Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Clinicopathological signature and prognostic value of RNA:m5C methyltransferases in gliomas

Peng Wang
Department of Neurosurgery, The Second Affiliated Hospital of Nanchang University;East China Institute of Digital Medical Engineering
Kai Huang
Department of Neurosurgery, The Second Affiliated Hospital of Nanchang University
Miaojing Wu
Department of Neurosurgery, The Second Affiliated Hospital of Nanchang University
Qing Hu
Department of Neurosurgery, The Second Affiliated Hospital of Nanchang University
Chuming Tao
Department of Neurosurgery, The Second Affiliated Hospital of Nanchang University
Kuangxun Li
School of Queen Mary, Nanchang University
Xingen Zhu (zxg2008vip@163.com)
The Second Affiliated Hospital of Nanchang University https://orcid.org/0000-0002-8556-0941

Research

Keywords: glioma, 5-methylcytosine methyltransferases, risk score, prognosis

Posted Date: February 4th, 2020

DOI: https://doi.org/10.21203/rs.2.22601/v1

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

Abstract

Background: Glioma is the most common primary intracranial tumor, accounting for the vast majority of intracranial malignant tumors. Aberrant expression of RNA:5-methylcytosine(m5C) methyltransferases has recently been the focus of research relating to the occurrence and progression of tumors. However, the prognostic value of RNA:m5C methyltransferases in glioma remains unclear. This study investigated RNA: m5C methyltransferase expression and defined its clinicopathological signature and prognostic value in gliomas. Methods: We systematically studied the RNA-sequence data of RNA:m5C methyltransferases underlying gliomas in the Chinese Glioma Genome Atlas (CGGA) and The Cancer Genome Atlas (TCGA) datasets and identified different subtypes using Consensus clustering analysis. Gene Ontology (GO) and Gene Set Enrichment analysis (GSEA) was used to annotate the function of these genes. Univariate Cox regression and the least absolute shrinkage and selection operator (LASSO) Cox regression algorithm analyses were performed to construct the risk score model. Kaplan-Meier method and Receiver operating characteristic (ROC) curves were used to assess the overall survival of glioma patients. Additionally, Cox proportional regression model analysis was developed to address the connections between the risk scores and clinical factors. Results: Consensus clustering of RNA:m5C methyltransferases identified three clusters of gliomas with different prognostic and clinicopathological features. Meanwhile, Functional annotations demonstrated that RNA:m5C methyltransferases were significantly associated with the malignant progression of gliomas. Thereafter, five RNA:m5C methyltransferase genes were screened to construct a risk score model which can be used to predict not only overall survival but also clinicopathological features in gliomas. ROC curves revealed the significant prognostic ability of this signature. In addition, Multivariate Cox regression analyses indicated that the risk score was an independent prognostic factor for glioma outcome. Conclusion: We demonstrated the role of RNA:m5C methyltransferases in the initiation and progression of glioma. We have expanded on the understanding of the molecular mechanism involved, and provided a unique approach to predictive biomarkers and targeted therapy.

Background

Traditional epigenetic modifications, including DNA methylation, histone modification and chromatin remodeling, target many biological processes that underlie the incidence and progression of cancer, including gliomas [1, 2]. In recent years, with the advent and development of high-throughput sequencing technologies coupled with direct RNA-sequencing technologies, the focus has shifted to the study of epigenetic modifications of RNA [3, 4]. Based on these sequencing technologies, published data reveals that RNA contains multiple dynamic modifications, among which the most studied are N6-methyladenosine (m^6A), 5-methylcytosine (m^5C), N1-methyladenosine (m^1A), N7-methylguanosine(N^7G), and ribose 2'-O-methylation as well as pseudouridine (Ψ) and inosine (I) [5–9]. The dynamic regulation and disorder of these RNA modifications are also significantly related to the occurrence, maintenance and progression of tumors [10, 11]. Among these RNA modifications, the m^6A was the first modification to be identified. Another well-studied modification of RNA is m^5C [12]. This post-transcriptional modification of

m^5C has been detected in most RNA species, including messenger RNAs (mRNA), mitochondrial ribosomal RNAs (rRNAs), transfer RNAs (tRNAs), enhancer RNAs (eRNAs), cytoplasmic RNAs and non-coding RNAs [12–15]. m^5C methylation of RNA is catalyzed by the NOL1/NOP2/sun domain RNA methyltransferase family and the DNA methyltransferase homologue TRDMT1 (formerly known as the DNA methyltransferase member DNMT2) in eukaryotes, but the function of the binding proteins and demethylases remains unclear, while there is evidence to suggest that YBX1 might be the binding protein for m^5C [11, 15–20]. The cellular functions and modifications of these enzymes contribute to our understanding of the mechanism of m^5C involvement in epigenetic inheritance related to various diseases, including tumors.

The NOL1/NOP2/sun domain RNA methyltransferase family consists of 7 members namely, NSUN1 (NOP2 nucleolar protein), NSUN2, NSUN3, NSUN4, NSUN5, NSUN6, and NSUN7. The biological function of these RNA:m^5C methyltransferases and the modifications they induce have revealed their importance in several aspects of protein biosynthesis, cell proliferation and differentiation, as well as on mitochondrial and nuclear gene expression [18, 20, 21]. Moreover. it has become increasingly clear that aberrant expression of RNA:m^5C methyltransferase may underlie the pathogenesis of several cancers. For instance, NSUN1, NSUN2, and NSUN4 were found to be overexpressed in a number of human cancers, including cancer of the breast, gallbladder, bladder, prostate and cervix [16, 21–24].

Glioma is the most common primary intracranial tumor, accounting for the vast majority of intracranial malignant tumors and is notorious for its high recurrence rate and resistance to treatment [2, 25, 26]. To date, no literature has reported the relationship between aberrant expression of these RNA:m^5C methyltransferases and clinicopathological features, as well as the prognostic value of these methyltransferases with respect to gliomas. It may be useful to study the biological role of these enzymes as they have potential as therapeutic targets for the treatment of glioma.

We comprehensively studied the expression profiles of the NOL1/NOP2/sun domain RNA methyltransferase family, which are RNA:m^5C methyltransferases, using the RNA sequencing data from the CGGA (n = 306) and TCGA (n = 616) datasets and aimed to investigate its prognostic value in glioma. In this study, we demonstrated the association between RNA and glioma malignant progression and constructed an RNA:m^5C methyltransferase-related risk score model to evaluate the patients with glioma. Surprisingly, this risk score model can effectively predict the malignancy and prognosis of glioma patients.

Materials And Methods

Acquisition of datasets and pre-processing

The CGGA RNA sequencing data (per kilobase of transcript per million mapped reads (RSEM)) and the relevant clinical information, such as WHO grade, *IDH*-mutation status, 1p/19q-codel status, histology, age, and survival information were downloaded from the CGGA data portal (http://www.cgga.org.cn/) as

the training set. Similarly, the TCGA RNA sequencing data (fragments per kilobase of transcript per million mapped reads (FPKM)) and clinical information were downloaded from the TCGA data portal (https://www.cancergenome.nih.gov/) and used as a verification set. Moreover, the somatic mutation data from glioma patients were also downloaded from the TCGA data portal as MAF files. After the removal of samples with missing data on survival and WHO grade, we obtained 306 (CGGA dataset) and 616 (TCGA dataset) glioma patients. The clinical information for the CGGA and TCGA datasets is listed in Additional file 5: Table S1.

The selection of seven RNA:m^{5}C methyltransferases

The seven RNA:m^{5}C methyltransferases were screened according to the method reported previously [11, 17, 18, 20]. Furthermore, the RNA expression data of seven RNA:5-methylcytosine methyltransferases were extracted from the downloaded CGGA and TCGA datasets together with the clinical information.

Consensus clustering and principal components analysis

Using R package “ConsensusClusterPlus” with unsupervised consensus clustering, we identified three subtypes based on the RNA:m^{5}C methyltransferase expression profiles of 306 patients with gliomas in CGGA dataset [27]. The appropriate number of subtypes was calculated using cumulative distribution function (CDF) and consensus matrices. Thereafter, we used principal components analysis (PCA) to detect differential gene expression amongst the three subtypes using R package pca3d and rgl [28].

Biological functional analysis

Function and interaction of the seven RNA:m^{5}C methyltransferases were predicted using the String website (https://string-db.org/) and the R package “corrplot”. Gene Ontology (GO) was performed with Metascape (http://metascape.org/) to annotate the function of differentially expressed genes in the different subtypes. Furthermore, The Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway enrichment, as well as GO analysis, was performed using Database for Annotation, Visualization and Integrated Discovery (DAVID; https://david.ncifcrf.gov/;version 6.8) and visualized in the imageGP website (http://www.ehbio.com/ImageGP/). Gene set enrichment analysis (GSEA; version 4.0.3) was performed using the JAVA program and downloaded from the website (http://software.broadinstitute.org/gsea/). The hallmark gene set (h.all.v6.0.symbols.gmt) was also downloaded. Thereafter, the hallmark gene set was determined to be significantly enriched following normalization (p-value <0.05) and a false discovery rate (FDR <0.25).

The Human Protein Atlas

The immunohistochemistry pathological specimen results from the five prognostic RNA:m^{5}C methyltransferases were downloaded from The Human Protein Atlas (https://www.proteinatlas.org/). Staining intensity, quantity and patients’ information can be inquiry online.

Construction and evaluation of risk score model

Univariate Cox regression analysis was performed to identify the genes significantly related to survival (P<0.05). Thereafter, six RNA:m^{5}C methyltransferases were screened with the least absolute shrinkage and selection operator (LASSO) multivariate Cox regression algorithm using the R package “glmnet” (version 3.0). Finally, the genes and coefficients in the risk score model were constructed based on the most suitable penalty parameter λ. The risk score formula we used was:

$$\text{Risk score} = \sum_{i=1}^{n}(Coef_i * Exp_i),$$

where $Coef_i$ is the coefficient and Exp_i is the normalized expression of each signature gene, The risk score system of five RNA:m^{5}C methyltransferases was constructed in the CGGA dataset, and evaluated with the CGGA and TCGA datasets. Patients were ranked into high-risk and low-risk groups using the median risk score. Moreover, the genomic alterations of these glioma patients were analyzed using the R package “Maftools” [29].

Statistical analysis

The RNA sequencing data were both log-transformed for the subsequent analysis. A Wilcoxon test was performed to contrast the expression of RNA:m^{5}C methyltransferases in gliomas stratified by clinicopathological features, while a Chi-square test was used for multivariate groups. Univariate, multivariate, LASSO Cox regression, and Kaplan-Meier analyses were performed using the R packages “glmnet” and “survival” [30, 31]. Receiver operating characteristic (ROC) curve analysis was performed to evaluate the prognostic efficacy of the risk score model using the R package “survivalROC”. All statistical analyses were conducted using R software (version 3.6.1) and SPSS (version 26). P<0.05 was considered statistically significant.

Results

The relationship between the aberrant expression of RNA:m^{5}C methyltransferases and clinicopathological characteristics of gliomas

m^{5}C modifications of RNA play an important role in the occurrence and progression of tumors and influence many biological functions. We comprehensively studied the relationship between each RNA:m^{5}C methyltransferase and the clinicopathological characteristics of gliomas, for example, WHO grades and isocitrate dehydrogenase (*IDH*)-mutant status. The relationship between aberrant expression of RNA:m^{5}C methyltransferases and WHO grades are shown using heatmaps (Fig. 1a, b). As shown in the heatmaps, the expression of almost all RNA:m^{5}C methyltransferases was significantly correlated with WHO grade. The significant differentially expressed RNA:m^{5}C methyltransferases include NOP2, NSUN2, NSUN4, NSUN5, NSUN6, and NSUN7. Thereafter we quantitatively analyzed the expression of these differentially expressed genes in the CGGA dataset (Fig. 1c) and verified our findings using the TCGA

dataset (Fig. 1d). As shown in the figure, NOP2, NSUN2, NSUN4, NSUN5, and NSUN7 were up-regulated and NSUN6 was down-regulated with an increase of WHO grade.

Mutation of the *IDH* gene has been reported in gliomas, particularly in low-grade gliomas (LGGs) and the prognostic value of the mutation has been confirmed by many authors in the literature. Based on these findings we investigated the relationship between aberrant expression of RNA:m^5C methyltransferases and *IDH*-mutant status in LGGs. The results showed different expression of NSUN3, NSUN4, NSUN5, NSUN6, and NSUN7 between *IDH*-mutant and *IDH*-wildtype status in both CGGA (Fig. 1e) and TCGA (Fig. 1f) datasets. In patients harboring *IDH* mutants, the expression of NSUN4, NSUN5, and NSUN7 increased, while the expression of NSUN3 and NUSN6 decreased. We also studied the expression of RNA:m^5c methyltransferases in glioblastomas stratified by *IDH* mutant status, and findings indicated that NSUN5, NSUN6, and NSUN7 were still differentially expressed. (Additional file 1: Figure S1A, B).

In addition, we predicted the gene mutation frequencies of seven RNA:m^5c methyltransferases in the cBioPortal of the Cancer Genomics Database and verified these findings using the TCGA dataset resulting in that mutations of these RNA:m^5c methyltransferases were rare (Additional file 1: Figure S1C; Additional file 5: Table S4). Even for the top-ranked RNA:m^5c methyltransferases like NOP2, the mutation frequencies were only 3%. It demonstrated that the aberrant expression of RNA:m^5C methyltransferases may not be generated by genetic mutation.

Interaction and unsupervised consensus analysis of these RNA:m^5C methyltransferases

To investigate the close connection between RNA:m^5C methyltransferases and clinicopathological features of gliomas, we systematically investigated the function, interaction, and correlation of RNA:m^5C methyltransferases. We found that all the RNA:m^5C methyltransferase genes were involved in various types of methylation, among which NOP2, NSUN3, NSUN4, and NSUN5 were mainly involved in rRNA methylation and NSUN2, NSUN6, as well as NSUN3 participated in tRNA methylation. Their function and interaction were supported by text mining and co-expression (Fig. 2a). Thereafter, a Pearson correlation analysis was performed to study the expression profile of seven RNA:m^5C methyltransferases in the CGGA (Fig. 2b) and TCGA (Additional file 2: Figure S2A) datasets. The expression of NOP2, NSUN4, and NSUN7 seemed to be closely linked. While the expression of NSUN6 was significantly negatively correlated with NSUN4, NSUN5 as well as NSUN7, and the expression of other genes were positively correlated with each other. These results were consistent with the quantitative analysis of RNA:m^5C methyltransferases expression profile in gliomas.

Based on the RNA:m^5C methyltransferase expression profiles of 306 patients with gliomas in the CGGA dataset, we used unsupervised consensus clustering analysis to identify three subtypes namely MC1, MC2, and MC3 (Fig. 2c). Clearly, k=3 seemed to be a relatively stable distinction of the samples in the CGGA dataset with clustering stability increasing from k=2 to k=9 (Additional file 2: Figure S2B-E). Furthermore, principal component analysis (PCA) was used to compare the transcription profiles of these

three subgroups. The results showed that they could be adequately divided into three distinct clusters (Fig. 2e). Next, we investigated the relationship between the RNA:m^{5}C methyltransferase expression profiles of these three subtypes and clinicopathological features of gliomas (Fig. 2d). In these three subtypes, MC2 subtype compared to MC3 subtype and MC3 subtype compared to MC1 subtype were significantly correlated with a higher grade (*P*<0.0001), *IDH*-wildtype status (*P*<0.0001), 1p/19q-noncodel status (*P*<0.0001), older average age at diagnosis (*P*<0.0001) and receipt of additional chemotherapy (*P*<0.01) (Additional file 5: Table S2). Moreover, we found significant differences in overall survival between the three groups (*P*<0.0001). The survival of patients who fell into the MC2 subtype was obviously shorter than for the other two subtypes (Fig. 2f). The results above indicating that consensus clustering of RNA:m^{5}C methyltransferases could identify subtypes with different clinicopathological features and prognosis in gliomas.

Functional annotation of the subtypes

To investigate the different clinicopathological features and overall survival rates of the three groups in gliomas, we annotated the biological processes of specific genes associated with the MC2 subtypes. A comparison was performed with the other two groups, in which 664 genes were up-regulated (log FC>1.5, normalized *P*-value<0.01 and FDR=0.05) and 645 genes were down-regulated (log FC<-1.5 and normalized *P*-value <0.01, FDR=0.05) in MC2 relative to MC1 and MC3 subtypes. GO analysis of the up-regulated genes revealed that “extracellular matrix organization”, “vasculature development”, “epithelial cell proliferation”, “cell-substrate adhesion” and “cellular response to tumor necrosis factor” were enriched in biological processes and pathways which might be highly correlated with malignant progression of gliomas. Top 20 significantly enriched biological processes were shown in the figure (Fig. 3a). The KEGG pathway analysis further revealed that these genes were also notably associated with tumor-relevant signaling pathways, for example, ECM-receptor interaction, Jak-STAT signaling pathway, and P53 signaling pathway, amongst others (Fig. 3b).

Moreover, GSEA was performed to investigate the hallmarks of tumors in the MC2 subtype. The results indicated that the tumor hallmarks, such as, P53 pathway (NES=1.720, *P*-value=0.013),P13K/AKT/mTOR signaling (NES=1.811, *P*-value=0.008), DNA repair (NES=2.050, *P*-value<0.001), and MTORC1 signaling ((NES=1.894, P-value=0.006) (Fig. 3c, d) were enriched in MC2 subtype. Combined with the above analysis, the subtypes identified by RNA:m5c methyltransferases were significantly associated with the malignant progression of gliomas.

Prognostic value of RNA:m^{5}C methyltransferases and construction of the risk score model by five RNA:m^{5}c methyltransferase genes

Based on the relationship between RNA:m^{5}C methyltransferases and malignant progression of gliomas, we further attempted to explore the prognostic role of RNA:m^{5}C methyltransferases in gliomas by a univariate survival analysis by Cox proportional hazards models of expression levels in the CGGA dataset, which was set up as a training dataset. We obtained six genes associated with prognosis

(*P*<0.01), among which NOP2, NSUN2, NSUN4, NSUN5, and NSUN7 acted as risk factors (HR>1), and NSUN6 played a protective role (HR<1) in gliomas (Fig. 4a). To improve the robustness of the six RNA:m^{5}C methyltransferases, these genes were selected to conduct an additional analysis by the LASSO Cox regression algorithm in the CGGA dataset (Fig. 4b). Five genes of RNA:m^{5}C methyltransferase genes and coefficients (Fig. 4c) were screened to construct the risk score model, and the formula for the risk score is as follows: 0.884 (expression value of NOP2) + 1.167 (expression value of NSUN4) + 0.190 (expression value of NSUN5) + (-0.161) (expression value of NSUN6) + 0.014 (expression value of NSUN7) both in the training dataset (CGGA) and the verification dataset (TCGA). In this risk score model, four genes (NOP2, NSUN4, NSUN5, and NSUN7) were cancer-promoting and NSUN6 was a cancer-suppressor gene. To better understand the role of these five prognostic genes in glioma, the K-M analysis was performed both in the CGGA dataset and TCGA dataset in which samples were classified by high or low expression according to the median gene expression level. All five of these RNA:m^{5}C methyltransferase genes were significantly correlated with OS (*P*<0.0001) (Fig. 4d; Additional file 3: Figure S3A). Moreover, we acquired the immunohistochemistry pathological specimens of the five RNA:m^{5}C methyltransferases from the website of The Human Protein Atlas, and these samples were taken from patients with a similar age and gender. We found that the protein expression of NOP2, NSUN4, NSUN5, NSUN7 in high-grade glioma tissues were much higher than those in low-grade glioma tissues, while NSUN6 was reversed (Fig. 4e). The above results demonstrate the prognostic value of these five genes in glioma.

The power of the prognostic value of the risk score model in gliomas

To obtain the predictive effect of the risk score model for clinical outcomes in patients with gliomas, the median of all patient' scores was used as a standard, and the data were divided into high- and low-risk groups both in the CGGA and TCGA datasets. The analyses indicated that the number of patients who died increased significantly as the risk score increased (Fig 5a, d). In addition, there was a significant difference in overall survival (*P*<0.0001) between the high-risk group and low-risk group (Fig. 5b, e). Thereafter, the ROC curve analyses at 1-year, 3-years, and 5-years for prognostic risk scores were performed to test the predictive efficiency of the risk model. The results showed that the risk score had high accuracy (the area under the curve (AUC) of all results in the ROC curve were >0.750) in distinguishing the OS of gliomas (Fig. 5c, f). Moreover, we performed further studies on the prognostic value of this signature in glioma patients stratified by WHO grade and *IDH*-mutant status. The results revealed that the risk score model could be used to divide glioma patients, in the CGGA dataset, into two distinct prognosis groups with different WHO grade subtypes (LGG and GBM) and *IDH*-mutant status subtypes (Fig. 5g-j). The similar results of the risk score model were also obtained using the TCGA dataset (Additional file 3: Figure S3B-E). From the comprehensive analyses above, we concluded that the prognostic efficacy of the risk score was accurate and stable.

The interrelation of the risk scores and clinicopathological characteristics in patients with gliomas

The expression of the five screened RNA:m^{5}C methyltransferases in low- and high-risk patients in the CGGA dataset are represented by heatmaps (Fig. 6a). We found statistically significant differences between the low-risk and high-risk groups both in the CGGA and TCGA datasets, based on WHO grade (P<0.0001), histology (P<0.0001), *IDH* status (P<0.0001), 1p/19q status (P<0.0001), age (P<0.0001), and receipt of additional chemotherapy (P<0.0001) (Additional file 5: Table S3). Thereafter, we explored the relationship between the risk scores and each clinicopathological characteristic. As shown in the figure, the risk scores were significantly different in these groups with the CGGA dataset (Fig. 6b) compartmentalized by WHO grade, *IDH* status, 1p/19q status, histology, age and receipt of additional chemotherapy and were verified in the TCGA dataset (Additional file 4: Figure S4A-G). Moreover, considering the importance of reported glioma-associated driver-gene alterations in glioma initiation and progression, including *ATRX*, P53 pathway (*TP53, MDM2, MDM4*), RB pathway (*CDK4, CDK6, CCND2, CDKNA/B, RB1*) and P13K/RTK pathway (*PIK3CA, PIK3R1, PTEN, EGFR, PDGFRA, NF1*), we obtained the somatic mutation data from the TCGA database. The mutational landscape of tumor driver-gene alterations between low- and high-risk patients with gliomas was rendered as a waterfall plot (Fig. 6c). The high-risk patients were characterized by more frequent alterations in tumor driver-genes than low-risk patients. This means high-risk patients harbored more progressive cancer. Combining the above results, the risk scores can predict not only the overall survival but also clinicopathological features.

Next, we investigated whether this risk score was an independent prognostic factor based on eight clinicopathological features. Univariate and multivariate Cox regression analyses were performed with the CGGA dataset. We observed that the risk score, age, WHO grade, *IDH* status, 1p/19q codel status, chemotherapy status and radiotherapy status were significantly correlated with prognosis using the univariate analysis (Fig. 7a). Multivariate analysis based on the above factors was performed and the risk score remained strongly associated with the OS (P<0.001, Fig. 7b). In the verification dataset (TCGA), we obtained similar results, including the same factors in the multivariate analysis, the risk score also remained strongly associated with the OS (P=0.002, Fig. 7c, d). The consensus results demonstrated that the risk score constructed by RNA:m^{5}C methyltransferases was a powerful and independent prognostic factor for glioma outcome. Based on this date, we explored the biological differences between the high-risk and low-risk groups. Three hundred and fifty-seven differentially up-regulated genes were obtained by variance and intersection analyses in the CGGA and TCGA datasets, which are represented by Venn diagrams (Additional file 4: Figure S4H). The GO analysis of the up-regulated genes revealed that “cell division”, “angiogenesis”, “cell adhesion”, “cell proliferation” and “extracellular matrix organization” were enriched in high-risk glioma patients (Fig. 7e). The KEGG pathway analysis further revealed that these genes were also notably associated with relevant signaling pathways in ECM-receptor interaction, cell cycle, PI3K-Akt signaling pathway, and P53 signaling pathway, amongst other (Fig. 7f).

Discussion

Increasingly, it has been demonstrated that the important role played by aberrant RNA epigenetic modifications in tumorigenesis and tumor progression, as well as patient prognosis. This suggests that

epigenetic regulators have a potential application in glioma diagnosis and prognosis. In the past, extensive literature reports involving glioma and 5-methylcytosine have focused on DNA methylation, which could serve as ideal biomarkers for cancer diagnosis [32–34]. In this research, we focus on RNA epigenetic modifications, investigating the aberrant expression of m^5C methyltransferase of RNA to explore whether RNA:m^5C methyltransferases also participated in glioma initiation and progression as well as being correlated with glioma prognosis. By analyzing expression profiles of RNA:m^5C methyltransferases from two open-access databases (TCGA and CGGA datasets), we identified three subtypes with different clinicopathological characteristics and prognoses. Moreover, the subtypes were closely correlated with tumor-related hallmarks, biological processes, and signaling pathways. Based on the features of RNA:m^5C methyltransferases, we constructed a related risk score algorithm that divided glioma patients into high- and low-risk groups to precisely predict clinical outcomes of glioma patients. Furthermore, univariate and multivariate Cox regression analyses were performed to demonstrate that it was an independent prognostic factor for glioma patients in addition to the WHO grade and IDH-mutant status as well as age at diagnosis. The risk score model based on five RNA:m^5C methyltransferases can serve as a potent prognostic signature and effectively stratify glioma patients based on risk scores and provide new insight into targeted therapy.

Of these seven RNA:m^5C methyltransferases, several genes have been reported to be involved in tumor progression across malignancies. According to the latest literature, NSUN5 plays an important role in ribosomal RNA cellular transformation and protein translation. Furthermore, NSUN5 epigenetic inactivation was associated with a better prognosis for glioma patients [35, 36]. NSUN2 is the most studied RNA:m^5C methyltransferases, and involved in the methylation of various tRNAs and mRNAs. NSUN2 can stabilize the mitotic spindle to promote tumor cell proliferation and was used to identify several targets reported in gallbladder carcinoma, bladder cancer and several tumors [16, 21, 22, 37]. NSUN1(also named NOP2, p120) encodes a protein specific to the nucleolus. It plays an important function in the synthesis of ribosomes and cell cycle of tumor proliferation [24, 38]. NSUN4 acts as cancer risk loci (Breast, Ovarian, and Prostate Cancer), and the identified MTERF4-NSUN4 axis plays a unique role in the biogenesis of mitochondrial ribosomes [23, 39, 40]. With respect to NSUN6 and NUSN7, there have been no reports on tumors and related mechanisms, and further studies are required. In this study, we found that the expression of the genes described above was increased as the WHO grade and age at diagnosis increased as well as with IDH-mutant status, indicating that the expression levels of these genes are highly correlated with malignant progression of gliomas. To the best of our knowledge, this study is the first to link these genes to the prognosis and clinical characteristics of gliomas, providing indications for further study of the molecular mechanisms involved.

For a comprehensive analysis and validation, we tested our m^5C RNA methyltransferase-related signature in a training set (CGGA) and a verification set (TCGA) respectively and found their prognostic value in predicting OS. The results of subsequent multivariate Cox regression analyses eliminated four variables other than grade and age, although among which IDH had been proved to be an effective prognostic indicator [41, 42]. This may be due to a deficiency in clinical information in the sample set, resulting in

conflicting or meaningless results. In addition, we explored the mutational landscape of tumor driver-gene alterations between low-risk and high-risk patients with gliomas[43, 44]. The results of obviously different driver-gene alterations associated with risk further prove the accuracy of this model. Furthermore, the differences in biological process analysis between high-risk groups and low-risk groups revealed that the risk associated biological processes were mainly involved in the regulation of cell cycle and cell proliferation and angiogenesis. Meanwhile, when considering the protein translation regulatory mechanism of NSUN4 and NSUN5 targeting ribosomal RNA and the effects of NOP2, NSUN2 on cell proliferation in other tumors, this may provide a new direction to determine the mechanism of malignant glioma progression in the context of these methyltransferases. It is unclear whether RNA:m^5C methyltransferases have potential clinical value for drug therapy, and this requires further study.

In conclusion, we firstly revealed the aberrant expression of RNA:m^5C methyltransferases and their relationship with clinicopathological signatures and prognostic value in glioma patients. A risk score model was built which could effectively stratify glioma patients into high-risk and low-risk so as to accurately predict survival. Additionally, we used a combination of multi-omics, multi-dataset, and multi-ethnics analyses to demonstrate the robustness of our results. However, this study still has various limitations and requires further optimization. Additional fundamental experiments are needed to reveal the molecular mechanism of m^5C RNA methyltransferases in glioma progression and the predictive efficiency of this signature needs to be tested for clinical application.

Conclusions

Our study demonstrated the significant correlation between the expression of RNA:m^5C methyltransferases and malignant progression of gliomas. Furthermore, we we constructed a m^5C RNA methyltransferase-related risk score model, which can effectively predict the prognosis of glioma patients. This study has significance for the role of RNA:m^5C methyltransferases in gliomas to enhance our understanding of the molecular mechanisms involved in the initiation and progression of gliomas and provides a unique approach to the discovery of predictive biomarkers and the selection of targeted therapy for the treatment of gliomas.

Abbreviations

TCGA: The Cancer Genome Altas; CGGA: Chinese Glioma Genome Atlas; m^5C: 5-methylcytosine; m^6A: N6-methyladenosine; *IDH*: isocitrate dehydrogenase; 1p/19q-codel: 1p/19q co-deletion; LGG: low-grade glioma; GBM: glioblastoma; PCA: principal component analysis; GO: Gene ontology; KEGG: The Kyoto Encyclopedia of Genes and Genomes; GSEA: Gene set enrichment analysis; LASSO: The least absolute shrinkage and selection operator; OS: overall survival; AUC: The area under the curve; ROC: Receiver operating characteristic

Declarations

Ethics approval and consent to participate

Not applicable

Consent for publication

Not applicable

Availability of data and materials

All data used and analyzed during the study are available in this published article and its additional files.

Competing interests

The authors declare that there have no competing interests.

Funding

This study is supported by the National Natural Science Foundation of China (Grant numbers: 81660420, 81960456) and the Natural Science Foundation of Jiangxi Province (grant number: 20171ACB20035).

Authors' contributions

PW and KH designed the research. PW, KH, CMT, and KXL contributed to the data collection and analysis, figures and tables, and were involved in manuscript writing. XGZ and KH performed the correction of the language and revision. All authors proofread and approved the final manuscript.

Acknowledgements

We thank all the participants who were involved in this study and Wiley Editing Services (http://wileyeditingservices.com/cn/) for providing editorial assistance.

Author details

[1]Department of Neurosurgery, The Second Affiliated Hospital of Nanchang University, No.1, Minde Road, Donghu District, Nanchang 330006, Jiangxi, China

[2]East China Institute of Digital Medical Engineering, NO.36, Guangxin Avenue, Shangrao 334000, Jiangxi, China.

[3]School of Queen Mary, Nanchang University, NO.461, Bayi Avenue, Nanchang 330000, Jiangxi, China.

References

1. Dawson MA, Kouzarides T: **Cancer epigenetics: from mechanism to therapy**. *Cell* 2012, **150**(1):12-27.

2. Gusyatiner O, Hegi ME: **Glioma epigenetics: From subclassification to novel treatment options**. *Semin Cancer Biol* 2018, **51**:50-58.
3. Helm M, Motorin Y: **Detecting RNA modifications in the epitranscriptome: predict and validate**. *Nat Rev Genet* 2017, **18**(5):275-291.
4. Frye M, Jaffrey SR, Pan T, Rechavi G, Suzuki T: **RNA modifications: what have we learned and where are we headed?** *Nat Rev Genet* 2016, **17**(6):365-372.
5. Li X, Xiong X, Yi C: **Epitranscriptome sequencing technologies: decoding RNA modifications**. *Nat Methods* 2016, **14**(1):23-31.
6. Meng S, Zhou H, Feng Z, Xu Z, Tang Y, Wu M: **Epigenetics in Neurodevelopment: Emerging Role of Circular RNA**. *Front Cell Neurosci* 2019, **13**:327.
7. Gilbert WV, Bell TA, Schaening C: **Messenger RNA modifications: Form, distribution, and function**. *Science* 2016, **352**(6292):1408-1412.
8. Morena F, Argentati C, Bazzucchi M, Emiliani C, Martino S: **Above the Epitranscriptome: RNA Modifications and Stem Cell Identity**. *Genes (Basel)* 2018, **9**(7).
9. Netzband R, Pager CT: **Epitranscriptomic marks: Emerging modulators of RNA virus gene expression**. *Wiley Interdiscip Rev RNA* 2019:e1576.
10. Delaunay S, Frye M: **RNA modifications regulating cell fate in cancer**. *Nat Cell Biol* 2019, **21**(5):552-559.
11. Popis MC, Blanco S, Frye M: **Posttranscriptional methylation of transfer and ribosomal RNA in stress response pathways, cell differentiation, and cancer**. *Curr Opin Oncol* 2016, **28**(1):65-71.
12. Hussain S, Aleksic J, Blanco S, Dietmann S, Frye M: **Characterizing 5-methylcytosine in the mammalian epitranscriptome**. *Genome Biol* 2013, **14**(11):215.
13. David R, Burgess A, Parker B, Li J, Pulsford K, Sibbritt T, Preiss T, Searle IR: **Transcriptome-Wide Mapping of RNA 5-Methylcytosine in Arabidopsis mRNAs and Noncoding RNAs**. *Plant Cell* 2017, **29**(3):445-460.
14. Trixl L, Lusser A: **The dynamic RNA modification 5-methylcytosine and its emerging role as an epitranscriptomic mark**. *Wiley Interdiscip Rev RNA* 2019, **10**(1):e1510.
15. Garcia-Vilchez R, Sevilla A, Blanco S: **Post-transcriptional regulation by cytosine-5 methylation of RNA**. *Biochim Biophys Acta Gene Regul Mech* 2019, **1862**(3):240-252.
16. Chen X, Li A, Sun BF, Yang Y, Han YN, Yuan X, Chen RX, Wei WS, Liu Y, Gao CC *et al*: **5-methylcytosine promotes pathogenesis of bladder cancer through stabilizing mRNAs**. *Nat Cell Biol* 2019, **21**(8):978-990.
17. Motorin Y, Lyko F, Helm M: **5-methylcytosine in RNA: detection, enzymatic formation and biological functions**. *Nucleic Acids Res* 2010, **38**(5):1415-1430.
18. Bohnsack KE, Hobartner C, Bohnsack MT: **Eukaryotic 5-methylcytosine (m(5)C) RNA Methyltransferases: Mechanisms, Cellular Functions, and Links to Disease**. *Genes (Basel)* 2019, **10**(2).

19. Yang T, Low JJA, Woon ECY: **A general strategy exploiting m5C duplex-remodelling effect for selective detection of RNA and DNA m5C methyltransferase activity in cells**. *Nucleic Acids Res* 2019.
20. Liu RJ, Long T, Li J, Li H, Wang ED: **Structural basis for substrate binding and catalytic mechanism of a human RNA:m5C methyltransferase NSun6**. *Nucleic Acids Res* 2017, **45**(11):6684-6697.
21. Frye M, Watt FM: **The RNA methyltransferase Misu (NSun2) mediates Myc-induced proliferation and is upregulated in tumors**. *Curr Biol* 2006, **16**(10):971-981.
22. Gao Y, Wang Z, Zhu Y, Zhu Q, Yang Y, Jin Y, Zhang F, Jiang L, Ye Y, Li H *et al*: **NOP2/Sun RNA methyltransferase 2 promotes tumor progression via its interacting partner RPL6 in gallbladder carcinoma**. *Cancer Sci* 2019, **110**(11):3510-3519.
23. Kar SP, Beesley J, Amin Al Olama A, Michailidou K, Tyrer J, Kote-Jarai Z, Lawrenson K, Lindstrom S, Ramus SJ, Thompson DJ *et al*: **Genome-Wide Meta-Analyses of Breast, Ovarian, and Prostate Cancer Association Studies Identify Multiple New Susceptibility Loci Shared by at Least Two Cancer Types**. *Cancer Discov* 2016, **6**(9):1052-1067.
24. Uchiyama B, Saijo Y, Kumano N, Abe T, Fujimura S, Ohkuda K, Handa M, Satoh K, Nukiwa T: **Expression of nucleolar protein p120 in human lung cancer: difference in histological types as a marker for proliferation**. *Clin Cancer Res* 1997, **3**(10):1873-1877.
25. Ostrom QT, Bauchet L, Davis FG, Deltour I, Fisher JL, Langer CE, Pekmezci M, Schwartzbaum JA, Turner MC, Walsh KM *et al*: **The epidemiology of glioma in adults: a "state of the science" review**. *Neuro Oncol* 2014, **16**(7):896-913.
26. Ostrom QT, Gittleman H, Farah P, Ondracek A, Chen Y, Wolinsky Y, Stroup NE, Kruchko C, Barnholtz-Sloan JS: **CBTRUS statistical report: Primary brain and central nervous system tumors diagnosed in the United States in 2006-2010**. *Neuro Oncol* 2013, **15 Suppl 2**:ii1-56.
27. Wilkerson MD, Hayes DN: **ConsensusClusterPlus: a class discovery tool with confidence assessments and item tracking**. *Bioinformatics* 2010, **26**(12):1572-1573.
28. Price AL, Patterson NJ, Plenge RM, Weinblatt ME, Shadick NA, Reich D: **Principal components analysis corrects for stratification in genome-wide association studies**. *Nat Genet* 2006, **38**(8):904-909.
29. Mayakonda A, Lin DC, Assenov Y, Plass C, Koeffler HP: **Maftools: efficient and comprehensive analysis of somatic variants in cancer**. *Genome Res* 2018, **28**(11):1747-1756.
30. Xiang ZJ, Wang Y, Ramadge PJ: **Screening Tests for Lasso Problems**. *IEEE Trans Pattern Anal Mach Intell* 2017, **39**(5):1008-1027.
31. Engebretsen S, Bohlin J: **Statistical predictions with glmnet**. *Clin Epigenetics* 2019, **11**(1):123.
32. Song CX, Yin S, Ma L, Wheeler A, Chen Y, Zhang Y, Liu B, Xiong J, Zhang W, Hu J *et al*: **5-Hydroxymethylcytosine signatures in cell-free DNA provide information about tumor types and stages**. *Cell Res* 2017, **27**(10):1231-1242.
33. Li W, Zhang X, Lu X, You L, Song Y, Luo Z, Zhang J, Nie J, Zheng W, Xu D *et al*: **5-Hydroxymethylcytosine signatures in circulating cell-free DNA as diagnostic biomarkers for human cancers**. *Cell Res* 2017, **27**(10):1243-1257.

34. Wu X, Zhang Y: **TET-mediated active DNA demethylation: mechanism, function and beyond**. *Nat Rev Genet* 2017, **18**(9):517-534.
35. Janin M, Ortiz-Barahona V, de Moura MC, Martinez-Cardus A, Llinas-Arias P, Soler M, Nachmani D, Pelletier J, Schumann U, Calleja-Cervantes ME *et al*: **Epigenetic loss of RNA-methyltransferase NSUN5 in glioma targets ribosomes to drive a stress adaptive translational program**. *Acta Neuropathol* 2019, **138**(6):1053-1074.
36. Heissenberger C, Liendl L, Nagelreiter F, Gonskikh Y, Yang G, Stelzer EM, Krammer TL, Micutkova L, Vogt S, Kreil DP *et al*: **Loss of the ribosomal RNA methyltransferase NSUN5 impairs global protein synthesis and normal growth**. *Nucleic Acids Res* 2019.
37. Hussain S, Benavente SB, Nascimento E, Dragoni I, Kurowski A, Gillich A, Humphreys P, Frye M: **The nucleolar RNA methyltransferase Misu (NSun2) is required for mitotic spindle stability**. *J Cell Biol* 2009, **186**(1):27-40.
38. Freeman JW, Hazlewood JE, Auerbach P, Busch H: **Optimal loading of scraped HeLa cells with monoclonal antibodies to the proliferation-associated Mr 120,000 nucleolar antigen**. *Cancer Res* 1988, **48**(18):5246-5250.
39. Camara Y, Asin-Cayuela J, Park CB, Metodiev MD, Shi Y, Ruzzenente B, Kukat C, Habermann B, Wibom R, Hultenby K *et al*: **MTERF4 regulates translation by targeting the methyltransferase NSUN4 to the mammalian mitochondrial ribosome**. *Cell Metab* 2011, **13**(5):527-539.
40. Spahr H, Habermann B, Gustafsson CM, Larsson NG, Hallberg BM: **Structure of the human MTERF4-NSUN4 protein complex that regulates mitochondrial ribosome biogenesis**. *Proc Natl Acad Sci U S A* 2012, **109**(38):15253-15258.
41. Yan H, Parsons DW, Jin G, McLendon R, Rasheed BA, Yuan W, Kos I, Batinic-Haberle I, Jones S, Riggins GJ *et al*: **IDH1 and IDH2 mutations in gliomas**. *N Engl J Med* 2009, **360**(8):765-773.
42. Waitkus MS, Diplas BH, Yan H: **Isocitrate dehydrogenase mutations in gliomas**. *Neuro Oncol* 2016, **18**(1):16-26.
43. Brennan CW, Verhaak RG, McKenna A, Campos B, Noushmehr H, Salama SR, Zheng S, Chakravarty D, Sanborn JZ, Berman SH *et al*: **The somatic genomic landscape of glioblastoma**. *Cell* 2013, **155**(2):462-477.
44. Ceccarelli M, Barthel FP, Malta TM, Sabedot TS, Salama SR, Murray BA, Morozova O, Newton Y, Radenbaugh A, Pagnotta SM *et al*: **Molecular Profiling Reveals Biologically Discrete Subsets and Pathways of Progression in Diffuse Glioma**. *Cell* 2016, **164**(3):550-563.

Supplementary Information

Additional file 1: Figure S1. The genetic alteration and differential expression of RNA:m^{5}C methyltransferases. **A**, **B** The differential expression of seven RNA:m^{5}C methyltransferases with different *IDH*-mutant status in glioblastoma. **C** The gene mutation frequencies of seven RNA:m^{5}C methyltransferases.

Additional file 2: Figure S2. Pearson correlation and unsupervised consensus analysis of selected RNA:m^5C methyltransferases. **A** Pearson correlation analysis of seven RNA:m^5C methyltransferase expression profiles in the TCGA dataset. **B, C** Consensus clustering cumulative distribution function for k=2 to 9. Relative change in area under CDF for k=2 to 9. **D, E** Consensus clustering matrix for k=2 and k=4.

Additional file 3: Figure S3. Kaplan-Meier overall survival curves for patients in the TCGA dataset. **A** Kaplan-Meier overall survival curves for each prognostic gene in the TCGA dataset. **B, C** Kaplan-Meier overall survival curves for patients in the TGGA dataset stratified by WHO grade. **D, E** Kaplan-Meier overall survival curves for patients in the TGGA dataset stratified by *IDH*-mutant status.

Additional file 4: Figure S4. A-G The relationship between the risk scores and each clinicopathological characteristic. **H** Variance and intersection analyses in the CGGA and TCGA datasets.

Additional file 5: Table S1. The clinicopathological information of glioma patients in the CGGA and the TCGA datasets. **Table S2.** The clinicopathological characteristics of glioma patients in three subtypes identified by selected RNA:m5C methyltransferases with the CGGA dataset. **Table S3.** The clinicopathological characteristics of glioma patients in low-risk and high-risk groups identified by five RNA:m5C methyltransferase genes. **Table S4.** The somatic mutation data of seven RNA:m5c methyltransferases from the TCGA dataset.

Figures

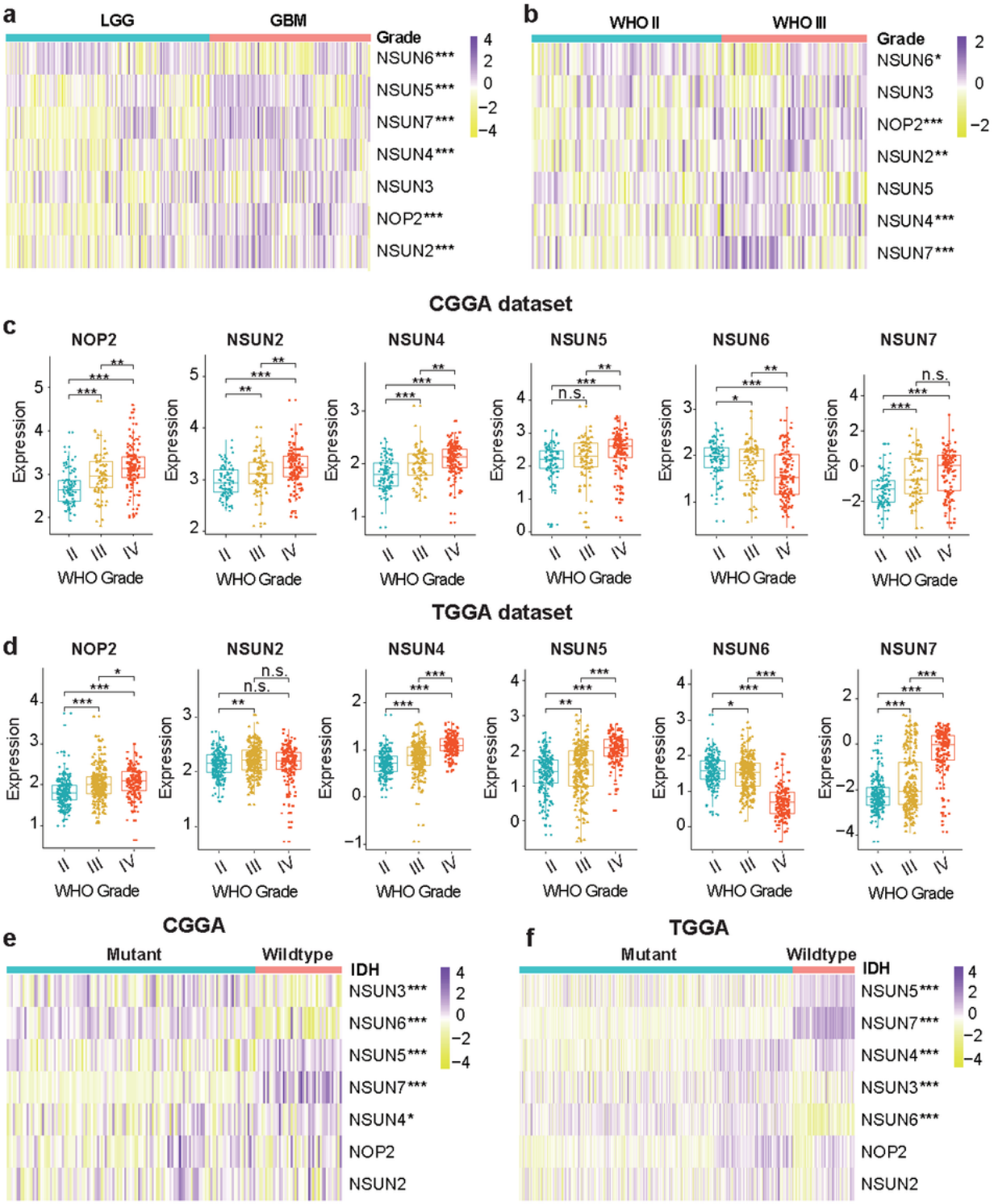

Figure 2

The relationship between the aberrant expression of RNA:m5C methyltransferases and clinicopathological characteristics of gliomas in the CGGA and TCGA datasets. a, b The differential expression of seven RNA:m5C methyltransferases with different WHO grades in gliomas. c, d The significant differential expression of RNA:m5C methyltransferases stratified by WHO grade in gliomas. e,

f The differential expression of RNA:m5C methyltransferases stratified by IDH-mutant status in low-grade gliomas. Significance: P*<0.05, P**<0.01, P***<0.001.

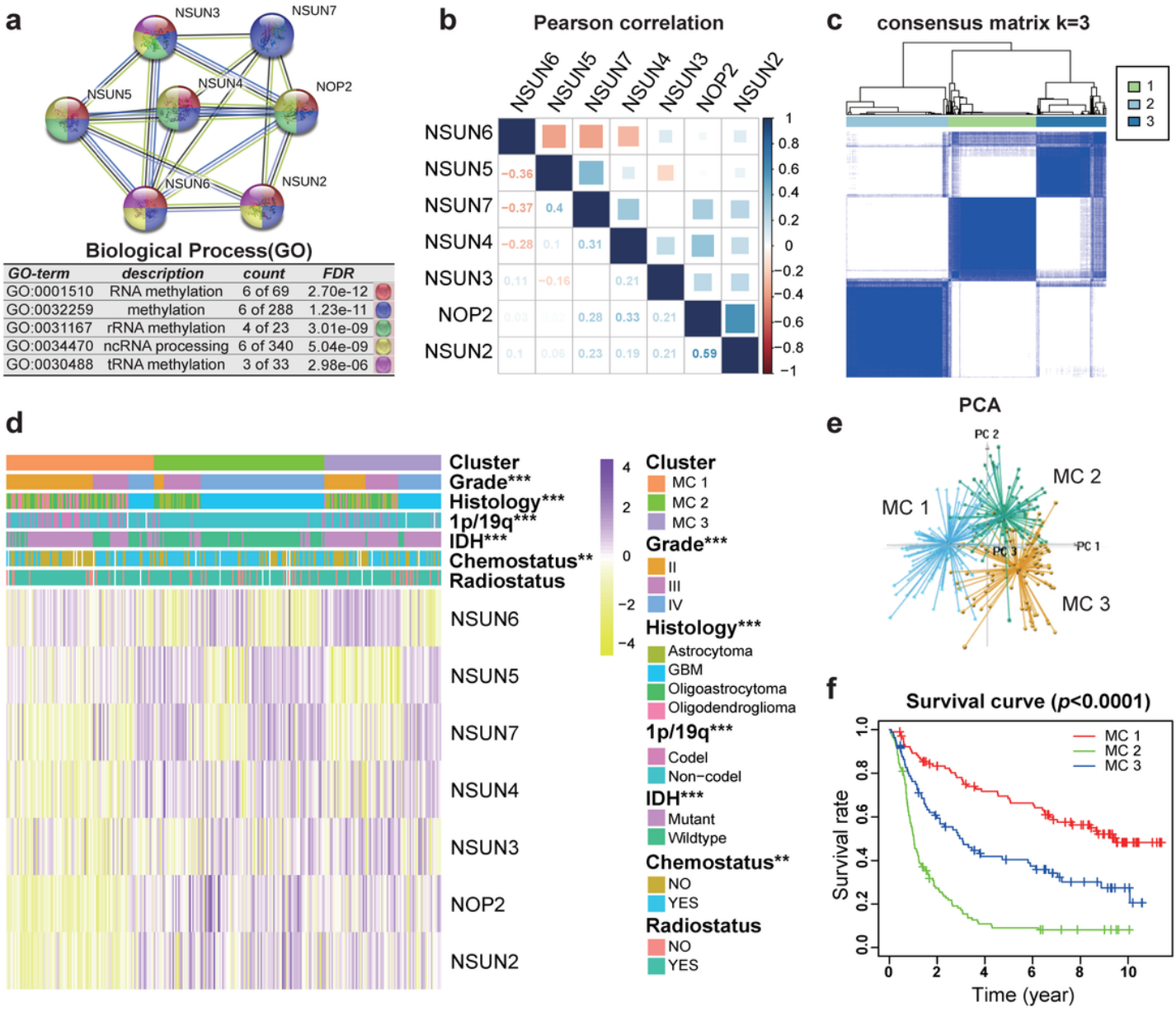

Figure 4

Interaction and unsupervised consensus analysis of selected RNA:m5C methyltransferases. a The function and interaction of seven RNA:m5C methyltransferases. b Pearson correlation analysis of seven RNA:m5C methyltransferase expression profiles in the CGGA dataset. c Consensus clustering matrix for the most suitable k=3. d The relationship between the RNA:m5C methyltransferase expression profiles of these three subtypes and clinicopathological features of gliomas. e Principal component analysis (3D) of the CGGA RNA-sequence profiles. f Kaplan-Meier overall survival curves for the glioma patients of three subtypes in the CGGA dataset. Significance: P*<0.05, P**<0.01, P***<0.001.

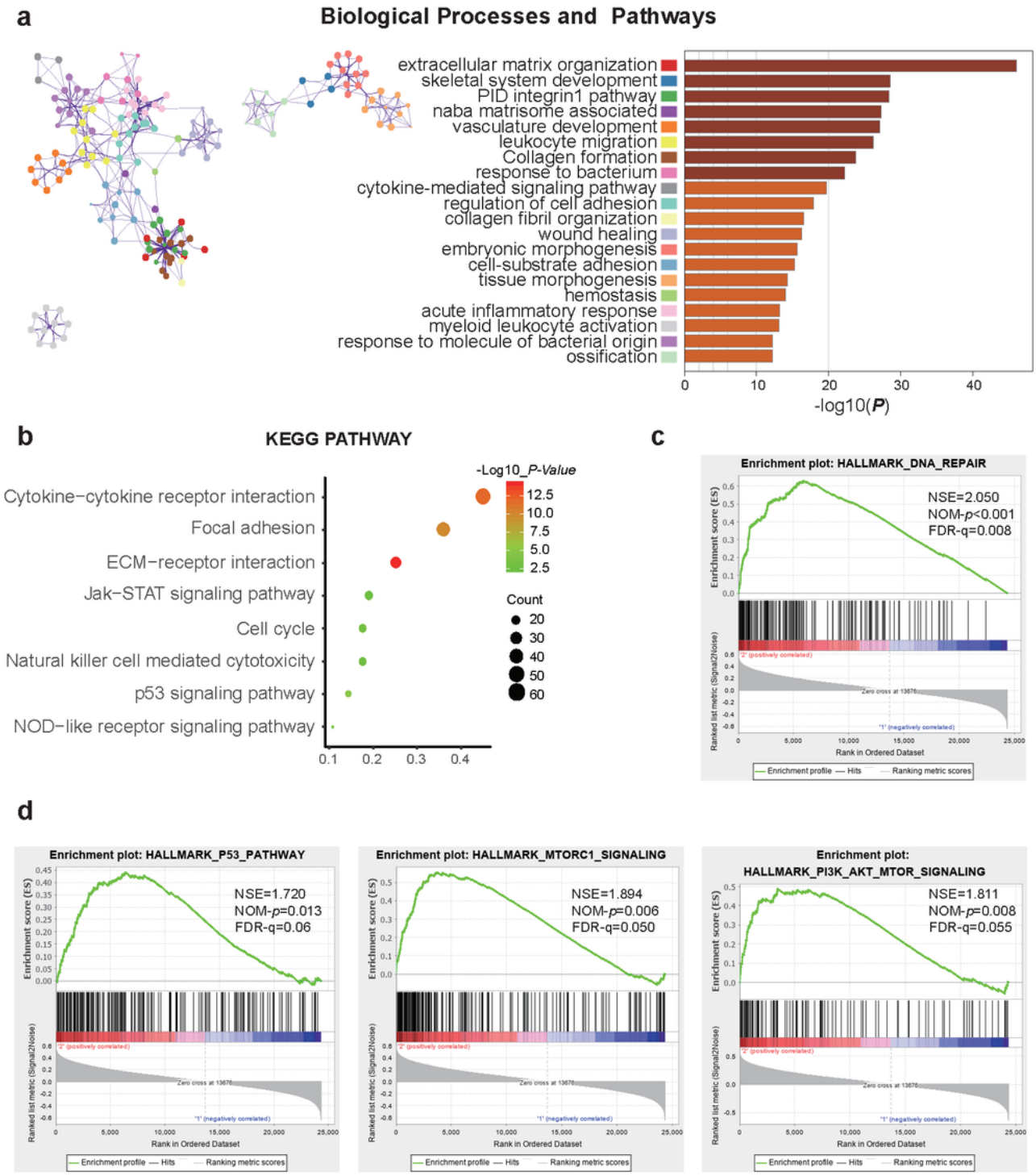

Figure 6

Function annotation of specific genes of the MC2 subtype. a Network and bar chart of 20 significantly enriched biological processes of up-regulated genes in the MC2 subtype. Each enriched node is presented in a different color. b KEGG pathway analysis of up-regulated genes in the MC2 subtype. c, d GSEA analysis of the MC2 subtype showed enrichment for various hallmarks of tumors.

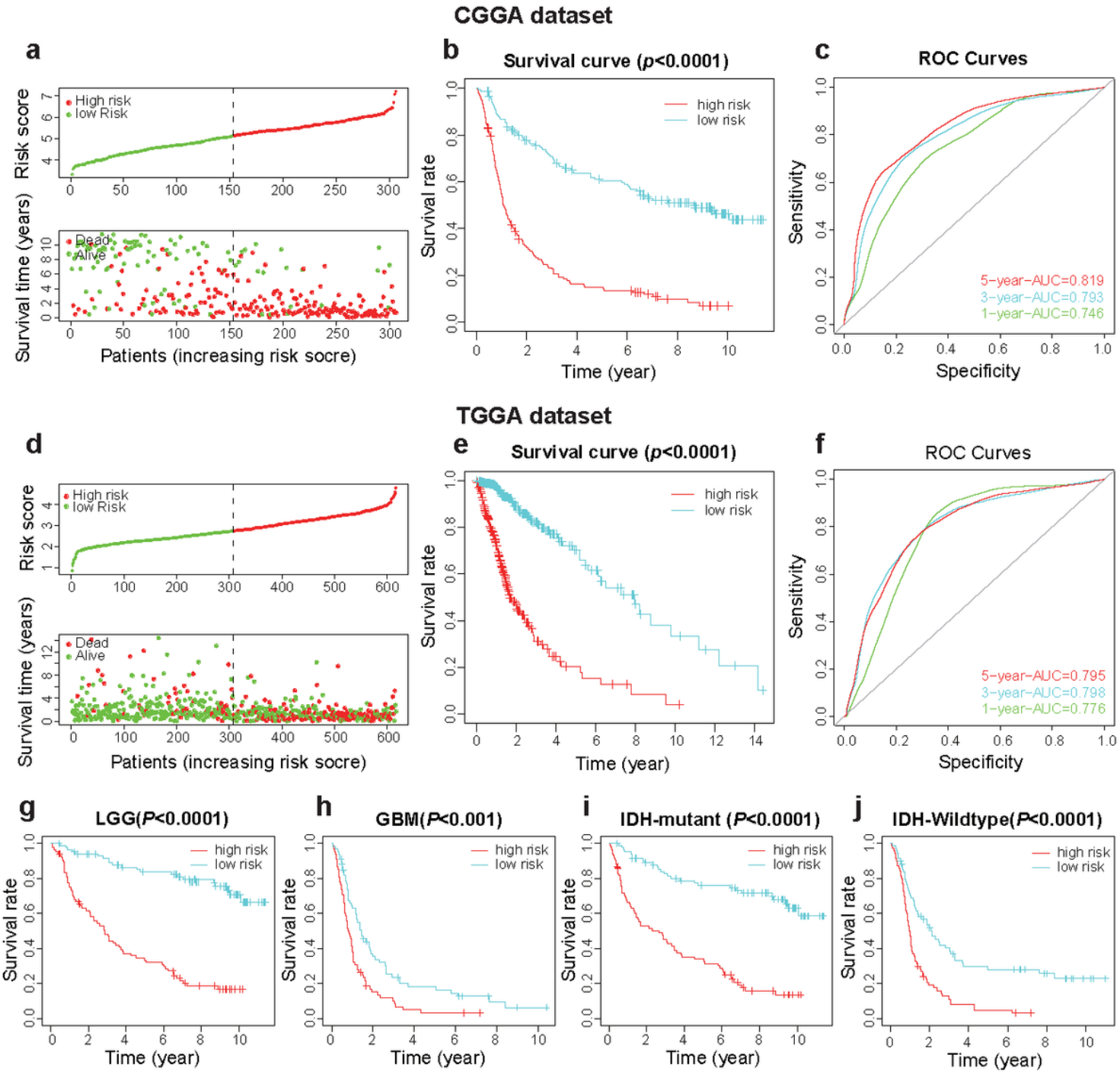

Figure 8

Construction of the risk score model using five RNA:m5C methyltransferase genes. a Univariate Cox regression analysis of seven RNA:m5C methyltransferases in the CGGA dataset. b, c Identification of five prognostic genes in the CGGA dataset and the coefficients constructed using the LASSO method. d Kaplan-Meier overall survival curves for each prognostic gene in the CGGA dataset. e Comparison of protein expression in immunohistochemical specimens of five prognostic RNA:m5C methyltransferases in glioma.

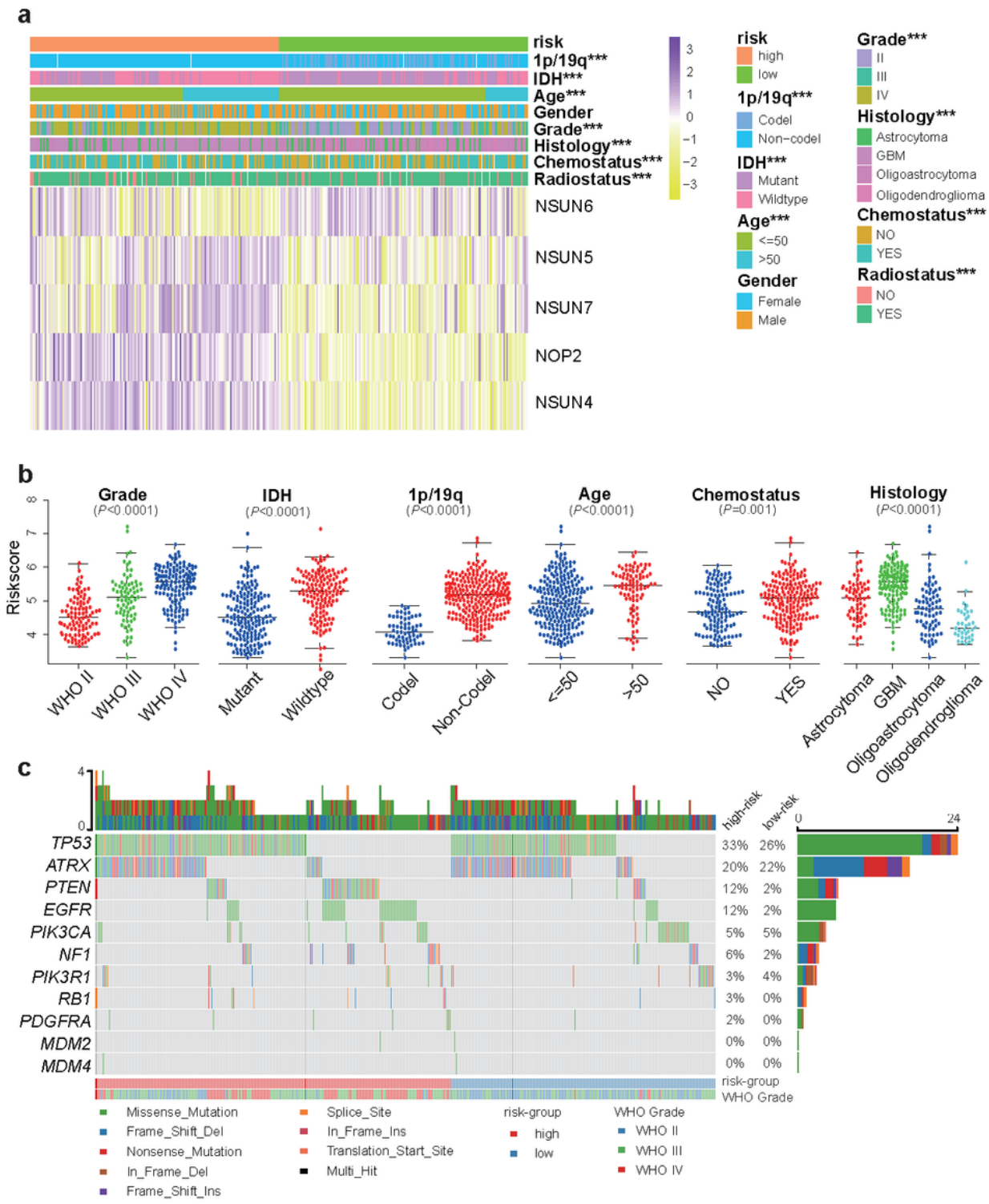

Figure 9

The prognostic value of the risk score model both in CGGA and TCGA dataset. a The patients with high-risk scores were correlated with a higher death rate and shorter survival time in gliomas. b Kaplan-Meier overall survival curves for patient data in the CGGA dataset. c The area under the curve (AUC) of ROC curves were 0.819, 0.793 and 0.746 in predicting 5-years, 3-years and 1-year OS events from the CGGA dataset respectively. d-f Validation of the risk score model in the TCGA dataset using the same analysis.

g- j Kaplan-Meier overall survival curves for patients in the CGGA dataset stratified by WHO grade and IDH-mutant status.

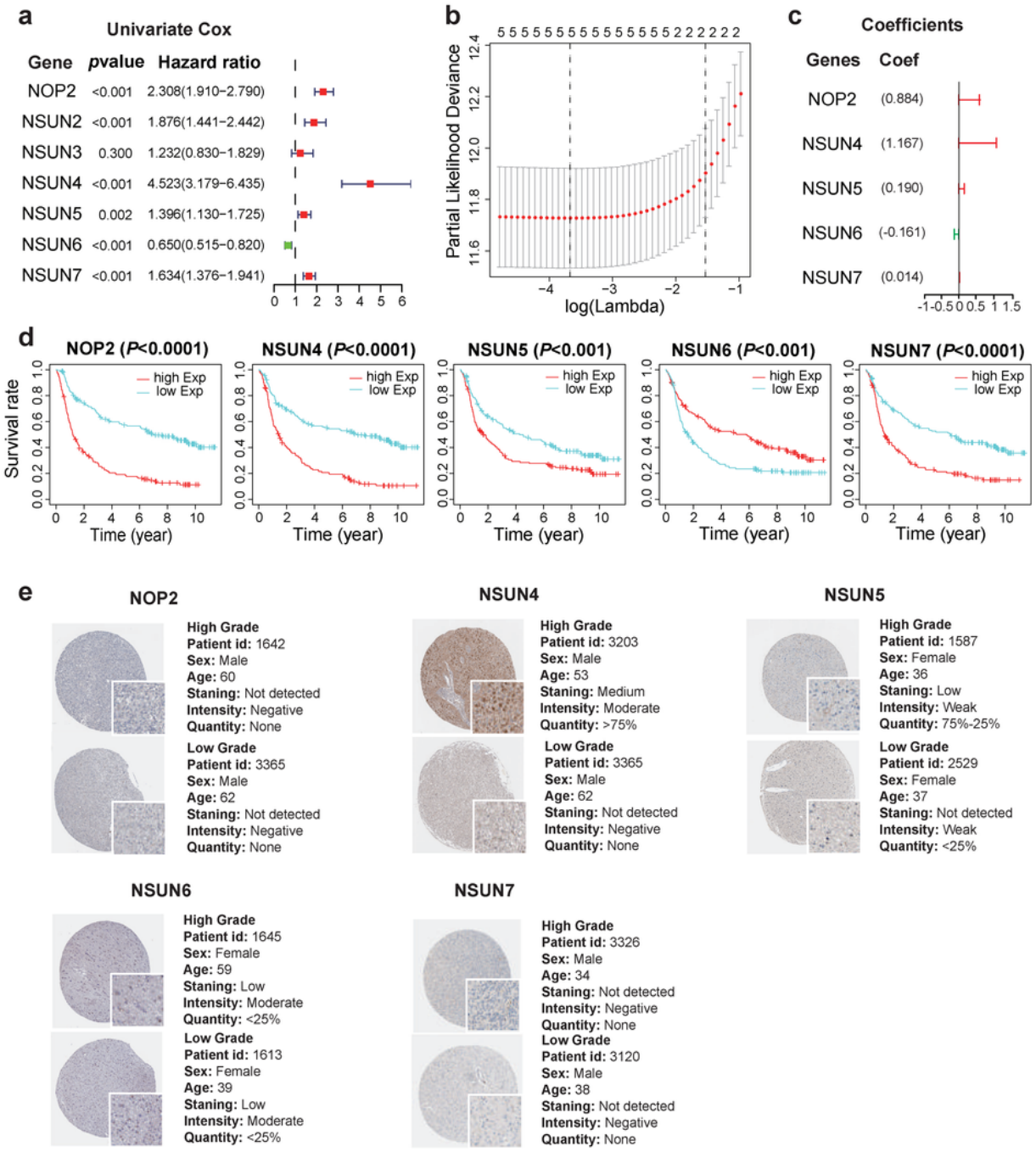

Figure 12

The interrelation of the risk scores and clinicopathological characteristics in gliomas, and the mutational landscape of tumor driver-gene alterations. a The relationship between five RNA:m5C methyltransferase expression profiles stratified by risk score and clinicopathological features of gliomas. b The relationship

between the risk scores and each clinicopathological characteristic. c The mutational landscape of tumor driver-gene alterations between low- and high-risk patients with gliomas. Significance: P*<0.05, P**<0.01, P***<0.001.

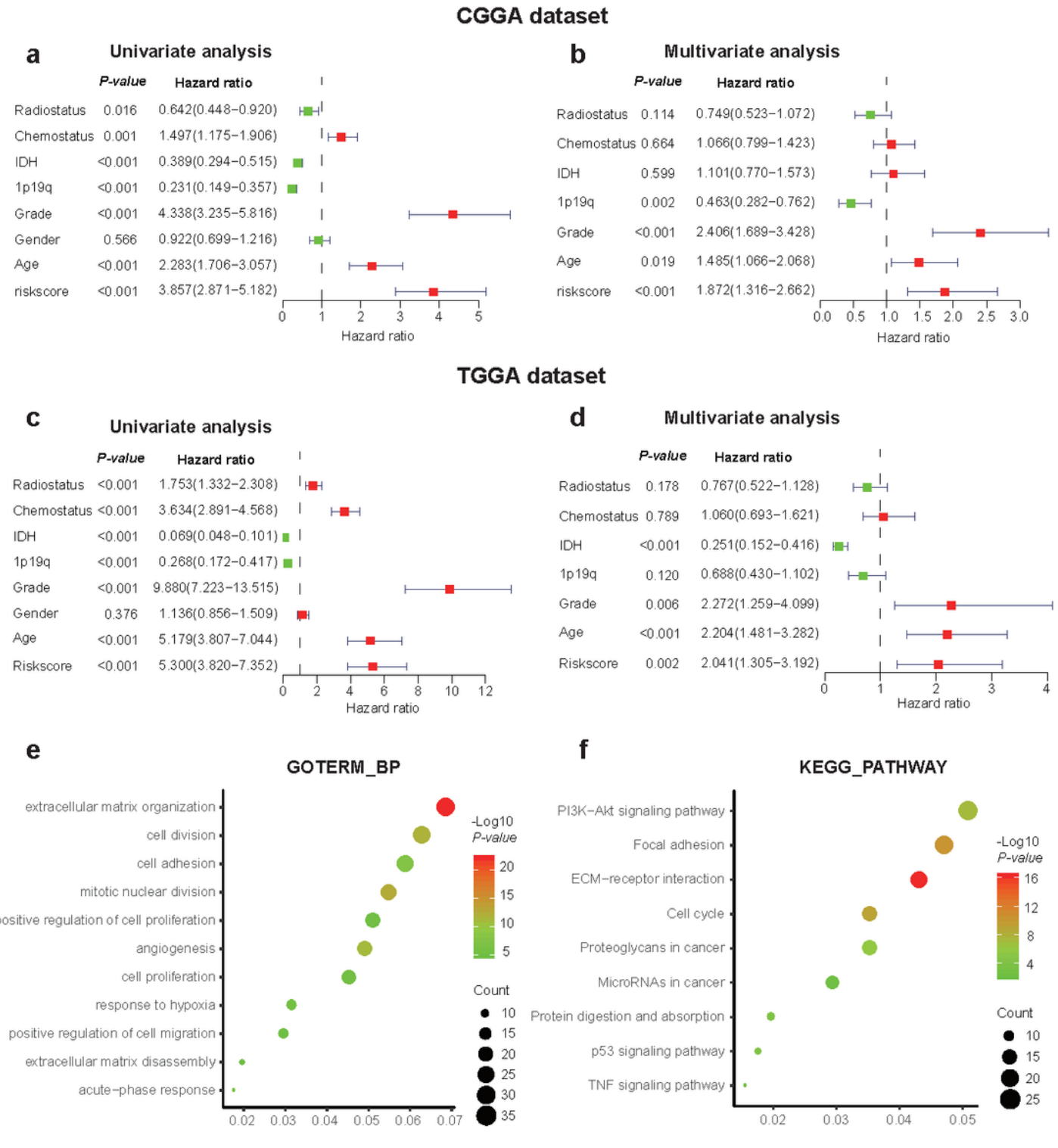

Figure 14

Univariate and multivariate Cox regression analyses of eight clinicopathological features and biological function analyses. a, b The risk score was an independent prognostic factor in the CGGA dataset. c, d The

risk score was an independent prognostic factor in the TGGA dataset. e, f GO and KEGG pathway analyses, respectively, of up-regulated genes in high-risk glioma patients.

Supplementary Files

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

- Additionalfile1.tif
- Additionalfile3.tif
- Additionalfile4.tif
- Additionalfile2.tif
- Additionalfile5.docx
- Additionalfile4.tif
- Additionalfile2.tif
- Additionalfile5.docx
- Additionalfile3.tif
- Additionalfile1.tif