Development and validation of a prognostic nomogram for lower-grade glioma based on an autophagy-related lncRNA signature

Type

Research paper

Keywords

glioma, nomogram, long noncoding RNA, autophagy, computational biology

Abstract

Introduction
Gliomas account for 75% of the primary malignant brain tumors. The prognosis and treatment planning vary in lower-grade gliomas (LGG) due to their heterogeneous clinical behaviors. The dysregulation of autophagy-related (ATG) lncRNAs plays a crucial role in LGG. We aimed to develop and validate an ATG lncRNA risk signature, and a survival nomogram with integration of novel prognostic for LGG patients.

Material and methods
Differentially expressed ATG lncRNAs were screened out based on TCGA and GTEx RNA-seq databases. ATG lncRNA prognostic signature was then established by Kaplan–Meier, univariate Cox proportional hazards regression, Least absolute shrinkage and selection operator (LASSO) regression and multivariate Cox proportional hazards regression, with its predictive value validated by time-dependent receiver operating characteristic (ROC) curves. Kaplan–Meier, univariate Cox regression and multivariate Cox proportional hazards regression were used to screen out clinical and molecular variables. A nomogram was developed and internally validated by ROC and calibration plots.

Results
An ATG lncRNA risk signature was constructed with six differentially expressed lncRNAs (LINC00599, LINC02609, AC021739.2, AL118505.1, AL354892.2, and AL590666.2). Based on the risk signature, a nomogram was developed by addition of the significant prognostic clinical variables (age and grade) and molecular variables (IDH status and MGMT status).

Conclusions
We identified an ATG lncRNA risk signature and develop a nomogram for individualized survival prediction in LGG patients. A user-friendly free online calculator to facilitate the use of this nomogram among clinicians is also provided: https://linstu2009.shinyapps.io/LGGPRODICTORapp/?_ga=2.3154800.1506830296.1588641469-159983587.1588641469.

Development and validation of a prognostic nomogram for lower-grade glioma based on an autophagy-related lncRNA signature

Running title: lncRNA-based nomogram in lower-grade gliomas

Abstract

Introduction：Gliomas account for 75% of the primary malignant brain tumors. The prognosis and treatment planning vary in lower-grade gliomas (LGG) due to their heterogeneous clinical behaviors. The dysregulation of autophagy-related (ATG) lncRNAs plays a crucial role in LGG. We aimed to develop and validate an ATG lncRNA risk signature, and a survival nomogram with integration of novel prognostic for LGG patients.

Material and methods: Differentially expressed ATG lncRNAs were screened out based on TCGA and GTEx RNA-seq databases. ATG lncRNA prognostic signature was then established by Kaplan–Meier, univariate Cox proportional hazards regression, Least absolute shrinkage and selection operator (LASSO) regression and multivariate Cox proportional hazards regression, with its predictive value validated by time-dependent receiver operating characteristic (ROC) curves. Kaplan–Meier, univariate Cox regression and multivariate Cox proportional hazards regression were used to screen out clinical and molecular variables. A nomogram was developed and internally validated by ROC and calibration plots.

Results: An ATG lncRNA risk signature was constructed with six differentially expressed lncRNAs (LINC00599, LINC02609, AC021739.2, AL118505.1, AL354892.2, and AL590666.2). Based on the risk signature, a nomogram was developed by addition of the significant prognostic clinical variables (age and grade) and molecular variables (IDH status and MGMT status).

Conclusions: We identified an ATG lncRNA risk signature and develop a nomogram for individualized survival prediction in LGG patients. A user-friendly free online calculator to facilitate the use of this nomogram among clinicians is also provided:

https://linstu2009.shinyapps.io/LGGPRODICTORapp/?_ga=2.3154800.1506830296.1588641469-159983587.1588641469.

Keywords: *glioma; autophagy; long noncoding RNA; nomogram; computational biology*

1. Introduction

Gliomas account for 75% of primary malignant brain tumors in adults and are associated with high mortality (1, 2). Lower-grade gliomas (LGGs), including diffuse low-grade and intermediate-grade gliomas (World Health Organization (WHO) grades II and III), show a considerably high morbidity (3, 4). Despite a rarer incidence and overall better prognosis for LGG than grade IV tumors and glioblastoma (GBM), 70% of LGGs can develop into GBM and lead to the death of patients within 10 years (5). Molecular alterations, which can be identified objectively, are now believed to serve as more important prognostic factors than histologic grading (3). The current gold standard treatment of glioma includes surgical resection followed by radiotherapy and chemotherapy (6). However, due to the heterogeneity of their clinical behaviors, the standard care of LGG has been debated, thus presenting a therapeutic challenge to physicians (4, 7). Therefore, searching for novel biomarkers for survival prediction and individualized treatment planning to improve the outcomes of LGG patients is necessary and urgent.

Induced by diverse cellular stresses, macroautophagy (autophagy hereafter) is a self-digestive process involving the formation and turnover of autophagosomes, which engulf cellular proteins and organelles for delivery to lysosomes. Unlike apoptosis that represents canonical type I programmed cell death, autophagy is a "double-edged sword", as it can contribute to

stability, survival and evasion of stress, where it is often referred to as “protective autophagy", as well as being toxic by promoting type II cell death (8, 9). These processes are regulated by evolutionarily conserved autophagy-related (ATG) genes (10, 11). The deregulation of ATG genes results in abnormal autophagy and is associated with a variety of pathological conditions, including cancer (12), with accumulating evidence demonstrating that autophagy is involved in the activities of glioma (7, 13). Specifically, long noncoding RNAs (lncRNAs), which are transcripts longer than 200 nucleotides (nt) without protein-coding capacity (14, 15), have been reported to regulate autophagy activity by changing the transcript levels of ATG genes (16, 17). Acting individually or cooperatively as competitive platforms for both miRNAs and mRNAs, lncRNAs are crucial regulators of ATG genes in autophagy regulatory networks (16, 18). For instance, the lncRNA PTENP1 induced cellular autophagy and apoptosis by decoying several ATG-targeting miRNAs, thus repressing the tumorigenic properties of hepatocellular carcinoma (19). Gu et al. revealed that the lncRNA DICER1-AS1 promoted the proliferation, autophagy and invasion of osteosarcoma cells by targeting ATG5 (20). Additionally, the lncRNA MEG3 was revealed to promote cisplatin-induced apoptosis via the inhibition of autophagy in human glioma cells (13). Thus, exploring ATG lncRNAs will be important to provide new insights into prognostic biomarkers and therapeutic interventions for LGG.

A nomogram is a useful and accessible tool for predicting survival and planning individualized treatments by providing an individualized estimate of survival rather than a group prediction (21). Although several nomograms for LGG survival have previously been established, they lack the integration of transcriptome data or a comprehensive inclusion of

novel prognostic factors. Thus, in this study, we aimed to develop and validate a prognostic nomogram for individualized survival prediction for LGG patients by integrating an ATG lncRNA risk signature with novel clinical and molecular prognostic factors (age, grade, isocitrate dehydrogenase (IDH) status, and O6-methylguanine-DNA methyltransferase (MGMT) status). In addition, a user-friendly online application was developed.

2. Materials and methods

2.1 Clinical data collection and processing

The study design was shown as a flow chart (Figure. 1). Two public databases, The Cancer Genome Atlas (TCGA) and Genotype-Tissue Expression (GTEx) Project, served as the primary sources of this study. TCGA (dataset ID: TCGA-LGG.htseq_fpkm) provided LGG lncRNA expression profiles as well as corresponding clinical information and molecular parameters, while GTEx (dataset ID: gtex_RSEM_gene_fpkm) offered lncRNA expression profiles of normal brain tissues. Specifically, the RNA sequencing (RNA-seq) expression profiles were downloaded from UCSC Xena (2020.3, https://xena.ucsc.edu/) (22). Clinical information and molecular parameters were downloaded from GlioVis (2020.3, http://gliovis.bioinfo.cnio.es) (23), including the following variables for each patient: tumor grade (grade II or grade III), age at diagnosis (>40 years old or < 40 years old), sex (male or female), IDH mutation status (IDH-mutant or IDH-wild type), MGMT status (methylated or unmethylated), survival/follow-up time in months (continuous) and survival status (alive or dead). By using data from Ensembl (https://uswest.ensembl.org/index.html) (24), we reannotated the gene symbols and extracted lncRNAs (including sense_overlapping, lincRNA,

3prime_overlapping_ncrna, processed_transcript, and antisense, sense_intronic (25)) from the original dataset. The lncRNA expression profiles from 529 normal brain samples and 529 LGG samples are presented by log2 (fpkm+1). Then, the two datasets were merged into one with normalization by using the "limma" package, version 3.42.0 (http://www.bioconductor.org/) (26) in R language, version 3.6.2 (https://cran.r-project.org/). LncRNAs with an expression value of 0 were removed. Eventually, 14086 lncRNAs among 1058 samples were found.

2.2 Identification of DElncRNAs

Principal component analysis (PCA) by the "stats" package, version 3.6.2 (https://cran.r-project.org/), was used to determine the visualized genetic distance and relatedness between normal brain tissue and LGG. Differentially expressed lncRNAs (DElncRNAs) between LGG and normal brain tissue were generated using the "limma" package. LncRNAs were considered to have statistically significant differences in expression if $|\log2(\text{fold-change})| \geq 2$ and false discovery rate (FDR) < 0.05. We performed hierarchical clustering based on the most variably expressed genes using Euclidean distance as the similarity metric and the complete linkage method as the between-cluster distance metric.

2.3 Identification of ATG DElncRNAs

ATG genes were extracted from the Human Autophagy Database (HADb, http://www.autophagy.lu/index.html) (27). All of the mRNA expression data were normalized by log2 transformation. Pearson correlation was applied to calculate the

correlation between the DElncRNAs and ATG genes. A DElncRNA with a correlation coefficient |r| > 0.4 and p value < 0.05 was considered to be an ATG DElncRNA.

2.4 Identification of a prognostic ATG lncRNA signature

Twenty-five LGG samples, in which survival, clinical or molecular subtype information was missing, were excluded. Using the "caret" package, version 6.0-85 (https://cran.r-project.org/) (28) in R language, the ATG DElncRNA profiles were then randomly divided into a training cohort (n = 252) and a validation cohort (n= 252). A survival analysis model was constructed based on the training cohort by the "survival" package, version 3.1-8 (https://cran.r-project.org/) in R language, while the validation cohort was used for model testing. The median expression level of the training cohort was used to split the ATG lncRNAs into high- and low-expression groups, followed by Kaplan–Meier (K-M) survival analysis to assess the survival differences between them. Univariate Cox proportional hazards regression models were used to assess the association between the ATG lncRNAs and the overall survival (OS) of LGG patients from the training cohort. P value < 0.05 was considered statistically significant. Least absolute shrinkage and selection operator (LASSO) regression, which avoids overfitting of the model in the risk signature according to the best lambda value, was performed to filter out the ATG lncRNAs that were significant in univariate Cox analysis. Subsequently, the joint effect of different covariates was assessed using multivariate Cox proportional hazards regression by the "step" function in R programming language, with results shown as forest plots. The relationships among the six

ATG lncRNAs and their coexpressed ATG genes were displayed by creating a Sankey diagram.

The prognostic prediction model was constructed based on the regression coefficient-weighted lncRNA expression, and a risk score formula was established as follows:

Risk score= $\sum_{i=1}^{n} \text{Expi} \times \text{Coei}$

In the formula, N is the number of selected ATG lncRNAs, with Expi being the expression value of each ATG lncRNA and Coei being the multivariate Cox regression coefficient. Next, the expression profile data of the corresponding ATG lncRNAs were extracted from the training cohort and substituted into the model to calculate the risk score of each patient. The patients were divided into high- and low-risk groups according to the median risk score value. K-M survival analysis was used to estimate the survival distributions. Receiver operating characteristic (ROC) analysis ("survivalROC" package, version 1.0.3, https://cran.r-project.org/) of the ATG lncRNA risk signature for predicting 1-, 3- and 5-year survival was carried out. Replication was carried out to internally validate the model by using the data from the validation cohort.

2.5 Functional enrichment analysis

Gene set enrichment analysis (GSEA, https://www.broad institute.org/gsea/index.jsp) was applied to identify the biological functions and pathways between the high- and low-risk groups based on the risk signature. Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways associated with the risk signature were further explored.

According to the GSEA website, an FDR of 0.25 is reasonable in the setting of exploratory discovery for the validation of the candidate hypothesis in future research, while a more stringent FDR may lead to overlooking potentially significant results. Thus, gene sets with an FDR < 0.25 in the high- and low-risk groups in the TCGA data set (TCGA-LGG.htseq_fpkm) were considered significantly different and were selected.

2.6 Development of a prognostic nomogram

In the training cohort, K-M survival analysis was first used to estimate the survival distributions of each clinical and molecular factor, followed by assessing the association between the OS and each factor using univariate Cox proportional hazards regression models. Next, the joint effect of different covariates was assessed using multivariate Cox proportional hazards regression by the "step" function in the R programming language. The results of both univariate Cox and multivariate Cox analyses are shown as forest plots. Receiver operating characteristic (ROC) analysis ("survivalROC" package) of the ATG lncRNA risk signature, as well as each clinical and molecular factor mentioned above for predicting the 1-, 3- and 5-year survival, was performed to evaluate the sensitivity and specificity of survival prediction. Nonsignificant variables (p value>0.05) were omitted. A nomogram was constructed based on the significant factors for predicting the survival of LGG patients. The survival ROC curve and calibration curve were used to assess the performance of the nomogram. The nomogram was internally validated using the validation cohort. To facilitate clinical use, a free online calculator for the final nomogram was established by the "DEnorm"

package (version 5.0.1, https://cran.r-project.org/) and published in "https://www.shinyapps.io/".

2.7 Statistical analysis

The R programming language (version 3.6.2) was used to perform statistical analyses, including PCA, K-M survival analyses, univariate and multivariate Cox regression models, LASSO regression and ROC curve analysis, as well as to draw figures, including heatmaps, boxplots, forest plots and calibration plots. Quantitative data are shown as the mean ± standard deviation (SD). Statistical differences between two groups were compared by the Wilcoxon test. A P value< 0.05 was considered statistically significant.

3. Results

3.1 Construction of a risk signature including six ATG DElncRNAs

To graphically determine the distribution of all 14086 lncRNAs within normal brain tissue and LGG, PCA was employed to show that the data had been normalized well and that the variation in the data were maximal (Figure 2a). Subsequently, 112 DElncRNAs with an expression ratio that differed between LGG and normal brain tissues by a factor of at least 2-fold were selected (Figure 2b and 2c). Each lncRNA with their median expression levels in LGG and normal brain tissues is shown in Figure 2d. A total of 232 ATG genes were downloaded from HADb, and 20 ATG DElncRNAs with Pearson correlation coefficient |r| >0.4 and p value< 0.05 were selected (Figure 2e). LGG patients in the TCGA dataset with detailed clinical information (age, grade, and gender) and molecular parameters (IDH status

and MGMT status) were randomly divided into a training cohort (n = 252) and a validation cohort (n = 252) (Table 1). Twelve ATG DElncRNAs of prognostic value were screened out by performing Kaplan-Meier analysis and univariate Cox analysis in the training cohort (Table S1). To select appropriate parameters for constructing a risk signature, LASSO regression was used and identified 10 ATG DElncRNAs (AC021739.2, AC093010.3, AL118505.1, AL121827.2, AL354892.2, AL355916.2, AL590666.2, LINC00599, LINC02609, and NEAT1) (Figures 3a and 3b). Eventually, only 6 ATG DElncRNAs (AC021739.2, AL118505.1, AL354892.2, AL590666.2, LINC00599, and LINC02609) remained following multivariate Cox regression analysis. AC021739.2, AL118505.1, AL354892.2, LINC00599, and LINC02609 were regarded as protective factors (hazard ratios (HRs) < 1), while AL590666.2 was the only risk factor (HR > 1) among these lncRNAs in LGG (Figure 3c). According to the HRs, a Sankey diagram was constructed to intuitively display the regulation of the six ATG lncRNAs on their coexpression genes, with five lncRNAs as protective factors and one as a risk lncRNA (Figure S1). Based on both univariate and multivariate Cox regression analyses, the six lncRNAs as novel prognostic biomarkers were suggested for further analysis.

3.2 Construction of the prognostic risk signature with six ATG lncRNAs in LGG

The six ATG lncRNAs (AC021739.2, AL118505.1, AL354892.2, AL590666.2, LINC00599, and LINC02609) were incorporated to develop a risk signature in the training cohort. The risk scores were produced using the formula mentioned in the methods as follows: risk

score= (-0.3899 × expression level of AC021739.2) + (-0.5685 × expression level of AL118505.1) + (-0.7595 × expression level of AL354892.2) + (-0.2849 × expression level of LINC00599) + (-0.3508 × expression level of LINC02609) + (0.5524 × expression level of AL590666.2). The samples in the training cohort were divided into high- and low-risk groups according to the median risk score. As shown in Figure 4a, higher risk scores suggested more deaths. In addition, along with the increasing risk scores, the expression levels of AC021739.2, AL118505.1, AL354892.2, LINC00599, and LINC02609 were decreased, whereas the expression level of AL590666.2 was increased. The K-M curve showed that the high-risk group was associated with poorer prognosis (Figure 4b). The ROC curve was used to evaluate the efficacy of the ATG lncRNA risk signature to predict 1-, 3-, and 5-year survival in LGG patients. The areas under the curve (AUCs) for 1-, 3-, and 5-year survival were 0.788, 0.857, and 0.687, respectively (Figure 4c-4e), indicating that the risk signature had good predictive performance.

3.3 Validation of the prognostic ATG lncRNA risk signature

The performance of the ATG lncRNA risk signature was internally tested in the validation cohort. The LGG samples in the validation cohort were divided into high- and low-risk groups according to the median risk score. In line with the results in the training cohort, downregulated expression levels of AC021739.2, AL118505.1, AL354892.2, LINC00599 and LINC02609, as well as upregulated expression of AL590666.2 and more deaths, were observed with higher risk scores (Figure 5a). Similarly, the K-M curve showed that patients in the high-risk group had a relatively unfavorable prognosis (Figure 5b). The AUCs for 1-,

3-, and 5-year survival were 0.906, 0.78, and 0.725, respectively (Figure 5c-5e), thereby confirming the good predictive efficacy of the risk signature.

3.4 Functional annotation and signaling pathway enrichment of the ATG lncRNA prognostic signature

GSEA was conducted to explore the biological functions and pathways associated with the ATG lncRNA risk signature in LGG patients. As a result, a total of 3373 GO functions were enriched in the high-risk group (FDR< 0.25, top 100 shown in Table S2), including autophagosomes, cell matrix adhesion and regulation of cell junction assembly functions (FDR< 0.05) (Figure S2a- S2c), while 181 GO functions were enriched in the low-risk group (FDR< 0.25) (Table S3), including ribosome assembly functions (FDR< 0.05) (Figure S2d). We also obtained 113 enriched KEGG pathways in the high-risk group (FDR< 0.25) (Table S4), including the regulation of autophagy, MAPK signaling pathway, and extracellular matrix (ECM) receptor interaction pathway (FDR< 0.05) (Figure S2e- S2g). For the low-risk group, 6 enriched KEGG pathways are shown in Table S5, including the ribosome pathway (FDR< 0.25) (Figure S2h).

3.5 Development and independent validation of a nomogram integrating the risk signature with clinical and molecular variables

The prognostic significance of clinical factors such as age, gender, and grade, as well as molecular parameters such as IDH status and MGMT status, were previously reported in glioma (3, 29-33). In the training cohort, the association of overall survival with these

variables was assessed using K-M survival analysis (Figure 6a-6e) and Cox proportional hazards regression in univariate and multivariable models (Figure 6f and 6g). The gender variation was eliminated from the models since the univariate analysis result showed no statistical significance between groups. ROC curves were also used to evaluate the prognostic accuracy of the risk signature and each variable. Due to space limitations, as well as the most commonly used indexes in clinical practice, we only show the predicted 1-, 3-, and 5-year survival rates for LGG patients. As shown in the ROC curves, the AUCs of the ATG lncRNA risk signature for predicting 1-, 3- and 5-year survival were 0.824 (Figure 6i), 0.901 (Figure 6j) and 0.700 (Figure 6k), respectively, which were higher than those of any clinical or molecular variables, except for that of IDH status in 1-year survival (0.867). Conclusively, the ATG lncRNA risk signature provided a more accurate survival prediction than other prognostic factors (age, grade, IDH status, and MGMT status), though they were contributing factors of survival and had good prognostic accuracy. Thus, a nomogram was developed by integrating the ATG lncRNA risk signature with novel prognostic clinical and molecular factors (age, grade, IDH status, and MGMT status). As shown in the nomogram, the probabilities of 1-, 3-, and 5-year survival could be quickly estimated as the total points by adding the points in each item (Figure 6h). ROC curves and calibration plots were used to evaluate the performance of the nomogram. The AUCs of the ROC curves for predicting 1-, 3- and 5-year survival were 0.877, 0.937 and 0.826, respectively, in the training cohort (Figure 7a) and 0.905, 0.914 and 0.732 in the validation cohort (Figure 7e). The calibration curves showed good agreement between the predictions and observations in the training cohort (Figure 7b-7d) and the validation cohort (Figure 7f-7h) for the probabilities of 1-, 3-

and 5-year survival. Considering that predictions for other time points are also important, free online software, established by the "DEnorm" package, for the developed nomogram was made available for easier clinical use: https://linstu2009.shinyapps.io/LGGPRODICTORapp/?_ga=2.3154800.1506830296.1588641469-159983587.1588641469.

4. Discussion

A recently published study had confirmed the involvement of aberrant expression of lncRNAs in glioma development by examining the lncRNA profiles from tumor and peritumoral tissues, without exploring the potential contribution of lncRNA expression to patients' survival (34). In our study, we identified six ATG lncRNAs (LINC00599, LINC02609, AC021739.2, AL118505.1, AL354892.2 and AL590666.2) that were of predictive value in LGG survival. Long intergenic nonprotein coding RNA 599 (LINC00599), also known as retinal noncoding RNA3, is located on human chromosome 8p23 and was first reported to be dynamically expressed during mouse retinal development (35). It is considered to regulate the differentiation of neurons and oligodendrocytes (36) and is now increasingly recognized as a critical player in a variety of cancers and considered an oncogene. For example, LINC00599 promotes the progression of both prostatic cancer and colorectal cancer (37). However, the functional roles of LINC00599 in glioma are controversial. Increased expression of LINC00599 was observed in glioma tissues and cell lines, where its silencing suppressed proliferation and invasion and induced cell cycle arrest involving the Akt/GSK3β pathway (38, 39). In contrast, it has been reported that LINC00599 is downregulated in GBM

cells, and its overexpression inhibits proliferation and induces apoptosis through the miR-185- 5p/KLF16 axis (40). In addition, Fu and colleagues revealed that the expression of LINC00599 was reduced in both LGG and GBM tissues and that it served as a tumor-suppressing lncRNA by inhibiting cell migration and invasion through the regulation of the EMT process (41). This finding is in agreement with the concept that LINC00599 is downregulated in LGG and serves as a protective factor for LGG survival. Despite accumulating studies on the molecular mechanisms of lncRNAs, knowledge of the remaining 5 lncRNAs (LINC02609, AC021739.2, AL118505.1, AL354892.2, and AL590666.2) is limited so far. In our study, we show that LINC02609, AC021739.2, AL118505.1 and AL354892.2 all have an HR < 1, indicating that they are positive predictors of LGG. Conversely, higher expression of AL590666.2 suggests an unfavorable prognosis in LGG. Therefore, these lncRNAs can be further explored for their potential roles in regulating autophagy and as prognostic markers and therapeutic targets of LGG. In addition, GSEA suggested that the ATG lncRNA risk signature was mainly related to autophagy and cell matrix adhesion, which was reported to be involved in the malignant transformation and local invasiveness of glioma cells (42), indicating the essential roles of our signature.

By integrating a set of novel prognostic factors, nomograms are useful and accessible tools for predicting survival and individualized treatment planning since they provide an individualized estimate of survival rather than a group prediction (21). Several nomograms for LGG patient survival have been established previously (4, 43, 44). Among these, the nomogram developed by Gittleman et al. represented the most comprehensive nomogram, including common essential prognostic variables such as sex, tumor grade, and age, as well

as some critical newly discovered factors including molecular subtype (IDH mutation, 1p/19q codeletion) and Karnofsky performance status (KPS) (4). Our nomogram included all the variables in Gittleman's nomogram, except for postoperative KPS, 1p/19q codeletion and sex. In fact, Gittleman did not identify a sex difference in LGG survival by analyzing data from both TCGA and the Ohio Brain Tumor Study (OBTS) but still kept it in the nomogram considering its clinical significance (4). In line with their results, sex was not statistically significant in our study. According to the published literature, a sex difference was more commonly observed in GBM (30, 45), whereas LGG incidence was nearly identical in males and females. Therefore, gender was not included in our nomogram. We did not include the KPS value due to the high amount of missing data in TCGA (up to 55.0%). Since 1p/19q codeletion is the most common genetic characteristic of only a specific type of glioma (oligodendroglioma) and is of predictive value in response to chemotherapy and radiation (3, 32, 46), we included IDH mutation status, which is regarded as a hallmark of LGG and characterizes the majority of LGG patients, instead of 1p/19q codeletion (32, 47). In addition, we added another important prognostic variable, MGMT status, which is a DNA repair protein correlated with prolonged survival in patients with diffuse gliomas (3, 32, 33) and was found to be statistically significant in our survival model. Most importantly, we deeply explored the TCGA transcriptome data of LGG tissue rather than simply associating clinical and molecular prognostic factors to establish a lncRNA prognostic signature, which seemed to have the best accuracy among all the prognostic factors according to the AUCs of the ROC curves and was independently validated. A nomogram was developed based on this ATG lncRNA risk signature, with integrations of novel clinical (age and grade) and molecular

(IDH status and MGMT status) prognostic factors, followed by independent validation demonstrating an accurate and stable performance by ROC curves and calibration plots. This is the first nomogram that comprehensively integrates an ATG lncRNA risk signature with novel clinical and molecular prognostic factors.

An important advantage of this study is the application of transcriptome data from the GTEx Project, which allows access to a much larger data set of normal brain tissue while minimizing measurement bias compared to other studies extracting data from several Gene Expression Omnibus (GEO) datasets. Furthermore, some classical statistical methods were applied to make the survival prediction convincing. First, to avoid overfitting of the model, LASSO regression was used to identify lncRNAs since it allows the model coefficients to become 0. This property is consistent with our expectation of identified biomarkers in clinical practice in that it is clinically efficient and economical to detect the least number of key biomarkers for diagnosis or prognosis prediction. Second, K-M survival analysis, univariate Cox proportional hazards regression models, multivariate Cox proportional hazards regression, and ROC curves were sequentially used to strictly screen out all potential prognostic factors, with ROC curves and calibration applied for the validation of both the risk signature and nomogram. Third, we internally validated the established risk signature and nomogram in the validation cohort, thereby testing their accuracy in predicting survival. Finally, user-friendly free online software was designed to facilitate the use of the nomogram by clinicians.

However, the present study also has certain limitations. First, our nomogram did not include information regarding treatment, such as the extent of surgical resection, chemotherapy and

radiotherapy. This is due to the lack of information on the extent of surgical resection in TCGA and the debated standard of care for LGG patients. Second, internal validation was used to evaluate the efficiency of the risk signature and nomogram rather than external validation because of the lack of integrated LGG lncRNA data from other databases, such as GEO. Finally, the validation cohort was based on 252 retrospective datasets from the TCGA database with the application of internal validation instead of external validation to test the accuracy of the risk signature and the nomogram. This is because there is a lack of or too little LGG lncRNA data from other databases, such as GEO, Chinese Glioma Genome Atlas (CGGA), and Repository of Molecular Brain Neoplasia Data (REMBRANDT). However, although we did not include other databases in our research as an external validation, the randomly grouped datasets from the TCGA database used for internal validation were from different institutions. Therefore, the signature is in fact independently validated and is still convincing. Considering all these above limitations, we may replicate our findings in larger cohorts, hopefully with the LGG clinical data bank built, or validate the five-lncRNA signature in future studies when integral datasets are available.

5. Conclusions

In conclusion, we complemented available genomic-based studies by identifying six ATG lncRNAs (LINC00599, LINC02609, AC021739.2, AL118505.1, AL354892.2, and AL590666.2) that may serve as potential prognostic biomarkers or therapeutic targets of LGG, followed by establishing of an ATG lncRNA risk signature. With integration of ATG lncRNA risk signature, novel clinical and molecular prognostic factors (age, grade, IDH

mutation status, and MGMT status), we developed and internally validated a nomogram, thereby providing healthcare practitioners with individualized survival estimates and facilitating treatment planning in LGG patients. To promote the clinical use of this model, a free online software for its implementation is provided as follows: https://linstu2009.shinyapps.io/LGGPRODICTORapp/?_ga=2.3154800.1506830296.1588641469-159983587.1588641469.

Ethics approval and consent to participate

Not applicable.

Patient consent for publication

Not applicable.

Availability of data and material

Data can be downloaded from UCSC Xena (https://xena.ucsc.edu/) and GlioVis (http://gliovis.bioinfo.cnio.es)

Statements and Declarations

The authors declare that they have no competing interests.

Funding

The present study was supported by The National Natural Science Foundation of China (grant nos. 81471279 and 81171138) and The Research Start-Up Fund of Wuxi School of Medicine, Jiangnan University (grant no. 1286010242190060).

Acknowledgements

The authors thank UCSC Xena and GlioVis for the availability of the data.

References

1. Lapointe S, Perry A, Butowski NA. Primary brain tumours in adults. Lancet (London, England). 2018;392(10145):432-46.
2. Wang H, Liu G, Li T, Wang N, Wu J, Zhi H. MiR-330-3p functions as a tumor suppressor that regulates glioma cell proliferation and migration by targeting CELF1. Archives of medical science : AMS. 2020;16(5):1166-75.
3. Brat DJ, Verhaak RG, Aldape KD, Yung WK, Salama SR, Cooper LA, et al. Comprehensive, Integrative Genomic Analysis of Diffuse Lower-Grade Gliomas. The New England journal of medicine. 2015;372(26):2481-98.
4. Gittleman H, Sloan AE, Barnholtz-Sloan JS. An independently validated survival nomogram for lower-grade glioma. Neuro-oncology. 2020;22(5):665-74.
5. Kiran M, Chatrath A, Tang X, Keenan DM, Dutta A. A Prognostic Signature for Lower Grade Gliomas Based on Expression of Long Non-Coding RNAs. Molecular Neurobiology. 2018;56(7):4786-98.
6. Peng Z, Liu C, Wu M. New insights into long noncoding RNAs and their roles in glioma. Molecular Cancer. 2018;17(1).
7. Ulasov I, Fares J, Timashev P, Lesniak MS. Editing Cytoprotective Autophagy in Glioma: An Unfulfilled Potential for Therapy. Trends in Molecular Medicine. 2020;26(3):252-62.
8. Emdad L, Bhoopathi P, Talukdar S, Pradhan AK, Sarkar D, Wang XY, et al. Recent insights into apoptosis and toxic autophagy: The roles of MDA-7/IL-24, a multidimensional anti-cancer therapeutic. Seminars in cancer biology. 2020;66:140-54.
9. Lin JZ, Lin N. A risk signature of three autophagy-related genes for predicting lower grade glioma survival is associated with tumor immune microenvironment. Genomics. 2021;113(1 Pt 2):767-77.
10. Mizushima N. Autophagy: process and function. Genes & development. 2007;21(22):2861-73.
11. Mizushima N, Yoshimori T, Ohsumi Y. The role of Atg proteins in autophagosome formation. Annual review of cell and developmental biology. 2011;27:107-32.
12. Mizushima N, Levine B, Cuervo AM, Klionsky DJ. Autophagy fights disease through cellular self-digestion. Nature. 2008;451(7182):1069-75.
13. Ma B, Gao Z, Lou J, Zhang H, Yuan Z, Wu Q, et al. Long non-coding RNA MEG3 contributes to cisplatin-induced apoptosis via inhibition of autophagy in human glioma cells. Mol Med Rep. 2017;16(3):2946-52.
14. Ulitsky I, Bartel DP. lincRNAs: genomics, evolution, and mechanisms. Cell. 2013;154(1):26-46.
15. Lin JZ, Lin N, Zhao WJ. Identification and validation of a six-lncRNA prognostic signature with its ceRNA networks and candidate drugs in lower-grade gliomas. Genomics. 2020;112(5):2990-3002.
16. Zhang J, Wang P, Wan L, Xu S, Pang D. The emergence of noncoding RNAs as Heracles in autophagy. Autophagy. 2017;13(6):1004-24.

17. Ebadi N, Ghafouri-Fard S, Taheri M, Arsang-Jang S, Parsa SA, Omrani MD. Dysregulation of autophagy-related lncRNAs in peripheral blood of coronary artery disease patients. Eur J Pharmacol. 2020;867:172852.
18. Khorkova O, Hsiao J, Wahlestedt C. Basic biology and therapeutic implications of lncRNA. Advanced Drug Delivery Reviews. 2015;87:15-24.
19. Chen C-L, Tseng Y-W, Wu J-C, Chen G-Y, Lin K-C, Hwang S-M, et al. Suppression of hepatocellular carcinoma by baculovirus-mediated expression of long non-coding RNA PTENP1 and MicroRNA regulation. Biomaterials. 2015;44:71-81.
20. Zhang Y, Li X, Zhou D, Zhi H, Wang P, Gao Y, et al. Inferences of individual drug responses across diverse cancer types using a novel competing endogenous RNA network. Molecular oncology. 2018;12(9):1429-46.
21. Gittleman H, Lim D, Kattan MW, Chakravarti A, Gilbert MR, Lassman AB, et al. An independently validated nomogram for individualized estimation of survival among patients with newly diagnosed glioblastoma: NRG Oncology RTOG 0525 and 0825. Neuro-oncology. 2017;19(5):669-77.
22. Goldman M, Craft B, Hastie M, Repečka K, McDade F, Kamath A, et al. The UCSC Xena platform for public and private cancer genomics data visualization and interpretation. bioRxiv. 2019.
23. Bowman RL, Wang Q, Carro A, Verhaak RG, Squatrito M. GlioVis data portal for visualization and analysis of brain tumor expression datasets. Neuro-oncology. 2017;19(1):139-41.
24. Cunningham F, Achuthan P, Akanni W, Allen J, Amode M R, Armean IM, et al. Ensembl 2019. Nucleic acids research. 2018;47(D1):D745-D51.
25. Wright MW. A short guide to long non-coding RNA gene nomenclature. Human genomics. 2014;8:7.
26. Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W, et al. limma powers differential expression analyses for RNA-sequencing and microarray studies. Nucleic Acids Research. 2015;43(7):e47-e.
27. Wang NN, Dong J, Zhang L, Ouyang D, Cheng Y, Chen AF, et al. HAMdb: a database of human autophagy modulators with specific pathway and disease information. Journal of cheminformatics. 2018;10(1):34.
28. Kuhn M. Building predictive models in R using the caret package. Journal of statistical software. 2008;28(5):1-26.
29. Pignatti F, van den Bent M, Curran D, Debruyne C, Sylvester R, Therasse P, et al. Prognostic factors for survival in adult patients with cerebral low-grade glioma. J Clin Oncol. 2002;20(8):2076-84.
30. Yang W, Warrington NM, Taylor SJ, Whitmire P, Carrasco E, Singleton KW, et al. Sex differences in GBM revealed by analysis of patient imaging, transcriptome, and survival data. Sci Transl Med. 2019;11(473).
31. Leu S, von Felten S, Frank S, Vassella E, Vajtai I, Taylor E, et al. IDH/MGMT-driven molecular classification of low-grade glioma is a strong predictor for long-term survival. Neuro-oncology. 2013;15(4):469-79.
32. Ostrom QT, Gittleman H, Truitt G, Boscia A, Kruchko C, Barnholtz-Sloan JS. CBTRUS Statistical Report: Primary Brain and Other Central Nervous System Tumors Diagnosed in the United States in 2011-2015. Neuro-oncology. 2018;20(suppl_4).
33. Stupp R, Brada M, van den Bent MJ, Tonn JC, Pentheroudakis G. High-grade glioma: ESMO Clinical Practice Guidelines for diagnosis, treatment and follow-up. Ann Oncol. 2014;25 Suppl 3:iii93-ii101.
34. Ding Y, Wang X, Pan J, Ji M, Luo Z, Zhao P, et al. Aberrant expression of long non-coding RNAs (lncRNAs) is involved in brain glioma development. Archives of medical science : AMS. 2020;16(1):177-88.
35. Blackshaw S, Harpavat S, Trimarchi J, Cai L, Huang H, Kuo WP, et al. Genomic analysis of mouse retinal development. PLoS Biol. 2004;2(9):E247.
36. Mercer TR, Qureshi IA, Gokhan S, Dinger ME, Li G, Mattick JS, et al. Long noncoding RNAs in neuronal-glial fate specification and oligodendrocyte lineage maturation. BMC Neurosci. 2010;11:14.

37. Tian C, Deng Y, Jin Y, Shi S, Bi H. Long non-coding RNA RNCR3 promotes prostate cancer progression through targeting miR-185-5p. Am J Transl Res. 2018;10(5):1562-70.
38. Zhu B, Zhang S, Meng N, Zhang H, Yuan S, Zhang J. Long non-coding RNA RNCR3 promotes glioma progression involving the Akt/GSK-3β pathway. Oncol Lett. 2019;18(6):6315-22.
39. Xu G, Wang H, Yuan D, Yao J, Meng L, Li K, et al. RUNX1-activated upregulation of lncRNA RNCR3 promotes cell proliferation, invasion, and suppresses apoptosis in colorectal cancer via miR-1301-3p/AKT1 axis in vitro and in vivo. Clin Transl Oncol. 2020.
40. Zhang L, Cao Y, Wei M, Jiang X, Jia D. Long noncoding RNA-RNCR3 overexpression deleteriously affects the growth of glioblastoma cells through miR-185-5p/Krüppel-like factor 16 axis. Journal of Cellular Biochemistry. 2018;119(11):9081-9.
41. Fu Q, Li S, Zhou Q, Yalikun K, Yisireyili D, Xia M. Low LINC00599 expression is a poor prognostic factor in glioma. Bioscience Reports. 2019;39(4).
42. Ku BM, Lee YK, Ryu J, Jeong JY, Choi J, Eun KM, et al. CHI3L1 (YKL-40) is expressed in human gliomas and regulates the invasion, growth and survival of glioma cells. Int J Cancer. 2011;128(6):1316-26.
43. Gorlia T, Wu W, Wang M, Baumert BG, Mehta M, Buckner JC, et al. New validated prognostic models and prognostic calculators in patients with low-grade gliomas diagnosed by central pathology review: a pooled analysis of EORTC/RTOG/NCCTG phase III clinical trials. Neuro-oncology. 2013;15(11):1568-79.
44. Wang Y, Xin S, Zhang K, Shi R, Bao X. Low GAS5 Levels as a Predictor of Poor Survival in Patients with Lower-Grade Gliomas. J Oncol. 2019;2019:1785042.
45. Ostrom QT, Rubin JB, Lathia JD, Berens ME, Barnholtz-Sloan JS. Females have the survival advantage in glioblastoma. Neuro-oncology. 2018;20(4):576-7.
46. Eckel-Passow JE, Lachance DH, Molinaro AM, Walsh KM, Decker PA, Sicotte H, et al. Glioma Groups Based on 1p/19q, IDH, and TERT Promoter Mutations in Tumors. N Engl J Med. 2015;372(26):2499-508.
47. Goodenberger ML, Jenkins RB. Genetics of adult glioma. Cancer Genet. 2012;205(12):613-21.

Legend

Fig. 1 Flow chart of study design.

Fig. 2 Screening of lncRNAs used for constructing the risk signature for lower-grade gliomas (LGG). (a) Principal components analysis of lncRNAs between LGG and normal brain tissues. (b) Volcano plot showed the distribution of differentially expressed lncRNAs (DElncRNAs) between LGG and normal brain tissues. (c) Heatmap exhibited the expression levels of the DElncRNAs. (d) Boxplot showed the expressions of DElncRNAs. The green and red boxes showed the DElncRNA expression in LGG and normal brain tissue, respectively. (e) The networks constructed by autophagy-related DElncRNAs (blue rectangle) and autophagy-related genes (yellow ellipse). Positive and negative Pearson coefficients were illustrated by red line and green line, respectively. The width of the line was proportional to the correlation

Fig. 3 Identification of the autophagy-related differentially expressed lncRNAs (ATG DElncRNA). (a) Log (Lambda) value of the 20 ATG DElncRNAs in least absolute shrinkage and selection operator (LASSO) model. (b) The most appropriate log (Lambda) value in the LASSO model. (c) Multivariate Cox regression analysis was performed and six ATG DElncRNAs (AC021739.2, AL118505.1, LINC00599, AL590666.2, LINC02609, and AL354892.2) were identified to for further construction of the risk signature

Fig. 4 Characteristics of the autophagy-related differentially expressed lncRNAs (ATG DElncRNA) risk signature in the training cohort. (a) lncRNA expression profiles, risk score

distributions and patient survival in the training cohort. (b) Survival curves for high-risk and low-risk groups classified by the risk signature in the training cohort. (c-e) Receiver operating characteristic (ROC) curves for the 1- (c), 3- (d), and 5- (e) year survival according to the ATG DElncRNA risk signature in the training cohort

Fig. 5 Efficacy of the autophagy-related differentially expressed lncRNAs (ATG DElncRNA) risk signature in the validation cohort. (a) LncRNA expression profiles, risk score distributions and patient survival in the validation cohort. (b) Survival curves for high-risk and low-risk groups classified by the risk signature in the validation cohort. (c-e) Receiver operating characteristic (ROC) curves for the 1- (c), 3- (d), and 5- (e) year survival according to the ATG DElncRNA risk signature in the validation cohort

Fig. 6 Assessment of the survival prognostic value of the risk signature, as well as clinical (grade, age, and gender) and molecular variables (IDH status and MGMT status) in LGG patients. (a-e) Kaplan-Meier survival curves showed the survival probabilities for LGG patients by grade (a), IDH status (b), MGMT status (c), age (d), and gender (e). (f-g) Univariate (f) and multivariate (g) Cox regression analyses evaluated the contribution of each variable to LGG survival. (h) Nomogram was developed by integrating the risk signature with age, grade, IDH status and MGMT status for predicting LGG survival. (i-k) ROC curves to evaluate the accuracy of each variable for predicting 1- (i), 3- (j), and 5- (k) year survival were shown with areas under curves (AUCs)

Fig. 7 Evaluation of the performance of the nomogram for survival prediction. (a) ROC curves showed the accuracy of the nomogram for predicting 1-, 3-, and 5- year survival in the training cohort. (b-d) Calibration curves showed the predicted values and the observed values of patient survival at 1- (b), 3- (c) and 5-(d) year in the training cohort. (e) ROC curves showed the accuracy of the nomogram for predicting 1-, 3-, and 5- year survival in the validation cohort. (f-h) Calibration curves showed the predicted values and the observed values of patient survival at 1- (f), 3- (g) and 5-(h) year survival in the validation cohort

Fig. S1 The relationships between the six autophagy-related differentially expressed lncRNAs and their co-expressed genes shown by Sankey diagram. The six autophagy-related lncRNAs were divided into protective and risk lncRNAs according to the hazard ratios

Fig. S2 Functional roles of the risk signature by the gene set enrichment analysis (GSEA). GO analysis showed gene sets related to autophagosome membrane (a), cell matrix adhesion (b) and regulation of cell junction assembly (c) were enriched in LGG patients with the high-risk score, while a gene set related to ribosome assembly (d) was enriched in LGG patients with a low risk score. KEGG showed that gene sets were enriched in the pathway of regulation of autophagy (e), MAPK signaling pathway (f), and ECM receptor interaction (g) in LGG patients with high- risk score, while a gene set related to ribosome (h) was enriched in LGG patients with a low risk score

Table 1. Clinicopathological characteristics of samples in the training cohort and the validation cohort.

Clinicopathological characteristics	Number of samples	
	Trainning cohort(n=252)	Validation cohort(n=252)
WHO Grade		
Grade II	109	116
Grade III	121	119
Unknown	22	17
Age		
Average value	44.01	42.52
Range	14-87	18-75
Unknown	32	25
Gender		
Male	87	109
Female	133	118
Unknown	32	25

Table S1. P value of the autophagy-related DElncRNAs with Kaplan–Meier analysis and Univariate Cox proportional hazards regression

gene	Kmpvalue	Unicox pvalue
AC021739.2	0.000341	3.08E-05
AC053503.1	7.32E-05	3.07E-06
AC093010.3	0.000113	0.001098
AL118505.1	0.000404	3.87E-09
AL121827.2	0.019342	3.05E-05
AL354892.2	9.71E-05	7.67E-08
AL355916.2	0.006057	0.025537
AL590666.2	0.003906	0.018444
LINC00599	0.00955	0.003334
LINC02283	0.002057	7.52E-08
LINC02609	0.009672	0.009522
NEAT1	0.023499	7.96E-05

Table S2. TOP 100 enriched GO functions in high risk group

NAME	SIZE	ES	NES	NOM p-val	FDR q-val
GO_PHOSPHATIDYLINOSITOL_3_5_BISPHOSPHATE_BINDING	28	0.561242	2.035787	0	0.016556
GO_T_CELL_RECEPTOR_SIGNALING_PATHWAY	198	0.571007	2.036016	0.005882	0.016623
GO_CELL_SUBSTRATE_JUNCTION	412	0.481075	2.034863	0	0.016637
GO_CELL_JUNCTION_ASSEMBLY	244	0.510005	2.031892	0	0.016713
GO_REGULATION_OF_GRANULOCYTE_CHEMOTAXIS	42	0.687937	2.036136	0	0.016723
GO_KERATAN_SULFATE_METABOLIC_PROCESS	34	0.703966	2.033007	0	0.016752
GO_NEGATIVE_REGULATION_OF_PEPTIDE_SECRETION	143	0.545066	2.033508	0	0.016756
GO_REGULATION_OF_ACTIN_FILAMENT_BASED_PROCESS	378	0.463275	2.02989	0	0.016795
GO_GOLGI_ASSOCIATED_VESICLE_MEMBRANE	109	0.46485	2.032095	0	0.016807
GO_REGULATION_OF_CELL_SUBSTRA	61	0.558547	2.029092	0	0.016836

TE_JUNCTION_ASSEMBLY					
GO_NEGATIVE_REGULATION_OF_ESTABLISHMENT_OF_PROTEIN_LOCALIZATION	193	0.480561	2.036274	0	0.016851
GO_PHOSPHATIDYLINOSITOL_3_PHOSPHATE_BINDING	39	0.603632	2.09271	0	0.016859
GO_NEGATIVE_REGULATION_OF_CYTOKINE_PRODUCTION	293	0.573397	2.030152	0	0.016872
GO_CELLULAR_RESPONSE_TO_MECHANICAL_STIMULUS	79	0.611249	2.026857	0	0.016873
GO_NEGATIVE_REGULATION_OF_RESPONSE_TO_WOUNDING	89	0.572212	2.024789	0	0.016951
GO_PLATELET_AGGREGATION	60	0.628384	2.027013	0.001866	0.016969
GO_COLLAGEN_CONTAINING_EXTRACELLULAR_MATRIX	407	0.578836	2.025475	0	0.016979
GO_COLLAGEN_BINDING	68	0.678635	2.036319	0	0.017019
GO_CELL_SUBSTRATE_JUNCTION_ORGANIZATION	98	0.556795	2.09355	0.001942	0.017022

GO_RESPONSE_TO_MECHANICAL_STIMULUS	211	0.51361	2.037198	0	0.017033
GO_CELLULAR_RESPONSE_TO_EXTRACELLULAR_STIMULUS	261	0.446498	2.027829	0	0.017058
GO_EPIBOLY	32	0.669239	2.027167	0	0.01711
GO_CELLULAR_GLUCAN_METABOLIC_PROCESS	74	0.551488	2.094342	0	0.017129
GO_GROWTH_FACTOR_BINDING	137	0.602065	2.037377	0	0.017166
GO_ACTIN_FILAMENT_BUNDLE	70	0.572732	1.964708	0	0.017168
GO_PHAGOCYTIC_VESICLE	131	0.604129	2.099622	0	0.017195
GO_POSITIVE_REGULATION_OF_DNA_BINDING_TRANSCRIPTION_FACTOR_ACTIVITY	258	0.478985	1.964856	0.002012	0.017211
GO_REGULATION_OF_HEMOPOIESIS	444	0.510226	1.967051	0.002004	0.017215
GO_RESPONSE_TO_COLD	49	0.5529	1.968744	0	0.017233
GO_ORGANELLE_DISASSEMBLY	99	0.393933	2.000475	0	0.017242
GO_POSITIVE_REGULATION_OF_EPITHELIAL_CELL_MIGRATION	163	0.501388	1.964088	0	0.017246

GO_CELL_MIGRATION_INVOLVED_IN_SPROUTING_ANGIOGENESIS	87	0.569169	1.968201	0.003899	0.017252
GO_ACTIN_CYTOSKELETON_REORGANIZATION	100	0.527181	1.963861	0	0.017254
GO_REGULATION_OF_GLYCOGEN_METABOLIC_PROCESS	34	0.585097	1.96719	0	0.017257
GO_REGULATION_OF_VASOCONSTRICTION	58	0.658629	2.038044	0	0.017258
GO_ORGANELLE_SUBCOMPARTMENT	376	0.402996	1.96774	0	0.017261
GO_GROWTH_FACTOR_RECEPTOR_BINDING	133	0.523218	1.964897	0	0.017279
GO_CELLULAR_TRANSITION_METAL_ION_HOMEOSTASIS	110	0.452484	1.968806	0	0.017296
GO_FAT_SOLUBLE_VITAMIN_METABOLIC_PROCESS	43	0.649742	1.96726	0	0.017303
GO_POSITIVE_REGULATION_OF_MYELOID_CELL_DIFFERENTIATION	92	0.627052	2.066803	0	0.017307
GO_NEGATIVE_REGULATION_OF_SIGN	31	0.580518	1.973933	0.001876	0.017308

AL_TRANSDUCTION_BY_P53_CLASS_MEDIATOR					
GO_NEGATIVE_REGULATION_OF_CYSTEINE_TYPE_ENDOPEPTIDASE_ACTIVITY	89	0.555556	2.000807	0	0.017312
GO_SYNCYTIUM_FORMATION	59	0.585111	1.978205	0	0.017318
GO_REGULATION_OF_ERBB_SIGNALING_PATHWAY	89	0.497655	1.97653	0	0.01732
GO_REGULATION_OF_BICELLULAR_TIGHT_JUNCTION_ASSEMBLY	22	0.750304	1.969197	0	0.017321
GO_PLACENTA_DEVELOPMENT	155	0.518545	2.038586	0	0.017324
GO_HEPARIN_BINDING	170	0.549429	1.964969	0	0.017331
GO_INTRINSIC_COMPONENT_OF_ENDOPLASMIC_RETICULUM_MEMBRANE	157	0.443332	1.970077	0.001992	0.017334
GO_RECEPTOR_CATABOLIC_PROCESS	36	0.543197	1.963301	0.002101	0.017334
GO_REGULATION_OF_SYSTEMIC_ARTERIAL_BLOOD_PRESSURE_MEDIATED_	49	0.621355	1.970477	0	0.017336

BY_A_CHEMICAL_SIGNAL					
GO_TYPE_2_IMMUNE_RESPONSE	38	0.729836	1.974052	0	0.017349
GO_CELLULAR_RESPONSE_TO_OSMOTIC_STRESS	40	0.567791	1.970663	0	0.01735
GO_PHOSPHATIDYLINOSITOL_METABOLIC_PROCESS	176	0.462833	1.982401	0	0.017354
GO_REGULATION_OF_PRODUCTION_OF_MOLECULAR_MEDIATOR_OF_IMMUNE_RESPONSE	140	0.610148	1.965173	0.004008	0.017365
GO_POSITIVE_REGULATION_OF_VACUOLE_ORGANIZATION	15	0.672656	1.98318	0	0.017366
GO_RESPONSE_TO_VIRUS	324	0.549361	1.978416	0.003868	0.017369
GO_LEUKOCYTE_CHEMOTAXIS	219	0.613631	1.974839	0.001927	0.017371
GO_RESPONSE_TO_MOLECULE_OF_BACTERIAL_ORIGIN	348	0.557128	1.980678	0.001927	0.017372
GO_POSITIVE_REGULATION_OF_VASCULATURE_DEVELOPMENT	216	0.586325	2.097044	0	0.017375

GO_RESPONSE_TO_PEPTIDE_HORMONE	433	0.408034	1.969336	0	0.017377
GO_PLATELET_ALPHA_GRANULE_LUMEN	66	0.654564	1.976582	0	0.017379
GO_REGULATION_OF_TUMOR_NECROSIS_FACTOR_MEDIATED_SIGNALING_PATHWAY	57	0.571639	2.001105	0.003929	0.017379
GO_CELL_ADHESION_MEDIATOR_ACTIVITY	60	0.591884	1.980116	0	0.017382
GO_REGULATION_OF_RESPONSE_TO_CYTOKINE_STIMULUS	183	0.595775	2.094889	0	0.017387
GO_RHO_PROTEIN_SIGNAL_TRANSDUCTION	203	0.488859	1.970913	0.003922	0.017389
GO_POST_TRANSLATIONAL_PROTEIN_MODIFICATION	363	0.38525	1.982594	0	0.01739
GO_RAS_PROTEIN_SIGNAL_TRANSDUCTION	446	0.433282	1.977478	0.001961	0.01739
GO_POSITIVE_REGULATION_OF_MAP_KINASE_ACTIVITY	253	0.462041	1.981245	0	0.017399
GO_EXTRACELLULAR_STRUCTURE_O	370	0.589636	2.021354	0	0.017403

RGANIZATION					
GO_REGULATION_OF_SYSTEMIC_ARTERIAL_BLOOD_PRESSURE	92	0.562044	1.975784	0	0.017405
GO_POSITIVE_REGULATION_OF_COLD_INDUCED_THERMOGENESIS	97	0.520958	1.986694	0	0.01741
GO_EPITHELIAL_CELL_APOPTOTIC_PROCESS	108	0.547505	1.969491	0.003854	0.017412
GO_POSITIVE_REGULATION_OF_SMOOTH_MUSCLE_CELL_MIGRATION	44	0.646477	1.979135	0	0.017415
GO_REGULATION_OF_EXTRINSIC_APOPTOTIC_SIGNALING_PATHWAY_VIA_DEATH_DOMAIN_RECEPTORS	58	0.637008	1.974223	0	0.017416
GO_RESPONSE_TO_STEROID_HORMONE	383	0.438781	1.988199	0	0.017422
GO_PEPTIDASE_REGULATOR_ACTIVITY	221	0.525822	1.965861	0	0.017425
GO_RUFFLE	170	0.487888	1.9749	0.003891	0.017426
GO_MICROVILLUS	84	0.536131	1.986188	0	0.017428
GO_INTRACELLULAR_REC	276	0.426416	1.972906	0	0.017429

EPTOR_SIGNALING_PATHWAY					
GO_REGULATION_OF_LIPID_METABOLIC_PROCESS	410	0.422613	1.980725	0	0.01743
GO_ENZYME_INHIBITOR_ACTIVITY	372	0.441698	1.987476	0	0.01743
GO_ACUTE_INFLAMMATORY_RESPONSE	107	0.642459	1.965262	0	0.017431
GO_INOSITOL_PHOSPHATE_MEDIATED_SIGNALING	56	0.555441	1.972326	0	0.017431
GO_LIPID_DROPLET	82	0.515753	2.065356	0	0.017434
GO_LIPID_TRANSPORTER_ACTIVITY	122	0.515524	1.979408	0	0.017435
GO_REGULATION_OF_T_CELL_RECEPTOR_SIGNALING_PATHWAY	39	0.704616	1.981707	0	0.017437
GO_INTEGRIN_BINDING	135	0.617678	2.03874	0	0.017437
GO_AMINOGLYCAN_METABOLIC_PROCESS	168	0.535746	1.98328	0	0.017443
GO_REGULATION_OF_GLUCAN_BIOSYNTHETIC_PROCESS	28	0.576106	1.97848	0	0.017447
GO_AZUROPHIL_GRANULE_LUMEN	90	0.594462	1.975098	0.001957	0.017447

GO_NEGATIVE_REGULATION_OF_SECRETION	232	0.509244	1.987058	0	0.017448
GO_NEGATIVE_REGULATION_OF_CYTOKINE_SECRETION	74	0.675757	2.067154	0	0.01745
GO_LAMELLIPODIUM	192	0.489031	1.981371	0	0.01745
GO_VASCULAR_ENDOTHELIAL_GROWTH_FACTOR_RECEPTOR_SIGNALING_PATHWAY	93	0.520131	1.970931	0.00381	0.01745
GO_TISSUE_MIGRATION	339	0.474941	1.961592	0	0.017453
GO_RESPONSE_TO_ESTROGEN	72	0.56475	2.018291	0	0.017457
GO_POSITIVE_REGULATION_OF_PROTEIN_LOCALIZATION_TO_CELL_SURFACE	19	0.639788	1.971108	0	0.017458
GO_NEGATIVE_REGULATION_OF_HYDROLASE_ACTIVITY	449	0.433331	1.976588	0	0.017459
GO_POSITIVE_REGULATION_OF_LEUKOCYTE_CHEMOTAXIS	86	0.635829	2.018765	0.001873	0.017466
GO_FICOLIN_1_RICH_GRANULE_MEMBRANE	60	0.659707	2.001215	0	0.017468

Table S3. Enriched GO functions in low risk group

NAME	SIZE	ES	NES	NOM p-val	FDR q-val
GO_PROTEIN_TARGETING_TO_MEMBRANE	191	-0.48049	-2.24347	0.003992	0.01843
GO_CYTOPLASMIC_TRANSLATION	98	-0.61306	-2.27214	0	0.022863
GO_RIBOSOME_ASSEMBLY	62	-0.66049	-2.10312	0	0.0247
GO_NUCLEAR_TRANSCRIBED_MRNA_CATABOLIC_PROCESS	207	-0.55866	-2.07574	0.001873	0.024959
GO_RIBONUCLEOPROTEIN_COMPLEX_SUBUNIT_ORGANIZATION	192	-0.53448	-2.09275	0	0.025232
GO_RIBOSOME	228	-0.6062	-2.10696	0.004	0.026858
GO_RIBOSOMAL_LARGE_SUBUNIT_BIOGENESIS	68	-0.68735	-2.07719	0	0.027156
GO_SPLICEOSOMAL_COMPLEX_ASSEMBLY	56	-0.63149	-2.13051	0	0.02803
GO_MATURATION_OF_LSU_RRNA	21	-0.73033	-2.06032	0	0.028154
GO_RIBOSOMAL_SMALL_SUBUNIT_BIOGENESIS	68	-0.66741	-2.11472	0	0.028521
GO_SNRNA_PROCESSING	36	-0.66135	-2.0477	0	0.030251

GO_TRANSLATIONAL_INITIATION	192	-0.55057	-2.13742	0.001905	0.032196
GO_ESTABLISHMENT_OF_PROTEIN_LOCALIZATION_TO_MEMBRANE	321	-0.37144	-2.017	0.002037	0.037769
GO_NUCLEAR_TRANSCRIBED_MRNA_CATABOLIC_PROCESS_NONSENSE_MEDIATED_DECAY	120	-0.68894	-2.0227	0.001942	0.038108
GO_RIBONUCLEOPROTEIN_COMPLEX_BIOGENESIS	417	-0.54035	-2.0088	0	0.039489
GO_POLYSOME	73	-0.60048	-2.13994	0.001934	0.041744
GO_MRNA_SPLICE_SITE_SELECTION	30	-0.66613	-1.9933	0	0.042847
GO_LARGE_RIBOSOMAL_SUBUNIT	117	-0.66087	-1.99741	0.005941	0.043215
GO_RIBOSOMAL_SUBUNIT	186	-0.64563	-1.98322	0.008114	0.043735
GO_MATURATION_OF_SSU_RRNA_FROM_TRICISTRONIC_RRNA_TRANSCRIPT_SSU_RRNA_5_8S_RRNA_LSU_RRNA	35	-0.66001	-1.9859	0	0.044647
GO_SNRNA_3_END_PROCESSING	30	-0.64969	-1.96689	0.002	0.045122

GO_SMALL_SUBUNIT_PROCESSOME	38	-0.66466	-1.97394	0	0.045952
GO_RIBOSOMAL_LARGE_SUBUNIT_ASSEMBLY	29	-0.73032	-1.96923	0	0.045967
GO_RIBOSOME_BIOGENESIS	289	-0.55207	-1.93556	0.005837	0.049691
GO_NBAF_COMPLEX	15	-0.72875	-1.95412	0	0.049698
GO_CYTOSOLIC_RIBOSOME	104	-0.71368	-1.92759	0.006048	0.049911
GO_RIBOSOMAL_SMALL_SUBUNIT_ASSEMBLY	19	-0.71972	-1.93693	0.001931	0.050372
GO_DNA_HELICASE_COMPLEX	15	-0.70093	-1.92901	0	0.050446
GO_NUCLEAR_EXOSOME_RNASE_COMPLEX	16	-0.76396	-1.93022	0	0.051276
GO_MATURATION_OF_SSU_RRNA	47	-0.62708	-1.91035	0.001942	0.051491
GO_INO80_TYPE_COMPLEX	25	-0.62175	-1.9034	0.003937	0.051517
GO_SMALL_RIBOSOMAL_SUBUNIT	73	-0.62197	-1.91461	0.01004	0.051568
GO_CYTOSOLIC_SMALL_RIBOSOMAL_SUBUNIT	44	-0.72424	-1.90546	0.004016	0.051658
GO_SNRNA_METABOLIC_PROCESS	45	-0.60853	-1.93777	0.003937	0.051679
GO_PROTEIN_LOCALIZATI	136	-0.5312	-1.91682	0.015748	0.0518

ON_TO_ENDOPLASMIC_RETICULUM					
GO_RRNA_METABOLIC_PROCESS	221	-0.54191	-1.91131	0.007752	0.052155
GO_REGULATION_OF_MRNA_POLYADENYLATION	17	-0.61744	-1.9185	0	0.052181
GO_ESCRT_COMPLEX	26	-0.50998	-1.92073	0.007952	0.052369
GO_ESTABLISHMENT_OF_PROTEIN_LOCALIZATION_TO_ENDOPLASMIC_RETICULUM	111	-0.6477	-1.89939	0.007905	0.052712
GO_PRERIBOSOME	77	-0.62277	-1.90574	0.007813	0.052732
GO_TRANSLATIONAL_ELONGATION	133	-0.46319	-1.93911	0.033268	0.052746
GO_CYTOPLASMIC_TRANSLATIONAL_INITIATION	31	-0.58886	-1.89717	0.001883	0.052782
GO_NCRNA_PROCESSING	374	-0.52259	-1.89137	0.009766	0.052823
GO_PROTEIN_TARGETING	425	-0.30207	-1.89203	0.003968	0.0536
GO_RNA_SPLICING_VIA_TRANSESTERIFICATION_REACTIONS	343	-0.45807	-1.93986	0.009434	0.054147
GO_TRANSLATION_INITIATION_FACTOR_ACTIVITY	51	-0.4763	-1.8838	0.007859	0.054715

GO_POLYSOMAL_RIBOSOME	32	-0.7682	-1.89204	0.00198	0.054785
GO_U2_TYPE_SPLICEOSOMAL_COMPLEX	93	-0.53824	-1.88524	0.005906	0.055112
GO_HISTONE_H3_ACETYLATION	59	-0.50965	-1.88061	0.005803	0.055415
GO_STRUCTURAL_CONSTITUENT_OF_RIBOSOME	162	-0.67042	-1.94065	0.010246	0.055611
GO_TRANSLATION_PREINITIATION_COMPLEX	18	-0.7125	-1.8672	0	0.061622
GO_VIRAL_GENE_EXPRESSION	192	-0.52075	-1.86758	0.015267	0.062602
GO_NUCLEAR_TRANSCRIBED_MRNA_CATABOLIC_PROCESS_EXONUCLEOLYTIC	35	-0.6123	-1.85864	0.007797	0.064142
GO_90S_PREIRIBOSOME	32	-0.61128	-1.85978	0.007767	0.064484
GO_VIRION_ASSEMBLY	39	-0.43279	-1.85456	0.00789	0.065451
GO_RESPIRATORY_CHAIN_COMPLEX_IV_ASSEMBLY	24	-0.57034	-1.85277	0.019763	0.065491
GO_TRANSLATION_ELONGATION_FACTOR_ACTIVITY	20	-0.57385	-1.85997	0.012	0.065563
GO_SMALL_NUCLEAR_RI	66	-0.53248	-1.85095	0.007874	0.065586

BONUCLEOPROTEIN_COMPLEX					
GO_TRANSLATION_FACTOR_ACTIVITY_RNA_BINDING	85	-0.41632	-1.84158	0.019417	0.068772
GO_EXORIBONUCLEASE_COMPLEX	26	-0.6274	-1.84234	0.001949	0.069289
GO_PROTEIN_MATURATION_BY_IRON_SULFUR_CLUSTER_TRANSFER	15	-0.64484	-1.84377	0.004057	0.069346
GO_RRNA_BINDING	62	-0.5234	-1.83846	0.02947	0.069931
GO_CHAPERONE_COMPLEX	22	-0.52839	-1.82641	0.009434	0.076538
GO_MRNA_MODIFICATION	22	-0.62497	-1.82752	0.001972	0.076843
GO_FORMATION_OF_CYTOPLASMIC_TRANSLATION_INITIATION_COMPLEX	16	-0.70601	-1.82141	0	0.07933
GO_COTRANSLATIONAL_PROTEIN_TARGETING_TO_MEMBRANE	99	-0.72461	-1.81843	0.005988	0.08072
GO_RNA_SPLICING	431	-0.41768	-1.81217	0.024904	0.08327
GO_EUKARYOTIC_48S_PREINITIATION_COMPLEX	15	-0.73597	-1.81246	0	0.084235
GO_MITOCHONDRIAL_RESPIRATORY_	95	-0.50974	-1.80943	0.044444	0.084279

CHAIN_COMPLEX_ASSEMBLY					
GO_IRON_SULFUR_CLUSTER_ASSEMBLY	22	-0.54778	-1.80088	0.022774	0.089932
GO_NUCLEOID	42	-0.52355	-1.79849	0.019455	0.090717
GO_MRNA_PROCESSING	498	-0.40716	-1.78769	0.045627	0.092635
GO_POSTSYNAPTIC_SPECIALIZATION_ORGANIZATION	34	-0.67327	-1.78794	0.015968	0.093654
GO_CYTOCHROME_COMPLEX_ASSEMBLY	34	-0.49464	-1.78917	0.032	0.093902
GO_TRANSLATION_REGULATOR_ACTIVITY_NUCLEIC_ACID_BINDING	109	-0.3963	-1.7904	0.013944	0.094174
GO_NCRNA_METABOLIC_PROCESS	450	-0.46841	-1.79178	0.015779	0.094191
GO_MRNA_CIS_SPLICING_VIA_SPLICEOSOME	34	-0.54017	-1.79253	0.009634	0.09475
GO_EUKARYOTIC_TRANSLATION_INITIATION_FACTOR_3_COMPLEX	16	-0.68698	-1.77829	0.003846	0.099917
GO_CYTOSOLIC_LARGE_RIBOSOMAL_SUBUNIT	57	-0.75558	-1.76673	0.004008	0.109939

GO_HISTONE_DEUBIQUITINATION	22	-0.60087	-1.76352	0.011429	0.111796
GO_RNA_POLYADENYLATION	46	-0.49275	-1.75518	0.023438	0.116335
GO_SM_LIKE_PROTEIN_FAMILY_COMPLEX	77	-0.50677	-1.75781	0.031373	0.116497
GO_MITOCHONDRIAL_GENE_EXPRESSION	160	-0.46009	-1.75576	0.061144	0.117205
GO_INHIBITORY_POSTSYNAPTIC_POTENTIAL	16	-0.75349	-1.74226	0.01046	0.123108
GO_DNA_TEMPLATED_TRANSCRIPTION_ELONGATION	111	-0.43912	-1.74231	0.030769	0.124478
GO_PROTEIN_ACETYLTRANSFERASE_COMPLEX	95	-0.48756	-1.74493	0.035294	0.124544
GO_PEPTIDYL_LYSINE_TRIMETHYLATION	41	-0.53053	-1.74333	0.023392	0.124847
GO_EXON_EXON_JUNCTION_COMPLEX	20	-0.56062	-1.73849	0.011742	0.125733
GO_VIRAL_BUDDING	25	-0.48639	-1.74507	0.024145	0.125902
GO_ATPASE_COMPLEX	82	-0.47688	-1.73439	0.043738	0.127328
GO_NCRNA_3_END_PROCESSING	48	-0.5108	-1.73567	0.022	0.127389

GO_EXCITATORY_SYNAPSE_ASSEMBLY	26	-0.68787	-1.73277	0.016	0.127698
GO_RNA_CAPPING	34	-0.48383	-1.72645	0.019646	0.132077
GO_SPLICEOSOMAL_COMPLEX	185	-0.47184	-1.72654	0.043478	0.133374
GO_REGULATION_OF_TRANSCRIPTION_ELONGATION_FROM_RNA_POLYMERASE_II_PROMOTER	31	-0.51849	-1.7158	0.033465	0.141145
GO_MITOCHONDRIAL_RNA_PROCESSING	16	-0.62328	-1.7162	0.007634	0.142233
GO_NEURON_CELL_CELL_ADHESION	16	-0.73569	-1.70679	0.004049	0.14791
GO_SAGA_TYPE_COMPLEX	28	-0.50231	-1.70685	0.030361	0.149355
GO_U1_SNRNP	21	-0.59628	-1.7079	0.019646	0.149486
GO_RNA_POLYMERASE_II_HOLOENZYME	81	-0.41571	-1.7033	0.054717	0.15076
GO_CATALYTIC_STEP_2_SPLICEOSOME	86	-0.50966	-1.70033	0.037255	0.153133
GO_REGULATION_OF_EXCITATORY_SYNAPSE_ASSEMBLY	15	-0.74903	-1.69871	0.006122	0.153862
GO_SNORNA_METABOLIC_PROCESS	16	-0.56597	-1.69566	0.019608	0.156082

GO_POSTSYNAPSE_ASSEMBLY	31	-0.63774	-1.68765	0.025	0.16482
GO_RNA_SURVEILLANCE	15	-0.6677	-1.68501	0.011719	0.165182
GO_RNA_3_END_PROCESSING	150	-0.42402	-1.68344	0.061538	0.16572
GO_REGULATION_OF_POSTSYNAPTIC_DENSITY_ORGANIZATION	16	-0.65842	-1.68531	0.020576	0.166421
GO_TRNA_METHYLATION	38	-0.57257	-1.6783	0.034	0.170823
GO_POSTSYNAPTIC_SPECIALIZATION_ASSEMBLY	21	-0.71853	-1.66276	0.026804	0.172121
GO_PRECATALYTIC_SPLICEOSOME	51	-0.56565	-1.66375	0.033865	0.172173
GO_RNA_METHYLTRANSFERASE_ACTIVITY	66	-0.50973	-1.66439	0.045726	0.172833
GO_MITOCHONDRIAL_RESPIRATORY_CHAIN_COMPLEX_IV_ASSEMBLY	20	-0.54388	-1.67438	0.035088	0.173026
GO_PEPTIDYL_LYSINE_ACETYLATION	167	-0.39595	-1.66493	0.04065	0.173565
GO_TRANSCRIPTION_ELONGATION_FROM_RNA_POLYMERASE_II_PROMOTER	83	-0.4032	-1.67495	0.055118	0.173905
GO_INTRACELLULAR_PRO	48	-0.36568	-1.66565	0.035433	0.174093

TEIN_TRANSMEMBRANE_TRANSPORT					
GO_PEPTIDYL_LYSINE_DIMETHYLATION	18	-0.54538	-1.67236	0.009804	0.174156
GO_METHYLTRANSFERASE_COMPLEX	113	-0.48412	-1.66651	0.083495	0.174404
GO_TRNA_THREONYLCARBAMOYLADENOSINE_METABOLIC_PROCESS	16	-0.5861	-1.66691	0.016	0.175365
GO_PRC1_COMPLEX	15	-0.63413	-1.66787	0.015444	0.175509
GO_SNORNA_BINDING	26	-0.52716	-1.66978	0.035225	0.176096
GO_MITOCHONDRIAL_TRANSLATION	134	-0.4493	-1.66822	0.093204	0.176609
GO_REGULATION_OF_TRANSLATIONAL_FIDELITY	17	-0.55445	-1.6549	0.025097	0.178565
GO_TRANSCRIPTION_ELONGATION_FROM_RNA_POLYMERASE_I_PROMOTER	30	-0.51936	-1.65576	0.032319	0.178865
GO_BITTER_TASTE_RECEPTOR_ACTIVITY	22	-0.6223	-1.65666	0.023576	0.179081
GO_MLL1_2_COMPLEX	29	-0.54225	-1.65071	0.052734	0.183087
GO_NEGATIVE_REGULATION_OF_MRNA_SPLICING_	21	-0.52078	-1.64912	0.017578	0.183835

VIA_SPLICEOSOME					
GO_SEH1_ASSOCIATED_COMPLEX	17	-0.53917	-1.64653	0.039526	0.185847
GO_MATURATION_OF_5_8S_RRNA	26	-0.59723	-1.64389	0.028169	0.186425
GO_METHYLATED_HISTONE_BINDING	66	-0.5017	-1.64483	0.064453	0.186605
GO_PROTEIN_PHOSPHATASE_TYPE_2A_COMPLEX	19	-0.50056	-1.64001	0.041152	0.188961
GO_TRNA_MODIFICATION	84	-0.51619	-1.64078	0.04	0.189325
GO_TRANSLATION_REGULATOR_ACTIVITY	139	-0.33466	-1.63855	0.048544	0.189656
GO_ADA2_GCN5_ADA3_TRANSCRIPTION_ACTIVATOR_COMPLEX	15	-0.5503	-1.63122	0.036122	0.194252
GO_IONOTROPIC_GLUTAMATE_RECEPTOR_SIGNALING_PATHWAY	26	-0.62379	-1.63199	0.037549	0.194577
GO_HISTONE_METHYLTRANSFERASE_COMPLEX	86	-0.49439	-1.62358	0.090373	0.195242
GO_PROTEIN_LOCALIZATION_TO_NUCLEOLUS	15	-0.6075	-1.63242	0.031558	0.195428
GO_O_METHYLTRANSFER	23	-0.45779	-1.62212	0.050485	0.195884

ASE_ACTIVITY					
GO_HISTONE_H4_ACETYLATION	64	-0.41691	-1.6288	0.065606	0.19608
GO_TRANSLATIONAL_TERMINATION	104	-0.45739	-1.62373	0.107422	0.196353
GO_SPLICEOSOMAL_TRI_SNRNP_COMPLEX	31	-0.55981	-1.62443	0.060784	0.196743
GO_REGULATION_OF_MRNA_PROCESSING	141	-0.36978	-1.63245	0.068359	0.196874
GO_RNA_METHYLATION	78	-0.48076	-1.6252	0.072978	0.197018
GO_U2_SNRNP	21	-0.52034	-1.6258	0.041339	0.197573
GO_NEGATIVE_REGULATION_OF_UBIQUITIN_DEPENDENT_PROTEIN_CATABOLIC_PROCESS	48	-0.36478	-1.62615	0.043137	0.198445
GO_REGULATION_OF_MRNA_SPLICING_VIA_SPLICEOSOME	102	-0.38697	-1.61917	0.043137	0.198493
GO_MITOCHONDRIAL_TRANSLATIONAL_TERMINATION	89	-0.49126	-1.61595	0.094488	0.201714
GO_RETROGRADE_TRANSPORT_ENDOSOME_TO_GOLGI	87	-0.35713	-1.61477	0.068136	0.202048
GO_N_TERMINAL_PROTEI	29	-0.46342	-1.61196	0.042389	0.204756

N_AMINO_ACID_MODIFICATION					
GO_PROTEIN_DNA_COMPLEX_DISASSEMBLY	19	-0.56443	-1.60893	0.038229	0.205056
GO_REGULATION_OF_CHROMATIN_SILENCING	26	-0.58609	-1.60903	0.039293	0.206287
GO_PEPTIDE_N_ACETYLTRANSFERASE_ACTIVITY	74	-0.48226	-1.60708	0.092338	0.206427
GO_PCG_PROTEIN_COMPLEX	47	-0.48965	-1.60922	0.053571	0.207409
GO_DEACETYLASE_ACTIVITY	45	-0.3911	-1.59965	0.03125	0.208882
GO_RRNA_MODIFICATION	35	-0.50801	-1.59991	0.057769	0.209833
GO_RETROMER_COMPLEX	20	-0.50312	-1.60082	0.059055	0.209925
GO_PRESPLICEOSOME	19	-0.57122	-1.60112	0.03	0.210882
GO_TRNA_BINDING	53	-0.47337	-1.6013	0.069307	0.211932
GO_POSTSYNAPTIC_MEMBRANE_ORGANIZATION	40	-0.4832	-1.60169	0.056112	0.212795
GO_NAD_DEPENDENT_PROTEIN_DEACETYLASE_ACTIVITY	16	-0.50427	-1.5911	0.038986	0.219707
GO_RNA_POLYMERASE_III_ACTIVITY	16	-0.52309	-1.58592	0.050881	0.220402
GO_NEUROMUSCULAR_PR	51	-0.43152	-1.58685	0.035782	0.220451

OCESS_CONTROLLING_BALANCE					
GO_NEGATIVE_REGULATION_OF_RNA_SPLICING	26	-0.46864	-1.58726	0.049213	0.221231
GO_N_ACETYLTRANSFERASE_ACTIVITY	93	-0.4209	-1.58904	0.094	0.221345
GO_APOLIPOPROTEIN_BINDING	17	-0.55518	-1.58728	0.044807	0.222564
GO_ORGANELLAR_RIBOSOME	87	-0.476	-1.57646	0.114286	0.228993
GO_REGULATION_OF_DNA_TEMPLATED_TRANSCRIPTION_ELONGATION	51	-0.44022	-1.57703	0.087475	0.229554
GO_CELL_DIFFERENTIATION_IN_HINDBRAIN	19	-0.63245	-1.57445	0.052314	0.230733
GO_RNA_POLYMERASE_III_COMPLEX	18	-0.51223	-1.57706	0.046875	0.230896
GO_U12_TYPE_SPLICEOSOMAL_COMPLEX	26	-0.53874	-1.57293	0.05098	0.231734
GO_TRNA_METHYLTRANSFERASE_ACTIVITY	34	-0.53994	-1.57736	0.071287	0.231798
GO_UBIQUITIN_DEPENDENT_PROTEIN_CATABOLIC_PROCESS_VIA_THE_MUL	22	-0.44701	-1.57097	0.047619	0.233278

TIVESICULAR_BODY_SORTING_PATHWAY					
GO_PROTEIN_TRANSMEMBRANE_IMPORT_INTO_INTRACELLULAR_ORGANELLE	33	-0.40029	-1.56656	0.064516	0.235804
GO_CEREBELLAR_PURKINJE_CELL_LAYER_DEVELOPMENT	24	-0.493	-1.56676	0.047035	0.236842
GO_TRANSCRIPTION_INITIATION_FROM_RNA_POLYMERASE_I_PROMOTER	37	-0.4606	-1.56746	0.073585	0.237217
GO_PROTEIN_ACYLATION	244	-0.32934	-1.56332	0.067485	0.23928
GO_NEGATIVE_REGULATION_OF_TRANSCRIPTION_REGULATORY_REGION_DNA_BINDING	21	-0.482	-1.56152	0.018036	0.240552
GO_CEREBELLAR_CORTEX_FORMATION	21	-0.58637	-1.55994	0.054902	0.241553
GO_MITOCHONDRIAL_LARGE_RIBOSOMAL_SUBUNIT	57	-0.49032	-1.55512	0.108738	0.247431
GO_POSITIVE_REGULATION_OF_TOR_SIGNALING	38	-0.36738	-1.55186	0.056974	0.248242

GO_CLEAVAGE_INVOLVED_IN_RRNA_PROCESSING	21	-0.5857	-1.55313	0.058939	0.249093
GO_REGULATION_OF_RNA_SPLICING	138	-0.35167	-1.55197	0.083499	0.249457

Table S4. Enriched KEGG pathways in high risk group

NAME	SIZE	ES	NES	NOM p-val	FDR q-val
KEGG_REGULATION_OF_ACTIN_CYTOSKELETON	213	0.532176	2.135438	0.001996	0.018736
KEGG_FOCAL_ADHESION	199	0.60119	2.159939	0	0.025345
KEGG_ECM_RECEPTOR_INTERACTION	84	0.706835	2.00091	0	0.027063
KEGG_GLYCOSAMINOGLYCAN_BIOSYNTHESIS_KERATAN_SULFATE	15	0.791504	2.005785	0	0.030407
KEGG_REGULATION_OF_AUTOPHAGY	35	0.538732	1.814661	0.00611	0.030557
KEGG_TOLL_LIKE_RECEPTOR_SIGNALING_PATHWAY	102	0.552883	1.815774	0.008048	0.031269
KEGG_PEROXISOME	78	0.435385	1.816214	0.006198	0.032209
KEGG_MAPK_SIGNALING_PATHWAY	266	0.432993	1.821241	0.003868	0.032285
KEGG_AMINO_SUGAR_AND_NUCLEOTIDE_SUGAR_METABOLISM	43	0.543255	1.843618	0.001938	0.032562
KEGG_VEGF_SIGNALING_PATHWAY	76	0.44881	1.829905	0.001953	0.032643
KEGG_NATURAL_KILLER_CELL_MEDI	132	0.547422	1.821792	0.009747	0.033062

ATED_CYTOTOXICITY					
KEGG_STARCH_AND_SUCROSE_METABOLISM	52	0.540293	1.796439	0	0.033358
KEGG_INTESTINAL_IMMUNE_NETWORK_FOR_IGA_PRODUCTION	45	0.769699	1.834458	0.003846	0.03362
KEGG_VIRAL_MYOCARDITIS	68	0.591547	1.823457	0.017045	0.033639
KEGG_RENAL_CELL_CARCINOMA	69	0.439259	1.830185	0.008197	0.033711
KEGG_PATHOGENIC_ESCHERICHIA_COLI_INFECTION	56	0.489727	1.844087	0.009709	0.033931
KEGG_B_CELL_RECEPTOR_SIGNALING_PATHWAY	75	0.553006	1.7975	0.029644	0.034212
KEGG_GALACTOSE_METABOLISM	26	0.546625	1.800383	0.005725	0.034318
KEGG_COMPLEMENT_AND_COAGULATION_CASCADES	69	0.68129	1.867047	0.001894	0.034443
KEGG_CYTOKINE_CYTOKINE_RECEPTOR_INTERACTION	263	0.59799	1.871848	0.003759	0.035106
KEGG_TYPE_I_DIABETES_MELLITUS	41	0.744304	1.845382	0.007692	0.035149
KEGG_LEUKOCYTE_TRA	116	0.596425	2.01487	0	0.035255

NSENDOTHELIAL_MIGRATION					
KEGG_SYSTEMIC_LUPUS_ERYTHEMATOSUS	56	0.769012	1.880306	0.003846	0.035923
KEGG_RIG_I_LIKE_RECEPTOR_SIGNALING_PATHWAY	70	0.482615	1.783497	0.003906	0.035972
KEGG_SMALL_CELL_LUNG_CANCER	84	0.544594	1.857676	0.010183	0.036378
KEGG_GLYCOLYSIS_GLUCONEOGENESIS	62	0.465965	1.846569	0.004082	0.036759
KEGG_LYSOSOME	121	0.489369	1.894862	0.011928	0.036949
KEGG_ARRHYTHMOGENIC_RIGHT_VENTRICULAR_CARDIOMYOPATHY_ARVC	74	0.560281	1.871955	0	0.037232
KEGG_AUTOIMMUNE_THYROID_DISEASE	50	0.72231	1.881859	0.003891	0.038122
KEGG_T_CELL_RECEPTOR_SIGNALING_PATHWAY	108	0.558937	1.93024	0.005929	0.038203
KEGG_LEISHMANIA_INFECTION	69	0.707043	1.847399	0.003914	0.038237
KEGG_INSULIN_SIGNALING_PATHWAY	137	0.436948	2.037023	0.00207	0.038649
KEGG_NICOTINATE_AND_	24	0.6386	1.899029	0.002105	0.038769

NICOTINAMIDE_METABOLISM					
KEGG_PENTOSE_PHOSPHATE_PATHWAY	26	0.474623	1.772345	0.00998	0.038914
KEGG_CELL_ADHESION_MOLECULES_CAMS	131	0.615927	1.919338	0.003929	0.039135
KEGG_PATHWAYS_IN_CANCER	324	0.451288	1.938272	0.003929	0.039624
KEGG_GLYCOSAMINOGLYCAN_DEGRADATION	21	0.60801	1.76597	0.013645	0.040456
KEGG_HEMATOPOIETIC_CELL_LINEAGE	85	0.643074	1.760825	0.009542	0.041222
KEGG_APOPTOSIS	87	0.532123	1.899153	0.002004	0.042294
KEGG_FC_GAMMA_R_MEDIATED_PHAGOCYTOSIS	96	0.495443	1.754226	0.026477	0.042473
KEGG_JAK_STAT_SIGNALING_PATHWAY	155	0.558674	1.942243	0.003906	0.043157
KEGG_TIGHT_JUNCTION	132	0.38537	1.733856	0.005837	0.044617
KEGG_PHENYLALANINE_METABOLISM	18	0.633864	1.74082	0.003802	0.045125
KEGG_TYROSINE_METABOLISM	42	0.51101	1.734968	0	0.045161
KEGG_HYPERTROPHIC_C	82	0.500714	1.743835	0.005629	0.04535

ARDIOMYOPATHY_HCM					
KEGG_GNRH_SIGNALING_PATHWAY	101	0.430955	1.727888	0.011299	0.04598
KEGG_ARGININE_AND_PROLINE_METABOLISM	54	0.483954	1.735242	0.005859	0.046049
KEGG_ABC_TRANSPORTERS	44	0.518584	1.713487	0.010438	0.04608
KEGG_PANTOTHENATE_AND_COA_BIOSYNTHESIS	16	0.647349	1.718925	0.015504	0.046206
KEGG_CHEMOKINE_SIGNALING_PATHWAY	188	0.484186	1.72411	0.011788	0.046517
KEGG_PPAR_SIGNALING_PATHWAY	69	0.480392	1.714984	0.011928	0.046545
KEGG_ALLOGRAFT_REJECTION	35	0.800731	1.720984	0.01165	0.046594
KEGG_ADIPOCYTOKINE_SIGNALING_PATHWAY	67	0.431698	1.704126	0.003839	0.047504
KEGG_GRAFT_VERSUS_HOST_DISEASE	37	0.780134	1.705041	0.017208	0.048156
KEGG_PRIMARY_IMMUNODEFICIENCY	35	0.699832	1.694537	0.031068	0.048653
KEGG_GLUTATHIONE_METABOLISM	49	0.466375	1.683739	0.016064	0.049096
KEGG_DILATED_CARDIOMYOPATHY	89	0.481324	1.695542	0.015355	0.049282

KEGG_CYTOSOLIC_DNA_SENSING_PATHWAY	54	0.514581	1.679721	0.018182	0.04943
KEGG_TRYPTOPHAN_METABOLISM	39	0.50625	1.684761	0.012245	0.049647
KEGG_GLYCOSPHINGOLIPID_BIOSYNTHESIS_GANGLIO_SERIES	15	0.609943	1.695923	0.021318	0.050022
KEGG_ANTIGEN_PROCESSING_AND_PRESENTATION	81	0.546542	1.685366	0.042389	0.050338
KEGG_N_GLYCAN_BIOSYNTHESIS	46	0.495394	1.68634	0.026639	0.05081
KEGG_RENIN_ANGIOTENSIN_SYSTEM	17	0.610401	1.662158	0.029644	0.054666
KEGG_VASCULAR_SMOOTH_MUSCLE_CONTRACTION	115	0.424752	1.638976	0.015534	0.063276
KEGG_NEUROTROPHIN_SIGNALING_PATHWAY	126	0.368647	1.630128	0.049281	0.067061
KEGG_NITROGEN_METABOLISM	23	0.569326	1.625885	0.021359	0.067842
KEGG_EPITHELIAL_CELL_SIGNALING_IN_HELICOBACTER_PYLORI_INFECTION	68	0.403531	1.62049	0.024857	0.069348
KEGG_ETHER_LIPID_METABOLISM	33	0.479678	1.610254	0.027613	0.072916

KEGG_ADHERENS_JUNCTION	73	0.421583	1.59993	0.047722	0.074717
KEGG_GLYCOSAMINOGLYCAN_BIOSYNTHESIS_HEPARAN_SULFATE	26	0.511454	1.601347	0.038388	0.075111
KEGG_O_GLYCAN_BIOSYNTHESIS	30	0.529865	1.60279	0.029354	0.075441
KEGG_AXON_GUIDANCE	129	0.425279	1.578991	0.048733	0.083028
KEGG_ASTHMA	28	0.722417	1.580982	0.046	0.083357
KEGG_GLYCEROPHOSPHOLIPID_METABOLISM	77	0.379531	1.575834	0.021526	0.083478
KEGG_PRION_DISEASES	35	0.478762	1.570973	0.054	0.084876
KEGG_FC_EPSILON_RI_SIGNALING_PATHWAY	79	0.417149	1.562325	0.046243	0.087301
KEGG_PANCREATIC_CANCER	69	0.431012	1.563278	0.054393	0.087892
KEGG_ACUTE_MYELOID_LEUKEMIA	57	0.414891	1.55866	0.068136	0.087903
KEGG_TYPE_II_DIABETES_MELLITUS	47	0.477193	1.549571	0.044444	0.091664
KEGG_NOD_LIKE_RECEPTOR_SIGNALING_PATHWAY	62	0.491906	1.544686	0.096045	0.092807
KEGG_PRIMARY_BILE_ACID_BIOSYNTHESIS	16	0.587258	1.532746	0.055441	0.097954

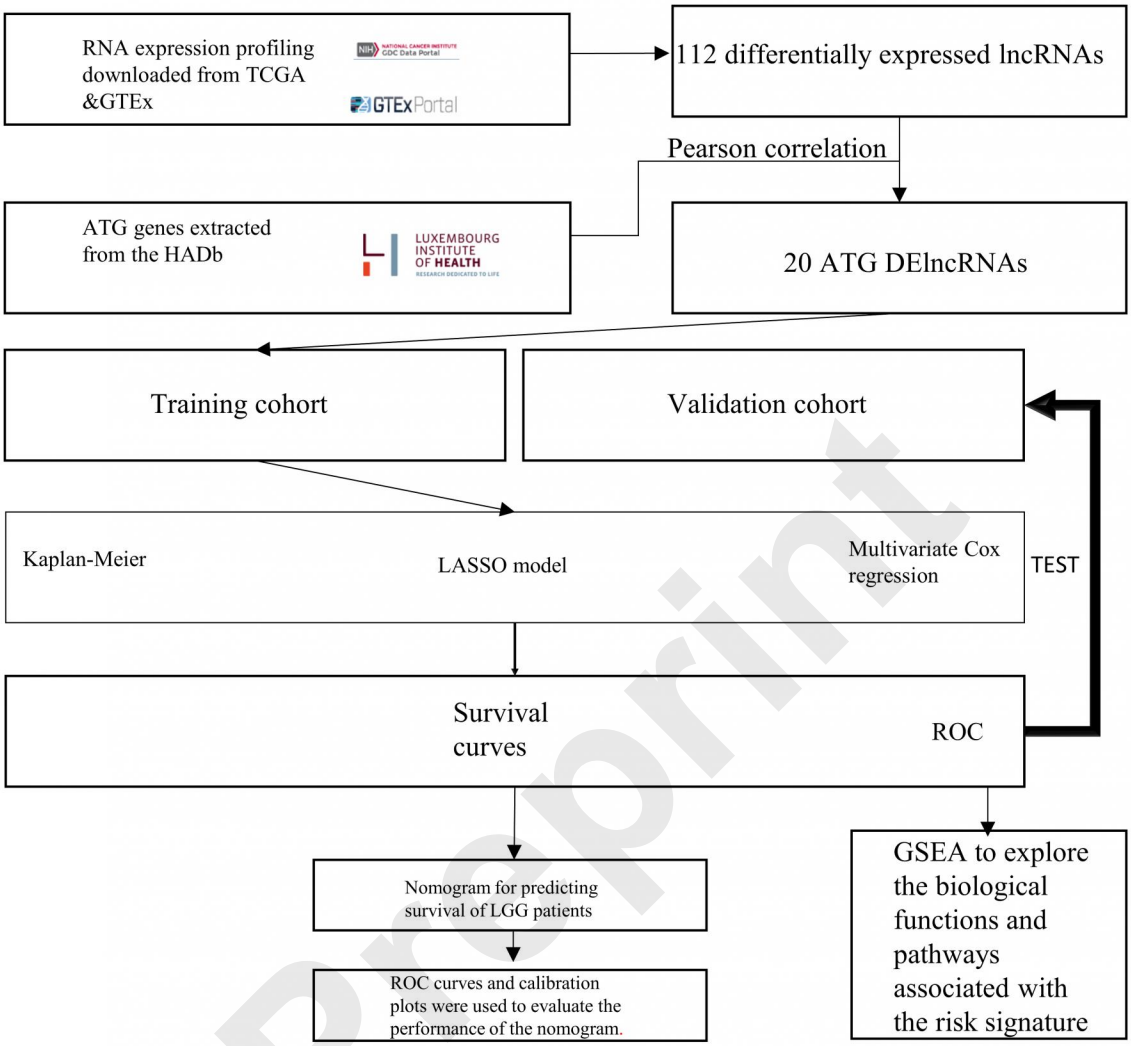
KEGG_ENDOCYTOSIS	181	0.320968	1.507206	0.075547	0.110194
KEGG_PROSTATE_CANCER	89	0.397656	1.509029	0.070833	0.110523
KEGG_ARACHIDONIC_ACID_METABOLISM	58	0.429252	1.489312	0.041096	0.119465
KEGG_BASAL_CELL_CARCINOMA	55	0.446184	1.485905	0.046843	0.120231
KEGG_MELANOGENESIS	101	0.376036	1.478702	0.045833	0.123579
KEGG_GLIOMA	65	0.38673	1.476176	0.084337	0.123878
KEGG_FRUCTOSE_AND_MANNOSE_METABOLISM	33	0.382101	1.469568	0.053254	0.124
KEGG_PHOSPHATIDYLINOSITOL_SIGNALING_SYSTEM	76	0.419983	1.470028	0.075547	0.125026
KEGG_GLYCOSYLPHOSPHATIDYLINOSITOL_GPI_ANCHOR_BIOSYNTHESIS	25	0.450008	1.471894	0.10241	0.125301
KEGG_MELANOMA	71	0.39926	1.464742	0.045908	0.125959
KEGG_BLADDER_CANCER	42	0.426059	1.454506	0.068093	0.127945
KEGG_RETINOL_METABOLISM	64	0.427193	1.457819	0.049242	0.128806
KEGG_FATTY_ACID_METABOLISM	42	0.372328	1.455226	0.063872	0.128814
KEGG_SNARE_INTERACTI	38	0.336303	1.431723	0.080827	0.142214

ONS_IN_VESICULAR_TRANSPORT					
KEGG_CALCIUM_SIGNALING_PATHWAY	177	0.421922	1.427873	0.117424	0.143381
KEGG_DRUG_METABOLISM_CYTOCHROME_P450	71	0.416455	1.422031	0.042969	0.145689
KEGG_GAP_JUNCTION	90	0.371276	1.395919	0.109589	0.160227
KEGG_INOSITOL_PHOSPHATE_METABOLISM	54	0.386076	1.393756	0.130178	0.160269
KEGG_ALPHA_LINOLENIC_ACID_METABOLISM	19	0.467995	1.396366	0.090349	0.161382
KEGG_VALINE_LEUCINE_AND_ISOLEUCINE_DEGRADATION	44	0.370271	1.397755	0.124211	0.161999
KEGG_MATURITY_ONSET_DIABETES_OF_THE_YOUNG	25	0.473197	1.380732	0.078838	0.16866
KEGG_ALDOSTERONE_REGULATED_SODIUM_REABSORPTION	42	0.414853	1.353165	0.131474	0.184841
KEGG_MISMATCH_REPAIR	23	0.540216	1.353886	0.204211	0.186215
KEGG_DRUG_METABOLISM_OTHER_ENZYMES	51	0.392828	1.354961	0.094675	0.187002

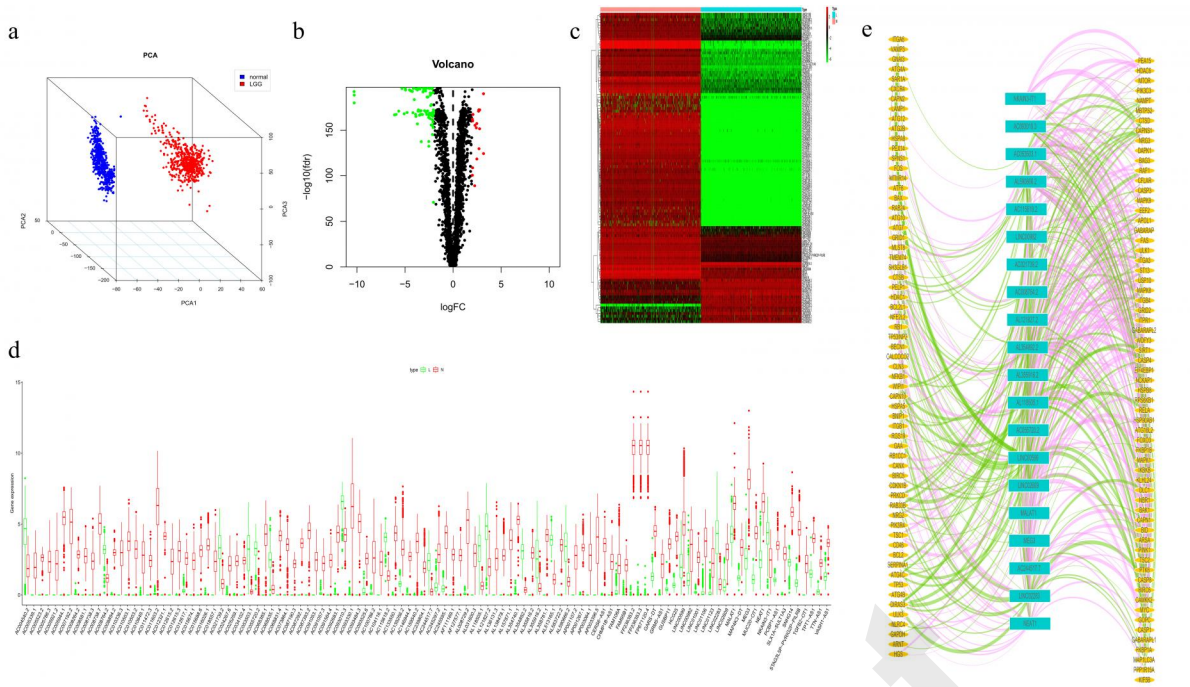
KEGG_BETA_ALANINE_METABOLISM	22	0.383454	1.340556	0.141717	0.19298
KEGG_ERBB_SIGNALING_PATHWAY	87	0.335174	1.326931	0.162162	0.202677
KEGG_PURINE_METABOLISM	154	0.291601	1.324655	0.106509	0.203209
KEGG_WNT_SIGNALING_PATHWAY	150	0.304075	1.306515	0.133056	0.216372
KEGG_METABOLISM_OF_XENOBIOTICS_BY_CYTOCHROME_P450	69	0.380683	1.294113	0.120921	0.224856
KEGG_GLYCOSPHINGOLIPID_BIOSYNTHESIS_LACTO_AND_NEOLACTO_SERIES	26	0.407436	1.28775	0.150476	0.228836
KEGG_OTHER_GLYCAN_DEGRADATION	16	0.484077	1.279699	0.220408	0.233321
KEGG_MTOR_SIGNALING_PATHWAY	52	0.315152	1.2739	0.191057	0.233728
KEGG_HEDGEHOG_SIGNALING_PATHWAY	56	0.36064	1.275143	0.141414	0.234884

Table S5. Enriched KEGG pathways in low risk group

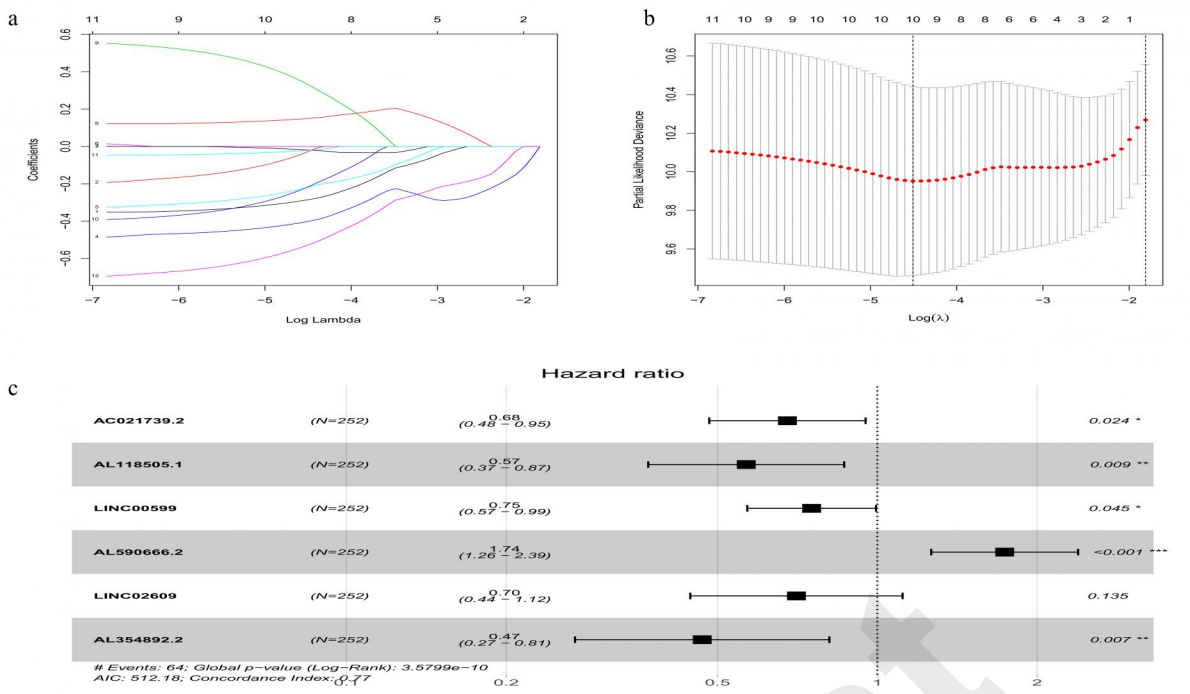
NAME	SIZE	ES	NES	NOM p-val	FDR q-val
KEGG_RIBOSOME	88	-0.79292	-1.87133	0	0.115899
KEGG_RNA_POLYMERASE	29	-0.53494	-1.77905	0.01232	0.126753
KEGG_SPLICEOSOME	127	-0.47047	-1.72958	0.054326	0.128102
KEGG_TERPENOID_BACKBONE_BIOSYNTHESIS	15	-0.68308	-1.68396	0.018987	0.135814
KEGG_SELENOAMINO_ACID_METABOLISM	25	-0.46771	-1.59004	0.052941	0.169369
KEGG_RNA_DEGRADATION	59	-0.46737	-1.59005	0.090361	0.203243



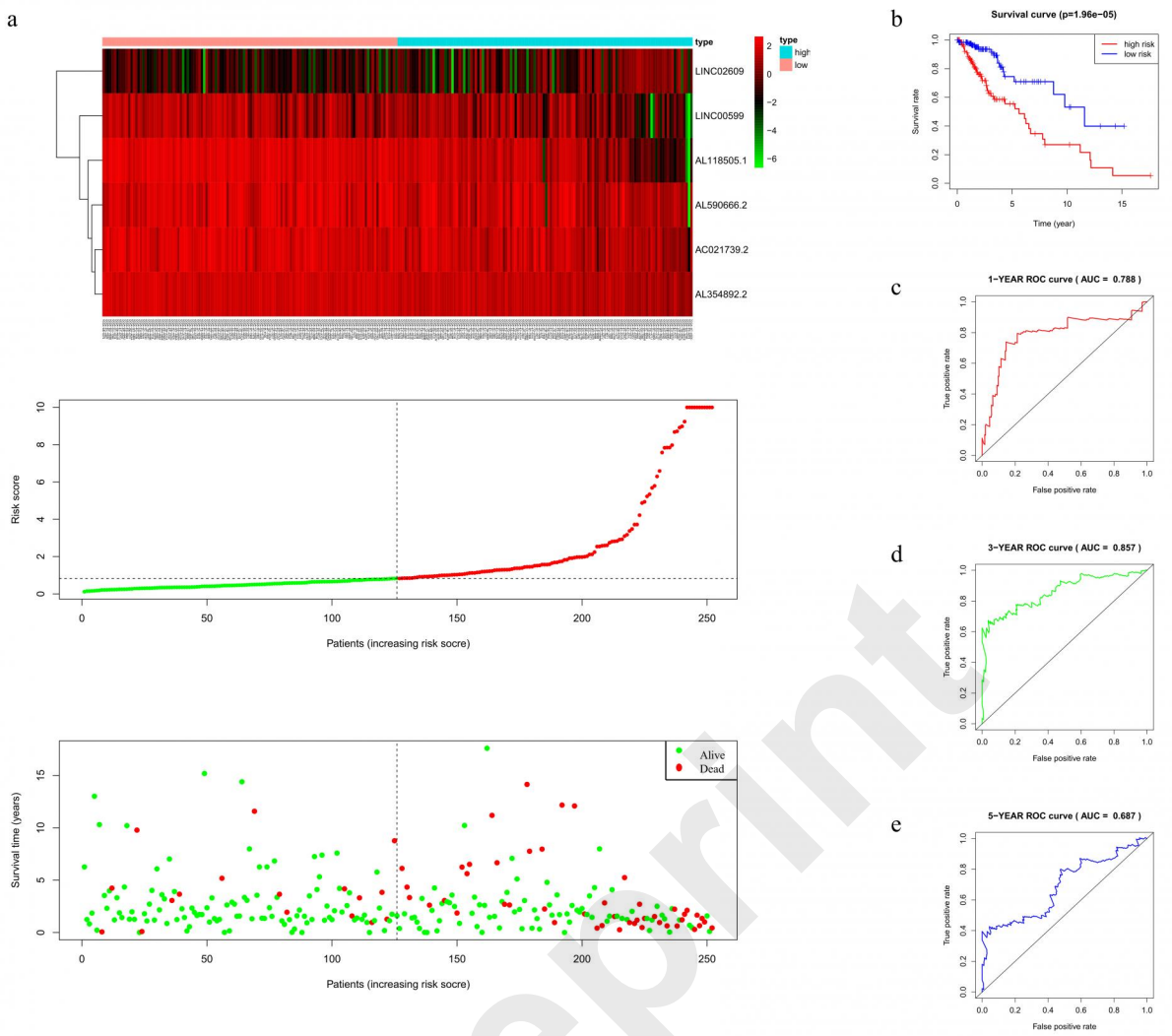
Flow chart of study design.



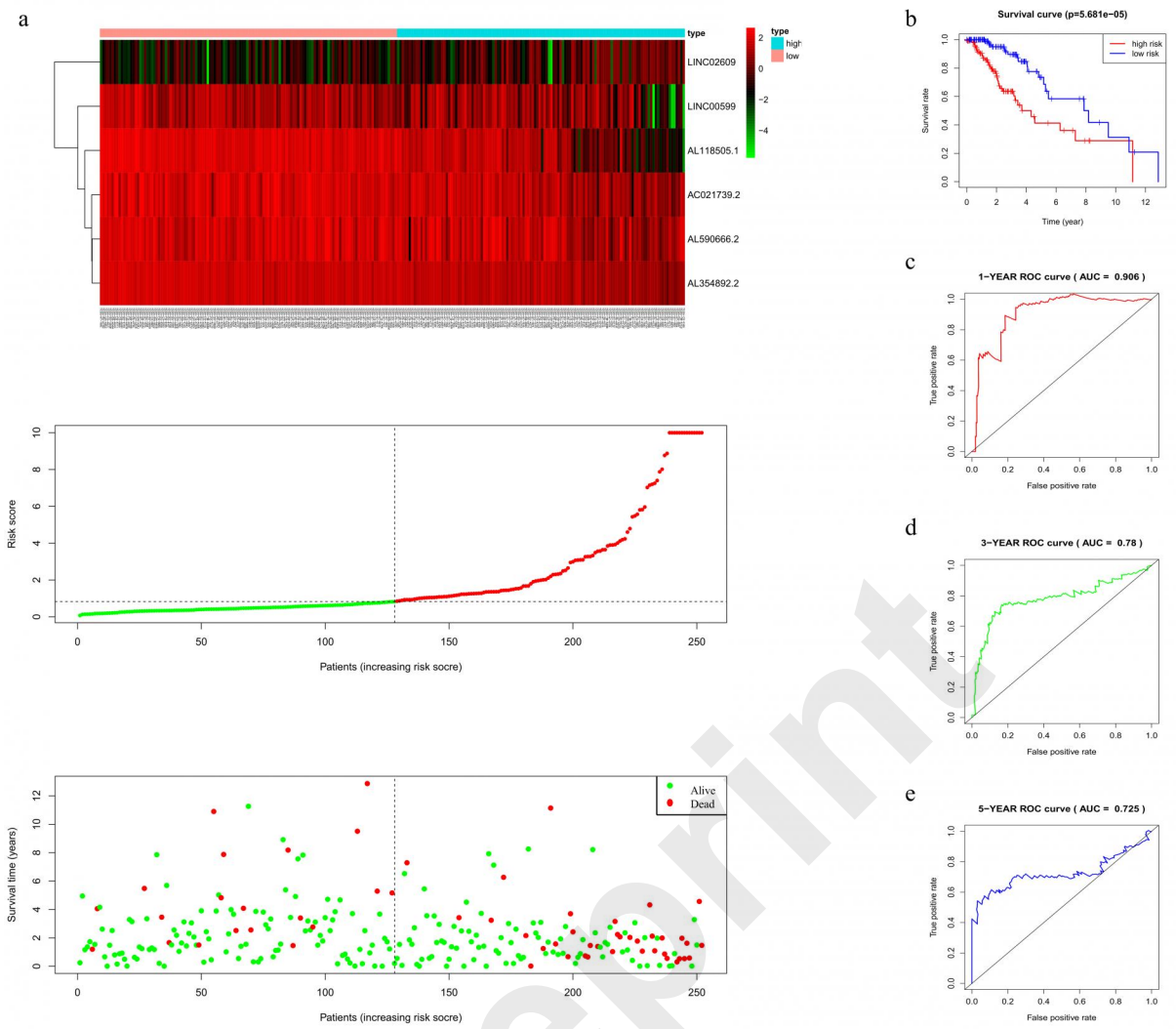
Screening of lncRNAs used for constructing the risk signature for lower-grade gliomas (LGG).



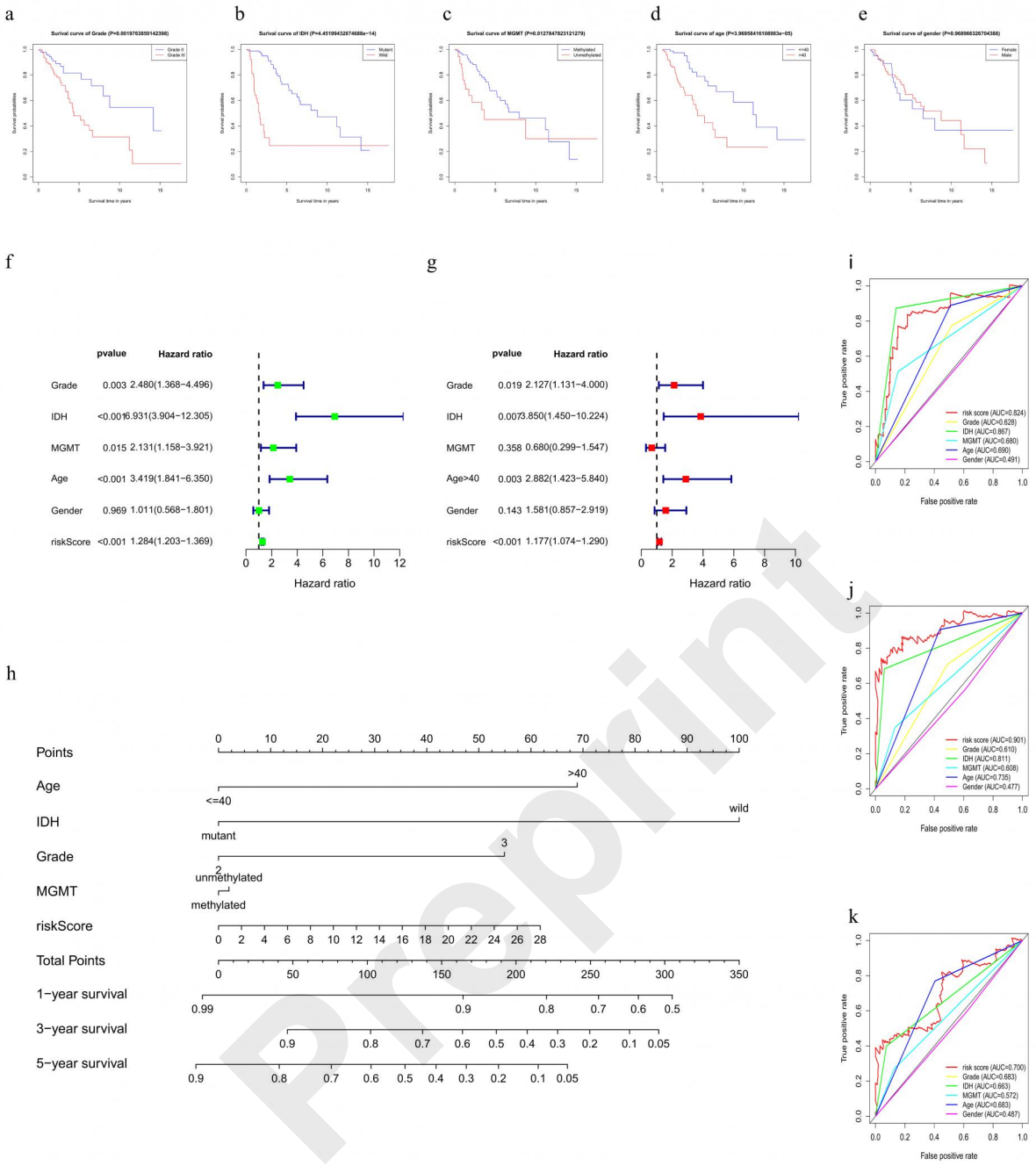
Identification of the autophagy-related differentially expressed lncRNAs (ATG DElncRNA).



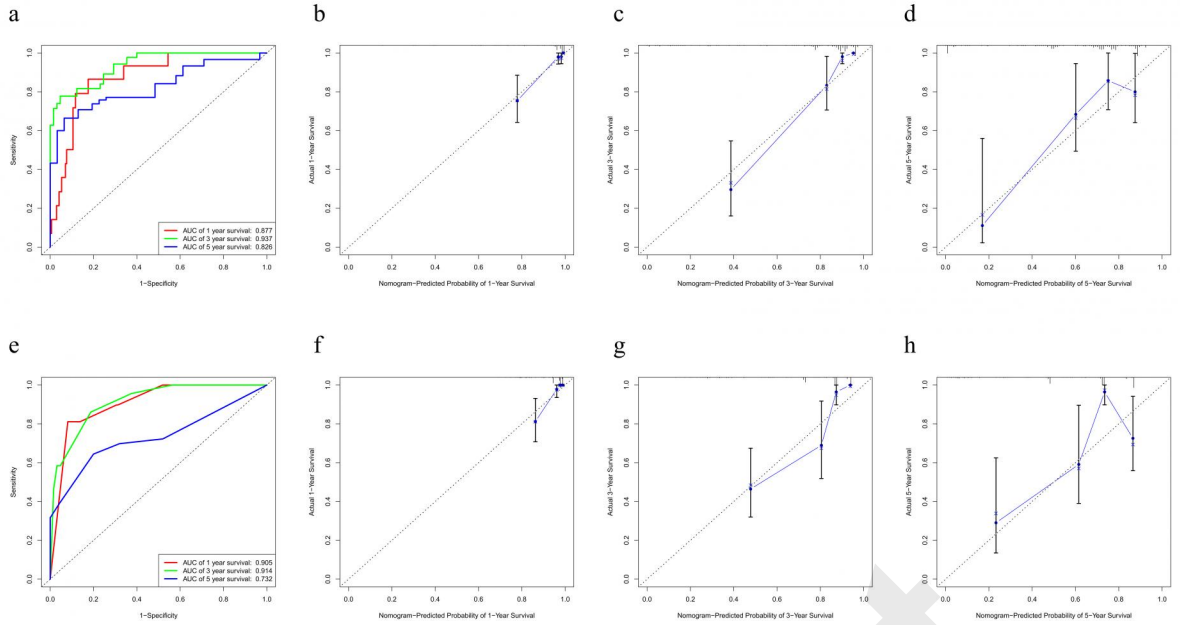
Characteristics of the autophagy-related differentially expressed lncRNAs (ATG DElncRNA) risk signature in the training cohort.



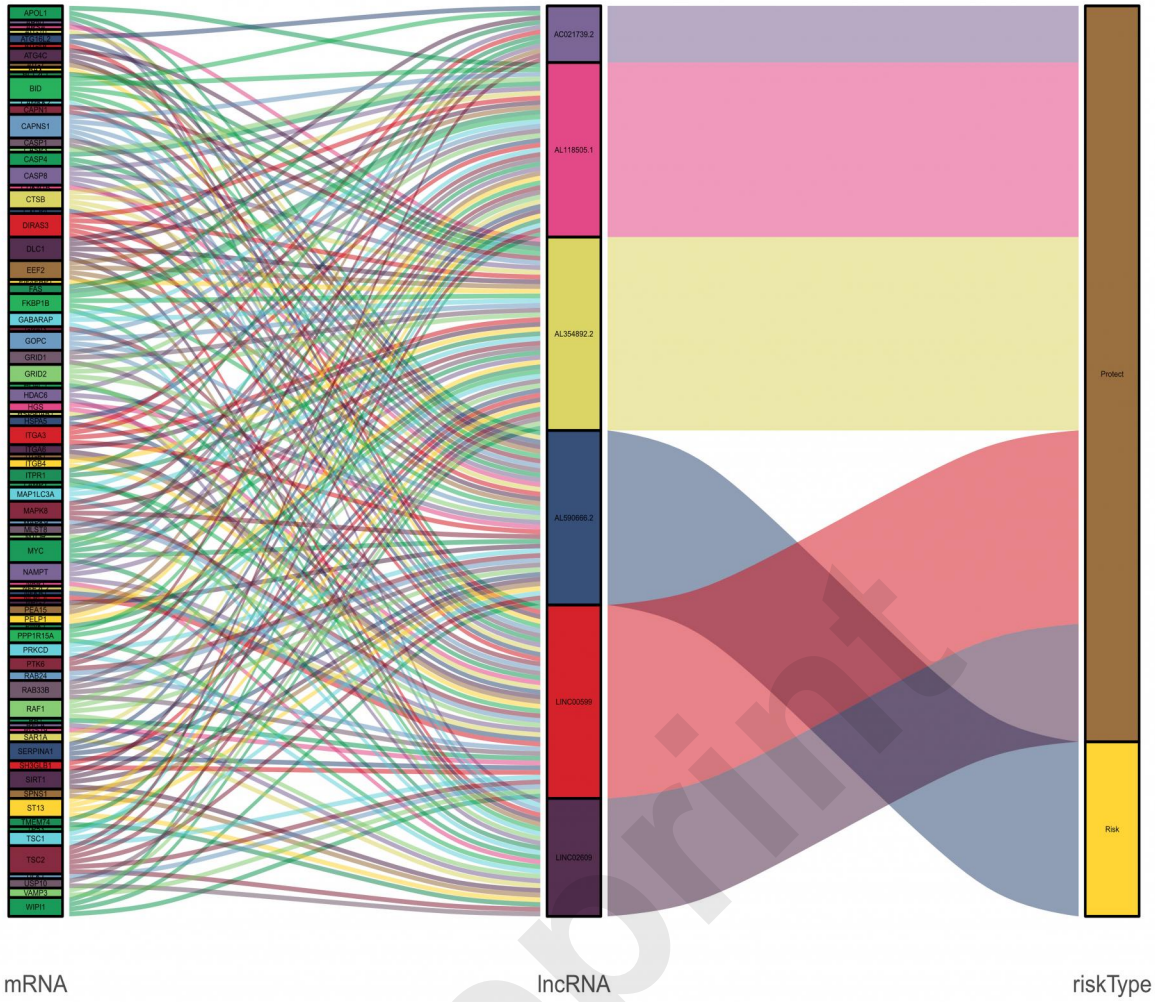
Efficacy of the autophagy-related differentially expressed lncRNAs (ATG DElncRNA) risk signature in the validation cohort.



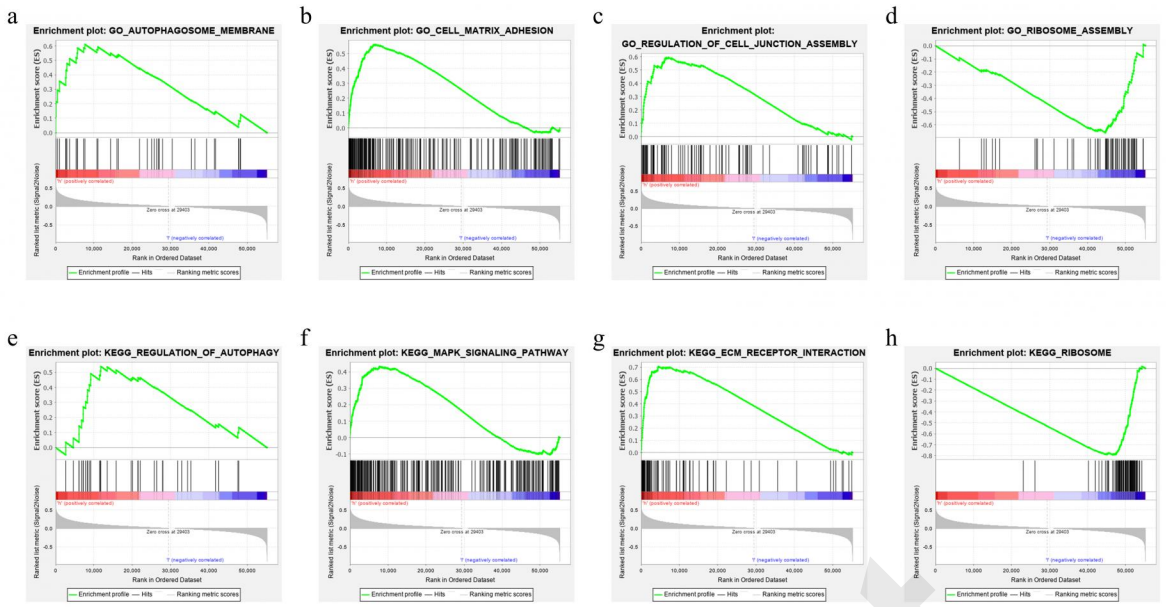
Assessment of the survival prognostic value of the risk signature, as well as clinical (grade, age, and gender) and molecular variables (IDH status and MGMT status) in LGG patients.



Evaluation of the performance of the nomogram for survival prediction.



The relationships between the six autophagy-related differentially expressed lncRNAs and their co-expressed genes shown by Sankey diagram



Functional roles of the risk signature by the gene set enrichment analysis (GSEA)