



OPEN ACCESS

**Edited by:**

Saber Imani,  
Affiliated Hospital of Southwest  
Medical University, China

**Reviewed by:**

Yu-Chan Chang,  
National Yang-Ming University, Taiwan  
Ping Zheng,  
The University of Melbourne, Australia  
Jun Li,  
Huazhong University of Science and  
Technology, China  
Mazaher Maghsoudloo,  
University of Tehran, Iran

**\*Correspondence:**

Jinsheng Hong  
13799375732@163.com  
Qiuyu Zhang  
qiuyu.zhang@fjmu.edu.cn  
Qiuyuan Yue  
circlesoo@sina.cn  
Fangqin Xue  
xuefangqingsl@sina.com

†These authors have contributed  
equally to this work

**Specialty section:**

This article was submitted to  
Molecular Diagnostics and  
Therapeutics,  
a section of the journal  
Frontiers in Molecular Biosciences

**Received:** 16 September 2021

**Accepted:** 19 November 2021

**Published:** 06 December 2021

**Citation:**

Zhang M, Chen H, Liang B, Wang X,  
Gu N, Xue F, Yue Q, Zhang Q and  
Hong J (2021) Prognostic Value of  
mRNAsi/Corrected mRNAsi  
Calculated by the One-Class Logistic  
Regression Machine-Learning  
Algorithm in Glioblastoma Within  
Multiple Datasets.  
Front. Mol. Biosci. 8:777921.  
doi: 10.3389/fmolb.2021.777921

# Prognostic Value of mRNAsi/ Corrected mRNAsi Calculated by the One-Class Logistic Regression Machine-Learning Algorithm in Glioblastoma Within Multiple Datasets

Mingwei Zhang<sup>1,2,3†</sup>, Hong Chen<sup>4†</sup>, Bo Liang<sup>5†</sup>, Xuezhen Wang<sup>1†</sup>, Ning Gu<sup>6</sup>, Fangqin Xue<sup>4\*</sup>,  
Qiuyuan Yue<sup>7\*</sup>, Qiuyu Zhang<sup>2\*</sup> and Jinsheng Hong<sup>1,3\*</sup>

<sup>1</sup>Department of Radiotherapy, Cancer Center, The First Affiliated Hospital of Fujian Medical University, Fuzhou, China, <sup>2</sup>Institute of Immunotherapy, Fujian Medical University, Fuzhou, China, <sup>3</sup>Key Laboratory of Radiation Biology of Fujian Higher Education Institutions, The First Affiliated Hospital, Fujian Medical University, Fuzhou, China, <sup>4</sup>Department of Gastrointestinal Surgery, Fujian Provincial Hospital, Fuzhou, China, <sup>5</sup>Nanjing University of Chinese Medicine, Nanjing, China, <sup>6</sup>Nanjing Hospital of Chinese Medicine Affiliated to Nanjing University of Chinese Medicine, Nanjing, China, <sup>7</sup>Department of Radiology, Fujian Cancer Hospital & Fujian Medical University Cancer Hospital, Fuzhou, China

Glioblastoma (GBM) is the most common glial tumour and has extremely poor prognosis. GBM stem-like cells drive tumorigenesis and progression. However, a systematic assessment of stemness indices and their association with immunological properties in GBM is lacking. We collected 874 GBM samples from four GBM cohorts (TCGA, CGGA, GSE4412, and GSE13041) and calculated the mRNA expression-based stemness indices (mRNAsi) and corrected mRNAsi (c\_mRNAsi, mRNAsi/tumour purity) with OCLR algorithm. Then, mRNAsi/c\_mRNAsi were used to quantify the stemness traits that correlated significantly with prognosis. Additionally, confounding variables were identified. We used discrimination, calibration, and model improvement capability to evaluate the established models. Finally, the *CIBERSORTx* algorithm and ssGSEA were implemented for functional analysis. Patients with high mRNAsi/c\_mRNAsi GBM showed better prognosis among the four GBM cohorts. After identifying the confounding variables, c\_mRNAsi still maintained its prognostic value. Model evaluation showed that the c\_mRNAsi-based model performed well. Patients with high c\_mRNAsi exhibited significant immune suppression. Moreover, c\_mRNAsi correlated negatively with infiltrating levels of immune-related cells. In addition, ssGSEA revealed that immune-related pathways were generally activated in patients with high c\_mRNAsi. We comprehensively evaluated GBM stemness indices based on large cohorts and established a c\_mRNAsi-based classifier for prognosis prediction.

**Keywords:** glioblastoma, mRNAsi, OCLR, prognosis, stemness indices

## 1 INTRODUCTION

Glioblastoma multiforme (GBM), is the most common and most malignant glial tumour (Young et al., 2015). There is no clear way to prevent GBM; the disease can be very difficult to treat, and a cure is often not possible. The typical treatment, which involves surgery, chemotherapy, and radiation therapy, (Gallego, 2015), may slow cancer progression and reduce signs and symptoms. However, cancer usually recurs despite treatment (Gallego, 2015). The most common length of survival following diagnosis is 12–15 months, with less than 3–5% of the patients surviving longer than 5 years (Gallego, 2015). Without treatment, the survival time is typically 3 months (McNeill, 2016). Therefore, developing and applying signatures or biomarkers that can effectively predict the prognosis of these patients is of vital importance. A good initial Karnofsky Performance Score (KPS), the methylation of the O6-methylguanine-DNA methyltransferase (*MGMT*) promoter, and mutations in isocitrate dehydrogenase 1 (*IDH1*) are associated with longer survival (Krex et al., 2007; Martinez et al., 2007; Burgenske et al., 2019; Chaddad et al., 2019). The above signatures or biomarkers can be used either alone or in combination to predict the prognosis of GBM (Molenaar et al., 2014). However, their predictive capacity is rather low and a new index is needed.

Stem-like cells, which are characterised by the self-renewal properties and therapeutic resistance, play crucial roles in various cancers, (Kaushal and Ramakrishna, 2020), especially in GBM (Wang et al., 2018). Although cancer stem-like cells are very important for prognosis in GBM, (Turaga et al., 2020), there are still some shortcomings and complications in quantifying these cells. The stemness features have been extensively studied using artificial intelligence and deep learning methods. (Pan et al., 2019). A good example is the calculation of the mRNA expression-based stemness index (mRNAsi) with the one-class logistic regression (OCLR) machine-learning algorithm (Sokolov et al., 2016; Malta et al., 2018). Tathiane M. Malta *et al.* used mRNAsi for the first time to reflect the degree of oncogenic dedifferentiation (Malta et al., 2018). They also found tumour heterogeneity at the single-cell level by measuring the mRNAsi and concluded that a lower mRNAsi correlated with better prognosis in various cancers. (Malta et al., 2018). The prognostic value of the mRNAsi differs among different cancers. Moreover, we have previously shown that the prediction performance of a single mRNAsi-based signature is not good in primary lower-grade glioma, (Zhang et al., 2020b), partly because the tissue biopsy samples are often mixed with non-tumour tissues (bulk tissues). This means that the expression data on which the mRNAsi is based may be contaminated with non-tumour information. Thus, tumour purity may solve this issue (Xia et al., 2020).

It remains unclear whether the mRNAsi is an independent prognostic indicator in GBM and whether the predictive capacity of mRNAsi is better than that of existing factors such as the mutational status of *IDH1* and the methylation status of *MGMT*. Previous studies have shown that the combination of clinical features with signatures or biomarkers can significantly improve prognosis prediction, (Zhang et al., 2020b; Zhang et al., 2020c), but this has not been verified with the mRNAsi, let alone the corrected mRNAsi (c\_mRNAsi), which is acquired using

‘Estimation of STromal and Immune cells in Malignant Tumours using Expression data’ (ESTIMATE) (Yoshihara et al., 2013) to calculate tumour purity. Whether c\_mRNAsi can predict GBM better than mRNAsi is unknown. Furthermore, although Tathiane M. Malta *et al.* (Malta et al., 2018) analysed cancer stemness quite extensively, this was done in almost 12,000 samples of 33 tumour types from only The Cancer Genome Atlas (TCGA) (Hoadley et al., 2014). Thus, overfitting was inevitable and the generalisation ability of the mRNAsi was not evaluated. Therefore, the prognostic value of the mRNAsi in GBM needs to be validated in other independent databases, such as the Chinese Glioma Genome Atlas (CGGA) and Gene Expression Omnibus (GEO) (Barrett et al., 2013).

In this study, we used mRNA expression data and the OCLR machine-learning algorithm to simultaneously examine the independent prognostic value of mRNAsi/c\_mRNAsi in TCGA, CGGA, and two GEO datasets. We compared mRNAsi/c\_mRNAsi directly and evaluated the model improvement ability. Then, we applied the latest *CIBERSORTx* tool (Newman et al., 2019) to evaluate the relationship between mRNAsi/c\_mRNAsi and immune cell infiltration and conducted single sample gene set enrichment analysis (ssGSEA) to comprehensively examine its prognostic value and relationship with the immune microenvironment.

## 2 MATERIALS AND METHODS

### 2.1 Data Acquisition

RNA-sequencing data (level 3) of 158 patients with GBM from TCGA and 279 patients with GBM from the CGGA were obtained. The data from TCGA were downloaded from the University of California Santa Cruz (UCSC) Xena website (<https://xena.ucsc.edu/>). Transcript abundances were measured in fragments per kilobase of transcript per million mapped reads (FPKM). We only included patients who had adequate clinical and pathological data. Then, to uncover the practicability and accuracy of independent prognostic factors for GBM, samples from the TCGA and CGGA cohorts were applied as training and validation cohorts, respectively. Moreover, we included two GEO datasets (GSE4412 (Freije et al., 2004) and GSE13041 (Lee et al., 2008)) with more than 100 samples and follow-up data as our external validation data. The characteristics of the patients from the databases are presented as means  $\pm$  standard deviations (continuous variables that satisfied the normal distribution), median, minimum, maximum and quartile (continuous variables that did not satisfy the normal distribution), and percentage (categorical variables), as appropriate.

### 2.2 mRNAsi/c\_mRNAsi Acquisition

The mRNAsi was calculated using the OCLR machine-learning algorithm (Malta et al., 2018). Tumour purity was evaluated with ESTIMATE (Yoshihara et al., 2013) and c\_mRNAsi was obtained by dividing the mRNAsi by tumour purity (Zhang et al., 2020b). The gene expression-based mRNAsi/c\_mRNAsi was represented using  $\beta$  values ranging from zero (no gene expression) to one (complete gene expression).

## 2.3 Analysis of Independent Prognostic Factors

### 2.3.1 The Relationship Between mRNAsi/c\_mRNAsi and Overall Survival (OS)

To explore the effect of mRNAsi/c\_mRNAsi on OS of patients with GBM, we used locally weighted scatterplot smoothing (Lowess) algorithm to flexibly evaluate the association of mRNAsi/c\_mRNAsi with OS. The results were obtained as a fitting smooth curve. When the curve was linear, mRNAsi/c\_mRNAsi was included as a continuous variable; otherwise, mRNAsi/c\_mRNAsi was included as a dichotomous variable in the subsequent analysis.

### 2.3.2 Survival Analysis

When the variables were analysed as dichotomous variables, the optimal cut-off for each index with the associated hazard of OS was identified by log-rank statistics in a *survfit* model, using the *cutp* function of the *survMisc* package. Then, patients with GBM were included into either the high or the low group according to the optimal cut-off. Next, Kaplan-Meier analysis with log-rank test was conducted to estimate the survival curves of each group and to compare the prognosis between different groups, by using the *survival* package.

### 2.3.3 Identification of Confounding Variables

Residual confounding variables refer to incomplete adjustment for factors related to both exposure and outcome (Kernan et al., 2000). The confounding variables that may influence the OS of patients with GBM need to be identified. To estimate the magnitude of the effect of mRNAsi/c\_mRNAsi on GBM, we used the Cox proportional hazards model. The regression coefficient changed more than 10% when the adjustment variables were included or not included or when those with  $p < 0.1$  in the univariate analysis with OS were considered as confounding variables to be adjusted (adjusted I/II model) (Kernan et al., 2000). The common covariates in TCGA were age, gender, IDH, radiotherapy, chemotherapy, and subtype. In addition, 1p19q and MGMTp were common covariates in CGGA but without subtype. Afterward, an interaction test and a stratified analysis (Soria et al., 2015) of the association between mRNAsi/c\_mRNAsi and OS were conducted in both the non-adjusted model and adjusted I model (identified confounders). A two-tailed  $p < 0.05$  was considered statistically significant. Empower ([www.empowerstats.com](http://www.empowerstats.com); X&Y solutions Inc., Boston, MA) and R (<http://www.R-project.org>) were used for the abovementioned statistical analyses.

## 2.4 Construction and Comparison of Prognostic Models

### 2.4.1 Model Establishment

After confirming the effect of mRNAsi/c\_mRNAsi on OS, we further evaluated and compared the benefit of five different models, including the prognostic model constructed by well-established clinical factors (model 1), model 1 integrated with mRNAsi (model 2)/c\_mRNAsi (model 3), and single mRNAsi (model 4)/c\_mRNAsi (model 5).

### 2.4.2 Model Evaluation and Nomogram

We used discrimination, calibration, and model improvement capability to assess the performance of the different models. Discrimination was evaluated through the receiver operating characteristic (ROC) curve, (Zhou et al., 2019), concordance index (C-index) (Harrell et al., 1996) and the prediction error and decision curve analysis (DCA) curves (Kerr et al., 2016). Notably, the enhanced bootstrap method with 500 resamples was used for internal validation (Wang et al., 2019). Discrimination and calibration were evaluated by apparent and adjusted C-index and Brier Score. Finally, model improvement capability was evaluated by applying net reclassification improvement (NRI) and integrated discrimination improvement (IDI) using the *survIDINRI* package (Pencina et al., 2008). After the best model was identified, the *regplot* package was employed to construct the nomogram.

### 2.4.3 External Validation

We applied the data from CGGA and GEO as external validation. In CGGA, as described above, we performed mRNAsi/c\_mRNAsi acquisition, independent prognostic factors analysis, and prognostic model construction and comparison. It should be noted that because the clinical information in TCGA and CGGA was not identical, common covariates were not the same. GSE4412 (Freije et al., 2004) and GSE13041, (Lee et al., 2008), which constitute GEO, were also applied for the external effect validation of mRNAsi/c\_mRNAsi on OS. Similarly, patients were divided into the high or low group based on the optimal cut-off, which was previously calculated using the same package. Kaplan-Meier analysis was employed to assess the two groups with the log-rank test. Afterwards, ROC analysis of time-independent outcomes was also performed.

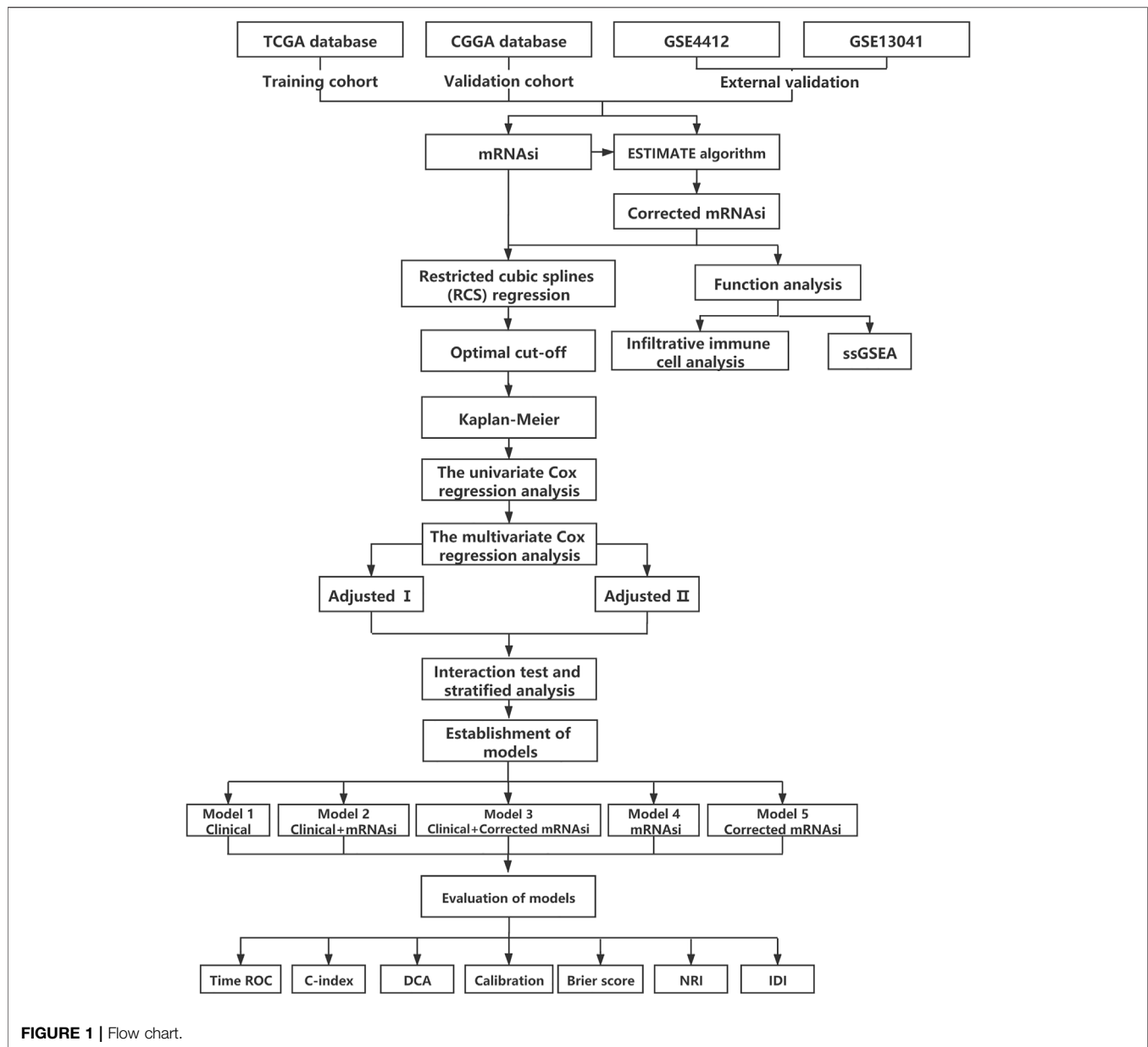
## 2.5 Function Analysis

### 2.5.1 Infiltrative Immune Cell Analysis

To characterise the abundance of 22 infiltrative immune cell types based on the expression matrix data of patients with GBM, the *CIBERSORTx* web tool (<https://cibersortx.stanford.edu/>) was applied (Newman et al., 2019). This tool uses batch correction to adjust the gene expression profile of the bulk of cells (mixture data) to eliminate possible cross-platform variations between the mixture data and the gene expression data of single cells (signature matrix) (Le et al., 2020). After enabling batch correction, performing the Bulk mode, and selecting the quantile normalisation algorithm, the absolute score for the proportion of 22 immune cell subsets in GBM samples was calculated. The samples with  $p < 0.05$  were enrolled for further analysis because of the high reliability of the inferred results (Ali et al., 2016). Wilcoxon rank-sum test was used to compare the differences in the proportion of the 22 infiltrative immune cell subtypes between the high and low groups. The Spearman correlation test was used to further explore the correlation of the two indexes with immune cell types.

### 2.5.2 Single Sample Gene Set Enrichment Analysis (ssGSEA)

The ssGSEA method, (Barbie et al., 2009), which is a modification of GSEA, (Subramanian et al., 2005), was



**FIGURE 1** | Flow chart.

developed to obtain an enrichment score for a single sample instead of two groups of samples. Here, the ssGSEA was used to compare differentially enriched hallmarks of cancer gene sets (Barbie et al., 2009). To identify key pathways in different groups, we chose to focus on 50 hallmark gene sets, which were designed to highlight gene sets contained in the Molecular Signatures Database (MSigDB), (Liberzon et al., 2015), one of the most widely used and comprehensive databases of gene sets for performing gene set enrichment analysis. The hallmarks of the gene sets effectively summarise most of the relevant information of the original founder sets and, by reducing both variation and redundancy, provide more refined and concise inputs for gene set enrichment analysis (Liberzon et al., 2015). Gene symbol profiles for *Homo sapiens* were

downloaded from the MSigDB. Then, the degree of association between each hallmark's ssGSEA profile was estimated using the *gsva* package. Next, differential analysis was performed with the *limma* package under the threshold of the absolute value of  $t > 1$  and adjusted  $p$  value  $< 0.05$ .

## 3 RESULTS

### 3.1 Patient Characteristics

An overview of the stemness indices-related signature development and validation workflow is presented in **Figure 1**. A total of 874 GBM samples (158 from TCGA as the training cohort, and 279 from CGGA and 437 from GEO as the validation



**TABLE 1** | Patient characteristics.

Character	Training cohort	External validation cohort	External validation GEO cohort	External validation GEO cohort
	TCGA ( <i>n</i> = 158)	CGGA ( <i>n</i> = 279)	GSE4412 ( <i>n</i> = 170)	GSE13041 ( <i>n</i> = 267)
Age	59.6 (13.6)	48.00 (39.5–57.0)	42.0 (33.0–54.0)	53.71 (13.8)
mRNAsi	0.56 (0.22)	0.50 (0.17)	0.60 (0.23)	0.42 (0.33–0.52)
c_mRNAsi	0.74 (0.25)	0.57 (0.16)	0.71 (0.25)	0.51 (0.15)
Male	102 (64.56%)	165 (59.14%)	64 (37.65%)	151 (63.18%)
IDH				
Wild type	148 (93.67%)	211 (75.63%)	NA	NA
Mutation	10 (6.33%)	68 (24.37%)	NA	NA
Radiotherapy				
No	29 (18.35%)	53 (19%)	NA	NA
Yes	129 (81.65%)	226 (81%)	NA	NA
Chemotherapy				
No	45 (28.48%)	49 (17.56%)	NA	NA
Yes	113 (71.52%)	230 (82.44%)	NA	NA

Data are presented as median (interquartile range) or *N* (%). TCGA, the cancer genome atlas; CGGA, chinese glioma genome atlas; GEO, gene expression omnibus; IDH, isocitrate dehydrogenase; NA, not applicable.

cohort) were obtained in our study. The patient characteristics are presented in **Table 1**.

## 3.2 mRNAsi/c\_mRNAsi Acts as an Independent Prognostic Factor

### 3.2.1 Patients With High mRNAsi/c\_mRNAsi GBM had a Better Prognosis

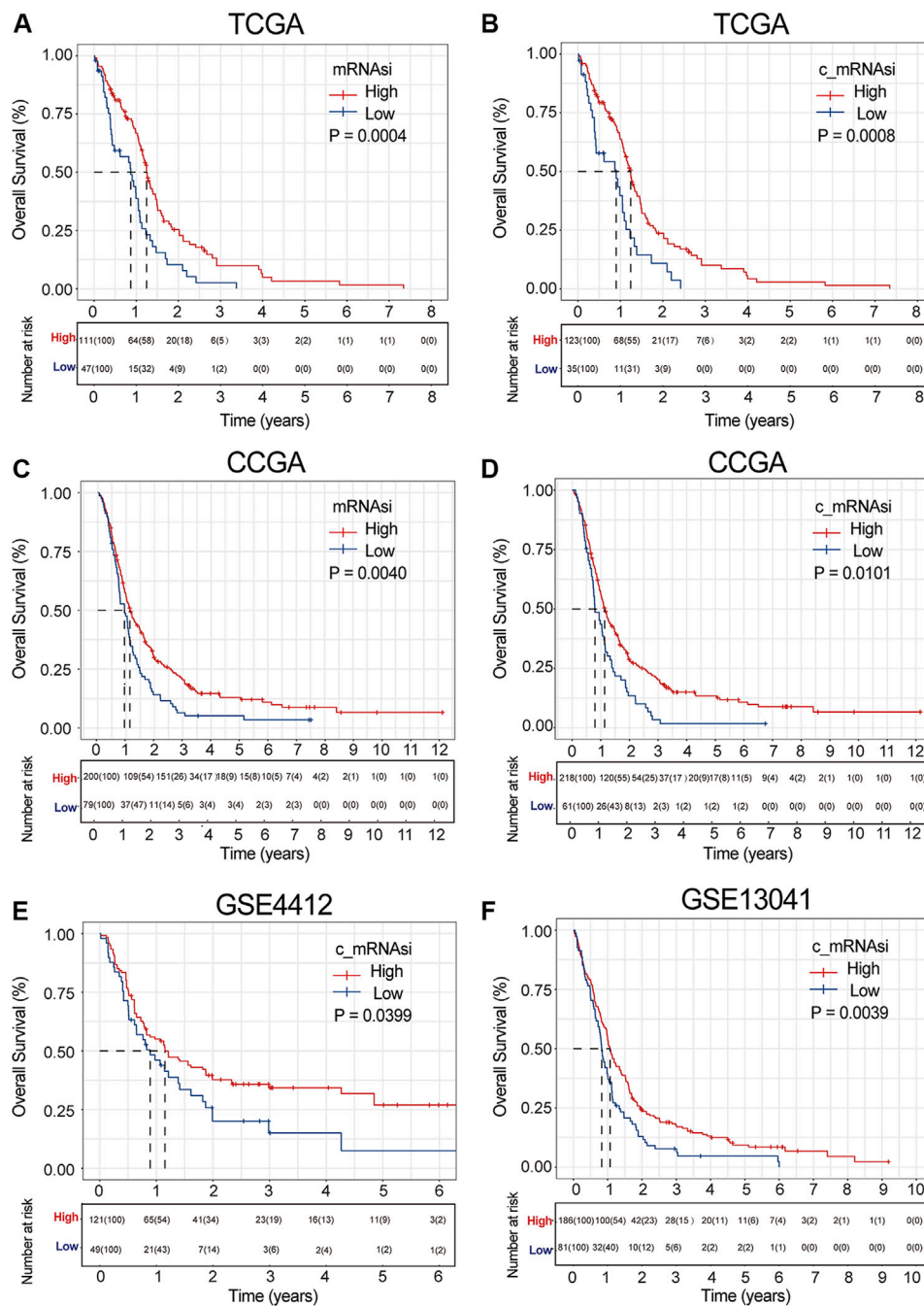
In the TCGA dataset, the relation between mRNAsi/c\_mRNAsi and OS was nonlinear. Therefore, they were considered dichotomous variables in subsequent analysis (**Supplementary Figures S1A,B**). A total of 158 samples were clustered into the high- (*n* = 111) or low- (*n* = 47) mRNAsi group based on the optimal cut-off value identified by *survMisc* package (**Supplementary Figure S2A**). Patients in the high-mRNAsi group had better OS than those in the low-mRNAsi group ( $p = 0.0003$ ) (**Figure 2A**). Similarly, 158 patients were clustered into the high- (*n* = 123) or low- (*n* = 35) c\_mRNAsi group based on the optimal cut-off value identified by the same package (**Supplementary Figure S2B**). Patients in the high-c\_mRNAsi group had better OS than those in the low-c\_mRNAsi group ( $p = 0.0008$ ) (**Figure 2B**). Moreover, we explored the relationship between mRNAsi/c\_mRNAsi and disease-specific survival/progression-free interval in TCGA, and found that the trends for disease-specific survival ( $p = 0.0028$ ) (**Supplementary Figures S3A,B**) and progression-free interval ( $p < 0.0001$ ) (**Supplementary Figures S3C,D**) were similar to that for OS.

To determine whether the mRNAsi/c\_mRNAsi-associated prognostic signature had a similar prognostic value in different populations, its prediction performance was validated externally in CGGA and GEO. Similarly, we considered mRNAsi/c\_mRNAsi as a dichotomous variable in CGGA according to the Lowess result (**Supplementary Figures S1C,D**). All samples in CGGA and GEO were clustered into the high- or low-mRNAsi/c\_mRNAsi group based on the optimal cut-off value identified by the same package (**Supplementary Figure S2C**). Consistent with the findings in TCGA, the Kalan-Meier curve in

CGGA revealed that patients in the high-mRNAsi/c\_mRNAsi group had better OS than those in the low-mRNAsi/c\_mRNAsi group ( $p = 0.0040$  and  $0.0011$ , respectively) (**Figures 2C,D**). GSE4412 has the transcriptional profiling of 170 GBM samples from 74 patients (Freije et al., 2004). A total of 170 GBM samples were divided into the high- (*n* = 121) or low- (*n* = 49) c\_mRNAsi group based on the optimal cut-off value (**Supplementary Figure S2D**), and the high-c\_mRNAsi group had better OS ( $p = 0.0400$ ) (**Figure 2E**). GSE13041 has 267 GBM samples from 239 patients, (Lee et al., 2008), which were divided into the high- (*n* = 186) or low- (*n* = 81) c\_mRNAsi group based on the optimal cut-off value (**Supplementary Figure S2**). The high-c\_mRNAsi group also had better OS ( $p = 0.0040$ ) (**Figure 2F**).

### 3.2.2 Identification of Confounding Variables

Given the possible interference of confounding variables, we carried out confounders identification and then adjusted for these potential confounding factors. In TCGA, we found that mRNAsi had to be adjusted for age through univariate analysis (**Figure 3A**). These covariates combined with common covariates (age, gender, IDH, radiotherapy, chemotherapy, and subtype) were enrolled into the adjusted II model. In the adjusted I model, after adjusting for confounders (age and IDH), mRNAsi was still associated with OS (hazard ratio (HR) = 0.561, 95% confidence interval (CI) 0.383–0.823,  $p = 0.003$ ) (**Figure 3B**). Furthermore, after adjusting for predominant clinical and prognostic factors (age, gender, IDH, radiotherapy, chemotherapy, and subtype) in the adjusted II model, mRNAsi independently predicted prognosis in TCGA (HR = 0.552, 95% CI 0.370–0.823,  $p = 0.004$ ) (**Figure 3C**). The interaction analysis revealed that gender played an interactive role in the association between mRNAsi and OS (**Supplementary Table S1**). Male patients had higher HRs between mRNAsi and OS (HR = 0.92; 95% CI, 0.11–7.59) than females (HR = 0.32; 95% CI, 0.17–0.61). In the same way, we found that only age should be adjusted on c\_mRNAsi through univariate analysis (**Figure 3D**), and this covariate combined

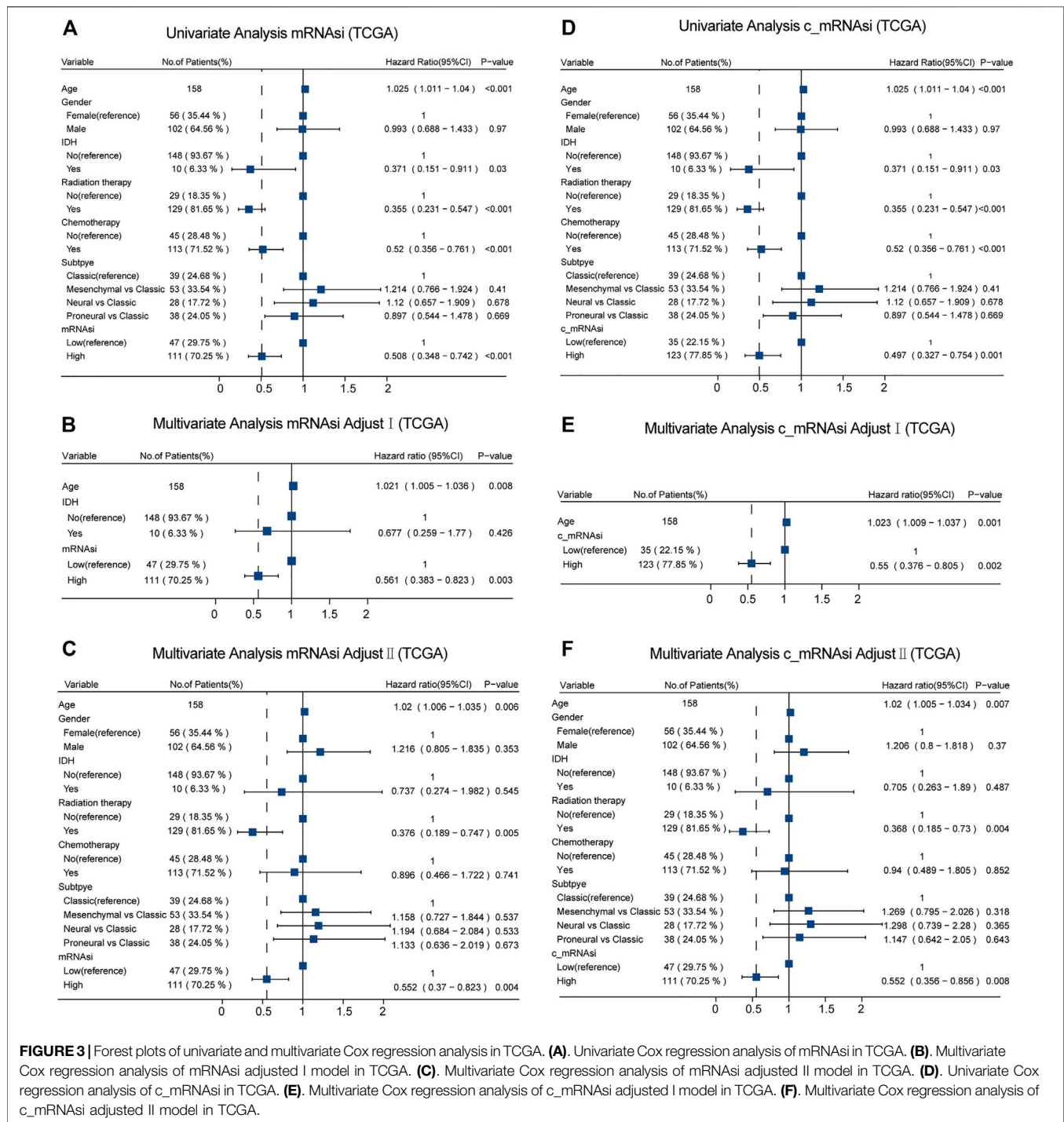


**FIGURE 2 |** Survival curve of mRNAsi/c\_mRNAsi on prognosis. **(A)**. Overall survival curve of mRNAsi in TCGA. **(B)**. Overall survival curve of c\_mRNAsi in TCGA. **(C)**. Overall survival curve of mRNAsi in CCGA. **(D)**. Overall survival curve of c\_mRNAsi in CCGA. **(E)**. Overall survival curve of c\_mRNAsi in GSE4412. **(F)**. Overall survival curve of c\_mRNAsi in GSE13041.

with common covariates were enrolled in the adjusted II model. In the adjusted I model, after adjusting for the confounder (age), c\_mRNAsi was still associated with OS (HR = 0.550, 95% CI 0.376–0.805,  $p = 0.002$ ) (Figure 3E). Furthermore, after adjusting for predominant clinical and prognostic factors in the adjusted II model, c\_mRNAsi independently predicted prognosis in TCGA (HR = 0.552,

95% CI 0.356–0.856,  $p = 0.008$ ) (Figure 3F). The effect of mRNAsi/c\_mRNAsi on OS was consistent across subgroups (Supplementary Table S1). Ultimately, mRNAsi/c\_mRNAsi was an independent prognostic factor for OS in patients with GBM.

Moreover, in CCGA, we identified different confounders (IDH, chemotherapy, and 1p19q on mRNAsi, as well as IDH



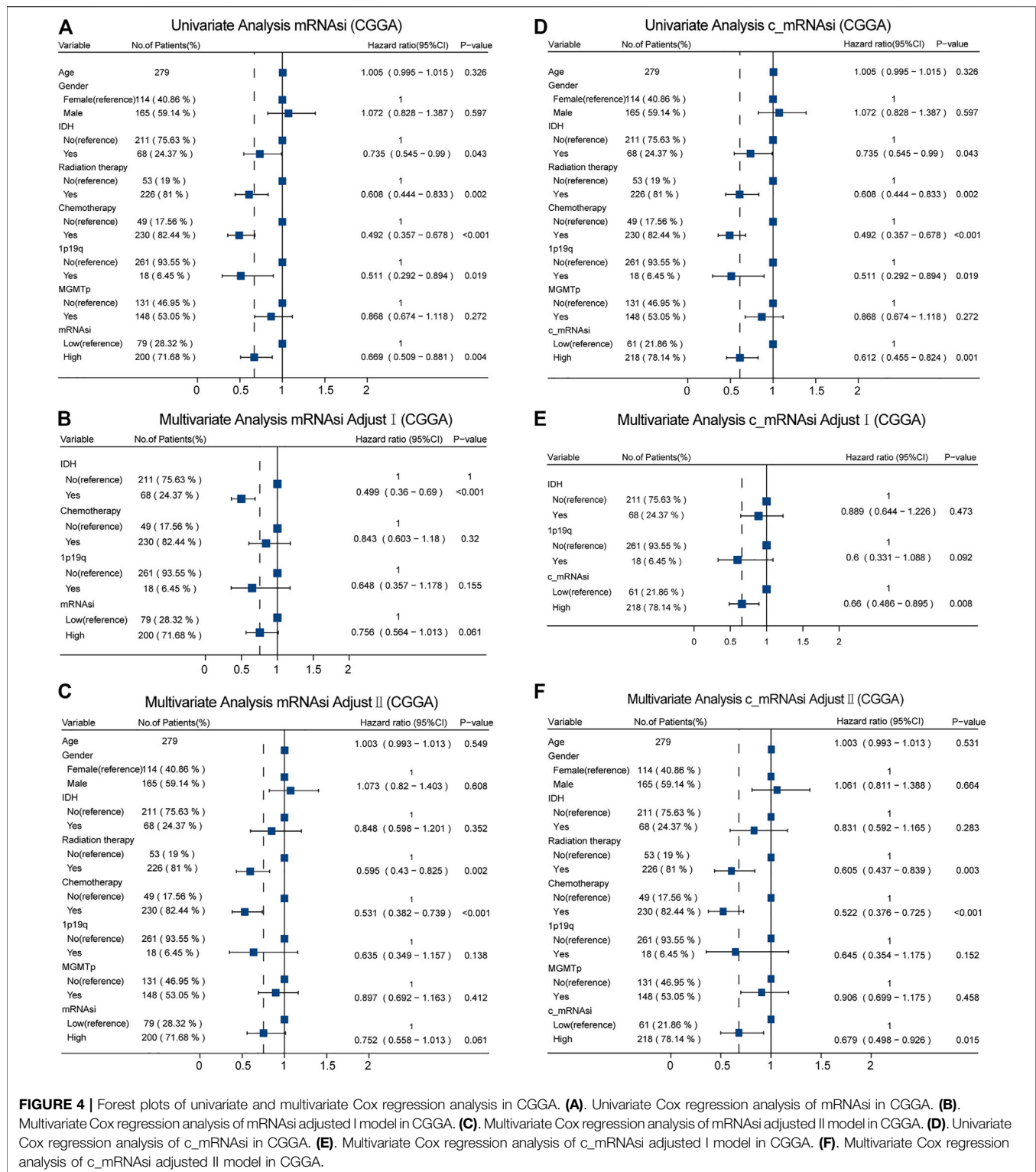
**FIGURE 3** | Forest plots of univariate and multivariate Cox regression analysis in TCGA. **(A)**. Univariate Cox regression analysis of mRNAsi in TCGA. **(B)**. Multivariate Cox regression analysis of mRNAsi adjusted I model in TCGA. **(C)**. Multivariate Cox regression analysis of mRNAsi adjusted II model in TCGA. **(D)**. Univariate Cox regression analysis of c\_mRNAsi in TCGA. **(E)**. Multivariate Cox regression analysis of c\_mRNAsi adjusted I model in TCGA. **(F)**. Multivariate Cox regression analysis of c\_mRNAsi adjusted II model in TCGA.

and 1p19q on c\_mRNAsi) that had to be adjusted through univariate analysis (Figures 4A,D), and these confounders (adjusted I model, Figures 4B,E) combined with common covariates (age, gender, IDH, radiotherapy, chemotherapy, 1p19q, and MGMTp) were enrolled in the adjusted II model (Figures 4C,F). We found that only c\_mRNAsi was an independent prognostic signature in patients with GBM in both the adjusted I and adjusted II models ( $p = 0.008, 0.015,$

respectively) (Figures 4E,F) and across stratified analyses (Supplementary Table S2).

### 3.3 Construction and Evaluation of Prognostic Models

We used discrimination, calibration, and model improvement capability to evaluate five established models. Models 2 and 3 had

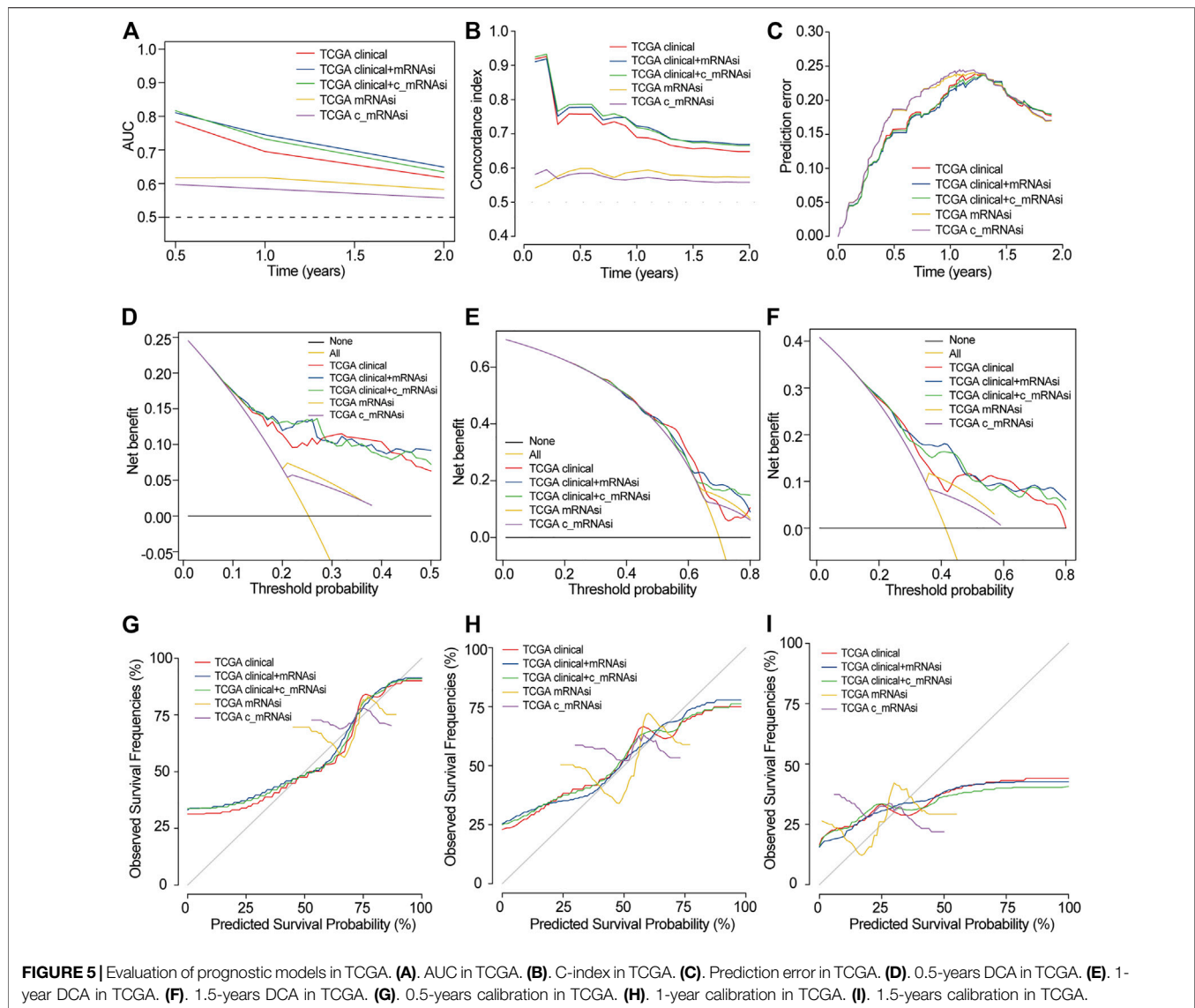


**FIGURE 4 |** Forest plots of univariate and multivariate Cox regression analysis in CGGA. **(A)** Univariate Cox regression analysis of mRNAsi in CGGA. **(B)** Multivariate Cox regression analysis of mRNAsi adjusted I model in CGGA. **(C)** Multivariate Cox regression analysis of mRNAsi adjusted II model in CGGA. **(D)** Univariate Cox regression analysis of c\_mRNAsi in CGGA. **(E)** Multivariate Cox regression analysis of c\_mRNAsi adjusted I model in CGGA. **(F)** Multivariate Cox regression analysis of c\_mRNAsi adjusted II model in CGGA.

a higher area under the curve (AUC), better C-index, and lower prediction error than the other models (Figures 5A–C). The apparent and adjusted C-index as well as the Brier scores in years 0.5-, 1-, and 1.5-years indicated that models 2 and 3 were better than the others (Supplementary Table S3). DCA showed that the

net benefit of models 2 and 3 in years 0.5 and 1 years was better than that of other models, but there was no significant difference in year 1.5 (Figures 5D–F). We found that the calibration of models 2 and 3 was better than that of other models in 0.5 and 1 years, while the calibration of the five models was poor in



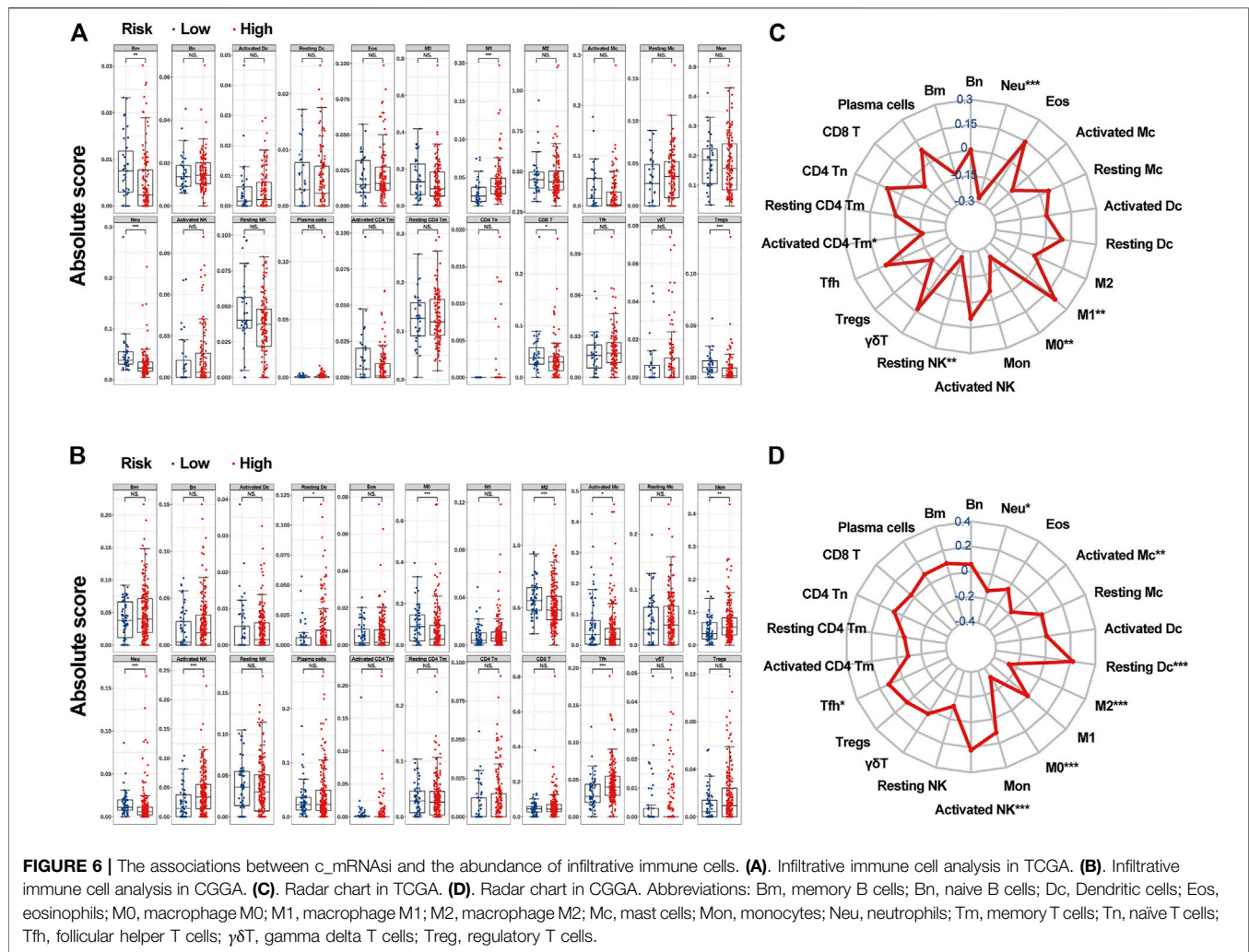


1.5 years (Figures 5G–I). As for model improvement capability, when model 1 was considered as the reference, the NRI and IDI of models 2 and 3 were both positive. In contrast, the NRI and IDI of models 4 and 5 were both negative, although there were no significant statistical differences (Supplementary Table S4). From the above results, we determined that models 2 and 3 had good discrimination and calibration in the OS prediction of patients with GBM.

In CGGA, we also built five models that were similar to those in TCGA. In TCGA, only 24 patients were followed up for more than 2 years, and the prognosis was very poor. Since the median follow-up time was only 12 months, the selected time points were 0.5, 1 and 1.5 years. In CGGA, since the 62 cases were followed up for more than 2 years, the time point was extended to 3 and 5 years. Here, models 2 and 3 also had a higher AUC and better C-index but had the same prediction error as the other models (Supplementary Figures S4A–C). DCA showed that the net benefit of model 2 in 0.5, 1, 1.5, and 3 years was higher than

that of other models, but there was no significant difference in 5 years (Supplementary Figures S4D–H). The calibration of models 2 and 3 was better than that of other models in 0.5 and 1 years, while the calibration of the five models was poor in 1.5, 3, and 5 years (Supplementary Figures S4I–M). The Brier scores indicated that models 2 and 3 were better than the others (Supplementary Table S5). As for model improvement capability, when model 1 was considered as the reference, the NRI and IDI of models 2 and 3 were both positive. In contrast, the NRI and IDI of models 4 and 5 were both negative (Supplementary Table S6). Because of the limited clinical data of GEO, we only compared models 4 and 5. The AUCs were 0.536–0.618, 0.544–0.589 in GSE4412 and GSE13041, respectively (Supplementary Figures S4N,O).

Combined with the above results, we found that mRNAsi/c\_mRNAsi is an independent prognostic factor in TCGA, but only c\_mRNAsi is an independent prognostic factor in CGGA. In TCGA, the comparison of the five models revealed that models 2



and 3 were the best, and there was little difference between these two models. In CGGA, model 3 performed the best among the five models. In GEO, there was no significant difference between the single *mRNAsi* and *c\_mRNAsi* models. Therefore, we finally decided to adopt model 3 (clinical factors integrated with the *c\_mRNAsi*) to predict OS and construct a nomogram in TCGA (Supplementary Figure S4). According to the nomogram, a representative patient with the total point of 286, the 0.5-years, 1-year, and 1.5-years survival rates were 82.6, 68.9, and 40.4%, respectively (Supplementary Figure S5).

### 3.4 Functional Analysis

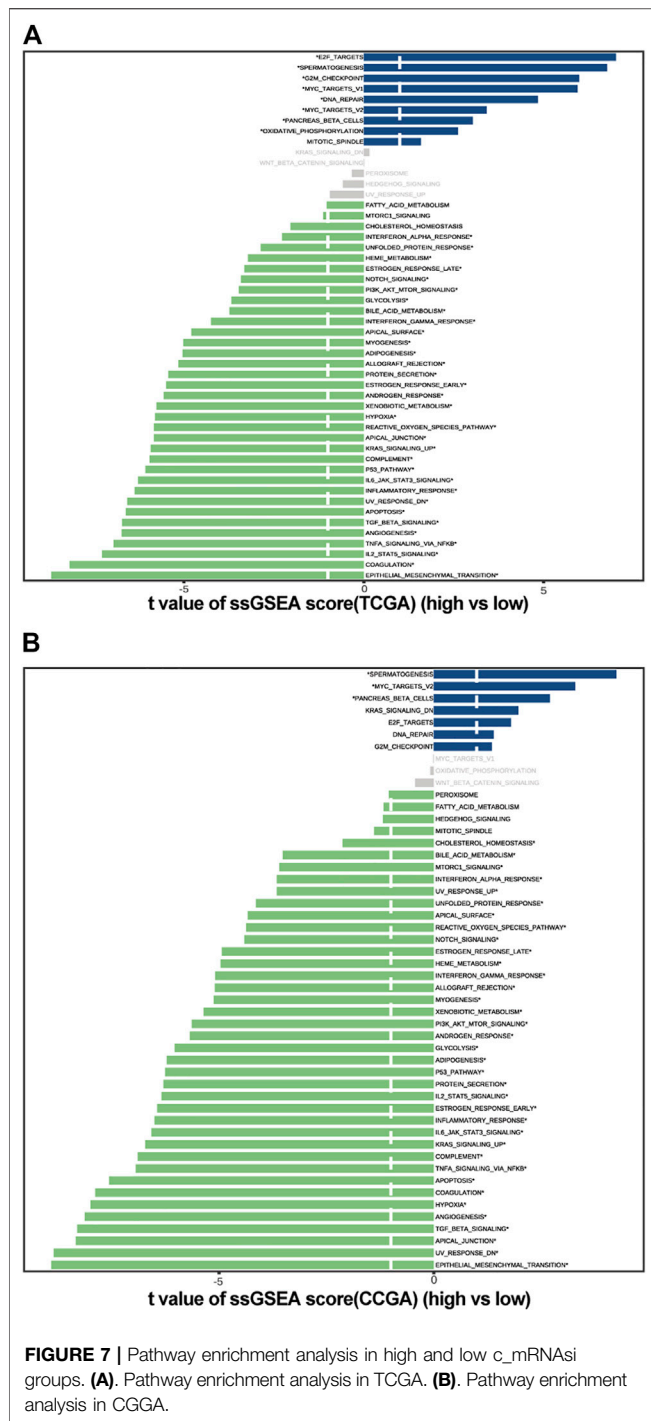
#### 3.4.1 Differential Abundance of Infiltrative Immune Cells

By applying the *CIBERSORTx* algorithm, the relative proportions of 22 immune cell subsets in GBM were acquired. A total of 158 patients with GBM from TCGA and 279 patients with GBM from the CGGA were enrolled for further analysis. In the TCGA dataset, the infiltration level of M1 macrophages was significantly higher in the high-*c\_mRNAsi* group, whereas the infiltration levels of memory B cells, neutrophils, CD8<sup>+</sup> T cells,

and regulatory T cells were significantly higher in the low-*c\_mRNAsi* group (Figure 6A). In CGGA, the infiltration levels of resting dendritic cells, monocytes, activated NK cells, and follicular T helper cells were significantly higher in the high-*c\_mRNAsi* group, whereas macrophages (M0 and M2), activated mast cells, and neutrophils were significantly higher in the low-*c\_mRNAsi* group (Figure 6B). Radar chart indicated that in the TCGA dataset, *c\_mRNAsi* was positively correlated with M1 macrophages and negatively correlated with neutrophils, M0 macrophages, resting NK cells, and activated memory CD4<sup>+</sup> T cells in the training cohort (Figure 6C). In the CGGA dataset, *c\_mRNAsi* was positively correlated with resting dendritic cells, activated NK cells, and follicular T helper cells, and negatively correlated with neutrophils, activated mast cells, and macrophages (M0 and M2) (Figure 6D).

#### 3.4.2 Pathway Enrichment Analysis

The ssGSEA was used to estimate the degree of enrichment of the MSigDB hallmark gene set in individual samples from the high- and low-*c\_mRNAsi* groups of both TCGA and CGGA. This allowed us to identify signalling pathways involved in GBM and



to estimate their degree of association with each group (high versus low). The results indicated that spermatogenesis, MYC targets v2, and pancreas beta cell pathways were involved significantly in the low-*c\_mRNAsi* group of both TCGA and CGGA, whereas unfolded protein response, haem metabolism, early and late oestrogen response, NOTCH signalling, glycolysis, bile acid metabolism, interferon  $\alpha/\gamma$  response, apical surface, myogenesis, adipogenesis, allograft rejection, androgen response, xenobiotic metabolism, hypoxia, reactive oxygen

species pathway, apical junction, KRAS signalling, complement, P53, IL6/JAK/STAT3 signalling, inflammatory response, UV response DN, apoptosis, TGF- $\beta$  signalling, angiogenesis, TNF $\alpha$  signalling via NF- $\kappa$ B, IL-2/STAT5 signalling, coagulation, and epithelial mesenchymal transition pathways were involved significantly in the high-*c\_mRNAsi* group of both TCGA and CGGA (Figures 7A,B).

## 4 DISCUSSION

GBM is composed of non-homogeneous cell populations exhibiting varying degrees of genetic and functional heterogeneity. Cancer stem cells are capable of sustaining tumours by manipulating genetic and non-genetic factors to metastasise, resist treatment, and maintain the tumour microenvironment (Saygin et al., 2019). Understanding the key traits and mechanisms of stemness of cancer stem-like cells provides opportunities to improve patient outcomes via improved prognostic models and therapeutics. However, tumour cells are usually comprised of a heterogeneous mixture of subclones, each of which may have its own distinct characteristics. Therefore, accurately assessing the make-up of the different cell states within a tumour biopsy is very important. Here, we calculated an mRNAsi, which was also corrected by tumour purity (*c\_mRNAsi*), based on the expression profile of 12,953 genes in 874 GBM samples from the TCGA, CGGA, and GEO public databases using the OCLR machine-learning algorithm. We found that, after confounding variable identification and interaction and stratified analyses, *c\_mRNAsi* remained an independent prognostic factor in both TCGA and CGGA, whereas mRNAsi was affected by gender in TCGA and was no longer an independent prognostic factor after adjustment in CGGA. Model 2 (clinical factors integrated with mRNAsi) and model 3 (clinical factors integrated with *c\_mRNAsi*) showed better calibration and discrimination than clinical factors alone in both TCGA and CGGA. Moreover, model 3 performed better than model 2, although there was no significant difference between the single mRNAsi and *c\_mRNAsi* models in GEO. Therefore, we concluded that *c\_mRNAsi* can be used as a new index for the construction of algorithms that predict the prognosis of patients with GBM. To explore the possible reasons for the difference in prognosis between the high- and low-*c\_mRNAsi* groups, we applied the *CIBERSORTx* algorithm to infer the abundance of immune infiltrating cells in TCGA and CGGA and found differential infiltration patterns across 5 and 8 clusters in TCGA and CGGA, respectively. Importantly, we found that high mRNAsi correlated significantly with high infiltration of immune activated cells, especially M1 macrophages, dendritic cells, monocytes, activated NK cells, and follicular T helper cells. Lastly, we screened the potential signalling pathways of the *c\_mRNAsi*-related signature and found that most of the pathways were immune-related.

In this study, we utilised the OCLR machine-learning algorithm to quantify mRNAsi and *c\_mRNAsi* for each GBM sample. Using OCLR, we previously identified undiscovered biological mechanisms associated with the dedifferentiated oncogenic state (Lian et al., 2019). Moreover, OCLR exhibited comparable performance with a more flexible and convenient



formulation to that with traditional support vector machine-based one-class predictors (Sokolov et al., 2016). We drew support from OCLR machine-learning algorithm to derive two distinct molecular metrics of stemness indices and finally selected c\_mRNAsi for subsequent validation analysis, owing to its observed prognostic significance in various data. Stemness indices had already been identified in several malignancies and had different prognostic values in ovarian cancer, (Kaipio et al., 2020), medulloblastoma, (Lian et al., 2019), colon cancer, (Tao et al., 2019), or acute myeloid leukaemia (Seneviratne et al., 2019). However, stemness indices are targeted at bulk tissues, a mixture of tumour tissue and normal tissue. Although some scholars developed new algorithms to adjust stemness indices, (Pan et al., 2019), the algorithms are complex. The ESTIMATE algorithm can adjust directly from transcriptome data, (Yoshihara et al., 2013), which is more convenient. In our multi-cohort screening, we examined the capacity of c\_mRNAsi, which is obtained after correcting the index by the tumour purity calculated with ESTIMATE, (Yoshihara et al., 2013), to predict OS. Our findings demonstrate that c\_mRNAsi can be implemented in clinical practice, something that has not been previously reported. Common clinical indicators of GBM include KPS, *MGMT*, *IDH1*, and epidermal growth factor receptor vIII (Burgenske et al., 2019; Chaddad et al., 2019). GBM-specific microRNAs, including miR-21 and miR-10b, have also been presented as biomarkers with promising prognostic values (Sasmita et al., 2018). Confounder identification and interaction tests could help us to better understand the relationship between the variables and disease. In our study, we first used the mRNAsi/c\_mRNAsi calculated by the algorithm as a variable to carry out residual confounder identification and interaction test with common clinical indicators, so as to minimise the impact of confounding factors on GBM OS as much as possible. Furthermore, discrimination and calibration are the most commonly used indicators in the evaluation of clinical prediction models. However, a systematic review found that while 63% of the studies on prediction models reported discrimination data, only 36% included calibration data (Wessler et al., 2015). In the present study, we report both discrimination and calibration in the training and validation cohorts. In addition, we did not only use the enhanced bootstrap test for internal validation, but also directly compared multiple models in two data sets (TCGA and CGGA) to minimise model overfitting. This significantly differs from traditional studies of clinical models, and allowed us to select the optimal model for prognosis prediction in patients with GBM.

Previous studies have shown that the higher the stemness indices scores, the worse the overall survival outcomes, (Pei et al., 2020), which is the opposite from our results. Therefore, we wanted to further explore the reasons for the different prognosis among the groups. Using gene-expression-based metrics, a recent study reported the association of stemness with immune cell infiltration and genomic, transcriptomic, and clinical parameters across 21 solid cancers (Miranda et al., 2019). Pervasive negative associations between cancer stemness and anticancer immunity have also been found (Miranda et al., 2019). In line with the current pancreatic cancer findings, we also analysed infiltrative immune cells in distinct

cohorts (TCGA and CGGA) using the *CIBERSORTx* algorithm and found that the high-c\_mRNAsi group exhibited significant immune suppression. Based on the expression data in TCGA and CGGA, we observed that c\_mRNAsi correlated negatively with infiltrating levels of immune cells. Promoting Treg overrepresentation and function induces systemic and intratumoural immunosuppression (Long et al., 2020). CD8<sup>+</sup> cytotoxic T lymphocyte cells, macrophages, Tregs, and other immune cells can respond to GBM treatment, including immunotherapy, to a certain extent (Choi et al., 2019). Meanwhile, high c\_mRNAsi was associated with the up-regulation of M1 macrophages, resting dendritic cells, monocytes, activated NK cells, and follicular T helper cells. In addition, we also found that general immune-related pathways were activated in the high-c\_mRNAsi group, which is consistent with previous findings (Zhang C. et al., 2020). Collectively, our results suggest that the better prognosis of patients with high c\_mRNAsi may be owing to the presence of more tumour stem cells and more tumour neoantigens, which results in the higher infiltration of tumour immune cells. Based on our findings, we propose c\_mRNAsi as a new marker for tumour immunotherapy in the future.

The stemness indices reflects the ability of self-renewal and unlimited proliferation. We found that significantly different enrichment pathways are mainly related to cell cycle, damage repair, proliferation, apoptosis, angiogenesis, glucose, lipid metabolism and energy metabolism through ssGSEA. The current view is that tumour stem cells are related to the inhibitory immune microenvironment (Alves et al., 2021). At present, it is found that the stemness indices is also related to IL6/JAK/STAT3, IL-2/STAT5, and TGF- $\beta$  signalling pathways (Zhang et al., 2019; Liu et al., 2021; Mo et al., 2021; Zhou et al., 2021). Interleukins are closely related to the proliferation and function of T cells (Ceuppens et al., 1988; Popmihajlov et al., 2012; Raeber et al., 2018). Tregs are a key source of TGF- $\beta$  ligands (Zhang Y. et al., 2020). Together, the pathways we enriched here are closely related to immune response.

There are several limitations in our study that need to be addressed in the future. First, we used four different datasets (TCGA, CGGA, GSE4412, and GSE13041) to test the prognostic value of mRNAsi/c\_mRNAsi and found the power and robustness of c\_mRNAsi. However, we could not definitively determine whether the c\_mRNAsi obtained from the bulk tumour sequencing/array could be utilised for all types of GBM samples from diverse genetic backgrounds. Furthermore, the study was based on public data. We should use our own sequencing data to verify the c\_mRNAsi and clinical model. Besides, the c\_mRNAsi signature could distinguish differential subpopulations with distinct prognosis. Whether stemness indices mediate poor immunotherapy response requires further investigation. There is still a long way before we can accomplish individualised classifications for treatment because other clinical and genetic/epigenetic factors must be considered and incorporated into treatment decision-making (Mansouri et al., 2019). In addition, due to the limitation of GEO data sources, we could only verify the prognostic signature of c\_mRNAsi, but could not identify confounding factors as in CGGA. Moreover, the c\_mRNAsi-related signature should be further validated in large samples of patients with GBM from



multiple centres to identify the associations not only with survival outcomes but also conventional drug responses, especially immunotherapy. Lastly, although we performed functional analysis and identified numerous differences in infiltrative immune cell abundance and the regulation of related pathways, specific experimental validations need to be designed to assess the real effect. Despite the above shortcomings, our work has certain advantages that cannot be ignored. We calculated the mRNAsi/c\_mRNAsi using a large number of samples and, for the first time, performed confounding variable identification and interaction and stratified analyses. Furthermore, we made a comprehensive comparison of several models, and proved the validity of our conclusions in multiple ways to verify the credibility of our results.

In conclusion, our study systematically assessed the GBM stemness indices based on multiple independent cohorts, providing a robust quantified mRNAsi/c\_mRNAsi reflective of stemness indices, and the associations with immune infiltration and immune related pathways. The c\_mRNAsi-based signature proved to be superior to other models in predicting OS prognosis, and may be a valuable classifier for uncovering distinct subgroups of stemness indices.

## DATA AVAILABILITY STATEMENT

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found in the article/**Supplementary Material**.

## AUTHOR CONTRIBUTIONS

MZ and JH conceived and designed the study. HC analyzed the data. BL authored the paper. XW prepared figures and tables. NG

supervised the study. FX and QY revised the manuscript. QZ audited the study.

## FUNDING

This study was supported by the National Natural Science Foundation of China (Nos. 82003386, 11601083, and U1805263), Fujian Provincial Committee of Natural Science and Technology (No. 2020J02039), National Collaboration Center in Immuno-Oncology (No. 2016sysbz02), Qihang Foundation of Fujian Medical University (No. 2018QH1088).

## SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fmolb.2021.777921/full#supplementary-material>

**Supplementary Figure 1** | Lowess results. **(A)**. mRNAsi with OS in TCGA. **(B)**. c\_mRNAsi with OS in TCGA. **(C)**. mRNAsi with OS in CGGA. **(D)**. c\_mRNAsi with OS in CGGA.

**Supplementary Figure 2** | Identification of optimal cut-off values. **(A)**. Optimal cut-off value of mRNAsi in TCGA. **(B)**. Optimal cut-off value of c\_mRNAsi in TCGA. **(C)**. Optimal cut-off value of mRNAsi in CGGA. **(D)**. Optimal cut-off value of c\_mRNAsi in CGGA.

**Supplementary Figure 3** | Survival curve of mRNAsi/c\_mRNAsi on prognosis. **(A)**. Disease-specific survival curve of mRNAsi in TCGA. **(B)**. Disease-specific survival curve of c\_mRNAsi in TCGA. **(C)**. Progression-free interval curve of mRNAsi in TCGA. **(D)**. Progression-free interval curve of c\_mRNAsi in TCGA.

**Supplementary Figure 4** | External validation in CGGA and GEO. **(A)**. AUC in TCGA. **(B)**. C-index in CGGA. **(C)**. Prediction error in CGGA. **(D)**. 0.5-year DCA in CGGA. **(E)**. 1-year DCA in CGGA. **(F)**. 1.5-year DCA in CGGA. **(G)**. 3-year DCA in CGGA. **(H)**. 5-year DCA in CGGA. **(I)**. 0.5-year calibration in CGGA. **(J)**. 1-year calibration in CGGA. **(K)**. 1.5-year calibration in CGGA. **(L)**. 3-year calibration in CGGA. **(M)**. 5-year calibration in CGGA. **(N)**. AUC in GSE4412. **(O)**. AUC in GSE13041.

**Supplementary Figure 5** | Construction of prognostic nomogram.

## REFERENCES

- Ali, H. R., Chlon, L., Pharoah, P. D. P., Markowitz, F., and Caldas, C. (2016). Patterns of Immune Infiltration in Breast Cancer and Their Clinical Implications: A Gene-Expression-Based Retrospective Study. *Plos Med.* 13 (12), e1002194. doi:10.1371/journal.pmed.1002194
- Alves, A. L. V., Gomes, I. N. F., Carloni, A. C., Rosa, M. N., da Silva, L. S., Evangelista, A. F., et al. (2021). Role of Glioblastoma Stem Cells in Cancer Therapeutic Resistance: a Perspective on Antineoplastic Agents from Natural Sources and Chemical Derivatives. *Stem Cell Res Ther* 12 (1), 206. doi:10.1186/s13287-021-02231-x
- Barbie, D. A., Tamayo, P., Boehm, J. S., Kim, S. Y., Moody, S. E., Dunn, I. F., et al. (2009). Systematic RNA Interference Reveals that Oncogenic KRAS-Driven Cancers Require TBK1. *Nature* 462 (7269), 108–112. doi:10.1038/nature08460
- Barrett, T., Wilhite, S. E., Ledoux, P., Evangelista, C., Kim, I. F., Tomashevsky, M., et al. (2013). NCBI GEO: Archive for Functional Genomics Data Sets-Update. *Nucleic Acids Res.* 41 (Database issue), D991–D995. doi:10.1093/nar/gks1193
- Burgenske, D. M., Yang, J., Decker, P. A., Kollmeyer, T. M., Kosel, M. L., Mladek, A. C., et al. (2019). Molecular Profiling of Long-Term IDH-Wildtype Glioblastoma Survivors. *Neuro-oncology* 21 (11), 1458–1469. doi:10.1093/neuonc/noz129
- Ceuppens, J. L., Baroja, M. L., Lorre, K., Van Damme, J., and Billiau, A. (1988/1950). Human T Cell Activation with Phytohemagglutinin. The Function of IL-6 as an Accessory Signal. *J. Immunol.* 141 (11), 3868–3874.
- Chaddad, A., Daniel, P., Sabri, S., Desrosiers, C., and Abdulkarim, B. (2019). Integration of Radiomic and Multi-Omic Analyses Predicts Survival of Newly Diagnosed IDH1 Wild-type Glioblastoma. *Cancers* 11 (8), 1148. doi:10.3390/cancers11081148
- Choi, B. D., Maus, M. V., June, C. H., and Sampson, J. H. (2019). Immunotherapy for Glioblastoma: Adoptive T-Cell Strategies. *Clin. Cancer Res.* 25 (7), 2042–2048. doi:10.1158/1078-0432.CCR-18-1625
- Freije, W. A., Castro-Vargas, F. E., Fang, Z., Horvath, S., Cloughesy, T., Liau, L. M., et al. (2004). Gene Expression Profiling of Gliomas Strongly Predicts Survival. *Cancer Res.* 64 (18), 6503–6510. doi:10.1158/0008-5472.can-04-0452
- Galleo, O. (2015). Nonsurgical Treatment of Recurrent Glioblastoma. *Curr. Oncol.* 22 (4), 273–281. doi:10.3747/co.22.2436
- Harrell, F. E., Jr., Lee, K. L., and Mark, D. B. (1996). Multivariable Prognostic Models: Issues in Developing Models, Evaluating Assumptions and Adequacy, and Measuring and Reducing Errors. *Statist. Med.* 15 (4), 361–387. doi:10.1002/(sici)1097-0258(19960229)15:4<361:aid-sim168>3.0.co;2-4

- Hoadley, K. A., Yau, C., Wolf, D. M., Cherniack, A. D., Tamborero, D., Ng, S., et al. (2014). Multiplatform Analysis of 12 Cancer Types Reveals Molecular Classification within and across Tissues of Origin. *Cell* 158 (4), 929–944. doi:10.1016/j.cell.2014.06.049
- Kaipio, K., Chen, P., Roering, P., Huhtinen, K., Mikkonen, P., Östling, P., et al. (2020). ALDH1A1-related Stemness in High-grade Serous Ovarian Cancer Is a Negative Prognostic Indicator but Potentially Targetable by EGFR/mTOR-PI3K/aurora Kinase Inhibitors. *J. Pathol.* 250 (2), 159–169. doi:10.1002/path.5356
- Kaushal, K., and Ramakrishna, S. (2020). Deubiquitinating Enzyme-Mediated Signaling Networks in Cancer Stem Cells. *Cancers* 12 (11), 3253. doi:10.3390/cancers12113253
- Kernan, W. N., Viscoli, C. M., Brass, L. M., Broderick, J. P., Brott, T., Feldmann, E., et al. (2000). Phenylpropanolamine and the Risk of Hemorrhagic Stroke. *N. Engl. J. Med.* 343 (25), 1826–1832. doi:10.1056/nejm200012213432501
- Kerr, K. F., Brown, M. D., Zhu, K., and Janes, H. (2016). Assessing the Clinical Impact of Risk Prediction Models with Decision Curves: Guidance for Correct Interpretation and Appropriate Use. *Jco* 34 (21), 2534–2540. doi:10.1200/JCO.2015.65.5654
- Krex, D., Klink, B., Hartmann, C., von Deimling, A., Pietsch, T., Simon, M., et al. (2007). Long-term Survival with Glioblastoma Multiforme. *Brain* 130 (Pt 10), 2596–2606. doi:10.1093/brain/awm204
- Le, T., Aronow, R. A., Kirshstein, A., and Shahriyari, L. (2020). A Review of Digital Cytometry Methods: Estimating the Relative Abundance of Cell Types in a Bulk of Cells. *Brief. Bioinformatics* 22. doi:10.1093/bib/bbaa219
- Lee, Y., Scheck, A. C., Cloughesy, T. F., Lai, A., Dong, J., Farooqi, H. K., et al. (2008). Gene Expression Analysis of Glioblastomas Identifies the Major Molecular Basis for the Prognostic Benefit of Younger Age. *BMC Med. Genomics* 1, 52. doi:10.1186/1755-8794-1-52
- Li, H.-P., Wang, J., Guo, S., Cai, X., and Xu, J.-W. (2019). Establishment and Verification of a Surgical Prognostic Model for Cervical Spinal Cord Injury without Radiological Abnormality. *Neural Regen. Res.* 14 (4), 713–720. doi:10.4103/1673-5374.247480
- Lian, H., Han, Y. P., Zhang, Y. C., Zhao, Y., Yan, S., Li, Q. F., et al. (2019). Integrative Analysis of Gene Expression and DNA Methylation through One-class Logistic Regression Machine Learning Identifies Stemness Features in Medulloblastoma. *Mol. Oncol.* 13 (10), 2227–2245. doi:10.1002/1878-0261.12557
- Liberzon, A., Birger, C., Thorvaldsdóttir, H., Ghandi, M., Mesirov, J. P., and Tamayo, P. (2015). The Molecular Signatures Database Hallmark Gene Set Collection. *Cel Syst.* 1 (6), 417–425. doi:10.1016/j.cels.2015.12.004
- Liu, S., Zhang, C., Wang, B., Zhang, H., Qin, G., Li, C., et al. (2021). Regulatory T Cells Promote Glioma Cell Stemness through TGF- $\beta$ -NF- $\kappa$ B-IL6-STAT3 Signaling. *Cancer Immunol. Immunother.* 70 (9), 2601–2616. doi:10.1007/s00262-021-02872-0
- Long, Y., Tao, H., Karachi, A., Gripping, A. J., Jin, L., Chang, Y., et al. (2020). Dysregulation of Glutamate Transport Enhances Treg Function that Promotes VEGF Blockade Resistance in Glioblastoma. *Cancer Res.* 80 (3), 499–509. doi:10.1158/0008-5472.CAN-19-1577
- Malta, T. M., Sokolov, A., Gentles, A. J., Burzykowski, T., Poisson, L., Weinstein, J. N., et al. (2018). Machine Learning Identifies Stemness Features Associated with Oncogenic Dedifferentiation. *Cell* 173 (2), 338–e15. doi:10.1016/j.cell.2018.03.034
- Mansouri, A., Hachem, L. D., Mansouri, S., Nassiri, F., Laperriere, N. J., Xia, D., et al. (2019). MGMT Promoter Methylation Status Testing to Guide Therapy for Glioblastoma: Refining the Approach Based on Emerging Evidence and Current Challenges. *Neuro-oncology* 21 (2), 167–178. doi:10.1093/neuonc/ny132
- Martinez, R., Schackert, G., Yaya-Tur, R., Rojas-Marcos, I., Herman, J. G., and Esteller, M. (2007). Frequent Hypermethylation of the DNA Repair Gene MGMT in Long-Term Survivors of Glioblastoma Multiforme. *J. Neurooncol.* 83 (1), 91–93. doi:10.1007/s11060-006-9292-0
- McNeill, K. A. (2016). Epidemiology of Brain Tumors. *Neurol. Clin.* 34 (4), 981–998. doi:10.1016/j.ncl.2016.06.014
- Miranda, A., Hamilton, P. T., Zhang, A. W., Pattnaik, S., Becht, E., Mezheyski, A., et al. (2019). Cancer Stemness, Intratumoral Heterogeneity, and Immune Response across Cancers. *Proc. Natl. Acad. Sci. USA* 116 (18), 9020–9029. doi:10.1073/pnas.1818210116
- Mo, F., Yu, Z., Li, P., Oh, J., Spolski, R., Zhao, L., et al. (2021). An Engineered IL-2 Partial Agonist Promotes CD8+ T Cell Stemness. *Nature* 597 (7877), 544–548. doi:10.1038/s41586-021-03861-0
- Molenaar, R. J., Verbaan, D., Lamba, S., Zanoni, C., Jeuken, J. W. M., Boots-Sprenger, S. H. E., et al. (2014). The Combination of IDH1 Mutations and MGMT Methylation Status Predicts Survival in Glioblastoma Better Than Either IDH1 or MGMT Alone. *Neuro-oncology* 16 (9), 1263–1273. doi:10.1093/neuonc/nou005
- Newman, A. M., Steen, C. B., Liu, C. L., Gentles, A. J., Chaudhuri, A. A., Scherer, F., et al. (2019). Determining Cell Type Abundance and Expression from Bulk Tissues with Digital Cytometry. *Nat. Biotechnol.* 37 (7), 773–782. doi:10.1038/s41587-019-0114-2
- Pan, S., Zhan, Y., Chen, X., Wu, B., and Liu, B. (2019). Identification of Biomarkers for Controlling Cancer Stem Cell Characteristics in Bladder Cancer by Network Analysis of Transcriptome Data Stemness Indices. *Front. Oncol.* 9, 613. doi:10.3389/fonc.2019.00613
- Pei, J., Wang, Y., and Li, Y. (2020). Identification of Key Genes Controlling Breast Cancer Stem Cell Characteristics via Stemness Indices Analysis. *J. Transl. Med.* 18 (1), 74. doi:10.1186/s12967-020-02260-9
- Pencina, M. J., D'Agostino, R. B., D'Agostino, Vasan, R. S., Jr., and Vasan, R. S. (2008). Evaluating the Added Predictive Ability of a New Marker: from Area under the ROC Curve to Reclassification and beyond. *Statist. Med.* 27 (2), 157–172. doi:10.1002/sim.2929
- Popmihajlov, Z., Xu, D., Morgan, H., Milligan, Z., and Smith, K. A. (2012). Conditional IL-2 Gene Deletion: Consequences for T Cell Proliferation. *Front. Immun.* 3, 102. doi:10.3389/fimmu.2012.00102
- Raeber, M. E., Zurbuchen, Y., Impellizzeri, D., and Boyman, O. (2018). The Role of Cytokines in T-Cell Memory in Health and Disease. *Immunol. Rev.* 283 (1), 176–193. doi:10.1111/imr.12644
- Sasmita, A. O., Wong, Y. P., and Ling, A. P. K. (2018). Biomarkers and Therapeutic Advances in Glioblastoma Multiforme. *Asia-pac J. Clin. Oncol.* 14 (1), 40–51. doi:10.1111/ajco.12756
- Saygin, C., Matei, D., Majeti, R., Reizes, O., and Lathia, J. D. (2019). Targeting Cancer Stemness in the Clinic: From Hype to Hope. *Cell stem cell* 24 (1), 25–40. doi:10.1016/j.stem.2018.11.017
- Seneviratne, A. K., Xu, M., Henao, J. J. A., Fajardo, V. A., Hao, Z., Voisin, V., et al. (2019). The Mitochondrial Transacylase, Tafazzin, Regulates AML Stemness by Modulating Intracellular Levels of Phospholipids. *Cell stem cell* 24 (4), 621–636. doi:10.1016/j.stem.2019.02.020
- Sokolov, A., Paull, E. O., and Stuart, J. M. (2016). One-class Detection of Cell States in Tumor Subtypes. *Pac. Symp. Biocomput* 21, 405–416. doi:10.1142/9789814749411\_0037
- Soria, J.-C., Felip, E., Cobo, M., Lu, S., Syrigos, K., Lee, K. H., et al. (2015). Afatinib versus Erlotinib as Second-Line Treatment of Patients with Advanced Squamous Cell Carcinoma of the Lung (LUX-Lung 8): an Open-Label Randomised Controlled Phase 3 Trial. *Lancet Oncol.* 16 (8), 897–907. doi:10.1016/s1470-2045(15)00006-6
- Subramanian, A., Tamayo, P., Mootha, V. K., Mukherjee, S., Ebert, B. L., Gillette, M. A., et al. (2005). Gene Set Enrichment Analysis: a Knowledge-Based Approach for Interpreting Genome-wide Expression Profiles. *Proc. Natl. Acad. Sci.* 102 (43), 15545–15550. doi:10.1073/pnas.0506580102
- Tao, Y., Kang, B., Petkovich, D. A., Bhandari, Y. R., In, J., In, G., et al. (2019). Aging-like Spontaneous Epigenetic Silencing Facilitates Wnt Activation, Stemness, and BrafV600E-Induced Tumorigenesis. *Cancer cell* 35 (2), 315–328. doi:10.1016/j.ccell.2019.01.005
- Turaga, S. M., Silver, D. J., Bayik, D., Paouri, E., Peng, S., Lauko, A., et al. (2020). JAM-A Functions as a Female Microglial Tumor Suppressor in Glioblastoma. *Neuro-oncology* 22, 1591–1601. doi:10.1093/neuonc/noaa148
- Wang, X., Prager, B. C., Wu, Q., Kim, L. J. Y., Gimple, R. C., Shi, Y., et al. (2018). Reciprocal Signaling between Glioblastoma Stem Cells and Differentiated Tumor Cells Promotes Malignant Progression. *Cell Stem Cell* 22 (4), 514–528. doi:10.1016/j.stem.2018.03.011
- Wessler, B. S., Lai Yh, L., Kramer, W., Cangelosi, M., Raman, G., Lutz, J. S., et al. (2015). Clinical Prediction Models for Cardiovascular Disease. *Circ. Cardiovasc. Qual. Outcomes* 8 (4), 368–375. doi:10.1161/circoutcomes.115.001693
- Xia, P., Li, Q., Wu, G., and Huang, Y. (2020). Identification of Glioma Cancer Stem Cell Characteristics Based on Weighted Gene Prognosis Module Co-expression Network Analysis of Transcriptome Data Stemness Indices. *J. Mol. Neurosci.* 70, 1512–1520. doi:10.1007/s12031-020-01590-z

- Yoshihara, K., Shahmoradgoli, M., Martínez, E., Vegesna, R., Kim, H., Torres-García, W., et al. (2013). Inferring Tumour Purity and Stromal and Immune Cell Admixture from Expression Data. *Nat. Commun.* 4, 2612. doi:10.1038/ncomms3612
- Young, R. M., Jamshidi, A., Davis, G., and Sherman, J. H. (2015). Current Trends in the Surgical Management and Treatment of Adult Glioblastoma. *Ann. Transl. Med.* 3 (9), 121. doi:10.3978/j.issn.2305-5839.2015.05.10
- Zhang, B., Ye, H., Ren, X., Zheng, S., Zhou, Q., Chen, C., et al. (2019). Macrophage-expressed CD51 Promotes Cancer Stem Cell Properties via the TGF- $\beta$ 1/sm $\alpha$ 2/3 axis in Pancreatic Cancer. *Cancer Lett.* 459, 204–215. doi:10.1016/j.canlet.2019.06.005
- Zhang, C., Chen, T., Li, Z., Liu, A., Xu, Y., Gao, Y., et al. (2020a). Depiction of Tumor Stemlike Features and Underlying Relationships with hazard Immune Infiltrations Based on Large Prostate Cancer Cohorts. *Brief. Bioinformatics* 22, bbaa211. doi:10.1093/bib/bbaa211
- Zhang, M., Wang, X., Chen, X., Guo, F., and Hong, J. (2020b). Prognostic Value of a Stemness Index-Associated Signature in Primary Lower-Grade Glioma. *Front. Genet.* 11, 441. doi:10.3389/fgene.2020.00441
- Zhang, M., Wang, X., Chen, X., Zhang, Q., and Hong, J. (2020c). Novel Immune-Related Gene Signature for Risk Stratification and Prognosis of Survival in Lower-Grade Glioma. *Front. Genet.* 11, 363. doi:10.3389/fgene.2020.00363
- Zhang, Y., Lazarus, J., Steele, N. G., Yan, W., Lee, H.-J., Nwosu, Z. C., et al. (2020d). Regulatory T-Cell Depletion Alters the Tumor Microenvironment and Accelerates Pancreatic Carcinogenesis. *Cancer Discov.* 10 (3), 422–439. doi:10.1158/2159-8290.CD-19-0958
- Zhou, J.-G., Liang, B., Jin, S.-H., Liao, H.-L., Du, G.-B., Cheng, L., et al. (2019). Development and Validation of an RNA-Seq-Based Prognostic Signature in Neuroblastoma. *Front. Oncol.* 9, 1361. doi:10.3389/fonc.2019.01361
- Zhou, M., Venkata, P. P., Viswanadhapalli, S., Palacios, B., Alejo, S., Chen, Y., et al. (2021). KDM1A Inhibition Is Effective in Reducing Stemness and Treating Triple Negative Breast Cancer. *Breast Cancer Res. Treat.* 185 (2), 343–357. doi:10.1007/s10549-020-05963-1

**Conflict of Interest:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

**Publisher's Note:** All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2021 Zhang, Chen, Liang, Wang, Gu, Xue, Yue, Zhang and Hong. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.