BMC Genomics

CrossMark

Exploring the key genes and signaling transduction pathways related to the survival time of glioblastoma multiforme patients by a novel survival analysis model

Yuan Xia[1], Chuanwei Yang[2,3], Nan Hu[4], Zhenzhou Yang[4], Xiaoyu He[5], Tingting Li[6*] and Le Zhang[1,6*]

Abstract

Background: This study is to explore the key genes and signaling transduction pathways related to the survival time of glioblastoma multiforme (GBM) patients.

Results: Our results not only showed that mutually explored GBM survival time related genes and signaling transduction pathways are closely related to the GBM, but also demonstrated that our innovated constrained optimization algorithm (CoxSisLasso strategy) are better than the classical methods (CoxLasso and CoxSis strategy).

Conclusion: We analyzed why the CoxSisLasso strategy can outperform the existing classical methods and discuss how to extend this research in the distant future.

Keywords: Least absolute shrinkage and selection operator (Lasso), Sure independence screening (SIS), Cox proportional hazards model (Cox), Glioblastoma multiforme (GBM), Signaling transduction pathway

Background

Glioblastoma multiforme (GBM) is the most common and malignant brain tumor [1–3]. Since GBM is high invasive and is mixed together with the healthy brain tissue, it is almost impossible to remove the tumor without causing serious consequences [4]. Moreover, GBM is very easy to relapse [5, 6]. The median survival and progression free survival time of GBM are 14.6 and 6.9 months, respectively. And the 5 year survival rate was 9.8 %[7]. Previous studies [8–10] indicated that gene mutation is one of the most important factors for GBM development. Therefore, gene expression analysis can not only be used to discover the underlying abnormality of gene expression associated with the GBM gene

mutation, but also be employed to discover gene signatures which could help us to investigate the related signaling transduction pathways. Results from the pathway analysis can lay the foundation for the GBM cancer targeted drug research in the future.

As one of the important survival analysis methods, the cox proportional hazards model [11] is broadly employed to investigate the connections between various covariates and the length of life. However, the classical cox proportional hazards model [12] can only process such survival data that the dimension of the factors (P) are less than the number of samples (N) [13] (we call it as P < <N type of data), but it is not able to handle the survival data that the dimension of the factors are greater than the number of samples such as the gene expression data [13] (we call it as P> >N type of data). To process P> >N type of data, Tibshirani et al., [14] integrated the Lasso algorithm, one of the constrained optimization methods, into the classical Cox

* Correspondence: tinalee@swu.edu.cn; zhanglcq@swu.edu.cn
[6]College of Mathematics and Statistics, Southwest University, Chongqing 400715, People's Republic of China
[1]College of Computer and Information Science, Southwest University, Chongqing 400715, People's Republic of China
Full list of author information is available at the end of the article

proportional hazards model [15] to select the key predictors. However, Fan et al., [16] pointed out if the number of predictors is much greater than the sample size (P> > N), a pre-cleaning step by a computationally expedient screening procedure is often preferred to increase the accuracy of the algorithm. Thus, Fan et al., [16] developed the sure independence screening (SIS) method by fitting marginal Cox regression models for each covariate and screening out several covariates by a pre-specified threshold. Nevertheless, reported by Hong et al., [17], marginal screening may encounter the difficulty in identifying these hidden and jointly important variables to incur false negatives. Therefore, Hong et al., [17] proposed a conditional SIS method to explore the potential predictors for the regular linear system, but not consider the survival data. On the other hand, developing a systematic approach to identify the target generic drug for the cancer treatment already becomes a popular research field [18, 19]. However, to our knowledge, there is no recent research discussing the incoherent connection between survival time and the target generic drug in detail.

To overcome the shortcomings of these previous research, this study proposed a multi-scale genes and signaling transduction pathways exploration platform (Fig. 1) with the following three innovations. Firstly, we innovatively analyzed the clinical GBM gene expression and survival time data [20] to investigate the incoherent relation between the signatures of genes and the survival time of GBM patient. Secondly, we not only integrated the constrained optimization method such as Lasso [15] into classical Cox proportional hazards model [13] to explore survival time related key genes by processing the P> > N type of data, but also used the SIS algorithm to improve the predictive accuracy. Thirdly, we employed KOBAS database [21] and hypergeometric test [22] to investigate the correlated GBM signaling transduction pathways regarding the explored survival time related key genes. And then, these survival time related signaling transduction pathway could help us to bridge the relation between the targeted drugs and the survival time for GBM patients.

The clinical GBM gene expression and survival data set used in this study is downloaded from the Georgetown Database of Cancer G-DOC [20], which has 54,675 features (P) and 227 samples (N). To handle such a P> > N type of data, we developed the CoxSisLasso strategy. It firstly integrated constrained optimized methods such as Lasso into the classical cox regression model to select the prior genes with potentially great impact on the patients' survival time. Secondly, conditioned on these genes selected by Lasso, conditional SIS method [23] is used to re-select the possible genes from these genes screened out in the first step. To bridge the relation between the targeted drugs and the survival time for GBM patients, we employed the KOBAS [21] application and the explored GBM survival time related key genes to investigate which signaling transduction pathways closely correlate with the GBM survival time.
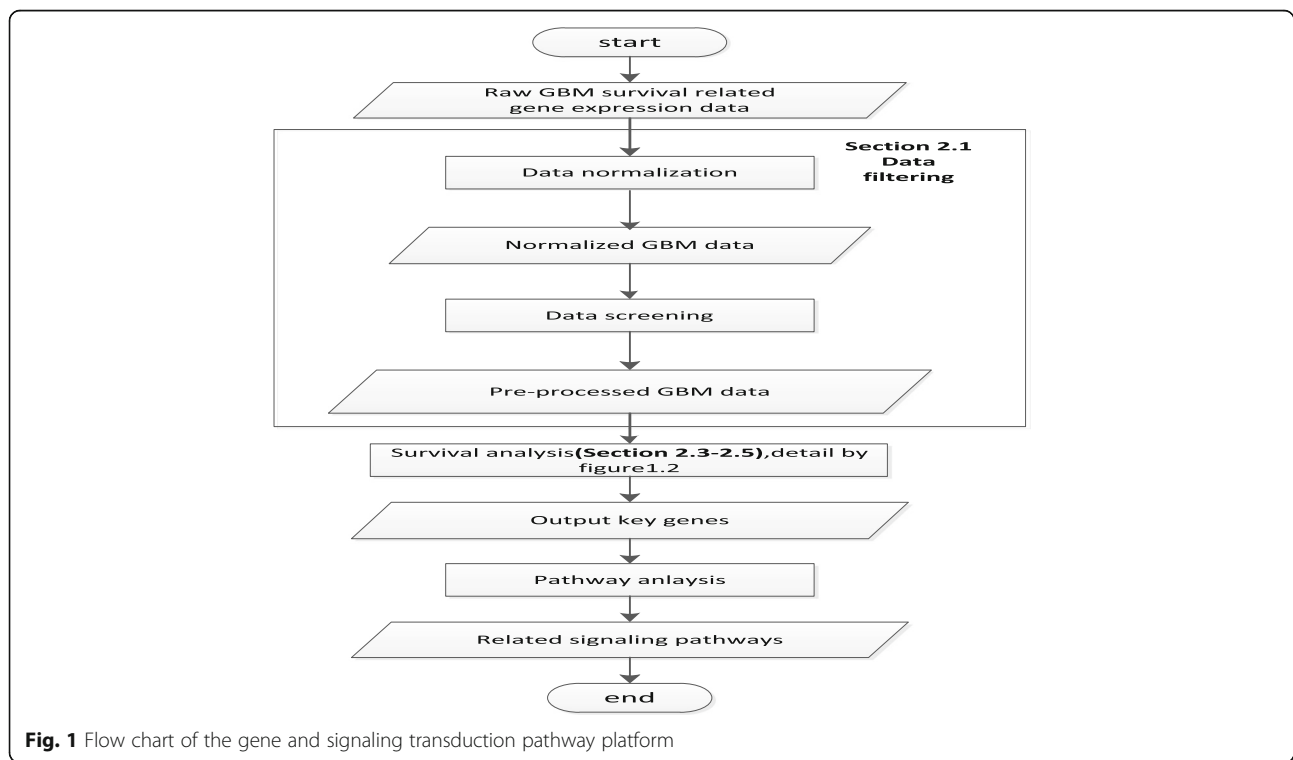

Fig. 1 Flow chart of the gene and signaling transduction pathway platform

In general, this study developed a multi-scale genes and signaling transduction pathways exploration algorithm that can not only investigate the molecular mechanism between the key genes and cancer patients' survival time, but also employ hypergeometric distribution based database (KOBAS) to look for the related signaling pathways in the proteomics level for the future targeted cancer therapy [24, 25]. Manually-reviewed experimental evidences showed that mutually explored GBM survival time related genes [26–38] and signaling transduction pathways [39–52] are closely related to the GBM. In addition, the research results demonstrate that our proposed CoxSisLasso strategy has the best predictive power and model fitting capacity compared to the CoxLasso and CoxSis strategy developed by Tibshirani et al.,[14] and Fan et al., [16], respectively. Finally, we theoretically analyze why the CoxSisLasso strategy outperforms CoxLass and CoxSis and discuss the further research.

Methods

Materials
We used a multi-study microarray database of GBM expression profiles ($n = 227$) from the Georgetown Database of Cancer G-DOC [20], based on the Affymetrix U133 plus 2.0 GeneChip microarray platform. The microarray datasets of GBM are listed by Table 1.

Data filtering
The original microarray datasets are normalized and preprocessed by R software package [53]. After preprocessing step, there are 227 samples and 54,675 genes left in the data matrix. Next, the interquartile range (IQR) threshold [54] is employed to screen out the genes with small variance value. After that, there are only 227

samples and 10,992 genes left in the GBM gene expression and survival time data matrix.

Cox proportional hazards model
Survival analysis [11, 55] works for the analysis of time duration until one or more events happen. As one of the widespread used survival analysis methods, the Cox proportional hazards model [13] is used to analyze the time-to-event data with both censored data and covariates, which assumes a semi-parametric form for the hazard as Eq. 1.

$$h_i(t) = h_0(t) \exp\ (x_i^T \beta) \tag{1}$$

where $h_i(t)$ is the hazard for patient i at time t, $h_0(t)$ is a shared baseline hazard function, β is an unknown p-dimensional regression coefficient vector and x_i is a vector of potential predictors for the ith individual. Based on the available samples, the estimator of the unknown parameter coefficients $\hat{\beta}$, can be obtained by maximizing the log-partial likelihood function as Eq. 2

$$\begin{aligned} \hat{\beta} &= \arg\max \log PL \\ &= \arg\max \sum_{k \in D} \left[x_k^T \beta - \log \left(\sum_{j \in R_k} \exp\left(x_j^T \beta\right) \right) \right] \end{aligned} \tag{2}$$

where D is the set of indices of the events and R_k denotes the set of indices of the individuals at risk at time t_k.

Since this study encounters the P> > N type of data, it is impossible to employ classical Cox proportional hazard regression method [13] to analyze the GBM gene expression data matrix directly. Therefore, the following sections propose three variable selection strategies to obtain the sparse regression coefficient.

Table 1 The illustration of microarray datasets of GBM

Data set name	Source site	Affymetrix platform	Sample size
GBM	DANA-FARBER CANCER INSTITUTE	HG-U133_Plus_2	2
GBM	NABTT/H. LEE MOFFITT CANCER CENTER	HG-U133_Plus_2	30
GBM	HENRY FORD HOSPITAL (RETRO)	HG-U133_Plus_2	62
GBM	M. D. ANDERSON CANCER CENTER	HG-U133_Plus_2	6
GBM	MSKCC/NEW YORK	HG-U133_Plus_2	2
GBM	NABTT/HENRY FORD HOSPITAL	HG-U133_Plus_2	7
GBM	NABTT/JOHNS HOPKINS	HG-U133_Plus_2	2
GBM	NIH NEURO-ONCOLOGY BRANCH	HG-U133_Plus_2	48
GBM	TJU	HG-U133_Plus_2	30
GBM	UCLA SCHOOL OF MEDICINE	HG-U133_Plus_2	7
GBM	UCSF	HG-U133_Plus_2	17
GBM	UNIV OF PITTSBURGH	HG-U133_Plus_2	9
GBM	UNIVERSITY OF WISCONSIN	HG-U133_Plus_2	5

Combined Cox and Lasso (CoxLasso) strategy

To obtain the sparse solution for the parameter β in the Cox proportional hazards model (Eq. 1), we have to integrate constrained optimization methods such as Lasso proposed by Tibshirani et al.,[14] into classical Cox proportional hazards model to minimize the negative log partial likelihood subject to the sum of the absolute values of the parameters being bounded by a constant as Eq. 3.

$$\hat{\beta} = \arg\min - \left\{ \sum_{k \in D} \left[x_k^T \beta - \log\left(\sum_{j \in R_k} \exp\left(x_j^T \beta \right) \right) \right] \right\}$$
$$\text{subject to } \sum_{j=1}^{p} \left| \beta_j \right| \leq t$$

$$(3)$$

It is equivalent to the following optimization problem

$$\hat{\beta} = \arg\min - \left\{ \sum_{k \in D} \left[x_k^T \beta - \log\left(\sum_{j \in R_k} \exp\left(x_j^T \beta \right) \right) \right] \right\} + \lambda \sum_{j=1}^{p} \left| \beta_j \right|$$

$$(4)$$

where λ is the tuning parameter to control the sparsity of the estimator. This research used the R package tool glmnet developed by Friedman et al.,[56] to implement the combined Cox and Lasso (CoxLasso) strategy (Fig. 2a) by using cross validation to choose the tuning parameter.

Combined Cox and SIS (CoxSis) strategy

Though directly integrating Lasso method into Cox model can process P> > N type of data, it may encounter problems with speed, stability, and accuracy, once the dimension of the covariates is ultra-high [23] . Therefore, it is often preferred to employ a simple and computationally efficient screening procedure to reduce the data dimensionality to a moderate size before using Lasso method. The combined Cox and SIS (CoxSis) strategy is illustrated by the following steps:

Step 1: Fit a marginal Cox regression model for each covariate x_m to obtain $\hat{\beta}_m$ by Eq. 5.1

$$\hat{\beta}_m = \arg\max \sum_{k \in D} \left[x_{km} \beta_m - \log\left(\sum_{j \in R_k} \exp\left(x_{jm} \beta_m \right) \right) \right] \quad (5.1)$$

Step 2: Rank the magnitudes of $\hat{\beta}_j, j = 1, 2, ..., p$ in decreasing order and keep the number of d top ranked covariates.
Step 3: Denote the index of selected covariates by Θ. Implement Lasso with the selected d covariates by minimizing Eq. 5.2

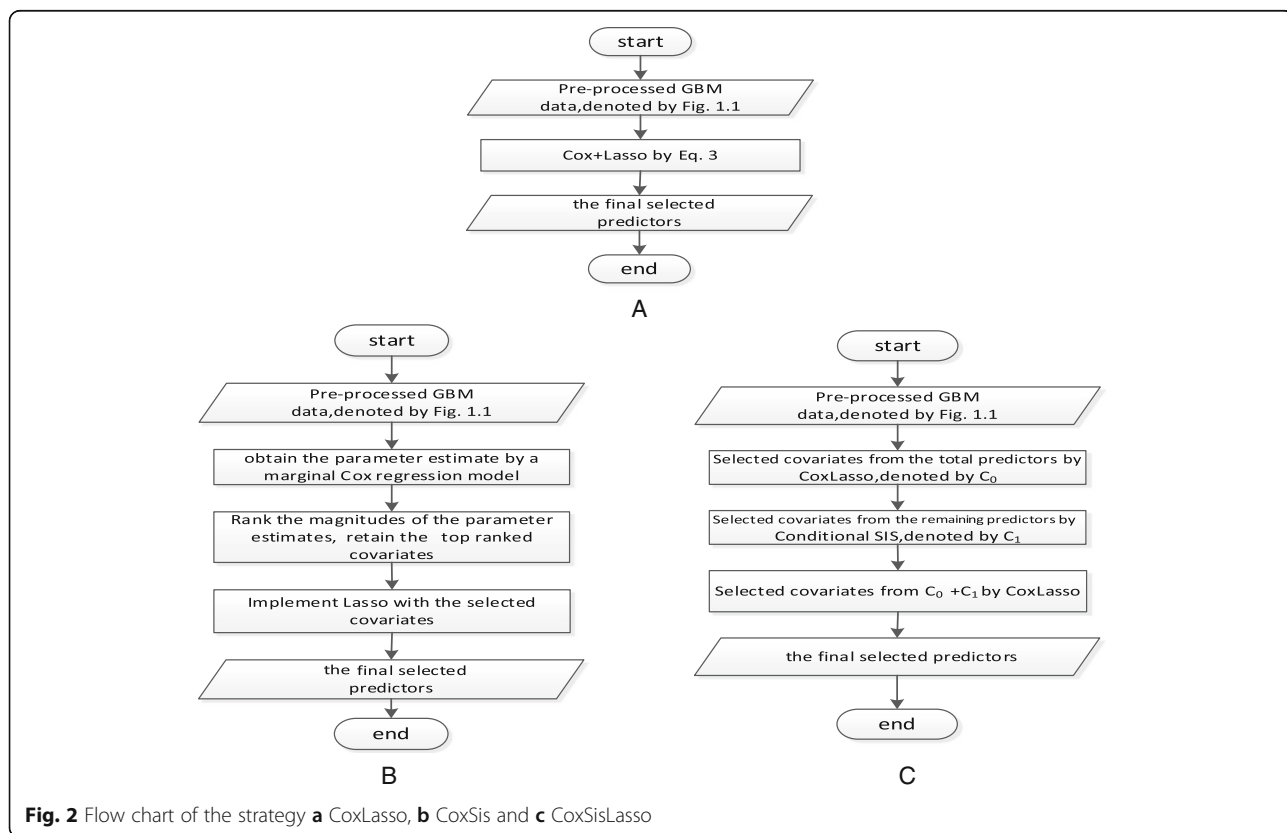**Fig. 2** Flow chart of the strategy **a** CoxLasso, **b** CoxSis and **c** CoxSisLasso

$$\min_{\beta_\Theta}\left\{-\sum_{k\in D}\left[x_{k,\Theta}^T\beta_\Theta-\log\left(\sum_{j\in R_k}\exp\left(x_{j,\Theta}^T\beta_\Theta\right)\right)\right]+\lambda\sum_{j\in\beta_\Theta}\left|\beta_j\right|\right\}$$
(5.2)

This study employs R package of SIS developed by Fan et al., [16] to implement the combined Cox and SIS (CoxSis) strategy (Fig. 2b).

Combined Cox, SIS and Lasso (CoxSisLasso) strategy

Recently, Barut et al., [57] proposed a conditional screening approach (Conditional SIS) to enhance the accuracy of SIS by using the prior knowledge of the key factors to select the predictors. Regarding to our P>>N type of data and the limitation of Lasso method in the stability and accuracy, this study proposed a combined Cox, SIS and Lasso (CoxSisLasso) strategy (Fig. 2c) to increase the predictive accuracy of the model as follows:

Step 1: Implement Lasso for the data. Denote the index of selected covariates with Lasso by C_0.
Step 2: Conditioned on the selected subset of covariates C_0, for each covariate x_m, $m\notin C_0$, fit the following Cox regression model by maximizing Eq. 6

$$\hat{\beta}_m=\underset{\beta_m}{\arg\max}$$
$$\left\{\sum_{k\in D}\left[x_{k,C_0}^T\beta_{C_0}+x_{k,m}\beta_m-\log\left(\sum_{j\in R_k}\exp\left(x_{j,C_0}^T\beta_{C_0}+x_{j,m}\beta_m\right)\right)\right]\right\}$$
(6)

Step 3: For a given threshold γ, keep the variables x_m, $m\notin C_0$ if $\left|\hat{\beta}_m\right|\ge\gamma$. Denote $C_1=\left\{m\notin C_0,\left|\hat{\beta}_m\right|\ge\gamma\right\}$. Then the augmented selected predictors are $C_0\cup C_1$.
Step 4: Implementing Lasso with the covariates in the set $C_0\cup C_1$ to select the final predictors.

For the threshold γ, Barut et al.,[57] proposed two procedures by controlling FDR and random decoupling to choose the proper level of threshold. Motivated by Zhao and Li [23], this study sets the threshold $\gamma = 1/p$, and p is the total number of all the covariates. Once the p-value of the Z-test of the covariate x_m, $m\notin C_0$ is less than the γ, we keep it as one of the important predictors.

Investigate potential signaling pathway regarding to the candidate genes related to the GBM survival time

After obtaining the explored GBM survival time related key genes by previous strategies, it is interesting for us to investigate which potential signaling pathways are closely related to these genes. And the potential pathways will be employed for the targeted drug therapy to treat the GBM cancer in the future.

KOBAS is a signaling transduction pathway database to identify statistically significantly enriched pathways using hypergeometric test [11]. In statistics, the hypergeometric test uses the hypergeometric distribution (Eq. 7) to calculate the statistical significance.

$$p(X=k)=\frac{\dbinom{K}{k}\dbinom{N-K}{n-k}}{\dbinom{N}{n}}$$
(7)

where N is the population size, K is the number of success states in the population, n is the number of draws, k is the number of observed successes.

Results

The explored GBM survival time related key genes by CoxLasso, CoxSis and CoxSisLasso strategy, respectively

Here, Table 2 shows the explored GBM survival time related key genes by CoxLasso, CoxSis and CoxSisLasso strategy, respectively. Also, the Venn plot (Fig. 3) indicates there are four common genes (AEBP1, GDNF, IL17RC and EIF3A) mutually selected by these three strategies, which closely correlate with the survival time of GBM patient validated by the manually-reviewed experimental evidences [26–38].

Firstly, AEBP1 (Adipocyte enhancer binding protein 1) was discovered as a transcriptional repressor [26]. It not only expresses at different levels in different organ and tissue types and its expression is relatively strong in brain [27], but also it can interact with tumor suppressor protein PTEN and inhibit its tumor-suppressing function [28]. AEBP1 can also negatively regulate IkB, resulting in the up-regulation of NF-kB and enhanced inflammatory response [29]. It is well known that both PTEN and NF-kB, closely related to the AEBP 1, are important players in GBM cancer progression. Moreover, previous research identified several genomic targets of AEBP1 playing vital roles in the survival of glioma cells [30].

Secondly, GDNF is a Glial Cell derived neurotrophic factor which promotes survival of neurons [31]. GDNF is not

Table 2 The explored genes for strategy CoxLasso, CoxSis and CoxSisLasso

Method	Key genes
CoxLasso	ARIH2, ZNF786, AEBP1, FOXG1, INTS1, GDNF, CUTC, SGCD, CCM2, IL17RC, EIF3A, CBLN1
CoxSis	YAP1, TRAF3IP2, AEBP1, GDNF, EAF2, ST5, IL17RC, EIF3A
CoxSisLasso	ARIH2, ZNF786, AEBP1, FOXG1, INTS1, GDNF, SGCD, IL17RC, EIF3A, CBLN1, SLC35D1, ELOVL2, CDCA7L, SNTB1, TELO2
CoxLasso, CoxSis and CoxSisLasso	AEBP1, GDNF, IL17RC, EIF3A

Fig. 3 Venn plot for the explored GBM genes
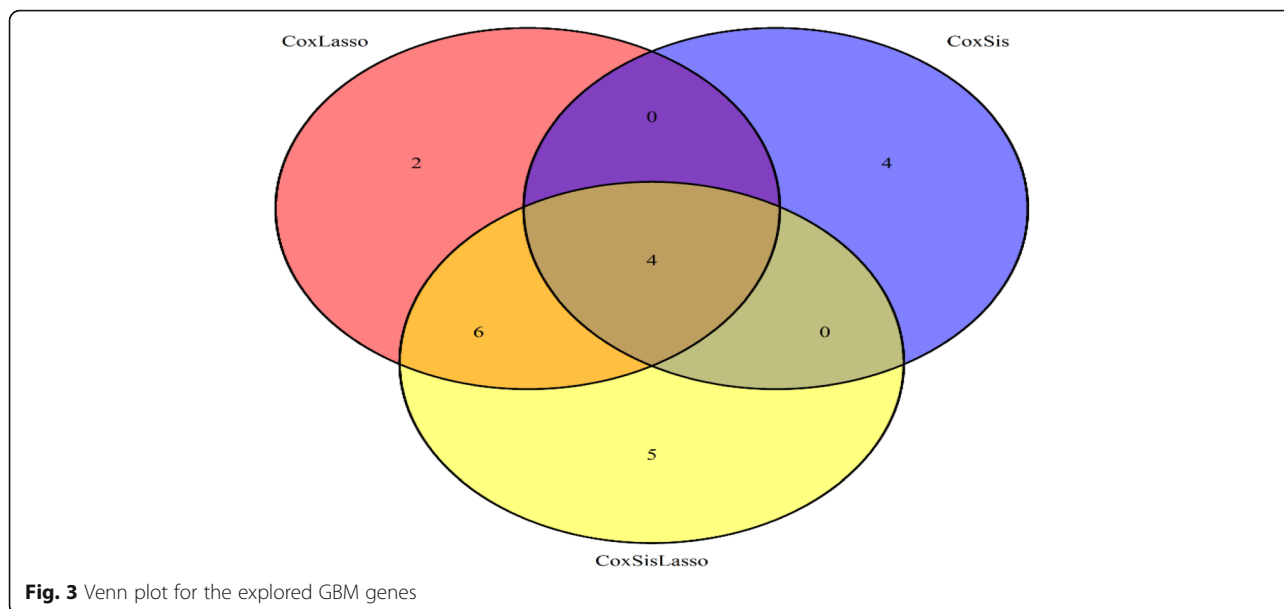
only identified as an important factor in macrophage infiltration into GBM, contributing to GBM progression [32], but also it can promote glioma cell invasion through its receptors that are present on invasive GBM cells [33].

Thirdly, EIF3A (Eukaryotic translation initiation factor subunit 3A) is not only expressed in all tissue types in human body and its expression is up-regulated in some type of cancers [34], but also it is important in regulating the expression of proteins involved in DNA repair pathway which is essential in drug sensitivity and resistance in cancer treatment [35, 36]. Especially, EIF3A is found to be overexpressed in some glioma patients [37].

Fourthly, Inlerleukin-17 receptor C (IL17RC) is a key molecule mediating interleukin 17 signaling. It is important in immune response and inflammation which are important in GBM progression [38].

Predictive performance comparison of survival time for each strategy

This study employs the idea of time-dependent receiver operating characteristic curve (ROC) for the censored
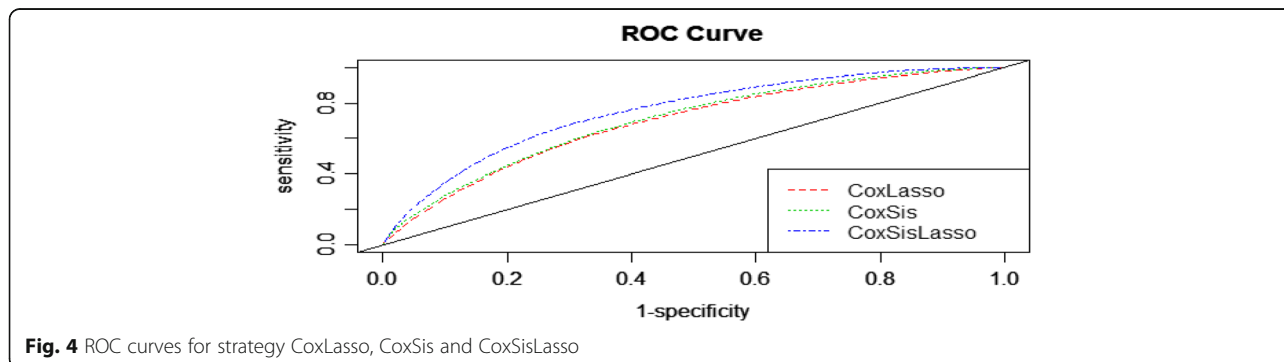
data and the area under the curve (AUC) [58, 59] to quantify the predicative accuracy for each strategy, when the outcome of interest is the survival time. The ROC curve depicts the sensitivity (Eq. 8.1) versus 1-specificity (Eq. 8.2) at each time t for any risk score function $x^T\beta$
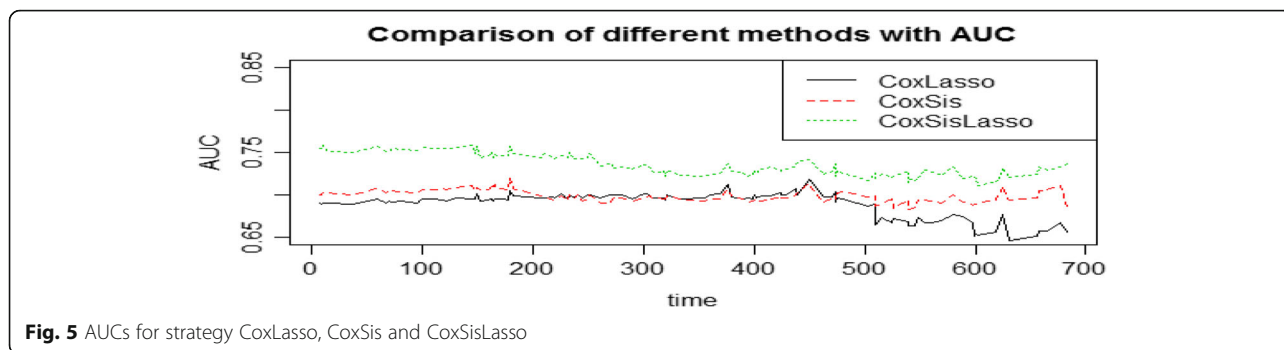
$$sensitivity(c, t|x^T\beta) = \Pr\{x^T\beta > c|\delta(t) = 1\} \quad (8.1)$$

$$specificity(c, t|x^T\beta) = \Pr\{x^T\beta \leq c|\delta(t) = 0\} \quad (8.2)$$

with c being the cut-off value and $\delta(t)$ is the event indicator at time t.

Figures 4 and 5 depicts the ROC curve at a specific predicted time 30 and AUC over a period of time respectively to quantify the performance of the three strategies to predict the survival time of the GBM patients. It demonstrates that CoxLasso and CoxSis strategy shares the similar predictive performance, whereas our proposed CoxSisLasso strategy has the best predictive accuracy since it not only has the greatest value of the sensitivity and 1-specificity (Fig. 4), but also the largest

Fig. 4 ROC curves for strategy CoxLasso, CoxSis and CoxSisLasso

Fig. 5 AUCs for strategy CoxLasso, CoxSis and CoxSisLasso

AUC value (Fig. 5). Furthermore, to assess the generalization ability of the proposed model, we randomly select 120 samples as the training samples and the rest 68 samples as the test samples. Figure 6 shows the AUCs of the three strategies for the testing samples. Both Figs. 5 and 6 turn out that our proposed CoxSisLasso method provides the largest AUC value with the best performance.

Model fitting performance comparison for each strategy

Table 3 summarizes the Cox regression results with the key genes selected by three strategies. R^2 is the statistic of the goodness of fit measure [60]. The concordance index [61] is a valuable measure of model discrimination in analyses involving survival time data. Greater R^2 and concordance index value imply better model fitting performance. Table 3 shows that R^2 and concordance index value of CoxSisLasso strategy (Table 3C) outperforms the other two (Table 3A & B). Moreover, by comparing results of Table 3C (CoxSisLasso) with the results of Table 3A (CoxLasso) and Table 3B (CoxSis), we found that CoxSisLasso not only can preserve the genes selected by the CoxLasso and CoxSis, but also it can introduce several statistically significant genes, which are potential for us to explore their relationships with GBM in the distant future.

The explored GBM survival time related signaling transduction pathways by CoxLasso, CoxSis and CoxSisLasso strategy, respectively

Here, Table 4 lists the explored GBM related signaling transduction pathways by CoxLasso, CoxSis and CoxSisLasso strategy, respectively. Also, the Venn plot (Fig. 7) indicates three explored GBM related signaling transduction pathways mutually selected by these three strategies.

And then, we employed the manually-reviewed experimental evidences [39–52] to demonstrate that these mutually explored signaling transduction pathways closely correlate with the survival time of GBM patient as following.

Firstly, mTOR (Mammalian target of rapamycin) is an important mediator of phosphatidyl-inositol-3 kinase (PI3K) pathway. And previous research turned out that constitutive activation of PI3K signaling is found in the majority of GBM patients [39]. Moreover, PI3K-akt-mTOR axis plays essential role in cell growth and proliferation [40]. Signaling of mTOR pathway is vital for cancer cell growth and survival in GBM patients [41]. Currently mTOR pathway inhibitors are under active investigation in preclinical experiments and in clinical trials for GBM treatment [42].

Secondly, TGF-beta (Transforming growth factor beta) is a secreted cytokine which signals through specific
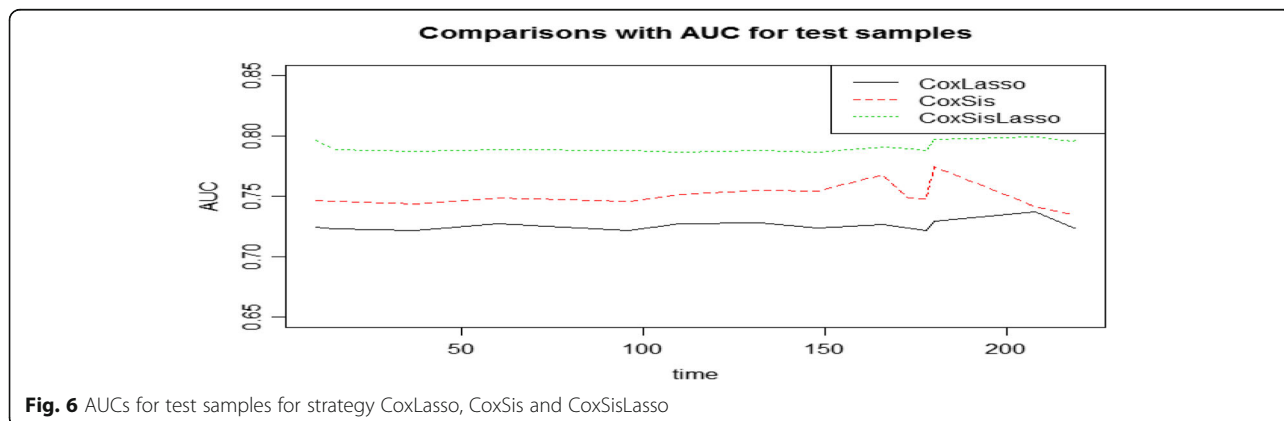

Fig. 6 AUCs for test samples for strategy CoxLasso, CoxSis and CoxSisLasso

Table 3 Model fitting results for strategy (A) CoxLasso, (B) CoxSis and (C) CoxSisLasso

Key genes	coef	exp(coef)	se(coef)	z	p-value
A					
ARIH2	0.28827	1.33412	0.20957	1.376	0.1690
ZNF786	0.73967	2.09524	0.31849	2.322	0.0202*
AEBP1	0.09910	1.10418	0.09315	1.064	0.2874
FOXG1	0.14722	1.15861	0.06712	2.193	0.0283*
INTS1	0.19661	1.21726	0.27385	0.718	0.4728
GDNF	−0.33054	0.71854	0.29059	−1.137	0.2553
CUTC	−0.03165	0.96885	0.27837	−0.114	0.9095
SGCD	0.12861	1.13724	0.20752	0.620	0.5354
CCM2	0.29707	1.34591	0.28104	1.057	0.2905
IL17RC	0.51024	1.66569	0.21579	2.364	0.0181*
EIF3A	−0.27131	0.76238	0.23337	−1.163	0.2450
CBLN1	−0.29685	0.74316	0.30079	−0.987	0.3237
$R^2 = 0.338$, Concordance = 0.687					
B					
YAP1	−0.28804	0.74973	0.12372	−2.328	0.019902*
TRAF3IP2	−0.39514	0.67358	0.20318	−1.945	0.051805
AEBP1	0.33103	1.39239	0.09268	3.572	0.000354***
GDNF	−1.09305	0.33519	0.30549	−3.578	0.000346***
EAF2	−0.53363	0.58647	0.21472	−2.485	0.012949*
ST5	−0.26305	0.76870	0.26139	−1.006	0.314240
IL17RC	1.02690	2.79240	0.22954	4.474	7.69e-06***
EIF3A	−0.40494	0.66702	0.21963	−1.844	0.065216
$R^2 = 0.375$, Concordance = 0.696					
C					
ARIH2	0.27310	1.31403	0.20654	1.322	0.186080
ZNF786	1.17873	3.25025	0.31923	3.692	0.000222***
AEBP1	0.20724	1.23027	0.09558	2.168	0.030151*
FOXG1	0.32694	1.38672	0.09262	3.530	0.000416***
INTS1	0.85607	2.35388	0.34015	2.517	0.011844*
GDNF	−0.48393	0.61636	0.35170	−1.376	0.168835
SGCD	−0.53359	0.58650	0.22601	−2.361	0.018233*
IL17RC	1.23644	3.44332	0.23552	5.250	1.52e-07***
EIF3A	−0.08224	0.92105	0.22885	−0.359	0.719339
CBLN1	−1.06495	0.34474	0.38400	−2.773	0.005548**
SLC35D1	−0.44547	0.64052	0.21196	−2.102	0.035579*
ELOVL2	−0.16161	0.85077	0.08395	−1.925	0.054210
CDCA7L	−0.38939	0.67747	0.11097	−3.509	0.000450***
SNTB1	−0.61372	0.54133	0.16747	−3.665	0.000248***
TELO2	−1.28721	0.27604	0.48148	−2.673	0.007507**
$R^2 = 0.515$, Concordance = 0.747					

Significance codes: 0 '***' 0.001 '**' 0.01 '*' 0.05

Table 4 The explored signaling transduction pathways for strategy CoxLasso, CoxSis and CoxSisLasso

Method	Key pathways
CoxLasso	1. internal ribosome entry pathway
	2. mtor signaling pathway
	3. TGF-beta signaling pathway
	4. p38 MAPK signaling pathway
	5. FoxO signaling pathway
CoxSis	1. Hippo signaling pathway
	2. internal ribosome entry pathway
	3. mtor signaling pathway
	4. TGF-beta signaling pathway
CoxSisLasso	1. internal ribosome entry pathway
	2. mtor signaling pathway
	3. TGF-beta signaling pathway
	4. Fanconi anemia pathway
CoxLasso, CoxSis and CoxSisLasso	1. internal ribosome entry pathway
	2. mtor signaling pathway
	3. TGF-beta signaling pathway

receptors and exerts its effect via intracellular Smad family proteins [43]. TGF-beta pathway controls GBM cell proliferation [44]. Its signaling contributes to the maintenance of tumor-initiating cells in GBM [45]. TGF-beta pathway is also involved in tumor invasion and metastasis in GBM patients [46]. Inhibition of TGF-beta pathway signaling reduced GBM cell proliferation and invasion in preclinical cell-based assays [47]. TGF-beta pathway inhibitors showed promising results in improving GBM patient survival in clinical trials [48].

Thirdly, IRE (Internal Ribosomal Entry) pathway is involved in the synthesis of some proteins during which protein synthesis is initiated from a start codon near an IRE site rather than by scanning the Kozak sequence. This IRE pathway is used in the translation of many eukaryotic genes including growth factors such as VEGF, FGF2 and PDGF [49] and transcription factors such as c-myc and hypoxia induced factor [50, 51] . Indeed, up-regulated expression of proto oncogene c-Jun in human GBM is mediated through a potent internal ribosomal entry site (IRES) in the 5′UTR of the c-Jun mRNA, and the upregulation of c-Jun contributes to the malignant properties of GBM cells [52].

Discussion
This study developed a multi-scale gene and signaling transduction pathway exploration platform based on the classical Cox proportional hazard model [12], constrained optimization method [14–16] and hypergeometric test to analyze P> > N type of GBM gene expression and survival time data (Table 1). Compared to the previous research
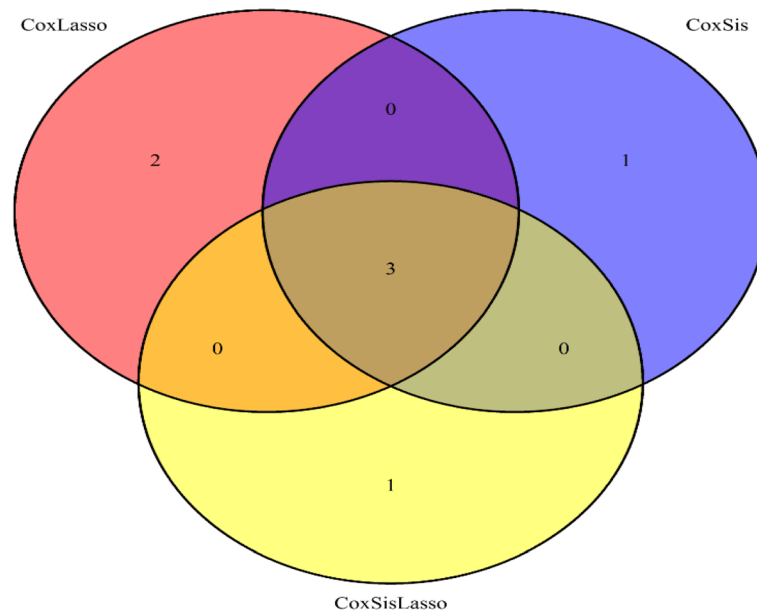
Fig. 7 Venn plot for the GBM related signaling transduction pathways

[14–16, 62], we proposed a novel CoxSisLasso strategy to investigate relationship between genes and GBM patients' survival time in molecular level as well as used KOBAS database [21] to look for the survival time related signaling transduction pathways.

On the one hand, manually reviewed experimental evidences validate that both mutually explored key genes [26–38] (Table 2) and signaling transduction pathways [39–52] (Table 4) are closely related to GBM. On the other hand, since CoxLasso strategy may encounter problems with speed, stability, and accuracy for processing high dimensional data [23], the CoxSis strategy is developed by employing a simple and computationally efficient screening procedure to reduce the dimensionality of the data to a moderate size before using Lasso method based on the previous work of Fan et al., [16]. Though classic marginal screening approach based CoxSis is theoretically proved to be capable of selecting all important predictors [56], it is difficult to identify these hidden predictors which jointly correlate with the response variable but not marginally. For this reason, we proposed the CoxSisLasso strategy, which not only uses the CoxLasso strategy to obtain a prior set of important predictors, but also incorporates the SIS [16] approach to select the important predictors regarding to the previous results. Figure 5 and 6 turned out that CoxSisLasso strategy has the best predictive power and model fitting capacity than both CoxLasso and CoxSis.

Conclusions

In general, this study innovatively developed a CoxSisLasso strategy to interrogate the connections between GBM gene expression and GBM patients' survival time as well as employed the KOBAS database [21] and hypergeometric test [21] to investigate the incoherent signaling transduction pathways and the survival time of GBM patient. Though the research results demonstrated the advantages of our algorithm, the current research still has several shortcomings such as the theoretically proof for the CoxSisLasso strategy, simulation study for the gene and pathway selection platform and so on. In the distant future, we will not only need improve our current CoxSisLasso algorithm, but also will employ the related pathway analysis theory [63] to explore the GBM survival time related proteins for the target drug study.

Availability of data and materials

Raw data (GBM expression profiles) are accessible through the Georgetown Database of Cancer G-DOC [61].

Authors' contributions

YX, LZ, XH and TL: Idea, implementation, testing and writing. CY, NH and ZY carried out experimental work for the manuscript. All authors read and approved the final manuscript.

Competing interests

The authors declare that they have no competing interests.

Consent for publication

Not applicable.

Ethics approval and consent to participate

Not applicable.

Author details

[1]College of Computer and Information Science, Southwest University, Chongqing 400715, People's Republic of China. [2]Systems Biology, the University of Texas MD Anderson Cancer Center, Houston, USA. [3]Breast Medical Oncology, the University of Texas MD Anderson Cancer Center, Houston, USA. [4]Cancer Center, Research Institute of Surgery, Daping Hospital, Third Military Medical University, Chongqing 400042, People's Republic of China. [5]Chongqing Zhongdi Medical Information Technology Co., Ltd, Chongqing 401320, People's Republic of China. [6]College of Mathematics and Statistics, Southwest University, Chongqing 400715, People's Republic of China.

References

1. Templeton A, Hofer S, Töpfer M, Sommacal A, Fretz C, Cerny T, Gillessen S. Extraneural spread of glioblastoma–report of two cases. Onkologie. 2008;31:192–4.
2. Tania A, Arkaitz C, Boris J, Guillermo V, Garry M, Raphael M, Luis A, Manuel G, Ismael GR. Cannabinoids induce glioma stem-like cell differentiation and inhibit gliomagenesis. J Biol Chem. 2007;282:6854–62.
3. Scott J, Rewcastle N, Brasher P, Fulton D, Hagen N, MacKinnon J, Sutherland G, Cairncross J, Forsyth P. Long-term glioblastoma multiforme survivors: a population-based study. Can J Neurol Sci. 1998;25:197–201.
4. Sasayama T, Nishihara K, Kondoh T, Hosoda K, Kohmura E. MicroRNA-10b is overexpressed in malignant glioma and associated with tumor invasive factors, uPAR and RhoC. Int J Cancer. 2009;125:1407–13.
5. Lassman AB, Iwamoto FM, Gutin PH, Abrey LE. Patterns of relapse and prognosis after bevacizumab (BEV) failure in recurrent glioblastoma (GBM). J Clin Oncol. 2008;26:431–6.
6. D'Amico A, Gabbani M, Dall'Oglio S, Cristofori L, Turazzi S, Sanzone E, Maluta S. Protracted administration of low doses of temozolomide (TMZ) in the treatment of relapse glioblastoma (GBM) enhances the antitumor activity of this agent. In: Asco Meeting. 2006. p. 810–3.
7. Gladson CL, Prayson RA, Liu WM. The pathobiology of glioma tumors. Ann Rev Pathol Mech Dis. 2010;5:33–50.
8. Yu L, Maximilian D, Nathan W, Bollen AW, Aldape KD, M Kelly N, Lamborn KR, Berger MS, David B, Brown PO. Gene expression profiling reveals molecularly and clinically distinct subtypes of glioblastoma multiforme. Proc Natl Acad Sci. 2005;102:5814–9.
9. Shumin D, Nutt CL, Betensky RA, Stemmer-Rachamimov AO, Denko NC, Ligon KL, Rowitch DH, Louis DN. Histology-based expression profiling yields novel prognostic markers in human glioblastoma. J Neuropathol Exp Neurol. 2005;64:948–55.
10. Bertram JS. The molecular biology of cancer. Mol Aspects Med. 2000;21:167–223.
11. Richards SJ. A handbook of parametric survival models for actuarial use. Scand Actuar J. 2012;2012:233–57.
12. Cox DR. Regression models and life-tables. J R Stat Soc. 1972;34:527–41.
13. Crichton N. Cox proportional hazards model. J Clin Nurs. 2002;11:723.
14. Tibshirani R. Regression shrinkage and selection via the Lasso. J R Stat Soc. 1996;58:267–88.
15. Tibshirani R. The lasso method for variable selection in the Cox model. Stat Med. 1997;16:385–95.
16. Fan J, Feng Y, Wu Y. High-dimensional variable selection for Cox's proportional hazards model. J Am Stat Assoc. 2010;105:205–17.
17. Hong HG, Wang L, He X. A data-driven approach to conditional screening of high dimensional variables. 2015. Manuscript.
18. Nelander S, Wang W, Nilsson B, She QB, Pratilas C, Rosen N, Gennemark P, Sander C. Models from experiments: combinatorial drug perturbations of cancer cells. Mol Syst Biol. 2008;4:1484–94.
19. Sergio Iadevaia YL, Morales FC, Mills GB, Ram PT. Identification of optimal drug combinations targeting cellular networks: integrating phospho-proteomics and computational network analysis. Cancer Res. 2010;70:6704–14.
20. The Georgetown Database of Cancer G-DOC. https://gdoc.georgetown.edu/gdoc/. Accessed 28 Apr 2016.
21. Xie C, Mao X, Huang J, Ding Y, Wu J, Dong S, Kong L, Gao G, Li CY, Wei L. KOBAS 2.0: a web server for annotation and identification of enriched pathways and diseases. Nucleic Acids Res. 2011;39:W316–22.
22. Rabinowitz L. Mathematical Statistics and data analysis. Elsevier; 2006.
23. Zhao SD, Li Y. Principled sure independence screening for Cox models with ultra-high-dimensional covariates. J Multivar Anal. 2012;105:397–411.
24. Takashi O. Drug target validation and identification of secondary drug target effects using DNA microarrays. Tanpakushitsu Kakusan Koso. 2007;52:1808–9.
25. Behr MA, Wilson MA, Gill WP, Salamon H, Schoolnik GK, Rane S, Small PM. Comparative genomics of BCG vaccines by whole-genome DNA microarray. Science. 1999;284:1520–3.
26. He GP, Muise A, Li AW, Ro HS. A eukaryotic transcriptional represser with carboxypeptidase activity. Nature. 1995;378:92–6.
27. Ro HS, Kim SW, Wu D, Webber C, Nicholson TE. Gene structure and expression of the mouse adipocyte enhancer-binding protein. Gene. 2002;280:123–33.
28. Zhang L, Reidy SP, Nicholson TE, Lee HJ, Majdalawieh A, Webber C, Stewart BR, Dolphin P, Ro HS. The role of AEBP1 in sex-specific diet-induced obesity. Mol Med. 2005;11:39–47.
29. Majdalawieh A, Zhang L, Ro HS. Adipocyte enhancer-binding protein-1 promotes macrophage inflammatory responsiveness by up-regulating NF-kappaB via IkappaBalpha negative regulation. Mol Biol Cell. 2007;18:930–42.
30. Ladha J, Sinha S, Bhat V, Donakonda S, Rao SM. Identification of genomic targets of transcription factor AEBP1 and its role in survival of glioma cells. Mol Cancer Res. 2012;10:25–35.
31. Yu T, Scully S, Yu Y, Fox GM, Jing S, Zhou R. Expression of GDNF family receptor components during development: implications in the mechanisms of interaction. Journal of Neurosci. 1998;18:4684–96.
32. Ku MC, Wolf SA, Respondek D, Matyash V, Pohlmann A, Waiczies S, Waiczies H, Niendorf T, Synowitz M, Glass R. GDNF mediates glioblastoma-induced microglia attraction but not astrogliosis. Acta Neuropathol. 2013;125:609–20.
33. Hoelzinger DB, Tim D, Berens ME. Autocrine factors that sustain glioma invasion and paracrine biology in the brain microenvironment. J Natl Cancer Inst. 2007;99:1583–93.
34. Saletta F, Rahmanto YS, Richardson DR. The translational regulator eIF3a: the tricky eIF3 subunit! Biochim Biophys Acta. 1806;2010:275–86.
35. Ji-Ye Y, Jie S, Zi-Zheng D, Qiong H, Mei-Zuo Z, De-Yun F, Hong-Hao Z, Jian-Ting Z, Zhao-Qian L. Effect of eIF3a on response of lung cancer patients to platinum-based chemotherapy by regulating DNA repair. Clin Cancer Res. 2011;17:4600–9.
36. R-Y L, Dong Z, Liu J, J-Y Y, Zhou L, Wu X, Yang Y, Mo W, Huang W, Khoo SK. Role of eIF3a in regulating cisplatin sensitivity and in translational control of nucleotide excision repair of nasopharyngeal carcinoma. Oncogene. 2011;30:4814–23.
37. Navani S. The human protein atlas. J Obstet Gynecol India. 2011;61:27–31.
38. Parajuli P, Mittal S. Role of IL-17 in Glioma Progression. Journal of Spine & Neurosurgery. 2013; Suppl 1:s1–004.
39. McLendon R, Friedman A, Bigner D, Van Meir EG, Brat JD. Comprehensive genomic characterization defines human glioblastoma genes and core pathways. Nature. 2013;455:1061–8.
40. Akhavan D, Mischel PS. mTOR Signaling in Glioblastoma: Lessons Learned from Bench to Bedside. Neuro Oncol. 2010;12:882–9.

41. Jhanwaruniyal M, Labagnara M, Friedman M, Kwasnicki A, Murali R. Glioblastoma: molecular pathways, stem cells and therapeutic targets. Cancers. 2015;7:538–55.

42. Arshawn S, Michael K. Targeting the PI3K/AKT/mTOR signaling pathway in glioblastoma: novel therapeutic agents and advances in understanding. Tumor Biol. 2013;34:1991–2002.

43. Zhang VE, Derynck R. **Smad-dependent and Smad-independent pathways in TGF-ß family signalling**. Nature. 2003;425:577–84.

44. Joan S, Hong-Van L, Lijian S, Anderson SA, Joan M. Integration of Smad and forkhead pathways in the control of neuroepithelial and glioblastoma cell proliferation. Cell. 2004;117:211–23.

45. Hiroaki I, Tomoki T, Yasushi I, Masamichi T, Nobuhito S, Keiji M, Kohei M. Glioma-initiating cells retain their tumorigenicity through integration of the Sox axis and Oct4 protein. J Biol Chem. 2011;286:41434–41.

46. Han J, Alvarezbreckenridge CA, Wang QE, Yu J. TGF-β signaling and its targeting for glioma treatment. Am J Cancer Res. 2015;5:945–55.

47. Roy LO, Poirier MB, Fortin D. Chloroquine inhibits the malignant phenotype of glioblastoma partially by suppressing TGF-beta. Invest New Drugs. 2015; 33:1020–31.

48. Bogdahn U, Hau P, Stockhammer G, Venkataramana NK, Mahapatra AK, Suri A, Balasubramaniam A, Nair S, Oliushine V, Parfenov V. Targeted therapy for high-grade glioma with the TGF-β2 inhibitor trabedersen: results of a randomized and controlled phase IIb study. Neuro Oncol. 2010;13:132–42.

49. Huez I, Créancier L, Audigier S, Gensac MC, Prats AC, Prats H. Two independent internal ribosome entry sites are involved in translation initiation of vascular endothelial growth factor mRNA. Mol Cell Biol. 1998;18:6178–90.

50. Stoneley M, Chappell S, Jopling CL, Dickens M, Macfarlane M, Willis A. c-Myc protein synthesis is initiated from the internal ribosome entry segment during apoptosis. Mol Cell Biol. 2000;20:1162–9.

51. Lang KJD, Andreas K, Goodall GJ. Hypoxia-inducible factor-1alpha mRNA contains an internal ribosome entry site that allows efficient translation during normoxia and hypoxia. Mol Biol Cell. 2002;13:1792–801.

52. Lior B, Revital K, Iris BD, Sivan O, Silke K, Peter H, Martin P, Anja-Katrin B, Lily V. Aberrant expression of c-Jun in glioblastoma by internal ribosome entry site (IRES)-mediated translational activation. Proc Natl Acad Sci U S A. 2012;109:E2875–84.

53. Bioconductor:open source software for boinformatics. http://www.bioconductor.org/. Accessed 28 Apr 2016.

54. Miller FP, Vandome AF, Mcbrewster J. Interquartile: Interquartile Range. 2010.

55. Singh R, Mukhopadhyay K. Survival analysis in clinical trials: basics and must know areas. Perspect Clin Res. 2011;2:145–8.

56. Friedman J, Hastie T, Tibshirani R. Regularization paths for generalized linear models via coordinate descent. J Stat Softw. 2010;33:1–22.

57. Barut E, Fan J, Verhasselt A. Conditional Sure Independence Screening. J Am Stat Assoc. 2016;111:1266–77.

58. Mcclish DK. Analyzing a portion of the ROC curve. Med Decis Making. 1989;9:190–5.

59. Pepe M. An interpretation for the ROC curve and inference using GLM procedures. Biometrics. 2000;56:352–9.

60. Myers SC, Jin L. R-squared around the world: new theory and new tests. Ssrn Electron J. 2004;79:257–92.

61. Kremers WK, Kremers WK. Concordance for survival time data: fixed and time-dependent covariates and possible ties in predictor and time. Mayo Foundation. 2007. http://www.mayo.edu/research/documents/biostat-80pdf/doc-10027891.

62. Simon N, Friedman JH, Hastie T, Tibshirani R. Regularization paths for Cox's proportional hazards model via coordinate descent. J Stat Softw. 2011;39:1–13.

63. Peng H, Peng T, Wen J, Engler DA, Matsunami RK, Su J, Zhang L, Chang CC, Zhou X. Characterization of p38 MAPK isoforms for drug resistance study using systems biology approach. Bioinformatics. 2014;30:1899–907.