

## Research Paper

# Machine learning-based identification of tumor-infiltrating immune cell-associated lncRNAs for improving outcomes and immunotherapy responses in patients with low-grade glioma

Nan Zhang<sup>1,2,8#</sup>, Hao Zhang<sup>1,8,9#</sup>, Wantao Wu<sup>1,3,8</sup>, Ran Zhou<sup>4</sup>, Shuyu Li<sup>5</sup>, Zeyu Wang<sup>1,8</sup>, Ziyu Dai<sup>1,8</sup>, Liyang Zhang<sup>1,8</sup>, Fangkun Liu<sup>1,8</sup>, Zaoqu Liu<sup>6</sup>, Jian Zhang<sup>7</sup>, Peng Luo<sup>7</sup>, Zhixiong Liu<sup>1,8</sup>, Quan Cheng<sup>1,8</sup>

1. Department of Neurosurgery, Xiangya Hospital, Central South University, China.
2. One-third Lab, College of Bioinformatics Science and Technology, Harbin Medical University, China.
3. Department of Oncology, Xiangya Hospital, Central South University, China.
4. Division of Neuroscience and Experimental Psychology, Faculty of Biology, Medicine and Health, University of Manchester, UK.
5. Department of Thyroid and Breast Surgery, Tongji Hospital, Tongji Medical College of Huazhong University of Science and Technology, China.
6. Department of Interventional Radiology, The First Affiliated Hospital of Zhengzhou, China.
7. Department of Oncology, Zhujiang Hospital, Southern Medical University, China.
8. National Clinical Research Center for Geriatric Disorders, Xiangya Hospital, Central South University, China.
9. Department of Neurosurgery, The Second Affiliated Hospital, Chongqing Medical University, China.

#These authors contributed equally to this work.

✉ Corresponding authors: **Dr. Zhixiong Liu**, Department of Neurosurgery, Xiangya Hospital, Center South University, Address: Changsha 410008, Hunan, P. R. China. E-mail: zhixiongliu@csu.edu.cn; **Dr. Quan Cheng**, Department of Neurosurgery, Xiangya Hospital, Center South University, Address: Changsha 410008, Hunan, P. R. China. E-mail: chengquan@csu.edu.cn.

© The author(s). This is an open access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>). See <http://ivyspring.com/terms> for full terms and conditions.

Received: 2022.04.22; Accepted: 2022.07.28; Published: 2022.08.08

## Abstract

**Rationale:** Accumulating evidence demonstrated that long noncoding RNAs (lncRNAs) involved in the regulation of the immune system and displayed a cell-type-specific pattern in immune cell subsets. Given the vital role of tumor-infiltrating lymphocytes in effective immunotherapy, we explored the tumor-infiltrating immune cell-associated lncRNA (TIIClncRNA) in low-grade glioma (LGG), which has never been uncovered yet.

**Methods:** This study utilized a novel computational framework and 10 machine learning algorithms (101 combinations) to screen out TIIClncRNAs by integratively analyzing the sequencing data of purified immune cells, LGG cell lines, and bulk LGG tissues.

**Results:** The established TIIClnc signature based on the 16 most potent TIIClncRNAs could predict outcomes in public datasets and the Xiangya in-house dataset with decent efficiency and showed better performance when compared with 95 published signatures. The TIIClnc signature was strongly correlated to immune characteristics, including microsatellite instability, tumor mutation burden, and interferon  $\gamma$ , and exhibited a more active immunologic process. Furthermore, the TIIClnc signature predicted superior immunotherapy response in multiple datasets across cancer types. Notably, the positive correlation between the TIIClnc signature and CD8, PD-1, and PD-L1 was verified in the Xiangya in-house dataset.

**Conclusions:** The TIIClnc signature enabled a more precise selection of the LGG population who were potential beneficiaries of immunotherapy.

Key words: Immunotherapy, low-grade glioma, lncRNA, immune checkpoint, immune infiltration

## Introduction

Low-grade glioma (LGG) is a heterogeneous group of neuroepithelial tumors derived from supporting glial cells [1] and includes grades I and II

out of four glioma grades classified by the World Health Organization (WHO) [2]. Given the superior malignancy of glioblastoma (grade IV glioma), The

Cancer Genome Atlas (TCGA) database classifies grades II and III gliomas as LGG. The outcomes of LGG are highly variable, depending on age at diagnosis, histological subtype, tumor size, etc. [3]. Surgical resection with postoperative radiotherapy and chemotherapy is the conventional treatment for LGG patients [4]. Despite its relatively benign biological characteristics, LGG remains incurable and slowly develops till premature death [1]. Numerous recent studies have focused on the tumor immune microenvironment (TIME) and the interactions between tumor and immune cells [5]. Immunotherapy based on the immune response, including immune checkpoint blockade (ICB) and adoptive cell transfer (ACT), has revolutionized the therapeutic outcomes for tumor patients. Researchers have made unremitting efforts to find more precise targets for the further development of immunotherapy. However, studies on immunotherapy of LGG remain to be enriched.

Long noncoding RNAs (lncRNAs), a group of noncoding RNAs with more than 200 nucleotides, are closely related to diversified biological functions [6]. Many studies demonstrated that lncRNAs played a critical role in regulating cellular biological processes through modulating gene expression at transcriptional, post-transcriptional, and epigenetic levels [7]. Furthermore, the latest research has found that lncRNAs participated in the immune system modulation and showed a cell-type-specific pattern in immune cell subsets [8]. Ranzan et al. investigated over 500 lncRNAs and identified T cell-specific lncRNAs regulating cell differentiation [9]. Another study found lncRNA SATB2-AS1 as an essential regulator in colorectal cancer progression and immune cell density [10]. Given their close association with immune cell infiltration [8, 10], lncRNAs have colossal potential to evaluate immunotherapy response and predict clinical outcomes. Thus, it is promising to incorporate these powerful lncRNAs to develop prognostic biomarkers based on bioinformatics technology. Besides, as scientific research has entered the big data era with the fast development of high-throughput sequencing technologies, machine learning has been gradually widely applied to extract essential knowledge from big data bioinformatics.

This study aimed to screen tumor-infiltrating immune cell-associated lncRNA (TIIClnc) through a novel computational framework. Via integrative analysis of sequencing data of purified immune cells, LGG cell lines, bulk LGG tissues, and machine learning algorithms, we established a TIIClnc signature to stratify LGG patients and predict the outcomes of immunotherapy.

## Materials and methods

### LGG patient and tumor cell line cohorts collection

Transcriptome data and clinical information of LGG patients were accessed from three databases, The Cancer Genome Atlas (TCGA, <https://portal.gdc.cancer.gov>) via Illumina-HiSeq platform, Chinese Glioma Genome Atlas (CGGA, <http://www.cgga.org.cn/>) via Illumina-HiSeq platform, and Gene Expression Omnibus (GEO, <http://www.ncbi.nlm.nih.gov/geo>) via Affymetrix Human Genome U133 Plus 2.0 Array platform. The total number of samples included in this study was 932, including 518 samples from the TCGA LGG dataset, 77 samples from the Xiangya in-house dataset, 169 samples from the CGGA LGG dataset, and 168 samples from the GSE108474 dataset. The immune-associated lncRNA signature was constructed through the TCGA LGG training dataset and subsequently validated using the Xiangya in-house dataset (Glioma tissues were collected and written informed consent was obtained from all patients. The included glioma tissues were approved by the Ethics Committee of Xiangya Hospital, Central South University). CGGA LGG and GSE108474 datasets were used as the external validating datasets. Transcriptome data of 10 LGG cell lines via the Affymetrix Human Genome U133 Plus 2.0 Array platform was accessed from GSE36133 (Cancer Cell Line Encyclopedia project (CCLE)).

### Purified immune cell line cohorts collection

Transcriptome data of 115 purified cell lines from 19 major immune cell types via Affymetrix Human Genome U133 Plus 2.0 Array platform was accessed from 16 datasets, including GSE27291, GSE27838, GSE28490, GSE13906, GSE23371, GSE25320, GSE28698, GSE28726, GSE49910, GSE51540, GSE59237, GSE37750, GSE39889, GSE42058, GSE6863, GSE8059, and processed as previously described [8].

### Preprocessing of transcriptome data

The Robust Multi-array Average (RMA) algorithm [11] from the R package *affy* was used to perform quantile normalization, background correction, and log<sub>2</sub> transformation of microarray data from the Affymetrix platform. The sequencing data fragments per kilobase million (FPKM) values were changed into transcripts per kilobase million (TPM) values. The probes of microarray data from the Affymetrix platform were renamed to obtain lncRNA expression profiles. After matching the annotation file of GENCODE (release 39) with the NetAffx annotation files (release 36), probe sets with Ensembl gene IDs as 'long non-coding RNA' was picked out.

1711 characteristic lncRNAs matching with 2019 probes in microarray data from the Affymetrix platform were selected for further analysis. Correspondingly, lncRNA expression profiles (IlluminaHiSeq platform) were accessed from the TCGA and CGGA databases.

### Tumor-infiltrating immune cell-associated lncRNA signature establishment

Through integrative lncRNA profiling analysis of purified immune cells, LGG cell lines, and bulk LGG cancer tissues, a novel computational framework was utilized to identify a tumor-infiltrating immune cell-associated lncRNA (TIIClnc) signature on account of a couple of machine learning algorithms. The details are as follows (Figure 1):

- The top 15% expressed lncRNAs (each immune cell line) were taken for candidate immune-related lncRNAs.
- The tissue specificity index (TSI) proposed by Yanai et al. [12] was applied to calculate the expression specificity of candidate lncRNAs for each cell type:

$$TSI_{lnc} = \frac{\sum_{i=1}^N (1 - x_{lnc,i})}{N-1}$$

where  $N$  represents the total number of immune cell types and  $x_{lnc,i}$  represents the expression level of immune cell  $i$  normalized by the maximal expression intensity of lncRNA in any immune cell type. TSI ranges from 0 to 1, in which lncRNA is defined as an immune cell-general lncRNA when the value is 0 or an immune cell-specific lncRNA when the value is 1. The highly expressed lncRNAs in all immune cell types were identified as immune-related intrinsic lncRNAs (ilncRNA).

- ilncRNAs differentially expressed between immune cell lines (upregulated) and LGG cell lines (downregulated) were determined as TIIClncRNAs.
- Univariate Cox regression analysis was subsequently used to filtrate the TIIClncRNAs with prognostic potential in the TCGA LGG dataset.
- 101 combinations of 10 machine learning algorithms, including Lasso, Ridge, stepwise Cox, CoxBoost, random survival forest (RSF), elastic network (Enet), partial least squares regression for Cox (plsRcox), supervised principal components (SuperPC), generalized boosted regression modeling (GBM), and survival support vector machine (survival-SVM) based on a 10-fold cross-validation were further used to screen out the most valuable TIIClnc

signature with the highest C-index.

- The TIIClnc signature was established based on the combination of RSF and CoxBoost. CoxBoost algorithm was used to screen out the most valuable TIIClncRNAs. RSF algorithm was further used to filtrate the most reliable model. Log-rank score test for splitting survival trees was conducted as previously described [13]. First, the  $x$ -variable  $x$  was assumed to be ordered as  $x_1 \leq x_2 \leq \dots \leq x_n$ . Then, the "ranks" for each survival time  $T_j$  ( $j \in [1, \dots, n]$ ) were computed. The obtained equation is as follows:

$$a_j = \delta_j - \sum_{k=1}^{\Gamma_j} \frac{\delta_k}{n - \Gamma_k + 1}$$

where  $\Gamma_k = \#\{t : T_t \leq T_k\}$  and  $\Gamma_j$  represents the index of the order for  $T_j$ . The log-rank score test came as follows:

$$TIIClnc \text{ signature} = S(x, c) = \frac{\sum_{x_k \leq c} (a_j - n_i \bar{a})}{\sqrt{n_i (1 - \frac{n_i}{n}) s_a^2}}$$

where  $\bar{a}$  and  $s_a^2$  represent the sample mean and sample variance of  $[a_j : j = 1, \dots, n]$ , respectively. The measure of node separation is determined using log-rank score splitting by  $|S(x, c)|$ . The best split is reached by maximizing this value over  $x$  and  $c$ .

### Annotation of immune-related characteristics for the TIIClnc signature

Seven types of immune modulators were collected [14]. T cell-inflamed gene expression profile (GEP), Cytotoxic activity (CYT), and interferon  $\gamma$  (IFN- $\gamma$ ) were calculated [15, 16]. Tumor mutation burden (TMB), microsatellite instability (MSI), T cell receptor (TCR) richness, TCR Shannon, and SNV Neoantigen were collected from the TCGA database. GATK4 was used to search for the SNPs and indels from the RNA sequencing data of the Xiangya in-house dataset. ANNOVAR was used to annotate the mutation information based on CRCh38 [17]. The tmb function of the R package maftools was applied further to calculate the TMB value of the Xiangya in-house dataset. The R package PreMSIm was used to predict the MSI value of the Xiangya in-house dataset. Six immune subtypes and immunophenoscore (IPS) were determined as previously described [14, 18]. The Tumor Immune Estimation Resource (TIMER) algorithm [19], single cell gene set enrichment analysis (ssGSEA) algorithm [18], Microenvironment Cell Populations-counter (MCPcounter) algorithm [20], and Estimation of Stromal and Immune cells in Malignant Tumours using Expression data (ESTIMATE) algorithm [21] were applied for calculating the abundance of immune infiltrating cells and ESTIMATE score.

Cancer immunity cycle, displaying the functional status of chemokines and immunomodulators, and 114 metabolic pathways were collected and calculated using gene set variation analysis (GSVA) [22-24]. TIME signatures independently developed by Kobayashi [25] and Bagaev [26] were collected and calculated using GSVA. Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) terms were also quantified using GSVA and gene set enrichment analysis (GSEA).

### **Predictive value of the TIIClnc signature for immunotherapy response**

The GSE35640 (melanoma) [27], GSE91061 (melanoma) [28], GSE78220 (melanoma) [29], Allen (melanoma) [30], Nathanson (melanoma) [31], IMvigor (urothelial carcinoma) [32], Braun (renal cell carcinoma) [33], GSE179351 (colorectal adenocarcinoma and pancreatic adenocarcinoma) [34], GSE165252 (esophageal adenocarcinoma) [35], and PRJNA482620 (glioblastoma) [36] datasets were used to predict the immunotherapy response, while the TIIClnc signature was calculated in each dataset. The GSE103668 (triple-negative breast cancer) [37] dataset was used to predict the targeted therapy response (cisplatin & bevacizumab). The subclass mapping was utilized to predict anti-PD-1 and anti-CTLA-4 immunotherapy responses [38]. The Tumor Immune Dysfunction and Exclusion (TIDE) algorithm was also used in this section [39].

### **RNA sequencing**

RNAstore-fixed tumor tissues of 77 LGG samples from the Xiangya in-house dataset (Table S1) were collected for RNA sequencing as previously described [40]. Briefly, total RNA was extracted from the tumor tissues using TRIzol (Sigma-Aldrich, CA, USA). RNA purity and RNA integrity were checked using the NanoPhotometer spectrophotometer (IMPLEN, CA, USA) and the RNA Nano 6000 Assay Kit of the Bioanalyzer 2100 system (Agilent Technologies, CA, USA), respectively. A total amount of 1 µg RNA per sample was used as input material for the RNA sample preparations. Sequencing libraries were generated using NEBNext® Ultra™ RNA Library Prep Kit for Illumina® (NEB, USA).

### **Immunohistochemistry staining**

Paraffin-embedded tumor tissues of 20 of 77 LGG samples used for RNA sequencing from the Xiangya in-house dataset were further collected for immunohistochemistry (IHC) staining. Briefly, tissue sections were placed in citric acid antigen repair buffer (pH 6.0) in a microwave oven for antigen retrieval. Slices were placed in a 3% hydrogen peroxide solution to block endogenous peroxidase.

3% bovine serum albumin (BSA) was used as a blocking reagent. The sections were incubated with primary antibodies against CD8 (Mouse, 1:1000, 66868-1-Ig, Proteintech, China), PD-1 (Rabbit, 1:800, 18106-1-AP, Proteintech, China), PD-L1 (Mouse, 1:1000, 66248-1-Ig, Proteintech, China). The sections were then incubated with a horseradish peroxidase-conjugated secondary antibody (1:200, GB23303, Servicebio, China). 3,3'-diaminobenzidine (G1211, Servicebio, China) was finally used for coloration, and hematoxylin was used for counterstaining cell nuclei.

### **RT-qPCR assay**

The primers of GAPDH (F ACAGCCTCAAG ATCATCAGC; R GGTCATGAGTCCTTCCACGAT), LOC101928134 (F GAGCGAGGGTGATTGTCCA; R GAAGAGGGGAAGGGGTTCTC), and LOC100133461 (F GAGAGACCTGCCCAAGCATT; R TCCAGGTTCTGCATGTGTCC) were designed using the primer 5.0. Total RNAs were extracted and reversely transcribed into cDNA by HiScript Q RT SuperMix for RT-qPCR. The expression levels of LOC101928134 and LOC100133461 were quantified using 2- $\Delta\Delta$ CT.

### **Cell Counting Kit-8 (CCK-8) assay**

The THP-1 cells were cultured in 1640 medium with 10% fetal bovine serum (FBS). The THP-1 cells were seeded in a 96-well plate at 10<sup>4</sup> cells/hole density. Three groups of the THP-1 cells transfected with siRNAs were cultured for 24 h, 48 h, and 72 h, respectively. Each group has four duplicated holes. The absorbance at 450 nm was measured after hatching under the condition of 37 °C and 5% CO<sub>2</sub>.

### **EdU assay**

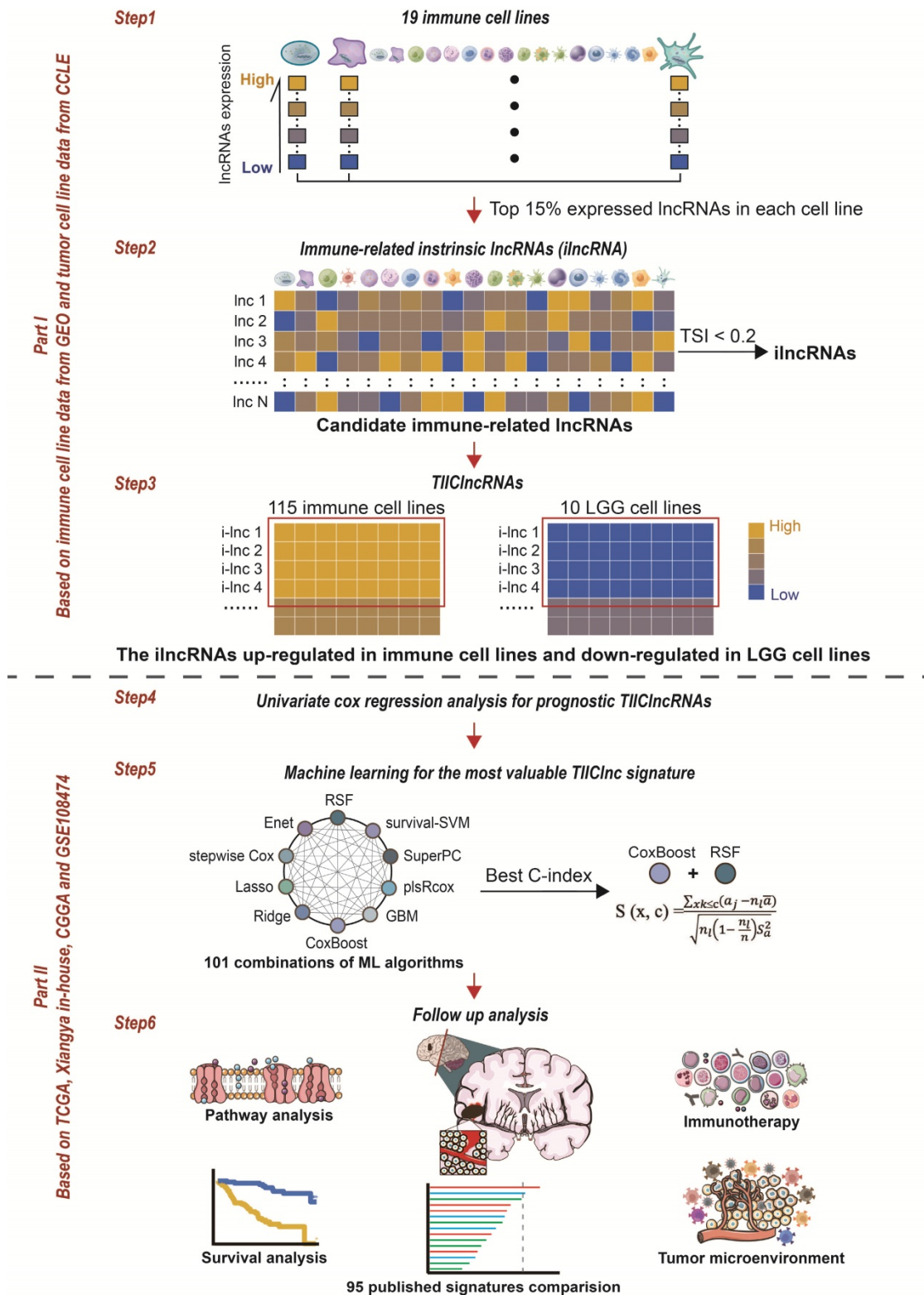
The EdU (5-ethynyl-2'-deoxyuridine) assay was performed according to the manufacturer's instructions (BeyoClick™ EdU Cell Proliferation Kit with Alexa Fluor 488, China). Three groups of the THP-1 cells transfected with siRNAs were incubated overnight with 50 µL 50 µM EdU medium and then fixed with 50 µL 4% paraformaldehyde. 100 µL 1x Apollo reaction solution was added for incubation for 30 min. Finally, the cells were incubated with 1 mL 1 × Hoechst 33342 reaction solution for 10 min and observed with a confocal microscope.

### **Transwell assay**

Three groups of the THP-1 cells transfected with siRNAs were centrifuged and resuspended using the serum-free medium. The density was adjusted to 10<sup>5</sup> cells/mL. 100 µL cell suspension was added to the upper chamber, and 500 µL 1640 with 10% FBS was added to the lower chamber. After culturing for 48 h,



the migrated THP-1 cells in the lower chamber were collected and counted by flow cytometry.



**Figure 1. The computational framework for establishing the TIIClnc signature.** The top 15% expressed lncRNAs were taken for candidate immune-related lncRNAs for each immune cell line. TSI was applied to calculate the expression specificity of candidate immune-related lncRNAs for each cell type. The highly expressed lncRNAs in all immune cell types were identified as immune-related ilncRNA. ilncRNAs significantly upregulated in immune cell lines and downregulated in LGG cell lines were defined as TIIClncRNAs. Univariate cox regression analysis was further used to screen out the prognostic TIIClncRNAs. All combinations of 10 machine learning algorithms, including RSF, Enet, Lasso, Ridge, stepwise Cox, CoxBoost, plsRcox, SuperPC, GBM, and survival-SVM, based on a 10-fold cross-validation were further used to screen out the most valuable TIIClnc signature with the highest C-index. The TIIClnc signature was finally generated based on the combination of RSF and CoxBoost. The association between the TIIClnc signature, prognosis, tumor immune microenvironment, and immunotherapy response was comprehensively investigated.

## Statistical analysis

Differentially expressed lncRNAs between immune and LGG cell lines were extracted by R packages 'limma'. Samples were grouped based on the cutoff value of the TIIClnc signature determined by the R package 'survminer'. The Kaplan-Meier survival plots were applied to estimate overall survival (OS) between the two TIIClnc signature groups using the R package 'survival.' C-index of OS was performed on the individual clinical variables, including the TIIClnc signature. The calibration curves of the TIIClnc signature were generated using the R package 'pec'. The predictive value of the TIIClnc signature for prognosis was measured with time-dependent receiver operating characteristic (ROC) curves using the R package 'timeROC.' For normally distributed variables, statistical differences between groups were determined by a two-tailed t-test while a one-way ANOVA test determined statistical differences among groups. For nonnormally distributed variables, statistical differences between groups were determined by a Wilcoxon test while a Kruskal-Wallis test determined statistical differences among groups. All the statistical analyses were performed in the R project, version 4.1.2.

## Results

### Identification of TIIClncRNAs

To comprehensively evaluate the immune cell-related lncRNA, 115 purified cell lines from 19 major immune cell types from 16 datasets were collected by searching for literature from 2007 to 2022 (Figure 1). 546 lncRNAs (top 15% expressed lncRNAs) were taken for candidate immune-related lncRNAs selection in each immune cell line. TSI score of the 546 lncRNAs was calculated to identify the ilncRNAs generally expressed in all 19 immune cell types (Table S2). Note that lncRNAs with a lower TSI score are generally highly expressed in different immune cell types, indicating their fundamental roles in immunity. 308 ilncRNAs were confirmed to be critical for regulating elemental immunity with the threshold of TSI < 0.2. By comparing the expression of 308 ilncRNAs in immune cell lines and LGG cell lines, 136 ilncRNAs were further found to be significantly upregulated in 115 immune cell lines and downregulated in 10 LGG cell lines. These 136 ilncRNAs were defined as LGG TIIClncRNAs.

### Development of the TIIClnc signature

To develop a prognostic TIIClnc signature, univariate Cox regression analysis regarding OS confirmed 46 prognostic TIIClncRNAs from 136 ilncRNAs in the TCGA LGG dataset. 10 machine

learning algorithms, including RSF, Enet, stepwise Cox, CoxBoost, plsRcox, Lasso, Ridge, SuperPC, GBM, and survival-SVM, were combined based on a 10-fold cross-validation to identify the most robust TIIClnc signature with the highest C-index in the TCGA LGG training dataset, Xiangya in-house validating dataset, and two external validating datasets (CGGA LGG and GSE108474) (Figure 2A). A final TIIClnc signature with the best performance was established based on the combined RSF and CoxBoost algorithms, which CoxBoost algorithm identified the 16 most valuable TIIClncRNAs (C7orf13, LINC00628, LINC01121, LOC100133461, LINC01134, TP53TG3HP, SLCO4A1-AS1, C1RL-AS1, LOC284395, TMEM72-AS1, PSMB8-AS1, DKFZp779M0652, LOC100506142, RPARP-AS1, CARD8-AS1, LOC101928134) (Figure 2B, Table S3) and RSF algorithm filtrated the most reliable model (Figure 2C). In accordance with our previous analyses, 16 TIIClncRNAs were generally expressed in all 19 immune cell types (Figure S1A). Besides, 16 TIIClncRNAs were significantly differentially expressed in immune cell lines (upregulation) and LGG cell lines (downregulation) (Figure S1B). Notably, LGG patients with high expression of 5 TIIClncRNAs had increased survival time, while LGG patients with high expression of 11 TIIClncRNAs had decreased survival time in the TCGA LGG dataset (Figure S2).

### In vitro validation of TIIClncRNAs

Among the 16 TIIClncRNAs, LOC100133461 and LOC101928134, with relatively higher expression in immune cell lines, were selected to further validate their potential roles in the TIME. RNA expression of LOC100133461 and LOC101928134 was significantly inhibited in three siRNAs groups of THP-1 cells, respectively (Figure S3A). The CCK-8 assay revealed the increased proliferation ability of THP-1 cells with inhibited expression of LOC100133461 and LOC101928134, respectively (Figure S3B and S3C). The EdU assay further proved the significantly increased proliferation ability of THP-1 cells with inhibited expression of LOC100133461 and LOC101928134, respectively (Figure S3D, S3E, S3F, S3G). Moreover, the Transwell assay revealed the significantly increased migration ability of THP-1 cells with inhibited expression of LOC100133461 and LOC101928134, respectively (Figure S3H, S3I, S3J, S3K).

### Prognostic value of the TIIClnc signature

LGG patients with high TIIClnc signature scores had decreased survival time in the TCGA LGG, Xiangya in-house, CGGA LGG, and GSE108474 datasets (Figure 2D). Consistently, time-dependent

ROC curves of 1-year, 2-year, 3-year, 4-year, and 5-year OS in the TCGA LGG dataset (AUC values, 0.913, 0.932, 0.960, 0.947, 0.938, respectively), Xiangya in-house dataset (AUC values, 0.755, 0.772, 0.827, 0.810, 0.861, respectively), CGGA LGG dataset (AUC values, 0.749, 0.779, 0.816, 0.800, 0.812, respectively), and GSE108474 dataset (AUC values, 0.656, 0.707, 0.715, 0.645, 0.642, respectively) confirmed the prognostic value of the TIIClnc signature (Figure 2E). The C-index of clinical factors in LGG patients was calculated. The univariate and multivariate Cox regression analysis was performed on clinical factors. The TIIClnc signature was found to be an independent prognostic factor like age, gender, grade, isocitrate dehydrogenase (IDH) status, 1p/19q status, methylation of O6-methylguanine-DNA methyltransferase (MGMT) status, with an advantage over all these clinical factors (Figure 3A, Table S4). Notably, the combination signature incorporating the TIIClnc signature, age, gender, grade, IDH status, 1p/19q status, and MGMT status showed better predictive efficacy in prognosis (Figure 3A). The calibration curves further proved the predictive accuracy of the TIIClnc signature (Figure 3B).

### Comparison of prognostic signatures in gliomas

Thanks to the development of next-generation sequencing and technologies for mining big data, predictive gene expression-based signatures have been widely explored and developed. For a comprehensive comparison of the performance of the TIIClnc signature with other signatures, the published signatures over the past ten years were systematically retrieved. Ultimately, 95 signatures (including mRNA and lncRNA signatures) were enrolled in this study (Table S5). These 95 signatures were closely related to different biological features, including immunotherapy response, immune infiltration, autophagy, ferroptosis, pyroptosis, stemness, epithelial-mesenchymal transition, hypoxia, glycolysis, epigenetics, N6-methyladenosine, and aging. Notably, the TIIClnc signature displayed better performance regarding C-index in the TCGA LGG (Figure 3C), Xiangya in-house (Figure 3D), CGGA LGG (Figure 3E), and GSE108474 (Figure 3F) datasets than almost all models.

### Immune characteristics related to the TIIClnc signature

To assess the role of the TIIClnc signature in the TIME of LGG, the relationship between the TIIClnc signature and immune cell infiltration and immune modulators was explored. Based on the TIMER algorithm, MCPcounter algorithm, and ssGSEA

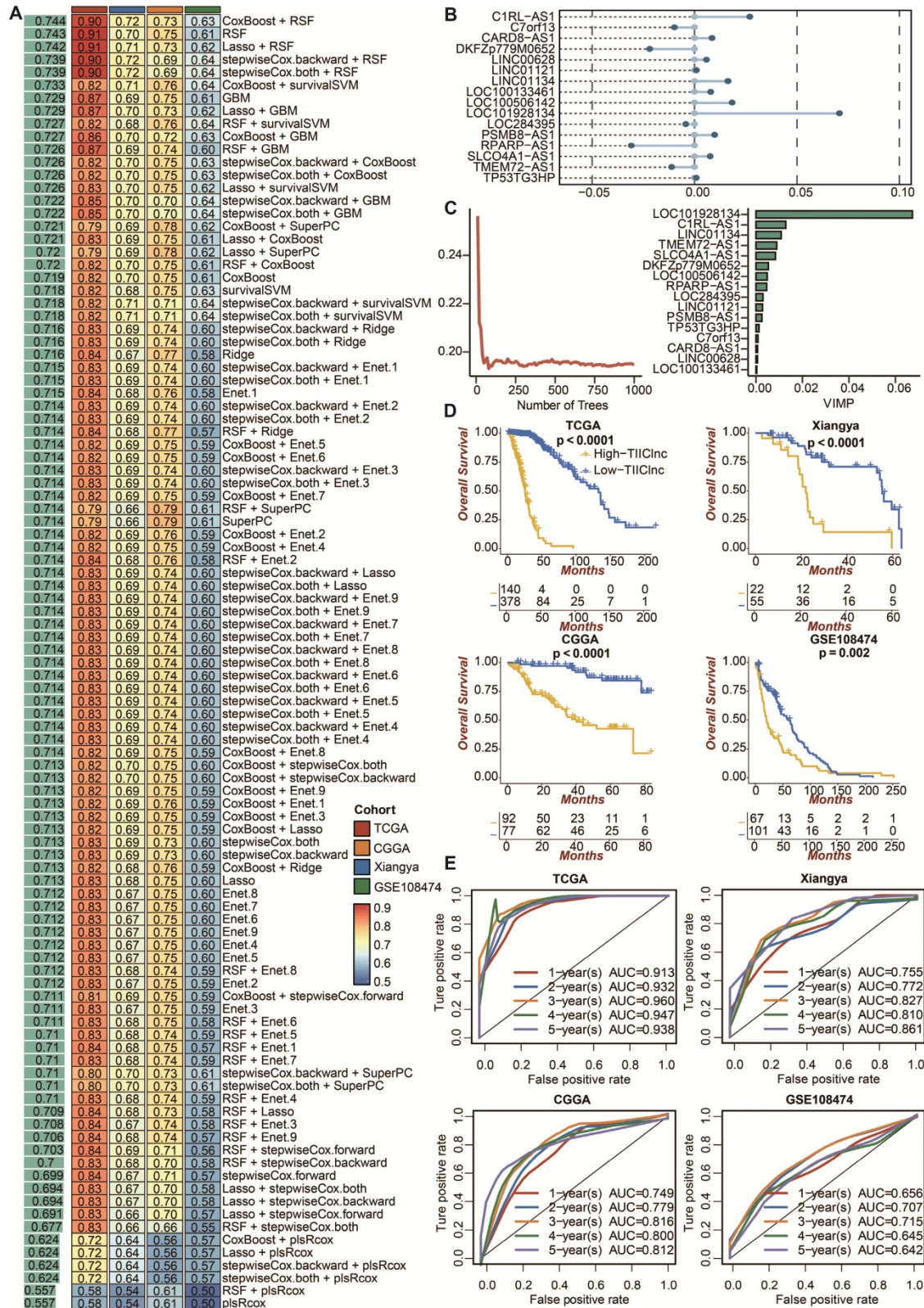
algorithm, the TIIClnc signature score was found to be positively correlated with almost all tumor immune infiltrating cells, such as T cells, T helper (Th) cells, natural killer (NK) cells, B cells, dendritic cell (DC), mast cells, myeloid-derived suppressor cell (MDSC), fibroblasts, macrophages, and regulatory T (Treg) cells in the TCGA LGG (Figure S4A), Xiangya in-house (Figure 4A), CGGA LGG (Figure S6A), and GSE108373 (Figure S7A) datasets. Based on the ESTIMATE algorithm, the TIIClnc signature score was positively correlated with a stromal score, immune score, and ESTIMATE score except for tumor purity in the TCGA LGG (Figure S4A), Xiangya in-house (Figure 4A), CGGA LGG (Figure S6A), and GSE108373 (Figure S7A) datasets. In addition, the TIIClnc signature score positively correlated with most immune modulators categorized into antigen presentation, cell adhesion, co-inhibitor, co-stimulator, ligand, receptor, and others in the TCGA LGG (Figure S4B), Xiangya in-house (Figure 4B), CGGA LGG (Figure S6B), and GSE108373 (Figure S7B) datasets. Notably, a strong correlation was observed between the TIIClnc signature score and classical immune checkpoint molecules, including IDO-1, LAG-3, PD-1, PD-L1, PD-L1, CTLA-4, BTLA, and TIGIT (Figure 4B, S4B, S6B, S7B).

### The TIIClnc signature is a predictive biomarker for immunotherapy response

To further explore the role of the TIIClnc signature in immunotherapy response, the relationship between the TIIClnc signature and several immunotherapy predictors was explored (Table S6). Notably, the MSI (Figure S5A), TMB (Figure S5B), CYT (Figure S5C), GEP (Figure S5D), TCR richness (Figure S5E), TCR Shannon (Figure S5F), SNV Neoantigen (Figure S5G), and IFN- $\gamma$  (Figure S5H) levels were all significantly higher in the high TIIClnc signature score group in the TCGA LGG dataset, all of which were determinants of better immunotherapy response. The significant expression differences of CYT (Figure 4C, S6C, S7C), GEP (Figure 4D, S6D, S7D), and IFN- $\gamma$  (Figure 4E, S6E, S7E) between two TIIClnc signature score groups were also verified in the Xiangya in-house, CGGA LGG, and GSE108474 datasets. Furthermore, the MSI high group was significantly associated with high TIIClnc signature scores (Figure 4F), and the TMB had a substantially higher level in the high TIIClnc signature score group (Figure 4G). Besides, a high TIIClnc signature score was more associated with lymphocyte depleted and inflammatory immune subtypes and high IPS score in the TCGA LGG dataset (Figure S5I), which the significant expression difference of IPS score between the two TIIClnc

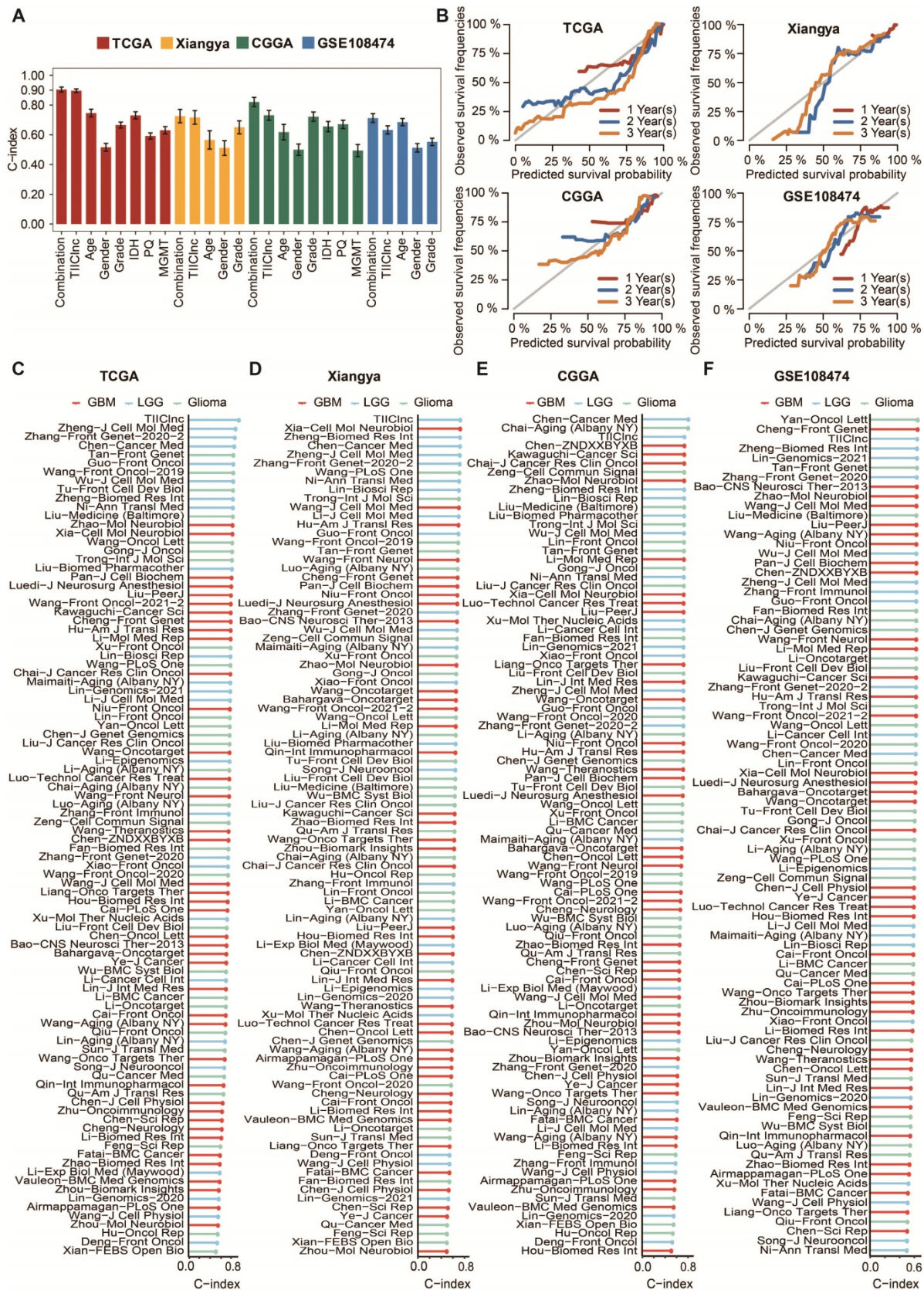


signature score groups was also verified in the Xiangya in-house, CGGA LGG, and GSE108474 datasets (Figure 4H, S6F, S7F).



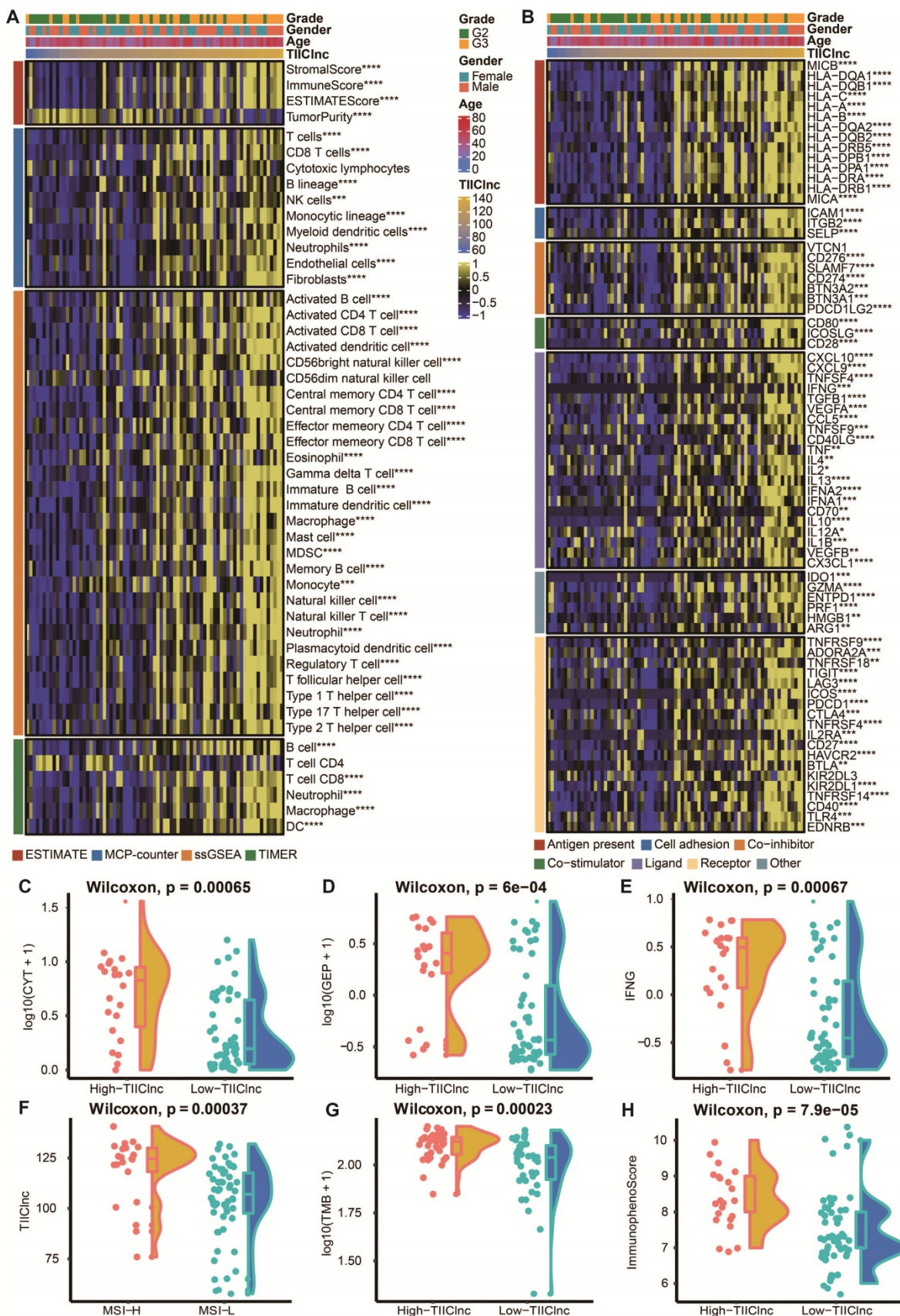
**Figure 2. The prognostic value of the TIICnc signature. A.** A total of 101 combinations of machine learning algorithms for the TIICnc signature via a 10-fold cross-validation framework. The C-index of each model was calculated across validation datasets, including TCGA LGG, Xiangya in-house, CGGA LGG, and GSE108474 datasets. **B.** The exhibition of the 16 most valuable TIICncRNAs based on the CoxBoost algorithm. **C.** The number of trees for determining the TIICnc signature with minimal error and the importance of the 16 most valuable TIICncRNAs based on the RSF algorithm. **D.** Kaplan-Meier survival curve of OS between patients with a high score of TIICnc signature and with a low score of TIICnc signature in the TCGA LGG, Xiangya in-house, CGGA LGG, and GSE108474 datasets. **E.** Time-dependent ROC curves of 1-year, 2-year, 3-year, 4-year, and 5-year OS in the CGGA LGG, Xiangya in-house, TCGA LGG, and GSE108474 datasets.





**Figure 3.** Comparison between the TIICnc signature and other models. **A.** The C-index of the TIICnc signature, other clinical factors, and the combination signature in the TCGA LGG, Xiangya in-house, CGGA LGG, and GSE108474 datasets. **B.** The 1-year, 2-year, and 3-year calibration curves of the TIICnc signature in the TCGA LGG, Xiangya in-house, CGGA LGG, and GSE108474 datasets. **C.** The C-index of the TIICnc signature and other models developed in the TCGA LGG dataset. **D.** The C-index of the TIICnc signature and other models developed in the Xiangya in-house dataset. **E.** The C-index of the TIICnc signature and other models developed in the CGGA LGG dataset. **F.** The C-index of the TIICnc signature and other models developed in the GSE108474 dataset.





**Figure 4.** Immune-related characteristics of the TIICinc signature in the Xiangya in-house dataset. **A.** Heatmap displaying the correlation between the TIICinc signature and immune infiltrating cells. **B.** Heatmap displaying the correlation between the TIICinc signature and immune modulator molecules. **C.** Box plot displaying the CYT levels between two TIICinc signature score groups. **D.** Box plot displaying the GEP levels between two TIICinc signature score groups. **E.** Box plot displaying the IFN- $\gamma$  levels between two TIICinc signature score groups. **F.** Box plot displaying the TIICinc levels between two MSI groups. **G.** Box plot displaying the TMB levels between two TIICinc signature score groups. **H.** Box plot displaying the IPS levels between two TIICinc signature score groups.

Moreover, the TIIClnc signature was directly established in immunotherapy datasets to evaluate its predictive value. In the IMvigor dataset, urothelial carcinoma patients with high TIIClnc signature scores had a prolonged survival time (Figure 5A). As expected, urothelial carcinoma patients with high TIIClnc signature scores were more likely to respond to anti-PD-L1 immunotherapy (Figure 5B). In the Braun dataset, renal cell carcinoma patients with high TIIClnc signature scores had prolonged survival time (Figure 5D). As expected, renal cell carcinoma patients with high TIIClnc signature scores tended to respond to anti-PD-1 immunotherapy (Figure 5E). In addition, patients in the GSE179351 (colorectal adenocarcinoma and pancreatic adenocarcinoma) (Figure 5C), and GSE165252 (esophageal adenocarcinoma) (Figure 5G) datasets with high TIIClnc signature scores were also more likely to respond to immunotherapy. Notably, patients in the GSE103668 (triple-negative breast cancer) dataset with high TIIClnc signature scores were more likely to respond to targeted therapy (Figure 5F). In the Allen dataset, melanoma patients with high TIIClnc signature scores had prolonged survival time (Figure 5H). As expected, melanoma patients with high TIIClnc signature scores tended to respond to anti-CTLA-4 immunotherapy (Figure 5I). In the GSE78220 dataset, melanoma patients with high TIIClnc signature scores had prolonged survival time (Figure 5J). As expected, melanoma patients with high TIIClnc signature scores tended to respond to anti-PD-1 immunotherapy (Figure 5K). In the Nathanson dataset, melanoma patients with high TIIClnc signature scores had prolonged survival time (Figure 5L), and melanoma patients with high TIIClnc signature scores tended to respond to anti-CTLA-4 immunotherapy (Figure 5M). In addition, patients in the GSE35640 (melanoma) (Figure 5N) and GSE91061 (melanoma) (Figure 5O) datasets with high TIIClnc signature scores were also more likely to respond to immunotherapy. Based on the TIDE algorithm, a high TIIClnc signature score was significantly associated with immune checkpoint inhibitors response in the TCGA LGG (Figure S8A), Xiangya in-house (Figure 5P), CGGA LGG (Figure S8C), and GSE108474 (Figure S8E) datasets. Based on the submap analysis, a high TIIClnc signature score was related to anti-CTLA-4 and anti-PD-1 immunotherapy responses in the Xiangya in-house dataset (Figure 5Q). In contrast, a high TIIClnc signature score was mainly associated with anti-PD-1 immunotherapy response in the TCGA LGG (Figure S8B), CGGA LGG (Figure S8D), and GSE108474 (Figure S8F) datasets. Notably, in an immunotherapy dataset of glioblastoma patients, PRJNA482620, patients with high TIIClnc signature scores also had prolonged survival time (Figure 5R).

The insignificant statistical analysis could be attributed to the small sample size.

Subsequently, the predictive value of the TIIClnc signature in immunotherapy response was explored in the Xiangya in-house dataset. Scatter plots of the TIIClnc signature score and CD8, PD-1, and PD-L1 demonstrated a significant positive correlation (Figure 6A). IHC images (Figure 6B) and H-scores (Figure 6C) further proved that the protein expressions of CD8, PD-1, and PD-L1 were dramatically higher in the high TIIClnc signature score group. Our research verified that patients with a high TIIClnc signature score could benefit more from immunotherapy than those with a low TIIClnc signature score.

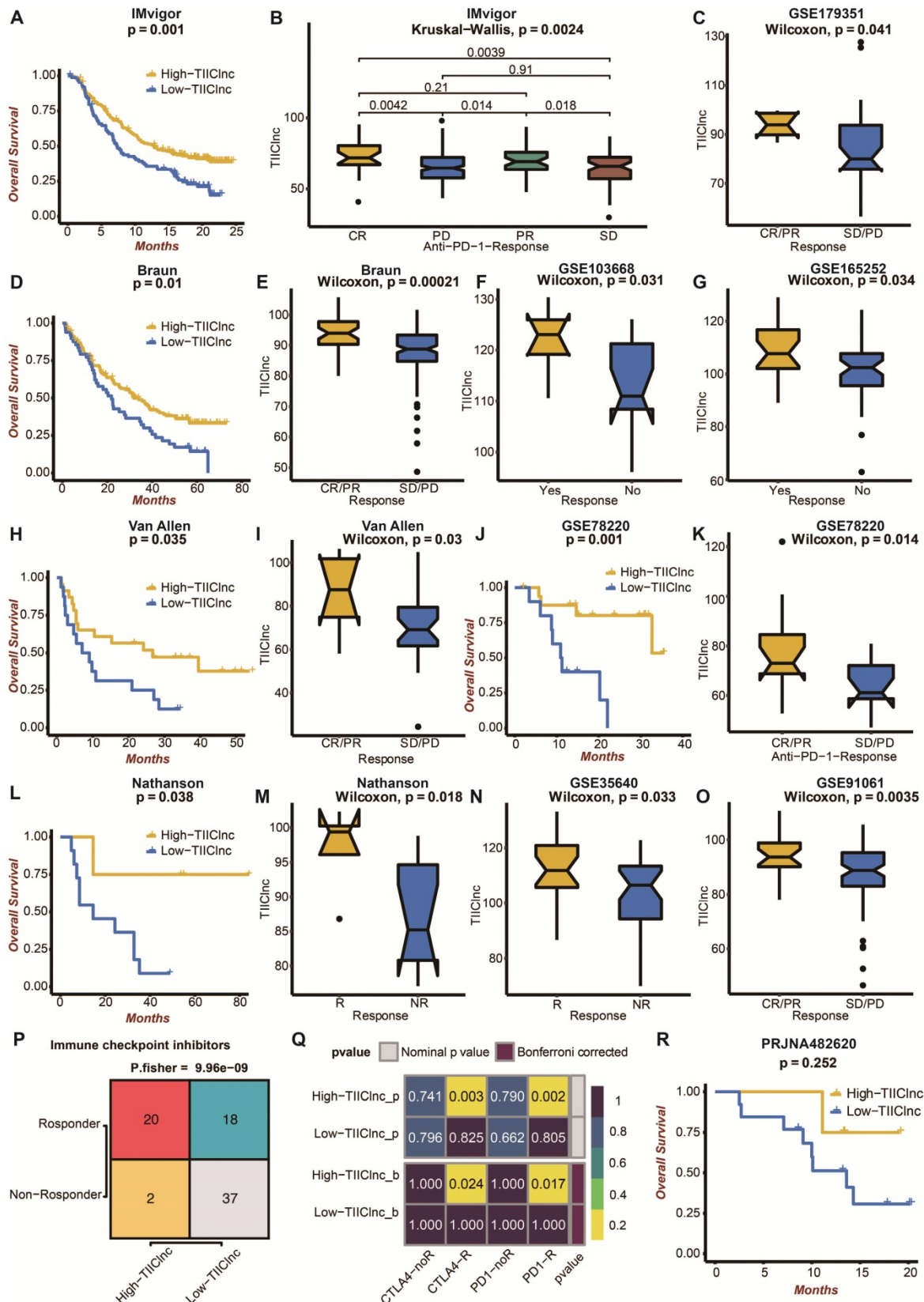
### Potential biological mechanisms related to the TIIClnc signature

The cancer immunity cycle was calculated to explore the potential biological mechanisms related to the TIIClnc signature. Of note, all of the seven steps in the cancer immunity cycle were more activated in the high TIIClnc signature score group, including antigen release (Step 1), cancer antigen presentation (Step 2), priming and activation (Step 3), tumor immune infiltrating cells recruitment (Step 4), immune cells infiltration (Step 5), cancer cells recognition by T cells (Step 6), and cancer cells killing (Step 7) in the TCGA LGG (Figure S9A), Xiangya in-house (Figure 7A), CGGA LGG (Figure S10A), and GSE108474 (Figure S11A) datasets. The TIIClnc signature score positively correlated with the metabolic activity of metabolites, such as retinoid, glycan, galactose, or glutathione, in the TCGA LGG (Figure S9B), Xiangya in-house (Figure 7B), CGGA LGG (Figure S10B), and GSE108474 (Figure S11B) datasets. In the GSVA of GO and KEGG terms, the TIIClnc signature score positively correlated with T cell activity, macrophage activity, Treg differentiation, regulation of fibroblast proliferation, DC migration, and pathways in cancer (Figure S9B, 7B, S10B, S11B). In the analysis of TIME signatures, a high TIIClnc signature score was associated with higher levels of innate immunity, T cells, IFN- $\gamma$  response, Treg, MDSC, recognition of tumor, proliferation, glycolysis developed by Kobayashi and antigen presentation, cytotoxic T and NK cells, anti-tumor microenvironment, checkpoint inhibition, Treg, granulocytes, MDSC, tumor promotive immune infiltrate, CAF, angiogenesis, tumor features developed by Bagaev in the TCGA LGG (Figure S9C and S9D), Xiangya in-house (Figure 7C and 7D), CGGA LGG (Figure S10C and S10D), and GSE108474 (Figure S11C and S11D) datasets. Additionally, in the GSEA of GO (Figure S9E, 7E, S10E, S11E) and KEGG (Figure S9F, 7F, S10F, S11F)



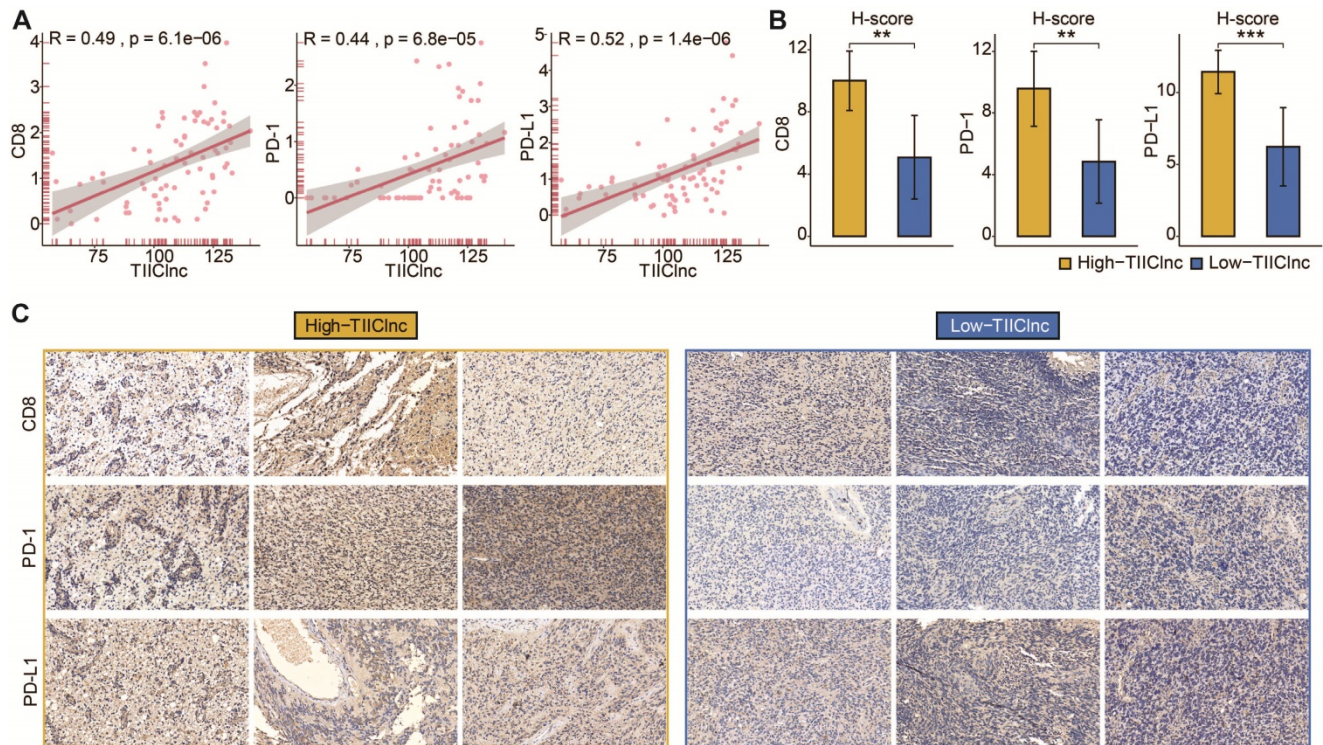
terms, T cell chemotaxis, T cell migration, response to IFN- $\gamma$ , regulation of macrophage activation, PD-L1 expression and PD-1 checkpoint pathway in cancer,

PI3K-Akt pathway, and NF-kappa B pathway, JAK-STAT pathway was more enriched in the high TIIClnc signature score group.



**Figure 5. Predictive value of the TIIClnc signature in immunotherapy response.** **A.** Kaplan-Meier survival curve of OS between patients with a high score of TIIClnc signature and a low score of TIIClnc signature in the IMvigor dataset. **B.** Box plot displaying the TIIClnc signature score in patients with different immunotherapy responses in the

IMvigor dataset. **C.** Box plot displaying the TIIClnc signature score in patients with different immunotherapy responses in the GSE179351 dataset. **D.** Kaplan-Meier survival curve of OS between patients with a high score of TIIClnc signature and a low score of TIIClnc signature in the Braun dataset. **E.** Box plot displaying the TIIClnc signature score in patients with different immunotherapy responses in the Braun dataset. **F.** Box plot displaying the TIIClnc signature score in patients with different immunotherapy responses in the GSE103668 dataset. **G.** Box plot displaying the TIIClnc signature score in patients with different immunotherapy responses in the GSE165252 dataset. **H.** Kaplan-Meier survival curve of OS between patients with a high score of TIIClnc signature and a low score of TIIClnc signature in the Allen dataset. **I.** Box plot displaying the TIIClnc signature score in patients with different immunotherapy responses in the Allen dataset. **J.** Kaplan-Meier survival curve of OS between patients with a high score of TIIClnc signature and a low score of TIIClnc signature in the GSE78220 dataset. **K.** Box plot displaying the TIIClnc signature score in patients with different immunotherapy responses in the GSE78220 dataset. **L.** Kaplan-Meier survival curve of OS between patients with a high score of TIIClnc signature and a low score of TIIClnc signature in the Nathanson dataset. **M.** Box plot displaying the TIIClnc signature score in patients with different immunotherapy responses in the Nathanson dataset. **N.** Box plot displaying the TIIClnc signature score in patients with different immunotherapy responses in the GSE35640 dataset. **O.** Box plot displaying the TIIClnc signature score in patients with different immunotherapy responses in the GSE91061 dataset. **P.** Contingency table between immunotherapy responses and TIIClnc signature score groups based on TIDE algorithm in the Xiangya in-house dataset. **Q.** Contingency table between immunotherapy responses (anti-PD-1 and anti-CTLA-4) and TIIClnc signature score groups based on submap analysis in the Xiangya in-house dataset. **R.** Kaplan-Meier survival curve of OS between patients with a high score of TIIClnc signature and a low score of TIIClnc signature in the PRJNA482620 dataset.



**Figure 6.** **A.** Scatter plot displaying the correlation between the TIIClnc signature score and CD8, PD-1, and PD-L1 in the Xiangya in-house dataset. **B.** Box plot displaying the H-score levels of CD8, PD-1, and PD-L1 based on IHC staining between two TIIClnc signature score groups in the Xiangya in-house dataset. The H-score was calculated by intensity score \* quantity score. As for intensity scores, 0, 1, 2, and 3 were considered negative, weak, moderate, and strong, respectively. As for quantity scores, 0, 1, 2, 3, and 4 represented <10%, 10-25%, 25-50%, 50-75%, >75% proportion of stained cells, respectively. H-score ranges from 0 to 12. **C.** Representative IHC staining images of CD8, PD-1, and PD-L1 in two TIIClnc signature score groups in the Xiangya in-house dataset.

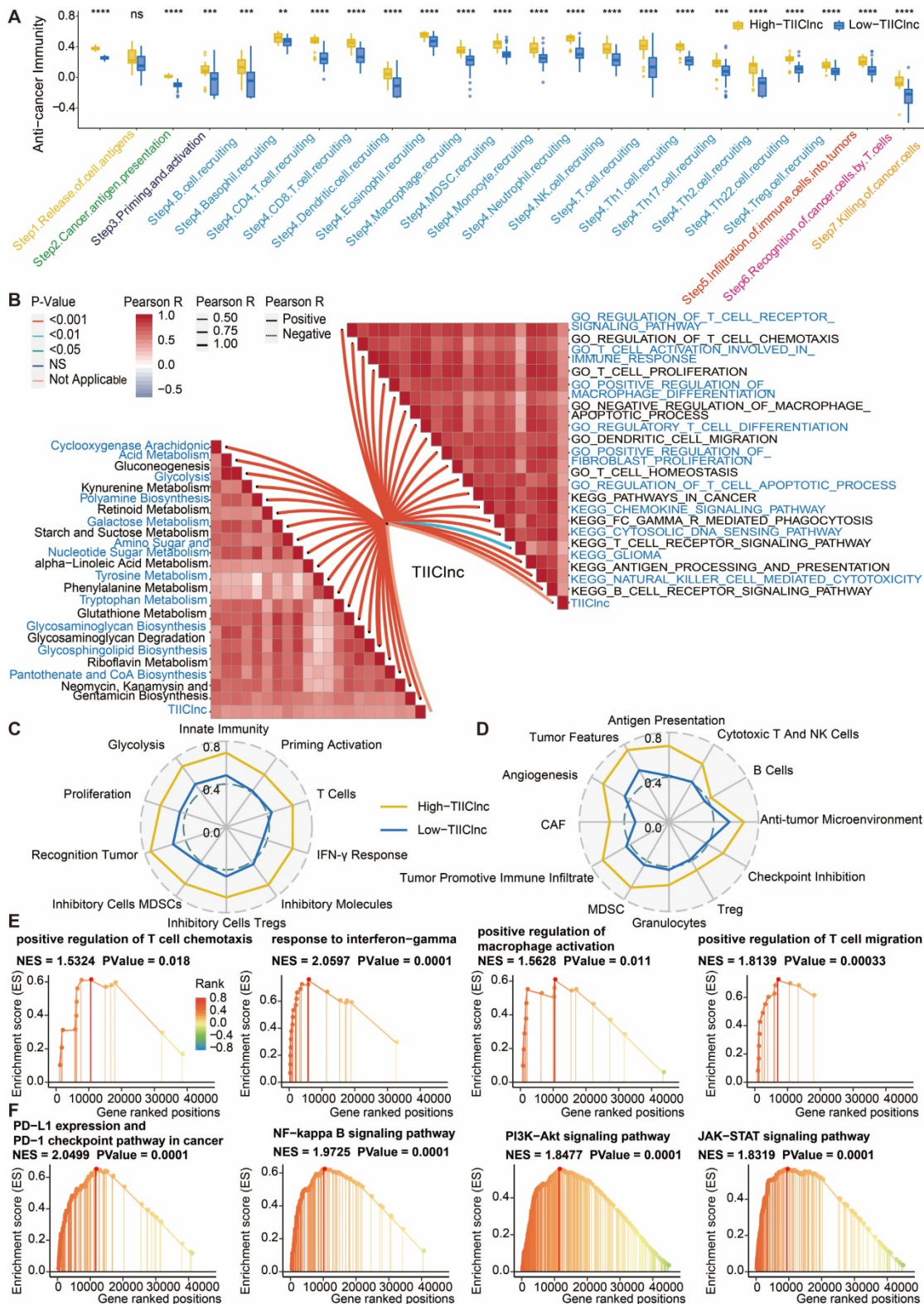
## Discussion

Compared with glioblastoma patients, LGG patients experienced a relatively prolonged survival time of 5-10 years. However, the treatment options for LGG are still limited to chemoradiotherapy and target therapy of tyrosine kinase receptor pathway [41]. The insufficient treatment options could sometimes lead to the overtreatment or undertreatment of LGG patients. Immunotherapy has become a leading star in the treatment of solid tumors. Although PD-1 has been widely detected among LGG patients, anti-PD-1 and anti-PD-L1 immunotherapies have not widely entered clinical trials of LGG patients partly due to the poor immunotherapy response and the results of several clinical trials (NCT03718767) remain to be announced [42]. For this reason, establishing reliable prognostic biomarkers to stratify LGG patients who

might benefit from immunotherapy is urgently needed.

This study utilized a novel computational framework to identify a robust and stable TIIClnc signature. Firstly, TSI was applied to screen out the ilncRNAs generally expressed in immune cells, indicating their fundamental roles in immune system regulation. Secondly, 101 combinations of 10 machine learning algorithms were used to identify the combination of CoxBoost and RSF as the optimal model based on the prognostic TIIClncRNAs, significantly reducing the dimensionality of variables and revealing underlying patterns, contributing to a more simplified and translational model. Therefore, the established TIIClnc signature conceivably precisely predicts LGG patients' prognosis and immune characteristics.





**Figure 7. Functional annotation of the TIICnc signature in the Xiangya in-house dataset. A.** Box plot displaying the cancer immunity cycle differences between two TIICnc signature score groups. **B.** Butterfly plot displaying the correlation between the TIICnc signature score and metabolic pathways, immune-related pathways based on GSEA of GO and KEGG terms. Immunogram radar plot displaying the correlation between the TIICnc signature score and TIME signatures developed by **C.** Kobayashi and **D.** Bagaev. **E.** GSEA of GO terms for the TIICnc signature score. **F.** GSEA of KEGG terms for the TIICnc signature score.

Among the 16 most valuable TIICncRNAs identified for the TIICnc signature, nine lncRNAs have been reported. C7orf13 was proved with an

inverse correlation with DNA methylation in glioblastoma [43]. LINC00628 contributes to lung adenocarcinoma by epigenetically interacting with the



LAMA3 promoter [44]. LINC01121 promotes cell proliferation, migration, and invasion of breast cancer via miR-150-5p/HMGA2 axis [45]. LINC01134 dictates hepatocellular carcinoma progression by interplaying with YY1 [46]. SLCO4A1-AS1 could predict poor prognosis and promote proliferation and metastasis of colorectal cancer via the EGFR/MAPK axis [47]. C1RL-AS1 drives the malignant phenotype of gastric cancer via the AKT/ $\beta$ -Catenin/c-Myc axis [48]. PSMB8-AS1 leads to pancreatic cancer progression via miR-382-3p/STAT1/PD-L1 axis [49]. RPARP-AS1/miR125a-5p axis promoted proliferation, migration, and invasion of colon Cancer [50]. CARD8-AS1 was found to regulate the metastasis of glioma [51]. The detailed reports of other 7 lncRNAs are lacking. Our research revealed that these 16 TIIClncRNAs were highly expressed in immune cells. Notably, LOC100133461 and LOC101928134 were found to inhibit the proliferation and migration ability of THP-1 cells, which was in accordance with the finding that they were hazardous markers in LGG patients. Thus, LOC100133461 and LOC101928134 were likely to facilitate tumor progression by suppressing the activity of monocytes in the TIME of LGG.

The predictive analysis demonstrated that the TIIClnc signature was a hazardous marker of OS in LGG patients. The ROC analysis further proved that the TIIClnc signature possessed high accuracy in predicting 1-year, 2-year, 3-year, 4-year, and 5-year OS of LGG patients. The stable performance of the TIIClnc signature in the TCGA LGG training dataset, Xiangya in-house dataset, and CGGA LGG, GSE108474 external validating datasets indicated the massive potential for the clinical application of the TIIClnc signature. IDH, 1p/19q, and MGMT status have long been determined as risk biomarkers for evaluating clinical strategies and outcomes of glioma patients [52, 53]. Notably, the TIIClnc signature was an independent risk factor with significantly superior performance than the above three risk biomarkers and clinical factors: age, grade, and gender. In addition, 95 published signatures of various functional gene combinations were retrieved for comparison. Few of these 95 signatures have been put into clinical practice [54]. Besides, many models displayed exemplary performance in the training dataset but weak performance in the validating datasets, indicating these models' poor universality and generalizability. Notably, according to the C-index assessment, the TIIClnc signature performed better than almost 95 signatures. It was conceivable that the feature gene selection and statistical prediction based on the best fit model performed by two combined machine learning algorithms ensured the stability and potential of our

TIIClnc signature.

The TIME is closely associated with the prognosis of brain tumors and the efficacy of immunotherapy [55, 56]. LGG patients with high TIIClnc signature score presented abundant immune cell infiltration, including T cells, Th cells, NK cells, B cells, DC, mast cells, MDSC, fibroblasts, macrophages, and Treg cells, all of which were related to anti-tumor or pro-tumor immunity in immunotherapy [57-62]. Immunotherapy has demonstrated considerable benefits in cancer patients with MSI-H [63]. TMB could enhance tumor immunogenicity and activate cytotoxic T cells; cancer patients with high TMB also benefit more from immunotherapy [64]. In this study, LGG patients with high TIIClnc signature scores were prone to higher MSI and TMB. Meanwhile, LGG patients with high TIIClnc signature scores were also associated with higher levels of predictors for better immunotherapy responses, including CYT, GEP, TCR richness, TCR Shannon, SNV Neoantigen, IFN- $\gamma$ , and IPS score. Moreover, the TIIClnc signature score was positively associated with classical immune checkpoint molecules, including IDO-1, LAG-3, PD-1, PD-L1, PD-L1, CTLA-4, BTLA, and TIGIT. For this reason, LGG patients with high TIIClnc signature scores are expected to benefit more from immunotherapy.

The TIIClnc signature was directly established in immunotherapy datasets to verify this hypothesis. The predictive analysis demonstrated that the TIIClnc signature was a favorable marker of OS in urothelial carcinoma (IMvigor dataset), renal cell carcinoma (Braun dataset), and melanoma (Allen, GSE78220, and Nathanson datasets) patients. As expected, urothelial carcinoma, renal cell carcinoma, and melanoma patients with high TIIClnc signature scores were prone to benefit more from anti-PD-1, anti-PD-L1, or anti-CTLA-4 immunotherapy, respectively. Consistently, melanoma patients in the GSE35640 and GSE91061 datasets, colorectal adenocarcinoma and pancreatic adenocarcinoma patients in the GSE179351 dataset, and esophageal adenocarcinoma patients in the GSE165252 dataset with high TIIClnc signature scores also benefit more from immunotherapy. Besides, triple-negative breast cancer patients in the GSE103668 dataset with high TIIClnc signature scores benefit more from targeted therapy. More importantly, glioblastoma patients with high TIIClnc signature scores in the PRJNA482620 dataset were related to relatively reduced survival time, which the small sample size could partly explain the insignificant results. Furthermore, in our Xiangya in-house dataset, the protein expression level of CD8, PD-1, and PD-L1 also increased in the high TIIClnc signature score group, suggesting that the TIIClnc

signature could potentially predict the response rate of immunotherapy in LGG.

Although the clinical significance of the TIIClnc signature in LGG is remarkable, some limitations in this study need to be issued. First, all included datasets were from single-center retrospective studies, and future TIIClnc signature validation should be performed in prospective multicenter cohorts. Second, the in-depth molecular mechanisms of how 16 identified most valuable TIIClncRNAs could influence the TIME and immunotherapy response of LGG should be further explored. Third, more immunotherapy cohorts of glioma patients are urgently needed to validate the TIIClnc signature in predicting immunotherapy response.

In conclusion, by integrative analysis of sequencing data of purified immune cells, LGG cell lines, and bulk LGG tissues based on a wealth of machine learning algorithms, a stable and robust TIIClnc signature to stratify LGG patients and predict the outcomes of immunotherapy was developed. A TIIClnc signature is a promising tool for personalized treatment and clinical management for individual LGG patients.

## Abbreviations

lncRNAs: long noncoding RNAs; TIIClncRNA: tumor-infiltrating immune cell-associated lncRNA; LGG: low-grade glioma; WHO: World Health Organization; TIME: tumor immune microenvironment; ICB: immune checkpoint blockade; ACT: adoptive cell transfer; TCGA: The Cancer Genome Atlas; CGGA: Chinese Glioma Genome Atlas; GEO: Gene Expression Omnibus; CCLE: Cancer Cell Line Encyclopedia project; RMA: Robust Multi-array Average; FPKM: fragments per kilobase million; TPM: transcripts per kilobase million; TSI: tissue specificity index; RSF: random survival forest; Enet: elastic network; plsRcox: partial least squares regression for Cox; SuperPC: supervised principal components; GBM: generalized boosted regression modeling; survival-SVM: survival support vector machine; GEP: T cell-inflamed gene expression profile; CYT: cytotoxic activity; IFN- $\gamma$ : interferon  $\gamma$ ; TMB: tumor mutation burden; MSI: microsatellite instability; TCR: T cell receptor; IPS: immunophenoscore; GSVA: gene set variation analysis; GO: Gene Ontology; KEGG: Kyoto Encyclopedia of Genes and Genomes; GSEA: gene set enrichment analysis; OS: overall survival; ROC: receiver operating characteristic; BSA: bovine serum albumin; FBS: fetal bovine serum; TIMER: Tumor Immune Estimation Resource; MCPcounter: Microenvironment Cell Populations-counter; ssGSEA: single cell gene set enrichment analysis; ESTIMATE: Estimation of STromal and Immune cells in

MAalignant Tumours using Expression data.

## Supplementary Material

Supplementary figures.

<https://www.thno.org/v12p5931s1.pdf>

Supplementary table 1.

<https://www.thno.org/v12p5931s2.xlsx>

Supplementary table 2.

<https://www.thno.org/v12p5931s3.xlsx>

Supplementary table 3.

<https://www.thno.org/v12p5931s4.xlsx>

Supplementary table 4.

<https://www.thno.org/v12p5931s5.xlsx>

Supplementary table 5.

<https://www.thno.org/v12p5931s6.xlsx>

Supplementary table 6.

<https://www.thno.org/v12p5931s7.xlsx>

## Acknowledgements

Financial support was provided by the National Natural Science Foundation of China (NO. 82073893, 82172685, and 81873635). Hunan Provincial Natural Science Foundation of China (NO.2022JJ20095). Hunan Provincial Health Committee Foundation of China (NO.202204044869).

## Author contributions

NZ, HZ, QC, ZL, and WW designed and drafted the manuscript. NZ, HZ, QC, JZ, PL, ZL, SL, FL, WW, RZ, and ZD wrote figure legends and revised the manuscript. NZ, HZ, and QC conducted data analysis. All authors have read and approved the final manuscript.

## Availability of data and materials

All data used in this work can be acquired from the Gene Expression Omnibus (GEO; <https://www.ncbi.nlm.nih.gov/geo/>), the Cancer Genome Atlas (TCGA) datasets (<https://xenabrowser.net/>), the Chinese Glioma Genome Atlas (CGGA) datasets (<http://www.cgga.org.cn/>), and the Cancer Cell Line Encyclopedia (CCLE) project (<https://portals.broadinstitute.org/ccle>).

## Code availability

Essential scripts for implementing machine learning-based integrative procedure are available on the Github website (<https://github.com/AweKevin/TIIClnc>).

## Ethical Approval

Glioma tissues were collected and written informed consent was obtained from all patients. The included glioma tissues were approved by the Ethics Committee of Xiangya Hospital, Central South

University.

## Competing Interests

The authors have declared that no competing interest exists.

## References

1. Youssef G, Miller JJ. Lower Grade Gliomas. *Curr Neurol Neurosci Rep.* 2020; 20: 21.
2. Louis DN, Ohgaki H, Wiestler OD, Cavenee WK, Burger PC, Jouvet A, et al. The 2007 WHO classification of tumours of the central nervous system. *Acta Neuropathol.* 2007; 114: 97-109.
3. Pignatti F, van den Bent M, Curran D, Debruyne C, Sylvester R, Therasse P, et al. Prognostic factors for survival in adult patients with cerebral low-grade glioma. *J Clin Oncol.* 2002; 20: 2076-84.
4. Zhang M, Wang X, Chen X, Zhang Q, Hong J. Novel Immune-Related Gene Signature for Risk Stratification and Prognosis of Survival in Lower-Grade Glioma. *Front Genet.* 2020; 11: 363.
5. Gao Z, Xu J, Zhang Z, Fan Y, Xue H, Guo X, et al. A Comprehensive Analysis of METTL1 to Immunity and Stemness in Pan-Cancer. *Front Immunol.* 2022; 13: 795240.
6. Zhou M, Zhang Z, Bao S, Hou P, Yan C, Su J, et al. Computational recognition of lncRNA signature of tumor-infiltrating B lymphocytes with potential implications in prognosis and immunotherapy of bladder cancer. *Brief Bioinform.* 2021; 22.
7. Fatica A, Bozzoni I. Long non-coding RNAs: new players in cell differentiation and development. *Nat Rev Genet.* 2014; 15: 7-21.
8. Sun J, Zhang Z, Bao S, Yan C, Hou P, Wu N, et al. Identification of tumor immune infiltration-associated lncRNAs for improving prognosis and immunotherapy response of patients with non-small cell lung cancer. *J Immunother Cancer.* 2020; 8.
9. Ranzani V, Rossetti G, Panzeri I, Arrigoni A, Bonnal RJ, Curti S, et al. The long intergenic noncoding RNA landscape of human lymphocytes highlights the regulation of T cell differentiation by linc-MAF-4. *Nat Immunol.* 2015; 16: 318-25.
10. Xu M, Xu X, Pan B, Chen X, Lin K, Zeng K, et al. lncRNA SATB2-AS1 inhibits tumor metastasis and affects the tumor immune cell microenvironment in colorectal cancer by regulating SATB2. *Mol Cancer.* 2019; 18: 135.
11. Irizarry RA, Hobbs B, Collin F, Beazer-Barclay YD, Antonellis KJ, Scherf U, et al. Exploration, normalization, and summaries of high density oligonucleotide array probe level data. *Biostatistics.* 2003; 4: 249-64.
12. Yanai I, Benjamin H, Shmoish M, Chalifa-Caspi V, Shklar M, Ophir R, et al. Genome-wide midrange transcription profiles reveal expression level relationships in human tissue specification. *Bioinformatics.* 2005; 21: 650-9.
13. Hothorn T, Lausen B. On the exact distribution of maximally selected rank statistics. *Computational Statistics & Data Analysis.* 2003; 43: 121-37.
14. Thorsson V, Gibbs DL, Brown SD, Wolf D, Bortone DS, Ou Yang TH, et al. The Immune Landscape of Cancer. *Immunity.* 2018; 48: 812-30 e14.
15. Ayers M, Lunceford J, Nebozhyn M, Murphy E, Loboda A, Kaufman DR, et al. IFN-gamma-related mRNA profile predicts clinical response to PD-1 blockade. *J Clin Invest.* 2017; 127: 2930-40.
16. Roh W, Chen PL, Reuben A, Spencer CN, Prieto PA, Miller JP, et al. Integrated molecular analysis of tumor biopsies on sequential CTLA-4 and PD-1 blockade reveals markers of response and resistance. *Sci Transl Med.* 2017; 9.
17. Wang K, Li M, Hakonarson H. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Res.* 2010; 38: e164.
18. Charoentong P, Finotello F, Angelova M, Mayer C, Efremova M, Rieder D, et al. Pan-cancer Immunogenomic Analyses Reveal Genotype-Immuno-phenotype Relationships and Predictors of Response to Checkpoint Blockade. *Cell Rep.* 2017; 18: 248-62.
19. Li T, Fan J, Wang B, Traugh N, Chen Q, Liu JS, et al. TIMER: A Web Server for Comprehensive Analysis of Tumor-Infiltrating Immune Cells. *Cancer Res.* 2017; 77: e108-e10.
20. Becht E, Giraldo NA, Lacroix L, Buttard B, Elarouci N, Petitprez F, et al. Estimating the population abundance of tissue-infiltrating immune and stromal cell populations using gene expression. *Genome Biol.* 2016; 17: 218.
21. Yoshihara K, Shahmoradgolgi M, Martinez E, Vegesna R, Kim H, Torres-Garcia W, et al. Inferring tumour purity and stromal and immune cell admixture from expression data. *Nat Commun.* 2013; 4: 2612.
22. Xu L, Deng C, Pang B, Zhang X, Liu W, Liao G, et al. TIP: A Web Server for Resolving Tumor Immunophenotype Profiling. *Cancer Res.* 2018; 78: 6575-80.
23. Chen DS, Mellman I. Oncology meets immunology: the cancer-immunity cycle. *Immunity.* 2013; 39: 1-10.
24. Rosario SR, Long MD, Affronti HC, Rowsam AM, Eng KH, Smiraglia DJ. Pan-cancer analysis of transcriptional metabolic dysregulation using The Cancer Genome Atlas. *Nat Commun.* 2018; 9: 5330.
25. Kobayashi Y, Kushihara Y, Saito N, Yamaguchi S, Kakimi K. A novel scoring method based on RNA-Seq immunograms describing individual cancer-immunity interactions. *Cancer Sci.* 2020; 111: 4031-40.
26. Bagaev A, Kotlov N, Nomi K, Svekolkin V, Gafurov A, Isaeva O, et al. Conserved pan-cancer microenvironment subtypes predict response to immunotherapy. *Cancer Cell.* 2021; 39: 845-65 e7.
27. Ulloa-Montoya F, Louahed J, Dizier B, Gruselle O, Spiessens B, Lehmann FF, et al. Predictive gene signature in MAGE-A3 antigen-specific cancer immunotherapy. *J Clin Oncol.* 2013; 31: 2388-95.
28. Riaz N, Havel JJ, Makarov V, Desrichard A, Urba WJ, Sims JS, et al. Tumor and Microenvironment Evolution during Immunotherapy with Nivolumab. *Cell.* 2017; 171: 934-49 e16.
29. Hugo W, Zaretsky JM, Sun L, Song C, Moreno BH, Hu-Lieskovan S, et al. Genomic and Transcriptomic Features of Response to Anti-PD-1 Therapy in Metastatic Melanoma. *Cell.* 2017; 168: 542.
30. Van Allen EM, Miao D, Schilling B, Shukla SA, Blank C, Zimmer L, et al. Genomic correlates of response to CTLA-4 blockade in metastatic melanoma. *Science.* 2015; 350: 207-11.
31. Nathanson T, Ahuja A, Rubinsteyn A, Aksoy BA, Hellmann MD, Miao D, et al. Somatic Mutations and Neopeptide Homology in Melanomas Treated with CTLA-4 Blockade. *Cancer Immunol Res.* 2017; 5: 84-91.
32. Mariathasan S, Turley SJ, Nickles D, Castiglioni A, Yuen K, Wang Y, et al. TGFbeta attenuates tumour response to PD-L1 blockade by contributing to exclusion of T cells. *Nature.* 2018; 554: 544-8.
33. Braun DA, Hou Y, Bakouny Z, Ficial M, Sant' Angelo M, Forman J, et al. Interplay of somatic alterations and immune infiltration modulates response to PD-1 blockade in advanced clear cell renal cell carcinoma. *Nat Med.* 2020; 26: 909-18.
34. Parikh AR, Szabolcs A, Allen JN, Clark JW, Wo JY, Raabe M, et al. Radiation therapy enhances immunotherapy response in microsatellite stable colorectal and pancreatic adenocarcinoma in a phase II trial. *Nat Cancer.* 2021; 2: 1124-35.
35. van den Ende T, de Clercq NC, van Berge Henegouwen MI, Gisbertz SS, Geijsen ED, Verhoeven RHA, et al. Neoadjuvant Chemoradiotherapy Combined with Atezolizumab for Resectable Esophageal Adenocarcinoma: A Single-arm Phase II Feasibility Trial (PERFECT). *Clin Cancer Res.* 2021; 27: 3351-9.
36. Zhao J, Chen AX, Gartrell RD, Silverman AM, Aparicio L, Chu T, et al. Immune and genomic correlates of response to anti-PD-1 immunotherapy in glioblastoma. *Nat Med.* 2019; 25: 462-9.
37. Birkbak NJ, Li Y, Pathania S, Greene-Colozzi A, Dreze M, Bowman-Colin C, et al. Overexpression of BLM promotes DNA damage and increased sensitivity to platinum salts in triple-negative breast and serous ovarian cancers. *Ann Oncol.* 2018; 29: 903-9.
38. Hoshida Y, Brunet JP, Tamayo P, Golub TR, Mesirov JP. Subclass mapping: identifying common subtypes in independent disease data sets. *PLoS One.* 2007; 2: e1195.
39. Jiang P, Gu S, Pan D, Fu J, Sahu A, Hu X, et al. Signatures of T cell dysfunction and exclusion predict cancer immunotherapy response. *Nat Med.* 2018; 24: 1550-8.
40. Zhang N, Zhang H, Wang Z, Dai Z, Zhang X, Cheng Q, et al. Immune Infiltrating Cells-Derived Risk Signature Based on Large-scale Analysis Defines Immune Landscape and Predicts Immunotherapy Responses in Glioma Tumor Microenvironment. *Front Immunol.* 2021; 12: 691811.
41. Yang K, Wu Z, Zhang H, Zhang N, Wu W, Wang Z, et al. Glioma targeted therapy: insight into future of molecular approaches. *Mol Cancer.* 2022; 21: 39.
42. Nduom EK, Weller M, Heimberger AB. Immunosuppressive mechanisms in glioblastoma. *Neuro Oncol.* 2015; 17 Suppl 7: vii9-viii14.
43. Etcheverry A, Aubry M, de Tayrac M, Vauleon E, Boniface R, Guenet F, et al. DNA methylation in glioblastoma: impact on gene expression and clinical outcome. *BMC Genomics.* 2010; 11: 701.
44. Xu SF, Zheng Y, Zhang L, Wang P, Niu CM, Wu T, et al. Long Non-coding RNA LINC00628 Interacts Epigenetically with the LAMA3 Promoter and Contributes to Lung Adenocarcinoma. *Mol Ther Nucleic Acids.* 2019; 18: 166-82.
45. Wang Z, Wang P, Cao L, Li F, Duan S, Yuan G, et al. Long Intergenic Non-Coding RNA 01121 Promotes Breast Cancer Cell Proliferation, Migration, and Invasion via the miR-150-5p/HMGA2 Axis. *Cancer Manag Res.* 2019; 11: 10859-70.
46. Rong Z, Wang Z, Wang X, Qin C, Geng W. Molecular interplay between linc01134 and YY1 dictates hepatocellular carcinoma progression. *J Exp Clin Cancer Res.* 2020; 39: 61.
47. Tang R, Chen J, Tang M, Liao Z, Zhou L, Jiang J, et al. lncRNA SLCO4A1-AS1 predicts poor prognosis and promotes proliferation and metastasis via the EGFR/MAPK pathway in colorectal cancer. *Int J Zhe J Biol Sci.* 2019; 15: 2885-96.
48. Zhen-Hua W, Yi-Wei G, Li-Qin Z, Jie-Yun Z, Zhe G, Wei-Jian G. Silencing of lncRNA C1RL-AS1 Suppresses the Malignant Phenotype in Gastric Cancer Cells via the AKT/beta-Catenin/c-Myc Pathway. *Front Oncol.* 2020; 10: 1508.
49. Zhang H, Zhu C, He Z, Chen S, Li L, Sun C. lncRNA PSMB8-AS1 contributes to pancreatic cancer progression via modulating miR-382-3p/STAT1/PD-L1 axis. *J Exp Clin Cancer Res.* 2020; 39: 179.
50. Ren Y, Zhao C, He Y, Min X, Xu H, Hu X. RPARP-AS1/miR125a-5p Axis Promotes Cell Proliferation, Migration and Invasion in Colon Cancer. *Oncotargets Ther.* 2021; 14: 5035-43.
51. Lin X, Jiang T, Bai J, Li J, Wang T, Xiao J, et al. Characterization of Transcriptome Transition Associates Long Noncoding RNAs with Glioma Progression. *Mol Ther Nucleic Acids.* 2018; 13: 620-32.



52. Eckel-Passow JE, Lachance DH, Molinaro AM, Walsh KM, Decker PA, Sicotte H, et al. Glioma Groups Based on 1p/19q, IDH, and TERT Promoter Mutations in Tumors. *N Engl J Med*. 2015; 372: 2499-508.
53. Wick W, Weller M, van den Bent M, Sanson M, Weiler M, von Deimling A, et al. MGMT testing—the challenges for biomarker-based glioma treatment. *Nat Rev Neurol*. 2014; 10: 372-85.
54. Liang X, Wang Z, Dai Z, Zhang H, Cheng Q, Liu Z. Promoting Prognostic Model Application: A Review Based on Gliomas. *J Oncol*. 2021; 2021: 7840007.
55. Tomaszewski W, Sanchez-Perez L, Gajewski TF, Sampson JH. Brain Tumor Microenvironment and Host State: Implications for Immunotherapy. *Clin Cancer Res*. 2019; 25: 4202-10.
56. Xu J, Gao Z, Liu K, Fan Y, Zhang Z, Xue H, et al. The Non-N(6)-Methyladenosine Epitranscriptome Patterns and Characteristics of Tumor Microenvironment Infiltration and Mesenchymal Transition in Glioblastoma. *Front Immunol*. 2021; 12: 809808.
57. Borst J, Ahrends T, Babala N, Melief CJM, Kastenmuller W. CD4(+) T cell help in cancer immunology and immunotherapy. *Nat Rev Immunol*. 2018; 18: 635-47.
58. Nair S, Dhodapkar MV. Natural Killer T Cells in Cancer Immunotherapy. *Front Immunol*. 2017; 8: 1178.
59. Hollern DP, Xu N, Thennavan A, Glodowski C, Garcia-Recio S, Mott KR, et al. B Cells and T Follicular Helper Cells Mediate Response to Checkpoint Inhibitors in High Mutation Burden Mouse Models of Breast Cancer. *Cell*. 2019; 179: 1191-206 e21.
60. Greene S, Robbins Y, Mydlarz WK, Huynh AP, Schmitt NC, Friedman J, et al. Inhibition of MDSC Trafficking with SX-682, a CXCR1/2 Inhibitor, Enhances NK-Cell Immunotherapy in Head and Neck Cancer Models. *Clin Cancer Res*. 2020; 26: 1420-31.
61. Zhang H, Luo YB, Wu W, Zhang L, Wang Z, Dai Z, et al. The molecular feature of macrophages in tumor immune microenvironment of glioma patients. *Comput Struct Biotechnol J*. 2021; 19: 4603-18.
62. Tanaka A, Sakaguchi S. Regulatory T cells in cancer immunotherapy. *Cell Res*. 2017; 27: 109-18.
63. Ganesh K, Stadler ZK, Cercek A, Mendelsohn RB, Shia J, Segal NH, et al. Immunotherapy in colorectal cancer: rationale, challenges and potential. *Nat Rev Gastroenterol Hepatol*. 2019; 16: 361-75.
64. McGranahan N, Furness AJ, Rosenthal R, Ramskov S, Lyngaa R, Saini SK, et al. Clonal neoantigens elicit T cell immunoreactivity and sensitivity to immune checkpoint blockade. *Science*. 2016; 351: 1463-9.