Genomics Proteomics Bioinformatics 17 (2019) 465–471

Genomics Proteomics Bioinformatics

www.elsevier.com/locate/gpb
www.sciencedirect.com

DATABASE

GliomaDB: A Web Server for Integrating Glioma Omics Data and Interactive Analysis

Yadong Yang [1,2,a], **Yang Sui** [1,2,b], **Bingbing Xie** [1,2,c], **Hongzhu Qu** [1,2,*,d] **Xiangdong Fang** [1,2,*,e]

[1] *CAS Key Laboratory of Genome Sciences and Information, Beijing Institute of Genomics, Chinese Academy of Sciences, Beijing 100101, China*
[2] *University of Chinese Academy of Sciences, Beijing 100049, China*

Received 9 February 2018; revised 22 March 2018; accepted 31 March 2018
Available online 5 December 2019

Handled by Hsien-Da Huang

KEYWORDS

Database;
Variations;
Methylation;
Network;
Survival analysis

Abstract Gliomas are one of the most common types of brain cancers. Numerous efforts have been devoted to studying the mechanisms of glioma genesis and identifying biomarkers for diagnosis and treatment. To help further investigations, we present a comprehensive **database** named GliomaDB. GliomaDB includes 21,086 samples from 4303 patients and integrates genomic, transcriptomic, epigenomic, clinical, and gene-drug association data regarding glioblastoma multiforme (GBM) and low-grade glioma (LGG) from The Cancer Genome Atlas (TCGA), Gene Expression Omnibus (GEO), the Chinese Glioma Genome Atlas (CGGA), the Memorial Sloan Kettering Cancer Center Integrated Mutation Profiling of Actionable Cancer Targets (MSK-IMPACT), the US Food and Drug Administration (FDA), and PharmGKB. GliomaDB offers a user-friendly interface for two main types of functionalities. The first comprises queries of (i) somatic mutations, (ii) gene expression, (iii) microRNA (miRNA) expression, and (iv) DNA **methylation**. In addition, queries can be executed at the gene, region, and base level. Second, GliomaDB allows users to perform **survival analysis**, coexpression **network** visualization, multi-omics data visualization, and targeted drug recommendations based on personalized **variations**. GliomaDB bridges the gap between

* Corresponding authors.
E-mail: quhongzhu@big.ac.cn (Qu H), fangxd@big.ac.cn (Fang X).
[a] ORCID: 0000-0003-2936-1574.
[b] ORCID: 0000-0003-1728-1593.
[c] ORCID: 0000-0002-8573-442X.
[d] ORCID: 0000-0001-7013-8409.
[e] ORCID: 0000-0002-6628-8620.

Peer review under responsibility of Beijing Institute of Genomics, Chinese Academy of Sciences and Genetics Society of China.

https://doi.org/10.1016/j.gpb.2018.03.008

glioma genomics big data and the delivery of integrated information for end users, thus enabling both researchers and clinicians to effectively use publicly available data and empowering the progression of precision medicine in glioma. GliomaDB is freely accessible at http://bigd.big.ac.cn/gliomaDB.

Introduction

Gliomas are the most common form of brain cancers and can be classified as Grade I–IV based on standards set by the World Health Organization (WHO). Grade I, II, and III gliomas are usually considered low-grade glioma (LGG), whereas Grade IV tumors are frequently termed high-grade glioma, which is also known as glioblastoma multiforme (GBM) (https://cancergenome.nih.gov/cancersselected/lowergradeglioma). In the United States of America, there were 23,820 estimated new cases and 17,760 estimated new deaths owing to diseases of the brain and nervous system in 2019 [1]. Glioma is one of the deadliest forms of human cancers, with a 5-year relative survival of 33% [2], and for GBM patients in particular, the median duration of survival is estimated to be 14 months after maximal surgical resection, radiotherapy, and chemotherapy [3].

Recent years have witnessed the rapid development of high-throughput technology, including microarray and next-generation sequencing. For example, The Cancer Genome Atlas (TCGA) [4–6] has been assembled from thousands of glioma cancer and noncancer samples. In addition, an enormous amount of data from independent studies has been deposited into Gene Expression Omnibus (GEO) [7,8]; both of these data aggregates provide an unprecedented opportunity for glioma research. For example, genomic profiling could be used to separate primary and secondary GBM, which are otherwise indistinguishable histologically [9,10]. Single cell sequencing technologies have been utilized for identifying tumor initiating cells in glioma and presenting a paradigm for interpretation of intra-tumor heterogeneity and personalized therapy [11]. *ATRX* has been associated with increased telomere length based on whole-genome data analysis; glioma molecular classification by *IDH* mutation status and 1p/19q codeletion were identified using clinically relevant molecular subsets [6]. Despite advances in glioma research, most studies use only a limited number of datasets because of insufficient ready-to-use resources. Moreover, the highly dispersed nature of data resources hindered the progression of precision medicine. Hence, an integrated database must be urgently established for the storage, retrieval, and analysis of big data in glioma.

In the recent past, several databases have been developed for the storage and analysis of big data in cancer. Some of the databases focus on pan-cancer expression analysis; for example, Gene Expression Profiling Interactive Analysis (GEPIA, http://gepia.cancer-pku.cn) [12] provides RNA sequencing (RNA-seq) data from 9736 tumors and 8587 normal samples from the TCGA and the Genotype-Tissue Expression (GTEx) projects and offers tools for differential analysis, similar gene analysis, correlation analysis, and dimensionality reduction. Cancer RNA-seq Nexus (http://syslab4.nchu.edu.tw/) [13] provides 28 types of cancer RNA-seq data from the TCGA and GEO. Moreover, this database provides functionalities for the differential analysis of genes and long noncoding RNAs (lncRNAs) as well as mRNA-lncRNA coexpression network analysis. Other databases specifically focus on glioma; some examples are given as follows. (1) The diffuse low-grade glioma (DLGG) database (http://db-gliomas-gradeii.net/) [14] provides 210 different fluid-attenuated inversion recovery (FLAIR) magnetic resonance (MR) images of DLGG patients at different levels of evolution and the tools for the analysis of clinical images. (2) GLIOMASdb (http://cgga.org.cn:9091/gliomasdb/) [15] provides RNA-seq data of 325 gliomas at different stages with different subtypes and identified progression-associated genes. (3) Xena (http://xena.ucsc.edu/) [16] provides numerous useful visualization and analysis tools for deposited omics data and secure analysis and visualization of private functional genomics data. (4) cBioPortal (http://www.cbioportal.org/) [17] provides simultaneous visualization of multiple types of genomic data from multiple data sources.

Although these databases or web tools provide abundant resources for the glioma scientific and clinical community, many additional features or functions that are often required by biologists and clinicians are not appropriately addressed by these tools and databases. For example, most of the dispersive datasets in GEO are not included in GEPIA, Cancer RNA-seq Nexus, DLGG, GLIOMASdb, Xena, or cBioPortal, which limits the data usage. While Cancer RNA-seq Nexus supports the query of a specific gene or lncRNA for the coexpression network, this database contains only the connections of the query gene/lncRNA but not the connections between all nodes. Moreover, while cBioPortal provides mutation annotation from OncoKB [18], CIViC [19], and My Cancer Genome [20], there lacks information on the mutation status for healthy populations and lacks customizable annotation of mutations for targeted drug recommendations. In addition, although GEPIA provides survival analysis based on gene expression profiles, the analysis is based on a single variable; an option to consider two or more variables (genes) is not available. Furthermore, none of these databases provide miRNA expression or DNA methylation profiles. Hence, to mitigate the aforementioned problems, we developed GliomaDB, an integrative database for glioma-related data, to complement the existing databases and web tools.

Implementation

GliomaDB codes were developed using an integrated development environment, Eclipse (http://www.eclipse.org). MySQL (https://www.mysql.com) is used to store and manage the metadata information of this database. For database connection and operation, MyBatis (http://www.mybatis.org) is used as a persistence framework. Spring (http://www.springsource.org) is used for the inversion of control containers. Java Server Pages is used to render the dynamic front pages. Struts (http://struts.apache.org) is used to manage the model-view-controller model web application. GliomaDB is hosted on a CentOS operating system with two servers, with Tomcat (http://

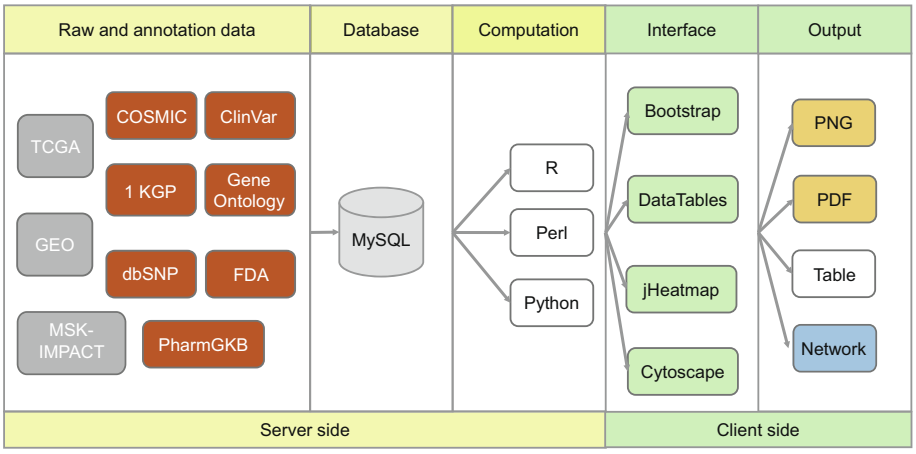


Figure 1 Scheme describing data processing, storage, and display for the GliomaDB visualization tool
Raw and annotation data from 10 public databases were stored in GliomaDB and then computed or analyzed using our in-house scripts, with outputs visualized in figures or tables. TCGA, The Cancer Genome Atlas; GEO, Gene Expression Omnibus; MSK-IMPACT, Memorial Sloan Kettering-Integrated Mutation Profiling of Actionable Cancer Targets; 1 KGP, 1000 Genomes Project; COSMIC, Catalogue of Somatic Mutations in Cancer; FDA, Food and Drug Administration.

tomcat.apache.org/) serving static and dynamic content and a MySQL (version 5.6.19) server providing the features for database management. All the plotting features in GliomaDB are implemented using R (version 3.4.2), Perl (version v5.10.1), and Python (version 2.7.12). The tables for this database are generated using DataTables (https://www.datatables.net) JavaScript library. The interactive heatmap and network are visualized with jHeatmap (https://jheatmap.github.io/jheatmap/) and Cytoscape (http://js.cytoscape.org/) JavaScript library, respectively (**Figure 1**).

Database content and usage

Database structure and organization

GliomaDB comprises four modules: search, analysis, team introduction, and statistics. The search module includes four aspects: genomic mutation, gene expression, miRNA expression, and DNA methylation. The analysis module contains four analytical perspectives: survival analysis, coexpression network visualization, cluster analysis, and variant-based targeted drug recommendation.

Data sources

Genomic variants, gene expression, miRNA expression, DNA methylation and clinical data of glioma patients were integrated from the TCGA (https://cancergenome.nih.gov/), MSK-IMPACT Clinical Sequencing Cohort [21], GEO (https://www.ncbi.nlm.nih.gov/geo/), and CGGA (http://www.cgga.org.cn/) projects; these data include tumor/normal tissue and blood samples. For the glioma sample, we selected only data from brain tissue, excluding data from cell lines, and all published data should be from after 2005. To continuously update the data, we developed a tool based on Entrez Programming Utilities (E-Utils) provided by the National Center for Biotechnology Information (NCBI) to automatically search for the newly updated datasets (see "Update" section in the "tutorial" page). For the genomic variants, we also integrated annotations from public resources, such as gene ontology (GO), the Catalogue Of Somatic Mutations In Cancer (COSMIC, http://cancer.sanger.ac.uk/cosmic/), the mutation frequency in different populations in the 1000 Genomes Project (https://www.genome.gov/27528684/1000-genomes-project/), and context information on the mutation in ClinVar (https://www.ncbi.nlm.nih.gov/clinvar/). Drug-responsive gene/variant data were collected from the Food and Drug Administration (FDA) (http://www.fda.gov/) and PharmGKB (https://www.pharmgkb.org).

Data preprocessing

For the expression profile generated with the microarray platform from the GEO database, we first convert the probe id to the gene symbol and then use the average value to represent the expression of a gene if there are more than one probes mapped to one gene. For the data from TCGA, we first download level 3 files from the Genomic Data Commons (GDC) data portal and then link the omics data to sample and patient information with the Application Program Interface (API) (https://docs.gdc.cancer.gov/API/Users_Guide/Getting_Started/#tools-for-communicating-with-the-gdc-api) provided by the GDC data portal.

Table 1 Statistics of omics data deposited in GliomaDB

Data category	No. of projects	No. of samples	No. of patients	No. of records
Somatic mutation	3	2490	1423	6,083,427
Gene expression	18	3309	3283	56,181,100
miRNA expression	3	733	715	1,072,792
DNA methylation	3	986	960	184,091,259

A

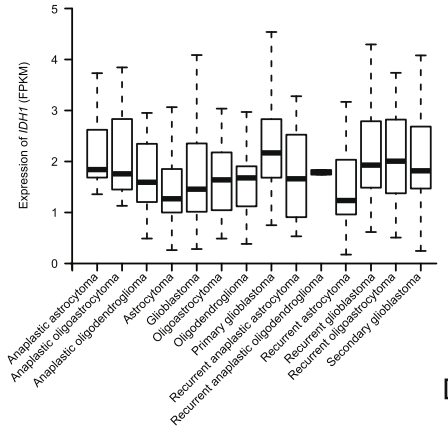
Search result for gene: *IDH1*

Gene symbol	*IDH1* (Gene ID : 3417)
Synonyms	HEL-216 | HEL-S-26 | IDCD | IDH | IDP | IDPC | PICD
Other names	MIM:147700 | HGNC:HGNC:5382 | Ensembl:ENSG00000138413 | HPRD:00984 | Vega:OTTHUMG00000132943
Genome context	2q34
Official full name	Isocitrate dehydrogenase (NADP(+)) 1, cytosolic
Gene type	Protein-coding
Gene summary	Isocitrate dehydrogenases catalyze the oxidative decarboxylation of isocitrate to 2-oxoglutarate. These enzymes belong to two distinct subclasses

Project accession	Project description	PMID
CGGA	The project explores brain tumor datasets over 2000 samples from Chinese cohorts. These datasets include the whole-exome sequencing (286), DNA methylation (159), mRNA sequencing (1018), mRNA microarray (301), microRNA microarray (198), and the matched clinical data.	23090983, 24550449, 22990979, 25415048, 25869098, 22960273
GSE13041-GPL570	Gene expression analysis of glioblastomas identified the major molecular basis for the prognostic benefit of younger age.	18940004
GSE13041-GPL8300	Gene expression analysis of glioblastomas identified the major molecular basis for the prognostic benefit of younger age.	18940004
GSE13041-GPL96	Gene expression analysis of glioblastomas identified the major molecular basis for the prognostic benefit of younger age.	18940004
GSE16011	Molecular subtypes were identified based on gene expression profiling on a large cohort of glioma samples (276) of all histological subtypes and grades.	19920198
GSE16581	Tumor biopsies from 43 female and 25 male subjects with sporadic meningioma were identified from the UCLA Neuro-oncology Program Tissue Bank through institutional review board approved protocols. 43 tumor samples were designated “benign” WHO I, 19 tumor samples were “atypical” WHO II, and 6 were “anaplastic” WHO III. Gene expression analysis was performed on the 68 tumor biopsies.	20015288
GSE2817	Unsupervised analysis using wavelet models of 30 diffuse glioma samples discovered new prognostic chromosomal patterns of gene expression.	17140431
GSE30074	Gene expression profiles of 30 primary medulloblastoma samples were analyzed with the aim of providing a new risk stratification approach based on expression profiles of medulloblastoma cohort.	22090452
GSE37418	A total of 76 pediatric medulloblastoma samples were analyzed, representing 4 expression classes.	22722829
GSE42669	Gene expression profiling experiments were conducted for 58 human glioblastoma samples using Affymetrix Human Gene 1.0 ST arrays according to manufacturer’s protocol.	23333277

Showing 1 to 10 of 17 entries

B



C

Sample accession	Expression value	Tissue	Platform	Quantification type	Molecule
CGGA_1001	1.29768	Glioblastoma	Illumina HiSeq 2000	FPKM	RNA
CGGA_1019	0.67193	Glioblastoma	Illumina HiSeq 2000	FPKM	RNA
CGGA_1023	0.71341	Glioblastoma	Illumina HiSeq 2000	FPKM	RNA
CGGA_1046	1.17449	Astrocytoma	Illumina HiSeq 2000	FPKM	RNA
CGGA_1053	1.61425	Glioblastoma	Illumina HiSeq 2000	FPKM	RNA
CGGA_1070	3.86966	Glioblastoma	Illumina HiSeq 2000	FPKM	RNA
CGGA_1071	1.98121	Astrocytoma	Illumina HiSeq 2000	FPKM	RNA
CGGA_1073	1.42823	Glioblastoma	Illumina HiSeq 2000	FPKM	RNA
CGGA_1074	2.21248	Glioblastoma	Illumina HiSeq 2000	FPKM	RNA
CGGA_1083	2.56327	Glioblastoma	Illumina HiSeq 2000	FPKM	RNA

Showing 1 to 10 of 325 entries

D

GliomaDB

Home / Sample

Information for sample: CGGA_1001

Project	GSE48865
Patient accession	E48865_CGGA_1001
Vital status	DEAD
Gender	Male
Age	12
Race	
Tumor location	
Tumor histology type	glioblastoma
KPS	
Grade	IV
Tissue	glioblastoma
Molecule	RNA

Figure 2 Example of gene expression search output with *IDH1*

A. The first query result of the gene expression of *IDH1* is shown as an example, including the gene summary information and projects with expression records of the query gene. **B.** Boxplot showing the expression of *IDH1* in different subtype of glioma from the CGGA project. **C.** Detailed gene expression information of *IDH1* in the CGGA project (only first 10 samples are listed). **D**. Information provided for a specific sample CGGA_1001 included in the project shown in panel C. CGGA, Chinese Glioma Genome Atlas.

Data statistics

GliomaDB integrates multi-omics data from 21 projects, which include 4303 patients and 21,086 samples. There are 6,083,427 records of single nucleotide variants (SNVs), and the corresponding annotations are based on hg19. There are 56,180,100 and 1,072,792 records in the gene expression and miRNA expression data, respectively. The DNA methylation data contain 27 K and 450 K data. The former has 27,578 CpG sites, and the latter has 485,578 CpG sites. There are 184,091,259 records in the DNA methylation data (Table 1). The variant/gene-related drug information collected from the FDA and PharmGKB contains data on the variant, PubMed ID, drug, disease, gene, *P* value, race, association, FDA guideline, etc. There are 77 targeted drugs and 6569 records regarding drug information.

Search

Four types of data can be queried in the search section: somatic mutation, gene expression, miRNA expression, and DNA methylation. GliomaDB provides a straightforward search

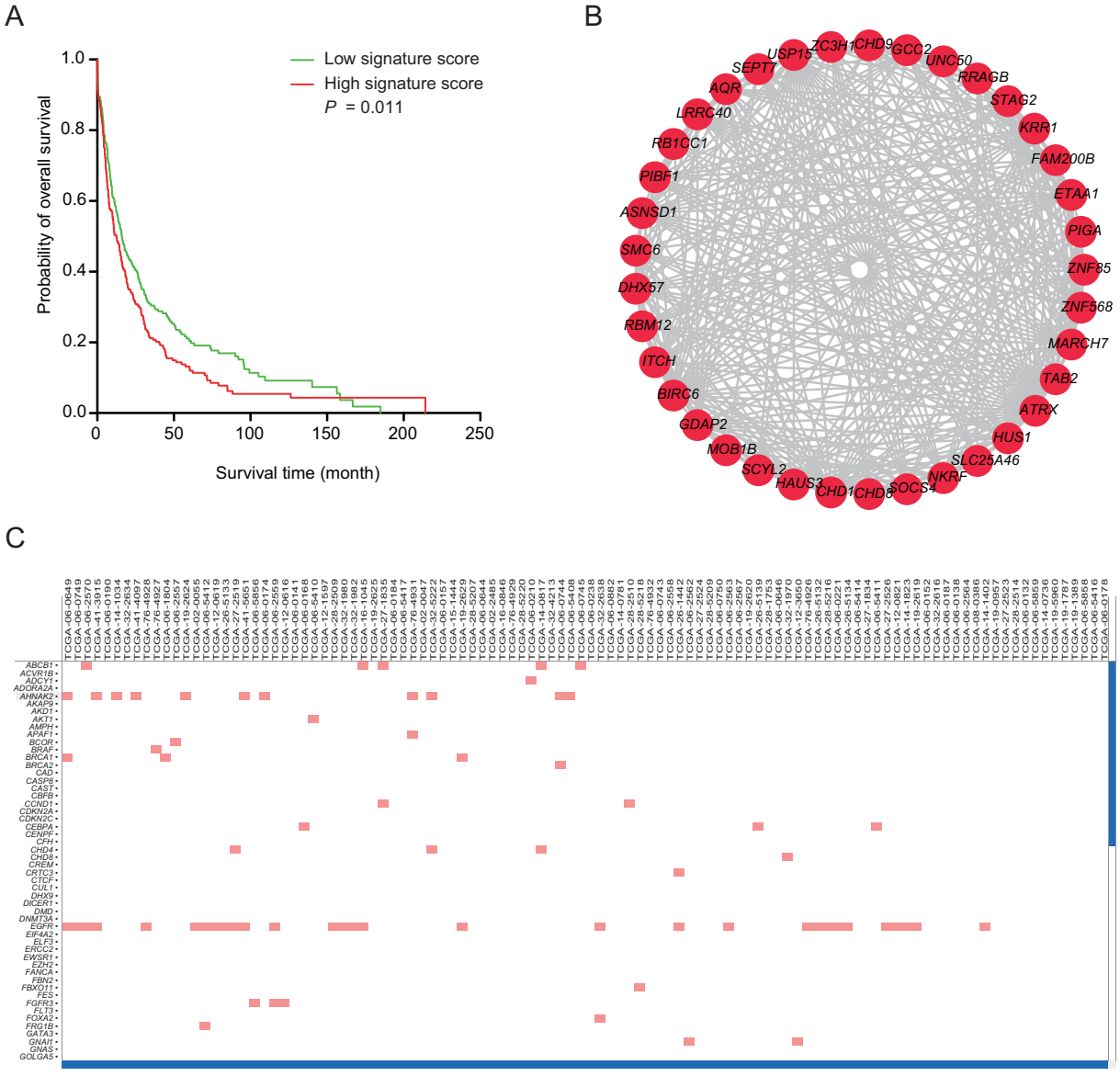


Figure 3 Example of GliomaDB analysis output

A. The overall survival analysis of a gene/genes of interest can be calculated and presented in a Kaplan–Meier plot. Here, we use expression of two genes, *KLF1* and *NF1*, in TCGA-LGG as an example. The mean score provided by coxph is used to split the samples. Samples with scores greater than the mean score of all samples are labeled as "high signature score", and the others are labeled as "low signature score". $P = 0.011$ indicates that the expression of *KLF1* and *NF1* is significantly associated with the overall survival of LGG patients in TCGA-LGG. **B**. The coexpression network of genes correlated with *ATRX*. All genes (shown in red solid circles) correlated with *ATRX* were firstly selected as a dataset and the relationships of each pair of genes in the dataset are visualized with the Cytoscape plugin in different layouts. **C**. The interactive heatmap visualization of the multi-omics data tested, including mutation, copy number variation, and expression profiles.

interface. Users can query by gene symbol (*e.g.*, *IDH1*), Ensembl ID (*e.g.*, ENSG00000138413), or gene ID (*e.g.*, 3417) in the "Gene name" search field. For the mutation query, chromosomal region, gene name, and dbSNP accession number are supported for the retrieval of somatic mutations. The results include the tumor sample and the matched normal sample, which could be further linked to detailed information about the corresponding patient. We also integrated the annotation for mutations from Oncotator [22]. This mutation information included the gene information and GO categories of the gene where the mutation is located, the somatic mutations from COSMIC located in the gene, the mutation frequency of the mutation in 1000 Genomes data, and the mutation information from ClinVar. In the gene and miRNA expression search section, the results are grouped by project, considering the incompatibility of expression value between different platforms (**Figure 2**). The results are presented in two steps. The first step shows the summary information of the query gene and projects containing the query gene in any of its samples (Figure 2A), and the second step shows the expression boxplot of different sample groups (Figure 2B) and detailed expression (Figure 2C) of the query gene in a specific project. Users can also obtain detailed information on the patient from whom the sample originated (Figure 2D) by clicking the sample accession number in the expression search results. In the methylation search section, chromosomal region, gene name, and cgid (Infinium MethylationEPIC probe ID) are supported for the retrieval of DNA methylation, and 450 K/27 K methylation data are included.

Analysis

GliomaDB includes four types of analysis: survival analysis, coexpression network, interactive heatmap visualization, and auxiliary targeted drug recommendation (**Figure 3**).

Survival analysis

Survival analysis based on gene expression levels is also widely used for predicting the clinical outcome of a given gene [23]. Therefore, the gene expression datasets were used for survival analysis. Single-gene or multiple-gene queries are both supported, and the results are presented in a Kaplan–Meier plot of two groups stratified by the mean score obtained by Cox regression (Figure 3A).

Coexpression network visualization

In each dataset, the Pearson's correlation coefficient is calculated for each of the two genes, which potentially denotes their regulation relationship. For the query of one gene, the resulting network includes edges between the query gene and each of other genes with a Pearson's correlation coefficient greater than 0.85 (Figure 3B).

Interactive heatmap view of multiple omics data

We integrated jHeatmap [24], an interactive web heatmap viewer built using JavaScript, to represent mutation, copy number variation, and expression profiles (Figure 3C).

Auxiliary targeted drug recommendation

We integrated the pharmacogenomics knowledge for personalized medicine from the FDA and PharmGKB [25] and offered a built-in interactive service for the retrieval of targeted drugs with either gene name/dbSNP accession ID/drug or a standard VCF file.

Conclusion and discussion

GliomaDB is a web server that has been developed for the integration of multiple omics data and interactive analysis in glioma studies. The data in GliomaDB are from TCGA, the GEO database, the MSK-IMPACT project, the FDA, and PharmGKB, with thousands of tumor and normal samples included. Data types include genome, transcriptome, miRNome, methylome, targeted drug, and genetic variation-drug association. GliomaDB is a time-saving, free, and intuitive tool for tapping the full potential of publicly available genomics big data, which enables biologists and clinicians without any programming experience to obtain ready-to-use multi-omics data and perform a diverse range of data analyses. GliomaDB is designed to complement existing tools, such as cBioPortal and GEPIA. It also has the potential to become a one-stop service for data query and analysis for the scientific and clinical community associated with the glioma field. In the future, we will not only continuously update multi-omics data from both glioma and normal samples, but also develop new analytical features for further exploration of the available big genomic data. We hope that GliomaDB would facilitate a better translation of data into knowledge and of knowledge to application.

Availability

GliomaDB is freely accessible at http://bigd.big.ac.cn/gliomaDB.

Authors' contributions

XF and HQ conceived the study and supervised the project. YY designed the system architecture. YY, YS, and BX wrote the source code. YY drafted the manuscript. XF revised the manuscript. All authors read and approved the final manuscript.

Competing interests

The authors have declared no competing interests.

Acknowledgments

This research was supported by the National Key R&D Program of China (Grant Nos. 2016YFC0901700, 2016YFC0901603, 2017YFC0907502, 2017YFC0908402, and 2017YFC0907405) and the Key Research Program of the Chinese Academy of Sciences, China (Grant No. KJZD-EW-L14). The authors would like to thank Enago (www.enago.cn) for the English language review.

References

[1] Siegel RL, Miller KD, Jemal A. Cancer statistics, 2019. CA Cancer J Clin 2019;69:7–34.

[2] Miller KD, Siegel RL, Lin CC, Mariotto AB, Kramer JL, Rowland JH, et al. Cancer treatment and survivorship statistics, 2016. CA Cancer J Clin 2016;66:271–89.

[3] Stupp R, Mason WP, van den Bent MJ, Weller M, Fisher B, Taphoorn MJ, et al. Radiotherapy plus concomitant and adjuvant temozolomide for glioblastoma. N Engl J Med 2005;352:987–96.

[4] Cancer Genome Atlas Research Network. Comprehensive genomic characterization defines human glioblastoma genes and core pathways. Nature 2008;455:1061–8.

[5] Brennan CW, Verhaak RG, McKenna A, Campos B, Noushmehr H, Salama SR, et al. The somatic genomic landscape of glioblastoma. Cell 2013;155:462–77.

[6] Ceccarelli M, Barthel FP, Malta TM, Sabedot TS, Salama SR, Murray BA, et al. Molecular profiling reveals biologically discrete subsets and pathways of progression in diffuse glioma. Cell 2016;164:550–63.

[7] Edgar R, Domrachev M, Lash AE. Gene Expression Omnibus: NCBI gene expression and hybridization array data repository. Nucleic Acids Res 2002;30:207–10.

[8] Barrett T, Wilhite SE, Ledoux P, Evangelista C, Kim IF, Tomashevsky M, et al. NCBI GEO: archive for functional genomics data sets–update. Nucleic Acids Res 2013;41:D991–5.

[9] Fuller GN, Scheithauer BW. The 2007 Revised World Health Organization (WHO) Classification of Tumours of the Central Nervous System: newly codified entities. Brain Pathol 2007;17:304–7.

[10] Ohgaki H, Kleihues P. Genetic alterations and signaling pathways in the evolution of gliomas. Cancer Sci 2009;100:2235–41.

[11] Wang L, Babikir H, Muller S, Yagnik G, Shamardani K, Catalan F, et al. The phenotypes of proliferating glioblastoma cells reside on a single axis of variation. Cancer Discov 2019;9:1708–19.

[12] Tang Z, Li C, Kang B, Gao G, Li C, Zhang Z. GEPIA: a web server for cancer and normal gene expression profiling and interactive analyses. Nucleic Acids Res 2017;45:W98–102.

[13] Li JR, Sun CH, Li W, Chao RF, Huang CC, Zhou XJ, et al. Cancer RNA-Seq Nexus: a database of phenotype-specific transcriptome profiling in cancer cells. Nucleic Acids Res 2016;44:D944–51.

[14] Parisot S, Duffau H, Chemouny S, Paragios N. Graph-based detection, segmentation & characterization of brain tumors. 2012 IEEE Conf Comput Vis Pattern Recognit 2012:988–95.

[15] Zhao Z, Meng F, Wang W, Wang Z, Zhang C, Jiang T. Comprehensive RNA-seq transcriptomic profiling in the malignant progression of gliomas. Sci Data 2017;4:170024.

[16] Goldman M, Craft B, Zhu J, Haussler D. Abstract 2584: The UCSC Xena system for cancer genomics data visualization and interpretation. Cancer Res 2017;77:2584.

[17] Gao J, Aksoy BA, Dogrusoz U, Dresdner G, Gross B, Sumer SO, et al. Integrative analysis of complex cancer genomics and clinical profiles using the cBioPortal. Sci Signal 2013;6:pl1.

[18] Chakravarty D, Gao J, Phillips SM, Kundra R, Zhang H, Wang J, et al. OncoKB: a precision oncology knowledge base. JCO Precis Oncol 2017;2017.

[19] Griffith M, Spies NC, Krysiak K, McMichael JF, Coffman AC, Danos AM, et al. CIViC is a community knowledgebase for expert crowdsourcing the clinical interpretation of variants in cancer. Nat Genet 2017;49:170–4.

[20] Kusnoor SV, Koonce TY, Levy MA, Lovly CM, Naylor HM, Anderson IA, et al. My Cancer Genome: evaluating an educational model to introduce patients and caregivers to precision medicine information. AMIA Jt Summits Transl Sci Proc 2016;2016:112–21.

[21] Zehir A, Benayed R, Shah RH, Syed A, Middha S, Kim HR, et al. Mutational landscape of metastatic cancer revealed from prospective clinical sequencing of 10,000 patients. Nat Med 2017;23:703–13.

[22] Ramos AH, Lichtenstein L, Gupta M, Lawrence MS, Pugh TJ, Saksena G, et al. Oncotator: cancer variant annotation tool. Hum Mutat 2015;36:E2423–9.

[23] Plitas G, Konopacki C, Wu K, Bos PD, Morrow M, Putintseva EV, et al. Regulatory T cells exhibit distinct features in human breast cancer. Immunity 2016;45:1122–34.

[24] Deu-Pons J, Schroeder MP, Lopez-Bigas N. jHeatmap: an interactive heatmap viewer for the web. Bioinformatics 2014;30:1757–8.

[25] Whirl-Carrillo M, McDonagh EM, Hebert JM, Gong L, Sangkuhl K, Thorn CF, et al. Pharmacogenomics knowledge for personalized medicine. Clin Pharmacol Ther 2012;92:414–7.