
	CALD1	SERPINH1	CALD1	NAMPT
	LAMB1	COL1A2	COL1A2	LAMC1
	SERPINH1	NAMPT	LAMC1	COL5A2
	COL1A2	LAMB1	NAMPT	COL1A2
	LAMC1	LAMC1	CLIC1	PDPN
	NAMPT	CLIC1	COL3A1	COL4A2
	ADAM12	COL4A2	COL4A2	GALNT2
	CAV1	ADAM12	PDPN	P4HB
	COL4A2	CAV1	P4HB	CXCR4
	CD93	NOX4	ADAM12	CLIC1
	PDPN	ENPEP	GALNT2	COL3A1

Top 20 genes are determined by six algorithms respectively in cytoHubba. The bold ones represent selected collagen genes; EPC: Edge percolated component.

Supplementary Information

Supplementary figure 1 The most significantly enriched GO terms and KEGG pathways of the yellow module; and the weighted co-expression network of 6 collagen genes (COL1A1, COL1A2, COL3A1, COL4A1, COL4A2, and COL5A2). (A) Top 6 significantly enriched BP annotations; (B) top 6 significantly enriched MF annotations; (C) top 6 significantly enriched CC annotations; (D) top 6 significantly enriched

KEGG pathways. The length of bars reflects the number of genes, and colors reflect p -value. (E) Weighted co-expression network of collagen genes in the yellow module, which was visualized in Cytoscape. Red and blue nodes stand for the collagen genes and other co-expressed genes.

Supplementary figure 2 mRNA levels of collagen genes in glioma (ONCOMINE). (A) The expression level of collagen genes in various cancers. (B) COL1A1, COL1A2, COL3A1, COL4A1, COL4A2, and COL5A2 mRNA levels were positively correlated with WHO grades based on the Sun Brain dataset from Oncomine. Different colors represent different grades (blue represented unassigned value).

Supplementary figure 3 Glioma grade plots of the collagen genes in the CGGA database. COL1A1 (A), COL1A2 (B), COL3A1 (C), COL4A1 (D), COL4A2 (E), and COL5A2 (F) mRNA levels were positively correlated with WHO grades based on the CGGA database. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$ or **** $p < 0.0001$.

Supplementary figure 4 The mRNA expression levels of collagen genes in different subgroup. (A) COL1A1, (B) COL1A2, (C) COL3A1, (D) COL4A1, (E) COL4A2, and (F) COL5A2. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$ or **** $p < 0.0001$. ns represents without statistical significant.

Supplementary figure 5 The protein-level of collagen genes in the Human Protein Atlas database (immunohistochemistry). The translational expression level of the 4 collagen genes was positively correlated with WHO grade in glioma samples. (A) COL1A1, (B) COL1A2, (C) COL4A1, (D) COL4A2. No data found for COL5A2 and COL3A1 was not detected in most of the samples.

Supplementary figure 6 The collagen gene expressions were positively correlated with stromal, immune score, and ESTIMATE scores in glioma patients. (A) GEO dataset, (B) CGGA dataset.

Supplementary figure 7 The correlation between the collagen genes and immune cell infiltration (TIMER) in GBM. (A) COL1A1, (B) COL1A2, (C) COL3A1, (D) COL4A1, (E) COL4A2, and (F) COL5A2. p -value < 0.05 represented statistically significant.

Figures

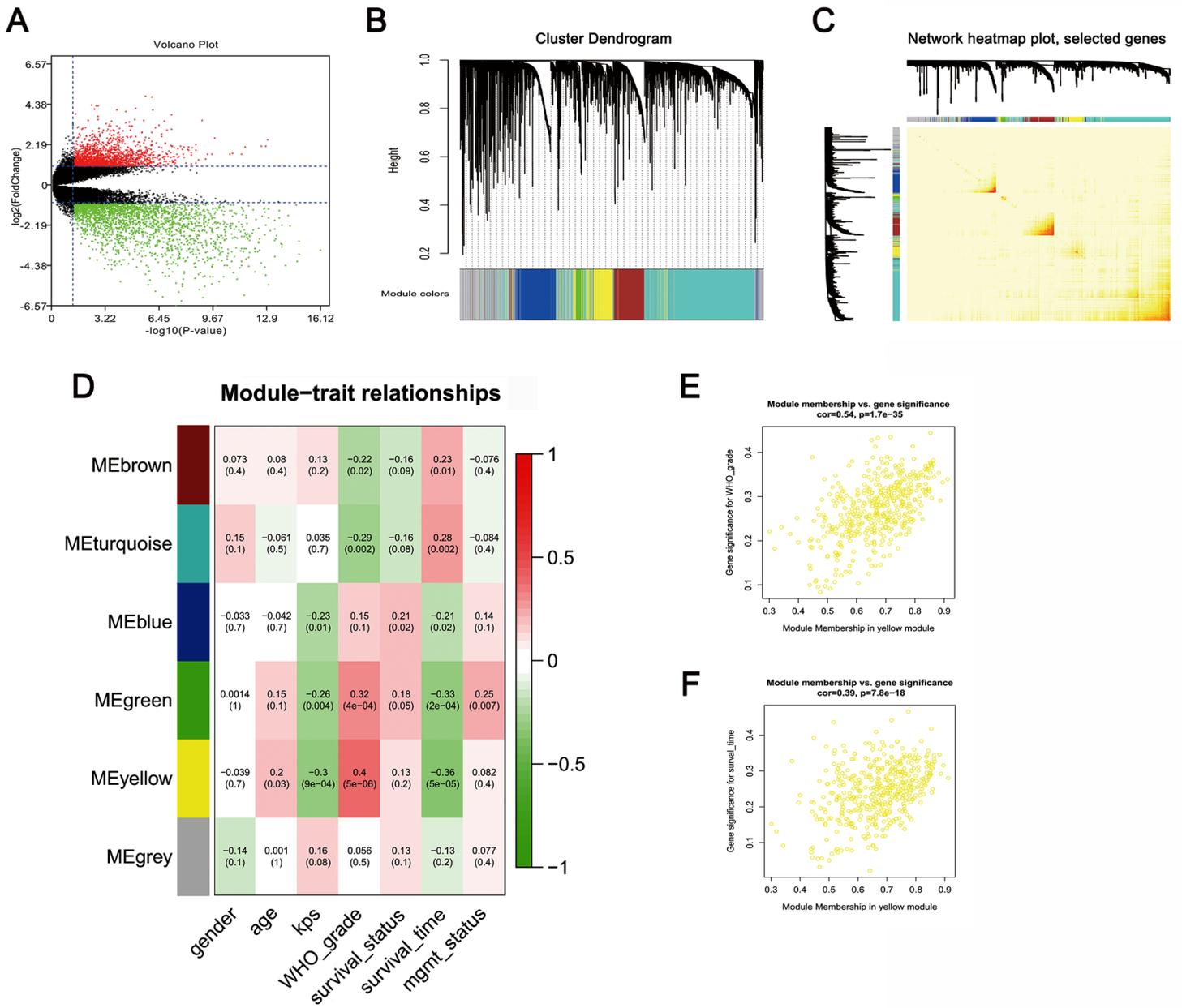


Figure 1

Identification of co-expression gene modules of glioma by WGCNA. (A) Volcano plots showing the DEGs between glioma samples and non-tumor samples. Red dots indicate up-regulated genes, while green dots down-regulated genes. (B) Cluster dendrogram of different gene modules. Genes were clustered into a module of high interconnection and marked by different colors in the horizontal bar (grey represented unassigned genes). (C) Network heatmap plot. Each column and row represent genes, low topological overlap showed light colors and higher topological overlap showed darker colors. (D) Relationships between modules and clinical traits. Each column corresponds to a clinical trait, and row corresponds to a module eigengene. Each unit includes the corresponding correlation coefficient and p-value in the first and second lines respectively. (E) and (F) indicate that the yellow module significantly correlated with glioma WHO-grade and survival time. Each dot represents a unique gene within the module, which was

plotted by MM on the x-axis and GS on the y-axis. Besides, the correlation value and p-value were displayed.

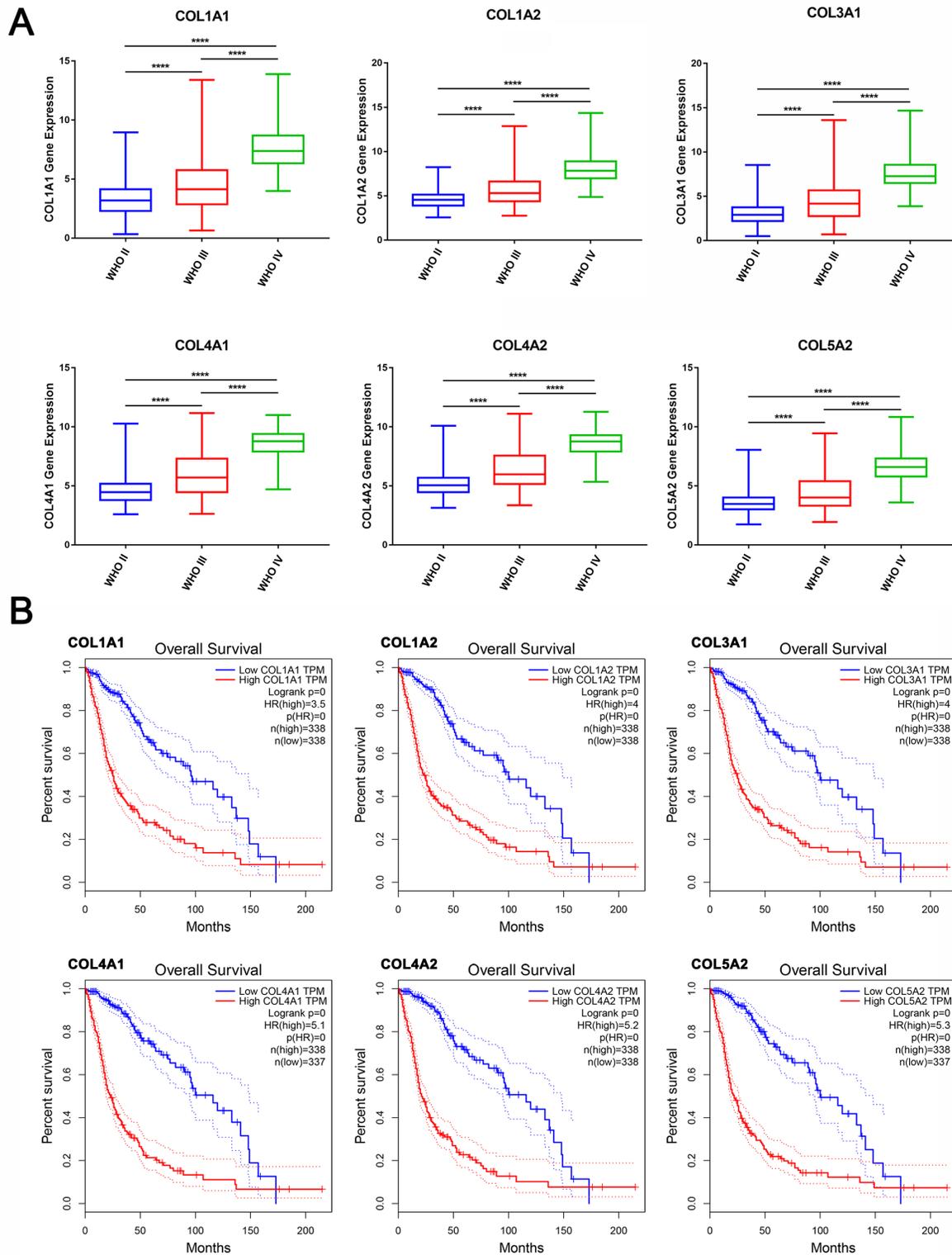


Figure 2

Increased COL1A1, COL1A2, COL3A1, COL4A1, COL4A2, and COL5A2 mRNA levels correlate with poor survival rates in glioma patients. (A) COL1A1, COL1A2, COL3A1, COL4A1, COL4A2, and COL5A2 mRNA levels were positively correlated with WHO grades based on TCGA database. (B) Kaplan-Meier analysis of

the relationships between collagen gene expression and OS time in glioma (performed in GEPIA database). Red line represents gene high expression in samples, and blue indicates low expression.

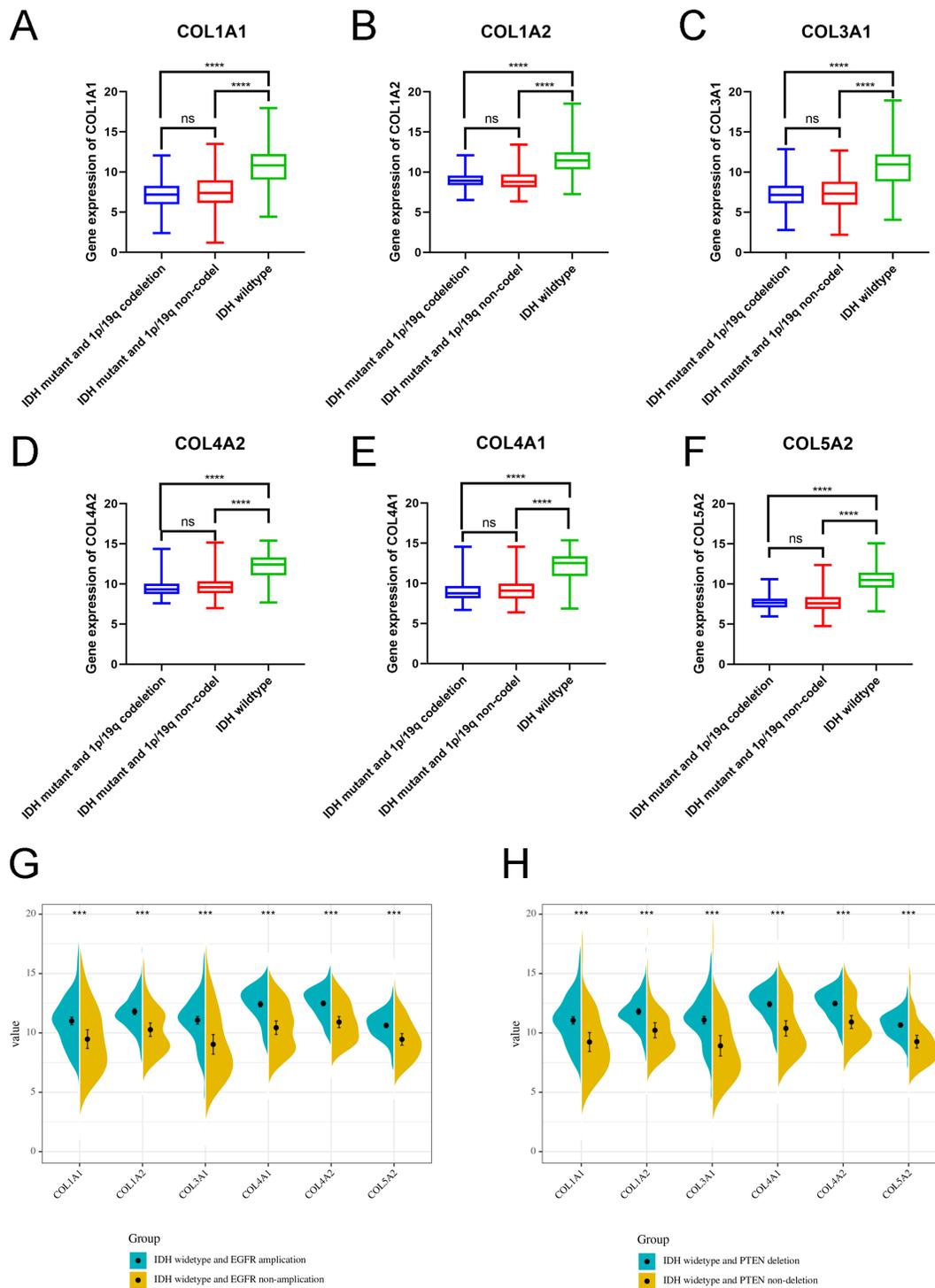


Figure 3

The mRNA expression levels of collagen genes in different subgroup of TCGA GBMLGG dataset. The expression of COL1A1 (A), COL1A2 (B), COL3A1 (C), COL4A1 (D), COL4A2 (E), and COL5A2 (F) in IDH mutant and 1p/19q codeletion, IDH mutant and 1p/19q non-codeletion, and IDH wild-type subgroups. (G)

and (H) The expression of the six genes in different IDH wild-type subgroup. *** $p < 0.001$ or **** $p < 0.0001$. ns represents without statistical significant.

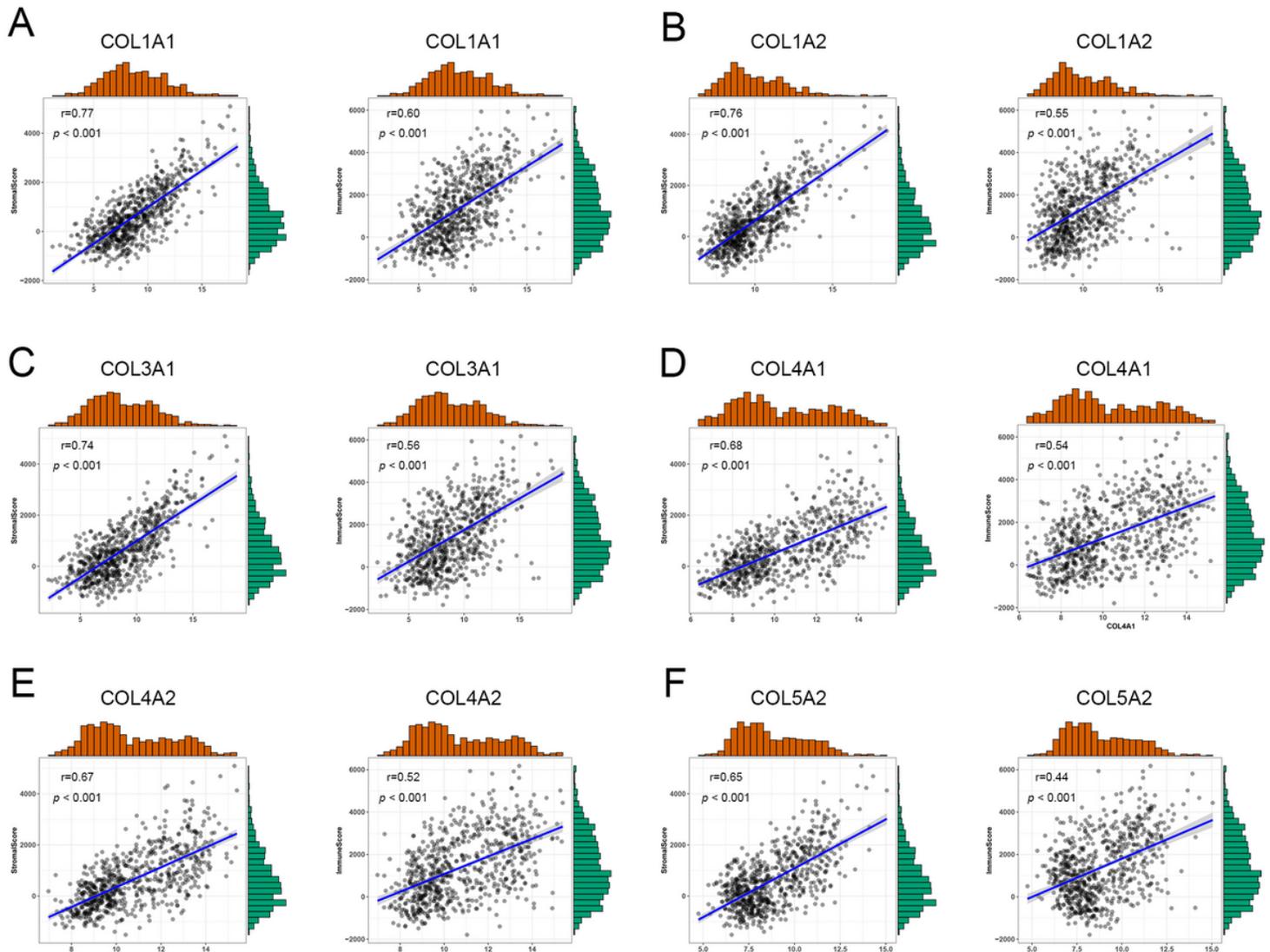


Figure 4

The collagen gene expressions were positively correlated with immunosuppressive properties. (A) Correlation between the collagen genes and immunosuppressive cell recruitment factors in TCGA dataset. (B) Correlation between the collagen genes and immunosuppressive factors in TCGA dataset. (C) Correlation between the collagen genes and immunosuppressive cell recruitment factors in the CGGA dataset. (D) Correlation between the collagen genes and immunosuppressive factors in the CGGA dataset. (E) Protein-protein interaction network of the collagen genes and immunosuppressive cell recruitment factors and immunosuppressive factors.

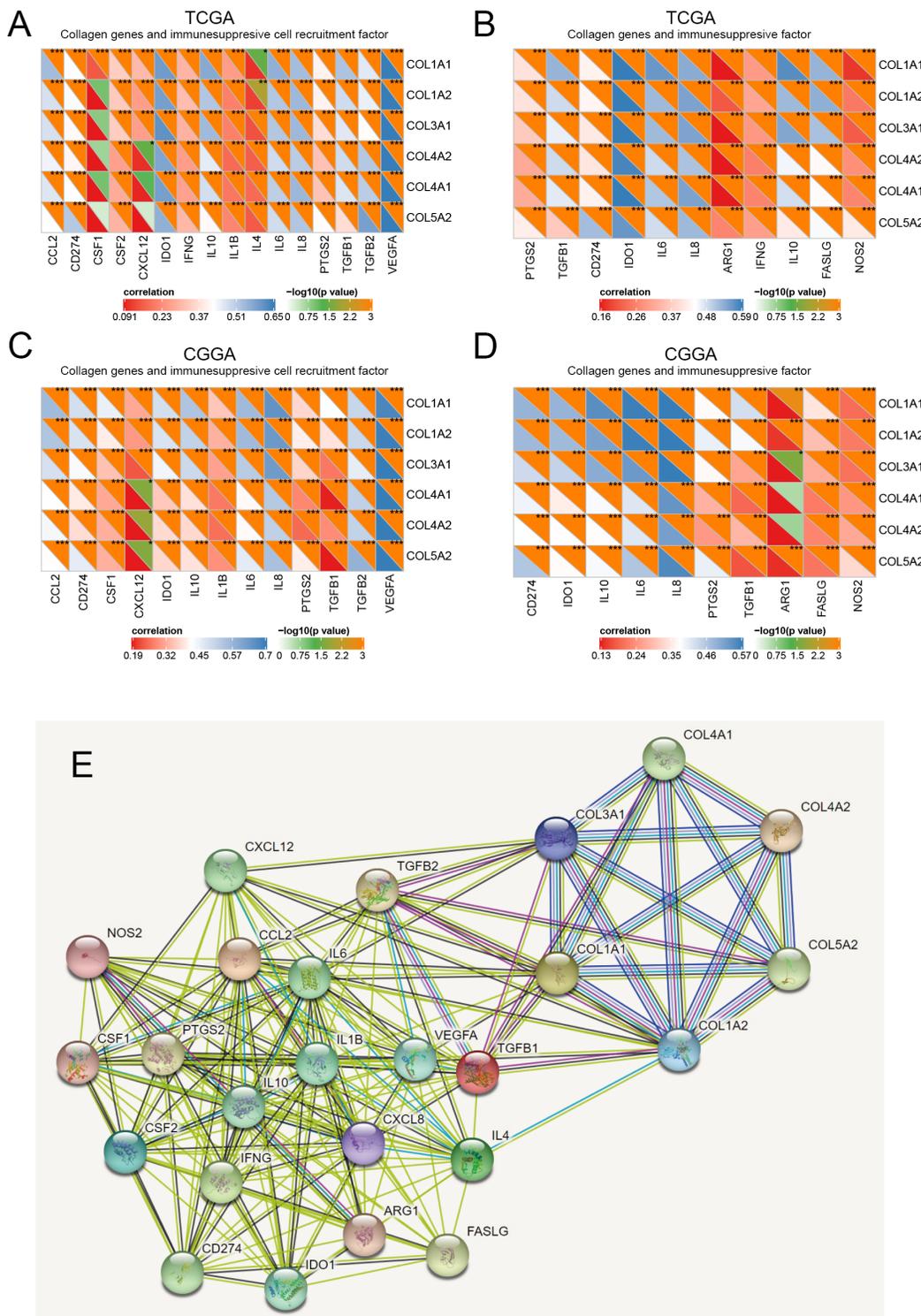


Figure 5

The collagen genes were positively correlated with significantly involved in the EMT process of Glioma. (A-C) Correlation between the collagen genes and biomarkers of epithelial-mesenchymal transition (EMT). (D) Protein-protein interaction network of the collagen genes and EMT related genes.

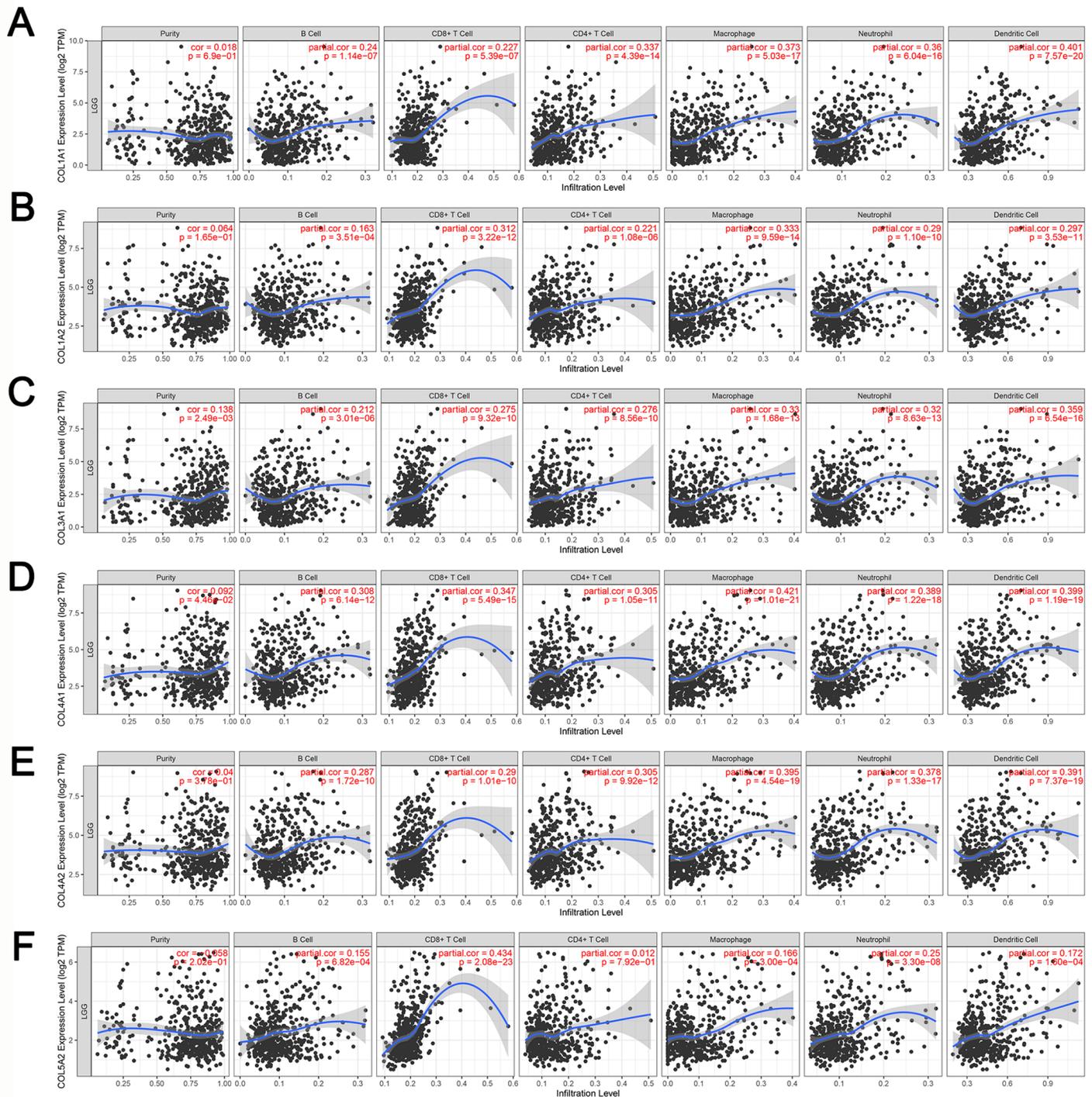


Figure 6

The correlation between the collagen genes and immune cell infiltration (TIMER) in LGGs. (A) COL1A1, (B) COL1A2, (C) COL3A1, (D) COL4A1, (E) COL4A2, and (F) COL5A2. p-value < 0.05 represented statistically significant.

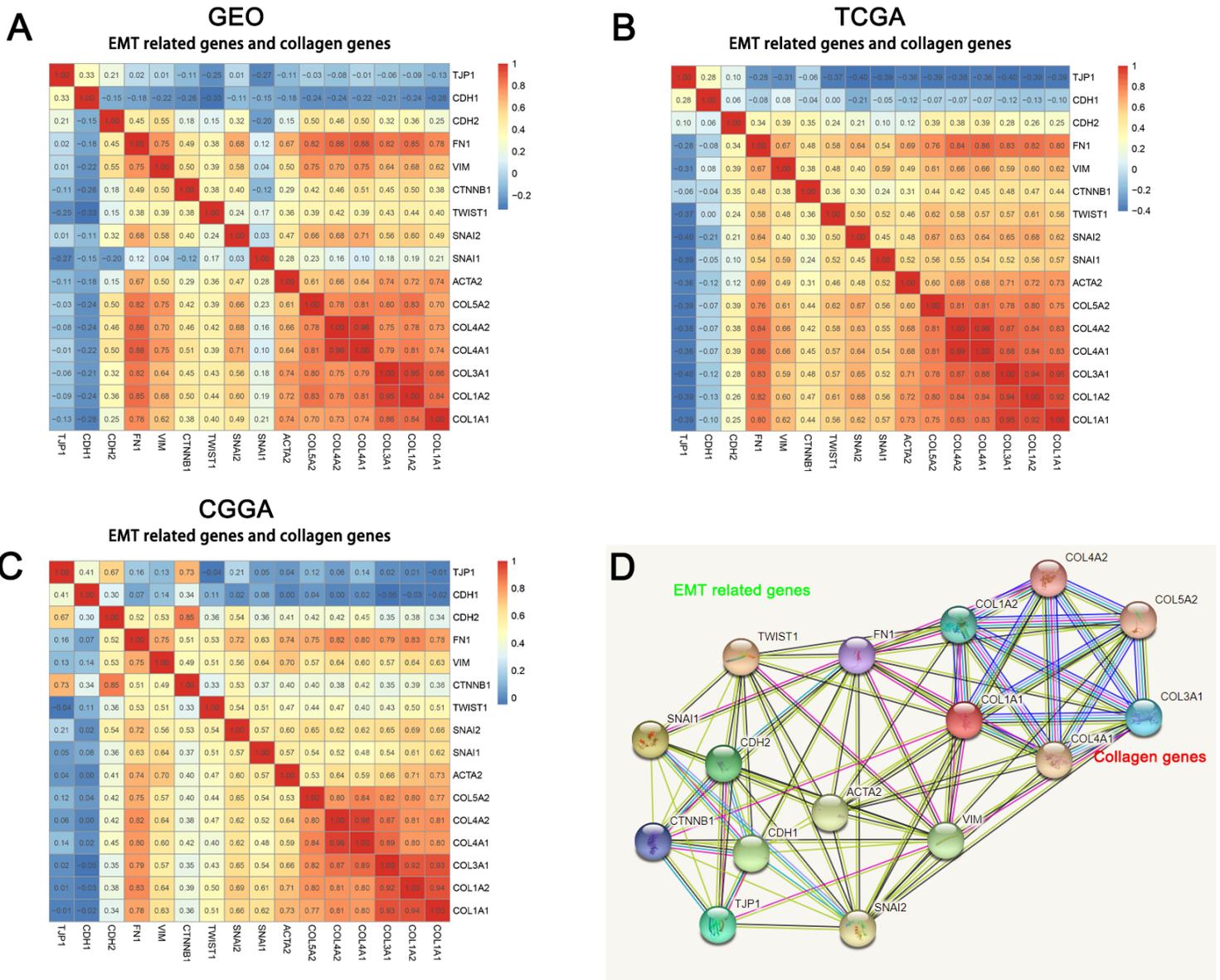


Figure 7

The collagen genes were significantly involved in the EMT process of Glioma. (A-C) Correlation between the collagen genes and biomarkers of epithelial-mesenchymal transition (EMT). (D) Protein-protein interaction network of the collagen genes and EMT related genes.

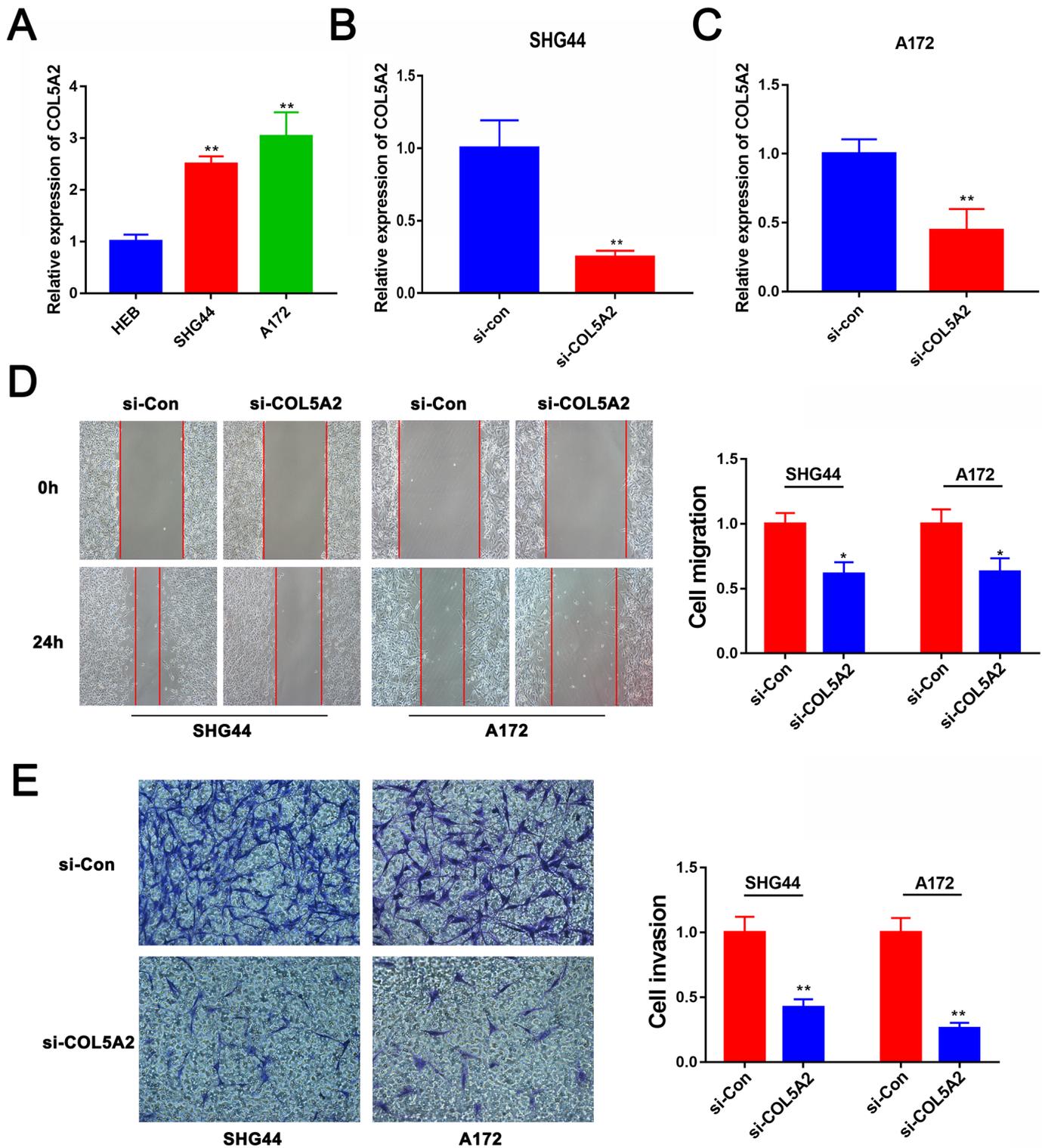


Figure 8

Knockdown of COL5A2 inhibiting migration and invasion. (A) Relative expression of COL5A2 in HEB, SHG44, and A172 cells measured by qRT-PCR. After COL5A2 siRNA transfected, SHG44 (B) and A172 (C) cells showed reduced COL5A2 level measured by qRT-PCR. (D) Wound healing assays indicated that the migratory ability of glioma cells was reduced after transfection. (E) Transwell assays revealed that the

invasiveness ability of glioma cells was inhibited after transfection. Data represented as mean value \pm SD. * $p < 0.05$, ** $p < 0.01$.

Supplementary Files

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

- [Supplementaryfigure2.tif](#)
- [Supplementaryfigure4.tif](#)
- [Supplementaryfigure7.tif](#)
- [Supplementaryfigure6.tif](#)
- [Supplementaryfigure3.tif](#)
- [Supplementaryfigure1.tif](#)
- [Supplementaryfigure5.tif](#)