

# Expression of lncRNAs Associated with GPCRs and Clinical Prognosis in GBM

**Hao Wang**

Neurosurgery Department, First Affiliated Hospital of Southern University of Science and Technology, Shenzhen Municipal People's Hospital

**Yaohui Wang**

Neurosurgery Department, First Affiliated Hospital of Southern University of Science and Technology, Shenzhen Municipal People's Hospital

**wei xiang** (✉ [xiangweirene@126.com](mailto:xiangweirene@126.com))

Neurosurgery Department, First Affiliated Hospital of Southern University of Science and Technology, Shenzhen Municipal People's Hospital, Shenzhen, Guangdong, 518020, China

---

## Research article

**Keywords:** Glioblastoma, G protein coupled receptor, lncRNA, Clinical outcomes

**Posted Date:** January 13th, 2021

**DOI:** <https://doi.org/10.21203/rs.3.rs-143404/v1>

**License:** © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

---

# Abstract

## Background:

Glioblastoma (GBM) is a primary malignant tumor of the central nervous system with a poor prognosis. Long non-coding RNAs (lncRNAs) play a variety of key regulatory roles in a variety of biological processes, and have an important influence on the occurrence and development of tumors by regulating the expression of target genes. However, their role in the prognosis of GBM is still lacking in accurate prognostic markers.

This study aims to establish an effective lncRNAs model to evaluate the prognosis of GBM.

## Methods:

We used data of mRNA, lncRNA and clinical follow-up from The Cancer Genome Atlas (TCGA) to conduct univariate analysis, clustering analysis, coding gene difference analysis, Kyoto Encyclopedia of Genes and Genomes (KEGG) analysis and Gene Ontology (GO) analysis for GBM patients. lncRNAs that are closely related to the survival and prognosis of GBM were found and a multiple regression model was constructed to calculate the risk score of the samples, so as to accurately predict the clinical prognosis of GBM patients.

## Results:

Through multiple systematic analysis, we found 5 lncRNAs that are closely related to the survival and prognosis of GBM, and these 5 lncRNAs can be used as independent prognostic factors. Through GO enrichment analysis and KEGG pathway analysis, we found that there is a close relationship between GBM and G protein coupled receptors. Therefore, 93 mRNAs associated with G-protein-coupled receptors and 5 lncRNAs associated with independent prognostic factors were selected to calculate the correlation, respectively. In other words, a tumor-specific lncRNAs/mRNAs co-expression network was constructed through biological prediction based on correlation analysis. The results showed that lncRNAs and mRNAs interregulated very closely in GBM patients. In addition, we also constructed a multiple regression model with these 5 lncRNAs which was able to calculate the risk score of each sample to accurately predict the clinical prognosis of GBM patients.

## Conclusion:

Our study found lncRNAs independently were related to the prognosis of GBM, and successfully constructed a multiple regression model related to lncRNAs, providing a new perspective for better evaluation of the role of lncRNAs in the clinical prognosis of GBM.

# Introduction

Glioblastoma originates from brain glial cells, and its morbidity and mortality rank first among the malignant tumors of the central nervous system. It is the most common primary malignant tumor of the central nervous system [1], accounting for about 31% of the tumors of the central nervous system and 81% of the primary malignant tumors of the central nervous system [2]. The glioblastoma is divided into I ~ IV level by WHO classification of central nervous system Tumors in 2007 [3]. I, II level is low level glioblastoma, III, IV level is high level glioblastoma, pathological level is higher, the worse prognosis. At present, the main therapeutic methods for glioblastoma include surgery, chemotherapy, radiotherapy, targeted therapy, etc. Surgery is still the main treatment, and some glioblastoma often need adjuvant postoperative chemoradiotherapy. However, the prognosis of patients with glioblastoma is still poor, among which, GBM has the highest mortality among adult brain tumors, with a 5-year survival rate of about 5% [4,5] and an average survival time of less than 18 months [6]. In recent years, although surgical operation, radiotherapy, chemotherapy, adjuvant therapy and other treatment methods have been continuously expanded [7], the prognosis of glioblastoma patients is still not optimistic [8]. This is largely due to the limited understanding of its molecular pathogenesis, so effective targeted therapies for GBM are currently almost unapproved. Therefore, it is important to identify effective targets as early as possible and explore the potential molecular mechanisms associated with the clinical prognosis of glioblastoma.

GPCRs are the most functionally diverse membrane protein family consisting of numerous receptors. More than 800 genes have been identified to be responsible for coding GPCRs [9]. Ligands such as light photons, odor molecules, hormones, neurotransmitters, chemokines, etc., can interact with the extracellular sites of GPCRs, creating molecular conformational changes in the transmembrane domains and intracellular sequences [10]. GPCRs are highly expressed in colon, prostate, and melanoma cancers, prostaglandin receptors are highly expressed in breast, colon, and prostate cancers, and chemokine receptors are highly expressed in lung and pancreatic cancers [11]. Lysophosphatidic acid (LPA) phosphorylated receptors, LPA1-LPA6 belongs to the GPCR family. The LPA receptors couple to multiple G proteins, G12/13, Gi, Gq, and probably Gs . LPA can induce the invasion of cancer cells through the Rho transduction pathway in breast and ovarian cancer cells [12]. G protein coupled receptor (G protein-coupled receptor 30 (GPR30)) is widely expressed in human normal tissues, including heart, brain, placenta, lung, liver, kidney, skeletal muscle, many recent studies also in malignant tumor, breast cancer, endometrial cancer, ovarian cancer found GPR30 expression, thus proving GPR30 participate in the occurrence of malignant tumor development [13].

However, these targets have not been considered as cancer targets of interest to a large extent. In recent years, GPCR mRNA signaling could characterize specific tumor types and was associated with survival, the expression of many cancer-related genes and pathways. Therefore, GPCRs can be used as functional oncogenes, biomarkers, surface antigens and pharmacological targets [14]. GPCR US28 was used as target for nanobody-directed photodynamic therapy (PDT) to kill the glioblastoma cells both in 2D and 3D cultures upon illumination with near-infrared light [15].

Non-coding RNA, including short RNA represented by miRNA, siRNA and piRNA and long non-coding RNA (lncRNA), among which lncRNA is the most transcribed. In recent years, studies on epigenetic regulation of glioblastoma have attracted great attention from scholars, especially the lncRNA closely related to cell regulation, which may provide a new treatment method at the molecular level for inhibiting the growth and recurrence of glioblastoma. In the past 30 years, lncRNA has been found to regulate gene expression through interaction with DNA, RNA, protein or chromatin remodeling [16]. For example, lncRNA can act as cell “address codes” to transfer protein complexes to appropriate positions on chromosomes and lead to subsequent activation or deactivation [17]. Compared with small interfering RNAs and microRNAs, lncRNAs have greater potential in target recognition and can be folded into more advanced structures, contributing to chromatin remodeling and transcription and post-transcription regulation [18].

The Cancer Genome Atlas (TCGA) is an open public data platform, including 38 kinds of 11, 000 cases of Cancer patient information. The mining and analysis of these huge data will help researchers comprehensively study the molecular mechanisms of various cancers, identify new therapeutic targets and tumor-related molecular markers, and thus improve the diagnostic ability and therapeutic effect of various human malignant diseases [19]. In this study, we selected glioblastoma brain samples from the TCGA database for differential expression analysis of gene transcriptional levels, including the data of coding RNA (namely mRNA) and lncRNA. For mRNA, GO and KEGG were used for enrichment analysis of the differentially expressed genes, respectively. By functional analysis of the differentially expressed genes in glioblastoma, it was found that the G protein receptor pathway was enriched significantly. For lncRNAs, we performed univariate COX regression analysis on glioblastoma samples with prognostic survival data in the TCGA database to screen out lncRNAs with significant prognostic differences. Next, lncRNAs with significant differences in expression and prognosis were selected, which were further screened by stepwise COX multivariate regression analysis as independent prognostic factors, and a risk model that could be used to predict the prognosis of patients was constructed, and lncRNA targets that might be related to the prognosis of glioblastoma were extracted. In the end, we selected GO and KEGG to enrich significantly different mRNA and expression data in glioblastoma samples that can act as independent prognostic factor lncRNA, and evaluated the association between them by constructing a co-expression network.

## Methods And Materials

### 2.1 Data collection and processing

RNA-SEQ raw data can be downloaded from TCGA (<https://tcga-data.nci.nih.gov/tcga/>). Then, the counts expression of the extracted RNA. According to the gene encoding proteins, the genes are divided into coding genes and non-coding genes, and corresponding counts are extracted for subsequent difference analysis.

### 2.2 Differential analysis

We also used the R package (R-3.6.3 <https://www.r-project.org/>) of "DESeq2" to analyze the differences between mRNA and lncRNA from the GBM sample and the normal sample. In addition to R distribution itself bring R package, other packages we used need to download and install on Bioconductor packages website (<http://www.bioconductor.org/>).

First, based on the R packages of "GGplot2", "Pheatmap", "Gplot", "Stringr", "Tidverse", "Tibble", "FactoMineR", "Factoextra", "String", "Factoiner", "Factoextra", etc., the RLE box plot, principal component plot (PCA) and cluster analysis plot were used to detect the quality of RNA sequencing data and remove the large discrete samples from the deviated groups. Then, all samples were standardized by using the VST function of DESeq2 package, and the differentmrna and lncRNA were screened by using the "DESeq" function. Finally, the mRNA annotation of the coding gene was carried out by using the "org.hs.eg.db" and the "clusterProfiler" R software package, and the lncRNA was annotated by the non-coding file gencode. V34, where the difference threshold was  $FDR < 0.01$  &  $P \text{ value} < 0.01$  &  $|\log_2FC| > 1$ .

### 2.3 Biological process and pathway enrichment analysis

Gene ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway analyses were performed using Toppgene (<https://toppgene.cchmc.org/>). The selection criteria are the number of genes that fall on a term/GO difference  $\geq 4$  &  $P \text{ value} < 0.05$ . The term/GO is drawn according to an enrich factor value sorted in descending order of size, taking the first 30 results.

Enrich factor = (the number of different genes in a term/the total number of different genes) / (the total number of genes in database term/the total number of genes in the database). The bar graph results of Top30 GO and KEGG were plotted by SigmaPlot 14.0.

### 2.4 univariate Cox proportional regression analysis

We chose 167 tumors samples from TCGA with control-survival information and began with the function rowSums in the R package (rowSums >1). A univariate Cox Proportional regression model was used to calculate the association between the expression of each lncRNA and OS.  $P \text{ value} < 0.05$  &  $HR1.5$  ( $HR > 1.5$  or  $HR < 0.67$ ) were used to analyze the effect of lncRNA on the prognosis of GBM. R package of GGplot2 is used for the volcanic diagram.

### 2.5 Venn analysis

The collection analysis software venny 2.1.0 (<https://bioinfogp.cnb.csic.es/tools/venny/index.html>) was used to draw the Venn diagram, the prognosis significantly lncRNA ( $p \text{ value} < 0.05$  &  $HR1.5$  ( $HR > 1.5$  or  $HR < 0.67$ )) and differentially expressed lncRNA ( $FDR < 0.01$  &  $p \text{ value} < 0.01$  &  $|\log_2FC| > 1$ ) take the intersection.

### 2.6 Muti Cox proportional regression analysis and stepwise Cox proportional hazards regression

Combined with the overall survival rate of GBM patients in TCGA, the R package of survival was used to perform multivariate COX regression analysis and stepwise COX multivariate regression analysis on the 12 lncRNAs to obtain lncRNAs (pvalue < 0.05). To make our model more optimized and practical, Multi-Cox proportional regression analysis and a stepwise Cox proportional hazards regression model was used. Finally, a risk score formula was calculated by taking into account of the expression of optimized lncRNAs and correlation estimated Cox regression coefficients: Risk Score =  $1.43 \times \text{AC021594.1} + 3.92 \times \text{AC105046.1} + 0.27 \times \text{CRNDE} + 0.92 \times \text{LINC01574} + 1.11 \times \text{LINC02320}$ . Patients with GBM were classified into the high- or low-risk group by ranking the given risk score. Differences in overall survival (OS) between the high- and low-risk group in the cohorts were assessed using Kaplan–Meier method and two-tailed log-rank test. A Cox proportional hazards regression model was used to identify independent prognostic factors.  $P < 0.05$  was set as significant difference in all statistical methods.

## 2.7 Construction of the LncRNA/mRNA Coexpression Network

To construct the LncRNA and mRNA coexpression network (CNC network), we calculated the Pearson correlation coefficient and R value to evaluate LncRNA-mRNA correlation [20].

The network construction procedure includes:

- (1) Preprocess data: 93 differentially encoded genes and expressions of the above 5 lncRNAs in 160 samples were selected from 3 KEGG and 2 GO enrichment analyses related to GPCRs.
- (2) R package of Psych was used to calculate the correlation between each lncRNA and 93 encoding genes Pearson, and top20 relationship pairs with significant  $FDR < 0.01$  &  $P < 0.01$  were selected.
- (3) Pearson correlation between lncRNAs involved in relation pair (2) was calculated, and significance  $FDR < 0.01$  &  $P < 0.01$  relation pair was selected.
- (4) Pearson correlation between the encoding genes involved in relation pairs was calculated, and Top100 relation pairs with significance  $FDR < 0.01$  &  $P < 0.01$  were selected.
- (5) Cytoscape (3.5.1) software was used to construct the network.

## 2.8 Statistical analysis

Differential analysis and normalization were mainly conducted by “DESeq2” package. Cox regression model or Kaplan-Meier analysis with log-rank test was performed by “survival” package. Student’s t test was used for continuous variables, while categorical variables were compared by  $\chi^2$  test. The Wilcoxon rank-sum test was a nonparametric statistical test mainly utilized for comparing two groups and the Kruskal-Wallis test was suitable when it comes to two or more groups. All statistical analysis was achieved in R software (Version 3.6.3). We considered  $p < 0.05$  to be statistically significant.

# Results

### 3.1 Extraction of RNA sequencing data

The RNA sequencing data, clinical information and survival information files of 155 GBM primary cancer tissues and 5 normal brain samples are all downloaded in TCGA. The TCGA includes data of 38 types of tumors, including 6 types of data (copy number, DNA methylation, gene expression RNAseq, phenotype, Somatic mutation, miRNA expression). In this study, we only took transcriptome data, including coding and non-coding RNA data.

### 3.2 Difference analysis of GBM samples

Violin diagram and box diagram were used to show the overall distribution of mRNA and LncRNA of all samples, and volcanic diagram and heat diagram were used to statistical and visual display the results of different mRNA and LncRNA. R package of pheatmap was used for the heatmap, and R package of ggplot2 was used for the box diagram, volcano diagram and violin diagram.

A heatmap can intuitively show the expression levels of different genes in each sample, cluster them according to the expression data of different genes, and obtain the relationship between the expression patterns of samples and genes, so as to draw a correlation conclusion with biological significance. In the heat map, each column represents one sample, including 155 GBM samples and 5 normal samples. Each line represents a gene (mRNA and LncRNA). Firstly, statistical software R was used to conduct statistical analysis of the differences between the mRNA and LncRNA of GBM samples and normal samples ( $FDR < 0.01$  &  $P \text{ value} < 0.01$  &  $|\log_2FC| > 1$ ). Finally, 547 differentially expressed mRNAs were selected from 19261 mRNAs, and 471 differentially expressed LncRNAs were selected from 11307 lncRNAs. Genes expressing similar expressions (mRNA and LncRNA) were clustered together with the samples. Black represents 0, indicating no change in gene expression level, red represents increased expression level, and green represents decreased expression level. The brightness of the color represents the degree to which gene expression levels rise or fall. It can be seen from the results that mRNA and LncRNA are mostly up-regulated in the 5 normal samples compared with GBM(Fig 1E,F).

The violin diagram and box diagram can be combined to show the distribution of gene expression abundance in each sample. The results showed that the abundance difference of mRNA and LncRNA in these 160 samples was less than 5%, that is to say, the abundance difference of coding and non-coding genes in normal and abnormal samples was not significant(Fig 1 A,C).

To be more specific, we also found by barplot that 430 of the 547 mRNAs were down-regulated and 117 were up-regulated, compared with the normal sample. Among the 471 differentially expressed lncRNAs, 422 were down-regulated and 49 were up-regulated (Fig 2A).

The volcano plot is used to show the significantly different expressions of the two sets of samples. The p-value and Fold change values obtained by accurate T-test statistical analysis were used to draw the volcano map between the two groups. In the volcano plot, one of the coordinates shows the negative log of P-values computed by the T-test, and the other shows the converted log of P-values compared to the

two conditions. Similarly, we can also find through the volcano map that compared with the normal sample, the GBM sample down-regulated more in the differentially expressed mRNA and LncRNA (Fig 1B,D).

### **3.3 Enrichment analysis of GO and KEGG pathways of DEGs**

We conducted GO and KEGG enrichment analysis of the differentially expressed genes respectively, and selected the GO and KEGG results of Top30. In the results of BP, we found that the main difference between gene enrichment in embryonic bone system development and formation, chemical synapses and synaptic transmission, eating behaviors, G protein coupled receptor pathway and G protein coupling the second messenger cyclic nucleotide receptor signaling pathways, neuropeptide signaling pathways, and other functions. In the results of MF, we found that the differentially expressed genes were mainly concentrated in neurotransmitter binding and receptor activity, G protein-coupled peptide receptor activity, G protein-coupled receptor activity and other functions. In the results of the KEGG pathway, we found that the differentially expressed genes were mainly concentrated in the interaction in stimulating nerve tissues, G-protein-coupled receptor ligand binding, G-protein-coupled receptor downstream signaling, and G-protein-coupled receptor pathways, among which G-protein-related pathways accounted for three of the first ten pathways. This indicates that there is a close relationship between GPCRs and GBM. More detailed GO and KEGG enrichment analysis results are shown in the Fig 3 below. This will help us further understand the role of DEGs in the genesis and development of GBM.

### **3.4 Univariate COX survival regression analysis**

In order to better classify the samples, there were 167 GBM tumor samples with prognostic survival data in TCGA, including 155 GBM primary cancer tissues and 12 GBM metastatic cancer tissues (one of which has no survival data), including 15110 lncRNAs in total, and 14480 lncRNAs were left after the deletion of some lncRNAs that appeared only a few times. Finally, a total of 642 lncRNAs with significant prognostic correlation were calculated by univariate COX survival regression analysis (Fig 2B).

### **3.5 LncRNA-based prognostic biomarkers in GBM patients**

In the significance analysis results of GBM samples, we obtained a total of 471 lncRNAs with significant differences, and 642 lncRNAs with significant prognostic differences in univariate COX survival regression analysis. The analysis software was used to take the intersection of the two (Venn diagram) to obtain 12 lncRNAs, which were not only differentially expressed in normal tissues and cancer tissues, but also had prognostic differences in cancer tissues (Fig 2C).

### **3.6 LncRNA with independent prognostic factor**

Prognostic biomarkers based on lncRNA in GBM patients, namely 12 lncRNAs obtained by intersection, were used for prognostic survival curves respectively. Univariate COX survival regression curve analysis showed that among the 12 lncRNAs (Fig 4), only AC097641.1 and AC117453.1 were  $HR < 1$ , is a protective factor. While other lncRNA  $HR > 1$ , is a risk factor. The prognosis of patients in the high-risk group was

significantly worse than that of patients in the low-risk group. Then, stepwise COX multivariate regression was used for the 12 lncRNAs obtained, and 5 lncRNAs ( $P < 0.05$ ) (Fig 5).

### 3.7 Multiple regression model risk Score survival analysis

We constructed a multiple regression model with 5 lncRNAs that could act as independent prognostic factors. The risk score of each sample was calculated based on a multiple regression model consisting of five lncRNAs (Risk Score  $= 1.43 \times \text{AC021594.1} + 3.92 \times \text{AC105046.1} + 0.27 \times \text{CRNDE} + 0.92 \times \text{LINC01574} + 1.11 \times \text{LINC02320}$ ) Then, according to the median risk score, all cases were divided into the high-expression group and the low-expression group, and the correlation between risk score and prognosis was calculated.  $P < 0.0001$  was significantly correlated with prognosis.

Survival analysis using scatter plot (Fig 6.A, B), heat map (Fig 6.C), and Kaplan-Meier plot (Fig 6.D), all showed that the prognosis of the high-risk group was significantly lower than that of the low-risk group. It is proved that the multiple regression model can accurately predict the clinical prognosis of GBM patients.

### 3.8 Coexpression network of coding gene/non-coding gene

Although more and more studies are trying to reveal the functional significance of lncRNAs, the biological role of most lncRNAs is still unknown. Biological processes and cellular regulatory networks are very complex and involve the interaction of various molecules, such as proteins, RNA and DNA. Therefore, we constructed a coexpression network of coding gene/non-coding gene based on 155 GBM samples and 5 normal brain samples in TCGA database to study the potential interaction between GPCR-related mRNAs and lncRNAs in each sample. The co-expression network consisted of 93 GPCR related mRNAs and 5 lncRNAs (Fig 7). The figure showed that 5 lncRNAs (AC021594.1, AC105046.1, CRNDE, LINC01574, LINC02320) were all related to a large number of targeted mRNAs. AC021594.1 can target 23 network nodes, AC105046.1 22 network nodes, CRNDE 20 network nodes, LINC01574 22 network nodes and LINC02320 23 network nodes. The co-expression network also shows that a coding gene can also target many network nodes. GRIN1 targets 23 network nodes, AC105046.1 targets 22 network nodes, CRNDE targets 20 network nodes, LINC01574 can target 22 network nodes, LINC02320 can target 23 network nodes, GABRA1 can target 19 network nodes, GABRG2 can target 18 network nodes, RASAL1 can target 18 network nodes, CHRM1 can target 16 network nodes, SST can target 16 network nodes. These results showed that lncRNAs and GPCR-related mRNAs in GBM patients interregulated very closely.

In addition, external data sets in CGGA were also used to verify the above results. In the CGGA database, one of the lncRNAs, namely CRNDE, could be found, and the results were consistent with the previous ones. In other words, the prognosis of the high-risk group was significantly lower than that of the low-risk group, which was significantly correlated with the prognosis (Fig 8).

## Discussion

GBM is one of the most common primary malignant tumors of the central nervous system. The etiology of this disease is unknown [21]. The standardized treatment options of GBM include surgical excision within the maximum safety range, postoperative adjuvant radiotherapy, and chemotherapy. The most common is the cytotoxic alkylating agent Temozolomide. Almost all GBM eventually relapsed [22]. Despite great progress in recent years, the mortality rate of GBM remains high. GBM survival rate is inversely proportional to age. Specifically, about 5% of GBM patients survive five years after diagnosis. The median survival time of untreated GBM patients is only about 3 months [23]. A better understanding of the genetic and molecular pathogenesis of GBM could lead to more effective treatments.

LncRNA does not participate in protein coding, many recent studies have found that LncRNA is involved in a variety of life activities in vivo, such as the diversity of embryonic stem cells, the regulation of cell cycle, and the occurrence and development of cancer [24]. Nowadays, more and more studies have confirmed that LncRNA can be used as a diagnostic marker for cancer and even as a therapeutic option [25].

The human genome contains more than 50,000 LncRNA genes [26]. lncRNAs play an important role in cell biology by regulating gene expression at different levels, including chromatin remodeling, transcription and post-transcription processing [27,28]. More and more evidence indicates that lncRNAs dysfunction may play a role in the occurrence and development of tumors [29,30]. Some studies have shown that abnormal LncRNA expression is related to the prognosis of GBM and can be used as a tumor-related predictor. Compared with protein-coding genes, lncRNAs have significant advantages as diagnostic and prognostic biomarkers [31]. In the global analysis of GBM samples, 584 of them are related to poor prognosis. 282 lncRNAs were associated with a better survival rate in GBM patients and were confirmed as prognostic biomarkers for GBM [32]. We attempted to use mRNA data, LncRNA data and clinical follow-up data from the tumor genome atlas to identify LncRNA related to GBM prognosis and provide potential therapeutic targets for treatment. lncRNAs have become novel biomarkers for the diagnosis and prognosis of various human cancers [33-36]. We speculated that the differential expression of lncRNAs might be used as biomarkers to determine the prognosis of GBM patients, and previous studies have shown that several lncRNAs, including MEG3, H19 and Gas5, are abnormally expressed in all GBM patients [37-40].

First, we downloaded all GBM-related transcriptomic sequencing data, clinical information and survival information files from TCGA, including 155 GBM primary cancer tissues and 5 normal samples. Then we conducted a significance analysis of the differentially expressed mRNA and LncRNA in GBM primary cancer tissues and normal sample tissues respectively. Among 19261 mRNAs, 547 mRNAs showed significant differences, 430 were down-regulated, and 117 were up-regulated, that is to say, in the normal sample, most of them are up-regulated. Among 11307 lncRNAs, 471 lncRNAs showed significant differences, 422 were down-regulated and 49 were up-regulated, that is to say, most of them were also up-regulated in the normal sample. Therefore, univariate COX survival regression analysis was performed on 167 GBM patients in TCGA, including 155 GBM primary cancer tissues and 12 GBM metastatic cancer tissues (one of which has no survival data), containing a total of 14480 lncRNAs, and finally 642 lncRNAs

with significant prognostic differences were screened out. Among 471 lncRNAs with significant difference and 642 lncRNAs with different prognosis, 12 lncRNAs (AC005632.5, AC021594.1, AC097641.1, AC105046.1, AC117453.1, CRNDE, HOTAIR, HOXC-AS2, HOXC-AS3, LINC01111, LINC01574, and LINC02320) were obtained from the intersection of the two (Venn diagram).

The 547 differentially expressed mRNAs obtained above were enriched by GO and KEGG respectively, and the GO and KEGG results of Top30 were selected. In the results of GO enrichment analysis, we found that the main difference between gene enrichment in embryonic bone system development and formation, chemical synapses and synaptic transmission, G protein coupled receptor pathway and G protein coupling the second messenger cyclic nucleotide receptor signaling pathways, neuropeptide signaling pathways, as well as the combination of neurotransmitters and receptors on activity, G protein coupling peptide receptor activity, G protein coupled receptor activity, etc. In the results of KEGG pathway analysis, we found that the differentially expressed genes were mainly concentrated in the interaction in stimulating neural tissue, G-protein-coupled receptor ligand binding, G-protein-coupled receptor downstream signaling, and G-protein-coupled receptor pathways, among which G-protein-related pathways accounted for three of the first ten pathways. Based on GO enrichment analysis and KEGG pathway analysis, we found that there was a close relationship between GBM and G protein-coupled receptors. This had not been found in previous studies.

G protein coupled receptors (GPCRs) is arguably the most common receptor in research, and also an important target in drug development. According to statistics, over 30% of clinical prescription drugs directly act on GPCR, which is the most widely used targeted gene family among drugs approved by The Food and Drug Administration (FDA). The structural variability of GPCR makes it a very specific protein, which makes members of this family the most important drug targets. GPCRs are also the largest family of cell membrane receptors involved in signal transduction [41], and GPCR-mediated signaling pathways include nerve conduction, embryonic development, regulation of the immune system, control of sight and smell, and other physiological activities [42]. GPCR can transfer various extracellular signals such as neurotransmitters and hormones into cells, thus playing an important role in the regulation of cell function, which is also closely related to the occurrence, development and metastasis of a variety of human tumors. The regulation of tumor by GPCR mainly involves the alteration of the level of the second messenger in the cell and then the realization of this process through the signal transduction pathway associated with it [43].

In order to fully understand how lncRNA affects GBM, we constructed a tumor-specific lncRNAs/mRNAs co-expression network based on biological prediction based on correlation analysis. In GO enrichment analysis and KEGG pathway, there were 93 mRNAs associated with GPCRs, and 5 lncRNAs (AC021594.1, AC105046.1, CRNDE, LINC01574, LINC02320) were all related to a large number of the GPCRs mRNAs. Then, stepwise COX multivariate regression was used and then we obtained five lncRNAs (AC021594.1, AC105046.1, CRNDE, LINC01574, LINC02320) that could act as independent prognostic factors ( $P < 0.05$ ). Then, multiple regression model was constructed to calculate the Risk Score of each sample (Risk Score

=1.43\*AC021594.1+3.92\*AC105046.1+ 0.27\*CRNDE + 0.92\*LINC01574 + 1.11\*LINC02320), so as to achieve the purpose of predicting the clinical prognosis of GBM patients.

The results of survival analysis showed that the prognosis of patients in the high-risk group was significantly lower than that of patients in the low-risk group, which proved that the multiple regression model established by us could predict the clinical prognosis of GBM patients more accurately. Studies have shown that colorectal neoplasia differentially expressed (CRNDE) play a role in promoting cancer in glioblastoma [44]. Compared with normal brain tissue, THE CRNDE of GBM patients was up-regulated 32 times, and the expression level of CRNDE was associated with glioblastoma grade [45]. Meanwhile, CRNDE can also down-regulate miR-186 and inhibit the expression of downstream target gene XIAP (X-linked apoptotic inhibitor) and PAK7 [P21 protein (Cdc42/Rac) -activated kinase 7], thus promoting the proliferation of human GSC [46]. In addition to these observations, Wang et al. [47] showed that CRNDE overexpression could promote the growth of glioblastoma cells through mammalian target in vitro and in vivo rapamycin (mTOR) signaling pathways.

In addition, in order to verify the predictive effect of these five lncRNAs on the prognosis of GBM patients, we used the data obtained from Chinese Glioma Genome Atlas (CGGA) as the validation set, downloaded and analyzed the corresponding CRNDE RNA-SEQ data and clinical information in CGGA, and obtained comparable results. Therefore, we speculated that the expression of these lncRNAs might become effective predictors of survival and potential prognostic biomarkers for GBM patients.

GBM is a primary malignant tumor of the central nervous system with a poor prognosis. Analysis of global GBM samples has found a variety of prognostic related lncRNAs, however, accurate prognostic markers are still lacking. We used mRNA data, lncRNA data and clinical follow-up data from TCGA to conduct univariate analysis, clustering analysis, coding gene difference analysis, KEGG analysis, GO analysis, etc., for GBM patients. Through multiple systematic analysis, we found 5 lncRNAs that are closely related to the survival and prognosis of GBM, and these 5 lncRNAs can be used as independent prognostic factors. In addition, we also constructed a multiple regression model with these 5 lncRNAs. The multiple regression model composed of these five lncRNAs was able to calculate the risk score of each sample to accurately predict the clinical prognosis of GBM patients. So we constructed an effective multiple regression model, which can effectively predict the clinical prognosis of GBM patients.

## Conclusion

In this study, GBM-specific mRNAs and lncRNAs were identified through bioinformatics analysis of tens of thousands of candidate RNAs. Finally, a multiple regression model related to lncRNAs was successfully constructed and it was found that these lncRNAs could be used as independent prognostic biomarkers for GBM. Our study provides a new perspective to better evaluate the role of lncRNAs in the clinical prognosis of GBM.

## Declarations

We declare that we have no financial and personal relationships with other people or organizations that can inappropriately influence our work, there is no professional or other personal interest of any nature or kind in any product, service and/or company that could be construed as influencing the position presented in, or the review of, the manuscript entitled “Expression of lncRNAs associated with GPCRs and clinical prognosis in GBM”.

We confirm that we would like support with depositing and managing our data, and we can also submit datasets to Springer Nature as part of our publisher's [Data Support Services](#). All data generated or analysed during this study are included in this published article.

**Ethics approval and consent to participate** There are no animal or human trials involved.

**Consent for publication** Not applicable.

**Competing interests** The authors declare that they have no competing interests.

**Funding** Not applicable

**Authors' contributions** All authors analyzed and interpreted the research data. All authors are the major contributors in writing the manuscript. Meanwhile, all authors read and approved the final manuscript.

**Acknowledgements** We appreciated Neurosurgery Department, First Affiliated Hospital of Southern University of Science and Technology, Shenzhen Municipal People's Hospital for supporting the study.

## Authors' details

<sup>1</sup> Hao Wang, Neurosurgery Department, First Affiliated Hospital of Southern University of Science and Technology, Shenzhen Municipal People's Hospital, Shenzhen, Guangdong, 518020, China.

<sup>1</sup> Yaohui Wang, Neurosurgery Department, First Affiliated Hospital of Southern University of Science and Technology, Shenzhen Municipal People's Hospital, Shenzhen, Guangdong, 518020, China

\* Wei Xiang, Neurosurgery Department, First Affiliated Hospital of Southern University of Science and Technology, Shenzhen Municipal People's Hospital, Shenzhen, Guangdong, 518020, China

## References

1. Fu S, Liang M, Wang Y, et al: **Dual-modified novel biomimetic nanocarriers improve targeting and therapeutic efficacy in glioma.** *ACS Appl Mater Interfaces* 2019, 11(2): 1841-1854.
2. Chintagumpala M, Gajjar A: **Brain tumors.** *Pediatr Clin North Am* 2015, 62(1) : 167-178
3. Komori T: **The 2016 WHO Classification of Tumours of the Central Nervous System: The Major Points of Revision** *Neurologia medico-chirurgica* 2017, 57(7): 301-311.
4. Ostrom Q T, Gittleman H, Stetson L, et al: **Epidemiology of gliomas.** *Cancer Treat Res* 2015, 163: 1-14.
5. Lapointe S, Perry A, Butowskin A: **Primary brain tumours in adults.** *The Lancet* 2018, 392: 432-446.

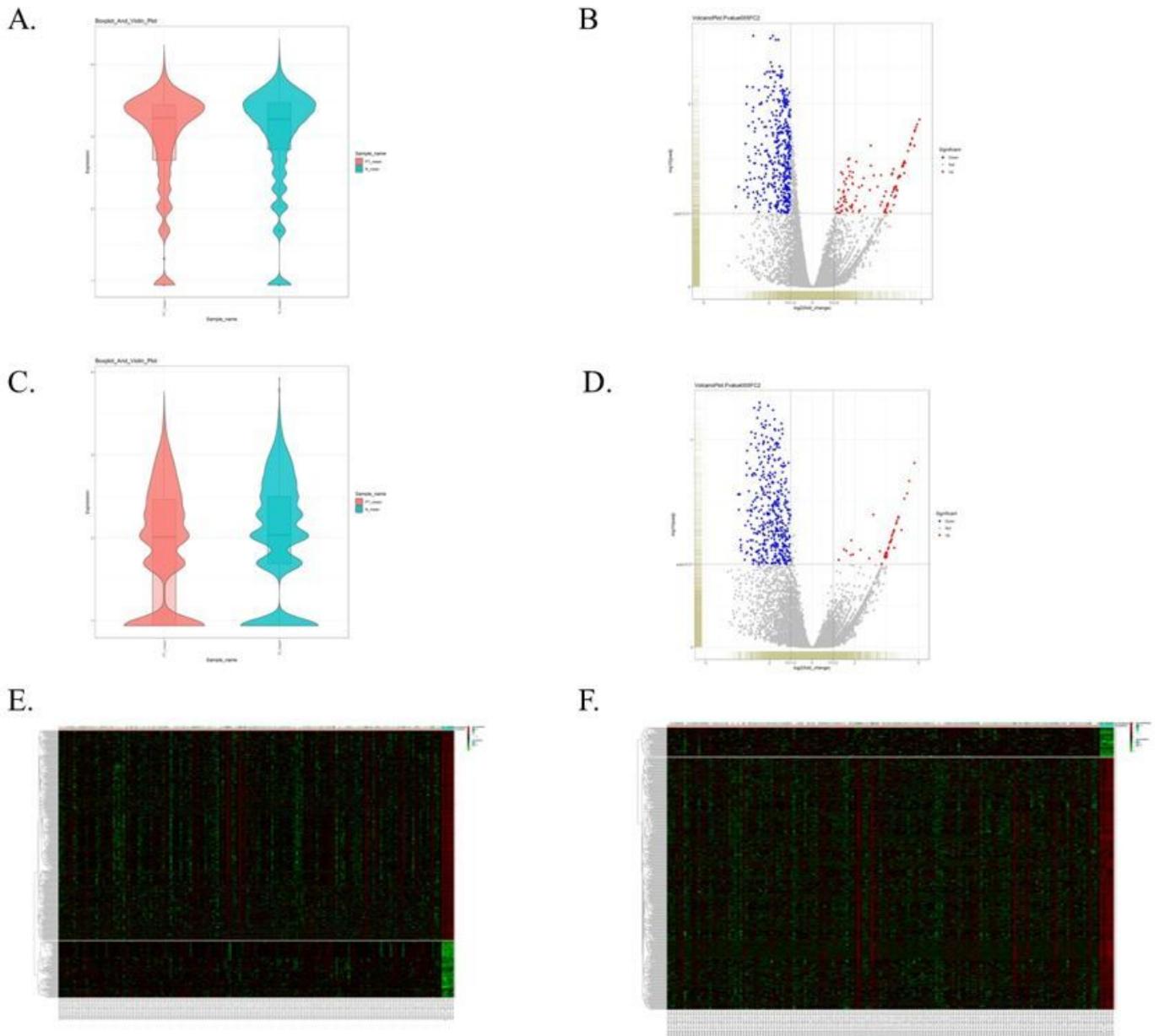
6. Raucher D, Dragojevic S, Ryu J: **Macromolecular Drug Carriers for Targeted Glioblastoma Therapy: Preclinical Studies, Challenges, and Future Perspectives.***Frontiers in Oncology* 2018, 8.
7. Mansouri A, Mansouri S, Hachem L D, et al: **The role of 5-aminolevulinic acid in enhancing surgery for high-grade glioma, its current boundaries, and future perspectives: A systematic review.***Cancer*
8. Delgado-Lopez PD, Corrales-Garcia EM: **Survival in glioblastoma: a review on the impact of treatment modalities.***Clin Transl Oncol* 2016, 18(11) : 1062-1071.
9. Duc NM, Kim HR, Chung KY: **Structural mechanism of G protein activation by G protein-coupled receptor.** *Eur J Pharmacol* 2015, 763: 214-222.
10. Brogi S, Tafi A, DLaurent D, et al: **Discovery of GPCR ligands for probing signal transduction pathways.***Frontiers in Pharmacology* 2014, 5: 255.
11. Rachel B S, Myriam M, Arun K, et al: **G Protein-Coupled Receptors in Cancer.***International Journal of Molecularences* 2016, 17(8): 1320.
12. Oyesanya RA, Greenbaum S, Dang D, et al: **Differential requirement of theepidermal growth factor receptor for G protein-mediated activation of transcription factors by lysophosphatidic acid.** *Mol Cancer* 2010, 8(9): 1-13.
13. Prossnitz ER, Arterburn JB, Smith HO, et al: **Estrogen signaling through the transmembrane G protein-coupled receptor GPR30.***Annu Rev Physiol* 2008,70: 165-190.
14. Gutierrez A N, Mcdonald P H: **GPCRs: Emerging anti-cancer drug targets.** *Cellular Signalling* 2018, 41: 65-74.
15. Groof T W M D, Mashayekhi V, Fan T S, et al: **Nanobody-Targeted Photodynamic Therapy Selectively Kills Viral GPCR-Expressing Glioblastoma Cells.***Molecular Pharmaceutics* 2019, 16(7).
16. Fatica A, Bozzoni I: **Long non-coding RNAs: new players in cell differentiation and development.** *Nat Rev Genet* 2014, 15(1): 7-21.
17. Batista PJ, Chang HY: **Long non-coding RNAs: cellular address codes in development and disease.***Cell* 2013, 152(6): 1298-1307.
18. Sahu A, Singhal U, Chinnaiyan AM: **Long non-coding RNAs in cancer: from function to translation.***Trends Cancer* 2015, 1(2): 93-109.
19. Hehuan, Zhu, Jun, et al: **Functional Long Noncoding RNAs (lncRNAs) in Clear Cell Kidney Carcinoma Revealed by Reconstruction and Comprehensive Analysis of the lncRNA-miRNA-mRNA Regulatory Network.***Medical Science Monitor International Medical Journal of Experimental & Clinical Research*
20. Liao Q, Liu CN, Yuan XY, et al: **Large-scale prediction of long non-coding RNA functions in a coding-noncoding gene co-expression network.***Nucleic Acids Res* 2011, 39: 3864-3878.
21. Fisher J L, Schwartzbaum J A, Wrensch M, et al: **Epidemiology of Brain Tumors.** *Neurologic Clinics* 2007, 25(4):867-890.
22. Arvold ND, Reardon DA: **Treatment options and outcomes for glioblastoma in the elderly patient.** *Clin Interv Aging* 2014, 9: 357-367.

23. Tamimi AF, Juweid M: **Epidemiology and Outcome of Glioblastoma.** *In: De Vleeschouwer S, editor. Glioblastoma. Brisbane (AU): Codon Publications*
24. Kretz M, Siprashvili Z, Chu C, et al: **Control of somatic tissue differentiation by the long non-coding RNA TINCR.***Nature* 2013,493(7431):231-235.
25. Kim J, Piao H L, Kim B J, et al: **Long noncoding RNA MALAT1 suppresses breast cancer metastasis.** *Nature Genetics* 2018, 50.
26. Iyer MK, Niknafs YS, Malik R, et al: **The landscape of long noncoding RNAs in the human transcriptome.***Nature Genetics* 2015, 47: 199-208.
27. Gutschner T, Diederichs S: **The hallmarks of cancer: a long non-coding RNA point of view.** *Rna Biology* 2012, 9(6): 703-719.
28. Ponting CP, Oliver PL, Reik W: **Evolution and functions of long noncoding RNAs,***Cell* 2009, 136(4): 629-641.
29. Yuchen S, Jian L, Liangzhao C, et al: **Long noncoding RNA SNHG12 facilitates the tumorigenesis of glioma through miR-101-3p/FOXP1 axis.***Gene* 2018, 676: S0378111918308965-.
30. Caley D, Pink R, Trujillano D, et al: **Long Noncoding RNAs, Chromatin, and Development.***The scientific world journal* 2014, 10(1): 90.
31. Hauptman N, Glavac D: **Long non-coding RNA in cancer.***International Journal of Molecular Sciences* 2013,14(3): 4655-4669.
32. Reon BJ, Anaya J, Zhang Y, et al: **Expression of lncRNAs in Low- Grade Gliomas and Glioblastoma Multiforme: An In Silico Analysis.***PLoS Medicine* 2016, 13: e1002192.
33. Chen G, Gu Y, Han P, et al: **Long noncoding RNA SBF2-AS1 promotes colorectal cancer proliferation and invasion by inhibiting miR-619-5p activity and facilitating HDAC3 expression.***J Cell Physiol* 2019, 234: 18688-96.
34. Youji, Yan, Zhongjun, et al: **Long non-coding RNA SNHG6 is upregulated in prostate cancer and predicts poor prognosis.***Molecular biology reports*
35. Liu W, Wang QP, Guo J: **Prognostic significance of long non-coding RNA DANCR expression in human cancers: A systematic review and meta-analysis.***Bioscience Reports*
36. Zeng S, Liu S, Feng J, Gao J, Xue F: **Upregulation of lncRNA AB073614 functions as a predictor of epithelial ovarian cancer prognosis and promotes tumor growth in vitro and in vivo.***Cancer Biomarkers: section A of Disease markers* 2019, 24: 421-28.
37. Reon BJ, Bruno T R K, Manjari K, et al: **LINC00152 Promotes Invasion through a 3'-Hairpin Structure and Associates with Prognosis in Glioblastoma.***Molecular Cancer Research* 2018, 16: 1470-82.
38. Cai J, Zhang J, Wu P, et al: **Blocking LINC00152 suppresses glioblastoma malignancy by impairing mesenchymal phenotype through the miR-612/AKT2/NF-κB pathway.***Journal of Neuro-Oncology* 2018, 140: 225-236.
39. Navarro G, Martínez-Pinilla E, Sánchez-Melgar A, et al: **A genomics approach identifies selective effects of trans-resveratrol in cerebral cortex neuron and glia gene expression.***PLoS One* 2017, 12(4):

e0176067.

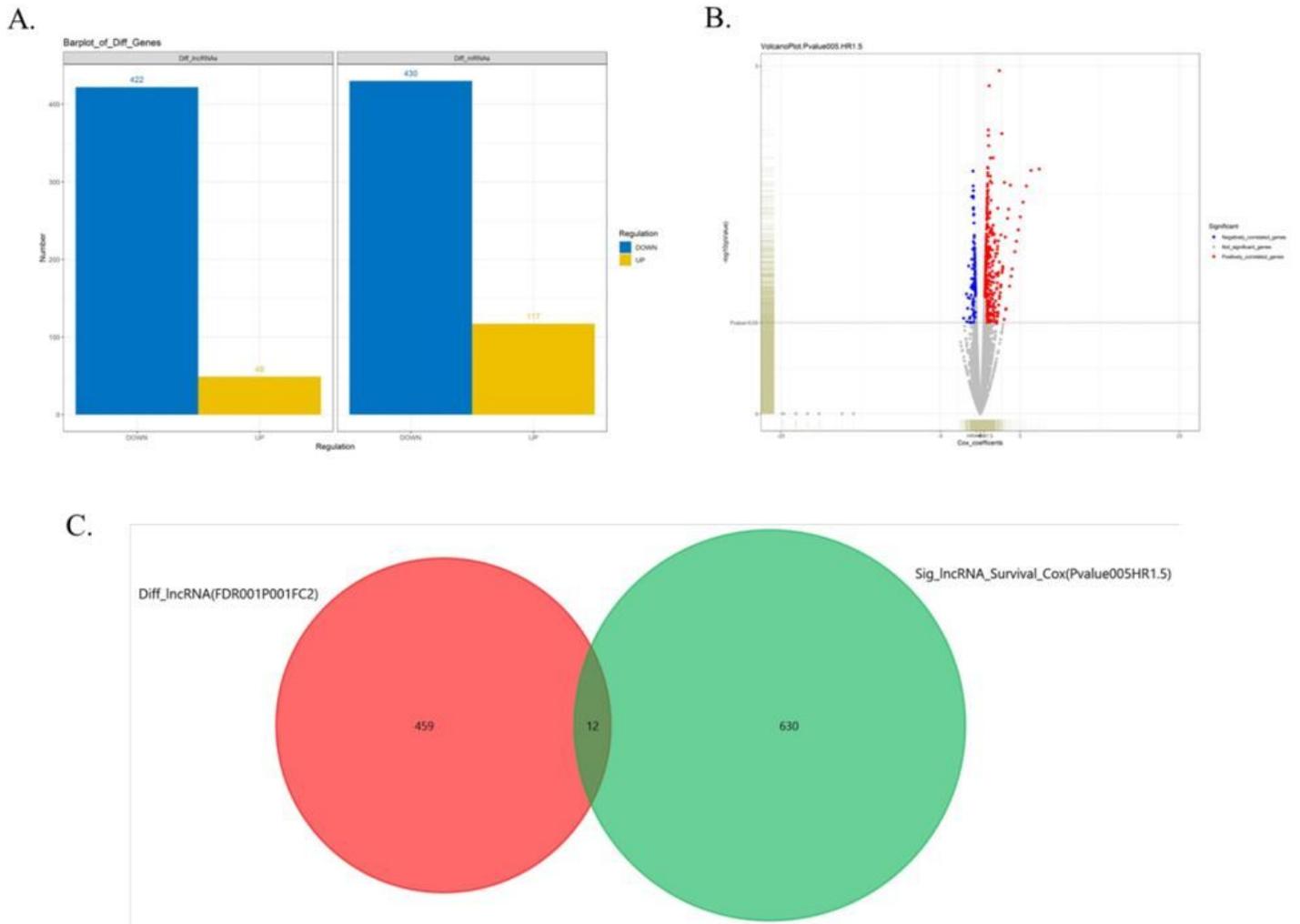
40. Quick Q, Skalli O: **Alpha-actinin 1 and alpha-actinin 4: contrasting roles in the survival, motility, and RhoA signaling of astrocytoma cells.** *Exp Cell Res* 2010, 316:1137-47.
41. Lappano R, Jacquot Y, Maggiolini M: **GPCR modulation in breast cancer.** *International Journal of Molecular Sciences* 2018, 19(12).
42. Li T, Yu B, Liu Z, et al: **Homocysteine directly interacts and activates the angiotensin II type I receptor to aggravate vascular injury.** *Nature Communications* 2018, 9(1): 11.
43. Fridlyand L E, Philipson L H: **Pancreatic Beta Cell G-Protein Coupled Receptors and Second Messenger Interactions: A Systems Biology Computational Analysis.** *Plos One* 2016, 11(5): e0152869.
44. Li D, Fei X, Dong Y, et al: **The long non-coding RNA CRNDE acts as a ceRNA and promotes glioma malignancy by preventing miR-136-5p-mediated downregulation of Bcl-2 and Wnt2.** *Oncotarget* 2017, 8: 88163-88178.
45. Zhang X, Sun S, Pu JK, et al: **Long non-coding RNA expression profiles predict clinical phenotypes in glioma.** *Neurobiology Disease* 2012, 48(1): 1-8.
46. Zheng J, Li X D, Wang P, et al: **CRNDE affects the malignant biological characteristics of human glioma stem cells by negatively regulating miR-186.** *Oncotarget* 2015, 6(28): 25339-25355.
47. Wang Y, Wang Y, Li J, et al: **CRNDE, a long non-coding RNA, promotes glioma cell growth and invasion through mTOR signaling.** *Cancer Letters* 2015, 367(2): 122-128.

## Figures



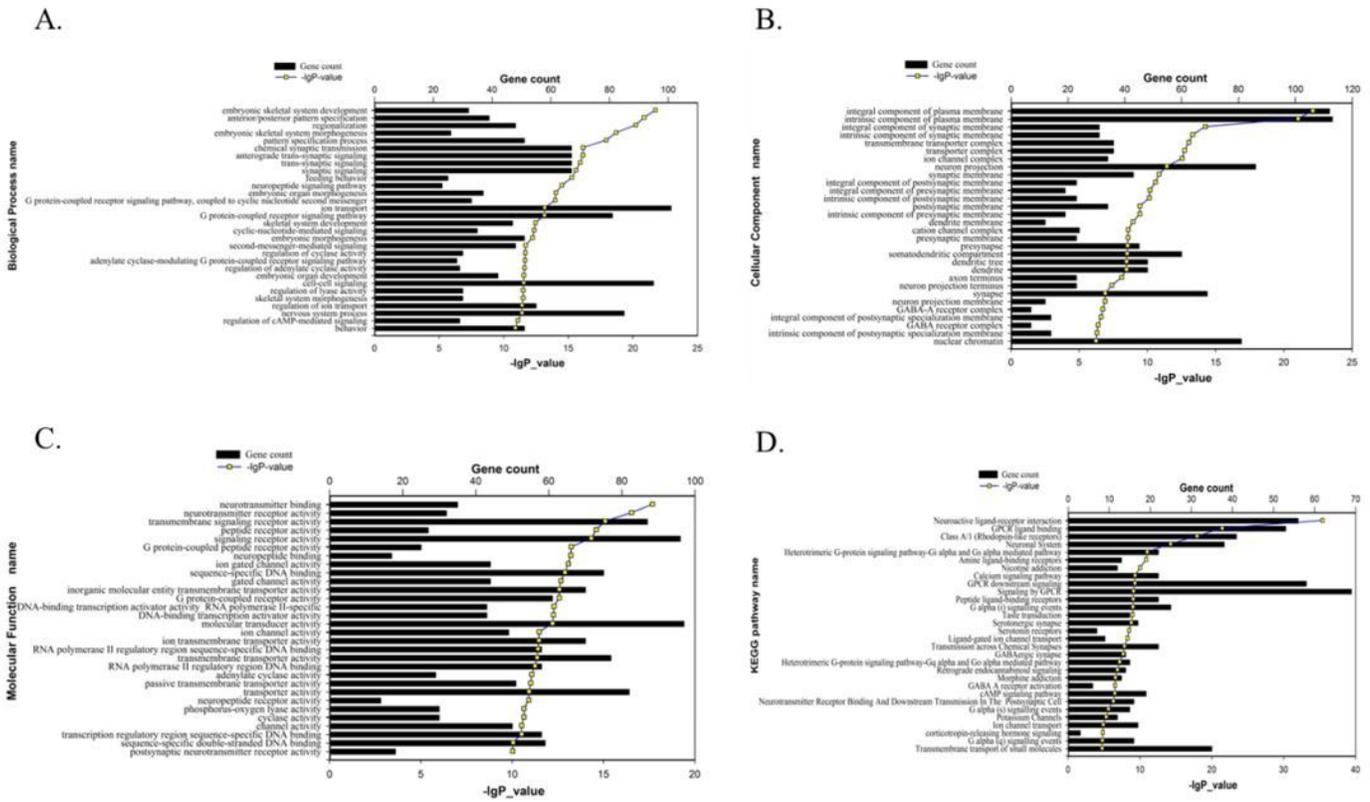
**Figure 1**

Difference analysis of GBM samples (A) Boxplot and violinplot of all mRNAs by tumor and normal group (B) The volcano plot is based on the expression values of all mRNAs detected by RNA SEQ (C) Boxplot and violinplot of all lncRNAs by tumor and normal group (D) The volcano plot is based on the expression values of all lncRNAs detected by RNA SEQ (E) Heatmap of the expression levels of the different mRNAs (F) Heatmap of the expression levels of the different lncRNAs



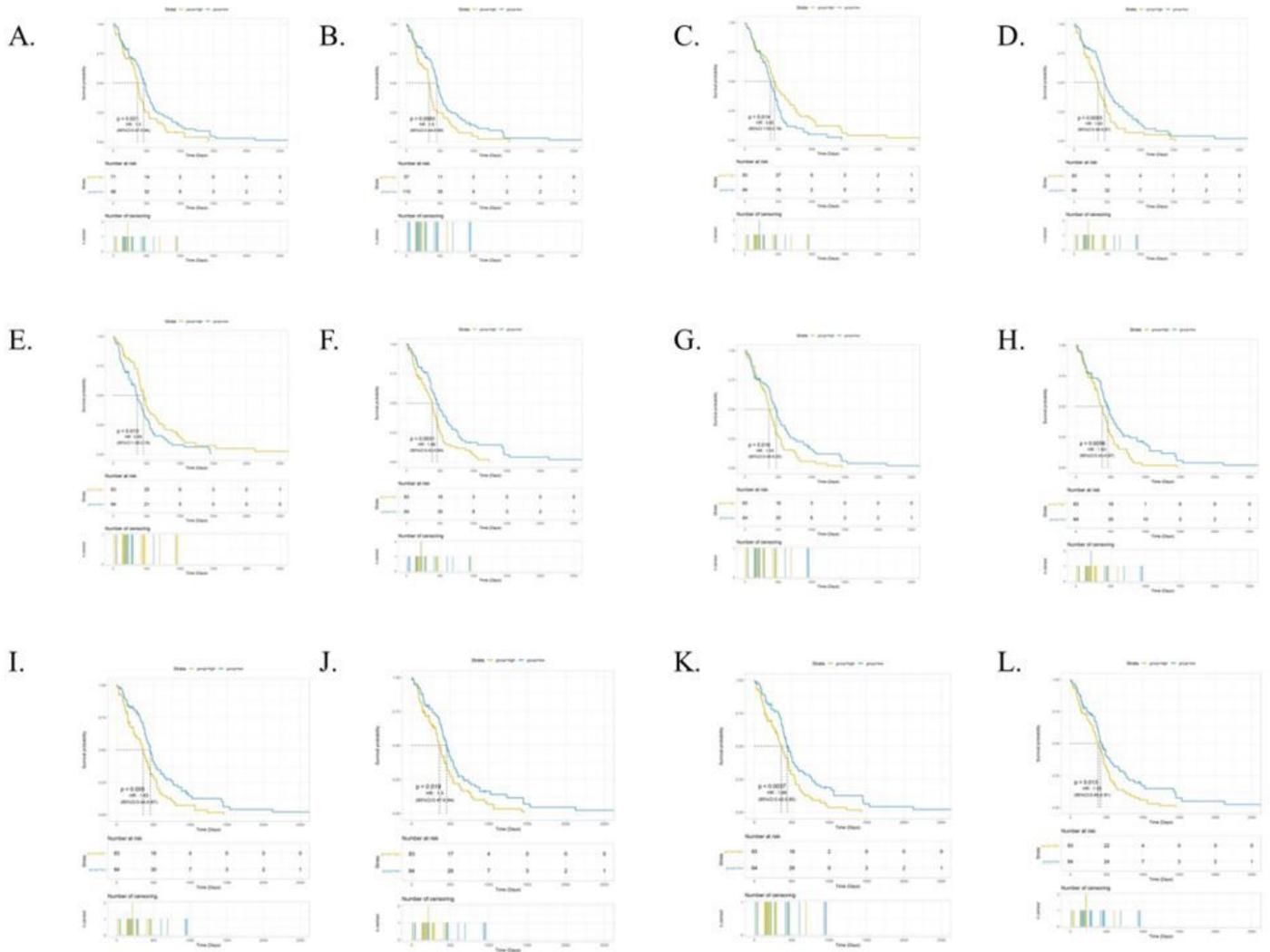
**Figure 2**

LncRNA-based prognostic biomarkers in GBM patients by univariate COX survival regression analysis (A) Barplot of the number of the different mRNAs and lncRNAs (B) The volcano plot of all significant survival lncRNAs (C) Venn plot



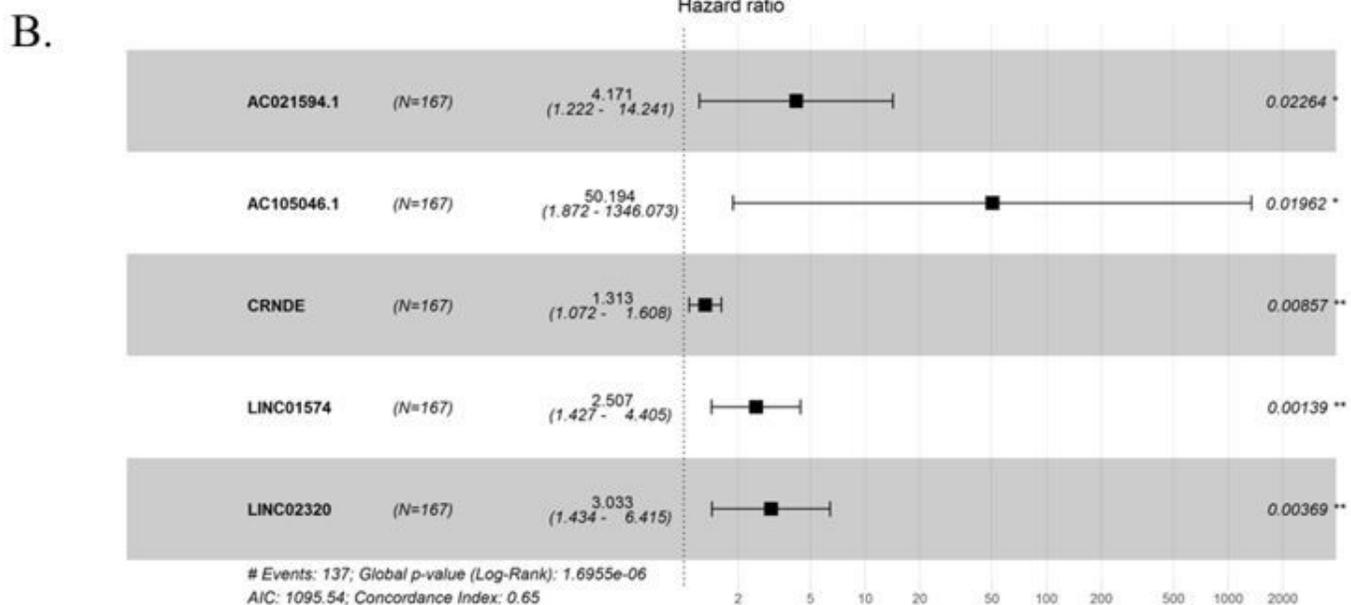
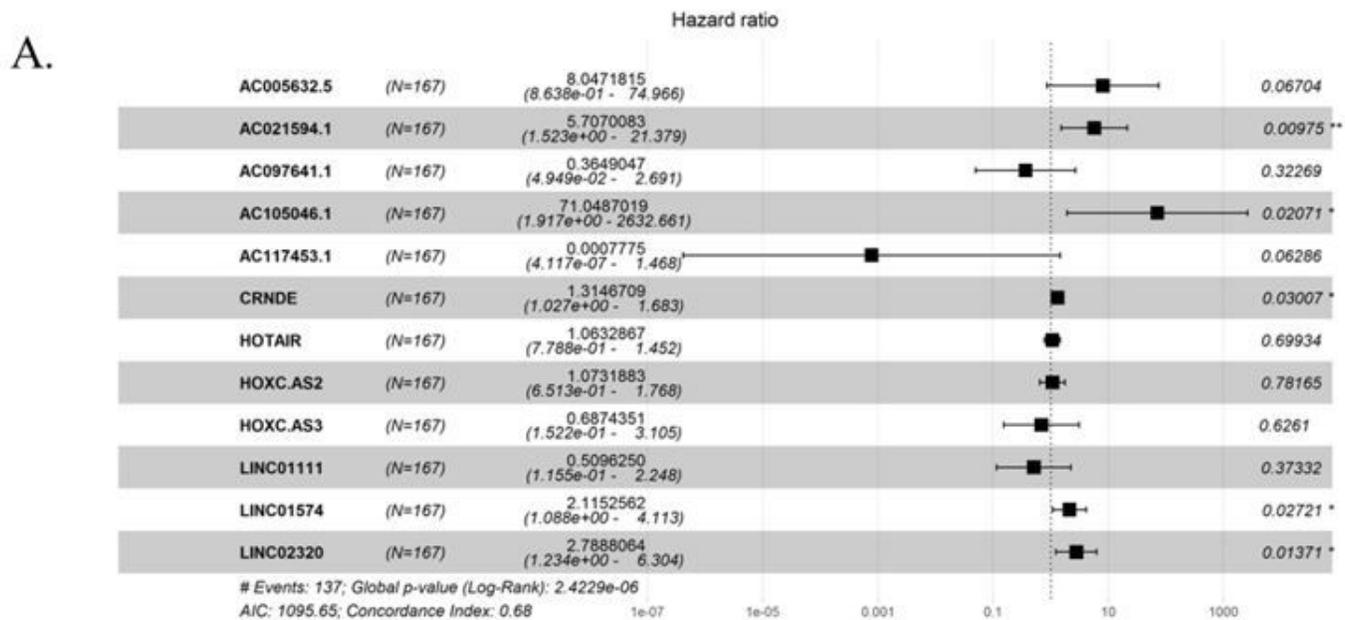
**Figure 3**

Enrichment analysis of GO and KEGG pathways of DEGs (A) Top30 barplot of GO BP enrichment annotation analysis (B) Top30 barplot of GO MF enrichment annotation analysis (C) Top30 barplot of GO CC enrichment annotation analysis (D) Top30 barplot of KEGG enrichment annotation analysis



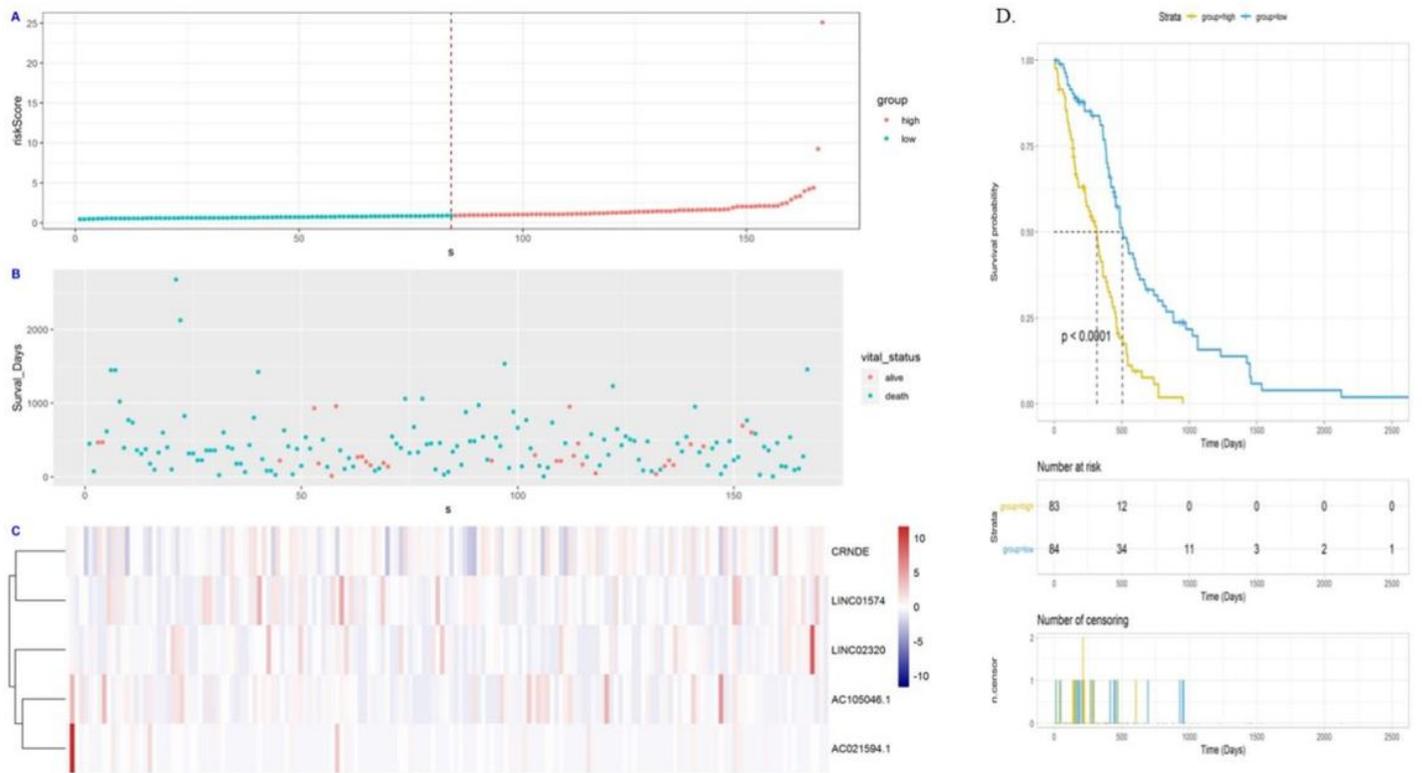
**Figure 4**

Kaplan-Meier curves of OS in all patients with GBM based on 12 lncRNAs expression (A) AC005632.5 (B) AC021594.1 (C) AC097641.1 (D) AC105046.1 (E) AC117453.1 (F) CRNDE (G) HOTAIR (H) HOxC-AS2 (I) HoxC-AS3 (J) LINC01111 (K) LINC01574 (L) LINC02320



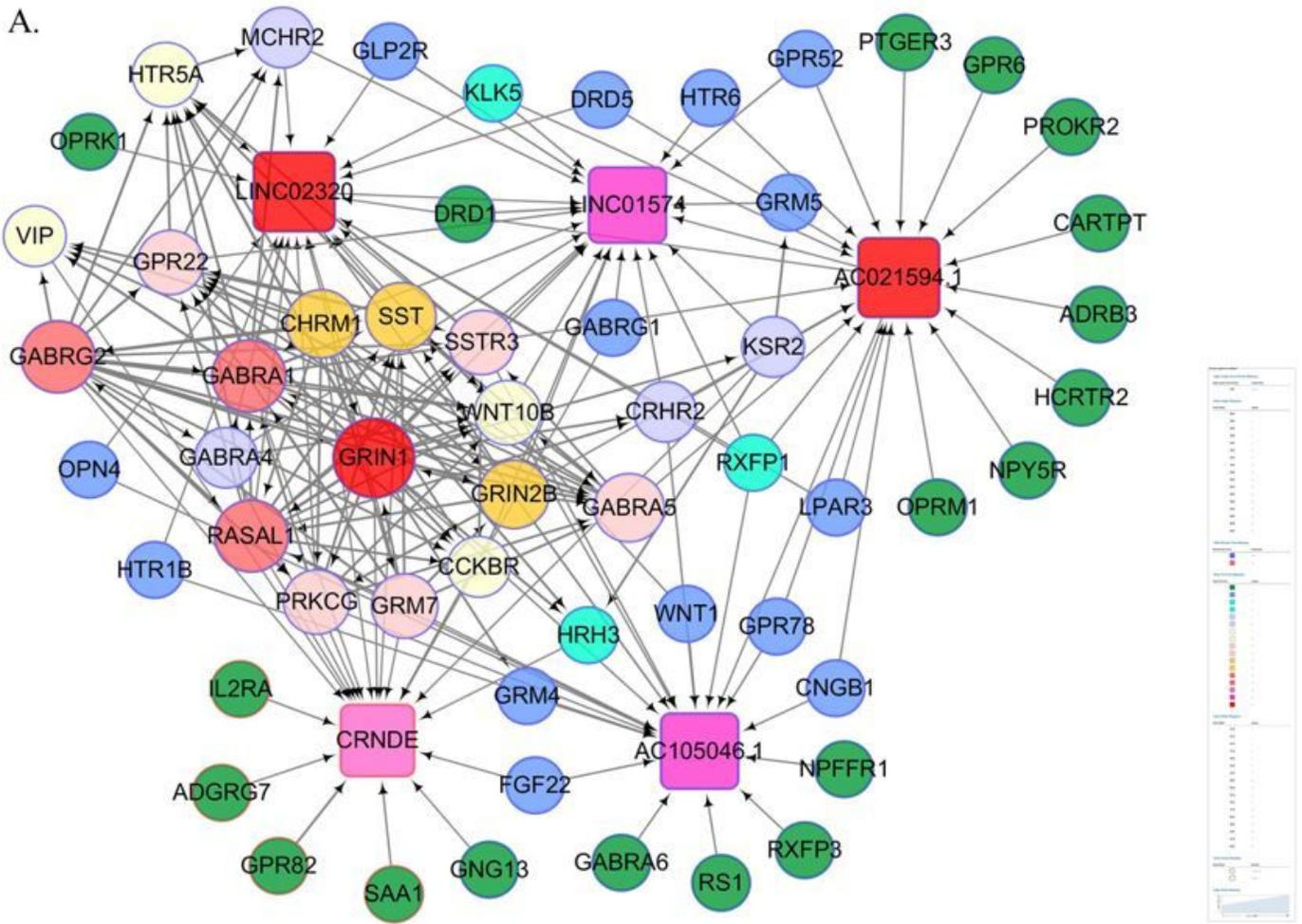
**Figure 5**

LncRNA with independent prognostic factor (A) Calculated by multivariate Cox regression model, hazard ratio with 95% confidence interval (95% CI) for independent 12 LncRNA signature were shown in forest plot (B) Calculated by stepwise regression method of multivariate Cox model, hazard ratio with 95% confidence interval (95% CI) for each independent 5 LncRNA signature were shown in forest plot



**Figure 6**

Multiple regression model risk Score survival analysis (A) the distribution of risk score (B) the plot of death status (C) heatmap of 5 lncRNAs expression (D) Kaplan-Meier curves of OS in GBM based on risk score



**Figure 7**

Coexpression network of coding gene/non-coding gene

A.

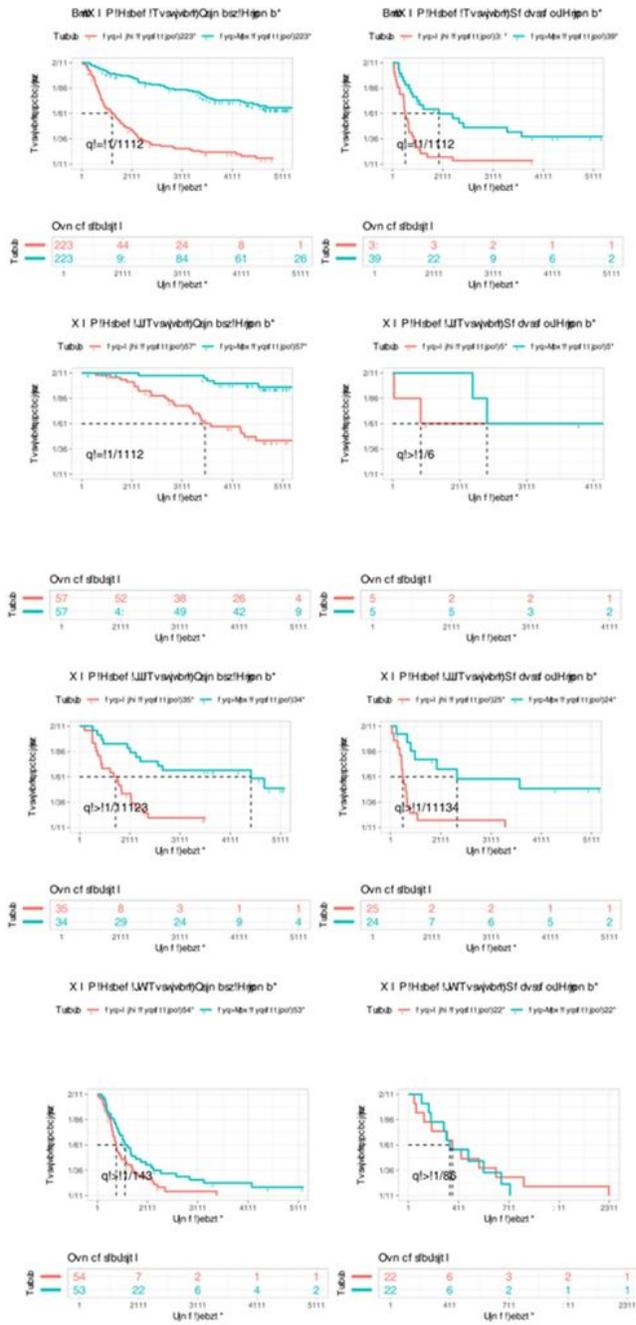


Figure 8

CRNDE in CGGA database