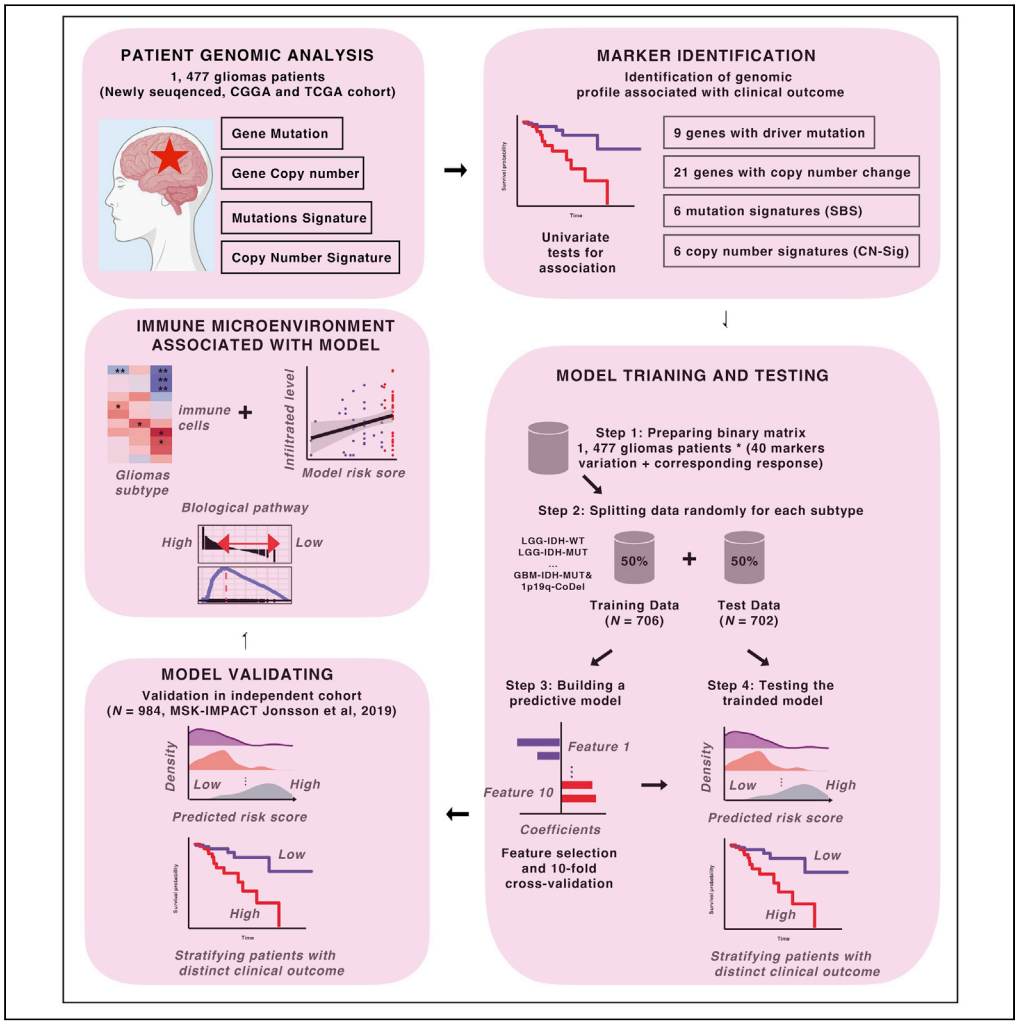
iScience

CellPress
OPEN ACCESS

Article

A prognostic risk model for glioma patients by systematic evaluation of genomic variations



Baifeng Zhang, Weiqing Wan, Zibo Li, ..., Liwei Zhang, Karin D. Rodland, Shirley X. Tsang

zhangbaifeng1@gmail.com (B.Z.)
zlwtt@aliyun.com (L.Z.)
sxtsang@biomatrix-intl.com (S.X.T.)

Highlights

Systematical evaluation of prognostic value of genomic features in over 1400 gliomas

SBS11 mutational signature is associated with glioma recurrence

SBS11 mutational signature predicts worse prognosis in low-grade gliomas

Prediction of high-risk gliomas with shortened survival

Zhang et al., iScience 25, 105681
December 22, 2022
https://doi.org/10.1016/j.isci.2022.105681

iScience

CellPress
OPEN ACCESS

Article

A prognostic risk model for glioma patients by systematic evaluation of genomic variations

Baifeng Zhang,[1,5,9,10,11,*] Weiqing Wan,[2,10] Zibo Li,[2,10] Zhixian Gao,[2] Nan Ji,[2] Jian Xie,[2] Junmei Wang,[3] Bin Wang,[4] Dora Lai-Wan Kwong,[1,5] Xinyuan Guan,[1,5] Shengjie Gao,[6] Yuanli Zhao,[2] Youyong Lu,[7] Liwei Zhang,[2,*] Karin D. Rodland,[8] and Shirley X. Tsang[9,*]

SUMMARY

The overall survival rate of gliomas has not significantly improved despite new effective treatments, mainly due to tumor heterogeneity and drug delivery. Here, we perform an integrated clinic-genomic analysis of 1, 477 glioma patients from a Chinese cohort and a TCGA cohort and propose a potential prognostic model for gliomas. We identify that SBS11 and SBS23 mutational signatures are associated with glioma recurrence and indicate worse prognosis only in low-grade type of gliomas and IDH-Mut subtype. We also identify 42 genomic features associated with distinct clinical outcome and successfully used ten of these to develop a prognostic risk model of gliomas. The high-risk glioma patients with shortened survival were characterized by high level of frequent copy number alterations including *PTEN*, *CDKN2A/B* deletion, *EGFR* amplification, less *IDH1* or *CIC* gene mutations, high infiltration levels of immunosuppressive cells and activation of G2M checkpoint and Oxidative phosphorylation oncogenic pathway.

INTRODUCTION

Gliomas, the most common type of cancer that starts in the glial cells of the brain, are clinically derived from various neural cells including astrocytes, oligodendrocytes, and ependymal cells. Glioblastoma (GBM), as the most aggressive and commonly occurring type of glioma, has an average length of survival following diagnosis of only 12 to 15 months and less than 3–7% of patients survive longer than five years.[1] The causes of most cases of glioblastoma remains unclear, and majority of glioblastoma diagnoses are *de novo* whereas others start as the low-grade type of gliomas (LGG) and progress into glioblastoma. To date, The Cancer Genome Atlas (TCGA) and other studies have performed large scale next-generation sequencing of the genome of gliomas patients and revealed the mutation landscape and intratumor heterogeneity of gliomas patients involved in tumorigenesis.[2–6] Several key genomic features such as the mutation of gene *IDH* and the deletion of chromosome arms 1p and 19q were identified as new biomarkers to stratify subgroups with distinct clinical outcomes and clinical treatment plans, this further reshaped and update the World Health Organization (WHO) classification of glioma.[5,7,8] However, these studies primarily focused on the somatic events among subtypes.

Numerous studies demonstrated mutation signature and copy number alterations signature documented the characteristics occurring throughout the whole life of cancer cells including DNA repair or exogenous processes such as chemotherapy treatment.[9,10] It was expected that the genomic signatures would have a great influence on the clinical outcome and treatment response of glioma patients. One recent study analyzing the mutational spectral following radiotherapy in glioma patients revealed that a radiotherapy-derived deletion signature was associated with worse clinical outcomes and may be used to predict sensitivity to radiation therapy.[11] Therefore, there is a need to identify and incorporate prognostic genomic signatures as additional molecular features that may enhance the treatments performance of gliomas.

Here, we sequence the whole exomes of a cohort of Chinese glioma patients and additionally also obtain the published large-scale genomic data from Chinese Glioma Genome Atlas (CGGA) and The Cancer Genome Atlas (TCGA).[2–4,12,13] We identify genomic variations, extract mutational and copy number

[1]Departments of Clinical Oncology, The University of Hong Kong-Shenzhen Hospital, Shenzhen 518053, China

[2]Department of Neurosurgery, Beijing Tiantan Hospital, Capital Medical University, No. 119 South 4th Ring West Road, Fengtai District, Beijing 100070, China

[3]Department of Neuropathology, Beijing Neurosurgical Institute, Beijing 100070, People's Republic of China

[4]Department of Neurosurgery, Peking University International Hospital, Peking University Eighth Clinical Medical School, Beijing 100084, China

[5]Departments of Clinical Oncology, Li Ka Shing Faculty of Medicine, The University of Hong Kong, Hong Kong, China

[6]BGI-Shenzhen, Shenzhen 518083, China

[7]Chinese Cancer Genome Consortium, Shenzhen 518083, China

[8]Biological Sciences Division, Pacific Northwest National Laboratory, Richland, WA 99354, USA

[9]BioMatrix, Rockville, MD, USA

[10]These authors contributed equally

[11]Lead contact

*Correspondence: zhangbaifeng1@gmail.com (B.Z.), zlwtt@aliyun.com (L.Z.), sxtsang@biomatrix-intl.com (S.X.T.)

https://doi.org/10.1016/j.isci.2022.105681

alteration signatures, and evaluate their clinical relevance in a total of 1, 477 patients. Our results generate a full picture of genomic variation signature and discover several additional genomic features including focal amplification or deletion and genomic signatures as potential prognosis markers. We further develop a prognostic risk model of glioma based on genomic features to stratify glioma patients into high-risk and low-risk two subgroups with significant distinct outcomes in Chinese and TCGA cohort ($p< 0.0001$). The performance of this model in the clinic was further validated in an independent cohort with 1,004 glioma patients from Memorial Sloan Kettering Cancer Center (MSKCC) via MSK-IMPACT target panel sequencing ($p< 0.05$).[14] Associated expression and immune microenvironment features of two distinct prognostic risk subgroups also were characterized.

RESULTS

Prognostic genomic mutation signatures of glioma patients

The 1,477 glioma patients in total, which comprise the 1,111 published TCGA whole exomes, 286 published Chinese exomes and 80 newly sequenced Chinese exomes in this study-yielded 269, 341 somatic SNVs (184, 242 non-silent mutations) and 200, 611 SCNA segments that were analyzed for downstream analysis (Tables S1, S2, S3, and S4).

The mutational signatures can reflect abnormal exposures or neoplastic progression. The Single Base Substitution (SBS) mutational signature analysis of gliomas was performed by stratifying the total somatic SNVs according to their trinucleotide mutational contexts and then applying multiple non-negative matrix factorization (NMF) iterations to decompose into optimal signatures with predefined etiologies in the Catalogue of Somatic Mutations in Cancer (COSMIC) database.[9] Of the 14 independent mutational signatures we identified, eleven COSMIC signatures and three newly identified signatures (SBS96F, SBS96G and SBS96H) showed substantial variation in the glioma patients and in the mutations attributed per patient (Figure S1A). Six mutational signatures (including SBS11, SBS15, SBS23, SBS44, SBS96F and SBS96G) were observed to be relevant to the prognosis of gliomas (Figure S1B). Of these signatures, SBS15, SBS44 and SBS96F indicated better clinical outcome. SBS15 and SBS44 were reported to be associated with defective DNA mismatch repair (MMR) and microsatellite instability (MSI) previously,[9] and SBS96F is the newly identified by T>G mutations at ATG and CTG trinucleotides (the mutated base is underlined) (Figure S1C). SBS11, SBS23 and SBS96G indicated worse clinical outcome. SBS96G was also newly identified and characterized predominantly by C>T mutations at ACG and CCA trinucleotides (Figure S1C). SBS11 was associated with prior treatment with the alkylating agent such as temozolomide in the previous study[9] and showed similar C>T mutational contexts to SBS23 (Figure 1A). Of interest, the substantial increase of SBS11 and SBS23 in absolute mutationnumber and composition fraction level was observed in Chinese cohort, GBM subtype, Recurrent and IDH-mutant gliomas patients (Figures 1A and 1C), indicated that the SBS11 and SBS23 could inform clinical treatment of gliomas patients. The shortened survival of gliomas patients with SBS11 and SBS23 mutations was observed in both Chinese ($p = 0.025$) and TCGA cohort ($p = 0.0002$), and in LGG subtype ($p = 0.00036$) but not in GBM subtype ($p = 0.57$), and only in IDH-mutant subtype ($p< 0.0001$) (Figures 1D and 1E), suggesting that SBS11 and SBS23 might be a prognostic signal in the early stage or IDH-mutated of glioma patients.

To identify copy number signatures, we analyzed the total SCNA segments from glioma patients and computed the genome-wide distributions of eight different features followed by previous studies[10,15]: Breakpoint count per 10 Mb (BP10MB), Breakpoint count per chromosome arm (BPArm), The observed absolute copy number (CN) of each segment. The absolute difference in copy number between adjacent segments across the genome (CNCP), Length of segments with oscillating copy number (OsCN), The log10 length of each segment (SS), The minimal number of chromosomes with 50% copy number variation (NC50), The burden of per chromosome (BoChr). Six copy number signatures as well as their important components were identified, namely CN-Sig1 to CN-Sig6 (Figure 2A). CN-Sig1 showed frequent breakpoints per 10 Mb, high copy number change point, high absolute copy number and high burden of copy number alterations in chromosome 7 and 12. This signature is caused by focal amplification of DNA segments located in chromosome 7 and 12. CN-Sig2 is characterized by one copy number change point and considerable oscillating copy number, suggesting a state of chromothripsis. CN-Sig3 is featured by zero absolute copy number and two copy number change point, indicating homozygous deletion. CN-Sig4 is represented by almost no breakpoints and zero copy change point, suggesting copy number neutral like. CN-Sig5 is characterized by few breakpoints and much more chromosomes with 50% CNA, reflecting the occurrence of whole genome duplication. CN-Sig6 is featured by one absolute copy number, one copy

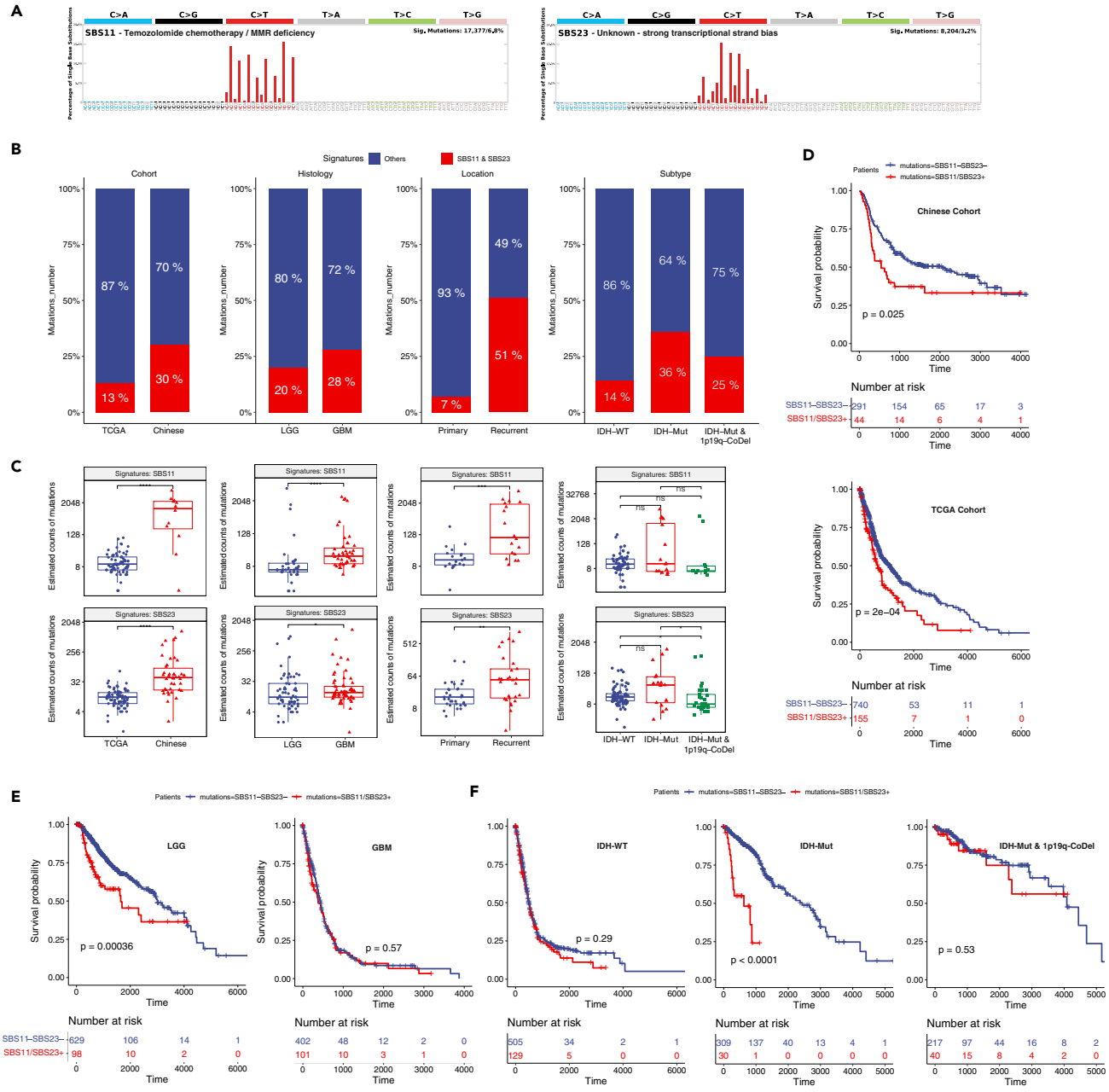


Figure 1. Prognostic implications of SBS11 and SBS23 mutational signature in gliomas

(A) Component weights of SBS11 and SBS23.

(B) Fraction difference of patients with SBS11 and SBS23 signature mutations among clinical phenotypes.

(C) Comparison of estimated number of SBS11 and SBS23 signature among clinical phenotypes. Wilcoxon signed-rank test: $p > 0.05$, $0.01 < P \leq 0.05$, $0.001 < P <= 0.01$, $0.0001 < P <= 0.001$ and $0 < P <= 0.0001$ were shown as ns, *, **, ***, ****. Data are represented as mean ± SEM.

(D) Kaplan-Meier survival analysis of glioma patients with SBS11 and SBS mutation or not in Chinese and TCGA cohort.

(E) Kaplan-Meier survival analysis of SBS11 and SBS mutation in LGG and GBM subtypes.

(F) Kaplan-Meier survival analysis of SBS11 and SBS mutation in IDH-wildtype (IDH-WT), IDH-mutant (IDH-Mut), and 1p/19q-codeleted IDH-mutant (IDH-Mut&1p19q-Co-Del) subtypes.

See also Figure S1.

number change point and one oscillating copy number, indicating loss of heterozygosity. Of these copy number signatures, only copy number neutral like CN-Sig4 indicated prolong survival ($p = 0.0011$) and other five copy number signatures indicated shorten survival ($p < 0.0001$) (Figure 2B), demonstrating that

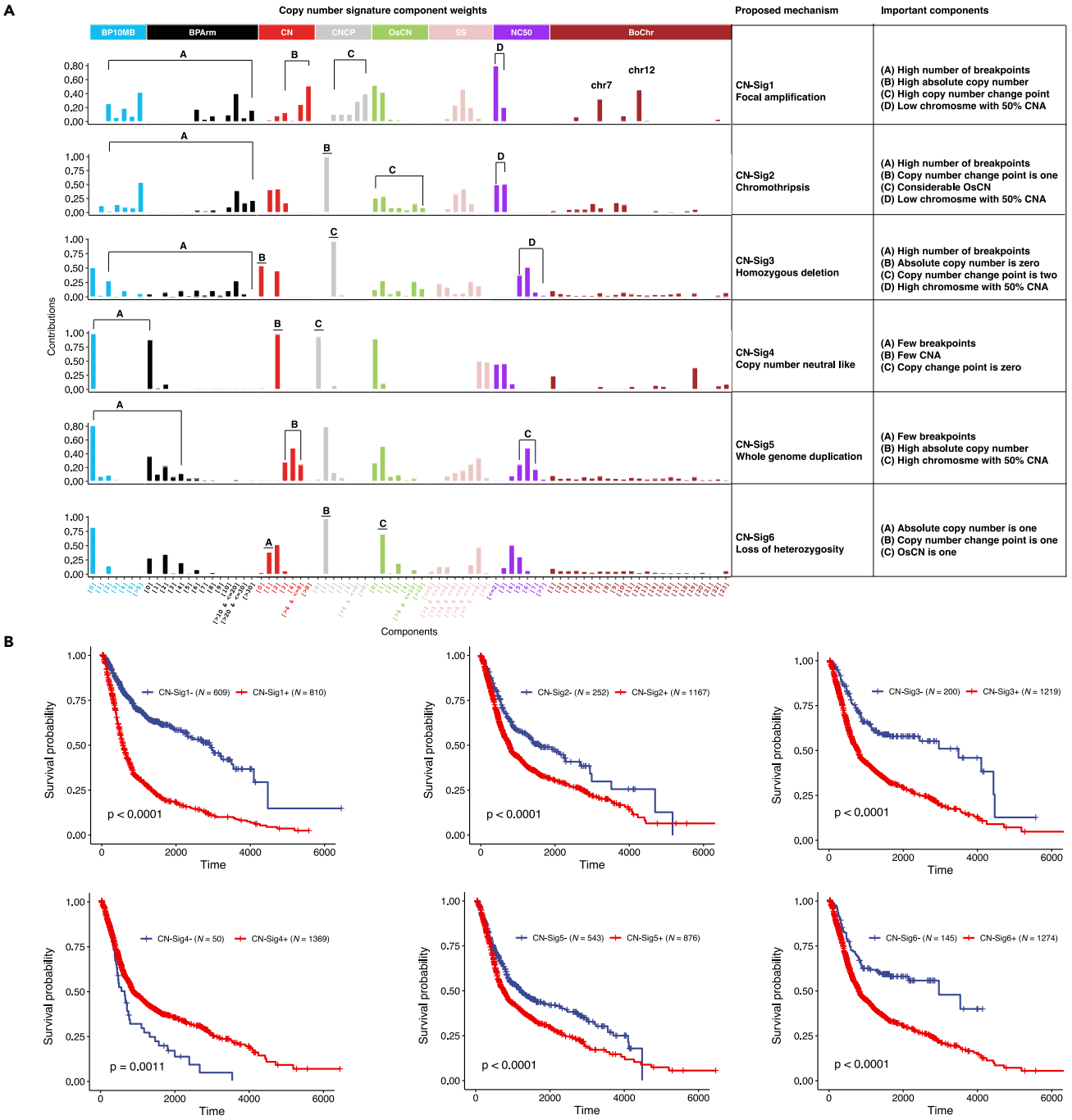


Figure 2. Copy number signatures of gliomas and their prognostic value

(A) Description of the defining component weights and proposed mechanisms for the six copy number signatures.

(B) Kaplan-Meier survival analysis of glioma patients with signature mutation or not for each copy number signatures.

the heterogeneity of copy number signatures in glioma patients and their promising role as potential prognostic signal.

Prognostic implications of significantly altered genes

Genes significantly altered by point mutations or copy number changes play an important role in regulating cellular growth and survival are likely to provide a selective growth advantage. Thirty significantly altered

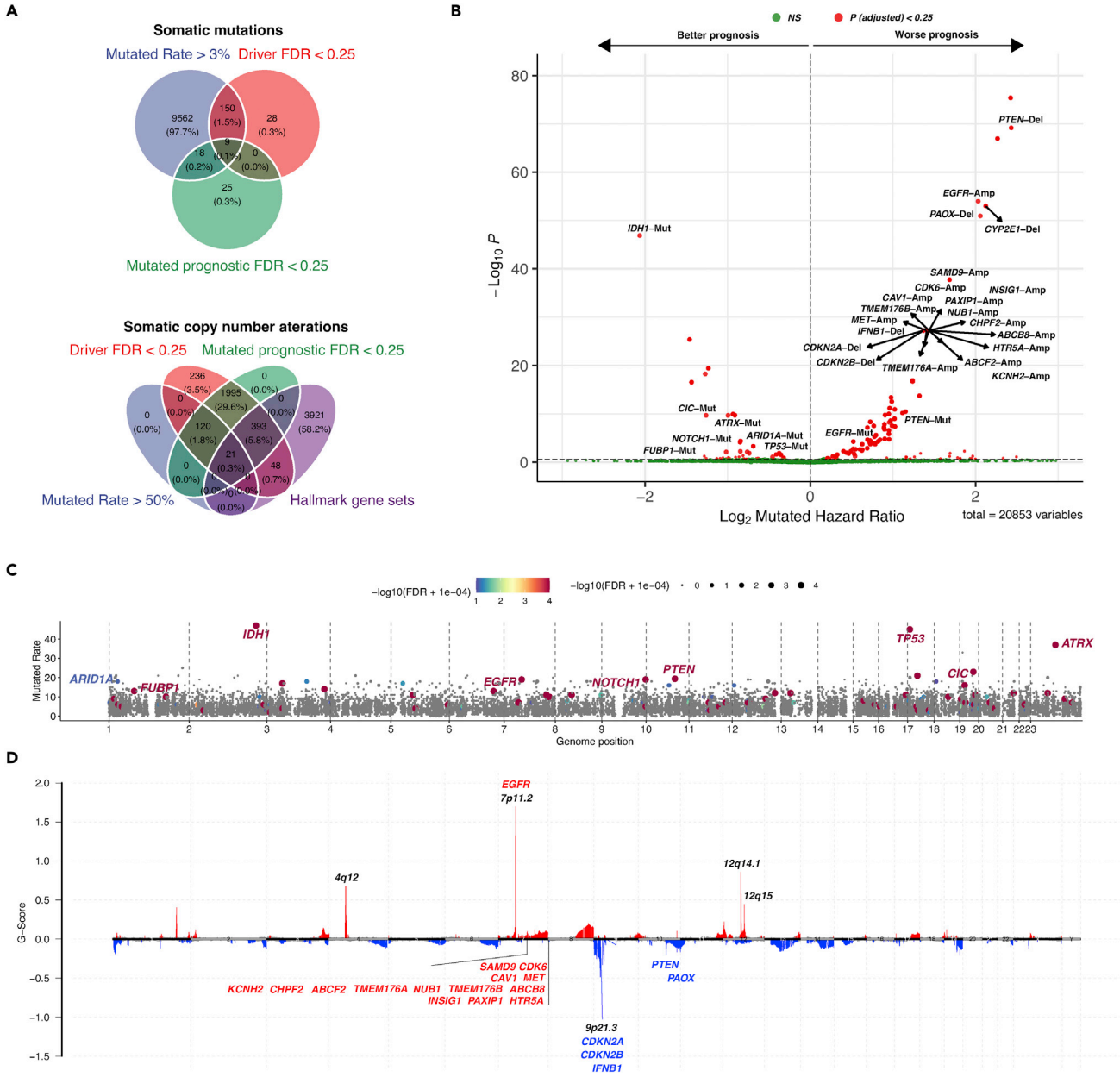


Figure 3. The identification of significantly altered gene and their prognostic implications

(A) The identification of 9 prognostic genes with driver mutations (driver FDR <0.25 reported by MutPanning software, above 3% mutated rate, prognostic FDR <0.25 reported by logrank test) and 21 prognostic genes with copy number changes (driver FDR <0.25 reported by GISTICv2 software, above 50% mutated rate, prognostic FDR <0.25 reported by logrank test).

(B) The distribution of *p* value (green: not significant/NS, red: *P* (adjusted/FDR) < 0.25) and Hazard Ratio of univariate survival analysis for genes with mutation or copy number changes.

(C) The visualization of genomic locations for significantly altered genes by mutation.

(D) The visualization of genomic locations for significantly altered genes by copy number changes (red: amplification, blue: deletion).

genes with prognostic value were identified in the data, including 9 genes with significant driver mutations occurred in over 3% of glioma patients (FDR <0.25, reported by MutPanning software[16]), and 21 genes located in frequently deleted or amplified region in over 50% of glioma patients (FDR <0.25, reported by GISTICv2 software[17]) (Figure 3A). Univariate survival analysis showed that *ARID1A*, *FUBP1*, *IDH1*, *NOTCH1*, *TP53*, *CIC* and *ATRX* gene mutations indicated better prognosis (Figures 3B and 3C) while *EGFR* mutations, *PTEN* mutations, amplification of chromosome 7p11.2 (*EGFR*) and 7q (*SAMD9*, *CDK6*,

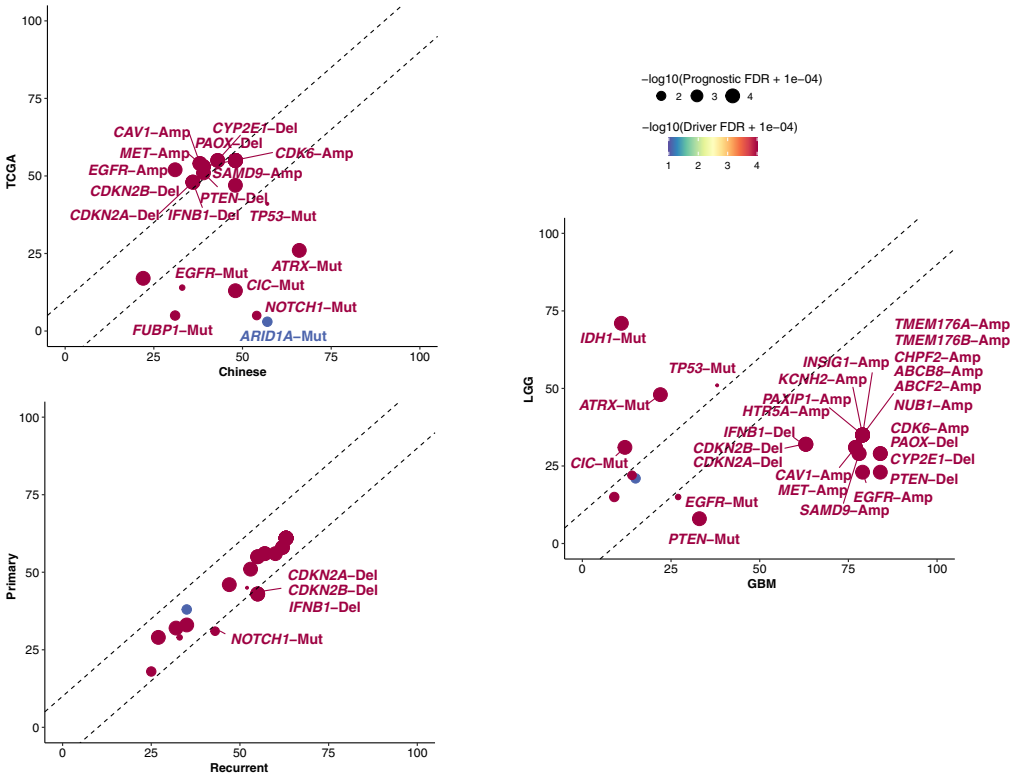


Figure 4. Comparison of occurring frequency among clinical subtypes (Highlighted with over 10% rate difference)

CAV1, *MET*, *KCNH2*, *CHPF2*, *ABCF2*, *TMEM176A*, *NUB1*, *TMEM176B*, *ABCB8*, *INSIG1*, *PAXIP1* and *HTR5A*), deletion of chromosome 9p21.3 (*CDKN2A*, *CDKN2B* and *IFNB1*) and deletion of *PTEN*, *PAOX* gene indicated worse prognosis (Figures 3B–3D).

Clinical disparities of significantly altered genes

To explore whether there are any clinical disparities of thirty prognostic significantly altered genes, we compared their fraction among multiple clinical subgroups including cohorts (TCGA versus Chinese), glioma subtypes (LGG versus GBM), tumor origin (Primary versus Recurrent) (Figure 4). TCGA cohorts showed higher copy number changed rate in genes whereas Chinese cohorts displayed higher mutation rate of genes, such as *NOTCH1* and *ARID1A*. Also, several gene mutations (*IDH1*, *TP53*, *ATRX* and *CIC*) with better prognosis were observed to be highly mutated in LGG subtypes and radiotherapy treated patients whereas almost all worse prognostic predictors showed higher rate in GBM subtypes.

Development of a prognostic risk model by systematic evaluation of prognostic predictor

To evaluate our prognostic predictors upon our data analysis using various genomic datasets, we conclude six mutational signatures, six copy number signatures and 30 significantly altered genes, we sought to determine if these prognostic predictors could be combined to develop a prognostic risk model of glioma patients. After the training and evaluation of 10-fold cross-validation, the final prognostic risk model consisted of ten markers (i.e., Negative markers (coefficient <0): *IDH1*-Mut, *CIC*-Mut, *TP53*-Mut and Positive markers (coefficient <0): *CDKN2B*-Del, CN-Sig1, CN-Sig3, CN-Sig6, *CDKN2A*-Del, *PTEN*-Del and *EGFR*-Amp ranked by their coefficients) (Figure 5A). By using these 10 biomarkers to concluded this "final model", the major subtypes patients in Chinese and TCGA cohort were separately stratified into high risk and low risk subgroups with predicted risk, which showed significant distinct overall survival rate (LGG-IDH-Mut&1p19q-Codeletion: $p = 0.0012$, LGG-IDH-Mut: $p = 0.00065$, GBM-IDH-Mut: $p = 0.031$, LGG-IDH-WT: $p < 0.0001$, GBM-IDH-WT: $p < 0.036$) (Figures 5B–5D). The genomic landscape of gliomas also demonstrated that negative markers displayed mutual exclusivity with positive markers (Figure 5B).

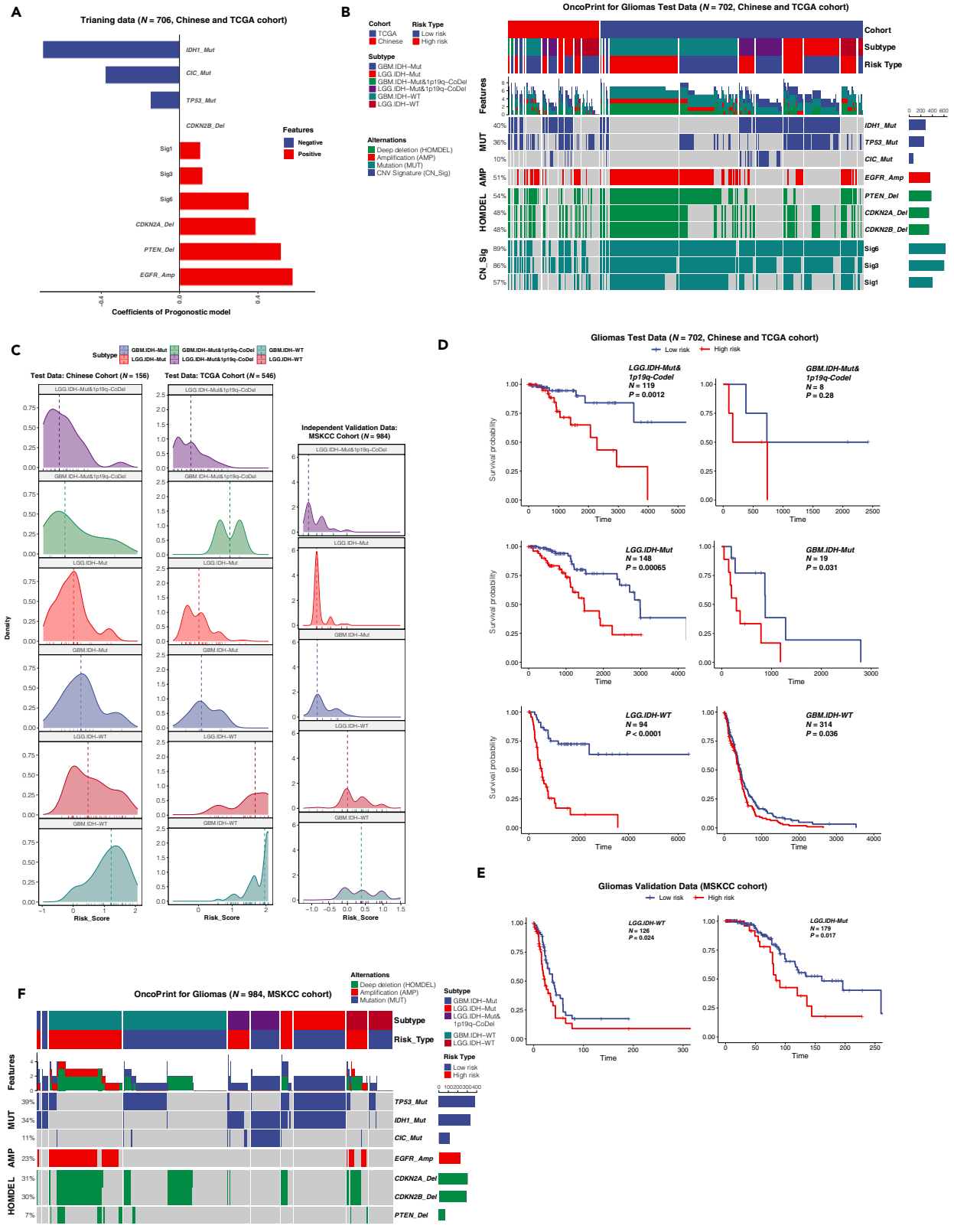

A
Trianing data (N = 706, Chinese and TCGA cohort)
IDH1_Mut
CIC_Mut
TP53_Mut
CDKN2B_Del
Sig1
Sig3
Sig6
CDKN2A_Del
PTEN_Del
EGFR_Amp
Features
Negative
Positive
Coefficients of Progonostic model
B
OncoPrint for Gliomas Test Data (N = 702, Chinese and TCGA cohort)
Cohort
Risk Type
Low risk
High risk
Subtype
GBM.IDH–Mut
LGG.IDH–Mut
GBM.IDH–Mut&1p19q–CoDel
LGG.IDH–Mut&1p19q–CoDel
GBM.IDH–WT
LGG.IDH–WT
Alternations
Deep deletion (HOMDEL)
Amplification (AMP)
Mutation (MUT)
CNV Signature (CN_Sig)
Features
MUT
AMP
HOMDEL
CN_Sig
C
Test Data: Chinese Cohort (N = 156)
Test Data: TCGA Cohort (N = 546)
Independent Validation Data: MSKCC Cohort (N = 984)
Density
Risk_Score
D
Gliomas Test Data (N = 702, Chinese and TCGA cohort)
Survival probability
Time
LGG.IDH-Mut& 1p19q-Codel N = 119 P = 0.0012
GBM.IDH-Mut& 1p19q-Codel N = 8 P = 0.28
LGG.IDH-Mut N = 148 P = 0.00065
GBM.IDH-Mut N = 19 P = 0.031
LGG.IDH-WT N = 94 P < 0.0001
GBM.IDH-WT N = 314 P = 0.036
E
Gliomas Validation Data (MSKCC cohort)
LGG.IDH-WT N = 126 P = 0.024
LGG.IDH-Mut N = 179 P = 0.017
F
OncoPrint for Gliomas (N = 984, MSKCC cohort)
Subtype
Risk_Type
TP53_Mut
IDH1_Mut
CIC_Mut
EGFR_Amp
CDKN2A_Del
CDKN2B_Del
PTEN_Del

Figure 5. The development and validation of proposed prognostic risk model

(A) Coefficients of final prognostic model based on training data from Chinese and TCGA cohort.

(B) Genomic landscape of prognostic features for high-risk and low-risk gliomas (GBM/LGG) in test data from Chinese and TCGA cohort.

(C) Distribution of predicted risk score and stratification of risk subgroups (Median: vertical dashed line, High: above median, Low: equal to or below median) in Chinese, TCGA and MSKCC-IMPACT gliomas cohort, respectively.

(D) Survival analysis of risk subgroup patients in Chinese and TCGA gliomas cohort, respectively.

(E) Survival analysis of risk subgroup patients in MSKCC-IMPACT gliomas cohort.

(F) Genomic landscape of available prognostic features in MSKCC-IMPACT gliomas cohort.

To validate model performance on an independent clinical dataset, we applied our model on MSK-IMPACT glioma dataset with about 1, 000 patients and created Kaplan-Meier curves for high/low risk subgroups. Because of the limited region of target panel sequencing in MSK-IMPACT cohort, only 7 of 10 markers including *TP53*-Mut, *IDH1*-Mut, *CIC*-Mut, *EGFR*-Amp, *CDKN2B*-Del, *CDKN2A*-Del, *PTEN*-Del was available and used. The high-risk subgroup was featured by risk positive markers (including *EGFR* amplification, *CDKN2A/B* deletion and *PTEN* deletion) and showed significantly shorten survival (LGG-IDH-Mut: $p = 0.017$, LGG-IDH-WT: $p = 0.024$, Figures 5E and 5F), suggesting that these risk positive markers might be useful as a potential therapeutic target for these patients with high risk.

Expression and immune features of subtypes of prognostic risk model

To explore whether there is any difference of tumor microenvironment among subtypes stratified by our model, we evaluated the correlation of our model risk scores with the infiltrated level of different immune component in TCGA gliomas cohort. We found that the infiltration levels of Neutrophils in GBM-IDH-WT, Plasma cells in LGG-IDH-Mut, M2 macrophages in LGG-ID-Mut&1p19q-Co-Del, $CD4^+$ memory resting T cells in LGG-IDH-WT positively correlated with our model prognostic risk scores (Figures 6A and 6B). In contrast, the infiltration levels of monocytes, eosinophils and activated Mast cells negatively correlated with risk score (Figures 6A and 6B). These observations indicated that the prognostic genomic features in our model also have a substantial association with tumor microenvironment.

Top 4 of 50 hallmark gene sets were also identified in the differential gene expression analysis of high-risk to low-risk patients in TCGA glioma patients and showed different enrichment between gliomas subtypes (Figure 5C): Oncogenic pathways such as G2M checkpoint (NES = 2.45, $P_{adjusted} = 0$) in high risk LGG-IDH-Mut, Interferon gamma response (NES = 2.32, $P_{adjusted} = 0$) in high risk LGG-IDH-Mut&1p19q-Co-Del and E2F targets (NES = 2.43, $P_{adjusted} = 0.001$) in high risk LGG-IDH-WT were significantly up-regulated whereas Oxidative phosphorylation (NES = 2.77, $P_{adjusted} = 0$) were significantly activated in high risk GBM-IDH-WT patients (Figure 6D), indicated diverse tumor supporting mechanisms between these glioma subtypes.

DISCUSSION

So far there is no comprehensive clinical prognostic or predictive classification for glioma and the survival rate of gliomas has not been significantly improved for decades. As a group of diseases with considerable molecular and tumor heterogeneity, gliomas still lack effective treatments and prognostic indicators that combine information on histology, tumor markers relative to outcome. In this study, we newly sequenced and established the largest genomic and clinical-pathological dataset of gliomas in China. We performed an integrated clinical-genomic analysis of gliomas in Chinese cohort and TCGA cohort to identify associations between somatic genomic alterations and key clinical phenotypes. In total, 42 markers were identified as prognostic indicators of glioma, including six mutational signatures, six copy number signatures, nine genes with driver mutations and twenty-one genes altered by deletions or amplifications in focal level.

Mutational signatures generally recorded cell-intrinsic process and previous exposures to exogenous factor. Single Base Substitution (SBS) Signatures have provided us deep understanding of the mutation process that molded a cancer genome. Notably, it was shown that SBS11, SBS23 or newly identified SBS96G signature mainly consisted of C>T mutations and predict worse prognosis only in LGG subtype patients, suggesting their potential as prognostic indicators in early stage of gliomas and a prevention strategy in avoiding exposures that cause C>T transversions. Several DNA mismatch-repair deficiency associated signatures, including SBS15 and SBS44, was also observed to prolong survival in our study, indicating that SBS15 and SBS44 may represent a pretreatment biomarker of susceptibility to PD1/PDL1 blockade immunotherapy in glioma patients.[18,19]

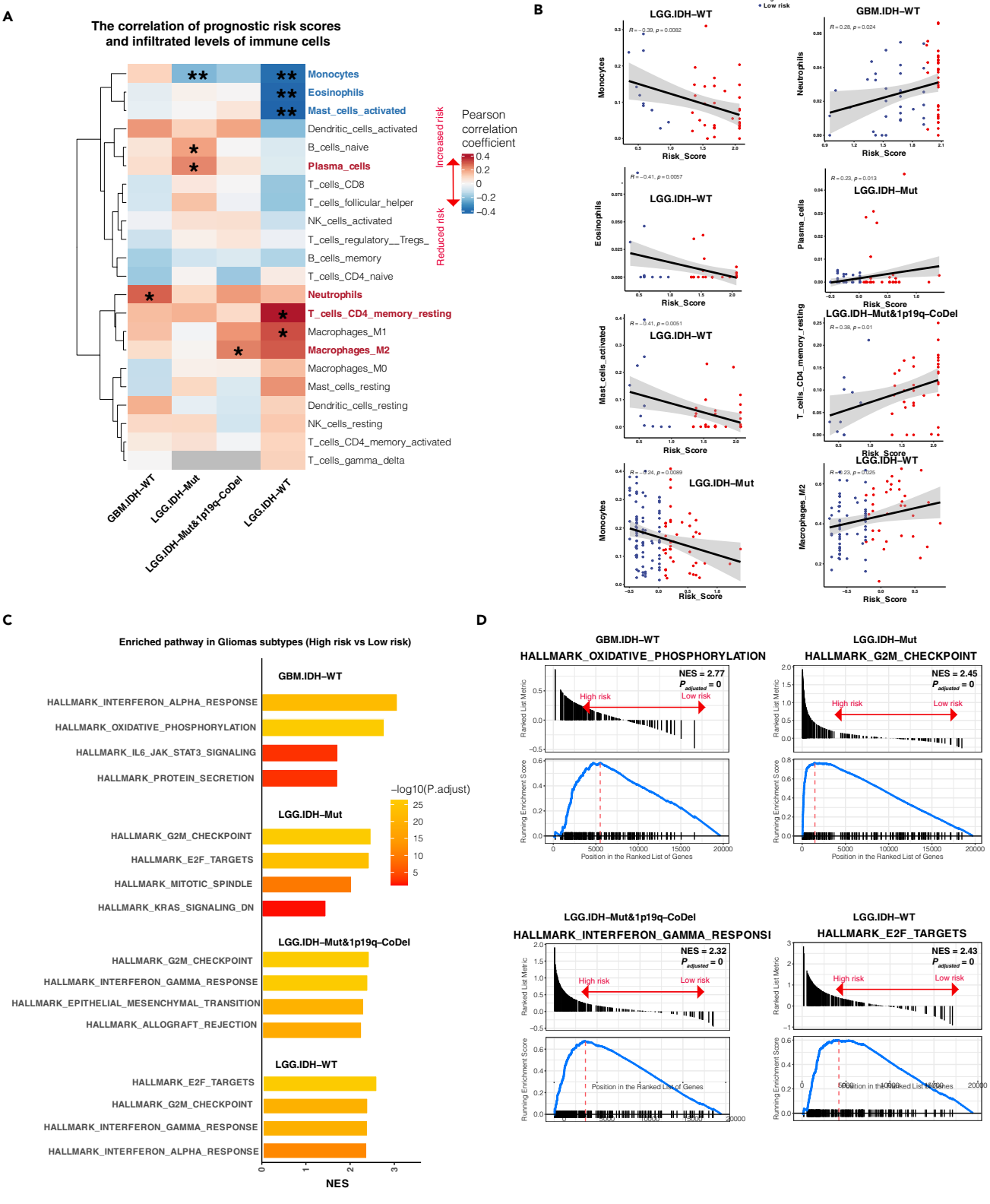

A
The correlation of prognostic risk scores and infiltrated levels of immune cells
Monocytes
Eosinophils
Mast_cells_activated
Dendritic_cells_activated
B_cells_naive
Plasma_cells
T_cells_CD8
T_cells_follicular_helper
NK_cells_activated
T_cells_regulatory__Tregs_
B_cells_memory
T_cells_CD4_naive
Neutrophils
T_cells_CD4_memory_resting
Macrophages_M1
Macrophages_M2
Macrophages_M0
Mast_cells_resting
Dendritic_cells_resting
NK_cells_resting
T_cells_CD4_memory_activated
T_cells_gamma_delta
GBM.IDH–WT
LGG.IDH–Mut
LGG.IDH–Mut&1p19q–CoDel
LGG.IDH–WT
Increased risk
Reduced risk
Pearson correlation coefficient
0.4
0.2
0
–0.2
–0.4
B
High risk
Low risk
LGG.IDH–WT
GBM.IDH–WT
LGG.IDH–Mut
LGG.IDH–Mut&1p19q–CoDel
Monocytes
Neutrophils
Eosinophils
Plasma_cells
Mast_cells_activated
T_cells_CD4_memory_resting
Macrophages_M2
Risk_Score
C
Enriched pathway in Gliomas subtypes (High risk vs Low risk)
GBM.IDH–WT
HALLMARK_INTERFERON_ALPHA_RESPONSE
HALLMARK_OXIDATIVE_PHOSPHORYLATION
HALLMARK_IL6_JAK_STAT3_SIGNALING
HALLMARK_PROTEIN_SECRETION
LGG.IDH–Mut
HALLMARK_G2M_CHECKPOINT
HALLMARK_E2F_TARGETS
HALLMARK_MITOTIC_SPINDLE
HALLMARK_KRAS_SIGNALING_DN
LGG.IDH–Mut&1p19q–CoDel
HALLMARK_G2M_CHECKPOINT
HALLMARK_INTERFERON_GAMMA_RESPONSE
HALLMARK_EPITHELIAL_MESENCHYMAL_TRANSITION
HALLMARK_ALLOGRAFT_REJECTION
LGG.IDH–WT
HALLMARK_E2F_TARGETS
HALLMARK_G2M_CHECKPOINT
HALLMARK_INTERFERON_GAMMA_RESPONSE
HALLMARK_INTERFERON_ALPHA_RESPONSE
–log10(P.adjust)
25
20
15
10
5
0
1
2
3
NES
D
GBM.IDH–WT
HALLMARK_OXIDATIVE_PHOSPHORYLATION
NES = 2.77
P adjusted = 0
LGG.IDH–Mut
HALLMARK_G2M_CHECKPOINT
NES = 2.45
P adjusted = 0
LGG.IDH–Mut&1p19q–CoDel
HALLMARK_INTERFERON_GAMMA_RESPONSE
NES = 2.32
P adjusted = 0
LGG.IDH–WT
HALLMARK_E2F_TARGETS
NES = 2.43
P adjusted = 0
High risk
Low risk
Ranked List Metric
Running Enrichment Score
Position in the Ranked List of Genes

Figure 6. Differential expression and immune compartments of prognostic risk subgroup patients in TCGA gliomas cohort

(A) The correlation heatmap of prognostic risk scores and infiltrated levels of immune cells within four major glioma subtypes (sample size >10). 0.01 $< P \leq$ 0.05, 0.001 $< P <$ =0.01 were shown as *, **.

(B) Significant association of infiltrated levels of immune cell types with predicted risk score in glioma patients.

(C) Significant differentially expressed gene sets in risk subgroup patients for four major glioma subtypes.

(D) The distribution of enrichment score for enriched gene sets.

Previous studies had shown that copy number signature exposures could predict both overall survival and the probability of drug-resistant relapse in ovarian carcinoma and prostate cancer.[10,15] In our analysis, the copy number signatures of glioma patients could be summarized into focal amplification, chromothripsis, homozygous deletion, copy number neutral like, whole genome duplication and loss of heterozygosity signatures. These signatures could represent the genomic complexity of profound copy number alterations profile in gliomas and could be used to stratify glioma patients into subgroups with significant distinct clinical outcome. These results demonstrated that the measurement of copy number signature provides a new efficient framework to predict patient response in clinical treatments of gliomas.

Of interest, for 42 markers as prognostic predictors, only ten of which was retained in our final prognostic risk model after the optimization of Feature Selection based on its contribution importance to model. Of these markers in final model, gene mutation markers (*IDH1*, *CIC*, *TP53*) decreased the risk of gliomas and were enriched in low-risk subgroup and LGG subtypes. In contrast, copy number changed markers increased the risk, and frequently occurred in high-risk subgroup and GBM subtypes. These data demonstrated that LGG patients showed distinct genomic variation features with GBM patients, suggesting that risk subgroups predicted by our prognostic risk model matched well with pathology subtypes of gliomas. Notably, we provided preliminary evidence that three deleted genes (*CDKN2A*, *CDKN2B* and *IFNB*) were associated with gliomas recurrence, which was also reported to be associated with early recurrence in meningiomas and to instigate chemotherapy-induced immunological dormancy in breast cancer.[20,21] Future studies are needed to characterize the mechanisms by which these markers interact to increase the risk of poor outcome, to establish the potential of applying to clinical management of glioma patients.

The good performance of our model on the independent MSKCC gliomas cohort demonstrated its feasible application on clinical dataset sequenced by limited target panel. Missed signature markers due to limited region, including CN-Sig1 focal amplification, CN-Sig3 homozygous deletion and CN-Sig6 loss of heterozygosity copy number signature will reduce the predicted risk score but also could be partially represented by remained markers in our model, such as *EGFR* amplification, *PTEN* deletion and *CDKN2A/B* deletion. Although the prognostic risk groups of gliomas were stratified using the genomic features of this model, the expression profile and immune compartments of different subgroups are distinct from each other. For the high-risk glioma patients with shortened survival, our data showed that their immunosuppressive tumor microenvironment had higher level of neutrophils, resting $CD4^+$ memory T cells and activated biological functional pathways involved in G2M checkpoint in LGG and Oxidative phosphorylation in GBM to support tumor growth, suggesting the shape or influence of genomic variations on immune microenvironment.

Limitations of the study

A potential limitation of the current study is that we only used one published dataset for validating the performance of our risk model, and this dataset is targeted sequencing with limited region, which may not fully represent the actual level of our model. This limitation may be avoided in future studies through validation on more datasets with whole-exome sequencing data. Also, the number of glioma patients in this study varies widely across molecular subtypes. Although we trained and test model separately by each subtype, the accuracy of prognosis prediction for specific subtype may be decreased by the insufficient number of patients. Finally, the conclusions in this study were mainly drawn by silico methods, which need to be further validated in further experiments and future large-scale prospective studies.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- RESOURCE AVAILABILITY

- Lead contact
- Materials availability
- Data and code availability
- EXPERIMENTAL MODEL AND SUBJECT DETAILS
- METHOD DETAILS
  - Variants calling
  - Mutation signature analysis
  - Copy number signature identification
  - Identification of significantly altered genes with prognostic implication
  - Training and validation of prognostic risk model for gliomas
  - Comparison of immune microenvironment features
  - Differential expression analysis
  - Gene set enrichment analysis
- QUANTIFICATION AND STATISTICAL ANALYSIS

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.isci.2022.105681.

ACKNOWLEDGMENTS

This study was supported by RGC Collaborative Research Fund (C4039-19GF), RGC Research Impact Fund (R1020-18F), Science and Technology Foundation of Shenzhen Grant (KQTD20180411185028798). X-Y.G. is a Sophie YM Chan Professor of Cancer Research. We thank Dr. Tao Jiang and Dr. Zheng Zhao for sharing the Chinese Glioma Genome Atlas (CGGA) whole exome raw sequencing data.

AUTHOR CONTRIBUTIONS

Study design: B.Z. and S.T. Bioinformatics analysis: B.Z. Published datasets download and processing: B.Z., W.W., and Z.L. Sample collection and pathological examination: Z.G., N.J., J.X., J.W., B.W., Y.Z., Y.L., and L.Z. Manuscript draft: B.Z. Manuscript revision: B.Z., D.K., X.G., K.R., and S.T. Approved the final manuscript: All authors.

DECLARATION OF INTERESTS

The authors declare no competing interests.

Received: August 22, 2021
Revised: November 18, 2022
Accepted: November 22, 2022
Published: December 22, 2022

REFERENCES

1. Ostrom, Q.T., Cioffi, G., Gittleman, H., Patil, N., Waite, K., Kruchko, C., and Barnholtz-Sloan, J.S. (2019). CBTRUS statistical report: primary brain and other central nervous system tumors diagnosed in the United States in 2012–2016. Neuro Oncol. *21*, v1–v100. https://doi.org/10.1093/NEUONC/NOZ150.

2. Cancer Genome Atlas Research Network, McLendon, R., Friedman, A., Bigner, D., Van Meir, E.G., Brat, D.J., Mastrogianakis, G.M., Olson, J.J., Mikkelsen, T., Lehman, N., Aldape, K., et al. (2008). Comprehensive genomic characterization defines human glioblastoma genes and core pathways. Nature *455*, 1061–1068. https://doi.org/10.1038/nature07385.

3. Brennan, C.W., Verhaak, R.G.W., McKenna, A., Campos, B., Noushmehr, H., Salama, S.R., Zheng, S., Chakravarty, D., Sanborn, J.Z., Berman, S.H., et al. (2013). The somatic genomic landscape of glioblastoma. Cell *155*, 462–477. https://doi.org/10.1016/j.cell.2013.09.034.

4. Hu, H., Mu, Q., Bao, Z., Chen, Y., Liu, Y., Chen, J., Wang, K., Wang, Z., Nam, Y., Jiang, B., et al. (2018). Mutational landscape of secondary glioblastoma guides MET-targeted trial in brain tumor. Cell *175*, 1665–1678.e18. https://doi.org/10.1016/j.cell.2018.09.038.

5. Cancer Genome Atlas Research Network, Brat, D.J., Verhaak, R.G.W., Aldape, K.D., Alfred Yung, W.K., Salama, S.R., Cooper, L.A.D., Rheinbay, E., Miller, C.R., Vitucci, M., Morozova, O., et al. (2015). Comprehensive, integrative genomic analysis of diffuse lower-grade gliomas. N. Engl. J. Med. *372*, 2481–2498. https://doi.org/10.1056/NEJMOA1402121.

6. Ceccarelli, M., Barthel, F.P., Malta, T.M., Sabedot, T.S., Salama, S.R., Murray, B.A., Morozova, O., Newton, Y., Radenbaugh, A., Pagnotta, S.M., et al. (2016). Molecular profiling reveals biologically discrete subsets and pathways of progression in diffuse glioma. Cell *164*, 550–563. https://doi.org/10.1016/J.CELL.2015.12.028.

7. Louis, D.N., Perry, A., Reifenberger, G., von Deimling, A., Figarella-Branger, D., Cavenee, W.K., Ohgaki, H., Wiestler, O.D., Kleihues, P., and Ellison, D.W. (2016). The 2016 World Health organization classification of tumors of the central nervous system: a summary. Acta Neuropathol. *131*, 803–820. https://doi.org/10.1007/S00401-016-1545-1.

8. Louis, D.N., Perry, A., Wesseling, P., Brat, D.J., Cree, I.A., Figarella-Branger, D., Hawkins, C., Ng, H.K., Pfister, S.M., Reifenberger, G., et al. (2021). The 2021 WHO

classification of tumors of the central nervous system: a summary. Neuro Oncol. *23*, 1231–1251. https://doi.org/10.1093/NEUONC/NOAB106.

9. Alexandrov, L.B., Kim, J., Haradhvala, N.J., Huang, M.N., Tian Ng, A.W., Wu, Y., Boot, A., Covington, K.R., Gordenin, D.A., Bergstrom, E.N., et al. (2020). The repertoire of mutational signatures in human cancer. Nature *578*, 94–101. https://doi.org/10.1038/s41586-020-1943-3.

10. Macintyre, G., Goranova, T.E., De Silva, D., Ennis, D., Piskorz, A.M., Eldridge, M., Sie, D., Lewsley, L.A., Hanif, A., Wilson, C., et al. (2018). Copy number signatures and mutational processes in ovarian carcinoma. Nat. Genet. *50*, 1262–1270. https://doi.org/10.1038/s41588-018-0179-8.

11. Kocakavuk, E., Anderson, K.J., Varn, F.S., Johnson, K.C., Amin, S.B., Sulman, E.P., Lolkema, M.P., Barthel, F.P., and Verhaak, R.G.W. (2021). Radiotherapy is associated with a deletion signature that contributes to poor outcomes in patients with cancer. Nat. Genet. *53*, 1088–1096. https://doi.org/10.1038/s41588-021-00874-3.

12. Zhao, Z., Zhang, K.-N., Wang, Q., Li, G., Zeng, F., Zhang, Y., Wu, F., Chai, R., Wang, Z., Zhang, C., et al. (2021). Chinese glioma genome Atlas (CGGA): a comprehensive resource with functional genomic data from Chinese glioma patients. Genom. Proteom. Bioinform. *19*, 1–12. https://doi.org/10.1016/J.GPB.2020.10.005.

13. Wang, J., Cazzato, E., Ladewig, E., Frattini, V., Rosenbloom, D.I.S., Zairis, S., Abate, F., Liu, Z., Elliott, O., Shin, Y.-J., et al. (2016). Clonal evolution of glioblastoma under therapy. Nat. Genet. *48*, 768–776. https://doi.org/10.1038/ng.3590.

14. Jonsson, P., Lin, A.L., Young, R.J., DiStefano, N.M., Hyman, D.M., Li, B.T., Berger, M.F., Zehir, A., Ladanyi, M., Solit, D.B., et al. (2019). Genomic correlates of disease progression and treatment response in prospectively characterized gliomas. Clin. Cancer Res. *25*, 5537–5547. https://doi.org/10.1158/1078-0432.CCR-19-0032.

15. Wang, S., Li, H., Song, M., Tao, Z., Wu, T., He, Z., Zhao, X., Wu, K., and Liu, X.S. (2021). Copy number signature analysis tool and its application in prostate cancer reveals distinct mutational processes and clinical outcomes. PLoS Genet. *17*, e1009557. https://doi.org/10.1371/journal.pgen.1009557.

16. Dietlein, F., Weghorn, D., Taylor-Weiner, A., Richters, A., Reardon, B., Liu, D., Lander, E.S., Van Allen, E.M., and Sunyaev, S.R. (2020). Identification of cancer driver genes based on nucleotide context. Nat. Genet. *52*, 208–218. https://doi.org/10.1038/s41588-019-0572-y.

17. Mermel, C.H., Schumacher, S.E., Hill, B., Meyerson, M.L., Beroukhim, R., and Getz, G. (2011). GISTIC2.0 facilitates sensitive and confident localization of the targets of focal somatic copy-number alteration in human cancers. Genome Biol. *12*, R41. https://doi.org/10.1186/GB-2011-12-4-R41.

18. Chalabi, M., Fanchi, L.F., Dijkstra, K.K., Van den Berg, J.G., Aalbers, A.G., Sikorska, K., Lopez-Yurda, M., Grootscholten, C., Beets, G.L., Snaebjornsson, P., et al. (2020). Neoadjuvant immunotherapy leads to pathological responses in MMR-proficient and MMR-deficient early-stage colon cancers. Nat. Med. *26*, 566–576. https://doi.org/10.1038/s41591-020-0805-8.

19. Mandal, R., Samstein, R.M., Lee, K.W., Havel, J.J., Wang, H., Krishna, C., Sabio, E.Y., Makarov, V., Kuo, F., Blecua, P., et al. (2019). Genetic diversity of tumors with mismatch repair deficiency influences anti-PD-1 immunotherapy response. Science *364*, 485–491. https://doi.org/10.1126/science.aau0447.

20. Sievers, P., Hielscher, T., Schrimpf, D., Stichel, D., Reuss, D.E., Berghoff, A.S., Neidert, M.C., Wirsching, H.-G., Mawrin, C., Ketter, R., et al. (2020). CDKN2A/B homozygous deletion is associated with early recurrence in meningiomas. Acta Neuropathol. *140*, 409–413. https://doi.org/10.1007/S00401-020-02188-W.

21. Lan, Q., Peyvandi, S., Duffey, N., Huang, Y.-T., Barras, D., Held, W., Richard, F., Delorenzi, M., Sotiriou, C., Desmedt, C., et al. (2018). Type I interferon/IRF7 axis instigates chemotherapy-induced immunological dormancy in breast cancer. Oncogene *38*, 2814–2829. https://doi.org/10.1038/s41388-018-0624-2.

22. Cibulskis, K., Lawrence, M.S., Carter, S.L., Sivachenko, A., Jaffe, D., Sougnez, C., Gabriel, S., Meyerson, M., Lander, E.S., and Getz, G. (2013). Sensitive detection of somatic point mutations in impure and heterogeneous cancer samples. Nat. Biotechnol. *31*, 213–219. https://doi.org/10.1038/nbt.2514.

23. Mayakonda, A., Lin, D.C., Assenov, Y., Plass, C., and Koeffler, H.P. (2018). Maftools: efficient and comprehensive analysis of somatic variants in cancer. Genome Res. *28*, 1747–1756. https://doi.org/10.1101/gr.239244.118.

24. Guo, X., Chen, F., Gao, F., Li, L., Liu, K., You, L., Hua, C., Yang, F., Liu, W., Peng, C., et al. (2020). CNSA: a data repository for archiving omics data. Database *2020*, baaa055. https://doi.org/10.1093/DATABASE/BAAA055.

25. Thorsson, V., Gibbs, D.L., Brown, S.D., Wolf, D., Bortone, D.S., Ou Yang, T.H., Porta-Pardo, E., Gao, G.F., Plaisier, C.L., Eddy, J.A., et al. (2018). The immune landscape of cancer. Immunity *48*, 812–830.e14. https://doi.org/10.1016/j.immuni.2018.03.023.

26. Liu, J., Lichtenberg, T., Hoadley, K.A., Poisson, L.M., Lazar, A.J., Cherniack, A.D., Kovatich, A.J., Benz, C.C., Levine, D.A., Lee, A.V., et al. (2018). An integrated TCGA pan-cancer clinical data resource to drive high-quality survival outcome analytics. Cell *173*, 400–416.e11. https://doi.org/10.1016/j.cell.2018.02.052.

27. Lichtenstein, L., Woolf, B., MacBeth, A., Birsoy, O., and Lennon, N. (2016). ReCapSeg: Validation of Somatic Copy Number Alterations for CLIA Whole Exome Sequencing.

28. Ignatiadis, N., Klaus, B., Zaugg, J.B., and Huber, W. (2016). Data-driven hypothesis weighting increases detection power in genome-scale multiple testing. Nat. Methods *13*, 577–580. https://doi.org/10.1038/nmeth.3885.

29. Liberzon, A., Birger, C., Thorvaldsdóttir, H., Ghandi, M., Mesirov, J.P., and Tamayo, P. (2015). The molecular signatures database hallmark gene set collection. Cell Syst. *1*, 417–425. https://doi.org/10.1016/J.CELS.2015.12.004.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Deposited data		
Raw whole-exome sequencing data and clinical information on the Chinese glioma patients	China National Center for Bioinformation-Beijing Institute of Genomics	https://ngdc.cncb.ac.cn/bioproject/browse/PRJCA001636
Somatic variants (including SNVs and SCNAs), normalized RNA expression profile, the immune cell fractions of tumor microenvironment and clinical information of TCGA glioma samples	Genomic Data Commons (GDC) data portal	https://gdc.cancer.gov/
Somatic mutational and clinical data in the MSKCC glioma samples	cBioPortal for Cancer Genomics	www.cbioportal.org/study?id=glioma_mskcc_2019
Raw whole-exome sequencing data and clinical information on the Chinese glioma patients newly seuqenced in this study	China National GeneBank DataBase (CNGBdb)	https://db.cngb.org/search/project/CNP0002128/
Software and algorithms		
Mutect2 (version 4.1.4)	Cibulskis et al.[22]	https://gatk.broadinstitute.org/hc/en-us/articles/360037593851-Mutect2
ReCapSeg	GATK	http://gatkforums.broadinstitute.org/categories/recapseg-documentation
SigProfiler	AlexandrovLab	https://github.com/AlexandrovLab/SigProfilerExtractor
MutPanning	Dietlein et al.[16]	https://www.genepattern.org/modules/docs/MutPanning#gsc.tab=0
GISTIC v2	Mermel et al.[17]	https://www.genepattern.org/modules/docs/GISTIC_2.0#gsc.tab=0
Maftools	Mayakonda et al.[23]	https://github.com/dentearl/mafTools
Sigminer	R package	NA
clusterProfiler	R package	NA
survminer	R package	NA
ggpubr	R package	NA

RESOURCE AVAILABILITY

Lead contact

Further information and requests for resources should be directed to and will be fulfilled by the lead contact, Baifeng Zhang (zhangbaifeng1@gmail.com).

Materials availability

This study did not generate new unique reagents.

Data and code availability

- The data that support the findings of this study have been deposited into CNGB Sequence Archive (CNSA)[24] of China National GeneBank DataBase (CNGBdb) with accession number CNP0002128.
- All original code has been deposited into CNGB Sequence Archive (CNSA)[24] of China National GeneBank DataBase (CNGBdb) with accession number CNP0002128.
- Any additional information required to reanalyze the data reported in this paper is available from the lead contact upon request.

EXPERIMENTAL MODEL AND SUBJECT DETAILS

The study was approved by the institutional review boards at Ethics Committee of Beijing Tiantan Hospital and informed consent was obtained from each participant. Raw whole-exome sequencing data and clinical information on the Chinese glioma patients were downloaded from the China National Center for Bioinformation-Beijing Institute of Genomics (BIGD accession number: PRJCA001636, https://ngdc.cncb.ac.cn/bioproject/browse/PRJCA001636).

Somatic variants (including SNVs and SCNAs) in the TCGA glioma samples with whole-exome sequencing data and clinical information were downloaded from the Genomic Data Commons (GDC) data portal (https://gdc.cancer.gov/).[25,26] Somatic mutational and clinical data in the MSKCC glioma samples were downloaded from the cBioPortal for Cancer Genomics (www.cbioportal.org/study?id=glioma_mskcc_2019). Tissue samples from 80 Chinese glioma patients were snap-frozen in liquid nitrogen immediately after surgical resection and preserved in liquid nitrogen (Table S2). Then, we performed whole exome sequencing using the Illumina HiSeq 4000 platform using pair-end sequencing strategy. After removing the adapters and quality control filtering, the sequencing reads were aligned to the reference human genome (hg19) using BWA.

METHOD DETAILS

Variants calling

All somatic SNVs were called out by the Mutect2 software (version 4.1.4).[22] All somatic SNVs were further filtered with the following parameters: a read depth of at least 10× in the germline and tumor samples, a maximum of two variant supporting reads in the germline, a minimum tumor variant allele frequency of 10% and a maximum germline variant allele frequency of 2%. The copy numberdata were segmented with the ReCapSeg software to identify the SCNAs, with all cohorts being processed by the same standard pipeline as described in GATK documentation provided by the Broad Institute (http://gatkforums.broadinstitute.org/categories/recapseg-documentation).[27]

Mutation signature analysis

We applied the computational framework named SigProfiler to decipher mutational signature profiles of 1,477 glioma patients and for assigning contributions of each signature to each patients, based on previous described methodology.[10] We feed the computational framework with the MAF profile of single nucleotide variants for all patients as input, and then run framework based on somatic mutations in sequence context and their distributions in each patient, and used multiple NMF iterations (10, 000–1, 000, 000) to decompose into signatures with predefined aetiologies that optimally explains the faction of each mutation context type in each mutational signature and estimates their activity to each sample.

Copy number signature identification

The extraction of copy number signatures was implemented by using an R package sigminer (https://cran.r-project.org/web/packages/sigminer/)[15] and could been summarized into the following three steps: (i) *Preprocessing*. The whole absolute copy number profiles of total 1, 477 glioma patients were summarized into patients-by-features matrices using genome-wide distribution of eight different features: Breakpoint count per 10 Mb (BP10MB), Breakpoint count per chromosome arm (BPArm), The observed absolute copy number of each segment (CN), The absolute difference in copy number between adjacent segments across the genome (CNCP), Length of segments with oscillating copy number (OsCN), The log10 length of each segments (SS), The minimal number of chromosome with 50% copy number variation (NC50), The burden of per chromosome (BoChr). These features were chosen as the distribution pattern of copy number events followed by previous described methodology.[10] (ii) *Signature identification*. The non-negative matrix factorization (NMF) algorithm with the input of the summarized patients-by-features matrices was used to extract signatures. As suggested, the optimal number of signatures as 6 is determined by trade-off of the mean sample cosine distance and average stability of solutions for a range of 2 to 10 after performing 50 runs. (iii) *Signature assignment*. The copy number segment records were assigned to extracted signatures and their expected absolute number in each patient were calculated.

Identification of significantly altered genes with prognostic implication

We feed MutPanning software with all somatic SNVs as input and identified 187 significantly somatic mutated genes followed by suggested criterial (FDR <0.25) according to mutations in unusual nucleotide contexts.[16] Based on the whole absolute copy number profiles, significantly copy number altered genomic

region in focal level were identified by GISTIC v2[17] and further analyzed and visualized by Maftools.[23] Total 2813 genes located on these focal amplified or deleted region were extracted (FDR <0.25). Overall, 3000 genes were finally defined as significantly altered genes for downstream analysis.

Training and validation of prognostic risk model for gliomas

In the search for potential prognostic markers, univariable survival analyses for 187 significantly somatic mutated genes, 2813 focal amplified or deleted genes, 16 single-base substitution (SBS) mutational signatures and 6 copy number alteration signatures separately were performed using Kaplan-Meier and log rank tests. Took account into statistic power and computation efficiency, altered genes should occur with above 3% for somatic mutation and over 50% for copy number changes in total 1, 477 patients (Chinese and TCGA cohort). Finally, nine somatic mutated genes, twenty-one copy number changed genes, six SBS mutational signatures and six copy number alteration signatures were defined as potential prognostic markers (FDR <0.25, log rank tests).

To build a multivariable prognostic prediction model, we generate potential prognostic marker*patient binary matrix data-each row corresponds to a patient and each column a marker as a covariate and corresponding response matrix with a column "time" of failure/censoring survival times and "status" a 0/1 indicator (1 means death, 0 means alive). Firstly, we separate gliomas into LGG-IDH-wildtype, LGG-IDH-mutant, LGG-1p/19q-co-deleted IDH-mutant, GBM-IDH-wildtype, GBM-IDH-mutant, GBM-1p/19q-codeleted IDH-mutant six major subtypes. For each subtype, we then randomly split the data into training set (50% for building a predictive model) and test set (50% for evaluating the model). A regularized cox model with 10-fold cross-validation was trained and implemented in *glmnet* R packge. Finally, after evaluation of regression coefficients, a prognostic risk prediction model with the combination of ten biomarkers including somatic gene mutation*IDH1*-Mut, *CIC*-Mut, *TP53*-Mut, somatic gene copy number alteration *CDKN2B*-Del, *CDKN2A*-Del, *PTEN*-Del, *EGFR*-Amp and copy number signature Sig1, Sig3, and Sig6 were successfully developed.

To visualize model performance in our dataset, risk scores for each patient were computed based on ten-marker prognostic prediction model and were used to split the patients into high (higher than median risk value) and low risk (equal to or lower than median risk value) subgroups for each molecular subtype. Kaplan-Meier curves were created for all subgroups.

Comparison of immune microenvironment features

The immune cell fractions of tumor microenvironment for TCGA glioma patients were estimated by CIBERSORT and downloaded from the Genomic Data Commons (GDC) data portal (https://gdc.cancer.gov/).[25] Then, we compared the difference of immune cell fractions between TCGA glioma patients with high risk and low risk predicted by our prognostic prediction model.

Differential expression analysis

Normalized RNA expression profile of TCGA glioma patients were also downloaded from the Genomic Data Commons (GDC) data portal (https://gdc.cancer.gov/).[25] The quantification and statistical inference of systematic gene expression changes between high risk and low risk TCGA glioma patients were performed by DESeq2 software followed by default parameter.[28]

Gene set enrichment analysis

We used a collection called Hallmark gene sets from The Molecular Signatures Database (MSigDB).[29] We generated a ranked gene list by their differential expression changes between risk subgroups and performed gene set enrichment analysis on this pre-ranked gene list to calculate the enrichment score for each Hallmark gene set. Whole procedure was implemented in clusterProfiler R package.[29]

QUANTIFICATION AND STATISTICAL ANALYSIS

Two-sided Mann-Whitney tests were performed with using Wilcoxon signed-rank test to generate the empirical *p* values and plotted with the R package "ggpubr". Data are plotted as mean ± SEM unless otherwise noted. Log-rank tests and appropriate Chi-square statistics were used to determine survival differences. These statistics and associated Kaplan-Meier curves were plotted using the R package "survival" and "survminer". *p* value <0.05 was considered statistically significant for all computational analysis unless otherwise stated.