Computational genomics of brain tumors: identification and characterization of glioma candidate biomarkers through multi-omics integrative molecular profiling

Lin Liu[1,2,3,4,#], Guangyu Wang[1,2,3,4,a,#], Liguo Wang[5], Chunlei Yu[1,2,3,4], Mengwei Li[1,2,3,4], Shuhui Song[1,2,3,4], Lili Hao[1,2,3,4], Lina Ma[1,2,3,4,*], Zhang Zhang[1,2,3,4,*]

[1] China National Center for Bioinformation, Beijing 100101, China

[2] National Genomics Data Center, Beijing 100101, China

[3] BIG Data Center & CAS Key Laboratory of Genome Sciences and Information, Beijing Institute of Genomics, Chinese Academy of Sciences, Beijing 100101, China

[4] University of Chinese Academy of Sciences, Beijing 100101, China

[5] Division of Biomedical Statistics and Informatics, Mayo Clinic College of Medicine, Rochester, Minnesota 55905, United States

[a] Present Address: The Methodist Hospital Research Institute, 6670 Bertner Ave, Houston, Texas 77030, United States

[#] These authors contributed equally.

*Corresponding authors: malina@big.ac.cn (M.L.N.), zhangzhang@big.ac.cn (Z.Z.)

Abstract

Glioma is one of the most common malignant brain tumors and exhibits low resection rate and high recurrence risk. Although a large number of glioma studies powered by high-throughput sequencing technologies have led to massive multi-omics datasets, there lacks of comprehensive integration of glioma datasets for uncovering candidate biomarker genes. In this study, we collected a large-scale assemble of multi-omics multi-cohort datasets from worldwide public resources, involving a total of 16,939 samples across 19 independent studies. Through comprehensive multi-omics molecular profiling across different datasets, we revealed that *PRKCG* (Protein Kinase C Gamma), a brain-specific gene detectable in cerebrospinal fluid, is closely associated with glioma. Specifically, it presents lower expression and higher methylation in glioma samples compared with normal samples. *PRKCG* expression/methylation change from high to low is indicative of glioma progression from low-grade to high-grade and high RNA expression is suggestive of good survival. Importantly, *PRKCG* in combination with *MGMT* is effective to predict survival outcomes after TMZ chemotherapy in a more precise manner. Collectively, *PRKCG* bears the great potential for glioma diagnosis, prognosis and therapy, and *PRKCG*-like genes may represent a set of important genes associated with different molecular mechanisms in glioma tumorigenesis. Accordingly, our study indicates the importance of computational integrative multi-omics data analysis and represents a data-driven scheme toward precision tumor subtyping and accurate personalized healthcare.

Keywords: glioma, multi-omics, *PRKCG*, biomarker, cerebrospinal fluid

Author Summary

Glioma is a type of brain tumors that represents one of the most lethal human malignancies with little chance for recovery. Nowadays, more and more studies have adopted high-throughput sequencing technologies to decode the molecular profiles of glioma from different omics levels, accordingly resulting in massive glioma datasets generated from different projects and laboratories throughout the world. Therefore, it has become crucially important on how to make full use of these valuable datasets for computational identification of glioma candidate biomarker genes in aid of precision tumor subtyping and accurate personalized treatment. In this study, we comprehensively integrated glioma datasets from all over the world and performed multi-omics molecular data mining. We revealed that *PRKCG*, a brain-specific gene abundantly expressed in cerebrospinal fluid, bears the great potential for glioma diagnosis, prognosis and treatment prediction, which has been consistently observed on multiple independent datasets. In the era of big data, our study highlights the significance of computational integrative data mining toward precision medicine in cancer research.

62 ## Introduction

63 Glioma, one of the serious central nervous system (CNS) tumors, represents ~80% of malignant
64 brain tumors (1, 2) and exhibits low resection rate and high recurrence risk (3). Since tumor
65 classification benefits accurate diagnosis and facilitates precise treatment, gliomas can be
66 classified, according to the histologic grading schemes, into LGG (astrocytoma,
67 oligodendroglioma and oligoastrocytoma) and GBM (glioblastoma multiforme) (4). Therefore,
68 identification of reliable molecular biomarkers for precise classification of different-grade
69 gliomas is crucial to aid tumor diagnosis, establish appropriate therapies, recognize prognostic
70 outcome and predict therapeutic response (5).

71 Powered by high-throughput sequencing technologies, a set of molecular biomarkers have been
72 discovered from different omics levels to assist glioma diagnosis and treatment (6, 7). Among
73 them, isocitrate dehydrogenase (*IDH*) mutation and 1p/19q co-deletion (codel) are two most
74 important genetic events for glioma grading (8-10). Patients with *IDH* mutation (*IDH*-mut)
75 have longer survival than those with *IDH* wild-type (*IDH*-WT) (11, 12). And the 1p/19q codel
76 is a distinctive feature of oligodendroglioma (13, 14). Furthermore, based on these two genetic
77 alterations, accumulated evidence suggested that gliomas can be divided into three subtypes
78 (*IDH*-mut & 1p/19q codel, *IDH*-mut & 1p/19q non-codel, and *IDH*-WT & 1p/19q non-codel),
79 which are associated with diverse clinical outcomes (15). Accordingly, in 2016, the World
80 Health Organization (WHO), in light of both histology and significant genetic events (mainly
81 by *IDH* and 1p/19q), divided gliomas into five categories (16, 17), including three LGGs
82 (diffuse astrocytoma, *IDH*-mut & 1p/19q non-codel; oligodendroglioma, *IDH*-mut & 1p/19q
83 codel; diffuse astrocytoma, *IDH*-WT & 1p/19q non-codel) and two GBMs (*IDH*-mut; *IDH*-
84 WT). Meanwhile, molecular markers at the transcriptome level have also been identified (18-
85 21); for example, an overexpression of epidermal growth factor receptor variant III (*EGFRvIII*)

86 has been reported to associate with malignant progression of GBM (22-24). In addition,
87 epigenetic modifications are also implicated in glioma (25-29). One classical biomarker is O6-
88 methylguanine-DNA-methyltransferase (*MGMT*) (30); patients with methylated *MGMT*
89 promoter have better clinical outcomes and are more sensitive to the alkylating chemotherapy
90 than those without methylated *MGMT* promoter (31-33).

91 Nowadays, there is an increasing number of high-throughput studies for better understanding
92 of glioma tumorigenesis (34-38), resulting in massive multi-omics datasets generated from
93 different projects and laboratories throughout the world. However, there lacks of
94 comprehensive integration of glioma datasets for computationally identifying and
95 characterizing candidate biomarkers. To this end, we collected a large-scale assemble of multi-
96 omics multi-cohort datasets from worldwide public resources and detected candidate biomarker
97 genes through comprehensive integrative molecular profiling on multiple independent datasets.
98 We revealed that *PRKCG*, a gene specifically expressed in brain and detectable in cerebrospinal
99 fluid (CSF), is closely associated with glioma, indicative of a potential biomarker for glioma
100 diagnosis, prognosis and treatment prediction.

101 Materials and Methods

102 Data collection

103 In this study, we collected a comprehensive assemble of multi-omics datasets (including
104 genomics, transcriptomics, DNA methylomics and proteomics) from The Cancer Genome Atlas
105 (TCGA, https://portal.gdc.cancer.gov/) (35), Genotype-Tissue Expression Portal (GTEx,
106 https://gtexportal.org/home/) (39), Gene Expression Omnibus (GEO,
107 https://www.ncbi.nlm.nih.gov/geo), Ivy Glioblastoma Atlas Project (Ivy GAP,
108 http://glioblastoma.alleninstitute.org) (40) and Chinese Glioma Genome Atlas (CGGA,

109 http://www.cgga.org.cn) (41, 42). Particularly, discovery datasets were derived from TCGA,
110 GTEx and large cohort studies in GEO (GSE83710, GSE16011 and GSE36278 for protein,
111 expression and methylation, respectively). As a result, a total of five discovery datasets and
112 fourteen validation datasets were obtained. For convenience, each dataset collected in this study
113 is assigned a unique accession number with the format: [D/V][*i*]-
114 [TCGA/GTEx/GEO/CGGA/Ivy GAP]-[E/V/P/M], where D/V in the first bracket represents the
115 dataset for discovery or validation, *i* in the second bracket indicates the dataset number, the
116 third bracket shows the data source (as mentioned above), and the last bracket indicates the data
117 type, namely, E for RNA expression, V for CNV, P for protein expression and M for DNA
118 methylation, respectively. The detailed information about all collected datasets was tabulated
119 in Table 1.

120 **Identification of brain-specific genes**

121 To identify brain-specific genes, we used the RNA-Seq dataset from GTEx (2016-01-15; v7)
122 (39), which contains 11,688 samples across 53 tissue sites of 714 donors. Considering that
123 several tissues have multiple different sites, gene expression levels were averaged over sites
124 that are from the same tissue. To reduce background noise, genes with maximal expression
125 levels smaller than 10 TPM (Transcripts Per Million) were removed. Finally, we obtained a
126 total of 15,176 gene expression profiles across 30 tissues (S1 Table).

127 Based on the expression levels across 30 tissues, we calculated the tissue specificity index τ
128 (43) for each gene to identify tissue-specific genes. τ is valued between 0 and 1, where 0
129 represents housekeeping genes that are consistently expressed in different tissues, and 1
130 indicates tissue-specific genes that are exclusively expressed in only one tissue (43). In this
131 study, brain-specific genes were defined as those genes that are maximally expressed in the

132 brain with τ>0.9. As a consequence, a list of the top 100 brain-specific genes ranked by the τ
133 index were obtained for further analysis (S2 Table).

134 Sample classification

135 To comprehensively study the potential of *PRKCG* in glioma diagnosis, we compared the
136 molecular profiles between normal and glioma samples, between LGG and GBM samples,
137 between primary GBM (pGBM) and recurrent GBM (rGBM) samples, and between glioma
138 samples with different anatomic features. We collected 122 GBM samples from the Ivy GAP
139 database (40) and grouped them according to their anatomic regions, namely, leading edge (LE,
140 the ratio of tumor/normal cells is about 1–3/100), infiltrating tumor (IT, the ratio of
141 tumor/normal cells is about 10–20/100), cellular tumor (CT, the ratio of tumor/normal cells is
142 about 100/1 to 500/1), pseudo-palisading cells around necrosis (PAN, the narrow boundary of
143 cells along the dead tissue), and microvascular proliferation (MVP, two or more blood vessels
144 sharing a common vessel wall of endothelial).

145 We investigated the prognostic role of *PRKCG* by dividing samples into subgroups based on
146 *PRKCG*'s expression level within all glioma samples and also within LGG and GBM samples,
147 respectively. When exploring the predictive role of *PRKCG*, we obtained DNA methylation
148 status (methylated and unmethylated) directly from the original study (35), which was defined
149 based on the beta value cutoff 0.3.

150 Identification of *PRKCG*-like genes

151 Genes that satisfy the following criteria were regarded as *PRKCG*-like genes: (1) Higher
152 methylation level of at least one CpG site (promoter region) in glioma samples than normal
153 samples; (2) Higher DNA methylation level in LGG samples than GBM samples; (3) Higher
154 expression level in LGG samples than GBM samples; and (4) Lower expression level in glioma

155 samples than normal samples. As a result, we obtained a total of 542 *PRKCG*-like genes, which
156 were further divided into two groups according to their correlations between gene expression
157 and methylation, namely, 114 genes with negative correlation and 297 genes with positive
158 correlation.

159 **Statistical analysis**

160 All statistical analyses were performed using R version 3.3.2. The Wilcoxon test was used for
161 the analysis of the difference in gene expression/methylation between tumor and normal
162 samples, and between different glioma subtypes. The statistical significance levels were coded
163 by *ns* (not significant) $p > 0.05$, * $p < 0.05$, ** $p < 0.01$ and *** $p < 0.001$. We performed the
164 survival analysis using the Kaplan-Meier method and estimated the statistical difference using
165 the log-rank test.

166 **Data availability statement**

167 All datasets integrated in this study were obtained from multiple public database resources (see
168 details in Table 1), which are freely available at ftp://download.big.ac.cn/glioma_data/.

169 **Results and Discussion**

170 ***PRKCG* is a brain-specific gene and detectable in cerebrospinal fluid**

171 Tissue-specific genes are believed to be crucial for identifying potential biomarkers with high
172 specificity (44-48). To identify candidate genes with brain specificity, we integrated expression
173 data from GTEx (D1-GTEx-E) (39), explored all genes' expression profiles and their tissue
174 specificity, and identified a list of top 100 brain-specific genes (S2 Table). To achieve the
175 detectability in the periphery, we assembled a total of 1,126 CSF-detectable proteins from GEO
176 (D2-GSE83710-P) (49), due to the critical significance of CSF as a feasible means to detect
177 genes expressed in human CNS (50, 51). After integrating brain-specific genes with CSF

178 proteins, we revealed that there are five brain-specific proteins that can be detected in CSF (Fig
179 1 and S1 Fig), in terms of fluorescence intensity from high to low, namely, *PRKCG, BCAN*,
180 *OPCML*, *GFAP* and *CAMK2A*, which are diversely expressed in different brain regions (S2
181 Fig). Specifically, *BCAN*, a member of the lectican family of chondroitin sulfate proteoglycans,
182 is highly expressed in glioma and may promote cell motility of brain tumor cells (52, 53). In
183 addition, the fusion event between *BCAN* and *NTRK1* (*BCAN-NTRK1*) is a potential glioma
184 driver and therapeutic target (54). *OPCML* encodes a member of the IgLON subfamily in the
185 immunoglobulin proteins and is down-regulated in gliomas and other brain tumors (55, 56).
186 *GFAP*, encoding one of the major intermediate filament proteins of mature astrocytes (57), can
187 be used to assess the differentiation state of astrocytoma (58). *CAMK2A* is a calcium
188 calmodulin-dependent protein kinase and reduced expression of *CAMK2A* is associated with
189 better survival in GBM (59, 60).

190 Remarkably, *PRKCG* (Protein Kinase C Gamma), a member of protein kinase C (PKC) family
191 located in 19q, exhibits higher fluorescence intensity than the other four genes (Fig 1 and S1
192 Fig). The expression profile of *PRKCG* across multiple brain developmental stages reveals that
193 its expression is extremely lower in the prenatal stages, but dramatically increases in the infancy
194 stages and is stabilized in the latter stages according to GenTree (S3 Fig) (61). Previous studies
195 have documented that unlike other PKC family members that are expressed in many tissues
196 aside from brain, *PRKCG* is brain-specifically expressed (62) and that mutations in *PRKCG* are
197 associated with spinocerebellar ataxia (63, 64). Additionally, it has been reported that PKC
198 signaling pathways contribute to the aggressive behavior of glioma cells (65) and atypical PKC
199 isozymes are fundamental regulators of tumorigenesis (66). To our knowledge, several genes
200 in 19q are closely associated with glioma (e.g., *TTYH1*, *UBE2S* (67, 68)). However, the
201 potential role of *PRKCG* in glioma remains unknown, and therefore, comprehensive molecular
202 characterization of *PRKCG* across multi-omics glioma datasets is highly desirable.

203 ***PRKCG* is significantly differentially expressed among normal, LGG and GBM samples**

204 We first investigated the expression pattern of *PRKCG* among normal, LGG and GBM samples
205 by using multiple discovery and validation datasets. We found that *PRKCG* expression is
206 significantly reduced in gliomas by contrast to normal samples (Fig 2A-F; p-value < 0.01,
207 Wilcoxon test). Furthermore, we discovered that *PRKCG* shows significantly different
208 expression profiles among different anatomic regions (Fig. 2G; p-value < 0.01, Wilcoxon test).
209 Strikingly, *PRKCG* expression is highest in LE (the outermost boundary of the tumor),
210 decreased in IT (the intermediate zone between the LE and the serious CT regions), and lowest
211 in the serious regions (CT, PAN and MVP) (see details in Materials and Methods). Consistently,
212 comparison between different-grade gliomas showed that *PRKCG* expression is significantly
213 lower in GBM samples than LGG samples (Fig 2H-J; p-value < 0.01, Wilcoxon test). We further
214 investigated its expression across pan-cancer samples. Although it has been documented that
215 *PRKCG* is up-regulated in colon cancer (69), the up-regulation in colon cancer is extremely
216 lower by comparison with glioma (LGG and GBM) (S4 Fig). Taken together, these results
217 presumably suggest that *PRKCG* is closely associated with glioma and its reduced expression
218 is coupled with glioma progression (Fig 2), highlighting its possible potential for glioma
219 diagnosis.

220 ***PRKCG* expression is highly sensitive to survival**

221 *PRKCG* expression change from high to low is indicative of progression from normal to glioma
222 and from LGG to GBM (Fig 2), implying that *PRKCG* expression is significantly associated
223 with glioma progression. Importantly, we observed that *PRKCG* expression is significantly
224 associated with survival rate, which is testified by multiple independent datasets (Fig 3).
225 Specifically, higher expression of *PRKCG* is indicative of longer overall survival in all glioma
226 samples (Fig 3A-B; p-value < 0.01, log-rank test). When separating LGG samples from GBM

227 samples, it is consistently observed that higher expression, albeit not statistically significant in
228 all examined datasets, tends to have longer overall survival in both LGG and GBM samples
229 (Fig 3C-F). Obviously, *PRKCG* expression has the potential capability to differentiate samples
230 with diverse survival states, which would be of critical significance for accurate glioma
231 subtyping, better therapeutic decisions and precision healthcare.

232 ***PRKCG* is significantly differentially methylated among normal, LGG and GBM samples**

233 Since *PRKCG* harbors two CpG sites (namely, cg26626089 and cg04518808) that are located
234 in the promoter region and covered in both HumanMethylation27 (27K) and
235 HumanMethylation450 (450K) BeadChip datasets, we then systematically investigated DNA
236 methylation profiles of these two sites among normal, LGG and GBM samples. Apparently, the
237 two sites show hypermethylation in GBM patients compared with normal samples (Fig 4),
238 which is more significant for cg26626089 (Fig 4A and 4C; p-value < 0.01, Wilcoxon test).
239 Furthermore, we examined the variation of methylation level by using whole-genome bisulfite
240 sequencing data of six GBM samples from TCGA and one normal sample from UCSC (2017
241 version; http://genome.ucsc.edu, last accessed on 12 May 2019). Consistently, most GBM
242 patients show higher methylation levels than normal samples (S5 Fig). In addition, considering
243 different-grade gliomas, both sites present much lower methylation levels in GBM samples than
244 LGG samples (Fig 4E-H; p-value < 0.01, Wilcoxon test). We further investigated its
245 methylation in pGBM and rGBM and obtained contradictory results in different populations;
246 *PRKCG* methylation shows no significant difference in the Chinese population (V10-CGGA-
247 M) (S6A and S6B Fig; p-value > 0.05, Wilcoxon test) but significantly difference in the
248 Switzerland population (V14-GSE60274-M) (70) (S6C and S6D Fig; p-value < 0.05, Wilcoxon
249 test). This is most likely caused by the population genetic difference and/or the small sample
250 size (both datasets have < 5 rGBM samples).

251 Collectively, *PRKCG* is significantly differentially expressed/methylated among normal, LGG
252 and GBM samples. Compared with normal samples, *PRKCG* presents lower expression and
253 higher methylation in glioma samples. With tumor malignancy, *PRKCG* methylation and
254 expression are both on the decrease (discussed below).

255 **Combined methylation signatures of *PRKCG* and *MGMT* are more effective in treatment**
256 **prediction**

257 It is known that *MGMT* encodes a DNA-repair protein and hypermethylation of *MGMT* is
258 associated with diminished DNA-repair activity, accordingly allowing the alkylating drug
259 temozolomide (TMZ) to have greater effect in GBM treatment (71, 72). In our study, we
260 obtained consistent results that patients with methylated *MGMT* are more sensitive to TMZ
261 treatment than those with unmethylated *MGMT* (Fig 5A; p-value < 0.01, log-rank test).

262 Considering that a single molecular biomarker might be lack of sufficient prediction power and
263 thus fail to determine the clinical therapeutic efficacy due to tumor heterogeneity (73), we
264 sought to examine the predictive potential of *PRKCG* for TMZ using 228 GBM samples with
265 matched DNA methylation and clinical data. We discovered that among the two CpG sites of
266 *PRKCG* (cg26626089 and cg04518808), cg26626089 is able to classify patients into two
267 groups with distinct survival advantages, as patients with methylated cg26626089 have
268 significantly longer survival than those with unmethylated cg26626089 (Fig 5B and S7A Fig).
269 By combining *PRKCG* (cg26626089) with *MGMT*, intriguingly, GBM patients receiving TMZ
270 treatment can be classified into four groups that exhibit significantly different survivals (Fig 5C
271 and S7B Fig; p-value < 0.01, log-rank test). The four groups, namely, *MGMT*-unmethylated +
272 *PRKCG*-unmethylated, *MGMT*-unmethylated + *PRKCG*-methylated, *MGMT*-methylated +
273 *PRKCG*-unmethylated, and *MGMT*-methylated + *PRKCG*-methylated, present gradually
274 improved longer survivals, as their 20-month OS rates are 0.18, 0.29, 0.39 and 0.51 (Fig 5C),

275 respectively, implying that the combined methylation signatures of *PRKCG* and *MGMT* might
276 guide more accurate GBM stratification and achieve better personalized therapeutic decisions.
277 Noticeably, elevated *MGMT* expression is associated with TMZ resistance (33). Similarly, we
278 consistently detected that the both-methylated group with better survival shows significantly
279 lower expression of *MGMT* (S7C Fig).

280 **Multi-omics molecular profiles of *PRKCG***

281 Based on multi-omics profiles of *PRKCG*, we explored the relationship between *PRKCG* and
282 classical molecular features/glioma grades. First, *PRKCG* is located on the chromosome
283 19q13.42, unifying previous findings that 1p/19q codel is closely associated with glioma.
284 Consistently, *PRKCG* CNV is closely associated with 19q status (Fig 6A and S8 Fig). Second,
285 *PRKCG* methylation (cg26626089) is associated with *IDH* status, agreeing well with the
286 finding that *IDH*-mut is an initiating event that remodels the glioma methylome, resulting in
287 extensive DNA hypermethylation (12, 74) and thus most likely indicating that *PRKCG*
288 methylation is a passenger of *IDH*-mut status. As LGG samples are always associated with
289 *IDH*-mut and GBM samples are associated with *IDH*-WT, it is not difficult to understand why
290 the methylation level of *PRKCG* is significantly lower in GBM than in LGG (Fig 4E-H). At the
291 same time, such higher expression level and higher methylation level lead to the suspicion
292 whether *PRKCG* expression in glioma is positively regulated by its DNA methylation or is
293 attributable to its CNV.

294 As expected, when considering CNV loss and gain separately, *PRKCG* CNV is positively
295 correlated with its expression, as observed in the CNV loss and gain groups, respectively (Fig
296 6B and 6C; p-value < 0.01, Spearman correlation = 0.26/0.32). However, such positive
297 correlation is absent when ignoring the difference of CNV status (Fig 6D). According to the
298 dosage effect theory (75), the CNV loss group should not express more *PRKCG* than the CNV

299 gain group. This implies that there is probably another factor rather than CNV to dominantly
300 regulate *PRKCG* expression in glioma. Although it contradicts the commonly accepted negative
301 association between gene expression and promoter CpG methylation, a large-scale pan-cancer
302 analysis has also revealed a positive correlation between promoter CpG methylation and gene
303 expression (76). Consistently, we did observe significant positive correlations between *PRKCG*
304 expression and CpG methylation within the promoter region (Fig 6E and 6F). This positive
305 regulation of CpG methylation is quite strong, which significantly improves *PRKCG* expression
306 in LGG samples; even these samples exhibit obvious CNV loss (Fig 6A). Thus, it is presumably
307 suggested that *PRKCG* is most likely regulated in different ways by DNA methylation, which
308 negatively regulates *PRKCG* expression from normal to tumor but positively regulates the
309 expression within tumor. However, gene expression regulation is a more complicated process
310 involving multiple factors aside from DNA methylation and CNV and more efforts on
311 comprehensive and in-depth molecular characterization of glioma are highly needed to
312 elucidate glioma pathogenesis.

313 **_PRKCG_-like genes may present heterogeneous roles in glioma tumorigenesis**

314 To better understand the regulation pattern of *PRKCG*, we further identified a total of 542
315 *PRKCG*-like genes that possess expression and DNA methylation patterns similar with *PRKCG*
316 (see Materials and Methods) and investigated whether these genes present heterogeneous roles
317 in glioma tumorigenesis (Fig 7A-B and S3 Table). Noticeably, some of these *PRKCG*-like
318 genes have already been reported to be closely related with glioma (77-81). For instance,
319 *AKAP6* (A-kinase anchoring protein 6) polymorphisms are associated with glioma
320 susceptibility and prognosis (80); Phosphorylated *SATB1* (SATB homeobox 1) contributes to
321 the invasive and proliferative of GBM cells and is associated with glioma prognosis (77);
322 Higher expression of *CDK17* (cyclin dependent kinase 17) is indicative of longer overall

323 survival (78); *PTPRM* (protein tyrosine phosphatase receptor type M) expression is
324 significantly reduced in GBM by contrast to LGG and higher expression is indicative of longer
325 overall survival (79); and *CHD5* (chromodomain helicase DNA binding protein 5) might act as
326 a tumor suppressor and its lower expression is associated with poor prognosis in glioma (81).
327 Among these *PRKCG*-like genes, we revealed that 114 genes show negative correlations
328 between methylation level and expression level, whereas 297 genes present positive
329 correlations (S3 Table). We further performed gene set enrichment analysis for these two
330 groups' genes. We found that the negatively-correlated genes are primarily enriched in the
331 MAPK and cGMP-PKG signaling pathways, which are essential for tumor cell proliferation
332 and differentiation (82, 83). On the contrary, the positively-correlated genes are significantly
333 involved in pathways relevant to cancer, inflammation and nerve synapse (Fig 7C). Thus,
334 *PRKCG*-like genes that exhibit the positive correlation between DNA methylation and
335 expression, presumably present heterogeneous roles in glioma tumorigenesis with complex
336 molecular mechanisms that need further extensive explorations both bioinformatically and
337 experimentally.

338 Conclusion

339 The rapid advancement of sequencing technologies enables large amounts of glioma data
340 generated from different projects and studies throughout the world. Therefore, it has become
341 crucially significant on how to make full use of these valuable data for computational
342 identification and characterization of glioma candidate biomarkers. In this study, we, for the
343 first time, assembled the most comprehensive collection of public glioma datasets with multi-
344 omics data types and different populations/countries. Through comprehensive molecular
345 profiling, we identified that *PRKCG*, a brain-specific gene detectable in CSF, is a potential
346 biomarker for glioma diagnosis, prognosis and treatment prediction. Specifically, it presents

347 lower expression and higher methylation in glioma samples than normal samples. *PRKCG*
348 expression/methylation change from high to low is indicative of glioma progression from low-
349 grade to high-grade and high RNA expression is suggestive of good survival. Importantly,
350 *PRKCG* in combination with *MGMT* is more effective to yield precise survival outcomes after
351 TMZ chemotherapy. In harmony with classical biomarkers, *PRKCG* as well as *PRKCG*-like
352 genes may play important and heterogeneous roles in glioma tumorigenesis. In the era of big
353 data, our findings highlight the importance of computational integrative multi-omics profiling
354 and represent a data-driven scheme toward precision tumor subtyping, accurate therapeutic
355 decisions and better personalized healthcare.

356 **Funding**

357 This work was supported by grants from The Strategic Priority Research Program of the
358 Chinese Academy of Sciences [XDA19050302, XDB13040500], National Key Research &
359 Development Program of China [2017YFC0907502, 2015AA020108], National Natural
360 Science Foundation of China [31871328], 13th Five-year Informatization Plan of Chinese
361 Academy of Sciences [XXH13505-05], K. C. Wong Education Foundation to Z.Z., the Youth
362 Innovation Promotion Association of Chinese Academy of Sciences [2019104] to M.L.N., and
363 International Partnership Program of the Chinese Academy of Sciences
364 [153F11KYSB20160008].

365 **Acknowledgements**

366 We thank Jun Yu, Songnian Hu, Fangqing Zhao, Yu Xue, Zheng Zhao, Jiabao Cao, Jian Sang,
367 Guangyi Niu, Man Li, and Yang Zhang for their valuable comments and discussions on this
368 work.

Author Contributions

Conceptualization, L.L. and W.G.Y.; Methodology, L.L., W.G.Y., and Z.Z.; Validation, L.L. and W.L.G.; Formal Analysis, L.L., W.G.Y., W.L.G., Y.C.L. and L.M.W.; Investigation, L.L. and W.G.Y.; Writing – Original Draft, L.L.; Writing – Review & Editing, M.L.N., W.L.G., S.S.H., H.L.L., and Z.Z.; Funding Acquisition, M.L.N. and Z.Z.; Supervision, M.L.N. and Z.Z.

Declaration of Interests

The authors declare no competing interests.

References

1. Schwartzbaum JA, Fisher JL, Aldape KD, Wrensch M. Epidemiology and molecular pathology of glioma. Nat Clin Pract Neurol. 2006;2(9):494-503; quiz 1 p following 16.

2. Maher EA, Furnari FB, Bachoo RM, Rowitch DH, Louis DN, Cavenee WK, et al. Malignant glioma: genetics and biology of a grave matter. Genes Dev. 2001;15(11):1311-33.

3. Stewart LA. Chemotherapy in adult high-grade glioma: a systematic review and meta-analysis of individual patient data from 12 randomised trials. Lancet. 2002;359(9311):1011-8.

4. 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(2):97-109.

5. Wesseling P, Kros JM, Jeuken JWM. The pathological diagnosis of diffuse gliomas: towards a smart synthesis of microscopic and molecular information in a multidisciplinary context. Diagnostic Histopathology. 2011;17(11):486-94.

6. Kim YW, Koul D, Kim SH, Lucio-Eterovic AK, Freire PR, Yao J, et al. Identification of prognostic gene signatures of glioblastoma: a study based on TCGA data analysis. Neuro Oncol. 2013;15(7):829-39.

7. Wiestler B, Capper D, Holland-Letz T, Korshunov A, von Deimling A, Pfister SM, et al. ATRX loss refines the classification of anaplastic gliomas and identifies a subgroup of IDH mutant astrocytic tumors with better prognosis. Acta Neuropathol. 2013;126(3):443-51.

8. Cohen AL, Holmen SL, Colman H. IDH1 and IDH2 mutations in gliomas. Curr Neurol Neurosci Rep. 2013;13(5):345.

9. 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(26):2499-508.

400 10. Waitkus MS, Diplas BH, Yan H. Biological Role and Therapeutic Potential of IDH
401 Mutations in Cancer. Cancer Cell. 2018;34(2):186-95.

402 11. Kloosterhof NK, Bralten LBC, Dubbink HJ, French PJ, van den Bent MJ. Isocitrate
403 dehydrogenase-1 mutations: a fundamentally new understanding of diffuse glioma? The Lancet
404 Oncology. 2011;12(1):83-91.

405 12. Turcan S, Rohle D, Goenka A, Walsh LA, Fang F, Yilmaz E, et al. IDH1 mutation is
406 sufficient to establish the glioma hypermethylator phenotype. Nature. 2012;483(7390):479-83.

407 13. Jenkins RB, Blair H, Ballman KV, Giannini C, Arusell RM, Law M, et al. A
408 t(1;19)(q10;p10) mediates the combined deletions of 1p and 19q and predicts a better prognosis
409 of patients with oligodendroglioma. Cancer Res. 2006;66(20):9852-61.

410 14. Cairncross G, Wang M, Shaw E, Jenkins R, Brachman D, Buckner J, et al. Phase III
411 trial of chemoradiotherapy for anaplastic oligodendroglioma: long-term results of RTOG 9402.
412 J Clin Oncol. 2013;31(3):337-43.

413 15. Mur P, Mollejo M, Hernandez-Iglesias T, de Lope AR, Castresana JS, Garcia JF, et al.
414 Molecular classification defines 4 prognostically distinct glioma groups irrespective of
415 diagnosis and grade. J Neuropathol Exp Neurol. 2015;74(3):241-9.

416 16. Molinaro AM, Taylor JW, Wiencke JK, Wrensch MR. Genetic and molecular
417 epidemiology of adult diffuse glioma. Nat Rev Neurol. 2019;15(7):405-17.

418 17. Louis DN, Perry A, Reifenberger G, von Deimling A, Figarella-Branger D, Cavenee
419 WK, et al. The 2016 World Health Organization Classification of Tumors of the Central
420 Nervous System: a summary. Acta Neuropathol. 2016;131(6):803-20.

421 18. Flynn JR, Wang L, Gillespie DL, Stoddard GJ, Reid JK, Owens J, et al. Hypoxia-
422 regulated protein expression, patient characteristics, and preoperative imaging as predictors of
423 survival in adults with glioblastoma multiforme. Cancer. 2008;113(5):1032-42.

424 19. Verhaak RGW, Hoadley KA, Purdom E, Wang V, Qi Y, Wilkerson MD, et al. Integrated
425 Genomic Analysis Identifies Clinically Relevant Subtypes of Glioblastoma Characterized by
426 Abnormalities in PDGFRA, IDH1, EGFR, and NF1. Cancer Cell. 2010;17(1):98-110.

427 20. Wang Q, Hu B, Hu X, Kim H, Squatrito M, Scarpace L, et al. Tumor evolution of glioma
428 intrinsic gene expression subtype associates with immunological changes in the
429 microenvironment. Cancer Cell. 2017;32(1):42-56.

430 21. Gao W-Z, Guo L-M, Xu T-Q, Yin Y-H, Jia F. Identification of a multidimensional
431 transcriptome signature for survival prediction of postoperative glioblastoma multiforme
432 patients. Journal of Translational Medicine. 2018;16(1).

433 22. Hatanpaa KJ, Burma S, Zhao D, Habib AA. Epidermal growth factor receptor in glioma:
434 signal transduction, neuropathology, imaging, and radioresistance. Neoplasia. 2010;12(9):675-
435 84.

436 23. Gan HK, Kaye AH, Luwor RB. The EGFRvIII variant in glioblastoma multiforme. J
437 Clin Neurosci. 2009;16(6):748-54.

438 24. Mukherjee B, McEllin B, Camacho CV, Tomimatsu N, Sirasanagandala S, Nannepaga
439 S, et al. EGFRvIII and DNA Double-Strand Break Repair: A Molecular Mechanism for
440 Radioresistance in Glioblastoma. Cancer Research. 2009;69(10):4252-9.

441 25. Zhang S, Zhao BS, Zhou A, Lin K, Zheng S, Lu Z, et al. m(6)A Demethylase ALKBH5
442 Maintains Tumorigenicity of Glioblastoma Stem-like Cells by Sustaining FOXM1 Expression
443 and Cell Proliferation Program. Cancer Cell. 2017;31(4):591-606 e6.

444 26. Esteller M. Cancer epigenomics: DNA methylomes and histone-modification maps. Nat
445 Rev Genet. 2007;8(4):286-98.

446 27. Park JY, Lee JE, Park JB, Yoo H, Lee SH, Kim JH. Roles of Long Non-Coding RNAs
447 on Tumorigenesis and Glioma Development. Brain Tumor Res Treat. 2014;2(1):1-6.

448 28. Zang L, Kondengaden SM, Che F, Wang L, Heng X. Potential Epigenetic-Based
449 Therapeutic Targets for Glioma. Front Mol Neurosci. 2018;11:408.

450 29. Malzkorn B, Wolter M, Riemenschneider MJ, Reifenberger G. Unraveling the Glioma
451 Epigenome-From Molecular Mechanisms to Novel Biomarkers and Therapeutic Targets. Brain
452 Pathology. 2011;21(6):619-32.

453 30. Binabaj MM, Bahrami A, ShahidSales S, Joodi M, Joudi Mashhad M, Hassanian SM,
454 et al. The prognostic value of MGMT promoter methylation in glioblastoma: A meta-analysis
455 of clinical trials. J Cell Physiol. 2018;233(1):378-86.

456 31. Wick W, Weller M, van den Bent M, Sanson M, Weiler M, von Deimling A, et al.
457 MGMT testing--the challenges for biomarker-based glioma treatment. Nat Rev Neurol.
458 2014;10(7):372-85.

459 32. Rivera AL, Pelloski CE, Gilbert MR, Colman H, De La Cruz C, Sulman EP, et al.
460 MGMT promoter methylation is predictive of response to radiotherapy and prognostic in the
461 absence of adjuvant alkylating chemotherapy for glioblastoma. Neuro Oncol. 2010;12(2):116-
462 21.

463 33. Hegi ME, Diserens AC, Gorlia T, Hamou MF, de Tribolet N, Weller M, et al. MGMT
464 gene silencing and benefit from temozolomide in glioblastoma. N Engl J Med.
465 2005;352(10):997-1003.

466 34. Cancer Genome Atlas Research N, Brat DJ, Verhaak RG, Aldape KD, Yung WK,
467 Salama SR, et al. Comprehensive, Integrative Genomic Analysis of Diffuse Lower-Grade
468 Gliomas. N Engl J Med. 2015;372(26):2481-98.

469 35. Ceccarelli M, Barthel FP, Malta TM, Sabedot TS, Salama SR, Murray BA, et al.
470 Molecular Profiling Reveals Biologically Discrete Subsets and Pathways of Progression in
471 Diffuse Glioma. Cell. 2016;164(3):550-63.

472 36. Zhao Z, Zhang K, Wang Z, Wang K, Liu X, Wu F, et al. A comprehensive review of
473 available omics data resources and molecular profiling for precision glioma studies. Biomed
474 Rep. 2019;10(1):3-9.

475 37. Cao H, Wang F, Li XJ. Future Strategies on Glioma Research: From Big Data to the
476 Clinic. Genomics Proteomics Bioinformatics. 2017;15(4):263-5.

477 38. Yang Y, Sui Y, Xie B, Qu H, Fang X. GliomaDB: A Web Server for Integrating Glioma
478 Omics Data and Interactive Analysis. Genomics Proteomics Bioinformatics. 2019;17(4):465-
479 71.

480 39. John Lonsdale JT, Mike Salvatore, Rebecca Phillips. The Genotype-Tissue Expression
481 (GTEx) project. Nat Genet. 2013;45(6):580-5.

482 40. Puchalski. RB, Shah. N, Miller. J, Dalley. R, Nomura. SR, Yoon. J-G, et al. An anatomic
483 transcriptional atlas of human glioblastoma. Science. 2018;360(6389):660-3.

484 41. Bao ZS, Chen HM, Yang MY, Zhang CB, Yu K, Ye WL, et al. RNA-seq of 272 gliomas
485 revealed a novel, recurrent PTPRZ1-MET fusion transcript in secondary glioblastomas.
486 Genome Res. 2014;24(11):1765-73.

487 42. Sun Y, Zhang W, Chen D, Lv Y, Zheng J, Lilljebjorn H, et al. A glioma classification
488 scheme based on coexpression modules of EGFR and PDGFRA. Proc Natl Acad Sci U S A.
489 2014;111(9):3538-43.

490 43. Yanai I, Benjamin H, Shmoish M, Chalifa-Caspi V, Shklar M, Ophir R, et al. Genome-
491 wide midrange transcription profiles reveal expression level relationships in human tissue
492 specification. Bioinformatics. 2005;21(5):650-9.

493 44. Klee EW. Data mining for biomarker development: a review of tissue specificity
494 analysis. Clin Lab Med. 2008;28(1):127-43, viii.

495 45. Vasmatzis G, Klee EW, Kube DM, Therneau TM, Kosari F. Quantitating tissue
496 specificity of human genes to facilitate biomarker discovery. Bioinformatics.
497 2007;23(11):1348-55.

498 46. Kim P, Park A, Han G, Sun H, Jia P, Zhao Z. TissGDB: tissue-specific gene database in
499 cancer. Nucleic Acids Research. 2018;46(D1):D1031-D8.

500 47. Prassas l, Chrystoja CC, Makawita S, Diamandis EP. Bioinformatic identification of
501 proteins with tissue-specific expression for biomarker discovery. BMC Medicine. 2012;10(39).

502 48. Mohammed A, Biegert G, Adamec J, Helikar T. Identification of potential tissue-
503 specific cancer biomarkers and development of cancer versus normal genomic classifiers.
504 Oncotarget. 2017;8(49):85692-715.

505 49. Sasayama D, Hattori K, Ogawa S, Yokota Y, Matsumura R, Teraishi T, et al. Genome-
506 wide quantitative trait loci mapping of the human cerebrospinal fluid proteome. Hum Mol
507 Genet. 2017;26(1):44-51.

508 50. Gate D, Saligrama N, Leventhal O, Yang AC, Unger MS, Middeldorp J, et al. Clonally
509 expanded CD8 T cells patrol the cerebrospinal fluid in Alzheimer's disease. Nature. 2020.

510 51. Khasawneh AH, Garling RJ, Harris CA. Cerebrospinal fluid circulation: What do we
511 know and how do we know it? Brain Circulation. 2018.

512 52. Sydney C. Gary CAZ, Veronica L. Chiang, Janette U. Gaw, Grace Gray, Susan
513 Hockfield. cDNA cloning, chromosomal localization, and expression analysis of human
514 BEHAB/brevican, a brain specific proteoglycan regulated during cortical development and in
515 glioma. GENE. 2000;256(1-2):139-47.

516 53. Phillips HS, Kharbanda S, Chen R, Forrest WF, Soriano RH, Wu TD, et al. Molecular
517 subclasses of high-grade glioma predict prognosis, delineate a pattern of disease progression,
518 and resemble stages in neurogenesis. Cancer Cell. 2006;9(3):157-73.

519 54. Cook PJ, Thomas R, Kannan R, de Leon ES, Drilon A, Rosenblum MK, et al. Author
520 Correction: Somatic chromosomal engineering identifies BCAN-NTRK1 as a potent glioma
521 driver and therapeutic target. Nat Commun. 2018;9:16187.

522 55. Reed JE, Dunn JR, du Plessis DG, Shaw EJ, Reeves P, Gee AL, et al. Expression of
523 cellular adhesion molecule 'OPCML' is down-regulated in gliomas and other brain tumours.
524 Neuropathol Appl Neurobiol. 2007;33(1):77-85.

525 56. Carminati. PO, Mello. SS, Fachin. AL, Junta. CM, Sandrin-Garcia. P, Carlotti. CG, et
526 al. Alterations in gene expression profiles correlated with cisplatin cytotoxicity in the glioma
527 U343 cell line. Genetics and molecular biology. 2010;33(1):159-68.

528 57. Horst M, Brouwer E, Verwijnen S, Rodijk M, de Jong M, Hoeben R, et al. Targeting
529 malignant gliomas with a glial fibrillary acidic protein (GFAP)-selective oncolytic adenovirus.
530 J Gene Med. 2007;9(12):1071-9.

531 58. van Bodegraven EJ, van Asperen JV, Robe PAJ, Hol EM. Importance of GFAP isoform-
532 specific analyses in astrocytoma. Glia. 2019;67(8):1417-33.

533 59. John S, Sivakumar KC, Mishra R. Bacoside A Induces Tumor Cell Death in Human
534 Glioblastoma Cell Lines through Catastrophic Macropinocytosis. Front Mol Neurosci.
535 2017;10:171.

536 60. Long H, Liang C, Zhang X, Fang L, Wang G, Qi S, et al. Prediction and Analysis of
537 Key Genes in Glioblastoma Based on Bioinformatics. Biomed Res Int. 2017;2017:7653101.

538 61. Shao Y, Chen C, Shen H, He BZ, Yu D, Jiang S, et al. GenTree, an integrated resource
539 for analyzing the evolution and function of primate-specific coding genes. Genome Res.
540 2019;29(4):682-96.

541 62. Saito N, Shirai Y. Protein kinase C gamma (PKC gamma): function of neuron specific
542 isotype. J Biochem. 2002;132(5):683-7.

543 63. Klebe S, Durr A, Rentschler A, Hahn-Barma V, Abele M, Bouslam N, et al. New
544 mutations in protein kinase Cgamma associated with spinocerebellar ataxia type 14. Ann
545 Neurol. 2005;58(5):720-9.

546 64. Yabe I, Sasaki H, Chen DH, Raskind WH, Bird TD, Yamashita I, et al. Spinocerebellar
547 ataxia type 14 caused by a mutation in protein kinase C gamma. Arch Neurol.
548 2003;60(12):1749-51.

549 65. do Carmo A, Balca-Silva J, Matias D, Lopes MC. PKC signaling in glioblastoma.
550 Cancer Biol Ther. 2013;14(4):287-94.

551 66. Reina-Campos M, Diaz-Meco MT, Moscat J. The Dual Roles of the Atypical Protein
552 Kinase Cs in Cancer. Cancer Cell. 2019;36(3):218-35.

553 67. Hu L, Li X, Liu Q, Xu J, Ge H, Wang Z, et al. UBE2S, a novel substrate of Akt1,
554 associates with Ku70 and regulates DNA repair and glioblastoma multiforme resistance to
555 chemotherapy. Oncogene. 2017;36(8):1145-56.

556 68. Jung E, Osswald M, Blaes J, Wiestler B, Sahm F, Schmenger T, et al. Tweety-Homolog
557 1 Drives Brain Colonization of Gliomas. J Neurosci. 2017;37(29):6837-50.

558 69. Catriona M. Dowling SLH, James J. Phelan, Mary Clare Cathcart, Stephen P. Finn,
559 Brian Mehigan, Paul McCormick, John C. Coffey, Jacintha, Kiely OSaPA. Expression of
560 protein kinase C gamma promotes cell migration in colon cancer. Oncotarget.
561 2017;8(42):72096-107.

562 70. Kurscheid S, Bady P, Sciuscio D, Samarzija I, Shay T, Vassallo I, et al. Chromosome 7
563 gain and DNA hypermethylation at the HOXA10 locus are associated with expression of a stem
564 cell related HOX-signature in glioblastoma. Genome Biol. 2015;16:16.

565 71. Fukushima T, Takeshima H, Kataoka H. Anti-glioma therapy with temozolomide and
566 status of the DNA-repair gene MGMT. Anticancer Res. 2009;29(11):4845-54.

567 72. Stupp R, Mason WP, van den Bent MJ, Weller M, Fisher B, Taphoorn MJB, et al.
568 Radiotherapy plus Concomitant and Adjuvant Temozolomide for Glioblastoma. New England
569 Journal of Medicine. 2005;352(10):987-96.

570 73. Li YH, Yu CY, Li XX, Zhang P, Tang J, Yang Q, et al. Therapeutic target database update
571 2018: enriched resource for facilitating bench-to-clinic research of targeted therapeutics.
572 Nucleic Acids Res. 2018;46(D1):D1121-D7.

573 74. Noushmehr H, Weisenberger DJ, Diefes K, Phillips HS, Pujara K, Berman BP, et al.
574 Identification of a CpG island methylator phenotype that defines a distinct subgroup of glioma.
575 Cancer Cell. 2010;17(5):510-22.

576 75. Henrichsen CN, Chaignat E, Reymond A. Copy number variants, diseases and gene
577 expression. Hum Mol Genet. 2009;18(R1):R1-8.

578 76. John CG Spainhour HSL, Soojin V Yi and Peng Qiu. Correlation Patterns Between DNA
579 Methylation and Gene Expression in The Cancer Genome Atlas. Cancer Informatics.
580 2019;18:1176935119828776.

581 77. Han S, Xia J, Qin X, Han S, Wu A. Phosphorylated SATB1 is associated with the
582 progression and prognosis of glioma. Cell Death Dis. 2013;4:e901.

583 78. Liu M, Xu Z, Du Z, Wu B, Jin T, Xu K, et al. The Identification of Key Genes and
584 Pathways in Glioma by Bioinformatics Analysis. J Immunol Res. 2017;2017:1278081.

585 79. Bourgonje AM, Verrijp K, Schepens JTG, Navis AC, Piepers JAF, Palmen CBC, et al.
586 Comprehensive protein tyrosine phosphatase mRNA profiling identifies new regulators in the
587 progression of glioma. Acta Neuropathologica Communications. 2016;4(1).

588 80. Zhang M, Zhao Y, Zhao J, Huang T, Wu Y. Impact of AKAP6 polymorphisms on Glioma
589 susceptibility and prognosis. BMC Neurology. 2019;19(1).

590 81. Wang L, He S, Tu Y, Ji P, Zong J, Zhang J, et al. Downregulation of chromatin
591 remodeling factor CHD5 is associated with a poor prognosis in human glioma. J Clin Neurosci.
592 2013;20(7):958-63.

593 82. Kennedy NJ, Cellurale C, Davis RJ. A radical role for p38 MAPK in tumor initiation.
594 Cancer Cell. 2007;11(2):101-3.

595 83. Wong JC, Fiscus RR. Essential roles of the nitric oxide (no)/cGMP/protein kinase G
596 type-Iα (PKG-Iα) signaling pathway and the atrial natriuretic peptide (ANP)/cGMP/PKG-Iα
597 autocrine loop in promoting proliferation and cell survival of OP9 bone marrow stromal cells.
598 Journal of Cellular Biochemistry. 2011;112(3):829-39.

599 84. Gravendeel LA, Kouwenhoven MC, Gevaert O, de Rooi JJ, Stubbs AP, Duijm JE, et al.
600 Intrinsic gene expression profiles of gliomas are a better predictor of survival than histology.
601 Cancer Res. 2009;69(23):9065-72.

602 85. Sturm D, Witt H, Hovestadt V, Khuong-Quang DA, Jones DT, Konermann C, et al.
603 Hotspot mutations in H3F3A and IDH1 define distinct epigenetic and biological subgroups of
604 glioblastoma. Cancer Cell. 2012;22(4):425-37.

605 86. Sun L, Hui AM, Su Q, Vortmeyer A, Kotliarov Y, Pastorino S, et al. Neuronal and
606 glioma-derived stem cell factor induces angiogenesis within the brain. Cancer Cell.
607 2006;9(4):287-300.

608 87. Griesinger AM, Birks DK, Donson AM, Amani V, Hoffman LM, Waziri A, et al.
609 Characterization of distinct immunophenotypes across pediatric brain tumor types. J Immunol.
610 2013;191(9):4880-8.

611 88. Gill BJ, Pisapia DJ, Malone HR, Goldstein H, Lei L, Sonabend A, et al. MRI-localized
612 biopsies reveal subtype-specific differences in molecular and cellular composition at the
613 margins of glioblastoma. Proc Natl Acad Sci U S A. 2014;111(34):12550-5.

614 89. Bredel M, Bredel C, Juric D, Harsh GR, Vogel H, Recht LD, et al. Functional network
615 analysis reveals extended gliomagenesis pathway maps and three novel MYC-interacting genes
616 in human gliomas. Cancer Res. 2005;65(19):8679-89.

617 90. Bredel M, Bredel C, Juric D, Duran GE, Yu RX, Harsh GR, et al. Tumor necrosis factor-
618 alpha-induced protein 3 as a putative regulator of nuclear factor-kappaB-mediated resistance to
619 O6-alkylating agents in human glioblastomas. J Clin Oncol. 2006;24(2):274-87.

620 91. Yan W, Zhang W, You G, Zhang J, Han L, Bao Z, et al. Molecular classification of
621 gliomas based on whole genome gene expression: a systematic report of 225 samples from the
622 Chinese Glioma Cooperative Group. Neuro-oncology. 2012;14(12):1432-40.

623 92. Lai RK, Chen Y, Guan X, Nousome D, Sharma C, Canoll P, et al. Genome-wide
624 methylation analyses in glioblastoma multiforme. PLoS One. 2014;9(2):e89376.

625 93. Mur P, Mollejo M, Ruano Y, de Lope AR, Fiano C, Garcia JF, et al. Codeletion of 1p
626 and 19q determines distinct gene methylation and expression profiles in IDH-mutated
627 oligodendroglial tumors. Acta Neuropathol. 2013;126(2):277-89.

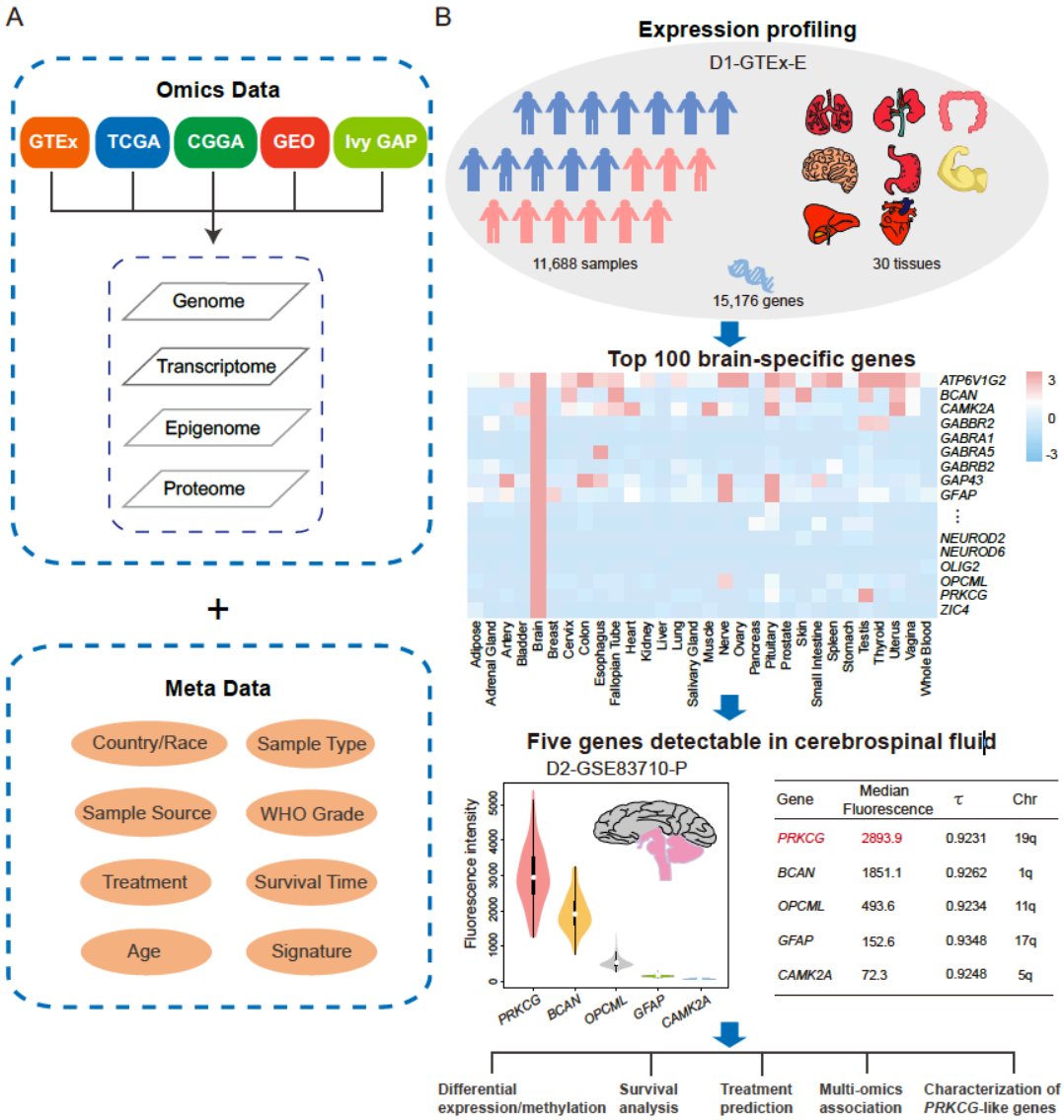
628 94. Zhang W, Yan W, You G, Bao Z, Wang Y, Liu Y, et al. Genome-wide DNA methylation
629 profiling identifies ALDH1A3 promoter methylation as a prognostic predictor in G-CIMP-
630 primary glioblastoma. Cancer Lett. 2013;328(1):120-5.

631

632

633

634 **Figure Legends**



635

636 **Fig 1 Datasets and bioinformatic analysis workflow. (A)** A comprehensive assemble of
637 multi-omics datasets and their corresponding meta data were integrated from GTEx, TCGA,
638 CGGA, GEO and Ivy GAP. **(B)** An integrative analysis workflow was adopted, including
639 detection of brain-specific genes, identification of CSF-detectable genes, ranking of candidate
640 genes in light of protein fluorescence. A series of bioinformatic analyses were performed,
641 including differential expression/methylation analysis, survival analysis, treatment prediction,
642 multi-omics association and characterization of *PRKCG*-like genes.

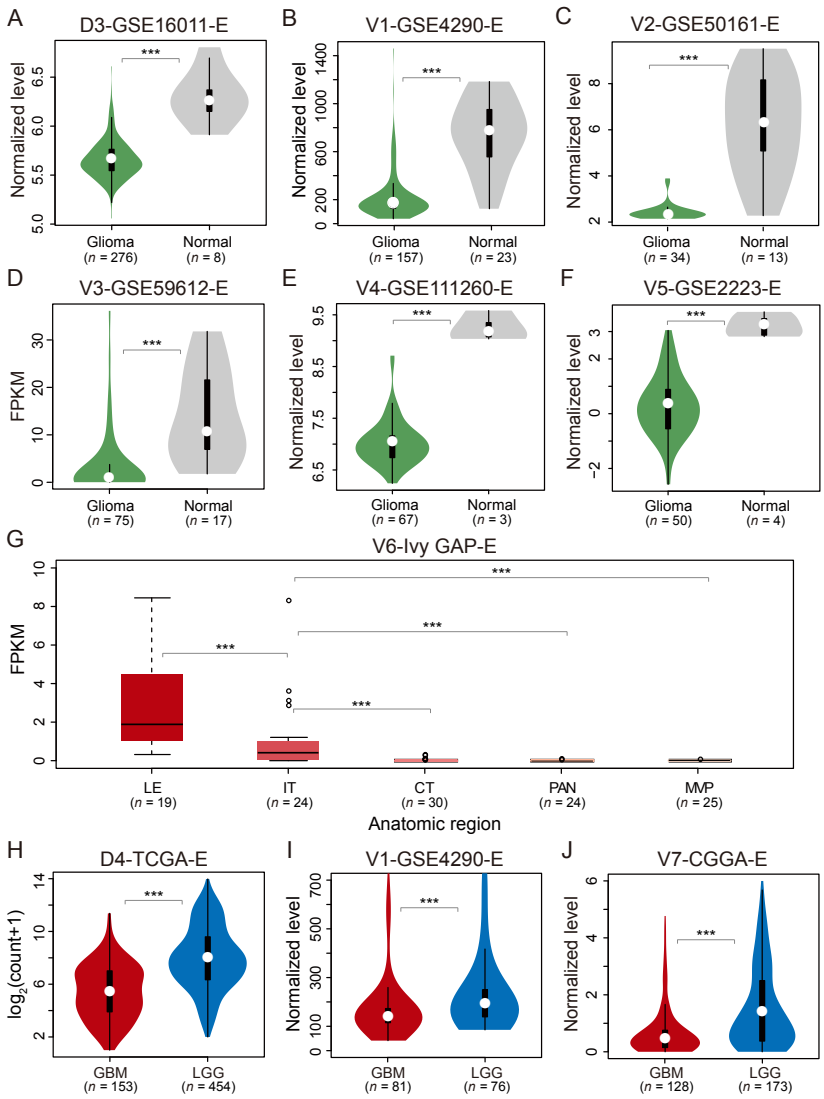
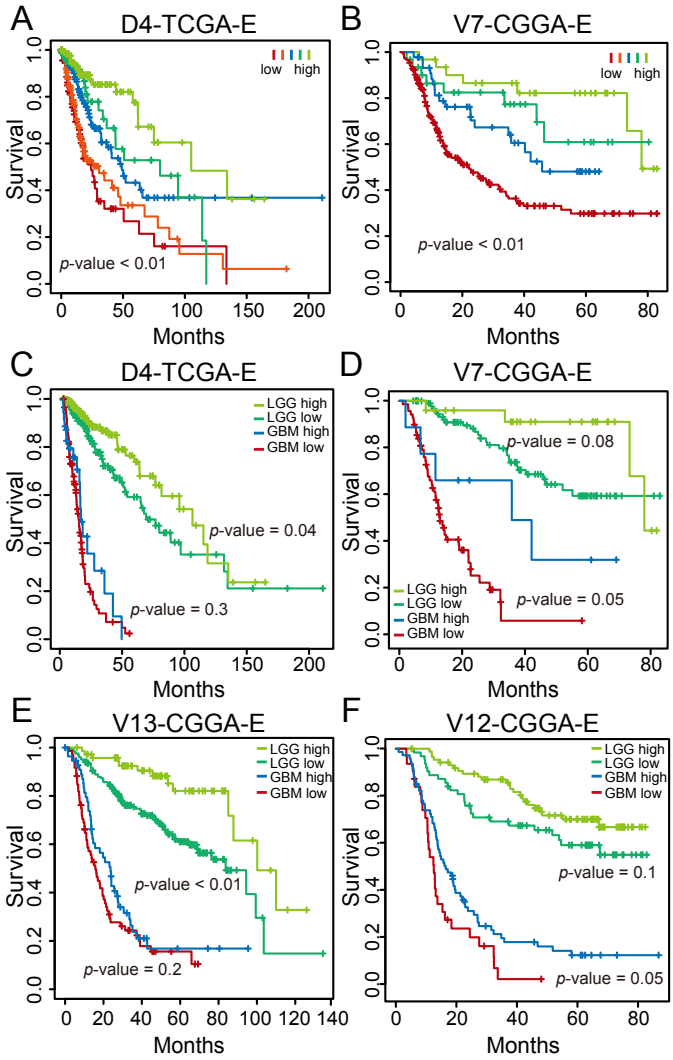


Fig 2 Expression profiles of *PRKCG* across normal, LGG and GBM samples. *PRKCG* expression profiles were compared between glioma and normal samples (D3-GSE16011-E in panel A [RMA normalized], V1-GSE4290-E in panel B [MAS5 normalized], V2-GSE50161-E in panel C [gcRMA normalized], V3-GSE59612-E in panel D, V4-GSE111260-E in panel E [RMA normalized], V5-GSE2223-E in panel F [Lowess normalized]), between different anatomic regions (V6-Ivy GAP-E in panel G), and between GBM and LGG samples (D4-TCGA-E in panel H, V1-GSE4290-E in panel I [MAS5 normalized] and V7-CGGA-E in panel J [Lowess normalized]). All the normalization methods labeled above were derived from and detailed in their corresponding publications, and all these datasets were made publicly

654 accessible at ftp://download.big.ac.cn/glioma_data/. The Wilcoxon tests were performed and

655 the statistical significance levels were coded by: *ns* $p>0.05$, * $p<0.05$, ** $p<0.01$ and ***

656 $p<0.001$.



657

658 **Fig 3 Expression profiles of *PRKCG* associated with survival.** Glioma samples were divided

659 into different groups based on *PRKCG* expression (panels A and B). LGG and GBM samples

660 were divided into two groups with high expression and low expression, respectively (panels C

661 to F). All these datasets can be publicly accessible at ftp://download.big.ac.cn/glioma_data/.

662 The log-rank tests were used to examine the statistical significance between different survival

663 curves.

664

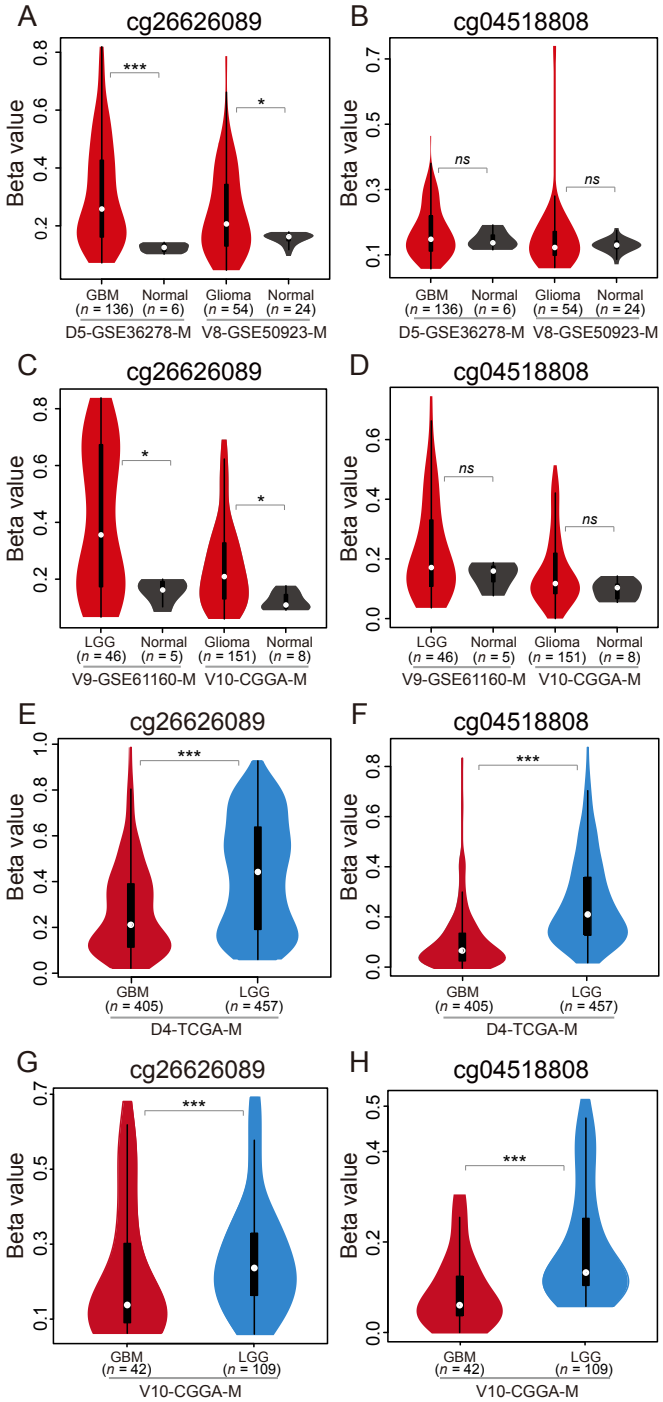
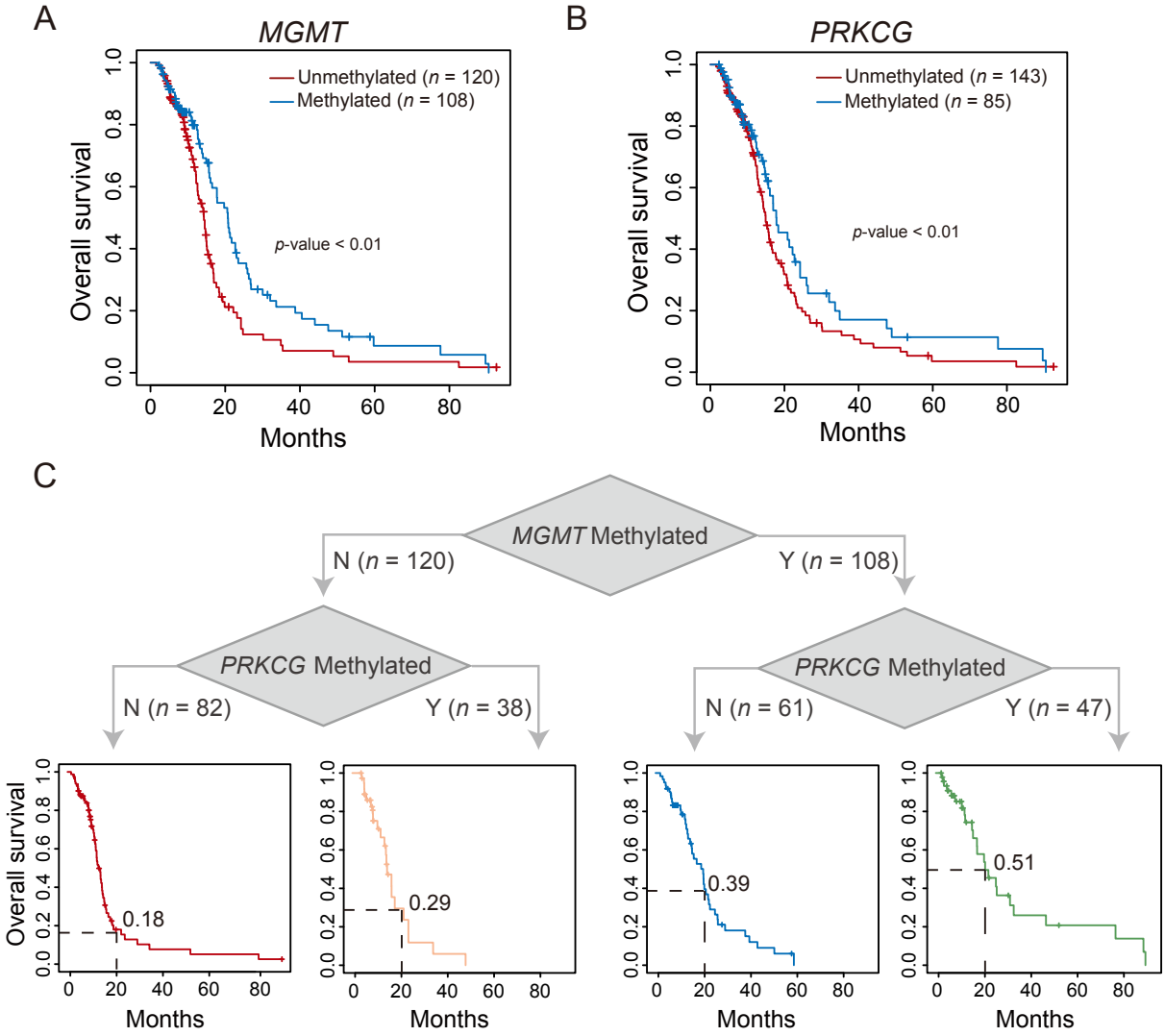


Fig 4 DNA methylation profiles across normal, LGG and GBM samples. *PRKCG* methylation profiles were compared between GBM and normal samples (panels A to D), and between LGG and GBM samples (panels E to H). All these datasets can be publicly accessible at ftp://download.big.ac.cn/glioma_data/. The Wilcoxon tests were used and their statistical significance levels were coded by: *ns* $p>0.05$, * $p<0.05$, ** $p<0.01$ and *** $p<0.001$.

671



672

673 **Fig 5 Combined DNA methylation signatures of *MGMT* and *PRKCG* for treatment**

674 **prediction. (A)** Kaplan-Meier survival curves for GBM patients with TMZ treatment based on

675 *MGMT* methylation. **(B)** Kaplan-Meier survival curves for GBM patients with TMZ treatment

676 based on *PRKCG* (cg26626089) methylation. **(C)** Kaplan-Meier survival curves for GBM

677 patients with TMZ treatment based on *MGMT* and *PRKCG* combined methylation signatures.

678

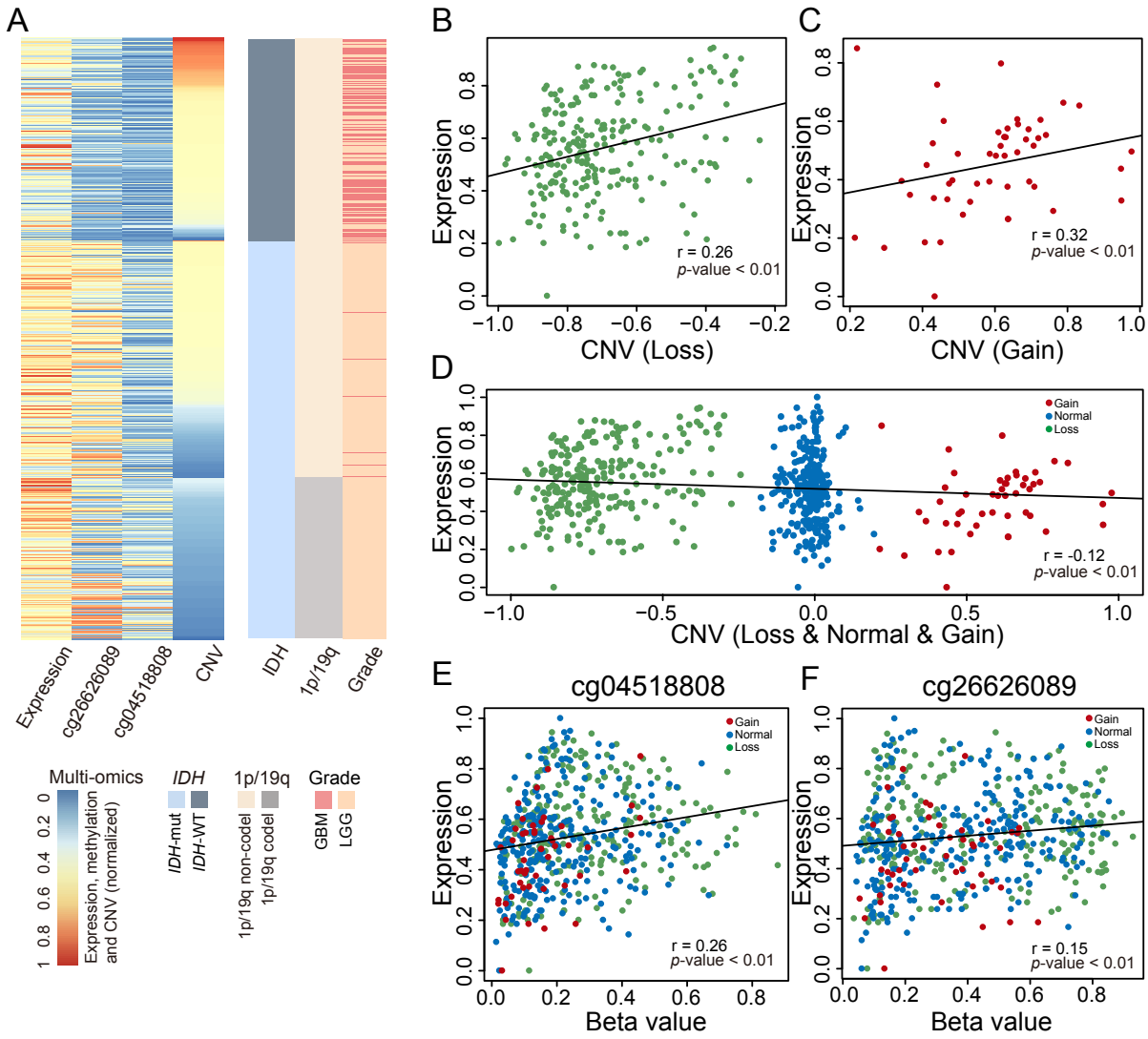
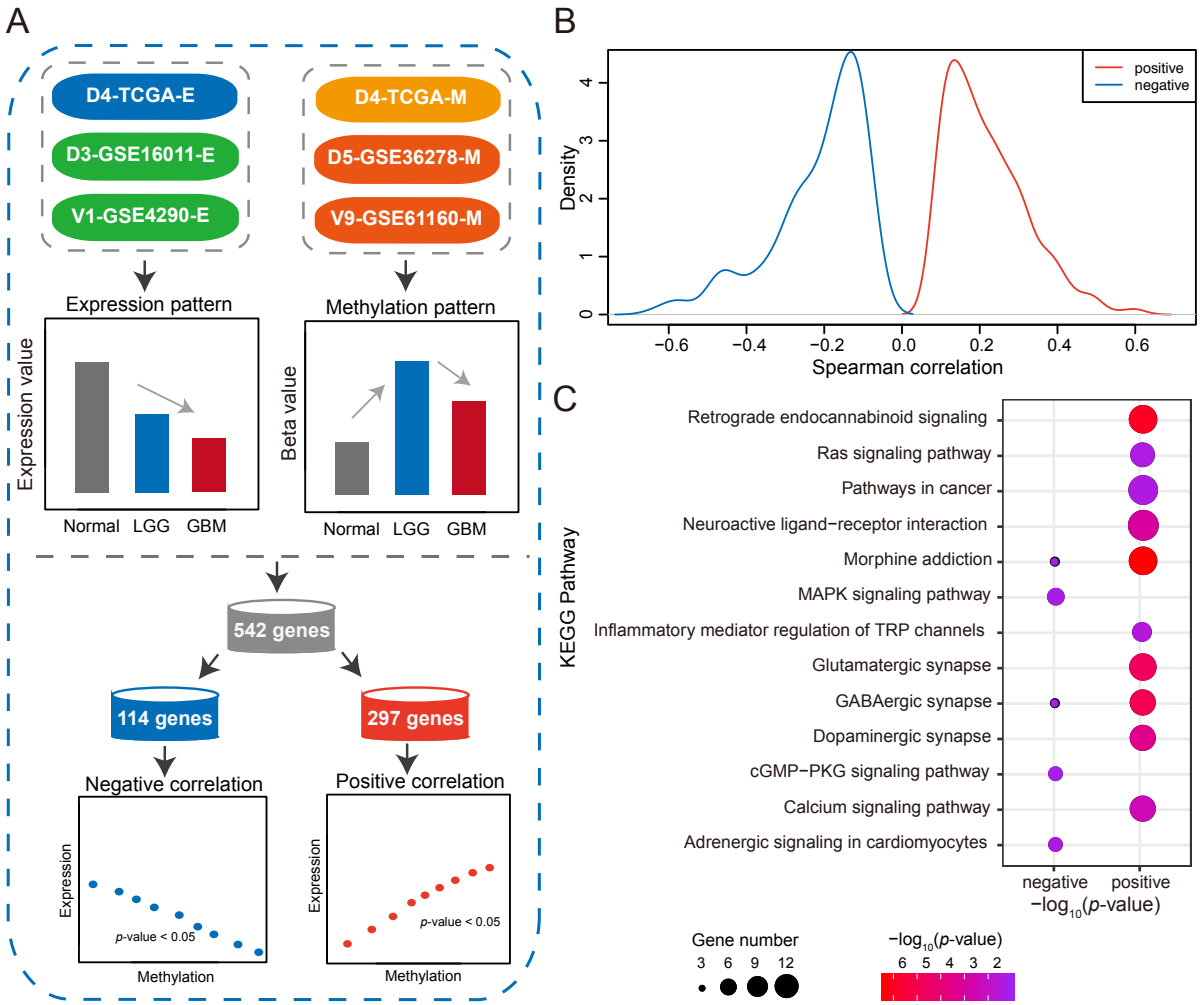


Fig 6 Multi-omics molecular profiles of *PRKCG*. (A) Association of *PRKCG*'s multi-omics signatures with *IDH*, 1p/19q status and WHO grade. **(B)** Correlation between *PRKCG* expression and CNV Loss. (**C**) Correlation between *PRKCG* expression and CNV Gain. (**D**) Correlation between *PRKCG* expression and all CNV status. **(E)** Correlation between *PRKCG* expression and DNA methylation of the CpG site cg04518808. **(F)** Correlation between *PRKCG* expression and DNA methylation of the CpG site cg26626089.



690

691 **Fig 7 Identification and characterization of *PRKCG*-like genes. (A)** Identification of

692 *PRKCG*-like genes, including two types of genes that possess negative and positive correlations

693 between expression and DNA methylation, respectively. **(B)** Spearman correlations of two

694 types of *PRKCG*-like genes. **(C)** The KEGG pathway enrichment of two types of *PRKCG*-like

695 genes.

696

697

698 **Tables**

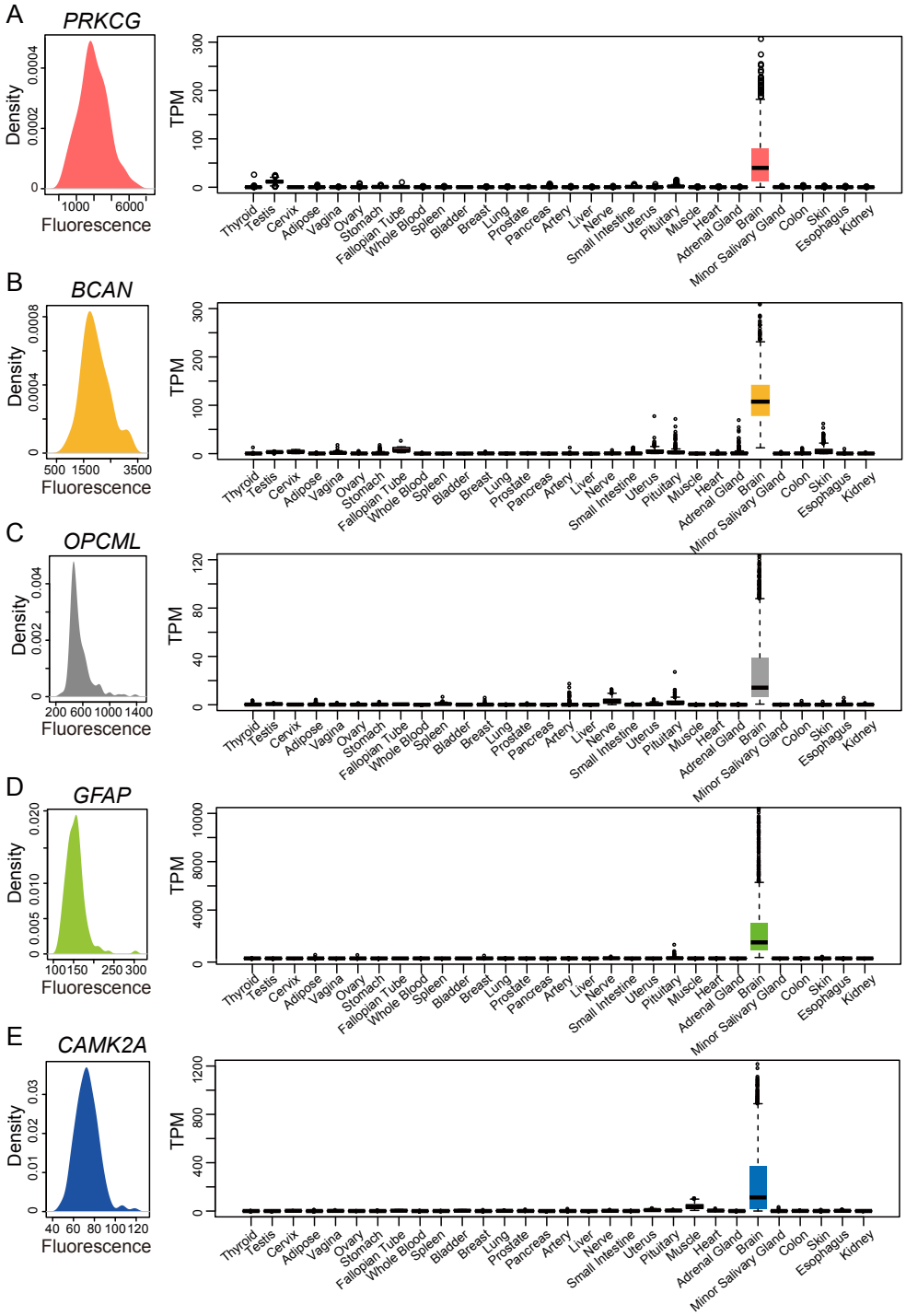
Table 1. Summary of multi-omics multi-cohort glioma datasets

Category	Accession number	Source	Omics type	# Samples	# Population country/race	Reference
Discovery	D1-GTEx-E	GTEx	Expression (RNA-Seq)	11,688	mostly white	(39)
	D2-GSE83710-P	GSE83710	Protein	133	Japan	(49)
	D3-GSE16011-E	GSE16011	Expression (Microarray)	284	Netherlands	(84)
	D4-TCGA-V D4-TCGA-E D4-TCGA-M D4-TCGA-M (TMZ treatment)	TCGA	CNV Expression (RNA-Seq) Methylation (27K+450K) Methylation (27K+450K)	1,018 607 862 228	mostly white	(35)
	D5-GSE36278-M	GSE36278	Methylation (450K)	142	Germany	(85)
Validation	V1-GSE4290-E	GSE4290	Expression (Microarray)	180	USA	(86)
	V2-GSE50161-E	GSE50161	Expression (Microarray)	47	USA	(87)
	V3-GSE59612-E	GSE59612	Expression (RNA-Seq)	92	USA	(88)
	V4-GSE111260-E	GSE111260	Expression (Microarray)	70	Norway	-
	V5-GSE2223-E	GSE2223	Expression (Microarray)	54	USA	(89, 90)
	V6-Ivy GAP-E	Ivy GAP	Expression (RNA-Seq)	122	unknown	(40)
	V7-CGGA-E	CGGA	Expression (Microarray)	301	China	(42, 91)
	V8-GSE50923-M	GSE50923	Methylation (27K)	78	USA	(92)
	V9-GSE61160-M	GSE61160	Methylation (450K)	51	Spain	(93)
	V10-CGGA-M	CGGA	Methylation (27K)	159	China	(94)
	V11-TCGA-M	TCGA	Methylation (WGBS)	6	white	-
	V12-CGGA-E	CGGA	Expression (RNA-Seq)	310	China	(41)
	V13-CGGA-E	CGGA	Expression (RNA-Seq)	667	China	-

V14-GSE60274-M	GSE60274	Methylation (450K)	68	Switzerland	(70)

Note:
CGGA: Chinese Glioma Genome Atlas, http://www.cgga.org.cn
GEO: Gene Expression Omnibus, https://www.ncbi.nlm.nih.gov/geo/
GTEx: Genotype-Tissue Expression, https://www.gtexportal.org
TCGA: The Cancer Genome Atlas, https://portal.gdc.cancer.gov
Ivy GAP: Ivy Glioblastoma Atlas Project, http://glioblastoma.alleninstitute.org

711 **Supporting information captions**

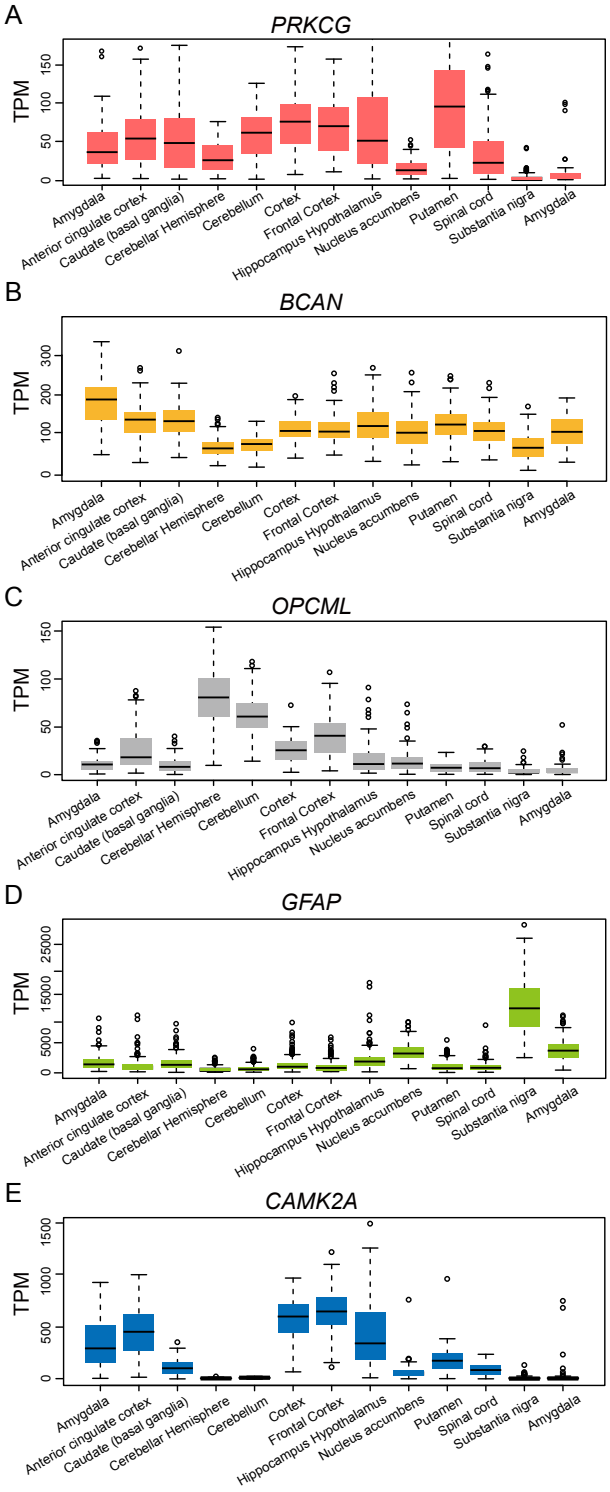


712

713 **S1 Fig. Protein level distributions in CSF and RNA expression profiles of *PRKCG* (A),**

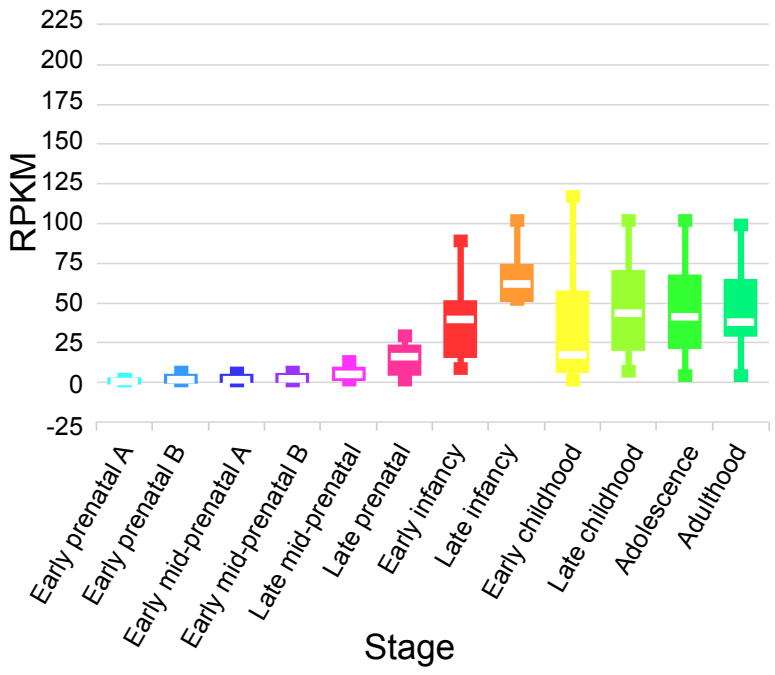
714 ***BCAN* (B), *OPCML* (C), *GFAP* (D), *CAMK2A* (E) across 30 normal human tissues.**

715



S2 Fig. Expression profiles of *PRKCG* (A), *BCAN* (B), *OPCML* (C), *GFAP* (D), and *CAMK2A* (E) across 13 human brain regions.

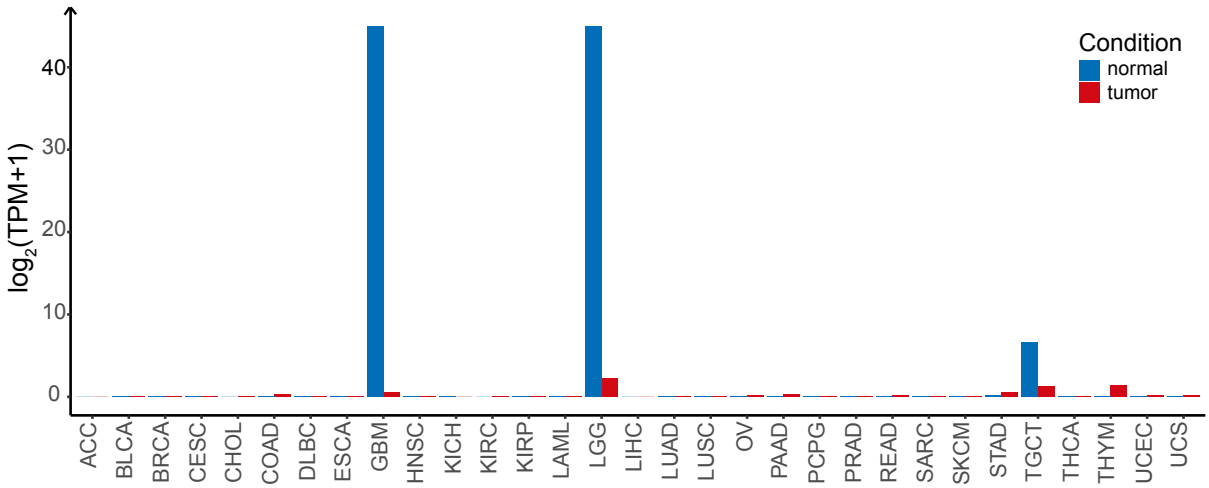
720



721

722 **S3 Fig. Expression profiles of *PRKCG* in brain developmental stages.**

723

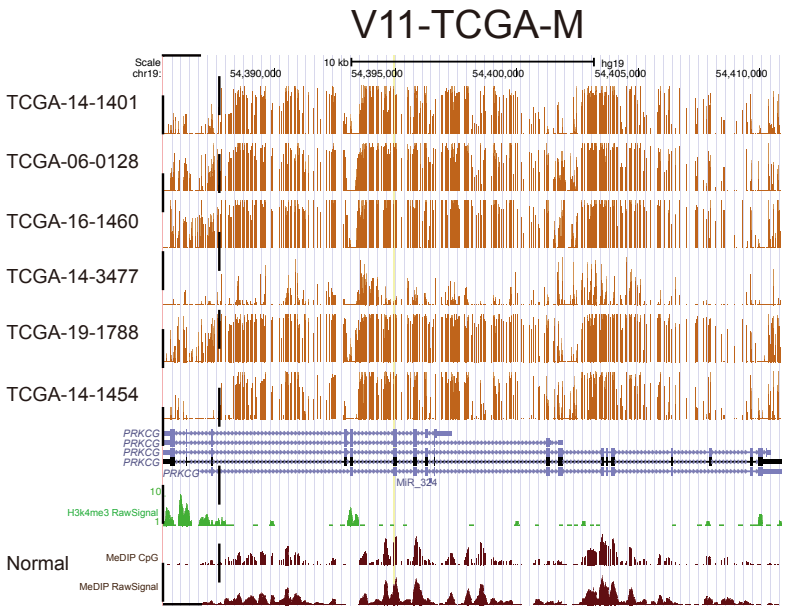


724

725 **S4 Fig. Expression profiles of *PRKCG* across 31 human tumor and normal tissues.**

726

727



728

729 **S5 Fig. Bisulfite DNA methylation profiles of *PRKCG* across six GBM samples and one**

730 **normal sample.**

731

732

733

734

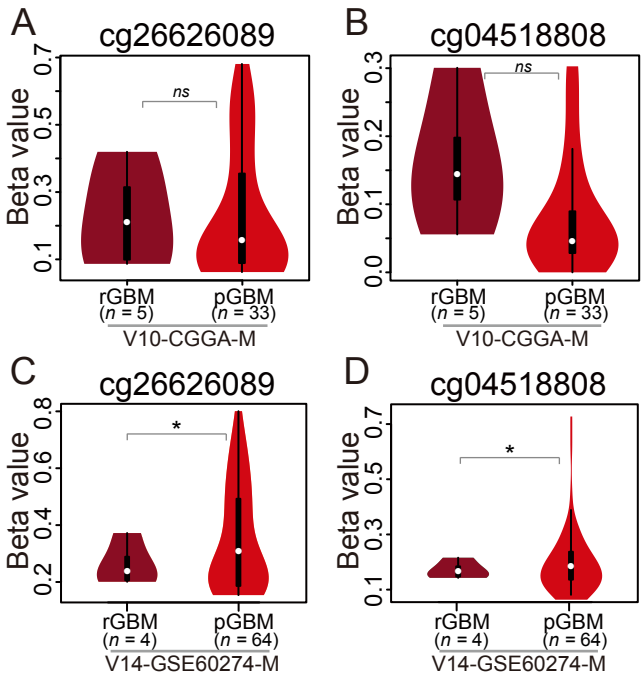
735

736

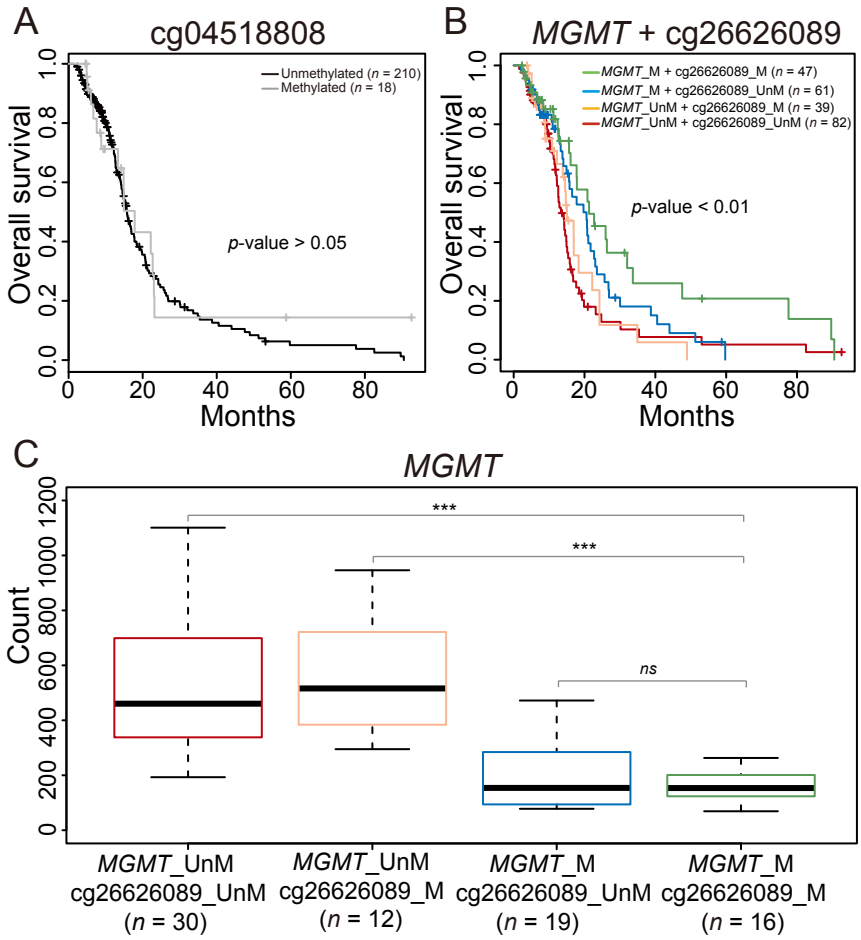
737

738

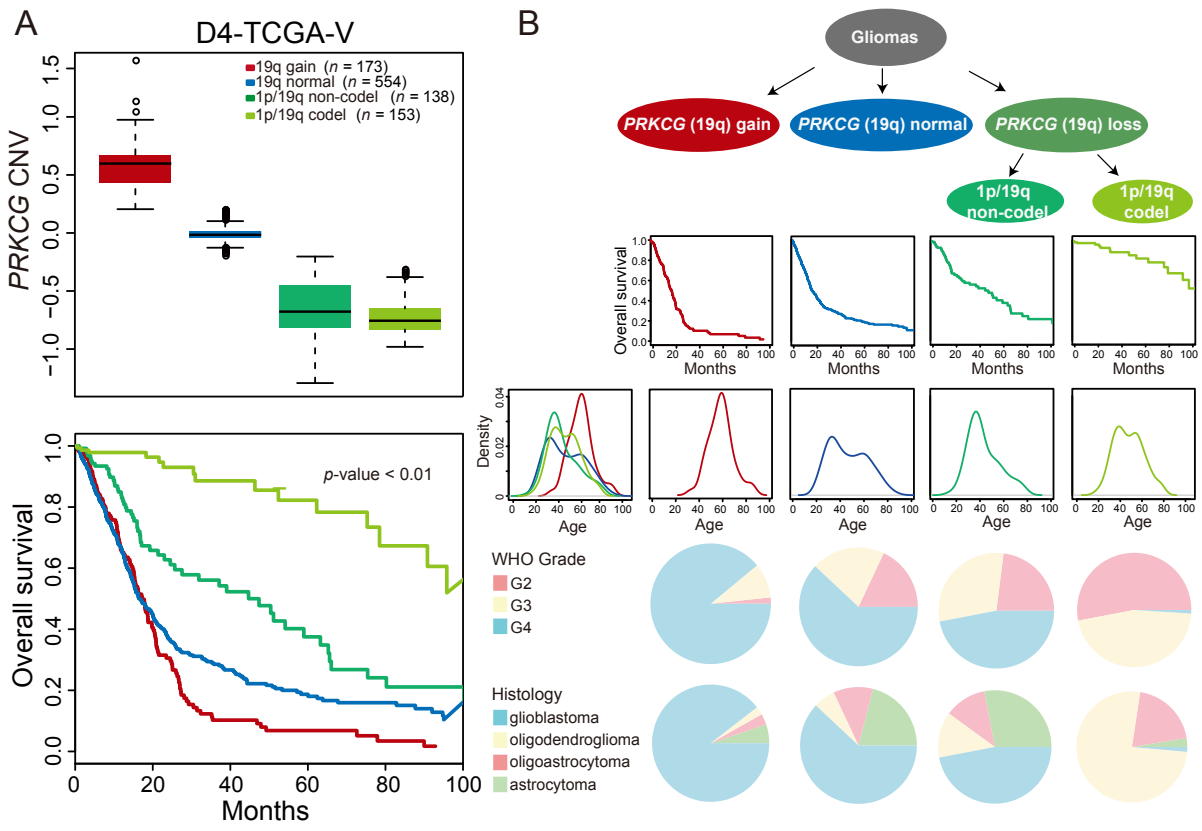
739



S6 Fig. DNA methylation profiles of *PRKCG* in recurrent GBM (rGBM) and primary GBM (pGBM) samples. *PRKCG* methylation profiles were compared between rGBM and pGBM samples (V10-CGGA-M in panels A and B and V14-GSE60274-M in panels C and D). All these datasets can be publicly accessible at ftp://download.big.ac.cn/glioma_data/. The Wilcoxon tests were used and the statistical significance levels were coded by: *ns* $p>0.05$, * $p<0.05$, ** $p<0.01$ and *** $p<0.001$.



S7 Fig. Predictive potential of *PRKCG* DNA methylation. (**A**) Kaplan-Meier survival curves for GBM patients with TMZ treatment based on *PRKCG* (cg04518808) methylation. (**B**) Methylation site cg26626089 in combination with *MGMT*, which were used to classify GBM patients into four groups. (**C**) Expression profiles of *MGMT* in the four groups.



S8 Fig. *PRKCG* CNV associated with survival. (A) Four groups of glioma samples were divided based on the 1p/19q status (19q gain, 19q normal, 1p/19q non-codel, and 1p/19q codel). **(B)** Kaplan-Meier survival probability, age, WHO grade and histology of the four groups.

S1 Table. 15,176 genes' RNA expression levels across 30 normal human tissues.

S2 Table. Tissue specificity τ values and maximum expression levels of the top 100 brain-specific protein-coding genes.

S3 Table. *PRKCG*-like genes and their annotations.