



Tumor Microenvironment Characterization in Glioblastoma Identifies Prognostic and Immunotherapeutically Relevant Gene Signatures

Jinsen Zhang¹ · Xing Xiao¹ · Xin Zhang¹ · Wei Hua¹

Received: 10 November 2019 / Accepted: 17 January 2020 / Published online: 31 January 2020
© Springer Science+Business Media, LLC, part of Springer Nature 2020

Abstract

Tumor microenvironment (TME) cells are important elements in tumor tissue. There is increasing evidence that they have important clinical pathological significance in predicting tumor clinical outcomes and therapeutic effects. However, no systematic analysis of TME cell interactions in glioblastoma (GBM) has been reported. We systematically analyzed the transcriptional sequencing data of GBM to find an immune gene marker to predict the clinical results of GBM. First, we downloaded the expression profiles and clinical follow-up information of GBM from The Cancer Genome Atlas (TCGA) and Gene Expression Omnibus (GEO). CIBERSORT was used to evaluate the infiltration mode of TME in 757 patients, systematically correlated TME phenotype with genomic characteristics and clinicopathological characteristics of GBM, defined four TME phenotypes, and TMEScore was constructed using algorithms such as random forest and principal component analysis. There is a significant correlation between TMEScore and age of onset. High TMEScore samples are characterized by immune activation, TGF pathway activation, and high expression of immune checkpoint genes, while low TMEScore samples are characterized by high-frequency IDH1 and MET mutations. Therefore, a comprehensive landscape depicting the TME characteristics of GBM may help explain GBM's response to immunotherapy and provide new strategies for cancer treatment. In this study, TMEScore can be used as a new prognostic marker to predict the survival of GBM patients, and as a potential predictor of immune checkpoint inhibitor response.

Keywords Tumor microenvironment · Immunity · TMEScore · TCGA

Introduction

Glioblastoma (GBM) is the most common highly degenerative primary intracranial tumor in adults, accounting for 14.9% of all

primary and central nervous system (CNS) tumors, 47.1% of malignant primary and central nervous systems tumors, and 56.1% of all gliomas (Thakkar et al. 2014). The prognosis of GBM is poor, and the relative survival rate is low, and only 5.5% of patients aged 55–64 survive for 5 years (Ostrom et al. 2017); the median survival time was only 3 months without treatment, and the median survival time after standard treatment was only 15 months (Malmstrom et al. 2012). GBM typically has the ability to severely invade and permeate normal surrounding tissue throughout the brain, including areas that control speech and motor function, making it impossible to perform a full resection. Infiltrating tumor cells are always retained in the surrounding brain and contribute to disease progression or recurrence. Tumors smaller than 5–6 cm and those that do not cross the mid-line, and supratentorial (cerebrum) and cerebellar tumors (persuadable to surgical resection) have been associated with favorable outcomes (Ellor et al. 2014; Walid 2008). Current standard therapy for GBMs encompasses maximally secured surgical resection followed by concomitant RT and TMZ chemotherapy (Stupp et al. 2009; Van Meir et al. 2010). Given the poor prognosis after standard treatment and the low

Jinsen Zhang and Xing Xiao contributed equally to this work.

Electronic supplementary material The online version of this article (<https://doi.org/10.1007/s12031-020-01484-0>) contains supplementary material, which is available to authorized users.

✉ Wei Hua
hs_huawei@126.com

Jinsen Zhang
760010228@qq.com

Xing Xiao
18211220058@fudan.edu.cn

Xin Zhang
xinzhanghs@126.com

¹ Department of Neurosurgery, Huashan Hospital, Fudan University, No.12 Wulumuqi Zhong Road, Shanghai 200040, China

level of targeted therapies in GBM, immunotherapy is a promising additional approach (Couzin-Frankel 2013; Farkona et al. 2016) and is currently undergoing intensive research. At the same time, several immune-related parameters have been reported for predicting the prognosis of patients with GBM (Cheng et al. 2016; Luoto et al. 2018; Mohme et al. 2018), further indicating that different immune status has a significant impact on the prognosis of GBM patients. Therefore, there is a need to systematically study the immunophenotype in the GBM microenvironment to better understand complex anti-tumor responses and to guide effective immunotherapy in GBM.

Different tumor microenvironments (TMEs) are formed in different stages of tumor development, and have multiple capabilities to induce adverse and beneficial consequences of tumorigenesis; immune cells can be activated to promote tumor growth and progression, most likely to be affected by the tumor microenvironment (Kim and Bae 2016). More and more research shows that TME plays a vital role in cancer progression and therapeutic response (Jiang et al. 2018; Zeng et al. 2018). For example, differences in the compositions of resident cell types within the TME, including cytotoxic T cells, helper T cells, dendritic cells (DCs), tumor-associated macrophages, mesenchymal stem cells, and associated inflammatory pathways, have been reported in patients with cancer (Fridman et al. 2017; Kalluri 2016; Mantovani et al. 2017; Turley et al. 2015). FOXP3 expression is important in GBM patient overall survival (Li et al. 2019). Piperi et al. reported that STAT3 impairs anti-tumor immunity (Piperi et al. 2019). At diagnosis, TME may reflect immune responses and chemotherapy benefits (Rosenberg et al. 2016), and the changes in the number of infiltrating CD8⁺ T cells, CD4⁺ T cells, macrophages, and fibroblasts in TME are associated with clinical and prognosis of various malignancies (Lee et al. 2014; Mariathasan et al. 2018; Nishino et al. 2017; Turley et al. 2015). There are several algorithms that can estimate the abundance of immune cells and other cells in TME (Becht et al. 2016; Newman et al. 2015; Yoshihara et al. 2013), although some studies have used these methods to explore the relationship between TME infiltration and clinical outcomes of tumors (Fu et al. 2018; Zeng et al. 2019), the comprehensive landscape infiltrated by TME has not been discussed in GBM so far.

In this study, based on clinical annotated GBM gene expression profile, computational algorithms were used to estimate the proportion of 22 immune cell types and cancer-related fibroblasts. We assessed the TME infiltration patterns from 757 GBM patients, and systematically correlated TME phenotypes with the genomic and clinicopathological characteristics of GBM. Therefore, we established a method to quantify the penetration pattern of TME (TMEScore). TMEScore was found to be a powerful prognostic biomarker and predictor of response to immunotherapy inhibitors.

Materials and Method

Data Collection and Processing

The expression data of TCGA-GBM's Affymetrix HT Human Genome U133a microarray platform were downloaded from the UCSC Xena database, containing a total of 539 samples 12,043 gene expression profiles at May 5, 2019, and the prognostic information of these samples at the same time was downloaded, no clinical data and follow-up time is less than 30 days samples were removed, and the final study included 500 GBM samples as training sets. Similarly, a set of chip dataset GSE13041 (Lee et al. 2008) were downloaded from the Gene Expression Omnibus (GEO) database, which integrates Affymetrix Human Genome U133A Array, Affymetrix Human Genome U133 Plus 2.0 Array, and Affymetrix Human Genome U95 Version 2 Array. A total of 267 GBM samples of expression data and clinical follow-up information, removal of follow-up time of less than 30 days, and finally included 257 cases as test sets. The sample statistics of the two sets of data are shown in Table 1. For the probe data, we used the R software package `hgu133plus2.db` to map the probe to GeneSymbol. Multiple probes correspond to the median expression of a gene, and probes that match multiple genes are removed.

Infiltrating Cell Score in TME

CIBERSORT is a deconvolution algorithm that uses a set of reference gene-expression values (a signature with 547 genes) considered a minimal representation for each cell type and, based on those values, infers cell type proportions in data from bulk tumor samples with mixed cell types using support vector regression. CIBERSORT is able to identify 22 kinds of human immune cells according to the gene expression data with high specificity and sensitivity, including B cells, T cells, NK cells, macrophages, DCs, and myeloid subset cells. In order to quantify the proportion of immune cells in GBM samples, CIBERSORT algorithm (Newman et al. 2015) was used to calculate the scores of 22 immune cells in TCGA-GBM and GSE13041 data sets with LM22 gene signature as the reference. Specifically, gene expression data were uploaded to the CIBERSORT website (<http://cibersort.stanford.edu/>), and scores of 22 immune cells were obtained using LM22 signature and 1000 permutation.

Consensus Clustering to Obtain Molecular Subtypes Associated with TME-Infiltrating Cells

Consensus clustering was performed using the `ConsensusClusterPlus` (Sorlie et al. 2001) package in R to determine subgroups of GBM based on the TME-infiltrating cells. As Zhang et al. (Zhang et al. 2018) evaluated the optimal

Table 1 Sample information of TCGA-GBM and GSE13041

Clinical features	TCGA	GSE13041
Event		
Alive	71	21
Dead	429	236
Platform		
GPL570	0	23
GPL8300	0	47
GPL96	500	187
New event type		
Unknown	153	257
LD	9	0
PD	227	0
RC	102	0
Age		
0~30	23	36
30~40	35	22
40~50	71	63
50~60	137	67
60~70	132	37
70~100	102	32
Gender		
Female	195	88
Male	305	142
Neoadjuvant		
No	479	0
Yes	21	0
Unknown	0	257
HC class distribution		
Mes	0	96
PN	0	76
Pro	0	58
ProMes	0	27
Unknown	500	0

number of clusters between $k=2-10$, repeated 1000 times to ensure the stability of the results, and used the R software package pheatmap for visualization.

Differentially Expressed Genes Associated with the TME Phenotype

To identify genes associated with TME cell-infiltrating patterns, the linear model was used to analyze the gene expression differences among phenotypic-related TME subgroups. Specifically, the R software package limma (Ritchie et al. 2015) was used to calculate the differential genes, and FDR < 0.05 was selected. In order to include more candidate genes, Foldchange was not restricted.

TME Phenotype-Related Differential Gene Re-Clustering

Nonnegative matrix factorization (NMF) is an unsupervised clustering method that is widely used in the discovery of tumor molecular subtypes based on genomics (Mirzal 2014; Yu et al. 2018). In order to further observe the relationship between the expression and phenotypes of TME phenotype-related difference genes, we used NMF method to re-cluster samples based on the expression profile of TME phenotype-related difference genes, and analyzed the clinical characteristics of the samples after re-clustering. The NMF method selects the standard “brunet” and performs 50 iterations. The number of clusters k is set to 2 to 10, and the average profile width of the common member matrix is calculated using the R package NMF (Ye et al. 2019), and the minimum member of each subclass is set to 10.

Dimension Reduction and Generation of TME Gene Signatures

In order to obtain a robust TME gene signatures, the prognostic value of each differentially expressed genes (DEG) were predicted, genes with prominent prognosis were selected, and the importance of these DEGs were evaluated by random forest algorithm. Specifically, the coph function of survival of R software package was used for univariate survival analysis, and the selection threshold was 0.05, and the genes with significant prognosis were included for randomForest feature selection using randomForest of R software package, and mtry of each segmentation set as 1–165 and ntree = 500, and the mtry value with the lowest error rate was selected as the optimal mtry value of the randomForest algorithm. Then ntree = 100 was selected according to the error rate of random forest. Finally, each DEG was ranked according to its importance, and 95% DEGs with cumulative importance > 95% were selected as candidate feature genes. k-means (Boillaud and Molina 2015) was used to divide these genes into five categories. Psych R package was used for principal component analysis of the expression profiles of 5 categories of genes, and the first principal component was extracted as signature score after 100 iterations. This approach has the advantage of focusing the score on the set with the largest block of well-correlated (or anticorrelated) genes in the set, while downweighting contributions from genes that do not track with other set members. For gene type j , signature score formula of samples is as follows:

$$S_j = \sum_{i=1}^{n_j} P c 1_i * Exp_i$$

j represents the j th class of the five types of genes, wherein n_j represents the number of genes of the j th gene,

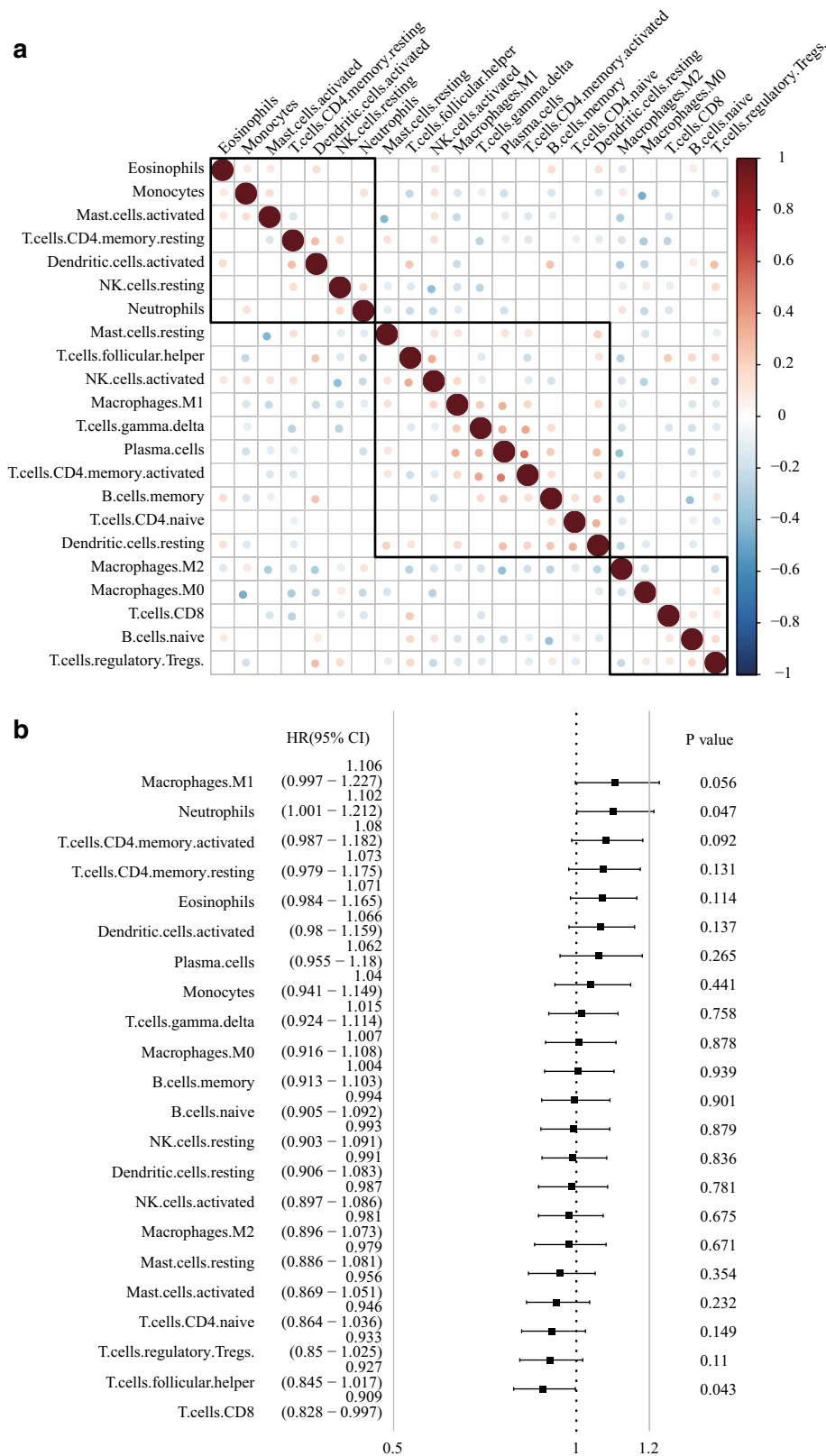


Fig. 1 a Correlation of 22 immune cells in TME, among which the dot size and color represented correlation, blue represented negative correlation, red represented positive correlation, and the white represented non-significant. **b** A forest map of 22 immune cells in

TME. **c** Heat map of the scores of 22 immune cells in TME, the higher the red, the lower the blue. **d** Distribution box map of 22 immune cell scores in four types of TMEC, red * indicates significant difference. **e** OS prognosis KM curves of four TMEC groups

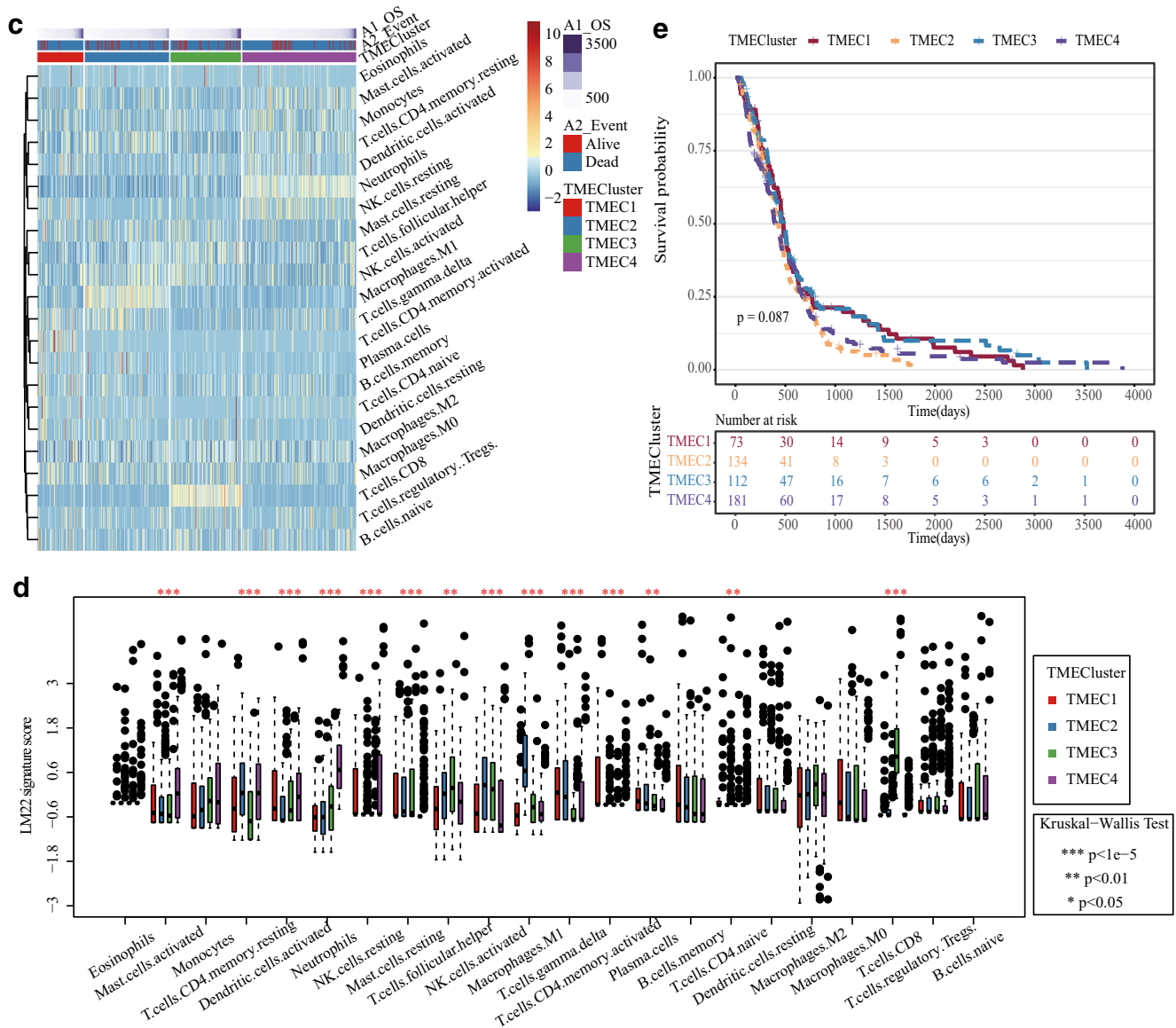


Fig. 1 continued

$Pc1_i$ represents the first principal component coefficient of the i th gene of the j th gene, and Exp_i indicates the first the i th gene expression level of the j -type gene.

Finally, the risk coefficient of each signature score is obtained by using the signature score of the five genes in each sample through multivariate regression. For any sample, the TMEScore formula is as follows:

$$TMEScore = \sum_{j=1}^5 S_j * \beta_j$$

where j represents the j th class of the five classes of genes, S_j represents the signature score of the class j gene of the sample, and β_j represents the risk regression coefficient of the signature score of the class j gene.

Relationship Between TMEScore and Clinical Features

To observe the relationship between TMEScore and clinical phenotype, the samples were divided into two groups based on the median TMEScore of the samples, which compared the prognostic differences between high TMEScore and low TMEScore, respectively. Similarly, the relationship between high and low TMEScore and age and gender is analyzed.

Relationship Between TMEScore and Immune-Related Gene Expression

In order to observe the relationship between TMEScore and immune-related genes, three types of immune-related genes were first collected: (1) immune activation genes CXCL10, CXCL9,

GZMA, GZMB, PRF1, IFNG, TBX2, TNF (where CD8A was not detected in TCGA-GBM data); (2) immunological checkpoint genes PDCD1, CTLA4, LAG3, PDCD1LG2 (IDO1, CD274, HAVCR2 are not detected in TCGA-GBM data); and (3) TGF/EMT pathway genes VIM, ACTA2, COL4A1, TGFBR2, ZEB1, CLDN3, SMAD9, TWIST1. The expression profiles of these genes were extracted to further analyze the differential expression of these three types of genes in high TMEScore and low TMEScore.

Relationship Between TMEScore and Tumor Genomic Variation

To observe differences in genomic variation between high TMEScore and low TMEScore samples, we downloaded SNP data from TCGA, removed intron and silent mutations, and used Fisher’s exact test to analyze the difference in the mutations between the two samples, selected the threshold $p < 0.05$.

Statistical Analysis

The normality of the variables was tested by the Shapiro-Wilk normality test (Ghasemi and Zahediasl 2012), unless otherwise specified. For the comparison of the two groups, the statistical significance of the normal distribution variables was estimated by the unpaired Student t test, and the non-normal distribution variables were analyzed by the Mann-Whitney U test. For comparisons between the two groups, the Kruskal-Wallis test and the one-way ANOVA were used as non-parametric and parametric methods, respectively (Hazra and Gogtay 2016). Correlation coefficients were calculated by Spearman and distance correlation analysis. The two-sided Fisher’s exact test was used to analyze the contingency table, and we used the Benjamini-Hochberg method to convert the p value to FDR. The Kaplan-Meier method was used to generate survival curves for the subgroups in each data set, and the log rank test was used to determine the statistical significance of the differences, with significance being defined as

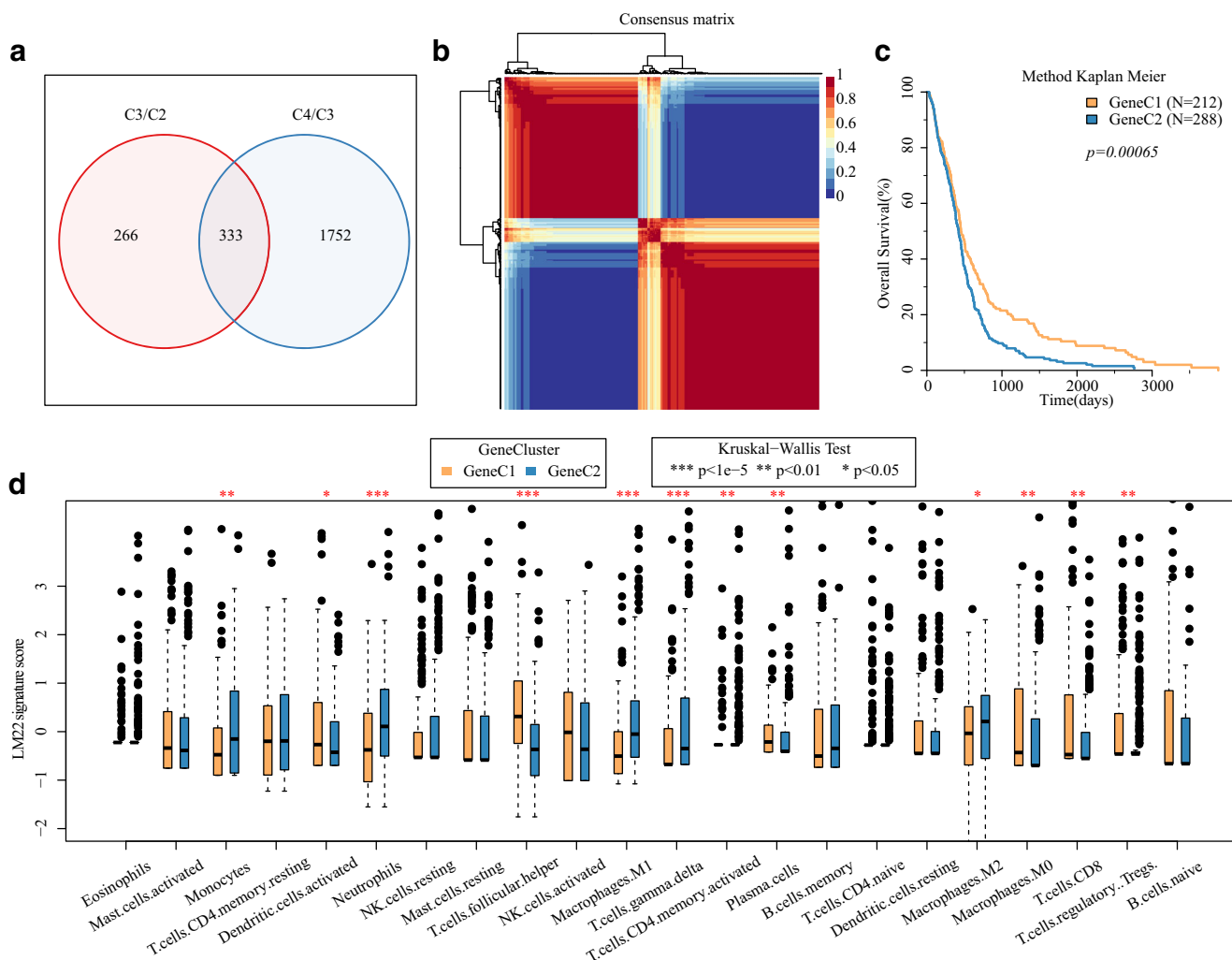


Fig. 2 a Venn diagram of differentially expressed genes between TMEC3 (C3), TMEC2 (C2), and TMEC4 (C4). b Heatmap of NMF algorithm consistency matrix. c Prognosis KM curve of GeneC1 and GeneC1. d

The scores of 22 immune cells were shown in the boxplot of GeneC1 and GeneC1 samples

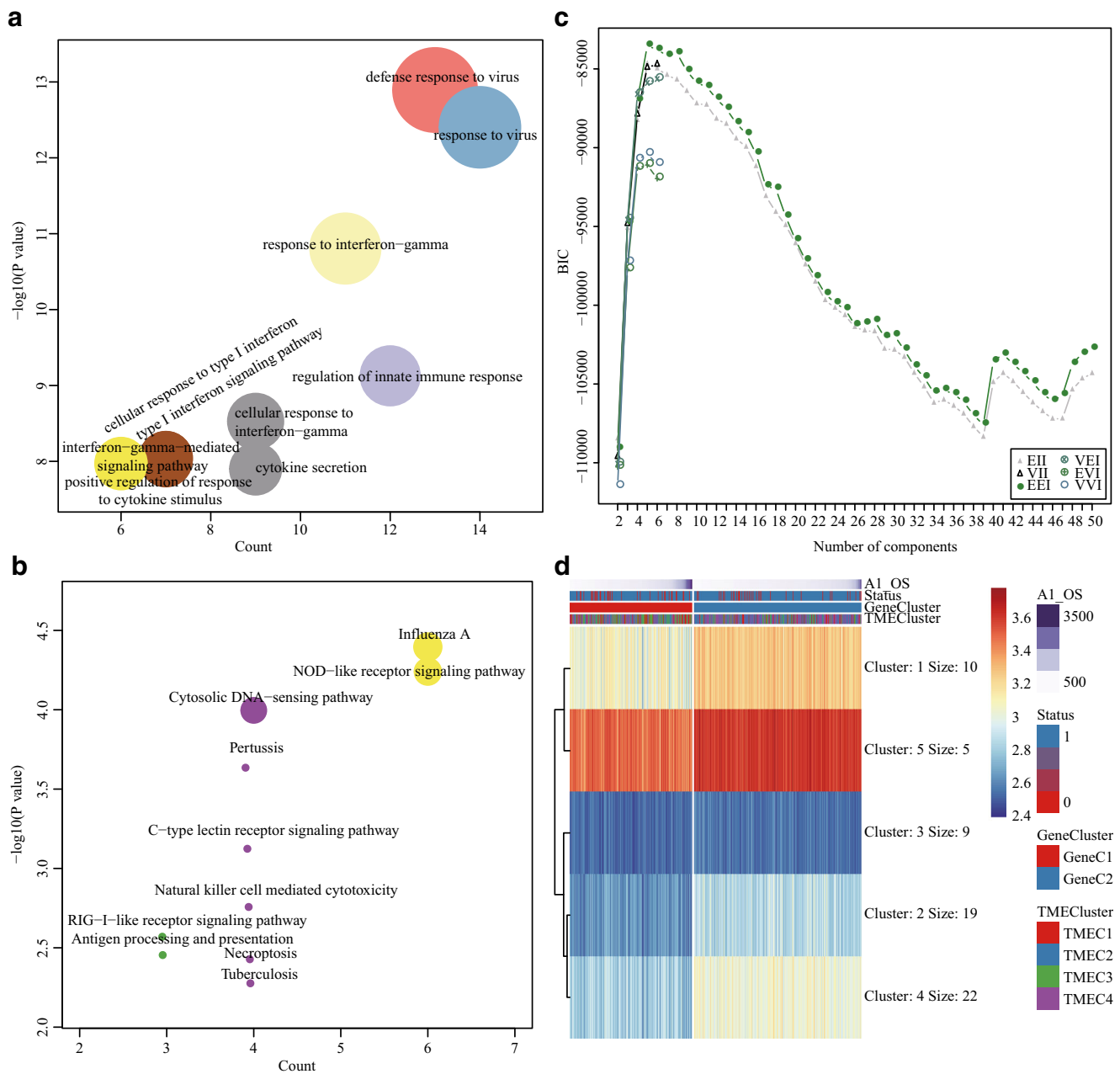


Fig. 3 **a** GO enrichment results of 65 genes. **b** KEGG pathway enrichment results of 65 genes. **c** BIC plot for models fitted to the 65 genes expression data. **d** k-means clustering results of 65 genes. **e** Heat map of the expression

$p < 0.05$. All of these analyses were performed in R 3.4.3, and all analyses were based on default parameters unless otherwise stated.

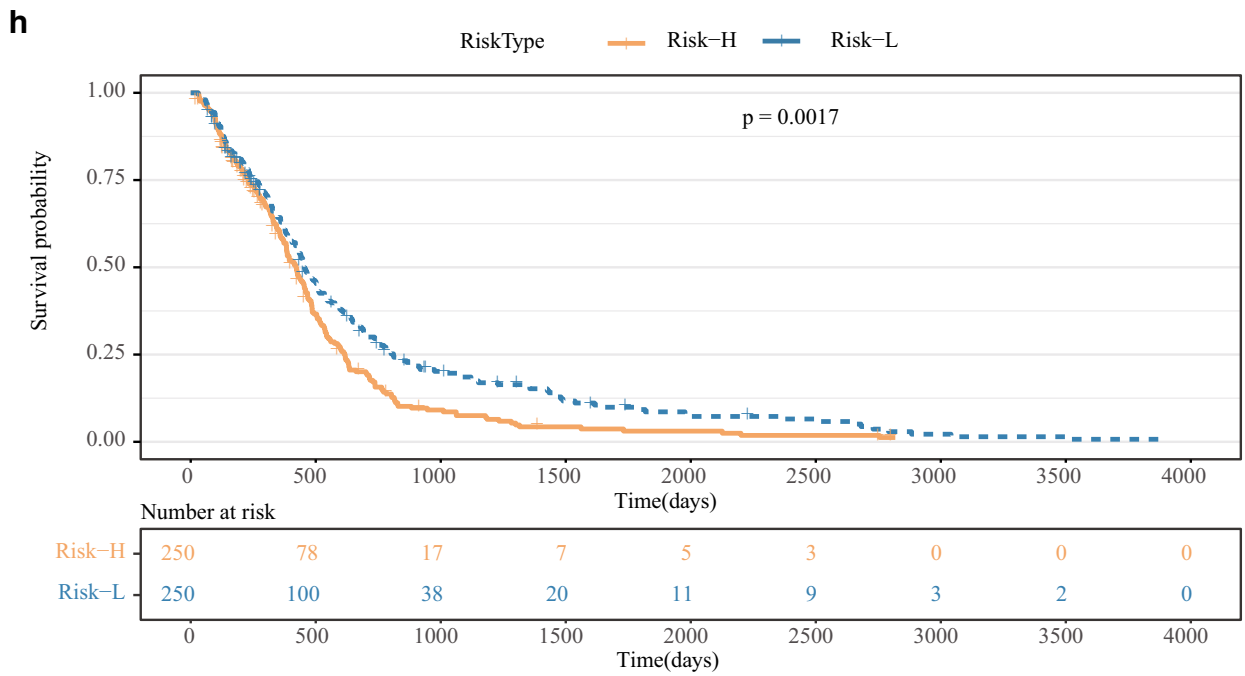
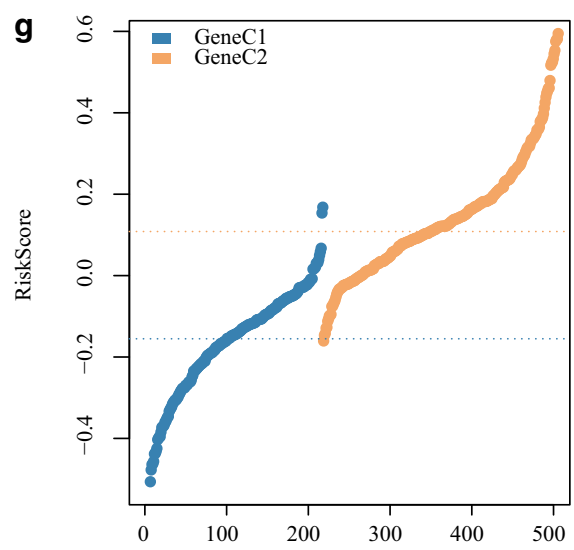
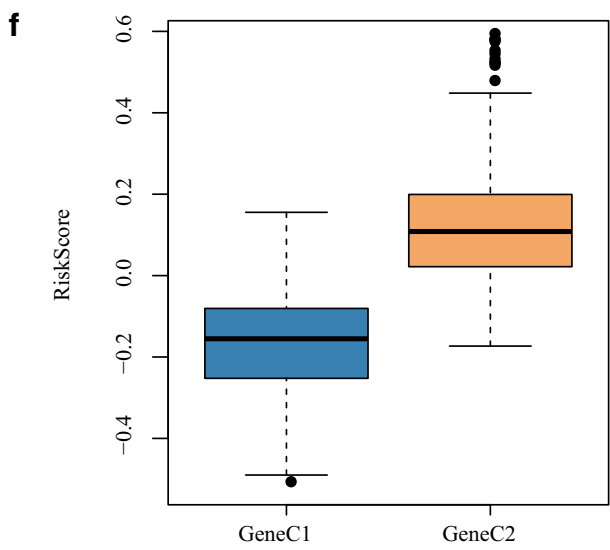
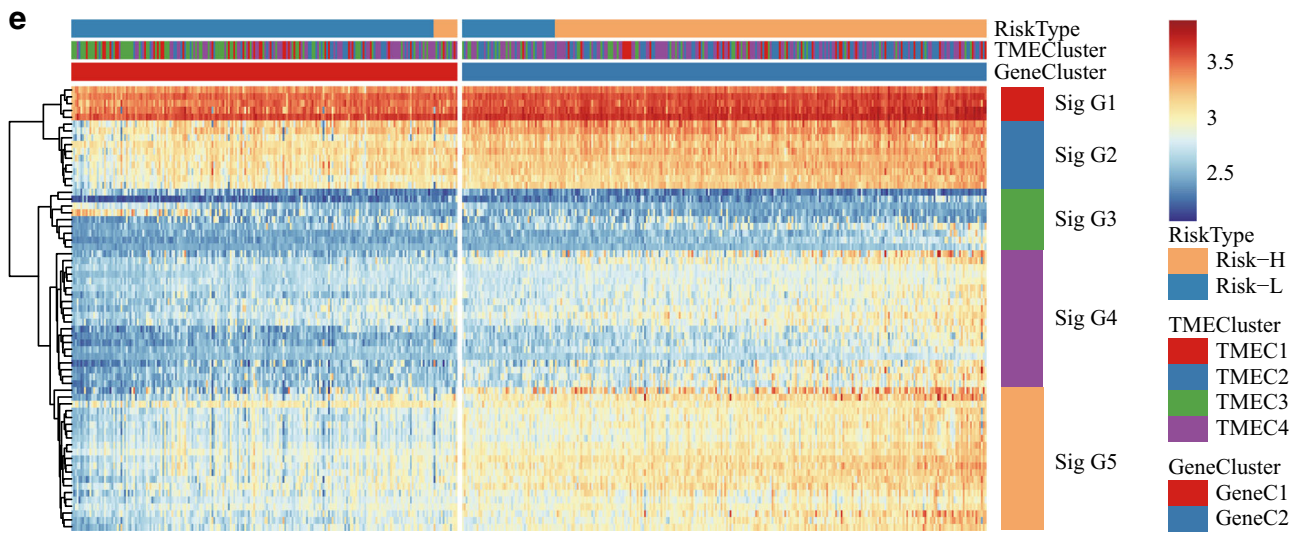
Results

TME Landscape in GBM

CIBERSORT was used to calculate 22 immune cell scores in the training set, and the correlation was analyzed between

levels of 65 genes. **f** Comparison of TME scores between GeneC1 and GeneC2. **g** TME score distribution of GeneC1 and GeneC2. **h** OS prognostic KM curve for Risk-H and Risk-L group samples

these scores. They were mainly divided into three categories, two with positive correlation and one with negative correlation (Fig. 1a); this suggests that there may be some specific communication mode between immune cells. The relationship between these 22 immune cell scores and prognosis were analyzed by Univariate cox, and found that neutrophils scores were associated with poor prognosis (log rank $p < 0.05$), while T cells CD8 scores were associated with better prognosis (log rank $p < 0.05$) (Fig. 1b). The scores of the top 4 immune cells with the highest prognostic value were selected for consistent



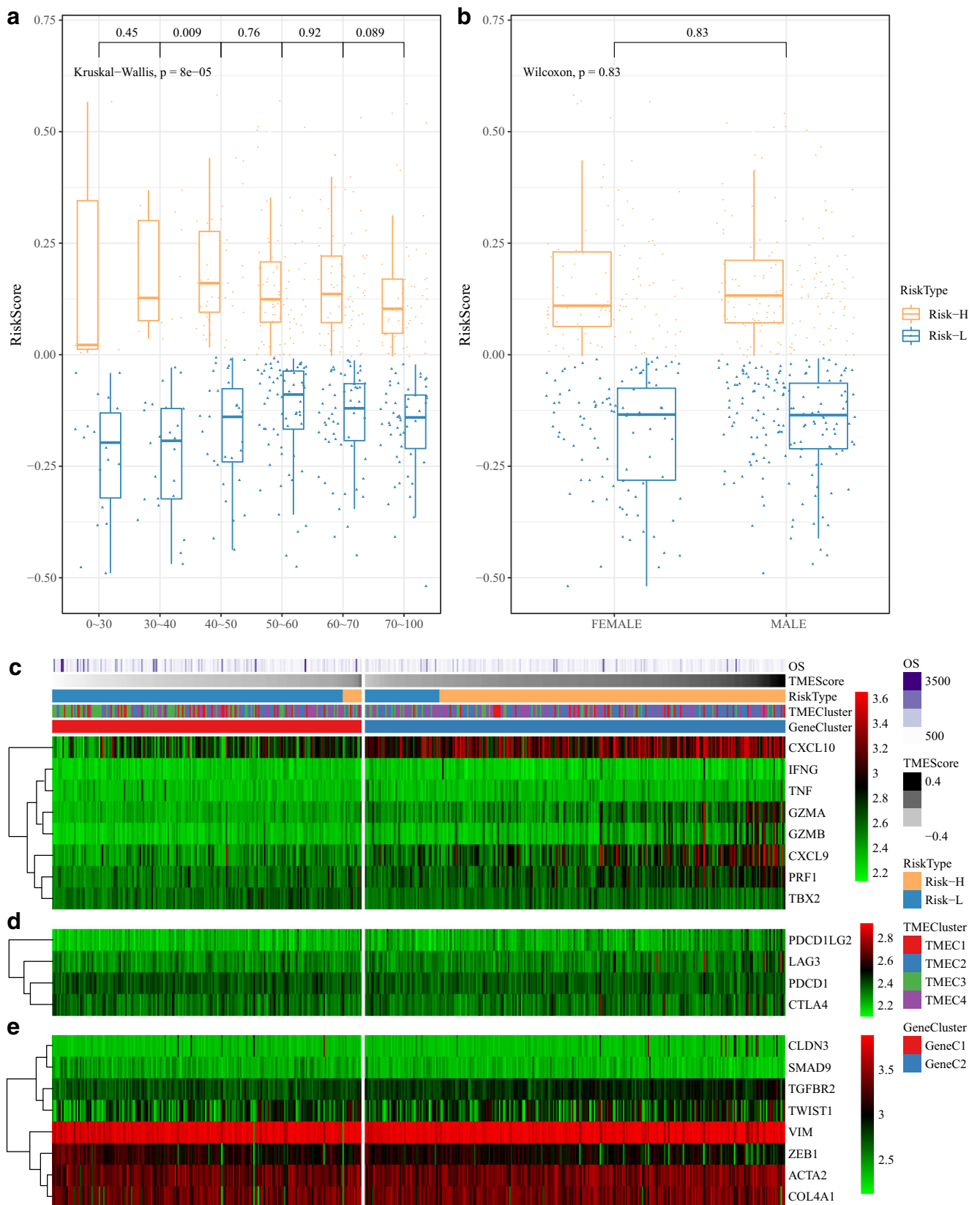


Fig. 4 **a** The relationship between age and TMEscore. **b** The relationship between Gender and TMEscore. **c** Heatmap of the immune activation gene expressed on the TCGA sample. **d** Heatmap of the immunological

checkpoint gene expressed on the TCGA sample. **e** Heatmap of the TGF pathway gene expressed on the TCGA sample

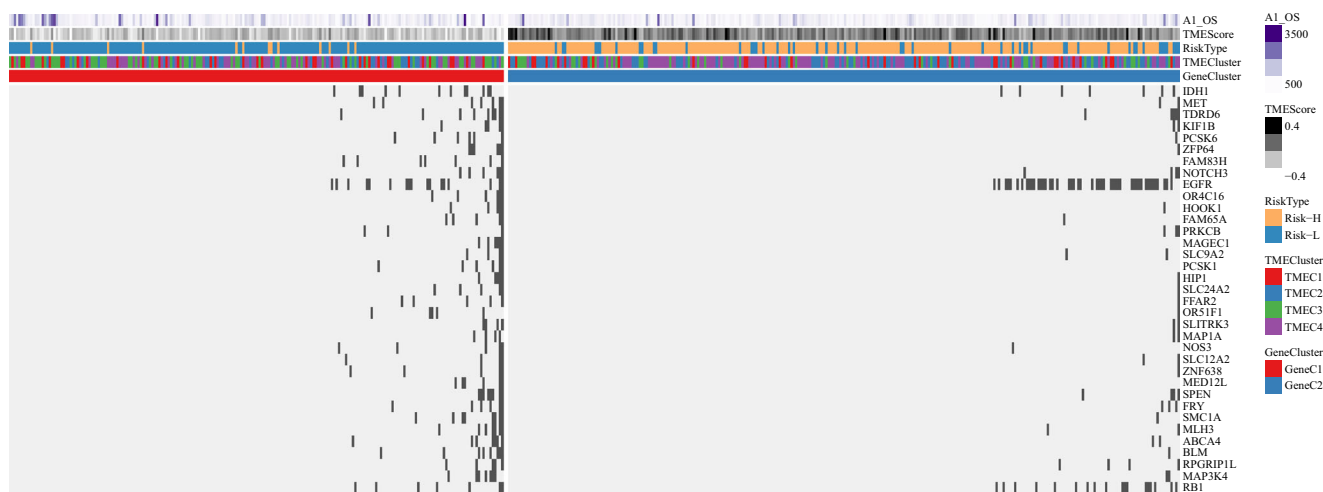


Fig. 5 The relationship between TME and genomic mutation characteristics, in which the horizontal axis represents the sample, the vertical axis represents the gene, the black rectangle represents the mutation, and the gray represents the unmutated

cluster analysis, and $k = 4$ was selected as the optimal number of clusters (Fig. S1). The four categories of TME scores are defined as TMEC1-TMEC4, and they have different patterns in 22 scores (Fig. 1c). Among them, we analyzed the distribution differences of these 22 kinds of scores in four types of samples, 14 of which (63.6%) had significant differences (Fig. 1d). Further analysis of the prognostic differences among the four samples showed no significant prognostic differences in their overall survival (Fig. 1e), but there are significant differences in the local (TMEC3 vs. TMEC2, log rank $p = 0.03$, Fig. S2). Similarly, we also get the optimal clustering results of four categories in the test set (Fig. S3). There were no significant differences in the overall survival of the four types of samples in the test set (Fig. S4A), But it is worth mentioning that these four types of samples overlap significantly with Phillips and colleagues suggested the following terms to be more descriptive of the cellular characteristics: “ProNeural” (PN) for HC1A, “Proliferative” (Pro) for HC2A, and “Mesenchymal” (Mes) for HC2B (Phillips et al. 2006) (Fig. S4B). Fourteen (63.6%) of the 22 scores in the four categories of PN, Pro, Mes, and ProMes have significant distribution differences (Fig. S4C). In short, TME score may be closely related to the occurrence and development of GBM.

Differential Expression of Genes Between TMEC

The differences in gene expression between those with the best prognosis (TMEC3) and those with the worst prognosis (TMEC2 and TMEC4) were analyzed in the training set. There were 599 DEGs in TMEC3-vs-TMEC2 and 2085 DEGs in TMEC4-vs-TMEC3, and 333 genes in the intersection of them were selected (Fig. 2a), Based on the 333 gene expression profile, TFGA samples were re-clustered using NMF to obtain stable class 2 samples, which were defined as GeneC1, GeneC2, respectively (Fig. 2b), There were

significant prognostic differences between GeneC1 and GeneC2 in overall survival (Fig. 2c). The distribution of GeneC1 and GeneC2 in 22 immune cell scores was compared, and there were significant differences between GeneC1 and GeneC2 in various TME (Fig. 2d), For example, GeneC1 with a good prognosis obtained significantly higher scores in T cell CD8 and T cell infertile helper than GeneC2.

Construction of the TME Signature

We first evaluated the correlation between 333 genes and prognosis, among which 166 genes (49.8%) were significantly correlated with prognosis. The expression profiles of these 166 genes were used for random forest dimension reduction analysis, and $n_{tree} = 100$ was selected according to the error rate of random forest (Fig. S5A). A total of 65 candidate genes were identified by selecting DEGs with cumulative importance $> 95\%$ (Fig. S5 BC). These 65 genes are mainly enriched in the immune-related GO Term and KEGG Pathway (Fig. 3a, b) and were clustered by k-means algorithm; the optimal number of clusters is 5 (Fig. 3c), defined as signature G1, signature G2, signature G3, signature G4, and signature G5, which contains 10, 19, 9, 22, and 5 genes, respectively, which have different patterns in each sample (Fig. 3d). As described in the “Methods” section, TMEscore is built to calculate the TMEscore for each sample in the training set as S1.txt. According to TMEscore, the samples were divided into Risk-H (high TMEscore) and Risk-L (low TMEscore) groups, and the relationship between the expression of 65 genes and Risk-H and Risk-L was shown in Fig. 3e, The five types of genes have different expression patterns in each sample, and their expressions are also different in the Risk-H and Risk-L groups. Among them, the samples of GeneC1 are mainly concentrated in the Risk-L group, while the samples of GeneC2 are mainly concentrated in the Risk-H group (Fig. 3f, g). Further analysis

of the prognostic differences between Risk-H and Risk-L, as expected, the prognosis of the Risk-H group was significantly worse than that of the Risk-L group (Fig. 3h).

Clinical and Transcriptome Characteristics of TMEScore

The relationship between TMEScore and age and gender were evaluated, and the results showed that there was a significant difference in the TMEScore of the age group sample (Fig. 4a), while no such significant difference was observed in gender group sample (Fig. 4b). To investigate the relationship between different TME and immune status, the relationship between the expression of immune activation genes (CXCL10, CXCL9, GZMA, GZMB, PRF1, IFNG, TBX2, TNF) and TMEC, GeneC, and TMEScore were analyzed. There were different expression patterns in different GeneC, TMEScore, and TMEC (Fig. 4c), and the expression level of Risk-H samples with poor prognosis such as CXCL10, CXCL9, and GZMB is significantly higher than that of prognosis (Fig. S6A). The relationship between the expression of immune checkpoint genes (PDCD1, CTLA4, LAG3, PDCD1LG2) and TMEC, GeneC, and TMEScore was analyzed (Fig. 4d); although the expression level was not high, there were different expression trends in different GeneC, TMEScore, and TMEC. PDCD1LG2 (PD-L2) was significantly higher in Risk-H than that in Risk-L group (Fig. S6B). Through analyzing the expression levels of TGF/EMT pathway genes VIM, ACTA2, COL4A1, TGFB2, ZEB1, CLDN3, SMAD9, and TWIST1 in TMEC, GeneC, and TMEScore, we can see that there are also different expression patterns in different GeneC, TMEScore, and TMEC (Fig. 4e); among them, the expression of ZEB1 was significantly lower in Risk-L than that in Risk-H, and the expression of other genes was significantly higher in Risk-H group than that in Risk-L (Fig. S6C). In addition, we also found the same phenomenon in the verification data set (Fig. S7). All in all, different TMEScore is closely correlated to the expression of immune genes.

TME Characteristics of Cancer Somatic Genomes

In order to observe the relationship between the distribution of somatic mutations and TMEScore, 35 genes with significant difference in mutation frequency in samples of Risk-H and Risk-L were finally identified by analyzing the mutation annotation files of TCGA-GBM cohort using Fisher's exact test (Fig. 5). Previous studies have reported links between genetic mutations and immune checkpoint blocking responses or resistance (Burr et al. 2017; George et al. 2017); however, few studies have been conducted in GBM. The mutation frequency of IDH1, MET, and other genes in Risk-L is significantly higher than that in Risk-H, while EGFR shows an opposite trend. These data can provide a new perspective for the study of the

formation mechanism of TME, and explore individual mutations and their role in cancer immunity and immunotherapy.

Discussion

Recently, tumor immunotherapy has achieved remarkable success in the treatment of advanced tumors (Maccio and Madeddu 2012; Sharma et al. 2011). A comprehensive understanding of GBM not only needs to focus on tumor cells but also on the tumor microenvironment (Leonardi et al. 2012; Yarchoan et al. 2017), which contains a variety of cell populations that interact with cancer cells and participate in various stages of tumorigenesis. Tumor-invasive immune cells and immune response in tumor microenvironment have attracted great attention of researchers and become promising therapeutic targets. Further research into how immune characteristics are associated with the development and progression of GBM can assist in the development of new and specific targeted therapeutic strategies, especially in the setting of combination therapies. In this study, the TME landscape in GBM was systematically analyzed. Among the 22 kinds of immune cells, many of them are related to the poor prognosis of GBM, and they have the potential of molecular stratification of GBM, which indicates that different immune characteristics of GBM in different tumor stages.

Therapeutic antibodies that block the PD-1/PD-L1 pathway induce a robust and long-lasting response in patients with a variety of cancers, including GBM (Hardcastle et al. 2017); however, these responses occur only in a small number of patients, and some studies have found that PD-1 expression, PD-L1 expression, MSI status, and mutation load are not effective biomarkers for predicting the benefit of immune checkpoint blockade (Roh et al. 2017). Therefore, it is important to develop biomarkers that predict the benefits of checkpoint immunotherapy, and the latest data supports the notion that TME plays a key role in checkpoint inhibitor immunotherapy (Cristescu et al. 2018). Here, we have elucidated the comprehensive landscape of the interaction between GBM clinical characteristics and infiltrating TME cells, and with the help of several computational algorithms, we have established a method to quantify the infiltrating mode of TME–TMEScore.

Comprehensive analysis shows that TMEScore is a prognostic biomarker of GBM, and high TMEScore is correlated with poor prognostic results. Age is a key prognostic factor in GBM (Lee et al. 2008); as expected, there is a significant association between TMEScore and the age of the patient. The immunological activation genes such as CXCL10, CXCL9, GZMB, and immunological checkpoint genes such as PDCD1LG2 are highly expressed in high TMEScore, suggesting the potential of TMEScore as a marker for the benefit of immunological checkpoints. The significantly high

expression of TGF pathway genes in high TMEScore indicates that TGF activation may result in a faster metastasis of GBM leading to poor clinical outcomes. IDH mutation is one of the causes of GBM (Cohen et al. 2013) and plays an important role in clinical treatment and prognosis. IDH1 mutation frequency in high TMEScore samples is significantly lower than that in low TMEScore samples, while EGFR gene shows an opposite trend. These data can provide a new perspective for studying the mechanism of TME formation and explore individual mutations and their role in GBM immunity and immunotherapy.

Although we use bioinformatics techniques to identify potential candidate immune gene markers involved in the development of GBM in large samples, further validation should be performed in a prospective cohort of immunotherapy to more fully define the cutoff values to be used. Second, given the heterogeneity of different tumor regions, it is appropriate to systematically assess immune infiltration at the tumor core and invading limbic areas, since not all patients with high TMEScore have greater immunotherapeutic benefits, and therefore more clinical factors should be included in the prediction model to improve accuracy. Finally, the results obtained by bioinformatics analysis alone are not sufficient, and experimental verification is needed to confirm these results. Therefore, further genetic and experimental studies with larger sample size and experimental verification are needed.

In summary, in this study, we systematically evaluated the TME infiltration pattern from 757 GBM patients and developed a TME infiltration model approach. The TME score was found to be a powerful prognostic biomarker and predictor for immune checkpoint inhibitor responses.

Funding This study was supported by Shanghai Committee of Science and Technology (No. 15441904500) and Shanghai Municipal Science and Technology Major Project (NO. 2018SHZDZX03).

References

- Becht E, Giraldo NA, Lacroix L, Buttard B, Elarouci N, Petitprez F, Selves J, Laurent-Puig P, Sautès-Fridman C, Fridman WH, de Reyniès A (2016) Estimating the population abundance of tissue-infiltrating immune and stromal cell populations using gene expression. *Genome Biol* 17:218. <https://doi.org/10.1186/s13059-016-1070-5>
- Boillaud E, Molina G (2015) Are judgments a form of data clustering? Reexamining contrast effects with the k-means algorithm. *J Exp Psychol Hum Percept Perform* 41:415–430. <https://doi.org/10.1037/a0038896>
- Burr ML et al (2017) CMTM6 maintains the expression of PD-L1 and regulates anti-tumour immunity. *Nature* 549:101–105. <https://doi.org/10.1038/nature23643>
- Cheng W, Ren X, Zhang C, Cai J, Liu Y, Han S, Wu A (2016) Bioinformatic profiling identifies an immune-related risk signature for glioblastoma. *Neurology* 86:2226–2234. <https://doi.org/10.1212/WNL.0000000000002770>
- Cohen AL, Holmen SL, Colman H (2013) IDH1 and IDH2 mutations in gliomas. *Curr Neurol Neurosci Rep* 13:345. <https://doi.org/10.1007/s11910-013-0345-4>
- Couzin-Frankel J (2013) Breakthrough of the year 2013. *Cancer immunotherapy Science* 342:1432–1433. <https://doi.org/10.1126/science.342.6165.1432>
- Cristescu R et al (2018) Pan-tumor genomic biomarkers for PD-1 checkpoint blockade-based immunotherapy. *Science* 362:eaar3593. <https://doi.org/10.1126/science.aar3593>
- Ellor SV, Pagano-Young TA, Avgeropoulos NG (2014) Glioblastoma: background, standard treatment paradigms, and supportive care considerations. *J Law Med Ethics* 42:171–182. <https://doi.org/10.1111/jlme.12133>
- Farkona S, Diamandis EP, Blasutig IM (2016) Cancer immunotherapy: the beginning of the end of cancer? *BMC Med* 14:73. <https://doi.org/10.1186/s12916-016-0623-5>
- Fridman WH, Zitvogel L, Sautès-Fridman C, Kroemer G (2017) The immune contexture in cancer prognosis and treatment. *Nat Rev Clin Oncol* 14:717–734. <https://doi.org/10.1038/nrclinonc.2017.101>
- Fu H et al (2018) Identification and validation of stromal Immunity predict survival and benefit from adjuvant chemotherapy in patients with muscle-invasive bladder. *Cancer Clin Cancer Res* 24:3069–3078. <https://doi.org/10.1158/1078-0432.CCR-17-2687>
- George S et al (2017) Loss of PTEN is associated with resistance to anti-PD-1 checkpoint blockade therapy in metastatic uterine. Leiomyosarcoma *Immunity* 46:197–204. <https://doi.org/10.1016/j.immuni.2017.02.001>
- Ghasemi A, Zahediasl S (2012) Normality tests for statistical analysis: a guide for non-statisticians. *Int J Endocrinol Metab* 10:486–489. <https://doi.org/10.5812/ijem.3505>
- Hardcastle J et al (2017) Immunovirotherapy with measles virus strains in combination with anti-PD-1 antibody blockade enhances antitumor activity in glioblastoma treatment. *Neuro-Oncology* 19:493–502. <https://doi.org/10.1093/neuonc/now179>
- Hazra A, Gogtay N (2016) Biostatistics series module 3: comparing groups: numerical variables. *Indian J Dermatol* 61:251–260. <https://doi.org/10.4103/0019-5154.182416>
- Jiang Y, Zhang Q, Hu Y, Li T, Yu J, Zhao L, Ye G, Deng H, Mou T, Cai S, Zhou Z, Liu H, Chen G, Li G, Qi X (2018) ImmunoScore Signature: A Prognostic and Predictive Tool in Gastric Cancer. *Ann Surg* 267:504–513. <https://doi.org/10.1097/SLA.0000000000002116>
- Kalluri R (2016) The biology and function of fibroblasts in cancer. *Nat Rev Cancer* 16:582–598. <https://doi.org/10.1038/nrc.2016.73>
- Kim J, Bae JS (2016) Tumor-associated macrophages and neutrophils in tumor microenvironment. *Mediators Inflamm* 2016:6058147. <https://doi.org/10.1155/2016/6058147>
- Lee K, Hwang H, Nam KT (2014) Immune response and the tumor microenvironment: how they communicate to regulate gastric cancer. *Gut Liver* 8:131–139. <https://doi.org/10.5009/gnl.2014.8.2.131>
- Lee Y, Scheck AC, Cloughesy TF, Lai A, Dong J, Farooqi HK, Liau LM, Horvath S, Mischel PS, Nelson SF (2008) Gene expression analysis of glioblastomas identifies the major molecular basis for the prognostic benefit of younger age. *BMC Med Genet* 1:52. <https://doi.org/10.1186/1755-8794-1-52>
- Leonardi GC et al (2012) The tumor microenvironment in hepatocellular carcinoma (review). *Int J Oncol* 40:1733–1747. <https://doi.org/10.3892/ijo.2012.1408>
- Li JQ et al (2019) A multi-element expression score is a prognostic factor in glioblastoma multiforme. *Cancer Manag Res* 11:8977–8989. <https://doi.org/10.2147/cmar.s228174>

- Luoto S, Hermelo I, Vuorinen EM, Hannus P, Kesseli J, Nykter M, Granberg KJ (2018) Computational characterization of suppressive immune microenvironments in glioblastoma. *Cancer Res* 78:5574–5585. <https://doi.org/10.1158/0008-5472.CAN-17-3714>
- Maccio A, Madeddu C (2012) Inflammation and ovarian cancer. *Cytokine* 58:133–147. <https://doi.org/10.1016/j.cyto.2012.01.015>
- Malmstrom A et al (2012) Temozolomide versus standard 6-week radiotherapy versus hypofractionated radiotherapy in patients older than 60 years with glioblastoma: the Nordic randomised, phase 3 trial. *Lancet Oncol* 13:916–926. [https://doi.org/10.1016/S1470-2045\(12\)70265-6](https://doi.org/10.1016/S1470-2045(12)70265-6)
- Mantovani A, Marchesi F, Malesci A, Laghi L, Allavena P (2017) Tumour-associated macrophages as treatment targets in oncology. *Nat Rev Clin Oncol* 14:399–416. <https://doi.org/10.1038/nrclinonc.2016.217>
- Mariathasan S et al (2018) TGFbeta attenuates tumour response to PD-L1 blockade by contributing to exclusion of T cells. *Nature* 554:544–548. <https://doi.org/10.1038/nature25501>
- Mirzal A (2014) Nonparametric Tikhonov regularized nmf and its application in cancer clustering. *IEEE/ACM Trans Comput Biol Bioinform* 11:1208–1217. <https://doi.org/10.1109/TCBB.2014.2328342>
- Mohme M et al (2018) Immunophenotyping of newly diagnosed and recurrent glioblastoma defines distinct immune exhaustion profiles in peripheral and tumor-infiltrating lymphocytes. *Clin Cancer Res* 24:4187–4200. <https://doi.org/10.1158/1078-0432.CCR-17-2617>
- Newman AM et al (2015) Robust enumeration of cell subsets from tissue expression profiles. *Nat Methods* 12:453–457. <https://doi.org/10.1038/nmeth.3337>
- Nishino M, Ramaiya NH, Hatabu H, Hodi FS (2017) Monitoring immune-checkpoint blockade: response evaluation and biomarker development. *Nat Rev Clin Oncol* 14:655–668. <https://doi.org/10.1038/nrclinonc.2017.88>
- Ostrom QT, Gittleman H, Liao P, Vecchione-Koval T, Wolinsky Y, Kruchko C, Barnholtz-Sloan JS (2017) CBTRUS statistical report: primary brain and other central nervous system tumors diagnosed in the United States in 2010–2014. *Neuro-Oncology* 19:v1–v88. <https://doi.org/10.1093/neuonc/nox158>
- Phillips HS et al (2006) Molecular subclasses of high-grade glioma predict prognosis, delineate a pattern of disease progression, and resemble stages in neurogenesis. *Cancer Cell* 9:157–173. <https://doi.org/10.1016/j.ccr.2006.02.019>
- Piperi C, Papavassiliou KA, Papavassiliou AG (2019) Pivotal role of STAT3 in shaping glioblastoma immune microenvironment. *Cells* 8:E1398. <https://doi.org/10.3390/cells8111398>
- Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W, Smyth GK (2015) Limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res* 43:e47. <https://doi.org/10.1093/nar/gkv007>
- Roh W et al (2017) Integrated molecular analysis of tumor biopsies on sequential CTLA-4 and PD-1 blockade reveals markers of response and resistance. *Sci Transl Med* 9:eaah3560. <https://doi.org/10.1126/scitranslmed.aah3560>
- Rosenberg JE et al (2016) Atezolizumab in patients with locally advanced and metastatic urothelial carcinoma who have progressed following treatment with platinum-based chemotherapy: a single-arm, multicentre, phase 2 trial. *Lancet* 387:1909–1920. [https://doi.org/10.1016/S0140-6736\(16\)00561-4](https://doi.org/10.1016/S0140-6736(16)00561-4)
- Sharma P, Wagner K, Wolchok JD, Allison JP (2011) Novel cancer immunotherapy agents with survival benefit: recent successes and next steps. *Nat Rev Cancer* 11:805–812. <https://doi.org/10.1038/nrc3153>
- Sorlie T et al (2001) Gene expression patterns of breast carcinomas distinguish tumor subclasses with clinical implications. *Proc Natl Acad Sci U S A* 98:10869–10874. <https://doi.org/10.1073/pnas.191367098>
- Stupp R et al (2009) Effects of radiotherapy with concomitant and adjuvant temozolomide versus radiotherapy alone on survival in glioblastoma in a randomised phase III study: 5-year analysis of the EORTC-NCIC trial. *Lancet Oncol* 10:459–466. [https://doi.org/10.1016/S1470-2045\(09\)70025-7](https://doi.org/10.1016/S1470-2045(09)70025-7)
- Thakkar JP, Dolecek TA, Horbinski C, Ostrom QT, Lightner DD, Barnholtz-Sloan JS, Villano JL (2014) Epidemiologic and molecular prognostic review of glioblastoma. *Cancer Epidemiol Biomark Prev* 23:1985–1996. <https://doi.org/10.1158/1055-9965.EPI-14-0275>
- Turley SJ, Cremasco V, Astarita JL (2015) Immunological hallmarks of stromal cells in the tumour microenvironment. *Nat Rev Immunol* 15:669–682. <https://doi.org/10.1038/nri3902>
- Van Meir EG, Hadjipanayis CG, Norden AD, Shu HK, Wen PY, Olson JJ (2010) Exciting new advances in neuro-oncology: the avenue to a cure for malignant glioma. *CA Cancer J Clin* 60:166–193. <https://doi.org/10.3322/caac.20069>
- Walid MS (2008) Prognostic factors for long-term survival after glioblastoma. *Perm J* 12:45–48. <https://doi.org/10.7812/tpp/08-027>
- Yarchoan M, Xing D, Luan L, Xu H, Sharma RB, Popovic A, Pawlik TM, Kim AK, Zhu Q, Jaffee EM, Taube JM, Anders RA (2017) Characterization of the immune microenvironment in hepatocellular carcinoma. *Clin Cancer Res* 23:7333–7339. <https://doi.org/10.1158/1078-0432.CCR-17-0950>
- Ye C, Toyoda K, Ohtsuki T (2019) Blind source separation on non-contact heartbeat detection by non-negative matrix factorization algorithms. *IEEE Trans Biomed Eng.* <https://doi.org/10.1109/TBME.2019.2915762>
- Yoshihara K, Shahmoradgoli M, Martínez E, Vegesna R, Kim H, Torres-García W, Treviño V, Shen H, Laird PW, Levine DA, Carter SL, Getz G, Stemke-Hale K, Mills GB, Verhaak RG (2013) Inferring tumour purity and stromal and immune cell admixture from expression data. *Nat Commun* 4:2612. <https://doi.org/10.1038/ncomms3612>
- Yu N, Gao YL, Liu JX, Shang J, Zhu R, Dai LY (2018) Co-differential gene selection and clustering based on graph Regularized multi-view nmf in cancer genomic data genes (Basel). 9:E586. <https://doi.org/10.3390/genes9120586>
- Zeng D, Li M, Zhou R, Zhang J, Sun H, Shi M, Bin J, Liao Y, Rao J, Liao W (2019) Tumor microenvironment characterization in gastric cancer identifies prognostic and immunotherapeutically relevant gene signatures. *Cancer Immunol Res* 7:737–750. <https://doi.org/10.1158/2326-6066.CIR-18-0436>
- Zeng D et al (2018) Gene expression profiles for a prognostic immunoscore in gastric cancer. *Br J Surg* 105:1338–1348. <https://doi.org/10.1002/bjs.10871>
- Zhang S et al (2018) Specific breast cancer prognosis-subtype distinctions based on DNA methylation patterns. *Mol Oncol* 12:1047–1060. <https://doi.org/10.1002/1878-0261.12309>

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.