

Identification and validation of immune pathway subtyping in glioblastoma to predict clinical prognosis and immunotherapy response in patients

Mingshi Zhang¹, Mingjun Li¹, Jinrui Liu¹, Zhicheng Gu¹, Yanmei Lu¹, Yu Long¹, Yuyi Hou²

¹Department of Neurosurgery, The First Affiliated Hospital of Jiamusi University, Jiamusi, China

²Department of Stomatology, The First Affiliated Hospital of Jiamusi University, Jiamusi, China

Submitted: 2 November 2022; **Accepted:** 16 January 2023

Online publication: 4 February 2023

Arch Med Sci

DOI: <https://doi.org/10.5114/aoms/159344>

Copyright © 2023 Termedia & Banach

Corresponding author:

Yuyi Hou

Department of Stomatology

The First Affiliated

Hospital of Jiamusi

University

Jiamusi, China

E-mail: houyuyi2022@

outlook.com

Abstract

Introduction: Glioblastoma (GBM) as a frequently diagnosed primary intracranial tumor has a significantly poor prognosis. Only a few studies have probed into the immune profile associated with GBM. This study explored the role of immune features of GBM in prognosis and immunotherapy response.

Material and methods: GBM samples were subtyped by evaluating 15 immune-related pathways and genes using consensus clustering. GISTIC2 analyzed copy number variations and the impute package was used to perform methylation analysis. Immune characteristics were unveiled by using ssGSEA, ESTIMATE, and CIBERSORT. Immunotherapy and chemotherapeutic drug responses were calculated with the TIDE and pRRophetic package respectively. Weighted gene co-expression network analysis (WGCNA), Cox regression, Lasso, and stepAIC were used to develop a prognostic immune score (IMscore) model.

Results: GBM was categorized into 3 subtypes: immune-deprived (Immune-D) (low enrichment of immune pathways and high enrichment of DNA damage repair pathways); stromal-enriched (Stromal-E) (high enrichment of immune pathways, oncogenic pathways and stromal pathways); and immune-enriched (Immune-E) (low enrichment of DNA damage repair pathways and high enrichment of immune pathways). Methylation differences were found in TWIST1, CDH2 and CDH1 among 3 subtypes. Immune-E responded better to immunotherapy, while Immune-D was more sensitive to chemotherapeutic drugs. This study established a prognostic model with five genes (OSMR, SPP1, CUL1, CTBP2, NGFR) for GBM.

Conclusions: Three subtypes had different prognosis and response to immunotherapy and chemotherapy. A five-gene prognostic model was robust to predict prognosis in GBM as well as pan-cancer. The subtyping and prognostic model may facilitate individualized prognosis management and personalized therapeutic intervention.

Key words: glioblastoma, immune-related pathways, immune subtyping, risk model, prognosis.

Introduction

In adults, glioblastoma (GBM) as a frequently detected primary central nervous system (CNS) malignancy has an incidence of approximately 5–7/100,000, with a significantly poor prognosis. The median survival

time of GBM patients is shorter than 15 months [1]. Safe tumor resection accompanied with post-operative concurrent radiotherapy is the most common clinical modality in GBM treatment, but still the recurrence rate is as high as 90% [2, 3]. Immune checkpoint inhibitors (ICIs) represented by PD-1/PD-L1 could successfully improve the prognosis of some GBM patients, but most patients do not benefit from them effectively [4, 5]. Findings of recent clinical trials indicated an urgent need to accurately identify potential GBM beneficiaries of immunotherapy. Given that no new therapeutic approach is able to prolong GBM survival [6, 7], it is important to actively explore the pathogenesis of GBM so as to discover new therapeutic targets for GBM and improve the prognosis of the patients.

GBM development is the result of a combination of factors internal to tumor cells that have an external tumor microenvironment (TME) [8, 9]. Exogenous components and intrinsic mechanisms of GBM cells determine the immunosuppressive state of the glioma microenvironment and are one of the greatest obstacles to achieving the effectiveness of immunotherapy [10]. Immune pathways such as antigen processing and presentation [11], natural killer cell mediated cytotoxicity [12], and T cell receptor signaling [13] have been reported to have great potential to confer the response to immunotherapy in glioblastoma. The interactions of secreted components from stromal cells and immune cells orchestrate TME and regulate the response to immunotherapy. A large amount of evidence has confirmed the role of oncogenic pathways such as Wnt signaling [14], PI3K-Akt signaling and TGF- β signaling [15] in promoting GBM development. In addition, DNA repair pathways have been uncovered to play critical roles in the maintenance of genomic stability that is responsible for tumorigenesis [16]. We supposed that by using the features of the above related pathways, we may establish an accurate subtyping system for classifying GBM patients and guiding immunotherapy.

To some extent, the failure of immunotherapy is related to inaccurate preclinical models, as pre-clinical studies do not consistently demonstrate drug responses that are therapeutically active in patients and are not able to classify patients correctly. To explore the immune features of GBM and improve immunotherapy for GBM patients, we collected immune and stroma-related pathways, used gene expression data to construct molecular subtyping of GBM immune using ConsensusClusterPlus. Gene mutation, methylation and immune state were analyzed to reflect the distance among molecular subtypes. Furthermore, we also performed validation analysis in the GBM immu-

notherapy dataset to provide new evidence for a deeper understanding of the immunomolecular mechanisms of GBM and individualized immunotherapy.

Material and methods

Patient data and gene expression profile data sources

Clinical and chip data of GBM patients from The Cancer Genome Atlas (TCGA) GDC API and the Chinese Glioma Genome Atlas (CGGA) database (<http://www.cgga.org.cn/>) were obtained. This study was approved by The First Affiliated Hospital of Jiamusi University and patients' consent was waived by the institution. The GBM datasets, including CGGA693 and CGGA325, were downloaded from the CGGA database. These two datasets were preprocessed in the following steps:

- 1) Downloading the normalized datasets;
- 2) Keeping the samples with complete survival time and status;
- 3) Merging the CGGA693 and CGGA325 datasets;
- 4) Using the limma package removeBatchEffect function to remove the batch effect between the two datasets [17], and the final dataset was named as CGGA.

After the two groups of data were preprocessed, there were 524 samples in TCGA and 374 samples in the CGGA dataset.

Subtyping of GBM patients

Fifteen immune-related pathways and corresponding genes belonging to DNA damage repair, stromal pathways, oncogenic pathways, and immune pathways were collected from a published study [18]. Then, in the TCGA and CGGA datasets, we used the single sample gene set enrichment analysis (ssGSEA) method to evaluate the scores of these 15 pathways, based on which the samples were clustered using the ConsensusClusterPlus R package. With the cumulative distribution function (CDF), we determined the optimal number of clusters. The optimal classification was determined based on CDF and samples' molecular typology was acquired through calculating the consistency matrix [19].

Analysis of immune genes and driver genes among the molecular subtypes

We obtained chemokine and receptor genes, MHC from published literature [20], and then compared the distribution of these genes as well as CTLA4 and PDCD1 in the different typologies of GBM and whether there was concordance in the different datasets.

Driver genes were obtained from existing studies; 159 out of 172 genes had copy data [21]. Next, we compared the mutation profiles of these 172 driver genes across GBM subtypes and differences in tumor mutation burden (TMB). We used GISTIC2 to analyze the variation of copy number in different fractions, with a ratio > 0.2 being defined as gain, ratio < 0.2 being defined as loss, and the rest being defined as diploid.

Pathway analysis

We obtained the genes corresponding to the HALLMARK_EPITHELIAL_MESENCHYMAL_TRANSITION (EMT) pathway from the h.all.v7.4.symbols.gmt file on the CGGA website, and then used the ssGSEA method to calculate the EMT score for both the TCGA and CGGA datasets. Next, the genes of cytolytic activity were collected from a previous paper [22]. The method of ssGSEA was applied to calculate the cytolytic activity score for both TCGA and CGGA datasets between the subtypes. ssGSEA in the GSVA R package first normalizes gene expression values of a given sample by rank order, and then calculates the enrichment score using the empirical CDF [23].

Methylation analysis between molecular subtypes and drug sensitivity analysis

The 450K methylation data of glioma samples from the TCGA dataset were downloaded, and then to complete the missing values [24], we used the KNN algorithm of the impute package. Further, 7 EMT promoting genes [25] were obtained from existing studies, and we analyzed the mean beta values of these genes across different subtypes and calculated the Pearson correlation between the gene expression values and the beta values of methylation. Next, the differences in chemotherapy and immunotherapy in molecular subtypes were compared. We used the TIDE (<http://tide.dfci.harvard.edu/>) software to assess the potential clinical effects of immunotherapy on our defined molecular subtypes, with higher TIDE prediction scores indicating a stronger likelihood of immune escape, which also suggested less immunotherapy benefit [26]. Finally, we performed drug sensitivity prediction for LGG with the pRRophetic package [27].

Weighted gene co-expression network analysis (WGCNA) among the molecular subtypes

Molecular subtype-correlated gene modules were defined by the WGCNA package [28]. The samples were first clustered to filter the co-expression modules. To determine the connection points of the co-expression network, it was specified that

the $\log(k)$ of the node with occurrence connectivity k had to be negatively correlated with the $\log(P(k))$ of the probability of occurrence of the node, with a correlation coefficient greater than 0.85. In the next step, the expression matrix was converted into an adjacency matrix, which was then transformed into a topology overlap matrix (TOM). Based on the TOM, we used the average-linkage hierarchical clustering method to cluster the genes according to the criteria of dynamic tree cut; the minimum number of genes per gene network module was set to 80. After this, the eigenvalues (eigen-genes) of each module were calculated, and then the modules were clustered, followed by merging close modules into a new module.

Construction and validation of prognostic model IMscore

In the TCGA dataset, the 15 pathways were first scored by one-way Cox regression analysis, followed by calculation of the Pearson correlation to select the positively correlated genes in the pathway. Next, for the pathway-related genes, we randomly divided the GBM samples in the TCGA dataset into a Train group and a validation (Test) group at a ratio of 3 : 2, and then reduced the number of pathway-related genes by the Lasso regression, which can better solve the problem of multicollinearity in regression analysis through compressing some coefficients and at the same time setting some coefficients to zero. The number of factors was determined when the coefficients of independent variables tend to zero, with the gradual increase of lambda. Next, the Akaike information criterion (AIC) was performed with stepwise regression, which could consider the model statistical fit as well as parameter number to fit it. Here, a smaller value indicates a model with better performance, meaning that the model may have a sufficient fit but with fewer parameters [29]. After that, in the TCGA Train dataset, the survminer package was used to find the best cutoff of the IMscore, based on which the GBM was divided into two groups. Finally, survival differences were compared by log-rank test between the two groups.

To verify the robustness of the IMscore model, we downloaded the expression profiles and clinical data of the remaining 32 cancer types in the TCGA database, and then used the IMscore model to calculate the IMscore of each cancer type to obtain the best cutoff. The survival curve analysis was performed on the high and low IMscore.

Statistical analysis

The limma package was used in group comparison, while differences in mean gene expression

among the samples were analyzed using the independent *t*-test. We used the MASS package for stepAIC analysis [29]. Enrichment pathway analysis was carried out using ssGSEA [23]. We performed KM survival analysis, univariate [30] and multivariate Cox proportional risk analysis and model development [31] in the Survival package, and visualized the results in the survminer package [32]. Lasso regression analysis was conducted in the glmnet package [32]. Unless otherwise specified, all statistical tests were bilateral and all statistical analyses were performed using R software (version 4.1.3, <https://www.r-project.org/>). $P < 0.05$ was defined as statistically significant.

Results

Glioma subtyping based on the pathway scores

We first used the limma package to remove the batch effect of CGGA693 and CGGA325 datasets, and the data before and after the removal of the batch effect were presented by principal component analysis (PCA) plots (Supplementary Figure S1). The results demonstrated that the samples of different datasets changed from independent distribution into uniform mixed distribution after removing the batch effect. Next, based on the 15 pathways and gene scores, it can be seen from the CDF Delta area curve that the clustering results were more stable when the cluster number was 3; therefore, the tri-correlation subtyping ($k = 3$) was obtained.

To more clearly demonstrate the pathway differences among the three subtypes, we drew a pathway heatmap based on the scores. GBM patients could be roughly classified into three subtypes, namely, immune-Deprived (Immune-D), stromal-enriched subtype (Stromal-E) and immune-enriched subtype (Immune-E), by the pathway scores. The Immune-E phenotype showed a higher immune pathway score, while Stromal-E phenotype had a significantly enriched stromal pathway score, and the oncogenic pathways were also significantly enriched. However, Immune-D was the opposite to that of Immune-E. As easily observed, there was consistency in the two datasets, and a clear boundary between different subtypes, supporting the reliability and prevalence of our subtypes in different datasets.

Comparison of immune scores between different GBM subtypes

The stromal and immune composition of different GBM subtypes was assessed according to the purity of the microenvironment and immune cell infiltration using ESTIMATE [33]. The results showed that in TCGA and CGGA subtypes Immune-

eScore was higher in Immune-E and Stromal-E than in Immune-D; StromalScore was the lowest in Immune-D and the highest in Stromal-E; ESTIMATEScore and TumorPurity were consistent in both subtypes (Figures 1 A–D). Subsequently, the ssGSEA method calculated EMT scores and cytolytic activity in the TCGA and CGGA datasets, and significant differences were detected among the three subtypes (Figures 1 E, F). Furthermore, in order to investigate the differences in immune cells and immune function of patients in different subtypes, we used the ssGSEA algorithm to calculate the relative abundance of 28 immune cells. The results showed that the TCGA and CGGA data were consistent, with the highest score in Immune-E (Figure 1 G). These results also had significant differences in the immune microenvironment among the three GBM subtypes. Furthermore, the differences in immune cells and functions of patients of different subtypes were assessed through calculating the relative abundance of 28 immune cells using the ssGSEA algorithm. Consistency in the TCGA and CGGA data was observed, with the highest score in Immune-E (Figure 1 G). These data indicated significant differences in the immune microenvironment among the three GBM subtypes.

Comparative analysis of the distribution of immune genes and MHC in the subtyping

To investigate whether the immune differences among the subtypes were caused by immune genes, we compared the expression differences of representative genes PDCD1 and CTLA4 in the TCGA and CGGA datasets, and it was observed that the two genes were expressed the highest in Immune-E (Figures 2 A, B). Further expression analysis on chemokine, MHC, and receptor genes showed that in Immune-E in different datasets, the expression of MHC genes was higher (Figure 2 C).

Comparison of cancer driver gene variants in different subtypes

Next, we explored the differences in immune microenvironment leading to GBM in different subtypes based on cancer driver genes. Fisher's test detected significant differences in 5 of the 172 cancer driver genes, with ZNF208 showing the highest mutation frequency in Immune-E typing (Figure 3 A), but there was no statistically significant difference in TMB (Figure 3 B). Meanwhile, the survival curve (KM) analysis in driver genes with mutations or in the wild-type samples showed that the mutant phenotypes of eight genes were significantly different from that of the wild type (Figure 3 C). Further inves-

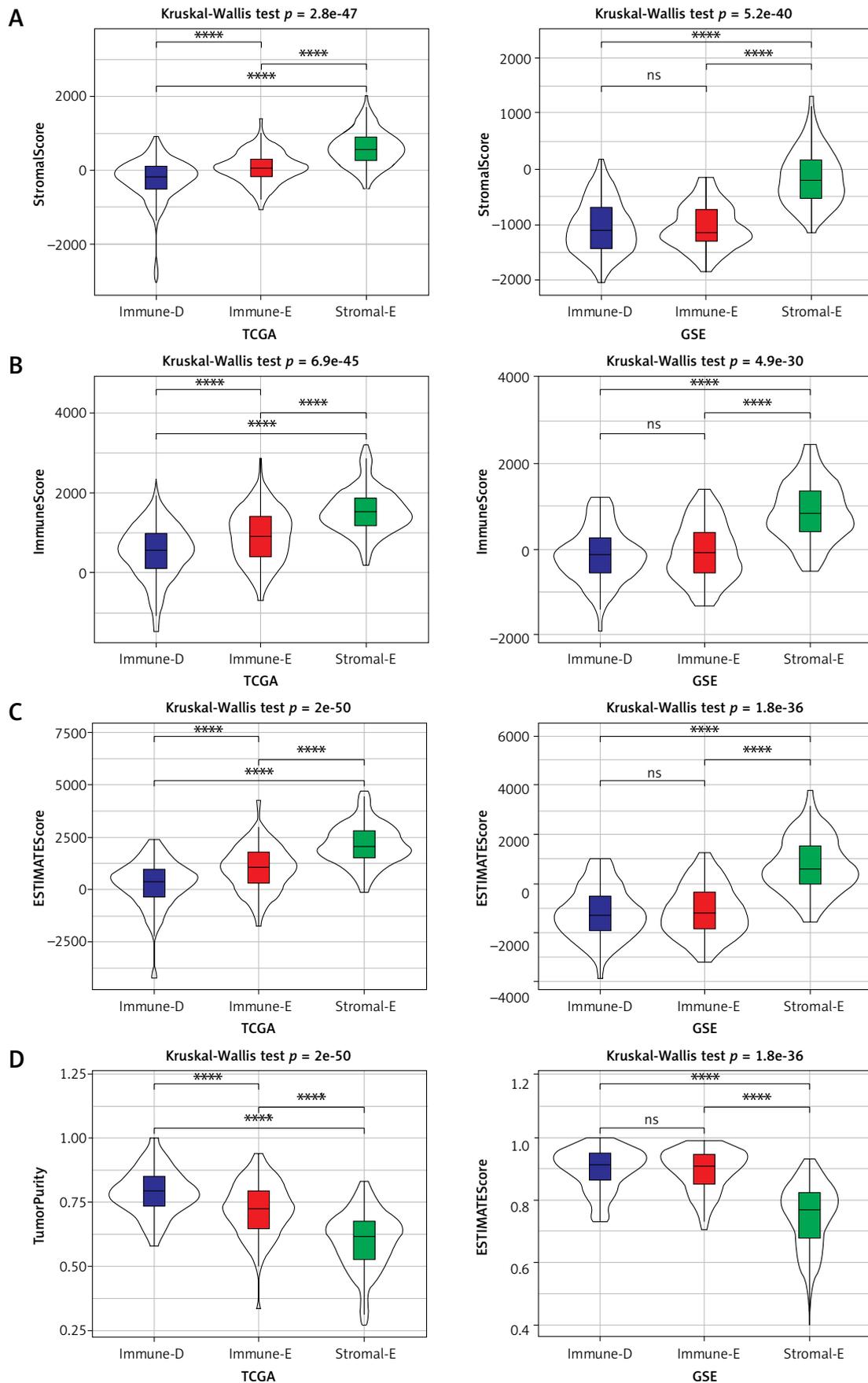


Figure 1. Comparison of immune scores and other characteristics among different subtypes of GBM. **A** – Comparison of the distribution of StromalScore among the subtypes. **B** – Comparison of the distribution of StromalScore among the subtypes. **C** – Comparison of the distribution of ESTIMATEScore among the subtypes. **D** – Comparison of the distribution of TumorPurity among the subtypes

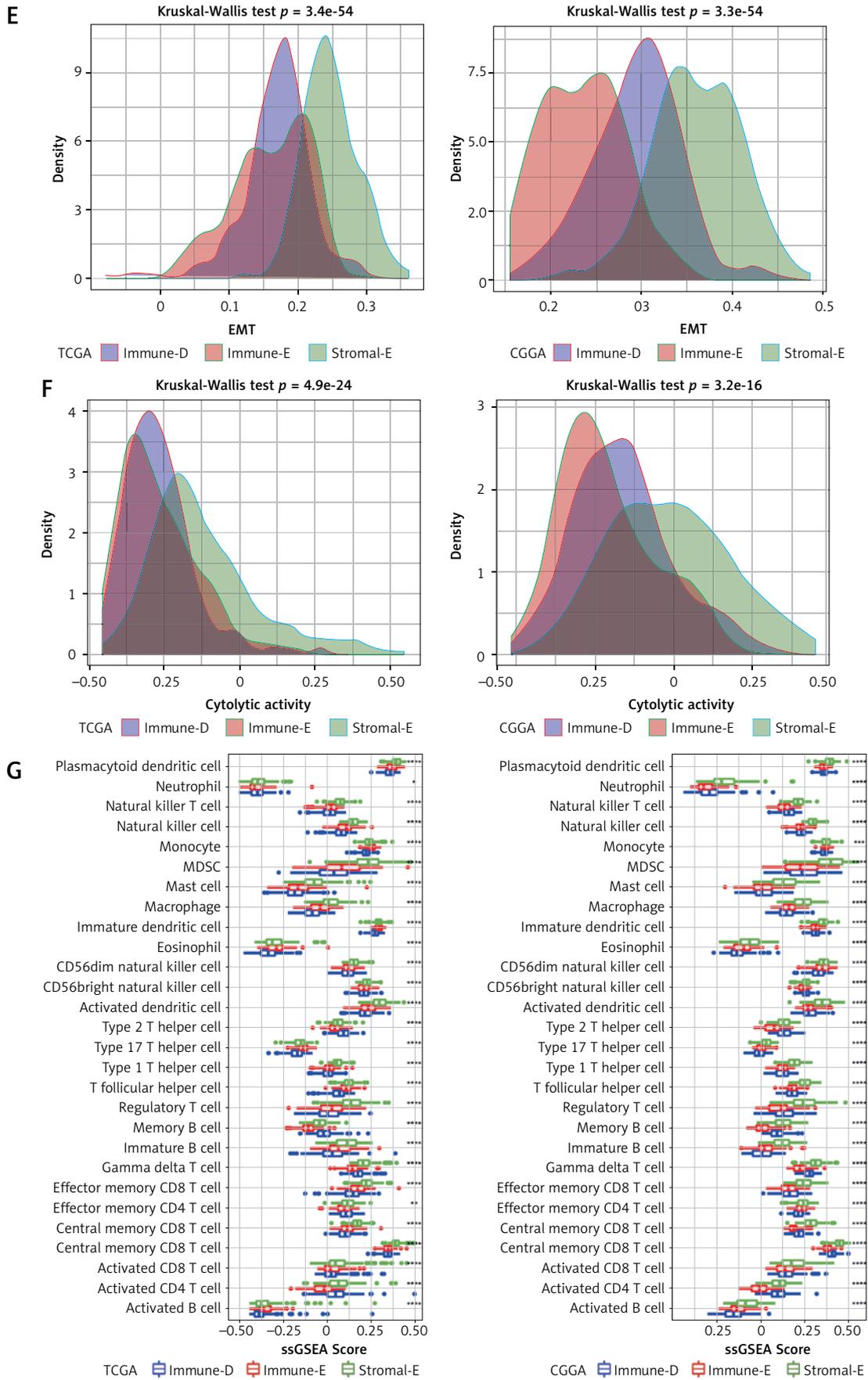


Figure 1 Cont. **E** – Comparison of the distribution of EMT scores among the subtypes. **F** – Comparison of the distribution of cytolytic activity score among the subtypes. **G** – Comparison of the distribution of immune cell scores among the subtypes. * $P < 0.05$; ** $p < 0.01$; *** $p < 0.001$; **** $p < 0.0001$. ns – not significant

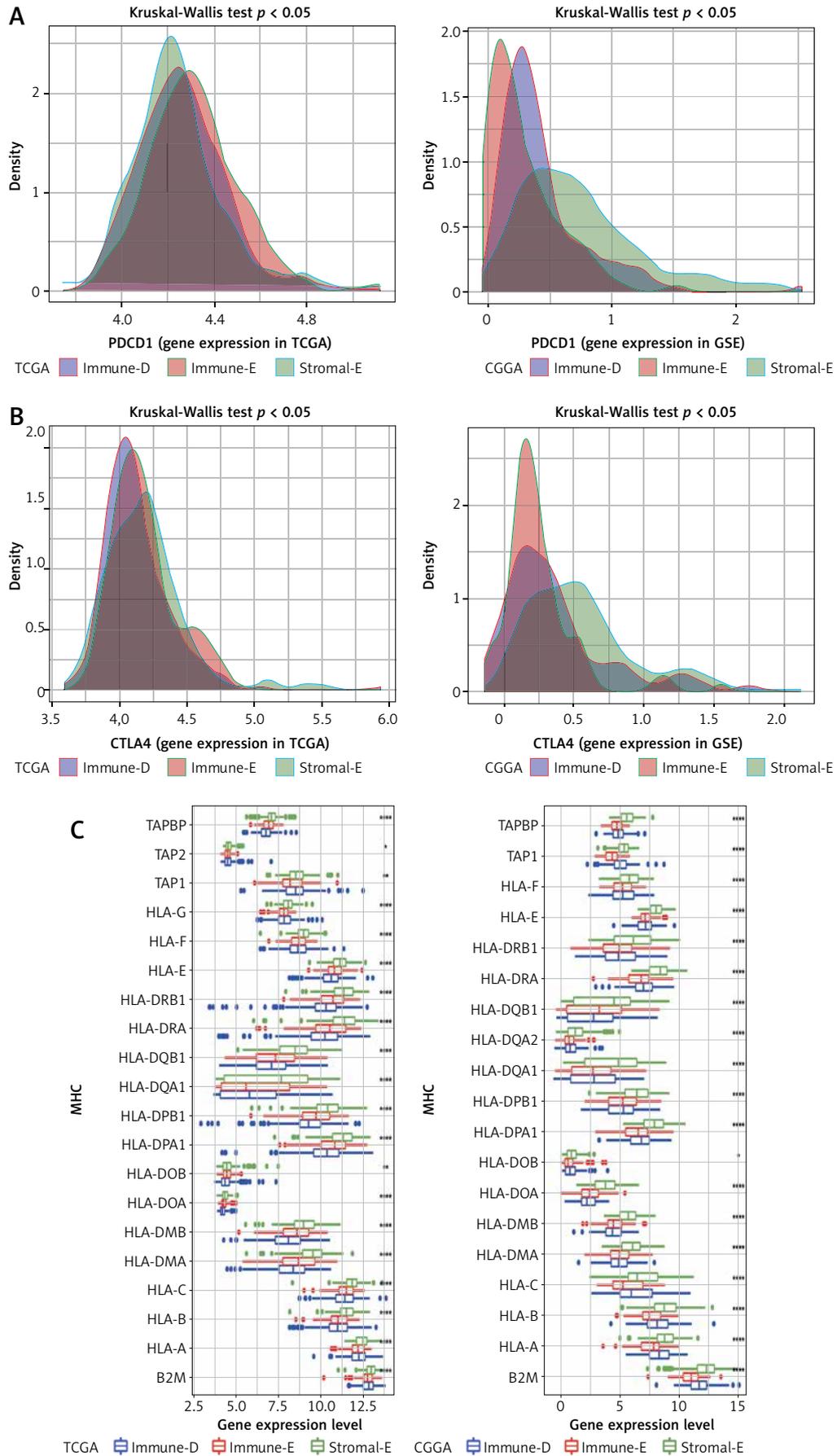


Figure 2. Comparative analysis of immune checkpoint genes and MHC in different subtypes of GBM. **A** – Density distribution of PDCD1 gene in different subtypes of GBM; **B** – Density distribution of CTLA4 gene in different subtypes of GBM; **C** – Comparative distribution of MHC gene in different subtypes of GBM. * $P < 0.05$; ** $p < 0.01$; *** $p < 0.001$; **** $p < 0.0001$. ns – not significant

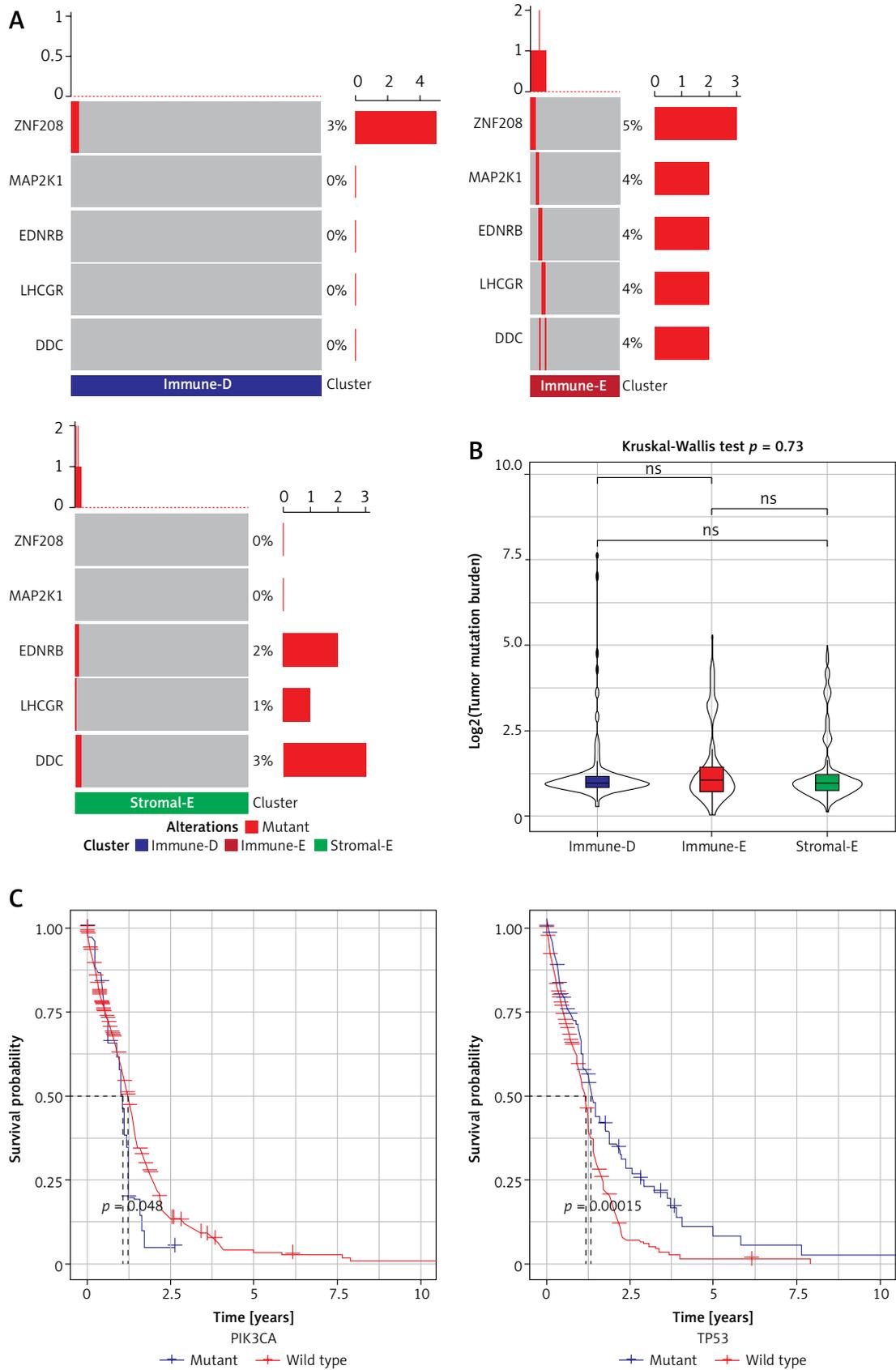


Figure 3. Mutation analysis of cancer driver genes in different subtypes of GBM. **A** – Mutation distribution of some driver genes; **B** – Difference analysis of TMB distribution in different subtypes; **C** – Mutation and wild-type survival analysis of some genes; ns – not significant

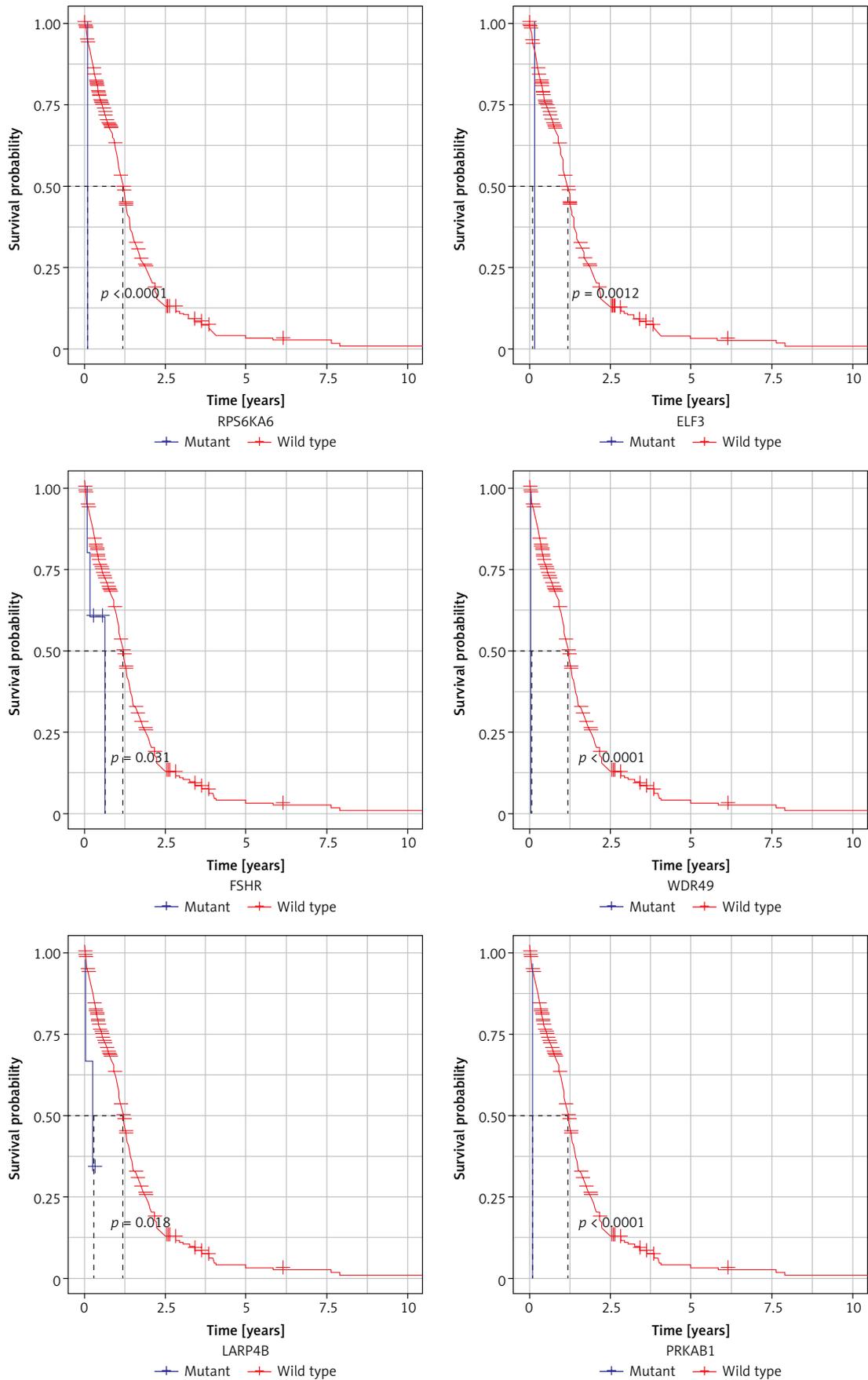


Figure 3. Cont. C – Mutation and wild-type survival analysis of some genes

tigation on the mutation pattern of cancer driver genes was performed with copy number variation (CNV) analysis. The results showed that all the 159 cancer driver genes had amplification deletions, and to more directly show the CNV, we plotted a waterfall diagram, and some results are shown in Figure 4 A. Additionally, we analyzed the TCGA data for LOH (loss of heterozygosity), NtAI (allelic imbalance extending to the telomere), LST (large-scale state transitions) and HRDScore for all three combined, and found that these scores were the lowest in the Stromal-E subtype and highest in the Immune-D subtype (Figures 4 B–E). Also, the expression of the genes corresponding to the subgroup in which gain occurred was higher, while that corresponding to the subgroup in which loss occurred was lower (Figure 4 F).

Clinical features analysis in different subtypes

In the TCGA and CGGA datasets, we compared the distribution of different clinical features in the three subtypes to determine whether the clinical features differed. The results revealed that Stromal-E had the most favorable prognosis, while that of the Immune-D was the worst (Figures 5 A, B). We also compared the distribution of IDH mutation in different subtypes, and found that it was significantly different in the three subtypes, while no significant differences were found for sex, MGMT, or age (Figures 5 C, D).

Methylation analysis and drug sensitivity prediction in different subtypes

To investigate the methylation of cancer driver genes, we collected 7 EMT-promoting genes [24] from an existing study, and then analyzed the differences in the mean methylation β values of these genes in the subtypes. There were no differences in the methylation beta values of VIM, ZEB2, and CLDN1, ZEB1 among the three subtypes, while CDH1 and TWIST1, CDH2 were greatly different (Figure 6 A). Pearson's correlation test demonstrated a clear negative correlation between the methylation degree and expression values of ZEB2 and TWIST1, while the other genes were less correlated (Figure 6 B). Furthermore, we analyzed the differential distribution of β values of cg probe of TWIST1 in the subtypes and observed that cg22498251 and cg20121142 were distributed significantly differently. Other loci are shown in Figure 6 C. The beta values of some cg loci had a significant negative correlation with TWIST1 gene expression, as shown in Figure 6 D.

As predicted by the TIDE online software, the TIDE score of Immune-E in the TCGA dataset was significantly lower than that in Stromal-E

and Immune-D, suggesting higher responsiveness of Immune-E to immunotherapy, whereas the immunotherapy was 74% responsive in Immune-E, which was much higher than that of Stromal-E and Immune-D. The results of the analysis on the CGGA dataset were in accordance with those of the TCGA dataset (Figures 7 A, B). After that, we predicted the responsiveness of different subtypes of TCGA and CGGA datasets to conventional chemotherapeutic drugs (sunitinib, cisplatin, sorafenib and pyrimethamine, crizotinib, paclitaxel), and both showed lower sensitivities of Immune-E (Figures 7 C, D).

WGCNA co-expression analysis

We clustered the GBM samples by the WGCNA package for identification of gene and clustering among different subtypes, and the co-expression network conformed to a scale-free network (Figure 8 A). To ensure that the network was scale-free, we chose $\beta = 14$ (Figures 8 B, C). By calculating the eigenvector values of the modules, we merged the closer modules by setting $\text{minModuleSize} = 80$, $\text{deepSplit} = 2$, $\text{height} = 0.25$, and acquired a total of 6 new modules (Figure 8 D). The gene statistics of each module are shown in Figure 8 E. Further correlation analysis of each module with molecular subtypes showed that the green module was significantly positively associated with Immune-E subtype, and that the brown module was significantly positively associated with Immune-E subtype but negatively associated with Immune-D and Stromal-E subtypes. The yellow module was significantly positively associated with the Stromal-E subtype but significantly negatively associated with the Immune-D subtype (Figure 8 F).

Construction and validation of prognostic model IMscore

To construct a survival model for accurately predicting the prognosis of GBM, we first carried out one-way Cox analysis on 15 pathway genes and obtained 6 prognosis-related pathways. The Pearson correlation between each pathway gene and the pathway was calculated to select genes showing a positive correlation ($\text{Corr} > 0$ and $p < 0.05$) in the pathway for further analysis. For the pathway-related genes, we randomly grouped the TCGA dataset data at a ratio of Train: Test = 3 : 2, and then performed one-way Cox analysis in the Train group to screen the prognosis-related genes, and collected a total of 53 genes. Next, applying Lasso regression and 10-fold cross-validation, we selected 7 genes at $\lambda = 0.1028$ as the target genes based on Lasso regression penalty coefficients. Furthermore, we used stepwise multifactorial regression analysis based

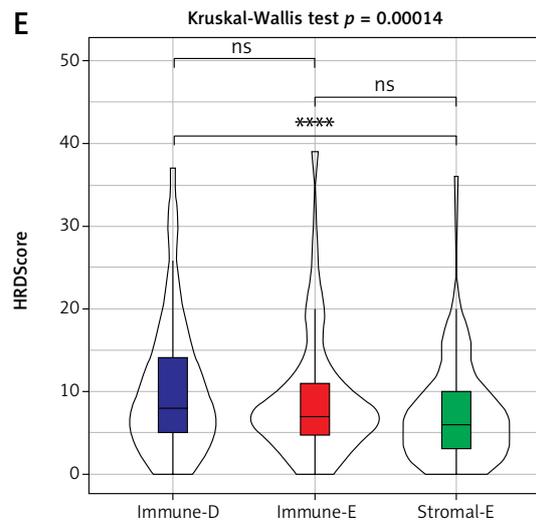
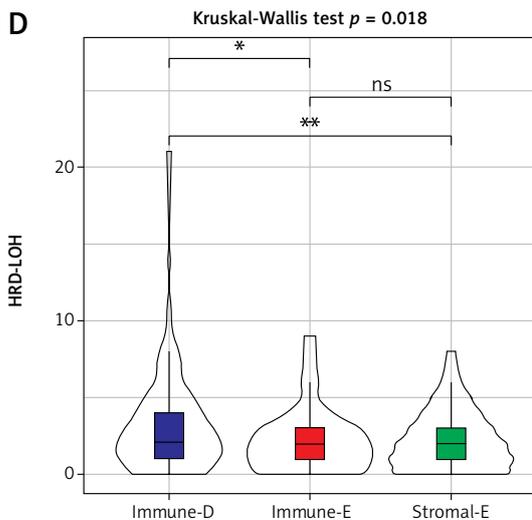
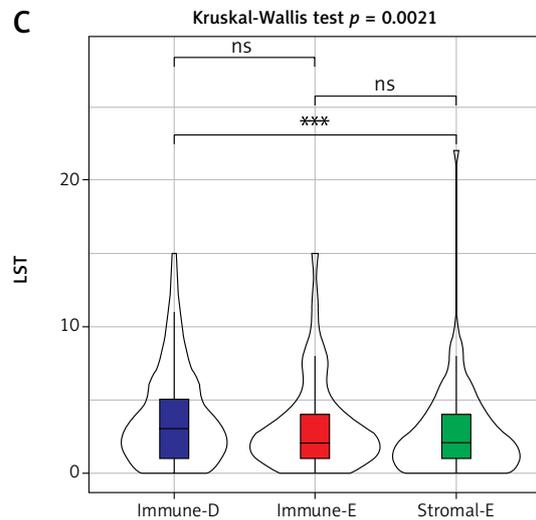
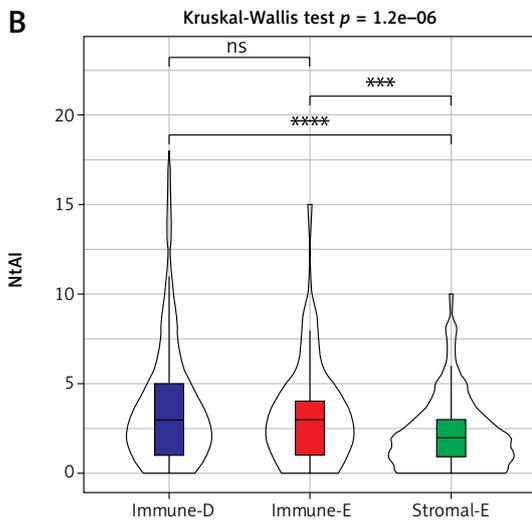
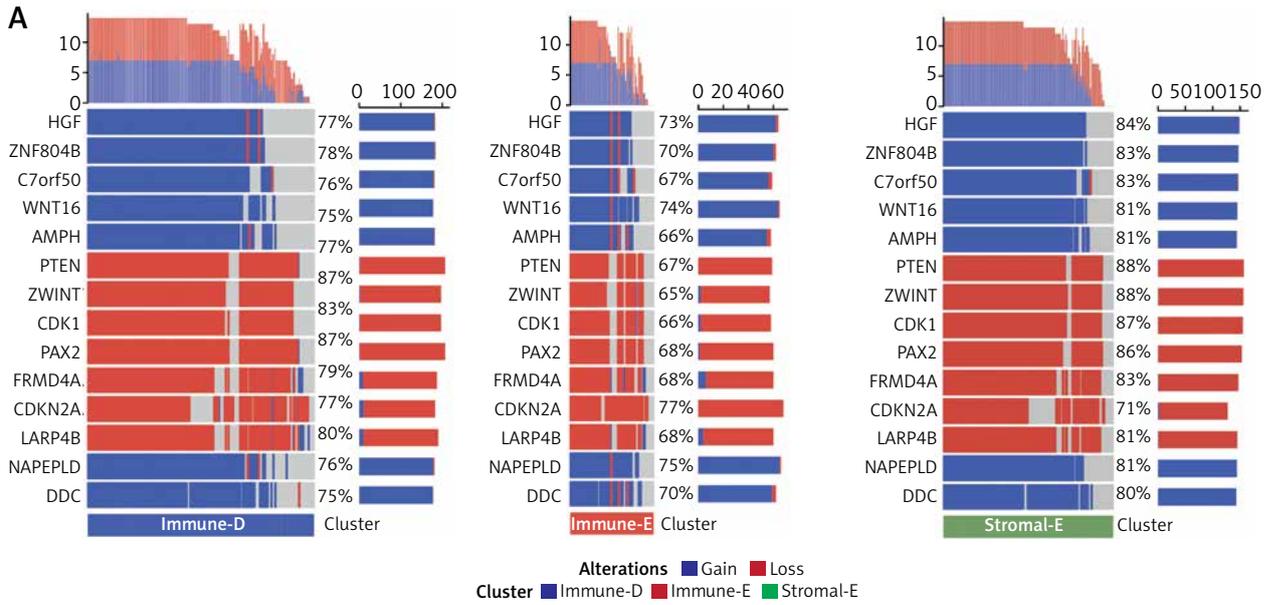


Figure 4. Analysis of gene copy number variation (CNV) in different subtypes of GBM. **A** – CNV distribution of some driver genes, **B–E** – Differences in HRD-related scores of different subtypes; **F** – differences in gene expression of some gene CNV subgroups. * $P < 0.05$; ** $p < 0.01$; *** $p < 0.001$; **** $p < 0.0001$. ns – not significant

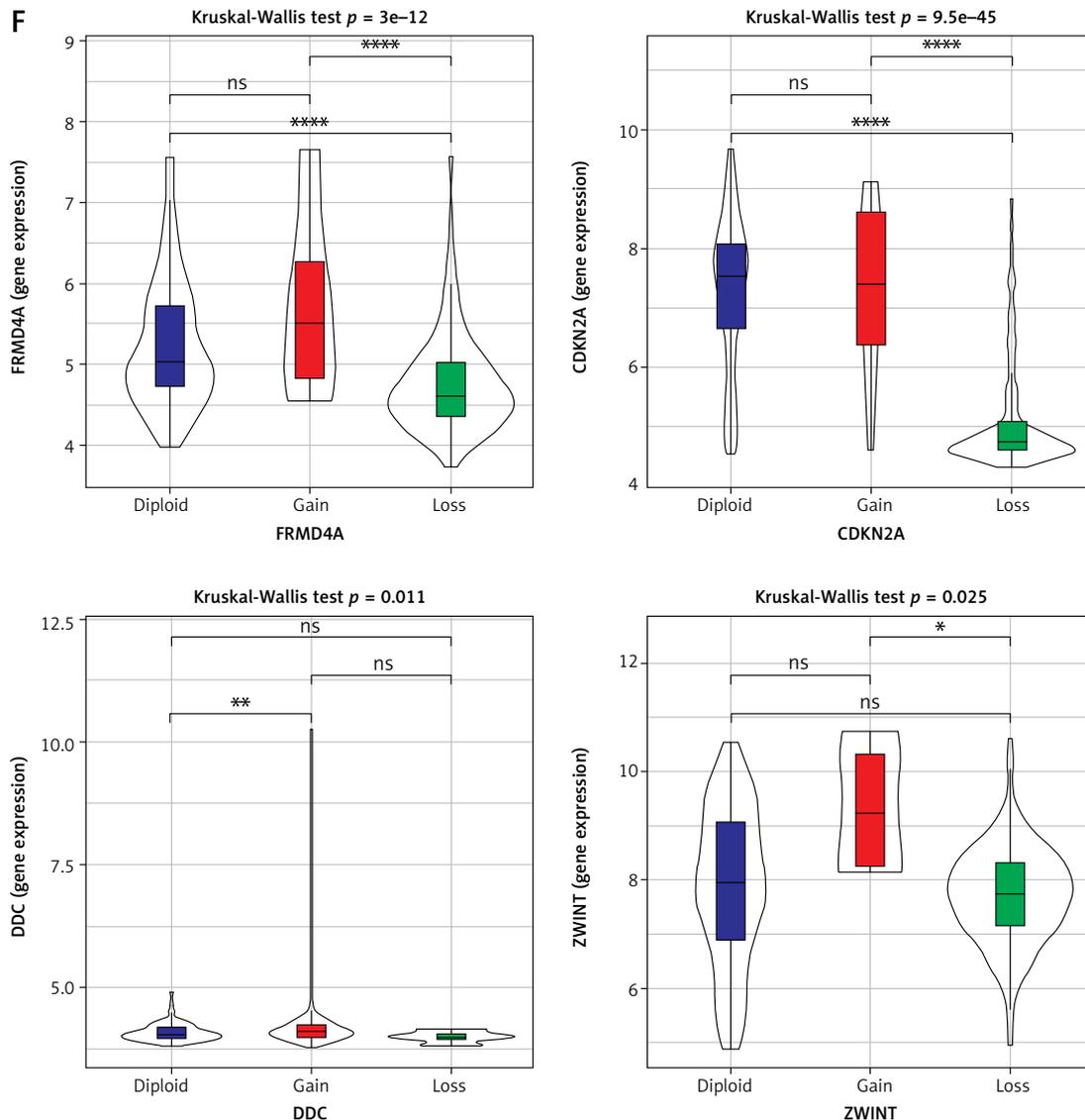


Figure 4. Analysis of gene copy number variation (CNV) in different subtypes of GBM. **A** – CNV distribution of some driver genes, **B–E** – Differences in HRD-related scores of different subtypes; **F** – differences in gene expression of some gene CNV subgroups. * $P < 0.05$; ** $p < 0.01$; *** $p < 0.001$; **** $p < 0.0001$. ns – not significant

on 7 genes from the Lasso analysis results, and finally screened 5 genes and calculated Lasso coefficients for each gene. According to the formula $IMscore = 0.116 * SPP1 + 0.151 * OSMR + 0.174 * NGFR - 0.5 * CTBP2 + 0.324 * CUL1$, IMscore was obtained.

Afterwards, the survminer package was employed to determine the optimal cutoff value, based on which GBM patients were divided into high and low groups. From the survival analysis results, the prognosis of the high-IMscore group was significantly lower than the low-IMscore one ($p < 0.001$), and the TCGA and CGGA datasets performed consistently (Figures 9 A, B). Also, we compared distribution differences of IMscore among the subtypes in different datasets, and observed that IMscore in Stromal-E was the lowest, while in the Immune-D subgroup it was the highest. There was consistency between TCGA and CGGA

datasets (Figures 9 C–F). To verify the stability of the model, we performed validation in TCGA pan-cancer, and it was found that the IMscore model had significant differences between high and low IMscore in all cancer types except ESCA, with a significantly worse prognosis of those with a high IMscore than with a low IMscore ($p < 0.05$) (Supplementary Figure S2).

Performance of IMscore for predicting response to immunotherapy

Immunotherapy-treated datasets CGGA 135222, ICGGA 91061, and Mvigor210 were subjected to IMscore calculation and on the website TIDE (<http://tide.dfci.harvard.edu/>). The predictive effect of IMscore and TIDE on treatment response was compared. Survival curves of IMscore and

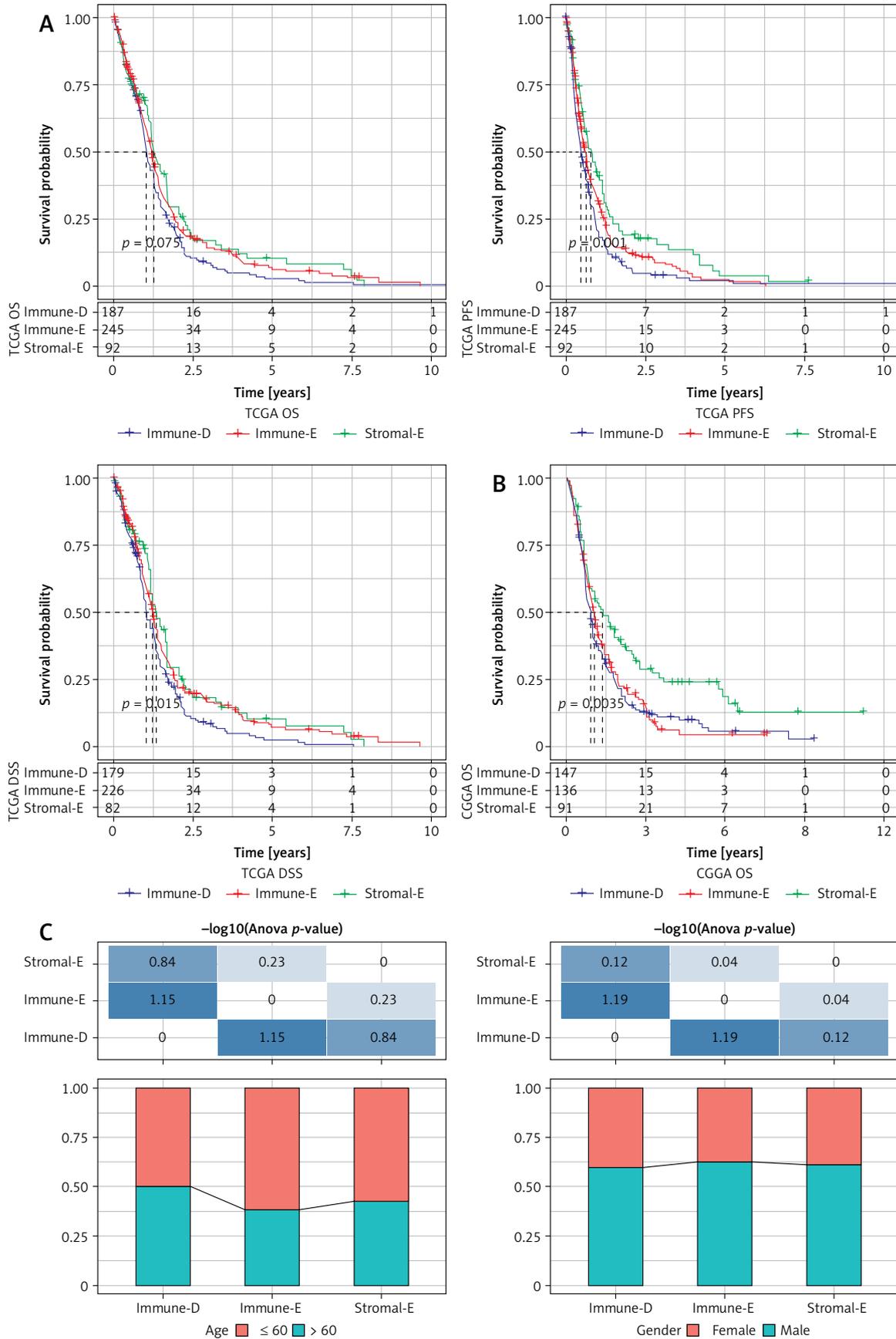


Figure 5. Distribution characteristics of each clinical variable in different molecular subtypes. **A** – Survival curve analysis of different subtypes of GBM in TCGA dataset; **B** – Survival curve analysis of different subtypes of GBM in GSE dataset; **C** – Distribution of clinical characteristics of different subtypes of GBM in TCGA dataset

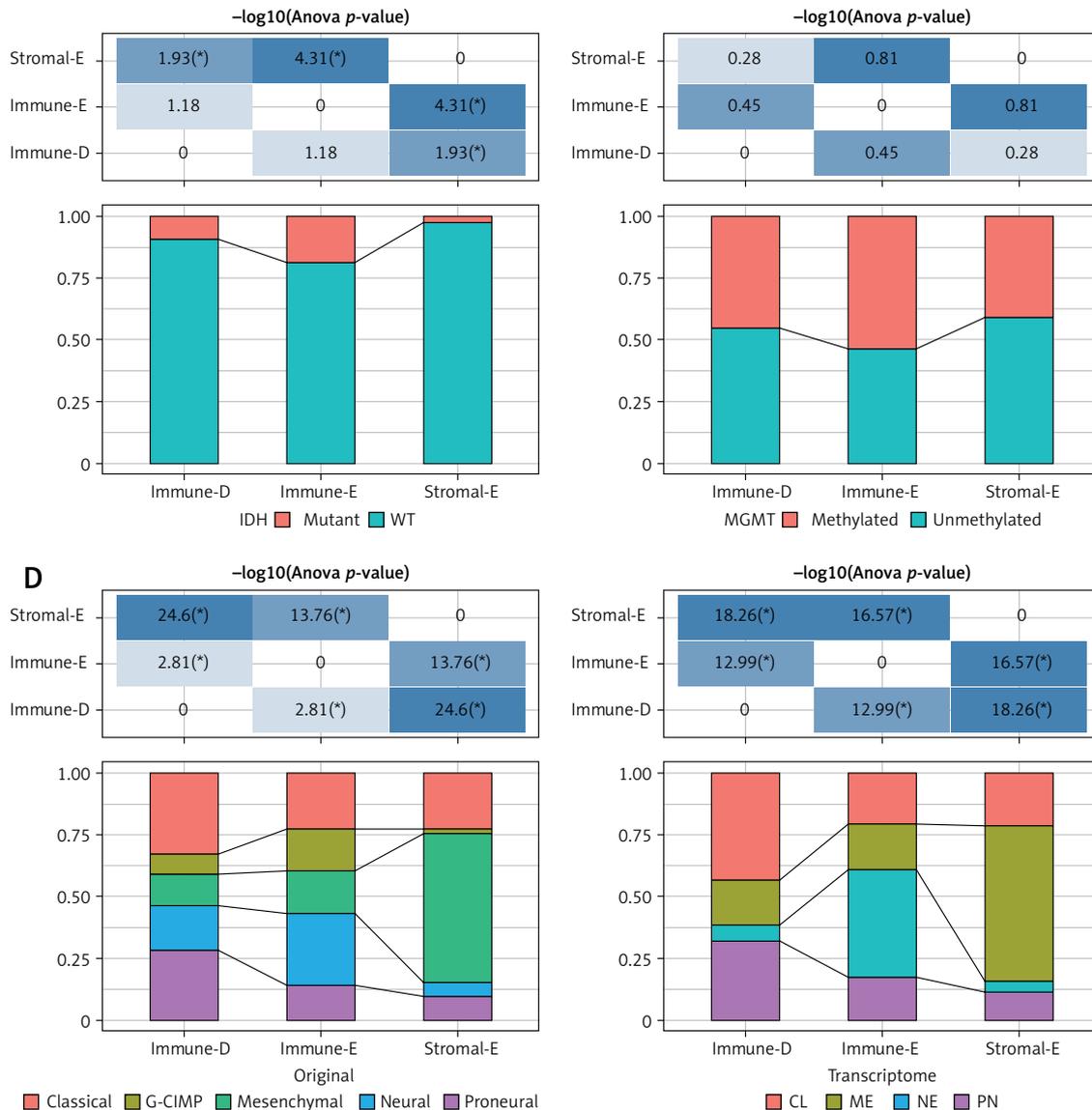


Figure 5. Cont. Distribution characteristics of each clinical variable in different molecular subtypes. **C** – Distribution of clinical characteristics of different subtypes of GBM in TCGA dataset; **D** – Distribution of subtypes in TCGA dataset compared with existing subtypes

TIDE scores and KM curves with median cutoff showed that in the IMvigor210 dataset, patients with low IMscores survived significantly better than those with high IMscores. In predicting immunotherapy responsiveness, the predictive power of IMscore was significantly stronger compared to TIDE score, as demonstrated by the fact that the area under the curve (AUC) of IMscore was significantly larger than that of TIDE score (Figure 10 C). In the CGGA 91061 dataset, patients with low IMscores survived significantly better than patients with high IMscores ($p < 0.05$) (Figures 10 D, E). For predicting immunotherapy responsiveness, IMscore was significantly more effective than TIDE score, as shown by a significantly larger AUC of IMscore than TIDE score (Figure 10 F). In the CGGA 135222 dataset, patients with low IMscores sur-

vived significantly better than those showing high IMscores ($p < 0.05$) (Figure 10 G, H). In terms of predicting immunotherapy responsiveness, IMscore was significantly better than TIDE score, as shown by a significantly larger AUC of IMscore than TIDE score (Figure 10 I). These results indicated that the IMscore model had a stronger performance than the TIDE score in terms of the prognostic and immunotherapy response predictions.

Discussion

GBM is a highly lethal brain tumor that responds poorly to immunotherapy and accounts for about 48.6% of primary malignant brain tumors, with a 5-year survival rate of 7.2% and a 1-year survival rate of around 42.8% [34]. GBM is difficult to treat and is associated with the diversity of the tumor

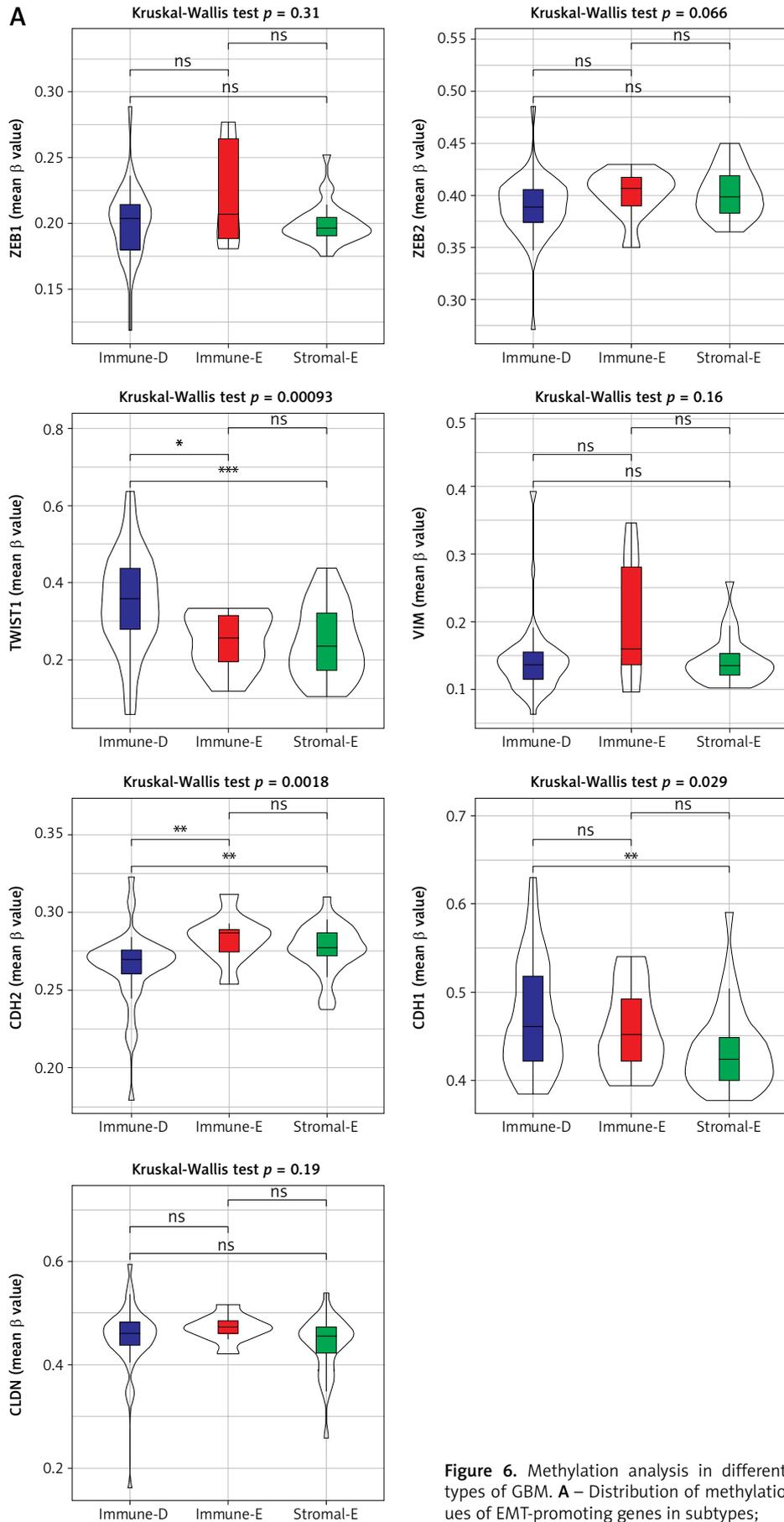


Figure 6. Methylation analysis in different subtypes of GBM. **A** – Distribution of methylation values of EMT-promoting genes in subtypes;

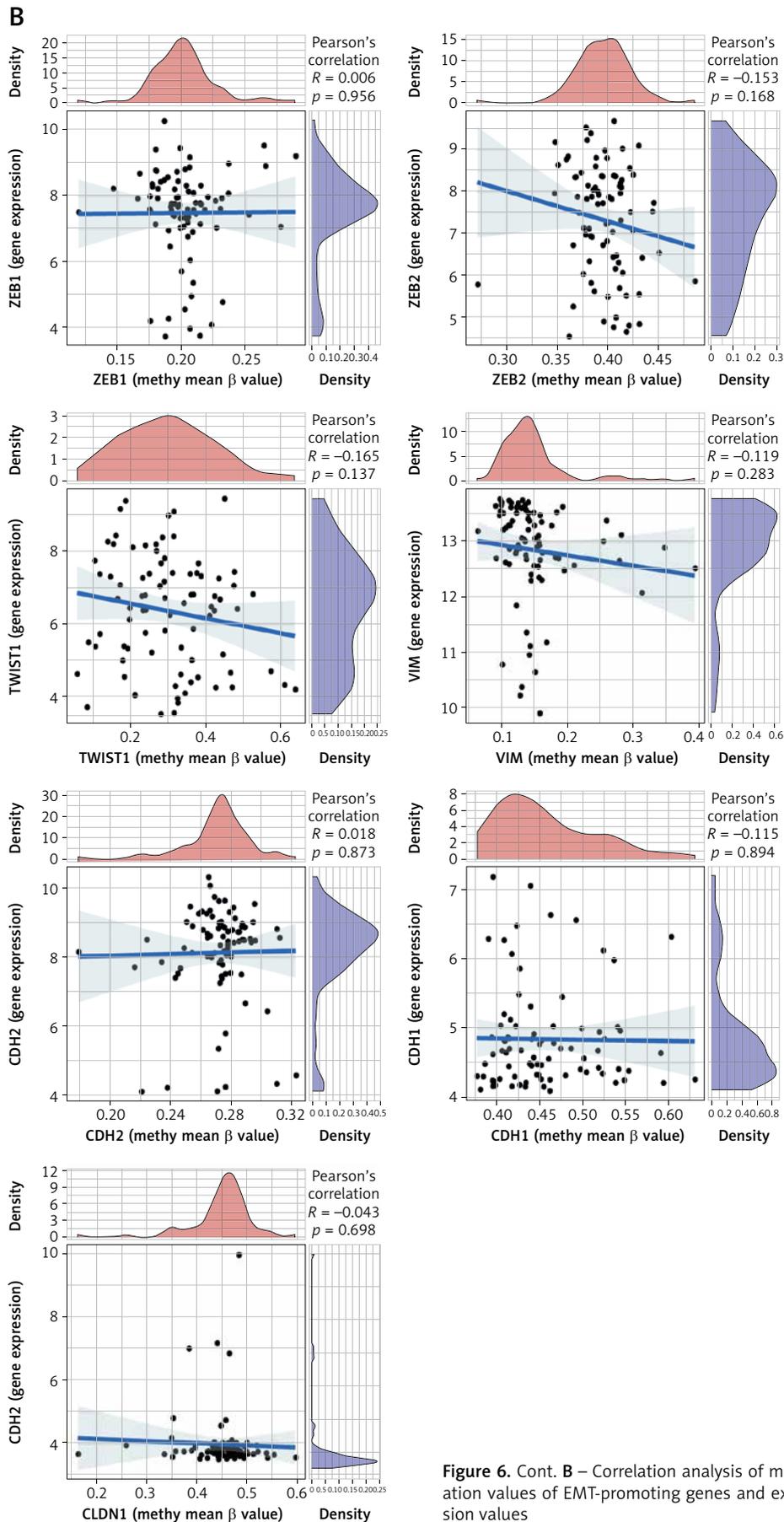


Figure 6. Cont. B – Correlation analysis of methylation values of EMT-promoting genes and expression values

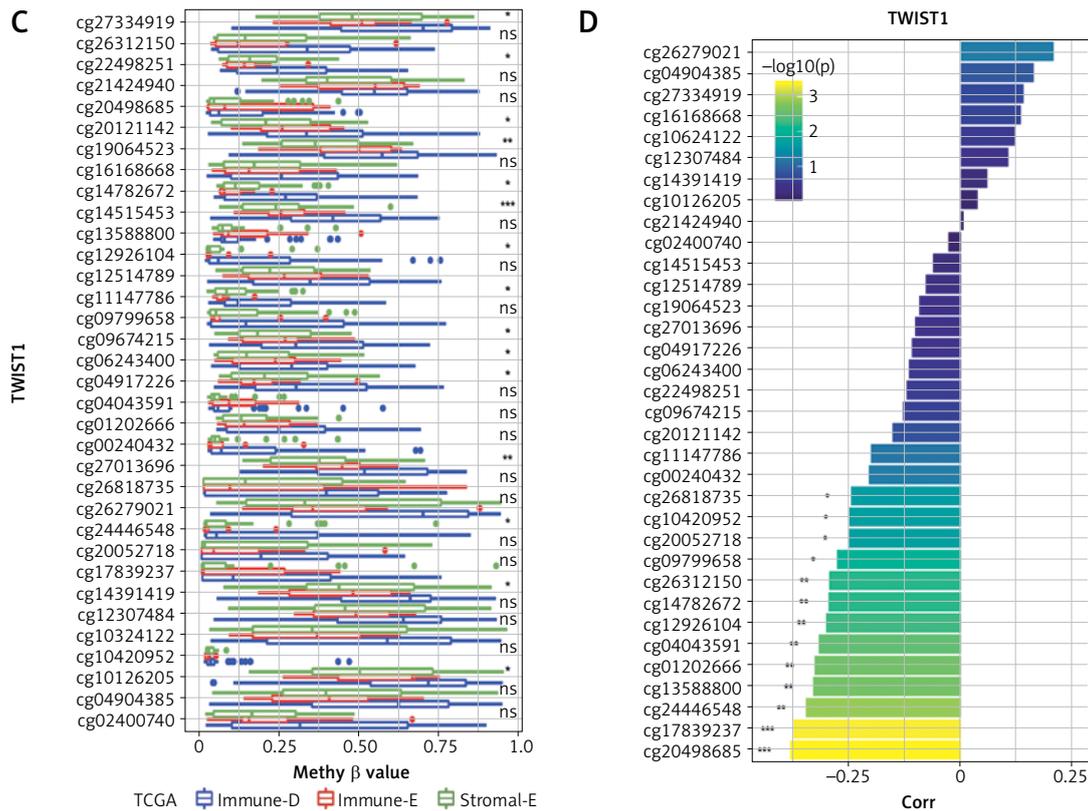


Figure 6. Cont. **C** – Differences in the distribution of β values of cg probe sites of TWIST1 gene in subtypes; **D** – Correlation of cg probe sites of TWIST1 beta values of probe loci correlated with TWIST1 gene expression. * $P < 0.05$; ** $p < 0.01$; *** $p < 0.001$; **** $p < 0.0001$. ns – not significant

microenvironment in which the GBM cell grows. The TME of GBM contains a range of non-tumor cells, including vascular cells, infiltrating and resident immune cells, and other glial cells, with special emphasis on various non-tumor components of the immune system, including stromal cells [35, 36]. Literature data suggested that GBM subtypes with molecular distinction exhibit differences in the microenvironment [37]. Data from GBM models suggested that driver mutations can create a unique GBM microenvironment where TAM can induce lymphocyte infiltration of tumors, positively or negatively influencing clinical treatment [38]. The results of clinical trials conducted so far have shown that the immune microenvironment of GBM was closely related to the outcome of immunotherapy [39, 40]. Therefore, it is important to actively explore immune-based GBM typing to understand the mechanism of tumorigenesis and precisely identify the potential beneficiary.

In the present study, we divided GBM patients into three groups by cluster analysis using immune-related pathways and gene expression. Immune composition, methylation, immunotherapy, and gene mutations varied significantly among the different subtypes, such findings being in accordance with those in a study conducted by

Lee *et al.* [41]. However, what factors contribute to the differences between different subtypes of GBM remained unclear. It has been shown that tumor-specific antigens or tumor-associated driver genetic mutations presented differently in different subtypes, resulting in a variety of molecular immune responses and leading to differential accumulation of immune cells [42, 43]. In this study, it was observed from CNV analysis that GBM tumor driver genes showed gene amplification deletions, and that LST, LOH, and NtAI scored differently among the subtypes, implying that variation in driver genes was intrinsic to different GBM subtypes. Interestingly, no significant difference was detected in TMB among different GBM subtypes, which was considered to be related to the therapeutic interventions, number selection, and sample size in clustering. These negative results also provided a new direction for clinical and basic research.

Methylation of GBM, especially MGMT methylation, has been extensively studied [44, 45]. Researchers have conducted methylation studies in different GBM subgroups, including age, gender and even race and studied in detail the correlation between the prognosis of different treatment regimens and MGMT methylation [46]. Consistent

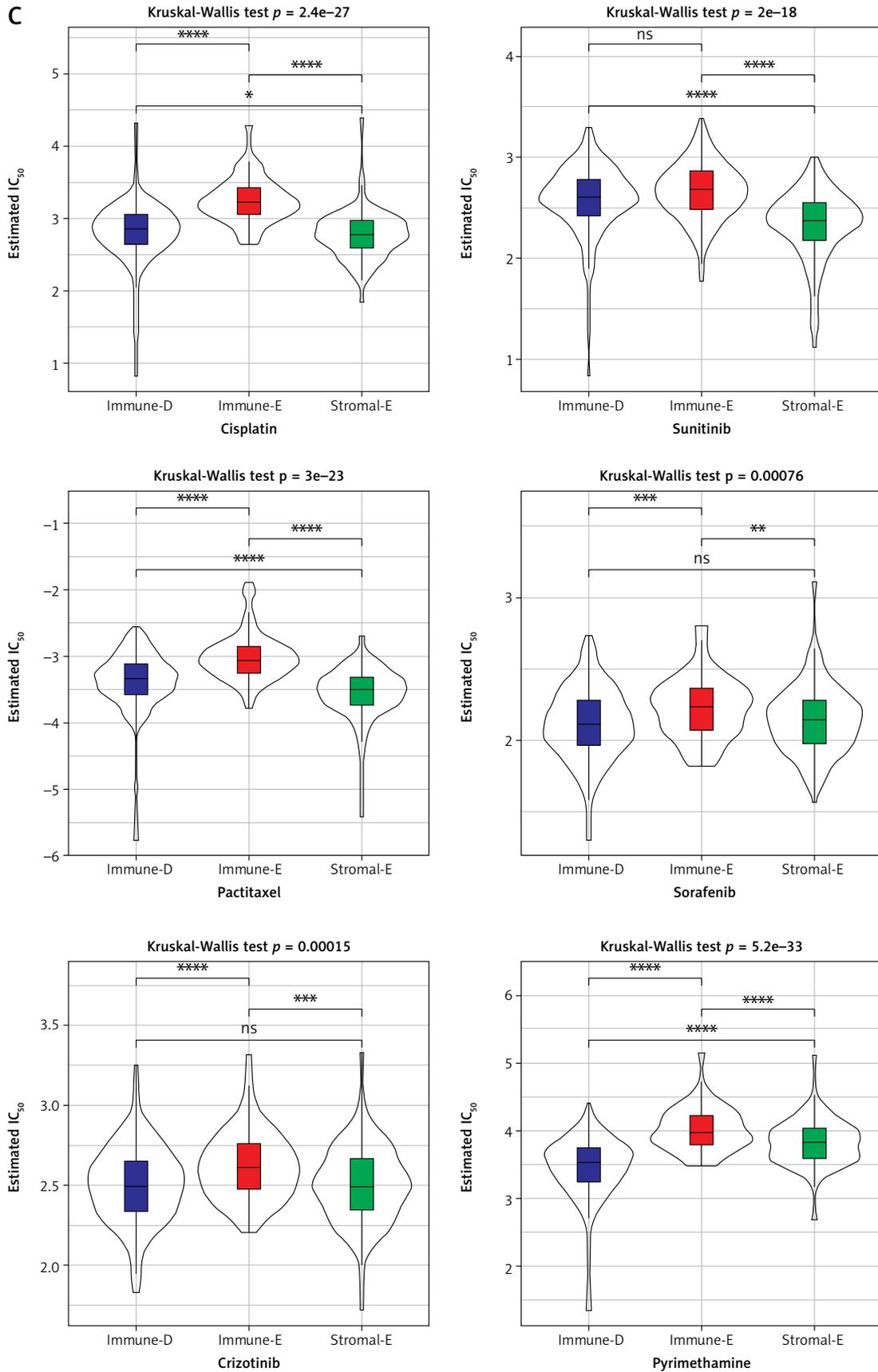


Figure 7. Cont. C – Differential analysis of drug IC_{50} for different subtypes of GBM in TCGA

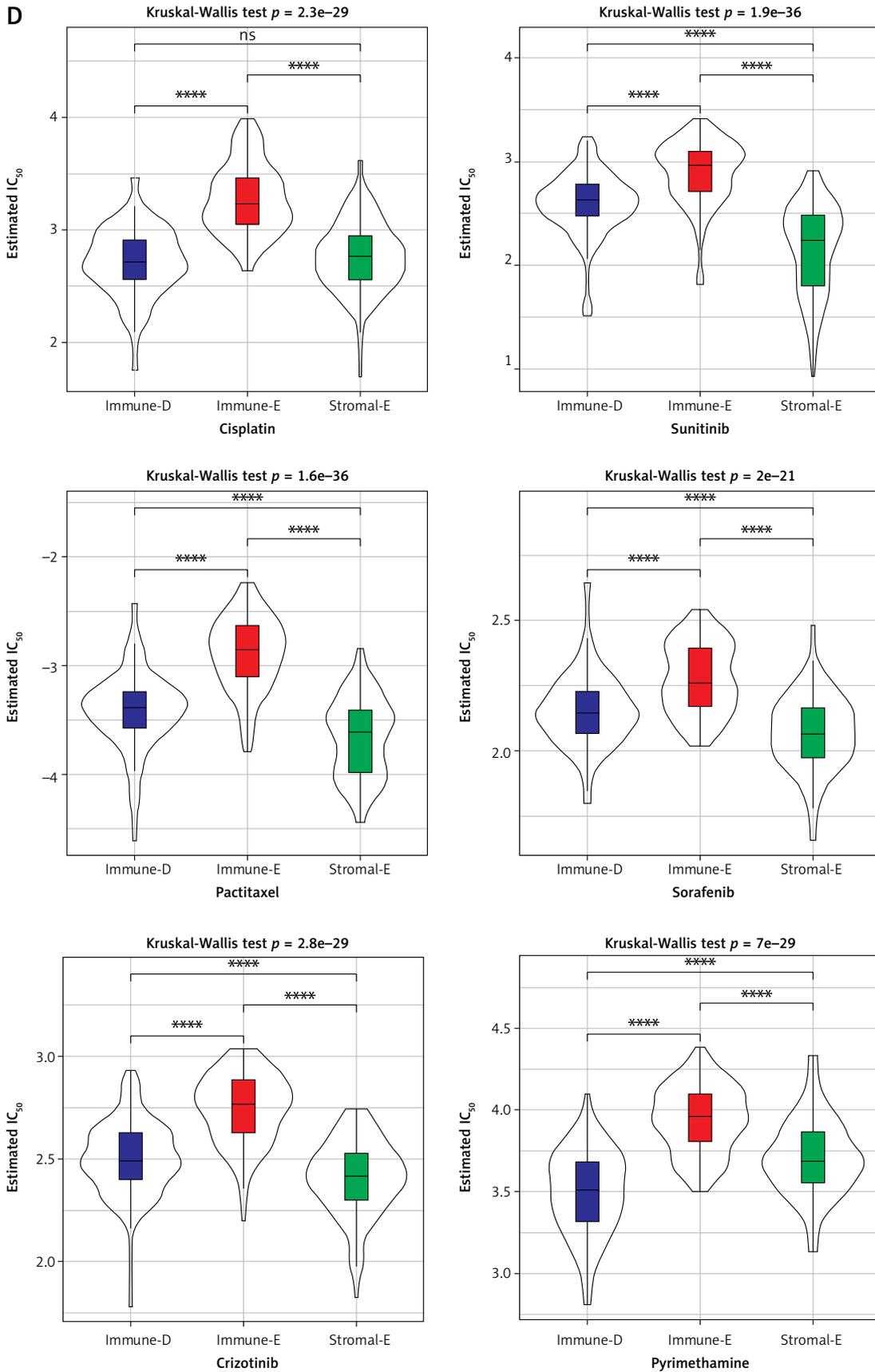


Figure 7. Cont. **D** – Differential analysis of drug IC_{50} for different subtypes of GBM in CGGA. * $P < 0.05$; ** $p < 0.01$; *** $p < 0.001$; **** $p < 0.0001$. ns – not significant

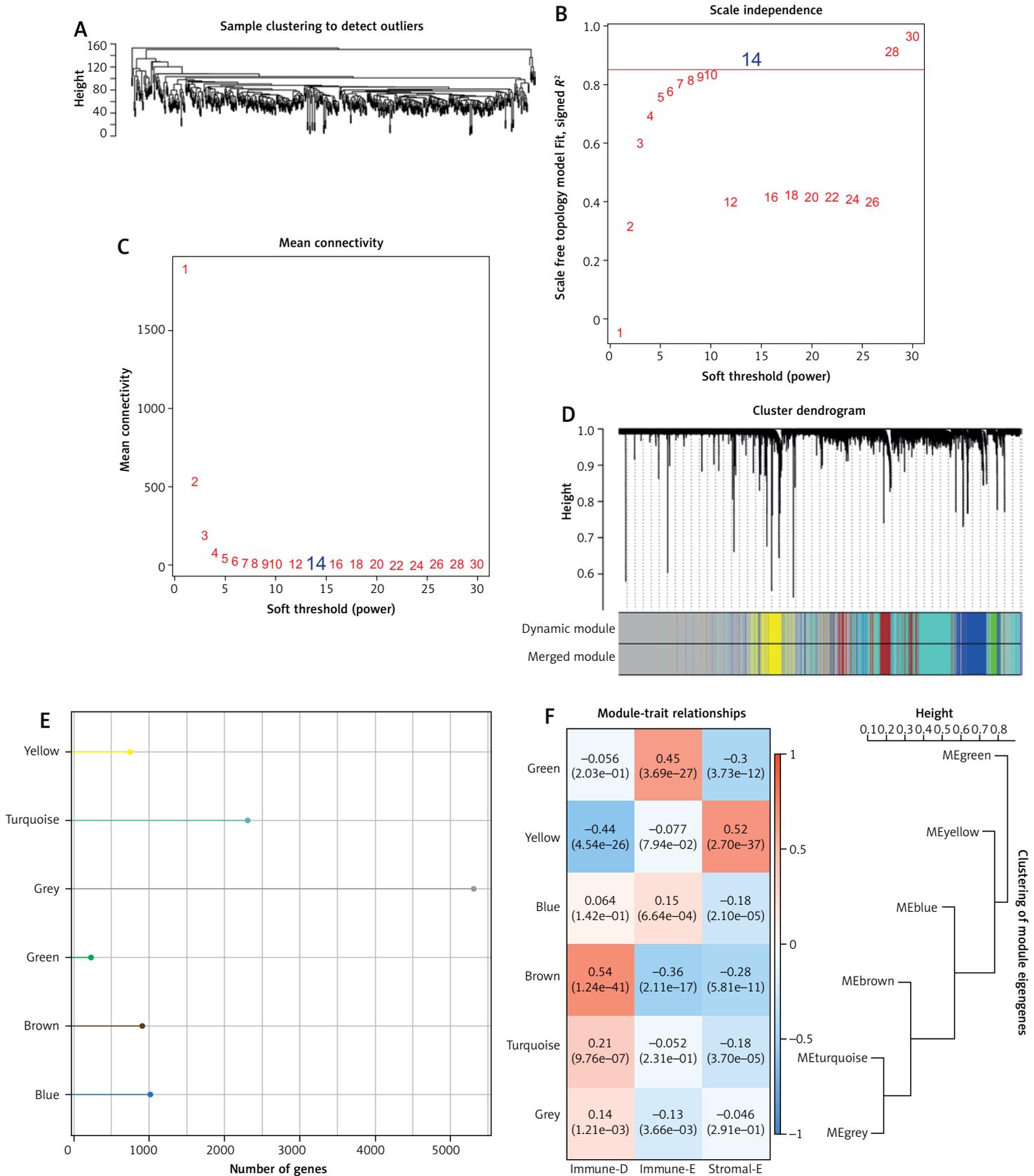


Figure 8. WGCNA analysis. **A** – Cluster tree of individual samples; **B** – Analysis of the scale-free fit index for various soft-thresholding powers (β); **C** – Analysis of the mean connectivity for various soft-thresholding powers; **D** – Dendrogram of all differentially expressed genes clustered based on a dissimilarity measure (1-TOM); **E** – The number of genes in each module; **F** – Correlation of the module feature vectors of each module with clinical data

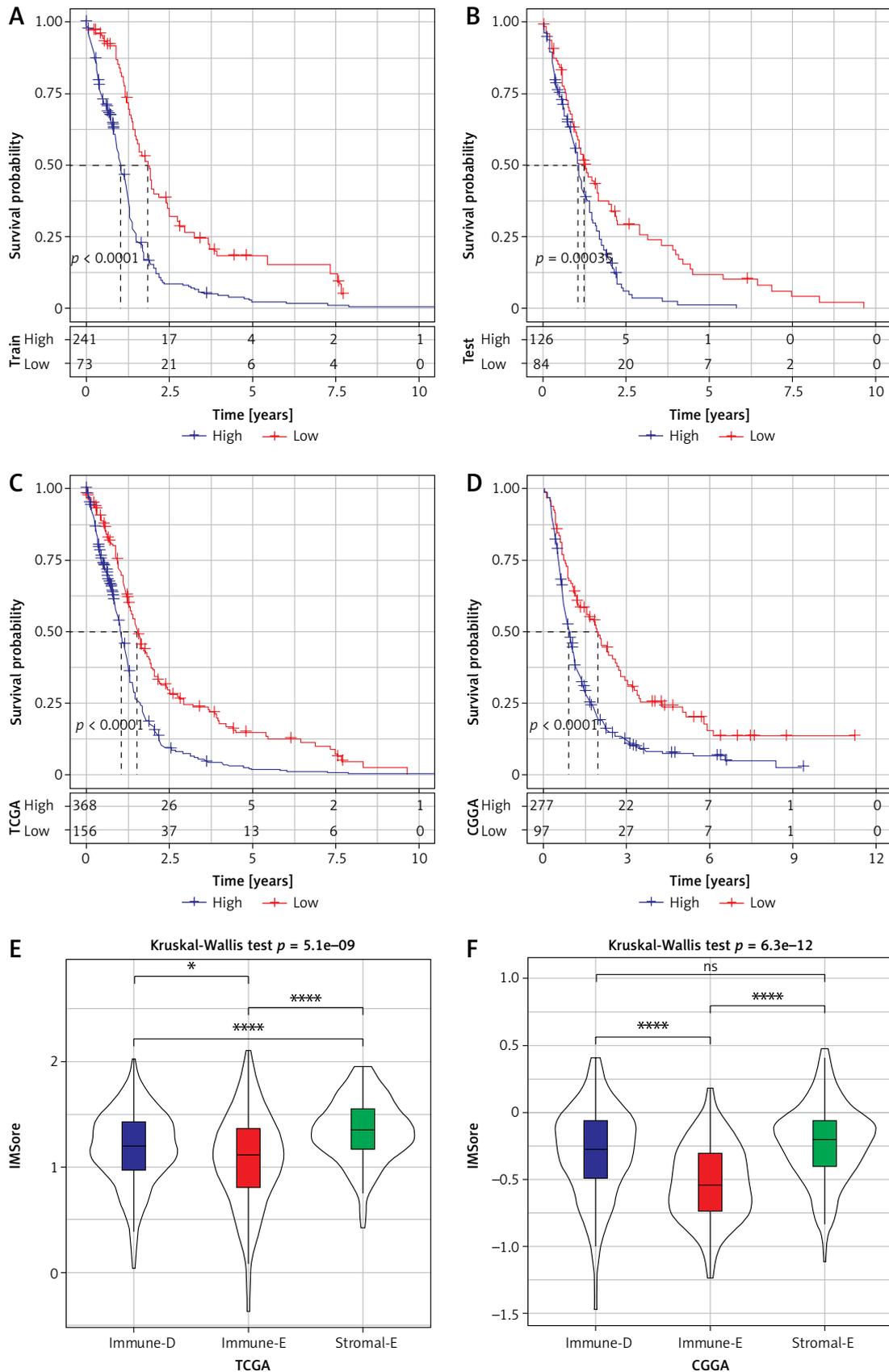


Figure 9. Construction of prognostic IMScore and survival analysis. **A** – KM survival curve of train dataset; **B** – KM survival curve of test dataset; **C** – KM survival curves of CGGA dataset; **D** – KM survival curves of TCGA dataset; **E** – Distribution of IMScore in TCGA dataset subtypes; **F** – Distribution of IMScore in subtypes of CGGA dataset. * $P < 0.05$; ** $p < 0.01$; *** $p < 0.001$; **** $p < 0.0001$. ns – not significant

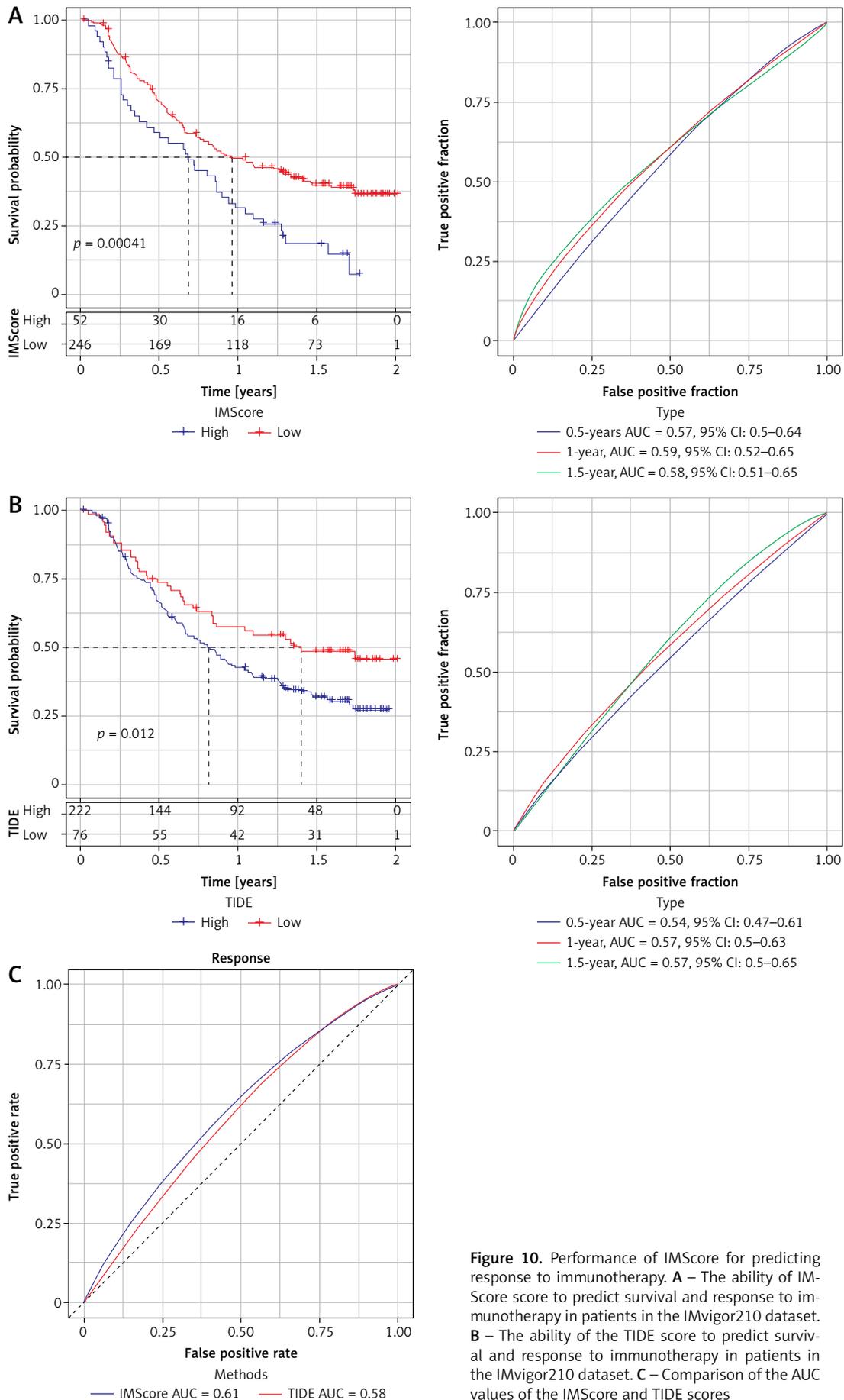


Figure 10. Performance of IMScore for predicting response to immunotherapy. **A** – The ability of IMScore score to predict survival and response to immunotherapy in patients in the IMvigor210 dataset. **B** – The ability of the TIDE score to predict survival and response to immunotherapy in patients in the IMvigor210 dataset. **C** – Comparison of the AUC values of the IMScore and TIDE scores

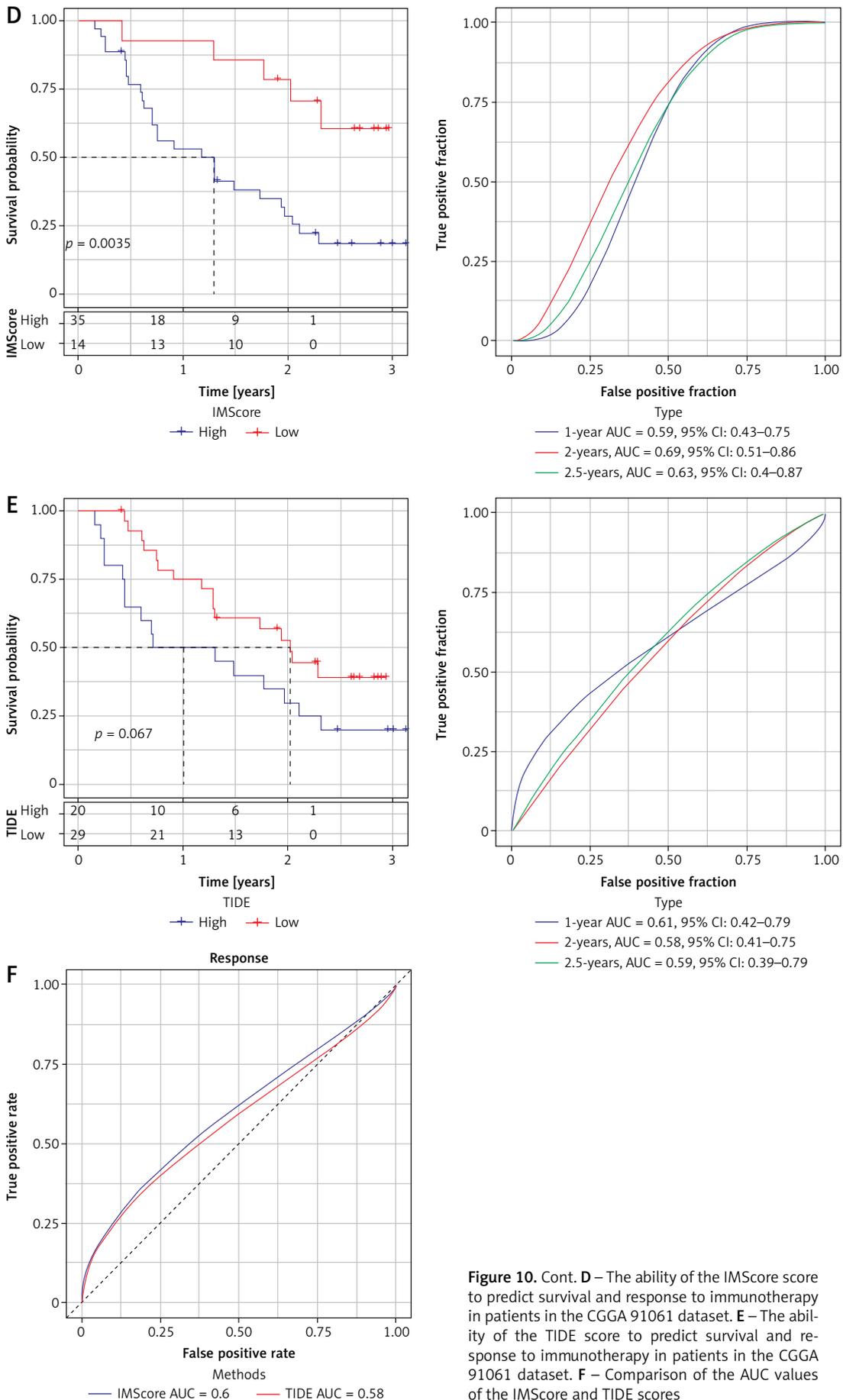


Figure 10. Cont. **D** – The ability of the IMScore score to predict survival and response to immunotherapy in patients in the CGGA 91061 dataset. **E** – The ability of the TIDE score to predict survival and response to immunotherapy in patients in the CGGA 91061 dataset. **F** – Comparison of the AUC values of the IMScore and TIDE scores

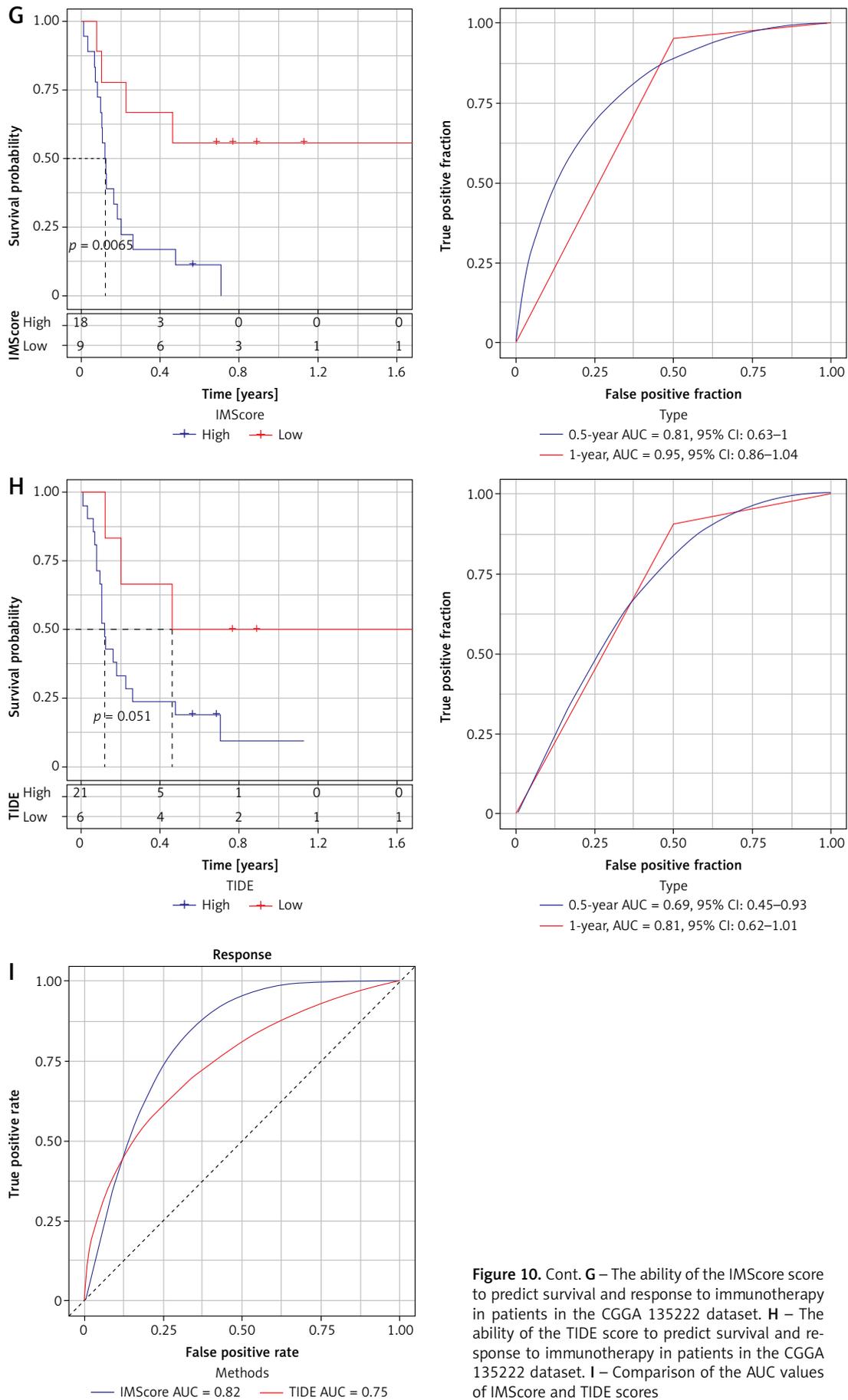


Figure 10. Cont. **G** – The ability of the IMScore score to predict survival and response to immunotherapy in patients in the CGGA 135222 dataset. **H** – The ability of the TIDE score to predict survival and response to immunotherapy in patients in the CGGA 135222 dataset. **I** – Comparison of the AUC values of IMScore and TIDE scores

with previous studies, we attempted to analyze the causes of different subtypes of GBM from the perspective of methylation, and found that CDH2, TWIST1, and CDH1 differed significantly among the different subtypes. Moreover, from Pearson's correlation test, a significant negative correlation could be observed between the degree of methylation and expression values of TWIST1 and ZEB2. The results provided new evidence for the causal relationship between methylation and differences in GBM among the subtypes. In addition to this, methylation of GBM genes can suppress the expression of certain genes, leading to cell death after alkylating agent injury, which was also correlated significantly with poor treatment response [47].

PD-1/PD-L1 checkpoint inhibitors have shown efficacy in the treatment of solid tumors. Immunotherapy has not achieved a breakthrough in GBM because of the weak immunogenic response but could suppress the immune microenvironment induced by immune cells and cytokines [48, 49]. A study showed that the expression of PD-L1 is positively correlated with glioma grade, with high expression of PD-1 /PD-L1 suggesting poor prognosis in GBM patients [50]. Emerging evidence indicated that higher stromal populations are associated with tumor progression through remodeling anti-tumor immunity and responsiveness to immunotherapy [51]. To better detect potential immune beneficiaries, we divided patients into three subtypes; a much higher proportion of positive responses occurred in Immune-E than in Stromal-E and Immune-D, suggesting that this group of patients was more responsive to immunotherapy. Overall, in GBM, it is necessary to establish a subgroup index containing multiple molecular markers for evaluating the therapeutic effects of PD-1/PD-L1 checkpoint blocker-based immunotherapy and conventional chemotherapy, which can prolong the survival time of GBM patients and achieve the goal of precision medicine.

The construction of risk models helps to identify GBM high-risk groups more accurately. In the current study, we successfully developed a prediction model via calculating the expression and expression coefficients of five genes and used the model to divide the patients into two groups of high and low risk. Among them, SPP1 was overexpressed in multiple cancers and was involved in tumor immunity via promoting invasion, migration, and cancer stem cell self-renewal to regulate tumor progression. In GBM cells, up-regulated SPP1 expression was associated with reactivation of GBM transcriptional regulation, tumor cell self-renewal and multipotency [52]. OSMR regulates brain tumor stem cell proliferation and GBM development, while OSMR deletion sensitizes

brain tumor stem cells to radiotherapy-induced cell death. Inhibition of OSMR has been found to improve GBM response to treatment and prolongs patients' survival [53]. CTBP2 is highly expressed in GBM, affecting the biological properties of GBM cells and tumor growth; moreover, it is associated with GBM prognosis [54]. Studies on NGFR and CUL1 in GBM are limited and demand further exploration. The brain was previously considered an immune-privileged organ due to the unique immune-related organization of the central nerve system: (i) lack of in situ dendritic cells and lymphatic flow; (ii) it has a blood-brain barrier to maintain the stability of the internal environment of the central nerve system [8, 55]. However, the development, treatment and prognosis of GBM have been found to be closely related to the immune function and the level of immune-related molecules in the patient's organism [56, 57]. How to effectively predict and precisely identify beneficiaries of immunotherapy is currently a great challenge to be addressed in clinical practice. Previous studies have used TIDE as a tool to estimate the possibility of tumor immune escape in the gene expression profile of tumor samples, representing an important tool for predicting immune response, but this predictor has limitations such as poor specificity [58]. Our IMscore scoring model outperformed the TIDE score in predicting immunotherapy response and patients' prognosis, providing a new strategy for the prediction of immunotherapy in GBM. The IMscore can distinguish patients with a good or poor prognosis, and it outperformed the TIDE score in predicting patients' prognosis. Immune-E subtype (the subtype most sensitive to immunotherapy) showed the lowest IMscore, suggesting that the low-IMscore group was more sensitive to immunotherapy than the high-IMscore group. With this subtyping and scoring system, clinicians may apply more precise treatment for the patients.

In conclusion, GBM samples from the TCGA and CGGA datasets were subtyped based on the scores of oncogenic pathways, stromal pathways, immune pathways, and DNA repair pathways. The subtyping system shows potential to guide clinical immunotherapy. GBM patients with Immune-E subtype may benefit more from immunotherapy. Furthermore, a prognostic model was built using prognosis-related pathway genes to predict overall survival, which also showed strong robustness and favorable results in pan-cancer, demonstrating a great potential for predicting immunotherapy response in cancer.

Conflict of interest

The authors declare no conflict of interest.

References

1. Lah TT, Novak M, Breznik B. Brain malignancies: Glioblastoma and brain metastases. *Semin Cancer Biology* 2020; 60: 262-73.
2. Tan AC, Ashley DM, López GY, Malinzak M, Friedman HS, Khasraw M. Management of glioblastoma: State of the art and future directions. *CA Cancer J Clin* 2020; 70: 299-312.
3. Di Nunno V, Franceschi E, Tosoni A, et al. Treatment of recurrent glioblastoma: state-of-the-art and future perspectives. *Exp Rev Anticancer Ther* 2020; 20: 785-95.
4. Yang T, Kong Z, Ma W. PD-1/PD-L1 immune checkpoint inhibitors in glioblastoma: clinical studies, challenges and potential. *Human Vaccines Immunotherapeutics* 2021; 17: 546-53.
5. Maghrouni A, Givari M, Jalili-Nik M, et al. Targeting the PD-1/PD-L1 pathway in glioblastoma multiforme: preclinical evidence and clinical interventions. *Int Immunopharmacol* 2021; 93: 107403.
6. Kim MM, Umemura Y, Leung D. Bevacizumab and glioblastoma: past, present, and future directions. *Cancer J* 2018; 24: 180-6.
7. Diaz RJ, Ali S, Qadir MG, De La Fuente MI, Ivan ME, Komotar RJ. The role of bevacizumab in the treatment of glioblastoma. *J Neurooncol* 2017; 133: 455-67.
8. Chen Z, Hambardzumyan D. Immunol microenvironment in glioblastoma subtypes. *Front Immunology*. 2018; 9: 1004.
9. Dapash M, Hou D, Castro B, Lee-Chang C, Lesniak MS. The interplay between glioblastoma and its microenvironment. *Cells* 2021; 10: 2257.
10. DeCordova S, Shastri A, Tsolaki AG, et al. Molecular heterogeneity and immunosuppressive microenvironment in glioblastoma. *Front Immunol* 2020; 11: 1402.
11. Srivastava S, Jackson C, Kim T, Choi J, Lim M. A Characterization of dendritic cells and their role in immunotherapy in glioblastoma: from preclinical studies to clinical trials. *Cancers* 2019; 11: 537.
12. Avril T, Vauleon E, Hamlat A, et al. Human glioblastoma stem-like cells are more sensitive to allogeneic NK and T cell-mediated killing compared with serum-cultured glioblastoma cells. *Brain Pathol* 2012; 22: 159-74.
13. Li L, Zhu X, Qian Y, et al. Chimeric antigen receptor T-cell therapy in glioblastoma: current and future. *Front Immunol* 2020; 11: 594271.
14. McCord M, Mukouyama YS, Gilbert MR, Jackson S. Targeting WNT signaling for multifaceted glioblastoma therapy. *Front Cellular Neurosci* 2017; 11: 318.
15. Shahcheraghi SH, Tchokonte-Nana V, Lotfi M, Lotfi M, Ghorbani A, Sadeghnia HR. Wnt/beta-catenin and PI3K/Akt/mTOR signaling pathways in glioblastoma: two main targets for drug design: a review. *Curr Pharmaceutical Design* 2020; 26: 1729-41.
16. Erasmus H, Gobin M, Niclou S, Van Dyck E. DNA repair mechanisms and their clinical impact in glioblastoma. *Mutat Res Rev Mutat Res* 2016; 769: 19-35.
17. Voß H, Schlumbohm S, Barwikowski P, et al. HarmonizR enables data harmonization across independent proteomic datasets with appropriate handling of missing values. *Nat Commun* 2022; 13: 3523.
18. Li L, Wang X. Identification of gastric cancer subtypes based on pathway clustering. *NPJ Precis Oncol* 2021; 5: 46.
19. Wilkerson MD, Hayes DN. ConsensusClusterPlus: a class discovery tool with confidence assessments and item tracking. *Bioinformatics* 2010; 26: 1572-3.
20. Ru B, Wong CN, Tong Y, et al. TISIDB: an integrated repository portal for tumor-immune system interactions. *Bioinformatics* 2019; 35: 4200-2.
21. Sauer G, Scholz D. On the purification and characterization of 2,3-PGase of red blood cells. *Folia Haematol* 1965; 83: 377-82.
22. Rooney MS, Shukla SA, Wu CJ, Getz G, Hacohen N. Molecular and genetic properties of tumors associated with local immune cytolytic activity. *Cell* 2015; 160: 48-61.
23. Barbie DA, Tamayo P, Boehm JS, et al. Systematic RNA interference reveals that oncogenic KRAS-driven cancers require TBK1. *Nature* 2009; 462: 108-12.
24. Wang S, Hu H. Impute the missing data using retrieved dropouts. *BMC Med Res Methodol* 2022; 22: 82.
25. Wang Y, Shi J, Chai K, Ying X, Zhou BP. The role of snail in EMT and tumorigenesis. *Curr Cancer Drug Targets* 2013; 13: 963-72.
26. Wang Q, Li M, Yang M, et al. Analysis of immune-related signatures of lung adenocarcinoma identified two distinct subtypes: implications for immune checkpoint blockade therapy. *Aging* 2020; 12: 3312-39.
27. Geeleher P, Cox N, Huang RS. pRRophetic: an R package for prediction of clinical chemotherapeutic response from tumor gene expression levels. *PLoS One* 2014; 9: e107468.
28. Langfelder P, Horvath S. WGCNA: an R package for weighted correlation network analysis. *BMC Bioinformatics* 2008; 9: 559.
29. Zhang Z. Variable selection with stepwise and best subset approaches. *Ann Transl Med* 2016; 4: 136.
30. Zhou X, Du J, Liu C, et al. A pan-cancer analysis of CD161, a potential new immune checkpoint. *Front Immunol* 2021; 12: 688215.
31. Jiang X, Yan Q, Xie L, et al. Construction and validation of a ferroptosis-related prognostic model for gastric cancer. *J Oncol* 2021; 2021: 6635526.
32. Fang Y, Huang S, Han L, Wang S, Xiong B. Comprehensive analysis of peritoneal metastasis sequencing data to identify LINC00924 as a prognostic biomarker in gastric cancer. *Cancer Manag Res* 2021; 13: 5599-611.
33. Li T, Fu J, Zeng Z, et al. TIMER2.0 for analysis of tumor-infiltrating immune cells. *Nucleic Acids Res* 2020; 48: W509-14.
34. Ostrom QT, Patil N, Cioffi G, Waite K, Kruchko C, Barnholtz-Sloan JS. CBTRUS Statistical Report: primary brain and other central nervous system tumors diagnosed in the United States in 2013-2017. *Neurooncology* 2020; 22 (12 Suppl 2): iv1-96.
35. Ma J, Chen CC, Li M. Macrophages/microglia in the glioblastoma tumor microenvironment. *Int J Mol Sci* 2021; 22: 5775.
36. Brandao M, Simon T, Critchley G, Giamas G. Astrocytes, the rising stars of the glioblastoma microenvironment. *Glia* 2019; 67: 779-90.
37. Wang Q, Hu B, Hu X, et al. Tumor evolution of glioma-intrinsic gene expression subtypes associates with immunological changes in the microenvironment. *Cancer Cell* 2017; 32: 42-56.e6.
38. Tang M, Xie Q, Gimble RC, et al. Three-dimensional bio-printed glioblastoma microenvironments model cellular dependencies and immune interactions. *Cell Res* 2020; 30: 833-53.
39. Medikonda R, Dunn G, Rahman M, Fecci P, Lim M. A review of glioblastoma immunotherapy. *J Neurooncol* 2021; 151: 41-53.
40. Lim M, Xia Y, Bettgowda C, Weller M. Current state of immunotherapy for glioblastoma. *Nature Rev Clin Oncol* 2018; 15: 422-42.
41. Sasmita AQ, Wong YP, Ling APK. Biomarkers and therapeutic advances in glioblastoma multiforme. *Asia-Pacific J Clin Oncol* 2018; 14: 40-51.

42. Verhaak RG, Hoadley KA, Purdom E, et al. Integrated genomic analysis identifies clinically relevant subtypes of glioblastoma characterized by abnormalities in PDGFRA, IDH1, EGFR, and NF1. *Cancer Cell* 2010; 17: 98-110.
43. Doucette T, Rao G, Rao A, et al. Immune heterogeneity of glioblastoma subtypes: extrapolation from the cancer genome atlas. *Cancer Immunol Res* 2013; 1: 112-22.
44. Maire CL, Fuh MM, Kaulich K, et al. Genome-wide methylation profiling of glioblastoma cell-derived extracellular vesicle DNA allows tumor classification. *Neurooncology* 2021; 23: 1087-99.
45. Binabaj MM, Bahrami A, ShahidSales S, et al. The prognostic value of MGMT promoter methylation in glioblastoma: a meta-analysis of clinical trials. *J Cell Physiol* 2018; 233: 378-86.
46. Mansouri A, Hachem LD, Mansouri S, et al. MGMT promoter methylation status testing to guide therapy for glioblastoma: refining the approach based on emerging evidence and current challenges. *Neurooncology* 2019; 21: 167-78.
47. Wenger A, Ferreyra Vega S, Kling T, Bontell TO, Jakola AS, Carén H. Intratumor DNA methylation heterogeneity in glioblastoma: implications for DNA methylation-based classification. *Neurooncology* 2019; 21: 616-27.
48. Simonds EF, Lu ED, Badillo O, et al. Deep immune profiling reveals targetable mechanisms of immune evasion in immune checkpoint inhibitor-refractory glioblastoma. *J Immunotherapy Cancer* 2021; 9: e002181.
49. Zhao J, Chen AX, Gartrell RD, et al. Immune and genomic correlates of response to anti-PD-1 immunotherapy in glioblastoma. *Nat Med* 2019; 25: 462-9.
50. Zeng J, See AP, Phallen J, et al. Anti-PD-1 blockade and stereotactic radiation produce long-term survival in mice with intracranial gliomas. *Int J Radiat Oncol Biol Phys* 2013; 86: 343-9.
51. Turley SJ, Cremasco V, Astarita JL. Immunological hallmarks of stromal cells in the tumour microenvironment. *Nat Rev Immunol* 2015; 15: 669-82.
52. Kijewska M, Kocyk M, Kloss M, et al. The embryonic type of SPP1 transcriptional regulation is re-activated in glioblastoma. *Oncotarget* 2017; 8: 16340-55.
53. Sharanek A, Burban A, Laaper M, et al. OSMR controls glioma stem cell respiration and confers resistance of glioblastoma to ionizing radiation. *Nat Commun* 2020; 11: 4116.
54. Chen L, Wang L, Qin J, Wei DS. CtBP2 interacts with ZBTB18 to promote malignancy of glioblastoma. *Life Sci* 2020; 262: 118477.
55. Desland FA, Hormigo A. The CNS and the brain tumor microenvironment: implications for glioblastoma immunotherapy. *Int J Mol Sci* 2020; 21: 1-19.
56. Huang B, Li X, Li Y, Zhang J, Zong Z, Zhang H. Current immunotherapies for glioblastoma multiforme. *Front Immunol* 2020; 11: 603911.
57. Kong Y, Feng ZC, Zhang YL, et al. Identification of immune-related genes contributing to the development of glioblastoma using weighted gene co-expression network analysis. *Front Immunol* 2020; 11: 1281.
58. Lin Z, Wang R, Huang C, et al. Identification of an immune-related prognostic risk model in glioblastoma. *Front Genet* 2022; 13: 926122.