Supplementary Information for

Elevated signature of a gene module co-expressed with CDC20 marks genomic instability in glioma

Yunqiu Zhang[a], Jiuyi Li[a], Kaikai Yi[b,c], Jing Feng[a], Zhengmin Cong[a], Zheng Wang[d,e], Yanfei Wei[a], Fan Wu[d,e], Wen Cheng[f], Ayaz Ali Samo[a], Paolo Salomoni[g], Qiong Yang[a], Yu Huang[h], Chunsheng Kang[b,c,i,1], Tao Jiang[d,e,i,1], and Xiaolong Fan[a,i,1]

[a]Beijing Key Laboratory of Gene Resource and Molecular Development, Laboratory of Neuroscience and Brain Development, Beijing Normal University, 100875 Beijing, China;
[b]Department of Neurosurgery, Tianjin Medical University General Hospital, 300052 Tianjin, China;
[c]Laboratory of Neuro-Oncology, Tianjin Neurological Institute, Department of Neurosurgery, Tianjin Medical University General Hospital and Key Laboratory of Neurotrauma, Variation, and Regeneration, Ministry of Education and Tianjin Municipal Government, 300052 Tianjin, China;
[d]Department of Neurosurgery, Beijing Tiantan Hospital, Capital Medical University, 100050 Beijing, China;
[e]Beijing Neurosurgical Institute, 100050 Beijing, China;
[f]Department of Neurosurgery, The First Hospital of China Medical University, 110001 Shenyang, China;
[g]German Center for Neurodegenerative Diseases, 53127 Bonn, Germany;
[h]Department of Medical Genetics, School of Basic Medical Sciences, Peking University Health Science Center, 100191 Beijing, China;
[i]Chinese Glioma Genome Atlas Network and Asian Glioma Genome Atlas Network.

[1]Correspondence:
xfan@bnu.edu.cn (X.F.); taojiang1964@163.com (T.J.); kang97061@tmu.edu.cn (C.K.).

This PDF file includes:

Supplementary information text
Figs. S1 to S19
Tables S1 to S11
References for SI reference citations

www.pnas.org/cgi/doi/10.1073/pnas.1814060116

Supplementary Information Text

Materials and Methods

Datasets used in this report

Transcriptome data of gliomas (GSE4290 (1), GSE16011 (2) and the Rembrandt (3)), human brain development (GSE25219 (4)) and cell types (GSE65000 (5)), neural cell types of mouse brain (GSE9566 (6)), human hematopoiesis (GSE24759 (7)), mouse hematopoiesis (GSE77098 (8)) were downloaded from Gene Expression Omnibus (https://www.ncbi.nlm.nih.gov/geo/) as of November 2016. Transcriptomes of *C. elegans* were obtained from the study by Kim et al (9). Transcriptome data of different development stages from early embryo to adult *Drosophila* were obtained from the study by Janic et al (10). The data of mRNA-seq, exome-seq, SNP6.0 analysis and the survival data of gliomas were obtained from the TCGA data portal. The data of 381 glioma samples released in 2013 constituted the training cohort, and the more recently released data from 301 grade II-III glioma samples constituted the validation cohort. The CGGA dataset was generated using 319 glioma samples collected at the Department of Neurosurgery, Beijing Tiantan Hospital, Capital Medical University in Beijing from patients treated between 2006 and 2009 (11). Written consent was obtained from all patients.

Identification of CDC20-M and CREBRF-M

Qlucore Omics Explorer 3.2 (Qlucore AB, Lund, Sweden) was used for gene co-expression module construction and differential gene expression between the glioma subtypes. The expression values were log2-transformed in all analyses. Using Pearson correlation coefficient (PCC) analysis in the GSE4290 dataset, 185 probe sets (139 genes) with their expression patterns highly correlated to CDC20 (202870_s_at) (at PCC not less than 70%) were identified as the CDC20-M. Protein-protein association network of CDC20-M shown in Fig 1A was generated using cytoscape (12). The genes shown in the network were filtered with a combined score > 0.99. The combined score was computed by combining the probabilities from the different evidence channels and corrected for the probability of randomly observing an interaction (13).

CREBRF expression was found to be consistently anti-correlated to CDC20 in all glioma datasets. CREBRF-M was constructed in GSE4290 including 157 probe sets (120 genes) at a PCC to CREBRF (235556_at) not less than 70%.

Stable and robust clustering of gliomas into subgroups with high or low expression of CDC20-M or CIN70

Using mRNA-seq data from the TCGA training cohort, we performed unsupervised consensus clustering based on partitioning around medoid (PAM) (14) to define stable and robust clusters, which was implemented in R package ConsensusClusterPlus. The expression values of the CDC20-M and CREBRF-M members were first log2 transformed and z-score standardized. Euclidean distance was also integrated into the clustering

analysis. The PAM clusters were iterated for 1000 times by random selection of 90% samples in each round. The resulting clusters for a given number of K were summarized in a consensus matrix. The K tested ranged from 2-10. The reasonable K was determined by the "cleanest" consensus matrix plot, and the best bimodal shape of CDF curve (close to 0 or 1). Unsupervised consensus clustering was also performed with CIN70 at K = 2.

Single sample prediction (SSP) of samples into the subtypes with distinct CDC20-M and CREBRF-M expression or with CIN70 expression

The mean expression profiles (the centroids) for CDC20-M and CREBRF-M members were calculated for the four subgroups of the TCGA training cohort shown in Fig. 2B. Samples in the TCGA validation cohort and the CGGA data set were then assigned to the nearest subtype/centroid using Spearman correlation. Chromosomal abnormalities, mutation burdens and survival analysis were subsequently performed in the individually assigned samples. SSP was also performed for the TCGA validation and CGGA datasets according to CIN70 signature-based clustering as shown in Fig. S8.

Survival analysis

Overall survival time was calculated from the date of surgery until death or the last follow-up contact. Kaplan-Meier survival curves were generated and analyzed with log-rank test using Prism 5.0 software.

Cox regression analysis

Cox regression analysis was performed with SPSS software, with survival time as the

dependent variable. CDC20-M score for each glioma sample was calculated as the average of normalized expression levels for CDC20-M members, and used as variables in multivariate Cox regression analysis.

Analysis of somatic copy number alterations and mutations

The level III SNP6.0 data of gliomas was downloaded from TCGA. The normalized data from cancer samples and their somatic controls were grouped according to the CDC20-M classification, and processed using NEXUS COPY NUMBER™ 8.0 (Biodiscovery Inc.). The implemented call algorithm was SNP-FASST2 (Fast Adaptive States Segmentation Technique) segmentation. It first calculated the log2 ratio of each probe across the whole genome and then arranged the ratios according to the probe position along the chromosome for each sample. The FASST2 algorithm was used to segment the genomes into regions of uniform ratios. Finally, the copy numbers of each region were inferred according to their log2 ratios. The log2 ratio threshold for gain, loss, high level gain and homozygous loss were 0.25, -0.2, 0.7 and -1.1, respectively. The gain and loss refer to single copy gain and single copy loss, respectively. High level gain indicates gain of two or more copies.

GISTIC2.0 (15) was used to analyze the arm-level and focal SCNAs of glioma samples. We used a noise threshold of 0.3, a broad length cutoff of 0.5 chromosome arms, a confidence level of 95% and a copy-ratio cap of 1.5. Copy number alterations of CDC20-M members were also analyzed by GISTIC2.0 under the same amplitude threshold as above. Similarly, we also analyzed the SCNAs and mutations in the core members of the TP53 pathway.

Mutation data were also downloaded from TCGA, the workflow for mutation detection

was MuTect2 (16). We analyzed the frequencies of Indels/frameshift mutations, synonymous mutations and non-synonymous mutations according to the annotation downloaded from TCGA portal.

Outlier samples with extreme data of SCNAs or mutations were excluded using a boxplot method (17, 18).

Gene set enrichment analysis (GSEA)

The enrichment of gene sets involved in DNA damage response across CDC20-M/CREBRF-M classified glioma subtypes was analyzed using GSEA (http://software.broadinstitute.org/gsea/index.jsp), which is a computational method determining whether an a priori defined set of genes shows statistically significant, concordant differences between two biological states (19). Gene sets were downloaded from http://software.broadinstitute.org/gsea/msigdb/genesets.jsp?collection=BP#. All parameters were set as default.

Gene dosage-dependent analysis

The gene dosage-dependent expression of the core members of CDC20-M (CDC20, AURKA, KIF2C, CENPF, CCNB1 and CCNB2) was analyzed among the gliomas with high or intermediate CDC20-M expression in the TCGA training cohort. Spearman's rank correlation analysis between the data of gene expression and gene dosage was performed.

Cell culture conditions

All cells were maintained at 37℃ in a humidified 5% CO_2 atmosphere. The patient

information of each cell line is listed in Table S9. The neuroblastoma cell line SK-N-SH (SK) was obtained from the National Infrastructure of Cell Line Resource (#3111C0001CCC000086). N5, N9, N33, PDX and SK cells were cultured in DMEM/F12 (Life Technologies, #11330-032) supplemented with 1% Penicillin-Streptomycin (Life Technologies, #15140122) and 10% FBS (Gemini, #900-108). N3 and N8 cells were cultured in DMEM/F12 (Life Technologies, #11330-032) supplemented with 1% Penicillin-Streptomycin (Life Technologies, #15140122), 1% B-27 (Life Technologies, #17504044) and 10% FBS (Gemini, #900-108). Human astrocytes (HA) (ScienCell, #1800) were cultured according to the manufacturer's protocol in astrocyte medium (ScienCell, #1801) containing 5% FBS, astrocyte growth supplement and 1% penicillin/streptomycin.

Real-time quantitative PCR

Total RNA was isolated using the RNeasy Plus Mini Kit (QIAGEN, #74134). Reverse transcription was performed with High Capacity RNA-to-cDNA Kit (Applied Biosystems, #4387406). TaqMan™ Universal Master Mix II, with UNG (Applied Biosystems, #4440038) was used for amplification using the Applied Biosystems 7500/7500 Fast Real-Time PCR Systems (Applied Biosystems). All samples were run in triplicate in parallel with UBC as the control. Using the $\Delta\Delta C_T$ method, relative transcript copy number for each transcript was normalized to UBC. The Taqman primer sets used for each transcript were: human CDC20 (CDC20 TaqMan® Gene Expression Assays, #Hs00426680_mH), human AURKA (AURKA TaqMan® Gene Expression Assays, #Hs01582072_m1), human KIF2C (KIF2C TaqMan® Gene Expression Assays, #Hs00901710_m1), and human UBC (UBC TaqMan® Gene Expression Assays, #Hs01871556_s1).

Identification of FGFR3-TACC3 fusion transcript

pLOC-FGFR3-TACC3 plasmid (a kind gift from Dr. Antonio Iavarone) was used as the positive control for detection of FGFR3-TACC3 fusion. cDNAs of HA, N5, N9, N33, PDX, N3, N8 and SK were generated as described above. Primer pair used for the FGFR3-TACC3 fusion screening was: FGFR3ex13-FW: 5'- GGCATGGAGTACTTGGCCTC - 3' and TACC3ex9-RV: 5'- GACGTCCTGAGGGAGTCTCA - 3', amplification conditions were 95℃-30 seconds/60℃-30 seconds/68℃-1 minute for 35 cycles, 68℃-5 minutes.

Cell survival and proliferation assay

2000 cells per well were plated in 96-well plates in a total volume of 100 μl medium described above. Each concentration of drug treatment was tested in six replicates. Fresh medium with or without drug was added every third day. Cell growth was measured using alamarBlue according to manufacturer's instruction (Thermo Fisher Scientific, #DAL1025). Briefly, cells were incubated with alamarBlue reagent (10 μl/100 μl medium) for 4 h at 37℃. Cell growth was then measured using a fluorescence plate reader with an excitation wavelength of 530 nm and an emission wavelength of 580 nm. The relative numbers of viable cells were expressed as the percentage of untreated cells.

Karyotype analysis

When the cultures in T75 flasks reached a confluence of 80%, Colcemid® (10 mg/mL) was added to a final concentration of 0.1 mg/mL. Following an incubation at 37°C for 4 h, changes in cell morphology were monitored using an inverted microscope. The cells were

detached using 1 mL of TrypLE™ Express (Life Technologies, #12605-010) for 5 min at room temperature. Then the cells were pelleted at 300 × g for 5 min. For the hypotonic treatment, 10 mL of pre-warmed 0.075 M KCl was slowly and carefully added, followed by an incubation at 37℃ for 15 min and fixation using methanol:acetic acid (3:1) solution. Prior to slide preparation, clean slides were placed in steam while the cells were resuspended in fresh fixative solution and dropped onto the surface of the slide. In order to obtain G-bands, the slides were aged at 70℃ for 3 h, and subsequently immersed in trypsin solution (0.2%, PH = 7.2) for 6 s, washed in saline solution and finally quickly rinsed in distilled water. The staining procedure was carried out using Giemsa (1:20), producing trypsin and Giemsa (GTG) bands. The band quality was evaluated under the microscope at a magnification of 100x and the lengths of trypsin treatment and Giemsa staining were adjusted in order to produce well stained bands.

Immunofluorescence staining

10^5 cells per well were plated on coverslips in a 6-well plate for 24 h. The cells were fixed with 4% paraformaldehyde for 15 min, washed 2 × 5 min with phosphate-buffered saline (PBS). Cells were permeabilized with 0.1% Triton X-100 for 5 min, washed 2 × 5 min with PBS and blocked with 5% fetal bovine serum (FBS) for 1 h. Cells were then incubated with one of the following primary antibodies at 4℃ overnight: mouse anti-α-tubulin (Sigma, #T5168, clone ID: B-5-1-2, at 1:2000 dilution in 5% FBS), rabbit anti-γ-tubulin (Sigma, #T3320, at 1:1000 dilution), rabbit anti-CENPB (abcam, #ab25734, at 1:100 dilution), or mouse anti-p-H2A.X (Thermo Fisher Scientific, #14-9865-82, clone ID: CR55T33, at 1:100 dilution). Following 3 × 5 min washing with 5% FBS, the coverslips were incubated

with two of the following secondary antibodies: Alexa Fluor anti-mouse 647 (for α-tubulin primary antibody), Alexa Fluor anti-rabbit 488 (for γ-tubulin or CENPB primary antibody), Alexa Fluor anti-mouse 488 (for α-tubulin or p-H2A.X primary antibody) or Alexa Fluor anti-rabbit 568 (for γ-tubulin or CENPB primary antibody) (all from Life Technologies, at 1:100 dilution) for 1 h. Finally, coverslips were washed 3 × 5 min with 5% PBS, and mounted with Antifade Mounting Medium with DAPI (Vector laboratories, #H-1200) according to the manufacturer's protocol.

The coverslips were examined under a Zeiss LSM700 confocal microscope. The images in Fig. 4C a and g panels are projection images of confocal image stacks of 1 μm interval.

Hematoxylin and eosin (HE) staining and immunohistochemical staining

Glioma specimens were fixed in 4% paraformaldehyde for 24 h and stored in 70% ethanol until paraffin embedding. Embedded tissue was processed to 5 μm sections and stained with hematoxylin and eosin (HE).

Immunohistochemical staining for the relevant proteins was performed on deparaffinized sections, rehydrated and treated in 10 mM citrate buffer (100℃, 10 min) for antigen repair. Subsequently, the sections were immersed in ethanol containing 3% hydrogen peroxide for 10 min to block endogenous peroxidase activity. The sections were incubated overnight at 4℃ with one of the following primary antibodies: rabbit anti-p-AURKA (Thermo Fisher Scientific, #44-1210G, at 1:100 dilution), mouse anti-p-H2A.X (Thermo Fisher Scientific, #14-9865-82, clone ID: CR55T33, at 5 μg/ml), and rabbit anti-CDC20 (Sigma, #HPA055288, at 1:50 dilution), followed by 3 × 5 min washing in PBS and incubation with the secondary anti-rabbit or anti-mouse antibodies conjugated with peroxidase. The images

were captured with Axio Imager 2 (Zeiss) after 3,3'-diaminobenzidine staining.

Grading of Ki-67 staining intensity

The Ki67 staining results were obtained from the Department of Pathology at Beijing Tiantan Hospital, Capital Medical University, with rabbit anti-Ki-67 mAb (ZSGB Bio, Cat# ZA-0502) as the primary antibody. The staining was assessed by observing 10 randomly selected fields, 100 cells were observed in each field. Samples containing <5%, 5%-30%, 30%-50% and >50% Ki67-positive cells were considered as intensity grades 0, 1, 2 and 3, respectively.

Analysis of cell cycle and MLN8237 treatment-induced cell polyploidization

The cell fraction in the S phase and quantification of MLN8237-induced polyploidization (with > 4N DNA content) were assessed in all glioma cell lines used and also in the HA cells. For detection of the cell fraction in the S phase, Click-iT™ EdU Alexa Fluor™ 488 Flow Cytometry Assay Kit (Invitrogen, #C10420) was employed. Briefly, following treatment with MLN8237 or DMSO for 72 h, cells were incubated with EdU at a concentration of 10 μM for 40 minutes. Cells were then dissociated, detection of EdU was subsequently performed according to the manufacturer's instructions. DNA was stained with Hoechst 33342 dye (Life Technologies, #H3570). Quantification of polyploidization of cells following MLN8237 treatment was also performed in living cells without EdU treatment. The stained cells were assessed by flow cytometry (ACEA NovoCyte^{TM}) using the pacific-blue channel to detect DNA content and analyzed using Flowjo 10.

Orthotopic patient derived xenograft glioma model and treatment with AURKA inhibitor MLN8237

5-week-old female BALB/C nude mice (Chinese Academy of Medical Science Cancer Institute) were used to establish intracranial glioma xenografts, and cells from an *IDH1*/*IDH2* wide type GBM were used as the cell model. A total of five hundred thousand cells were injected to each mouse under the guidance of a stereotactic instrument at coordinates relative to bregma: 2.0 mm posterior, 2.0 mm lateral, 3.0 mm ventral. Bioluminescence imaging was used to detect intracranial tumor growth on days 10, 17, 24 and 31. In the AURKA inhibition studies, mice were randomly divided into two groups (9 mice in the control group and 10 mice in the treatment group), receiving every other day, two 200 μl applications of 25 mg/ml L-Arginine (Sigma, #A8094) or 200 μl of 25 mg/kg MLN8237 (Selleck, #S1133) diluted in L-Arginine by oral gavage. When all mice in the control group were dead, the surviving mice in the treatment group were sacrificed and the brains were extracted and fixed in 10% formalin for 24 h, embedded in paraffin, and sectioned into 5 μm slices. All experiments were conducted in accordance with animal protocols approved by the Institutional Animal Care and Use Committee at Tianjin Medical University. The data were normalized to bioluminescence detected at the initiation of treatment for each animal. Kaplan-Meier survival curves were plotted to analyze the survival outcomes between the control and the AURKA inhibited groups.

Statistical analyses

For statistical analyses, all experiments were replicated at least 3 times as indicated in the figure legends. The mean and standard deviation (SD) of the results are shown in the graphs.

The two-tailed Student's t-test was employed to determine the statistical significance of differences between two groups unless otherwise described in the text. The one-way ANOVA was performed for comparisons in more than two groups. Survival analysis of patients or mice was performed using Kaplan-Meier plots. A log-rank (Mantel-Cox) test was used to compare the survival curves. Fisher's exact two-tailed test was performed for comparing the percentage of lagging chromosomes and centrosomes amplification errors. The statistical analyses were performed using GraphPad Prism 7 software or SPSS statistics 22. Statistical significances were established at $P < 0.05$, $P < 0.01$ and $P < 0.001$, labeled as *, ** and *** respectively.

Supplemental Figures

Fig. S1

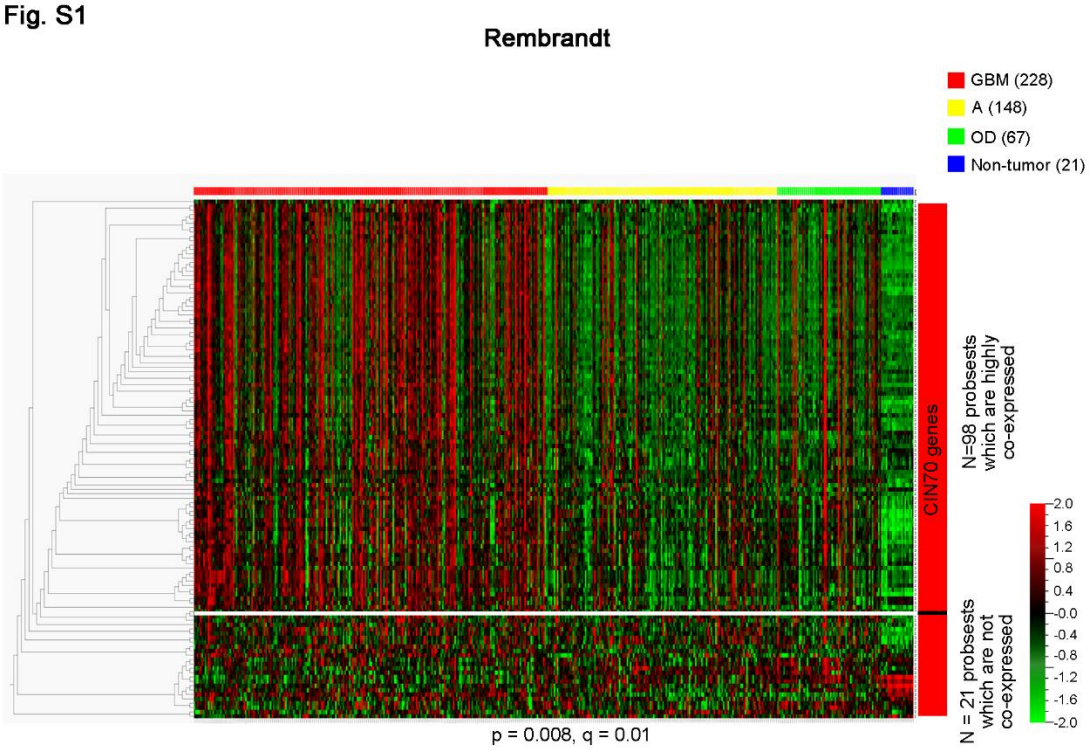


Fig. S1. Only a subset of CIN70 members are co-expressed in glioma transcriptomes. Supervised clustering based on histological classification between glioma samples and the CIN70 expression was performed in the REMBRANDT dataset. Heatmap showed that only a subset of CIN70 genes are co-expressed. Similar patterns were observed in GSE16011 and TCGA datasets. GBM, glioblastoma multiforme; A, astrocytoma; OD, oligodendroglioma.

Fig. S2

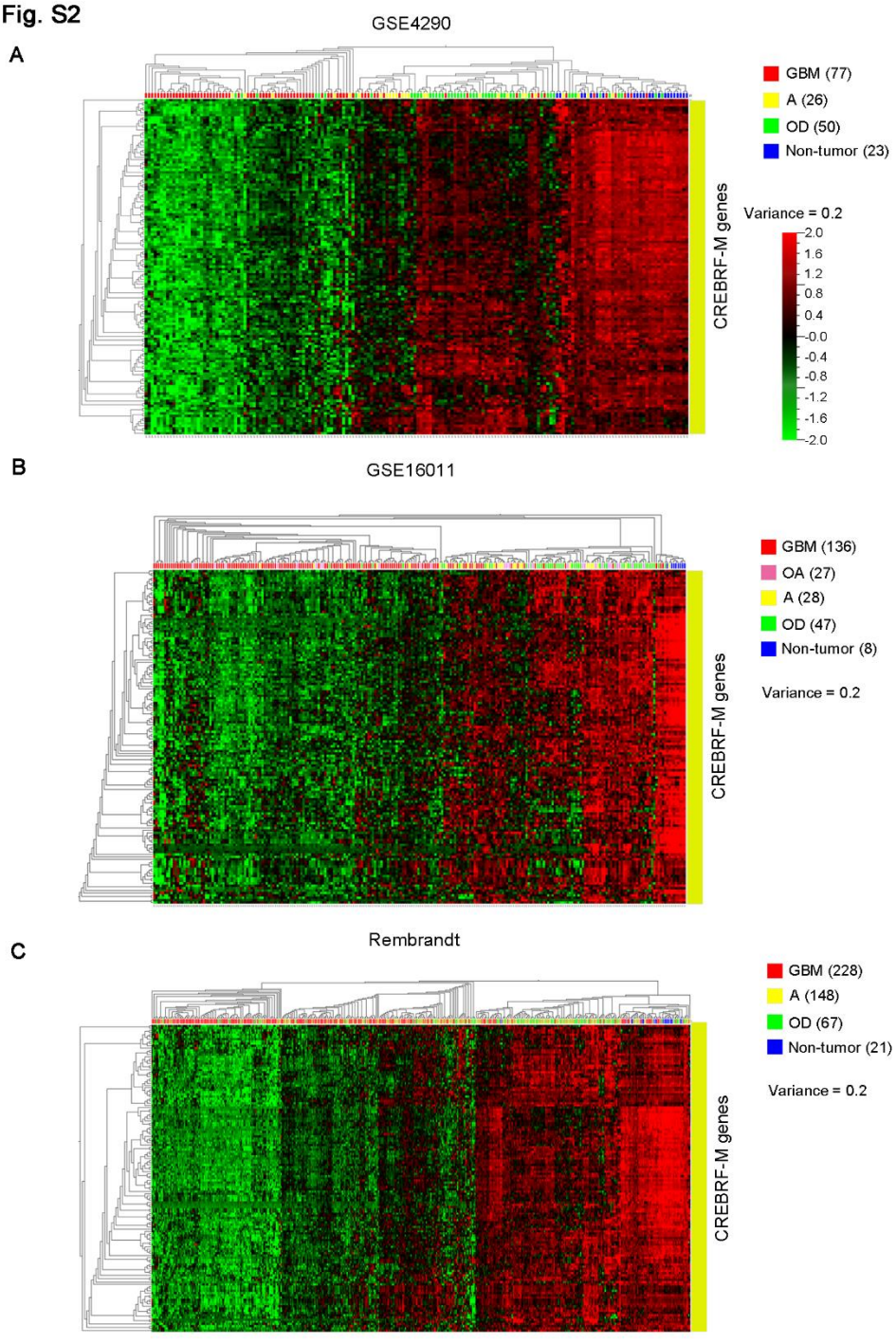

A
GSE4290
GBM (77)
A (26)
OD (50)
Non-tumor (23)
Variance = 0.2
2.0
1.6
1.2
0.8
0.4
-0.0
-0.4
-0.8
-1.2
-1.6
-2.0
CREBRF-M genes
B
GSE16011
GBM (136)
OA (27)
A (28)
OD (47)
Non-tumor (8)
Variance = 0.2
CREBRF-M genes
C
Rembrandt
GBM (228)
A (148)
OD (67)
Non-tumor (21)
Variance = 0.2
CREBRF-M genes

Fig. S2. Enriched expression of CREBRF-M in normal brain tissues.

CREBRF expression was negatively correlated to CDC20-M in the transcriptome datasets examined in this study. CREBRF-M was identified in GSE4290 as the genes co-expressed with CREBRF (235556_at) at a PCC not less than 70%. Heatmaps of unsupervised hierarchical clustering between CREBRF-M and glioma samples in GSE4290 **(A)**, GSE16011 **(B)** and REMBRANDT **(C)**, with the exclusion of genes with standard deviations below 20% of the maximal standard deviation, show enriched expression of CREBRF-M in normal tumor brain tissue samples. GBM, glioblastoma multiforme; OA, oligoastrocytoma; A, astrocytoma; OD, oligodendroglioma.

Fig. S3

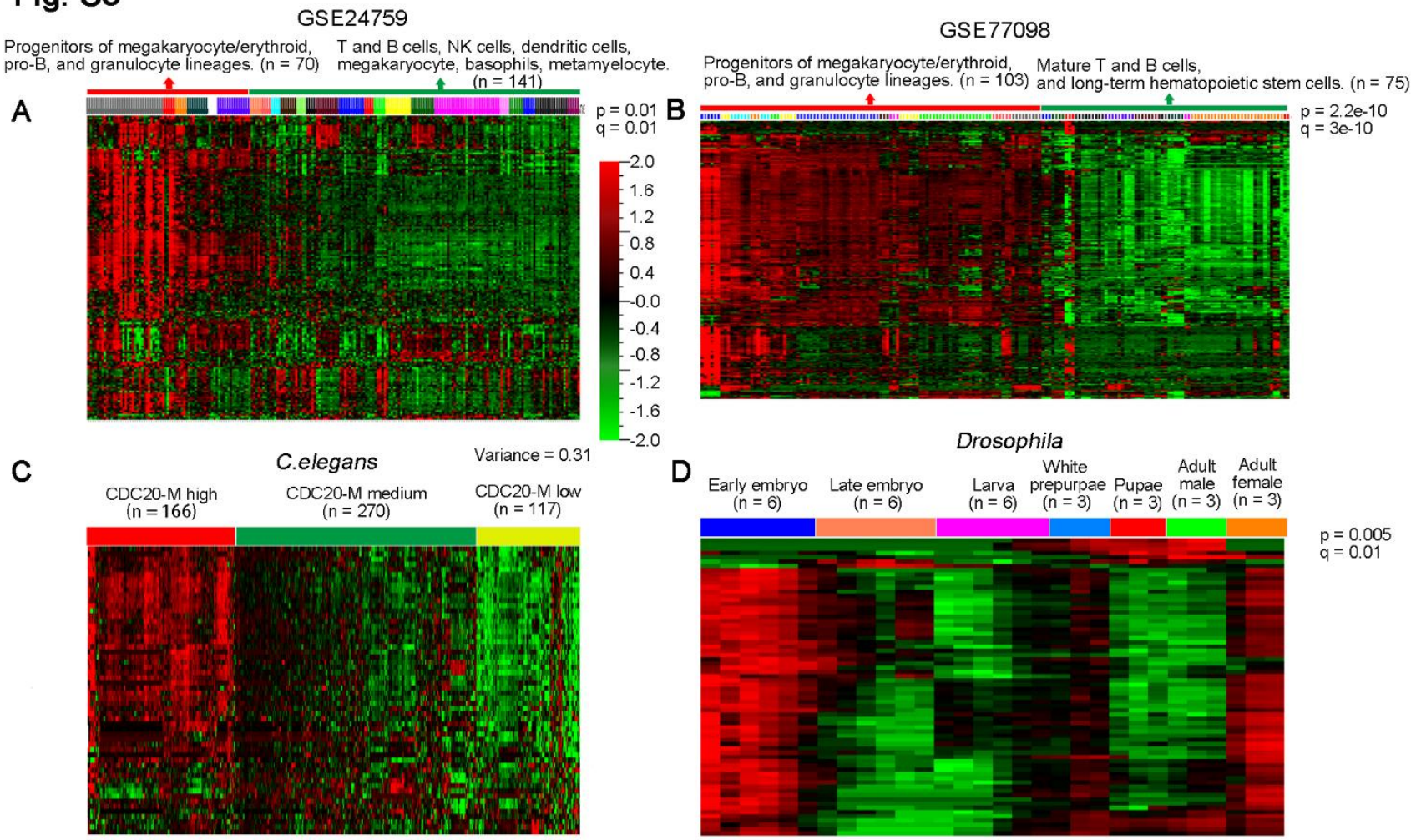

GSE24759
Progenitors of megakaryocyte/erythroid, pro-B, and granulocyte lineages. (n = 70)
T and B cells, NK cells, dendritic cells, megakaryocyte, basophils, metamyelocyte. (n = 141)
A
p = 0.01
q = 0.01
2.0
1.6
1.2
0.8
0.4
-0.0
-0.4
-0.8
-1.2
-1.6
-2.0
GSE77098
Progenitors of megakaryocyte/erythroid, pro-B, and granulocyte lineages. (n = 103)
Mature T and B cells, and long-term hematopoietic stem cells. (n = 75)
B
p = 2.2e-10
q = 3e-10
C
C.elegans
Variance = 0.31
CDC20-M high (n = 166)
CDC20-M medium (n = 270)
CDC20-M low (n = 117)
D
Drosophila
Early embryo (n = 6)
Late embryo (n = 6)
Larva (n = 6)
White prepurpae (n = 3)
Pupae (n = 3)
Adult male (n = 3)
Adult female (n = 3)
p = 0.005
q = 0.01

Fig. S3. Characterization of CDC20-M as a species-conserved cell proliferation marker.

(**A**) and (**B**) Heatmaps of supervised hierarchical clustering analysis of CDC20-M and samples performed in the transcriptome datasets of steady state human hematopoiesis (GSE24759) (**A**) and murine hematopoiesis (GSE77098) (**B**) are shown. Enriched CDC20-M expression was found in the progenitors of megakaryocyte/erythroid and granulocyte lineages and in pro-B progenitors. (**C**) Heatmap of unsupervised hierarchical clustering of *C. elegans* cultured under various growth conditions, or carrying different mutations. Samples were clustered according to CDC20-M expression. Variance = 0.31. (**D**) Heatmap of supervised clustering between CDC20-M expression and samples in the different stages of *Drosophila* life cycle is shown. Early embryo samples show the highest expression of CDC20-M. Enriched CDC20-M signature also indicates a high extent of cell proliferation.

Fig. S4

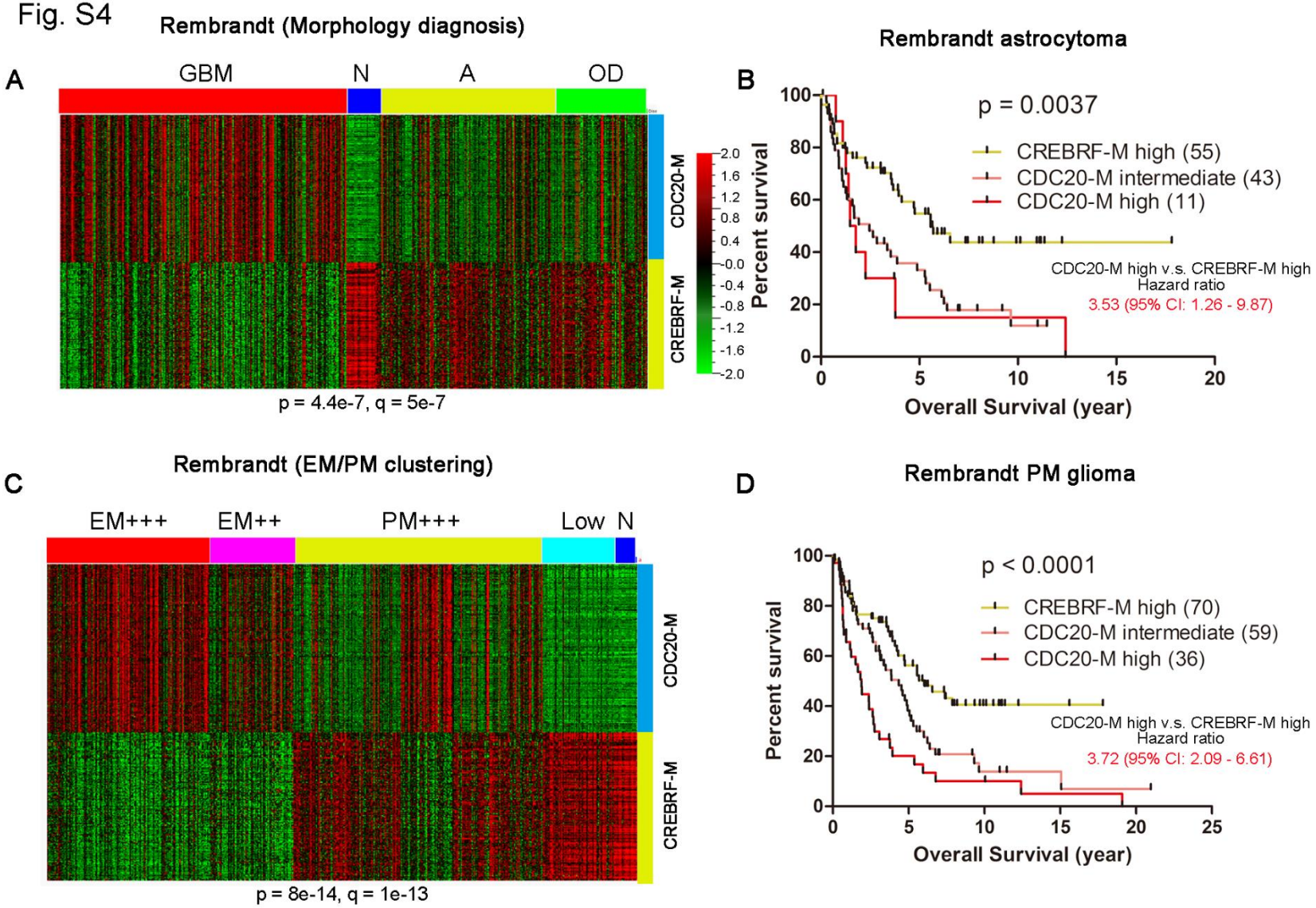
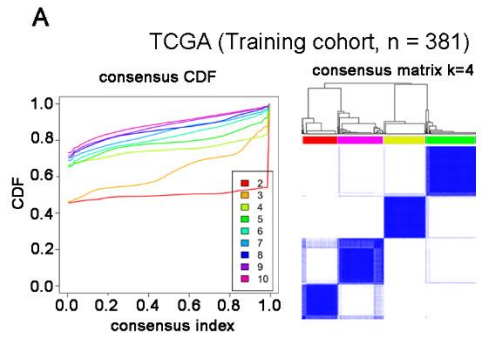
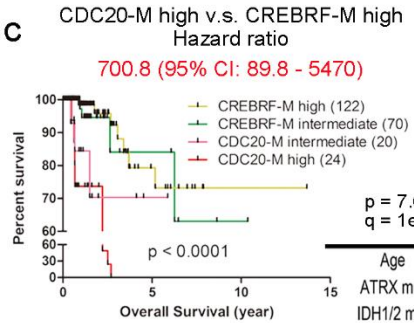

Rembrandt (Morphology diagnosis)
A
GBM
N
A
OD
CDC20-M
CREBRF-M
2.0
1.6
1.2
0.8
0.4
0.0
-0.4
-0.8
-1.2
-1.6
-2.0
p = 4.4e-7, q = 5e-7
Rembrandt astrocytoma
B
p = 0.0037
CREBRF-M high (55)
CDC20-M intermediate (43)
CDC20-M high (11)
CDC20-M high v.s. CREBRF-M high
Hazard ratio
3.53 (95% CI: 1.26 - 9.87)
Percent survival
Overall Survival (year)
Rembrandt (EM/PM clustering)
C
EM+++
EM++
PM+++
Low
N
CDC20-M
CREBRF-M
p = 8e-14, q = 1e-13
Rembrandt PM glioma
D
p < 0.0001
CREBRF-M high (70)
CDC20-M intermediate (59)
CDC20-M high (36)
CDC20-M high v.s. CREBRF-M high
Hazard ratio
3.72 (95% CI: 2.09 - 6.61)
Percent survival
Overall Survival (year)

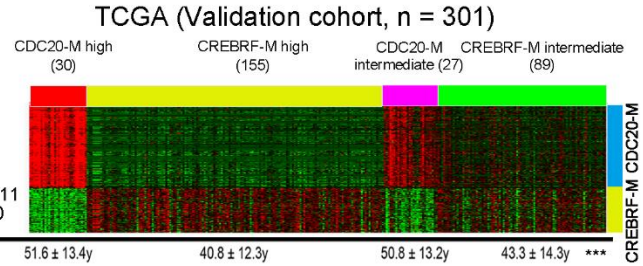
Fig. S4. High CDC20-M expression marks poor prognosis in the same morphological subtypes or the EM/PM molecular subtypes.

Heatmaps of CDC20-M and CREBRF-M expression profiles in glioma samples in Rembrandt dataset sorted according to morphological diagnosis (**A**), or EM/PM clustering (**C**) are shown. Among the grade II/III astrocytomas, gliomas with high or intermediate level of CDC20-M expression were associated with poorer prognosis compared to gliomas with high level of CREBRF-M expression (**B**), a similar trend was found among the PM gliomas **(D)**. The hazard ratio between patients with high CDC20-M glioma and patients with high CREBRF-M gliomas is shown in panels (**B**) and (**D)**, marked in red.

Fig. S5

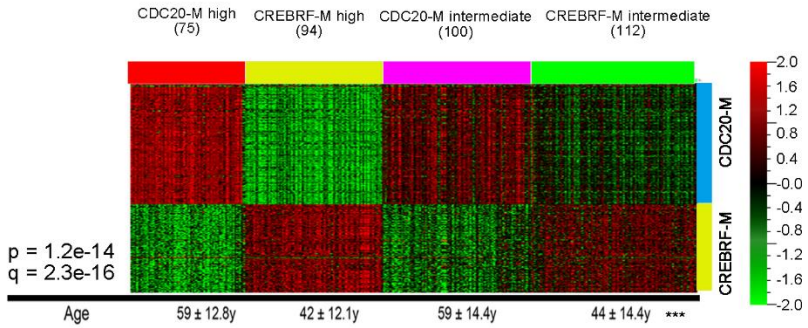






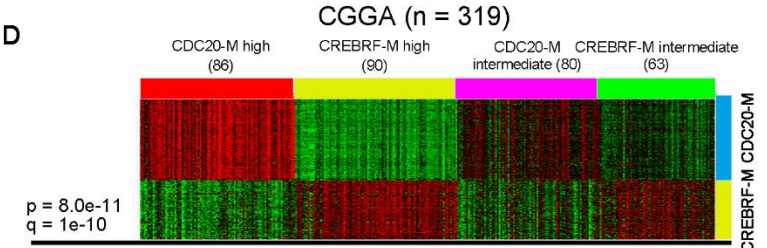
Age	51.6 ± 13.4y	40.8 ± 12.3y	50.8 ± 13.2y	43.3 ± 14.3y ***
ATRX mut	26.7%	36.0%	19.2%	35.2%
IDH1/2 mut	43.3%	92.0%	50.0%	89.8% ***
1p19q co-del	3.3%	38.0%	23.1%	43.2% ***
GBM	0.0%	0.0%	0.0%	0.0%
Astrocytoma	46.7%	31.3%	50.0%	29.5%
Oligoastrocytoma	16.7%	21.3%	11.5%	15.9%
Oligodendroglioma	16.7%	28.7%	15.4%	34.1%
Unknown	20.0%	18.7%	23.1%	20.5%
EM	46.7%	12.0%	50.0%	9.1% ***
PM	16.7%	77.3%	46.2%	90.9% ***
EMlowPMlow	0.0%	10.0%	0.0%	0.0% **
Unknown	36.7%	0.7%	3.8%	0.0% ***

B



Age	59 ± 12.8y	42 ± 12.1y	59 ± 14.4y	44 ± 14.4y ***
ATRX mut	5.3%	30.9%	9.0%	43.8% ***
IDH1/2 mut	3.9%	80.9%	8.0%	77.7% ***
1p19q co-del	0.0%	28.7%	0.0%	24.1% ***
GBM	62.7%	3.2%	66.0%	4.5% ***
Astrocytoma	2.7%	17.0%	13.0%	27.7% ***
Oligoastrocytoma	2.7%	24.5%	4.0%	22.3% ***
Oligodendroglioma	1.3%	48.9%	3.0%	34.8% ***
Unknown	30.7%	6.4%	14.0%	10.7% ***
EM	77.3%	6.4%	77.0%	17.9% ***
PM	22.7%	69.1%	22.0%	79.5% ***
EMlowPMlow	0.0%	24.5%	1.0%	2.7% ***

D



Age	47.2 ± 12.6y	41.2 ± 9.0y	45.0 ± 12.7y	40.7 ± 11.2y ***
IDH1 mut	27.9%	73.3%	33.8%	76.2% ***
1p19q co-del	2.3%	41.1%	10.0%	34.9% ***
GBM	84.9%	6.7%	57.5%	23.8% ***
Astrocytoma	7.0%	33.3%	18.8%	25.4% ***
Oligoastrocytoma	5.8%	33.3%	21.3%	31.7% ***
Oligodendroglioma	2.3%	26.7%	2.5%	19.0% ***
EM	66.3%	4.4%	51.3%	9.5% ***
PM	31.4%	35.6%	37.5%	73.0% ***
EMlowPMlow	1.2%	56.7%	7.5%	15.9% ***
Unknown	1.2%	3.3%	3.8%	1.6% ***

Fig. S5. Unsupervised consensus clustering and SSP of gliomas with CDC20-M and CREBRF-M signature, and characteristics of each subtype in the TCGA training and validation cohorts, and in the CGGA dataset.

(**A**) Results of unsupervised consensus clustering generated using partitioning around medoid between glioma samples, and the CDC20-M and CREBRF-M signature in the TCGA training cohort. The number of clusters (K) were individually assessed. (Left) Consensus CDF plots corresponding to the consensus matrices in the K ranging from 2 to 10. (Right) Consensus matrices for K = 4, which were the most stable clusters for TCGA training cohort. (**B**) The upper part shows the same heatmap as in Fig. 2B. Age at diagnosis, genomic characteristics, morphological diagnosis and EM/PM subtypes of each CDC20-M subgroup are shown in the lower part. ***: $p < 0.001$ as analyzed in one-way ANOVA test for age comparison, and Fisher´s exact test for all the other comparisons. (**C**) (Left) Kaplan-Meier plot of the overall survival for patients from each molecular subtype of the TCGA validation cohort is shown. The overall survival data were analyzed using log-rank tests. The hazard ratio between patients with high CDC20-M and patients with high CREBRF-M gliomas is shown on top of the survival curves. (Right) The upper part shows the heatmap of supervised clustering of the CDC20-M and CREBRF-M signature in the TCGA validation cohort. Samples were assigned to the nearest centroid as determined by Spearman correlation. Age at diagnosis, genomic characteristics, morphological diagnosis and EM/PM subtypes in each CDC20-M subgroup are shown in the lower part. ***: $p < 0.001$ as analyzed in one-way ANOVA test for age comparison, and Fisher´s exact test for all the other comparisons. (**D**) The upper part shows the same heatmap as in Fig. 2C. Age at diagnosis, genomic characteristics, morphological diagnosis and EM/PM subtypes in

each CDC20-M subgroup in the CGGA dataset are shown in the lower part. ***: $p < 0.001$ as analyzed in one-way ANOVA test for age comparison, and Fisher´s exact test for all the other comparisons.

Fig S6.

A

CDC20-M cluster	TCGA cluster				
	LGr1	LGr2	LGr3	LGr4	NA
CDC20-M high	22	0	13	70	0
CDC20-M intermediate	19	1	14	92	1
CREBRF-M intermediate	55	16	104	24	2
CREBRF-M high	46	73	107	15	8

B

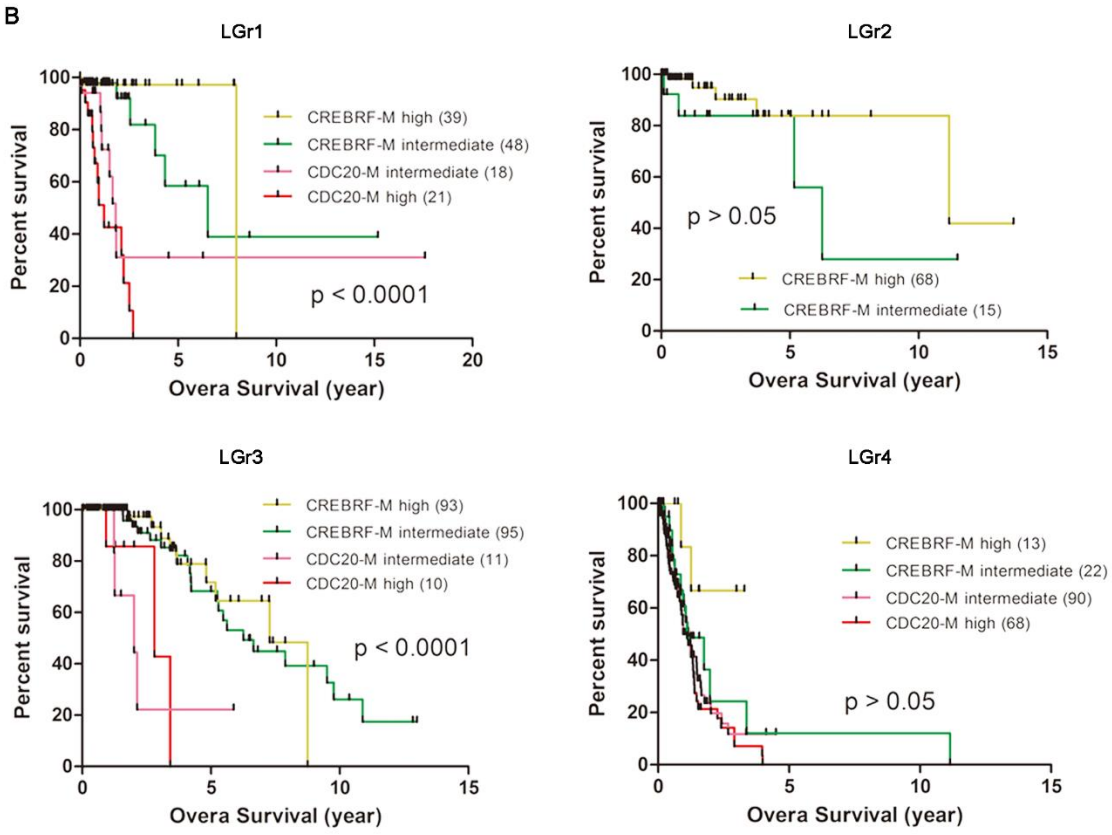


Fig. S6. CDC20-M signature-based clustering distinguishes the lower grade gliomas with poorer from those with better prognosis within the *IDH* mutated LGr1 and LGr3 clusters.

(A) Cross-labeling of lower grade glioma samples by the TCGA clustering or by the CDC20-M/CREBRF-M clustering. **(B)** According to CDC20-M/CREBRF-M clustering, the overall survival data were reanalyzed using log-rank tests. In the good prognostic LGr1 and LGr3 clusters, patients of gliomas with high or intermediate CDC20 expression showed worse prognosis compared to patients of gliomas with high or intermediate CREBRF-M expression.

Fig. S7

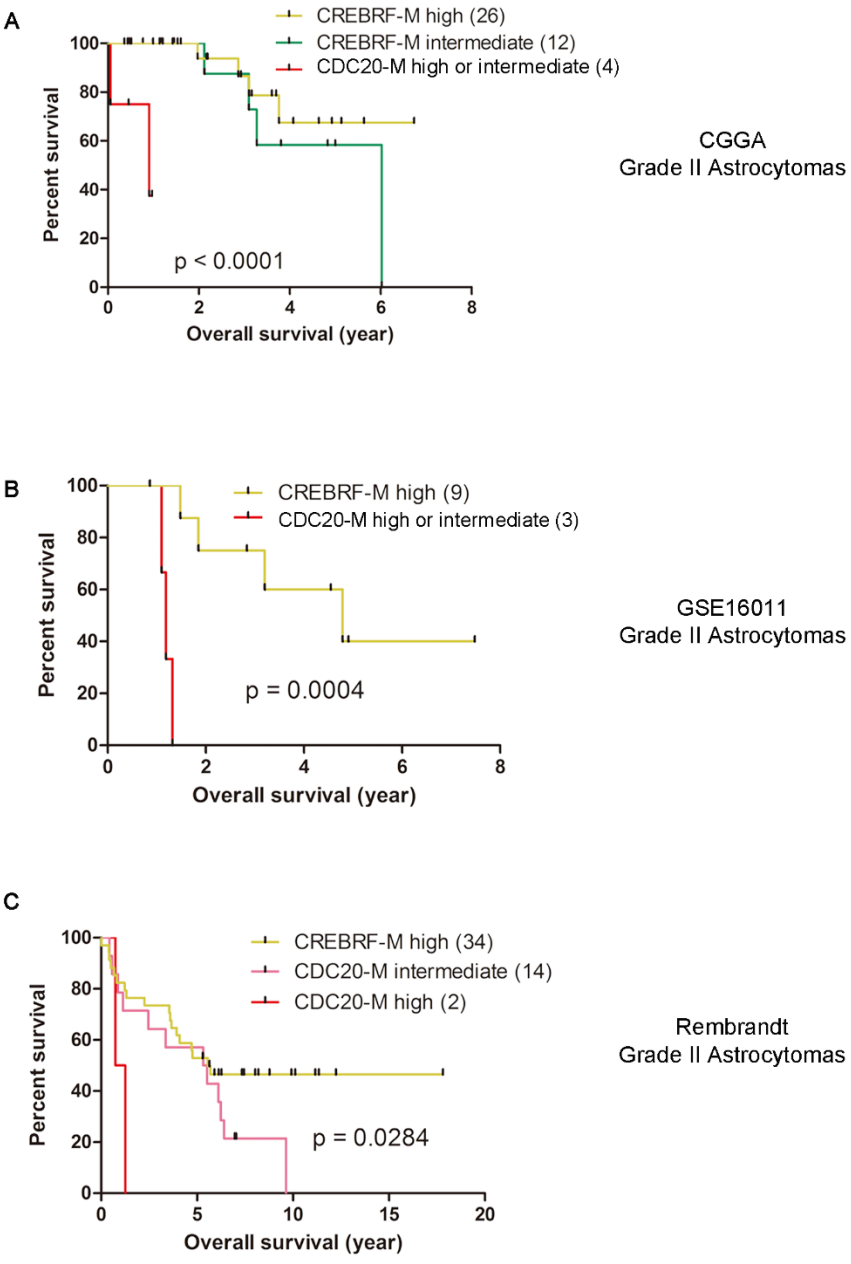
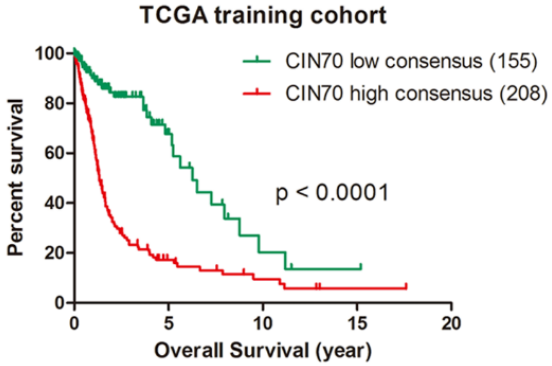


Fig. S7. Rapid progression of grade II astrocytomas with high CDC20-M expression. Overall survival analysis was performed in patients with grade II astrocytomas stratified according to the CDC20-M/CREBRF-M clustering. P values of log-rank test are shown.

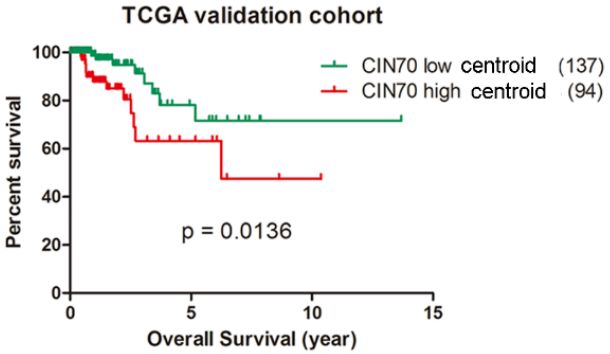
Fig. S8

A



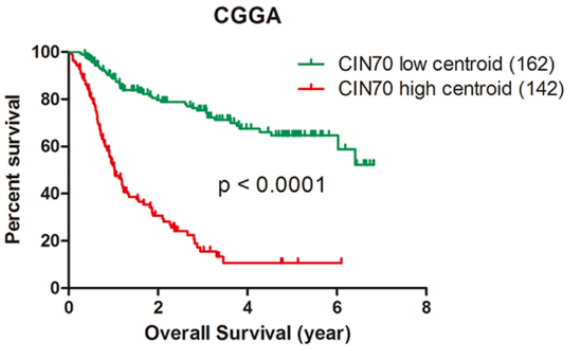
Hazard ratio: 3.055 (95% CI: 2.190 - 4.124)

B



Hazard ratio: 3.077 (95% CI: 1.260 - 7.511)

C



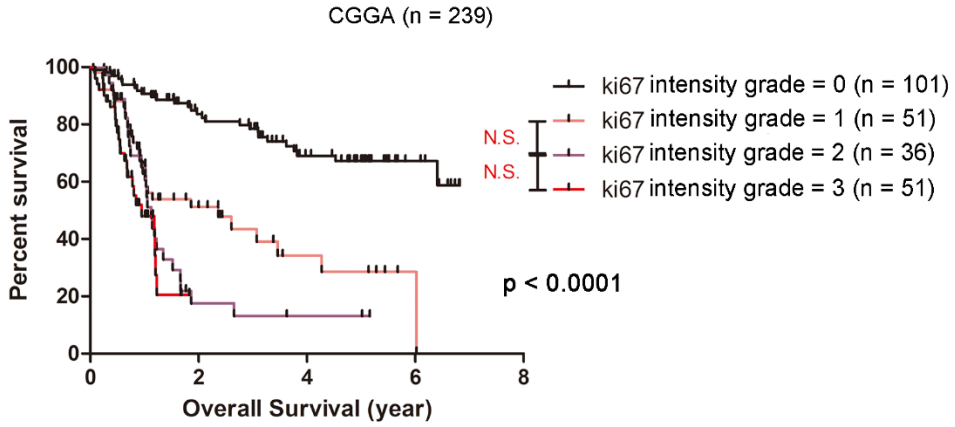
Hazard ratio: 6.172 (95% CI: 4.264 - 8.932)

Fig. S8. Kaplan-Meier plot of the overall survival for patients from the TCGA training (A), TCGA validation (B), and CGGA (C) datasets classified according to CIN70 signature.

Samples in the TCGA training dataset were classified using CIN70 signature-based unsupervised consensus clustering, and served as a basis for the calculation of centroid for CIN70 high or low signature. Subsequently, CIN70 high or low gliomas in the TCGA validation and CGGA dataset were identified using SSP. The overall survival data were analyzed using log-rank test. The hazard ratio between patients with high CIN70 glioma and patients with low CIN70 gliomas is shown under the survival curves of each dataset.

Fig. S9

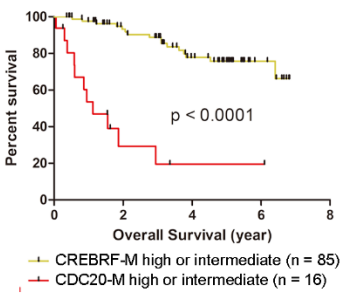
A



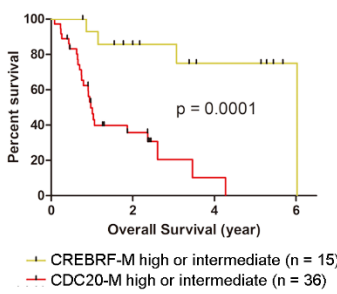
B

CDC20-M cluster	Ki67 intensity grades			
	= 0	= 1	= 2	= 3
CDC20-M high	6	18	12	30
CDC20-M intermediate	10	18	15	14
CREBRF-M intermediate	22	9	7	6
CREBRF-M high	63	6	2	1

Ki67 intensity grade = 0 (101)



Ki67 intensity grade = 1 (51)



Ki67 intensity grade = 2 (36)

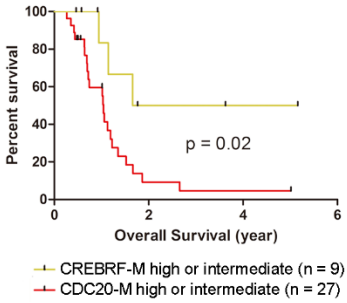


Fig. S9. CDC20-M signature outperforms Ki-67 immunohistochemical staining in predicting prognosis of glioma patients.

(A) Overall survival data of the 239 CGGA glioma patients were analyzed according to the extent of Ki-67 staining. Samples containing <5%, 5%-30%, 30%-50% and >50% Ki67-positive cells were categorized into intensity grades 0, 1, 2 and 3, respectively. **(B)** Cross-labelling between CDC20-M/CREBRF-M clustering and Ki-67 staining intensity subgroups. Within all subgroups of Ki-67 staining intensity grades of 0 to 2, gliomas with high CDC20-M expression consistently showed poorer overall survival compared to those with high CREBRF-M expression. The overall survival data were analyzed using log-rank test. Gliomas with Ki-67 staining intensity grade 3 were excluded from the survival analysis because too few CREBRF-M high or intermediate samples were found in this subgroup.

Fig. S10

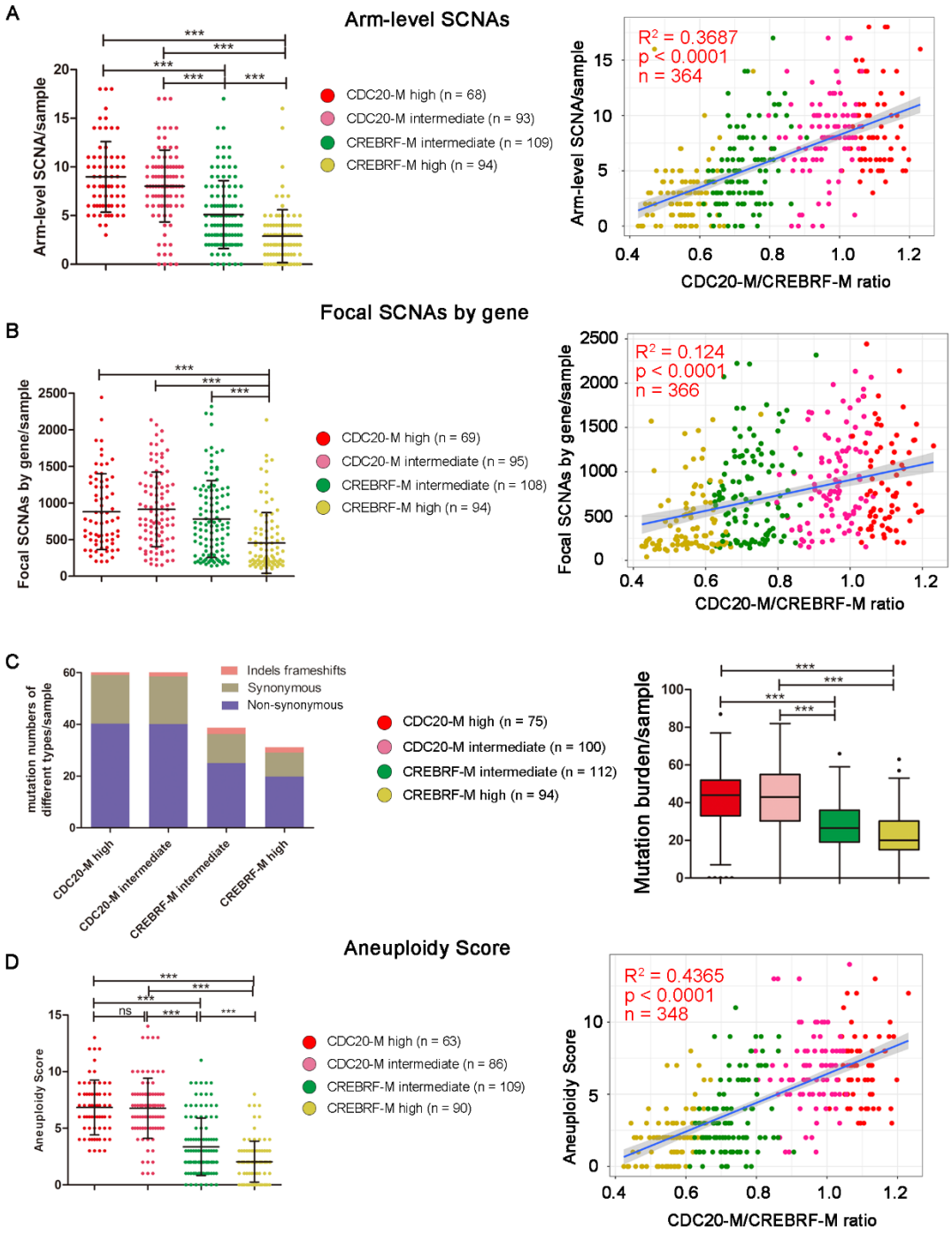
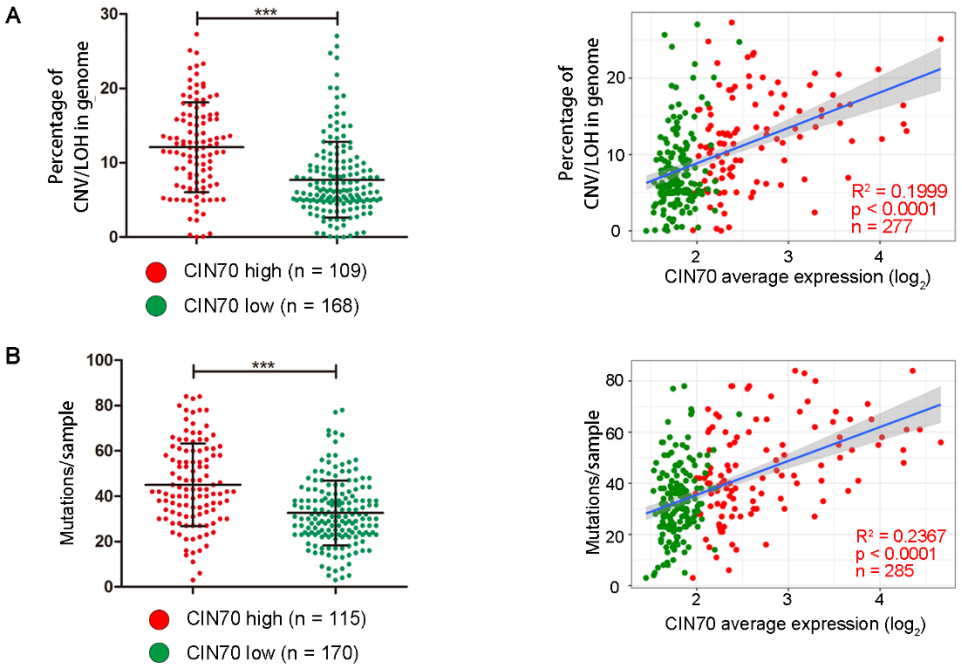

A
Arm-level SCNAs
Arm-level SCNA/sample
CDC20-M high (n = 68)
CDC20-M intermediate (n = 93)
CREBRF-M intermediate (n = 109)
CREBRF-M high (n = 94)
R² = 0.3687
p < 0.0001
n = 364
Arm-level SCNA/sample
CDC20-M/CREBRF-M ratio
B
Focal SCNAs by gene
Focal SCNAs by gene/sample
CDC20-M high (n = 69)
CDC20-M intermediate (n = 95)
CREBRF-M intermediate (n = 108)
CREBRF-M high (n = 94)
R² = 0.124
p < 0.0001
n = 366
Focal SCNAs by gene/sample
CDC20-M/CREBRF-M ratio
C
Indels frameshifts
Synonymous
Non-synonymous
mutation numbers of different types/sample
CDC20-M high
CDC20-M intermediate
CREBRF-M intermediate
CREBRF-M high
CDC20-M high (n = 75)
CDC20-M intermediate (n = 100)
CREBRF-M intermediate (n = 112)
CREBRF-M high (n = 94)
Mutation burden/sample
D
Aneuploidy Score
CDC20-M high (n = 63)
CDC20-M intermediate (n = 86)
CREBRF-M intermediate (n = 109)
CREBRF-M high (n = 90)
R² = 0.4365
p < 0.0001
n = 348
Aneuploidy Score
CDC20-M/CREBRF-M ratio

Fig. S10. High CDC20-M signature correlates to the extent of arm-level or focal SCNA, non-synonymous mutation and aneuploidy.

(A) (Left) Extent of arm-level SCNAs per sample in the CDC20-M defined glioma subgroups of the TCGA training cohort. ***: $p < 0.001$, unpaired two-sided Student's t-test. (Right) Linear regression analysis between the ratio of CDC20-M/CREBRF-M and arm-level SCNAs in all samples analyzed in the left panel. Regression coefficient (R^2), P value and number of samples (N) are indicated. **(B)** (Left) Extent of focal-level SCNAs by gene per sample in the CDC20-M defined glioma subgroups of the TCGA training cohort. ***: $p < 0.001$, unpaired two-sided Student's t-test. (Right) Linear regression analyses between the ratio of CDC20-M/CREBRF-M and focal-level SCNAs by gene in all samples analyzed in the left panel. Regression coefficient (R^2), P value and number of samples (N) are indicated. **(C)** (Left) Average numbers of indels/frameshifts, synonymous and non-synonymous mutations per sample in the CDC20-M/CREBRF-M clusters in the TCGA training cohort. (Right) Box plot of mutation burden (non-synonymous and indels/frameshifts mutations) in each CDC20-M/CREBRF-M group of TCGA training cohort. ***: $p < 0.001$ as analyzed in one-way ANOVA test. **(D)** (Left) Extent of aneuploidy score in the CDC20-M defined glioma subgroups of the TCGA training cohort. ***: $p < 0.001$, unpaired two-sided Student's t-test. (Right) Linear regression analysis between the ratio of CDC20-M/CREBRF-M and aneuploidy score in all samples analyzed in the left panel. Regression coefficient (R^2), P value and number of samples (N) are indicated.

Fig. S11



C TCGA validation cohort

Percentage of CNV/LOH in genome < 10% (n = 169)					P value
CREBRF-M high	68.0% (115)	94.7% (160)	CIN70 low	76.3% (129)	< 0.0001
CREBRF-M intermediate	26.6% (45)				
CDC20-M intermediate	4.1% (7)	5.3% (9)	CIN70 high	23.7% (40)	
CDC20-M high	1.2% (2)				

D TCGA validation cohort

Mutation/sample < 40 (n = 178)					P value
CREBRF-M high	64.0% (114)	93.3% (166)	CIN70 low	71.3% (127)	< 0.0001
CREBRF-M intermediate	29.2% (52)				
CDC20-M intermediate	5.6% (10)	6.7% (12)	CIN70 high	28.1% (51)	
CDC20-M high	1.1% (2)				

Fig. S11. Glioma subtypes defined by the CIN70 signature are less distinct in the extent of their chromosomal abnormalities and mutation burden.

(A) and **(B)** (Left) the extent of chromosomal abnormalities (A) or mutation burden (B) in the CIN70 signature-defined glioma subgroups in the TCGA validation cohort. ***: $p < 0.001$ as analyzed using unpaired two-sided Student's t-test. (Right) Linear regression analysis between the average $\log_2$ expression of CIN70 signature and the extent of chromosomal abnormalities (A) or mutation burden (B) in glioma samples. The same color codes were used in all panels. Regression coefficient (R^2), P value and number of samples (N) are indicated. **(C)** CIN70 signature-based glioma subtypes were less distinct in the extent of chromosomal abnormalities compared to the CDC20-M-based glioma subtype. This table shows number of samples in each subtype with a percentage of CNV/LOH in genome < 10% in the TCGA validation cohort. P value was analyzed using Fisher´s exact test. **(D)** CIN70 signature-based glioma subtypes were less distinct in the extent of mutation burden compared to the CDC20-M-based glioma subtype. This table shows number of samples in each subtype with less than 40 mutations in the TCGA validation cohort. P value was analyzed using Fisher´s exact test.

Fig. S12 A CDC20-M high (67 samples with SNP 6.0 data)

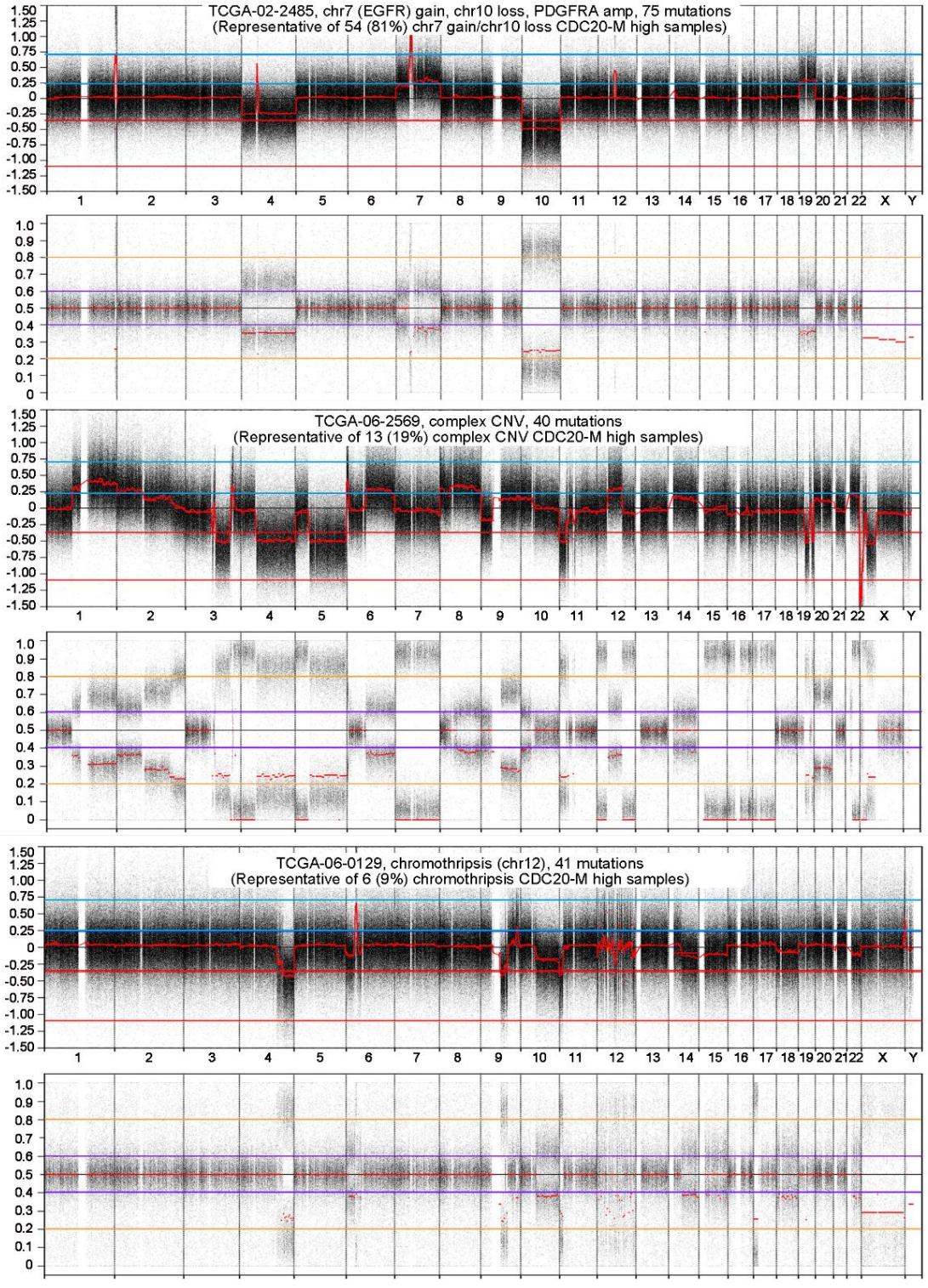


Fig. S12 B CREBRF-M high (87 samples with SNP 6.0 data)

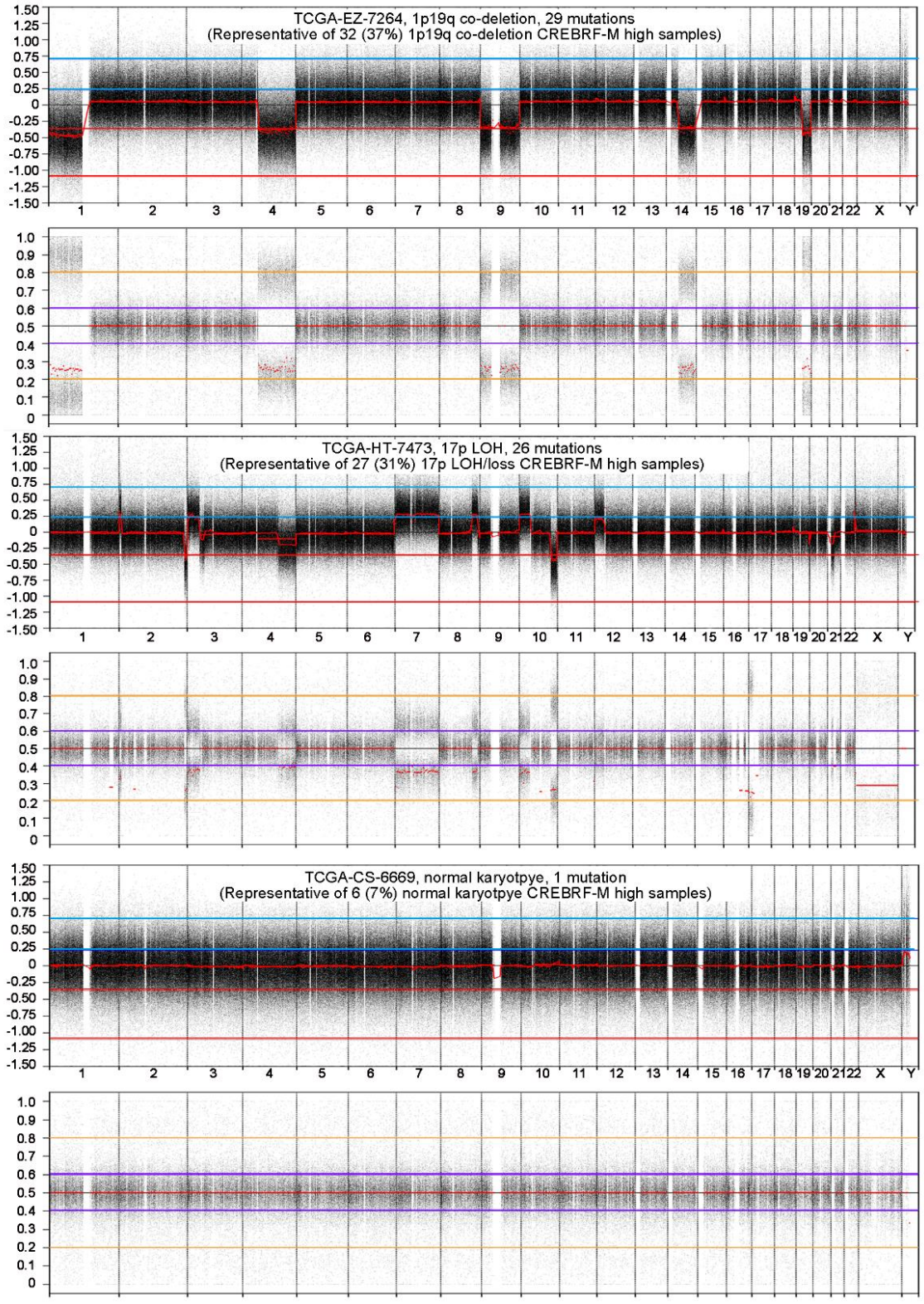


Fig. S12. Distinct patterns and severities of chromosomal abnormalities between gliomas with high CDC20-M expression and gliomas with high CREBRF-M expression.

The whole chromosome view of three representative CDC20-M high glioma samples **(A)** and three representative CREBRF-M high glioma samples **(B)** are shown. For each sample, the upper plot shows the log ratio of copy number aberrations and the lower plot shows allelic events. In the upper plot, each dot corresponds to a probe on the array. The "call" thresholds for single and higher copy number gain or loss are shown as blue and red horizontal lines, respectively in the copy number (log ratio) plots. The red line in the plot are the output of the call algorithm. For the B allele frequency plot in the bottom, the brown and purple lines mark the thresholds for LOH and allelic imbalance regions, respectively.

Fig. S13

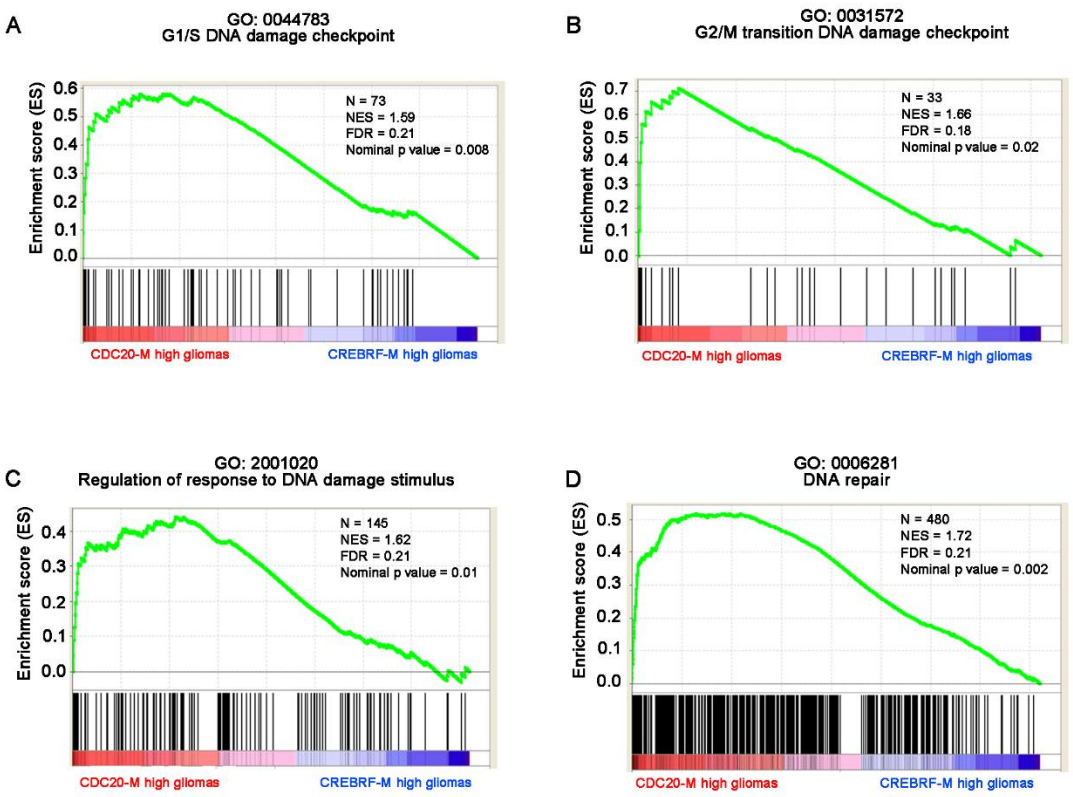
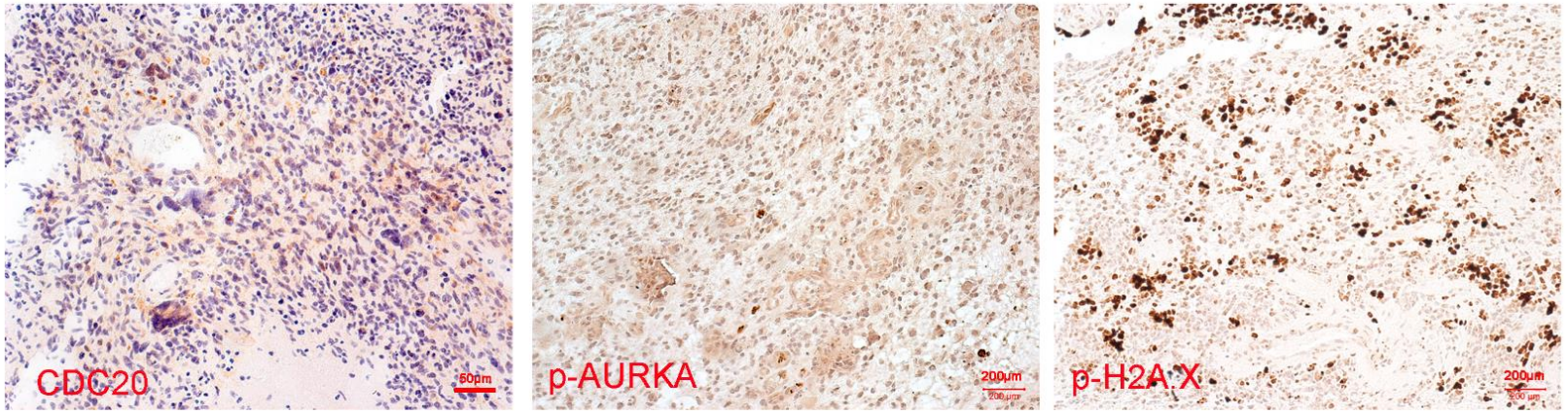


Fig. S13. GSEA plots demonstrating enrichment of the DNA damage response signatures in the CDC20-M high gliomas.

Gene sets involved in G1/S DNA damage checkpoint **(A)**, G2/M DNA damage checkpoint **(B)**, regulation of response to DNA damage stimulus **(C)** and DNA repair **(D)** are enriched in CDC20-M high gliomas. Gene sets were obtained from GO Biological Process Ontology (http://www.geneontology.org/). Number of genes, normalized enrichment scores (NES), false discovery rate (FDR) and nominal p value of gene set are indicated in each panel. Data are derived from comparing the transcriptomes of the CDC20-M high gliomas with the transcriptomes of the CREBRF-high gliomas in the TCGA training cohort.

Fig. S14

CDC20-M high



CREBRF-M high

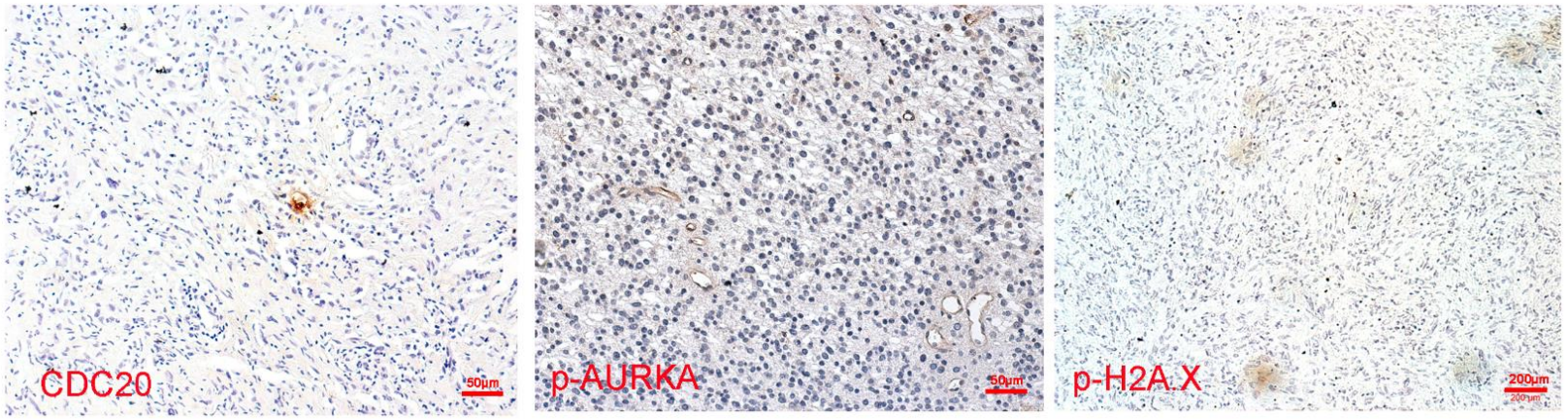


Fig. S14. CDC20-M high gliomas show higher expression of CDC20, p-AURKA and p-H2A.X.

Representative images of immunohistochemical staining of CDC20, p-AURKA and p-H2A.X in glioma tissues classified according to CDC20-M classification in CGGA. Scale bars are indicated in figures. Results are representative of two independent experiments (n = 3 to 6 samples for each staining per group).

Fig. S15

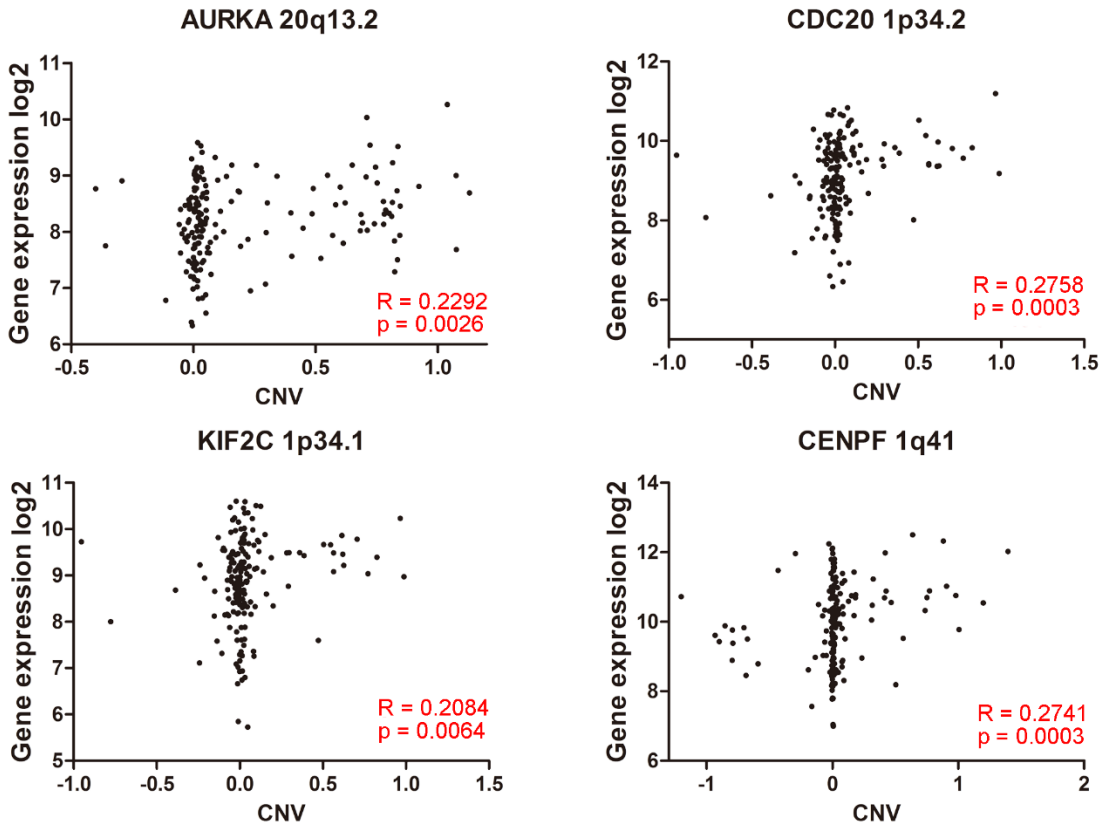


Fig. S15. Gene dosage-dependent expression of AURKA, CDC20, KIF2C and CENPF in CDC20-M high/intermediate gliomas of TCGA training cohort.

Core members of CDC20-M with SCNAs resulting in gene dosage-dependent expression in gliomas with high or intermediate expression levels of CDC20-M were identified using Spearman's rank correlation analysis. Correlation coefficients (R) and their P values are indicated.

Fig. S16

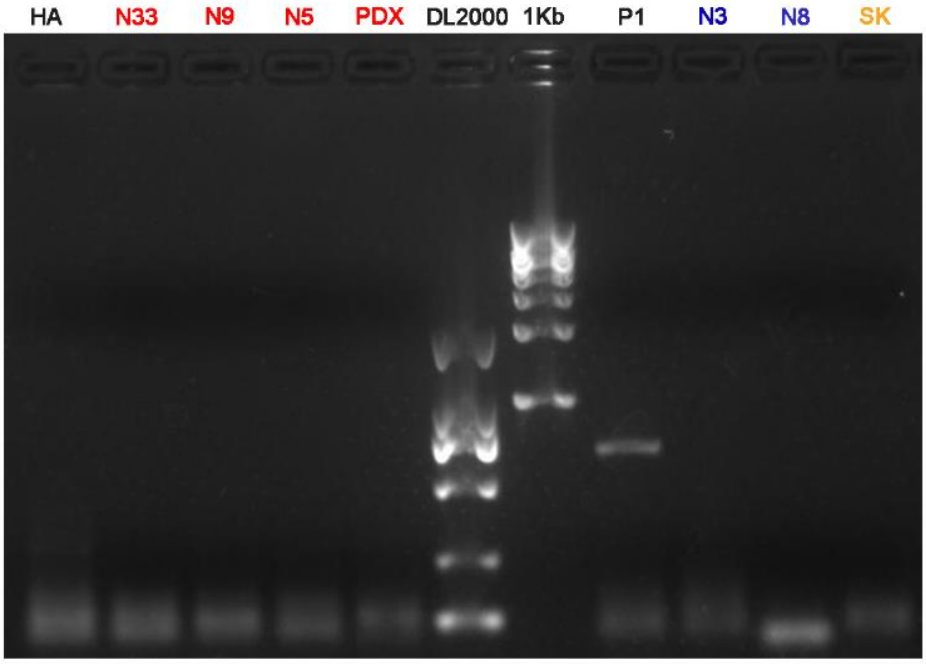
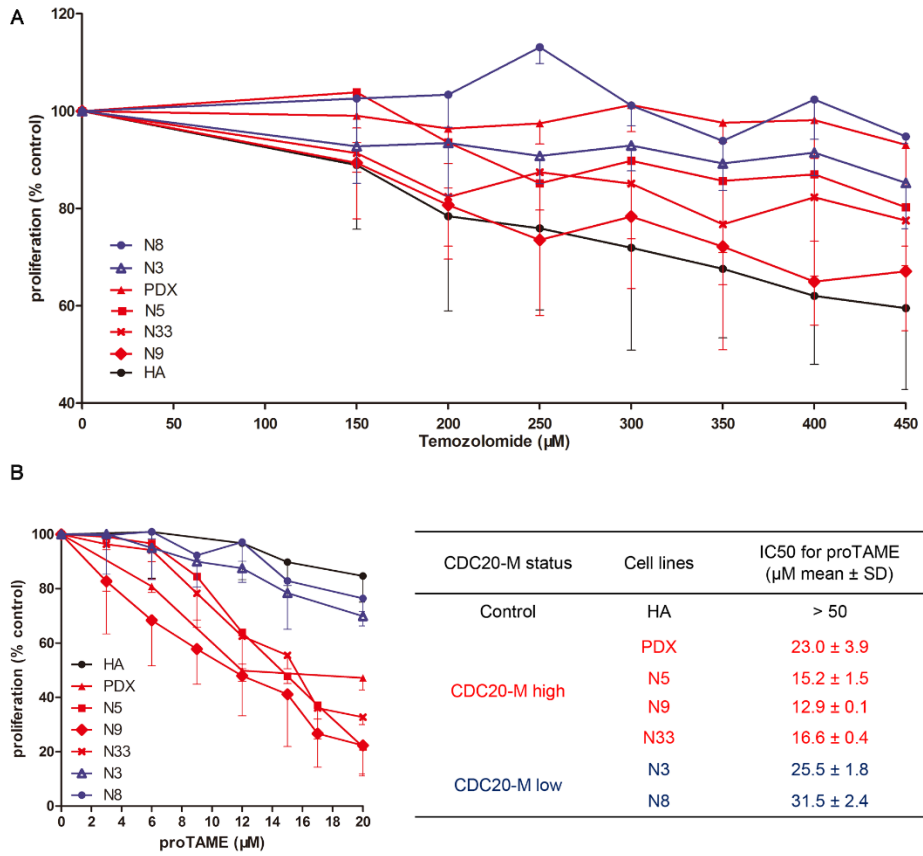


Fig. S16. FGFR3-TACC3 fusion transcript was not detected in all cell lines analyzed in this study.

FGFR3-TACC3 specific RT-PCR with cDNA derived from the cell lines were performed. pLOC-F3T3 plasmid (20), encoding FGFR3-TACC3 fusion transcript, was used as the positive control (P1).

Fig. S17

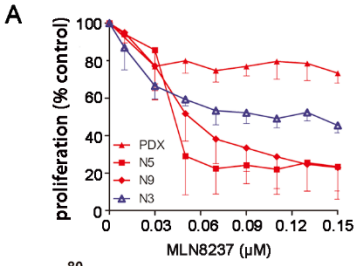


CDC20-M status	Cell lines	IC50 for proTAME (μM mean ± SD)
Control	HA	> 50
CDC20-M high	PDX	23.0 ± 3.9
	N5	15.2 ± 1.5
	N9	12.9 ± 0.1
	N33	16.6 ± 0.4
CDC20-M low	N3	25.5 ± 1.8
	N8	31.5 ± 2.4

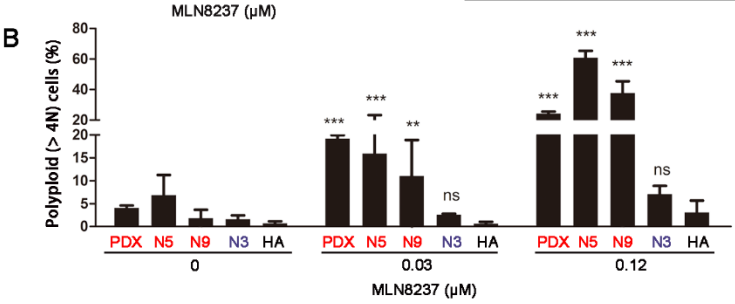
Fig. S17. Sensitivity of glioma cell lines to temozolomide and proTAME.

(A) Log-phase glioma cells were treated with temozolomide at varying concentrations for three days. Glioma lines were not more sensitive than controls to treatment with temozolomide. Data shown are averages of three independent experiments. Each concentration contained six replicates in one experiment. Error bars indicate SD. **(B)** (Left) Log-phase cell lines were treated with proTAME at varying concentrations for 24 h. CDC20-M high glioma cell lines were more sensitive to proTAME compared with CDC20-M low glioma cell lines and HA. Data shown are averages of three independent experiments, each concentration was assessed in six replicates in one experiment, and error bars indicate SD. (Right) The table shows the summary of IC50 concentrations of proTAME for all cell lines tested.

Fig. S18



CDC20-M status	Cell lines	IC50 for MLN8237 (μM mean ± SD)
Control	HA	> 0.15
CDC20-M high	PDX	> 0.15
	N5	0.04 ± 0.02
	N9	0.05 ± 0.01
	N33	0.05 ± 0.00
CDC20-M low	N3	0.11 ± 0.03
	N8	> 0.15



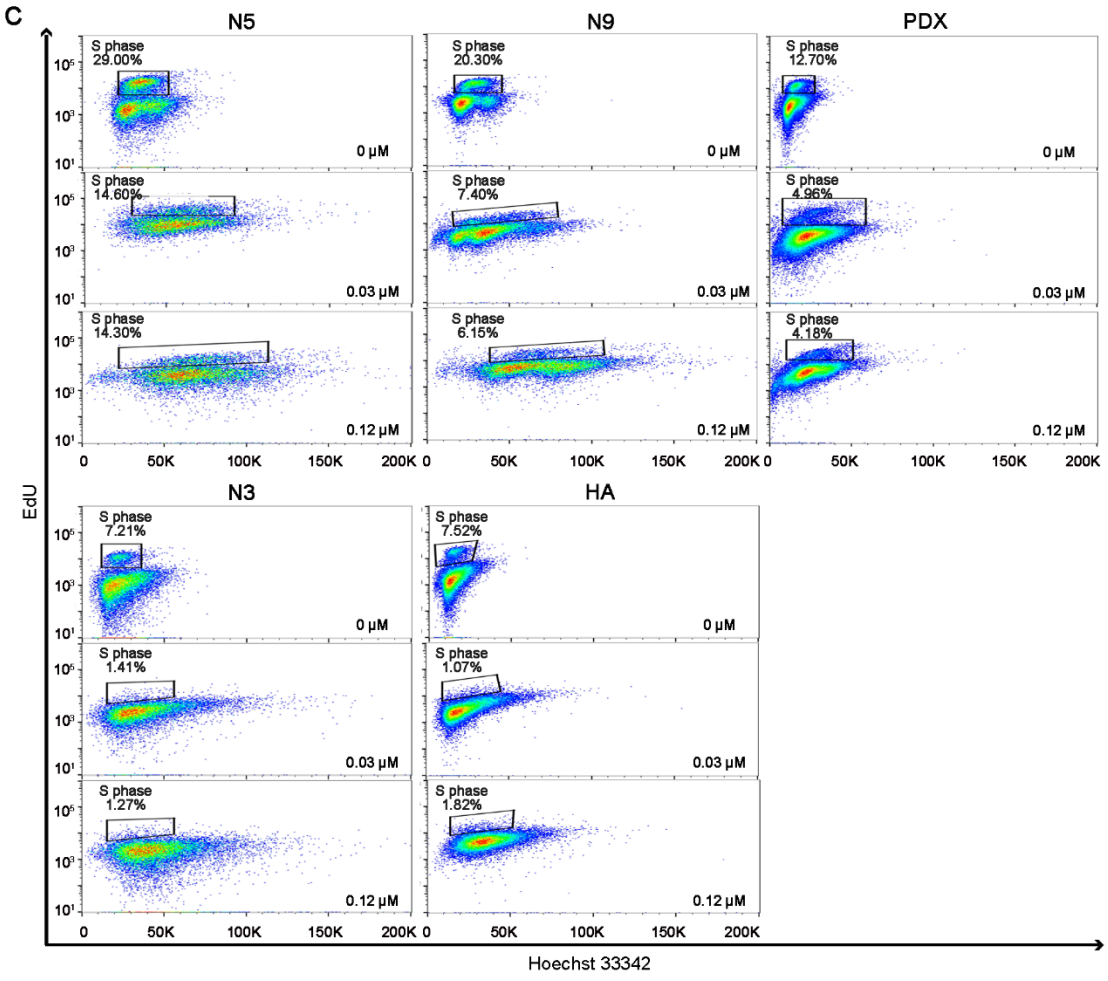


Fig. S18. AURKA inhibition induced proliferation arrest and polyploidization in CDC20-M high glioma cells.

(A) (Left) Log-phase glioma cell lines were treated with MLN8237 at varying concentrations for 12 days. Most of the CDC20-M high glioma cell lines were more sensitive to MLN8237 compared with CDC20-M low glioma cell lines. Data shown are averages of three independent experiments. Each concentration was assessed in six replicates in one experiment, and error bars indicate SD. (Right) The table shows the summary of IC50 concentrations of MLN8237 for all cell lines tested. **(B)** Quantification of cells showing polyploidization (with > 4N DNA content) for glioma cell lines cultured with DMSO (control) or MLN8237 for 72 h. Error bars in the plots represent the mean ± SD of three independent experiments. ***: $p < 0.001$; **: $p < 0.01$, and ns: $p > 0.05$, as compared to HA in each concentration; two-way ANOVA test. **(C)** Cell proliferative index (% of EdU-positive S phase cells) and polyploidization were assessed in N5, N9, PDX, N3 and HA cells with or without MLN8237 treatment at the indicated doses for three days. Results are representative of technical triplicates.

Fig. S19

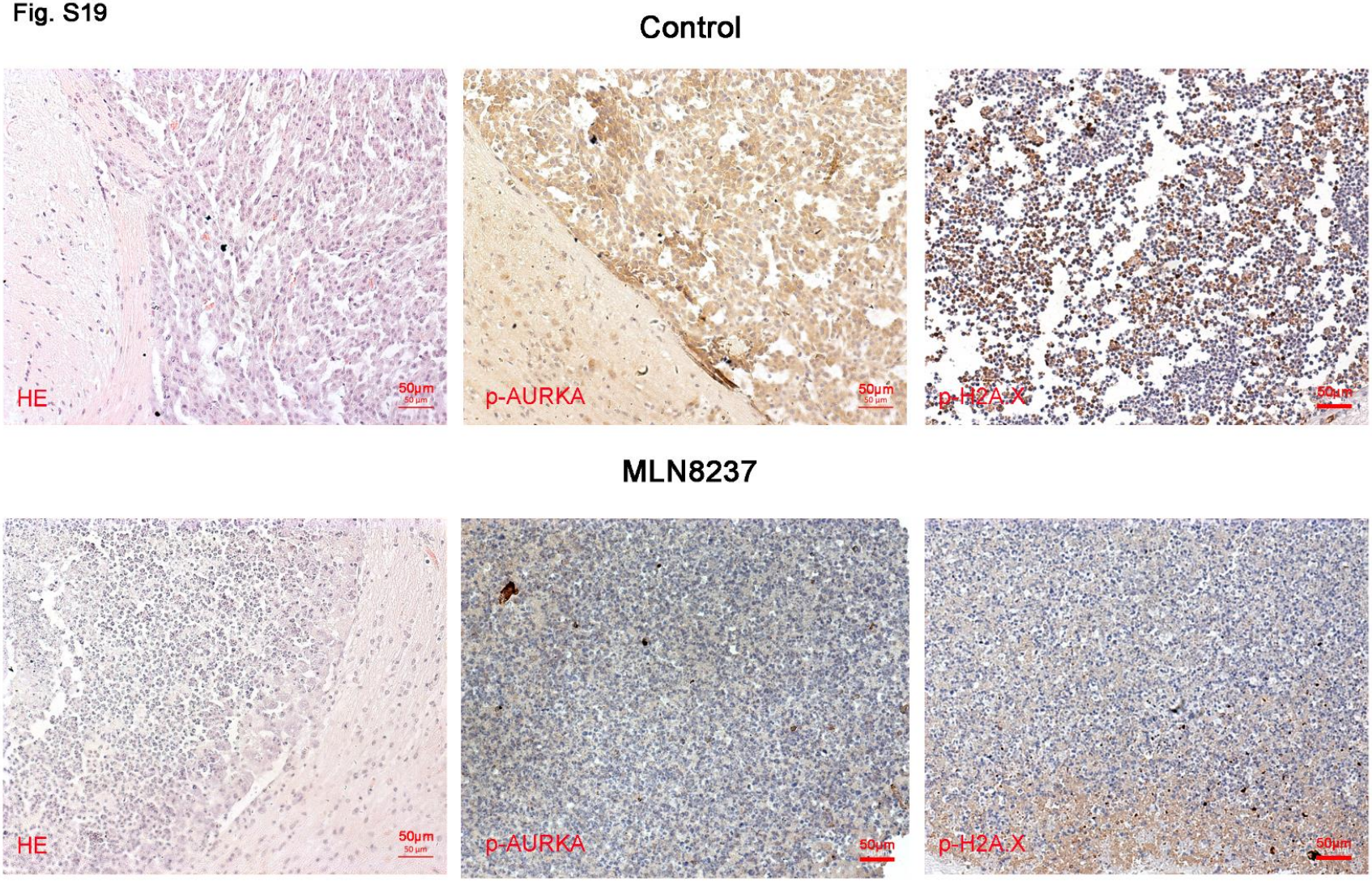

Control
HE
50μm
p-AURKA
50μm
p-H2A.X
50μm
MLN8237
HE
50μm
p-AURKA
50μm
p-H2A.X
50μm

Fig. S19. Reduced expression of p-AURKA and p-H2A.X in CDC20-M high PDX model following MLN8237 treatment.

Representative images of HE staining and immunohistochemical staining of p-AURKA and p-H2A.X in xenograft sections from mice treated with vehicle or MLN8237. Results are representatives of two independent experiments, with three xenografts in each group.

Table S1. Characteristics of the datasets used for the development and characterization of CDC20-M signature-based identification of GI in adult diffuse gliomas.

Dataset	Country	Platform	Samples
Rembrandt	America	HG-U133Plus2.0	521 glioma samples (GBM 228, Astrocytoma 148, Oligodendroglioma 67, Mixed 11, unknown 67), 21 epileptic samples
GSE16011	Holland	HG-U133Plus2.0	238 glioma samples (GBM 136, Astrocytoma 28, Oligodendroglioma 47, Oligoastrocytoma 27), 8 normal brain tissue samples
GSE4290	America	HG-U133Plus2.0	157 glioma samples (GBM 81, Astrocytoma 26, Oligodendroglioma 50), epileptic samples
TCGA training cohort	America	mRNA-seq	381 glioma samples (GBM 160, Astrocytoma 69, Oligodendroglioma 91, Oligoastrocytoma 60, unknown 1)
		SNP6.0	376 glioma samples (GBM 156, Astrocytoma 69, Oligodendroglioma 91, Oligoastrocytoma 60)
		Exome-seq	371 glioma samples (GBM 154, Astrocytoma 66, Oligodendroglioma 90, Oligoastrocytoma 60, unknown 1)
TCGA validation cohort	America	mRNA-seq	301 glioma samples (Astrocytoma 100, Oligodendroglioma 82, Oligoastrocytoma 54, unknown 65)
		SNP6.0	294 glioma samples (Astrocytoma 99, Oligodendroglioma 81, Oligoastrocytoma 54, unknown 60)
		Exome-seq	293 glioma samples (Astrocytoma 99, Oligodendroglioma 81, Oligoastrocytoma 54, unknown 59)
CGGA	China	mRNA-seq	319 glioma samples (GBM 140, Astrocytoma 67, Oligodendroglioma 40, Oligoastrocytoma 72)

The platforms, countries of origin, and sample composition of six glioma datasets used in our study are summarized.

Table S2. Members of CDC20-M listed according to the tightness of their co-expression to CDC20. The genes are listed in descending order according to their PCC to CDC20 expression.

CDC20-M was identified as the top 185 probe sets (containing 139 genes) most correlated to CDC20 in GSE4290. Members shared between CDC20-M and CIN70 are marked in red. The shared members of the 20 genes most co-expressed with CDC20 in GSE4290, GSE16011 and Rembrandt datasets are marked in yellow. Gene functions are categorized according to DAVID bioinformatics resources.

Gene Symbol	PCC with CDC20 (%)	Chromosome Location	Regulators of DNA repair and damage checkpoints	Regulators of cell cycle, centrosome function, spindle organization and microtubule-based movement
CDC20	100	1p34.1		✓
TYMS	86.28	18p11.32	✓	
NCAPH	84.86	2q11.2		✓
TACC3	84.69	4p16.3		✓
KIF2C	84.4	1p34.1		✓
NUF2	84.34	1q23.3		✓
KIAA0101	84.21	15q22.31	✓	
CENPF	83.8	1q41		✓
FOXM1	83.77	12p13	✓	✓
AURKA	83.69	20q13		✓
SMC4	83.42	3q26.1		✓
CCNB2	83.37	15q22.2		✓
PRC1	83	15q26.1		✓
SHCBP1	82.83	16q11.2		✓
RRM2	82.71	2p25-p24	✓	
CCNB1	82.7	5q12		✓
CDCA3	82.61	12p13		✓
NUSAP1	82.53	15q15.1		✓
PTTG1	82.27	5q35.1	✓	✓
BIRC5	82.2	17q25		✓
ASPM	82.15	1q31		✓
TPX2	82.13	20q11.2		✓
BUB1B	82.1	15q15		✓
MLF1IP	82.04	4q35.1		✓
CCNA2	81.99	4q27		✓
UBE2C	81.92	20q13.12		✓
GINS1	81.58	20p11.21	✓	
TIMELESS	81.5	12q13.3	✓	✓
FANCI	81.46	15q26.1	✓	✓
PCNA	81.36	20pter-p12	✓	
ECT2	81.12	3q26.1-q26.2		✓

Table S2 continued.

Gene Symbol	PCC with CDC20 (%)	Chromosome Location	Regulators of DNA repair and damage checkpoints	Regulators of cell cycle, centrosome function, spindle organization and microtubule-based movement
CENPE	80.84	4q24-q25		✓
KIF20A	80.73	5q31		✓
KIF14	80.7	1q32.1		✓
GAS2L3	80.21	12q23.1		✓
CENPA	80.21	2p23.3		✓
CDK1	80.17	10q21.1	✓	✓
CHEK1	80.01	11q24.2	✓	✓
CDCA8	79.99	1p34.3		✓
TTK	79.73	6q13-q21		✓
CDK2	79.56	12q13	✓	✓
E2F7	79.02	12q21.2	✓	✓
NCAPG	78.93	4p15.33		✓
CKS2	78.81	9q22		✓
HMGB2	78.58	4q31		✓
LRR1	78.34	14q21.3		
DTL	78.21	1q32	✓	✓
HMMR	78.19	5q33.2-qter		✓
SGOL2	77.98	2q33.1		✓
MAD2L1	77.97	4q27		✓
HJURP	77.89	2q37.1		✓
BUB1	77.83	2q14		✓
MCM2	77.82	3q21	✓	✓
PBK	77.8	8p21.2		✓
ESPL1	77.77	12q		✓
STIL	77.64	1p32		✓
TRIP13	77.57	5p15.33		✓
MELK	77.56	9p13.2		✓
NEK2	77.47	1q32.2-q41		✓
RNASEH2A	77.25	19p13.2	✓	
MKI67	77.25	10q26.2		✓
DLGAP5	77.18	14q22.3		✓
CENPK	76.45	5q12.3		✓
CDCA7L	76.37	7p15.3		✓
MCM6	76.36	2q21	✓	✓
CKAP2L	76.29	2q13		✓
TK1	76.24	17q23.2-q25.3		✓
MCM8	76.22	20p12.3	✓	✓
KIF11	76.07	10q24.1		✓

Table S2 continued.

Gene Symbol	PCC with CDC20 (%)	Chromosome Location	Regulators of DNA repair and damage checkpoints	Regulators of cell cycle, centrosome function, spindle organization and microtubule-based movement
CASC5	76.03	15q14		✓
GTSE1	75.96	22q13.2-q13.3		✓
DEPDC1	75.96	1p31.2		
CEP55	75.62	10q23.33		✓
DEPDC1B	75.58	5q12.1		✓
RBBP8	75.58	18q11.2	✓	✓
RACGAP1	75.46	12q13.12		✓
FAM64A	75.19	17p13.2		✓
SPAG5	75.14	17q11.2		✓
ZWINT	75.01	10q21-q22		✓
KPNA2	74.84	17q24.2		
LMNB1	74.61	5q23.2		✓
NCAPG2	74.56	7q36.3		✓
MCM3	74.46	6p12	✓	✓
CDC6	74.44	17q21.3	✓	✓
TOP2A	74.39	17q21-q22	✓	✓
RFC2	74.13	7q11.23	✓	✓
GINS2	74.09	16q24.1	✓	
CDC45	73.97	22q11.21	✓	✓
ASF1B	73.89	19p13.12	✓	✓
POLE2	73.85	14q21-q22	✓	
GJC1	73.75	17q21.31		
MYBL2	73.73	20q13.1		✓
DBF4	73.62	7q21.3	✓	✓
KIF15	73.61	3p21.31		✓
UBE2S	73.32	19q13.43		✓
SPC24	73.16	19p13.2		✓
NDC80	73.02	18p11.32		✓
EME1	72.99	17q21.33	✓	
IGF2BP3	72.97	7p11		
DNMT1	72.96	19p13.2	✓	
KIFC1	72.91	6p21.3		✓
CASP2	72.88	7q34-q35		
FBXO5	72.87	6q25.2	✓	✓
MND1	72.82	4q31.3	✓	
KIF4A	72.74	Xq13.1		✓
CDKN2C	72.69	1p32	✓	✓
C11orf82	72.55	11q14.1		✓

Table S2 continued.

Gene Symbol	PCC with CDC20 (%)	Chromosome Location	Regulators of DNA repair and damage checkpoints	Regulators of cell cycle, centrosome function, spindle organization and microtubule-based movement
DTYMK	72.29	2q37.3		✓
DSN1	72.2	20q11.23		✓
RFC4	72.16	3q27	✓	✓
NRM	72.12	6p21.33		
DDX39A	71.86	19p13.12		✓
PDIA4	71.67	7q35		
ZNF850	71.63	19q13.12		
AURKB	71.6	17p13.1		✓
ZNF765	71.58	19q13.42		
PARPBP	71.56	12q23.2	✓	✓
CDKN3	71.41	14q22	✓	✓
CDCA7	71.36	2q31		✓
BUD31	71.27	7q22.1		
KNTC1	71.24	12q24.31		✓
USP1	71.18	1p31.3	✓	
FANCD2	71.07	3p26	✓	✓
CDCA2	71.02	8p21.2		✓
SNRPB	70.96	20p13		
TMEM48	70.92	1p32.3		
LMNB2	70.75	19p13.3	✓	
PTBP1	70.59	19p13.3		
KIF23	70.56	15q23		✓
CDC25A	70.53	3p21	✓	✓
TCF3	70.38	19p13.3	✓	
BRCA1	70.32	17q21	✓	✓
GPX7	70.26	1p32	✓	
RAE1	70.22	20q13.31		✓
GTF2IRD2	70.2	7q11.23		
RFC3	70.18	13q13.2	✓	✓
CENPW	70.08	6q22.32		✓
POC1A	70.07	3p21.2		✓

Table S3. The top 120 genes most correlated to CREBRF expression in glioma transcriptome dataset GSE4290.
The genes are listed in descending order according to their PCC with CREBRF expression.

Gene Symbol	PCC with CREBRF (%)	Chromosome Location
CREBRF	100.00	5q35.1
IGIP	81.27	5q31
CBX7	80.66	22q13.1
MXI1	79.74	10q24-q25
ZBTB44	78.53	11q24.3
GTF2H5	78.03	6q25.3
DCUN1D5	77.86	11q22.3
ADRBK2	77.77	22q12.1
HLF	77.33	17q22
CHIC1	77.31	Xq13.2
SEC62	77.08	3q26.2
RPS6KA5	76.42	14q31-q32.1
NEBL	76.39	10p12
TOM1L2	76.16	17p11.2
TEF	76.11	22q13.2
JMY	76.02	5q14.1
FAIM2	75.99	12q13
KIAA0240	75.77	6p21.1
REPS2	75.72	Xp22.2
LOC283588	75.49	14q32.11
AKAP6	75.47	14q12
ZRANB1	75.47	10q26.13
YPEL3	75.45	16p11.2
MAPK10	75.39	4q22.1-q23
GRAMD1B	75.38	11q24.1
SCAPER	75.29	15q24
CPEB3	75.21	10q23.32
AGXT2L1	75.12	4q25
LOC283713	75.11	15q13.1
ZNF540	75.05	19q13.12
SEP8	75.02	5q31
SCAMP1	74.75	5q14.1
SGSM1	74.70	22q11.23
CRYZL1	74.64	21q21.3
KSR2	74.47	12q24.22-q24.23
HDAC4	74.44	2q37.3
PIAS1	74.41	15q
FSTL5	74.26	4q32.3

Table S3 continued.

Gene Symbol	PCC with CREBRF (%)	Chromosome Location
PPP1R12B	74.23	1q32.1
UQCRB	73.89	8q22
PPM1A	73.87	14q23.1
KIAA0368	73.87	9q31.3
FBXO9	73.84	6p12.3-p11.2
PURA	73.74	5q31
FRY	73.64	13q13.1
PPM1L	73.45	3q26.1
PLEKHM3	73.35	2q33.3
DNAJC12	73.27	10q22.1
GTDC1	73.27	2q22.3
C7orf41	73.10	7p14.3
SH3BGRL2	73.08	6q14.1
GABARAPL3///GABARAPL1	73.02	15q26.1///12p13.2
SLK	72.81	10q24.33
RELN	72.80	7q22
IKZF5	72.80	10q26
PPP1R3E	72.65	14q11.2
MYCBP2	72.65	13q22
FAM190B	72.61	10q23.1
FBXL20	72.53	17q12
BZRAP1	72.53	17q22-q23
NR1D2	72.47	3p24.2
NEGR1	72.37	1p31.1
HMGCLL1	72.35	6p12.1
EZH1	72.33	17q21.1-q21.3
ACADSB	72.30	10q26.13
PCGF5	72.19	10q23.32
RPL37	72.10	5p13
PELI3	72.10	11q13.2
RAB11FIP2	72.10	10q26.11
SPTAN1	72.03	9q34.11
HERC1	71.86	15q22
ADD3	71.85	10q25.2
TUB	71.77	11p15.5
GABARAPL1	71.71	12p13.2
KIAA1107	71.69	1p22.1
NAP1L3	71.65	Xq21.3-q22
MLLT6	71.61	17q21
PLA2G6	71.57	22q13.1
AVPI1	71.52	10q24.2

Table S3 continued.

Gene Symbol	PCC with CREBRF (%)	Chromosome Location
TTBK2	71.51	15q15.2
GARNL3	71.45	9q33.3
MINOS1	71.41	1p36.13
ANKRD46	71.36	8q22.2
GABRG1	71.35	4p12
SKP1	71.32	5q31
FBRSL1	71.32	12q24.33
OIP5-AS1	71.29	15q15.1
CDS2	71.15	20p13
TBRG1	71.12	11q24.2
USP46	71.10	4q12
LOC157562	71.07	8q22.3
PAPOLG	70.94	2p16.1
AKR1C1	70.87	10p15-p14
ZMYND11	70.82	10p14
ABLIM1	70.81	10q25
YPEL4	70.77	11q12.1
NALCN	70.70	13q32.3
NAP1L2	70.70	Xq13
CRY2	70.65	11p11.2
SERP2	70.65	13q14.11
RUNDC3A	70.62	17q21.31
NTRK2	70.62	9q22.1
SLC12A6	70.60	15q13
ZBTB4	70.46	17p13.1
ADARB2	70.44	10p15.3
CAB39L	70.37	13q14.2
MKX	70.36	10p12.1
ATP6V1G2	70.33	6p21.3
PPP1R3F	70.32	Xp11.23
POU6F1	70.32	12q13.13
KIAA1377	70.31	11q22.1
VPS13D	70.30	1p36.22
EIF1	70.26	17q21.2
RABGGTB	70.19	1p31
NCAM1	70.18	11q23.1
NCOA2	70.15	8q13.3
GABBR1	70.12	6p21.31
MAGEE1	70.12	Xq13.3
CCDC85A	70.05	2p16.1
ARHGAP32	70.04	11q24.3

Table S4. Results of multivariate Cox regression analysis for CIN70 score, age at the diagnosis, IDH1 mutation, 1p19q co-deletion, MKI67 score (the log2 RNA expression of Ki-67) in four large glioma datasets. CIN70 score was prognostic in the CGGA and REMBRANDT cohorts, but not in the other two cohorts.

Dataset	Variate	P value	Hazard ratio	95% CI for Hazard ratio		Dataset	Variate	P value	Hazard ratio	95% CI for Hazard ratio	
				Lower	Upper					Lower	Upper
TCGA (n = 363)	CIN70 score	.078	1.367	.965	1.937	**GSE16011 (n = 235)**	CIN70 score	.202	1.579	.782	3.186
	age	.000	1.057	1.042	1.072		age	.000	1.026	1.014	1.039
	IDH1 mutation	.000	.336	.196	.576		IDH1 mutation	.152	.766	.531	1.103
	1p19q co-deletion	.225	.626	.293	1.335		1p19q co-deletion	.000	.304	.184	.503
	MKI67 score	.531	.948	.804	1.119		KPS	.000	.970	.960	.980
							MKI67 score	.625	.901	.594	1.367
CGGA (n = 304)	CIN70 score	.000	2.375	1.496	3.770	**Rembrandt (n = 403)**					
	age	.161	1.012	.995	1.028		CIN70 score	.000	2.920	1.676	5.085
	IDH1 mutation	.000	.432	.280	.665		age	.000	1.280	1.216	1.348
	1p19q co-deletion	.001	.280	.136	.580		1p19q co-deletion	.015	.572	.364	.897
	MKI67 score	.656	.939	.712	1.239		MKI67 score	.020	.589	.378	.919

Table S5. Copy number and mutation status of core components in the TP53 pathway based on SNP6.0 and WES data of CDC20-M high glioma samples in TCGA training cohort.

Sample	TP53	CDKN2A	MDM2	MDM4	TP53 pathway changes
TCGA-06-5416	mutation	CN Gain+mutation	mutation	mutation	4
TCGA-06-2569	LOH+mutation		CN Gain	CN Gain	3
TCGA-26-5136	mutation	CN Gain	CN Loss		3
TCGA-14-1825	2 mutations	Homozygous Copy Loss			2
TCGA-19-1390	2 mutations	Homozygous Copy Loss			2
TCGA-02-2483	LOH+mutation	CN Loss			2
TCGA-E1-5304	LOH+mutation	CN Loss			2
TCGA-06-2567	mutation	CN Loss			2
TCGA-14-0817	mutation	CN Loss			2
TCGA-19-4065	mutation	CN Loss			2
TCGA-28-1753	mutation	CN Loss			2
TCGA-28-5215	mutation	CN Loss			2
TCGA-06-0744	mutation	Homozygous Copy Loss			2
TCGA-76-4925	mutation	Homozygous Copy Loss			2
TCGA-06-0129	mutation			mutation	2
TCGA-14-0871	mutation			CN Gain	2
TCGA-28-5218		CN Loss		CN Gain	2
TCGA-32-4213		CN Loss		High Copy Gain, LOH	2
TCGA-DU-6403		CN Loss		High Copy Gain	2
TCGA-FG-6692		CN Loss		CN Gain	2
TCGA-06-0187		Homozygous Copy Loss	High Copy Gain		2
TCGA-12-0821		Homozygous Copy Loss		High Copy Gain, LOH	2
TCGA-14-0787		Homozygous Copy Loss		CN Gain	2
TCGA-15-0742		Homozygous Copy Loss		CN Gain	2
TCGA-76-4932		Homozygous Copy Loss		CN Gain	2
TCGA-HT-8011		Homozygous Copy Loss		High Copy Gain	2

Table S5 continued.

Sample	TP53	CDKN2A	MDM2	MDM4	TP53 pathway changes
TCGA-02-2485	2 mutations				1
TCGA-06-2559	2 mutations				1
TCGA-41-5651	CN Loss				1
TCGA-26-5133	CN Loss+mutation				1
TCGA-06-0130	mutation				1
TCGA-06-0644	mutation				1
TCGA-06-5858	mutation				1
TCGA-12-0619	mutation				1
TCGA-14-1034	mutation				1
TCGA-19-1787	mutation				1
TCGA-28-2509	mutation				1
TCGA-32-2616	mutation				1
TCGA-06-0745		CN Loss			1
TCGA-06-5408		CN Loss			1
TCGA-06-5413		CN Loss			1
TCGA-06-5414		CN Loss			1
TCGA-12-5295		CN Loss			1
TCGA-27-1831		CN Loss			1
TCGA-27-1837		CN Loss			1
TCGA-27-2524		CN Loss			1
TCGA-27-2528		CN Loss			1
TCGA-28-5204		CN Loss			1
TCGA-28-5209		CN Loss			1
TCGA-76-4928		CN Loss			1
TCGA-DU-6406		CN Loss			1
TCGA-06-0125		Homozygous Copy Loss			1
TCGA-06-0211		Homozygous Copy Loss			1
TCGA-12-3650		Homozygous Copy Loss			1
TCGA-12-3652		Homozygous Copy Loss			1
TCGA-12-5299		Homozygous Copy Loss			1
TCGA-26-5134		Homozygous Copy Loss			1
TCGA-27-2523		Homozygous Copy Loss			1

Table S5 continued.

Sample	TP53	CDKN2A	MDM2	MDM4	TP53 pathway changes
TCGA-28-5208		Homozygous Copy Loss			1
TCGA-41-2572		Homozygous Copy Loss			1
TCGA-76-4931		Homozygous Copy Loss			1
TCGA-06-0157				High Copy Gain, LOH	1
TCGA-06-0174			High Copy Gain		1
TCGA-06-0686			High Copy Gain		1
TCGA-06-5856			High Copy Gain		1
TCGA-12-3653				CN Gain	1
TCGA-19-2624			High Copy Gain		1
TCGA-26-5135			High Copy Gain		1
TCGA-06-2557					0
gene altered in all	**40.6%**	**66.7%**	**13.0%**	**21.7%**	**98.6%**
pathway altered in all		**98.6%**			

Table S6. Copy number and mutation status of core components in the TP53 pathway based on SNP6.0 and WES data of CDC20-M intermediate glioma samples in TCGA training cohort.

Sample	TP53	CDKN2A	MDM2	MDM4	TP53 pathway changes
TCGA-27-2521	LOH+mutation	Homozygous Copy Loss	LOH		3
TCGA-06-0210	mutation	CN Loss		High Copy Gain, LOH	3
TCGA-28-5207	mutation		CN Gain	High Copy Gain	3
TCGA-DU-7010	mutation	CN Gain		CN Gain	3
TCGA-12-1597		CN Gain	CN Gain	High Copy Gain	3
TCGA-19-2625	2 mutations			CN Gain	2
TCGA-DU-6392	3 mutations	mutation			2
TCGA-CS-4943	LOH+mutation	Homozygous Copy Loss+mutation			2
TCGA-16-0846	mutation			CN Gain	2
TCGA-06-2558	mutation		LOH		2
TCGA-06-0743	mutation	Homozygous Copy Loss			2
TCGA-06-2563	mutation	Homozygous Copy Loss			2
TCGA-28-5216	mutation	Homozygous Copy Loss			2
TCGA-27-1830	mutation	CN Loss			2
TCGA-32-1970	mutation	CN Loss			2
TCGA-FG-5963	mutation	CN Loss			2
TCGA-HT-7606	mutation	CN Loss			2
TCGA-06-0750		CN Loss		High Copy Gain, LOH	2
TCGA-06-5411		CN Loss		High Copy Gain, LOH	2
TCGA-28-5220		Homozygous Copy Loss		High Copy Gain	2
TCGA-19-2619		Homozygous Copy Loss		CN Gain	2
TCGA-06-1804		CN Loss		CN Gain	2
TCGA-HT-8104		CN Loss		CN Gain	2
TCGA-27-2526		CN Loss	High Copy Gain		2

Table S6 continued.

Sample	TP53	CDKN2A	MDM2	MDM4	TP53 pathway changes
TCGA-14-0790		Homozygous Copy Loss	CN Gain		2
TCGA-32-2615		CN Loss, Homozygous Copy Loss	CN Gain		2
TCGA-06-2561		CN Loss	CN Gain		2
TCGA-06-5415		CN Loss	CN Gain		2
TCGA-14-1823		CN Loss	CN Gain		2
TCGA-14-2554		CN Loss	CN Gain		2
TCGA-06-5417	2 mutations				1
TCGA-76-4929	2 mutations				1
TCGA-DU-8165	2 mutations				1
TCGA-HT-7469	2 mutations				1
TCGA-27-1835	CN Loss+mutation				1
TCGA-32-2634	LOH				1
TCGA-02-0055	mutation				1
TCGA-06-0184	mutation				1
TCGA-06-0190	mutation				1
TCGA-06-0238	mutation				1
TCGA-06-2570	mutation				1
TCGA-12-0618	mutation				1
TCGA-27-2519	mutation				1
TCGA-41-3915	mutation				1
TCGA-DH-5140	mutation				1
TCGA-DU-7013	mutation				1
TCGA-06-0747		Homozygous Copy Loss			1
TCGA-32-2632				High Copy Gain, LOH	1
TCGA-06-0168				CN Gain	1
TCGA-06-0138			High Copy Gain		1
TCGA-19-0957			High Copy Gain		1
TCGA-41-2571			High Copy Gain		1
TCGA-19-5960			CN Loss		1
TCGA-06-2565		Homozygous Copy Loss			1
TCGA-28-1747		Homozygous Copy Loss			1

Table S6 continued.

Sample	TP53	CDKN2A	MDM2	MDM4	TP53 pathway changes
TCGA-28-2514		Homozygous Copy Loss			1
TCGA-DU-5852		Homozygous Copy Loss			1
TCGA-32-2638		CN Loss+mutation			1
TCGA-02-2486		CN Loss			1
TCGA-06-0158		CN Loss			1
TCGA-06-0219		CN Loss			1
TCGA-06-0649		CN Loss			1
TCGA-06-0878		CN Loss			1
TCGA-06-2562		CN Loss			1
TCGA-06-2564		CN Loss			1
TCGA-06-5418		CN Loss			1
TCGA-06-5859		CN Loss			1
TCGA-14-0789		CN Loss			1
TCGA-14-1402		CN Loss			1
TCGA-14-1829		CN Loss			1
TCGA-16-1045		CN Loss			1
TCGA-19-2620		CN Loss			1
TCGA-19-2629		CN Loss			1
TCGA-26-5139		CN Loss			1
TCGA-32-1982		CN Loss			1
TCGA-32-5222		CN Loss			1
TCGA-41-4097		CN Loss			1
TCGA-76-4926		CN Loss			1
TCGA-CS-6186		CN Loss			1
TCGA-DU-5847		CN Loss			1
TCGA-DU-6402		CN Loss			1
TCGA-DU-6405		CN Loss			1
TCGA-DU-7006		CN Loss			1
TCGA-06-0178		CN Gain			1
TCGA-02-0047					0
TCGA-06-0141					0
TCGA-06-5410					0
TCGA-06-5412					0
TCGA-08-0386					0
TCGA-14-0781					0
TCGA-19-1389					0
TCGA-27-1832					0
TCGA-28-2513					0
TCGA-28-5213					0
TCGA-DB-5274					0
gene altered in all	**33.7%**	**61.1%**	**15.8%**	**14.7%**	**88.4%**
pathway altered in all		**88.4%**			

Table S7. Copy number and mutation status of core components in the TP53 pathway based on SNP6.0 and WES data of CREBRF-M intermediate glioma samples in TCGA training cohort.

Sample	TP53	CDKN2A	MDM2	MDM4	TP53 pathway changes
TCGA-HT-7475	2 mutations	CN Loss		CN Gain	3
TCGA-IK-7675	mutation	CN Loss	CN Loss		3
TCGA-DB-5281	2 mutations	CN Loss			2
TCGA-DB-5277	CN Gain, LOH+mutation	CN Loss			2
TCGA-E1-5305	mutation		CN Gain		2
TCGA-06-0221	mutation	CN Loss			2
TCGA-DU-6407	mutation	CN Loss			2
TCGA-DU-7301	mutation	CN Loss			2
TCGA-FG-5965	mutation	CN Loss			2
TCGA-HT-7476	mutation	CN Loss			2
TCGA-CS-5395		CN Loss		High Copy Gain	2
TCGA-27-1834		CN Loss		High Copy Gain, LOH	2
TCGA-DU-7292		CN Loss		High Copy Gain, LOH	2
TCGA-FG-8185		CN Loss	mutation		2
TCGA-DH-5143	2 mutations				1
TCGA-DU-7298	2 mutations				1
TCGA-DU-7304	2 mutations				1
TCGA-DU-8167	2 mutations				1
TCGA-FG-6690	2 mutations				1
TCGA-FG-8182	2 mutations				1
TCGA-FG-8188	2 mutations				1
TCGA-FN-7833	2 mutations				1
TCGA-HT-7604	2 mutations				1
TCGA-HT-7879	2 mutations				1
TCGA-HT-8013	2 mutations				1
TCGA-HT-8114	2 mutations				1
TCGA-HT-8563	2 mutations				1
TCGA-HT-8564	2 mutations				1
TCGA-HW-8319	2 mutations				1
TCGA-DB-5275	LOH+mutation				1
TCGA-CS-4942	mutation				1
TCGA-CS-5396	mutation				1
TCGA-CS-6665	mutation				1
TCGA-DB-5273	mutation				1

Table S7 continued.

Sample	TP53	CDKN2A	MDM2	MDM4	TP53 pathway changes
TCGA-DH-5142	mutation				1
TCGA-DU-5872	mutation				1
TCGA-DU-6395	mutation				1
TCGA-DU-6396	mutation				1
TCGA-DU-6399	mutation				1
TCGA-DU-6408	mutation				1
TCGA-DU-6542	mutation				1
TCGA-DU-7007	mutation				1
TCGA-DU-7008	mutation				1
TCGA-DU-7015	mutation				1
TCGA-DU-8163	mutation				1
TCGA-DU-8166	mutation				1
TCGA-E1-5302	mutation				1
TCGA-E1-5303	mutation				1
TCGA-E1-5322	mutation				1
TCGA-FG-6691	mutation				1
TCGA-FG-7636	mutation				1
TCGA-HT-7472	mutation				1
TCGA-HT-7478	mutation				1
TCGA-HT-7483	mutation				1
TCGA-HT-7601	mutation				1
TCGA-HT-7676	mutation				1
TCGA-HT-7686	mutation				1
TCGA-HT-7690	mutation				1
TCGA-HT-7858	mutation				1
TCGA-HT-7873	mutation				1
TCGA-HT-8106	mutation				1
TCGA-HT-8108	mutation				1
TCGA-HT-8111	mutation				1
TCGA-HW-8320	mutation				1
TCGA-FG-6688				High Copy Gain	1
TCGA-CS-5394				LOH	1
TCGA-CS-4941			High Copy Gain, LOH		1
TCGA-06-0171		CN Loss			1
TCGA-06-0749		CN Loss			1
TCGA-06-0882		CN Loss			1
TCGA-15-1444		CN Loss			1
TCGA-CS-5397		CN Loss			1

Table S7 continued.

Sample	TP53	CDKN2A	MDM2	MDM4	TP53 pathway changes
TCGA-DU-5854		CN Loss			1
TCGA-DU-7012		CN Loss			1
TCGA-DU-8158		CN Loss			1
TCGA-DU-8168		CN Loss			1
TCGA-FG-5964		CN Loss			1
TCGA-DU-6393		LOH			1
TCGA-28-2510					0
TCGA-32-1980					0
TCGA-CS-5390					0
TCGA-CS-6666					0
TCGA-CS-6668					0
TCGA-DU-5870					0
TCGA-DU-5874					0
TCGA-DU-6394					0
TCGA-DU-6397					0
TCGA-DU-6404					0
TCGA-DU-6410					0
TCGA-DU-7009					0
TCGA-DU-7014					0
TCGA-DU-7018					0
TCGA-DU-7290					0
TCGA-DU-8161					0
TCGA-DU-8164					0
TCGA-E1-5318					0
TCGA-FG-7634					0
TCGA-FG-7637					0
TCGA-FG-7638					0
TCGA-HT-7467					0
TCGA-HT-7468					0
TCGA-HT-7471					0
TCGA-HT-7481					0
TCGA-HT-7482					0
TCGA-HT-7616					0
TCGA-HT-7677					0
TCGA-HT-7687					0
TCGA-HT-7692					0
TCGA-HW-8322					0
TCGA-IK-8125					0
gene altered in all	**54.5%**	**21.8%**	**3.6%**	**5.5%**	**70.9%**
pathway altered in all		**70.9%**			

Table S8. Copy number and mutation status of core components in the TP53 pathway based on SNP6.0 and WES data of CREBRF-M high glioma samples in TCGA training cohort.

Sample	TP53	CDKN2A	MDM2	MDM4	TP53 pathway changes
TCGA-DU-5855	mutation	CN Loss		mutation	3
TCGA-26-1442	mutation	CN Loss			2
TCGA-DU-7019	mutation	CN Loss			2
TCGA-HT-7474	mutation	CN Gain			2
TCGA-HT-7688	mutation			CN Gain	2
TCGA-CS-6667	2 mutations				1
TCGA-DB-5270	2 mutations				1
TCGA-HT-7609	2 mutations				1
TCGA-HT-7610	2 mutations				1
TCGA-HT-7611	2 mutations				1
TCGA-HT-7602	3 mutations				1
TCGA-DU-6401	CN Loss+mutation				1
TCGA-CS-4938	mutation				1
TCGA-CS-5393	mutation				1
TCGA-CS-6290	mutation				1
TCGA-DB-5276	mutation				1
TCGA-DB-5280	mutation				1
TCGA-DU-5871	mutation				1
TCGA-DU-7011	mutation				1
TCGA-DU-7299	mutation				1
TCGA-DU-7306	mutation				1
TCGA-DU-7309	mutation				1
TCGA-E1-5307	mutation				1
TCGA-FG-6689	mutation				1
TCGA-FG-8191	mutation				1
TCGA-HT-7470	mutation				1
TCGA-HT-7473	mutation				1
TCGA-HT-7479	mutation				1
TCGA-HT-7485	mutation				1
TCGA-HT-7603	mutation				1
TCGA-HT-7689	mutation				1
TCGA-HT-7693	mutation				1
TCGA-HT-7855	mutation				1
TCGA-HT-7880	mutation				1
TCGA-HT-7902	mutation				1
TCGA-HT-8018	mutation				1
TCGA-HT-8105	mutation				1
TCGA-HW-7489	mutation				1

Table S8 continued.

Sample	TP53	CDKN2A	MDM2	MDM4	TP53 pathway changes
TCGA-HW-7490	mutation				1
TCGA-HW-8321	mutation				1
TCGA-14-0736		CN Loss			1
TCGA-DH-5144		LOH			1
TCGA-DU-7302		CN Loss			1
TCGA-EZ-7264		CN Loss			1
TCGA-06-0132					0
TCGA-CS-4944					0
TCGA-CS-6669					0
TCGA-CS-6670					0
TCGA-DB-5278					0
TCGA-DB-5279					0
TCGA-DH-5141					0
TCGA-DU-5849					0
TCGA-DU-5851					0
TCGA-DU-6400					0
TCGA-DU-7294					0
TCGA-DU-8162					0
TCGA-E1-5311					0
TCGA-E1-5319					0
TCGA-FG-5962					0
TCGA-FG-7641					0
TCGA-FG-7643					0
TCGA-FG-8181					0
TCGA-FG-8186					0
TCGA-FG-8187					0
TCGA-HT-7480					0
TCGA-HT-7605					0
TCGA-HT-7607					0
TCGA-HT-7608					0
TCGA-HT-7620					0
TCGA-HT-7680					0
TCGA-HT-7681					0
TCGA-HT-7684					0
TCGA-HT-7691					0
TCGA-HT-7694					0
TCGA-HT-7695					0
TCGA-HT-7854					0
TCGA-HT-7874					0
TCGA-HT-7875					0
TCGA-HT-7877					0

Table S8 continued.

Sample	TP53	CDKN2A	MDM2	MDM4	TP53 pathway changes
TCGA-HT-7884					0
TCGA-HT-8010					0
TCGA-HT-8113					0
TCGA-HT-8558					0
TCGA-HW-7486					0
TCGA-HW-7487					0
TCGA-HW-7491					0
TCGA-HW-7493					0
TCGA-HW-7495					0
gene altered in all	**45.5%**	**9.1%**	**0.0%**	**2.3%**	**50.0%**
pathway altered in all			**50.0%**		

Table S9. The patient information of each glioma cell line

Cell line	Histology	Grade	Gender	Age
PDX	GBM	Ⅳ	M	46
N5	GBM	Ⅳ	F	49
N9	GBM	Ⅳ	M	53
N33	GBM	Ⅳ	M	34
N3	AOA	Ⅲ	M	31
N8	OA	Ⅱ	M	47

GBM, glioblastoma multiforme; AOA, anaplastic oligoastrocytoma; OA, oligoastrocytoma; F, female; M, male.

Table S10. Frequent CNA and mutation in CDC20-M members in gliomas with high or intermediate CDC20-M signature.

TCGA Samples / CDC20-M members status	CDC20-M high or intermediate	CREBRF-M intermediate	CREBRF-M high	P value
Percentage of samples harboring CNA	93.5% (159/170)	77.5% (86/111)	39.4% (37/94)	**< 0.0001**
Number of CNAs per changed sample (mean ± SD)	18 ± 13	9 ± 9	4 ± 5	**< 0.0001**
Percentage of samples harboring mutations	24.7% (42/170)	10.8% (12/111)	14.9% (14/94)	0.0081
Number of mutations per changed sample (mean ± SD)	1.7 ± 1.5	1.3 ± 0.7	1.2 ± 0.4	0.3497

Copy number amplifications (CNA) and nonsynonymous mutations in the members of CDC20-M were analysed according to CDC20-M/CREBRF-M clustering in the TCGA training cohort. Chi-square test was used to compare the percentages of samples harboring CNA or mutations, and one-way ANOVA test was used to compare the number of CNAs or mutations per sample.

Table S11. Summary of CNAs and mutations in the CDC20-M members in TCGA training cohort classified according to the CDC20-M/CREBRF-M clustering.

CDC20-M members	CDC20-M high or intermediate (170)		CREBRF-M intermediate (111)		CREBRF-M high (94)	
	Number of samples with amplification	Number of samples with mutation	Number of samples with amplification	Number of samples with mutation	Number of samples with amplification	Number of samples with mutation
ASF1B	49	1	8		2	
ASPM	21	5	6	1	1	1
AURKA	45		10	1	2	
AURKB	2		1		0	
BIRC5	11		4		0	
BRCA1	6	4	1		0	
BUB1	6	1	0	1	0	
BUB1B	2		0		1	
BUD31	126		36		10	
CASC5	0	1	0	1	0	1
CASP2	124	1	40		12	1
CCNA2	1		0		2	
CCNB1	5		1		0	
CCNB2	3		2		1	
CDC20	15		4		0	
CDC25A	8		3		0	
CDC45	6		3		0	
CDC6	4		1		0	
CDCA2	8		10		1	
CDCA3	16		20		6	
CDCA7	7		0		0	
CDCA7L	119	1	24		4	
CDCA8	16		4		0	
CDK1	0		1		0	
CDK2	13		2		0	
CDKN2C	15	4	4		0	1
CDKN3	2		1		0	1
CENPA	7	1	0		0	
CENPE	0	3	1		0	
CENPF	19	3	5		1	1
CENPK	4	1	0	1	0	
CENPN	2		1		0	
CENPU	2		0		0	
CENPW	1		1		0	
CEP55	0		0		0	
CHEK1	3	1	16		10	

Table S11 continued.

CDC20-M members	CDC20-M high or intermediate (170)		CREBRF-M intermediate (111)		CREBRF-M high (94)	
	Number of samples with amplification	Number of samples with mutation	Number of samples with amplification	Number of samples with mutation	Number of samples with amplification	Number of samples with mutation
CKAP2L	6		0		0	
CKS2	14		5		0	1
DBF4	128		33		6	
DDIAS	4		10		5	
DDX39A	50		8		2	
DEPDC1	17		4		0	
DEPDC1B	4	1	0		0	
DLGAP5	2	1	1		0	
DNMT1	52		6		2	1
DSN1	44		9		1	
DTL	20		5		1	
DTYMK	7	1	1		1	
E2F7	10	1	2		0	
ECT2	12		4	1	2	
EME1	6		4		0	
ESPL1	14	2	2		0	1
FAM64A	1	1	1		0	1
FANCD2	12	1	6		1	
FANCI	4	1	3		1	
FBXO5	3		1		0	
FOXM1	17	1	18		5	1
GAS2L3	9		3		0	
GINS1	42		9		2	
GINS2	2		1		0	
GJC1	6		1		0	
GPX7	16		4		0	
GTF2IRD2	127		27		3	
GTSE1	3	2	2		0	
HJURP	5		0	1	0	
HMGB2	2		0		0	
HMMR	6		0		0	
IGF2BP3	120		24		4	
KIAA0101	3		1		1	
KIF11	0		0		0	
KIF14	19		5		2	
KIF15	8	2	3		0	
KIF20A	6	1	0	1	0	
KIF23	3		1		1	1

Table S11 continued.

CDC20-M members	CDC20-M high or intermediate (170)		CREBRF-M intermediate (111)		CREBRF-M high (94)	
	Number of samples with amplification	Number of samples with mutation	Number of samples with amplification	Number of samples with mutation	Number of samples with amplification	Number of samples with mutation
KIF2C	15		4		0	
KIF4A	0	1	0		0	
KIFC1	2		2		0	
KNTC1	11		7		0	
KPNA2	8		4		0	
LMNB1	6		1		0	
LMNB2	53		9	1	2	
LRR1	2		1		0	
MAD2L1	1		0		0	
MCM2	9	1	1		0	
MCM3	0	1	2		0	
MCM6	7		0		0	1
MCM8	45		9		2	
MELK	12	1	2		2	
MKI67	0	3	0	1	0	1
MND1	2	1	0		0	
MYBL2	43		8		2	
NCAPG	2		1		0	
NCAPG2	125		36		11	
NCAPH	7		0		0	
NDC1	16	1	4		0	
NDC80	9		2		3	
NEK2	20		5		1	
NRM	0		4		0	
NUF2	20		8		1	
NUSAP1	3		1		1	
PARPBP	9		3		0	
PBK	8		11		1	
PCNA	45		8		2	
PDIA4	124		38		12	
POC1A	10		3	1	0	
POLE2	2	1	1		0	
PRC1	5		3		1	
PTBP1	54		14		3	
PTTG1	6		0		1	
RACGAP1	14		3		0	
RAE1	45		10		2	
RBBP8	10		0		1	1

Table S11 continued.

CDC20-M members	CDC20-M high or intermediate (170)		CREBRF-M intermediate (111)		CREBRF-M high (94)	
	Number of samples with amplification	Number of samples with mutation	Number of samples with amplification	Number of samples with mutation	Number of samples with amplification	Number of samples with mutation
RFC2	129		26		3	
RFC3	4		1		1	
RFC4	16	1	3		1	
RNASEH2A	49	1	7		2	
RRM2	7		2		0	
SGO2	6	3	1		0	
SHCBP1	7		2		1	
SMC4	10		2	1	0	1
SNRPB	44	1	10		2	
SPAG5	4	3	1		0	
SPC24	53		7		2	
STIL	16	1	4		0	
TACC3	10		2		1	
TCF3	52	1	9		2	
TIMELESS	15	1	3	2	0	
TK2	2		1		0	
TOP2A	5	1	1		0	
TPX2	45	1	9		1	1
TRIP13	6		2	1	1	
TTK	1	1	0		0	
TYMS	10		4		3	
UBE2C	45		9		2	
UBE2S	42		3		0	
USP1	17		4		0	
ZNF765	40		4		0	
ZNF850	46		3		1	
ZWINT	0	1	2		0	

Supplementary References

1. Sun L, *et al.* (2006) Neuronal and glioma-derived stem cell factor induces angiogenesis within the brain. *Cancer cell* 9(4):287-300.
2. Gravendeel LA, *et al.* (2009) Intrinsic gene expression profiles of gliomas are a better predictor of survival than histology. *Cancer research* 69(23):9065-9072.
3. Madhavan S, *et al.* (2009) Rembrandt: helping personalized medicine become a reality through integrative translational research. *Mol Cancer Res* 7(2):157-167.
4. Kang HJ, *et al.* (2011) Spatio-temporal transcriptome of the human brain. *Nature* 478(7370):483-489.
5. Florio M, *et al.* (2015) Human-specific gene ARHGAP11B promotes basal progenitor amplification and neocortex expansion. *Science (New York, N.Y* 347(6229):1465-1470.
6. Cahoy JD, *et al.* (2008) A transcriptome database for astrocytes, neurons, and oligodendrocytes: a new resource for understanding brain development and function. *J Neurosci* 28(1):264-278.
7. Novershtern N, *et al.* (2011) Densely interconnected transcriptional circuits control cell states in human hematopoiesis. *Cell* 144(2):296-309.
8. de Graaf CA, *et al.* (2016) Haemopedia: An Expression Atlas of Murine Hematopoietic Cells. *Stem Cell Reports* 7(3):571-582.
9. Kim SK, *et al.* (2001) A gene expression map for Caenorhabditis elegans. *Science* 293(5537):2087-2092.
10. Janic A, Mendizabal L, Llamazares S, Rossell D, & Gonzalez C (2010) Ectopic expression of germline genes drives malignant brain tumor growth in Drosophila. *Science* 330(6012):1824-1827.
11. Bao ZS, *et al.* (2014) RNA-seq of 272 gliomas revealed a novel, recurrent PTPRZ1-MET fusion transcript in secondary glioblastomas. *Genome Res* 24(11):1765-1773.
12. Su G, Morris JH, Demchak B, & Bader GD (2014) Biological network exploration with Cytoscape 3. *Current protocols in bioinformatics* 47:8 13 11-24.
13. Szklarczyk D, *et al.* (2017) The STRING database in 2017: quality-controlled protein-protein association networks, made broadly accessible. *Nucleic acids research* 45(D1):D362-D368.
14. Monti ST, P.; Mesirov, J.; Golub, T. (2003) Consensus Clustering: A Resampling-Based Method for Class Discovery and Visualization of Gene Expression Microarray Data. *Machine Learning* 52:91-118.
15. Mermel CH, *et al.* (2011) GISTIC2.0 facilitates sensitive and confident localization of the targets of focal somatic copy-number alteration in human cancers. *Genome Biol* 12(4).
16. Cibulskis K, *et al.* (2013) Sensitive detection of somatic point mutations in impure and heterogeneous cancer samples. *Nature biotechnology* 31(3):213-219.
17. Mcgill R, Tukey JW, & Larsen WA (1978) Variations of Box Plots. *Am Stat* 32(1):12-16.

18. Frigge M, Hoaglin DC, & Iglewicz B (1989) Some Implementations of the Boxplot. *Am Stat* 43(1):50-54.
19. Subramanian A*, et al.* (2005) Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles. *Proceedings of the National Academy of Sciences of the United States of America* 102(43):15545-15550.
20. Di Stefano AL*, et al.* (2015) Detection, Characterization, and Inhibition of FGFR-TACC Fusions in IDH Wild-type Glioma. *Clinical Cancer Research* 21(14):3307-3317.